Comparison and Combination of Textual and Visual Features for Interactive Cross-Language Image Retrieval

Pei-Cheng Cheng[1], Jen-Yuan Yeh[1], Hao-Ren Ke[2],
Been-Chian Chien[3], and Wei-Pang Yang[1, 4]

[1] Department of Computer & Information Science, National Chiao Tung University,
1001 Ta Hsueh Rd., Hsinchu, Taiwan 30050, R.O.C.
{cpc, jyyeh, wpyang}@cis.nctu.edu.tw
[2] University Library, National Chiao Tung University,
1001 Ta Hsueh Rd., Hsinchu, Taiwan 30050, R.O.C.
claven@lib.nctu.edu.tw
[3] Department of Computer Science and Information Engineering,
National University of Tainan,
33, Sec. 2, Su Line St., Tainan, Taiwan 70005, R.O.C.
bcchien@mail.nutn.edu.tw
[4] Department of Information Management, National Dong Hwa University,
1, Sec. 2, Da Hsueh Rd., Shou-Feng, Hualien, Taiwan 97401, R.O.C.
wpyang@mail.ndhu.edu.tw

Abstract. This paper concentrates on the user-centered search task at ImageCLEF 2004. In this work, we combine both textual and visual features for cross-language image retrieval, and propose two interactive retrieval systems – T_ICLEF and VCT_ICLEF. The first one incorporates a relevance feedback mechanism based on textual information while the second one combines textual and image information to help users find a target image. The experimental results show that VCT_ICLEF had a better performance in almost all cases. Overall, it helped users find the topic image within a fewer iterations with a maximum of 2 iterations saved. Our user survey also reported that a combination of textual and visual information is helpful to indicate to the system what a user really wanted in mind.

1 Introduction

The ImageCLEF campaign [2] under the CLEF[1] (Cross-Language Evaluation Forum) conducts a series of evaluations on systems which are built to accept a query in one language and to find images with relevant captions in different languages. In 2004, three tasks were proposed based on different domains, scenarios, and collections: (1) *the bilingual ad hoc retrieval task*, (2) *the medical retrieval task*, and (3) *the user-centered search task*.

This paper concentrates on the user-centered search task. The task follows the scenario that a user is searching with a specific image in mind, but without any key information about it. The goal is to determine whether the retrieval system is being used

[1] The official website is available at http://clef.iei.pi.cnr.it:2002/.

C. Peters et al. (Eds.): CLEF 2004, LNCS 3491, pp. 793 – 804, 2005.

in the manner intended by the designers as well as to determine how the interface helps users reformulate and refine their search topics. We proposed two systems: (1) *T_ICLEF*, and (2) *VCT_ICLEF* to address the task. T_ICLEF is a cross-language image retrieval system, which is simply enhanced with a relevance feedback mechanism; VCT_ICLEF is practically T_ICLEF but provides a color table that allows users to indicate color information about the target image. Our principal objective is to compare and to combine textual and visual features under an interactive cross language image searching situation.

In the following sections, the overview of the proposed interactive search process is described. Section 2 introduces previous work on query reformulation. Sections 3-4 illustrate the proposed methods for the interactive search task, and some preliminary results are presented in Section 5. Finally, we finish with a conclusion in Section 6.

1.1 Overview of the Proposed Interactive Search Process

Fig. 1 shows an overview of the proposed interactive search process. Given an initial query, $Q = (Q_T, Q_I)$, in which Q_T denotes a Chinese text query, and Q_I stands for a query image, the system performs cross-language image retrieval and returns a set of "relevant[2]" images to the user. The user then evaluates the relevance of the returned images, and gives a relevance value to each of them. This process is called relevance feedback. At the following stage, the system invokes the query reformulation process to derive a new query, $Q' = (Q'_T, Q'_I)$. The new query is believed to be closer to the user's information need. Finally, the system performs once again image retrieval according to Q'. The process iterates until the user finds the target image.

