

FUNCTIONAL SEMI-AUTOMATED SEGMENTATION OF RENAL DCE-MRI SEQUENCES

Chevallier B.⁽¹⁾, Ponvianne Y.⁽²⁾, Collette J.L.⁽¹⁾, Mandry D.⁽²⁾, Claudon M.⁽²⁾ and Pietquin O.^(1,2)

⁽¹⁾ SUPELEC-Metz campus, IMS Research Group, 2 rue E. Belin, 57070 Metz (France)

⁽²⁾ IADI, INSERM, ERI 13, Tour Drouet CHU Nancy Brabois 54500 Vandoeuvre-les-Nancy - France
Nancy University Nancy - Tour Drouet CHU Nancy Brabois 54500 Vandoeuvre-les-Nancy - France

ABSTRACT

In dynamic contrast-enhanced magnetic resonance imaging (DCE-MRI), segmentation of internal kidney structures is essential for functional evaluation. Manual morphological segmentation of cortex, medulla and cavities remains difficult and time-consuming especially because the different renal compartments are hard to distinguish on a single image. We propose to test a semi-automated method to segment internal kidney structures from a DCE-MRI registered sequence. As the temporal intensity evolution is different in each of the three kidney compartments, pixels are sorted according to their time-intensity curves using a k-means partitioning algorithm. No ground truth is available to evaluate resulting segmentations so a manual segmentation by a radiologist is chosen as a reference. We first evaluate some similarity criteria between the functional segmentations and this reference. The same measures are then computed between another manual segmentation and the reference. Results are similar for the two types of comparisons.

Index Terms— Image segmentation, biomedical magnetic resonance imaging, biomedical image processing.

1. INTRODUCTION

In DCE-MRI renal anatomical compartment identification is essential for functional kidney evaluation. Segmentation of cortex, medulla and cavities is usually performed manually by a radiologist and can be quite time-consuming and fastidious. Some semi-automated methods like thresholding, active contours and region-based techniques have been applied in the medical field but few to renal DCE-MRI [1]. Furthermore it remains difficult to distinguish internal renal structures from a single image because of highly changing contrast during perfusion. To delineate each of the three anatomical compartments radiologists select images that seem to be the most relevant in different phases of the perfusion. They anyway use only very few frames of the whole temporal sequence. Thus a slight error in registration or any through-plane motion can lead to great variations in functional results.

Some authors use time-intensity series to sort the renal voxels according to their functional profile. Yet very few validation results for real data have been exposed except qual-

itative consistency with manual segmentation or comparison between renograms [2].

We propose to test a semi-automated method for functional segmentation of renal cortex, medulla and pelvocaliceal cavities based on k-means clustering of pixel temporal contrast evolution. The proposed technique requires less manual intervention and may offer more robustness and reproducibility thanks to the possibility to use the whole sequence or at least a great part of it and not only a few frames. The resulting segmentations will be compared quantitatively with anatomical manual ones. As a reference point, discrepancy measures between two manual segmentations will be performed too.

2. METHOD FOR FUNCTIONAL SEGMENTATION

Let us suppose we have a DCE-MRI sequence with N_T images. Each of the N_T images is first registered to a reference image I_R in order to correct respiratory motion. We get then the temporal evolution of contrast for each pixel of I_R .

Our purpose is not to extract kidney but to compare only its internal segmentation in three regions of interest (ROI), which are cortex, medulla and cavities (see figure 1 for examples of frames and figure 3 for ROI masks determined on the same kidney). Consequently a global mask with N pixels is created before functional segmentation. The N pixels of I_R are then classified in K clusters $C_j, 1 \leq j \leq K$ according to their temporal contrast evolution. Supervised classification cannot be considered here because we do not have access to enough manually segmented sequences for building up a training data set. Among unsupervised classification methods, k-means partitioning algorithm is well suited to our problem because of its simplicity and because vectors to be classified are strongly correlated.

