Semantic Gap Reduction in Content Based Image Retrieval: A Survey

Nirali V Patel , Ghanshyam I Prajapati
PG Scholar (IT), SVMITBharuch, India
niralipatel2691@gmail.com
Department of IT, SVMITBharuch, India
giprajapati612@gmail.com

Abstract— The explosive growth of image data leads to the need of research and development of image retrieval. Content Based Image Retrieval (CBIR) is the approval image retrieval system by which the target image to be retrieved on the basis of useful features of the given image. So, in large collection, users of different domains face a problem of retrieving images relevant to the user query. To overcome this problem, different techniques are adopted. The significance of content based image retrieval system depends on the adopted features to represent images in the knowledge base. Using low-level features cannot give satisfactory results in many cases recovery; especially when high-level concepts in the user's mind are not easily expressible in terms of low-level features, i.e. semantic gap. The need to improve the precision of image retrieval systems and reduce the semantic gap is high in view of the growing need for image retrieval. We first present semantic extraction methods, and then the key technologies for reducing the semantic gap, i.e. object-ontology, machine learning, generating semantic relevance feedback templates and web image retrieval are discussed.

Keywords— Semantic Gap, Semantic Feature Extraction, Object Ontology, Relevance Feedback

Introduction

The rapid development of multimedia and communication technology has lead to increased demands for multimedia information. Efficient image and video retrieval tools are needed to hold such a large database. The efficient image retrieval and classification tools are needed by users from several domains like remote sensing, medicine, crime prevention etc. Therefore the need of efficient image retrieval tools to retrieve the images from large database is get an important part of the present time cutting edge technology. In text keyword establish approach first the images are manually annotated by a set of keywords, and then using these keywords the image retrieval is executed. There are two disadvantages with text based approach:

The manual keyword annotation process takes more than necessary time.
Annotation inaccuracy due to human perception subjectivity changes from person to person.

To get over these disadvantages of text keyword based approach, content based image retrieval (CBIR) was presented in early 1980s. In CBIR systems, images are indexed by their visual content like color, texture, and shape in the form of feature vectors rather than a set of keywords. Some previously developed CBIR systems are QBIC [2], Photobook [3], and Virage [4], which stands for image contents as a function of attributes of image, like color, texture and shape. In such systems the retrieval is done by comparing the set of features of a query image with the set of features of the images in the database. The CBIR systems which use only low level features for image retrieval can be conceived as the computer centric systems. The performance of such system is not that much given satisfaction because of two basic problems:

The semantic gap
The human perception subjectivity

The semantic gap is because of differs between the information that one can extract from the visual data and the interpretation in a real word situation [5]. Humans run to use high-level features called concepts such as keywords to understand contents of the images and to calculate their similarity among them. While the features extracted from image using image processing techniques are mostly low-level features such as color, texture, shape etc. There is no straight mapping between the high-level features and the low-level features, which bridges the gap between the features used by the humans and the computer system. Therefore even though there are much research attempts on the development of the CBIR systems, the performance of CBIR systems is still unequal because of the mismatch between the user semantic concepts and system generated low level image features. The second one is the human perception subjectivity. The human perception may differ from person to person. Different persons, or the same persons under the different circumstances, may comprehend the same image differently. In order to stamp down these problems, image retrieval systems should be directed to support high-level querying and browsing. A comprehensive Survey on CBIR can be elaborated in [6]-[7]. Liu [et.al] [8] laid out an elaborated survey on several methods for reduction of semantic gap in CBIR.

