Content-Based Medical Image Retrieval Using Low-Level Visual Features and Modality Identification

Juan C. Caicedo, Fabio A. Gonzalez and Eduardo Romero

BioIngenium Research Group
National University of Colombia
{jccaicedoru,fagonzalezo,edromero}@unal.edu.co
<http://www.bioingenium.unal.edu.co>

Abstract. This paper presents the image retrieval results obtained by the BioIngenium Research Group, in the frame of the ImageCLEFmed 2007 edition. The applied approach consists of two main phases: a pre-processing phase, which builds an image category index and a retrieval phase, which ranks similar images. Both phases are based only on visual information. The experiments show a consistent frame with theory in content-based image retrieval: filtering images with a conceptual index outperforms only-ranking-based strategies; combining features is better than using individual features; and low-level features are not enough to model image semantics.

1 Introduction

Designing and modeling methods for medical image search is a challenging task. Hospitals and health centers are surrounded by a large number of medical images with different types of contents, which are mainly archived in traditional information systems. In the last decade, content-based image retrieval methods have been widely studied in different application domains [1] and particularly, research in the medical field has taken special interest. The ImageCLEFmed is a retrieval challenge in a collection of medical images [2], which is organized yearly to stimulate the development of new retrieval models for heterogeneous document collections containing medical images as well as text. The BioIngenium Research Group at the National University of Colombia participated in the retrieval task of the ImageCLEFmed 2007 edition [3], using only visual information.

Some important issues for retrieving in heterogeneous image collections are a coherent image modeling and a proper problem understanding. Different modalities of medical images (radiography, ultrasound, tomography, etc.) could be discriminated using basic low level characteristics such as particular colors, textures or shapes as they are at the base of most image analysis methods. Traditional approaches are mainly based on low-level features which describe the visual appearance of images, because those descriptors are general enough to

represent heterogeneous contents [4]. Histogram features and global descriptors have been used to build similarity measures between medical images, obtaining poor results in heterogeneous collections because they do not fully describe the content's semantic.

This work attempts to introduce some slight additional information in the retrieval process by the use of a filtering method. Our approach, firstly try to identify a general modality for each image in the database in a pre-processing phase and then uses histogram features for ranking. This two-phase approach makes use of low-level features to describe image contents and a classification model to recognize the modality associated to one image. We accepted the system should recognize 11 general image modalities so that the retrieval algorithm was forced to a subset of images conceptually related to the query. In this paper some details about the system and the model used by the BioIngenium Research Group to participate in the ImageCLEFmed 2007 edition are presented and discussed. The reminder of this paper is organized as follows: Section 2 presents the two-phase proposed approach. Section 3 presents and discusses the results obtained in the challenge evaluation and Section 4 contains some concluding remarks and future work.

2 Proposed Approach

The image retrieval process consists of two main phases: pre-processing phase and retrieval phase. Both phases are described as follows.

2.1 Pre-processing Phase

The pre-processing phase is composed of two main components: a feature extraction model and a classification model. The input of the pre-processing phase is the original image database, i.e. images from the ImageCLEFmed collection, with more than 66,000 medical images. The output of the pre-processing phase is an index relating each image to its modality and a feature database. This scheme is shown in Figure 1.

The Feature Extraction Model The feature extraction model operates on the image database to produce two kind of features: histogram features and meta-features. Histogram features are used to build the feature database, which is used in the retrieval phase to rank similar images. Meta-features are a set of histogram descriptors, which are used as the input to the classification model to be described later. Histogram features used in this system are [5,4,6]:

- *Gray scale and color histogram (Gray and RGB)*
- *Local Binary Partition histogram (LBP)*
- *Tamura texture histogram (Tamura)*
- *Sobel histogram (Sobel)*
- *Invariant feature histogram (Invariant)*

Fig. 1. Preprocessing phase: The input corresponds to a medical image database. The phase produces as output the feature database and the classified images. This phase uses a low-level feature extraction framework and a classification model based on a multilayer perceptron or a support vector machine.

Metafeatures are calculated from histogram features in order to reduce the dimensionality. These metafeatures are the four moments of the moment generating function (mean, deviation, skewness and kurtosis) and the entropy of the histogram. Each histogram has five associated metafeatures, meaning a total of 30 meta-features with information of color, texture, edges and invariants.

Classification Model Since the data set contains different type of images with different modalities, the proposed approach first attempts to identify the modality of a given image. This restricts the query results to contain images with the same modality as the query image. The classifier is not applied to the raw information of the histograms, since the dimensionality of the feature vector will be very high. Instead, the set of metafeatures are used to reduce the dimensionality, with some information loss. A machine-learning approach is used to classify images in the database. First a training set was selected from the database composed of 2,500 images in 11 categories, each category corresponding to a general image modality . Image modalities are described in Table 1.

Table 1. Image categories.

