NCTU-ISU's Evaluation for the User-Centered Search Task at ImageCLEF 2004

Pei-Cheng Cheng[a], Jen-Yuan Yeh[a], Hao-Ren Ke[b], Been-Chian Chien[c,d] and Wei-Pang Yang[a,e]

[a]Department of Computer & Information Science, National Chiao Tung University,
1001 Ta Hsueh Rd., Hsinchu, TAIWAN 30050, R.O.C.
{cpc, jyyeh, gis91610, wpyang}@cis.nctu.edu.tw

[b]University Library, National Chiao Tung University,
1001 Ta Hsueh Rd., Hsinchu, TAIWAN 30050, R.O.C.
claven@lib.nctu.edu.tw

[c]Department of Information Engineering, I-Shou University
1, Sec. 1, Hsueh Cheng Rd., Ta-Hsu Hsiang, Kaohsiung, TAIWAN 84001, R.O.C.
cbc@isu.edu.tw

[d]Department of Computer Science and Information Engineering, National University of Tainan
33, Sec. 2, Su Line St., Tainan, TAIWAN 70005, R.O.C.

[e]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. 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.

Keywords: Interactive search; Cross-Language image retrieval; Relevance feedback; User behavior.

1 Introduction

The ImageCLEF campaign under the CLEF[1] (Cross-Language Evaluation Forum) is conducting a series of evaluations on systems which are built to accept a query in a language and to find relevant images with captions in different languages. In this year, three tasks[2] are proposed based on different domains, scenarios, and collections. They are 1) *the bilingual ad hoc retrieval task*, 2) *the medical retrieval task*, and 3) *the user-centered search task*. The first is to perform bilingual retrieval against a photographic collection in which images are accompanied with captions in English. The second is, given an example image, to find out similar images from a medical image database which consists of images such as scans and x-rays. The last one aims to assess user interaction for a known-item or target search.

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. Previous work (e.g, [Kushki04]) has shown that interactive search helps improve recall and precision in the retrieval task. However, in this paper, the goal is to determine whether the retrieval system is being used in the manner intended by the designers as well as to determine how the interface helps users reformulate and refined 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 enhanced with the relevance feedback mechanism; VCT_ICLEF is practically T_ICLEF but provides a color table which allows users to indicate color information about the target image.

In the following sections, the design of our systems is first described. We then introduce the proposed methods for the interactive search task, and present our results. Finally, we finish with a conclusion and a discus-

[1] The official website is available at http://clef.iei.pi.cnr.it:2002/.
[2] Please refer to http://ir.shef.ac.uk/imageclef2004/ for further information.

sion of future work.

2 Overview of the Interactive Search Process

The overview of the interactive search process is shown in Fig. 1. 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 the Cross-Language image retrieval, and returns a set of "relevant" images to the user. The user then evaluates the relevance of the returned images, and gives a relevance value to each of them. The process is called relevance feedback. In our system, the user is only able to indicate as an image "*non-relevant*," "*neutral*," and "*relevant*." 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 more corresponding to the user's need. Finally, the system performs again image retrieval according to Q'. The process iterates until the user finds the target image.

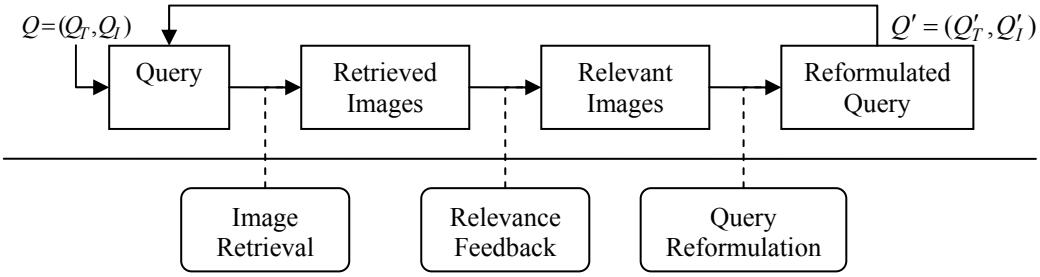


Fig. 1. The overview of the user-centered search process.

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 create both an image and a query representation by representing them as a vector in the vector space model [Salton83]. 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 as mentioned before. 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 while an image vector representation, for example, P_I and Q_I, is represented with color histogram.

