

# Meiji University at ImageCLEF2008 Photo Retrieval Task: Evaluation of Image Retrieval Methods Integrating Different Media

Kosuke Yamauchi, Takuya Nomura, Keiko Usui, Yusuke Kamoi, Miki Eto and Tomohiro Takagi

*Department of Computer Science*

*Meiji University*

*1-1-1 Higashimita, Tama-ku, Kawasaki-shi, Kanagawa, 214-8571, Japan*

Email: {yamauchi, takagi} @cs.meiji.ac.jp

## Abstract

This paper describes the participation of the Human Interface Laboratory of Meiji University in the ImageCLEF2008 photo retrieval task. We submitted eight retrieval runs taking two main approaches. The first approach combined Text-Based Image Retrieval (TBIR) and Context-Based Image Retrieval (CBIR). The second approach applied query expansion using conceptual fuzzy sets (CFS). A CFS is a method that uses the expression of meaning depending on the context, which an ordinary fuzzy set does not recognize. A conceptual dictionary is necessary to perform query expansion using CFS, and this is constructed by clustering. We propose here the use of query expansion with CFS, pseudo relevance feedback (PRF), and other techniques, for image retrieval that integrates different media, and we verify the utility of the system by explaining our experimental results.

## Categories and Subject Descriptors

H.3 [Information Storage and Retrieval]: H.3.1 Content Analysis and Indexing; H.3.3 Information Search and Retrieval; H.3.4 Systems and Software; H.3.7 Digital Libraries;

## Keywords

Information Retrieval, Image Retrieval, Query Expansion, Conceptual Fuzzy Sets, Fuzzy Clustering

## 1 INTRODUCTION

We present here our participation in the ImageCLEF2008 photo retrieval task. The task deals with answering 39 queries of variable complexity from a repository of 20,000 photographic images in the IAPR TC-12 photographic collection. The details of this task can be found in ImageCLEFphoto 2008 [1].

We submitted eight retrieval runs this year taking two main approaches.

The first approach is an image retrieval system that integrates Text-Based Image Retrieval (TBIR) and Context-Based Image Retrieval (CBIR). We are taking this approach in order to overcome the difficulties that arise with these methods individually. For example, using TBIR, a person's subjectivity can easily be introduced, and furthermore, images without text cannot be retrieved. Using CBIR with the present technology, it is very difficult to see only the contents of the image and to make the computer understand what the image is. This method is also less accurate than TBIR. We expect that these difficulties can be avoided by combining TBIR and CBIR, and the accuracy of the image retrieval can be improved by the synergistic effect of different media used to solve these problems.

The second approach is query expansion using Conceptual Fuzzy Sets (CFS). Current search engines are powerful. However, several words are difficult to search. Takagi et al. [2] proposed resolving this problem by query expansion depending on the context using CFS. A CFS is a method that uses meaning expression depending on the context, which a fuzzy set does not recognize. We propose a system that depends more on query context than on query expansion for improving the packaging method of a conceptual fuzzy set.

This paper is organized as follows. In section 2, CFS and a method of query expansion using CFS are presented. Section 3 describes the details of all eight retrieval systems that we submitted. Section 4 describes the results of our experiment. In Section 5, we consider the results and discuss our study.

## 2 CONCEPTUAL FUZZY SETS

In this section, we explain what the conceptual fuzzy sets (CFS) is. Next, the method necessary to construct a conceptual dictionary in order to perform a query expansion using CFS is described, and finally, the query expansion technique is described.

### 2.1 Conceptual Fuzzy Sets

A CFS is based on the use theory of meaning propounded by Wittgenstein for the expression of meaning of a concept. According to this theory, the meaning of a word can be expressed by another word. Thus, the meaning of a word can be expressed by other words associated with one another. Also, the expression of the meaning of a word by another word makes a closed circular system. In CFS, the meaning of a word is expressed by the set of words and their activity values.

