AUTOMATIC BAYESIAN SEGMENTATION OF HUMAN FACIAL TISSUE
USING 3D MR-CT FUSION BY INCORPORATING MODELS OF
MEASUREMENT BLURRING, NOISE AND PARTIAL VOLUME

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES
OF
MIDDLE EAST TECHNICAL UNIVERSITY

BY

EMRE ŞENER

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR
THE DEGREE OF DOCTOR OF PHILOSOPHY
IN
ENGINEERING SCIENCES

SEPTEMBER 2012

Approval of the thesis:

AUTOMATIC BAYESIAN SEGMENTATION OF HUMAN FACIAL TISSUE USING 3D MR-CT FUSION BY INCORPORATING MODELS OF MEASUREMENT BLURRING, NOISE AND PARTIAL VOLUME

submitted by **Emre Şener** in partial fulfillment of the requirements for the degree of **Doctor of Philosophy in Engineering Sciences Department, Middle East Technical University** by,

Prof. Dr. Canan Özgen   _____
Dean, Graduate School of **Natural and Applied Sciences**

Prof. Dr. Murat Dicleli   _____
Head of Department, **Engineering Sciences Dept., METU**

Assoc. Prof. Dr. Utku Kânoğlu   _____
Supervisor, **Engineering Sciences Dept., METU**

Assoc. Prof. Dr. Ü. Erkan Mumcuoğlu   _____
Co-Supervisor, **Informatics Inst., METU**

Examining Committee Members:

Prof. Dr. Serdar Akyar   _____
Dahili Tıp Bilimleri / Radyodiagnostik Anabilim Dalı
Ankara Üniversitesi Tıp Fakültesi

Assoc. Prof. Dr. Utku Kânoğlu   _____
Engineering Sciences Dept., METU

Assoc. Prof. Dr. Hakan İ. Tarman   _____
Engineering Sciences Dept., METU

Assist. Prof. Dr. Senih Gürses   _____
Engineering Sciences Dept., METU

Assist. Prof. Dr. Didem Gökçay   _____
Medical Informatics, Informatics Inst., METU

Date: _____

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name, Last name : Emre Şener

Signature :

ABSTRACT

AUTOMATIC BAYESIAN SEGMENTATION OF HUMAN FACIAL TISSUE
USING 3D MR-CT FUSION BY INCORPORATING MODELS OF
MEASUREMENT BLURRING, NOISE AND PARTIAL VOLUME

Şener, Emre

Ph.D., Department of Engineering Sciences

Supervisor: Assoc. Prof. Dr. Utku Kânoğlu

Co-Supervisor: Assoc. Prof. Dr. Ü. Erkan Mumcuoğlu

September 2012, 102 pages

Segmentation of human head on medical images is an important process in a wide array of applications such as diagnosis, facial surgery planning, prosthesis design, and forensic identification. In this study, a new Bayesian method for segmentation of facial tissues is presented. Segmentation classes include muscle, bone, fat, air and skin. The method incorporates a model to account for image blurring during data acquisition, a prior helping to reduce noise as well as a partial volume model. Regularization based on isotropic and directional Markov Random Field priors are integrated to the algorithm and their effects on segmentation accuracy are investigated. The Bayesian model is solved iteratively yielding tissue class labels at every voxel of an image. Sub-methods as variations of the main method are generated by switching on/off a combination of the models. Testing of the sub-methods are performed on two patients using single modality three-dimensional (3D) images as well as registered multi-modal 3D images (Magnetic Resonance and Computerized Tomography). Numerical, visual and statistical analyses of the methods are conducted. Improved segmentation accuracy is obtained through the use of the proposed image models and multi-modal data. The

methods are also compared with the Level Set method and an adaptive Bayesian segmentation method proposed in a previous study.

Keywords: Medical Image Segmentation, Bayesian, Data Fusion, Partial Volume, Deblurring, Human Facial Tissue, Directional Prior, Magnetic Resonance Imaging, Computed Tomography

ÖZ

İNSAN YÜZ DOKULARININ BULANIKLAŞMA, GÜRÜLTÜ VE KİSMI
HACİM MODELLERİ İÇEREN BAYESÇİ 3B MR-BT GÖRÜNTÜ
BİRLEŞMESİ YÖNTEMİ KULLANILARAK OTOMATİK BÖLÜTLENMESİ

Şener, Emre

Doktora, Mühendislik Bilimleri Bölümü

Tez Yöneticisi: Doç. Dr. Utku Kânoğlu

Ortak Tez Yöneticisi: Doç. Dr. Ü. Erkan Mumcuoğlu

Eylül 2012, 102 pages

Tıbbi görüntülerde insan kafasının bölütlenmesi tıbbi teşhis, yüz ameliyatı planlaması, protez tasarımı ve adli tıpta kimlik teşhisi gibi geniş bir yelpazeye yayılan uygulamalarda önemli bir işlemdir. Bu çalışmada, yüzdeki dokuların bölütlenmesi için yeni bir Bayesçi yöntem sunulmaktadır. Kullanılacak bölütleme sınıfları kas, kemik, yağ, hava ve deri olarak belirlenmiştir. Yöntem, görüntü alımında oluşan görüntü bozulma modeli, gürültü azaltmaya yardımcı önsel bilginin yanı sıra, kısmi hacim modeli de içermektedir. Eşyönlü ve yönsel Markov Rasgele Alanı önsel bilgilerine dayalı bölütleme düzenlemesi yönteme dahil edilmiştir ve bölütleme doğruluğu üzerindeki etkileri araştırılmıştır. Bayesçi modeli görüntünün her vokselindeki doku sınıfını verecek şekilde yinelemeli olarak çözülmüştür. Model kombinasyonlarını açıp kapatarak ana yöntemin alt yöntemleri elde edilmiştir. Alt yöntemlerin denemesi tek modalitede üç-boyutlu (3B) görüntüler ve birbirine çakıştırılmış çoklu-modalitede 3B görüntülerde (Manyetik Rezonans ve Bilgisayarlı Tomografi) yapılmıştır. Yöntemlerin sayısal, görsel ve istatistiksel analizi sunulmuştur. Önerilen görüntü modelleriyle ve çoklu modalite veri kullanımıyla bölütleme doğruluğu iyileştirilmiştir. Ayrıca

yöntemler daha önceki bir çalışmada sunulan Level Set ve adaptif Bayesçi bölütlemeleriyle karşılaştırılmıştır.

Anahtar Kelimeler: Tıbbi Görüntü Bölütleme, Bayesçi, Veri Birleşmesi, Kısmi Hacim, Görüntü Netleştirme, İnsan Yüz Dokusu, Yönsel Önbilgi, Manyetik Rezonans Görüntüleme, Bilgisayarlı Tomografi

This thesis is dedicated to my mother and grandmother

for their love, endless support and patience.

ACKNOWLEDGEMENTS

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

2D	Two-dimensional
3D	Three-dimensional
ANOVA	Analysis of variance
ASSD	Average symmetric surface distance
CT	Computerized tomography
EM	Expectation maximization
FBP	Filtered back projection
FM	Finite mixture
ICM	Iterated conditional modes
HU	Hounsfield unit
MAP	Maximum *a-posteriori*
MR	Magnetic resonance
MRF	Markov random field
MRI	Magnetic resonance imaging
MSSD	Maximum symmetric surface distance
PV	Partial Volume
PVD	Partial Volume Difference
POCS	Projections onto convex sets
RMSSSD	Root mean square symmetric surface distance
ROI	Region of Interest

CHAPTER 1

INTRODUCTION

1.1   Motivation

Segmentation of medical images containing human head tissues is an active research area. An accurately obtained head model is useful in several applications. For example, the masticatory muscles are the biggest muscles in human face and they play an important role in chewing and facial expressions [1]. Simulation of facial expressions to aid in craniofacial surgery planning is another possible application and it was the aim of a study by Gladilin *et al.* [2] in which tissue deformations according to a biomechanical model are calculated by the finite element method. The mimic muscle model and an example of facial expression simulation obtained in [2] are shown in Figure 1.1. These techniques are also applicable to character animation which can be used in gaming and movie industries. Another simulation study fused motion captured animations and muscle activation signals in order to simulate speech [3]. Three-dimensional (3D) models used in [3] were obtained from a Magnetic Resonance (MR) scan and manually segmented. In another face modelling study, Zhang *et al.* [4] suggested a 3D facial biomechanical model to be used in expression animations. In this model, the skull structure was included to constrain the skin motion. The muscles were classified into groups such as linear and sheet-like.

Measurements on face tissues are important to diagnose, analyse and treat patients with neuromuscular disorders. Farrugia *et al.* [5] aimed to do measurements on

1

facial muscles such as tongue, pterygoids and masseter muscles which are smaller and harder to detect and are less accessible for procedures such as biopsy than limb muscles. Conditions such as facial and tongue muscle atrophy, fibrosis in fat can be observed better with these measurements. In patients with hemifacial microsomia which is a congenital disorder that affects the lower half of the face, calculating the muscle volumes is also a critical measurement and segmenting these muscles manually were found to be time consuming [6].

(a)   (b)
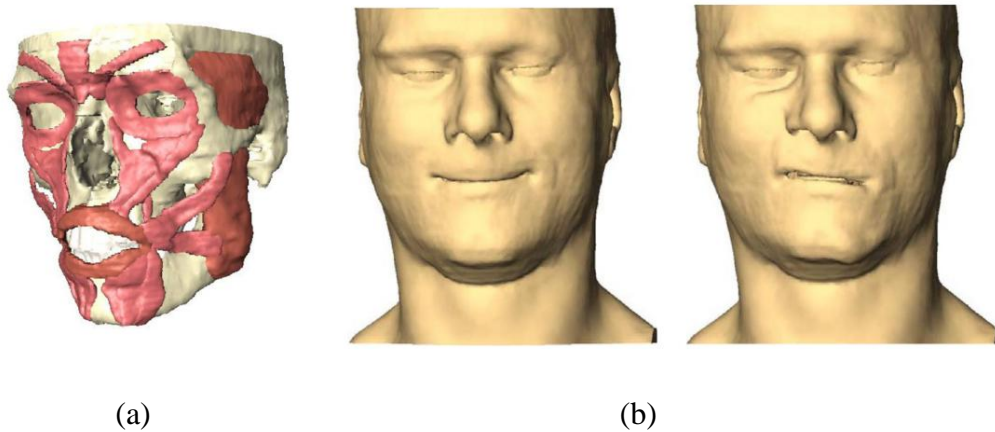
Figure 1.1 Facial expression simulation.

(a) Mimic muscle model and (b) sample face expression simulations (original image by [2])

Maxillofacial surgery is a type of surgery which focuses on treating defects and abnormalities in hard tissues of the face and jaws. The demand from surgeons for preoperative predictions on person specific models has led to studies such as [7]. Although Computerized Tomography (CT) or MR images are generally used to segment facial tissues, there have also been studies which calculate masseter muscle volume from ultrasonography and analyse its relationship with facial morphology [8]. According to these studies, stronger muscles tend to have a correlation with features of the short face syndrome such as parallelism of jaws. *Craniofacial Surgery Planner* software was the product of a study on simulation and prediction of craniofacial surgeries [9]. Meshes of the bones and soft-tissues were generated by a thresholding algorithm. The artefacts of the algorithm were

removed by eliminating small connected regions. The removal of these regions were performed with manual supervision.

Applications which use facial models are not only limited to medical. Facial reconstruction carries a lot of importance in forensics when there is little evidence to work with. De Greef and Willems [10] mentioned obtaining facial features such as nose projection, eye protrusion and mouth width being critical in identification of an unknown body. Another application in forensics or anthropology is parameterizing the properties of head tissues. Stephan [11] stated that the masseter muscle plays an important role in facial reconstruction processes because of its position and size. Stephan [11] also claimed that despite its importance, the predictions on the masseter muscle in facial reconstruction methods had been superficial.

The reconstruction of the head from medical images is found to also have a critical role in dental treatments. Zepa *et al.* [12] obtained masseter muscle, medial pterygoid muscle, mandible volumes and sizes from manually segmented CT images. Zepa *et al.* [12] found that there were correlations between the facial tissue sizes and dental patient groups.

The mentioned studies employ manual segmentation of volumes or at least slices. Manual segmentation of 3D images or several two-dimensional (2D) slices is a time consuming process. Automatic and semi-automatic algorithms help to reduce the time spent by the users. The motivation of this study is the lack of variety of automatic segmentation methods specialized on human face muscles. Therefore, it is aimed to develop an algorithm which employs the fusion of CT and MR images.

Magnetic Resonance Imaging (MRI) and CT were used mainly for anatomic imaging of human head. MRI has good soft-tissue contrast, but for bone and air segmentation, CT is certainly superior. CT has a better resolution than MRI and it

3

also provides some soft-tissue contrast. MR-CT image fusion has been attempted by a few studies such as [13] and [14], in which the fusion was in the form of extracting bone-air information from CT images and skin/soft-tissue information from MR images. However, considering that both modalities provide soft-tissue information and air-bone information (to a lesser extent by MRI), a fusion scheme where all the available information from both modalities is utilized for every tissue type is certainly desirable [15].

The Bayesian Markov Random Field segmentation model proposed in this study is an extension of the model described in [15], which was motivated by methods in [16]. In [15], Kale *et al.* developed Level Set and Bayesian methods by MR-CT image fusion which also included modelling of partial volume classes. A sample segmentation result taken from [15] is shown in Figure 1.2. Although Bayesian results were quite good and better than Level Set, there were some shortcomings: (i) Segmentation of very thin (one pixel thick) structures was achieved by using an ad hoc mean intensity shift adaptively for thin structure regions. A more systematic method is necessary by accounting for image resolution; (ii) the resolution difference between MR and CT images was not considered; (iii) correlated CT noise could not be integrated into the segmentation model with the partial volume model adapted; (iv) more sophisticated priors that could improve the segmentation results of very thin structures were not tested.

Here, methods are provided addressing all of the above issues in order to provide further improvements and make them more general. In that regard, a Bayesian segmentation method which includes models of image resolution, noise and partial volume is proposed. The method is developed further by including image fusion as well as uniform and directional priors.

4
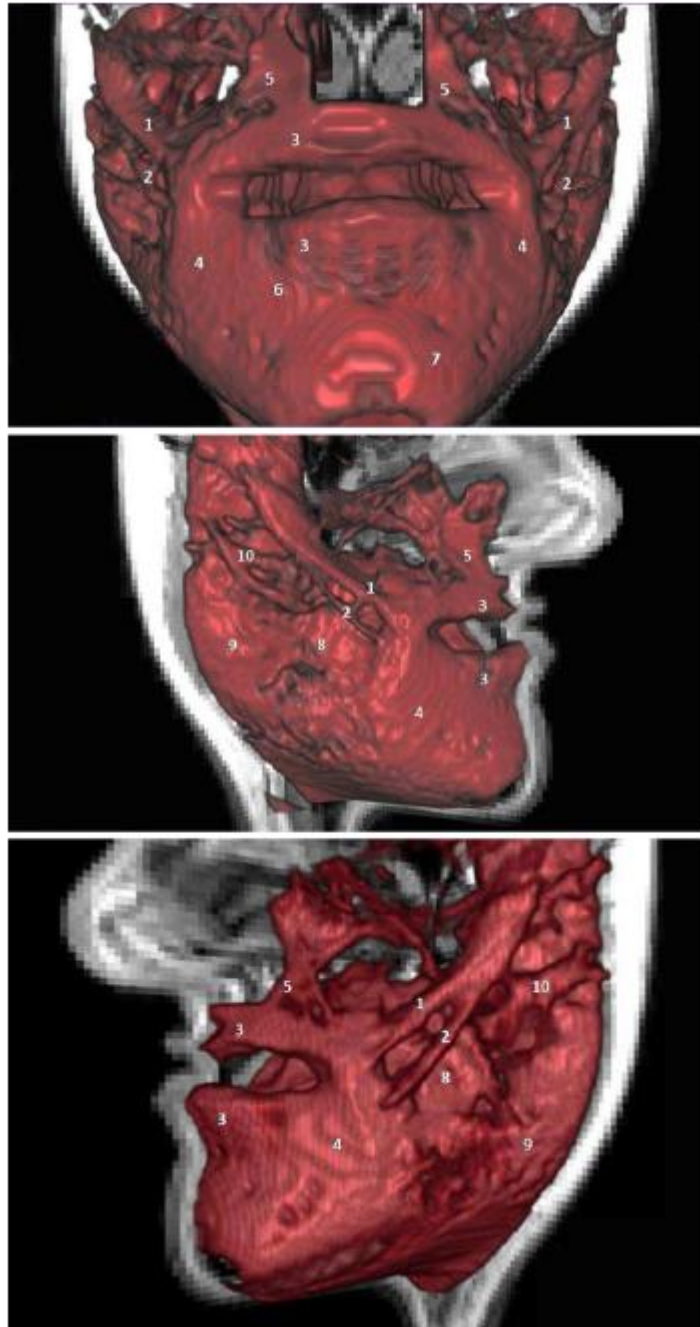
Figure 1.2 Volume rendering of segmented facial muscles.

Segmentation is performed by the Bayesian method proposed in [15] with muscle region indices which correspong to: (1) zygomaticus minor, (2) zygomaticus major, (3) orbicularis oris, (4) depressor anguli oris, (5) levator labii superioris, (6) depressor labii inferioris, (7) mentalis, (8) buccinators, (9) masseter, (10) ductus parotideus (original images by [15])

1.2 Related Work and Background

There are several types of classifications of automatic image segmentation algorithms. The methods are categorized according to various criteria in several image segmentation review and survey studies ([17], [18], [19]). The categories of the methods which were employed in human head segmentation are summarized in this section.

Thresholding techniques are some of the simplest labelling methods and used often in image segmentation. They have sub categories such as shape-based, clustering and spatial [20]. Thresholding has been used in a study combined with morphological operations in order to segment skull in MR images [21]. Brain and scalp were segmented first. Then, thresholding and closing/opening operations were performed on the remaining regions. The resulting segmentation was then masked with the prior brain and scalp segmentations.

There are also other methods such as watershed transform which is a transformation that models a flooding process on an image with morphological operations [22]. Region growing is another class of methods where the segmentation spreads from a seed point often selected by the user [23]. Marching cubes algorithm is similar to region growing but it is generally used to obtain isosurfaces. Olszewski *et al.* [13] used this method to segment bone, skin and other tissues from CT and MR images separately in order to use in aid of maxillofacial orthognathic surgery which is performed to correct facial bone conditions. According to this study, the surgery was to be planned by simulation of jaw movements using motion curves, inverse kinematics and collision detection. Although using single modalities or multiple modalities separately often yields reasonable segmentation accuracy, an extensive review on 3D image fusion processes in orthodontics and orthognathic surgery planning concludes that image fusion is the most accurate method for analysis [14]. The importance of orthognathic surgery simulation even led to studies which modelled the head by

using the data obtained by a laser scanner instead of a medical image [24]. The analysis of the tissue movements was made in [24] in order to plan the operative methods.

Clustering methods are often used in image segmentation. *k-means* and fuzzy *c-means* are among these clustering algorithms [25]. Banerjee *et al.* [26] used fuzzy *c-means* approach for multimodal segmentation. They formed features of voxels from average grey values in the voxel neighbourhood and busyness value which is affected by the orientation of the voxel. Segmentation was initially performed separately in CT and MR images and was assessed visually. A fusion operator was also defined in order to process the data from multimodal sources. The images were registered and fused prior to segmentation and fuzzy *c-means* was applied to obtain the tissue classes. Pham [27] suggested an algorithm which uses adaptive clustering according to edges. The algorithm simultaneously updates an edge field which adds robustness and accuracy.

Level Set is a method which is used in calculating and tracking curves and surfaces [28]. In a general segmentation algorithm that could be applied to medical image segmentation, Gibou and Fedkiw [29] suggested a hybrid *k-means*, level set and nonlinear diffusion algorithm. These methods were combined to achieve high computational speed of *k-means* and robustness of level set. The number of sub segmentations was adjusted by diffusion pre-processing parameters. Another approach to segmentation problems is solution by active contours and snakes [30]. These are models that find the contours in an image by minimizing an energy function which tries to snap to boundaries while preserving a regularized shape. In one of the studies in which this approach was used, Yezzi *et al.* [31] improved the curve evolution with binary and ternary flows by introducing constraints. Although the contour methods often use gradients to find boundaries, [31] allowed the capturing of these boundaries without very high gradients.

Variational models form another category of segmentation algorithms. The following are some of the studies on segmentation using variational principles. Samson *et al.* suggested two variational methods that use contour and region information which integrated image restoration into classification [32]. The methods included an edge-preserving regularization function and were applied on multispectral images with supervision. Holtzman Gazit *et al.* [33] aimed to segment thin structures in medical images by combining Chan-Vese minimal variance and geodesic active surface model. They used three terms which represent the edge information, minimal variance and active surface energy function. The solution was obtained by level-set method and a hierarchical approach was integrated to the method in order to segment multiple objects.

There are quite number of studies on semi-automatic segmentation tools which decrease the user interaction time significantly. Yushkevich *et al.* [34] developed the software *ITK Snap* based on active contour models which uses Level Set method. The method is based on seed points which were input by the user. Weights of active contour energy terms were also supplied. The software was deemed to be capable of full manual segmentation. Kan *et al.* [35] proposed a variety of 3D editing tools for image segmentation. The tools included hole-filling with 3D morphological closing with line elements, point bridging, surface dragging. A quantitative analysis of these tools with respect to intra-operator and inter-operator use was supplied. One of the studies on semi-automatic methods [36] aims to obtain 3D appearance of the lip muscles from MRI using the software *3D Slicer* which is a framework for image visualisation and processing [37]. This segmentation result is then to be used to help in the diagnosis of facial nerve palsy and facial expression restoring treatment planning. Live-wire is a widely used tool in image editing software. Barrett *et al.* [38] proposed a variation of live-wire to be used in segmentation of 2D medical images. The proposed live-wire variation tracks the boundaries of regions using recent user input in addition to gradient information. Here, a similar approach as in [39] that required minimal user input

by adaptively transforming the user provided seed points to the neighbouring slices with success.

The term *regularization* usually points to an extra term or step used in segmentation algorithms which introduces a spatial relationship between voxels. This term/step regularizes the classification such that voxels close to each other affect the labelling of each other. This effect may be through the use of the labels in the previous iteration step, gradients, edges, the labels or other features of the image. The spatial relationships between neighbouring voxels usually carry important information in image segmentation. Therefore, taking these relationships into account is critical in the accuracy of a method. This claim is backed up by several studies which use the neighbourhood information. One of such methods is the inclusion of Markov Random Fields (MRF) in the segmentation formulation [40]. MRF is a model used for adding spatial dependency and regularization. Chiverton *et al.* [41] suggested using an adaptive MRF regularization in order to solve the partial volume (PV) problem in a fully automatic algorithm, and integrated a regularization scheme which uses the gradients for locally adaptive neighbourhood continuity into the Bayesian formulation. A clustering algorithm was then implemented to calculate parameters such as tissue class means in MR data. Rezaeitabar and Ulusoy [42] suggested using a region growing algorithm which employed a MRF model and works with manually placed seed points. MRF modelled spatial relationships between a voxel and its neighbours as well as the voxels at the same location in other MR images which are registered to each other. The registered images were used for training after manual segmentation of a reference image. Region growing was performed on a slice and the seed point was transferred to other slices bounds of which were decided by the user.

Statistical and probabilistic models find themselves a place in image segmentation. Bayesian and finite mixture (FM) are among the primary examples of such models. The equations to obtain the parameters of the model cannot be

solved directly in most practical situations. Expectation maximization algorithm is one of the methods which can be used to calculate the parameters in those cases. In a study by Zhang *et al.* [43], FM model is used for segmenting MR head images. A hidden MRF with Gaussian distribution and expectation maximization (EM) algorithm were integrated. The model parameters were obtained by EM and labelling was done by applying Iterated Conditional Modes (ICM) to the maximum *a-posteriori* (MAP) criterion. The method was also modified to include bias correction. Another statistical approach was suggested in [44] to segment the head into nine tissue types such as brain tissues, muscle, fat, eye, etc. Pre-processing was performed on the MR image to separate brain, bone and scalp. EM was used to segment brain tissues. The remaining classes were segmented using spatial and statistical prior and applying filling and morphological operations. Jiang *et al.* [45] proposed a method for deblurring of spiral CT images. The method maximizes edge-to-noise ratio by quantifying the noise effect using a discrepancy measure and EM algorithm and is primarily focused on improving cochlea features on images.

1.3 Bayesian Segmentation and Markov Random Field Priors

Statistical models which originate from Bayes' rule have been employed in several studies in image segmentation. Laidlaw *et al.* [46] uses a Bayesian approach and accounts in the PV effects. They assigned vector features to voxels and obtained tissue mixtures by fitting multi-dimensional histograms. These histograms helped in preventing the co-linearity of means in one-dimensional histograms. Hurn *et al.* [47] focused on hierarchical Bayesian formulation for fusion of multimodal images for segmentation. MAP estimates were used and the estimations were done by the ICM method. The method was tested on synthetic and real data for CT and high-resolution/low-resolution MR image fusion.

Discretization of a continuous anatomical region by a voxelated volume usually results in some voxels, especially the ones close to tissue boundaries, which contain information about multiple tissues. These voxels are deemed to contain partial volumes and their segmentation labels are referred as partial volume classes. Shattuck *et al.* [16] proposed a method which includes a PV model to segment MR head images. The MR image was subject to bias correction and skull removal pre-processes. A MAP classifier with a Gibbs prior was used to calculate the voxel labels. The probability density function for voxels which contain partial volume information was calculated by marginalization of the PV ratio of tissue types.

The study by Ruan *et al.* [48] modelled the mixture tissue classes as a Gaussian function and applied a MAP estimator. MRF was also included to ensure spatial coherence. The MRF model had an additional component based on the shape of the local intensity pattern. The voxels of the image were first classified into pure and mixed tissue types and a second step was executed to reclassify the mixed classes into pure classes. Efforts to integrate segmentation and registration in an MRF framework by MAP estimation were also made [49]. The proposed method used joint class histograms similar to the joint intensity histograms in mutual information theory. Non-rigid registration was also implemented but it was found to be ineffective in this particular case. The algorithm was applied at multiple levels and voxels at the same coordinate were granted parent-child relationship with each other through the hierarchy levels as well as spatial relationship with the neighbour voxels at the same level in the Markov prior. ICM was used to obtain the segmentation labels which minimize the likelihood model based on Gaussian mixture model. Van Leemput *et al.* [50] studied on a framework in which voxels were initially classified into pure tissue types and performed PV classification on the down-sampled version of the pure class segmentation. This was done by simultaneously approximating model parameters during classification.

There are variations of MRF priors that depend on the location and shape of the region to be segmented. Priors favouring especially thin structures have been suggested in multiple studies. Most of these priors use a tensor to enhance the effects of the orientations in the image. Wong *et al.* [51] used such MRF priors in conjunction with Bayesian segmentation for segmentation of tubular structures. The method estimates local structures with an orientation tensor constructed by a quadrature filter and give more weight to the direction of orientations in the MRF prior. Descoteaux *et al.* [52] worked on the segmentation of pituitary gland and thin sinus bones to aid in surgery simulations. The sinus bones are thin sheet-like structures. Therefore, the Hessian matrix at each voxel is used to determine the sheetness measure. A geometric flow model is then applied to obtain the thickness of the sheet structure. In the study by Frangi *et al.* [53], a vessel enhancement filter was proposed that is constructed from Hessian matrix eigenvalues. Different measures were calculated according to the eigenvalues and second order ellipsoid modelling of vessels. The vesselness (or tube-likeness) of each voxel was calculated at multiple scales. Another multi-scale study [54] focused on adaptive orientation selection based on Hessian matrix eigenvalues applied on Gaussian smoothed images. The method was tested on thin structures such as brain vessels on MR images and liver vessels on CT images.

The usage of Hessian matrix and quadratic filters was employed in several more studies. Lorenz *et al.* [55] segmented line-like structures in 2D and 3D by calculating the measure of similarity of a line structure from eigenvalues of the Hessian matrix. The centreline was extracted by using this method and body of the structure is obtained by applying active contours. Westin *et al.* [56] classified the local structures using 3D quadratic filters. The filters were obtained from orientation tensors and voxels in the image were categorized as belonging to planar, linear or isotropic structures. Adaptive thresholding was applied depending on the local structure type.

