Classification of CT figures in biomedical articles based on body segments

Zhiyun Xue, Sameer Antani, L. Rodney Long, Dina Demner-Fushman, George R. Thoma
Lister Hill National Center for Biomedical Communications
National Library of Medicine
Bethesda, MD, USA

***Abstract*—Figures in biomedical articles provide important information that can be utilized to enrich user experience in biomedical article retrieval. One method to improve retrieval performance is to categorize figures into various modalities. We have previously used a hierarchical classification strategy that significantly improves retrieval performance. In this paper, we extend the hierarchy and add body segment classification, i.e., classifying the figures in CT (computed tomography) modality into different body segments, such as head, abdomen, pelvis, or thorax. To address the large variety of article images, we extracted a wide set of feature types (feature vector length of 2321) and applied a multi-class SVM classifier. Feature selection was applied to reduce the feature vector to length 50. Evaluation of the proposed method on a dataset consisting of 2465 figures from a subset of open access biomedical articles from the National Library of Medicine's (NLM) PubMed Central® repository achieves classification accuracy of over 90%. This demonstrates its effectiveness and potential to become a vital component in biomedical document retrieval systems such as OpenI, a multimodal biomedical literature search system developed at NLM.**

Keywords— figure classification; content-based image retrieval; CT image classification; biomedical article retrieval

I. INTRODUCTION

Authors frequently use figures in biomedical articles to illustrate cases and demonstrate findings. Figures can be clinical images, graphs, diagrams, screenshots, and photos, all of which provide rich visual information. This information is broadly considered useful for research and education in medical science. Therefore, it is of great significance if biomedical publication retrieval systems can provide the capability of searching figures effectively. Although figure captions and related descriptions of figures in the articles provide valuable information about the content of figures and can be utilized in figure searching, they alone cannot effectively represent the visual information in the figures. Hence, there is a clear need to develop a multi-modal searching strategy which combines both text and visual features, in particular by exploiting the visual content of figures. In previous work, we have published information about our prototype multimodal system called OpenI[1] which currently provides access to over 1.3 million figures from over 450,000 biomedical publications. The reader is referred to [1] and [2] for details on the OpenI system and other comparable medical information retrieval systems.

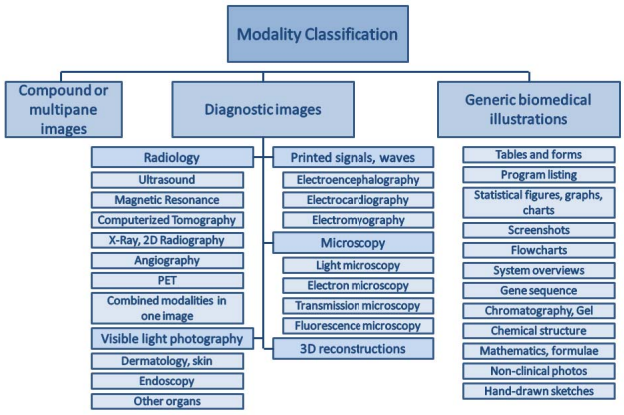

Fig. 1. ImageCLEF modality hierarchy

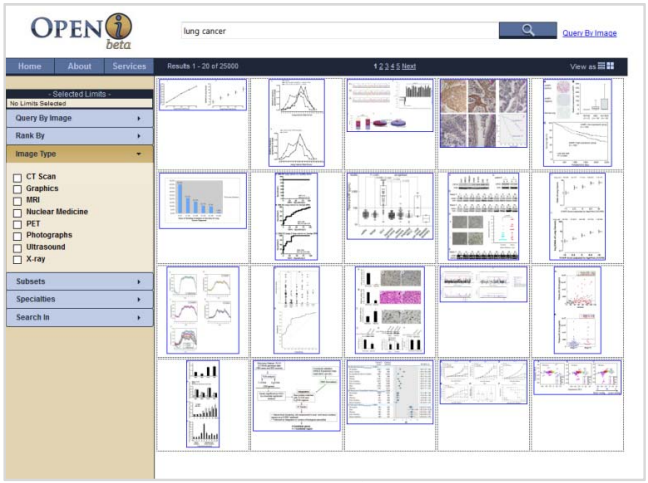

Fig. 2. OpenI image type filter

[1] http://openi.nlm.nih.gov/

978-0-7695-5089-3 2013
U.S. Government Work Not Protected by U.S. Copyright
DOI 10.1109/ICHI.2013.17

