Volume 5, Issue 3, March 2015 ISSN: 2277 128X
International Journal of Advanced Research in Computer Science and Software Engineering
Research Paper
Available online at: www.ijarcsse.com

A Multimodal Approach for Image Annotation and Searching

Adnan Sddiqui, Nishchol Mishra, Jitendra Singh Verma
School of IT, RGPV Bhopal
M.P., India

Abstract— Several models have been prepared for searching images which utilize different techniques. For each of new technique different outcomes are obtained which have probability to search the required image as per user requirement on the basis of query. So targeting the most probable paper makes the very low precision. The basic component here is to collect different features of the image so that searching can be made more effective. The optimization of the searching it is required that all the visual features of the image should utilize. In this paper, the visual content of image is use for annotation with multimodal approach. Proposed work shows the better result as compare to previous works.

Index Terms— Automatic image annotation, image Tagging, Image Query, image retrieval, annotation models.

I. INTRODUCTION

Traditional approach is used for searching the information over the internet, i.e., searching the information on the basis of textual data. But compared to text, visual content is easy to identify. As the technology grows rapidly, huge amount of image database, audio and video database has been stored over the internet and is continuously managed by w3c. Now, one can also search the information that he / she desires not only on the basis of keywords, but also through images or audio / video files. A number of search engines are
configured in such a way that they can easily search the information on the basis of the image provided.

A. Image Retrieval Problem

Now, almost in all the aspects of our life including banking, marketing, businesses, government, education, health care, crime prevention, surveillance, engineering, architecture, journalism, fashion, graphic design and historical research, images are used for efficient services. An image database is a huge collection of images. An image database is a system where image data is integrated and stored [1]. Image data include the raw images and information extracted from images by automated or computer assisted image analysis.

B. Text-Based Image Retrieval and Content-Based Image Retrieval

In text-based retrieval, images are indexed using keywords, subject headings or classification codes, which in turn are used as retrieval keys during search and retrieval [2]. Text-based retrieval is non-standardized because different users employ different keywords for annotation. Text descriptions are sometimes subjective and incomplete because they cannot depict complicated image features very well. Examples are texture images that cannot be described by text. Textual information about images can be easily searched using existing technology, but requires humans to personally describe every image in the database. This is impractical for very large databases, or for images that are generated automatically, e.g. from surveillance cameras. It is also possible to miss images that use different synonyms in their descriptions. Systems based on categorizing images in semantic classes like "cat" as a subclass of "animal" avoid this problem, but still face the same scaling issues [6].
The Content Based Image Retrieval (CBIR) technique uses image content to search and retrieve digital images. Content-based image retrieval systems were introduced to address the problems associated with text-based image retrieval. Content based image retrieval is a set of techniques for retrieving semantically-relevant images from an image database based on automatically-derived image features [12]. The main goal of CBIR is efficiency during image indexing and retrieval, thereby reducing the need for human intervention in the indexing process. The computer must be able to retrieve images from a database without any human assumption on specific domain (such as texture vs. non-texture, or indoor vs. outdoor).

II. RELATED WORK

Content based image retrieval for general-purpose image databases is a highly challenging problem because of the large size of the database, the difficulty of understanding images, both by people and computers, the difficulty of formulating a query, and the issue of evaluating results properly. A number of general-purpose image search engines have been developed. In the commercial domain, QBIC [7] is one of the earliest systems. Recently, additional systems have been developed such as VIR [10]. In the academic domain, MIT Photobook [8] is one of the earliest systems.

Berkeley Columbia Visualseek and Webseek [9] are some of the recent well known systems. The common ground for CBIR systems is to extract a signature for every image based on its pixel values and to define a rule for comparing images. The signature serves as an image representation in the "view" of a CBIR system. The components of the signature are called features. One advantage of a signature over the original pixel values is the significant compression of image representation.

Existing general-purpose CBIR systems roughly fall into three categories depending on the approach to extract signatures: histogram, color layout and region-based search. There are also systems that combine retrieval results from individual algorithms by a weighted sum matching metric [13].

