Approaching Archival Photographic Collections with Advanced Image Retrieval Technology

Lei Wang
School of Computer Science and Software Engineering
University of Wollongong, NSW, Australia, 2500
E-mail: leiw@uow.edu.au

July 29, 2012

Abstract

Nowadays, the volume of archival photographic collections is increasing with an unprecedented pace, posing significant challenges to traditional text-based image management systems. Content-based image retrieval (CBIR) technology has made considerable progress after two decades of intensive research. Instead of using text annotation, it directly uses visual content to conduct image search, making it more efficient in handling large collections. This paper introduces a prototype image retrieval system that applies the state-of-the-art CBIR technology to the archival photographic collection of National Archives of Australia, and discusses the advantages that it can bring forth. From the demonstrated result, it can be expected that CBIR technology will make important contributions to archival management and research in the digital era.

1 Introduction

During the last two decades, digital imaging equipments, computing technology and the Internet have made significant progress. It have become astonishingly easy, convenient and efficient for us to preserve and record information using digital images. The advert of mobile platforms and its ubiquitous use have made this situation more pronounced in the last several years. Also, the fast development of the Internet speeds up the dissemination, collection and sharing of images. These factors together have largely contributed to the increasing volumes of collections in libraries, museums, archives, galleries and social media websites.

Archival photographic collection is one of the typical examples. The sizes of collections are expanding with an unprecedent pace with more images collected and richer topics covered. This trend is expected to last for a sufficiently long period. Nevertheless, the increase on collection size does not necessary lead to better visual information access service for the public and the society. Instead, while we enjoy the convenience and efficiency in recording and collecting images, the fast increasing volume of image databases makes the access, browsing and retrieval of images more and more difficult. This situation has posed significant challenges to traditional text-based image management systems [10], which are currently widely used for archival photographic collections. The primary requirement for such systems is that every single image has to be annotated in advanced, and image browsing, search and retrieval are all conducted upon the associated annotations. The advantages of this approach lies at that through annotating images, it effectively links images to high-level semantic concepts used by humans and that the well-developed text processing techniques can be employed.

Nevertheless, the text-based systems are experiencing the following difficulties:

- With the increasing number of collected images, it becomes more and more expensive, laborious and time-consuming, if not intractable, to ask human annotators exhaustively annotate each image;
- Text annotations are often limited in expressive power. However, image content is becoming more diverse with the expanding collection size. In this situation, text annotations become less effective in accurately characterizing each image [1];
- Due to the subjectivity of human perception, people annotate same images differently. As a result, a query submitted by a user may not be necessarily consistent with the annotations associated with the relevant images, which makes the search fail. An example of this case is shown in Figure 1.

These problems call for a more efficient image retrieval technology to be used.

Content-based image retrieval (CBIR in short) technology provides us with a powerful tool to effectively address the above issues. It has been intensively researched during the last two decades and has achieved significant progress [3, 12]. The basic idea of CBIR is to directly use image content to evaluate the similarity between two images to conduct search and retrieval. This removes the above three difficulties in the following ways:

- Human annotator is replaced by computer, which handles large-sized image collections more efficiently;
- Text annotations are replaced by visual features describing the colors, textures, shapes in each image, which can deal with diverse image content;

[1] An adage says "A picture is worth a thousand words".

Figure 1: The above images can be annotated in various ways. As a result, the annotations associated with them may not agree well with the queries submitted by users who are going to find these images. Image courtesy of National Archives of Australia.

- Image retrieval is conducted by comparing the visual features associated with images, which is free of human perception subjectivity.

There has been a large body of literature on CBIR, including all sorts of approaches, methods and algorithms studying different issues related to CBIR [2]. Among of them, two fundamental issues have received much attention. One is how to characterize image content with different visual features, and the other is to how to evaluate image similarity in a way consistent with human perception. In addition to these two, another issue specifically related to large-sized image collections is how to find the images matching a query efficiently.

