

Contextual Sub Image submission for the Retrieval of Information Using Image Search in Web Mining

Dr V.Meenakshi¹, Pricilla Mellina Rose², Vidhyalakshmi³

Assistant Professor, Department of Computer Applications, R.V. Govt. College, Chengalpet, Tamil Nadu, India¹

M.Phil Student, Department of Computer Science, Mother Teresa Women's University Chennai, India²

Assistant Professor, Department of Civil Engineering, SRM University, Ramapuram, Tamil Nadu, India³

ABSTRACT: The project is mainly concerned with retrieval of information from web source. This is probably achieved in the way of Question and Answering System. Many servers like Google, yahoo and other servers use match retrieval algorithm to fetch the probably best answering system. Normally text will be given as a question tag in the space provided in the web search engine and probable answer will be enriched with multi-media inclusive of image, audio and video. The system is used to fetch the content of the multimedia from Google API's. Since, the data is got from the web, it is necessary to process and rank and re-rank the search results. Fetching the information through query based on image leads to the image search in Web Mining. Search procedures let us search information using image instead of keywords as text. A picture from the desktop, tablet or mobile phone is to be uploaded and the search engines will show all the required web pages from the internet that is having similar type of images. Based on classifiers for all pre-defined attributes, each and every image is represented by an attribute feature which consists of the responses from the classifiers. In the paper, we propose a novel sub-image submission for the image search. Hence in addition to the image submitted, the web frame work will automatically ask for the contextual based sub-image submission thereby narrowing the search. The proposed sub-image submission significantly improves both the accuracy as well as the efficiency of the information retrieval through image search.

KEYWORDS: Image search, textual and visual context, contextual Image, contextual re-ranking

I. INTRODUCTION

Searching for the information leads to two main schemes of searching the images on the Web. In the first method, images are searched for by a query in the form of a textual keyword. This method is widely adapted in general search in web. In the second method, example-based scheme allows the users to search for similar images by providing an uploaded example image serving as query. This method is used in mobile search technology.



Fig1.1 Image Search based on keyword or text

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 8, August 2016

The first scheme of search based on keyword-based scheme is illustrated in Fig 1.1. In this method images are searched for by a query in the form of a textual keyword. This scheme can be seen as a direct extension of the widely adopted general Web search.

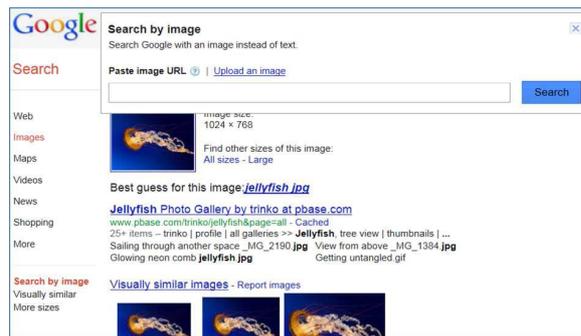


Fig. 1.2 Image Search based on search by image

The second, example-based scheme allows the users to search for similar images by providing an uploaded example image serving as query. This scheme is illustrated in Figure 1.2. This type of fetching data using image has been deployed in more and more commercial Web search engines (e.g. TinEye) and this scheme is also used in mobile search technology (e.g. Google Goggles).

In this paper, we propose a novel image search scheme based on contextual image search. This approach is different from the conventional image search schemes in which a separate interface (e.g., image input box) is used to allow users to submit a sub-image which again act as a sub-query. After getting the new sub-image the search scheme enable the user to narrow down the search images so that they could get the required information in a quick way.



Fig.1.3 Google image retrieval

Fig 1.3 shows the conventional retrieval image information retrieval using image search. This can be done by downloading image, using URL etc. Merely making use of the single image input is not sufficient to search the required intent hence, this new approach explores the contextual new sub-image information to understand the search intent better with two key steps: Augmented sub-image search results in re-ranking using context of the sub-image, and this expects to have better search results. Beyond contextual Web search, the context in this case is much richer and includes images rather than texts. This image search with sub-image input allow users to select an image as a search query from Web pages or from other documents. Experiments shows the proposed schemes make the sub-image search very much convenient and the quick search results are more expected and relevant for any user. The proposed query-

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 8, August 2016

specific sub-image significantly improves both the efficiency and accuracy of image re-ranking and information retrieval.

