Shape-based tumor retrieval in mammograms using Relevance-Feedback techniques

Stylianos D. Tzikopoulos[1], Harris V. Georgiou[1], Michael E. Mavroforakis[2], and Sergios Theodoridis[1]

[1] National and Kapodistrian University of Athens, Dept. of Informatics and Telecommunications, Panepistimiopolis, Ilissia, Athens 15784, Greece
[2] University of Houston, Department of Computer Science, PGH 219, 4800 Calhoun Rd., Houston, TX 77204-3010, USA

Abstract. This paper presents an experimental "morphological analysis" retrieval system for mammograms, using Relevance-Feedback techniques. The features adopted are first-order statistics of the Normalized Radial Distance, extracted from the annotated mass boundary. The system is evaluated on an extensive dataset of 2274 masses of the DDSM database, which involves 7 distinct classes. The experiments verify that the involvement of the radiologist as part of the retrieval process improves the results, even for such a hard classification task, reaching the precision rate of almost 90%. Therefore, Relevance-Feedback can be employed as a very useful complementary tool to a Computer Aided Diagnosis system.

1 Introduction

Over the recent years, content-based image retrieval (CBIR) systems are gaining in importance [1,2,3]. Such systems extract visual features from the "query" image, e.g. color, texture or shape and perform a comparison of it with the available images in a database, using specific *similarity measures*. The most *similar* images are returned to the user.

The scenario described above uses low-level features, which are not capable of capturing the image semantics, e.g. the high-level semantic concept that is meaningful for a user. This is known as the *semantic gap*. In order to address this gap, Relevance Feedback techniques have been developed since the early and mid-1990's [4,5]. In such a system, the user interacts with the search engine and marks the images that he perceives as relevant or non-relevant. Taking into account this feedback information, the engine "learns" and improved results are returned to the user during the next iteration.

The search engine is usually a classifier, trained by the relevant and non-relevant samples, labelled by the user [6]. Support Vector Machines (SVMs) [7] are often chosen for this classification scheme. They allow fast learning and evaluation, they do not need restrictive assumptions regarding the data, they are flexible and they turn out to be less sensitive, compared to density-based learners with respect to the problem of class imbalance. Therefore, many Relevance Feedback schemes use the 2-class SVMs for the classification step [8,9,10,11].

The type of the patterns that the search engine returns to the user for labelling is of high importance. For example in [11], the user labels patterns that have been classified as relevant with high confidence, e.g., the furthest patterns to the positive (relevant) side of the classifier. This can be easier for the user, but gives no useful information to the system, leading to slow convergence. More popular approaches adopt the active technique [9]. This approach provides the user with the most informative patterns, e.g. the patterns closest to the decision boundary, in order to improve the speed of the convergence.

Besides image retrieval, Relevance Feedback can also be employed to other systems. In fact, Relevance Feedback was first introduced for the retrieval of text documents [12], music [13], 3D objects [14] and more recently it was used for medical image retrieval [15,16,17,18]. In such a context, the aim of a retrieval system is to function in conjunction with a Computer Aided Diagnosis (CAD) system. The radiologists can be provided with relevant past cases -according to the query-, along with proven pathology and other information, making the diagnosis more reliable. Relevance Feedback seems an ideal scheme for the improvement of the performance of medical image retrieval systems, as it incorporates the radiologist's judgement, in order to capture the some higher-level semantic concepts of the medical images. The judgement of such an expert is the result of a very complex and vague procedure, combining a multitude of quantitative and qualitative facts, as well as the radiologist's experience, and therefore should be taken into consideration.

From these works, only the one in [18] is referred to mammograms. However, it restricts to a small number of images and focuses on micro-calcification clusters, in contrast to our work, which is based on a larger database and focuses on masses. Furthermore, the whole approach is different, as it will become clear in the following sections.

In this work, a Relevance Feedback scheme for the retrieval of mammograms is presented. The system retrieves mammograms containing masses of the same morphology as the query image. The system is tested on a dataset of 2274 masses of the DDSM database [19], that originate from 7 distinct classes. The adopted features for the shape description are first-order statistics of the mass boundary. The latter is given as part of the annotation of the database. The obtained results are promising, according to specific statistical measures and they show a convergence of the relevant retrieved images, reaching the success rate of almost 90%.

