# Optimize color feature extraction of Iraqi paper currency for classification application

**Shaimaa Hameed Abd, Dr.Shaimaa Hameed Shaker, Dr. Ivan A. Hashim**

**Abstract--One of the most important and efficient requirements in image clustering, classification, retrieval and indexing is extraction of different color features from images. Untiled now, color features are one of the most popular and widely used visual features. In this paper, three types of color features are used in order to find the most effective and efficient features can be used to classify the Iraqi paper currency, these are color moment (MVS), color histogram (CH) and color coherence vector (CCV). In MVS features three type of color moment are extracted, these are $\mu_i$ ,$\sigma_i$, $\delta_i$ .For CH, each image described by its colors and the pixels number in each color and the number of vectors for this features is equally to 256 vector. The last extracted color feature is CCV feature which considered an improvement technique for CH features, its produce double number of vectors based on the define color range. The final step that performed on the extracted features is features minimization process. In which the features vectors from the previous step is minimized into less set without effect on the classification performance. The simulation result for the MVS was equally to 97.86%. while for CH feature 98.7% and the result for CCV features is equally to 100%.**

***Shaimaa Hameed Abd**, Computer Engineering Department, University of Technology, Baghdad, Iraq.*
Dr. ***Shaimaa Hameed Shaker**, Computer Engineering Department, University of Technology, Baghdad, Iraq.*

***Dr. Ivan A. Hashim,** Electrical Engineering Department, University of Technology, Baghdad, Iraq.*

***Index Terms*—color features, features extraction, color histogram, color coherent vector, CBIR.**

## I. Introduction

In recognition process, CBIR system and image processing, feature extraction is a process in which, reducing dimensionality of recognition process. When input data to processing system algorithms is too large to be processed or organize, its necessary to transformed these huge data into less one (deal with only important one), which collectively called feature vector. This transformation process called feature extraction. after features extraction is selected carefully then only relevant information are stored from large input data. Images captured usually contain different important features such as color, size and other hidden features.

On the other hand, Color feature is one of the most important kind of features that make recognition process easy this could have attributed that color easy to recognized by humans and colors are the most commonly applicable visual features in images retrievals application and in different image processing filed. Beside this, color features are easy to extract, matched, recognized, and its effectiveness found in indexing and searching for specific color image in large databases [1].

For currency identification process, there's an important step Before feature extraction step these step should be holed first, which called image-object extraction. In object extraction step is related to extract important objects from background (unimportant objects) in the input image.

Quad histogram method is one of the popular techniques that used for features extraction. Is used to determine the homogenous area with different size, and the quad tree decomposition is used on input images. The quad histogram method shows acceptable efficiency and accuracy in image retrieval [2].

for (CBIR) application it depends on different characteristics about the image, for example the color, shape and textures of the image. CBIR widely applied in a wide range of application such as finical transaction, internet, searching and biomedicine. The researchers used a technique that called Multistage CBIR. This technique is constructed of three-layer feed forward, each layer responsible for specific comparison. The first layer compare color features, second layer compare texture features and the third layer compare shape features [3].

The main aim of this research is to explain the operation of the color moment (MVS), color histogram (CH) and color coherence vector (CCV). Apply these features on the Iraqi dinar (IQD), and also determine how features are selected and minimized in order to perfectly identify the Iraqi dinar (IQD).

## II. Color Moment

The color moments are used for color feature extraction. The distribution of colors in the image describe by the supposition that the color moments can be translate as probability of distribution. the probability distributions are described by a number of unique moments (e.g., normal distributions are differentiated by their mean and variance) [4],[5].
In different research, it shown that the most common of color distribution information is controlled by the lower order moments [6]. these color moments are mean, variance and skewness [7].

The first order color moment called mean $(\mu i)$, which described as the average color in the image and it can be calculated by the (1) [8]:

$$\mu_i = \frac{1}{N}\sum_{j=1}^{N} f_{ij} \quad (1)$$

where $f_{ij}$ is the value of the *i*-th color component of the image pixel *j*, and *N* is the number of pixels in the image.

