Image Annotation System Using Visual and Textual Features

Swapnil Barai and Alfonso F. Cardenas

Computer Science Department

University of California, Los Angeles, USA

Abstract

We present an automated Image Annotation system called I-Tag which uses both visual and textual information of the images and recommends relevant tags for them. The automatic generation of metadata would allow image searches and content-based image retrieval (CBIR) to be more effective. We use state of the art tools on text based retrieval and image content based retrieval to retrieve similar images. We use an open source image retrieval engine called FIRE [8], to perform content based image search. We use the visual features as well as the title of the image to recommend tags for the image with high relevance. Experiments are conducted on approximately 5000 images downloaded from FLICKR to compare the accuracy of the recommended tags. Our experiments show that the mean average precision of the recommended tags is 69%.

Index terms: image annotation, tagging, image retrieval, content-based image retrieval.

1. Introduction

Image annotation also known as image tagging is the process of assigning metadata in the form of captions and keywords to digital images. Developing effective methods for automated annotation of images and the consequent retrieval of images aided by such annotations is a relatively new and highly active area due to its wide importance in the internet, and still an open research area [1]–[35]. Major search engines, such as Google, MSN, and Yahoo! that provide image search depend heavily on the name of the images, their textual description and the text surrounding the images, found on the Web pages containing these images. Hence computer assisted annotation of images would guarantee their visibility as search engine result.

Manual image annotation has been in existence for quite a while. Image sharing websites such as Picasa[34], Flickr[13], etc. give their users the ability to annotate the images while uploading them. These annotations are generally one or two words describing the image. However the major drawback of this technique is that the tags are user specific. Various attempts have been made towards automating this manual process [1]. Some analyze the information generally present in the tags that people give to their images, and henceforth recommend tags for their images [10]. Some World Wide Web image search engines attempt to automate annotation by using text from the image filename and text near the image on a webpage. However, search results using this method usually contain many irrelevant images. Most of the real world systems just rely on the textual features that are present with the image such as the title, the text present around an image on the webpage from where it is uploaded [11]. Also Google has made attempts towards this by involving the user in understanding the semantics of the images by developing an Image-Tag game Image Labeler [11]. However they have not come up with an Image tag recommendation system yet.

Efforts have been devoted also for more computer-aided automatic annotation based on image features. To annotate an image, different visual features, for example, color, texture, and shape are extracted. Then, using some classification algorithm, an automated system will be able to annotate the image. An image annotation system is very closely related to image retrieval system. An image retrieval system can retrieve images using the text in the annotated images which is more naturally represented in comparison to non-annotated images. There are three basic image annotation approaches: free text annotation, keywords and ontologies [1]. Keywords annotation is the use of arbitrarily chosen keywords or keywords from controlled vocabularies and is the most widely used. Free text annotation does not have any restriction on how to use text to annotate image. Ontologies are the use of a large classification system that classifies different aspects of life into hierarchical categories [23]. ImageCLEF provides the visual concept detection task [25] and medical image annotation task [24] for this area. Both task aims to annotate images using ontologies. There have also been many efforts to reduce the annotation retrieval gap [1],[2],[3],[4],[7] which refers to the difference in metadata generated by the human experts and those by systems for the same image. Automatic annotation of images using Statistical Learning Approach [5] and hidden Markov models [6] has also been studied. However reliable automatic image annotation still remains an open problem.

We present in this paper an automated Image Annotation system called I-Tag which uses both visual and textual information in the title of the images and recommends relevant tags for them. We believe that

by accounting for both visual as well as such textual features of images it is possible to reduce the annotation retrieval/semantic gap and provide highly relevant tags for them. [35] tries to utilize both visual and textual features, but therein the textual features refer to the text surrounding the image on the webpage from where it is taken, and this would lead to the semantic gap problem.

2. Dataset

The dataset consists of 5000 images downloaded from Flickr [13]. Flickr is a website that allows users to upload images to it and provides online image management. Through Flickr, users can share their images with other Flickr users as well as other people in the world.

