Hierarchical segmentation using equivalence test (HiSET): Application to DCE image sequences

Fuchen Liu [a,g,h], Charles-André Cuenod [a,d,e], Isabelle Thomassin-Naggara [c,f,e], Stéphane Chemouny [g], Yves Rozenholc [a,b,*]

[a] *Université Paris Descartes, Université Sorbonne Paris Cité (USPC), France*
[b] *Faculté de Pharmacie de Paris - EA bioSTM, France*
[c] *Université Pierre et Marie Curie - Sorbonne Université, France*
[d] *Hôpital Européen Georges Pompidou (HEGP), Assistance Publiques - Hôpitaux de Paris (APHP), France*
[e] *UMR-S970, PARCC, France*
[f] *Hôpital Tenon - APHP, France*
[g] *Intrasense®, France*
[h] *MAP5, UMR CNRS 8145, France*

ARTICLE INFO

ABSTRACT

Dynamical contrast enhanced (DCE) imaging allows non invasive access to tissue micro-vascularization. It appears as a promising tool to build imaging biomarkers 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 a large region of interest when it is functionally homogeneous.

We propose a novel method for automatic segmentation of DCE image sequences into functionally homogeneous regions, called *DCE-HiSET*. Using an observation model which depends on one parameter *a* and is justified *a posteriori*, DCE-HiSET is a hierarchical clustering algorithm. 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 reduce complexity and takes advantage of the regularity of anatomical features, while the second recovers (spatially) disconnected homogeneous structures at a larger (global) scale.

Given a minimal expected homogeneity discrepancy for the multiple equivalence test, both steps stop automatically by controlling the Type I error. This provides an adaptive choice for the number of clusters. Assuming that the DCE image sequence is functionally piecewise constant with signals on each piece sufficiently separated, we prove that DCE-HiSET will retrieve the exact partition with high probability as soon as the number of images in the sequence is large enough. The minimal expected homogeneity discrepancy appears as the tuning parameter controlling the size of the segmentation. DCE-HiSET has been implemented in C++ for 2D and 3D image sequences with competitive speed.

© 2018 Elsevier B.V. All rights reserved.

1. Introduction

1.1. DCE imaging

DCE (dynamic contrast enhanced) imaging using computed tomography (CT), magnetic resonance imaging (MRI) or ultrasound imaging (US) appears promising as it can monitor the local changes in microcirculation secondary to the development of new vessels (neo-angiogenesis). DCE-MRI and DCE-CT (called also CT-perfusion) have been extensively tested alone or in combination with other techniques (Winfield et al., 2016) in pathological conditions such as cancer, ischemia and inflammation, in various tissues including brain (Bergamino et al., 2014), breast (Chen et al., 2010), prostate (Sanz-Requena et al., 2016), heart (Bakir et al., 2016; Nee et al., 2009), kidney (Woodard et al., 2015), liver (Raj and Juluru, 2009; Chen and Shih, 2014), genital organs, gastrointestinal tract, bone (Michoux et al., 2012) and placenta (Frias et al., 2015). They show a great potential to: 1/ detect and characterize lesions (Sanz-Requena et al., 2016; Ferre et al., 2016; Bhooshan et al., 2010); 2/ personalize treatment including new targeted drugs, radiotherapy and mini invasive surgery; 3/ monitor and optimize treatments

---

during the follow-up (Padhani and Khan, 2010; Wang et al., 2014) or after heart, liver or kidney transplants (Khalifa et al., 2013).

DCE imaging follows the local distribution of a contrast agent after intravenous injection, using sequential acquisition (Ingrisch and Sourbron, 2013). Taken at the same cross-section of the patient during the entire acquisition, the images of the *DCE image sequence* discretely reveal, at each voxel, a *time curve* (TC) made of noisy intensities. TCs analysis involves either hardly reproducible abstract parameters (Medved et al., 2004; Tuncbilek et al., 2005, 2004) or physiological parameters after a deconvolution with respect to the patient's Arterial Input Function (AIF)[1] (Brix et al., 2004; Chung et al., 2016; Tofts, 1997; Tofts et al., 1995; Axel, 1980; Fieselmann et al., 2011; Ostergaard et al., 1996; Brochot et al., 2006; Schmid et al., 2009; Comte et al., 2017).

1.2. Limitations

However, DCE imaging suffers from two technical issues, that hamper its use: movements and low SNR. Movements (such as breathing, cardiac motion and non-periodic motions like bowel peristalsis) can be addressed by motion correction (Glocker et al., 2011; Sotiras et al., 2009). Movements and registration is not our concern and we assume further that the DCE image sequences have already been registered. Low SNR, which is our concern, affects any estimation. It is a consequence of minimizing the total X-ray dose received by the patient during this multi-image acquisition in DCE-CT or of the short delay between two consecutive images in DCE-MRI.

To improve SNR, either TCs are averaged over large regions of interest (ROI) manually drawn by the radiologist or denoised with spatial filtering techniques (e.g. PCA). However, large ROI could result in a lack of homogeneity. In this case, the gain in variance resulting from averaging is obtained at the price of a high bias which may compromise further analysis. Similarly, filtering techniques applied to the observed noisy TCs could trigger a tricky trade-off between information and noise when the noise structure has heavy tails or when the TC shows high frequencies. The latter is a typical consequence of injecting the contrast agent as a bolus. From a statistical point-of-view, as a result of the bias-variance decomposition of the estimation risk, this issue can be solved using adaptive statistical procedures (e.g. threshold wavelet decomposition or multi-scale comparison) with respect to the (unknown) smoothness of the (unknown) underlying signal. However, as the functional information in DCE sequences results from a convolution with the AIF, the true signals can be expected to be smooth enough. In this context, a multi-scale comparison, known to be adaptive with respect to Hölder regularity, provides a good bias-variance trade-off while having low level of complexity.

The fixed cross section of a DCE image sequence defines a fixed spatial (2D or 3D) domain made up of voxels. At each voxel, a TC is observed in the time domain, leading to the 2D+T or 3D+T representations. After registration, a DCE image sequence reveals tissues (organs, vessels, bones, muscles, fat, etc) as having homogenous functional properties. Hence, tissues can be considered as spatially static functional objects. As shown by Zheng et al. (2009) for breast tumors and on our real data in Section 3, even heterogeneous tumors or metastases have only a small number of functional behaviors. Thus, their homogeneous sub-parts can also be treated as static functional objects. It leads to segmenting the DCE image sequence into regions (of voxels) showing homogeneous TCs. Within each of these regions, SNR can be improved by averaging without the loss of temporal information. This alternative is known as *DCE image sequence segmentation*.

Segmentation is known to be a useful tool for common image analysis to get SNR improvement, it is an even more important preprocessing steps for DCE-imaging analysis. Indeed, in addition to the expected SNR improvement, it allows to: 1/ reveal the functional anatomy that is hardly visible on static images (Hanson et al., 2017), therefore allowing the correct selection of different tissues (Irving et al., 2016b; Heye et al., 2013); 2/ summarize the functional information, hence reducing the amount of parameter extractions; 3/ increase the conspicuity of the images and facilitate the fit of the TCs (Hou et al., 2014; Selvan et al., 2017; Giannini et al., 2010); 4/ analyze genuine tissues and obtain correct parameters; 5/ provide and guarantee the detection, measurement and characterization of lesions in clinical practice (Baumgartner et al., 2005); 6/ improve the communication between clinicians by providing synthetic pictures.

In this segmentation framework, unfortunately, due to the diversity of tissues and their associated behaviors in response to the contrast agent injection, the number of regions/objects is unknown and clearly not predictable. Moreover, tissues could be filled, surrounded or interrupted by air or water but could also appear at several disconnected locations due to their geometry (e.g. colon). Thus, their number may vary from a few (a dozen) to many (several hundred). Therefore, a good segmentation should achieve an adaptive partitioning of the spatial domain into potentially disconnected functionally homogenous subsets/tissues. Here, adaptation is expected with respect to both the partition size and the subset shapes.

1.3. Overview of this paper

To achieve the goals above, we propose a new method called *DCE-HiSET* where HiSET stands for *Hierarchical Segmentation using Equivalence Test*. Assuming constant noise level, DCE-HiSET employs a multiple equivalence test derived from a multi-resolution comparison in the time domain, known to be adaptive up to the unknown Hölder regularity of the unknown underlying signal using only $O(\log(n))$ comparisons. More specifically, given two TCs (or two averages of TCs), the equivalence test rejects when they are not sufficiently separated to be considered distinguishable. Thus, it allows to bind subsets that are functionally homogenous under this alternative. Considering the *p*-value of the multiple equivalence test as a dissimilarity measure, DCE-HiSET is a bottom-up hierarchical clustering algorithm, consisting of two steps: one local and one global: at each iteration the most functionally homogeneous pair of subsets are bound.

In the local step (growth of regions), starting from a partition made up of all voxels, only subsets that are spatially neighbors may be aggregated into larger clusters.

In the global step (merging regions), starting from the partition resulting from the local step, all clusters may be aggregated into larger clusters, regardless of whether they are neighbors or not.

These two steps ensure that *DCE-HiSET* will benefit from: 1/ the spatial regularity which can be expected inside tissues; 2/ a computational complexity which is lower than quadratic in the number of voxels; 3/ the opportunity to recover non-connected homogenous tissue.

DCE-HiSET depends on two intuitive parameters: α, the significance level of the multiple equivalence test, controlling the probability of mistakenly binding two regions in the iterative process; δ, the smallest expected homogeneity discrepancy between two (unobservable true) TCs. Empirically, the segmentation results show little sensitivity to α; hence, it can be easily fixed as a meta-parameter. Parameter δ is a human interpretable parameter given it is meaningful on noiseless signals. It defines a kind of (multi-resolution) tube in which TCs are considered similar. Given α and δ, both steps automatically stop through a proper control of the

---

[1] The AIF is the main blood flow arriving to the tissue of interest, often the aorta or a large artery.

Type I error, providing an adaptive choice of the number of clusters.

For an image sequence made up of functionally homogenous piecewise regions with signals on each piece that are different enough with respect to n, the number of images in the sequence, DCE-HiSET is theoretically proven to retrieve the true underlying partition with high probability.

For a real DCE sequence made of n images, we propose a model of the observed intensities of the sequence, which depends only on one parameter a. DCE-HiSET is applied on the transformed sequence after a variance stabilization which ensure a constant noise level equal to 1.

Reading guide: Hereafter, we review the literature related to image segmentation in the context of DCE image sequence. In Section 2, we describe our statistical model and objective (Section 2.1), the multiple equivalence test (Section 2.2) and the two-step clustering procedure with its theoretical properties (Section 2.3), along with the parameters involved (Section 2.4). In Section 3, we first introduce the material used for evaluation, including both synthetic and real DCE image sequences (Section 3.1) and state-of-the-art competitors (Section 3.2). Then, we present a comparison using the synthetic DCE image sequence (Section 3.3) and a study on real DCE image sequences, including parameter influence and model validation (Section 3.4). Technical details and proofs are relegated to Appendices together with a table of our notations with their explanations.

1.4. Prior approaches to DCE image sequence segmentation

Region (object)-based segmentation has been investigated for detecting lesions (Chen et al., 2006; Irving et al., 2016a; Tartare et al., 2014), or for retrieving internal structure of organs using prior knowledge on the number of tissues in the organ of interest (Lecoeur et al., 2009; Li et al., 2012). All these works use clustering-based image segmentation methods generally together with a supervised refinement step requiring a training dataset. Here, we focus only on an unsupervised approach that does not require a such training dataset. Moreover, we do not expect to have a prior knowledge of the number of tissues.

TC representation in a DCE image sequence may be either parametric (Agner et al., 2013; Tartare et al., 2014; Irving et al., 2016a; Shou et al., 2016) or non-parametric (Li et al., 2012). The latter is known to provide a representation adaptive to the unknown TC regularity. This is not the case for parametric representations which are using a fixed number of descriptors.

Direct clustering of the TCs had been proposed using k-means (Kachenoura et al., 2011; Li et al., 2012) or fuzzy c-means (FCM) (Chen et al., 2006). It leads to segmenting the spatial domain by using the resulting TC labels. These approaches require an extra post-processing step (e.g. hole-filling) (Chen et al., 2006). To incorporate spatial and time domain structures into one global procedure, two main types of segmentation have been proposed. The first consists of segmenting features obtained from binding spatial information (voxel coordinates) to the TCs (Comaniciu and Meer, 2002). The second considers the TCs to be distributed as a Gaussian mixture model (GMM) showing some regularity over the spatial domain ensured by a Markov Random Field (MRF) prior (Chatzis and Varvarigou, 2008; Wu et al., 2006).

These previous approaches focused only on partial (or binary) segmentation. Despite being potentially adaptable to our purpose (complete segmentation of the spatial domain), to our knowledge, these methods have never been used for this objective.

Even for classic static images (grey, color, texture), few non-supervised methods have been proposed to solve the more complex problem of complete image segmentation. We review their main ideas and apply them when possible to DCE image sequence

segmentation. These can be broadly classified into three categories: *model-based, graph-based* and *hybrid*.

Model-based methods describe the feature space as a mixture of models over the spatial domain. Minimizing the within-cluster distance, the segmentation is obtained by k-means (Li et al., 2012) or FCM (Chen et al., 2006). Maximizing the posterior probability of a GMM with MRF prior, the segmentation is achieved by Expectation-Maximization (EM) (Celeux et al., 2003) or Bayesian sampling (MCMC) (Tu and Zhu, 2002). k-means and FCM tend to find clusters with comparable shapes while the use of GMM allows clusters to have different shapes. All these methods are strongly sensitive to their initialization step (seeds or initial partition) and suffer from the use of a pre-specified number of clusters. They tend to fail in presence of clusters with 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 in the feature space, obtained by binding spatial coordinates and color information. Each individual voxel (or pixel) is assigned to a mode by minimizing a criterion computed on bound features. They do not need a predefined number of clusters; however they do need global kernel bandwiths for both time and spatial domain. Their choices are not automatic and require a strong expertise in the image to be segmented. Unfortunately, the use of global bandwidth cannot provide adaption in a general framework.

Graph-based methods treat image segmentation as a graph-partitioning problem. The weighted graph is constructed from an image by considering voxels (or pixels) as nodes. The edges reflect the similarity between features, spatial distance or even both (Shi and Malik, 2000; Zelnik-Manor and Perona, 2004). The partition is the result of a global minimization using the eigenvalues of a Laplacian matrix. Graph-based methods are able to handle feature space with more complex structures. However, they also 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 a cost function that depends on the number of clusters which is allowed to vary in a predefined set of values. Such approach has been used to segment DCE image sequences of prostate tumor (Tartare et al., 2014), with the number of clusters ranging between 1 and 5 only. In the context of tumor segmentation, Irving et al. (2016a) used a supervised step to discriminate between tissues/behaviors.

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). These hybrid methods have been applied to DCE image sequence (Irving et al., 2016a; McClymont et al., 2014). The local step over-segments the image into local homogenous clusters, called supervoxel (or superpixel). The global step merges these superpvoxels to obtain the final result. Use of supervoxels is expected to reduce the sensitivity to noise and provide better segmentation, as mentioned in Tao et al. (2007). Hybrid methods share similar flaws with the methods they are based on in each step (see above).

2. Method

2.1. Statistical observation model and objective

For DCE-MRI, at each voxel location x on a finite grid $\mathcal{X}$ describing the image cross-section, the intensities $\Phi^x(t_j) \in \mathbf{R}_+$, $j = 1, \ldots, n$ are observed at the n acquisition times, $t_1, t_2, \ldots, t_n$. Up to the baseline, defined as the intensity before the arrival of the contrast agent, these intensities account for the amount of contrast agent particles within the voxel and within the acquisition delay.