Category	Examples	Category	Examples	Category	Examples
Angiography	98	Histology	401	Magnetic Resonance	382
Ultrasound	183	Organ photo	196	Tomography	364
Endoscopy	137	Patient photo	171	Drawing	117
Gamagraphy	159	Radiography	344		

This dataset was used as training set for two classifiers. The first classifier is a Support Vector Machine (SVM) with the Gaussian kernel [7]. The second classifier is a Multilayer Perceptron (MP) with one hidden layer and a variable number of neurons, 30 inputs and 11 outputs. Each classifier had a training phase in which the hyper-parameters (complexity for the SVM and number of hidden neurons for the MP) were tuned, using 10-fold cross validation. A test set of images was used to calculate an estimate of the classification error on unseen instances.

Table 2 shows the performance of the best classification models in both training and test sets.

Table 2. Performance of the modality classification models on training and test sets.

	Parameters	Training set error	Test set error
Multilayer Perceptron	Hidden nodes: 40	11.83%	20.78%
Support Vector Machine	$\gamma = 1, \lambda = 2$	18.69%	22.10%

2.2 Retrieval Phase

Fig. 2. Retrieval phase: A query image is received as input and a set of relevant images is generated as output. This phase uses the same feature extraction framework as in the pre-processing phase, but only applied to the query image. It also uses the previously trained classification model on the query image to select the subset of images to rank.

The image ranking process starts by receiving the query. The first step is to classify this image in order to restrict the search only to images with the same modality, herein called the filtering method. Then, the relevance ranking is calculated using different similarity measures.

Filtering Images in the database are filtered according to the modality of the query image. For this purpose, the query image is classified, using the model trained in the pre-processing phase.

Ranking Images are represented in this phase as histograms so that distances are calculated using similarity measures. In this work, five different similarity measures were tested: Euclidean distance, Relative Bin Deviation, Relative Deviation, Chi-square distance and Jhensen-Shannon Divergence; the last four specifically designed for histogram comparison.

The scheme has been adapted from our previous work on content-based image retrieval in a histopathology-image domain [8,9]. In that work, all the combinations of histogram types and similarity measures were tested to choose the best-performing similarity measure for each type of histogram.

Specifically, for each feature-similarity pair the retrieval performance on a set of images was calculated. The better feature-similarity combinations for each histogram feature are shown in Table 3. The similarity measures that produced the best results were Jensen-Shannon Divergence and Relative Bin Deviation. These similarities are defined as follows:

Jensen-Shannon Divergence

$$D_{JSD}(H, H') = \sum_{m=1}^M H_m \log \frac{2H_m}{H_m + H'_m} + H'_m \log \frac{2H'_m}{H'_m + H_m} \quad (1)$$

Relative Bin Deviation

$$D_{rbd}(H, H') = \sum_{m=1}^M \frac{\sqrt{(H_m - H'_m)^2}}{\frac{1}{2} (\sqrt{H_m} + \sqrt{H'_m})} \quad (2)$$

where M is the number of bins in the histogram and H_m is the value of the m -th bin.

Table 3. Feature-metric pairs defined as similarity measures on image contents.

Feature-similarity	Feature-similarity
Gray-RelativeBinDeviation	LBP-RelativeBinDeviation
RGB-RelativeBinDeviation	Tamura-RelativeBinDeviation
Sobel-JhensenShannon	Invariant-JhensenShannon

Similarity Measure Combination The combination of multiple similarity measures may produce better results than using the individual measures. To combine similarity measures we used a Cross Category Feature Importance (CCFI) scheme [10]. This scheme uses the probability distribution of metafeatures to calculate a weight for each similarity measure. The combined similarity measure is:

$$s(x, y) = \sum_{f \in F} \omega(f) s_f(x, y) \quad (3)$$

where x and y are images, F is the feature set, $s_f(\cdot)$ is the similarity measure associated to the feature f and $\omega(f)$ is the importance factor for that similarity measure. The CCFI calculates each ω in the following way:

$$\omega(f) = \sum_{c_j \in J} p(c_j | f)^2 \quad (4)$$

Since we have some categories predefined in the database, we can calculate the weight of each feature using the probability class distribution of features. There are two classifications produced by different classifiers: SVM classification and MP classification. In each case the probability distribution varies according to the final classification. That means that the weights calculated in the scenario of the SVM classifier are different of those calculated in the scenario of the MP classifier.

3 Results and Discussion

3.1 Experimental Settings

We sent eight runs for evaluation that are divided into two groups: one using the MP classifier and the other using the SVM classifier. That is to say, the filtering method in the retrieval phase depends on the selected classifier. As each group of experiments have four runs, they correspond to four different strategies in the ranking method. Although our system have six similarity measures implemented, we sent three runs using only three of them individually: RGBHisto-RBD, Tamura-RBD, Sobel-JS. The fourth run corresponds to the similarity measure combination, that operates with the six implemented similarity measures.

3.2 Results

The results of our eight experiments are shown in Table 4, sorted out by MAP. In this table, the column *Run* shows the name of the sent experiment, following a three-parts convention: (1) *UNALCO* to identify our group at the National University of Colombia; (2) an identifier for the classification model used, *nni* for the multilayer perceptron and a *svmRBF* for the support vector machine; and (3) the name of the filtering method used: RGB histogram (RGBHisto),

Sobel histogram (Sobel), Tamura histogram (Tamura), and lineal combination of features (FeatComb).