Cross Media Relevance Model[3] where the vision information of each image was denoted as blob set which is to manifest the semantic information of image. However, blob set in CMRM was erected based on discrete region clustering which produced a loss of vision features so that the annotation results were too perfect .In order to compensate for this problem , a Continuous Relevance Model (CRM) was proposed in[4].

III. PROPOSED WORK

Proposed work can be divided into different modules based on the steps of calculation from the user query to final output on the screen. In fig3.1 it can be seen that there are three different modules. First phase utilizing the features of the retrieval images, after that find the distance from one image feature to the other based on the different values of the features and final ranking of those feature combination is prepared.

Second makes comparison to including query pre-processing and fetch annotated images from the database and make initial ranking of the images based on the most similar feature value with query image. Both module use same technique, one for annotated image and other for query image.

A. First Module and second Module

First and second module consist the Image query, pre-processing and feature generation phase. Both include same technique.

*1) Input image***:** First module takes the input image from user for image

Annotation of.First module used for analyzing the related image then make an initial rank of the image based on single features. And second module utilizes training images to extract features for comparison to query image.

2) Pre-Processing: Here the entered image is need to be convert into common format for the system. This can be understood as the images are of different size and color format so conversion of this is necessary because as image is a matrix and in order to perform operation on it there dimension should be same.

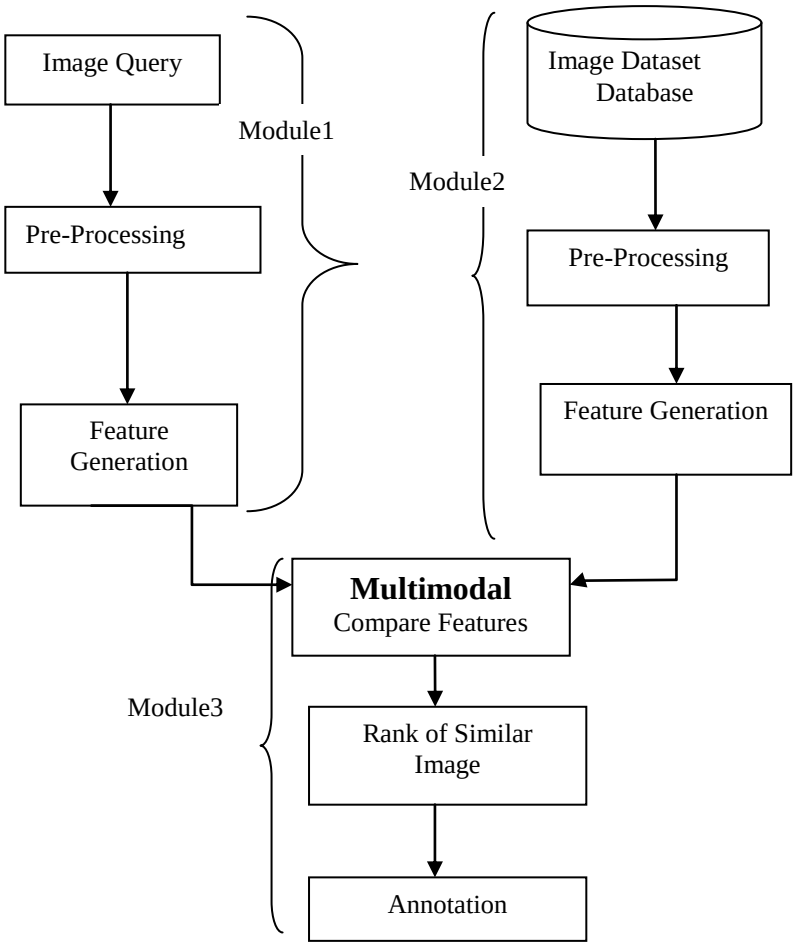

Fig. 3.1 Represent block diagram of Proposed Work for Annotation

3) *Feature Generation***:** Here image visual feature are retrieve from image where they are analyzed for finding that whether those image are close to each other or not. Different feature which are use for developing are:
There are 7 global features extracted, including

