UNIVERSITÉ PARIS DESCARTES | USPC Université Sorbonne Paris Cité

Thèse en convention CIFRE ANRT avec la société INTRASENSE

présentée par

Fuchen LIU

pour obtenir le grade de

Docteur en Mathématiques Appliquées - Statistiques

Hierarchical clustering using equivalence test: Application on automatic segmentation of Dynamic Contrast Enhanced image sequence

---

réalisée sous la direction de

Yves Rozenholc et **Charles-André Cuénod**

préparée au laboratoire *UMR CNRS 8145 - MAP5* de l'Université PARIS DESCARTES

au sein de l'École Doctorale ED 474 - Frontière du Vivant

Soutenue publiquement le 11 Juillet 2017 devant le jury composé de :

Bouveyron Charles	Examinateur	Université Paris Descartes, MAP5
Bridal S. Lori	Examinateur	UPMC / CNRS
Chemouny Stéphane	Membre invité	INTRASENSE
Cuénod Charles-André	Directeur de Thèse	Université Paris Descartes / HEGP
Felblinger Jacques	Rapporteur	Université de Lorraine / CHU de Nancy-Brabois
Friboulet Denis	Rapporteur	INSA de Lyon
Guillevin Rémy	Examinateur	Université de Poitiers
Robin Stéphane	Rapporteur	AgroParisTech / INRA
Rozenholc Yves	Directeur de Thèse	Université Paris Descartes

Résumé

L'imagerie de perfusion permet un accès non invasif à la micro-vascularisation tissulaire. Elle apparaît comme un outil prometteur pour la construction de biomarqueurs d'imagerie pour le diagnostic, le pronostic ou le suivi de traitement anti-angiogénique du cancer. Cependant, l'analyse quantitative des séries dynamiques de perfusion souffre d'un faible rapport signal sur bruit (SNR). Le SNR peut être amélioré en faisant la moyenne de l'information fonctionnelle dans de grandes régions d'intérêt, qui doivent néanmoins être fonctionnellement homogènes.

Pour ce faire, nous proposons une nouvelle méthode pour la segmentation automatique des séries dynamiques de perfusion en régions fonctionnellement homogènes, appelée DCE-HiSET. Au coeur de cette méthode, HiSET (*Hierarchical Segmentation using Equivalence Test* ou *Segmentation hiérarchique par test d'équivalence*) propose de segmenter des caractéristiques fonctionnelles ou signaux (indexées par le temps par exemple) observées discrètement et de façon bruité sur un espace métrique fini, considéré comme un paysage, avec un bruit sur les observations indépendant Gaussien de variance connue. HiSET est un algorithme de clustering hiérarchique qui utilise la p-valeur d'un test d'équivalence multiple comme mesure de dissimilarité et se compose de deux étapes. La première exploite la structure de voisinage spatial pour préserver les propriétés locales de l'espace métrique, et la seconde récupère les structures homogènes spatialement déconnectées à une échelle globale plus grande. Etant donné un écart d'homogénéité δ attendu pour le test d'équivalence multiple, les deux étapes s'arrêtent automatiquement par un contrôle de l'erreur de type I, fournissant un choix adaptatif du nombre de régions. Le paramètre δ apparaît alors comme paramètre de réglage contrôlant la taille et la complexité de la segmentation. Théoriquement, nous prouvons que, si le paysage est fonctionnellement constant par morceaux avec des caractéristiques fonctionnelles bien séparées entre les morceaux, HiSET est capable de retrouver la partition exacte avec grande probabilité quand le nombre de temps d'observation est assez grand.

Pour les séries dynamiques de perfusion, les hypothèses, dont dépend HiSET,

sont obtenues à l'aide d'une modélisation des intensités (signaux) et une stabilisation de la variance qui dépend d'un paramètre supplémentaire a et est justifiée a posteriori. Ainsi, DCE-HiSET est la combinaison d'une modélisation adaptée des séries dynamiques de perfusion avec l'algorithme HiSET. A l'aide de séries dynamiques de perfusion synthétiques en deux dimensions, nous avons montré que DCE-HiSET se révèle plus performant que de nombreuses méthodes de pointe de clustering.

En terme d'application clinique de DCE-HiSET, nous avons proposé une stratégie pour affiner une région d'intérêt grossièrement délimitée par un clinicien sur une série dynamique de perfusion, afin d'améliorer la précision de la frontière des régions d'intérêt et la robustesse de l'analyse basée sur ces régions tout en diminuant le temps de délimitation. La stratégie de raffinement automatique proposée est basée sur une segmentation par DCE-HiSET suivie d'une série d'opérations de type érosion et dilatation. Sa robustesse et son efficacité sont vérifiées grâce à la comparaison des résultats de classification, réalisée sur la base des séries dynamiques associées, de 99 tumeurs ovariennes et avec les résultats de l'anapathologie sur biopsie utilisés comme référence. Finalement, dans le contexte des séries d'images 3D, nous avons étudié deux stratégies, utilisant des structures de voisinage des coupes transversales différentes, basée sur DCE-HiSET pour obtenir la segmentation de séries dynamiques de perfusion en trois dimensions.

Cette thèse de doctorat a été supportée par un contrat CIFRE de l'ANRT (Association Nationale de la Recherche et de la Technologie) avec la société INTRASENSE qui conçoit, développe et commercialise des solutions de visualisation et d'analyse d'imagerie médicale, incluant Myrian®. L'outil DCE-HiSET a été porté sur Myrian® et testé complètement fonctionnel.

Abstract

Dynamical contrast enhanced (DCE) imaging allows non invasive access to tissue micro-vascularization. It appears as a promising tool to build imaging biomarker for diagnostic, prognosis or anti-angiogenesis treatment monitoring of cancer. However, quantitative analysis of DCE image sequences suffers from low *signal to noise ratio* (SNR). SNR may be improved by averaging functional information in large regions of interest, which however need to be functionally homogeneous.

To achieve SNR improvement, we propose a novel method for automatic segmentation of DCE image sequence into functionally homogeneous regions, called *DCE-HiSET*. As the core of the proposed method, HiSET (*Hierarchical Segmentation using Equivalence Test*) aims to cluster functional (e.g. with respect to time) features or signals discretely observed with noise on a finite metric space considered to be a landscape. HiSET assumes independent Gaussian noise with known constant level on the observations. It uses the p-value of a multiple equivalence test as dissimilarity measure and consists of two steps. The first exploits the spatial neighborhood structure to preserve the local property of the metric space, and the second recovers (spatially) disconnected homogeneous structures at a larger (global) scale. Given an expected homogeneity discrepancy δ for the multiple equivalence test, both steps stop automatically through a control of the type I error, providing an adaptive choice of the number of clusters. Parameter δ appears as the tuning parameter controlling the size and the complexity of the segmentation. Assuming that the landscape is functionally piecewise constant with well separated functional features, we prove that HiSET will retrieve the exact partition with high probability when the number of observation times is large enough.

In the application for DCE image sequence, the assumption is achieved by the modeling of the observed intensity in the sequence through a proper variance stabilization, which depends only on one additional parameter a. Therefore, DCE-HiSET is the combination of this DCE imaging modeling step with our statistical core, HiSET. Through a comparison on synthetic 2D DCE image sequence, DCE-HiSET has been proven to outperform other state-of-the-art clustering-based

methods.

As a clinical application of DCE-HiSET, we proposed a strategy to refine a roughly manually delineated ROI on DCE image sequence, in order to improve the precision at the border of ROIs and the robustness of DCE analysis based on ROIs, while decreasing the delineation time. The automatic refinement strategy is based on the segmentation through DCE-HiSET and a series of erosion-dilation operations. The robustness and efficiency of the proposed strategy are verified by the comparaison of the classification of 99 ovarian tumors based on their associated DCE-MR image sequences with the results of biopsy anapathology used as benchmark.

Furthermore, DCE-HiSET is also adapted to the segmentation of 3D DCE image sequence through two different strategies with distinct considerations regarding the neighborhood structure cross slices.

This PhD thesis has been supported by contract CIFRE of the ANRT (Association Nationale de la Recherche et de la Technologie) with a french company INTRASENSE, which designs, develops and markets medical imaging visualization and analysis solutions including Myrian®. DCE-HiSET has been integrated into Myrian® and tested to be fully functional.

Contents

List of Figures

List of Tables

Remerciements

Chapter 1

Introduction

We start this introduction with a detailed context of cancer, perfusion imaging and their relationship, in order to help readers to understand the observations and the statistical challenge which was at the origin of this work. Then we introduce the statistical core of this thesis, which aims to address this challenge in a general framework and the several applications we have tackled using our development.

1.1 Medical context

1.1.1 Cancer

Being responsible to 8.2 million deaths and 14.1 million new cases all over the world in 2012 [Ferlay et al., 2013], cancer has been one of the most mortal and prevalent diseases in the world during decades. Several different factors such as the development of new therapeutic methodology and the improvement of detection and monitoring techniques for therapeutic follow-up provide strong hopes for cancer treatment. On one hand, more targeted and personalized therapeutic care with respect to patient's states can significantly reduce the mortality. On the other hand, more advanced and comprehensive detection and monitoring tools might help provide earlier and more accurate diagnosis and detailed feedback about treatment used.

In this context, as an indispensable tool for radiologists to detect tumors and to monitor treatments, recent medical imaging tools, such as DCE-imaging, should pursue more standardized operation protocol as well as more robust analysis techniques, in order to more precisely and efficiently characterize the effect of therapies or drugs on tumors in clinical studies. With currently widely used imaging modalities, the development of medical imaging analysis techniques capable of exhibiting subtle functional information in an accurate and robust way plays one of the principle roles in the fight against cancer.

1.1.1.1 Treatments

In clinical practice, cancer can be treated by surgery, radiotherapy, chemotherapy, hormonal therapy, and targeted therapy. More recently, studies have shown that angiogenesis inhibitors controlling for example the VEGF has potential in treatment to many types of cancer, at least associated with other drugs.

Surgery

Surgery is an effective way to cure cancer showing localized tumors. However, once the cancer has metastasized to other sites of the body, a complete surgical removal is usually impossible since single metastatic tumorous cell can potentially grow into a new tumor. Unfortunately, localized tumors may be invasive cancer.

Radiotherapy

Radiation therapy consists in irradiating tumors with ionizing radiation to kill tumorous cells or at least stop their division by damaging their genetic material, and shrink tumors by stopping their growth. As surgery, radiation therapy works efficiently to localized tumors. Although in practice radiation therapy is provided in repeated small doses, such that most normal cells can recover from the effects of radiation, the damage to nearby healthy tissue can not be ignored. Moreover, radiotherapy is less effective when tumors are hypoxic because oxygen is an efficient radio-sensitizer.

Chemotherapy

Chemotherapy consists in destroying tumorous cells with cytotoxic drugs and is typically applied along with radiotherapy or surgery when tumor metastasis has occurred. However, due to the complex characteristics of the tumorous vascular network (Figure 1.1), efficient targeting of the tumor with cytotoxic drugs could be a very difficult task. Most forms of chemotherapy target rapidly dividing cells and are not specific to tumorous cells. Therefore, chemotherapy might harm healthy tissue, especially tissues that have a high replacement rate. Moreover, during chemotherapy, patients can develop resistance to the cytotoxic drugs, which reduces the therapeutic effects.

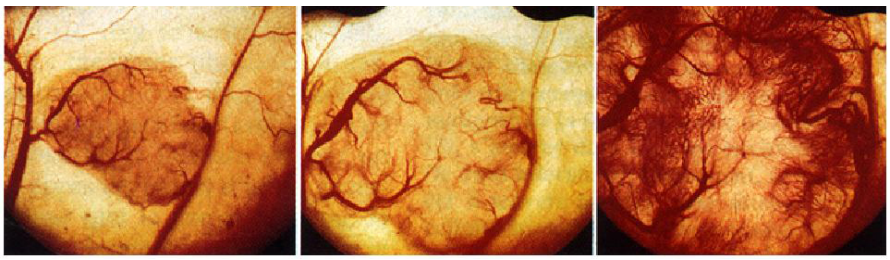


Figure 1.1: Three tumors with their vascular structures: the left one illustrates a solid tumor with relatively simple vascular structure; the middle one illustrates a tumor growing from the left one due to tumorous angiogenesis, more capillaries and small vessels are formed during its growth; the right one illustrates an invasive tumor with an completely disorganized micro-vascular structure.

Hormonal therapy

Hormonal therapy consists in inhibiting the growth of certain types of tumorous cell by manipulating levels and activities of certain hormones. During the treatment, patients may suffer several side effects such as headache, nausea and tiredness, and increased risk of endometrial cancer, blood clots and stroke.

Targeted therapy

Targeted therapy consists in blocking the cell pathways (proteins or other molecules)

that can lead to cancers using agents targeting these pathways. Targeted therapy has to be adapted to issues such as patient's state and type of tumors to be effective. However, researches show that cancer cells may become resistant to the drug.

Anti-angiogenic therapy

Angiogenesis is the process through which new blood vessels are generated from pre-existing ones [Birbrair et al., 2015]. In order to transform from a benign state to a malignant one, tumors need a dedicated blood supply to provide the oxygen and other essential nutrients they require. Therefore, tumors undergo an 'angiogenic switch' that breaks the balance between pro and anti-angiogenic factors in tumorous angiogenesis. Tumor vascular network is extended in a disorganized and inundant way.

From this perspective, anti-angiogenic therapy has been introduced in recent decades to regulate tumorous angiogenesis. It reduces the production of pro-angiogenic factors, prevent them binding to their receptors and block their actions. New therapeutic strategies have entered trials to combine anti-angiogenic inhibitor with chemotherapy to enhance delivery of cytotoxic molecules [Jain, 2005]. However, since many factors are involved in angiogenic phenomena, patients can develop resistance to anti-angiogenic therapy [Eikesdal and Kalluri, 2009]. With respect to these issues, indicators of tumorous vascular structure can play a significant role to personalized anti-angiogenic treatment to monitor their effects and to detect the development of drug resistance.

1.1.1.2 Detection and Monitoring

The choice of therapy depends upon the location, the grade of the tumor, the stage of the disease, as well as the general state of the patient. Meanwhile, each option has its specific targets, limitations and undesirable side effects. Therefore, for each individual, in order to avoid unnecessary burden brought by inefficient or even inappropriate therapy, critical decisions must be made based on both accurate

information obtained at diagnosis and later during the monitoring of the treatment leading to a personalized treatment. In addition, pre-clinical studies (REMISCAN) which aim at improving the efficiency of a new drug also require accurate tumor characterization and monitoring techniques. These techniques are expected to be able to assess morphological, molecular or functional indicators, which are specific and predictive, of tumor state.

1.1.1.3 Indicators in cancer therapy

In order to help therapists and doctors in the personalization of the treatment, several tools are available.

Morphological

In clinical practice, morphological features of tumor, such as tumor size, have been considered to be indicators of the tumor's therapeutic sensitivity [Stanley et al., 1977] and of the metastasis probability [Koscielny et al., 1984]. The RECIST criteria [Eisenhauer et al., 2009] is currently a clinical standard for the evaluation of cancer therapy, which classify tumors into three categories: stable, in regression or in progression. However, by definition, these criteria do not provide functional information on the tumor. For instance, according to this criteria, a tumor with a large necrotic core would be identified to be of equal risk as a highly vascularized tumor of the same size.

Molecular

Since certain specific molecules are produced along with the growth and functional modifications of tumors, monitoring the evolution of the concentration of those molecules can provide valuable information about the type and the stage of cancer [Ludwig and Weinstein, 2005]. Therefore, new technologies that can simultaneously assess the concentration of thousands proteins has been developed and achieved a major improvement in diagnosis and tumor classification [Kulasingam and Diamandis, 2008].

Functional

As mentioned above, the assessment of tumor angiogenesis is able to reveal massive informations about functional modification of tumors. Quantification of the blood circulation inside tumorous vascular network became a proper indicator for cancer therapy. However, due to the complexity of tumorous vascular network during tumor growth (Figure 1.1), the blood circulation in the capillaries and small peripheral vessels, which is defined as the micro-circulation, appear as the target to evaluate.

A common way to achieve this purpose is to visualize micro-vascular perfusion in order to characterize micro-vascular properties. The perfusion can be tracked by an imaging modality that measures the signal changes in the tissue of interest. The patterns of these changes are connected to the underlying perfusion. According to pathologists, tumorous tissue differs from normal tissue and usually shows heterogeneity, leading to similar variations in its micro-vascular properties. However, due to the growth of the tumor, one can expect the tumor to exhibit homogeneous sub-regions. Therefore, finding the sub-regions in tumorous tissue with similar perfusion patterns provides valuable information about tumor structure and its angiogenesis.

1.1.2 Perfusion Imaging

1.1.2.1 Technique

Among the new techniques, dynamic contrast enhanced (DCE) imaging using computed tomography (DCE-CT), magnetic resonance imaging (DCE-MRI) or ultrasonic imaging (DCE-US) is very promising, because it can non-invasively monitor the local changes in microcirculation secondary to the development of new vessels (neo-angiogenesis) induced by the growth of cancers. DCE imaging follows and analyzes the local distribution of a contrast agent in the blood circulation system after intravenous injection, using a sequential acquisition [Ingrisch and

Sourbron, 2013]. The imaging modalities (CT, MRI and US) allow to have a 2D or 3D vision of a volume of the body during the acquisition. The DCE images taken at the same cross-section of patient during the entire acquisition period compose a *DCE image sequence*. Since each image in DCE image sequence has a thickness, every pixel in image has a volume. Therefore, DCE image consists of *voxels* instead of pixels as in general image.

1.1.2.2 Modalities

Computed Tomography

The CT scanner is a device that measures the absorption by the tissues of X-rays emitted by a ray tube that rotates around the patient providing a Radon transform of the image [Herman, 2009]. After inversion of the Radon transform that is specific to each brand of device, it is possible to have a 2D or 3D vision reconstruction of a patient. The scanner allows to observe all areas of the body and all types of tissues (organs, bone, blood vessels, etc.).

Magnetic Resonance Imaging

With respect to CT, MRI is a less invasive tool based on the principle of nuclear magnetic resonance and uses some properties of atomic nuclei when exposed to magnetic fields. This tool is mainly dedicated to the imaging of the central nervous system, muscles, heart and tumors, it can observe the "soft" tissues with higher contrast than the scanner. Nevertheless, the CT scanner is not yet possible to be completely replaced by the MRI since some tissues such as bones are not observable with MRI, and moreover MRI does not allow precise localization at high time frequency, which indeed is needed for following the injection of a contrast agent bolus.

Ultrasonic imaging

Ultrasonic imaging uses high frequency sound waves that are reflected by tissue to produce images. Ultrasonic imaging is a non-invasive modality, which can be

brought directly and quickly to the patient in real time with a relatively low cost. However, it may provide less anatomical detail than CT and MRI. Moreover, it performs very poorly when there is a gas between the transducer and the organ of interest [Iftimia et al., 2011], due to the extreme differences in acoustic impedance.

1.1.2.3 Protocol

A contrast agent intravenously injected following the blood circulation is sketched in Figure 1.2, which illustrates the sequential image acquisition. The imaging acquisition starts before the injection in order to measure the pre-injection baseline consisting of non-uniform gray levels across the image without enhancement induced by contrast agent. The contrast agent is then sequentially followed during and after injection to monitor the kinetic of the vessel and tissue enhancements over time. The contrast agent arrives with the oxygenated blood through the aorta and main arteries. Its concentration within unit volume voxels is measured for the first time when it passes through the cross-section of the image. This measured concentration within a unit volume voxel inside the aorta is called the arterial input function (AIF). Afterwards, the contrast agent is distributed within the arterial system to enter the microvascular network made of capillaries within the tissues. The exchange within the tissue of oxygen, nutriments (such as glucose), and wastes (such as CO2) as well as contrast agent occurs at the microvascular level, and the concentration of contrast agent during this exchange is measured in all tissue voxels inside the image cross-section. Later, the contrast agent returns to the venous system with the deoxygenated blood. Each type of normal and pathological tissue has its own and specific microvascular architecture and physiology, leading to specific functional behavior, and therefore different temporal enhancement behavior after contrast injection.

This course of contrast agent is monitored by CT, MRI or US, which allows observations and therefore analysis at the voxel level at the same voxel grid for every image in the sequence. At each voxel of the image sequence, it provides a *time intensity curve* (TIC) that is the sum of the “baseline” grey level (before the

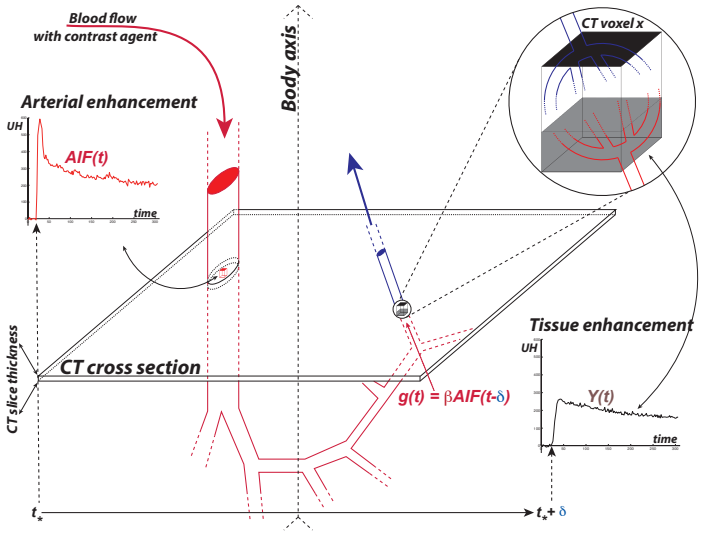


Figure 1.2: DCE-imaging and contrast agent circulation. The patient's body is materialized by the mixed arrow.

injection) and the varying effect of the contrast agent related to its concentration (after the injection). Removing the pre-injection phase and subtracting the baseline from the time-intensity curve, we can construct a *time enhancement curve* (TEC) only for the enhancement induced by contrast agent during the the post-injection phase. In Figure 1.2, the red curve represents the TEC for a voxel in the aorta, and the black curve represents TEC for a voxel in the tissue of interest such as tumor. Two types of curves reveal different characteristics of tissues and both are of interest of further analysis. In the sequel, without more precision, *time curve* (TC) will refer either to TIC or TEC, depending on the study purpose.

1.1.2.4 Perfusion analysis

The TCs from the DCE image sequences can be analyzed using qualitative or quantitative approaches. In qualitative analysis, descriptive features such as maximum enhancement (ME), area under the curve (AUC) and time to peak (TTP) are directly derived from the observed TCs of voxels. These measures are used either individually [Tuncbilek et al., 2005; Medved et al., 2004; Tuncbilek et al., 2004] or as a combination [Lavini et al., 2006]. However, they are not

normalized and are sensitive to the variations of patient's physiology (stress, blood pressure, heartbeat, etc.) and the acquisition protocols such as the number of images in the sequence, the amount of injected contrast agent, and the total acquisition duration. In quantitative analysis, pharmacokinetic models such as standard and extended Tofts [Tofts et al., 1995; Tofts, 1997] and Brix models [Brix and et al., 2004] and blood flow models [Axel, 1980; Fieselmann et al., 2011; Ostergaard et al., 1996] are commonly used in order to remove dependencies existing in qualitative analysis. These models describe the diffusion of contrast agent from blood plasma into the extracellular extravascular space (EES) using different assumptions and simplifications. Well-defined physiological parameters, such as tissue blood flow, blood volume and vessel permeability, are extracted by the deconvolution of the tissue response to the patient's AIF, through 1) parametric estimation in pharmacokinetic models [Brochot et al., 2006], 2) semi-parametric techniques based on adaptive smoothing of each observed TC [Schmid et al., 2009] and 3) non-parametric estimation based on Laplace deconvolution in the blood flow model [Comte et al., 2014].

1.2 Issues in DCE imaging

1.2.1 Issues for analysis of DCE image sequences

DCE imaging, however, suffers from several issues, which hamper its large diffusion. First of all, consensus is lacking for the analysis strategy. Secondly, voluntary and involuntary movements (such as breathing and cardiac motion, and non-periodic motions such as bowel peristalsis) limit the quality of the analysis, and can be addressed by motion correction methods [Glocker et al., 2011; Sotiras et al., 2009]. Finally, low SNR is a common issue in dynamic in vivo imaging due to the short time allowed to record the signal within the image sequence and affects any quantitative estimation[1]. For CT, because of the need to minimize the tissue irradiation during

[1] any quantitative estimation is aimed to solve an ill-posed deconvolution

the entire sequence, using low Kvs and MAs to reduce the total X ray dose leads to very poor quality for individual images. For MRI, high-speed sequences are obtained with a high cost in signal. The larger resolution we want without changing magnetic field, fewer protons arrive at the each voxel, therefore, the lower SNR is.

1.2.2 SNR improvement via segmentation

To improve SNR when analyzing DCE images, either large manual regions of interest (ROI) are drawn by the radiologist or image sequence is denoised by spatial filtering techniques. However, large ROI could end up with a lack of homogeneity by mixing different types of TCs, and, filtering techniques could trigger a tricky balance between information and noise when the TC shows variation of high frequency and induce a low spatial resolution. This work focus on this last issue.

DCE image sequence reveals tissues (organs, tumors, metastasis, vessels, bones, muscles, fat, etc.) having homogeneous properties that can be considered from an image processing point-of-view as objects in a scene. Consequently, one can think to recover these tissues/objects. It leads to segment the DCE image sequence into regions (of voxels) showing homogeneous TCs. This provides the opportunity to improve SNR without loss of spatial and temporal information, by averaging the TCs in these recovered regions. This alternative is known as *DCE image sequence segmentation*. In this context, DCE image sequence is considered as a spatial (2D or 3D) domain made of voxels. At each voxel, one can observe a TC in time domain. Hence, one needs a double representation of both spatial structure (voxel) and time structure (curve). Segmentation in this context is to obtain a partition of the spatial domain into subsets having their time structure homogeneous.

Depending on the plain positions of the DCE image sequences, the amount of tissues could potentially vary from only a few to many. These tissues could be filled, surrounded or interrupted by air, water or pure geometry effects (e.g. colon) but also appear at several disconnected locations in the image. Therefore, DCE image segmentation should have the ability in presence of noise to recover an

unknown number of homogeneous regions that could be numerous and spatially disconnected. Moreover, DCE image segmentation should take advantage of functional information embedded in the image sequence in order to preserve as much as possible the temporal information. From a statistical point-of-view, this is ensured by the use of adaptive methods (in time domain). However, let us point out that, as the functional information results from the convolution with the AIF, one can expect the signals to be smooth enough in Hölder spaces.

1.3 Statistical core of this PhD thesis

We present the statistical core of this PhD thesis in a general context. In this work, we propose a new method, called *Hierarchical Segmentation using Equivalence Test (HiSET)*, aiming to cluster functional (e.g. with respect to time) features or signals discretely observed with noise on a finite metric space considered to be a landscape and where the noise on the observations is assumed independent Gaussian with known constant variance. HiSET employs a multiple equivalence test derived from a multi-resolution comparison test in the time domain, which is known to be adaptive to the unknown Hölder regularity of signal, in order to compare two elements in the landscape. Considering the p-value of the multiple equivalence test as dissimilarity measure, HiSET is a bottom-up hierarchical clustering algorithm consisting of two steps: one local and one global.

In the local step (region growing), starting from a partition made of all elements of the metric space, only spatial neighbors may be aggregated into larger clusters. One can view the resulting partition of this local step as the collection of spatially adaptive super-elements.

In the global step (region merging), starting from the partition resulting from the local step, all clusters may be aggregated into larger clusters, regardless of their disconnectedness.

HiSET depends on two intuitive parameters: δ that controls the expected

homogeneity discrepancy between two functional features and α the significance level (upper bound of the Type I error) of the multiple equivalence test, which controls the probability to stop too late in the iterative process. Given α and δ, both steps stop automatically through a proper control of the Type I error, providing an adaptive choice of the number of clusters. Homogeneity discrepancy is ensured as parameter δ controls whether the difference of two functional features is close to 0 or not, regardless any modeling of these features. In this spirit, it is a human interpretable parameter as it is meaningful on noiseless observations.

For a set made of functionally piecewise homogenous subsets, HiSET is proven to be able to retrieve the true partition with high probability when the number of observation times is large enough.

In the application for DCE image sequence, the assumption is achieved by the modeling of the observed intensity in the sequence through a proper variance stabilization, which depends only on one additional parameter a. Combing this DCE imaging modeling step with our statistical core, HiSET, provides *DCE-HiSET*.

1.4 Medical applications

1.4.1 2D and 3D segmentation

We have applied DCE-HiSET on both 2D and 3D DCE image sequence segmentation. For 3D sequences, we have considered two strategies, a 2D-to-3D strategy and a direct 3D strategy. The 2D-to-3D strategy aims to propagate 2D-regions obtained from a first 2D-segmentation of each slice by doing the complete 3D-segmentation starting from the partition made of all 2D-partitions. The direct 3D strategy applies DCE-HiSET directly considering neighbors in the 6 directions (4 in the slice plus 2 to go from slice to slice).

1.4.2 ROI refinement and tumor classification

We also used DCE-HiSET as a preprocessing step to refine a roughly manually delineated ROI made by cliniciens. Followed by a series of erosion-dilation process, DCE-HiSET helps remove the regions only located at or connected to the border of the manual ROI while preserving the homogeneous regions connected to the interior of the ROI. We apply the result of this refinement to a series of DCE image sequences of 99 ovarian tumors graded with respect to their aggressiveness according to the biology from a biopsy. Using several classification models covering unsupervised and supervised approaches, we show that the classification with respect to their grade of these ovarian tumors based on the functional features observed in the ROI was improved after ROI refinement.

1.5 Reading guide

In Chapter 2, we review the existing clustering methods that have been applied to DCE image sequence segmentation. In Chapter 3, we describe the statistical model (§3.1), the multiple equivalence test (§3.2) and the two-step clustering procedure with its theoretical properties (§3.3). In Chapter 4, we apply and evaluate the proposed method on 2D DCE image sequence. We compare the proposed method to other state-of-the-art methods with synthetic DCE image sequence and study the parameter influence on the segmentation result with real DCE image sequences, along with the model verification. Additionally, in Chapter 5, an automatic strategy aiming to refine the manual ROI is described with an evaluation of robustness and efficiency. In Chapter 6, we evaluate the proposed method on 3D DCE image sequence. We explore and evaluate several segmentation strategies based on the proposed method. This work being supported by a CIFRE contract in the company Intrasense, Chapter 7 is devoted to present shortly the company, its medical imaging software Myrian®, together with how this work has been implemented in Myrian® and how the company may benefit from this implementation.

Chapter 2

Previous state of the art

2.1 Previous works related to DCE image sequence segmentation

Region (object)-based segmentation has been investigated for detecting lesions [Agner et al., 2013; Chen et al., 2006; Irving et al., 2016; McClymont et al., 2014; Shi et al., 2009; Shou et al., 2016; Stoutjesdijk et al., 2012; Tartare et al., 2014], or for retrieving internal structure of organs using a prior knowledge on the number of tissue in the organ of interest [Abdelmunim et al., 2008; Chevaillier et al., 2011; Lecoeur et al., 2009; Li et al., 2012a]. With respect to the requirement of training dataset, we divide all classification methods, which have been used in these works to classify voxels into classes (regions), into two categories: supervised and unsupervised. Supervised approaches require training dataset to build the classifiers that are used to classify new dataset. Unsupervised approaches do not require training dataset and are referred in this work as clustering-based image segmentation methods. Supervised approaches are usually used in conjunction with unsupervised approaches to achieve a more specific objective.

2.1.1 One-stage unsupervised approaches

Chen et al. [2006] used a fuzzy c-means (FCM) clustering to segment a predefined breast ROI into two classes: lesion and non-lesion. In this work, DCE image sequence has six acquisition times: one for pre-contrast and five for post-contrast. FCM takes enhancements at post-contrast times as input. After FCM, the soft membership map is binarized (two classes) with a empirically chosen threshold. And hole filling is needed to get a closed lesion.

Shi et al. [2009] also used FCM with two classes to segment suspicious breast lesion. After the hole filling operation to remove isolated voxels within lesion region, the level-set (LS) method was used to refine the boundary of lesion. In this work, each sequence has 10 images: one for pre-contrast and 9 for post-contrast, but only the standard deviation and maximum enhancement computed for every voxel are used for FCM.

Lecoeur et al. [2009] investigated the segmentation of lesion and internal structure in brain for multimodal (multi-channel as RGB) imaging, which is very similar to DCE imaging in the sense of image sequence. This work introduced a hierarchical segmentation method that iteratively uses graph cut (GC) to separate two classes/tissues (object and background), until expected number of classes/tissues are segmented. The energy function to be minimized in GC contains region information and border information. For region information, each class is modelled by a multivariate (univariate for each channel) Gaussian distribution. Border information is provided by the spectral gradient.

Chevaillier et al. [2011] proposed a two step method for the segmentation of internal structure (compartments) in kidney. In the first step, it constructs a graph structure with automatically determined number of nodes from all voxels, using a growing neural gas algorithm. In the second step, it merges the nodes into three classes using two thresholds of qualitative parameters based on time curves, which have 256 acquisition times.

For the same purpose, Li et al. [2012a] used a wavelet-based k-means clustering

with k equal to 5. Wavelet coefficients were selected by hard thresholding with minimizing Stein's unbiased risk. It used cosine distance of selected wavelet coefficient as distance for k-means.

Shou et al. [2016] developed soft null hypotheses in testing procedure to classify (identify) brain lesion voxels with enhancement out of other voxels without, leading to a binary classification: lesion and non-lesion. Several null hypotheses were proposed to quantify the qualitative behavior of time series. PCA was used to reduce the dimension. The third and fourth principle components (PCs) are considered as the representative of enhancement and are further used in testing. In this work, soft null hypotheses were compared to the clustering based on a mixture of k normal distributions with k varying. This work also proposed two points to argue the failure of clustering: there is a continuous change in intensity from non enhancement to strong enhancement, and lesion only accounts for a small portion of voxels that do not have high influence on the estimation of the parameters for mixture distributions.

For breast lesion segmentation, Agner et al. [2013] proposed to use a spectral embedding (SE) to represent time intensity curves with only 3 dimensions in order to have a stronger gradient and to provide more descriptive region statistics for active contour (AC) model, which is later used to define the boundary of lesion. Regional information is modelled by a mixture of two multivariate Gaussian distributions, one for lesion and the other for non-lesion. In this work, the proposed method, SE-AC, was compared to FCM-AC that uses FCM in conjunction with AC model.

Tartare et al. [2014] used spectral clustering to partition a ROI on prostate into clusters. Considering a k-nearest neighbor graph, the computed affinity matrix incorporates the spatial information of image. The number of clusters equals to the number of eigenvectors used in spectral representation of time curves and is automatically estimated using a normalized modularity criterion. With this number, voxels are clustered using k-means. Additionally, clusters were labelled by comparing the averaged curves to AIF.

McClymont et al. [2014] used a combination of two methods: mean shift (MS)

clustering and GC for breast lesion detection and delineation. MS clustering is able to produce spatially contiguous regions by treating spatial coordinates as part of input along with feature domain and does not need a predefined number of clusters. After removing clusters corresponding to vessels and non-enhanced tissues, a region adjacent graph was constructed considering each remaining cluster as a vertex. GC was made to separate lesion from surroundings. Like some earlier works, DCE image sequence has only 4 or 5 acquisition times: one for pre-contrast and others for post-contrast. For each voxel, MS used a feature vector consisting of its spatial coordinate and its enhancement at post-contrast times. GC used a measure based on the mean enhancement during post-contrast.

2.1.2 Multi-stage approaches mixing unsupervised and supervised methods

Stoutjesdijk et al. [2007, 2012] proposed a three-step strategy for the segmentation of lesion and subdivision in breast. Firstly, Otsu thresholding [Otsu, 1979] was applied on the first enhancement to delineate a lesion in breast. Secondly, mean shift (MS) clustering was used to subdivide the lesion into contiguous and homogeneous clusters using enhancement parameters. Finally, classifiers such as linear discriminative analysis (LDA) and support vector machine (SVM) were trained on kinetic parameters to select the most malignant ROI.

Irving et al. [2014, 2016] used a supervoxel over-segmentation to retrieve local homogenous regions and a parts-based graphical model to include global anatomical relationships. To get supervoxels, an adapted k-means clustering method, simple linear iterative clustering (SLIC), was used on the data represented by PCA with the first three PCs. SLIC is based on a new distance combining the feature distance and spatial distance. The number of supervoxels and the compactness parameter are required to be defined. Then supervised classifier LDA was trained to assign one of three classes (tumor, bladder and lumen) to each supervoxel. Supervoxel connectivity was represented by a graph and belief propagation was used to identify

three nodes for three classes from all supervoxel candidates. In addition, these works pointed out that supervoxels can be produced using methods including SLIC, MS, normalized cut (NC), etc.

2.1.3 Discussion

In summary, these works focused only on the partial segmentation for lesions and organs. However, the approaches that have been investigated in [McClymont et al., 2014; Irving et al., 2014, 2016] revealed a two-step framework that can be generalized to DCE image sequence segmentation, which, to remind, is a complete segmentation of the entire spatial domain in image. In this framework, the over-segmented supervoxels are generated using methods such as MS or SLIC in the first step, and then fully connected conditional random fields (CRF) or normalized cut (NC, multi-class version of GC for binary) propagates the supervoxels into a final partition of the entire spatial domain. In other words, the method according with this framework will also be a local-global method, as what we expected for our method. Hence, they appear to be excellent competitors.

2.2 Clustering methods for DCE image sequence segmentation

In this section, we will first describe the issues of DCE image sequence segmentation. Then, considering that DCE image sequence segmentation is a special case of general image segmentation, we will broaden the survey of literature and review the clustering methods, mentioned in the works from Section 2.1, in the perspective of image segmentation. Roughly speaking, we divide these clustering methods into three categories and discuss about their limitations. Finally, among these methods, we will introduce the technical details, according to the original problem setting, about five of them, which will be compared to our method later in Section 4.2.

2.2.1 Issues of DCE image sequence segmentation

In this thesis, we focus only on unsupervised approach that does not require a training dataset. Clustering-based methods involved in previous works provide only a partial (or binary) segmentation of the DCE image sequence. Despite being potentially adaptable to our purpose (complete segmentation of the spatial domain), these methods have never been used for this objective to our knowledge. Moreover, we do not expect to have a prior knowledge on the number of regions since we want to deal with large images potentially covering several organs, tumors or metastasis that are known to be heterogeneous.

With respect to time structure representation in DCE image sequence, both parametric and non-parametric approaches have been investigated. Parametric representations use a fixed number of qualitative descriptors based on curves (slope, ME, AUC, etc), or, of eigenvectors resulting from a principle component analysis (PCA) [Irving et al., 2016; Shou et al., 2016] or from a spectral embedding [Agner et al., 2013; Tartare et al., 2014] for all the curves. By construction, these approaches can not be adaptive to the curve regularity that may change from voxel to voxel. Non-parametric representations such as wavelet coefficient [Li et al.,

2012a] aim to handle this problem, as they are known to be adaptive in Sobolev spaces.

As mentioned above, segmenting DCE image sequence requires at the same time to take into account the similarity in the time domain as well as the regularity in the spatial domain. To get around the complexity of such combined approach, several works have proposed a direct clustering of the time domain features using k-means [Kachenoura et al., 2011; Li et al., 2012a] or fuzzy c-means (FCM) [Chen et al., 2006]. However, not taking into account the spatial structure, an extra post-processing step (e.g. hole-filling) is required to regularize the resulting segmentation [Chen et al., 2006] due to the presence of noise. To incorporate spatial and time domain structure in one global procedure, two main types of approaches arising from the general context of image segmentation have been investigated in the context of DCE image sequence: 1/ features are obtained by binding spatial information (voxel coordinates) to the time domain features [Comaniciu and Meer, 2002]; 2/ the time domain features are represented over the spatial domain as a finite mixture of simple objects (e.g. Gaussian model) and the regularity condition is incorporated as Markov random field (MRF) prior [Chatzis and Varvarigou, 2008; Wu et al., 2006].

2.2.2 Categories of clustering-based method

As mentioned above, the previous works focused only on partial segmentation. However, other clustering-based techniques that do not deal with DCE image sequences but only with classical (grey, color, texture) static images have considered the complete segmentation. We review their main ideas to offer a broader picture to readers about image segmentation. One can broadly classify clustering methods for image segmentation into three categories: model-based, graph-based and hybrid.

2.2.2.1 Model-based methods

Model-based methods aim to describe the feature space as a mixture of models. For this purpose, k-means [Li et al., 2012a] and FCM [Chen et al., 2006] tend to minimize an objective function corresponding to the within-cluster distance. Gaussian mixture models (GMMs) have been used in conjunction with MRF and formulate the maximum a posterior probability as objective function that is estimated by expectation-maximization (EM) [Celeux et al., 2003] or gradient descent/ascent [Chatzis and Varvarigou, 2008], with mean-field-like approximation, or by Bayesian sampling (MCMC) [Tu and Zhu, 2002]. k-means and FCM tend to find clusters with comparable shapes while EM in conjunction with GMMs allows clusters to have different shapes. In common, they all suffer from the choice of initial partition and from the use of a pre-specified number of clusters. They tend to fail in presence of clusters having complex and unknown shapes, as previously mentioned in [Zelnik-Manor and Perona, 2004].

Mean shift (MS) [Comaniciu and Meer, 2002] and quick shift (QS) [Vedaldi and Soatto, 2008] aim to find the modes of the distribution of feature space, and then assign each individual pixel to a mode by minimizing a criterion that is computed on features obtained by binding spatial coordinates and color information. MS and QS do not need a predefined number of clusters, however do need kernel bandwidths for both time and spatial domain. The choice of bandwidths is not automatic and requires domain expertise.

2.2.2.2 Graph-based methods

Graph-based methods treat image segmentation as a graph-partitioning problem. The weighted graph is constructed from image by considering each pixel as a node and by connecting each pair of pixels with an edge [Shi and Malik, 2000; Zelnik-Manor and Perona, 2004]. The weight on the edge reflects the similarity between two pixels with respect to either features or spatial distance or even both. Making use of the eigenvalue system of Laplacian matrix, these methods minimize a

global criterion to get a partition. Graph-based methods are able to handle feature space with more complex structure. However they do require the knowledge of the number of clusters and some scale parameters to compute affinity [Shi and Malik, 2000]. Zelnik-Manor and Perona [2004] addressed these two requirements through a self-tuning local scaling and a maximization of the cost function corresponding to the number of clusters ranging within a predefined set of values. Such approach has been used in DCE image sequences for prostate tumor segmentation [Tartare et al., 2014], with the number of clusters ranging only from 1 to 5. In the context of tumor segmentation in DCE image sequence, Irving et al. [2016] used a supervised step to discriminate tumorous tissue.

