International Journal of Science and Research (IJSR)
ISSN: 2319-7064
Index Copernicus Value (2016): 79.57 | Impact Factor (2017): 7.296

Content Based Image Retrieval System using HSV Color Space to Support Fake News Analysis

S. D. Samantaray[1], Ashish Tyagi[2]

[1]Professor, Department of Computer Engineering, GBPUAT Pantnagar, Uttarakhanda, India

[2]Department of Computer Engineering, GBPUAT Pantnagar, Uttarakhanda, India

Abstract: ***The Content Based Image Retrieval (CBIR) system is applicable for the retrieval of digital images by using their visual contents such as color, shape and texture. In our proposed we had used color as a visual content for the retrieval of images. HSV color space has been used to extract the color feature from the image. This paper presents an approach to overcome from the semantic gap challenge which comes from the high-level human semantics. This paper present a technique to overcome from such challenge and the technique is to applied an elliptical mask on the image and divide the image into five regions i.e., top left, top right, bottom left, bottom right and centre of the image. For the similarity matching chi-square distance has used. This paper also presents an application implementation of the CBIR system for the analysis of fake news. The retrieval results were obtained by applying the HSV histogram on the dataset of around 1025 images of various categories like flower, buses, politician, Egyptian pyramid and nature related images.***

Keywords: Content Based Image Retrieval, chi-square distance, HSV color space, semantic gap

1. Introduction

The main concern in CBIR is the need for an effective and efficient feature extraction method for image representation, which conforms to the subjective human perception. This subjectivity transpires at all semantic levels while analyzing images because different users in the same situation or the same user in different circumstances may investigate or classify the same image differently. This inconsistency between image retrieval, by using low-level image features and high-level human semantics, is termed as the "semantic gap".

In this study, we limited our goal to developing a general comprehensive algorithm for retrieving the images based on primitive visual features. Color is the most expansively used visual feature for image retrieval. Color features are relatively robust to the viewing angle, translation, and rotation of the regions of interest in an image. In general, two types of image features are used to describe an image: 1) color feature 2) holistic structure features. The difference between these features is not always distinct. If spatial distribution is considered when extracting color features of an image, then the color features can be considered as holistic structure color features. Thus, we used an approach to retrieve images based on color feature extraction by using color descriptor named as color histogram (CH).

The CH represents the distribution of color contents effectively in an image when the color pattern is unique compared with the remaining the data set. The CH is easy and fast to compute and very robust for translation and rotation about the view axis. It is used for image retrieval by many commercial systems, such as QBIC, and academic systems such as NETRA, RETIN, KIWI, and Image Minor. To calculate the color histogram we used HSV color model in our proposed work and we had used the 8 bins for Hue channel, 12 bins for saturation channels and 3 bins for Value channels and according to it our each image is represent as a vector which contains 288 values of floating point number.

The second main concern in CBIR is the selection of an effective and efficient similarity matching function, which can do the matching of the two feature vectors of the images and gives the accurate result. There are various choices available for the similarity matching for the feature vectors of the images like Euclidean distance, Manhattan distance and chi-square distance etc. but in our proposed work we had chosen chi-square distance for our similarity matching function.

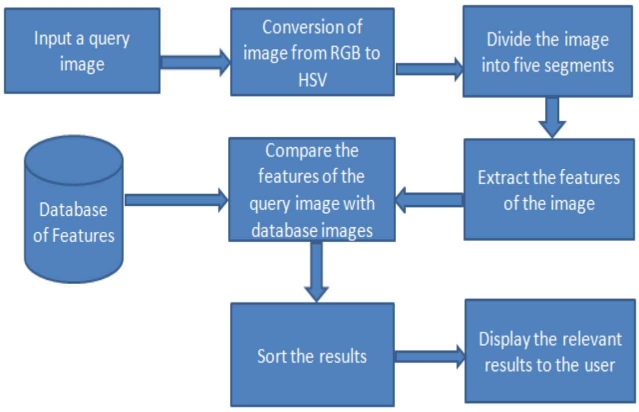


Figure 1: Basic block diagram for CBIR.