- 225-dimensional Block-wise color moments. Each image is split into 5-by-5 blocks, and 9-dimensional color moment features are extracted from each block.
- 64-dimensional HSV color histogram. A 64-block size feature vector is generated in HSV color space for each image.
- 144-dimensional Color autocorrelogram. HSV color moments are quantized into 36 bins with 4 different pixels pair distances.
- 256-dimensional RGB color histogram. A 256-dimensional histogram feature vector is extracted in RGB color space.
- 75-dimensional Edge distribution histogram. Each image is divided into 5 blocks and 15-dimensional EDH features are extracted.
- 128-dimensional Wavelet texture. 128-dimensional features are extracted using the mean and standard deviation of the energy distribution of each sub-band at different levels.
- Corner feature of the 256 dimensional image of gray format.

Here let take an image I whose above feature are maintain as I = {f1, f2, f3……fn}. Where each fn is a matrix of numeric values.

B. Third module

Third module comprises features comparison and annotation based on reranking .Ranking assigned on the basis of distances between combinations of features.
1) Multimodal Compare Feature: A modality can be viewed as a description to image such as color, edge, texture, etc. This method is usually called "multimodal fusion" or "multimodality learning". Sometimes it is also named "late fusion", whereas the approach of using concatenated high-dimensional global feature vector is named "early fusion" [14].
Here feature value from different images are compare and store in the matrix. So each modal is the representation of feature values difference. In the similar fashion other features also have its own modal matrix, so combination of all is term as multimodal. This can be understans as let image I1, I2, I3 and I4 have feature vector for above features are {f11, f12, f13, f14, f15 }, {f21, f22, f23, f24, f25}, {f31, f32, f33, f34, f35 }, {f41, f42, f43, f44, f45 }.

Loop 1:m // m represent number of modals
Loop 1:n // n represent number of images
M(m, n) = distance(Qm, fmn) // Q is for query image
EndLoop
EndLoop
Weight = {w1, w2, w3, w4, w5}

Now multiply multimodal matrix with weight matrix cell wiseW1*distance(Q1,f14). Finally sum all values row wise so that a single value is generate for each image. The obtained vector will identify the most similar image base on the distance of feature values.
*2) Annotation***:** Here once the most similar image is obtain from the dataset it is required to assign the tags from the fetched image. As the similarity of the image is found on the basis of the image visual features so the tags attach with top similar image is assign to the query image.
Algorithm for Image Annotation by Multimodal:
Input: Query, Dataset, W
OutPut: Rank

1. *Query = Pre_process(Query)*
2. *Q[1, m] = feature_generation(Query)*
3. *Loop 1:n // N is number of image in dataset*
4. *I = Dataset(n)*
5. *I = Pre_process(I)*
6. *M[n, m] = feature_generation(I)*
7. *EndLoop*
8. *Loop 1:n*
9. *Loop 1:m*
10. *M[n, m] = W[1,m]*Distance(Q[1,m], M[n, m])*
11. *EndLoop*
12. *EndLoop*
13. *R = Sum(M) // This will add values row wise*
14. *Rank = Min(R)*
15. *Q_tag = Tags(Rank) // tags present at rank position in dataset is assign*

Algorithm Pseudo code of image annotation

Output of this algorithm is a single position of the image which represents the rank of the fetched images based on the input image.

IV. IMAGE RETRIEVAL

In image retrieval, when user enter one query for image retrieval first it analyze the keywords of the related image then make a rank of the image. So for analyzing the keywords of the query it need to pre-process it. This can be understood as:

A. Text Pre-Processing

Data preprocessing reduces the size of the input text documents significantly. It involves activities like sentence boundary determination, natural language specific stop word elimination. Stop-words are functional words which occur frequently in the language of the text (for example a, the, an, of etc. in English language), so that they are not useful for classification. Here whole query is read and put all words in the vector. Now again read the file which contain stop words then remove similar words from the vector. Once the data is pre-process then it will be the collection of the words that may be in the image keyword list. For example let one query is in text vector Qd[] ={a1, k1, k3, a2, k5, a3, a4, f2……..an} and let the stop words collection is S[] = {a1,a2,a3,…….am}. Then the vector obtain after the Pre-Processing is D[] = {k1, k3, k5, k2,…….kx}.