In the last several years, the Bag-of-features (BoF in short) model has been developed for image recognition, demonstrating superior recognition performance to existing methods in the literature [1]. It has been applied to image retrieval and again shown excellent retrieval performance [11]. The BoF model integrates advanced visual features and image similarity measures and supports fast search. It is considered as the pillar of support for the state-of-the-art CBIR technology. In this paper, we first introduce the BoF model in the context of image recognition, and then demonstrate a prototype system that employs this model to conduct retrieval on the archival photographic collection of National Archives of Australia.

Before the end of this section, it is instructive to discuss the most critical issue, called "semantic gap", of the CBIR approach. It means the gap between the low-level visual features used by computers and the high-level concepts used by humans. As a result, two images similar in terms of visual features do not necessarily have similar visual content, and vice versa. This has long been a standing issue in the area of image recognition. Much research work has been developed to reduce this gap in the literature, but there is still a long way to go to completely solve this problem [4].

2 The Bag-of-Features model

The BoF model originates from the field of information retrieval [7]. It was applied to the area of image recognition about ten years ago, and has been significantly developed since then [11, 1]. Its basic idea can be illustrated by Figure 2. In the BoF model, an image is viewed as a "document" and each small-sized patch of the image is viewed as a "word" in the document. The content of an document can be inferred from the occurrence of different words in the document. Applying the same idea to image recognition, the BoF model aims to infer the content of an image from the occurrence of different types of image patches. For example, if the patches of "nose" or "eye" are observed in an image, then it highly likely contains a face.

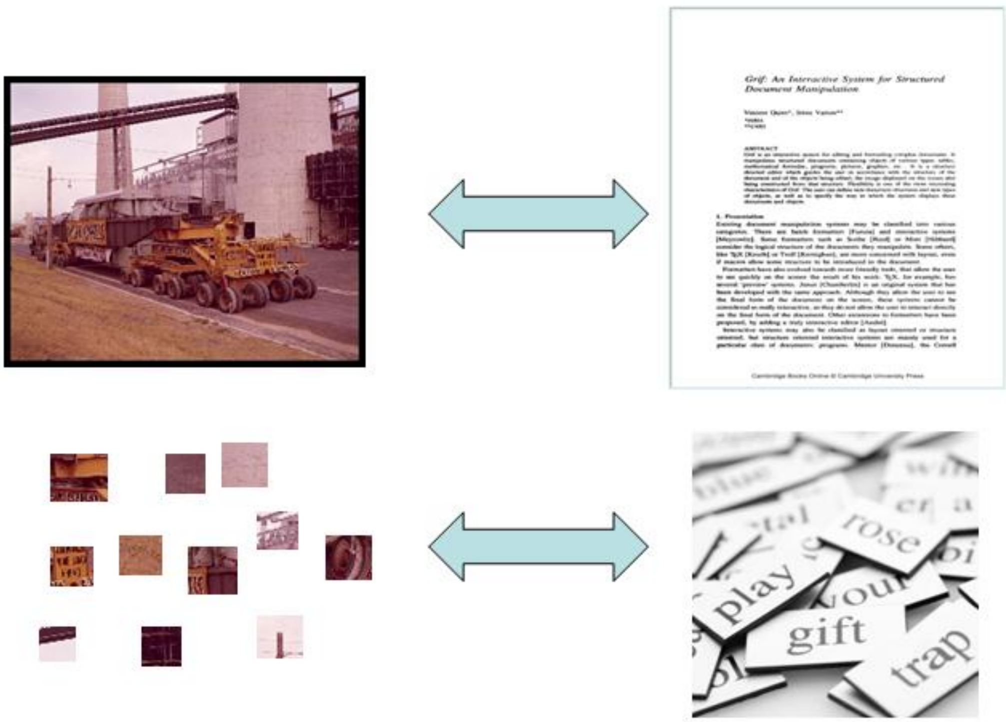

Figure 2: In the BoF model, an image is viewed as a "document" and each small-sized patch of the image is viewed as a "word" in the document.