IEEE computer society

For research and development of multimodal searching, our group has been participating in the ImageCLEF [2] medical retrieval track, ImageCLEFmed. ImageCLEFmed focuses on medical image annotation and retrieval and aims to provide support and resources for the evaluation of visual information retrieval systems. Some of our techniques evaluated in the benchmark have been integrated into the OpenI system. Since filtering search results by image modality is not only a desirable feature for end users, but also significantly improves retrieval performance, modality classification became one of the subtasks of ImageCLEFmed 2012 [3]. This modality classification task aimed to evaluate the state of the art in figure classification for a subset of the open access articles from NLM's PubMed Central®[3], which is a free archive of open access biomedical and life sciences journal literature at the National Library of Medicine. The images were to be categorized into one of thirty-one modalities as shown in Figure 1. Our ITI (Image and Text Integration) group investigated various approaches for modality classification, and the best of our submitted runs was ranked fifth of all the submissions [4]. Compared to the flat classification strategy that classifies all classes together, a hierarchical classification strategy which utilizes the hierarchical structure of class taxonomy achieved better performance. The class taxonomy used in ImageCLEF 2012 aims at completeness, that is, it tries to cover all of the classes of images that appear in the biomedical literature However, to the end users of biomedical retrieval systems, such as clinicians and patients, the importance of each class varies. For example, users are likely more interested in diagnostic images than illustration figures, and may want to limit their search to a certain imaging modality, such as CT or X-ray, rather than a certain type of illustration, such as flow charts or system diagrams. In OpenI, we implemented a filter which allows users to limit their search to eight image types: CT scan, graphics, MRI, nuclear medicine, PET, photographs, ultrasound, and X-ray, as shown in Figure 2. For any medical image class taxonomy, there is also a likely need/interest to add a new level: body segment classes. That is, to further classify the diagnostic image figures, into different body segments, such as head, abdomen, pelvis, or thorax; this is the goal of our current work. In the following, we report on our initial progress towards reaching this goal. We focus on CT modality, as CT scans are frequently used in hospitals to examine abnormalities in various body locations and are subsequently often shown in biomedical documents for illustration and discussion.

The rest of the paper is organized as follows. We first present the proposed method in Section II. Then we describe experimental tests on a subset of data from ImageCLEFmed and discuss the results in Section III. Section IV draws the conclusions and provides directions for future work.

II. METHOD

Even for the same modality, the figures have very large variation with respect to image size, intensity illumination, window setting, viewing direction, anatomical position, pathology abnormalities, organ entirety (amount of the organ visible), arrow annotation, etc. For our initial feasibility evaluation, we limit our efforts to axial view cross sections and the images containing one whole/near-whole object (object entirety). With respect to object entirety, Figure 3 and Figure 4 give several examples of images that are included or not included in the study, respectively. The data used in this study are from ImageCLEF collections. Since the proposed method is a supervised classification method, a ground truth dataset needs to be created first. To label each figure (i.e., to identify which body segment the image belongs to) in the ground truth dataset, we were guided by the online LUMEN Cross-Section Tutorial[4] which uses CT slices from the Visible Human Project [5]. The tutorial divides the body into six regions: head and neck, upper limb, thorax, abdomen, pelvis, and lower limb, but we considered only five of them: head and neck, thorax, abdomen, pelvis, and lower limb, because the cross sections of upper limb overlap with those of thorax. In addition, the lower limb region defined in our approach doesn't contain the section overlapped with the pelvis region. The guideline for how to define body segment partition is given as follows: 1) the separation between the head and neck region and the thorax region is decided by the first presence of the clavicle in the cross section image; 2) the separation between the thorax region and the abdomen region is decided by the first disappearance of the inferior lobe in the cross section image; 3) the separation between the abdomen region and the pelvis region is decided by the first presence of the ilium in the cross section image; 4) the separation between the pelvis region and the lower limb region is decided by the first disappearance of the ischial tuberosity in the cross section image. We discuss the generation of ground truth data more in the results section. After creating the ground truth dataset, a number of features were calculated from the images. These features and corresponding labels were then used to train a supervised classifier. We will discuss these steps below.

