IMAGE RETRIEVAL BASED ON THE COMBINATION OF RGB AND HSV'S HISTOGRAMS AND COLOUR LAYOUT DESCRIPTOR

Javier Poveda Figueroa[1,*], Vladímir Robles Bykbaev[2]

Abstract

In this paper we present the first stage of a new approach to improve the precision and recall of the content-based image retrieval task. To do this, we aim to combine three colour features, RGB and HSV histograms, and MPEG-7 Colour Layout Descriptor. To perform the combination, we propose to use an approximation based on Borda Voting-Schemes. Under that the Borda Voting-Schemes needs at least three votes to perform the combination, we intend to use the K-Nearest Neighbors methods to select the candidate images, given a query image. In the second stage, we'll implement our approach using at least three image databases.

Keywords: HSV histogram, RGB histogram, Colour layout fescriptor, Borda Voting Schemes, KNN.

Resumen

En este artículo se presenta la primera fase de experimentación de una nueva propuesta para mejorar la precisión y cobertura de la recuperación de imágenes basada en contenidos. Para realizar esta tarea, se propone combinar tres características de color, como son los histogramas RGB y HSV y el Descriptor de Distribución de Color del estándar MPEG-7. A fin de llevar a cabo la combinación, se plantea emplear una aproximación basada en Esquemas de Votación Borda. En virtud de que dichos esquemas requieren al menos tres votos, se propone usar la técnica de los K - Vecinos más cercanos, con el objetivo de seleccionar las imágenes candidatas a partir de imagen de consulta. En la segunda etapa se implementará nuestra propuesta empleando al menos tres bases de datos de imágenes.

Palabras clave: Histograma HSV, Histograma RGB, Colour Layout Descriptor, Borda Voting Schemes, KNN.

[1,*]*Colaborador del Laboratorio de Investigación en Sistemas Informáticos e Inteligencia Artificial, Carrera de Ingeniería de Sistemas, Universidad Politécnica Salesiana, sede Cuenca. Autor para correspondencia: jpoveda@est.ups.edu.ec*
[2]*Máster en Inteligencia Artificial, Reconocimiento de Formas e Imagen Digital, Ingeniero en Sistemas, Estudiante de Doctorado en Informática, Universidad Politécnica de Valencia, Encargado del Laboratorio de Investigación en Sistemas Informáticos e Inteligencia Artificial, Universidad Politécnica Salesiana, sede Cuenca*

1. Introduction

Nowadays, the Content-Based Image Retrieval (CBIR) is a technique that has several areas of application. For example, in the medical area, the CBIR applications help to detect non-melanoma skin lesions [1], breast cancer [2], while in forensics area it is possible to detect altered images [3]. Another application that successfully uses CBIR is a device that tries to help in navigation, and object location for blind people [4]. These successfully developed applications are a real example that CBIR techniques have a broad scope of applications.

The constant growth of the visual information, like pictures and videos, makes necessary to have efficient methods to retrievethis kind of information in a proper way [5].

The MPEG-7 standard defines a set of visual descriptors to extract several image features. An example of these features is: image edges and color distribution [6], the shape of objects in the image [7], the texture characteristics [8], etc.

With this approach, we try to improve the level of precision or recall in the images retrieval task, using the combination of three colour features: RGB and HSV's histograms, and MPEG-7 Colour Layout Descriptor. To conduct the combination, we use a new alternative, the Borda Voting-Schemes.

The rest of the paper is organized as follows. In the Section 2 we review the fundamentals of the colour features like RGB and HSV histograms and MPEG-7 Colour Layout Descriptor (CLD). The K - Nearest Neighbors method is described in Section 3. In Section 4 we review the Borda Voting-Schemes. In Section 5 we describe our approach to combine the RGB, HSV and CLD features to perform the image retrieval task. Conclusions and future work are reviewed in the Section 6.

2. Colour features fundamentals

2.1 RGB histogram

RGB is the acronym for Red, Blue and Green. This histogram is the most used histogram in computer graphics and it uses its red, green and blue components to create a new color [5]. This color histogram combines red, green and blue colors to form a new color. To form a new color it is necessary to increase the values of one or more of the components of the RGB components [9].

