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

IJESRT

INTERNATIONAL JOURNAL OF ENGINEERING SCIENCES & RESEARCH TECHNOLOGY

COMPUTATIONAL MODEL FOR COLOR CLASSIFICATION BASED ON K-NEAREST NEIGHBOR COLOR MINING TECHNIQUE

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

ABSTRACT

In this thesis, we are developed a effective and accurate methodology for identification and recognition of human vision system for example camera and robots both are performed the colour classifications action in practical life.

The objective of the proposed methodology is colour classification of the image based on hybrid concepts of K-NN data mining technique and pixel processing. K-NN technique is one of the vital classification approach of data mining because it is analyzed the pixel with neighbour pixel with 8 connected quantifiable manner of the image for colour mining on the behalf of K=8 value. We know that pixel processing is a essential part of image processing.

In colour mining we calculated particular RGB colour with its combinations value like Red Colour and its all possible combinations and after that analysed pixels and calculated some variable like that colour variance, maximum colour, minimum colour, median colour, area of colour and also colour quality by using MATLAB tools.

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

INTRODUCTION

Recently the development of digital color for computer vision technology that doing the classification of the images. MPEG-7 [4, 15] 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 color 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 [15, 16] 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 and efficiency.

The MPEG-7 visual descriptors [1, 5, 7, 15] define and classification of a varied set of images like landscape images, cityscape images, skyscape images etc. which describe the various subsets of visual contents. These descriptors include color, texture, and shape descriptors which describe different features of visual content like the landscape image objects. CSD (Color Structure Descriptor) is one type of color descriptor that based on color histograms, but aims at identifying localized color distributions using a small structuring window. To guarantee interoperability, the color structure descriptor is bound to the Hue-Min-Max-Difference (HMMD) color space and users can efficiently browse.

DATA MINING ISSUES

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.

- 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

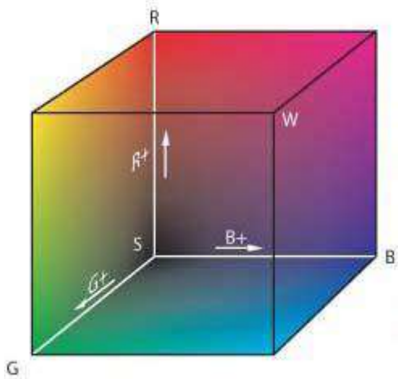
ISSN: 2277-9655
Scientific Journal Impact Factor: 3.449
(ISRA), Impact Factor: 2.114

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].

INTRODUCTION TO COLOR SPACE

RGB: The RGB color space is one of the most widely used color models. The name of the model comes from the initials of the three additive primary colors: Red, Green, and Blue. It is defined as the unit cube in the Cartesian coordinate system as can be seen in the following figure.



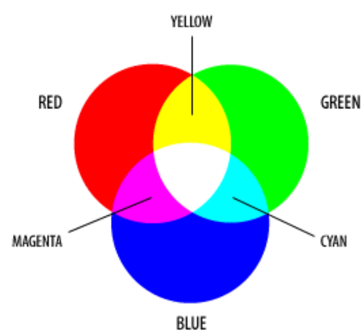


Figure. 1 Unit cube in the Cartesian coordinate system of the RGB Color Space

To indicate in which proportion it is mixed each color, a value is assigned to each of the primary colors where 0 value means that it does not intervene in the mixture. It is quite common to code each primary color with one byte. Figure 2.4.1 shows the geometry of the RGB color model for specifying colors using a Cartesian coordinate system. The grayscale spectrum, i.e. those colors made from equal amounts of each primary, lies on the line joining the black and white vertices This is an additive model, i.e. the colors present in the light add to form new colours, and is appropriate for the mixing of colored light for example. The image on the left of figure 2.2.1.2 shows the additive mixing of red, green and blue primaries to form the three secondary colours yellow (red + green), cyan (blue + green) and magenta (red + blue), and white ((red + green + blue).

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

The RGB model is used for color monitors and most video cameras.

