A Novel and Fast Method for Cluster Analysis of DCE-MR Image Series of Breast Tumors

Mojgan Mohajer*[a], Gunnar Brix[b], Karl-Hans Englmeier[a]
[a]Institute for Biological and Medical Imaging (IBMI), Helmholtz Zentrum Muenchen, Ingolstaedter-Landstr. 1, D-85764 Oberschleissheim, Germany www.helmholtz-muenchen.de
[b]Federal Office for Radiation Protection (BfS), Department of Medical Radiation Hygiene and Dosimetry, Ingolstaedter-Landstr. 1, D-85764 Oberschleissheim, Germany

ABSTRACT

A novel approach is introduced for clustering tumor regions with similar signal-time series measured by dynamic contrast-enhanced (DCE) MRI to segment the tumor area in breast cancer. Each voxel of the DCE-MRI dataset is characterized by a signal-time curve. The clustering process uses two describer values for each pixel. The first value is L_2-norm of each time series. The second value r is calculated as sum of differences between each pair of $S(n-i)$ and $S(i)$ for $i = \{0 \ldots n/2\}$ where S is the intensity and n the number of values in a time series. We call r reverse value of a time series. Each time series is considered as a vector in an n-dimensional space and the L_2-norm and reverse value of a vector are used as similarity measures. The curves with similar L_2-norms and similar reverse values are clustered together. The method is tested on breast cancer DCE-MRI datasets with N = 256 x 256 spatial resolution and n = 128 temporal resolution. The quality of each cluster is described through the variance of Euclidean distances of the vectors to the mean vector of the corresponding cluster. The combination of both similarity measures improves the segmentation compared to using each measure alone.

Keywords: DCE-MRI, Time Series, Segmentation, Clustering, L_p-norm

1. INTRODUCTION

In the last decade dynamic contrast-enhanced (DCE) MRI has become an important functional imaging method for tumor detection and characterization that complements established morphological imaging approaches[1]. Temporal change of the signal intensity measured after the administration of a paramagnetic contrast agent (CA) reflects the underlying change in the CA concentration and thus provides information on tissue microcirculation and microvasculature. In breast cancer, for example, this information can be used to improve tumor diagnosis and therapy management. The purpose of this work is to distinguish tumor areas that are characterized by a similar enhancing pattern. Each voxel in the DCE MRI dataset is related to a signal curve (SC) which demonstrates the enhancing pattern for this voxel. The clusters of similar SCs can be used for further pharmacokinetic analysis for a more accurate characterization of tumor microcirculation. Actually a-priori is not known how many different patterns are in a tumor area or other surrounding tissue. The other undefined topic is how to define the similarity or dissimilarity between SCs. The conventional clustering methods, especially those which are based on Euclidian distance as dissimilarity measure, are not satisfactory. Therefore, there are a variety of different clustering methods which are used to cluster DCE-MRI signal curves. Here is a summery of these methods: neural network is implemented for classification of carcinoma tumors, which is based on training data with three defined classes[2]. As it is usual for training-based clustering there is a need for a possible large number of training data and a well-selected expert's knowledge. Two groups[3, 4] use pre-defined characteristic of SCs as similarity measures which are basically derived from the classification of initial enhancement and the late pattern. The schema of classification using pre- and post-pattern has been widely adopted following its initial clinical evaluation Kuhl et al[5]. In both methods the classification of extracted features is done according to the

*Mojgan.mohajer@helmholtz-muenchen.de; phone +49-89-31874454; fax +49-89-31873017; www.helmholtz-muenchen.de

Medical Imaging 2010: Biomedical Applications in Molecular, Structural, and Functional Imaging, edited by Robert C. Molthen, John B. Weaver, Proc. of SPIE Vol. 7626, 76260R · © 2010 SPIE
CCC code: 1605-7422/10/$18 · doi: 10.1117/12.843967

fixed thresholds. Both methods are successful to segment and classify the tumors but they are too rough to distinguish between different areas inside a tumor. Other methods are vector quantization[6] and minimization of energy function[7]. The first one has too many parameters and gets more complex for high dimensional vector spaces, while the latter one suffers from general optimization problems and is again too steady and complex for high dimensional vector spaces. In our method we describe the SC with two features, which are defined by using all intensity values and temporal information in the SC. The final clustering proceeds in the two-dimensional feature space. The introduced method is fast, simple and robust in comparison to previous methods.

