Published in final edited form as:

Proc Int Conf Image Proc. 2008 October 12; 2008: 3000–3003. doi:10.1109/ICIP.2008.4712426.

COMPUTER-AIDED DIAGNOSIS AND VISUALIZATION BASED ON CLUSTERING AND INDEPENDENT COMPONENT ANALYSIS FOR BREAST MRI

A. Meyer-Baese, O. Lange, T. Schlossbauer, and A. Wismüller

Florida State University, Electrical and Computer Engineering, 2525 Pottsdamer Street, Tallahassee, FL 32310-6046

Abstract

Computer-aided diagnosis and simultaneous visualization based on independent component analysis and clustering are integrated in an intelligent system for the evaluation of small mammographic lesions in breast MRI. These techniques are tested on biomedical time-series representing breast MRI scans and enable the extraction of spatial and temporal features of dynamic MRI data stemming from patients with confirmed lesion diagnosis. By revealing regional properties of contrast-agent uptake characterized by subtle differences of signal amplitude and dynamics, these methods provide both a set of prototypical time-series and a corresponding set of cluster assignment maps which further provide a segmentation with regard to identification and regional subclassification of pathological breast tissue lesions. Both approaches lead to an increase of the diagnostic accuracy of MRI mammography by improving the sensitivity without reduction of specificity.

Index Terms

Unsupervised clustering; topographic ICA; computer-aided diagnosis; visualization; breast magnetic resonance imaging; kinetics enhancement

1. INTRODUCTION

Magnetic resonance (MR) is an emerging and promising new modality for detection and further evaluation of clinically, mammographically and sonographically occult cancers [YDA84,HWP89]. The lesion differential diagnosis in dynamic protocols is based on the assumption that benign and malignant lesions exhibit different enhancement kinetics. In [KMK⁺99] was shown that the shape of the time-signal intensity curve represents an important criterion in differentiating benign and malignant enhancing lesions in dynamic breast MR imaging. The results indicate that the enhancement kinetics, as represented by the time-signal intensity curves visualized in Figure 1, differ significantly for benign and malignant enhancing lesions and thus represent a basis for differential diagnosis. In breast cancers, plateau or washout-time courses (type II or III) prevail. Steadily progressive signal intensity time courses (type I) are exhibited by benign enhancing lesions. Also, these enhancement kinetics is shared not only by benign tumors but also by fibrocystic changes [KMK⁺99].

The small lesion evaluation is performed by an automated computer-aided diagnosis system based on preprocessing of the signal-intensity time-courses (SITC), segmentation of signal-intensity time courses and then automated evaluation of the SITCs based on an unsupervised classifier and ICA techniques. Figure 2 visualize the flow diagram of such a CAD system.

2. MATERIAL AND METHODS

A total of 40 patients, all female and age range 48–61, with solid breast tumors were examined. All patients had histopathologically confirmed diagnosis from needle aspiration/excision biopsy and surgical removal. Breast cancer was diagnosed in 31 out of the total 40 cases.

MRI was performed with a 1.5 T system (Magnetom Vision, Siemens, Erlangen, Germany) equipped with a dedicated surface coil to enable simultaneous imaging of both breasts. The patients were placed in a prone position. First, transversal images were acquired with a STIR (short TI inversion recovery) sequence (TR=5600 ms, TE=60 ms, FA=90°, IT=150 ms, matrix size 256×256 pixels, slice thickness 4 mm). Then a dynamic T1 weighted gradient echo sequence (3D fast low angle shot sequence) was performed (TR=12 ms, TE=5 ms, FA=25°) in transversal slice orientation with a matrix size of 256×256 pixels and an effective slice thickness of 4mm.

The dynamic study consisted of 6 measurements with an interval of 83 s. The first frame was acquired before injection of paramagnetic contrast agent (gadopentatate dimeglumine, 0.1 mmol/kg body weight, Magnevist™, Schering, Berlin, Germany) immediately followed by the 5 other measurements. The initial localization of suspicious breast lesions was performed by computing difference images, i.e. subtracting the image data of the first from the fourth acquisition. As a preprocessing step to clustering and ICA, each raw gray level time-series $S(\tau)$, $\tau \in \{1, \dots, 6\}$ was transformed into a pixel time course (PTC) of relative signal reduction $x(\tau)$ for each voxel, the pre-contrast scan at $\tau = 1$ serving as reference.

