Information Sciences 178 (2008) 4301–4313

Contents lists available at ScienceDirect

Information Sciences

journal homepage: www.elsevier.com/locate/ins

Image retrieval model based on weighted visual features determined by relevance feedback

Woo-Cheol Kim, Ji-Young Song, Seung-Woo Kim, Sanghyun Park *

Department of Computer Science, Yonsei University, 134 Sinchon-dong Seodaemun-gu, Seoul 120-749, Republic of Korea

ARTICLE INFO

Article history:
Received 7 March 2007
Received in revised form 6 May 2008
Accepted 23 June 2008

Keywords:
Image retrieval
Keyword-based image retrieval
Content-based image retrieval
Relevance feedback
Multimedia database

ABSTRACT

An accurate and rapid method is required to retrieve the overwhelming majority of digital images. To date, image retrieval methods include content-based retrieval and keyword-based retrieval, the former utilizing visual features such as color and brightness, and the latter utilizing keywords that describe the image. However, the effectiveness of these methods in providing the exact images the user wants has been under scrutiny. Hence, many researchers have been working on relevance feedback, a process in which responses from the user are given as feedback during the retrieval session in order to define a user's need and provide an improved result. Methods that employ relevance feedback, however, do have drawbacks because several pieces of feedback are necessary to produce an appropriate result, and the feedback information cannot be reused. In this paper, a novel retrieval model is proposed, which annotates an image with keywords and modifies the confidence level of the keywords in response to the user's feedback. In the proposed model, not only the images that have been given feedback, but also other images with visual features similar to the features used to distinguish the positive images are subjected to confidence modification. This allows for modification of a large number of images with relatively little feedback, ultimately leading to faster and more accurate retrieval results. An experiment was performed to verify the effectiveness of the proposed model, and the result demonstrated a rapid increase in recall and precision using the same amount of feedback.

1. Introduction

Information on the internet is shifting from text-based to multimedia-based with large amounts of visual and audio data. The development of tools such as digital cameras and scanners, which convert analog data into digital data, has accelerated the increase in multimedia information on the internet, and widened internet bandwidth has dramatically improved access. These changes have demonstrated the need for current internet search systems to improve their search engines to include multimedia data such as images, music, and videos. Among these, images are most numerous, requiring a more efficient searching technique.

The current image search technology is keyword-based. Keyword-based searching uses the file name by which the image has been stored or keywords describing the image. The limitation of this method is that if no keyword has been associated with an image or if the keyword associated with an image does not describe the image properly, the accuracy of the search result is poor. Entering a keyword(s) for each image manually is a short-term solution, but considering the rapid increase in the number of images, this cannot be an ultimate solution.

* Corresponding author. Tel.: +82 2 2123 5714; fax: +82 2 365 2579.
E-mail addresses: twelvepp@cs.yonsei.ac.kr (W.-C. Kim), jysong@cs.yonsei.ac.kr (J.-Y. Song), kimsw@cs.yonsei.ac.kr (S.-W. Kim), sanghyun@cs.yonsei.ac.kr (S. Park).

0020-0255/$ - see front matter
doi:10.1016/j.ins.2008.06.025

Content-based searching is another method used for image retrieval and it uses visual features such as color distribution and brightness as search conditions. The limitation of a content-based search is the fact that a visual feature recognized by a computer can be different from the same feature recognized by a person [3,15]. Also, users who are accustomed to text-based searching may experience difficulty using images as a search condition. The digital visual features of an image of the same object may also vary considerably, depending on the conditions under which the image was taken.

Therefore, neither the content-based method nor the keyword-based method provides satisfactory precision. We therefore suggest the notion of relevance feedback as an additional method to improve these retrieval methods in which the users present the system with several example answers from an initial retrieval result. The system then uses these answers to modify the search conditions and return a more refined result [14,19].

Previous studies that have employed relevance feedback have improved the retrieval result accuracy by comparing the similarity of images using the visual features shared among the positive images. However, the feedback process must be iterated to yield a correct result, and the feedback information acquired cannot be reused [14,19]. Recently, to achieve acceptable retrieval results when the same queries are given, Hoi et al. kept a log of feedback information [10].

Studies that have integrated relevance feedback with keyword-based searches have improved the accuracy of their results by refining the degree of association between an image and its keyword by automatically using feedback data. That is, the keyword becomes more relevant or less relevant to an image as the feedback process iterates, thereby increasing the keyword confidence of positive images and decreasing the keyword confidence of negative images [9,11,12]. The limitation of this system is that only the confidence in the images that have received feedback is modified, leaving other images unmodified.

To solve the problems generated when applying relevance feedback to a keyword-based search, this paper proposes a new model which integrates keyword-based and content-based searches. In the proposed model, not only the images which have received feedback, but also other images with visual features similar to the features used to distinguish positive images are subjected to confidence modification. This enables modification of a large number of images with only a few pieces of feedback, ultimately leading to faster and more accurate retrieval results.

This paper is organized as follows. In Section 2, current studies on image retrieval will be introduced. Section 3 will explain the proposed search model in detail. In Section 4, the performance of the proposed model will be evaluated experimentally.

2. Related work

Image retrieval is usually classified into content-based searches and keyword-based searches. A content-based search extracts visual features such as color, texture, and shape from the image as search conditions [1,3,6,15]. QBIC [6], Virage [15], and VisualSEEK [1] are the best known content-based search systems. QBIC uses color, texture, sample images, and sketches as queries and is employed mainly in vast image or video databases. Virage uses color layout, texture, and the contour of an object as visual features; this system is being applied in face recognition and in the retrieval of ophthalmologic images. VisualSEEK searches images using top, bottom, right, and left color relation, which are the most important visual features people use in recognizing an image.

Most of the content-based search systems utilize several visual features to increase their retrieval accuracy. However, the visual features calculated by a computer program can be different from the visual features people recognize, limiting accuracy improvement. Moreover, users who are accustomed to a text-based search engine may experience difficulties in using images as a search condition [3].

