A Survey on Context-based Image Retrieval with Application for Demonstration: Searching Screenshot in Video

NG Yiu Wai, *MSc in Information Technology (HKPolyU)*

***Abstract*—This paper aims to provide a brief introduction on existing technologies for context-based image retrieval (CBIR). Several survey papers in recent decades are discussed. An application is built to demonstrate the basic elements of CBIR.**

***Index Terms*—Context-based Image Retrieval**

I. Introduction

Multimedia database querying is always an important topic in computer vision and database management. In modern society, anybody could generate high-resolution images, for example taking photos using mobile phone, and share images through the Internet. Since the Internet becomes popular in the 1990s, the volume of digital images generated by users grows together with the Internet. New technologies are needed to manage this huge number of images. A survey paper in year 1999 by Yong Rui, Thomas S. Huang and Shih-Fu Chang [1] summarized the existing methods for image retrieval in multimedia database at that time. About one decade later, another survey paper written by Ritendra Datta, Dhiraj Joshi and James Z. Wang [2] is published to summarize the new trends of CBIR. In recent years, significant achievements are made in machine learning algorithms. A survey conducted in year 2013 by T. Dharami and I. Laurence Aroquiaraj summarized how to apply unsupervised machine learning in CBIR [3]. In this paper, some important ideas of CBIR are introduced. An application is created to demonstrate the concepts and technologies.

II. Architecture of Traditional CBIR System

The image retrieval system before 1990s mainly adopts text-based approach. Images are manually clustered into groups and tagged with key words. The tag words are used as key for index building and searching. This approach is not suitable nowadays because the classification relies on human. It is almost impossible to handle a huge volume of images.

Content-based image retrieval is proposed to solve this problem. It aims to analyze images by software, but not human, and automatically builds key for indexing. Images are indexed and managed based on their contents, i.e. intensity and spatial information of pixels. Main components of CBIR system include:

A. Features Extraction

Features are distinct characteristics of an image to identify it from others. Due to the large size of an image, it is inefficient to classify images by comparing every pixel. Features vector is a list of elements which represents the content of an image. The size of features vector should be much smaller than the original image, which allows fast comparison and searching. In CBIR, the features vector is generated based on pixels of an image. Different algorithms are developed to automatically extract features vector from image.

.

B. High Dimensional Indexing

Features vectors are usually in high dimension, which is not appropriate for building search tree. Some dimension reduction techniques are developed to lower the length of features vector. In other words, extracting features from a features vector. A common approach is using principal component analysis. By studying the distribution of features' numerical values, a reduced size features vector could be extracted for index building and searching.

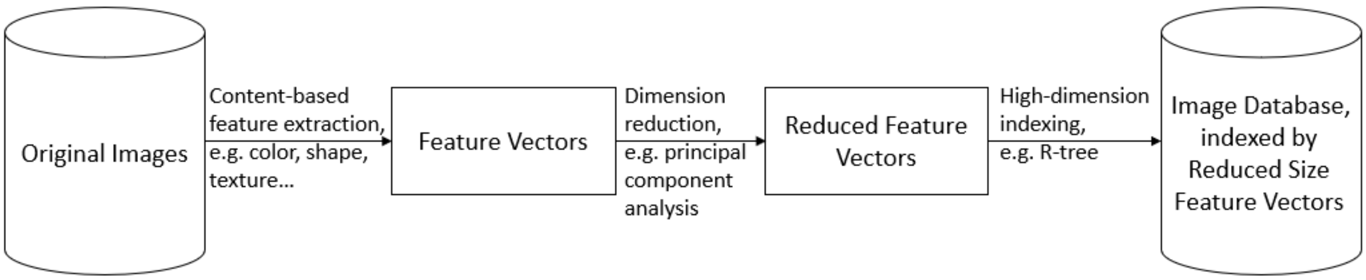

Figure 1: Architecture of traditional CBIR system.

Searching could be performed by comparing the features vectors. A well-designed CBIR system should supports similarity search. Images with similar content should generate similar features vectors. A popular and efficient approach to search similar features vectors is k-nearest-neighbor querying. It is an important topic in database querying, but out of the scopes of this paper.

III. FEATURES EXTRACTION

There are multiple methods for extracting features from images. Below are some well-studied methods.

A. Color

Color is the intensity value of each pixel. Color-based methods is simple and robust. One of the most commonly used color-based method is color histogram. It uses the distribution of intensity values in different color channels to represent an image. A similar technique is color moment, which transforms the distribution to some representative values by considering their derivations.

The main drawback of color-based method is the spatial information are not included. Performing a similarity search purely based on color-based features may return irrelevant images. It is because the correlation between pixels cannot be represented by color-based features. An improved method is called color layout, which divides image to blocks before color-features is extracted. A certain degree of spatial information is retained in this method.

B. Texture

Texture is the visual patterns of a group of pixels in a region. Texture-based methods usually involves capturing the contrast of pixel in a region, i.e. the change of intensity among nearby pixels. The repeating pattern of intensity changes is texture, which could be formulated to values by wavelet transform. Wavelet transform is designated to record the wave-like oscillation of amplitudes. It is a good method to capture the repeating pattern of intensity changes. Then texture-based features are extracted.

C. Shape

Shape is one of the most important features for humans to recognize objects. Shape-based methods could be divided into two categories: boundary-based and region-based. A representative boundary-based method is Fourier descriptor. The outer shapes of objects are captured by sine and cosine functions using Fourier transform. For region-based method, a common method is moment invariant. A point of the image is selected to be the center. Moment is calculated by multiplying the geometric location and intensity of all points.