In queueing theory, it is usual to model the arrival increments as a Poisson distribution. Neglecting the baseline and assuming independent increments, namely independent arrivals of contrast agent particles, the amount of contrast agent particles $\Phi^x(t_j)$ within the voxel x at time t_j can also be assumed to be Poisson distributed. We denote by $\phi^x(t_j)$ its expectation such that $\Phi^x(t_j) \sim \mathcal{P}(\phi^x(t_j))$. Here, the Poisson distribution is used to model the signal enhancements due to the contrast agent injection. In this context, $\phi^x(t_j)$ may be considered large. This leads to the following Gaussian approximation of the distribution of $\Phi^x(t_j)$ together with its classical variance stabilization:

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

To relax our assumptions and provide more flexibility in the choice of the intensity distribution (e.g. heavier tails), we suppose that there exists $0 < a < 1$ such that

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

with a variance stabilized by the following:

$$\frac{\left(\Phi^x(t_j)\right)^a}{a} - \frac{\left(\phi^x(t_j)\right)^a}{a} \sim \mathcal{N}(0, 1). \tag{3}$$

In other words, denoting

$$I^x(t_j) := \frac{\left(\Phi^x(t_j)\right)^a}{a} \quad \text{and} \quad i^x(t_j) := \frac{\left(\phi^x(t_j)\right)^a}{a}, \tag{4}$$

we assume that the following model describes the transformed intensities:

$$I^x(t_j) = i^x(t_j) + \eta_j^x, \qquad j = 1, \ldots, n \text{ and } x \in \mathcal{X}, \tag{5}$$

where η_j^x are standard Gaussian random variables independent with respect to time index j (thanks to the production of one image at each time and to the significant delay between two consecutive image acquisitions). Spatial independence is much harder to justify due to the presence of spatial artifacts (Graves and Mitchell, 2013; Herman, 2009). Nevertheless, we make the following assumption in order to derive our mathematical construction.

Modeling Assumption 1. Random variables η_j^x in (5) are standard Gaussian, independent with respect to both spatial location x and time index j.

Furthermore, the *time curve* (TC) refers to the transformed version $i^x(.)$. Then, $I^x := (I^x(t_1), \ldots, I^x(t_n))$ appears as a discretely observed noisy version of the unobservable true TC at time $t_1, \ldots, t_n$, that is of $i^x := (i^x(t_1), \ldots, i^x(t_n))$.

Our statistical objective is 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 by answering the question "are $i^x(.)$ and $i^y(.)$ equal or not?" from their discrete observations I^x and I^y.

Integrating a more complex dependence in the spatial structure and extending our construction is certainly feasible but is far beyond the scope of this paper. Moreover thanks to the quality of our results and the study of the empirical residuals (see Section 3.4.2), only small improvements may be expected from a more realistic modelization of the spatial dependency. Nonetheless, we can provide insights why these artifacts are having such small effects in the clustering. If they affect the full sequence on a small spatial domain, they will be integrated into the underlying unknown signals $i^x(.)$ for those x in this domain and then be automatically compensated. Otherwise, when all images of the sequence are not affected in the same way[2], the resulting spatial dependence varies from image to image and compensate by considering strategies based on the whole TC as in our construction.

The contrast agent is always injected a few seconds after acquisition starts, such that the baseline grey level may be estimated. Depending on their objective: tissue separation or analysis of the enhancements, radiologists or clinicians will focus on either the "time intensity curve", which is made from the original intensities, or the "time enhancement curve", obtained after the removal of the estimated baseline at each voxel.

2.2. Equivalence test and dissimilarity measure

Given a set of voxels $X \subset \mathcal{X}$, we consider the average discrete TCs

$$\bar{I}^X := \frac{1}{|X|} \sum_{x \in X} I^x \quad \text{and} \quad \bar{i}^X := \frac{1}{|X|} \sum_{x \in X} i^x, \tag{6}$$

where $|X|$ is the cardinality of X. Given another set Y such that $X \cap Y = \emptyset$, we consider the scaling factor $\rho^2(X, Y) := |X|^{-1} + |Y|^{-1}$ and the normalized difference $D^{XY} := (\bar{I}^X - \bar{I}^Y) / \rho(X, Y)$. Under Modeling Assumption 1, D^{XY} follows a multivariate Gaussian distribution $\mathcal{N}(d^{XY}, \mathrm{Id}_n)$ with mean $d^{XY} := (\bar{i}^X - \bar{i}^Y) / \rho(X, Y)$.

The TCs of X and Y will be considered as similar if their difference d^{XY} is statistically not different from the zero vector. For this purpose, we consider an *equivalence test* of the hypotheses

$$\mathcal{H}_0 : d^{XY} \neq \vec{0} \quad \text{versus} \quad \mathcal{H}_1 : d^{XY} = \vec{0}, \tag{7}$$

whose alternative (research) hypothesis is the null hypothesis of the conventional test of equality. The term *equivalence* highlights that equality is aimed to be shown[3]

As d^{XY} belongs to $\mathbb{R}^n$, using one Gaussian equivalence test (Wellek, 2010) per time index would involve n tests, leading to an important multiplicity problem[4]. In order to control this, we modify the multiple test of Baraud et al. (2003, 2005), which use a dyadic decomposition of the time indexes into $\lfloor \log_2 n \rfloor$ time partitions. Our modifications, closely related to the Haar wavelet decomposition, ensure that the involved orthogonal projections are independent to control our clustering procedure. For $K = 0, \ldots, K_0$ with $K_0 := \lfloor \log_2 n \rfloor - 1$, the construction of the time partitions and the projection of D^{XY} onto the K-th partition, denoted by $\Pi_K D^{XY}$, are described in Appendix A. Starting from $D_{K_0+1}^{XY} = D^{XY}$, we consider the K-th *residual after projection*: $D_K^{XY} = \Pi_K D^{XY} - \Pi_{K-1} D^{XY}$, for $K = K_0 + 1, \ldots, 0$ with $\Pi_{-1} D = 0$. By construction, the D_K^{XY} for $K = K_0, \ldots, 0$ are orthogonal, which ensures their independence under Assumption 1, thanks to Cochran's theorem. Considering K-th *rescaled residual after projection* $\bar{D}_K = \Sigma_K^{-1/2} D_K$ where Σ_K is the diagonal matrix defined by (A.1), our test statistics are thereby $\|\bar{D}_K^{XY}\|_n^2$ instead of $\|\Pi_K D^{XY}\|_n^2$ as in Baraud et al. (2003, 2005). Under Assumption 1,

$$\|\bar{D}_K^{XY}\|_n^2 \sim \chi^2(2^{K-1}, \|\bar{d}_K^{XY}\|_n^2) \quad \text{for} \quad K = 1, \ldots, K_0$$

and $\|\bar{D}_0^{XY}\|_n^2 \sim \chi^2(1, \|\bar{d}_0^{XY}\|_n^2)$ where

- $\|u\|_n$ denotes the Euclidian norm of vector u,

---

[2] 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

[3] Equivalence testing derives from the need of pharmacists to prove that a drug-copy was making as well as the original drug up to a given tolerance. This is exactly our context here but for signals instead of drugs. In a classical testing construction with a given level of risks, even if all signals would be same a proportion close to the risk level will be declared different leading to many isolated voxels.

[4] a coordinate-by-coordinate equivalence test will compare each coordinate of the difference to 0, leading to n comparisons each realized at a level α/n to ensure a global level α using a Bonferroni correction for simplicity. As a consequence the power of the test will be strongly reduced. Instead, the multi-resolution construction uses only $\log_2 n$ tests with a Bonferroni-corrected level $\alpha/\log_2 n$ ensuring a much smaller lost of power as $\log_2 128 \ll 128$.

- $\chi^2(\mu, \lambda)$ is the non-central chi-squared distribution with μ degrees of freedom and non-centrality parameter λ,
- $\bar{d}_K^{XY}$ denotes the K-th rescaled residual after projection of d^{XY}.

We consider the union-intersection *equivalence* test of the hypotheses

$$\mathcal{H}_0 = \bigcup_{K=0}^{K_0} \mathcal{H}_0^K \quad \text{versus} \quad \mathcal{H}_1 = \bigcap_{K=0}^{K_0} \mathcal{H}_1^K. \qquad (8)$$

In other words, $\mathcal{H}_1$ is accepted if and only if all $\mathcal{H}_1^K$ are.

In this setting, ideally, one would like to chose $\mathcal{H}_0^K$ as $\|\bar{d}_K^{XY}\|_n^2 \neq 0$ and $\mathcal{H}_1^K$ as $\|\bar{d}_K^{XY}\|_n^2 = 0$. Unfortunately, such hypotheses are not well separated statistically and one has to consider $\mathcal{H}_0^K$ of the form $\|\bar{d}_K^{XY}\|_n^2 > n\delta_K^2$. The quantity $n\delta_K^2$ is called *equivalence margin*. It defines the discrepancy between two unobservable true TCs that we are ready to tolerate.

As the following convergence holds

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

it means that $\mathcal{H}_1$ is rejected if the rescaled projected differences of the residuals shows some energy on any dyadic partition. In other words, two observations are considered to come from the same signal if one cannot detect a given level of energy in the rescaled residuals of their difference on any partition.

Before we go any further, we would like to comment on the use of multiple test instead of one simple test such as one based on the Euclidian distance between discretely observed TCs. The multiple test has been shown (Baraud et al., 2003) to be optimal to detect departure from the zero signal without the knowledge of neither the departure nor its Hölder regularity. It is optimal in the sense that it achieves, up to a logarithmic factor, the same rate of detection as when the regularity of the departure is known, which is naturally an easier case. Moreover, any procedure which is based on only one partition may be shown to be sub-optimal. The simple test based on the Euclidian distance between discretely observed TCs is that of $\mathcal{H}_0^K$ against $\mathcal{H}_0^K$ with $K = 0$, hence it is sub-optimal.

In order to define the p-value associated with our multiple equivalence test, we recall the following result.

Proposition 1 (Berger and Hsu, 1996). *If R_K is a rejection region at level α, the union-intersection test with rejection region $R = \bigcap_{K=0}^{K_0} R_K$ is of level α.*

Here, denoting Λ_K a random variable following a shifted $\chi^2(2^{\bar{K}}, n\delta_K^2)$ where $\bar{K}$ equals $K - 1$ if $K > 0$ and 1 otherwise, the p-value associated with the K-th test is

$$p_K(X, Y) := \mathbf{P}\left(\Lambda_K \leqslant \|\bar{D}_K^{XY}\|_n^2\right), \qquad (9)$$

and the following proposition holds.

Corollary 1. *The p-value of the union-intersection test defined by (8) with rejection region $R = \bigcap_{K=0}^{K_0} R_K$ is*

$$p(X, Y) := \max_K p_K(X, Y). \qquad (10)$$

Despite our effort, we did not find this corollary in the literature, hence we provide its proof, which is reported in Appendix B. □

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.

In the Gaussian setting, when the null is equality to zero on one given partition, it can be shown that the p-values are a monotonic transformation of an Euclidian distance. Hence, the rejection region can be defined by a distance. For the construction of Baraud et al. (2003), one can derive a multi-resolution distance as

the maximum of monotonic transformations of the p-values. In our equivalence setting, where equality to zero is the alternative, we have not been able to derive a such distance and it is not clear if a monotonic transformation of the p-values can provide such a distance due to the presence of the equivalence margins which acts as thresholds.

2.3. Clustering using equivalence test

A general setup of the clustering problem is to produce a partition of $\mathcal{X}$ into ℓ^* subsets $C_1 \ldots C_{\ell^*}$, 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 $\mathcal{H}_1$ so that the associated p-value is small.

Our bottom-up hierarchical clustering algorithm is based on one main loop that aggregates two most similar clusters at each iteration. At iteration $\bar{\ell}$, assuming we have at hand a partition $\mathcal{P}^{\bar{\ell}}$ made of $\ell = |\mathcal{X}| - \bar{\ell}$ clusters $C_1^{\bar{\ell}}, \ldots, C_{\ell}^{\bar{\ell}}$ together with indicators $\mathbb{1}_{s,s'}^{\bar{\ell}}$, $1 \leqslant s, s' \leqslant \ell$, equal to 1 if $C_s^{\bar{\ell}}$ and $C_{s'}^{\bar{\ell}}$ are neighbors (defined hereafter) and 0 otherwise. Then the new partition is obtained by merging the two clusters indexed by s_1 and s_2 such that

$$(s_1, s_2) = \arg \min_{\{(s,s') | 1 \leqslant s, s' \leqslant \ell \text{ and } \mathbb{1}_{s,s'}^{\bar{\ell}} = 1\}} p\left(C_s^{\bar{\ell}}, C_{s'}^{\bar{\ell}}\right). \qquad (11)$$

We denote the *minimum dissimilarity* at iteration $\bar{\ell}$ by

$$p(\bar{\ell}) := \min_{\{(s,s') | 1 \leqslant s, s' \leqslant \ell \text{ and } \mathbb{1}_{s,s'}^{\bar{\ell}} = 1\}} p\left(C_s^{\bar{\ell}}, C_{s'}^{\bar{\ell}}\right). \qquad (12)$$

If the minimum may be achieved on more than one couple (s, s'), we chose the one with the smallest lexical order.

The neighbors of a subset C, denoted $\mathcal{V}(C)$, is specified by the following: In the running partition, when two subsets are merged in one new cluster, the neighbors of the latter are those of both subsets. We now have to clarify the initialisation of each step.

The first step, called *local*, aims to take into account the spatial regularity existing in the images. It provides a preliminary partition made of connected clusters. It starts with the partition made of the $|\mathcal{X}|$ singletons: $\mathcal{P}^0 := \{\{x\}, x \in \mathcal{X}\}$. In this partition, two clusters are neighbors if and only if they are spatially connected on the (2D or 3D) grid $\mathcal{X}$.

The second step, called *global*, aims to recover clusters with spatially disconnected sub-structures. It use as input the output partition of the first step. In this step, all clusters are neighbors of each other; hence all indicators are always set to one.

2.3.1. Automatic selection of number of clusters

We already highlighted that to achieve proper segmentation, the main concern is the protection of the functional homogeneity inside each cluster. Therefore, the choice of the number of clusters should correspond to the expected level of homogeneity, but not the other way around. In DCE image sequences, when targeting a large area like the abdomen, the number of clusters can range from a dozen to a few thousands.

In order to automatically stop the iterations and to select a final partition, we introduce a *control function*, denoted by $c_\alpha(\ell)$, such that the iterations stop as soon as $p(\bar{\ell}) > c_\alpha(\ell)$. Starting from $\mathcal{P}^0$ with $p(0) \leqslant c_\alpha(|\mathcal{X}|)$, 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 at iteration $|\mathcal{X}| - \ell^{loc}$ with a lower value for the minimum dissimilarity as the neighborhood structure is larger. Once 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^*. The pseudo coded algorithm is given in Appendix E.

The following definition specifies the notions of functional separation for two subsets and for a partition.

Definition 1.

1. Two subsets X and Y of $\mathcal{X}$ are called "δ-separated" if their unobservable true TCs satisfy $\|d_K^{XY}\|_n^2 \geq n\delta^2$ for at least one value K, $0 \leq K \leq K_0$.
2. $\mathcal{X}$ is a "δ-partition of size ℓ" if there exists a partition of $\mathcal{X}$ into δ-separated subsets, $C_1, \ldots, C_\ell$.

We have now the right setup to provide theoretical results for our procedure. The following two theorems control the probability of, first, stopping too late − that is binding two δ-separated subsets − and, second, binding wrongly − that is binding two non-separated subsets before two δ-separated ones. Thus, their corollary provides a control of the probability to recover the exact underlying partition.

Theorem 1. *Stopping too late = binding too much - Under Assumption 1, 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*

$$c_\alpha(\ell) = \left(\frac{2\alpha}{\ell(\ell-1)}\right)^{\frac{1}{K_0+1}}, \quad \text{for every } 1 < \ell \leq |\mathcal{X}|. \tag{13}$$

The proof of Theorem 1 is given in Appendix C. □

Theorem 2. *Wrongly binding - Under Assumption 1, if $\mathcal{X}$ in a δ-partition of size ℓ_0, the probability that along the iterations two subsets of two δ-separated clusters merge before any pair of non separated subsets is lower than $\beta_n = |\mathcal{X}|^3 (n/2)^{-\kappa}$ for $\kappa > 1$ as soon as*

$$n\delta_K^2 \geq \max\left\{2^{1+\bar{K}/2}; 2(K + \kappa \log \frac{n}{2}); 1.57\sqrt{\kappa}\right\}$$

which is ensured, when $n \geq 5$, by

$$n\delta_K^2 \geq \max\left\{\sqrt{n}, 2(1 + \kappa \log 2)\log_2 \frac{n}{2}\right\}. \tag{14}$$

A direct consequence of Theorems 1 and 2 is the following result which controls the probability to achieve an exact recovery of the segmentation.

