ISSN: 2277-9655
Impact Factor: 1.852

IJESRT

INTERNATIONAL JOURNAL OF ENGINEERING SCIENCES & RESEARCH TECHNOLOGY

A Framework For Annotating Images and its Respective Tags

A. Vijay Vasanth[*1], R. Premnivas[2], S. Rajalakshmi[3]

[*1]Senior Assistant Professor, Department of Computer Science and Engineering,, Christ College of Engineering & Technology, Pondicherry University, Pondicherry, India

[2, 3]M.Tech final year Students, Department of Computer Science and Engineering, Christ College of Engineering & Technology, Pondicherry University, Pondicherry, India

vizeyvasanth@yahoo.com

Abstract

The vast resource of pictures available on the web and the fact that many of them naturally co-occur with topically related documents and are captioned we focus on the task of automatically generating captions for images, here the model learns to create captions from a database of news articles, and the pictures embedded in them, and their captions, and consists of two stages. Content selection identifies what the image and accompanying article are about, whereas surface realization determines how to verbalize the chosen content. We approximate content selection with a probabilistic image annotation model that suggests keywords for an image. In the Proposed system extensive features are extracted from the database images and stored in the feature library. The extensive features set is comprised of shape features along with the color, texture and the contour let features, which are utilized in this work. When a query image is given, the features are extracted in the similar fashion. Subsequently, GA-based similarity measure is performed between the query image features and the database image features.

Keywords: Image Content, Tag, Retrieval, Query

Introduction

Image processing is any form of signal processing for which the input is an image, such as a photograph or video frame; the output of image processing may be either an image or a set of characteristics or parameters related to the image. Most image-processing techniques involve treating the image as a two-dimensional signal and applying standard signal-processing techniques to it.

Image processing is a method to convert an image into digital form and perform some operations on it, in order to get an enhanced image or to extract some useful information from it. It is a type of signal dispensation in which input is image, like video frame or photograph and output may be image or characteristics associated with that image. Usually Image Processing system includes treating images as two dimensional signals while applying already set signal processing methods to them.

It is among rapidly growing technologies today, with its applications in various aspects of a business. Image Processing forms core research area within engineering and computer science disciplines too.

Image processing basically includes the following three steps.

- Importing the image with optical scanner or by digital photography.
- Analyzing and manipulating the image which includes data compression and image enhancement and spotting patterns that are not to human eyes like satellite photographs.

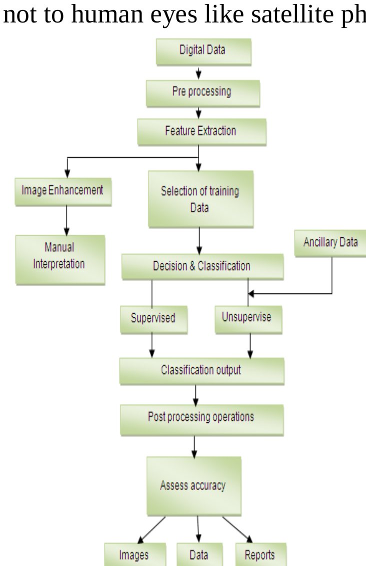

Fig 1: Image Formation

Literature Survey

Genetic Algorithms

Genetic Algorithms (GAs) are adaptive heuristic search algorithm based on the evolutionary ideas of natural selection and genetics. As such they represent an intelligent exploitation of a random search used to solve optimization problems. Although randomized, GAs are by no means random, instead they exploit historical information to direct the search into the region of better performance within the search space. The basic techniques of the GAs are designed to simulate processes in natural systems necessary for evolution, especially those follow the principles first laid down by Charles Darwin of "survival of the fittest.". Since in nature, competition among individuals for scanty resources results in the fittest individuals dominating over the weaker ones.
GAs simulates the survival of the fittest among individuals over consecutive generation for solving a problem. Each generation consists of a population of character strings that are analogous to the chromosome that we see in our DNA. Each individual represents a point in a search space and a possible solution. The individuals in the population are then made to go through a process of evolution.

