Web-Scale Image Retrieval and Its Novel Applications

Sung-Eui Yoon
Division of Web Science and Technology and Dept. of Computer Science
KAIST

Abstract

Content-based image retrieval has been studied more than two decades, but surprisingly has not been widely deployed for web-scale image databases. This is mainly caused by two factors: low scalability and lack of commercial applications. In this poster we review our recent research activities for developing web-scalable image retrieval and its novel applications to address various problems. First we discuss scalability issues to be addressed for handling web-scale image databases. We also describe our hashing technique to drastically lower down the memory and computational requirements for web-scale image retrieval. Second we explain its applications to the fields of image copyright and editing, and discuss benefits that it brings over prior techniques developed without considering or utilizing web-scale image databases. Finally, we discuss remained problems towards our goal.

1 Introduction

WWW (World Wide Web) is increasing its presence and impacts to our lives in various aspects. Recently, web science has been considered as a new entity among academic research fields [Hendler et al. 2008]. Nonetheless, there has been rather less attentions on studying multimedia, especially images and videos, in the context of web science. In this poster we focus on web-scale image retrieval and their applications to various problems that can utilize or consider images available at WWW.

Thanks to rapid advances of digital camera and various image processing tools, we can easily create new pictures and images for various purposes. This in turn results in a huge amount of images available online through various image/video sharing websites such as flickr, YouTube, and facebook. For example, flickr contains more than five billion images and flickr members update more than three thousand image every minute [1].

[1] http://blog.flickr.net/en/2010/09/19/5000000000/

These huge image databases pose significant challenges in various applications (e.g., computer vision or image retrieval applications), in terms of scalability. Moreover, this huge amount of image datasets can enable novel approaches to address existing well-known problems (e.g., image editing) and even enable novel applications that have been not realized. At a high level, a huge amount of images available on the web enables a novel, massive-scale data-driven paradigm that has not been extensively studied in relevant academic fields such as image processing and computer vision.

In addition to the huge amount of available data, these image datasets are not created in a random manner. These data are voluntarily shared by web users and reflect their social behavior and interactions. Also, they form collective intelligence on various topics. As a result, by investigating and utilizing them, we can observe various social phenomena that has not been available through small sets of controlled data, and even design a new paradigm for better interactions between WWW and users in terms of sharing and utilizing images and other multimedia datasets.

In this poster, we explain our approaches towards web-scale image retrieval techniques and their applications, to handle and utilize web-scale image databases in efficient and effective manners. We first discuss the scalability issue and our approach addressing it in Sec. 2, followed by studying novel applications enabled by these recent advancements in Sec. 3. Finally, Sec. 4 presents our conclusion and future directions.

2 Scalability

Virtually all of image retrieval techniques first represent each individual image in their databases with an image descriptor. The most well-known image descriptors include SIFT [Lowe 2004] and GIST [Oliva and Torralba 2001]. These descriptors are accurate and have been demonstrated to work well in practice. They are unfortunately high-dimensional vector data (e.g. hundreds to thousands), leading to computational and memory-related complications. For example, when there are one billion images, the memory requirement of 384 dimensional, uncompressed GIST representations for those images is about 1.5 terabytes.

Most image retrieval techniques rely on efficient similarity search. As similarity search, nearest neighbor search techniques have been widely studied. Initially tree-based techniques [Friedman et al. 1977] such as kd-trees have been actively developed to accelerate nearest neighbor search. These tree-based techniques, however, have shown improvements only on low-dimensional data, but are not scalable to high-dimensional data points. In order to overcome this problem, hashing techniques have been receiving growing attentions and considered as an efficient solution for such high-dimensional data points [Torralba et al. 2008; Weiss et al. 2008].