Textual Vector Representation

Let W ($|W| = n$) the set of significant keywords in the corpus, C ($|C| = m$) the set of categories which are defined in the corpus, and Y ($|Y| = k$) the set of publication years of all images, for an image P, it textual vector representation (i.e., P_T) is defined as Eq. (1),

$$P_T = < w_{t_1}(P_T), ..., w_{t_n}(P_T), w_{c_1}(P_T), ..., w_{c_m}(P_T), w_{y_1}(P_T), ..., w_{y_k}(P_T) > \quad (1)$$

where the first n dimensions the weighting of a keyword t_i in P_T, which is measured by TF-IDF [Salton83], as computed in Eq. (2); the following n+1 to n+m dimensions indicate whether P belongs to a category c_i, which is shown as Eq. (3); the final n+m+1 to n+m+k dimensions present whether P was published in y_i, which is defined as Eq. (4).

$$w_{t_i}(P_T) = \frac{tf_{t_i,P_T}}{\max tf} \times \log \frac{N}{n_{t_i}} \tag{2}$$

$$w_{c_i}(P_T) = \begin{cases} 1 & \text{if } P \text{ belongs to } c_i, \\ 0 & \text{otherwise} \end{cases} \tag{3}$$

$$w_{y_i}(P_T) = \begin{cases} 1 & \text{if } P \text{ was published in } y_i, \\ 0 & \text{otherwise} \end{cases} \tag{4}$$

In Eq. (2), $\frac{tf_{t_i,P_T}}{\max tf}$ 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. (3) and Eq. (4), both of them compute the weighting of the category and the temporal feature as a Boolean value. In other words, considering a category c_i and a year y_i, $w_{c_i}(P_T)$ and $w_{y_i}(P_T)$ are set to 1 if and only if P belongs to c_i and P was published in y_i respectively.

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 up it in a dictionary. The dictionary that we use is pyDict[3]. Up to now, it is hard to determine the translation as correctly as possible. 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 can not cover as many useful search terms as possible 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 is taken from a thesaurus – WordNet [Miller95]. For each English translation, we include its synonyms, hypernyms, and hyponyms into the query.

Now, it comes out a new problem. Assume $AfterExpansion(Q_T) = \{e_1, ..., e_h\}$ 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. 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. (5),

$$AfterDisambiguity(Q_T) = \{e_i, e_j \mid e_i, e_j \in AfterExpansion(Q_T) \\ \& \; e_i, e_j \text{ have a significant cooccurrence}\} \tag{5}$$

After giving the definition of $AfterDisambiguity(Q_T)$, for a query Q, its textual vector representation (i.e., Q_T) is defined in Eq. (6),

$$Q_T = < w_{t_1}(Q_T), ..., w_{t_n}(Q_T), w_{c_1}(Q_T), ..., w_{c_m}(Q_T), w_{y_1}(Q_T), ..., w_{y_k}(Q_T) > \tag{6}$$

where $w_{t_i}(Q_T)$ is the weighting of a keyword t_i in Q_T, which is measured as Eq. (7), $w_{c_i}(Q_T)$ 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. (8), 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.

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

In Eq. (7), W is the set of significant keywords as defined before, $\frac{tf_{t_i,Q_T}}{\max tf}$ stands for the normalized frequency of t_i in $AfterDisambiguity(Q_T)$, maxtf is the maximum number of occurrences of any keyword in $AfterDisambiguity(Q_T)$, N indicates the number of images in the corpus, and n_{t_i} denotes the number of images in whose caption t_i appears. Similar to Eq. (3) and Eq. (4), both Eq. (8) and Eq. (9) compute the weighting of the category and the temporal feature as a Boolean value.

$$w_{t_i}(Q_T) = \begin{cases} \frac{tf_{t_i,Q_T}}{\max tf} \times \log \frac{N}{n_{t_i}} \end{cases} \tag{7}$$

$$w_{c_i}(Q_T) = \begin{cases} 1 & \text{if } \exists j, e_j \in AfterDisambiguity(Q_T) \text{ and } e_j \text{ occurs in } c_i, \\ 0 & \text{otherwise} \end{cases} \tag{8}$$

$$w_{y_i}(Q_T) = \begin{cases} 1 & \text{if } Q_T \text{ contains ″Y年以前,″ and } y_i \text{ is before Y}, \\ 1 & \text{if } Q_T \text{ contains "Y年之中,″ and } y_i \text{ is before Y}, \\ 1 & \text{if } Q_T \text{ contains ″Y年以後,″ and } y_i \text{ is before Y}, \\ 0 & \text{otherwise} \end{cases} \tag{9}$$