GAs are based on an analogy with the genetic structure and behavior of chromosomes within a population of individuals using the following foundations:

- Individuals in a population compete for resources and mates.
- Those individuals most successful in each 'competition' will produce more offspring than those individuals that perform poorly.
- Genes from `good' individuals propagate throughout the population so that two good parents will sometimes produce offspring that are better than either parent.

Thus each successive generation will become more suited to their environment. The purpose of edge detection in general is to significantly reduce the amount of data in an image, while preserving the structural properties to be used for further image processing. Several algorithms exists, and this worksheet focuses on a particular one developed by John F. Canny (JFC) in 1986 . Even though it is quite old, it has become one of the standard edge detection methods and it is still used in research. The aim of JFC was to develop an algorithm that is optimal with regards to the following criteria:
1. **Detection:** The probability of detecting real edge points should be maximized while the probability of falsely detecting non-edge points should be minimized. This corresponds to maximizing the signal-to-noise ratio.
2**. Localization:** The detected edges should be as close as possible to the real edges.
3. **Number of responses:** One real edge should not result in more than one detected edge (one can argue that this is implicitly included in the first requirement). With JFC's mathematical formulation of these criteria, Canny's Edge Detector is optimal for a certain class of edges (known as step edges).The images used throughout this worksheet are generated using this implementation.

Briefly, Alpaydin mentions optical character recognition, speech recognition, and encoding/decoding as example applications of k-means (Alpaydin). However, a survey of the current literature on the subject offers a more in depth treatment of some other practical applications, such as "data detection … for burst-mode optical receiver" (Zhao, et al.), and recognition of musical genres (Turnbull and Elkan), which are specialized examples of what Alpaydin mentions.

As Zhao, Nehorai, and Porat describe "burst-mode data-transmission systems," a "significant feature of burst-mode data transmissions is that due to unequal distances between" sender and receivers, "signal attenuation is not the same" for all receivers. Because of this, "conventional receivers are not suitable for burst-mode data transmissions." The importance, they note, is that many "high-speed optical multi-access network applications, [such as] optical bus networks and WDMA optical star networks" can use burst-mode receivers (Zhao).

In their paper, they provide a "new, efficient burst-mode signal detection scheme" that utilizes "a two-step data clustering method based on a K-means algorithm." They go on to explain that "the burst-mode signal detection problem" can be expressed as a "binary hypothesis," determining if a bit is 0 or 1. Further, although they could use maximum likelihood sequence estimation (MLSE) to determine the class, it "is very computationally complex, and not suitable for high-speed burst-mode data transmission." Thus, they use an approach based on k-means to solve the practical problem where simple MLSE is not enough (Zhao).

The paper from Douglas Turnbull and Charles Elkan is quite a bit more concise, and not as full of jargon as the Zhao article, so its example may prove more meaningful. In it, they show that although "the classification of music by genre is difficult to automate," they are able to "achieve human-level accuracy with fast training and classification" by using "radial basis function (RBF) networks using a combination of unsupervised and supervised initialization methods." Furthermore, they note that the initialization methods they use are "hundreds of times" faster than using "RBF networks trained with gradient descent." They are able to achieve this goal in part with help from "an improved method for initializing the k-means clustering algorithm which is useful for both unsupervised and supervised initialization methods" (Turnbull).

The new approach Turnbull and Elkan use to initialize k-means is what they call Subset Furthest First (SFF). They note that a problem with (normal) Furthest First is "that it tends to find the outliers in the data set." By using only a subset, they found that there are less outliers that can be found, and "thus, the proportion of no outlier points obtained as centers is increased" (Turnbull).

Existing System

CBIR takes an image as a query and identifies the matched images based on the visual similarity between the query image and gallery images. Various visual features, including both global features (e.g., color, texture, and shape) and local features.

TBIR represents the visual content of images by manually assigned keywords/ tags. It allows a user to present his/her information need as a textual query and find the relevant images based on the match between the textual query and the manual annotations of images.