2.1. Exploratory Data Analysis Methods

The employed classifier – the neural gas network [MBS93] – is based on grouping image pixels together based on the similarity of their intensity profile in time (i.e., their time courses).

It utilizes a neighborhood–ranking of the reference vectors $\mathbf{w}_i$, corresponding here to the prototypical PTCs sharing similar temporal characteristics, for the given data vector $\mathbf{x}$.

The learning rule for the “neural gas” network is [MBS93]

$$\mathbf{w}_i(t+1) = \mathbf{w}_i(t) + \varepsilon(t) \exp\{-k_i(\mathbf{x}, \mathbf{w}_i/\lambda)\}(\mathbf{x}(t) - \mathbf{w}_i(t)) \quad (1)$$

where $k_i = 0, \dots, N-1$ represents the rank index describing the “neighborhood–ranking” of the reference vectors $\mathbf{w}_i$ to the data vector $\mathbf{x}$ in a decreasing order, N is the number of units in the network, and λ determines the number of neural units changing their synapses with every iteration. The step size $\varepsilon \in [0, 1]$ describes the overall extent of the modification.

ICA is employed to examine the spatio-temporal signal behavior and to visualize areas with high temporal correlation with the extracted signal components. For lesion segmentation we employ standard ICA techniques such as JADE, TDSEP, and topographic ICA [MB03].

Topographic ICA represents a generative model which combines topographic mapping with ICA. As in all topographic mappings, the distance in the representation space given by the topographic grid is related to the distance of the represented components. This distance is defined for topographic ICA by the mutual information implied by higher–order correlations [HH01]. Thus, a natural distance measure is given in the context of ICA. Traditional topographic mapping methods define distance either based on the Euclidian distance or correlation. The ICA distance measure enables the definition of a topography even if the Euclidian distances are all equal as it is the case with an orthogonal vector space.

2.2. Segmentation Methods

In the following, we will present two segmentation methods for the evaluation of signal intensity time courses for the differential diagnosis of enhancing lesions in breast MRI.

The first segmentation method is the clinical standard method to analyze dynamic MRI of the breast and is based on carefully choosing a region of interest (ROI) surrounding the contrast enhancing lesion. For all the voxels belonging to this ROI, an average signal intensity (SI) time curve was computed. In the present study, we selected only lesions with an initial contrast enhancement $\geq 50\%$ for comparative analysis between the standard evaluation method and unsupervised clustering. Thus, we use a semiautomatic segmentation method to determine the ROI including all voxels of a lesion with an initial contrast enhancement of $\geq 50\%$.

Figure 3 illustrates the described segmentation method.

The second segmentation method based on exploratory data analysis techniques, unsupervised clustering or ICA, uses the ROI of the previous segmentation method while all the voxels within this ROI are subject to this analysis. This allows the clinical radiologist a more detailed view of the signal curves by partitioning the ROI. This segmentation method reveals regional properties of contrast-agent uptake characterized by subtle differences of signal amplitude and dynamics. For every single cluster or independent component a mean percentage signal intensity change (PSIC)-curve is computed which contains only the signal information of the signal time curves that are assigned to this particular cluster or IC. For example, if we perform clustering for $N=4$ clusters, then we obtain 4 cluster-specific PSIC-curves, as shown in Figure 4(a) and (b).

3. RESULTS

All lesions ($n=40$) with an initial signal increase $\geq 50\%$ after contrast injection were included in the comparative analysis of the conventional method and cluster analysis. Histological findings were malignant in 31 and benign in 9 lesions. Lesion size was determined as the number of voxels with an initial SI increase of $\geq 50\%$.

Clustering and ICA results were evaluated by (i) qualitative visual inspection of corresponding cluster-specific time-signal intensity curves for the “neural gas” network and the ICA techniques, and (ii) receiver operating characteristic (ROC) analysis.

We have determined the optimal number of clusters and thus the optimal number of characteristic curves for lesion segmentation and classification. A small number will lead to a characteristic curve being very close to the average curve, and thus not able to subdifferentiate the lesion. A larger number, however, makes the characteristic curve prone to noise sensitivity. Our simulation results demonstrated that four clusters are adequate for a correct identification of the time-signal intensity curve types as shown in Figure 4.

The visualization of the segmentation results based on topographic ICA for the same lesion are shown in Figure 4(c) and (d). By considering four independent components, we obtain a better insight into the heterogeneous structure of this lesion.

