IOSR Journal of Computer Engineering (IOSR-JCE)
e-ISSN: 2278-0661,p-ISSN: 2278-8727, Volume 16, Issue 5, Ver. VIII (Sep – Oct. 2014), PP 40-49
www.iosrjournals.org

Optimized Hypergraph Based Social Image Search Using Visual-Textual Joint Relevance Learning

Arya S
(Dept. of Computer Science and Engineering, Mahatma Gandhi University, Kerala, India)

Abstract: *Recent years have witnessed a great success of social media websites. Tag-based image search is an important approach to access the image content of interest on these websites. However, the existing ranking methods for tag-based image search frequently return results that are irrelevant or lacking in diversity. Most of the existing methods estimate the relevance of images by using tags and visual characteristics either separately or sequentially. The proposed system uses an approach that utilize simultaneously both visual information and textual information in real time to estimate the relevance of user tagged image. The method used to determine the relevance estimation is the hypergraph learning approach. The hypergraph is a generalization of a graph in which an edge in the hypergraph can be connected to any number of vertices. In the proposed method each social image can be represented by the bag-of-visual words and bag-of-textual words, which can be obtained from the textual content and visual content of the particular image. A hypergraph can be constructed in which the vertices represent the social images for ranking and the each hyperedge represents the visual words or tags that are obtained from the image. In the hypergraph learning scheme, both the visual content and tag information are taken into consideration at same time. Different from the method used by the traditional hypergraph, in the proposed system a social image hypergraph is constructed where vertices represent the images and hyperedges represent the visual or textual terms. The set of pseudo-positive images are used to achieve the learning, where the weight of hyperedges are updated throughout the learning process. Thus only the most relevant images are given to the user.*

Index Terms*: Hypergraph learning, social image search, tag, visual-textual.*

I. Introduction

There is an explosion of social media content available online, such as Flickr, Youtube, google and Zooomr. Such media repositories promote users to collaboratively create, evaluate and distribute media information. They also allow users to annotate their uploaded media data with descriptive keywords called tags. The explosive growth of multimedia and network technologies have lead to the rapid development of the social media in recent years. The growth of social media websites lead to the extensive research of efforts that have been dedicated to tag-based social image search. Both the visual information and the tags are considered for the searching social media websites. The visual information and tags plays a major role in searching of images in social media websites**.** Fig. 1 illustrates a social image and is associated with the user-provided tags. These valuable metadata can greatly useful to facilitate the organization and search of the social media.

The images can be easily retrieved for a given query by indexing the images with its associated tags. However, since user-provided tags are usually noisy and incomplete, simply applying text-based retrieval approach may lead to unsatisfactory results. The tag-based social image search cannot achieve satisfactory results because of the high amount of noise present in the user-provided tags. Most of the tags are incorrectly spelled so they can be considered as irrelevant. In a study it is reported only 50% of the tags provided by the Flicker users are related to the images the rest are irrelevant tags.

The lack of an optimal ranking strategy is another reason for the unsatisfactory search results. Currently, Flickr website provides two ranking options for tag-based social image search. One is "most recent", which ranks images based on their uploading time, and the other is "most interesting", which ranks the images based on each image's "interestingness" in flickr, a measure that integrates the information of click-through, comments, etc. These two methods are known as time-based ranking and interestingness-based ranking, respectively. They both rank images according to measures (interestingness or time) that are not related to relevance and it results in many irrelevant images in the top search results. These two methods do not consider visual content and textual content of the images. So they are not based on the relevance measures. Thus, the search results are not efficient in terms of relevance. Therefore, a ranking approach that is able to explore both the tags and images' content is desired to provide users better social image search results. A good ranking should take both visual and textual contents.

Fig. 1 An example of a social image with its associated tags

.

Present day visual search engines, like Google and Yahoo, rely on textual information associated to the visual data, such as textual image descriptions in HTML documents. As these information sources do not originate from the visual content of the image, they often provide inadequate descriptions for retrieval. When retrieval of visual data is performed on the basis of textual information only, the accuracy of the search results is likely to be suboptimal.

Several algorithms have been proposed to improve the efficiency of social image search. Most of the existing methods use tags and visual content either separately or sequentially. In the separated approach for social image search only the visual information is used to calculate the relevance score. But the sequential approach uses both tags and visual information sequentially. In this method tags are first used to generate the initial relevance score then these scores are refined using the visual information. Sequentially or separately using these two information sources is suboptimal for social-image search.