The paper is organized as follows: Sec. 2 provides a rigorous literature survey. The methodology used in this paper is presented in Sec. 3. Other existing methods are mentioned briefly in this section as well. Sec. 3 also describes the proposed algorithm, whereas the experimentation is given in Sec. 4. Finally, Sec. 5 concludes the paper and throws light on the scope of future work.

2. Literature Survey

Zhang et al. extracted low-level features by using a 3-dimensional dominant color vector (H, S. and V) and a 24-

dimensional Gabor feature vector. This paper proposed a new approach for digital image retrieval by using intermediate semantic features and multi-step search. This approach suggests a new direction from the existing image retrieval approaches, which works with high- or low-level semantic features. Unlike the existing systems in literature, the proposed system was capable of capturing regional and global features using semantic and low-level features. The results suggested that this system had notable advantages and is more promising compared with the existing techniques. In addition, it has a powerful SQL-based retrieval interface to support semantic and low-level retrieval.

Li et al. represented the image by computing the HSV histogram as a color feature, pyramid wavelet transform (PWT) by using the Haar wavelet as a texture feature (in YCbCr color space), and an edge histogram as a shape feature (in YCbCr color space). This study focused on solving the small sample size problem and improving the capability of a kernel machine compared to traditional SVM-based RFs.

Hong et al. used visual features, such as CM and wavelet moments, for computing feature vectors and comparing the query image and database images by using Mahalanobis distance.

Gosselin and Cord exploited the color and texture information using the L*a*b* space and Gabor filters, respectively. Tests were conducted on the generalist COREL photo database containing 50,000 pictures. This method merged all the semantic information based on binary annotations provided by users during retrieval sessions. Therefore, the kernel matrix framework, which offers acceptable properties of matrices and efficient combinations with kernel-based techniques for image retrieval classifiers, was adopted.

Rui et al. used the CH and CM and co-occurrence matrix for the COREL test set in an interactive approach using relevance feedback. Similarity matrix used for CH vectors was CH intersection and that for CM was Euclidean distance (ED). The authors also used the MESL test set for which the visual features used are the CH, CM, Tamura, co-occurrence matrix, Fourier descriptors, and Chamfer shape descriptors. Chamfer matching was used for Chamfer shape representation, whereas weighted ED was used for the remaining features.

Broilo and Natale presented the feature vector comprising of 32 CH bins, 9 CMs, 16 edge histograms and wavelet texture energy values. The HVS color space was used to extract the CM, and the RB space was used for the CH. Weighted ED was used for similarity matching. The image retrieval problem was formulated as an optimization problem and was solved by using particle swarm optimization.

Jiang et al. used a 9-dimensional CM in the LUV color space, a 64-dimensional CM in the HSV space, a 10-dimensional coarseness vector, and 8-dimensional directionality (Tamura's) texture features for low-level feature vectors. The experiments were conducted on COREL gallery. Generalized manifold ranking-based image retrieval for images was proposed in this study. The proposed method was observed to better than the existing learning methods such as MRBIR, whether or not the query image was in the database, which makes it more suitable for real applications.

Bian and Tao used the CH in the HSV color space, a 128-dimensional CCV in the L*a*b* space, and a 9- dimensional CM feature in the LUV color space. The texture and shape features were extracted using a wavelet transform and the edge directional histogram, respectively. Wenjin Chena et al. reported the improvement of an Image Guided Decision Support (IGDS) system, which was shown to reliably discriminate among malignant lymphomas and leukameia that are sometimes confused with each other during routine microscopic evaluation.

3. Methodology and Proposed Algorithm

The primitive features required to compute an image feature vector are color and holistic structure features. If spatial distribution is considered when extracting color features of an image, then the resultant feature vector is considered to give high retrieval accuracy. Thus, at the very first step we have to choose an image descriptor by which we could retrieve the features

Figure 2: The pipeline of an image descriptor

In our proposed we had chosen color histogram as our image descriptor and we used HSV color space to calculate the Color Histogram of the images. Now, after this our job is that to apply this image descriptor to each image and calculate the features vector from each image and store them in database.

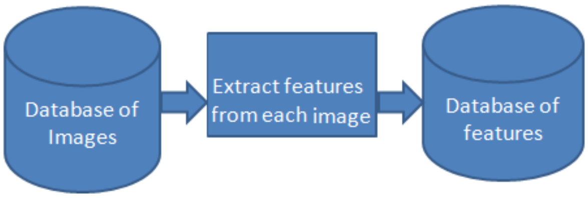


Figure 3: Representation of extracting feature vector.