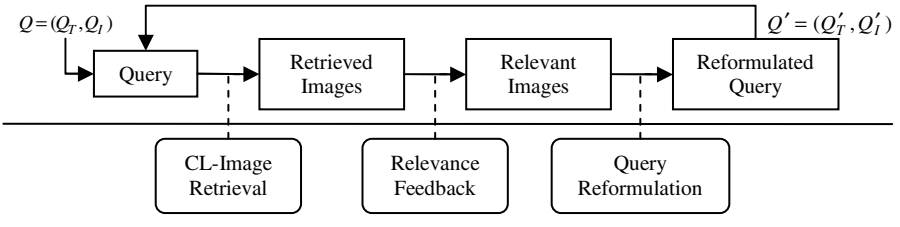


Fig. 1. Overview of the proposed user-centered search process

2 Previous Work

Previous work on image retrieval usually exploits low-level features, for example, *color*, *texture*, *shape*, etc., extracted from an image to measure its similarity to the query (e.g., [4]). However, the retrieval performance is always limited due to the gap between semantic concepts which are explained as humans' perceptions, and low-level image features used to represent an image. Recently, relevance feedback has

[2] The degree of relevance is judged by the similarity metric used in the retrieval system.

been successfully employed to alleviate the above-mentioned problem (e.g., [3] [8] [9] [12]).

Previous work (e.g., [7]) has shown that interactive search helps improve recall and precision in the retrieval task. Some work defined a new weighted query by associating more significant features with larger weights, and less important ones with smaller weights. The strategy is mostly used (e.g., [5], [12], [13]). For example, [13] proposed a low-level feature-based relevance feedback framework, in which for each feature *i*, an ideal query, q_i, is modeled as a weighted sum of all positive examples, which is shown in Eq. (1)

$$q_i = \frac{\pi^T Y_i}{\sum_{j=1}^{n} \pi_j} \tag{1}$$

where Y_i an $n \times K_i$ training matrix for feature *i*, which is obtained by stacking the first *n* positive examples, K_i the length of feature *i*. The *n*-dimension vector, $\pi = [\pi_1, ..., \pi_n]$, represents a relevance degree for the *n* positive images. Ciocca *et al.* (2002) [1] proposed a novel query reformulation method for relevance feedback. After the relevant images are selected, they contribute their features to a new query feature vector when their similarities to the average of all relevant images are significantly large. The new query feature vector is the average of the contributing features.

There are still other researches which address relevance feedback with probabilistic models. For example, Cox *et al.* (2000) [3], Vasconcelos and Lippman (1999) [16], Meilhac and Nastar (1999) [9], employed Bayesian estimation to update the probability distribution of all images. The main idea is to consider feedback examples as a sequence of independent queries and to try to minimize the retrieval errors by Bayes' rule.

3 Cross-Language Image Retrieval

In this section, we describe how to create the representation for an image or a query, and how to compute the similarity between an image and the query on the basis of their representations.

3.1 Image/Query Representations

We represent both an image and a query as a vector in the vector space model [14]. First of all, we explain the symbols used in the following definitions of representations. $P = (P_T, P_I)$ denotes an image where P_T and P_I stand for the captions of *P* and the image *P* respectively, and $Q = (Q_T, Q_I)$ represents a query, which is defined in Section 1.1. In our proposed approach, a textual vector representation, such as P_T and Q_T, is modeled in terms of three distinct features – *term*, *category*, and *temporal* information, whilst an image vector representation, for example, P_I and Q_I, is represented with a color histogram.

Textual Vector Representation

Let W ($|W| = n$) the set of significant keywords in the corpus, C ($|C| = m$) the set of categories defined in the corpus, and Y ($|Y| = k$) the set of publication years of all images. For an image P, its textual vector representation (i.e., P_T) is defined as Eq. (2),

$$P_T = < w_{t_1}(P_T),...,w_{t_n}(P_T),w_{c_1}(P_T),...,w_{c_m}(P_T),w_{y_1}(P_T),...,w_{y_k}(P_T) > \quad (2)$$

where the first n dimensions indicate the weighting of a keyword t_i in P_T, which is measured by TF-IDF [14], as computed in Eq. (3); the following n+1 to n+m dimensions indicate whether P belongs to a category c_i, which is shown as Eq. (4); the final n+m+1 to n+m+k dimensions express whether P was published in year y_i, which is defined as Eq. (5).

$$w_{t_i}(P_T) = \frac{tf_{t_i,P_T}}{\max tf} \times \log \frac{N}{n_{t_i}} \quad (3)$$

$$w_{c_i}(P_T) = \begin{cases} 1 & \text{if } P \text{ belongs to } c_i, \\ 0 & \text{otherwise} \end{cases} \quad (4)$$

$$w_{y_i}(P_T) = \begin{cases} 1 & \text{if } P \text{ was published in the year } y_i, \\ 0 & \text{otherwise} \end{cases} \quad (5)$$

In Eq. (3), $\frac{tf_{t_i,P_T}}{\max tf}$ stands for the normalized frequency of t_i in P_T, maxtf is the maximum number of occurrences of any keyword in P_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. Regarding Eq. (4) and Eq. (5), both of them compute the weighting of the category and the temporal feature as a Boolean value.

