A Semantic Content-Based Retrieval Method for Histopathology Images

Conference Paper · January 2008
DOI: 10.1007/978-3-540-68636-1_6 · Source: DBLP

3 authors, including:

Fabio A. González
National University of Colombia
209 PUBLICATIONS **3,524** CITATIONS
SEE PROFILE

Eduardo Romero
National University of Colombia
209 PUBLICATIONS **938** CITATIONS
SEE PROFILE

A Semantic Content-Based Retrieval Method for Histopathology Images

Juan C. Caicedo, Fabio A. González, Eduardo Romero
{jccaicedoru, fagonzalezo, edromero}@unal.edu.co

BioIngenium Research Group
Universidad Nacional de Colombia

Abstract. This paper proposes a model for content-based retrieval of histopathology images. The most remarkable characteristic of the proposed model is that it is able to extract high-level features that reflect the semantic content of the images. This is accomplished by a *semantic mapper* that maps conventional low-level features to high-level features using state-of-the-art machine-learning techniques. The semantic mapper is trained using images labeled by a pathologist. The system was tested on a collection of 1502 histopathology images and the performance assessed using standard measures. The results show an improvement from a 67% of average precision for the first result, using low-level features, to 80% of precision using high-level features.

1 Introduction

Medical images have been supporting clinical decisions in health care centres during the last decades, for instance the Geneve University Hospital reported a production rate of 12.000 daily images during 2.002 [9]. Traditional medical image database systems store images as a complementary data of textual information, providing the most basic and common operations on images: transfer and visualization. Usually, these systems are restricted to query a database only through keywords, but this kind of queries limits information access, since it does not exploit the intrinsic nature of medical images.

A recent approach to medical image database management is the retrieval of information by content, named Content-Based Image Retrieval (CBIR)[9] and several systems such as ASSERT [11], IRMA [7] or FIRE [4] work following this approach. These systems allow evaluation of new clinical cases so that when similar cases are required, the system is able to retrieve comparable information for supporting diagnoses in the decision making process. One drawback of current CBIR systems is that they are based on basic image features that capture low-level characteristics such as color, textures or shape. This approach fails to capture the high-level patterns corresponding to the semantic content of the image, this may produce poor results depending on the type of images the system deals with.

On the other hand, it is well known that the diagnosis process in medicine is mainly based on semantic or semiotic knowledge, difficult issues to deal with

when image knowledge contents has to be organized for retrieval tasks. To extract image semantics is a great challenge because of the semantic gap [8], that is to say, the existing distance between conceptual interpretation at a high level and the low-level feature extraction, which is possible using conventional image processing techniques.

The problem of extracting semantic features from images may be approached from two different perspectives: an analytic approach and an inductive approach. The analytic approach requires to understand, with the help of an expert, what a given pattern is; then a model to decide whether the pattern is present or not is built, based on this knowledge. On the other hand, the inductive approach, or machine-learning approach, requires to collect enough image samples where the pattern is present or absent, and to train a model able to discriminate both situations. The inductive approach has many advantages: it just relays on the expert for labeling the samples; the model may be easily retrained when new data is available; and there is not need for dealing directly with the complexity of the patterns. In this work, the inductive approach is followed.

This paper presents the design, implementation and evaluation of a new method for semantic feature extraction of a basal-cell-carcinoma database. The whole system is modeled as to map a set of low-level features into high-level semantic properties for a collection of basal-cell-carcinoma images, which were previously annotated by an expert pathologist. The reminder of this paper is organized as follows. In Section 2, the problem of content-based retrieval in histopathology is introduced. In Section 3, the model for feature extraction is presented. Methods for compare images are in Section 4, Section 5 presents results of the experimental evaluation and some concluding remarks are presented in Section 6.

2 The problem of accessing histopathology images by content

Medical practice constantly requires access to reference information for the decision making process in diagnostics, teaching and research. Previous works have designed CBIR systems for medical image databases providing services such as query-by-example, query-by-regions or automatic image annotations among others [13,10]. This kind of tools helps physicians to take informed decisions and to make case-based reasoning.