In this paper, I propose a hypergraph-based approach to utilize simultaneously the visual information and tags for image relevance learning. In the proposed method, each social image is represented by bag-of-textual-words and bag-of visual- words features, which are generated from the tags and the visual content of the image, respectively. A hypergraph is constructed, in which the vertices denote the social images for ranking, and each visual word or tag generates a hyperedge. In such a hypergraph learning scheme, both the visual content and the tag information are taken into consideration at the same time. Different from the method by using the traditional hypergraph learning approaches that adopts fixed hyperedge weights, we further learn the weights which indicate the importance of different visual words and tags. In this way, the effects of the informative visual words and tags can be enhanced. In the learning process, we first identify a set of pseudo relevant samples based on tags. Then, we calculate the relevance scores of images by iteratively updating them and the weights of hyperedges.

II. Related Work

This section briefly introduce the related work on social image search

1) *Social Image Search*

The image search technology has witnessed a great advance in last decade. Different from the general web images, social images are usually associated with a set of user-provided descriptors called tags, and thus tag-based image search can be easily accomplished by using these descriptors as index terms. Since user-provided tags are usually very noisy and irrelevant this search frequently results in unsatisfactory search results. In comparison with the extensive studies on how to help users better perform tagging or mining tags for other applications, the literature regarding tag-based image search is still very sparse. Most of such studies focuses on how to refine the tags of images or measure their relevance levels. Li et al. proposed a tag relevance learning method which is able to assign each tag a relevance score, and they have shown its application in tag-based image search. Li et al .proposed an optimization scheme for tag refinement based on the visual and semantic connection between images. Sun and Bhowmick proposed a method to measure the tag clarity score based on the query language model and the collection language model. These methods can help tag-based image search by improving the tags' quality, but they cannot deal with the aforementioned lack-of-diversity problem.

2) Traditional image retrieval framework

Content-based image retrieval (CBIR) has been popular for many years. The task requires to find the given query image from a given image collection that is similar by the search engine. Traditional methods for CBIR are based on a vector space model. These methods represent an image as a set of features and the difference between two images is measured through a similarity function between their feature vectors. While there have been no large-scale, standardized evaluations of image retrieval systems, most image retrieval systems are based on features representing color [6], texture, and shape that are extracted from the image pixels.

The most straightforward approach to find matching images is 'Nearest neighbor' search. It contains the implicit assumption that for each feature the class posterior probabilities are approximately constant for matching and non-matching images. However nearest neighbor search has two major drawbacks. First, the nearest neighbor might assign equal weight to both the relevant features and irrelevant features. Thus, the retrieval accuracy will suffer dramatically if a large number of features of an image are irrelevant to the query. Therefore we can prefer many images similar with respect to the irrelevant features. It is reasonable to select a subset of features before the nearest neighbor search. But most feature selection techniques require large amounts of labeled data. Applying feature selection to the retrieval problem becomes rather difficult since usually only a small number of (image) query examples are given.

3) Separated Methods

In separated methods the tags or the visual contents are utilizes separately in order to calculate the relevance of the social image. The relevance score of the social image is calculated only by using the visual content or the textual content of the image. Thus the search result may not be sufficiently good.

4) Sequential Methods

In sequential approach visual information and tags and employed sequentially for social image search. However, in most of the existing methods textual-content based analysis is performed first and then the visual content based analysis is performed next. In the relevance based ranking method used for social image search first the relevance scores are calculated based on the tags of the images and then these calculated relevance scores are refined using the visual content of the images. Though more than half of the tags are noisy there are also meaningful tags that are useful for the searching of the image. Therefore, separately or sequentially using two information is suboptimal for social image search.We summarize these separated and sequential schemes in Fig.2, where separated methods can be further divided into textual content only methods and visual content-only methods.

III. Hypergraph Analysis

Before explaining the proposed method we are going to give a brief introduction to the hypergraph analysis. In a simple graph, samples are represented by vertices and an edge links the two related vertices. Learning tasks can be performed on a simple graph. For instance, assuming that samples are represented by feature vectors in a feature space, an undirected graph can be constructed by using their pair wise distances, and graph-based semi-supervised learning approaches can be performed on this graph to categorize objects. It is noted that this simple graph cannot reflect higher-order information. Compared with the edge of a simple graph, a hyperedge in a hypergraph is able to link more than two vertices. For clarity, we first illustrate several important notations and their definitions throughout the paper in Table I.

