A Survey of Tag Completion for Efficient Image Retrieval Based on TBIR

Mr. Wakchaure Sujit R., Prof. Shamkuwar Devendra O.

Abstract— **Tag-based image retrieval often used to increase performance to retrieving images with the help of search engines. Image retrieval based on user-provided image tags on the photo sharing websites. A requirement for effective searching and retrieval of images in rapid growing online image databases is that each image has accurate and useful annotation. For the superior image search and retrieval experiences availability the quality of manual tags must be better, but previous study shows that the manual tagging to the images are inconsistent and inefficient to finding appropriate image. Missing and ambiguous tags degrades the performances of TBIR, so in order to that problem, we study completion of tags to overcome that problem. The main goal is to automatically fill in the missing tags as well as correct noisy tags for given images. With the help of tag matrix we find the optimal solution for tag matrix, and consistency with tags of that image and its visual content similarity. This system efficient and helpful when tagging images and it is a dynamic approach for optimization procedure.**

Index Terms—**TBIR, Tag completion, image annotation, image retrieval.**

I. INTRODUCTION

Social tagging became popular with the launch of sites like Delicious and Flickr. Since then, different social systems have been built that support tagging of a variety of resources. Given a particular web object or resource, tagging is a process where a user assigns a tag to an object. On Delicious, a user can assign tags to a particular ookmarked URL. On Flickr, users can tag photos uploaded by them or by others. Whereas Delicious allows each user to have her personal set of tags per URL, Flickr has a single set of tags for any photo. On blogging sites like Blogger, Wordpress, Livejournal, blog authors can add tags to their posts. On micro-blogging sites like Twitter, hash tags are used within the tweet text itself. On social networking sites like Facebook, Orkut, etc., users often annotate parts of the photos. Users can also provide tagging information in other forms like marking something as “Like” on Facebook. Upcoming event sites can allow users to comment on and tag events. Recently, tripletags (tags of the format namespace: key=value (e.g., geo:lat=53.1234) are becoming popular. Such a syntax can improve the usability of tags to a large extent. Using rel-tags1, a page can indicate that the destination of that hyperlink is an author-designated tag for the current page. Rel-tags have been used by various implementation sites to tag blogs, music, links, news articles, events, listings, etc. Citation websites have tags attached to publication entries. Cataloging sites like LibraryThing and Shelfari allow users to tag books. Social news sites like Digg, SlashDot allow users to attach tags to news stories. Yelp, CitySearch and other such business/product reviews sites allow users to attach their reviews and other users to select tags to rate reviews too. Multimedia objects like podcasts, live casts, videos and music can also be tagged on sites like Youtube, imeem, Metacafe, etc. On Yahoo! Answers, you can tag an answer as positive or negative depending on how helpful it was. Tags are often used to collect such binary or multi-valued ratings or categorical decisions from users. Tags are omni-present on the web. But what led to the emergence of tagging based systems? As we shall see in this section, tags are a better way of generating metadata and prevent problems associated with fixed taxonomies in social systems.

Image retrieval has been widely studied from two paradigms: *content-based* image retrieval (CBIR) [10], [41]and keyword/tag-based image retrieval [12] (TBIR). The former requires users to formulate a query using an example image. The retrieval system then returns the set of images that best matches the given example based on visual content, i.e., low level features such as color and texture. TBIR or *Annotation-based* image retrieval, on the other hand, enables users to formulate naturally semantic queries using textual keywords. In order to support this retrieval paradigm, many *automatic image annotation* techniques have been proposed, which assign a few relevant keywords to an unannotated image to describe its visual content for image indexing and retrieval.

CBIR finds visual similarity between query image and images of database. CBIR systems not so efficient due to the semantic gap between the low-level visual features used to represent images and the high-level semantic meaning behind images. With the rapid advance in the technology of digital imaging, there is an explosive growth in the amount of available image data in our daily lives. This trend urgently necessitates the development of effective retrieval technology for large volume of images. The prevalence of digital photography devices (e.g., digital cameras, mobile phones) has led to over 200 billion images accessible online and the number is continuously growing (Yahoo!, 2010). Owing to increasing popularity of tagging activities in social media sharing platforms (e.g., Flickr), many of these services enable users to annotate images with tags. The availability of such tags as metadata has given rise to opportunities to build novel and superior tag-based techniques to enhance significantly our ability to understand social images and to retrieve them effectively and efficiently.

