AUTOMATED KIDNEY SEGMENTATION IN 3D ULTRASOUND IMAGERY, AND ITS APPLICATION IN COMPUTER-ASSISTED TRAUMA DIAGNOSIS

by

Mahdi Marsousi

A thesis submitted in conformity with the requirements for the degree of Doctor of Philosophy
Graduate Department of Electrical and Computer Engineering
University of Toronto

© Copyright 2016 by Mahdi Marsousi
Abstract

Automated Kidney Segmentation in 3D Ultrasound Imagery, and its Application in Computer-Assisted Trauma Diagnosis

Mahdi Marsousi

Doctor of Philosophy

Graduate Department of Electrical and Computer Engineering

University of Toronto, 2016

Due to the limitations of emergency healthcare/technology, blunt abdominal bleeding causes a large number of preventable deaths each year. To save a trauma patient’s life, a rapid diagnosis is required, which is not always available in emergency situations. This is the thesis of this PhD research that a computer-assisted algorithm based on 3D ultrasound imagery, as a portable imaging modality, provides a systematic solution to facilitate rapid diagnosis of trauma patients in emergency situations by first responders (ie. paramedics).

3D ultrasound imagery, which is a portable imaging system, is selected as the preferred imaging modality for trauma diagnosis, because it can be carried to the location of emergency situation. Therefore, by eliminating the need of moving an unstable patient to an imaging room, rapid trauma diagnosis is achievable. Compared to 2D sonography, 3D ultrasound imaging facilitates automated detection and localization of internal organs. This is essential, specifically in the sense that ultrasonographers are not always present at emergency situations. Hence, a computer-assisted solution is essential to guide first responders to perform trauma diagnosis using a 3D ultrasound device.

An abdominal bleeding has a high tendency to align around the right kidney. The right-upper-quadrant view of sonography shows the entire kidney shape, and therefore, it is considered as the most relevant internal view to trauma diagnosis. Paramedics usually lack proper knowledge to find the right-upper-quadrant view, to detect the kidney shape, and to detect an internal bleeding using an ultrasound imaging device. Hence, computer-assisted algorithms are required to perform these tasks. The focus of this thesis is to introduce automated methods to detect and segment the kidney shape. The detected kidney shape will be used for two purposes: (a) it is used to calculate the ultrasound probe’s misalignment with respect to the right-upper-quadrant view, which is used to guide the operator to move the probe toward the correct alignment on the patient’s body; (b) the detected kidney shape is used to initialize the kidney segmentation process, and thereby, an automated kidney segmentation approach is achieved. The kidney segmentation output can be used to automatically detect an internal bleeding.
Acknowledgements

I would like to express my sincere appreciation to my supervisor Professor Konstantinos N. Plataniotis for his support, consultations, and invaluable advices. I believe this PhD research would be impossible to accomplish without having his kind supports. I feel deeply thankful for all his supports during my PhD studies.

I would like to thank both Professor Stergiopoulos and Defense Research and Development Canada for their technical supports on this PhD research. Especially, I would like to thank Prof. Stergiopoulos for sharing information, data set, and resources related to my PhD research. Indeed, their support was very important toward the achievements during my PhD research.

I also would like to thank the department of electrical engineering of the University of Toronto for providing all the resources and financial supports during my PhD program. I acknowledge that the excellent academical environment provided at the University of Toronto has been very influential and positively effective for me to progress in my PhD research.

I also appreciate the helps and collaborations of my colleagues and friends during my PhD studies. Specially, I should thank my lab-mates: Dohyoung Lee, Xingyu Li, Amir Aghaei, and Mohammad-Shahin Mahanta. Moreover, I have especially benefited from the truly professional support from the technical staff, Jim Prall. Furthermore, I would like to thank my friend Bruno Korst for providing great and exciting DSP lab-experiments during the DSP course, ECE431F.

I would like to thank my parents, Batoul Zandkarimi and Rasoul Marsousi, for their support and patience. Also, I would like to express my gratitude to my mother-in-law, Masoumeh Mofidi, my brother-in-law, Mohammad-Malid Mofidi, my sister-in-law, Haleh Mofidi, and her husband, Mohammad Nabizadeh for all their warm and great emotional supports.

Last, but not least, I would like to dedicate this thesis to my family, my wife, Hoda, and my daughter, Zahra, for their love, patience, and great understanding. Their love provided me with a great strength to focus on this thesis.
## Contents

1 Introduction ........................................ 1
1.1 Background ........................................... 2
1.2 Motivation ............................................. 5
1.3 Why 3D Ultrasound Imagery? ......................... 7
1.4 Problem Definition ...................................... 9
1.5 Challenges ............................................. 10
1.6 Volumetric Ultrasound Dataset ......................... 12
  1.6.1 Actual ultrasound volumes (Ds3DUsIs-AUVs) .......... 12
  1.6.2 Ground truth data (Ds3DUsIs-GTs) .................. 15
  1.6.3 Ultrasound volume simulator (Ds3DUsIs-UVS) ......... 16
1.7 Prior Works ............................................ 17
  1.7.1 Internet-based FAST ................................ 18
  1.7.2 Computer-assisted FAST ............................ 19
1.8 Contributions .......................................... 20
1.9 Thesis Organization ..................................... 22

2 Kidney detection ........................................ 23
2.1 Introduction ........................................... 24
2.2 Shape model ............................................ 27
  2.2.1 Introduction ........................................ 27
  2.2.2 Proposed shape model: complex-valued implicit shape model (CVISM) .......... 30
2.3 Speckle reduction ....................................... 35
  2.3.1 Introduction ........................................ 35
  2.3.2 Proposed speckle reduction for 3D ultrasound images:
      mixture of filters based on local variation (MFLV) .................. 40
2.4 Volume enhancement .................................................. 44
  2.4.1 Contrast Enhancement ........................................... 44
  2.4.2 Global-to-local thresholding .................................. 46
2.5 Voxel Classification with SANNs ................................... 49
  2.5.1 Training .......................................................... 52
  2.5.2 Classification .................................................... 54
2.6 Shape-to-volume registration ........................................ 57
  2.6.1 Introduction ..................................................... 57
  2.6.2 Proposed solution .............................................. 58
2.7 Shape-based and atlas-based kidney detection ...................... 63
2.8 Application of kidney detection in computer assisted ultrasonic probe placement for trauma diagnosis ......................... 65
  2.8.1 Introduction ..................................................... 65
  2.8.2 Initial probe placement ....................................... 66
  2.8.3 Correcting probe placement ................................... 67
  2.8.4 Calculating probe misalignment ............................... 67
  2.8.5 Implementation .................................................. 68
2.9 Experimental results .................................................. 69
  2.9.1 Hardware and software setup ................................ 69
  2.9.2 Implementation details ........................................ 69
  2.9.3 3D ultrasound dataset ........................................ 70
  2.9.4 Comparison metrics ........................................... 71
  2.9.5 Evaluating the kidney detection methods .................... 71
    2.9.5.1 Sensitivity analysis to kidney shape misalignment and occlusion ........ 72
    2.9.5.2 Evaluating kidney detection accuracy ...................... 77
    2.9.5.3 Computational time analysis ............................... 81
2.10 Summary and Concluding Remarks .................................. 82
3 Automated kidney segmentation ....................................... 84
  3.1 Introduction ....................................................... 85
  3.2 3D segmentation methods: review .................................. 88
    3.2.1 Deformable models .......................................... 90
      3.2.1.1 Continuous explicit representation ...................... 91
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.2.1.1.1</td>
<td>Parametric model (3D snake):</td>
<td>91</td>
</tr>
<tr>
<td>3.2.1.1.2</td>
<td>Geodesic model:</td>
<td>93</td>
</tr>
<tr>
<td>3.2.1.1.3</td>
<td>B-spline surface:</td>
<td>94</td>
</tr>
<tr>
<td>3.2.1.1.4</td>
<td>Superquadrics:</td>
<td>95</td>
</tr>
<tr>
<td>3.2.1.1.5</td>
<td>Active rays:</td>
<td>98</td>
</tr>
<tr>
<td>3.2.1.1.6</td>
<td>Probabilistic approaches:</td>
<td>100</td>
</tr>
<tr>
<td>3.2.1.2</td>
<td>Discrete explicit representation</td>
<td>104</td>
</tr>
<tr>
<td>3.2.1.2.1</td>
<td>Deformable meshes:</td>
<td>104</td>
</tr>
<tr>
<td>3.2.1.2.2</td>
<td>T-surface:</td>
<td>107</td>
</tr>
<tr>
<td>3.2.1.2.3</td>
<td>Spring level-set:</td>
<td>108</td>
</tr>
<tr>
<td>3.2.1.3</td>
<td>Implicit representation</td>
<td>110</td>
</tr>
<tr>
<td>3.2.1.3.1</td>
<td>Edge-based level-set:</td>
<td>110</td>
</tr>
<tr>
<td>3.2.1.3.2</td>
<td>Level-set without-edge (regional level-set):</td>
<td>112</td>
</tr>
<tr>
<td>3.2.1.3.3</td>
<td>Regional level-set with shape-prior:</td>
<td>114</td>
</tr>
<tr>
<td>3.2.1.3.4</td>
<td>Parametric level-Set:</td>
<td>116</td>
</tr>
<tr>
<td>3.2.1.3.5</td>
<td>Active Surface:</td>
<td>117</td>
</tr>
<tr>
<td>3.2.2</td>
<td>Review of segmentation methods: Conclusion</td>
<td>118</td>
</tr>
<tr>
<td>3.3</td>
<td>Proposed kidney segmentation</td>
<td>120</td>
</tr>
<tr>
<td>3.4</td>
<td>Experimental results</td>
<td>125</td>
</tr>
<tr>
<td>3.4.1</td>
<td>Hardware and software setup</td>
<td>125</td>
</tr>
<tr>
<td>3.4.2</td>
<td>3D ultrasound dataset</td>
<td>125</td>
</tr>
<tr>
<td>3.4.3</td>
<td>Comparison metrics</td>
<td>125</td>
</tr>
<tr>
<td>3.4.4</td>
<td>Parameter analysis</td>
<td>126</td>
</tr>
<tr>
<td>3.4.5</td>
<td>Evaluating kidney segmentation methods</td>
<td>128</td>
</tr>
<tr>
<td>3.5</td>
<td>Summary and Concluding Remarks</td>
<td>131</td>
</tr>
<tr>
<td>4</td>
<td>Conclusions and future work</td>
<td>132</td>
</tr>
<tr>
<td>4.1</td>
<td>Research summary</td>
<td>133</td>
</tr>
<tr>
<td>4.2</td>
<td>Discussions</td>
<td>136</td>
</tr>
<tr>
<td>4.3</td>
<td>Limitations</td>
<td>138</td>
</tr>
<tr>
<td>4.4</td>
<td>Future work</td>
<td>139</td>
</tr>
<tr>
<td>4.5</td>
<td>List of publications</td>
<td>140</td>
</tr>
</tbody>
</table>
List of Tables

1.1 Summarizing advantages and disadvantages of prior arts on computer- and internet-assisted FAST examination. .................................................. 19

2.1 Summarizing advantages and disadvantages of state-of-the-art of kidney detection in 3D ultrasound images. .................................................. 25

2.2 Parts of the proposed shape-based and atlas-based kidney detection methods. ................. 26

2.3 Summarizing advantages and disadvantages of different representations of shape models. . 30

2.4 Summarizing advantages and disadvantages of prior arts in speckle reduction in ultrasound images. ................................................................. 40

2.5 Comparison of the speckle reduction methods, including the proposed method (MFLV), SRAD, and sparse representation via dictionary learning (SR-DL), in terms of quality enhancement based on the $SNR_A$ metric and computational time. ................. 42

2.6 This table provides the list of parameters of the SANNs associated with the MATLAB neural network toolbox. ............................................................ 55

2.7 Implementation methods of the parts of the proposed shape-based and atlas-based kidney detection methods in MATLAB. ................................................. 70

2.8 This table shows mean ($\mu$) and standard deviation ($\sigma$) of estimating the kidney shape deformations, using both the shape-based and atlas-based kidney detection methods. Errors are calculated in voxels for translations and degrees for orientations. ......................... 73

2.9 Comparing accuracy ($ACC_{KD}$), sensitivity ($Sens_{KD}$), and specificity ($Spec_{KD}$) of the kidney detection methods using the evaluation set of actual ultrasound volumes of the healthy volunteers ($Ds3DUsIs-AUVs-ESH$). ......................................................... 78

3.1 Summarizing characteristics, including Manual Initialization (Man. Init.), automated segmentation (Auto. Seg.), methodology, advantages, and disadvantages, of the kidney segmentation methods for 3D ultrasound images. ........................................ 87
3.2 List of parameters of the implemented CVRLS-SP method for segmenting the kidney shape in ultrasound volumes. ......................................................... 127

3.3 Comparing $DSC$ and $ACC$, sensitivity $SENS$, specificity $SPEC$, $HD$ and $MD$ of the segmentation methods: the proposed method of this paper, Marsousi et al-EMBC14 (M-EMBC14), Noll et al., and MRF-AC. $\mu$ and $\sigma$ are mean and standard deviation, respectively. The units of $DSC$, $SENS$ and $SPEC$ are pixels, and the units of both $HD$ and $MD$ are millimeters (mm). ................................................................. 128
List of Figures

1.1 This figure shows a protocol for trauma diagnosis in trauma patients. The focus of this thesis is specified by blue (darker) background color. ........................................ 2

1.2 Displaying the six abdominal views of FAST examination [1]. The region (2) corresponds to the Morison’s pouch (RUQ) view, which is the focus of this PhD thesis. ............... 3

1.3 This figure shows a 2D Morison’s pouch (RUQ) view, and specifies liver, right kidney, and free-fluid regions [1]. ................................................................. 4

1.4 This figure summarizes the motivations of this PhD thesis. ........................................ 7

1.5 This figure shows a 2D ultrasound and a 3D ultrasound image. ..................................... 7

1.6 This figure lists advantages and disadvantages of 3D ultrasound imaging, CT imaging, and 2D imaging for designing a computer-assisted trauma diagnostic solution. ............. 8

1.7 This figure defines the problem to be addressed in this PhD thesis. As shown in figure, the 3D ultrasound (3D US) imaging device is connected to the personal computer using the ultrasound research interface (URI) protocol. ................................. 10

1.8 This figure shows the challenges toward designing the computer-assisted diagnostic system for FAST examination. ................................................................. 11

1.9 This figure shows (a) the curved-array ultrasonic probe (RM6C H48671ZG transducer) for real-time 4D abdominal imaging, (b) the GE Voluson e ultrasound device used to generate actual ultrasound volumes of this thesis, and (c) the 3D scan procedure applied by the ultrasound device to generate volumetric images [2]. ........................ 13

1.10 This figure provides details of the dataset of 3D ultrasound images (Ds3DUsIs) of this thesis. ................................................................. 15

1.11 This figure shows three examples of actual 3D ultrasound images of Ds3DUsIs using 3D planar view. (a) shows a with-kidney image of a healthy volunteer, (b) shows a without-kidney image of a healthy volunteer, and (c) shows a with-kidney image of an abnormal patient with in-body fluid. ........................................ 16
1.12 This figure shows three examples of actual 3D ultrasound images with their ground truths of $Ds3DUsIs$ using 3D planar view. The green regions show the ground truth data.

1.13 This figure shows three examples of simulated ultrasound volumes. $\theta_z$, $\mathbf{X}_c = [x_c, y_c, z_c]$, and $\lambda_{oc}$ are rotation over the z-axis, the center point of the kidney shape, and visibility ratio, respectively.

1.14 This figure shows the contributions of this PhD thesis.

2.1 This figure shows three structural regions of the kidney shape in 3D ultrasound images, including renal medulla, pyelocalyceal system, and kidney capsule.

2.2 This figure shows the block diagram of registering a training shape on the reference shape.

2.3 This figure shows the block diagram of generating the kidney shape model.

2.4 This figure shows 2D slices of the generated regions of the Kidney shape model (CVISM). The right figure shows the renal medulla region, $\psi_{RM}$, and the left figure shows the combination of kidney capsule, $\psi_{KC}$, and pyelocalyceal system, $\psi_{PS}$.

2.5 This figure shows a with-kidney image from the actual ultrasound dataset ($Ds3DUsIs-AUVs$) in the left side, and its image histogram in the right side.

2.6 This figure shows the block diagram of reducing speckle in 3D ultrasound images using the MFLV method. $LPF$ stands for Low-Pass Filtering.

2.7 This figure shows speckle reduction results using three methods, including MFLV (the proposed method), SRAD, and sparse representation via dictionary learning (SR-DL). Results of the SRAD method are for three different number of iterations, including $\#itr : 2$, $\#itr : 4$, and $\#itr : 6$. The bottom row displays zoom-in of the de-speckling result of a part of the kidney shape.

2.8 This figure shows the output of contrast enhancement using the LHE method for different values of $rad_2 \in \{1, 3, 5, 7, 9\}$, compared to the HE method. $t_c$ is the computational time in seconds.

2.9 This figure shows the output of global-to-local thresholding (top-row) and the enhanced volume (bottom-row) for three values of $rad_3 \in \{1, 2, 3\}$, compared to the LHE result. $t_c$ is the computational time in seconds.

2.10 This figure shows three examples of enhancing ultrasound volumes using the proposed preprocessing. In the right figure, $V_{eh}$, red and blue colors are used to visualize positive and negative values of voxels, respectively.

2.11 This figure shows single-layer and multi-layer structures of artificial neural networks.
2.12 This figure shows how the partitions of the kidney shape are designated in the 2D space. The green region corresponds to the partition $B_{2,1}$. .................................................. 54

2.13 This figure shows the block diagram of training the SANNs. ........................................ 54

2.14 This figure shows voxel classification for different number of texture features. The dark voxels correspond to the renal medulla region, the bright voxels correspond to kidney capsule and pyelocalyceal system regions, and the gray voxels are non-kidney voxels. .... 56

2.15 This figure shows the RCNCC metric versus two registration parameters, $\theta_x$ and $\theta_z$, for a 3D ultrasound image ($\theta_y = 0$ and $s = 1$). The brighter (yellow) regions correspond to desirable registration solutions. The gray stars are seed point, and the red star is the selected seed point. Four iterative updates of the registration parameters using the Gradient Descent method are shown with black arrows. ........................................ 62

2.16 This figure shows the block diagram of shape-based kidney detection approach. .............. 64

2.17 This figure shows the block diagram of atlas-based kidney detection approach. .............. 64

2.18 The processing pipeline of the two phases of the proposed computer-assisted ultrasound probe placement. ................................................................. 66

2.19 This figure provides sensitivity analysis of the shape-based and atlas-based kidney detection methods toward the kidney shape orientation over the $x$-axis. The blue lines in the top sub-figures shows detected orientation of the kidney shape in simulated ultrasound volumes over the $x$-axis, $\Delta \theta^d_x$, versus the actual orientation of the kidney shape over the $x$-axis, $\Delta \theta^a_x$ (gray line with the slope one). The green lines in the bottom figures show the kidney shape visibility versus the actual orientation of the kidney shape over the $x$-axis, $\Delta \theta^v_x$. The red lines in the bottom figures show the kidney shape detection versus the actual orientation of the kidney shape over the $x$-axis, $\Delta \theta^d_x$. .................................................. 74

2.20 This figure provides sensitivity analysis of the shape-based and atlas-based kidney detection methods toward the kidney shape orientation over the $y$-axis. The blue lines in the top sub-figures shows detected orientation of the kidney shape in simulated ultrasound volumes over the $y$-axis, $\Delta \theta^d_y$, versus the actual orientation of the kidney shape over the $y$-axis, $\Delta \theta^a_y$ (gray line with the slope one). The green lines in the bottom figures show the kidney shape visibility versus the actual orientation of the kidney shape over the $y$-axis, $\Delta \theta^v_y$. The red lines in the bottom figures show the kidney shape detection versus the actual orientation of the kidney shape over the $y$-axis, $\Delta \theta^d_y$. .................................................. 74
2.21 This figure provides sensitivity analysis of the shape-based and atlas-based kidney detection methods toward the kidney shape orientation over the $z$-axis. The blue lines in the top sub-figures show detected orientation of the kidney shape in simulated ultrasound volumes over the $z$-axis, $\Delta \theta_{z}^{dt}$, versus the actual orientation of the kidney shape over the $z$-axis, $\Delta \theta_{z}^{ac}$ (gray line with the slope one). The green lines in the bottom figures show the kidney shape visibility versus the actual orientation of the kidney shape over the $z$-axis, $\Delta \theta_{z}^{ac}$. The red lines in the bottom figures show the kidney shape detection versus the actual orientation of the kidney shape over the $z$-axis, $\Delta \theta_{z}^{ac}$.

2.22 This figure provides sensitivity analysis of the shape-based and atlas-based kidney detection methods toward the kidney shape translation over the $x$-axis. The blue lines in the top sub-figures show detected translation of the kidney shape in simulated ultrasound volumes over the $x$-axis, $\Delta t_{x}^{dt}$, versus the actual translation of the kidney shape over the $x$-axis, $\Delta t_{x}^{ac}$ (gray line with the slope one). The green lines in the bottom figures show the kidney shape visibility versus the actual translation of the kidney shape over the $x$-axis, $\Delta t_{x}^{ac}$. The red lines in the bottom figures show the kidney shape detection versus the actual translation of the kidney shape over the $x$-axis, $\Delta t_{x}^{ac}$.

2.23 This figure provides sensitivity analysis of the shape-based and atlas-based kidney detection methods toward the kidney shape translation over the $y$-axis. The blue lines in the top sub-figures show detected translation of the kidney shape in simulated ultrasound volumes over the $y$-axis, $\Delta t_{y}^{dt}$, versus the actual translation of the kidney shape over the $y$-axis, $\Delta t_{y}^{ac}$ (gray line with the slope one). The green lines in the bottom figures show the kidney shape visibility versus the actual translation of the kidney shape over the $y$-axis, $\Delta t_{y}^{ac}$. The red lines in the bottom figures show the kidney shape detection versus the actual translation of the kidney shape over the $y$-axis, $\Delta t_{y}^{ac}$.

2.24 This figure provides sensitivity analysis of the shape-based and atlas-based kidney detection methods toward the kidney shape translation over the $z$-axis. The blue lines in the top sub-figures show detected translation of the kidney shape in simulated ultrasound volumes over the $z$-axis, $\Delta t_{z}^{dt}$, versus the actual translation of the kidney shape over the $z$-axis, $\Delta t_{z}^{ac}$ (gray line with the slope one). The green lines in the bottom figures show the kidney shape visibility versus the actual translation of the kidney shape over the $z$-axis, $\Delta t_{z}^{ac}$. The red lines in the bottom figures show the kidney shape detection versus the actual translation of the kidney shape over the $z$-axis, $\Delta t_{z}^{ac}$.
2.25 This figure shows the output of both shape-based and atlas-based kidney detection methods for three arbitrary simulated ultrasound volumes. ............................... 77

2.26 This figure shows the receiver operating characteristic (ROC) curve, for detecting the with-kidney and without-kidney images of the evaluation set of actual ultrasound volumes of the healthy volunteers (Ds3DUsIs-AUVs-ESH), using the shape-based method, the atlas-based method, and a random guess. ............................... 78

2.27 This box-plot represents statistical variation of the accuracy of detecting the center point of the kidney shape in the with-kidney images of the Ds3DUsIs-AUVs-ESH dataset for the four kidney detection methods. ............................... 79

2.28 The table in this figure represents the statistical comparison of the kidney detection methods for estimating the center-point of the kidney shape in the with-kidney images of the evaluation set of actual ultrasound volumes of the healthy volunteers (Ds3DUsIs-AUVs-ESH). $h$ is the state of the t-test analysis for each pair of methods, based on the significance level of 5%. For each non-diagonal cell of the table, $h = 1$ means the corresponding method on the left column performs better than the corresponding method in the top row, $h = -1$ means the corresponding method in the top column performs better than the corresponding method in the left row, and $h = 0$ means the experimental data of the corresponding methods do not show a significant difference. ............................... 80

2.29 This figure shows the steps of shape-based and atlas-based kidney detection methods for five with-kidney actual images of the Ds3DUsIs-AUVs-ESH dataset. Each row belongs to an actual ultrasound volume. ............................... 80

2.30 This bar-graph shows the total computational time of the shape-based and atlas-based methods and the computational times of three parts of the shape-based and atlas-based methods, including LHE, shape-to-volume registration, and voxel classification, which implementational approach highly affects their computational time. ............................... 82

3.1 This figure categorizes the commonly used segmentation methods in medical imaging. ............................... 89

3.2 Displaying control points and sample points in the B-Spline surface model [3]. ............................... 95

3.3 3D superquadrics. (a) $\epsilon_1 = \epsilon_2 = 0.1$; (b) $\epsilon_1 = \epsilon_2 = 0.5$; (c) $\epsilon_1 = \epsilon_2 = 1$; (d) $\epsilon_1 = \epsilon_2 = 2$; (e) $\epsilon_1 = \epsilon_2 = 4$ [4]. ............................... 96

3.4 Displaying the geometry of the deformable superquadric surface [5]. ............................... 97
3.5 Displaying the steps of superquadrics with model-recovery-model and model-select. (a) shows a 3D input range data; (b) shows the initialization step with seeds. (c) shows an intermediate step of model-recovery; and (d) shows the final model after model-selection stage [6].

3.6 Displaying radial rays and the segmentation surface of the active rays approach [7].

3.7 Displaying the automated landmarking process, proposed by Frangi et al. [8]. Based on using the natural coordinate, the principal modes of variation only account for nonrigid deformations, and pose or size differences are not considered in the statistical shape modeling.

3.8 (a) shows the alignment of vertices in triangulation model versus simple mesh. (b) shows the geometry of simplex mesh and the simplex angle [9].

3.9 The solid and dashed contours show T-surface before and after deformations, respectively. Yellow nodes show aligned nodes on cells edges. Red and white points show 'on' vertices and newly added vertices, respectively. (a) shows the deformation process. (b) second phase of re-parametrization. (c) first phase of re-parametrization and boundary cells with gray color [10].

3.10 Displaying a particle, vertices and springs (gray) [11].

3.11 This figure represents the iterative process of image segmentation with the regional level-set method.

3.12 This figure represents the iterative process of image segmentation with the regional level-set method.

3.13 This figure lists the segmentation methods along with their advantages, disadvantages, rate of use in medical image segmentation, and references (examples) of papers which used the segmentation methods in 3D ultrasound imagery.

3.14 This figure the block diagram of the proposed CVRLS-SP method for segmenting the kidney shape into its structural regions.

3.15 Displaying TP, TN, FP and FN regions, and the Hausdorff distance (HD).

3.16 The 3D surface, plotted in this figure, represents the average kidney segmentation accuracy, based on the DSC metric, for the training volumes of the dataset of actual ultrasound volumes (Ds3DUltrasound-AUms-Tr).
3.17 This figure shows results of the kidney segmentation methods, including (a) the proposed method of this paper, (b) Marsousi et al-EMBC14, (c) Noll et al., and (d) MRF-AC, of three ultrasound volumes of the evaluation set. The green, red and yellow colors represent FN, FP, and TP regions, respectively.

3.18 Displaying automated kidney segmentation process of three with-kidney images with low, middle, and high accuracy of kidney segmentation as, $DSC = 0.70$, $DSC = 0.83$, and $DSC = 0.86$ respectively. The three images represent three different kidney shape’s postures with different scales and orientations. The sub-figures from left to right show original volume, detected kidney shape, segmented kidney shape, comparison of AS vs. GT, and 3D rendered AS. In the second and third sub-figures from left, the blue and red regions indicate renal medulla and pyelocalyceal system/kidney capsule, respectively. In the second sub-figure from right side, the yellow, red, and green colors indicate TP, FP, and FN regions, respectively.

3.19 This box-plot displays statistical variations of the accuracy of kidney segmentation, based on the Dice’s similarity coefficient, using the proposed CVRLS-SP method, Marsousi et al-EMBC14, Noll et al. [12], and MRF-AC [13] methods.

3.20 The table in this figure represents the statistical comparison of the kidney segmentation accuracy of the methods, including the proposed CVRLS-SP, Marsousi et al-EMBC14 (M-EMBC14), Noll et al., and MRF-AC. The results are obtained using the with-kidney images of the evaluation set of actual ultrasound volumes of the healthy volunteers ($Ds3DUIs-AUVs-ESH$). $h$ is the state of the t-test analysis for each pair of methods, based on the significance level of 5%. For each non-diagonal cell of the table, $h = 1$ means the corresponding method on the left column performs better than the corresponding method in the top row, $h = -1$ means the corresponding method in the top column performs better than the corresponding method in the left row, and $h = 0$ means the experimental data of the corresponding methods do not show a significant difference.
## List of Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1D</td>
<td>One dimensional</td>
</tr>
<tr>
<td>2D</td>
<td>Two dimensional</td>
</tr>
<tr>
<td>3D</td>
<td>Three dimensional</td>
</tr>
<tr>
<td>4D</td>
<td>Four dimensional</td>
</tr>
<tr>
<td>3D US</td>
<td>Three dimensional ultrasound</td>
</tr>
<tr>
<td>4-DOF</td>
<td>Four degrees of freedom</td>
</tr>
<tr>
<td>AAM</td>
<td>Active appearance model</td>
</tr>
<tr>
<td>ANN</td>
<td>Artificial neural network</td>
</tr>
<tr>
<td>ASM</td>
<td>Active shape model</td>
</tr>
<tr>
<td>CAD</td>
<td>Computer-assisted diagnostic</td>
</tr>
<tr>
<td>CC</td>
<td>cross correlation</td>
</tr>
<tr>
<td>CDF</td>
<td>Cumulative density function</td>
</tr>
<tr>
<td>CSI</td>
<td>Class separability index</td>
</tr>
<tr>
<td>CT</td>
<td>Computer tomography</td>
</tr>
<tr>
<td>CVISM</td>
<td>Complex-valued implicit shape-model</td>
</tr>
<tr>
<td>CVRRLS-SP</td>
<td>Complex-valued regional level-set approach via shape prior</td>
</tr>
<tr>
<td>Ds3DUis</td>
<td>Dataset of 3D ultrasound images</td>
</tr>
<tr>
<td>Ds3DUis-AUVs</td>
<td>Dataset of 3D ultrasound images - actual ultrasound volumes</td>
</tr>
<tr>
<td>Ds3DUis-GTs</td>
<td>Dataset of 3D ultrasound images - ground truths</td>
</tr>
<tr>
<td>Ds3DUis-UVS</td>
<td>Dataset of 3D ultrasound images - ultrasound volume simulator</td>
</tr>
<tr>
<td>FAST</td>
<td>Focus assessment with sonography for trauma</td>
</tr>
<tr>
<td>FFS</td>
<td>Fully formed speckle</td>
</tr>
<tr>
<td>FIR</td>
<td>Finite impulse response</td>
</tr>
<tr>
<td>G-SVMs</td>
<td>Gabor-support vector machines</td>
</tr>
<tr>
<td>G2LTh</td>
<td>Global-to-local thresholding</td>
</tr>
<tr>
<td>GPU</td>
<td>Graphical processing unit</td>
</tr>
<tr>
<td>GVF</td>
<td>gradient vector flow</td>
</tr>
<tr>
<td>HDLSS</td>
<td>high-dimension-low-sample-size</td>
</tr>
<tr>
<td>HE</td>
<td>Histogram equalization</td>
</tr>
</tbody>
</table>
ICP: Iterative Closest Point
IRFAST: Image-registered FAST
MFLV: Mixture of filters based on local variation
MRF-AC: Markov random field and active contours
NCC: normalized cross correlation
NRLR: Non-randomly distributed with long-range order
NRSR: Non-randomly distributed with short-range order
LHE: Localized histogram equalization
LUQ: Left-upper-quadrant
OSRAD: Oriented speckle removing anisotropic diffusion
PDE: Partial difference equation
PDM: Point distribution model
RCNCC: Regularized complex normalized cross-correlation
ROC: Receiver operating characteristic
RUQ: Right-upper-quadrant
SAD: sum of absolute difference
SANN: Spatially aligned neural networks
SNR: Signal to noise ratio
SpringLS: Spring Level-set
SR-DL: sparse representation via dictionary learning
SRAD: speckle removing anisotropic diffusion
STAPLE: Simultaneous truth and performance level estimation
SVD: Singular value decomposition
TGC: Time gain compensation
URI: Ultrasound research interface
### List of Important Symbols

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$V$</td>
<td>input ultrasound volume</td>
</tr>
<tr>
<td>$[s_x, s_y, s_z]$</td>
<td>specifies the size of ultrasound along the $x$-, $y$-, and $z$- axes</td>
</tr>
<tr>
<td>$\Omega_V$</td>
<td>3D image domain of the ultrasound volume</td>
</tr>
<tr>
<td>$\bar{X}$</td>
<td>coordinate of a voxel in a 3D image domain</td>
</tr>
<tr>
<td>$G_V$</td>
<td>ground truth data of the input volume $V$, as a binary volume, specifying voxel’s membership of $V$ to the kidney shape</td>
</tr>
<tr>
<td>$\Psi$</td>
<td>representing the kidney shape model (CVISM)</td>
</tr>
<tr>
<td>$ST_{\tilde{\rho}_{st}}$</td>
<td>specifies a similarity transformation</td>
</tr>
<tr>
<td>$\tilde{\rho}_{st}$</td>
<td>a vector of the similarity transformation parameters</td>
</tr>
<tr>
<td>$V_{dn}$</td>
<td>de-speckled volume</td>
</tr>
<tr>
<td>$\text{nib}(X, rad)$</td>
<td>specifying a set of voxels, covered by box with the center point, $\bar{X}$, and radius, $rad$.</td>
</tr>
<tr>
<td>$V_{eh}$</td>
<td>enhanced volume</td>
</tr>
<tr>
<td>$L$</td>
<td>output of voxel classification with SANNs</td>
</tr>
<tr>
<td>$\Gamma^*$</td>
<td>the value of the RCNCC metric</td>
</tr>
<tr>
<td>$S_T$</td>
<td>a set of seed points for the shape-to-volume registration</td>
</tr>
<tr>
<td>$\phi$</td>
<td>level-set function</td>
</tr>
<tr>
<td>$\phi$</td>
<td>level-set function of the shape prior</td>
</tr>
<tr>
<td>$\Re{}$</td>
<td>Real part extractor for complex numbers</td>
</tr>
<tr>
<td>$\Im{}$</td>
<td>Imaginary part extractor for complex numbers</td>
</tr>
<tr>
<td>$AT_{\tilde{\rho}_{af}}$</td>
<td>specifies an affine transformation</td>
</tr>
<tr>
<td>$\tilde{\rho}_{af}$</td>
<td>a vector of the affine transformation parameters</td>
</tr>
<tr>
<td>$H(.)$</td>
<td>is the Heaviside function</td>
</tr>
<tr>
<td>$\delta(.)$</td>
<td>is the delta Dirac function</td>
</tr>
</tbody>
</table>
Chapter 1

Introduction
1.1 Background

Abdominal trauma causes a large number of preventable death each year [14]. Abdominal trauma is referred to injuries to the abdomen, either blunt or penetrating, which usually results in severe blood loss and death if its treatment process is delayed. An abdominal trauma can be caused by many different reasons, though, motor vehicle collisions, sport injuries, child abuse, and most importantly gunshot wounds are among the most frequent causes of abdominal trauma. An abdominal trauma patient may have early symptoms, including abdominal pain, nausea, vomiting, blood in the urine, and fever [15]. An abdominal trauma with massive internal bleeding results in a great amount of blood in abdomen, and it quickly threatens the trauma patient’s life [16]. To save a trauma patient’s life, he or she should go immediately under diagnosis and treatment.

![Diagram of trauma diagnosis protocol](image)

Figure 1.1: This figure shows a protocol for trauma diagnosis in trauma patients. The focus of this thesis is specified by blue (darker) background color.

Although several imaging modalities can be used for trauma diagnosis, ultrasound imagery is the preferred one for hemodynamically unstable patients [17]. Hemodynamics is the fluid dynamics of blood flow. Hemodynamic parameters include heart rate, blood pressure, central venous pressure, pulmonary artery pressure, pulmonary artery occlusion pressure, and cardiac output. Hemodynamic instability refers to abnormalities of any of these parameters [18]. Detecting hemodynamic instability is easy when hemodynamic parameters largely deviate from its normal range. Some techniques to evaluate...
Figure 1.2: Displaying the six abdominal views of FAST examination [1]. The region (2) corresponds to the Morison’s pouch (RUQ) view, which is the focus of this PhD thesis.

hemodynamic instability are listed below [19]:

- Alterations in pulse (heart beat) is often an indication of blood loss or dehydration;

- Blood pressure or mean arterial pressure is an appropriate indicator of clinical instability;

- Extremes of temperature is considered as a sign of clinical instability.

Paramedics and first responders are capable to detect hemodynamic instability in such a situation. However, detecting hemodynamic instability is not easy in an early stage of hemorrhage. Muniz et al. [20] showed a computer-assisted solution can be used to help paramedics in early identification of a hemodynamically unstable patient.

Computer Tomography (CT) imaging provides the highest imaging quality of internal organs, however a patient should be under a stable condition to be moved to the CT scanning room [17]. This restricts the utility of CT imaging for hemodynamically unstable patients, and leaves no option other than using ultrasound imagery, as a portable imaging device, for diagnosing hemodynamically unstable patients in emergency situations [17]. In addition, other attributes of ultrasound imagery such as non-invasiveness and affordability add to its suitableness for trauma diagnosis in emergency situations [15]. A common protocol of trauma diagnosis is shown in Fig. 1.1. Also, Fig. 1.1 shows the focus of this thesis research.

Ultrasound imagery has been widely used in trauma diagnosis, also called as Focused Assessment with Sonography for Trauma (FAST). FAST is a rapid bedside ultrasound examination to find free fluids,
as an indication of internal bleeding, around abdominal organs (hemoperitoneum) [21]. An internal free-fluid has a “low-echoic” characteristic, and does not scatter ultrasonic waves. Therefore, it appears as a dark region in sonography. In trauma diagnosis, ultrasonographers look for irregular dark regions in internal views corresponding to abdominal bleeding [22]. In general, six internal views are associated with abdominal trauma diagnosis, which are listed here:

1. Parasternal view to detect pericardial fluid;
2. Right-upper-quadrant view (RUQ) to detect fluid around the hepatorenal interface and right chest;
3. Right paracolic gutter;
4. Left upper quadrant (LUQ) view to detect fluid around the splenorenal interface and left chest;
5. Left paracolic gutter;
6. Longitudinal and transverse pelvis views to find trauma around bladder.

These six views are shown in Fig. 1.2. Among these six views of FAST examination, the Morison’s pouch view (RUQ view) has a higher sensitivity to peritoneal free-fluid of a trauma patient in supine position [23]. This is because free fluids due to abdominal bleeding commonly move and locate in the upper right quadrant view of a trauma patient in supine position [24]. Thus, the focus of this thesis is on the Morison’s pouch (RUQ) view (see Fig. 1.2). Figure 1.3 shows a two-dimensional (2D) ultrasound image of the Morison’s pouch view of a trauma patient. In Fig. 1.3, liver, right kidney, and free-fluid are specified.

Figure 1.3: This figure shows a 2D Morison’s pouch (RUQ) view, and specifies liver, right kidney, and free-fluid regions [1].
1.2 Motivation

Computer-assisted diagnostic (CAD) systems are becoming very popular in medical applications [25, 26], such as breast-cancer detection [27] and cardiology [28]. Attempts to engage CAD systems in medical applications have drastically increased due to recent advancements in parallel computation technology, which facilitate developing near real-time diagnostic tools [29]. In the line with these efforts, this thesis is focused on developing computer-assisted algorithms to be used in rapid diagnosis of abdominal trauma patients in emergency situations.

As discussed earlier in Sec 1.1, FAST examination is the preferred diagnosis approach for hemodynamically unstable trauma patients, especially in emergency situations. This is because ultrasound imagery, as a portable device, can be carried to the room or place in which an unstable patient is under care of physicians. Trauma physicians and ultrasonographers usually are trained to perform FAST examinations. However, in remote regions where a trauma patient requires emergency care and rapid diagnosis, trained physicians and ultrasonographers are not always accessible to conduct FAST examination [30]. Instead, paramedics are those who are sent to emergency situations to provide first aids to (trauma) patients. But, paramedics usually lack the following knowledge, restraining them from conducting the FAST examination:

- proper radiological knowledge to correctly place the ultrasonic probe on the trauma patient’s body to scan relevant views of trauma diagnosis,
- adequate anatomical understanding to recognize internal organs, and to search for free fluids in the spaces between internal organs in acquired ultrasound images,
- enough experience and training to distinguish free-fluids (due to an internal bleeding) from fluid-carrying organs.

Thus, trauma patients are kept under regular care until they are transported (usually with an ambulance) to a referral hospital. This puts a long delay in the process of trauma diagnosis of the trauma patient, which could cost the patient’s life. This problem becomes more serious in situations with massive casualty, such as massive car collisions, natural disasters, and wars. Because of limited capacities of surgical operations in these situations, all patients must be quickly diagnosed, and only true trauma patients should be transported to surgical rooms. Otherwise, a wrongly selected patient with stable condition might be taken to the surgery room, instead of a true trauma patients which would cost at the patient’s life [30]. The lack of facilities and trained physicians in emergency situations result in a
large number of preventable deaths every year [30].

The motivation of this PhD research is that a computer-assisted solution can be used to fill the above-mentioned gap by facilitating trauma diagnosis by paramedics. Such a solution offers rapid diagnosis of trauma patients in emergency situations, and can potentially lead to reduce the current mortality rate due to abdominal trauma. A computer-assisted trauma diagnosis can be designed to assist operators in the following ways:

- guiding them to correctly place the ultrasonic probe on the patient’s body to acquire internal views corresponding to FAST examination,
- analyzing acquired ultrasound images to decide whether a free-fluid, as an indication of internal bleeding, exists or not.

The primary assistant that a computed-assisted trauma diagnosis can provide, is to guide the FAST examiner to move and/or rotate the ultrasonic probe toward the correct alignment on the patient’s body to acquire images corresponding to FAST examination. Without providing such a CAD tool, it is impossible for a non-skilled operators (ie. paramedics) to properly find correct views (such as the Morison’s pouch). Hence, providing computer-assisted probe placement is the primary requirement in designing a computer-assisted trauma diagnosis.

The task of detecting and classifying fluids into abdominal bleeding due to trauma and fluid-carrying organs requires a good anatomical knowledge. The abdominal location of a detected fluid is an important feature to decide whether it is a free-fluid due to abdominal bleeding or a fluid-carrying organ. The location of a detected fluid can be obtained from its adjacency with other abdominal organs. For example, in a Morison’s pouch image of a trauma patient, a free-fluid due to abdominal bleeding usually place around the interface of the right kidney and liver (See Fig. 1.3). Therefore, it is necessary that a computer-assisted solution is engaged to automatically detect and segment relevant organs to trauma diagnosis in acquired ultrasound images (i.e the right kidney and liver in the Morison’s pouch view).

Due to recent advancements in parallel computing, real-time 3D ultrasound imaging and volumetric data processing are becoming realizable, which facilitates the development of the computer-assisted system based on the above-mentioned requirements. Recent works reported volumetric imaging speed of 32 volumetric ultrasound images per second [31] which is a break-through record, compared to the conventional technology which provides a single 3D ultrasound image each four seconds. With this speed of 3D ultrasound imaging, smooth and fast navigation of the ultrasonic probe toward the correct placement on the patient’s body is possible. On the other hand, recent advancements in the multi-thread computation using multi-core processors and GPU programming facilitates real-time 3D image
Chapter 1. Introduction

Motivations

Saving trauma patients' lives
Recent development of real-time 3D ultrasound imagery
Recent advancements in parallel computation

Developing computer-assisted trauma diagnosis system facilitates FAST-examination by paramedics.
Designing computer-assisted probe placement is only possible using a real-time 3D ultrasound imaging device.
Implementing real-time automated algorithms to detect organs and free-fluids are only possible by high computational capability.