The RGB color histogram can be represented on the Cartesian coordinate system, which is illustrated in Figure 1. As it shows a new color can be formed by changing the values of the RGB component. For example, if the magenta color is needed, it is necessary to combine green and blue to create it [9].
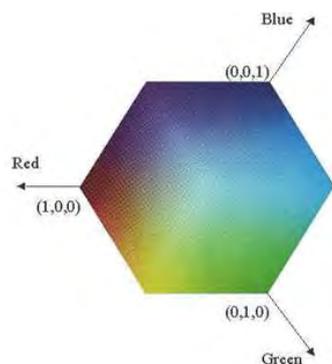
Figure 1. RGB coordinate system where the "x" axis represents red colors, "y" axis represents green colors and "z" represents blue colors [10]. As this figure shows changing the values of RGB component can form a new color.

2.2 HSV histogram

HSV color histogram is the acronym of Hue, Saturation and Value; this color histogram is

the closest perception of the human eye. As [11] explains, the human eye perceives colors by the excitation of two cells of the eye, which are rods and cones. *"Excitation of the cone cells leads to perception of color, while rod cells help in perception of various shades of gray. [...] The HSV color histogram separates the luminance component (Intensity) of a pixel color form its chrominance components (Hue and saturation)."* This representation works as the human eye because it works like the separation of the rods and cones.

On the representation of the components for this histogram Hue represents the chromatic component, saturation represents the predominance value of a hue a color and Value represents the intensity of the color [10].

The generation of colors of the HSV color space is represented in the next way. According to [12], Hue defines the color by changing its angle; hue is defined as an angle in the range $[0, 2\pi]$. Saturation is the depth or purity of the color and is measured as a radial distance from the central axis with value between 0 at the center to 1 at the outer space. Finally, value is represented by the vertical central axis.

In Figure 2 is illustrated how we represent an HSV color histogram.

2.3 Colour layout descriptor

This MPEG-7 visual descriptor represents the spatial layout of color images in a very compact form [6]. To extract the CLD descriptor, we assume the original image is represented in RGB color space, and the necessary steps to perform this task are the next:

– *CLD block calculation.* In this step the image is divided in 64 blocks of pixels. The width and height of each block are calculated
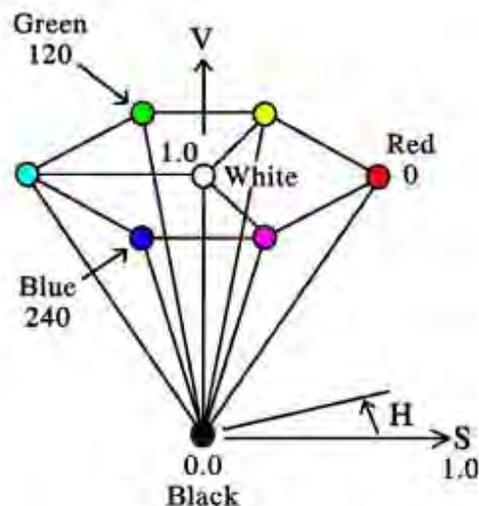
Figure 2. Representation of a HSV color descriptor. The value V is represented by the main axis orthogonal of the plane. The angle represents the Hue value, while radius represents the level of Saturation (purity of colour) [13].

by dividing the width and height of the image for 8, respectively. This process is performed for the three components of RGB colour space and we obtain 3 matrices of 64 blocks.

– *Representative colour selection.* Each block must have only one colour. To select the representative colour, the values of all pixels in the block are averaged (Figure 3). After that, each 8x8 matrix is transformed to YCbCr colour space [1]

(a) Original image   **(b)** Image with representative colour selected
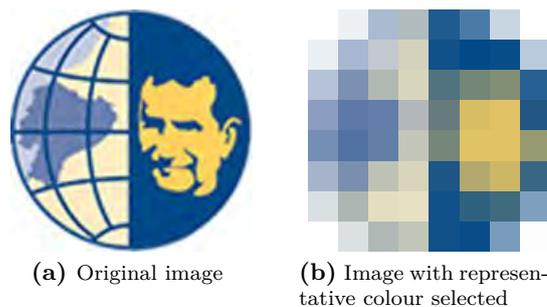
Figure 3. Result to applying representative colour selection in image in RGB colour space.

---

[1] The YCbCr is a family of color spaces used in digital video and image transmission. The component Y represents the luminance and Cb and Cr are the blue-difference and red-difference chrominance components.

– *DCT (Discrete Cosine Transform)*. In this step the three colour components of image (Y, Cb and Cr) are transformed by 8x8 DCT to obtain three 8x8 DCT matrices of coefficients [6]. To calculate the DCT in a 2D matrix, the following formula is used:

$$B_{pq} = \alpha_p \alpha_q \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} A_{mn} \cos \frac{\pi(2m+1)p}{2M}$$
$$\cos \frac{\pi(2n+1)q}{2N} \quad (1)$$

Where:

- B_{pq} is the coefficient for *p,q* position of the DCT matrix.

