NCTU_DBLAB@ImageCLEFmed 2005: Medical Image Retrieval Task

Pei-Cheng Cheng[1], Been-Chian Chien[2], Hao-Ren Ke[3], and Wei-Pang Yang[14]

[1] Department of Computer & Information Science, National Chiao Tung University,
1001 Ta Hsueh Rd., Hsinchu, TAIWAN 30050, R.O.C.
{cpc, wpyang}@cis.nctu.edu.tw
[2] Department of Computer Science and Information Engineering,
National University of Tainan
33, Sec. 2, Su-Lin Street, Tainan, 700, Taiwan, R.O.C.
bcchien@mail.nutn.edu.tw
[3] Institute of Information Management and University Library, National Chiao Tung University,
1001 Ta Hsueh Rd., Hsinchu, TAIWAN 30050, R.O.C.
claven@lib.nctu.edu.tw
[4] Department of Information Management, National Dong Hwa University
1, Sec. 2,Da Hsueh Rd., Shou-Feng, Hualien,Taiwan, R.O.C.
wpyang@mail.ndhu.edu.tw

Abstract

In this article, we describe the used technologies and experimental results for the medical retrieval task at ImageCLEF 2005. The topics of competition this year contain both semantic queries and visual queries. The content-based approach containing four image features and the text-based approach using word expansion are developed to accomplish the mission. The experimental results show that the text-based approach has higher precision rate than content-based approach. Further, the results of combining both the content-based and text-based approaches are better than those using only one of the approaches. We summarize that the consideration on the image of visual queries can provide more human semantic perception and improve the efficiency for medical image retrieval.

ACM Categories and Subject Descriptors: Information Storage and Retrieval; Artificial Intelligence;

Free Keywords: Medical Image retrieval; content based image retrieval;

1. Introduction

In this paper, we present the research experience of the NCTU group at ImageCLEFmed 2005. The dataset of medical image retrieval task contains about 50,000 images totally from the Casimage, MIR, PEIR, and PathoPIC datasets. Each image of collection contains annotations in XML format. The majority of the annotations are in English but a significant number is also in French and German, with a few cases that do not contain any annotation at all. The queries of this task were formulated with example images and a short textual description explaining the research goal.

The participants were requested to accomplish the task in either fully automatic retrieval or retrieval with manual feedback. The used query information can consider the example images or the textual description only, or combine the images and textual together. The main purpose is to evaluate the retrieval of medical images from heterogeneous and multilingual document collections containing images as well as text.

For handling the medical image retrieval task, the NCTU group gives two primitives: content-based approach and text-based approach. The combination of the two approaches using similarity weight was also discussed. The content-based approach in this work uses four image features, **Facade scale image feature, Gray Histogram layout**, **Coherence Moment and Color histogram**, extracted from the images directly. The text-based approach processes the annotations based on the vector space model and word expansion using Wordnet[Miller95]. The mixed retrieval of visual and textual is done by combining the content-based approach and the text-approach with different weights adjusting.

In Section 2, the image features for the content-based approach are described. Section 3 illustrates the text-based approach that processes the multilingual annotations and translation. Section 4 describes our submissions at ImageCLEF 2005 and the ranking results. Here, we give an explanation and a discussion on our experimental results. Finally, section 5 provides concluding remarks and future direction are for medical image retrieval.

2. Image features

This section describes the features used in the paper for the ImageCLEF 2005 evaluation. In an image retrieval system, image features are extracted from pixels of image. For fast response time, image features used must be concise, and for precision, image features is used must contain meaningful information to represent the image itself. Image feature can retrieval the image in visual. In this paper we adopt several image features we have proposed [Cheng 04] which image features have good performance in medical image application.

In designing the image features, to emphasize the contrast of an image and handle images with less illuminative influence, we normalize the value of a pixel before quantization. In [Cheng 04] we proposed a relative normalization method. First, we cluster a whole image into four clusters by the K-means clustering method [Han01]. We sort the four clusters ascendant according to their mean values. After clustering, we shift the mean of the first cluster to value 50 and the fourth cluster to 200; then, each pixel in a cluster is multiplied by a relative weight to normalize. Let m_{c1} be the mean value of cluster 1 and m_{c4} is the mean value of cluster 4. The normalization formula of pixel p(x,y) is defined in Eq. (1).

$$p(x,y)_{normal} = (p(x,y) - (m_{c1} - 50)) \times \frac{200}{(m_{c4} - m_{c1})} \quad . \tag{1}$$

After normalization, we scale an image into common 128*128 pixels and extract image features.