The RGB color model relates very closely to the way we perceive color with the r, g and b receptors in our retinas. RGB uses additive color mixing and is the basic color model used in television or any other medium that projects color with light. It is the basic color model used in computers and for web graphics, but it cannot be used for print production. The secondary colors of RGB – cyan, magenta, and yellow – are formed by mixing two of the primary colors (red, green or blue) and excluding the third color. Red and green combine to make yellow, green and blue to make cyan, and blue and red form magenta. The combination of red, green, and blue in full intensity makes white.

YCbCr: The YCbCr is a legacy color space of the precedent analogue color television and MPEG standards. It is defined by a linear transformation of RGB color space:

Luminance Y = 0.299*R + 0.587*G + 0.114*B
Blue chrominance Cb = 0.169*R – 0.331*G + 0.500*B
Red chrominance Cr = 0.500*R – 0.419*G – 0.081*B

The difference between YCbCr and RGB is that YCbCr represents color as brightness and two color difference signals, while RGB represents color as red, green and blue. For the Monochrome color representation, Y component alone in the YCrCb is used The MPEG-7 standard recommends using this color space for the Color Layout Descriptor.

CMY: The CMY (cyan-magenta-yellow) model is a subtractive model appropriate to absorption of colors, for example due to pigments in paints. Whereas the RGB model asks what is added to black to get a particular color, the CMY model asks what is subtracted from white. In this case, the primaries are cyan, magenta and yellow, with red, green and blue as secondary colors. When a surface coated with cyan pigment is illuminated by white light, no red light is reflected, and similarly for magenta and green, and yellow and blue. The relationship between the RGB and CMY models is given by:

HSV: The HSV color space attempts to describe perceptual color relationships more accurately than RGB, while remaining computationally simple. It is defined as a cylinder

Fig. 2. consists of Hue (H), Saturation (S) and Value (V). Hue is represented by the angle from 0 to 360º and specifies one color family from another. Saturation (= [0,1]) specifies how pure a color is and Value (= [0,1]) specifies how bright or dark a color is.

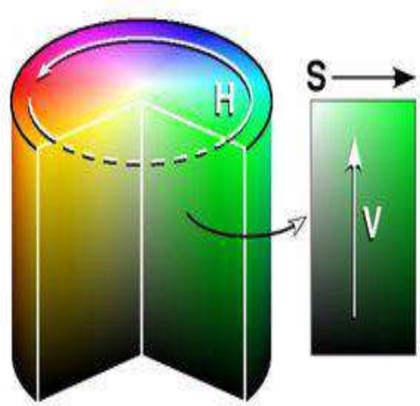
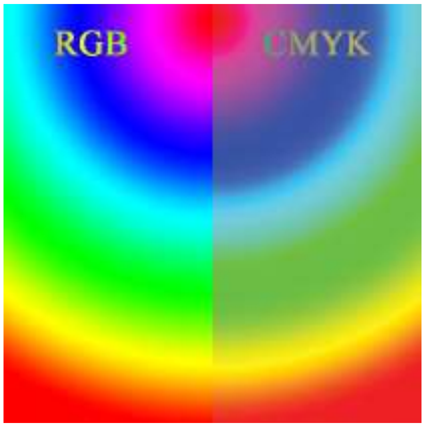


Figure 2 HSV Color Space

CMYK: The 4-colour CMYK model used in printing lays down overlapping layers of varying percentages of transparent cyan (C), magenta (M) and yellow (Y) inks. In addition a layer of black (K) ink can be added. The CMYK model uses the subtractive color model.

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



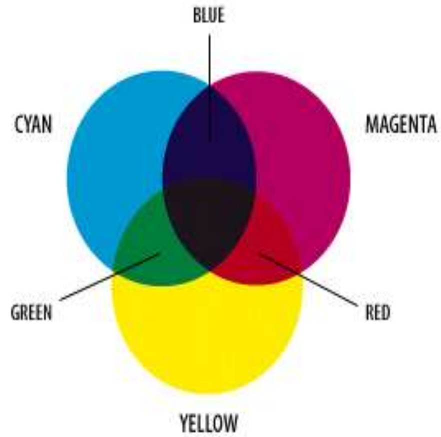


Figure 3 The CMYK Colors Model.

HSI: As mentioned above, color may be specified by the three quantities hue, saturation and intensity. This is the HSI model, and the entire space of colors that may be specified in this way is shown in figure2.3.1.

