

Multi-Modal Image Retrieval by Integrating Web Image Annotation, Concept Matching and Fuzzy Ranking Techniques

Ja-Hwung Su, Bo-Wen Wang, Tien-Yu Hsu, Chien-Li Chou, and Vincent S. Tseng

Abstract

Traditional image retrieval aims at bridging visual images and human concepts through visual or textual descriptions. However, it is still a challenging issue to reduce the gap between the images and user's intentions. To this end, a considerable number of studies in the field of multimedia mining have been conducted on how to effectively meet the user's requirements for image retrieval over the past few decades. However, there remain some problems unsettled. For content-based image retrieval, it is not easy to identify the user's interest by using visual descriptions only. For textual-based image retrieval, the modern search engines incur the problems of high manual tagging cost and low automated tagging precision. Moreover, precise image retrieval also leads unsatisfied results because of the rough human concepts. To catch user's concept well, we propose a novel approach, namely Intelligent SeMantic Image explorER (iSMIER), which considers the requirements of usability, intelligence and effectiveness simultaneously. Based on the proposed web image annotation, concept matching and fuzzy ranking methods, the users can obtain the desired images from the image collection easily and effectively. Through empirical evaluations, our annotation models can deliver high accuracy for serving semantic image retrieval.

Keywords: Image retrieval, image annotation, fuzzy set, cross media, information retrieval.

1. Introduction

Advanced image capturing devices recently enable a large increase in image archives. How to acquire the desired images for users from an image repository has been brought to researchers' attention over the past few years. For an image retrieval system, the primary goal is

to bridge visual images and human concepts through visual or textual descriptions. In general, classical image retrieval systems can be categorized into two types: *Content-Based Image Retrieval (CBIR)* and *Textual-Based Image Retrieval (TBIR)* systems. Behind *CBIR* [4][6][7][15] systems, the major notion is to represent the image by a set of visual features. Through content comparisons, the user can obtain the desired images from a set of image collections. In addition to *CBIR*, another solution for semantic image retrieval is *Textual-Based Image Retrieval (TBIR)*, such as [3], [5], Google, Yahoo, etc. Based on the captions of images, the image search engines can find the images most relevant to user's query terms. Figure 1 is an example for illustrating the paradigms of *CBIR* and *TBIR*.



Figure 1. Examples of *CBIR* and *TBIR*.

Although *CBIR* and *TBIR* have been proposed for a long time, they still suffer from some problems. Imagine that you are browsing a web page and taking a look at a preferred image. How could you obtain the similar images from a large image archive? The most popular way is to submit the linguistic terms or visual images to the *TBIR* or *CBIR* systems, respectively. Thus you can obtain a set of relevant images. Whatever the system is, there still exist some problems unsettled. For *TBIR*, first, the user has to stop browsing for image retrieval. Second, the user has to precisely define what she/he wants by the specific query terms. Third, the gap between the user's interest and images is too large to narrow due to the ambiguous query. To capture the user's intention, *TBIR* has to associate the high-level concepts with low-level visual features semantically. Although manual annotation is a useful solution instinctively, it needs prohibitive cost especially for a large-scale data set.

Corresponding Author: Vincent S. Tseng is with the Department of Computer Science and Information Engineering, National Cheng Kung University, No. 1, Ta-Hsueh Road, Tainan, Taiwan, ROC.

E-mail: tsengsm@mail.ncku.edu.tw

Manuscript received 14 Dec. 2009; revised 21 Jan. 2010; accepted 12 Mar. 2010.

Hence, some recent studies made attempts to approximate a near-optimal solution for image annotation. Unfortunately, it is still limited in the diverse relations between visual features and human concepts. For CBIR, the user has to download the image and submit it as a query to CBIR systems. So far, few studies are really successful in presenting the image by a set of visual features because of visual diversity problem. So-called visual diversity stands for that, a visual feature is shared by different kinds of images. As a result, it is difficult for image retrieval systems to identify the concept of an image precisely. Let us take Figure 1 as an example. Obviously, the visual diversity heavily exists in the resulting images because the concepts of Ship, Sky, Billiards, Mountain, SubSea and Surfing share the same colors.

To deal with such above problems, we propose a novel approach, namely *Intelligent SeMantic Image explorER* (*iSMIER*), to capture user's interested images using effective image annotation, concept matching and fuzzy ranking techniques. The benefits of *iSMIER* over TBIR or CBIR are three-folded:

1. **Usability:** It is not necessary for the user to stop browsing and define the query term any more. Through the right-mouse-click of the interested web image, the query procedure becomes very easy.
2. **Intelligence:** Without concerning the query terms, free your query to fetch your preferred images by the automated captions.
3. **Effectiveness:** Our proposed annotation approaches can achieve high quality of image annotation. That is, the correctly annotated keywords in *iSMIER* are very helpful to semantic image retrieval. Moreover, the proposed fuzzy ranking approach also enhances semantic image retrieval.

The experimental evaluations reveal that our proposed approaches can precisely capture user's intention and further support semantic image retrieval. The remaining of this paper is organized as follows. The previous work is described in Section 2. In Section 3, we present the proposed method in great detail. The related experimental evaluations are illustrated with Section 4. Finally, the conclusion is stated in Section 5.

2. Related Work

As can be concluded from above, the benefits of our proposed *iSMIER* mainly relies on image annotation techniques including visual- and textual-based annotation models. Without excellent annotations, we cannot obtain good results by *iSMIER*. As a result, we describe the details of the past literatures for image annotation in the followings.

Visual-based annotation: Some previous work made

attempts to find what visual features are shared by similar images in the image repositories. Chang *et al.* [4] provided an ensemble of binary classifiers, namely CBSA (Content-Based Soft Annotation), to predict an associated label membership for each image. By assuming that regions in an image can be described using a small vocabulary of blobs, Jeon *et al.* [10] proposed probabilistic models to predict the probability of generating a word given the blobs in an image. These segmentation-based approaches focus on the analysis of either the whole image or divided regions without considering the concept-objects in the image. Besides, On the basis of image segmentation, Wang *et al.* [25] proposed a method named SIMPLicity that classifies the images into predefined categories to assist the image retrieval. For SIMPLicity, one critical issue is that the restricted range of user-defined categories and expensive artificial cost will strongly limit the effort of the image annotation. Based on the categorical and numerical attributes with respect to captions and contents of regions, Pan *et al.* [14] developed MMG model to annotate the image by CCD (Cross-modal Correlation Discovery) algorithm. CRM (Continuous Relevance Model) was developed by Lavrenko *et al.* [13] to annotate a video on the basis of statistics. They segmented each sequential key-frame into several rectangle regions and then extracted the related visual features from these segmented regions. The annotation of each image is yielded soon after calculating the related probabilities with Gaussian Mixture Function. For visual-based annotation models, the visual annotations are limited in the predefined concepts.

Textual-based annotation: Another solution for enhancing the accuracy of image annotation is to make use of the textual information in the compound web page, such as file name, URL, title, surrounding text and frequent keywords [17][18][26]. WebSEEK [16] is a search engine that extracts the URL and some tags to detect the specific classification of the images. Without considering visual features, the context of a web page may narrow the gap between the visual images and user's concepts. However, there still exist many unsolved problems. First, frequent inappropriate textual information leads to the distorted annotations. Second, the appropriate similarity function between concepts and images is difficult to derive. At last, too many redundant candidate keywords increase the computation cost during the annotation procedure. Cheng *et al.* [3] supplied hierarchical clustering approach to support web image categorization and web image annotation by word N-grams for extracting feature-terms. The *tf-idf* is used as a weighting scheme to judge the similarity between property set of an image and vocabulary set of a view. However, the result depending on *tf-idf* and N-grams

cannot represent the real similarity since the statistics of keyword-term occurrence or co-occurrence only shows a part of the relationship between concepts and images. High complexity also limits its practicality.