2.1 Facade scale image feature

The pixel values of an image are trivial and straight-forward features. For computational efficiency, images are always scaled to a common small size and compared using the Euclidean distance. [Keysers04] has shown that in the optical character recognition and medical image retrieval based on facade image features have obtained excellent results. In this work we scale down an image into 8×8 pixels to form a 64 feature vectors as facade scale image feature.

2.2 Gray Histogram layout

Histogram [Swain91] is a prime image feature for image information retrieval. Histogram method is invariant in image rotation, it is easy to implement and have good result in color image indexing. Because radiotherapy medical image only consists of gray level, the spatial relationship becomes very important. Medical images always contain particular anatomic regions (lung, liver, head, and so on); therefore, similar images have similar spatial structures. We divide an image into nine sections and calculate their histogram respectively. After normalization, the gray values are quantized into 16 levels for computational efficiency.

In the gray histogram, the gray value may be quantized into several bins to improve the similarity between adjacent bins. We set an interval range δ to extend the similarity of each gray value. The histogram layout feature estimates the probability of each gray level that appears in a particular area. The probability equation is defined in Eq. (2), where δ is set to 10, where p_j is a pixel of the image, and m is the total number of pixels. The gray histogram layout of an image has a total of 144 bins.

$$h_{c_i}(I) = \frac{\sum_{j=1}^{m} \frac{[p_j - \frac{\delta}{2}, p_j + \frac{\delta}{2}] \cap c_i}{\delta}}{m} \quad . \tag{2}$$

2.3 Coherence Moment

One of the problems to design an image representation is the semantic gap. The state-of-the-art technology still cannot reliably identify objects. The coherence moment feature attempts to describe the features from the human's viewpoint in order to reduce the semantic gap.

We cluster an image into four classes by the K-means algorithm. After clustering an image into four classes, we calculate the number of pixels (COH_κ), mean value of gray value (COH_μ) and standard variance of gray value (COH_ρ) in each class. For each class, we group connected pixels into eight directions as an object. If an object is bigger than 5% of the whole image, we denote it as a big object; otherwise it is a small object. We count how many big objects (COH_o) and small objects (COH_ν) are in each class, and use COH_o and COH_ν as parts of image features.

Since we intend to know how the reciprocal effects among pixels, so we use the smooth method on the image. If the spatial distribution of pixels of two images is similar, they will also be similar after smoothing. If their spatial distributions are quite different, they may have a different result after smoothing. After smoothing, we cluster an image into four classes and calculate the number of big objects (COH_τ) and small objects (COH_ω). Each pixel will be influenced by its neighboring pixels. Two close objects of the same class may be merged into one object. Then, we can analyze the variation between the two images before and after smoothing. The coherence moment of each class form a seven-feature vector, (COH_κ, COH_μ, COH_ρ, COH_o, COH_ν, COH_τ, COH_ω). The coherence moment of an image is a 56-feature vector that combines the coherence moments of the four classes.

2.4 Color histogram Features

Color histogram [Swain91] is a basic method and has good performance for representing image content. The color histogram method gathers statistics about the proportion of each color as the signature of an image. In our work, the colors of an image are represented in the HSV (Hue/ Saturation/ Value) space, which is believed closer to human perception than other models, such as RGB (Red/ Green/ Blue) or CMY (Cyan/ Magenta/ Yellow). We quantize the HSV space into 18 hues, 2 saturations, and 4 values, with additional 4 levels of gray values; as a result, there are a total of 148 (i.e., 18×2×4+4) bins. Let C ($|C| = m$) a set of colors (i.e., 148 bins), P_I (Q_I) is represented as Eq. (3), which models the color histogram $H(P_I)$ ($H(Q_I)$) as a vector, in which each bucket h_{c_i} counts the ratio of pixels of P_I (Q_I) in color c_i.