II. RELATED WORK

Image search has now-a-days become an increasingly important research topic in order to facilitate access to the information on rapidly growing collections of images over the Web and maximize the benefit for the users. People use their personal websites between friends using message service, or they share via social networks like Facebook, Tweet, Flickr etc., Google, Bing or Yahoo are used to search Internet through images, text input etc., Now-a-days, many interesting researches are coming for the searches related to search using images. Regarding this many different researches have been introduced to retrieve images.

Personalized Image search is the one which search for the particular images of any User intention on the Web. For searching images, any user might provide the necessary and apt query terms like keyword, image file, and then the system can search for the images similar to the query. The similarity used for search criteria can be Meta tags, distribution of color in images, region/shapes attributes, etc. Web image search engines namely Google searches for images rely on the surrounding text features. It is cumbersome and complicated for any web based image search engines to interpret the search of the user's intention only by giving query comprise of keywords. This leads to incorporate noise and high ambiguity in search of results which are extremely not necessary in the context of the users. It became mandatory for using visual information to solve the problem of ambiguity in the text-based image retrieval process. In case of Google search, text box search will auto complete when the user is typing similar added keywords. This method differs from the user intention while searching. So in order to avoid this kind of difficulties, it is mandatory to use visual information to solve the uncertainty in text-based retrieval of the image. To retrieve exact matching, and acquire user's intention we can allow them text query with extended or related images as a suggestion. The context-or text based retrieval of image is done by Purves et al., 2010[1]. Since now no search for image is globally accepted for its effectiveness. Context based image retrieval by directly accessing the color value and the corresponding intensities of an image, texture and shape information is used by Enser, 2000[2]. Enser et al., 2007[3] operates on low-level features of an image for some specialized applications. Palacio et al., 2011[3] submitted traditional search engine to provide tools to add the meaning of an image that minimized semantic gap in the context of spatial queries.

Visual Search ranking methods based on unsupervised method are defined as Classification based, clustering based and graph based. The classification based unsupervised image search is done by R.Yan et al., 2003[4]. W.H.Hsu et al., 2006[5] worked on visual search ranking methods based on clustering method. They search the wider re-ranking via information bottleneck principle. X.Tian et al., 2010[6] ranked the visual search based on graph method. He used Bayesian video search re-ranking method. X.Tian et al., 2010[7] also did active re-ranking for the web image search. X.Tian et al., 2011[8] did the research based on learning to judge image search results. B. Geng, et al., 2012[9] worked on unified context based model for web image retrieval. L. Yang and A. Hanjalic 2010[10], L. Yang and Hanjalic 2011[11], L. Yang and Hanjalic 2012[12] did their researches based on supervised re-ranking for web image search, Learning from search engine and human supervision for web image search and prototype based image search re-ranking respectively in the consecutive years.

III. PROPOSED METHOD

Web-Image search engines mainly rely on text features surrounding the image. It is difficult to predict user's intention only with the query they give and this leads to ambiguous, noisy search results from search engines which does not provide satisfactory result. In this work, we provide a novel image search approach. This only requires the user to click with one query image and to narrow down the search by using another sub-image from the database retrieved and the search is re-ranked based on the visual content. The image retrieval using query image is done as follows. (1) The query image of the users search is categorized into predefined adaptive weight categories, which reflect user's search intention. (2) Based on visual content of the image selected by the user using image clustering mechanism. (3) Sub-

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 8, August 2016

Image submission is used as a more relevant image the user is searching for. All these steps are automated from the user without extra effort.

The major components of this work includes

- User Interface
- Analysis of Query
- Selecting the Query
- Web search navigator
- Result of the query

3. Image Content Extraction

Extracting Image information from the descriptions assigned to images, features of image content can be extracted.