We represent the relation between tags and images by a tag matrix, where each row corresponds to an image and each column corresponds to a tag. Each entry in the tag matrix is a real number that represents the relevance of a tag to an image. Similarly, we represent the partially and noisy tagged images by an observed tag matrix, where an entry is marked as 1 if and only if image is annotated by keyword tag.

Proposed System

The project aims at developing a new TBIR system works, in two processes namely On-line and Off-line.

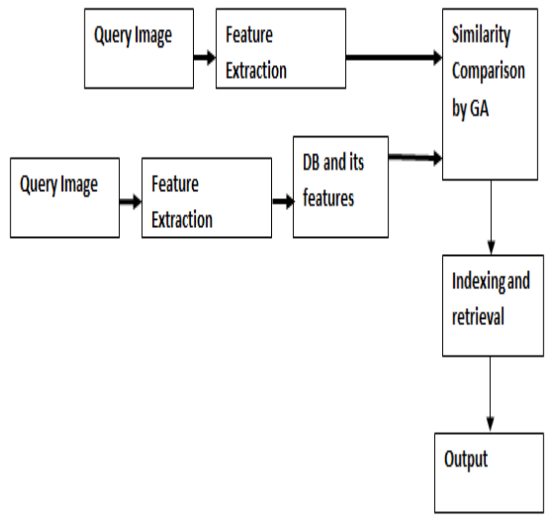

Fig 2: Proposed Architecture

The functions in off-line process are (i) image collection from standard image databases such as Flicker dataset, (ii) extraction of features such as regions and mean value of Red(R), Green (G) and Blue (B) components of an image and (iii) store the images and its features into the database. In On-line process, the user gives the query image for retrieval. This system identifies contour regions and extracts mean value of R, G and B components of the query image. A heuristics approach along with GA is used to compute the similarity between the query and candidate images in the database. Finally, the resultant images are displayed.

Methodology

There are four modules involve in this proposed system. The modules are as follows

- Automatic tagging
- Feature Extraction
- Shape Feature Extraction
- GA based Similarity Measure

A. Automatic Tagging

Given a database of images with some initially assigned tags, the proposed algorithm first generates a tag matrix denoting the relation between the images and initially assigned tags. It then automatically completes the tag matrix by updating the relevance score of tags to all the images. The completed tag matrix will be used for tag based image search or image similarity search.

B. Feature Extraction

Color is an important dimension of human visual perception that allows discrimination and recognition of visual information. Color features are relatively easy to extract and match, and have been found to be effective for indexing and searching of color images in image databases. Here we have considered the mean values of the pixel colors as the color feature of an image. Thus, each pixel of a color image is represented by a vector in Pi = [Ri Gi Bi].

C. Shape Feature Extraction

Shape representation generally looks for effective ways to capture the essence of the shape features that make it easier for a shape to be stored, transmitted, compared against and recognized. These features must also be independent of translation, rotation, and scaling of the shape. To extract the shape features from the image, initially the image in RGB color space so converted to gray scale image. Then median filter is applied to the converted gray scale image to remove the noises. The noise free image is subjected to clustering to detect different shapes present in the image'**s-** means clustering treats each object as having a location in space

D. GA based Similarity Measure

Genetic algorithms (GAs) within the field of evolutionary computation are robust computational and search procedures modeled on the mechanics of natural genetic systems. GA has the ability to improve search performance and they have the following advantages over traditional search methods: (i) directly work with a coding of the parameter set; (ii) search process is carried out from a population of points; (iii) probabilistic transition rules are used instead of deterministic ones. The fitness is based on the SED between the query image and the generated chromosome's feature sets.

Implementation & Experimental Results

The implementation part defines the short communication about the concept.