In the above, we introduce how to create a textual vector representation for P_T. As for a query Q, one problem is that since Q_T is given in Chinese, it is necessary to translate Q_T into English, which is the language used in the image collection. We first perform the word segmentation process to obtain a set of Chinese words. For each Chinese word, it is then translated into one or several corresponding English words by looking it up in a dictionary. The dictionary that we use is pyDict[3]. Up to now, it is still hard to determine the correct translation; therefore, we tend to keep all English translations in order not to lose the consideration of any correct word.

Another problem is the so-called short query problem. A short query usually cannot cover many useful search terms because of the lack of sufficient words. We address this problem by performing the query expansion process to add new terms to the original query. The additional search terms are taken from a thesaurus – WordNet [10]. For each English translation, we include its *synonyms*, *hypernyms*, and *hyponyms* into the query.

[3] An English/Chinese dictionary written by D. Gau, which is available at http://sourceforge.net/projects/pydict/.

It comes out as a new problem. Assume $AfterExpansion(Q_T) = \{e_1,...,e_h\}$ is the set of all English words obtained after query translation and query expansion, 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 exploit *word co-occurrence relationships* to determine final query terms. The main idea is 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 for monolingual image retrieval. So far, we have a set of search terms, $AfterDisambiguity(Q_T)$, which is presented as Eq. (6),

$$AfterDisambiguity(Q_T) = \{e_i, e_j \mid e_i, e_j \in AfterExpansion(Q_T) \\ \& \ e_i, e_j \text{ have a significant cooccurrence}\} \tag{6}$$

After giving the definition of $AfterDisambiguity(Q_T)$, for a query Q, its textual vector representation (i.e., Q_T) is defined in Eq. (7),

$$Q_T = < w_{t_1}(Q_T),...,w_{t_n}(Q_T), w_{c_1}(Q_T),...,w_{c_m}(Q_T), w_{y_1}(Q_T),...,w_{y_k}(Q_T) > \tag{7}$$

where $w_{t_i}(Q_T)$ is the weighting of a keyword t_i in Q_T, which is measured as Eq. (8), $w_{c_i}(Q_T)$ indicates whether there exists an $e_j \in AfterDisambiguity(Q_T)$ and it also occurs in a category c_i, which is shown as Eq. (9), and $w_{y_i}(Q_T)$ presents whether there is an $e_j \in AfterDisambiguity(Q_T)$, e_j is a temporal term, and e_j satisfies a condition caused by a predefined temporal operator.

In Eq. (8), $\frac{tf_{t_i,Q_T}}{\max tf}$ 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. Eq. (9) and Eq. (10) compute the weighting of the category and the temporal feature as a Boolean value.

$$w_{t_i}(Q_T) = \begin{cases} \frac{tf_{t_i,Q_T}}{\max tf} \times \log \frac{N}{n_{t_i}} \end{cases} \tag{8}$$

$$w_{c_i}(Q_T) = \begin{cases} 1 & \text{if } \exists j, e_j \in AfterDisambiguity(Q_T) \text{ and } e_j \text{ occurs in } c_i, \\ 0 & \text{otherwise} \end{cases} \tag{9}$$

$$w_{y_i}(Q_T) = \begin{cases} 1 & \text{if } Q_T \text{ contains "Y年以前," and } y_i \text{ is BEFORE Y,} \\ 1 & \text{if } Q_T \text{ contains "Y年之中," and } y_i \text{ is IN Y,} \\ 1 & \text{if } Q_T \text{ contains "Y年以後," and } y_i \text{ is AFTER Y,} \\ 0 & \text{otherwise} \end{cases} \tag{10}$$