1.4 Aims and Scope

Here, it is aimed to segment the facial tissues with an automatic method. A Bayesian model is selected for this purpose because it was concluded that the multimodal image fusion, noise modelling and neighbourhood relationships can be integrated well within that model. The formulated Bayesian rule includes deblurring (for each modality, using as input the resolution sigma of each system), denoising, multi-modal image information and MRF prior components. The optimal labelling is obtained with a MAP classifier. The problem in this state becomes similar to an image restoration problem. The solution is found iteratively through substitution of the restored intensities by label mean intensities in the formulation. The set of voxel labels which minimize the model function is taken as the segmentation result. Partial volume effect is added by introducing mixed classes to the iterations. The results are analysed over multiple regions of interest on two subjects. Pre-processes such as intensity correction and registration are applied as well as post-processes like connected components, local morphological opening/closing. The effects of isotropic and adaptive directional priors are also investigated.

Although the results of the previous study [15] which used MRF prior, multi modality fusion and partial volume modelling were very promising, it has some drawbacks. It required some parameter tuning, specifically, mean tissue intensity shift for very thin structure regions, difference in resolution of MRI and CT was not modelled, CT image noise was not modelled properly, and potentially more useful other priors were not tested. In this study, significant model extensions of the mentioned Bayesian method were investigated: (i) resolution deblurring model for MRI and CT which also eliminates system or scan setting dependant parameter tuning; (ii) more realistic CT image noise model (correlated); (iii) an adaptive directional prior particularly targeted for very thin structures. The first two extensions required a different type of partial volume model which was also investigated.

The human head contains tissues such as muscles, skin, tendons, fat, bones, marrow, teeth, vessels, veins, facia, nerves, glands and ducts. Groups of these tissues have similar densities/intensities on MR and CT images. One of these groups includes muscles, skin, vessels, nerves, glands, ducts and tendons. This group of tissues are gathered under the "muscle" class. The only exception to this grouping is the skin. Since it covers the head, this anatomical information is used to segment the skin with a basic method and separate it from the other muscle-like tissues. Bones, teeth and skull form the "bone" class. Although marrow has a different intensity than bone, it is contained in the bones and is also included in the bone class. Air and fat are easier to distinguish from these groups and each other. Therefore, they are assigned their own classes. With these groups, the main tissue classes can be listed as: muscle, bone, fat, air and skin. The skin is the special case, it being classified initially as muscle and segmented as skin later. Segmentation of other tissues in the head such as eyes is beyond the scope of this study.

CHAPTER 2

METHOD

2.1 Human Data for Algorithm Test

The CT images used in this study were acquired by using standard protocols from patients who were already required to have a full head CT scan as part of their diagnosis and treatment planning. The CT scans were performed by using standard CT protocols. The MR images were then acquired. The imaging protocols have been approved by the institutional ethics committee, and the subjects gave their informed consent prior to the MRI scans.

The CT and MR scans were taken from two subjects. First subject was a 32 year-old female in prosthetic dental treatment. The axial CT slices of her scan spanned from orbita inferior to maxillar and mandibular region. The scanning of this region is regularly used in this kind of treatment. The other patient was a 21 year-old male with chronic headache and suspected sinus pathology. Paranazal sinus CT protocol was applied and the scan covered the head region which starts at upper frontal sinus and ends at mandibular region. During MR imaging, full heads of the subjects were scanned.

2.1.1 Image Parameters

The CT images were acquired by Philips mx8000IDT (Philips Medical Systems, Best, Netherlands) which is a 16 channel multi-detector computerized tomography device. The settings used when obtaining the images were:

- Dose parameters of 120 kV and 221 mA-s,
- Rotation of 0.5 s,
- Collimation of 16x0.75,
- Slice thickness of 1.5 mm,
- Reconstructed slice thickness 0.75 mm,
- Field of view (FOV) of 250 mm,
- Matrix size 512x512,
- and reconstructed voxel size 0.49x0.49x0.75 mm, in x, y and z directions respectively.

The MR images were acquired by a Philips Intera 1.5 Tesla device (Philips Medical Systems, Best, Netherlands) and head coil. T1-weighted protocols were successfully used previously [1], [13] and [15]. The T1-weighted (3D, Fast-Field Echo) protocol parameters were:

- TR of 25 ms,
- TE of 4.6 ms (in-phase),
- Flip angle of 30 degrees,
- Field of view of 240 mm,
- Matrix size of 256x256,
- Measured voxel size of 0.94x0.94x1.88 mm (in x, y and z directions),
- Reconstructed voxel size of 0.94x0.94x0.94 mm (in x, y and z directions),
- Transverse slice orientation,
- and 15 minutes of scan duration.

2.1.2 Regions of Interest

A total of sixteen regions of interest (ROI) were used to test the segmentation method. Eleven of them belong to the first patient and 5 belong to the second. The ROIs were formed by cropping 5 consecutive slices with an approximate size of 40x40 pixels (corresponding to a 37x37mm area). The regions were mostly selected so that they include thin facial muscles although some regions include larger tissues for testing. 16 regions and the tissues they contain are given in Table 2.1. The ground truth segmentation of these ROIs are performed manually by a physician who is an expert in radiology. ITK Snap [34] is used for manual segmentation.

Table 2.1 ROIs and the tissues they contain

ROI Index	Subject	Contained Muscles/Tissues
1	1	levator labii superioris, buccinators, zygomaticus major/minor muscles
2	1	masseter, buccinator, zygomaticus major/minor muscles, ductus parotideus (partial), facial blood vessels
3	1	masseter, buccinator, zygomaticus major/minor muscles, ductus parotideus (majority), facial blood vessels
4	1	buccinator, zygomaticus major/minor muscles, ductus parotideus (proximal segment), facial blood vessels
5-11	1	masseter, zygomaticus major/minor, orbicularis oculi muscles, facial blood vessels
12-13	2	masseter, buccinator, zygomaticus major/minor muscles, facial blood vessels, maxillary bone
14-16	2	maxillary bone, maxillary sinus, zygomatic bone, masseter, zygomaticus major/minor (proximal segment), orbicularis oculi muscles, facial blood vessels

2.1.3   Pre-processing

The intensities in the MR images have spatial variations or "a bias field" because of the magnetic field non-uniformity. The multiplicative bias value at each voxel must be known to correct the intensities. There were many studies which propose methods to calculate and correct the field such as [16] and [57]. The method presented in [57] is known as N3. N3 was proposed as an iterative approach to estimate the bias field and tissue intensities.

In order to apply the proposed segmentation method in this study, the CT and MR images registered to each other after bias field correction. The transformation which brings the images to the same coordinate system was applied to the CT image so that both images are at equal resolutions. The details of the bias correction and rigid/non-rigid registration methods are explained in Appendix A and Appendix B.

2.2   Bayesian Segmentation

In Bayesian model of segmentation, images are assumed to be random variables. The intensities of the pixels in a 3D image, $\vec{I}$, are written in vector form of size $M \times 1$ where M is the number of pixels. Similarly, the segmentation class labels are denoted by a vector, $\vec{\theta}$. The Bayes' rule according to this model can be written as

$$p\left(\vec{\theta} / \vec{I}\right) = \frac{p(\vec{I} / \vec{\theta}) p(\vec{\theta})}{p(\vec{I})} . \qquad\qquad 2.1$$

$p(\vec{\theta} / \vec{I})$ and $p(\vec{I} / \vec{\theta})$ are called posterior probability and likelihood respectively. $p(\vec{\theta})$ is the term that contains the prior information on class labels. The main goal of segmentation is maximizing the posterior probability. Maximizing the

posterior probability $p(\vec{\theta}/\vec{I})$ is called MAP estimation and it differs from maximizing the likelihood term $p(\vec{I}/\vec{\theta})$ with maximum likelihood (ML) by the inclusion of the prior term. Maximizing an expression is equivalent to maximizing its logarithm. Therefore, in order to simplify Equation 2.1, the logarithms of both sides are taken. The equation then becomes:

$$\arg\max_{\vec{\theta}} \log p(\vec{\theta}/\vec{I}) = \arg\max_{\vec{\theta}} \left[\log p(\vec{I}/\vec{\theta}) + \log p(\vec{\theta}) - \log p(\vec{I}) \right]. \qquad 2.2$$

$p(\vec{I})$ is independent of $\vec{\theta}$ and can be ignored for the maximization. In the case of Gaussian distribution, the likelihood can be written as

$$p(\vec{I}/\vec{\theta}) = g(\vec{I}, \vec{\lambda}, \Sigma) = \frac{1}{\sqrt{(2\pi)^k |\Sigma|}} e^{-\frac{1}{2}(\vec{I}-\vec{\lambda})^T \Sigma^{-1}(\vec{I}-\vec{\lambda})}, \qquad 2.3$$

where $\vec{\lambda}$ is the vector of mean values of the segmentation classes and it is of size M. Σ is the covariance matrix of $\vec{I}$. If constants in Equation 2.3 are ignored and a Gibbs [58] prior is used, maximizing posterior probability is equivalent to minimizing the following function

$$f = (\vec{I}-\vec{\lambda})^T \Sigma^{-1}(\vec{I}-\vec{\lambda}) + \beta \sum_{i=1}^{M} C(I_i), \qquad 2.4$$

where i are pairs of voxels in the neighbourhood of a voxel, namely cliques. The cliques are summed in the neighbourhood of a voxel to obtain the Gibbs prior of that voxel. The second term is the sum of the Gibbs priors which contains prior neighbourhood potentials of the cliques of all voxels. β is the weight of the prior term. Basic Bayesian segmentation is obtained through the optimization of the cost function in Equation 2.4. Modifications and additions to the basic formulation will be explained in the following sections.

2.3   Restoration Based Segmentation Model

Imaging systems may introduce distortion to their output in forms of transformations, blurring and/or noise. The process which is used to model and rectify the distortion is called image restoration [59]. Image restoration problem is defined as obtaining a high-resolution image from N lower resolution images. In the segmentation context and with the images being written in vector form, the high-resolution image is the vector of class means, $\vec{\lambda}$, and the low-resolution image is the image to be segmented, $\vec{I}$. In image restoration the low-resolution images originate from $\vec{\lambda}$ and they are the results of geometric transformation, blurring, decimation and noise addition on $\vec{\lambda}$. The relation between $\vec{\lambda}$ and $\vec{I}_k$ can be written as

$$\vec{I}_k = D_k B_k T_k \vec{\lambda} + \vec{E}_k \qquad (k = 1, .., N)$$

2.5

where T_k, B_k, D_k and $\vec{E}_k$ are transformation, blurring, decimation matrices and additive noise respectively [59]. Combining these matrices into a single matrix as $A_k = D_k B_k T_k$ yields

$$\vec{I}_k = A_k \vec{\lambda} + \vec{E}_k .$$

2.6

Here, in transforming the image restoration problem to a segmentation problem, three assumptions are made. (i) The size of $\vec{\lambda}$ in each dimension is assumed to be equal to the sizes of $\vec{I}_k$ since the aim of segmentation methods is usually finding one label for each pixel. (ii) The pixel labels are required to be in the same coordinate system with the image. Therefore, the transformation T is identity. The first two assumptions leave only the effects of the blurring matrix B_k in A_k. The Bayesian segmentation model given in Equation 2.4 already has an integrated noise model. Therefore, $\vec{E}_k$ is also ignored. (iii) It is assumed that the original

image $\vec{\lambda}$ is composed of pixels with intensities equal to the mean intensity values of the tissue type that the pixels contain. Therefore, the elements of $\vec{\lambda}$ are allowed only to have the values of mean intensities of the class labels μ_k. The class labels are chosen from the label set $\mu = \{\mu_{air}, \mu_{bone}, \mu_{fat}, \mu_{muscle}\}$.

Since only the effects of blurring remain in A_k, that effect is integrated into the Bayesian segmentation model by substituting $\vec{\lambda}$ with $A_k \vec{\lambda}$ in Equation 2.4. The minimization of this equation with respect to $\vec{\lambda}$ gives the segmentation solution with MAP as

$$\vec{\lambda}_{MAP} = \arg \min_{\vec{\lambda}} [(\vec{I} - A\vec{\lambda})^T W^{-1} (\vec{I} - A\vec{\lambda}) + C(\vec{\lambda})]. \qquad 2.7$$

Here W is the autocorrelation of the noise and $C(\vec{\lambda})$ is the prior potentials in the Gibbs prior.

As an example of single image input, Equation 2.4 can be rewritten as a function of segmentation labels for the MR image as

$$f(\vec{\lambda}) = (\vec{I}_{MR} - A_{MR}\vec{\lambda})^T W_{MR}^{-1} (\vec{I}_{MR} - A_{MR}\vec{\lambda}) + C(\vec{\lambda}). \qquad 2.8$$

For notation convenience, the likelihood and prior terms of the posterior probability function will be separated as

$$f(\vec{\lambda}) = u(\vec{\lambda}) + C(\vec{\lambda}). \qquad 2.9$$

The matrices and vectors in $u(\vec{\lambda})$ can be shown in a more explicit form as follows

$$A_{MR} = \begin{bmatrix} A_{11} & \cdots & A_{1M} \\ \vdots & \vdots & \vdots \\ A_{M1} & \cdots & A_{MM} \end{bmatrix}, \quad W_{MR} = \begin{bmatrix} W_{11} & \cdots & W_{1M} \\ \vdots & \vdots & \vdots \\ W_{M1} & \cdots & W_{MM} \end{bmatrix}, \qquad 2.10$$

$$\vec{I}_{MR} = \begin{bmatrix} i_1 \\ \vdots \\ i_M \end{bmatrix} \text{ and } \vec{\lambda} = \begin{bmatrix} \lambda_1 \\ \vdots \\ \lambda_M \end{bmatrix},$$

where M is the number of pixels in the MR image and $\lambda_j \in \mu$ is the mean intensity of the selected label for pixel j. Any combination of λ_j defines a segmentation of the whole image. The chosen segmentation is the set of labels that minimize $u(\vec{\lambda})$.

The CT and MR information are used together by introducing the data and parameters of a second image in the formulation. This can be achieved by registration of the images, which is explained in Appendix B. Ideally, the images should be kept in their original resolutions and the transformation matrix obtained by registration should be included in the formulation. However, in this study it is assumed that the CT and MR images have the same height and width in terms of pixel numbers. This is achieved by using the transformation applied version of the registered image. The size of the output label set is also set to the same height and width of the CT and MR images. This leaves matrix A to only include the effects of blurring. For CT-MR fusion, the image vector $\vec{I}$ is obtained by concatenation of the vectors of two images as

$$\vec{I} = \begin{bmatrix} \vec{I}_{MR} \\ \vec{I}_{CT} \end{bmatrix}. \qquad\qquad 2.11$$

With the concatenation, the likelihood term becomes a multivariate Gaussian. However, the CT and MR images are uncorrelated. Therefore, the covariance matrix of the multivariate Gaussian is block-diagonal. Then, the segmentation result obtained from the extended log-likelihood (Equation 2.4) with the multivariate Gaussian is simplified to

$$\vec{\lambda}_{MAP} = \arg\min_{\vec{\lambda}}[(\vec{I}_{MR} - A_{MR}\vec{\lambda})^T W_{MR}^{-1}(\vec{I}_{MR} - A_{MR}\vec{\lambda})$$

$$+ (\vec{I}_{CT} - A_{CT}\vec{\lambda})^T W_{CT}^{-1}(\vec{I}_{CT} - A_{CT}\vec{\lambda}) + C(\vec{\lambda})].$$

2.12

This becomes an optimization problem with M parameters which are the labels of the M pixels in the image. The optimization is performed by using the Iterated Conditional Modes (ICM) [60]. It is a deterministic algorithm which aims to do local optimizations sequentially until convergence. The ICM algorithm updates the label of a pixel at a time according to the minimization of the cost function and goes through all pixels in the image. The process continues until the iteration in which only a small portion of pixels change labels.

2.3.1 Measurement Noise Model and Autocorrelation Matrix

The autocorrelation matrices used in the restoration formulation are related to the noise models of the images. The noise structure in MR images can be assumed as independent noise [16]. The studies on the measurement noise of the detector bins in the CT devices show that there is no correlation for the data noise [61]. However, this is valid only for the raw data. Since access to raw data is not always possible, the noise model of the image is needed. The CT images are constructed using a method called Filtered Back Projection (FBP). The method is based on projecting sensor measurements onto an image and involves summation of the back projected values when calculating pixel intensity values. Hence, noise model of the reconstructed CT image is correlated [62].

In order to calculate the autocorrelation matrix for a CT image, a homogeneous region in the image is selected as shown in Figure 2.1. The autocorrelation for a horizontal shift q and vertical shift r at voxel coordinate (x, y) is calculated by

$$R_{pq} = \sum_{x=1}^{m}\sum_{y=1}^{n}[I(x, y) - \mu][I(x+q, y+r) - \mu],$$

2.13

23

where m and n are the width and height of the selected region respectively, and μ is the estimated mean within the region.

Figure 2.1 A selected homogeneous region to calculate autocorrelation

When the intensity values of the pixels of the image are scaled to the interval [0, 1], the autocorrelations for different q, r values are $R_{00} = 0.0467$, $R_{10} = 0.0032$, $R_{20} = 0.0012$ and $R_{30} = 0.0003$ for the selected region in Figure 2.1. The magnitude continues to decline rapidly for p and/or q values that are greater than 3. The autocorrelation is also calculated along the lines above and below the horizontal center and in the y-direction (vertical). The values obtained can be represented in spatial coordinates as follows:

$$R = \begin{bmatrix} R_{22} & R_{12} & R_{02} & R_{12} & R_{22} \\ R_{21} & R_{11} & R_{01} & R_{11} & R_{21} \\ R_{20} & R_{10} & R_{00} & R_{10} & R_{20} \\ R_{21} & R_{11} & R_{01} & R_{11} & R_{21} \\ R_{22} & R_{12} & R_{02} & R_{12} & R_{22} \end{bmatrix}$$

2.14

$$= \begin{bmatrix} 0 & 0 & 0.0022 & 0 & 0 \\ 0 & 0.0061 & 0.0089 & 0.0061 & 0 \\ 0.0012 & 0.0032 & 0.0467 & 0.0032 & 0.0012 \\ 0 & 0.0061 & 0.0089 & 0.0061 & 0 \\ 0 & 0 & 0.0022 & 0 & 0 \end{bmatrix}.$$

The zero values in the matrix and the values for $q \geq 3$ or $r \geq 3$ which remain outside the matrix are found to be negligible. Next, the autocorrelation matrix is constructed using the values of the R matrix. In raster scanned pixel order representation, the autocorrelation matrix (for an R matrix of zero values for $q, r \geq 2$) becomes a banded matrix:

$$W = \begin{bmatrix} R_{00} & R_{10} & & & & R_{11} & R_{01} & R_{11} & \\ R_{10} & R_{00} & R_{10} & & & R_{11} & R_{01} & R_{11} & \\ & \ddots & \ddots & \ddots & & & \ddots & \ddots & \ddots \\ & & & & & & & & \\ & & & & & & & & \\ \ddots & \ddots & \ddots & & & & \ddots & \ddots & \ddots \\ & R_{11} & R_{01} & R_{11} & & & R_{10} & R_{00} & R_{10} \\ & & R_{11} & R_{01} & R_{11} & & & R_{10} & R_{00} \end{bmatrix}.$$

2.15

The matrix W should be modified so that the values corresponding to spatially unrelated pixel pairs are set to zero. Although the pixel at the end of a line and the pixel at the start of a next line have consecutive indices, they are on separate parts of an image. This raster scan format of image pixels in vector format and the neighbourhood relationships are described in detail in section 2.4.2.

The autocorrelation matrices W are obtained with the noise parameters which are used in creating the artificial data. The matrix is 3-banded for correlated (coloured) CT noise and diagonal for independent MR noise case.

2.3.2 Image Blurring Estimation and Matrix

The smoothing matrix A is formed in raster scan pixel representation from the kernels with which the artificial images are smoothed. This formation is similar to the formation of autocorrelation matrices. The σ values are found as $\sigma_{CT} = 0.8$ pixels (0.75mm) and $\sigma_{MR} = 1.3$ pixels (1.22mm) with the assumption of isotropic Gaussian blurring. These are approximate values experimentally measured from tissue boundaries in CT and MR images (such as bone-muscle). The 5x5 kernels corresponding to these σ values are calculated as

$$A_{CT} = \begin{bmatrix} 0 & 0.05 & 0.011 & 0.05 & 0 \\ 0.05 & 0.052 & 0.114 & 0.052 & 0.05 \\ 0.011 & 0.114 & 0.252 & 0.114 & 0.011 \\ 0.05 & 0.052 & 0.114 & 0.052 & 0.05 \\ 0 & 0.05 & 0.011 & 0.05 & 0 \end{bmatrix}$$

$$A_{MR} = \begin{bmatrix} 0.009 & 0.024 & 0.032 & 0.024 & 0.009 \\ 0.024 & 0.058 & 0.077 & 0.058 & 0.024 \\ 0.032 & 0.077 & 0.104 & 0.077 & 0.032 \\ 0.024 & 0.058 & 0.077 & 0.058 & 0.024 \\ 0.009 & 0.024 & 0.032 & 0.024 & 0.009 \end{bmatrix}$$

2.16

2.3.3 Partial Volume

The intensity values of partial volume voxels can be calculated by linear interpolation of intensities of the pure tissue classes with respect to the ratio that they occupy when the effects of blurring due to system resolution, noise and other

modifications caused by the sensors are ignored. The PV model presented by [16] proposed the marginalization of the partial volume classes in the minimization function. However, the marginalization of the minimization function is not possible in the case of correlated noise and blurring model that we considered here. Therefore, it was decided to model the partial classes by introducing a finite set of PV classes with mean values that fall between pure classes. The mean values are calculated by

$$\mu_{AB} = \alpha\mu_A + (1-\alpha)\mu_B,$$ 2.17

where A and B are pure tissue classes, α is the fraction of the voxel that is occupied by tissue class A $(0 \le \alpha \le 1)$.

In order to decide on the specific partial volume classes to be used, the mean intensities of the pure tissue classes in the CT and MR images are plotted (Figure 2.2). In the cases where only the CT or MR image is used, some partial volume classes may cause potential problems. For example, using a bone-fat partial volume class will result in overlapping intensities with a muscle-fat class since muscle intensity is between bone and fat intensities in both modalities. Therefore, partial volumes (for single modality cases) are only defined for pure classes that have adjacent intensity intervals (e.g. air-bone, bone-muscle, muscle-fat PV classes in MR).

The intensity plot for the CT-MR fusion case (Figure 2.3) yields a 2D graph where potential overlaps caused by partial classes are at a minimum. Here, as an implementation, more partial volume classes are used between fat-muscle than other mixed classes since those are the pure classes which have the longest boundary between themselves. In order to balance the accuracy and computation complexity, using 2 partial volume classes between most combinations of pure tissue classes is found to be ideal. Only between fat-muscle pure classes, 6 partial classes are used. Partial volume classes of some pure class combinations such as

27

air-muscle are not used because they do not have a common boundary in an ideal image. All classes are listed in Table 2.2. Note that in the ICM update of the segmentation, each pixel is updated by the class which gives the minimum cost function value. Since the cost for every class is calculated, the computation times are directly proportional to the number of classes used.

Table 2.2 Number of pure and partial volume classes used for single image and fusion cases

Tissue Class	Class Count
MR or CT (Single image)	
Pure classes	4
Bone-muscle	2
Muscle-fat	6
Total	12
MR+CT (Fusion)	
Pure classes	4
Bone-muscle	2
Muscle-fat	6
Air-fat	2
Air-bone	2
Total	16
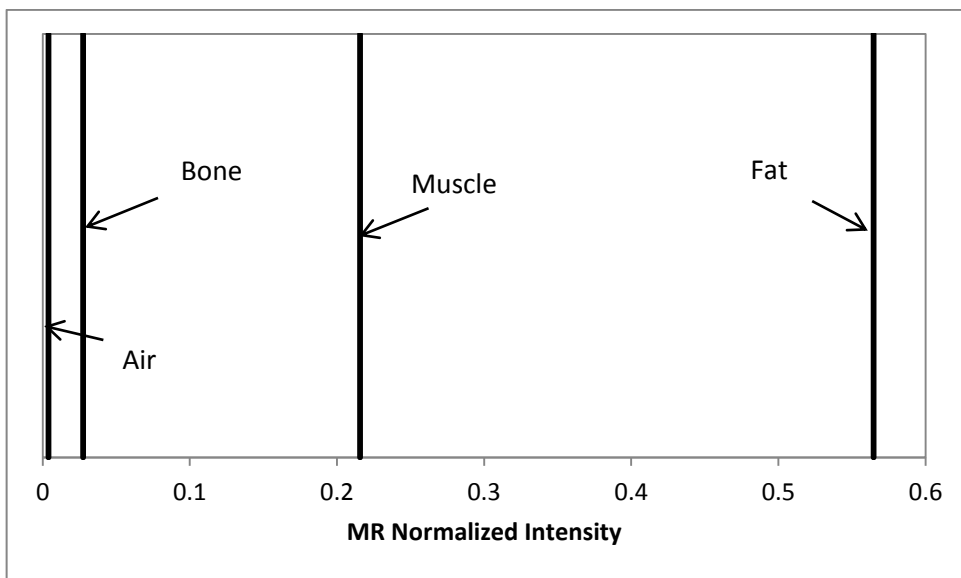
(a)
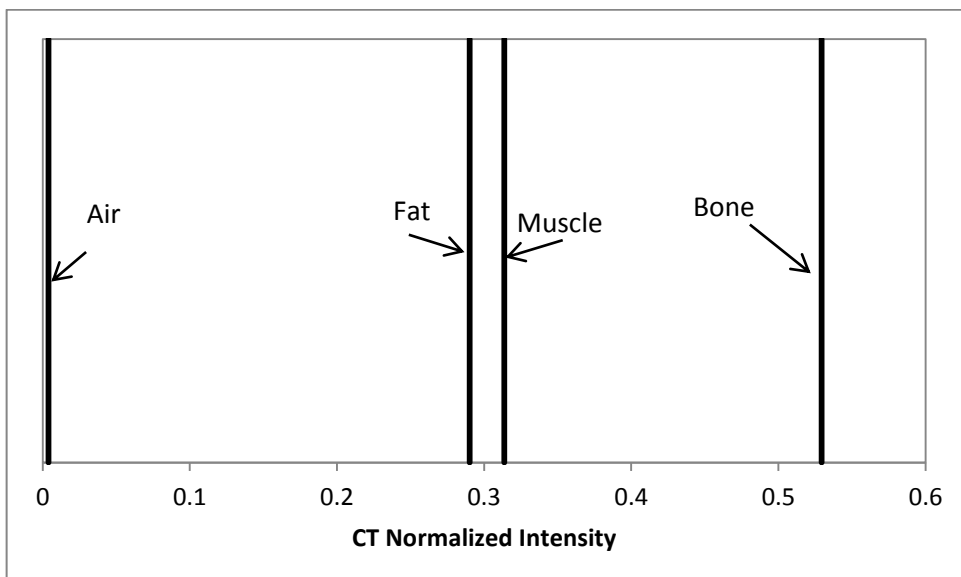
(b)

Figure 2.2 Relative positions of pure classes on a normalized intensity scale in (a)
MR and (b) CT modality
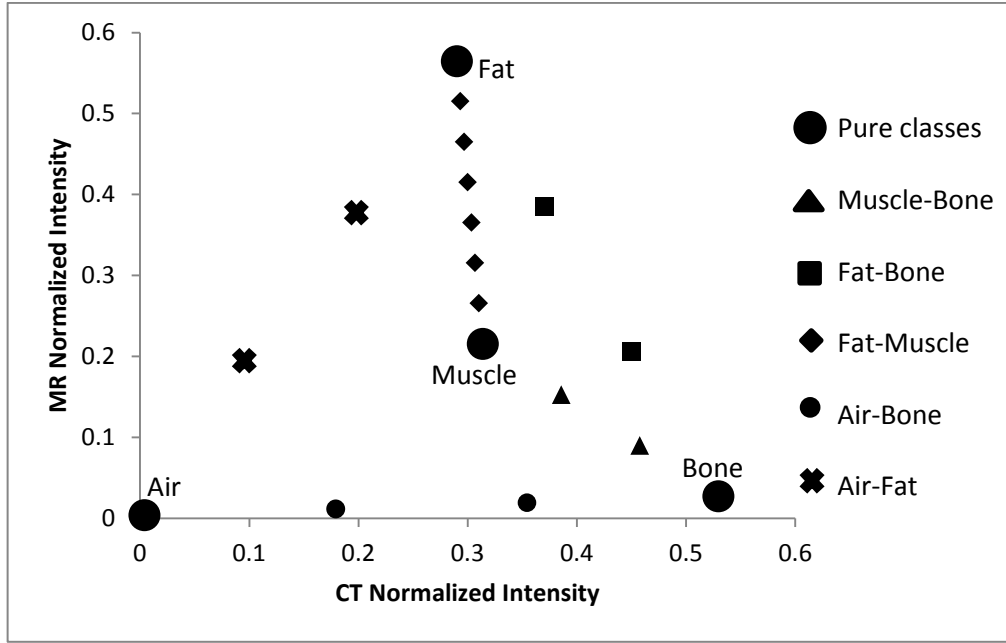
Figure 2.3 Pure and partial volume classes used in the method and corresponding MR-CT normalized intensities

2.3.4 Regularization

A Potts prior has been used by Shattuck *et al.* [16] in the segmentation of brain tissues with a partial volume model. The prior sums the weighted label relationship function for a pixel neighbourhood. The relationship function returns the value *-2* for identical labels, *-1* for labels that share a tissue type (e.g. muscle-fat and pure muscle) and *+1* for not having any common tissues. The function can be written as

$$z(j,k) = \begin{cases} +1, \text{ if } j \text{ and } k \text{ do not have common tissues,} \\ -1, \text{ if } j \text{ and } k \text{ are not of equal type but have a} \\ \quad \text{common tissue type} \\ -2, \text{ if } j \text{ and } k \text{ are of the same tissue type .} \end{cases} \qquad 2.18$$

Here, j and k are indices of the neighbouring pixels. For thin structures and regions close to boundaries, isotropic regularization may have disadvantages. A pixel belonging to a thin structure may be surrounded by mostly background pixels and only a few foreground pixels. This will result in the background pixels affecting the cost function more than foreground pixels and the labelling process may become inaccurate. A solution to this problem is using *a priori* knowledge based on the structural orientation of the pixel. Wong *et al.* [51] used directional priors to obtain vessel segmentation. They focused on using Hessian or orientation matrices at each pixel to obtain structural orientation and assigning higher priority to the neighbour pixel in that direction. It is aimed to combine that approach with the Potts prior. In that regard, as the first step, the Hessian matrix at a pixel is calculated by

$$H = \begin{bmatrix} H_{xx} & H_{xy} \\ H_{xy} & H_{yy} \end{bmatrix}, \qquad\qquad 2.19$$

where H_{xx} and H_{yy} are the second partial derivatives in the x and y directions respectively and H_{xy} and H_{yx} are the mixed partial derivative in the same directions. These derivatives are calculated by finite difference. The eigenvalues e_1 and e_2 and the corresponding eigenvectors $\vec{v}_1$ and $\vec{v}_2$ of the Hessian matrix are then obtained such that $e_1 \geq e_2$. The orientation of the jth pixel, $\vec{o}_j$, is selected as the orthogonal direction to $\vec{v}_1$ if the following conditions are satisfied:

$$e_1 \geq T \quad \text{and} \quad \frac{e_1}{e_2} \geq 2 \qquad\qquad 2.20$$

where T is a threshold value which adjusts the selection of the orientations. $s_{j,t}$ form a mask on the image to indicate whether isotropic or directional prior is to be used and they are defined as binary values according to eigenvalue conditions at pixels of an image as follows

$$s_{j,t} = \begin{cases} 1, & \text{if conditions in 2.20 are true for modality } t , \\ 0, & \text{otherwise,} \end{cases} \qquad 2.21$$

where index t stands for modality type (CT or MR). The decision method of the regularization type and orientation that will be used at any pixel is given in Figure 2.4. According to this process, if $s_{j,CT}$ and/or $s_{j,MR}$ is equal to 1, directional prior in the orientation direction is to be used. If both of them are zero, the pixel is checked for another orientation condition. In order to check this condition, the orientations in a 5x5 neighbourhood of the pixel are normalized to unit length and vector summed. If the magnitude of the sum is greater than 0.25, which is chosen empirically, the use of directional prior is confirmed. In the opposite case, it is decided that the uniform prior is the prior function choice. The 0.25 magnitude and $T = 0.024$ thresholds are chosen empirically so that most boundaries are marked for directional regularization.