Fig. 3. Images containing a whole or near-whole cross section

[2] http://www.imageclef.org/

[3] http://www.ncbi.nlm.nih.gov/pmc/

[4] http://www.meddean.luc.edu/lumen/meded/grossanatomy/x_sec/

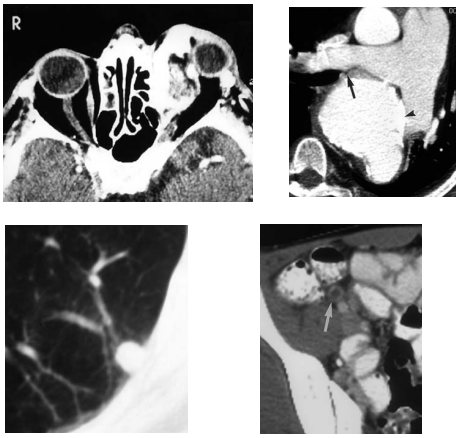

Fig. 4. Images containing a partial cross section

A. Feature Extraction

There are many different features proposed in the literature whose effectiveness is application dependent. For most cases, using one or a few feature types is inadequate. Therefore, to achieve satisfactory classification results, we first apply various types of descriptors that have been shown to be effective for other applications, i.e. we represent the image characteristics using a large number of features. Then, a feature selection procedure is used to remove irrelevant and redundant features. Fourteen types of feature descriptors are applied to represent the visual characteristics of figures. Each feature descriptor is described briefly below (several descriptors that are closely related are introduced together, such as CEDD feature descriptor and FCTH feature descriptor). For detailed techniques, please refer to the referenced literature. Please note that in this paper, we focus on visual features only. Combining visual features with text features, such as captions, is planned for future work.

- Tamura descriptor: Tamura features [6] are texture features based on studies of human visual perception. They consist of six features corresponding to six visually meaningful texture properties: coarseness, contrast, directionality, line-likeness, regularity, and roughness. However, through experiments, the authors [6] found only three of the six features have strong correlation with human perception. They are: coarseness, contrast, and directionality, which are used in this paper.
- CEDD and FCTH: CEDD (color and edge directivity descriptor) [7] and FCTH (fuzzy color and texture histogram) [8] are two descriptors used by the Lucene image retrieval (LIRE) library for image indexing and retrieval. Both features incorporate color and texture information in one histogram which is computed by combining three fuzzy units. The first and second fuzzy units, the parts for color information representation, are the same for CEDD and FCTH. CEDD and FCTH differ in the third fuzzy unit, which captures texture information. Both features are compact, and their sizes are limited to less than 72 bytes per image.
- GLCM: GLCM (gray-level co-occurrence matrix) is a well-known texture analysis method. Five of the 14 features proposed by Haralick [9] are used for our application: maximum probability, contrast, entropy, uniformity, and inverse difference moment.
- Color moments: moments and the related invariants are widely used in image pattern analysis. Color moments [10] consist of the first, second, and third moments of each color channel.
- CLD and EHD: CLD (color layout descriptor) and EHD (edge histogram descriptor) are MPEG-7 features [11]. CLD captures the spatial layout of the dominant colors on an image grid consisting of 8 by 8 blocks and is represented using DCT (discrete cosine transform) coefficients. EHD represents the local edge distribution in the image, i.e. the relative frequency of occurrence of five types of edges (vertical, horizontal, 45-degree diagonal, 135-degree diagonal, and non-directional) in the sub-images.
- Bag of SIFT: SIFT (scale invariant feature transform) [12] features are local features that are relatively invariant to translation, scaling, orientation, and image noise. The SIFT algorithm consists of four major steps. First, it detects the maxima and minima in the scale space. Then, it identifies key points by removing those extrema with low contrast. Then, it assigns an orientation to each key point. Finally, it computes the local image gradient feature measured relative to the orientation of the key point, to provide invariance to rotation. For each key point, a vector of length 128 features distinctively represents the neighborhood around it. The extracted SIFT features of key points are then clustered and the images are represented by a bag of these quantized features, similar to the method of "bag of words" in text retrieval. For details of the method, please refer to [13].
- LBP: The LBP (local binary pattern) [14] operator is a texture descriptor that is robust against illumination changes. The texture information in the image is represented by a histogram of binary patterns. The binary patterns are generated by thresholding the relative intensity between the central pixel and its neighboring pixels. Because of its computational simplicity and efficiency, LBP and its extensions have been successfully used in various computer vision applications. Two versions of LBP based features [15] are extracted: one is calculated using the original LBP definition and the local contrast measure, the other is obtained with the joint distribution of LBP and local variance in a circularly symmetric neighborhood.
- Local color histogram: the image is divided into blocks, and, for each block, a color histogram is computed. The feature is a cascade of the histograms of all the blocks.
- Primitive length, edge frequency, and autocorrelation: primitive length [16], edge frequency [16], and autocorrelation [16] are well-known texture analysis

