

Content Based Image Retrieval for Dynamic Time Series Data

Birgit Lessmann^{1,2}, Tim W Nattkemper², Johannes Huth²,
Christian Loyek², Preminda Kessar³, Michael Khazen³, Linda Pointon³,
Martin O. Leach³ and Andreas Degenhard¹

¹Theoretical Physics, Physics Department, University of Bielefeld, 33615 Bielefeld

²Applied Neuroinformatics, Technical Faculty, University of Bielefeld, 33615 Bielefeld

³Clinical MR Research Group, Royal Marsden Hospital,

Institute of Cancer Research, Sutton, Surrey, UK

Email: lessmann@physik.uni-bielefeld.de

Abstract. Content based image retrieval (CBIR) systems in the field of medical image analysis are an active field of research. They allow the user to compare a given case with others in order to assist in the diagnostic process. In this work a CBIR system is described working on datasets which are both time- and space-dependent. Different possible feature sets are investigated, in order to explore how these datasets are optimally represented in the corresponding database.

1 Introduction

In clinical practice, the amount of digitised image data acquired with respect to diagnostic decision making is steadily increasing. In particular three or four dimensional acquisition strategies give rise to large amount of data to be processed. One example is the acquisition of a series of images utilising dynamic contrast enhanced magnetic resonance imaging (DCE-MRI), regarded as a powerful technique for the early detection of breast cancer [1]. In DCE-MRI the uptake of a contrast agent is monitored over a selected time interval to access characteristics of the uptake behaviour in tissue [2]. This allows for detecting abnormalities and, additionally, classifying the detected lesions by their uptake characteristics. Malignant tumours are typically characterised by a signal intensity curve showing a strong and fast uptake often followed by a washout. Contrarily, benign tumours typically possess a slight or moderate uptake behaviour of the contrast agent [2]. The acquired image series contains both, the temporal and the spatial information, resulting in a high dimensional data set. Software tools designed to assist the radiologist in his or her daily clinical work are therefore an active field of research. One particular class of assistant tools are the content based image retrieval (CBIR) systems based on the concept of query by example (QBE), i.e. searching database entries similar to a given query. Here we present a CBIR system for the interactive analysis of DCE-MRI images.

2 Methods

2.1 Content Based Image Retrieval

CBIR systems are based on the concept of query by example (QBE). Given an example case as input, the system output are those cases which are, by a chosen measure, most similar to the query case. CBIR systems for various purposes are currently under development. One type of example are retrieval systems for histopathological applications, usually based on texture, or modality independent retrieval systems for automatic categorisation of image data [3].

2.2 Feature Vector Computation

We evaluate datasets acquired with different acquisition protocols, an additional challenge often unavoidable in daily clinical work. We consider a database containing DCE-MRI data with four and also five post-contrast images. We compare methods that differ in the treatment of these varying datasets with respect to feature vector computation. During the retrieval step the feature vectors of the different cases are compared. To achieve comparable feature vectors for our heterogeneous database we apply different methodologies. An established method to describe signal enhancement in DCE-MRI is the use of radiological meta-features [4], which do not depend on the dimension of the time-series. However, meta features, in general, strongly depend on the particular application, motivating the search for feature vectors applicable in a more generalised context. In this work, we propose and compare possible substitutions for meta features. In the following paragraphs the sets of feature vectors analysed in this work are described in detail.

Truncation In a first approach we truncate a six dimensional time series by omitting the last time-point, so there is the same minimum number of time-points for each sequence.

Extrapolation In a second approach the five dimensional time series are extrapolated to six dimensional ones.

Wavelet features Based on a Discrete Wavelet Transform (DWT) a more sophisticated approach is considered. Wavelet analysis is a method of data analysis to access the localised and scale dependent information in signals [5]. With respect to time series analysis, wavelet-based methods are quite established [6]. We have wavelet-transformed the five or six-dimensional time series and then truncated the achieved series of coefficients to the length of five. I.e. in case of a six-dimensional time series one coefficient is omitted. In contrast the truncation in the time domain, the information of the last time-point, if available, is still partly included in some of the remaining coefficients. Here the Daubechies-4-wavelet [5] is employed for our computations.

Radiological Meta-Features Another type of feature vectors is based on established kinetic meta features used in MR radiology. We have evaluated the CBIR-system using the features percentage enhancement, washout ratio, steepest slope and time-to-peak enhancement as described in [4].

3 Database

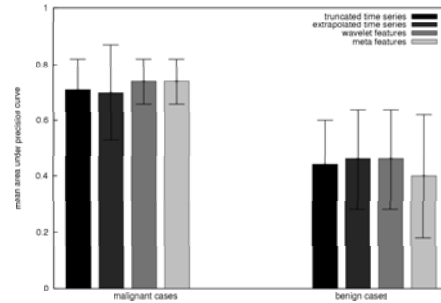
Our analysis includes 27 cases, i.e. 27 time series of DCE-MR images. Each single image in this time series is acquired within an acquisition time of 90 sec. Two images are acquired before and four or five images after injection of the contrast agent. Only one pre-contrast image was included in the further analysis. In 10 cases only four post-contrast images are available resulting in a shorter time series of images. 14 cases are histopathologically classified as malignant and 13 as benign. Segmentation of tumour tissue was performed within a pre-processing step by applying the dynamic pyramid linking algorithm in 3d. This usually results in a decomposition of a tumour into several parts. The average time series computed for each of the latter are then stored as entries in the database, normalised with the pre-contrast value. The database comprises a total number of approximately 3000 entries.