TABLE I
NOTATIONS AND DEFINITIONS

NOTATION	DEFINITION
$\mathcal{X} = (x_1, x_2, \ldots, x_n)$	$\mathcal{X}$ indicates the image set and x_i indicates the i^{th} image.
f_i^{bow}	The $n_c \times 1$ bag-of-visual-words feature of vector for x_i.
f_i^{tag}	The n_t bag-of-textual-words feature of vector for x_i.
n_c	The size of visual codebook.
n_t	The number of employed tags
$G = (V, \mathcal{E}, w)$	G indicates a hypergraph, and V, $\mathcal{E}$ and w indicate the set of vertices, the set of edges, and the weights of hyperweights, respectively.
N	The number of images in hypergraph learning.
V	The set of n vertices of the hypergraph.
$\mathcal{E}$	The set of edges of the hypergraph that contains n_e elements, where n_e is the number of edges.
$w = [w_1, w_2, \ldots, w_{ne}]$	The $n_e \times 1$ weight vector of the hyperedges in the hypergraph.
$\delta(e)$	The degree of edge e.
D_v	The $n \times n$ diagonal matrix of the vertex degrees
D_e	The $n_e \times n_e$ diagonal matrix of the edge degrees.
H_i	The $n \times n_e$ incidence matrix for i-th hypergraph.

K	The number of the selected pseudo-relevant images.
Y	The $n \times 1$ label vector for hypergraph learning. The elements of the pseudo-relevant images are set to 1, and the others are 0.
F	The $n \times 1$ to-be-learned relevance score vector.

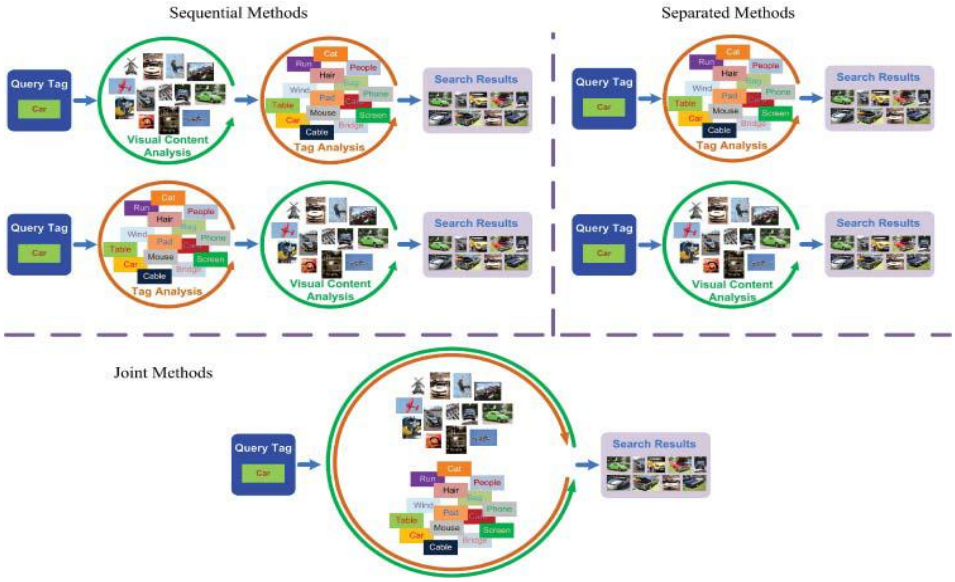


Fig.2. Illustration of different social image search methods.

A hypergraph $G = (V, \mathcal{E}, w)$ is composed by a vertex set *V*, an edge set $\mathcal{E}$, and the weights of the edges *w*. Each edge *e* is given a weight *w(e)*. The hypergraph *G* can be denoted by a $|V| \times |\mathcal{E}|$ incidence matrix **H** with entries defined as:

$$h(v,e) = \begin{cases} 1 \ if \ v \in e \\ 0 \ if \ v \notin e \end{cases} \quad (1)$$

For a vertex $v \in V$,its vertex degree can be estimated by:

$$d(v) = \sum_{e \in \mathcal{E}} w(e)\, h(e) \quad (2)$$

The degree of an hyperedge $e \in \mathcal{E}$ in a hypergraph can be calculated by:

$$\delta(e) = \sum_{v \in V} h(v,e). \quad (3)$$

Diagonal matrix of the vertex and diagonal matrix of the edge vertex can be denoted by denoted by $\boldsymbol{D_v}$ and $\boldsymbol{D_e}$ respectively. Let W denote the diagonal matrix of the hyperedge weights

$$W(i,j) = \begin{cases} w(i) & if \ i = j \\ 0 & otherwise \end{cases} \quad (4)$$