The rest of the paper is organized as follows: In section 2, the mammographic image database used is presented. The shape features extracted and the classifier used are described in detail in section 3. Section 4 presents the results obtained, while discussion and conclusion are summarized in section 5.

2 Dataset

The methodology presented in this work was applied to images of the Digital Database of Screening Mammography (DDSM), that are provided and described in [19], available online freely for scientific purposes. This database consists of 2589 cases and each case corresponds to four mammograms, representing the two breasts at craniocaudal (CC) and mediolateral (MLO) oblique views. The images of the database are the result of a digitization procedure, using three different scanners. All the images are analysed by expert radiologists and the corresponding abnormalities (4775) are annotated as calcifications (2201), masses (2556) or other abnormalities (18). Each abnormality has been associated by pixel-level ground truth information, provided by a radiologist, and a complete description of the abnormality is given, including diagnostic assessment, subtlety and proven pathology, as described in [19]. In addition, in case of calcifications, information about the type and the distribution are available and in case of masses, the description includes the mass shape and the mass margin.

The spatial resolution of the images is 50 μm/pixel or 43.5 μm/pixel and the bit depth is 12 or 16 bits, according to the scanner used. A typical image of the database is shown at figure 1.

Fig. 1. Left MLO image of the case 3088 of the database.

3 Methods

3.1 Feature Extraction

For the case of a mass abnormality of a mammogram, a detailed pixel-based ground-truth description is available, as figure 2(a) shows. In addition, characterization of the mass shape is given, including the following classes: Architectural Distortion, Irregular, Lobulated, Lymph Node, Oval, Round, Tubular or Other. Our goal is to predict the class of the shape, using the morphological characteristics of the boundary. The features adopted for this purpose are first-order statistics of the Normalized Radial Distance, as presented in [20,21]. Obviously, other features may also be used. Our choice is justified, by the simplicity of these two features and by the high accuracy that is finally obtained.

First, the centroid of the mass is calculated, using cumulative distributions of the projections in both x and y axes. Then, the radial distance sequence of the centroid and the border pixels are extracted, using equation (1).

$$d\left(i\right)=\sqrt{\left(\left(x\left(i\right)-X_0\right)^2+\left(y\left(i\right)-Y_0\right)^2\right)},\ i=1,2,...,N \tag{1}$$

where the point (X_0, Y_0) is the centroid and $x\left(i\right), y\left(i\right)$ are the parametric data series of length N, corresponding to the border pixels.

In order to overcome problems associated with the non-convex shape of the boundary, in relation to the mass centroid, the Radial Distance Function samples are calculated from the mass perimeter towards the centroid (and not vice versa, as commonly used), as it has been pointed out in [22]. In addition, a pre-processing stage of image dilation [23], using a 3×3 mask, is applied before acquiring the borderline curve, so that to avoid problems with edge-following techniques in cases of sharp corners. The boundary is traced at pixel level, starting from the pixel that corresponds to angle 0. The sequence of the pixels continues to follow the mass boundary counter-clockwise as figure 2(a) illustrates. Finally, a normalization step is adopted using equation (2). The Normalized Radial Distance extracted from the mass boundary of figure 2(a) is shown at figure 2(b).

$$d_n\left(i\right)=\frac{d\left(i\right)-d_{min}}{d_{max}-d_{min}} \tag{2}$$

Seven simple curve features are now extracted from the Normalized Radial Distance, meaning the d_n sequence, using the following equations (3)–(9).

