Automatic segmentation of subcutaneous mouse tumors by multiparametric MR analysis based on endogenous contrast

Magn Reson Mater Phy (2015) 28:363–375
DOI 10.1007/s10334-014-0472-1

RESEARCH ARTICLE

Automatic segmentation of subcutaneous mouse tumors by multiparametric MR analysis based on endogenous contrast

Stefanie J. C. G. Hectors · Igor Jacobs · Gustav J. Strijkers · Klaas Nicolay

Received: 19 July 2014 / Revised: 26 October 2014 / Accepted: 29 October 2014 / Published online: 27 November 2014

Abstract
Object Contrast-enhanced T_1-weighted imaging is usually included in MRI procedures for automatic tumor segmentation. Use of an MR contrast agent may not be appropriate for some applications, however. We assessed the feasability of automatic tumor segmentation by multiparametric cluster analysis that uses intrinsic MRI contrast only.
Materials and methods Multiparametric MRI consisting of quantitative T_1, T_2, and apparent diffusion coefficient (ADC) mapping was performed in mice bearing subcutaneous tumors ($n = 21$). *k*-means and fuzzy *c*-means clustering with all possible combinations of MRI parameters, i.e. feature vectors, and 2–7 clusters were performed on the multiparametric data. Clusters associated with tumor tissue were selected on the basis of the relative signal intensity of tumor tissue in T_2-weighted images. The optimum segmentation method was determined by quantitative comparison of automatic segmentation with manual segmentation performed by three observers. In addition, the automatically segmented tumor volumes from seven separate tumor data sets were quantitatively compared with histology-derived tumor volumes.
Results The highest similarity index between manual and automatic segmentation ($SI_{manual,automatic} = 0.82 \pm 0.06$) was observed for *k*-means clustering with feature vector $\{T_2, ADC\}$ and four clusters. A strong linear correlation between automatically and manually segmented tumor volumes ($R^2 = 0.99$) was observed for this segmentation method. Automatically segmented tumor volumes also correlated strongly with histology-derived tumor volumes ($R^2 = 0.96$).
Conclusion Automatic segmentation of mouse subcutaneous tumors can be achieved on the basis of endogenous MR contrast only.

Keywords Tumor segmentation · Multiparametric MRI · Endogenous contrast · Cluster analysis

S. J. C. G. Hectors (✉) · I. Jacobs · G. J. Strijkers · K. Nicolay
Biomedical NMR, Department of Biomedical Engineering, Eindhoven University of Technology, P.O. Box 513, 5600 MB Eindhoven, The Netherlands
e-mail: s.j.c.g.hectors@tue.nl

G. J. Strijkers
Biomedical Engineering and Physics, Academic Medical Center, University of Amsterdam, Amsterdam, The Netherlands

Introduction

Accurate tumor volume delineation is often necessary in (pre)clinical cancer research and clinical practice. Tumor size measurements are widely used to assess tumor treatment efficacy [1]. Furthermore, image-guided cancer-treatment techniques, including MRI-guided high-intensity focused ultrasound (HIFU) and radiotherapy, require accurate tumor demarcation to define the target volume for treatment. Currently, tumor delineation is usually performed by manual segmentation of the tumor tissue, which is time-consuming and prone to inter-observer variability. Automatic tumor segmentation would facilitate more robust, rapid, objective, and reproducible delineation of the tumor tissue.

Multiple methods for (semi)automatic tumor segmentation based on MRI images have been described for different tumors, including breast [2–6], prostate [7, 8], brain [9–11], and head and neck [12] tumors. A variety of segmentation algorithms were used in these studies, including volume growing, threshold-based methods,

and clustering algorithms, for example fuzzy *c*-means and fuzzy connectedness. Contrast-enhanced imaging is commonly included in MRI procedures for automated tumor segmentation. However, if tumor segmentation has to be performed shortly before treatment, the presence of the gadolinium (Gd) contrast agent during the treatment might have adverse effects. It has recently been reported that presence of Gd during HIFU treatment may be inappropriate, because Gd-induced magnetic susceptibility changes in the tissue cause inaccuracies in the temperature maps acquired during treatment [13]. Furthermore, tumor ablation with HIFU could induce entrapment of Gd in the treated tissue, possibly resulting in prolonged retention of Gd by the body and potential release of free Gd^{3+} ions from the chelates [14]. These free Gd^{3+} ions have been strongly associated with the development of nephrogenic systemic fibrosis (NSF) [15]. The disadvantages of contrast-enhanced T_1-weighted imaging for treatment planning are, however, not restricted to HIFU treatment. The presence of the MR contrast agent may also be inappropriate if treatment evaluation with MRI has to be performed immediately after treatment, because the effect of the contrast agent on the post-treatment MRI data could lead to misinterpretation of the effects of treatment. Furthermore, apart from treatment-related disadvantages of injection of Gd contrast agents, in clinical practice, injection of Gd contrast agent is contraindicated for patients with severely impaired renal function, because of the increased risk of NSF [16]. Cancer patients are especially vulnerable to renal toxicity as a result of contrast agents, because they are frequently exposed to nephrotoxic chemotherapeutics and may be dehydrated because of chemotherapy-related and cancer-related nausea and vomiting [17].

There have been few studies of automatic tumor segmentation methods based on endogenous contrast MRI. Hsieh et al. [18] reported a method for automatic segmentation of meningiomas on the basis of T_1 and T_2-weighted images. Although successful segmentation was achieved for most tumors, the method failed in approximately 20 % of cases. The authors reported that noticeable edema was present in the brain in these failure cases, indicating that the method could not distinguish between tumor and edematous tissue. A recent preclinical study demonstrated semi-automatic size measurement for tumors subcutaneously inoculated in the neck region of mice [19]. A threshold-based method was applied to T_2-weighted images to segment the tumor tissue from the surrounding tissue. Additional manual delineation had to be performed to exclude regions where the segmentation had failed.

These previous reports on automated tumor segmentation based on endogenous contrast suggest that a more advanced MR procedure is required to improve the accuracy of automatic tumor segmentation. Specifically, we hypothesize that addition of apparent diffusion coefficient (ADC) mapping to endogenous contrast MRI methods may result in improved discrimination between tumor tissue and surrounding tissue, for example edema. Cancerous tissue typically has a lower ADC than non-cancerous tissue, because of high cell density, and thus increased effects of diffusion-hindering obstacles, for example cell membranes and other macromolecular structures [20]. In contrast, the peri-tumoral edematous tissue usually has a high ADC because of low cellularity. It has been shown that the combination of ADC maps and T_2-weighted images improves the sensitivity and specificity of detection of prostate cancer by visual inspection [21].

The purpose of this study was to assess whether accurate automatic tumor segmentation could be achieved on the basis of combined analysis of quantitative T_1, T_2, and ADC maps. Clustering-based algorithms were used as segmentation method, because these algorithms are particularly suitable for segmentation on the basis of multiparametric data. The purpose of the study was to assess the general feasibility of clustering algorithms based on endogenous MRI data for automatic tumor segmentation. Therefore, as a first step, the implemented algorithms were applied to segmentation of subcutaneous mouse tumors. To determine the optimum clustering method for automatic delineation of the tumor tissue, *k*-means clustering and fuzzy *c*-means clustering were performed with all possible combinations of MR parameters, i.e. feature vectors. Furthermore, the number of clusters was varied for each feature vector. The optimum segmentation method, i.e. the method that yielded the best segmentation of the tumor, was determined by quantitative comparison of automatic tumor segmentation and manual tumor segmentation performed by three experienced observers. Quantitative correlation analysis was also performed between automatically segmented tumor volumes and histology-derived tumor volumes.