Corollary 2. Exact partitioning. *Under Assumption 1, 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.*

Let us point out that, for $n \leq 2000$ and $\kappa \geq 2$, the maximum in (14) is always achieved at the second term. If δ_K is chosen to be 1, then for $n = 100$ one can take $\kappa = 11$. In this case for a DCE image sequence with 2D-domain $\mathcal{X}$ of size 512×512 or equivalently with 3D-domain $\mathcal{X}$ of size $25 \times 100 \times 100$ the probability β_n of a wrong binding along all the iterations is less than 0.0037. For a 2D-domain of size 256×256, this probability drops to less than 5.8×10^{-5}. These values are realistic for DCE-MRI. For DCE-CT, usually n is smaller around 30. Then one has to accept larger values of δ_K to keep having a large enough κ. Naturally, these values are provided by theoretical upper-bounds, in practice, β_n may be much smaller. However, the latter theorem together with its corollary shows that for realistic DCE image sequence one can expect almost no error in the segmentation if tissue are sufficiently separated.

These two theorems do not provide an understanding of the benefit of using 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 only aggregating neighbors, the local step also offers the opportunity to take into account a possible regularity existing over the domain of $\mathcal{X}$ between TCs.

The proof of Theorem 2 is a direct consequence of the following lemma that controls the probability of making a mistake, at any step of the iteration, by binding two subsets δ-separated before two subsets sharing the same unobservable true TC.

Lemma 1. *Binding δ-separated before unseparated sets - The probability, that two δ-separated clusters merge before two subsets of one cluster do, is less than $|\mathcal{X}|^2 (n/2)^{-\kappa}$ for $\kappa > 1$ as soon as $n\delta_K^2$ satisfies one of the inequalities given in Theorem 2.*

The proof of Lemma 1 and Theorem 2 are given in Appendix D. □

The lines of proof of Lemma 1 use that K_0 −the number of individual tests which define the multiple test− is of order $\log_2 n$. Adapting this proof when the number of individual tests is 1 would lead to much larger δ_K. Thus, the classical hierarchical clustering based on the Euclidean distance should lead to very poor segmentation, not being able to recover clusters when their TCs are not sufficiently separated. This is illustrated in Appendix F on a simulated image sequence.

2.3.2. Adaptation to hierarchical clustering

In the conventional bottom-up hierarchical clustering (Ackerman and Ben-David, 2011), the linkage function is defined to ensure that the minimum dissimilarity is non-decreasing along the iterations. This is not guaranteed for HiSET so far. However, this could be fixed by considering the corrected dissimilarities $\bar{p}(.,.)$ defined as follows. Assuming C_1 and C_2 have been merged into C, then for C' in $\mathcal{V}(C)$,

$$\bar{p}(C, C') := \begin{cases} \max\{\bar{p}(C_1, C'), p(C, C')\} & \text{if } C' \in \mathcal{V}(C_1) \setminus \mathcal{V}(C_2); \\ \max\{\bar{p}(C_2, C'), p(C, C')\} & \text{if } C' \in \mathcal{V}(C_2) \setminus \mathcal{V}(C_1); \\ \max\{\min\{\bar{p}(C_1, C'), \\ \quad \bar{p}(C_2, C')\}, p(C, C')\} & \text{otherwise.} \end{cases} \tag{15}$$

Thus, the minimum dissimilarity function (with correction) becomes

$$\bar{p}(\bar{\ell}) := \min_{\{(s,s')|1 \leq s,s' \leq \ell \text{ and } \mathbb{1}_{s,s'}^{\bar{\ell}}=1\}} \bar{p}\left(C_s^{\bar{\ell}}, C_{s'}^{\bar{\ell}}\right). \tag{16}$$

The selection of the number of clusters for both local and global clustering with a corrected minimum dissimilarity function is illustrated in Fig. 1.

2.4. Parameter interpretation

So far, we have introduced three parameters: 1/ the "model" factor a in the variance stabilization transformation; 2/ the equivalence margin δ defining the homogeneity discrepancy; 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 only involved in the definition of, respectively, the minimum dissimilarity function (16) and the control function (13). Their influence on the segmentation is studied in Section 3.4.

3. Experiments and results

3.1. Material

DCE-HiSET has been implemented using C++ code wrapped in R. Computation times for our examples range from a few seconds to a few dozens of seconds when δ is decreasing. DCE-HiSET has been evaluated, first on one synthetic DCE image sequence, and second on two real DCE-MR image sequences.

The synthetic sequence made by radiologists consists of 120 images on a grid $\mathcal{X}$ of 112x112 voxels. It contains eleven clusters of
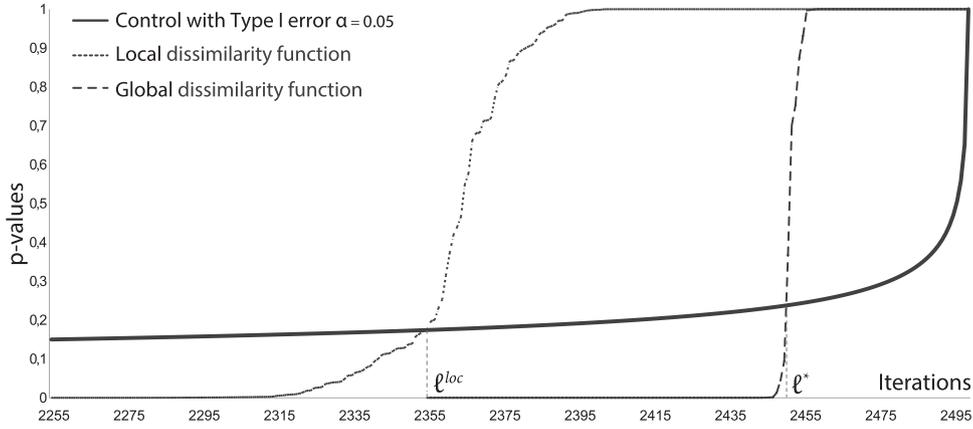
Fig. 1. Selection of the number of clusters for both local and global clustering - Solid curve: the control function (13). Dotted (resp. dashed) curve: the corrected minimum dissimilarity (16) for the local (resp. global) clustering step. When the dotted curve reaches the solid one, at iteration ℓ^{loc}, the local clustering stops. Using the resulting partition, the global clustering starts from ℓ^{loc} with the dashed curve. When the dashed curve reaches the solid one, at iteration ℓ^*, the algorithm stops, providing the final partition. Minimum dissimilarity functions are shown even after they reach the control function to illustrate their typical behavior.
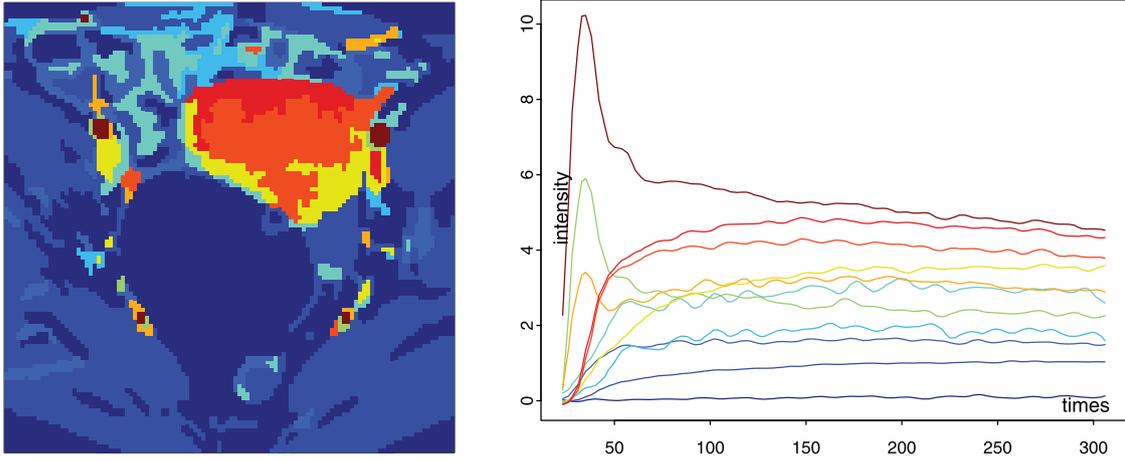

Fig. 2. Synthetic DCE image sequence: (left) The ground-truth segmentation of $\mathcal{X}$; (right) The true enhancement curves, $i^x(t)$, associated with the 11 clusters using corresponding colors.

various sizes and shapes, representing real anatomical structures and their complexity, see Fig. 2. Independent standard Gaussian noises are added to the intensities at each time and at every voxel.

The real DCE-MR image sequences of two 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, acquired over 305 seconds, see Fig. 3. With the delayed injection of the contrast agent bolus, their n_0 ($n_0 = 12$ and 10) first images show only the grey level baseline up to the noise level. In both sequences, a ROI around the tumor has been manually drawn by an experienced radiologist after acquisition.

Each real DCE-MR image sequence is handled as follows:

1. The original intensity $\Phi^x(t_j)$, for each time t_j and voxel x, is processed as in (4) to obtain $I^x(t_j)$.
2. The baseline intensity, b^x, of each voxel x is estimated by averaging the n_0 first intensities during the baseline phase, hence providing $\hat{b}^x$.
3. The baseline intensities are removed from the intensities to obtain the enhancements after baseline phase. For simplicity, the enhancements are also denoted $I^x(t_j)$:

$$I^x(t_j) \leftarrow \frac{I^x(t_j) - \hat{b}^x}{\sqrt{1 + 1/n_0}}, \quad \text{for } j = n_0 + 1, \ldots, n.$$

These are Gaussians with a variance equal to 1 and are independent as soon as Assumption 1 holds for $I^x(t_j)$.

4. The image sequence is segmented from the $I^x(t_j)$ for $j = n_0 + 1, \ldots, n$ and $x \in \mathcal{X}$.
5. After segmentation, for each cluster C, the average intensity $\bar{I}^C$ is computed from (6) and normalized residuals are defined as

$$\hat{\xi}^x_j := \frac{I^x(t_j) - \bar{I}^C(t_j)}{\sqrt{1 - 1/|C|}}, \quad \text{for } j = n_0 + 1, \ldots, n$$

for all x in cluster C. These follow Gaussian distribution $\mathcal{N}(0, 1)$ as soon as Assumption 1 holds for $I^x(t_j)$.

The synthetic sequence directly provides 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.

3.2. Competitors

To compare DCE-HiSET on the synthetic image, we considered the following 6 competitors from the 3 categories previously discussed in Section 1.4:

- model-based: k-means, HMRF-FCM (Chatzis and Varvarigou, 2008) and mean shift (MS) (Comaniciu and Meer, 2002);
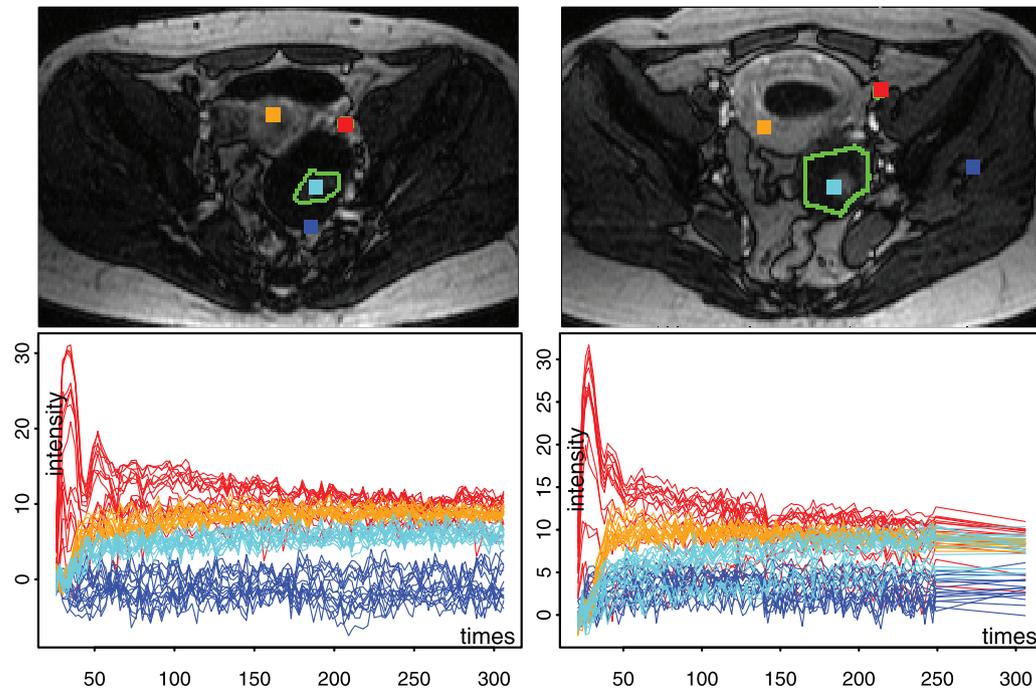- graph-based: normalized cut (NC) (Shi and Malik, 2000);

Fig. 3. DCE-MRI image sequences of two female pelvis with ovarian tumors - each column shows one sequence. Top: image obtained at time t_{30} (*after arterial phase*) with the tumor ROI (*green*) together with four 4x4 squared neighborhoods (*red, 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 $a = 0.45$. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

- hybrid: MS followed by normalized cut (MS-NC) (Tao et al., 2007), SLIC (Achanta et al., 2010; Irving et al., 2014) followed by normalized cut (SLIC-NC) (Irving et al., 2016a).

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

Competitor details

All competitors, except for MS, require the number of clusters as input parameter, and we implemented them according to the recommendations of their authors.

k-means was used with Euclidean distance between TCs. 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 random initializations and picked the one with the best value of the objective function.

For other competitors, a PCA decomposition of the features has been used first with 3 eigenvectors then 6. Three was enough to explain 95% of the variance; 6 corresponds to the number of tests used in our multiple equivalence test for the sake of fairness. We observe no significant difference between these two choices and report results only with six eigenvectors.

HMRF-FCM requires an input partition and is highly sensitive to its choice. It uses the *k*-means result as input and then optimizes 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).

MS-NC, NC uses the output of MS as input; therefore it can only reduce the number of clusters from the MS step.

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

When reviewing the literature, we realized that not only were some methods not available as a package but that some details on the parameter selection were also missing. To compensate for the latter, we used an Oracle approach: for each method and each parameter, the range of the latter had been manually chosen to cover

under- and over-fitting. Inside these ranges, the parameters were chosen to provide the best value of the objective function. This approach led to the following parameter ranges.

- when required, the number of clusters was allowed to vary from 7 to 15.
- the multiplicative constant ε for the entropy penalty controlling the fuzziness in HMRF-FCM was allowed to vary from 0.2 to 4 by increments of 0.1.
- MS and MS-NC use a temporal and a spatial bandwidth, denoted respectively bw_t and bw_s. We have let bw_t vary from 0.03 to 0.1 by increments of 0.01 and $r_{ts} := bw_t/bw_s$ from 3 to 15 by increments of 0.5.
- SLIC uses a size and a compactness of supervoxels, denoted respectively sv_s and sv_c. We have let sv_s vary from 3 to 10 by step of 0.5 and sv_c from 0 to 0.05 by increments of 0.005.

Competitor implementations

All competitors and DCE-HiSET were run on the same laptop (macbook pro with i5 core 2.5 Ghz with 16 GB of RAM) to provides comparable running times.