Importance

Figure 1.4: This figure summarizes the motivations of this PhD thesis.

Figure 1.5: This figure shows a 2D ultrasound and a 3D ultrasound image.

Perhaps, without the new advancement in parallel computation, developing a trauma diagnosis system was impossible. Thus, the recent technological advancements in 3D ultrasound imaging and parallel computation have also motivated this PhD thesis. The motivations of this PhD thesis are summarized in Fig. 1.4. In the next section, we discuss the advantages of 3D ultrasound imagery compared to 2D ultrasound imaging and CT imaging.

1.3 Why 3D Ultrasound Imagery?

Although CT imaging has been recognized as the most accurate imaging modality to diagnose trauma patients, ultrasound imagery is the preferred imaging modality for diagnosing hemodynamically unstable trauma patients. This is because moving an unstable patient under resuscitational operations to a scanning room with CT imaging system puts a large risk on the patient’s life. Ultrasound imagery is portable and can be carried to any place. Despite imaging resolution and quality of ultrasound imagery
### Figure 1.6:}
This figure lists advantages and disadvantages of 3D ultrasound imaging, CT imaging, and 2D imaging for designing a computer-assisted trauma diagnostic solution.

Ultrasound imaging does not impose any risk to patients, whereas X-ray radiation in CT imaging limits its utility for some patients, like pregnant patients;

- Since ultrasound devices are portable, unstable patients are not required to be moved from a resuscitation room to an imaging room;

- Ultrasound imaging provides immediate results, which accelerates triaging, whereas both CT and MRI imaging are considered as off-line imaging modalities.

<table>
<thead>
<tr>
<th>3D ultrasound imaging</th>
<th>CT imaging</th>
<th>2D ultrasound imaging</th>
</tr>
</thead>
<tbody>
<tr>
<td>GE Voluson e ultrasound machine</td>
<td>GE Revolution CT imaging machine</td>
<td>GE Vscan 2D pocket-size ultrasound device</td>
</tr>
<tr>
<td>Portable</td>
<td>Bulky and fixed</td>
<td>Portable</td>
</tr>
<tr>
<td>Good for developing automated algorithms for detecting and localizing organs</td>
<td>Perfect for developing automated algorithms for detecting and localizing organs</td>
<td>Not appropriate for developing automated algorithms</td>
</tr>
<tr>
<td>Real-time (using new technology of computation)</td>
<td>Off-line</td>
<td>Real-time</td>
</tr>
<tr>
<td>Noninvasive</td>
<td>Prohibited for some patients</td>
<td>Noninvasive</td>
</tr>
<tr>
<td>Marginal quality and clearness of organs in acquired images</td>
<td>Higher quality and resolution of images</td>
<td>Poor quality of imaging</td>
</tr>
<tr>
<td>Affordable</td>
<td>Expensive</td>
<td>Low price</td>
</tr>
</tbody>
</table>

⊕: Perfect matching with the requirements of computer-assisted trauma diagnosis.
⊕: Marginal matching with the requirements of computer-assisted trauma diagnosis.
○: Poor matching with the requirements of computer-assisted trauma diagnosis.
Compared to 2D ultrasound imaging, 3D ultrasound is less popular, especially in FAST examination, because [33]:

- 2D ultrasound imagery has been used for a long while, while 3D ultrasound is relatively new, and many physicians are not trained to use 3D ultrasound imagery.

- 3D ultrasound data has a higher structural complexity, which makes it difficult for radiologists to interpret acquired 3D ultrasound images without using computer-assisted tools.

Despite the fact that 2D ultrasound imaging has been used as the popular imaging tool for FAST examination, 3D ultrasound imagery is essential for designing a computer-assisted trauma diagnostic solution. Compared to MRI and CT imaging modalities which provide internal views with fixed coordinate with respect to the patient’s body, ultrasound imagery does not have a fixed coordinate with respect to the patient’s body. In ultrasound imagery, computer-assisted ultrasonic probe placement can be provided by estimating the ultrasonic probe’s alignment relative to internal organs or structures of interest. Detecting and localizing internal organs or structures are only possible using a 3D ultrasound imaging device, and this is the reason why designing computer-assisted ultrasonic probe placement requires 3D ultrasound imagery [34]. In addition, 3D ultrasound imagery offers a higher detectability to abdominal bleeding over 2D ultrasound imagery, because 3D ultrasound provides internal views that could not be seen with 2D ultrasound imaging, and therefore, a hidden internal bleeding from 2D ultrasound imaging can be detected using 3D ultrasound imagery [33]. Furthermore, 3D ultrasound imagery facilitates measuring the volume of an internal bleeding, which is not achievable by 2D ultrasound imagery [33]. Hence, 3D ultrasound imagery helps trauma experts to know the size and volume of internal bleeding. Figure 1.5 shows both a 2D ultrasound image and a 3D ultrasound image. The advantages and disadvantages of 3D ultrasound imagery, CT scan, and 2D ultrasound imagery for designing a computer-assisted solution for trauma diagnosis of hemodynamically unstable patients in emergency situations are summarized in Fig. 1.6. As shown in Fig. 1.6, 3D ultrasound imagery has a better matching with the requirements of designing a computer-assisted trauma diagnostic solution.

1.4 Problem Definition

The problem to be addressed in this thesis is to design a computer-assisted solution to help paramedics to conduct FAST examination of trauma patients in emergency situations. This computer-assisted solution is based on 3D ultrasound imagery to scan abdominal organs and views, corresponding to the FAST examination. To simplify the problem, the Morison’s pouch (RUQ) view is only considered at
Figure 1.7: This figure defines the problem to be addressed in this PhD thesis. As shown in figure, the 3D ultrasound (3D US) imaging device is connected to the personal computer using the ultrasound research interface (URI) protocol.

this stage (future works can concentrate on other views to improve the solution’s effectiveness). The computer-assisted trauma diagnostic solution includes the following parts:

- providing a computer-assisted mechanism to guide operators to find the correct ultrasonic probe placement, corresponding to the Morison’s pouch (RUQ) view, on the patient’s body,
- automatically detecting and segmenting the kidney shape in 3D ultrasound images,
- automatically detecting and segmenting fluid regions.

Figure 1.7 shows the problem to be addressed in this PhD thesis. As shown in Fig. 1.7, in the computer-assisted trauma diagnosis of this thesis, acquired 3D ultrasound images using the 3D ultrasound device are sent to a fast personal computer (laptop), in which the computer-assisted tools are implemented and installed.

1.5 Challenges

The design of a computer-assisted solution for FAST examination face many challenges, which can be categorized into (a) imaging modality related challenges, (b) hardware/software based challenge, and (c) experimentation based challenge. These challenges should be adequately addressed to provide a reliable solution for trauma diagnosis. Figure 1.8 summarizes the challenges toward the design of the computer-assisted diagnostic system for FAST examination. These challenges are discussed in the following paragraphs.

Processing ultrasound imagery, as the targeted modality of this thesis, has been recognized as a highly challenging task among the biomedical signal processing society. Speckle, which is due to the nature of ultrasound imaging [35], reduces both the spatial resolution and contrast quality of imaging internal structures in ultrasound images [36]. Speckle in synthetic aperture images is multiplicative noise [35], and is generated due to the artifact caused by interference signals from randomly distributed
Chapter 1. Introduction

Challenges of computer-assisted FAST examination

- Image modality related challenges
  - Ultrasound specific challenges
    - Partial occlusion of kidney shape
  - Organ-specific challenge
  - Ultrasonic probe misalignment
- Hardware/software based challenges
  - Supporting multi-core processors
  - Using effective and efficient algorithms
- Experimentation based challenge
  - Supporting fast GPU
  - Using parallel computation

Figure 1.8: This figure shows the challenges toward designing the computer-assisted diagnostic system for FAST examination.

scatters [37]. In particular, in ultrasound imaging, speckle is caused by random interactions of scattered ultrasonic waves from tissues [38].

As an ultrasound beam penetrates deep into the body through tissue layers, its amplitude is attenuated due to absorption, reflection, and scattering at interfaces. This beam attenuation is compensated in most of ultrasound devices by applying the *time gain compensation* (TGC) technique [39]. TGC equally amplifies echoes in the same depth, however it does not linearly operate for similar tissues in different depths. As a result, regions of the same tissue type may appear inhomogeneous, and the intensity distribution of different tissues overlap each other [39]. This inhomogeneity of intensity profile of tissues is an obstacle toward detecting and segmenting the kidney shape.

In addition to challenges associated with the nature of ultrasound imagery, other challenges exist, which are related to the scan of Morison’s pouch view. Due to the adjacency of the right kidney and liver with ribs, which highly scatter ultrasonic waves, dark shadows are often projected on the kidney and liver shapes. Also, the ultrasonic probe’s misalignment, which could happen due to operators’ inexperience, results in partial visualization of the kidney shape. As a result of these two problems, the kidney shape is partially occluded by shadows, which adds up to the complexity level of kidney detection and segmentation.

The kidney detection module is a very important part of the computer-assisted trauma diagnosis, because the detected kidney's alignment in 3D ultrasound volumes will be used to provide computer-assisted probe placement. Despite computational complexity of 3D image processing for automated
kidney detection is very high, the computer-assisted solution targeted in this thesis must operate in a near real-time fashion. This requires both hardware and software considerations as follows:

- a fast portable computer is required to obtain the highest possible computational speed;

- the algorithms must be efficiently developed.

Parallel computation can be used as a remedy to decrease computational time of kidney detection, however it requires the developed algorithms to be highly parallelizable. Also, to use parallel computation, the hardware setting should be capable to support computation with fast graphical processing units (GPUs).

Finally, computerized algorithms designed for trauma diagnosis should be carefully evaluated through experimentation to ensure they could be used in real-scenarios. This requires access to ultrasound volumes acquired from both normal volunteers and trauma patients. However, access to such images are restricted due to confidentiality of patients’ information. Also, collected ultrasound volumes should provide a wide range of variability in terms of morphological variations and different postures of the kidney shape to ensure the computerized algorithms are robust enough to be applied on new patients. In addition, correct metrics should be selected to evaluate the performance of algorithms.

1.6 Volumetric Ultrasound Dataset

Ultrasound volumetric dataset is a collection of 3D images either acquired from an ultrasound device or synthetically generated to simulate 3D images acquired from an ultrasound machine. In this thesis, a dataset of 3D ultrasound images (Ds3DUsIs) is used for three purposes, including training, sensitivity analysis to parameters of the introduced algorithms, and evaluation of the introduced methods compared to state-of-the-art. Ds3DUsIs consists of three types of data, including actual ultrasound volumes (Ds3DUsIs-AUVs), ground truths (Ds3DUsIs-GTs), and RUQ ultrasound volume simulator (Ds3DUsIs-UVS). The actual ultrasound volumes are acquired using a 3D ultrasound machine from both normal volunteers and abnormal subjects. In the rest of this section, we discuss Ds3DUsIs in details. The detailed structure of Ds3DUsIs is provided in Fig. 1.10.

1.6.1 Actual ultrasound volumes (Ds3DUsIs-AUVs)

An actual ultrasound volume is a 3D image acquired from a normal volunteer or a trauma patient using a 3D ultrasound device.
**Definition 1.1.** A 3D Ultrasound volume is defined as, \( V \in \mathbb{R}^{s_x \times s_y \times s_z} \), where \( \{s_x, s_y, s_z\} \) specify the size of \( V \), such that \( V(\vec{X}) \) maps each voxel, \( \vec{X} \), belonging to the image domain, \( \Omega_V \), into an integer value from 0 to 255. This mapping is expressed as \( V : \vec{X} = [x, y, z]^T \mapsto y \), where \( \vec{X} \in \Omega_V \) and \( y \in [0, 1, \cdots, 255] \).

The set of actual ultrasound volumes (\( Ds3DUsIs-AUVs \)) are acquired using the GE Voluson e machine [2]. The ultrasound volumes have been acquired under all ethical considerations, and healthy volunteers and abnormal subjects agreed to use their images for researching purposes [40]. The GE Voluson e machine operates in 2D B-Mode, 3D static, and real-time 4D modes. The GE Voluson e machine generates 3D ultrasound images using a mechanical wobbling transducer. The acquisition time of a 3D static volume is adjustable from 7.5 to 15 seconds. “The longer the acquisition time, the better the spatial resolution will be” [2]. This ultrasound device supports variety of transducers. For 3D abdominal imaging, a particular curved-array ultrasonic probe is recommended (RM6C H48671ZG), which can send and receive ultrasound beams between 1 MHz to 7 MHz with the view angle of 60°. The acquired volumes are stored in “KRETZFILE 1.0” format, which is a specific format of GE ultrasound volumes. Figure 1.9 shows the GE Voluson e ultrasound device and the RM6C ultrasound probe used to generate the actual ultrasound volumes of \( Ds3DUsIs-AUVs \).

![Figure 1.9](image_url)

**Figure 1.9:** This figure shows (a) the curved-array ultrasonic probe (RM6C H48671ZG transducer) for real-time 4D abdominal imaging, (b) the GE Voluson e ultrasound device used to generate actual ultrasound volumes of this thesis, and (c) the 3D scan procedure applied by the ultrasound device to generate volumetric images [2].

\( Ds3DUsIs-AUVs \) includes 50 ultrasound volumes acquired from eight healthy volunteers and eight abnormal subjects. The abnormal subjects show typically in-body fluids, while their clinical situations are not critical, and 3D ultrasound scanning have been scheduled well in advance. Note that the
process of 3D ultrasound image acquisition is currently slow (7.5 to 15 seconds per volume), and it cannot be directly applied on trauma patients in hemodynamically unstable condition. Though, 3D ultrasound images acquired from the abnormal subjects with in-body fluids resemble 3D ultrasound images containing abdominal bleeding of trauma patients. The 50 ultrasound volumes are comprised of 29 ultrasound volumes from the Morison’s pouch (RUQ) view, visualizing the right kidney shape and referred as “with-kidney” images, and 21 ultrasound volumes from other internal views, which do not show the right kidney shape and referred as “without-kidney” images. Eight ultrasound volumes of the 29 with-kidney images have been acquired from abnormal subjects, and the rest 21 volumes have been acquired from healthy volunteers. All the without-kidney images have been acquired from healthy volunteers.

Remark 1.1. The use of both with-kidney and without-kidney images allows us to evaluate the accuracy of detecting images with the kidney shape from images without the kidney shape.

All the “with-kidney” images have the same size of \( \{ s_x = 178 \, \text{px}, s_y = 250 \, \text{px}, s_z = 178 \, \text{px} \} \), and their pixel resolution is \( 0.056 \, \text{mm} \), providing an approximate depth of view of \( 178 \, \text{px} \times 0.056 \, \text{mm} \approx 10 \, \text{cm} \).

Figure 1.12 shows three actual ultrasound volumes, including (a) a “with-kidney” image of a healthy volunteer, (b) a “without-kidney” image of a healthy volunteer, and (c) a “with-kidney” image of an abnormal patient.

Ds3DUsIs-AUVs is divided into training and evaluation sets. 12 volumetric images, consisting of six “with-kidney” an six “without-kidney” images, are used for training purposes (Ds3DUsIs-AUVs-Tr). The evaluation set consists of two parts: volumes of healthy volunteers, and volumes of abnormal subjects. The evaluation set of healthy volunteers (Ds3DUsIs-AUVs-ESH) consists of 15 “with-kidney” images and 15 “without-kidney” images. The evaluation set of abnormal subjects (Ds3DUsIs-AUVs-ESA) consists of eight “with-kidney” images.

Remark 1.2. In abnormal subjects with in-body fluids, free-fluids around the right kidney can change the morphology of the RUQ view. Thus, using the images of abnormal subjects, the robustness of kidney shape detection against morphological changes can be evaluated.

Remark 1.3. No subject has images in both training and evaluation sets.

Remark 1.4. The kidney shape of ultrasound images of each volunteer in the dataset is morphologically different from the kidney shape of other volunteers, and therefore, the evaluation of kidney detection and segmentation methods using the with-kidney images reflects the robustness of kidney detection and segmentation methods under morphological variations of the kidney shape.
1.6.2 Ground truth data (Ds3DUsIs-GTs)

For each with-kidney volume, at least one ground truth data has been manually generated using TurtleSeg [41]. A ground truth data of a with-kidney volume, $G_V \in \mathbb{R}^{s_x \times s_y \times s_z}$, is a binary volumetric data, which specifies voxels’ memberships to the kidney shape, such that a voxel with coordinates $\vec{X} = [x, y, z]^T$ belongs to the kidney shape if $G_V(\vec{X}) = 1$, and the voxel does not belong to the kidney shape if $G_V(\vec{X}) = 0$.

Two trainees have been engaged to generate the ground truth data. The trainees have been asked to re-generate ground truth data of each volume for three times. The final ground truth data of each with-kidney volume is created by calculating the average shape. Based on the Dice’s coefficient (DSC) [42], we calculated the inter- and intra-observer agreements of the ground truth data, and the average inter-observer agreement, and intra-observer agreements of trainees #1 and #2 are 0.8345, 0.9172, and 0.9166, respectively.
1.6.3 Ultrasound volume simulator (Ds3DUsIs-UVS)

The number of ultrasound volumes in the dataset of actual ultrasound volume is limited, and the actual ultrasound dataset does not provide a wide range of shape variability to evaluate the robustness of kidney detection and segmentation. As we will discuss later, we are interested to examine the robustness of the proposed method against a wide range of deformations and occlusions of the kidney shape in ultrasound volumes.

Remark 1.5. Although more versatility in the dataset could be achieved by increasing the number of ultrasound volumes in the dataset, there will always remain gaps in the combination of deformations and occlusions of the kidney shape that the dataset would provide.

Thus, in this thesis, a 3D ultrasound image simulator is designed and used, mimicking actual 3D ultrasound images of the RUQ view, in which the alignment, posture, and occlusion rate of the kidney
Figure 1.13: This figure shows three examples of simulated ultrasound volumes. \( \theta_z, \vec{X}_c = [x_c, y_c, z_c] \), and \( \lambda_{oc} \) are rotation over the z-axis, the center point of the kidney shape, and visibility ratio, respectively.

shape is adjustable. The simulator uses the following factors to generate a simulated ultrasound volume of the Morison’s pouch (RUQ) view:

1. a kidney shape model (which will be introduced in 2.2.2) is used to model the kidney’s anatomical structure;

2. desired deformation and occlusion are applied on the kidney shape model.

3. the diaphragm, which appears as a bright curve in the Morison’s pouch view, is modeled as a hemisphere attached to the kidney’s shape.

4. speckle noise and random intensity profile variability are added to model the ultrasound-specific challenges (see Sec. 1.5).

The simulator’s inputs are orientation parameters, \([\theta_x^{sim}, \theta_y^{sim}, \theta_z^{sim}]\), scaling parameter, \(s^{sim}\), translation parameters, \([t_x^{sim}, t_y^{sim}, t_z^{sim}]\), and kidney’s shape visibility ratio, \(\lambda_{oc}\) where \(0 \leq \lambda_{oc} \leq 100(\text{percent})\), and the simulator output is a simulated ultrasound volume. Figure 1.13 shows three examples of simulated ultrasound volumes.

1.7 Prior Works

In this section, prior works related to computer-assisted FAST examination are reviewed. In general, the related works to computer-assisted FAST examination can be categorized into internet-based FAST and computer-assisted FAST. The former uses internet to apply a remote supervision of an expert physician for a paramedic to conduct FAST examination, and the latter uses computerized algorithms to assist
Chapter 1. Introduction

FAST examiners to conduct trauma diagnostic procedure. In the following, brief reviews on related works to these two category are presented.

1.7.1 Internet-based FAST

The use of internet to provide remote medical services, also called tele-medicine, has been increased during the recent years [43–45]. One application of internet in medicine is tele-sonography for FAST examination, in which ultrasound images acquired by paramedics in pre-hospital areas are sent to referral centers over the internet for taking remote medical decisions. In the referral centers, the received images are live-streamed for specialists to remotely guide paramedics to conduct diagnosis of trauma patients in emergency situations [46–49]. Boniface et al. [46] assessed the utility of tele-sonography to remotely guide FAST examination. In this research, a paramedic and a model patient were placed in an examination room, while a FAST expert in another room was viewing the acquired ultrasound images and providing verbal commands to modify the ultrasonic probe placement. In this research, four views including the right upper quadrant (RUQ), left upper quadrant (LUQ), subxiphoid view of the heart, and transabdominal pelvic views have been examined [46]. The paramedic was given an instruction to initially place the ultrasound probe on the intersection of subxiphoid horizontal line and the right midaxillary line. When the FAST expert felt the image has an adequate quality, the acquired image was recorded. In another research, Ogedegbe et al. [48] developed a mobile FAST examination to be used in ambulances, in which two ultrasound-trained physicians conducted the FAST scan, and the acquired images were sent to a remote FAST expert. The tele-sonography FAST platform of [48] comprised of two-way voice and one-way video communication to send real-time ultrasound images to a remote expert. The system supported 3G and 4G LTE connectivity to transfer acquired images to the referral center [48].

Ito et al. [50] reported a more advanced design for tele-FAST examination. In [50], a wearable robot was designed to conduct FAST examination in remote areas. The tele-senography robot system, with four degrees of freedom (4-DOF), was equipped with ultrasonic probes at FAST-related positions on the patients’ bodies, and it provided a fine-tuning capability of the probes alignments by a remote physician. Thereby, physicians were able to remotely conduct FAST examination in emergency situations.

Although there are indications of success in the use of tele-sonography for FAST examination, the following limitations reduce its utility in emergency situations:

- tele-sonography requires fast and reliable internet connectivity, which is not usually accessible in remote regions;

Table 1.1: Summarizing advantages and disadvantages of prior arts on computer- and internet-assisted FAST examination.

<table>
<thead>
<tr>
<th>Authors</th>
<th>Tele-echo</th>
<th>Computer-assisted</th>
<th>Methodology</th>
<th>Advantages</th>
</tr>
</thead>
<tbody>
<tr>
<td>Boniface et al. [46]</td>
<td>✓</td>
<td></td>
<td>Using two-way voice channel and one-way video channel to transfer vocal commands/feedback from expert and ultrasound images taken by paramedic.</td>
<td>Can be used to examine all views related to FAST examination. - Requires dedicated physicians to remotely conduct trauma diagnosis,</td>
</tr>
<tr>
<td>Ogedegbe et al. [48]</td>
<td>✓</td>
<td></td>
<td>Two-way voice and one-way video communication to send real-time ultrasound images to the remote expert using satellite, 3G, 4G, or LTE connectivity.</td>
<td>- Wider range of support in remote regions, - Can be used to examine all views related to FAST examination. - Very expensive data transfer,</td>
</tr>
<tr>
<td>Ito et al. [50]</td>
<td>✓</td>
<td>✓</td>
<td>Using a wearable robot echographer controlled by a remote physician at referral center, and transferring ultrasound images and control signals of the robot over the internet</td>
<td>- Remote physicians have full control on FAST examination - Requires a heavy robot to be placed on the patient’s body, - Requires a reliable internet connectivity, - Needs a dedicated physician at a referral center</td>
</tr>
<tr>
<td>Vosburgh et al. [52]</td>
<td>✓</td>
<td></td>
<td>Registering ultrasound images on CT images using real-time ultrasonic probe tracker to guide FAST examiners</td>
<td>- Provides easier interpretation of acquired ultrasound images for FAST examiners. - Does not rely on internet connectivity. - Requires many facility installment as well as CT scan of patients.</td>
</tr>
</tbody>
</table>

- a free expert might not be available to remotely conduct diagnosis when a patient requires triage.

Therefore, tele-Sonography is not always reliable for FAST examination in emergency situations, especially when a large number of patients are in the urgent need of triage (eg. natural disasters and massive vehicular collisions [51]).

### 1.7.2 Computer-assisted FAST

To our best of knowledge, the only computer-assisted FAST examination was reported in Vosburgh et al. [52]. Vosburgh et al. [52] utilized a computer-assisted system which provided an image-registered FAST (IRFAST) examination using CT scans, aiming to improve accuracy and reliability of trauma diagnosis. In the first step, a CT scan volume was acquired from a patient, and then anatomical components were segmented in the CT volume. The segmentation task was performed using prior anatomical information of organs’ shapes of interest. During the trauma diagnosis operation, the ultrasonic probe alignment
Chapter 1. Introduction

was estimated and tracked by a positioning sensor, and the probe alignment was used to register the ultrasonic anatomical view on the CT anatomical model [52]. This technique helped FAST examiners to easily find optimal positions of the ultrasonic probe. The need to install facilities and additional sensors makes this approach to be only useful in resuscitation rooms, however it is not applicable to be used in remote emergency situations.

1.8 Contributions

This PhD thesis introduces computer-assisted algorithms to be used in a computer-assisted trauma diagnostic system. The system-level contributions of this thesis are listed below,

- Designing a computer-assisted ultrasonic probe placement to be used in conjunction with a 3D ultrasound research platform to assist operators during trauma diagnosis,
- Proposing a processing pipeline to automatically detect and segment the right kidney shape,

As theoretical contributions of this thesis, we introduce two kidney detection methods, including shape-based and atlas-based kidney detection methods, and an automated kidney segmentation approach. The kidney detection methods are designed to answer two questions: (a) does the kidney shape exist in each input ultrasound volume, and (b) if the kidney shape exists, what is alignment in the ultrasound volume. The shape-based kidney detection method uses shape prior knowledge to detect the kidney shape from non-kidney structures in ultrasound volumes. The atlas-based kidney detection method uses both shape and texture prior knowledge to detect the kidney shape. We introduce a fast method to reduce speckle interference in 3D ultrasound images, called the mixture of filters based on local variation (MFLV). We propose a new approach to create shape model of abdominal organs in 3D ultrasound volumes, which engages complex-valued numbers to implicitly model voxels belonging to the structural regions of the kidney shape, and it is called complex-valued implicit shape-model (CVISM). The CVISM is used to generate a kidney shape model, containing its anatomical regions. We also propose an approach based on artificial neural network classifiers to classify voxels into background and structural regions of the kidney shape, and it is called spatially aligned neural networks (SANNs). To detect the kidney shape, we propose a shape-to-volume registration method based on a new similarity metric, called the regularized complex normalized cross-correlation (RCNCC), which finds the best alignment of the kidney shape model on the image data. The RCNCC metric provides two features: (a) robustness against partial occlusion of the kidney shape, and (b) a high specificity to the actual kidney shape within non-kidney structures in ultrasound volumes. This thesis also introduces an automated
approach, based on the level-set framework, to segment the kidney shape into its structural regions. The proposed kidney segmentation method of this thesis is called the complex-valued regional level-set approach via shape prior (CVRLS-SP). The theoretical contributions of this thesis are summarized as,

- Introducing shape-based and atlas-based kidney detection methods, including the following parts:
  - a 3D shape-based kidney detection method based on shape-to-volume affine registration using a new similarity metric, regularized complex normalized cross-correlation (RCNCC),
  - a 3D atlas-based kidney detection method based on spatially aligned neural networks (SANNs),
  - a new complex-valued implicit shape-model (CVISM) to represent abdominal organs with its structural regions for 3D ultrasound imagery,
  - a fast speckle reduction method, called the mixture of filters based on local variation (MFLV), for 3D ultrasound images,

- Introducing a new segmentation method, so-called the complex-valued regional level-set approach via shape prior (CVRLS-SP), to automatically segment the kidney shape into its structural regions in 3D ultrasound images.
1.9 Thesis Organization

The organization of this thesis is as follows. In chapter 2, the shape-based and atlas-based kidney detection methods are introduced. Chapter 3 represents the automated kidney segmentation method. Finally, chapter 4 provides discussions and conclusions on the thesis.
Chapter 2

Kidney detection
2.1 Introduction

As discussed in 1.4, kidney detection is the most important part of the proposed computer-assisted solution for trauma diagnosis. In this thesis, kidney detection is referred to the process of searching for the kidney shape in a 3D ultrasound image to answer the following questions:

1. does the kidney shape exist?

2. if the kidney shape exists, what is its alignment (orientation/scale/translation) with respect to a reference alignment in a volumetric ultrasound image?

In this thesis, the kidney detection module is required for two purposes:

1. estimating ultrasonic probe misalignment to provide probe navigational commands to guide paramedics in conducting FAST scan;

2. initializing kidney segmentation close to the actual kidney shape in ultrasound volumes to remove the need of operator’s interaction for manual initialization of the segmentation process. This provides a fully automated kidney segmentation approach.

The task of organ detection in medical images has been recognized as the main challenge toward designing computer-assisted medical diagnosis solutions [53–57]. Specifically, this task becomes more difficult when the imaging modality is sonography, because its imaging resolution and quality are lower than other imaging modalities, such as MRI and CT. In particular, kidney detection in 3D ultrasound images faces the following challenges:

• the clarity of kidney shape is reduced by speckle interference;

• the boundaries of kidney shape is not clear due to low-contrast intensity profile;

• similar tissue layers of the kidney shape may appear inhomogeneously;

• the kidney shape might be partially occluded due to either ultrasonic probe misalignment or shadows of the ribs projected over the kidney shape.

The task of kidney segmentation in 2D ultrasound images has been investigated by many researchers, such as Xie et al. [58], Wu and Sun [59], Hafizah and Supriyanto [60], and Huang et al. [61]. Xie et al. [58] proposed a 2D kidney segmentation method using texture and shape priors. In their method, features from inside and outside of the segmented contour are extracted with Gabor filters. A kidney shape model was then generated using a set of training kidney shapes, based on the so-called point
Table 2.1: Summarizing advantages and disadvantages of state-of-the-art of kidney detection in 3D ultrasound images.

<table>
<thead>
<tr>
<th>Authors</th>
<th>Methodology</th>
<th>Advantages</th>
<th>Disadvantages</th>
</tr>
</thead>
<tbody>
<tr>
<td>Noll et al. [12]</td>
<td>- Applies volume enhancement, - uses radial rays to find the center point of the kidney shape.</td>
<td>- Robust against speckle and low-contrast intensity profile</td>
<td>- Low specificity: may wrongly detect other structures</td>
</tr>
<tr>
<td>Marsousi et al. - EMBC14 [63]</td>
<td>- Applies volume enhancement, - uses template matching algorithm to fit a kidney shape model on image data.</td>
<td>- Low computational cost, - applies shape information</td>
<td>- Not robust against kidney deformation</td>
</tr>
<tr>
<td>Ardon et al. [64]</td>
<td>- Convolving a learnt-based kernel with different poses with input image, - using a SVM classifier to decide which pose/alignment of kernel provides a better matching with the kidney shape in an ultrasound volume.</td>
<td>- Using a classifier based on texture knowledge to detect the kidney bounding box.</td>
<td>- A large set of training ultrasound volumes is required to train SVM classifier, - not capable to decide whether a 3D ultrasound image contains the kidney’s shape or not.</td>
</tr>
</tbody>
</table>

Three works have reported kidney detection in 3D ultrasound imagery [12, 63, 64]. Noll et al. [12] proposed a method to detect and segment the kidney’s shape in 3D ultrasound volumes. In their method, ultrasound volumes were first enhanced to reduce speckle noise and to improve intensity contrast. Then, a deformable model method based on radial rays was used to detect the kidney’s shape. Marsousi et al.-EMBC14 [63] proposed a shape-based kidney detection/segmentation method. A kidney shape model was generated using a set of manually segmented kidney shapes in a few training 3D ultrasound images. First, each input 3D ultrasound image was preprocessed to be enhanced, and then, a template matching algorithm was applied to search for the best fit of the shape model on the enhanced image. If the matching value was greater than a specific threshold, the method decided that the kidney’s shape...
Table 2.2: Parts of the proposed \textit{shape-based} and \textit{atlas-based} kidney detection methods.

<table>
<thead>
<tr>
<th>Method</th>
<th>Shape model</th>
<th>Preprocessing</th>
<th>Voxel classification with SANNs</th>
<th>Shape-to-volume registration</th>
<th>Decision making</th>
</tr>
</thead>
<tbody>
<tr>
<td>Shape-based</td>
<td>✓</td>
<td>✓ ✓ ✓ ✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Atlas-based</td>
<td>✓</td>
<td>✓ ✓ ✓ ✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>

existed in the 3D ultrasound image, or not. The applied template matching in [63] might not be robust against kidney shape deformations. Ardon \textit{et al.} [64] proposed a method to automatically segment the kidney shape in 3D ultrasound images. Their method adopted kidney detection to automatically extract a bounding box of the kidney shape, and the extracted bounding box was used to initialize kidney segmentation based on an implicit deformation framework. For kidney detection in an input 3D ultrasound image, a learnt-based kernel with different poses was convoluted with the image to find a good matching with the kidney shape of the image. A SVM classifier was trained and used to decide which pose and alignment of the convolution kernel provided a better matching with the kidney shape in a 3D ultrasound image. The methodology, advantages, and disadvantages of the state-of-the-art are summarized in Table 2.1. Because of the shortfalls of these three methods described in Table 2.1, they did not satisfy the requirements of the targeted medical application of this thesis. Therefore, in this research, we decided to design a new kidney detection method to address the challenges, discussed in Sec. 1.5, and to meet the requirements of real-time kidney detection in 3D ultrasound images.

In this thesis, two methods are introduced to detect the kidney shape in 3D ultrasound images. The first method, which is named \textit{“shape-based”} kidney detection method, uses statistical knowledge of the kidney shape to perform kidney detection. The second method, which is named \textit{“atlas-based”} kidney detection method, uses both texture information and statistical representation of the kidney shape to perform kidney detection. We will show through experimentation that the \textit{atlas-based} method provides higher accuracies of detecting the kidney shape and estimating its orientation and alignment, compared to the \textit{shape-based} kidney detection method. However, the computational cost of the \textit{atlas-based} kidney detection method is higher than the \textit{shape-based} kidney detection method. In this thesis, we represent both the \textit{shape-based} and \textit{atlas-based} kidney detection methods because of the following reasons:

- the \textit{atlas-base} method needs a very powerful portable computer, supporting an expensive graphical processing unit (GPU) and an extreme Intel processor, to perform the kidney detection task in a real-time fashion, whereas the \textit{shape-based} kidney detection method can be run on a less expensive hardware setting.
- we will show how the combination of both the \textit{shape-based} and \textit{atlas-based} kidney detection methods
can be used to provide a reliable computer-assisted probe placement for ultrasound scan of the RUQ view.

The parts of shape-based and atlas-based kidney detection methods are shown in Table 2.2. As shown in Table 2.2, both methods are common in many parts including shape model, speckle reduction, shape-to-volume registration, and decision making. The atlas-based kidney detection method uses prior texture knowledge of the kidney shape to classify voxels into kidney and non-kidney voxels using spatially aligned neural network classifiers \(^1\) (SANNs). In the shape-based kidney detection method, 3D ultrasound images are enhanced to improve discrimination between tissue and non-tissue voxels. The higher computational cost of the atlas-based kidney detection method attributes to extracting texture features and applying multiple neural network classifiers.

In the rest of this chapter, all the parts (Table 2.2), including shape model, speckle reduction, contrast enhancement and thresholding, voxel classification with SANNs, shape-to-volume registration, and decision making, are introduced in details. Then, we represent processing pipelines of the shape-based and atlas-based kidney detection methods. Finally, experimental results, conclusions, and remarks are provided.

### 2.2 Shape model

#### 2.2.1 Introduction

Using shape prior knowledge has shown to be effective for robust object segmentation in medical images [66]. By definition, shape model in medical applications refers to a mathematical representation of anatomical knowledge of an internal organ or structure of interest. In general, mathematical representations of shape models used in literature can be categorized into explicit representation, implicit representation, and label-based representation. In the explicit representation, a shape model is explicitly represented by a set of voxels, in which each voxel has a spatial distribution in the space. This shape representation is quite popular, and is known as the point distribution model (PDM) [67]. In the implicit representation, a shape model is defined by signed distance maps, in which voxels belonging to the shape model receive positive values and voxels outside the shape model receive negative values. This approach of shape model representation, also known as level-set approach, is topologically flexible and can be used to represent detached structures [68]. In the label-based representation, labels are assigned for each region of a multi-regional structure or organ, and the probability that a voxel receives a specific

\(^1\)Neural Network classifier refers to a non-linear classifier which constitutes of inter-connected neurons in different layers and classifies input signals into output classes [65].
label from a set of training data is calculated and assigned to the voxel in the shape model [69]. The label-based representation is also known as the multi-atlas representation based on simultaneous truth and performance level estimation (STAPLE) [70].

In the explicit representation, a shape model in the 3D space is defined by a set of \( N_X \) points, also known as landmarks. Landmarks are usually selected on the object’s boundary, such that the object shape can be depicted by connecting adjacent landmarks with straight lines or curves. Assume a set of training shapes exists, such that each training shape is explicitly outlined by \( N_X \) points in the 3D space (i.e. the 3D Cartesian grid) as \( \mathbf{x}_i = [x_{i1}, x_{i2}, \cdots, x_{iN_x}, y_{i1}, y_{i2}, \cdots, y_{iN_x}, z_{i1}, z_{i2}, \cdots, z_{iN_x}] \). Thus, each \( \mathbf{x}_i \)'s dimension is \( 3N_x \). For \( N_{tr} \) training shapes, the mean shape, \( \bar{x} \), is simply formed by calculating the average of the training shapes as follows,

\[
\bar{x} = \frac{1}{N_{tr}} \sum_{i=1}^{N_{tr}} \mathbf{x}_i. \quad (2.1)
\]

Then, the covariance matrix, \( S \), of the training shapes are calculated as follows,

\[
S = \frac{1}{N_{tr} - 1} \sum_{i=1}^{N_{tr}} (\mathbf{x}_i - \bar{x})(\mathbf{x}_i - \bar{x})^T. \quad (2.2)
\]

Then, singular value decomposition\(^2\) (SVD) is performed on \( S \) to extract its eigen-values and eigen-vectors \( (v_m) \), where \( m \in [1, \ldots, 3N_x] \). \( c \) dominant eigen-vectors \( (c << 3N_x) \) are selected, as modes of variations. Then, the shape model is represented as a linear combination of the mean shape and the modes of variations as follows [72],

\[
\mathbf{x} = \bar{x} + \sum_{m=1}^{c} w_m \cdot v_m, \quad (2.3)
\]

where \( w_m \)'s are linear weights. \( c \) is chosen such that the summation of the selected eigen-values (corresponding to the selected eigen-vectors) reaches a certain ratio of the total summation of eigen-values. This ratio is usually selected between 0.9 to 0.98 [73]. Although the explicit representation is very popular for shape modeling in medical images, it suffers from numerical instability, inability of capturing high curvature locations, inflexibility in topological changes, and difficulty in selecting landmarks in 3D space [74].

In implicit representation, a shape model is defined by a signed distance function\(^3\), \( \phi_s \), such that

---

\(^2\)Singular value decomposition in linear algebra is a process of factorization of a matrix \( M \) into the form \( U \Sigma V^T \), where \( U \) is a unitary matrix, \( \Sigma \) is a rectangular diagonal matrix with eigen-values on diagonal entries, and \( V \) is a unitary matrix of eigen-vectors [71].

\(^3\)Signed distance function in image processing refers to the distance of a given point \( \tilde{X} \) in an image domain, \( \Gamma \), from an object boundary, with the positive sign determining \( \tilde{X} \) is inside the boundary, negative sign determining \( \tilde{X} \) is outside the boundary, and zero-level determining \( \tilde{X} \) is on the object boundary [75].
Voxels belonging to the shape model receive positive values $\phi_s(\vec{X}) \geq 0$, where $\vec{X}$ specifies coordinate of a voxel in the 3D Cartesian grid. Assume a set of training shapes exists, such that each training shape is defined by a separate signed distance (level-set) function, $\phi_{s,i}$, where $i \in [1, \cdots, N_{tr}]$. The mean level-set function, $\overline{\Psi}$, is calculated by the average of the training level-set functions as,

$$\overline{\Psi} = \frac{\sum_{i=1}^{N_{tr}} \phi_{s,i}}{N_{tr}}. \quad (2.4)$$

Mean-offset functions, $\{\hat{\phi}_{s,1}, \hat{\phi}_{s,2}, \cdots, \hat{\phi}_{s,N_{tr}}\}$, are calculated by reducing the mean level-set function from each training level-set function, $(\hat{\phi}_{s,i} = \phi_{s,i} - \overline{\Psi})$. Then, each mean-offset function is vectorized, $\{\vec{f}_{s,i} = \text{vectorize}(\hat{\phi}_{s,i})|i \in [1, 2, \cdots, N_{tr}]\}$, and the vectors are horizontally concatenated to form the shape-variability matrix, $S = [\vec{f}_{s,1}, \vec{f}_{s,2}, \cdots, \vec{f}_{s,N_{tr}}]$. Then, eigen-vectors and eigen-values of $S$ are calculated using singular value decomposition (SVD), and $c$ most significant eigen-vectors, corresponding to the greater eigen-values, are selected and reshaped back from the vector form into the original domain of training shapes, $\{\psi_1, \psi_2, \cdots, \psi_c\}$. Finally, a shape model is represented as a linear combination of the shapes of variations and the mean level-set function as,

$$\Psi = \overline{\Psi} + \sum_{m=1}^{c} w_m \cdot \psi_m, \quad (2.5)$$

where $w_m$s are linear weights. $c$ is chosen such that the summation of the selected eigen-values (corresponding to the selected eigen-vectors) reaches a certain ratio of the total summation of eigen-values. The implicit representation provides topological flexibility and capability of modeling sharp edges. However, this method suffers from high computational cost, and inability to represent multi-regional structures and organs of interest.

In label-based (STAPLE) representation, a shape model is defined by a label map, $L(\vec{X})$, which assigns a label $l \in [0, 1, \cdots, N_l]$ to each voxel, where $N_l$ is the number of regions. The label, $l = 0$, specifies voxels belonging to the background. Assume a set of training data exists, such that each one is defined as a label map, $L_i(\vec{X})$, and their combination is $L = \{L_1, L_2, \cdots, L_{N_{tr}}\}$, where $N_{tr}$ is the number of training label maps. Then, the label maps are used to calculate a probability for each voxel to be labeled as $L(\vec{X}) = l$ in the shape model, using the following formulation [76],

$$p(L(\vec{X}) = l|L) = \frac{p(L(x) = l) \prod_{i=1}^{N_{tr}} p(L_i(\vec{X})|L(x) = l)}{\sum_{l'=1}^{N_l} p(L(x) = l') \prod_{i=1}^{N_{tr}} p(L_i(\vec{X})|L(x) = l')}, \quad (2.6)$$

where $p(a|b)$ is the conditional probability of the event $a$, given the event $b$. The label-based repre-
Table 2.3: Summarizing advantages and disadvantages of different representations of shape models.

<table>
<thead>
<tr>
<th>Representation Name</th>
<th>References</th>
<th>Advantages</th>
<th>Disadvantages</th>
</tr>
</thead>
<tbody>
<tr>
<td>Explicit representation (PDM)</td>
<td>[67,72]</td>
<td>- providing closed-form representation of object’s boundary</td>
<td>- numerical instability, - inability of capturing high curvature locations, - inflexibility in topological changes, - selection of 3D landmarks is tedious</td>
</tr>
<tr>
<td>Implicit representation</td>
<td>[68,75]</td>
<td>- topologically flexible, easy modeling of sharp edges</td>
<td>- high computational cost, - inability to represent multi-regional organs</td>
</tr>
<tr>
<td>Label-based (STAPLE) (multi-Atlas multi-region)</td>
<td>[69,70,76]</td>
<td>- Capable to model multi-regional organs</td>
<td>- numerical instability, - high computational cost</td>
</tr>
</tbody>
</table>

sentation provides a framework to define multi-regional structures and organs, while modeling shape variations based on probabilistic representation. STAPLE provides the best representation for kidney detection, and it inspires our simplified representation of shape models for multi-regional organs, which will be introduced in the next sub-section.

