# A Content-Based Retrieval System for Endoscopic Images

Shunren Xia, Weirong Mo, and Zanchao Zhang

Key Lab of BME of Ministry of Education, Zhejiang
University, Hangzhou 310027, P. R. China.

srxia@mail.bme.zju.edu.cn

## Abstract

Background: Content-based medical image retrieval is now getting more and more attention in the world, a feasible and efficient retrieving algorithm for clinical endoscopic images is urgently required. Methods: Based on the study of single feature image retrieving techniques, including color clustering, color texture and shape, a new retrieving method with multi-features fusion and relevance feedback is proposed to retrieve the desired endoscopic images. Results: A prototype system is set up to evaluate the proposed method's performance and some evaluating parameters such as the retrieval precision & recall, statistical average position of top 5 most similar image on various features, etc. are therefore given. Conclusions: The algorithm with multi-features fusion and relevance feedback gets more accurate and quicker retrieving capability than the one with single feature image retrieving technique due to its flexible feature combination and interactive relevance feedback.

**Keyword**: Endoscopic images; multi-features fusion; relevance feedback; content-based image. retrieval.

## I. Introduction

CBIR [1] (Content-Based Image Retrieval) is a term that describes the process of retrieving desired images from a large collection on the basis of different visual or semantic features (color, texture, shape and semantic, etc.) that can be automatically extracted from the images themselves. The most frequent query paradigm is query by example. Comparing with manual retrieval work, CBIR shortens the retrieval time greatly. It is helpful in CAD (Computer-Aided Diagnosis) since doctors can refer to previous diagnosed cases before making their final decision when confused by clinical difficulties. Some well known commercial systems such as QBIC [2] and research system such as Photobook [3] are available now. On the request of CBIR in medical field, only a few prototypes [4, 5] have been implemented and evaluated and most of them concentrate on specified imaging fields. For example, ASSERT [6] is a typical CBIR application on high resolution lung CT images specially. Until now there is no practical CBIR for clinical endoscopic images yet, this paper presents a method by using multi-feature fusion and relevance feedback to retrieve endoscopic images and the experimental results are given at the same time.

## II. CBIR Technique Based on Multi-feature Fusion

As we have known, low-level visual features such as color, texture and shape, etc. can be used to characterize image's visual and statistical features. According to preliminary statistical results of endoscopic images' color distribution in HSV color space and characteristic of human perception, uneven quantization is adopted to characterize each H, S, V component as follows [7]:

Component H is unevenly quantified into 6 levels with two levels for 0-60, two levels for 330-360 and two levels for the rest parts respectively; Component V is evenly quantified into eight levels and S into 4 levels. Therefore endoscopic images' color feature can be represented by a clustered 192bin (6x8x4) histogram. To compensate the missing color information when using Haralick's gray co-occurrence matrix, a kind of color co-occurrence modal is set up here. For each image in HSV color space, cluster it to 192bin just like what color feature extraction does. When constructing model, four different angles ($0^o$, $45^o$, $90^o$, $135^o$) with three different pixel intervals (1, 3, 9) are considered. So total 12 color co-occurrence matrices with size of 192x192 are set up. Then two statistical features i. e. mean and standard derivation of the 4 most efficient textural parameter, namely contrast, energy, correlation and entropy, are extracted from each matrix to construct a textural feature vector $[\mu_{CON}, \sigma_{CON}, \mu_{ENG}, \sigma_{ENG}, \mu_{COR}, \sigma_{COR}, \mu_{ENT}, \sigma_{ENT}]$. After inner-feature Gaussian normalization [8], every 8bin vector has identical range and physical meaning. As for shape features, it is calculated from endoscopic image's ROI (Region of Interest), which is generally corresponding to some clinical abnormality and is outlined by the physician. Twelve ROI's shape features such as area, X's and Y's eccentricity, the ratio of long axis to short axis, the degree of circularity and the first three central moments of X's and Y's axis, etc. are chosen based on mutually outlined ROI after removing background [9]. Similarly, inner-feature Gaussian normalization is used. The detailed algorithm refers to [7].

Although using only one feature can implement CBEIR function completely, the performance is somewhat unacceptable due to the limitation of single feature. For example, color histogram, with robustness to the transformation of shift, rotation and scale, is sensitive to the variation of illumination. Its neglect of spatial information sometimes makes the retrieval results differ from the expectation. Like color histogram, texture and shape have its own limitations respectively. In a word, single feature cannot contain all visual properties of the image to be considered. The better solution is to merge multiple features by adjusting the weights of each feature component automatically with respect to special application requirement.

