ISSN: 2277-9655
Scientific Journal Impact Factor: 3.449
(ISRA), Impact Factor: 2.114

IJESRT

INTERNATIONAL JOURNAL OF ENGINEERING SCIENCES & RESEARCH TECHNOLOGY

A REVIEW OF COLOR MINING AND IMAGE PROCESSING FIELD IN COLOR CLASSIFICATION

Sanjay Singh Bhadoria*, Rahul Deshmukh, Pharindra Kumar Sharma
* M.tech MIST Bhopal,
HOD MIST Bhopal, Ph.D Scholar NIU Noida UP
Ph.D Scholar NIU Noida UP

ABSTRACT

We are analyzed of different-2 colour mining technique for colour classification in this paper. Too many methods are used in colour classification like that histogram, PCA, LDA, ICA, KNN, K-means etc. All of methods either belongs to image processing or mining fields. But our proposed work is depended on two key field area first is colour mining and second is that image processing. By using KNN mining method we classified colors and related colors after that applying pixel processing method for accurate colour classification. Here pixel processing method also plays a filtering role. We are saying on the basis of our proposed work is much better than previous method.

KEYWORDS: Data mining, color mining, color image, data mining technique, K-NN, correlation, PCA.

INTRODUCTION

Recently, the rapid development of digital image and video devices catalyzed the growth of digital multimedia contents. In response to the need of efficient management of multimedia database systems, MPEG-7 [1] has been developed as a standard for multimedia content description to facilitate interoperability between different searching engines and different databases for multimedia data searching and content exchange. For describing multimedia contents, low level features are usually used since they can be generated efficiently. Among the low level image features, color is the most expressive and effective in visual content description.

Several color descriptors [2] have been defined in MPEG-7 for describing color features of multimedia contents. These color descriptors are applicable to a broad range of applications according to their complexity (in terms of extraction and similarity measure) and efficiency (in terms of descriptor size and accuracy). Among the color descriptors, Dominant Color Descriptor (DCD) [2], is a compact color descriptor, it stores the dominant colors, their percentages, and several optional parameters. Since DCD stores only the dominant colors (max. 8 colors) instead of a color histogram, the storage requirement is very effective with relatively small redundancy. Unfortunately, DCD performance is the worst among various color descriptors in its existing usage as reported in [2-3]. Its performance is even worse than a more compact descriptor, the Color Layout Descriptor (CLD). The performance of DCD is quite limited due to the lack of spatial color distribution information and the color accuracy problem. Among various MPEG-7 color visual descriptors, Color Structure Descriptor (CSD) always gives the best retrieval rate. The main reason is the capture of spatial color distribution information in CSD. However, CSD have redundancy as it uses a fixed color space for the histogram representation. The most compact CSD configuration takes about 32 bytes for each image. The size is approximately 4 times that of CLD and about double of DCD. This might be negligible but, for mobile or other bandwidth concerned applications, every bit should be saved. To address this problem we can combine the dominant color features of DCD and the spatial color distribution structure of CSD to design a new color descriptor, called Dominant Color Structure Descriptor (DCSD), which maintains the compactness of dominant colors while significantly improving the retrieval accuracy by using structured dominant color histograms for introducing spatial distribution information of the image.

ABOUT DATA MINING

Mining methodology and user interaction issues: These reflect the kinds of knowledge mined the ability to mine knowledge at multiple granularities, the use of domain knowledge, ad hoc mining, and knowledge visualization.

ISSN: 2277-9655
Scientific Journal Impact Factor: 3.449
(ISRA), Impact Factor: 2.114

- Mining different kinds of knowledge in databases: Since different users can be interested in different kinds of knowledge, data mining should cover a wide spectrum of data analysis and knowledge discovery tasks, including data characterization, discrimination, association, classification, clustering, trend and deviation analysis, and similarity analysis. These tasks may use the same database in different ways and require the development of numerous data mining techniques.
- Interactive mining of knowledge at multiple levels of abstraction: Since it is difficult to know exactly what can be discovered within a database, the data mining process should be interactive. For databases containing a huge amount of data, appropriate sampling techniques can first be applied to facilitate interactive data exploration. Interactive mining allows users to focus the search for patterns, providing and refining data mining requests based on returned results. Specifically, knowledge should be mined by drilling down, rolling up, and pivoting through the data space and knowledge space interactively, similar to what OLAP can do on data cubes. In this way, the user can interact with the data mining system to view data and discovered patterns at multiple granularities and from different angles[25,29].
- Incorporation of background knowledge: Background knowledge, or information regarding the domain under study, may be used to guide the discovery process and allow discovered patterns to be expressed in concise terms and at different levels of abstraction. Domain knowledge related to databases, such as integrity constraints and deduction rules, can help focus and speed up a data mining process, or judge the interestingness of discovered patterns[12,13].
- Data mining query languages and ad hoc data mining: Relational query languages (such as SQL) allow users to pose ad hoc queries for data retrieval. In a similar vein, high-level data mining query languages need to be developed to allow users to describe ad hoc data mining tasks by facilitating the specification of the relevant sets of data for analysis, the domain knowledge, the kinds of knowledge to be mined, and the conditions and constraints to be enforced on the discovered patterns. Such a language should be integrated with a database or a data warehouse query language, and optimized for efficient and flexible data mining [4,7,8].