Recent study shows tags given by users to the images are inconsistent and unreliable in describing the visual content of images in Flickr data [13]. With that study, the manually annotated tags tend to be noisy and incomplete, results poor performance of TBIR. This was observed in [14], where, on average, less than 10 percent of query words were used as image tags, implying that many useful tags were missing in the database. In this work, we cover this challenge by automatically filling in the missing tags and correcting the noisy tags referred as tag completion problem.

Study shows TBIR uses the visual content of images by manually assigned keywords/tags to cover the problem of CBIR It allows a user to present his 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. TBIR is usually more accurate in identifying relevant images [24] by overcoming the problem arising from the semantic gap. TBIR also useful in document retrieval problem and therefore can be efficiently implemented using the inverted index technique. On the dependency on availability and quality of manual tags efficiency of TBIR is dependent.

With the help of automatic image annotation techniques [17], [20] with the use of visual content we complete this problem. Unfortunately, the annotated tags for most Web images are incomplete and noisy, so it is difficult to directly apply automatic image annotation technique.

We use matrix completion, in that we represent the relation between tags and images by a tag matrix, row represents image and each column represents to a tag. Each entry in the tag matrix is a real number that represents the relevance of a tag to an image. Similarly, we represent the partially and noisy tagged images by an observed tag matrix, where an entry (i,j) is marked as 1 if and only if image i is annotated by keyword/ tag j. On other hand the tag information, we also compute the visual similarity between images based on the extracted visual features. We search optimal tag matrix that is consistent with the observed tag matrix and the visual similarity between images. Our proposed algorithm is effective in comparison to the state-of-the-art algorithms for automatic image annotation.

The rest of the paper is organized as follows. Section 2 reviews the related work on automatic image annotation. In Section 3, we describes problem of tag completion, algorithm. Section 4 describes automatic image annotation and tag-based search. 5. Finally, we conclude the paper with future work discussion.

II. REALATED WORK

In recent work shows, the major difference between our work and aforementioned efforts is that a textual document contains much redundancy of words to conveys its semantic whereas images are usually associated with only few tags. Furthermore, redundancy of tags is minimal in many social image tagging systems. Particularly, in Flickr, a tag cannot be assigned more than once to the same image. Moreover, the tags are assigned by different users with different motivations and different criteria for determining the degree of relatedness of a tag to an image. All these differences demand systematic investigation of the impact of different formulations on image search ranking.

More recent work shows improvement of automatic image annotation in considering visual features. Other than guessing annotated tags for the image, many algorithms [11] have been developed to predict annotations for individual areas within an image. So from this, performance of automatic image annotation is far from being satisfactory. Many algorithms have been proposed for automatic image annotation [18] . They can roughly be grouped into two major categories, depending on the type of image representations used. The first group of approaches is based upon global image features [23], such as color moment, texture histogram, etc. The second group of approaches adopts the local visual features.

Several recent works on image annotation are based on distance metric learning. Monay and Gatica-Perez [19] proposed annotating the image in a latent semantic space. Zhuang and Hoi proposed a two-view learning algorithm for tag reranking. Li et al. proposed a neighbour voting method for social tagging. Similarly to the classification-based approaches, these methods require clean and complete image tags, making them unsuitable for the tag completion problem.

A. Linguistic Classification of Tags

Based on linguistics, tags can be classified as follows

- Functional: Tags that describe the function of an object. (e.g., weapon)
- Functional collocation: These are defined by function but in addition, they have to be collected in a place (and/or time). (e.g., furniture, tableware)
- Origin collocation: Tags that describe why things are together? (e.g., garbage, contents, dishes (as in "dirty dishes" after a meal)).
- Function and origin: Tags that decribe why an object is present, what is the purpose, or where did it come from. (e.g., "Michelangelo" and "medieval" on an image of a painting by Michelangelo)

III. TAG COMPLETION

We first present a framework for tag completion, and then describe an efficient algorithm for solving the optimization problem related to the proposed framework.