Materials and methods

Tumor model

CT26.WT mouse colon carcinoma cells (American Type Culture Collection; ATCC; CRL-2638) were cultured as a monolayer at 37 °C and 5 % CO_2 in RPMI-1640 medium (Invitrogen, Breda, The Netherlands), supplemented with 10 % fetal bovine serum (Greiner Bio-One, Alphen a/d Rijn, The Netherlands) and 50 U/ml penicillin and streptomycin (Lonza Bioscience, Basel, Switzerland). Early passages (5–10) of the original ATCC batch were used for inoculation.

Ten to twelve-week-old Balb/c mice ($n = 21$; Charles River, Maastricht, The Netherlands) were inoculated with

2×10^6 CT26.WT cells subcutaneously in the right hind limb. Approximately 10 days after inoculation, tumors became palpable in all animals.

MRI measurements

MRI measurements were performed with a 6.3-T scanner (Bruker BioSpin, Ettlingen, Germany) using a 3.2-cm-diameter quadrature birdcage RF coil (Rapid Biomedical, Rimpar, Germany). The mice were positioned in a custom-made cradle, equipped with a mask for anesthetic gas (1–2 % isoflurane). Respiration was monitored with a balloon sensor. Animal temperature was monitored and maintained at body temperature by use of a warm water pad. For reduction of susceptibility artefacts in the echo planar imaging (EPI) sequences, the tumor-bearing limb was covered with degassed ultrasound gel (Aquasonic® 100; Parker Laboratories). Artefacts were further reduced by local shimming of the tumor-bearing limb.

The multi-slice imaging procedure, covering the whole tumor, consisted of fat-suppressed T_2-weighted imaging (spin-echo, echo time TE = 30 ms, repetition time TR = 1,000 ms, number of averages NA = 1), quantitative T_1 mapping (inversion recovery look-locker EPI, TE = 8 ms, TR = 10,000 ms, inversion time = 30 ms, flip angle = 20°, pulse separation = 400 ms, number of points = 15, NA = 2), T_2 mapping (MLEV-prepared GE-EPI [22], 7 TEs ranging from 1 to 82 ms, TR = 2,000 ms, NA = 2), and ADC mapping (double spin-echo prepared EPI, TE = 41 ms, TR = 4,000 ms, *b*-values 0, 100, 200, and 400 s/mm^2, three orthogonal directions, NA = 4). All acquired images had a matrix size of 128 × 128, FOV of 4 × 4 cm^2 and 1 mm slice thickness. Irrespective of tumor size, a minimum of twelve slices was acquired. For tumors extending outside these twelve slices, the number of slices was increased (to a maximum of sixteen slices) to cover the entire tumor volume.

Image processing and generation of maps

Image analysis was performed in Mathematica 8.0 (Wolfram Research, Champaign, IL, USA). Maps were calculated on a pixel-by-pixel basis in each slice. Values were only determined for pixels with substantial signal intensity (>0.05 times the maximum signal intensity) in the T_2-weighted images. T_1 maps were generated as described elsewhere [23]. T_2 maps were calculated from mono-exponential fitting of the multi-echo data. For generation of ADC maps, mono-exponential fitting was performed by use of the signal intensities at the different *b*-values for each diffusion-encoding direction separately. Next, ADC values of the different directions were averaged to obtain the final (orientation-invariant) ADC value for each pixel.

Automatic tumor segmentation

A flow chart of the automatic tumor segmentation method is shown in Fig. 1. Tumor segmentation based on the acquired maps was performed by either *k*-means or fuzzy *c*-means clustering. Fuzzy c-means clustering is a soft version of *k*-means clustering, in which each pixel has a certain amount of membership of each cluster. In applications where there is overlap between the clusters in the data set, fuzzy *c*-means clustering is more suitable than *k*-means [24]. Clustering was performed with all possible feature vectors, namely $\{T_1\}$, $\{T_2\}$, {ADC}, $\{T_1$, ADC$\}$, $\{T_2$, ADC$\}$, $\{T_1, T_2\}$ and $\{T_1, T_2$, ADC$\}$ and with different numbers of clusters, ranging from 2 to 7. For fuzzy *c*-means clustering the fuzziness index was set to 2. The clustering algorithms were iterated until the method converged or the maximum number of iterations of 50 was reached.

The segmentation method was performed on large rectangular regions around the tumor-bearing limb. These regions contained tumor tissue and surrounding muscle, oedema, and bone. Before segmentation, pixels within the ultrasound gel were excluded from the defined rectangular regions by application of a threshold to the T_2-weighted images. This led to segmentation of the hyperintense ultrasound gel and the relatively hypo-intense tissue. To correct for scaling differences between the different parameters, features were normalized (mean = 0, standard deviation SD = 1) before clustering was performed.

Prior knowledge of specific characteristics of the tumor tissue is needed to enable the algorithm to automatically assign clusters as tumor or non-tumor tissue. In the example of the subcutaneous tumors presented here, we made use of the typically higher signal intensity of the tumor tissue compared with the surrounding muscle tissue in T_2-weighted images. An initial tumor mask was generated by a threshold-based segmentation method that was based on the T_2-weighted images. The threshold-based tumor segmentation is illustrated on the right of Fig. 1. For this method, the T_2-weighted images from the different animals were normalized by dividing the pixel intensities by the average signal intensity of the ultrasound gel surrounding the tumor-bearing limb. Next, a histogram of the normalized signal intensities in the T_2-weighted images averaged for all animals was generated. A clear peak that, on the basis of visual inspection, belonged to the tumor tissue was observed in this averaged histogram. The minimum and maximum intensity values for this peak were found to be 0.45 and 0.65, respectively. Subsequently, all pixel