2.2.2.3 Hybrid methods

Hybrid methods consist of two steps – one local and one global – each derived from a model- or graph- based method described above [Hedjam and Mignotte, 2009; Tao et al., 2007]. Such hybrid methods have been applied to DCE image sequence [Irving et al., 2016; McClymont et al., 2014]. The local step over-segments the image into local homogeneous clusters, also referred as "superpixels" [Achanta et al., 2010] (or "supervoxel" in the context of DCE image sequence [Irving et al., 2016]), then the global step merges these superpixels into global clusters. The use of over-segmented homogeneous regions, instead of pixels, may help reduce the sensitivity to noise and hence result in better segmentation performance, as mentioned in [Tao et al., 2007]. However, they share similar defaults with the methods used in each step as described above.

2.2.3 k-means

k-means is a model-based method and aims to partition n observations into k clusters in which each observation belongs to the cluster with the nearest mean.

Let $(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n)$ denote a set of observations, each of which is a d-dimensional real vector. k-means clustering aims to partition the n observations into k sets

$\mathcal{P} = \{C_1, C_2, \ldots, C_k\}$ so as to minimize the within-cluster sum of squares (WCSS),

$$\arg\min_{\mathcal{P}} \sum_{i=1}^{k} \sum_{\mathbf{x} \in C_i} \|\mathbf{x} - \boldsymbol{\mu}_i\|^2,$$

where $\boldsymbol{\mu}_i$ is the mean of observations included in C_i. Given an initial set of k means $m_1^{(0)}, \ldots, m_k^{(0)}$, the algorithm proceeds with iterations alternating between two steps:

Assignment step: Assign each observation to the cluster whose mean yields the least WCSS.

$$C_i^{(t)} = \{x_p : \|x_p - m_i^{(t)}\|^2 \leq \|x_p - m_j^{(t)}\|^2, \qquad \forall j, 1 \leq j \leq k\},$$

where each x_p is assigned to only one cluster in $\mathcal{P}^{(t)}$, even if it could be assigned to two or more of them. Since the sum of squares is the squared Euclidean distance, this is intuitively the "nearest" mean.

Update step: Calculate the new means to be the centroids of the observations in the new clusters.

$$m_i^{(t+1)} = \frac{1}{|C_i^{(t)}|} \sum_{x_j \in C_i^{(t)}} x_j,$$

where $|C_i^{(t)}|$ is the cardinality of set $C_i^{(t)}$ indicating the number of observations within it. Since the arithmetic mean is a least-squares estimator, this also minimizes the WCSS objective.

The algorithm has converged when the assignments no longer change. Since both steps optimize the WCSS objective, and there only exists a finite number of such partitions, the algorithm must converge to a (local) optimum. However, there is no guarantee that the global optimum is achieved using this algorithm.

The initialization of k-means clustering is usually in a random fashion. Either the k means are randomly picked from all observations, or each observation is randomly assigned to a cluster among k clusters and then the means are calculated. In practice, several initializations are usually taken in order to have a minimal WCSS to start with for further iterations. However, due to the randomness, all clusters resulting from k-means tend to have the compatible sizes.

2.2.4 Hidden Markov random field with fuzzy c-means clustering (HMRF-FCM)

Hidden Markov random field (HMRF) models have been widely used for image segmentation, as they appear naturally in problems where a spatially constrained clustering scheme, taking into account the mutual influences of neighboring sites, is asked for. As another model-based method, fuzzy c-means (FCM) clustering has also been successfully applied in several image segmentation applications. Chatzis and Varvarigou [2008] proposed a novel approach formulated using an FCM-type algorithm regularized by Kullback-Leibler (KL) divergence information, and facilitated by employing a mean-field-like approximation of the MRF (prior) distribution.

Hidden Markov random field Let $X = \{1, \ldots, n\}$ be a finite set of sites. For every site $j \in X$, its state $x_j \in S$ where S is a set of all possible states. We denote a configuration of state values of X as $\mathbf{x} = (x_j)_{j \in X}$ and its space as the product space $\mathcal{X} = S^n$. A strictly positive probability distribution $p(\mathbf{x})$, $\mathbf{x} \in \mathcal{X}$, is called a *random field* [Maroquin et al., 1987]. A random field $p(\mathbf{x})$ is considered as a *Markov random field* if $j \in X$ is only conditionally dependent to its neighbors $\partial_j \subset X$

$$p(x_j|\mathbf{x}_{X-\{j\}}) = p(x_j|\mathbf{x}_{\partial_j}), \qquad \forall j \in X.$$

The joint distribution of a MRF is given by

$$p(\mathbf{x}|\beta) = \frac{\exp(-U(\mathbf{x}|\beta))}{\sum_{\mathbf{x}\in\mathcal{X}} \exp(-U(\mathbf{x}|\beta))},$$

where β is a parameter known as the inverse supercritical temperature and $U(\mathbf{x}|\beta)$ is the energy function in form of

$$U(\mathbf{x}|\beta) = \sum_{C\in\mathcal{C}} V_C(\mathbf{x}|\beta).$$

$\mathcal{C}$ is a class of subsets of the sites that contains all neighboring sites

$$\mathcal{C} = \{C \subset X | \forall j, \ell \in C, j \in \partial_\ell, \ell \in \partial_j\}.$$

$C \in \mathcal{C}$ is a clique and V_C stands for the clique potential associated with the clique C.

The calculation of $p(\mathbf{x}|\beta)$ involves all $\mathbf{x} \in \mathcal{X}$, which in general is hardly feasible in term of computational complexity. A simple yet successful solution is the pseudo-likelihood approximation of MRF [Besag, 1975]

$$p(\mathbf{x}|\beta) = \prod_{j=1}^{n} p(x_j|\mathbf{x}_{\partial_j}; \beta),$$

where

$$p(x_j|\mathbf{x}_{\partial_j}; \beta) = \frac{\exp(-\sum_{C \ni j} V_C(\mathbf{x}|\beta))}{\sum_{x_j \in S} \exp(-\sum_{C \ni j} V_C(\mathbf{x}|\beta))}.$$

Let $\mathbf{y}_j \in \mathbb{R}^d$ be the observation on each site $j \in X$. With a prior MRF $p(\mathbf{x})$,

$$p(\mathbf{y}, \mathbf{x}) = p(\mathbf{y}|\mathbf{x})p(\mathbf{x})$$

is known as a HMRF model [McLachlan and Peel, 2000]. By assuming

$$p(\mathbf{y}|\mathbf{x}) = \prod_{j=1}^{n} p(\mathbf{y}_j|x_j)$$

with $p(\mathbf{y}_j|x_j)$ being multivariate Gaussian, we have

$$p(\mathbf{y}|\mathbf{x}; \boldsymbol{\theta}) = \prod_{j=1}^{n} p(\mathbf{y}_j|x_j; \boldsymbol{\theta}_{x_j}) = \prod_{j=1}^{n} \mathcal{N}(\mathbf{y}_j|\boldsymbol{\mu}_{x_j}, \boldsymbol{\Sigma}_{x_j}), \tag{2.1}$$

where $\boldsymbol{\mu}_{x_j}$ and $\boldsymbol{\Sigma}_{x_j}$ are the mean and covariance matrix of the emission distribution of the hidden state at site j of the HMRF model, and $\boldsymbol{\theta}_{x_j}$ contains the elements of the $\boldsymbol{\mu}_{x_j}$ and $\boldsymbol{\Sigma}_{x_j}$.

In addition, given an estimate $\hat{\mathbf{x}}$ of $\mathbf{x}$, the mean-field-like approximation of MRF is defined as

$$p(\mathbf{x}|\beta) = \prod_{j=1}^{n} p(x_j|\hat{\mathbf{x}}_{\partial_j}; \beta), \tag{2.2}$$

which has been shown to be an interesting alternative to Bayesian sampling (e.g. Markov-chain Monte Carlo) [Chalmond, 1989] for the computation of the posterior probabilities.

Fuzzy c-means clustering Considering the typical problem of clustering n multivariate data points into k clusters, the standard FCM algorithm defines the fuzzy objective function, based on the distance between data point and the cluster center, as

$$J_\phi = \sum_{i=1}^{k}\sum_{j=1}^{n} r_{ij}^{\phi} d_{ij},$$

where $\phi \geq 1$ is a weighting exponent on each fuzzy membership function r_{ij} to control their degree of fuzziness, and is called the fuzzifier of the clustering algorithm. Introducing a regularization term based on entropy, a different objective function is proposed in [Miyamoto and Mukaidono, 1997]

$$J_\lambda = \sum_{i=1}^{k}\sum_{j=1}^{n} r_{ij} d_{ij} + \lambda \sum_{i=1}^{k}\sum_{j=1}^{n} r_{ij} \log r_{ij}.$$

Here, the entropy term works as the fuzzifier and the parameter λ is the model's degree of fuzziness for the fuzzy membership values. Another FCM variant proposed a regularization by KL information [Ichihashi et al., 2001] instead of entropy with objective function defined as

$$J_\lambda = \sum_{i=1}^{k}\sum_{j=1}^{n} r_{ij} d_{ij} + \lambda \sum_{i=1}^{k}\sum_{j=1}^{n} r_{ij} \log\left(\frac{r_{ij}}{\pi_i}\right), \tag{2.3}$$

where π_i is the prior probability of the i-th cluster.

FCM-type treatment of the HMRF model The fuzzy clustering-type treatment is formulated by introducing a HMRF-oriented modification of the regularized-by-KL-information fuzzy objective function (2.3),

$$Q_\lambda = \sum_{i=1}^{k}\sum_{j=1}^{n} r_{ij} d_{ij} + \lambda \sum_{i=1}^{k}\sum_{j=1}^{n} r_{ij} \log\left(\frac{r_{ij}}{\pi_{ij}}\right). \tag{2.4}$$

We define the fuzzy membership function r_{ij} in (2.3) as the point-wise posteriors with mean-field-like approximation (2.2)

$$p(x_j = i|\mathbf{y}_j) = p(x_j = i|\mathbf{y}_j, \hat{\mathbf{x}}_{\partial_j}).$$

The distance function d_{ij} is defined as the negative log-likelihood of the i-th state with respect to the j-th observation

$$d_{ij}(\boldsymbol{\theta}_i) = -\log p(\mathbf{y}_j|x_j = i; \boldsymbol{\theta}_i). \tag{2.5}$$

π_{ij} is the point-wise prior probability also based on the mean-field-like approximation (2.2)

$$\pi_{ij} = p(x_j = i|\hat{\mathbf{x}}_{\partial_j}; \beta) = \frac{\exp(-\sum_{C\ni j} V_C(\tilde{\mathbf{x}}_{ij}|\beta))}{\sum_{h=1}^{k} \exp(-\sum_{C\ni j} V_C(\tilde{\mathbf{x}}_{hj}|\beta))},$$

where $\tilde{\mathbf{x}}_{ij} = (x_j = i, \hat{\mathbf{x}}_{\partial_j})$. Eventually, the objective function (2.4) becomes

$$Q_\lambda(\mathbf{\Psi}) = -\sum_{i=1}^{k}\sum_{j=1}^{n} r_{ij} \log p(\mathbf{y}_j|x_j = i; \boldsymbol{\theta}_i) + \lambda \sum_{i=1}^{k}\sum_{j=1}^{n} r_{ij} \log\left(\frac{r_{ij}}{\pi_{ij}}\right). \tag{2.6}$$

where $\mathbf{\Psi} = \{R, \boldsymbol{\theta}, \beta\}$ and $R = \{r_{ij}\}$.

HMRF model parameters estimation using the HMRF-FCM algorithm
To obtain an estimate of the HMRF model parameters through the HMRF-FCM algorithm, given a model fitting dataset, we have to iteratively minimize the fuzzy objective function Q_λ, given by (2.6), over R, $\boldsymbol{\theta}$, and β, in a coordinate descent fashion. From t-th to $(t+1)$-th iteration, HMRF-FCM algorithm, on the basis of a mean-field-like approximation of the MRF prior, consists of the following steps.

1. Estimate $\mathbf{x}^{(t)}$ by assigning each site to the state to maximize its fuzzy membership function
$$x_j^{(t)} = \arg\max_{i=1}^{k} r_{ij}^{(t)}.$$

2. Compute the point-wise prior probabilities $\pi_{ij}^{(t)}$
$$\pi_{ij}^{(t)} = p(x_j = i|\mathbf{x}_{\partial_j}^{(t)}; \beta^{(t)}) = \frac{\exp(-\sum_{C\ni j} V_C(\tilde{\mathbf{x}}_{ij}^{(t)}|\beta))}{\sum_{h=1}^{k} \exp(-\sum_{C\ni j} V_C(\tilde{\mathbf{x}}_{hj}^{(t)}|\beta))},$$
where $\tilde{\mathbf{x}}_{ij}^{(t)} = (x_j = i, \hat{\mathbf{x}}_{\partial_j}^{(t)})$ and is the current estimate of the neighborhood of site j.

3. Compute the distance $d_{ij}^{(t)}$ by combining (2.1) and (2.5),
$$d_{ij}^{(t)} = \frac{w}{2}\log(2\pi) + \frac{1}{2}\log|\mathbf{\Sigma}_i^{(t)}| + \frac{1}{2}(\mathbf{y}_j - \boldsymbol{\mu}_i^{(t)})^T \mathbf{\Sigma}_i^{(t)^{-1}} (\mathbf{y}_j - \boldsymbol{\mu}_i^{(t)}).$$

4. Compute the fuzzy membership functions $r_{ij}^{(t+1)}$ from the derivation of (2.2)
$$r_{ij}^{(t+1)} = \frac{\pi_{ij}^{(t)} \exp\left(-(1/\lambda) d_{ij}^{(t)}\right)}{\sum_{h=1}^{k} \pi_{hj}^{(t)} \exp\left(-(1/\lambda) d_{hj}^{(t)}\right)}.$$

5. Update the estimator $\boldsymbol{\mu}_i^{(t+1)}$ and $\boldsymbol{\Sigma}_i^{(t+1)}$ from the derivation of (2.2)

$$\boldsymbol{\mu}_i^{(t+1)} = \frac{\sum_{j=1}^n r_{ij}^{(t)} \mathbf{y}_j}{\sum_{j=1}^n r_{ij}^{(t)}}, \qquad \boldsymbol{\Sigma}_i^{(t+1)} = \frac{\sum_{j=1}^n r_{ij}^{(t)} (\mathbf{y}_j - \boldsymbol{\mu}_i^{(t)})(\mathbf{y}_j - \boldsymbol{\mu}_i^{(t)})^T}{\sum_{j=1}^n r_{ij}^{(t)}}.$$

6. Update the inverse temperature parameter $\beta^{(t+1)}$ from the derivation of (2.2)

$$\beta^{(t+1)} = \arg\max_{\beta} \sum_{i=1}^{k} \sum_{j=1}^{n} \log p(x_j = i | \mathbf{x}_{\partial_j}^{(t)}; \beta),$$

   which depends the choice of the energy function.

7. In case of convergence,

$$|Q_\lambda(\boldsymbol{\Psi}^{(t+1)}) - Q_\lambda(\boldsymbol{\Psi}^{(t)})| / Q_\lambda(\boldsymbol{\Psi}^{(t)}) < T_c,$$

   where T_c is the convergence threshold, exit; otherwise $t \leftarrow t+1$ and return to step 1.

For image segmentation For the application on image segmentation problem, we impose HMRF on the image with sites standing for pixel and states for segment labels. Given a neighborhood system (4- or 8-neighbor), the energy function is defined as

$$U(\mathbf{x}|\beta) = -\beta \sum_{j \in X} \sum_{\ell \in \partial_j} \delta(x_j - x_\ell), \qquad \delta(x_j - x_\ell) = \begin{cases} 1, & \text{if } x_j = x_\ell \\ 0, & \text{otherwise,} \end{cases}$$

leading to a reformulation of MRF prior $\pi_{ij}^{(t)}$ as

$$\pi_{ij}^{(t)} = \frac{\exp(\beta^{(t)} \sum_{\ell \in \partial_j} \delta(i - x_\ell^{(t)}))}{\sum_{h=1}^k \exp(\beta^{(t)} \sum_{\ell \in \partial_j} \delta(h - x_\ell^{(t)}))}.$$

In practice, given a neighborhood system, user needs to define the regularization parameter λ in (2.6) and the convergence threshold T_c (or the maximum number of iterations, considering the computational cost).

2.2.5 Mean shift (MS) clustering

Comaniciu and Meer [2002] proposed a model-based clustering approach based on the mean shift procedure, which is an adaptive gradient ascent method widely used for density estimation problem. This approach is attractive because it produces spatially contiguous regions and does not require a predefined number of clusters.

Considering $\{\mathbf{x}_i \in \mathbf{R}^d, i = 1, \ldots, n\}$ as a set of feature vector in d-dimensional space, the mode of the density $f(\mathbf{x})$ closest to $\mathbf{x}_i$ can be located at the maxima of kernel density estimator

$$\hat{f}(\mathbf{x}) = \frac{1}{nh^d} \sum_{i=1}^{n} K\left(\frac{\mathbf{x} - \mathbf{x}_i}{h}\right),$$

where $K(\mathbf{x})$ is a multivariate kernel with profile function $k(x)$ and h is bandwidth parameter. We use radially symmetric kernel satisfying

$$K(\mathbf{x}) = c_{k,d} k(\|\mathbf{x}^2\|),$$

where $c_{k,d}$ is the normalization constant making $K(\mathbf{x})$ integrate to one and assumed to be strictly positive. Starting with $\mathbf{y}_{i,1} = \mathbf{x}_i$, the maxima is reached at the stationary point $y_{i,c}$ of an iterative process

$$\mathbf{y}_{i,j+1} = \frac{\sum_n^{i=1} \mathbf{x}_i k(\|\frac{\mathbf{y}_j - \mathbf{x}_i}{h}\|^2)}{\sum_n^{i=1} k(\|\frac{\mathbf{y}_j - \mathbf{x}_i}{h}\|^2)}.$$

For image, a joint-domain kernel $K_{h_s h_r}$ is defined, so as to combine the spatial and the feature domain, as the product of two radially symmetric kernels and the Euclidean metric allows a single bandwidth for each domain

$$K_{h_s h_r}(\mathbf{x}) = \frac{c}{h_s^{d_s} h_r^{d_r}} k\left(\left\|\frac{\mathbf{x}^s}{h_s}\right\|^2\right) k\left(\left\|\frac{\mathbf{x}^r}{h_r}\right\|^2\right),$$

where

- $\mathbf{x}^s$ and $\mathbf{x}^r$ are the spatial and feature components of $\mathbf{x}$, with respectively d_s and d_r dimension;
- $d_s + d_r = d$;

- h_s and h_r are the bandwidth parameters for the spatial and feature domains respectively;
- c is the normalization constant.

After finding the associated mode y_i for voxel x_i, $i = 1, \ldots, n$, we group together all y_i that are closer than h_s in spatial domain and h_r in the feature domain, to form the clusters $\{C_\ell\}_{\ell=1,\ldots,\ell^*}$. Finally, the label ℓ is assigned to each x_i if $x_i \in C_\ell$.

In practice, an Epanechnikov or a (truncated) normal kernel always provides satisfactory performance, so the user only needs to set the bandwidth parameter $\mathbf{h} = (h_s, h_r)$, which, by controlling the size of the kernel, determines the resolution of the mode detection.

2.2.6 Normalized cut (NC)

Shi and Malik [2000] proposed normalized cut for image segmentation by representing the image as a graph. In graph clustering problem, we seek to partition the set of vertices into disjoint sets $C_1, C_2, \ldots, C_k$, where by some measure the similarity among the vertices in a set C_i is high and, across two different sets C_i and C_j is low.

From an image grid $\mathcal{V} = \{1, 2, \ldots, n\}$, we construct the graph $G = (\mathcal{V}, E)$ by taking each pixel $i \in \mathcal{V}$ as a node and define the edge weight w_{ij} between node i and j as the product of a feature similarity term and spatial proximity term:

$$w_{ij} = \exp^{\frac{-\|\mathbf{y}_i - \mathbf{y}_j\|_2^2}{\sigma_y}} \times \begin{cases} \exp^{\frac{-\|\mathbf{x}_i - \mathbf{x}_j\|_2^2}{\sigma_x}} & \text{if } \|\mathbf{x}_i - \mathbf{x}_j\|_2 < r \\ 0 & \text{otherwise,} \end{cases} \tag{2.7}$$

where $\mathbf{x}_i$ is the spatial location of node i, and $\mathbf{y}_i$ is a d-dimensional feature vector based on pixel intensity.

Partitioning a graph into two disjoint sets relates to the minimization of the following objective function

$$\min_{\mathbf{e}} \mathrm{Ncut}(\mathbf{e}) = \min_{\mathbf{e}} \frac{\mathbf{e}^T(D - W)\mathbf{e}}{\mathbf{e}^T D\mathbf{e}}, \qquad \mathbf{e}^T D\mathbf{1} = 0,$$

where W is an $n \times n$ symmetrical matrix with $W(i,j) = w_{ij}$ and D is an $n \times n$ diagonal matrix with $D(i,i) = \sum_j w_{ij}$. This is realized by solving the generalized eigenvalue system of Laplacian matrix $D - W$

$$(D - W)\mathbf{e} = \lambda D\mathbf{e},$$

and the eigenvector with the second smallest eigenvalue of the system is the solution of the bi-partitioning (Ncut) problem.

In order to achieve a k-partition, the proposed algorithm performs a recursive two-way Ncut. It consists of the following steps:

1. Given a set of features, set up a weighted graph $G = (\mathcal{V}, E)$, compute the weight on each edge, and summarize the information into W and D.

2. Solve $(D - W)\mathbf{e} = \lambda D\mathbf{e}$ for eigenvectors with the smallest eigenvalues.

3. Use the eigenvector with the second smallest eigenvalue to bipartition the graph by finding the splitting point along the eigenvector such that Ncut is minimized.

4. Decide if the current partition should be subdivided by checking the stability of the cut, and make sure Ncut is below the pre-specified value, or the expected number of segments is reached.

5. Recursively repartition the segmented parts if necessary.

In practice, user needs to define the number of expected segments, the kernel window σ_x and σ_y, and the radius of neighborhood r in (2.7).

2.2.7 Simple linear iterative clustering (SLIC)

Achanta et al. [2010] proposed a new approach dedicated to produce high quality, compact and nearly uniform superpixels in an efficient way. For an image grid with n pixels, an expected number of superpixels k is defined such that each one consists of approximately n/k pixels and is centered at every $s = \sqrt{n/k}$ pixels in

the grid. Let $\mathbf{x}_i$ be the spatial location of voxel i and $\mathbf{y}_i$ be a d-dimensional feature vector based on pixel intensity. A novel distance measure D_s is defined, in between voxel i and a superpixel center c, as the sum of feature dissimilarity and the spatial distance normalized by the grid interval, s, of superpixel centers:

$$D_s = \|\mathbf{x}_c - \mathbf{x}_i\|_2 + \frac{m}{s}\|\mathbf{y}_c - \mathbf{y}_i\|_2, \tag{2.8}$$

where m is a parameter allowing users to control the compactness of a superpixel. The greater the value of m is, the more spatial proximity is emphasized and the more compact the superpixel is.

To be specific, the SLIC algorithm consists of the following steps:

1. Sample k regularly spaced cluster centers.

2. Perturb every cluster center in a $r \times r$ neighborhood to the position with lowest gradient that is computed as

$$G(x, y) = \|\mathbf{y}_{x+1,y} - \mathbf{y}_{x-1,y}\|_2^2 + \|\mathbf{y}_{x,y+1} - \mathbf{y}_{x,y-1}\|_2^2,$$

   where $\mathbf{y}_{x,y}$ is the feature vector of the pixel at position (x, y).

3. Assign every pixel to the cluster whose center is the nearest one to it according to the distance measure D_s defined in (2.8), while the pixel is also within the $2s \times 2s$ squared neighborhood around the cluster center.

4. Compute new cluster centers and residual error E that is the L_1 distance between previous centers and updated centers.

5. If $E \leq$ threshold, stop; otherwise, repeat step 3 and 4.

6. Reassign the small segments, which have the same label as a large cluster but disconnected to it, to their largest neighboring cluster, in order to enforce connectivity.

In practice, user needs to define the number of superpixels k, the compactness parameter m, the radius of neighborhood r for center perturbation and the threshold for the convergence.

Chapter 3

DCE-HiSET

3.1 DCE statistical model and objective

In this section, we take MRI as an example and describe how to adapt a very simple and general statistical model to intensity observed in MR images. Then, we propose the independence assumption on this model and state the statistical objective based on it.

For MRI, at each voxel location x on a finite grid $\mathcal{X}$ describing the image cross-section,

$$\Phi^x(t_j) \in \mathbf{R}_+, \qquad j = 1, ..., n$$

are observed at the n acquisition times, $t_1, t_2, ..., t_n$. Since the observed intensity, at time t_j, $\Phi^x(t_j)$ is proportional to the amount of contrast agent arriving in the voxel x at time t_j (with or without the baseline effect of the tissue), one can imagine that $\Phi^x(t_j)$ follows a Poisson distribution with parameter $\phi^x(t_j) \in \mathbf{R}_+$:

$$\Phi^x(t_j) \sim \mathcal{P}(\phi^x(t_j)).$$

3.1.1 First assumption

In this context, $\phi^x(t_j)$ may be considered large due to the injected contrast agent.

Assumption 1 *We assume that $\phi^x(t_j)$ follows Poisson distribution that can be approximated by a Gaussian distribution:*

$$\Phi^x(t_j) \sim \mathcal{P}(\phi^x(t_j)) \approx \mathcal{N}(\phi^x(t_j), \phi^x(t_j)). \tag{3.1}$$

According to Assumption 1, we can rewrite $\Phi^x(t_j)$ as

$$\Phi^x(t_j) := \phi^x(t_j) + \sigma_j^x \varepsilon_j^x, \quad \text{where} \quad \sigma_j^x = \sqrt{\phi^x(t_j)} \quad \text{and} \quad \varepsilon_j^x \sim \mathcal{N}(0,1). \tag{3.2}$$

Hence, we have an initial heteroscedastic model as (3.2) and the variance of observed intensity depends on both spatial location of voxel x and time index j.

3.1.2 Variance Stabilization

We use a function $f(\cdot) : \mathbf{R} \to \mathbf{R}$ to transform $\Phi^x(t_j)$ such that the variance of the transformed intensity $f(\Phi^x(t_j))$ always equals to 1 regardless of x and t_j,

$$f(\Phi^x(t)) \sim \mathcal{N}(f(\phi^x(t)), 1). \tag{3.3}$$

With the first-order Tylor series expansion of function f, we have

$$f(\Phi^x(t)) - f(\phi^x(t)) = (\Phi^x(t) - \phi^x(t)) \cdot f'(\phi^x(t)) + o(\sigma_t^x).$$

In order to make the right side of last equality has variance as 1,

$$\begin{aligned} \operatorname{Var}\left[f(\Phi^x(t)) - f(\phi^x(t))\right] &= 1 \\ \operatorname{Var}\left[\Phi^x(t) - \phi^x(t)\right] \cdot f'(\phi^x(t))^2 &= 1 \\ \phi^x(t) \cdot f'(\phi^x(t))^2 &= 1 \\ f'(\phi^x(t)) &= \left[\phi^x(t)\right]^{-\frac{1}{2}}, \end{aligned} \tag{3.4}$$

we deduce that

$$f(\phi^x(t)) = 2\left[\phi^x(t)\right]^{\frac{1}{2}}.$$

In summary, the classical variance stabilization associated to Assumption 1 is

$$2\sqrt{\Phi^x(t_j)} - 2\sqrt{\phi^x(t_j)} \sim \mathcal{N}(0,1).$$

3.1.3 Second (Generalized) assumption

In practice, however, the Poisson model may appear not flexible enough since the variance of the observed intensity $\Phi^x(t_j)$ could be not exactly $\phi^x(t_j)$.

Assumption 2 *We relax Assumption 1, with respect to the Gaussian distribution, to*

$$\Phi^x(t_j) \sim \mathcal{N}\left(\phi^x(t_j), [\phi^x(t_j)]^{1+\tau}\right), \quad \textit{where} \quad -1 < \tau < 1. \tag{3.5}$$

To make the same variance stabilization as in Assumption 1, we deduce a general form of f,

$$f(\psi) = \frac{2}{1-\tau}\psi^{\frac{1-\tau}{2}}.$$

Therefore, by denoting $a := (1-\tau)/2$, we relax (3.3) such that there exists $0 < a < 1$,

$$\Phi^x(t_j) \sim \mathcal{N}\left(\phi^x(t_j), [\phi^x(t_j)]^{2-2a}\right), \tag{3.6}$$

whose variance is stabilized by the following relation:

$$\frac{(\Phi^x(t_j))^a}{a} - \frac{(\phi^x(t_j))^a}{a} \sim \mathcal{N}(0,1). \tag{3.7}$$

3.1.4 Intensity model

Considering a fixed a, we denote

$$J^x(t_j) := \frac{(\Phi^x(t_j))^a}{a} \quad \text{and} \quad q^x(t_j) := \frac{(\phi^x(t_j))^a}{a}, \tag{3.8}$$

and link them by the following model

$$J^x(t_j) = q^x(t_j) + \eta_j^x, \qquad j = 1, \ldots, n \text{ and } x \in \mathcal{X}, \tag{3.9}$$

where η_j^x are standard normal variables independent with respect to time index j (thanks to the production of one image at each time and to the significant pause in between the consecutive acquisition times), however not necessarily independent with respect to the spatial index x (MRI and CT being known to produce spatial artifacts [Graves and Mitchell, 2013; Herman, 2009] that translate into noise correlations).

Because contrast agent is always injected a few seconds after acquisition starts, we rewrite

$$q^x(t) = b^x + q_0^x(t),$$

where $q_0^x(t) = 0$ for $t < t^*$ and b^x to be baseline grey-level at voxel x, and assume that there exists n_0 such that $q_0^x(t_j) = 0$ for all $j = 1, \ldots, n_0$ at $x \in \mathcal{X}$. From now on, whenever *time curve* (TC) refers to time intensity curve (resp. time enhancement curve), we denote q^x (resp. q_0^x) by i^x and assume that we observe

$$I^x(t_j) = i^x(t_j) + \eta_j^x, \qquad j = 1, \ldots, n \text{ and } x \in \mathcal{X}. \tag{3.10}$$

In this model, $I^x := (I^x(t_1), \ldots, I^x(t_n))$ appears as a discretely observed noisy version of the unobservable true TC, i^x, at time $t_1, \ldots, t_n$.

For CT, we propose the same model for the enhancement, or even for the intensity after a linear transformation in order to take into account the possible non-positive values.

Despite being not able to justify this model from a physical point-of-view, we obtain from our experiments not only excellent results but also *a posteriori* justification of such model (see Section 4.5.1).

In DCE imaging, spatial artifacts that affect a single image are not affecting the full sequence in the same way. For example, one can think of radial artifacts whose direction or center may vary from image to image, or of band artifacts that do not appear at the same coordinate. As a consequence, such spatial artifacts (when they exist) are not affecting the entire TC and their effects may be strongly reduced (especially considering the global noise level of such DCE image sequences) or even (randomly) compensated. Therefore, we will make the following assumption:

Assumption 3 *Random variables η_j^x in* (3.10) *are independent with respect to both spatial location x and time index j.*

3.1.5 Objective

We describe our statistical objective as to build a partition of $\mathcal{X}$ made of ℓ non-overlapping clusters (regions),

$$\mathcal{X} = C_1 \cup C_2 \cup \ldots \cup C_\ell,$$

such that $x, y \in \mathcal{X}$ belong to the same cluster if and only if $i^x(.) = i^y(.)$. We propose to achieve this objective by answering the question "are $i^x(.)$ and $i^y(.)$ equal or not?" based on their discrete observations I^x and I^y.

3.2 Equivalence test and dissimilarity measure

In this section, based on the proposed statistical model and assumption of independence, we build the multiple equivalence test to compare two observed TCs, in order to prove their equality. We adopt the p-value of this test as the dissimilarity measure, which is further used in the proposed clustering algorithm in Section 3.3.

Given a set of voxels $X \subset \mathcal{X}$, we denote

$$\bar{I}^X := \frac{1}{|X|} \sum_{x \in X} I^x \quad \text{and} \quad \bar{i}^X := \frac{1}{|X|} \sum_{x \in X} i^x,$$

where $i^x = (i^x(t_1), \ldots, i^x(t_n))$ and $|X|$ is the cardinality of X. It follows Gaussian distribution

$$\bar{I}^X \sim \mathcal{N}\left(\bar{i}^x, \frac{1}{|X|}\mathrm{Id}_n\right). \tag{3.11}$$

Given another set Y such that $X \cap Y = \emptyset$, the difference between two averages is

$$\bar{I}^X - \bar{I}^Y \sim \mathcal{N}\left(\bar{i}^X - \bar{i}^Y, \left(\frac{1}{|X|} + \frac{1}{|Y|}\right)\mathrm{Id}_n\right). \tag{3.12}$$

We define the normalized difference as

$$D^{XY} := \frac{\bar{I}^X - \bar{I}^Y}{\rho(X,Y)}, \tag{3.13}$$

which also follows Gaussian distribution under Assumption 3,

$$D^{XY} \sim \mathcal{N}(d^{XY}, \mathrm{Id}_n), \tag{3.14}$$

where

$$d^{XY} := \frac{\bar{i}^X - \bar{i}^Y}{\rho(X,Y)} \qquad \text{and} \qquad \rho^2(X,Y) := \frac{1}{|X|} + \frac{1}{|Y|}. \tag{3.15}$$

We propose to decide that the average TCs of X and Y are similar if d^{XY} is not different from the zero vector.

3.2.1 Statistical hypothesis test on equivalence

In statistical hypotheses testing, the research or alternative hypothesis represents what the study aims to put in evidence. The burden of proof is on the alternative

in the sense that it is established only if there is enough evidence in its favor. In conventional (two-sided) comparative study,

$$\mathcal{H}_0 : d^{XY} = 0 \quad \text{versus} \quad \mathcal{H}_1 : d^{XY} \neq 0, \tag{3.16}$$

the research hypothesis is the hypothesis of difference, $\mathcal{H}_1$.

However, in order to construct clusters in a merging process, we want to prove the equivalence between TCs of voxels such that we can merge the voxels together. This fits the goal of equivalence test [Walker and Nowacki, 2010] that aims to demonstrate equivalence. Therefore the burden of proof rests on equivalence. In essence, the null and alternative hypotheses in equivalence test are simply those of a conventional comparative study reversed,

$$\mathcal{H}_0 : d^{XY} \neq 0 \quad \text{versus} \quad \mathcal{H}_1 : d^{XY} = 0. \tag{3.17}$$

The term "equivalence" recalls that one proves in this context the equality, which defines an equivalence relation, however equivalence is not defined in the strict sense, but rather to mean that two random variables are close enough, up to a given margin, such that their means cannot be distinguishable. This margin is called *'equivalence margin'*.

3.2.2 Equivalence test

We denote the normalized difference D^{XY} defined in (3.13) and its expectation d^{XY} in (3.15) as

$$D^{XY} = \{D_j^{XY}, j = 1, 2, \ldots, n\}, \qquad d^{XY} = \{d_j^{XY}, j = 1, 2, \ldots, n\}, \tag{3.18}$$

where

$$D_j^{XY} = \frac{\frac{1}{|X|}\sum_{x\in X} I^x(t_j) - \frac{1}{|Y|}\sum_{y\in Y} I^y(t_j)}{\rho(X,Y)}, \quad d_j^{XY} = \frac{\frac{1}{|X|}\sum_{x\in X} i^x(t_j) - \frac{1}{|Y|}\sum_{y\in Y} i^y(t_j)}{\rho(X,Y)}. \tag{3.19}$$

3.2.2.1 Gaussian case

The construction of equivalence test under Gaussian assumption can be found in [Wellek, 2010]. Given the equivalence margin δ, the null and alternative hypothesis of equivalence test for each $j = 1, \ldots, n$ are defined as

$$\mathcal{H}_0^j : d_j^{XY} < -\delta \quad \text{or} \quad d_j^{XY} > \delta \quad \text{versus} \quad \mathcal{H}_1^j : -\delta \leqslant d_j^{XY} \leqslant \delta. \tag{3.20}$$

According to (3.19), we have

$$D_j^{XY} - d_j^{XY} \sim \mathcal{N}(0, 1). \tag{3.21}$$

Therefore, the confidence interval of d_j^{XY} with confidence level α is

$$[D_j^{XY} - \Psi^{-1}(1 - \alpha/2), \quad D_j^{XY} + \Psi^{-1}(1 - \alpha/2)], \tag{3.22}$$

where $\Psi(.)$ is the cumulative distribution function of the standard normal distribution and its inverse function $\Psi^{-1}(.)$ is the quantile function. If the confidence interval (3.22) is included in $[-\delta, \delta]$, we will accept the equivalence hypothesis $\mathcal{H}_1^j$, otherwise we accept the non-equivalence hypothesis $\mathcal{H}_0^j$, as shown in Figure 3.1.

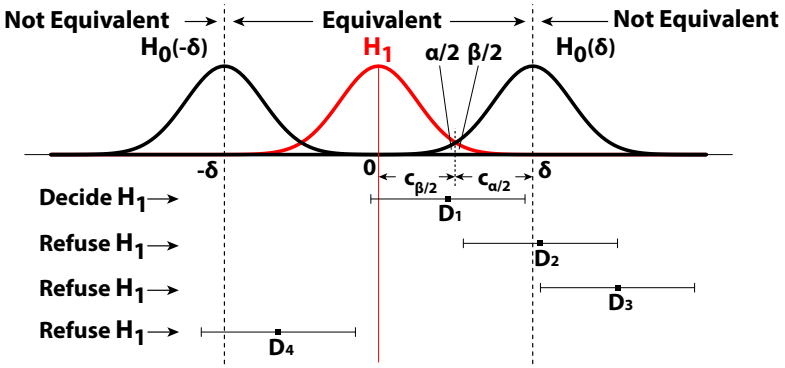


Figure 3.1: Equivalent test in Gaussian case: if the confidence interval is included in $[-\delta, \delta]$, equivalence is accepted. $c_a = \Psi^{-1}(1 - a)$.

3.2.2.2 Chi-square case

Using equivalence test under Gaussian assumption in our model, we would have to run n comparisons (one per time index) and face a multiplicity problem, which

becomes important when n is large. To control this multiplicity, we follow the works of [Baraud et al., 2005] and [Durot and Rozenholc, 2006], and use dyadic decomposition of the time indexes to project the D^{XY} onto sub-spaces such that only $\lfloor \log_2 n \rfloor$ tests are needed to perform the comparison.

Denoting $K_0 := \lfloor \log_2 n \rfloor - 1$, we first define an almost regular partition $\mathcal{T}^{K_0}$ of the index set $\{1, \ldots, n\}$ into 2^{K_0} sets,

$$\mathcal{T}^{K_0} = \left\{ T_r, r = 1, \ldots, 2^{K_0} \right\} \quad \text{where} \quad T_r = \left\{ j \mid \frac{r-1}{2^{K_0}} < \frac{j}{n} \leqslant \frac{r}{2^{K_0}} \right\}.$$

Then, we build $K_0 + 1$ partitions of $\{1, \ldots, n\}$, denoted by $\mathcal{T}^K$ and of size 2^K for $K = 0, \ldots, K_0$, by grouping sets T_r in a pairwise fashion:

$$\mathcal{T}^K := \left\{ T_k^K = \bigcup_{(k-1)/2^K < r/2^{K_0} \leqslant k/2^K} T_r,\ k = 1, \ldots, 2^K \right\}.$$

Regular projection For each K, we consider the projection of D^{XY}, denoted by $\Pi_K D^{XY}$, onto the vectors with constant component on each T_k^K of $\mathcal{T}^K$:

$$\Pi_K D^{XY} := (\underbrace{m_1^K, \ldots, m_1^K}_{n_1^K}, \underbrace{m_2^K, \ldots, m_2^K}_{n_2^K}, \ldots, \underbrace{m_{2^K}^K, \ldots, m_{2^K}^K}_{n_{2^K}^K}),$$

where

$$m_k^K = \frac{1}{n_k^K} \sum_{j \in T_k^K} D_j^{XY}, \quad n_k^K = |T_k^K|.$$

Then

$$\|\Pi_K D^{XY}\|_n^2 = \sum_{k=1}^{2^K} n_k^K (m_k^K)^2 = \sum_{k=1}^{2^K} \left(\frac{1}{\sqrt{n_k^K}} \sum_{j \in T_k^K} D_j^{XY} \right)^2.$$

Orthogonal projection Instead of using a test based on $\|\Pi_K D^{XY}\|_n^2$ as in [Baraud et al., 2005, 2003], we consider orthogonal projections to ensure their independence, which will be used later to control our clustering procedure. More precisely, starting from $R_0 = D^{XY}$,

$$\Pi_0 R_0 := (\underbrace{m_1^0, m_1^0, ..., m_1^0}_{n}), \quad m_1^0 = \frac{1}{n} \sum_{j=1}^{n} D_j^{XY},$$

we consider the K-th *residual after projection*, R_K, for $K = 1, \ldots, K_0$, as

$$R_K := R_{K-1} - \Pi_{K-1} R_{K-1} = (D_1^{XY} - m_1^{K-1}, D_2^{XY} - m_1^{K-1}, ..., D_n^{XY} - m_{2^{K-1}}^{K-1}).$$

The projection on the interval set $\{T_k^K, k = 1, ..., 2^K\}$, which are the dyadic decomposition of $\{T_k^{K-1}, k = 1, ..., 2^{K-1}\}$, is

$$\Pi_K R_K := (\underbrace{m_1^K - m_1^{K-1}}_{n_1^K}, \underbrace{m_2^K - m_1^{K-1}}_{n_2^K}, \ldots, \underbrace{m_{2^K}^K - m_{2^{K-1}}^{K-1}}_{n_{2^K}^K}).$$

The square norm of the orthogonal projection $\Pi_K R_K$ is therefore

$$\|\Pi_K R_K\|_n^2 = \sum_k^{2^K} n_k^K (m_k^K - m_{\lceil k/2 \rceil}^{K-1})^2 = \sum_k^{2^K} (\underbrace{\sqrt{n_k^K}(m_k^K - m_{\lceil k/2 \rceil}^{K-1})}_{M_k^K})^2,$$

where

$$\begin{aligned} M_k^K &= \frac{1}{\sqrt{n_k^K}} \sum_{j \in T_k^K} D_j^{XY} - \frac{\sqrt{n_k^K}}{n_{\lceil k/2 \rceil}^{K-1}} \sum_{j \in T_{\lceil k/2 \rceil}^{K-1}} D_j^{XY} \\ &\sim \mathcal{N}\left(\frac{1}{\sqrt{n_k^K}} \sum_{j \in T_k^K} d_j^{XY} - \frac{\sqrt{n_k^K}}{n_{\lceil k/2 \rceil}^{K-1}} \sum_{j \in T_{\lceil k/2 \rceil}^{K-1}} d_j^{XY}, \quad 1 + \frac{n_k^K}{n_{\lceil k/2 \rceil}^{K-1}} \right). \end{aligned}$$

Unfortunately, $\|\Pi_K R_K\|_n^2$ does not follow a chi-square distribution since the variance of M_k^K is not equal to 1. In order to get a statistic out of D^{XY} that follows a chi-square distribution, we normalize M_k^K into $\bar{M}_k^K$

$$\bar{M}_k^K := \sqrt{\frac{n_{\lceil k/2 \rceil}^{K-1}}{n_{\lceil k/2 \rceil}^{K-1} + n_k^K}} M_k^K \sim \mathcal{N}(\cdots, 1),$$

such that the normalized residual after projection having its Euclidean norm as

$$\|\Pi_K \bar{R}_K\|_n^2 := \sum_{k=1}^{2^K} (\bar{M}_k^K)^2 = \sum_{k=1}^{2^K} \left(\sqrt{\frac{n_k^K n_{\lceil k/2 \rceil}^{K-1}}{n_{\lceil k/2 \rceil}^{K-1} + n_k^K}} \left(m_k^K - m_{\lceil k/2 \rceil}^{K-1} \right) \right)^2.$$

In addition, we define

$$\|\Pi_0 \bar{R}_0\|_n^2 := \|\Pi_0 R_0\|_n^2 = \|\Pi_0 D^{XY}\|_n^2 = n(m_1^0)^2.$$

In this way, for each K, $\bar{R}_K$ is orthogonal to $\bar{R}_{K-1}$, which ensures the independence of $(\bar{R}_0, \dots, \bar{R}_{K_0})$ under Assumption 3, thanks to Cochran's theorem. Under Assumption 3, we have therefore

$$\|\Pi_0 \bar{R}_0\|_n^2 \sim \chi^2(1, \|\Pi_0 \bar{r}^{XY}\|_n^2), \qquad \|\Pi_K \bar{R}_K\|_n^2 \sim \chi^2(2^K, \|\Pi_K \bar{r}_K^{XY}\|_n^2)$$

where

- $\|u\|_n$ denotes the Euclidian norm of vector u,
- $\chi^2(\mu, \lambda)$ is the non-central chi-squared distribution with μ degrees of freedom and non-centrality parameter λ,
- $\bar{r}_K^{XY}$ denotes the K-th normalized residual after orthogonal projection of d^{XY} with additional normalization.

