

**Detection of Carbon Composition in hypo-eutectoid steel based on color
texture of Iron-Carbon image**Mr. Sharadchandra M. Kawale¹, Dr. Anilkumar N. Hohambe²^{1,2}Department of Computer Science & Engineering, College of Engineering, Osmanabad

Abstract — The fundamental data clustering problem may be defined as the process of grouping the data objects into classes or clusters, so that objects within a cluster have high similarity in comparison to one another but are very dissimilar to objects in other clusters. This paper proposes solution to find percentage of carbon contents in hypo-eutectoid steel through extraction of color features of various images. A color quantization is focuses on color as feature and considers HVS space. Image pixel color is quantized into number of colors and histogram of these colors has calculated. To form clusters of images k-means algorithm is applied and proposed solution for finding carbon percentages in hypo-eutectoid steel.

Keywords- Retrieval, Histogram, Clustering, HSV, K-Means, hypo-eutectoid steel.

I. INTRODUCTION

Color is the most important visual era for image and scene analysis. Due to color humans can easily understand real world scene images containing multiple objects. Images of various applications are converted to digital form and stored in image databases for later use. A wide range of possible applications that require video and image database are: web searches, crime prevention, the military, music video clips, news broadcasting, home entertainment, education and training, cultural heritage, geographical information systems (GIS) and remote sensing, medical diagnosis, journalism and advertising, fashion and interior design [1, 2]. Rich information is hidden in this data collection which is potentially useful. A major challenge with these fields is to extract meaning from the data they contain i.e. to discover structure and find patterns. Exploring and analyzing the vast volume of image data is becoming increasingly difficult.

Data clustering is the process of dividing data elements into classes or clusters so that items in the same class are as similar as possible, and items in different classes are as dissimilar as possible. Data mining is the process of sorting through large data sets to identify patterns and establish relationships to solve problems through data analysis. Data mining tools allow enterprises to predict future trends [3].

Clustering analyzes data objects without consulting a known class label. In general, the class labels are not present in the training data simply because they are not known to begin with. Clustering can be used to generate such labels. The objects are clustered or grouped based on the principle of maximizing the intra class similarity and minimizing the interclass similarity [1, 3].

In this paper we propose a data mining approach to cluster the images based on color feature for finding percentage of carbon composition in hypo-eutectoid steel. The concept of color histogram is used to obtain the features. RGB color space is converted to Hue, Saturation and Value (HSV) color space. Based on Hue, Saturation and Value, image is quantized to number of colors and histogram of these colors is formed. A K-means clustering algorithm is applied to cluster the images. The rest of the paper is organized as follows: In section two we provide an overview of image retrieval, histogram, HSV color and clustering on hypo-eutectoid steel images. Also we present the rules for quantization of the HSV color model and the calculation of histogram values.

These guidelines include complete descriptions of the fonts, spacing, and related information for producing your proceedings manuscripts.

II. LITERATURE SURVEY**2.1 Image retrieval:**

An image retrieval system is a system for browsing, searching and retrieving images from a large database of digital images. Following are some methods of image retrievals. An image is a representation of an object and represented by a matrix of intensity value. It is sampled at points known as pixels and represented by color intensity in Red, Green and Blue (RGB) color model. 8 bits are used to represent each pixel of RGB components separately. Thus total 24 bits are required to represent one pixel of the image. In image retrieval feature extraction is the process of interacting with images and performs extraction of meaningful information of images. The measurements or properties used to classify the objects are called Features, and the types or categories into which they are classified are called classes. Low-level visual features such as color, texture and shape often employed to search relevant images based on the query image. A n-dimensional feature vector represent an image where n is the selected number of extracted features.[1]

Color information is the most widely used feature for image retrieval because of its strong correlation with the underlying image objects. A commonly used one is the RGB space because most digital images are acquired and

represented in this space. However, due to the fact that RGB space is not perceptually uniform, color space such as HSV (Hue, Saturation, and Value), HSL (Hue, Saturation and Luminance), CIE $L^*u^*v^*$ and CIE $L^*a^*b^*$ tend to be more appropriate for calculating color similarities. Color Histogram [1] [4] is the commonly and very popular color feature used in many image retrieval system. The mathematical foundation and color distribution of images can be characterized by color moments. Color Coherence Vector (CCV) has been proposed to incorporate spatial information into color histogram representation [5].