Keyword-based searches use descriptive keywords as a search condition. The image retrieval accuracy will increase significantly if the keyword describes an image accurately. Although the best case, in terms of accuracy, is to enter the keywords manually, it is impossible for people to enter accurate keywords for each image considering the size of current image databases. This is why there have been many studies on automatic keyword annotation [2,5,7,8,18].

Cheng and Chien annotated a new image with a keyword by regional clustering [2]. The new image is first divided into several regions based on visual features, and then the keyword of the cluster most similar to each region is given to the image. Jeon et al. suggested a semantic hierarchy for refining confidence in a keyword [7]. For example, since a cat and dog are both pets, images annotated with cat and dog will have high confidence when considering the keyword 'pet'. Feng et al. suggested applying two different systems to one image. Once two keyword sets are produced, the keywords included in both sets are given high confidence, while the keywords included in only one set are given low confidence [5]. KBIR has a disadvantage in that the keyword can be varied according to the perspectives of users. Zhang et al. used the term feedback (a method that changes keywords to keywords with a similar meaning throughout the feedback process in order to enhance retrieval accuracy) to enhance retrieval accuracy [18]. Recently, ALIPR [8] proposed a real-time computerized annotation system based on the pixel information stored in the image. When given a new image, this system extracts 15 relevant keywords using D2-clustering on a set of feature vectors obtained from the new image and the training images, and lets users choose several keyword from among the extracted 15 keywords.

Basically, the automatic keyword annotation systems mentioned above employ a method that connects the keyword to other images through visual similarity calculated by various programs. Once every keyword has been annotated, a keyword-based search can be carried out, thereby improving the search speed compared to content-based searches. However, these systems still face the issue of low retrieval accuracy. For this reason, there has been great interest in research on using relevance feedback in image retrieval in order to increase retrieval accuracy [14,19]. Relevance feedback refers to a method in which users evaluate the initial retrieval result to define their query more accurately and achieve a better retrieval result. Relevance feedback can be employed both in content-based and keyword-based searches.

Relevance feedback can be employed in a content-based search in two ways. The first is to modify the information representing the query image so that it is more similar to the positive images and less similar to the negative images, according to the feedback given by the user. The second is to give higher weight to more distinguishing features in order to reflect the user feedback when calculating similarity [14,16,17].

In keyword-based searches, a keyword is added or removed according to the feedback from the user [9], or the confidence in a keyword is modified [2,10]. This technique will not immediately provide the users with an improved retrieval result, but as feedback data accumulate, the confidence level of irrelevant keywords will become lower and that of relevant keywords will become higher. A keyword can be added or removed from the image. However, there are some limitations in how the confidence level can be modified when only the confidence level of images given as feedback is modified. In order to improve overall retrieval accuracy, the feedback process must iterate several times.

To solve the problems generated when applying relevance feedback to keyword-based searches, this paper proposes a new model that integrates keyword-based and content-based searches. In the proposed model, images that have had feedback supplied, as well as other images with visual features similar to the features used to distinguish the positive images are subjected to confidence modification. This allows the modification of many images with a limited amount of feedback, ultimately providing faster and more accurate retrieval results.

3. The proposed image retrieval model

In this section, the proposed model is described. The model can be divided into two parts. The first is an image collection portion which collects image data into an image database. The other part is an image retrieval portion which searches for the images that the user wants.

Image collection uses various methods to create metadata for image retrieval. That is, all images have to be annotated in order to employ keyword-based retrieval while visual features such as color, shape, and texture have to be extracted from images for content-based retrieval. Since this paper employs both retrieval methods, we need to annotate the images with keywords and extract visual features from the images. These tasks are performed in Fig. 1①.

The image retrieval is composed of three modules. First, the keyword-based image retrieval module (Fig. 1②) searches a list of ranked images that include at least one of the query keywords received from the user. These ranked images are called the *initial retrieval result*. This technique shows low accuracy when a keyword annotated to an image does not represent the content of the image accurately enough to provide a satisfactory retrieval result. So, to improve accuracy, the retrieval system receives feedback information, which indicates the images that are relevant and those that are not. The image reordering module (Fig. 1③) is used to analyze the images given as feedback, and then provides the user with a rearranged result. As a last step, the confidence modification module modifies the keyword confidence of the images, raising the confidence level of positive images and lowering that of negative images. Furthermore, modifying the confidence level of other images with similar visual features to distinguish relevant images from irrelevant images enables the system to modify more images with an equal amount of feedback.

In this section, we describe the image retrieval process mentioned above in detail. The properties of images that must be stored for keyword-based and content-based searches will be defined in Section 3.1. In Section 3.2, automatic keyword annotation, which creates the initial value of the keyword and confidence, is explained. In Section 3.3, the keyword-based search

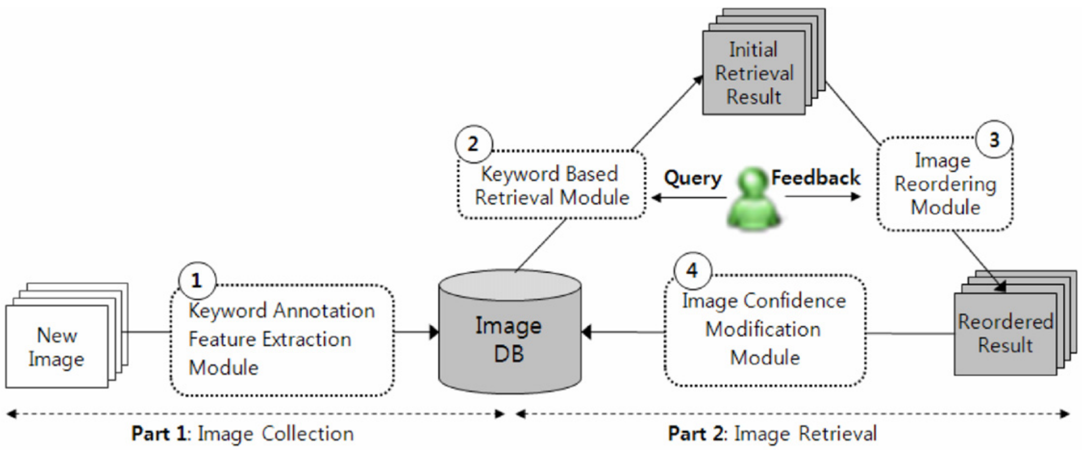

Fig. 1. Proposed image search model.

process is described and in Section 3.4, the creation of an improved, re-arranged result through relevance feedback is explained. Modifying the image confidence to provide improved image retrieval results is discussed in Section 3.5.