$$D[] = Qd[] - S[]$$

For Example: Qd[] = {'Every',' morning', 'Ram',' study',' for',' two',' hour',' and',' during',' this',' time',' his',' mother',' give',' him',' one',' glass',' milk',' with','bread','jam','in ','breakfast'}

B. After pre-processing

Now D[] = {'Ram',' hour',' time',' glass',' milk','bread','jam','breakfast'}

After this pre-processing, now keywords of the image are retrieved from each image and put in a vector which is compare with the D[] vector on the basis of the most similar image keywords a list is maintain which contain the priority list.

If query vector D[] = {k1, k3, k5}. Then image keyword vector IM [] = { (K1, k2, k3), (K5, k2, k8),(K1, k4, k6, k2), (K1, k3, k5, k7),(K3, k5, k8) }. Now by comparing these vector it is found that the most similar vector is store in initial list vecorIL[].

IL[] ←similar_keywords(D,IM)

So IL[] vector have [4, 1, 5, 2, 3]. This is consider as the new order of the images as per query keywords, this is term as rank of the image. Now image obtain from the new rank can be rearrange and display for the output.

V. RESULT AND ANALYSIS

This work mostly focuses on image annotation and Retrieval from the large dataset. In order to better understand the proposed algorithm some experiments are done on it by fetching relevant images from the collection based on the example query. This section, first introduce experimental settings, Dataset description, evaluation parameter and then present the experimental results that validate the effectiveness of the approach.

A. Requirements:

In This work all algorithms and utility measures were implemented using the 2.27 GHz Intel Core i3 machine, equipped with 4 GB of RAM, and running under Windows 7 Professional. MATLAB 2012a is the simulator use for the implementation of this work.

B. Data Set

In order to conduct the experiment an artificial dataset which is a collection of images from different category are utilize. As images are of different format so first it is necessary to make it in readable format for experiment tool MATLAB. Now this collection of images of different category is shown in table I. It has been tried to collect image from different category so that maximum test is done on the proposed feature algorithms.

Table I Dataset of Different category.

Category	Examples
Objects	Ipod, map
Animal	Butterfly, Gorrila
Scene	TajMahal, Hotel Taj
Person	Barack Obama, Lena

C. Evaluation Parameter

We adopt Normalized Discounted Cumulated Gain [46] as the performance evaluation measure. As images are ranked based on the different query pass by the user for this one evaluation parameter Normalized Discounted Cumulated Gains use which can find that proposed work is effective against the previous work or not.

Here let the query text entered by the user 'INDIA TAJ' then as per the pass query text images will be pop up, now let for top five images if Normalized Discounted Cumulated Gain is
Given below for this result. Then first it need that from the top five images how many images are relevant then other are consider as the irrelevant images

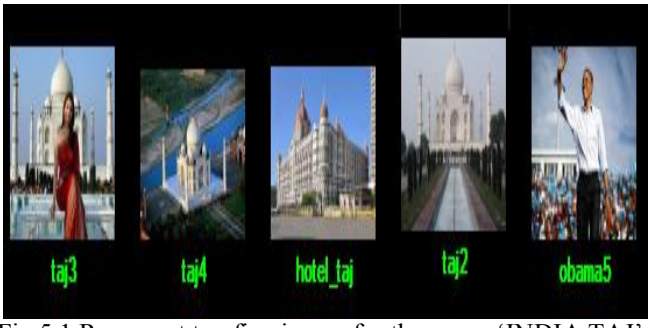

Fig.5.1 Represent top five image for the query 'INDIA TAJ'.