1. The mean value is estimated using equation (3)

$$m=\frac{1}{N}\sum_{i=1}^{N}d_n\left(i\right) \tag{3}$$

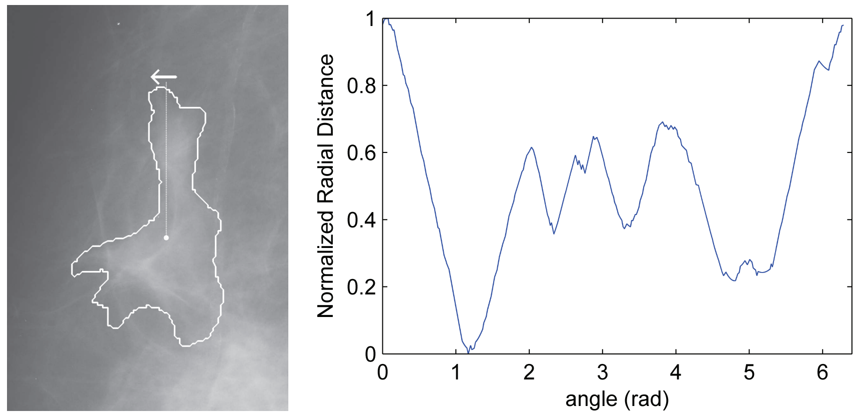
2. The standard deviation is calculated using equation (4)

$$s=\sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(d_n\left(i\right)-m\right)^2} \tag{4}$$

3. The mass circularity is extracted according to equation (5)

$$C=\frac{P^2}{A} \tag{5}$$

where A is the area and P the perimeter of the mass.



(a) (b)

Fig. 2. a) The segmented mass of the right MLO mammogram of case 3129, annotated as Irregular, the mass centroid and the starting pixel on the boundary, and b) the corresponding normalized radial distance signal.

4. The entropy is calculated using equation (6)

$$E = \sum_{i=1}^{B} P_k \log\left(P_k\right) \tag{6}$$

   where P_k are the probability values estimated using a $B = 100$ bins histogram.
5. The area ratio parameter is estimated by equation (7)

$$A_R = \frac{1}{m \cdot N} \sum_{i=1}^{N} \left(d_n\left(i\right) - m\right) \tag{7}$$

   where $A_R = 0$, $\forall d\left(i\right) \leq m$.
6. The zero-crossing count Z_c is computed as a count of the number of times the d_n signal crossed the average d_n value.
7. The roughness index R is calculated by dividing the d_n signal into small segments of equal length and then estimating a roughness index for each one of them according to equations (8) and (9)

$$R\left(j\right) = \sum_{i=j}^{L+j} \left|d\left(i\right) - d\left(i+1\right)\right|, \; j = 1, \ldots, \lceil \frac{N}{L} \rceil \tag{8}$$

$$R = \frac{\sum_{j=1}^{k} \left(R\left(j\right) + \frac{L_{k+1}}{L} \cdot R\left(k+1\right)\right)}{k + \frac{L_{k+1}}{L}}, \; k = \lceil \frac{N}{L} \rceil - 1 \tag{9}$$

where $R(j)$ is the roughness index for the j segment, $L = 16$ the number of boundary points in each segment, k denotes the last segment of length L and N is the total number of boundary points available.

3.2 Classification

For the classification of the masses of the database at step 0 of the Relevance Feedback procedure, a simple Euclidean minimum distance classifier [24] is used. The Euclidean distance E_i of query pattern q from all the available patterns p_i, where $i = 1, 2, \ldots, N_t$, is estimated according to equation (10).

$$E_i = \sqrt{\sum_{j=1}^{S} \left(q(j) - p_i(j)\right)^2} \tag{10}$$

where N_t is the number of all the patterns of the database and $S = 7$ is the dimension of the feature space.

On the next steps of the process, an SVM classifier [25,26] is trained according to the feedback of the user. In the simple SVM case, the system returns the most "confident" relevant patterns for labelling, while in the active SVM case, the system returns the most "ambiguous" patterns for labelling, as described in section 1.

4 Experiments and Results

For the evaluation of the Relevance Feedback scheme, the 2556 masses, included in the database, are used. Note that apart from the detailed boundary of each mass, a classification of the shape of the masses in the following classes is also available: Architectural Distortion, Irregular, Lobulated, Lymph Node, Oval, Round, Tubular or Other. The masses corresponding to the Architectural Distortion class are excluded from the experiments, as this characterization can be extracted mainly by a comparison between the pair of breasts and not from a mass boundary itself. We further exclude 27 masses that have been annotated to belong to more than one classes. Thus the resulting size of the database is 2274 masses with 7 distinct classes. In order to evaluate the performance of the retrieval results at each round of the RF, the *precision curve* [27] is used. The precision at each round is defined as $pr = \frac{R}{N}$, where $N = 10$ is the total number of returned images to the user and R are the relevant images among them.