General fuzzy sets express phenomena without a clear boundary. However, when applying them to various realistic problems, they are not situation-dependent. The meaning of a concept changes in situation-dependent phenomena, and cannot correspond to a fixed expression. This problem occurs when the generality of knowledge is not obtained because the ambiguity of a fuzzy set is fixed, and the mechanism including background is not given. The essence of this problem is the expression of meaning.

Conceptual fuzzy sets solve situation-dependent problems by a recall mechanism (i.e., combined fuzzy sets) and the general versatility of knowledge by making a closed circular system based on the use theory of meaning.

In our system, we created a conceptual fuzzy set by the superpositioning of concepts. The concept component of a fuzzy set is called a prototype concept, and a set of prototype concepts is called a concept dictionary. The concept depending on the input context can be generated by superpositioning a prototype concept that is similar to the input.

### 2.2 Conceptual Dictionary

We constructed a conceptual dictionary to perform query expansion using conceptual fuzzy sets. The following methods are commonly used to construct a conceptual dictionary.

- i) Manual construction.
- ii) Clustering corpora, assuming one cluster is one concept.
- iii) Using an existing edifice of knowledge (e.g., Wordnet), and making a concept from words included in each directory, etc.

However, the first method is a very tedious task, and the third method cannot be used to make a concept if the existing edifice of knowledge cannot be found. Thus, in our experiment, we constructed a concept dictionary by clustering.

We compared the precision of query expansion by constructing two concept dictionaries. We show each method for constructing these concept dictionaries below.

#### 2.2.1 Conceptual Dictionary using Fuzzy C-means

We used the fuzzy C-means method to construct the conceptual dictionary. The construction process was as follows.

1. From the data included in the corpus, make a word vector from the annotation text and a region ID vector from an image. Make a feature vector by combining the region ID vector and the word vector.
2. Cluster the feature vectors using the fuzzy C-means method; in this case, we set the number of clusters to be 1,000.
3. The fuzzy C-means method was used for clustering, and all vectors belonged to all clusters. Therefore, we set the threshold based on the membership value and limit the number of vectors that belong to one cluster. In addition, in a reorganization of the Term Frequency-Inverse Document Frequency (TF-IDF) method, we weight the feature vectors taking into consideration the degree of belonging to a cluster. First, we define  $E_j$  as a clustered element (i.e., feature vector in our experiment), and  $j$  expresses an identifier. Here,  $E_j$  has the same value of the degree of belonging to each cluster  $I$  (in our experiment, it was the same as the membership value), we define this as  $Belong(E_j, I)$ , and we define  $TF$  as the multiplication of feature vectors taking into consideration  $Belong(E_j, I)$ .  $TF$  is obtained with Eq. (1).

$$TF = \sum_{E_j \in I} Belong(E_j, I) \times E_j \quad (1)$$

4. We define the word set from 3. As the prototype concept.

### 2.3 Query Expansion using CFS

The technique of query expansion using CFS is described here. The flow of query expansion is shown below.

1. Extract a word vector from <title> and <narr> of topic, and extract a region ID vector from the image. Then, make a feature vector by combining the word vector and the region ID vector.
2. Calculate the degree of similarity of the feature vector of the query and each prototype concept in a concept dictionary using the cosine measure.
3. Make the query's concept by overlapping a number of similar prototype concepts  $N$ . The overlap of the prototype concept is calculated by Eq. (2) Here, *Similarity* is the cosine measure,  $C_i$  is the number of prototype concept  $i$ , and we set  $N = 5$ .

$$NewConcept = \sum_i^N Similarity (C_i, Query) \times C_i \quad (2)$$

4. Extract words that have a high score, and append this to the query. In our experiment, we extracted 15 words and appended these words to the query.

## 3 SYSTEM DESCRIPTION

In this section, details of the retrieval system that we submitted are described.

### 3.1 Integration System

This system is the bare system used as the base of all submitted systems, and it is considered to be the final retrieval result that integrates the retrieval results of TBIR and CBIR. This system was built based on the system of CINDY [3], which participated in the ImageCLEF2006 photo retrieval task. First, the details of TBIR and CBIR are explained, and then the method of integration is described.