The advantage of independent component analysis (ICA) over clustering lies in the increase in specificity with about 30% compared to unsupervised clustering for the diagnosis of fibroadenomas. The results demonstrated that ICA identifies both lesion morphology as well as lesion-specific dynamic enhancement patterns.

Results obtained from the clustering and ICA techniques were also evaluated by an ROC analysis. In the ROC curve, we compare both segmentation methods with the gold standard

(histology). From Table 1, we see that the explorative data analysis-based segmentation outperforms the threshold-based. The best results are achieved by the neural gas network, topographic ICA and Jade.

4. CONCLUSION

We introduced an automatic lesion segmentation and classification system based on unsupervised clustering and ICA techniques. The performed ROC-analysis shows that both the unsupervised clustering as well as the ICA techniques represent a valuable tool for supporting radiologic diagnosis in dynamic breast MR imaging. While clustering is in general faster than ICA, the main benefit of ICA lies in an increase in specificity for the diagnosis of benign lesions. This result is quite promising since it eliminates the need of additional morphological criteria for an accurate diagnosis as it has been demonstrated in conjunction with clustering techniques. However, the most important advantage lies in the potential of increasing the diagnostic accuracy of MRI mammography by improving the sensitivity without reduction of specificity for the data sets examined.

Acknowledgments

Thanks to NIH Grant 5K25CA106799-03 for funding.

5. REFERENCES

- HH01. Hyvarinen A, Hoyer P. Topographic independent component analysis. *Neural Computation* 2001;13(7):1527–1558. [PubMed: 11440596]
- HWP89. Heywang S, Wolf A, Pruss E. Mri imaging of the breast: Fast imaging sequences with and without gd-dtpa. *Radiology* 1989;171(2):95–103. [PubMed: 2648479]
- KMK+99. Kuhl CK, Mielcareck P, Klaschik S, Leutner C, Wardelmann E, Gieseke J, Schild H. Dynamic breast mr imaging: Are signal intensity time course data useful for differential diagnosis of enhancing lesions? *Radiology* 1999;211(2):101–110. [PubMed: 10189459]
- MB03. Meyer-Bäse, A. *Pattern Recognition for Medical Imaging*. Elsevier Science/Academic Press; 2003.
- MBS93. Martinetz T, Berkovich S, Schulten K. Neural gas network for vector quantization and its application to time-series prediction. *IEEE Transactions on Neural Networks* 1993;4(4):558–569. [PubMed: 18267757]
- YDA84. Yousef E, Duchesneau R, Alfid R. Magnetic resonance imaging of the breast. *Radiology* 1984;150(2):761–766. [PubMed: 6695077]

Fig. 1.

Schematic drawing of the time-signal intensity curve types [KMK⁺99]. Type I corresponds to a straight (Ia) or curved (Ib) line; enhancement continues over the entire dynamic study. Type II is a plateau curve with a sharp bend after the initial upstroke. Type III is a washout time course $\frac{SI_c - SI}{SI}$ where SI is the precontrast signal intensity and SI_c is the postcontrast signal intensity. In breast cancers, plateau or washout-time courses (type II or III) prevail. Steadily progressive signal intensity time courses (type I) are exhibited by benign enhancing lesions.

Fig. 2.

Diagram of the CAD system employed for the small mammographic lesion evaluation: after a signal preprocessing, the data are segmented based on two different segmentation methods (threshold- and explorative data analysis-based) and the obtained time–signal intensity curves are compared to the four Kuhl classes and automatically assigned to a class by a clustering or ICA.

Fig. 3.
Conventional segmentation method based on thresholding.

(a)

Fig. 4. Segmentation based on “neural gas” network and topographic ICA applied to data set $\#1$ (malignant lesion, DCIS) and resulting in four clusters. (a) shows the cluster distribution for each slice ranging from 6 to 8 and (b) visualizes the representative time—signal intensity time curves for each cluster for the “neural-gas” network and (c) and (d) correspondingly for the topographic ICA.

Results of the comparison between threshold-based segmentation and explorative data analysis-based segmentation for the 40 small lesions. It is illustrated the average area under the curve (AUC) for 20 different ROC runs using the same parameters but different algorithms' initializations.

Table 1

	Conv. Meth.	Neur.-Gas	TopolCA	JADE	TDSEP
AUC	0.6207	0.8388	0.8192	0.8101	0.8183