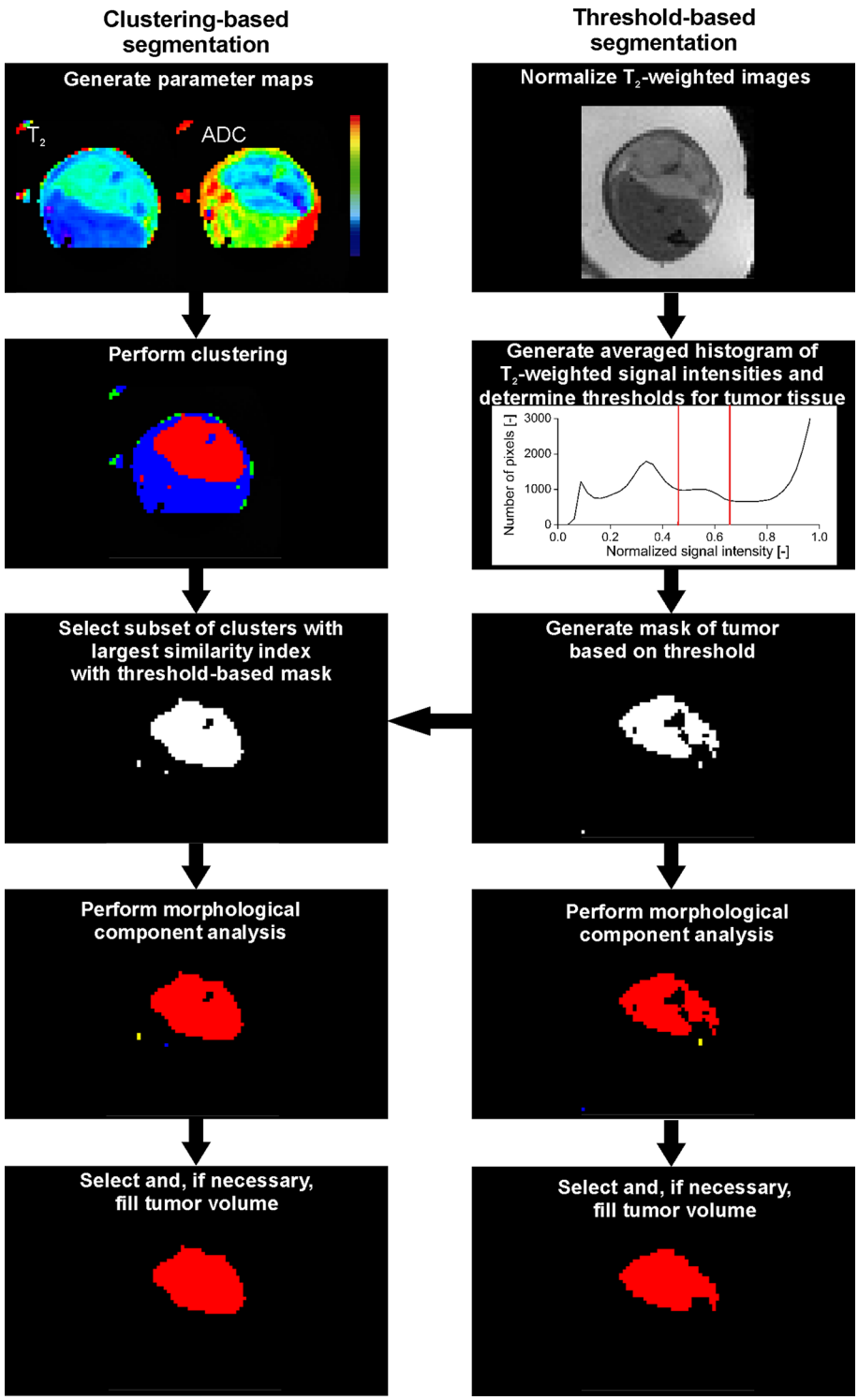

Fig. 1 Flowchart of the clustering-based and threshold-based tumor segmentation methods. The tumor mask resulting from the threshold-based segmentation method was used for automatic selection of the tumor clusters in the clustering-based segmentation method. For details see text

values between these minimum and maximum intensities were set to 1, which resulted in a tumor mask that could be used as input for the clustering-based algorithm to enable automatic determination of the subset of clusters belonging to the tumor tissue. The similarity index (SI) between the threshold-based tumor segmentation and the clustering-based segmentation method was determined for all possible subsets of clusters. The SI is a measure of the area defined as tumor tissue by both methods relative to the total segmented area, and was determined by use of Eq. 1 [9, 25].

$$SI_{threshold,clustering} = \frac{2 * S_{threshold} \cap S_{clustering}}{2 * S_{threshold} \cap S_{clustering} + S_{threshold} \cap \overline{S_{clustering}} + S_{clustering} \cap \overline{S_{threshold}}} \quad (1)$$

In this equation $S_{threshold}$ and $S_{clustering}$ are the tumor segmentations resulting from threshold-based and clustering-based segmentation, respectively. The $\cap$ symbol represents intersection.

The tumor clusters were determined by automatic selection of the subset of clusters for which the $SI_{threshold,clustering}$ was maximum. To eliminate small groups of pixels outside the large tumor volume that were assigned to (one of) the tumor cluster(s), connected components analysis was performed on all pixels that were assigned to the tumor cluster(s) (Mathematica command morphological components). This connected components method segments an image into different groups of connected pixels. The largest component resulting from this analysis was assumed to be the tumor. Region growing was applied on this component to include pixels within the tumor that were not assigned to the tumor cluster(s), e.g. because of naturally occurring necrotic tumor tissue that has MR values different from those of viable tumor tissue and may, therefore, initially not be selected as tumor tissue by the segmentation method. To compare the performance of the clustering-based segmentation methods with that of the threshold-based segmentation method, the same procedure of morphological component analysis and region growing was also applied to the generated threshold-based tumor mask.

Manual tumor segmentation

Manual segmentation was performed by three independent observers (O_1, O_2, and O_3). All three observers have substantial experience (>4 years) in MRI of subcutaneous mouse tumors. Regions of interest (ROIs) around the tumor tissue were drawn on the 3rd echo images of the T_2 mapping procedure, because this image has approximately the same echo time (28 ms) as the spin-echo T_2-weighted image (30 ms). ROI definition was performed on the images from the T_2 mapping rather than on the spin-echo T_2-weighted images, because, although maximum effort was taken to minimize the effects of susceptibility artefacts in the EPI acquisitions, minor geometric distortions could still be present in the EPI acquisitions. These distortions could lead to a slight misregistration between the spin-echo T_2-weighted imaging and the EPI-based T_1, T_2, and ADC acquisitions.

Evaluation of automatic segmentation methods

To determine the optimum automatic segmentation method, the tumor delineations resulting from the different segmentation methods were quantitatively compared with the manual tumor delineations. The performance of the segmentation methods was assessed by calculation of the specificity, sensitivity, and similarity index ($SI_{manual,automatic}$) between manual and automatic tumor segmentation. An SI >0.7 is indicative of good segmentation [26].

The sensitivity, specificity, and $SI_{manual,automatic}$ were determined by use of Eq. 2, 3 and 4, respectively:

$$\text{Sensitivity} = \frac{TP}{TP + FN} \quad (2)$$

$$\text{Specificity} = \frac{TN}{TN + FP} \quad (3)$$

$$SI_{manual,automatic} = \frac{2 * TP}{2 * TP + FP + FN} \quad (4)$$

In these equations, TP is the number of true-positive pixels, FP is the number of false-positive pixels, TN is the number of true-negative pixels, and FN is the number of false-negative pixels.

The sensitivity, specificity, and $SI_{manual,automatic}$ values of all assessed segmentation methods were determined for each observer separately. The optimum segmentation method was defined as the method for which the $SI_{manual,automatic}$ value, averaged for the different observers, was maximum.

Assessment of inter-observer variability

Manual segmentation by the different observers were compared by calculation of the above-mentioned SI. The SI between two different manual segmentations was assessed by use of the equation:

$$SI_{O_i,O_j} = \frac{2 * S_i \cap S_j}{2 * S_i \cap S_j + S_i \cap \overline{S_j} + S_j \cap \overline{S_i}}, \quad (5)$$

in which S_i and S_j are the manual segmentations of observers O_i and O_j, respectively.

Comparison of the optimum automatic segmentation method with histology

The segmentation method that was considered optimum on the basis of the above measures of performance was run on seven separate MRI data sets of the same tumor model that had been acquired by use of the same MRI procedure as described above. For these data sets, histological sections of the tumors excised directly after acquisition of the MRI

Table 1 Average similarity index between the three manual tumor segmentations

Observer	O_1	O_2	O_3
O_1	1	0.86 ± 0.04	0.84 ± 0.04
O_2	–	1	0.87 ± 0.04
O_3	–	–	1

Values are mean ± SD ($n = 21$)

data were available. The 5-µm-thick cryosections covered the entire tumor volume and were cut with a 300-µm inter-section distance. The cryosections were stained for nicotinamide adenine dinucleotide diaphorase (NADH) activity, which is a marker for cell viability (the procedure is described elsewhere [27]). All sections were imaged with brightfield microscopy consisting of mosaic acquisition at 5× magnification. Analysis of the microscopy images was done in Mathematica. ROIs were manually drawn around the entire tumor tissue on each section. Subsequently, histology-derived tumor volumes were calculated by multiplication of the sum of the areas of tumor tissue in all sections with the inter-section distance.