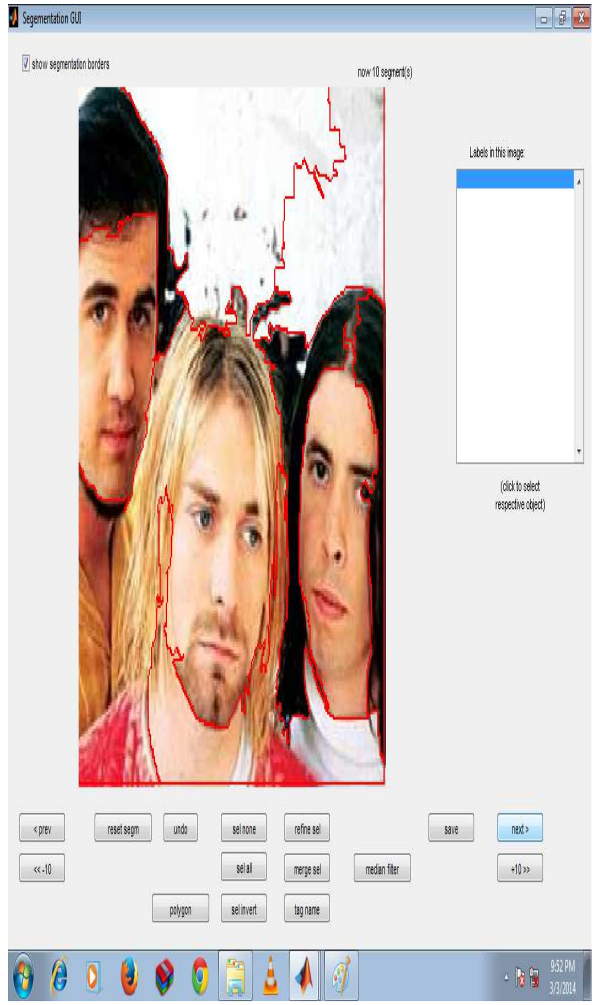

Fig 3: Annotation of Images

It first generates a tag matrix denoting the relation between the images and initially assigned tags. It then automatically completes the tag matrix by updating the relevance score of tags to all the images. The completed tag matrix will be used for the retrieval purpose.

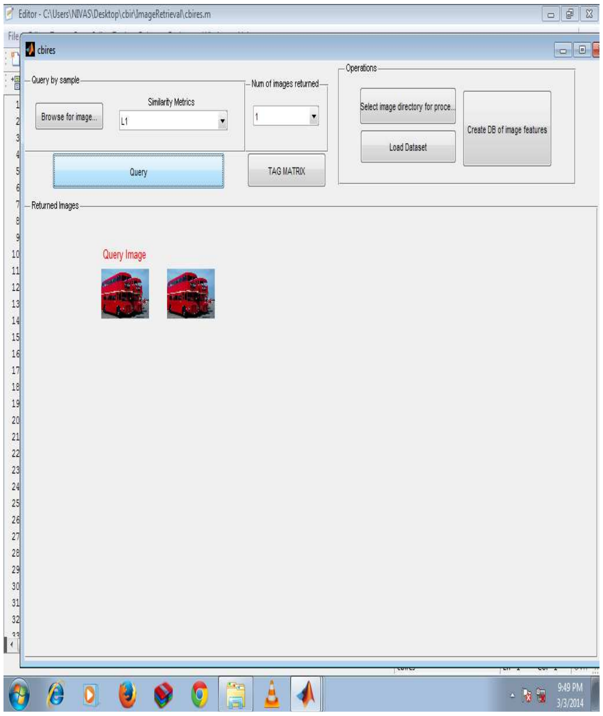

Fig 4: Image Retrieval

It finds portions such that objects within each cluster are as close to each other as possible and as far from objects in other clusters as possible. The image is converted from 2D vector to 1 D vector. After the K-means algorithm again the 1D vector is converted to 2 D vectors and then the canny algorithm is used for the detection of edges present in all the clustered sets of the images. Canny algorithm consists of five steps: smoothing, finding gradients, on maximum suppression. Double thersholding and edge tracking by hysteresis.