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 about the time dimension. Take, for example, the above query which 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 this year, 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 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 [Swain91] is a basic method and has good performance for representing image content. The color histogram method gathers statistics about the proportion of each color as the signature of an image. In our work, the colors of an image are represented in the HSV (Hue/ Saturation/ Value) space, which is believed closer to human perception than other models, such as RGB (Red/ Green/ Blue) or CMY (Cyan/ Magenta/ Yellow). We quantize the HSV space into 18 hues, 2 saturations, and 4 values, with additional 4 levels of gray values; as a result, there are a total of 148 (i.e., 18×2×4+4) bins. Let C ($|C| = m$) a set of colors (i.e., 148 bins), P_I (Q_I) is represented as Eq. (10), which models the color histogram $H(P_I)$ ($H(Q_I)$) as a vector, in which each bucket h_{c_i} counts the ratio of pixels of P_I (Q_I) in color c_i.

$$\begin{aligned} P_I &= < h_{c_1}(P_I), ..., h_{c_m}(P_I) >, \\ Q_I &= < h_{c_1}(Q_I), ..., h_{c_m}(Q_I) > \end{aligned} \tag{10}$$

Previous work models that each pixel is only assigned into 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. (11),

$$h_{c_i}(P_I) = \frac{\sum_{p \in P_I} \frac{|\alpha_p - \beta_p|}{\delta}}{|P_I|} \tag{11}$$

where $[\alpha_p, \beta_p]$ is equivalent to $c_i \bigcap [p - \frac{\delta}{2}, p + \frac{\delta}{2}]$. Fig. 2 gives an example to explain what we called a partial pixel. In the figure, c_{i-1}, c_i, c_{i+1} stand for a color bin, a solid line indicate the boundary of c_i, p is the value of a pixel, $[p - \frac{\delta}{2}, p + \frac{\delta}{2}]$ denotes the interval range δ, the shadow part, $[\alpha_p, \beta_p]$, is the intersection of $[p - \frac{\delta}{2}, p + \frac{\delta}{2}]$ and c_i, each contribution of the pixel to c_i and c_{i-1} is computed as $\frac{|\alpha_p - \beta_p|}{\delta}$ and $\frac{|(p - \delta/2) - \alpha_p|}{\delta}$. It is clear that a pixel has its contributions not only to c_i but also to its neighboring bins.

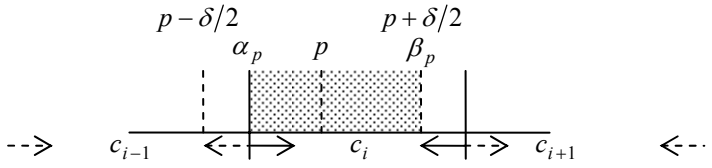


Fig. 2. The illustration of the idea of the partial pixel.

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 proposed 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

$$P_T = < w_{t_1}(P_T), ..., w_{t_n}(P_T), w_{c_1}(P_T), ..., w_{c_m}(P_T), w_{y_1}(P_T), ..., w_{y_k}(P_T) >,$$
$$Q_T = < w_{t_1}(Q_T), ..., w_{t_n}(Q_T), w_{c_1}(Q_T), ..., w_{c_m}(Q_T), w_{y_1}(Q_T), ..., w_{y_k}(Q_T) >, \tag{12}$$
$$Sim_1(P,Q) = \frac{\vec{P}_T \cdot \vec{Q}_T}{|\vec{P}_T||\vec{Q}_T|}$$

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

$$Sim_2(P,Q) = \alpha \cdot Sim_1(P,Q) + \beta \cdot Sim_3(P,Q) \tag{13}$$

where

$$P_I = < h_{c_1}(P_I), ..., h_{c_m}(P_I) >,$$
$$Q_I = < h_{c_1}(Q_I), ..., h_{c_m}(Q_I) >,$$
$$Sim_3(P,Q) = \frac{H(P_I) \bigcap H(Q_I)}{|H(Q_I)|} = \frac{\sum_i \min(h_{c_i}(P_I), h_{c_i}(Q_I))}{\sum_i h_{c_i}(Q_I)}$$

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

4 Interactive Search Mechanism

4.1 User Interface

Fig. 3 and Fig. 4 demonstrate the user interfaces which are 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. The system then returns a set of "relevant" images on the basic foundation of the similarity of the query and each image in the collection. In our design, the system first returns 80 images for the initial search, but 40 images in the later iterations. This is because that in the initial search the system does not catch an idea about what the user want exactly, a set of more images may induce the user to mark more relevant images, and it will give helps to the system when reformulating the query.