Statistics

All data are presented as mean ± SD. The $SI_{manual,automatic}$ values from the different segmentation methods were statistically compared by use of a paired t test. The effect of different observers on the $SI_{manual,automatic}$ values was assessed by ANOVA for repeated measures. If the effect of observer on $SI_{manual,automatic}$ proved statistically significant, Bonferroni post-hoc tests were performed to compare the $SI_{manual,automatic}$ values of the individual observers. For all tests, the level of significance was set to $\alpha = 0.05$.

The automatically and manually segmented tumor volumes were compared by linear regression that consisted of fitting the data points to the line $y = a * x + b$. A similar analysis was performed for comparison of tumor volumes from histology with automatically segmented tumor volumes. Outliers that significantly affected the linear regression were identified by calculation of the Cook's distance for each data point. A data point was regarded as an outlier if its Cook's distance was greater than $4/(n - k - 1)$, where n is the number of data points and k is the number of fitted parameters (Eq. 2) [28].

Results

Inter-observer variability

Three observers manually delineated the tumor tissue on the T_2-weighted images of all tumors. Inter-observer variability was assessed by calculation of the similarity index (SI). The SI values between the different manual segmentations are listed in Table 1, and show good agreement among segmentation by the different observers.

Performance of *k*-means clustering

Average sensitivity and specificity for the *k*-means methods with the different feature vectors as a function of the

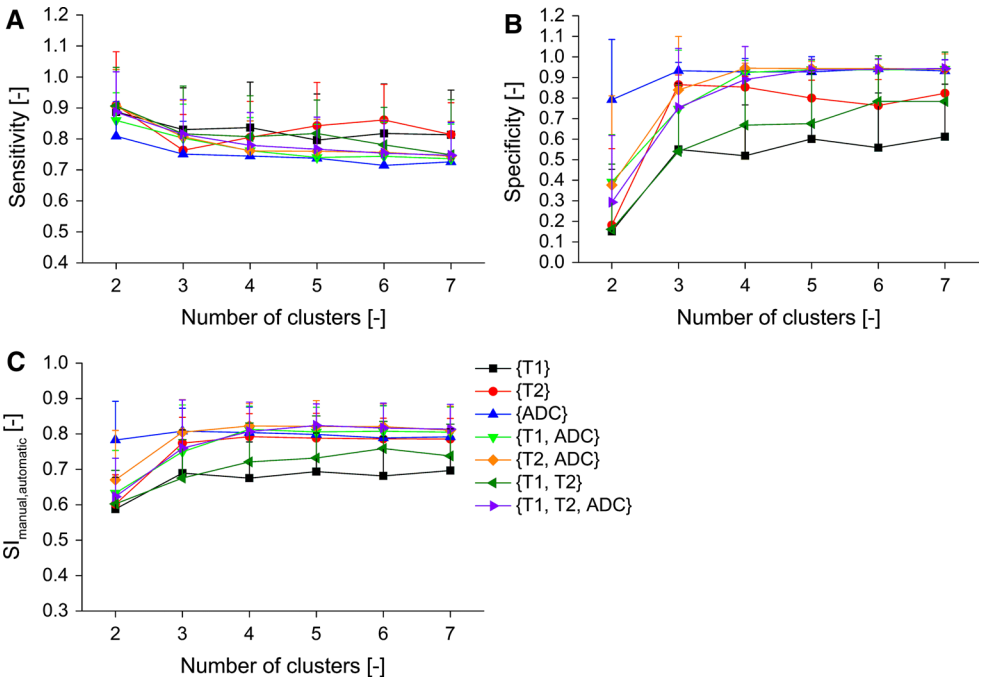

Fig. 2 Mean sensitivity (**a**), specificity (**b**), and $SI_{manual,automatic}$, i.e. similarity index between manual and automatic tumor segmentation (**c**) averaged over the observers for the different feature vectors (*bottom right*) as function of the number of clusters for the *k*-means clustering-based segmentation method. The *error bars* represent the standard deviation among the different animals ($n = 21$)

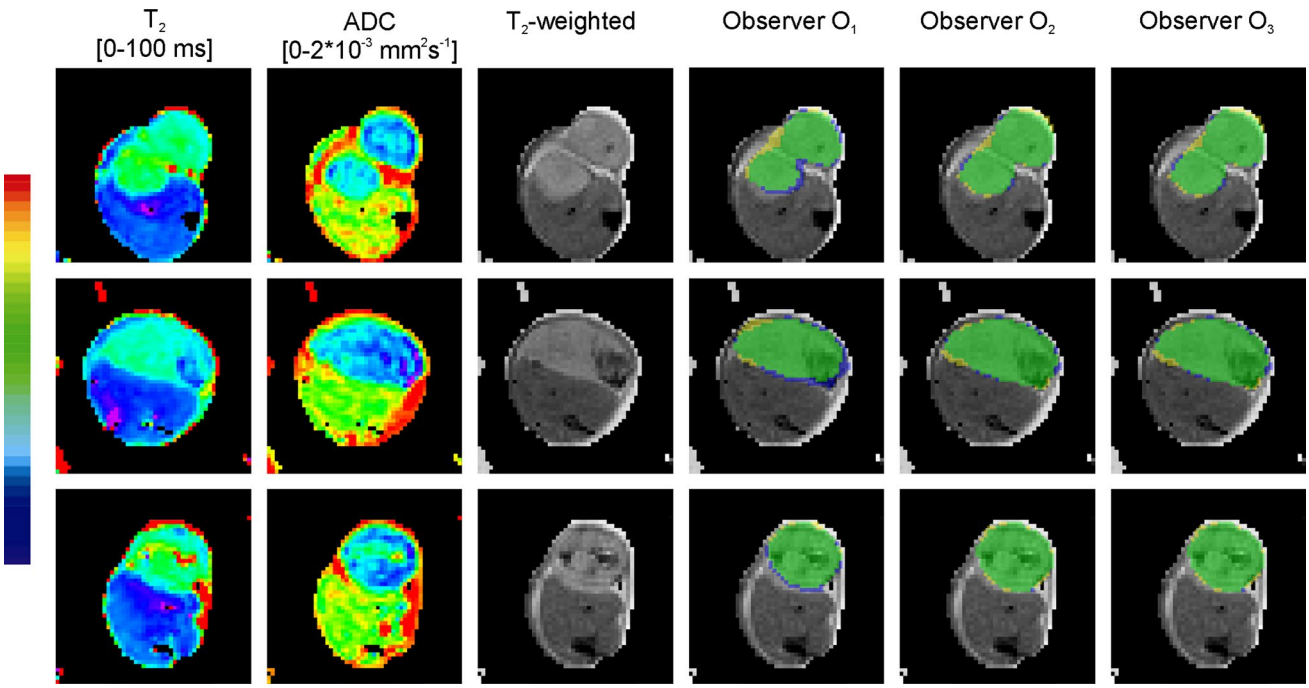

Fig. 3 Three representative examples of the results of *k*-means clustering with feature vector {T_2, ADC} and four clusters. The T_2 and ADC maps of the tumor-bearing limbs are shown in the *first* and *second columns*, respectively. The *third column* shows the corresponding T_2-weighted images. In the *fourth*, *fifth*, and *sixth columns* the automatic tumor segmentation (*blue*) and manual segmentation (*yellow*) for observers O_1, O_2, and O_3, respectively, are overlaid on the T_2-weighted image

number of clusters are displayed in Fig. 2a, b, respectively. An increase in specificity of the automatic segmentation method with increasing number of clusters was usually observed. Selection of the optimum *k*-means method for automatic tumor delineation was based on the $SI_{manual,automatic}$ values, which are displayed in Fig. 2c for the different methods. The highest $SI_{manual,automatic}$ value (0.82 ± 0.06) was observed for feature vector {T_2, ADC} with four clusters. For this *k*-means method the sensitivity (Fig. 2a) and specificity (Fig. 2b) were 0.76 ± 0.10 and 0.95 ± 0.02, respectively. The $SI_{manual,automatic}$ value for this method was statistically significantly larger than those for most of the other *k*-means methods. Only the *k*-means methods with feature vector {T_1, ADC} with four clusters, feature vector {T_1, T_2, ADC} with 4 and 5 clusters, and feature vector {T_2, ADC} with 3, 5, and 6 clusters did not have a statistically significantly lower $SI_{manual,automatic}$ than the optimized *k*-means method with feature vector {T_2, ADC} and four clusters. The $SI_{manual,automatic}$ value for the automatic segmentation using *k*–means with feature vector {T_2, ADC} and four clusters was higher than 0.7 (indicative of good agreement; as discussed in the "Materials and methods" section) for 20 out of 21 tumors.

