ISSN: 2277-9655
Impact Factor: 1.852

IJESRT

INTERNATIONAL JOURNAL OF ENGINEERING SCIENCES & RESEARCH TECHNOLOGY

Review of Image Classification Techniques Based on LDA, PCA and Genetic Algorithm

Mukul Yadav[*1], Gajendra Singh Chandel[2], Ravindra Gupta[3]
[*1,2,3] Department of Computer Science & Engineering, S.S.S.I.S.T, Sehore, M.P, India
Mukulyadav86@gmail.com

Abstract

Image classification is play an important role in security surveillance in current scenario of huge Amount of image data base. Due to rapid change of feature content of image are major issues in classification. The image classification is improved by various authors using different model of classifier. The efficiency of classifier model depends on feature extraction process of traffic image. For the feature extraction process various authors used a different technique such as Gabor feature extraction, histogram and many more method on extraction process for classification. We apply the FLDA-GA for improved the classification rate of content based image classification. The improved method used heuristic function genetic algorithm. In the form of optimal GA used as feature optimizer for FLDA classification. The normal FLDA suffered from a problem of core and outlier problem. The both side kernel technique improved the classification process of support vector machine.FLDA perform a better classification in compression of another binary multi-class classification. Directed acyclic graph applied a graph portion technique for the mapping of feature data. The mapping space of feature data mapped correctly automatically improved the voting process of classification.

Keywords: Image Classification, FLDA, GA, PCA.

Introduction

The goal of content-based image retrieval is to retrieve the images that users want to search; it has become a research hot issue in recent decades. However, the performance of the state-of-art CBIR systems is still unsatisfactory due to semantic gap between image low level feature and user's image perception. Although many efforts have been carried out to bridge the semantic gap, the improvement gained is unsatisfactory. Researchers argue that the key to effectively improve CBIR performance lies in the ability to access the image at the level of objects, the reason why the users retrieve image is to search for images containing particular objects of interest, so the ability to represent, index and match images at the level of objects is critical. Therefore, region-based image representation is a feasible strategy to precisely index and retrieve images at object level. As conventional color features used in CBIR, there are color histogram, color correlogram, and dominant color descriptor (DCD) [1]. Color histogram is the most commonly used color representation, but it does not include any spatial information. Color correlogram describes the probability of finding color pairs at a fixed pixel distance and provides spatial information. Therefore color correlogram yields better retrieval accuracy in comparison to color histogram. Color auto correlogram is a subset of color correlogram, which captures the spatial correlation between identical colors only. Since it provides significant computational benefits over color correlogram, it is more suitable for image retrieval. Image classification is basically the task of classifying the number of images into (semantic) categories based on the available training data. The objective of digital image classification procedure is to categorize the pixels in an image into land over cover classes [7]. The output is thematic image with a limited number of feature classes as opposed to a continuous image with varying shades of gray or varying colors representing a continuous range of spectral reflectance. The range of digital numbers in different bands for particular features is known as a spectral pattern or spectral signature. A spectral pattern can be composed of adjacent pixels or widely separated pixels. Digital image classification technique can generally be classified into two types: Unsupervised classification Techniques and Supervised classification Techniques. Classification approaches deal poorly on content based image classification tasks being one of the reasons of high dimensionality of the feature space. Most proposed Content based color image classification techniques automatically extract low-level features (e.g. color, texture, shapes and layout of objects) to measure the similarities among images by comparing their feature

differences[9]. Although, color, texture and shape features have been used for describing image content. Color is the most widely and popularly used low-level visual features and is invariant to image size and orientation. Color correlogram defines the probability of finding color pairs at a fixed pixel distance and provides spatial information. Hence color correlogram contains a better retrieval accuracy in comparison to color histogram. On the other hand color auto correlogram is a subset of color correlogram, which captures the spatial correlation between identical colors only. In Section II related work. The Section III discusses Problem formulation IV Used Approach followed by a conclusion in Section V.

Related Work

[1] In this title the author explained that an adaptive fuzzy c-means algorithm was developed and applied to the segmentation and classification of multicolour fluorescence in situ hybridization (M-FISH) images , which can be used to detect chromosomal abnormalities for cancer and genetic disease diagnosis. The algorithm improves the classical fuzzy c-means algorithm (FCM) by the use of a gain field, which models and corrects intensity in homogeneities caused by a microscope imaging system, flairs of targets (chromosomes), and uneven hybridization of DNA. Other than directly simulating the in homogeneously distributed intensities over the image, the gain field regulates centres of each intensity cluster.