Related work

The subject of image retrieval has been an active research area for various decades and becomes progressively concerning area in recent years. Various reviews on image retrieval have been released.In 2008, Yu Xiaohong depicts that the basic components of content-based image retrieval system are color, texture, shape and semantic. The semantic-based image retrieval is a good way to resolve the “semantic gap” problem, so the semantic-based image retrieval method is emphasized in his paper. Other associate techniques such as relevance feedback and performance evaluation also discussed.In 2008, Prof. Sharvari Tamane, proposed a new system for image retrieval using high level features which is on the basis of taking out low level features such as color, shape and texture features and discourse them into high level semantic features using fuzzy production rules, by using image mining technique. Main benefit of this proposed method is the potential of retrieval using high level semantic features.In 2010, Lijun Zhao and Jiakui Tang compare different visual feature combinations in retrieval experiments. They depict that only low-level features for CBIR cannot attain a satisfactory measurement performance, since the user's high-level semantics cannot be easily explicated by low-level features. With a view to reduce the gap between user query concept and low-level features in CBIR, a multi-round relevance feedback (RF) strategy on the basis of both support vector machine (SVM) and feature similarity is followed to gain the user's requirement. This implementation can make better the performance by increasing the number of feedback.In 2012, Anuja Khodaskar proposed efficient content-based image retrieval (CBIR) systems are accurate characterization of visual information. Traditional image retrieval method has some restrictions**.** With a view to get the better the retrieval accuracy of content-based image retrieval systems, research direction has been shifted from designing sophisticated low-level feature extraction algorithms to reducing the semantic gap between the visual features and the richness of human semantics this paper introduce the technique for effective CBIR with high level semantic features by employing object ontology. Object-Ontology supplies a better quality definition of high-level query concepts. In 2013, Jisha.K.P, Thusnavis Bella Mary. I, Dr.A.Vasuki, concentrate on the semantic based image retrieval system using Gray Level Co-occurrence Matrix (GLCM) for texture feature extraction. The images are retrieved granting to user satisfaction and thereby reduce the semantic gap between low level features and high level features.

Semantic feature extraction

Feature Extraction is characterized by two procedures, Text Based Features and Visual Features. In the first system Keywords and annotations are separated and in the second method two sorts of components, General elements, for example, Color, Texture and Shape and Domain Specific Features, for example, human confronts and fingerprints can be removed. General components of the inquiry picture and different pictures put away in the database are removed in view of their numeric pixel values. These general components can be either extracted from the whole picture or from locales acquired by picture object's division. At that point the CBIR framework stores these pixel values in reduced structure into partitioned database known as feature database which is otherwise called Image Signature [5]. Eakins specified three levels of questions in CBIR [7].

Level 1: CBIR frameworks recover pictures by low level components, for example, shading, surface, shape or the spatial area. Cases of such questions may incorporate discover pictures with red shading protests or discover more pictures that resemble this picture. [7]

Level 2: Logical Features, for example, the Identity of the Items Shown are utilized for picture recovery. Examples of such queries might include “find an image of Ice" and “find an image of the Dense Forest”. To answer inquiries at this level, reference to some outside store of information is ordinarily needed. [7]

Level 3: Abstract attributes, for example, the importance of the scenes portrayed, including a lot of high level thinking about the significance and motivation behind the articles in a picture are utilized. Examples of such queries might be “find images of Scottish folk dancing” or “find a image depicting suffering". Accomplishment in noting inquiries at this level can hold complex reasoning. [7]

The most critical gap present is the semantic gap between levels 1 and 2. Level 1 framework can't look the semantic pictures viably with intelligent elements like picture of Land region having ice. Numerous low level feature extraction calculations have been created. Here we concentrate on these feature extraction strategies.