At this stage we have bunch of feature vectors and now we have to match them with the query image for the retrieval of results. So, now we have to define a similarity function for the matching and we have chosen chi-square distance as our similarity function. In figure it is shown that how matching has done by using chi-square distance as a similarity function in this process of matching each vector of the database image has matched with the query image vector and it gives

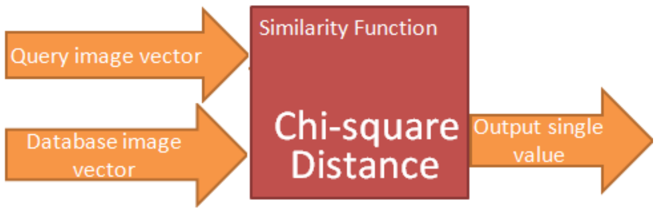


Figure 4: Showing the matching using chi-square distance

a single value as a result of matching of two vectors and we have to store that value in a different vector in a sorted manner.

Proposed Algorithm

Input: An Image
Output: Similar Images
Algorithm:
Step1: Load the database image Dn.
Step2: Transform a RGB image into HSV color space.
Step3: Divide the image into 5 segments top left, top right, bottom left, bottom right and centre.
Step4: Compute the feature vector Fi = $\{f_h\}$
Where: f_h is the HSV histogram.
Step5: Compute the feature vector F_D of all database images.
Where: $F_D = \{F_1, F_2, \ldots\ldots, F_{1025}\}$
Step6: Save F_D
Step7: Input a query image Q_i
Step8: Compute a feature vector $F_q = \{f_h\}$
Step9: Comparison of features of query image with the database image features this comparison is done by chi-square distance. This is formulated as follow:

$$\frac{1}{2}\sum_{i=1}^{n}\frac{(x_i-y_i)^2}{(x_i+y_i)}$$

Where, n is a number of bins in the histogram.
X_i is a value of first bin.
Y_i is a value of second bin.
Step10: After comparison saved the results and sort them.
Step11: Show the results to the user.

4. Results and Discussion

The retrieval of results are based on color descriptor and these results indicated that that color structure based features gives more information than texture features. In our dataset we have various categories of images like flower, buses, politicians etc. the details of the dataset is shown in the following table. In our proposed work the retrieval of the images on the basis of their visual contents by the use of their primitive features under which we have various choices like color feature, shape feature and texture etc. but in our proposed work we had chosen color feature for the retrieval of images and to extract the color feature we had chosen HSV color space. Along with the retrieval images we will also shown the histogram representation of the retrieved images which gives the more clear understanding that how the images are similar with each other.

Table 1: Dataset description

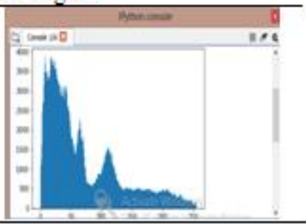
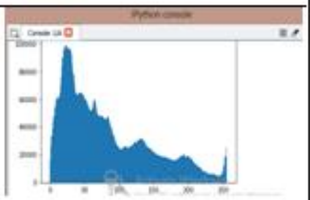
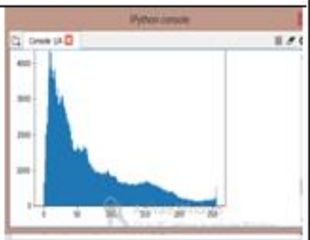
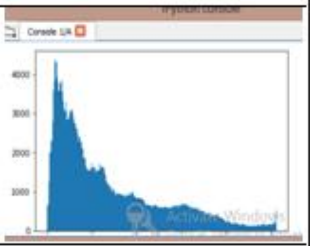
Parameters	Dataset
Number of images	1025
Number of vectors per image	5
Floating point numbers	288
Hue channels	8
Saturation channels	13
Value channels	3
Total floating point numbers	1440

In our proposed work we had taken a dataset of 1025 images which contain the images from various categories like politicians, flowers, buses, Egyptian pyramids etc.