The experiments were carried out according to the following scenario:

- The user chooses a mass from the database as query image
- Repeat for steps 0 (no feedback yet) to 10 (user gave feedback 10 times)
  - The system returns to the user 10 images for evaluation and the precision is estimated
  - The system returns to the user 10 images to label
  - The user labels a subset of the images, as "relevant" or "non-relevant"

- The system is re-trained, using the feedback of the user as new information

The above scenario is repeated for all the images of the database, in order to achieve more focused results. The system uses the simple SVM scheme [11], according to which the user labels the most confident relevant patterns, or the active SVM scheme [9], where the user labels the most uncertain patterns that lie closest to the classifier's decision boundary. In addition, the user is modeled as follows:

– The 'patient' user, that marks all the patterns returned by the system at each step as relevant or non-relevant, that can lead to imbalanced training sets.
– The less 'patient' user, that marks up to four relevant and four non-relevant patterns, among the patterns that the system returns at each step.
– The 'impatient' user, that marks up to three relevant and three non-relevant patterns, among the patterns that the system returns at each step.
– The 'lazy' user, that marks up to two relevant and two non-relevant patterns, among the patterns that the system returns at each step.

The average precision achieved at each iteration step for all the above configurations is shown in figure 3. Note that all the curves start from the same point at step 0, as no information is given from the user. At step 1, the simple and active techniques of the same type of user achieve equal precision rate, as the available images at step 0 for each user type are the same for these two scenarios. However, at step 1 the user of the active scenario provides more informative feedback than the one of the simple scenario, leading to a quicker convergence of the classifier. This is the reason for the fact that active SVM outperforms the simple SVM at steps greater than or equal to 2, always for the same type of user. The maximum precision rate of 89.7% is observed for the case of active scenario that the user marks up to 4 relevant and 4 non-relevant patterns and not for the 'patient' user, because probably the latter one creates sometimes imbalanced training sets.

Figure 4 shows the average precision for each one of the 7 classes for the scenario of active SVM and a usual user -not 'patient', nor 'lazy'- labeling at most 3 relevant and 3 non-relevant images at each step. The size of each class is given in parenthesis, indicating the number of the masses in the database that belong to each class. It is obvious that all the precision curves, except for the one of the Tubular class (only 4 samples), are increasing monotonically. Therefore, the retrieval process gives better results, on average, for all the classes as the Relevance Feedback proceeds.

5 Discussion and Conclusion

In this work, Relevance Feedback has been employed as a complementary tool to a Computer Aided Diagnosis system, that retrieves masses with similar shape

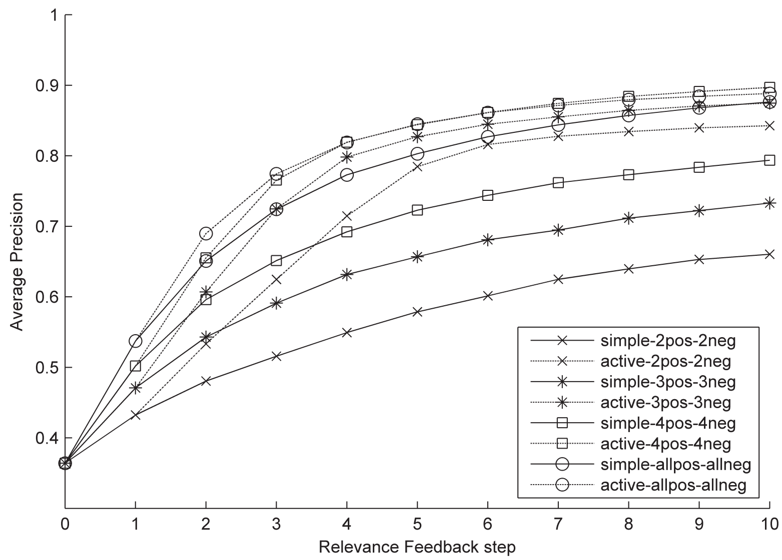


Fig. 3. Average precision at different steps of the RF procedure.