Torralba et al. [2008] suggested that encoding high-dimensional data points into binary codes based on hashing techniques enables higher scalability thanks to both its compact data representation and efficient indexing mechanism. Under an assumption that similar high-dimensional data points are mapped to similar binary codes, we can efficiently identify approximate nearest neighbors by looking into only those similar binary codes (based on the Hamming distance), instead of accessing memory-intensive high-dimensional data.

Prior hashing approaches can be broadly classified into data-independent and data-dependent techniques. Hashing functions used in data-independent techniques are chosen independently from the input points. The most widely known techniques in this category is Locality-Sensitive Hashing (LSH) [Indyk and Motwani 1998]. This technique is extended to various metric spaces [Datar et al. 2004; Charikar 2002; Jain et al. 2008; Chum et al. 2008; Raginsky and Lazebnik 2009]. Recent research attentions have been shifted to developing data-dependent techniques to consider the distribution of data points and thus design better hashing functions. Notable examples include spectral hashing [Weiss et al. 2008], semi-supervised hashing [Wang et al. 2010], and joint optimization [He et al. 2011].

In all of these existing hashing techniques, hyperplanes are used to partition the data points (located in the original data space or a kernel space) into two sets and assign two different binary codes (e.g., -1 or $+1$) depending on which set each point is assigned to. Departing from this conventional approach, we have explored a novel hypersphere-based scheme, *spherical hashing*, for computing binary codes [Heo et al. 2012]. Intuitively, hyperspheres provide much stronger power in defining a tighter closed region in the original data space than hyperplanes. For example, at least $d+1$ hyperplanes are needed to define a closed region for a d-dimensional space, while only a single hypersphere can form such a closed region even in an arbitrarily high dimensional space. One can find that hyperplanes in a kernel space are able to map to non-linear hashing functions. However, we have found that the proposed simple spherical hashing in the original space achieves more spatially coherent partitioning than the non-linear hashing functions used in recent works [He et al. 2011; Joly and Buisson 2011; Raginsky and Lazebnik 2009].

In order to highlight benefits of our method, we have tested our method against different benchmarks that consists of one to 75 million image feature points with varying dimensions. We have also compared our method with many state-of-the-art techniques and found that our method significantly outperforms all the tested techniques, confirming the superior ability of defining closed regions with tighter bounds compared to conventional hyperplane-based hashing functions [Heo et al. 2012].

3 Applications

In this section we consider two well-known problems, image watermarking and scene completion, in the field of image processing and computer vision. In each problem, many techniques have been proposed and demonstrated to work well. Nonetheless, by approaching the same problems in the context of web-scale image databases, new technical challenges arise and in turn novel solutions emerge to address them.

3.1 Image Watermarking

One of the many challenging problems caused by easy image processing and modification technologies is the security problem. By the nature of digital data, it is very easy to copy, modify, redistribute the original image data. In order to address the security problem related to images, image watermark techniques have been studied actively in the last decade [Zheng et al. 2007].

The main concept of image watermarking is to embed visually imperceptible patterns on images so that a copyright holder of images can claim his or her ownership by extracting those patterns. Therefore, most image watermark techniques focus on extracting the embedded watermark patterns in a highly accurate manner

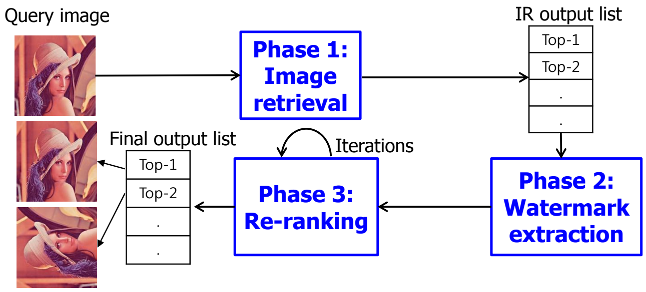

Figure 1: This figure shows an overview of the IRIW framework. Courtesy of authors of the IRIW paper [Jun et al. 2011].

against many different image attack scenarios (e.g., geometric transformation, cropping, and noise addition).

