Available online at www.worldnewsnaturalsciences.com

World News of Natural Sciences

An International Scientific Journal

WNOFNS 30(2) (2020) 96-103 EISSN 2543-5426

Optimizing Space Complexity using Color Spaces in CBIR Systems for Medical Diagnosis

S. Pradeep Kumar Kenny

Department of Computer Science, St. Joseph University, Dimapur, Nagaland, India

E-mail address: kenny_isles@yahoo.co.in

ABSTRACT

Content based Image Retrieval systems are now used in various automated systems because they largely produce accurate results as they do not depend on the metadata for telling what the image is but rather define it on the basis contents of the image like color, shape, texture and spatial locations. Content based Image retrieval systems have a repository of similar images and when a query image is presented to system it matches similar images in the database. This process aids in various applications like security checks to medical diagnosis. But all CBIR systems in common have to store the images which take a huge space. Here in this work, a unique approach is being devised to reduce the space complexity for a CBIR system used for detecting cervical cancer. When it comes to medical image it is not the question of how to reduce space, but along with it, the original contents of the image also has to be preserved.

***Keywords*:** Content Based Image retrieval, Cervical Cancer, Color Space, Data Mining

1. INTRODUCTION

Content Based Image Retrieval is an application of computer vision to solve the image retrieval problem. This problem is based on how to search and retrieve images from a database with the least amount of time and with maximum efficiency. Content based means searching and retrieving an image using it contents like shape, color, texture and spatial location rather than the metadata as most traditional system do. The Content based retrieval system thrives because the metadata method can often be misleading as it depends on the metadata creator to define the content. Here on the other hand the content is defined using accurate mathematical

(Received 10 March 2020; Accepted 01 April 2020; Date of Publication 02 April 2020)

methods thus increasing accuracy and the ability to match more similar images in the database when presented with a query image.

In the field of data mining it is a well established fact that the more data you have, the more precise your predictions can be. For example, you have a medical database and you need to diagnose a patient having a form of cancer. When the CBIR system is presented with the symptoms it retrieves similar cases from the database. If the database size is large then the accuracy rate goes up even higher, thus giving the doctor a look into many such cases and hence helping him prescribe the best medication possible. But with the larger database comes the problem of storage.

Here, in this work, a unique approach is being formulated to save storage space by using a different color space. In the previous work we had formulated and tested a content based image retrieval system using combination feature set. The whole CBIR system was developed with accuracy and efficiency in mind. Like for segmentation a watershed segmentation using morphological operators was used for preserving the edges, thereby increasing the accuracy. A number of such techniques were implemented. A need was felt to find a method to reduce the space complexity of the 917 images in the sample database as it would increase efficiency of the overall system being developed.

Section II spells out the unique approach used to tackle this problem. Section III gives the results and Section IV concludes the work.

2. COLOR SPACE

In the previous work, a lot of attention was paid to textural features to improve the accuracy of the CBIR model. Here, the concentration is on reducing storage space and improve efficiency.

Color is formed when white light from the sun or any other source bounces of the surface and wherein certain wavelength gets subtracted hence forming the different hues which we see in the rainbow. Colour as we see is present in wavelengths measuring 390 to 700 nm. To represent this color in the computer for the purpose of display or computation a color space concept is used.

A color space is an abstract mathematical model which describes colors as a set of tuples of numbers. There are various color models available like RGB, CMYK, YUV, HSV, etc. All these have various variants and are used for varied applications. RGB is used in monitors of both, television and computers. CMYK is used for printing, YUV is used as television standards, and HSV is used for high quality image editing works.

In this work, the images are stored in HSI color space which is variant of HSV color space. In doing this, the amount of storage reduces dramatically.

2. 1. HSI Image

HSI Image is a combination of Hue, Saturation and Intensity of an image. It is a variant of HSV image where the V stands for velocity. A diagrammatic model of HSI image is shown below in **Figure 1**.

The HSI color model can be best represented an inverted cone. The central core line is the intensity. Intensity ranges from 0 to 1. The values in between these are the grayshades showing the intensity of the Hue which can be any color of the color cube which is a combination of primary and secondary colors. Saturation is the diluation of how bright and light the Hue should be. In short hue tells the color, intensity tells the brightness and saturation can be said as the sharpness of the color.