We used the *k*-means implementation provided in R[5] by the function kmeans. Depending on the number of clusters, the computation time ranged from 10 to 90 min for the 250 runs.

We were not able to find available code for HMRF-FCM. Therefore, we used a C-implementation wrapped in R with computational time varying from 1.5 to 10 min 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[6] in Matlab[7] with computation time varying from 0.3 s to 33 s, depending on the values of bandwidths. The C-implementation of NC[8] wrapped in Matlab

---

[5] https://cran.r-project.org/

[6] http://a-asvadi.ir/

[7] http://www.mathworks.com

[8] http://www.timotheecour.com/

resulted in computation time always less than 1 s. The same held for SLIC[9] using its *C*-implementation wrapped in Matlab with computation time depending only 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.

3.3. Evaluation on the synthetic sequence

Evaluation criterion For the synthetic sequence, the accuracy of segmentation results has been measured by the Fowlkes–Mallows Index (FM) (Fowlkes and Mallows, 1983),

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

where the number of voxel pairs is denoted N_{11} when both voxels are classified into the same cluster in both partitions; N_{10} when they are in the same cluster of the first partition but two different clusters in the second partition; N_{01} when they are in two different clusters in the first partition but the same cluster in the second partition. This index lies between 0 and 1: the closer it is to 1, the more similar the segmentations. It accounts for the proportion of pairs of voxels which are in the same cluster in both partitions with respect to the geometric average of being at least in a same cluster for one partition. That is the proportion of well classified pairs by both methods with respect to the average number of pairs which could be well classified by at least one method. Here the use of the geometrical mean instead of the classical arithmetical mean provides more weight to the largest value. As already noted, keeping the cluster homogeneous with respect to the time intensity curve is the top concern, especially as the number of clusters cannot be considered as known. In this context, Fowlkes-Mallows Index provides a good trade-off for measuring how partitions show the same type of results or not. It is indeed a better end-point than the number of clusters given that 1/ in our context, it is unknown; 2/ two partitions can have the same number of clusters and show strong variations in the construction of their clusters.

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

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 *error-map* as the indicator function of the set

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

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

While the Fowlkes-Mallows index accounts for pairs of similar voxels which are or not in the same cluster of the second partition, the error map counts for voxels of each cluster A (in the first partition) which are not in the *best* cluster B (in the second partition), where the best cluster B is defined as to cover the most of A and vice-versa. Hence, together they offer two different points-of-view on how close two partitions are. We considered more indices; however they all showed similar behavior to one of the two described above. Therefore, for sake of simplicity, we choose the ones with intuitive interpretation.

Behavior of DCE-HiSET We used DCE-HiSET with a fixed value of $\alpha = 0.001$ 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 Fig. 4. We observed

that the FM and wFM indexes achieve their maximum for the same value, denoted δ^*. Moreover, they remain stable and higher than 80% in a large range of δ from 0.5 to 2. The choice of a fix value for α is justified further in Section 3.4.1 as this parameter has only a small influence on the final result with respect to δ. Indeed, it only controls the probability that the final partition is too small and not the size of the partition itself.

Fig. 5 shows the segmentation results with 3 values of δ, including δ^*. Clearly, small values of δ result in over-segmentation (ℓ^* too large): large homogeneous regions are split into sub-regions with irregular borders; however, these sub-regions remain large. In contrast, large values of δ enforce under-segmentation: true clusters start to merge; however the border geometry is not changed, ensuring a good recognition of most of the structures. Nevertheless, as one can expect when thinking of bias-variance trade-off, it is clearly more risky to use too small values than too large ones.

For a fully automatic procedure with automatic selection of δ 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 $\delta^* = 2\delta_0$.

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 the true enhancement curves by 2/3 and 4/3, while the noise level was kept fixed with a standard deviation equal to 1. In this setting, both the separation distance and the SNR are affected by the multiplicative factor and the segmentation is all the harder the smaller the multiplicative factor becomes. Fig. 6 shows that both FM and wFM indexes achieve optimal value for the same value δ^*, indicating that the key factor is indeed the ratio between the separation distance and the SNR. Clearly, the harder the segmentation, the smaller is the range around the optimal value when relaxing the index maximization. The error maps corresponding to δ^* are given in Fig. 7.

Comparison with competitors Information and Oracle performances of all competitors together with DCE-HiSET are summarized in Table 1.

For *k*-means, Fig. 8 shows the result when $k = \ell^* = 11$ that is the true number of clusters, together with the best FM and wFM values achieved when *k* varies. The weaknesses of *k*-means are clearly visible: on one hand, having two or more initial centers in one true cluster (see $\ell^* = 10$ or 11 in Fig. 8) results in pulverized clusters (known as the *pepper and salt* effect); on the other hand, when no center reaches small clusters, the latter are artificially merged into other clusters, leading to a high FM but low wFM (see $\ell^* = 8$ in Fig. 8).

If the initial partition given by *k*-means has no pulverized cluster, HMRF-FCM barely improves the segmentation of *k*-means. Otherwise, it regularizes the pulverized clusters by grouping their voxels into a "panther texture" when ε becomes large enough (see Fig. 8). However (not presented here), if one cluster has already been split in the initial partition provided by *k*-means, HMRF-FCM is unable to fix it. Thus, the increase of FM and wFM induced by HMRF-FCM is very small for a given *k*. Moreover, the best FM and wFM indexes, when *k* and ε vary, are unchanged with respect to *k*-means. For each of MS, MS-NC and SLIC-NC (see Fig. 9), as for DCE-HiSET, the best results for both indexes are achieved at same parameter sets, showing a good and expected property of stability with respect to the size of the clusters. MS recovered the right number of clusters. High FM (0.994) and relatively low wFM (0.916) indicate that mistakes have been made in small clusters. MS-NC recovers only 10 clusters with worse indexes than for MS alone. Indeed, two clusters resulting from the best result of MS have merged during the NC step. SLIC-NC has neither pulverized
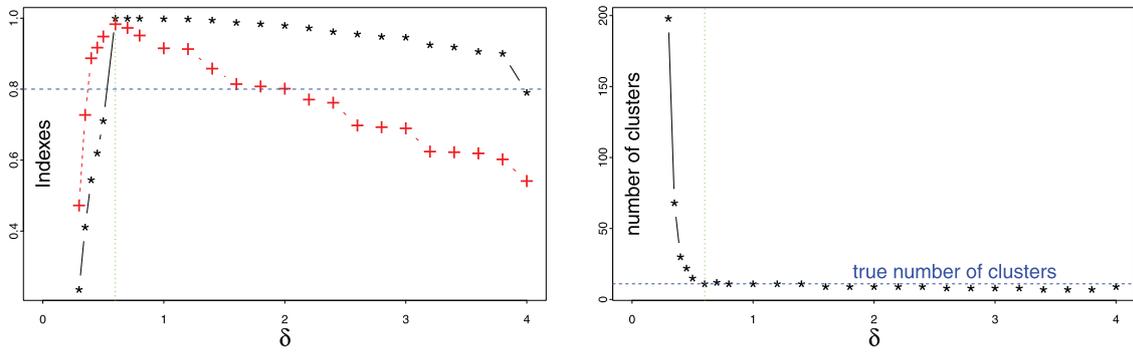
---

[9] http://www.vlfeat.org/

Fig. 4. 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). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)
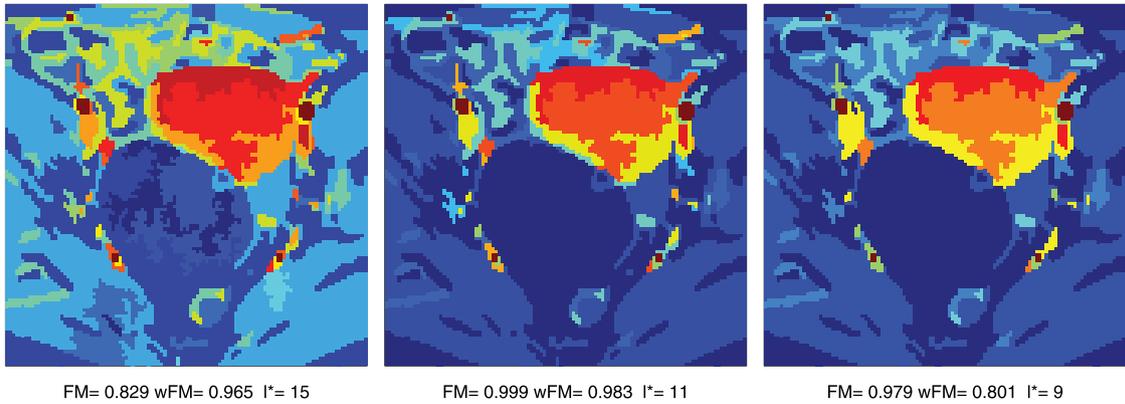

FM= 0.829 wFM= 0.965 I*= 15   FM= 0.999 wFM= 0.983 I*= 11   FM= 0.979 wFM= 0.801 I*= 9

Fig. 5. 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 $\alpha = 0.001$.
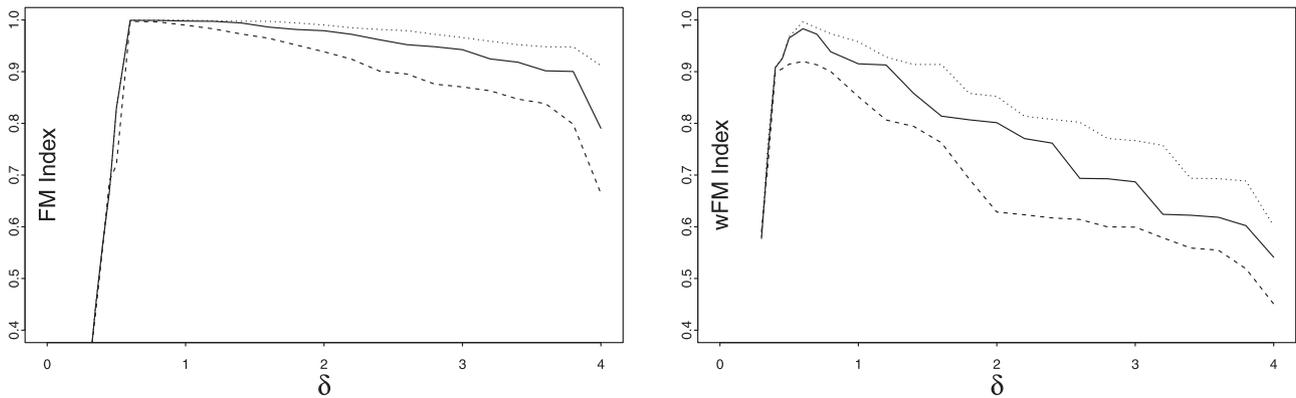

Fig. 6. FM index (left) and weighted FM index (right) when true enhancement are multiplied by 4/3 (dotted), 1 (solid) and 2/3 (dashed).
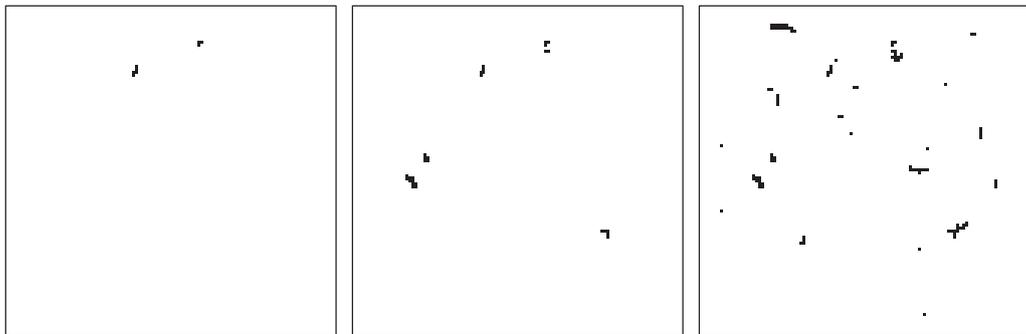

Fig. 7. Error maps for wFM-optimal value of δ when true enhancement are multiplied by 4/3 (left), 1 (middle) and 2/3 (right).

Table 1

Fowlkes–Mallows index with its weighted version: for each competitor, the input parameters are provided with their associated used range. The Oracle, given in parentheses, achieves the highest value of the index over the parameter ranges.

Method	Parameter	Range by Step	Highest FM	Highest wFM
k-means	k	7–15 by 1	0.993 ($k = 8$)	0.827 ($k = 10$)
HMRF-FCM	k	7–15 by 1	0.993 $\left(\begin{array}{c} k = 8 \\ \varepsilon = 0.5 \end{array}\right)$	0.827 $\left(\begin{array}{c} k = 10 \\ \varepsilon = 1 \end{array}\right)$
	ε	0.2–4 by 0.5		
MS	bw_t	0.03–0.1 by 0.01	0.994	0.919
	r_{ts}	3–14 by 0.5	($bw_t = 0.06$, $r_{ts} = 11$)
MS-NC	k	7–15 by 1	0.993	0.852
	bw_t	0.03–0.1 by 0.01	($k = 10$, $bw_t = 0.06$, $r_{ts} = 11$)
	r_{ts}	3–14 by 0.5		
SLIC-NC	k	7–15 by 1	0.983	0.821
	sv_s	3–10 by 0.5	($k = 12$, $sv_s = 3$, $sv_c = 0$)
	sv_c	0–0.05 by 0.005		
DCE-HiSET	δ	0.2–4 by 0.1	**0.999** ($\delta = 0.6$)	**0.983** ($\delta = 0.6$)

FM = 0.714 wFM = 0.802 l* = 11   FM = 0.993 wFM = 0.759 l* = 8   FM = 0.855 wFM = 0.827 l* = 10

lda=0.2: FM=0.714 wFM =0.803 l*=11   lda=0.5: FM=0.72 wFM =0.806 l*=11   lda=2: FM=0.723 wFM =0.805 l*=11
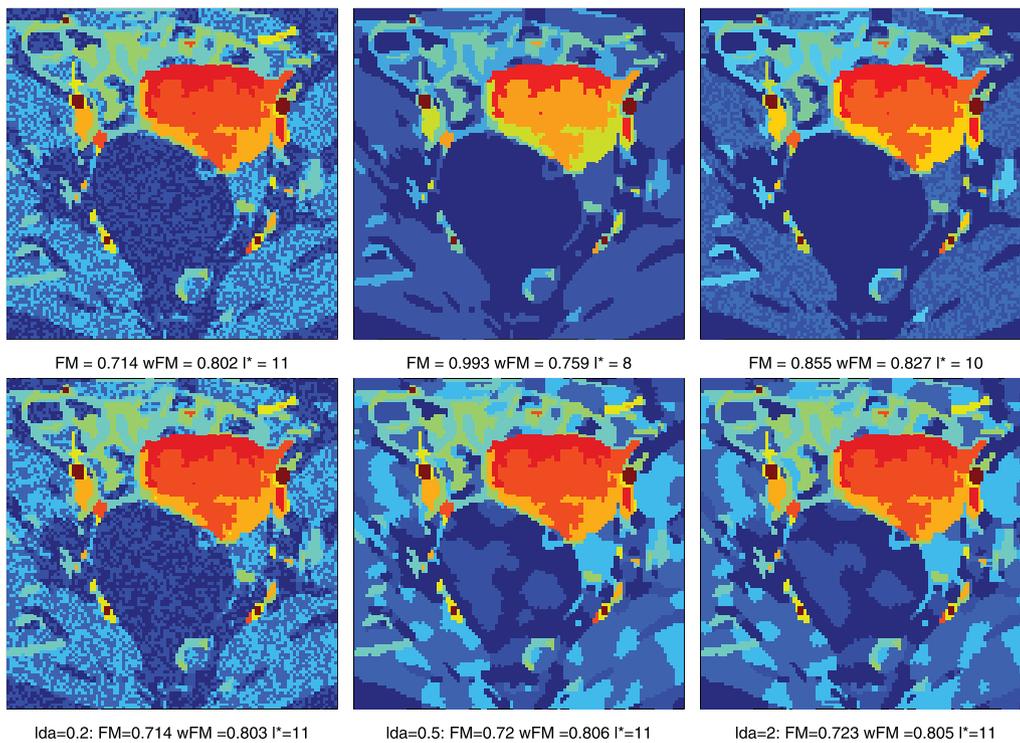
Fig. 8. Oracle results - Top: Best result of k-means among 250 runs with true number of clusters and with: $k = 11$ (left); highest FM index (middle); highest wFM index (right). Bottom: Results of HMRF-FCM initialized by the result of k-means with $k = 11$ and with: $\varepsilon = 0.2$ (left); $\varepsilon = 0.5$ (middle); $\varepsilon = 2$ (right).
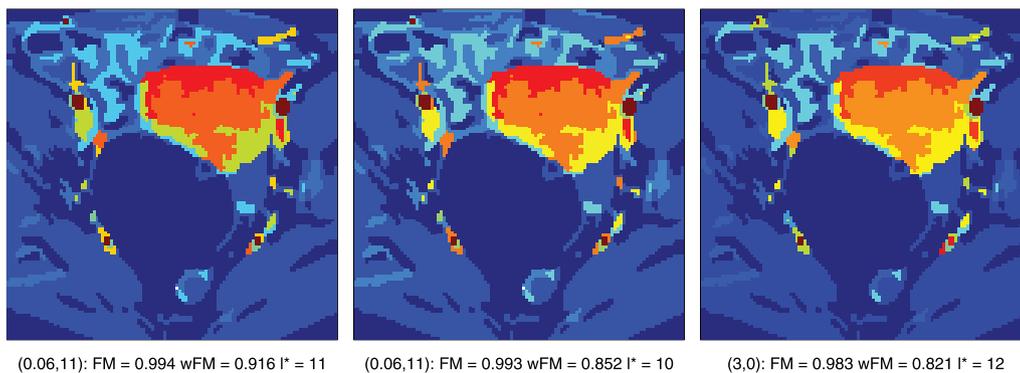