[2] In this title author explained that an improvement to the NBNN image classification algorithm that increases classification accuracy and improves its ability to scale to large numbers of object classes. The key observation is that only the classes represented in the local neighbourhood of a descriptor contribute significantly and reliably to their posterior probability estimates. Instead of maintaining a separate search structure for each class's training descriptors, we merge all of the reference data together into one search structure, allowing quick identification of a descriptor's local neighbourhood. They showed an increase in classification accuracy when we ignore adjustments to the more distant classes and show that the run time grows with the log of the number of classes rather than linearly in the number of classes as did the original.

[6] in this title author explained the Wavelet based Multi Class image classification using Neural Network, A feature extraction and classification of multiclass images by using Haar wavelet transform and back propagation neural network. The wavelet features are extracted from original texture images and corresponding complementary images. The features are made up of different combinations of sub-band images, which offer better discriminating strategy for image classification and enhance the classification rate.

[4] In this title author explained text-image co-occurrence data become available on the web mining on those data is playing an increasingly important role in web applications they consider utilizing description information to help image classification and propose a novel image classification method focusing on text-image co-occurrence data. In general, there are three main steps in our system: feature extraction, training classifiers and classifier fusion. In feature extraction phase, several features are extracted including not only visual features such as color, shape, texture, but also text features. In the process of training classifiers, visual and text classifiers are trained separately with SVM model [10].

[5] In this title author explained that automatic image annotation is an important but highly challenging problem in semantic-based image retrieval. they formulate image annotation as a supervised learning image classification problem under region-based image annotation framework. In region-based image annotation, keywords are usually associated with individual regions in the training data set. This research applies a novel simple decision tree (SDT) algorithm in our image annotation system, which can classify a large number of training data faster and more effectively. The proposed SDT algorithm is experimented on image annotation Corel data sets. Compared to classical algorithms, SDT accelerates the operation speed of the algorithm, and the classification accuracy remains robustness [4].

[3] SNMFCA: Supervised NMF-based Image Classification and Annotation A novel supervised nonnegative matrix factorization based framework for both image classification and annotation (SNMFCA). The framework consists of two phrases: training and prediction. In the training phrase, two supervised nonnegative matrix factorizations for image descriptors and annotation terms are combined together to identify the latent image bases, and represent the training images in the bases space. These latent bases can capture the representation of the images in terms of both descriptors and annotation terms. Based on the new representation of training images, classifiers can be learnt and built.

[9] In this title author explained a novel approach for content based color image classification using Support Vector Machine (SVM). Traditional classification approaches deal poorly on content based image classification tasks being one of the reasons of high dimensionality of the feature space. In this paper, color image classification is done on features extracted from histograms of color components. The benefit of using color image histograms are better efficiency, and insensitivity to small changes in camera view-point i.e.

ISSN: 2277-9655
Impact Factor: 1.852

translation and rotation. As a case study for validation purpose, experimental trials were done on a database of about 500 images divided into four different classes has been reported and compared on histogram features for RGB, CMYK, Lab, YUV, YCBCR, HSV, HVC and YIQ color spaces. Results based on the proposed approach are found encouraging in terms of color image classification accuracy.

[7] in this paper author explained the Local Naive Bayes Nearest Neighbor for Image Classification An improvement to the NBNN image classification algorithm that increases classification accuracy and improves its ability to scale to large numbers of object classes. The key observation is that only the classes represented in the local neighbourhood of a descriptor contribute significantly and reliably to their posterior probability estimates.

[8] in thie title author focus on Research on Image Classification Based on a Combination of Text and Visual Features A text-image co-occurrence data become available on the web, mining on those data is playing an increasingly important role in web applications. Utilizing description information to help image classification and propose a novel image classification method focusing on text-image co-occurrence data. In general, there are three main steps in our system: feature extraction, training classifiers and classifier fusion.

