SHOW AND TELL: A Seamlessly Integrated Tool For Searching with Image Content And Text

Zhiyong Zhang Carlos Rojas Olfa Nasraoui
Hichem Frigui
Department of Computer Engineering and Computer Science
University of Louisville
Speed School of Engineering, Louisville, KY 40292
{olfa.nasraoui}@louisville.edu

Abstract

In this paper, an image search tool that combines keyword and image content feature *querying* and *search* is presented. The developed search tool tries to bridge the gap between *commercial search engines*, and *CBIR* (Content Based Image Retrieval) systems developed mostly in the academic field. The tool is implemented by building on and extending the open source text-based search engine *Nutch* and its powerful *Lucene* based crawling and indexing capabilities. Several user friendly search options are provided to allow users to query the index using not only words, but also by *showing* an image example, as well as image feature descriptions. The search system applies a variety of *image* and *text* mining principles to allow indexing both the *content* and *textual* annotations in a homogenous manner, and uses clustering to organize the search results and help users navigate through large collections of images on various websites. The proposed system promises to help users such as scientists, educators and students search through a variety of images, while taking advantage of available text and annotations when available. The *modular and open source* framework promises an easy extension of the system to *special* collections and more specialized scientific image database retrieval tasks by the simple addition of *plugins* that are non-invasive with respect to the indexing and search engine. The plugins could for example extract rich information from an image such as the presence of certain events such as "*flare*", "*coronal loop*" or "*sun spot*".

1 introduction

Current large scale image search engines like Google , Yahoo, etc. base their image search capabilities mainly on text features associated to pictures. The use of keywords may be a very effective way to search for images if every image is described exhaustively and the image-searcher has a clear idea of which descriptors to use. However, keywords that are associated to images can be incomplete. Moreover, the user may not have a clear description of what he/she is looking for in words. For all these reasons, the inclusion of image features (color, texture, shape, ..., etc) seems like a reasonable way to improve the image search results. On the other hand, current research in CBIR (Content Based Image Retrieval) such as *Blobworld* [29], [7], [5], *FIDS*, *SIMBA*, *WEBSEEk*, and *Simplicity* [33], [18], ..., etc, have illustrated the merits of exploiting low level image features for image retrieval. However, on *large scale Web resources*, most CBIR systems' performance may become questionable from a scalability point of view. Moreover, they generally do not provide keyword based search functionalities which may be useful in certain search scenarios. As can be seen from Table 1, the two methods (*search engine and CBIR*) have the potential to complement each other. If combined together into one search tool, then this tool could tap on the combined strengths of a wealth of well-developed text retrieval techniques and the robustness and scalability of current Search Engine technologies to handle image retrieval. A combined tool also promises to alleviate the weaknesses of either type of search working in isolation. In this paper, we present a seamless

Table 1: Comparison of Keyword and Content Based Search

Type	Framework	Weakness	Strength
Keyword-Based	Search Engines	Information Expression Difficulties	Well-Developed Text Retrieval Methods;Scalability;Easy to Use
Content-Based	CBIR	Semantic Gap and Scalability	Ability to express the user's information need using content

Table 2: Comparison of Different Image Search Systems

System Name	Indexing		Querying			System Name	Indexing		Querying		
	textual	content	text	img example	img feature		textual	content	text	img example	img feature
Google	√		√			WebSeer	√	√	√		
QBIC		√		√	√	FIRE	√	√		√	
BlobWorld		√		√		Webface	√	√		√	
SIMPLicity		√		√		*"Show and Tell"*	√	√	√	√	√

combination of the two approaches in an integrated image search system, which can be accessed through: *http://webmining.spd.louisville.edu:8080*.

2 Related Work and Contributions

In addition to the above mentioned CBIR systems, some other systems include *QBIC* [12], *WALRUS* [22], *CIRES* [15], *C-BIRD* [20], and [30] [19] [13]. All these CBIR systems are based on image *content* features such as color, texture, shape, structure, ..., etc. They seldom take any textual information or image annotations into consideration. Some image retrieval systems like *WebSeer* [31], *FIRE* [11], *Webfaces* [1] and [8] [6] have recently started considering their image source from Web collections and textual information, but the way they use textual information is mostly for building internal relationships or models between a set of training images and corresponding textual information by using machine learning, classification, clustering, or Latent Semantic Indexing. The textual information in these systems remains hidden to users; that is, users are not allowed to query the system using these textual keywords. Hence, they share limitations that are similar to the above mentioned CBIR systems when it comes to the user interface. Users in these systems are only allowed to query by submitting an example query image [16], or by progressively browsing a set of images [35]. For a comparison, see Table 2.