(0.06,11): FM = 0.994 wFM = 0.916 l* = 11   (0.06,11): FM = 0.993 wFM = 0.852 l* = 10   (3,0): FM = 0.983 wFM = 0.821 l* = 12

Fig. 9. Oracle results - MS (left); MS-NC (middle); SLIC-NC (right).
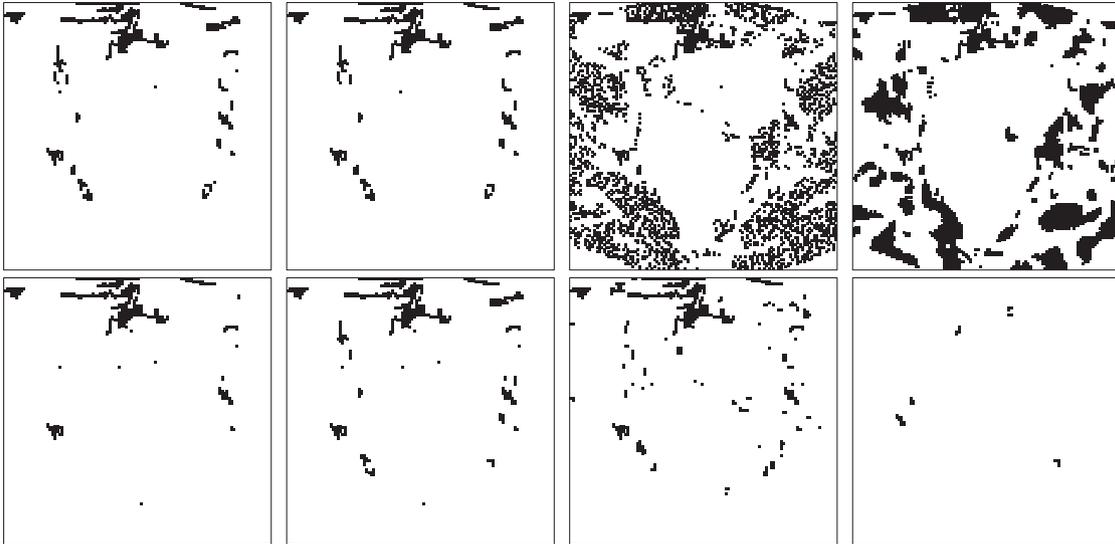
mentation as DCE-HiSET does.

Fig. 10. Error maps - From left to right: (top) *k*-means and HMRF-FCM with highest FM; *k*-means and HMRF-FCM with highest wFM; (bottom) best results of MS, MS-NC, SLIC-NC (here the best results are achieved on same parameter sets for both indexes) and the (adaptive) result of DCE-HiSET.

issues, as with *k*-means, nor mistakes in small clusters, as with MS, thanks to the first step with SLIC. However, due to the supervoxel size limitation in SLIC, voxels in small clusters may be spread in the surrounding large clusters and/or narrow parts of clusters tend to be thicker.

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

Fig. 10 shows the error maps of the Oracle of each competitor together with the error map of DCE-HiSET to compare how segmentation errors (misclassified voxels) are distributed.

From this study of the synthetic sequence, we see clearly that DCE-HiSET outperforms all competitors with respect to both FM and wFM indexes while using only a single parameter. DCE-HiSET shows moreover a very stable behavior. The error maps show clearly that all competitors have trouble with small and/or complex clusters while DCE-HiSET makes 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.

3.4. Experiment on real DCE-MRI sequences

3.4.1. Parameter influence

Fig. 11 shows the influence of parameters α and δ on the segmentation size. Clearly α has a negligible effect on the partition size compared to that of δ, in both local and global steps. Hence, we fixed $\alpha = 0.001$ for the rest of this study. Of note, we also observed that the cluster shape is not much influenced by α either; α only controls the probability of stopping too late and not how clusters are built along the iterations.

3.4.2. Model verification and selection of parameter a

In order to justify our model given by Eqs. 2–(5), we studied the distribution of the normalized residuals $\hat{\xi}_j^x$ when δ and a vary (see Fig. 12). We observed that the distribution tails of $\hat{\xi}_j^x$ are heavier than those of a Gaussian distribution when a is large (Poisson's case: $a = 0.5$) and are lighter when a is too small ($a = 0.4$); this remains true for all values of δ. For the intermediate value

$a = 0.45$, the distribution shows a behavior close to Gaussian (or at least sub-Gaussian i.e. having only Gaussian tails). In this case, δ becomes the proper tuning parameter to obtain a residual distribution close to (sub-) Gaussian. Let us point out that, thanks to the use of tests, only tails are of interest. Extension of our construction to the sub-Gaussian case would require many more theoretical developments beyond the scope of this paper. Similarly, adaptation with respect to a or improving the model to get a better variance stabilization would be an objective of future investigation. Nevertheless, the quality of our empirical experiments shows that the benefit of this will be probably small.

We continue this study with a fixed value $a = 0.45$.

3.4.3. Segmentation results

The two real DCE-MR image sequences have been fully segmented into homogeneous regions that are highly consistent with anatomical structures as shown in Fig. 13. The smaller δ is, the more details in anatomical structures can be observed in image sequence. This is highlighted (see Fig. 14) in the ROIs defined by the squares that contain the manually segmented ROIs, shown in black on Fig. 13. The corresponding estimated TCs inside each cluster within the ROI together with their corresponding size are shown in Fig. 15. By segmenting the full DCE sequence, these estimated TCs are obtained by averaging TCs which do not necessarily belong to the ROI but do belong to the same homogeneous cluster. As a benefit, the SNR observed for these estimated TCs is strongly improved, providing a real opportunity for further analysis and comparisons. From these figures, one can clearly understand the advantages of DCE-HiSET. 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 the functional discrepancy between them.

4. Discussion

In this paper, we proposed a new method, DCE-HiSET, for segmenting the DCE image sequence into homogeneous regions with respect to the TC observed in each voxel. With this approach, the low SNR of DCE image sequence is significantly improved by averaging the TCs of voxels in homogeneous regions.
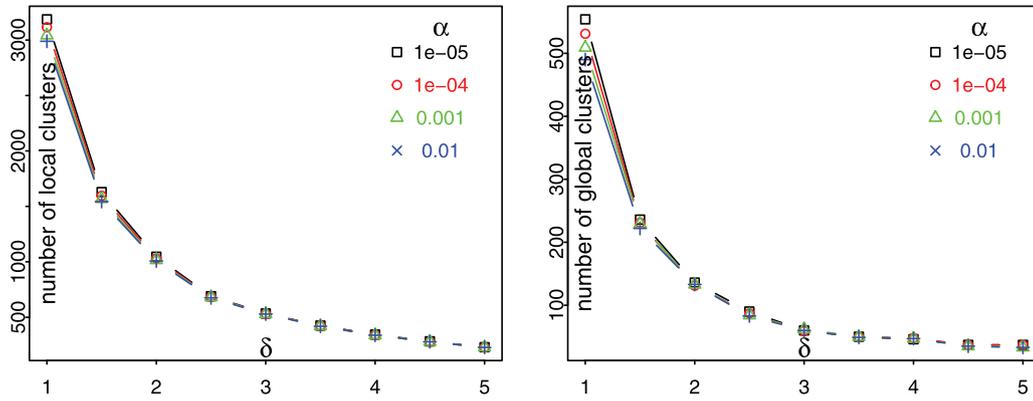
Fig. 11. Number of clusters when α and δ vary for (*left*) local and (*right*) global steps.
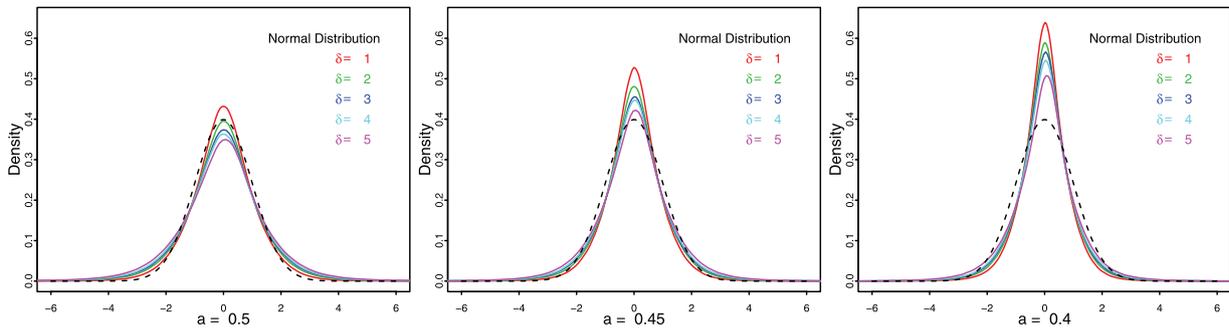

Fig. 12. Normalized residuals - estimated densities for $\delta \in \{1, 2, 3, 4, 5\}$ and normal density as reference (black and dashed) when using a variance stabilization with $a=0.5$ (left); 0.45 (middle); 0.4 (right).
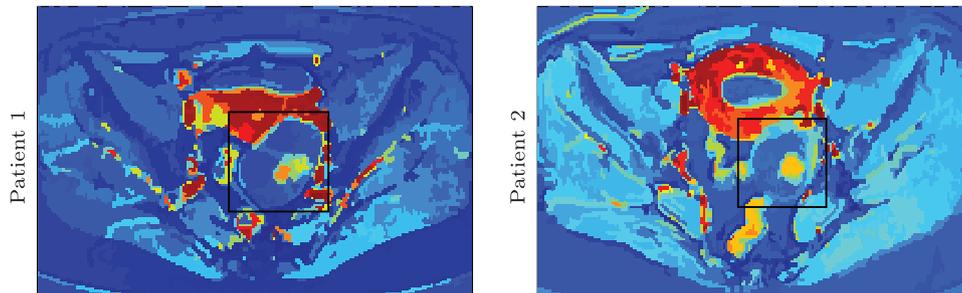

Fig. 13. Segmentation results of the two real DCE-MR image sequences using $\delta = 3$.

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 up of functionally piecewise homogeneous regions, our algorithm is proven theoretically 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, highly consistent with real anatomy, regardless of their shape, size and disconnectedness. The total computation complexity is controlled during the local step by the neighborhood structure, and during the global step by the small size of the partitions.

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 and error map. 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- sequences. However, in 3D, slice thickness and/or distance between slices can be very different from the voxel size in the imaging cross-section. These differences can be taken into account 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 optimizing of a simple criterion, which can provide a fully automatic procedure for each DCE image sequence. Used in conjunction with a registration tool (Glocker et al., 2011; Sotiras et al., 2009), DCE-HiSET can provide the right segmentation tool of an iterative registration/segmentation/labelling process during which, along the iterations, both the estimated signals and the registration are learned.
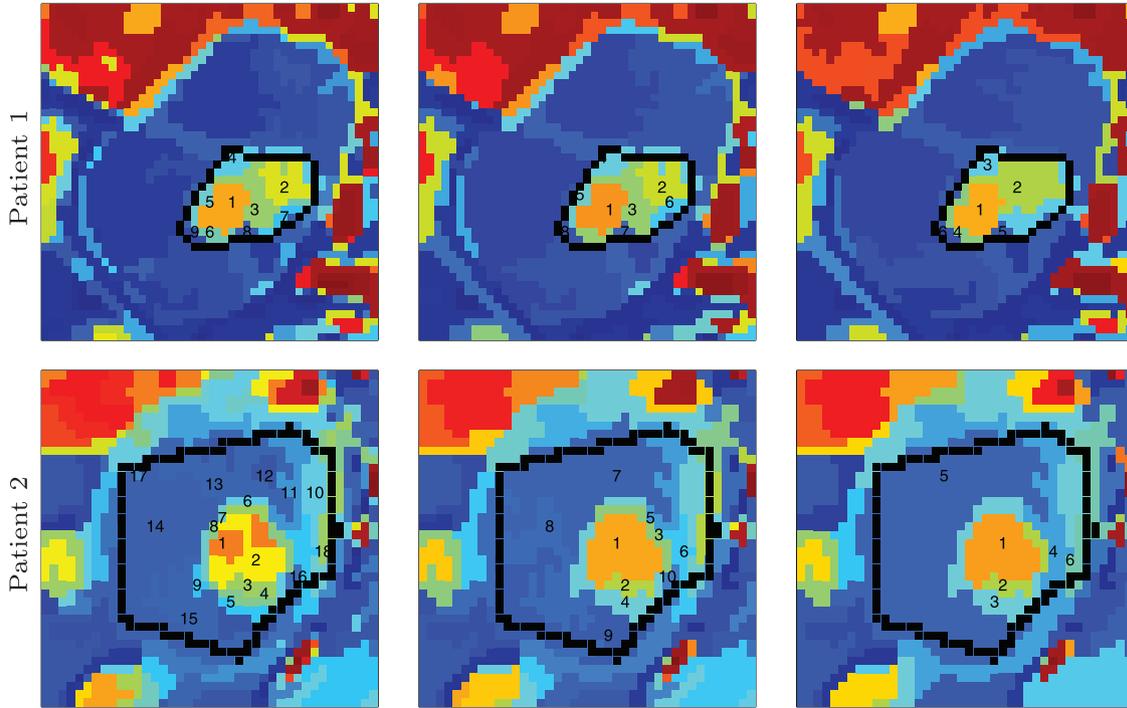
Fig. 14. Squared zoom in the DCE-MR image sequences with δ equals 2 (left); 3 (middle); 4 (right). The manually segmented ROI appears in black and clusters inside are numbered. The first sequence (top) shows from 9 to 6 clusters. The second (bottom) from 18 to 6.
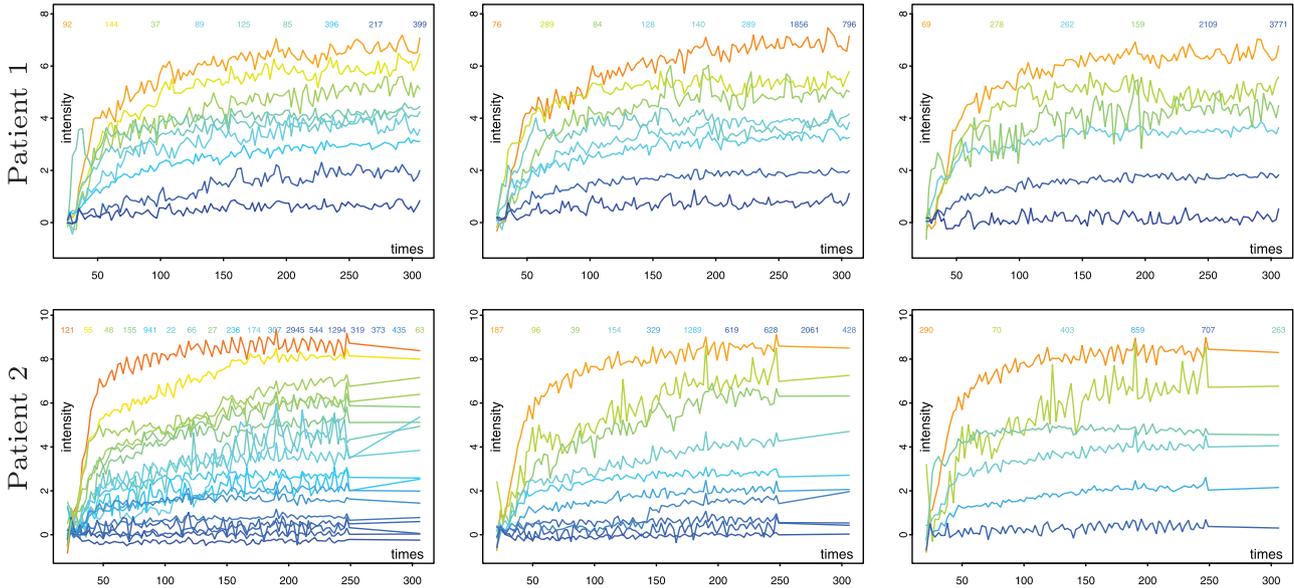

Fig. 15. Average TCs (after variance stabilization) in the clusters shown inside the manually segmented ROIs (see Fig. 14) with δ=2, 3 and 4 from left to right. Size of the corresponding clusters are given at the top of each subfigure with corresponding colors.

