International Journal On Engineering Technology and Sciences – IJETS™
ISSN(P): 2349-3968, ISSN (O): 2349-3976
Volume V, Issue IX, September - 2018

Image Object Retrieval Using Conventional Approaches: A Survey

Amitha I C
Research Scholar
Department of IT
Kannur University
amithachandran@gmail.com

Prof. Dr. N. K. Narayanan,
Principal, College of Engineering Vadakara,
nknarayanan@gmail.com

Abstract

Efficient object retrieval is considered as more challenging problem in the area of big data, since targets usually only occupies small regions on images. Object retrieval differs from traditional image retrieval. Object retrieval focuses on retrieving sub-image level object, which usually appears in different background context. One of the main challenges in sub-image level object retrieval is to retrieve the small queried images from a cluttered scene. If the object is very small and the scene is cluttered with many other objects too[1]. Object retrieval is a fundamental problem with numerous applications on product search, archive video search, database consistency checking, theft detection etc.

1. Introduction

The aim of the survey is to identify strength and weaknesses of conventional object retrieval methods. The input to the object retrieval system can be achieved in two different ways either a text or an image.

Object recognition in still images has been widely studied in the field of computer vision. Tasks may involve classification, retrieval and even detection. Classification, also called object categorization, consists in classifying images into categories by giving them the labels of the objects they contain. For each object category, images are attributed the value1or 0depending on whether or not they contain the object. Following the same idea, we can perform object retrieval .Users can query the system to retrieve images which contain the objects they are looking for. This should be based upon an efficient and fast index structure to ensure a reasonable response time, particularly when dealing with large databases and/ or complex models [2].

Object recognition involves many challenges starting with the definition of the training dataset and the choice of the visual descriptors, moving to the way the computer learns and the construction of a reliable classifier. Some researchers prefer to use a bottom-up approach tracing the very low information in the signal and trying to interpret it so as to get powerful models. Others find that it is more intuitive to use a top-down approach. It deals with uncertainty of instance labels. An image is viewed as a bag of multiple features which are the local visual signatures. The bag will have only one label according to whether or not it includes at least one positive instance. It follows that it is only certain for a negative bag that there are no objects. Using a weak learning approach will also help to train images without much knowledge about the objects inside so there will be no need to construct a ground truth per object location [2].

An object is viewed as a tangible concept. Good recognition comes with a good description, especially one that uses multi-criteria such as shape, texture, scale and color. Image descriptors are indeed the raw material and the basic data for learning. In order to cover the difference in the nature of the objects to be learned and at the same time the intra-class variability of the same object, a multiple description scheme is needed. It is then up to the learning algorithm to choose a descriptor or a combination of many descriptors that best suits a given category. Interpretability could derive from this fact. Interpretability aims to reduce the semantic gap that exists between human knowledge and the computational representation of the models learned. Computer models are usually too abstract for users to understand where bad results come from. By generating interpretable models, we somehow create a link between the numerical representation of objects and our visual representation. Not only does interpretability enhance our understanding of outputted results but it is also a very effective tool for user interactivity. It allows users to comprehend what the generic model is composed of and to choose, in different situations, the visual patches that best matches their needs. Therefore, interpret-ability is a means to achieve generality. Users can perform object retrieval in large database collections which may contain heterogeneous data from different sources [2].

This paper gives a survey on conventional object retrieval approaches. First of all different feature representation methods used for image object retrieval.

Second, have a look on different conventional feature selection methods. Third, a comparative study on all the conventional approaches have been used for image object retrieval.

2. Conventional Image Object Retrieval Methods

In [3], illumination invariant object recognition was achieved by normalizing the three color bands. They employed the compressed histogram of the chromaticity to arrive at a valuable representation of an object which can facilitate high retrieval accuracy. The first shortcoming of this method lies in the usage of a uniform quantization scheme in obtaining the chromaticity, which is not in agreement with the perception of the human vision system. In this method, they developed an approach using the CIE UCS transform to circumvent this problem. Second, instead of using uncompressed images to achieve the illumination invariant indexing and retrieval, they carry out our indexing process directly in the DCT domain by using several coefficients from each macro-block. Third, in light of the special properties of the normalized chromaticity histogram frames, the foundation of the ensuing low-pass filtering, an additional step is inserted to render this frame smoother thus resulting in a better data reduction. Fourth, in order to facilitate efficient retrieval during data query phase, which is of utmost importance in digital libraries, the 36-dimensional model vectors as the indices of model images in digital libraries are clustered by use of vector quantization techniques. This clustering strategy reduces the searching space by order of magnitude. Desirable results have been observed from our experiments using the proposed color-object-indexing/retrieval algorithm.