From now on, we denote $\bar{d}_0^{XY} := d^{XY}$. For $K = 0, \dots, K_0$, we expect ideally that

$$\mathcal{H}_0^K : \|\Pi_K \bar{d}_K^{XY}\|_n^2 \neq 0, \quad \text{versus} \quad \mathcal{H}_1^K : \|\Pi_K \bar{d}_K^{XY}\|_n^2 = 0.$$

Statistically, such hypotheses are not well separated and one has to consider

$$\mathcal{H}_0^K : \|\Pi_K \bar{d}_K^{XY}\|_n^2 > n\delta_K^2, \quad \text{versus} \quad \mathcal{H}_1^K : \|\Pi_K \bar{d}_K^{XY}\|_n^2 \leqslant n\delta_K^2. \tag{3.23}$$

where $n\delta_K^2$ is equivalence margin. It defines the discrepancy between two unobservable true TCs that we are ready to tolerate. Without prior knowledge of where or how such discrepancies would appear, we therefore choose $\delta_K = \delta$ that does not depend on K. As

$$\frac{1}{n}\|\Pi_K \bar{r}_K^{XY}\|_n^2 \xrightarrow[n\to\infty]{} \frac{1}{T}\int_0^T \left(\Pi_K \bar{r}_K^{XY}(t)\right)^2 dt,$$

it means that we are looking at any dyadic partition where the projected difference of the normalized residual shows enough energy in order to decide $\mathcal{H}_0$. In other words, two observations are considered to come from the same signal if one cannot detect a given level of energy in the residual of their difference on any considered partition. The use of multiple dyadic partitions ensures that the classical (non equivalence) version of this test is adaptive to the unknown Hölderian regularity of the difference between the signals [Baraud et al., 2003].

3.2.3 Multiple equivalence test

We will accept $\mathcal{H}_1$ if and only if all $\mathcal{H}_1^K$ are accepted, where $K \in \mathcal{K}$ and $\mathcal{K} = \{1, \ldots, n\}$ for Gaussian case, and $\mathcal{K} = \{0, \ldots, K_0\}$ for Chi-sqaure case. Following the idea of multiple test in [Baraud et al., 2005, 2003], we build a multiple *equivalence* test as

$$\mathcal{H}_0 = \bigcup_{K \in \mathcal{K}} \mathcal{H}_0^K \quad \text{versus} \quad \mathcal{H}_1 = \bigcap_{K \in \mathcal{K}} \mathcal{H}_1^K. \tag{3.24}$$

By definition in [Berger and Hsu, 1996], this multiple test is also an *intersection-union test* (IUT).

In order to define the p-value associated to the multiple equivalence test, we recall the result of Berger and Hsu:

Proposition 1 (Berger and Hsu [1996]) *If $\mathcal{R}_K$ is a rejection region at level α, the IUT with rejection region $\mathcal{R} = \bigcap_{K \in \mathcal{K}} \mathcal{R}_K$ is of level α.*

For Gaussian case, given $\Lambda \sim \mathcal{N}(0, \delta)$, the p-value associated with the K-th equivalence test is

$$p_K(X, Y) = \mathbf{P}\left(\Lambda \leqslant |D_K^{XY}|\right). \tag{3.25}$$

For Chi-square case, given $\Lambda \sim \chi^2(2^K, n\delta^2)$, the p-value associated with the K-th equivalence test is

$$p_K(X, Y) = \mathbf{P}\left(\Lambda \leqslant \|\Pi_K \bar{R}_K^{XY}\|_n^2\right), \tag{3.26}$$

For both cases, the following corollary holds:

Corollary 1 *The p-value of the IUT defined by (3.24) with rejection region $\mathcal{R} = \bigcap_{K \in \mathcal{K}} \mathcal{R}_K$ is*

$$p(X, Y) := \max_{K \in \mathcal{K}} p_K(X, Y). \tag{3.27}$$

Proof of Corollary 1: Given an observation, the p-value is the largest significance level α such that $\mathcal{H}_0$ is accepted. Considering the rejection region for the IUT in form $\mathcal{R}^\alpha = \bigcap_{K \in \mathcal{K}} \mathcal{R}_K^\alpha$, the p-value of the IUT (3.24) becomes

$$
\begin{aligned}
p(X,Y) &= \sup\{\alpha \mid D^{XY} \notin \mathcal{R}^{\alpha}\} \\
&= \sup\left\{\alpha \mid D^{XY} \notin \bigcap_{K\in\mathcal{K}} \mathcal{R}_K^{\alpha}\right\} \\
&= \sup\left\{\alpha \mid \bigcup_{K\in\mathcal{K}} \left\{D^{XY} \notin \mathcal{R}_K^{\alpha}\right\}\right\} \\
&= \sup\left\{\alpha \mid \exists K \in \mathcal{K} \text{ with } \underbrace{D^{XY} \notin \mathcal{R}_K^{\alpha}}_{p_K(X,Y)\geqslant\alpha}\right\} \\
&= \sup\left\{\alpha \mid \alpha \leqslant \max_{K\in\mathcal{K}}(p_K(X,Y))\right\} \\
&= \max_{K\in\mathcal{K}} p_K(X,Y).
\end{aligned}
$$

The closer $p(X,Y)$ is from 0, the more similar the average TCs on X and Y are. Our hierarchical clustering is based on this observation and uses the $p(X,Y)$ as dissimilarity measure.

3.3 HiSET: Clustering using equivalence test

In this section, we will describe the two-step clustering procedure and the corresponding theoretical properties, along with the parameters involved.

A general setup of the clustering problem is to produce a partition of $\mathcal{X}$ into ℓ^* subsets

$$\mathcal{X} = C_1 \cup \ldots \cup C_{\ell^*}, \quad C_s \cap C_{s'} = \emptyset, \quad 1 \leqslant s, s' \leqslant \ell^*. \tag{3.28}$$

The subsets are called clusters such that elements share similar properties in one cluster and have different properties from cluster to cluster. In the context of DCE image sequence, we expect voxels to have the "same" TC in one cluster and different TCs otherwise. Using the previously introduced equivalence test, here "same" means that two voxels have the difference of their TCs under $\mathcal{H}_1$. Moreover, we need to take into account the local property, in spatial sense, of DCE image sequence.

In order to retrieve local homogeneous clusters first, we incorporate hierarchical clustering with region growing and aggregate voxels in a bottom-up manner where only neighboring voxels with smallest dissimilarity value of time curves are considered. In this local clustering process, dissimilarities are computed as p-values only for spatial neighbors and considered as 1 otherwise. By controlling the type I error with an user-defined significance level of multiple equivalence test, we provide an automatic way to stop the local clustering that also determines both the number of local clusters and the partition. This local clustering step is followed by a global one, which aims at recovering global (disconnected) homogeneous cluster with several local homogeneous sub-clusters. To this end, starting with the previous partition resulting from local clustering, we adapt the same hierarchical clustering to region merging without consideration of the neighborhood structure while computing the dissimilarity. Therefore, the local homogeneous clusters are merged again in a bottom-up way regardless of their disconnectedness. Global clustering is also stopped by the same significance level of multiple equivalence test and the final partition with global clusters is produced.

The complexity of the local clustering is controlled by the spatial localization while the complexity of the global clustering is controlled thanks to the expected small size of the partition obtained after the local phase.

3.3.1 Local clustering

Given an image sequence on a grid $\mathcal{X}$, the initial partition

$$\mathcal{P}^0 := \{\{x\}, x \in \mathcal{X}\}$$

is made of $N = |\mathcal{X}|$ clusters: each voxel is considered as a singleton cluster. In this partition, two clusters are neighbors if and only if they are spatially connected on the 2D or 3D grid $\mathcal{X}$. For each cluster $C = \{x\}$ in the partition $\mathcal{P}^0$, we define its neighbor set $\mathcal{V}(C)$ as the union of the singleton clusters (voxels) around it in the four directions (north, south, east and west), if they are available on the grid. Supposing x has a two-dimensional coordinates (x_1, x_2) on the grid, each direction is defined as

$$n = (0, 1), \quad s = (0, -1), \quad e = (1, 0), \quad w = (-1, 0).$$

Hence,

$$\mathcal{V}(C) := \{\{x + n\}, \{x + s\}, \{x + e\}, \{x + w\}\}. \tag{3.29}$$

For each pair of clusters C_s and $C_{s'}$ in $\mathcal{P}^0$, we define indicators as

$$\mathbb{1}^0_{s,s'} = 1 \Leftrightarrow C_s \in \mathcal{V}(C_{s'}) \Leftrightarrow C_{s'} \in \mathcal{V}(C_s), \tag{3.30}$$

and their p-value corresponding to each equivalence test is given as (3.25) for Gaussian case and as (3.26) for Chi-square case. Then the dissimilarity between two clusters C_s and $C_{s'}$ is defined as

$$p(C_s, C_{s'}) := \max_{K \in \mathcal{K}} \{p_K(C_s, C_{s'})\}. \tag{3.31}$$

For each pair of clusters C_s and $C_{s'}$ in $\mathcal{P}^0$ having $\mathbb{1}^0_{s,s'} = 0$,

$$p(C_s, C_{s'}) := 1. \tag{3.32}$$

The minimum dissimilarity value of $\mathcal{P}^0$ is therefore denoted as

$$p(0) := \min_{\{(s,s')|1\leqslant s,s'\leqslant N \text{ and } \mathbb{1}^0_{s,s'}=1\}} \{p(C_s, C_{s'})\} \tag{3.33}$$
$$= \min_{\{(s,s')|1\leqslant s,s'\leqslant N \text{ and } \mathbb{1}^0_{s,s'}=1\}} \left\{\max_{K\in\mathcal{K}} \{p_K(C_s, C_{s'})\}\right\}.$$

At iteration $\bar{\ell}$, the number of clusters has been decreased by 1 for every past iteration. Hence, we have at hand a partition $\mathcal{P}^{\bar{\ell}}$ made of $\ell = N - \bar{\ell}$ clusters

$$\mathcal{P}^{\bar{\ell}} := \left\{C_1^{\bar{\ell}}, C_2^{\bar{\ell}}, \ldots, C_\ell^{\bar{\ell}}\right\}, \tag{3.34}$$

together with indicators defined as

$$\mathbb{1}^{\bar{\ell}}_{s,s'} = 1 \Leftrightarrow C_s \in \mathcal{V}(C_{s'}) \Leftrightarrow C_{s'} \in \mathcal{V}(C_s). \tag{3.35}$$

Then the new partition $\mathcal{P}^{\bar{\ell}+1}$ is obtained by merging the two clusters indexed by s_1 and s_2 having the minimum dissimilarity in $\mathcal{P}^{\bar{\ell}}$,

$$(s_1, s_2) = \arg \min_{\{(s,s')|1\leqslant s,s'\leqslant \ell \text{ and } \mathbb{1}^{\bar{\ell}}_{s,s'}=1\}} p\left(C_s^{\bar{\ell}}, C_{s'}^{\bar{\ell}}\right). \tag{3.36}$$

Because the minimum may be achieved on more than one couple (s, s'), one needs to define a strategy to pick one couple, and we use the smallest lexical order. Once two clusters C_s and $C_{s'}$ satisfying (3.36) are merged into a new cluster C, the partition is updated accordingly to

$$\mathcal{P}^{\bar{\ell}+1} = \mathcal{P}^{\bar{\ell}} \setminus \{C_{s_1}, C_{s_2}\} \cup \{C\}, \tag{3.37}$$

and the neighbor set of C is then defined as

$$\mathcal{V}(C) = (\mathcal{V}(C_s) \cup \mathcal{V}(C_{s'})) \setminus \{C\}. \tag{3.38}$$

Moreover, for every C' in $\mathcal{P}^{\bar{\ell}+1}$, if C_s or $C_{s'}$ is a neighbor of C', then $\mathcal{V}(C')$ is updated into

$$\mathcal{V}(C') \setminus \{C_s, C_{s'}\} \cup \{C\}, \tag{3.39}$$

meaning that C_s or $C_{s'}$ appearing in the neighbor set of C' is replaced by the new cluster C made of their union. In addition, we denote

$$p(\bar{\ell}) := \min_{\{(s,s')|1\leqslant s,s'\leqslant \ell \text{ and } \mathbb{1}^{\bar{\ell}}_{s,s'}=1\}} p\left(C_s^{\bar{\ell}}, C_{s'}^{\bar{\ell}}\right) \tag{3.40}$$

the so-called *minimum dissimilarity function* at iteration $\bar{\ell}$. Furthermore, the dissimilarity values concerning this emerging cluster C and its neighbor clusters in $\mathcal{V}(C)$ are computed as (3.31).

In local clustering, we start from a partition made of all voxels as singletons and then build successive partitions by merging two clusters having minimum dissimilarity value at each iteration. During the iterations, a (bottom-up) hierarchy of partitions with decreasing sizes from N to 1 is provided. Through thresholding the type I error with the significant level of multiple equivalence test, which will be introduced in Section 3.3.3, the local clustering phase is stopped at certain iteration $\bar{\ell}^{loc}$ and the number of local clusters ℓ^{loc} is automatically determined. It provides a preliminary partition made of only spatially connected clusters.

3.3.2 Global clustering

The global step takes the output of the local step as input and uses the same iteration step as the local one. Following the past iterations of local clustering phase, global clustering starts with the partition resulting from $\bar{\ell}^{loc}$-th iteration and continues the same iteration as in local clustering except that the neighbor set $\mathcal{V}(C)$ of each cluster C is changed.

The reason that we need to change the neighbor set is because we want to retrieve the global homogeneous clusters which are split into several disconnected local homogeneous clusters. Due to the way in which we define the neighbor set in local clustering, disconnected clusters are forbidden to merge together. After local clustering stops, we can not merge the neighbors anymore. Moving to a global scale of clustering, in order to still keep a relatively low complexity, we can consider the new neighbor set of a cluster C as the neighbors of its neighbors, meaning that

$$\mathcal{V}'(C) := \bigcup_{C' \in \mathcal{V}(C)} \{\mathcal{V}(C')\} \setminus \{C\}. \tag{3.41}$$

Or, since the number of local clusters is expected to be small enough, we can discard the complexity and consider all other clusters except C as the neighbors of

C,

$$\mathcal{V}'(C) := \mathcal{P} \setminus \{C\}. \tag{3.42}$$

We start from a partition made of the ℓ^{loc} local clusters and then continue building successive partitions by merging two clusters having minimum dissimilarity value at each iteration. During the iterations, the same type of partition hierarchy is raised while partition size continues decreasing from ℓ^{loc} to 1. With the same threshold of significance level, the global clustering is stopped at iteration $\bar{\ell}^*$ and the number of global clusters ℓ^* is automatically determined.

3.3.3 Automatic selection of number of clusters

In order to automatically stop the iterations and to select a final partition, we introduce a so-called "*control function*", denoted by $c_\alpha(\ell)$, such that the iterations stop as soon as $p(\bar{\ell}) > c_\alpha(\ell)$. This control function is a direct result from the control of the probability that the algorithm stops too early. Prior to the theoretical details regarding the probability control, we first introduce a couple of definitions.

Definition 1

1. *Two subsets X and Y of $\mathcal{X}$ are called "δ-separated" if their unobservable true TCs satisfy $\mathcal{H}_0^K$ for at least one value $K \in \mathcal{K}$.*

2. *$\mathcal{X}$ is a "δ-partition of size ℓ" if there exists a partition of $\mathcal{X}$ into ℓ δ-separated subsets, $C_1, \ldots, C_\ell$.*

Theorem 1 - Stopping too late - *Under Assumption 3, if $\mathcal{X}$ in a δ-partition of size ℓ_0, the probability that $\ell^* < \ell_0$ is lower than $\alpha > 0$ as soon as*

$$c_\alpha(\ell) = \left(\frac{2\alpha}{\ell(\ell-1)}\right)^{\frac{1}{M}}, \quad \text{for every } 1 < \ell \leqslant |\mathcal{X}|, \tag{3.43}$$

where $M = |\mathcal{K}|$ is the number of tests in the multiple equivalence test.

Proof of Theorem 1: We want to control $\mathbf{P}_{\mathcal{H}_0}(p(\bar{\ell}) \leqslant c_\alpha(\ell))$ the probability of a false merge at iteration $\bar{\ell}$ with clusters $C_1, \ldots, C_\ell$. This is the probability that

given $1 \leqslant i < j \leqslant \ell$ there exists $K \in \mathcal{K}$ for two clusters C_i and C_j such that they are δ-separated:

$$\begin{aligned}
\mathbf{P}_{\mathcal{H}_0}(p(\bar{\ell}) \leqslant c_\alpha(\ell)) &= \mathbf{P}_{\mathcal{H}_0}\left[\min_{1\leqslant i<j\leqslant \ell}\{p(C_i, C_j)\} \leqslant c_\alpha(\ell)\right] \\
&= \mathbf{P}_{\mathcal{H}_0}\left[\bigcup_{1\leqslant i<j\leqslant \ell}\{p(C_i, C_j) \leqslant c_\alpha(\ell)\}\right] \\
&= \mathbf{P}_{\mathcal{H}_0}\left[\bigcup_{1\leqslant i<j\leqslant \ell}\left\{\max_{K\in\mathcal{K}}\{p_K(C_i, C_j)\} \leqslant c_\alpha(\ell)\right\}\right] \\
&\leqslant \sum_{1\leqslant i<j\leqslant \ell} \mathbf{P}_{\mathcal{H}_0}\left[\max_{K\in\mathcal{K}}\{p_K(C_i, C_j)\} \leqslant c_\alpha(\ell)\right] \\
&= \sum_{1\leqslant i<j\leqslant \ell} \mathbf{P}_{\mathcal{H}_0}\left[\bigcap_{K\in\mathcal{K}}\{p_K(C_i, C_j) \leqslant c_\alpha(\ell)\}\right]. \qquad (3.44)
\end{aligned}$$

As $D_K^{C_iC_j}$ and $\Pi_K \bar{R}_K^{C_iC_j}$ both are independent with respect to K, by assumption for the former and by construction for the latter (as orthogonal projections of a Gaussian vector, thanks to Cochran's theorem), $p_K(C_i, C_j)$ are also independent with respect to K. Moreover they have the same uniform distribution as p-values coming from absolutely continuous distributions. Hence,

$$\begin{aligned}
\mathbf{P}_{\mathcal{H}_0}[p(\bar{\ell}) \leqslant c_\alpha(\ell)] &\leqslant \frac{\ell(\ell-1)}{2} \prod_{K\in\mathcal{K}} \mathbf{P}_{\mathcal{H}_0}\left[p_K(C_1, C_2) \leqslant c_\alpha(\ell)\right] \\
&= \frac{\ell(\ell-1)}{2} \left(c_\alpha(\ell)\right)^M. \qquad (3.45)
\end{aligned}$$

Controlling the probability of false merge by α leads to

$$\frac{\ell(\ell-1)}{2} \left(c_\alpha(\ell)\right)^M = \alpha,$$

and we deduce thereby

$$c_\alpha(\ell) = \left(\frac{2\alpha}{\ell(\ell-1)}\right)^{\frac{1}{M}}.$$

□

Starting from $\mathcal{P}^0$ with $p(0) \leqslant c_\alpha(N)$, the local clustering recursively merges two clusters at each iteration until $p(\bar{\ell})$ exceeds $c_\alpha(\ell)$, resulting in a partition $\mathcal{P}^{loc}$ made of ℓ^{loc} clusters. Then the global clustering starts with a lower value of the minimum dissimilarity at iteration $N - \ell^{loc}$, thanks to the switch of neighborhood structure

that involves much more dissimilarities. Again the global clustering recursively merges two clusters at each iteration until $p(\bar{\ell})$ exceeds $c_\alpha(\ell)$ again, resulting in a final partition $\mathcal{P}^*$ and producing a final number of clusters ℓ^*.

Having a stop criterion as the control function, we now need to control the probability of stopping too late. This is obtained in the following theorem.

Theorem 2 - Stopping too early - *Under Assumption 3, if $\mathcal{X}$ in a δ-partition of size ℓ_0, the probability that $\ell^* > \ell_0$ is lower than β_n.*

- *For Gaussian case: When n is small, β_n goes to 0 faster than $|\mathcal{X}|^3\eta_\delta(n)/2$, where $\eta_\delta(n) := n\left(1 - \Psi_0\left(\delta/2\right) + \Psi_\delta\left(\delta/2\right)\right)^n$.*
- *For Chi-square case: When n goes to infinity, β_n goes to 0 faster than $|\mathcal{X}|^3 n^{-\eta_\delta(n)+1}\log_2 n$, where $\eta_\delta(n) := \min\left(\frac{3}{2\log 2}n\delta^2, \frac{1}{\sqrt{2}}n\delta\right)$.*

The proof of Theorem 2 is a direct consequence of the following lemma that controls how dissimilarities are organized between subsets δ-separated or sharing same unobservable true TCs.

Lemma 1 *For Gaussian case, the probability, that two δ-separated clusters merge before two subsets of one cluster do, goes to 0 faster than $|\mathcal{X}|^2\eta_\delta(n)/2$. For Chi-square case, when n goes to infinity, the probability, that two δ-separated clusters merge before two subsets of one cluster do, goes to 0 faster than $|\mathcal{X}|^2 n^{-\eta_\delta(n)+1}\log_2 n$.*

Proof of Lemma 1: Assuming that we have at hand $C_1, ..., C_\ell$ δ-separated and at least one supplementary set C_0 such that there exists $1 \leqslant \ell' \leqslant \ell$ with $i^{C_0} = i^{C_{\ell'}}$, we want to compute a lower bound of the probability to have

$$p(C_0, C_{\ell'}) < \min_{\substack{0 \leqslant j < j' \leqslant \ell \\ (j,j') \neq (0,\ell')}} p(C_j, C_{j'}). \tag{3.46}$$

Hence if a merge occurs, it will happen with high probability between two sets having same true TC instead of sets δ-separated.

Let us denote by $\chi^2_{2^K, n\delta^2}(.)$ the cumulative distribution function (CDF) of a $\chi^2(2^K, n\delta^2)$ random variable. We recall that the dissimilarity $p_K(C_i, C_j) =$

$\chi^2_{2^K,n\delta^2}(\|\Pi_K \bar{R}_K^{C_iC_j}\|_n^2)$ and that they are independent w.r.t K thanks to orthogonal projections and Cochran's theorem. For sake of simplicity, we assume that $\ell' = 1$. The probability that (3.46) occurs is controlled as following

$$\begin{aligned}
&\mathbf{P}\left[\max_{K\in\mathcal{K}} p_K(C_0,C_1) < \min_{\substack{0\leqslant j<j'\leqslant \ell \\ (j,j')\neq(0,1)}} \max_{K'\in\mathcal{K}} p_{K'}(C_i,C_j)\right] \qquad (3.47)\\
&= \prod_{K\in\mathcal{K}} \mathbf{P}\left[p_K(C_0,C_1) < \min_{\substack{0\leqslant j<j'\leqslant \ell \\ (j,j')\neq(0,1)}} \max_{K'\in\mathcal{K}} p_{K'}(C_i,C_j)\right]\\
&= \prod_{K\in\mathcal{K}} \left\{1 - \mathbf{P}\left[p_K(C_0,C_1) \geqslant \min_{\substack{0\leqslant j<j'\leqslant \ell \\ (j,j')\neq(0,1)}} \max_{K'\in\mathcal{K}} p_{K'}(C_i,C_j)\right]\right\}\\
&= \prod_{K\in\mathcal{K}} \left\{1 - \mathbf{P}\left[\bigcup_{\substack{0\leqslant j<j'\leqslant \ell \\ (j,j')\neq(0,1)}} \left\{p_K(C_0,C_1) \geqslant \max_{K'\in\mathcal{K}} p_{K'}(C_i,C_j)\right\}\right]\right\}\\
&\geqslant \prod_{K\in\mathcal{K}} \left\{1 - \left(\frac{\ell(\ell-1)}{2} + \ell - 1\right)\mathbf{P}\left[p_K(C_0,C_1) \geqslant \max_{K'\in\mathcal{K}} p_{K'}(C_i,C_j)\right]\right\}\\
&= \prod_{K\in\mathcal{K}} \left\{1 - \left(\frac{\ell}{2}+1\right)(\ell-1)\prod_{K'\in\mathcal{K}} \mathbf{P}\left[p_K(C_0,C_1) \geqslant p_{K'}(C_i,C_j)\right]\right\} \qquad (3.48)
\end{aligned}$$

For Gaussian case, (3.48) equals

$$\prod_{K\in\mathcal{K}} \left\{1 - \left(\frac{\ell}{2}+1\right)(\ell-1)\prod_{K'\in\mathcal{K}} \mathbf{P}(\underbrace{|D_K^{C_0C_1}|}_{\sim\mathcal{N}(0,1)} \geqslant \underbrace{|D_{K'}^{C_iC_j}|}_{\sim\mathcal{N}(\delta,1)})\right\}.$$

Given $X \sim \mathcal{N}(0,1)$ and $Y \sim \mathcal{N}(\delta,1)$,

$$\begin{aligned}
\mathbf{P}(X \geqslant Y) &= \min_{y>0} \left[\mathbf{P}(X \geqslant Y, Y \geqslant y) + \mathbf{P}(X \geqslant Y, Y < y)\right]\\
&\leqslant \min_{y>0} \left[\mathbf{P}(X \geqslant y) + \mathbf{P}(Y < y)\right]\\
&= 1 - \Psi_0(\delta/2) + \Psi_\delta(\delta/2), \qquad (3.49)
\end{aligned}$$

where $\Psi_\mu(.)$ is the CDF of normal distribution with mean equal to μ and variance equal to 1. Therefore, as $\ell \leqslant |\mathcal{X}|$,

$$\begin{aligned}
&\mathbf{P}\left[\max_{K\in\mathcal{K}} p_K(C_0,C_1) < \min_{\substack{0\leqslant j<j'\leqslant \ell \\ (j,j')\neq(0,1)}} \max_{K'\in\mathcal{K}} p_{K'}(C_i,C_j)\right] \\
&\geqslant \prod_{K\in\mathcal{K}}\left\{1-\left(\frac{\ell}{2}+1\right)(\ell-1)\left(1-\Psi_0\left(\delta/2\right)+\Psi_\delta\left(\delta/2\right)\right)^n\right\} \\
&\geqslant \prod_{K\in\mathcal{K}}\left\{1-\frac{|\mathcal{X}|^2}{2}\left(1-\Psi_0\left(\delta/2\right)+\Psi_\delta\left(\delta/2\right)\right)^n\right\} \\
&= \left(1-\frac{|\mathcal{X}|^2}{2}\left(1-\Psi_0\left(\delta/2\right)+\Psi_\delta\left(\delta/2\right)\right)^n\right)^n \\
&\geqslant 1-\frac{n}{2}|\mathcal{X}|^2\left(1-\Psi_0\left(\delta/2\right)+\Psi_\delta\left(\delta/2\right)\right)^n.
\end{aligned}$$

For Chi-square case, (3.48) equals

$$\prod_{K\in\mathcal{K}}\left\{1-\left(\frac{\ell}{2}+1\right)(\ell-1)\prod_{K'\in\mathcal{K}}\mathbf{P}(\underbrace{\|\Pi_K\bar{R}_K^{C_0C_1}\|_n^2}_{\sim\chi^2(2^K)} \geqslant \underbrace{\|\Pi_{K'}\bar{R}_{K'}^{C_iC_j}\|_n^2}_{\sim\chi^2(2^{K'},n\delta^2)})\right\}.$$

We then consider the following concentration inequalities:

Proposition 2 (Massart [2003]) *Let D be a non-central χ^2 variable with μ degrees of freedom and non-centrality parameter $\lambda \geqslant 0$, then for all $x > 0$*

$$\mathbf{P}[D \geqslant (\mu+\lambda^2)+2\sqrt{(\mu+2\lambda^2)x}+2x] \leqslant e^{-x}, \tag{3.50}$$

$$\mathbf{P}[D \leqslant (\mu+\lambda^2)-2\sqrt{(\mu+2\lambda^2)x}] \leqslant e^{-x}. \tag{3.51}$$

Given $X\sim\chi^2(J)$ and $Y\sim\chi^2(L,n\delta^2)$,

$$\begin{aligned}
\mathbf{P}(X\geqslant Y) &= \min_{y>0}\left[\mathbf{P}(X\geqslant Y, Y\geqslant y)+\mathbf{P}(X\geqslant Y, Y<y)\right] \\
&\leqslant \min_{y>0}\left[\mathbf{P}(X\geqslant y)+\mathbf{P}(Y<y)\right] \\
&= \mathbf{P}(X\geqslant \underbrace{J+2\sqrt{Jx_1}+2x_1}_{=y}) + \mathbf{P}(Y<\underbrace{L+n\delta^2-2\sqrt{(L+2n\delta^2)x_2}}_{=y}) \\
&\leqslant e^{-x_1}+e^{-x_2},
\end{aligned} \tag{3.52}$$

with x_1 and x_2 satisfying

$$J + 2\sqrt{Jx_1} + 2x_1 = L + n\delta^2 - 2\sqrt{(L + 2n\delta^2)x_2} \Leftrightarrow x_2 = \frac{\left(L - J + n\delta^2 - 2\sqrt{Jx_1} - 2x_1\right)^2}{4(L + 2n\delta^2)}.$$

Hence, right side term of (3.52) is minimized to $2e^{-x_1}$ if

$$x_1 = \frac{\left(L - J + n\delta^2 - 2\sqrt{Jx_1} - 2x_1\right)^2}{4(L + 2n\delta^2)}.$$

Solving this equation and choosing the largest root provide that

$$\begin{aligned} x_1 &= \left(L + \frac{3n\delta^2}{2}\right) + \sqrt{J(L + 2n\delta^2)} \\ &\quad + \frac{1}{2}\left(3L^2 - J^2 + 6JL + 10(J + L)n\delta^2 + 8n^2\delta^4 + \frac{4\sqrt{J}}{\sqrt{L + 2n\delta^2}}(2L^2 + 7Ln\delta^2 + 6n^2\delta^4)\right)^{1/2}. \end{aligned}$$

As $\mathbf{P}(X \geqslant Y)$ is the largest when J is maximum ($J = 2^{K_0} \geqslant n/4$) and L is minimum ($L = 2^0 = 1$), we have that $\mathbf{P}(X \geqslant Y) \leqslant 2e^{-x}$ with

$$\begin{aligned} x &= \left(1 + \frac{3n\delta^2}{2}\right) + \sqrt{2^{K_0}(1 + 2n\delta^2)} \\ &\quad + \frac{1}{2}\left(3 - 2^{2K_0} + 6 \cdot 2^{K_0} + 10(2^{K_0} + 1)n\delta^2 + 8n^2\delta^4 + \frac{2^{K_0/2+2}}{\sqrt{1 + 2n\delta^2}}(2 + 7n\delta^2 + 6n^2\delta^4)\right)^{1/2} \\ &\geqslant \min\left(\frac{3}{2}n\delta^2, \frac{1}{\sqrt{2}}n\delta\right). \end{aligned} \tag{3.53}$$

Therefore, as $\ell \leqslant |\mathcal{X}|$,

$$\begin{aligned} &\mathbf{P}\left[\max_{K\in\mathcal{K}} p_K(C_0, C_1) < \min_{\substack{0\leqslant j<j'\leqslant \ell \\ (j,j')\neq(0,1)}} \max_{K'\in\mathcal{K}} p_{K'}(C_i, C_j)\right] \\ &\geqslant \prod_{K\in\mathcal{K}}\left\{1 - 2^{K_0+1}\left(\frac{\ell}{2} + 1\right)(\ell - 1)e^{-x(K_0+1)}\right\} \\ &\geqslant \prod_{K\in\mathcal{K}}\left\{1 - 2^{K_0}|\mathcal{X}|^2 e^{-x(K_0+1)}\right\} \\ &= \left(1 - 2^{K_0}|\mathcal{X}|^2 e^{-x(K_0+1)}\right)^{K_0+1} \\ &\geqslant \left(1 - \frac{n}{2}|\mathcal{X}|^2 e^{-x\log_2 n}\right)^{\log_2 n} \\ &\geqslant 1 - \frac{n\log_2 n}{2}|\mathcal{X}|^2 e^{-x\log_2 n} \\ &= 1 - |\mathcal{X}|^2\frac{\log_2 n}{2}n^{1-x/\ln 2} \\ &\geqslant 1 - |\mathcal{X}|^2\frac{\log_2 n}{2}n^{1-\min(\frac{3}{2}n\delta^2, \frac{1}{\sqrt{2}}n\delta)/\ln 2}. \end{aligned}$$

The last inequality is deduced by (3.53). As $|\mathcal{X}|$ does not depend on n, the right-hand term goes to 1 at a rate faster than $|\mathcal{X}|^2 n^{-\eta_\delta(n)+1} \log_2 n$. □

Proof of Theorem 2: We denote the final δ-separated clusters as $C_1, C_2, \ldots, C_{\ell^*}$. Let us assume that, at iteration $\bar{\ell} = |\mathcal{X}| - \ell^* - 1$, we have clusters $C_0^{\bar{\ell}}, C_1^{\bar{\ell}}, C_2^{\bar{\ell}}, \ldots, C_{\ell^*}^{\bar{\ell}}$ such that the unobservable true TCs on $C_0^{\bar{\ell}}$ and on $C_1^{\bar{\ell}}$ are equal while $C_1^{\bar{\ell}}, C_2^{\bar{\ell}}, \ldots, C_{\ell^*}^{\bar{\ell}}$ are δ-separated. In such context, $C_0^{\bar{\ell}}$ and $C_1^{\bar{\ell}}$ should be merged into one to get the final partition into ℓ^* δ-separated clusters $C_1, C_2, \ldots, C_{\ell^*}$, where $C_1 = C_0^{\bar{\ell}} \cup C_1^{\bar{\ell}}$ and $C_i = C_i^{\bar{\ell}}$ for $i = 2, \ldots, \ell^*$. According to the definition of control function in (3.43), we should not have

$$\min_{0 \leqslant i,j \leqslant \ell^*} p(C_i^{\bar{\ell}}, C_j^{\bar{\ell}}) \geqslant c_\alpha(\ell^* + 1)$$

to ensure that the iterations do not stop.

Let us consider ℓ, $1 < \ell \leqslant \ell^*$, such that $p(C_0^{\bar{\ell}}, C_\ell^{\bar{\ell}})$ is the largest. Then we introduce a *phantom* cluster $C_{\ell^*+1}^{\bar{\ell}}$, δ-separated from $C_1^{\bar{\ell}}, C_2^{\bar{\ell}}, \ldots, C_{\ell^*}^{\bar{\ell}}$, such that $p(C_\ell^{\bar{\ell}}, C_{\ell^*+1}^{\bar{\ell}}) = c_\alpha(\ell^* + 1)$ together with $p(C_0^{\bar{\ell}}, C_{\ell^*+1}^{\bar{\ell}}) \geqslant p(C_0^{\bar{\ell}}, C_\ell^{\bar{\ell}})$. This construction is always possible by adding one voxel to $\mathcal{X}$ with a TC satisfying both conditions. It ensures that $C_{\ell^*+1}^{\bar{\ell}}$ cannot merge with $C_0^{\bar{\ell}}$ or $C_1^{\bar{\ell}}$ and has no effect in the coming iteration. Thereby, according to Lemma 1,

$$p(C_0^{\bar{\ell}}, C_1^{\bar{\ell}}) < \min_{1 \leqslant i,j \leqslant \ell^*+1} p(C_i^{\bar{\ell}}, C_j^{\bar{\ell}}) = c_\alpha(\ell^* + 1)$$

with probability larger than $1 - |\mathcal{X}|^2 \eta_\delta(n)/2$ for Gaussian case and $1 - |\mathcal{X}|^2 n^{-\eta_\delta(n)+1} \log_2 n$ for Chi-square case.

Similar construction can be repeated for previous iterations. First, we consider the two subsets C_0, C_1 of the current partitions having same unobservable true TC and the minimal dissimilarity (such that they should merge first). Second, for other subsets having same unobservable true TC, we keep the original TC of one of these subsets and change the TCs of the other subsets to be δ-separated. Using again a phantom cluster having its dissimilarity with C_1 equals to the control function at the given iteration, $p(C_0^{\bar{\ell}}, C_1^{\bar{\ell}})$ is smaller than the dissimilarity in question with probability larger than $1 - |\mathcal{X}|^2 \eta_\delta(n)/2$ for Gaussian case and $1 - |\mathcal{X}|^2 n^{-\eta_\delta(n)+1} \log_2 n$ for Chi-square case.

Finally, as $\ell \leqslant |\mathcal{X}|$, a Bonferroni control ensures that all subsets with similar unobservable true TCs merge before the minimum dissimilarity reaches the control function with probability larger than

$$1 - |\mathcal{X}|^3 \eta_\delta(n)/2 \qquad \text{for Gaussian case,} \tag{3.54}$$

$$1 - |\mathcal{X}|^3 n^{-\eta_\delta(n)+1} \log_2 n \qquad \text{for Chi-square case,} \tag{3.55}$$

which ends the proof of Theorem 2. □

A direct consequence of Theorems 1 and 2 together with Lemma 1 is the following result.

Corollary 2 *Under Assumption 3, if $\mathcal{X}$ is a δ-partition, it will be exactly recovered with probability larger than $1 - \alpha - \beta_n$ with β_n defined in Theorem 2.*

To give an understanding of how negligible β_n is, assuming that $\mathcal{X}$ is a 2D-image made of 1000 by 1000 voxels or equivalently a 3D-volume of 25 slices made of 200 by 200 voxels for each slice, then for $n = 5$ (in Gaussian case), $\beta_n \leqslant 10^{-9}$ as soon as $\delta \geqslant 8$; for $n = 15$ (in Chi-square case), $\beta_n \leqslant 10^{-19}$ as soon as $\delta \geqslant 1$. To be noticed, δ is defined in different way for two cases.

These two theorems are not providing an understanding of the benefit to use two steps in our algorithm. However, one can easily understand that the local step does not suffer from the combinatorial complexity of the global step, thanks to the neighborhood structure. Moreover, by aggregating only neighbors, the local step also offers the opportunity to take into account a possible regularity existing over the domain of $\mathcal{X}$ between TCs.

3.3.4 Adaptation to hierarchical clustering

Both local and global clustering behave like a bottom-up hierarchical clustering, which starts from a finest partition and merges one pair of clusters while moving up the hierarchy step by step. There are two major aspects in hierarchical clustering [Ackerman and Ben-David, 2011]: metric and linkage function. Metric is a measure of dissimilarity between pairs of voxels. Linkage function defines the dissimilarity

of sets of voxels based on the dissimilarities of all pairs of voxels. Therefore, in hierarchical clustering, one needs to compute the dissimilarity of every pair of elements at the beginning to get a dissimilarity matrix. At each iteration of merging, the two clusters C_1 and C_2 with minimum dissimilarity among all are merged together and the dissimilarity between the emerging cluster $C = C_1 \cup C_2$ and every other cluster C' are compute with linkage function using the same dissimilarity matrix. Above all, linkage function guarantees that the dissimilarity between C and C' is larger than the ones between C_1 (or C_2) and C', leading to an increasing minimum dissimilarity of iteration while moving up the hierarchy.

Comparing hierarchical clustering with our local clustering, the main difference is that we do not need compute all the dissimilarities between every pair of voxels at the beginning since we only consider the neighbors. As a consequence, we do not have the linkage function that is based on the full dissimilarity matrix from the beginning. For both voxels and clusters, we use the same multiple equivalence test to compute the dissimilarity as the p-value of the test. Moreover, at each iteration, we add some new dissimilarities concerning the emerging cluster and its neighbors. Therefore, we cannot guarantee that the minimum dissimilarity of iteration $p(\bar{\ell})$ defined in (3.40) is increasing with respect to $\bar{\ell}$, as with the linkage function in hierarchical clustering.