There are several reasons for choosing to work with Flickr. Firstly, they have a large number of images available for download since they have a lot of users uploading images to their website every day. This makes it easier to build our data set. Also along with the images the metadata associated with them is also available. Flickr provides an API for developers to easily access its information, such as the images, user information, and image information. Flickrj is an API kit developed for Java to access the Flickr API and is a Java API which wraps the REST-based Flickr API. It covers most of the functionality that Flickr API provides. Images in the interesting photo section, along with their titles, tags, and number of favorite, are downloaded directly using Flickrj.

The downloaded images are chosen such that at least one of the following words, "Clouds", "Flower", "Tree", "Beach" , "Sunset", "Winter", "Night", "Grass", "Dog", and "Water", in their title or tags are chosen to build up the Flickr data set to be used in the experiment. We also downloaded images having two of these keywords occurring together in their title or tags such as, "Beach" and "Water" or "Winter" and "Night" etc. Similarly we also had in our dataset images with triplets of these keywords and images with quadruples of these keywords occurring together in their title or tags. We conducted separate experiments on these sets of images and observed the variation of results. We do not use more than quadruples of keywords as there was no significant improvement in our results after that. Using interesting photos and restricting the images to have certain keywords ensured the quality of the photos and reduced the noise of the dataset. The reason for using these keywords is because they appear more frequently in the title or the tag of an image and therefore, the data set is easier to build. However, our system should work well with other less popular terms and image data sets – depending of course in the quality of tags assigned by the public to those images.

3. Image annotation system and implementation

In this section, we describe the image annotation system I-TAG that we have built, the components of the system and their implementation, and content-based image retrieval system highlights.

3.1 Overview

The objective of the textual and content based image annotation system is to allow the users to give images to the I-TAG tool and get back relevant tags for them as the result. Fig. 1 shows a high level block diagram illustrating our system. There are three main tasks involved in building an annotation system for images. The first task is to build a dataset for the annotation system, comprising of images from Flickr. This is done by the administrator of the system. The images for the dataset are downloaded from Flickr along with their respective title and tags. Then the user of the system gives an image to the system as input, which is to be annotated automatically by the system, referred to as the Query Image X. The next task for the administrator is to find images in the downloaded dataset, that are visually similar to the query image X. To perform this task state of the art content-based image retrieval engine is used that can extract visual features of a given image and return images that are visually similar to it. Also along with each of these similar images it generates a similarity score ranging from 0 to 1. Higher similarity score implies that the image is more similar to the original query image. The administrator orders these images based on this similarity score. Then the administrator extracts textual features (tags and/or titles) of all these similar images and the query image itself and forms a list of all the keywords. Finally the administrator finds the most frequently occurring keywords in this list and recommends it as the tags for the query image X.

3.2 Image Database

First we download images along with their tags from Flickr and build up our dataset as mentioned earlier. Obtaining this information is relatively easy since Flickr provides an API for developers. Flickrj [14] is an API kit developed for Java to access the Flickr API. It is a Java API which wraps the REST-based Flickr API. We store the image id, image title and the image tags in a MySQL database [9]. Next we divide the dataset into testdata (for which to generate its tags) and traindata i.e. we store them as two different MySQL tables. The testdata table has the following columns: image id, image title and tags that are to generated. The traindata table has the following columns: image id, image title, original tags (tags that are downloaded along with the image from flickr) and recommended tags (this column would be empty initially). The images

in the traindata form the dataset to submit to the image retrieval engine FIRE. The images in the testdata form the set of query images X, i.e. those images for which our tool has to recommend tags. The tags that have been downloaded along with these test-data images from Flickr are the Original or the 'Ideal' tags and serve as the comparison base in the evaluation of the tags recommend by I-TAG for these images.

3.3 Content Based Image Retrieval

The Flexible Image Retrieve Engine (FIRE)[8] is used as a "black box" to perform content-based image search in our I-TAG system. FIRE has participated in the ImageCLEF since and has shown to be one of the best among other content based image retrieval system (CBIR) participants in both photo and medical retrieval tasks. It is available for download on the FIRE website. The initial screen of FIRE shows several randomly chosen from the database to present to the user. The user can choose one of the images to start a query, ask the system to present more random images or upload one's own image as a query image. Unique to FIRE, but not to most of the several available CBIR systems, is the flexibility for a user to upload an arbitrary set or database of images and then obtain from FIRE the images in it most similar to an arbitrary query image or images provided by the user.