#### 3.1.1 Text-based Image Retrieval

We used Apache Lucene [4] as the TBIR method. Apache Lucene is an all-text retrieval engine developed as open source software. In Lucene, we can use text retrieval with TF-IDF, but we built Okapi BM25 [5] into Lucene for our system.

#### 3.1.2 Content-based Image Retrieval

CBIR is depicted in Figure 1.

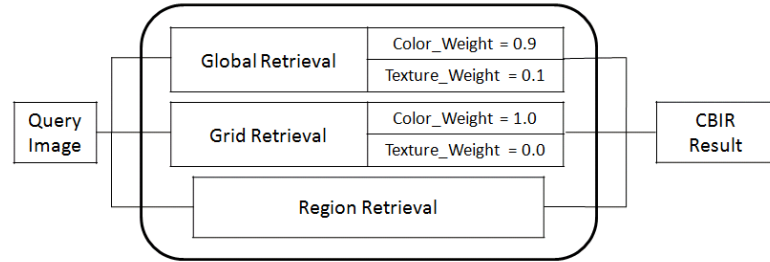


Fig. 1 CBIR

CBIR consists of three retrieval modules called global retrieval, grid retrieval, and region retrieval. The retrieval results of the three retrieval modules are integrated by Eq. (3), and this integration result is considered to be the final retrieval result of CBIR.

$$\begin{aligned}
 CBIR \ result &= global\_result * global\_weight \\
 &+ grid\_result * grid\_weight \\
 &+ region\_result * region\_weight
 \end{aligned} \quad (3)$$

where *result* indicates the retrieval result of each retrieval module, and *weight* indicates weight. In this system, each weight is as follows:  $global\_weight = 0.3$ ,  $grid\_weight = 0.6$ ,  $region\_weight = 0.1$ . The details of the three retrieval modules are described as follows.

- Global Retrieval

In the global retrieval module, first, the feature values of color (lab) and texture (wavelet transform) are extracted from the whole image. The distance scale between the images uses Earth Mover's Distance for the color and Euclidean distance for the texture. In this module, the weights of color and texture are as follows: Color\_Weight = 0.9, Texture\_Weight = 0.1.

- Grid Retrieval

In the grid retrieval module, first, an image is divided into a 3×3 block. And, feature values of color (lab average, standard deviation) and texture (wavelet transform) are extracted for each block. The distance scale between the images uses Euclidean distance for both color and texture. In this module, the weight of color and texture are as follows: Color\_Weight = 1.0, Texture\_Weight = 0.0.

- Region Retrieval

The region retrieval module first segments an image with the JSEG algorithm [6]. Then feature values are extracted from all regions. These features are normalized into Z-scores and combined into a 47-dimensional vector. The distance scale between the images uses the cosine measure. Table 1 lists the adopted features and their dimensions.

Table 1 Feature values of region (numbers in parentheses indicate number of dimensions)

<b>Color</b>	RGB average (3), standard deviation (3)
	Lab average (3), standard deviation (3)
<b>Texture</b>	Wavelet transform (24)
<b>Shape</b>	Z-Fourier descriptors (8)
<b>Position</b>	X and Y coordinates of median point (2)
<b>Size</b>	Number of pixels (1)

### 3.1.3 Integration System

Here, an integrated part of the retrieval result is described. Figure 2 outlines the flow of the Integration System.

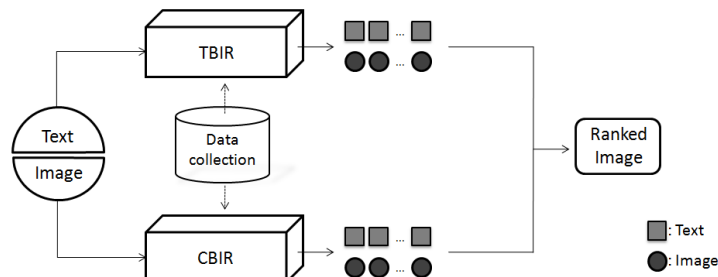


Fig. 2 Integration System

If a query is input, it is divided into an annotation and three images, and the annotation is passed to the TBIR, and the images are passed to the CBIR. TBIR described in 3.1.1 is executed using <title> and <narr> from the query annotation. At the same time, CBIR described in 3.1.2, is executed. Finally, the retrieval results of TBIR and CBIR are integrated by Eq. (4), and the final retrieval result is obtained.