The effect of the T threshold on orientations is shown in Figure 2.5. A threshold value of 0.02 results in some pixels in homogeneous regions having orientations. If the threshold is increased to 0.028, the orientations on apparent edges begin to disappear. The empirically obtained threshold of 0.024 yields the most acceptable results on visual inspection. A sample set of orientations after orientation post-processing that are calculated on a sample MR slice are shown in Figure 2.6.

The orientations are used in obtaining the relationship with the neighbouring pixels through an orientation similarity function [51]. This function which returns the relationship between two pixels is

$$h(j,k) = \exp\left(-\frac{\delta(\vec{u}_{jk}, \vec{o}_j)}{2\sigma_h^2} \right), \qquad 2.22$$

where $\vec{u}_{jk}$ is the unit vector in the direction from pixel j to pixel k and $\vec{o}_j$ is the orientation at pixel j. $\delta(\vec{u}, \vec{o})$ is the orientation discrepancy function which

returns decreasing values for increasing compatibility between $\vec{u}$ and $\vec{o}$, and is given by Wong *et al.* [51] as

$$\delta(\vec{u}, \vec{o}) = 1 - \left| \vec{u}^T \vec{o} \right|.$$
2.23

The combination of the orientation similarity and the Potts prior functions is achieved by multiplying them. The reciprocal of the distance between the pixels is used as the weight of the Potts prior. Summation of the combined directional priors for the neighbourhood of a pixel gives the total effect of the proximity labelling. The use of directional prior aims to enhance the effect of pixels in the orthogonal direction to the eigenvectors corresponding to the dominant eigenvalues of the Hessian matrix. Then, final version of the optimized segmentation becomes

$$\vec{\lambda}_{MAP} = \arg \min_{\vec{\lambda}} [(\vec{I}_{MR} - A_{MR}\vec{\lambda})^T W_{MR}^{-1} (\vec{I}_{MR} - A_{MR}\vec{\lambda})$$

$$+ (\vec{I}_{CT} - A_{CT}\vec{\lambda})^T W_{CT}^{-1} (\vec{I}_{CT} - A_{CT}\vec{\lambda})$$
2.24

$$+ \beta \sum_{j=1}^{M} \sum_{k=1}^{N_j} h(j,k) z(j,k)],$$

where M is the number of pixels, N_j is the number of neighbourhood pixels around the jth pixel and $z(j,k)$ is the Potts prior.

2.4   Application of the Method

The details of the application of the method such as label means, localized version of the ICM, metrics to be calculated from the segmentation results and ground truth and some parameters are given in the following subsections.
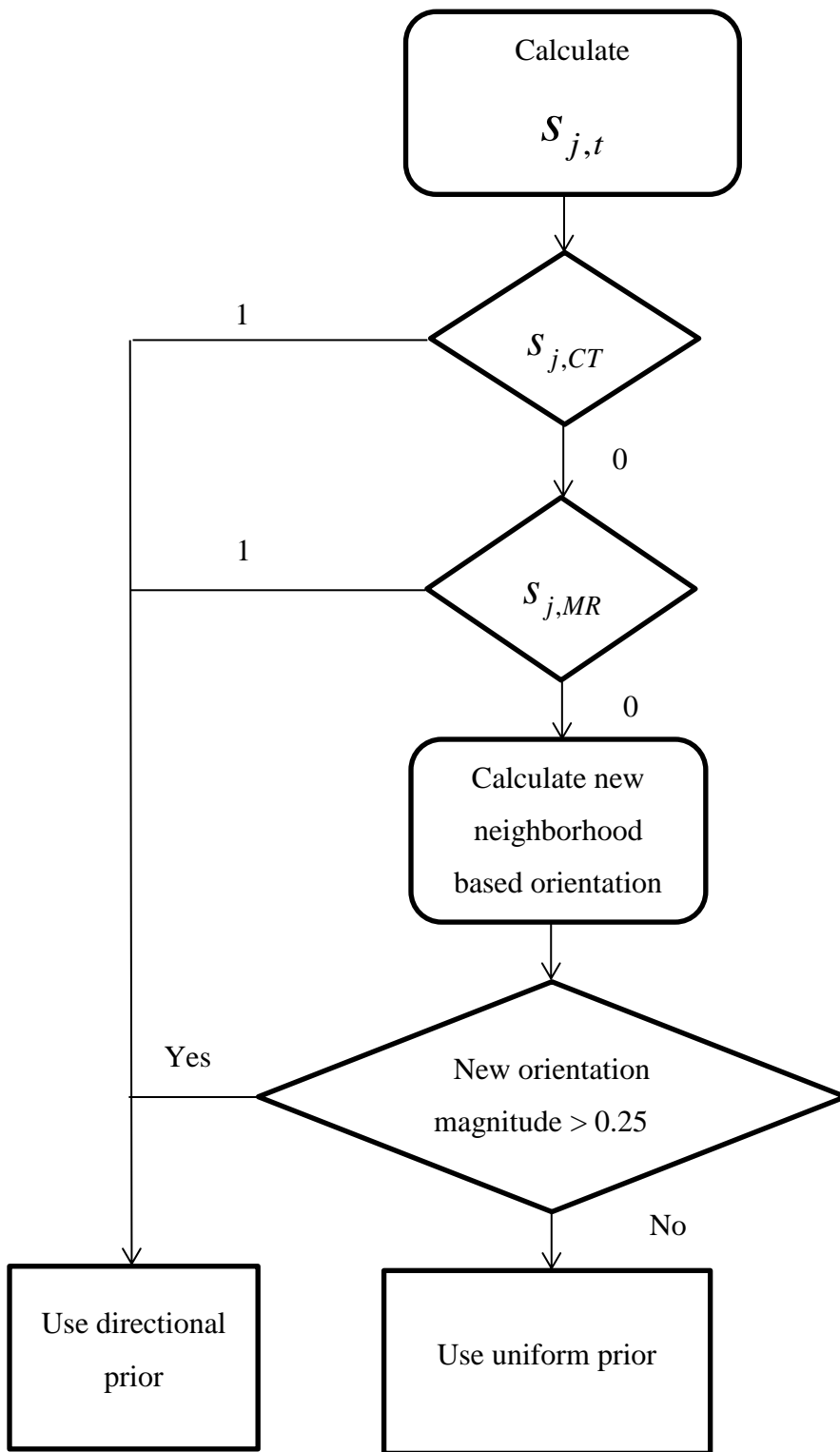
Figure 2.4 Decision workflow for prior selection depending on pixel orientations
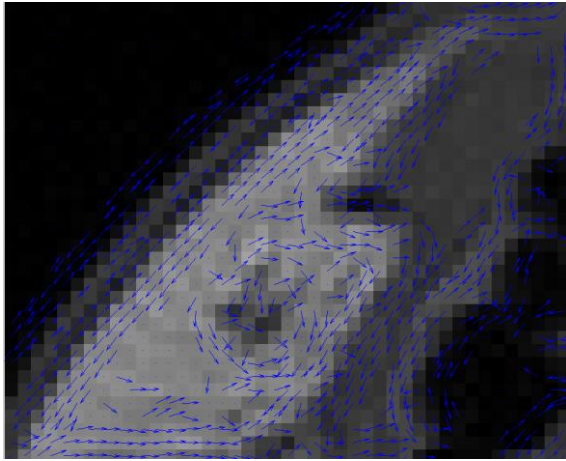
(a)
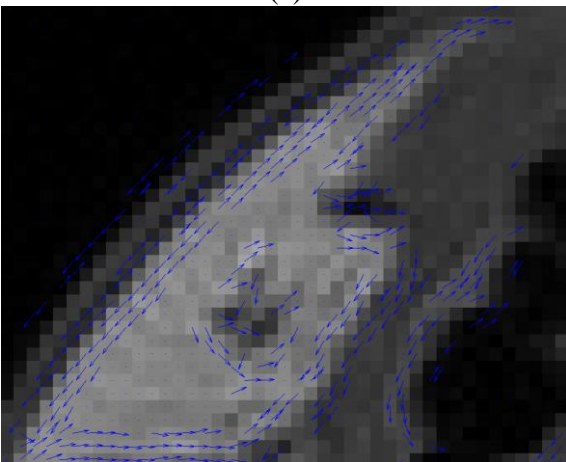
(b)
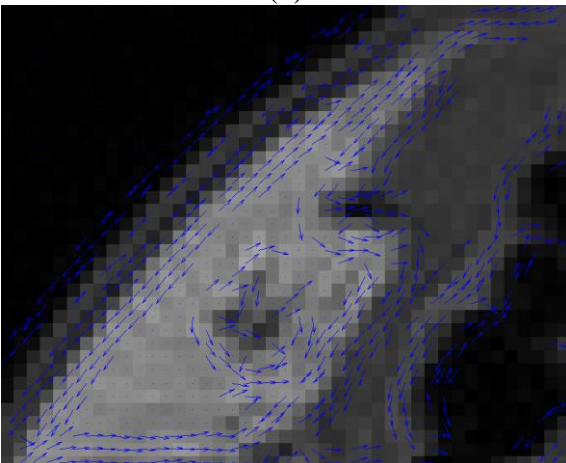
(c)

Figure 2.5 The orientations calculated from a MR slice with different threshold

values, T, (a) 0.02, (b) 0.028 and (c) 0.024
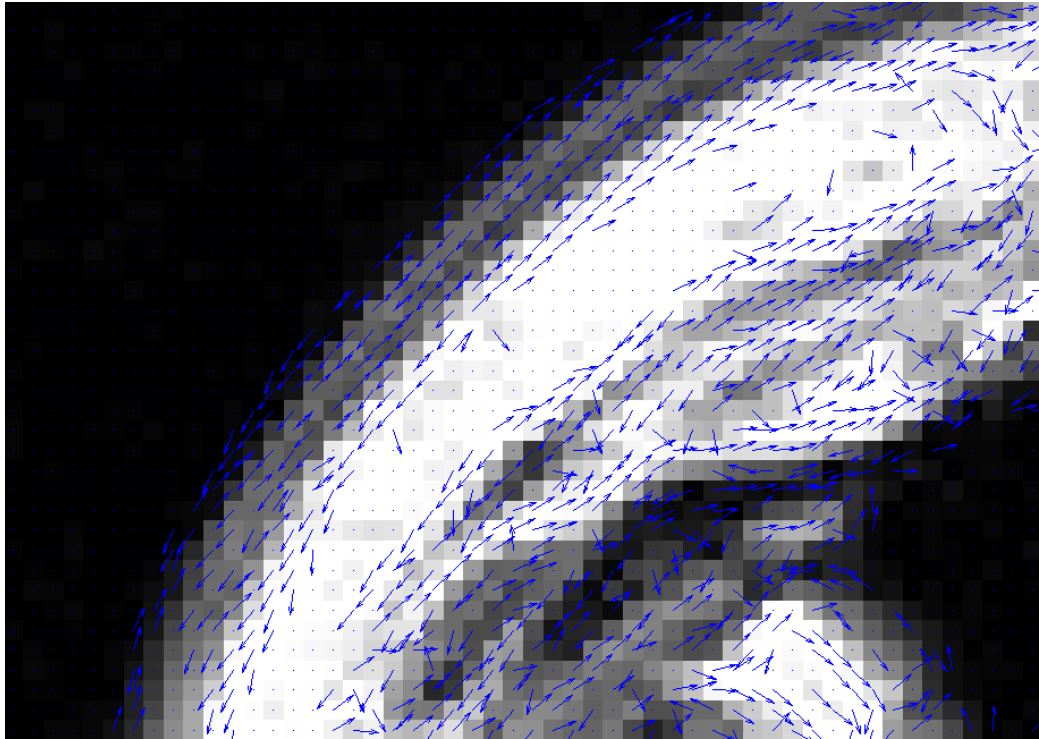
Figure 2.6 Orientations calculated using the Hessian matrix of pixels on an MR slice

2.4.1　Setting the Tissue Class Mean Values

The intensities in the images are normalized such that the minimum and maximum intensities correspond to 0 and 1 respectively. The mean intensities for the fat and muscle classes are obtained from the homogeneous regions that are marked for autocorrelation calculations in section 2.3.1. The bone and air mean intensities are obtained from the intensity histogram of the volume. The normalized mean intensities of the tissue classes are shown in Table 2.3.

Table 2.3 Pure tissue types and their normalized mean intensities

Pure Tissue Type	MR Normalized Intensity	CT Normalized Intensity
Bone	0.027	0.588
Air	0	0
Muscle	0.196	0.247
Fat	0.608	0.215

2.4.2 Localized Iterated Conditional Modes

As explained in section 2.3.1, a 5×5 matrix can be used for the autocorrelation matrix since the values for further sizes can be neglected. Similarly, resolution extent is also within this limit. Therefore, only the 5×5 neighbourhood of a voxel affects the calculated label of that voxel in Iterated Conditional Mode (ICM) iterations. This neighbourhood grid is shown in Figure 2.7. The voxels in the grid are numbered starting from top left. In Equation 2.24, the size of I and λ becomes 25×1 and the size of the square matrices A and W become 25×25. The matrices and vectors are formed according to the grid numbering in raster scan format. The blurring model matrix and autocorrelation matrix are constructed such that the 5×5 kernel elements are spread along diagonal directions and the elements corresponding to non-neighbour relations are set to zero, e.g. the 10th column of the 11th row should be set to zero (Figure 2.7) because despite having consecutive indices, they are not in neighbourhood of each other.

The inverse of W matrix has to be calculated only once for the whole image. As an additional optimization, Equation 2.24 is written in an explicit form such that its parts which are affected by the label of the center pixel, which is pixel 13 of the 5×5 region, and the parts which are not affected by it are separated as

37

$$f(\vec{\lambda}) = f_c + f_{nc},$$ <div align="right">2.25</div>

where f_c are the terms that are functions of the label of the center pixel and f_{nc} are the remaining terms. f_{nc} is calculated once for a pixel at every iteration. Then, for each class to be tried, f_c is obtained according to the class mean intensity. The pixel is labelled with the class which yields the minimum cost function value. After this process is repeated for every pixel, the current labels of the pixels compared to the labels of the previous iteration. The iterations are terminated if 98% of the pixels do not change labels.

1	2	3	4	5
6	7	8	9	10
11	12	13	14	15
16	17	18	19	20
21	22	23	24	25

Figure 2.7 Localized voxel grid for ICM in raster scan format. All 25 pixels are used to update the center pixel only

2.4.3   Metrics to Evaluate Segmentation Performance

The metric choice for segmentation evaluation is based on the metrics which were used in a liver segmentation competition [63]. Especially, the metrics that aim to evaluate the contour accuracy are considered. One of those metrics of that type is

Average Symmetric Surface Distance (ASSD). The initial step for this metric is to obtain the border pixels of the test and reference images. Border pixel of a tissue type is defined as a pixel with a neighbour pixel of different tissue types. For the pixels of the first border, the closest pixel of the other border is found using Euclidean distance. This process is repeated for the second border and all distances are averaged.

The Root Mean Square Symmetric Surface Distance (RMSSSD) is a similar metric to ASSD. The only difference is that the squared distances are averaged and the square root of the average is the metric. Maximum Symmetric Surface Distance (MSSD) is the metric which returns the maximum of distance obtained in the calculation of ASSD.

Dice coefficient measures the ratio of the intersection and union of two sets which are the ground truth labels and labels obtained by the segmentation process. Although Dice coefficient is widely used in segmentation and it is an efficient evaluation metric, it has some drawbacks for thin and small structures. Even few inaccurate pixels affect drastically the Dice coefficient in small structures drastically while the effect becomes minimal for large tissues with similar contour segmentations errors. This is because of the fact that the ratio of the number of false positive and negative pixels to the number of total pixels is smaller in large areas.

Another possible metric to be calculated is the Partial Volume Difference (PVD). The partial volume ratios of synthetic image pixels are known during the construction of the image. An error metric is to be calculated as the difference of these known partial volume ratios and the partial volumes found by the segmentation method. Applicability on real data depends on the manual segmentation method and whether the manual method allows partial volume pixels.

The results are presented in RMSSSD and Dice metrics because of the parallelism in RMSSSD and ASSD results. The lack of sub-voxel manual segmentation prevented the PV metric from being used. In order to convert the RMSSSD and Dice metrics into a single value, these metrics are first normalized in a [0, 100] interval. This is done by selecting a region in which there is a mixture of thick and thin muscles and the ratio of the thick muscles is slightly favoured so that the method performs well for all cases. ROI 2 is selected when these conditions are considered. A sample slice of ROI 2 is shown in Figure 2.8. The average Dice coefficient of all methods in this ROI (0.914) is assigned a score of 90. A Dice score of 1.0 is given a perfect 100 score. Any Dice score obtained is converted to this score by interpolation or extrapolation. Similarly, the average RMSSSD metric (0.829) is assigned 90 as the combined score. The RMSSSD value for a perfect segmentation is 0, so the 100 score is matched with that value. Any RMSSSD score obtained is converted to this score by interpolation or extrapolation. The combined score is then calculated as the average of the normalized Dice and RMSSSD scores.

2.4.4 Simulation Data

Here, before proceeding with the testing of the methodology with real data, method is tested on synthetic data which would also help in determining the optimum algorithm parameters. For one of the synthetic images, two thin elliptic strips which represent the muscle tissue are created and connected to each other. The background is assumed to be fat tissue. The pixels which are inside the muscle region are marked in high resolution and down-sampled to a lower resolution by pixel averaging. The down-sampling yields an image with pixel values in the interval of [0, 1] and this image represents the partial volume ratio of the muscle tissue at every pixel. These ratios are to be used to give the effect of partial volumes to the synthetic CT and MR images. The high and low resolution muscle masks are given in Figure 2.9.
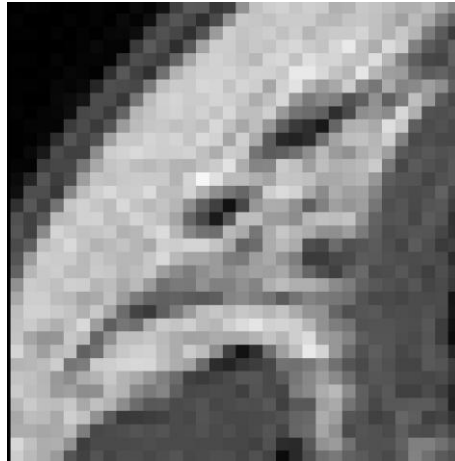
Figure 2.8 A slice from ROI 2, the region around which the combined scores are calculated

<div align="center">(a)   (b)</div>
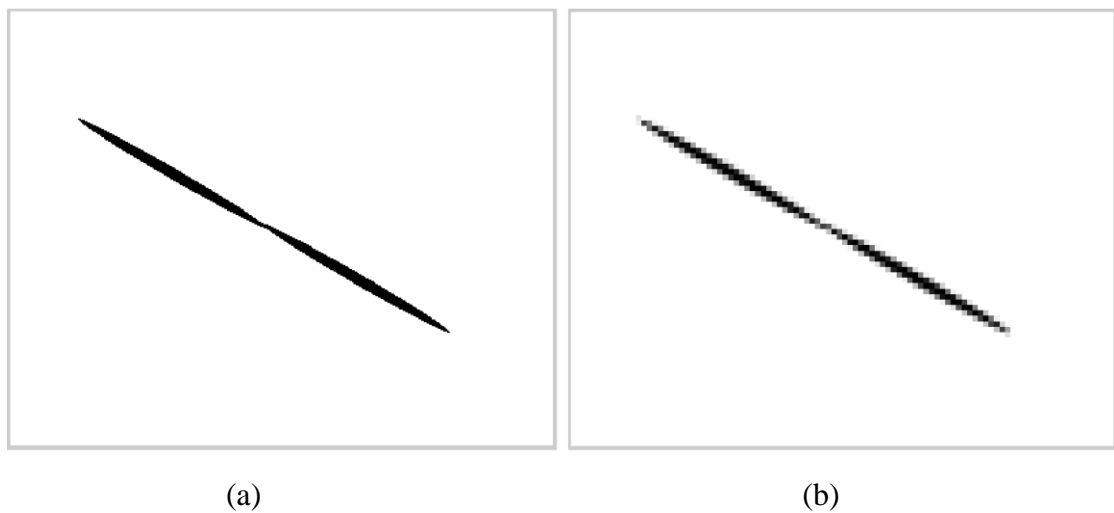
Figure 2.9 (a) High resolution and (b) low resolution of synthetic muscle

The CT and MR images corresponding to the muscle-fat partial volume maps are created by setting the pixel intensities to the weighted sum of the mean intensities. For example, the intensity of a pixel in the CT image with muscle volume ratio α_{muscle} and fat volume ratio $1-\alpha_{muscle}$ is set to $\alpha_{muscle}\mu_{muscle_{CT}} + (1-\alpha_{muscle})\mu_{fat_{CT}}$.

The images are then smoothed by 5×5 Gaussian kernels with the σ values obtained in section 2.3.2.

To complete the synthetic data, independent noise is added to the MR image while coloured noise is added to the CT image using the estimated noise statistics of patient data. The maximum artificial noise magnitudes are 5% and 15% of the image intensities for the CT and MR images respectively. Although the noise percentages appear to be low, the magnitudes of the added noise are approximately 30% of the difference between muscle and fat mean intensities. The blurred and noise added images can be seen in Figure 2.10. Another set of artificial data is created to model thin muscles surrounded by fat and bone in 3D. Similar to the first artificial set, the labels are prepared in higher resolution than the final image. The down-sampled, smoothed and noise added CT and MR versions of the original label image are shown in Figure 2.11.

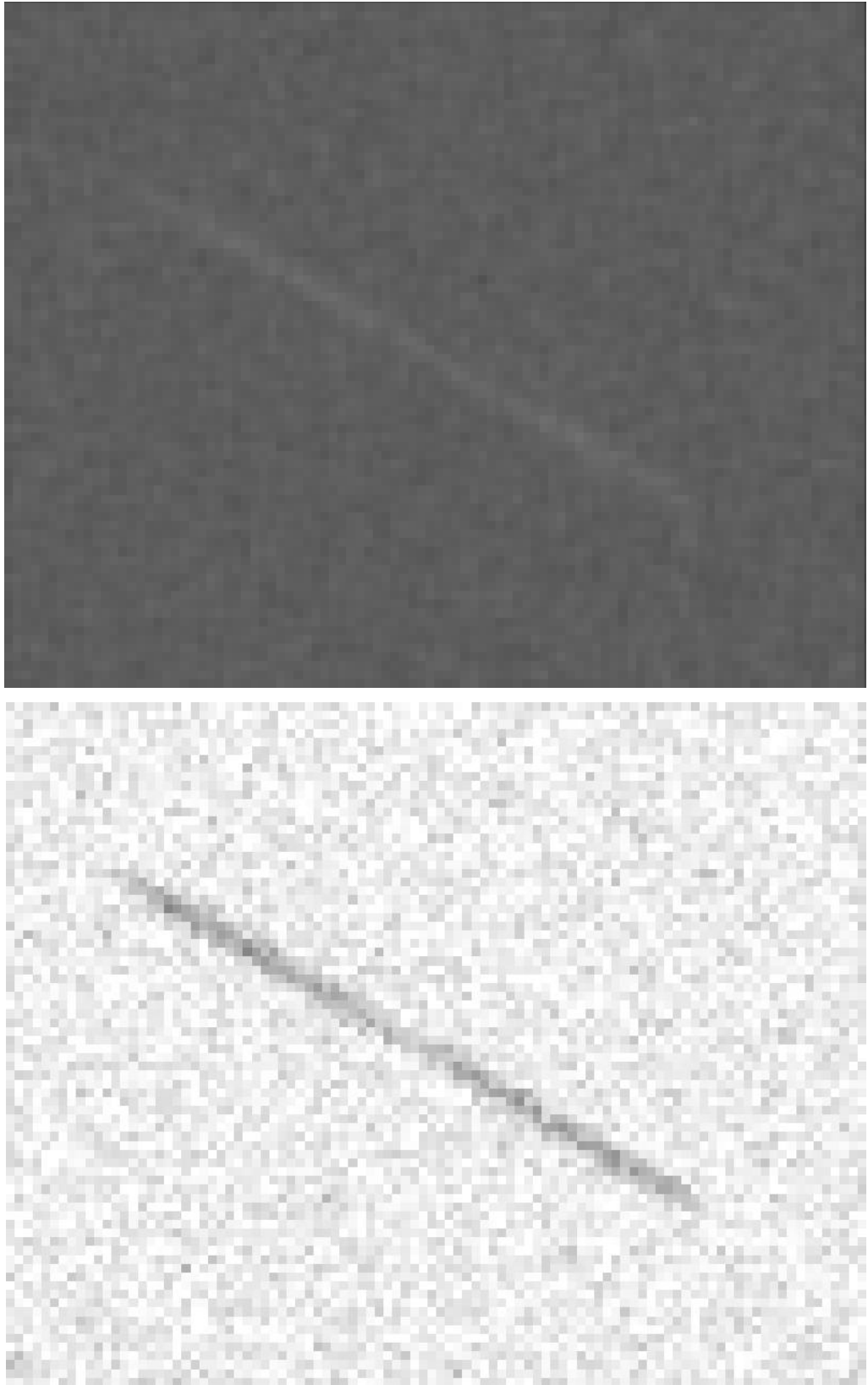(THIS SPACE IS LEFT BLANK INTENTIONALLY.)