Different methods have been developed in the literature to sample informative image patches from an image [8]. Intuitively speaking, informative pathes often correspond to those in which pixel intensities show significant changes, and these patches often contain important visual cues for recognition. Figure 3 shows some informative image patches sampled from an image. As seen, they are often "T"-junctions or "L"-junctions displaying considerable intensity changes. In addition, recent research shows that simply doing a raster scan over the pixels to densely sample image pathes can even produce better recognition performance [5]. This method has become more popular due to its simplicity and effectiveness.

Figure 3: This figure shows some informative image patches sampled from an image.

A fundamental difference of image retrieval from information retrieval lies at that for images, there is no well-defined words. To use the BoF model, "visual word" has to be generated. In image retrieval, image patches are firstly sampled from each image in a collection and its visual content is then characterized with a kind of descriptor [9]. All the image patches are pooled together and clustered into multiple groups. Each group contains visually similar image patches and they are viewed as a "visual word". Some examples of visual words are displayed in Figure 4. A collection of visual words are called "visual dictionary", which mimics the dictionary used in information retrieval. Once a visual dictionary is obtained, each image can be represented by a histogram showing the frequency of the occurrence of each visual word in this image. Then two images can be compared by using the associated histograms.

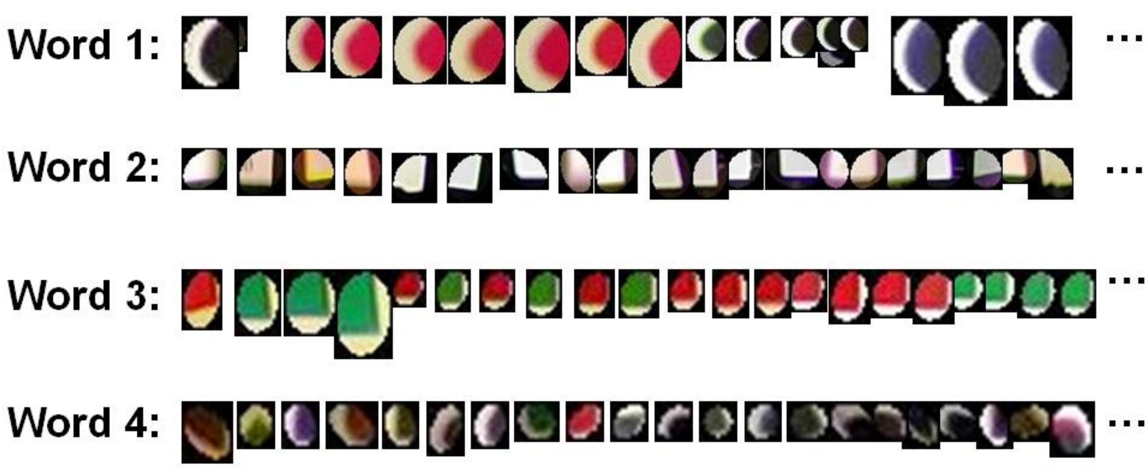

Figure 4: This figure shows examples of "visual words".

Structurally, the above histogram is same to that used in information retrieval, and supports fast search by nature. For image retrieval with the BoF model, the number of visual words is usually much larger than the image patches extracted from an image. As a result, the obtained histogram is very sparse, that is, the number of non-empty bins is very small. This makes “invert file”, a technique widely used in information retrieval to achieve fast search, still applicable to image retrieval [11]. This technique can significantly reduce the number of images to be compared and swiftly find the relevant images from a large-sized database.