To be mentioned, with regard to $w_{y_i}(Q_T)$, three operators – "BEFORE," "IN," and "AFTER" – are defined to take into account a query such as "1900 年以前拍攝的愛丁堡城堡的照片 (Pictures of Edinburgh Castle taken before 1900)," which also concerns time. Take, for example, the above query that targets only images taken before 1900; a part of the textual vector of the above query about the temporal feature is given in Table 1, it gives an idea that P_1 will be retrieved since its publication year was in 1899 while P_2 will not be retrieved because of its publication year, 1901. Note that in our current implementation, we only consider *years* for the temporal feature. Hence, for a query like "1908 年四月拍攝的羅馬照片 (Photos of Rome taken in April 1908)," "四月 (April)" is treated as a general term, which only contributes its effect to the term feature.

Table 1. An example which shows how the time operators work while considering the time dimension

Year	…	1897	1898	1899	1900	1901	1902	…
P_1	0	0	0	1	0	0	0	0
P_2	0	0	0	0	0	1	0	0
Q_T	1	1	1	1	0	0	0	0

Image Vector Representation

Color histogram [15] is a basic method and has good performance for representing the visual contents of an image. 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 to be closer to human perceptions 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_1 (Q_1) is represented as Eq. (11), which models the color histogram $H(P_1)$ ($H(Q_1)$) as a vector, in which each bucket h_{c_i} counts the ratio of pixels of P_1 (Q_1) in color c_i.

$$P_1 = < h_{c_1}(P_1),...,h_{c_m}(P_1) >,\ Q_1 = < h_{c_1}(Q_1),...,h_{c_m}(Q_1) > \tag{11}$$

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. (12),

$$h_{c_i}(P_1) = \frac{\sum_{p \in P_1} \frac{|\alpha_p - \beta_p|}{\delta}}{|P_1|} \tag{12}$$

Fig. 2 gives an example to explain what we call a partial pixel. In the figure, c_{i-1}, c_i, and c_{i+1} stand for a color bin, a solid line indicates 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.

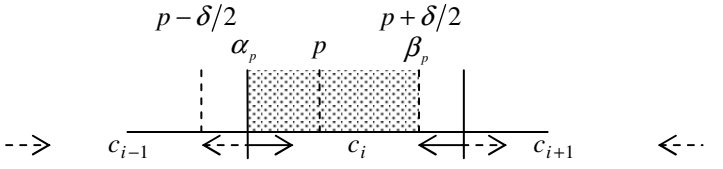


Fig. 2. An illustration of the partial pixel idea

3.2 Similarity Metric

While a query $Q=(Q_T, Q_I)$ and an image $P=(P_T, P_I)$ are represented in terms of a textual and an image vector representation, we propose two strategies to measure the similarity between the query and each image in the collection. In the following, we briefly describe the proposed strategies: Strategy 1, which is exploited in the system T_ICLEF, only takes into account the textual similarity while Strategy 2[4], which combines the textual and the image similarity, is employed in the system VCT_ICLEF.

Strategy 1 (T_ICLEF): Based on the textual similarity

$$Sim_1(P,Q)=\frac{\vec{P}_T\cdot\vec{Q}_T}{|\vec{P}_T||\vec{Q}_T|} \tag{13}$$

Strategy 2 (VCT_ICLEF): Based on both the textual and the image similarity

$$Sim_2(P,Q)=\alpha\cdot Sim_1(P,Q)+\beta\cdot Sim_3(P,Q)$$

where

$$Sim_3(P,Q)=\frac{H(P_I)\bigcap H(Q_I)}{|H(Q_I)|}=\frac{\sum_i \min(h_{c_i}(P_I), h_{c_i}(Q_I))}{\sum_i h_{c_i}(Q_I)} \tag{14}$$

[4] In our implementation, α is set to 0.7, and β is set to 0.3.

4 Interactive Search Mechanism

In this section a user interface for each proposed system is introduced. Then, the proposed query reformulation methods are described regarding textual and visual queries.

4.1 User Interface

Fig. 3 and Fig. 4 demonstrate the user interfaces designed for the user-centered search task at ImageCLEF 2004. Both systems have a search panel on the top, which allows users to type a Chinese query. In the display area, a pull-down menu below each image allows users to feedback the relevance of each image, which is provided as "*non-relevant*," "*neutral*," and "*relevant*." In our design, the system first returns 80 images for the initial search, but 40 images in later iterations. This is because in the initial search the system does not develop an idea about what the user wants exactly. A further set of images may induce the user to mark more relevant images and to assist the system to reformulate the query.