After getting back the result from FIRE, the user can start a new query by choosing other random images or uploading his own image. The user can refine the retrieval result by giving feedback to FIRE using three buttons below the image, the user can tell FIRE if the image is relevant or not: the green plus sign means the image is relevant, the red minus sign means the image is not relevant and the grey sign is chosen by default and means that the user does not have any opinions on that image. The user can look at the information of a particular image in the retrieval result, the full size image and the information of the visual features extracted.

Feature Extraction and Image Retrieval

For each image, FIRE extracts five visual features: color histogram, Tamura features, global texture descriptor, invariant feature histograms, and sparse patch histograms. Invariant feature refers to a feature that does not change when transformations are applied to an image. The transformations considered are translation, rotation, and scaling [30]. The invariant feature histogram is used in FIRE.

To decide which image in the database should be retrieved for an image query X, the distance score between the query image and each image I in the database is calculated. Each feature is assigned a weight and the distance score is the weighted sum of all the features.

We have also manipulated the code of FIRE such that it returns the name of the similar images as well in the final result. This is required specifically for our tool, as once we get the name of these similar images then only can we retrieve their textual information (title and tags) from our database.

3.4 Tag Recommendation

Now we can recommend tags for the query images X s. Using FIRE we got the most similar images (from the train dataset) to the query image. We use the image titles of these similar images to form a list of all the tags that have been associated with these similar images. We then find the most popular tags among this list and add them to the recommended tags column in the test dataset for that particular query image.

To improve our results we have also added the title of the similar images as well as the title of the query image to the list that we formed of all the associated tags. This helps as follows: for example the query image has title 'beach volleyball', but due to some reason the tag 'volleyball' is not a part of the most popular tags list and hence is not recommended. Thus by adding the title of the query image we are increasing the chances of the tag 'volleyball' to be returned as one of the recommended tags. The same reason holds good for why we are adding the titles of the similar images to our list. To summarize the tool should use all the data that is available to it for the task of recommending tags.

4. Experiments and Results

This section explains the four experiments performed over our dataset. The images in the test dataset act as the query images for our system. Similar images for these query images are searched from the images in the train dataset. The experiments compare the tags that are recommended for the query images by our system, with the "Ideal Tags" of these images or the tags that were associated with the Image on Flickr. Our goal is to show that the tags recommended by our system are accurate and relevant.

We have four different sets of images in our dataset. The first one comprises the images that have at least one of the following words, "Clouds", "Flower", "Tree", "Beach" , "Sunset", "Winter", "Night", "Grass", "Dog", and "Water", in their title or tags. The second has images with two of these keywords occurring together in their title or tags such as, "Beach" and "Water", "Winter" and "Night" etc. Similarly third and fourth datasets have images with triplets of these keywords and images with quadruples of these keywords occurring together in their title or tags. We refer to these sets of images as 'One-Word queried images', 'Two-Word queried images', 'Three–Word queried images' and 'Four-Word queried images' respectively. We have

conducted separate experiments on these sets of images and observed the variation of results.

For each of these sets of images we divide them into test-dataset and train-dataset. Test-dataset has the images that are uploaded to our system one by one for annotation i.e., they contain all the Query images. Each query image is then uploaded to our system. Next the administrator of the system finds images in the train dataset that are visually similar to the query image using FIRE. FIRE retrieves the ten most similar images to the query image, based upon a similarity score. After retrieving the ten most similar images to the query image, the administrator passes the titles of all these similar images to our database which has the textual information of all these images. Next using the titles of these similar images the administrator finds the tags that are associated with these images in the traindata table and forms a list that contains the titles and tags of the similar images. To the list we also add the title of the query image. This is done to improve the accuracy of recommended tags. Now from this list thus formed, we find out the most frequently occurring keywords and recommend them as tags for the query image. Finally we store these recommended tags in the "recommended tags" column of the testdata table for the respective query image. We repeat these steps for each of the query images in the four different sets of images we have. The following example illustrates all this visually:

Suppose the user gives the following image as input to the system.
Title: Beach Volleyball
Ideal/Original Tags: Beach, Volleyball, Sport, Sand
Recommended Tags: To be recommended by the annotation system

This image becomes our Query image X and has been selected from the test dataset. Table 1 shows a snapshot of the test dataset (before the annotation of the image). The tags that have been downloaded along with these test-data images from Flickr are the Original or the 'Ideal' tags. It is clearly visible that this image is a part of the One-Word queried images set, as it contains only one keyword 'Beach', out of our list of ten keywords, in either its title or tags.