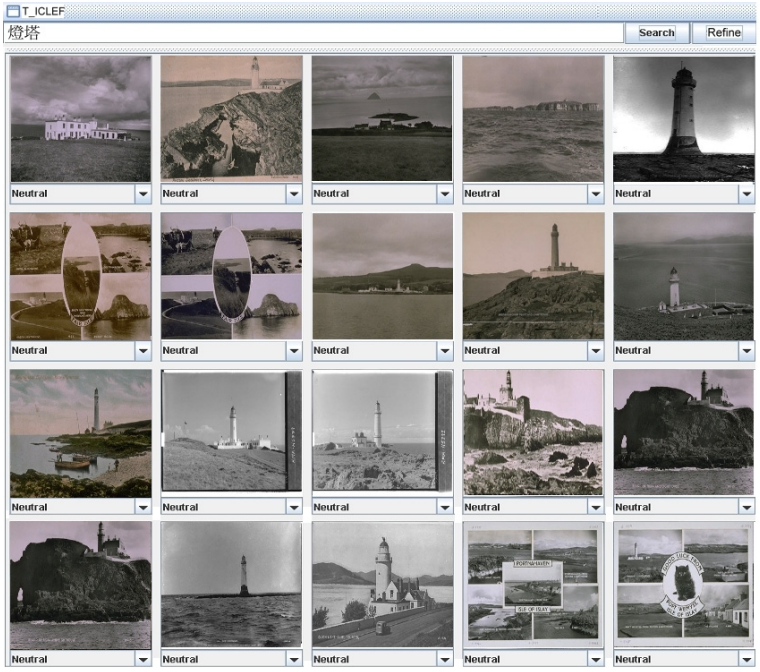


Fig. 3. The interface of T_ICLEF.

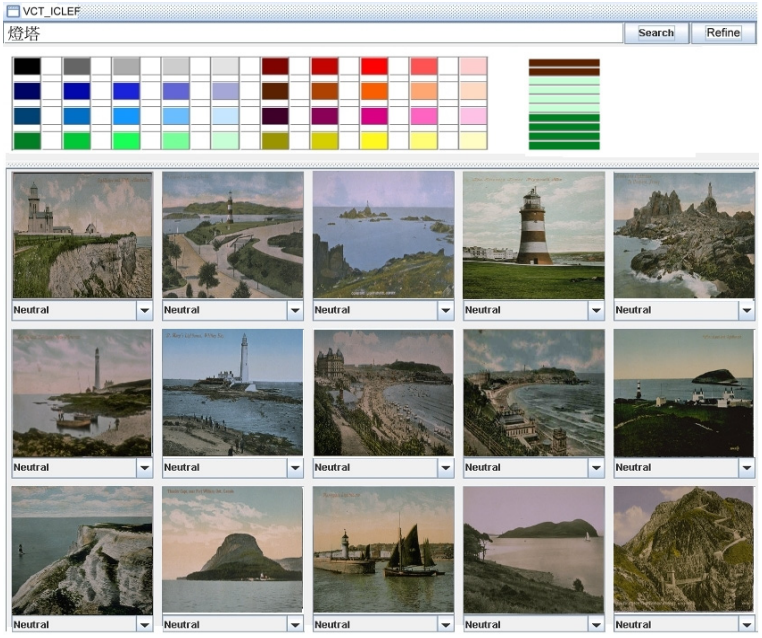


Fig. 4. The interface of VCT_ICLEF.

In the display area, a pull-down menu below each image assists users in feedback of the relevance of each image. In fact, it is the color table shown in VCT_ICLEF which distinguishes the two systems. Users can provide color information, which is considered as an image in the system, 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.

4.2 Query Reformulation

As mentioned in Section 2, 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 again Cross-Language image retrieval based on the new query.

Recall that we denote as the original query $Q = (Q_T, Q_I)$ and the new query $Q' = (Q'_T, Q'_I)$; as for Q'_T, we exploit a practical method, as shown in Eq. (14), for query reformulation. This mechanism, which has been suggested by [Rocchio65], is achieved with a weighted query by adding useful information which is extracted from relevant images as well as decreasing useless information which is 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 the average of the relevant images. We do not take into account the irrelevant images for Q'_I since in our observations, there is always a large difference among the non-relevant images in which we believe that adding the irrelevant information to Q'_I will make no contribution.

$$Q'_T = \alpha \cdot Q_T + \frac{\beta}{|REL|} \sum_{P_T \in REL} P_T - \frac{\gamma}{|NREL|} \sum_{P_T \in NREL} P_T \tag{14}$$

$$Q'_I = \frac{1}{|REL|} \sum_{P_I \in REL} P_I \tag{15}$$

In Eq. (14) and Eq. (15), α, β, $\gamma \geq 0$ are parameters, *REL* and *NREL* stands for the sets of relevant ad irrelevant images which are marked by the user.