Let be $\{x_i, 1 \leq i \leq N\}$ the N pixels of I_R to be classified. A N_T -components vector $X_i = (P_1 X_{i1}, \dots, P_{N_T} X_{iN_T})$ is associated with each pixel, where X_{ip} is the contrast at time p for the pixel x_i and (P_1, \dots, P_{N_T}) a weight vector. In order to get the optimal partition $C = \{C_j, 1 \leq j \leq K\}$ the k-means algorithm minimizes the cost function corresponding to the global distortion over classes

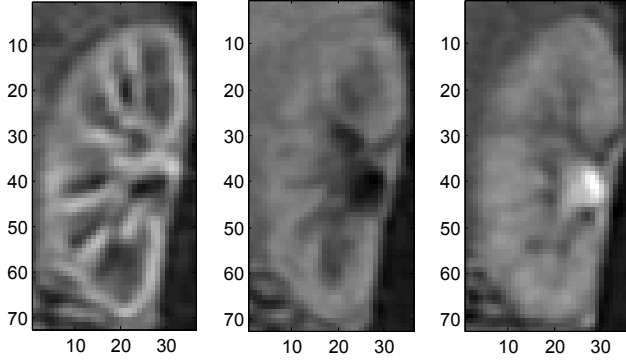


Fig. 1. Example of frames during arterial peak, filtration and late phase

$$I(C) = \sum_{j=1}^K \sum_{x_i \in C_j} \|X_i - \mu_j\|^2 \quad (1)$$

where μ_j is the centroid or prototype vector of all the points x_i in the cluster C_j , i.e.

$$\mu_j = \sum_{x_i \in C_j} X_i \quad (2)$$

The weight vector is determined according to the time-intensity curve of the entire kidney in order to enforce the more significant perfusion phases (baseline, arterial peak, filtration and late phase) and to reduce the influence of equilibrium phase, where the three compartments have similar contrast (see figure 2).

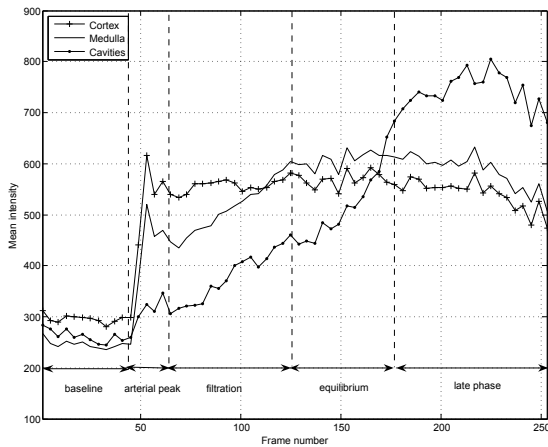
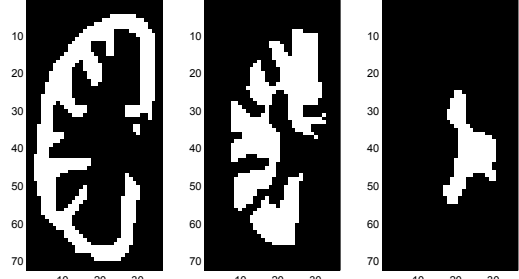
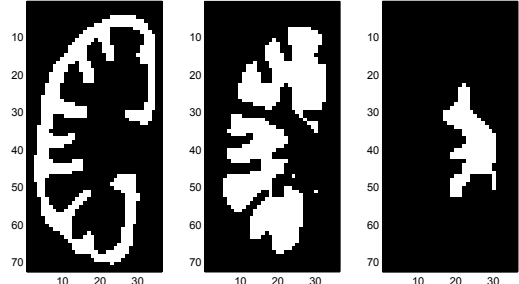


Fig. 2. Typical time-intensity curves for cortex, medulla and cavities

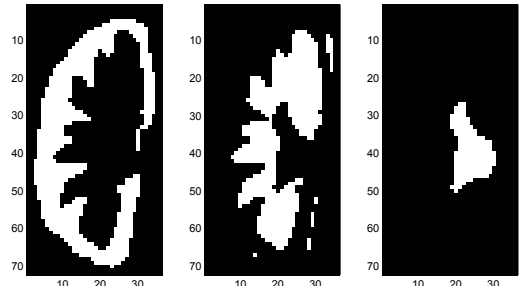
Every cluster is then associated with the anatomical compartment of the reference anatomical segmentation that has the more common pixels with it. This allows getting three areas corresponding to the three anatomical compartments but an observer could also easily merge some of the K clusters to get the same result.