In fact, it is the color table shown in VCT_ICLEF that distinguishes the two systems. Users can provide color information to help the system determine the best query strategy. According to the experimental results, VCT_ICLEF has a better performance by exploiting color information for searching.

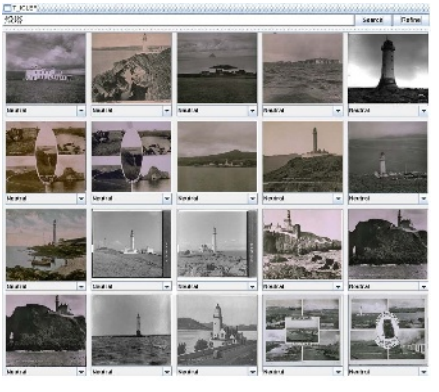


Fig. 3. The user interface of T_ICLEF

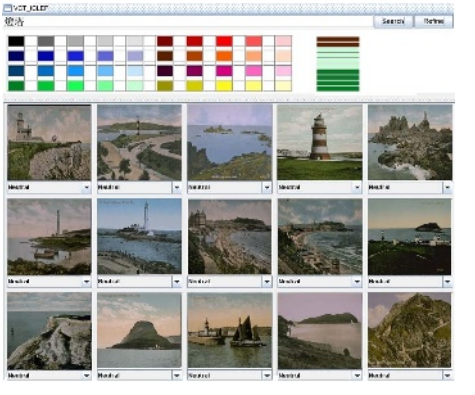


Fig. 4. The user interface of VCT_ICLEF

4.2 Query Reformulation

As mentioned before, in the relevance feedback process, the user evaluates the relevance of the returned images, and gives a relevance value (i.e., non-relevant, neutral, and relevant) to each of them. At the next stage, the system performs query reformulation to modify the original query on the basis of the user's relevance judgments, and invokes cross-language image retrieval again based on the new query.

Recall that we denote the original query as $Q = (Q_T, Q_I)$ and the new query $Q' = (Q'_T, Q'_I)$; regarding Q'_T, we exploit a practical method, as shown in Eq. (15),

for query reformulation. This mechanism, which has been suggested by [11], is achieved with a weighted query by adding useful information extracted from relevant images as well as decreasing useless information derived from non-relevant images to the original query. Regarding Q_I', it is computed as the centroid of the relevant images, which is defined as their average. We do not take into account the irrelevant images for Q_I', since in our observation there is always a large difference among the non-relevant images. Empirically, adding the irrelevant information to Q_I' makes no helpful contribution.

$$Q_T' = \alpha \cdot Q_T + \frac{\beta}{|REL|} \sum_{P_{i_T} \in REL} P_{i_T} - \frac{\gamma}{|NREL|} \sum_{P_{j_T} \in NREL} P_{j_T} \tag{15}$$

$$Q_I' = \frac{1}{|REL|} \sum_{P_I \in REL} P_I \tag{16}$$

In Eq. (15) and Eq. (16), $\alpha, \beta, \gamma \geq 0$ are parameters, *REL* and *NREL* stand for the sets of relevant ad irrelevant images marked by the user.

5 Evaluation Results

In this section we present our evaluation results for the user-centered search task at ImageCLEF 2004. The collection used for evaluation is the St. Andrews historic photographs. For detailed information about the St. Andrews Collection, the topic images, and the evaluation methodology, please refer to [2].

5.1 The Searchers' Backgrounds

There are 8 people involved in the task, including 5 male and 3 female searchers. Their average age is 23.5, with the youngest 22 and the oldest 26. Three of them major in computer science, two major in social science and the others are librarians. In particular, three searchers have experiences in participating in projects about image retrieval. All of them have an average of 3.75 years (with a minimum of 2 years and a maximum of 5 years) accessing online search services, specifically for Web search. On average, they search approximately 4 times a week, with a minimum of once and a maximum of 7 times. However, only a half of them have experiences in using image search services, such as Google images search.