In [4] In this work, we focus on automatic extraction of object boundaries from Canny edge field for the purpose of content-based indexing and retrieval over image and video databases. A multi-scale approach is adopted where each successive scale provides further simplification of the image by removing more details, such as texture and noise, while keeping major edges. At each stage of the simplification, edges are extracted from the image and gathered in a scale-map, over which a perceptual sub segment analysis is performed in order to extract true object boundaries. The analysis is mainly motivated by Gestalt laws and our experimental results suggest a promising performance for main objects extraction, even for images with crowded textural edges and objects with color, texture, and illumination variations. Finally, integrating the whole process as feature extraction module into MUVIS framework allows us to test the mutual performance of the proposed object extraction method and subsequent shape description in the context of multimedia indexing and retrieval. A promising retrieval performance is achieved, and especially in some particular examples, the experimental results show that the proposed method presents such a retrieval performance that cannot be achieved by using other features such as color or texture.

In [5] a novel local structural approach is explained, in this method for object retrieval in a cluttered and occluded environment without identifying the outlines of an object. It works by first extracting consistent and structurally unique local neighborhood from inputs or models and then voting on the optimal matches employing dynamic programming and a novel hypercube-based indexing structure. The proposed concepts have been tested on a database with thousands of images and compared with the six nearest-neighbors shape description with superior results.

In [6] they described an approach to generalize the concept of text-based search to non textual information. In particular, they elaborate on the possibilities of retrieving objects or scenes in a movie with the ease, speed, and accuracy with which Google retrieves web pages containing particular words, by specifying the query as an image of the object or scene. In this approach, each frame of the video is represented by a set of viewpoint invariant region descriptors. These descriptors enable recognition to proceed successfully despite changes in viewpoint, illumination, and partial occlusion.

Vector quantizing these region descriptors provides a visual analogy of a word, which we term a visual word. Efficient retrieval is then achieved by employing methods from statistical text retrieval, including inverted file systems, and text and document frequency weightings. The final ranking also depends on the spatial layout of the regions. Object retrieval results are reported on the full length feature films. They discussed three research directions for the presented video retrieval approach and review some recent work addressing them: 1) building visual vocabularies for very large-scale retrieval; 2) retrieval of 3-D objects; and 3) more thorough verification and ranking using the spatial structure of objects.

In [7] they have discussed on the effectiveness of region-based representation for content-based image retrieval. One common weakness of region-based approaches is that perform detection using low level visual features within the region and the homogeneous image

regions have little correspondence to the semantic objects. Thus, the retrieval results are often far from satisfactory. In addition, the performance is significantly affected by consistency in the segmented regions of the target object from the query and database images. Instead of solving these problems independently, this paper proposes region-based object retrieval using the generalized Hough transform (GHT) and adaptive image segmentation. The proposed approach has two phases. First, a learning phase identifies and stores stable parameters for segmenting each database image. In the retrieval phase, the adaptive image segmentation process is also performed to segment a query image into regions for retrieving visual objects inside database images through the GHT with a modified voting scheme to locate the target visual object under a certain affine transformation. The learned parameters make the segmentation results of query and database images more stable and consistent. Computer simulation results show that the proposed method gives good performance in terms of retrieval accuracy, robustness, and execution speed.

3. Comparison of the Conventional Approaches

In [3], Jie Wei proposed a color object indexing and retrieval scheme. Here they have taken the advantages of chromaticity, to achieve illumination invariant color object indexing and retrieval. Their indexing and retrieval is on compressed domain instead of uncompressed dataset. The entire object indexing and retrieval scheme is carried out by four simple steps. First perform a direct manipulation in compressed domain; second an invariant object indexing; third clustering of modeled objects and finally a newly proposed indexing and retrieval algorithm. This method is not immediately applicable in cluttered environment. Performance of the proposed method was evaluated on both Swain's image database and Assorted image databases. The proposed indexing / retrieval method works directly in the JPEG-compressed domain.

In [4], Serkan Kiranyaz et al focuses on extraction of object boundaries from Canny edge field for the purpose of content based indexing and retrieval over image and video databases. They proposed a systematic approach which Performs automatic object extraction by sub segment analysis over (object) boundaries in order to achieve visual content extraction. A multi-scale approach is adopted where each successive scale provides further simplification of the image by removing more details, such as texture and noise, while keeping major edges. At each stage of the simplification, edges are extracted from the image and gathered in a scale-map, over which a perceptual sub segment analysis is performed in order to extract true object boundaries. Between the query frame and a particular frame in the database, the feature vectors of all segments are used for similarity distance calculation with the following matching criteria: for each CL segment in the query frame, a "matching" CL segment in the compared frame is found. The analysis is mainly motivated by Gestalt laws and their experimental results suggest a promising performance for main objects extraction, even for images with crowded textural edges and objects with color, texture, and illumination variations.