2.2 Histogram:

The brightness histogram $h_f(z)$ of an image provides the frequency of the brightness value z in the image- the histogram of an image with L gray-levels are represented by a one dimensional array with L elements. The histogram usually provides the global information about the image. It is invariant to translation and rotation around the viewing axis and varies slowly with changes of view angle, and scale [1].

To define discrete color histograms, quantization of a given color space into a finite number of color cells required. Each of them corresponds to a histogram bin. The color histogram of an image is then constructed by counting the number of pixels that fall in each of these cells. There are many different approaches to color quantization, including vector quantization, clustering, and neural networks [6].

2.3 HSV Color Model:

Instead of a set of color primaries, the HSV model uses color descriptions that have a more intuitive appeal to a user. To give a color specification, a user selects a spectral color and the amount of white and black that is to be added to obtain different shades, tints and tones. Color parameters in this model are Hue (H), Saturation (S) and Value (V). The 3-D representation of the HSV model is derived from the RGB cube. If we imagine viewing the cube along the diagonal from the white vertex to the origin, we see an outline of the cube that has a hex cone shape. The boundary of the hex cone represents the various hues and it is used at the top of the HSV hex cone. In the hex cone, saturation is measured along a horizontal axis and value is along the vertical axis through the center of the hex cone [1,4].

Hue is represented as an angle about the vertical axis, ranging from 0° at red through 360° . Vertices of the hexcone are separated by 60° intervals. Yellow is at 60° , Green at 120° and Cyan opposite red at $H=180^\circ$. Complementary colors are 180° apart. Blue at 240° and Magenta at 300° .

Saturation (S) varies from 0 to 1. It is represented in this model as the ratio of the purity of a selected hue to its maximum purity as $S=1$. Value V varies from 0 at the apex of the hexcone to 1 at the top. The apex represents black. Proposed conversion of RGB to HSV color model.

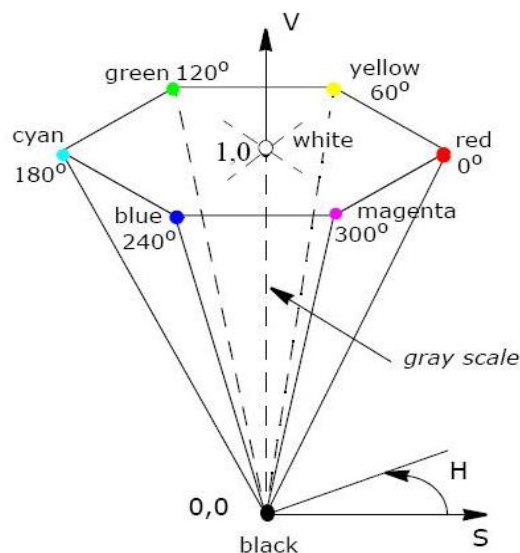


Fig.1 - HSV Color Model

2.4 Clustering:

There are techniques such as clustering for unsupervised learning or class discovery that attempt to divide data sets into naturally occurring groups without a predetermined class structure. The cluster analysis is a partitioning of data into meaningful subgroups (clusters), which the number of subgroups and other information about their composition or representatives are unknown. Cluster analysis does not use category labels that tag objects with prior identifiers i.e. we don't have prior information about cluster seeds or representatives. The objective of cluster analysis is simply to find a convenient and valid organization (i.e. group) of the data [5]. Intelligently classifying image by content is an important way to mine valuable information from large image collection. Reference [8] explores the challenges in image grouping into semantically meaningful categories based on low-level visual features. The SemQuery approach proposes a general framework to support content-based image retrieval based on the combination of clustering and querying of the heterogeneous features [9].

Reference [10] describes data mining and statistical analysis of the collections of remotely sensed image. Large images are partitioned into a number of smaller more manageable image tiles. Then those individual image tiles are processed to extract the feature vectors. The concept of fuzzy ID3 decision tree for image retrieval was discussed in [11]. ID3 is a decision tree method based on Shannon's information theory. Given a sample data set described by a set of attributes and an outcome, ID3 produces a decision tree, which can classify the outcome value based on the values of the given attributes like Color, Texture and Spatial Location. Image dataset were defined in 10 classes (concepts): grass, forest, sky, sea, sand, firework, sunset, flower, tiger and fur. At each level of the ID3 decision tree, the attribute with smallest entropy is selected from those attributes not yet used as the most significant for decision making.