3.1. Image and property

In order to employ the proposed model, an image must have three properties: (1) a keyword for the keyword-based search, (2) keyword confidence reflecting the degree of relevance between the image and the keyword, and (3) a low level visual feature for the content-based search. The definitions of image property are defined below:

Definition 1. Image I:

$$I = [F, (K, C)].$$

- F is the set of low level visual features, which express the color, pattern, and texture of an image. A system has n visual features extracted from an image. If we define the jth property of the image I as f_i or *Feature*(I,j), F can be represented as $\{f_i = Feature(I,j) | 1 \leqslant j \leqslant n\}$. VF_j is jth visual feature used in this model.
- K is the set of keywords representing the content of an image. An image has m different keywords at most. The jth keyword linked to image I is represented as k_i or *Keyword*(I,j). Defining all the keywords that a set of images can have as W, K can be represented as $\{k_i = Keyword(I,j) | 1 \leqslant j \leqslant m, k_jW\}$.
- C is the set of confidence levels representing the degree of accuracy by which the keyword describes the image. The confidence of the jth keyword in association with image I is represented as c_j or *Confidence*(I,k_j) and has a minimum value of MINCONF and a maximum value of MAXCONF. Also, the confidence of keyword kw, which is the jth keyword of image I associated with image I, is *Confidence*$(I,kw(=k_j))$. If keyword kw is not associated with image I, the value of *Confidence*(I,kw) is 0.

3.2. Automatic keyword annotation

Several methods can be used to automatically annotate keywords including (1) finding a suitable keyword by analyzing the context in which the image is stored and (2) using a content-based search method. Content-based searches show a relatively high accuracy and robustness with regard to the environment from which the image comes. Therefore, this paper employs automatic keyword annotation through a content-based search.

Automatic keyword annotation with a content-based search requires a training set T, which is a set of images associated with keywords with high confidence. The step of automatic keyword annotation for new images using the training set T occurs as follows. First, the similarity in visual features between new images I_{new} and the images I_i belonging to training set T, is calculated using *FeatureSimilarity* (I_{new},I_i). FeatureSimilarity ranges from 0 to 1 and as the value approaches 1, the visual features of the two images become more similar.

Next, in order to select a keyword to be associated with a new image I_{new}, the system calculates the *Confidence*(I_{new},kw) of each kw from set W, a set of all keywords. Formula (1) is used to calculate the confidence level. Now, each image can have m keywords at most, and thus the top m keywords based on confidence level are selected. These keywords form a set for a new image I_{new}. If any of these m keywords have lower confidence than MINCONF, the keyword is considered to be irrelevant and is removed from the keyword set of I_{new}:

Formula (1). Confidence of keyword kw associated with a new image I_{new}:

$$Confidence(I_{\text{new}}, kw) = \frac{\sum_{I_i \in T}(FeatureSimilarity(I_{\text{new}}, I_i) \times Confidence(I_i, kw))}{\sum_{I_i \in T} FeatureSimilarity(I_{\text{new}}, I_i)}. \tag{1}$$

3.3. Keyword-based image retrieval

Once all images in the database are annotated using the automatic keyword annotation described in Section 3.2, keyword-based retrieval can be performed. The keyword-based image retrieval process is as follows: (1) The user enters the query keywords as a search condition; (2) The database retrieves all images that contain one or more query keywords; and (3) The retrieved images are arranged based on the sum of keyword confidence and are presented to the user in the order of the sum of confidences. Defining the query keywords as Q, the sum of confidences is calculated using *ConfidenceSum* (Q,I_i) described in Formula (2):

Formula (2). Sum of confidence between image I_i and query set Q:

$$ConfidenceSum(I_i, Q) = \sum_{kw \in Q} Confidence(I_i, kw). \tag{2}$$

The systems for keyword annotation and keyword-based image retrieval must consider keyword synonyms. For synonym processing, our system utilizes the dictionary that maintains sets of synonyms with their representative keywords. More specifically, our system only annotates representative keywords during the process of automatic keyword annotation, and replaces user-specified keywords with their representative keywords before initiating the process of keyword-based image retrieval.

3.4. Relevance feedback

If automatic keyword annotation has high accuracy, it could provide satisfactory results by applying the process discussed in Section 3.3. However, automatic keyword annotation generally shows a low accuracy level and users are not very likely to receive the retrieval result they wanted with a simple keyword search. Therefore, this paper has employed relevance feedback in order to increase retrieval accuracy.

Relevance feedback reorders a set of images retrieved as an initial result set using feedback information from the user. There are two types of feedback: positive image and negative image. The former represents images of interest to the user, while the latter indicates images not of interest to the user. We can enhance the retrieval result by analyzing user feedback and recalculating the similarity between the query and images.

3.4.1. Visual features used in relevance feedback

This paper is based on the idea that the visual feature used to distinguish positive and negative images depends on the query keyword. For instance, if a keyword '*forest*' is used as a query, the user is likely to focus on the color rather than the shape or pattern when submitting his or her feedback. The accuracy of an image retrieval result can be increased by retrieving images which contain colors similar to the positive image and by increasing the confidence of the keyword'forest' associated with the images.