However, this may be fixed by considering the corrected dissimilarity $\bar{p}(.,.)$ defined as following. Assuming C_1 and C_2 have been merged into C, then for C' in $\mathcal{V}(C)$,

$$\bar{p}(C, C') := \begin{cases} \max\{\bar{p}(C_1, C'), p(C, C')\} & \text{if } C' \in \mathcal{V}(C_1) \setminus \mathcal{V}(C_2); \\ \max\{\bar{p}(C_2, C'), p(C, C')\} & \text{if } C' \in \mathcal{V}(C_2) \setminus \mathcal{V}(C_1); \\ \max\left\{\min\{\bar{p}(C_1, C'), \bar{p}(C_2, C')\}, p(C, C')\right\} & \text{otherwise.} \end{cases} \tag{3.56}$$

Consequently, the minimum dissimilarity function (with correction) becomes

$$\bar{p}(\bar{\ell}) := \min_{\{(s,s') | 1 \leqslant s, s' \leqslant \ell \text{ and } \mathbb{1}^{\bar{\ell}}_{s,s'} = 1\}} \bar{p}\left(C_s^{\bar{\ell}}, C_{s'}^{\bar{\ell}}\right). \tag{3.57}$$

These corrections only occur when the minimum dissimilarity decreases after merging, thereby we use the same control function defined in (3.43). The selection of number of clusters for both local and global clustering with corrected minimum dissimilarity function is illustrated in Figure 3.2.

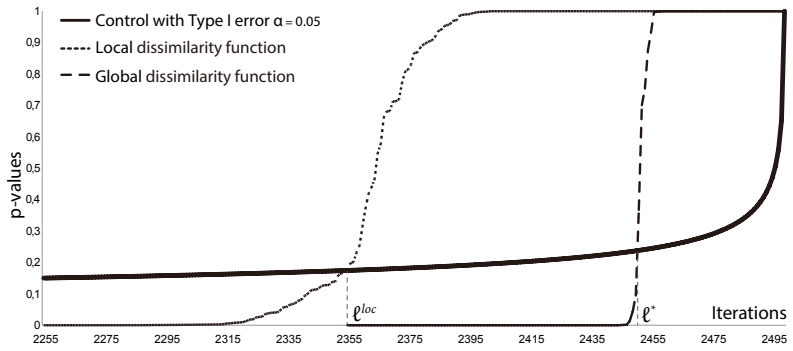


Figure 3.2: Selection of number of clusters for both local and global clustering: The solid curve is the control function defined in (3.43). The dotted (resp. dashed) curve is the minimum dissimilarity function (with correction) defined in (3.57) for the local (resp. global) clustering step. When the dotted curve reaches the solid one, ℓ^{loc} is defined and the local clustering stops. The global clustering starts from ℓ^{loc} with the dashed curve. When the dashed curve reaches the solid one, ℓ^* is defined and the algorithm stops, providing the final partition. Both local and global minimum dissimilarity functions are shown even after they reach the control function to illustrate their typical behavior.

3.3.5 Parameter interpretation

So far, we have introduced three parameters: 1/ the "model" factor a in the variance stabilization transformation; 2/ the homogeneity discrepancy δ defining the equivalence margin and 3/ the significance level α of the multiple equivalence test. Parameter a plays a role before the clustering starts, while δ and α are used during the clustering and respectively involved in the definition of the minimum dissimilarity function (3.57) and the control function (3.43). To be more specific,

parameter α controls whether p-values are considered small or not, regardless how the data are normalized. In this spirit, unlike the thresholds used in regions-growing and -merging methods [Tilton et al., 2012; Tremeau and Borel, 1997], it appears as a meta-parameter in the sense that its choice is not data-driven. Homogeneity discrepancy is ensured as parameter δ controls whether the difference of two unobservable true TCs is close to 0 or not, regardless any modelling of these signals. In this spirit, it is a human interpretable parameter as it is meaningful on noiseless signals.

3.3.6 Algorithm

The algorithm consists of two steps: local clustering and global clustering. After distinct definitions of the initial partition $\mathcal{P}$ and of the neighborhood structure $\mathcal{N}$, two steps share the same main loop to iteratively merge clusters and the same control procedure to stop and to select the number of clusters.

Input $\mathcal{P}$, α and δ.

Initialization $\bar{p} = 0$, $\ell = |\mathcal{P}|$, $\mathcal{N} := \{\mathcal{V}(C), C \in \mathcal{P}\}$ and $p(X, Y)$ for $\forall X, Y \in \mathcal{P}$.

Iterations

while $\bar{p}(\bar{\ell}) < c_\alpha(\ell)$ **do**

Find (s_1, s_2) satisfying (3.36);

New cluster: $C \leftarrow C_{s_1} \cup C_{s_2}$;

Update partition: $\mathcal{P}^{\bar{\ell}+1} \leftarrow \mathcal{P}^{\bar{\ell}} \setminus \{C_{s_1}, C_{s_2}\} \cup \{C\}$;

Update $\mathcal{N}$ as (3.38) and (3.39);

Compute new dissimilarities as in (3.56) if $C' \in \mathcal{V}(C)$, 1 otherwise;

$\bar{\ell} \leftarrow \bar{\ell} + 1$;

Update $\bar{p}(\bar{\ell})$ as in (3.57).

end while

Number of clusters: $\ell^* \leftarrow N - \bar{\ell}$.

Return $\mathcal{P}^{\bar{\ell}}$

3.3.7 Generalization

HiSET is not only for DCE image sequences, but also can be applied on any finite metric space, considered as a landscape, which functional features are discretely observed on and can be modeled by (3.10) as soon as Assumption 3 is fulfilled. By only adapting the multiple equivalence test, HiSET may be applied in various types of models where a signal is observed on a landscape such as multimodality images, vector-valued graphs and consumption (electricity, water, etc.) using geographical information or user types.

3.4 Illustration of DCE-HiSET

In this section, the procedure of HiSET is described in a graphical way. Given a very small image sequence with only 5 by 5 voxel grid for each image, we build a 25 by 25 matrix reserved for dissimilarities between voxels. However, not the dissimilarity of every pair of voxels has to be computed since we only consider the neighbors in local clustering step. Therefore, we only need to compute less than 100 (25x4) dissimilarities and put them at the corresponding positions in the matrix. Among all dissimilarities available in the matrix, we choose the minimal one and then merge the corresponding pair of voxels together to have a cluster, for instance, voxel 13 and 18 in Figure 3.3. As for the matrix, we remove the (13th and 18th) rows and columns related to the two voxels. For the emerging cluster (C), we locate its neighbors (8, 12, 14, 17, 19 and 23) and compute the dissimilarities between them and C. In the matrix, we add a new column and a new row for C and put the new dissimilarities at the right positions (black dots). This is the first iteration that is followed by others in the same way.

We simulate a *Chessboard* image sequence (see Figure 3.4) with 2 clusters, 100 times and Gaussian noise of level $\sigma = 1$. The true TCs are simply defined as the straight lines: from 1 to -1 for the black cluster and from -1 to 1 for the red one.

First, we run local clustering with $\delta = 1$. The reason why we use a such equivalence margin δ is that the largest difference between two true TCs is only 2 considering the noise level is already 1. Controlling the Type I error $\alpha < 0.001$ (see Figure 3.5-a), we stop local clustering with 100 local clusters exactly as expected (see Figure 3.5-b,c). Then we run global clustering from these local clusters and control the type I error in the same way. At the end, only 2 global clusters are left (see Figure 3.5-e,f).

From this example, HiSET shows the ability to handle the variation in time dimension coming from the noise and the homogeneity in spatial dimensions with both local and global scales. However, there is a great chance that this perfect result may only come from the significant separation between two clusters at the

starting and end times. In order to reduce this possibility and explore the limitation of this method, we remove the most separated parts of both clusters and use only the images between $t = 25$ and $t = 75$.

From the observed TCs in Figure 3.6-a, we can tell two clusters are totally mixed up with respect to noise level $\sigma = 1$. After the local clustering (Figure 3.6-b) using same parameter setting as before, we find fewer local clusters due to the mistakes made at the border of clusters. These mistakes lead to the connections between two or more local clusters from same global cluster. At the end of global clustering (Figure 3.6-c), we manage to almost recover one cluster and split the other one into two clusters. In this much more difficult situation, the underlying information in image sequence is heavily disrupted by lower SNR since the largest difference between two true time curves is only 1, which also equals the noise level.

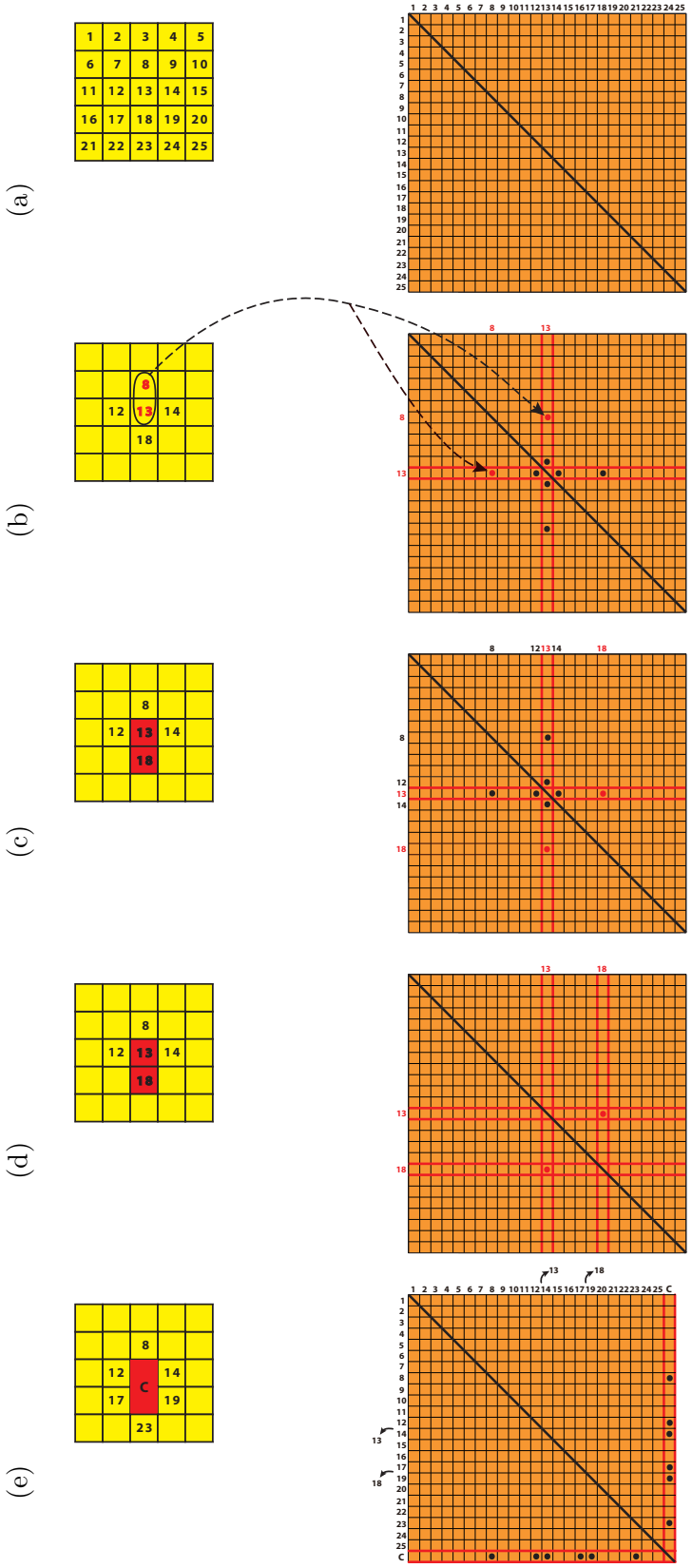
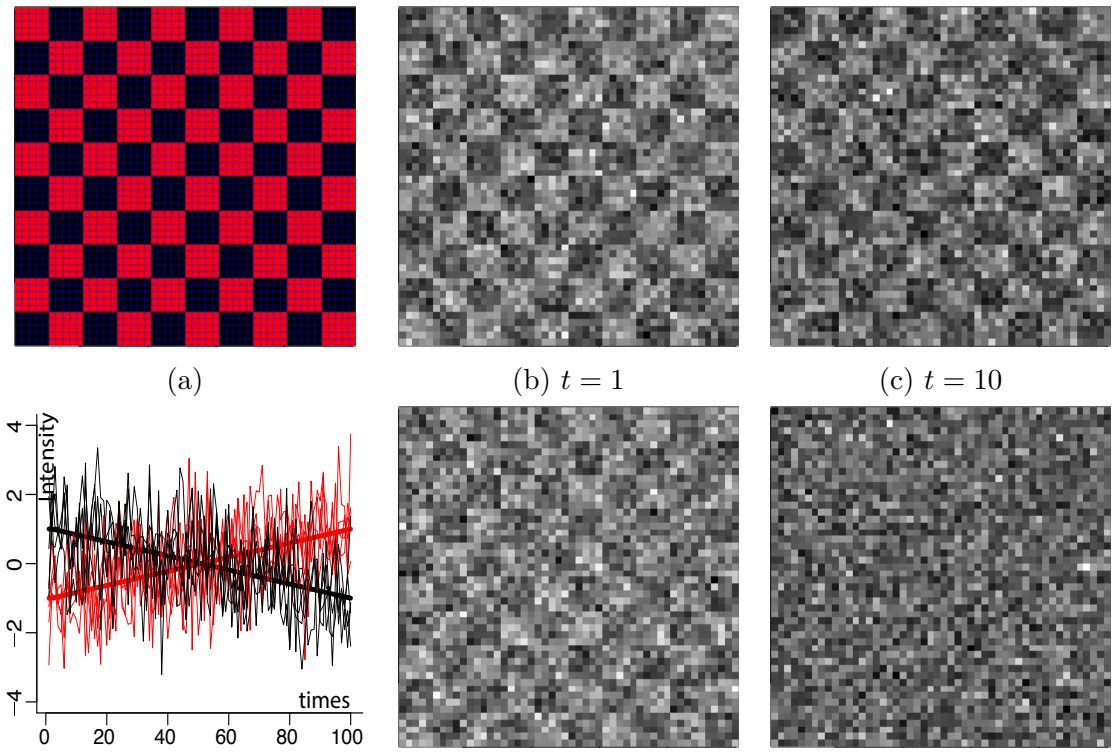


Figure 3.3: Illustration of one iteration in clustering process: (a) build the dissimilarity matrix; (b) compute the dissimilarities between neighbors; (c) locate the minimum dissimilarity; (d) merge the voxels and shrink the matrix; (e) compute and add new dissimilarities.



(a) (b) $t = 1$ (c) $t = 10$

(d) (e) $t = 20$ (f) $t = 50$

Figure 3.4: Chessboard image sequence with 50 by 50 voxels and only two clusters: (a) the red one and the black one; (b,c,e,f) the image at $t = 1, t = 10, t = 20$ and $t = 50$ among 100 times; (d) the straight lines are the true TCs for two clusters and the noisy ones are the observations randomly picked from the sequence.

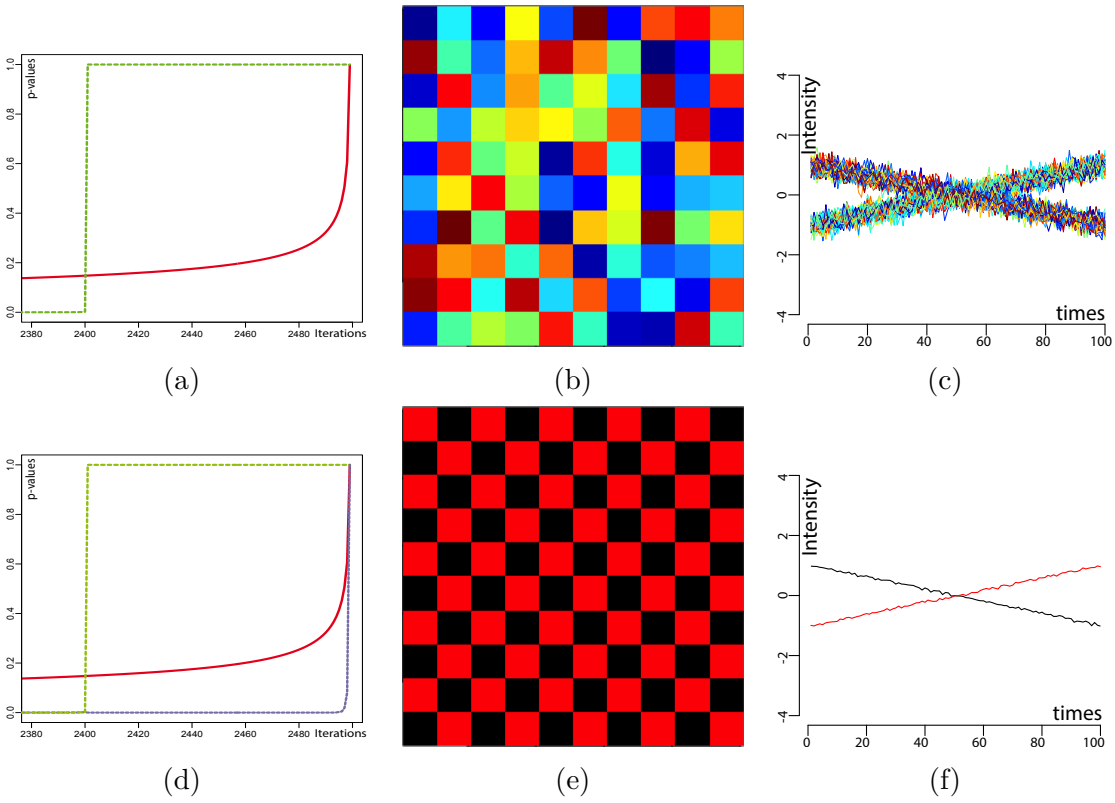
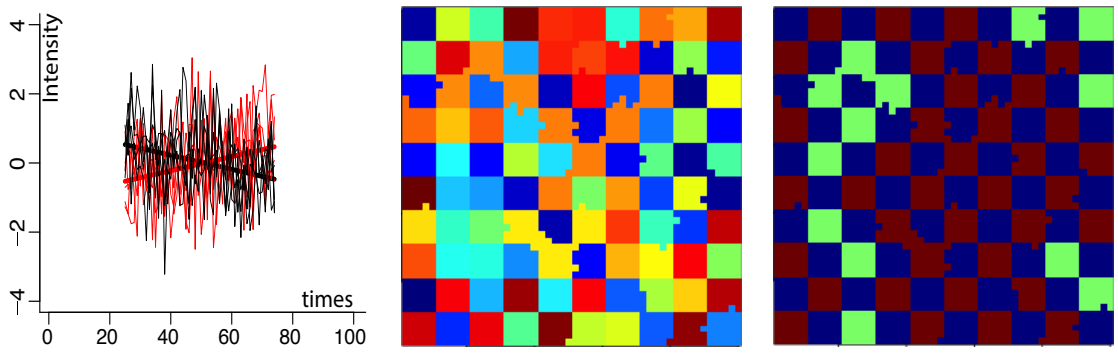


Figure 3.5: Local and global segmentation of Chessboard image sequence: (a) the selection of number of local clusters: the red curve is the the control function in term of iteration and the green is the dissimilarity function during local clustering; (b) the color map of 100 local clusters from local clustering; (c) the averaged (estimated) TCs of local clusters in (b); (d) selection of number of global clusters: the purple curve, additional to (a), is the dissimilarity function during global clustering. (e) the color map of 2 global clusters from global clustering; (f) the average (estimated) TCs of global clusters.



(a) (b) (c)

Figure 3.6: Local and global segmentation of shrunken Chessboard image sequence: (a) the shrunken true and observed observed TCs; (b) local clustering of shrunken chessboard with 2 clusters: the color map of 89 local clusters from local clustering; (c) global clustering of shrunken chessboard with 2 global clusters: the color map of 3 global clusters from global clustering.

Chapter 4

Segmentation of 2D DCE-MR image sequence

In this chapter, we evaluate the proposed method on 2D DCE-MR image sequence. We compare the proposed method to other state-of-the-art methods with synthetic DCE image sequence and study the parameter influence on the segmentation result with real DCE image sequences, along with the model verification.

4.1 Materials: DCE image sequences and ROIs

We have retrospectively analyzed 99 DCE-MR image sequences obtained from female patients having ovarian masses between 1st January 2008 and 28th February 2010 and described in a previous study [Thomassin-Naggara et al., ress]. Our institutional ethics committee approved this retrospective study and granted a waiver of informed consent.

All images were obtained at a single institution on the same 1.5 T scanner (Sonata, Siemens, Erlangen, Germany). In addition to the conventional high-resolution T1 and T2 weighted images, an axial or axial oblique DCE T1-weighted gradient-echo sequence (2D FLASH) was obtained (TR = 27, TE = 2.24, flip angle = 80°, slice thickness = 5 mm, interslice gap = 1 mm, FOV = 400×200, matrix = 256×128, NEX = 1, number of slices = 3) through the tumour at the level

where "solid tissue" was observed on non-enhanced MR images. Saturation bands were placed superior and inferior to eliminate in-flow artefacts. Gadolinium chelate (DOTAREM®, Guerbet, Aulnay, France) was given at a dose of 0.2 $ml \cdot kg^{-1}$ via a Power Injector at a rate of 2 $ml \cdot s^{-1}$ and flushed by 20 ml of saline. Images were obtained at 2.4 s intervals beginning 10 s before the bolus injection, for a total of 312 s and a total of 130 time frames.

DCE-HiSET has been implemented using C++ code wrapped in R and costs computation time for our examples ranging from few seconds to few dozens of seconds, while δ is decreasing. We evaluate it, first, on one synthetic DCE image sequence, and second, on two real DCE-MR image sequences.

The synthetic sequence consists of 120 images on a grid $\mathcal{X}$ of 112x112 voxels. It was made, by radiologists, of eleven clusters having various size and shape, aiming to represent real anatomical structures and their complexity, see Figure 4.1. A random Gaussian noise with variance equal to 1 is added on every voxel at each time.

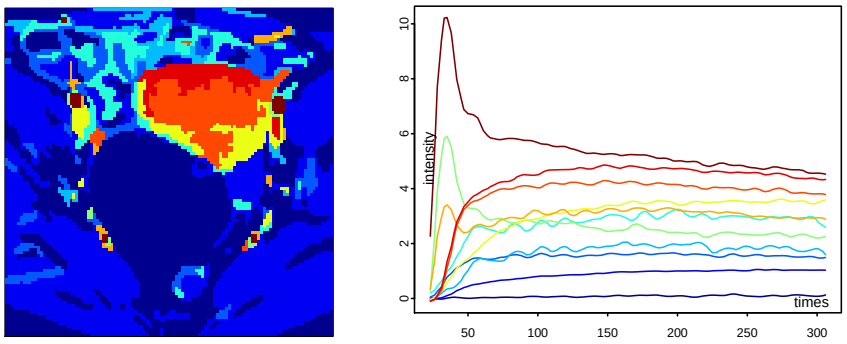


Figure 4.1: Synthetic DCE image sequence: (left) The ground-truth segmentation of $\mathcal{X}$; (right) The true enhancement curves, $i^x(t)$, associated to the 11 clusters using corresponding colors.

The two real DCE-MR image sequences on female pelvis with ovarian tumors [Thomassin-Naggara et al., 2012] consist of respectively 130 and 107 images on a grid $\mathcal{X}$ of 192x128 voxels and are obtained during 305 seconds, see Figure 4.2. The contrast agent is injected with a delay ensuring that the first n_0 images show, up

to the noise level, only the grey level baseline without any enhancement induced by the contrast agent. The value of n_0 is respectively 12 and 10 for the two sequences. Moreover, in both sequences, a ROI around the tumor has been manually drawn by an experienced radiologist after the acquisition of the sequences.

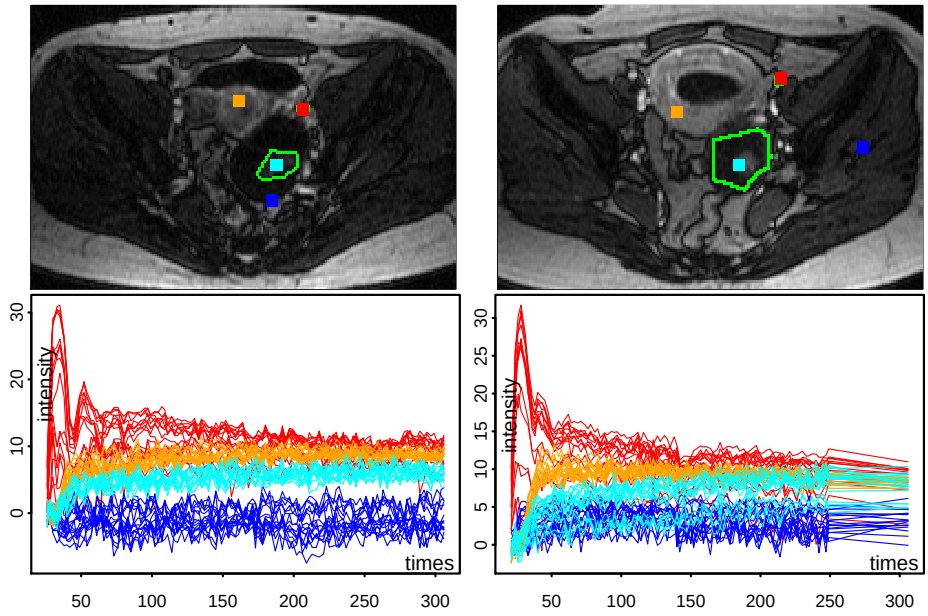


Figure 4.2: Two DCE-MRI image sequences on female pelvis with ovarian tumors - each column shows one sequence. Up: image obtained at time t_{30} (*after arterial phase*) with the tumor ROI (green) together with four 4x4 squared neighborhoods (*r*ed, cyan, orange and blue), the red ones covering the iliac artery identified by the radiologist. Bottom (with corresponding colors): the sets of 16 time enhancement curves, $I^x(t_j)$, observed in the four squares after variance stabilization using $a = 0.45$.

Each real DCE-MR image sequence is handled in following steps:

1. The sequence is transformed with function $f(\psi) = \psi^a/a$ (see §4.5.1 for the choice of a) providing the $J^x(t_j)$, defined in (3.8), for $j = 1, ..., n$ and $x \in \mathcal{X}$.
2. The baseline b^x of each voxel $x \in \mathcal{X}$ is estimated by averaging the $J^x(t_j)$, for $j = 1, ..., n_0$, providing $\hat{b}^x$.
3. For $x \in \mathcal{X}$ and $j = n_0 + 1, ..., n$, $I^x(t_j) := \frac{J^x(t_j) - \hat{b}^x}{\sqrt{1+1/n_0}}$ are the enhancements.

They are Gaussians with variance equal to 1 and are independent as soon as Assumption 3 holds for $J^x(t_j)$.

4. The image sequence is segmented using the $I^x(t_j)$ for $j = n_0 + 1, \ldots, n$ and $x \in \mathcal{X}$.

5. After segmentation, normalized residuals are defined as

$$\hat{\xi}_j^x := \frac{I^x(t_j) - \bar{I}^C(t_j)}{\sqrt{1 - 1/|C|}},$$

for all x in cluster C and $j = n_0 + 1, ..., n$. They follow Gaussian distribution $\mathcal{N}(0, 1)$ as soon as Assumption 3 holds for $J^x(t_j)$.

The synthetic sequence provides directly the $I^x(t_j)$ defined at the third step with $n_0 = 0$. Only the fourth and fifth steps are applied to this sequence.

4.2 Competitors

To compare with DCE-HiSET, we considered the following 6 clustering competitors from the 3 categories previously discussed in Section 2.2:

- model-based: k-means, HMRF-FCM [Chatzis and Varvarigou, 2008] and mean shift (MS) [Comaniciu and Meer, 2002];
- graph-based: normalized cut (NC) [Shi and Malik, 2000];
- hybrid: MS followed by normalized cut (MS-NC) [Tao et al., 2007], SLIC [Achanta et al., 2010] followed by normalized cut (SLIC-NC).

Due to the noisy nature of image sequence, we employed NC only as the second step of hybrid method such that the noise level has been already reduced by the first step.

Competitor details All competitors, except for MS, require as input parameter the number of clusters.

k-means was using Euclidean distance between TCs. A decomposition on the first PCA-eigenvectors in the feature space (of the TCs) was used for other

competitors. For sake of fairness, we used three and six eigenvectors. Three was chosen as the number of eigenvectors needed to cover 95% variance explanation and six as the number of tests used in our multiple equivalence test. We observe no significant difference between two choices and report further the results using six eigenvectors.

It is well known that k-means is highly sensitive to initialization, hence, for each value of k, we ran k-means with 250 different initializations and picked among these runs the one with best value regarding the objective function[1].

HMRF-FCM requires an input partition and is highly sensitive to this choice. It used the k-means result as input and then optimized a regularized objective function of a FCM-type rather than a EM-type [Celeux et al., 2003], which has been proven to be effective for vector-valued image [Chatzis and Varvarigou, 2008].

In MS-NC, NC uses the output of MS as input and can only reduce the number of clusters from the MS step to the required number.

By construction SLIC provides an over-segmentation which is used as the initialization of NC in SLIC-NC.

Tuning parameters of competitors We used value from 7 to 15 for the number of clusters when it was required by a competitor. Other input parameters were

- for HMRF-FCM, a multiplicative constant ε for the entropy penalty controlling the fuzziness. We used ε from 0.2 to 4.
- for MS and MS-NC, the temporal and the spatial bandwidth, bw_t and bw_s. We used bw_t from 0.03 to 0.1 and $r_{ts} := bw_t/bw_s$ from 3 to 15.
- for SLIC-NC, the size and the compactness of supervoxels, sv_s and sv_c, required by SLIC. We used sv_s from 3 to 10 and sv_c from 0 to 0.05.

[1]For the synthetic image sequence, due to the cluster size varying from thousands to only dozens of voxels, the probability that one and only one initial center is chosen in each cluster, ensuring a proper convergence of k-means, can be computed using multinomial distribution and is smaller than $2 \cdot 10^{-9}$.

Competitor implementations We used the `kmeans` implementation for k-means in R^2. Depending on the number of clusters, the computation time ranges from 10 to 90 minutes for the 250 runs.

We were not able to find available code for HMRF-FCM and used a C-implementation wrapped in R with computational time varying from 1.5 to 10 minutes given the initial partition, depending on both the convergence rate (the number of iterations needed to reach convergence) and the number of clusters.

We used the implementation of MS[3] in Matlab[4] with computation time varying from 0.3 seconds to 33 seconds, depending on the values of bandwidths. The C-implementation of NC[5] wrapped in Matlab costs computation time always less than 1 second. Same held for SLIC[6] using its C-implementation wrapped in Matlab with computation time only depending on the supervoxel size. For MS-NC, the computation time depends on the gap between the number of clusters resulting from MS and the one expected by NC.

4.3 Evaluation on the synthetic sequence

For the synthetic sequence, segmentation results accuracy is measured by Fowlkes-Mallows Index (FM) [Fowlkes and Mallows, 1983],

$$\mathrm{FM} = \frac{N_{11}}{\sqrt{(N_{11}+N_{10})(N_{11}+N_{01})}},$$

where $N_{..}$ is the number of voxel pairs that are classified into: one same cluster in both partitions (N_{11}); the same cluster in the first partition but two different clusters in the second partition (N_{10}); two different clusters in the first partition but the same cluster in the second partition (N_{01}).The value of this index is in between 0 and 1. The closer it is to 1, the more similar the segmentations are.

[2] `https://cran.r-project.org/`
[3] `http://a-asvadi.ir/`
[4] `http://www.mathworks.com`
[5] `http://www.timotheecour.com/`
[6] `http://www.vlfeat.org/`

To take into account the effect of cluster size on the accuracy measure, we also consider a weighted version, wFM. For wFM, a pair of voxels (x_1, x_2) is counted for $w_1 \cdot w_2$ instead of 1, where $w_i = \frac{|\mathcal{X}|}{|C_i|}$ with C_i being the cluster to which x_i belongs in the first partition, for $i = 1, 2$.

In order to provide a representation of the difference between two partitions $(C_1, \ldots, C_\ell)$ and $(D_1, \ldots, D_{\ell'})$, we also compute the so-called *error-map* as the indicator function of the set

$$\bigcup_{i=1}^{\ell} (C_i \setminus D_{j_i}) \quad \cup \quad \bigcup_{j=1}^{\ell'} \left(D_j \setminus C_{i_j}\right)$$

where $i_j = \arg\max_i |D_j \cap C_i|$ and $j_i = \arg\max_j |C_i \cap D_j|$.

We used DCE-HiSET with $\alpha = 0.001$ and various values of δ to segment the synthetic image sequence. The effect of the latter on the segmentation accuracy and on the number of clusters is illustrated in Figure 4.3. We observed that the FM and wFM indexes remain stable and higher than 0.8 in a large δ-range from 0.5 to 2. Clearly small values of δ result in over-segmentation (ℓ^* too large) while large values enforce under-segmentation. Nevertheless, as one can expect when thinking of bias-variance tradeoff, it is clearly more risky to use too small values than too large ones. Following the slope heuristic [Baudry et al., 2012; Birgé and Massart, 2007], the fast decrease of the number of clusters when δ increases may provide an automatic choice of δ: one can detect the value δ_0 of δ where the relative slope, defined as $(\ell^*(\delta^-) - \ell^*(\delta^+))/\ell^*(\delta^-)$, is less than a fixed value and then define the optimal δ as $\delta^* = 2\delta_0$.

Figure 4.4 shows the segmentation results with 3 values of δ, including δ^*. From this figure, we observe clearly that when δ is chosen too small large homogeneous regions are split into large sub-regions having irregular borders. As expected when δ is too large, true clusters start to merge, however the geometry of the borders is not changed, ensuring a good recognition of most of the structures.

In order to study the effect of TC separations on the segmentation results and the stability of DCE-HiSET with respect to the parameter δ, we have multiplied

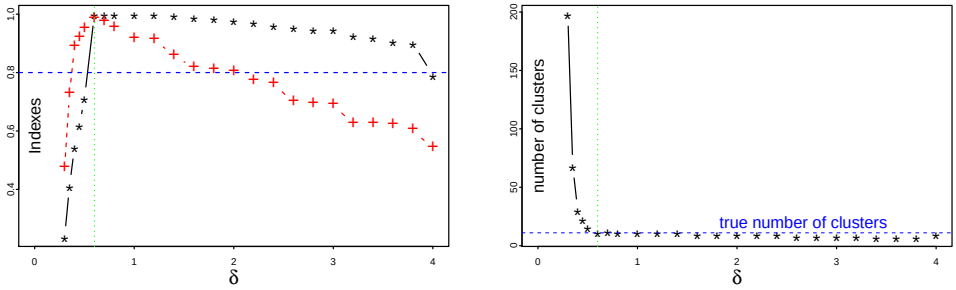


Figure 4.3: Synthetic image sequence segmentation using DCE-HiSET when δ varies: (Left) Fowlkes-Mallows Index (black stars) and its weighted version (red crosses) - (Right) number of clusters. Result with best indexes is achieved at $\delta = 0.6$ (green dashed line).

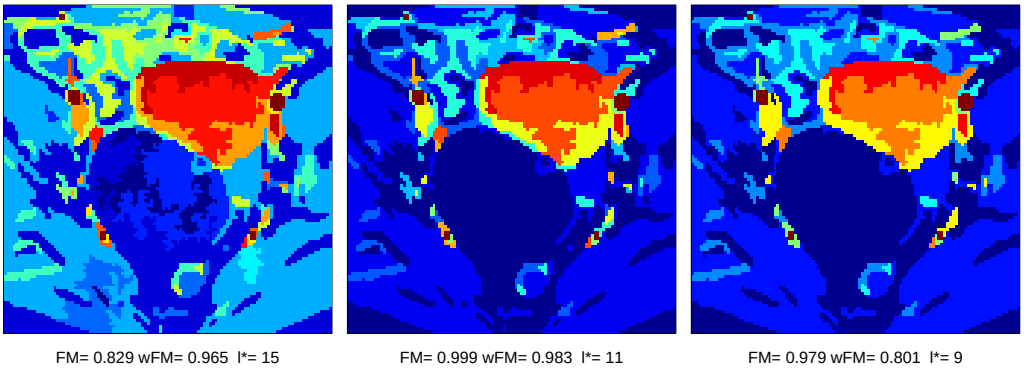


Figure 4.4: Segmentation results of DCE-HiSET of synthetic image sequence with δ equals to 0.5 (left), $\delta^* = 0.6$ (middle) and 2.0 (right) when $\alpha = 0.001$.

the true enhancement curves by 2/3 and 4/3. The noise level was kept at the same value with standard deviation equal to 1 in order to change the SNR. Figure 4.5 shows that both FM and wFM indexes, when δ varies, are stable on a large δ-range. Moreover, the higher the SNR is, the larger this δ-range is. Surprisingly only the maximum of wFM is affected by the SNR. The error maps corresponding to the best wFM are given in Figure 4.6.

We now provide a comparison with the competitors on the synthetic sequence. Information and performance of all competitors together with DCE-HiSET are summarized in Table 4.1. More details of each competitor follow.

For k-means, we present in Figure 4.7 the result when $k = \ell^* = 11$ that is the

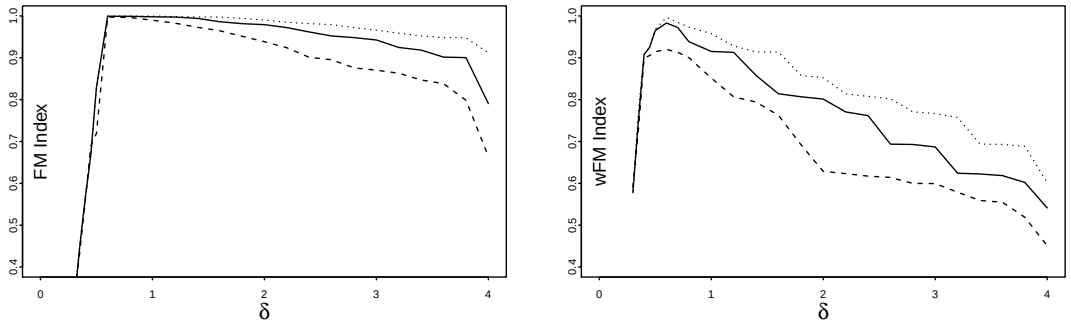


Figure 4.5: FM index (left) and weighted FM index (right) when true enhancement curves, $i^x(t)$ (Fig. 4.1), are multiplied by 1 (solid), 2/3 (dashed) and 4/3 (dotted).

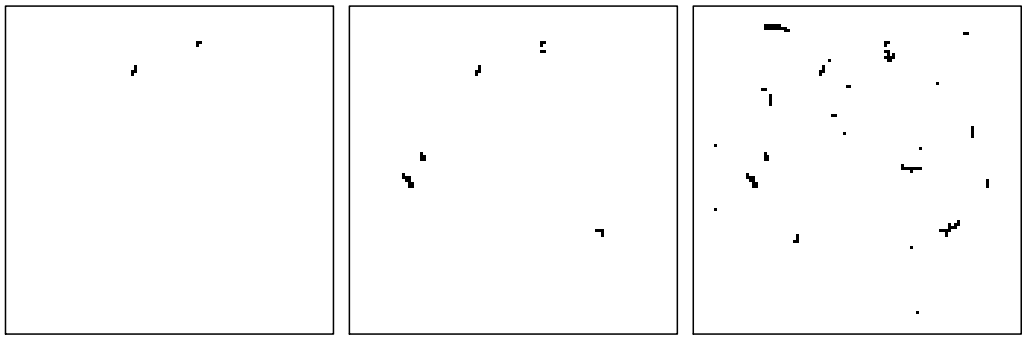


Figure 4.6: Error maps when the SNR decreases by multiplying the true enhancements by 4/3 (left), 1 (middle) and 2/3 (right).

true number of clusters, together with the best FM and wFM values achieved when k varies. From these results, one can clearly see the default of k-means. On one hand, having two or more initial centers in one true cluster (see $\ell^* = 10$ or 11 in Fig. 4.7) leads resulting clusters to be pulverized (so-called *pepper and salt* effect). On the other hand, when no center reaches small clusters, the latter are artificially merged, leading to a high FM but low wFM (see $\ell^* = 8$ in Fig. 4.7). We did not use any automatic choice of k since the best results, with respect to both FM and wFM indexes, achieved by k-means are already worse than DCE-HiSET.

If the initial partition given by k-means has no pulverized cluster, HMRF-FCM barely improves the segmentation of k-means. Otherwise, HMRF-FCM regularized the pulverized clusters by grouping their voxels up to some kind of "panther texture" when ε becomes large enough (see Fig. 4.7). However, if one cluster has been

Method	Parameter	Range	Highest FM	Highest wFM
k-means	k	7 – 15	0.993 ($k = 8$)	0.827 ($k = 10$)
HMRF-FCM	k	7 – 15	0.993 ($k = 8$, $\varepsilon = 0.5$)	0.827 ($k = 10$, $\varepsilon = 1$)
	ε	0.2 – 4		
MS	bw_t	0.03 – 0.1	0.994	0.919
	r_{ts}	3 – 14	($bw_t = 0.06$, $r_{ts} = 11$)	
MS-NC	k	7 – 15	0.993	0.852
	bw_t	0.03 – 0.1	($k = 10$, $bw_t = 0.06$, $r_{ts} = 11$)	
	r_{ts}	3 – 14		
SLIC-NC	k	7 – 15	0.983	0.821
	sv_s	3 – 10	($k = 12$, $sv_s = 3$, $sv_c = 0$)	
	sv_c	0 – 0.05		
DCE-HiSET	δ	0.2 – 4	**0.999** ($\delta = 0.6$)	**0.983** ($\delta = 0.6$)

Table 4.1: The parameter information and the performance of all competitors and of DCE-HiSET.

already split in the initial partition provided by k-means, HMRF-FCM is unable to fix it. As a consequence, the increase of FM and wFM induced by HMRF-FCM is very small for given k. Moreover, the best FM and wFM indexes, when k and ε vary, have kept unchanged with respect to k-means.

For each of MS, MS-NC and SLIC-NC, like for DCE-HiSET, the best results for both indexes are achieved at a same parameter set.

