

Retrieving Biomedical Images through Content-Based Learning from Examples Using Fine Granularity

Hao Jiang^{*,a}, Songhua Xu^{*,b,1}, Francis C.M. Lau^a

^aDepartment of Computer Science, The University of Hong Kong,
Pokfulam Road, Hong Kong, China

^bBiomedical Science and Engineering Center, Oak Ridge National Laboratory,
Oak Ridge, Tennessee 37831, USA

ABSTRACT

Traditional content-based image retrieval methods based on learning from examples analyze and attempt to understand high-level semantics of an image as a whole. They typically apply certain case-based reasoning technique to interpret and retrieve images through measuring the semantic similarity or relatedness between example images and search candidate images. The drawback of such a traditional content-based image retrieval paradigm is that the summation of imagery contents in an image tends to lead to tremendous variation from image to image. Hence, semantically related images may only exhibit a small pocket of common elements, if at all. Such variability in image visual composition poses great challenges to content-based image retrieval methods that operate at the granularity of entire images. In this study, we explore a new content-based image retrieval algorithm that mines visual patterns of finer granularities inside a whole image to identify visual instances which can more reliably and generically represent a given search concept. We performed preliminary experiments to validate our new idea for content-based image retrieval and obtained very encouraging results.

Keywords: Biomedical image search, content-based image retrieval, learning from online examples.

1. INTRODUCTION

We introduce a new content-based image search algorithm for retrieving biomedical images based on learning from examples. Compared with traditional content-based image retrieval algorithms, our new method performs its visual concept learning on a finer granularity than that of entire images. Our algorithm is capable of finding images in an unannotated image collection through inferring the relevance of an image with respect to a given text query on the granularity of visual patch sets. The relatedness between an image search text query and an unannotated image is derived through referring to example images acquired from online image search on this fine granularity. Using these example images harvested from the Web, we can estimate an unannotated image's relevance to a text query through the visual translation performed by individual visual patch sets. To cope with noise in the online image search results, we further leverage the ranking scores associated with the online reference images to more precisely represent the user's image search intent using the noisy set of sample reference images. According to the estimated image-to-text-query relevance,

*: To be considered as equal first authors.

¹: Contact him at Oak Ridge National Laboratory, Oak Ridge, TN 37831-6173, USA, Email: xus1@ornl.gov.

Notice: This manuscript has been authored by UT-Battelle, LLC, under Contract No. DE-AC05-00OR22725 with the U.S. Department of Energy. The United States Government retains and the publisher, by accepting the article for publication, acknowledges that the United States Government retains a non-exclusive, paid-up, irrevocable, world-wide license to publish or reproduce the published form of this manuscript, or allow others to do so, for United States Government purposes.

for any text in the image search query, our algorithm can search across an unannotated personal photo collection to find semantically most related images.

In the remaining of this paper, we first take a brief look at the work most related to our study. We introduce our new algorithm for inferring image-to-text-query relevance for image retrieval and ranking in Sec. 3. In Sec. 4, we discuss a prototype system that implements our algorithm along with the experiment results obtained using the prototype system and we conclude this paper in Sec. 5.

2. RELATED WORK

2.1 Image Re-Ranking

Our work is situated in the broad area of image search studies, which is one of the most heated topics in today's computing technology research. Abundant commercial image search engines, such as Google Image Search, Yahoo! Image Search, and PicSearch (<http://www.picsearch.com>), as well as research prototype systems, e.g. WebSeek [1] and IGroup [2], are available today that rank or re-rank images upon a user submitted query by matching meanings of candidate images with the automatically parsed intent of a user query. Among the general field of image search studies, our work in this paper is most closely related to the design of image re-ranking algorithms.

One of the most closely related work to our study here is Lin et al.'s algorithm on re-ranking web image search results using a probabilistic image re-ranking model [3]. In comparison with their method, our image re-ranking algorithm as described in this paper estimates the image-to-query relevance through both analyzing the text information associated with an image and exploring the visual content of an image. Cai et al. [4] proposed a hierarchic image clustering method through jointly analyzing visual, textual, and link information of an image. However, their study is proposed for clustering images, which cannot be directly applied for retrieving or re-ranking images. Also related to our work is Zhou and Huang's image retrieval algorithm [5], which jointly uses the keyword and the visual content of an image for image retrieval. Recently, Jing and Baluja [6] adapted Google's PageRank algorithm for product image search through modifying the conventional Google PageRank algorithm to suit the image domain. Their approach is based on image similarities estimated from low-level visual features proposed in [7].