Figure 2.10 Blurred and noise-added artificial (top) CT and (bottom) MR images
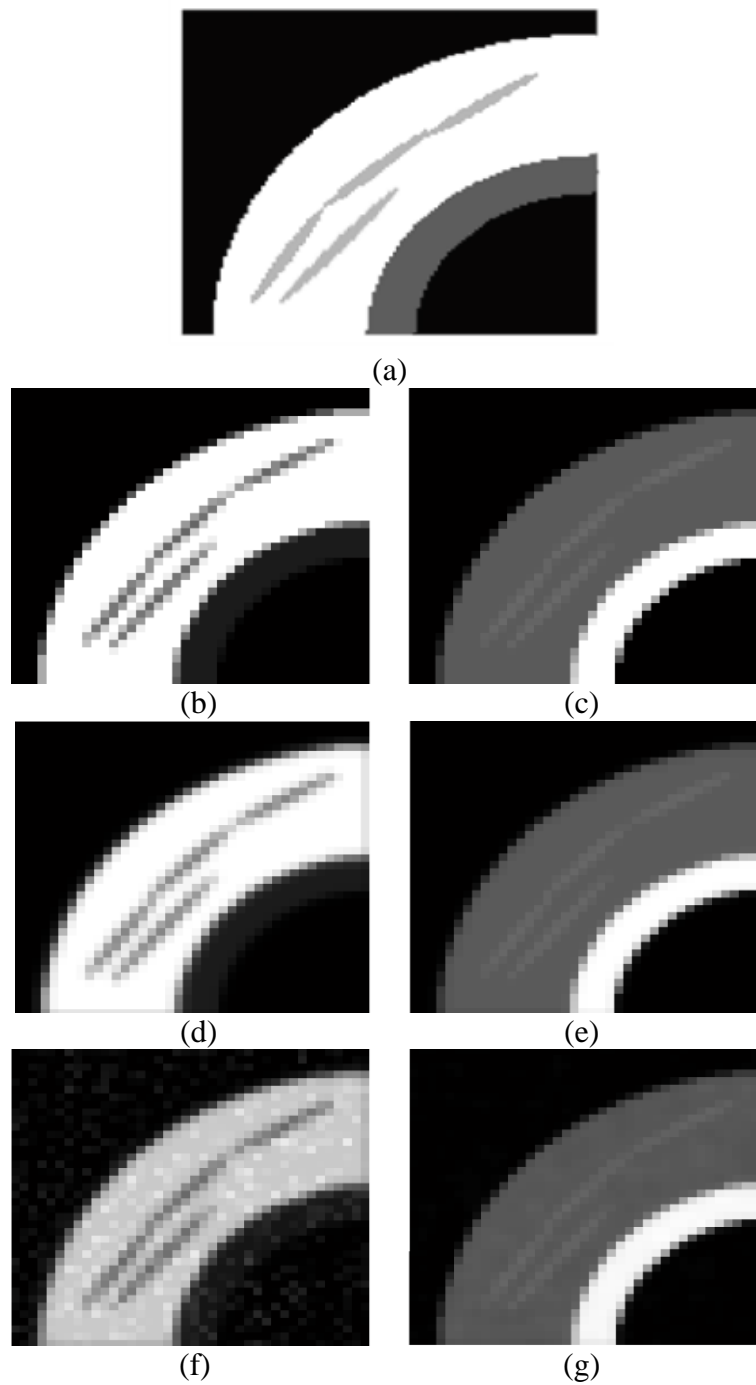
Figure 2.11 Synthetic image creation steps. (a) High resolution labeled image, (b) down-sampled MR, (c) down-sampled CT, (d) blurred MR, (e) blurred CT, (f) noise-added MR and (g) noise-added CT simulations

2.4.5 Setting the Parameters

The coefficient of the prior term β (for both uniform and directional priors) is one of the most important parameters of the method presented here. It is a weight value to set the balance between the likelihood and prior terms in the cost function. Setting the value too low may result in noisy segmentations which are based mostly on the intensity value of a pixel while setting its value too high may cause some regions to overgrow as iterations progress. In order to optimize the value of the parameter, several runs are made within an interval. The metrics presented in section 2.4.3 are calculated and placed on a graph to visualize the optimal solution. The graphs of different metrics vs. β values for CT only, MR only and MR+CT fusion cases with directional MRF prior are given in Figure 2.12, Figure 2.13 and Figure 2.14 respectively. For each case, the metrics are normalized and summed to obtain a unified metric. The optimal β values which optimize these metrics are then obtained. Dice coefficient for fat tissue was not used because it did not vary significantly with β.

The optimal β values for uniform and directional neighbourhood models and for different image modalities are given in Table 2.4.

Table 2.4 Optimal β values for various modality and MRF prior cases

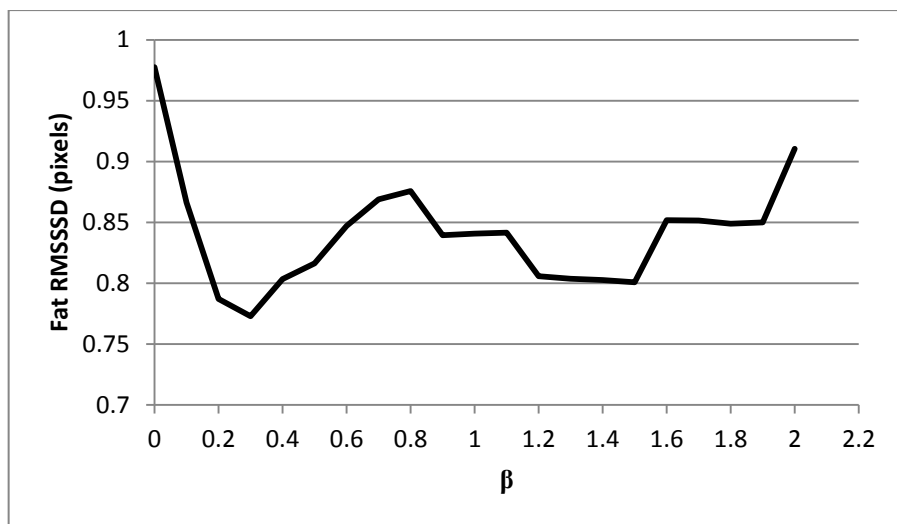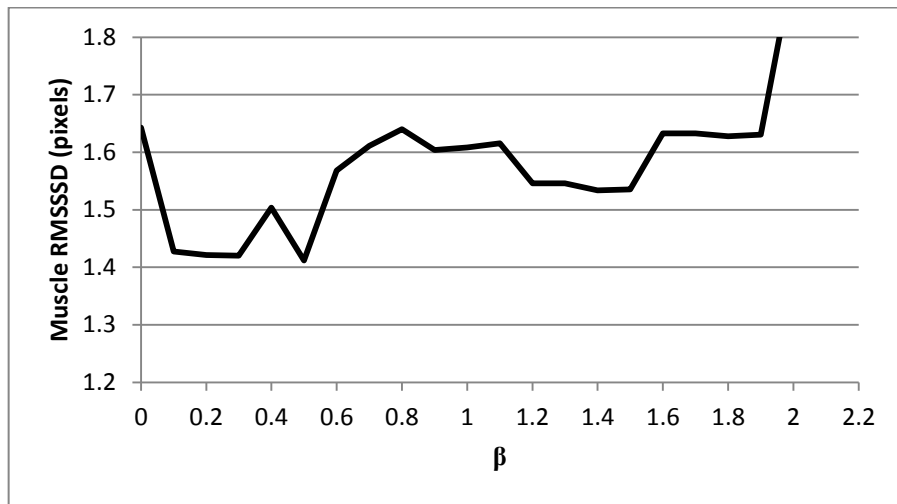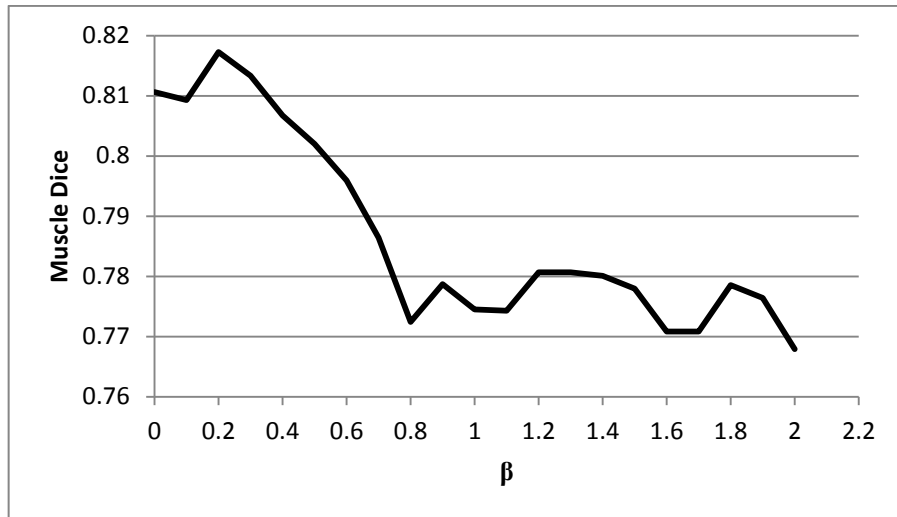| Modality	MRF Prior	β
CT	Directional	0.19
MR	Directional	0.18
CT+MR	Directional	0.22
All	Uniform	0.32

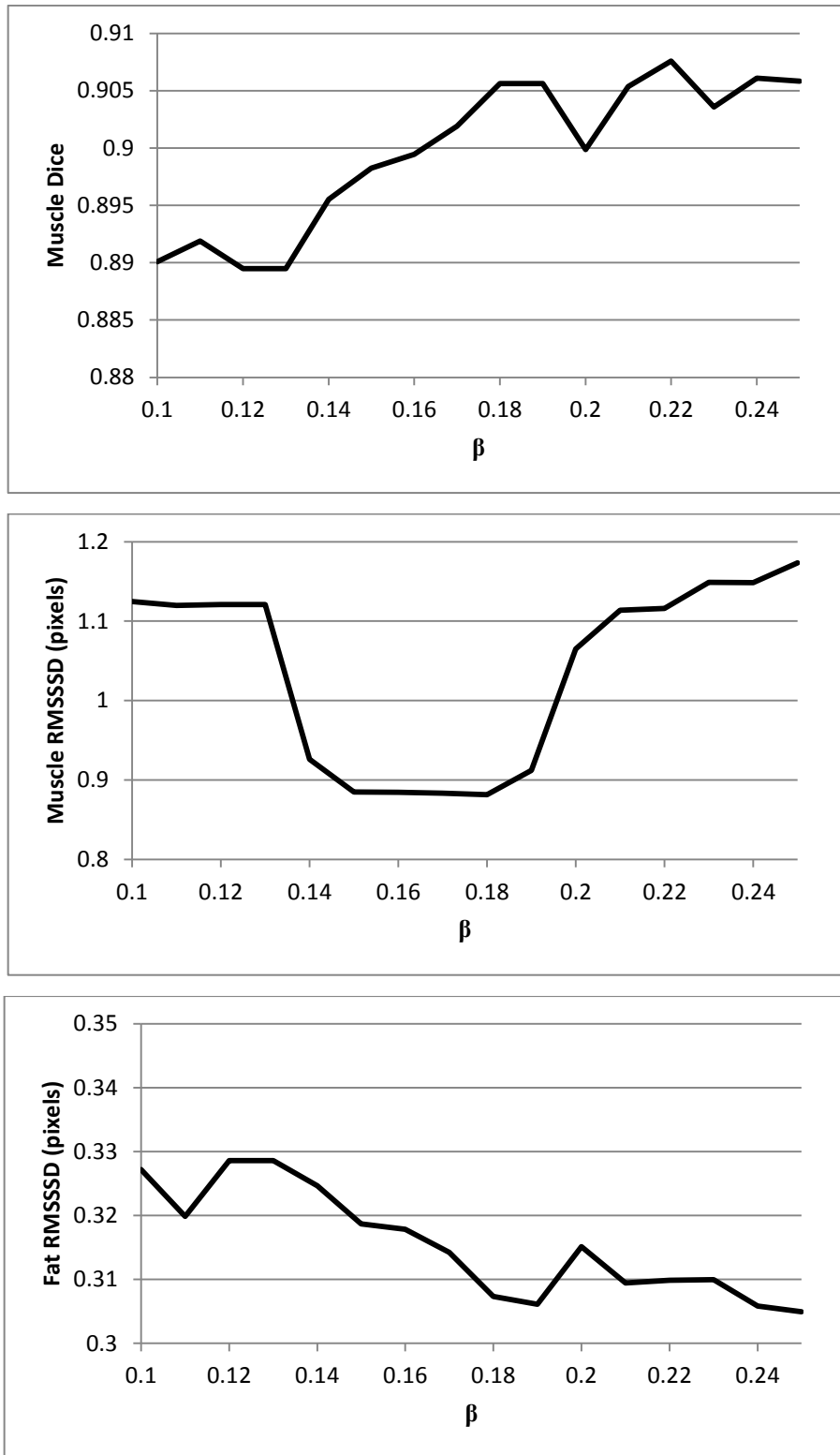**Figure 2.12** Three metrics vs. β in CT only case for the directional prior

Figure 2.13 Three metrics vs. β in MR only case for the directional prior
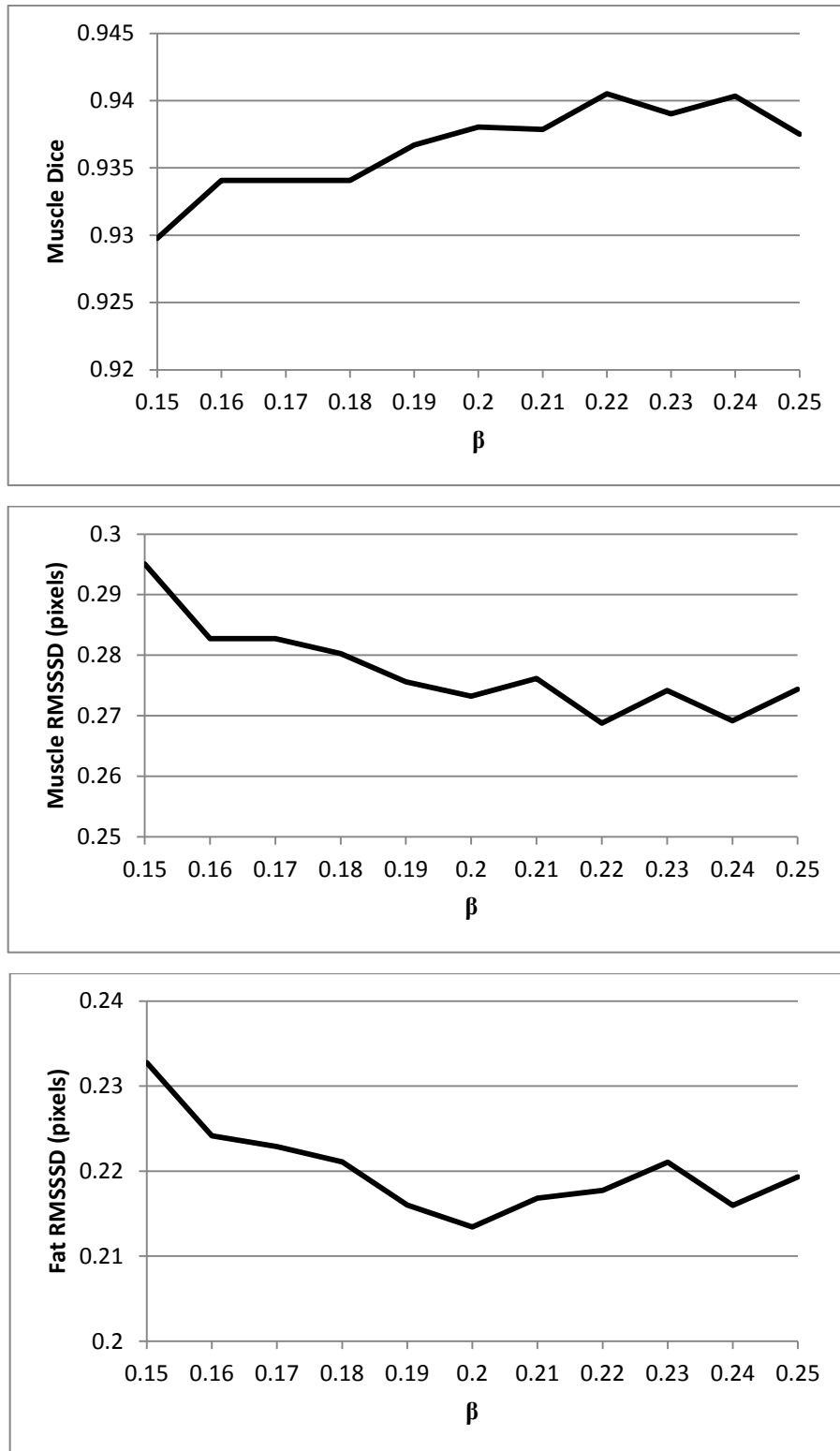
Figure 2.14 Three metrics vs. β in CT+MR fusion case for the directional prior

2.4.6 Post Processing

The intensity values belonging to skin voxels are similar to muscle intensities in both modalities. The primary aim of this study is proposing and testing a segmentation method for facial muscles. Therefore, the skin is removed with a post-process in order to be able to assess the muscle segmentation more accurately. The removal of the skin is done by using morphological erosion operation and pixel connectivity of the muscle-like tissue segmentation result.

Bones contain marrow, which has muscle-like intensity and is usually classified as muscle tissue by the method. Therefore, a simple filling algorithm is performed for the bone class in order to change the labelling of the marrow from muscle to bone.

Another post-process step is the label reassignment of the pixels on bone boundary. The pixels on the bone-fat interface are sometimes misclassified as muscle pixels. Therefore, the bone segmentation is binarized and expanded by one voxel. The voxels to which the expansion occurs are given new labels with a voting scheme in their neighbourhood. They are assigned the label that appears the most times in the adjacent voxels.

The segmentation result may contain noisy small components. A post-process step is performed to remove components of one pixel size. This is performed on the 3D segmentation result so that thin structures with continuity in the axial direction do not get removed.

CHAPTER 3

RESULTS

3.1   Test Cases

The method will be applied by using a combination of blurring estimation model, partial volume class modelling, single image/fusion modelling and regularization components. The effect of the contribution of these components will be examined on test images. The eight main cases and the models and images used in them are shown in Table 3.1. The CT and MR image cases only have regularization model as a variable because the lack of other components yield very low metric results and they were excluded from the comparisons. Each case is assigned with an index and a short name. The short names encode which models are used in the case and they are given in Table 3.1. The three letters represent the blurring, PV and regularization models respectively. The test cases will be denoted either by index number or short name. The initialization of the algorithm is performed by using the case where none of the models are used and is not included in Table 3.1.

3.2   Results

Sixteen ROIs were defined in Table 2.1. Dice, RMSSD, combined scores and standard deviations of these scores that are calculated of 8 cases over these ROIs are given in Table 3.2. The individual results of muscle, fat and bone regions in ROIs are given in from Figure 3.1 to Figure 3.9. The air results do not differ

significantly between the different methods. Hence, they are not provided in detail. The calculated mean and standard deviation values over all ROIs are given in Figure 3.10 to Figure 3.12 also.

Table 3.1 Test cases: models and image modalities used in them. In the second and third columns, '+' indicates that the corresponding model is used and '-' indicates the opposite case. The letters of the short names represent the blurring, PV and regularization models.

Case Index	Blurring Estimation (A matrix)	PV Model	Images Used	Regularization Model (prior)	Short Name
1	+	+	MR	Uniform	ooU
2	+	+	MR	Directional	ooD
3	+	+	CT	Uniform	ooU
4	+	+	CT	Directional	ooD
5	-	-	MR+CT (Fusion)	Uniform	xxU
6	-	+	MR+CT (Fusion)	Uniform	xoU
7	+	+	MR+CT (Fusion)	Uniform	ooU
8	+	+	MR+CT (Fusion)	Directional	ooD

Table 3.2 (From top to bottom) Muscle, fat and bone score averages and standard deviations over all ROIs (RMSSSD values are in pixels)

Case Index	Case	Dice Mean	Dice Std.	RMSSSD Mean	RMSSSD Std.	Combined Mean	Combined Std.
1	MR (ooU)	0.767	0.111	1.428	0.589	77.894	9.862
2	MR (ooD)	0.777	0.103	1.417	0.602	78.538	9.473
3	CT (ooU)	0.750	0.106	1.483	0.636	76.545	9.480
4	CT (ooD)	0.750	0.106	1.471	0.628	76.646	9.394
5	Fusion (xxU)	0.839	0.079	1.324	0.556	82.694	7.591
6	Fusion (xoU)	0.857	0.057	1.033	0.275	85.496	4.546
7	Fusion (ooU)	0.868	0.048	0.912	0.293	86.864	4.010
8	Fusion (ooD)	0.865	0.050	0.899	0.286	86.771	4.159

Case Index	Case	Dice Mean	Dice Std.	RMSSSD Mean	RMSSSD Std.	Combined Mean	Combined Std.
1	MR (ooU)	0.913	0.022	0.607	0.067	91.295	1.351
2	MR (ooD)	0.914	0.021	0.607	0.061	91.349	1.248
3	CT (ooU)	0.850	0.089	1.845	1.520	80.143	14.173
4	CT (ooD)	0.850	0.089	1.844	1.518	80.159	14.141
5	Fusion (xxU)	0.929	0.022	0.633	0.113	92.062	1.385
6	Fusion (xoU)	0.930	0.021	0.627	0.102	92.152	1.475
7	Fusion (ooU)	0.927	0.023	0.572	0.078	92.310	1.603
8	Fusion (ooD)	0.925	0.024	0.571	0.072	92.230	1.665

| Case Index	Case	Dice Mean	Dice Std.	RMSSSD Mean	RMSSSD Std.	Combined Mean	Combined Std.
1	MR (ooU)	0.550	0.247	4.545	3.591	46.459	30.238
2	MR (ooD)	0.554	0.238	4.448	3.345	47.314	27.970
3	CT (ooU)	0.789	0.247	0.754	0.339	83.224	13.740
4	CT (ooD)	0.789	0.247	0.757	0.344	83.190	13.739
5	Fusion (xxU)	0.738	0.299	0.922	0.568	79.222	16.180
6	Fusion (xoU)	0.779	0.308	1.201	2.041	79.942	27.115
7	Fusion (ooU)	0.799	0.252	0.705	0.362	84.105	14.415
8	Fusion (ooD)	0.800	0.252	0.690	0.344	84.217	14.397
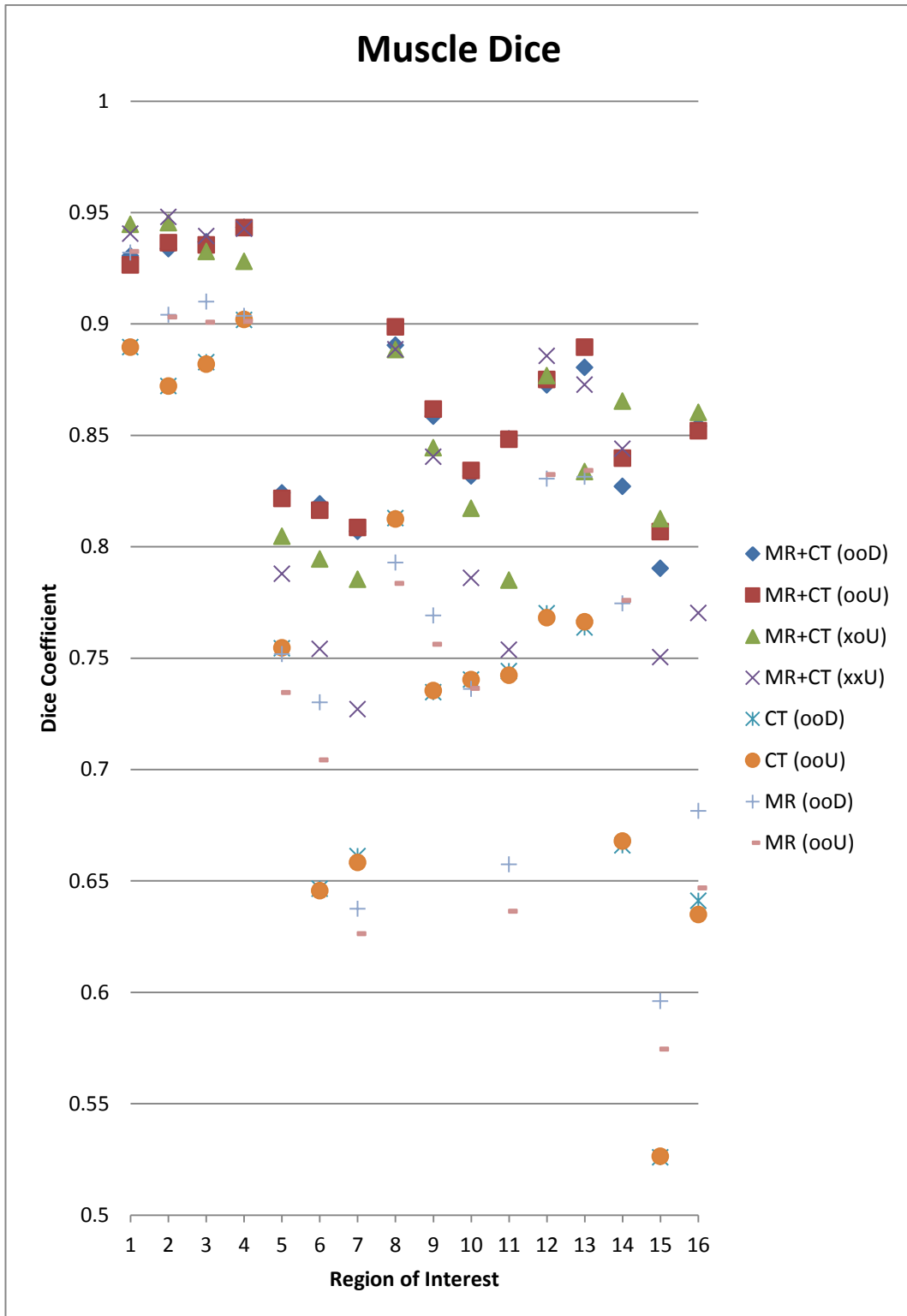
Figure 3.1 Dice coefficient results for muscle in all ROIs

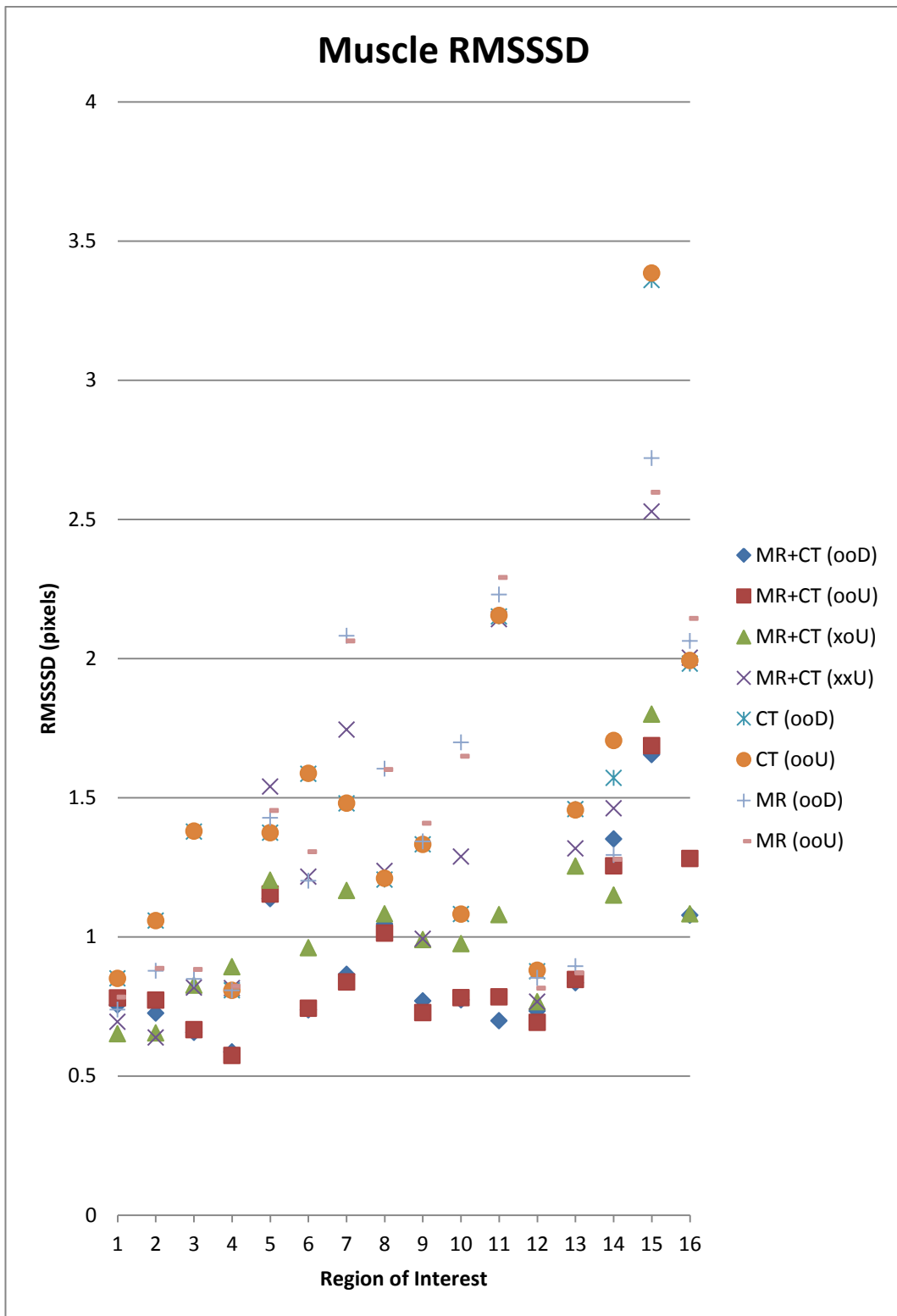**Figure 3.2** RMSSSD results for muscle in all ROIs