Finally, by only adapting the multiple equivalence test, HiSET can be applied to various types of models where a signal is observed on a spatial field, such as multimodality images, but also such as electrical consumptions observed in a region or a country.

Appendix A. Dyadic decomposition of the time indexes

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

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

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 fash-

ion:

$$\mathcal{T}^K := \left\{ T_k^K = \bigcup_{(k-1)/2^K < r/2^{k_0} \leqslant k/2^K} T_r, \; k = 1, \ldots, 2^K \right\}.$$

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

$$\Pi_K D^{XY} := (\underbrace{m_1^K, \ldots, m_1^K}_{|T_1^K|}, \ldots, \underbrace{m_{2^K}^K, \ldots, m_{2^K}^K}_{|T_{2^K}^K|}), \quad m_k^K = \frac{1}{|T_k^K|} \sum_{j \in T_k^K} D^{XY}(t_j).$$

For $K \geq 1$,

$$D_K = (\underbrace{m_1^K - m_1^{K-1}, \ldots, m_1^K - m_1^{K-1}}_{|T_1^K|}, \underbrace{m_2^K - m_1^{K-1}, \ldots, m_2^K - m_1^{K-1}}_{|T_2^K|}, \ldots,$$

$$\underbrace{m_{2^K-1}^K - m_{2^{K-1}}^{K-1}, \ldots, m_{2^K-1}^K - m_{2^{K-1}}^{K-1}}_{|T_{2^K-1}^K|}, \underbrace{m_{2^K}^K - m_{2^{K-1}}^{K-1}, \ldots, m_{2^K}^K - m_{2^{K-1}}^{K-1}}_{|T_{2^K}^K|}).$$

Under Assumption 1, the variance of $m_{2k-i}^K - m_k^{K-1}$, with $k = 1, \ldots, 2^{K-1}$ and $i = 0, 1$ is

$$(\sigma_{2k-i}^K)^2 = \frac{1}{|T_{2k-i}^K|} + \frac{1}{|T_k^{K-1}|} = \frac{|T_{2k-i}^K| + |T_k^{K-1}|}{|T_{2k-i}^K||T_k^{K-1}|},$$

and we consider the deterministic diagonal matrix $\Sigma_K^{1/2}$ with diagonal

$$\left(\underbrace{[\sigma_1^K |T_1^K|]^{-1}, \ldots, [\sigma_1^K |T_1^K|]^{-1}}_{|T_1^K|}, \underbrace{[\sigma_2^K |T_2^K|]^{-1}, \ldots, [\sigma_2^K |T_2^K|]^{-1}}_{|T_2^K|}, \ldots, \right. \tag{A.1}$$

$$\left. \underbrace{[\sigma_{2^K-1}^K |T_{2^K-1}^K|]^{-1}, \ldots, [\sigma_{2^K-1}^K |T_{2^K-1}^K|]^{-1}}_{|T_{2^K-1}^K|}, \underbrace{[\sigma_{2^K}^K |T_{2^K}^K|]^{-1}, \ldots, [\sigma_{2^K} |T_{2^K}^K|]^{-1}}_{|T_{2^K}^K|} \right)$$

together with the vector $\bar{D}_K = \Sigma_K^{1/2} D_K$ called K-th rescaled residual after projection. Similarly, for $K = 0$, we have

$$D_0 = (\underbrace{m_1^0, \ldots, m_1^0}_{n})$$

and, as the variance of m_1^0 is $1/n$, we define $\bar{D}_0 := D_0$. The $\bar{D}_K$ are independent as the D_K are. Denoting by $\|u\|$ the Euclidian norm of a vector u of $\mathbb{R}^n$, the normalisation ensures that for $K \geq 1$

$$\|\bar{D}_K\|_n^2 = \sum_{k=1}^{2^{K-1}} \left(\left[\frac{m_{2k-1}^K - m_k^{K-1}}{\sigma_{2k-1}^K} \right]^2 + \left[\frac{m_{2k}^K - m_k^{K-1}}{\sigma_{2k}^K} \right]^2 \right)$$

$$= \sum_{k=1}^{2^{K-1}} \left(\sqrt{\frac{|T_{2k-1}^K||T_k^{K-1}|}{|T_{2k-1}^K| + |T_k^{K-1}|}} (m_{2k-1}^K - m_k^{K-1})^2 \right.$$

$$\left. + \sqrt{\frac{|T_{2k}^K||T_k^{K-1}|}{|T_{2k}^K| + |T_k^{K-1}|}} (m_{2k}^K - m_k^{K-1})^2 \right).$$

taking into account the 2^{K-1} linear links, for $k = 1, \ldots, 2^{K-1}$

$$(|T_{2k-1}^{K-1}| + |T_{2k}^K|) m_k^{K-1} = |T_{2k-1}^K| m_{2k-1}^K + |T_{2k}^K| m_{2k}^K, \tag{A.2}$$

it follows that

$$\|\bar{D}_K\|_n^2 \sim \chi^2(2^{K-1}, \|\bar{d}_K\|_n^2)$$

where

- $\chi^2(\mu, \lambda)$ is the non-central chi-squared distribution with μ degrees of freedom and non-centrality parameter λ,
- $\bar{d}_K$ denotes the Kth residual after projection of $\bar{d} = \mathbb{E}(D)$.

For $K = 0$, we have moreover $\|\bar{D}_0\|_n^2 \sim \chi^2(1, \|\bar{d}_0\|_n^2)$.

Appendix B. Proof of Corollary 1

Given an observation, the p-value is the largest significance level α such that $\mathcal{H}_0$ is accepted. Considering rejection region for the IUT of the form $R^\alpha = \bigcap_K R_K^\alpha$, the p-value of the IUT (8) becomes

$$p(X, Y) = \sup\{\alpha \mid D^{XY} \notin R^\alpha\} = \sup \left\{ \alpha \mid D^{XY} \notin \bigcap_K R_K^\alpha \right\}$$

$$= \sup \left\{ \alpha \mid \bigcup_K \left\{ D^{XY} \in \bar{R}_K^\alpha \right\} \right\} = \sup \left\{ \alpha \mid \exists K \text{ with } \underbrace{D^{XY} \in \bar{R}_K^\alpha}_{p_K(X,Y) \geqslant \alpha} \right\}$$

$$= \sup \left\{ \alpha \mid \alpha \leqslant \max_K (p_K(X, Y)) \right\} = \max_K p_K(X, Y).$$

Appendix C. Proof of Theorem 1 - Stopping too late

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 $1 \leq i < j \leq \ell$ there exists $K \in \{0, \ldots, K_0\}$ for two clusters C_i and C_j such that they are δ-separated:

$$\mathbf{P}_{\mathcal{H}_0}(p(\bar{\ell}) \leqslant c_\alpha(\ell)) = \mathbf{P}_{\mathcal{H}_0} \left[\min_{1 \leqslant i < j \leqslant \ell} \left\{ p(C_i, C_j) \right\} \leqslant c_\alpha(\ell) \right]$$

$$= \mathbf{P}_{\mathcal{H}_0} \left[\bigcup_{1 \leqslant i < j \leqslant \ell} \left\{ p(C_i, C_j) \leqslant c_\alpha(\ell) \right\} \right]$$

$$= \mathbf{P}_{\mathcal{H}_0} \left[\bigcup_{1 \leqslant i < j \leqslant \ell} \left\{ \max_K \{ p_K(C_i, C_j) \} \leqslant c_\alpha(\ell) \right\} \right]$$

$$\leqslant \sum_{1 \leqslant i < j \leqslant \ell} \mathbf{P}_{\mathcal{H}_0} \left[\max_K \{ p_K(C_i, C_j) \} \leqslant c_\alpha(\ell) \right]$$

$$= \sum_{1 \leqslant i < j \leqslant \ell} \mathbf{P}_{\mathcal{H}_0} \left[\bigcap_K \{ p_K(C_i, C_j) \leqslant c_\alpha(\ell) \} \right]. \tag{C.1}$$

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

$$\mathbf{P}_{\mathcal{H}_0}(p(\bar{\ell}) \leqslant c_\alpha(\ell)) \leqslant \frac{\ell(\ell-1)}{2} \prod_K \mathbf{P}_{\mathcal{H}_0}[p_K(C_1, C_2) \leqslant c_\alpha(\ell)]$$

$$= \frac{\ell(\ell-1)}{2} (c_\alpha(\ell))^{K_0+1}. \tag{C.2}$$

Controlling the probability of false merge by α leads to

$$\frac{\ell(\ell-1)}{2} (c_\alpha(\ell))^{K_0+1} = \alpha,$$

and we deduce thereby

$$c_\alpha(\ell) = \left(\frac{2\alpha}{\ell(\ell-1)} \right)^{\frac{1}{K_0+1}}.$$

Appendix D. Proof of Theorem 2 - Wrongly binding

The proof of Theorem 2 is a consequence of the Lemma 1 which control the probability to bind two δ-separated clusters before non separated ones at a given iteration of the hierarchy.

Proof of Lemma 1. We recall that $\bar{K} := K - (K > 0)$. Assuming that we have at hand a partition $C_1, \ldots, C_\ell$, we consider the sets of couples (j, j') with $1 \leq j < j' \leq \ell$ such that C_j and $C_{j'}$

- have same intensity that we denote $\mathcal{S}_0$,
- are δ-separated that we denote $\mathcal{D}_\delta$. $\square$

We are looking for an upper-bound for the probability to have two δ-separated subsets merging before any two pair of subsets having same intensity, that is

$$P\left[\min_{(j,j')\in\mathcal{S}_0} p(C_j,C_{j'}) > \min_{(j,j')\in\mathcal{D}_\delta} p(C_j,C_{j'})\right]$$

$$\leq P\left[(\min_{(j,j')\in\mathcal{S}_0} p(C_j,C_{j'}) > \min_{(j,j')\in\mathcal{D}_\delta} p(C_j,C_{j'}))\bigcap(\min_{(j,j')\in\mathcal{S}_0} p(C_j,C_{j'}) \leq \varepsilon)\right]$$

$$+ P\left[(\min_{(j,j')\in\mathcal{S}_0} p(C_j,C_{j'}) > \min_{(j,j')\in\mathcal{D}_\delta} p(C_j,C_{j'}))\bigcap(\min_{(j,j')\in\mathcal{S}_0} p(C_j,C_{j'}) > \varepsilon)\right]$$

$$\leq P\left[(\varepsilon > \min_{(j,j')\in\mathcal{D}_\delta} p(C_j,C_{j'}))\bigcap(\min_{(j,j')\in\mathcal{S}_0} p(C_j,C_{j'}) \leq \varepsilon)\right]$$

$$+ P\left[\min_{(j,j')\in\mathcal{S}_0} p(C_j,C_{j'}) > \min_{(j,j')\in\mathcal{D}_\delta} p(C_j,C_{j'})\Big|\min_{(j,j')\in\mathcal{S}_0} p(C_j,C_{j'}) > \varepsilon\right]$$

$$\times P\left[\min_{(j,j')\in\mathcal{S}_0} p(C_j,C_{j'}) > \varepsilon\right]$$

$$\leq P\left[\varepsilon > \min_{(j,j')\in\mathcal{D}_\delta} p(C_j,C_{j'})\right] + P\left[\min_{(j,j')\in\mathcal{S}_0} p(C_j,C_{j'}) > \varepsilon\right]$$

$$\leq \sum_{(j,j')\in\mathcal{D}_\delta} P\left[\varepsilon > p(C_j,C_{j'})\right] + 1 - P\left[\bigcup_{(j,j')\in\mathcal{S}_0} p(C_j,C_{j'}) \leq \varepsilon\right]$$

$$\leq |\mathcal{D}_\delta|P[\varepsilon > p(A,B)] + 1 - P[\ p(C,D) \leq \varepsilon]$$

$$\leq \frac{\ell(\ell-1)}{2}P[p(A,B) < \varepsilon] + P[\ p(C,D) > \varepsilon]$$

where (A, B) (respectively (C, D)) represents any pair of subsets being δ-separated (respectively having same intensity) in the partition $C_1,\ldots,C_\ell$.

