Functional segmentation of renal DCE-MRI sequences using vector quantization algorithms

Béatrice Chevaillier · Damien Mandry · Jean-Luc Collette · Michel Claudon · Marie-Agnès Galloy · Olivier Pietquin

Received: date / Accepted: date

Abstract In dynamic contrast-enhanced magnetic resonance imaging, segmentation of internal kidney structures like cortex, medulla and cavities is essential for functional assessment. To avoid fastidious and time-consuming manual segmentation, semi-automatic methods have been recently developed. Some of them use the differences between temporal contrast evolution in each anatomical region to perform functional segmentation. We test two methods where pixels are classified according to their time-intensity evolution. They both require a vector quantization stage with some unsupervised learning algorithm (K-means or Growing Neural Gas with targeting). Three or more classes are thus obtained. In the first case the method is completely automatic. In the second case, a restricted intervention by an observer is required for merging. As no ground truth is available for result evaluation, a manual anatomical segmentation is considered as a reference. Some discrepancy criteria like overlap, extra pixels and similarity index are computed between this segmentation and a functional one. The same criteria are also evaluated between the reference and another manual segmentation. Results are comparable for the two types of comparisons, proving that anatomical segmentation can be performed using functional information.

Keywords Image segmentation · Vector quantization · Biomedical image processing · Biomedical magnetic resonance imaging · Image sequence analysis · Clustering methods.

1 Introduction

Perfusion dynamic contrast-enhanced magnetic resonance imaging (DCE-MRI) with injection of an exogenous contrast agent such as gadolinium chelates is a widely used MRI technique providing functional information for renal assessment [1, 2]. Several functional parameters like the glomerular filtration rate or the differential renal function [3] can thus be evaluated non-invasively from time-intensity curves of different anatomical compartments.

[1] IMS Research Group, Supélec Metz Campus, Metz, France
[2] U947, Inserm, Nancy, France
[3] IADI, Nancy-Université, Nancy, France
Tel.: (33)-387-764-770
Fax: (33)-387-764-700
E-mail: Olivier.Pietquin@supelec.fr
The segmentation of kidney in regions of interest (ROI) defining the studied compartments is then essential for functional evaluation and detection of diseases affecting different parts of this organ. Manual segmentation of cortex, medulla and pelvo-calceal cavities is generally performed by a radiologist. This operation can be quite time-consuming and fastidious, especially if the number of slices is high. Furthermore, for currently used T1 weighted sequences, it cannot be done on a single image because all the compartments are not enhanced during the same perfusion phase for physiological reasons: contrast for cavities is maximum during late perfusion phase whereas cortex and medulla can only be distinguished just after contrast agent enters the kidney (see figure 1). So radiologists have first to review the whole sequence to select the most relevant images. If the chosen frames are not well registered or in case of through-plane motion, functional results can vary greatly. Compartment delineation is anyway delicate because images are highly noisy and blurred, and remains operator-dependant.

Few popular automated or semi-automated segmentation methods like intensity thresholding or region-growing algorithms, that are usually used in medical field, have been applied and evaluated for DCE-MRI of kidney [4]. Anyway problems due to the use of only a few frames remain. As the functional profile varies between the anatomical compartments, a classification of pixels according to their temporal evolution can be performed: such a technique could improve the robustness and reproducibility of the segmentation because the whole sequence is utilized. In [5], a level set approach is used: segmentation of cortex, medulla and pelvis is performed in two steps by minimizing an energy functional taking into account both spatial correlation among pixels in the same image and temporal correlation across the whole sequence. In [6] a graph cut algorithm using a temporal Markov model for time-intensity curves allows separation of cortex, medulla and collecting system; operator intervention during the process is necessary for seeds positioning. In [7] a multi-step analysis including successive registrations and segmentations is proposed: pixels are classified using a K-means partitioning algorithm applied to their time-intensity curves.

Regarding validation, no ground truth is available for real data. In [8] a segmentation error with manual reference segmentation is computed by adding up the volumes of false-positive (oversegmented) and false-negative (undersegmented) voxels. Nevertheless the validation consists mostly in qualitative coherence with segmentation by radiologists or in comparisons between the corresponding compartment volumes or the renograms [9].