From Table 2, we can see that the earliest systems focused either on the textual part or the image content part, but always *exclusively* one or the other. For other systems that combine text and content together, the final *interface* for a user's query turns out to be either content based or keyword based, where one feature is *exclusive* of the other. In contrast, the proposed system, that we named *"Show and Tell"*, is distinguished by the following contributions: (1) We build an image search engine based on modifying the *Lucene* based open source search engine "nutch". In this way, we extend nutch's powerful capability of indexing and searching to image retrieval. (2) We encode the *low level image features* into *text* and then combine the color features with textual annotations (or tags) of the image to yield seamlessly integrated and indexable representations of the images in a text-like domain. (3) We not only provide the query *by keyword* capability but also provide the *query by example image* compatibility and most importantly, we provide the *query by image feature* (currently illustrated by color features) capability. (4) Our real-time clustering capability can help the user locate the desired images more quickly.

3 Indexing Images -Combining Colorwords and Annotations

Current indexing and retrieval techniques for text documents, such as Web pages, pdf documents, etc, are nearly full-fledged. Inspired by this framework, we decided to encode the image features into *text-like keywords*, and therefore tap on the strengths of the *storing*, *indexing* and *retrieval* techniques used for processing text documents. We then build the index based on the keywords encoded from the image features. We name the encoded color feature keywords "*colorwords*". Just like a web page or pdf document may contain many keywords, an image may contain many colorwords. The ranking of resulting documents is usually based on the TF-IDF method. Similarly, if an image contains a high frequency of a specific colorword, and a user searches for that specific colorword, then the ranking for that image would be higher. On the other hand, if a colorword, is contained in too many images, then it will not be a good descriptor for searching images. After getting the image histogram, we then quantize them to get the corresponding colorwords

The image *textual* information or annotations are extracted from World Wide Web image collections by

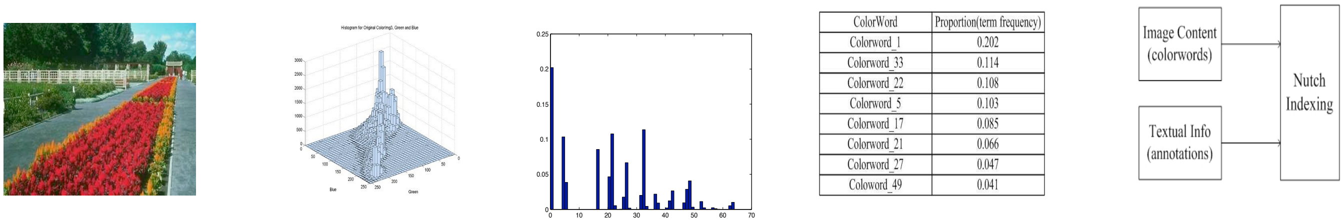

Figure 1: Indexing Images:from left to right: original image, color histogram 2-D projection, color histogram on 64 bins, *colorwords* and their corresponding frequencies (*TF*), and finally indexing colorwords and textual annotations

parsing the *Image URL* and *Image Anchor Text* (text that describes the *link* to an image). Generally, the image URL can contain important information about the image category and image name information, which provide good image annotations, while the Anchor text can provide a description of the link to the image from inside a Web page. Certainly these sources may not be perfect. We left for future work, the extraction of annotations from "*ALT*" tags, more text located in the vicinity of the anchor text, as well as other tags that may be available when images are embedded in XML documents.

Figure 1 illustrates how the information obtained after parsing an image's contents, i.e. its *colorwords*, and parsing the image *textual* annotations from the parent HTML file, is fed to *Nutch* for building an inverted searchable index.

4 Querying the Searchable Index

After the colorwords have been extracted from the image and indexed, we correlate the colorwords with human beings' linguistic color descriptions such as *red, blue, brown*, ..., etc. We use an *image color palette* containing 12 colors as shown in Table 4. Notice that the color definitions are not categorical or clear-cut.

Table 3: Image Color Palette

Red	Green	Blue	Cyan	Magenta	Yellow	Brown	Orange	Violet	Gray	Black	White

Each of the above colors forms a group of several similar colorwords and each colorword may belong to multiple groups. Here the traditional information retrieval technique *TF-IDF* plays a reincarnated role of matching and ranking according to *textual* and *image content* (except that *image content* has been mapped to *feature words*).