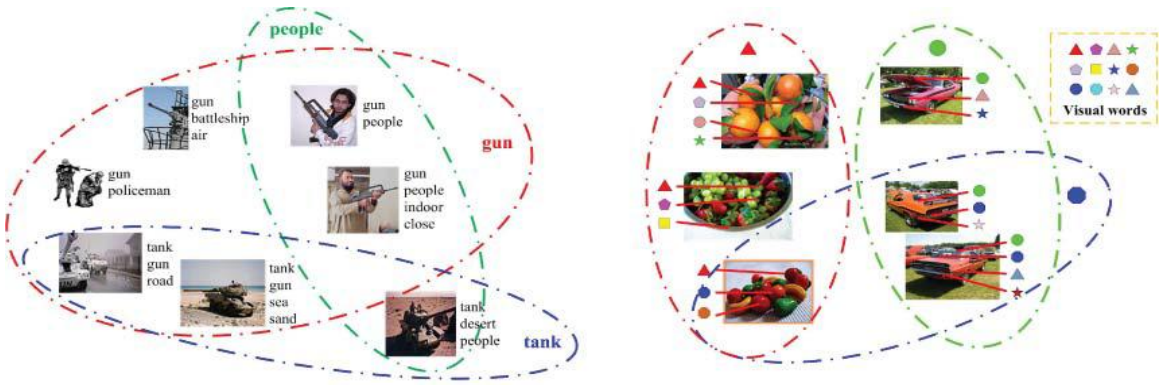


Fig. 3. Examples of hyperedge construction. (a) Example of textual hyperedge construction, where three hyperedges are generated by tags "people," "gun," and "tank." (b) Example of visual hyperedge construction, where three hyperedges are generated by three visual words.

IV. Optimized Visual-Textual Relevence Learning

This section explains the proposed hypergraph-based visual-textual joint relevance learning approach by simultaneously using both the visual content and the textual information. The algorithm for visual-textual joint relevance learning is shown in Algorithm 1.

Algorithm Visual-Textual Joint Relevance Learning Method for Social Image Search.

Input: The image set for re-ranking $\mathcal{X} = (x_1, x_2, \ldots\ldots, x_n)$, reference image y

Output: The relevance score vector f for image re-ranking

Step 1. Hypergraph Construction

1. Regard each social image in the social image set $\mathcal{X} = (x_1, x_2, \ldots\ldots, x_n)$ as a vertex in the hypergraph.
2. Generate a bag-of-visual-words for each image

3. Generate a bag-of-visual-words for the reference image
4. Construct hyperedges by using bag-of-visual-words. There are n_c visual hyperedges in total.
4. For each image, the tags are ranked by and only top min (n_l, n_i) tags are left for further processing. Here ni is the number of tags in xi, and nl is set as 100 in our experiments.
5. Generate bag-of-textual-words for the reference image.
6. Generate a bag-of-textual-words for each image.
7. Construct hyperedges by using bag-of-textual-words the same textual words are connected by one hyperedge. There are n_t textual hyperedges in total degrees $\mathbf{D}v$ and $\mathbf{D}e$, the initial weights of all hyperedges w, respectively
8. Generate the incidence matrix $\mathbf{H}_i$, the diagonal matrices of the vertex degrees and the hyperedge.

Step 2. Pseudo-Relevant Sample Selection

The Relevance Distance is employed to estimate the semantic relevance of an image x_i to the query tag t_q .

Step 3. Relevance Learning on Hypergraph

Conduct semi-supervised learning on the hypergraph structure to find score vector f and the weights for hyperedge w.

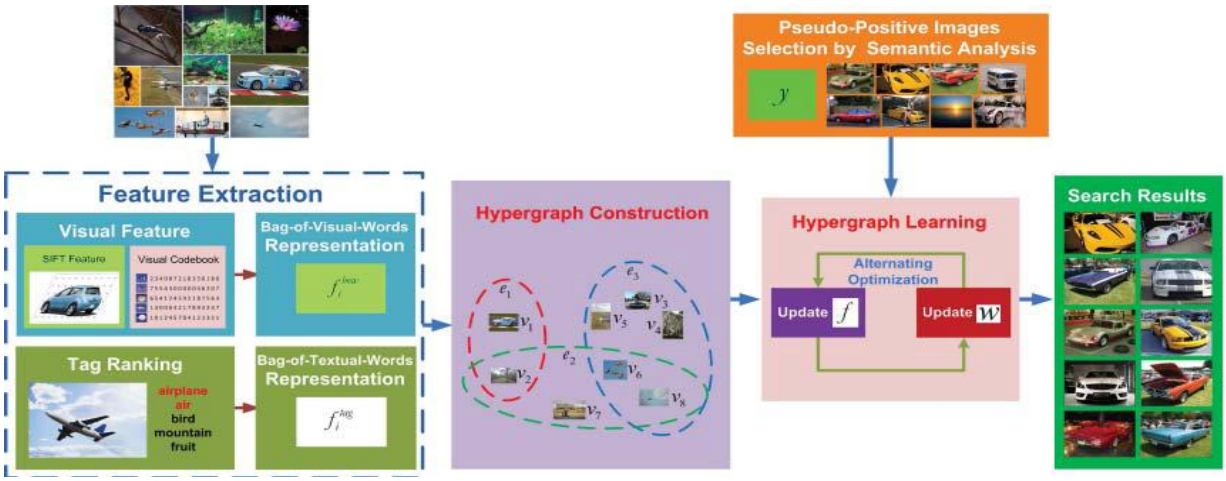


Fig.4. Schematic illustration of the proposed optimized social image search using visual textual joint relevance learning.