Even with drastic advances on image watermarking, the state-of-the-art image watermark techniques have certain false negative and false positive probabilities. As a result, a high number of false negative and false positive cases can occur, if we attempt to identify copyright-violated images solely based on image watermark techniques for web-scale image databases such as flickr. Furthermore, extracting watermark patterns and matching those patterns against the watermark pattern of the input query image can take prohibitive time for a large-scale image database consisting of millions of images or more.

In order to efficiently and accurately identify images that are modified or are the exactly same images from a query image in large-scale image databases, we presented a novel, Image Retrieval based Image Watermark (IRIW) framework [Jun et al. 2011]. Instead of exhaustively scanning and extracting watermark patterns from all the images in the image database, IRIW framework first identifies similar images given a query image by using a SIFT-based image retrieval method. Then the framework extracts watermark patterns only from those similar images and measures watermark pattern similarities against the query image. Finally, the framework re-ranks images by considering both image and watermark pattern similarities against the query image, in order to place images that are more likely to be copyright-violated in higher ranks in a final image list. Fig. 1 shows an overview of the IRIW framework. We also proposed to use a SIFT-aware image watermark method that does not embed watermark patterns on image regions where we get SIFT features, in order to minimize negative effects on our SIFT-based image retrieval method.

In order to verify the benefits of our method, we tested our method in an image database that consists of 10 K images. We found that our method improves the performance of searching copyright-violated images given a query image by more than two orders of magnitude over the exhaustive method that searches those copyright-violated images by accessing all the images in the database. More importantly, our method improves the accuracy of search results by reducing ratios of false negative and false positive cases up to two times over the exhaustive method. The performance and accuracy improvements of our method is mainly caused by identifying similar images based on image retrieval and by checking watermark similarities only for those images.

3.2 Scene Completion

When a photographer takes a picture, the picture may contain undesirable objects. In this case, the photographer can remove such objects and fill the region (i.e. hole) where the taken objects are located. These kinds of operations known as scene completion are quite common in practice and various image editing tools support such operations.

Scene completion has been extensively studied in the field of image processing. Prior approaches fill the hole of an image with other parts of the image. At a high level, these approaches identify similar patterns and transfer those patterns to the hole. However, these prior approaches may leave noticeable artifacts, when the hole takes a rather large region of the image or there are not many repeating patterns in the image.

Recently, Hays and Efros [2007] proposed a novel approach for scene completion using millions of photographs. At a high level, this approach identifies a small set of similar images among millions of images available on the web to the given image. For each similar image to the given image we then align these two images in terms of the boundary of the hole and blend them seamlessly based on the well-known Poisson blending technique [Pérez et al. 2003]. This data-driven scene completion technique showed that one can easily create visually-pleasing scene-completed images.

The main factor leading this technique to work well in wider cases over the traditional approaches is that it identifies and utilizes visually similar images on the web to the given image, while prior techniques try to address the same problem utilizing only the given image. This technique demonstrates a strong potential of the web-scale data-driven paradigm for addressing existing problems that work only with given images.

4 Conclusion

In this poster we have explained our recent research activities towards developing web-scale image retrieval and its applications to existing problems in the field of digital watermarking and image processing. Nonetheless, there are many technical challenges that we need to overcome and interesting questions that need to be answered:

1. **Lowering down costs:** Supporting web-scale image retrieval is not cheap. It requires a large set of commodity hardware with a huge power consumption. In this poster we have considered lowering down memory and computational requirements. Nonetheless, further technical improvements are required to enable cost-effective web-scale image retrieval.

2. **Developing commercial applications:** There have been drastic technical advances. Unfortunately, there has been little attentions on developing useful user interfaces and applications that help people to address their problems related to visual data. For example, current image retrieval techniques can identify similar images to the given image, but commercial needs of identifying visually or semantically similar images have not been discussed deeply. In order to realize commercial successes for web-scale image retrieval, it is critical to identify novel applications with an effective and convenient user interfaces describing what they want to find with visual data. Currently, we are looking into utilizing web-scale image retrieval for advertising commercial products.