In order to reflect the user's feedback in the confidence of a keyword, the visual feature, which plays a critical role in distinguishing positive and negative images, must be decided first. This paper employs DP (discrimination power) of visual features for this purpose. Discrimination power of VF_j is a numerical value which represents how well the positive and negative images are distinguished when rearranging the images according to VF_j. That is, the difference between the average f_j of positive images and the f_j of each initial retrieval result image is calculated, and then the initial retrieval result is rearranged according to the difference. Smaller differences between the two values correlate to a higher rank. Formally DP_j, the discrimination power of VF_j, is calculated as shown below:

Formula (3). Discrimination power of VF_j:

$$DP_j = \frac{Po_j + Ne_j}{N_p + N_n}. \tag{3}$$

where N_p and N_n represent the number of positive and negative images selected by the user, respectively. Po_j and Ne_j represent the number of positive images among the top N_pth images and the number of negative images among the bottom N_nth images, respectively, when the initial retrieval result is rearranged according to the average f_j of the positive images. The DP_j value becomes higher as VF_j distinguishes the positive and negative images more accurately. That is, VF_j has strong discriminatory power if it assigns relevant and irrelevant images into two opposite sections (relevant images at the top and irrelevant images at the bottom) of the list. So, if we go back to the previous instance, we can see that the visual feature for color has a relatively higher discrimination power than the others, such as shape or pattern, when '*forest*' is used as a keyword.

3.4.2. Rearrangement of images

After calculating DP_j as shown above, the initial retrieval result has to be rearranged and presented to the user. To do this, the virtual query image I_{avg} is first created by averaging each visual feature of the positive images.

Next, the similarity between each image I_i included in the initial retrieval result and the virtual image I_{avg} is calculated using the similarity function *WeightedFeatureSimilarity* ($I_{\text{avg}}, I_i, \{DP_1, DP_2, \ldots, DP_n\}$). This similarity function calculates the weight of each feature using Formula (4) shown below, and reflects the weight of each feature in calculating the similarity of the two images. In Formula (4), n is the number of visual features, DP_j and w_j are the discrimination power and weight of VF_j, respectively:

Formula (4). Similarity between I_i and I_{avg}.

(a) Weight of VF_j

$$w_j = \frac{DP_j}{\sum_{k=1}^{n} DP_k}. \tag{4}$$

(b) Similarity using VF_j

$ns_j =$ normalized similarity between I_i and I_{avg} using VF_j.

(c) WeightedFeatureSimilarity

$$WeightedFeatureSimilarity(I_{\text{avg}}, I_i, \{DP_1, \ldots, DP_n\}) = \sum_{j=1}^{n} w_j \times ns_j.$$

When the initial retrieval result is rearranged according to the above similarity function which reflects the weight of each visual feature, the images with visual features similar to the images selected as positive by the user are ranked higher than the images that have different visual features. Through this, the user is given a retrieval result including the images that have been given positive feedback among all images containing the query keyword.

3.5. Confidence modification of a keyword

The image retrieval system proposed in this paper uses feedback information to modify the confidence of keywords in the initial retrieval result, which we call the *extended feedback method*.

First, for positive images, the confidence of each keyword given as a query is increased to a certain degree. The upper limit of the confidence must be set, and if the increased confidence exceeds MAXCONF, the confidence will be set as MAXCONF. If a positive image does not have a query keyword, the keyword is annotated with the confidence value MINCONF. Next, for the negative images, the confidence of each keyword given as a query is decreased to a certain degree. If the confidence becomes lower than MINCONF, the keyword is considered irrelevant to the image and thus removed from the image. Lastly, the system selects the images ranked highly, but which have not been given feedback from the user, and modifies the confidence of the selected images. These selected images are called *additional images*. However, the increment at this time must be smaller than that of positive images since one cannot be sure of the accuracy of additionally selected images.

By using the extended feedback method, the confidence level of an image can be modified faster with the same number of pieces of feedback. However, confidence modification using extended feedback will not always improve the accuracy of image retrieval. For instance, suppose that among a set of images associated with keyword *kw*, the keywords of all the relevant images have a MAXCONF value, while the keywords of all the irrelevant images have a value between MINCONF and MAXCONF. The confidence of a correct keyword cannot become higher in this case, thereby increasing the confidence of an irrelevant keyword. This will result in a decrease in the image retrieval accuracy.

To deal with such a problem, this paper proposes to apply the extended feedback method selectively. That is, for each query keyword, we first calculate its average confidence from additional images. If the average is lower than the predetermined threshold value, the extended feedback method can be employed. If not, we consider that confidence modification has been performed well enough to provide a satisfactory result, and thus do not employ the extended feedback method.

4. Experiment

In this section, the effectiveness of the proposed relevance feedback method will be verified through an experiment.

4.1. Environment

In this experiment, 9281 images used in CalTech image research [4] were selected and used. The 9281 images are associated with a suitable keyword and have visual features which clearly describe the object. For the experiment, a Pentium 4 2.8 GHz computer with 1 GB of memory and an 80 GB hard disk operating with Window XP was utilized.

The visual features of each image were extracted by MPEG-7 XM software [13] and five visual features (color layout, color structure, homogeneous texture, edge histogram, and region shape) were simultaneously used. The similarity function developed in XM software was adopted for this study.

The size of the training set was 360 images, which is about 4% of the total number of images. Keywords were associated with the other 8921 images automatically using the training set during the automatic keyword annotation step. A total of 9281 images, including the training set, were used as inputs to a database.

Performance evaluation is based on two methods. One is the naive feedback method [11] developed by Lu et al. This method simply modifies the confidence level of images which have received feedback. The second method is the extended feedback method proposed in this paper which modifies the confidence level of additional images.

4.2. Definition of terms

The image retrieval result should be manually checked to determine experimental accuracy. However, checking the accuracy of all the automatic keyword annotations and retrieval results manually is extremely labor-intensive. To solve this matter, we regarded the keywords annotated objectively to CalTech images as *manual keywords* so that the system itself can check the accuracy automatically. A manual keyword describes the content of an image accurately and therefore has a confidence of MAXCONF.