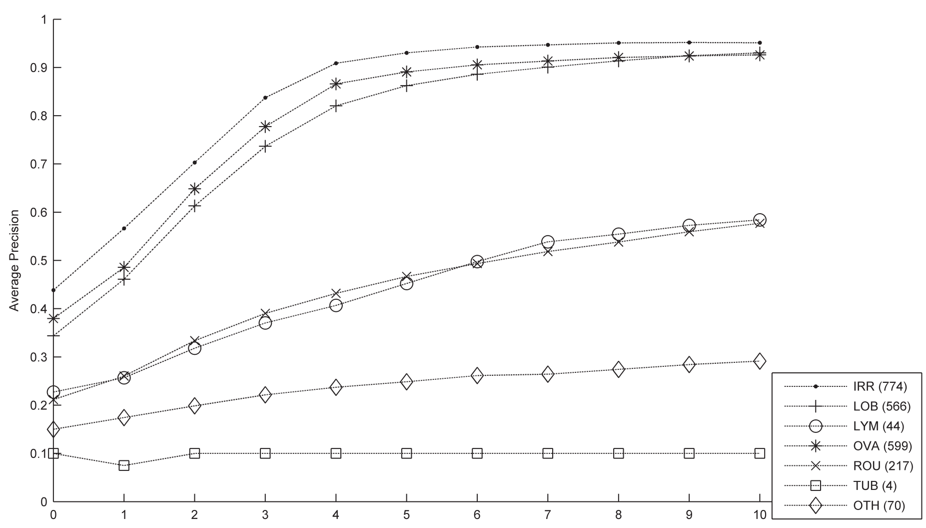


Fig. 4. Average precision per class for active-3pos-3neg scenario.

as the query one. The judgement of the radiologist is considered to be of high importance to such a sensitive system as a medical application, where the errors should be eliminated and therefore it is suggested to be taken into consideration. The results, which almost reach 90% precision rate, show that the retrieval process can be improved significantly, when the radiologist is incorporated in the retrieval process, even for a hard classification task of 7 classes, using features of first-order statistics.

The system converges much faster when the user is more actively involved in the process, by labeling more samples as ”relevant” or ”non-relevant”. In addition, the active technique converges faster to better results than the simple one, while the average precision for each class (figure 4) follows the rules of the Relevance Feedback scheme. The mammographic dataset used for the evaluation is rather extensive, consisting of the large number of 2274 masses, categorized in 7 distinct classes; these facts ensure that the results presented are very useful, reliable and consistent.

The system is also available online for any user at [28].

References

1. Smeulders, A., Worring, M., Santini, S., Gupta, A., Jain, R.: Content-based image retrieval at the end of the early years. IEEE Transactions on pattern analysis and machine intelligence **22**(12) (2000) 1349–1380
2. Gevers, T., Smeulders, A.: Content-based image retrieval: An overview, Prentice Hall (2004)
3. Datta, R., Li, J., Wang, J.: Content-based image retrieval: approaches and trends of the new age. In: Proceedings of the 7th ACM SIGMM international workshop on Multimedia information retrieval, November, Citeseer (2005) 10–11
4. Zhou, X., Huang, T.: Relevance feedback in image retrieval: A comprehensive review. Multimedia systems **8**(6) (2003) 536–544
5. Rui, Y., Huang, T., Ortega, M., Mehrotra, S.: Relevance feedback: A power tool for interactive content-based image retrieval. IEEE Transactions on circuits and systems for video technology **8**(5) (1998) 644–655
6. Crucianu, M., Ferecatu, M., Boujemaa, N.: Relevance feedback for image retrieval: a short survey. State of the Art in Audiovisual Content-Based Retrieval, Information Universal Access and Interaction, Including Datamodels and Languages
7. Vapnik, V.: Estimation of dependencies based on empirical data. Series in Statistics. Springer (1982)
8. Hong, P., Tian, Q., Huang, T.: Incorporate support vector machines to content-based image retrieval with relevant feedback. In: Proceedings of the 7th IEEE International Conference on Image Processing, Citeseer (2000) 750–753
9. Tong, S., Chang, E.: Support vector machine active learning for image retrieval. In: Proceedings of the ninth ACM international conference on Multimedia, ACM New York, NY, USA (2001) 107–118
10. Liu, R., Wang, Y., Baba, T., Uehara, Y., Masumoto, D., Nagata, S.: SVM-based active feedback in image retrieval using clustering and unlabeled data. (2008) 954–961
11. Drucker, H., Shahrary, B., Gibbon, D.: Relevance feedback using support vector machines. (2001) 122–129