$$\begin{aligned}
 \text{Final retrieval result} &= \text{TBIR\_result} * \text{TBIR\_weight} \\
 &+ \text{CBIR\_result} * \text{CBIR\_weight}
 \end{aligned}
 \tag{4}$$

where *result* indicates the retrieval result of each retrieval system, and *weight* indicates the weight. In this system, each weight is as follows: *TBIR\_weight* = 0.5, *CBIR\_weight* = 0.5.

### 3.2 Integration System + CFS

This system combined CFS with the Integration System. Figure 3 outlines the flow of the Integration System + CFS.

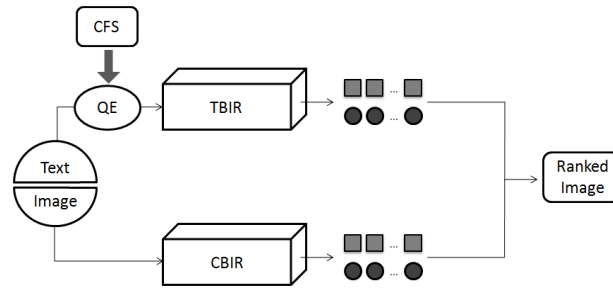


Fig. 3 Integration System + CFS

If a query is input, query expansion using CFS is performed when TBIR is executed. The details of query expansion using CFS are described in 2.3 of Section 2. The result of TBIR using query expansion is integrated with the result of CBIR, and the final retrieval result is obtained. In this system, each weight is as follows:  $TBIR\_Weight = 0.5$ ,  $CBIR\_Weight = 0.5$ .

### 3.3 Inter Media Pseudo Relevance Feedback

Inter Media Pseudo Relevance Feedback (IMPRF) is a query expansion system that is performed using an annotation of the image obtained through the retrieval result of CBIR, and that also executes TBIR. This system is based on the system of IPAL [7], which participated in the ImageCLEF2006 photo retrieval task. Figure 4 outlines the flow of IMPRF.

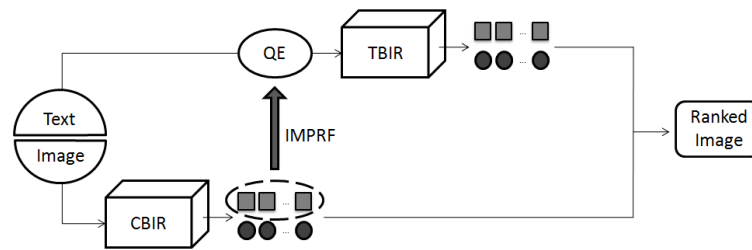


Fig. 4 IMPRF

When a query is input, first CBIR is executed. Next, query expansion is performed using the retrieval result of CBIR. Query expansion is performed by extracting 15 words from <TITLE>, <DESCRIPTION>, <NOTES> and <LOCATION> of the annotations of the top 5 images of the retrieval result of CBIR. TBIR is executed using the words obtained by the query expansion. The retrieval results of TBIR and CBIR are integrated, and the final retrieval result is obtained. In this system, each weight is as follows:  $TBIR\_weight = 0.5$ ,  $CBIR\_weight = 0.5$ .

### 3.4 IMPRF + PRF

This system combines pseudo relevance feedback (PRF) with IMPRF and is outlined in Fig. 5.