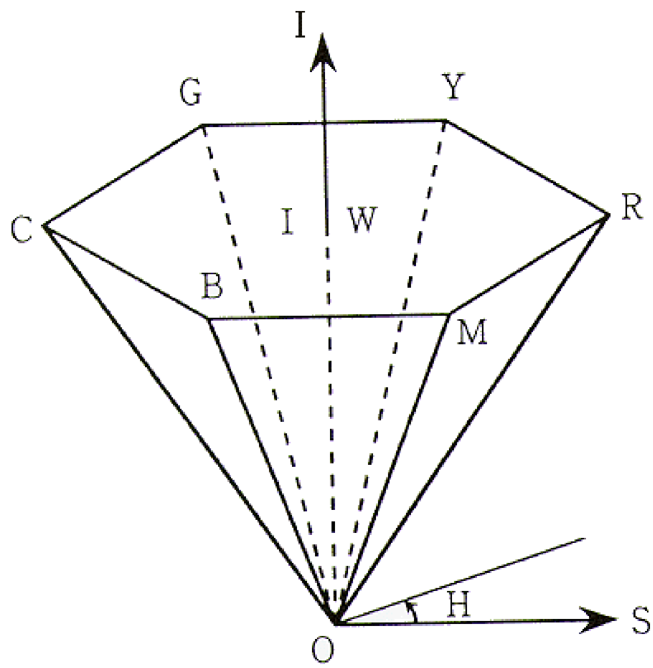


Figure 1. HSI color model

The reason for choosing HSI color model is that it is good for high quality computer graphics and for medical images a high quality image is a must.

2. 2. Conversion to HIS color space

With a lot of potential color models available the most widespread and most commonly used color model is the RGB color model. **Figure 2** shows the RGB color model. RGB stands for Red Green and Blue which the primary colors of the color space.

The input images are normally in this color model and these need to be converted to HSI color space. The pseudo code of this is given below.

1) Separate the three components namely H, S and I.
2) Obtain the binary image from the saturation component.
3) Multiply the binary image with hue component and obtain the product image.
4) Take the complement of intensity or value component.
5) Obtain the binary image from the complement image.
6) Merge the output of step 5 and step 3 to obtain the grey scale image.