12. Salton, G.: Automatic information organization and retrieval. McGraw Hill Text (1968)
13. Chen, G., Wang, T., Herrera, P.: Relevance Feedback in an Adaptive Space with One-Class SVM for Content-Based Music Retrieval. Proceedings of the International Conference on Audio, Language and Image Processing, 2008. ICALIP 2008.
14. Papadakis, P., Pratikakis, I., Trafalis, T., Theoharis, T., Perantonis, S.: Relevance Feedback in Content-based 3D Object Retrieval: A Comparative Study. Volume 5. (2008)
15. Rahman, M., Bhattacharya, P., Desai, B.: A framework for medical image retrieval using machine learning and statistical similarity matching techniques with relevance feedback. IEEE transactions on Information Technology in Biomedicine **11**(1) (2007) 58–69
16. Zhang, Q., Tai, X.: A Relevance Feedback Method in Medical Image Retrieval Based on Bayesian Theory. In: BioMedical Engineering and Informatics, 2008. BMEI 2008. International Conference on. Volume 1. (2008)
17. Song, W., Huangshan, C., Hua, T.: Analytic Implementation for Medical Image Retrieval Based On FCM Using Feature Fusion With Relevance Feedback. In: Proceedings of the 2nd International Conference on Bioinformatics and Biomedical Engineering, 2008. ICBBE 2008. (2008) 2590 – 2595
18. El-Naqa, I., Yang, Y., Galatsanos, N., Nishikawa, R., Wernick, M.: A similarity learning approach to content-based image retrieval: application to digital mammography. IEEE Transactions on Medical Imaging **23**(10) (2004) 1233–1244
19. Heath, M., Bowyer, K., Kopans, D., Moore, R., Kegelmeyer, P.: The digital database for screening mammography. (2000) 212–218
20. Kilday, J., Palmieri, F., Fox, M.: Classifying mammographic lesions using computerized image analysis. IEEE Transactions on Medical Imaging **12**(4) (1993) 664–669
21. Bruce, L., Adhami, R.: Classifying mammographic mass shapes using the wavelet transform modulus-maxima method. IEEE Transactions on Medical Imaging **18**(12) (1999) 1170–1177
22. Georgiou, H., Mavroforakis, M., Dimitropoulos, N., Cavouras, D., Theodoridis, S.: Multi-scaled morphological features for the characterization of mammographic masses using statistical classification schemes. Artificial Intelligence in Medicine **41**(1) (2007) 39–55
23. Gonzalez, R., Woods, R.: Digital Image Processing. Prentice Hall (2007)
24. Theodoridis, S., Koutroumbas, K.: Pattern Recognition. 4th edn. Academic Press (2009)
25. Mavroforakis, M.E., Theodoridis, S.: A geometric approach to support vector machine (svm) classification. IEEE Trans Neural Netw **17**(3) (May 2006) 671–682
26. Mavroforakis, M.E., Sdralis, M., Theodoridis, S.: A geometric nearest point algorithm for the efficient solution of the SVM classification task. IEEE Transactions on Neural Networks **18**(5) (2007) 1545–1549
27. Luo, J., Nascimento, M.: Content-based sub-image retrieval using relevance feedback. (2004) 9
28. Image Processing Techniques for Mammographic Images, `http://mammo.di.uoa.gr`