Consider a vector L as the list of image represent the relevance by 1 and irrelevant by 0 so if the first image is relevant then first element in the vector is 1, if the second image is relevant then second element in the vector is 1, if the third image is irrelevant then third element in the vector is 0.
So for above query let L = [1 1 0 1 0] Then put this value in the Normalized Discounted Cumulated Gain formula where P = 5. Zp is the total sum if all the values in the L vector i represent the position in the result such that i = {1,2,3,4,5}.
The NDCG measure is computed as

$$NDCG@P = Zp \sum_{i=1}^{p} \frac{2^{l(i)} - 1}{\log(i+1)}$$

By entering the query and searching the desired image it was obtained that they can be categorize into few levels such as relevant or not. It can be further categorized into most relevant, relevant, less relevant, and irrelevant.
Result of NDCG

Table II Values of NDCG@10 and NDCG@5 by Different query.

NDCG Results		
Query	**NDCG@10**	**NDCG@5**
lion animal	0.53	0.461
sea water	0.87	1
hill water	0.795	0.764
young president	1	1
butterfly insect	0.8007	0.9202

From the above table II it is found that the including of the modal annotation has increased the efficiency of image re-ranking. In different categories of the images, one can find the improved results. As the annotation can improve identification of the image more accurately so a perfect combination of the features is obtained by combining each of them at different category of the image present in the dataset.

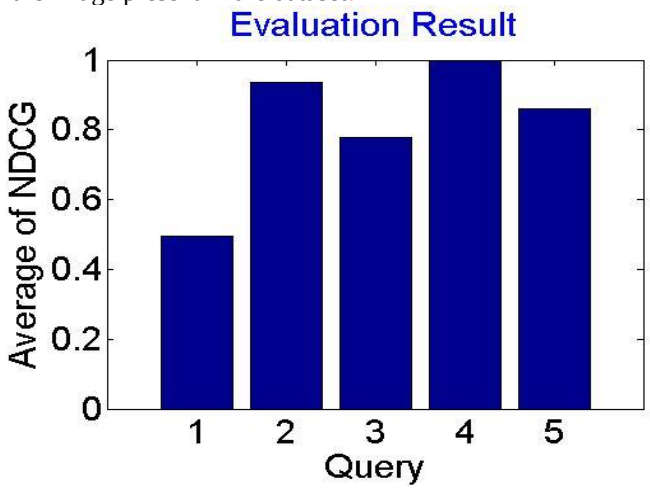

Fig5.2 Represent the bar graph of the Average NDCG values with different query

It is clear from the above fig5.2 that the values at different NDCG levels increase with sharp annotation and query. So the utilization of the annotation has increase the efficiency of the work then the visual features. Above results justify as the dataset contain image of different environment and location that can test the algorithm combination very well.

Now as the new query image is passed into the system for the annotation then top ranked images which is based on the feature vector of the image is selected and annotation present in that image is assigned to the query image. So for the testing the dataset contain set of images that are not present in the training set and that image of testing dataset should not contain any kind of tags or annotation. As in previous work such as cross media relevance modal [3] image is passed with tag which contains some keywords.

Table III Annotation Results obtain for different category of input images.

Category	Precision	Recall	F-measure
Objects	0.75	0.5	0.599
Animal	0.857	0.75	0.806
Scene	1	0.428	0.599
Person	0.75	0.33	0.459

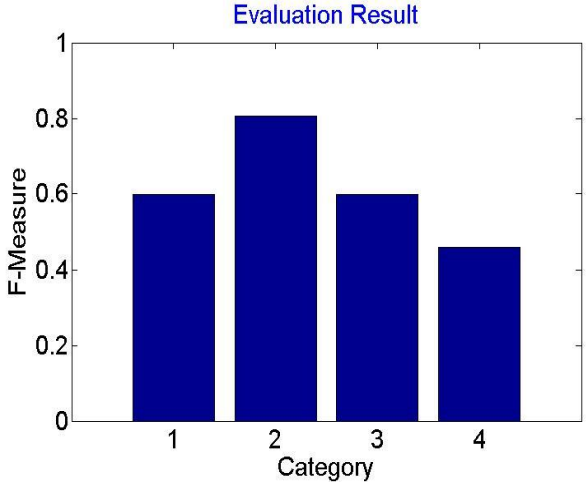

Fig 5.3 Graph of F-measure values for different category of images.