$$P_I = < h_{c_1}(P_I),\ldots,h_{c_m}(P_I) >, \quad Q_I = < h_{c_1}(Q_I),\ldots,h_{c_m}(Q_I) > \tag{3}$$

In many previous studies, each pixel is only assigned a single color. Consider the following situation: I_1, I_2 are two images, all pixels of I_1 and I_2 fall into c_i and c_{i+1} respectively; I_1 and I_2 are indeed similar to each other, but the similarity computed by the color histogram will regard them as different images. To address the problem, we set an interval range δ to extend the color of each pixel and introduce the idea of a partial pixel as shown in Eq.(4),

$$h_{c_i}(P_I) = \frac{\sum_{p \in P_I} \frac{|\alpha_p - \beta_p|}{\delta}}{|P_I|} \tag{4}$$

Let c_{i-1}, c_i, and c_{i+1} stand for a color bin, a solid line indicate the boundary of c_i, p is the value of a pixel, $[p - \frac{\delta}{2}, p + \frac{\delta}{2}]$ denotes the interval range δ, the shadow part, $[\alpha_p, \beta_p]$, is the intersection of $[p - \frac{\delta}{2}, p + \frac{\delta}{2}]$ and c_i. The contributions of the pixel to c_i and c_{i-1} are computed as $\frac{|\alpha_p - \beta_p|}{\delta}$ and $\frac{|(p - \delta/2) - \alpha_p|}{\delta}$, respectively. It is clear that a pixel has its contributions not only to c_i but also to its neighboring bins.

Using the modified color histogram, the similarity of two color images q and d is defined in Eq. (5):

$$\text{SIMcolor(H(q), H(d))} = \frac{\text{H(q)} \cap \text{H(d)}}{|\text{H(q)}|} = \frac{\sum_{i=1}^{n} \min(h_i(q), h_i(d))}{\sum_{i=1}^{n} h_i(q)} \ . \tag{5}$$

2.5 color/gray feature

The medical image collection of the ImageCLEF 2005 evaluation contains gray and color images. In color images, users are usually attracted by the change of colors more than the positions of objects. Thus, the effective feature in query a color image is different from query a gray image. The image have an obvious feature is whether the image is color or gray value image. When the user queries an image by example, the system first determines whether the example is color or grayscale. We calculate the color histogram, if the four bins of gray values occupy more than 80% of the whole image, we decide that the query image is gray; otherwise it is color. If the input is a color image, then we set the weight parameter denoted by “C”; if the query image is detected gray valued image, we use the weight parameter denote by “G” as show in Table 1 and Table 2.

3. Textual Vector Representation

In the ImageCLEFmed collections of annotations are in English, French and German. The overall multilingual search process is show in Fig. 1. Given an initial query Q, the system performs the cross-language retrieval, and returns a set of relevant documents to user. We use the representation expressing a query as a vector in the vector space model [Salton83]. The Textual Vector Representation is defined as following. Let W ($|W| = n$) be the set of significant keywords in the corpus. For a document D, its textual vector representation (i.e., D_T) is defined as Eq. (6),

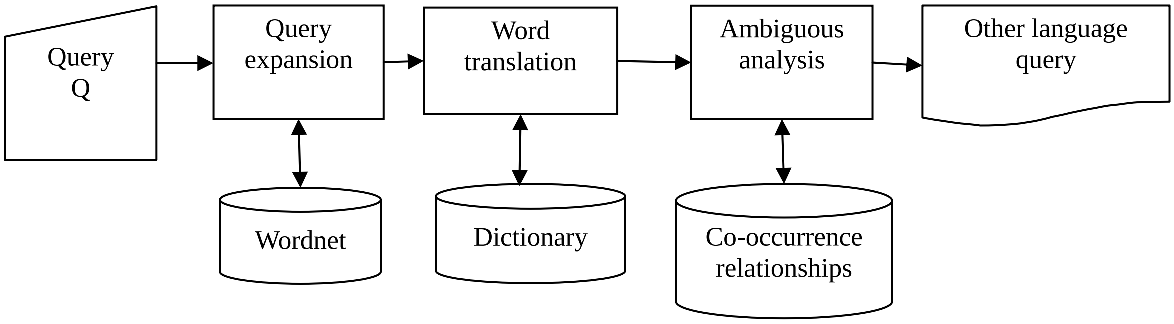

Figure 1: Text based multilingual query translation flowchart

$$D_T = < w_{t_1}(D_T),..., w_{t_n}(D_T) > \quad (6)$$

where the n dimensions indicate the weighting of a keyword t_i in D_T, which is measured by TF-IDF [Salton83], as computed in Eq.(7);