In addition to above, another type of approaches annotates the feedback images by using query terms [5][23]. However, this kind of methods depends heavily on the quantity and quality of user's feedback. Moreover, its idea is instinctive and ignores the accumulated information of user's interaction. Wang *et al.* [22] annotated an image by both of visual and textual search. The results are still limited in the predefined categories and the performance is dependent upon the quality of visual and textual search. Wong *et al.* [24] made use of additional metadata (called Exchangeable Image File Format, EXIF), such as aperture, exposure time, subject distance, focal length, and fire activation, to tag the images. Nevertheless it cannot be applied to the images that do not contain such additional metadata. The web image annotation approach proposed by Feng *et al.* [8] involves two SVM classifiers that classify images by visual features and textual information. Tseng *et al.* proposed a hybrid approach to enhance image annotation [20][21]. Tian *et al.* [19] exploited multi-context analysis to accomplish the image classification by reducing the feature dimensionalities. According to the notion of classification, the discovered semantics still cannot escape from the user-predefined range.

After reviewing the related work for image annotation, another important issue is how to effectively retrieve the images by visual contents. Traditional CBIR paradigm intends to approximate the optimal similarity functions and useful visual features to achieve the precise image retrieval [7][15]. Unfortunately, the precisely processing is very difficult to deliver the user's interest. The similar experienced phenomenon also exists in most of the other AI research fields. Hence, fuzzy set theory has been adopted by more and more recent intelligent systems due to its simplicity and similarity to human reasoning [1][2][9][11]. FIRST (Fuzzy Image Retrieval SysTem) proposed by Krishnapuram *et al.* [12] uses Fuzzy Attributed Relational Graphs (FARGs) to represent images where each node in the graph represents an image region and each edge represents a relation between two regions. Every query is converted to a FARG to compare with the FARGs in the database. Chen *et al.* [6] proposed UFM (unified feature matching) to retrieve the images. In this study, an image is represented several segmented regions based on a fuzzy feature. Nevertheless, the above two fuzzy approaches are limited in region segmentation due to region segmentation is not very robust. Therefore, in this paper, we propose a new fuzzy matching (also called fuzzy ranking) technique to touch the user's mind without

region segmentation. The above fuzzy image retrieval still cannot exactly deliver the diverse human-concepts hidden in visual-contents. To effectively attack the lack of above, this paper presents a novel image retrieval browser, namely *Intelligent SeMantic Image explorER (iSMIER)*, which integrates image annotation and fuzzy set to reach the high quality of image retrieval.

3. Proposed Method

3.1 Overview of Intelligent SeMantic Image explorER (iSMIER)

As depicted in Figure 2, the goal of *iSMIER* is to provide the users with easy-and-free query environment to obtain the desired images. Basically, successful retrieval for *iSMIER* relies on two phases, namely off-line training and on-line query phases, which include annotation-based learning, conceptual information processing, content-based preprocessing, category mapping and visual fuzzy ranking. Once the correct captions for the browsing web image are derived by the proposed annotation models, textual-based query is automatically triggered to find the matching concepts and to rank the relevant images. The whole procedure is briefly described in the followings.

I. **Training Phase:** The major concern of this phase is to construct annotation, concept and content prediction modules, with respect to Annotation-based Learning, Conceptual Information Preprocessing and Content-based Preprocessing, to support query phase. For annotation-based learning module, a number of web pages containing annotated images are collected as the learning set. Upon the learning set, the proposed annotation models can be generated to facilitate the web image tagging. For conceptual information processing module, a set of keywords referred to the image categories are collected as the feature-keywords. Then the term frequency (defined as *tf* in this dissertation) and the inverse document frequency (defined as *idf* in this dissertation) for each feature-keyword are calculated to construct the proposed conceptual information model. The conceptual information model is helpful to bridging the target categories and the user's query term. For content-based preprocessing module, all visual images in the database are transformed into fuzzy sets to accelerate the visual fuzzy ranking. Once three modules are ready, query phase can work smoothly.

II. **Query Phase:** This is another important phase to present the novelty of this work. Based on the annotations derived, how to simplify the query procedure to hunt user's interested images effectively and efficiently is our next concern. The query starts with that a user is browsing a web page and interested in an embedded image. She/he submits the interested image as

a query to the annotation model by clicking the interested image. Next, the system predicts the relevant captions and associates the predicted captions with the image categories by category mapping module. Finally the images for each resulting category are sorted by visual fuzzy ranking module.

The details of how to perform the semantic image retrieval are described in the succeeding subsections.

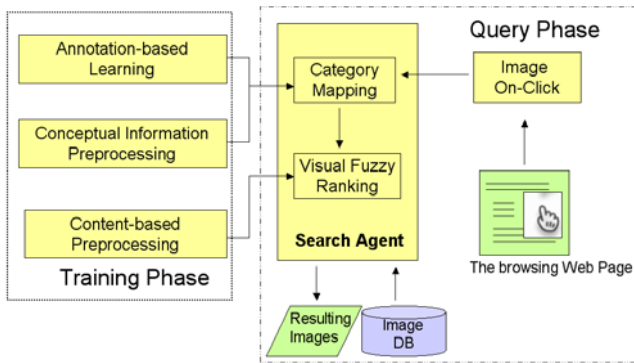


Figure 2. The framework of *iSMIER*.

3.2 Training Phase

3.2.1 Annotation-based Learning

Indeed, annotation-based learning is very important for the proposed semantic image retrieval. The goal of this module is to generate the prediction model for web image annotation. If the web image can be tagged correctly, the automated semantic query can work smoothly. On the other hand, without correct annotations, the user cannot find the desired images by the automated semantic query. As shown in Figure 3, this module can be divided into three models, namely $Model_{MACK}$ (*Mining Affinities of Captions and Keywords*), $Model_{ISD}$ (*Image Sense Disambiguation*) and $Model_{FIM}$ (*Fusion of ISD and MACK*).

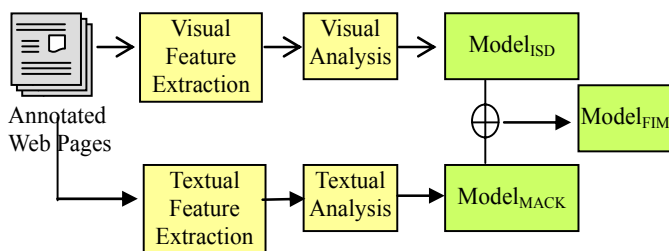


Figure 3. Workflow of annotation-based Learning.

A. Construction of $Model_{MACK}$ (*Mining Affinities of Captions and Keywords*)

In principle, what $Model_{MACK}$ concerns is to discover the affinities of image captions and webpage keywords. A web page always contains several images and some available textual information. The textual information includes articles, anchors, HTML tags, link information,

image names, etc. Thus there exist some affinities between image captions and textual information. From human viewpoint, each keyword existing in textual information would be a potential caption to an image. Nevertheless, not every keyword is suitable to describe an image in real applications. That is, the keywords existing in different textual information would be assigned the positiveness and negativeness. The positiveness and negativeness denote the degrees of being a correct caption and a noise caption, respectively. It motivates us to discover the positive degree and negative degree for each keyword from the web pages. In this model, the selected attributes are shown in Table 1. The construction of this model is expressed as the following three steps.

Table 1. The definitions of the selected attributes for $Model_{MACK}$.

Parameter	Attribute	Domain
N	File Name	{boolean}
T	Page Title	{boolean}
A	ALT Tag	{boolean}
S	Surrounding Text	{boolean}
U	URL	{boolean}

Step 1: Process the training web pages

Before building the textual-based annotation model, we have to process the web pages one by one. First, the potential keywords are extracted from each annotated web page (also called training pages) through removing the stop words. Next, each potential keyword has to be stemmed. Finally, if the keywords are the pre-annotated captions, they are collected as positive keywords. Otherwise, the others are collected as negative keywords. The processing example is shown in Figure 4.

Step 2: Generate positive and negative tables

To clarify the positive degree and negative degree of each potential keyword, we have to infer the positive weight and negative weight of the referred attribute for each potential keyword. Hence, in this step, the primary task is to collect the textual information of the pre-annotated captions. The positive and negative keywords are defined as follows.

Definition 1. Suppose that a number of keywords are extracted from the m^{th} training web page after stop-word removal and word stemming. For the extracted keywords, if the n^{th} keyword is contained in the training ground-truth for the m^{th} training web page, it is defined as positive keyword K_{mn} . Otherwise, it is defined as negative keyword NK_{mn} .