2. MATERIALS AND METHOD

2.1 Patients

The algorithm was tested on the basis of DCE-MRI datasets acquired in female patients with breast cancer (ages 42-87 years, mean age 61.11 years). Breast lesions were proved by histology following breast biopsy/surgery.

2.2 MR Imaging

Data acquisition was performed with a 1.5 T MR system. The patients lay prone on the breast coil with the arms extended above the head. On the basis of a static images with the following parameters: T_1-weighted FLASH image, TR = 12 ms, TE = 5 ms, flip angle $\alpha = 35°$, slice thickness TH = 4 mm, matrix size = 256 × 256, field of view (FOV) = 320 mm; two transaxial section were defined (FOV = 320 mm, TH = 6 mm) crossing the lesion and the aorta for further evaluation. Strongly T_1-weighted images (TR = 10 ms, TE = 4.1 ms, $\alpha = 12°$, raw matrix size = 256 × 128, matrix size = 256 × 256, T_{rec} = 125 ms) were acquired from the same cross *s* before, during, and after intravenous administration of 0.1 mmol Gd-DTPA (MAGNEVIST; Schering AG, Berlin, Germany) per kg body weight at a constant rate over 30 s with an infusion pump (CAI 626P/Tomojet; Doltron AG, Uster, Switzerland). A total 128 acquisitions were performed over a period of 6.9 min, with one acquisition each 3.25 s.

2.3 Feature description

In DCE-MRI an image series with *n* time frames is acquired before, during and after administration of a CA resulting in a signal curve (SC) with *n* data points for each voxel. These curves can be considered as vectors in an *n*-dimensional vector space. The SCs are described through two features, L_p-norm and reverse value. These two values make the basis for the segmentation and clustering of SCs.

The L_p-norm $|x|_p$ for $p = 1,2,...$ of a vector $\mathbf{x} = (x_1, x_2, ..., x_n)$ is defined as[8]:

$$|x|_p = \left(\sum_i |x_i|^p \right)^{\frac{1}{p}} \tag{1}$$

The parameter *p* is set to two for our method. The reverse value r for a vector $\mathbf{x} = (x_1, x_2, ..., x_n)$ is defined as:

$$\mathrm{r} = \sum_{i=a}^{k} x_{n-k} - x_k \tag{2}$$

Where *a* and *n* indicate the start and end indices of the signal curve which are considered in calculation of r-value. The middle point of *a* and *n* is the value of *k*. Figures-1 is the schematic illustration of reverse value for two SCs. The blue line demonstrates the reverse line and the shadowed area shows the reverse value. It must be taken into account that the colored area shows only the absolute value, but the value itself can be negative. The different area colors in the first SC (from left) in figure 1 demonstrate the partial negative and positive summation in the reverse value for this curve. The reverse value can be calculated for each curve as a partial reverse value according to the starting point *a* and end point *n*.

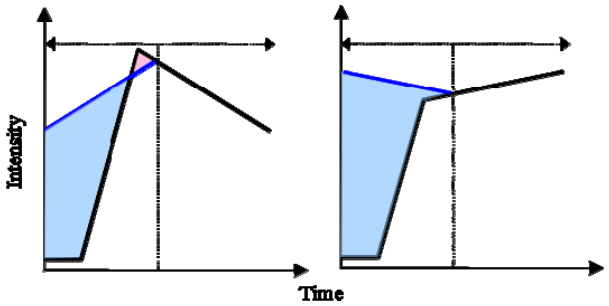

Figure 1. A schematic illustration of reverse value for two different curves is demonstrated. The blue line is the reverse line and the area in color is an illustration of absolute value for reverse value. The reverse value can partially or totally be negative. The different area colors depict the negative and positive parts.