3 The prototype CBIR system

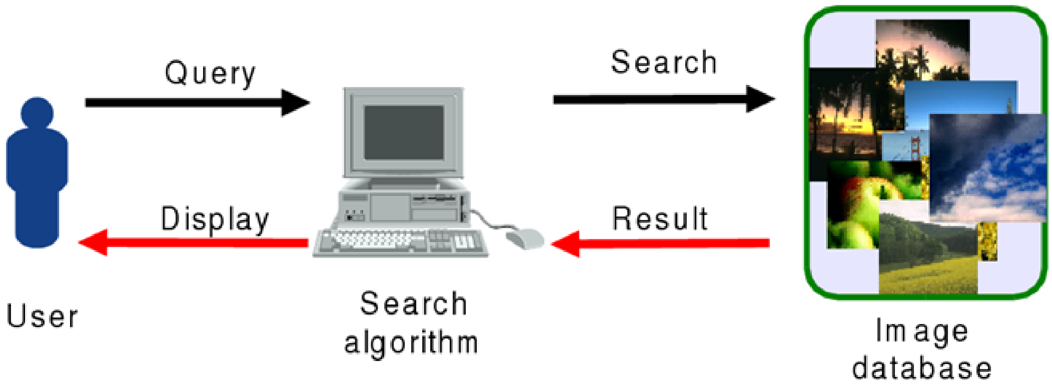

Figure 5: The illustration of our CBIR system.

In the following, this paper introduces the prototype system that applies the BoF model based CBIR approach to the archival photographic collection of National Archives of Australia. The structure of the system is illustrated in Figure 5. The collection of National Archives of Australia contains more than 220, 000 images and photos that cover a wide range of topics and span a long period of time. In building this system, a subset of 28, 000 images are randomly selected for use. Following the BoF model, image patches are densely sampled from each image and described using the method in [6], and in total around 61.4 million image patches are obtained. By applying an appropriate algorithm, they are clustered to generate 5, 000 visual words. Accordingly, each image is represented as a histogram of 5000 bins. Based on the histograms, an invert file is created to achieve fast search. A measure is predefined to evaluate the similarity of two images based on the associated histograms. A similarity score is assigned to each database image involved in the comparison. Sorting the scores in a descending order gives the top relevant images. These steps are illustrated in Figure 6. The whole system can be run on a laptop. Once an example of query is submitted, the retrieval result can be obtained in seconds.

The following gives some examples of retrieval result. As shown in Figure 7, the interface of the demonstration system is partitioned into left and right panels. The left one shows the images randomly selected from the image database, and the right one displays the retrieval result for

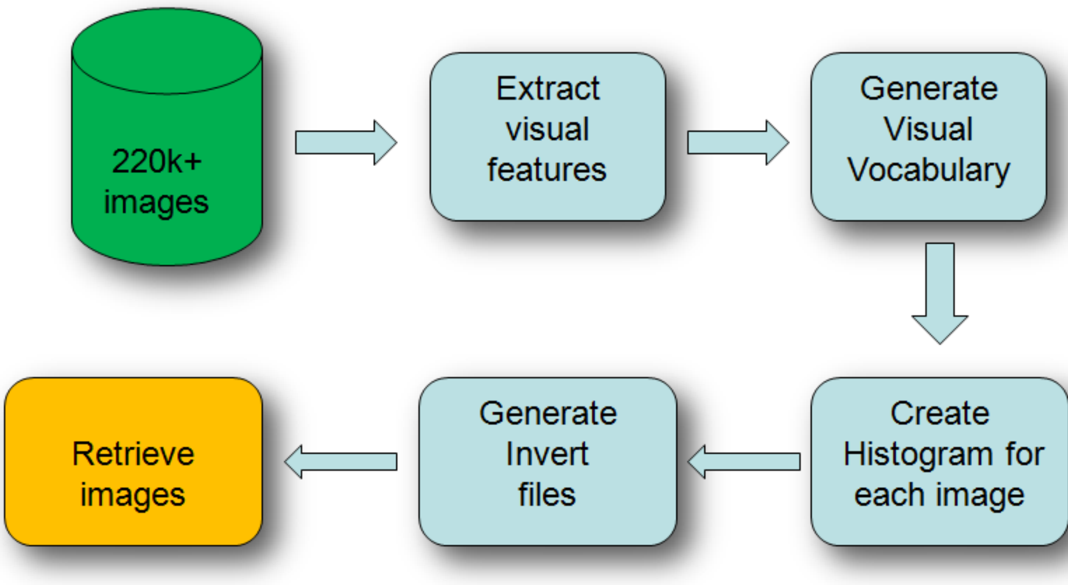

Figure 6: The key steps of building our CBIR system.