In this article we propose to improve the method for functional segmentation of cortex, medulla and cavities that has been described in [10] and to compare it with the technique developed in [11]. Renal pixels are first classified in several clusters according to their temporal contrast evolution. As we do not have access to enough manual anatomical segmentations to build up a training data set, supervised classification cannot be considered in our case. Two
vector quantization algorithms, namely K-means [12] and a variant of the Growing Neural Gas [13], are used here to reduce information and to get K typical curves, or prototypes, that represent the whole set of time-intensity curves. Depending on the results of vector quantization step, a simple intervention by an operator may be required to merge some clusters. The first validation step for the induced segmentation consists in a visual inspection in order to check the segmentation consistency. Regarding quantitative assessment, as no ground truth is available for result evaluation, a manual morphological segmentation by a radiologist is then chosen as the reference. Some similarity criteria like overlap, extra pixels and similarity index are evaluated between this segmentation and functional ones. As a comparison, the same criteria are then computed between the reference and a manual segmentation by another radiologist. Resulting segmentations are also compared with those obtained by simple intensity thresholding.

2 Theory: vector quantization of time-intensity curves of kidney pixels

Time-intensity curves of the renal pixels are obtained from a registered DCE-MRI sequence with Nr frames (registration method is detailed in 3.1). Three frames from such a sequence, corresponding to different perfusion phases (arterial peak, filtration and late phase), can be seen in figure 1. These curves, similar to those in figure 4, may be normalized before vector quantization (see section 3.3).

Let \( \{ x_i \}_{1 \leq i \leq N} \) be the \( N \) kidney pixels and \( \xi_i = (\xi_{i1}, \ldots, \xi_{iN_T}) \) the \( N_T \) components vector associated with pixel \( x_i \), where \( \xi_{ip} \) is its contrast at time \( p \). The \( N \) vectors \( \xi_i \) can be considered as samples of an underlying probability distribution over a manifold \( X \subseteq \mathbb{R}^{N_T} \) with a density of probability \( p(\xi) \). The purpose of vector quantization techniques is classically to find, using the \( \xi_i \), a discrete set \( w = \{ w_j \}_{1 \leq j \leq K} \subseteq X \) of prototypes (or nodes) that maps the distribution as "accurately" as possible according to some criterion. Let be \( w(\xi) = \arg\min_{w_i} \{ \| \xi - w_i \|^2 \} \). The position of the prototypes is optimal when it minimizes the cost function, or distorsion,

\[
E(w) = \int_X \| w(\xi) - \xi \|^2 p(\xi) d\xi = \sum_{j=1}^K E_j
\]

where

\[
E_j = \int_{V_j} \| w_j - \xi \|^2 p(\xi) d\xi \quad \text{and} \quad V_j = \{ \xi \in X : w(\xi) = w_j \}
\]

\( V_j \) is the so-called Voronoi cell around \( w_j \): it consists of all points of \( X \) that are closer to \( w_i \) than to any other \( w_i \). The set of \( \{ V_j \}_{1 \leq j \leq K} \) is a partition of \( X \). Once \( w \) has been determined, a first clustering \( C = \{ C_1, \ldots, C_K \} \) of the \( \{ \xi_i \}_{1 \leq i \leq N} \) is then built up so that \( C_j = \{ \xi_i : \xi_i \in V_j \} \). This clustering induces a partition of the corresponding pixels \( \{ x_i \}_{1 \leq i \leq N} \).

In this paper we use two vector quantization algorithms: K-means and a variant of the classical Growing Neural Gas (GNG). Their main characteristics are described respectively in sections 2.1 and 2.2.
2.1 \textit{K}-means algorithm

In order to get the optimal set \( w \), the \( K \)-means algorithm minimizes the cost function corresponding to the global distortion over classes [12]:

\[
E(w) = \sum_{j=1}^{K} \sum_{\xi_i \in V_j} \|\xi_i - w_j\|^2
\]

where

\[
w_j = \frac{1}{N_j} \sum_{\xi_i \in V_j} \xi_i \quad \text{and} \quad N_j = |\{\xi_i : \xi_i \in V_j\}|
\]

The prototype \( w_j \) is thus the centroid of the \( N_j \) points \( \xi_i \) in the Voronoï cell \( V_j \). This cost function is the discrete version of equation 1.