Figure 3.3 Combined metric results for muscle in all ROIs
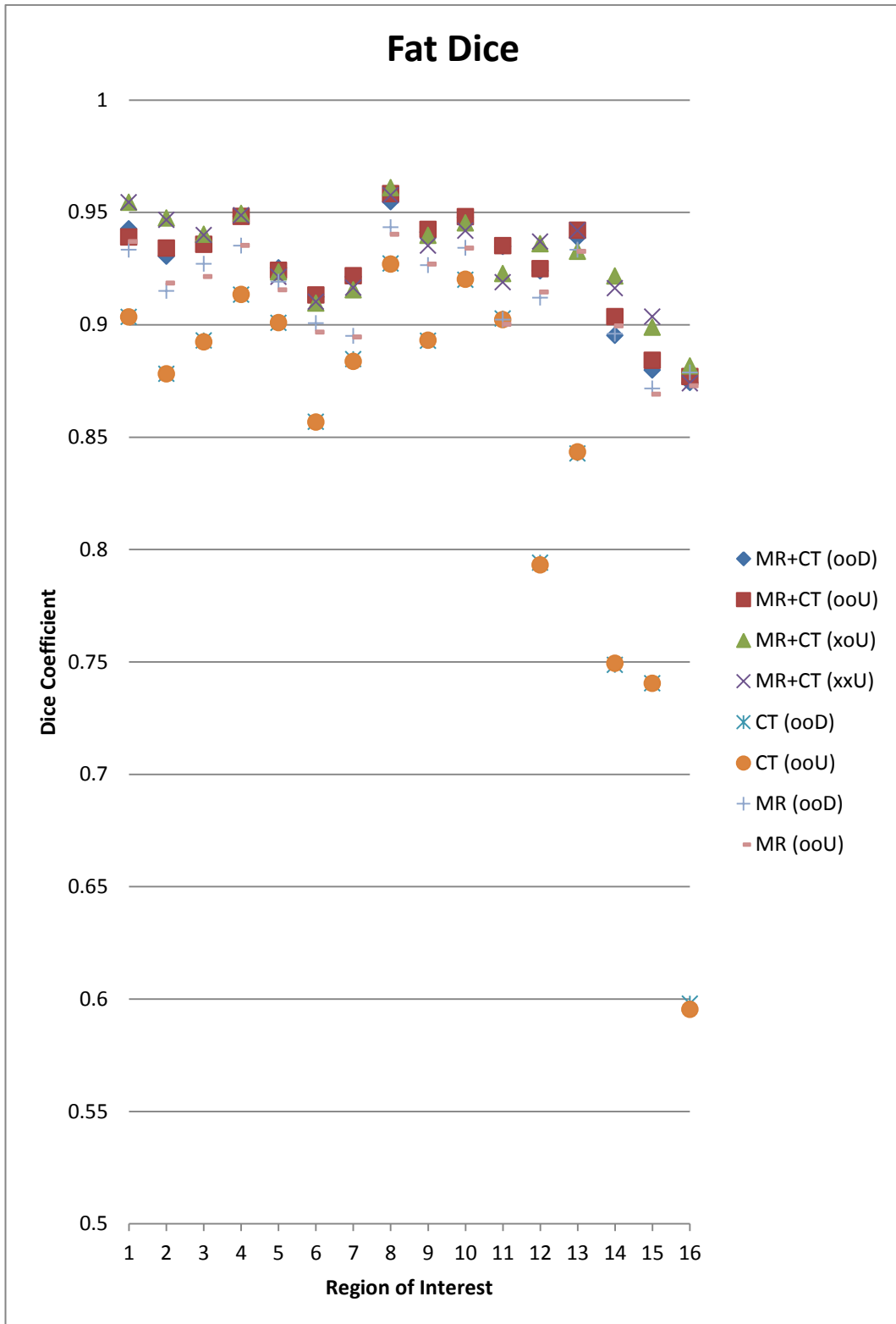
Figure 3.4 Dice coefficient results for fat in all ROIs

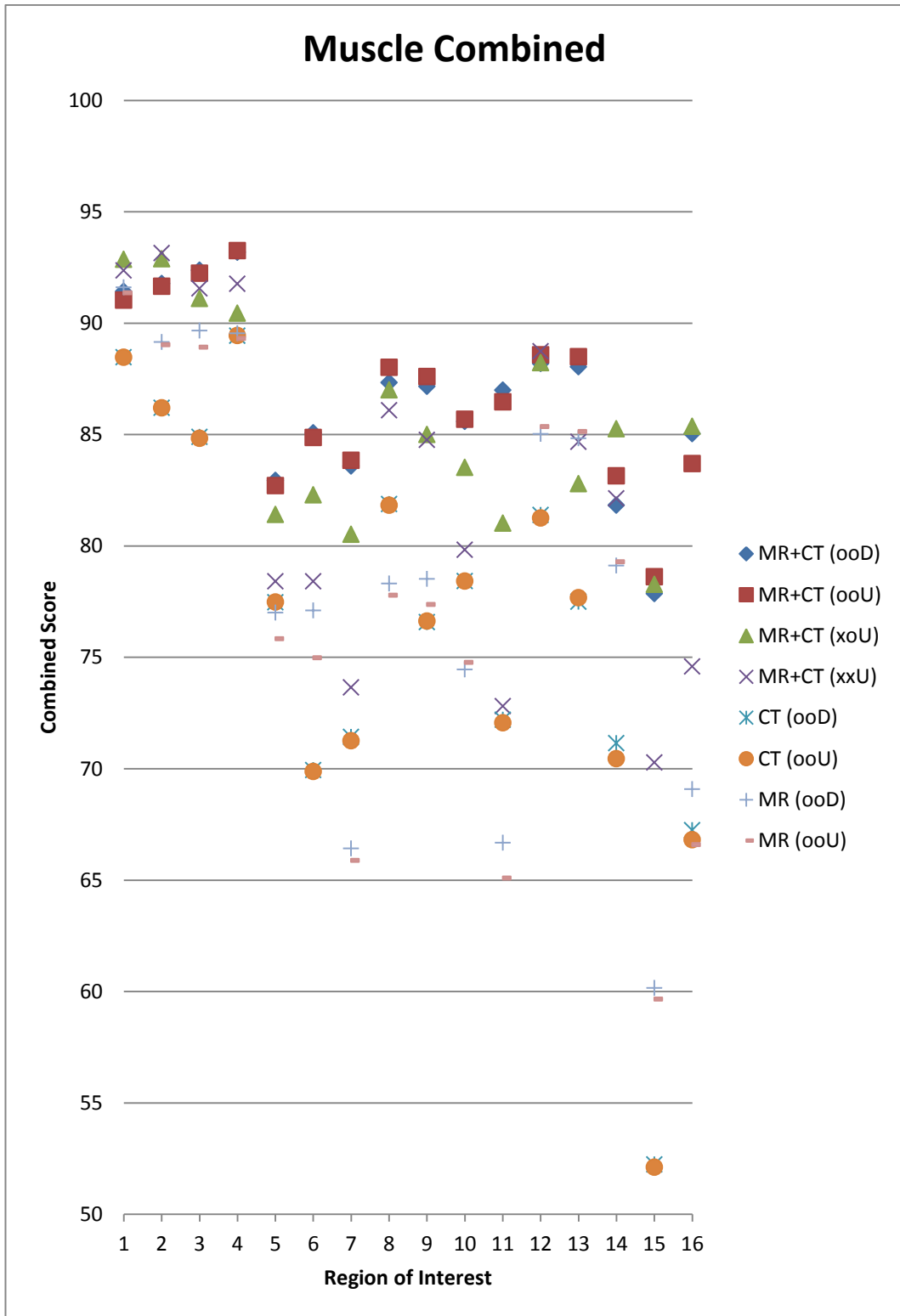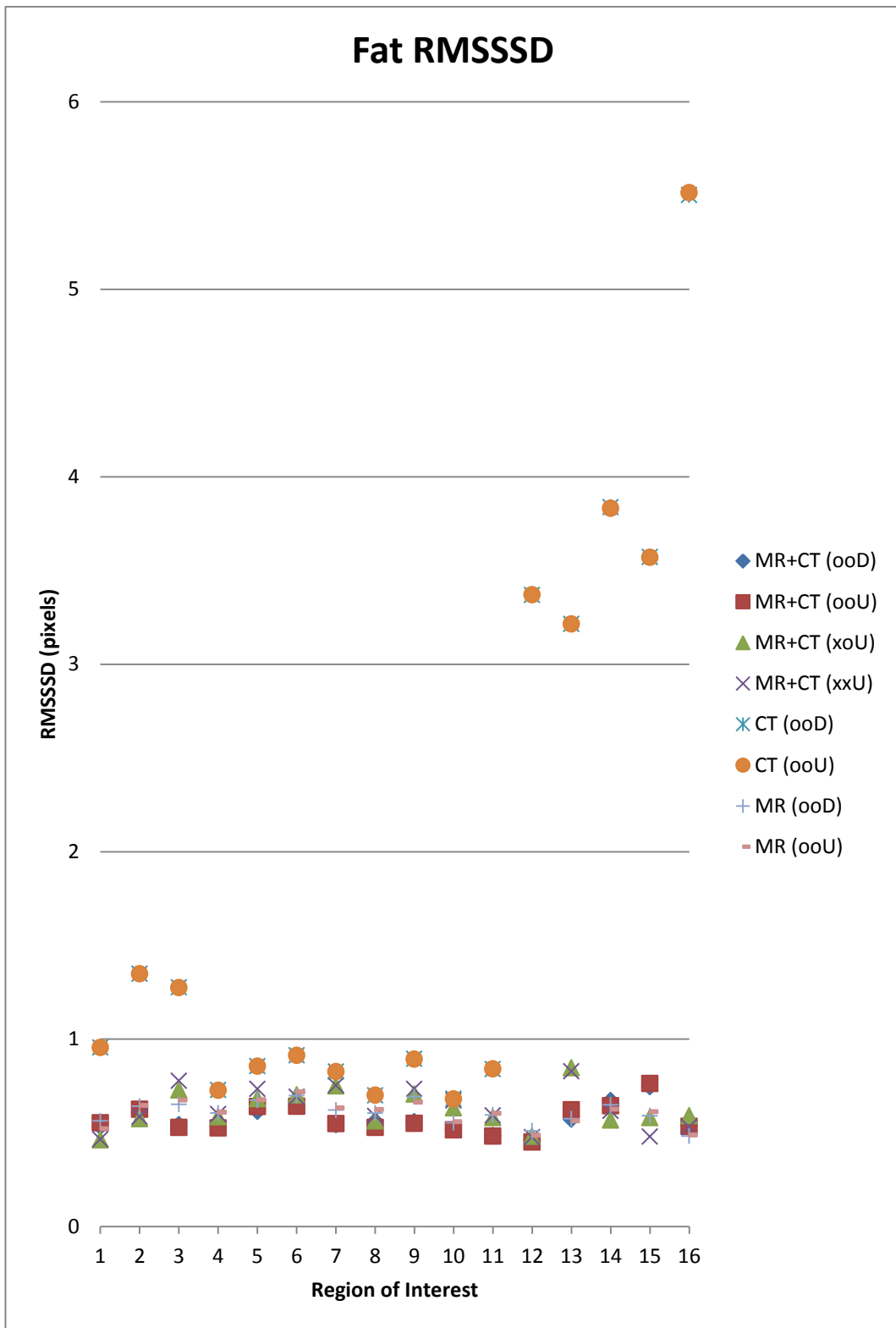**Figure 3.5** RMSSSD results for fat in all ROIs
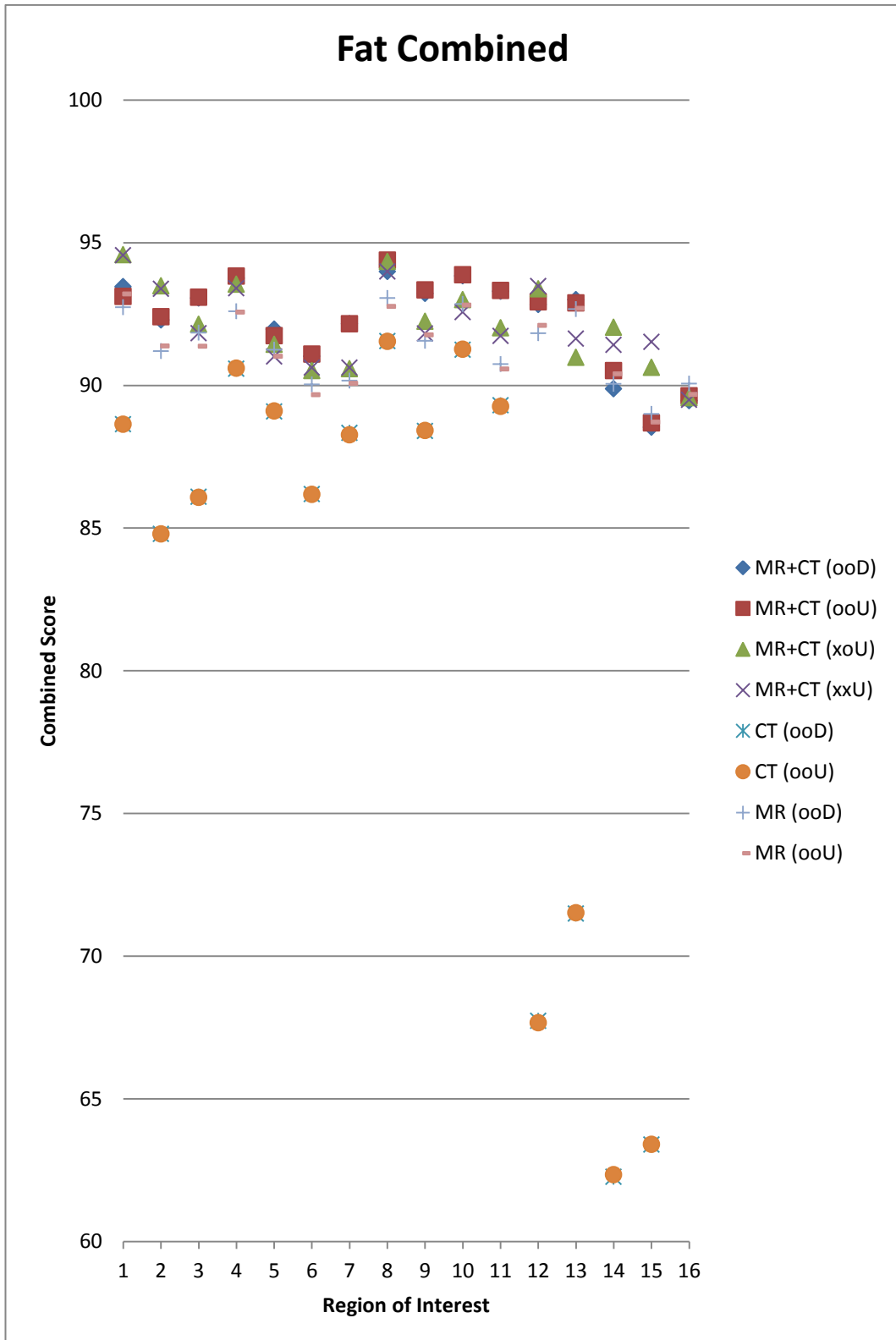
Figure 3.6 Combined metric results for fat in all ROIs
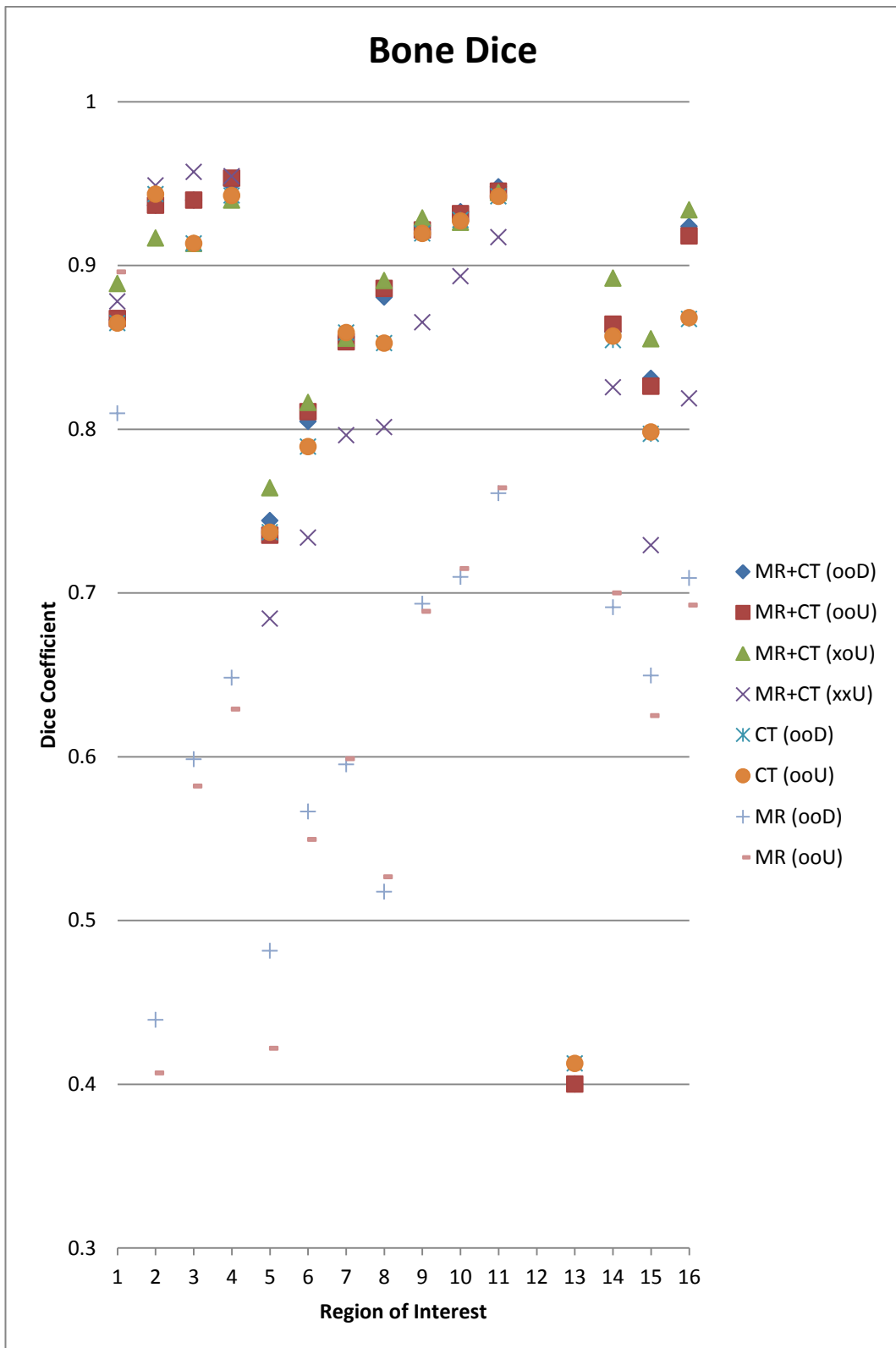
Figure 3.7 Dice coefficient results for bone in all ROIs
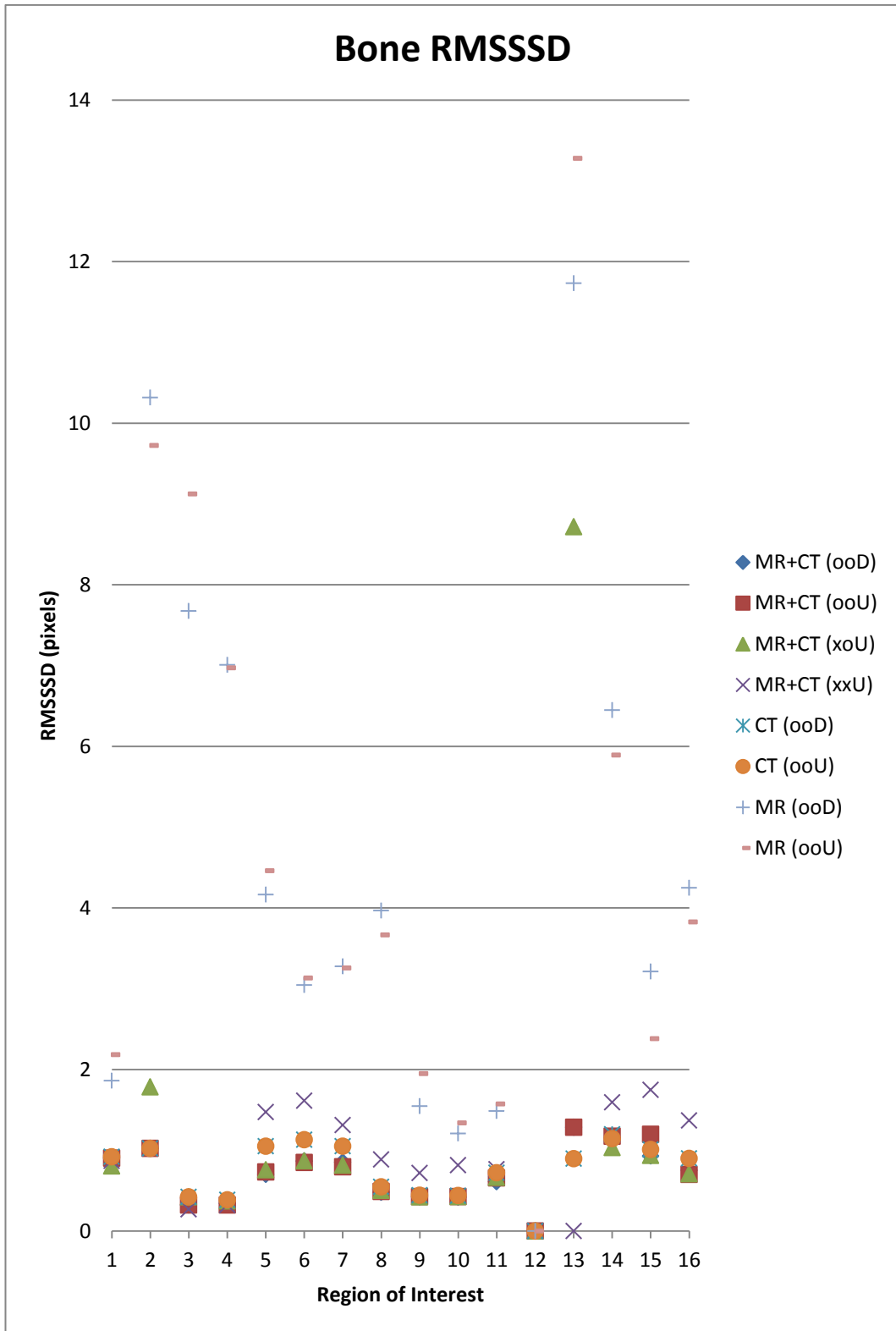
Figure 3.8 RMSSSD results for bone in all ROIs
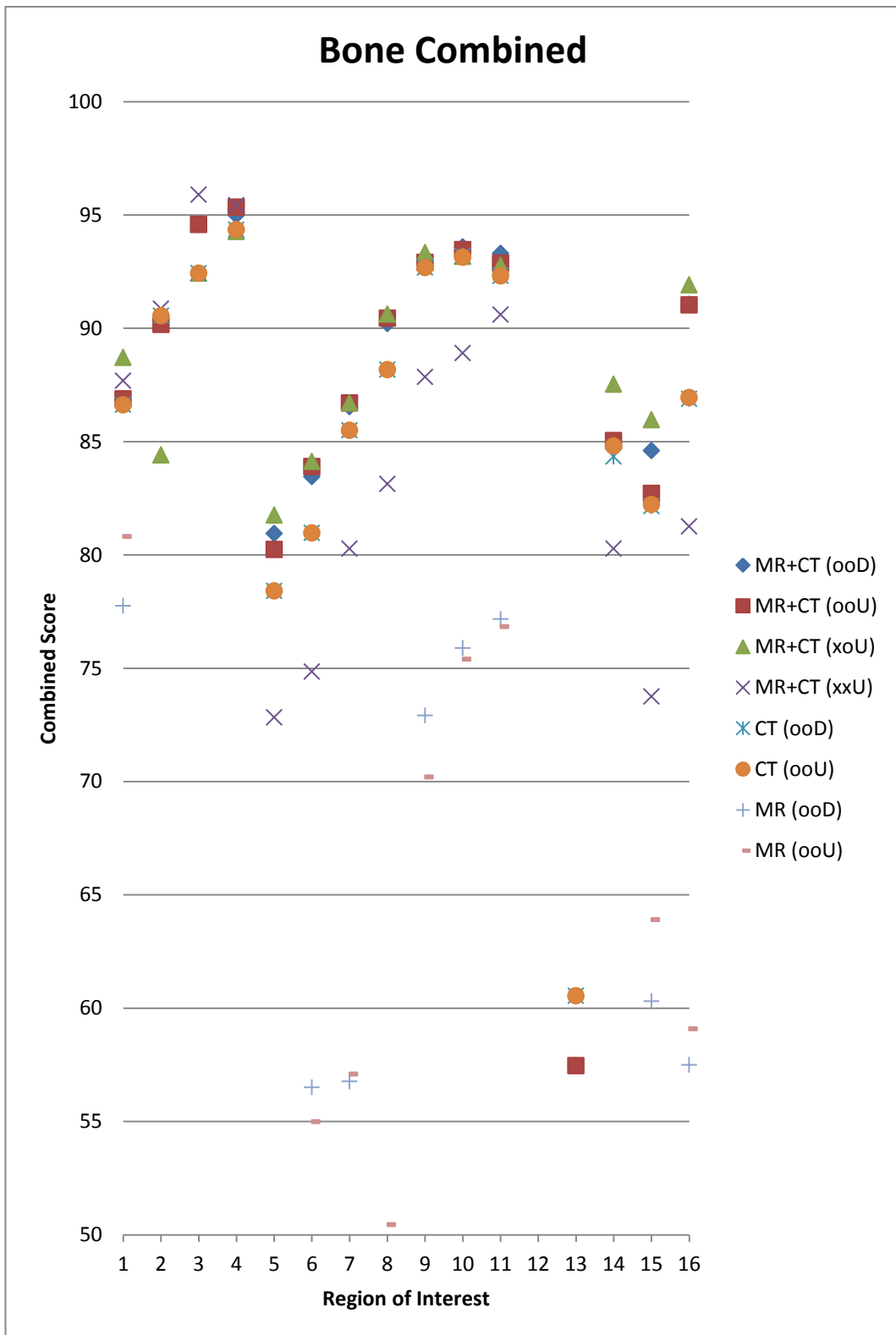
Figure 3.9 Combined metric results for bone in all ROIs
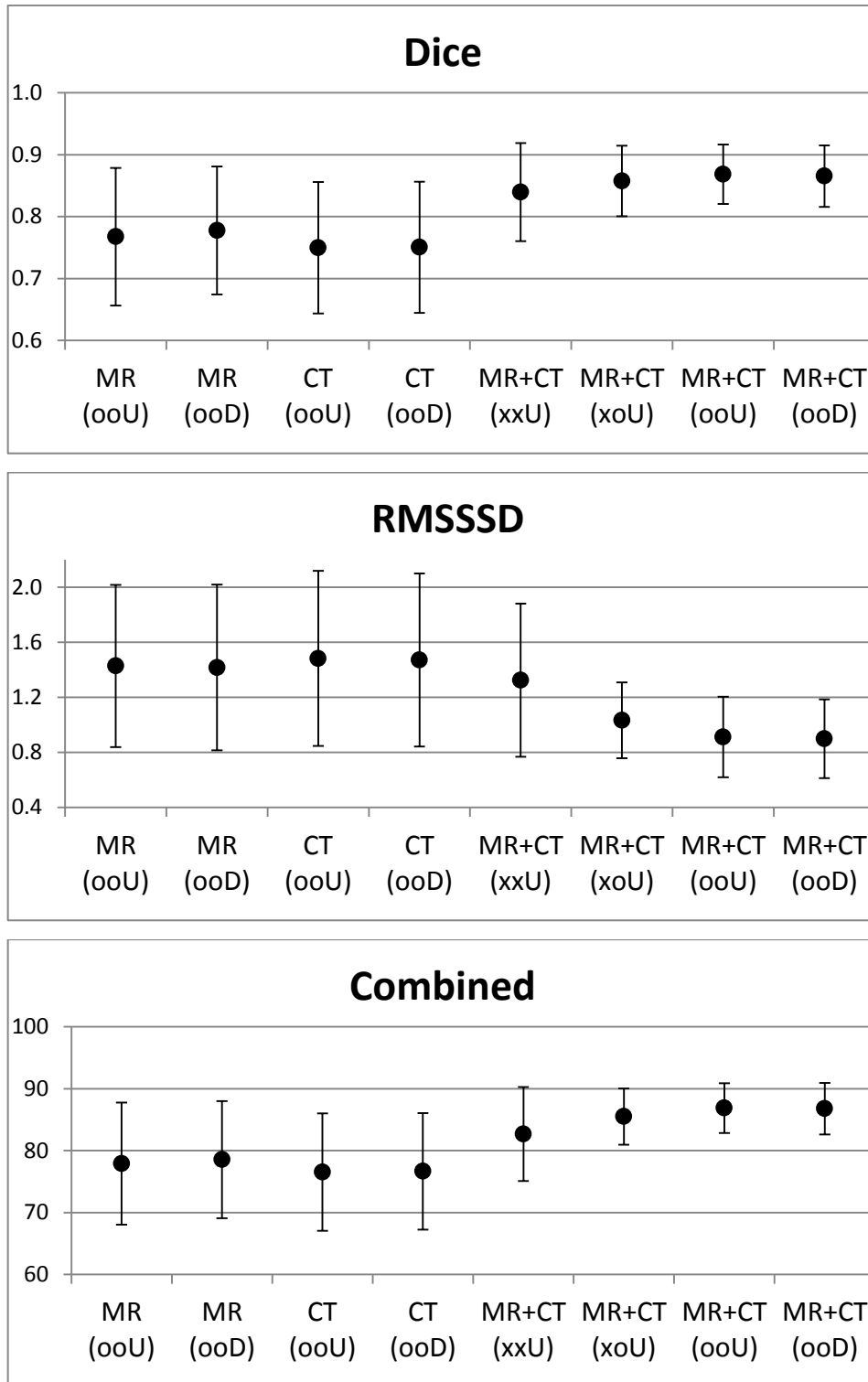
Figure 3.10 Mean and standard deviation plot of different metrics (calculated over all ROIs) for 8 test cases (Muscle class)
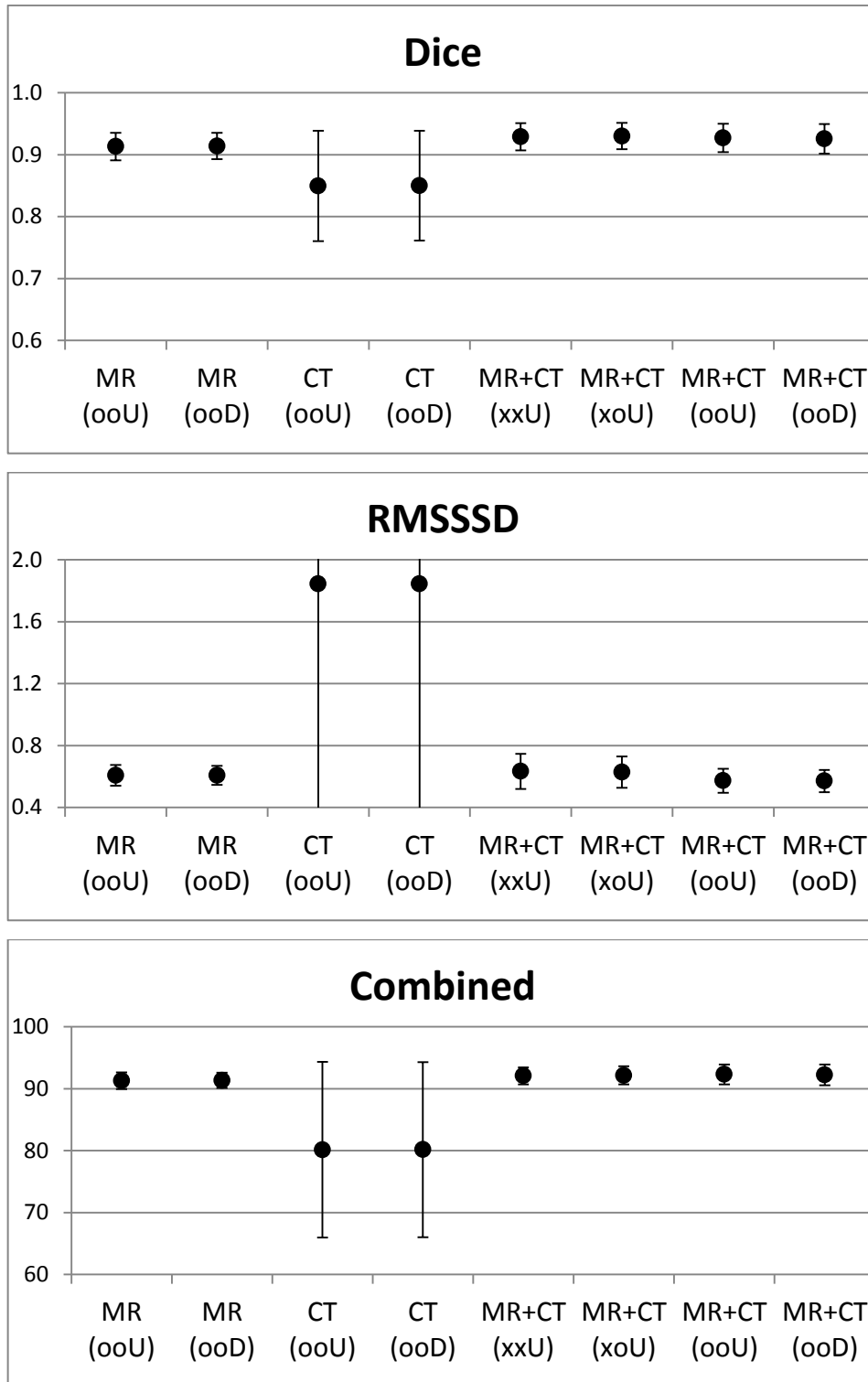
Figure 3.11 Mean and standard deviation plot of different metrics (calculated over all ROIs) for 8 test cases (Fat class)
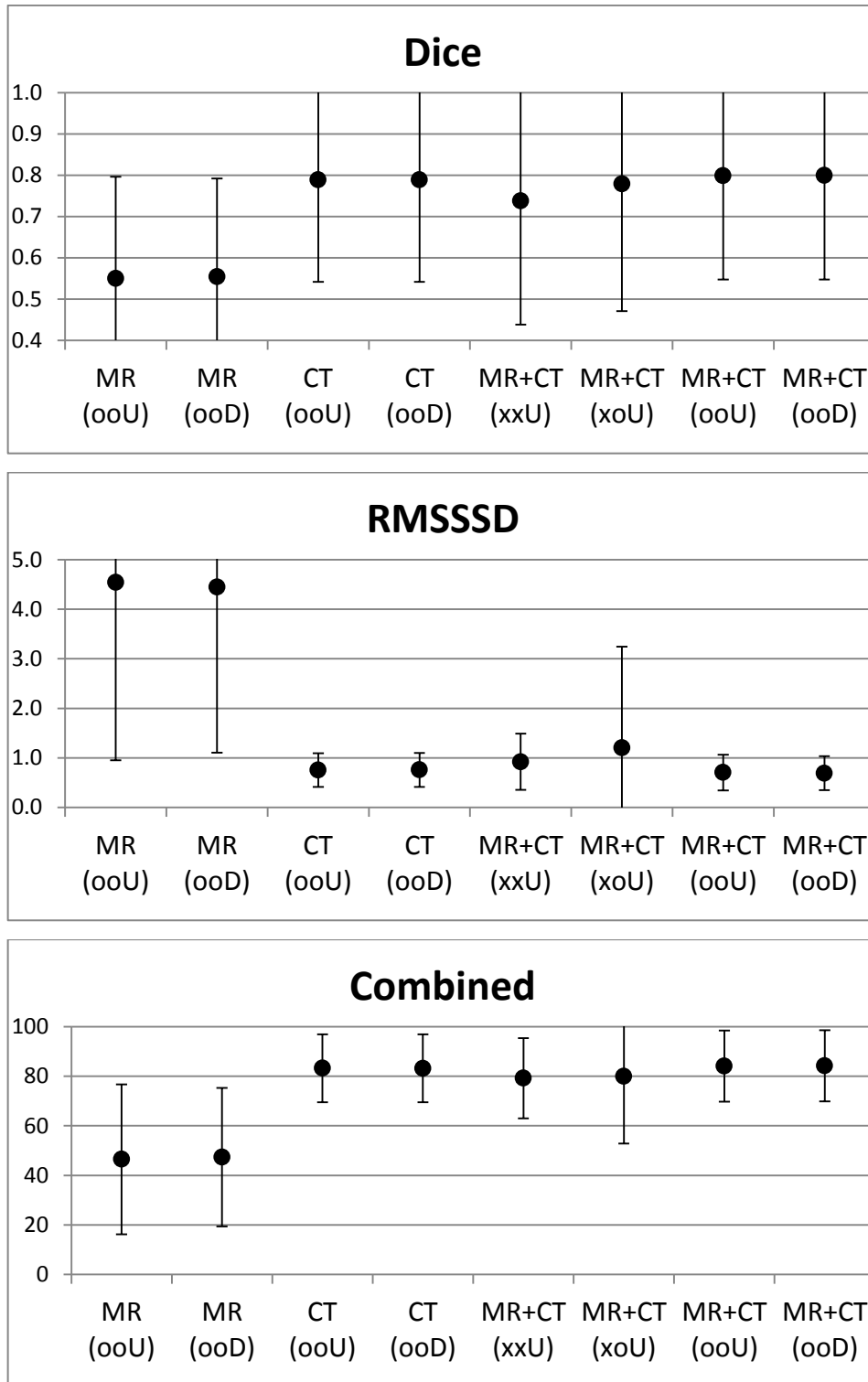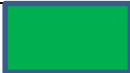
Figure 3.12 Mean and standard deviation plot of different metrics (calculated over all ROIs) for 8 test cases (Bone class)