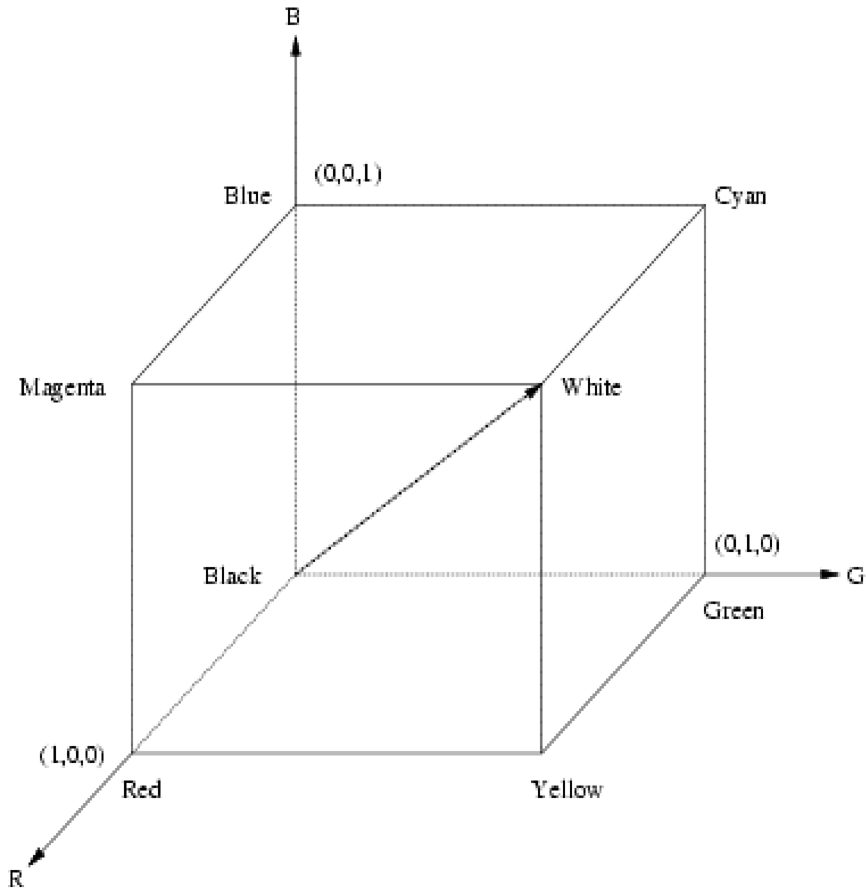
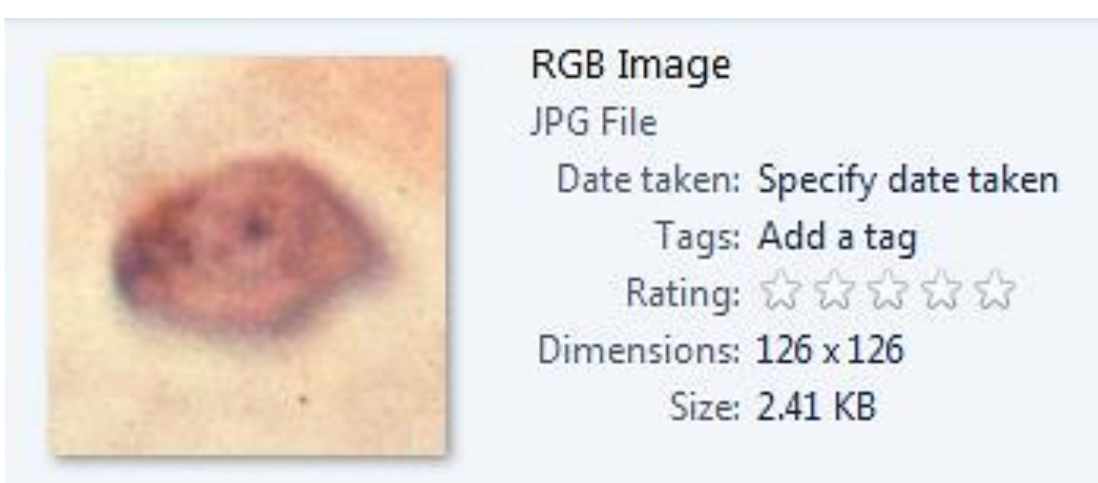
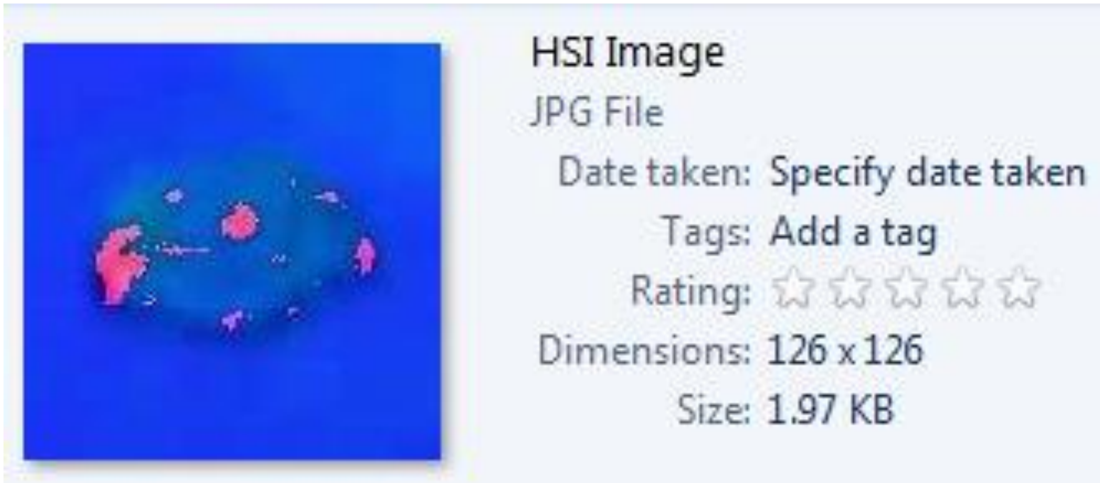


Figure 2. RGB Color Model

After applying the above pseudocode on the RGB image the result would be as shown in **Figure 3**.