An important hypothesis of this work is that domain-specific knowledge may improve the performance of a CBIR system. Particularly, the proposed system deals with histopathology images, so the particularities of this kind of images need to be studied and understood. A basic concept in histology is that there exist four basic types of tissue: epithelial, connective, muscle, and nerve[6]. With very few exceptions, all organs contain a different proportion of these four basic tissues. In general, histological techniques highlight these tissues with few colours since dyes are designed to specifically arise a particular tissue feature. In terms of image processing, histological images are distinguished by having more or less

homogeneous textures or repeated patterns, which may be used to characterise the image. Main information in histological images lyes on repeated patterns of textures, with particular edges and slight color differences.

Histopathology images used in this work were acquired to diagnose a special skin cancer called basal-cell carcinoma. Slides were obtained from biopsy samples which were fixed in paraphin, cut to a 5 mm thickness, deposited onto the glass slides and finally colored with Hematoxilin-Eosin. The whole collection is close to 6.000 images associated with clinical cases. A subset of the collection consisting of 1.502 images were annotated and organized in semantic groups by a pathologist. The groups are representative of the semantic categories that are relevant in the scenario of a content-based image retrieval system, according to the expert. The identified groups are not disjoint because of the nature of the image contents.

3 Feature extraction

This section is devoted to describe how low-level features are transformed into semantic characteristics. The whole process starts by a conventional low-level feature extraction phase that reduces image dimensionality: histograms of predefined edge, texture and color characteristics. Dimensionality is further reduced using statistical descriptors of the histogram up to a fourth order along with its entropy. The resulting feature vector, herein called meta-features, grossly describes the underlying probability distribution associated with each different histogram. Once images are expressed as meta-features, a semantic mapper transforms them into semantic features. This mapper is devised for capturing the pathologist knowledge and is composed of 19 basic components, each specialized upon different concepts previously defined by an expert pathologist. Figure 1 illustrates the feature extraction process.

3.1 Low-level feature extraction

A very convenient approach to face feature extraction consists in using a statistical frame: images are modeled as random variables so that histograms stand for the probability distribution of any of the selected features i.e. edges, textures and colors. Histogram features have been traditionally used in content-based image retrieval to calculate similarity measures and to rank images [3]. The following histogram features were explored:

- *Gray histogram*: Luminance intensities in a 256 scale.
- *Color histogram*: In the RGB color model with a partition space of $8 \times 8 \times 8$
- *Local binary partition*: A local texture analysis to determine neighbor dominant intensities
- *Tamura texture histogram*: Composed of contrast, directionality and coarseness
- *Sobel histogram*: Edge detection
- *Invariant feature histograms*: Local invariant transformations such as rotation and translation

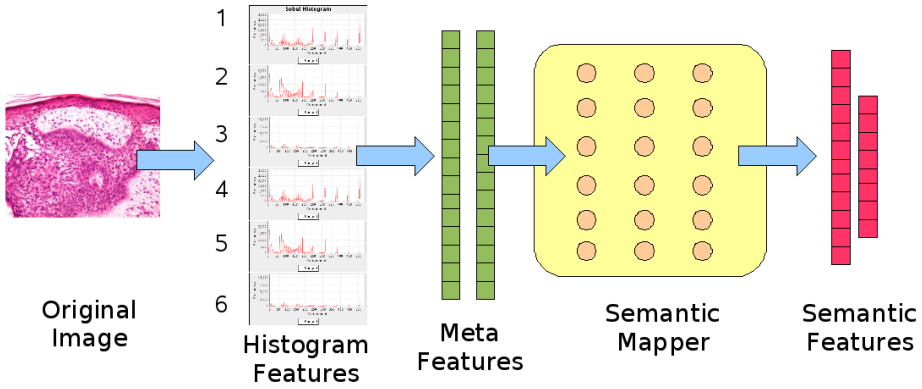

Fig. 1. Feature extraction model. Histogram features are extracted from the original image. Then histograms are processed to obtain meta features, which are the input for the semantic mapper to produce semantic features