Problem Formulation

Image Classification and retrieval is current research trend in computer vision. The application of image classification used in various field such as remote sensing, Photo Gallery and Medical diagnosis. In concern of classification, the rate of classification depends on the feature attributes of image data and depends on behavior of classifier. In process of survey study paper and journal of image classification based on various data mining approach and neural network. Some Classification based on binary classes and some other one are multilevel classification. The binary classification generate a issue for MSE(Mean Square Error) and degrade the rate of prediction in classification such as Decision Tree , Navie Bayes etc. In another approach multilevel classification generates large number of confusion matrix and suffered rate of classification. we are motivated for doing work in the field of image classification and improved the rate of classification. Following problems are minimized in our dissertation with the help of RBF Network.

1. Semantic Gap between Input Image and Classified Image
2. Core Point and Outlier Problem in DAG-SVM
3. Feature Extraction and Negative Query Processor

Approach Used

In this paper we Proposed an improved technique for fusion based image classification. In process of improved classification technique is proposed and we have used LDA technique. The Linear discriminate analysis is kernel based classifier, here kernel play a role of hyper plane. The size and efficiency of hyper plane decide the efficiency of classification. In classification process Linear discriminate analysis suffered two types of problem in region classification one is core point problem and another is outlier of feature point, in single class classification. In the process of image classification the lower content of visual feature such as color texture and dimensions. The feature extractor process extracts the feature of image database and store in the form of matrix. The processing of feature mapping convert into vector form for processing of fused property of classifier. The process of feature selection based on genetic algorithm. The optimizations of feature selection process is improved the classification rate of fisher linear discriminate analysis. The process of classification discuss in this section is LDA, FLDA and kernel based FLDA classifier. The kernel based FLDA classifier used Gaussian kernel for classification analysis. The selection process of feature in fused method is used genetic algorithm.

Linear Discriminant Analysis (LDA):

The Linear Discriminant Analysis originally designed by Fisher for taxonomic classification. LDA searches for those vectors in the underlying space that best discriminate among classes (rather than those that best describe the data) [11]. More formally, given a number of independent features relative to which the data is described, LDA creates a linear combination of these which yields the largest mean differences between the desired classes. It tries to find an optimal reducing-dimensionality linear projection that maximizes the scatter of all projected samples. However, for classification, the between class scatter should be maximized, while the within-class scatter should be minimized. if a data set is categorized, it makes sense to use the class information to build a more desirable projection space to improve discrimination while reducing the dimensionality of the feature space. LDA is an example of a class specific method, in the sense that it tries to “shape” the scatter in order to make it more favorable for classification. This method seeks the projections that maximize the ratio of the between-class scatter to the within-class scatter in the projection space.

ISSN: 2277-9655
Impact Factor: 1.852

Principal Component Analysis (PCA)

Principal Component Analysis is one of the simplest and common statistical methods to reduce high-dimensional data. It consists of a transformation from the original variables into a new set of uncorrelated variables, called principal components (PCs). These new variables are linear combinations of the original ones ranked in decreasing order of importance and the number of PCs is determined by the cumulative variance explained by the first components. If the original variables are highly correlated, then the first few PCs will account for most of the variation and the remaining PCs can be discarded with slight loss of information [12]. The more PCs are used the better the reconstructed data fits the original noisy data. Given a n-dimensional vector representation of each image, the PCA method can be used to find a subspace whose basis vectors correspond to the maximum-variance directions in the original space . All images are projected onto the subspace to find a set of weights that describes the contribution of each vector. An unknown image can be identified by two steps: First, project that image onto the subspace to obtain its set of weights; second, compare set of weights of the unknown image to set of weights of known images.

Genetic Algorithms

For the process of separation of class of image feature separation for generation of FLDA mining by RBF classification, this classification whole class in two sections, in one section we classified only higher semantic vale and another section of class contain lower value of class. The process of searching of data according to given feature set of data set we used genetic algorithm for better searching of classified class and finally generated optimized feature. Here we discuss process of genetic algorithm.

Genetic Algorithm (GA), first introduced by John Holland in the early seventies, is the powerful stochastic algorithm based on the principles of natural selection and natural genetics, which has been quite successfully, applied in machine learning and optimization problems. To solve a Problem, a GA maintains a population of individuals (also called strings or chromosomes) and probabilistically modifies the population by some genetic, operators such as selection, crossover and mutation, with the intent of seeking a near optimal solution to the problem. Coding to Strings in GA[5,6], each individual in a population is usually coded as coded as a fixed-length binary string. The length of the string depends on the domain of the parameters and the required precision. For example, if the domain of the parameter x is [2,5] and the precision requirement is six places after the decimal point, then the domain [2,5] should be divided into 7,000,000 equal size ranges.