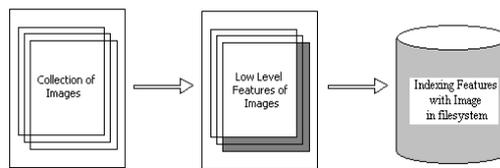


Fig 3.1 Extracting Image Content

Figure 3.1 shows the typical image content or the low-level features extraction. The main idea is to extract the necessary low-level features taken directly from the image itself. These low-level features cover texture, color and shape as well as their spatial as well as spatial-temporal distribution (e.g. Enser 2000[13], Liew and Law 2008[14], Datta et al. 2008[15]). These features can be extracted directly from the pixels of an image. Initially, each and every image of the image collection is gone through an algorithm which extracts low-level features of the image. Then the index is created and all the extracted low-level features are referenced with an identifier referencing the actual image with the low-level features stored within the index. These low-level feature are intended to capture the visual properties of the selected image, either globally or locally and is done for a small group of pixels (Datta et al. 2008, Deselaers et al. 2008[16]). Meiappane et al.,[17] requires the user only to click on one query image with the minimum effort and images from the database retrieved by text-based search are re-ranked based on both visual and textual content.

IV. DATA SET AND EXPERIMENT

Experiments

Here, the effectiveness of the proposed sub-image search is demonstrated by means of an experimental set-up and the publicly available Web Queries is taken as dataset. We used re-ranking methods to narrow down our search and to come up with a best possible result.

Experimental setup

To make the experimental set-up, we initially used Web Queries consists of dataset comprising of large amount of images and the diverse images for search queries. The dataset consists of total of 300 queries. For every query, hundreds of images are retrieved and downloaded with the help of web-search engine that resulted in a total of more than 65000 images. And to illustrate the effectiveness of the proposed method, the sub-image based search baseline from search engine is compared with the supervised and unsupervised re-ranking methods. Supervised approaches are used which include the supervised re-ranking. The unsupervised re-ranking methods include the PRF re-ranking, random walk re-ranking, and Bayesian re-ranking. In the PRF re-ranking, top-ranked images in the initial search result for a given query are used as positive samples and the negative samples are selected from the background images. For Bayesian re-ranking, the best performing local learning consistency and pair-wise ranking distance are used.

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 8, August 2016

Tasks: We first divide the image set into training test and test set by time $T_T = T - (1 \text{ year})$, where T is the end timepoint of the dataset. That is, for each topic, the test set consists of the images in the last one year of the database, and the training set T_B comprises the other images, which are used to learn the image occurrence patterns.

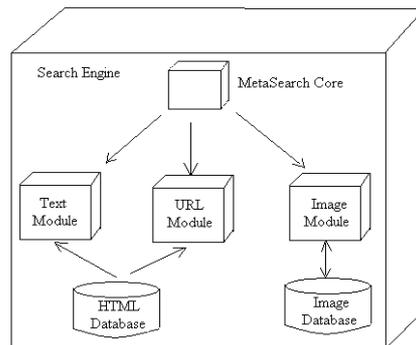


Fig 4.1 image retrieval using meta data

Fig 4.1 shows the existing image search and meta-search core using sub image as an additional source of information. An image is simply a data structure which holds physical attributes also called as primitive or low-level features. The primitive features are color, texture, shape and its spatial and spatial-temporal distribution. Primitive features do not have any inherent meaning in itself. They are nothing but bits representing an image. Its meaning relates to a picture which is a human interpretation of how these features are actually assembled and in what context they were retrieved or created. Hence there is neither a well-defined nor an objectively quantifiable attributes like color or spatial distribution of shapes in the image, but a property assigned by humans through combining objective and subjective knowledge in a socio-cognitive process. Because such a meaning cannot be derived from the primitive features themselves and it is a well-known technique to manually or automatically assign title, descriptions, keywords/tags or other textual descriptions, which then textually represent an image.