A. Hypergraph Construction

In order to represent higher order information we are using hypergraph. Compared to the edges of a simple graph, a hypergraph is able to link more than two vertices. A hypergraph $G = (V, \mathcal{E}, w)$ is composed by a vertex set V, an edge set $\mathcal{E}$, and the weights of the edges w. Each social image set $\mathcal{X} = \{x_1, x_2, \ldots\ldots, x_n\}$ as a vertex in the hypergraph $G = (V, \mathcal{E}, w)$. Let n indicate the total number of images in $\mathcal{X}$, and thus the generated hypergraph has n vertices. To create the hyperedges visual information and textual information are extracted from an image. For the visual content of each social image, the bag-of-visual-words representation is employed for image description. To generate the bag-of-visual-words representation, a dense set of uniformly distributed points are first identified for each social image, and the local SIFT descriptors on these points are extracted. Let n_c indicate the size of the bag-of-visual-words codebook. Each image xi is represented by an $n_c \times 1$ feature vector f_i^{bow}, where $f_i^{bow}(k, 1) = 1$ indicates that x_i contains at least one data point belonging to the k-th visual code.

In order to construct the hyperedge for each image x_i , the associated tags are first ranked. The probability density function is used to estimate the initial relevance scores for these tags. Then a random walk over a tag similarity graph is performed to refine the relevance scores. Only the refined tags are left for further processing.

Next generate the bag of-textual-words representation for each image by using the n_t tags. Each image x_i is represented by an $n_t \times 1$ feature vector f_i^{tag}, where $f_i^{tag}(k, 1) = 1$ indicates that *xi* contains the *k*-th selected tag. With the help of the feature vectors such as bag-of-visual-words and bag of-textual-words the hypergraph is constructed. Fig. 3 provides an example to show how visual and textual hyperedges are constructed. In the example, there are three hyperedges constructed by visual words or tags respectively. The hypergraph construction and learning is illustrated in Fig.4.

B. Social Image Relevance Learning Formulation on Hypergraph

In the above step we have constructed a hypergraph in which the vertices denotes the social image. Next step is the searching of the image in the hypergraph. The searching process is known as binary classification problem. Then we calculate the relevance

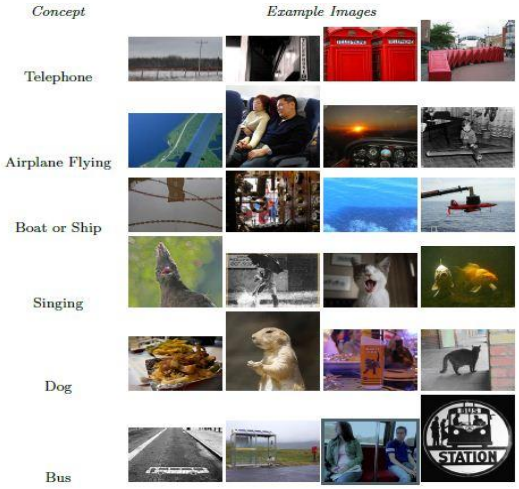


Fig.5. Several example images and their associated tags used in our experiments

scores among the different vertices in the hypergraph, and the transductive inference is also formulated as a regularization framework arg min$\{\Omega(f) + \lambda R_{temp} + \mu\psi(\omega)\}$. Here the regularizer term $\Omega(f)$ on the hypergraph structure applies the formation of $\Omega(f)$ in the social image search task. $\Omega(f)$ indicates that highly related vertices should have close label results, which is defined as:

$$\frac{1}{2}\sum_{i=1}^{n_e}\sum_{v\in V}\frac{w_i h(u,e_i)\, h(v,e_i)}{\delta(e_i)}\left(\frac{f(u)}{\sqrt{d(u)}} - \frac{f(v)}{\sqrt{d(v)}}\right)^2 \quad (5)$$