2.2 Content-Based Image Retrieval

Our research also belongs to content-based image retrieval studies. Plenty of commercial tools and prototype systems are available today for content-based image retrieval, e.g. Gupta and Jain's Virage system [8], and Wang et al.'s SIMPLIcity system [9], just to name a few. Liu et al. [10] offered a high-level overview on content-based image retrieval research from the semantics understanding's perspective. Adrian et al. [11] proposed a content-based image retrieval method following an ontology driven approach, which can significantly improve image retrieval precision. Kidambi and Narayanan [12] introduced a human computer integrated approach for content-based image retrieval, which aims at improving the accuracy of image retrieval using user feedback. Their approach is especially useful when automatic methods fail to interpret the semantics of an image. Hsiao et al. [13] represented an image as a set of visual bag-of-words. Given their image representation, they then applied a text analysis procedure followed by a relevance feedback scheme to understand image content for image retrieval. People have also compiled surveys on domain specific content-based image retrieval methods and systems, e.g. Muller et al.'s survey on content-based image retrieval for medical applications [14].

3. METHOD

Given a user-given textual query for searching in an image database, our method first establishes a reference image set from online image examples. We then extract the visual features of each image in the database and estimate each visual

feature's relevance to the textual query according to the visual examples provided by the reference image set. After that we can return a rank list of images to the user, which are ranked according to the images' relevance to the query.

3.1 Establishing Reference Image Set from Online Examples

To establish the reference image set for a given query, Q , we consider both visual features in the image and its accompanying text. The query Q is first forwarded to Google Search and Google Image Search respectively. The first three hundred non-duplicating webpages and images are saved respectively. All these images will form the reference image set, $G(Q)$. In the image search phase, we also save the text content of a source webpage that contains a reference image in $G(Q)$. Figure 1 shows the workflow to establish the reference image set for a given textual query.

3.2 Estimating Relevance to Query of Visual Patches

In our method, visual patches of an image are defined based on Speeded-Up Robust Features (SURF). With these visual patches, we further introduce the concept of visual patch sets. A visual patch set is defined as a set of visual patches detected from an image whose circular regions overlap. Two visual patch sets are identical if they contain the same set of constituent visual patches. Any visual patch set that occurs more than five times in the example image set will be considered a frequent visual patch set. Let $\phi_i(Q)$ denote a frequent visual patch set that is detected from the example image set $G(Q)$. All the frequently occurring visual patch sets detected from the example image set $G(Q)$ are organized as a collection, which is denoted as $\Phi(Q) = \{\phi_1(Q), \phi_2(Q), \dots, \phi_{z(Q)}(Q)\}$, assuming there are $z(Q)$ frequent visual patch sets contained in the example image set $G(Q)$.

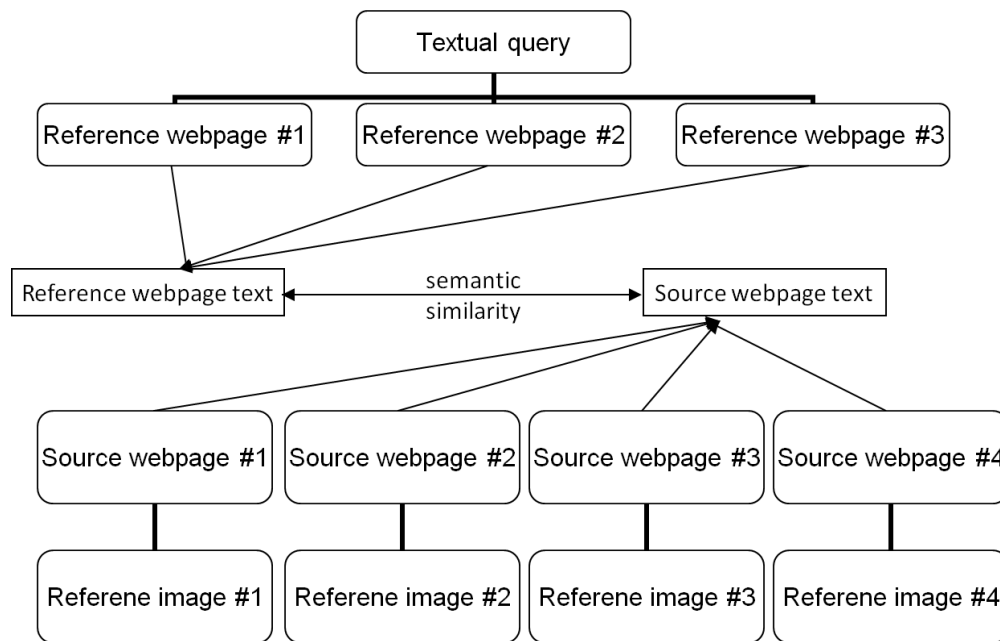


Figure 1: The computing workflow to establish the reference image set for a given textual query.