Therefore, images have both manual and automatic keywords. For a keyword *kw*, $Confidence^{M}(I_i, kw)$ denotes the confidence of *kw* treating *kw* as a manual keyword. If *kw* is not a manual keyword of I_i, $Confidence^{M}(I_i, kw)$ returns 0. Similarly, $Confidence^{A}(I_i, kw)$ denotes the confidence of *kw* treating *kw* as an automatic keyword. If *kw* is not an automatic keyword of I_i, $Confidence^{A}(I_i, kw)$ returns 0. In addition, Set^{A}_{kw} is defined as a set of images that have *kw* as their automatic keyword, while Set^{M}_{kw} is defined as a set of images that have *kw* as their manual keyword.

4.3. Definition of recall and precision

In this section, we define the recall and precision used for measuring the retrieval accuracy. *Recall* is the proportion of retrieved accurate data among the accurate data related to the query within a database. This shows the capability of the system to retrieve relevant data. The proportion of accurate data among the retrieved data is called *precision*.

Recall and precision are shown in Fig. 2. Set *A* represents all of the related images within a database when a query has been submitted, while set *B* represents all of the retrieved images by a given query. That is, for a given query keyword *kw*, set *A* can be represented as Set^{M}_{kw}, which is a group of images annotated by manual keyword *kw*, and set *B* can be represented as Set^{A}_{kw}, which is a group of images annotated by automatic keyword *kw*. The exact ones the user wants among the images retrieved by the system are placed in the space where the two circles overlap and can be represented as the intersection of Set^{M}_{kw} and Set^{A}_{kw}.

The above definition of recall and precision cannot be directly applied to our system since each keyword may have a different confidence level. For these reasons, we propose new recall and precision measurements which consider keyword confidence as well.

For a given keyword *kw*, its recall is now defined as the rate between the sum of the confidences of the images having *kw* as their manual keyword and the sum of the confidences of the images to which *kw* is accurately associated as their automatic keyword. That is, recall is calculated as shown below:

Formula (5). Recall of keyword *kw*:

$$Recall(kw) = \frac{\sum_{I_i \in \{Set^{M}_{kw} \cap Set^{A}_{kw}\}} Confidence^{A}(I_i, kw)}{\sum_{I_i \in \{Set^{M}_{kw}\}} Confidence^{M}(I_i, kw)}. \tag{5}$$

As expressed in Formula (6), the precision of keyword *kw* can be defined as the proportion of the sum of the confidences of the images having *kw* as their automatic keyword and the sum of the confidences of the images to which *kw* is accurately associated as their automatic keyword:

Formula (6). Precision of keyword *kw*:

$$Precision(kw) = \frac{\sum_{I_i \in \{Set^{M}_{kw} \cap Set^{A}_{kw}\}} Confidence^{A}(I_i, kw)}{\sum_{I_i \in \{Set^{A}_{kw}\}} Confidence^{A}(I_i, kw)}. \tag{6}$$

The recall and precision of the entire system can be calculated by the average recall and average precision of all the keywords registered in the system, as shown in Formulas (7) and (8). Here, *W* and $|W|$ express the set of keywords used in the system and the size of *W*, respectively:

Formula (7). Recall of entire system:

$$Recall = \frac{\sum_{kw \in W} Recall(kw)}{|W|}. \tag{7}$$

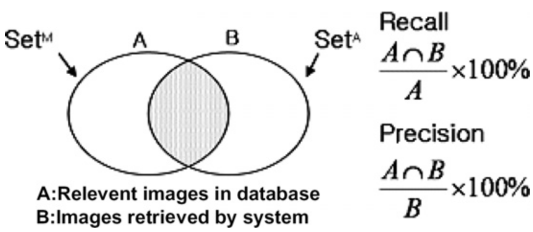

Fig. 2. Recall and precision.

Formula (8). Precision of entire system.

$$Precision = \frac{\sum_{kw \in W} Precision(kw)}{|W|}. \tag{8}$$

4.4. Parameter decision

An initial experiment must be performed to decide two critical parameter values: $\text{Threshold}^{\text{Size}}$ and $\text{Threshold}^{\text{EXTFB}}$. The parameter $\text{Threshold}^{\text{Size}}$ determines the number of images selected as additional images in the extended feedback method, and the parameter $\text{Threshold}^{\text{EXTFB}}$ determines whether to select additional images in the extended feedback method.

Experiments were performed to investigate the recall and precision of the system, while receiving feedback from 0 to 7000 times. Three positive and negative images were chosen by the user (i.e., N_p and N_n). Keyword confidence ranged from a MINCONF value of 0 to a MAXCONF value of 5.

4.4.1. $\text{Threshold}^{\text{Size}}$ parameter decision

Greater $\text{Threshold}^{\text{Size}}$ values result in more images being selected as additional images. As a consequence, the confidence of more keywords rises, thus quickly increasing the recall of the system. However, the possibility of irrelevant images being included becomes higher as the value grows, resulting in lower precision. To prevent this, an optimal parameter value for both recall and precision should be decided. In Figs. 3 and 4, the graph shows the growth pattern of recall and precision as $\text{Threshold}^{\text{Size}}$ increases in feedback image size multiples of 0.5–1.5. The x-axis denotes the number of queries and the y-axis shows the value of recall and precision.

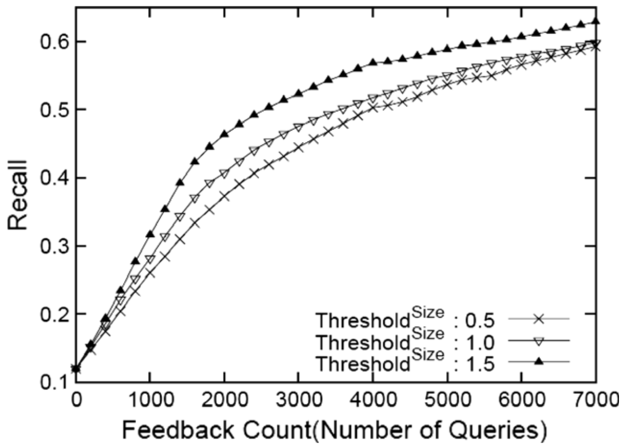

Fig. 3. Comparison of recall values for deciding $\text{Threshold}^{\text{Size}}$.