(a) Anatomical manual segmentation (OP1)



(b) Anatomical manual segmentation (OP2)



(c) Functional semi-automated segmentation

Fig. 3. Example of cortex, medulla and cavities segmentations

3. EXPERIMENT

3.1. Materials

Eight two-dimensional DCE-MRI sequences of normal kidney perfusion with 256 low resolution frames (initial matrix size 256×256 , pixel size between 1,172 mm and 1,875 mm) are used. A rectangular area covering kidney was selected: corresponding matrix sizes varie between 47×35 and 84×59 . To correct respiratory motion a rigid registration including translations and rotation was performed before segmentation; mutual information was used as a similarity measure because of fast changing contrast during perfusion [3]. Nevertheless through-plane motions remain and acquisition is very noisy. Three images of different perfusion phases can be seen on figure 1.

Test	OP2	Auto	Auto	Auto
Clusters	3	5	6	7
WCP	69.8	74.3	75.4	75.9
PO	71.8	70.9	75.3	74.9
PE	9.2	17.0	19.5	17.3
SI	0.79	0.75	0.77	0.77
MD	0.6	0.7	0.7	0.6

(a) Cortex

Test	OP2	Auto	Auto	Auto
Clusters	3	5	6	7
WCP	84.4	79.2	78.4	75.7
PO	84.0	79.7	77.0	76.7
PE	56.6	63.1	55.4	52.1
SI	0.70	0.67	0.66	0.68
MD	1.0	1.3	1.2	1.1

(b) Medulla

Test	OP2	Auto	Auto	Auto
Clusters	3	5	6	7
WCP	74.9	59.1	63.1	68.2
PO	73.9	57.2	60.9	66.2
PE	16.1	7.0	8.0	18.0
SI	0.77	0.69	0.70	0.71
MD	0.8	0.7	0.6	1.1

(c) Cavities

Test	OP2	Auto	Auto	Auto
Clusters	3	5	6	7
WCP	74.9	71.3	72.7	73.5

(d) Percentage of well classified pixels in global kidney

Test	OP1	Auto	Auto	Auto
Clusters	3	5	6	7
WCP	89.7	78.2	80.9	81.7
PO	89.0	74.0	79.3	79.2
PE	36.1	35.6	36.8	34.1
SI	0.79	0.70	0.73	0.74
MD	0.7	0.9	0.8	0.8

(a) Cortex

Test	OP1	Auto	Auto	Auto
Clusters	3	5	6	7
WCP	60.3	68.2	68.0	66.4
PO	60.5	69.9	67.1	67.9
PE	11.8	32.4	25.8	22.4
SI	0.70	0.69	0.68	0.71
MD	0.80	1.2	1.1	1.0

(b) Medulla

Test	OP1	Auto	Auto	Auto
Clusters	3	5	6	7
WCP	81.8	63.6	66.9	74.1
PO	82.2	63.3	65.9	72.8
PE	32.4	9.9	11.5	19.3
SI	0.77	0.72	0.73	0.76
MD	0.9	0.6	0.6	0.9

(c) Cavities

Test	OP1	Auto	Auto	Auto
Clusters	3	5	6	7
WCP	74.9	72.6	73.5	73.8

(d) Percentage of well classified pixels in global kidney

Table 1. Similarity measures for segmentations of the three ROI, where OP1 is considered as a reference

3.2. Morphological manual segmentation

Two experienced radiologists (OP1, OP2) reviewed the dynamic registered sequences in order to delineate three ROI which are the cortex, the medulla and the pelvo-caliceal cavities. The procedure was as follows:

- visualization of the sequence,
- selection of a frame in the late phase with cavities well delineated and segmentation of the cavities,
- selection of the cortical enhancement peak, allowing segmentation of the cortex and the medulla (by difference with the cavities previously underlined).

The entire kidney mask is the common area of the two entire kidneys delineated by the two radiologists and the three internal ROI are included in this global mask. The same mask is used for functional segmentation. An example of such manual segmentations is shown on figure 3(a) et (b).

Table 2. Similarity measures for segmentations of the three ROI, where OP2 is considered as a reference

3.3. Comparing morphological and functional segmentations

For each of the eight subjects a morphological manual segmentation is chosen as a reference. The test segmentation can be the other morphological manual segmentation or a functional segmentation obtained by the proposed method.

The two segmentations to be compared can be considered as two binary maps R (reference) and T (test) with label 1 inside the ROI and 0 outside. Four types of pixels can then be defined, according to their labels in the two maps:

Pixel type	Label in R	Label in T
True positive (TP)	1	1
False Negative (FN)	1	0
False Positive (FP)	0	1
True Negative (TN)	0	0

Four similarity measures between the reference segmen-

tation and the test segmentation are then computed for each ROI:

- percentage overlap $PO = 100 \times TP / (TP + FN)$, i.e. percentage of pixels of the reference ROI recovered in test segmentation,
- 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,
- similarity index $SI = (2 \times TP) / (TP + FN + FP)$. SI is sensitive to both differences in size and location [4]. For instance two equally sized ROI 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.
- mean distance (in pixel) between contours of test and reference segmentation (MD).