A. A Framework for Tag Completion

Fig. 1 describes the tag completion task. Given a binary image-tag matrix, our aim is to complete the tag matrix automatically with real numbers that shows the probability of assigning the tags to the images. Given the completed tag matrix, we can run TBIR to efficiently and accurately detect the relevant images for textual query.

B. Optimization

Prior to finding optimization problem, we develop a subgradient descent-based approach (Algorithm 1). As comparing to the other optimization approaches such as Newton's method and interior point methods, the subgradient descent approach is advantageous in that its computational complexity per iteration is significantly lower, making it suitable for large image datasets.

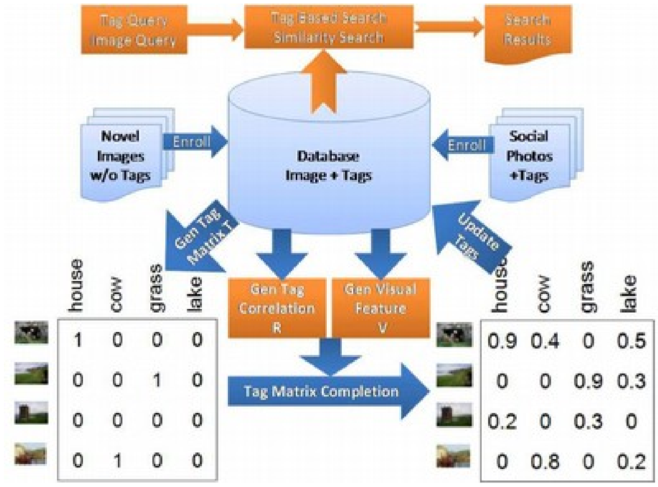

Fig. 1.Tag matrix completion for image search. Assigned tags for the given database, our algorithm creates a tag matrix states relation between images an d initially assigned tags. It completes tag matrix automatically with the help of updating significant score of tags to all images. The Completed tag matrix will be used for TBIR.

Algorithm 1. Tag Completion Algorithm (TMC)

1: INPUT:

- Observed tag matrix: $\hat{T}$ ε $IR^{n \times m}$
- Parameters: γ, η, λ, and μ
- Convergence threshold: ε

2: OUTPUT: the complete tag matrix T

3: Compute the tag correlation matrix R= $\hat{T}^{T}\hat{T}$

4: Initialize wl = 1d, T1 = $\hat{T}$, and t = 0

5: **repeat**

6: Set t = t + 1 and stepsize ηt = 1/t

7: Compute $\hat{T}_{t+1}$ and $\hat{w}_{t+1}$ according to (8)

8: Update the solutions T_{t+1} and $\hat{w}_{t+1}$ according to (9) and (10)

9: **until** convergence: $\|\mathcal{L}(T_t, w_t) - \mathcal{L}(T_{t+1}, w_{t+1})\| \leq \varepsilon \| \mathcal{L}(T_t, w_t)\|$

C. Discussion

The objective of this work is to complete the tag matrix for all the images, it belongs to the category of transductive learning. In order to turn a transductive learning method into an inductive one, one common approach is to retrain a prediction model based on outputs from the transduction method [2]. A similar approach can be used for the proposed approach to make predictions for out-of-samples.

IV. Experiments

Automatic image annotation and tag-based image retrieval are the two tasks for tag matrix. Three benchmark datasets are used in the proposed study.

- Corel dataset [7]. It consists of 4,993 images, each image has 5 tags and 260 unique keywords used in this dataset.
- Labelme photo collection. It consists of 2,900 online photos, annotated by 495 nonabstract noun tags. The maximum number of annotated tags per image is 48.

Fig. 2. Illustration of the single-tag-based image search. The word on the left is the query and images on its right are the search results. The images displayed in the three rows are the results returned by the proposed TMC method, the TagProp method, and the TagRel method, respectively. The blue outlines are the results for the proposed methods, the white lines are the results for the baseline methods.

- Flickr photo collection. It consists of one million images that are annotated by more than 10,000 tags.

The maximum number of annotated tags per image is 76. Since most of the tags are only used by a small number of images, we reduce the vocabulary to the first 1,000 most popular tags used in this dataset, which reduces the database to 897,500 images.