Then our system extracts the visual features of the query image X and using the CBIR FIRE system finds visually similar images to it, in the images in the train dataset of our system. FIRE returns the following images are the four most similar images in the train dataset to the query image X.

Title: Volleyball 123X104

Next the titles of these similar images are sent to the train dataset table, to get the Tags associated with the respective images. Table 2 shows a snapshot of the Train Dataset table.

Image Id	Image Title	Ideal Tags	Recom. Tags
1	Beach Volleyball	Beach,Volleyball, Sport,Sand	

Table 1: Snapshot of Test Dataset Table before the recommendation process

Image Id	Image Title	Tags
1	Volleyball	Players,Shot,Ball,Sport
2	123X104	Beach,Net,Volleyball
3	Game	Sand,Volleyball,Jump,Sport
4	Volleyball Match	Sport, Net, Beach, Shot

Table 2: Snapshot of Train Dataset Table

Now the administrator forms a list of Tags and titles of the similar images and also including the title of the query image. Hence the list is now:
Beach Volleyball, Volleyball, 123X104, Game,
Volleyball Match, Players, Shot, Ball, Sport, Beach, Net,

Volleyball, Sand, Volleyball, Jump, Sport, Sport, Net, Beach, Shot.

Finally from the list thus formed, the administrator finds out the most frequently keywords and recommends them as tags for the query image to the user, as shown below.

Query/Input Image
Title: Beach Volleyball

Output, recommended tags:
1) Volleyball, 2) Beach, 3)Shot, 4) Sport

The test dataset after this is:

Image ID	Image Title	Ideal Tags	Recom. Tags
1	Beach Volleyball	Beach, Volleyball, Sport, Sand	Volleyball, Beach, Shot, Sport

Table 3: Snapshot of Test Dataset Table after the recommendation process

Result evaluation
The results are evaluated using precision, average precision and mean average precision (MAP)

Precision = (Number of recommended tags) ∩ (Number of original/'Ideal' Tags) / (Number of recommended tags)

Average precision refers to the average of the precision for the five query images of the same word. Here is the definition of average precision:

Average Precision = Sum of Precision / Number of Query images

MAP is the average of all the average precisions. It is calculated for each sets of images, i.e. One-Word Images, Two-word Images, Three-word images and Four-word Images.

MAP = Sum of Average Precision of all query images / Number of Query images

Usually, recall and F-score are also used to evaluate the result of retrieval systems. However, they are not used here as they do not have much applicability.

Recall = (Number of recommended tags) ∩ (Number of original/'Ideal' Tags) / (Number of original/'Ideal' Tags)

$F\beta$ = (1 + β2) (Precision x Recall) / (β2 Precision + Recall)

The number of original/Ideal tags that any test-image has, is totally dependent on the user who has uploaded that image (thus is quite variable). As recall is significantly dependent on the number of original/ideal tags, its value does not us tell much about the efficiency of our system. Also F-score is dependent on recall, hence it is also not a useful measure of the efficiency of our system.

Precision results
The following table shows the results for Three-word query images, i.e. images that have any three of the following keywords "Clouds", "Flower", "Tree", "Beach", "Sunset", "Winter", "Night", "Grass", "Dog", and "Water" occurring together in their title or tags. We have approximately 1200 Three-Word query images (divided into testdata and traindata) in our dataset.

Query Image having keywords:	Average Precision (%)
Winter,Night,Tree	67
Beach,Sunset,Water	68
Flower,Tree,Grass	68

Table 4: Results of Three-Word Query Images

Thus the Mean Average Precision for the Three-Word query images is Sum of all Precisions/3= **67.6%.**

The comparison of results for One-word, Two-word, Three-word and Four-Word query images are shown below in Fig. 2. We observe good precision values in recommending tags for the images. We can easily see that the precision value for Four-Word query images is as high as 69%. Our remarks are qualitative opinions because there is no comparable data set in the literature for comparison.