### 2.2.2 Proposed shape model: complex-valued implicit shape model (CVISM)

In this thesis, a new mathematical representation, namely the complex-valued implicit shape model (CVISM), is represented for modeling multi-regional organ shape in 3D ultrasound images. In 3D ultrasound images, an organ appears with multiple structural regions, if some parts of its structure have low-echoic characteristic (dark appearance) and some others have high-echoic characteristic (bright appearance). Therefore, a representation supporting three labels including $bg$: background, $br$: bright tissue, and $dr$: dark tissue suffices for shape modeling of multi-regional organs in 3D ultrasound images. In CVISM, an implicit complex-valued representation is used such that real values are assigned to bright (or dark) regions of an organ of interest, imaginary values are assigned to dark (or bright) regions of the organ of interest, and zero magnitudes (both real and imaginary parts zero) are assigned to background (structures not belonging to the organ of interest). Although, the STAPLE shape modeling [69] can be used to represent three labels, including $bg$, $br$, and $dr$, to represent an organ’s shape model, CVISM tailors complex-valued numbers to provide easier formulation and implementation of the organ’s shape model for real-time medical diagnosis applications using 3D ultrasound imagery. Now, we focus on the kidney shape to show how the CVISM can be used to represent the kidney’s triple structural regions in 3D ultrasound images.

Consider the anatomical structure of the kidney’s shape, shown in Fig. 2.1, where three structural regions are specified, including the kidney capsule, renal medulla, and pyelocalyceal system. Voxels belonging to the renal medulla region are dark, and voxels belonging to the kidney capsule and pyelo-
calyceal system are bright. CVISM is adopted to generate a kidney shape model which includes these three structural regions. We select positive real values to represent the kidney capsule and pyelocalyceal system regions, and negative imaginary values to represent the renal medulla region.

**Remark 2.1.** There is no difference between two cases as follows:

- assigning real values to renal medulla, and imaginary values to kidney capsule/pyelocalyceal system;
- assigning imaginary values to renal medulla, and real values to kidney capsule/pyelocalyceal system.

We will explain that the output of the image enhancement module in the shape-based kidney detection method is a volumetric image which assigns positive and negative values to tissue and non-tissue voxels, respectively (sub-section 2.4.2), and the output of the voxel classification with SANNs in the atlas-based kidney detection method is a volumetric image which assigns positive and negative values to kidney and non-kidney voxels, respectively (sub-section 2.5). To match the shape model with the outputs of image enhancement and voxel classification modules, positive real values are assigned to the kidney capsule and pyelocalyceal system regions because their voxels receive positive values in the outputs of image enhancement and voxel classification modules, and negative imaginary values are assigned to the renal medulla region because its voxels appear with negative values in the outputs of image enhancement and voxel classification modules. Therefore, the kidney’s CVISM is represented as follows,

$$\Psi(\vec{X}) = \left(\psi_{PS}(\vec{X}) + \psi_{KC}(\vec{X})\right) - i \cdot \psi_{RM}(\vec{X}),$$

(2.7)

where $\psi_{PS}$, $\psi_{RM}$, and $\psi_{KC}$ are real-positive functions defining the voxels’ memberships in the pyelocalyceal system, renal medulla, and kidney capsule regions, respectively. $\vec{X}$ represents coordinate of voxels in the 3D domain of the kidney’s CVISM, ($\vec{X} \in \Omega_{\Psi}$, where $\Omega_{\Psi}$ is the 3D domain of the kidney’s CVISM).
Remark 2.2. The CVISM representation in equation (2.7) assigns positive values to voxels which are members of a region of interest, and assigns zeros to voxels which are not members of the region of interest. This representation is different from the implicit representation using the level-set framework [75], which assigns the zero-level to voxels on segmented boundaries, and assigns positive and negative values to voxels inside and outside segmented regions, respectively.

To generate the kidney’s structural regions, the shape model of the renal medulla region, $\psi_{RM}$, is generated using a set of training shapes, and then, it is used to form the shape models of kidney capsule and pyelocalyceal system regions, $\psi_{KC}$ and $\psi_{PS}$.

Assume a set of training shapes are selected from $Ds3DUsIs-GTs$, corresponding to the with-kidney images of $Ds3DUsIs-AUVs-Tr$ (see Sec 1.6.1). The training shapes are defined as, $\{G_{V,i}|i \in [1, 2, \cdots, N_{tr}]\}$, where $N_{tr}$ is the number of training shapes. Each training shape is a binarized volume, so that $G_{V,i}(\vec{X}) \in \{0, 1\}$ and $\vec{X} \in \Omega_V$, where $\Omega_V$ is the 3D domain of the volumetric ultrasound images. The training shapes are used to generate a mean-shape, representing the average of renal medulla’s shape variation. In the training shapes, the delineated renal medulla region has different alignment/orientation/scale. The training shapes should be well aligned on each other to ensure the generated mean shape only represents the statistical average of renal medulla’s shape variation of the training shapes. To achieve this goal, a reference training shape of interest is arbitrarily selected and the other training shapes are registered on the reference shape via the similarity transformation

A similarity transformation in the 3D space can be defined by seven parameters, including three rotations over the x-, y-, and z-axes ($\theta_x, \theta_y, \text{and} \theta_z$, respectively), three translations over the x-, y-, and z-axes ($t_x, t_y, \text{and} t_z$, respectively), and a scaling factor, $s$. We define a vector of similarity transformation parameters as, $\vec{p}_{ST} = [\theta_x, \theta_y, \theta_z, s, t_x, t_y, t_z]$. Now, we define the similarity transformation, $ST_{\vec{p}_{ST}}\{\vec{X}\}$, of a voxel $\vec{X}$ in the 3D space:

**Definition 2.1.** A 3D similarity transformation with a parameter vector, $\vec{p}_{ST}$, maps a voxel $\vec{X}$ into another voxel $\vec{Y}$ in the same 3D domain as follows,

$$\vec{Y} = ST_{\vec{p}_{ST}}\{\vec{X}\} = M(\theta_x, \theta_y, \theta_z, s) \cdot \vec{X} + [t_x, t_y, t_z]^T \quad (2.8)$$

and,

$$M(\theta_x, \theta_y, \theta_z, s) = S \cdot M_{\theta_x} \cdot M_{\theta_y} \cdot M_{\theta_z} \quad (2.9)$$

where $S$ is a diagonal matrix with diagonal elements equal $s$. $M_{\theta_x}$, $M_{\theta_y}$, and $M_{\theta_z}$ are 3D rotation
matrices over the $x$-, $y$-, and $z$-axes, respectively [77].

Assume a reference shape is selected as $G_{V,\text{ref}}$, and we want to align a training shape $G_{V,i}$ on $G_{V,\text{ref}}$. This is performed through a shape-to-shape registration process using a global deformation based on the similarity transformation. The problem is to find a vector of similarity transformation parameters, $\bar{p}_{st,i}$, such that it deforms $G_{V,i}$ to maximize the similarity between $G_{V,i}$ and $G_{V,\text{ref}}$. Therefore, this problem is formulated as a searching process for a sub-optimal $\bar{p}_{st,i}$ using the following non-convex minimization problem,

$$\bar{p}^*_{st,i} = \min_{\bar{p}_{st}} \text{Err}(G_{V,\text{ref}}, G_{V,i}, \bar{p}_{st}),$$

and,

$$\text{Err}(G_{V,\text{ref}}, G_{V,i}, \bar{p}_{st}) = \int \int \int_{X = [x,y,z] \in \Omega_V} (G_{V,\text{ref}}(X) - G_{V,i}(ST_{\bar{p}_{st}}\{X\}))^2 dx dy dz.$$  (2.11)

$\int \int \int_{D} \{ \cdot \} dx dy dz$ is the triple-integral, where $D$ specifies the 3D domain of the integral. $\text{Err}$ is the error function between the reference and source shapes, and $\bar{p}^*_{st,i}$ is a sub-optimal solution of the shape-to-shape registration, as a vector of similarity transformation parameters. Equation (2.10) is iteratively solved based on the Gradient descent method, in which each iteration consists of two steps as follows,

- updating translation parameters $(t_x, t_y, t_z)$ such that the mass center of the target shape is placed on the mass center of the reference,
- updating scaling and rotation parameters $(s, \theta_x, \theta_y, \text{and} \theta_z)$.

The iterative process stops when the stopping criteria, $Err_{itr} - Err_{itr-1} < \epsilon$, is achieved, where $Err_{itr}$, $Err_{itr-1}$, and $\epsilon$ are the current error function, the previous error function, and the stopping threshold. The output of this process is a registered shape, $G_{V,i}^{\text{reg}}$, where $G_{V,i}^{\text{reg}}(X) = G_{V,i}(ST_{\bar{p}^*_{st,i}}\{X\}), \forall X \in \Omega_V$.

The block diagram of registering a training shape on the reference shape is shown in Fig. 2.2.

The mean shape of the renal medulla region, $\psi_{RM}$, is generated by averaging the registered training shapes, $\{G_{V,i}^{\text{reg}}| i \in [1,2,\cdots,N_{tr}]\}$, as follows,

$$\psi_{RM}(X) = \frac{1}{N_{tr}} \cdot \sum_{i=1}^{N_{tr}} G_{V,i}^{\text{reg}}(X), \forall X \in \Omega_V.$$  (2.12)

After generating the mean shape of the renal medulla region, $\psi_{RM}$, it is used to generate the shape models of the kidney capsule and pyelocalyceal system regions. The pyelocalyceal system region is

---

5Shape-to-shape registration is referred to a process that deforms a shape (source) to maximize its similarity with the second shape (reference).

6Global deformation is referred to a transformation that is similarly applied on all voxels in the image domain.
encompassed by the renal medulla region. To generate the shape model of the pyelocalyceal system, we apply a morphological operator \textit{imfill} [78] to fill the encompassed region by the renal medulla, \( \psi_{\text{fill}} \). Then, the filled region, corresponding to the pyelocalyceal system region, is extracted by \( \psi_{\text{fill}} - \psi_{\text{RM}} \). The voxels of the pyelocalyceal system region are set to \( \psi'_{\text{PS}} \). To create the kidney capsule region, we utilize another morphological operator, \textit{imdilate}, which dilates \( \psi_{\text{fill}} \) to obtain \( \psi_{\text{dil}} \). The kidney capsule is then achieved by \( \psi_{\text{dil}} - \psi_{\text{fill}} \), and its voxels are assigned with \( \psi'_{\text{KC}} \). To avoid any bias of the shape model toward a specific region, the triple regions, including \( \psi_{\text{RM}}, \psi'_{\text{KC}}, \text{and} \psi'_{\text{PS}} \), should have the same integral of magnitudes. Therefore, we select the integral of magnitudes of the renal medulla region as the reference, and the other two regions are adjusted as follows,

\[
\psi_{\text{PS}}(\vec{X}) = \psi'_{\text{PS}}(\vec{X}) \cdot \frac{\iiint_{\vec{Y}=[x,y,z]\in\Omega_{\Phi}} \psi_{\text{RM}}(\vec{Y})dxdydz}{\iiint_{\vec{Y}=[x,y,z]\in\Omega_{\Phi}} \psi'_{\text{PS}}(\vec{Y})dxdydz}, \forall \vec{X} \in \Omega_{\Phi},
\]

and,

\[
\psi_{\text{KC}}(\vec{X}) = \psi'_{\text{KC}}(\vec{X}) \cdot \frac{\iiint_{\vec{Y}=[x,y,z]\in\Omega_{\Phi}} \psi_{\text{RM}}(\vec{Y})dxdydz}{\iiint_{\vec{Y}=[x,y,z]\in\Omega_{\Phi}} \psi'_{\text{KC}}(\vec{Y})dxdydz}, \forall \vec{X} \in \Omega_{\Phi},
\]

where \( \Omega_{\Phi} \) is the 3D domain of the kidney’s shape model (CVISM). By placing the generated \( \psi_{\text{RM}}, \psi_{\text{KC}}, \text{and} \psi_{\text{PS}} \) in equation (2.7), the kidney’s shape model is obtained. The block diagram of generating the

Figure 2.2: This figure shows the block diagram of registering a training shape on the reference shape.
2.3 Speckle reduction

2.3.1 Introduction

Ultrasound imagery suffers from low signal to noise ratio (SNR) [79]. Although ultrasonographers are able to extract information from ultrasound images in the presence of noise, it is difficult for computerized algorithms to operate under such a low SNR [80]. There are three main reasons which cause ultrasound images to have low SNR, as follows:

- **Additive noise**: a transmitted ultrasonic signal is in a pulsed format and has a short duration in time, and therefore its spectrum is wide. As a result, high frequency noise can easily affect the receiving ultrasonic signal, scattered from tissue layers [79];

- **Multiplicative interference (speckle)**: the interference of energy from randomly scattered ultrasonic signals from different tissue layers results in speckle in ultrasound images [80].

- **Nonlinear distortion**: ultrasonic signal is highly distorted when traveling deep through tissue layers [79].
Since the nonlinear distortion and additive noise are insignificant compared to the multiplicative noise, the noisy signal is generally assumed to be resulted from speckle interference \[80\]. The speckle pattern in ultrasound imagery can be categorized into three classes as follows \[81\]:

- **Fully formed speckle (FFS)**: is formed by many fine randomly scattered signals within the resolution cell of the pulse-echo system, and the amplitude of the back-scattered signal can be modeled by the Rayleigh distribution (with a constant SNR of 1.92) \[82\]. Blood cells are examples of this type of scatterers \[81\].

- **Non-randomly distributed with long-range order (NRLR)**: refers to a spatially varying back-scattered signal from a finite number of scatterers, and can be modeled with K-distribution (with a SNR below 1.92). The lobules\(^7\) in liver parenchyma is an example of this type of scatterers \[83\].

- **Non-randomly distributed with short-range order (NRSR)**: refers to a spatially invariant coherent structure (such as organ surfaces and blood vessels), and the back-scattered signals can be modeled by the Rician distribution (with a SNR above 1.92) \[84\].

According to the characteristics of these categories of speckle pattern, the following points should be considered for enhancing ultrasound volumes of the Morison’s pouch view:

- A 3D ultrasound image acquired from the Morison’s pouch view is comprised by blood vessels, organ surfaces (\textit{i.e.} liver and kidney), lobules of liver, and probably blood cells of a free-fluid due to an abdominal bleeding. Thus, the intensity profile of an ultrasound volume of the Morison’s pouch view is a complicated mixture of distribution models (Fig. 2.5). Therefore, image enhancement methods, which assume distribution models for the intensity profile (\textit{e.g.} Gaussian or Rayleigh distributions), are subject to failure.

- Ultrasound volumes should be processed in each local region to classify voxels into FFS (\textit{i.e.} blood cells) and NRLR/NRSR (\textit{i.e.} blood vessels and organ surfaces). Then, replace FFS voxels by local mean values, and keep the other voxels, corresponding to NRLR/NRSR class, unaltered \[81\].

For decades, researchers have tried to reduce speckle artifact in ultrasound images \[35,37,80,85,86\]. For the sake of unity of mathematical representation of prior arts, assume original ultrasonic signal, its noisy version, and its estimated noise-free signal at a voxel in the domain of the ultrasound image are denoted by \(f\), \(g\), and \(\hat{f}\), respectively, such that \(g = f \cdot n\) where \(n\) is a multiplicative disturbance (\textit{i.e.}

\[^{7}\text{Lobules of Liver is a small division of the liver defined at the histological scale.}\]
Figure 2.5: This figure shows a with-kidney image from the actual ultrasound dataset \((Ds3DUsIs-AUVs)\) in the left side, and its image histogram in the right side.

Speckle interference. Note that most of B-mode ultrasound imaging devices perform log compression to achieve the desired dynamic range for display \([87]\). The log compression also converts the multiplicative interference term into an additive interference term, however, this transformation results in partial information lost and distortions of the intensity profile. Therefore, speckle reduction in log-compressed ultrasound images does not need to re-apply the log-transformation to convert the multiplicative term into an additive term.

Lee \([85]\) proposed the local linear minimum mean-square error (LLMMSE) method to reduce noise in digital images, where each voxel’s intensity is adjusted to follow a desirable local mean, \(\mu_d\), and desirable local variance, \(\sigma_d^2\), using the following equation,

\[
\hat{f} = \mu_d + \sqrt{\frac{\sigma_d^2}{\sigma_g^2}} (g - \mu_g),
\]

(2.15)

where \(\sigma_g^2\) and \(\mu_g\) are local variance and local mean of \(g\). Kuan et al. \([86]\) proposed an optimal filter as follows,

\[
\hat{f} = \mu_g + \frac{\sigma_f^2}{\sigma_g^2} \cdot (g - \mu_g) = \mu_g + k \cdot (g - \mu_g),
\]

(2.16)

where \(k = \frac{\sigma_f^2}{\sigma_g^2}\), and \(\sigma_f^2\) is the local variance of the original signal, \(f\). In the case that noise is uncorrelated additive, we can write \(\sigma_f^2 = \sigma_g^2 - \sigma_n^2\), where \(\sigma_n^2\) is the non-stationary noise variance. In the multiplicative noise, \(g = f \cdot n\), where \(n\) is independent from \(f\), \(\sigma_f^2\) and \(\sigma_g^2\) have the following relationship,

\[
\sigma_g^2 = \sigma_n^2 (\sigma_f^2 + \mu_f^2) + \sigma_f^2 \mu_n^2,
\]

(2.17)

and by assuming \(\mu_g = \mu_f\), \(\sigma_f^2\) can be estimated as follows,

\[
\sigma_f^2 = \frac{\sigma_g^2 - \sigma_n^2 \mu_g^2}{\sigma_n^2 + \mu_n^2}.
\]

(2.18)
Both of the Lee and Kuan et al. filters work in the same way, as they linearly combine the center voxel intensity in a filter window with the average intensity of the window, in which the mixture depends on the variation inside the moving window [35]. Some other methods, such as Forst filter [88] and Gamma maximum a Posteriori (MAP\(^8\)) [90], were designed based on a similar scheme of Lee and Kuan et al. filters. These methods suffered from at least two drawbacks as follows:

- they are sensitive to the size of the moving window, such that a too large window over-smooths the image data and blurs edges, while a too small window is not able to reduce noise;
- when a window contains edges, its variance increases, resulting in a higher smoothness. Consequently, edges of objects in the image are smoothed.

The anisotropic diffusion method was proposed by Perona and Malik [91] to preserve edges from being blurred (diffused in homogenous regions). This method was an iterative approach which starts at the noisy signal’s intensity and changes using the diffusion equation as follows,

\[
\begin{aligned}
\frac{\partial u}{\partial t} &= \text{div}(c\nabla u) = c \cdot \text{div}(\nabla u) + \nabla c \cdot \nabla u \\
\end{aligned}
\tag{2.19}
\]

where \(c\) is the diffusion coefficient at each voxel in the image domain, \(\text{div}(\cdot)\) is the divergence operator, and \(\nabla\) is the gradient operator. By setting the value of \(c\) equal to 0 at edges and 1 at homogenous regions, edges remained sharp while non-edge regions are smoothed. Two suggested functions for \(c\) were \(c = \frac{1}{1+(x/k)^2}\) and \(c = \exp[-(x/k)^2]\). Yu and Acton [35] showed Kuan’s method (eq. (2.16)) and Perona and Malik method (eq. (2.19)) are closely related by considering \(c = (1-k)\). The Perona and Malik method performs well in images which are corrupted by additive noise. However, the Perona and Malik method undesirably amplifies the strength of speckle interference in ultrasound images. Yu and Acton [35] proposed the speckle removing anisotropic diffusion (SRAD) method, which not only preserves image edges, but also enhances edges by allowing diffusion on both sides of the edges. In fact, SRAD provides an enhancement to the Perona and Malik method, such that the diffusion coefficient is itself a function of an edge detector, \(q\), for speckled imagery. \(c(q)\) and \(q\) are calculated as follows [35],

\[
c(q) = \exp[-(q^2 - q_0^2)q_0^2(1 + q_0^2)],
\tag{2.20}
\]

and,

\[
q = \sqrt{0.5((\nabla f/f)^2 - 0.0625(\nabla^2 f/f)^2)}/(1 + 0.25(\nabla^2 f/f)^2),
\tag{2.21}
\]

\(^8\)A maximum a Posteriori probability (MAP\(^8\)) estimates an unobserved quantity on the basis of empirical data [89].
where \( q_0 \) is the speckle scale function, and is iteratively calculated as \( q_0 = q'_0 \exp[-\rho t] \), where \( q'_0 \) and \( \rho \) are constant values, and \( t \) is the iteration number. \( q \) takes high and low values at edges and homogenous regions, respectively. The original SRAD method was developed for 2D ultrasound imagery, and it was extended to 3D ultrasound images by Sun et al. [37]. Later on, Krissian et al. [80] added orientation sensitivity to the SRAD method, and called it Oriented-SRAD (OSRAD). The OSRAD allows different levels of filtering across the image edges and in the direction of high curvature boundaries. Although SRAD and OSRAD have proven their effectiveness in speckle reduction in ultrasound images, their high computational complexity is a bottleneck toward real-time operations. As reported in Krissian et al., the computational time of SRAD for a 3D ultrasound image was about 30 seconds [80].

Recently, some researchers applied sparse representation via dictionary learning [92, 93] to reduce speckle interference in ultrasound imagery [94–96]. These methods train an over-complete dictionary over the patches extracted from the noisy ultrasound image, and de-noise the image through sparse coding with a similarity constraint on the smoothed image. Assume, \( D \) is an over-complete matrix with \( N_D \) columns. Each column of \( D \) is an atom, \( d_i \), with a unity norm-\( l_2 \). Also, image patches\(^9\) are extracted from an ultrasound volume, \( F = [f_1, \ldots, f_{N_f}] \), where \( \{f_j|j \in [1, \ldots, N_f], f_j \in \mathbb{R}^{n_f}\} \).

Patches are extracted using a patch extraction operator, \( F = P_{Ext}(V,n_f) \), where \( V \) is a 3D ultrasound image. Another operator is used to reconstruct a volume from patches, as \( V = V_{Rec}(F) \). According to the definition of sparse coding, we want to find a sparse vector, \( x_j \), with a few number of non-zero coefficients, such that the reconstruction error, \( \epsilon \), of an image patch, \( f_j \), is minimized, \( f_j = D \cdot x_j + \epsilon \).

The average reconstruction error of the image patches depends on how well the dictionary represents the image patches. Dictionary learning has been designed to learn atoms which perfectly represent patches of a particular image (while the dictionary might not represent well other images). By assuming that noise signal is spatially uncorrelated, dictionary learning only trains atoms on the non-noisy data of image patches. Therefore, \( D \) is expected to better representative image data rather than noise. This objective is formulated through an optimization problem as follows,

\[
< D, X > = \underset{D, X}{\arg\min} \| P_{Ext}(V) - D \cdot X \|_F + \lambda \| V_{Rec}(D \cdot X) - V \|_2^2 \quad \text{subject to} \quad \| x_j \|_0 < \Gamma_X, \ \forall j \in [1, \ldots, N_f],
\]

(2.22)

where \( \| . \|_F, \| . \|_2, \| . \|_0 \) are Frobenius norm, norm-2, and norm-zero operators, respectively, and \( \Gamma_X \) is the maximum number of non-zero coefficients. This problem is solved in two steps as follows [92],

1. using K-SVD method [93] to find \( D \) and \( X \),

2. reconstructing the smoothed image as \( V_{dn} = \frac{1}{\lambda + 1} \cdot (\lambda V + V_{Rec}(D \cdot X)) \).

\(^9\)An image patch is a vector which is generated by vectorizing a box of voxels from the image domain [92].
Table 2.4: Summarizing advantages and disadvantages of prior arts in speckle reduction in ultrasound images.

<table>
<thead>
<tr>
<th>Method</th>
<th>References</th>
<th>Advantages</th>
<th>Disadvantages</th>
</tr>
</thead>
<tbody>
<tr>
<td>LLMMSE</td>
<td>[85, 86]</td>
<td>- simplicity</td>
<td>- sensitive to the window size</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- blurring edges</td>
<td></td>
</tr>
<tr>
<td>Anisotropic diffusion</td>
<td>[91]</td>
<td>- preserving edges</td>
<td>- amplifies speckle interference</td>
</tr>
<tr>
<td>SRAD, OSRAD</td>
<td>[35, 37, 80]</td>
<td>- enhancing edge information</td>
<td>- massive computational cost</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- reducing speckle noise</td>
<td></td>
</tr>
<tr>
<td>Sparse representation via dictionary learning (SR-DL)</td>
<td>[94–96]</td>
<td>- using dictionary of atoms which are more representative to image data rather than noise - possibly preserving edges</td>
<td>- massive computational cost</td>
</tr>
</tbody>
</table>

The idea behind using sparse representation over dictionary learning for de-noising is to collect redundant information throughout an image to recover image data from uncorrelated noise. Although its utility in ultrasound image de-speckling has been shown in [94–96], no physical justification has been provided to explain why sparse coding based on dictionary learning can reduce multiplicative noise (speckle interference) in ultrasound images. In addition, the mathematical complexity of this method for speckle reduction in 3D ultrasound images is extremely high, compared to the other methods including SRAD and OSRAD, because it must go through an iterative process of over-complete dictionary learning. There are some other methods proposed to reduce speckle interference in ultrasound images, however they either suffer from high computational cost, or have weak performance in 3D ultrasound image enhancement [97]. Table 2.4 summarizes advantages and disadvantages of commonly used speckle reduction methods for enhancing 3D ultrasound images.

2.3.2 Proposed speckle reduction for 3D ultrasound images: mixture of filters based on local variation (MFLV)

In this subsection, a new fast speckle reduction method for 3D ultrasound images is introduced, which is called the mixture of filters based on local variation (MFLV). The proposed method is inspired by the SRAD method in the sense that different smoothing strengths are applied for homogenous regions and edges. The idea is to classify voxels into homogenous (characterized as FFS, such as blood cells) and edge (characterized as NRLR/NRSR, such as organ’s surface and blood vessels). After classifying voxels, homogenous voxels are strongly smoothed while edge voxels are weakly smoothed to preserve edge information.
Assume for an input 3D ultrasound image, $V \in \mathbb{R}^{s_x \times s_y \times s_z}$, with the size of $[s_x, s_y, s_z]$, we want to estimate the membership of voxels to *homogenous* and *edge* classes. To reduce the effect of speckle interference in estimating the membership of voxels, this estimation is performed in a low resolution, $V_{lr} \in \mathbb{R}^{r \cdot s_x \times r \cdot s_y \times r \cdot s_z}$, where $r$ specifies the image resolution ($r = \frac{1}{2^p}$ and $p$ must be an integer number) [98]. The local variation for each voxel, $\tilde{X}$, belonging to the low-resolution image domain, $\Omega_{V_{lr}}$, is estimated as follows,

$$\sigma^2_{rad}(\tilde{X}) = \frac{1}{(2 \cdot rad + 1)^3 - 1} \sum_{\tilde{Y} \in \text{nib}(\tilde{X}, rad)} \left( V_{lr}(\tilde{X}) - V_{lr}(\tilde{Y}) \right)^2,$$

where $rad$ specifies the radius of a box covering neighbor voxels, $\text{nib}(\tilde{X}, rad)$ is a set of voxels located inside a box of radius $rad$ and centered at $\tilde{X}$. The number of voxels located inside the box of radius $rad$ is $(2 \cdot rad + 1)^3$. Note that the summation in equation (2.23) is divided by $(2 \cdot rad + 1)^3 - 1$ to provide an unbiased estimation of local variation. By assuming that the image contains voxels belonging to both the *homogenous* and *edge* classes, the membership of voxels are obtained by dividing the local variance to the maximum local variance. By this selection, voxels with high local variations receive membership values close to 1, and homogenous voxels receive membership values close to 0. The memberships of voxels in $\Omega_{V_{lr}}$, are obtained as follows,

$$\alpha_{lr}(\tilde{X}) = \frac{\sigma^2_{rad}(\tilde{X})}{\max\left\{\sigma^2_{rad}(\tilde{X}) | \tilde{X} \in \Omega_{V_{lr}}\right\}},$$

where $\alpha_{lr}$ specifies the membership of voxels in the *homogenous* and *edge* classes in the low resolution, $r$. In the next step, the memberships of voxels in the original resolution, $\alpha(\tilde{X})$, where $\tilde{X} \in \Omega_V$, are obtained by up-sampling $\alpha_{lr}$ with the factor $\frac{1}{r} = 2^{p'}$, using linear interpolation [98].

After estimating memberships of voxels, the input volume, $V$, is convolved with two Gaussian finite impulse response (FIR) filters, with zero-mean and variances $\sigma_{low}$ and $\sigma_{high}$, in the spatial domain, and the results are $V_{\sigma_{low}}$ and $V_{\sigma_{high}}$, respectively, where $\sigma_{high} >> \sigma_{low}$. The Gaussian FIR filter with $\sigma_{high}$ results in a highly smoothed image, $V_{\sigma_{high}}$, whereas the Gaussian FIR filter with $\sigma_{low}$ provides a weak smoothness and preserves image edges in $V_{\sigma_{low}}$. Therefore, a de-speckled volume can be achieved by mixing the output of these two filtered volumes, based on the calculated memberships, $\alpha$. A voxel with a low membership value (close to 0) should be highly smoothed, and therefore, its intensity should be assigned based on the corresponding value from $V_{\sigma_{high}}$, whereas a voxel with a high membership value
Figure 2.6: This figure shows the block diagram of reducing speckle in 3D ultrasound images using the MFLV method. LPF stands for Low-Pass Filtering.

Table 2.5: Comparison of the speckle reduction methods, including the proposed method (MFLV), SRAD, and sparse representation via dictionary learning (SR-DL), in terms of quality enhancement based on the $SNR_A$ metric and computational time.

<table>
<thead>
<tr>
<th>Method</th>
<th>Region #1 $SNR_A$</th>
<th>Region #2 $SNR_A$</th>
<th>Region #3 $SNR_A$</th>
<th>$SNR_A$ Average improvement (%)</th>
<th>Computational time (sec)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SRAD (#itr: 2)</td>
<td>4.5858</td>
<td>5.1739</td>
<td>2.8118</td>
<td>9.66</td>
<td>24.14</td>
</tr>
<tr>
<td>SRAD (#itr: 4)</td>
<td>4.8258</td>
<td>5.4379</td>
<td>2.9297</td>
<td>14.98</td>
<td>49.09</td>
</tr>
<tr>
<td>SRAD (#itr: 6)</td>
<td>4.9978</td>
<td>5.5441</td>
<td>3.0083</td>
<td>18.1092</td>
<td>66.04</td>
</tr>
<tr>
<td>SR-DL</td>
<td>5.3378</td>
<td>5.6375</td>
<td>2.9037</td>
<td>20.1153</td>
<td>3339.9</td>
</tr>
<tr>
<td>MFLV</td>
<td>5.1387</td>
<td>5.9380</td>
<td>3.1182</td>
<td>23.49</td>
<td>4.96</td>
</tr>
</tbody>
</table>

(close to 1) receives its intensity from $V_{\sigma_{\text{low}}}$. The de-speckled volume is achieved as follows,

$$V_{\text{dn}}(\vec{X}) = \alpha(\vec{X}) \cdot V_{\sigma_{\text{low}}} (\vec{X}) + \left(1 - \alpha(\vec{X})\right) \cdot V_{\sigma_{\text{high}}} (\vec{X}), \quad \forall \vec{X} \in \Omega_V. \quad (2.25)$$

The MFLV method consists of six computational blocks, including (1) 3D image down-sampling, (2) calculating memberships based on local variations, (3) 3D image up-sampling (4) 3D convolution of $V_{\sigma_{\text{low}}}$, (5) 3D convolution of $V_{\sigma_{\text{high}}}$, and (6) mixing filters based on membership values. In contrast with edge preserving methods such as anisotropic diffusion and SRAD, the MFLV method does not rely on iterative optimization of partial deference equations (PDE), and therefore, the proposed method operates faster. The block diagram of the MFLV method is shown in Fig. 2.6.

The MFLV method is developed in MATLAB, and its parameters are set as $rad = 1$ and $pr = 1$ ($r = \frac{1}{2pr} = 0.5$). We compare the speckle reduction methods, including the MFLV, SRAD, and SR-DL methods, for an actual 3D ultrasound image of the RUQ view, in terms of image quality enhancement and computational time. The results of speckle reduction using the methods are shown in Fig. 2.7. Accordingly, the MFLV method shows a better improvement compared to the other methods in terms of smoothing homogenous regions while preserving the kidney shape’s details. As shown in Fig. 2.7, the smoothness of the SRA method increases by increasing the number of iterations. To quantitatively
Figure 2.7: This figure shows speckle reduction results using three methods, including MFLV (the proposed method), SRAD, and sparse representation via dictionary learning (SR-DL). Results of the SRAD method are for three different number of iterations, including $\#itr: 2$, $\#itr: 4$, and $\#itr: 6$. The bottom row displays zoom-in of the de-speckling result of a part of the kidney shape.

evaluate the image quality enhancement of the speckle reduction methods, the $SNR_A$ metric is used [97], which measures the ratio of mean to standard deviation of the intensity values for regions of interest. We calculate $SNR_A$ for the following regions:

- **Region#1**: the combination of kidney capsule and pyelocalyceal system regions, which is supposed to be bright in the ultrasound volume,
- **Region#2**: the region outside the kidney shape,
- **Region#3**: the renal medulla region, which is supposed to be dark in the ultrasound volume.

Note that these three regions are obtained from the detected kidney shape using the atlas-based kidney detection method of this thesis. The calculated $SNR_A$ of the three regions for the original ultrasound volume are $Region\#1-SNR_A = 4.2493$, $Region\#2-SNR_A = 4.5797$, and $Region\#3-SNR_A = 2.6011$. The comparison of the methods are demonstrated in Table 2.5. The $SNR_A$ of the SRAD method is
calculated for three different number of iterations, including two, four, and six. Accordingly, the proposed method (MFLV) provides a higher quality of speckle reduction for region #2 and region #3 compared to the SRAD and SR-DL methods, while the computational time of the MFLV method is lower than the other methods. Also, the MFLV method provides a higher average improvement (23.49%). The combination of speckle reduction quality and faster operation of the MFLV method, makes it more convenient for enhancing 3D ultrasound images.

2.4 Volume enhancement

The quality of the de-speckled volume, $V_{dn}$, is still low due to two reasons:

- low-contrast intensity profile of tissue layers,
- inhomogeneous intensity profile of similar tissue layers.

These two problems may reduce the separability of the kidney shape from its surrounding tissues in 3D ultrasound images. To address these problems, two preprocessing modules, namely the localized histogram equalization (LHE) and global-to-local thresholding (G2LTh), are sequentially applied on the de-speckled volume, which will be introduced in the next two sub-sections.

2.4.1 Contrast Enhancement

The localized histogram equalization (LHE) method is applied to improve the de-speckled volume’s contrast. The LHE method calculates image histogram at each voxel, and uses it to estimate the cumulative density function (CDF) of the intensity levels in the neighborhood of the voxel. Then, the estimated CDF is used to obtain local transformations for voxels’ intensity levels to expand highly concentrated intensity levels in the local histogram, and to shrink wide regions in the local histogram [99]. The output of this process is an image with an improved contrast. Note that the intensity profile of tissue layers is inhomogeneous throughout the ultrasound volume, and therefore, the localized CDF is applied to effectively enhance the image contrast. This fact is shown in Fig. 2.8, which displays the output of LHE versus the output of histogram equalization (HE).

Assume $V_{dn}$ is a de-speckled volume, and $Hist(V_{dn}, rad_2, \vec{X}, i)$, where $i \in \{0, 1, \cdots, 255\}$ is an intensity level, provides the local histogram of voxels inside a box, specified by $nib(\vec{X}, rad_2)$ with radius $rad_2$ and centered at $\vec{X}$ [99]. The local CDF of voxel $\vec{X}$ is obtained as follows,

$$CDF_{rad_2}(\vec{X}, i) = \left[ \frac{Hist(V_{dn}, rad_2, \vec{X}, i)}{(2 \cdot rad_2 + 1)^3} \right], \forall \vec{X} \in \Omega_V,$$ (2.26)
Chapter 2. Kidney detection

(a) De-speckled Volumes ($V_{dn}$)

$CSI = 1.89$

(b) Histogram equalization

$CSI = 1.82, \ t_c = 0.12(\text{sec})$

(c) LHE with $rad_2 = 1$

$CSI = 1.07, \ t_c = 0.89(\text{sec})$

(d) LHE with $rad_2 = 3$

$CSI = 1.18, \ t_c = 0.98(\text{sec})$

(e) LHE with $rad_2 = 5$

$CSI = 1.45, \ t_c = 1.16(\text{sec})$

(f) LHE with $rad_2 = 7$

$CSI = 1.76, \ t_c = 1.76(\text{sec})$

(g) LHE with $rad_2 = 9$

$CSI = 2.01, \ t_c = 2.69(\text{sec})$

Figure 2.8: This figure shows the output of contrast enhancement using the LHE method for different values of $rad_2 \in \{1, 3, 5, 7, 9\}$, compared to the HE method. $t_c$ is the computational time in seconds.

where $\lfloor \cdot \rfloor$ is the nearest integer function, and $0 \leq CDF_{rad_2}(\tilde{X}) \leq 255$ (note that in this equation, 255 is equivalent of 1 in the general definition of CDF). The denominator of equation (2.26) divides the local histogram at each point by the number of voxels inside $\text{nib}(\tilde{X}, rad_2)$. The contrast enhanced volume is then obtained using the following equation,

$$V_{LHE}(\tilde{X}) = CDF_{rad_2}(\tilde{X}, V_{dn}(\tilde{X})).$$  \hspace{2cm} (2.27)

In $V^{LHE}$, voxels corresponding to dark homogenous regions in $V$ can receive high intensity levels, which may cause incorrect decisions. To solve this, we apply voxel-wise multiplication of $V^{LHE}$ with an exponential function of $V^{dn}$, as follows:

$$V'_{LHE}(\tilde{X}) = V_{LHE}(\tilde{X}) \cdot \left(1 - \exp(-\beta \cdot V_{dn}(\tilde{X})) \right),$$  \hspace{2cm} (2.28)

where $\beta$ is the correction coefficient. The LHE method has a massive computational cost, as it requires computing the CDF at each voxel, and its computational cost is proportional to the radius of the box, $rad_2$. Though, its mathematical derivation allows us to use parallel computation to reduce its computational time. Thus, the LHE method is developed using parallelized computation based on GPU-programming (CUDA). Figure 2.8 shows the output images, the class separability index (CSI), and computational time of the LHE for $\beta = 1$ and different values of $rad_2$, compared to the HE method.
calculated CSI show the separability of the Region#1 and Region#3 (Sec. 2.3.2), and is calculated as 

\[ CSI = \frac{(\mu_3 - \mu_1)^2}{\sigma_1 \cdot \sigma_2} \]  

where \( \mu_3 \) and \( \mu_1 \) are mean intensity levels of the Region#1 and Region#3, respectively, and \( \sigma_3 \) and \( \sigma_1 \) are standard deviations of voxels’ intensity levels of the Region#1 and Region#3, respectively. The higher CSI relates to the higher separability of the Region#1 and Region#3 in the ultrasound volume. The CSI of the de-speckled volume, \( V_{dn} \), is \( CSI = 1.89 \). According to Fig. 2.8, the output of LHE with \( rad_2 = 9 \) shows an improved intensity contrast to \( CSI = 2.01 \). Thus, the LHE method with \( rad_2 = 9 \) is selected to improve the image contrast.

2.4.2 Global-to-local thresholding

The contrast enhanced volume still has an inhomogeneous intensity profile for similar tissues, which may result in incorrect kidney detection. We want to perform an operation on the contrast enhanced volume such that tissue voxels receive positive values and non-tissue voxels receive negative values, and thereby the problem of inhomogeneous intensity profile will be addressed. This is obtained by combining a global thresholding operation with a local thresholding operation, namely global-to-local thresholding (G2LTh). In the global thresholding operation, two threshold values are calculated based on the image histogram, such that voxels with the intensity levels bellow the first threshold, \( th_{low} \), belongs to the non-tissue class, and voxels with the intensity levels above the second threshold, \( th_{high} \), belongs to the tissue class. Then, voxels with intensity levels between \( th_{low} \) and \( th_{high} \) are classified into tissue and non-tissue classes using a local thresholding method which is based on localized information of the intensity profile.

The ordinary Kriging method [100] provides local thresholding using spatial covariance and indicator Kriging to locally classify voxels into two regions. Assume for a voxel \( \tilde{X} \) in the image domain, \( \Omega_V \), a set of neighbor voxels are defined as \( \{ \tilde{Y} \} \in \text{nib}(\tilde{X}, rad_3) \), where \( rad_3 \) defines the box’s radius covering the voxels in the set. By assuming that intensity levels of voxels are locally stationary and have a stationary spatial covariance in ultrasound images [101], we can write 

\[ \text{Cov}(V_{LHE}'(\tilde{Y}_k), V_{LHE}'(\tilde{Y}_j)) = C(\|\tilde{Y}_k - \tilde{Y}_j\|) \]  

where \( \tilde{Y}_j \) and \( \tilde{Y}_k \) are two voxels in \( \text{nib}(\tilde{X}, rad_3) \). For each voxel in the image domain, \( \Omega_V \), a regression model is fitted as follows [102],

\[ V_{LHE}'(\tilde{X}) = \lambda_0 + \sum_{k=1}^{N_{LT}} \lambda_k V_{LHE}'(\tilde{Y}_k) \]  

where \( N_{LT} \) is the number of voxels in \( \text{nib}(\tilde{X}, rad_3) \), which is equal \( N_{LT} = (2 \cdot rad_3 + 1)^3 \), and \( \{ \lambda_k \} \)s are coefficients of the regression model. Now, the estimation error’s variance is minimized subject to the
estimation error equals zero, by finding the coefficients $\lambda_k$s as follows [102],

$$\begin{cases}
\sum_{k=1}^{N_{LT}} \lambda_k \times C(\|Y_k - Y_j\|) + \mu = C(\|Y_k - Y_0\|), \quad \forall j \in \{1, \cdots, N_{LT}\}, \\
\sum_{k=1}^{N_{LT}} \lambda_k = 1.
\end{cases} \tag{2.30}$$

Equation (2.30) is re-written as follows,

$$\begin{bmatrix}
\lambda_1 \\
\vdots \\
\lambda_{N_{LT}} \\
\mu
\end{bmatrix} =
\begin{bmatrix}
C(\|\bar{Y}_2 - \bar{Y}_1\|) & \cdots & C(\|\bar{Y}_1 - \bar{Y}_{N_{LT}}\|) & 1 \\
\vdots & \ddots & \vdots & \vdots \\
C(\|\bar{Y}_{N_{LT}} - \bar{Y}_1\|) & \cdots & C(\|\bar{Y}_{N_{LT}} - \bar{Y}_{N_{LT}}\|) & 1 \\
1 & \cdots & 1 & 0
\end{bmatrix}^{-1}
\begin{bmatrix}
C(\|\bar{Y}_1 - \bar{Y}_0\|) \\
\vdots \\
\vdots \\
C(\|\bar{Y}_{N_{LT}} - \bar{Y}_0\|)
\end{bmatrix}. \tag{2.31}$$

To solve equation (2.31), the stationary covariance function, $C(d)$, where $d$ is the distance between two points $\bar{Y}_k$ and $\bar{Y}_j$, should be estimated for each input volume [100]. After calculating the coefficients $\lambda_k$s, the probability $P\left(th, \bar{X} | \text{nib}(\bar{X}, \text{rad}_3)\right) \equiv Pr\{V_{eh}(\bar{X} < th)\}$, where $V_{eh}$ is the desired enhanced volume, is calculated for all voxels in the image domain, for both the global threshold values, $th_{low}$ and $th_{high}$, as follows,

$$\begin{cases}
P\left(th_{low}, \bar{X} | \text{nib}(\bar{X}, \text{rad}_3)\right) = \sum_{i=1}^{n} \lambda_i \times i(th_{low}, \bar{Y}_i), \\
P\left(th_{high}, \bar{X} | \text{nib}(\bar{X}, \text{rad}_3)\right) = \sum_{i=1}^{n} \lambda_i \times i(th_{high}, \bar{Y}_i),
\end{cases} \tag{2.32}$$

and,

$$i(th, \bar{Y}) = \begin{cases}
1, & \text{if } V_{LHE}'(\bar{Y}) \leq th, \\
0, & \text{otherwise}.
\end{cases} \tag{2.33}$$

$P\left(th_{low}, \bar{X} | \text{nib}(\bar{X}, \text{rad}_3)\right)$ represents the probability that the voxel $\bar{X}$ belongs to the non-tissue class, and $1 - P\left(th_{high}, \bar{X} | \text{nib}(\bar{X}, \text{rad}_3)\right)$ represents the probability that the voxel $\bar{X}$ belongs to the tissue.
\[
L_{G2L}(\vec{X}) = \begin{cases} 
0, & \text{if } V_{LHE}(\vec{X}) \leq \text{th}_{\text{low}} \\
0, & \text{if } \text{th}_{\text{low}} \leq V_{LHE}(\vec{X}) \leq \text{th}_{\text{high}} \& P\left(\text{th}_{\text{low}}, \vec{X}|\text{nib}(\vec{X}, \text{rad}_3)\right) > 1 - P\left(\text{th}_{\text{high}}, \vec{X}|\text{nib}(\vec{X}, \text{rad}_3)\right) \\
1, & \text{if } \text{th}_{\text{low}} \leq V_{LHE}(\vec{X}) \leq \text{th}_{\text{high}} \& P\left(\text{th}_{\text{low}}, \vec{X}|\text{nib}(\vec{X}, \text{rad}_3)\right) < 1 - P\left(\text{th}_{\text{high}}, \vec{X}|\text{nib}(\vec{X}, \text{rad}_3)\right) \\
1, & \text{if } V_{LHE}(\vec{X}) \geq \text{th}_{\text{high}} 
\end{cases}
\] (2.34)

Finally, the enhanced volume is obtained using the following equation,

\[
V_{eh}(\vec{X}) = \begin{cases} 
\frac{-(\text{th}_{\text{high}} - V_{LHE}(\vec{X}))}{\text{th}_{\text{high}}}, & \text{if } L_{G2L}(\vec{X}) = 0, \\
\frac{(V_{LHE}(\vec{X}) - \text{th}_{\text{low}})}{1 - \text{th}_{\text{low}}}, & \text{if } L_{G2L}(\vec{X}) = 1.
\end{cases}
\] (2.35)

Figure 2.9 shows the output of the G2LTh method and the enhanced volume for three values of \(\text{rad}_3 \in \{1, 2, 3\}\). Accordingly, the \(\text{rad}_3 = 1\) provides a better enhancement with \(\text{CSI} = 2.1487\) and lower computational time \(t_c = 0.95\). Figure 2.10 shows steps of ultrasound volume enhancement, including speckle reduction, LHE, G2Lth, and volume enhancement output for three actual ultrasound volumes.
Chapter 2. Kidney detection

Figure 2.10: This figure shows three examples of enhancing ultrasound volumes using the proposed preprocessing. In the right figure, $V_{ch}$, red and blue colors are used to visualize positive and negative values of voxels, respectively.