A. Automatic Image Annotation

In this context, this work randomly separate each dataset into two group collections. One group consisting of 80 percent of images is used as training data, and the other collection consisting of 20 percent of images is used as testing data. We repeat this 20 times. Each run creates a new separation of the

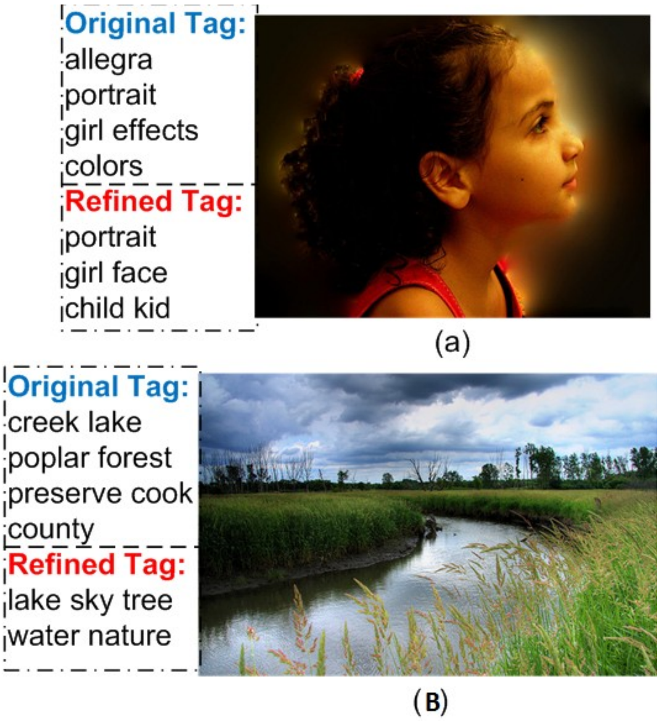

Fig. 3. Illustration of Exemplar tag refinement results from our search approach

collections. We report the result based on the average over the 20 trials. To run the proposed algorithm for automatic image annotation, we simply view test images as special cases of partially tagged images, i.e., no tag is observed for test images. We have to apply algorithm for completion matrix for training and test images. We then rank the tags for test images in the descending orderbased on their relevance scores in the completed tag matrix, and return the top ranked tags as the annotations for the test images.

B. Tag-Based Image Retrieval

Unlike the experiments for image annotation where each dataset is divided into a training set and a testing set, for the experiment of tag-based image retrieval, we will include all the images from the dataset except the queries as the gallery images for retrieval. Similarly to the previous experiments, we vary the number of observed tags from 1 to 4. Similarly to the previous experiments, we only compare the proposed algorithm to TagProp and TagRel because the other approaches were unable to handle the partially tagged images. Below, we first present the results for queries with single tag, and then the results for queries consisting of multiple tags.

V. Conclusion

This survey clarifies a tag matrix completion method for image tagging and image retrieval. We will optimize the relation of 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 will create two sets for tag completion in this work, i.e., automatic image annotation and tag-based image retrieval. Unreliable and inconsistent manual tags are efficiently overcome with the help of these techniques. Operational and experimental results on three open benchmark datasets show that the proposed method significantly outperforms several state-of-the-art methods for automatic image annotation.

References

[1] A.B. Goldberg, X. Zhu, B. Recht, J. Xu, and R.D. Nowak, "Transduction with Matrix Completion: Three Birds with One Stone," *Proc. Neural Information Processing Systems Foundation Conf.*, pp. 757-765, 2010.

[2] A. Gammerman, V. Vovk, and V. Vapnik, "Learning by Transduction," *Proc. Conf. Uncertainty in Artificial Intelligence,* pp. 148-155, 1998.

[3] A. Ioffe, "Composite Optimization: Second Order Conditions, Value Functions and Sensityvity," *Analysis and Optimization of Systems*, vol. 144, pp. 442-451, 1990.

[4] A. Makadia, V. Pavlovic, and S. Kumar, "A New Baseline for Image Annotation," *Proc. 10th European Conf. Computer Vision,* pp. 316-329, 2008.

[5] A. Rakhlin, "Applications of Empirical Processes in Learning Theory: Algorithmic Stability and Generalization Bounds," PhD thesis, MIT, 2006.