For each visual patch set ϕ_k , we need to estimate its relevance to an image search query Q . Denote the relevance to the query Q of the example image G_i by $r_i(Q)$. Then, for each visual patch set ϕ_k , we derive its relevance to the query Q as follows.

First we estimate the salience of each visual patch set contained in an example image G_i . To this end, we borrow the traditional Term Frequency–Inverse Document Frequency (TF–IDF) scheme in the text processing domain [17]. Let $m_j(\phi_k)$ be the number of times the visual patch set ϕ_k occurs in the j -th example image $G_j(Q)$. Adopting the standard form

of the TF term definition [17], we define the Term Frequency (TF) for the visual patch set φ_k as $(1 + \ln m_j(\varphi_k))$. Let $y(\varphi_k)$ be the number of images in the example image collection $G(Q)$ that contain the visual patch set φ_k and $\zeta(Q)$ be the total number of images in the example image collection. Again following the conventional IDF term definition, we define the Inverse Document Frequency (IDF) term for φ_k as $\ln(1 + \zeta(Q)/y(\varphi_k))$. Overall, we define the TF-IDF term for the visual patch set φ_k as $(1 + \ln m_j(\varphi_k)) \ln(1 + \zeta(Q)/y(\varphi_k))$. Then we calculate the relative salience of the visual patch set φ_k among all the visual patch set contained in the reference image G_j as follows:

$$\rho(\varphi_k, G_j) = \frac{(1 + \ln m_j(\varphi_k)) \ln(1 + \frac{\zeta(Q)}{y(\varphi_k)})}{\sum_{\varphi_i \in \Phi(Q)} (1 + \ln m_j(\varphi_i)) \ln(1 + \frac{\zeta(Q)}{y(\varphi_i)})}$$

Second, according to the relative salience of each visual patch set contained in G_j , we then proportionally attribute G_j 's relevance score to the query Q to individual visual patch sets present in the image. That is, the fraction of query relevance score allocated for the visual patch set φ_k is $r_j(Q)\rho(\varphi_k, G_j)$. Recall that $r_j(Q)$ is the j -th example image G_j 's relevance score to the query Q . Finally, we derive φ_k 's overall relevance score to the query, denoted as $\theta(\varphi_k, Q)$, by summing up the fractional relevance scores φ_k acquires from all the example images, i.e.:

$$\theta(\varphi_k, Q) = \sum_{j=1}^{\zeta(Q)} r_j(Q)\rho(\varphi_k, G_j)$$

3.3 Estimating Relevance to Query of Unannotated Images

With every visual patch set's query relevance $\theta(\varphi_k, Q)$ estimated through the above equation, we can now estimate an unannotated image I_x 's image-to-query relevance through summing the query relevance of all the visual patch sets carried in I_x . Finally, given an input query Q , each search candidate image I_x will be assigned a query relevance score as estimated above. Those candidate images whose estimated query relevance scores are above a certain threshold will be returned as image search results for the query. All the search result images are also ranked in a descendent order according to each image's estimated query relevance score.