For example, as shown in Tables 2 and 3, the training

textual information can be divided into two tables, namely *positive table* and *negative table*. These two tables can be viewed as the textual-based annotation model. In these tables, K_{mn} indicates the n^{th} positive keyword in the m^{th} page and NK_{mn} indicates the n^{th} negative keyword in the m^{th} page. For each attribute defined in Table 1, the value is 1 if the keyword is extracted from the referred attribute.

Table 2. Example of the positive table.

id	keyword	N	T	S	A	U
1	K_{11}	1	0	1	1	0
2	K_{12}	1	1	0	0	0
3	K_{21}	0	1	1	1	0
4	K_{22}	0	0	1	1	1
5	K_{31}	0	0	1	1	1
6	K_{41}	1	1	1	1	0
7	K_{51}	1	0	0	1	1
8	K_{61}	0	0	0	0	1
9	K_{62}	1	1	0	1	1
10	K_{63}	1	1	1	1	1

Table 3. Example of the negative table.

id	keyword	N	T	S	A	U
1	NK_{11}	0	0	1	1	0
2	NK_{12}	1	1	0	0	1
3	NK_{21}	0	1	1	0	0
4	NK_{22}	0	0	1	1	1
5	NK_{31}	0	0	1	1	1
6	NK_{41}	1	1	1	1	0
7	NK_{51}	1	0	0	1	1
8	NK_{61}	0	1	0	0	1
9	NK_{62}	1	1	0	1	1
10	NK_{63}	0	0	0	0	1

Step 3: Generate positive and negative patterns

From the positive and negative tables, we can discover a set of positive patterns psf and a set of negative patterns nsf . The sets of positive i -length- and negative j -length-patterns can be defined as psf^i and nsf^j , respectively. For example, from Table 2, $psf^1 = \{\{N\}, \{T\}, \{S\}, \{A\}, \{U\}\}$. For T, the positive supports of $\{T\}$, $\{T, U\}$, $\{T, S, A\}$, $\{T, S, A, U\}$ and $\{F, T, S, A, U\}$ are 5/10, 2/10, 3/10, 1/10 and 1/10, respectively. The other supports of positive patterns and negative patterns can be deduced as the same as this example

B. Construction of $Model_{ISD}$ (Image Sense Disambiguation)

The primary goal of $Model_{ISD}$ is to disambiguate the image sense (also called caption in this work) while a set of senses is shared with a number of images. As shown in Figure 5, the annotated images in this construction have to be grouped into several clusters first by calculating the visual similarities. Next, two related measures, called $Ssense_Frequency$ (sf) and $Entropy$, have to be calculated. They are defined as the following.

Definition 2. Consider that a training data set D is divided into t clusters, $D = \{C_1, C_2, \dots, C_t\}$, and there is a set of senses. Assume that a cluster contains several images and each image is assigned several senses. Hence a cluster can be viewed as a collection of senses, $C_j = \{se_1, se_2, \dots, se_g\}$. The entropy of the j^{th} cluster can be defined as:

$$Entropy^j = \sum_{1 \leq i \leq g} \log\left(\frac{|C_j|}{sf_i^j}\right) \quad (1)$$

where sf_i^j stands for the frequency of the i^{th} sense in the j^{th} cluster.

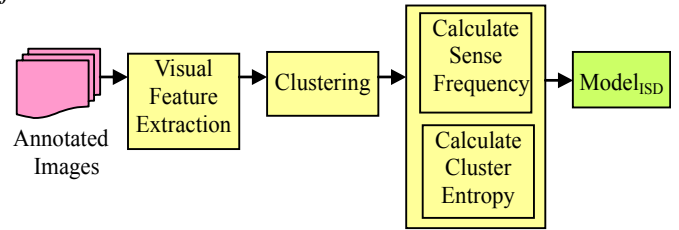


Figure 5. Construction of $Model_{ISD}$.

Let us take Figure 6 as an example to show the $Model_{ISD}$. Assume that 5 images containing 3 unique senses $\{\text{bear, grass, flower}\}$ are grouped into the j^{th} cluster. The frequency set of $\{\text{bear, grass, flower}\}$ is $\{3, 3, 1\}$ and $Entropy^j$ is $\log(7/3) + \log(7/3) + \log(7/1) = 1.583$.

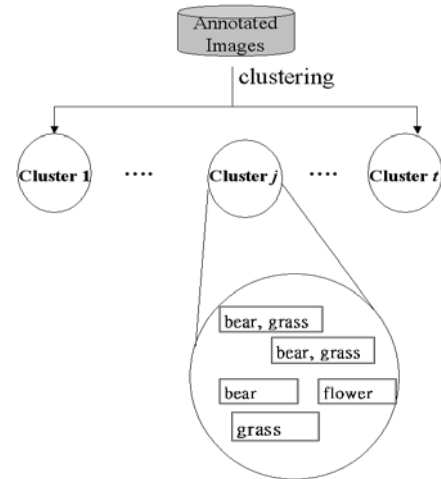


Figure 6. Example of $Model_{ISD}$.

3.2.2 Conceptual Information Preprocessing

As mentioned above, the aim of this module is to construct a model, called $Model_{CMM}$ (Concept Matching Model), for associating the image categories and the predicted web image captions. To this end, each image category has to be featured by a set of relevant keywords that are gathered from a large number of related web pages. As shown in Figure 7, the procedure is divided

into four steps.

Step 1. Collect the relevant web pages.

Behind this model, the major notion is to reflect the linguistic features of target image categories. To carry out this notion, we have to collect the relevant keywords as the linguistic features (metadata) from a corpus. Unfortunately, there is no general corpus to cover all words. It motivates us to utilize the web as a near-complete corpus. In this step, the web is regarded as a large-scale corpus. For each target image category, the category term is submitted to the search engine, such as Google, and top 10 matching web pages are returned as the most-relevant web pages.

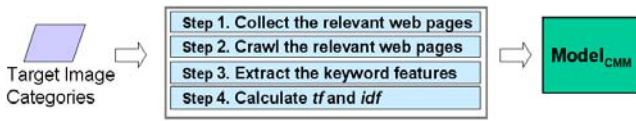


Figure 7. Workflow of constructing Model_{CMM}.

Step 2. Crawl the relevant web pages.

For each target image category, we analyze the related contents page by page. Because the matching web pages are ranked by Google, the words in the contents is highly related to the submitted image category-term. From the embedded words, the stop words are removed and the remaining words are stemmed.

Step 3. Extract the feature-keywords.

After crawling the relevant web pages, the remaining keywords (also called term) for each category are collected as the feature-keywords in this step.

Step 4. Calculate term frequency (*tf*) and inverse document frequency (*idf*).

Although the keywords are identified in the previous steps, not every keyword is related to the category. The basic information to weight the keywords includes the term frequency and the inverse document frequency. For term frequency, if the term frequency is high, the term is good enough to present the category. By contrast, the inverse document frequency represents the related distinctness of a term. If the inverse document frequency of the term is low, the importance of the term is high. Finally, for each category, top 2000 feature-keywords in this work are retained to serve the category mapping in the query phase.

Definition 3. Assume that there are y categories in the database $CA = \{ca^1, ca^2, \dots, ca^y\}$ and the i^{th} category ca^i contains a set of feature-keywords $\{fk_1^i, fk_2^i, \dots, fk_N^i\}$. The term frequency for the keyword fk_j^i is defined as

$$tf_j^i = \frac{\text{count}(fk_j^i)}{\sum_{1 \leq x \leq N} \text{count}(fk_x^i)}, \quad (2)$$

and the inverse document frequency for the keyword fk_j^i is

$$idf_j^i = \log\left(\frac{|CA|}{|\sum_{1 \leq m \leq y} df_j^m|}\right), \quad (3)$$

where

$$df_j^m = \begin{cases} 1, & \text{if } fk_j^i \in ca^m \\ 0, & \text{otherwise} \end{cases}$$

Thus the feature value of fk_j^i is defined as

$$T_fu_j^i = tf_j^i * idf_j^i. \quad (4)$$

3.2.3 Content-based Preprocessing

In addition to above modules, this module is produced to achieve efficient fuzzy image retrieval. For efficiency, the high dimensional visual features are reduced into fuzzy sets to accelerate the image retrieval. In this module, we integrate similarity function and membership function to transform the visual features into fuzzy sets, called Model_{Fset}. As shown in Figure 8, the primary task of Model_{Fset} can be decomposed into the following subtasks.