$SI_{manual,automatic}$ values between manual segmentation by each observer and automatic segmentation with *k*-means clustering with feature vector {T_2, ADC} and four clusters were 0.85 ± 0.05, 0.82 ± 0.08, and 0.80 ± 0.07 for observers O_1, O_2, and O_3, respectively. These high $SI_{manual,automatic}$ values are indicative of good agreement between automatic segmentation and individual manual segmentation. Statistical analysis showed there was a significant effect of observer on the $SI_{manual,automatic}$ values ($P < 0.001$). The $SI_{manual,automatic}$ values for observer O_1 were statistically significantly higher than those for observers O_2 and O_3. Furthermore, $SI_{manual,automatic}$ values obtained for observer O_2 were statistically significantly higher than those for observer O_3.

Figure 3 shows three representative examples of automatic tumor segmentation with *k*-means clustering with feature vector {T_2, ADC} and four clusters. On the right-hand side of the figure, the automatic tumor segmentation and the manual tumor segmentation for each observer are overlaid on the T_2-weighted images. Close agreement between the manual and automated tumor segmentation was obtained for all observers. The algorithm accurately segmented the tumor tissue, even for tumors with a heterogeneous appearance on the T_2-weighted images, for example the tumors in the 2nd and 3rd row of Fig. 3. For these heterogeneous tumors, the (most likely necrotic) regions within the tumor with a low signal intensity on the T_2-weighted images were usually not assigned to one of the tumor clusters. However, they were included

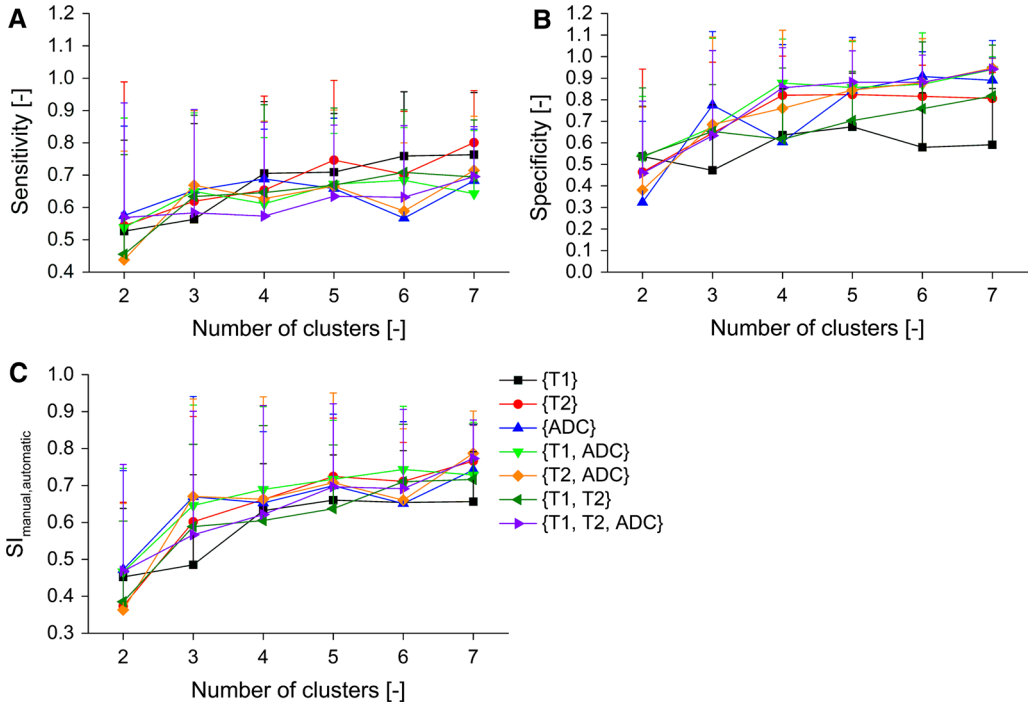

Fig. 4 Mean sensitivity (**a**), specificity (**b**), and $SI_{manual,automatic}$, i.e. the similarity index between the manual and automatic tumor segmentation (**c**), averaged over the observers for the different feature vectors (*bottom right*) as function of the number of clusters for the fuzzy *c*-means clustering-based segmentation method. The *error bars* represent the standard deviation among the different animals ($n = 21$)

in the final segmented tumor volume by the region growing algorithm that was applied in the last step of the algorithm (Fig. 1; discussed in the "Materials and methods" section).

Performance of fuzzy c-means clustering

Average sensitivity, specificity, and $SI_{manual,automatic}$ values for the different fuzzy *c*-means methods are shown in Fig. 4a–c, respectively. All values were generally lower than those obtained for *k*-means clustering (Fig. 2). For fuzzy *c*-means clustering, the highest $SI_{manual,automatic}$ was noted for feature vector $\{T_2, ADC\}$ and seven clusters. For this fuzzy *c*-means method, the $SI_{manual,automatic}$ was 0.79 ± 0.11, which tended to be lower ($P = 0.053$) than the $SI_{manual,automatic}$ for the optimum *k*-means method.

Performance of threshold-based segmentation method

For the threshold-based segmentation method, which only made use of T_2-weighted images, a mean $SI_{manual,automatic}$ of 0.69 ± 0.14 was obtained, which is significantly lower than the $SI_{manual,automatic}$ values for both the optimized *k*-means ($P < 0.001$) and fuzzy *c*-means ($P < 0.01$) clustering methods.

Comparison of tumor volumes from automatic and manual segmentation

Figure 5a shows the tumor volumes resulting from *k*-means clustering with feature vector $\{T_2, ADC\}$ and four clusters versus the average tumor volumes derived from manual tumor segmentation by the three observers. A high linear correlation ($R^2 = 0.99$) was observed between manually and automatically segmented tumor volumes. However, the slope of the linear fit was 0.88, indicative of either underestimation of the tumor volume in the automatic segmentation or overestimation of the tumor volume in the manual segmentation. Figure 5b shows a Bland–Altman plot of the difference between the tumor volume in manual and automatic segmentation resulting from the optimized segmentation method versus the mean of the manually and automatically segmented tumor volumes. This plot further illustrates that the automatically segmented tumor volumes were consistently smaller than the manually segmented tumor volumes.