an example of query selected by a user, which is highlighted by a red square. Figure 7 shows the result of searching for an image related to "Wine Testing" [2]. As seen in the right panel, a number of identical copies are found from the image database. This retrieval result shows that the duplicated copies in a large-sized database can be easily found by using advanced CBIR techniques. The similar result can also be obtained for near-duplicated copies where colors, textures or other visual cues have been slightly modified. This helps to better manage image databases by grouping identical copies and avoiding collecting duplicated copies. In addition, when there are identical images in a database, the images can be found and annotated in one shot, which will save the cost on image annotation.

Figure 8 shows the retrieval result for an image of "The Governor General Lord Casey". The example of query is highlighted in the left panel, showing the scene of investiture. Imagine that an archival researcher finds this image somewhere and is keen to know more information about it. This can be well solved by using the CBIR approach, as demonstrated by the retrieval result in the right panel. A number of (non-identical) images related to the same event are retrieved, providing more information on it. Another retrieval result is show in Figure 9. The example of query talks about "The International Federation of Business and Professional Women held their 24th board meeting at the Canberra Rex Hotel in Canberra. Delegates enjoy a fashion parade". Again, more information on the event in the query can be obtained via the retrieval result.

[2] Note that the CBIR system does not need to know the information of "Wine Testing". This information is mentioned here only to explain the query.

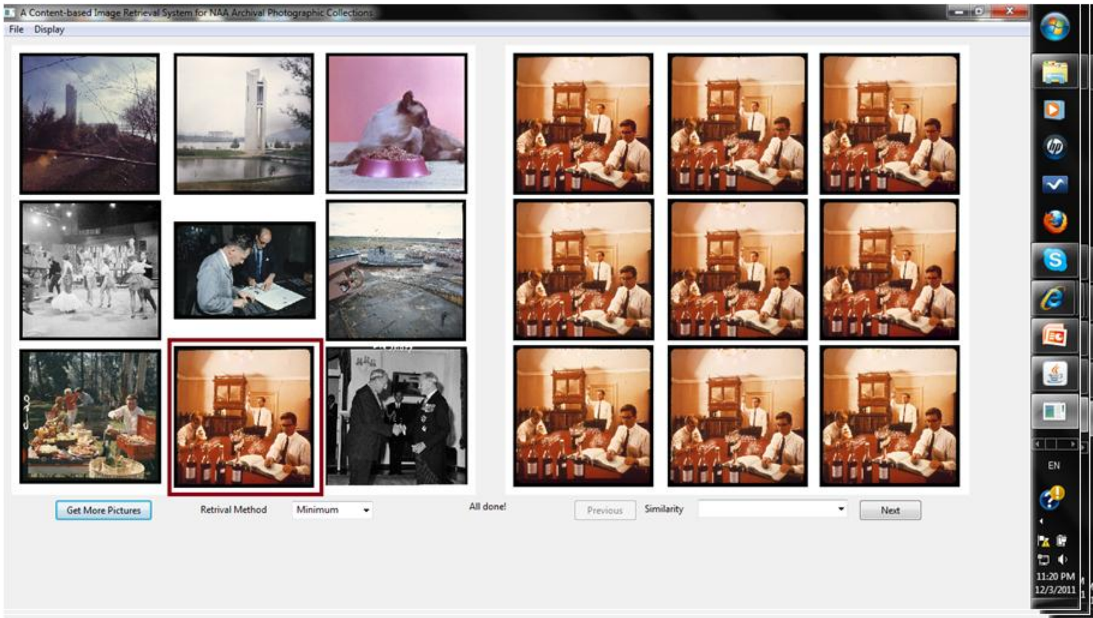

Figure 7: Example of retrieval result (I)