- *M* is the number of rows of the matrix.

- *N* is the number of columns of the matrix.

- α_p and α_q are constants which value are calculated as follows:

$$\alpha_p \begin{cases} \frac{1}{\sqrt{M}}, & p=0 \\ \sqrt{\frac{2}{M}}, & 1 \le p \le M-1 \end{cases}$$

$$\alpha_q \begin{cases} \frac{1}{\sqrt{N}}, & p=0 \\ \sqrt{\frac{2}{N}}, & 1 \le p \le N-1 \end{cases}$$

– *Zigzag reordering*. The final step consists on zigzag reading of the DCT coefficients (Figure 4). This should be done in this way because the nonzero coefficients always appear in the top-left corner of the coefficients matrix. Usually, the CLD employs the firs 6 coefficients of the Y colour component, and 3 coefficients of *Cb* and *Cr* colour components.

3. K - Nearest Neighbors

K-Nearest Neighbors, also known as KNN, is a supervised machine learning method, which
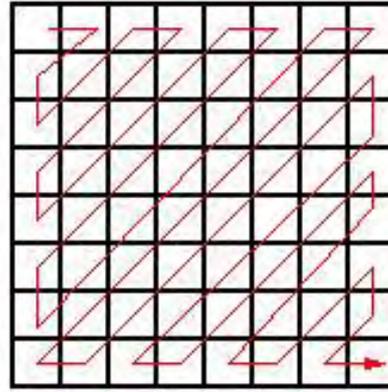
Figure 4. Zigzag scanning applied to Y, Cb and Cr components of the representative colour image.

classifies the data of a corpus. With this classifier a set of data can be classified in order to discover which elements are from the set of data.

To solve the classification problem the algorithm uses a series of clusters of data, which each one represent a category. When a new kind of data appears, the KNN algorithm tries to recognize which category the data belongs.

Figure 5 is illustrated how the KNN algorithm works [14].
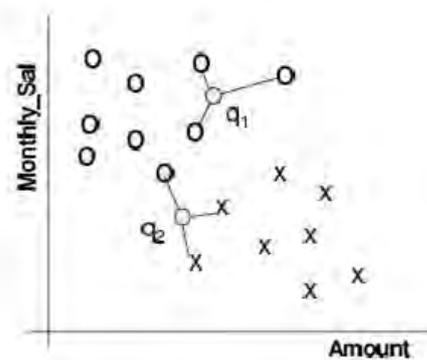
Figure 5. Representation of a KNN algorithm which main goal is to discover if the data, which we are seeking, is from Monthy-Sal or Amount.

To find to which set of elements the new data belongs to, a distance formula is used, which permits the calculation of the distance

of one element to another, to calculate we use equation 2 [14].

$$D_{a-b} = \sqrt{\sum_{i=1}^{N}(a_i - b_i)^2} \qquad (2)$$

Equation 2. Euclidean distance to calculate the distance between two elements, where b_i represents the position of the first element and a_i represents the position of the second element [15].

4. Borda voting scheme

Borda voting - scheme is a technique, which can help to classify the data of a corpus; this method has its origins in France it the 18th century. It was proposed by Charles Borda because the voters did bad decisions in the process of voting; for this reason Borda proposed a method by which the voters must vote for the candidates by its merits [16].

Following the Borda scheme this method consists in a group of candidates, which vote for the elements of a set of data by its relevance. For this method we have a set of candidates X=$\{x_1, x_2, x_3, \ldots, x_n\}$ where $n \geq 3^2$ is and m voters, where $m \geq 3$ [13]. To calculate the voting results it is necessary to follow the next steps:

Firstly, we choose a photograph and we have three categories, which gives us 3 options per each candidate. In the Table 1 shows the options and the candidates of each element.