Before to go further, we recall the following concentration inequalities for shifted-χ^2:

Proposition 2 (cf Birgé). *Let D be a non-central χ^2 variable with μ degrees of freedom and non-centrality parameter $\lambda \geq 0$, then for all $x > 0$*

$$\mathbf{P}[D \geqslant \mu + \lambda^2 + 2\sqrt{(\mu+2\lambda^2)x} + 2x] \leqslant e^{-x}, \tag{D.1}$$

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

As a consequence the ε-quantile of a $\chi^2(2^{\bar{K}}, n\delta_K^2)$ satisfies

$$\chi^{-2}_{2^{\bar{K}},n\delta_K^2}(\varepsilon) \geq 2^{\bar{K}} + (n\delta_K^2)^2 - 2\sqrt{(2^{\bar{K}}+2(n\delta_K^2)^2)\log(1/\varepsilon)}.$$

Let us first focus on the term $P[p(C,D) > \varepsilon]$ for $(C,D) = (C_j,C_{j'})$ with $(j,j') \in \mathcal{S}_0$. By construction and thanks to the independence of the $\bar{D}_K^{C,D}$

$$P[p(C,D) > \varepsilon] = 1 - P\left[\max_K P(\chi^2(2^{\bar{K}}, n\delta_K^2) \leq \|\bar{D}_K^{C,D}\|^2) \leq \varepsilon\right]$$

$$= 1 - \prod_K P\left[\|\bar{D}_K^{C,D}\|^2 \leq \chi^{-2}_{2^{\bar{K}},n\delta_K^2}(\varepsilon)\right]$$

$$= 1 - \prod_K P\left[\chi^2(2^{\bar{K}}) \leq \chi^{-2}_{2^{\bar{K}},n\delta_K^2}(\varepsilon)\right]$$

Setting $(n\delta_K^2)^2 = a_K^4 2^{\bar{K}}$ for some a_K to be made precise later, we get
$P[p(C,D) > \varepsilon]$

$$\leq 1 - \prod_K\left(1 - P\left[\chi^2(2^{\bar{K}}) > 2^{\bar{K}} + a_K^4 2^{\bar{K}} - 2\sqrt{(2^{\bar{K}}+2a_K^4 2^{\bar{K}})\log\frac{1}{\varepsilon}}\right]\right).$$

For $K = 0,\ldots,K_0$ let us define $y_K > 0$ as the solution of

$$2^{\bar{K}} + 2\sqrt{2^{\bar{K}}y_K} + 2y_K = 2^{\bar{K}} + a_K^4 2^{\bar{K}} - 2\sqrt{(2^{\bar{K}}+2a_K^4 2^{\bar{K}})\log\frac{1}{\varepsilon}},$$

i.e.

$$y_K^2 = \frac{1}{2}\left(\sqrt{2a_K^4 2^{\bar{K}} + 2^{\bar{K}} - 4\sqrt{(2a_K^4 2^{\bar{K}} + 2^{\bar{K}})\log\frac{1}{\varepsilon}} - 2^{\bar{K}/2}}\right)$$

$$= 2^{\bar{K}/2-1}\left(B_K\sqrt{1 - \frac{4}{2^{\bar{K}/2}B_K}\sqrt{\log\frac{1}{\varepsilon}}} - 1\right) \tag{D.3}$$

with $B_K^2 = 4(2a_K^4 + 1)$. Then, using the properties of the logarithm, we get

$$P[p(C,D) > \varepsilon] \leq 1 - \prod_K\left(1 - e^{-y_K}\right) = 1 - \exp\left[\sum_K \log\left(1 - e^{-y_K}\right)\right]$$

$$\leq 1 - \exp\left[\log\left(1 - \sum_K e^{-y_K}\right)\right] = \sum_K e^{-y_K}$$

$$\leq \gamma\frac{e}{e-1} \leq 1.6\gamma, \tag{D.4}$$

where the first inequality of the last line holds if $y_K \geq K - \log\gamma$. Let us now assume that

$$2^{\bar{K}}B_K^2 \geq (2^8/9)\log(1/\varepsilon) \qquad \text{and} \qquad a_K^2 \geq 2. \tag{D.5}$$

It follows from these conditions that

$$1 - \frac{4}{2^{\bar{K}/2}B_K}\sqrt{\log\frac{1}{\varepsilon}} \geq 1/4 \qquad \text{and} \qquad (B_K/2) - 1 = \sqrt{2a_K^4 + 1} - 1 \geq a_K^2,$$

so that (D.3) leads to $y_K^2 \geq 2^{\bar{K}/2-1}a_K^2$. Putting all our conditions together, we see that $P[p(C,D) > \varepsilon] < 1.6\gamma$ if

$$(n\delta_K^2)^2 = a_K^4 2^{\bar{K}} \geq \max\left\{2^{\bar{K}+2}; 4(K - \log\gamma)^2; (2^5/9)\log(1/\varepsilon) - 2^{\bar{K}-1}\right\}. \tag{D.6}$$

Now, if A and B are two δ-separated subsets then

$$P[p(A,B) < \varepsilon] = P\left[\max(U_0,\ldots,U_{K_0}) < \varepsilon\right]$$

where $U_0,\ldots,U_{K_0}$ are $K_0 + 1$ independent random variables more concentrated on the right of the interval $[0,1]$ than uniforms. We recall that $n/2 < 2^{K_0+1} \leq n$. It follows that

$$P[p(A,B) < \varepsilon] = \prod_{k=0}^{K_0} P[U_k < \varepsilon] = \varepsilon^{K_0+1} = 2^{-(K_0+1)\log_2(1/\varepsilon)} < (n/2)^{\log_2\varepsilon}.$$

Hence the overall probability to have two δ-separated subsets merging before any two pair of subsets having same intensity is less than

$$1.6\gamma + (\ell^2/2)(n/2)^{\log_2\varepsilon}.$$

At this stage one has the choice of γ and ε but it seems reasonable to choose the two terms of the previous bound of the same order, i.e. $1.6\gamma = (\ell^2/2)(n/2)^{\log_2\varepsilon}$ which leads to the bound $\ell^2(n/2)^{\log_2\varepsilon}$. Finally setting $\varepsilon = 2^{-\kappa}$ for κ large enough, we get the bound $\ell^2(n/2)^{-\kappa}$ which can be made small by a convenient choice of κ. Since $\ell \geq 2$, (D.6) holds if

$$n\delta_K^2 \geq \max\left\{2^{1+\bar{K}/2}; 2(K + \kappa\log\frac{n}{2}); 1.57\sqrt{\kappa}\right\}.$$

As $2\log(n/2) \geq 1.57$ for $n \geq 5$ and since $\bar{K} \leq K_0 - 1 \leq \log_2 n - 2$, it is enough in this case to have

$$n\delta_K^2 \geq \max\left\{\sqrt{n}, 2(1 + \kappa\log 2)\log_2\frac{n}{2}\right\}$$

to ensure the overall probability to have two δ-separated subsets merging before any two pair of subsets having same intensity is less than $\ell^2(n/2)^{-\kappa}$ when $\kappa \geq 1$.

Proof of Theorem 2. Along the iterations of the hierarchical clustering, it is enough to bound the probability to wrongly binding at each step using the previous lemma. That is at most $|\mathcal{X}|$ times. $\square$

Appendix E. 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 $X, Y \in \mathcal{P}$.

Iterations

while $\bar{p}(\bar{\ell}) < c_\alpha(\ell)$ **do**
   Find (s_1, s_2) satisfying (11);
   New cluster: $C \leftarrow C_{s_1} \cup C_{s_2}$;
   Update partition: $\mathcal{P}^{\bar{\ell}+1} \leftarrow \mathcal{P}^{\bar{\ell}} \setminus \{C_{s_1}, C_{s_2}\} \cup \{C\}$;
   New neighbor: $\mathcal{V}(C) := \mathcal{V}(C_{s_1}) \cup \mathcal{V}(C_{s_2}) \setminus \{C_{s_1}, C_{s_2}\}$;
   Compute new dissimilarities as in (15) for $C' \in \mathcal{V}(C)$;
   Update $\mathcal{N} \leftarrow \mathcal{N} \setminus \{\mathcal{V}(C_{s_1}), \mathcal{V}(C_{s_2})\} \cup \mathcal{V}(C)$;
   $\bar{\ell} \leftarrow \bar{\ell} + 1$;
   Update $\bar{p}(\bar{\ell})$ as in (16).
end while
Number of clusters: $\ell^* \leftarrow N - \bar{\ell}$.

Appendix F. HiSET v.s. Classic hierarchical clustering

In order to demonstrate the strength of HiSET over the classical hierarchical clustering based on the Euclidian distance, we compare both methods on a simulated image sequence consisting of three clusters (see Fig. F.16), called latter Chessboard. Standard Gaussian noise has been added to the true enhancement curves to provide the observations. Each image in the sequence is of size 55×55 and the sequence contains 100 images.

Fig. F.17 illustrates the segmentation results. Given the true number of clusters, the hierarchical clustering based on the Euclidian distance binds red and green clusters and misclassifies several voxels. Indeed the separation distance between the true curves is not large enough with respect to the noise level, providing an illustration of the benefit of using a multiple test. Only, one voxel (in white) is identified as a third cluster. Without any knowledge, HiSET is able to recover all clusters up to a mistake on one voxel.

Appendix G. Notations

|	
$\mathcal{X}$	the voxel grid corresponding to cross section of the image
x, y	voxels of the grid
X, Y	subsets of voxels
$|X|, |Y|$	cardinalities
$t_1, \ldots, t_n$	the acquisition times
$\Phi^x(t_j)$	the observed intensity at time t_j and voxel x
$\phi^x(t_j)$	the expectation of the intensity at time t_j and voxel x
$I^x(t_j)$	the transformed observed intensity at time t_j and voxel x
$i^x(t_j)$	the expectation of the transformed intensity at time t_j and voxel x
η_j^x	an additive standard Gaussian noise
$\bar{I}^X$	empirical mean of the I^x over X
$\bar{i}^X$	expectation of the empirical mean of i^x over X
$\rho(X, Y)$	scaling factor $|X|^{-1} + |Y|^{-1}$
D^{XY}	normalized difference of $\bar{I}^X$ and $\bar{I}^Y$
d^{XY}	expectation of normalized difference of $\bar{i}^X$ and $\bar{i}^Y$
K	a time index partition reference
K_0	the reference of the finest time index partition
$\Pi_K D^{XY}$	projection over the K-th partition
D_K^{XY}	$\Pi_K D^{XY} - \Pi_{K-1} D^{XY}$
Σ_K	covariance matrix of D_K^{XY}
$\bar{D}_K$	scaled version of D_K^{XY}
$\bar{d}_K$	expectation of $\bar{D}_K$
$\|u\|_n$	Euclidian norm of u in $\mathbb{R}^n$
$C_1, \ldots, C_\ell$	some clusters
$\mathcal{N}(\mu, \sigma^2)$	Gaussian distribution with mean μ and variance σ^2
$\mathcal{N}(d, \Sigma)$	multivariate Gaussian distribution with mean d and covariance matrix Σ
$\chi^2(d, \lambda)$	shifted chi-squared distribution
$\mathcal{H}_0^K$	the null hypothesis of the equivalence test over the K-th partition
$\mathcal{H}_1^K$	the alternative of the equivalence test over the K-th partition
$p_K(X, Y)$	the p-value over the K-th partition when testing $\bar{i}^X(.) = \bar{i}^Y(.)$
$p(X, Y)$	the p-value of the multiple equivalence test of $\bar{i}^X(.) = \bar{i}^Y(.)$
$\bar{\ell}$	$|\mathcal{X}| - \ell$
$p(\bar{\ell})$	the minimum dissimilarity at iteration $\bar{\ell}$
$c_\alpha(\ell)$	the control function when the clustering is made of ℓ clusters (iteration $\bar{\ell}$)
$\mathcal{P}^{loc}$	the final partition after the local step
ℓ^{loc}	the size of $\mathcal{P}^{loc}$
$\mathcal{P}^*$	the final partition after the global step
ℓ^*	the size of $\mathcal{P}^*$
ℓ_0	the size of the true underlying partition
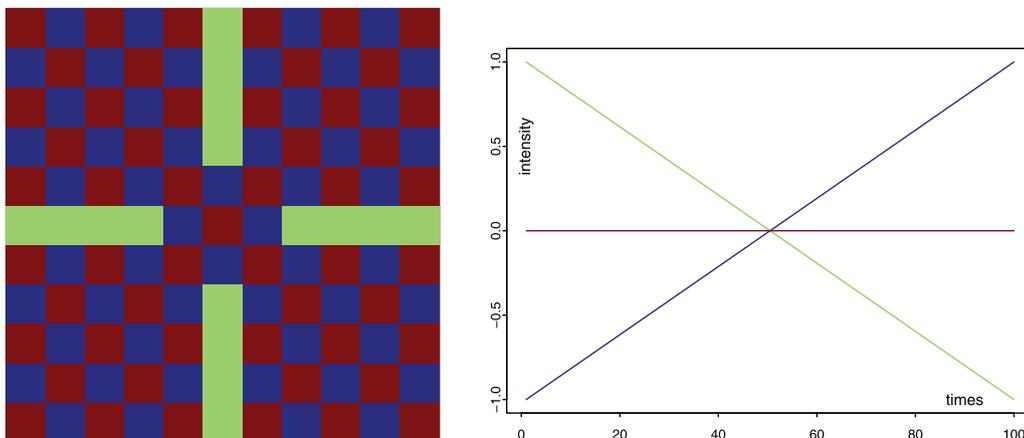
Fig. F1. Chessboard image sequence: (left) The ground-truth segmentation of $\mathcal{X}$; (right) The true enhancement curves, $i^x(t)$, associated to the 3 clusters using corresponding colors.
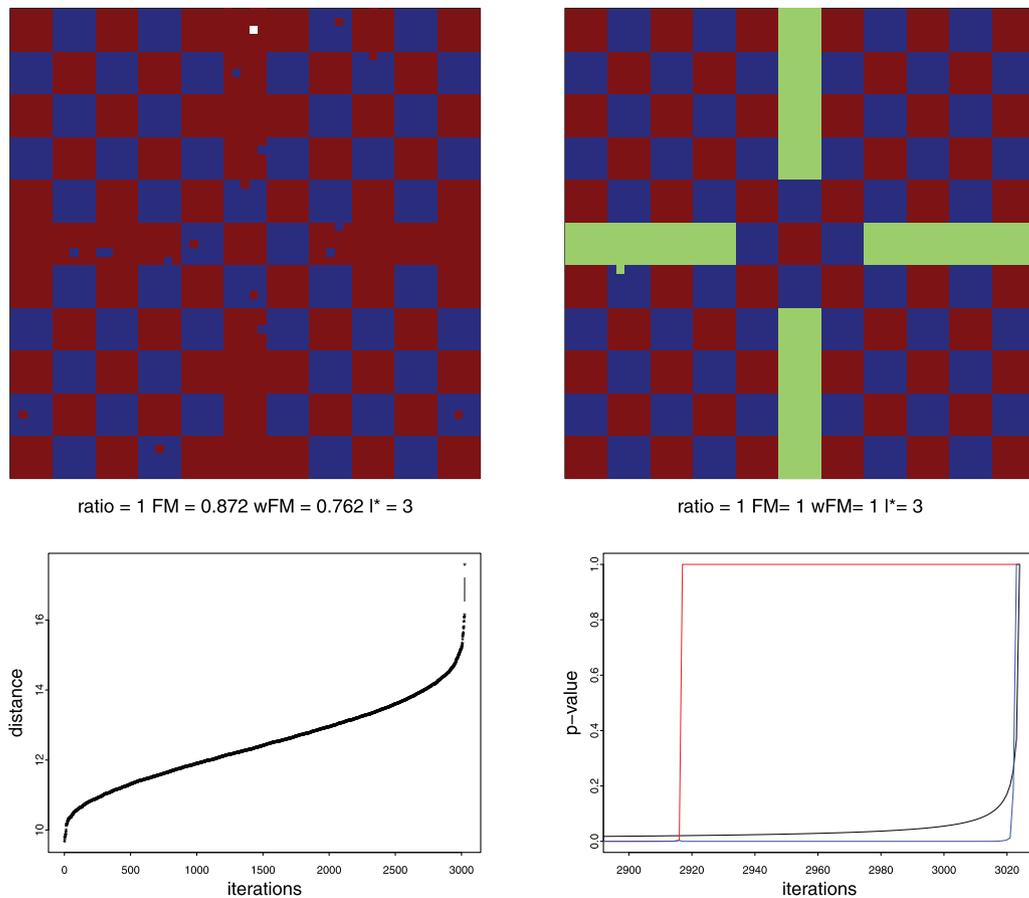
Fig. F2. Segmentation results for the chessboard image sequence: hierarchical clustering based on the Euclidian distance (left top) and its linkage function (left bottom); HiSET (right top) and the corresponding dissimilarity function together with the control function (right bottom).

References

Achanta, R., Shaji, A., Smith, K., Lucchi, A., Fua, P., Süsstrunk, S., 2010. SLIC Super-pixels. Technical Report. EPFL.

Ackerman, M., Ben-David, S., 2011. Discerning linkage-based algorithms among hierarchical clustering methods. In: Proceedings of the Twenty-Second international joint conference on Artificial Intelligence, IJCAI'11. AAAI Press, pp. 1140–1145.

Agner, S.C., Xu, J., Madabhushi, A., 2013. Spectral embedding based active contour (SEAC) for lesion segmentation on breast dynamic contrast enhanced magnetic resonance imaging. Med. Phys. 40 (3).

Axel, L., 1980. Cerebral blood flow determination by rapid-sequence computed tomography: theoretical analysis. Radiology 137 (3), 679–686.