$$w_{t_i}(D_T) = \frac{tf_{t_i,D_T}}{\max tf} \times \log \frac{N}{n_{t_i}} \quad (7)$$

In Eq.(7), $\frac{tf_{t_i,D_T}}{\max tf}$ stands for the normalized frequency of t_i in D_T, max tf is the maximum number of occurrences of any keyword in D_T, N indicates the number of documents in the corpus, and n_{t_i} denotes the number of documents in whose caption t_i appears.

In the above, we introduce the textual vector representation for documents. As for q query Q, one problem is that since Q_T is given in English, it is necessary to translate Q_T into French and German, which is the languages used in the document collection.

A short query usually cannot cover as many useful search terms as possible because of the lack of sufficient words. We perform the query expansion process to add new terms to the original query. The additional search terms is taken from a thesaurus – WordNet [Miller95]. For each expansion English term, it is then translated into one or several corresponding French and German words by looking it up in a dictionary[1].

Assume $AfterExpansion(Q_T) = \{e_1,...,e_h\}$ is the set of all English words obtained after query expansion and query translation, it is obvious that $AfterExpansion(Q_T)$ may contain a lot of words which are not correct translations or useful search terms. To resolve the translation ambiguity problem, we define *word co-occurrence relationships* to determine final query terms. If the co-occurrence frequency of e_i and e_j in the corpus is greater than a predefined threshold, both e_i and e_j are regarded as useful search terms.

So far, we have a set of search terms, $AfterDisambiguity(Q_T)$, which is presented as Eq.(8),

$$AfterDisambiguity(Q_T) = \{e_i, e_j \mid e_i, e_j \in AfterTranslation(Q_T) \\ \& \, e_i, e_j \text{ have a significant co-occurrence}\} \quad (8)$$

After giving the definition of $AfterDisambiguity(Q_T)$, for a query Q, its textual vector representation (i.e., Q_T) is defined in Eq. (9),

$$Q_T = < w_{t_1}(Q_T),..., w_{t_n}(Q_T) > \quad (9)$$

where $w_{t_i}(Q_T)$ is the weighting of a keyword t_i in Q_T, which is measured as Eq.

(10), $w_{c_i}(Q_T)$ indicates whether there exists an $e_j \in AfterDisambiguity(Q_T)$.

In Eq.

---

[1] http://www.freelang.net/

(10), W is the set of significant keywords as defined before, $\frac{tf_{t_i,Q_T}}{\max tf}$ stands for the normalized frequency of t_i in $AfterDisambiguity(Q_T)$, maxtf is the maximum number of occurrences of any keyword in $AfterDisambiguity(Q_T)$, N indicates the number of images in the corpus, and n_{t_i} denotes the number of images in whose caption t_i appears.

$$w_{t_i}(Q_T) = \left\{ \frac{tf_{t_i,Q_T}}{\max tf} \times \log \frac{N}{n_{t_i}} \right. \qquad (10)$$

4. Submissions to the ImageCLEF 2005 Evaluation

The entire ImageCLEFmed *library* consists of multiple collections (e.g., Casimage, PEIR, MIR, PathoPIC). Each *collection* is organized into cases that represent a group of related images and annotations. Each *case* consists of a group of images and an optional annotation. Each *image* is part of a case and has optional associated annotations, which consist metadata (e.g., HEAL tagging), and/or a textual annotation.

In the practice for medical images retrieval system, the doctor query by semantic sentence that is more convenience. After textual query, the system retrieves related images by annotation for user to browse. The user can further more rank the images in visually. Combine the textual and visually features will help the user find the desire image more precise. In ImageCLEF 2005, medical image retrieval task contains 25 queries for evaluation, the queries mixed visual image and semantic textual to retrieve desire images. The visually queries use image example to find similar images, each topic contain at least one image example. The semantic textual queries allow user query by a sentence, which some semantic concept are hard to derive from images directly. The goal of this task is to examine how the visual feature can improve the query result.