A set of meta-features are calculated from the information of the histogram h as follows (and k is a index for histogram bins):

- *Mean*: $\sum_k kh(k)$.
- *Deviation*: $\sum_k (k-\mu)h(k)$
- *Skewness*: $\frac{\mu^3}{\sigma^3}$, the third central moment.
- *Kurtosis*: $\frac{\mu^4}{\sigma^4} - 3$, the fourth central moment.
- *Entropy*: $-\sum_k h(k)ln[h(k)]$.

All meta-features are calculated on each of the six histogram features, which amounts to a total of 30 meta-features per image.

3.2 Semantic mapper

Overall, every pathologist follows a standard training addressed to strength out both diagnosis precision and velocity. An efficient management of these two complementary issues is based on four classic steps: look, see, recognize and understand [2]. A pathologist that evaluates an image is basically looking for patterns and his/her decisions are usually based on the presence or absence of a specific pattern. These patterns are associated with concepts, that is, pathologists give meaning to the image or in other words, they "understand" the image contents. Patterns may correspond to simple low-level features, but in most cases they are a complex combination of them. Features are usually made up of many of this kind of patterns and are called high-level or semantic features. The main hypothesis of this work is that using semantic features we can achieve a better CBIR performance than using low-level features.

The core of the proposed semantic model is the semantic mapper. Since groups are non disjoint, the semantic mapper is not a single classifier but a

model of many learning algorithms identifying the different groups to which the image belongs. This mapper is composed of 18 Support Vector Machine (SVM) classifiers [12], each specialized on deciding whether or not one image belongs to one of the eighteen possible classes, and a extra classifier to detect noise. When the image representation is processed through this semantic mapper, meta-features are individually processed by each of the 19 classifiers. In this model, each classifier outputs a score value indicating whether the image belongs to its group or not. With the output of each classifier, the semantic model builds a semantic feature vector containing the membership degree of one image to every semantic group.

3.3 Semantic mapper training

The dataset used to train each classifier is composed of 1.502 images, organized in 19 different groups (corresponding to the 19 different categories defined by the pathologist). The training dataset is composed of meta-features with their corresponding labels and each group has a specific dataset. Each dataset is entailed with exactly the same attributes except for the class label which can only have two possible values: positive if the example belongs to this group and negative otherwise. In most of the groups there is a considerable amount of imbalance between negative and positive classes, this is solved by resampling the class with less elements. Each dataset is split, using a stratified sampling approach, into two subsets: 20% is used for testing and 80% for training and validation. A 10-fold cross validation scheme on the training set is used to evaluate the classification rate. The test dataset is set aside and used at the end for calculating the final error rate of the classifier.

The very basic unit of a mapper is a SVM classifier, which is provided with different parameters (herein called hyper-parameters) such as the regularization parameter, the type of kernel, and the kernel-specific parameters. Extensive experimentation was performed to select the best set of hyper-parameters for each SVM classifier.

4 Metrics for image content

Similarity evaluation of image contents is achieved using metrics. For image retrieval, metrics are designed to detect differences between the available features. This work uses metrics for two type of contents: low-level features and semantic features as follows.

4.1 Low-level feature metrics

Since low-level features are histograms, they require metrics evaluating differences between probability distributions. Evaluated metrics were Relative Bin Deviation D_{rbd}, Jensen-Shannon Divergence D_{JSD} and Euclidean Distance L_2.

For each feature, we experimentally found the most appropriate metric capturing the feature topology in the image collection, obtaining a set of feature-metric pairs able to rank histopathology images. Many features can be evaluated in an individual metric using a linear combination approach of the feature-metric pairs. If x and y are images; F_k is a function to extract a low-level feature k; and M_k is a metric able to compare the feature k, then, a metric to evaluate many low level features is:

$$d(x, y) = \sum_k w_k M_k \left(F_k(x), F_k(y)\right)$$

where w_k is a factor that controls the relative importance of each feature-metric pair. The best values for all w_k were found by exhaustive search.