Steel is an alloy of iron (Fe) and carbon (C), carbon by weight. Carbon is the most cost-effective alloying material for iron. It was found that the carbon content in pearlite depends on carbon content of the steel and temperature of transformation, which are both determining the percentage of free ferrite. The influence of free ferrite has been eliminated by using cooling rates fast enough for obtaining a fully pearlite structure in steels between 0.20 and 0.82%C [13]. Hence it has been observed that pearlite and ferrite as components in hypo-eutectoid steel and percentage of pearlite and ferrite refers the percentage of carbon in hypo-eutectoid steel.

III. PROPOSED WORK

Carbon identification through pearlite volume fraction has been difficult in mechanical metallurgy specimen testing. The same problem, for unknown carbon percentage of hypo-eutectoid steel has been solved through image processing techniques.

Sample images of hypo-eutectoid steel have been considered for identification the carbon percentages through pearlite identification. The image is in form of color and it is having size of 256x256 bit in length. The image is converted from RGB to HSV form. At time of HSV formation we require the gray scale image to find carbon percentages through pearlite identification. To do this we have only concentrate on value in HSV conversion. The resulting matrix of value is used to find percentage of whiteness and darkness present in image. The set carbon percentages through pearlite in image are taken for the comparison of percentage of carbon and pearlite values.

For this we have been use the sample images of hypo-eutectoid steel from data base is taken. This predefined percentage of carbon in hypo-eutectoid steel. We propose the carbon percentage in four range namely 0, 0.2, 0.4 and 0.8.

Then histogram is applied color patterns. Every pixel is having different intensity value. K-mean algorithm has been applied to performing sampling for best described features. After finding intensity concentration of different pixel in image we can find different classes from the image. Every class has different concentration. Percentage of class is obtained for finding hypo-eutectoid steel.

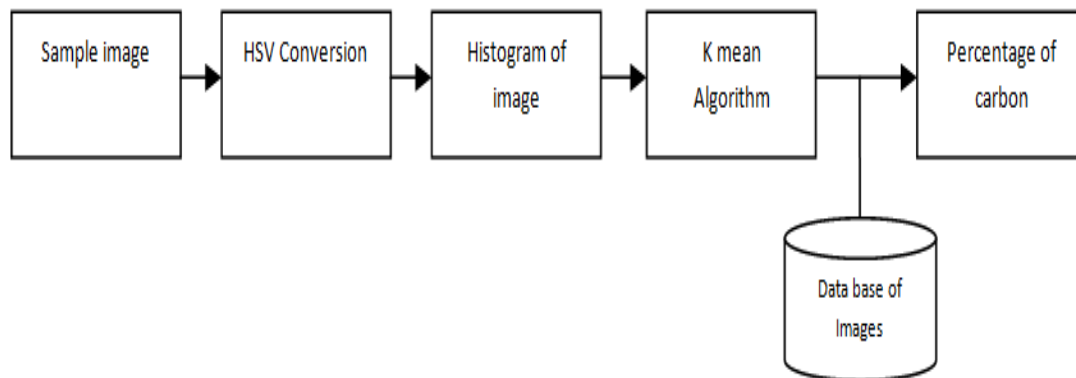


Fig. 2 - Block diagram of process to find carbon percentages

HSV means Hue-Saturation-Value, where the Hue is the color. And since color is not an easy thing to separate or compare, Hue is often represented as a circular angle (between 0.0 to 1.0 when stored as floats). Being a circular value means that 1.0 is the same as 0.0. For example, a Hue of 0.0 is red, a Hue of 0.25 would be green, a Hue of 0.5 is blue, a Hue of 0.75 is pink, and a Hue of 1.0 would be the same as a Hue of 0.0 which is red (again). Saturation is the greyness, so that a Saturation value near 0 means it is dull or grey looking whereas as a Saturation value of 0.8 might be a very strong color (eg: red if Hue is 0). And Value is the brightness of the pixel, so 0.1 is black and 0.9 is white. Unfortunately, there are different ways to represent HSV colors, such as whether a full brightness V of 1.0 should be bright white or a bright color.

Histograms are the basis for numerous spatial domain processing techniques. Histogram manipulations can be used for image enhancement. The inherent information contained in the histogram of an image can be used for image processing applications like image compression and segmentation. The histogram of a digital image with intensity values in the range $[0, L-1]$ is a discrete function $h(r_k) = n_k$, where r_k is the k th intensity value and n_k is the number of pixels in the image with intensity r_k . The histogram of a digital image gives conclusive evidence about the quality of the image.

Histogram Equalization is a digital image processing technique used to increase the contrast of an image. The transformation function uniformly spreads out the most frequent intensity values to improve the global contrast. The method is useful when both the background and foreground are either bright or dark.