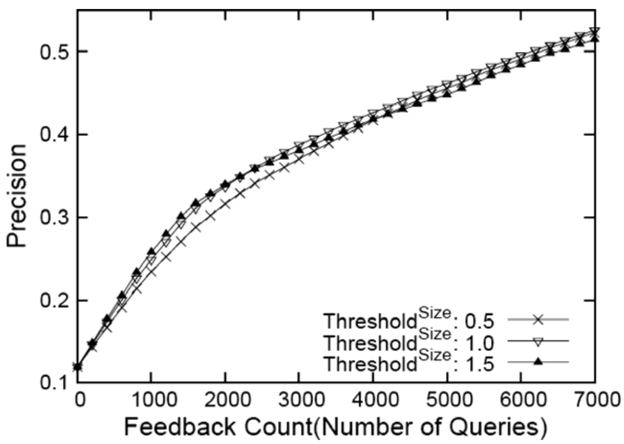

Fig. 4. Comparison of precision values for deciding $\text{Threshold}^{\text{Size}}$.

As shown in Fig. 3, the recall becomes larger when Threshold^{Size} increases. However, notice from Fig. 4 that the increment in Threshold^{Size} makes the precision higher only when the number of pieces of feedback is relatively small (i.e., less than 1500 in the figure). Beyond about 1500 counts of feedback, the increment of Threshold^{Size} does not bring significant benefit. When Threshold^{Size} equals 1, both precision and recall are maximized as a whole. Considering the trade-off between recall and precision, we set the base value for Threshold^{Size} to 1 in subsequent experiments.

4.4.2. Threshold^{EXTFB} parameter decision

Next, the value of Threshold^{EXTFB} is examined. Threshold^{EXTFB} is the value that determines if the confidence modification is performed well enough by comparing the confidence average of additional images with MAXCONF. Supposing the MAXCONF value is 5 and the Threshold^{EXTFB} value is 0.8, we use a value of 4, the product of MAXCONF and 0.8, as a standard to determine how to employ the extended feedback method.

By changing the Threshold^{EXTFB} value from 0.6 to 0.8, the recall and precision values can be investigated. When the Threshold^{EXTFB} value decreases, the system stops the extended feedback process earlier. As a result, its defect rate, e.g., a slow increment of precision, can be reduced. In contrast, the recall at this time rises slowly because the system loses chances to modify the confidence of accurate keywords. Therefore, we have to decide an optimal value of Threshold^{EXTFB} to maximize both recall and precision.

In Fig. 5, after 7000 feedback events, the recall reaches a maximum when Threshold^{EXTFB} is 0.8. When Threshold^{EXTFB} is 0.7 and 0.8, the recall is satisfactory although slightly reduced. Moreover, as shown in Fig. 6 (especially in Fig. 6b), the precision becomes lowest when Threshold^{EXTFB} is 0.8. This is because the extended feedback method continues to process images after the confidence of a relevant keyword has reached MAXCONF, while irrelevant images are selected to modify the confidence. Fig. 6 indicates that the precision is largest when Threshold^{EXTFB} is 0.7. Consequently, in this paper the Threshold^{EXTFB} value was fixed at 0.7 considering the trade-off between recall and precision.

4.5. Experiments for performance evaluation

4.5.1. Experiment 1: Recall and precision according to the size of the training set

The more images there are in the training set, the better the automatic keyword annotation result becomes. In order to verify the effect of the size of the training set, an experiment was performed with different training set sizes (360, 920 and 1840). Training set sizes of 360, 920 and 1840 represent about 4%, 10% and 20% of the entire image set size used in our experiments, respectively.

To compare the two feedback methods, extended feedback and naive feedback, we first measured their precision and recall values and then presented their differences in percentage form as shown in Formulas (9) and (10). Here, RE (ExtendedFeedback) and RE (NaiveFeedback) denote the recall values of the extended feedback method and the naive feedback methods, respectively. In a similar manner, PR (ExtendedFeedback) and PR (NaiveFeedback) express the precision values of the extended feedback method and the naive feedback method, respectively:

Formula (9). Equation for the growth rate of recall:

$$\text{Growth rate of recall} = \frac{RE(ExtendedFeedback) - RE(NaiveFeedback)}{RE(NaiveFeedback)} \times 100. \tag{9}$$

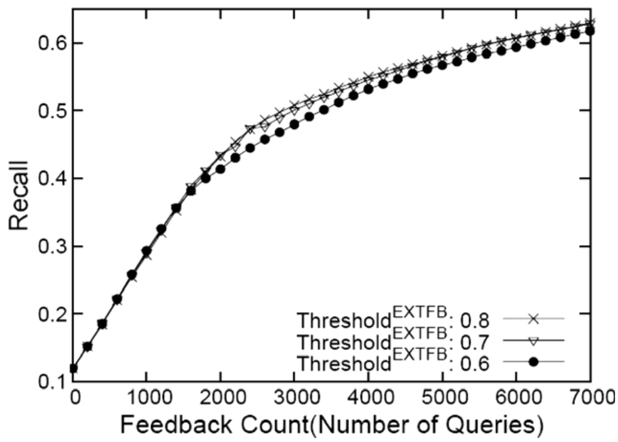

Fig. 5. Comparison of recall values for deciding Threshold^{EXTFB}.