Where the vector f is to-be-learned relevance score vector. Eq. (5) further turns into:

$$\Omega(f) = \sum_{i=1}^{n_e}\sum_{v\in V}\frac{w_i h(u,e_i)\, h(v,e_i)}{\delta(e_i)} \times\left(\frac{f^2(u)}{d(u)} - \frac{f(u)f(v)}{\sqrt{d(u)d(v)}}\right)$$

$$= \sum_{u\epsilon v} f^2(u)\sum_{i=1}^{n_e}\frac{w_i h(u,e_i)}{d(u)}\sum_{v\in V}\frac{h(v,e_i)}{\delta(e_i)} - \sum_{i=}^{n_e}\sum_{u,v\in V}\frac{f(u)h(u,e_i)w_i\ h(v,e_i)f(v)}{\sqrt{d(u)d(v)}\ \delta(e_i)}$$

$$= f^T(I-\theta)f \quad (6)$$

Where $\theta = D_v^{-\frac{1}{2}} HWD_e^{-1}\ H^T D_v^{-\frac{1}{2}}$. Let $\Delta = I - \theta$, where Δ is the normalized hypergraph Laplacian. Thus we rewrite the regularizer $\Omega(f)$ as:

$$\Omega(f) = f^T\Delta f \quad (7)$$

In the constructed hypergraph, all the hyperedges are initialized with an identical weight. However, the hyperedges are with different effects a1so there exists a lot of uninformative visual words and tags for a given query. Therefore, performing a weighting or selection on the hyperedges will be helpful. Here we integrate the learning of the hyperedge weights into the formulation.

C. Pseudo-Relevant Sample selection

Pseudo-relevant samples are used in the hypergraph learning algorithm. In this method we simply estimate the relevance of image *x* to tag t_q and all tags of *x*.

$$s(x, t_q) = \frac{1}{n}\sum_{t\in\tau_i} s_{tag}(t_q, t)$$

Where τ_i is the tag set of x_i . Thus all the social images that are associated with the tag are ranked in the descending order and the top K results are the pseudo-relevant images.

V. Experiment Result

A. Experimental Settings

The experiment is conducted using flicker dataset and google dataset. The collected dataset is based on their popularity of tags such as *apple .fruit, car, fish, bird, room, camera, building, water, tree , cow, box, flower, chicken, motorcycle, telephone, mobile, watch, waterfall, weapon, gun, lion ,tiger, rice, swimmer, lake, horse, hockey, forest, spider, crow, frog, pen, book, bmw, furniture, clock, lotus, leaf, table, turtle.* These tags are employed to search images and the top 1000 searching results for each query tag are collected with their associated information. There are 5000 images and 2000 unique tags in total. Fig. 5 shows some example images and their associated tags used in our experiments. In our experiment the images with different meanings are also taken into consideration. We compare the following methods:

1) Hypergraph-based relevance learning with hyper edge weight estimation i.e. without the real time processing. The method is denoted as "HG-WE."
2) Optimized social image search using visual-textual joint relevance learning, i.e., the proposed approach. The method is performs the search in real time to find the required images.

TABLE II
THE *NDCG*@20 RESULTS OF TWO METHODS

Query	**HG-WE**	**Optimized real time search**
apple	0.5759	0.7183
beach	0.7609	1.0000
bird	0.9576	0.9653
bmw	0.6224	0.7267
car	0.8095	0.7991
chicken	0.7609	1.0000
cow	0.9065	1.0000
eagle	1.0000	1.0000
flower	0.7288	0.8888
forest	0.9460	0.9951
fruit	0.4642	0.9397
furniture	0.8255	0.8560
hair	0.5630	0.9346
horse	1.0000	0.9865
lion	0.6971	0.9175
rainbow	0.8035	0.9093
spider	0.6155	0.8306
telephone	0.7034	0.9204
turtle	0.8069	0.9537
watch	0.6832	0.9669
waterfall	0.7623	0.9249
weapon	0.5253	0.6338

We set the size of the tag and the visual directory to 1000, i.e., $n_c=n_t=1000$. We set K to 100 for pseudo-relevant sample selection. i.e., we take 100 pseudo-relevant images. The parameter n_1 is set as 10. For the above two methods, I randomly select 1000 images that are not associated with query tags as negative in the learning process. The Normalized Discounted Cumulative Gain (NDCG) is employed for performance evaluation.

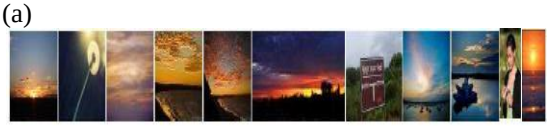
B. Experimental Results and Discussion

Table II illustrates the *NDCG*@20 comparison of two methods. Here we illustrate not only the NDCG measurements of each query but also the average NDCG measurements of the 22 queries. From the results we have the following observations:

1) The proposed method achieves better performance than the "HG-WE" method.
2) The proposed method achieves best results for most queries. Among the two methods our proposed method achieves best average performance. This shows that the proposed method can greatly improve the performance of hypergraph learning.

From the table II we can see that our proposed approach outperforms the existing method. Fig. 6 demonstrates the top 10 results obtained by 2different methods for the example query *sunset*. From the figure we can understand that the proposed method is superior.

