1
Introduction

In this chapter, we discuss overall flow and applications of image search techniques.

1.1 Introduction to Image Search

We identify various information by using search tools such as Google. Many well-established techniques perform the search by matching texts or labels that users type. This kind of text-based search has been extremely successful, since a huge amount of information is represented by texts.

Image search or context-based image retrieval looks at the content of images and identifies similar images that a user provided.

Image search or Content-Based Image Retrieval (CBIR) has been studied for many decades, since images or other multimedia (e.g., video) are common representations in addition to texts. Simply speaking, the main goal of image search is to identify similar images given a user-specified image or other types of inputs such as sketch or labels (Fig. 1.1).

There have been constant demands for image search, but its demands become more pronounced in these days, since people take photos frequently and record them for later use. This trend is supported by developments of digital cameras and easy of sharing photos by using mobile phones and various social media.

Additionally, mobile phones are parts of our lives now and we commonly use them for various tasks including taking pictures. It is rather inconvenient to type texts on those mobile phones, and thus leads people to communicate information through images and videos more. These trends are compounding together and generating ever-growing needs for accessing and sharing information through various multimedia including images.

Key components of image search include image representations, indexing or organization methods of images, matching methods among images. We typically represent input and DB images in a particular image representation such as SIFT and CNN features that

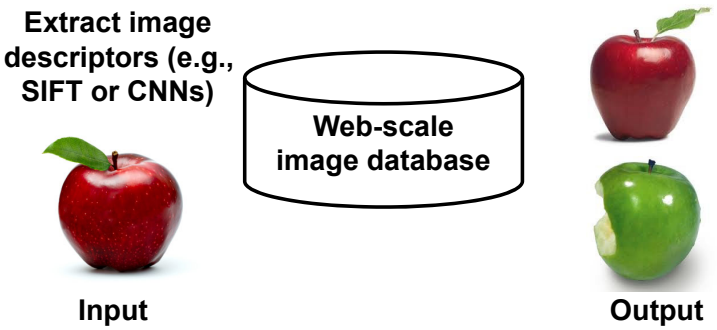

Figure 1.1: Image search starts with an input image and identifies similar images from a large-scale image database. Image search identifies such images by looking into their content. As a result, image search is also known as Content-Based Image Retrieval (CBIR).

encode content information stably across different illumination level and so on. For efficient search, we represent images in a compact and efficient database representation. We then go through all the images in the DB and identify similar images with the given query image based on matching techniques. Additionally, we perform a high level tasks such as classification or localization of particular objects in query or searched images.

Image search has various applications such as searching desired products and photos, image stitching, improving techniques of object/scene/location recognitions, robot motion planning, etc., where some forms of visual information needs to be matched. This is discussed in Ch. 1.2.

Related topics and resources. This book aims for a self-contained article, but we explain other available resources and topics that can help you to deepen your understanding on the subject.

1. Computer vision. Computer vision targets to reconstruct three dimensional (3D) models from 2D images. Many techniques (e.g., image representation) for image search have been developed from the computer vision community. An excellent, online book written by Szeliski on the topic is available [1].

2. Conference proceedings related to image search. Many papers related to image search or retrieval appear at various conferences including CVPR, ICCV, ECCV, ACM MM, ACM SIGGRAPH, etc. Recent techniques related to machine learning are from NeurIPS, ICML, etc. Some of computer vision papers are available at `http://www.cvpapers.com`.

[1] Richard Szeliski. *Computer Vision: Algorithms and Applications*. Springer-Verlag New York, Inc., 1st edition, 2010

1.2 *Applications of Image Search*

Image search is one of fundamental tools across different fields and thus there are numbers applications utilizing it. We list some of them

Figure 1.2: The left shows a photo album related to a landmark. The right shows a reconstructed 3D model in a form of point clouds. The reconstruction process is related to matching similar parts of photos. The image is excerpted from the work of Snavely et al.

here. YOON: List applications in a detailed manner including cross modal search.

1. **3D reconstruction.** One can wish to reconstruct a 3D model out of related photos. For example, suppose that you visited a Colosseum in Rome and took many photos of the landmark. At your home, browsing your album of the photos, you may want to see the overall structure of the Colosseum in a different view from ones taken in your photos. This can be started with reconstructing the 3D model of the Colosseum. Typically, reconstructing a 3D model from a photo album begins by matching corresponding parts of the images, and thus can be realized by techniques related to image search. Once we build a 3D model of a scene, one can navigate the scene and better retrieve images related to the scene (e.g., Photo tourism [2]).

2. **Time-lapse photography or rephotography.** Time-lapse photography captures images of certain views in a much lower frequency than that commonly used to capture them (Fig. 1.3; the image is excerpted from the paper [3]). As a result, by seeing those captured sequence, we can see slowly moving or changing phenomena in a highlighted manner. Rephotography captures the same scene with a time lag to show "now and then" looks of the scene. To support such techniques, it is critical to capture the same scene in the same projection, which is a tight requirement. Instead, if we can identify matching points between images, and compute their correspondences and warp them into a canonical space, we can support such techniques in a more relaxed manner.

3. **Edit transfers to other similar images.** Suppose that we have to modify a particular part of a frame in a video and you need to edit other frames in a similar way. In this case, transferring those edits to similar parts in other frames is critical for enabling efficient edits (Fig. 1.3). To enable such techniques, we also need to identify matching features or pixels.

[2] Noah Snavely, Steven M. Seitz, and Richard Szeliski. Photo tourism: Exploring photo collections in 3d. *ACM Trans. Graph.*, 2006

[3] Xiaoyong Shen, Xin Tao, Chao Zhou, Hongyun Gao, and Jiaya Jia. Foremost regional matching for internet scene images. *ACM Transactions on Graphics*, 2016

Figure 1.3: (a) shows photos identified from internet. By matching corresponding parts of the photo, we can project those photos in a canonical view and thus create time-lapse video shown in (b), or transfer an edit in a photo into another photo (e). This figure is excerpted from the work of Shen et al.