Figure 3.10 and Figure 3.11 show that CT-only cases (cases 3 and 4) yields the worst muscle and fat segmentation results. MR-only case (cases 1 and 2) results are better because of the better soft tissue contrast of the modality. CT-MR fusion (cases from 5 to 8) resulted in progressively better scores for muscle and fat as PV and resolution models are added. For the bone class, MR-only cases performed the worst as expected. CT images are superior in distinguishing bone tissues from other tissues and this can be observed by the higher scores obtained by the CT-only cases. Although MR images have very poor bone-air contrast, they unexpectedly contribute to the bone scores in the fusion cases.

Although numerical results are important for the analysis of the method, visual observation is also critical for the algorithm accuracy assessment. In that regard visual inspections on the regions of interest are made. The tissue classes are shown as coloured overlays on the original CT and MR images. The colours corresponding to each tissue class are shown in Table 3.3 and a sample segmentation result is shown in Figure 3.13.

Table 3.3 Segmentation labels and their colors

Muscle	
Bone	
Skin	
Air	
Fat	
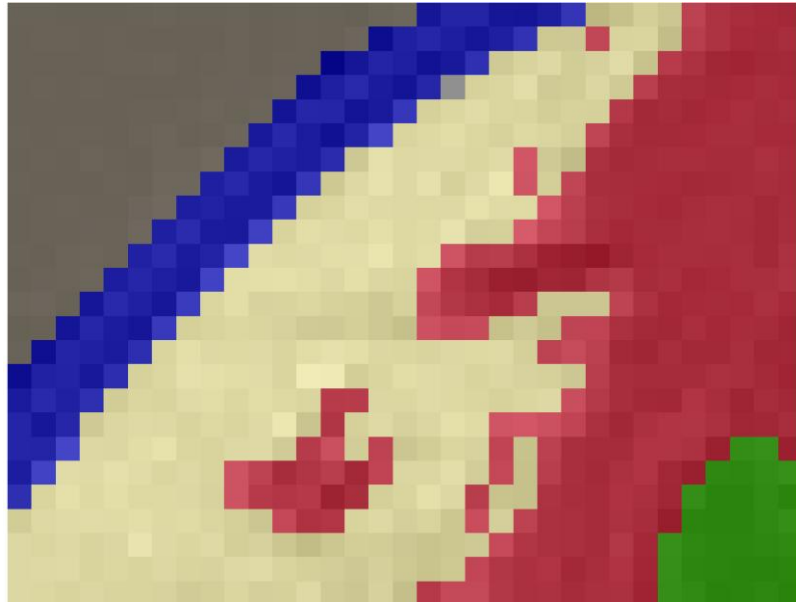
Figure 3.13 Sample ground truth slice colored with segmentation labels

(a)

(b)

(c)

(d)

Figure 3.14 ROI 6 original (a) CT and (b) MR; ground truth overlay on original
(c) CT and (d) MR

Figure 3.15 ROI 6 Segmentation results for the cases 1 through 8. (a) MR (ooU),
(b) MR (ooD), (c) CT (ooU) (d) CT (ooD), (e) Fusion (xxU), (f) Fusion (xoU), (g)
Fusion (ooU) and (h) Fusion (ooD)

The original CT and MR images and those images with overlays of manually segmented muscle (red) and bone (green) classes are shown in Figure 3.14. Thin muscles and muscle-like tissues exist in this region. The segmentation results for each case listed in Table 3.1 are shown in Figure 3.15. MR only cases (case 1 and 2) have inferior bone segmentation accuracy and mostly inaccurate segmentations of bone-air regions. The CT only cases (3 and 4) provide superior bone-air segmentation results, but are unable to identify big parts of thin muscle structures. In the absence of resolution model (blurring matrix), although CT and MR fusion (the cases 5 and 6) yields improved results over the cases 1-4, there are still big gaps between thin muscle parts which should form a single structure according to the ground truth segmentation. The cases 7 and 8 for which the blurring matrix is non-identity are quite successful in connecting the previously disconnected very thin muscle structures. The gap between them is reduced to a single pixel.

(a) (b)

(c) (d)
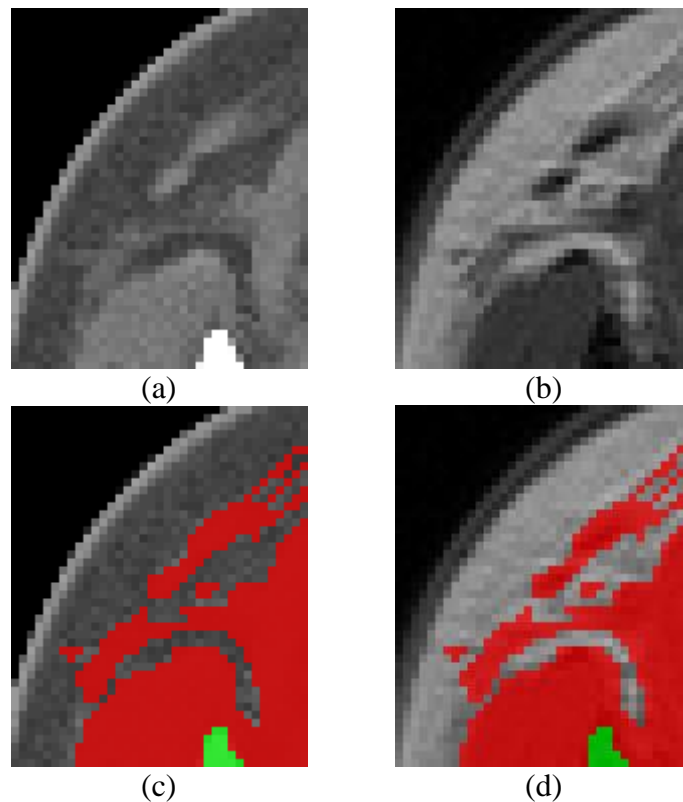
Figure 3.16 ROI 3 original (a) CT and (b) MR; ground truth overlay on original (c) CT and (d) MR
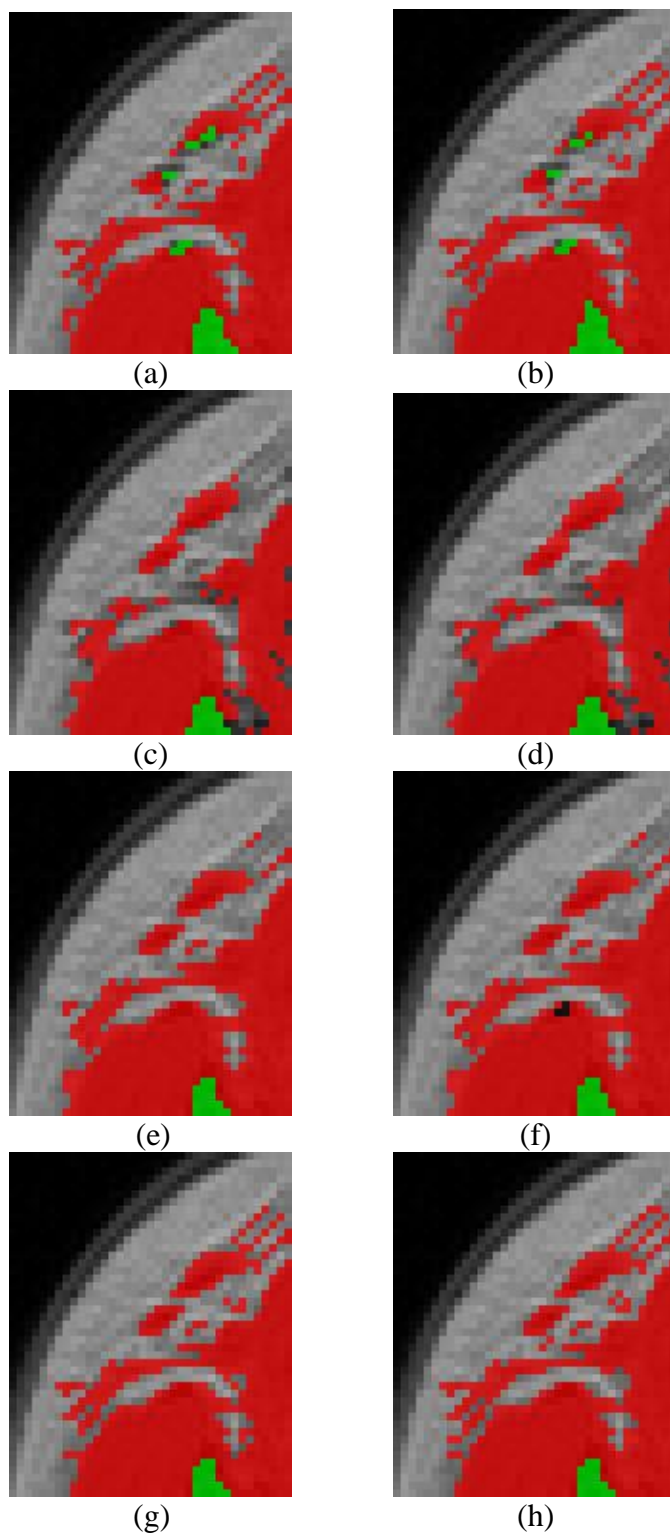
Figure 3.17 ROI 3 segmentation results for the cases 1 through 8. (a) MR (ooU), (b) MR (ooD), (c) CT (ooU) (d) CT (ooD), (e) Fusion (xxU), (f) Fusion (xoU), (g) Fusion (ooU) and (h) Fusion (ooD)

Smaller structures between the main thin muscle structure and bone are partially segmented in these cases. Finally, when case 7 (ooU) and case 8 (ooD) are compared, the latter seems to provide better connectivity of a very thin muscle structure, but on the other hand produced some spurious thin muscle pixels that don't exist in the ground truth.

The segmentation results in ROI 3 show a similar trend to the results of ROI 6 (Figure 3.16 and Figure 3.17). The thin muscle strips on the top right portion of the image are not segmented accurately in cases 1 through 6. The accuracy is enhanced with the addition of blurring matrix to the model (cases 7 and 8). Cases 7 (ooU) and 8 (ooD) are almost identical, which suggests that prior type did not make any difference.

(a)   (b)
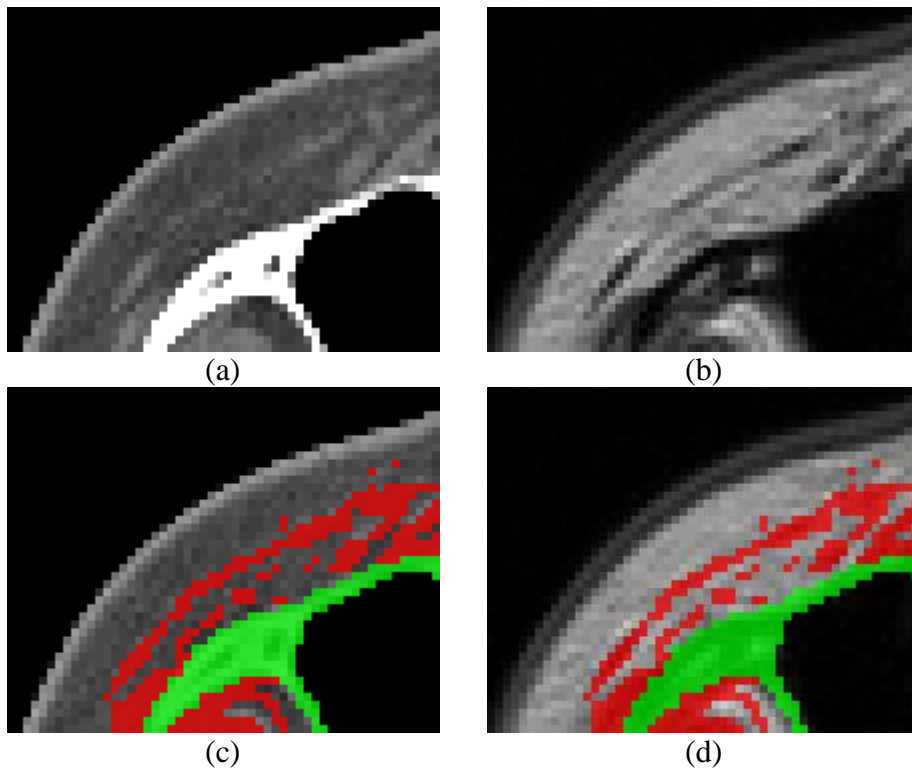
(c)   (d)
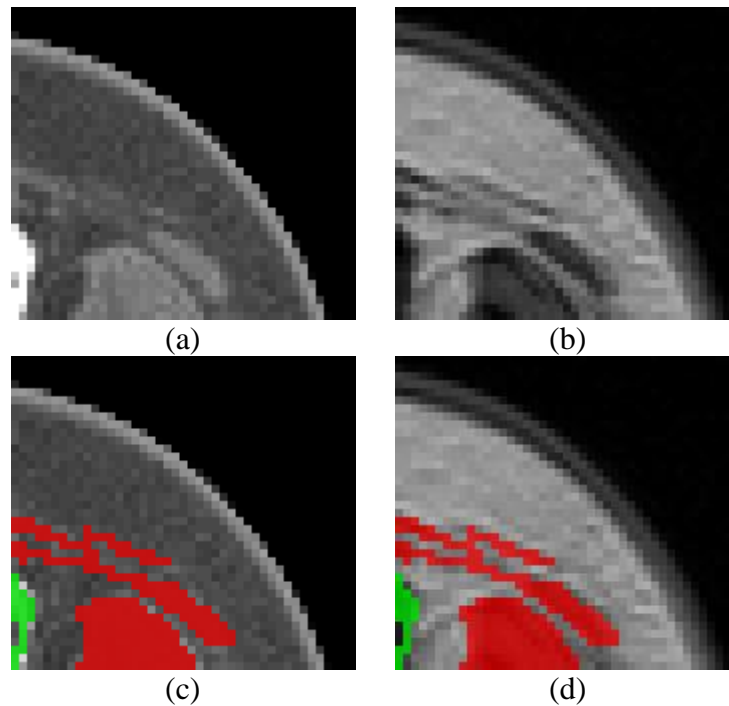
Figure 3.18 ROI 8 original (a) CT and (b) MR; ground truth overlay on original (c) CT and (d) MR
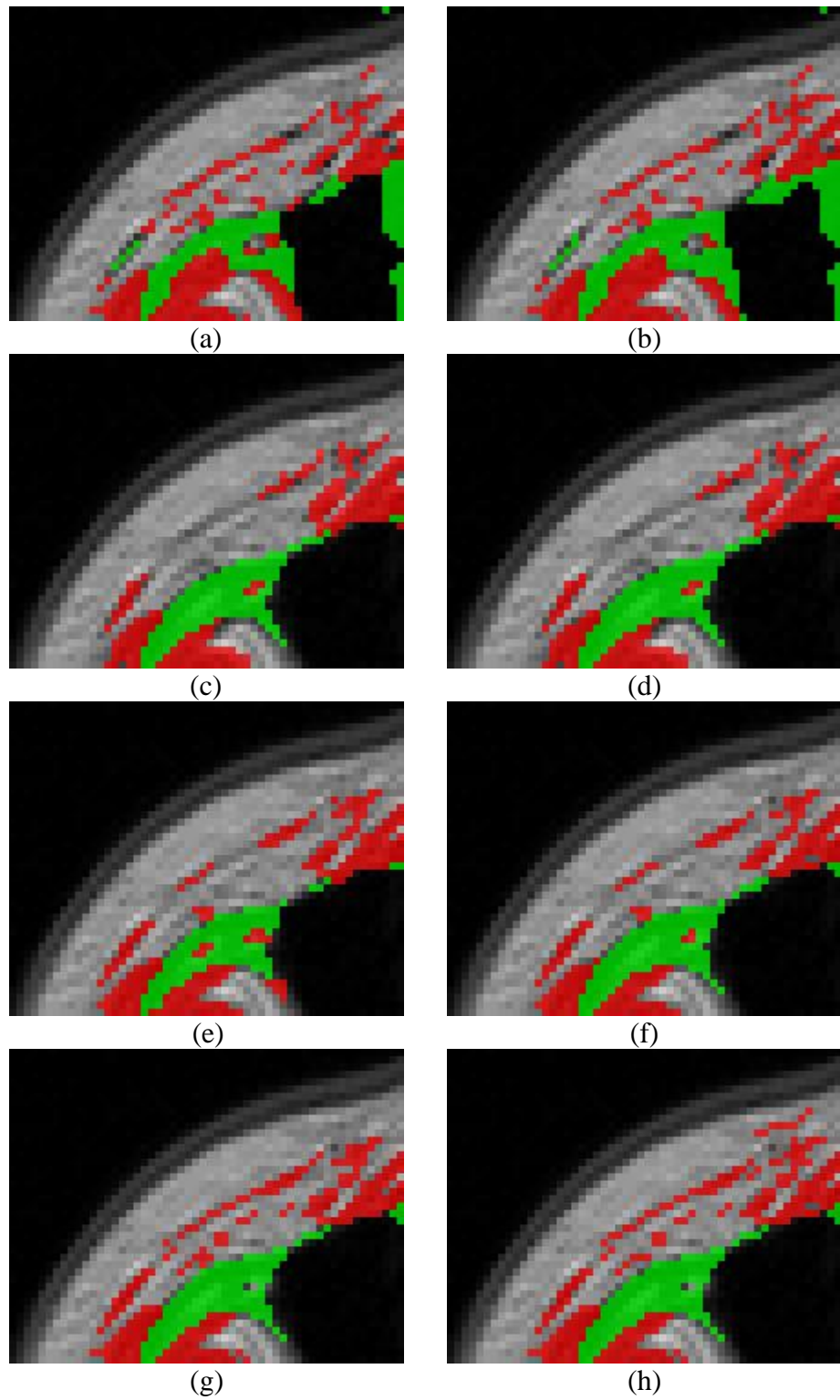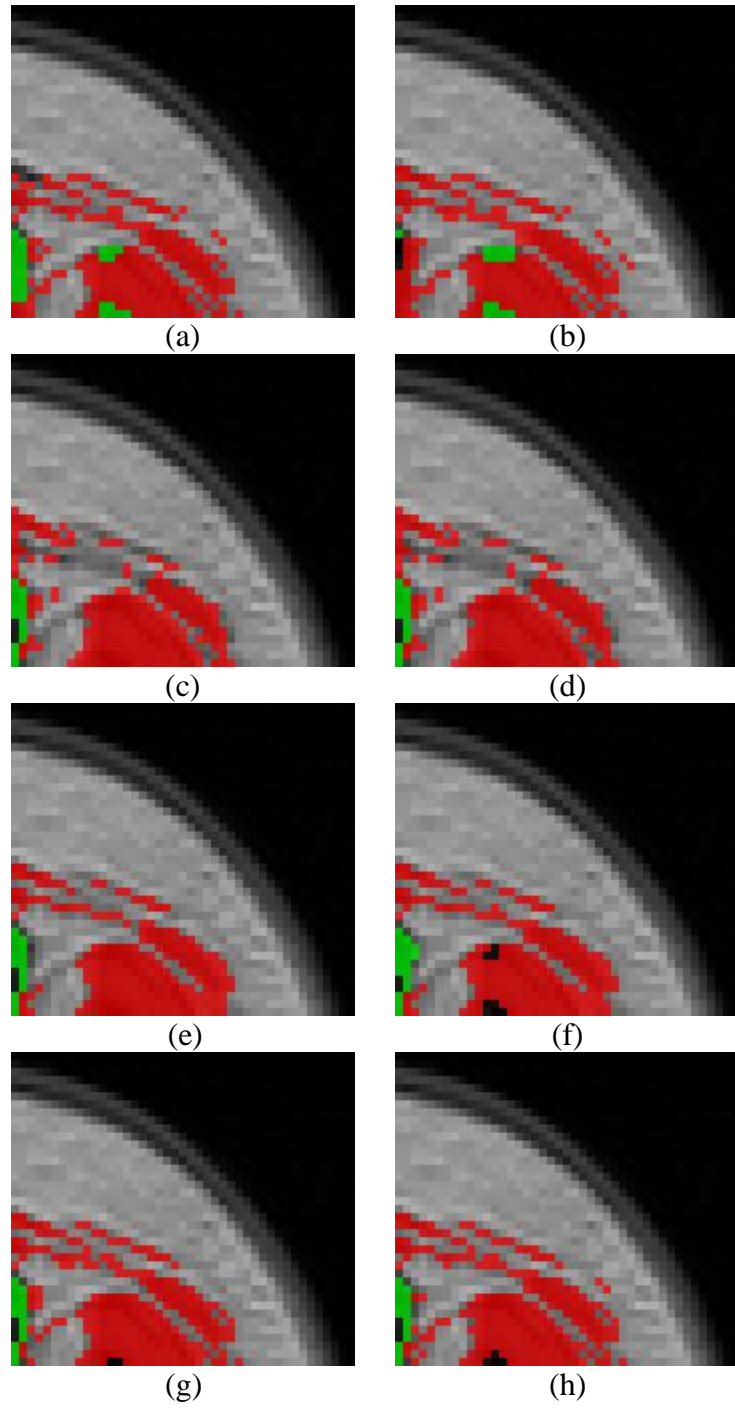
Figure 3.19 ROI 8 segmentation results for the cases 1 through 8. (a) MR (ooU),
(b) MR (ooD), (c) CT (ooU) (d) CT (ooD), (e) Fusion (xxU), (f) Fusion (xoU), (g)
Fusion (ooU) and (h) Fusion (ooD)

In ROI 8 (Figure 3.18 and Figure 3.19), although cases 6, 7 and 8 progressively yield better connectivity of very thin muscles, some over-segmentations of muscle occur on the outer regions of the thin muscles and near the bone. This results in very similar numerical scores among cases 6, 7 and 8.