A. Color

Color is a vital visual variable. Taking into account the diverse applications diverse shading spaces are accessible, for example, RGB, LAB, LUV, HSV (HSL), YCrCb and the tint min-max difference (Gee). Be that as it may, among all models, the HSV (Hue, Immersion and Value) model is considered as the most suitable for speaking to color on the grounds that it is perceptual to the client [7]. In picture recovery framework, color histogram of inquiry picture is figured and after that contrasted and the color histograms of different pictures put away in the database and afterward those pictures are recovered whose shading histograms match those of the inquiry most nearly. Color histogram is represented to as the extent of pixels of every color represented to in the color [8]. At the question time, the client can either indicate color extent, for example, "red 35%" or determine a sample picture from which color histogram is figured. Swain and Ballard proposed a matching technique in 1991, which is known as histogram intersection. Sticker and Orengo proposed color moments system to supplant the color histogram which lessens the quantization impacts [6],[8]. It concentrates on the thought that color dissemination can be described by its moments. Smith and Chang proposed color set as an approximation to the color histogram, to execute it they changed over the (R, G, B) color space into a perceptually uniform space, for example, HSV, and afterward quantized the changed color space into M receptacles [1], [13].

B. Texture

Texture implies visual examples having homogeneity property furthermore; these can't come about because of a solitary shading or power. Cases of such surfaces are mists, trees, blocks, hair and fabric. Haralick et al. proposed in 1970 the co-occurrence matrix which is built in view of the introduction and separation between picture pixels. Through the co-occurrence matrix it is anything but difficult to figure the difference, coarseness, directionality and normality, periodicity, directionality and irregularity to comprehend the texture meaning. Smith and Chang proposed the new approach which utilizes the statistics (mean and variance) removed from the wavelet sub bands as the texture representation. Gross et al. proposed the new strategy which utilizes the wavelet transform with KL expansion and Kohonen maps to perform texture analysis. Thyagarajan et al. joined the wavelet transform with a co-occurrence matrix to exploit both statistics based furthermore, transform based texture analysis [13], [3].

C. Shape

Regular articles are principally perceived by their shape. A number of components normal for article shape are figured for each article recognized inside of each put away picture. For the most part, shape representations can be separated into two classes, boundary based and region based. The previous employments just the external limit of the shape while the last uses the whole shape region.[7] The best representation for these two classes are Fourier descriptor and moment variants. The primary thought of Fourier Descriptor is to utilize the Fourier transformed boundary as the shape feature. The principle thought of Moment invariants is to utilize region based moments, which are invariant to changes as the shape feature. Inquiries to the framework can be entered either as example image or as a sketch [13].

semantic gap reduction

Those retrieval systems which are based on low-level features do not match human recognition, and considered deficient and unpredictable. This is known as semantic gap [2]. These systems are not ready to discover images according to user's requirement such as "Discover pictures with rocks." Some researchers have attempted to fill the semantic gap by proposing new techniques. These techniques are explained below:

A. Object Ontology:

Object Ontology utilizes semantics of an image to characterize it. This technique, characterize different levels for assigning low-level image features. Every level characterizes the region attribute that is surely understood to people and considered as the intermediate level descriptor for an image [5]. For example, sky can be defined as topmost blue region. Similarly, blue region can be defined as "blue high", "blue medium", and "blue lo". Images can be classified into different categories by mapping such descriptors to high-level semantics based on our knowledge. For example "sky" can be defined as region of "light blue "color and "upper" spatial location. Quantization of shading and surface component is the key in such frameworks. To support semantic-based image retrieval, a more viable and broadly utilized approach to quantize color data is by color naming. Berk, Brownstone and Kaufman proposed a shading naming framework „CNS" which quantizes the tone values into set of essential hues, for example, red, orange, brown, yellow, green, blue and purple, black, grey and white. Contrasted with color, texture naming framework is not yet available.

B. Machine Intelligence

This technique is suitable for complex semantics however hard to actualize. This strategy utilizes two sorts of machine learning tools that are supervised or unsupervised, related low-level features with inquiry ideas. This technique predicts the estimation of creation on the premise of inputs that user give.
In supervised machine insight, number of pictures is gathered and on premise of section measure a paired classifier is prepared to distinguish semantic class name. Bayesian Classifier is a vital technique in which pictures database are naturally arranged into general sorts as indoor or open air (city/scene, and so forth.) Another technique is neural system as per which the client chooses 11 classes (ideas): block, cloud, hide, grass, ice, street, rock, sand, skin, trees and water. At that point a great deal of data is brought into the neural system classifiers to set up the connection between low level components of a picture and its abnormal state semantics. Choice tree is another procedure to get semantic elements. Choice tree is built on the bases of pertinence of pictures data to the question and after that utilized as a model to recognize pictures from databases into two classes: relevant and irrelevant. In unsupervised learning, the point is to describe how data is sorted out or gathered that user entered. No result; It just depicts the info's association data. In this technique, similar images are relegated to groups. Each group is appointed a name, which boosts the possibility of getting similar images of that particular group.