5 Evaluation Results

In this section, we present our evaluation results for the user-centered search task at ImageCLEF 2004.

5.1 The St. Andrews Collection

At the ImageCLEF 2004, the St. Andrews Collection[5] is used for the evaluation purpose in which the majority of images (82%) are in black and white. It is indeed a subset of the St. Andrews University Library photographic collection from which 10% (i.e., 28,133) images have been used. All images have an accompanying textual description which is composed of 8 distinct fields[6], including 1) Record ID, 2) Title, 3) Location, 4) Description, 5) Date, 6) Photographer, 7) Categories, and 8) Notes. In total, the 28,133 captions consist of 44,085 terms and 1,348,474 word occurrences; the maximum caption length is 316 words, but on average 48 words in length. All captions are written in British English, and around 81% of captions contain text in all fields. In most cases, the caption is a grammatical sentence of about 15 words.

5.2 The User-Centered Search Task

The goal of the user-centered search task is to study whether the retrieval system is being used in the manner

[5] Please refer to http://ir.shef.ac.uk/imageclef2004/guide.pdf for a detail description.
[6] In this paper, almost all fields were used for indexing, except for 1 and 8. The words were stemmed and stop-words were removed as well.

intended by the system designers and how the interface helps users reformulate and refine their search requests, given that a user searches with a specific image in mind but without knowing key information thereby requiring them to describe the image instead. At the ImageCLEF 2004, the interactive task is using an experimental procedure similar to iCLEF 2003[7].

In brief, given two interactive Cross-Language image retrieval systems – T_ICLEF and VCT_ICLEF, and the 16 topics shown in Fig. 5, 8 users are asked to test each system with 8 topics. For a system/topic combination, a total of 4 searchers will test the system. Users are given a maximum of 5 mins only to find each image. Topics and systems will be presented to the user in combinations following a latin-square design to ensure user/topic and system/topic interactions are minimized. Moreover, in order to know the search strategies used by searchers, we also conducted an interview after the task.

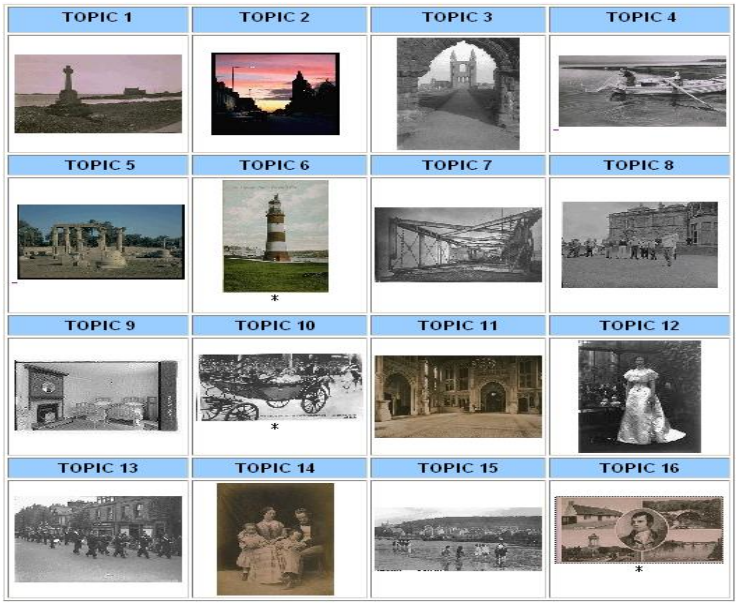


Fig. 5. The 16 topic images used in the user-centered search task.

5.3 Searcher Background

There are 8 people involved in the task, including 5 male and 3 female searchers. The average age of them is 23.5, with the youngest of 22 and the oldest of 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, in Web search. In average, they search about 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. In addition, all of them are native speaker of Chinese, but all learned English before. Most of them report that his/her English ability is acceptable or good.