Conclusion

The main contributions of this work are the identification of the problem in existing work. In this project there are two processes namely On-line and Off-line process which is used as storing and retrieving and the full image annotation will be listed and displaying the object which presented in the content and by using the genetic algorithm is to improve the search performance and that will provide relevant images.

References

[1] L.Wu, R. Jin and A.K. Jain"Tag completion for image retrieval" ieee transactions on pattern analysis and machine intelligence, vol. 35, no. 3, march 2013.

[2] C. Desai, D. Ramanan, and C.C. Fowlkes, "Discriminative Modelsfor Multi-Class Object Layout," Int'l J. Computer Vision, vol. 95,PP.1-12,2011

[3] L. Wu, L. Yang, N. Yu, and X.-S. Hua, "Learning to Tag,"Proc.18th Int'l Conf. World Wide Web,pp. 361-361, 2009.

[4] L. Yang, R. Jin, L. Mummert, R. Sukthankar, A. Goode, B. Zheng,S.C.H. Hoi, and M. Satyanarayanan, "A Boosting Framework forVisuality-Preserving Distance Metric Learning and Its Application to Medical Image Retrieval," IEEE Trans. Pattern Analysis andMachine Intelligence,vol. 32, no. 1, pp. 30-44, Jan. 2010.

[5] M.E.I. Kipp and G.D. Campbell, "Patterns and Inconsistencies in Collaborative Tagging Systems: An Examination of Tagging Practices,"Ann. General Meeting Am. Soc. for Information Scienceand Technology,vol. 43, pp. 1-18, 2006.

[6] M. Guillaumin, T. Mensink, J. Verbeek, and C. Schmid, "TagProp:Discriminative Metric Learning in Nearest Neighbor Models forImage Auto-Annotation," Proc. 12th IEEE Int'l Conf. ComputerVision,pp. 309-316, 2009.

[7] M.-L. Zhang and Z.-H. Zhou, "ML-LNN: A Lazy Learning Approach to Multi-Label Learning,"Pattern Recognition,vol. 40,pp. 2038-2048, 2007.

[8] M.S. Lew, "Content-Based Multimedia Information Retrieval:State of the Art and Challenges," ACMTrans.Multimedia Computing, Comm. and Applications,vol. 2, pp. 1-19, 2006.

[9] N.Zhou,W.K.Cheung,G.Qiu,andX.Xue,"AH ybrid Probabilistic Model for Unified Collaborative and Content-Based Image Tagging," IEEE Trans. Pattern Analysis and Machine Intelligence,vol. 33, no. 7 pp. 1281-1294, July 2011.

[10] P. Wu, S.C.-H. Hoi, P. Zhao, and Y. He, "Mining Social Images with Distance Metric Learning for Automated Image Tagging,"Proc. Fourth ACM Int'l Conf. Web Search and Data Mining,pp. 197-206, 2011.

[11] R. Datta, W. Ge, J. Li, and J.Z. Wang, "Toward Bridging the Annotation-Retrieval Gap in Image Search by a Generative Modeling Approach,"Proc. 14th Ann. ACM Int'l Conf. Multimedia,pp. 977-986, 2006.

[12] S.C.H. Hoi, W. Liu, M.R. Lyu, and W.-Y. Ma, "Learning Distance Metrics with Contextual Constraints for Image Retrieval,"Proc.IEEE Conf. Computer

ISSN: 2277-9655
Impact Factor: 1.852

Vision and Pattern Recognition,pp. 2072-2078,2006.
[13] S.L. Feng, R. Manmatha, and V. Lavrenko, “Multiple Bernoulli
[14] S. Shalev-Shwartz and N. Srebro, “SVM Optimization: Inverse Dependence on Training Set Size,”Proc. 25th Int’l Conf. Machine Learning,pp. 928-935, 2008.
[15] X. Li, C.G.M. Snoek, and M. Worring, “Learning Social Tag Relevance by Neighbor Voting,”IEEE Trans. Multimedia, vol. 11,no. 7, pp. 1310-1322, Nov. 2009.