5.2 Results

We are interested in which system helps searchers find a target image most effectively. We summarize the average number of iterations[5] and the average time spent by a searcher for each topic in Fig. 5. In the figure, it does not give information in the

[5] Please note that our system does not have an efficient performance; since for each iteration it spent about 1 minute to retrieve relevant images, approximately 5 iterations is performed within the time limit.

case that all searchers did not find the target image. (For instance, regarding topic 2, all searchers failed to complete the task by using T_ICLEF within the definite time.) The figure shows that overall VCT_ICLEF helps users find the image within a fewer iterations with a maximum of 2 iterations saved. For topics 2, 5, 7, 11, 15 and 16, no searcher can find the image by making use of T_ICLEF. With regard to topics 10 and 12, VCT_ICLEF has a worse performance. In our observation, the reason is that most images (82%) in the corpus are in black and white, once the user gives imprecise color information, VCT_ICLEF needs to take more iterations to find the image consequently.

Table 2 presents the number of searchers who failed to find the image for each topic. It is clear that VCT_ICLEF outperforms T_ICLEF in almost all cases. Considering topic 3, we believe that it is caused by the same reason we mentioned above for topics 10 and 12. Finally, we give a summary of our proposed systems in Table 3. The table illustrates that while considering those topics that at least one searcher completed the task, T_ICLEF cost additional 0.4 iterations and 76.47 seconds. By using VCT_ICLEF, on average 89% of searchers successfully found the image, while when using T_ICLEF, around 56.25% of searchers were successful.

To show the effects of color information used in VCT_ICLEF, we take Fig. 3 and Fig. 4 as examples. Regarding topic 6, the query used was "燈塔 (Lighthouse)." For T_ICLEF, it returned a set of images corresponding to the query; however, the target image could not be found in the top 80 images. Since topic 6 is a color image, while we searched the image with color information using VCT_ICLEF, the image was found in the first iteration. We conclude that color information can assist the user to tell the system what he is searching for. For an interactive image retrieval system, it is necessary to provide users not only an interface to issue a textual query but also an interface to indicate the system the visual information of the target.

Table 2. Number of searchers who did not find the target image for each topic

Topic	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
T_ICLEF	1	4	1	0	4	0	4	1	0	1	4	0	0	0	4	4
VCT_ICLEF	1	0	2	0	0	0	0	0	0	0	2	0	0	0	2	0

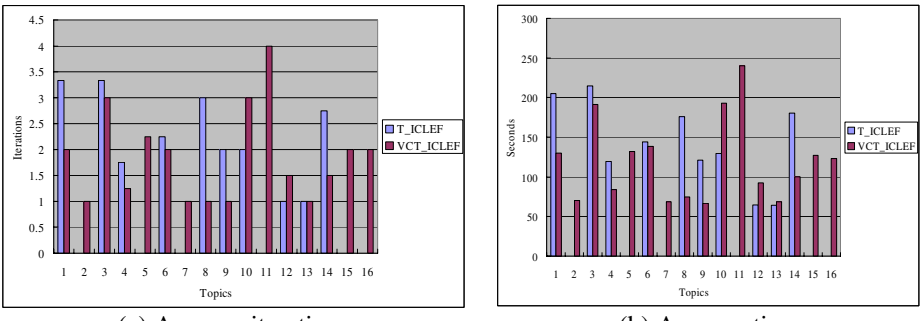
Table 3. Average steps to find the target image, and the average spent time

	Avg. Iterations (Not including not found)	Avg. Spent Time for each topic	Avg. percent of searchers who found the target image (#/4×100%)
T_ICLEF	2.24	208.67s	56.25%
VCT_ICLEF	1.84	132.20s	89.00%

5.3 Search Strategies

In our survey of search strategies exploited by searchers, we found that 5 searchers thought that additional color information about the target image was helpful to indicate to the system what they really wanted. Four searchers preferred to search the image with a text query first, even when using VCT_ICLEF. They then considered color

information for the next iteration in the situation that the target image was in color but the system returned images all in black and white. When searching for a color image, 3 searchers preferred to use color information first. Moreover, 2 searchers hoped that in the future, users can provide a textual query to indicate color information, such as "黃色 (Yellow)." Finally to be mentioned, in our systems the user is allowed to provide a query consisting of temporal conditions. However, since it is hard to decide in which year the image was published, no one used a query which contains temporal conditions.