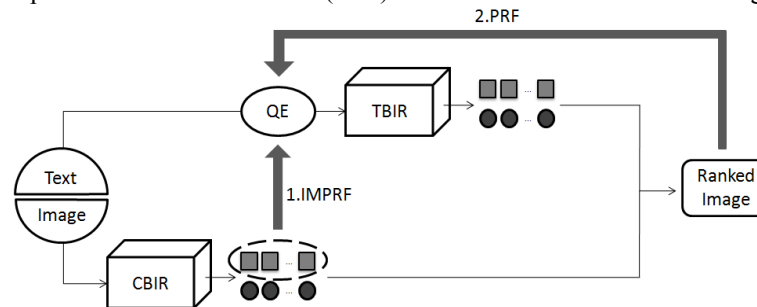


Fig. 5 IMPRF + PRF

Once the result of integration is similarly obtained with IMPRF as explained in 3.3, PRF is performed from the integration result. PRF is carried out by extracting 15 words from <TITLE>, <DESCRIPTION>, <NOTES>, and <LOCATION> of the annotations of the top 5 images of the retrieval result. TBIR is executed using words obtained by PRF. The retrieval results of TBIR and CBIR are integrated, and the final retrieval result is obtained. In this system, each weight is as follows:  $TBIR\_weight = 0.5$ ,  $CBIR\_weight = 0.5$ .

### 3.5 IMPRF + CFS

This system combines CFS with IMPRF and is outlined in Fig. 6.

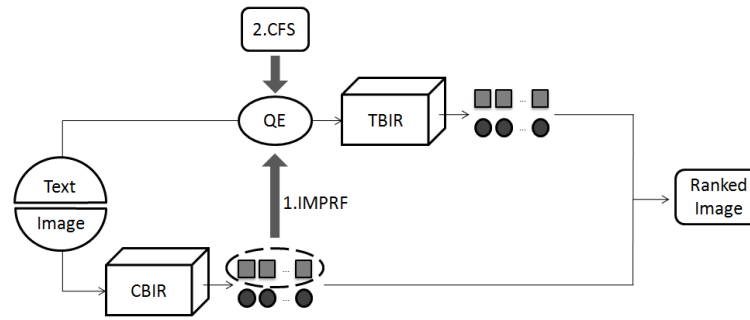


Fig. 6 IMPRF + CFS

If a query is input, first CBIR is executed. Next, query expansion is performed using the retrieval result of CBIR. Query expansion is carried out by extracting 15 words from annotations of the top 5 images obtained in the retrieval result of CBIR. Moreover, query expansion is performed using CFS for the words obtained by query expansion, and TBIR is executed. Finally, the retrieval results of TBIR and CBIR are integrated, and the final retrieval result is obtained. In this system, each weight is as follows:  $TBIR\_weight = 0.5$ ,  $CBIR\_weight = 0.5$ .

### 3.6 CMPRF + CFS + PRF

This system is Cross Media Pseudo Relevance Feedback (CMPRF), in which IMPRF is used twice. Figure 7 outlines the flow of CMPRF + CFS + PRF.

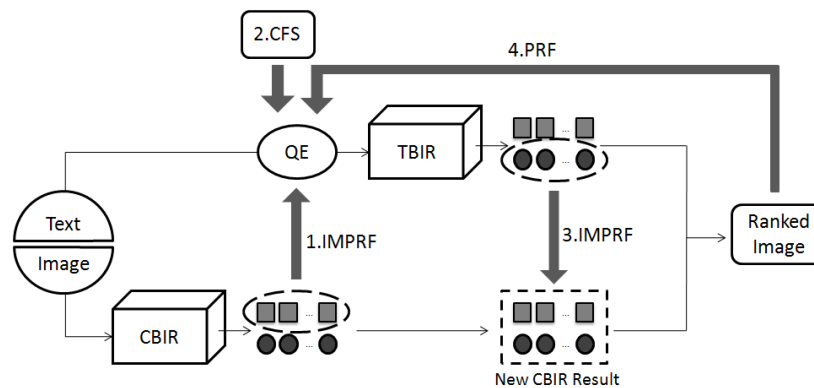


Fig. 7 CMPRF + CFS + PRF

When a query is input, CBIR is executed first. Query expansion is performed next using the retrieval result of CBIR. Query expansion is performed by extracting 15 words from the annotations of the top 5 images of the retrieval result of CBIR. Moreover, query expansion is performed using CFS for the words obtained by query expansion, and TBIR is executed. Next, feature values are extracted from the top 5 images of the retrieval result of TBIR, the average of the feature values is obtained from these, and CBIR is executed again. The retrieval results of TBIR and CBIR are integrated, and a retrieval result is obtained. Moreover, PRF is performed using this retrieval result. TBIR is executed again using words obtained by PRF. Finally, the retrieval results of TBIR and CBIR are integrated, and the final retrieval result is obtained. In this system, each weight is as follows:  $TBIR\_weight = 0.5$ ,  $CBIR\_weight = 0.5$ .