Thus it is clearly visible that as the number of keywords increase in the query images, the precision values for the annotations also improve as our results show in Fig. 2. This can be accounted for the fact, that as the number of keywords increase in the query image, the image becomes more specific and users ought to tag it with similar keywords and hence our system gets better results

During the course of the experiments another important observation was made. Until now we were assuming the tags for the query images given by the users on Flickr as 'ideal tags', and we were trying to compare the tags recommended by our system for that image with its 'ideal tags' and calculate the precision. However it might be that these ideal tags originally given by the user(s) are not very accurate. Hence we manually tried to observe the relevancy of tags (as a neutral user) recommended by our system for a sample set of 100 images. We found out that for every 1 in 10 images the precision of tags was actually better than the precision calculated based upon the 'ideal tags'. Hence the precision goes up by 1.1% incrementally, and to a total level of 70.1% (for Four-word queried images).

Comparison to other annotation approaches is very difficult because their image test data sets and experiments details are not published.

.5.Conclusions

In this project we have developed an automatic image annotation tool I-TAG which uses both visual and textual features and recommends relevant tags for them with high relevance. We showed that by accounting for both visual and textual features of the image that we are able to annotate the images with high precision. We chose the Flexible Image Retrieval Engine (FIRE) to perform similar image search based on visual features. FIRE is one of the best content based image retrieval systems in the ImageCLEF photographic retrieval task.

The quality or accuracy of the image similarity by FIRE, or any other system that could be used instead, has of course a direct influence on the quality and accuracy of the tags. We have used MYSQL as our database, to store the textual information associated with images and extract this information when required.

To annotate an image, first the user gives an image to the system which is called the query image. Then our system finds images in our database that are visually most similar to the given image using FIRE. Then the administrator of the system extracts the textual information i.e. titles and tags of these similar images and forms a list. To this list, the title of the query image is also added. Finally the most frequently occurring keywords in this list are recommended as tags for the query image.

We have downloaded a dataset of 5000 images from FLICKR between the dates of August 2009 and October 2009. The title or tags of these images has at least one of the following keywords "Clouds", "Flower", "Tree", "Beach" , "Sunset", "Winter", "Night", "Grass", "Dog", and "Water" for restricting the dataset.

Experiments were conducted to evaluate the accuracy of our annotation system. The tags recommended by our system are compared with the original/ideal tags of these images provided by users to calculate precision. Testing with independent user judges

should be done in the future. We have tested four kinds of image sets: One-word query images, two-word query images, three-word query images and four-word query images. Experiments are performed separately on these four sets of images. Our results based on average precision and mean average precision are shown in Fig. 2, indicating maximum accuracy for four-word query images.

6.References

[1] A. Hanbury, "A Survey of Methods for Image Annotation", Journal of Visual Languages and Computing, vol. 19, Issue 5, pp. 617-627, October 2008.

[2]Datta, R., Ge, W., Li, J., and Wang, J. Z. 2006. "Toward bridging the annotation-retrieval gap in image search by a generative modeling approach", In Proceedings of the 14th Annual ACM international Conference on Multimedia (Santa Barbara, CA, USA, October 23 - 27, 2006), MULTIMEDIA '06. ACM, New York, NY, 977-986.

[3] Irfanullah, Nida Aslam, Kok-Keong Loo,Roohullah, "Semantic Annotation Gap: Where to put Responsibility?", JDCTA: International Journal of Digital Content Technology and its Applications, Vol. 3, No. 1, pp. 94 -97, 2009

[4] L. Hollink, G. Schreiber, J. Wielemaker, and B. Wielinga, "Semantic annotation of image collections", In Workshop on Knowledge Markup and Semantic Annotation, KCAP'03, 2003.

[5] Baigang Huang, Junshan Li, Shuangyan Hu, "Automatic Annotation of Images by a Statistical Learning Approach," *International Conference on Natural Computation*, vol. 4, pp. 477-480, 2008 Fourth International Conference on Natural Computation, 2008.