(a) Average iterations (b) Average time

Fig. 5. The average number of iterations and time spent searching for each topic

6 Conclusions

We participated in the user-centered search task at ImageCLEF 2004. In this paper, we proposed two interactive cross-language image retrieval systems – T_ICLEF and VCT_ICLEF. The first one is implemented with a practical relevance feedback approach based on textual information while the second one combines textual and image information to help users find a target image. The experimental results show that VCT_ICLEF has a better performance than T_ICLEF in almost all cases. Overall, VCT_ICLEF helps users find the image within a fewer iterations with a maximum of 2 iterations saved.

In the future, we plan to investigate user behaviors to understand in which cases users prefer a textual query as well as in which situations users prefer to provide visual information for searching. Besides, we also intend to implement a SOM (Self-Organizing Map) [6] on image clustering, which we believe that it can provide an effective browsing interface to help searchers find a target image.

Acknowledgements

This research was partially supported and sponsored by National Science Council, R.O.C. under NSC93-2213-E-009-044.

References

1. Ciocca G., Gagliardi I., Schettini R.: Quicklook2: An integrated multimedia system. International Journal of Visual Languages and Computing, Special Issue on Querying Multiple Data Sources, Vol. 12, No. 1 (2001), 81-103.
2. Clough P., Sanderson M., Müller H.: The CLEF Cross Language Image Retrieval Track (ImageCLEF) 2004. Proceedings of the CLEF 2004 Workshop, Bath, UK (2004), 459-473.
3. Cox I. J., Minka T. P., Papathomas T. V., Yianilos P. N.: The Bayesian image retrieval system, pichunter: Theory, implementation, and psychophysical experiments. IEEE Transactions on Image Processing, Special Issue on Digital Libraries, Vol. 9, No. 1 (2000), 20-37.
4. Flickner M., Sawhney H., Niblack W., Ashley J., Huang Q., Dom B., Gorkani M., Hafner J., Lee D., Petkovic D., Steele, D., Yanker P.: Query by Image and Video Content: The QBIC system. IEEE Computer, Vol. 28, No. 9 (1995), 23-32.
5. Ishikawa Y., Subramanya R., Faloutsos C.: Mindreader: Query databases through multiple examples. Proceedings of 24th VLDB Conference, New York, (1998), 218-227.
6. Kohonen T.: The Self-Organizing Map. Neurocomputing, Vol. 21, No. 1-3 (1998), 1-6.
7. Kushki A., Androutsos P., Plataniotis K. N., Venetsanopoulos A. N.: Query Feedback for Interactive Image Retrieval. IEEE Transactions on Circuits and Systems for Video Technology, Vol. 14, No. 5 (2004), 644-655.
8. Lu Y., Hu C., Zhu X., Zhang H., Yang Q.: A unified framework for semantics and feature based relevance feedback in image retrieval systems. Proceedings of 8th ACM Multimedia International Conference, Los Angeles, CA (2000), 31-37.
9. Meilhac C., Nastar C.: Relevance feedback and category search in image databases. IEEE Int. Conf. Multimedia Computing and Systems (1999), 512-517.
10. Miller G.: WordNet: A Lexical Database for English. Communications of the ACM (1995), 39-45.
11. Rocchio J. J.: Relevance feedback in information retrieval. In Salton G. (ed.): The SMART Retrieval System: Experiments in Automatic Document Processing, Englewood Cliffs, NJ: Prentice-Hall (1971), 313-323.
12. Rui Y., Huang T. S.: Relevance feedback: A power tool for interactive content-based image retrieval. IEEE Circuits Syst. Video Technol., Vol. 8, No. 5 (1999), 644-655.
13. Rui Y., Huang T. S.: A novel relevance feedback technique in image retrieval. Proceedings of the 7th ACM International Conference on Multimedia, Orlando, FL (1999), 67-70.
14. Salton G., McGill M. J. (eds.): Introduction to Modern Information Retrieval. McGraw-Hill (1983).
15. Swain M. J., Ballard D. H.: Color Indexing. International Journal of Computer Vision, Vol. 7 (1991), 11-32.
16. Vasconcelos N., Lippman A.: Learning from user feedback in image retrieval systems. Proc. NIPS'99, Denver, CO (1999).