### 3.7 TBIR + CFS

This system combines CFS with TBIR and uses a text only approach. In section 2.2.1, the use of word vector and region ID vector was explained in terms of constructing the conceptual dictionary. However, the conceptual dictionary is constructed using only a word vector because this system uses text only. In addition, when a query expansion is performed, a word vector is generated from <title> and <narr> of the query, and this is assumed to be a feature vector. TBIR is executed as explained in 3.1.1.

### 3.8 Auto Feature Weighting

Auto Feature Weighting (AFW) is an approach using image only. This system examines how similar three query images are when retrieving, and the weight of feature values is dynamically changed according to the similarity. Figure 8 is an outline of the flow of AFW.

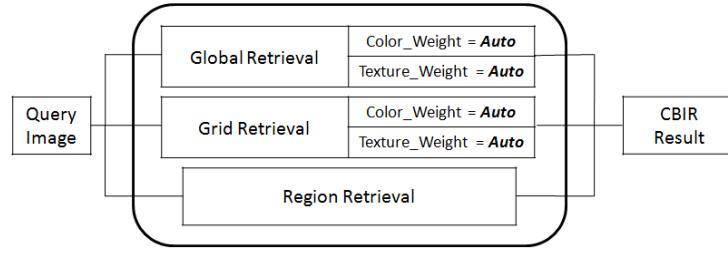


Fig. 8 Auto Feature Weighting

One way this system differs from CBIR explained in 3.1.2 is that the weights of color and texture of the global retrieval and grid retrieval change dynamically. This is represented by *Auto* in Figure 8.

If three query images are input, the distance between each query image in color and texture is calculated. The weight is obtained from the input distance. Because the similarity of the color and texture features is small if the distance is large, the weight is reduced. Because the similarity of the features is large if the distance is small, the weight is increased.

In addition, an inverse number of the average value is obtained by Eq. (5). The reason for this is that a large weight is given for the features that look alike because they look alike due to the small value to calculate the distance, as stated above.

$$\begin{cases} ave\_color\_similarity = \frac{1}{\left( \frac{color\_similarity_{1-2} + color\_similarity_{1-3} + color\_similarity_{2-3}}{3} \right)} \\ ave\_texture\_similarity = \frac{1}{\left( \frac{texture\_similarity_{1-2} + texture\_similarity_{1-3} + texture\_similarity_{2-3}}{3} \right)} \end{cases} \quad (5)$$

where *similarity* indicates the distance between each query image, and the subscript indicates the compared images. For example with  $color\_similarity_{1-2}$ , the distance in the color feature between the first and second images of the query image is shown.

Weight is obtained by Eq. (6) from  $ave\_color\_similarity$  and  $ave\_texture\_similarity$  that were calculated by Eq. (5).

$$\begin{cases} Weight_{color} = \frac{ave\_color\_similarity}{ave\_color\_similarity + ave\_texture\_similarity} \\ Weight_{texture} = \frac{ave\_texture\_similarity}{ave\_color\_similarity + ave\_texture\_similarity} \end{cases} \quad (6)$$

Whenever a query is input, these operations are performed, and CBIR is executed after calculating the weights.

## 4 EXPERIMENTAL RESULTS

Table 2 lists the experimental results of the systems submitted.