For the case of query-by image example, the system should return results that are ranked by similarity with the query images. Besides the widely used Euclidean distance and Quadratic distance, some other distance measure like Cosine angle distance [23], Earth Mover's Distance, [28] and other distance measures have been explored in several systems [27][24] [32] [34] [21].We have used Cosine and Euclidean distance successfully for search. We have also added realtime clustering functionalities to the search interface. Clustering was used in [36], [2], [3] and [9] for different purposes. We use it to obtain a more user friendly interface. The realtime clustering is helpful to quickly locate the images that a user is seeking. This will play the role of *zooming* in search. We use K-means to cluster the image color features. Choosing the number of clusters is beyond the scope of this paper. Hence we chose $k = \sqrt{N}$, for N result images.

As mentioned above,we have implemented our system by using Nutch through crawling *Washington* database, *Bigfoto*, and the SIMPLICity annotated database that has been stored on our server for experimentation. Table 4 lists he syntax and examples of six different querying modes implemented so far, including querying by text keywords, image features (currently color), and image example, as well as Boolean combinations thereof. For some snapshots of query examples, see Figure 2.

Table 4: Query Syntax and Examples. Note: by default, Boolean "AND" is used

Query Type	Examples	Query Type	Examples
Keyword	*roses*	+(-)Keyword +(-)Image Color	*roses -imgcolor:red*
Image Color	*imgcolor:red*	+(-)Keyword +Image Example	*roses imgURL:http://xxx.jpg*
Image Example	*imgURL:http://xxx.jpg*	+(-)Image Color +Image Example	*imgcolor:red imgURL:http://xxx.jpg*