2.4 Segmentation and clustering

The main idea of the algorithm is that the similarity among curves is determined by L2-norm and the reverse value. The interval of calculated l_2 and r values is divided into k and m sub-intervals of equal length, respectively. The curves with L2-norms and reverse values being in the same sub-intervals are selected as members of the same cluster. This approach results in $m \times n$ clusters. For a subsequent (tracer-kinetic) analysis of the curves, it is necessary to use relative $(S(i)-S_0)/S_0$ (with S_0 as the pre-contrast signal) rather than the absolute signal intensities. On the other hand clustering with L2-norm and reverse value separates the soft tissue and other parts of the image successfully from tumor area. To overcome this shortcoming, in a further step, the relative intensity values are used in a bottom up process on $(m \times n)$ clusters to join the similar clusters together. Two clusters are joined to a new cluster that the Euclidean distances between the mean vectors of these two clusters are minimum among all clusters. The variance of Euclidean distances between the vectors inside a cluster and the mean vector of the cluster serves as a measurement for the quality of clustering. A large value of the variance indicates more dissimilarity among the curves inside a cluster.

3. RESULT

In an experiment, the L2-norm and reverse value are studied as separate describers for clustering of intensity curves and results are compared with the introduced clustering method using both the L2-norm and reverse value together. Figure-2 (a-f) depicts results of clustering by dividing the L2-norm interval (the interval between minimum and maximum value of l_2 for that dataset) into equal sub-intervals for a patient. The number of intervals in a, b, c, d, e, and f is 1, 2, 4, 8, 32, and 64 respectively. As it can be seen, dividing L2-norm interval into sub-intervals reduces the distance of the curves inside a cluster to the mean curve of the cluster. However, by increasing the number of sub-intervals to more than 16, the distance to mean curve does not further decrease. Therefore, 16 sub-intervals are chosen for division of L2-norm interval. Figure-3 (g-j) depicts results of clustering by dividing the reverse value interval (the interval between minimum and maximum value of r for that dataset) into equal sub-intervals. The number of intervals in g, h, i, and j is 1, 2, 4, and 8 respectively. In this case division onto 8 sub-intervals is the optimum. Figure-3 (k-l) summarizes the result of the introduced cluster analysis by dividing L2-norm interval into 16 sub-intervals and reverse value interval into 8 sub-intervals. Figure-3 (k) shows the projection of the cluster on the plane of L2-norm and distance to mean curve of the cluster while figure-3 (l) demonstrates the same cluster points projected on the plane of reverse value and distance to mean curve of the cluster. Comparing the result presented in figure-3 (k-l) to the results of clusters presented in figure-3 (a-j), the distance to mean curve of the cluster is markedly reduced.

In figure-3 the results of clustering for two patients are depicted as gray level images. Each cluster is demonstrated in a different gray level. The patient 1 (left) has a large tumor and so more clusters, while in the case of patient 2 (right) there are fewer clusters.

Figure-4 and figue-5 show some selected clusters in a breast tumor obtained by introduced algorithm for patient 1 and 2 from figure-3. The pixels inside each cluster are marked in red color and the related curves of the pixels inside the cluster are drawn in the next row. The mean curve of each cluster is drawn in red. The segmentation of the tumor into regions

with a different enhancement pattern is clearly visible. In figure-5 can be observed that strongly noisy signal curves from other parts like heart can result in similar l_2 and r values and are selected with pixels from tumor area for a cluster. Therefore it is of interest to segment the tumor area manually or automatically before the clustering.