Table 1. Voters with its different options

	Candidate 1	Candidate 2	Candidate 3
Option 1	2	2	3
Option 2	3	3	1
Option 3	1	3	2

Secondly, we represent the data taken by the candidates to form a square matrix, which is formed by the number of options we have; in this case we have a 3 x 3 matrix. We fill the matrix with 1 in each row that the candidates have preference and 0 in the rest of positions [17].

$$\begin{bmatrix} 0 & 0 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 1 & 1 & 0 \end{bmatrix}$$

Finally, we sum all the elements of each row and we get an array of data, which shows us the order of classification of the image [18].

$$\begin{bmatrix} 2 \\ 4 \\ 3 \end{bmatrix} \begin{bmatrix} c1 \\ c2 \\ c3 \end{bmatrix}$$

5. Experiment desing

The process for this experiment is divided in two parts. The first part of this experiment is the extraction of features of the images, quantization of the images features, shuffle of the images features and getting the train and test sets for experimentation. In the second part we use the KNN algorithm and the Borda voting - schemes to get the image candidate for classification.
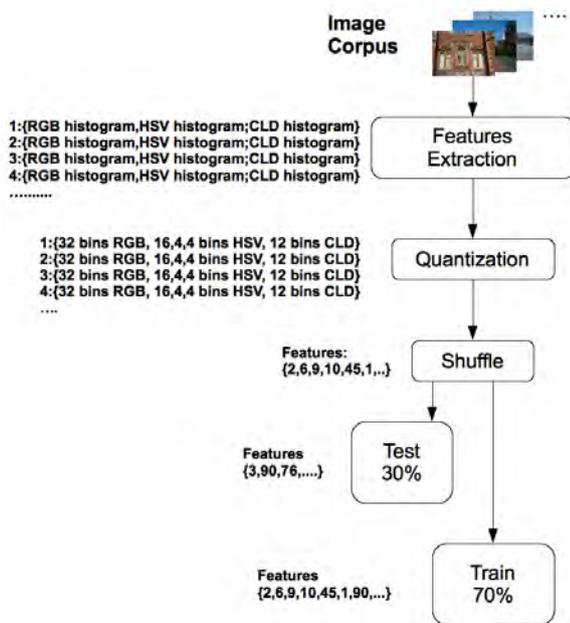
The first part goes as follows:

–*Features Extraction:* It helps to extract the bins of all images, apply a conversion method to transform images to three descriptors, in this case HSV and CLD, and RGB descriptors, so that we have three descriptors to start our work.

– *Quantization:* The quantization method reduces the dimensionality of the image features. This method is used to improve the calculation process of the distance between two features set.

– *Shuffle:* The objective of this process is to mix the image set of features as a necessary stage for the experiment, because the system must not learn the image features in the same order each time.

– *Test and Train for experimentation:* This process is used for dividing the image feature set in two subsets to do experimentation. This two set of images features are the training set and the testing set. The training set has 70 % of the features, which is used as the base to compare the features to find a new one, and the testing set which is going to be compared with the training set of images to find to which category it fits in. In Figure 6 is explained the first part of the process.


Figure 6. Representation of the extraction of features of the image corpus, quantization, shuffle of the images and get the test and train sets of images features.

The second part goes as follows:

– *KNN Algorithm:* It is used for performing the classification process based on each feature set. To do this process we take each image features of the test set and compare them to each image features of the training set. As

a result the system returns a set of votes and candidates for each descriptor.

– *Borda voting - schemes:* This ensemble method is used to classify the image, to do this step we use the votes we obtained on the earlier step. We use the candidates and votes that we obtained of the KNN process and generate a matrix with each result we had in the previous process, then we sum the matrixes and we get an array of data which results in the classification order of the candidate. As a result the system can classify the image in a category. Figure 7 explains the second part of the experimentation design.
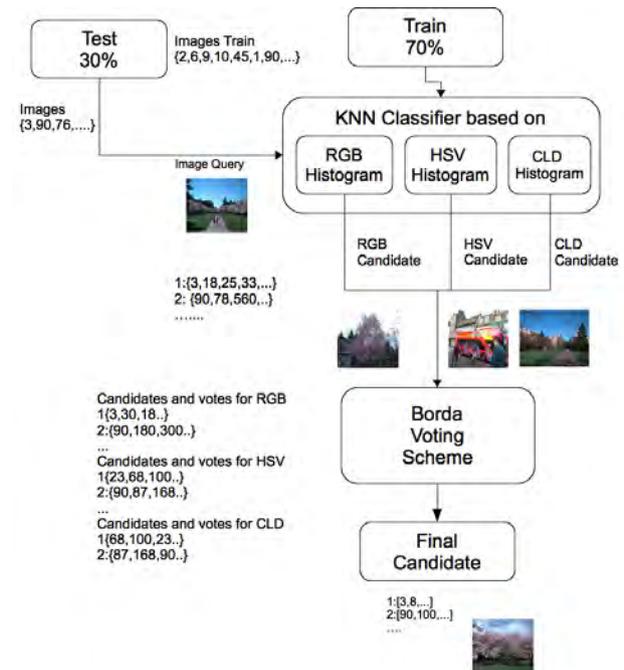
Figure 7. The second stage of the experiment process.