methods which use statistical rules to describe the spatial distribution and relation of gray values.

- Image width to height ratio

The length of each feature type is given in Table I.

TABLE I. LENGTH OF FEATURES

Feature	Length	Feature	Length
CEDD	144	Edge frequency	25
FCTH	192	Tamura descriptor	18
CLD	16	Color moments	3
EHD	80	Primitive length	5
Bag of SIFT	256	width to height ratio	1
LBP1	256	Local color histogram	1024
LBP2	256	Autocorrelation coefficients	25
GLCM	20		
	Combined:	2321	

B. Feature Selection

As stated in Section II.A, a large number of features are extracted and the overall feature length is over 2300. Therefore, we also applied a feature selection procedure to remove redundant and irrelevant feature variables, with the goal of reducing training/testing time and possibly improving classification performance. Generally, a feature selection algorithm contains four main stages: subset generation, subset evaluation, stopping criteria, and result validation. There are many feature selection methods proposed in the literature based on different search strategies and evaluation measures. For a good survey on feature selection methods, please refer to [17]. We employed the methods implemented in WEKA[5], Java-based open source machine learning software.

C. Classification

Image classification methods can be generally categorized into two broad groups: supervised classifiers and unsupervised classifiers. In contrast to unsupervised classifiers, supervised classifiers require a stage of learning/training for the classifier parameters, but can generally achieve better performance. There are many kinds of supervised classifiers including support vector machines (SVM), decision trees, Bayesian networks, neural networks, and Boosting. Among these classifiers, SVMs are the leading methods, and achieve very good generalization performance on a wide range of applications. Basically, for linearly separable binary class patterns, SVM finds the optimal decision hyperplane, which has the largest distance to the nearest training samples in the different classes. For patterns that are not linearly separable, SVM maps the original data into a new high dimensional space using kernel functions and obtains the maximum margin separating decision surface in the new space. To extend the SVM algorithm to multi-class cases, several simple and effective combination methods can be used. We used the method of *one-against-one*[18]. This method combines all pair-wise comparisons of binary SVM classifiers. The N-class case is divided into $N(N-1)/2$ two-class cases. A binary SVM classifier is trained for each pair of classes. When presented a test case, each binary SVM classifier gives one vote to the winning class. The class having the largest number of votes is then assigned as the label of the test case. The sequential minimal optimization (SMO) algorithm [19], is an algorithm widely used for SVM training because of its efficiency in solving the optimization problem arising from the derivation of the SVM. We use WEKA, which incorporates SMO, for our application.

[5] http://www.cs.waikato.ac.nz/ml/weka/

III. RESULTS AND DISCUSSION

For supervised classification, manual annotation is required to generate the labeled dataset used to train and evaluate the classifier. To create the labeled dataset (ground truth), we use text searching on figure captions and article text referencing the figures to extract the candidate images for each class (body segment) from the ImageCLEF dataset and then manually filter out the wrong cases by visual examination. Specifically, the figures whose captions or text snippets in the article that discuss the figure contain the text query ("CT" and the name of body segment, for example, head/neck/brain for "head and neck" class, or thorax/chest/lung for "thorax" class) are retrieved as candidates. The visual examination is carried out under the guidelines described in Section II. The final number of images for each body segment is given in Table II (because the obtained number of images in the lower limb class is very limited, we discarded this class and consider four classes only: head and neck, thorax, abdomen, and pelvis). The total number of figures in the dataset is 2465.