Figure 2.11: This figure shows single-layer and multi-layer structures of artificial neural networks.

### 2.5 Voxel Classification with SANNs

In section 2.4, a volume enhancement approach is introduced to improve separability of tissue and non-tissue voxels. However, this method is not able to separate voxels belonging to the kidney shape from non-kidney voxels. In this section, another method to enhance ultrasound volumes is represented which employs neural network classifiers to classify voxels into kidney and non-kidney classes. We will show such a classification improves separability of voxels belonging to the kidney shape from the rest of volume, which contributes to a higher accuracy of kidney detection.

Artificial neural network (ANN) is a widely used classifier, which provides the ability of classifying non-linearly distributed data structures. The basic computational element in ANN is called artificial neuron, and consists of a set of weighted inputs, a summation operator, and an activation function. The
relation between inputs and output of an artificial neuron is defined as follows [103] (Fig. 2.11.(a)),

\[ y_i = f\left( \sum_j w_{i,j} \cdot y_j \right) \]  

(2.36)

where \( f(.) \) is an activation function, \( y_i \) and \( y_j \) are outputs of nodes \( i \) and \( j \), respectively, and \( w_{i,j} \) is a weighted connection of the node \( j \) to the node \( i \) [103]. A neural network with an input and an output layer is called a single layer neural network (Fig. 2.11.(b)). The input layer does not perform any computation, and it only distribute inputs through the output layer. In the single layer structure, each neuron at the output layer is responsible to do computations, and specify the desired output based on its weighted inputs. Multi-layer neural network can be used to provide higher level of classification complexity, by adding hidden layers to the neural network structure (Fig. 2.11.(c)). The number of hidden layers, \( N_{HL} \), is an important parameter to determine the complexity of neural network. Deep learning methods have been recently proposed to train multi-layer neural networks [104].

Voxel classification using artificial neural network has been widely used in medical image processing [105]. Such a task comprises texture-feature extraction, training a neural network classifier, and classifying voxels using the trained neural network classifier. However, due to the high variation of texture information throughout the kidney shape, a single neural network classifier might not be able to correctly classify voxels. On the other hand, using a deep neural network classifier with multiple hidden layers might also over-fit the classifier on the training data.

Zhan and Shen [106] proposed an approach to segment the prostate shape in 3D ultrasound images based on spatially placed Gabor-support vector machines (G-SVMs). According to Zhan and Shen [106], the prostate tissue may have very different texture features throughout the prostate shape, making it difficult for a single G-SVM classifier to correctly label all prostate tissue. To address this problem, they divided the prostate surface into multiple sub-surfaces, and then, a G-SVM classifier was trained and attached to each sub-surface.

In the proposed method by Zhan and Shen, the ultrasonic probe placement was estimated, and used to initiate a 3D deformable surface. Then, voxels around each sub-surface of the 3D deformable surface were labeled using the corresponding G-SVM classifier, and the label information were used to evolve the 3D deformable surface until the prostate shape was segmented. This method made some assumptions to estimate the probe placement [106], and supposed that the prostate shape existed in the ultrasound volume. This method was not able to differentiate between prostate and non-prostate structures. Because of these limitations, this method is not usable for the application of this thesis.
By inspiring from the idea of Zhan and Shen [106] to use multiple spatially placed classifiers, this thesis introduces a voxel classification scheme, consisting of multiple neural network classifiers, in which each neural network classifier is spatially aligned on the image domain, and is trained to detect voxels of a specific partition of the kidney shape. This voxel classification scheme is named spatially aligned neural networks (SANNs). This topology helps to reduce underlying texture information complexity of each neural network classifier, and improves the overall performance of voxel classification. Since each neural network classifier is spatially aligned in the image domain, and is trained to classify voxels of a particular partition of the kidney shape, it requires that the kidney shape in a query ultrasound volume has a specific alignment and orientation to match with the SANNs. However, the kidney shape can have any random alignment/orientation in ultrasound volumes due to misalignment of the ultrasonic probe. To achieve a correct performance of the SANNs, ultrasound volumes must be transformed to re-align their kidney shapes on the reference alignment/orientation of the SANNs. This task is called volume-to-shape registration. The voxel classification with the SANNs comprises the following steps:

- registering the ultrasound volume on the expected alignment of the kidney shape,
- extracting texture-features with 3D Gabor filters\(^\text{10}\),
- classifying voxels with the SANNs.

One difficulty facing this approach is to perform the volume-to-shape registration to align the kidney shape in an ultrasound volume on the desired alignment/orientation of the kidney shape. In the ideal case, the kidney shape alignment/orientation in the ultrasound volume is known, and a 3D transformation can be easily obtained to perform the desired registration task. However, in a real scenario, not only the alignment/orientation of the kidney shape is unknown, but also we are not sure whether the ultrasound volume contains the kidney shape or not. The success of this method ties on properly solving this problem.

The second problem is to properly select the size and number of SANNs classifiers. Selecting a too large size for the sub-surfaces, reduces the number of SANNs classifiers, and results in increasing the variation of the texture information for each neural network classifier, making it difficult for the classifier to correctly classify kidney and non-kidney voxels. On the other hand, selecting a too small size for the partitions increases the number of SANNs classifiers, and results in insufficient number of training samples for each classifier, causing a weak training process of the SANNs classifiers. Therefore,\(^\text{10}\)

\(^{10}\)3D Gabor filters are linear filters, which their frequency and orientation representations are similar to those of the human visual system, and they have been found to be particularly appropriate for texture representation and discrimination.
Chapter 2. Kidney detection

the number of SANNs classifiers should be appropriately selected. In the rest of this section, we introduce
two processes of training and classification.

2.5.1 Training

This subsection introduces a procedure to train the SANNs based on a set of training ultrasound volumes.
A sub-set of with-kidney volumes from the training actual ultrasound volumes (Ds3DUsIs-AUVs) is se-
lected, \( \{ V_{i}^{tr} \mid i \in [1, \cdots, N_{tr}] \} \), and one volume is selected as the reference volume, \( V_{ref}^{tr} \). Then, for each of
the training volumes, a similarity transformation with the parameters, \( \vec{p}_{st,i}^{*} \), is obtained by solving equa-
tion (2.10) (Sec 2.2.2). The training volumes are first de-speckled by the MFLV method (Sec 2.3.2), and
then the similarity transformations are applied to register the training volumes on the reference volume,
and the new set of registered volumes is achieved,
\[
\{ V_{reg}^{tr}(\vec{X}) = V_{i}^{dn}(ST_{\vec{p}_{st,i}}(\vec{X})) \mid \forall \vec{X} \in \Omega_{V} \mid i \in [1, \cdots, N_{tr}] \}
\]
(hint: see Sec 2.2.2 to find the definition of variables and functions related to the similarity transfor-
mation). We also used the introduced kidney segmentation method in Sec 3.3 to finely segment the
kidney shape of the training volumes into their structural regions, including the renal medulla and the
combination of kidney capsule and pyelocalyceal system. The kidney segmentation results are required
to generate desired labels for training the SANNs. The segmentation result of the \( i^{th} \) training volume
is, \( A_{i}^{tr} \), in which specifies the structural regions as follows,

\[
A_{i}^{tr}(\vec{X}) = \begin{cases} 
-1, & \vec{X} \in \text{renal medulla} \\
1, & \vec{X} \in \text{pyelocalyceal system or kidney capsule} \\
0, & \text{otherwise}
\end{cases}
\]  

(2.37)

The segmentation results are also transformed by their corresponding similarity transformations to obtain
the registered segmentation volumes, \( \{ A_{i}^{reg} \mid i \in [1, \cdots, N_{tr}] \} \).

The objective of training SANNs is to discriminate voxels belonging to the kidney shape from non-
kidney voxels, based on texture information of the kidney shape in ultrasound volumes. The 3D Gabor
filters are recognized as an effective tool to extract meaningful texture information from images [107].
We use the 3D Gabor filters to extract texture information from ultrasound volumes. A 3D Gabor filter
is defined as follows [107],

\[
g_j(\vec{X}) = g(\vec{X}; \theta_{y}, \theta_{z}, F, \sigma) = \hat{g}(\vec{X}) \cdot \exp \left\{ j2\pi F [\sin \theta_{y} \cos \theta_{z}, \sin \theta_{y} \sin \theta_{z}, \cos \theta_{y}] \cdot \vec{X} \right\},
\]  

(2.38)
are $16$ the texture feature volumes, as distinctive feature volumes, the total number of texture feature volumes

\[
\text{texture feature volumes have both real and imaginary parts. By separating real and imaginary parts of}
\]

\[
\{\text{The Gabor filters are applied on the registered volumes and a set of texture feature volumes are achieved,}
\]

\[
\text{set of } 16 \text{ 3D Gabor filters are selected with } F = 2, \sigma \in \{0.4, 0.6\}, \theta_z \in \{0, \frac{\pi}{4}, \frac{2\pi}{3}, \frac{3\pi}{4}\} \text{ and } \theta_y \in \{0, \frac{\pi}{2}\}.
\]

The Gabor filters are applied on the registered volumes and a set of texture feature volumes are achieved, 
\[
\{V_{i,j}^{g} = V_{i}^{reg} \ast g_j(\bar{X}) \mid i \in [1, \cdots, N_{tr}], j \in [1, \cdots, 16]\}. \quad \text{Since the Gabor filters are complex-valued, the}
\]

texture feature volumes have both real and imaginary parts. By separating real and imaginary parts of the texture feature volumes, as distinctive feature volumes, the total number of texture feature volumes are $16 \times 2 = 32$.

Assume the kidney shape is partitioned into $N_{x}^{SANN}, N_{y}^{SANN},$ and $N_{z}^{SANN}$ number of splits along the

\[
x, y, \text{ and } z\text{-axes, respectively, which makes the total number of } N^{SANN} = N_{x}^{SANN} \cdot N_{y}^{SANN} \cdot N_{z}^{SANN}
\]

partitions. Also, we consider a 50% pixel overlapping for aligning the boxes, specifying partitions of the kidney shape. For the kidney shape’s size of $W_{x} \times W_{y} \times W_{z}$ and its top-left corner point coordinated at

\[
\bar{X}_{B} = [x_{B}, y_{B}, z_{B}]^T, \text{ voxels belonging to the partitions are specified as follows}
\]

\[
B_{k_{x}, k_{y}, k_{z}} \equiv \bigcup \{ \bar{X} = [x, y, z]^T \} : \begin{cases}
    x_{B} + \frac{W_{x}(k_{x} - 1)}{N_{x}^{SANN} + 1} \leq x < x_{B} + \frac{W_{x}(k_{x} + 1)}{N_{x}^{SANN} + 1} \\
y_{B} + \frac{W_{y}(k_{y} - 1)}{N_{y}^{SANN} + 1} \leq y < y_{B} + \frac{W_{y}(k_{y} + 1)}{N_{y}^{SANN} + 1} \\
z_{B} + \frac{W_{z}(k_{z} - 1)}{N_{z}^{SANN} + 1} \leq z < z_{B} + \frac{W_{z}(k_{z} + 1)}{N_{z}^{SANN} + 1}
\end{cases}
\]

(2.40)

where $B_{k_{x}, k_{y}, k_{z}}$ is a subset of the image domain, $\Omega_{V}$, covering voxels belonging to a partition, $[k_{x}, k_{y}, k_{z}]$, in which $k_{x} \in [1, 2, \cdots, N_{x}^{SANN}]$, $k_{y} \in [1, 2, \cdots, N_{y}^{SANN}]$, and $k_{z} \in [1, 2, \cdots, N_{z}^{SANN}]$ specify the partition number along the $x$, $y$, and $z$-axes, respectively. Figure 2.12 shows a 2D version of partitioning the kidney shape for $N_{x}^{SANN} = 2$ and $N_{y}^{SANN} = 2$. In Fig. 2.12, the partition $B_{2,1}$ is highlighted as the green region.

For each partition, the subset $B_{k_{x}, k_{y}, k_{z}}$ of voxels from the texture feature volumes are extracted, 
\[
\{V_{i,j}^{g} \mid i \in [1, \cdots, N_{tr}], j \in [1, \cdots, 32], k_{x} \in [1, 2, \cdots, N_{x}^{SANN}], k_{y} \in [1, 2, \cdots, N_{y}^{SANN}], k_{z} \in [1, 2, \cdots, N_{z}^{SANN}]\}, \text{ and vectorized to form } \bar{v}_{i,j,k_{x},k_{y},k_{z}}. \text{ Then, the vectors are horizontally concatenated to be grouped based on their registered volume number, } i, \text{ and partition number, } [k_{x}, k_{y}, k_{z}], \text{ so we}
\]

obtain feature matrices, $F_{i,k_{x},k_{y},k_{z}} = [\bar{v}_{i1,k_{x},k_{y},k_{z}}, \cdots, \bar{v}_{i32,k_{x},k_{y},k_{z}}]$, which each one has 32 columns. Also, for each partition, desired labels are extracted as a subset $B_{k_{x},k_{y},k_{z}}$ of voxels from the registered segmentation volumes, 
\[
\{A_{i,j}^{reg} \mid i \in [1, \cdots, N_{tr}], k_{x} \in [1, 2, \cdots, N_{x}^{SANN}], k_{y} \in [1, 2, \cdots, N_{y}^{SANN}], k_{z} \in [1, 2, \cdots, N_{z}^{SANN}]\}. \text{ The volumetric labels are vectorized as, } \bar{v}_{i,k_{x},k_{y},k_{z}}. \text{ Finally, the feature matri-
\]
Figure 2.12: This figure shows how the partitions of the kidney shape are designated in the 2D space. The green region corresponds to the partition $B_{2,1}$.

Figure 2.13: This figure shows the block diagram of training the SANNs.

ces and vectorized labels of registered training volumes are vertically concatenated to form, $F_{k_x, k_y, k_z} = [F^T_{1,k_x,k_y,k_z}, \ldots, F^T_{N_{irr},k_x,k_y,k_z}]^T$ and $\vec{l}_{k_x,k_y,k_z} = [\vec{l}^T_{1,k_x,k_y,k_z}, \ldots, \vec{l}^T_{N_{irr},k_x,k_y,k_z}]^T$, where $F_{k_x,k_y,k_z}$ and $\vec{l}_{k_x,k_y,k_z}$ are called input features and targeted labels of the partition $[k_x,k_y,k_z]$. The input features and targeted labels of each partition are used to train its corresponding neural network classifier, $NET_{k_x,k_y,k_z}$. The main parameters of training the SANNs are listed in Table 2.6. The block diagram of training the SANNs is shown in Fig. 2.13.

2.5.2 Classification

The objective of voxel classification is to use the trained SANNs to classify voxels into (a) non-kidney voxels, (b) renal medulla voxels, and (c) the combination of kidney capsule and pyelocalyceal system voxels. Assume $V$ is an input volume, and $V^{dn}$ is its de-speckled volume. $V^{dn}$ should be transformed such
Table 2.6: This table provides the list of parameters of the SANNs associated with the MATLAB neural network toolbox.

<table>
<thead>
<tr>
<th>Parameter Name</th>
<th>Description</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>(N_y^{SANN}, N_y^{SANN}, N_x^{SANN})</td>
<td>Number of partitions along the x-, y-, and z-axes</td>
<td>([2, 2, 1])</td>
</tr>
<tr>
<td>(N_y)</td>
<td>Number of texture features per voxel</td>
<td>32</td>
</tr>
<tr>
<td>(N_{HL})</td>
<td>Number of hidden layers</td>
<td>1</td>
</tr>
<tr>
<td>(N_{NL})</td>
<td>Number of neurons per layers</td>
<td>75</td>
</tr>
<tr>
<td>trainFcn</td>
<td>Training function</td>
<td>Scaled conjugate gradient back-propagation [108]</td>
</tr>
<tr>
<td>transferFcn</td>
<td>Activation function</td>
<td>Hyperbolic tangent sigmoid (tansig)</td>
</tr>
</tbody>
</table>

that its kidney shape is aligned on the expected alignment/orientation of the kidney shape. This step is necessary to ensure the SANNs classifiers are correctly placed on the kidney shape in the transformed volume, and the voxel classification process correctly classifies kidney voxels from non-kidney ones. To solve this problem, we apply a step from the introduced shape-to-volume registration method, called finding the best seed point (Sec 2.6.2), which searches within a pre-defined set of similarity transformation parameters to find a similarity transformation, \(ST_{reg}\), that approximately fits the kidney shape model on the kidney shape in the ultrasound volume, \(V^{dn}\). Then, the inverse of the similarity transformation, \(ST_{reg}^{-1}\), is applied on \(V^{dn}\) to align its kidney shape on the expected alignment/orientation of the kidney shape, and the registered volume is achieved, \(V^{reg}(\vec{X}) = V^{dn}(ST_{reg}^{-1}\{\vec{X}\}), \forall \vec{X} \in \Omega_V\). Afterward, \(V^{reg}\), is convolved with the Gabor filters of equation (2.38), \(\{g_j|j \in [1, 2, \cdots, 16]\}\), to obtained feature volumes, \(\{V^g_j = V^{reg} * g_j(\vec{X})|i \in [1, \cdots, N_{tr}], j \in [1, \cdots, 16]\}\). For each partition, the subset \(B_{x,y,z}\) of voxels from the texture feature volumes are extracted, \(\{V^g_j \in [1, \cdots, 16], k_x \in [1, 2, \cdots, N_{SANN}]\}\), and vectorized to form \(\vec{v}_{j,k_x,k_y,k_z}\). Then, the vectors are horizontally concatenated to be grouped based on their partition number, \([k_x, k_y, k_z]\), so we obtain the feature matrices, \(F_{k_x,k_y,k_z} = [\vec{v}_{i,1,k_x,k_y,k_z}, \cdots, \vec{v}_{i,32,k_x,k_y,k_z}]\). For each partition, the corresponding neural network classifier, \(NET_{k_x,k_y,k_z}\), are applied on the input matrix to estimate label vectors as, \(\vec{p}^{reg}_{k_x,k_y,k_z} = NET_{k_x,k_y,k_z}(F_{k_x,k_y,k_z})\). Then, each label vector is reshaped into a 3D box, and is placed in the corresponding location of the image domain specified by \(B_{x,y,z}\) to obtain \(L^{reg}_{k_x,k_y,k_z}\). Note that \(L^{reg}_{k_x,k_y,k_z}\) has the same size of \(V^{dn}\), and has non-zero values for voxels inside the region specified by \(B_{x,y,z}\). We also generate \(W_{k_x,k_y,k_z}\) in which its voxels inside \(B_{x,y,z}\) receive ones, and its voxels outside \(B_{x,y,z}\) receive zeros. The generated labels are mixed to create, \(L^{reg}\), using the following equation,

\[
L^{reg}(\vec{X}) = \frac{\sum_{j=1}^{N_{SANN}} L_{k_x,k_y,k_z}(\vec{X})}{\sum_{j=1}^{N_{SANN}} W_{k_x,k_y,k_z}(\vec{X})}, \forall \vec{X} \in \Omega_V. \tag{2.41}
\]

Finally, the output of voxel classification, \(L\), is achieved by applying the similarity transformation
Figure 2.14: This figure shows voxel classification for different number of texture features. The dark voxels correspond to the renal medulla region, the bright voxels correspond to kidney capsule and pyelocalyceal system regions, and the gray voxels are non-kidney voxels.

on $L^{reg}$, as $L(\tilde{X}) = L^{reg}(ST_{p_{st}}\{\tilde{X}\}), \forall \tilde{X} \in \Omega_V$. Figure 2.14 shows the voxel classification performance versus different number of texture features, $N_g \in \{8, 12, 16, 24, 32, 40, 48\}$. According to Fig. 2.14, the CSI measure, which indicates the separability of kidney and non-kidney voxels, improves by increasing the number of texture features, $N_g$, however, the computational time is also increased by increasing $N_g$. The number of texture features is set to $N_g = 32$, as a trade off between the separability quality and the computational time of voxel classification. The steps of voxel classification are summarized as follows:

1. performing speckle reduction using the MFLV method to achieve, $V^{dn}$,

2. applying the first step of shape-to-volume registration, finding the best seed point, based on the shape-based kidney detection method to find $ST_{p_{st}}^{-1}$ and to obtain $V^{reg}$,

3. extracting partitions, vectorizing, and horizontally concatenating them to obtain $F_{k_x,k_y,k_z}^s$,

4. applying neural network classifiers on $F_{k_x,k_y,k_z}^s$, to achieve label vectors, $\vec{l}_{k_x,k_y,k_z}^{reg}$,

5. reforming label vectors to achieve $L_{k_x,k_y,k_z}$, and creating $W_{k_x,k_y,k_z}$ for all the partitions,

6. finding $L^{reg}$ using equation (2.41),

7. finding $L$ by applying the similarity transformation, $ST_{p_{st}}$, on $L^{reg}$.
2.6 Shape-to-volume registration

2.6.1 Introduction

3D image registration in abdominal medical images is the process of matching two 3D images of the same internal structure/organ taken at different times, different viewpoints, different subjects, and different sensors [66, 109]. The 3D image to be deformed is called the “source”, and the 3D image to fit the source on it is called the “target”. Medical image registration is categorized based on the type of source and target into shape-to-shape registration, shape-to-volume registration, and volume-to-volume registration [73]. Note that a shape refers to a specific representation of the organ’s shape (see Sec 2.2), and a volume refers to a 3D image.

Image registration can be performed using area-based and feature-based approaches [109]. In the area-based approach, a searching algorithm is applied to find the best similarity of an object of interest (source) with an image data (target). The object of interest is deformed by orientation, scale, and translation to fit it on the image data. In literature, this approach is also referred as template matching. In the feature-based approach, a searching algorithm is applied to find corresponding points on the source and target which have similar localized features. Features can be either region-features, line-features, or point-features [109]. Two famous region-features are scale-invariant feature transform (SIFT) [110] and speeded up robust features (SURF) [111], which are both local feature detector and descriptors. Despite the wide use of these features in object detection and image registration, they are not robust for ultrasound images, due to the low-quality and resolution of ultrasound images [112]. Therefore, we use the area-based approach to perform image registration in 3D ultrasound images.

In general, image registration consists of three main components:

- **transformation model**: is a geometrical deformation to transform a 3d image, which is either applied on the entire voxels in the same way (global or rigid deformation), or is applied differently on local neighborhoods (local or non-rigid deformation);

- **similarity metric**: is an equation that assigns a similarity rate to the source and target, such as sum of absolute difference (SAD), cross correlation (CC), and normalized cross correlation (NCC) [109].

- **optimization algorithm**: defines a process to estimate the parameters of transformation model to maximize the similarity metric between the source and target.

The objective of using image registration in this thesis is to find a global deformation for the kidney
shape model (CVISM), which fits the deformed CVISM on the kidney shape in ultrasound volumes. We call this task as the “kidney detection” process in 3D ultrasound images. Since we want to register the kidney shape model on the volumetric ultrasound data, it is called shape-to-volume registration. The shape-to-volume registration task is performed to answer two questions:

- does the kidney shape exist in an ultrasound volume or not?
- and if the kidney shape exists, what is its alignment/orientation in the ultrasound volume?

To ensure the detected shape is actually the kidney shape in 3D ultrasound images, the transformation model should be restricted to global (rigid) deformations to avoid free deformations of the CVISM such that it fits on any other object than the actual kidney shape. In addition, solving image registration problem under global deformation constraint is much faster than applying non-rigid deformations. These are two reasons of selecting global deformation.

The kidney shape in 3D ultrasound images might be partially visible due to a misalignment of the ultrasonic probe alignment. To perform a successful kidney detection, the registration method should be robust against partial shape occlusion. Some researchers have attempted to perform image registration of occluded shapes [113,114]. Kaneko et al. [113] proposed the selective cross-correlation (SCC) to detect partially occluded objects in images by masking non-concordant pixels in template and image from the similarity metric calculation. In the SCC method, mask coefficients are generated based on the data consistency between template and image. However, the kidney shape in an ultrasound volume is usually deformed by rigid and elastic deformations, and checking voxels coherency between the shape model and the ultrasound volume, results in masking out majority of voxels from the similarity metric calculation. This may lead to an incorrect registration task, and thus, it is not applicable for kidney detection in 3D ultrasound.

### 2.6.2 Proposed solution

We introduce a shape-to-volume registration based on a new similarity metric, named the regularized complex normalized cross-correlation (RCNCC), to provide registration’s robustness against the kidney shape’s partial occlusion and non-kidney structures resembling the kidney shape. In the next paragraphs, we explain the components of the shape-to-volume registration process, including a global transformation model, the RCNCC similarity metric, and an optimization strategy.

We apply the 3D similarity transformation as the global deformation model, with the same definition provided in Sec 2.2.2. Accordingly, the similarity transformation is defined as $ST_{\vec{p}_{st}}$, where $\vec{p}_{st}$ is the vector of seven similarity transformation parameters as, $\vec{p}_{st} = [\theta_x, \theta_y, \theta_z, s, t_x, t_y, t_z]^T$, where $\theta_x, \theta_y,$
\( \theta, \) and \( s \) are orientations over the \( x-, y-, \) and \( z- \) axes and scaling factor, respectively, and \( t_x, t_y, \) and \( t_z \) are translation parameters over the \( x-, \) \( y-, \) and \( z- \) axes, respectively (see Definition 2.1). We break the vector of similarity transformation parameters into two sub-vectors as, \( \vec{p}_{st,1} = [\theta_x, \theta_y, \theta_z, s]^T \) and \( \vec{p}_{st,2} = [t_x, t_y, t_z]^T, \) such that \( \vec{p}_{st,2} \) contains orientation and scaling parameters, and \( \vec{p}_{st,2} \) contains translation parameters. We aim to find the registration parameters, \( \vec{p}_{st} \), such that the transformed CVISM, \( \Psi(ST_{\vec{p}_{st}} \{ \vec{Y} \}) \), where \( \vec{Y} \in \Omega_{\Psi} \), obtains its maximum similarity with the kidney shape in an (enhanced) ultrasound volume, \( V^{eh} \). The optimization problem of our shape-to-volume registration is defined as follows,

\[
\langle \Gamma^*, \vec{p}_{st,1}^*, \vec{p}_{st,2}^* \rangle = \max_{\vec{p}_{st}} \left\{ \Gamma(ST_{\vec{p}_{st,1}}, \vec{p}_{st,2}) \right\},
\]

(2.42)

where \( \Gamma^*, \vec{p}_{st,1}^*, \) and \( \vec{p}_{st,2}^* \) are the RCNCC value, orientation/scale parameters, and translation parameters corresponding to the maximum similarity metric, respectively. The RCNCC similarity metric, \( \Gamma \), is defined as follows,

\[
\Gamma(\vec{p}_{st,1}, \vec{p}_{st,2}) = \max_{\vec{p}_{st,2} \in \Omega_{V}^{\text{ub}}} \left\{ \frac{\max\{0, \Re\{\Sigma_f(\vec{p}_{st,1}, \vec{p}_{st,2})\}\} \cdot \max\{0, \Im\{\Sigma_f(\vec{p}_{st,1}, \vec{p}_{st,2})\}\} \cdot \Lambda(\vec{p}_{st,1}, \vec{p}_{st,2})}{\Sigma_{II}^f(\vec{p}_{st,1}) \cdot \Sigma_{II}^R(\vec{p}_{st,1})} \right\},
\]

(2.43)

\[
\Sigma_f(\vec{p}_{st,1}, \vec{p}_{st,2}) = \iiint_{\vec{Y} = [x, y, z]^T \in \Omega_{\Psi}} (V^{eh}(\vec{p}_{st,2} + \vec{Y}) \cdot \Psi(ST_{\vec{p}_{st,1}} \{ \vec{Y} \}) dxdydz,
\]

(2.44)

\[
\Sigma_{II}^f(\vec{p}_{st,1}) = \iiint_{\vec{Y} = [x, y, z]^T \in \Omega_{\Psi}} (\Re\{\Psi(ST_{\vec{p}_{st,1}} \{ \vec{Y} \})\})^2 dxdydz,
\]

(2.45)

\[
\Sigma_{II}^R(\vec{p}_{st,1}) = \iiint_{\vec{Y} = [x, y, z]^T \in \Omega_{\Psi}} (\Im\{\Psi(ST_{\vec{p}_{st,1}} \{ \vec{Y} \})\})^2 dxdydz,
\]

(2.46)

\[
\Lambda(\vec{p}_{st,1}, \vec{p}_{st,2}) = K + \iiint_{\vec{Y} = [x, y, z]^T \in \Omega_{\Psi}} \left\{ V(\vec{p}_{st,2} + \vec{Y}) > 0 \right\} \cdot \left\{ \Psi(ST_{\vec{p}_{st,1}} \{ \vec{Y} \}) > 0 \right\} dxdydz
\]

(2.47)

where \( \vec{p}_{st,2} \in \Omega_{V}^{\text{ub}} \), and \( \Omega_{V}^{\text{ub}} \) is a sub-domain of the ultrasound volume (will be explained in the next paragraphs). \( \Lambda(\vec{p}_{st,1}, \vec{p}_{st,2}) \) is the regularization factor. \( \max\{0, a\} \) is a function that zeros negative values of \( a \). \( K \) is a constant number, preventing the registration process from trapping in trivial answers with small values of the regularization factor. \( ST_{\vec{p}_{st,1}} \{ \vec{Y} \} \) applies orientation/scaling transformation (without translation). As the transformed shape model moves out of the ultrasound pyramid, its shape is partially occluded. This results in a reduced cross-correlation value, preventing the registration process to detect the kidney shape partially aligned outside the ultrasound pyramid. To address this problem, the regularization factor is applied to boost the similarity metric at locations where the transformed kidney shape is partially occluded. In equation (2.43), the regularization factor is calculated as the
intersection volume of the ultrasound pyramid and the transformed shape model, and the ultrasound pyramid is specified by \( \{ V(\vec{p}_{st,1} + \vec{Y}) > 0 \} \). In equation (2.43), the first term (\( \Re\{\cdot\} \)) and second term (\( \Im\{\cdot\} \)) of the numerator receive higher values at locations in \( V^{eh} \) where the data structure is similar to kidney capsule/pyelocalyceal system and renal medulla, respectively. By multiplying these two terms, only locations similar to all the kidney structural regions receive local maximum values, and equation (2.43) picks the global maximum as a candidate kidney shape’s center point, \( \vec{p}_{st,2}^* \). In equation (2.43), the second and third terms in the denominator normalize the numerator of equation (2.43), to avoid any registration’s bias toward the intensity profiles of the transformed CVISM. Although equation (2.42) searches for \( \vec{p}_{st} \), including both \( \vec{p}_{st,1} \) and \( \vec{p}_{st,2} \), its formulation implies two levels of maximization process as follows:

1. **inner-maximization**: for a value of \( \vec{p}_{st,2} \), the formula of the RCNCC metric, defined in equation (2.43), performs an exhaustive search to find a value of \( \vec{p}_{st,2} \) that maximizes the RCNCC metric, \( \Gamma \). This search is performed within a sub-domain, \( \Omega_{V}^{sub} \). \( \Omega_{V}^{sub} \) can be adaptively selected to ensure a robust performance of the optimization method is achievable;

2. **outer-maximization**: equation (2.42) searches for a value of \( \vec{p}_{st,1} \) that maximizes the maximum RCNCC metric, \( \Gamma \), of equation (2.43).

Therefore, the left side of the equality formula in equation (2.43), \( \vec{p}_{st,1} \) is without star-sign (which means it is not necessarily an optimal value of \( \vec{p}_{st,1} \)), while \( \vec{p}_{st,2} \) has a star-sign which dictates its value is optimal such that it maximizes the right side of equation (2.43). In the left side of equation (2.42), both the sub-vectors of registration parameters have star-signs, indicating the maximization problem in the right side of equation (2.42) aims to find optimal values of both of the sub-vectors, \( \vec{p}_{st,1}^* \) and \( \vec{p}_{st,2}^* \), as well as the maximum RCNCC value, \( \Gamma^* \). After introducing the transformation model and the similarity metric, now, we proceed to introduce an optimization algorithm that finds a solution for equation (2.42).

According to the Hadamard’s definition of well-posed problems, almost all realistic medical image registration scenarios, including the non-convex optimization problem\(^{11}\) in equation (2.42), are ill-posed [66]. Thus, no close-form solutions can be derived ensuring the global optimum answer of the registration parameters is attainable. On the other words, each solution may obtain a different answer of the registration parameters [66]. Though, this does not mean a good maximum answer is not achievable! To

\(^{11}\)A real-valued function is called convex (concave upward) if each line segment between two points on the graph of the function lies above or on the graph [115]. For the optimization problem of the proposed shape-to-volume registration, assume \( \vec{p}_{st,1}^* \) and \( \vec{p}_{st,1}^2 \) are two points on the graph of the RCNCC function, \( \Gamma \) (equation (2.43)). For each value of a parameter, \( t \in [0, 1] \), \( \Gamma(t \cdot \vec{p}_{st,1}^2 + (1 - t) \cdot \vec{p}_{st,1}^2, \vec{p}_{st,2}^2) \leq t \cdot \Gamma(\vec{p}_{st,1}^2, \vec{p}_{st,2}^2) + (1 - t) \cdot \Gamma(\vec{p}_{st,1}^2, \vec{p}_{st,2}^2) \). This condition is only valid in local neighborhoods of maximum solutions of the equation (2.42), and is not valid for the entire space of \( \vec{p}_{st} \). Therefore, the optimization problem of equation (2.42) is called a non-convex problem.
find a good maximum solution for equation (2.42), we design an optimization strategy, consisting of two steps: *initialization* and *iterative improvement*. We assume the non-convex optimization problem in equation (2.42) has a local convexity around its good maximum answers (see Fig. 2.15). Therefore, the non-convex optimization problem of equation (2.42) is broken down as follows:

- **initialization**: finding a good initialization that places the non-convex optimization problem of equation (2.42) in a locally convex area,

- **iterative improvement**: applying an iterative solution that moves the answer of registration parameters toward a good maximum solution.

Given an input 3D ultrasound image, the optimization strategy starts with the *initialization* step, also called *finding the best seed point*, in which a set of seed points, \( S_T = \bigcup_{j=1}^{N_T} \vec{p}_{st,j} \), are tested to choose the best seed point, which can lead the optimization solution to find a desirable solution. Here, a seed point is referred to a vector of orientation/scale parameters, \( \vec{p}_{st,j} \). A desirable solution refers to a registered kidney shape model which fits inside the kidney shape in a 3D ultrasound image. By increasing the number of seed points, \( N_T \), the probability of placing the optimization problem in a locally convex area related to a better solution increases, however the computational time increases too. Therefore, the proper selection of \( N_T \) is important to maintain both the optimality and efficiency of the solution. To reduce the space of possible answers, we set imaging constraints on the parameters of \( \vec{p}_{st,1} \) as, \(-\frac{\pi}{4} < \theta_x < \frac{\pi}{4}, -\frac{\pi}{2} < \theta_z < \frac{\pi}{2}, \) and \( 0.5 < s < 2 \). To reduce the number of seed points, we ignored changes of \( \theta_y \), and seed points are generated for \( \theta_y = 0 \). Through experimentation, we found a good \( S_T \) is achieved by combining \( \theta_x \in \{-\frac{\pi}{6}, 0, \frac{\pi}{6}\}, \theta_z \in \{-\frac{\pi}{6}, -\frac{\pi}{3}, 0, \frac{\pi}{3}, \frac{\pi}{2}\}, \) and \( s \in \{0.75, 1, 1.5\} \), and therefore, we have \( N_T = 5 \times 3 \times 3 = 45 \). For an input ultrasound volume, we calculate \( \Gamma(\vec{p}_{st,1}, \vec{p}_{st,2}) \) for all the seed points in \( S_T \), on the entire 3D domain of the ultrasound volume (ie. \( \Omega^\text{sub}_V = \Omega_V \)), to ensure the *initialization* step considers all possible placements of the kidney shape in the ultrasound volume, and the selected seed point properly initializes the solution of equation (2.42).

In the *iterative improvement* step, equation (2.42) is iteratively solved using the Gradient Descent method to provide a good approximation of the global maximum answer of \( \Gamma \) (if exists). Although the Gradient Descent method finds a local maximum answer of equation (2.42), the initialization with the selected seed point leads the Gradient Descent method to provide a desirable solution. In the *Iterative Improvement* step, we assume \( \vec{p}_{st,2}^* \), obtained in the *Initialization* step, is close to the center point of the kidney shape. Therefore, at each iteration of the Gradient Descent method, \( \Omega^\text{sub}_V \) of equation (2.43) is limited to a small neighborhood of the previous value of \( \vec{p}_{st,2}^* \), such that \( \Omega^\text{sub}_V = \text{nib} (\vec{p}_{st,2}^*, \text{rad}_4) \), where \( \text{rad}_4 \) defines the radius of the searching box for \( \vec{p}_{st,2}^* \) in the current iteration of the Gradient Descent
Figure 2.15: This figure shows the RCNCC metric versus two registration parameters, $\theta_x$ and $\theta_z$, for a 3D ultrasound image ($\theta_y = 0$ and $s = 1$). The brighter (yellow) regions correspond to desirable registration solutions. The gray stars are seed point, and the red star is the selected seed point. Four iterative updates of the registration parameters using the Gradient Descent method are shown with black arrows.

method (Hint: nib(.,.) is defined in Sec 2.3.2). This reduction of the searching domain provides two advantages as follows,

- reducing the computational complexity of the optimization process,
- increasing the robustness of the optimization algorithm, by eliminating unwanted large changes of the answer due to undesired interferences.

Remark 2.3. The proposed optimization algorithm is very similar to semi-automated organ segmentation methods (such as the active appearance model (AAM) [116]), where an operator is required to select an initial seed point to manually initialize the organ segmentation process. In the proposed method of this thesis, the applied initialization step mimics a visual system of a trained operator to select an initialization of the kidney detection process. Thereby, the proposed solution is independent from an operator (i.e., paramedic) interaction.

Figure 2.15 shows an example of the RCNCC variation with respect to two registration parameters, $\theta_x$ and $\theta_z$, for a 3D ultrasound image. In Fig. 2.15, the RCNCC level is mapped with colors from (dark) blue to (bright) yellow, in which the yellow regions correspond to desirable registration solutions. The seed points are shown with gray stars, and the selected seed point is shown with the red star. As shown in Fig. 2.15, the selected seed point is close to a yellow region which enables the Gradient Descent
Algorithm 1: The proposed shape-to-volume registration method

Input: \( V^{eh}, \Psi, K \);
Output: \( \Gamma^*, \vec{p}^\ast_{st} \);

begin
  \set\Gamma^l = -1, \Gamma^* = 0, \epsilon = 0.1;
  \set\delta p_i = \{+\delta s, -\delta s, +\delta \theta_x, -\delta \theta_x, +\delta \theta_y, -\delta \theta_y, +\delta \theta_z, -\delta \theta_z\};
  \textbf{Step I: Initialization}
  \text{Find } \tilde{p} = \max \{\Gamma(\vec{p}_{st,1}, \vec{p}^\ast_{st,2})\} \text{ for } \vec{p}_{st,1} \in S_T; \\
  \textbf{Step II: Iterative Improvement}
  \text{while } \Gamma^* - \Gamma^l > \epsilon \text{ do}
  \text{for } i \in 1, \ldots, 8 \text{ do}
    \vec{p}_i = \vec{p}_{st,1} + \delta p_i; \\
    \Psi_T = 3D\text{SimilarityTransform}(\Psi, ST \vec{p}_i); \\
    \Sigma_I = 3D\text{ComplexCorr}(V^{eh}, \Psi_T); \\
    \Sigma_{II} = \text{Sum}(\Re\{\Psi_T\}), \Sigma_{II} = \text{Sum}(\Im\{\Psi_T\}); \\
    \Lambda = K + 3D\text{Corr}(|V^{eh}| > 0, \{|\Psi_T| > 0\}); \\
    [\Gamma_i, \vec{p}_{st,2}^\ast] = \max(\max(0, \Re\{\Sigma_I\}), \max(0, \Im\{\Sigma_I\}) \Lambda \Sigma_{II} \Sigma_{II}^\ast); \\
  \text{Update Similarity Transformation: } \vec{p}_{st,1} = \frac{\sum_{i=1}^8 \vec{p}_i \Gamma_i}{\sum_{i=1}^8 \Gamma_i}; \\
  i^* = \max_i \{\Gamma_i\}; \\
  \vec{p}_{st,1}^\ast = \vec{p}_i^*; \\
  \vec{p}_{st,2}^\ast = \vec{p}_{st,2}^\ast; \\
  \Gamma^\ast = \Gamma^*; \\
  \Gamma^l = \Gamma^*; \\
end

method to find a desirable solution. Algorithm 1 shows the pseudo-code of the proposed kidney’s shape registration method. As can be seen in Algorithm 1, it is comprised of the \textit{initialization} and \textit{iterative improvement} steps. The \textit{iterative improvement} step consists of an outer-loop of iterations and an inner-loop, containing a 3D similarity transformation, a 3D complex correlation (\textit{3DComplexCorr}), a 3D correlation (\textit{3DCorr}), two 3D summations, and finding the maximum of \( \Gamma_i \)'s. The most time consuming tasks are the \textit{3DCorr} and \textit{3DComplexCorr}, and each one is repeated for \( \text{itr} \times 8 \) times. After finding \( \Gamma^* \), \( \vec{p}_{st,1}^\ast \), and \( \vec{p}_{st,2}^\ast \), if \( \Gamma^* \) is greater than a threshold, \( \Gamma^* > \Gamma^\text{th} \), we decide that the kidney shape exists in the ultrasound volume, and the calculated parameters of the similarity transformation are returned as the outputs of the kidney detection process, \( \vec{p}_{st}^\ast = [\vec{p}_{st,1}^\ast]^T, [\vec{p}_{st,2}^\ast]^T]^T \).