Table 4. Automatic runs using only visual information.

Run	Relevant	MAP	R-prec	P10	P30	P100
UNALCO-nni_FeatComb	644	0.0082	0.0149	0.020	0.0144	0.0143
UNALCO-nni_RGBHisto	530	0.0080	0.0186	0.0267	0.0156	0.0153
UNALCO-nni_Sobel	505	0.0079	0.0184	0.020	0.0167	0.0187
UNALCO-nni_Tamura	558	0.0069	0.0167	0.0233	0.0156	0.0153
UNALCO-svmRBF_Sobel	344	0.0056	0.0138	0.0033	0.0133	0.0133
UNALCO-svmRBF_FeatComb	422	0.0051	0.0077	0.010	0.0089	0.0093
UNALCO-svmRBF_RGBHisto	368	0.0050	0.0103	0.0133	0.010	0.0093
UNALCO-svmRBF_Tamura	375	0.0048	0.0109	0.0067	0.010	0.010

The general ranking of our runs follows what is currently considered as true. Firstly, the MP classifier used for image filtering together with a feature-combination strategy for image ranking, shows the best MAP score in this set of runs. In all cases, the MP classifier shows better performance than the SVM to filter images, which is in general consistent with the error rates obtained in the training phase (Table 2). Tamura texture shows the worst results in both filtering strategies. In general, the feature combination approach performs better than individual similarity measures, suggesting that the combination strategy using the *Cross Category Feature Importance* scheme is a useful approach that effectively combine features based on their probability distribution.

3.3 Discussion

The performance of the proposed approach in the competition is actually not enough for medical image retrieval. This could be explained, in particular by the fact that a restricted set of features was used¹ and, in general, by the fact that visual features alone are not enough for achieving a good retrieval performance.

In general, results behave as we expected: low-level features are still poor to describe the medical image semantics. Nevertheless, those results show that our scheme is consistent with general concepts in content-based image retrieval. First, the feature combination strategy performs better than the individual feature approach, suggesting that visual concepts can be modeled by mixing low-level features. Second, the filtering strategy allows a better retrieval than a simple one i.e. only-visual approaches (GE_GIFT and DEU_CS groups). Furthermore, a good filtering strategy allows identification of more relevant images. In fact, the SVM classification model performs poorer than the MP classification model

¹ Content-based image retrieval systems such as GIFT and FIRE use considerably more visual features than our approach.

in the training and testing sets and this could be related to the worst retrieval performance of the SVM-based runs.

4 Conclusions and Future Work

Content-based medical image retrieval is still a challenging task that needs new and clever methods to implement useful and effective systems. This paper discusses the main components of an image-retrieval system based on a two-phase strategy to build an image category index and to rank relevant images. This system is completely based on low-level visual information and makes not use of textual data. In general, obtained results match well with what one would expect, not only because of the well known semantic gap but because of the consistency in feature combination and filtering quality.

The future work at our lab will aim to take full advantage of all information into the collection, i.e. to involve textual data. Although textual data alone has demonstrated to be successful for image retrieval, we are very interested in models that mix up textual and visual data to improve the performance of our retrieval system.

References

1. Santini, S., Gupta, A., Jain, R.: Content based image retrieval at the end of the early years. Technical report, Intelligent Sensory Information Systems, University of Amsterdam (2000)
2. Müller, H., Michoux, N., Bandon, D., Geissbühler, A.: A review of content based image retrieval systems in medical applications clinical benefits and future directions. *International Journal of Medical Informatics* **73** (2004) 1–23
3. Müller, H., Deselaers, T., Kim, E., Kalpathy-Cramer, J., Deserno, T.M., Hersh, W.: Overview of the imageclef 2007 medical retrieval and annotation tasks. *Cross-Language Retrieval in Image Collections (ImageCLEF)* (2007)
4. Deselaers, T.: Features for Image Retrieval. PhD thesis, RWTH Aachen University. Aachen, Germany (2003)
5. Siggelkow, S.: Feature Histograms for Content-Based Image Retrieval. PhD thesis, Albert-Ludwigs-Universität Freiburg im Breisgau (2002)
6. Mark S. Nikson, A.S.A.: Feature Extraction and Image Processing. Elsevier (2002)
7. Schölkopf, B., Smola, A.: Learning with kernels. Support Vector Machines, Regularization, Optimization and Beyond. The MIT Press (2002)
8. Caicedo, J.C., Gonzalez, F.A., Romero, E., Triana, E.: Design of a medical image database with content-based retrieval capabilities. In: *Advances in Image and Video Technology. IEEE Pacific Rim Symposium on Image Video and Technology. PSIVT 2007.* (2007)
9. Caicedo, J.C., Gonzalez, F.A., Romero, E., Triana, E.: A semantic content-based retrieval method for histopathology images. In: *Information Retrieval Technology: Theory, Systems and Applications. Proceedings of the Asia Information Retrieval Symposium, AIRS2008.* (2008)
10. Wettschereck, D., Aha, D.W., Mohri, T.: A review and empirical evaluation of feature weighting methods for a class of lazy learning algorithms. *Artificial Intelligence Review* **11** (1997) 273–314