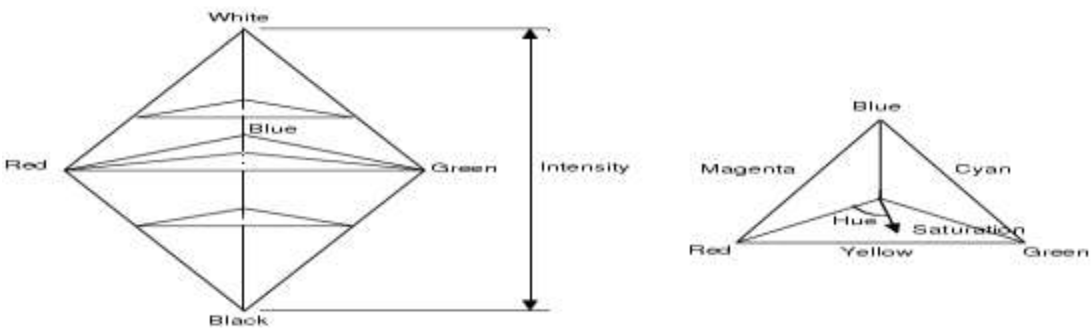


Figure 4 The CMYK Colors Model shows The HSI model, showing the HSI solid on the left, and the HSI triangle on the right, formed by taking a horizontal slice through the HSI solid at a particular intensity.

Hue is measured from red, and saturation is given by distance from the axis. Colours on the surface of the solid are fully saturated, i.e. pure colours, and the greyscale spectrum is on the axis of the solid. For these colors, hue is undefined.

Conversion between the RGB model and the HSI model is quite complicated. The intensity is given by

$$I = \frac{R+G+B}{3},$$

where the quantities R, G and B are the amounts of the red, green and blue components, normalized to the range [0,1]. The intensity is therefore just the average of the red, green and blue components. The saturation is given by:

$$S = 1 - \frac{\min(R,G,B)}{I} = 1 - \frac{3}{R+G+B}\min(R,G,B)$$

where the min(R,G,B) term is really just indicating the amount of white present. If any of R, G or B are zero, there is no white and we have a pure color. The expression for the hue, and details of the derivation may be found in reference.

REVIEW OF LITERATURE

INTRODUCTION

Text classification is a supervised learning task of assigning natural language text documents to one or more predefined categories or classes according to their contents. While it is a classical problem in the

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

field of information retrieval for a half century, it has recently attracted an increasing amount of attention due to the ever-expanding amount of text documents available in digital form. Its applications span a number of areas including auto-processing of emails, filtering junk emails, cataloguing Web pages and news articles, etc.

A large number of techniques have been developed for text classification, including Nearest Neighbor, neural networks, regression, rule induction, and Support Vector Machines (SVM), Joachims. Among them SVM has been recognized as one of the most effective text classification methods.

Yang & Liu gave a comparative study of many algorithms. As supervised learning methods, most existing text classification algorithms require sufficient training data so that they obtained classification model can generalize well. When the number of training data in each class decreases, the classification accuracy of traditional text classification algorithms.

PROPOSED WORK

Color classification is an important issue in image mining. 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.

EXPERIMENTAL RESULT OF THE COMPUTATIONAL MODEL FOR COLOR CLASSIFICATION BASED ON K- NEAREST NEIGHBOR COLOR MINING TECHNIQUE.

In the evaluation of data mining, it also has calculated different-different colors by proposed computational model. In this algorithm we are regarded as the two most important aspects and therefore both of them should be considered at the same time. In order to verify color classification of algorithm proposed in this paper, a great number of experiments on an image are performed. Our query applies on 100 color images which are composed of flower, tree, architecture of earth, architecture of city, lands and etc.

Some images are following that come from query database

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

Below tables are showed the result of color classification base on red number of pixels.