4.1 Results

Evaluation of the results has been done on the basis query image that we had applied as an input image so let's take a case study for an image which illustrate how all the working has been done. In this case study we are showing the histogram representation of query images and the histogram of resultant images are shown in the table 2 and we took a case study of politician images. Each histogram containing the 1440 bins for an image as we mentioned earlier that there are 8 bins for the hue channel, 12 bins for the saturation channel and 3 bins for the value channel so, according to this we have total 288 bins for one image but our image has been partitioned into five regions by using an elliptical at the centre which breaks the image into five regions named top left, top right, bottom left, bottom right and centre of the image. So, each image is defined by the 1440 bins of histogram

Table2: Histogram Representation

Input image	Histogram
	
Results	
	
	
	

5. Summary and Conclusion

The proposed work used HSV color model in order to extract the color features from the images and on the basis of these features the retrieval of similar images has to be done and as we had divided our image into five segments the accuracy gets increased. In proposed work we had used chi-square distance as a similarity function which gives much accurate results in case of histogram matching chi-square distance helps in increase the accuracy of the system. After we get the similarity vector then we sort that vector and retrieve those results which are relevant for us or retrieve the number of images that we want.

The main concern in CBIR is the need for an effective and efficient feature extraction method for image representation, which conforms to the subjective human perception. This subjectivity transpires at all semantic levels while analyzing images because different users in the same situation or the same user in different circumstances may investigate or classify the same image differently. This inconsistency between image retrieval, by using low-level image features and high-level human semantics, is termed as the semantic gap.

There are various applications where we could implement our proposed work like if we talk about the image based news where we are having an image and on the basis of that image the whole being manipulated but by the use of this system we can retrieve similar kind of images if database consist all the related images and then we can check the authenticity of that image. Second application would be we can use our system to diagnose the disease in vegetables or plants as we know there are some fix pattern in the disease of plants or vegetables like if talk about black spot disease then leaf of plants having some black spot on the basis of these signs we can diagnose the diseases in plants or vegetables.

6. Acknowledgement

The author would like to thank Prof. S.D.Samantaray for his valuable discussions and guidance. The authors would also like to thank the anonymous reviewers for their valued suggestions in improving the quality of this paper.

References

[1] Pranoti Mane and Dr. N. G. Bawane, “Comparative Performance Evaluation of Edge Histogram Descriptors and Color Structure Descriptors in Content Based Image Retrieval”, IJCA Proceedings on NCIPET 2013, No.6, pp. 5-9, December 2013.

[2] Y. Liu, D. Zhang, G. Lu and W. Y. Ma, “A survey of content-based image retrieval with high-level semantics”, Pattern Recogn. vol.40, issue1, pp. 262-282, January 2007.

[3] Hamid Sadegi & A.Raie, “Approximated chi-square distance for the histogram matching in facial image analysis”, 10th Iranian Conference on Machine Vision and Image Processing, Nov. 22-23, 2017; Isfahan Univ. of Technology, Isfahan, Iran.

[4] K.Prasanthi Jasmine & P.Rajesh Kumar, “Color histogram and DBC histogram Co-Occurrence matrix for content based image retrieval”,I.J. Information Engineering and Electronic Business, 2014, 6, 47-54 Published Online December 2014 in MECS.

[5] Pranoti Mane and Narendra Bawane, “An approach to explore the role of color models and color descriptors in the optimization of semantic gap in content based image retrieval”, International Journal of Computer Applications (0975 – 8887),Volume 104 – No.14, October 2014.

[6] T.V.Madhusuhana Rao, Dr.S.Pallam Setty & Dr.Y.Srinivas, “An efficient system using for medical image retrieval using gamma distribution”, I.J. Image,

Graphics and Signal Processing, 2015, 6, 52-58 Published Online May 2015 in MECS.

[7] Yong Rui, Thomas S. Huang, Fellow, IEEE, Michael Ortega, and Sharad Mehrotra, “Relevance Feedback: A power tool for interactive content based image retrieval”, IEEE transactions on circuits and systems for video technology, vol. 8, no. 5, september 1998.