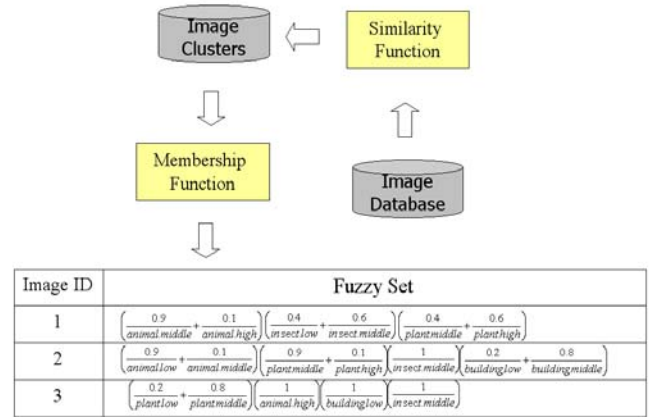


Figure 8. Workflow of constructing Model_{Fset}.

I. Group the images into several clusters.

In this step, visual features with respect to *Color Layout* and *Homogeneous Texture* are extracted to serve the Euclidean Distance computations. At last, the images for each category are grouped into equal group number by K-means. In this paper, the cluster quantity for each category is 8.

II. Transform the visual features into fuzzy sets.

Due to the high dimensionality of visual features, traditional visual comparison is time consuming. To overcome such obstacles, our intent is to reduce the visual dimensionality by transforming visual features into fuzzy sets. Figure 9 depicts the detailed procedure of how to transform visual features into fuzzy sets, called Algorithm Trans_Fset. The whole process can be

elaborated on the following two steps.

Step 1. Similarity calculation: For each image in the database, we find the top s relevant clusters by visual comparisons. The major idea behind this step shown in Figure 10 is to identify the categories relevant to the images. It can prevent the errors caused by the precise identification.

Input: The images in database D , a set of categories with the related clusters and a set of membership functions

Output: Transaction Table T containing images with fuzzy sets

Algorithm Trans_Fset

1. Define cardinality k ;
2. **for** each image $I_j \in D$ **do**
3. Calculate distances and discover the top k closer clusters;
4. Calculate the count cnt_{ca_i} ($0 \leq cnt_{ca_i} \leq k$) of each category ca^i from the most-relevant k clusters;
5. **for** each category with $cnt_{ca_i} \neq 0$ **do**
6. Convert cnt_{ca_i} of ca^i into a fuzzy set $f_{ca^i}^j$ denoted as $(\frac{M_1^{ca^i}}{R_1^{ca^i}} + \frac{M_2^{ca^i}}{R_2^{ca^i}} + \dots + \frac{M_n^{ca^i}}{R_n^{ca^i}})$ by employing the given membership functions, where $R_n^{ca^i}$ is the n^{th} fuzzy region of ca_i and $M_n^{ca^i}$ is the fuzzy membership value in region $R_n^{ca^i}$;
7. $F^j = \cup f_{ca^i}^j$;
8. **end for**
9. $T = \cup F^j$;
10. **end for**
11. **return** T ;

Figure 9. Algorithm Trans_Fset.

Step 2. Fuzzy set calculation: After similarity calculations, the top s relevant clusters are determined and the cardinality of clusters for each category is further counted. The fuzzy sets of each image in database are implied by our proposed membership functions based on [9]. From Line 4 to 9 in Figure 9, the transaction table T will be generated by these fuzzilized images. For example, assume that s is 20. Table 4 is an example showing that, a dataset contains three images. Each image contains several categories and the referred cardinality, respectively. Next, the quantitative cardinality of categories will be transformed into fuzzy sets. In Table 4, the cardinality set of clusters to {animal, insect, building, plant} is {7, 3, 5, 5} for the 1st image. They are defined as {(animal, 7), (insect, 3), (building, 5), (plant, 5)}. Let us take building as an example, it is further converted into the fuzzy set $(\frac{0.0}{building\ low} + \frac{0.5}{building\ middle} + \frac{0.5}{building\ high})$ by using the given membership functions, as shown in Figure 11. Table 5 is

an example showing the transformed fuzzy sets, where *category.level* is called a fuzzy region, e.g., *building.low*.

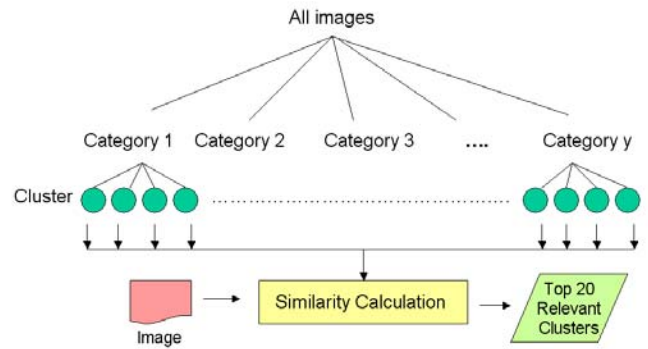


Figure 10. Similarity calculation.

Table 4. Example of the conceptualized table.

Image ID	Category
1	(animal, 7), (insect, 3), (building, 5), (plant, 5)
2	(animal, 2), (plant, 7), (insect, 6), (building, 5)
3	(animal, 5), (insect, 4), (plant, 2), (city, 6), (figure, 3)

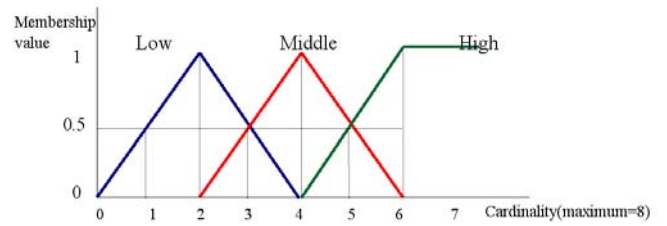


Figure 11. Fuzzy membership functions for cardinality attribute.

Table 5. Example of the fuzzilized table.

Image ID	Fuzzy Set
1	$(\frac{1}{animal\ high}) (\frac{0.5}{insect\ low} + \frac{0.5}{insect\ middle})$ $(\frac{0.5}{building\ middle} + \frac{0.5}{building\ high}) (\frac{0.5}{plant\ middle} + \frac{0.5}{plant\ high})$
2	$(\frac{1}{animal\ low}) (\frac{1}{plant\ high}) (\frac{1}{insect\ high})$ $(\frac{0.5}{building\ middle} + \frac{0.5}{building\ high})$
3	$(\frac{0.5}{animal\ middle} + \frac{0.5}{animal\ high}) (\frac{1}{insect\ middle})$ $(\frac{1}{plant\ low}) (\frac{1}{city\ high}) (\frac{0.5}{figure\ low} + \frac{0.5}{figure\ middle})$

3.3 Query Phase

To make the query easy and free, depicted in Figure 12, the users only need to click the interested web image to submit the textual and visual information to the search agent. After category mapping, the most-relevant categories can be derived. Eventually the images are ranked by visual fuzzy ranking. For search agent, two

major modules namely *category mapping* and *visual fuzzy ranking* are described as follows in detail.

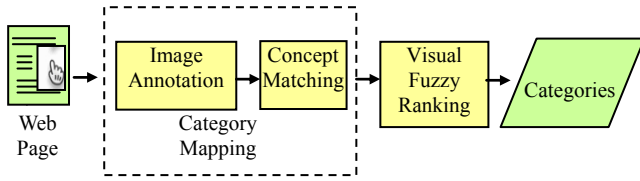


Figure 12. Workflow of query phase.

3.3.1 Category Mapping

In fact, it is not easy to achieve the high quality of semantic image retrieval because the gap between the user’s intention and visual features is large. To aim at this issue, we propose the module for category mapping to meet the user’s semantic demand well. In general, this module is triggered by a web image clicked by the user. Two major tasks of this module, called *image annotation* and *concept matching*, are performed to bridge the user’s interest and image categories. That is, after the image annotation work, a set of captions is submitted to serve the concept matching work. At last, the potential categories for the user’s interest can be inferred.