### 2.7 Shape-based and atlas-based kidney detection

In this section, we introduce two strategies, including shape-based and atlas-based approaches, to detect the kidney shape in 3D ultrasound images. The shape-based method only uses shape prior information of the kidney shape, based on the generated kidney CVISM, to detect the kidney shape. The atlas-
based method uses both prior shape and texture information of the kidney shape to perform the kidney detection task. The block diagram of the shape-based kidney detection approach is shown in Fig. 2.16. The block diagram of the atlas-based kidney detection approach is shown in Fig. 2.17. Accordingly, compared to the shape-based kidney detection method, the atlas-based approach has one more step (voxel classification), which adds texture information of the kidney shape to improve the accuracy of kidney detection. Adding prior texture information in the processing pipeline of atlas-based kidney detection increases the computational cost, and needs more expensive hardware configuration to address the real-time operation, required for trauma diagnosis.

Figure 2.16: This figure shows the block diagram of shape-based kidney detection approach.

Figure 2.17: This figure shows the block diagram of atlas-based kidney detection approach.
2.8 Application of kidney detection in computer assisted ultrasonic probe placement for trauma diagnosis

2.8.1 Introduction

As discussed in Sec 1.4, in ultrasound imaging for triaging, regions around abdominal organs are scanned to find indications of internal bleeding and abnormalities. To view abdominal organs with an ultrasound probe, a high level of anatomical understanding and knowledge is required to know where to put the ultrasonic probe on the patient’s body to scan relevant internal views. Therefore, a well-trained radiologist is needed to conduct abdominal ultrasound scans for triaging. Although skilled radiologists and advanced imaging devices are available at referral hospitals where patients are in hemodynamically stable conditions, these facilities are not usually accessible in emergency situations such as resuscitative bedside, pre-hospital environments, and smaller referral centers, resulting in a large number of preventable deaths [51].

We introduce a computer-assisted ultrasonic probe placement method, which can be installed as a software module in an ultrasound research platform [117, 118]. In the proposed solution of this thesis, anatomical knowledge of shape and alignment of the kidney shape are fed into the process of ultrasound imaging to facilitate rapid diagnostic assessment of trauma patients by paramedics who lack anatomical knowledge. The probe placement solution has been proposed under the ethics committee protocol [40]. By recording a volumetric region (3D image) and by detecting the kidney shape, rapid diagnostic assessment by paramedics is possible. The recorded 3D image can be sent to a referral center or a computerized diagnostic module to make a decision. Because the proposed solution does not need live-streaming of ultrasound images, and merely a final 3D image is enough to be sent to a remote expert for making a diagnostic decision, a low-speed internet connectivity suffices. Alternatively, in the absence of a remote expert or unavailability of an internet connectivity in emergency situation, automated diagnosis can be utilized to make a diagnostic decision.

In the proposed solution of this thesis, the ultrasonic probe is navigated by processing 3D ultrasound images, without using information of positioning sensors or another imaging modality. This is essential in triage where time consuming calibration may cost patients’ lives. The idea is to feed anatomical knowledge of the organ’s location/shape into the imaging process to compensate for the lack of anatomical knowledge of paramedics. The process of probe placement consists of two phases: initial probe placement, and correcting probe placement. The block diagram of the proposed ultrasonic probe
Select an Organ of Interest
Specify Age and Gender of Patient
Visually Guide the Paramedic how to Place Ultrasound Probe on a Related Anatomical Model
Waiting for CSoCIPP from Paramedic

Phase I

Phase II

Anatomical knowledge of organ’s location

Anatomical knowledge of organ’s shape

Calculate Organ’s Shape Misalignment
Organ Exists?

Yes

Detect and Locate the Organ of Interest

Yes

Send Probe Navigational Command to Paramedic

Waiting for CSoPPR from Paramedic

No

Organ’s Shape Misaligned?

Yes

Calculate Probe Misalignment

Use Acquired Image For Triaging

No

Figure 2.18: The processing pipeline of the two phases of the proposed computer-assisted ultrasound probe placement.

2.8.2 Initial probe placement

In the first phase, anatomical knowledge of the kidney’s location is used to guide a paramedic to place the ultrasound probe on a location on a patient’s skin, which minimizes the probe misalignment with respect to a correct probe alignment for imaging the RUQ view. The human anatomy differs by age and gender among patients, and to correctly guide operators to find an initial probe placement, the age and gender of patients should be specified. Based on the age and gender of a patient, an anatomical model in the supine position is displayed, and graphical instructions of initial probe placement are displayed on the anatomical model (Fig. 2.18). The intersection of the horizontal sub-xiphoid and right midaxillary lines optimizes probe placement for the RUQ view, and the probe marker should be directed toward the cephalad [119]. Based on [119], we created animations to guide operators to locate the initial probe placement of the RUQ view (an animation frame for a mid-age male model is shown inside Fig. 2.18 Phase-I). The first phase continues until the paramedic gives a confirmation signal of completing initial probe placement (CSoCIPP). Due to inexperience of paramedics and human anatomy variability among patients, initial probe placement only provides a partial kidney visibility, and further adjustments are required to obtain a correct image of the RUQ view.
2.8.3 Correcting probe placement

Upon receiving the CSoCIPP signal, the second phase starts to fine-tune the ultrasound probe alignment for imaging the RUQ view. In the second phase, a process of probe realignment is iterated until the correct probe placement is obtained. Each iteration consists of image acquisition, kidney detection, calculating probe misalignment, sending a probe navigational command, and waiting for the paramedic’s confirmation. An iteration begins with acquiring a 3D ultrasound image, and the acquired image is processed to detect the kidney shape inside the image. If the kidney shape is not detected, the paramedic is referred back to redo the first phase. Otherwise, the alignment of the detected kidney shape is compared with a reference alignment to find the kidney shape misalignment in the acquired 3D image. Because the alignment of the kidney shape inside images is directly linked to the ultrasound probe alignment, the calculated organ’s shape misalignment is used to estimate its corresponding ultrasound probe misalignment. After estimation of the ultrasound probe misalignment, a navigational command is sent to the paramedic to move or rotate the probe toward the correct placement. Afterward, the process waits for a confirmation signal of performing probe re-alignment (CSoPPR) from the paramedic. This process repeats until a correct view is obtained (Fig. 2.18).

2.8.4 Calculating probe misalignment

As discussed in Sec. 2.8.3, the calculated kidney shape misalignment in an acquired 3D ultrasound image is used to estimate the probe misalignment, and the estimated probe misalignment is used to generate a probe navigational command. Because the probe misalignment changes the viewpoint, it results in the kidney shape’s misalignment inside acquired 3D ultrasound images. Assume the output of the kidney detection process is given as, $\vec{p}_{st} = [\theta_x^*, \theta_y^*, \theta_z^*, s, t_x^*, t_y^*, t_z^*]$. We select translation and orientation parameters of $\vec{p}_{st}$ to form a 3D rigid-body transformation as, $\vec{p}_{rb} = [\theta_x^*, \theta_y^*, \theta_z^*, t_x^*, t_y^*, t_z^*]$. The rigid-body transformation is defined as, $ST_{\vec{p}_{rb}}$. In the rigid-body transformation, the size of the kidney shape with respect to the reference shape is ignored. This selection is because the size of the kidney shape has nothing to do with the probe misalignment. We need to determine the transformation, $P_{up}$ (probe projection matrix), which projects the rigid-body transformation, $ST_{\vec{p}_{rb}}$, from the 3D image domain into the real 3D space of the ultrasound probe, $T_{up} = P_{up} \cdot ST_{\vec{p}_{rb}}$. The probe projection matrix depends on the settings of the imaging device, and therefore, for each device, $P_{up}$ is required to be calculated once
Given a few pairs of $T_{up}$s and $ST_{\vec{r}_p}s$, $P_{up}$ is obtained by solving the following equation,

$$\hat{P}_{up} = \min_{P_{up}} \sum_{k=1}^{N_{up}} \|T^k_{up} - P_{up} \cdot ST^k_{\vec{r}_p}\|_F,$$  \hspace{1cm} (2.48)

where $\|\cdot\|_F$ and $N_{up}$ are the Frobenius norm and the number of pairs of $T_{up}$s and $ST_{\vec{r}_p}$s.

---

**Algorithm 2: Calculating the Probe Projection Matrix**

**Select patient model and 3D ultrasound device:**

**Initialize:** $T^k_{up} = \text{eye}(4)$ for $k \in \{1, 2, 3, 4\}$ where $\text{eye}(4)$ is an identity matrix of size $4 \times 4$;

**Step-1:** Find the Morison’s pouch view;

**Step-2:** Move (slide) the probe slightly toward cephalad direction, record the movement as $T^1_{up}(1, 4)$, acquire a 3D image, and calculate and save it as $ST^1_{\vec{r}_p}$;

**Step-3:** Return the probe back to the Morison’s pouch view, and move (slide) the probe slightly toward caudad direction, record it as $T^2_{up}(1, 4)$, acquire a 3D image, and calculate and save $ST^2_{\vec{r}_p}$;

**Step-4:** Return the probe back to the Morison’s pouch view, rotate the probe (clockwise), record it as $T^3_{up}(1 : 3, 1 : 3) = R_{\theta_z}$ where $R_{\theta_z}$ is the rotation matrix of probe $\theta_z$, acquire a 3D image, and calculate and save $ST^3_{\vec{r}_p}$;

**Step-5:** Return the probe back to the Morison’s pouch view, rotate the probe (counter-clockwise), record it as $T^4_{up}(1 : 3, 1 : 3) = R_{\theta_z}$ where $R_{\theta_z}$ is the rotation matrix of probe $\theta_z$, acquire a 3D image, and calculate and save $ST^4_{\vec{r}_p}$;

**Step-6:** Calculate $P_{up}$ using equation (2.48).

---

### 2.8.5 Implementation

To meet practical expectations for triaging, the following implementation-related questions are addressed:

1. which platform can support requirements of the proposed solution; and
2. how to design a user-friendly interface that guides operators to conduct triaging. Our solution requires an ultrasound device that provides real-time external access to the acquired 3D ultrasound images. Recently, imaging devices have been developed under the titles of ultrasound research interface (URI) platform and ultrasound advanced open platform (ULA-OP) [120,121], which provide real-time access to acquired 3D images. A graphical user interface (GUI) is designed to visually guide operators to find the initial probe placement, to show acquired 3D images, and to display ultrasound probe navigational commands. The 3D ultrasound images are displayed by multiplanar reformatting and ray-casting techniques [122]. The GUI also provides tools for operators to give the CSoCIPP and CSOPPR signals.
2.9 Experimental results

In this section, we represent experiments to evaluate the proposed shape-based and atlas-based kidney detection methods, compared to the Marsousi et al-EMBC14 [63], and Noll et al. [12].

2.9.1 Hardware and software setup

We have used a DELL-precision workstation computer, with an Intel Xeon E5-1660 processor (containing 16 cores of 3.00 GHz), and a NVIDIA Quadro K2200 video processing card (containing 640 GPU-cores and 4GB memory). This configuration provides a high capability of parallel processing. A comparable configuration is affordable in a portable format (ex. DELL-precision mobile workstation). The proposed solution is intended to be used in conjunction with a real-time 3D ultrasound imaging device, currently under-development [123]. The proposed method of this paper is currently under-development, and therefore, MATLAB (R2015b) has been chosen as a suitable environment to develop the mathematical formulation, and to derive experimental analysis.

2.9.2 Implementation details

Both of the proposed shape-based and atlas-based kidney detection methods have massive computational loads, and their computational time should be minimized to be applicable for real-time interaction by paramedics for trauma diagnosis. The total computational time of the proposed shape-based kidney detection method is the summation of the computational times of the following parts: speckle reduction, \(LHE\), \(G2LTh\), and shape-to-volume registration, including finding the best seed point and iterative improvement. The total computational time of the proposed atlas-based kidney detection method is the summation of the computational times of the following parts: speckle reduction, \(LHE\), \(G2LTh\), finding the best seed point, voxel classification, and iterative improvement. We have used multi-thread programming (parfor), MATLAB’s built-in GPU functions, and CUDA programming to minimize the computational time of the proposed solution. The MFLV method has been developed using MATLAB Built-in GPU-accelerated 3D Gaussian (imgaussfilt3) and 3D Box (imboxfilt3) filters. The LHE is computationally massive, and has been implemented using GPU-CUDA programming. The G2LTh module has shown a faster operation using single-thread programming. The most computationally demanding part of the shape-based kidney detection method is the shape-to-volume registration task, consisting of finding the best seed point and iterative improvement. The finding the best seed point module requires 45 3DComplexCorrs, and each iteration of the optimization with the Gradient Descent method performs eight...
3DCorr, and eight 3DComplexCorrs. We have used multi-thread programming (parfor) to distribute the 3D correlations between eight CPU-cores. Each core interfaces with the GPU to run a 3DCorr and a 3DComplexCorr. For the atlas-based kidney detection method, voxel classification has even more computational load than the shape-to-volume registration task. The voxel classification requires 3D convolution of the ultrasound volume with 16 complex-valued Gabor filters, two 3D transformations, and applying the SANN classifiers. The 3D convolutions have been performed using CUDA-GPU programming, 3D transformations have been performed using single-core programming, and the SANN classifiers have used MATLAB built-in neural network classifier based on GPU and multi-threading. Since MATLAB R2015b does not have a GPU-accelerated 3D-correlation function, we have developed both 3DCorr and 3DComplexCorr functions using GPU-CUDA programming, and called them within the MATLAB codes. Table 2.7 shows four implementation methods to develop the parts in MATLAB, including single-core sequential programming (Single-thread), parfor, gpuArray, and the combination of parfor and gpuArray.

Table 2.7: Implementation methods of the parts of the proposed shape-based and atlas-based kidney detection methods in MATLAB.

<table>
<thead>
<tr>
<th>Module name</th>
<th>Single-thread</th>
<th>parfor</th>
<th>gpuArray</th>
<th>parfor &amp; gpuArray</th>
</tr>
</thead>
<tbody>
<tr>
<td>MFLV</td>
<td>✓</td>
<td></td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>LHE</td>
<td></td>
<td></td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>G2LTh</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Finding the best seed point</td>
<td></td>
<td></td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>Voxel classification</td>
<td>✓</td>
<td></td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>Iterative optimization</td>
<td></td>
<td></td>
<td></td>
<td>✓</td>
</tr>
</tbody>
</table>

2.9.3 3D ultrasound dataset

The training set (Ds3DUsIs-AUVs-Tr), including six with-kidney and six without-kidney images, are used (a) for generating the kidney shape model, (b) for training the SANNs, and (c) for sensitivity analysis to kidney shape deformation and occlusion. The evaluation set of healthy volunteers (Ds3DUsIs-AUVs-ESH), including 15 with-kidney and 15 without-kidney images, as well as the evaluation set of abnormal
Chapter 2. Kidney detection

71 patients (Ds3DUsIs-AUVs-ESA), including 8 with-kidney images, are used to evaluate the performance of the kidney detection methods. Also, each with-kidney image in the training and evaluation sets has at least one 3D ground truth (Ds3DUsIs-GTs). We also used the ultrasound volume simulator (Ds3DUsIs-UVS), mimicking actual ultrasound images of the RUQ view, to examine the performance of the proposed shape-based and atlas-based kidney detection methods under any desired combination of kidney shape deformation and partial occlusion. For more information on the dataset, the readers are referred to Sec. 1.6.

2.9.4 Comparison metrics

To evaluate the kidney detection methods, we use accuracy, sensitivity, and specificity metrics as,

\[
\begin{align*}
\text{Acc}_{KD} &= \frac{N_{TP} \cdot N_{TN}}{N_{TP} \cdot N_{TN} + N_{FP} \cdot N_{FN}}, \\
\text{Sens}_{KD} &= \frac{N_{TP}}{N_{TP} + N_{FN}}, \\
\text{Spec}_{KD} &= \frac{N_{TN}}{N_{TN} + N_{FP}},
\end{align*}
\]

(2.49)

where \(N_{TP}, N_{TN}, N_{FP}, \) and \(N_{FN}\) are numbers of true positive, true negative, false positive, and false negative kidney detections, respectively.

We also use the class separability index (CSI) to measure the detectability of with-kidney images from without-kidney images for the shape-based and atlas-based kidney detection methods. The CSI metric is defined based on calculated \(\Gamma^*\)s of shape-to-volume registration, as follows,

\[
\text{CSI} = \frac{N_{wk}(\mu - \mu_{wk})^2 + N_{wok}(\mu - \mu_{wok})^2}{\sigma_{wk}^2 + \sigma_{wok}^2},
\]

(2.50)

where \(\sigma_{wk}, \sigma_{wok}, \mu, \mu_{wk}, \mu_{wok}, N_{wk},\) and \(N_{wok}\) are standard deviations of \(\Gamma^*\)s for the with-kidney and without-kidney images, mean of \(\Gamma^*\)s of all samples, means of \(\Gamma^*\)s of the with-kidney and without-kidney images, and the numbers of with-kidney and without-kidney images, respectively.

2.9.5 Evaluating the kidney detection methods

To compare accuracy and robustness of the kidney detection methods, including the proposed shape-based and atlas-based methods, Marsousi et al.-EMBC14, and Noll et al., we set up two experiments. In the 1st experiment, the simulated ultrasound volumes are used to evaluate the robustness of the kidney detection methods against the kidney shape’s translation, orientation, and occlusion. In the 2nd experiment, the evaluation set, containing 30 actual ultrasound volumes, is used to compare accuracy, sensitivity, and specificity of the kidney detection methods. For the kidney detection methods, a true positive detection must satisfy \(d = ||\hat{X}^{dt} - \hat{X}^{ac}|| < d_{min}\), where \(\hat{X}^{dt}\) and \(\hat{X}^{ac}\) are detected and actual center points of the kidney shape in an ultrasound volume, and \(d_{min}\) is set as \(d_{min} = 35\)px to ensure true
positive detected center points are aligned inside the actual kidney shape. For the proposed method of this paper, true positive ($TP_{KD}$), true negative ($TN_{KD}$), false positive ($FP_{KD}$), and false negative ($FN_{KD}$) detections are determined as,

$$\text{Decision} = \begin{cases} 
TP_{KD}, & V_{wok} : (\Gamma^* > \Gamma^{th} \cap d < d_{min}) \\
TN_{KD}, & V_{wok} : \Gamma^* < \Gamma^{th} \\
FN_{KD}, & V_{wok} : \Gamma^* < \Gamma^{th} \\
FP_{KD}, & V_{wok} : \Gamma^* > \Gamma^{th} \cup V_{wk} : d > d_{min}
\end{cases}$$

(2.51)

where $V_{wk}$ & $V_{wok}$ refer to with-kidney and without-kidney images, respectively. We empirically set $\Gamma^{th} = 3.5$ for both the atlas-based and shape-based kidney detection methods, based on the training dataset ($Ds3DUsIs-AUVs-Tr$).

### 2.9.5.1 Sensitivity analysis to kidney shape misalignment and occlusion

In this experiment, we analyze the sensitivity of the shape-based and atlas-based kidney detection methods to the kidney shape’s misalignment and occlusion. The dataset of actual ultrasound volumes ($Ds3DUsIs-AUVs$) does not provide enough variability and flexibility to analyze tolerance and robustness of the proposed kidney detection methods against the kidney shape occlusion and deformation. Therefore, simulated ultrasound volumes ($Ds3DUsIs-UVS$) are used to able the analyze of the proposed methods under any arbitrary kidney shape deformation and occlusion. In the simulated images, translation over $x$, $y$, and $z$-axes are defined as $\{\Delta t_x, \Delta t_y, \Delta t_z \in [-100, -95, \cdots, 95, 100]\}$. Based on the nature of the ultrasound imaging, the probe orientations over $y$- and $z$-axes, so-called rocking and tilting [119] respectively, are limited to small angels, whereas the probe orientation over $x$-axes, so-called rotating [119], is not limited. However, we assume the initialization step of probe placement prevents the kidney shape from being disoriented by large angels. Thus, simulated ultrasound volumes of the RUQ view are generated with the kidney shape’s orientations over the $x$, $y$, and $z$-axes by, $\{\Delta \theta_x, \Delta \theta_y, \Delta \theta_z \in [-45^\circ, -42.5^\circ, \cdots, 42.5^\circ, 45^\circ]\}$. For each of the deformation parameters $\Delta t_x$, $\Delta t_y$, $\Delta t_z$, $\Delta \theta_x$, $\Delta \theta_y$, and $\Delta \theta_z$, the kidney shape visibility is measured from 0 to 1, and the simulated images are fed into the shape-based and atlas-based kidney detection methods. In figures 2.19, 2.20, 2.21, 2.22, 2.23, 2.24 the sensitivity analysis of the proposed shape-based and atlas-based kidney detection methods with respect to the kidney shape orientation over the $x$, $y$, and $z$-axes, and the kidney shape translation over the $x$, $y$, and $z$-axes, respectively, are demonstrated. In each of the figures, the result of the shape-based and atlas-based methods are shown in the left-side (a) and right-side (b), respectively. Accordingly, the shape-based kidney detection method shows wider ranges of detectability toward all the six orientations and translations of the kidney shape. Though, the atlas-based method shows better
matches with the unit-slope lines (dashed gray) in all the six figures for the detected ranges. This can be interpreted that the accuracy of estimating the kidney shape deformation is higher for the atlas-based method. Table 2.8 shows the mean and standard deviation of estimating the kidney shape deformations, including $\Delta t_y$, $\Delta t_z$, $\Delta \theta_x$, $\Delta \theta_y$, and $\Delta \theta_z$, for both the shape-based and atlas-based methods. Accordingly, the atlas-based method provided higher accuracy of estimation for the kidney shape deformations, except for the orientation and translation over the $x$ axis. This might be because of the insufficiency of the number of Gabor features, representing texture information over the $x$ axis, for voxel classification. We see in Fig. 2.19 that both the methods have weak estimation of $\Delta \theta_x$. Figure 2.25 shows three examples of applying both the shape-based and atlas-based kidney detection methods on simulated volumes.

This analysis also can reflect the usability of the proposed shape-based and atlas-based kidney detection methods in the computer-assisted probe placement task, which is required to help operators to find the correct placement of the ultrasonic probe corresponding to the RUQ view. Based on the results represented in both the figures and table 2.8 in this sensitivity analysis, we recommend the following strategy to apply both the shape-based and atlas-based methods to maximize the effectiveness of the proposed solution as follows:

- **Phase II-1:** At the beginning of the phase II of the computer-assisted probe placement, the kidney shape might be misaligned in terms of orientation and translation, with high deviations with respect to the reference alignment. Therefore, the shape-based kidney detection is a better choice for early steps of probe placement, because it provides a wider range of detectability.

- **Phase II-2:** After a few iterations of the computer-assisted probe placement using the atlas-based kidney detection method, the kidney shape’s misalignment gets smaller and smaller. This makes the use of the atlas-based kidney detection to be more efficient, cause the kidney shape is more likely aligned within the detectability range of the atlas-based kidney detection method, and using the atlas-based method provides a higher accuracy of estimating the kidney shape deformations, at least over the $y$- and $z$- axes.

Table 2.8: This table shows mean ($\mu$) and standard deviation ($\sigma$) of estimating the kidney shape deformations, using both the shape-based and atlas-based kidney detection methods. Errors are calculated in voxels for translations and degrees for orientations.

<table>
<thead>
<tr>
<th></th>
<th>$\Delta \theta_x$ (degree) $[\mu, \sigma]$</th>
<th>$\Delta \theta_y$ (degree) $[\mu, \sigma]$</th>
<th>$\Delta \theta_z$ (degree) $[\mu, \sigma]$</th>
<th>$\Delta t_x$ (voxel) $[\mu, \sigma]$</th>
<th>$\Delta t_y$ (voxel) $[\mu, \sigma]$</th>
<th>$\Delta t_z$ (voxel) $[\mu, \sigma]$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Shape-based</td>
<td>[13.2766, 9.8131] [6.7643, 4.3881]</td>
<td>[2.5169, 1.9578] [4.5000, 5.0648]</td>
<td>[3.1579, 2.3632] [4.9000, 2.7319]</td>
<td>[4.6667, 5.4746] [4.6667, 5.4746]</td>
<td>[1.4737, 1.5044] [1.8000, 1.3992]</td>
<td></td>
</tr>
<tr>
<td>Atlas-based</td>
<td>[16.3514, 11.260] [4.8353, 3.9326]</td>
<td>[2.2224, 1.1531] [4.6667, 5.4746]</td>
<td>[1.4737, 1.5044] [1.8000, 1.3992]</td>
<td>[4.6667, 5.4746] [4.6667, 5.4746]</td>
<td>[1.4737, 1.5044] [1.8000, 1.3992]</td>
<td></td>
</tr>
</tbody>
</table>
Figure 2.19: This figure provides sensitivity analysis of the shape-based and atlas-based kidney detection methods toward the kidney shape orientation over the $x$-axis. The blue lines in the top sub-figures show detected orientation of the kidney shape in simulated ultrasound volumes over the $x$-axis, $\Delta \theta_{\text{det}}^x$, versus the actual orientation of the kidney shape over the $x$-axis, $\Delta \theta_{\text{act}}^x$ (gray line with the slope one). The green lines in the bottom figures show the kidney shape visibility versus the actual orientation of the kidney shape over the $x$-axis, $\Delta \theta_{\text{vis}}^x$. The red lines in the bottom figures show the kidney shape detection versus the actual orientation of the kidney shape over the $x$-axis, $\Delta \theta_{\text{det}}^x$.

Figure 2.20: This figure provides sensitivity analysis of the shape-based and atlas-based kidney detection methods toward the kidney shape orientation over the $y$-axis. The blue lines in the top sub-figures show detected orientation of the kidney shape in simulated ultrasound volumes over the $y$-axis, $\Delta \theta_{\text{det}}^y$, versus the actual orientation of the kidney shape over the $y$-axis, $\Delta \theta_{\text{act}}^y$ (gray line with the slope one). The green lines in the bottom figures show the kidney shape visibility versus the actual orientation of the kidney shape over the $y$-axis, $\Delta \theta_{\text{vis}}^y$. The red lines in the bottom figures show the kidney shape detection versus the actual orientation of the kidney shape over the $y$-axis, $\Delta \theta_{\text{det}}^y$. 
Figure 2.21: This figure provides sensitivity analysis of the shape-based and atlas-based kidney detection methods toward the kidney shape orientation over the $z$-axis. The blue lines in the top sub-figures shows detected orientation of the kidney shape in simulated ultrasound volumes over the $z$-axis, $\Delta \theta_{z}^{\text{det}}$, versus the actual orientation of the kidney shape over the $z$-axis, $\Delta \theta_{z}^{\text{act}}$ (gray line with the slope one). The green lines in the bottom figures show the kidney shape visibility versus the actual orientation of the kidney shape over the $z$-axis, $\Delta \theta_{z}^{\text{act}}$. The red lines in the bottom figures show the kidney shape detection versus the actual orientation of the kidney shape over the $z$-axis, $\Delta \theta_{z}^{\text{act}}$.

Figure 2.22: This figure provides sensitivity analysis of the shape-based and atlas-based kidney detection methods toward the kidney shape translation over the $x$-axis. The blue lines in the top sub-figures shows detected translation of the kidney shape in simulated ultrasound volumes over the $x$-axis, $\Delta t_{x}^{\text{det}}$, versus the actual translation of the kidney shape over the $x$-axis, $\Delta t_{x}^{\text{act}}$ (gray line with the slope one). The green lines in the bottom figures show the kidney shape visibility versus the actual translation of the kidney shape over the $x$-axis, $\Delta t_{x}^{\text{act}}$. The red lines in the bottom figures show the kidney shape detection versus the actual translation of the kidney shape over the $x$-axis, $\Delta t_{x}^{\text{act}}$. 
Chapter 2. Kidney detection

Figure 2.23: This figure provides sensitivity analysis of the shape-based and atlas-based kidney detection methods toward the kidney shape translation over the $y$-axis. The blue lines in the top sub-figures shows detected translation of the kidney shape in simulated ultrasound volumes over the $y$-axis, $\Delta t_y^{dt}$, versus the actual translation of the kidney shape over the $y$-axis, $\Delta t_y^{ac}$ (gray line with the slope one). The green lines in the bottom figures show the kidney shape visibility versus the actual translation of the kidney shape over the $y$-axis, $\Delta t_y^{ac}$. The red lines in the bottom figures show the kidney shape detection versus the actual translation of the kidney shape over the $y$-axis, $\Delta t_y^{ac}$.

Figure 2.24: This figure provides sensitivity analysis of the shape-based and atlas-based kidney detection methods toward the kidney shape translation over the $z$-axis. The blue lines in the top sub-figures shows detected translation of the kidney shape in simulated ultrasound volumes over the $z$-axis, $\Delta t_z^{dt}$, versus the actual translation of the kidney shape over the $z$-axis, $\Delta t_z^{ac}$ (gray line with the slope one). The green lines in the bottom figures show the kidney shape visibility versus the actual translation of the kidney shape over the $z$-axis, $\Delta t_z^{ac}$. The red lines in the bottom figures show the kidney shape detection versus the actual translation of the kidney shape over the $z$-axis, $\Delta t_z^{ac}$.
2.9.5.2 Evaluating kidney detection accuracy

In this part, we evaluate the accuracy, sensitivity, and specificity of the kidney detection methods, including the proposed shape-based and atlas-based kidney detection methods, Marsoussi et al.-EMBC14 [63], and Noll et al. [12] methods. The evaluation sets of actual ultrasound volumes of both the healthy volunteers (Ds3DUsIs-AUVs-ESH) and the abnormal subjects (Ds3DUsIs-AUVs-ESA) are used to conduct this evaluation. The results of the kidney detection of the methods for the actual ultrasound volumes of the healthy volunteers are provided in Table 2.9. According to Table 2.9, the atlas-based method provides an improved kidney detection accuracy ($ACC_{KD} = 93.33\%$) over the state-of-the-art, and the shape-based method stands in the second position with $ACC_{KD} = 86.67\%$. The higher accuracy of the atlas-based kidney detection method attributes to the use of both the shape and texture prior knowledge in the kidney detection process.

Figure 2.26 shows the receiver operating characteristic (ROC) curve for detecting the with-kidney and without-kidney images of the evaluation set of actual ultrasound volumes of the healthy volunteers (Ds3DUsIs-AUVs-ESH), using the shape-based method, the atlas-based method, and a random guess. Figure 2.26 demonstrates that both the shape-base and atlas-based kidney detection methods provide acceptable discrimination between with-kidney and without-kidney images, while the atlas-based method performs slightly better than the shape-based approach. Also in figure 2.26, the area under the ROC curve (AUC) for each of the atlas-based and shape-based kidney detection methods estimates the probability that the kidney detection method would assign a higher score to a randomly selected with-kidney image than a randomly selected without-kidney image. The AUC of the atlas-based method is higher than the AUC of the shape-based method. The better performance of the atlas-based method can be attributed to the higher separability of kidney from non-kidney voxels in the enhanced volumes of the
Table 2.9: Comparing accuracy ($\text{ACC}_{\text{KD}}$), sensitivity ($\text{Sens}_{\text{KD}}$), and specificity ($\text{Spec}_{\text{KD}}$) of the kidney detection methods using the evaluation set of actual ultrasound volumes of the healthy volunteers ($\text{Ds3DUsIs-AUVs-ESH}$).

<table>
<thead>
<tr>
<th>Method</th>
<th>$N_{TP}$</th>
<th>$N_{TN}$</th>
<th>$N_{FP}$</th>
<th>$N_{FN}$</th>
<th>$\text{ACC}_{\text{KD}}$ (%)</th>
<th>$\text{Sens}_{\text{KD}}$</th>
<th>$\text{Spec}_{\text{KD}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Shape-based</td>
<td>14</td>
<td>12</td>
<td>3</td>
<td>1</td>
<td>86.67</td>
<td>0.9333</td>
<td>0.8000</td>
</tr>
<tr>
<td>Atlas-based</td>
<td>15</td>
<td>13</td>
<td>2</td>
<td>0</td>
<td>93.33</td>
<td>1</td>
<td>0.8667</td>
</tr>
<tr>
<td>Marsousi et al-EMBC14</td>
<td>12</td>
<td>13</td>
<td>2</td>
<td>3</td>
<td>83.33</td>
<td>0.8000</td>
<td>0.8667</td>
</tr>
<tr>
<td>Noll et al.</td>
<td>9</td>
<td>11</td>
<td>4</td>
<td>6</td>
<td>66.67</td>
<td>0.6000</td>
<td>0.7333</td>
</tr>
</tbody>
</table>

Figure 2.26: This figure shows the receiver operating characteristic (ROC) curve, for detecting the with-kidney and without-kidney images of the evaluation set of actual ultrasound volumes of the healthy volunteers ($\text{Ds3DUsIs-AUVs-ESH}$), using the shape-based method, the atlas-based method, and a random guess.

In patients with abdominal trauma, the free fluids around the right kidney due to bleeding change the morphology of the RUQ view. Thus, using the images of abnormal subjects, the robustness of the organ’s shape detection methods against morphological changes in ultrasound images are evaluated. Also, the deformed kidney shape in the ultrasound volumes of abnormal subjects mimic deformations of damaged kidney due to injury. We applied the kidney detection methods on the evaluation set of actual ultrasound volumes of abnormal subjects ($\text{Ds3DUsIs-AUVs-ESA}$), and the kidney detection accuracies are $\frac{5}{8} \cdot 100\% = 62.50\%$, $\frac{3}{8} \cdot 100\% = 37.50\%$, $\frac{2}{8} \cdot 100\% = 25.00\%$, and $\frac{1}{8} \cdot 100\% = 12.50\%$ for the atlas-based kidney detection, shape-based kidney detection, Marsousi et al-EMBC14 [63], and Noll et al. [12] methods, respectively. Accordingly, the atlas-based method provides a higher accuracy of detection for the with-kidney images of the abnormal subjects. The detection accuracy for the with-kidney images of the trauma patients is lower than for the subset of healthy volunteers, because the right kidney’s shape and its surrounding tissues in the RUQ view of trauma patients appear distorted in the 3D ultrasound images compared to those of the normal subjects, which makes it more difficult for the kidney detection
methods to correctly detect the kidney shape. More importantly, ultrasound volumes of abnormal subjects were taken with a different settings of the imaging device with respect to the imaging device setting for scanning the healthy volunteers. Thus, the kidney shape in ultrasound volumes of abnormal subjects appear much bigger than the kidney shape in acquired volumes from healthy volunteers.

We use the class separability index (CSI in equation (2.50)) to compare the separability power of the \textit{with-kidney} images from the \textit{without-kidney} images using both the shape-based and atlas-based methods. The calculated CSI of the shape-based and atlas-based kidney detection methods are 13.8164 and 19.9663, respectively. Accordingly, compared to the shape-based method, the atlas-based method has a higher CSI, confirming that the atlas-based method provides a better separability of the \textit{with-kidney} images from the \textit{without-kidney} images. The higher CSI of the atlas-base method also indicates that the atlas-based method may provide a higher accuracy of separating \textit{with-kidney} from \textit{with-kidney} images in new ultrasound volumes.

We also compare the Euclidean distance errors of the center points of the detected kidney shapes, with respect to their actual center points, for the \textit{with-kidney} images in the \textit{Ds3DUsIs-AUVs-ESH} dataset. The variations of Euclidean distance errors of the methods are shown in box-plot graph of Fig. 2.27. In the box-plot graphs of Fig. 2.27, the central red marks represent the median values, the edges of the blue boxes specify the 25th and 75th percentiles, the edges of whiskers show most extreme data points not considering outliers, and outliers are plotted by red plus signs. To represent the statistical variation of computational time, the distance between the 25th percentile and 75th percentile, so-called the inter-quartile range, is selected because it is the most significant basic robust measure of scale. The mean and standard deviation of the Euclidean distance error of the atlas-based kidney
Figure 2.28: The table in this figure represents the statistical comparison of the kidney detection methods for estimating the center-point of the kidney shape in the with-kidney images of the evaluation set of actual ultrasound volumes of the healthy volunteers (Ds3DUsIs-AUVs-ESH). $h$ is the state of the t-test analysis for each pair of methods, based on the significance level of 5%. For each non-diagonal cell of the table, $h = 1$ means the corresponding method on the left column performs better than the corresponding method in the top row, $h = -1$ means the corresponding method in the top column performs better than the corresponding method in the left row, and $h = 0$ means the experimental data of the corresponding methods do not show a significant difference.

Figure 2.29: This figure shows the steps of shape-based and atlas-based kidney detection methods for five with-kidney actual images of the Ds3DUsIs-AUVs-ESH dataset. Each row belongs to an actual ultrasound volume.
detection, shape-based kidney detection, Noll et al. [12], and Marsousi et al-EMBC14 [63] methods are 18.5599 ± 8.3413, 21.9365 ± 6.7656, 30.1491 ± 12.1050, and 31.1480 ± 16.5811, respectively. Accordingly, the atlas-based method provides a lower mean of the Euclidean distance error, compared to the other methods. These results show that the atlas-based kidney detection method provides a more accurate estimation of the kidney shape’s alignment in the actual ultrasound volumes. These results also match with our experimental findings in the sensitivity analysis of Sec 2.9.5.1, in which results with simulated ultrasound volumes demonstrate improved alignment estimations of the kidney shape using the atlas-based method, compared to the shape-based approach.

We also performed statistical analysis, based on the t-test method, to determine the significance of comparing kidney detection results. The result of statistical analysis is shown in Fig. 2.28, which shows the state of test \( h \) and the \( p \)-value for each test. For each cell in the table shown in Fig. 2.28, assume the method-a and method-b are corresponding left column and top row, respectively. Based on this assumption, the state of test, \( h \), can take values from \( \{-1, 0, 1\} \), in which a \(-1\) means method-b performs better than method-a, a \(1\) means method-a performs better than method-b, and a \(0\) means we cannot conclude which method performs better. The results in Fig. 2.28 are obtained based on the 5% significance level. Accordingly, the atlas-based method performs better than the Noll et al. [12], and Marsousi et al-EMBC14 [63] methods.

Figure 2.29 shows five examples of kidney detection with the shape-based and atlas-based methods. As shown in Figure 2.29, both of the shape-based and atlas-based kidney detection methods provide more or less similar detection outputs, at-least in these five examples. However, note that because of the higher CSI of the atlas-based kidney detection method, it is expected to be a more reliable approach to distinguish between with-kidney and without-kidney images. This advantage of the atlas-based method is achieved because of using both of the shape and texture prior knowledge of the kidney shape from training ultrasound volumes.

\subsection*{2.9.5.3 Computational time analysis}

The computational time of the proposed shape-based and atlas-based kidney detection methods should be about few seconds to allow real-time interaction with operators to provide an applicable computer assisted-probe placement. Note that the speckle reduction process with the MFLV method is performed in the half resolution, and therefore, the computational time is less than what is reported in Sec. 2.3.2 (the computational time of the MFLV method in Table 2.5 is 4.96 (sec), which is obtained using the GPU accelerated functions for the full size volume). The computational time of the proposed solution is measured for actual ultrasound images. Figure 2.30 demonstrates the total computational times of
Chapter 2. Kidney detection

Figure 2.30: This bar-graph shows the total computational time of the shape-based and atlas-based methods and the computational times of three parts of the shape-based and atlas-based methods, including LHE, shape-to-volume registration, and voxel classification, which implementational approach highly affects their computational time.

The computational time of the MFLV and G2LTh modules do not depend on the implementational method. As shown in Figure 2.30, the combination of GPU and multi-threading (multiple CPU cores) provides the lowest computational time for both the shape-based and atlas-based methods. The shape-based method can detect the kidney shape in less than ten seconds, which satisfies the requirement of the targeted application of this thesis (computer-assisted probe placement). In order to reduce the computational time of the atlas-based method to meet the requirements of computer-assisted probe placement, a portable computer with a higher capability of parallel computation is required.

2.10 Summary and Concluding Remarks

In this chapter, we represented two approaches, including shape-based and atlas-based methods, to detect the kidney shape in 3D ultrasound volumes. The proposed methods were able to decide whether the kidney shape exists in an ultrasound volume or not, and if the kidney shape exists, what is its alignment in terms of orientation, scaling, and translation, with respect to the reference kidney shape. The proposed shape based kidney detection method used shape prior knowledge of the kidney shape to perform kidney detection, and the atlas-based kidney detection used both shape and texture prior information to detect the kidney shape in volumetric ultrasound images. Through experimentation, the atlas-based method
showed more reliable kidney detection and more accurate kidney shape localization\textsuperscript{12}, at the cost of excessive computational complexity. We also discussed the utility of the proposed shape-based and atlas-based kidney detection methods in providing a computer-assisted probe placement solution to be used by paramedics in emergency situations, where trauma patients’ lives tie to rapid diagnosis and medical services. In the next chapter, we will show how the proposed kidney detection method of this chapter can be used to provide a fully automated kidney segmentation process.

\textsuperscript{12}Here, localization means to find the alignment of the kidney shape inside an ultrasound volume.
Chapter 3

Automated kidney segmentation
Chapter 3. Automated kidney segmentation

3.1 Introduction

The task of segmenting the kidney’s shape in three-dimensional (3D) ultrasound images has a paramount medical significance in computer-assisted diagnosis of trauma patients, suffering from abdominal bleeding in emergency situations. As explained in Sec 1.1, the right upper quadrant (RUQ) view in abdominal Sonography, visualizing the right kidney, has been recognized as the most relevant view to the trauma diagnosis, because a free fluid due to an abdominal bleeding in a trauma patient laid in the supine position, has a high tendency to locate around the upper boundary of the right kidney. Thus, correct segmentation of the kidney’s shape from its surrounding tissues in ultrasound images is essential for trauma diagnosis.

Although the task of kidney segmentation has been extensively investigated in applications where imaging modalities such as CT and MRI scanners were used [124–126], there are relatively few works in automated kidney segmentation for ultrasound imagery. This is because automated kidney segmentation in ultrasound imagery face many challenges, discussed in Sec 1.5.

Kidney segmentation in 2D ultrasound images has been reported by some researchers, such as Xie et al. [58], Huang et al. [61], and some others [61,127–130]. However, they are either not applicable or not easily extendable into 3D kidney segmentation. Xie et al. [58] proposed a kidney segmentation method using texture and shape priors. In their method, features from inside and outside of the segmented contour are extracted with Gabor filters. A kidney shape model is then generated using a set of training kidney shapes based on the so-called point distributed model (PDM) approach [67]. In Xie et al. method, a level-set function is initialized with a manually aligned shape model. Their shape model is progressively evolved using energy minimization, driven by two-sided features, until a convergence is attained. The method was shown to be robust against gaps within the kidney boundary. In their method, the level-set function should be manually initialized close to the actual kidney shape, and therefore, their method provided semi-automated kidney segmentation. Huang et al. [61] proposed a 2D ultrasound segmentation method which used a super-ellipsoid as a parametric shape model to represent the kidney shape, and then, it applied a region-based level-set propagation to segment the kidney. Since their method did not consider texture information and detailed model of the kidney shape, it required manual supervision to prevent the elliptical shape from being fitted on non-kidney structures.