Comparison of tumor volumes from automatic segmentation and histology

Figure 6 shows a correlation plot between histology-derived tumor volumes and tumor volumes derived from

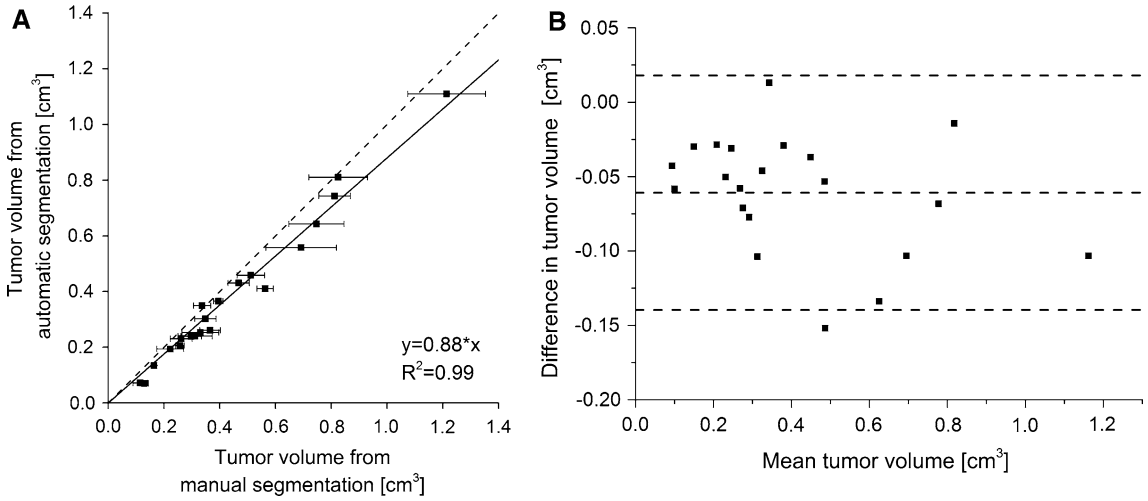

Fig. 5 **a** Tumor volumes derived from automatic tumor segmentation with *k*-means clustering with feature vector {T_2, ADC} and four clusters versus average tumor volumes derived from manual tumor segmentation. The *error bars* represent the standard deviation for the manually segmented tumor volumes of the different observers ($n = 3$). The fit to the data points is plotted as the *solid line*. The corresponding fit data are shown *bottom right*. The line of identity (*dashed line*) is added as visual reference. **b** Bland–Altman plot of the differences between tumor volume for the manually and automatically segmented tumors (i.e. automatically segmented tumor volume minus the observer-averaged manually segmented tumor volume) versus the mean of the manually and automatically segmented tumor volumes. The *dashed lines* represent the mean and 95 % confidence interval for the differences between tumor volumes

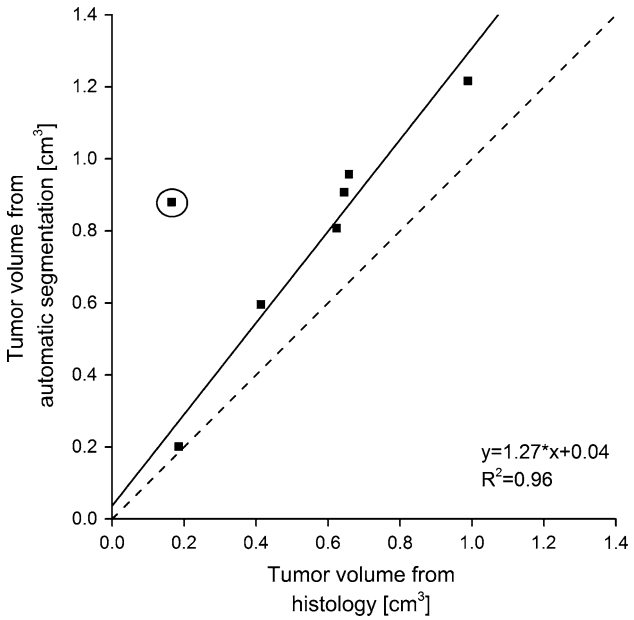

Fig. 6 Tumor volumes resulting from automatic segmentation with *k*-means clustering with feature vector {T_2, ADC} and four clusters versus tumor volumes derived from whole-tumor based histology. The fit to the data points is plotted as the *solid line*. The corresponding fit data are shown *bottom right*. The line of identity (*dashed line*) is added as visual reference. One data point was identified as an outlier and is indicated with a *black circle*

automatic segmentation of seven separate data sets. A strong linear correlation ($R^2 = 0.96$) was observed after omitting a data point that was identified as an outlier. For all tumors, the automatically segmented tumor volume was larger than the tumor volume derived from histology.

Discussion

An automatic segmentation method for delineation of tumor tissue based on endogenous MR contrast has been described. For optimization of the segmentation method, automatic tumor delineations were quantitatively compared with manual segmentation performed by three observers. The largest overlap between automatic and manual segmentation was observed for *k*-means clustering with feature vector {T_2, ADC} and four clusters. Visual inspection of the four clusters originating from this *k*-means method showed that three clusters corresponded to tumor tissue, peritumoral edema, and muscle tissue. The fourth cluster usually originated from small regions of a few pixels in muscle or tumor tissue. The bone tissue in the tumor-bearing limb was typically excluded from analysis, since parameter values were not calculated for bone pixels, because the signal intensity for these pixels was at noise level (as discussed in the "Materials and methods" section). The $SI_{manual,automatic}$ of *k*-means clustering with {T_2, ADC} and four clusters was not statistically significantly better than the *k*-means method with feature vectors {T_1, ADC} (four clusters) and {T_1, T_2, ADC} (4 and 5 clusters), which indicated that clustering based on ADC and either T_1 or T_2 worked equally well for segmentation of these tumors.

Fuzzy *c*-means clustering did not improve the tumor segmentation compared with *k*-means clustering, which implies that the contrast between the clusters of the different tissue types was well-defined. In clinical applications of automated tumor segmentation, fuzzy *c*-means clustering is widely used for different types of tumor [3–5, 9, 10]. For human tumors, the tissue around the tumor is usually very different from the surrounding tissue (muscle, oedema, bone) in the subcutaneous tumor model here. The contrast between tumor and surrounding tissue could be less well-defined, which may explain the additive value of fuzzy *c*-means clustering in the clinic. Furthermore, the resolution of clinical MR images is usually lower than that in our study ($0.31 \times 0.31 \times 1$ mm^3), leading to larger partial volume effects at the tumor borders and a less well-defined demarcation between tumor and surrounding tissue. In such cases, a soft segmentation method, for example fuzzy *c*-means, could prove beneficial.

A strong linear correlation was observed between tumor volumes from the optimized automated segmentation and manual segmentation (Fig. 6). However, the automatically segmented tumor volumes were, on average, smaller than those from the manual segmentation, which corresponds to the finding that the sensitivity of the optimized segmentation method was lower (0.76) than its specificity (0.95). Although the partial volume effects were relatively small, as stated above, they could partly explain the smaller tumor volumes for the automatic segmentation. Visual inspection of the segmentation revealed that pixels at the tumor rim were usually included in manual segmentation, whereas they were not always included in the automatic segmentation because of partial volume effects with adjacent muscle or edematous tissue. These partial volume effects should be carefully taken into account when the segmentation method is used for image-guided treatment planning. It would be advisable to include a safety margin around the segmented tumor volume to prevent undertreatment. Safety margins are, usually, already included in image-guided therapy, for example during HIFU procedures [29], to take into account the possible presence of occult tumor cells outside the tumor mass that are not visible with the imaging method used. The deviation between the automatically and manually segmented tumor volumes was also partly caused by errors in the manual tumor segmentation. Manual tumor delineation was based on T_2-weighted images only. The border between tumor tissue and peri-tumoral edema was sometimes hardly visible, which led to incorrect inclusion of edematous tissue in manual segmentation of several tumors, and emphasizes the need for automatic segmentation methods. Possible susceptibility artefacts that arise from EPI that was used as readout in the mapping acquisitions most likely did not cause the observed mismatch between the automatically and manually segmented tumor volumes, because both the manual and automatic segmentation were performed on the basis of EPI images. No apparent geometrical differences between the EPI images of the T_1, T_2, and ADC mapping procedures were observed. In addition, susceptibility artefacts at air–tissue interfaces were small, anyway, because of application of ultrasound gel to the tumor-bearing limb (as discussed in the "Materials and methods" section).