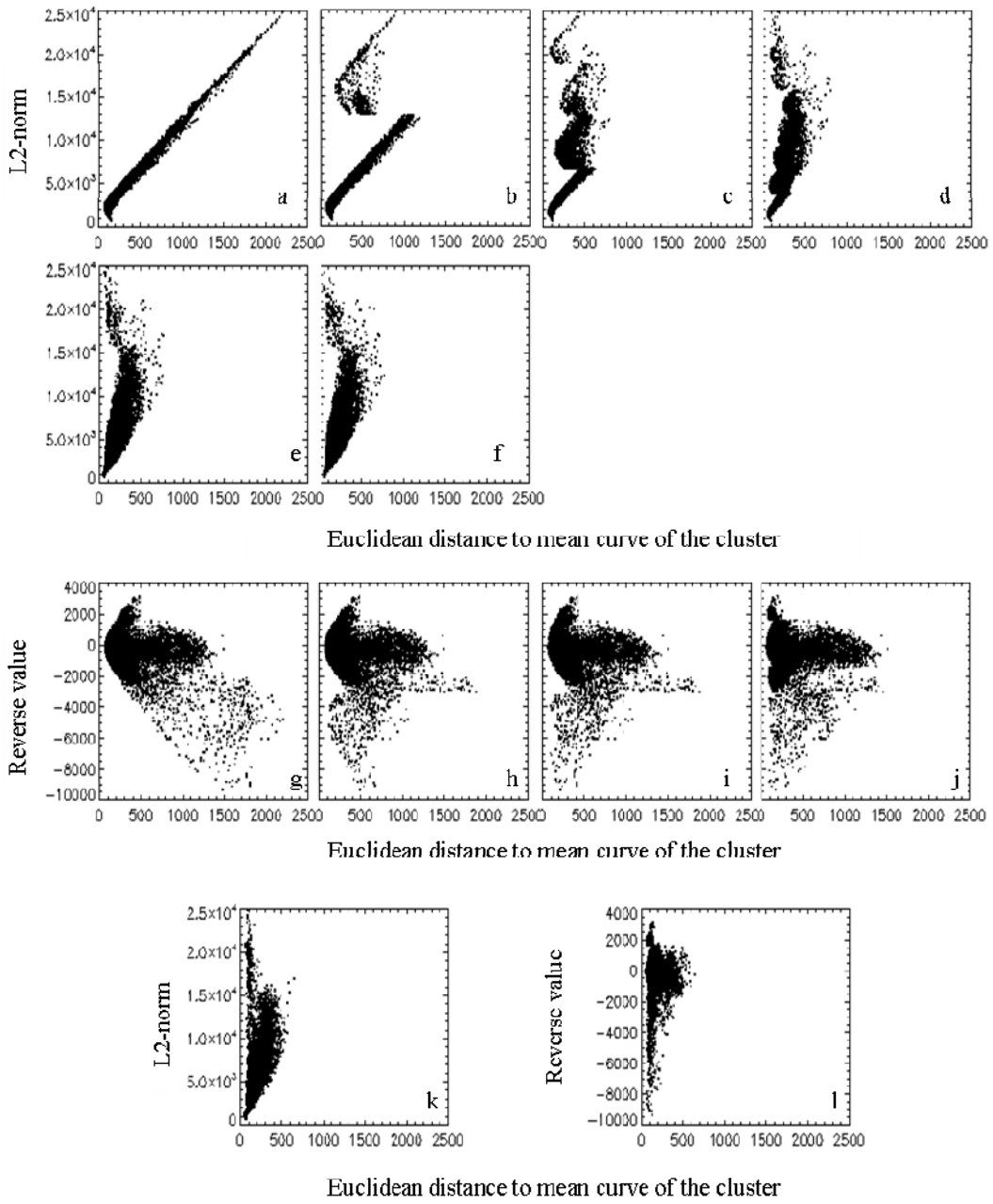

Figure 2. (a-f) The L2-norm interval is divided into equal sub-intervals to study how the number of proper sub-intervals influences the clustering process. The number of intervals in a, b, c, d, e, and f is 1, 2, 4, 8, 32, and 64 respectively. As can be seen, the use of more than 16 intervals does not change the quality of clusters. (g-j) Depict results of clustering by dividing the reverse value interval into equal sub-intervals. The number of intervals in g, h, i, and j is 1, 2, 4, and 8 respectively. (k-l) show the result of a cluster analysis by dividing L2-norm interval into 16 sub-intervals and reverse value interval into 8 sub-intervals, (k) shows the projection of the cluster points on the plane of L2-norm and distance to mean curve of the cluster while (l) demonstrates the same cluster points projected on the plane of reverse value and distance to mean curve of the cluster.

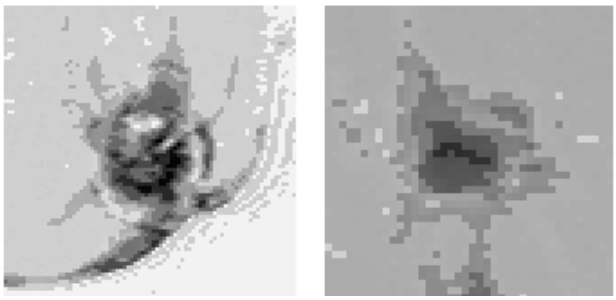

Figure 3. This figure shows the result of clustering for two patients. Each cluster is demonstrated in a different gray level. The patient 1 (left) has a large tumor and so more clusters, while in the case of patient 2 (right) there are fewer clusters.