For MS, the best result has been achieved at $bw_t = 0.06$ and $r_{ts} = 11$. With these values, MS recovered the right number of clusters, as shown in Figure 4.8. High FM (0.994) and relatively low wFM (0.916) indicate that mistakes have been made in small clusters, for instance, two voxels (one red and one orange) are spatially disconnected from their clusters. This reveals the sensibility of MS to noise artifacts.

For MS-NC, the best result has been achieved also at $bw_t = 0.06$ and $r_{ts} = 11$. It leads to only 10 clusters with worse indexes (see Fig. 4.8) than for MS alone. Indeed, two clusters resulting from the best result of MS have merged during the

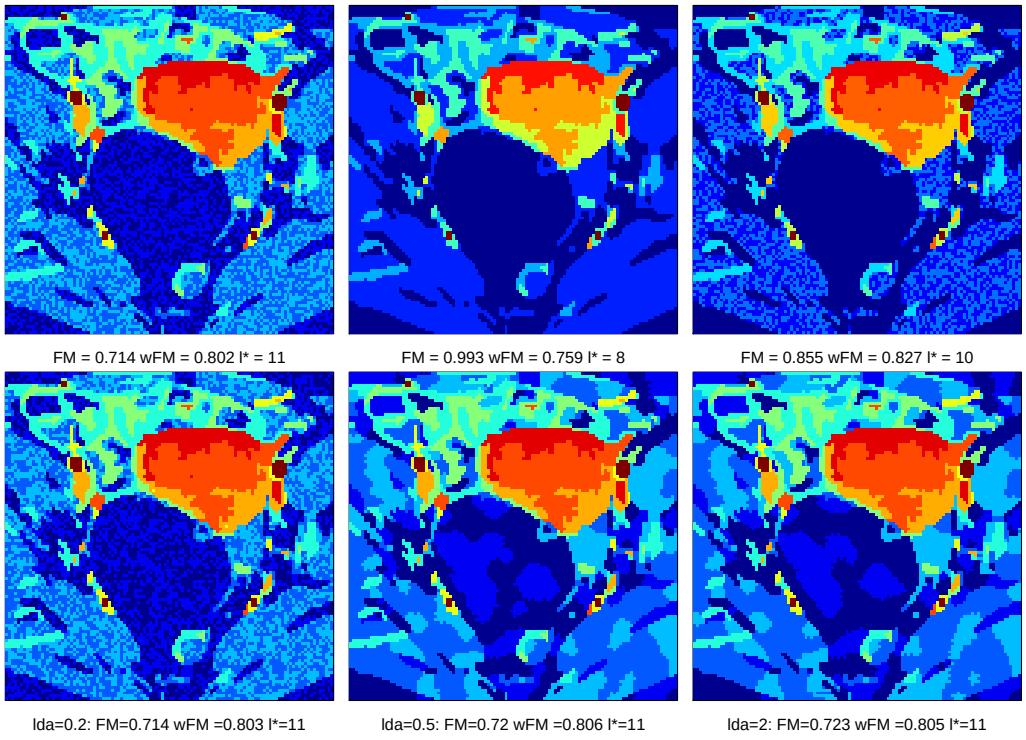


Figure 4.7: Competitors' best results on synthetic sequence. Top: Best result of k-means among 250 runs given the true number of clusters $k = 11$ (left); with highest FM index (middle) and with highest wFM index (right). Bottom: Results of HMRF-FCM, initialized by the result of k-means with $k = 11$ and using $\varepsilon = 0.2$ (left), $\varepsilon = 0.5$ (middle) and $\varepsilon = 2$ (right).

NC step while others remain unchanged.

For SLIC-NC, the best result has been achieved at $sv_s = 3$ and $sv_c = 0$ with 12 clusters, providing FM=0.983 and wFM=0.821 (see Fig. 4.8). It has neither the pulverized issue like k-means nor the noise sensitive issue in MS, thanks to the first step with SLIC. However, due to the supervoxel size limitation in the SLIC step, voxels in the clusters of small size may be spread in the surrounding large clusters and/or narrow parts of clusters tend to have thicker shapes (see Fig. 4.8).

Despite their weakness in term of indexes, as DCE-HiSET, MS, MS-NC and SLIC-NC show the ability to recover regions highly consistent to the ground-truth segmentation.

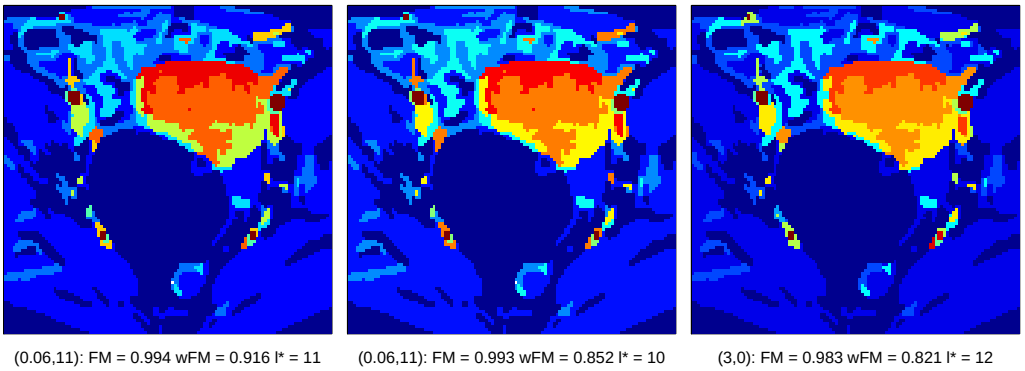


Figure 4.8: Competitors' best results on synthetic sequence. Left: MS (bw_t,r_{ts})=(0.06,11). Middle: MS-NC (bw_t,r_{ts},ℓ^*)=(0.06,11,11). Right: SLIC-NC (sv_s,sv_c,ℓ^*)=(3,0,12).

Figure 4.9 shows the error maps of best results achieved by each competitor in order to compare their behaviors in term of locations of the segmentation errors (misclassified voxels).

From this study of the synthetic sequence, we see clearly that DCE-HiSET

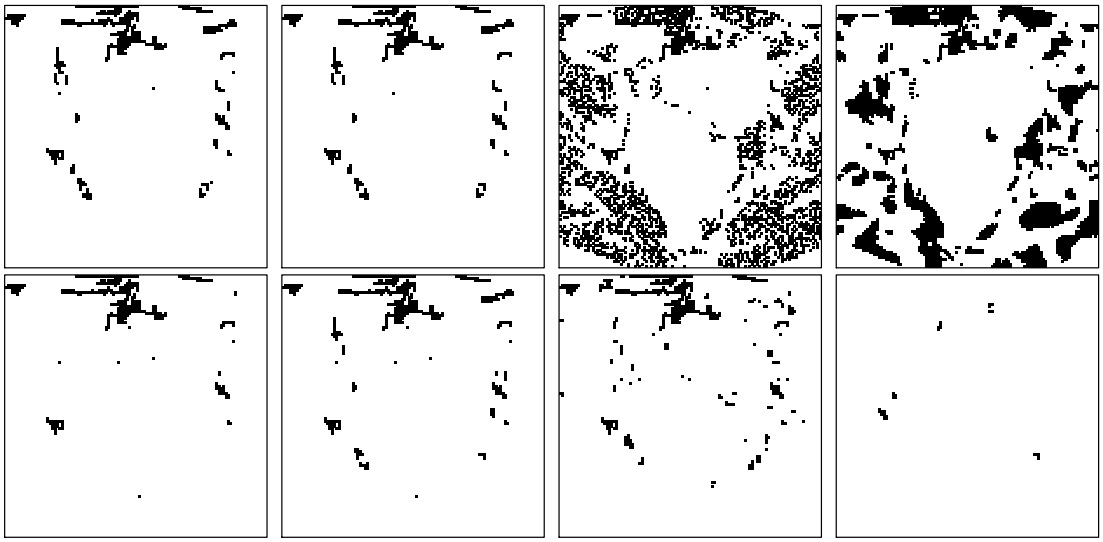


Figure 4.9: Error maps on synthetic sequence. From left to right, (Top) k-means with highest FM, HMRF-FCM with the highest FM, k-means with highest wFM, HMRF-FCM with highest wFM, (Bottom) best (FM- and wFM) results for MS, MS-NC, SLIC-NC and DCE-HiSET.

outperforms all others with respect to both FM and wFM indexes with only a single parameter having a very stable behavior. The error maps show clearly that all competitors have trouble with small and/or complex clusters while DCE-HiSET makes much fewer errors in these areas, thanks to the multi-resolution comparison. Moreover, when the SNR decreases the segmentation errors of DCE-HiSET expand regularly all over the image.

4.4 Comparison of the robustness with respect to SNR

In this section, the effect of noise level on segmentation result is investigated for all competitors. In the same way than the study on the stability of HiSET in last section, we have multiplied the true enhancement curves of the synthetic image sequence by different ratios varing from 1/2 to 7/3, while the noise level is remained at the same value with standard deviation equal to 1, such that the SNR is decreasing. For each ratio, the best segmentation result with the highest FM index of six methods are compared with each other along with the corresponding error map (Figure 4.10-4.14). For HiSET, δ is fixed at 1 and α is fixed at 0.001 for all ratios. For all other five methods, the parameters are freely tuned (within a reasonable range) such that the highest FM index is achieved, except that the true number of clusters is given for k-means and HMRFFCM. Apparently, HiSET results fewer mistakes, even for the smaller ratios (lower SNR).

Additionally, the values of FM and weighed FM index are given, respectively, in Table 4.2 and 4.3. In term of FM index, MS, MSNC, SLICNC and HiSET show much more stability than k-means and HMRFFCM with respect to SNR. The accuracies of the former four methods are also much superior to the latter two and all are close to 1 (Figure 4.15). Considering also the weighed FM index, HiSET shows a more stable and better performance than other methods.

ratio	k-means	HMRFFCM	MS	MSNC	SLICNC	HiSET
*1/2	0.566	0.581	0.961	0.961	0.969	0.972
*2/3	0.655	0.674	0.990	0.989	0.983	0.990
*1	0.714	0.722	0.994	0.994	0.988	0.998
*4/3	0.881	0.882	0.994	0.994	0.988	1.000
*5/3	0.882	0.808	0.994	0.994	0.989	1.000
*2	0.855	0.855	1.000	1.000	0.985	1.000
*7/3	0.657	0.660	1.000	1.000	0.985	1.000

Table 4.2: FM index of the best result of each method given the true enhancements are multiplied by different ratios.

ratio	k-means	HMRFFCM	MS	MSNC	SLICNC	HiSET
*1/2	0.636	0.649	0.703	0.703	0.803	0.788
*2/3	0.689	0.699	0.820	0.805	0.861	0.852
*1	0.802	0.806	0.916	0.914	0.879	0.915
*4/3	0.889	0.895	0.894	0.858	0.883	0.958
*5/3	0.895	0.823	0.919	0.919	0.865	0.977
*2	0.805	0.814	0.962	0.916	0.901	0.998
*7/3	0.701	0.701	0.986	0.918	0.900	1.000

Table 4.3: Weighted FM index of the best result of each method given the true enhancements are multiplied by different ratios.

4.5 Experiment on real sequences

Figure 4.16 shows the influence of parameters α and δ on the segmentation. Clearly α has a neglectable effect on the number of clusters while δ has a strong influence on it during both the local and global steps. Hence, we fixed $\alpha = 0.001$ for the rest of this study. Let us point out that we observed that the cluster shape is not much influenced by α neither.

4.5.1 Model justification and parameter a selection

In order to justify our modelling given by equations (3.6-3.9), we studied the distribution of the normalized residuals $\hat{\xi}_j^x$ when δ and a vary (see Fig. 4.17). We observed that, while δ varies, the $\hat{\xi}_j^x$-distribution tails are heavier than the Gaussian when a is large (Poisson's case: $a = 0.5$) and are lighter when a is too small ($a = 0.4$), showing a proper behavior for intermediate values (e.g. $a = 0.45$). In

this last case, δ becomes the right tuning parameter controlling the discrepancy between TCs. Clearly, the residuals do not have a perfect Gaussian behavior as their distribution is, even for intermediate values, more concentrated than Gaussian. However, first, thanks to the use of tests, only tails are of interest, second, our construction may be extended to the sub-Gaussian case (only the independence of the tests being a concern). Yet such generalization would require more theoretical developments that are beyond the scope of this paper. Similarly, adaptation with respect to a or improving the model for variance stabilization would be of interest for future works. Nevertheless, the quality of our empirical experiments shows that the benefit will be probably small.

We continue this study fixing $a = 0.45$.

4.5.2 Segmentation results

The two real DCE-MR image sequences have been fully segmented into homogeneous regions that are highly consistent to the anatomical structure as shown in Figure 4.18. The smaller δ is, the more details in anatomical structures can be observed in image sequence. This is highlighted inside the squares (in black on Figure 4.18) that contain the manually segmented ROIs (see Fig. 4.2). The corresponding estimated TCs inside each cluster within the ROI together with their corresponding size are shown in Figure 4.20. Thanks to the opportunity to segment a full DCE sequence, these estimated TCs are obtained by averaging TCs which do not necessarily belong to the ROI but to the same homogeneous cluster. As a benefit, the SNR observed for these estimated TCs is strongly improved, providing a true opportunity for further analysis and comparisons. From these figures, one can clearly understand what DCE-HiSET is doing. It is indeed providing a piecewise constant representation of the DCE image sequence in functionally homogeneous regions, where δ controls the size of the pieces and functional "step" between them.

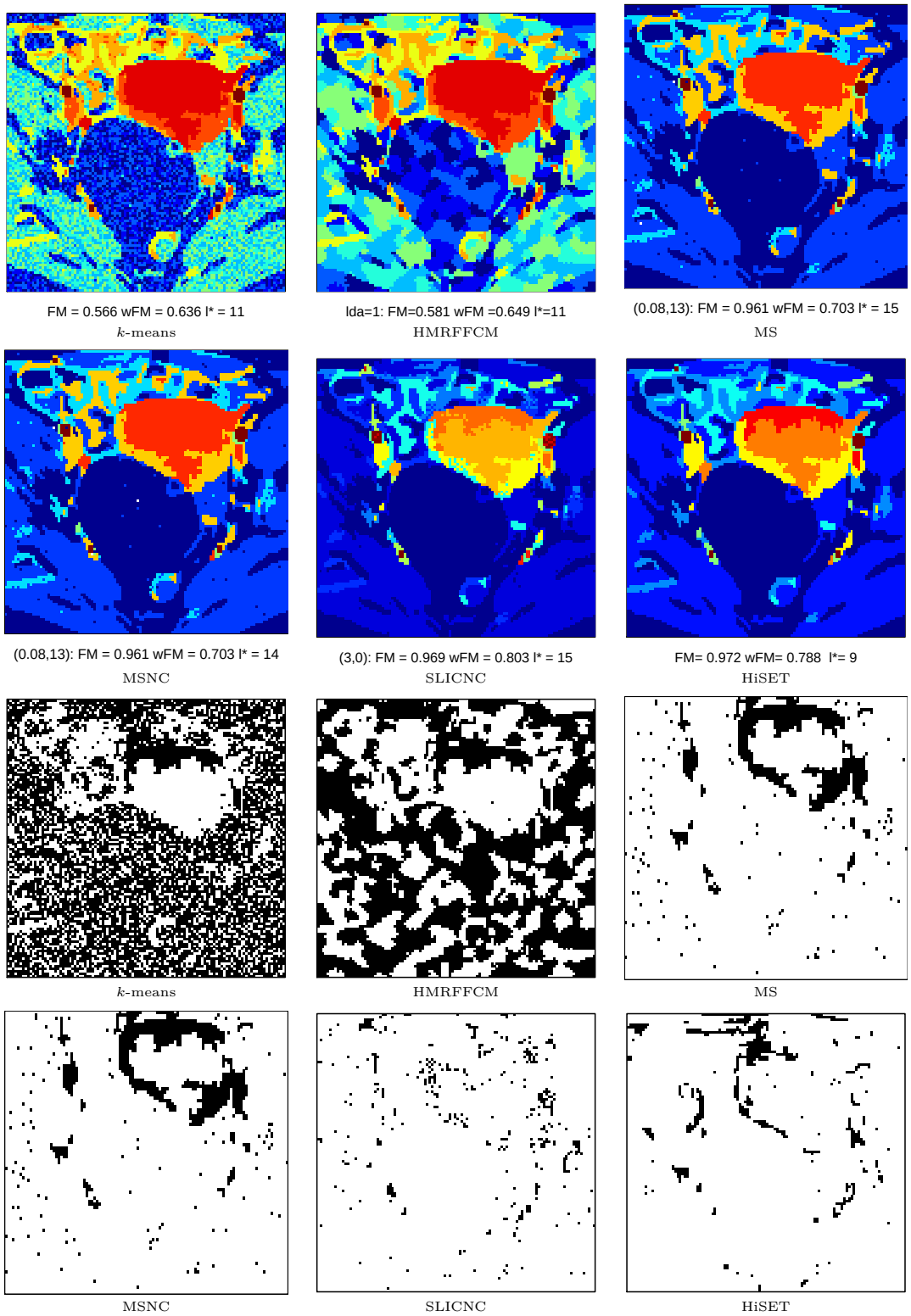


Figure 4.10: Segmentation results (the first two rows) and error maps (the last two rows) of six methods when the true enhancements are multiplied by 1/2. The result with highest FM index is given for each method except HiSET. The right number of clusters is given for k-means and HMRFFCM. For HiSET, $\delta = 1$ and $\alpha = 0.001$.

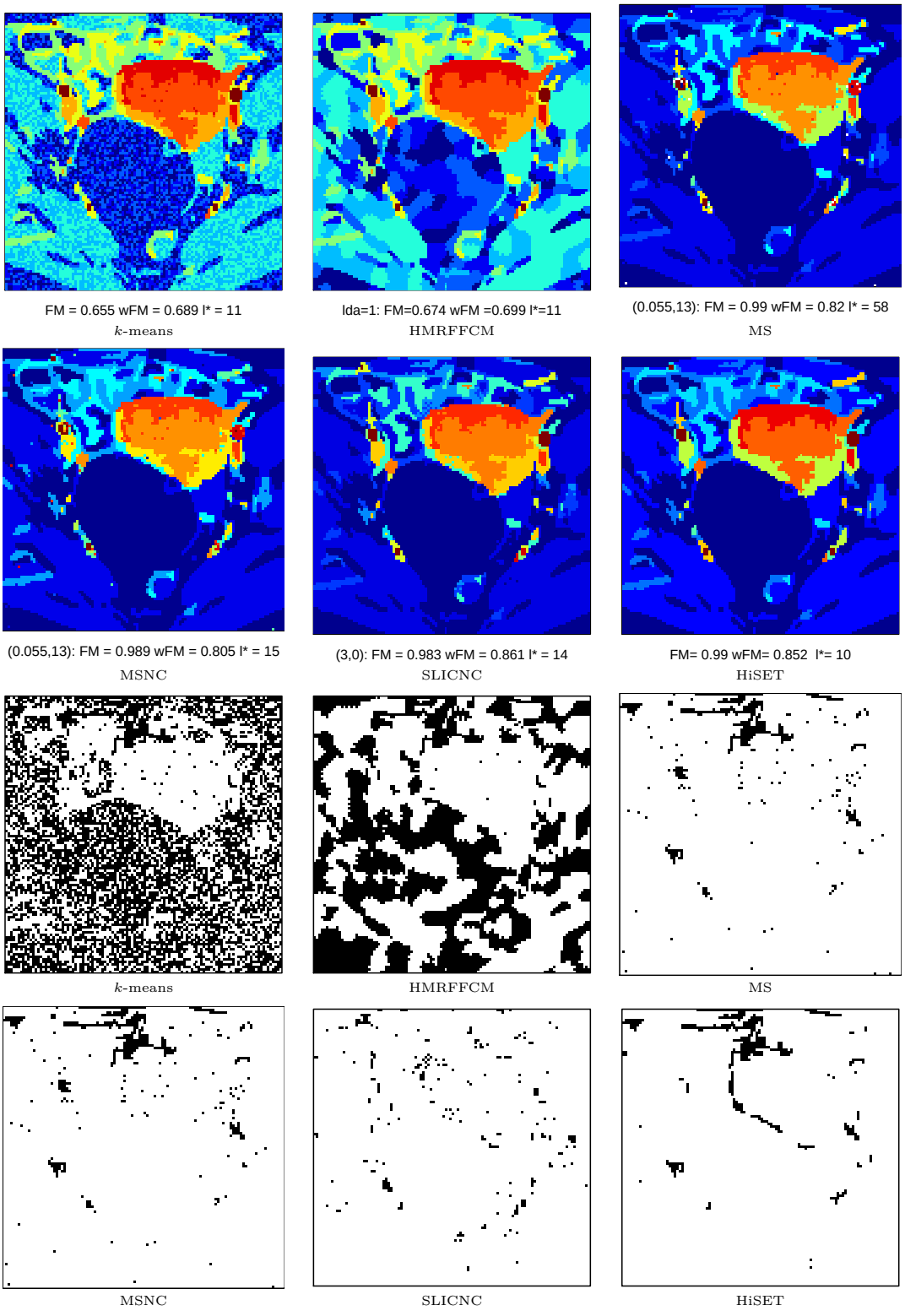


Figure 4.11: Segmentation results (the first two rows) and error maps (the last two rows) of six methods when the true enhancements are multiplied by 2/3. The result with highest FM index is given for each method except HiSET. The right number of clusters is given for k-means and HMRFFCM. For HiSET, $\delta = 1$ and $\alpha = 0.001$.

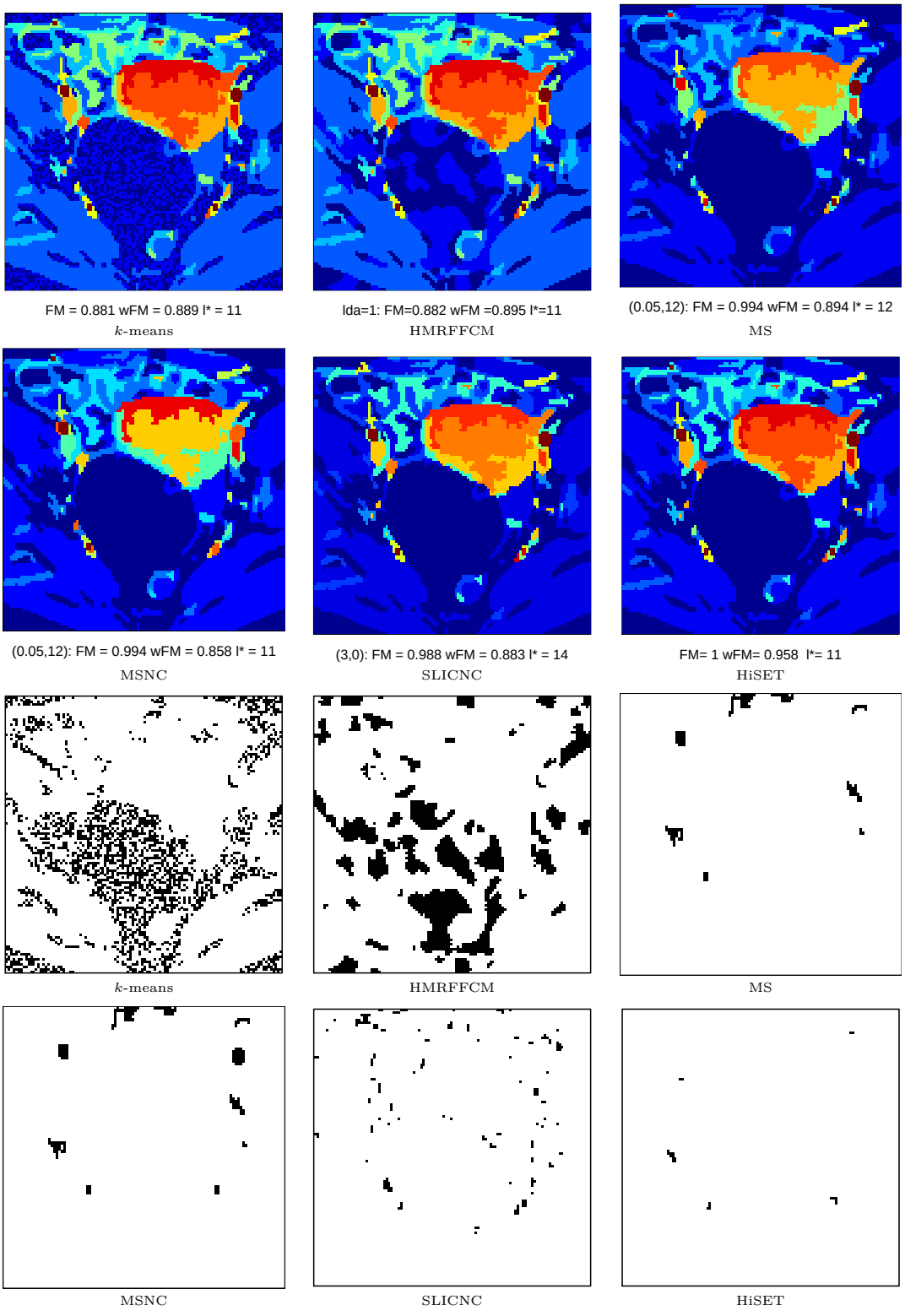


Figure 4.12: Segmentation results (the first two rows) and error maps (the last two rows) of six methods when the true enhancements are multiplied by 4/3. The result with highest FM index is given for each method except HiSET. The right number of clusters is given for k-means and HMRFFCM. For HiSET, $\delta = 1$ and $\alpha = 0.001$.

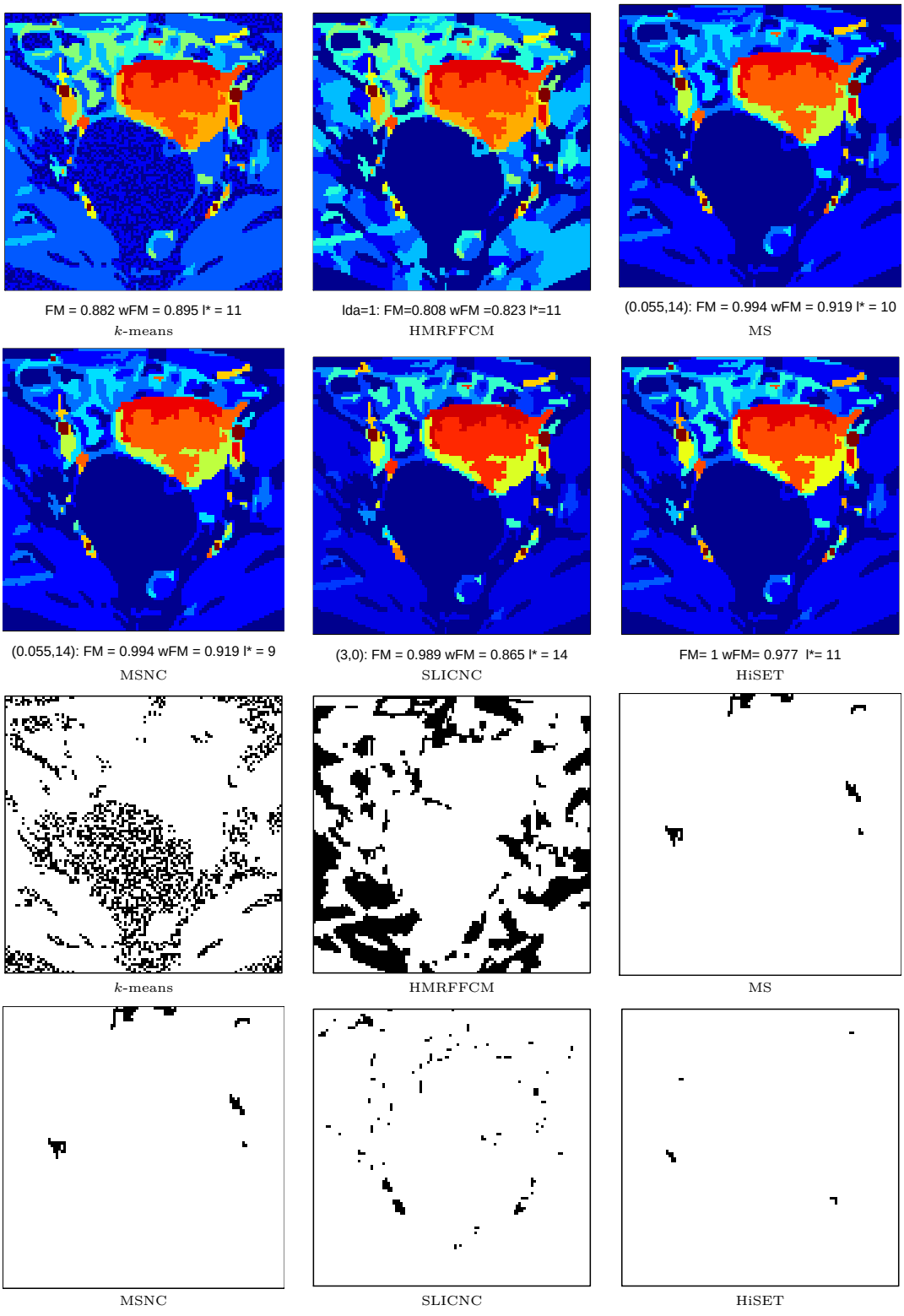


Figure 4.13: Segmentation results (the first two rows) and error maps (the last two rows) of six methods when the true enhancements are multiplied by 5/3. The result with highest FM index is given for each method except HiSET. The right number of clusters is given for k-means and HMRFFCM. For HiSET, $\delta = 1$ and $\alpha = 0.001$.

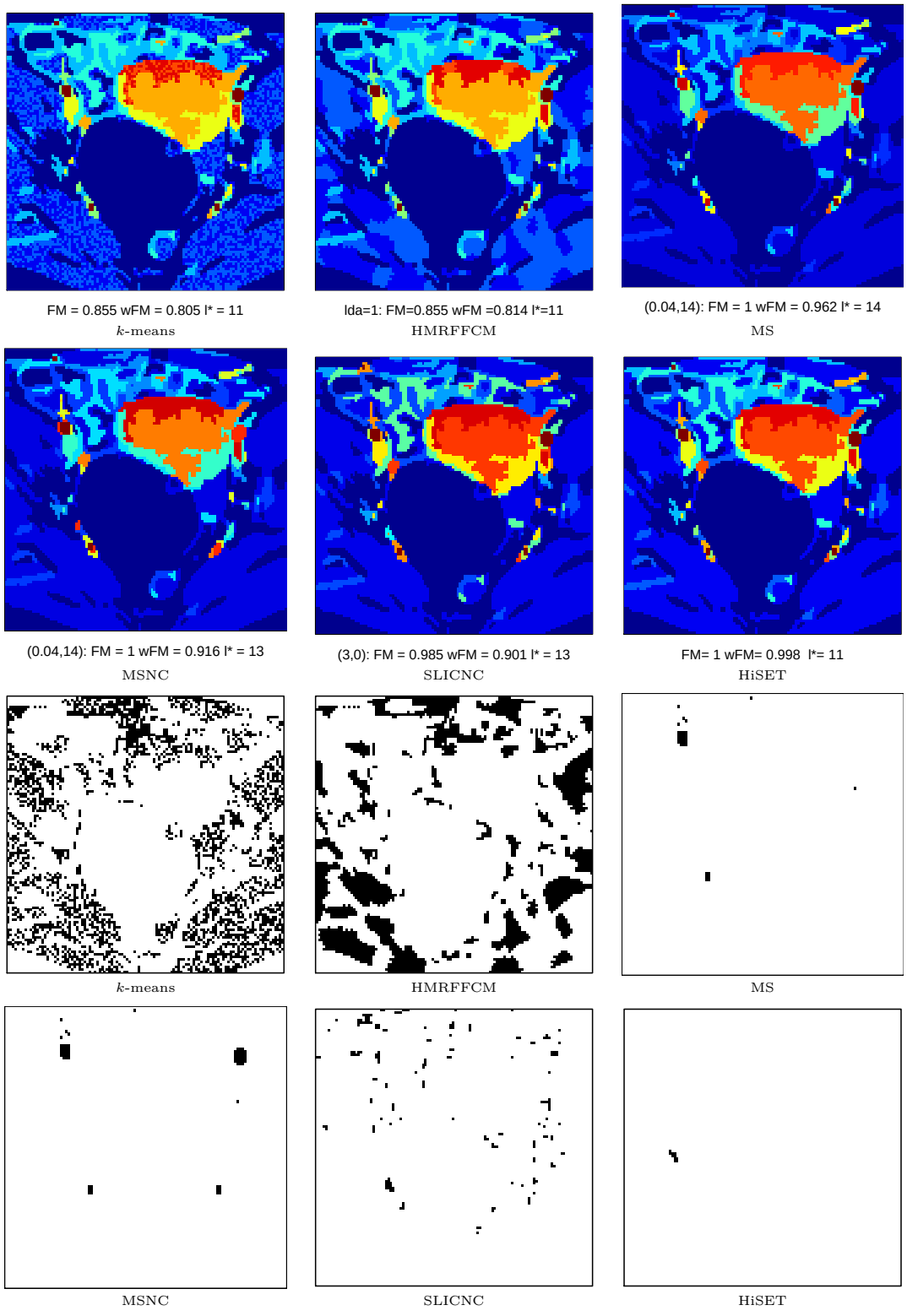


Figure 4.14: Segmentation results (the first two rows) and error maps (the last two rows) of six methods when the true enhancements are multiplied by 2. The result with highest FM index is given for each method except HiSET. The right number of clusters is given for k-means and HMRFFCM. For HiSET, $\delta = 1$ and $\alpha = 0.001$.

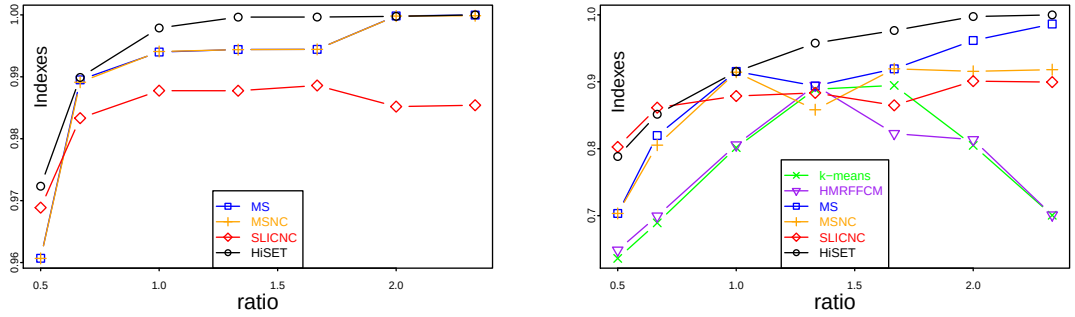


Figure 4.15: FM index (left) and weighted FM index (right) of best result of each method when true enhancement curves are multiplied by ratios from 1/2 to 7/3. Only four methods (MS, MSNC, SLICNC and HiSET) are shown for FM and all methods are shown for weighted FM index.

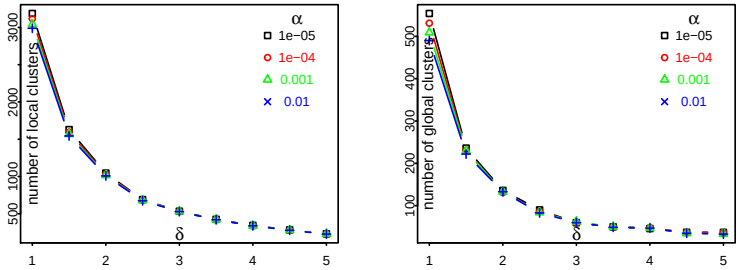


Figure 4.16: Number of clusters when α and δ vary for (*left*) local and (*right*) global steps.

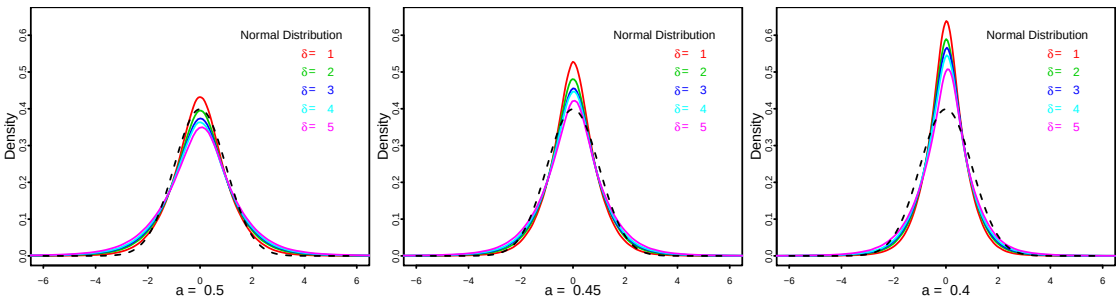


Figure 4.17: Normalized residual distributions (with δ=1/2/3/4/5, one color for each) and normal distribution density (black and dashed) with a=0.5/0.45/0.4 of variance stabilization.

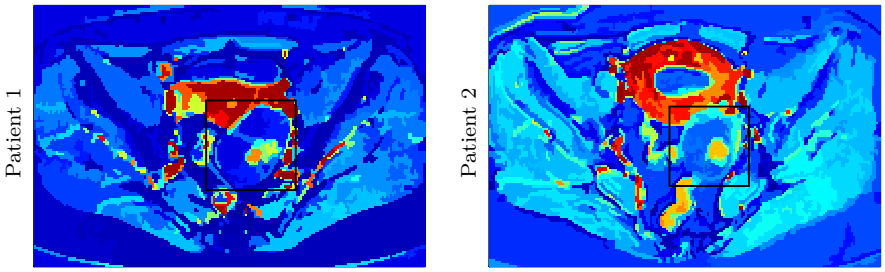


Figure 4.18: Segmentation results of the two real DCE-MR image sequences using $\delta = 3$.

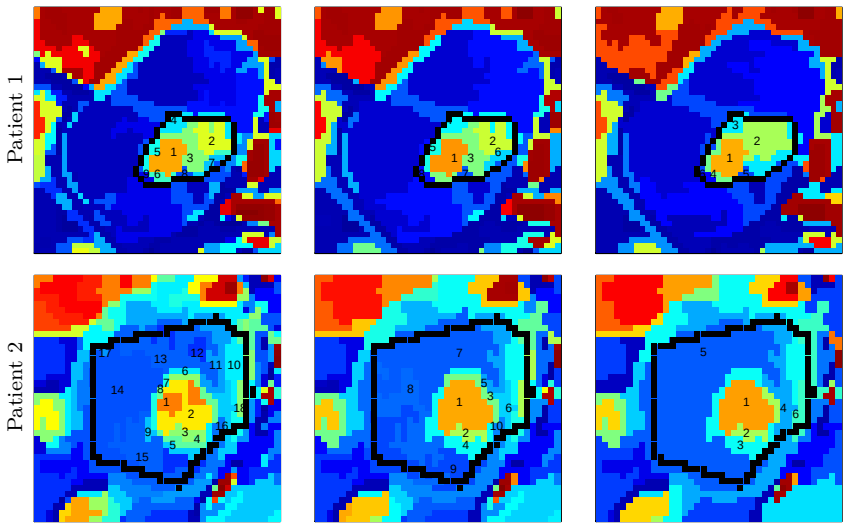


Figure 4.19: Squared zoom for the DCE-MR image sequence with δ equals 2 (left), 3 (middle) and 4 (right). Manually segmented ROI appears in black and clusters inside are numbered. The first sequence shows from 9 to 6 clusters (top). The second from 18 to 6 (bottom).

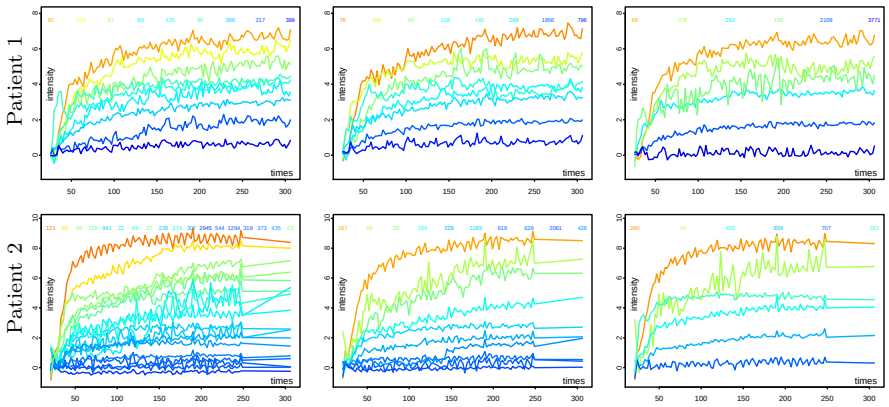


Figure 4.20: Average TCs (after variance stabilisation) of the clusters inside the manually segmented ROIs (see Fig. 4.19) with δ=2, 3 and 4 from left to right. Size of the corresponding clusters are given at the top of each subfigure with corresponding colors.

Chapter 5

Automatic refinement of manual ROI

In this chapter, we propose an automatic strategy to refine the manually delineated ROI based on DCE image sequence segmentation. As an application of the proposed clustering method, the strategy is described with an evaluation of robustness and efficiency through classification of ovarian tumors.

In medical imaging, regions of interest (ROIs) are often manually drawn by defining their boundary on one image such that qualitative or quantitative analysis can be performed, in order to retrieve information regarding the tissues of interest. The boundary of ROI should be drawn along the significant boundaries between the tissues. However, the identification of the precise expected boundary of ROI by only visual inspection can be a very challenging and time-consuming task.

For analysis of the image sequence obtained from DCE imaging, ROIs are manually delineated either on one of the images of the dynamic sequence (often an image during the injection phase giving high contrast between tissues), or on an image of high anatomical quality coming from a different sequence (high-resolution T1 or T2 weighted images) and then pasted to all other images in the DCE image sequence. Comparing to individual image, the precision issue is amplified for image sequence. Instead of an intensity value, each voxel has functional information

varying along the acquisition time. Information contained in two voxels could be homogeneous at one time but not at another time. Hence, the ROI that is delineated on the reference image to contain all desired information could include undesired information on other images. In addition, the ROI may also shift away from the intended region (tissue) due to movements occurring during the image acquisition period or due to registration errors between the high-resolution image and the dynamic sequence. Therefore, semi-automatic techniques are required to decrease the delineation time as well as to improve the robustness of analysis based on ROI.