QueryidNumber : 60
Total number of documents based on the queries: 40000
Retrieved : 40000
Relevant : 7658
R.d-Retrieval: 6709

Interpolated Recall-Precision Average		
	Precision	Average
Interpolated at	0.0	0.8675
	0.1	0.7564
	0.2	0.6543
	0.3	0.5385
	0.4	0.4987
	0.5	0.4865
	0.6	0.3647
	0.7	0.2576
	0.8	0.2064
	0.9	0.1533
1.0	0.0213	

Average Precision (Non-interpolated)
Averaged over queries 0.4325

Precision after R docs retrieved		
	Docs	
Precision at	5	0.7655
	10	0.7324
	15	0.6834
	20	0.6775
	30	0.6234
	100	0.4958
	200	0.3876
	500	0.2543
1000	0.1435	

Exact : 0.4653

Fig. 4.2 Search engine average precision interpolated and non-interpolated

Fig 4.2 shows search engine's average precision interpolated and non-interpolated result.

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 8, August 2016

Performance evaluation

We compared the proposed method with Bayesian re-ranking and text-based search baseline to demonstrate the effectiveness. Fig. 4.3 shows the overall performance averaged over all the queries. Because the variant of Bayesian re-ranking performs the best among all the seven graph-based ranking methods including Bayesian variants and random walk, we use it as the representative of Bayesian and the graph-based re-ranking methods. From this result, we could see the proposed sub-image method consistently outperforms the text-based search baseline and Bayesian re-ranking at all truncation levels. Specifically, the proposed Subimg obtains about considerable relative improvement compared with the text-based search baseline and the Bayesian ranking, respectively.

In comparison, Bayesian reranking gains only a marginal 4.1% relative improvement over the text-based search baseline method. The results indicates that the proposed sub-image learning-to-rerank method can learn a good ranking function and that the ranking model can be generalized across a broad range of queries. Moreover, the proposed sub-img method is considerably good method, since it requires less computational cost than the graph-based reranking methods that also includes Bayesian reranking, which requires computation of iterative nature. The performance improvement with the added advantage of less computational cost clearly indicates that the sub-img method is a promising paradigm for the Web image search. This conclusion is supported by the performance comparison on all three search engines namely Yahoo and Google. We could also see that for all the search engines the sub-img method improves the performance over baseline search.

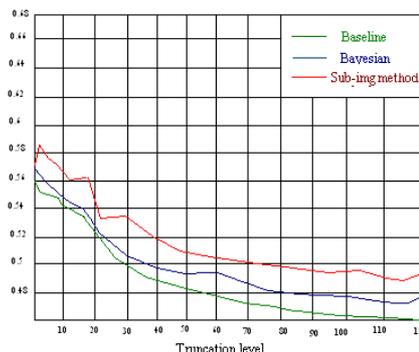


Fig. 4.3 Page No.36 Phd Thesis Yang.pdf

From the fig 4.3 we could see moreover, the proposed method performs consistently better than the Bayesian ranking method on all the search engines. It can be observed that the sub-img method promotes highly relevant images (the images marked by the red rectangle). The Bayesian reranking has the inability over the sub-img method to do the same by the reasons mentioned here. Bayesian reranking mainly relies on the initial ranking, while in sub-img, the initial ranking is regarded as one of features, the weights of which can be adjusted to the use cases automatically. Second, we use mainly the visual neighborhood structure instead of visual similarity. This increases the problems introduced by the imperfection of visual similarity estimation which tend to make the existing methods unpredictable in many cases. For instance, the promotion of the image marked by blue rectangle by Bayesian reranking attributes relates to the deficiency of visual similarity estimation.

We also performed a statistical significance test to verify whether the improvement of the sub-img method is statistically significant. The p-values of the t-test of learning-to-rerank over the text-based search baseline as well as the Bayesian reranking method in terms of NDCG@40 and NDCG@100 are done. They are computed by modeling the performance improvement for each query taking NDCG difference between two methods as a t-test. From this result we could see that the improvement of the sub-img method is statistically significant over the rest. We can see that on the dataset the performance is boosted after applying the proposed sub-img method to rerank the image search results. The sub-img method improves the performance over baseline by 8.5% and over Bayesian reranking by 5.1% in terms of NDCG@100.