(a)



(b)

Figure 3. (a) RGB Image, **(b)** HSI Image

Figure 3(a) shows a cervical cancer cell represented with RGB color space with 2.41 KB of storage space and Figure 3(b) shows a cervical cancer cell represented in HSI color space with a storage space of 1.97 KB which amounts to about 18.67% reduction of storage space. Although the reduction seems meagre it has a significant impact in databases containing thousands of images.

2. 3. Dataset

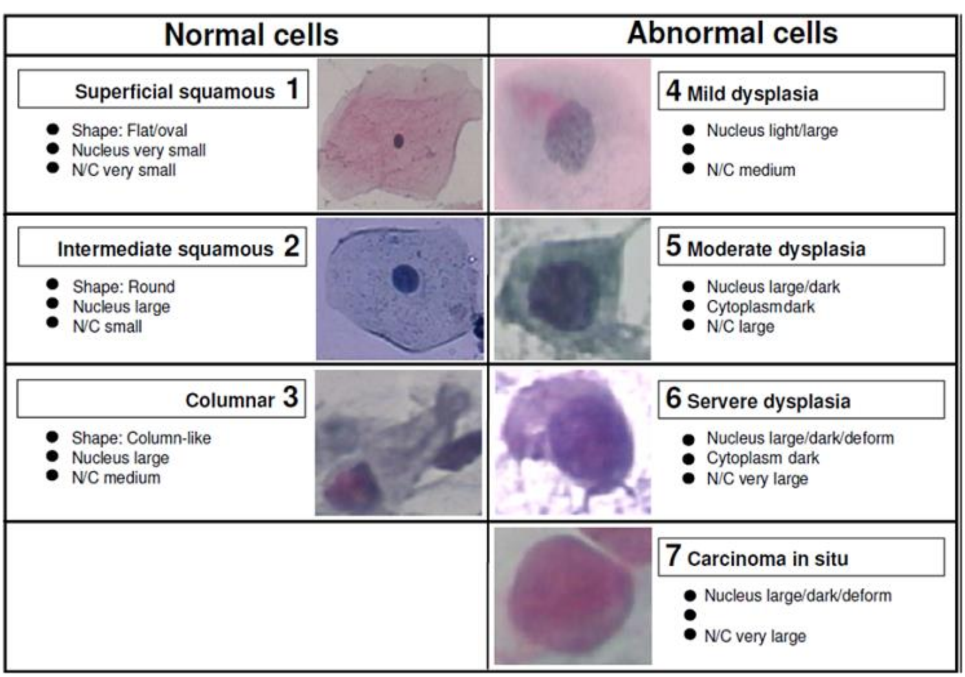


Figure 4. Distribution of the Herlev Dataset

To prove that this technique is really effective, it has also been tested on a large dataset. The images used here are from the Herlev Dataset. This dataset contains a total of 917 images A distribution of the Images is shown in **Figure 4**.

