Tag Based Image Retrieval Using Optimized Tag Completion Matrix

Dinla O.K, Anu Joseph

Abstract—**An image retrieval system is a computer system for browsing, searching and retrieving images from a large database of digital images. Most common methods of image retrieval utilize some method of adding metadata such as keywords, or descriptions to the images so that retrieval can be performed over the annotation words. Most of social image search engines are based on keyword/tag matching. This type of image retrieval is tag-based image retrieval. TBIR is not only efficient but also effective. The performance of TBIR is highly dependent on the availability and quality of manual tags. Manual image annotation is time-consuming, laborious, expensive and manual tags are often unreliable and inconsistent; to address this challenge, there has been a large amount of research done on automatic image annotation. In addition, since users tend to choose general and ambiguous tags in order to minimize their efforts in choosing appropriate words, tags that are specific to the visual content of images tend to be missing or noisy, leading to a limited performance of TBIR. To address this challenge, the problem of tag completion is used, where the goal is to automatically fill in the missing tags as well as correct noisy tags for given images. Image-tag relation is represented by a tag matrix, and search for the optimal tag matrix consistent with both the observed tags and the visual similarity. Haar Wavelet transform and k-means clustering is used in the proposed system. The proposed method is significantly more effective than the state-of-the-art algorithms and it is computationally efficient and scales well to large databases.**

Index Terms—**Automatic Image Annotation, haar wavelet transformation, TBIR, K Means clustering.**

I. Introduction

With the increasing growth of computer technology, rapidly declining cost of storage and ever-increasing access to the Internet, digital acquisition of information has become increasingly popular in recent years. Image retrieval is the process of searching and retrieving images from a huge dataset. As the images grow complex and diverse, retrieval of the right images becomes a difficult challenge. Due to the explosive growth of digital technologies, ever increasing visual data are created and stored. Nowadays, visual data are as common as textual data. There is an urgent need of effective and efficient tool to find visual information on demand. A large amount of research has been carried out on image retrieval (IR).

Image retrieval is an important topic in the field of pattern recognition and artificial intelligence. In general, IR research efforts can be divided into two types of approaches. The first
type of approach focuses on content based image retrieval (CBIR), where images are automatically indexed and retrieved with low level content features like colour, shape and texture. Content-based image retrieval is opposed to concept-based approaches. The term "content" in this context might refer to colours, shapes, textures, or any other information that can be derived from the image itself. CBIR is desirable because most web-based image search engines rely purely on metadata and this produces a lot of garbage in the results. However, recent research has shown that there is a significant gap between the low level content features and semantic concepts used by humans to interpret images. In addition, it is impractical for general users to use a CBIR system because users are required to provide query images.

The second approach of image retrieval is keyword/tag-based image retrieval. To overcome the limitations of CBIR, TBIR represents the visual content of images by manually assigned keywords / tags. It allows a user to present his/her information need as a textual query and find the relevant images based on the match between the textual query and the manual annotations of images. Compare to CBIR, TBIR is usually more accurate in identifying relevant images by alleviating the challenge arising from the semantic gap. TBIR is also more efficient in retrieving relevant images than CBIR because it can be formulated as a document retrieval problem and therefore can be efficiently implemented using the inverted index technique. However, the performance of TBIR is highly dependent on the availability and quality of manual tags. In most cases, the tags are provided by the users who upload their images to the social media sites, and are therefore often inconsistent and unreliable in describing the visual content of images. In order to minimize the effort in selecting appropriate words for given images, many users tend to describe the visual content of images by general, ambiguous, and sometimes inappropriate tags, as explained by the principle of least effort. As a result, the manually annotated tags tend to be noisy and incomplete, leading to a limited performance of TBIR.

Manually tagging images is both subjective and time consuming. The annotator cannot always remember the tags that he annotated earlier to the similar images and sometimes he may also assign incorrect tags. These incorrect tags will affect the efficiency of image retrieval. So it would be nice if the tagging process could be done automatically. A

Manuscript received April , 2014.

***Dinla O.K**, Department of Computer Science & Engineering, Calicut University, Kozhikode, India, 97744093662*

***Anu Joseph**, Department of Computer Science & Engineering, KMCT College of Engineering, Kozhikode, India, 8281666512.*

ISSN: 2278 – 7798

requirement for effective searching and retrieval of images in rapid growing image databases is that each image has accurate and useful annotation. The tags which describe the content of images can help users easily manage and access large-scale image datasets. With these metadata, the manipulations of image data can be easier to be accomplished, such as browsing, indexing and retrieval. Automatic social image tagging plays a critical role in modern tag-based image retrieval systems. Existing tagging methods usually contain dirty and incomplete tags, which severely limit the performance of these tagging methods. The limitation of current automatic image annotation approaches motivates us to develop a new computational framework for tag completion.