Implementation is detailed in [12]: the main drawback of \( K \)-means is that it easily gets trapped in local minima, with the final results depending on initialization step. Moreover the final number of prototypes has to be predefined: that may be difficult especially for complex shaped distributions in high dimensional spaces (see section 2.3).

2.2 GNG with targeting algorithm (GNG-T)

The GNG-T algorithm [14] is a variant of the classical GNG algorithm [13] that minimizes a cost function that tends towards the distortion defined in equation 1. A network consisting of both a set of nodes and a graph structure that preserves the topology of the underlying probability distribution is iteratively built. This graph is made up of a set of connections between pairs of nodes that are considered as topological neighbours. It approximates the induced Delaunay triangulation of the prototypes on the manifold \( X \) and may facilitate the analysis of the probability distribution [15]. Edges are drawn up using a competitive Hebbian learning rule: the main idea is, for each input \( \xi_i \), to link the two best matching prototypes. Not only the winner, i.e. the closest prototype of the current data, but also all its direct topological neighbors are modified at each step. Influence of initialization is thus reduced. The number of nodes is iteratively adapted during algorithm execution. While a fixed prior lattice, that may be not suitable to the data, has to be chosen for other algorithms like self-organizing maps [16], no prior knowledge about the topology of the distribution is required here, and GNG-T can adapt very easily to distributions with complex shape (see section 2.3). All implementation details can be found in [14].

2.3 Comparative examples of results

Results of vector quantization for the same synthetic probability distribution consisting of a two-dimensional Gaussian mixture with three clusters are shown in figures 2(a) and 2(b) for \( K \)-means, and 2(c) and 2(d) for GNG-T. This illustrates what can be obtained for the real high dimensional data we will use. For \( K \)-means, the number \( K \) of clusters is predefined and one Gaussian cluster is traditionally associated to each node. In the proposed example it can be observed that the distribution is well described with three prototypes that allow suitable separation of the associated clusters, whereas quantization with four nodes is less satisfying: significant parts of the two clusters in the right part of the figure are mixed up in the Voronoï cell of a single prototype. For real data with unknown probability distribution
Fig. 2. Examples of vector quantization for Gaussian mixtures with three clusters: results for the same samples for $K$-means with $K = 3$ (a) and $K = 4$ (b) with associated Voronoi cells (large solid lines), and by GNG-T with complete graph (c) and with final partition (large solid lines) after non-scientific edges have been broken (d) (small dots ), crosses ($\times$) and plus signs ($+$) represent samples of the distribution, large dots are the resulting nodes, linked with edges for GNG-T only. In (e) and (f), the two lateral clusters have ten times as many samples as the central one; resulting nodes (large dots) for $K$-means with $K = 3$ in (e) can be compared with real centers of the underlying generating process (large dots) in (f).

may be difficult to redefine an adequate $K$ value. Moreover, let us suppose that one of the clusters is represented by a weak percentage of the $N$ data vectors $\xi_i$. If it is relatively close to the others, this cluster could be erroneously associated with another one because the corresponding samples would then have little weight in the global cost function. It would thus not appear in the final partition, even if the $K$ value is suitable. Such an example can be seen in figure 2(e): the central cluster is mixed with the right one, whereas the left one is represented by two prototypes (real centers of the underlying generating process can be seen in figure 2(f)).
On the contrary, for GNG-T, in an ideal case, every class should rather correspond to a connected set of nodes than to a single prototype. That is why the $K$ value can be fairly higher than the number of clusters. Moreover it adapts to the local density distribution and is thus less sensitive to both under-representation of one cluster for suitable $T$ values and to initial conditions. In the example shown in figure 2(c), three connected subgraphs corresponding to the three clusters should be obtained. Because the classes are not obviously separable and because of noise, a single connected network is obtained: to get the final clusters, some edges have to be broken, like on figure 2(d), according to rules that may depend greatly on the application (the proposed rules for kidney segmentation are described in section 3.2). Each cluster of the final partition consists in the union of the Voronoï cells of all the nodes of a given subgraph.

