Chapter 54

A Combination of Spatial Pyramid and Inverted Index for Large-Scale Image Retrieval

Vinh-Tiep Nguyen  
*Vietnam National University, Ho Chi Minh City, Vietnam*

Thanh Duc Ngo  
*Vietnam National University, Ho Chi Minh City, Vietnam*

Minh-Triet Tran  
*Vietnam National University, Ho Chi Minh City, Vietnam*

Duy-Dinh Le  
*Vietnam National University, Ho Chi Minh City, Vietnam*

Duc Anh Duong  
*Vietnam National University, Ho Chi Minh City, Vietnam*

**ABSTRACT**

Large-scale image retrieval has been shown remarkable potential in real-life applications. The standard approach is based on Inverted Indexing, given images are represented using Bag-of-Words model. However, one major limitation of both Inverted Index and Bag-of-Words presentation is that they ignore spatial information of visual words in image presentation and comparison. As a result, retrieval accuracy is decreased. In this paper, the authors investigate an approach to integrate spatial information into Inverted Index to improve accuracy while maintaining short retrieval time. Experiments conducted on several benchmark datasets (Oxford Building 5K, Oxford Building 5K+100K and Paris 6K) demonstrate the effectiveness of our proposed approach.

DOI: 10.4018/978-1-5225-5204-8.ch054
1. INTRODUCTION

Image retrieval refers to the task of finding images of a certain object in image datasets. In a typical scenario, users provide a target object. The retrieval system then returns a ranked list of images containing the object of interest. With its wide range of applications in real-life, image retrieval has gained interest in recent years. Although several approaches have been introduced in recent years, image retrieval remains a challenging problem. Beside difficulty due to the variations of visual content in the images (i.e. same object or scene has various appearances in different viewpoints, environmental conditions), a yet another problem is to keep retrieval time reasonably short, especially when the retrieved dataset is large. Hence, reducing retrieval time while maintaining accuracy is always required.

There are many works have been proposed to address image retrieval. Most of them rely on the Bag-of-Words (BoW) model (Liu, 2013; Sivic & Zisserman, 2003) of which each image is represented by a single compact vector. There are three reasons to make the success of BOW model. Firstly, this model takes advantage of the power of local feature descriptor, specifically in this case the SIFT features. Secondly, the model represents image / video by a single vector so that comparison between two images or shots are simplify to comparing to BOW vectors using a dissimilarity score. Thirdly, representing image as a large dimensional vector can take advantage of the sparse space of the data to be able to search effectively. To achieve a compact representation, local features are quantized into clusters whose centers are representative visual words. Based on those visual words, a histogram of occurrence counts of words is generated to present each image. Using the generated histograms of images, an inverted index of the images are formulated for fast retrieval. One index term is corresponding to one visual word. A posting list is created with each term. The list includes identities of images having the term (i.e. the visual word) inside. Thus, given an query image, a set of images having similar visual words with the query can be rapidly obtained by the union of posting lists. Each posting list here is of one visual word of the query. Images are then ranked by counting the number of visual words shared by the query image and indexed images. An image having larger number of shared words will be ranked higher in the list. By using such methodology, images can be returned to users in a short time.

However, the major limitation of the standard BoW model and Inverted Index is that they ignore an important source of information to differentiate. By counting the number of words but not considering locations of words or their co-occurrences in images, spatial information from the extracted words is totally discarded. Hence, in some cases, two images with different content can be considered as similar if we just index and count shared visual words using their global histograms. An example is given in Figure 1.

Our objective in this paper is to investigate an approach which can take advantage of the spatial information from visual words in images to increase accuracy while keeping retrieval time short. The key idea is to consider the distribution of words in different regions of the images. Images having more similar word distributions in regions with those of the query image are considered more relevant. Multiple inverted indexes are used to rapidly obtain candidate image sets for regions, thus help to maintain fast retrieval speed. The rest of this paper is organized as follows. Section II presents several works related to our study. In Section III, we introduce standard approach which used as baseline and then present our proposed approach in detail. Section IV shows experimental results with discussions. Finally, we conclude the paper in Section V.

www.igi-global.com/e-resources/library-recommendation/?id=1

Related Content

Optimal Image Watermarking using Hybrid Optimization Algorithm
www.igi-global.com/article/optimal-image-watermarking-using-hybrid-optimization-algorithm/151506?camid=4v1a

Evaluating an Evolutionary Particle Swarm Optimization for Fast Fuzzy C-Means Clustering on Liver CT Images
www.igi-global.com/chapter/evaluating-an-evolutionary-particle-swarm-optimization-for-fast-fuzzy-c-means-clustering-on-liver-ct-images/108407?camid=4v1a

Watermarking of Digital Images With the Substitution of Low-Value Bits to Increase Capacity

Applications of Kernel Theory to Speech Recognition
www.igi-global.com/chapter/applications-kernel-theory-speech-recognition/24825?camid=4v1a