[6] Ghoshal, A., Ircing, P., and Khudanpur, S. 2005. Hidden Markov models for automatic annotation and content-based retrieval of images and video. In *Proceedings of the 28th Annual international ACM SIGIR Conference on Research and Development in information Retrieval* (Salvador, Brazil, August 15 - 19, 2005). SIGIR '05. ACM, New York, NY, 544-551

[7] Stephan Bloehdorn, Kosmas Petridis, Carsten Saathoff, Nikos Simou, Vassilis Tzouvaras, Yannis S. Avrithis, Siegfried Handschuh, Ioannis Kompatsiaris, Steffen Staab, Michael G. Strintzis – "Semantic Annotation of Images and Videos for Multimedia Analysis". In ESWC pp. 592-607, 2005.

[8] Deselaers, T., "Features for Image Retrieval", *Computer Science Department*, Aachen, Germany, RWTH Aachen University, 12/2003.

[9] MySQL http://www.mysql.com/

[10] Sigurbjörnsson, B. and van Zwol, R. 2008. Flickr tag recommendation based on collective knowledge. In *Proceedings of the 17th international Conference on World Wide Web* (Beijing, China, April 21 - 25, 2008). WWW '08. ACM, New York, NY, 327-336

[11] Behold http://www.behold.cc/

[12] Google Image Labeler http://images.google.com/imagelabeler/

[13] Flickr. http://www.flickr.com/

[14] Flickrj. http://flickrj.sourceforge.net/

[15] M.Grubinger, P. Clough, A. Hanbury, and H. Muller, "Overview of the ImageCLEFphoto 2007 Photographic Retrieval", Advances in Multilingual and Multimodal Information Retrieval: 8th Workshop of the Cross-Language Evaluation Forum, CLEF 2007, Budapest, Hungary, September 19-21, 2007.

[16] C. Faloutsos, R. Barber, M. Flickner, J. Hafner, W. Niblack, D. Petkovic, W. Equitz, "Efficient and effective querying by image content", J. Intell. Inf. Syst. 3 (3–4), 1994.

[17] A. Pentland, R.W. Picard, S. Scaroff, "Photobook: content-based manipulation for image databases", Int. J. Comput. Vision 18 (3), 1996.

[18] J. R. Bach, C. Fuller, A. Gupta, A. Hampapur, B. Horowitz, R. Humphrey, R. Jain, and C.-F. Shu, "The Virage image search engine: An open framework for image management", In I. K. Sethi and R. C. Jain, (editors), IS&T/SPIE Proceedings of Storage & Retrieval for Image and Video Databases IV, vol. 2670, pp. 76-87, San Jose, 1996.

[19] J.R. Smith, S.F. Chang, "VisualSeek: a Fully Automatic Content-Based Query System", Proceedings of the Fourth ACM International Conference on Multimedia, pp. 87–98, 1996.

[20] Elisa Drelie Gelasca, Joriz De Guzman, Steffen Gauglitz, Pratim Ghosh, JieJun Xu, Emily Moxley, Amir M. Rahimi, Zhiqiang Bi and B. S. Manjunath, "CORTINA: Searching a 10 Million + Images Database", Technical Report, VRL, ECE, University of California, Santa Barbara, Sep. 2007.

[21] J.Z. Wang, J. Li, G. Wiederhold, "SIMPLIcity: Semantics-Sensitive Integrated Matching for Picture Libraries", IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 23, no. 9, pp. 947–963, 2001.

[22] Gass, T., T. Weyand, T. Deselaers, and H. Ney, "FIRE in ImageCLEF 2007: Support Vector Machines and Logistic Regression to Fuse Image Descriptors in for Photo Retrieval", Advances in Multilingual and Multimodal Information Retrieval 8th Workshop of the Cross-Language Evaluation Forum, CLEF 2007, vol.

5152, Budapest, Hungary, Springer, September 2007.

[23] E. Hyvonen, A. Styrman, S. Saarela, "Ontology-based image retrieval", Proceedings of XML Finland Conference, pp. 27-51, 2002.

[24] T. Deselaers, T. Deserno, N. Muller, "Automatic medical image annotation in ImageCLEF 2007 Overview, results, and discussion", Pattern Recognition Letters, vol. 29, issue 15, pp. 1988-1995 November 2008.