A. Image Annotation

A.1 Image Annotation by Model_{MACK}

How to determine the captions from a set of keywords extracted from a testing/browsing web page is the main concentration in this model. To make the determination more effective, each potential keyword has to be assigned two degrees, namely textual positiveness and textual negativeness, by the following definition.

Definition 4. Consider that a keyword is extracted from q attributes in a testing/browsing web page. Hence, q is the length of the maximum pattern. For this keyword, the referred textual positiveness is defined as:

$$T_positiveness = \sum_{i=1}^q \min(sup(psf^i)), \quad (5)$$

where $sup(psf^i)$ stands for the supports of the positive i -length-patterns, which is calculated by the positive table. And the textual negativeness is defined as:

$$T_negativeness = \sum_{j=1}^q \max(sup(nsf^j)), \quad (6)$$

where $sup(nsf^j)$ stands for the supports of the negative j -length-patterns, which is calculated by the negative table. The textual degree of the keyword is $TDegree=(T_positiveness-T_negativeness)$.

Table 6 is an instance for the keywords extracted from a browsing web image. In this case, assume that a testing web page contains 5 keywords $\{CK_1, CK_2, CK_3, CK_4, CK_5\}$ and the user’s interested web image. Suppose that CK_1 is extracted from File Name, Surrounding Text and

ALT tag. The referred attribute set is $\{N, S, A\}$ and $q=3$. According to Tables 2 and 3, the related $T_positiveness$ is $[\min(sup(N), sup(S), sup(A))] + [\min(sup(N, S), sup(S, A), sup(N, A))] + [\min(sup(N, S, A))]$ $= [\min(0.6, 0.6, 0.8) + \min(0.3, 0.6, 0.5) + (0.3)] = 1.2$. Similarly, $T_negativeness$ is $[\max(sup(N), sup(S), sup(A))] + [\max(sup(N, S), sup(S, A), sup(N, A))] + [sup(N, S, A)] = [\max(0.4, 0.5, 0.6) + \max(0.1, 0.4, 0.3) + (0.1)] = 1.1$. Finally, the $TDegree=1.2-1.1=0.1$. Suppose that the attribute set referred to another keyword CK_2 is $\{N, T\}$ and $q=2$. According to Definition 4, $TDegree=[\min(0.6, 0.5) + (0.4)] - [\max(0.4, 0.5) + (0.3)] = 0.1$. The resulting table for each keyword degree is shown in Table 7. If the threshold of selecting the predicted captions, defined as α , is set to be 0, the resulting caption set is $\{CK_1, CK_2, CK_3\}$.

Table 6. Example of attribute keywords for the browsing web image.

Keyword#	F	N	T	S	A	U
CK ₁	3	1	0	1	1	0
CK ₂	6	1	1	0	0	0
CK ₃	2	0	1	1	1	0
CK ₄	6	0	0	1	1	1
CK ₅	4	0	0	1	1	1

Table 7. Example of resulting degrees for extracted keywords.

Keyword#	$T_positiveness$	$T_negativeness$	$TDegree$
CK ₁	1.2	1.1	0.1
CK ₂	0.9	0.8	0.1
CK ₃	1.1	1.1	0
CK ₄	1.2	1.3	-0.1
CK ₅	1.2	1.3	-0.1

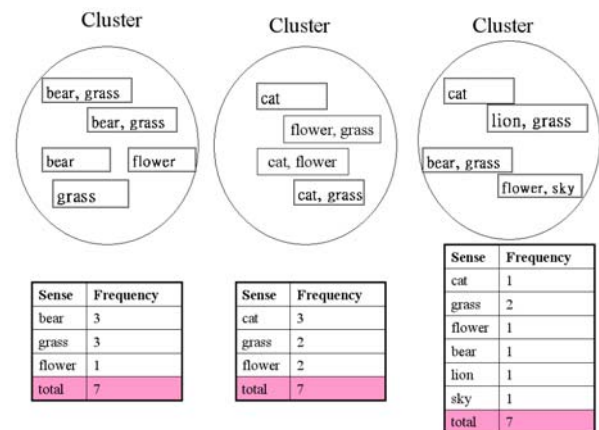


Figure 13. Example of three relevant clusters to the browsing web image.

A.2 Image Annotation by Model_{ISD}

Based on the entropy and frequency of each sense, all we want to do is to prune the ambiguous senses of an

image. The detailed procedure starts with that an image QI is submitted to this model. First, a set of the most-relevant cluster set CS to QI is determined by visual distance calculations. Assume that the referred distance to the j^{th} nearest cluster $C_j \in CS$ is defined as dc^j . Therefore the visual degree of the i^{th} sense can be defined as:

$$VDegree = \sum_{1 \leq j \leq |CS|} \left[\left(\frac{sf_i^j}{\sum_{1 \leq v \leq |C_j|} sf_v^j} \right) \times Entropy^j \times \left(\frac{\sum_{1 \leq k \leq |D|} dc^k}{dc^j} \right) \right], \quad (7)$$

where $\left(\frac{\sum_{1 \leq k \leq |D|} dc^k}{dc^j} \right)$ indicates the normalized distance for the j^{th} cluster and $\left(\frac{sf_i^j}{\sum_{1 \leq v \leq |C_j|} sf_v^j} \right)$ indicates the normalized frequency for the i^{th} sense in the j^{th} cluster.

For example, assume that there are $|D|$ clusters in Model_{ISD}, $\left(\sum_{1 \leq k \leq |D|} dc^k \right) = 30114$ and $|CS|=3$. Figure 13 shows that three clusters contain 13 images and 6 unique senses. If the referred distance set for $\{C_1, C_2, C_3\}$ is $\{713, 770, 1020\}$, the normalized distance set is $\{42.24, 39.11, 29.52\}$. Additionally, the entropy set for $\{C_1, C_2, C_3\}$ is $\{1.58, 1.46, 4.68\}$ and the frequency set for each sense in the related clusters is also illustrated with Figure 13. As a result, the visual degree for each sense can be derived as shown in Table 8

Table 8. Example of the resulting visual degrees.

Sense	VDegree
cat	$[(0/7)*(1.58)*(42.24)]+[(3/7)*(1.46)*(39.11)]+[(1/7)*(4.68)*(29.52)]=44.21$
grass	$[(3/7)*(1.58)*(42.24)]+[(2/7)*(1.46)*(39.11)]+[(2/7)*(4.68)*(29.52)]=84.38$
flower	$[(1/7)*(1.58)*(42.24)]+[(2/7)*(1.46)*(39.11)]+[(1/7)*(4.68)*(29.52)]=45.58$
bear	$[(3/7)*(1.58)*(42.24)]+[(0/7)*(1.46)*(39.11)]+[(1/7)*(4.68)*(29.52)]=48.34$
lion	$[(0/7)*(1.58)*(42.24)]+[(0/7)*(1.46)*(39.11)]+[(1/7)*(4.68)*(29.52)]=19.74$
sky	$[(0/7)*(1.58)*(42.24)]+[(0/7)*(1.46)*(39.11)]+[(1/7)*(4.68)*(29.52)]=19.74$

A.3 Image Annotation by fusion model Model_{FIM}

This prediction, namely FIM (*Fusion of ISD and MACK*), fuses above two models to enhance the prediction quality. The mixed degree φ of each candidate caption is:

$$\varphi = w(\text{normalized}(TDegree)) + (1-w)(\text{normalized}(VDegree)), \quad (8)$$

where w denotes the weight of the annotation models. According to φ , the better results can be derived to serve semantic image retrieval. Note that, if the degrees of two keywords are equal, we adopt the related occurrences on the web page to determine the referred ranks.