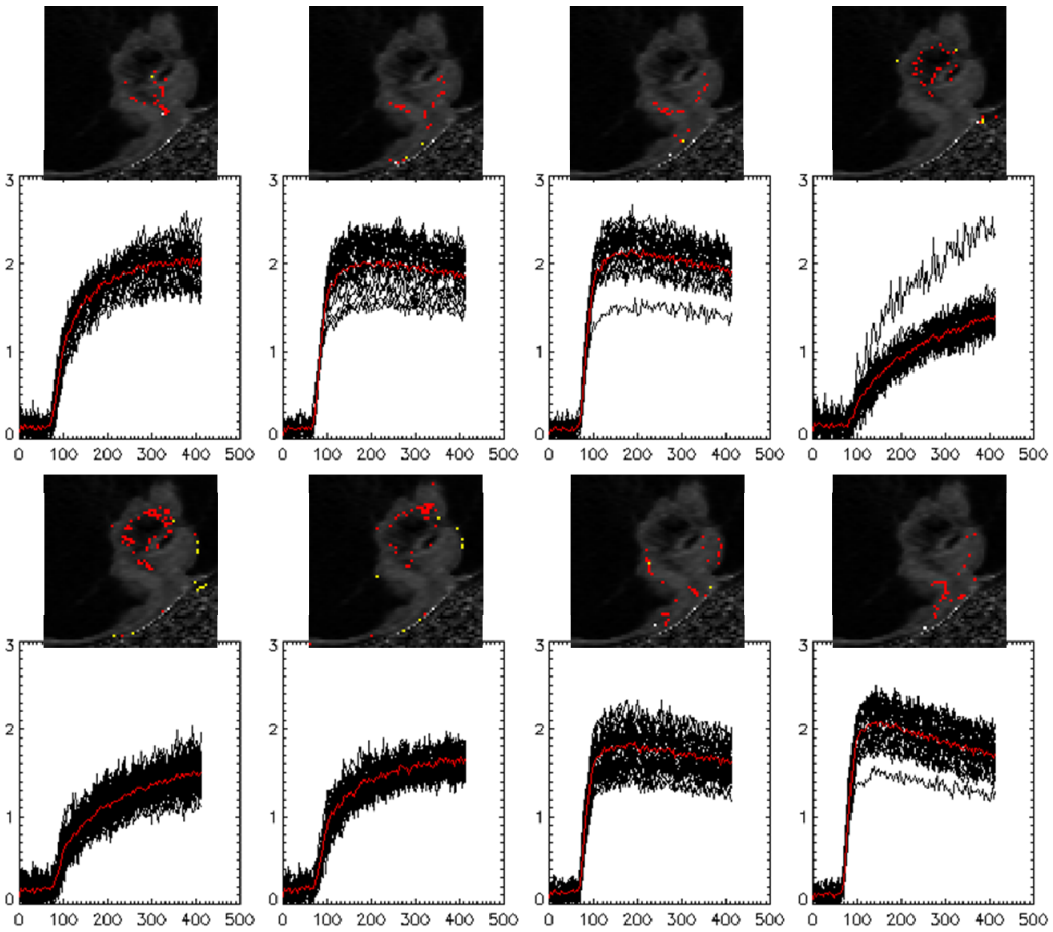

Figure 4. The curves inside each cluster are illustrated for some clusters after clustering of the SCs of patient 1 from figure 3. The clustering is done with 16 sub intervals in L_2 interval and 8 in reverse value. The red curve indicates the mean curve of each cluster.