All submissions of participants in this task were classified into automatic runs and manual runs. The automatic runs means that the system at the query process without human manual intervened. In the automatic category, the methods can be classified into three sub-categories: Text only, Visual only and Mixed retrieval (visual and textual) according to the feature used. The category "Text only" means that systems use textual feature only to retrieve relevant images. "Visual only" category means that systems only use visual image feature without combine textual annotation to retrieve similar images. Mixed retrieval means the systems combine the visual and textual feature to retrieve images.

In this task, we have submitted ten runs for the mixed retrieval of automatic runs and six runs for the visual only of automatic runs. In the content-based approach, we combine four proposed image features by weighted adjusting to retrieve related images. The weight of features we set at the system initial and do not have any further user intervention while query is processing. Table 1 lists the query result of visual only runs and the setting weight of four image features. Table 2 lists the result of mixed retrieval runs and the setting weight of image features and textual features. The different of each runs is the weighted setting of features.

The query topics contain color and gray images. We first examine the queries image is color or gray image by **color/gray feature**. According to the image is color or gray set different weight for image features. In the Table 1, "C" denotes that query image is color image and "G" denotes that query image is gray image. We submit six runs for visual only category. The run, "nctu_visual_auto_a8", has the better result in our experiment. The weight of each feature are set equal, it means that four image features have the same importance. The result also shows that visual only approach has a bottleneck because the query topics contain semantic queries.

Table 1: The query result of visual only runs and the weight of visual image features

Submission runs	The weight of Image features								Result	
Image features	Coherence		Gray HIS		Color HIS		Facade		MAP	Rank of runs
detected color or gray	C	G	C	G	C	G	C	G		
nctu_visual_auto_a1	0.3	0.2	0.3	0.5	1	0.2	1	1	0.0628	14
nctu_visual_auto_a2	0.3	0.2	0.5	0.3	0.3	0.5	1	1	0.0649	10
nctu_visual_auto_a3	0.5	1	0.5	1	1	0.5	1	1	0.0661	8
nctu_visual_auto_a5	0.1	0.2	0.1	0.5	1	0.5	0.5	1	0.0631	13
nctu_visual_auto_a7	0.3	0.2	0.3	0.5	1	0.2	1	0.5	0.0644	11
nctu_visual_auto_a8	1	1	1	1	1	1	1	1	0.0672	7

The setting weights of mixed runs and results are listed in the table 2. The result of run8, run9 and run10 illustrate that combine the visual and textual feature will get better results than single features. Run8 assume that the significant of visual and textual feature are equal. Run9 emphasizes the weight of visual features and Run10 emphasizes the weight of textual features. The result shows that text-based approach is better than content-based approach, but the content-based approach can improve the textual result.

Table 2: The result of mixed retrieval runs and the weight of visual image features and textual features

Submission runs	The weight of Image features													
Image features	Coherence		Gray HIS		Color HIS		Facade		visual		textual		MAP	Rank
Detected color or gray	C	G	C	G	C	G	C	G	C	G	C	G		
nctu_visual+Text_auto_1	0.3	0.2	0.3	0.5	1	0.2	1	1	1	1	0.8	0.1	0.2276	10
nctu_visual+Text_auto_2	0.3	0.2	0.5	0.3	0.3	0.5	1	1	1	1	0.8	0.1	0.2127	14
nctu_visual+Text_auto_3	0.5	1	0.5	1	1	0.5	1	1	1	1	0.8	0.1	0.2286	9
nctu_visual+Text_auto_4	0.3	0.2	0.3	0.5	1	0.2	1	1	1	1	1	0.2	0.2389	3
nctu_visual+Text_auto_5	0.1	0.2	0.1	0.5	1	0.5	0.5	1	1	1	0.8	0.1	0.2246	12
nctu_visual+Text_auto_6	0.3	0.2	0.3	0.5	1	0.2	1	1	1	1	1	1	0.2318	7
nctu_visual+Text_auto_7	0.3	0.2	0.3	0.5	1	0.2	1	0.5	1	1	0.8	0.1	0.2265	11
nctu_visual+Text_auto_8	1	1	1	1	1	1	1	1	1	1	1	1	0.2324	6
nctu_visual+Text_auto_9	1	1	1	1	1	1	1	1	1	1	0.1	0.1	0.0906	22
nctu_visual+Text_auto_10	1	1	1	1	1	1	1	1	0.1	0.1	1	1	0.1941	15