3 Experimental setup

3.1 Materials

The data consist in eight two-dimensional low resolution DCE-MRI sequences of normal kidney perfusion with 256 frames (5 right and 3 left kidneys). Patient age varies between 3 weeks and 51 years. Examinations were performed on a whole-body 1.5 T MR-scanner (General Electric Healthcare). A 3D ultrast character echo LAVA sequence (with $T_1$ weighting) was used with the following acquisition parameters: 15° flip angle, TR/TE 2.3 ms/1.1 ms. Interval between acquisitions is approximately 1.5 to 2 sec for the first 5 min and 9 sec for the last 6 min. For this study, only the slice with the largest proportion of renal tissue will be segmented. The initial matrix size $256 \times 256$ with pixel size between 1.172 mm and 1.875 mm for a 10 mm slice thickness. A rectangular area covering kidney was then selected (size between $47 \times 35$ and $84 \times 59$). As examination duration is about 11 minutes, kidney is moving because of respiration or more generally untimely motions of the patient. In-plane movements were eliminated by a registration algorithm including rigid transformations (translations and rotations). The similarity measure was mutual information because of fast and highly contrast modifications during perfusion [17]. All images of the sequences were registered on a reference frame (cortical peak): this avoids error accumulation in transformation computation when using iterative methods, where each frame is registered on the previous one. Mutual information is evaluated with Parzen windows: the algorithm is similar to the one developed in [18]. Besides through-plane motions could not be corrected because of a too important slice thickness. No other pre-processing treatment is applied.

3.2 Creation of the three final compartments

The vector quantization step leads to $K$ prototypes that describe the density probability of time-intensity curves of kidney pixels. Typical temporal evolution in the main perfusion phases (baseline, arterial peak, filtration, equilibrium and late phase) can be seen in figure 3. If only $K = 3$ prototypes are sufficient to describe adequately the distribution, the prototypes resulting from $K$-means algorithm are expected to be the mean temporal normalized evolution of the three sought compartments (cortex, medulla and cavities). Nevertheless, relatively great differences between time-intensity curves can be noticed inside each ROI (see figure 4), even after normalization. Euclidean distance between two vectors of two different compartments might be smaller than the disparity within a single one. Higher values
for $K$ can thus be predefined and tested. If nodes are intended to be manually merged by an operator to get the three final ROIs, the $K$ value must remain small: for instance, $K \leq 6$ is reasonable for a very easy and unambiguous grouping.

With GNG-T, in an ideal case, three connected subgraphs corresponding to the three compartments should be obtained, but as explained in section 2.3, vector quantization results most of time in a single connected network with a relatively high number of nodes (10 up to about 30). In this case, a manual grouping cannot be conceivable. Some physiologically based criteria are thus used to break properly the edges of the one-class graph in order to get the final expected ROIs. The procedure, suitable for healthy kidneys, is the following:

- Cavities extraction: cavities should be the brighter anatomical compartment in late phase (see figure 3). So all nodes whose average intensity during this perfusion phase is superior to a first threshold $t_1$ and that are directly connected in the GNG-T graph are supposed to be cavities.
- Cortex and medulla separation: filtration rate depends on tissue nature and is used to separate cortex from medulla. The filtration phase is automatically detected on the average time-intensity curve of the whole kidney for 7 of the 8 studied cases and is manually adjusted for the last one. The slope of time-intensity curves is then computed using standard linear regression for all remaining prototypes: a node is considered as cortex if the corresponding slope is smaller than a second threshold $t_2$, otherwise it is attributed to medulla.