D. Machine Learning

The common goal of traditional features extraction algorithms is mapping intensity and spatial information of image pixels to features using well-studied mathematical functions. However, machine learning methods choose another approach: clustering images to groups by using the intensity and spatial information of pixels. The features extracted representing the grouping information of an image. Some well-known clustering algorithms are D-EM Algorithm and Support Vector Machines.

IV. High Dimensional Indexing

Features are the representative information of an image. Image retrieval means comparing a target features vector with other features vectors in database, then retrieves the image with matched features vector. Matching means the distance between two vectors is very small, or even equals to zero. However, the features vectors are usually in high dimension. It is a computational expensive to compare high dimension data. The dimension of features vectors should be reduced to speed up the indexing and searching process.

A. Dimension Reduction

The most common method of dimension reduction is principal component analysis. It is an orthogonal transformation to map features into a set of linearly uncorrelated values. After transformation, low-dimensional features will represent most of the information. High-dimension features are less important. They could be omitted when comparing feature vectors for image searching. A famous principal component analysis technique is called Karhunen–Loève transform.

B. Search Tree for Multi-Dimension Keys

Comparing feature vectors by calculating the distance is expensive, even if the dimensions of vectors are reduced. Search tree should be built using the feature vectors to reduce the number of comparisons in searching. Some popular multi-dimensional index trees are k-d tree, quad-tree, R-tree and their variants.

Another technique for indexing is using the lower boundary of distance [4]. The computation cost of lower boundary is cheaper than computing distance, which could speed up comparison and searching. The k-nearest-neighbor querying techniques could be applied to improve the efficiency of image retrieval.

V. Experiment: Application for Searching Screenshot in Video

To demonstrate the basic components of CBIR, an application for searching screenshot in video is created. Features vectors are extracted from frames of a video. After that, search tree is built using the feature vectors. User could input any image to check whether it is a screenshot and when does it appear. The algorithms used in this application include:

- Feature extraction technique is color layout. Frames are divided into rectangular blocks. Features extracted are the average intensity of pixels in each block for every color channel.
- Dimension reduction technique is discrete cosine transform. Only the top 16-dimensions features are retained to speed up the indexing and searching process.
- Indexing technique is octree, which is a variant of quad-tree. The top 3-dimensions features are used for building index tree. The lower-boundary function is calculated using the top 3-dimensions features only.
- Distance function is mean square error. Or in other words, it is the Euclidean distance between two vectors.

The experiment is success. Screenshots from a video could be identified by comparing the feature vectors. It is faster than sequential search because a search tree is built. Details of algorithms and data structure could be found in the demonstration report.

In one of the samples, a video with length = 26 seconds is chosen as the input data. 113 frames in the video are selected for feature extraction. An octree tree is built using the feature vectors. Finally, a screenshot is selected for searching it among all frames. By comparing the features vectors, it is found that the screenshot appears at time = 18 seconds. 14 comparison are performed, which is much less than the total number of vectors = 113. It is an evidence that images could be retrieved by using features vectors. Searching in indexed feature vectors is faster than sequential search. This illustrative sample could be found in the demonstration report.

VI. Conclusion and Other Topics

Content-based image retrieval techniques are important for managing the images. The core elements are feature extraction and high-dimension indexing. Different algorithms are developed in past decades to improve CBIR's performance.

CBIR systems are implemented since the 1990s. Comparing with text-based image retrieval system, it is highly automated and more efficient in indexing images. It makes the maintenance of huge volume multimedia databases possible. However, the rapid growth of images collection never stops. New techniques are invented to further improve the speed and accuracy of multimedia databases.

For example, one direction for improving dimension reduction is mixing text-based and content-based image retrieval. The numerical features vectors are high-dimensional, which requires multiple computational operations for calculating distance between two vectors. It is one of the bottlenecks for fast image searching. Recently a paper [5] studies on how to efficiently combine text index with image database, i.e. using bag-of-words as index of CBIR system. Images are clustered into groups using their content-based features, then mapped into bag-of-words for indexing. Images could be retrieved using the multi-dimensional text invert index.

References

[1] Rui, Y., Huang, T. S., & Chang, S. F. (1999). Image retrieval: Current techniques, promising directions, and open issues. *Journal of visual communication and image representation*, *10*(1), 39-62.

[2] Datta, R., Joshi, D., Li, J., & Wang, J. Z. (2008). Image retrieval: Ideas, influences, and trends of the new age. *ACM Computing Surveys (Csur)*, *40*(2), 5.

[3] Dharani, T., & Aroquiaraj, I. L. (2013). A survey on content based image retrieval. In *2013 International Conference on Pattern Recognition, Informatics and Mobile Engineering* (pp. 485-490). IEEE.

[4] Korn, F., Sidiropoulos, N., Faloutsos, C., Siegel, E., & Protopapas, Z. (1996). Fast nearest neighbor search in medical image databases. In *1996 International Conference on Very Large Data Bases.*

[5] Siedlaczek, M., Wang, Q., Chen, Y. Y., & Suel, T. (2018, December). Fast Bag-Of-Words Candidate Selection in Content-Based Instance Retrieval Systems. In *2018 IEEE International Conference on Big Data (Big Data)* (pp. 821-830). IEEE.