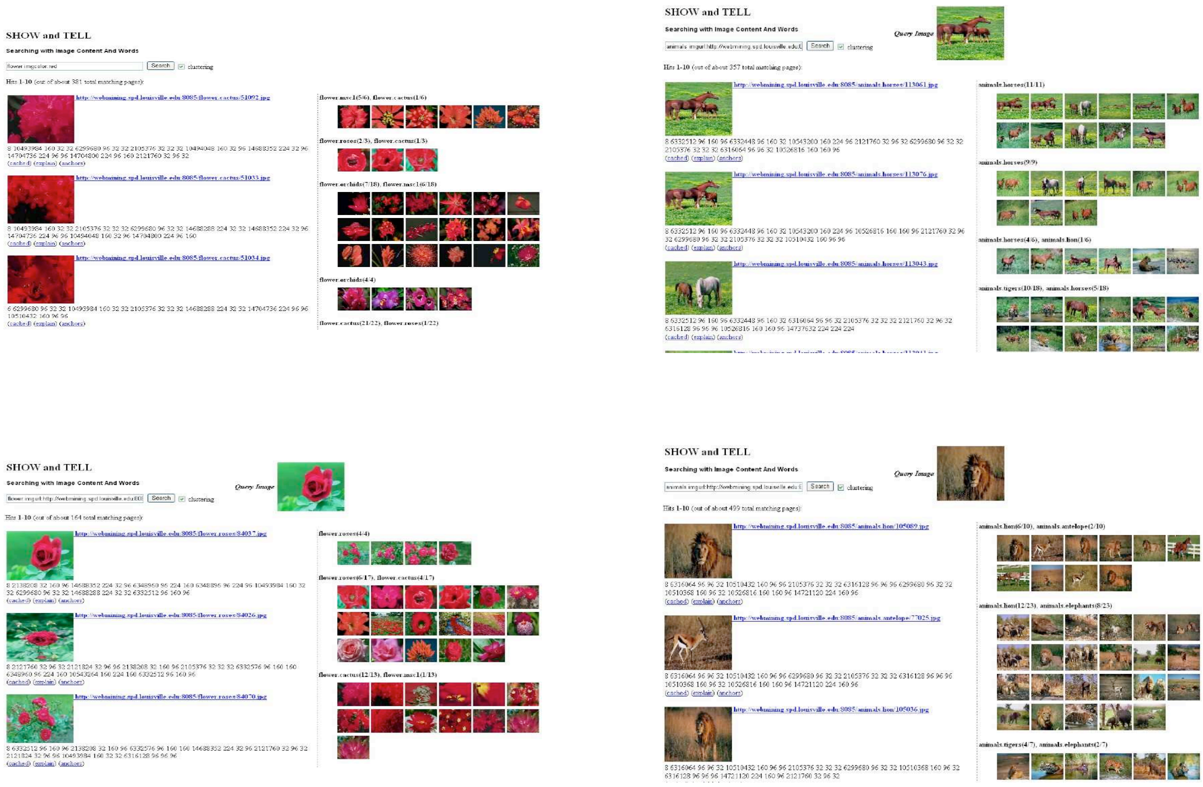

Figure 2: Query Example Snapshots
Queries Clockwise From Topleft:
flower imgcolor:red;
animals imgURL:http:webmining.spd.louisville.edu:8085/animals.horses/113061.jpg;
animals imgURL:http://webmining.spd.louisville.edu:8085/animals.lion/105089.jpg;
flower imgURL: http://webmining.spd.louisville.edu:8085/flower.roses/84037.jpg

5 Conclusions And Future Work

We have described our preliminary experience in developing a search tool benefiting from an integration of text and image content in querying and indexing. This framework promises to inherit the scalability and performance of powerful text search engines. In the future, we will consider crawling more world wide web images to get more annotations and extending our method to include image texture, shape and other features (possibly some that are application or collection dependent) to improve the search capabilities. Our search system applies a variety of image and text mining principles to allow extraction and indexing of both image content and textual annotations in a homogenous manner, and uses clustering to organize the search results and help users navigate through large collections of images on various websites. The proposed system promises to help users such as scientists, educators and students search through a variety of images, while taking advantage of available text and annotations when available. Adapting the proposed system to specific scientific needs is easy by extending the current and open source framework through the modular development of plugins that can extract richer and more sophisticated image content attributes that go far beyond color and texture to allow queries such as "*solar soho imgSolarEvent: Coronal-Mass-Ejection*" or by showing the system an example that is *like* the image to find such as "*solar soho imgSolarEvent: Coronal-Mass-Ejection imgurl:http://www.nasa.gov/images/imageX.jpg*".

References

[1] R. Baeza-Yates, J. R. del Solar, R. Verschae, C. Castillo, and C. Hurtado. Content-based image retrieval and characterization on specific web collections. *Conference on Image and Video Retrieval (CIVR), Springer LNCS*, pages 189–198, 2004.

[2] K. Barnard, P. Duygulu, and D. Forsyth. Clustering art. *Computer Vision and Pattern Recognition*, pages 435–439, 2001.

[3] K. Barnard and D. Forsyth. Learning the semantics of words and pictures. *International Conference on Computer Vision*, 2:408–415, 2001.

[4] M. Cafarella and D. Cutting. Building nutch: Open source search. *ACM Queue*, 2(5), April 2004.

[5] C. Carson, S. Belongie, H. Greenspan, and J. Malik. Blobworld: Image segmentation using expectation-maximization and its application to image querying. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(8), August 2002.

[6] C. Carson and V. E. Ogle. Storage and retrieval of feature data for a very large online image collection. *IEEE Computer Society Bulletin of the Technical Committee on Data Engineering*, 19(4), December 1996.

[7] C. Carson, M. Thomas, S. Belongie, J. Hellerstein, and J. Malik. Blobworld: A system for region-based image indexing and retrieval. *Proc. Third International Conf. Visual Information Systems*, pages 509–516, 1999.

[8] M. L. Cascia, S. Sethi, and S. Sclaroff. Combining textual and visual cues for content-based image retrieval on the world wide web. *Proceedings of the IEEE Workshop on Content - Based Access of Image and Video Libraries(CBAIVL)*, 1998.

[9] Y. Chen, J. Z. Wang, and R. Krovetz. An unsupervised learning approach to content-based image retrieval. *Proc. IEEE International Symposium on Signal Processing and its Applications*, pages 197–200, July 2003.

[10] Y. Deng, B. S. Manjunath, a. M. C. Kenney, and H.Shin. An efficient color representation for image retrieval. *IEEE Transactions on Image Processing*, 10(1):140–147, Jan 2001.

[11] T. Deselaers, T. Weyand, D. Keysers, W. Macherey, and H. Ney. Fire in imageclef 2005: Combining content-based image retrieval with textual information retrieval. *Working Notes of the CLEF Workshop, Vienna, Austria*, September 2005.

[12] M. Flickner, H. Sawhney, W. Niblack, J. Ashley, and et al. Query by image and video content: the qbic system. *IEEE computer*, 28(9):23–32, Sept 1995.

[13] Y. Gong. Advancing content-based image retrieval by exploiting image color and region features. *Multimedia Systems*, 7(6), 1999.