[8] Manesh Kokare, B.N.Chatterji & P.K.Biswas, “A Survey on current content based image retrieval methods”, JETE Journal of Research Vol 48, Nos 3 & 4, May-August 2002, pp 261-271

[9] Y. Liu, D. Zhang, G. Lu and W. Y. Ma, “A survey of content-based image retrieval with high-level semantics”, Pattern Recogn. vol.40, issue1, pp. 262-282, January 2007.

[10] Arnold W. M. Smeulders, Marcel Worring, Simone Santini, Amarnath Gupta,and Ramesh Jain, “Content based image retrieval at the end of the early years.”, IEEE Trans.Pattern Anal. Mach. Intell 22(12), pp.1349–1380, 2000.

[11] Manesh Kokare, B.N. Chatterji & P.K.Biswas, “Comparison of similarity metrics for Texture Image Retrieval”.

[12] Cristian Jos´e L´opez Del Alamo_,Lizeth Joseline Fuentes P´erez†, Luciano Arnaldo Romero Calla‡, Wilber Roberto Ramos Lov´on§, “A novel approach for image feature extraction using HSV model color and filters wavelet”, 2013 XXXIX Latin American Computing Conference (CLEI)

[13] P. P. Mane, N. G. Bawane and A. Rathi, “Image retrieval using primitive feature extraction with hybrid approach”, National Conference ACCET-15, India, pp. 99–102, 2015.

[14] M.Kunaver, student member, IEEE, J.F.Tasic, Member IEEE “Image feature extraction- an overview” Serbia & Montenegro, Belgrade, November 22-24, 2005

[15] Jingjin Guo, Lizhen Liu*, Wei Song, Chao Du, 2Xinlei Zhao “The Study Of Image Feature Extraction and Classification” Information and Engineering College, Capital Normal University, Beijing 100048, P. R. China 2Foreign Language College, Capital Normal University, Beijing 100048, P. R. China, December 14, 2017.

[16] Kuo-Kun Tseng1, Jiaqian Li1 and Lantian Wang1 “A New Statistical-Based Algorithm for Medical Image Feature Extraction”, Information and Engineering College, Capital Normal University, September 21, 2017.

[17] Sidheswar Routray, Arun Kumar Ray and Chandrabhanu Mishra “Analysis of Various Image Feature Extraction Methods against Noisy Image: SIFT, SURF and HOG”, 2017 Second International Conference on Electrical, Computer and Communication Technologies (ICECCT), November 23, 2017.

[18] Wu Chen1 , Wang Fubin ,Chen Xianzhong.et.al “RGB Color Decomposition and Image Feature Extraction of Flame Image in Rear of Sintering Machine”, Proceedings of the 36th Chinese Control Conference July 26-28, 2017, Dalian, China.

[19] Zhenmin Zhu, Ruichao Song and Shiming Chen “Zhenmin Zhu, Ruichao Song and Shiming Chen”, 31st Youth Academic Annual Conference of Chinese Association of Automation Wuhan, China, November 11, 2016.

[20] Rajkumar Goel, Vineet Kumar.et.al (2017) “A Review of Feature Extraction Techniques for Image Analysis”, International Journal of Advanced Research in Computer and Communication Engineering”, February 2, 2017.

[21] Hamid Sadhegi & Abolghasem-A.Raie (2017) “Approximated Chi-Square Distance for Histogram Matching in Facial Image Analysis: Face and Expression Recognition”, 10th Iranian Conference on Machine Vision and Image Processing, Nov. 22-23, 2017; Isfahan Univ. of Technology, Isfahan, Iran

[22] E. Boutellaa, M. Bengherabi.et.al(2015) “Face verification using local binary patterns generic histogram adaptation and chi-square based decision”

[23] Samyeul Noh(2012) “χ 2 Metric Learning for Nearest Neighbor Classification and Its Analysis”, 21st International Conference on Pattern Recognition (ICPR 2012) November 11-15, 2012. Tsukuba, Japan.

[24] Jasman Pardede , Benhard Sitohang.et.al(2017) “Comparison of Similarity Measures in HSV Quantization for CBIR”, 2017 International Conference on Data and Software Engineering (ICoDSE).

[25] Nalini Pasumarthi & Lakshmi Malleswari(2015) “An Empirical Study and Comparative Analysis of Content Based Image Retrieval (CBIR) Techniques with various similarity measures”.