Among these criteria only SI does not depend on the choice of a reference segmentation but the same values appear twice in the results table to make the comparison easier.

4. RESULTS

Examples of two manual anatomical segmentations and of a functional semi-automated segmentation are shown on figure 3: the total number of pixels varies here between 604 and 750 for cortex, 652 and 667 for medulla, 159 and 204 for cavities. Segmentations are tested for a predefined number of clusters K that varies between 5 and 7. The k-means algorithm minimizes global distortion, so cavities whose surface is much smaller than cortex or medulla are often not identified for $K = 3$ or 4 because they may have little weight in cost function. On the other hand pixels in the same anatomical compartment (particularly in cortex or medulla) can have relatively different time-intensity curves and are split up in several clusters. For $K \geq 8$ some clusters that have very few pixels might be not relevant; moreover it may become difficult for the observer to merge them unambiguously.

Tables 1 and 2 show means over the eight subjects and the percentage of globally well classified pixels (WCP) for each type of ROI and for the entire kidney regions respectively for OP1 and OP2 as a reference. WCP for a given ROI is the total number of TP pixels over the eight subjects, divided by the total number of pixels for this ROI. WCP for entire kidney is the total number of TP for all the ROI over the eight subjects, divided by the total number of classified pixels. WCP is a little different from average PO inasmuch as cases with small kidneys have less influence on WCP.

Discrepancy measures between two manual segmentations are similar to those observed between automatic and any of the manual segmentations. In particular globally WCP are almost the same. Nevertheless cavities are rather underestimated by automatic segmentation in comparison with manual ones but as a result, pixels classified as cavities in

functional segmentation are most of the time recognized as cavities by radiologists. The worst results are obtained for medulla, with a particularly high PE, but this is generally true both for manual and automatic segmentations: medulla is an intermediate area with complex shape and remains anyway difficult to segment manually.

Increasing the number of clusters does not systematically improve each criteria, even if the global WCP and most of the time SI become a little larger. For instance an increase in PO for cavities is coupled with an increase in PE.

5. CONCLUSION AND PERSPECTIVES

We have tested a method for functional segmentation of kidney structures from DCE-MRI sequences and compared the resulting segmentations with manual ones. Discrepancy measures between two manual segmentations are similar to those observed between automatic and any of the manual segmentations, indicating that the semi-automated method might be suitable for renal segmentation in DCE-MRI. Functional curves are directly obtained from the prototype vectors resulting from the k-means algorithm. Moreover generated time saving is considerable: manual segmentation requires between 12 and 15 minutes for one sequence versus about twenty seconds for automatic segmentation and cluster merging by an observer. As a further work, we intend to optimize the weight vector and select automatically the number of clusters. Pixels that remain spatially isolated after merging (see figure 3(c)) are likely misclassified and would require a specific treatment. We plan also to do similar tests in case of automatic kidney extraction.

6. REFERENCES

- [1] N. Michoux, J.P. Vallee, A. Pechere-Bertschi, X. Montet, L. Buehler, and B. Beers, "Analysis of contrast-enhanced mr images to assess renal function," *MAGMA Magnetic Resonance Materials in Physics, Biology and Medicine*, vol. 19, no. 4, pp. 167 – 179, 2006.
- [2] T. Song, V.S. Lee, H. Rusinek, J.B. Sajaous, and A.F. Laine, "Registration and segmentation of dynamic three-dimensional mr renography based on fourier representations and k-means clustering," in *International Society for Magnetic Resonance in Medicine (ISMRM) 13 th Scientific Meeting*, 2005.
- [3] J.P.W. Pluim, J.B.A. Maintz, and M.A. Viergever, "Mutual-information-based registration of medical images: a survey," *IEEE Trans. Med. Imaging*, vol. 22, no. 8, pp. 986 – 1004, 2003.
- [4] A.P. Zijdenbos, B.M. Dawant, R.A. Margolin, and A.C. Palmer, "Morphometric analysis of white matter lesions in mr images: method and validation," *IEEE Trans. Med. Imaging*, vol. 13, no. 4, pp. 716 – 24, 1994.