K-means is one of the simplest unsupervised learning algorithms in which each point is assigned to only one particular cluster. The procedure follows a simple, easy and iterative way to classify a given data set through a certain number of clusters (assume k clusters) fixed a priori [12]. The procedure consists of the following steps:

Algorithm:

Step 1: Set the number of cluster k

Step 2: Determine the centroid coordinates

Step 3: Determine the distance of each object to the centroids

Step 4: Group the object based on minimum distance

Step 5: Continue from step 2, until convergence that is no object move from one group to another.

Proposed method will gives percentage of carbon in hypo-eutectoid steel through given image.

IV. CONCLUSION

We proposed a new framework based on the color feature of image to group images in an unsupervised manner. The concept of color histogram is used to obtain the features. Hypo-eutectoid steel images are converted to HSV color space. Based on hue, saturation and value, the image is and histogram of these 54 colors is formed. A K-means, clustering algorithm is applied to cluster the images to find percentage of carbon in hypo-eutectoid steel by considering ferrite and pearlite percentage in specimen.

REFERENCES

- [1] G. Eason, B. Noble, and I. N. Sneddon, "On certain integrals of Lipschitz-Hankel type involving products of Bessel functions," *Phil. Trans. Roy. Soc. London*, vol. A247, pp. 529–551, April 1955. (references)
- [2] Manish Maheshwari*, Dr. Mahesh Motwani, Dr. Sanjay Silakari, "New Feature Extraction Technique for Color Image Clustering" *International Journal of Computer Science and Electronics Engineering (IJCSEE)* Volume 1, Issue 1 (2013) ISSN 2320–4028.
- [3] H.J.Zhang et al., "Video Parsing, Retrieval and Browsing: an Integrated and Content-Based Solution", *Proc. ACM Multimedia 95*, San Francisco, Nov 95.
- [4] Nishi Sharma, Mr. Anurag Joshi, "Color Based Image Segmentation Using Data Mining Functionalities" *International Journal of Advanced Research in Computer Science & Technology (IJARCST 2014)* ISSN : 2347 – 8446 Vol. 2, Issue 2, Ver. 2.
- [5] Wayne Niblack, Ron Barber, William Equitz, Myron Flickner, Eduardo H. Glasman, Dragutin Petkovic, Peter Yanker, Christos Faloutsos, Gabriel Taubin: "The QBIC Project: Querying Images by Content, Using Color, Texture, and Shape", *Storage and Retrieval for Image and Video Databases (SPIE)* 1993: 173-187.
- [6] Greg Pass, Ramin Zabih, Justin Miller, "Comparing Images Using Color Coherence Vectors", *ACM Multimedia 1996*: 65-73.
- [7] H.J. Zhang and D. Zhong, "A Scheme for visual feature-based image indexing", *Proceedings of SPIE conference on storage and retrieval for image and video databases III*, pp36-46, 1995.
- [8] Y. Uehara, S. Endo, S. Shiitani, D. Masumoto, and S. Nagata, "A computer-aided Visual Exploration System for Knowledge Discovery from Images", In *Proceedings of the Second International Workshop on Multimedia Data Mining (MDM/KDD'2001)*, San Francisco, CA, USA, August, 2001.
- [9] Gholamhosein Sheikholeslami, Wendy Chang, Aidong Zhang, "SemQuery: Semantic Clustering and Querying on Heterogeneous Features for Visual Data", *IEEE Trans. Knowl. Data Eng.* 14(5): 988-1002 (2002)
- [10] Krzysztof Koperski, Giovanni Marchisio, Selim Aksoy, and Carsten Tusk, "Applications of Terrain and Sensor Data Fusion in Image Mining", *IEEE* 2002, pp 1026-1028
- [11] Ying Liu¹, Dengsheng Zhang¹, Guojun Lu¹, Wei-Ying Ma², "Deriving High-Level Concepts Using Fuzzy-Id3 Decision Tree for Image Retrieval", *IEEE* 2005, pp 501-504.
- [12] Shamik Sural, Gang Qian and Sakti Pramanik, "SEGMENTATION AND HISTOGRAM GENERATION USING THE HSV COLOR SPACE FOR IMAGE RETRIEVAL" 0-7803-7622-6/0202002 IEEE.
- [13] P. HOUIN, A. SIMON and G. BECK "Relationship between Structure and Mechanical Properties of Pearlite between 0.2% and 0.8%C" *UDC 669.112.22 : 539.4 : 539.52* September 22, 1980.