We believe that this is our own small list for further investigation on web-scale image retrieval, and web science will play a major role on developing technology and applications related to web-scale image retrieval, in addition to considering their social impacts.

Acknowledgements

We would like to thank all the members of Scalable Graphics Lab. This work was supported in part by WCU program (R31-30007) under MEST/NRF of Korea, BK, MSRA, KMCC, DAPA/ADD (UD080042AD), Adobe, MEST/NRF (2011-0030822), and MCST/KOCCA/CT/R&D 2011.

References

CHARIKAR, M. S. 2002. Similarity estimation techniques from rounding algorithms. In *STOC*.

CHUM, O., PHILBIN, J., AND ZISSERMAN, A. 2008. Near duplicate image detection: min-hash and tf-idf weighting. In *BMVC*.

DATAR, M., IMMORLICA, N., INDYK, P., AND MIRROKNI, V. S. 2004. Locality-sensitive hashing scheme based on p-stable distributions. In *SoCG*.

FRIEDMAN, J. H., BENTLEY, J. L., AND FINKEL, R. A. 1977. An algorithm for finding best matches in logarithmic expected time. *ACM TOMS 3*, 3, 209–226.

HAYS, J., AND EFROS, A. A. 2007. Scene completion using millions of photographs. *ACM Trans. Graph. 26* (July).

HE, J., RADHAKRISHNAN, R., CHANG, S.-F., AND BAUER, C. 2011. Compact hashing with joint optimization of search accuracy and time. In *CVPR*.

HENDLER, J., SHADBOLT, N., HALL, W., BERNERS-LEE, T., AND WEITZNER, D. 2008. Web science: an interdisciplinary approach to understanding the web. *Commun. ACM 51* (July), 60–69.

HEO, J., LEE, Y., HE, J., CHANG, S.-F., AND YOON, S.-E. 2012. Spherical hashing. In *IEEE Computer Vision and Pattern Recognition (CVPR)*.

INDYK, P., AND MOTWANI, R. 1998. Approximate nearest neighbors: toward removing the curse of dimensionality. In *STOC*.

JAIN, P., KULIS, B., AND GRAUMAN, K. 2008. Fast image search for learned metrics. In *CVPR*.

JOLY, A., AND BUISSON, O. 2011. Random maximum margin hashing. In *CVPR*.

JUN, J. Y., KIM, K., HEO, J.-P., AND YOON, S.-E. 2011. Iriw: Image retrieval based image watermarking for large-scale image databases. In *Int. Workshop on Digital-forensics and Watermarking*.

LOWE, D. 2004. Distinctive image features from scale-invariant keypoints. *IJCV 60*, 2, 91–110.

OLIVA, A., AND TORRALBA, A. 2001. Modeling the shape of the scene: a holistic representation of the spatial envelope. *IJCV*.

PÉREZ, P., GANGNET, M., AND BLAKE, A. 2003. Poisson image editing. *ACM Trans. Graph. 22* (July), 313–318.

RAGINSKY, M., AND LAZEBNIK, S. 2009. Locality sensitive binary codes from shift-invariant kernels. In *NIPS*.

TORRALBA, A., FERGUS, R., AND WEISS, Y. 2008. Small codes and large image databases for recognition. In *CVPR*.

WANG, J., KUMAR, S., AND CHANG, S.-F. 2010. Semi-supervised hashing for scalable image retrieval. In *CVPR*.

WEISS, Y., TORRALBA, A., AND FERGUS, R. 2008. Spectral hashing. In *NIPS*.

ZHENG, D., LIU, Y., ZHAO, J., AND SADDIK, A. E. 2007. A survey of rst invariant image watermarking algorithms. *ACM Computing Survey 39*, 2, 5.