Some other papers have reported kidney segmentation in 3D ultrasound volumes [12,13,63,64,116]. Fernandez and Lopez [13] proposed a segmentation method based on a combination of Markov random field and active contours (MRF-AC) to segment the kidney’s shape in 3D ultrasound images. Their
method segmented the kidney’s shape in all 2D slices of an ultrasound volume, and then, a 3D surface was built-up by concatenating the segmented 2D contours. As a result, information along the $z$–axis was not directly used in the segmentation task, resulting in a shape discontinuity along the $z$–axis. In addition, the method in [13] required an operator to initialize the kidney segmentation task. In another attempt, Cerrolaza et al. [116] applied an active appearance model (AAM) method based on Gabor features to segment the kidney’s shape in 3D ultrasound images. In this method, a point distribution model (PDM) of the kidney’s shape was generated from a set of training shapes. The PDM was placed on a 3D ultrasound image, and iteratively deformed to segment the kidney’s shape. Their method calculated smoothed directional gradients at each PDM’s landmark, using 3D Gabor filters, to move landmarks toward the kidney boundary. This method required an operator to define the major axis of the kidney’s shape, and therefore, it was semi-automated. Ardon et al. [64] proposed a method to automatically segment the kidney shape in 3D ultrasound images. Their method adopted kidney detection to automatically extract bounding box of the kidney shape, and the extracted bounding box was used to initialize kidney segmentation based on an implicit deformation framework [131]. For kidney detection in an input 3D ultrasound image, a learnt-based kernel with different poses was convoluted with the input image to find a good matching with the kidney shape of the input image. A SVM classifier was trained and used to decide which pose and alignment of the convolution kernel provided a better matching with the kidney shape in a 3D ultrasound image. The fitted kernel was then used to initialize the deformable model to segment the kidney shape. To improve the segmentation accuracy, their solution provided a tool for operators to insert landmark points on the kidney shape, indicating points situated inside or outside the kidney shape, and the points were directly imposed in the evolution formula of the implicit deformable model.

Noll et al. [12] proposed a method to detect and segment the kidney’s shape in 3D ultrasound volumes. In their method, ultrasound volumes were first enhanced to reduce speckle noise and improve intensity contrast. Then, a deformable model method based on radial rays was used to detect the kidney’s shape. Afterward, the detected kidney’s shape was used to initiate the fast marching method [132], and then, the output of the fast marching method was used to initiate a level-set function. The level-set function was evolved using edge-based information [133] to segment the kidney’s shape. Marsousi et al. [63] proposed a shape-based kidney detection/segmentation method. A kidney shape model was generated using a set of manually segmented kidney shapes in a few training 3D ultrasound images. First, each input 3D ultrasound image was preprocessed to be enhanced, and then, a template matching algorithm was applied to search for the best fit of the shape model on the enhanced image. If the matching value was greater than a specific threshold, the method decided that the kidney’s shape exists in the 3D
ultrasound image. If the kidney’s shape was detected, the kidney shape model was used to initialize a level-set function, and finally, the kidney shape was segmented by evolving region-based level-set method. This method had two drawbacks: (a) the applied template matching in [63] might not be robust against kidney shape deformations, and (b) the applied region-based level-set might not be robust against kidney intensity inhomogeneity. Noll et al. [12] and Marsousi et al. [63] were distinctive among all the other kidney segmentation methods as they were fully-automated. Table 3.1 summarizes characteristics of the kidney segmentation methods. According to the disadvantages of the methods listed in Table 3.1, none of the prior arts of kidney segmentation in 3D ultrasound imagery matches with the requirements of the targeted application of this thesis (ie. computer-assisted trauma diagnosis in emergency situation).

In the rest of this chapter, we first provide a literature review on 3D segmentation methods which are commonly used in medical applications, and then, we propose a new segmentation approach to be used for automated kidney segmentation in 3D ultrasound images.

Table 3.1: Summarizing characteristics, including Manual Initialization (Man. Init.), automated segmentation (Auto. Seg.), methodology, advantages, and disadvantages, of the kidney segmentation methods for 3D ultrasound images.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Marsousi et al. [63]</td>
<td>✓</td>
<td></td>
<td>Template Matching + Level-Set</td>
<td>Low comput. cost</td>
<td>Not robust against kidney deformation/occlusion</td>
</tr>
<tr>
<td>Noll et al. [12]</td>
<td></td>
<td>✓</td>
<td>Radial Ray Trace + Fast Marching + Level-Set</td>
<td>Robust against speckle and low-contrast intensity profile</td>
<td>High-comput. cost, not robust against kidney deformation</td>
</tr>
<tr>
<td>MRF-AC [13]</td>
<td>✓</td>
<td></td>
<td>2D Active Contour + MRF + 3D Reconstruction</td>
<td>Low comput. cost</td>
<td>Discontinuity along z-axis</td>
</tr>
<tr>
<td>Ardon et al. [64]</td>
<td>✓</td>
<td>✓</td>
<td>Convolutional Kernel based SVM + Deformable Model</td>
<td>Moderate robustness against kidney deformation</td>
<td>Requires a large training set of ultrasound volumes</td>
</tr>
<tr>
<td>Cerrolaza et al. [116]</td>
<td>✓</td>
<td></td>
<td>3D PDM + 3D Gaussian AAM</td>
<td>Using texture information</td>
<td>Needs operator’s intervention</td>
</tr>
</tbody>
</table>
3.2 3D segmentation methods: review

In contrast with 2D medical image processing, medical image processing in 3D domain provides higher valued information about the state of the nature, and increases the level of certainty of automated diagnosis [134], [135]. 3D image processing improves the quality of volume calculation, measurement, and quantitative analysis [136]. For example, the Left Ventricular volume, which is an essential heart quantity, has been shown to be more accurately estimated from 3D echocardiography volumes rather than using 2D echocardiography images [137]- [138]. As another example, image guided surgery is facilitated by merging 3D pre-operative or intra-operative images and reality [136]. Due to the recent processor’s enhancements, such as speed and parallelism capability, the realization of near real-time 3D segmentation becomes more affordable [139], [140]. Hence, we have been witnessing a rapid growth of researchers’ contributions in 3D medical image processing.

Many works have been performed to overcome challenges toward 3D medical image segmentation. Freedman et al. [141] categorized 3D segmentation challenges as follows:

1. the object of interest is usually spread out in the volume, and edges are not strong;
2. in addition to the desired object, other undesired objects exist in the 3D image which might or might not resemble the object of interest;
3. there is no significant difference between pixels’ intensity profile of desired object and undesired objects in the 3D image;
4. the desired object and undesired objects commonly have similar shapes.

To be more general, we add two more difficulties as follows:

1. interfering artefacts obscure the clarity of the object of interest;
2. varying intensity levels of pixels through similar tissues of an object of interest makes the segmentation process even harder.

Depending on the type of medical imaging modality, some of these challenges are more significant. 3D ultrasound images are well known to have poor resolution, low quality, unclear edges of objects, high level of speckle interference, and inhomogeneous intensity profile of the object boundary [142]. In order to correctly select a segmentation method for a medical application, not only a complete understanding on the application-based challenges is required, but also a good knowledge on segmentation methods’ capabilities is needed.
Figure 3.1: This figure categorizes the commonly used segmentation methods in medical imaging.

Many survey papers have been published to characterize and compare 3D segmentation approaches. Some of these papers are application-specific surveys [142–146], investigating related publications to a particular application. Some other survey papers are method-based surveys, which mostly focus on categorizing segmentation methods, and comparing their characteristics, such as mathematical representation and topological flexibility [147–153]. Based on our literature survey, we categorize 3D segmentation methods into two main categories, including region-based methods and deformable models. Each of these two categories can be categorized into sub-groups, shown in Fig. 3.1. Each sub-category of segmentation methods has special characteristics, making it robust against some challenges and weak against some other ones.

Deformable models are referred to surfaces which progressively deform based on differential equations, aiming to fit data inside images [154]. Deformable models are generally categorized into explicit representation and implicit representation methods. Region-based methods perform volumetric segmentation in which each voxel is assigned with a regional label, compared to deformable models which use surface modeling to separate different regions. The region-based methods can be classified into five categories including, thresholding [155], region growing [156–158], clustering [159], watershed [160–162], and graph-cuts [163–165]. In general, deformable models are more controllable in terms of shape and topology, whereas region-based methods have less control on the geometry of labeled voxels. Since deformable models provide more geometrical control over the segmented region, they are more suitable for segmenting the kidney shape in 3D ultrasound images. Therefore, the focus of this literature review is on deformable models.
3.2.1 Deformable models

The deformable models emerged in mid 1980’s when the elastically deformable model was proposed by Terzopoulos et al. [166]. Soon after, Kass et al. [167] proposed the active contour model, also known as the “Snake”, which has attracted a lot of attentions among the image processing society, and afterward, deformable models have diverged into many different classes of segmentation methods. Each class of deformable models can be characterized based on two main attributes: surface representation model, and surface evolution approach [147].

The surface representation model plays an important role in forming behaviors of deformable models. Deformable models which apply a parametric model to represent voxels’ coordinates on the surface model are classified as explicit representation [151]. In contrast, deformable models which define an object boundary by an iso-value surface of a function are classified as implicit representation [147].

In the explicit representation category, the surface is represented as \( \vec{v}(s, r) = [x(s, r), y(s, r), z(s, r)]^T \) where \( x, y, \) and \( z \) are continuous coordinate functions, and \( s, r \in [0, 1] \) form the parametric domain [168]. (digitized format \( \vec{v}(s_k, r_l) : s_k = k/N_s \text{ where } k \in [1, ..., N_s] \text{, and } r_l = l/N_r \text{ where } l \in [1, ..., N_r] \).) In the explicit representation, both shape flexibility and computational complexity are proportionally related to the number of parameters [147]. The explicit representation category can be divided into two sub-categories: continuous models, and discrete models [169]. The continuous representation provides the ability to compute differential equations, surface normals, and curvatures for the deformable surface [169]. 3D Snake, B-Spline surface, superquadrics, and probabilistic models are in the continuous representation sub-category. In the discrete models sub-category, the surface is formed with a set of connected voxels. Particle systems, spring-mass model, triangulation, T-surface, simplex mesh and SpringLS are in the discrete explicit representation sub-category.

In the implicit representation category, the surface is represented as \( \vec{v} = [(x, y, z) : \phi(t, x, y, z) = k]^T \) where \( \phi(.) \) is a level-set function, \( t \) is the time, and \( x, y, \) and \( z \) are coordinate functions, and \( k \) refers to a specific level. This representation is also known as the level-set [170]. Segmentation methods which apply the implicit representation are topologically flexible. In some medical applications in which regions of interest are apart in the image, the topological flexibility of the implicit representation becomes very advantageous. However, in some other medical applications which a single connected region is targeted, the topological flexibility may result in incorrect segmentation [171]. In addition, the implicit representation adds up one extra dimension to the segmentation domain, and therefore, its computational cost is high [172].

The second important aspect of deformable models is the mathematical approach to handle the
deformable surface evolution. Most of deformable models define an energy minimization problem to evolve their surfaces. The energy functional should contain at least two terms: regularization term, and external energy [147]. The regularization term, also known as the internal energy, controls the smoothness (and/or stiffness) of the surface. The external energy is required to fit the surface on the image data. The total energy, $E_{\text{total}} = E_{\text{reg}} + E_{\text{ext}}$, is a functional which is usually solved using the Euler-Lagrange equation. The Euler-Lagrange solution is then discretized, based on partial difference equations (PDE), to evolve the deformable surface.

Deformable models apply more or less similar internal energies, whereas many different external energies have been proposed to accommodate application-specific segmentation requirements. Internal energies are usually defined by minimizing first- (and/or second-) order derivative(s) of the deformable surface. In general, external energies can be categorized into edge-based and region-based [173]. Edge-based external energies apply image gradient information. Two drawbacks associated with edge-based external energies are sensitivity to noise, and dependency on the initial surface [173]. On the other side, region-based external energies are more robust against noise, and are less dependable to the surface initialization. However, region-based external energies require prior knowledge of regions’ intensity profiles. In addition, region-based methods are sensitive to intensity variability of regions [173]. In the rest of this section, we provide brief introduction on the deformable model segmentation methods.

3.2.1.1 Continuous explicit representation

3.2.1.1.1 Parametric model (3D snake): The active contour model, also known as the snakes, was first introduced by M. Kass et al. [167] in two-dimensional space. Soon after the introduction of 2D snake, it was extended into 3D segmentation [168], [174]. The snake method uses a parametric explicit representation to form a segmentation surface. In the parametric model, a surface is represented as a mapping function from a 2D parametric space into a 3D space, $\Omega = [0, 1] \times [0, 1] \rightarrow \mathbb{R}^3$, as follows [175],

$$v(s, r) = [\bar{x}(s, r), \bar{y}(s, r), \bar{z}(s, r)]^T,$$

where $s$ and $r$ are elements of the parametric space. The surface evolution is controlled by minimizing an energy functional, defined as follows [176],

$$E(v) = \int_{\Omega} \left( w_s \left\| \frac{\partial v}{\partial s} \right\|^2 + w_r \left\| \frac{\partial v}{\partial r} \right\|^2 + \left( w_{ss} \left\| \frac{\partial^2 v}{\partial s^2} \right\|^2 + w_{rr} \left\| \frac{\partial^2 v}{\partial r^2} \right\|^2 \right) + 2 \cdot w_{sr} \left\| \frac{\partial^2 v}{\partial r \partial s} \right\|^2 + P(v(r, s)) \right) dsdr,$$

where $P(v(r, s))$ is the external potential which attracts the surface toward objects’ boundaries.
equation (3.2), the first term, containing Lagrange multipliers \(w_s\) and \(w_r\), controls the surface smoothness by minimizing the first-order derivative. The second term in the functional, containing Lagrange multipliers \(w_{ss}\) and \(w_{rr}\), controls the surface stiffness like a thin plate \([175]\). Using the Euler-Lagrange equation, an iterative surface evolution is obtained as follows \([175]\),

\[
\frac{\partial v}{\partial t} - w_s \frac{\partial^2 v}{\partial s^2} - w_r \frac{\partial^2 v}{\partial r^2} + 2w_{rs} \frac{\partial^4 v}{\partial r^2 \partial s^2} + w_{ss} \frac{\partial^4 v}{\partial s^4} + 2w_{rr} \frac{\partial^4 v}{\partial r^4} = F_{ext}(v(s, r)),
\]

where \(F_{ext}(v(s, r))\) denotes the external force. According to the derived minimization problem, the surface starts with an initial state, \(v(s, r, t = 0)\), and then, it iteratively shrinks until an external force stops the surface evolution. One regular choice for the external force is the image gradient (edge-based force), \(F_{ext} = \frac{\nabla V}{|\nabla V|}\), where \(V \in \mathbb{R}^3\) is a 3D image. As a challenge toward using image gradient, points on the surface located inside and away from object boundary are both affected by the gradient vectors, and they do not move toward the desirable edges of the object of interest. One approach to address this problem is to use an inflation force (balloon force), as \(F_{balloon} = -\gamma_{balloon} \vec{n}\), where \(\vec{n}\) is an inward normal vector to the surface at each surface point \(v(s, r)\) \([175]\). As a problem of using the balloon force, its weight should be correctly selected, or otherwise, the edge-based force would not be able to stop the surface from exceeding into a non-desirable region.

The gradient vector flow (GVF) is a remedy to the problem of the image gradient force \([177]\). The GVF extends image gradient vectors into homogenous regions away from edge-based information. Let's define the GVF force as \(F_{GVF}(\vec{v}) = [u_x(\vec{v}), u_y(\vec{v}), u_z(\vec{v})]\). In order to propagate gradient vectors through homogenous regions, an energy functional is formulated as follows \([178]\),

\[
\epsilon(F_{GVF}) = \int \int \mu (\frac{\partial u_x}{\partial x} + \frac{\partial u_y}{\partial y} + \frac{\partial u_z}{\partial z})^2 + \frac{\partial u_x}{\partial x}^2 + \frac{\partial u_y}{\partial y}^2 + \frac{\partial u_z}{\partial z}^2 + \frac{\partial u_x}{\partial x} \frac{\partial u_y}{\partial y} + \frac{\partial u_y}{\partial y} \frac{\partial u_z}{\partial z} + \frac{\partial u_z}{\partial z} \frac{\partial u_x}{\partial x} + |\nabla V|^2 |F_{GVF} - \nabla V|^2 dxdydz.
\]

The second term of equation (3.4) forces the GVF to follow gradient vectors in regions with higher gradient amplitudes, \(|\nabla V|\). In other regions, the first term smooths the transition of gradient vectors from object boundary into homogenous regions. The energy functional of equation (3.4) is solved by the Euler-Lagrange equation \([178]\). The 3D snake is flexible in terms of surface smoothness and robustness against noise. However, its computational cost proportional increases by increasing the number of sample points on the surface. To increase segmentation details, more sample points are required on the surface.
which results in an exponential increment of its computational cost. Therefore, except a few cases (such as [179]), this method has not been used for 3D medical image segmentation.

3.2.1.1.2 Geodesic model: The Geodesic method is a particular format of 3D snake, which applies geometric strategy to evolve the 3D surface. In the Geodesic method, the second-order derivative (stiffness controlling term) is eliminated. Having this modification, the smoothness of the surface is decreased, however, it allows to derive a geometric surface evolution [180]. The energy minimization functional is formulated as follows [181],

\[
E(v) = \int \int \left( w_s \left\| \frac{\partial v}{\partial s} \right\|^2 + w_r \left\| \frac{\partial v}{\partial r} \right\|^2 \right) \\
+ 2 \left( w_{sr} \left\| \frac{\partial^2 v}{\partial r \partial s} \right\|^2 \right) + \lambda g \left( |\nabla V(v(r,s))| \right)^2 dsdr,
\]

(3.5)

where \( g (|\nabla V|) \) is a function of the image gradient. By taking some mathematical derivations, the energy functional of equation (3.5) can be re-formulated into the following minimization problem [180],

\[
E(v) = \int \int g (|\nabla V(v(r,s))|) \left| \frac{\partial v}{\partial s} \times \frac{\partial v}{\partial r} \right| dsdr,
\]

(3.6)

where \( \int \int \left| \frac{\partial v}{\partial s} \times \frac{\partial v}{\partial r} \right| dsdr \) is the surface area. By solving equation (3.6) using the Euler-Lagrange equation, the following surface flow equation is obtained [182],

\[
\frac{\partial v}{\partial t} = \left( g_{\nabla V}(\kappa + \alpha_{balloon}) - \nabla g_{\nabla V} \cdot \vec{N} \right) \vec{N},
\]

(3.7)

where \( \kappa \) is the arithmetic mean of the principal curvature, and \( \alpha_{balloon} \) is a constant speed term which is inspired by the balloon force [175]. \( g_{\nabla V} = g (|\nabla V(v)|) \) can be defined as follows,

\[
g (|\nabla V|) := \frac{1}{1 + \| \nabla G_\sigma * V \|^2},
\]

(3.8)

where \( G_\sigma \) is a Gaussian FIR filter. In homogenous regions \( g_{\nabla V} \) is approximately equal to one, whereas in edge regions, \( g_{\nabla V} \) is close to zero. Based on the minimization equation (3.6), the portion of surface area corresponding to homogenous regions is minimized since \( g_{\nabla V} \) is close to its maximum value. Therefore, the surface moves in homogenous regions. In edge regions, \( g_{\nabla V} \) is close to zero, and thus, the area of the corresponding portion of the surface in edge region is not minimized, which results in stopping the surface in edge regions. \( g_{\nabla V} \) is also called the stopping factor. As an advantage of the geometric model
over the parametric model, the geometric model is free of parametrization. However, the geometric model is sensitive to local minima, and also fails to stop at weak edges [183].

### 3.2.1.1.3 B-spline surface:

The B-Spline surface was first proposed by Liao and Medioni [184] as an extension of B-Spline Snake [185] into the 3D space. In the B-Spline surface, the explicit representation is parametrically modeled using B-Spline functions as follows [186],

\[
\mathbf{v}(s, r) = \sum_i \sum_j Q_{i,j} B_3^i(s) B_3^j(r),
\]

where \( Q_{i,j} \)s are control points, \( B_3^i(s) \) and \( B_3^j(r) \) are the B-Spline functions (see Fig. 3.2). This representation provides easy calculation of differential characteristics such as surface normals or curvatures of the surface [186]. In addition, the B-Spline surface implicitly has surface smoothness constraint, and therefore, the internal energy term is not needed [3]. This fact contributes in a reduced computational cost. By replacing bi-cubic b-spline functions in equation (3.9), it is represented in a matrix format as follows [3, 187],

\[
\mathbf{v}_{i,j} = \mathbf{s} \mathbf{M}_R \mathbf{Q}_{R_{i,j}} \mathbf{M}_R^T \mathbf{r}^T,
\]

where

\[
\mathbf{M}_R = \begin{bmatrix}
-1/6 & 1/2 & -1/2 & 1/6 \\
1/2 & -1 & 1/2 & 0 \\
-1/2 & 0 & 1/2 & 0 \\
1/6 & 2/3 & 1/6 & 0
\end{bmatrix},
\]

\[
\mathbf{Q}_{R_{i,j}} = \begin{bmatrix}
Q_{i+1,j+1} & Q_{i+1,j} & Q_{i+1,j+1} & Q_{i+1,j+2} \\
Q_{i-1,j-1} & Q_{i-1,j} & Q_{i-1,j+1} & Q_{i-1,j+2} \\
Q_{i+1,j-1} & Q_{i+1,j} & Q_{i+1,j+1} & Q_{i+1,j+2} \\
Q_{i+2,j+1} & Q_{i+2,j} & Q_{i+2,j+1} & Q_{i+2,j+2}
\end{bmatrix},
\]

\[
\mathbf{s} = \begin{bmatrix}
s_1^3 & s_1^2 & s_1 & 1 \\
s_2^3 & s_2^2 & s_2 & 1 \\
\vdots & \vdots & \vdots & \vdots \\
s_N^3 & s_N^2 & s_N & 1
\end{bmatrix},
\]

\[
\mathbf{r} = \begin{bmatrix}
r_1^3 & r_1^2 & r_1 & 1 \\
r_2^3 & r_2^2 & r_2 & 1 \\
\vdots & \vdots & \vdots & \vdots \\
r_N^3 & r_N^2 & r_N & 1
\end{bmatrix},
\]
where \( s_k = k/m \) and \( r_k = k/n \). In the B-Spline surface evolution, external forces on surface points are projected into control points. This results in a low-pass filtering effect on the external force, and provides robustness of surface evolution against noise. At each iteration of surface evolution, control points are updated using the following equation,

\[
\Delta Q_{i,j} = \alpha M_R^{-1} \left( s^T s \right)^{-1} s^T F_{i,j} r \left( r^T r \right)^{-1} \left( M_R^T \right)^{-1},
\]

where \( \alpha \) is the updating weight of control points. \( F_{i,j} \) is a matrix of external forces on surface points.

This method has received a great attention in segmenting left ventricle in echocardiography images [186,188–191].

3.2.1.1.4 Superquadrics: Superquadrics were first applied for 3D segmentation by Solina and Bajcsy [192]. A superquadric surface is a parametric representation which applies least square error minimization to fit the superquadric surface on an image data. It was first introduced by [192] to parametrically represent a single object in an input 3D image. A superquadric surface is explicitly defined as follows [192],

\[
S(s,r) = \begin{bmatrix}
a_x \cos^2(\pi(s - 0.5)) \cos^2(2\pi(r - 0.5)) \\
a_y \cos^1(\pi(s - 0.5)) \sin^2(2\pi(r - 0.5)) \\
a_z \sin^2(\pi(s - 0.5))
\end{bmatrix},
\]

where \( 0 \leq s, r \leq 1 \) are two independent angle parameters. \( a_x, a_y, a_z \) are superquadric size in \( x-, y- \) and \( z- \) coordinates, respectively. \( \epsilon 1 \) and \( \epsilon 2 \) are squareness parameters in latitude and longitude planes [192]. Figure 3.3. shows the effect of squareness parameters on the superquadric surface. The equation (3.13)
Chapter 3. Automated kidney segmentation

Figure 3.3: 3D superquadrics. (a) $\epsilon_1 = \epsilon_2 = 0.1$; (b) $\epsilon_1 = \epsilon_2 = 0.5$; (c) $\epsilon_1 = \epsilon_2 = 1$; (d) $\epsilon_1 = \epsilon_2 = 2$; (e) $\epsilon_1 = \epsilon_2 = 4$ [4].

can be represented in an implicit equation [192],

$$ F(x, y, z) = \left( \left( \frac{x}{a_x} \right)^{\frac{2}{\epsilon_1}} + \left( \frac{y}{a_y} \right)^{\frac{2}{\epsilon_2}} \right)^{\frac{\epsilon_2}{\epsilon_1}} + \left( \frac{z}{a_z} \right)^{\frac{2}{\epsilon_1}} = K, \quad (3.14) $$

where $K = 1$, $K < 1$ and $K > 1$ correspond to voxels on the surface, inside the surface, and outside the surface, respectively. Therefore, the voxels on the surface are implicitly represented as $F(x, y, z) = 1$. To generalize the model, rotations $\theta_x, \theta_y, \theta_z$ and translation $t_x, t_y, t_z$ of voxels respect to the origin are modeled as follows [192],

$$ \begin{bmatrix} v(r, s) \\ 1 \end{bmatrix} = \begin{bmatrix} R_{3 \times 3}(\theta_x, \theta_y, \theta_z) & t_x \\ 0 & 0 & 0 & 1 & t_y \\ 0 & 0 & 0 & 1 & t_z \end{bmatrix} \begin{bmatrix} S(r, s) \\ 1 \end{bmatrix}, \quad (3.15) $$

where $R_{3 \times 3}(\theta_x, \theta_y, \theta_z)$ is the 3D rotation matrix. The model in equation (3.15) has 11 parameters as $\Lambda = [a_x, a_y, a_z, \epsilon_1, \epsilon_2, \theta_x, \theta_y, \theta_z, t_x, t_y, t_z]$. To fit the superquadric surface on the data, the following minimization problem needs to be solved [192],

$$ \min_{\Lambda} \iiint |1 - F(x, y, z, \Lambda)| \, dx \, dy \, dz. \quad (3.16) $$

The fitness accuracy of the superquadric surface highly depends on its initialization. As recommended in [192], the superquadric surface is initialized by an 3D ellipsoid ($\epsilon_1 = 1, \epsilon_2 = 1$).

Terzopolous and Metaxas [5] proposed the deformable superquadrics by including a local deformation
Chapter 3. Automated kidney segmentation

Figure 3.4: Displaying the geometry of the deformable superquadric surface [5].

Based on splines to the superquadric surface (Figure 3.4). Each point on the deformable superquadric surface is represented as follows,

\[
v(r, s) = \vec{c} + R_{3 \times 3}(\theta_x, \theta_y, \theta_z) \times (S(r, s) + d(r, s)),
\]

and,

\[
d(r, s) = \sum_i \sum_j Q_{i,j} B^i_s(r) B^j_r(s),
\]

where \(\vec{c}\) is the object center, \(B_i(s)\) and \(B_j(r)\) are splines corresponding to \(s\) and \(r\) parametric coordinates, respectively. \(Q_{i,j}\)s are node points and specify the degree of freedom [5]. The bazier spline is utilized by Zhou and Kambhamettu [4] because it easily models high curvatures without erratic oscillations. Since internal organs usually have regular shapes, the deformable superquadric surface has been widely used to segment internal organs, such as cardiac segmentation in MRI and CT images [193], ventricular segmentation in echocardiograms [194], prostate segmentation in ultrasound images [195] and vasculature modeling [196].

The original superquadric model was extended by Leonardis et al [6] to segment objects with multiple parts. This method contains two intertwined stages: model-recovery and model-selection. In the model-recovery stage, a set of seeds are placed in an image to initiate the procedure. Then, they are independently allowed to iteratively grow up until the whole data is recovered. Afterward, the model-selection stage is applied to select the minimum number of superquardics which provides a low recovery error (Figure 3.5). This multi-part segmentation approach is followed by later works [197], [198], [199].
Chapter 3. Automated kidney segmentation

3.2.1.1.5 Active rays: Despite several advantages that active contour models provide for image processing tasks, they have some shortcomings as follows,

- active contour models may trap in non-desired local edges;
- there is no control over surface samples' orders, and samples crossings may occur during energy minimization [200]

To address these two shortcomings, the 2D active rays model was proposed by Denzler et al. [201] which contains principles of active contour models. Recently, this approach is extended to 3D segmentation by Steger et al. [7]. In this approach, \( N_{Ray} = N_s \times N_r \) rays, centered at \( \bar{c}_{Ray} \), are evenly distributed in a sphere of radius, \( r_{Ray} \) (Figure 3.6). Each ray is evenly divided into \( m_{Ray} \) bins. Thus, the coordinate of each surface sample on its corresponding ray \((s, r)\) is defined as follows,

\[
v(s, r) = \bar{c}_{Ray} + \left( \frac{r_{Ray}}{m_{Ray}} \right) \bar{d}_{Ray}(s, r),
\]

(3.19)

where \( l(s, r) \) is a label which specifies the assigned bin on the \((s, r)\) ray. \( \bar{d}_{Ray}(s, r) \) is a vector with a unit-norm \( \mathbb{2} \) which specifies a direction for its corresponding ray, and is obtained as, \( \bar{d}_{Ray}(s, r) = R_{\theta_z(s)} R_{\theta_y(r)} \left[ \begin{array}{ccc} 1 & 0 & 0 \end{array} \right]^T \), where \( \theta_z(s) = \pi(s - 0.5) \) and \( \theta_y(r) = 2\pi(r - 0.5) \).

The Ray-Based Segmentation is formulated in the form of minimizing an energy function, \( E_{total} \). Two energy terms are defined (1) an appearance-based energy \( E_{appearance} \), and (2) a shape-based energy \( E_{shape} \). The appearance (external) energy term is defined to catch image information, whereas the shape
(internal) energy term is derived to regulate the segmentation surface. The total energy is defined as follows,

$$E_{total}(l) = \lambda_{Ray} E_{appearance}(l) + E_{shape}(l),$$  \hspace{1cm} (3.20)

where,

$$E_{appearance} = \sum_{s} \sum_{r} CA(\vec{v}(s, r)), \hspace{1cm} (3.21)$$

and,

$$E_{shape} = \sum_{s} \sum_{r} \sum_{s=N_{Nib}}^{s+N_{Nib}} \sum_{r=r-N_{Nib}}^{r+N_{Nib}} CS(l(s, r), l(s_n, r_n)), \hspace{1cm} (3.22)$$

where $\lambda_{Ray}$ controls the influence of the image information in the segmentation result. $(2 \times N_{Nib} + 1) \times (2 \times N_{Nib} + 1)$ is the number of neighbor rays. The greater $N_{Nib}$ results in more smoothness of the segmented surface. $CS$ is the cost function of the shape term, and $CA$ is the cost function of the image information. These two cost functions are defined as follows,

$$CS(l(s, r), l(s_n, r_n)) = \left( \frac{r_{Ray}}{m_{Ray}} |l(s, r) - l(s_n, r_n)| \right)^{\gamma_{Ray}}, \hspace{1cm} (3.23)$$

and,

$$CA(l(n_{\theta_z}, n_{\theta_y})) = \exp\left\{ -V_{\nabla G_\sigma}(\vec{v}(s, r)) \right\}, \hspace{1cm} (3.24)$$

and,

$$V_{\nabla G_\sigma} = \nabla(V * G_\sigma), \hspace{1cm} (3.25)$$

where $\gamma_{Ray}$ is a parameter which controls the shape stiffness. $G_\sigma$ is a Gaussian filter with zero mean and standard deviation of $\sigma$. Also, $*$ is the convolution operator. The optimization problem is formulated
by minimizing the total energy,

$$
\hat{l} = \arg\min_l \{E_{total}(l)\}.
$$

To solve equation (3.26), three methods can be used: 1) gradient descent, 2) belief-propagation, and 3) graph-cut. As a drawback of this approach, labels should be initialized close to the actual boundaries [7].

### 3.2.1.1.6 Probabilistic approaches:

In order to incorporate prior shape knowledge into segmentation based on explicit representation, Cootes et al. [202] [203] introduced the active shape model (ASM) which builds a statistical model covering both typical shape and typical variability. This method is also known as the smart snake method. ASM was first used for 3D medical image analysis by Hill et al. [204]. In ASM, a set of manually landmark-ed training shapes are used to create a Point Distributed Model (PDM) [62], in which each point in the model represents a particular part of the shape, $x_i = (x_{i,1}, y_{i,1}, z_{i,1}, \cdots, x_{i,K}, y_{i,K}, z_{i,K})$, where $i \in \{1, \cdots, N\}$ corresponds to the $i^{th}$ training shape, and $K$ is the number of landmarks. The active shape model consists of three major steps:

1. aligning training shapes,
2. creating the point distributed model,
3. iteratively fitting the shape on the image data.

In order to find the statistics of landmark points, all the training shapes must be aligned. A shape alignment is modified using a rigid body deformation consisting of scale ($s$), rotation ($\vec{\theta}$), and translation ($\vec{t}$). Each shape is modified to minimize the weighted distance between each shape, $x_i$, and a mean shape, $x_{mean}$, as,

$$
E_j = (x_{mean} - M(s_j, \vec{\theta}_j)x_j - \vec{t}_j)^TW(x_{mean} - M(s_j, \vec{\theta}_j)x_j - \vec{t}_j),
$$

where $M(.,.)$ is a matrix of deformation with parameters $s_j$ and $\vec{\theta}_j$ as scale and rotations of $j^{th}$ shape, respectively. $\vec{t}_j$ is the translation vector of $j^{th}$ shape. $W$ is a matrix of weights which emphasizes landmark points with lower variations. At the first iteration, the mean shape is initiated with the first training shape. Then, all the training shapes are aligned on the mean shape. Afterward, the mean shape is calculated again. The two latest steps are repeated until a convergence in the mean shape is achieved.

The points distributed model is a cloud of possible shapes in a $3n \times N$ domain, and is called the allowable shapes domain. The allowable shapes domain has a cloud center, $\vec{x} = \frac{1}{N} \sum_{i=1}^{N} x_i$. For each shape, the difference vector is $dx_i = x_i - \vec{x}$. The covariance matrix is then calculated as,

$$
S = \frac{1}{N} \sum_{i=1}^{N} dx_i dx_i^T.
$$
The principal component analysis (PCA) [205] is applied to extract $k$ eigenvectors associated with largest eigenvalues, $P = [P_1, \cdots, P_k]$. Each shape is approximated with a linear combination of selected eigenvectors, $x_i = x + Pb$, where $b$ is a vector of coefficients. In order to fit the shape model on the image data, an external energy is calculated at each point, and is mapped into the set of adjustments as, $dX = (dX_1, dY_1, dZ_1, \cdots, dX_n, dY_n, dZ_n)^T$. The shape center adjustment, $dX_c = [dX_c, dY_c, dZ_c]$, is attained by,

$$dX_c = \frac{1}{K} \left[ \sum_{i=1}^{K} dX_i, \sum_{i=1}^{K} dY_i, \sum_{i=1}^{K} dZ_i \right],$$

(3.29)

which should be modified at each iteration. Then, the shape posture including orientation and scale are modified based on $dX$. Afterward, the residual adjustment, $dx$, is corrected by local movements of the shape points using the following equation,

$$dx = M((s(1 + ds))^{-1}, -(\theta + d\theta))y - x$$

(3.30)

where $y = M(s, \theta)x + dX - dX_c$. Finally, $dx$ is mapped into $db$ by $db = P^Tdx$. This process continues until a convergence is achieved.

After introducing the ASM method in 1992, many attempts have been performed to improve its performance for segmentation. The first attempt was made by Cootes et al. [206] which applied a multi-resolutional approach to reduce the computational cost, and to improve the segmentation accuracy and robustness against trapping in local minimum solutions. This method starts by fitting a shape model on the smoothed image in the coarsest resolution. The fitted model is then passed through finer resolutions until the best segmentation is achieved in the original image resolution. Therefore, the shape model is the same for different scales, and changes in levels of detailed shape information are not considered in different resolutions.

In the classical ASM, the number of samples in the training set is limited, while each sample has a high dimensionality, $N << 3 \times K$. This problem is known as high-dimension-low-sample-size (HDLSS). Therefore, its eigen-vectors are not capable to represent full range of shape variations. This problem is addressed by Davatzikos et al. [207]. Davatzikos et al. proposed a multi-resolutional approach based on wavelet transform in which the entire shape is hierarchically divided into a set of sub-bands. Each sub-band has less number of landmarks, $K' << K$, and therefore, the number of samples, $N$, is enough to represent shape variations of each sub-band. Considering a mother wavelet function, $\varphi(t)$, each wavelet function is defined as $\psi_{a,b} = \sqrt{2^a} \varphi(2^{-a}t - b)$ where $a$ and $b = 2^p$ are dilation and translation parameters. We define landmarks’ coordinates as $(u_{i,j}, v_{i,j}, w_{i,j})$ where $i = [1, \cdots, N]$ and $j = [1, \cdots, K]$. By
applying the wavelet transform on the landmarks, we obtain wavelet coefficients as, \( c_i = W(u_i) = [c_{i,j}] \), \( d_i = WT(v_i) = [d_{i,j}] \), and \( g_i = WT(w_i) = [g_{i,j}] \), where \( j = [1, \cdots, L] \), and \( WT(.) \) is the wavelet transform. All the coefficients of each shape are gathered in \( W_i \), and then, it is divided into \( B = 2^p \) sub-bands, \( W^{(b)}_i \) where \( b = [1, \cdots, B] \). Afterwards, their statistics are computed as \( \mu^{(b)} \) and \( C^{(b)} \) which are mean and covariance matrix, respectively. For each sub-band, the matrix of eigen-vectors, \( \Phi^{(b)} \), and matrix of eigenvalues, \( \Lambda^{(b)} \), are obtained using the PCA method. Eigen-vectors related to the first few bands represent relatively global shape appearance, whereas higher indices represent more localized shape variability. Upon the arrival of a new shape, \( a \), it is first aligned on the mean shape model by \( T(a) \). Then, its wavelet coefficients are divided into \( B \) bands. Each band is then approximated using \( \hat{s}^{(b)} = \mu^{(b)} + \Phi^{(b)}q^{(b)} \), where \( q^{(b)} \) is the parameter vector. Then, the inverse wavelet transform is applied to return the coefficients in the original domain. Its inverse transform, \( T^{-1}(W^1(\hat{s})) \), provides the input shape approximation which matches the shape model. Although this approach addresses the HDLSS problem, it is limited to single-object structures.

Cerrolaza et al. [208, 209] extended the hierarchical approach into a multi-object framework. This method still lacked supporting inter-object relationships as each shape piece was independently represented. This work was improved by Cerrolaza et al. [210] in which inter-object relationships were considered in the hierarchical shape model. Suppose the general shape consists of \( M \) single-objects. All samples at the resolution level, \( r \), can be represented as, \( x^r = (x^r_1, \cdots, x^r_M) \). The single-objects at the resolution level \( r \) has \( [k^r_1, \cdots, k^r_M] \) landmarks. At each resolution level, a specific relationship between single-objects are defined in which single-objects are divided into disjoint subsets, \( S_d \) where \( d = [1, \cdots, d_r] \), \( \bigcup_{d=1}^{d_r} S_d = M \) and \( \bigcap_{d=1}^{d_r} S_d = \emptyset \). At lower resolution levels, objects are more jointly defined to preserve inter-object relationships to impose general restrictions on the relative positions between single-objects. As the resolution increases, the number of disjoint subsets increases until each single-object has a particular disjoint subset. Therefore, in higher resolutions inter-object relationships are less considered, and the segmentation is more focused on capturing more shapes’ details.

The task of manual selection of 3D landmarks is very tedious and time-consuming. Some researches have been delivered to provide solutions for automated 3D landmarking [211] [8] [212]. Brett and Taylor [211] proposed an automated approach to define 3D PDM landmarks. In this method, 3D surfaces were represented by densely triangulated polyhedral surface. They used the Iterative Closest Point (ICP) algorithm to obtain correspondences among the surfaces in the training set. Afterward, a mean shape was generated as a surface represented by dense triangulations. Then, triangle decimation was applied to optimize the number of triangulations (sparse triangulations), and their nodes were selected as the final landmarks. As a drawback of this method, the resultant shape model was subject to folding due to
Figure 3.7: Displaying the automated landmarking process, proposed by Frangi et al. [8]. Based on using the natural coordinate, the principal modes of variation only account for nonrigid deformations, and pose or size differences are not considered in the statistical shape modeling.

the nature of ICP which was highly localized and did not consider connectivity constraints. This work was improved by Kaus et al. [213] in which external energy and internal energy terms were applied on a triangular mesh representation model in binarized volumes, to drive the mesh model toward a desired object boundaries while maintaining the mesh connectivity.

To avoid the problems associated with using ICP, Frangi et al. [8] proposed an automated 3D PDM based on multi-resolution free-form nonrigid registration, which defined a continuous deformation field using a set of B-Spline basis functions. The deformation contained global and local transformations. For global transformation, the quasi-affine transform based on 9 parameters was used. To capture details, a local transformation based on a B-spline free-form was used, in which spacing between nodes were hierarchically changed in a coarse-to-fine manner. To find correspondences between shapes in the training set, two similarity measures were used, including label consistency and $\kappa$ statistics. A shape model was initiated by a randomly selected shape from the training set. Then, all training shapes were registered on the shape model. Afterward, the shape model was updated with the average of all the registered shapes. The generated shape model was then used to automatically extract surface landmarks using the marching cubes algorithm [214]. In the marching cubes algorithm, the surface curvature was utilized to insert more landmarks around relevant edges and relatively less landmarks in flat areas.

The ASM method provides a framework to incorporate statistical knowledge of the organ’s shape in the segmentation process, which addresses many challenges associated with medical image segmentation tasks. Therefore, it has been widely used in 3D ultrasound segmentation [215–217]. However, it does not provide a connection between object appearance and the statistical shape model.

Cootes et al. [218] proposed the active appearance model (AAM) which generates a statistical appearance model by modeling texture variations. In this method, the allocated correspondences were used to align the training set images on a reference shape using the Procrustes analysis. The wrapping model
of each training image was then generated to match the appearances on a reference image to generate a shape-free patch, \( g_{\text{mean}} \). The modes of variations were generated using the eigen-analysis. The shape and appearance models were described as,

\[
\begin{align*}
    \mathbf{x} &= \mathbf{x}_{\text{mean}} + Q_s c \\
    \mathbf{g} &= \mathbf{g}_{\text{mean}} + Q_g c
\end{align*}
\]  

(3.31)

where \( c \) is a vector of controlling parameters, \( \mathbf{x}_{\text{mean}} \) and \( \mathbf{g}_{\text{mean}} \) are the mean shape and the mean texture in a mean shaped patch, respectively. \( Q_s \) and \( Q_g \) are matrices describing the modes of variation obtained from the training set [218]. This approach was first developed for face recognition application. Guo et al. [219] proposed a novel approach for segmenting ultrasound images by incorporating sparse representation in the AAM framework. In this method, sparse representation based on learnt-dictionaries [220] was used to locate the initial pose of the targeted object. The AAM method facilitates incorporating shape and texture prior knowledge of internal organs and structures of interest in the segmentation process, which provides robust segmentation for 3D ultrasound images. Therefore, it has been widely used for ultrasound image segmentation [221–224].