TABLE II. NUMBER OF SAMPLES

Body Segment	Head & neck	Thorax	Abdomen	Pelvis
Number of images	412	1088	542	423

To improve computation efficiency, we applied one attribute selection method provided in WEKA (AttributeSelectedClassifier in the meta-classifier in which the dataset is reduced by attribute selection before being passed to a classifier). Specifically, the *feature evaluator* is set to "CfsSubsetEval" which estimates the value of a subset of attributes by considering the individual predictive ability of each feature along with the degree of redundancy between the features; the *search method* is set to "BestFirst" which searches the space of feature subsets by greedy hill-climbing augmented with a backtracking facility. This procedure reduces the length of the feature vector to 50. These 50 attributes belong to the features of CEDD, CLD, EHD, FCTH, LCH, primitive length, bag of SIFT, Tamura descriptor, and image width to height ratio. Therefore 14 types of features are reduced to 9 types. For classification, the SMO method was applied and the default values in WEKA were used (polynomial kernel with exponent being 1). We evaluated the classification using ten-fold cross-validation. Table III and Table IV show the evaluation results for the classification with and without the step of feature selection, respectively. The measures used for evaluation are: true positive (TP) rate, false positive (FP) rate, precision,

recall, F-score, and ROC area. Comparing Table III and Table IV, the one with feature selection has only a slight performance drop when using only 50 selected attributes instead of the original 2321. These results demonstrate the effectiveness of the proposed method (92.3% accuracy without feature selection and 91.8% accuracy with feature selection). The method can be easily extended to other modalities under the "Radiology" category shown in Figure 1, such as X-ray.

TABLE III. CLASSIFICATION RESULTS (WITHOUT FEATURE SELECTION)

	TP rate	FP rate	Precision	Recall	F-score	ROC area
Head&neck	0.934	0.003	0.982	0.934	0.958	0.988
Thorax	0.952	0.037	0.953	0.952	0.953	0.967
Abdomen	0.891	0.048	0.84	0.891	0.865	0.939
Pelvis	0.875	0.02	0.9	0.875	0.887	0.965
Average	0.923	0.031	0.924	0.923	0.923	0.964
Accuracy	92.3%					

TABLE IV. CLASSIFICATION RESULTS (WITH FEATURE SELECTION)

	TP rate	FP rate	Precision	Recall	F-score	ROC area
Head&neck	0.976	0.006	0.971	0.976	0.973	0.993
Thorax	0.943	0.028	0.964	0.943	0.954	0.969
Abdomen	0.876	0.056	0.815	0.876	0.844	0.925
Pelvis	0.853	0.021	0.894	0.853	0.873	0.964
Average	0.918	0.029	0.92	0.918	0.919	0.963
Accuracy	91.8%					

IV. CONCLUSION AND FUTURE WORK

Figure searching is one vital component in an article retrieval system, since biomedical researchers regularly add figures into their publications to demonstrate clinical study and research results. Since classification of figure types can facilitate feature searching, it has become an important recent research topic. In this paper, we extend our previous work on figure type classification and propose a new method to automatically classify CT figures into four major categories of body segments: head and neck, thorax, abdomen and pelvis. Fourteen different types of image feature descriptors are employed for characterizing the visual properties of figures. We also performed feature selection to obtain a tractable set of features. We used a supervised multi-class classifier based on the SVM algorithm. The method was tested on a dataset of 2465 figures, yielding an overall performance of over 90% accuracy. Future research directions include two main aspects. We will incorporate text features in addition to the image features used in our current study. We will also apply the method to other suitable modalities.

ACKNOWLEDGMENT

This research was supported by the Intramural Research Program of the National Institutes of Health (NIH), National Library of Medicine (NLM), and Lister Hill National Center for Biomedical Communications (LHNCBC).