The inter-observer variability was assessed by calculation of the SI between the different manual segmentations (Table 1). Strong agreement (average SI of 0.86 ± 0.02) was observed between the tumor delineations of the different observers. Visual inspection of the manual segmentation showed that differences between the manual tumor delineations were mainly caused by the aforementioned inclusion of peri-tumoral edema. In a future study, the accuracy of manual tumor delineation could be improved by inclusion of diffusion-weighted images as a visual reference. The contrast between tumor and edematous tissue is usually better on diffusion-weighted images than on T_2-weighted images, because of the large difference between apparent diffusion coefficient of water in tumor tissue and edema [30]. In addition, manual segmentation could also be performed on contrast-enhanced T_1-weighted images. These tumor delineations could then be compared with automatic segmentation by the proposed intrinsic MR contrast-based method to assess whether our method could rival tumor delineation based on contrast-enhanced MRI, which, in the clinic, is regarded as the best method. This study can be regarded as a feasibility study in which we determined whether automatic tumor segmentation based on intrinsic MR contrast can be performed for a subcutaneous tumor model. Because T_2-weighted images, rather than contrast-enhanced T_1-weighted images, are usually used for manual segmentation of these subcutaneous tumors (e.g. Refs. [31–33]), manual tumor delineation from T_2-weighted images were considered as the best method in this study.

Selection of the clusters associated with the tumor tissue was performed on the basis of a threshold-based segmentation method (Fig. 1 and discussed in the "Materials and methods" section). $SI_{manual,automatic}$ was statistically significantly lower for threshold-based segmentation than for clustering-based segmentation. Although a clear peak that corresponded to tumor tissue was observed in the averaged histogram of signal intensities in the T_2-weighted images (Fig. 1), this peak considerably overlapped with other peaks in the histogram. Consequently, next to the tumor tissue, regions of other types of tissue were also incorrectly assigned as tumor tissue by the threshold-based segmentation. In addition, for individual animals, the tumor peak in the histogram could be slightly shifted compared with the histogram averaged for all animals, resulting in incomplete segmentation of the tumor by the threshold-based

segmentation. Nevertheless, threshold-based segmentation provided a suitable initial tumor segmentation from which the tumor clusters could be selected in the clustering-based segmentation. However, care must be taken when defining the thresholds for the T_2-weighted signal intensity values. To address this, we evaluated the effect of varying the thresholds on the $SI_{manual,automatic}$ values of the optimized segmentation method. Application of a smaller range of normalized intensity values (0.5–0.6) as threshold did not affect the $SI_{manual,automatic}$ values, whereas a larger range (0.4–0.7) did lead to a lower performance of the method ($SI_{manual,automatic} = 0.78 \pm 0.11$ compared with $SI_{manual,automatic} = 0.82 \pm 0.06$ for the threshold range used in this study (0.45–0.65)). This finding suggests it is better to choose the threshold intensity range conservatively. A range which is too broad leads to larger overlap of the tumor peak with peaks of other tissue and consequently to substantial misclassification of surrounding non-tumor tissue as tumor tissue.

The threshold-based segmentation may, however, not be applicable to some (clinical) tumor types for which the contrast between tumor and surrounding tissue is poorly defined on T_2-weighted images. For those tumor types, the threshold-based segmentation method may be applied to T_1-weighted or diffusion-weighted images, if the contrast between tumor and other tissue is more visible on those images. While the threshold-based method is likely to require re-definition of the set thresholds for each tumor type, automatic selection of clusters could possibly also be performed on the basis of features more specific to tumors in general, for example shape, size [18], and tissue homogeneity [9]. Recently, Linguraru et al. [34] showed that liver tumors can be accurately detected on contrast-enhanced computed tomography images by using a set of features that describe the intensity, shape, size, and homogeneity of identified objects of interest in the liver. Apart from their potential utility for cluster selection, such tumor features could also be used to select the tumor volume from the results from morphological component analysis. In this study, the largest component was assumed to be the tumor volume, but this procedure would not be applicable to data sets in which more than one tumor is present. In such cases, the tumor volumes could be automatically selected on the basis of the above-mentioned tumor features.

Furthermore, clustering-based segmentation may also be performed without input of prior knowledge on tumor tissue characteristics. The expert observer or radiologist could then manually select the cluster(s) associated with tumor tissue and subsequently the tumor volume(s). Although this would imply some user interaction, manual selection of the tumor clusters may still yield a more accurate and faster segmentation than tumor segmentation based purely on laborious manual delineations.

Manual segmentation was regarded as the best method for tumor demarcation in this study. In other reports on automatic segmentation of tumors, manual segmentation is also often used as reference for evaluation of the performance of the segmentation method presented [3, 6, 9, 10, 12]. However, the above-discussed inter-observer variability might suggest that manual segmentation is not the best reference for optimization of automatic segmentation methods. Ideal validation of the proposed segmentation method would consist of spatial correlation analysis between the segmented tumors and histological tumor sections. However, spatial registration between MRI and histology is known to be very challenging and would have required an intermediate MRI scan of the excised tumors and a robust anatomical reference [35]. Instead of assessment of the spatial correlation between MRI and histology, in this study whole-tumor histology was used to calculate the tumor volume of a small set of tumors. A strong correlation was observed between automatically segmented tumor volumes and histology-derived tumor volumes. However, the tumor volumes from histology were consistently smaller than the automatically segmented tumor volumes. This finding can most probably be explained by tissue shrinkage during preparation of the tumors for cryosectioning. In addition, histological processing can induce substantial tissue deformation, which may explain the outlier observed in Fig. 6. Visual inspection of the automatically segmented tumor volume corresponding to the outlier showed that the algorithm had accurately identified the tumor tissue. Tumor weight could provide an alternative measure of tumor volume. Strong correlations between tumor mass and MRI-derived measurements of tumor burden in mouse lung [36] and subcutaneous [19] tumor models have been described elsewhere.

Clinical application of the automatic segmentation method seems feasible, because the proposed MRI procedures are already clinically available. However, as indicated above, different properties of the surrounding tissue and the typically lower spatial resolution in clinical MR imaging could affect the accuracy of the proposed algorithm. Furthermore, the typical lower field strength of clinical scanners could affect the observed contrast between tumor and surrounding tissue, because most MR parameters, for example T_1 and T_2, depend on the magnetic field strength. Because of these potential challenges regarding application of the proposed method from the subcutaneous mouse tumor model to human tumors, the algorithm must be tested for different tumors to gain insight in its clinical applicability. Alteration of the algorithm to enable segmentation of a broader range of tumor types may include use of tumor features other than the relative signal intensity in the T_2-weighted images for automatic selection of the clusters associated with the tumor tissue, as indicated above.

Apart from having potential clinical utility, the proposed method may also be useful for a variety of preclinical applications. Analysis of preclinical MRI data for different tumor models is usually performed on the basis of manually drawn ROIs. An observer-independent tumor volume measure could increase reproducibility in these preclinical studies. Application of the algorithm to several orthotopic animal tumor models is necessary to assess the suitability of the method in preclinical MR cancer research in general, because segmentation of these orthotopic tumors may, similarly to human tumors, require adaptation of the algorithm.

For practical application of the proposed method to a larger variety of tumor types, the implemented algorithms should be integrated into software with a graphical user interface (GUI) that facilitates user-friendly MRI data processing and subsequent automatic tumor segmentation. This GUI should enable user interaction, for example addition of information on tumor characteristics and manual adjustment of automatic segmentation. In addition, previous knowledge derived from automatic segmentation of different tumor types could be stored by the software to improve the accuracy of the automatic segmentation of subsequent tumors.