Substantial efforts have been made on the topics regarding (semi-)automatic ROI delineation (or selection). Craciunescu et al. [2012] and Heye et al. [2013] evaluated the impact of tumor ROI selection on the assessment of intra- and inter-patient variability. Ortiz et al. [2014], Górriza et al. [2009] and Liu et al. [2014] proposed ROI selection methods for diagnosis purpose. Zöllner et al. [2009], Li et al. [2012b] and Stoutjesdijk et al. [2007] applied clustering-based methods on ROI segmentation in order to improve the perfusion analysis based on the segmented ROIs. Forbes et al. [2006] investigated the ROI selection for both diagnosis and perfusion analysis.

We propose a new method to automatically refine and to improve the efficiency of ROI selection in DCE image sequences. Instead of considering each image in the sequence individually and choosing the potential optimal reference, we take into account the homogeneity of functional information.

Through DCE-HiSET, we can recover all functionally homogenous regions in the image grid. Manipulating the ROI with an automatic pipeline of erosion and dilation, we preserve the homogeneous regions connected to the interior of the ROI as the refined ROI and remove the regions only located at or connected to the border of manual ROI. In consequence, refined ROI has a very optimized boundary and is highly consistent to the tissue expected with respect to the functional information relating to its perfusion or diffusion behavior. In addition, we automatically filter the TC with non physiological behavior which are mostly generated by motion.

Finally, to evaluate the efficiency of this new semi-automated ROI selection method, we tested it on a set of DCE-MRI sequences of 99 ovarian tumors assuming that a better ROI selection will help in differentiating benign, borderline and malignant tumor when analyzing the TC of the voxels found within the ROI.

5.1 Materials

For every dynamic sequence, experienced radiologist (IT) manually delinea- ted a ROI of suspicious tumor tissue by drawing coarsely and rapidly the limits vertices of a polygon on a single image of the series. Several examples of manually delineated ROI are shown in the first column of Figure 5.7.

After surgery, the tumors have been characterized on pathology as benign, borderline or malignant. Borderline is an intermediate state in between benign and malignant. Among all, 49 are categorized as benign, 9 borderline and 41 malignant.

All the image and ROI processing have been made on R software (https://www.r-project.org/), after an import of original DICOM images. All statistical analyses have also been made on R software. No motion correction has been used for any DCE image sequence.

5.2 Filtration of average TCs

For every resulting cluster from DCE-HiSET, we compute the average TC from the observed TCs (after variance stabilization) of all voxels within. While keeping all observed TCs in the cluster functionally homogeneous, the SNR of the average TC is significantly improved (Figure 5.1). Therefore, the average TC can be considered as a denoised version of every observed TC in the cluster.

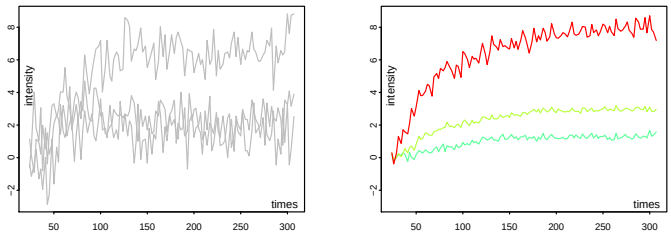
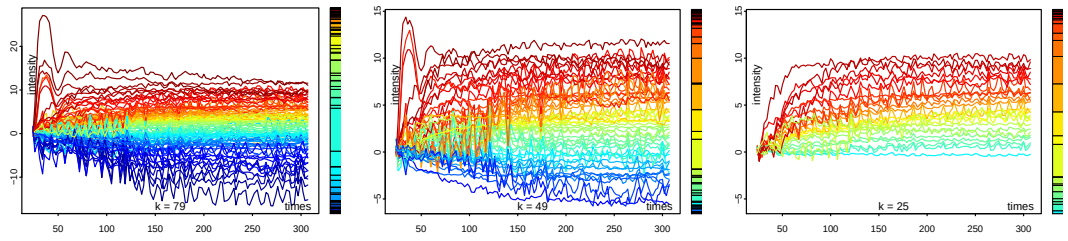


Figure 5.1: The improvement of SNR induced by HiSET. Left: Enhancement TCs of three randomly selected voxels. Right: Average enhancement TCs of clusters, to which the voxels belong, segmented by HiSET.



(a) (b) (c)

Figure 5.2: Average enhancement TCs of clusters segmented by HiSET in a DCE image sequence. (a) ones of clusters in the entire DCE image sequence; (b) ones of clusters intersecting with manual ROI; (c) ones remained after filtration. Colors are assigned to clusters according to the AUC of the corresponding average TC. Red relates to the large AUCs and blue to the low ones. k stands for the total number of clusters. The size of each cluster (scaled by the size of image) is given as the length of its color bar on the color scale on the right.

In order to get ride of clusters of voxels having aberrant behavior after injection, mostly due to motion artifacts, and adding noise to the analysis, a set of temporal

filters have been used (Figure 5.3). Since after contrast injection the tissues can not have a negative enhancement as compare to the pre-injection baseline (as opposed to Dynamic Susceptibility Contrast MR perfusion where the first pass of contrast agent induces a signal drop), negative enhancing clusters have been filtered. Since after the first pass where rapid rise and drop in signal can be observed, the signal changes are very slow, clusters with bumps, spikes of pits (with signal going down and then up again) have been filtered. Additionally, clusters of very small size (less than 5 voxels within) have also been filtered, as well as the ones with typical arterial behavior.

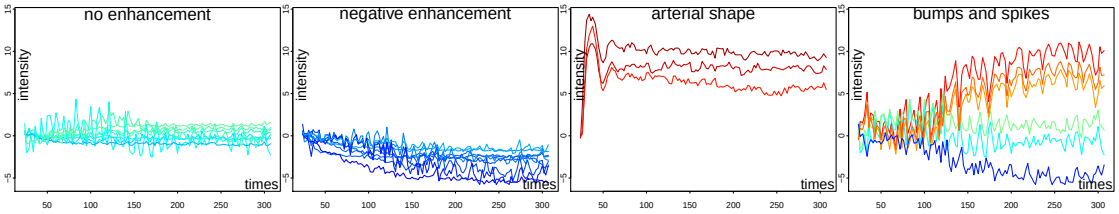


Figure 5.3: The filters remove from Figure 5.2 the average TCs showing (a) no enhancement, (b) negative enhancement, (c) arterial behavior, and (d) bumps or spikes, where (b) and (d) being due to motion artifacts.

After the filtration, only the average TCs having meaningful enhancement are automatically preserved for further qualitative analysis. For some studies, especially the ones with large manual ROI (for instance in Figure 5.2), the number of average TCs is significantly reduced, so is the amount of artifacts.

5.3 Automatic refinement of ROI

Given a roughly drawn manual ROI on a DCE-MR image sequence, following processes are done to obtain an automatically refined ROI:

1. segment the full image sequence into local clusters using HiSET (only local clustering step) with the parameter setting;
2. erode the ROI by two voxels in both horizontal and vertical directions, to get eroded ROI;
3. retrieve all the local clusters intersecting the eroded ROI and fuse these clusters to have the initial refined ROI;
4. erode again the initial refined ROI by one voxel in both directions, to get refined-eroded ROI;
5. identify disconnected segments of refined-eroded ROI induced by erosion and discard the small ones (pruning);
6. dilate the remaining segments of refined-eroded ROI by one voxel in both directions, to get final refined ROI.

Given a DCE image sequence and a manual ROI, the evolution of the ROI along the process pipeline is illustrated in Figure 5.4. Within the refined ROI, the filtration of average TCs is illustrated in Figure 5.5.

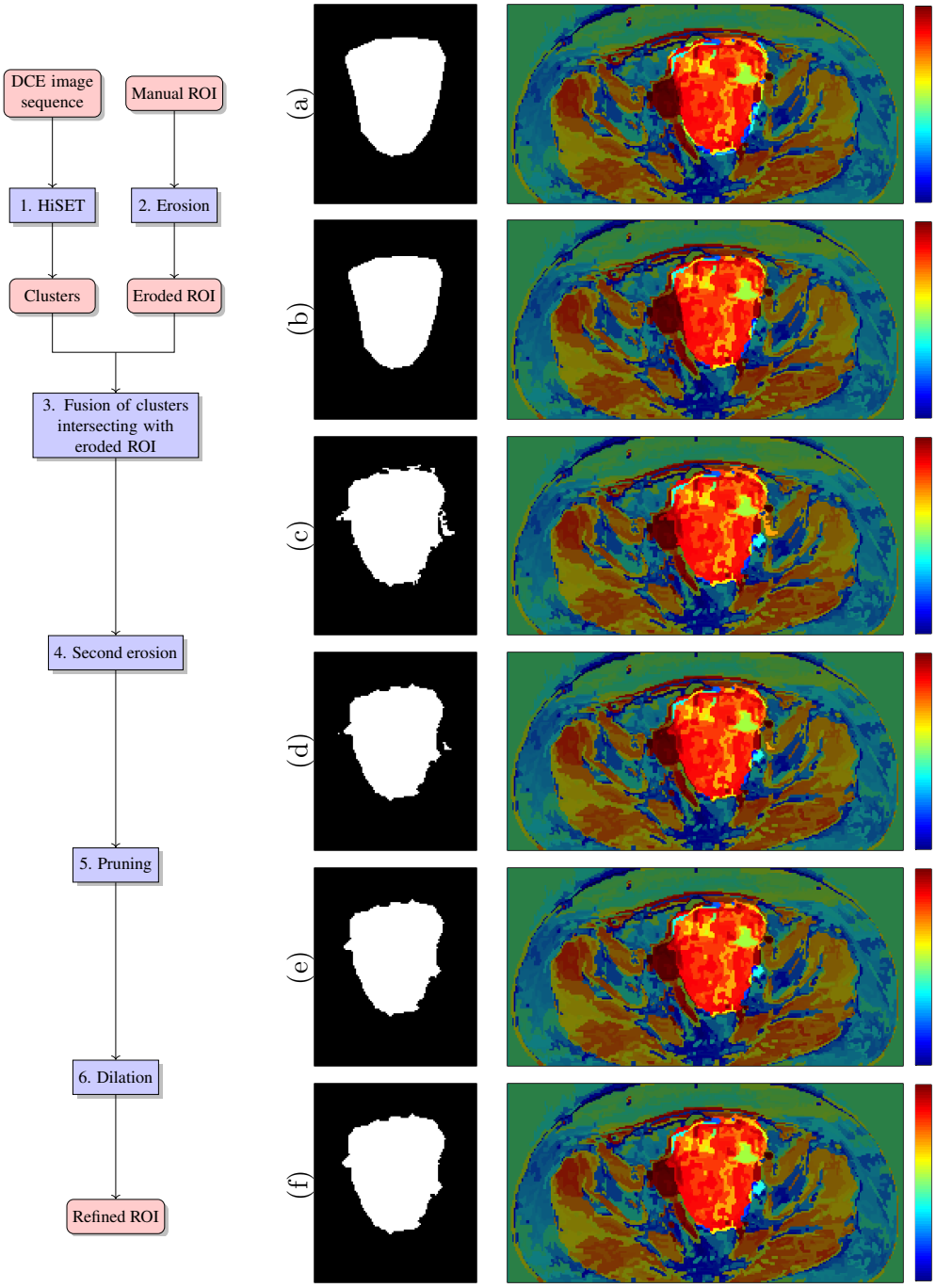


Figure 5.4: The evolution of ROI during refinement process. From top to bottom: (a) manually delineated ROI by locating vertices of a polygon; (b) erosion of manual ROI by two voxels; (c) aggregate all clusters resulting from the segmentation by HiSET and intersecting eroded ROI; (d) a second erosion of previous ROI; (e) discard the small segments disconnected with the main one of previous ROI; (f) dilation of previous ROI to have final refined ROI. Left: binary mask of each ROI. Right: transparent mask of each ROI on color map of segmentation result. Colors are assigned to clusters according to the AUC of the corresponding average TC. Red relates to the large AUCs and blue to the low ones.

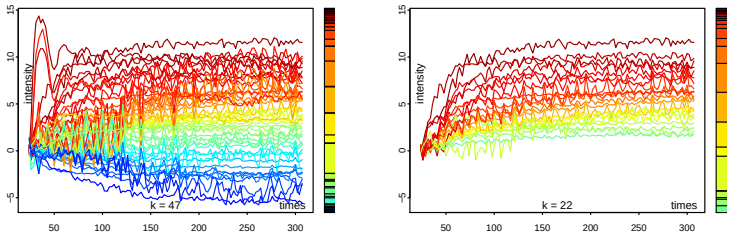


Figure 5.5: Filtration of average TCs within the refined ROI. Left: all average TCs; Right: filtered average TCs. Colors are consistent to Figure 5.2. k stands for the number of average TCs. The size of each cluster is given as the length of its color on the color bar on the right.

5.4 Test of the automatic ROI refinement on ovarian tumors

To test the effectiveness of auto-refinement of ROI on DCE-imaging analysis, we investigate on clinical data whether the classification of the ovarian tumors of our sequences will be improved from crude manual ROI to refined ROI. We tested this strategy taking into account that Thomassin et al. have previously shown that analyzing DCE TCs of ovarian tumors can help in differentiating benign, borderline and malignant tumors [Thomassin-Naggara et al., 2008, 2011].

5.4.1 Extraction of features from average TCs

For the automatic refinement of ROI, segmentation has to be implemented with time intensity curves to distinguish different tissues having the same enhancement but different baselines. To study the vascular properties of tissue, however, we are only interested in the enhancement induced by contrast agent. Therefore, a second segmentation based on time enhancement curves is required for perfusion analysis, which is based on the qualitative characteristics extracted from the resulting average TCs.

For every average TC, the qualitative characteristics that we consider to perform and automatic classification include the *maximum enhancement* (ME) and

enhancement amplitude (EA) normalized by the peak of AIF (which is identified as the average TC with highest enhancement in Figure 5.2-a), and the angle of slope during the early (wash-in) or the late (washout) phase (Figure 5.6-a). To obtain the measurements of these parameters, we smooth the average TC, $\bar{I}^C$, of each cluster C with local polynomial regression fitting [Cleveland et al., 1991] to have $\tilde{I}^C$. ME is naturally defined as the normalization of the largest enhancement of the smoothed TC during the entire acquisition time

$$\mathrm{ME}^C := \max_j \tilde{I}^C(t_j) \Big/ \max_j \mathrm{AIF}(t_j).$$

Then, we fit a piecewise linear model (one piece for the early phase and the other for the late one) on the average TC

$$\hat{I}^C(t_j) := \begin{cases} a_1^C + b_1^C * t_j & \text{if } t_j \leq t_{j_0}, \\ a_2^C + b_2^C * t_j & \text{if } t_j > t_{j_0}, \end{cases}$$

where t_{j_0} is the end of the time interval for the first linear model. The interception of two phases is then defined as the moment where the piecewise linear model has the least square error

$$\hat{t}_{j_0}^C := \arg\min_{t_{j_0}} \sum_j (\bar{I}^C(t_j) - \hat{I}^C(t_j))^2.$$

EA is then defined as the normalization of the enhancement at the interception on the smooth TC

$$\mathrm{EA}^C := \tilde{I}^C(\hat{t}_{t_0}).$$

The slope angle during the early phase, SA_e is defined as the angle in between the fitted line of this phase and the horizontal axis

$$\mathrm{SA}_e^C := \arctan(b_1^C) * 180/\pi,$$

and the slope angle during the late phase, SA_l, is defined as the angle in between the fitted line of this phase and the vertical axis

$$\mathrm{SA}_l^C := \arctan(b_2^C) * 180/\pi + 90.$$

For the TCs corresponding to malignant tissue, the enhancement is expected to arrive more quickly with higher amplitude, hence the slope angle during the early phase is expected to be larger. The slope during the late phase is downside, thereby larger than 90 degree. For the TCs corresponding to benign tissue, the enhancement is expected to arrive rather slowly with lower amplitude, hence the slope angle during the early phase is expected to be smaller. The slope during the late phase is upside, thereby smaller than 90 degree. Unifying the contrary effects of the slope angles during the early and the late phase, we define the *composite angle* (CA) as the angle between two slopes

$$\mathrm{CA}^C := 90 - \mathrm{SA}_e^C + \mathrm{SA}_l^C = (\arctan(b_2^C) - \arctan(b_1^C)) * 180/\pi,$$

such that it is larger for normal tissue and smaller for tumorous tissue.

In an image sequence, the (manual or refined) ROI usually covers several tissues having heterogeneous enhancement behaviors. And the number of tissues is usually varying from sequence to sequence. Therefore, we consider the statistics of the qualitative characteristics of average TCs as the representative features of ROI for further classification. For each ROI, we consider four features: the weighted (by the size of cluster) mean of, respectively, ME (wmME), EA (wmEA) and CA (wmCA),

$$\mathrm{wmME} := \frac{\sum_C N^C \mathrm{ME}^C}{N^{roi}}, \quad \mathrm{wmEA} := \frac{\sum_C N^C \mathrm{EA}^C}{N^{roi}}, \quad \mathrm{wmCA} := \frac{\sum_C N^C \mathrm{CA}^C}{N^{roi}},$$

where N^{roi} is the number of voxels in ROI and $N^C = |C \cap N^{roi}|$; as well as the span of smoothed average TCs during the late phase (SP),

$$\mathrm{SP} = \max_C \tilde{I}^C(t_n) - \min_C \tilde{I}^C(t_n),$$

which can be regarded as a measure of heterogeneity (Figure 5.6-b).

5.4.2 Classification of the ovarian tumors

Finally, we apply, on these extracted features from all 99 DCE-MRI sequences, both unsupervised and supervised classification methods, including k-means [Hartigan

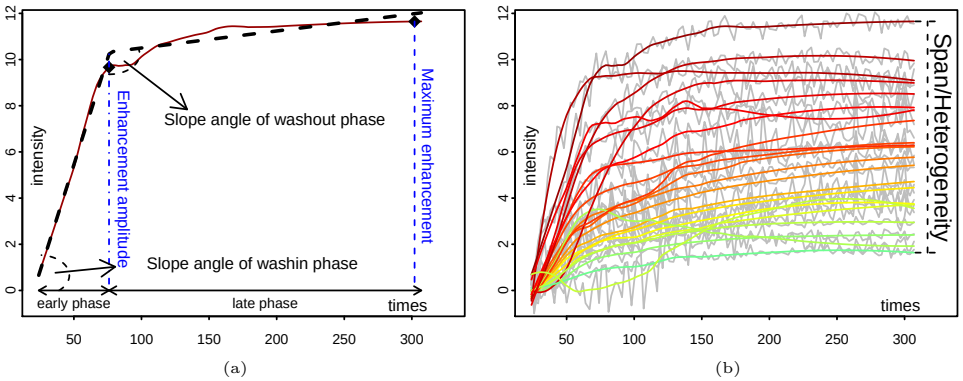


Figure 5.6: Qualitative features extracted from average TCs include (a) maximum enhancement (ME), enhancement amplitude (EA), and the slope angle during early and late phase made on one smoothed TC, and (b) Span/heterogeneity (SP) at the end of late phase based on the smoothed version (colored) of the average TCs (gray).

and Wong, 1979], linear discriminative analysis (LDA) [Ripley, 1996] and support vector machine (SVM) [Hsu and Lin, 2002] in order to separate the different types of ovarian tumors. For all three methods, we perform a binary classification (potentially benign vs. potentially malignant). Additionally, we also classify all cases into three classes (potentially benign, borderline and malignant) with k-means and SVM. Therefore, we investigate in total five classification models: k-means with 2 classes, k-means with 3 classes, LDA with 2 classes, SVM with 2 classes and SVM with 3 classes. For each model, the results from three cases: 1) manual ROI, 2) manual ROI with filtration of average TCs and 3) refined ROI with filtration, are compared with each other in term of classification interpretation and performance.

5.4.3 Quality evaluation of classification

For binary classification, we count *true positive* (TP) for correctly classified benign tumors and *true negative* (TN) for correctly classified malignant ones. Based on TP and TN, *receiver operating characteristic* (ROC) curve is reported to illustrate the specificity and sensitivity of classification. Naturally, the area under the ROC curve (ROC_{auc}) is considered as a measure of classification performance with its

confidence interval (ROC_{ci})

Alternatively, we define another measure, the *classification accuracy* (ACC), as the percentage of correctly classified tumors, with the sum of TP and TN as numerator and by the size of population classified as dominator. Unlike ROC_{auc}, ACC can be generalized to the case of classification with three classes, where the numerator is the amount of correctly classified tumors for all classes.

To evaluate the classification method, we also perform the leave-one-out (loo) cross-validation. Again, ACC, ROC_{auc} and ROC_{ci} are reported as, respectively, ACC^{loo}, ROC^{loo}_{auc} and ROC^{loo}_{ci} for leave-one-out cross-validation.

5.5 Results

5.5.1 Auto-refinement of ROI

In the implementation of the proposed auto-refinement strategy, the full segmentation of DCE image sequence using HiSET takes the most of the processing time, which is about 45 s in our experiments. This time is only slightly affected by the parameter settings, but, significantly affected by the size of image grid, or of the sub-grid (containing the manual ROI with enough margin for each side) chosen to be segmented. In this work, HiSET is directly implemented on the entire image grid for all DCE image sequences, with the same parameter setting: $a = 0.45$, $\delta = 3$ and $\alpha = 0.0001$. Other steps of the strategy take less than 1 s all combined, which is negligible with respect to HiSET.

As an example, Figure 5.7 shows the results of ROI auto-refinement for 6 DCE image sequences, which are selected to have various size of corresponding manual ROIs. Both manual ROI and refined ROI are shown in a binary way such that one can observe the difference between two at the boundary. A transparent mask of refined ROI is also shown on top of the color map of the segmentation such that the heterogeneity within the ROI is highlighted.

As a reflection of the heterogeneity within the tissue covered by ROI, the number of clusters resulted from the segmentation and intersected with the ROI is investigated in Figure 5.8. From the 17 DCE image sequences in Figure 5.8, one can observe that the manual ROI intersects a relatively large amount of clusters having different enhancement behaviors. However, from one third to half of them correspond to non-anatomical features, movements, or arteries, which are considered to be artifacts for quantitative analysis of perfusion behaviors of tissue of interest. Therefore, they are excluded through the filtration of average TCs. Additionally, for some cases, the auto-refinement also removes a small amount (relative to filtration) of clusters that intersect the manual ROI with small overlaps at the boundary. These clusters are considered to be included by mistake during

the manual delineation.

Through the filtration for all 99 DCE image sequences, the number of the average TCs within the manual ROI is reduced from 25.7 in average (ranging from 7 to 66) to 14.2 in average (ranging from 3 to 44), leading to a 44.5% decrease. After the auto-refinement, this number is further reduced to 11.8 in average (ranging from 2 to 39), which represents a 9.7% decrease additional to the filtration.

5.5.2 Classification of the ovarian tumors using ROIs

Figure 5.9 and Figure 5.11 illustrate the classification results of the ovarian tumors covered by ROIs in three cases using all five classification models based on two features introduced above: wmME and wmCA.

For binary classification models (Figure 5.9), only potential benign and malignant tumors are considered, leading to a population of size 90. In the manual ROI, even after filtration, benign and malignant tumors are heavily mixed, especially for the ones having large wmCA and small wmME (left-top corner of Figure 5.9-a,b), leading to a tough separation. This phenomenon is significantly improved in refined ROI (Figure 5.9-c), such that much fewer tumors are classified into the wrong decision area.For refined ROI with filtration, the separations between two classes identified by three methods align much better than the ones from the other two types of ROI.

The performance of binary classification using each method for three ROIs is also compared in term of ROC (Figure 5.10), illustrating a significant improvement induced by the refinement of ROI. Figure 5.12 illustrates another comparison of classification performance, in term of ACC, of the bivariate classification introduced above, along with a trivariate one based on three features: wmEA, wmCA and SP. From both bivariate and trivariate classifications, one can observe significant improvements of ACC induced by the filtration of average TCs and the ROI refinement. Complete result of classification performance in term of different metrics, for different models, combinations of features and ROIs are summarized in

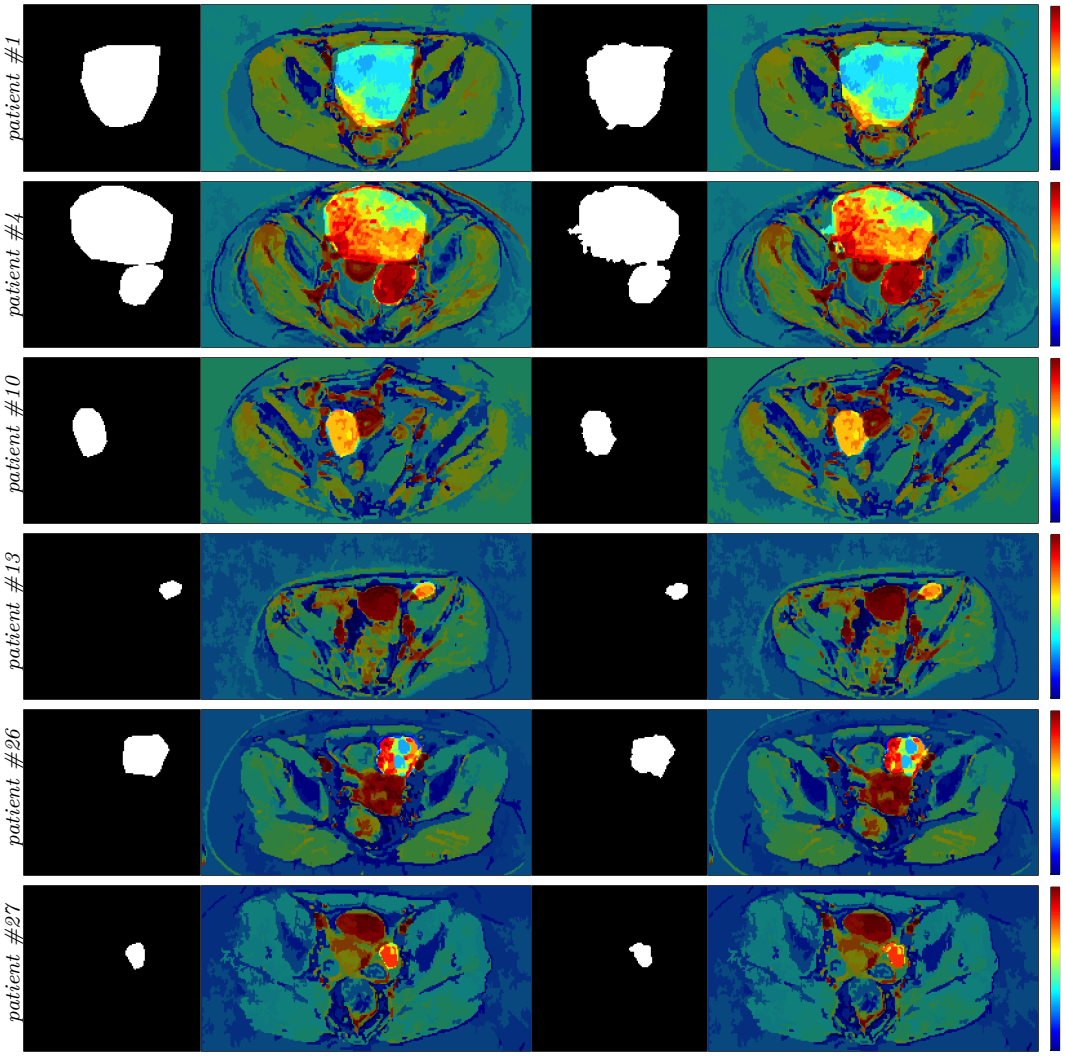


Figure 5.7: The results of auto-refinement of ROIs, for 6 DCE image sequence, with various size. Each row shows the results for one DCE image sequence and its patient number is given at the beginning of the row. From left to right, (a) cropped binary masks of manual ROIs; (b) transparent masks on color maps of segmentation results of manual ROIs; (c) cropped binary masks of refined ROIs; (d) transparent masks on color maps of segmentation results of refined ROIs. Colors are assigned to clusters according to the AUC of the corresponding average TC. Red relates to the large AUCs and blue to the low ones.

Table 5.1 and Table 5.2. The improvement of the classification performance induced by ROI refinement is consistent in term of different metrics and classification models.

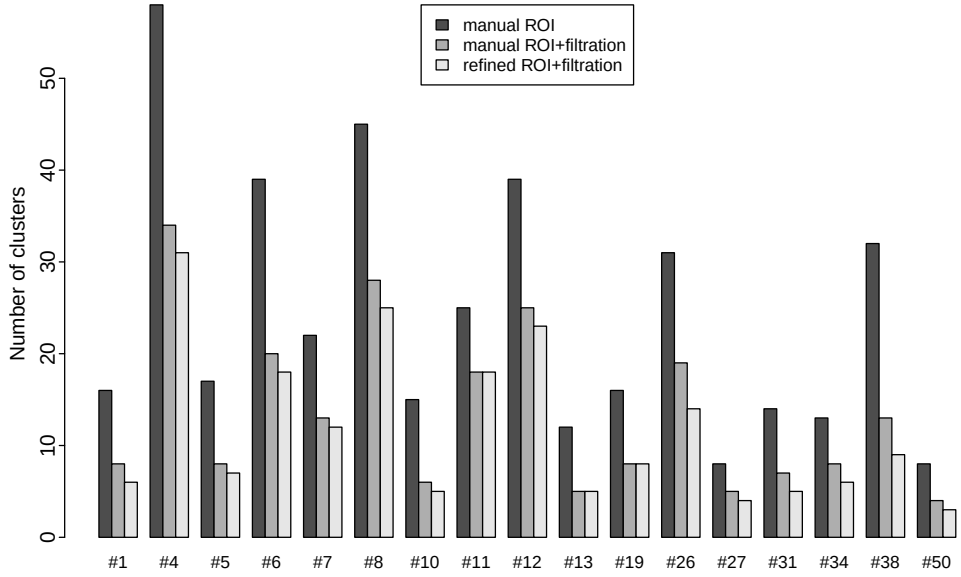


Figure 5.8: Number of clusters resulted from manual ROI, manual ROI with filtration of average TCs and refined ROI with filtration, in 17 DCE image sequences.

Among three classification methods, SVM outperforms the other two. In binary case of SVM, ACC is improved from 71.1% for manual ROIs to 78.9% by filtration, and then to 84.4% by ROI refinement. Similarly, ROC_{auc} is improved from 0.78 (95% ROC_{ci} = 0.69-0.88) for manual ROIs to 0.94 (95% ROC_{ci} = 0.89-0.99) for refined ROI. Additionally, in case of three classes, ACC is improved from 67.7% for manual ROIs to 72.7% by filtration, and then to 79.8% by ROI refinement.

5.6 Discussion

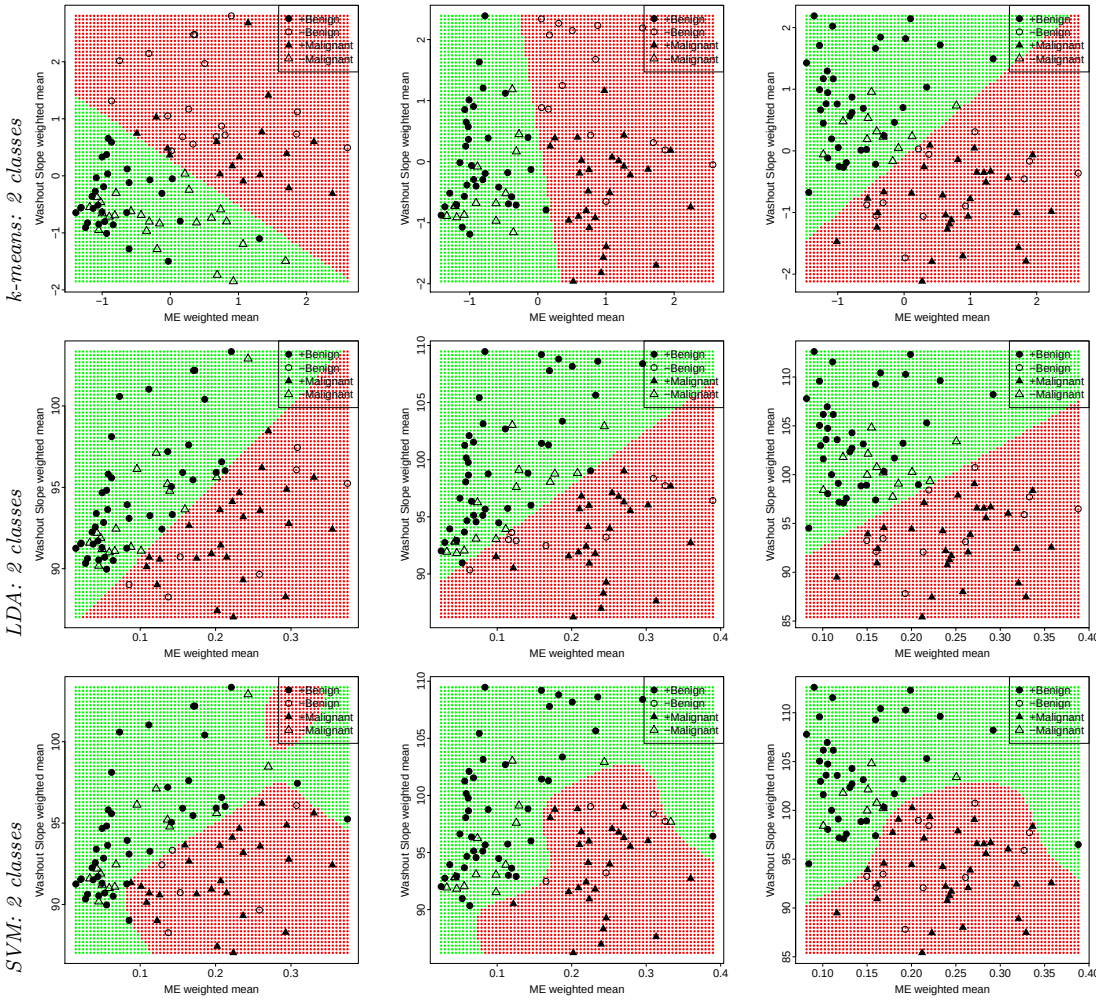
The proposed strategy aims to automatically refine the manually and roughly delineated ROI, based on an automatic segmentation of DCE image sequence using HiSET, to ease the process of ROI selection in a robust way. In our experiments, the auto-refinement takes about 45 s per DCE image sequence, most of which is cost by the HiSET while other steps of the strategy take less than 1 s all combined. The segmentation time is significantly affected by the size of image grid. Therefore,

Model	ROI type	ACC(%)	ACC^{loo}(%)	ROC_{auc}	ROC_{ci}	ROC^{loo}	ROC^{loo}_{ci}
k-means_2	Manual ROI	64.4	64.4	0.69	0.58-0.8	0.69	0.58-0.8
	Manual ROI + Filtration	67.8	67.8	0.72	0.61-0.83	0.72	0.61-0.83
	Refined ROI + Filtration	75.6	75.6	0.82	0.73-0.9	0.82	0.73-0.9
LDA_2	Manual ROI	66.7	66.7	0.73	0.63-0.84	0.71	0.6-0.82
	Manual ROI + Filtration	67.8	67.8	0.74	0.63-0.84	0.71	0.59-0.82
	Refined ROI + Filtration	77.8	76.7	0.83	0.75-0.92	0.81	0.72-0.91
SVM_2	Manual ROI	**67.8**	72.2	**0.78**	0.68-0.88	0.76	0.66-0.86
	Manual ROI + Filtration	**71.1**	72.2	**0.75**	0.64-0.86	0.8	0.71-0.9
	Refined ROI + Filtration	**82.2**	82.2	**0.92**	0.86-0.98	0.85	0.76-0.94
k-means_3	Manual ROI	45.5	45.5				
	Manual ROI + Filtration	50.5	50.5				
	Refined ROI + Filtration	53.5	53.5				
SVM_3	Manual ROI	62.6	71.7				
	Manual ROI + Filtration	65.7	70.7				
	Refined ROI + Filtration	72.7	71.7				

Table 5.1: Performance of classifications based on two features: wmME and wmCA.

Model	ROI type	ACC(%)	ACC^{loo}(%)	ROC_{auc}	ROC_{ci}	ROC^{loo}	ROC^{loo}_{ci}
k-means_2	Manual ROI	72.2	71.1	0.73	0.62-0.83	0.73	0.62-0.83
	Manual ROI + Filtration	73.3	73.3	0.75	0.65-0.86	0.75	0.65-0.86
	Refined ROI + Filtration	75.6	75.6	0.8	0.7-0.89	0.8	0.7-0.89
LDA_2	Manual ROI	71.1	65.6	0.74	0.64-0.85	0.7	0.59-0.82
	Manual ROI + Filtration	74.4	74.4	0.78	0.69-0.88	0.75	0.65-0.86
	Refined ROI + Filtration	77.8	76.7	0.86	0.79-0.94	0.83	0.75-0.92
SVM_2	Manual ROI	**71.1**	77.8	**0.78**	0.69-0.88	0.88	0.81-0.95
	Manual ROI + Filtration	**78.9**	82.2	**0.82**	0.73-0.92	0.84	0.75-0.93
	Refined ROI + Filtration	**84.4**	82.2	**0.94**	0.89-0.99	0.93	0.88-0.98
k-means_2	Manual ROI	49.5	48.5				
	Manual ROI + Filtration	51.5	51.5				
	Refined ROI + Filtration	52.5	51.5				
SVM_3	Manual ROI	67.7	67.7				
	Manual ROI + Filtration	72.7	74.7				
	Refined ROI + Filtration	79.8	73.7				

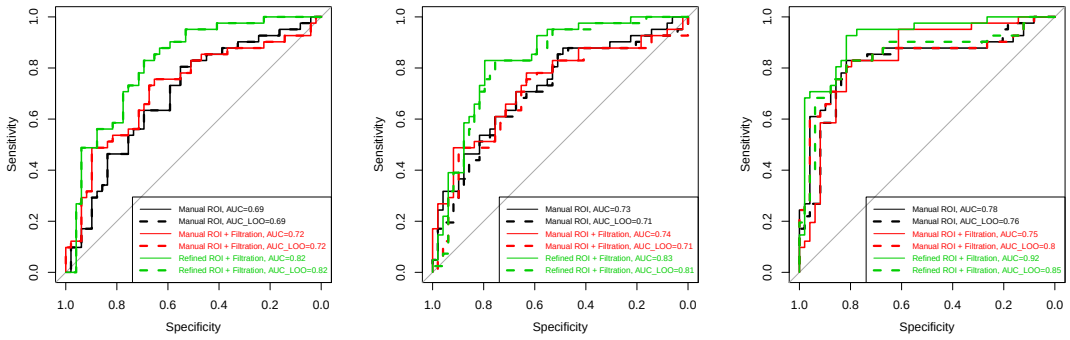
Table 5.2: Performance of classifications based on three features: wmEA, wmCA and SP.



(a) Manuel ROI (b) Manuel ROI + Filteration (c) Refined ROI + Filteration

Figure 5.9: Binary classifications of ovarian tumor (potentially benign and malignant) based on two features: wmME and wmCA. From left to right, classification is based on (a) manual ROI, (b) manual ROI with filtration of average TCs and (c) refined ROI with filtration. From top to bottom, three classification method are used: k-means, LDA and SVM. Each type of tumor is represented with a shape, which is filled if the tumor is correctly classified (with '+' in legend) by the model, otherwise ('-') is left unfilled. Decision area of each class is covered by dots in green for benign, blue for borderline and red for malignant.

it is natural to segment a sub-grid containing the manual ROI with enough margin on each side, in order to reduce the processing time. However, a direct segmentation on the entire image grid is more convenient and the segmentation result can be

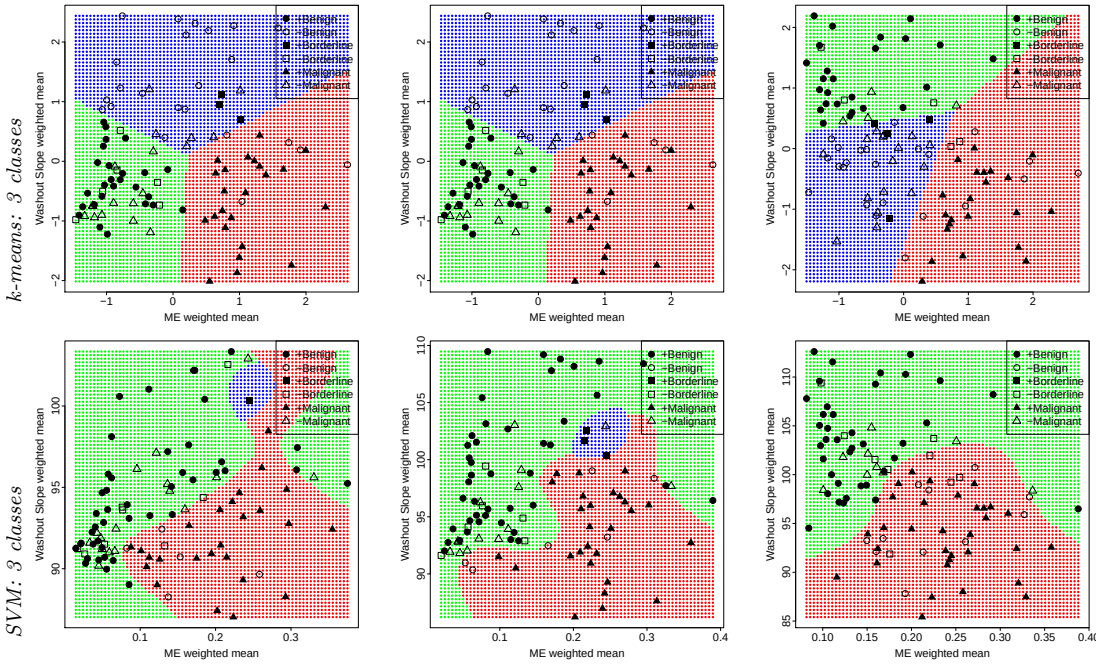


(a) k-means (b) LDA (c) SVM

Figure 5.10: Performance of binary classification of ovarian tumor (potentially benign and malignant), based on two features: wmME and wmCA. From left to right, three classification methods are used: (a) k-means, (b) LDA and (c) SVM. For each method, ROCs of both the classification (solid thinner line) and the leave-one-out cross-validation (dashed thicker line) are shown for three ROIs (black for manual ROI, red for manual ROI with filtration and green for refined ROI with filtration). Specially, for k-means, the ROC of the classification and the one of cross-validation are overlapped for each ROI.

always reused for other manual ROIs concerning other tissues in the same DCE image sequence.