6. Conclusions and future work

The main goal of this project is to improve the precision and recall of the system, so that we can get best results at discovering and recognizing the patterns of the image.

Given that MPEG-7 standard defines a complete set of visual and audio descriptors, is

important to analyze if is possible to combine other visual descriptors like Edge Histogram, Dominant Color, etc. In the same way, it is important to study more combination methods like AdaBoost, Decision Trees, etc.

For future work we will start the experiments we designed in this paper with different image databases of different themes in order to improve the system.

Ackowledgment

References

[1] L. Ballerini, X. Li, R. Fisher, and J. Rees, "A query-by-example content-based image retrieval system of non-melanoma skin lesions," *Medical Content-Based Retrieval for Clinical Decision Support*, pp. 31–38, 2010.

[2] D. Tahmoush, "CBIR for mammograms using medical image similarity," *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, vol. 7628, p. 8, 2010.

[3] T. Kijak, T. Furon, and L. Amsaleg, "Deluding image recognition in SIFT-based CBIR systems," in *Multimedia in Forensics, Security and Intelligence (MiFor)*, Florence, Italy, October 29 2010.

[4] S. Caperna, C. Cheng, V. Fan, A. Luthra, B. O'Leary, J. Sheng, A. Sun, L. Stearns, R. Tessler, P. Wong, and J. Yeah, "A nav-igation and object location device for the blind," *University of Maryland*, 2009.

[5] N. Singhai and S. Shandilya, "A survey on: Content based image retrieval systems," *International Journal of Computer Applications IJCA*, vol. 4, no. 2, pp. 22–26, 2010.

[6] R. Balasubramani and V. Kannan, "Efficient use of MPEG-7 color layout and edge histogram descriptor in CBIR systems," *Global Journal of Computer Science and Technology*, vol. 9, no. 4, pp. 157–163, 2009.

[7] S. Berretti, A. Del Bimbo, and P. Pala, "Retrieval by shape similarity with perceptual distance and effective indexing," *IEEE Transactions on Multimedia*, vol. 2, no. 4, pp. 225–239, 2000.

[8] T. Deselaers, D. Keysers, and H. Ney, "Features for image retrieval: an experimental comparison," *Information Retrieval*, vol. 11, no. 2, pp. 77–107, 2008.

[9] K. Mekaldji, S. Boucherka, and C. S, "Color quantization and its impact on color histogram based image retrieval," in *Procedings of the Second Conference Internationale sur l'Informatique et ses Applications (CIIA'09)*, Saida, Algeria, May 3 - 4 2009.

[10] J. Miralles, *Tutorial de GIMP*. [Online]. Available: http://sites.google.com/site/tutorialdegimp/011---teoria-del-color-for%macion-y-mezcla-de-colores -rgb-y-cmyk

[11] S. Jeong, "Histogram-based color image retrieval," Stanford University, Palo Alto, CA, Psych221/EE362 Project Report, 2001.

[12] A. Vadivel, S. Sural, and A. Majumdar, "Human color perception in the HSV space and its aplication in histogram generation for image retrieval," in *SPIE Procedings seetings*, San José CA,United States of America, 2005.

[13] S. Sural, G. Quian, and S. Pramatik, "Segmentation and histogram generation using the HSV color space for image retrieval," in *Procedings International conference on Image Processing*, 2002.

[14] University of Auckland. Department of Computer Science, *CBIR: Color Features*. [Online]. Available: http://www.cs.auckland.ac.nz/compsci708s1c/lectures/Glect-html/topic3c708FSC.htm

[15] P. Cunningham and S. Delany, "K-nearest neighbors," Report of the UDC School of Computer Science and Informatics, Dublin, Ireland, Tech. Rep., 2007.

[16] Carleton College, *Computer Science Comps Project. Netflix prize.* [Online]. Available: http://cs.carleton.edu/cs_comps/0910/netflixprize/final_results/knn/index.html

[17] V. Robles, "Esquemas de votación borda aplicados al etiquetado de roles semánticos," Master's thesis, Universidad Politécnica de Valencia, Valencia, Spain, 2010.

[18] ——, "Borda based voting schemes for semantic role labeling," in *13th International Conference on Text, Speech, and Dialogue. Lecture Notes in Computer Science*, Brno, Czech Republic, September 2010.