In this paper, it is hypothesized that the performance of image retrieval can be improved by automatic image annotation. In this work automatic image annotation can be done for both missing tags as well as noisy tags. The goal is to automatically fill in the missing tags as well as correct noisy tags for given images. Image-tag relation is represented by a tag matrix, and search for the optimal tag matrix consistent with both the observed tags and the visual similarity. In this work a novel algorithm is proposed for solving this optimization problem. Extensive empirical studies show that the proposed algorithm is significantly more effective. The proposed algorithm is computationally efficient and scales well to large databases. Main aim of this study is to identify and assign keyword or tag to the image based on its visual content that is effective for fast image retrieval. In this work, challenge is addressed by automatically filling in the missing tags and correcting the noisy ones. This problem is referred as the tag completion problem. One way to complete the missing tags is to directly apply automatic image annotation techniques to predict additional keywords/tags based on the visual content of images.

The rest of this paper is organized as follows: In Section 2, gives an overview of automatic image annotation. Section 3 provides a detailed description for the proposed framework and algorithm. Section 4 summarizes the experimental results on automatic image annotation and tag-based search. Section 5 concludes this study with suggestions for future work.

II. AUTOMATIC IMAGE ANNOTATION

Due to several challenges faced by the Content Based Image retrieval System as mentioned in the previous section, an alternative image retrieval technique has been recognized and
used widely in recent times. Previously image annotation for the purpose of indexing and later retrieval of the requisite image or images from the image collections was done manually. However this method is considerably laborious and
expensive, hence an automatic image annotation technique is used to annotate images based on its content. Automatic Image Tagging also known as Automatic Image Annotation is an annotation technique which assigns tags or captions to an image relevant to an image to increase the efficiency of image retrieval and classification. The idea with automatic image tagging is that tags are automatically captioned and assigned to the digital image. These tags should describe every important part or aspect of the image and its context. Automatic image tagging can be done based on the visual content of the image, contextual information, or using a mixture of these two approaches. Thus the main objective of Automatic Image Annotation technique is to bridge the semantic gap between the image retrieval and the image semantics understandable by humans.

There are three types of image annotation: manual, automatic and semi automatic. Manual annotation needs users to enter some descriptive keywords when perform image browsing. Automatic annotation detects and labels semantic content of images with a set of keywords automatically. In case of Semiautomatic annotation, it needs user's interaction to provide an initial query and feedback for image annotation while browsing.

A. *Purpose of Automatic Image Annotation*

The search and the retrieval operation of requisite image by a search engine like goggle is usually performed based on text without much emphasis on the content of the image. Therefore assigning suitable keywords to images is a vital task. This is achieved by automatic image annotation, which assigns specific words or keywords to an image relevant to the content of the image thereby bridging the semantic gap and achieving multi-level image annotation automatically. Besides, automatic image annotation technique is used as it is less tedious and minimizes time requirement for annotation of
images compared to manual annotations.
This can be achieved in two steps.

1. Find an accurate keyword relevant to the query image.
2. Given one keyword, annotate the image to describe the details of this image.

Flickr is one of the most comprehensive image resources on the web. As a photo management and sharing application, it provides user with the ability to tag, organize and share their photos online. Flickr offers a public API which can be used to write applications that use Flickr in some way.

The objective of Automatic Image Annotation is to search over user contributed photo sites like Flickr which have accumulated rich human knowledge and billions of photos, then associate surrounding tags from those visually similar Flickr photos for the unlabeled image. For an unlabelled image, photos in the social media are extracted by the automatic image annotation technique; the annotations associated with the images are expanded, and then ranks are associated with the images, based on their relevance to the given image in terms of visual consistency and semantic consistency. Then amongst the expanded annotations, candidate annotations are retrieved and assigned to the requisite image.

B. *Advantages of Automatic Image Annotation*

1. High performance and less time-consuming.
2. It can perform effective manipulation for exponentially growing photo collection.
3. It is more flexible than the manual photo tagging.
4. Robust.

III. PROPOSED SYSTEM

Proposed System integrates both TBIR and CBIR. First step in the proposed system is to generate a partially observed tag matrix. The goal is to complete the tag matrix that is to retrieve an optimised tag matrix. So that this optimised tag matrix can be used to improve the performance of TBIR. Besides the tag information, the visual similarity between images is computed based on the extracted visual features and search for the optimal tag matrix that is consistent with both the observed tag matrix and the pair wise visual similarity between images. In case of incomplete or noisy tags, CBIR is used to automatically filling missing tags that is to complete the tag matrix. After completing the tag matrix images are retrieved using TBIR. In the proposed system haar wavelet transformation and k means clustering is used. The proposed
method aims to improve the performance of CBIR using wavelet decomposition by haar wavelet. After that features are extracted using f-norm theory. The use of progressive retrieval strategy is to provide balance between computational complexity and retrieval accuracy.