B. Concept Matching

After implying the captions of the browsing web

image, the next step is to connect the predicted captions and the image categories. Indeed, it is a challenging issue for computing the similarity of two linguistic terms. Without effective concept matching, the user's intention cannot be delivered by the predicted captions. Note that, a predicted caption can also be viewed as a user's query term Q in the followings. Figure 14 shows the procedure of finding the relevant categories to the user's query term. First, the query term is submitted to search engine like Google, and a set of extracted feature-keywords are gained, as shown in lines 1-3 of Figure 14. Next, the matrix L_{CA-FK} is initialized. Lines 7-15 show the similarity calculation. For each category, the referred degree is accumulated by the feature values of matching feature-keywords. That is, if finding the same feature-keyword in the feature sets of both query-term Q and target-category ca^i , the feature value rv_i in the matrix L_{CA-FK} is accumulated by $(tf_j^Q * T_fu_j^i)$. The similarity calculation ends with that all feature-keywords of Q are employed to find the matching feature-keyword of each category. At last, each target category is assigned the related degree. For each query term derived by annotation models, the degree related to each category is accumulated into the ranking matrix. According to the ranking matrix, top 5 relevant categories are selected as the results.

Input: A query term Q and a set of image categories $CA=\{ca^1, ca^2, \dots, ca^y\}$ and ca^i contains a set of feature-keywords

Output: The ranking matrix L

Algorithm Concept_Match

1. submit Q as a term to the search engine and collect the relevant web pages $\{r_1, r_2, \dots, r_{10}\}$;
2. crawl the web pages;
3. extract the keyword features $FK=\{fk_1^Q, fk_2^Q, \dots, fk_z^Q\}$;
4. let a matrix $L_{CA-FK}=[rv_y]$, where rv denotes the accumulated feature value for each category;
5. **for** $j=1$ to y **do**
6. initialize $rv_j=0$;
7. **for** $j=1$ to z **do**
8. **for** $i=1$ to y **do**
9. **for** each keyword $fk \in ca^i$ **do**
10. **if** $fk = fk_j^Q$ **then**
11. $rv_i = rv_i + (tf_j^Q * T_fu_j^i)$;
12. **break**;
13. **end if**
14. **end for**
15. **end for**
16. **return** the ranking matrix L ;

Figure 14. Algorithm Concept_Match.

3.3.2 Visual Fuzzy Ranking

Although the user's preferred image categories can be

inferred successfully by the above procedure, another problem to face is the visual similarities between the browsing web image and images in the database. From visual viewpoint, it is because a category perhaps contains diverse contents that the visual ranking is actually necessary. To deal with the scalability and uncertainty problems for image retrieval, we present a novel fuzzy ranking method in this work, as stated in the followings.

In this stage, the membership value set of the browsing image has to be calculated by *Similarity Calculation* and *Fuzzy Set Calculation* mentioned in Section 3.2.3. Next, based on the top 5 most-relevant categories predicted, we compare the browsing web image and the images of top 5 categories by transformed fuzzy sets. The aspect behind this stage is to view the membership values as the conceptual features transformed by visual contents. Given a browsing web image QI and an image DI in the database. The fuzzy similarity between QI and DI is defined as:

$$FSIM_{QI-DI} = \sum_{ca^i \in CA} \sum_{1 \leq x \leq p} |QI.M_{R_x^{ca^i}}^{ca^i} - DI.M_{R_x^{ca^i}}^{ca^i}|, \quad (9)$$

where $QI.M_{R_x^{ca^i}}^{ca^i}$ and $DI.M_{R_x^{ca^i}}^{ca^i}$ denote the membership values in fuzzy region $R_x^{ca^i}$ for image QI and DI , respectively. Based on Equation 9, the most similar images for top 5 categories can be determined. From the aspect of computation cost, the search space is reduced into top 5 categories and the visual retrieval is accordingly accelerated without scanning the whole database.

4. Experimental Evaluations

In this paper, three off-line preprocessing modules and an on-line search agent have been presented for annotating images, conceptualizing query and ranking results, respectively. In this section, we will evaluate our proposed image retrieval by some experiments. In general, the experimental evaluations were conducted in three major aspects: 1) evaluations of the proposed annotation models including $Model_{ISD}$, $Model_{MACk}$ and $Model_{FIM}$, 2) examinations of the proposed concept matching method and 3) demonstrations of the system prototype for visual fuzzy ranking.

4.1 Evaluations of annotation models

In fact, our proposed semantic image retrieval relies heavily on the effectiveness of web image annotation. Without effective annotation, the user's interest cannot be delivered by visual images exactly. To this end, it is necessary to show the robustness of our proposed web image annotation models by following experimental results. For this experiment, the experimental data is a

collection of 10 categories gathered from Google, including *Bear, Cat, Dog, Lion, Tiger, Flower, Grass, Sky, Snow* and *Water*. Each category contains 100 unique web images occurring in 100 different web pages. 50% of experimental data is selected as the training set and the others are adopted to serve the testing experiments. For visual-based model, the used visual features are *Color Layout* and *Homogeneous Texture*. To investigate the effectiveness achieved by our proposed models, three measures, namely *precision*, *recall* and *F-measure*, are used in the experiments. They are defined as:

$$precision = \frac{|Correct|}{|Predicted|} * 100\%, \quad recall = \frac{|Correct|}{|Relevant|} * 100\%,$$

$$F - measure = \frac{2 * precision * recall}{precision + recall},$$

where *Correct* is the correct annotation set, *Predicted* is the resulting annotation set and *Relevant* is the ground-truth annotation set. For example, assume that the ground-truth annotation set is {tiger, grass, sky} and the resulting keyword set is {water, tiger, lion, grass, snow}. Accordingly, F-measure is $(2 * 2/5 * 2/3) / (2/5 + 2/3) = 51\%$.

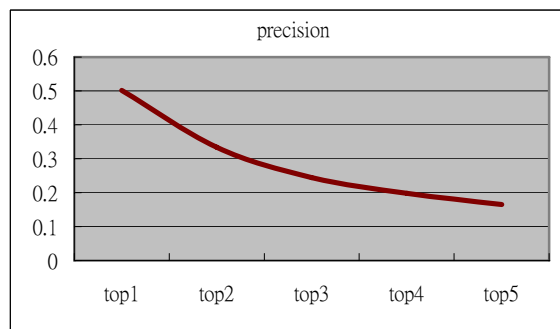


Figure 15. The precisions of $Model_{MACk}$ under different top 5 results.

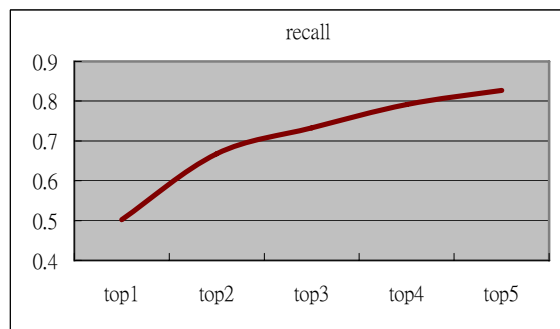


Figure 16. The recall of $Model_{MACk}$ under different top 5 results.

For textual-based annotation, the potential keywords whose *TDegree* cannot exceed the threshold, $\alpha=0$, are truncated. Figures 15 and 16 show that the textual-based annotation model can correctly annotate at least 50%

images for top 1 result in terms of precision and overall more than 80% correct captions for the testing images can be derived within 5 results in terms of recall. In this experiment, some more details are stated as follows. First, the amount of keywords extracted from the web page is almost large, i.e., 2000. Hence, the experimental results reveal that, our proposed textual-based annotation model is very promising for web image annotation within 5 results. Second, some correct predicted keywords are automatically identified incorrect results because of ground-truth diversity and ambiguity. From this viewpoint, the precision and recall could be better if testing manually. Figure 17 further depicts that our proposed textual-based models is more effective than the aged methods including DT (DecisionTree) [21] and textual-based SVM (SVMt) [8] under top 2 results in terms of F-measures. It reveals that: First, classification-based annotators cannot achieve the good annotation quality in dealing with vast textual information. Second, the proposed annotation model performs well even facing the diverse textual information on the Web.

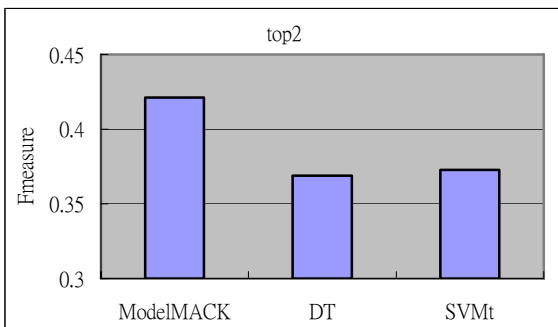


Figure 17. The F-measures of ModelMACK, DT and SVM under top 2 results.