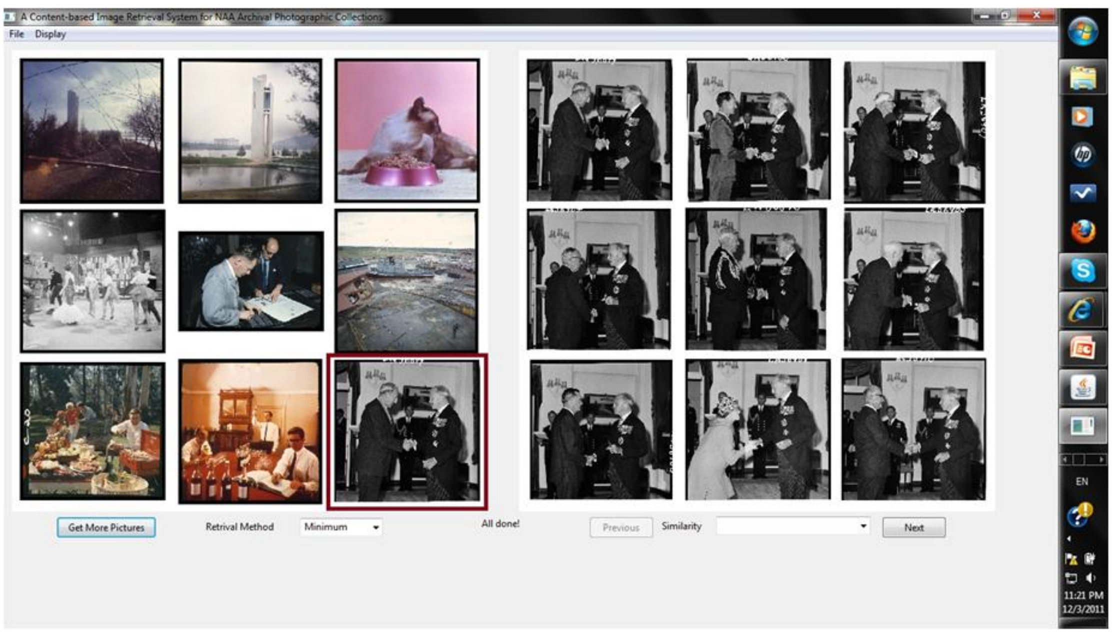

Figure 8: Example of retrieval result (II)

4 Discussion and future work

The above results only preliminarily demonstrate the power of advanced CBIR technology. Two issues are worth mentioning here. One is that the above retrieval conducts search based on the visual content of an whole image. We are extending the search to a region-based mode, that is, searching for images that contain a particular building, person, sign or even a generic object.