[25] T. Deselaers, and A. Hanbury, "The Visual Concept Detection Task in ImageCLEF 2008", CLEF Workshop 2008 / Evaluating Systems for Multilingual and Multimodal Information Access, Aarhus, Denmark, Springer, 2009

[26] Li, J. and Wang, J. Z. 2008. Real-Time Computerized Annotation of Pictures. IEEE Trans.

[27] Pon, R. K., Cárdenas, A. F., Buttler, D., Critchlow, T., "iScore: Measuring the Interestingness of Articles in a Limited User Environment", IEEE Symposium on Computational Intelligence and Data Mining 2007, Honolulu, HI, April 1-5, 2007.

[28] 3. F. Tushabe, M. H. F. Wilkinson (2008): "Content-based Image Retrieval Using Combined 2D Attribute Pattern Spectra", in book: Advances in Multilingual and Multimodal Information Retrieval by Lecture Notes in Computer Science, Springer, Vol. 5152, ISBN 978-3-540-85759-4.

[29] www.wikipedia.com

[30] T. Deselaers, "Features for image retrieval", diploma thesis, RWTH Aachen University, Aachen, Germany, 2003.

[31] H. Tamura, S. Mori, T. Yamawaki, "Textural Features Corresponding to Visual Perception", IEEE Transaction on Systems, Man, and Cybernetcs, Vol. SMC-8, No. 6, pp. 460–472, June 1978.

[32] J. Puzicha, Y. Rubner, C. Tomasi, J. Buhmann, "Empirical Evaluation of Dissimilarity Measures for Color and Texture", International Conference on Computer Vision, Corfu, Greece, 1999.

[33] http://images.search.yahoo.com/

[34] http://picasa.google.com/

[35] X. Li, L. Chen, L. Zhang, F. Lin, and W. Ma, "Image annotation by large-scale content-based image retrieval," in Proceedings of the 14th Annual ACM international Conference on Multimedia, 2006.
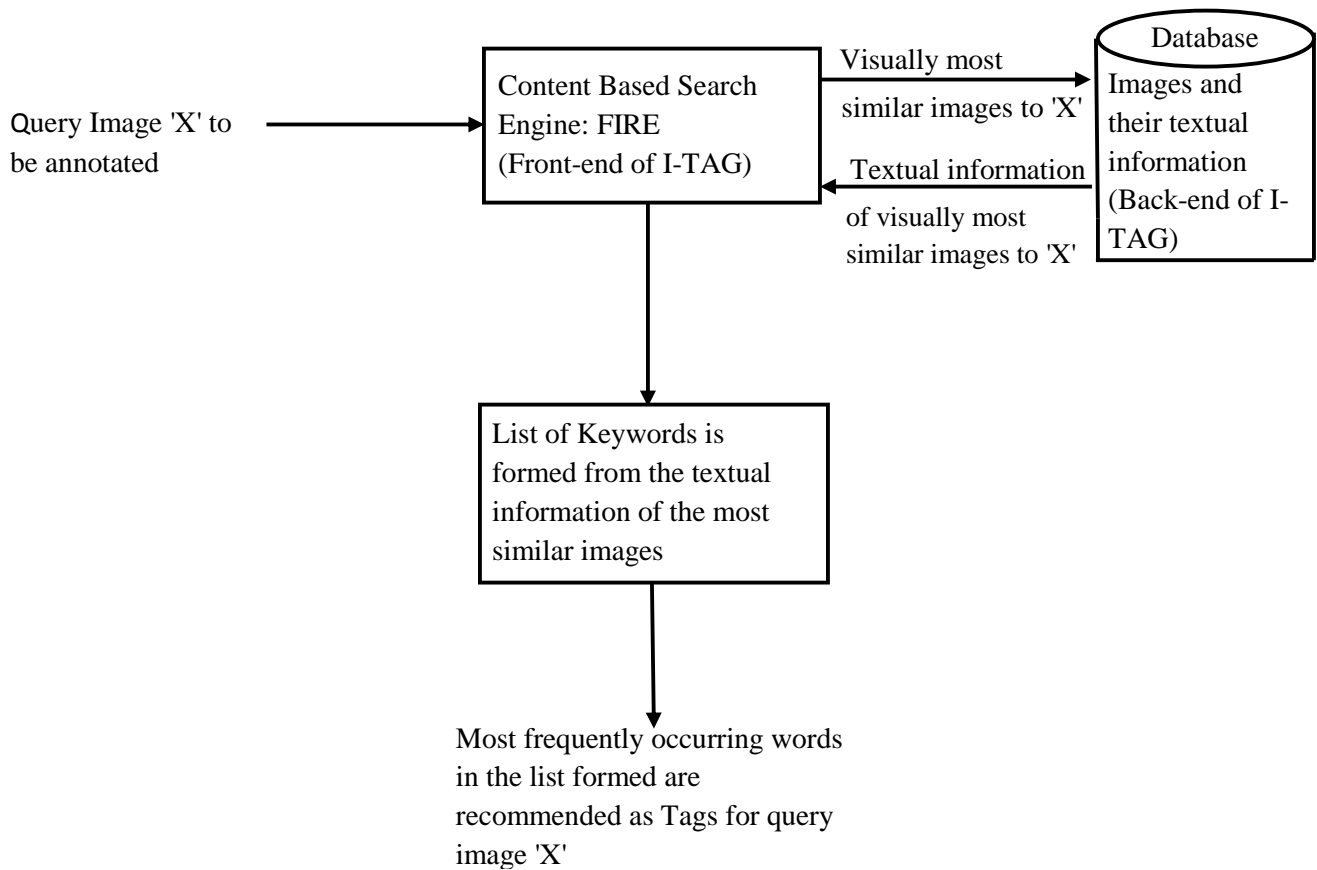
Query Image 'X' to be annotated → Content Based Search Engine: FIRE (Front-end of I-TAG)