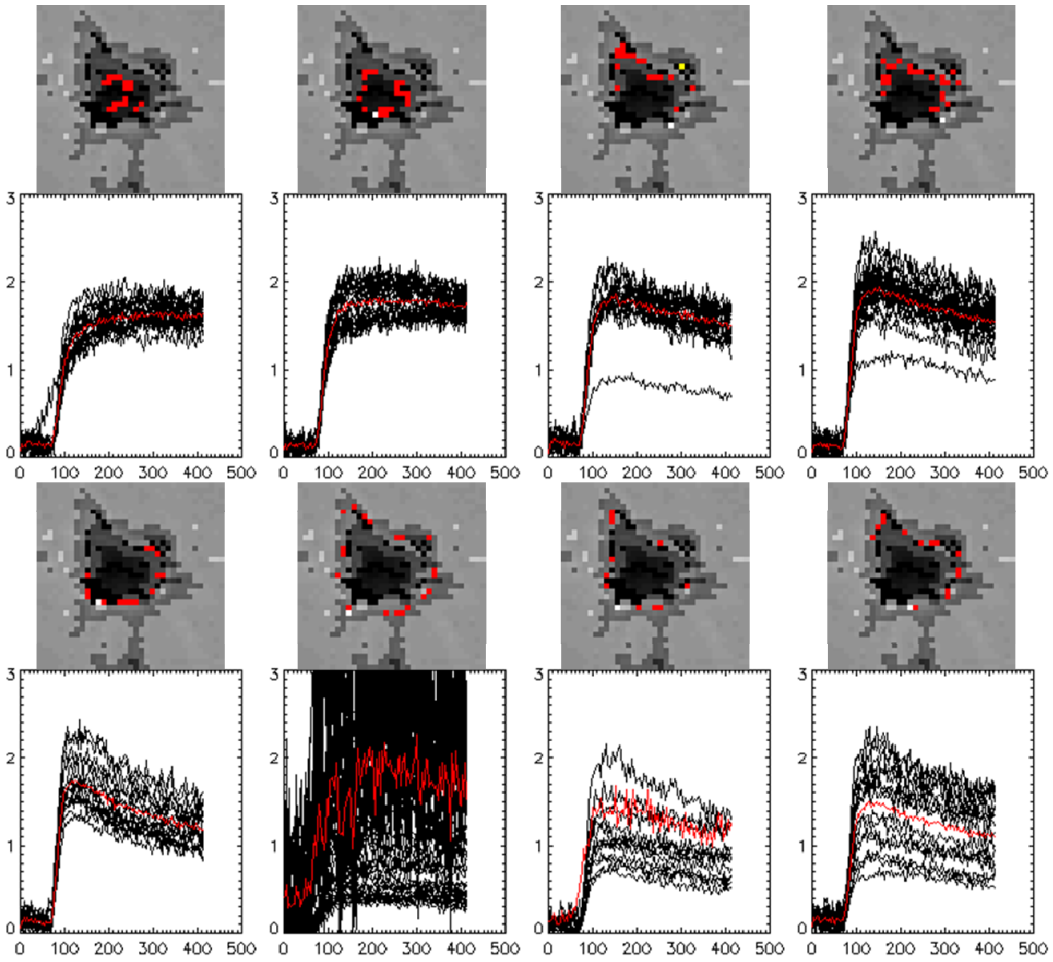

Figure 5. The curves inside each cluster are illustrated for some clusters after clustering of the SCs of patient 2 from figure 3. The clustering is done with 16 sub intervals in L_2 interval and 8 in reverse value. The red curve indicates the mean curve of each cluster. It can be observed that strongly noisy signal curves from other parts like heart can result in similar l_2 and r values and are selected with pixels from tumor area for a cluster.

4. CONCOLUSION

A new and fast method for fully automated segmentation and clustering of DCE-MRI signal curves is introduced. The advantage of this method compared to other existing methods beside its simplicity and fastness is that there are not a fix number of clusters. The area defined by L_2 and r rages is divided into fixed number of sub areas but the number of clusters is automatically defined according to the dispersion of points inside this area. The size of sub areas finally indicates the strictness of clustering.

The results presented in this study demonstrate the advantage of combining the L2-norm and the reverse value compared to the usage of each parameter alone. This method can be used as a pre-segmentation method for further analysis of the curves, for example, for pharmacokinetic analysis of concentration-time curves requiring a high signal-to-noise ratio of the data. Further evaluation of method is necessary to examine its robustness. Moreover, possible advantages of dividing the intervals into sub-intervals of different length as compared to equal length have to be evaluated. The Euclidean distance as a proper similarity measure also has to be discussed in more details.