C. Relevance Feedback

Relevance feedback techniques work online and taking into account aims of user. Relevance feedback technique is not attainable in some domain yet this system includes user's correspondence. Relevance feedback mechanism lives up to expectations when the user enters the question as an image, picture or content. The system provides the initial results of user retrieval and make image as "relevant" and "irrelevant." Machine learning algorithm learns user input and selector gets other images [4]. The procedure is repeated until the user is fulfilled with the outcomes.
A typical scenario for RF in CBIR is:
(1) The system provides the initial results of retrieval through query, sketches, etc.
(2) User judges the above results concerning whether and to what degree they are relevant (positive samples)/irrelevant (negative samples) to the query.
(3) Machine learning algorithm is connected to learn user feedback. Then go back to (2).
(2) & (3) are repeated until the user is fulfilled by the outcomes.

D. Web Image Retrieval
This system has some specialized contrast from image retrieval in other application. Some extra data on the Web is accessible to ease semantic-based image retrieval. For instance, the URL of image file often has a clear hierarchical structure including some information about the image such as image category [12]. In addition, the HTML record additionally contains some valuable data in image title, ALT-

tag, the enlightening content surrounding the image, hyperlinks, and so on. Then again, such data can just clarify images to a certain extend. Existing Web image seeking, for example, Google and AltaVista hunt images based on textual evidences only. In spite of the fact that these methodologies can find numerous important images, they can't confirm whether the retrieved images really contain the query concepts so the retrieval precision is poor. The outcome is that users need to experience the whole rundown to locate the desired images. This is a tedious process as the returned results always contain different subjects which are combined. To improve Web image retrieval performance, researchers are attempting to consolidate the evidences from textual data and visual image contents.

E. Semantic Template

Semantic Template is an arrangement of general characteristics that are figured from the quantity of images stored in the database. It follows a link between the characteristics of high and low-level features. It usually characterized as a component of concept representative figured from a group of test images. Chang et al. presented semantic visual template (SVT) to interface the low-level image feature to high level. Semantic visual template (SVT) image feature links low-level to high level concepts for feature retrieval. To produce SVT, the user first characterizes the format for a particular idea by determining articles and their spatial and worldly limitations, the weights allotted to every feature for each object. This initial phase of consultation is given to the system. Through user interaction, the system eventually meets to a small set of queries that match the concept in the mind of the user i.e. Recall ratio maximizes. SVT generation relies on upon the connection with the user and requires user understanding top to bottom of image attributes.

PERFORMANCE ANALYSIS

Numerous techniques are accessible for measuring the execution of image retrieval systems. The most widely recognized assessment measures utilized as a part of image retrieval systems are precision and recall [9]. These are generally displayed as precision vs. recall graph. The standard meanings of these two measures are given by taking following equations.

Precision (P) is defined as the ratio of the number of relevant images retrieved to the number of total retrieved images [8].

$$\text{Precision} = \frac{\text{Number of Relevant Images Retrieved}}{\text{Total number of Images Retrieved}}$$

Recall (R) is defined as the number of retrieved relevant images over the total number of relevant images available in the database [8].

$$\text{Recall} = \frac{\text{Number of relevant images retrieved}}{\text{Numbet of relevnt images in the database}}$$

High precision implies that less relevant images are returned or more relevant images are retrieved, while high recall implies couples of relevant images are missed.