A. *Framework for tag completion*

Fig 1 illustrates the framework for tag completion. This is a framework for tag matrix completion and its application to image search.

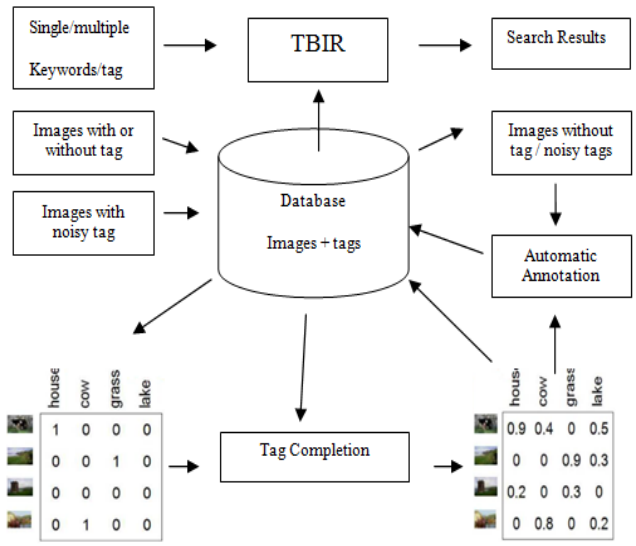

Fig.1 Framework for tag completion

Framework of tag completion can be described as follows. Images assigned with tags are uploaded to the database and they are trained based on these manually assigned tags. If images are uploaded to the database with missing tags and noisy tags CBIR is used to extract the visual content. These visual contents and manually assigned tags are used for automatic image annotation. When images are uploaded, image tag relation is represented as a partially observed tag matrix. Visual similarities of the images are used to complete the tag matrix. Searching can be done efficiently based on TBIR.

The proposed system is divided into two parts: learning and querying. Learning is the process of training the uploaded images in the database based on the tag associated with the images. In case of un-annotated images and for the images having incorrect tag, CBIR or the visual features of the trained images are used for annotation. Main steps in learning are:

1. Partial matrix creation
2. Image decomposition
3. Feature Extraction
4. K Means clustering
5. Regenerating Tag Completion Matrix
6. Automatic Annotation

First step in learning process is to create a tag matrix M'^{nxm} where n represent the number of images and m *represents* the number of tags. M'_{ij} is set to 1 if the image i is associated with tag j, otherwise it is set to 0, which shows the relationship between images and tags. Next step is to read rgb values of the image and create red, green and blue matrices. After creating red, green and blue matrices image is decomposed using haar wavelet transformation [3]. Feature is extracted by using F-Norm theory [4]. K-mean clustering is used to form the cluster of images and similarity matching is done using f-norm theory. Output of the learning process will be an optimised tag matrix M^{nxm} which consist of real values which represent the similarity of images. This optimised tag matrix is called completed tag matrix and this can be used for image annotation. When an image is uploaded without tag or with incorrect tag, that image is taken as a query image and search for the image that has highest similarity with the query image. After training process query image is assigned with the tag of the image having highest similarity.

B. *Algorithm*

1. Creation of partially observed tag matrix
2. Indexing
   a) Read all images from database
   b) Red, green, blue matrix generation
   c) Matrix decomposition using Haar Wavelet transformation
   d) Create feature vector using F-Norm theory
   e) Create document using feature vector
   f) K means clustering
   g) Add document to cluster
3. Rank Regeneration
4. Tag Completion

IV. EXPERIMENTAL RESULT

Fig2. Illustrates single-tag-based image search. The word on the right is the query and images on its left are the search results. The images displayed in the two rows are the results returned by the proposed System, the existing TMC method [1], respectively. The first row is the results for the proposed method and second row shows the results for the TMC method.

ISSN: 2278 – 7798

Fig. 1 Retrieved images based on proposed system and existing system

The performance of image retrieval is basically measured in terms of its precision and recall. Precision measures the retrieval accuracy; it is the ratio between the number of relevant images retrieved and the total number of images retrieved. Recall measures the ability to retrieve all relevant images in the database. It is the ratio between the number of relevant images retrieved and all of the relevant images in the database. They are defined as follows.

$$\text{Precision} = \frac{\text{Number of relevant images retrieved}}{\text{Total number of images retrieved}}$$

$$\text{Recall} = \frac{\text{Number of relevant images retrieved}}{\text{Total number of relevant images}}$$

In the proposed system Mean Average Precision@*K* is used for evaluation.