Visually most similar images to 'X' →

Database

Images and their textual information (Back-end of I-TAG)

Textual information of visually most similar images to 'X'

List of Keywords is formed from the textual information of the most similar images

Most frequently occurring words in the list formed are recommended as Tags for query image 'X'
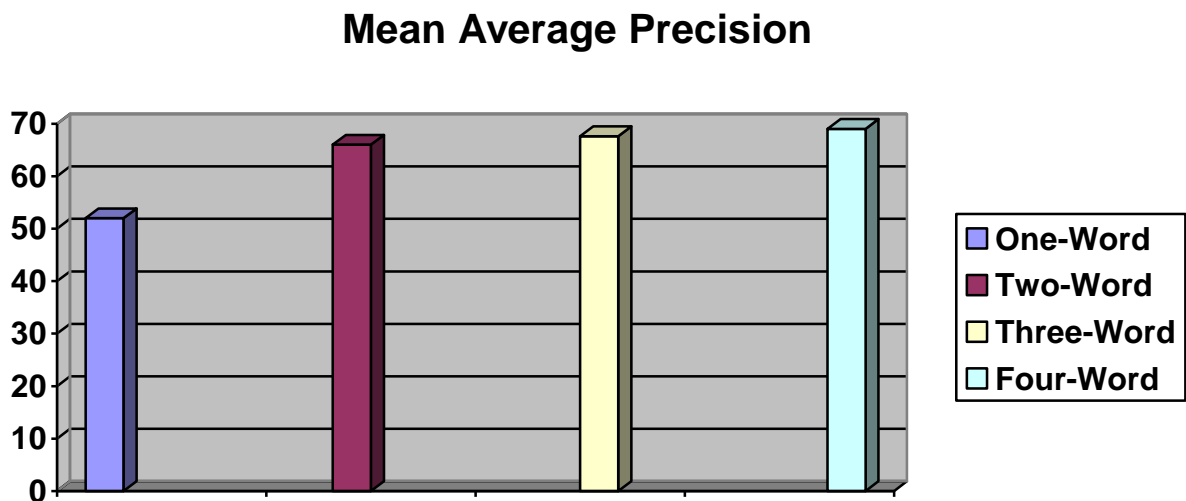
Fig 1: I-Tag system overview

Mean Average Precision

Fig 2: Mean Average Precision comparison