[14] E. H. O. GOSPODNETIC. *Lucene in Action.* Manning Publications Co., Greenwich, CT, 2005.

[15] Q. Iqbal and J. Aggarwal. Cires: A system for content-based retrieval in digital image libraries. pages 205–210. 7 International Conference on Control Automation, Robotics and Vision (ICARCV), December 2002.

[16] V. Kovalev and S. Volmer. Color co-occurrence descriptors for querying-by-example. *Proc. of the 5th Int'l Conf. on Multimedia Modeling*, pages 32–38, October 1998.

[17] W. K. Leow and R. Li. Adaptive binning and dissimilarity measure for image retrieval and classification. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2:234–239, 2001.

[18] J. Li and J. Wang. Automatic linguistic indexing of pictures by a statistical modeling approach. *IEEE Trans. on PAMI*, 25(9), 2003.

[19] X. Li, S.-C. Chen, M.-L. Shyu, and B. Furht. An effective content-based visual image retrieval system. *Proceedings of the 26th International Computer Software and Applications Conference on Prolonging Software Life: Development and Redevelopment*, 2002.

[20] Z.-N. Li, O. R. Zaiane, and B. Yan. C-BIRD: Content-based image retrieval from digital libraries using illumination invariance and recognition kernel. In *DEXA Workshop*, pages 361–366, 1998.

[21] Q. Lv, M. Charikar, and K. Li. Image similarity search with compact data structures. *In Proceedings of the 13th ACM Conference on Information and Knowledge Management (CIKM'04), Washington,D.C.*, November 2004.

[22] A. Natsev, R. Rastogi, and K. Shim. WALRUS: a similarity retrieval algorithm for image databases. SIGMOD, Philadelphia, PA, 1999.

[23] G. Qian, S. Sural, Y. Gu, and S. Pramanik. Similarity between euclidean and cosine angle distance for nearest neighbor queries. *SAC'04*, March 2004.

[24] S. Ravela and R. Manmatha. On computing global similarity in images. *Proceedings of the 4th IEEE Workshop on Applications of Computer Vision (WACV'98)*, 1998.

[25] J. Ren, Y. Shen, and L. Guo. A novel image retrieval based on representative colors. *Image and Vision Computing New Zealand*, 2003.

[26] K. S. A. R. Rohit Khare, Doug Cutting. Nutch: A flexible and scalable open-source web search engine. In *CommerceNet Labs Technical Report 04-04*, November 2004.

[27] Y. Rubner, J. Puzicha, C. Tomasi, and J. M. Buhmann. Empirical evaluation of dissimilarity measures for color and texture. *Computer Vision and Image Understanding*, 84(1):25–43, October 2001.

[28] Y. Rubner, C. Tomasi, and L. Guibas. The earth mover's distance as a metric for image retrieval. *International Journal of Computer Vision*, 40(2):99–121, 2000.

[29] S.Belongie, C.Carson, H.Greenspan, and J.Malik. Color and texture-based image segmentation using em and its application to content-based image retrieval. *IEEE ICCV*, pages 675–682, Jan 1998.

[30] T. K. SHIH, J.-Y. HUANG, C.-S. WANG, J. C. HUNG, and C.-H. KAO. An intelligent content-based image retrieval system based on color , shape and spatial relations. *Proc. Natl. Sci. Counc. ROC(A)*, 25(4), 2001.

[31] M. J. Swain, C. Frankel, and V. Athitsos. Webseer: An image search engine for the world wide web. *IEEE Computer Vision and Pattern Recognition Conference*, 1997.

[32] K. Vu, K. A. Hua, and J. Oh. A noise-free similarity model for image retrieval systems. *Proc. of IS&T/SPIE conference on Storage and Retrieval for Media Databases*, pages 1–11, Jan 2001.

[33] J. Z. Wang, J. Li, and G. Wiederhold. Simplicity: Semantics-sensitive integrated matching for picture libraries. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 23(9), Sept 2001.

[34] W. Xiaoling and X. Kanglin. Application of the fuzzy logic in content-based image retrieval. *JCS&T*, 5, April 2005.

[35] C. C. Yang. Content based image retrieval: A comparison between query by example and image browsing map approaches. *Journal of Information Science*, 30(3):257–270, 2004.

[36] R. Zhang and Z. M. Zhang. A clustering based approach to efficient image retrieval. *Proceedings of the 14th IEEE International Conference on Tools with Artificial Intelligence (ICTAI'02)*, 2002.