(a)



(b)

Fig.6. Top results obtained by different methods for the query *sunset*.(a) Hypergraph based visual-textual relevance learning with hyperedge i.e., HG-WE (b) optimized real time visual-textual relevance learning i.e., the proposed method.

(a)

(b)

Fig. 7. Top results obtained by different methods for the query *apple*. (a) Hypergraph based visual-textual relevance learning with hyperedge i.e., HG-WE (b) optimized real time visual-textual relevance learning i.e., the proposed method.

.

We further investigate the performance of the proposed method on query with multiple meanings. The proposed method is not limited by multiple meanings, and any image with one meaning from all different meanings is regarded as relevant. In our method, the pseudo-relevant sample selection procedure is not limited to any special meaning. Therefore, the final ranking list can preserve images from all different meanings, and users can obtain searching results with different meanings. In our experiments, the query *apple* and the query *jaguar* are two queries with more than one meaning. Fig. 7 demonstrates the top 10 results obtained by different methods for an example query *apple*. The proposed method can return relevant results with different meanings, which demonstrates the superiority of the proposed approach.

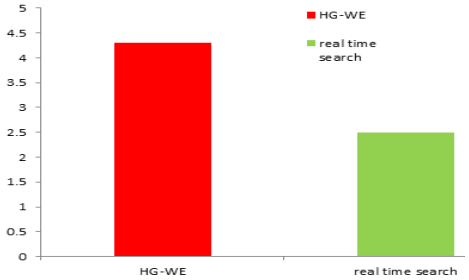


Fig.8. Computational cost comparison.

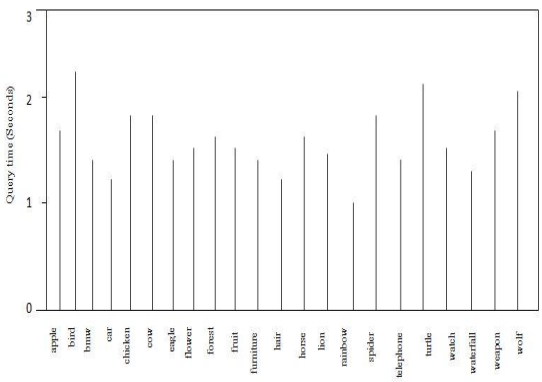


Fig.9. Computational cost for each query tag of the proposed method.

C. On the Parameter n1

To filter the noise tags in the hypergraph construction procedure the parameter *n1* is employed. When *n1* is small the performance of the proposed method is worst. When *n1* increase, the image search performance in terms of NDCG@20 becomes better while growth speed becomes slow. The image search performance is relatively steady, when *n1* is greater than 10. That is when *n1* is too small, only a few tags are kept for further processing, which may lead to information lost by removing most of the tags. With the increase of *n1*, more tags are selected, which can employ more meaningful tags for further processing and improve the image search performance. When *n1* is large enough, most of meaningful tags have been selected, and continuing to select more tags may not only involve useful tags but also bring in more noise tags, which could reduce the image search performance.

D. On the Running Time Comparison

The cost per query of existing system and proposed method is analyzed in this section and result is shown in Fig.8. The computational cost for each query tag of the proposed method is provided in Fig. 9. The computational costs are recorded on a PC with Pentium 4 2.0GHz and 4G memory. We draw following conclusions.

1) The proposed method has higher computational cost but better performance compared to the other method.

The proposed method is most efficient with worst performance.

VI. Conclusion

This paper proposes an approach that simultaneously utilizes both textual information and visual information in real time for social image search. In the proposed method first the hyperedges of the hypergraph is constructed using the visual content and tags of the image. Then the relevance learning procedure is performed on the hypergraph structure. Different from the conventional hypergraph learning algorithms, our approach learns not only the relevance scores among images but also the weights of hyperedges. The effects of uninformative visual words and tags can be minimized by performing the learning of hyperedge weights. To test the performance of the proposed system, we conducted experiments on the datasets of flickr and google. The experimental results shows that the proposed method achieved better performance than other existing methods. Besides the relevance performance, video search based on semantic concept detectors is critically dependent on tagged example image. Social tagged images can act as a training resource for concept-based video search which is my future work.

Reference

[1] Y. Gao, M. Wang, H. Luan, J. Shen, S. Yan, and D. Tao, "Tag-based social image search with visual-text joint hypergraph learning," in *Proc. ACM Conf. Multimedia*, 2011, pp. 1517 1520.

[2] Z.-J. Zha, L. Yang, T. Mei, M. Wang, and Z. Wang, "Visual query suggestion," in *Proc. ACM Conf. Multimedia*, 2009, pp. 15–24.