4 Results

A statistical evaluation of the different feature sets is performed. All segments of each tumour for each case are selected as a query example and the entries corresponding to the remaining cases as database content. For each given query example the number of retrieved malignant and benign segments are counted. For a query example classified as malignant the retrieved malignant segments are labelled as true positive (TP) and the retrieved benign ones as false positive (FP) and vice versa. From the values of TP and FP the precision of each query is computed. The four different methodologies are then compared to each other in a statistical evaluation. The results for the best ten matches are shown in figure 1. For each tumour an average precision curve is computed. Figure 1 presents the mean areas under the precision curve. In the averaging process the results for the different segments are weighted regarding the size of the segment. Since this analysis is based on only retrieving the best ten matches, recall values cannot be computed. Since the database comprises 1678 feature vectors describing malignant tissue and 1073 feature vectors describing benign tissue, the *a priori* probabilities of retrieving malignant or benign cases are 0.61 and 0.39 respectively. Regarding these values the results of malignant tissue are significantly better than those of benign tissue. In fact, the results of the benign lesions are quite near to the *a priori* probability.

As can be seen in figure 1 the retrieval results for the truncation and extrapolation procedure are quite similar. The wavelet based features show a very good performance in retrieving malignant tissue and benign cases. The meta features show a performance similar to those of the wavelet features in retrieving

Fig. 1. Mean areas under the precision curve for the for different methods. Left: malignant cases, Right: benign cases



the malignant cases, but provide a performance inferior to the others in case of benign entries. However, regarding the statistical fluctuations measured none of the feature sets can definitely be determined as the best suited one. A further observation is, that the standard deviations of the benign cases are much higher than those of the malignant cases. This is assumed to be related to the following observation. Starting from a database containing the segments of 25 malignant and 17 benign lesions, we count the number of lesions providing retrieval results exceeding the *a priori* probabilities of retrieving malignant or benign cases. The results are shown in Table 1.

In most of the malignant cases the averaged area under the precision curve exceed the *a priori* probability. In case of benign lesions the retrieval results seem to be quite randomly distributed. This leads to the conclusion that malignant tissue presumably shows a more characteristic uptake kinetic than benign tissue. In [7] it was shown, that malignant tumours often contain regions showing a benign uptake characteristic. Thus a lot of segments in the database derived from malignant tumours are difficult to be distinguished from the benign ones.

5 Discussion and Conclusion

We proposed and evaluated a CBIR system for DCE-MR images capable of handling inhomogeneous data sets. Different techniques are proposed to accomplish this task, including a wavelet based scheme. In a statistical analysis we showed that the feature vector sets achieve quite comparable results. However, among the approaches to be compared with the meta features the wavelet based approach shows the smallest standard deviation with respect to malignant cases. This observation corresponds to the fact that, regarding this approach, all lesions except one achieve a retrieval result exceeding the *a priori* probability. Furthermore, an advantage of the wavelet based approach is that it does not require an *a priori* decision for truncating a selected time point or extrapolating an additional time point, to achieve homogeneity necessary for database retrieval. In particular the latter observation is desirable in clinical practice. The observation that the retrieval results of benign tissue only slightly exceed the *a priori* values

Table 1. The number of lesions reaching a retrieval results higher than the *a priori* probability.

method	malignant cases	benign cases
	total: 25	total: 17
truncation	21	8
extrapolation	21	9
wavelet features	24	9
meta features	24	6

is assumed to be related to tumour inhomogeneity. As shown in [7] malignant tumours also contain tissue showing benign characteristics. Thus a significant amount of tumour segments in the database are classified as malignant while showing benign signal enhancement. Thus, we conclude that the representation of a tumour in a database through the result of a segmentation procedure is insufficient for retrieval tasks. Further analysis with respect to tumour features will be necessary to find a representation scheme satisfying the requirements within the clinical practice.

Acknowledgements. The authors would like to thank Oliver Schulz, Max-Planck-Institute Dortmund, for providing software resources. The work was funded by the University of Bielefeld and the federal state Nordrhein-Westfalen, Germany.

References

1. UK MRI Breast Screening Study Advisory Group. Magnetic resonance imaging screening in women at genetic risk of breast cancer: imaging and analysis protocol for the UK multicentre study. *Magn Res Imag* 2000;18:765–776.
2. Heywang-Koebrunner SH, Beck R. Contrast-Enhanced MRI of the Breast. Berlin: Springer-Verlag; 1996.
3. Lehmann TM, Gueld MO, Deselaers T, Keysers D, Schubert H, Spitzer K, et al. Automatic Categorization of medical images for content-based retrieval and data-mining. *Comp Med ImagGraph* 2005;29:143–155.
4. Szabo BK, Aspelin P, Wiberg MKristofferson, Bone B. Analysis of kinetic and morphologic diagnostic criteria. *Acta Radiologica* 2003;44:379–386.
5. Daubechies I. Ten Lectures on Wavelets. SIAM: CBMS-NFS Series Appl. Math.; 1991.
6. Percival DB, Walden AT. Wavelet Methods for Time Series Analysis. Cambridge, UK: Cambridge University Press; 2000.
7. Twellmann Thorsten, Lichte Oliver, Nattkemper TimW. An Adaptive Tissue Characterisation Network for Model-Free Visualisation of Dynamic Contrast-Enhanced Magnetic Resonance Image Data. *IEEE Transactions on Medical Imaging* 2005;24(10):1256–1266.