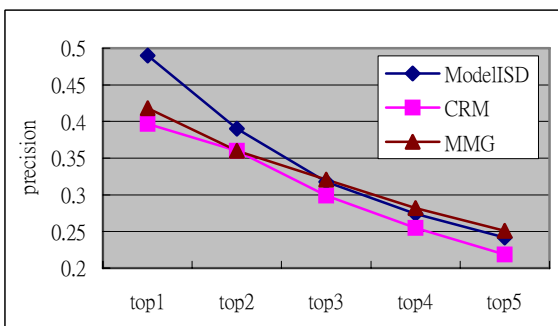


Figure 18. The precisions of ModelISD, CRM and MMG under different top 5 results.

For visual-based annotation, the training data is grouped into 25 clusters. In this evaluation, for each testing image, top 4 most-relevant clusters are selected to calculate the visual degree. Figures 18, 19 and 20 reveal

that Model_{ISD} we propose can bring out the better results than other existing approaches, including MMG [14], CRM [13] and visual-based SVM (SVMv) [8]. The experimental results deliver some aspects. First, classification-based annotator (SVMv) is better than statistics-based ones (CRM and MMG). Second, our proposed visual-based annotation model can tag images more successfully than other current annotators by discovering associations between visual features and captions.

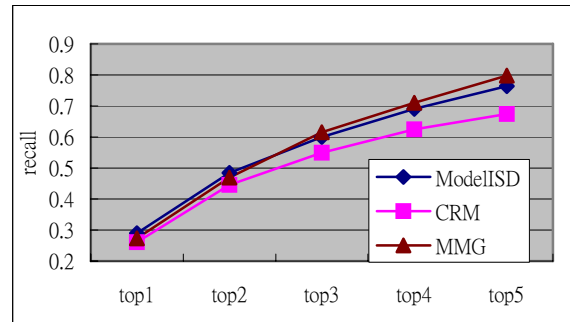


Figure 19. The recall of Model_{ISD}, CRM and MMG under different top 5 results.

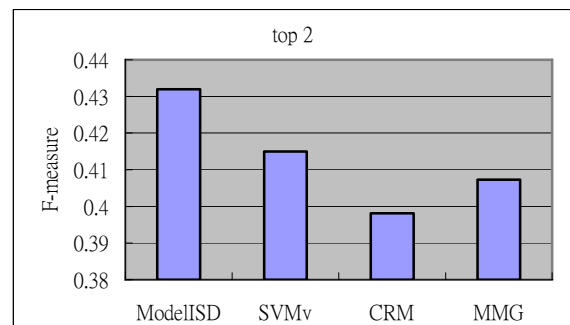


Figure 20. The F-measures of Model_{ISD}, CRM, MMG and SVM under top 2 results.

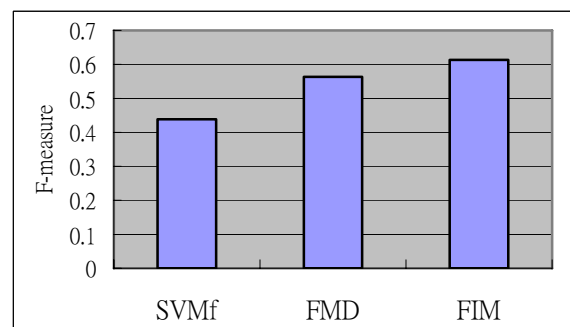


Figure 21. The F-measures for different hybrid annotation approaches.

For hybrid annotation, Figure 21 shows that the results of FIM are still better than other well-known methods including fusion-based SVM (SVMf) [8] and FMD [21].

In this evaluation, SVMf is classification-based fusion method and FMD is the fusion one fusing classification- and statistics-based methods. The experimental results say that, classification only method is worst than hybrid fusion methods. Overall textual-based approaches perform better than visual-based ones. Whatever the annotation model is, our approaches can achieve the better quality of web image annotation. To conclude from above, its reliability can provide semantic image retrieval with good support for hunting the user's interest.

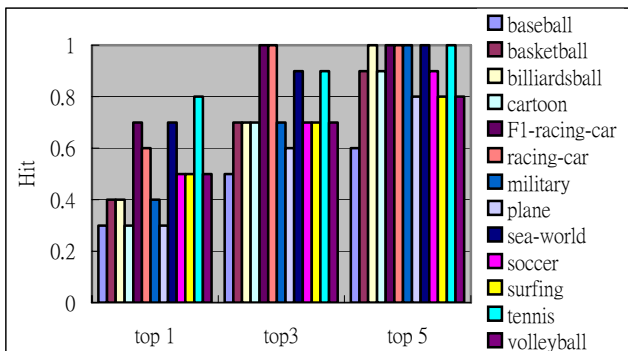


Figure 22. The hits of concept matching for different categories.

4.2 Examinations of concept matching model

After confirming the effectiveness of annotation models, the next evaluation we want to show is the promise of connecting the predicted captions and image categories in the database by our proposed concept matching model. In other words, if the category of the clicked image can be predicted successfully, the user's intention can be presented. The experimental data contains the collection of 13 image categories in real life. For each category, top 2000 most-relevant keywords that came from the top 10 returned results of Google search are adopted as feature keywords. Also for each category, we collected top 10 most-relevant keywords from Wikipedia as the query data. For this experiment, another measure to evaluate the concept matching is defined as:

$$Hit = \begin{cases} 100\%, & \text{if the returned } k \text{ results contain the query term} \\ 0, & \text{otherwise} \end{cases}$$

Hit represents coverage for the correctly returned categories over the resulting ones. Figure 22 shows the experimental results of concept matching in terms of *Hit*. From Figure 22, we can know that most users' interests can be captured in top 5 results. On average, the hit rate can reach 90% for top 5 results. It delivers an aspect that our proposed concept-matching technique is very robust even facing diverse user's queries. Although most hits are larger than 80%, the hit of baseball is 60%. The proper interpretation is that the selected feature-keyword set is not good enough to present the conceptual category

"baseball". It leaves us some room to improve in the future. In summary, our proposed concept matching model can bridge the user's interest to target categories successfully.

4.3 Demonstrations of system prototype

Figure 23 is an elaborate screen snapshot of *iSMIER*. It illustrates a scenario that the user is surfing the web through the proposed *iSMIER*. As she/he takes a look at a preferred image about F1 racing, she/he can submit the interested web image to the search agent by the right-mouse-click. Once the search agent receives the clicked image, annotation module tags the image and delivers the captions to category mapping module. Then the related categories are sorted and presented to the user. Meanwhile, the query image is transformed into fuzzy sets and the relevant images are ranked by calculating fuzzy similarities. Figure 24 is a resulting example for the search results by the proposed *iSMIER*. From viewpoint of system intelligence and effectiveness, it exhibits the proposed *iSMIER* can effectively capture the user's intention from visual images to human concepts. From viewpoint of system usability, the user can obtain the desired images related to her/his browsing web image without downloading any images.

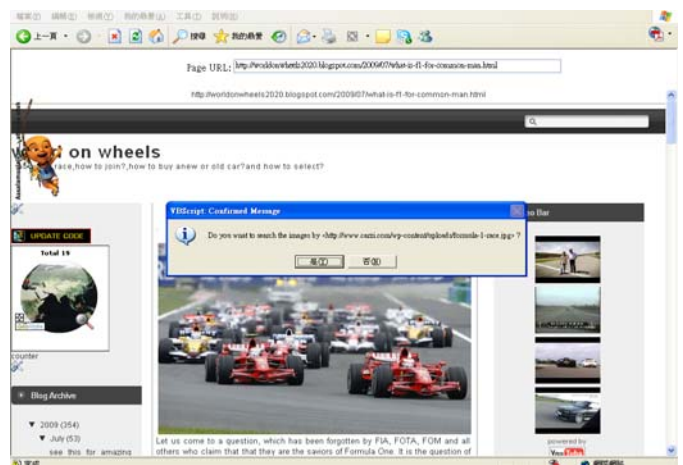


Figure 23. The screen snapshot of *iSMIER*.