Conclusion

We have shown that time-efficient, automatic segmentation of tumor tissue growing subcutaneously in the mouse hindleg can be achieved on the basis of endogenous MR contrast only, without the need of injection of a contrast agent. We believe this automatic segmentation method will be beneficial for various clinical and preclinical applications.

Acknowledgments This research was performed within the framework of CTMM, the Center for Translational Molecular Medicine (www.ctmm.nl), project VOLTA (grant 05T-201).

Conflict of interest The authors have no conflict of interest.

Ethical standards All animal experiments were performed according to the Directive 2010/63/EU of the European Parliament and approved by the Animal Care and Use Committee of Maastricht University.

References

1. Partridge SC, Gibbs JE, Lu Y et al (2005) MRI measurements of breast tumor volume predict response to neoadjuvant chemotherapy and recurrence-free survival. AJR Am J Roentgenol 184:1774–1781
2. Alderliesten T, Schlief A, Peterse J et al (2007) Validation of semiautomatic measurement of the extent of breast tumors using contrast-enhanced magnetic resonance imaging. Invest Radiol 42:42–49
3. Chen W, Giger ML, Bick U (2006) A fuzzy c-means (FCM)-based approach for computerized segmentation of breast lesions in dynamic contrast-enhanced MR images. Acad Radiol 13:63–72
4. Kannan SR, Ramathilagam S, Devi P et al (2012) Improved fuzzy clustering algorithms in segmentation of DC-enhanced breast MRI. J Med Syst 36:321–333
5. Pang Y, Li L, Hu W et al (2012) Computerized segmentation and characterization of breast lesions in dynamic contrast-enhanced MR images using fuzzy c-means clustering and snake algorithm. Comput Math Methods Med 2012:634907
6. Vignati A, Giannini V, De Luca M et al (2011) Performance of a fully automatic lesion detection system for breast DCE-MRI. J Magn Reson Imag 34:1341–1351
7. Vos PC, Barentsz JO, Karssemeijer N et al (2012) Automatic computer-aided detection of prostate cancer based on multiparametric magnetic resonance image analysis. Phys Med Biol 57:1527–1542
8. Ozer S, Langer DL, Liu X et al (2010) Supervised and unsupervised methods for prostate cancer segmentation with multispectral MRI. Med Phys 37:1873–1883
9. Harati V, Khayati R, Farzan A (2011) Fully automated tumor segmentation based on improved fuzzy connectedness algorithm in brain MR images. Comput Biol Med 41:483–492
10. Liu J, Udupa JK, Odhner D et al (2005) A system for brain tumor volume estimation via MR imaging and fuzzy connectedness. Comput Med Imag Graph 29:21–34
11. Fathi Kazerooni A, Mohseni M, Rezaei S et al (2014) Multiparametric (ADC/PWI/T2-w) image fusion approach for accurate semi-automatic segmentation of tumorous regions in glioblastoma multiforme. Magn Reson Mater Phy. doi:10.1007/s10334-014-0442-7
12. Lodder WL, Gilhuijs KG, Lange CA et al (2013) Semi-automated primary tumor volume measurements by dynamic contrast-enhanced MRI in patients with head and neck cancer. Head Neck 35:521–526
13. Hijnen N, Elevelt A, Grüll H (2012) Effects of MRI Contrast Agents during HIFU ablation therapy. In: Proceedings of the 3rd MRgFUS symposium, Washington DC, USA ID P-116-EA
14. Hijnen NM, Elevelt A, Grüll H (2013) Stability and trapping of magnetic resonance imaging contrast agents during high-intensity focused ultrasound ablation therapy. Invest Radiol 48:517–524
15. MacNeil S, Bains S, Johnson C et al (2011) Gadolinium contrast agent associated stimulation of human fibroblast collagen production. Invest Radiol 46:711–717
16. Kuo PH, Kanal E, Abu-Alfa AK et al (2007) Gadolinium-based MR contrast agents and nephrogenic systemic fibrosis. Radiology 242:647–649
17. Amet S, Deray G (2012) Renal toxicity of contrast agents in oncologic patients. Bull Cancer 99:295–307
18. Hsieh TM, Liu YM, Liao CC et al (2011) Automatic segmentation of meningioma from non-contrasted brain MRI integrating fuzzy clustering and region growing. BMC Med Inform Decis Mak 11:54
19. Montelius M, Ljungberg M, Horn M et al (2012) Tumour size measurement in a mouse model using high resolution MRI. BMC Med Imaging 12:12
20. Herneth AM, Guccione S, Bednarski M (2003) Apparent diffusion coefficient: a quantitative parameter for in vivo tumor characterization. Eur J Radiol 45:208–213
21. Lim HK, Kim JK, Kim KA et al (2009) Prostate cancer: apparent diffusion coefficient map with T2-weighted images for detection—a multireader study. Radiology 250:145–151
22. Levitt M, Freeman R (1981) Compensation for pulse imperfections in NMR spin-echo experiments. J Magn Reson 43:56–80

23. Karlsson M, Nordell B (1999) Phantom and in vivo study of the Look-Locher T1 mapping method. Magn Reson Imag 17:1481–1488
24. Omran MGH, Engelbrecht AP, Salman A (2007) An overview of clustering methods. Intell Data Anal 11:583–605
25. Zijdenbos AP, Dawant BM, Margolin RA et al (1994) Morphometric analysis of white matter lesions in MR images: method and validation. IEEE Trans Med Imag 13:716–724
26. Khayati R, Vafadust M, Towhidkhah F et al (2008) Fully automatic segmentation of multiple sclerosis lesions in brain MR FLAIR images using adaptive mixtures method and Markov random field model. Comput Biol Med 38:379–390
27. Hectors SJ, Jacobs I, Strijkers GJ et al (2014) Multiparametric MRI analysis for the identification of high intensity focused ultrasound-treated tumor tissue. PLoS ONE 9:e99936
28. Belsley DA, Kuh E, Welsch RE (2005) Regression diagnostics: identifying influential data and sources of collinearity. Wiley, New York
29. Schmitz AC, Gianfelice D, Daniel BL et al (2008) Image-guided focused ultrasound ablation of breast cancer: current status, challenges, and future directions. Eur Radiol 18:1431–1441
30. Spuentrup E, Buecker A, Adam G et al (2001) Diffusion-weighted MR imaging for differentiation of benign fracture edema and tumor infiltration of the vertebral body. Am J Roentgenol 176:351–358
31. Zhang CC, Yan Z, Giddabasappa A et al (2014) Comparison of dynamic contrast-enhanced MR, ultrasound and optical imaging modalities to evaluate the antiangiogenic effect of PF-03084014 and sunitinib. Cancer Med 3:462–471
32. Zhen Z, Tang W, Chuang YJ et al (2014) Tumor vasculature targeted photodynamic therapy for enhanced delivery of nanoparticles. ACS Nano 8:6004–6013
33. Zhang L, Zhou H, Belzile O et al (2014) Phosphatidylserine-targeted bimodal liposomal nanoparticles for in vivo imaging of breast cancer in mice. J Controll Release 183:114–123
34. Linguraru MG, Richbourg WJ, Liu J et al (2012) Tumor burden analysis on computed tomography by automated liver and tumor segmentation. IEEE Trans Med Imag 31:1965–1976
35. Alic L, Haeck JC, Bol K et al (2011) Facilitating tumor functional assessment by spatially relating 3D tumor histology and in vivo MRI: image registration approach. PLoS ONE 6:e22835
36. Tidwell VK, Garbow JR, Krupnick AS et al (2012) Quantitative analysis of tumor burden in mouse lung via MRI. Magn Reson Med 67:572–579