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

Color Images	Number of Red Color Pixels	Overall percentage (in %)
1	211	.82
2	77	.30
3	4598	17.96
4	13699	53.51
5	2691	10.51
6	547	2.14
7	384	1.50
8	8561	33.44
9	4967	19.40
10	10437	40.77
11	21600	84.38

Table1 Number of Red Color pixels and its Percentage

Color Images	Number of Green Color Pixels	Overall percentage(in %)
1	5800	22.66
2	112	.44
3	19093	74.58
4	8731	34.11
5	21769	85.05
6	5119	20.00
7	17263	67.43
8	6829	26.68
9	19916	77.80
10	12719	49.68
11	2076	8.11

Table2 Number of Green Color pixels and its Percentage

Color Images	Number of Blue Color Pixels	Overall percentage(in %)
1	19469	76.05
2	24130	94.26
3	243	.95
4	2619	10.23
5	1014	3.96
6	18277	71.39
7	7209	28.16
8	9826	38.38
9	104	.41
10	2118	8.27
11	1697	6.63

Table3 Number of Green Color pixels and its Percentage

Below tables are showed the result of color classification base on quantity of color value.

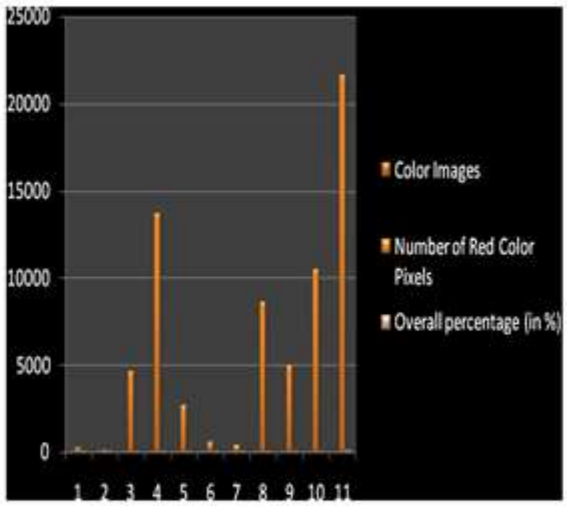
Color Images	Quantity of Red Color (in %)	Quantity of Green Color (in %)	Quantity of Blue Color (in %)
1	6.75	23.81	69.44
2	3.28	3.94	92.78
3	6.38	89.26	4.36
4	54.19	28.41	17.4
5	12.69	29.94	57.37
6	7.13	24.79	68.08

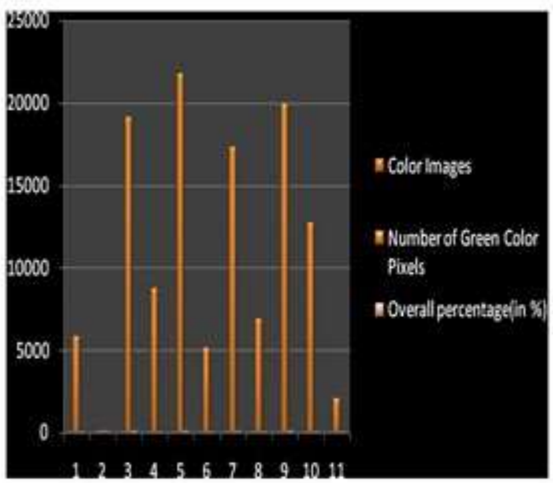
ISSN: 2277-9655
Scientific Journal Impact Factor: 3.449
(ISRA), Impact Factor: 2.114

7	6.91	89.57	3.52
8	23.81	26.91	49.28
9	17.84	81.48	0.68
10	47.16	51.29	1.55
11	93.58	5.17	1.25

Table4 Number of Red, Green and Blue Color quantity

Now we are showed graph of color classification.





CONCLUSION AND FUTURE WORK

The conclusion of the thesis is color classification by using two effective approach: K-NN and pixel processing. We first efficiently extracted three major color components of the image by K-NN method i.e., RED, GREEN, and BLUE and after that apply pixel analyzed methods. Our proposed work is much better than previous work because we apply first color mining and second pixel analyzed method both are top classification methods. In previous work only one of the method are used for classification.

In future we will be adding two approach fuzzy and GA of our proposed work for better color classification.