Conclusion and Future Scope

In this paper we discuss content based image retrieval using image classification technique. we proposed a hybrid method for image classification. Our experimental result shows that better result in compression of old and traditional method of image classification. FLDA-GA improved the classification rate of content based image classification. The improved method used heuristic function genetic algorithm. In the form of optimal GA used as feature optimizer for FLDA classification. The normal FLDA suffered from a problem of core and outlier problem. The both side kernel technique improved the classification process of support vector machine.

FLDA perform a better classification in compression of another binary multi-class classification. Directed acyclic graph applied a graph portion technique for the mapping of feature data. The mapping space of feature data mapped correctly automatically improved the voting process of classification. But Directed acyclic graph suffered a little bit problems with mapping of space data into feature selection process. Performance of result evaluation shows that our FLDA-GA is better classifier in compression of SVM.

References

[1] *H Cao, HW Deng, YP Wang "Segmentation of M-FISH Images for Improved Classification of Chromosomes With An Adaptive Fuzzy C-means Clustering Algorithm " IEEE 2012. Pp 235-244.*

[2] *Sancho McCann David G. Lowe "Local Naive Bayes Nearest Neighbor for Image Classification" IEEE 2011. Pp 887-896.*

[3] *Liping Jing, Chao Zhang and Michael K. Ng "SNMFCA: Supervised NMF-based Image Classification and Annotation" in IEEE 2011. Pp 356-361.*

[4] *L Tian, D Zheng, C Zhu ". Research on Image Classification Based on a Combination of Text and Visual Features" IEEE 2011. Pp 458-464.*

[5] *Sai Yang and Chunxia Zhao "A Fusing Algorithm of Bag-Of-Features Model and Fisher Linear Discriminative Analysis in Image Classification" in IEEE International Conference on Information Science and Technology 2012. Pp 564-571.*

[6] *Ajay Kumar Singh, Shamik Tiwari , V.P. Shukla " Wavelet based Multi Class image classification using Neural Network" in International Journal of Computer Applications Volume 37– No.4, January 2012.*

ISSN: 2277-9655
Impact Factor: 1.852

[7] *Sancho McCann , David G. Lowe "Local Naive Bayes Nearest Neighbor for Image Classification" in IEEE 2012. Pp 1345-1352.*

[8] *Lexiao Tian, Dequan Zheng and Conghui Zhu "Research on Image Classification Based on a Combination of Text and Visual Features" in Eighth International Conference on Fuzzy Systems and Knowledge Discovery 2011.*

[9] *S Agrawal, NK Verma, P Tamrakar "Content Based Color Image Classification using SVM" IEEE 2011. Pp 670-678.*

[10] *M. Simard, S. S. Saatchi, G. De Grandi, "The use of decision tree and multiscale texture for classification of JERS-1 SAR data over tropical forest," IEEE Trans. Geosci. Remote Sens., vol. 38, no. 5, Sep. 2000. Pp 2310–2321.*

[11] *M. Pesaresian, J. A. Benediktsonn, "A new approach for the morphological segmentation of high-resolution satellite imagery," IEEE Trans. Geosci. Remote Sens., vol. 39, no. 2, 2001. Pp 309–320.*

[12] *J. A. Benedktsonn, J. A. Palmason, J. R. Sveinsson, "Classification of hyperspectral data from urban areas based on extended morphological profiles," IEEE Trans. Geosci. Remote Sens., vol. 43, no. 3, 2005. Pp 480–491.*

[13] *M. Dalla Mura, A. Villa, J. A. Benediktsson, J. Chanussot,, L. Bruzzone, "Classification of hyperspectral images by using extended morphological attribute profiles and independent component analysis," IEEE Geosci. Remote Sens. Lett., vol. 8, no. 3, 2011. Pp 542–546.*

[14] *M. Dalla Mura, J. AtliBenediktsson, B. Waske, L. Bruzzone, "Morphological attribute profiles for the analysis of very high resolution images," IEEE Trans. Geosci. Remote Sens., vol. 48, no. 10,2010. Pp 3747–3762.*