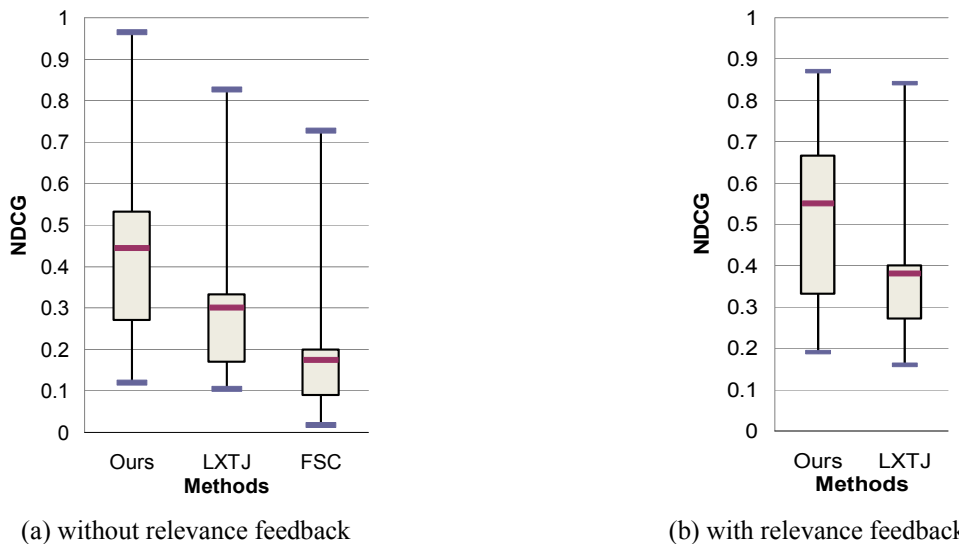


Figure 2: NDCG scores of our image search experiments: (a) without relevance feedback; (b) with relevance feedback from users. We illustrate the key statistics of all the NDCG scores using boxplots.

4. EXPERIMENT RESULTS

We evaluate the image search and ranking quality by adopting the metric of normalized discounted cumulative gain (NDCG), which is a standard measure for ranking quality assessments [17]. Eight graduate students were invited to participate in our experiments. For each query result, all participating users were asked to assign a numerical label to each of the top 50 search result images to offer their respective opinions on these images' relevance to the corresponding queries. The numeric label, in the range of [0,1], represents the image's relevance to the query (1: “fully relevant”; 0: “fully irrelevant”). By averaging all the subjects' opinions regarding these images' query relevances, we compute the NDCG score for each query result respectively.

The image database used in our experiments consists of 6787 images, where around 80% of them are medical images while the rest are images outside the biomedical domain; these latter images were randomly acquired from Picasa and Flickr to serve as negative samples or noise. We compared the measured performance of our algorithm with two state-of-the-art content-based image retrieval algorithms which we refer to as “LXTJ” [16] and “FSC” [15] in the following discussion. All our experiments were performed on a Windows XP desktop computer with a 2.66GHz Intel Core 2 Duo CPU and 2GB main memory.





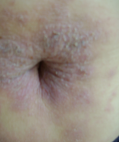














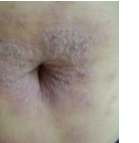








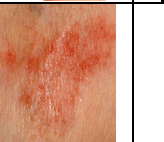
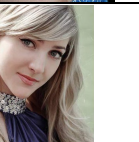
Methods	Top ten image search results				
LXTJ					
					
FSC					
					
Ours					
					

Figure 3: Top ten image search results for the query keyword “dermatitis” by three algorithms.

We asked every subject in our study to perform 30 medical-related image searches. For each subject, his or her 30 image search queries were freely chosen by the subject himself or herself. For each selected image search keyword, we run all three search algorithms against our image database. After that, we evaluate the NDCG score for each query performed by each algorithm respectively. In Figure 2(a), we report the NDCG scores obtained by the three algorithms respectively over all image querying sessions. All these results are generated fully automatically, i.e. no user feedback is provided for the algorithms. Figure 2(b) reports the respective NDCG scores of image search experiments performed over our image dataset by our algorithm and LXTJ with a three-round user feedback. From these results under both experimental settings, we can clearly see that our algorithm consistently outperforms the other two algorithms. Figure 3 shows the top ten search results of an example query by these three algorithms, from which we can see our new algorithm performs noticeably superiorly to the two leading content-based image retrieval algorithms for retrieving biomedical images.

In our new content-based image retrieval method introduced above for searching unannotated biomedical images, the key novelty lies in its decomposition of every example image into a set of visual patches. Based on this decomposition, we then extract frequently occurring visual patch sets that commonly exist between multiple example images. These visual patch sets serve as reliable instances visually embodying the intent of a given image query, which is not always clearly expressed by the query text itself. Once we have identified these low-level visual instances, our algorithm is able to perform content-based image retrieval through learning from examples on the fine granularity of visual patch sets rather than on the entire images. Such a difference in our algorithm design as compared with traditional methods empowers our method to be capable of retrieving biomedical images more reliably and comprehensively.

5. CONCLUSION

This study introduces a new content-based image search algorithm that performs image content analysis and understanding at finer granularities of the visual patch sets rather than the entire images. The algorithm first decomposes example images into small sets of visual patches, and then mines frequently occurring visual patch sets commonly exhibited across multiple example images. The purpose of the second step is to identify reliable and generically applicable visual instances, in the form of frequently occurring common visual patch sets, which express the target image retrieval intent in a more reusable way. Experiment results successfully validate the effectiveness of our new algorithm for content-based image retrieval in the biomedical domain.