The two thresholds are initialized so that cortex represents about 50% of the global kidney and cavities 20%. The only manual intervention consists in their coarse tuning by an operator. As the algorithm is very fast, the adjustment can be done in real time. Let us notice that the second criterion would not help to separate cavities from other compartments: indeed a theoretically unexpected but fairly high arterial peak can be seen in figure 4(c) and may be observed in all compartments because of great kidney vascularization and partial volume effect. The use of topological structure for compartment building avoids gathering together some nodes with similar contrast in late phase but whose temporal evolution differs sufficiently during the rest of the sequence.
3.3 Choice of a vector normalization method

A priori K-means could be performed directly on the time-intensity vectors $\xi_i$ stemming from the DCE-MRI sequence like in [10]. Nevertheless each vector is divided by its norm, in order to take into account rather the shape of the time-intensity curves and to attenuate illumination inhomogeneities due to acquisition. For GNG-T, as it is convenient to keep similar target values, all intensities $I$ are first replaced by $(I - I_B)/(I_L - I_B)$, where $I_B$ is the mean value for baseline and $I_L$ the mean value during late phase for the time-intensity curve of entire kidney. This operation allows to obtain a similar dynamic for all the kidneys while keeping relative order of intensities, which is not necessarily the case for the first normalization.

3.4 Manual reference segmentations by radiologists

Two experienced radiologists (OP1 and OP2) performed manual segmentation. They had both to extract the whole kidney mask and to segment the desired internal structures (cortex, medulla and cavities). The manual segmentations were performed as follows:

1. visualization of the complete registered sequence,
2. selection of a frame in the late phase where cavities are the brightest compartment and delineation of the corresponding ROI,
3. identification of the cortical peak frame and extraction of the cortex,
4. segmentation of medulla by removing cortex and cavities from the whole kidney.

The global mask was then defined as the common area between the two manually delineated entire kidneys, including the three ROIs previously segmented. This mask was used for functional segmentation as well in order to discard background pixels. Two segmentations by OP1 and OP2 can be seen in figure 5(a) and (b): the variations are due essentially to a different choice of the two frames and to some subjectivity of the operator for both kidney contour delineation and extraction of its internal structures.

3.5 Quantitative validation

One of the manual segmentations is chosen as a reference. All the other segmentations (functional ones obtained thanks to the proposed semi-automated methods or the other manual
one) will be compared to this reference. Any segmentation of one of the compartments is considered as a binary map with label 1 inside the ROI and label 0 outside. Let be R the reference segmentation and T the tested one. Four types of pixels can be distinguished by taking into account their labels in R and T:

<table>
<thead>
<tr>
<th>Pixel type</th>
<th>Label in R</th>
<th>Label in T</th>
</tr>
</thead>
<tbody>
<tr>
<td>True Positive (TP)</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>False Negative (FN)</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>False Positive (FP)</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>True Negative (TN)</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Three discrepancy criteria are then computed for each compartment:

- Percentage overlap \( PO = 100 \times TP / (TP + FN) \), i.e. percentage of pixels of the reference ROI that are in the test ROI too.
- Percentage extra \( PE = 100 \times FP / (TP + FN) \), i.e. the number of pixels that are in the test ROI while they are out of reference ROI, divided by the number of pixels in the reference ROI. Perfect segmentation would give \( PO = 100\% \) and \( PE = 0\% \). High values for both \( PO \) and \( PE \) for a given segmentation can for instance point out some oversegmentation of the corresponding compartment. A high \( PE \) associated to a weak \( PO \) can indicate a globally wrong position of the ROI.
- Similarity index \( SI = (2 \times TP) / (2 \times TP + FN + FP) \). \( SI \) is sensitive to both differences in size and location [19]. For instance two equally sized ROIs that share half of their pixels would yield \( SI = 1/2 \). A ROI covering another that is twice as little would give \( SI = 2/3 \). For a perfect segmentation the SI value would be 1.

### 4 Results and discussion

Examples of two manual segmentations and of two functional ones are shown in figure 5. For this case, size of final compartments varies between 532 and 700 pixels for cortex, 375 and 559 for medulla, 161 and 217 for cavities. Tests are performed for the following options:

- vector quantization by \( K \)-means applied on normalized vectors \( \xi \) so that \( \| \xi \| = 1 \) and manual merging for \( K = 3 \) to 6.
- vector quantization by GNG-T and merging by manual tuning of two independent thresholds.