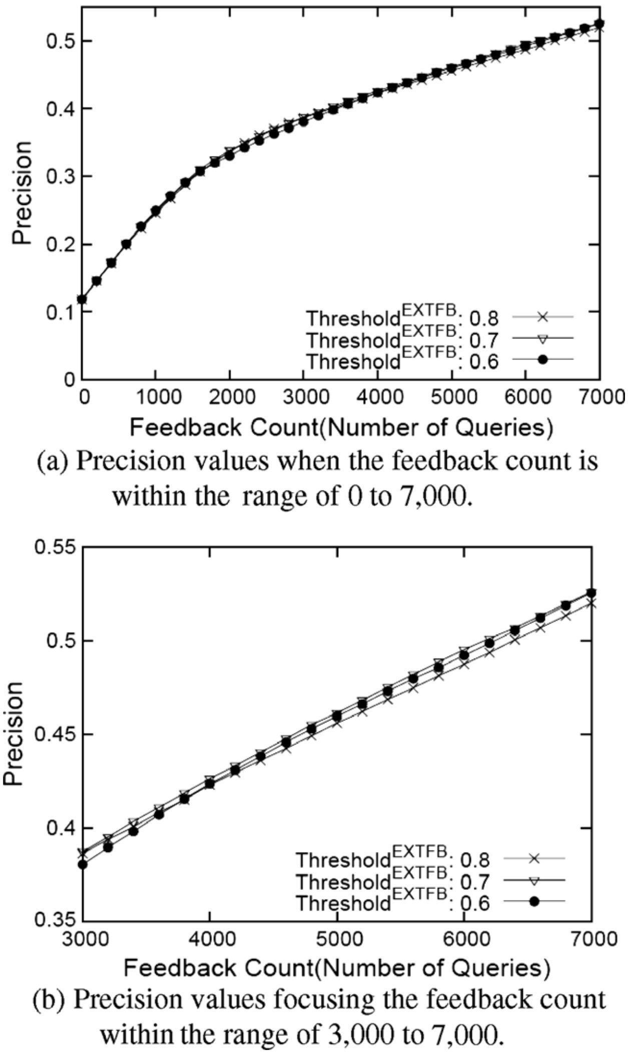

Fig. 6. Comparison of precision values for deciding Threshold^{EXTFB}.

Formula (10). Equation for the growth rate of precision:

$$\text{Growth rate of precision} = \frac{PR(ExtendedFeedback) - PR(NaiveFeedback)}{PR(NaiveFeedback)} \times 100. \tag{10}$$

Figs. 7 and 8 show the growth rates of recall and precision, respectively, with different training set sizes. As shown clearly in both figures, the improvement of the extended feedback method over the naive feedback method is more remarkable in smaller training sets. This is because the confidence is modified more rapidly by the extended feedback method when the keyword of the initial system is not accurate. That is, the confidence modification has been proven to be more effective when the initial images of a database are associated with irrelevant automatic keywords.

4.5.2. Experiment 2: Recall and precision according to feedback size

The most effective way of modifying images is to give feedback to every retrieved image. However, leaving feedback information for every initial retrieval result is practically impossible considering the large amount of images in a database. Therefore, the user selects only a few images from the first result and submits these as feedback information.

The proposed method's performance was compared with that of the naive feedback method while varying the feedback size. The different feedback sizes used in the experiment were 4, 6, and 8. N_p and N_n were set to half of the feedback size. The feedback sizes of 4, 6, and 8 were about 0.04%, 0.06%, and 0.08% of the number of total images, respectively.

Figs. 9 and 10 show the performance result according to the three different feedback sizes. Both figures indicate that the gap between the two methods becomes wider when there is only a small amount of feedback. That is, as the proportion of

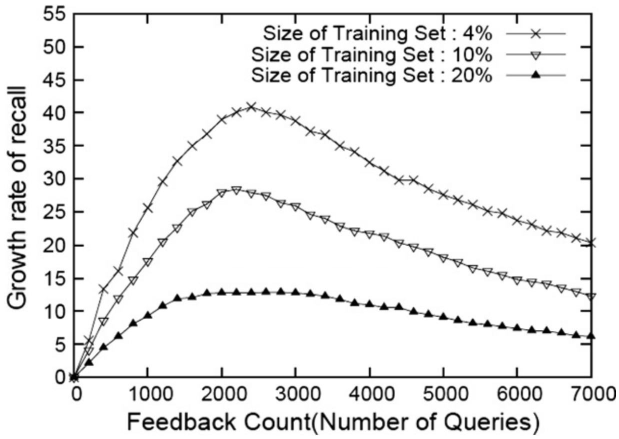

Fig. 7. Comparison of the growth rate of recall according to different training set sizes.

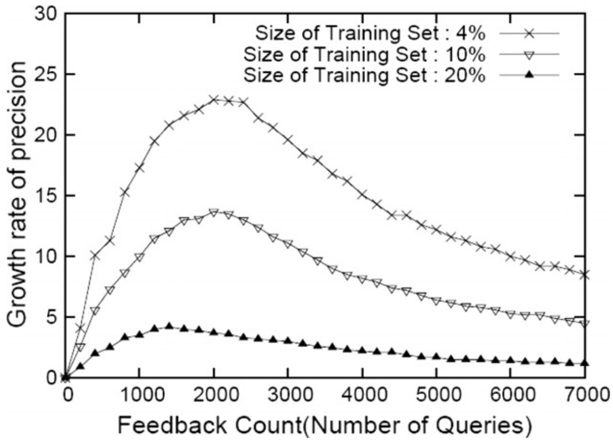

Fig. 8. Comparison of the growth rate of precision according to different training set sizes.

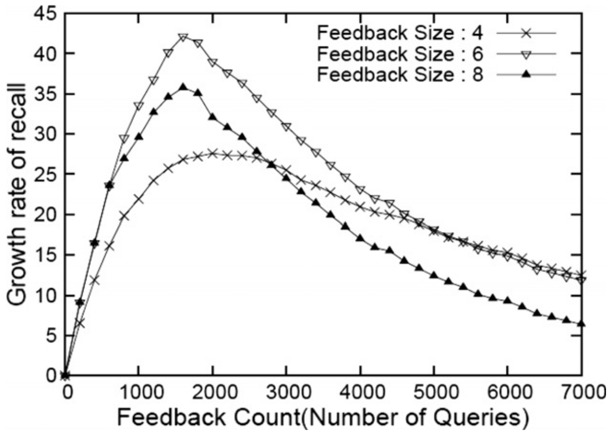

Fig. 9. Comparison of the growth rate of recall according to different feedback sizes.