REFERENCES

[1] Jackson, A., Buckley, D. L., Parker, G.J.M., [Dynamic Contrast-Enhanced Magnetic Resonance Imaging in Oncology], Springer, Berlin & Heidelberg & New York, (2003).

[2] Lucht, R., Delorme, S., Brix, G., "Neural network-based segmentation of dynamic MR mammographic images," Magnetic Resonance Imaging 20(2), 147-154 (2002).

[3] Lavini, C., de Jonge, M.C., van de Sande, M.G.H., Tak, P.P., Nederveen, A.J., Maas, M., "Pixel-by-pixel analysis of DCE MRI curve patterns and an illustration of its application to the imaging of the musculoskeletal system," Magnetic Resonance Imaging 25 (5), 604-612 (2007).

[4] Kubassova, O., Boyle, R.D., Radjenovic, A., "Quantitative analysis of dynamic contrast-enhanced MRI datasets of the metacarpophalangeal joints," Academic Radiology 14(10), 1189-1200 (2007).

[5] Kuhl, C.K., Mielcareck, P., Klaschik, S., Leutner, C., Wardelmann, E., Gieseke, J., Schild, H.H., "Dynamic breast MR imaging: are signal intensity time course data useful for differential diagnosis of enhancing lesions?," Radiology 211, 101-110 (1999).

[6] Wismueller, A., Meyer-Baese, A., Lange, O., Schlossbauer, T., Kallergi, M., Reiser, M.F.., Leinsinger, G., "Segmentation and classification of dynamic breast magnetic resonance image data," Journal of Electronic Imaging 15, (2006).

[7] Zheng, Y., Baloch, S., Englander, S., Schnall, M.D., Shen, D., "Segmentation and classification of breast tumor using dynamic contrast-enhanced MR images," Lecture Notes in Computer Science 4792, 393-401 (2007).

[8] Horn, R. A., Johnson, C. R., [Norms for vectors and matrices], Cambridge University Press, England, (1990).