CONCLUSION

In this paper different semantic feature extraction techniques and strategies for reducing semantic gap are described. There are three traditional strategies to extract the image semantic feature. Firstly, semantic feature can be extracted in view of image processing and domain knowledge which contains three key procedures that are image segmentation, object recognition, and object relation analysis, Secondly, we can get image semantics from manual marks or human connection. Thirdly, we can remove semantics from external data, for example, document name, URL, content close to the image or metadata data. Systems for reducing semantic gap are outlined Ontology-based algorithms are anything but difficult to outline and are suitable to applications with simple semantic features. Machine learning methods are utilized to learn more complex semantics. Because of its simplicity in usage and the unconstrained mapping from low-level features to high-level concepts utilizing decision tree which is a very effective tool for image retrieval. RF has been ended up being effective in expanding image retrieval precision. The issue is that most current systems requires around five or considerably more iterations before it merges to a stable execution level, yet users are irritated and may surrender after a few tries. Semantic templates utilizes qualities separated from a group of similar images which simplicity to retrieve image of same kind. Web image retrieval utilizes extra information to encourage the retrieval handled turned into an active research area.

REFERENCES

[1] A. Smeulders, et al., “Content-based image retrieval at the end of the early years”, IEEE Trans. PAMI, 22(12): pp. 1349-1380, 2000.
[2] B. Luo, X. Wang, and X. Tang, “A world wide web based image search engine using text and image content features,” in *IS&T/SPIE Electronic Imaging, Internet Imaging IV*, 2003.
[3] J. Cui, F. Wen, and X. Tang, “Real time google and live Image search re-ranking,” in *ACM Multimedia*, 2008.
[4] ——, “Intentsearch: interactive on-line image search reranking,” in *ACM Multimedia*, 2008.
[5] X. Tang, K. Liu, J. Cui, F. Wen, and X. Wang, “Intentsearch: Capturing user intention for one-click internet image search,” *T-PAMI*, vol. 34, no. 7, pp. 1342 – 1353, july 2012.
[6] F. Jing, C. Wang, Y. Yao, K. Deng, L. Zhang, and W.-Y. Ma, “Igroup: web image search results clustering,” in *ACM Multimedia*, 2006.
[7] S. Wang, F. Jing, J. He, Q. Du, and L. Zhang, “Igroup: Presenting web image search results in semantic clusters,” in *ACMCHI*, 2007.
[8] X. Wang, K. Liu, and X. Tang, “Query-specific visual Semantic spaces for web image re-ranking,” in *CVPR*, june 2011.
[9] Li, Qingyong, Hu, Hong, and Shi, Zhongzhi, “Semantic Feature Extraction Using Genetic Programming in Image Retrieval”, IEEE, Proceedings of the 17th International

Conference on Pattern Recognition (ICPR‟04).
[10] Eidenberger, Horst, and Breiteneder, Christian, "Semantic Feature Layers in Content-based Image Retrieval: Implementation of Human World Features", Vienna University of Technology, Institute of Software Technology and Interactive Systems, Favoritenstrasse 9 11 – 188/2, A-1040 Vienna, Austria.
[11] Chen, Zhijun, "Semantic Research on Content-Based Image Retrieval", IEEE, 2010.
[12] Kaur, Harpreet, and Jyoti, Kiran, "Survey of Techniques of High Level Semantic Based Image Retrieval", International Journal of Research in Computer and Communication technology, IJRCCT, ISSN 2278-5841, Vol 2, Issue 1, January 2013.
[13] Pattanaik, Swapnalini, and Bhalke, D.G., "Beginners to Content Based Image Retrieval", International Journal of Scientific Research Engineering &Technology (IJSRET), Volume 1 Issue2, pp 040-044, May 2012.
[14] Deb, Sagarmay, and Zhang, Yanchun. "An Overview of Content-based Image Retrieval Techniques", IEEE, 18th International Conference on Advanced Information Networking and Application (AINA‟04), 2004.