Given the retrieval images sorted by similarity according to CBEIR technique with single feature described above, rearrange them according to the comprehensive similarity measure [3] and return the most similar $n$ images to the user. Certainly, it is necessary to perform inter-feature Gaussian normalization [8] within single feature before constructing the comprehensive similarity measure because the sorted images retrieved by different single features have different similarity weight when fusing different features.

After inter-feature Gaussian normalization, the probability of distance between example image and query image limited to within [0, 1] will be more than 99%. Then different weights can be reliably assigned to different single features. For example, color feature may be more efficient than other two features according to the preliminary experiment [7]. Users can express their concern by adjusting the weight for different single features until the retrieval results are acceptable. All images given can be characterized by using $N$ features, the normalized, weighted similarity between example $Q$ and query image $Q_i$ can be defined as the expression (1).

$$Sim(Q,Q_i)=(\sum_{k=1}^{N} w_k \times s_k)/\sum_{k=1}^{N} s_k \quad . \tag{1}$$

Where $w_k$ is the weight of the $k$ th feature, $s_k$ is the normalized similarity of the $k$ th feature and $Sim(Q,Q_i)$ is the combined similarity. Practically, the weights of color, texture and shape can be assigned by human-machine interaction according to user's concern.

## III. CBIR Technique Based on Relevance Feedback

The retrieval results based on single low-level visual features above cannot always satisfy with user's semantic requirements [1]. So the interactive relevance feedback is introduced to guide the system to search for desirable result.

For every image in the retrieval results, the user judges whether or not it is similar to the sample image and returns the judgment to the system. The system analyzes user's judgment and then executes next round of retrieval procedure by adjusting the related parameters until the results match user's requirement properly. The drawback of feedback algorithm is time-consuming. Although the method of clustering [10] by modifying the distribution of geometry center has been proved to get better retrieval performance, it requires a lot of iterations to achieve the convergence. Using artificial network [11] can bridge the semantic gap somewhat but its computational complexity is also high. In order to reduce iterative times, a modified feedback algorithm is given below:

Given $n$ is the number of the most similar images, $R^i=\{R_1^i,R_2^i,\ldots,R_j^i,\ldots R_n^i\}$ stands for the retrieved image set sorted by single feature $f_i$ and $R=\{R_1,R_2,\ldots,R_j,\ldots R_n\}$ is the merged image set sorted by the combined similarity measure. For every image $R_j$ in $R$, $Cov_j$ represents the similarity coefficient judged by the user and $w_i$ represents similarity weight. $Cov_j$ and $w_i$ are defined in (2) and (3).

$$Cov_j=\begin{cases}1, & if\ similar \quad .\\ 0, & if\ uncertain \quad .\\ -1, & if\ dissimilar \quad .\end{cases} \tag{2}$$

$$w_i=\begin{cases}\sum_{j=1}^{n} Cov_j^i, & \sum_{j=1}^{n} Cov_j^i>0 \quad .\\ 0, & \sum_{j=1}^{n} Cov_j^i\le 0 \quad .\end{cases} \tag{3}$$

Where $\sum_{j=1}^{n} Cov_j^i$ stands for the contribution of the feature $f_i$. The greater the $\sum_{j=1}^{n} Cov_j^i$ is, the more concern about feature $f_i$ does. After being revised, $w_i$ will be put into next iteration until the result satisfies with the user's demands.

## IV. Experimental Results and Discussions

In order to evaluate the algorithm in this paper, a prototype system, which includes 800 standard clinical endoscopic images with resolution of 256x256x24bits, is set up. The system can retrieve images not only by single low-level features but also by any combination of single features automatically and manually. The precision and recall are adopted here to evaluate retrieval performance [12] and defined as below.

$$prcision = \frac{number\ of\ relevant\ items\ retrieved}{number\ of\ all\ items\ retrieved} \qquad (4)$$

$$recall = \frac{number\ of\ relevant\ items\ retrieved}{number\ of\ all\ relevant\ items} \qquad (5)$$

To quantitively evaluate precision (expressed as *P* for short) and recall (expressed as *Q* for short) of single feature, five independent experts with strong clinical background are chosen to judge whether the image is similar or not and the decision is dominated by the majority of vote. Meanwhile, four different images denoted as *A*, *B*, *C* and *D* are randomly selected from the endoscopic images database. Table 1 lists the precision and recall based on single low-level features and feature fusion respectively. Table 2 lists the precision and recall after different feedback times.