Figure 24. The final results of *iSMIER* by fuzzy ranking.

5. Conclusions

In this paper, we have presented a novel image retrieval system named *iSMIER* using image annotation, concept matching and fuzzy ranking techniques. Through this system, the users do not have to define the query terms or download the web image to search the relevant images. The main contribution of this study can be concluded as that we propose a cross-media retrieval paradigm from visual domain to semantic domain, which includes: 1) effective annotation for conceptualizing web images, 2) intelligent concept matching for enhancing semantic image retrieval, 3) efficient fuzzy ranking for accelerating the image retrieval and 4) friendly interface for simplifying the query procedure. Through experimental evaluation, *iSMIER* is shown to have good performance through integrating image annotation, concept matching and fuzzy ranking modules we propose. In the future, we will further apply *iSMIER* to other multimedia retrieval applications.

Acknowledgment

This research was supported by National Science Council, Taiwan, R.O.C. under grant no. NSC 98-2631-H-006-001.

References

- [1] G. Acampora, V. Loia, "Fuzzy control interoperability and scalability for adaptive domotic framework," *IEEE Transactions on Industrial Informatics*, vol. 1, no. 2, pp. 97-111, 2005.
- [2] G. Acampora, C. S. Lee, M. H. Wang, "FML-Based Ontological Agent for Healthcare Application with Diabetes," In *Proc. of Web Intelligence/IAT Workshops*, 2009.
- [3] P. J. Cheng and L. F. Chien, "Personalized Image Browsing and Annotation on the Web Using Query Taxonomy," In *Proc. of International Conf. on Digital Archive Technologies*, 2002.
- [4] E. Chang, K. Goh, G. Sychay, and G. Wu, "CBSA: content-based soft annotation for multimodal image retrieval using Bayes Point Machines," *IEEE Transactions on Circuits and Systems for Video Technology, Special Issue on Conceptual and Dynamical Aspects of Multimedia Content Description*, vol. 13, Issue 1, pp. 26-38, 2003.
- [5] Z. Chen, W. Y. Liu, C. H. Hu, M. J. Li, and H. J. Zhang, "iFind: A Web Image Search Engine," In *Proc. of the 24th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, 2001.
- [6] Y. Chen and J. Z. Wang, "A Region-Based Fuzzy Feature Matching Approach to Content-Based Image Retrieval," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, Issue 9, pp. 1252-1267, 2002.
- [7] C. Djeraba, "Association and Content-Based Retrieval," *IEEE Transactions on Knowledge and Data Engineering*, vol. 15, no. 1, pp. 118-135, 2003.
- [8] H. Feng, R. Shi, and T. S. Chua, "A bootstrapping framework for annotating and retrieving WWW images," In *Proc. of the 12th annual ACM international conf. on Multimedia Technical session 15*, 2004.
- [9] T. P. Hong, K. Y. Lin, and S. L. Wang, "Mining Fuzzy Generalized Association Rules from Quantitative Data under Fuzzy Taxonomic Structures," *International Journal of Fuzzy Systems*, vol. 5, no. 4, pp. 239-246, 2003.
- [10] J. Jeon, V. Lavrenko, and R. Manmatha, "Automatic Image Annotation and Retrieval using Cross-Media Relevance Models," In *Proc. of the 26th annual international ACM SIGIR conference on Research and development in information retrieval*, 2003.
- [11] A. Kandel, *Fuzzy Expert Systems*, CRC Press, Boca Raton, pp. 8-19, 1992.
- [12] R. Krishnapuram, S. Medasani, S. H. Jung, Y.-S. Choi, and Rajesh Balasubramaniam, "Content-Based Image Retrieval Based on a Fuzzy Approach," *IEEE Transactions on Knowledge and Data Engineering*, vol. 16, no. 10, pp. 1185-1199, 2004.
- [13] V. Lavrenko, S. L. Feng, and R. Manmatha, "Statistical Models for Automatic Video Annotation and Retrieval," In *Proc. of the International Conf. on Acoustics, Speech and Signal Processing*, 2004.
- [14] J. Y. Pan, H. J. Yang, C. Faloutsos, and P. Duygulu, "Automatic multimedia cross-modal correlation discovery," In *Proc. of the 10th ACM Int'l Conf. Knowledge discovery and data mining*, 2004.
- [15] J. R. Smith and S.-F. Chang, "VisualSEEK: A fully automated content-based image query system," In *Proc. of the 4th ACM international Conference on Multimedia*, 1996.
- [16] J. R. Smith and S. F. Chang, "An Image and Video Search Engine for the World-Wide Web," In *Proc. of IS&T/SPIE Symposium on Electronic Imaging: Science and Technology (EI'97) - Storage and Retrieval for Image and Video Databases*, 1997.
- [17] H. M. Sanderson and M. D. Dunlop, "Image retrieval by hypertext links," In *Proc. of ACM SIGIR*, 1997.
- [18] H. T. Shen, B. C. Ooi, and K.L. Tan, "Giving meaning to WWW images," In *Proc. of the 8th*

annual ACM international conference on Multimedia, 2000.

- [19] Y. H. Tian, T. Huang, and W. Gao, "Exploiting multi-context analysis in semantic image classification," *Journal of Zhejiang University SCIENCE*, vol. 6, no 11, pp. 1268-1283, 2005.
- [20] V. S. Tseng, J. H. Su, J. H. Huang, and C. J. Chen, "Integrated Mining of Visual Features, Speech Features and Frequent Patterns for Semantic Video Annotation," *IEEE Transactions on Multimedia*, vol. 10, no. 1, pp. 260-267, 2008.
- [21] V. S. Tseng, J. H. Su, B. W. Wang, and Y. M. Lin, "Web Image Annotation by Fusing Visual Features and Textual Information," In *Proc. of the 22nd ACM Symposium on Applied Computing*, 2007.
- [22] X. J. Wang, "Annotating Images by Mining Image Search Results," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 30, no. 11, pp. 1919-1932, 2008.
- [23] L. Wenyin, S. Dumais, Y. Sun, H. Zhang, M. Czerwinski, and B. Field, "Semi-automatic image annotation," In *Proc. of International Conference on HCI*, 2001.
- [24] R. C. F. Wong and C. H. C. Leung, "Automatic Semantic Annotation of Real-World Web Images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 30, no. 11, pp. 1933-1944, 2008.
- [25] J. Z. Wang, J. Li, and G. Wiederhold, "SIMPLiCity: Semantics-Sensitive Integrated Matching for Picture Libraries," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, no. 9, pp. 947-963, 2001.
- [26] H. Xu, X. Zhou, and L. Lin, "WISA: A Novel Web Image Semantic Analysis System," In *Proc. of ACM SIGIR*, 2008.



Ja-Hwung Su received his Ph.D. degree in the Department of Computer Science and Information Engineering at National Cheng Kung University. Currently, he is a Postdoctoral Fellow in the Department of Computer Science and Information Engineering at National Cheng Kung University. His research interests include multimedia mining, web mining,

information retrieval and data warehousing.



Bo-Wen Wang received his B.S. and M.S. degrees in the Department of Computer Science and Information Engineering at National Chiao Tung University in 1988 and 1990, respectively. He is currently a Ph.D. student in the Department of Computer Science and Information Engineering at National

Cheng Kung University. His research interests include data mining and multimedia mining.



Tien-Yu Hsu received his Ph.D. degree in the Department of Computer Science and Information Engineering at National Chiao Tung University. Currently, he is an Assistant Curator in the Information Science Department at National Museum of Natural Science, Taiwan, R.O.C. His research interests include digital museum, digital archive, E-learning, digital contents analysis and knowledge management.



Chien-Li Chou received his B.S. degree in the Department of Computer Information Science at National Chiao Tung University. He is currently a M.S. student in the Department of Computer Science and Information Engineering at National Cheng Kung University. His research interests include data mining and multimedia mining.



Vincent S. Tseng is currently a professor in the Department of Computer Science and Information Engineering as well as the director for Institute of Medical Informatics at National Cheng Kung University, Taiwan. He received the Ph.D. degree from National Chiao Tung University, Taiwan. Dr. Tseng's main research interest is data mining and he has published more than 170 papers. He is a member of IEEE and ACM.