3.3 Statistical Analysis of the Results

For statistical comparison, the normality of the distribution of the combined scores averaged over all 16 ROIs is checked using Shapiro-Wilk test. The calculated p values for all 8 cases are given in Table 3.4.

Table 3.4 Results of Shapiro-Wilk normality test

Case Index	Case Name	p
1	MR (ooU)	0.581
2	MR (ooD)	0.791
3	CT (ooU)	0.370
4	CT (ooD)	0.330
5	Fusion (xxU)	0.629
6	Fusion (xoU)	0.905
7	Fusion (ooU)	0.266
8	Fusion (ooD)	0.469

All methods are found to have normal distributions ($p > 0.05$), enabling the use of ANOVA test. The significant differences between the means of methods are calculated with this test. Although means of cases 7 and 8 are slightly better than case 6, ANOVA test revealed that these three cases do not have significant difference from each other. The comparison of muscle class scores are given in Figure 3.20.

In spite of these statistical results, by qualitative analysis of ROI 3, 6 and 8, it is revealed that case 7 (ooD) has provided some very thin muscle pixels which were not labelled so in the ground truth labelling by the expert. In close analysis, those regions demonstrate in fact a presence of very subtle thin muscle-like structures (a continuous strip of pixels with different intensity than fat). It is suspected that this could be misjudgement by the expert, since the expert was not trained to take into account the effect of system resolution (Experts look for a clear separation of intensity from that of fat.). Therefore, the labelling of ground truth can be considered questionable because of expert's knowledge of the effect of system resolution (and PV effect combined). Another pitfall of this evaluation is obviously not consulting a group of experts as opposed to just one expert.
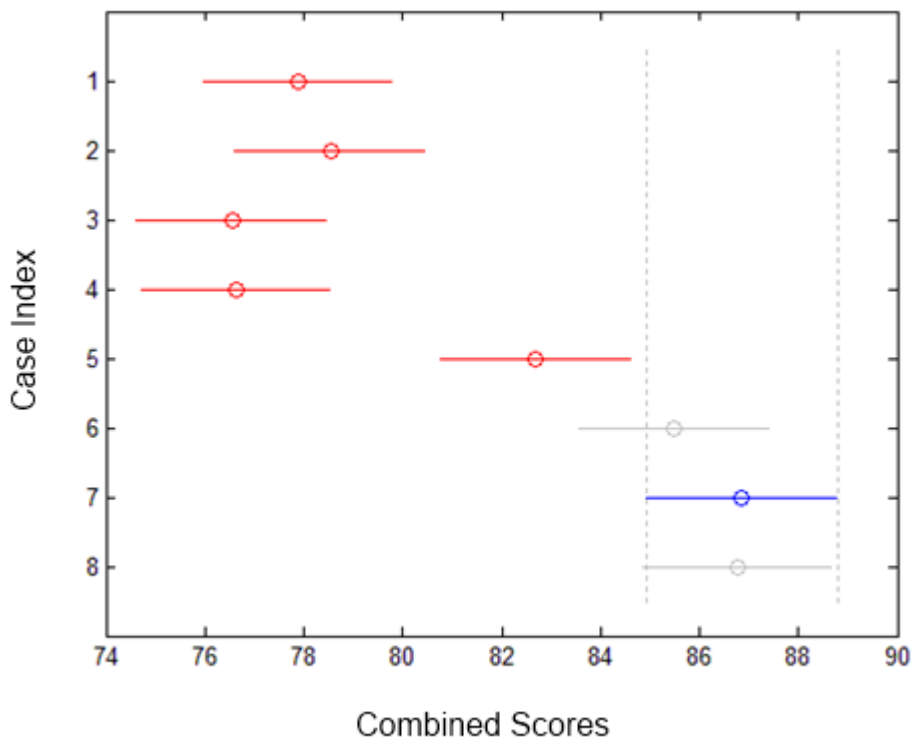
Figure 3.20 Bar chart representation of muscle scores. Bars represent means and standard deviations

3.4 Computational Cost

There are various factors affecting the computation time of the method. The most important ones are the size of the image and the number of partial volumes used. The times spent at different steps of the method are given in Table 3.5. It takes 10 iterations on average for the method to converge to a solution. For the standard fusion case, the segmentation of five slices of a 50x50 ROI takes approximately 5 seconds. The computation time increases linearly with image size. Therefore, it can be assumed that five slices of a 256x256 would take approximately two minutes.

Although the method cannot be called fast, it should be noted that the speed of the algorithm is bound to increase by optimization of the MATLAB code or code porting from MATLAB to C++.

Table 3.5 Computation times of the steps of the algorithm for five slices of 50x50 pixel image on a 2.27 GHz i5 processor

Type of process	Time	Number of times the process is repeated
Computation of orientations	0.23 s	1
Calculation of f_{nc}	0.17 s	Iteration count
Calculation of f_c	0.02 s	(Iteration count) x (Total number of labels)
Postprocesses	0.14 s	1

CHAPTER 4

CONCLUSION, DISCUSSION AND FUTURE WORK

A method to segment the human head facial tissues and especially thin muscles was developed. The Bayesian segmentation model which includes MR-CT fusion, blurring and noise models was formulated and tested. The effect of each model was investigated. Although the greatest contribution was made by the addition of fusion, the other models such as partial volume model, blurring matrix and prior models helping to reduce image noise were all enhancements to the segmentation accuracy. Two models of MRF priors were also tested. There was not a significant difference between isotropic and directional prior cases. Therefore, using isotropic prior is more advantageous since there are fewer parameters to optimize.

Although the quantitative and statistical analysis yielded very similar results for the uniform and directional prior cases, thin muscle structures in the segmentation result which did not exist in the ground truth were observed in the qualitative analysis. These may simply be segmentation inaccuracies or they can be the result of an inaccurate ground truth. In either case, the method should also be tested by ground truth segmentations made by multiple experts instead of segmentations made by a single expert.

A similar comparison can also be made for the isotropic prior case of this study and the adaptive Bayesian case in [15]. The results of [15] were obtained in 3D. This makes it possible that continuity of thin structures in the axial direction is

preserved better. The images were segmented in a different manner here slice by slice. Extension of the algorithm presented here to the third dimension may increase the segmentation accuracy. The adaptive Bayesian method proposed by Kale *et al.* [15] achieved quite similar results as presented here (the accuracies of the two methods are within *2%* of each other on combined scores). However, Emre Kale *et al.* [15] requires more parameters to be tuned which are medical imaging system and data acquisition setting dependant.

Although the computation time is greater here when compared to [15], it can be improved by better code implementations. If the MATLAB code is implemented emphasizing MATLAB's matrix computation strengths or if the algorithm is re-coded in a programming language such as C++, the performance should increase by orders of magnitude. Even more speed enhancement is possible through the use of parallel processing with multi-threading or libraries like Nvidia Compute Unified Device Architecture (CUDA) [64] which use the parallel processing power of Graphics Processing Units (GPU).

Although it is not a standard practice to have both CT and MR scans of the same patient, the proposed method should be tested on more subjects. Having more datasets would also mean that some of them could be used as training sets for more accurate parameter optimization (as opposed to simulation). In addition, the algorithm can be tested with other imaging modalities and MR imaging sequences (T2, spin-echo, etc.). However, it should be noted that using a modality other than CT may result in loss of information regarding bone and air. The segmentation success of using low-dose CT as opposed to a normal dose should also be tested when radiation is a concern, especially in the case of a full head scan.

The calculations were carried out with normalized CT intensities instead of Hounsfield units (HU). Ideally, it would be better to use HU, which would be system independent. The intensity values in Hounsfield units are in similar ranges

on different CT settings and images. Therefore, it is possible to generalize the mean tissue intensities.

The autocorrelation of the CT and MR images were obtained by a manual method, and this process needs to be repeated for imaging systems and data acquisition settings other than the ones used in this study. This process can be switched for a more robust method. The means of class labels and their standard deviations which were obtained similarly can be calculated by methods such as EM algorithm.

The estimation model of the resolution σ is assumed to be constant throughout the volume, where in fact it is typically variable in the field of view for spiral CT systems [65]. However, the results presented here indicate that even a constant σ assumption provided improvements in the segmentation output.

The methods and tests can be carried over to the 3D to include the data coherence in the axial direction. In that case, the computational cost would increase, but not dramatically since the number of processed voxels would be the same. Only the kernel sizes convolved around the voxels would change.

The bias correction, image registration and pre-processing steps are also important for the accuracy of the segmentation algorithm when multimodal fusion is used. Improvements on the pre-processing methods are expected to improve the segmentation performance.

Partial volume ratio estimates at each pixel are a natural result of the method. Validation in terms of the partial volume ratios can be performed if these estimates are also available for the ground truth segmentation as in the simulated images.

The advantages and disadvantages of using orientation tensors over Hessian tensors in the directional prior should be investigated. Orientation tensors have been used in various recent studies on the segmentation of thin structures, and they may have a positive effect on the method.

Automatic segmentation methods almost never yield perfect results. Perfect result itself is not easily quantifiable because the interpretation of a medical image can be subjective to an extent. Even when an agreement on what ground truth segmentation should be is reached, manual correction (called "editing") of the segmentation results is usually required. A segmentation editing method which suits the thin structures may decrease the time spent on this step dramatically. A tool may be needed even in the case of accurate automatic segmentation if the separation of tissue subtypes such as different muscles and the other tissue types categorized by the method as muscle class (i.e. veins, vessels, tendon, etc.) is required. The performance of the automatic segmentation is still one of the most critical factors for the subsequent manual editing methods in decreasing the time spent by the experts.

REFERENCES

[1]   H. P. Ng, S. H. Ong, S. Huang, J. Liu, K. W. C. Foong, P. S. Goh and W. L. Nowinski, "Salient features useful for the accurate segmentation of masticatory muscles from minimum slices subsets of magnetic resonance images," *Machine Vision and Applications,* vol. 21, no. 4, pp. 449-467, 2008.

[2]   E. Gladilin, S. Zachow, P. Deuflhard and H. C. Hege, "Anatomy- and physics-based facial animation for craniofacial surgery simulations," *Medical & Biological Engineering & computing,* vol. 42, no. 2, pp. 167-70, 2004.

[3]   E. Sifakis and A. Selle, "Simulating speech with a physics-based facial muscle model," *Proceedings of the ACM SIGGRAPH/Eurographics Symposium On Computer Animation,* pp. 261-270, 2006.

[4]   Y. Zhang, E. C. Prakash and E. Sung, "A new physical model with multilayer architecture for facial expression animation using dynamic adaptive mesh," *IEEE Transactions On Visualization and Computer Graphics,* vol. 10, no. 3, pp. 339-52, 2004.

[5]   M. E. Farrugia, G. M. Bydder, J. M. Francis and M. D. Robson, "Magnetic resonance imaging of facial muscles," *Clinical radiology,* vol. 62, no. 11, pp. 1078-86, 2007.

[6]   C. Huisinga-Fischer, "Precision and accuracy of CT-based measurements of masticatory muscles in patients with hemifacial microsomia," *Dentomaxillofacial Radiology,* vol. 33, no. 1, pp. 12-16, 2004.

[7]   W. Mollemans, F. Schutyser, N. Nadjmi, F. Maes and P. Suetens, "Predicting soft tissue deformations for a maxillofacial surgery planning system: from computational strategies to a complete clinical validation," *Medical Image Analysis,* vol. 11, no. 3, pp. 282-301, 2007.

[8]   P. C. Benington, J. E. Gardener and N. P. Hunt, "Masseter muscle volume measured using ultrasonography and its relationship with facial morphology," *European Journal of Orthodontics,* vol. 21, no. 6, pp. 659-70, 1999.

[9] M. Meehan, M. Teschner and S. Girod, "Three-dimensional simulation and prediction of craniofacial surgery," *Orthod Craniofac Res. Suppl 1,* vol. 6, pp. 102-7, 2003.

[10] S. De Greef and G. Willems, "Three-dimensional cranio-facial reconstruction in forensic identification: latest progress and new tendencies in the 21st century," *Journal of forensic sciences,* vol. 50, no. 1, pp. 12-7, 2005.

[11] C. N. Stephan, "The human masseter muscle and its biological correlates: A review of published data pertinent to face prediction," *Forensic Science International,* vol. 201, no. 1-3, pp. 153-9, 2010.

[12] K. Zepa, I. Urtane, Z. Krisjane and G. Krumina, "Three-dimensional assessment of musculoskeletal features in Class II and Class III patients," *Stomatologija,* vol. 11, no. 1, pp. 15-20, 2009.

[13] R. Olszewski, M. B. Villamil, D. G. Trevisan, L. P. Nedel, C. M. D. S. Freitas, H. Reychler and B. Macq, "Towards an integrated system for planning and assisting maxillofacial orthognathic surgery," *Computer Methods and Programs in Biomedicine,* vol. 91, no. 1, pp. 13-21, 2008.

[14] J. M. Plooij, T. J. Maal, P. Haers, W. A. Borstlap, A. M. Kuijpers-Jagtman and S. J. Bergé, "Digital three-dimensional image fusion processes for planning and evaluating orthodontics and orthognathic surgery. A systematic review," *International Journal of Oral and Maxillofacial Surgery,* vol. 40, no. 4, pp. 341-52, 2011.

[15] E. H. Kale, E. U. Mumcuoglu and S. Hamcan, "Automatic segmentation of human facial tissue by MRI-CT fusion: A feasibility study," *Computer Methods and Programs in Biomedicine,* p. In press (available online), 2012.

[16] D. W. Shattuck, S. R. Sandor-Leahy, K. A. Schaper, D. A. Rottenberg and R. M. Leahy, "Magnetic resonance image tissue classification using a partial volume model," *NeuroImage,* vol. 13, no. 5, pp. 856-76, 2001.

[17] N. R. Pal and S. K. Pal, "A review on image segmentation techniques," *Pattern Recognition,* vol. 26, no. 9, pp. 1277-1294, 1993.

[18] D. J. Withey and Z. J. Koles, "A review of medical image segmentation: methods and available software," *International Journal of Bioelectromagnetism,* vol. 10, no. 3, pp. 125-148, 2008.

[19] R. M. Haralick and L. G. Shapiro, "Image segmentation techniques," *Computer Vision, Graphics, and Image Processing,* vol. 29, no. 1, pp. 100-132, 1985.

[20] B. Sankur and M. Sezgin, "Image thresholding techniques: A survey over categories," *Pattern Recognition,* 2001.

[21] B. Dogdas, D. Shattuck and R. M. Leahy, "Segmentation of skull in 3D human MR images using mathematical morphology," *Human Brain Mapping,* vol. 26, no. 4, pp. 273-285, 2005.

[22] S. Beucher, "The watershed transformation applied to image segmentation," *Scanning Microscopy International,* vol. 6, pp. 299-314, 1991.

[23] R. Adams and L. Bischof, "Seeded region growing," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 16, no. 6, pp. 641-647, 1994.

[24] N. Noguchi and M. Goto, "Computer simulation system for orthognathic surgery," *Orthodontics & Craniofacial Research,* vol. 6 Suppl 1, pp. 176-8, 2003.

[25] R. Nock and F. Nielsen, "On weighting clustering," *IEEE Transactions On Pattern Analysis and Machine Intelligence,* vol. 28, no. 8, pp. 1223-35, 2006.

[26] S. Banerjee, D. Mukherjee and D. Dutta Majumdar, "Fuzzy c-means approach to tissue classification in multimodal medical imaging," *Information Sciences,* vol. 115, no. 1-4, pp. 261-279, 1999.

[27] D. L. Pham, "Unsupervised tissue classification in medical images using edge-adaptive clustering," *Engineering in Medicine and Biology Society,* vol. 12, no. 2, pp. 634-637, 2003.

[28] S. Osher and R. Fedkiw, Level set methods and dynamic implicit surfaces, New York: Springer-Verlag, 2002.

[29] R. Fedkiw and F. Gioub, "A fast hybrid k-Means level set algorithm for

segmentation," *4th Annual Hawaii International Conference on Statistics and Mathematics,* pp. 281-291, 2005.

[30] M. Kass, A. Witkin and D. Terzopoulos, "Snakes: Active Contour Models," *International Journal of Computer Vision,* vol. 1, no. 4, pp. 321-331, 1988.

[31] J. Anthony Yezzi, A. Tsai and A. Willsky, "Medical image segmentation via coupled curve evolution equations with global constraints," *IEEE Workshop On Mathematical Methods in Biomedical Image Analysis,* pp. 12-19, 2000.

[32] C. Samson, L. Blanc-Féraud, G. Aubert and J. Zerubia, "Two variational models for multispectral image classification," *Lecture Notes in Computer Science,* vol. 2134, pp. 344-356, 2001.

[33] M. Holtzman-Gazit, R. Kimmel, N. Peled and D. Goldsher, "Segmentation of thin structures in volumetric medical images," *IEEE Transactions On Image Processing,* vol. 15, no. 2, pp. 354-63, 2006.

[34] P. A. Yushkevich, J. Piven, H. C. Hazlett, R. G. Smith, S. Ho, J. C. Gee and G. Gerig, "User-guided 3D active contour segmentation of anatomical structures: significantly improved efficiency and reliability," *NeuroImage,* vol. 31, no. 3, pp. 1116-28, 2006.

[35] Y. Kang, K. Engelke and W. A. Kalender, "Interactive 3D editing tools for image segmentation," *Medical Image Analysis,* vol. 8, no. 1, pp. 35-46, 2004.

[36] R. Olszewski, Y. Liu, T. Duprez, T. M. Xu and H. Reychler, "Three-dimensional appearance of the lips muscles with three-dimensional isotropic MRI: in vivo study," *International Journal of Computer Assisted Radiology and Surgery,* vol. 4, no. 4, pp. 349-52, 2009.

[37] D. Gering, A. Nabavi, R. Kikinis, W. Grimson, N. Hata, P. Everett, F. Jolesz and W. Wells III, "An integrated visualization system for surgical planning and guidance using image fusion and interventional imaging," *International Conference of Medical Image Computing and Computer Assisted Intervention,* vol. 2, pp. 809-819, 1999.

[38] W. A. Barrett and E. N. Mortensen, "Interactive live-wire boundary

extraction," *Medical Image Analysis,* vol. 1, no. 4, pp. 331-41, 1997.

[39] E. Sener, E. U. Mumcuoglu, K. E. H and A. V, "BT ve MR görüntülerini 3 boyutta çakıştırarak ve birlikte kullanarak, otomatik ve yarı-otomatik bölütleme yöntemleri geliştirilmesi ve bu yöntemlerin kafa modeli oluşturma ve yüz mimik kaslarının bölütlenmesine uygulanması," *5.National Conference of Medical Informatics,* 2008.

[40] H. Rue and L. Held, Gaussian Markov random fields: theory and applications, CRC Press, 2005.

[41] J. P. Chiverton and K. Wells, "Adaptive partial volume classification of MRI data," *Physics In Medicine and Biology,* vol. 53, no. 20, pp. 5577-94, 2008.

[42] Y. Rezaeitabar and I. Ulusoy, "Automatic 3D segmentation of individual facial muscles using unlabeled prior information," *International Journal of Computer Assisted Radiology and Surgery,* vol. 7, no. 1, pp. 35-41, 2012.

[43] Y. Zhang, M. Brady and S. Smith, "Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm," *IEEE Transactions On Medical Imaging,* vol. 20, no. 1, pp. 45-57, 2001.

[44] T. Tasdizen, D. Weinstein and J. N. Lee, "Automatic Tissue Classification for the Human Head from Multispectral MRI," *Technical Report,* 2004.

[45] M. Jiang, G. Wang, M. W. Skinner, J. T. Rubinstein and M. W. Vannier, "Blind deblurring of spiral CT images," *IEEE Transactions on Medical Imaging,* vol. 22, no. 7, pp. 837-845, 2003.

[46] D. H. Laidlaw, K. W. Fleischer and A. H. Barr, "Partial-volume Bayesian classification of material mixtures in MR volume data using voxel histograms," *IEEE Transactions On Medical Imaging,* vol. 17, no. 1, pp. 74-86, 1998.

[47] M. A. Hurn, K. V. Mardia, T. J. Hainsworth, J. Kirkbride and E. Berry, "Bayesian fused classification of medical images," *IEEE Transactions on Medical Imaging,* vol. 15, no. 6, pp. 850-858, 1996.

[48] S. Ruan, C. Jaggi, J. Xue, J. Fadili and D. Bloyet, "Brain tissue classification of magnetic resonance images using partial volume modeling," *IEEE Transactions On Medical Imaging,* vol. 19, no. 12, pp. 1179-87, 2000.

[49] P. P. Wyatt and J. A. Noble, "MAP MRF joint segmentation and registration of medical images," *Medical Iimage Analysis,* vol. 7, no. 4, pp. 539-52, 2003.

[50] K. Van Leemput, F. Maes, D. Vandermeulen and P. Suetens, "A unifying framework for partial volume segmentation of brain MR images," *IEEE Transactions On Medical Imaging,* vol. 22, no. 1, pp. 105-19, 2003.

[51] W. C. K. Wong, A. C. S. Chung and S. C. H. Yu, "Local orientation smoothness prior for vascular segmentation of angiography," *Lecture Notes In Computer Science,* vol. 3022, pp. 353-365, 2004.

[52] M. Descoteaux, M. Audette, K. Chinzei and K. Siddiqi, "Bone enhancement filtering: application to sinus bone segmentation and simulation of pituitary surgery," *Computer Aided Surgery,* vol. 11, no. 5, pp. 247-55, 2006.

[53] A. F. Frangi, W. J. Niessen, K. L. Vincken and M. A. Viergever, "Multiscale vessel enhancement filtering," *Medical Image Computing and Computer-Assisted Intervention, Lecture Notes in Computer Science,* vol. 1496, pp. 130-7, 1998.

[54] Y. Sato, S. Nakajima, N. Shiraga, H. Atsumi, S. Yoshida, T. Koller, G. Gerig and R. Kikinis, "3D multi-scale line filter for segmentation and visualization of curvilinear structures in medical images," *Medical Image Analysis,* vol. 2, no. 2, pp. 143-168, 1998.

[55] C. Lorenz, I. Carlsen, T. M. Buzug, C. Fassnacht and J. Weese, "Multi-scale line segmentation with automatic estimation of width, contrast and tangential direction in 2D and 3D medical images," *Proceedings of the First Joint Conference on Computer Vision, Virtual Reality and Robotics in Medicine and Medial Robotics and Computer-Assisted Surgery,* pp. 233-242, 97.

[56] C.-F. Westin, A. Bhalerao, H. Knutsson and R. Kikinis, "Using local 3D structure for segmentation of bone from computer tomography images,"

Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 794-800.

[57] J. Sled, A. Zijdenbos and A. Evans, "A nonparametric method for automatic correction of intensity nonuniformity in MRI data," *IEEE Transactions on Medical Imaging,* vol. 17, no. 1, pp. 87-97, 1998.

[58] S. Geman and G. D, "Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 6, no. 6, pp. 721-741, 1984.

[59] M. Elad and A. Feuer, "Restoration of a single superresolution image from several blurred, noisy, and undersampled measured images," *IEEE Transactions On Image Processing,* vol. 6, no. 12, pp. 1646-1658, 1997.

[60] J. Besag, "On the statistical analysis of dirty pictures," *Journal of the Royal Statistical Society. Series B (Methodological),* vol. 48, no. 3, pp. 259-302, 1986.

[61] J. L. T. L. H. Wang and L. Z, "Penalized weighted least-squares approach to sinogram noise reduction and image reconstruction for low-dose X-ray computed tomography," *IEEE Transactions on Medical Imaging,* vol. 25, no. 10, pp. 1272-83, 2006.

[62] P. Razifar, M. Sandström, H. Schnieder and B. Långström, "Noise correlation in PET, CT, SPECT and PET/CT data evaluated using autocorrelation function: a phantom study on data, reconstructed using FBP and OSEM," *BMC Medical Imaging,* vol. 5, pp. 1-23, 2005.

[63] X. Deng and G. Du, "Editorial : 3D segmentation in the clinic : A grand challenge ii - liver tumor segmentation," *MICCAI Workshop on 3D Segmentation in the Clinic,* 2008.

[64] NVIDIA, "NVIDIA CUDA Parallel Computing Platform," [Online]. Available: http://www.nvidia.com/object/cuda_home_new.html. [Accessed 12 10 2012].

[65] G. Schwarzband and N. Kiryati, "The point spread function of spiral CT,"

Physics in Medicine and Biology, vol. 50, pp. 5307-5322, 2005.

[66] L. Brown, "A survey of image registration techniques," *Computing Surveys,* vol. 24, no. 4, pp. 325-376, 1992.