Bakir, M., Wei, J., Nelson, M.D., Mehta, P.K., Haftbaradaran, A., Jones, E., Gill, E., Sharif, B., Slomka, P.J., Li, D., Shufelt, C.L., Minissian, M., Berman, D.S., Bairey Merz, C.N., Thomson, L.E., 2016. Cardiac magnetic resonance imaging for myocardial perfusion and diastolic function-reference control values for women. Cardiovasc. Diagn. Ther. 6 (1), 78–86. doi:10.3978/j.issn.2223-3652.2015.09.03.

Baraud, Y., Huet, S., Laurent, B., 2003. Adaptive tests of linear hypotheses by model selection. Ann. Stat. 31 (1), 225–251.

Baraud, Y., Huet, S., Laurent, B., 2005. Testing convex hypotheses on the mean of a Gaussian vector. application to testing qualitative hypotheses on a regression function. Ann. Stat. 33 (1), 214–257.

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

Baumgartner, C., Gautsch, K., Bohm, C., Felber, S., 2005. Functional cluster analysis of ct perfusion maps: a new tool for diagnosis of acute stroke? J. Digit Imaging 18 (3), 219–226. doi:10.1007/s10278-004-1048-9.

Bergamino, M., Bonzano, L., Levrero, F., Mancardi, G.L., Roccatagliata, L., 2014. A review of technical aspects of t1-weighted dynamic contrast-enhanced magnetic resonance imaging (DCE-MRI) in human brain tumors. Phys. Med. 30 (6), 635–643. doi:10.1016/j.ejmp.2014.04.005.

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

Bhooshan, N., Giger, M.L., Jansen, S.A., Li, H., Lan, L., Newstead, G.M., 2010. Cancerous breast lesions on dynamic contrast-enhanced mr images: computerized characterization for image-based prognostic markers. Radiology 254 (3), 680–690. doi:10.1148/radiol.09090838.

Birgé, L., Massart, P., 2007. Minimal penalties for Gaussian model selection. Probab. Theory Relat. Fields 138 (1), 33–73.

Brix, G., Kiessling, F., Lucht, R., Darai, S., Wasser, K., Delorme, S., Griebel, J., 2004. Microcirculation and microvasculature in breast tumors. Magn. Reson. Med. 52, 420–429.

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

Celeux, G., Forbes, F., Peyrard, N., 2003. Em procedures using mean field-like approximations for Markov model-based image segmentation. Pattern Recognit. 36, 131–144.

Chatzis, S.P., Varvarigou, T.A., 2008. A fuzzy clustering approach toward hidden Markov random field models for enhanced spatially constrained image segmentation. IEEE Trans. Fuzzy Syst. 16 (5), 1351–1361.

Chen, B.B., Shih, T.T., 2014. DCE-MRI in hepatocellular carcinoma-clinical and therapeutic image biomarker. World J. Gastroenterol. 20 (12), 3125–3134. doi:10.3748/wjg.v20.i12.3125.

Chen, W., Giger, M.L., Bick, U., Newstead, G.M., 2006. Automatic identification and classification of characteristic kinetic curves of breast lesions on DCE-MRI. Med. Phys. 33 (8), 2878–2887.

Chen, W., Giger, M.L., Newstead, G.M., Bick, U., Jansen, S.A., Li, H., Lan, L., 2010. Computerized assessment of breast lesion malignancy using DCE-MRI robustness study on two independent clinical datasets from two manufacturers. Acad. Radiol. 17 (7), 822–829. doi:10.1016/j.acra.2010.03.007.

Chung, S., Shah, B., Storey, P., Iqbal, S., Slater, J., Axel, L., 2016. Quantitative perfusion analysis of first-pass contrast enhancement kinetics: application to MRI of myocardial perfusion in coronary artery disease. PLoS ONE 11 (9).

Comaniciu, D., Meer, P., 2002. Mean shift: a robust approach toward feature space analysis. IEEE Trans. Pattern Anal. Mach. Intell. 24 (5), 603–619.

Comte, F., Cuenod, C.-A., Pensky, M., Rozenholc, Y., 2017. Laplace deconvolution on the basis of time domain data and its application to dynamic contrast-enhanced imaging. J. Royal Stat. Soc.: Ser. B (Stat. Methodol.) 79 (1), 69–94. doi:10.1111/rssb.12159.

Ferre, R., Aldis, A., AlSharif, S., Omeroglu, A., Mesurolle, B., 2016. Differentiation of fibroadenomas and pure mucinous carcinomas on dynamic contrast-enhanced MRI of the breast using volume segmentation for kinetic analysis: a feasibility study. AJR Am. J. Roentgenol. 206 (2), 253–258. doi:10.2214/AJR.15.14709.

Fieselmann, A., Kowarschik, M., Ganguly, A., Hornegger, J., Fahrig, R., 2011. Deconvolution-based CT and MR brain perfusion measurement: theoretical model revisited and practical implementation details. Int. J Biomed. Imaging 2011, 20.

Fowlkes, E.B., Mallows, C.L., 1983. A method for comparing two hierarchical clusterings. J. Am. Stat. Assoc. 78 (383), 553–569.

Frias, A.E., Schabel, M.C., Roberts, V.H., Tudorica, A., Grigsby, P.L., Oh, K.Y., Kroenke, C.D., 2015. Using dynamic contrast-enhanced MRI to quantitatively characterize maternal vascular organization in the primate placenta. Magn. Reson. Med. 73 (4), 1570–1578. doi:10.1002/mrm.25264.

Giannini, V., Vignati, A., Morra, L., Persano, D., Brizzi, D., Carbonaro, L., Bert, A., Sardanelli, F., Regge, D., 2010. A fully automatic algorithm for segmentation of the breasts in DCE-MR images. Conf. Proc. IEEE Eng. Med. Biol. Soc. 2010, 3146–3149. doi:10.1109/IEMBS.2010.5627191.

Glocker, B., Sotiras, A., Komodakis, N., Paragios, N., 2011. Deformable medical image registration: setting the state of the art with discrete methods. Annu. Rev. Biomed. Eng. 13, 219–244.

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

Hanson, E., Eikefjord, E., Rorvik, J., Andersen, E., Lundervold, A., Hodneland, E., 2017. Workflow sensitivity of post-processing methods in renal DCE-MRI. Magn. Reson. Imaging 42, 60–68. doi:10.1016/j.mri.2017.05.003.

Hedjam, R., Mignotte, M., 2009. A hierarchical graph-based markovian clustering approach for the unsupervised segmentation of textured color images. In: Proceedings of the 2009 16th IEEE International Conference on Image Processing (ICIP). IEEE, pp. 1365–1368.

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., 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. doi:10.1148/radiol.12120255.

Hou, Z., Wang, Y., Thng, C.H., Ng, Q.S., Goh, V., San Koh, T., 2014. Automatic region-of-interest segmentation and registration of dynamic contrast-enhanced images of colorectal tumors. Phys. Med. Biol. 59 (23), 7361–7381. doi:10.1088/0031-9155/59/23/7361.

Ingrisch, M., Sourbron, S., 2013. Tracer-kinetic modeling of dynamic contrast-enhanced MRI and CT: a primer. J. Pharmacokinet. Pharmacodyn. 40 (3), 281–300.

Irving, B., Cifor, A., Papiez, B.W., Franklin, J., Anderson, E.M., Brady, S.M., Schnabel, J.A., 2014. Automated colorectal tumour segmentation in DCE-MRI using supervoxel neighbourhood contrast characteristics. Med. Image Comput. Comput. Assist. Interv. 17 (Pt 1), 609–616.

Irving, B., Franklin, J., Papież, B., Anderson, E., Sharma, R., Gleeson, F., Brady, S.M., Schnabel, J., 2016a. Pieces-of-parts for supervoxel segmentation with global context: application to DCE-MRI tumour delineation. Med. Image Anal. 32, 69–83.

Irving, B., Franklin, J.M., Papiez, B.W., Anderson, E.M., Sharma, R.A., Gleeson, F.V., Brady, S.M., Schnabel, J.A., 2016b. Pieces-of-parts for supervoxel segmentation with global context: application to DCE-MRI tumour delineation. Med. Image Anal. 32, 69–83. doi:10.1016/j.media.2016.03.002.

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

Khalifa, F., Abou El-Ghar, M., Abdollahi, B., Frieboes, H.B., El-Diasty, T., El-Baz, A., 2013. A comprehensive non-invasive framework for automated evaluation of acute renal transplant rejection using DCE-MRI. NMR Biomed 26 (11), 1460–1470. doi:10.1002/nbm.2977.

Lecoeur, J., Ferre, J.-C., Collins, D.L., Morrissey, S.P., Barillot, C., 2009. Multi channel MRI segmentation with graph cuts using spectral gradient and multidimensional gaussian mixture model. In: Pluim, J.P.W., Dawant, B.M. (Eds.), Proceedings of the SPIE-Medical Imaging. Vol. 7259. Orlando, United States. 72593X1-72593X11

Li, S., Zöllner, F.G., Merrem, A.D., Peng, Y., Roervik, J., Lundervold, A., Schad, L.R., 2012. Wavelet-based segmentation of renal compartments in DCE-MRI of human kidney: initial results in patients and healthy volunteers. Comput. Med. Imaging Graph. 36, 108–118.

McClymont, D., Mehnert, A., Trakic, A., Kennedy, D., Crozier, S., 2014. Fully automatic lesion segmentation in breast MRI using mean-shift and graph-cuts on a region adjacency graph. J. Magn. Reson. Imaging 39, 795–804.

Medved, M., Medved, M., Karczmar, G., Yang, C., Dignam, J., Gajewski, T.F., Kindler, H., Vokes, E., MacEneany, P., Mitchell, M.T., Stadler, W.M., 2004. Semi-quantitative analysis of dynamic contrast enhanced MRI in cancer patients: variability and changes in tumor tissue over time. Magn. Reson. Imaging 20, 122–128.

Michoux, N., Simoni, P., Tombal, B., Peeters, F., Machiels, J.P., Lecouvet, F., 2012. Evaluation of DCE-MRI postprocessing techniques to assess metastatic bone marrow in patients with prostate cancer. Clin. Imaging 36 (4), 308–315. doi:10.1016/j.clinimag.2011.10.002.

Nee, G., Jehan-Besson, S., Brun, L., Revenu, M., Hamon, M., Hamon-Kerautret, M., Perrin, M., 2009. A MRI myocardial perfusion analysis tool. Conf. Proc. IEEE Eng. Med. Biol. Soc. 2009, 4387–4390. doi:10.1109/IEMBS.2009.5333479.

Ostergaard, L., Sorensen, A.G., Kwong, K.K., Weisskoff, R.M., Gyldensted, C., Rosen, B.R., 1996. High resolution measurement of cerebral blood flow using intravascular tracer bolus passages. part ii: experimental comparison and preliminary results. Magn. Reson. Med. 36 (5), 726–736.

Padhani, A.R., Khan, A.A., 2010. Diffusion-weighted (DW) and dynamic contrast-enhanced (DCE) magnetic resonance imaging (MRI) for monitoring anticancer therapy. Target Oncol. 5 (1), 39–52. doi:10.1007/s11523-010-0135-8.

Raj, A., Juluru, K., 2009. Visualization and segmentation of liver tumors using dynamic contrast MRI. Conf. Proc. IEEE Eng. Med. Biol. Soc. 2009, 6985–6989. doi:10.1109/IEMBS.2009.5333859.

Sanz-Requena, R., Marti-Bonmati, L., Perez-Martinez, R., Garcia-Marti, G., 2016. Dynamic contrast-enhanced case-control analysis in 3T MRI of prostate cancer can help to characterize tumor aggressiveness. Eur. J. Radiol. 85 (11), 2119–2126. doi:10.1016/j.ejrad.2016.09.022.

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

Selvan, A.N., Cole, L.M., Spackman, L., Naylor, S., Wright, C., 2017. Hierarchical cluster analysis to aid diagnostic image data visualization of ms and other medical imaging modalities. Methods Mol. Biol. 1618, 95–123. doi:10.1007/978-1-4939-7051-3_10.

Shi, J., Malik, J., 2000. Normalized cuts and image segmentation. IEEE Trans. Pattern Anal. Mach. Intell. 22 (8), 888–905.

Shou, H., Shinohara, R.T., Liu, H., Reich, D.S., Crainiceanu, C.M., 2016. Soft null hypotheses: a case study of image enhancement detection in brain lesions. J. Comput. Graph. Stat. 25 (2), 570–588.

Sotiras, A., Komodakis, N., Glocker, B., Deux, J.-F., Paragios, N., 2009. Graphical models and deformable diffeomorphic population registration using global and local metrics. In: Proceedings of MICCAI, pp. 672–679.

Tao, W., Jin, H., Zhang, Y., 2007. Color image segmentation based on mean shift and normalized cuts. IEEE Trans. Syst. Man Cybern 37 (5), 1382–1389.

Tartare, G., Hamad, D., Azahaf, M., Puech, P., Betrouni, N., 2014. Spectral clustering applied for dynamic contrast-enhanced mr analysis of time–intensity curves. Comput. Med. Imaging Graph. 38 (8), 702–713.

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

Tofts, P.S., 1997. Modeling tracer kinetics in dynamic GD-DTPA MR imaging. J. Magn. Reson. Imaging 7, 91–101.

Tofts, P.S., Berkowitz, B., Schnall, M.D., 1995. Quantitative analysis of dynamic GD-DTPA enhancement in breast tumors using a permeability model. Magn. Reson. Med. 33, 564–568.

Tu, Z., Zhu, S.-C., 2002. Image segmentation by data-driven Markov chain Monte Carlo. IEEE Trans. Pattern Anal Mach Intell 24 (5), 657–673.

Tuncbilek, N., Karakas, H.M., Altaner, S., 2004. Dynamic MRI in indirect estimation of microvessel density, histologic grade, and prognosis in colorectal adenocarcinomas. Abdom. Imaging 29, 166–172.

Tuncbilek, N., Karakas, H.M., Okten, O.O., 2005. Dynamic contrast enhanced MRI in the differential diagnosis of soft tissue tumors. Eur. J. Radiol. 53, 500–505.

Vedaldi, A., Soatto, S., 2008. Quick shift and kernel methods for mode seeking. In: Proceedings of the European Conference on Computer Vision (ECCV), pp. 705–718.

Wang, C.H., Yin, F.F., Horton, J., Chang, Z., 2014. Review of treatment assessment using DCE-MRI in breast cancer radiation therapy. World J. Methodol. 4 (2), 46–58. doi:10.5662/wjm.v4.i2.46.

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

Winfield, J.M., Payne, G.S., Weller, A., deSouza, N.M., 2016. DCE-MRI, DW-MRI, and mrs in cancer: challenges and advantages of implementing qualitative and quantitative multi-parametric imaging in the clinic. Top Magn. Reson. Imaging 25 (5), 245–254. doi:10.1097/RMR.0000000000000103.

Woodard, T., Sigurdsson, S., Gotal, J.D., Torjesen, A.A., Inker, L.A., Aspelund, T., Eiriksdottir, G., Gudnason, V., Harris, T.B., Launer, L.J., Levey, A.S., Mitchell, G.F., 2015. Segmental kidney volumes measured by dynamic contrast-enhanced magnetic resonance imaging and their association with CKD in older people. Am. J. Kidney Dis. 65 (1), 41–48. doi:10.1053/j.ajkd.2014.05.017.

Wu, Q., Salganicoff, M., Krishnan, A., Fussell, D.S., Markey, M.K., 2006. Interactive lesion segmentation on dynamic contrast enhanced breast MRI using a markov model. Med. Imaging 2006: Image Process. 6144 (1), 1487–1494.

Zelnik-Manor, L., Perona, P., 2004. Self-tuning spectral clustering. In: Proceedings of the Advances in Neural Information Processing Systems 17. MIT Press, pp. 1601–1608.

Zheng, Y., Englander, S., Baloch, S., Zacharaki, E.I., Fan, Y., Schnall, M.D., Shen, D., 2009. Step: spatiotemporal enhancement pattern for mr-based breast tumor diagnosis. Med. Phys. 36 (7), 3192–3204.