From above fig5.3 it is shown that f-measure value is always above the 0.5 which is quite impressive as image is inserted in the system without any tag, so searching is done only by the visual feature. Combination of color and texture features isused in the working outcomes with effective values in annotation.

VI. CONCLUSION

World Wide Web has necessitated the users to make use of automated tools to locate desired information resources and to follow. Web image re-ranking has been widely used to reduce the user searching time on the internet; its success mainly depend on the accuracy of image features similarities. The integration of image-based features has recently attracted a lot of attention. In this paper, we have given an overview of recent works in this field, then compared them and discussed some of their limitations. The main objective of this study is to tackle this problem in an adaptable and effective way. This work presents utilization of visual features for ranking the image as both make the re-ranking process more powerful, which is shown in results.

REFERENCES

[1] Meng Wang, Hao Li, Dacheng Tao, Ke Lu, and Xindong Wu "Multimodal Graph-Based Reranking for Web Image Search. IEEE Transaction on image processing Vol. 21, NO. 11, November 2012.

[2] KirtiYadav, SudhirSinghImproving Web Image Search Re-Ranking Using Hybrid ApproachIJARCSSE. Volume 4, Issue 6,June 2014.

[3] Jeon J, Lavrenko V, Manmatha R.," Automatic image annotation and retrieval using cross-media relevance models". Proc. of Int. ACM SIGIR Conf. On Research and Development in Information Retrieval, Toronto, Canada, 119-126,Jul. 2003.

[4] Lavrenko V, Manmatha R, Jeon J." A model for learning the semantics of pictures". Proceedings of the Seventeenth Annual Conference on Neural Information Processing Systems, 2003.

[5] I. Zukerman, D. W. Albrecht & A. E. Nicholson. Predicting user's requests on the WWW. Proc. of the seventh international conference on User modeling, pages 275{284, 1999.

[6] J. Dom_enech, J. Sahuquillo, J. A. Gil & A. Pont. The Impact of the Web Prefetching Architecture on the Limits of Reducing User's Perceived Latency. Proc. of the International Conference on Web Intelligence, 2006.

[7] M. Flickner, H. Sawhney, W. Niblack, J. Ashley, Q. Huang, B. Dom, M. Gorkani, J. Hafner, D. Lee, D. Petkovic, and P. Yanker, "Query by image and video content: The QBIC system," IEEE Computer, vol. 28, no 9, pp.23-32, Sep. 1995.

[8] A. Pentland, R. Picard, and S. Sclaroff , "Photobook: Content based manipulation of image databases," International Journal of Computer Vision, vol.18, no 3, pp.233–254, June 1997.

[9] J. Smith and S. Chang, "Visualseek: A Fully Automated Content-Based Image Query System," Proceedings of the 4th ACM international conference on Multimedia table of contents, Boston, Massachusetts, United States, Nov. 1996, pp. 87-98.

[10] A. Gupta, and R. Jain, "Visual information retrieval," Comm. Assoc. Comp. Mach., vol. 40, no. 5, pp. 70–79, May. 1997.

[11] J. Li, J. Wang, and G. Wiederhold, "Integrated Region Matching for Image Retrieval," In Proceedings of the 2000 ACM Multimedia Conference, Los Angeles, October 2000, pp. 147-156.

[12] R. Murumkar, Mr. C.M. Jadhav, Ms. Swati." An Effective Image Search Reranking Based On Prototype" IJESRT, 3(6): June, 2014.

[13] F. Long, H. Zhang, H. Dagan, and D. Feng, "Fundamentals of content based image retrieval," in D. Feng, W. Siu, H. Zhang (Eds.): "Multimedia Information Retrieval and Management. Technological Fundamentals and Applications," Multimedia Signal Processing Book, Chapter 1, Springer-Verlag, Berlin Heidelberg New York, 2003, pp.1-26.

[14] C. G. Snoek, M. Worring, and A. W. Smeulders, "Early versus late fusion in semantic video analysis," in Proc. ACM Multimedia, 2005, pp. 399–402.