PREVIOUS WORK

A new Dominant Color Structure Descriptor (DCSD) is proposed in this paper. It is designed to provide an efficient way to represent both color and spatial structure information with single compact descriptor. The descriptor combines the compactness of Dominant Color Descriptor (DCD) and the retrieval accuracy of Color Structure Descriptor (CSD) to enhance the retrieval performance in a highly efficient manner. The feature extraction and similarity measure of the descriptor are designed to address the problems of the existing descriptors while utilize the advantages ofthem. Experimental results show that DCSD has a significant improvement on both retrieval performance and descriptor size over DCD. An eight-color DCSD (DCSD 8) gives an Averaged

Normalized Modified Retrieval Rate (ANMRR) of 0.0993 using MPEG-7 common color dataset, outperforming compact configurations of Scalable Color Descriptor and Color Structure Descriptor with smaller descriptor size.

Ka-Man Wong, Lai-Man Po, and Kwok-Wai Cheung "DOMINANT COLOR STRUCTURE DESCRIPTOR FOR IMAGE RETRIEVAL" 1-4244-1437-7/07/2007 IEEE.

Text Categorization (TC), also known as Text Classification, is the task of automatically classifying a set of text documents into different categories from a predefined set. If a document belongs to exactly one of the categories, it is a single-label classification task; otherwise, it is a multi-label classification task. TC uses several tools from Information Retrieval (IR) and Machine Learning (ML) and has received much attention in the last years from both researchers in the academia and industry developers. In this paper, we first categorize the documents using KNN based machine learning approach and then return the most relevant documents.

ISSN: 2277-9655
Scientific Journal Impact Factor: 3.449
(ISRA), Impact Factor: 2.114

Vishwanath Bijalwan1, Vinay Kumar2, Pinki Kumari3 and Jordan Pascual4, " KNN based Machine Learning Approach for Text and Document Mining", International Journal of Database Theory and Application Vol.7, No.1 (2014), pp.61-70

PROPOSED WORK

Color classification is an important in image processing. Classification has been widely applicable in different areas of science, technology, social science, biology, economics, medicine and stock market. Classification problem appears in other different field like pattern recognition, statistical data analysis, bio-informatics, etc. There exist many classification methods in the literature.

Input:- Take a query color image
Output:-KNN Data Mining based color classification

Step 1: Read an query image.
Step 2: Normalized the unique image size.
Step 3: After that calculate image size, array of pixels on particular color.
Step 4: Apply k-nearest neighbor data mining technique on those array of pixels on particular color.
Step 5: After step 4 we are classified colors in the image.

In above algorithm after executing the 4th steps, we get the classification of colors in images based on K-NN data mining technique. On those resulted output are presented step5.

CONCLUSION AND FUTURE WORK

The conclusion of our proposed work is better because our result it is depended on two filed. These two filed are work parallely and definitely getting high accuracy result.

REFERENCES

1. Arvind Sharma, Nishchol Mishra, Pharindra Kumar Sharma "Image Retrieval Algorithm Based on Quantized DCT Coefficients" Computational Intelligence and Communication Networks (CICN), 2011 International Conference IEEE, Page(s): 431-434, 7-9 Oct. 2011
2. Pharindra Kumar Sharma, Dr. Sanjeev Sharma, Nishchol Mishra "An Efficient Method for Landscape Image Classification and Matching Based on MPEG-7 Descriptors", IJCTA Vol 2 (4), 987-992, JULY-AUGUST 2011.
3. "Image Retrieval Using Local Compact DCT-based Representation" ˇ Stˇep´an Obdrˇz´alek1 and Jiˇr´ı Matas1,2 1 Center for Machine Perception, Czech Technical University, Prague, CZ 2 Centre for Vision Speech and Signal Processing, University of Surrey, Guildford, UKDAGM'03, 25th Pattern Recognition Symposium September 10-12, 2003, Magdeburg, Germany Springer-Verlag.
4. "Enhancement of Color Images by Scaling the DCT Coefficients", Jayanta Mukherjee, Senior Member, IEEE, and Sanjit K. Mitra, Life Fellow, IEEE, IEEE TRANSACTIONS ON IMAGE.
5. K. A. Vidhya and G. Aghila, "A Survey of Naïve Bayes Machine Learning approach in Text Document Classification", International Journal of Computer Science and Information Security, vol. 7, no. 2, (2010), pp. 206-211.
6. Z. Abdullah and M. S. Hitam, "Features Extraction Algorithm from sgml for Classification", Journal of Theoretical and Applied Information Technology, vol. 3, (2007), pp. 72-78.
7. L. Wang, X. Zhao, "Improved knn Classification Algortihm Research in Text Categorization", In the Proceedings of the 2nd International Conference on Communications and Networks (CECNet), (2012), pp. 1848-1852.
8. A. McCallum and K. Nigam, "A Comparison of Event Models For Naïve Bayes Text Classification", In The Proceedings Of The Workshop On Learning For Text Categorization, (1998), pp. 41-48.
9. W. Wang, D. B. Do and X. Lin, "Term Graph Model for Text Classification", In the Proceedings of the International Conference on Advanced Data Mining and Applications, (2005), pp. 19-30.
10. T. M. Mitchell, "Machine Learning", Carnegie Mellon University, McGraw-Hill Book Co, (1997).