[3] J. Shen, D. Tao, and X. Li, "QUC-tree: Integrating query context information for efficient music retrieval," *IEEE Trans. Multimedia*, vol. 11, no. 2, pp. 313–323, Feb. 2009.

[4] J. Shen, D. Tao, and X. Li, "Modality mixture projections for semantic video event detection," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 18, no. 11, pp. 1587–1596, Nov. 2008.

[5] Cilibrasi, R., Vitanyi, P.M.B.: The google similarity distance. *IEEE Trans. Knowl. Data Eng*.**19**, 370–383 (2007)

[6] Clarke, C.L.A., Kolla, M., Cormack, G.V., Vechtomova, O., Ashkan, A., Büttcher, S., MacKinnon,I.: Novelty and diversity in information retrieval evaluation. In: Proceedings of the 31st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 659–666. ACM, New York (2008)

[7] Hsu, W.H., Kennedy, L.S., Chang, S.-F.: Video search re-ranking via information bottleneck principle. In: Proceedings of ACM Multimedia, pp. 35–44 (2006)

[8] J. Shen and Z. Cheng, "Personalized video similarity measure," *ACM Multimedia Syst. J.*, vol. 15, no. 7, pp. 421–433, 2011.

[9] M. Kato, H. Ohshima, S. Oyama, and K. Tanaka, "Can social tagging improve web image search?" in *Web Information Systems Engineering* (Lecture Notes in Computer Science). vol. 5175. Berlin, Germany: Springer-Verlag, 2008, pp. 235–249.

[10] K. Yang, M. Wang, X. S. Hua, and H. J. Zhang, "Social image search with diverse relevance ranking," in *Proc. ACM Conf. Multimedia Model.*, 2009, pp. 174–184.

[11] D. Liu, M. Wang, X.-S. Hua, and H.-J. Zhang, "Semi-automatic tagging of photo albums via exemplar selection and tag inference," *IEEE Trans. Multimedia*, vol. 13, no. 1, pp. 82–91, Feb. 2011.

[12] X. Li, C. G. M. Snoek, and M. Worring, "Learning tag relevance by neighbor voting for social image retrieval," in *Proc. ACM Conf. Multimedia Inf. Retr.*, 2008, pp. 180–187.

[13] D. Liu, M. Wang, L. Yang, X. S. Hua, and H. Zhang, "Tag quality improvement for social images," in *Proc. IEEE Int. Conf. Multimedia*, Jul. 2009, pp. 350–353.

[14] A. Smeulders, M. Worring, S. Santini, A. Gupta, and R. Jain, "Contentbased image retrieval at the end of early years," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 12, pp. 1349–1380, Dec. 2000.

[15] D. Zhou, J. Huang, and B. Schokopf, "Learning with hypergraphs: Clustering, classification, and embedding," in *Proc. Adv. Neural Inf. Process. Syst.*, 2007, pp. 1–8.

[16] R. H. V. Leuken, L. Garcia, X. Olivares, and R. Zwol, "Visual diversification of image search results," in *Proc. ACM Int. World Wide Web Conf.*, 2009, pp. 341–350.

[17] K. Jarvelin and J. Kekalainen, "Cumulated gain-based evaluation of IR techniques," *ACM Trans. Inf. Syst.*, vol. 20, no. 4, pp. 422–466, 2002.

[18] M. Campbell, A. Haubold, M. Liu, A. Natsev, J. Smith, J. Tesic, L. Xie, R. Yan, and J. Yang. IBM Research TRECVID-2007 Video Retrieval System. TRECVID Workshop, Gaithersburg, USA, November, 2007.
[19] X. Li, C. Snoek, and M. Worring. Learning Tag Relevance by Neighbor Voting for Social Image Retrieval. In Proceedings of the ACM International Conference on Multimedia Information Retrieval, pages 180{187, Vancouver, Canada, October 2008.
[20] Y.-G. Jiang, C.-W. Ngo, and J. Yang, "Toward optimal bag-of-features for object categorization and semantic video retrieval," in *Proc. ACM Int. Conf. Image Video Retr.*, 2007, pp. 494–501.
[21] L. Wu, X. Hua, N. Yu, W. Ma, and S. Li, "Flickr distance," in *Proc. ACM Conf. Multimedia*, 2008, pp. 31–40.
[22] A. Sun and S. S. Bhowmick, "Image tag clarity: In search of visual representative tags for social images," in *Proc. SIGMM Workshop Social Media*, 2009, pp. 19–26
[23] Y. Yang, F. Nie, D. Xu, J. Luo, Y. Zhuang, and Y. Pan, "A multimedia retrieval framework based on semi-supervised ranking and relevance feedback," *IEEE Trans. Pattern Anal. Mach. Intel.*, vol. 34, no. 4, pp. 723–742, Apr. 2012.