From table 1 and table 2, it can be easily found: 1) The precision based on color (HSV color space) clustering is higher and the recall is lower because color clustering according to histogram neglects the spatial information. 2) The precision based on texture is lower than the one based on color clustering in a whole. However, the precision and recall of the most similar 10 images are better than the one of the following images. This means texture is particularly good at charactering image's local region when user concerns more about local texture. 3) The precision based on shape is also lower but recall is higher because locating ROI can narrow the searching range with the same or similar shape characters. 4) The precision and recall based on the fusion of color, texture and shape (weights equals to 5, 1 and 1 respectively in our experiment) are both higher than any one based on single visual features. 5) The results after 2 or 3 times of feedback are clearly better than what only 1 time feedback does. But additional feedback after 3 times cannot improve the retrieval performance substantially. So the iteration can be ended after some limited times.

Table 1. The precision (*P*) & recall (*R*) based on different feature algorithms

| Feature | Image *A* | | Image *B* | | Image *C* | | Image *D* | | Average Value | |
|---|---|---|---|---|---|---|---|---|---|---|
| | *P* | *R* | *P* | *R* | *P* | *R* | *P* | *R* | *P* | *R* |
| Color | 0.84 | 0.57 | 0.78 | 0.58 | 0.65 | 0.55 | 0.91 | 0.48 | 0.795 | 0.545 |
| Texture | 0.25 | 0.63 | 0.15 | 0.59 | 0.11 | 0.65 | 0.19 | 0.65 | 0.175 | 0.630 |
| Shape | 0.25 | 0.65 | 0.22 | 0.66 | 0.16 | 0.68 | 0.18 | 0.68 | 0.202 | 0.668 |
| Fusion | 0.90 | 0.50 | 0.77 | 0.54 | 0.85 | 0.49 | 0.80 | 0.52 | 0.830 | 0.511 |

Table 2. The precision (*P*) & recall (*R*) based on relevance feedback algorithms

| Feedback times | Image *A* | | Image *B* | | Image *C* | | Image *D* | | Average Value | |
|---|---|---|---|---|---|---|---|---|---|---|
| | *P* | *R* | *P* | *R* | *P* | *R* | *P* | *R* | *P* | *R* |
| 1 | 0.62 | 0.42 | 0.58 | 0.46 | 0.30 | 0.55 | 0.50 | 0.49 | 0.500 | 0.480 |
| 2 | 0.75 | 0.60 | 0.60 | 0.62 | 0.55 | 0.66 | 0.55 | 0.66 | 0.613 | 0.635 |
| 3 | 0.75 | 0.72 | 0.65 | 0.78 | 0.65 | 0.79 | 0.65 | 0.72 | 0.675 | 0.750 |
| 4 | 0.77 | 0.78 | 0.69 | 0.80 | 0.68 | 0.78 | 0.68 | 0.78 | 0.705 | 0.785 |

Admittedly, the key of applying multi-feature fusion is to modify low-level feature's weights. When inappropriate weights are passed into the iteration, system has to spend more time to converge. As the convergence approaches the expectation, the more iterations do not improve accuracy significantly.

Average rank represents the compactness of result set [12]. Table 3 lists the average rank of the most four similar images after three times of feedback. The result via multi-feature fusion and relevance feedback is much closer to the user's requirements.

**Table 3.** Statistical average position of top 5 most similar image on various features

| Features | Color | Texture | Shape | Fusion | Feedback |
|---|---|---|---|---|---|
| 1st | 1.12 | 1.20 | 1.20 | 1.00 | 1.00 |
| 2nd | 4.60 | 5.33 | 6.10 | 2.75 | 2.37 |
| 3rd | 6.80 | 7.17 | 7.20 | 4.01 | 3.63 |
| 4th | 7.60 | 13.00 | 8.10 | 5.25 | 4.37 |

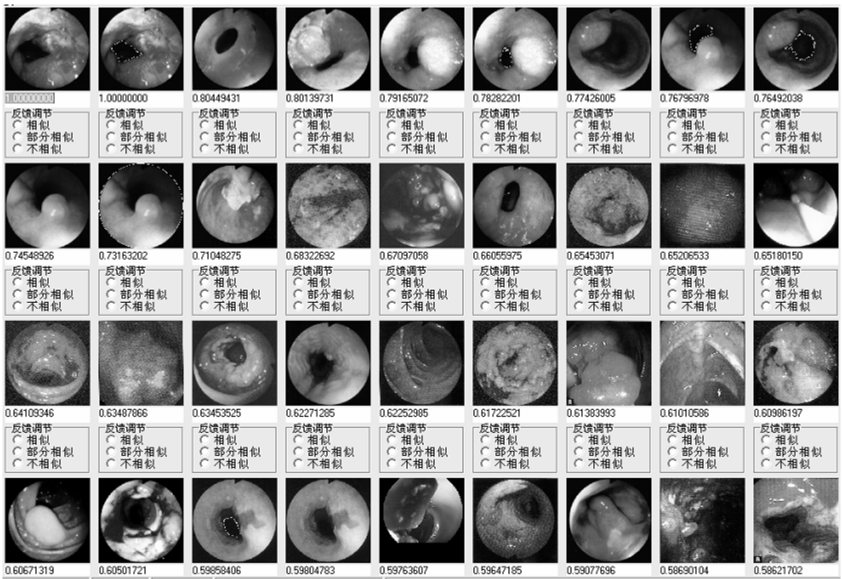


**Fig. 1.** The interface of endoscopic images retrieval system

Figure 1 outlines the prototype system. The top-left most is the example image. The subsequent images are the retrieved images and are listed from left to right, from top to bottom by their similarity.