[6] S.L. Feng, R. Manmatha, and V. Lavrenko, "Multiple Bernoulli Relevance Models for Image and Video Annotation," Proc. *IEEE Conf. Computer Vision and Pattern Recognition*, pp. 1002-1009, 2004.

[7] P. Duygulu, K. Barnard, J.F.G. de Freitas, and D.A. Forsyth, "Object Recognition as Machine Translation: Learning a Lexicon for a Fixed Image Vocabulary," *Proc. Seventh European Conf. Computer Vision*, pp. 97-112, 2002.

[8] M. Guillaumin, T. Mensink, J. Verbeek, and C. Schmid, "TagProp: Discriminative Metric Learning in Nearest Neighbor Models for Image Auto-Annotation," *Proc. 12th IEEE Int'l Conf. Computer Vision*, pp. 309-316, 2009.

[9] L. Wu, L. Yang, N. Yu, and X.-S. Hua, "Learning to Tag," *Proc. 18th Int'l Conf. World Wide Web*, pp. 361-361, 2009.

[10] L. Yang, R. Jin, L. Mummert, R. Sukthankar, A. Goode, B. Zheng, S.C.H. Hoi, and M. Satyanarayanan, "A Boosting Framework for Visuality-Preserving Distance Metric Learning and Its Application to Medical Image Retrieval," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 32, no. 1, pp. 30-44, Jan. 2010.

[11] 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.

[12] L. Chen, D. Xu, I.W. Tsang, and J. Luo, "Tag-Based Web Photo Retrieval Improved by Batch Mode Re-Tagging," *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pp. 3440-3446, 2010.

[13] R. Datta, W. Ge, J. Li, and J.Z. Wang, "Toward Bridging the Annotation-Retrieval Gap in Image Search by a Generative Modeling Approach," *Proc. 14th Ann. ACM Int'l Conf. Multimedia*, pp. 977-986, 2006.

[14] P. Heymann, G. Koutrika, and H. Garcia-Molina, "Can Social Bookmarking Improve Web Search?" *Proc. Int'l Conf. Web Search and Web Data Mining*, pp. 195-206, 2008.

[15] C. Liu, J. Yuen, and A. Torralba, "Nonparametric Scene Parsing via Label Transfer," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 33, no. 12, pp. 2368-2382, Dec. 2011.

[16] M. Guillaumin, T. Mensink, J. Verbeek, and C. Schmid, "TagProp: Discriminative Metric Learning in Nearest Neighbor Models for Image Auto-Annotation," *Proc. 12th IEEE Int'l Conf. Computer Vision,* pp. 309-316, 2009.

[17]] D. Metzler and R. Manmatha, "An Inference Network Approach to Image Retrieval," *Proc. Int'l Conf. Image and Video Retrieval*, pp. 42-50, 2004.

[18] 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.

[19] F. Monay and D. Gatica-Perez, "PLSA-Based Image Auto-Annotation: Constraining the Latent Space," *Proc. 12th Ann. ACM Int'l Conf. Multimedia*, pp. 348-351, 2004.

[20] L. Wu, S.C. Hoi, J. Zhu, R. Jin, and N. Yu, "Distance Metric Learning from Uncertain Side Information with Application to Automated Photo Tagging," *Proc. 17th ACM Int'l Conf. Multimedia*, 2009.

[21] S.C.H. Hoi, W. Liu, M.R. Lyu, and W.-Y. Ma, "Learning Distance Metrics with Contextual Constraints for Image Retrieval," *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pp. 2072-2078, 2006.

[22] Z. Wu, Q. Ke, J. Sun, and H.-Y. Shum, "Scalable Face Image Retrieval with Identity-Based Quantization and Multireference Reranking," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 33, no. 10 pp. 1991-2001, Oct. 2011.

[23] K.-S. Goh, E.Y. Chang, and B. Li, "Using One-Class and Two-Class Svms for Multiclass Image Annotation," *IEEE Trans. Knowledge and Data Eng*., vol. 17, no. 10, pp. 1333-1346, Oct. 2005.

[24] G. Zhu, S. Yan, and Y. Ma, "Image Tag Refinement towards Low-Rank, Content-Tag Prior and Error Sparsity," *Proc. ACM Int'l Conf. Multimedia*, pp. 461-470, 2010.