Let us note that if GNG-T were performed on normalized vectors, the merging criteria defined in section 3.2 may be not valuable any more since the transformation is liable to modify relative intensity values. In order to test visually the qualitative consistency of segmentations, the contours of each ROI are superimposed on images from the DCE-MRI sequence (pixels out of kidney mask are set to zero and appear in black). The frames are selected in the perfusion phases during which each compartment is visible at best, i.e. arterial peak for cortex and medulla and late phase for cavities.

Visual consistency between these frames and the functional segmentations is very good for both semi-automated methods. On the contrary an example for cortex in figure 6 shows that simple intensity thresholding on a single frame does not result in a satisfying segmentation similar to the manual one: in figure 6(a) left, threshold is too low so that a great part of medulla is selected. Increasing its value does not improve the results because some cortex is then not recovered (figure 6(a) middle and right). A similar phenomenon can be observed.
Fig. 5 Examples of anatomical manual segmentations by OP1 (a) and OP2 (b), and of functional ones after vector quantization by K-means (c) and by GNG-1 (d) for a given kidney: for each segmentation, cortex is drawn on the left frame, medulla on the middle one and cavities on the right one.

Fig. 6 Segmentations of cortex (a) and cavities (b) by intensity thresholding for three increasing threshold values (from left to right).

For cavities in figure 6(b). This method will thus not be further considered nor quantitatively tested.

Quantitative comparisons between segmentations are presented in table 1 and in figures 7 and 8. Segmentations by OP1 are chosen as references. First of all, average results for segmentations by OP2, for the functional ones described above are presented in table 1. They can also be compared with results for K-means with $K = 7$ without normalization, which is the best of the methods presented in [10]. The means over the eight cases of the three similarity criteria defined in section 3.5 are evaluated. The percentage of well classified pixels is computed too: for segmentations of one compartment, it is the sum of $TP$ pixels over the eight kidneys divided by the total number of pixels for this kind of ROI. For global kidney, it is the sum of all $TP$ pixels, whatever the compartment, divided by the sum of all kidney pixels. This percentage is a little different from average overlap since little kidneys have less influence on it. Concerning $PO$ (overlap) and $PE$ (extra pixels), these criteria have to be examined together in order to highlight real improvements of the whole segmentation: most of time an increase of overlap goes indeed together with an augmentation of extra pixels, and vice versa. For instance, in the first four columns of table 1(a) and (b), one can see that both $PO$ and $PE$ values raise for cortex, whereas they decrease for medulla: this evolution points
Table 1: Discrepancy measures for segmentations of the three ROIs and global results (reference: OP1) for OP2. K-means with K = 5 to 6 with normalization (KM5n to KM6n), GNG-T and K-means with K = 7 without normalization (KM7). Results for K-means with K = 3 and for GNG-T appear in bold if they are equal or better than those for OP2 (bold italic). Criteria are percentage of well classified pixels (WCP), overlap (PO), extra pixels (PE) and similarity index (SI).

<table>
<thead>
<tr>
<th>Test</th>
<th>OP2</th>
<th>KM3n</th>
<th>KM4n</th>
<th>KM5n</th>
<th>KM6n</th>
<th>GNG-T</th>
<th>KM7</th>
</tr>
</thead>
<tbody>
<tr>
<td>WCP (%)</td>
<td>69.8</td>
<td>83.6</td>
<td>86.6</td>
<td>88.8</td>
<td>89.2</td>
<td><strong>83.7</strong></td>
<td>75.9</td>
</tr>
<tr>
<td>PO (%)</td>
<td>71.8</td>
<td>82.8</td>
<td>85.2</td>
<td>88.3</td>
<td>88.5</td>
<td><strong>83.2</strong></td>
<td>74.9</td>
</tr>
<tr>
<td>PE (%)</td>
<td>9.2</td>
<td>19.3</td>
<td>25.4</td>
<td>29.0</td>
<td>27.8</td>
<td>21.9</td>
<td>11.2</td>
</tr>
<tr>
<td>SI</td>
<td>0.79</td>
<td>0.82</td>
<td>0.82</td>
<td>0.81</td>
<td>0.82</td>
<td><strong>0.81</strong></td>
<td>0.77</td>
</tr>
</tbody>
</table>