The second color moment called variance $(\boldsymbol{\sigma\, i})$,which is computed by taking square root of the variance of the color

distribution, the second order color moment is described in (2) [9],[10].

$$\sigma_i = \sqrt{\frac{1}{N}\sum_{j=1}^{N}(f_{ij} - \mu_i)^2} \quad (2)$$

while the there'd color moment is skewness ($s_i$), and it described by the (3) [11].

$$s_i = \sqrt[3]{\frac{1}{N}\sum_{j=1}^{N}(f_{ij} - \mu_i)^3} \quad (3)$$

Since only 9 (three moments for each of the three color components), so that the mean($\boldsymbol{\mu i}$)variance ($\boldsymbol{\sigma i}$) and skewness ( $\boldsymbol{\delta i}$) are computed for each $\boldsymbol{i}$equal to (1,2,3) [7].

## III. Color Histogram

In CBIR application, color histogram (CH) is widely applied and considered most popular applied color features. This color features are computed by applying CH on RGB image. However, its computation caricaturized by easy and efficiency. [12].

The CH represents by a plot that describe the color content of the image, its counts the number of pixels in each color level in an image [13], the Figure (1) show the CH diagram of RGB image.

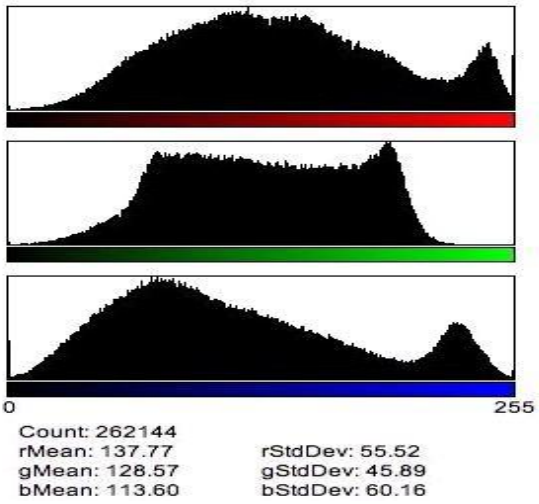


Figure (1): Color histogram diagram of three channels (Red, Green and Blue).

The CH represents of the image is rotation, translation and scale invariant, so that it is very appropriate for color-based CBIR. But there is drawback of the use CH, they do not take in account of space information concerning the color. This may lead to unexpected errors [14]. Table (I) summarized difference between the global color histogram (GCH) and local color histogram (LCH) [15].

Table (I): the difference between the GCH and LCH.

| LCH | GCH |
|---|---|
| 1. Segments the image into blocks and the compute the CH for each block | 1. Represent the single image in one CH |
| 2.when comparing two images, distance is computed by comparing their CH, between a region in one image, and between the same region location but in different image. | 2. the use of GCH the color image will be encoded with it is CH. |
| 3. the final distance between the two images will computed by summing all these distance. | 3. the final distance between two images will computed by calculating the distance between their CH |

The CH are flexible to contrast in different color space like RGB, HSV and other color space and in any dimension. Any pixel in the image can be described according to it is color space, in CH the pixels described by its location in each bin [16].

## IV. Color Coherence Vector (CCV)

CH used for featured extraction in many image processing applications. The CH have different advantages such as simplicity, efficiency and insensitivity. however, to their advantage, the CH have lack of special information of pixels, this lead to similar color distribution for different images [17], for this reason, there is another method that combine the histogram–based method. the special information can be classifying each pixel in
the color region to be either coherent or incoherent based on these pixels, in which are part of color region or not. The CCV store both coherent and incoherent pixels within each color, and provide separation between the both value.

The CCV technique is derived from CH method. The Color Coherence vector takes in the consideration some special information about the pixels within some color coherence area. The CCV classify the pixels in two type either coherence or incoherent. The coherent pixels are a part of defined sizable, contiguous color region, whereas incoherent pixels are not a part from this region. A CCV perform this classification for each color image. The most important feature about CCV is prevent coherent pixels in one color image from matching incoherent pixels in another image. This would allow to exact

distinctions that cannot be detected by color histogram [18].