[67] C. R. J. Maurer and J. M. Fitzpatrick, "A review of medical image registration in Interactive Image-Guided Neurosurgery," in *Interactive Image-Guided Neurosurgery (R. J. Maciunas, ed.)*, Park Ridge, IL: American Association of Neurological Surgeons, 1993, pp. 17-44.

[68] J. B. A. Maintz and M. A. Viergever, "A survey of medical imaging registration," *Medical Image Analysis,* vol. 2, no. 1, pp. 1-36, 1998.

[69] F. Maes, A. Collignon, D. Vandermeulen, G. Marchal and P. Suetens, "Multimodality image registration by maximization of mutual information," *IEEE transactions on medical imaging,* vol. 16, no. 2, pp. 187-98, 1997.

[70] P. Viola and W. Wells, "Alignment by maximization of mutual information," *International Journal of Computer Vision,* vol. 24, no. 2, pp. 137-154, 1997.

[71] C. Studholme, D. Hill and D. Hawkes, "An overlap invariant entropy measure of 3D medical image alignment," *Pattern Recognition,* vol. 32, pp. 71-86, 1999.

[72] M. J. D. Powell, "An efficient method for finding the minimum of a function of several variables without calculating derivatives," *Computer Journal,* vol. 7, pp. 152-162, 1964.

[73] W. H. Press, S. Teukolsky and W. F. B. Vetterling, Numerical recipes in C, Cambridge University Press, Second Edition, 1992.

[74] Y. Huang, C. Chen, C. Tsai, C. Shen and L. Chen, "Survey on block matching motion estimation algorithms and architectures with new results source," *Journal of VLSI Signal Processing Systems,* vol. 42, no. 3, pp. 297-320, 2006.

[75] G. Xiao, J. M. Brady, J. A. Noble, M. Burcher and R. English, "Nonrigid registration of 3-D free-hand ultrasound images of the breast," *IEEE Transactions On Medical Imaging,* vol. 21, no. 4, pp. 405-12, 2002.

[76] B. Likar and F. Pernus, "A hierarchical approach to elastic registration based on mutual information," *Image and Vision Computing,* vol. 19, no. 1-2, pp. 33-44, 2001.

[77] M. D. Buhmann, Radial basis functions: theory and implementations, Cambridge University Press, 2003.

[78] C. D. Lloyd, Local models for spatial analysis, Taylor & Francis, 2007.

[79] M. Jiang, G. Wang, M. W. Skinner, J. T. Rubinstein and M. W. Vannier, "Blind deblurring of spiral CT images," *IEEE Transactions On Medical Imaging,* vol. 22, no. 7, pp. 837-45, 2003.
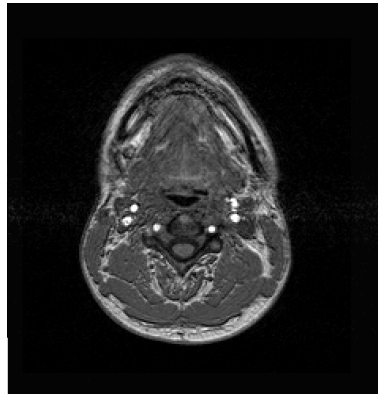
APPENDIX A

## MR INTENSITY NON-UNIFORMITY CORRECTION

The intensities in the MR images have spatial variations because of the magnetic field non-uniformity. The multiplicative bias value at each voxel must be known to correct the intensities. Shattuck *et al.* [16] modelled the multiplicative bias field as a slowly varying field. The MR volume was divided into small blocks and a parametric model on tissues was fitted to the histograms of the blocks. A regularization scheme was applied to the grid of bias values. The bias field for every voxel were then calculated with interpolation using B-splines.

A method called *non-parametric non-uniform intensity normalization* (N3) was proposed for intensity correction in MR images [57]. N3 was claimed to be independent of MR pulse sequence and irregularities in the image such as pathologies. The method was based on the correction of the frequencies of the image.

N3 is chosen as the intensity correction method in this study. The bias field is calculated and applied to the original MR image. The effect of the correction in two slices can be seen in Figure A.1. One of the slices is taken from the mandible level and the other from the region between the lips. The intensities at the lower part of this image, especially below the mouth level, had lower intensities than the upper part. The bias values in this region are found to be greater than 1.0, thus yielding brighter intensity values.
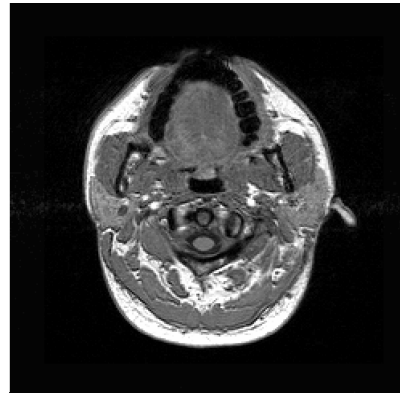
(a)   (b)
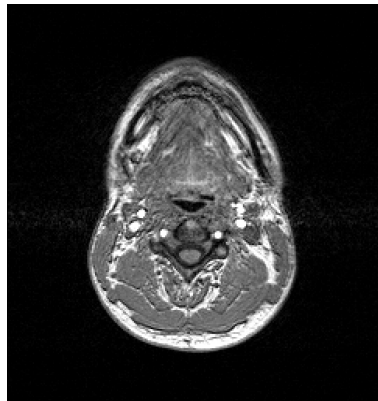
(c)   (d)

Figure A.1 Effect of N3 intensity correction. Slices from (a) mandible and (b) mouth levels before correction. Slices from (c) mandible and (d) mouth levels after correction.

APPENDIX B

MR-CT REGISTRATION

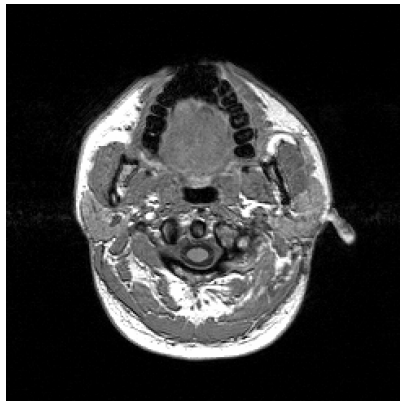The CT and MR images used in segmentation have different orientations and voxel sizes originally. The process known as registration is applied in order to bring the coordinate systems to an alignment. Image registration aims to find an optimal transformation between two images so that the points on the images are mapped onto each other spatially and with respect to intensity. Two 3D images I_1 and I_2 can be mapped to each other by

$$I_2(x, y, z) = g(I_1(f(x, y, z))), \qquad \text{B1}$$

where $I_1(x, y, z)$ and $I_2(x, y, z)$ are the functions which map their coordinates to intensity values, f is the coordinate transformation and g is the intensity transformation [66]. In the case of medical images, the mapped points correspond to the same anatomical point on each image [67]. Image registration can be classified according to nature of registration basis, nature of transformation, modalities involved and subject [68]. A classification based on the transformation type divides image registration methods into two categories which are called rigid and non-rigid methods. In rigid registration only transformations during which parallel lines stay parallel are used. Non-rigid registration methods involve other types of transformation that are more complex. In this study, registration schemes composed of rigid and non-rigid components are used.

Rigid Registration

The original CT and MR images are misaligned because of the unknown position relationship between the images and tilting of the head around every axis. Since the images are taken from the same patient, structural similarity is very high except for possible soft tissue deformations. The situation can be different if the images are taken pre-operation and post-operation. However, it is aimed to use both CT and MR images for segmentation and it is reasonable to assume that the images are obtained at similar dates with no major structural differences. CT image will be registered to MR image. Therefore, it will have its resolution lowered. This helps in simulating the fusion of low-resolution CT and MR images.

For registration purposes, the transformed image will be called as floating image and the other image as target image. In rigid registration the floating image is free to be translated along and rotated around the 3 axes. Scaling is not needed because the voxel sizes in millimetres are known. The transformation which will align two images are demonstrated in Figure B.1. This becomes an optimization problem with parameters as translations t_x, t_y and t_z ; and rotations r_x, r_y and r_z which can be combined into transformation T. The optimization of this transformation is to be made according to a cost function which maximizes the similarity between the images.
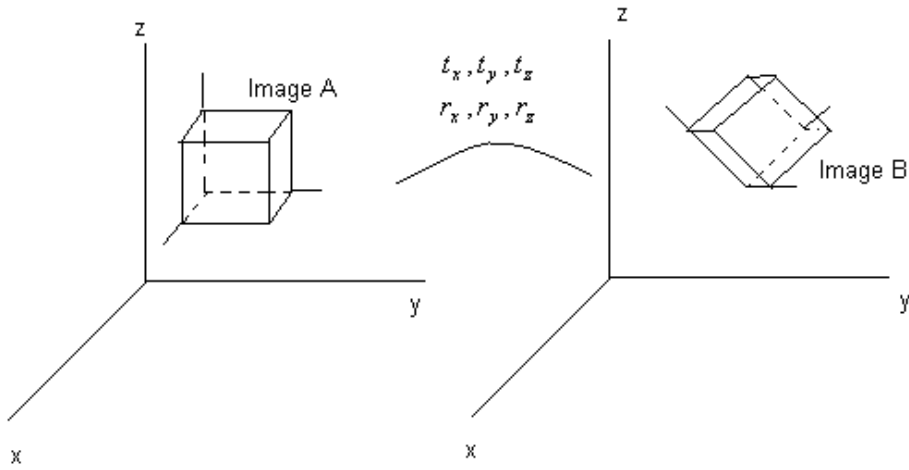
Figure B.1 The transformation between the registered images

Similarity Metric

The similarity metric for the rigid registration process is based on maximization of mutual information [69]. This is a statistical method which calculates the joint probability distributions of two images and measures the degree of dependence between them. One of the most important properties of the mutual information metric is that it is robust to changes in intensity values. Therefore, it is a powerful tool for multi-modality registration. Mutual information (MI) originates from information theory [70]. MI of two variables I_1 and I_2 is defined as the amount of information contained in the variables about each other;

$$MI(I_1, I_2) = H(I_1) + H(I_2) - H(I_1, I_2), \qquad \text{B2}$$

where $H(I_1)$ and $H(I_2)$ denote the marginal entropies and $H(I_1, I_2)$ denotes the joint entropy of I_1 and I_2. In image registration presented here, with reference to Figure B.1, suppose that I_1 is the MR image and TI_2 is the transformation applied CT image. The mutual information metric is enhanced further into

normalized mutual information. Normalized mutual information (NMI) $G(I_1, TI_2)$ is a modified version of MI. NMI is overlap invariant [71] meaning that the measure is robust when the overlap between I_1 and transformed image TI_2 is not optimum. NMI is calculated by

$$G(I_1, TI_2) = \frac{H(I_1) + H(TI_2)}{H(I_1, TI_2)}. \qquad \text{B3}$$

The entropies $H(I_1)$, $H(TI_2)$ and $H(I_1, TI_2)$ are calculated by

$$H(\bullet) = -\sum p(\bullet) \log p(\bullet), \qquad \text{B4}$$

where $p(I_1)$ and $p(TI_2)$ are marginal intensity probabilities and $p(I_1, TI_2)$ is the joint intensity probability. Joint intensity probability is obtained from the joint histogram which is a multi-dimensional histogram of the images.

In order to obtain the joint histogram, the intensities on the images are divided into intervals which are called bins. For example, if the [0, 1] intensity range is divided into 10 bins, intensity range [0, 0.1] is assigned to bin 1, the range [0.1, 0.2] is assigned to bin 2, etc. If the bin counts for the two images are M and N, the joint histogram is initialized as a 2-dimensional array with zero values. Then the intensity bins of two voxels at the same coordinate in each image are paired. The corresponding counter in the joint histogram is increased by one for every intensity pair. This process is repeated for every voxel to yield the counters and the histogram is normalized into a probability distribution.

Search Method

The maximization of the mutual information of the two images is achieved by using Powell's conjugate gradient descent method [72]. A one-dimensional search is performed at every iteration of the algorithm. The direction of the search at each step is found by combining the directions used in the previous steps. The new

direction is added to the list of search vectors while a previous one is deleted. One of the most important properties of this method is its ability to work without the function gradient [73] and it is not possible to supply an algorithm with the gradient of a function in most practical cases such as the MI function used here.

Partial Volume Interpolation (PVI)

The optimization procedure tests the cost function with floating point numbers. Translating an image with floating point numbers require interpolation of pixel values. When calculating mutual information the pixel values of the floating and target images are required. Using interpolated pixel values from the floating image and exact values from the target image may result in several local minima. Partial Volume Interpolation (PVI) is implemented to overcome this difficulty [69]. Each pixel in the target image is perturbed by a random amount less than the size of a pixel. This provides the smoothness of the cost function. Note that the Partial Volume term in this context is slightly different from the PV effect described in section 2.3.3. In the context of interpolation PV implies the usage of non-integer pixel coordinates in registration.

Non-rigid Registration

Although the images are aligned to an extent after rigid registration (Figure B.2), there are parts of the image with inaccurate registrations (Figure B.3). In order to correct the inaccuracies in the registration, a non-rigid scheme is to be employed. One of the most studied image registration algorithms is block matching.
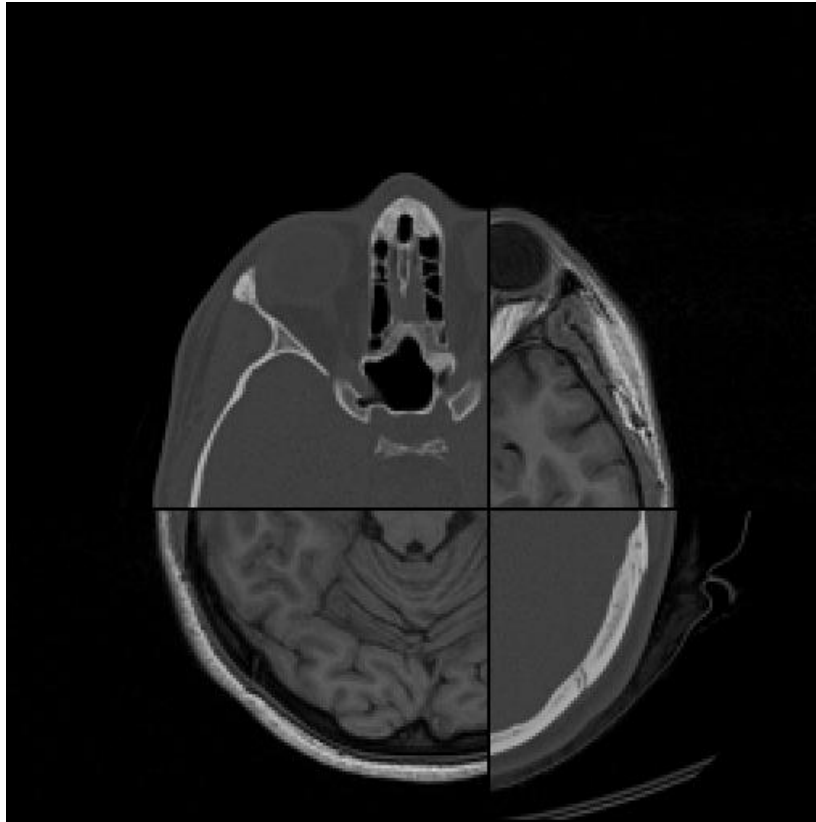
Figure B.2 MR (top right, bottom left) and CT (top left, bottom right) images are shown together after registration
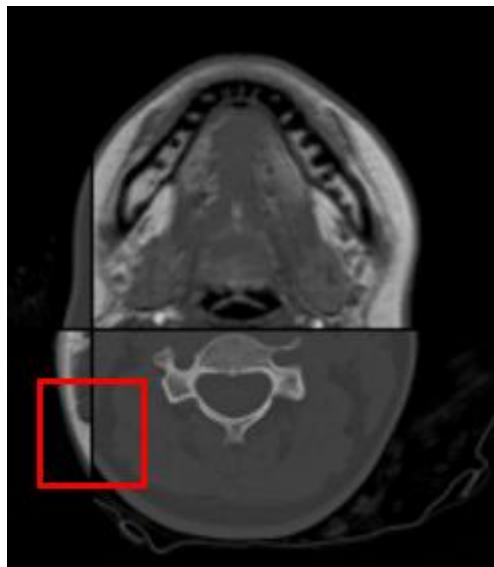

Figure B.3 A slice with where mis-registration is more apparent

Block Matching

Block matching is the process in which part of one image is searched on the image to determine the motion in the image. Although block matching is generally used in video compression using motion estimation [74], medical applications of the method exist such as ultrasound image registration [75]. In block matching, a sample block centered on a coordinate in the floating image is searched in the search window centered on the same coordinate of the target image based on a similarity metric. For an image of $256 \times 256 \times 256$ voxels, the sample block size was chosen as $24 \times 24 \times 24$ empirically. Although a smaller blocks size would provide a finer deformation field, it was concluded through testing that the field became highly irregular due to the small amount of information contained in the block.
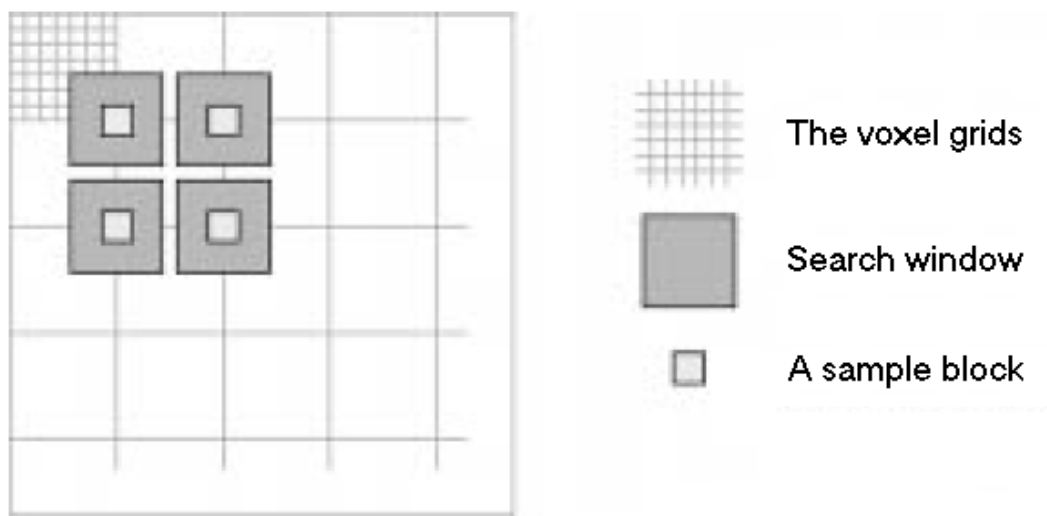

Figure B.4 A sample block and search window (original image from [75])

Deformations found by rigid registration account for the larger part of the deformations obtained during the whole registration process. Therefore, the deformations to be calculated by the non-rigid part are minimal in comparison.

This leads to the possible use of smaller search windows. The size of the search windows was chosen as 8x8x8 allowing 8 voxel translations along each axis. The translations are performed with integer intervals. The similarity metric chosen for block matching is correlation coefficient (CC). Among other alternatives there are Sum of Squared Differences (SSD), Sum of Absolute Differences (SAD), Correlation Ratio (CR) and Mutual Information (MI). SSD and SAD are not suitable as multimodal metrics. Although MI is a multimodal metric, the blocks do not contain a lot of information. Therefore, MI is also ruled out because MI works best with more information.

Likar *et al.* [76] proposed a hierarchical approach to elastic registration in which images are divided in exponentially increasing number of subparts and registered locally. Block matching in [76] is carried out with a 3 level hierarchical scheme. This reduces the number of computations substantially. However, the scheme used in here is a slightly altered version of the mentioned method. As a difference, same number of blocks is used in each level. The blocks are sampled down and then matched while allowing coarse translations. For the next hierarchy level, block sizes are doubled and translations are limited to half of the previous level.

When this process is performed for all blocks, a deformation field $s(x, y)$ at every block center (x, y) is obtained as

$$s(x, y) = \begin{bmatrix} x_d \\ y_d \end{bmatrix}. \qquad \qquad \text{B5}$$

Transformations in images are performed in a reverse fashion. Applying transformation directly to an image results in empty voxels in the transformed image. Therefore, the transformation is reversed and applied to an empty image to find the source pixel coordinate in the original image. This process ensures that all the pixels in the transformed image contain information. Considering this, the reversed deformation field is written as

$$RF(x + x_d, y + y_d) = \begin{bmatrix} -x_d \\ -y_d \end{bmatrix}, \qquad \text{B6}$$

such that it shows the deformations from the target image to the floating image. The data points of this field are at irregular points so a linear interpolation method cannot be used to obtain the deformations at every voxel of the target image. Radial Basis Functions (RBF) are widely used to approximate functions with scattered data [77].

Radial Basis Functions

A function can be approximated as the sum of N radial basis functions

$$f(x) = \sum_{i=1}^{N} \sum_{i=1}^{N} \omega_i \phi\left(\|x - x_i\|\right), \qquad \text{B7}$$

where x_i are the N centers and ω_i are the corresponding weights. The centers are the points where the function values are known. For the basis function ϕ there are several choices, such as linear ($\phi(r) = r$), cubic spline ($\phi(r) = r^3$), Gaussian ($\phi(r) = e^{-r^2/c}$), thin-plate spline ($\phi(r) = r^2 \log r$) etc. where $r = \|x - x_i\|$. Thin-plate splines are widely used because of the smoothness they provide [78]. Therefore, they are selected to be used for the interpolation.

The function values at the centers are known as

$$f_j = f(x_j) \qquad j = 1, .., N, \qquad \text{B8}$$

which can also be approximated by the sum of radial function as

$$f_j = f(x_j) = \sum_{i=1}^{N} \omega_i \phi\left(\|x_j - x_i\|\right). \qquad \text{B9}$$

The weights ω_i are the only unknowns in this set of linear equations. The equations can be rewritten in matrix form as

$$\begin{bmatrix} \phi_{11} & .. & \phi_{1N} \\ \vdots & .. & \vdots \\ \phi_{N1} & .. & \phi_{NN} \end{bmatrix} \begin{bmatrix} \omega_1 \\ \vdots \\ \omega_N \end{bmatrix} = \begin{bmatrix} f_1 \\ \vdots \\ f_N \end{bmatrix}, \qquad \text{B10}$$

where the basis function values are $\phi_{ij} = \phi\big(\|x_j - x_i\|\big)$. This linear system can be solved to obtain the unknown basis function weights.

The function values in these formulations are scalar, whereas the block translations are 3D vectors. In order to utilize the RBFs 3 systems are obtained each representing a component of the vector. The three approximating functions become

$$d_x(P) = \sum_{i=1}^{N} \omega x_i \, \phi\big(\|P - P_i\|\big),$$

$$d_y(P) = \sum_{i=1}^{N} \omega y_i \, \phi\big(\|P - P_i\|\big), \qquad \text{B11}$$

$$d_z(P) = \sum_{i=1}^{N} \omega z_i \, \phi\big(\|P - P_i\|\big),$$

where P is the coordinate of a voxel in the image and P_i are the reversed block centers.

Non-Rigid Registration Result

The block matching registration is applied to the images so that the transformations are added to the transformations found in rigid registration. To represent the final transformations a slice is shown before (Figure B.4) and after (Figure B.6) non-rigid registration. Although the deformations during non-rigid transformations throughout the volume do not exceed a couple of pixels, they are

very important for the accuracy of the fusion segmentation method. Therefore, application of block matching is critical for the success of the overall algorithm.
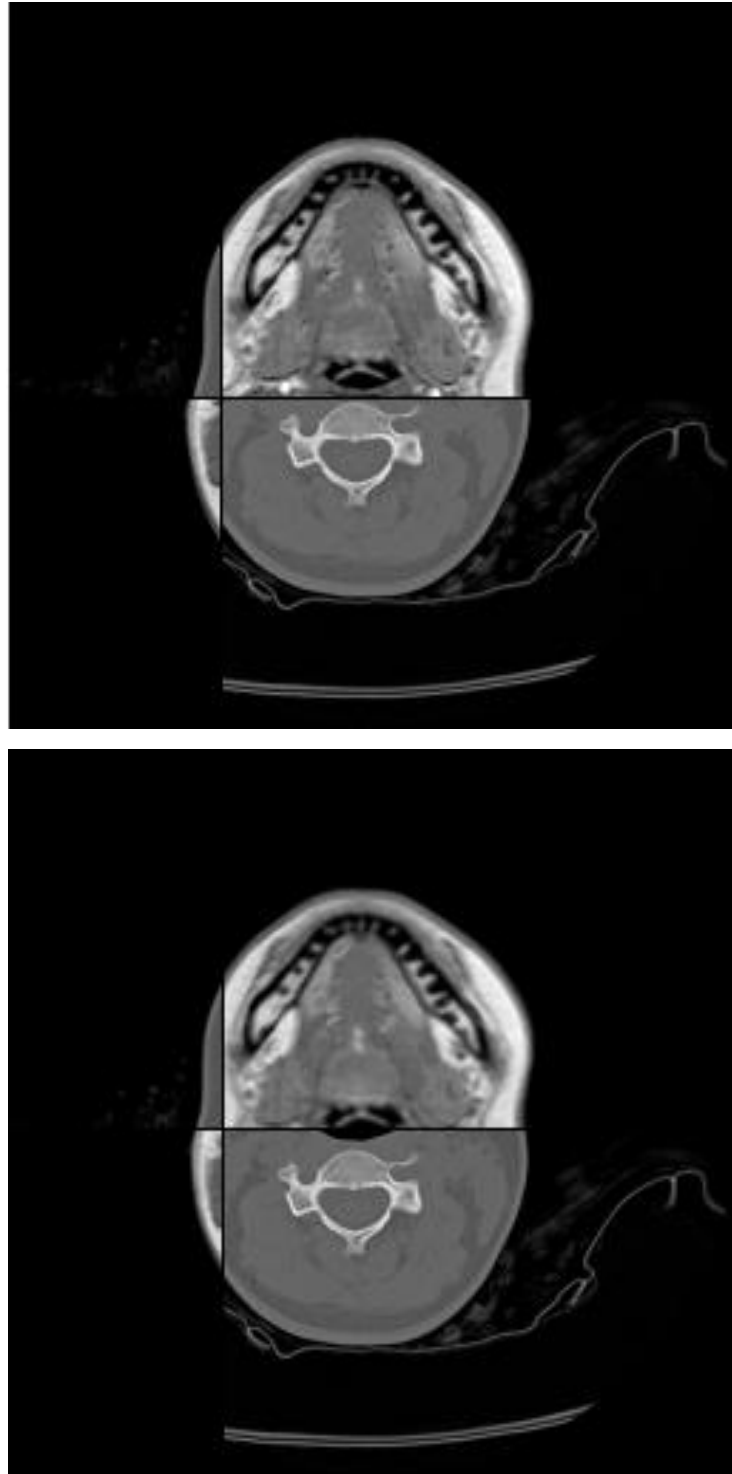

Figure B.5 MR slice after (a) rigid registration and (b) non-rigid registration

CURRICULUM VITAE

PERSONAL INFORMATION

Surname, Name: Şener, Emre

Nationality: Turkish (TC)

Date and Place of Birth: 16 September 1976 , İzmir

Marital Status: Single

Phone: +90 536 274 15 41

email: sener_emre@metu.edu.tr

EDUCATION

Degree	Institution	Year of Graduation
MS	METU Engineering Sciences	2003
BS	METU Mechanical Engineering	2000
High School	Bornova Anatolian High School, İzmir	1994

WORK EXPERIENCE

Year	Place
2012-	Ekstrem Bir Computer
2011-2012	Turkish Army
2009-2010	B-Proje Engineering
2007-2008	TUBITAK Project in Electrical Engineer Dpt. METU
2004-2006	B-Proje Engineering
2002-2003	Makrosam

FOREIGN LANGUAGES

Fluent English

PUBLICATIONS

Journals

Mumcuoğlu E.Ü., Bozkurt F.M., Aslan M., Şener E., Uğur Ö., Computerized Scar Detection on Renal Cortical Scintigraphy Images, Nuclear Medicine Communications, 2011

Eraslan A.N., Şener E., Argeşo H., Stress Distributions in Energy Generating Two-layer Tubes Subjected to Free and Radially Constrained Boundary Conditions, International Journal of Mechanical Sciences, 2003

National Conferences

Şener E., Mumcuoğlu E.Ü., Kale E., Akgün V., Automatic and Semi-Automatic Segmentation of Facial Mimic Muscles Using CT and MR Images, 5th TurkMIA Congress on Medical Informatics, Turkey, 2008

Şener E., Eraslan A.N., Analytical Thermoplastic Solution of a Linearly Hardening Cylinder with Fixed Ends, 13th TUMTMK, National Mechanical Sciences Congress, Turkey, 2003