3. EXPERIMENTAL RESULTS

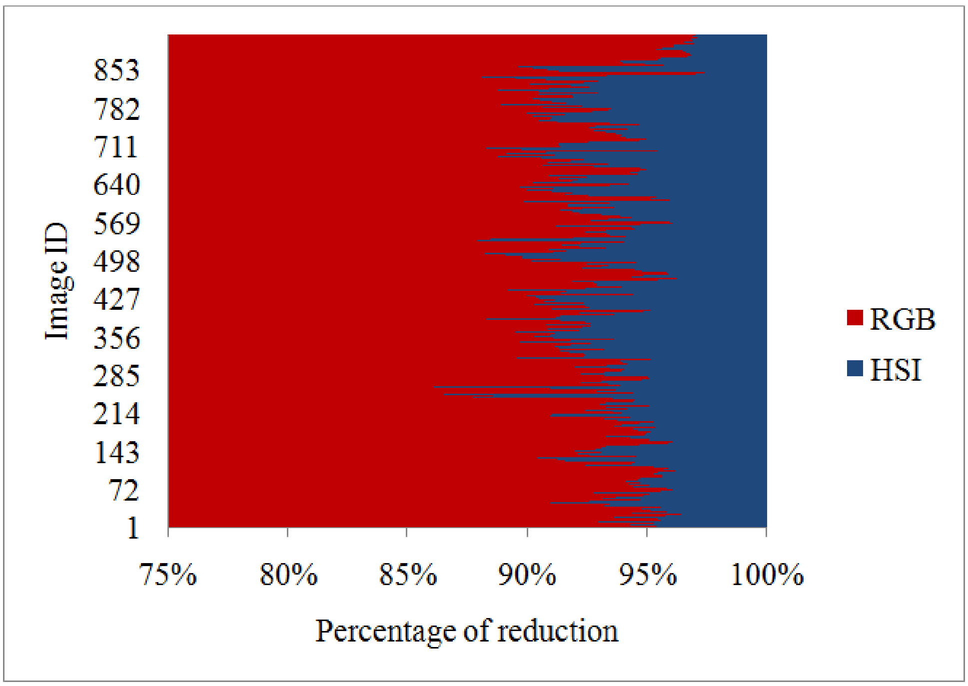


Figure 5. Percentage of Reduction of RGB versus HSI Color Spaces

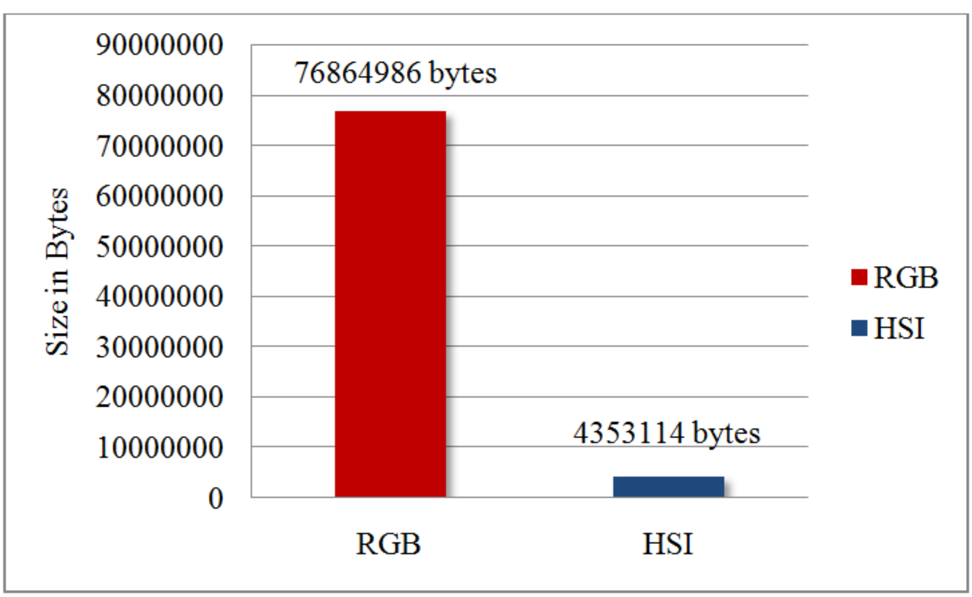


Figure 6. Total File Sizes of the Images Present in the Herlev Data Set Represented by the Two Color Spaces

The images in the Herlev dataset are represented in RGB color space. Hence using the above algorithm it is converted into HSI color space. The percentage of reduction for each of the image in the Herlev Dataset is shown in a stacked graph in **Figure 5** and the percentage of reduction for the images is shown in graphical format in **Figure 6**. As shown in Figure 5 the RGB color space takes about 76864986 bytes (76.86 MB) to store the images whereas the HSI color space takes only 4353114 bytes (4.35 MB) to store the images. This reduces the storage space by 94.34%.

4. CONCLUSION

The technique of swapping color spaces proves to improve storage space complexity by a whopping 93.34% and thus proving the efficiency of the Content Based Image Retrieval System. The other important advantage is that it doesn't consume much computation time to convert from one color model to another, thus providing us with an efficient and robust system.