### 3.2.1.2 Discrete explicit representation

The idea of engaging discretized deformable surface was first developed by Terzopoulos et al. [166]. The idea was designed based on a 3D mesh of vertices which were bonded to each other by imaginary springs as,

\[
f_{\text{spring}}(m, n, k) = \kappa(\|\vec{X}_{m,n} - \vec{X}_{m,n,k}\| - r_0),
\]

(3.32)

where \( \kappa \) is the spring constant, \( r_0 \) is the natural length of the spring, \( \vec{X}_{m,n} \) is a vertex surrounded by \( \vec{X}_{m,n,k} \) neighboring vertices. \( m \in \{1, \cdots, M\} \) and \( n \in \{1, \cdots, N\} \) form a 2D parametric space to describe a 3D surface. The equilibrium of internal and external forces guides the deformable surface to smoothly extract an object of interest. Their proposed method obeys a Newtonian law of movement to actively deform the mesh surface aiming to track a deforming object in a 3D image. Afterward, many methods have been designed based on the discretized deformable surface model. In general, we categorize them based on the surface representation model. Meshes are represented by connected surface elements. Each surface element is formed with \( K \) vertices. Meshes with \( K = 3, K = 4, \) and \( K = 5 \) are known as triangulation, quadrilateral and simplex mesh, respectively.
3.2.1.2.1 Deformable meshes: After the emergence of discretized deformable models, Vasilescu and Terzopoulos [225] reorganized the discrete deformable surface into interconnected triangulations by adjustable springs. Their method offered hierarchical subdivision and merging of triangles to adaptively tune the surface density aiming to minimize the number of meshes around homogenous regions, and to precisely acquire high-curvature data structures. Soon after, this idea was carried on by Huang and Goldgof [226]. In their method, every three nodal points (vertices) created a planar patch, and neighboring vertices were connected to each other by imaginary springs. They used a physical model to move vertices to capture data structures as,

\[ m \frac{d^2 \vec{X}_{m,n}}{dt^2} + r \frac{d \vec{X}_{m,n}}{dt} + f^{ext}_{m,n} = f^{int}_{m,n}, \]  

(3.33)

and,

\[ f^{int}_{m,n} = \sum_k c_k \left( \frac{\| \vec{X}_{m,n} - \vec{X}_{m,n,k} \| - r_0}{\| \vec{X}_{m,n} - \vec{X}_{m,n,k} \|} \right) \left( \vec{X}_{m,n} - \vec{X}_{m,n,k} \right), \]  

(3.34)

where \( f^{ext}_{m,n} \) is the external force created by a spring between \( \vec{X}_{m,n} \) and its adjacent data point, and \( f^{int}_{m,n} \) is the internal force by springs between \( \vec{X}_{m,n} \) and its neighboring vertices, \( \vec{X}_{m,n,k} \). \( c_k \) is the stiffness coefficient. The triangular mesh model allowed their method to adaptively add and remove vertices while preserving the overall shape structure [226]. If the distance between two neighboring vertices was getting too large (\( \| \vec{X}_{m,n} - \vec{X}_{m,n,k} \| > t_{high} \)) or too small (\( \| \vec{X}_{m,n} - \vec{X}_{m,n,k} \| < t_{low} \)), then a new vertex was added, or an existing vertex was removed from the mesh.

Delingette [227] proposed the simplex meshes, and then provided more details on its medical applications in [9]. In simplex mesh, each vertex is connected to 3 neighboring vertices, \( \{ \vec{X}_{m,n,0}, \vec{X}_{m,n,1}, \vec{X}_{m,n,2} \} \) (Fig. 3.8a). The simplex angle, \( \varphi_{m,n} = \angle(\vec{X}_{m,n}, \vec{X}_{m,n,0}, \vec{X}_{m,n,1}, \vec{X}_{m,n,2}) \) is then defined as follows,

\[ \sin(\varphi_{m,n}) = \frac{r_{m,n}}{R_{m,n}} \text{sign}(X_{m,n} \cdot \vec{N}_{m,n}), \]  

(3.35)

where \( R_{m,n} \) is the radius of the sphere encompassing four vertices, \( \{ \vec{X}_{m,n}, \vec{X}_{m,n,0}, \vec{X}_{m,n,1}, \vec{X}_{m,n,2} \} \), and \( r_{m,n} \) is the radius of the circle encompassing three vertices, \( \{ \vec{X}_{m,n,0}, \vec{X}_{m,n,1}, \vec{X}_{m,n,2} \} \) which is centered at \( C_{m,n} \). \( \vec{N}_{m,n} \) is a normal vector to the plane of the three neighboring vertices (Fig. 3.8b). The projection of \( \vec{X}_{m,n} \) on the plane of the three neighboring vertices, \( \vec{F}_{m,n} \), is represented by three parameters \( \{ \epsilon_{m,n,0}, \epsilon_{m,n,1}, \epsilon_{m,n,2} \} \) as,

\[ \vec{F}_{m,n} = \epsilon_{m,n,0} \vec{X}_{m,n,0} + \epsilon_{m,n,1} \vec{X}_{m,n,1} + \epsilon_{m,n,2} \vec{X}_{m,n,2}, \]  

\( \{ \epsilon_{m,n,0} + \epsilon_{m,n,1} + \epsilon_{m,n,2} = 1 \} \).  

(3.36)
\{\epsilon_{m,0}, \epsilon_{m,1}\} \text{ are called the metric parameters. Based on equation (3.36), each vertex can be represented by its metric parameters and its simplex angle as follows,}

\[
\begin{align*}
\vec{X}_{m,n} &= \vec{F}_{m,n} + L(r_{m,n}, d_{m,n}, \varphi_{m,n}) \vec{N}_{m,n}, \\
L(r_{m,n}, R_{m,n}, \varphi_{m,n}) &= \frac{(r_{m,n}^2 + D_{m,n}^2) \tan(\varphi_{m,n})}{\sqrt{r_{m,n}^2 + (r_{m,n}^2 + D_{m,n}^2) \tan^2(\varphi_{m,n}) + r_{m,n}}}, \\
d_{m,n} &= \| C_{m,n} - F_{m,n} \|,
\end{align*}
\]

(3.37)

with

\[
\epsilon = 1 \text{ if } |\varphi_{m,n}| < \pi/2,
\]

and \( \epsilon = -1 \) if \( |\varphi_{m,n}| > \pi/2. \)

The entire mesh surface is completely represented by three parameters \( \{\epsilon_{m,0}, \epsilon_{m,1}, \varphi_{m,n}\} \), up to a scale and isometry. The simplex mesh deformation is controlled by equation (3.37). The internal force is derived to maintain the surface continuity, and consists of two terms, \( f_{n,m}^{\text{int}} = f_{n,m}^{Tangent} + f_{n,m}^{\text{Normal}} \). The goal of the tangent term, \( f_{n,m}^{Tangent} \), is to control the position of each vertex based on its neighboring vertices as,

\[
\begin{align*}
    f_{n,m}^{Tangent} &= (\epsilon_{m,n,0} - \frac{1}{3}) \vec{X}_{m,n,0} + (\epsilon_{m,n,1} - \frac{1}{3}) \vec{X}_{m,n,1} \\
    &+ (\epsilon_{m,n,2} - \frac{1}{3}) \vec{X}_{m,n,2},
\end{align*}
\]

(3.38)

where \( \epsilon_{m,n,k} = \frac{1}{3} \) and \( k \in \{0, 1, 2\} \) refers to the relaxed position. The normal term, \( f_{n,m}^{\text{Normal}} \), controls the mean curvature of the simplex mesh, and is achieved as follows,

\[
    f_{n,m}^{\text{Normal}} = (L(r_{m,n}, d_{m,n}, \varphi_{m,n}) - L(r_{m,n}, d_{m,n}, \tilde{\varphi}_{m,n})) \vec{N}_{m,n}
\]

(3.39)

where \( \tilde{\varphi}_{m,n} \) is the reference simplex angle. One possible option for \( \tilde{\varphi}_{m,n} \) is the simplex angle continuity which equals the average of the neighboring vertices’ simplex angles as,

\[
\tilde{\varphi}_{m,n} = \sum_{m',n' \in \{d_{m',n'} < R_{vac}\}} \varphi_{m',n'},
\]

(3.40)

where \( d_{m',n'} = \| \vec{X}_{m,n} - \vec{X}_{m',n'} \| \), and \( R_{vac} \) is the radius of a spherical region centered at each vertex. Increasing \( R_{vac} \) results in a smoother mesh shape with less surface curvatures. The external force, \( f_{n,m}^{\text{ext}} \), is derived to push vertices to their closest data points, and is defined as follows,

\[
    f_{n,m}^{\text{ext}} = G \left( \frac{(M_{Cl}(m,n) - \vec{X}_{m,n}) \cdot \vec{N}_{m,n}}{D_{ref}} \right) \vec{N}_{m,n},
\]

(3.41)
where $D_{ref}$ is the cut-off distance, $G(.)$ is the derivative operator, and $M_{Cl}(m,n)$ is the closest data point to $X_{m,n}$. The advantage of simplex mesh over triangulations is the parametric framework which is provided by the simplex mesh to represent the shape surface, and results in more controllability of the surface curvature and smoothness. As a disadvantage compared to triangulations, it is more sophisticated to add or to remove vertices to control the surface density through homogenous to high curvature data points.

In order make the simplex mesh to be more suited for medical image segmentation, Montagnat and Delingette [228] applied a shape constraint on the simplex meshes, and used global deformations based on a similarity transformation into the motion equation. The shape constraint (prior shape model) was defined as a set of reference simplex angles, $\{\varphi_0^{m,n}\}$. Equation (3.41) can be modified to follow the prior shape as,

$$\tilde{\varphi}_{m,n} = \varphi_0^{m,n} + \sum_{m',n' \in \{d_{m',n'} < R_{vac}\}} \lambda_{m',n'}(\varphi_0^{m,n} - \varphi_{m',n'}),$$

(3.42)

where $\sum \lambda_{m',n'} = Const < 1$ maintains the convergence of the mesh surface. The use of deformable mesh in ultrasound image segmentation has been reported in some works [229–231].

### 3.2.1.2.2 T-surface:

McAnerney and Terzopoulos [232] proposed a new derivation of discretized deformable models, called T-Surface. T-Surface consisted of three components:

1. a parametric deformable surface explicitly defined by nodes,
2. 3D affine cell image decomposition (3D ACID) formed with triangular mesh elements,
3. a re-parametrization process which provides topological adaptivity.

A 3D grid of uniform cubes were generated, and each cube was subdivided into six tetrahedral cells. The deformable surface was formed by a set of nodes, which were located on the edges of grid cells at
Chapter 3. Automated kidney segmentation

Figure 3.9: The solid and dashed contours show T-surface before and after deformations, respectively. Yellow nodes show aligned nodes on cells edges. Red and white points show 'on' vertices and newly added vertices, respectively. (a) shows the deformation process. (b) second phase of re-parametrization. (c) first phase of re-parametrization and boundary cells with gray color [10].

The beginning of each deformation step. The grid vertices were specified to be 'on' if they were located inside the deformable surface, and 'off' if they were outside the T-Surface. Boundary cells were defined as those cells with both 'on' and 'off' cells. The surface nodes deformed under the Laplacian framework of the 3D snake, in which internal and external forces interactively deformed the T-Surface to move the T-Surface closer to the object of interest. As the surface evolved, the nodes freely moved beyond edges of the boundary cells (Fig 3.9a). Every $M$ steps, a re-parametrization was derived to reallocate T-Surface nodes. The re-parametrization process was performed in two phases: (1) recalculating correspondences between the T-Surface and the grid cells, and selecting new nodes locating on edges of boundary cells (Fig 3.9c); (2) updating status of grid vertices by turning them 'on' or 'off' based on their new alignment inside or outside the T-Surface (Fig 3.9b). The re-parametrization process facilitated topological flexibility by merging collided T-Surfaces or by splitting a T-Surface into two individual T-Surfaces. Despite the topological flexibility that T-Surface provided, only few works reported the use of T-surface in 3D ultrasound image segmentation [233]. This might be because of the sensitivity of re-parametrization process to speckle interference in ultrasound volumes.

3.2.1.2.3 Spring level-set: Spring Level-set (SpringLS) [11, 234, 235] is a new generation of the discrete deformable surfaces, which makes a bridge between implicit representation (level-set) and deformable meshes. The implicit representation provides topologically flexible, which is free of re-parametrization. On the other side, deformable meshes are suitable for model registration. SpringLS develops a framework to take the advantages of both deformable meshes and implicit representation models. In SpringLS, the surface is represented by disconnected triangular surface elements (springls), which each springl is centered at a particle, $\vec{P}_{n,m}$, and has three vertices, $\vec{X}_{n,m,k}$ where $k \in \{0, 1, 2\}$. Each particle is linked to a correspondence point, $\vec{a}_{n,m}$, as a prior shape model, which is responsible to bound the shape deformation to an allowance shape domain. The SpringLS surface is implicitly tracked
by a level-set function based on calculating a clamped distance function for each spring. By associating
the level-set to springs, three advantages are provided:

- the sign of the level-set function specifies regions inside and outside the segmentation model;
- the signed level-set indicates when new springs are required to be added or removed;
- the iso-surface obtained from the zero level-set is a watertight surface representation model.

The particles and vertices are evolved under an energy minimization problem which considers both rigid
and fluid characteristics of the segmentation target. The evolutionary equations can be split into two
terms: advection and relaxation. In the advection step, particles and vertices are perturbed to move
toward desired data points as,

\[
\frac{\partial \vec{P}_{n,m}(t)}{\partial t} = \lambda_p \vec{N}_{m,n} G(\vec{P}_{n,m}(t)) + \lambda_\sigma \vec{\sigma}(\vec{P}_{n,m}(t)) + \lambda_A A(\vec{P}_{n,m}(t), \vec{a}_{n,m}) \vec{P}_{n,m}(t),
\]

\[
\frac{\partial \vec{X}_{n,m,k}(t)}{\partial t} = \lambda_p \vec{N}_{m,n} G(\vec{P}_{n,m}(t)) + \lambda_\sigma \vec{\sigma}(\vec{P}_{n,m}(t)) + \lambda_A A(\vec{P}_{n,m}(t), \vec{a}_{n,m}) \vec{X}_{n,m,k}(t),
\]

where \(G(.)\), which is obtained by the gradient vector flow (GVF) [177], is a pressure in the normal
direction at each spring. \(\vec{\sigma}(.)\) is an external velocity field. \(G(.)\) and \(\vec{\sigma}(.)\) both control the fluid
deformation. \(A(\vec{P}_{n,m}(t), \vec{a}_{n,m})\) is a local affine transformation and is used to model rigid deformation of
the SpringLS. After the advection step, the relaxation step is applied on vertices based on spring-mass
equations, while particles are kept fixed. Each vertex is re-coordinated using,

\[
\vec{X}_{n,m,k} = l_{n,m,k} M(\vec{b}_{n,m}) \vec{s}_{n,m} + \vec{P}_{n,m},
\]

\[
\vec{s}_{n,m} = \frac{1}{l_{n,m}} M(-\vec{b}_{n,m}) \left( \vec{X}_{n,m,k} - \vec{P}_{n,m} \right),
\]

where \(\|\vec{P}_{n,m} - \vec{X}_{n,m,k}\|\) is the spring length, and \(\vec{s}_{n,m}\) is the tangent vector. \(\vec{b}_{n,m} \in \mathbb{R}^3\) is the vector of
axis-angles. An internal force is derived to attract vertices toward its particle based on the spring force,

\[
\vec{f}_{n,m,k}^{int} = \kappa_{int} (l_{n,m,k} - 2 \ast r_0) M(\vec{b}_{n,m}) \vec{s}_{n,m},
\]

where \(\kappa_{int}\) is the internal spring constant. An external force is derived to bind each vertex to its
neighboring vertices as,

\[
\vec{f}_{n,m,k}^{ext} = \kappa_{ext} \sum_{[n''m''k''] \in \{d_{n'',m'',k'}^{n,m,k} < R_{vac}\}} \left( \vec{X}_{n,m,k} - \vec{X}_{n'',m'',k'} \right),
\]
where $\kappa_{ext}$ is the external spring constant, and $R_{vac}$ is the radius of a vicinity in which neighboring vertices of a vertex are placed (Fig 3.10). $d_{m,n,k}^{n,m,k'}$ is the Euclidean distance between $\vec{X}_{n,m,k}$ and $\vec{X}_{n',m',k'}$. By applying a gradient descent approximation, the relaxation minimization problem is solved by iteratively updating the following parameters,

$$
\begin{align*}
\rho^{iter+1}_{n,m,k} &= \rho^{iter}_{n,m,k} - \lambda \left( M \left( b_{n,m} \right) s_{n,m} \right) \left( f^{ext}_{n,m,k} + f^{int}_{n,m,k} \right), \\
\end{align*}
$$

(3.47)

and,

$$
\begin{align*}
\rho^{iter+1}_{n,m,k} &= \rho^{iter}_{n,m,k} - \lambda \sum_{m} \rho^{iter}_{n,m,k} \cdot \rho^{iter}_{n,m,k} \cdot f^{ext}_{n,m,k}.
\end{align*}
$$

(3.48)

Figure 3.10: Displaying a particle, vertices and springs (gray) [11].

The signed level-set $\phi(\vec{X})$ for each voxel in the space domain, $\vec{X} \in \mathbb{R}^3$ is calculated as,

$$
\begin{align*}
\phi^{itr+1}(\vec{X}) &= \phi^{itr}(\vec{X}) - \Delta \kappa \cdot \phi^{itr}(\vec{X}) \left( \omega(\vec{X}) \cdots \nabla \omega(\vec{X}) \cdot f^{ext}(\vec{X}) \right). \\
\end{align*}
$$

(3.49)

where $\omega(\vec{X})$ is an unsigned level-set, and is calculated as $\omega(\vec{X}) = \arg\min_{d_{max}} d_{1,1}(\vec{X}, \cdots, d_{N,M}(\vec{X})$. $d_{m,n}(\vec{X})$ is the clamped unsigned distance of $\vec{X}$ to $\vec{P}_{m,n}$. The calculated level-set is used to control the evolution of particles under the advection step. If $\left( \phi(\rho^{iter}_{n,m}) - \phi(\rho^{iter-1}_{n,m}) \right) \cdot \nabla \phi(\rho^{iter-1}_{n,m}) \cdot \vec{N}_{m,n} \geq 0$, the particle and vertices positions are modified by equation (3.43).

3.2.1.3 Implicit representation

3.2.1.3.1 Edge-based level-set: The level-set method was first applied in [170, 236] to extract regions of interest in 2D images. Then, it was extended to 3D segmentation by Caselles [181]. The level set function, $\phi(\vec{X}, t)$, assigns a level to each voxel, $\vec{X}$, where $\vec{X} \in \Omega_V$, and $\Omega_V$ is the 3D image domain. $t$ is the time parameters, and in the level-set evolution, $t$ corresponds to the iteration number. Also, $\phi(\vec{X}, t = 0)$ specifies the initial value of the level-set function at each voxel, $\vec{X} \in \Omega_V$. The level-set of voxels placed on the segmentation surface receive zero, the level-set of voxels placed inside...
the segmentation surface receive some positive values, and the level-set of voxels placed outside the segmentation surface receive some negative values.

In the primitive formulation of the level-set method and its evolution, \( \phi(\vec{X}, t) \) was iteratively evolved along its normal vector fields by \( F(\vec{X}) \), which controlled the speed of changing the level-set function at each time-point, and stopped changing at edges. The speed function, \( F(\vec{X}) \), consisted of two terms, \( F(\vec{X}) = F_A(\vec{X}) + F_G(\vec{X}) \), where \( F_A(\vec{X}) \) was similar to the balloon force, and \( F_G(\vec{X}) \) maintained the smoothness of the segmented region. The \( F_A(\vec{X}) \) itself consisted of two terms, including \( \vec{V}(\vec{X}) \) and \( \vec{S}(\vec{X}) \), where \( \vec{V}(\vec{X}) \) was the normal speed vector pointing toward the zero level-set surface at each voxel, \( \vec{X} \in \Omega_V \), and \( \vec{S}(\vec{X}) \) was another speed vector pointing toward edges in the image. Here, we briefly introduce the underlying formulation of primitive level-set method.

For the sake of simplicity, assume the speed function is only formed by \( F_A \). The level-set evolution equation at iteration, \( t \), is written as follows,

\[
\phi_t(\vec{X}, t) + \vec{V}(P) \cdot \nabla \phi(\vec{X}, t) + \vec{S}(\vec{X}) \cdot \nabla \phi(\vec{X}, t) = 0
\]

where \( \nabla \) is the gradient operator, and \( \phi_t(\vec{X}, t) \) is the derivative of the level-set function with respect to \( t \) at the location \( \vec{X} \in \Omega_V \). In equation (3.50), the normal term of the first speed term is only required, and therefore, equation (3.50) can be simplified as follows,

\[
\phi_t(\vec{X}, t) + V(\vec{X})|\nabla \phi(\vec{X}, t)| + \vec{S}(\vec{X}) \cdot \nabla \phi(\vec{X}, t) = 0
\]

The speed function is now calculated based on the intensity level at each voxel, aiming to reduce the level-set evolution speed at locations corresponding to objects with higher intensity levels as,

\[
V(\vec{X}) = \exp(-K_2 \left| \nabla \left( \{ G_\sigma * I \} (\vec{X}) \right) \right|),
\]

where \( G_\sigma \) is the Gaussian function with zero-mean and standard deviation of \( \sigma \), and \( \exp(.) \) is the exponential function. This definition is used to attain a smooth reduction in the speed of level-set evolution at voxels with the zero level-set, also called “front”, approaching bright regions (presumably corresponding to the objects of interest). The term \( S(\vec{X}) \) was defined as,

\[
S(\vec{X}) = K_1 \nabla \left| \nabla \left( \{ G_\sigma * I \} (\vec{X}) \right) \right|.
\]

\( V(\vec{X}) \) and \( S(\vec{X}) \) can be interpreted as follows: homogeneous regions receive high values of \( V(\vec{X}) \), leading
to an evenly growing of the front, while their $S(\vec{X})$s are small. When the front approaches an object’s boundary, the $V(\vec{X})$ decreases and $S(\vec{X})$ increases, pushing the front toward the objects of interest. The numerical implementation was described in [237].

The value of level-set function was required to be calculated for each single voxel at each iteration, resulting in an excessive computational cost for 3D segmentation. The solution of this problem was proposed by Sethian [238], by applying a fast marching approach in the level-set framework. The idea was to sweep the front ahead in an upwind fashion by limiting the level-set update to the voxels in a narrow band around the front, and to march this narrow band forward, while freezing the rest of voxels. This method became soon the main algorithm of implementing level-set segmentation methods. The edge-based level-set was very popular in late 19th, and many papers have been published to address the segmentation task in ultrasound imagery. Some examples of using edge-based level-set in 3D ultrasound segmentation are [12,239–242].

3.2.1.3.2 Level-set without-edge (regional level-set): Chan and Vese [243] proposed a modified framework for image segmentation, based on the conventional level-set, that provided robustness against noise and low-contrast edges. Their method was designed based on the Mumford-Shah functional [244], which applied regional information to derive the stopping term for segmenting regions. Thus, their method is also well-known as the regional level-set method. Here, we provide a brief explanation on the representation model and level-set evolution of the regional level-set.

Assume a segmentation problem of a 3D image, $V : \Omega_V \mapsto \mathbb{R}$, where $\Omega_V$ is the 3D image domain, and a level-set function is assigned to $V$, such that voxels inside, outside, and on the segmentation surface are specified based on the levels of $\phi(\vec{X})$, as follows,

\[
\begin{align*}
\phi(\vec{X}) &> 0, \quad \text{where } \vec{X} \in \Omega_V \text{ is inside the segmentation surface}, \\
\phi(\vec{X}) &< 0, \quad \text{where } \vec{X} \in \Omega_V \text{ is outside the segmentation surface}, \\
\phi(\vec{X}) &= 0, \quad \text{where } \vec{X} \in \Omega_V \text{ is on the segmentation surface}.
\end{align*}
\]

(3.54)

In order to involve regional information of the voxels into the Mumford-Shah framework, two important functions are defined as follows [243],

\[
H(\phi) = \begin{cases} 
1, & \text{if } \phi \geq 0, \\
0, & \text{if } \phi < 0,
\end{cases} \quad \delta(\phi) = \frac{d}{d\phi}H(\phi),
\]

(3.55)

where $H$ is called the Heaviside function, and $\delta$ is the delta Dirac function. By adopting, $H$, and $\delta$, we
can define the length and area of the segmented region, using the following equations,

\[
\text{Length}\{\phi = 0\} = \iiint_{\vec{X} \in \Omega_{\nu}} \nabla H\left(\phi(\vec{X})\right) dxdydz = \iiint_{\vec{X} \in \Omega_{\nu}} \delta\left(\phi(\vec{X})\right) \left|\nabla \phi(\vec{X})\right| dxdydz, \quad (3.56)
\]

and,

\[
\text{Area}\{\phi \geq 0\} = \iiint_{\vec{X} \in \Omega_{\nu}} H\left(\phi(\vec{X})\right) dxdydz. \quad (3.57)
\]

Assume the mean intensity levels of voxels inside and outside the segmentation surface are \(c_1\) and \(c_2\), respectively. The variation of intensity levels of voxels inside and outside the segmentation regions can be written as follows,

\[
\text{Var}_{\text{in}} = \iiint_{\{\vec{X} | \phi(\vec{X}) > 0, \forall \vec{X} \in \Omega_{\nu}\}} \left|V(\vec{X}) - c_1\right|^2 dxdydz = \iiint_{\vec{X} \in \Omega_{\nu}} \left|V(\vec{X}) - c_1\right|^2 H\left(\phi(\vec{X})\right) dxdydz, \quad (3.58)
\]

and,

\[
\text{Var}_{\text{out}} = \iiint_{\{\vec{X} | \phi(\vec{X}) < 0, \forall \vec{X} \in \Omega_{\nu}\}} \left|V(\vec{X}) - c_1\right|^2 dxdydz = \iiint_{\vec{X} \in \Omega_{\nu}} \left|V(\vec{X}) - c_1\right|^2 \left(1 - H\left(\phi(\vec{X})\right)\right) dxdydz. \quad (3.59)
\]

Then, the energy functional \(F(c_1, c_2, \phi)\) can be written as,

\[
F(c_1, c_2, \phi) = \mu \cdot \text{Length}\{\phi = 0\} + \nu \cdot \text{Area}\{\phi \geq 0\} + \lambda_1 \cdot \text{Var}_{\text{in}} + \lambda_2 \cdot \text{Var}_{\text{out}}, \quad (3.60)
\]

where \(\mu, \nu, \lambda_1,\) and \(\lambda_2\) are Lagrangian multipliers, which control the effects of the four terms in the energy functional, governing the level-set evolution. In equation (3.60), the first term controls the smoothness of the front (see Sub-section 3.2.1.3.1), the second term is a constant speed, the third and four terms are external energies, pushing the segmented region toward the desired objects. \(c_1\) and \(c_2\) are defined as follows,

\[
\left\{
\begin{array}{l}
  c_1(\phi) = \text{average}\left\{V(\vec{X})|\phi(\vec{X}) \geq 0\right\}, \\
  c_2(\phi) = \text{average}\left\{V(\vec{X})|\phi(\vec{X}) < 0\right\}.
\end{array}
\right. \quad (3.61)
\]

Thus, \(c_1\) and \(c_2\) are calculated as follows,

\[
c_1(\phi) = \frac{\iiint_{\vec{X} \in \Omega_{\nu}} V(\vec{X}) H\left(\phi(\vec{X})\right) dxdydz}{\iiint_{\vec{X} \in \Omega_{\nu}} H\left(\phi(\vec{X})\right) dxdydz}, \quad (3.62)
\]
Figure 3.11: This figure represents the iterative process of image segmentation with the regional level-set method.

and,

$$c_2(\phi) = \frac{\iiint_{\vec{X} \in \Omega} V(\vec{X}) \left(1 - H(\phi(\vec{X}))\right) dxdydz}{\iiint_{\vec{X} \in \Omega} \left(1 - H(\phi(\vec{X}))\right) dxdydz}. \quad (3.63)$$

The Euler-Lagrange equation is used to solve the energy functional of equation (3.60), and the following PDE equation is obtained:

$$\frac{\partial \phi}{\partial t}(\vec{X}) = \delta(\phi(\vec{X})) \cdot \left[ \mu \cdot \text{div} \left( \frac{\nabla \phi(\vec{X})}{|\nabla \phi(\vec{X})|} \right) - \nu - \lambda_1 \cdot \left(V(\vec{X}) - c_1\right)^2 - \lambda_2 \cdot \left(V(\vec{X}) - c_2\right)^2 \right], \quad (3.64)$$

where $t$ is added to the level-set function, denoting the iteration number, for the use in an iterative solution of equation (3.64). The block diagram of the iterative solution of the regional level-set evolution is shown in Fig. 3.11. The active contour without-edge (regional level-set or region-based level-set) approach provides a higher robustness against noise and low-contrast edges, compared to the edge-based level-set. Therefore, after the introduction of regional level-set, it has attained a great interest among researchers with the medical society for segmenting organs and structures of interest in 3D ultrasound images [245–247].

### 3.2.1.3.3 Regional level-set with shape-prior:

In medical applications, applying prior shape knowledge is desired, because it provides more robustness against challenges of medical image segmentation. Chan and Zhu [248] proposed an extension to the Chan and Vese framework [243], by adding shape prior constraint in the regional level-set functional. In the Chan and Zhu method [248], another energy term is defined as the shape energy as,

$$F_{shape}(\phi, \psi) = \iiint_{\vec{X} \in \Omega} \left(H(\phi(\vec{X})) - H(\psi(\vec{X}))\right)^2 dxdydz, \quad (3.65)$$
where $\psi$ is a level-set function, based on a signed distance function, representing a prior shape. Also, consider $\psi(\vec{X})$ is connected to a default shape of interest, $\psi_0$, using a geometrical transformation, such as a similarity transformation, $ST_{\vec{p}}$ (see Sec 2.2.2 for a complete definition of the similarity transformation). Then, the relation between $\psi$ an $\psi_0$ is written as follows,

$$\psi(\vec{X}) = \psi_0(ST_{\vec{p}}\{\vec{X}\}).$$  \hspace{1cm} (3.66)

Considering equation (3.66), the shape energy can be written as follows,

$$F_{\text{shape}}(\phi, \psi, \vec{p}) = \iint_{\vec{X} \in \Omega_v} (H(\phi(\vec{X})) - H(\psi_0(ST_{\vec{p}}\{\vec{X}\})))^2 dxdydz,$$ \hspace{1cm} (3.67)

By adding the shape energy to the regional level-set functional, the following energy formula is obtained,

$$F(c_1, c_2, \phi, \psi) = \iint_{\vec{X} \in \Omega_v} (H(\phi(\vec{X})) - H(\psi_0(ST_{\vec{p}}\{\vec{X}\})))^2 dxdydz$$

$$+ \mu \iint_{\vec{X} \in \Omega_v} \delta(\phi(\vec{X})) \left| \nabla \phi(\vec{X}) \right| dxdydz + \nu \iint_{\vec{X} \in \Omega_v} H(\phi(\vec{X})) dxdydz$$

$$+ \lambda_1 \iint_{\vec{X} \in \Omega_v} \left| V(\vec{X}) - c_1 \right|^2 H(\phi(\vec{X})) dxdydz$$

$$+ \lambda_2 \iint_{\vec{X} \in \Omega_v} \left| V(\vec{X}) - c_1 \right|^2 (1 - H(\phi(\vec{X}))) dxdydz.$$ \hspace{1cm} (3.68)

The Euler-Lagrange equation is used to solve the energy functional of equation (3.68), and the following PDE equation is obtained:

$$\frac{\partial \phi}{\partial t}(\vec{X}) = \gamma \cdot \delta(\phi(\vec{X})) \cdot \left( H(\phi(\vec{X})) - H(\psi_0(ST_{\vec{p}}\{\vec{X}\})) \right)$$

$$+ \delta(\phi(\vec{X})) \cdot \left[ \mu \cdot \text{div} \left( \frac{\nabla \phi(\vec{X})}{\left| \nabla \phi(\vec{X}) \right|} \right) - \nu - \lambda_1 \cdot \left| V(\vec{X}) - c_1 \right|^2 - \lambda_2 \cdot \left| V(\vec{X}) - c_2 \right|^2 \right].$$ \hspace{1cm} (3.69)

where $\gamma$ is a Lagrangian multiplier, controlling the effect of the shape prior constraint on the regional level-set evolution. $t$ is added to the level-set function, denoting the iteration number, for the use in an iterative solution of equation (3.69). In each iteration of the level-set evolution process, three steps should be taken as follows,

1. updating the level-set function, $\phi$,

2. updating the average intensity levels, $c_1$ and $c_2$,

3. updating the similarity transformation parameters, $\vec{p}$, to ensure the transformed shape-prior, $\psi$, follows the segmented region, $\phi$. 


Figure 3.12: This figure represents the iterative process of image segmentation with the regional level-set method.

This procedure connects the segmentation region to the transformed shape prior, and ensures the segmented region stays similar to the shape prior, as it goes through the iterative evolution. The block diagram of the iterative solution of the regional level-set with shape-prior evolution is shown in Fig. 3.12.

The level-set method with shape-prior provides a well-suited framework for segmenting internal structures and organs with known shape information. Thus, this method has been widely used in ultrasound image segmentation [249–251].

### 3.2.1.3.4 Parametric level-Set:

Bernard et al. [252] proposed a parametric representation of the level-set function based on a linear combination of B-Spline basis functions. The parametric level-set method used B-spline basis functions to interpolate the level-set function in spatial domain. This provides a discrete representation of the level-set function, and simplifies the energy minimization problem into discrete convolutions. Since B-spline is separable, the energy minimization turns into three 1D convolutions which are mathematically efficient, compared to the conventional level-set. Because of using the B-spline representation of the level-set function, the segmentation result is expected to be smoother than the conventional level-set method. Also, the use of parametric representation to form the level-set function provides a more robust platform to segment medical images in the presence of noise and artifacts. The B-spline level-set function is defined as follows [252],

$$\phi(x) = \sum_{k^3} c[k] \beta(\frac{x}{h} - k),$$  \hspace{1cm} (3.70)

where $\beta(.)$ is the 3-D uniform symmetric quadratic B-spline. The knots are defined on a grid with a regular spacing by $h$. The coefficients of B-spline representation are gathered in $c[k]$. This function is separable and is written in the following form,

$$\beta(x) = \beta(x) \times \beta(y) \times \beta(z).$$  \hspace{1cm} (3.71)
The general expression of the energy functional for region-based level-set can be formulated as follows,

\[ F(\phi) = \lambda_1 \int_{\Omega} g_{\text{in}}(\vec{X}, \phi(\vec{X}))H(\phi(\vec{X}))dxdydz + \lambda_2 \int_{\Omega} g_{\text{out}}(\vec{X}, \phi(\vec{X}))(1 - H(\phi(\vec{X})))dxdydz + \lambda_3 \int_{\Omega} g_c(\vec{X}, \phi(\vec{X}))\|\nabla \phi(\vec{X})\| dxdydz. \]

In equation (3.72), the first, second, and third terms are energy terms related to the inside, outside, and on the segmentation surface, respectively. Also, \( g_{\text{in}}(\cdot) \), \( g_{\text{out}}(\cdot) \), and \( g_c(\cdot) \) are functions of intensity levels of voxels in object, background, and over the contour in the image, \( f \). \( \lambda_1, \lambda_2 \) and \( \lambda_3 \) are regulation parameters, controlling the influence of each energy term in the surface evolution. By combining (3.70)-(3.72), the following minimization problem is obtained,

\[ \frac{\partial F}{\partial c[k_0]} = \int_{\Omega} w(\vec{x})\beta(x - k_0) \times \beta(y - k_0) \times \beta(z - k_0)dxdydz. \]  

(3.72)

where,

\[ w(\vec{x}) = \lambda_1 \left( \frac{\partial g_{\text{in}}(\vec{x}, \phi(\vec{x}))}{\partial \phi(\vec{x})}H(\phi(\vec{x})) + g_{\text{in}}(\vec{x}, \phi(\vec{x}))\delta(\phi(\vec{x})) \right) + \lambda_2 \left( \frac{\partial g_{\text{out}}(\vec{x}, \phi(\vec{x}))}{\partial \phi(\vec{x})}(1 - H(\phi(\vec{x}))) - g_{\text{out}}(\vec{x}, \phi(\vec{x}))\delta(\phi(\vec{x})) \right) + \lambda_3 \left( \frac{\partial g_c(\vec{x}, \phi(\vec{x}))}{\partial \phi(\vec{x})}\|\nabla \phi(\vec{x})\| - \text{div}\left(g_c(\vec{x}, \phi(\vec{x}))(\nabla \phi(\vec{x}))\|\nabla \phi(\vec{x})\|\right) \right). \]  

(3.73)

For more information on the discretization of equation (3.73), readers are referred to [252, 253].

### 3.2.1.3.5 Active Surface:

The level-set approach adds an extra dimension to the segmentation problem. For a 3D surface, its level-set function belongs to the four-dimensional (4D) space. This increases the computational cost of segmentation. The active surface method [172] provides a 3D segmentation approach with lower computational complexity, using 2D implicit representation of the 3D segmentation surface. In the 3D space, a surface function can be mathematically represented as,

\[ x = g(y, z), \]  

(3.74)

where \( g : \mathbb{R}^2 \mapsto \mathbb{R} \). The active surface representation function, \( f \), is defined as,

\[ f(x, y, z) = x - g(y, z). \]  

(3.75)

The zero-value function \( f(x, y, z) = 0 \) corresponds to the 3D segmentation surface. Barbosa et al. [254] proposed the B-spline explicit active surface, which applies the B-spline basis functions to define the surface function. This approach has three advantages to the prior works,
1. This approach combines characteristics of implicit and explicit formulations in the active surface framework, and thereby, it facilitates employing both local and global region-based energy terms, which are commonly used in the level-set framework.

2. It provides the ability of controlling the smoothness of the segmentation surface by setting the scale factor of the B-spline formulation.

3. The B-spline formulations allow separability of representation into three 1D functions. Thus, its underlying formulation of the surface evolution is simply based on 1D convolutions, and therefore, its computational costs is low-enough to be used in real-time operations.

The B-spline active surface representation is defined as follows,

\[ x_0 = g(y, z) = \sum_{k \in \mathbb{Z}^{n-1}} c[k] \beta(\frac{x^*}{h} - k), \]  

where \( x^* \in \mathbb{R}^2 \) is a pixel of coordinates \([y, z] \), and \( \beta(x^*) \) is the 2D uniform cubic symmetric B-spline, which is separable into two 1D functions, \( \beta(x^*) = \beta(y)\beta(z) \). The knots are located on a rectangular grid with spacing of \( h \). Also, \( c[k] \) are the coefficients of the B-spline representation. The region-based energy formulation of the level-set framework is applied for the surface evolution of B-spline active surface. For more information, readers are referred to [254]. Since this method has been recently introduced, it has not been widely used in medical image processing yet. Though, a few papers have reported the use of active surface model in medical applications [255–257].

### 3.2.2 Review of segmentation methods: Conclusion

We described deformable models commonly used in medical image processing applications. This review provided a clear vision on advantages and disadvantages of each category of the image segmentation methods. The table shown in Fig. 3.13 lists the segmentation methods and summarizes their advantages and disadvantages. The regional level-set (active contour without-edge) is the most popular method in medical image segmentation because of the following reasons,

- it provides topological flexibility to segment multi-structural organs;
- it easily incorporates prior shape models to segment internal organs of interest;
- it provides an easy implementation for 3D image segmentation.

In the next section, we introduce a new segmentation approach, based on the regional level-set with shape prior, which provides the ability to segment multi-regional organs in 3D ultrasound images.
### Table 3.13: Segmentation Methods and Their Characteristics

<table>
<thead>
<tr>
<th>Name</th>
<th>Advantages</th>
<th>Disadvantages</th>
<th>Rate of use</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Region-based methods</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Thresholding</td>
<td>- Simplicity of development into 3D segmentation,</td>
<td>- Can easily leak into non-desired regions through blurred boundaries,</td>
<td>3</td>
<td>[132-142]</td>
</tr>
<tr>
<td></td>
<td>- Fast if working with mean intensity level of voxels.</td>
<td>- Sensitive to initialization,</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Does not support topological constraints on the segmentation region.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Watershed</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Clustering</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Deformable growing</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Active shape model</td>
<td>Adding statistical shape (&amp; texture) information of an organ of interest.</td>
<td>Requires manual initialization of the shape model close to object of interest.</td>
<td>2</td>
<td>[199-201]</td>
</tr>
<tr>
<td>Active appearance model</td>
<td></td>
<td></td>
<td>2</td>
<td>[206-209]</td>
</tr>
<tr>
<td><strong>Explicit representation</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Continuous models</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3D snake</td>
<td>- Flexibility of surface smoothness,</td>
<td>High comp. time since it requires dense sample points for accurate 3D segment.</td>
<td>4</td>
<td>[156]</td>
</tr>
<tr>
<td></td>
<td>- Robustness against noise.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B-spline surface</td>
<td>Its surface is implicitly smooth, and thus, no need to use internal energy.</td>
<td>Topological inflexibility.</td>
<td>3</td>
<td>[163],</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>[166-169]</td>
</tr>
<tr>
<td>Geodesic active contour</td>
<td>Free of parametrization</td>
<td>- Traps in local minima,</td>
<td>4</td>
<td>[157]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Fails to extract weaker edges</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Superquadrics</td>
<td>Providing local and global deformations.</td>
<td>Topological inflexibility.</td>
<td>3</td>
<td>[175-177]</td>
</tr>
<tr>
<td>Active rays</td>
<td>Avoids crossing of surface samples</td>
<td>Must be initialized close to the object.</td>
<td>4</td>
<td>[183]</td>
</tr>
<tr>
<td><strong>Probabilistic models</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Active shape model</td>
<td></td>
<td></td>
<td>2</td>
<td>[199-201]</td>
</tr>
<tr>
<td>Active appearance model</td>
<td></td>
<td></td>
<td>2</td>
<td>[206-209]</td>
</tr>
<tr>
<td><strong>Discrete models</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Deformable mesh</td>
<td>Free of parametrization</td>
<td>High degree of surface freedom, needs constrains to reduce possible deform.</td>
<td>3</td>
<td>[215-217]</td>
</tr>
<tr>
<td>T-surface</td>
<td>Providing topological flexibility by merging and splitting T-Surfaces.</td>
<td>Its re-parametrization process is sensitive to speckle interference.</td>
<td>4</td>
<td>[219]</td>
</tr>
<tr>
<td>Spring level-set</td>
<td>Bridge between implicit and explicit representation models.</td>
<td>High complexity of prior shape modeling</td>
<td>4</td>
<td>[223]</td>
</tr>
<tr>
<td><strong>Implicit representation</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Active surface</td>
<td>- Combines characteristics of implicit and explicit formulations,</td>
<td>- Does not provide ability to segment multi-regional structures.</td>
<td>3</td>
<td>[243-245]</td>
</tr>
<tr>
<td></td>
<td>- Easy control of surface smoothness</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Level-set methods</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Edge-based level-set</td>
<td>- Free of parametrization,</td>
<td>- Exceeds to non-desired regions from low contrast edges on objects of interest.</td>
<td>2</td>
<td>[42,158,</td>
</tr>
<tr>
<td></td>
<td>- Topological flexibility.</td>
<td></td>
<td></td>
<td>227–230]</td>
</tr>
<tr>
<td>Regional level-set</td>
<td>- Robust against low-contrast edges</td>
<td>- Does not include shape prior knowledge</td>
<td>1</td>
<td>[233-235]</td>
</tr>
<tr>
<td>Regional level-set with</td>
<td>- Supports shape prior knowledge in 3D image segmentation</td>
<td>- Not robust against intensity profile inhomogeneity</td>
<td>2</td>
<td>[237-239]</td>
</tr>
<tr>
<td>shape prior</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Parametric level-set</td>
<td>- Simplifies energy minimization problem into discrete convolutions.</td>
<td>- Difficult modelling of prior shapes.</td>
<td>4</td>
<td>[241]</td>
</tr>
<tr>
<td></td>
<td>- Smooth segmentation</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Figure 3.13:** This figure lists the segmentation methods along with their advantages, disadvantages, rate of use in medical image segmentation, and references (examples) of papers which used the segmentation methods in 3D ultrasound imagery.
3.3 Proposed kidney segmentation

In this section, we introduce a new segmentation method, based on the regional level-set with shape prior, which is named complex-valued regional level-set with shape prior (CVRLS-SP), and is designed to segment the kidney shape into its multiple structural regions in 3D ultrasound images. Also, we will show how to incorporate the proposed atlas-based kidney detection method in the CVRLS-SP to provide a fully automated kidney segmentation approach. Before proceeding to the introduction of the proposed method, we recommend readers to read Sec. 3.2.1.3.2 and Sec. 3.2.1.3.3, which provides detailed mathematical definitions for some equations used in the proposed kidney segmentation.