ACKNOWLEDGEMENT

Songhua Xu performed this research as a Eugene P. Wigner Fellow and staff member at the Oak Ridge National Laboratory, managed by UT-Battelle, LLC, for the U.S. Department of Energy under Contract DE-AC05-00OR22725. The authors thank Dr. Vladimir A. Protopopescu for his helpful comments on the manuscript.

REFERENCES

- [1] SMITH, J. and CHANG, S.-F. 1997. Visually searching the web for content. *IEEE Multimedia* 4, 3, 12–20.
- [2] JING, F., WANG, C., YAO, Y., DENG, K., ZHANG, L., and MA, W.-Y. 2006. IGroup: a web image search engine with semantic clustering of search results. In *MULTIMEDIA '06: Proc. of the 14th Annual ACM International Conference on Multimedia*. ACM, New York, NY, USA, 497–498.
- [3] LIN, W.-H., JIN, R., and HAUPTMANN, A. 2003. Web image retrieval re-ranking with relevance model. *Proc. of IEEE/WIC International Conference on Web Intelligence*, 242–248.
- [4] CAI, D., HE, X., LI, Z., MA, W.-Y., and WEN, J.-R. 2004. Hierarchical clustering of www image search results using visual, textual and link information. In *MULTIMEDIA '04: Proceedings of the 12th annual ACM international conference on Multimedia*. ACM, New York, NY, USA, 952–959.

- [5] ZHOU, X. S. and HUANG, T. 2002. Unifying keywords and visual contents in image retrieval. *IEEE Multimedia* 9, 2, 23–33.
- [6] JING, Y. and BALUJA, S. 2008. Pagerank for product image search. In *WWW '08: Proceedings of the 17th international conference on World Wide Web*. ACM, New York, NY, USA, 307–316.
- [7] LOWE, D. G. 2003. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision* 20, 91–110.
- [8] GUPTA, A. and JAIN, R. 1997. Visual information retrieval. *Communications of the ACM* 40, 5, 70–79.
- [9] WANG, J. Z., LI, J., and WIEDERHOLD, G. 2000. Simplicity: Semantics-sensitive integrated matching for picture libraries. In *VISUAL '00: Proceedings of the 4th International Conference on Advances in Visual Information Systems*. Springer-Verlag, London, UK, 360–371.
- [10] LIU, Y., ZHANG, D., LU, G., and MA, W.-Y. 2007. A survey of content-based image retrieval with high-level semantics. *Pattern Recognition* 40, 1, 262–282.
- [11] ADRIAN, P., CHRISTOPHE, M., and PIERRE-ALAIN, M. 2007. Ontology driven content based image retrieval. In *CIVR '07: Proceedings of the 6th ACM international conference on Image and video retrieval*. ACM, New York, NY, USA, 387–394.
- [12] KIDAMBI, P. and NARAYANAN, S. 2008. A human computer integrated approach for content based image retrieval. In *ICCOMP'08: Proceedings of the 12th WSEAS international conference on computers*. World Scientific and Engineering Academy and Society (WSEAS), Stevens Point, Wisconsin, USA, 691–696.
- [13] HSIAO, J.-H., CHEN, C.-S., and CHEN, M.-S. 2008. A novel language-model-based approach for image object mining and re-ranking. In *ICDM '08: Proceedings of the Eighth IEEE International Conference on Data Mining*. IEEE Computer Society, Washington, DC, USA, 243–252.
- [14] MULLER, H., MICHOUX, N., BANDON, D., and GEISSBUHLER, A. 2004. A review of content-based image retrieval systems in medical applications—clinical benefits and future directions. *International Journal of Medical Informatics* 73, 1, 1–23.
- [15] FENG, H., SHI, R., and CHUA, T.-S. 2004. A bootstrapping framework for annotating and retrieving www images. In *MULTIMEDIA '04: Proceedings of the 12th Annual ACM International Conference on Multimedia*. ACM, New York, NY, USA, 960–967.
- [16] LIU, Y., XU, D., TSANG, I.W., and LUO, J. 2011. Textual query of personal photos facilitated by large-scale web data. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 33:1022–1036.
- [17] SANGER, J., and FELDMAN, R., *The Text Mining handbook: advanced approaches in analyzing unstructured data*, Cambridge University Press, 2006.