(a) Cortex

<table>
<thead>
<tr>
<th>Test</th>
<th>OP2</th>
<th>KM3n</th>
<th>KM4n</th>
<th>KM5n</th>
<th>KM6n</th>
<th>GNG-T</th>
<th>KM7</th>
</tr>
</thead>
<tbody>
<tr>
<td>WCP (%)</td>
<td>84.8</td>
<td>72.2</td>
<td>60.5</td>
<td>62.1</td>
<td>62.2</td>
<td>73.0</td>
<td>72.7</td>
</tr>
<tr>
<td>PO (%)</td>
<td>84.0</td>
<td>72.8</td>
<td>66.1</td>
<td>62.7</td>
<td>62.8</td>
<td>73.0</td>
<td>72.7</td>
</tr>
<tr>
<td>PE (%)</td>
<td>56.6</td>
<td><strong>36.2</strong></td>
<td>30.3</td>
<td>24.9</td>
<td>24.9</td>
<td><strong>33.7</strong></td>
<td>32.1</td>
</tr>
<tr>
<td>SI</td>
<td>0.70</td>
<td><strong>0.70</strong></td>
<td>0.67</td>
<td>0.66</td>
<td>0.66</td>
<td><strong>0.71</strong></td>
<td>0.68</td>
</tr>
</tbody>
</table>

(b) Medulla

<table>
<thead>
<tr>
<th>Test</th>
<th>OP2</th>
<th>KM3n</th>
<th>KM4n</th>
<th>KM5n</th>
<th>KM6n</th>
<th>GNG-T</th>
<th>KM7</th>
</tr>
</thead>
<tbody>
<tr>
<td>WCP (%)</td>
<td>74.9</td>
<td>70.5</td>
<td>71.5</td>
<td>69.1</td>
<td>71.2</td>
<td>68.7</td>
<td>68.2</td>
</tr>
<tr>
<td>PO (%)</td>
<td>73.9</td>
<td>69.1</td>
<td>69.8</td>
<td>65.9</td>
<td>69.5</td>
<td>69.8</td>
<td>66.2</td>
</tr>
<tr>
<td>PE (%)</td>
<td><strong>16.1</strong></td>
<td><strong>14.5</strong></td>
<td>16.2</td>
<td>13.9</td>
<td>14.2</td>
<td><strong>11.1</strong></td>
<td>18.0</td>
</tr>
<tr>
<td>SI</td>
<td>0.77</td>
<td>0.75</td>
<td>0.75</td>
<td>0.74</td>
<td>0.74</td>
<td>0.77</td>
<td>0.71</td>
</tr>
</tbody>
</table>

(c) Cavities

<table>
<thead>
<tr>
<th>Test</th>
<th>OP2</th>
<th>KM3n</th>
<th>KM4n</th>
<th>KM5n</th>
<th>KM6n</th>
<th>GNG-T</th>
<th>KM7</th>
</tr>
</thead>
<tbody>
<tr>
<td>WCP (%)</td>
<td>74.9</td>
<td>77.5</td>
<td>77.0</td>
<td>76.5</td>
<td>77.1</td>
<td><strong>77.6</strong></td>
<td>73.5</td>
</tr>
</tbody>
</table>

(d) Global kidney

out a slight better overlap for cortex but also some kind of increasing oversegmentation of this compartment to the detriment of the medulla.

Quantitative comparison results between manual reference segmentation and any of the functional ones are similar to those between the same reference and a segmentation by another radiologist, whatever the type of ROI. In other words, any anatomical segmentation and any functional one are as similar as the two anatomical ones. Nevertheless the results are less good for K-means without normalization.