The number of clusters resulted from the segmentation and intersected with the ROI corresponds to the heterogeneity within the tissue covered by the ROI. In average, nearly half of the clusters in the manual ROI correspond to non-anatomical features, movements, or arteries, which are considered to be artifacts for quantitative analysis of perfusion behaviors of tissue of interest. Therefore, they are excluded through the filtration of average TCs. Additionally, the auto-refinement also removes a small amount (relative to filtration) of clusters that intersect the manual ROI with small overlaps at the boundary. These clusters are considered to be included by mistake during the manual delineation, mostly due to drawing errors or even more unmanageably due to displacement of the delineated area due to motion effects during the imaging time. Despite that the most of reduction on the number of clusters comes from the filtration, the ROI refinement plays a more

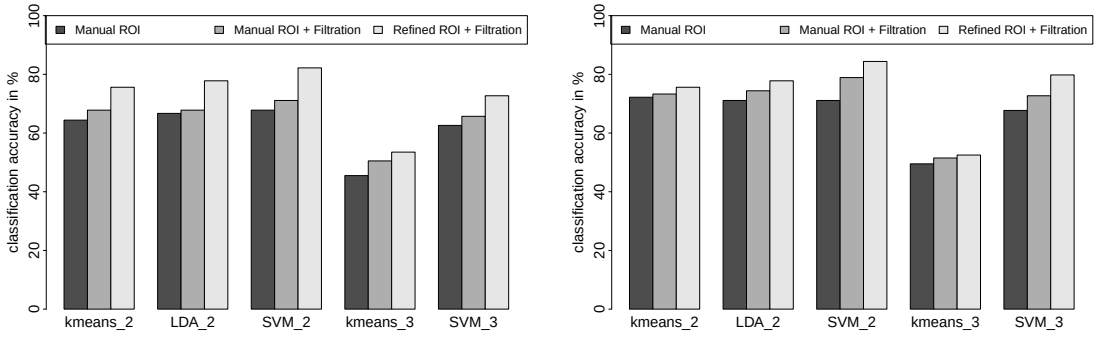


(a) Manuel ROI (b) Manuel ROI + Filteration (c) Refined ROI + Filteration

Figure 5.11: Classifications of ovarian tumor with three classes (benign, borderline and malignant), based on two features: wmME and wmCA. From left to right, classification is based on (a) manual ROI, (b) manual ROI with filtration of average TCs and (c) refined ROI with filtration. Two classification methods are used: k-means (top) and SVM (bottom). Each type of tumor is represented with a shape, which is filled if the tumor is correctly classified (with '+' in legend) by the model, otherwise ('-') is left unfilled. Decision area of each class is covered by dots in green for benign, blue for borderline and red for malignant.

important role for improving the tumor classification.

Conventionally, the evaluation of (semi-)automatic ROI construction strategy based on image segmentation usually involves a comparison against the ground-truth, which is a ROI delineated manually and carefully by experienced radiologist [Liu et al., 2014; Stoutjesdijk et al., 2007]. In this way, a quantitative index such as DICE [Taha and Hanbury, 2015] can be used to measure the robustness and the effectiveness of the strategy. However, in this work, the proposed strategy takes a roughly delineated ROI as an input and aims to refine it only at the boundary,



(a) wmME + wmCA

(b) wmEA + wmCA + SP

Figure 5.12: The classification accuracy (ACC) in percentage. The classifications are based on (a) two features: wmME and wmCA; (b) three features: wmEA, wmCA and SP. Each group of bars corresponds to one of five classification models, including k-means with 2 classes, k-means with 3 classes, LDA with 2 classes, SVM with 2 classes and SVM with 3 classes. In each group, classifications are made individually on manual ROI, manual ROI with filtration of average TCs and refined ROI with filtration.

in order to include the homogeneous regions extended from the center of ROI to outside and also to exclude the regions corresponding to movements and features not interested. Therefore, the resulting conventional measures will only vary in a small range corresponding to the inclusion and the exclusion of the voxels only around the boundary, while the center part of ROI always remains unchanged, especially for the large ROIs. Alternatively, we evaluate the effectiveness of the proposed strategy on DCE-imaging analysis through the classification of the ovarian tumors, as in [Ortiz et al., 2014; Górriza et al., 2009; Liu et al., 2014; Forbes et al., 2006].

Using different combinations of features and models, the classification performance is significantly improved from manual ROI to refined ROI. Using the same method, such as k-means or SVM, the binary classification performs better than the one with 3 classes. Being the intermediate state between benign and malignant, borderline could be classified into either of other two classes in practice, leading to a source of uncertainty of classification with 3 classes. Additionally, due to the

unbalance between the size of borderline classes and others, k-means with 3 classes performs significantly worse than with only 2 classes since it tends to identify clusters with compatible sizes. Moreover, the supervised methods such as LDA and SVM should be carefully tuned to avoid underfitting or overfitting the model. In brief, the presence of borderline tumors and the lack of more training/testing data are the limitations for this work. Therefore, we focus more on proving that the refined ROI improves the classification in different models, instead of the model selection problem.

Chapter 6

Segmentation on 3D DCE-MR image sequence

In this chapter, we evaluate the proposed method on 3D DCE-MR image sequence, also referred as DCE volume. We explore and evaluate several segmentation strategies based on the proposed method.

6.1 Materials

We have two 3D DCE image sequences, with 5 phases for each, from two different patients. Unlike the DCE image sequences that we have used in previous experiments, each one of these two has the following characteristics:

- consists of 5 time points, one time for each phase among 5 phases:
  - Baseline phase: before the injection of contrast agent;
  - Arterial phase 1: after the injection of contrast agent and before the observed intensity within the main arterial reaches the maximum;
  - Arterial phase 2: after the observed intensity within the main arterial reaches the maximum and before the contrast agent arrives in portal vein;

  - Portal phase: during the arrival of contrast agent in portal vein;
  - Late phase: after the contrast agent passing through the portal vein.

- consists of 132 slices on a grid of 512×512 voxels for each, covering the whole abdominal area. Hence, it has three dimensions in space.
- Voxel size is 0.66 x 0.66 x 3.4 mm for the first and 0.78 x 0.78 x 3.4 mm for the second.
- Inter-slice space is 1.7mm for both.

This kind of DCE volumes, with only a few time points (one for each phase), is more common in clinical practice than the one with higher acquisition frequency, since it provides higher resolution, produces higher SNR and requires less operational time. From two DCE volumes, we select three sub-volumes (see Figure 6.1) for further experiment. The information regarding sub-volumes are detailed in Table 6.1.

Sub-volume	Patient	Target area	Size (in voxel)
# 1	Patient 1	Liver	$280 \times 210 \times 70$
# 2	Patient 1	Kidney	$100 \times 100 \times 40$
# 3	Patient 2	Upper abdomen	$400 \times 320 \times 80$

Table 6.1: Summary information of sub-volumes.

6.2 Segmentation strategies

Due to the small number of times, we use Gaussian test instead of Chi-square test, leading to a small enough probability (3.55) of having a wrong clustering. All strategies require the same transformation to stabilize the variance, which has been previously defined in Section 3.1 and verified for 2D case in Section 4.5.1. The SNR is consistent cross slices, thereby, we can choose parameter a using only one slice.

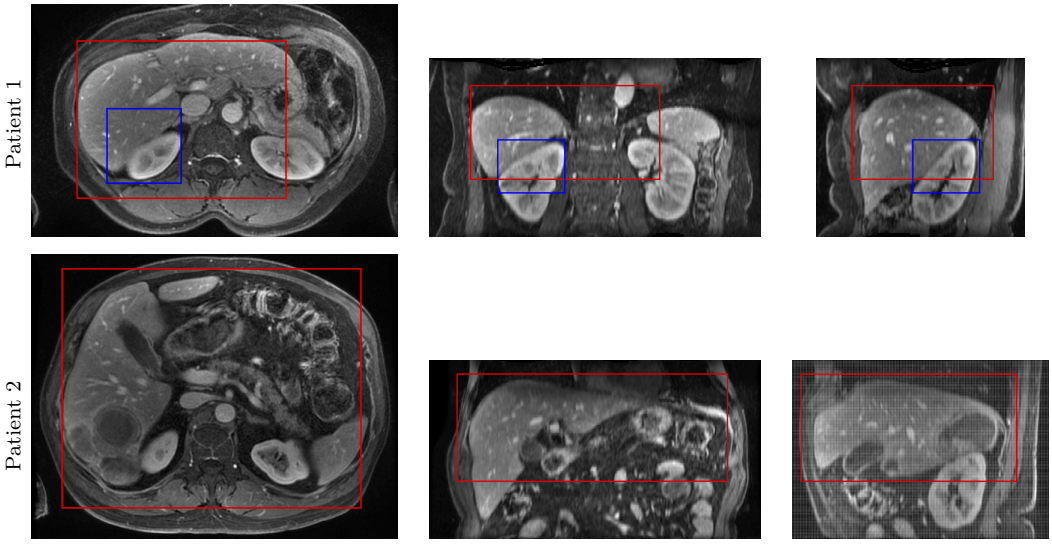


Figure 6.1: Multi-view of two DCE volumes at portal phase: axial plane (left column), coronal plane (middle column), sagittal plane (right column). For both patients, sub-volumes are selected (surrounded by color frames in three plans) for further evaluation.

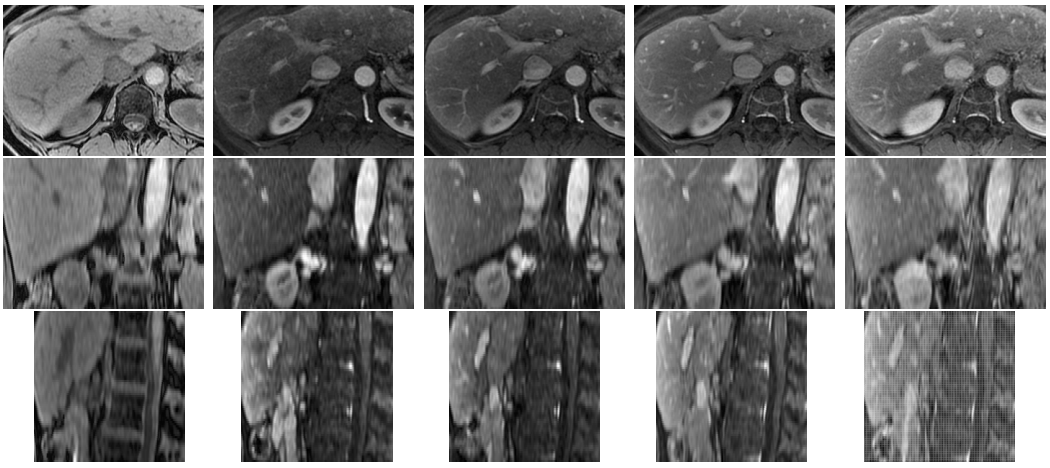


Figure 6.2: Multi-view of 3D DCE image sequence: axial plane (top row), coronal plane (middle row), sagittal plane (bottom row). From left to right, each column corresponds to a phase.

As shown in Figure 6.3, the residual densities of segmentations on different slices with the same combination of a and δ (while α is fixed) behave very similarly.

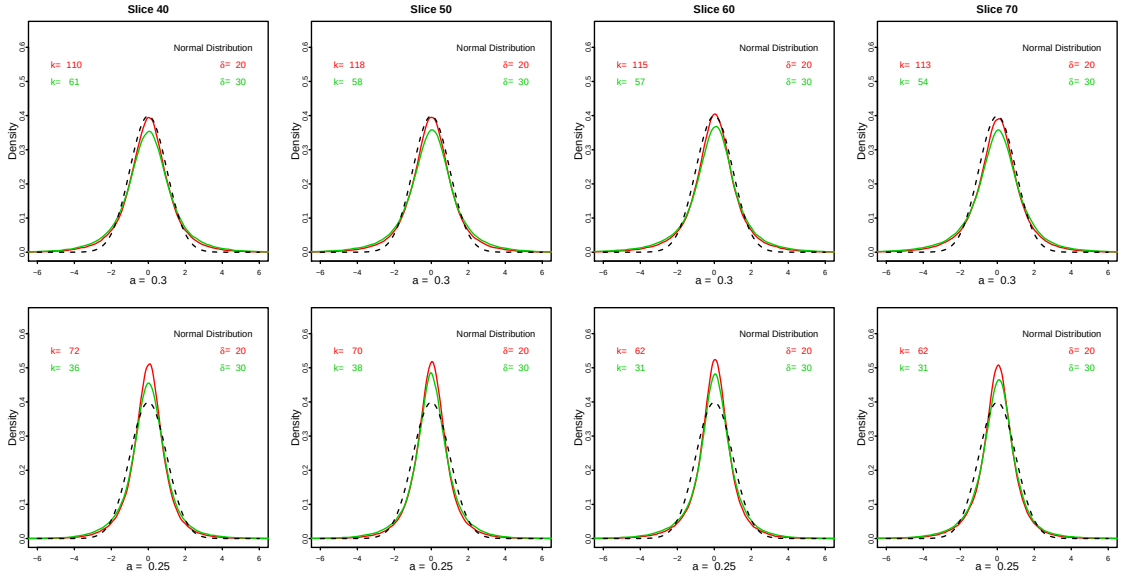


Figure 6.3: Residual densities after segmentations of different slices using the same sets of values of a and δ. Each column corresponds to one slice. The first row shows the segmentations based on a variance stabilization transform with $a = 0.3$ while the second with $a = 0.25$. For both values of a, two values of δ, 20 and 30, are used in HiSET.

6.2.1 2D-to-3D strategy

Considering that the thickness of slice is much larger than the size of voxel in image, we give priority to the neighborhood structure appearing within each slice, and then deal with the neighborhood structure appearing cross slices. In this spirit, we apply HiSET individually for every slice and then recover 3D segmentation by the propagation of all regions resulting from 2D segmentations of all slices.

The 2D-to-3D strategy consists of the following steps:

1. Run 2D HiSET for every slice. A set of clusters of voxel are obtained for each slice.

2. Define the neighborhood structure within the same slice: For one cluster, its neighbors are all the clusters which the neighboring voxels of its member voxels belong to except itself.

3. Define the neighborhood structure cross slices: two clusters coming from adjacent slices and overlapping (for at least one voxel) on the grid are neighbors of each other.

4. 3D propagation of all clusters of all slices with new neighborhood structures defined in step 2 and 3.

3D propagation of clusters in slices is essentially equivalent to HiSET that we have described in Section 3.3, except that the input is a set of clusters, instead of voxels, but still with a neighborhood structure. The same routine of local and global clustering is proceeded.

6.2.2 Direct 3D strategy

We equally treat the neighborhood structure within slice and cross slices, meaning that every voxel in volume has its adjacent voxels as its neighbors. HiSET is directly applied on all voxels available in DCE volume with a 3D neighborhood structure.

6.2.3 From intensity-based to enhancement-based segmentation

For every cluster resulting from segmentation based on intensity, the averaged intensity at the first phase is subtracted from the averaged intensity at the other phases, to have the averaged enhancements induced by the contrast agent. An additional clustering of the clusters based on their averaged enhancements can reduce the complexity of anatomical features by merging ones having same enhancement but different baselines, such as the different parts of livers. This phenomena can be observed from Figure 6.4: on the top left, the liver is split into several parts (with colors varying from dark red to orange), but on the top right, these parts are merged into only a couple of clusters (blue). In addition, two segmentations offer an hierarchical (with only two levels) representation of clusters.

In Figure 6.4, one can also observe that the clusters, from intensity-based segmentation, having the same enhancement could have quite different baselines. In practice, by simply aggregating them all as proposed above could result a mixture of different features, especially for poorly enhanced ones such as fat, muscle, water and air. Therefore, a compromise of intensity-based and enhancement-based segmentation has been taken, called intermediate segmentation. In intermediate segmentation, only the clusters having same enhancement, which is defined by enhancement-based segmentation, and close enough baselines, which is defined by a given threshold, are merged together. To be specific, for all the clusters labelled in the same vertical line in Figure 6.4-d, a hierarchical clustering based on baseline is conducted such that the clusters having their averaged baselines closer than h are considered to be one cluster in intermediate segmentation. The aggregations of clusters are highlighted by ellipses surrounding the cluster labels in Figure 6.4-d,f,h, given different values of h. The corresponding intermediate segmentation results are respectively shown in Figure 6.4-c,e,h. One can observe that, as h increases, more and more clusters from intensity-based segmentation are aggregated and the intermediate segmentation approaches to enhancement-based segmentation (i.e. red/orange regions in liver).

Enhancement-based segmentation can be incorporated into volume segmentation strategy at one of the following three stages:

- at voxel level - the enhancement is computed for every voxel prior to clustering;
- at slice level - the average enhancement is computed for every cluster resulted in every slice after slice segmentation, prior to 3D propagation;
- at volume level - the average enhancement is computed for every cluster resulted in volume after 3D propagation.

In practice, the last two stages are preferred since the enhancement is already averaged within homogeneous cluster, leading to a significant improvement of SNR with respect to individual voxels. On the other hand, intermediate segmentation

can be only combined at the last two stages since it requires intensity-based and enhancement-base enhancement.

6.2.4 Tuning of parameter δ

One can use different values of δ for each step of segmentation strategy, in order to have a trade-off between homogeneity and details of anatomical features. We discriminate them with distinct notation, as in Figure 6.6:

- δ_l: local clustering,
- δ_g: global clustering,
- δ_{3D}: 3D propagation,
- δ_{enh}: enhancement clustering.

Commonly, the following relations are respected during the parameter tuning

$$\delta_l \leq \delta_g \leq \delta_{3D} \leq \delta_{enh}.$$

For 2D-to-3D strategy, δ_l and δ_g are used (equally or not) for 2D segmentation, based on intensity, of all slices, and then δ_{3D} is used for 3D propagation of all clusters in slices. The enhancement clustering using δ_{enh} can be combined either after 2D segmentation for all slices, or after 3D propagation for the volume, in order to have clusters with homogeneous enhancement. For 3D strategy, δ_l and δ_g are used (equally or not) for 3D segmentation of the entire volume based on intensity, and then the enhancement clustering is followed if necessary by specifying δ_{3D}.

Finally, two volume segmentation strategies are summarized as the diagram shown in Figure 6.6.

6.3 Results

The 2D-to-3D strategy is implemented on DCE volume #1 and #3 while the 3D strategy is implemented on DCE volume #2. In all instances of implementation,

$a = 0.3$ and $\alpha = 0.001$ are used. Figure 6.7 includes three segmentation results, one for each volume, and the δ-parameters are set to:

- for volume #1, $\delta_l = 10$, $\delta_g = 20$, $\delta_{3D} = 20$ and $\delta_{enh} = 40$;
- for volume #2, $\delta_l = \delta_g = 20$ and $\delta_{enh} = 40$;
- for volume #3, $\delta_l = \delta_g = 20$, $\delta_{3D} = 20$ and $\delta_{enh} = 40$.

Clearly, one can observe that, from coronal and sagittal planes, the 2D-to-3D strategy is not able to always capture the regular shape of cluster crossing slices. It is due to the fact that, in this strategy, the neighborhood structure is only defined within slice but not cross slices. Therefore, the border of a cluster in one slice would not match the border in adjacent slices in a continuous fashion, leading to the appearance of "horizontal lines" in coronal and sagittal planes.

On the other hand, by defining neighborhood structure in all three spatial dimensions equally, the 3D strategy constructs clusters having regular shapes within and cross slices. However, its computational cost is more important than the alternative strategy.

As in 2D segmentation of DCE image sequence, the larger δ is, the fewer homogeneous regions are resulted, meaning that the fewer anatomical features are segmented from DCE volume.

As shown in Figure 6.9, δ_l and δ_g significantly effect the small local features (clusters) in slices (axial plan). With a larger δ_l (or δ_g) value, these features are more likely to be merged into surrounding large features, which happens, for instance, on the the vessels passing through the liver. Due to the irreversible nature of the segmentation strategy, these small features are drastically omitted and can not be recovered in the following steps. Therefore, it is highly recommended to start the strategy with a small value of δ_l, if the small features are also interested. Additionally, by doing so, the result of this step (local clustering) is always reusable as an input for further tuning of other δs, leading to a great benefit of computational time by avoiding to repeat this step that is the most time-consuming one among all steps.

On the other hand, δ_{3D} and δ_{enh} focus more on cluster propagation in volume. Most of the clusters in volume consist of only a small amount of clusters in slices, and by using larger δ_{enh} (equivalent for δ_{3D}), this amount increases slowly and uniformly within the volume, as shown in Figure 6.8. For every cluster among hundreds are segmented in volume, its sub-clusters in slices could be distributed in a scattered way within the volume made of millions of voxels and, however, have the shapes already determined in 2D segmentation step. This explains the hardly visible distinction among the segmentation results in Figure 6.10 and Figure 6.11.

6.4 Discussion

In this work, the segmentation strategies are only evaluated with three sub-volumes. To be noticed, they can surely be used directly on the entire DCE volume, in which case, one can expect an improvement in term of SNR as the amount of equivalent observations may increase in a given cluster.

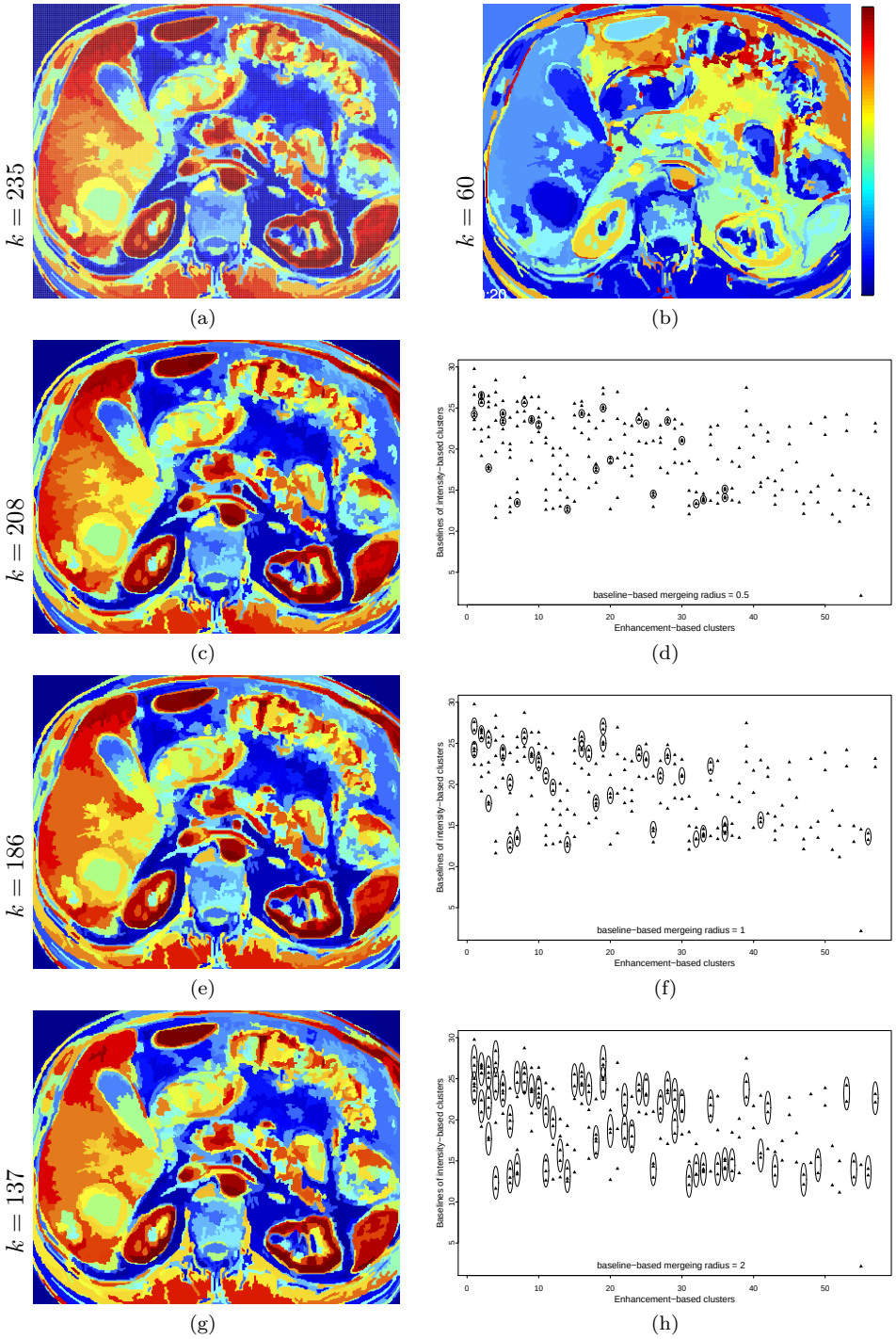


Figure 6.4: For one slice given $a = 0.3$ and $\alpha = 0.001$: (a) intensity-based segmentation using $\delta_l = \delta_g = 20$; (b) enhancement-based segmentation based on average enhancement of clusters from last segmentation using $\delta_{enh} = 40$; (c-d) intermediate segmentation using $h = 0.5$; (e-f) intermediate segmentation using $h = 1$; (h-g) intermediate segmentation using $h = 2$. In (a-c,e,h), colors are arranged according to AUC of average intensity/enhancement curves from red (high AUC) to blue (low AUC), and the number of clusters in each segmentation is given on the left. In (d,f,g), same scatter is shown with every solid triangle standing for a cluster from intensity-based segmentation and ellipses highlight the cluster aggregations in intermediate segmentation.

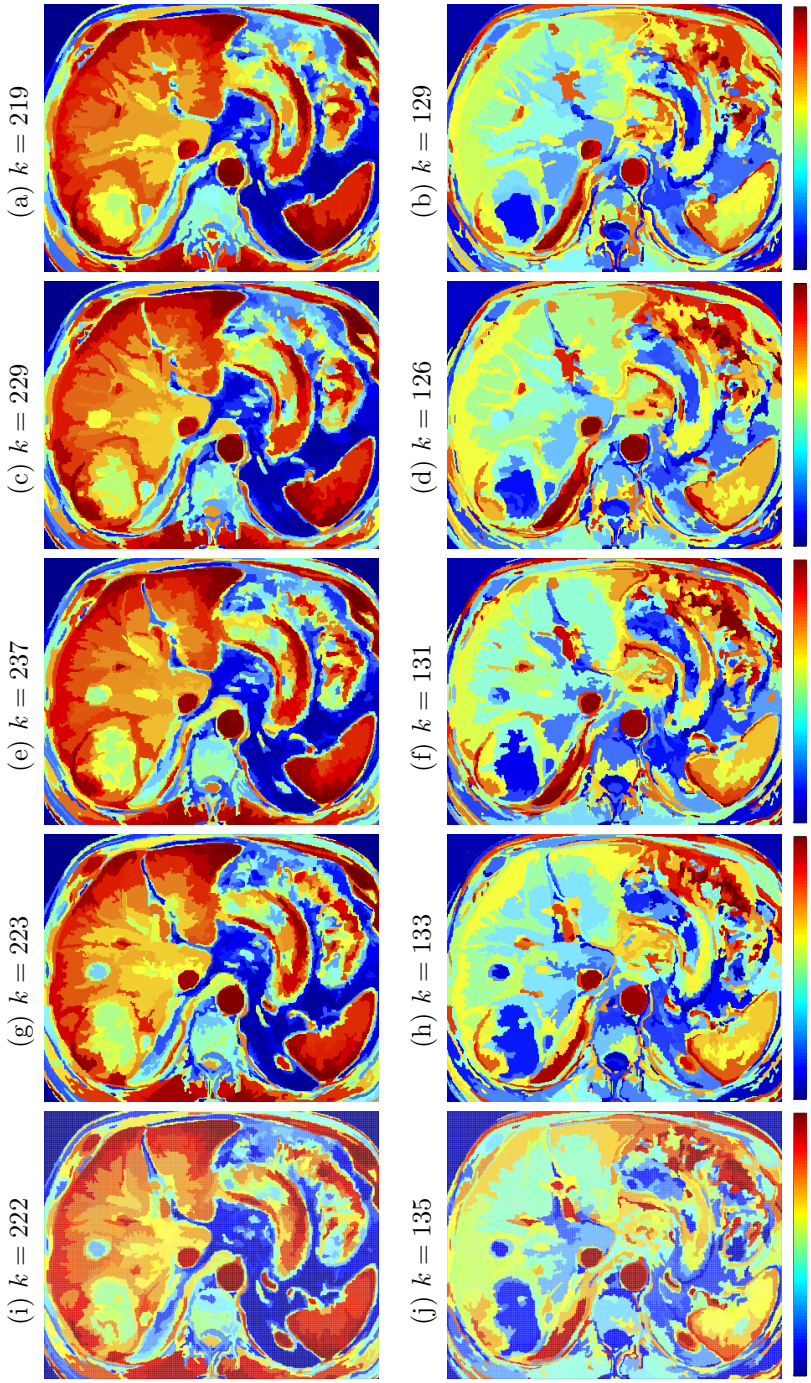


Figure 6.5: For one slice given $a = 0.3$ and $\alpha = 0.001$: (a) intensity-based segmentation using $\delta_l = \delta_g = 20$; (b) enhancement-based segmentation based on average enhancement of clusters from last segmentation using $\delta_{enh} = 40$; (c-d) intermediate segmentation using $h = 0.5$; (e-f) intermediate segmentation using $h = 1$; (h-g) intermediate segmentation using $h = 2$. In (a-c,e,h), colors are arranged according to AUC of average intensity/enhancement curves from red (high AUC) to blue (low AUC), and the number of clusters in each segmentation is given on the left. In (d,f,g), same scatter is shown with every solid triangle standing for a cluster from intensity-based segmentation and ellipses highlight the cluster aggregations in intermediate segmentation.

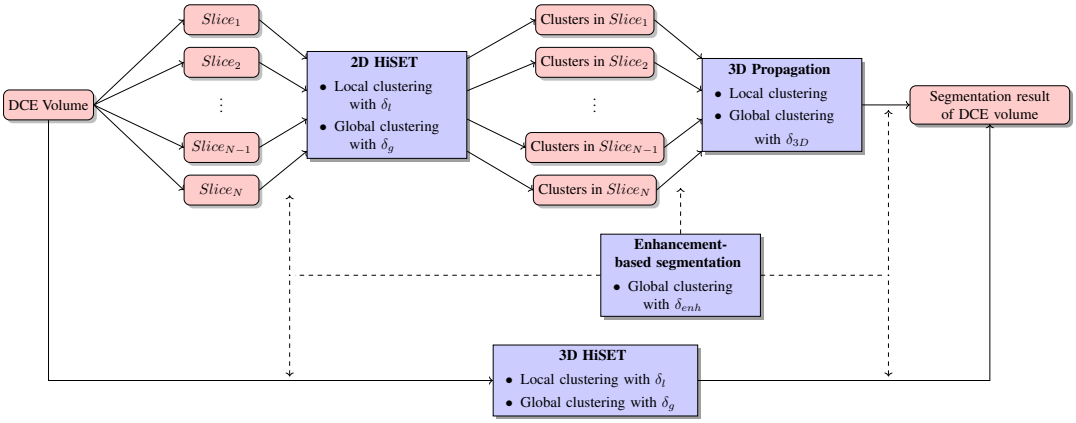


Figure 6.6: DCE volume segmentation strategies.

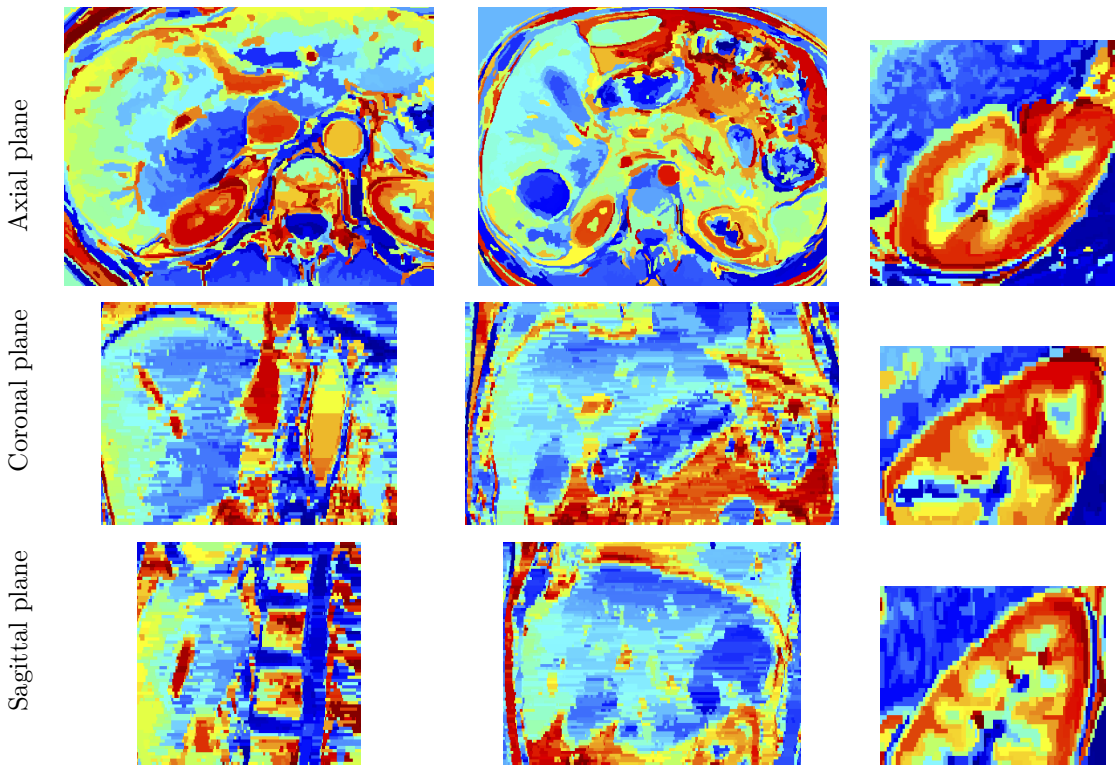


Figure 6.7: Multi-view of 3D segmentations of three DCE volumes with $a = 0.3$ and $\alpha = 0.001$. Left column: sub-volume #1, $\delta_l = 10$, $\delta_g = 20$, $\delta_{3D} = 20$ and $\delta_{enh} = 40$; Middle column: sub-volume #3, $\delta_l = \delta_g = 20$, $\delta_{3D} = 20$ and $\delta_{enh} = 40$; Right column: sub-volume #2, $\delta_l = \delta_g = 20$ and $\delta_{enh} = 40$.

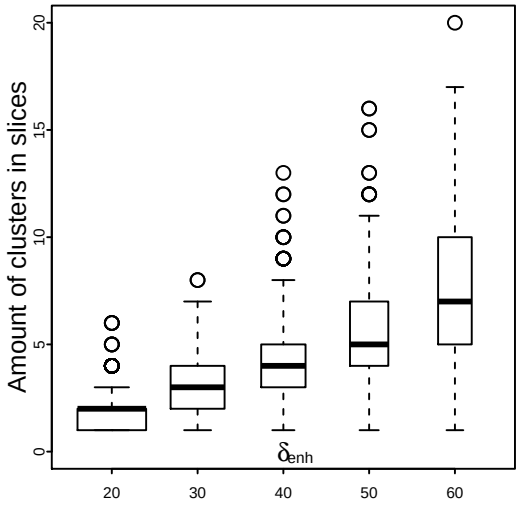


Figure 6.8: The number of clusters in slices of which every cluster in volume consists, after volume segmentation of DCE volume #1 with $a = 0.3$, $\alpha = 0.001$, $\delta_l = 10$, $\delta_g = 20$, $\delta_{3D} = 20$ and different δ_{enh} for enhancement clustering at the end. The width of box is proportional to the number of clusters in volume.

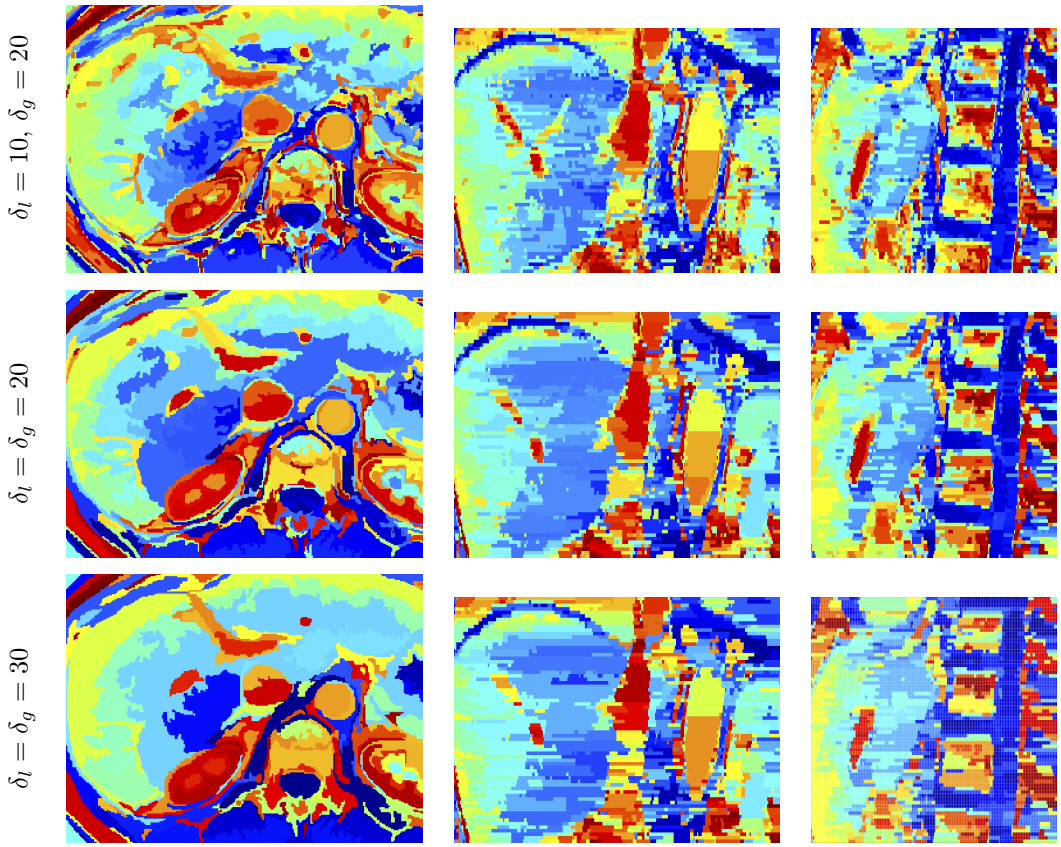


Figure 6.9: Effect of δ_l and δ_g: multi-view of 2D-to-3D segmentations of DCE volume #1 with $a = 0.3$, $\alpha = 0.001$, $\delta_{3D} = 40$ for volume segmentation based on intensity of slice segmentation, $\delta_{enh} = 60$ for enhancement clustering at the end, and different δ_l and δ_g for slice segmentation.

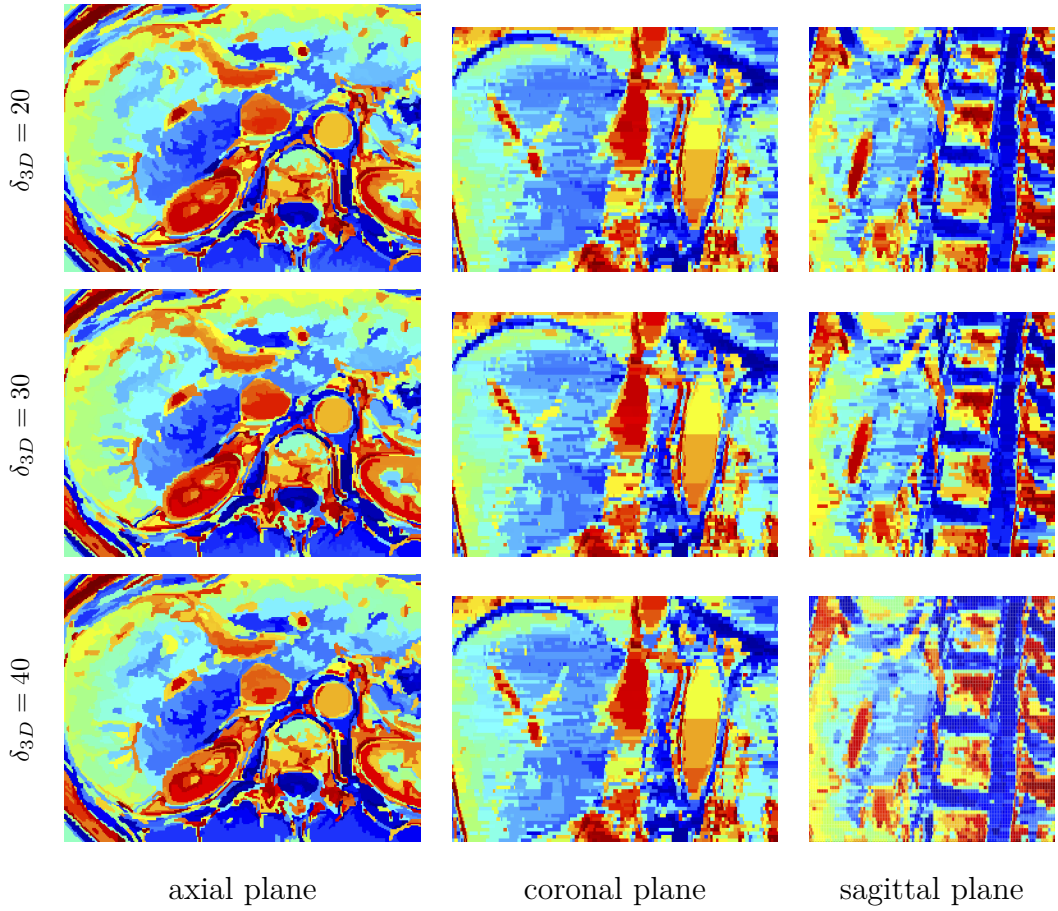


Figure 6.10: Effect of δ_{3D}: multi-view of 2D-to-3D segmentations of DCE volume #1 with $a = 0.3$, $\alpha = 0.001$, $\delta_l = 10$, $\delta_g = 20$ for slice segmentation, $\delta_{enh} = 40$ for enhancement clustering at the end, and different values of δ_{3D}.