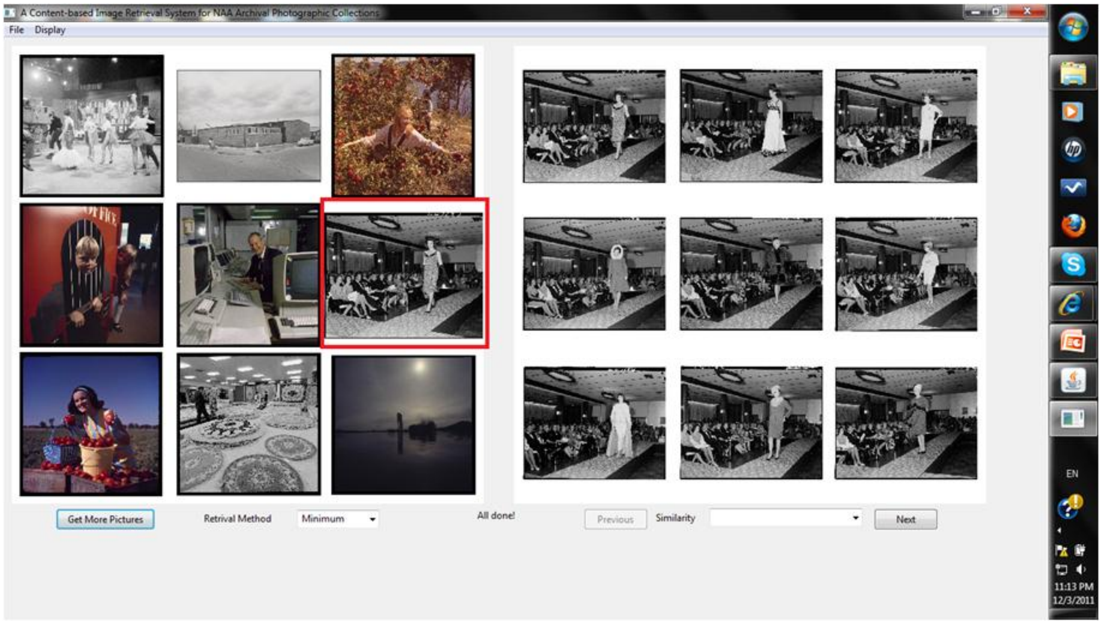

Figure 9: Example of retrieval result (III)

This is usually more desired when searching over archival photographic collections, and will provide more help to archival research. Region-based image retrieval has been studied in the literature, but the development of the BoF model makes such a retrieval mode easier to realize and more efficient [11]. Certainly, to achieve high accuracy on generic archival photographic collections, more research needs to done. Another issue is that the above retrieval conducts search based on a given example of query. In many cases, however, users may not have an example image beforehand and will have to use text-based queries. To address this issue, a retrieval system that combines traditional text-based method and the new content-based one is desired. We believe that a successful image retrieval system shall be the one that fully integrates text annotations and visual content, explores the synergy of them and freely switches between them. This will also be considered in the future work of this paper.

Acknowledgments

This work is supported by the Australian Research Council Linkage Grant LP0991757 with the Partner Organization of National Archives of Australia.

References

[1] Gabriela Csurka, Christopher R. Dance, Lixin Fan, Jutta Willamowski, and Cédric Bray. Visual categorization with bags of keypoints. In ECCV, 2004.

[2] Ritendra Datta, Dhiraj Joshi, Jia Li, and James Z. Wang. Image retrieval: Ideas, influences, and trends of the new age. ACM Comput. Surv., 40(2):5:1–5:60, May 2008.

[3] Myron Flickner, Harpreet S. Sawhney, Jonathan Ashley, Qian Huang, Byron Dom, Monika Gorkani, Jim Hafner, Denis Lee, Dragutin Petkovic, David Steele, and Peter Yanker. Query by image and video content: The qbic system. IEEE Computer, 28(9):23–32, 1995.

[4] David A. Forsyth and Jean Ponce. Computer Vision: A Modern Approach. Prentice Hall Professional Technical Reference, 2002.

[5] Frederic Jurie and Bill Triggs. Creating efficient codebooks for visual recognition. In ICCV, pages 604–610, 2005.

[6] Svetlana Lazebnik, Cordelia Schmid, and Jean Ponce. Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories. In Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Volume 2, CVPR '06, pages 2169–2178, Washington, DC, USA, 2006. IEEE Computer Society.

[7] Christopher D. Manning, Prabhakar Raghavan, and Hinrich Schtze. Introduction to Information Retrieval. Cambridge University Press, New York, NY, USA, 2008.

[8] Krystian Mikolajczyk and Cordelia Schmid. Scale & affine invariant interest point detectors. International Journal of Computer Vision, 60(1):63–86, 2004.

[9] Krystian Mikolajczyk and Cordelia Schmid. A performance evaluation of local descriptors. IEEE Transactions on Pattern Analysis & Machine Intelligence, 27(10):1615–1630, 2005.

[10] Yong Rui and Thomas S. Huang. Image retrieval: Current techniques, promising directions and open issues. Journal of Visual Communication and Image Representation, 10:39–62, 1999.

[11] J. Sivic and A. Zisserman. Video Google: A text retrieval approach to object matching in videos. In Proceedings of the International Conference on Computer Vision, volume 2, pages 1470–1477, October 2003.

[12] Arnold W. M. Smeulders, Marcel Worring, Simone Santini, Amarnath Gupta, and Ramesh Jain. Content-based image retrieval at the end of the early years. IEEE Trans. Pattern Anal. Mach. Intell., 22(12):1349–1380, December 2000.