References

[1] B.V. Dhandra, R. Hegadi, M. Hangarge, and V.S. Malemath. Analysis of Abnormality in Endoscopic images using Combined HSI Color Space and Watershed Segmentation. 18th International Conference on Pattern Recognition (ICPR'06). IEEE *Xplore*: 18 September 2006. DOI:10.1109/ICPR.2006.268

[2] Su Zhang, Wei Yang, Yi-Lun Wu, Rui Yao, and Shi-Dan Cheng, Abnormal region detection in gastroscopic images by combining classifiers on neighboring patches. *Machine Learning and Cybernetics 2009 International Conference on*, vol. 4, pp. 2374-2379, 2009.

[3] Shuqiong Wu, Megumi Nakao, and Tetsuya Matsuda, Automatic GrabCut based lung extraction from endoscopic images with an initial boundary. *Signal Processing (ICSP) 2016 IEEE 13th International Conference on*, pp. 1374-1378, 2016.

[4] Vasileios Charisis, Leontios J. Hadjileontiadis, Christos N. Liatsos, Christos C. Mavrogiannis, and George D. Sergiadis, Abnormal pattern detection in Wireless Capsule Endoscopy images using nonlinear analysis in RGB color space. *Engineering in Medicine and Biology Society (EMBC) 2010 Annual International Conference of the IEEE*, pp. 3674-3677, 2010.

[5] Sareena, Ajay Mittal, and Manvjeet Kaur, Computer-aided-diagnosis in colorectal cancer: A survey of state of the art techniques. *Inventive Computation Technologies (ICICT) International Conference on*, vol. 1, pp. 1-6, 2016.

[6] Corina Barbalata and Leonardo S. Mattos, Laryngeal Tumor Detection and Classification in Endoscopic Video. *Biomedical and Health Informatics IEEE Journal of*, vol. 20, no. 1, pp. 322-332, 2016

[7] Seung-Hwan Bae and Kuk-Jin Yoon, Polyp Detection via Imbalanced Learning and Discriminative Feature Learning. *Medical Imaging IEEE Transactions on*, vol. 34, no. 11, pp. 2379-2393, 2015.

[8] Masoud S. Nosrati, Rafeef Abugharbieh, Jean-Marc Peyrat, Julien Abinahed, Osama Al-Alao, Abdulla Al-Ansari, and Ghassan Hamarneh, Simultaneous Multi-Structure Segmentation and 3D Nonrigid Pose Estimation in Image-Guided Robotic Surgery. *Medical Imaging IEEE Transactions on*, vol. 35, no. 1, pp. 1-12, 2016.

[9] Ming Liu, Jue Jiang, and Zenan Wang, Colonic Polyp Detection in Endoscopic Videos With Single Shot Detection Based Deep Convolutional Neural Network. *Access IEEE*, vol. 7, pp. 75058-75066, 2019.

[10] Yaqub Jonmohamadi, Yu Takeda, Fengbei Liu, Fumio Sasazawa, Gabriel Maicas, Ross Crawford, Jonathan Roberts, Ajay K. Pandey, and Gustavo Carneiro, Automatic Segmentation of Multiple Structures in Knee Arthroscopy Using Deep Learning. *Access IEEE*, vol. 8, pp. 51853-51861, 2020.

[11] J. Bernal, J. Sánchez, and F. Vilariño. Towards automatic polyp detection with a polyp appearance model. *Pattern Recognition*, vol. 45(9), pp. 3166, 2012.

[12] Chia Hsiang Wu and Mei Yun Su, Specular Highlight Detection from Endoscopic Images for Shape Reconstruction. *Applied Mechanics and Materials*, vol. 870, pp. 357, 2017.

[13] Masoud S. Nosrati, Jean-Marc Peyrat, Julien Abinahed, Osama Al-Alao, Abdulla Al-Ansari, Rafeef Abugharbieh, and Ghassan Hamarneh, *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2014*, vol. 8674, pp. 324, 2014.

[14] Shipra Suman, Nicolas Walter, Fawnizu Azmadi Hussin, Aamir Saeed Malik, Shiaw Hooi Ho, Khean Lee Goh, and Ida Hilmi. *Neural Information Processing*, vol. 9489, pp. 373, 2015.