REFERENCES

1. "Content-Based Image Retrieval Using Color and Edge Direction Features" Jianlin Zhang School of Software Nanchang University Nanchang, Wensheng Zou Institute of Computer Technology Engineering Nanchang University Nanchang, 978-1-4244- 5848-6/10 , 2010 IEEE.
2. "A Review of Content-Based Image Retrieval", G. Rafiee, S.S. Dlay, and W.L. Woo School of Electrical, Electronic and Computer Engineering, Newcastle University, England, United Kingdom, 2010 IEEE.
3. Huiyu Zhou*, Abdul H. Sadka, Mohammad R. Swash, Jawid Azizi and Abubakar S.Umar, School of Engineering and Design, Brunel University, "Content Based Image Retrieval and Clustering: A Brief Survey" Recent Patents on Electrical Engineering 2009, Bentham Science Publishers Ltd 2, 187-199 187.
4. THE 2D-DCT COEFFICIENT STATISTICAL BEHAVIOUR: A COMPARATIVE ANALYSIS ON DIFFERENT TYPES OF IMAGE SEQUENCES M. Mitreaa,b*, F.Prêteuxb, A. Vlada,c, C. Fetitab aFaculty of Electronics and Telecommunications, "Politehnica" University, Bucharest–Romania bARTEMIS Project Unit, GET / INT, Evry–France cThe Research Institute for Artificial Intelligence, Romanian, Academy Journal of Optoelectronics and Advanced Materials Vol. 6, No. 1, March 2004, p. 95 – 102.
5. "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.
6. "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 PROCESSING, VOL. 17, NO. 10, OCTOBER 2008.
7. "Review of shape representation and description techniques", Dengsheng Zhang∗, Guojun Lu Gippsland School of Computing and Info. Tech., Monash University, Churchill, Vic 3842, Australia, 16 July 2003.
8. "Shape Classification Using the Inner-Distance" Haibin Ling, Student Member, IEEE, and David W. Jacobs, Member, IEEE, IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, VOL. 29, NO. 2, FEBRUARY 2007.
9. Zyad Shaaban and Thawar Arif, Faculty of Information Technology, Applied Science University, Amman , "Shape Representation and Retrieval Using Radial Basis Probabilistic Neural Networks", Jordan IJCSNS International Journal of Computer Science and Network Security, VOL.8 No.12, December 2008..
10. "Shape-based Image Retrieval" RYSZARD S. CHORA´S University of Technology & Life Sciences Faculty of Telecommunications & Electrical Engineering S. Kaliskiego 7, 85-796 Bydgoszcz POLAND Proceedings of the 7th WSEAS International Conference on SIGNAL PROCESSING (SIP'08), Istanbul, Turkey, May 27-30, 2008.
11. "A Descriptive Algorithm for Sobel Image Edge Detection", O. R. Vincent, Clausthal, O. Folorunso, University of Technology, Germany and University of Agriculture, Abeokuta, Nigeria, proceedings of Informing Science & IT Education Conference (InSITE) 2009.
12. SEGMENTATION AND OBJECT RECOGNITION USING EDGE DETECTION TECHNIQUES, Y.Ramadevi, T.Sridevi, B.Poornima, B.Kalyani Department of CSE , Chaitanya Bharathi Institute of Technology Gandipet, Hyderabad, International Journal of Computer Science & Information Technology (IJCSIT), Vol 2, No 6, December 2010.

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

13. ”Morphological Background Detection and Enhancement of Images With Poor Lighting”, Angélica R. Jiménez-Sánchez, Jorge D. Mendiola-Santibañez, Iván R. Terol-Villalobos, Gilberto Herrera-Ruíz, Damián Vargas-Vázquez, Juan J. García-Escalante, and Alberto Lara-Guevara IEEE TRANSACTIONS ON IMAGE PROCESSING, VOL. 18, NO. 3, MARCH 2009 .
14. “Application for Morphological Image Processing”, Author: Dan Campbell UW-Madison Computer Engineering/Computer Science,the university of Wisconsin Madison, Course: ECE533,2006.
15. 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.
16. Pharindra Kumar Sharma, Shalabh Agarwal, Piyush Shrivastava. “Image Enhancement Based on Color Histogram and DCT Approach”, IJCTA Vol 2 (4), 999-1002, JULY-AUGUST 2011.