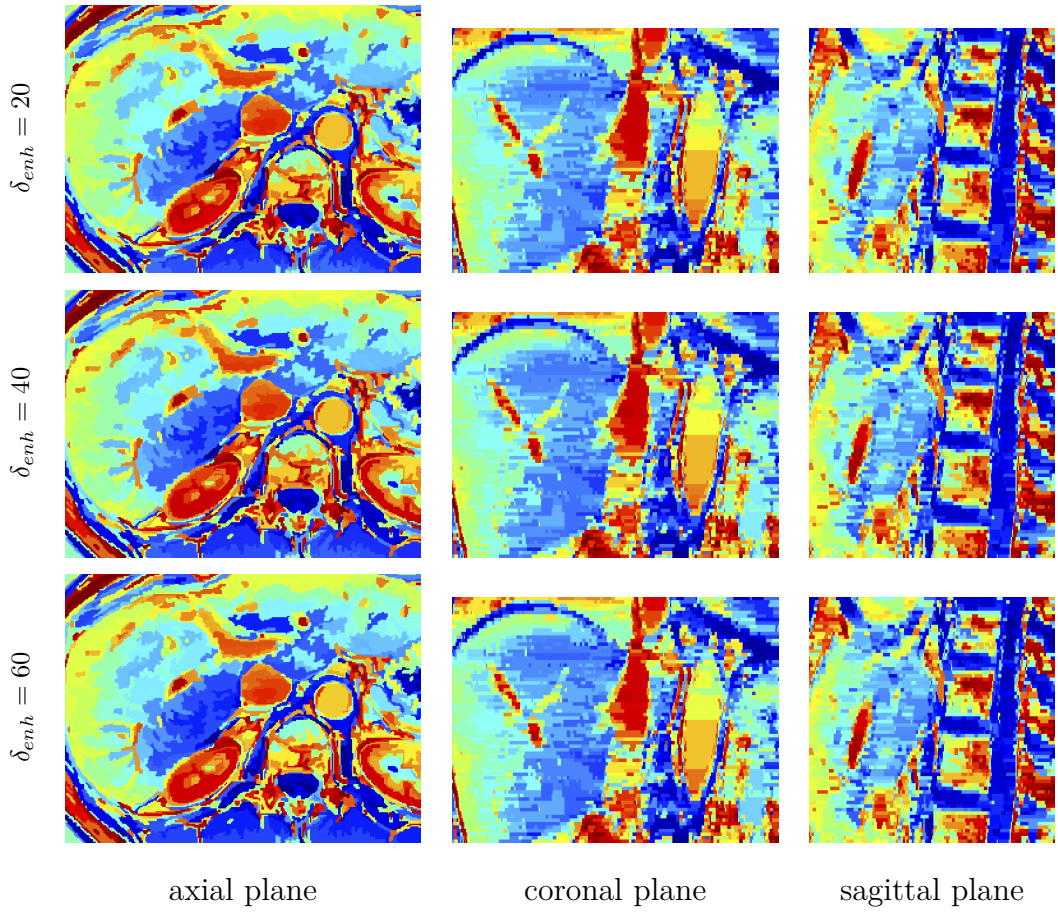
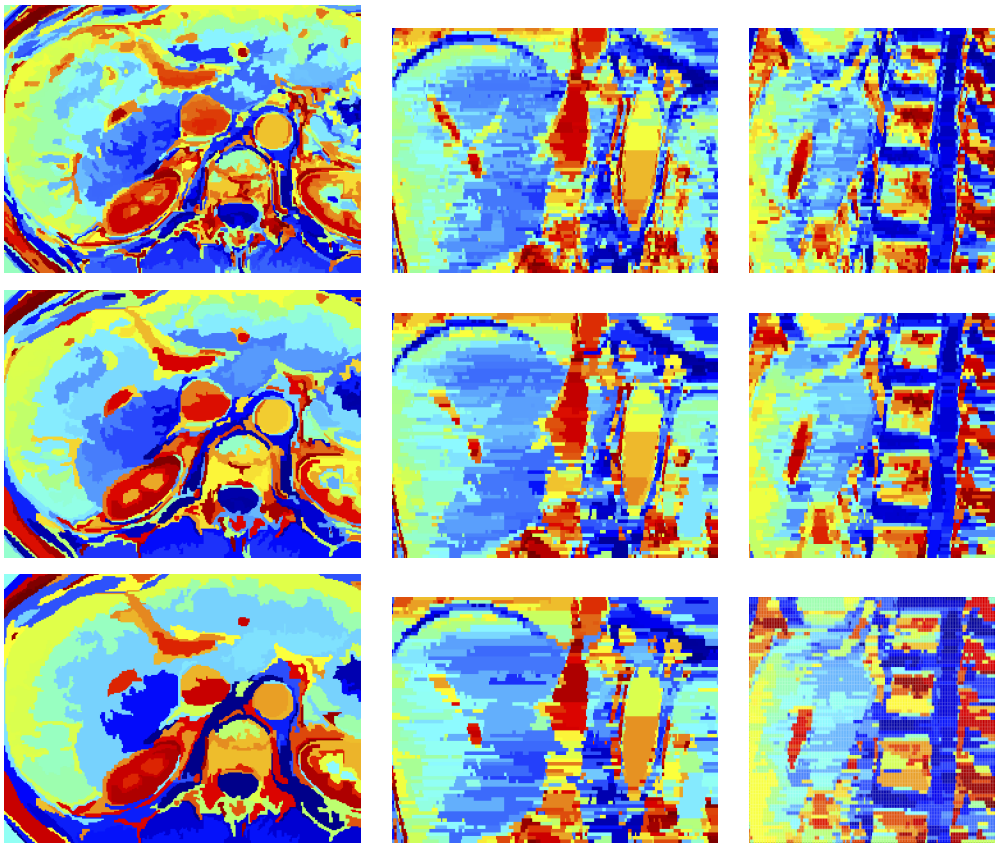


Figure 6.11: Effect of δ_{enh}: multi-view of 2D-to-3D segmentations of DCE volume #1 with $a = 0.3$, $\alpha = 0.001$, $\delta_l = 10$, $\delta_g = 20$ for slice segmentation, $\delta_{3D} = 20$ for volume segmentation based on intensity of slice segmentation, and different values of δ_{enh}.



axial plane coronal plane sagittal plane

Figure 6.12: Multi-view of 2D-to-3D segmentations of DCE volume #1 with $a = 0.3$ and $\alpha = 0.001$. Top row: $\delta_l = 10$, $\delta_g = 20$, $\delta_{3D} = 20$ and $\delta_{enh} = 40$; Middle row: $\delta_l = \delta_g = 20$, $\delta_{3D} = 30$ and $\delta_{enh} = 50$; Bottom row: $\delta_l = \delta_g = 30$, $\delta_{3D} = 40$ and $\delta_{enh} = 60$.

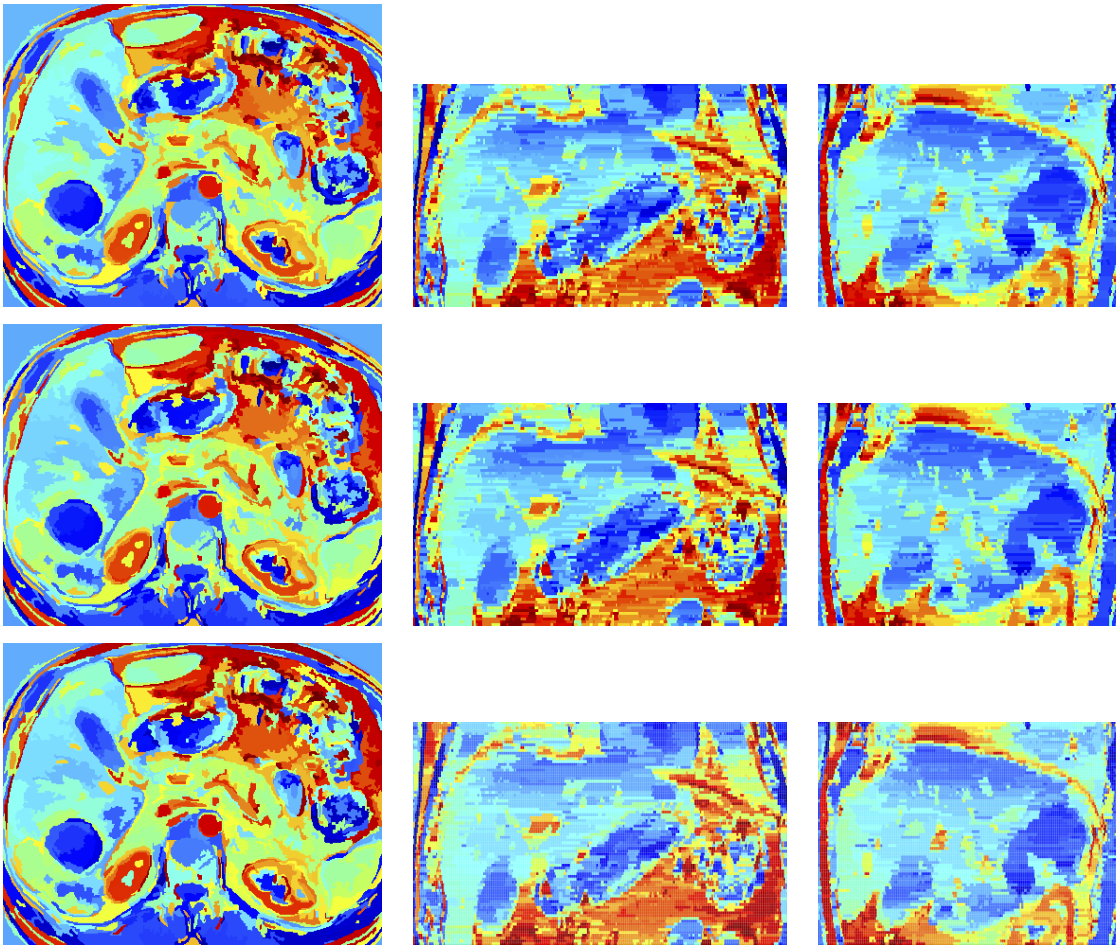
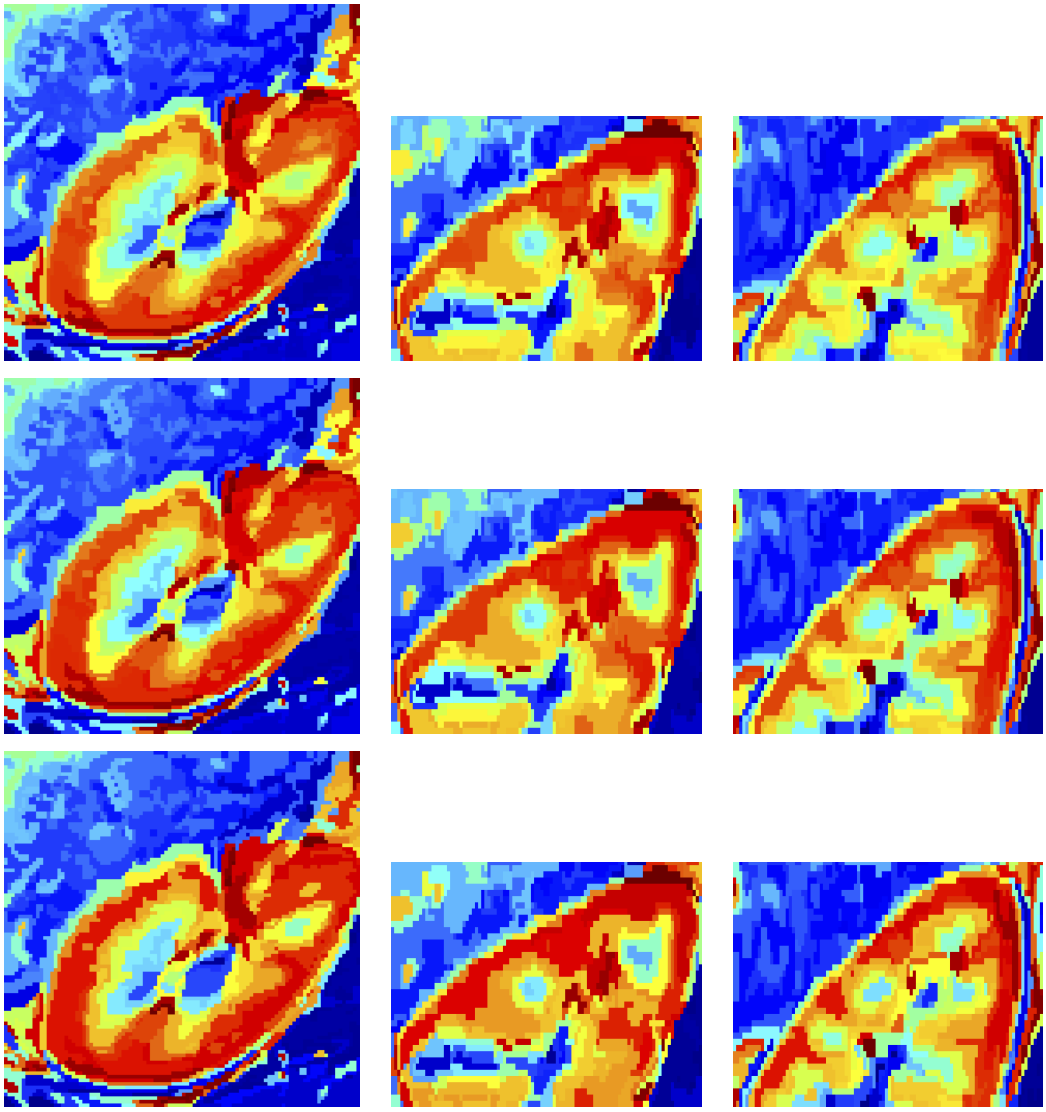


Figure 6.13: Multi-view of 2D-to-3D segmentations of DCE volume #3 with $a = 0.2$ and $\alpha = 0.001$. Top row: $\delta_l = 10$, $\delta_g = \delta_{3D} = 20$ and $\delta_{enh} = 40$; Middle row: $\delta_l = \delta_g = 20$, $\delta_{3D} = 30$ and $\delta_{enh} = 50$; Bottom row: $\delta_l = \delta_g = 30$, $\delta_{3D} = 40$ and $\delta_{enh} = 60$.



axial plane coronal plane sagittal plane

Figure 6.14: Multi-view of direct 3D segmentation of DCE volume #2 with $a = 0.3$ and $\alpha = 0.001$. Top row: $\delta_l = 10$, $\delta_g = \delta_{3D} = 20$ and $\delta_{enh} = 40$; Middle row: $\delta_l = \delta_g = 20$, $\delta_{3D} = 30$ and $\delta_{enh} = 50$; Bottom row: $\delta_l = \delta_g = 30$, $\delta_{3D} = 40$ and $\delta_{enh} = 60$.

Chapter 7

Implementation in Myrian®

Intrasense helps healthcare organizations enhance their medical imaging visualization and analysis in more than 40 countries. Founded in 2004, Intrasense is the specialist of medical imaging software for advanced visualization, quantitative analysis, workflow and qualitative clinical evidence on multi-modal medical images. The company designs, develops and markets Myrian®, a unique vendor-neutral viewer, advanced imaging layer for the visualization and advanced processing of multimodal medical images such as MRI, CT scans, PET, X-rays, and more. Over the last 10 years the company has gained an unrivalled know how to develop and support innovative and user friendly medical imaging software solutions. Developed with leading academic and industrial research partners, Myrian® solutions (platform, SDK and clinical modules) combine agility, innovation and customer experience focus, servicing healthcare professional for an improved patient's care. With over 800 Customer sites around the world, Myrian® is certified as a medical device in over 40 countries in Canada, the United States (FDA), Europe (CE) and Asia. Intrasense SA is listed on NYSE Alternext (FR0011179886-ALINS).

Myrian® combines and uses the various medical images to extract information that is vital to the patient's care, to the rapid appraisal of the efficiency of treatment and to the assessment of drug candidates in oncology, chronic diseases and other pathologies. It is a new-generation workstation for the interpretation of results

and the advanced 3D/4D post processing. It can be adapted to every modality, every organ and every pathology. Myrian® solution is enriched by advanced clinical modules each targeting one organ or one pathology with tools such as segmentation engines, specific measurement algorithms, simulation or examination tools, optimized user interface and more.

Myrian® provides also its own software development kit (SDK), which is a set of C++ interfaces and libraries that allow third-party developers to extend the possibilities of the platform by the mean of binary add-ons (DLL). Using add-ons, one can create viewports, toolboxes, tools, viewport overlays, volume processors, detection engines to generate ROI or annotations. Within the SDK, the SDK Library provides more than 100 C++ classes to help developers create their own add-ons. Those classes address various fields: user interface, algorithms for medical imaging, mathematics, file I/O, interaction with the operating system, localization and many more.

DCE-HiSET is integrated using SDK as a volume operator (VOP) into Myrian®. This operator is defined as a segmentation tool for DCE image sequence that can be used for different modules. The operator can deal with DCE image sequence with both large and small number of times. For small number of times (<8), the multiple equivalence test used for computing the dissimilarity between two TCs is based on Gaussian test. On the other hand, for large number of times, the multiple equivalence test is based on Chi-square test. Thereby, the choice of test has been made in automatic way depending on the number of times in the sequence.

For parameters, α is fixed to 0.0001 since its minor influence on the segmentation result. a and δ are left to users to define and tune. In addition to the parameters defined in Chapter 3, the interface of this VOP also requires other two parameters: the last time index of baseline phase and the slices to be processed in parallel. The former defines the baseline phase and the latter refers to a specific slice if the input is a positive integer, to all slices otherwise.

For a DCE image sequence of spatial dimensions 256×256, the VOP of DCE-HiSET takes approximately 20 seconds to produce a segmentation map and an

averaged (denoised) version of sequence. For a DCE volume of spatial dimensions $256 \times 256 \times 40$, 3D segmentation can proceed in parallel with four threads for example, one per slice. In this way, the VOP takes approximately 3 minutes (10 times 20 seconds) to produce segmentation maps for all slices. Then the 3D propagation takes about a dozen seconds. Apparently, the total time can be further reduced with more threads running in parallel for 3D segmentation. For larger spatial dimension, the resampling technique can be used before segmentation with DCE-HiSET such that the total amount of voxels is reduced while the information is mostly preserved. Additionally, the interactive visualization tools are also developed using SDK to illustrate the segmentation results, including the average TCs in segmented regions and the selection of ROIs.

Conclusion

In this work, we proposed a new method, DCE-HiSET, to segment the DCE image sequence into homogeneous regions with respect to the TC observed in each voxel. In this way, the low SNR of DCE image sequence is significantly improved by averaging the TCs of voxels in homogeneous regions.

Using a dissimilarity measure based on multiple equivalence test, the homogeneity of TC is directly controlled by the equivalence margin that is the discrepancy tolerance between true but unobservable TCs. The number of regions is automatically determined by this equivalence margin and by the significant level of the testing procedure, the latter having a much smaller influence. At least for synthetic DCE image sequence made of functionally piecewise homogeneous regions, our algorithm is proven to be able to retrieve the unknown true partition with high probability as soon as the number of images is large enough.

Consisting of a local and a global clustering step, DCE-HiSET can retrieve the homogeneous regions, which are highly consistent to real anatomy, regardless of their shape, size and disconnectedness. The total computation complexity is controlled during the local step by the small number of neighbors concerned while during the global step by the small size of the partition.

Through comparison on a synthetic DCE image sequence with a relatively small number of regions, DCE-HiSET outperforms other clustering-based methods with respect to Fowlkes-Mallows indexes. Moreover, our implementation in C++ code wrapped in R is comparable to the best competitors in term of computation time.

DCE-HiSET can be used for both 2D and 3D DCE image sequences. However,

in 3D, slice thickness and/or inter-slice distance can be very different from voxel size in the imaging cross-section. Out of this concern, we proposed two strategies having different neighborhood structure cross slices, but both based on HiSET.

As an application, this work also proposed a strategy to refine the ROI manually delineated by clinicians in DCE image sequence, to improve the precision at the border of ROI. Based on a hierarchical segmentation (HiSET) of image sequence, the strategy consists of a series of erosion and dilation to remove the regions corresponding to movements, artifacts and unexpected tissues initially covered by the manual ROI. To evaluate the effect of the proposed strategy on the robustness of perfusion analysis, we apply it to the classification of ovarian tumors from 99 studies of female patients. With different classification models, the result shows a consistent improvement of classification accuracy induced by the ROI refinement.

DCE-HiSET may be extended to take into account such difference by weighting the p-values with respect to the direction of the neighbors. Other extensions of DCE-HiSET include the automatic tuning of parameters a and δ by the optimization of a simple criterion, which may provide a fully automatic procedure for each DCE image sequence. DCE-HiSET may be the proper segmentation part in conjunction with a proper registration tool [Glocker et al., 2011; Sotiras et al., 2009] of an iterative registration process for DCE image sequence, in which along the iterations both the signals and the registration on those signals are learned. Moreover, by only adapting the multiple equivalence test, HiSET may be applied in various types of models where a signal is observed on a landscape such as multimodality images, vector-valued graphs and consumption (electricity, water, etc.) using geographical information or user types.

Due to the multidisciplinary nature of this PhD project, I am working closely with radiologists, mathematicians and software engineers, to cover the vast scope of the project. During the development and the application of DCE-HiSET, I manage to achieve different tasks sequentially and simultaneously. Through the interaction with radiologists, I was able to understand the clinical challenges

and reformulate them into mathematic problems. After defining properly the problems, I investigated the literature of different disciplines for the state-of-the-art approaches to address the problems. At the same time, I collected image data from various heterogeneous sources and kept detailed records regarding the image data, for further experiments and tests. Proposing our own theory and assumptions, I developped the prototype of the segmentation algorithm and its variations in different programming languages, as well as extensive tools for image processing, visualisation, analysis, classification and interactive illustration of results. The algorithm has been validated through both mathematical proofs and intensive experiments on heterogeneous image datasets. Meanwhile, I managed to transform the prototype into compatible module to integrate into company's software platform Myrian®. I also spent tremendous amount of effort to prepare, present and discuss the work in laboratory seminars, company meetings, as well as national and international conferences. I also produced annual research reports and wrote papers for the publications regarding the progress of this work.

During three years of CIFRE PhD, I have not only built strong skills to communicate with collaborators from different backgrounds and the ability to adapt initiatively the dialogue for more efficient presentation and discussion, but also developed deep knowledge of machine learning algorithms and the ability of autonomous learning and research. Working closely with Research and Development team of Intrasense, I have gained a strong knowledge about software development, learnt about the structure of the company and also appreciated the culture of the company and the spirit of teamwork. All these will be valuable assets for my future professional career.

Bibliography

Abdelmunim, H., Farag, A., Miller, W., and AboelGhar, M. (2008). A kidney segmentation approach from dce-mri using level sets. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, pages 1–6, Anchorage, AK.

Achanta, R., Shaji, A., Smith, K., Lucchi, A., Fua, P., and Süsstrunk, S. (2010). Slic superpixels. Technical Report 149300, EPFL.

Ackerman, M. and Ben-David, S. (2011). Discerning linkage-based algorithms among hierarchical clustering methods. In *IJCAI'11 Proceedings of the Twenty-Second international joint conference on Artificial Intelligence*, volume 2, pages 1140–1145. AAAI Press ©2011.

Agner, S. C., Xu, J., and Madabhushi, A. (2013). Spectral embedding based active contour (seac) for lesion segmentation on breast dynamic contrast enhanced magnetic resonance imaging. *Medical Physics*, 40(3).

Axel, L. (1980). Cerebral blood flow determination by rapid-sequence computed tomography: theoretical analysis. *Radiology*, 137(3):679–86.

Baraud, Y., Huet, S., and Laurent, B. (2003). Adaptive tests of linear hypotheses by model selection. *The Annals of Statistics*, 31(1):225–251.

Baraud, Y., Huet, S., and Laurent, B. (2005). Testing convex hypotheses on the mean of a gaussian vector. application to testing qualitative hypotheses on a regression function. *The Annals of Statistics*, 33(1):214–257.

Baudry, J.-P., Maugis, C., and Michel, B. (2012). Slope heuristics: overview and implementation. *Statistics and Computing*, 22(2):455–470.

Berger, R. L. and Hsu, J. C. (1996). Bioequivalence trials, intersection-union tests and equivalence confidence sets. *Statistical Science*, 11(4):283–319.

Besag, J. (1975). Statistical analysis of nonlattice data. *Statistician*, 24:179–195.

Birbrair, A., Zhang, T., Wang, Z.-M., Messi, M. L., Mintz, A., and Delbono, O. (2015). Pericytes at the intersection between tissue regeneration and pathology. *Clinical Science*, 128(2):81–93.

Birgé, L. and Massart, P. (2007). Minimal penalties for gaussian model selection. *Probability Theory and Related Fields*, 138(1):33–73.

Brix, G. and et al., F. K. (2004). Microcirculation and microvasculature in breast tumors. *Magnetic Resonance in Medicine*, 52:420–429.

Brochot, C., Bessoud, B., Balvay, D., Cuénod, C.-A., Siauve, N., and Bois, F. Y. (2006). Evaluation of antiangiogenic treatment effects on tumors' microcirculation by bayesian physiological pharmacokinetic modeling and magnetic resonance imaging. *Magnetic Resonance Imaging*, 24(8):1059–1067.

Celeux, G., Forbes, F., and Peyrard, N. (2003). Em procedures using mean field-like approximations for markov model-based image segmentation. *Pattern Recognition*, 36:131–144.

Chalmond, B. (1989). An iterative gibbsian technique for reconstruction of m-ary images. *Pattern Recognition*, 22(6):747–761.

Chatzis, S. P. and Varvarigou, T. A. (2008). A fuzzy clustering approach toward hidden markov random field models for enhanced spatially constrained image segmentation. *IEEE Transactions on Fuzzy Systems*, 16(5):1351–1361.

Chen, W., Giger, M. L., Bick, U., and Newstead, G. M. (2006). Automatic identification and classification of characteristic kinetic curves of breast lesions on dce-mri. *Medical Physics*, 33(8):2878–2887.

Chevaillier, B., Mandry, D., Collette, J.-L., Claudon, M., Galloy, M.-A., and Pietquin, O. (2011). Functional segmentation of renal dce-mri sequences using vector quantization algorithms. *Neural Process Letters*, 34(1):71–85.

Cleveland, W. S., Grosse, E., and Shyu, W. M. (1991). Local regression models. In Chambers, J. M. and Hastie, T. J., editors, *Statistical Models in S*, chapter 8. Chapman and Hall/CRC, 1 edition.

Comaniciu, D. and Meer, P. (2002). Mean shift: A robust approach toward feature space analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(5):603–619.

Comte, F., Cuenod, C. A., Pensky, M., and Rozenholc, Y. (2014). Laplace deconvolution on the basis of time domain data and its application to dynamic contrast enhanced imaging. *Journal of the Royal Statistical Society*, 15(3):397–408.

Craciunescu, O. I., Yoo, D. S., Cleland, E., Muradyan, N., Carroll, M. D., MacFall, J. R., Barboriak, D. P., and Brizel, D. M. (2012). Dynamic contrast-enhanced mri in head-and-neck cancer: The impact of region of interest selection on the intra- and interpatient variability of pharmacokinetic parameters. *International Journal Radiation Oncology Biology Physics*, 82(3):e345–e350.

Durot, C. and Rozenholc, Y. (2006). An adaptive test for zero mean. *Mathematical Methods of Statistics*, 15(1):26–60.

Eikesdal, H. P. and Kalluri, R. (2009). Drug resistance associated with antiangiogenesis therapy. In *Seminars in cancer biology*, volume 19(5).

Eisenhauer, E. A., Therasse, P., Bogaerts, J., Schwartz, L. H., Sargent, D., Ford, R., Dancey, J., Arbuck, S., Gwyther, S., Mooney, M., Rubinstein, L., Shankar, L.,

Dodd, L., Kaplan, R., Lacombe, D., and Verweij, J. (2009). New response evaluation criteria in solid tumours: revised recist guideline (version 1.1). *European journal of cancer*, 45(2):228–247.

Ferlay, J., Soerjomataram, I., Ervik, M., Dikshit, R., Eser, S., Mathers, C., Rebelo, M., Parkin, D., Forman, D., and Bray, F. (2013). Cancer incidence and mortality worldwide: Iarc cancerbase no. 11. Technical report, International Agency for Research on Cancer.

Fieselmann, A., Kowarschik, M., Ganguly, A., Hornegger, J., and Fahrig, R. (2011). Deconvolution-based ct and mr brain perfusion measurement: Theoretical model revisited and practical implementation details. *International Journal of Biomedical Imaging*, 2011(14):1–20.

Forbes, F., Peyrard, N., Fraley, C., Georgian-Smith, D., Goldhaber, D. M., and Raftery, A. E. (2006). Model-based region-of-interest selection in dynamic breast mri. *Journal of Computer Assisted Tomography*, 30:675–687.

Fowlkes, E. B. and Mallows, C. L. (1983). A method for comparing two hierarchical clusterings. *Journal of the American Statistical Association*, 78(383):553–569.

Glocker, B., Sotiras, A., Komodakis, N., and Paragios, N. (2011). Deformable medical image registration: Setting the state of the art with discrete methods. *Annual Review of Biomedical Engineering*, 13:219–244.

Górriza, J., Lassla, A., Ramíreza, J., Salas-Gonzaleza, D., Puntonetb, C., and Lang, E. (2009). Automatic selection of rois in functional imaging using gaussian mixture models. *Neuroscience Letters*.

Graves, M. J. and Mitchell, D. G. (2013). Body mri artifacts in clinical practice: a physicist's and radiologist's perspective. *Journal of Magnetic Resonance Imaging*, 38(2):269–287.

Hartigan, J. A. and Wong, M. A. (1979). A k-means clustering algorithm. *Applied Statistics*, 28:100–108.

Hedjam, R. and Mignotte, M. (2009). A hierarchical graph-based markovian clustering approach for the unsupervised segmentation of textured color image. In *IEEE International Conference on Image Processing (ICIP)*.

Herman, G. T. (2009). *Fundamentals of Computerized Tomography: Image Reconstruction from Projections*. Springer.

Heye, T., Merkle, E. M., Reiner, C. S., Davenport, M. S., Horvath, J. J., Feuerlein, S., Breault, S. R., Gall, P., Bashir, M. R., Dale, B. M., Kiraly, A. P., and Boll, D. T. (2013). Reproducibility of dynamic contrast-enhanced mr imaging part ii. comparison of intra- and interobserver variability with manual region of interest placement versus semiautomatic lesion segmentation and histogram analysis. *Radiology*, 266(3):812–821.

Hsu, C.-W. and Lin, C.-J. (2002). A comparison on methods for multi-class support vector machines. *IEEE Transactions on Neural Networks*, 13:415–425.

Ichihashi, H., Miyagishi, K., and Honda, K. (2001). Fuzzy c-means clustering with regularization by k-l information. In *10th IEEE International Conference on Fuzzy System*, pages 924–927.

Iftimia, N., Brugge, W. R., and Hammer, D. X. (2011). *Advances in Optical Imaging for Clinical Medicine*. John Wiley and Sons.

Ingrisch, M. and Sourbron, S. (2013). Tracer-kinetic modeling of dynamic contrast-enhanced mri and ct: a primer. *Journal of Pharmacokinetics and Pharmacodynamics*, 40(3):281–300.

Irving, B., Cifor, A., Papież, B. W., Franklin, J., Anderson, E. M., Brady, S. M., and Schnabel, J. A. (2014). Automated colorectal tumour segmentation in dce-mri using supervoxel neighbourhood contrast characteristics. In *Medical Image Computing and Computer-Assisted Intervention - MICCAI 2014*, volume 8673, pages 609–616.

Irving, B., Franklin, J. M., Papież, B. W., Anderson, E. M., Sharma, R. A., Gleeson, F. V., Brady, S. M., and Schnabel, J. A. (2016). Pieces-of-parts for supervoxel segmentation with global context: Application to dce-mri tumour delineation. *Medical Image Analysis*, 32:69–83.

Jain, R. (2005). Normalization of tumor vasculature: an emerging concept in antiangiogenic therapy. *Science*, 307(5706):58–62.

Kachenoura, N., Cluzel, P., Frouin, F., Toledano, D., Grenier, P., Cuenod, C. A., and Balvay, D. (2011). Evaluation of an edge-based registration method: application to magnetic resonance first-pass myocardial perfusion data. *Magnetic Resonance Imaging*, 29(6):853–860.

Koscielny, S., Tubiana, M., Lê, M. G., Valleron, J., Mouriesse, H., Contesso, G., and Sarrazin, D. (1984). Breast cancer: relationship between the size of the primary tumour and the probability of metastatic dissemination. *British journal of cancer*, 49(6):709–715.

Kulasingam, V. and Diamandis, E. P. (2008). Strategies for discovering novel cancer biomark- ers through utilization of emerging technologies. *Nature clinical practice Oncology*, 5(10):588–599.

Lavini, C., de Jonge, M. C., van de Sande, M. G., Tak, P. P., Nederveen, A. J., and Maas, M. (2006). Pixel-by-pixel analysis of dce mri curve patterns and an illustration of its application to the imaging of the musculoskeletal system. *Magnetic Resonance Imaging*, 25:604–612.

Lecoeur, J., Ferre, J.-C., Collins, D. L., Morrissey, S. P., and Barillot., C. (2009). Multi channel mri segmentation with graph cuts using spectral gradient and multidimensional gaussian mixture model. In Pluim, J. P. W. and Dawant, B. M., editors, *SPIE-Medical Imaging*, volume 7259, pages 72593X1–72593X11, Orlando, United States.

Li, S., Zöllner, F. G., Merrem, A. D., Peng, Y., Roervik, J., Lundervold, A., and Schad, L. R. (2012a). Wavelet-based segmentation of renal compartments in dce-mri of human kidney: Initial results in patients and healthy volunteers. *Computerized Medical Imaging and Graphics*, 36:108–118.

Li, S., Zöllner, F. G., Merrem, A. D., Peng, Y., Roervik, J., Lundervold, A., and Schad, L. R. (2012b). Wavelet-based segmentation of renal compartments in dce-mri of human kidney: Initial results in patients and healthy volunteers. *Computerized Medical Imaging and Graphics*, 36:108–118.

Liu, H., Liu, Y., Zhao, Z., Zhang, L., and Qiu, T. (2014). A new background distribution-based active contour model for three-dimensional lesion segmentation in breast dce-mri. *Medical Physics*, 41:082303.

Ludwig, J. A. and Weinstein, J. N. (2005). Biomarkers in cancer staging, prognosis and treatment selection. *Nature reviews Cancer*, 5(11):845–856.

Maroquin, J., Mitte, S., and Poggio, T. (1987). Probabilistic solution of ill-posed problems in computational vision. *Journal of the American Statistical Association*, 82:76–89.

Massart, P. (2003). *Concentration Inequalities and Model Selection*. Springer.

McClymont, D., Mehnert, A., Trakic, A., Kennedy, D., and Crozier, S. (2014). Fully automatic lesion segmentation in breast mri using mean-shift and graph-cuts on a region adjacency graph. *Journal of Magnetic Resonance Imaging*, 39:795–804.

McLachlan, G. and Peel, D. (2000). *Finite Mixture Models*. Probability and Statistics Series. New York: Wiley.

Medved, M., Medved, M., Karczmar, G., Yang, C., Dignam, J., Gajewski, T. F., Kindler, H., Vokes, E., MacEneany, P., Mitchell, M. T., and Stadler, W. M. (2004). Semiquantitative analysis of dynamic contrast enhanced mri in cancer patients: variability and changes in tumor tissue over time. *Magnetic Resonance Imaging*, 20:122–128.

Miyamoto, S. and Mukaidono, M. (1997). Fuzzy c-means as a regularization and maximum entropy approach. In *Proceedings of 7th International Fuzzy System Association World Congress*, volume 2, pages 86–92.

Ortiz, A., Górriz, J. M., Ramírez, J., and Martinez-Murcia, F. J. (2014). Automatic roi selection in structural brain mri using som 3d projection. *PLoS ONE*, 9(4):e93851.

Ostergaard, L., Sorensen, A. G., Kwong, K. K., Weisskoff, R. M., Gyldensted, C., and Rosen, B. R. (1996). High resolution measurement of cerebral blood flow using intravascular tracer bolus passages. part ii: Experimental comparison and preliminary results. *Magnetic Resonance in Medicine*, 36(5):726–36.

Otsu, N. (1979). A threshold selection method from gray-level histograms. *IEEE Transactions on Systems, Man, and Cybernetics*, 9(1):62–66.

Ripley, B. D. (1996). *Pattern Recognition and Neural Networks*. Cambridge University Press.

Schmid, V. J., Whitcher, B., Padhani, A. R., and Yang, G.-Z. (2009). Quantitative analysis of dynamic contrast-enhanced mr images based on bayesian p-splines. *IEEE Transactions on Medical Imaging*, 28(6):789–798.

Shi, J. and Malik, J. (2000). Normalized cuts and image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(8):888–905.

Shi, J., Sahiner, B., Chan, H.-P., Paramagul, C., Hadjiiski, L. M., Helvie, M., and Chenevert, T. (2009). Treatment response assessment of breast masses on dynamic contrast-enhanced magnetic resonance scans using fuzzy c-means clustering and level set segmentation. *Medical Physics*, 36:5052–5063.

Shou, H., Shinohara, R. T., Liu, H., Reich, D. S., and Crainiceanu, C. M. (2016). Soft null hypotheses: A case study of image enhancement detection in brain lesions. *Journal of Computational and Graphical Statistics*, 25(2):570–588.

Sotiras, A., Komodakis, N., Glocker, B., Deux, J.-F., and Paragios, N. (2009). Graphical models and deformable diffeomorphic population registration using global and local metrics. In *MICCAI*, pages 672–679.

Stanley, J., Shipley, W., and Steel, G. (1977). Influence of tumour size on hypoxic fraction and therapeutic sensitivity of lewis lung tumour. *British journal of cancer*.

Stoutjesdijk, M. J., Veltman, J., Huisman, H., Karssemeijer, N., Barentsz, J. O., Blickman, J. G., and Boetes, C. (2007). Automated analysis of contrast enhancement in breast mri lesions using mean shift clustering for roi selection. *Journal of Magnetic Resonance Imaging*, 26:606–614.

Stoutjesdijk, M. J., Zijp, M., Boetes, C., Karssemeijer, N., Barentsz, J. O., and Huisman, H. (2012). Computer aided analysis of breast mri enhancement kinetics using mean shift clustering and multifeature iterative region of interest selection. *Journal of Magnetic Resonance Imaging*, 36:1104–1112.

Taha, A. A. and Hanbury, A. (2015). Metrics for evaluating 3d medical image segmentation: analysis, selection, and tool. *BMC Medical Imaging*.

Tao, W., Jin, H., and Zhang, Y. (2007). Color image segmentation based on mean shift and normalized cuts. *IEEE Transactions on Systems, Man, and Cybernetics*, 37(5):1382–1389.

Tartare, G., Hamad, D., Azahaf, M., Puech, P., and Betrouni, N. (2014). Spectral clustering applied for dynamic contrast-enhanced mr analysis of time–intensity curves. *Computerized Medical Imaging and Graphics*, 38(8):702–713.

Thomassin-Naggara, I., Balvay, D., Aubert, E., Daraï, E., Rouzier, R., Cuenod, C. A., and Bazot, M. (2012). Quantitative dynamic contrast-enhanced mr imaging analysis of complex adnexal masses: a preliminary study. *European Radiology*, 22(4):738–745.

Thomassin-Naggara, I., Bazot, M., Daraí, E., Callard, P., Thomassin, J., and Cuenod, C. (2008). Epithelial ovarian tumors: value of dynamic contrast-enhanced mr imaging and correlation with tumor angiogenesis. *Radiology*, 248:148–159.

Thomassin-Naggara, I., Soualhi, N., Balvay, D., and Cuenod, C. A. (In Press). Quantifying tumour vascular heterogeneity with dce-mri in complex adnexal masses. *Journal of Magnetic Resonance Imaging*.

Thomassin-Naggara, I., Toussaint, I., Perrot, N., Rouzier, R., Cuenod, C., Bazot, M., and Daraí, E. (2011). Characterization of complex adnexal masses: value of adding perfusion- and diffusion-weighted mr imaging to conventional mr imaging. *Radiology*, 258(3):793–803.

Tilton, J. C., Tarabalka, Y., Montesano, P. M., and Gofman, E. (2012). Best merge region-growing segmentation with integrated nonadjacent region object aggregation. *IEEE Transactions on Geoscience and Remote Sensing*, 11:4454–4467.

Tofts, P. S. (1997). Modeling tracer kinetics in dynamic gd-dtpa mr imaging. *Journal of Magnetic Resonance Imaging*, 7:91–101.

Tofts, P. S., Berkowitz, B., and Schnall, M. D. (1995). Quantitative analysis of dynamic gd-dtpa enhancement in breast tumors using a permeability model. *Magnetic Resonance in Medicine*, 33:564–568.

Tremeau, A. and Borel, N. (1997). A region growing and merging algorithm to color segmentation. *Pattern Recognition*, 30(7):1191–1203.

Tu, Z. and Zhu, S.-C. (2002). Image segmentation by data-driven markov chain monte carlo. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(5):657–673.

Tuncbilek, N., Karakas, H. M., and Altaner, S. (2004). Dynamic mri in indirect estimation of microvessel density, histologic grade, and prognosis in colorectal adenocarcinomas. *Abdom Imaging*, 29:166–72.

Tuncbilek, N., Karakas, H. M., and Okten, O. O. (2005). Dynamic contrast enhanced mri in the differential diagnosis of soft tissue tumors. *Europeen Journal Radiology*, 53:500–505.

Vedaldi, A. and Soatto, S. (2008). Quick shift and kernel methods for mode seeking. In *European Conference on Computer Vision (ECCV)*.

Walker, E. and Nowacki, A. S. (2010). Understanding equivalence and noninferiority testing. *Journal of General Internal Medicine*, 26(2):192–196.

Wellek, S. (2010). *Testing Statistical Hypotheses of Equivalence and Noninferiority, Second Edition*. Chapman and Hall/CRC.

Wu, Q., Salganicoff, M., Krishnan, A., Fussell, D. S., and Markey, M. K. (2006). Interactive lesion segmentation on dynamic contrast enhanced breast mri using a markov model. *Medical Imaging 2006: Image Processing*, 6144(1).

Zelnik-Manor, L. and Perona, P. (2004). Self-tuning spectral clustering. In *Advances in Neural Information Processing Systems 17*, pages 1601–1608. MIT Press.

Zöllner, F. G., Sance, R., Rogelj, P., Ledesma-Carbayo, M. J., Rørvik, J., Santos, A., and Lundervold, A. (2009). Assessment of 3d dce-mri of the kidneys using non-rigid image registration and segmentation of voxel time courses. *Computerized Medical Imaging and Graphics*, 33:171–181.