Table 2 Experimental results of photo retrieval task

RunID	Language	Modality	$P@20$	CR20	MAP	GMAP
Integration System	EN - EN	TXTIMG	0.4282	0.3975	0.2940	0.1823
Integration System + CFS	EN - EN	TXTIMG	0.4218	0.3297	0.2966	0.1697
IMPRF	EN - EN	TXTIMG	0.3949	0.3223	0.2711	0.1522
IMPRF + PRF	EN - EN	TXTIMG	0.4038	0.2861	0.2921	0.1368
IMPRF + CFS	EN - EN	TXTIMG	0.4231	0.3226	0.3032	0.1676
CMPRF + CFS + PRF	EN - EN	TXTIMG	0.3974	0.2858	0.2870	0.1364
TBIR + CFS	EN - EN	TXT	0.4051	0.3966	0.3015	0.1844
AFW	EN - EN	IMG	0.0051	0.0212	0.0005	0.0002

The Integration System had the highest accuracy in  $P@20$  and  $CR20$ . As for this result, it was understood that integrating different media, i.e., TBIR and CBIR, was conducive to the higher retrieval result. Moreover, when looking at the mean average precision ( $MAP$ ), which can comprehensively assess how well the retrieval system performs, the top three systems perform query expansion using CFS. This indicates that performing query expansion using CFS produces a higher retrieval result. Here, the notable conclusion is TBIR + CFS. Although this system approach uses text only, it retrieves with very high accuracy on the whole. However, it is clear that of the systems we submitted, the best accuracy is obtained using IMPRF + CFS. Although an approach using image only is not very accurate, it is understood that accuracy improves by integrating different media. Furthermore, when query expansion is used abundantly with one system (e.g., CMPRF + CFS + PRF), query expansion does not produce good results, and a lot of the same words appear. Although query expansion is one technique for increasing retrieval accuracy, it is apparent that greater frequency of query expansion will not result in higher accuracy.

## 5 DISCUSSION

We demonstrated that retrieval accuracy improved by performing query expansion using CFS in image retrieval that integrates different media.

However, future tasks remain, and these are as follows.

It is necessary to improve the accuracy because the accuracy of CBIR is low. Because it is clear that accuracy improves by integrating TBIR and CBIR, it is assumed that higher accuracy of CBIR will lead to a further rise in overall accuracy. In addition, it is thought that accuracy will increase by revising the method of integration because simply adding TBIR and CBIR is not sufficient (e.g., Eq. 4). In other words, the method of integration plays a very important role when TBIR and CBIR are integrated.

Also, another problem is how to determine the initial point of clustering when the conceptual dictionary is constructed. This system sets the initial point at random. It is thought that accuracy will fall outside the accuracy range if another corpus is used, but there is the possibility that accuracy will be higher or lower than the results reported here when the conceptual dictionary is constructed anew. It is necessary to consider using another fuzzy clustering technique if the accuracy changes each time the conceptual dictionary is constructed.

## REFERENCES

- [1] ImageCLEFphoto2008, "<http://www.imageclef.org/2008/photo>"
- [2] Yan-Chen Chen, Hiroshi Sekiya, and Tomohiro Takagi, "Conceptualized Queries for Information Retrieval," *Proc. of North American Fuzzy Information Processing Society Annual Conference (NAFIPS)*, 2007.
- [3] M.M. Rahman, Varun Sood, Bipin C. Desai, and Prabir Bhattacharya, "CINDI at ImageCLEF 2006: Image Retrieval & Annotation Tasks for the General Photographic and Medical Image Collections," *CLEF 2006 Working Notes*, 2006.
- [4] Apache Lucene, "<http://lucene.apache.org/>"
- [5] Stephen E. Robertson, Steve Walker, Susan Jones, Micheline Hancock-Beaulieu, and Mike Gatford, "Okapi at TREC-3," *In Proceedings of the Third Text Retrieval Conference (TREC)*, 1994.
- [6] Y. Deng and B. S. Manjunath, "Unsupervised segmentation of color-texture regions in images and video," *Transactions of Pattern Analysis and Machine Intelligence 2001*, 2001.
- [7] Nicolas Mailliot, Jean-Pierre Chevallet, Vlad Valea, and Joo Hwee Lim, "IPAL Inter-Media Pseudo-Relevance Feedback Approach to ImageCLEF 2006 Photo Retrieval," *CLEF 2006 Working Notes*, 2006.