4.2 Semantic metric

Semantic features are codified in a vector per image, in which each position represents a value of membership degree to the corresponding group. These values are produced by each component of the semantic mapper and are scaled to fit the $[0, 1]$ range. Each image may belong to many groups at the same time, providing information about the content and interpretation of the overall scene. To compare images in a semantic way, the Tanimoto coefficient was selected, which is a generalization of the Jaccard coefficient [1]. In this problem, Tanimoto coefficient can interpret, how many positions in the semantic vectors are showing coincidences, emphasizing the similarity between concepts shared by both images. Given two semantic vectors A and B, each with 19 positions, the Tanimoto coefficient assigns a similarity score to the associated images as follows:

$$T(A, B) = \frac{A \cdot B}{\|A\|^2 + \|B\|^2 - A \cdot B}$$

5 Experimentation and results

Müller et al [5] presents a framework to evaluate CBIR systems in order to report comparable results from different research centers in a standardized way. The most representatives of those performance measures are precision and recall. Since precision can be measured for different values of recall, the average precision of the n-th result is reported to compare experiments, named $P(n)$. Also a precision vs recall graph may be drawn, which provides information about the behavior of the system in many points. Also, the rank of relevant results is used for measuring performance; in this work, the rank of the first relevant result (Rank1) and the average, normalized rank (NormRank) were used.

Each experiment was configured to select 30 random queries in the collection, through a query-by-example approach. Results associated to each query were evaluated as relevant or irrelevant against the ground truth, and performance measures were averaged to obtain the final result of the experimentation.

Table 1 shows the performance of the CBIR system. In one case, only low-level features were used. In the other case, semantic features were used. For all

the measures, the semantic features outperform the low-level features. This is corroborated by the precision vs. recall graph (Fig. 2).

Model	Rank1	NormRank	P(1)	P(20)	P(50)	P(100)
Low level features	8.22	0.28	0.67	0.30	0.21	0.16
Semantic features	1.96	0.07	0.80	0.59	0.51	0.45

Table 1. Performance measures for image retrieval

The low-level-feature system performance serves a bottom line to assess the real contribution of incorporating domain knowledge to the system. The results corroborate the hypothesis that this domain knowledge, represented on the semantic features, greatly improve the performance of the system.