REFERENCES

[1] D. Demner-Fushman, S. Antani, M. Simpson, G. R. Thoma, Design and development of a multimodal biomedical information retrieval system, Journal of Computing Science and Engineering, 6(2):168-177, June 2012.

[2] P. Ghosh, S. Antani, L. R. Long, G. R. Thoma, Review of medical image retrieval systems and future directions, 24th International Symposium on Computer-Based Medical Systems, pp. 1-6, Bristol, UK, June 2011.

[3] H. Müller, A. G. S. Herrera, J. Kalpathy-Cramer, D. Demner-Fushman, S. Antani, I. Eggel, Overview of the ImageCLEF 2012 Medical Image Retrieval and Classification Tasks. CLEF (Online Working Notes/Labs/Workshop) 2012.

[4] M. S. Simpson, D. You, M. M. Rahman, D. Demner-Fushman, S. Antani, G. R. Thoma, ITI's participation in the ImageCLEF 2012 Medical Retrieval and Classification Tasks. *CLEF 2012 Working Notes, Rome, Italy, September 2012.*

[5] V. Spitzer, M. J. Ackerman, A. L. Scherzinger, and D. Whitlock, The visible human male: A technical report, Journal of the American Medical Informatics Association, 3:118-130, 1996.

[6] P. Howarth, S. Ruger, Robust texture features for still image retrieval, IEEE Proceedings of Vision, Image and Signal Processing, 152(6): 868-874, December 2005.

[7] S.A. Chatzichristofis, Y.S. Boutalis, CEDD: Color and edge directivity descriptor: A compact descriptor for image indexing and retrieval, In: Gasteratos, A., Vincze, M., Tsotsos, J.K. (eds.) Proceedings of the 6th International Conference on Computer Vision Systems. Lecture Notes in Computer Science, 5008:312-322, Springer- Verlag Berlin Heidelberg, 2008.

[8] S.A. Chatzichristofis, Y.S. Boutalis, FCTH: Fuzzy color and texture histogram: A low level feature for accurate image retrieval, In: Proceedings of the 9th International Workshop on Image Analysis for Multimedia Interactive Services, 191-196, 2008.

[9] R. M. Haralick, K. Shanmugam, and I. Dinstein, Textural features for image classification, IEEE Trans. on Systems, Man, and Cybernetics, SMC-3(6): 610-621, November 1973

[10] M. Stricker, and M. Orengo, Similarity of color images, In SPIE Conference on Storage and Retrieval for Image and Video Databases III, 2420:381-392, Feb. 1995.

[11] M. Lux, Caliph & Emir: MPEG-7 photo annotation and retrieval, Proceedings of the seventeen ACM international conference on Multimedia, 925-926, 2009, Beijing, China.

[12] D.G. Lowe, Distinctive image features from scale-invariant keypoints, International Journal of Computer Vision, 60 (2): 91–110, 2004.

[13] M.M. Rahman, S.K. Antani, G.R. Thoma, Biomedical CBIR using "bag of keypoints" in a modified inverted index, 24th International Symposium on Computer-Based Medical Systems (CBMS), 1-6, June 2011.

[14] G. Zhao, M. Pietikäinen, Dynamic texture recognition using local binary patterns with an application to facial expressions, IEEE Trans. Pattern Analysis and Machine Intelligence, 29(6):915-928, 2007.

[15] T. Menp, The local binary pattern approach to texture analysis: extensions and applications, PhD Thesis, University of Oulu, 2003

[16] G.N. Srinivasan, G. Shobha, Statistical texture analysis, Proceedings of World Acad. Sci. Eng. Technol., 36:1264–1269, 2008.

[17] L. C. Molina, L. Belanche, A. Nebot, Feature selection algorithms: a survey and experimental evaluation, Proceedings of IEEE International Conference on Data Mining, 306- 313, 2002

[18] C. Hsu, C. Lin, A comparison of methods for multiclass support vector machines, IEEE Transactions on Neural Networks, 13(2):415-425, Mar 2002.

[19] J. Platt, Sequential Minimal Optimization: A Fast Algorithm for Training Support Vector Machines, 1998. http://research.microsoft.com/apps/pubs/default.aspx?id=69644