Mean Average Precision (map)

For a given query, Average Precision (ap) is the average of the precision values obtained when each relevant image is retrieved for that query; if a relevant image is not retrieved at all, its corresponding precision is 0. Hence ap emphasizes the relevant images to be ranked higher and is capped by the recall. map is the mean of aps for a set of queries. Mean Average Precision@*K* is the mean of aps for a set of queries among the top-*K* retrieved images for a given query. Based on the experimental result we can understand that the MAP@20 for the proposed method is 90. So from this result we can say that proposed system performs efficiently than TMC.

V. Conclusion

In this paper, we present a new image search engine framework which consists of Automatic Image Annotation (AIA), Tag Based Image Retrieval (TBIR), Content Based Image Retrieval (CBIR), and Tag matrix techniques. The AIA plays an important role in the proposed framework since the annotation of the tags is done using Tag Completion Algorithm and the tags are annotated automatically. Thus it minimizes the work of the user to add the tags manually by himself. TBIR thus will give an increase in efficiency in retrieving the image from the database than the former system in which the tags are annotated manually. We proposed a tag matrix completion method for image tagging and image retrieval by using haar wavelet transformation and k means clustering. Image tag relation is represented as a tag matrix, and our aim is to optimize the tag matrix by minimizing the difference between tag-based similarity and visual content-based similarity. The proposed method falls into the category of semi-supervised learning in that both tagged images and untagged images are exploited to find the optimal tag matrix. We evaluate the proposed method for tag completion by performing two sets of experiments, i.e., automatic image annotation and tag-based image retrieval. This automatic image annotation and image retrieval method is suitable for missing tags as well as noisy tags.

REFERENCES

[1] Lei Wu, Rong Jin and Anil K.Jain, "Tag Completion for Image Retrieval" IEEE Transaction on Pattern Analysis and Machine Intelligence VOL. 35, NO. 3, MARCH 2013.

[2] M.S. Lew, "Content Based Multimedia Information : State of the Art and Challenges," ACM Trans. Multimedia Computing, Comm and applications, vol.2,pp. 1-19, 2006.

[3] Pasnur M.A., P. S. Malge "Image Retrieval Using Modified Haar Wavelet Transform and K Means Clustering", *IJETAE,* Vol 3, Issue 3, March 2013

[4] M. Wegmuller, J. P. von der Weid, P. Oberson, and N. Gisin, "An Approach of Image Retrieval Using Content Based Retrieval System," in *International Journal of Advanced Research in Computer Science and Software Engineering*. Volume 3, Issue 2, February 2013.

[5] E. Akbas and F.T.Y. Vural, "Automatic Image Annotation by Ensemble of Visual Descriptors," Proc. IEEE Conf. Computer Vision and Pattern Recognition, pp. 1-8, 2007.

[6] R. Jin, J.Y. Chai, and L. Si, "Effective Automatic Image Annotation via a Coherent Language Model and Active Learning," Proc. 12th Ann. ACM Int'l Conf. Multimedia, pp. 892-899, 2004.

[7] Y.-G. Jiang, C.-W. Ngo, and J. Yang, "Towards Optimal Bag-of-Features for Object Categorization and Semantic Video Retrieval,"Proc. Sixth ACM Int'l Conf. Image and Video Retrieval, pp. 494-501,2007.

[8] L. Wu and N. Yu, "Visual Language Modeling for Image Classification," Proc. Int'l Workshop Multimedia Information Retrieval, pp. 115-124, 2007.

[9] B. Russell, A. Torralba, K. Murphy, and W. Freeman, "Labelme: A Database and Web-Based Tool for Image Annotation," Int'l J. Computer Vision, vol. 77, pp. 157-173, 2008.

[10] Mary Czerwinski2, LiuWenyin1, Hong, Jiang Zhang1, "Semi-Automatic Image Annotation," U.S. Patent 5 668 842, Sept. 16, 2000.

[11] Y. Liu, R. Jin, and L. Yang, "Semi-Supervised Multi-Label Learning by Constrained Non-Negative Matrix Factorization,"Proc. 21st Nat'l Conf. Artificial Intelligence, pp. 421-426, 2006.

Dinla O.K , B.Tech. Degree in Computer Science and Engineering from SNGCET, Kerala in 2010 and pursuing M.Tech in Computer Science and Engineering from KMCT College of Engineering, research area is artificial intelligence, Recent publication is Tag Based Image Retrieval using Haar Wavelet Transform and K Means Clustering.

Anu Joseph MTech. Assistant Professor,*Department of Computer Science & Engineering, KMCT College of Engineering*