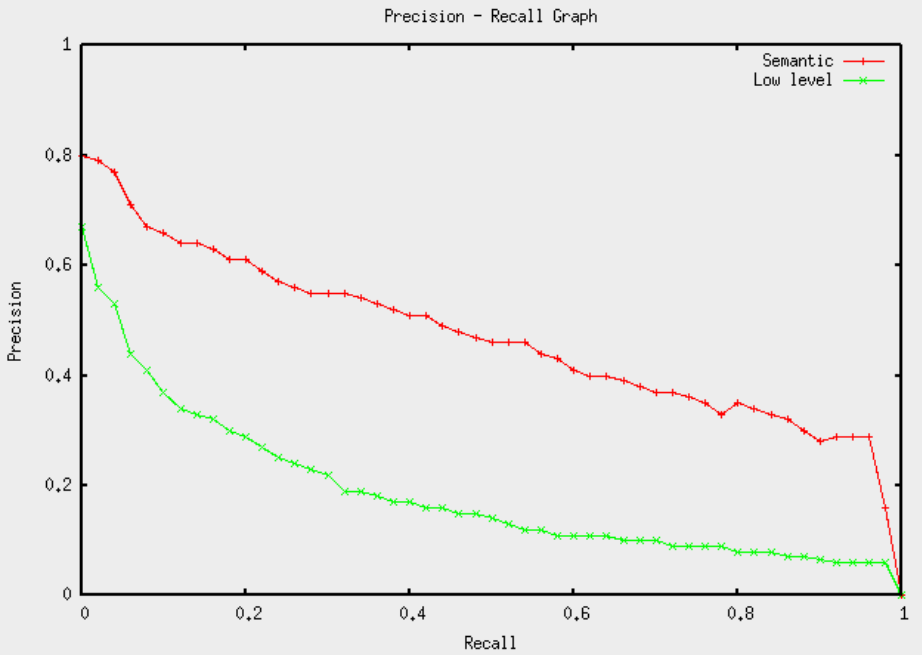

Fig. 2. Precision vs Recall graph comparing the system performance using two types of features: low-level features and semantic features

6 Conclusions and future work

The paper presented a novel approach to represent histopathology knowledge, which is naturally included into a CBIR system. The approach allows to bridge

the semantic gap preserving in the same context both, the low-level features and the semantic features. This was accomplished by a semantic mapper based on state-of-the-art machine-learning algorithms. On the other hand, the experimental results demonstrate the main hypothesis of this work, that is to say, taking into account the semantic content of images highly improves the performance of the CBIR system.

The future work includes exploring richer semantic representations, using other low-level features, performing a more extensive evaluation with a larger bank of images and additional pathologists to test the system.

References

1. Sing T. Bow, editor. *Pattern Recognition and Image Preprocessing*. Marcel Dekker. Inc, 2002.
2. Gianni Bussolati. Dissecting the pathologists brain: mental processes that lead to pathological diagnoses. *Virchows Arch*, 448(6):739–743, 2006.
3. Thomas Deselaers. *Features for Image Retrieval*. PhD thesis, RWTH Aachen University. Aachen, Germany, 2003.
4. Thomas Deselaers, Tobias Weyand, Daniel Keysers, Wolfgang Macherey, and Hermann Ney. Fire in imageclef 2005: Combining content-based image retrieval with textual information retrieval. *Image Cross Language Evaluation Forum*, 2005.
5. Müller H, Müller W, Marchand-Maillet S, McG Squire D, and Pun T. A framework for benchmarking in visual information retrieval. *International Journal on Multimedia Tools and Applications*, 21:55–73, 2003.
6. Luiz Carlos Junqueira and José Carneiro. *Basic Histology, Tenth Edition*. MacGraw Hill, 2003.
7. Thomas Lehmann, Mark Güld, Christian Thies, Benedikt Fischer, Klaus Spitzer, Daniel Keysersa, Hermann Neya, Michael Kohnen, Henning Schubert, and Berthold Weinb. The irma project: A state of the art report on content-based image retrieval in medical applications. *In Korea-Germany Workshop on Advanced Medical Image*, pages 161–171, 2003.
8. Ying Liu, Dengsheng Zhang, Guojun Lu, and Wei-Ying Ma. A survey of content-based image retrieval with high-level semantics. *Pattern Recognition*, 40:262–282, 2007.
9. Henning Müller, Nicolas Michoux, David Bandon, and Antoine Geissbuhler. A review of content based image retrieval systems in medical applications clinical bene ts and future directions. *International Journal of Medical Informatics*, 73:1–23, 2004.
10. Euripides Petrakis and Christos Faloutsos. Similarity searching in medical image databases. *IEEE Transactions on Knowledge and Data Engineering*, 9, 1997.
11. Chi-Ren Shyu, Carla Brodley, Avinash Kak, Akio Kosaka, Alex Aisen, and Lynn Broderick. Assert: A physician-in-the-loop content-based retrieval system for hrct image databases. *Computer Vision and Image Understanding*, 75:111–132, 1999.
12. Alexander J. Smola and Bernhard Schölkopf. *Learning with kernels: Support Vector Machines, Regularization, Optimization, and Beyond*. The MIT Press, 2002.
13. James Z. Wang. Region-based retrieval of biomedical images. *International Multimedia Conference Proceedings of the eighth ACM international conference on Multimedia*, 2000.