5. Conclusions and future work

ImageCLEF 2005 medical image retrieval task offers a good test platform to evaluate the ability of image retrieval technologies. There are totally 112 runs submitted for this task. The results of the evaluation show that the method we proposed is excellent. Our best result rank by MAP is 3, there are one better system than us.

In the experiment result, we find that content-based approach retrieving similar images rely on visual feature, which has less semantic expansion. The text-based approach has better performance than content-based approach. Combine the textual and visual features will get best result.

The results in the medical retrieval task show that weighted setting between the features is very important. The variation between different settings of weight is extreme. Suitable weight adjusting will improves the results.

References

[Manjunath96] B. S. Manjunath and W. Y. Ma, “Texture features for browsing and retrieval of large image data” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, (Special Issue on Digital Libraries), Vol. 18 (8), August 1996, pp. 837-842.

[Flickner95] M. Flickner, H. Sawhney, W. Niblack, J. Ashley, Q. Huang, B. Dom, M. Gorkani, J. Hafner, D. Lee, D. Petkovic, D. Steele, P. Yanker, Query by Image and Video Content: The QBIC system, *IEEE Computer* 28 (9) 1995, pp. 23-32.

[Carson99] C. Carson, M. Thomas, S. Belongie, J. M. Hellerstein, J. Malik, Blobworld: A system for region-based image indexing and retrieval, in: D. P. Huijsmans, A. W. M. Smeulders (Eds.), *Third International Conference On Visual Information Systems* (VISUAL' 99), no. 1614 in Lecture Notes in Computer Science, Springer{Verlag, Amsterdam, The Netherlands, 1999, pp. 509-516.

[Belongie98] S. Belongie, C. Carson, H. Greenspan, J. Malik, Color and texture based image segmentation using EM and its application to content-based image retrieval, in: *Proceedings of the International Conference on Computer Vision (ICCV'98)*, Bombay, India, 1998, pp. 675-682.

[Squire99] D. M. Squire, W. Muller, H. Muller, and J. Raki. Content-Based Query of Image Databases, Inspirations from Text Retrieval: Inverted Files, Frequency-Based Weights and Relevance Feedback. *In Scandinavian Conference on Image Analysis*, Kangerlussuaq, Greenland, June 1999, pp. 143-149,.

[Wang01] J. Z. Wang, J. Li, and G. Wiederhold. SIMPLIcity: Semantics-Sensitive Integrated Matching for Picture Libraries. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 23(9), September 2001, pp. 947-963.

[Han01] J. Han and M. Kamber, Data Mining: Concepts and Techniques. *Academic Press*, San Diego, CA, USA, 2001.

[Keysers04] D. Keysers, W. Macherey, H. Ney, and J. Dahmen. Adaptation in Statistical Pattern Recognition using Tangent Vectors. *IEEE transactions on Pattern Analysis and Machine Intelligence*, 26(2), February 2004, pp. 269-274.

[Swain91] Swain M.J. and D. H. Ballard, “Color Indexing”, *International Journal of Computer Vision,* Vol. 7, 1991, pp.11-32.

[Cheng 04] P.C. Cheng, B.C. Chien, H.R. Ke, and W.P. Yang, *“KIDS’s evaluation in medical image retrieval task at ImageCLEF 2004”*, Working Notes for the CLEF 2004 Workshop September, Bath,UK , pp. 585-593.

[Boser92] B. Boser, I. Guyon, and V. Vapnik A training algorithm for optimal margin classifiers. *In Proceedings of the Fifth Annual Workshop on Computational Learning Theory*, 1992.
[Salton83] G. Salton and M. J. McGill, *Introduction to Modern Information Retrieval*. New York: McGraw-Hill, 1983.