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 8, August 2016

V. CONCLUSION

Fetching the information through query based on the image leads to the image search in Web Mining. Based on classifiers for all pre-defined attributes, each and every image is represented by an attribute feature which consists of the responses from the classifiers. In the paper, we presented a novel sub-image submission for the image search. Hence in addition to the image submitted first time, the web frame work will automatically ask for the contextual based sub-image submission thereby narrow down the search. The proposed sub-image submission significantly improves both the accuracy as well as the efficiency of the information retrieval through image search.

From this work, we see many possibilities to further innovate, explore and also to extend the learning through various paradigm. The features of unsupervised ranking methods could be employed to reconstruct the ranking features. However, in some cases, the response time may not be critical and for those cases, we could use fusion of multiple unsupervised ranking features in the learning process for a better result. Our experimental results reported the re-ranking feature selection and improved the performance of the learned re-ranking model. This proposed work if deployed more and more Web search commercial engines will be benefitted and it will increase the potential usage of mobile search scenario also.

REFERENCES

- [1] Purves, R.P., Edwards, A., Fan, X., Hall, M., Tomko, M., Automatically generating keywords for georeferenced images. In: *Proceedings of the international conference on Multimedia information retrieval*, ACM Press, 2010.
- [2] Enser, P., Visual image retrieval: seeking the alliance of concept-based and content-based paradigms. *Journal of Information Science*, 26(4), 199–210. doi:10.1177/016555150002600401, 2000.
- [3] Palacio, D., Cabanac, G., Sallaberry, C., Hubert, G., On the evaluation of Geographic Information Retrieval systems. *International Journal on Digital Libraries*, 11(2), 91–109, 2011.
- [4] R. Yan, A. G. Hauptmann, and R. Jin. Multimedia search with pseudorelevance feedback. In CIVR, 2003.
- [5] W. H. Hsu, L. S. Kennedy, and S.-F. Chang. Video search reranking via information bottleneck principle. In ACM Multimedia, 2006.
- [6] X. Tian, L. Yang, J. Wang, Y. Yang, X. Wu, and X.-S. Hua. Bayesian videosearch reranking. In ACM Multimedia, 2008.
- [7] X. Tian, D. Tao, X.-S. Hua, and X. Wu. Active reranking for web imagesearch. *IEEE Trans. Img. Proc.*, 19:805–820, March 2010.
- [8] X. Tian, Y. Lu, L. Yang, and Q. Tian. Learning to judge image search results. In ACM Multimedia, 2011.
- [9] L. Yang, B. Geng, A. Hanjalic, and X.-S. Hua. A unified context model for web image retrieval. *ACM Transactions on Multimedia Computing, Communications and Applications*, 8(3):28, 2012.
- [10] L. Yang and A. Hanjalic. Supervised reranking for web image search. In ACM Multimedia, pages 183–192, 2010.
- [11] L. Yang and A. Hanjalic. Learning from search engine and human supervision for web image search. In ACM Multimedia, 2011.
- [12] L. Yang and A. Hanjalic. Prototype-based image search reranking. *IEEE Transactions on Multimedia*, 14(3-2):871–882, 2012.
- [13] Enser, P., Visual image retrieval: seeking the alliance of concept-based and content-based paradigms. *Journal of Information Science*, 26(4), 199–210. doi:10.1177/016555150002600401, 2000.
- [14] Liew, A.W.-C., Law, N.-F., Content-Based Image Retrieval, In: Khosrow-Pour, M. (Ed.) *Encyclopedia of Information Science and Technology*, Second Edition, IGI Global, pp. 744–749, 2008.
- [15] Datta, R., Joshi, D., Li, J., Wang, J.Z., Image retrieval: Ideas, Influences, and Trends of the New Age. *ACM Computing Surveys*, 40(2), 1–60. doi:10.1145/1348246.1348248, 2008.
- [16] Deselaers, T., Keysers, D., Ney, H., Features for image retrieval: an experimental comparison. *Information Retrieval*, 11(2), 77–107. doi:10.1007/s10791-007-9039-3, 2008.
- [17] Meiappane, S. Monesh, S. Pradeep Kumar and U. Murugan. One Click Image Search Re-Ranking Based On User Preference *International Research Journal of Engineering and Technology*, 3(3): 226 –231