## V. Conclusions

According to the results illustrated above, the utilization of multi-feature and relative feedback greatly improves the CBIR's robustness, flexibility and accuracy. In the next step, we will further compare the retrieval result and the retrieval speed in a larger image collection. The methodology in this paper can also be embedded into the PACS, RIS, CAD and other clinical image management information systems and we are sure that the technique will enable the system to provide a new image retrieval tool.

## Acknowledgements

The authors are grateful to acknowledge the support of National Natural Science Foundation of China (No. 60272029) and Provincial Natural Science Foundation of Zhejiang (No. M603227 ) in the research. The authors also much appreciate the anonymous reviewers for their generous and valuable comments on this paper.

## References


[1] Smeulders, M., Worring, M., Santini: Content-based Image Retrieval at The End of Early Years [J]. IEEE Transaction on Pattern Analysis and Machine Intelligence, Vol.22 (12), (2000) 1349-1380;

[2] Flickner, M., Sawhney, H., Niblack, W., Ashley, J., Huang, Q., Dom. B.: Query by Image and Video Content: the QBIC System [J]. IEEE Computer, Vol.28 (9), (1995) 23-32.

[3] Pentland, A., Picard, R., Sclaroff, S.: Photobook: Tools for Content-based Manipulation of Image Database [J]. International Journal of Computer, Vol. 18(3), (1996) 233-254.

[4] Müller, H., Michoux, N., Bandon, D.: A Review of Content-based Image Retrieval Systems in Medical Applications-clinical Benefits and Future Directions [J]. International Journal of Medical information, Vol. 73(1), (2004) 1-23.

[5] Liu, Y., Lazar, A., Rothfus, W.E., Buzoiano, M.,Kanade, T.: Classification-driven feature space reduction for semantic-based neuroimage retrieval. VISIM. 2001.

[6] Shyu, C.R., Brodley, C.E., Kak, A.C., Kosaka, A., Aisen, A.M., Broderick, L.S.: ASSERT: A Physician-in-the-loop Content-based Retrieval System for HRCT Image Databases [J]. Computer Vision and Image Understanding (special issue on content-based access for image and video libraries), Vol. 75(1), (1999) 111-132.

[7] Shunren, X., Weirong, M., Xichao, Ch.: An Endoscopic Image Retrieval System Based on Color Clustering Method [C]. SPIE Conference on Multi-spectral Image Processing and Pattern Recognition. (2003) 410-413.

[8] Ortega, M.: Supporting similarity queries in MARS[C]. In: the 5th ACM Conference. On Multimedia, (1997) 403-413.

[9] Shunren X., Krishnan S. M., Tjoa M. P., Goh Peter M. Y.: A Novel Methodology forExtracting Colon's Lumen from Colonoscopic Images, Journal of Systemics, Cybernetics and Informatics [J], Vol. 1(2), (2003)7-12.

[10] Tomas, M., Lehmann, al, Berthold, Wein B.: Content-based Image Retrieval in Medical Application: A Novel Multi-step Approach [C]. SPIE. (2000) 312-320.

[11] Müller, H., Squire, D.M., Pun, T.: Learning from user behavior in image retrieval: application of the market basket analysis, Int. J. Computer Vision. Vol. 56, (2004) 65–77.

[12] Henning, M., Antoine, R., Jean-Paul, V.: Comparing Feature Sets for content-based Image Retrieval in a Medical Case Database [C]. SPIE Conference on Medical Imaging. (2004) 99-109.


Shunren Xia received the Ph. D. degree in signal & information from Southeast University, China in 1992. He is a now Professor of Biomedical Engineering at Zhejiang University, China. His research interests have included the content-based image retrieval, computer-aided diagnosis on medical images and medical image processing, etc.

Weirong Mo received the Master degree in BME in Zhejiang University, China in 2005. He is currently majoring in his Ph. D. degree in National University of Singapore,

Zanchao Zhang received his Bachelor degree in BME in Zhejiang University, China in 2001. He is now an undergraduate student in Zhejiang University.