The implicit deformable model, so-called the level-set method, has been widely used to segment abdominal organs, in which voxels inside and outside segmented regions receive positive and negative values in a Lipschitz function, $\phi(\vec{X})$, respectively. As discussed in Sec. 3.2.2, the regional level-set with shape prior provides a well-suited framework for segmenting anatomical structures in medical images. This is because

- its representation model is non-parametric,
- it provides topological flexibility,
- it is robust against low-contrast edges,
- it facilitates the use of shape prior in the segmentation process of anatomical structure.

Though, the conventional level-set framework does not provide segmentation of multiple regions for anatomical structures. In the targeted application of this thesis, we are interested to segment the kidney shape into its structural regions, including the renal medulla, kidney capsule, and pyelocalyceal system. This is important because (a) it provides more information from the patients for the clinical use, and (b) involving structural regions improves the segmentation specificity to the kidney shape. The multi-phase level-set (MPLS) approach was proposed to support segmentation of anatomical structures with multiple regions [258]. In the MPLS, regions are distinguished by their different average levels, however, this is not an option in 3D ultrasound images.

The CVRLS-SP is designed to support the representation of anatomical structures into two classes plus background (similar to the structure of the kidney CVISM). In Sec. 2.2.2, we showed the renal medulla region appears with dark voxels, while the kidney capsule and pyelocalyceal system regions appear with bright voxels. Therefore, a regional level-set representation model, supporting three regions, including $bg$: background, $br$: bright tissue, and $dr$: dark tissue, in conjunction with a shape prior model
(based on the kidney CVISM), suffices the segmentation of the kidney shape into its three anatomical regions. More specifically, it would be enough to assign $dr$ to the renal medulla, $br$ to the kidney capsule and pyelocalyceal system regions, and $bg$ to the background voxels. Such a multi-regional representation is simply provided using complex-valued representation, such that voxels with real-positive values of the level-set function correspond the kidney capsule and pyelocalyceal system regions, voxels with imaginary-positive values of the level-set function correspond to the renal medulla, and voxels with negative real and negative imaginary parts belong to the non-kidney regions. This is mathematically defined as follows,

\[
\begin{aligned}
(br) : \vec{X} &\in \text{pyelocalyceal system or kidney capsule, if } \Re\{\phi(\vec{X})\} > 0, \\
(dr) : \vec{X} &\in \text{renal medulla, if } \Im\{\phi(\vec{X})\} > 0, \\
(bg) : \vec{X} &\in \text{background, otherwise,}
\end{aligned}
\]

where $\vec{X} = [x, y, z]^T \in \Omega_V$, and $\Omega_V$ is the 3D image domain. $\Re\{}$ and $\Im\{}$ are real-part and imaginary-part extracting operators. Also, we apply the shape prior information in the regional level-set framework for the following reasons:

- to ensure that the segmentation result stays similar to the kidney shape, with a controllable amount of local deformations;
- to prevent the segmentation process from leaking into non-kidney regions through gaps of the kidney capsule region.

In the prior arts, the initialization of the shape prior has been known as a manual task in which an operator’s intervention was required. In this thesis, we employ the atlas-based (or shape-based) kidney detection method to register the kidney CVISM on the kidney shape in an ultrasound volume, and use the fitted CVISM to initialize the shape prior of the CVRLS-SP. Thereby, we eliminate the need of manual intervention, and obtain a fully automated kidney segmentation approach. Remembering from Sec. 2.6.2, the outputs of the shape-to-volume registration are $\vec{p}_{st,1}$ and $\vec{p}_{st,2}$, in which $\vec{p}_{st,1}$ specifies the orientation/scaling of the kidney shape in the ultrasound volume with respect to the kidney CVISM, and $\vec{p}_{st,2}$ specifies the center point of the kidney shape in the ultrasound volume. The shape prior, $\phi_s$, is formed from the detected kidney shape as follows,

\[
\phi_s(\vec{X}) = 2\left(\left\{\Psi \left(ST_{\vec{p}_{st,1}} \vec{X} + \vec{p}_{st,2}\right)\right\} > 0\right) - (1 + i), \quad \forall \vec{X} \in \Omega_V
\]

where $ST_{\vec{p}_{st,1}} \vec{X} + \vec{p}_{st,2}$ places the kidney CVISM, $\Psi$, on the detected kidney shape. The level-set function is also initiated by the shape prior as, $\phi(X,t=0) = \phi_s(\vec{X})$, where $t$ is the artificial time, specifying
In equation (3.79), the real and imaginary parts carry structural information of the kidney shape. In fifth and sixth terms are external energies, and they apply regional information in the front propagation. Fourth terms of equation (3.79) are internal energies, and they control the shape smoothness, while the ensure the evolving shape of the level-set function follows the kidney shape model [248]. The third and each iteration. The first two terms of equation (3.79) are shape prior constraints, and are applied to

{\begin{align*}
F(c_1, c_2, \phi) &= \gamma \iiint (H(\Re\{\phi(\vec{X})\}) - H(\Re\{\phi_s(AT_{\vec{p}})\{\vec{X}\}\}))^2 dxdydz \\
&+ \gamma \iiint (H(\Im\{\phi(\vec{X})\}) - H(\Im\{\phi_s(AT_{\vec{p}})\{\vec{X}\}\}))^2 dxdydz \\
&+ \mu \iiint \delta(\phi(\vec{X})) \left| \nabla \Re\{\phi(\vec{X})\} + i \nabla \Im\{\phi(\vec{X})\} \right| dxdydz \\
&+ \nu \iiint |H(\phi(\vec{X}))| dxdydz \\
&+ \lambda \iiint \left| V_{dn}(\vec{X}) - c_{br} \right|^2 (H(\Re\{\phi(\vec{X})\}) + (1 - H(\Im\{\phi(\vec{X})\}))) dxdydz \\
&+ \lambda \iiint \left| V_{dr}(\vec{X}) - c_{dr} \right|^2 (H(\Im\{\phi(\vec{X})\}) + (1 - H(\Re\{\phi(\vec{X})\}))) dxdydz,
\end{align*}}

(3.79)

where \( \delta(\cdot) \) is the Dirac delta function, and \( \delta(\phi(\vec{X})) \) specifies voxels on the zero-level surface of \( \phi \). \( H(\cdot) \) is the Heaviside function, and \( H(\phi(\vec{X})) \) and \( 1 - H(\phi(\vec{X})) \) specify regions inside and outside the zero-level surface of \( \phi \), respectively. \( c_{br} \) and \( c_{dr} \) are the average intensity levels of the kidney capsule/pyelocalyceal system and the renal medulla regions, respectively. \( AT_{\vec{p}} \) is the affine transformation, and \( \vec{p} = [\theta_x, \theta_y, \theta_z, s_x, s_y, s_z, t_x, t_y, t_z]^T \) is the vector of affine transformation parameters, including three orientations, \{\theta_x, \theta_y, \theta_z\}, three scaling factors, \{s_x, s_y, s_z\}, and three translation factors, \{t_x, t_y, t_z\}. The affine transformation is applied to match the prior shape with the level-set function at each iteration. The first two terms of equation (3.79) are shape prior constraints, and are applied to ensure the evolving shape of the level-set function follows the kidney shape model [248]. The third and fourth terms of equation (3.79) are internal energies, and they control the shape smoothness, while the fifth and sixth terms are external energies, and they apply regional information in the front propagation. In equation (3.79), the real and imaginary parts carry structural information of the kidney shape. In
equation (3.79), $\gamma$ is the Lagrangian multiplier of the shape prior ($\gamma > 0$), and controls the influence of shape prior constraint on the level-set evolution. $\nu$ is a Lagrangian multiplier, and works like a constant speed, similar to the balloon force in 3D snake (Sec 3.2.1.1.1). $\mu$ is a Lagrangian multiplier, controlling the smoothness of the zero level-set surface. Increasing the value of $\mu$ results in a smoother segmentation, however it reduces from the level of details in the segmented region. $\lambda$ is the Lagrangian multiplier of the regional information, and controls the influence of the external energy on the level-set evolution. $c_{br}$ and $c_{dr}$, which are the average intensity of pyelocalyceal system/kidney capsule and renal medulla, are calculated using the following equations,

$$
c_{br} = \frac{\iiint_{\bar{Y}=\{x,y,z\} \in \Omega} H(\Re\{\phi(\bar{Y})\})V_{dn}(\bar{Y})dxdydz}{\iiint_{\bar{Y}=\{x,y,z\} \in \Omega} H(\Re\{\phi(\bar{Y})\})dxdydz},
$$

(3.80)

and,

$$
c_{dr} = \frac{\iiint_{\bar{Y}=\{x,y,z\} \in \Omega} H(\Im\{\phi(\bar{Y})\})V_{dn}(\bar{Y})dxdydz}{\iiint_{\bar{Y}=\{x,y,z\} \in \Omega} H(\Im\{\phi(\bar{Y})\})dxdydz}.
$$

(3.81)

The equation (3.79) is iteratively solved using the Euler-Lagrange equation, as follows,

$$
\frac{\partial \phi}{\partial t}(\bar{X}) = 2\gamma \left( H(\Re\{\phi(\bar{Y})\}) - H(\Re\{\phi_s(\mathbf{AT}_{\mathbf{p}_{af}}\{\bar{X}\})\}) \cdot \delta(\Re\{\phi(\bar{X})\}) \right)
+ 2\gamma \left( H(\Im\{\phi(\bar{Y})\}) - H(\Im\{\phi_s(\mathbf{AT}_{\mathbf{p}_{af}}\{\bar{X}\})\}) \cdot \delta(\Im\{\phi(\bar{X})\}) \right)
+ \mu \delta(\phi(\bar{X})) \text{div} \left( \frac{\nabla \phi}{|\nabla \phi|} \right) - \nu
+ \lambda \cdot \left((V_{dn}(\bar{X}) - c_{br})^2 - (V_{dn}(\bar{X}) - c_{dr})^2\right) \cdot \delta(\Re\{\phi(\bar{X})\})
+ \lambda \cdot \left((V_{dn}(\bar{X}) - c_{dr})^2 - (V_{dn}(\bar{X}) - c_{br})^2\right) \cdot \delta(\Im\{\phi(\bar{X})\})
$$

(3.82)

In equation (3.82), $\nu$ acts as a constant velocity. Since the level-set function is initiated close to the actual boundaries, the constant velocity is not needed, and $\nu$ is zeroed. The parameters of the affine transformation should be updated each $N_{itr} \geq 1$ iterations. By reducing $N_{itr}$, the shape prior will have a smoother follow of the level-set function, providing a more robust segmentation against noise and gaps within the kidney capsule region. However, reducing $N_{itr}$ increases the computational cost. The value of $N_{itr}$ can be selected based on a trade-off between the robustness and computational time. We empirically set $N_{itr} = 5$. The parameters of the affine transformation are obtained by solving the
following minimization problem,

\[ \tilde{p}_{af}^t = \arg\min_{\tilde{p}_{af}} \int \int \int_{\Omega_V} \left( H(\Re\{\phi(\tilde{X})\}) - H(\Re\{\phi_s(\tilde{X})\}) \right)^2 + \left( H(\Im\{\phi(\tilde{X})\}) - H(\Im\{\phi_s(\tilde{X})\}) \right)^2 d\tilde{x} d\tilde{y} d\tilde{z} \]

(3.83)

where \( \tilde{p}_{af}^t \) is the updated vector of affine transformation parameters at iteration \( t \). This equation is solved for each parameter in \( \tilde{p}_{af} \) while fixing the other parameters, using the Gradient Descent method (similar as calculating parameters of the similarity transformation in Sec. 2.2.2). Figure 3.14 shows the block diagram of the proposed kidney segmentation method (CVRLS-SP). The iterative process continues until the convergence metric, \( \Sigma \Delta \phi(t) \), is smaller than a small constant, \( \epsilon \). The convergence metric is defined as follows,

\[ \Sigma \Delta \phi(t) = \sqrt{\int \int \int_{\Omega_V} \left( \phi(\tilde{X}, t) - \phi(\tilde{X}, t-1) \right)^2 d\tilde{x} d\tilde{y} d\tilde{z}}. \]

(3.84)

The convergence metric in eq. (3.84) calculates the integral of changes in the updated level-set function at iteration, \( t \), with respect to the previous iteration, \( t - 1 \). If the integral of changes, \( \Sigma \Delta \phi(t) \) is smaller than the threshold value, \( \epsilon \), then the iterative evolution of the level-set propagation stops, which means the level-set function has reached the desired region of interest.
3.4 Experimental results

In this section, we represent experiments to evaluate the proposed kidney segmentation method (CVRLS-SP), compared to the Marsousi et al-EMBC14 [63], Noll et al. [12], and MRF-AD [13] methods.

3.4.1 Hardware and software setup

We use the same hardware and software configuration that was discussed in 2.9.1 for evaluating the kidney detection methods.

3.4.2 3D ultrasound dataset

The training set ($Ds3DUsIs-AUVs-Tr$), including six with-kidney and six without-kidney images, are used to conduct parameters analysis of the CVRLS-SP method. The with-kidney images of the evaluation set of healthy volunteers ($Ds3DUsIs-AUVs-ESH$), including 15 with-kidney images, are used to evaluate the performance of the kidney detection methods. Also, each with-kidney image in the training and evaluation datasets has at least one 3D ground truth ($Ds3DUsIs-GTs$).

3.4.3 Comparison metrics

To evaluate the kidney segmentation methods, we use three types of metrics: (a) overlap measures, (b) sensitivity/specificity, and (c) error measures. The overlap metrics provide a measure of shapes’ similarities, while the sensitivity and specificity measures show how much a segmentation method is sensitive and specific to the kidney’s shape variability. The error measures estimate how close a segmented region is to its ground truth data. Assume $GT$ and $AS$ are binarized ground truth and automated segmentation, respectively. We use the Dice’s similarity coefficient ($DSC$), and accuracy measure ($ACC$) [259] to measure the overlap of ground truth and automated segmentation result. The overlap measures, sensitivity ($SENS$) and specificity ($SPEC$) are defined as,

\[
\begin{align*}
DSC &= \frac{2TP}{2TP + FN + FP}, \\
ACC &= (100\%) \cdot \frac{TP + TN}{TP + TN + FP + FN}, \\
SENS &= \frac{TP}{TP + FN}, \\
SPEC &= \frac{TN}{TN + FP},
\end{align*}
\]

(3.85)

where $TP$, $TN$, $FP$, and $FN$ are true positive, true negative, false positive, and false negative segmentation regions, respectively. The segmentation regions are denoted in Fig. 3.15. Note that voxels
outside the ultrasound pyramid are excluded from the scoring process. The \( DSC \) metric does not use \( TN \) in its calculation, whereas in the \( ACC \) metric, \( TN \) is accounted, and since \( TN \) is usually greater than \( TP \), \( ACC \) is greater than \( DSC \). To analyze the parameters of the proposed kidney segmentation, we use the \( DSC \) measure. We use two metrics to calculate segmentation errors: mean distance (\( MD \)) and Hausdorff distance (\( HD \)) as,

\[
\begin{align*}
MD &= \frac{1}{|AS|} \int_{p' \in AS} e(p', GT) dp', \\
HD &= \max_{p \in AS} \{ e(p, GT) \},
\end{align*}
\]

where \( e(p, GT) \) is the minimum \( L-2 \) norm distance of a voxel \( p \) to \( GT \). The Hausdorff distance measures the maximum distance between corresponding points on \( GT \) and \( AS \) surfaces, whereas the mean distance compares \( GT \) and \( AS \) by calculating the average distances from corresponding points, without extracting their surfaces [260]. The Hausdorff distance emphasizes outliers, while the mean distance measures the average segmentation error. These two metrics have been widely used in literature to measure the segmentation error in medical images [261,262].

### 3.4.4 Parameter analysis

The proposed CVRLS-SP method for segmenting the kidney shape has four parameters, \( \gamma, \mu, \nu, \) and \( \lambda \). Since the level-set function is initialized by the registered kidney CVISM through the kidney detection process, we assume the initial level-set function is close enough to the regions of the kidney shape. Thus, the parameter \( \nu \), which operates like a balloon force to help the segmentation region to escape from homogenous regions, is not required, and we set \( \nu = 0 \). For the sake of simplicity, we set \( \lambda = 1 \), and thereby, we narrow down the parameter analysis into two parameters, including \( \gamma \) and \( \mu \). These two parameters are analyzed to find their ranges of robust operation for kidney segmentation. To ensure the segmentation results are not biased toward the parameter analysis, parameter analysis is performed.
Figure 3.16: The 3D surface, plotted in this figure, represents the average kidney segmentation accuracy, based on the DSC metric, for the training volumes of the dataset of actual ultrasound volumes ($Ds3DUsIs-AUVs-Tr$).

Table 3.2: List of parameters of the implemented CVRLS-SP method for segmenting the kidney shape in ultrasound volumes.

<table>
<thead>
<tr>
<th>Parameter Name</th>
<th>Description</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\lambda$</td>
<td>Lagrange multiplier of controlling shape prior influence on level-set evolution</td>
<td>1</td>
</tr>
<tr>
<td>$\mu$</td>
<td>Lagrange multiplier controlling the smoothness of the Front propagation</td>
<td>0.1</td>
</tr>
<tr>
<td>$\nu$</td>
<td>Lagrange multiplier of a constant speed of the Front propagation</td>
<td>0</td>
</tr>
<tr>
<td>$\gamma$</td>
<td>Lagrange multiplier controlling the regional information</td>
<td>0.05</td>
</tr>
<tr>
<td>$N_{itr}$</td>
<td>Number of iterations which the affine transformation parameters of the shape prior is updated</td>
<td>5</td>
</tr>
<tr>
<td>$t_{max}$</td>
<td>Maximum number of iterations</td>
<td>50</td>
</tr>
<tr>
<td>$\epsilon$</td>
<td>A constant criteria to stop the iterative process of CVRLS-SP evolution</td>
<td>10</td>
</tr>
</tbody>
</table>

using the six with-kidney images of the training set of actual ultrasound volumes ($Ds3DUsIs-AUVs-Tr$). The CVRLS-SP method are applied on the training volumes for $\gamma \in \{0, 0.025, 0.050, \cdots, 0.575\}$ and $\mu \in \{0, 0.025, 0.050, \cdots, 0.5\}$. The result is shown in Figure 3.16, which displays a 3D surface of the average kidney segmentation accuracy, based on the DSC metric, representing the parameter analysis for $\gamma$ and $\mu$. Accordingly, we set $\gamma = 0.05$ and $\mu = 0.1$. The parameters of the implemented CVRLS-SP for segmenting the kidney shape are listed in Table 3.2.


3.4.5 Evaluating kidney segmentation methods

This experiment is designed to compare the kidney segmentation methods. We apply the methods on the with-kidney images of the evaluation set of healthy volunteers (Ds3DUis-AUVs-ESH), and calculate DSC, ACC, SENS, SPEC, HD, and MD metrics for each one of the four segmentation methods. For each with-kidney image, the accuracy of segmenting the kidney shape is calculated for the entire kidney shape, including the renal medulla, pyelocalyceal system, and kidney capsule regions. This is because Marsousi et al-EMBC14, Noll et al., and MRF-AC, are not capable to discriminate the kidney’s structural regions in kidney segmentation. Table 3.3 represents mean, $\mu$, and standard deviation, $\sigma$, of the accuracy metrics of segmenting the kidney’s shape in the with-kidney images of the evaluation set. According to Table 3.3, the average accuracy by the proposed CVRLS-SP method, $DSC = 0.8143 \pm 0.0408$, is better than the average DSC of the other three methods. According to Table 3.3, the proposed method of this paper provides higher accuracy (ACC), higher sensitivity (SENS), higher specificity (SPEC), and lower HD error among the methods. Figure 3.17 shows kidney segmentation results of three with-kidney images for the segmentation methods. As shown in Fig. 3.17, the automated segmentation obtained by the proposed method provides the best match with the ground truth data. The Marsousi-EMBC14’s method suffers from improper initializations of the segmentation process, and the Noll et al. method leaks into undesired regions through gaps within the kidney’s boundary. According to Table 3.3, the proposed CVRLS-SP method possess the minimum Hausdorff distance error, $HD = 8.41$ (mm), and the minimum mean distance error, $MD = 0.29$ (mm), stating that the segmented kidney shape using the CVRLS-SP method provides the least deviation from the actual kidney shape in the actual ultrasound volumes (Ds3DUis-AUVs-ESH). Comparing the methods based on the Hausdorff distance error in millimeters might have more clinical relevance for medical doctors than comparing the segmentation accuracy based on the DSC metric, which is more important from the image processing perspective. Figure 3.18 shows the proposed kidney segmentation procedure for three with-kidney images with different postures of the kidney shape.

Table 3.3: Comparing DSC and ACC, sensitivity SENS, specificity SPEC, HD and MD of the segmentation methods: the proposed method of this paper, Marsousi et al-EMBC14 (M-EMBC14), Noll et al., and MRF-AC. $\mu$ and $\sigma$ are mean and standard deviation, respectively. The units of DSC, SENS and SPEC are pixels, and the units of both HD and MD are millimeters (mm).

<table>
<thead>
<tr>
<th>Method</th>
<th>DSC</th>
<th>ACC (%)</th>
<th>SENS</th>
<th>SPEC</th>
<th>HD (mm)</th>
<th>MD (mm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CVRLS-SP</td>
<td>0.8143</td>
<td>0.0408</td>
<td>97.48</td>
<td>0.72</td>
<td>0.7863</td>
<td>0.0814</td>
</tr>
<tr>
<td>M-EMBC14</td>
<td>0.5792</td>
<td>0.0832</td>
<td>94.37</td>
<td>1.37</td>
<td>0.5525</td>
<td>0.1110</td>
</tr>
<tr>
<td>Noll et al.</td>
<td>0.4207</td>
<td>0.0793</td>
<td>85.48</td>
<td>2.77</td>
<td>0.7024</td>
<td>0.0912</td>
</tr>
<tr>
<td>MRF-AC</td>
<td>0.5921</td>
<td>0.1457</td>
<td>93.10</td>
<td>2.74</td>
<td>0.6947</td>
<td>0.2132</td>
</tr>
</tbody>
</table>
Figure 3.17: This figure shows results of the kidney segmentation methods, including (a) the proposed method of this paper, (b) Marsousi et al-EMBC14, (c) Noll et al., and (d) MRF-AC, of three ultrasound volumes of the evaluation set. The green, red and yellow colors represent FN, FP, and TP regions, respectively.

Figure 3.18: Displaying automated kidney segmentation process of three with-kidney images with low, middle, and high accuracy of kidney segmentation as, $DSC = 0.70$, $DSC = 0.83$, and $DSC = 0.86$ respectively. The three images represent three different kidney shape’s postures with different scales and orientations. The sub-figures from left to right show original volume, detected kidney shape, segmented kidney shape, comparison of AS vs. GT, and 3D rendered AS. In the second and third sub-figures from left, the blue and red regions indicate renal medulla and pyelocalyceal system/kidney capsule, respectively. In the second sub-figure from right side, the yellow, red, and green colors indicate TP, FP, and FN regions, respectively.

Also, the accuracy of segmenting the structural regions of the kidney shape using the proposed CVRLS-SP method are calculated, and the segmentation results of the renal medulla region and the combination of the kidney capsule and pyelocalyceal system regions, as the mean plus/minus the standard deviation of calculated DSCs, are $DSC = 0.7618 \pm 0.0508$ and $DSC = 0.6907 \pm 0.0422$, respectively. Accordingly, the segmentation accuracies of each individual region is lower than the segmentation accuracy of the entire kidney shape ($DSC = 0.8143 \pm 0.0408$). This is because the segmented structural regions interfere each other at some parts of the kidney shape. Note that the prior art methods were
Chapter 3. Automated kidney segmentation

Figure 3.19: This box-plot displays statistical variations of the accuracy of kidney segmentation, based on the Dice’s similarity coefficient, using the proposed CVRLS-SP method, Marsousi et al-EMBC14, Noll et al. and MRF-AC methods.

Figure 3.20: The table in this figure represents the statistical comparison of the kidney segmentation accuracy of the methods, including the proposed CVRLS-SP, Marsousi et al-EMBC14 (M-EMBC14), Noll et al., and MRF-AC. The results are obtained using the with-kidney images of the evaluation set of actual ultrasound volumes of the healthy volunteers (Ds3DUsIs-AUVs-ESH). $h$ is the state of the t-test analysis for each pair of methods, based on the significance level of 5%. For each non-diagonal cell of the table, $h = 1$ means the corresponding method on the left column performs better than the corresponding method in the top row, $h = -1$ means the corresponding method in the top column performs better than the corresponding method in the left row, and $h = 0$ means the experimental data of the corresponding methods do not show a significant difference.

In figure 3.19, the box-plot representation is used to show statistical variations of kidney segmentation accuracy of the methods, based on the DSC metric. Accordingly, the minimum DSC of the proposed method of this paper is equal or higher than the maximum DSC of the other methods. We also performed statistical analysis, based on the t-test method, to determine the significance of comparing kidney segmentation results using the proposed CVRLS-SP, Marsousi et al-EMBC14 (M-EMBC14), Noll et al., and MRF-AC methods. The result of statistical analysis is shown in Fig. 3.20, which shows the state of test ($h$) and the $p$-value for each test. For each cell in the table shown in Fig. 3.20, assume the method-a and method-b are corresponding left column and top row, respectively. Based on this assumption, the state of test, $h$, can take values from $\{-1, 0, 1\}$, in which a $-1$ means method-b performs better than method-a, a 1 means method-a performs better than method-b, and a 0 means we cannot conclude which method performs better. The results in Fig. 3.20 are obtained based on the 5%
significance level. Accordingly, the proposed kidney segmentation method (CVRLS-SP) has obtained better segmentation accuracy, compared to the other methods.

3.5 Summary and Concluding Remarks

In this chapter, we first discussed the state-of-the-art in kidney segmentation in 3D ultrasound imagery, explained their methodologies, and summarized their advantages and disadvantages pertinent to the application of this thesis, which is automated kidney segmentation in 3D ultrasound images. Then, we categorized 3D segmentation methods which have been commonly used in medical applications, and provided detailed introduction to each category of segmentation methods. We summarized the methods with their pros and cons, and selected the category of implicit representation methods, in particular the regional level-set with shape-prior method, as the best matching approach to the targeted application.

After discussing the existing methods and their characteristics, we introduced a new segmentation algorithm to accommodate the requirements of the targeted application of this thesis. The new segmentation method was designed based on the regional level-set with shape-prior method, aiming to add the capability of segmenting internal structures of interest (i.e. the kidney shape) into its multi-structural regions (i.e. the renal medulla and the combination of kidney capsule and pyelocalyceal system) in 3D ultrasound images. We discussed how this capability can provide both clinical advantage and higher segmentation accuracy. The proposed segmentation method was named the complex-valued regional level-set based on shape prior, because it used the complex-values to represent both bright tissues and dark tissues of an organ of interest, and thereby, we showed this representation was sufficient to separate a multi-regional structure/organ of interest from its surrounding tissues. We explained that by using the atlas-based (or shape-based) kidney detection module to initialize the CVRLS-SP, the proposed kidney segmentation will be fully automated. This is an essential requirement for the application of this thesis, in which paramedics without required anatomical and radiological understanding for triaging are the targeted users to utilize the proposed solution in emergency situations, and therefore, the proposed method should not rely on the operator’s knowledge. Hence, the solution should be able to autonomously segment the kidney shape in ultrasound volumes. The utility of the CVRLS-SP method in kidney segmentation was evaluated through experimentation using actual ultrasound volumes. The parameters of the CVRLS-SP method were determined through parameter analysis using the training set of actual ultrasound volumes. We compared the segmentation accuracy of the proposed method with the state-of-the-art using similarity, sensitivity, specificity, and error metrics. The experimental results confirmed the improved segmentation accuracy of the kidney shape using the proposed CVRLS-SP method of this thesis, compared to the state-of-the-art.
Chapter 4

Conclusions and future work
4.1 Research summary

Every year, abdominal trauma causes a large portion of preventable deaths. A trauma patient with massive abdominal bleeding becomes soon hemodynamically unstable, and needs emergency diagnosis and medical services to save his/her life. Advanced medical services for curing trauma injuries are usually available in hospitals and referral clinical centers, however, such medical services are not always provided for trauma patients in remote areas, resulting in a high volume of fatality. The shortage of medical services in emergency situations for trauma patients includes (a) the lack of diagnostic devices, and (b) inexperience of paramedics in conducting rapid diagnosis.

In this thesis, we introduced image processing algorithms based on 3D ultrasound imagery to be used for developing a computer-assisted diagnosis of trauma patients in emergency situations. The algorithms was designed by incorporating prior anatomical knowledge of the kidney shape to guide paramedics, who may lack required anatomical knowledge to conduct trauma diagnosis, in two ways: (a) to conduct an abdominal sonography associated with the trauma diagnosis, and (b) to make a clinical decision in a remote emergency situation where skilled trauma specialists are not present.

We discussed the importance of 3D ultrasound imagery to facilitate the design of a computer-assisted diagnostic system for trauma patients in remote areas. The portability, affordability, real-time imaging, and non-intrusiveness of 3D ultrasound imagery were some important reasons of its privilege over other imaging modalities, such as CT and MRI, for designing a computer-assisted trauma diagnostic system. However, processing 3D ultrasound imagery is a challenging task, because it faces many difficulties associated to the nature of ultrasound scan. Poor resolution, speckle interference, low-contrast edges and discontinuous boundaries of organs, inhomogeneous intensity profile of similar tissue layers are among the most important challenges of ultrasound image processing.

In this thesis, we focused on the right-upper-quadrant (RUQ) view of the abdominal ultrasound scan, because this view has shown the highest sensitivity to abdominal bleeding. We discussed that detecting and segmenting the kidney shape constitute the main key for designing the proposed trauma diagnostic solution, for at least two reasons as follows:

- an abdominal bleeding has the tendency to locate around the right kidney in a trauma patient laid in the supine position;

- the right kidney has a unique structure among internal organs and it is entirely visible in the RUQ view, and thus, it can be used as the landmark to find the correct alignment of the ultrasonic probe for scanning the RUQ view.
Therefore, in this thesis, we introduced algorithms to automatically detect and segment the kidney shape in 3D ultrasound images.

The importance of kidney detection in the proposed trauma diagnosis solution attributes to the following reasons:

- it is required for designing a computer-assisted ultrasonic probe placement;
- it facilitates automated kidney segmentation.

To accommodate these two requirements, kidney detection should well address the following questions:

- does the kidney shape exist in an ultrasound volume?
- if it exists, what is its alignment in terms of orientation, scale, and location with respect to the reference alignment of the kidney shape?

To address these two queries, we introduced two kidney detection methods in chapter 2, including the shape-based and atlas-based kidney detection methods. The shape-based method only used shape prior information of the kidney shape, extracted from training ultrasound volumes, to perform the kidney detection task, while the atlas-based method used both shape prior and texture information of the kidney shape, extracted from training ultrasound volumes, to detect the kidney shape in 3D ultrasound images. The combination of shape prior and texture information made the atlas-based method to be more specific to the kidney shape, however, its sensitivity to the kidney shape orientation and translation was reduced, compared to the shape-based method. In addition, we showed through experimentation that the atlas-based method provided a more accurate estimation of the kidney shape misalignment in terms of both orientation and translation. In the other side, the shape-based method has shown a wider range of detectability of the kidney shape in terms of orientation and translation. Accordingly, we designed a strategy for computer-assisted probe placement to efficiently use the capabilities of both the shape-based and atlas-based methods as follows:

- The shape-based method should be used at the beginning of the computer-assisted probe placement process, because the probe placement may have a large deviation from the correct alignment, resulting in a wide range of kidney shape orientation and translation. In such a wide range of deformations of the kidney shape, the shape-based methods provides a better detectability.

- After a few iterations of the computer-assisted probe placement, the kidney shape misalignment, with respect to the reference alignment in the RUQ view, is expected to be reduced. Then, the
Chapter 4. Conclusions and future work

atlas-based method should be used, because it provides a higher accuracy of kidney misalignment estimation.

In addition to the utility of both the shape-based and atlas-based kidney detection methods in computer-assisted probe placement, the computational complexity of each of these two methods is a constraint to select one method for kidney detection. The atlas-based kidney detection method has an excessive computational cost, and needs a more advanced computer to conduct real-time operations, compared to the shape-based method. In the computer workstation of this research, which was used to implement methods based on parallel computation, the computational time of the shape-based volume was 8 seconds per volume, while the atlas-based method spent about 30 seconds to detect the kidney shape in a volumetric ultrasound image. Thus, the computational time of the shape-based method was close to the requirement of real-time operations, whereas the atlas-based approach needed faster processors with higher capabilities of parallel computation to reduce its computational time. Conclusively, both the shape-based and atlas-based methods are valuable in the proposed computer-assisted trauma diagnostic solution.

The proposed shape-based and atlas-based kidney detection methods consisted of the following parts:

• the CVISM modeling of the kidney shape,
• the MFLV method to reduce speckle interference,
• volume enhancement by combining the LHE and G2LTh methods,
• voxel classifications based on the SANNs,
• rigid shape-to-volume registration, consisting of:
  
  – finding the best seed point to initialize the optimization process,
  – iterative optimization algorithm based on the Gradient Descent method.

We explained both the higher accuracy of kidney detection and higher computational complexity of the atlas-base method were attributed to the use of voxel classification with SANNs based on 3D Gabor texture features, which were not used in the shape-based method. We explained the advantage of the CVISM shape representation for modeling abdominal organs in 3D ultrasound imagery. The CVISM used complex-valued implicit representation to implicitly model voxels inside and outside multi-structural regions of the kidney shape, including the renal medulla, kidney capsule, and pyelocalyceal system. This characteristic made the CVISM shape modeling to be distinctive among the many existing shape representation approaches. We proposed the MFLV method to reduce speckle noise in 3D ultrasound
Chapter 4. Conclusions and future work

images. The MFLV method provided a comparative speckle reduction, if not saying better, compared to the state-of-the-art, with a much lower computational cost. The proposed voxel classification based on SANNs applied a set of spatially aligned neural network classifiers to classify voxels into structural regions of the kidney shape. We also introduced a new similarity metric, the RCNCC metric, for the shape-to-volume registration task, which provided high sensitivity and specificity of shape registration to the kidney shape by the following considerations,

- the RCNCC metric allowed incorporating structural regions of the kidney shape into the registration task, increasing the specificity of shape registration to the kidney shape;
- the RCNCC metric included a regularization term to ensure the registration algorithm is able to detect partially visible kidney shapes.

In chapter 3, we introduced a new segmentation method to meet the requirements of the proposed computer-assisted trauma diagnostic solution. The new segmentation method adds the capability of multi-regional segmentation to the existing state-of-the-art method (ie. the regional level-set with shape-prior method). The proposed segmentation method was named as the complex-valued regional level-set based on shape prior, because it used complex-values to represent both bright tissues and dark tissues of an organ of interest, and thereby, we showed this representation was sufficient to separate a multi-regional structure/organ of interest from its surrounding tissues. We discussed how this capability can provide both clinical advantage and higher segmentation accuracy. We explained the proposed kidney segmentation method was fully automated, because the CVRLS-SP was initialized by the detected kidney shape using one of the introduced kidney detection methods (preferably the atlas-based approach). Noteworthily, automated kidney segmentation is an essential requirement for computer-assisted trauma diagnostic solution, in which paramedics without required anatomical and radiological understanding for triaging are the targeted users to utilize the proposed solution in emergency situations, and therefore, the proposed method should not rely on the operator’s knowledge. Hence, the proposed kidney segmentation met the requirement of the targeted application of this thesis.

4.2 Discussions

The introduced method in chapter 3 provided automated kidney segmentation in 3D ultrasound images. This was achieved by employing one of the kidney detection methods, which eliminated the need of an operator to initialize the segmentation process. In a real-time operation, one of the kidney detection methods is applied on ultrasound volumes, acquired at the rate of one volume per three seconds (based
The detected kidney shape is then segmented using the proposed kidney segmentation module, in an off-line process. Thus, the computational cost of the kidney segmentation module is not a concern, but the kidney detection module should be fast enough to operate in a real-time scenario. The proposed shape-based and atlas-based kidney detection methods have been developed using multi-core multi-threading and GPU-CUDA programming. The computational time of the shape-based method has been reduced from an average of 80 seconds to an average of eight seconds. The atlas-based method has gained much faster operation using parallel computation, and its current computational time using the applied computer workstation is about 30 seconds, which is slower than what we need for real-time operation. Based on the number of cores (i.e., 640 GPU-cores, each one running at 1.2GHz) and the memory bandwidth of the current GPU card, we expect the computational time of the atlas-based method falls below 10 seconds using a more powerful graphic card (e.g., NVIDIA Quadro M6000 which has 2400 GPU cores, each one running at 1.6GHz).

As discussed in section 2.9, the combination of the shape-based and atlas-based kidney detection methods provides both high accuracy and robustness against kidney shape deformation and occlusion. The good performance of the proposed kidney detection methods not only matters but also it is essential in the overall performance of the computer-assisted trauma diagnosis, because any error in kidney detection may either delay or prevent an operator from finding the correct placement of the ultrasonic probe for acquiring the Morison’s pouch view, resulting in the failure of rapid diagnosis of a trauma patient.

The reported kidney segmentation accuracy of the proposed method of this paper in Table 3.3, based on the average DSC metric, is 0.8143. As reported in Sec. 1.6.2, the inter-observer agreement of two trainees based on the DSC metric, is 0.8345. Thus, the accuracy of the proposed automated method is only 2.02% less than what can be achieved by trained operators.

The reported results in Table 2.9, the box-plot of Fig. 2.27, and the statistical analysis shown in Fig. 2.28, confirm the proposed shape-based and atlas-based kidney detection methods improve the kidney detection accuracy of the state-of-the-art. This improvement can be attributed to the following characteristics of the proposed shape-based and atlas-based kidney detection methods:

- the designed preprocessing components, including the MFLV speckle reduction, volume enhancement using LHE and G2LTh, and voxel classification with SANNs, overcome ultrasound challenges, and provide an enhanced volumetric image for robust and accurate kidney shape registration;

- the CVISM includes details of structural regions of the kidney shape, which reduces the probability of false-positive detection of none-kidney structures;
the new RCNCC similarity metric improves the accuracy of shape registration by considering structural regions of the kidney shape.

According to the results reported in Table 3.3, Fig. 3.19, and Fig. 3.20, the proposed CVRLS-SP kidney segmentation method outperforms the state-of-the-art. The better performance of the CVRLS-SP method attributes to the following characteristics of the proposed method:

- the level-set initialization by the fitted CVISM on the detected kidney shape in ultrasound volumes prevents the segmentation to trap in local answers;
- the proposed solution uses shape prior to regulate the segmentation result to be close to the kidney shape model, and prevents the segmentation to leak into non-kidney regions;
- the multi-regional segmentation method based on the definition of complex-valued regional level-set includes structural details of the kidney shape in the segmentation task, which enforces the level-set evolution to follow the kidney shape’s structure, and thereby a higher accuracy is achieved.

The improved segmentation result using the proposed CVRLS-SP method positively affects the performance of computerized free-fluid detection. A free fluid in Morison’s pouch view appears as a dark region placed in the interface of the kidney and liver. When a computerized algorithm searches for dark regions in an ultrasound volume of the Morison’s pouch view, it might detect multiple dark regions, caused by many different reasons such as shadows of ribs, blood vessels, and/or free-fluid. To correctly select the dark region corresponding to a free-fluid, its adjacency with the kidney shape should be taken into the account. Therefore, incorrect segmentation of the kidney shape may result in false-negative or false-positive decisions.

### 4.3 Limitations

The proposed shape-to-volume registration method has a high computational cost, because it is designed based on 3D convolutions. To reduce the computational time, we have applied two strategies as follows: (a) using parallel computation, and (b) reducing the image resolution to \(\text{res} = 0.25\). As a limitation, the resolution reduction may result in losing shape information, and may cause false decision. This problem might be addressed through time when more capable computations become affordable for medical applications.

The voxel classification based on SANNs requires a pre-processing step to find and register the kidney shape inside the ultrasound volume on the expected kidney shape alignment. This is performed
by applying some steps of the shape-based kidney detection method, including volume enhancement and
finding the best seed point, and the obtained seed point is used to transform the ultrasound volume to
register its kidney shape on the reference alignment. However, finding the best seed point is itself subject
to fail, and its failure results in an incorrect performance of the atlas-based kidney detection method.

The selection of the set of seed point, discussed in Sec 2.6.2, is influential in the performance of the
proposed shape-to-volume registration. The denser the seed points are selected, the better detection can
be achieved. However, increasing the number of seed point proportionally increases the computational
time of the proposed shape-to-volume registration.

The proposed CVRLS-SP kidney segmentation method uses an estimated mean of intensity level
for each of the multi-regional structures, $c_{br}$ and $c_{dr}$ (see Sec 3.3). However, the nature of ultrasound
imaging implies that similar tissue layers may have different intensity levels. This may result in reduced
kidney segmentation accuracy.

Shadows of ribs projected on the kidney shape partially occlude the kidney shape, and may result
in a kidney detection failure. A shadow appears as a dark ribbon, which partially obscures the kidney
shape in ultrasound images. Although the RCNCC metric is robust to the kidney shape occlusion to
some extend, we have observed reduced accuracy of estimating the kidney shape alignment and kidney
shape segmentation in a few number of ultrasound volumes.

4.4 Future work

As a continuation of this research, we can work on the proposed solution to increase its robustness and
to reduce its computational time. Some possible extensions to the proposed solution are listed below,

- The voxel classification relies on some steps of the shape-based kidney detection, which may nega-
tively affect the performance of the atlas-based method. A new process can be designed to perform
volume to volume registration of ultrasound volumes on a reference volume/shape, and the regis-
tered volume can be used for voxel classification. The conventional SIFT features are not robust
enough to be used in ultrasound volumes. Some recent studies have reported promising results
using new features [263]. Such a process can remove the dependency of the atlas-based method
from the shape-based approach.

- The existence of shadows in ultrasound images may reduce the accuracy and robustness of kidney
shape detection and segmentation. We have worked on a shadow detection and removal method.
As its primitive results are not convincing in all the actual ultrasound volumes of the dataset, it is
not included in this thesis. Further research is required to improve the shadow detection/removal operation, and thereby, the accuracy of the proposed solution is expected to increase.

- The low quality of the actual ultrasound volumes in the dataset is partially due to the use of the ultrasound imaging device which synthetically generates 3D ultrasound volumes. We believe by using a real 3D ultrasound device, the quality of ultrasound volumes will be increased, which would contribute in increasing the robustness and accuracy of the proposed solution.

- In patients with kidney injury and/or kidney diseases, such as kidney tumor, polycystic lesions, diabetes, and kidney stone, the kidney shape might be morphologically changed. As a future work, these problems can be taken into consideration to ensure correct kidney detection can be achieved for a wider range of patients.

- The computational time of the proposed CVRLS-SP kidney segmentation is high, because it uses sequential programming. We can investigate the utility of GPU programming to accelerate the segmentation process using the CVRLS-SP method.

- As another path for future work, we will attempt to incorporate localized regional information in the CVRLS-SP method to improve its robustness against inhomogeneous intensity profile.

- We will use the designed kidney segmentation to develop an automated free-fluid detection algorithm.

### 4.5 List of publications

The publications, including journal papers, conference papers, and a US-patent, during my PhD studies at the University of Toronto is listed below,

- **Journal papers:**

- **Conference papers:**


• US-Patents:

Bibliography