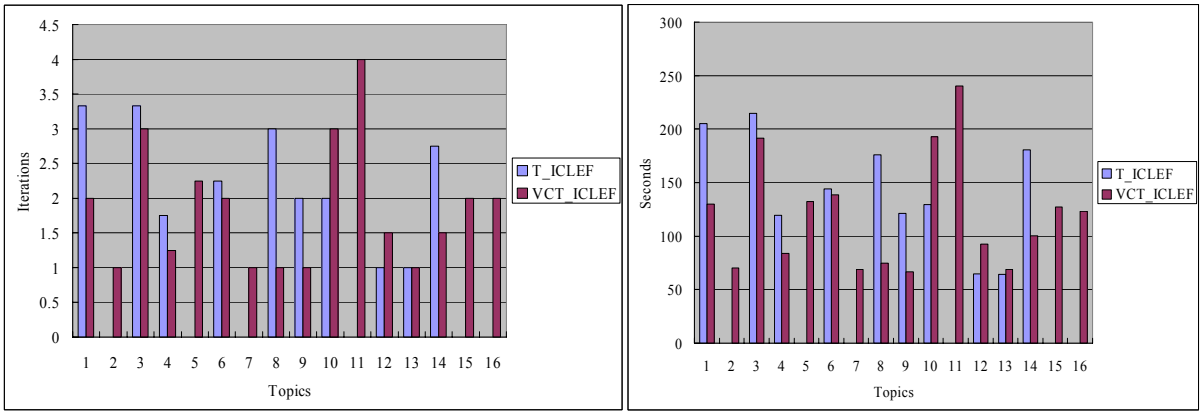
5.4 Results

We are interested in which system helps searchers find a topic image efficiently. We summarize the average number of iterations[8] and the average time spent by a searcher for each topic in Fig. 6. In the figure, it does not give information in the 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 top-

[7] Please refer to http://terral.lsi.uned.es/iCLEF/2003/guidelines.htm for further information.
[8] Please note that our system does not have a efficient performance; since for each iteration it spent about 1 minute to retrieve relevant images, approximately 5 iterations is performed within the time limit.

ics 2, 5, 7, 11, 15 and 16, no searcher can find the image by making use of T_ICLEF. Furthermore, with regard to topics 10 and 12, VCT_ICLEF has a worse performance. In our observations, it is because that most images (82%) in the corpus are in black and white, once the user gives imprecise color information, VCT_ICLEF needs to cost 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 search completed the task, T_ICLEF cost additional 0.4 iterations and 76.47 secs. Besides, by using VCT_ICLEF, on average, 89% of searchers successfully found the image while by using T_ICLEF, there is around 56.25% of searchers who did not fail the task.



(a) Average iterations (b) Average time

Fig. 6. The average number of iterations and the average time spent by a search for each topic.

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%

To show the effect of color information used in VCT_ICLEF, we take Fig. 3 and Fig. 4 for example. 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 by using VCT_ICLEF, the image was found in the first iteration. We conclude that color information can help a user indicate to 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.

Finally, we give an example as Fig. 7 to show that the proposed interactive mechanism works effectively. The query is "長鬍鬚的男人 (A man with a beard)". The evaluated system is T_ICLEF; after 2 iterations of relevance feedback, it is obviously that we can improve the result by our feedback method.

(a) Results after the initial search (b) Results after 2 feedback iterations

Fig. 7. Results for the query "長鬍鬚的男人 (A man with a beard)".

5.5 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 the system what they really wanted. Four searchers preferred to search the image with a text query first, even though by 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 in which temporal conditions were contained.

6 Conclusion

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) [Kohonen98] on image clustering, which we believe that it can provide an effective browsing interface to help searchers find a target image.

References

[Kohonen98] T. Kohonen, "The Self-Organizing Map," *Neurocomputing*, Vol. 21, No. 1-3, pp. 1-6, 1998.

[Kushki04] A. Kushki, P. Androutsos, K. N. Plataniotis, and A. N. Venetsanopoulos, "Query Feedback for Interactive Image Retrieval," *IEEE Transactions on Circuits and Systems for Video Technology*, Vol. 14, No. 5, 2004.

[Miller95] G. Miller, "WordNet: A Lexical Database for English," *Communications of the ACM*, pp. 39-45, 1995.

[Rocchio65] J. J. Rocchio, and G. Salton, "Information Search Optimization and Iterative Retrieval Techniques," *Proc. of AFIPS 1965 FJCC*, Vol. 27, Pt. 1, Spartan Books, New York, pp. 293-305, 1965.

[Salton83] G. Salton, and M. J. McGill, "*Introduction to Modern Information Retrieval*," McGraw-Hill, 1983.

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