In order to test the retrieval performance, they have indexed two image databases. The first database is the binary shape database, which contains 1400 binary images, is basically used to examine the effectiveness of the proposed feature extraction method. They observed a significant retrieval performance, i.e., 85% or higher Recall for 78% average Precision level. The second database is composed of 400 natural images carrying various content (i.e., sports, cars, objects, wild-life scenery, animals, outdoor city-scape, etc.). Among several retrieval experiments, they observed a range of 21%–78% Recall versus 8%–100% of Precision level. During the indexing of the natural image database bearing 400 images, five iterations (scales) are used and the average indexing time per image is calculated as 9.82 s. No iterations (0 scales) are used for the binary image database and the average indexing time per image is 1.21 s.

In [5], the proposed method aims to retrieve model objects that best match each of the overlapping objects in the query. The success of the system implies that shape analysis can be built on realistic input data instead of making the assumption of pre segmentation. The proposed system has three major parts. First, a technique has been found to extract unique local structure consistently when enough or distinguishable portion of the object can be observed. Second, sufficient and invariant features of each local structure have been derived to represent each structure. Third, an indexing and voting scheme have been designed to retrieve appropriate models with respect to the input. The proposed local structure has been demonstrated to be more effective than 6NN since it employs more than just the "nearest" attribute. The proposed system has been tested with the logo database from the work of Huet and

Hancock. The database consists of 2,711 simple characters and 984 complex trademarks and cartoons. The system performance was compared with the six-nearest neighbor (6NN) method since 6NN had been shown effectiveness for local structure representation. The comparison was conducted in terms of recall which is the ratio of the number of relevant items retrieved over the total number of relevant items available in the database. Each recall curve was generated with different scheme and formula.

In [6], they described an approach to generalize the concept of text-based search to non textual information. In particular, they elaborate on the possibilities of retrieving objects or scenes in a movie with the ease, speed, and accuracy with which Google retrieves web pages containing particular words, by specifying the query as an image of the object or scene. In this approach, each frame of the video is represented by a set of viewpoint invariant region descriptors. These descriptors enable recognition to proceed successfully despite changes in viewpoint, illumination, and partial occlusion. This method demonstrates an application of text -retrieval techniques for efficient visual search for objects in videos. Probabilistic models from statistical text analysis and machine translation have been also adapted to the visual domain in the context of object category recognition and scene classification.

In [7], they have discussed on the effectiveness of region-based representation for content-based image retrieval. One common weakness of region-based approaches is that perform detection using low level visual features within the region and the homogeneous image regions have little correspondence to the semantic objects. Thus, the retrieval results are often far from satisfactory. In addition, the performance is significantly affected by consistency in the segmented regions of the target object from the query and database images. Instead of solving these problems independently, this paper proposes region-based object retrieval using the generalized Hough transform (GHT) and adaptive image segmentation. The proposed approach has two phases. First, a learning phase identifies and stores stable parameters for segmenting each database image. In the retrieval phase, the adaptive image segmentation process is also performed to segment a query image into regions for retrieving visual objects inside database images through the GHT with a modified voting scheme to locate the target visual object under a certain affine transformation. This paper presented an object search method using the GHT based on content aware image segmentation.

4. Conclusion

This paper gives an idea about different conventional image object retrieval schemes and their performance measures. Among the discussed conventional approaches “Efficient Visual Search for Objects in Videos Visual search using text-retrieval methods can rapidly and accurately locate objects in videos despite changes in camera viewpoint, lighting, and partial occlusions” by Josef Sivic and Andrew Zisserman gives the best text object retrieval from a given video input.

5. References

[1] Chen, Zhiyong, Wei Zhang, Bin Hu, Xiaochun Can, Si Liu, and Dan Meng. "Retrieving Objects by Partitioning." IEEE Transactions on Big Data 3, no. 1 (2017): 44-54.

[2] Rebai, Ahmed, Alexis Joly, and Nozha Boujemaa. "BLasso for object categorization and retrieval: Towards interpretable visual models." Pattern Recognition 45.6 (2012): 2377-2389.

[3] Wei Jie. "Color object indexing and retrieval in digital libraries." IEEE transactions on image processing 11.8 (2002): 912-922.

[4] Kiranyaz, Serkan, Miguel Ferreira, and Moncef Gabbouj. "Automatic object extraction over multiscale edge field for multimedia retrieval." IEEE Transactions on Image Processing 15.12 (2006): 3759-3772.

[5] Maylor, K. H. "Part-Based Object Retrieval in Cluttered Environment."

[6] Sivic, Josef, and Andrew Zisserman. “Efficient visual search for objects in videos." Proceedings of the IEEE 96.4 (2008): 548-566.

[7] Chung, Chi-Han, Shyi-Chyi Cheng, and Chin-Chun Chang. "Adaptive image segmentation for region-based object retrieval using generalized Hough transform." Pattern recognition 43.10 (2010): 3219-3232.