For all types of comparisons, similarity index is of a same order of magnitude. The percentage of globally well classified pixels is anyway a little higher for most of functional segmentations and varies between 74.4 % and 77.6 % for K-means with normalization and GNG-T, while its value is 74.9 % for manual segmentations. The worst scores for all criteria are globally obtained for medulla: cortico-medullary boundary is anyway not very sharp and this complex shaped compartment remains particularly delicate to delineate even manually. Moreover errors on internal cortex and cavities boundaries are both transferred to medulla because of its internal position in renal structure.

Concerning K-means applied on normalized vectors, very satisfying results are directly obtained with K = 3 for almost all kidneys, and are not necessarily improved on average by
increasing $K$, maybe because the distribution could be comparable to the one presented in the example of section 2.3 (figure 2). This aims to show that in most cases, normalized time-intensity curve is a discriminatory feature to directly distinguish the three types of pixels. Furthermore, no merging step is required. This was not the case for $K$-means without normalization [10]: for $K = 3$ or 4, segmentations were often not satisfying, above all because cavities whose surface is much smaller than cortex or medulla were not always recovered (a similar example of such a “cluster loss” in the case of three close together clusters for a two-dimensional distribution can be seen in figure 2(c)). Normalization leads very likely to a better homogeneity of temporal evolution within each compartment and to a better separability, reducing thus the problem due to under-representation of cavities. The results for $K$-means with $K = 3$ and for GNG-T are very similar: they are slightly better for the latter because its flexibility is a little more suited to the treatment of hard cases. In table 1 the discrepancy criteria values appear in bold if they are equal or better than the corresponding ones for manual segmentations (slanted bold). These results show that each of the three compartments is well identified, i.e., the complete segmentation is effectively satisfying. Among the different tested variants, these two methods are the most suitable for functional segmentation of healthy kidneys.

In a second step, OP1 still being the reference, dispersion of results can be observed in figure 7 for $SI$ and in figure 8 for $PO$ and $PE$: only $K$-means with $K = 3$ is represented among methods with normalization, since the resulting segmentations are the best ones. Each kidney is tested only once. Variance over the eight cases is much higher for $K$-means with $K = 7$ without normalization than for manual segmentations. On the other hand, it is most of time lower for the two other functional methods, particularly for $K$-means with $K = 3$.

5 Conclusion

Vector quantization of time-intensity curves stemming from renal perfusion DCE-MRI sequences makes possible the separation of kidney internal structures. In case the entirely automated segmentation of internal renal structures from global kidney mask by $K$-means with three clusters is not satisfying, quantization with a higher number of prototypes followed by a merging step offers more flexibility to process hard cases. Quantitative comparison results between manual reference segmentation and any of the functional ones are similar to those between the same reference and a segmentation by another radiologist. Fur-
Moreover the mean functional curves are nearly the same for any technique. The proposed methods require very little manual intervention: for \( K \)-means, very satisfying results are directly obtained with \( K = 3 \), so that no merging step is required; for GNG-I, the only manual intervention consists in a coarse real-time tuning of two independent thresholds: the first to extract cavities by breaking the most suitable edges, i.e. by adding or removing a relatively numerous set of pixels with homogeneous temporal behavior, the second to separate in the same way medulla from cortex. GNG-I can be considered as an alternative solution offering even more flexibility than \( K \)-means: the vector quantization step allows to take into account the whole temporal evolution and to reduce noise before using physiological properties during the most characteristic perfusion phases. Dispersion of similarity criteria is lower for these two functional methods than for manual segmentation. Time-saving for both semi-automated methods is considerable: manual delineation of compartments lasts about 10 minutes for one sequence, functional segmentations about only 20 seconds. Additional tests will be carried out on a larger data base of three-dimensional acquisitions with both healthy and pathological kidneys. For the latter, as new temporal evolutions may appear, an adaptation of merging criteria after quantization by GNG-I as well as a search for optimal \( K \) value may be necessary; it is even probable that expected differences of pixel representativeness will justify the use of GNG-I rather than \( K \)-means.

References