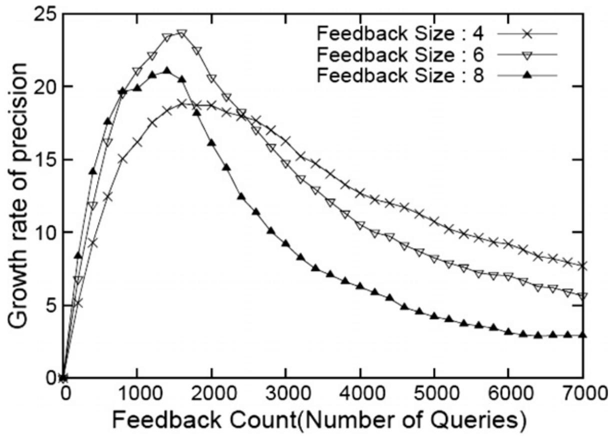

Fig. 10. Comparison of the growth rate of precision according to different feedback sizes.

feedback to total images becomes smaller, the extended feedback method shows better results than the naive feedback method. When the feedback size is 4, the growth rate of recall and precision is as high as 13% and 8%, respectively, after 7000 feedback sessions.

5. Conclusion

In this paper, a new model, which incorporates relevance feedback into keyword-based searches, has been proposed. In the proposed model, not only the images that have been given positive feedback, but also other images with visual features similar to the features used to distinguish positive images, are subjected to confidence modification. This enables the modification of a large number of images with only a few pieces of feedback which ultimately leads to faster and more accurate retrieval results.

The main contribution of this paper can be summarized with two points. (1) A novel search model that integrates keyword-based searches, relevance feedback, and content-based feedback, which ultimately increases retrieval accuracy and shortens retrieval time, has been proposed. (2) A method of automatically deciding the optimum weighted similarity function for content-based searches has been proposed using positive and negative images selected during the relevance feedback process.

A performance evaluation was performed to verify the effectiveness of the proposed method. Experimental results show that recall and precision increase up to 13% and 8%, respectively, with the same amount of feedback.

References

[1] J.R. Bach, C. Fuller, A. Gupta, A. Hampapur, B. Horowitz, R. Humphrey, R. Jain, C. Shu, The virage image search engine: an open framework for image management, in: I.K. Sethi, R.C. Jain (Eds.), Proceedings of the Storage and Retrieval for Still Image and Video Databases, vol. IV, SPIE, 1996, pp. 76–87.
[2] P.J. Cheng, L.F. Chien, Effective image annotation for search using multi-level semantics, Lecture Notes in Computer Science 2911 (2003) 230–242.
[3] S. Deb, Y. Zhang, An overview of CBIR techniques, in: Proceedings of the 18th International Conference on Advanced Information Networking and Applications, AINA'04, IEEE CS Press, 2004, pp. 59–64.
[4] L. Fei-Fei, R. Fergus, P. Perona, Learning generative visual models from few training examples: an incremental Bayesian approach tested on 101 object categories, in: Proceedings of the IEEE CVPR Workshop on Generative-Model Based Vision, CVPRW'04, IEEE Computer Society, 2004, p. 178.
[5] H. Feng, T. Chua, A learning based approach for annotating large online image collection, in: Y. Ping, P. Chen (Eds.), Proceedings of the 10th International Multimedia Modeling Conference, MMM'04, 2004, pp. 249–256.
[6] M. Flinker, H. Samhey, 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 (1995) 23–32.
[7] J. Jeon, V. Lavrenko, R. Manmatha, Automatic image annotation and retrieval using cross-media relevance models, in: Proceedings of the 26th Annual International ACM SIGIR Conference, SIGIR'03, 2003, pp. 119–126.
[8] J. Li, J.Z. Wang, Real-time computerized annotation of pictures, IEEE Transactions on Pattern Analysis and Machine Intelligence 30 (2008).
[9] W. Liu, S. Dumais, Y. Sun, H. Zhang, M. Czerwinski, B. Field, Semi-automatic image annotation, in: Proceedings of the Human–Computer Interaction, HCII'01, 2001, pp. 326–333.
[10] S.C.H. Hoi, M.R. Lyu, R. Jin, A unified log-based relevance feedback scheme for image retrieval, IEEE Transactions on Knowledge and Data Engineering 18 (2006) 509–524.
[11] Y. Lu, C. Hu, X. Zhu, H.J. Zhang, Q. Yang, A unified framework for semantics and feature based relevance feedback in image retrieval systems, in: Proceedings of the ACM Multimedia, ACM Press, 2000, pp. 31–38.
[12] O. Marques, N. Barman, Semi-automatic semantic annotation of images using machine learning techniques, in: D. Fensel, K.P. Sycara, J. Mylopoulos (Eds.), Proceedings of the International Semantic Web Conference ISWC'03, Springer, 2003, pp. 550–565.
[13] MPEG-7 Video Group, MPEG-7 Visual Part of the eXperimentation Model (version 9.0) January 2001, ISO/IEC JTCI/SC29/WG11 N3914.
[14] Y. Rui, T.S. Huang, M. Ortega, S. Mehrotra, Relevance feedback: a power tool for interactive content-based image retrieval, IEEE Transactions on Circuits and Systems for Video Technology 8 (1998) 644–655.

[15] J.R. Smith, S.F. Chang, VisualSeek: a fully automated content-based image query system, in: Proceedings of the ACM International Conference on Multimedia, 1996, pp. 87–93.
[16] W.T. Wong, F.Y. Shih, J. Liu, Shape-based image retrieval using support vector machined, Fourier descriptors and self-organizing maps, Information Sciences 177 (2007) 1878–1891.
[17] J. Zachary, S.S. Iyengar, J. Barhen, Content based image retrieval and information theory: a general approach, Journal of the American Society for Information Science and Technology 52 (2001) 840–852.
[18] C. Zhang, J.Y. Chai, R. Jin, User term feedback in interactive text-based image retrieval, in: Proceedings of the 28th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, 2005, pp. 51–58.
[19] X. Zhou, T.S. Huang, Relevance feedback in content-based image retrieval: some recent advances, Information Sciences 148 (2002) 129–137.