The starting stage of computing a CCV issimilar to computing CH. First blurring the image, a bit by replacing each pixel value with the average value in a small neighborhood area (an average of center pixel value with the 8 adjacent pixels), moreover, the color space is defined with only n distinct colors in the image. the follow step is classification of pixels within a given limited color space as coherent or incoherent pixels. Name the number of coherent pixels for $j^{th}$ discrete color $\alpha j$ and the number of incoherent pixels $\beta j$. The total pixel number with that color is $(\alpha j + \beta j)$, and the CH would describe image as in (4).

$$( \alpha 1 + \beta 1, \alpha 2 + \beta 2, .., \alpha n + \beta n ) \quad (4)$$

Instead, for each color compute the pair $(\alpha j, \beta j)$ which called the coherence pair for the jth color. The CCV pairs vector for the color image describedby (5).

$$\mathbf{(}( \alpha 1, \beta 1\mathbf{),} (\alpha 2 , \beta 2),\ldots, (\alpha n , \beta n\mathbf{) )} \quad \mathbf{(5)}$$

## V. Features Selection

Features selection process is widely applied in different applications, as a tool for removing irrelevant or unwanted features. There's no stander or single algorithm for feature selection process because of each application have a requirement and methods different from other. The feature selectin process subject can be divided as following:

### A. Characteristics of Chose Feature Selection

Feature selection methods are choosing based on different data set characteristics; these are:

- Data type: different type of data can be used, but for feature the data type lie in two areas, either for features or for class labels as represent below:
  - features: value for each feature can be either continuous (C), discrete (D), or nominal including Boolean (N).
  - Class label: depend on feature selection method that can be either deals with only binary class (two class) or multiple (more than two) class
- Data size: data size deals with the following:
  - If the method is performed on small training set.
  - Or the method is performed on large data set.
- Noise: noise is typically insert through the feature selection operation. however, the noise insertion can lead to:
  - Misclassification
  - Conflicting data

### B. Basic Steps for Feature Selection Methods

Feature selection methodology try to find best subset features among multiple subsets. There are four important steps for features selection methods [19] that shown in Fig. (2), these are as following:

- ❖ a generation procedure: to generate the next candidate subset, this procedure is also called as a search procedure. This step generates subset of features for evaluation procedure. It begins either with no features, with all features or with random features.

- ❖ an evaluation function: to evaluate the subset under examination;
- ❖ a stopping criterion: to decide when to stop. This step is basically depending on the output of the both generation and evaluation. The stopping rule based on
- ❖ generation procedure as the following:
  - whether a predefined number of features are selected,
  - and whether a predefined number of iterations reached.

The stopping rule based on evaluation function as bellow:

- whether addition (or deletion) of any feature does not produce a better subset.
- whether an optimal subset according to some evaluation function is obtained.

- ❖ a validation procedure to check whether the subset is valid. Most of commonly used method for this validation procedure were based on:
  - using artificial datasets
  - using real-world datasets.

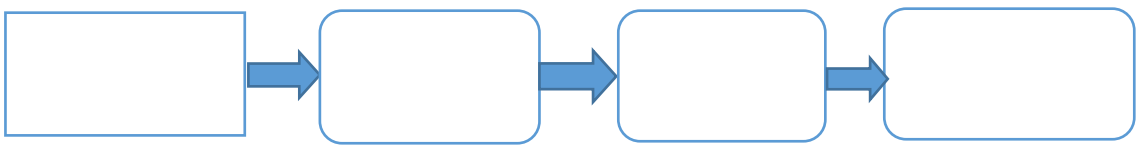

Figure (2): Basic steps for features selection method

## VI. Proposed Method

The proposed method consists of three stages. The first stage is image acquisition or image capturing. The second stage is image-object extraction and the last stage is feature extraction.

### A. Image Acquisition

The first step is image acquisition or image capturing, in which all images samples that used for both training and testing the overall algorithm is introduced. In this stage different devices can be employed for capturing image some of them can be either scanner, camera, phone camera or these images can be taken from any website could be acceptable in both size and resolutions. In this research, images capturing done by using the scanner, which made by brother company, and it have the specification that shown in Table (II).

The scanned image in this research is stored in a database named as IQPC. This database contains about 280 images about Iraqi currencies with different denominations. The scanned image for each denomination is captured in different position, direction and rotation angles. the scanned banknotes are shown in Fig. (3).

Table (II): Technical specification for the scanner.

| | |
|---|---|
| **Optical Resolution** | **1200 x 2400 dpi** |
| **Interpolated Resolution** | **19200 x 19200 dpi** |
| **Gray Scale Depth** | **8 bit** |
| **Color Depth** | **24 bit** |
| **Color Depth (Internal)** | **30 bit** |
| **Model** | **Brother/ MFC-J470DW** |

Figure (3): The scanned image of 500 IQD.

### B. Image-Object Extraction

ImageObject extraction is an important step in different image processing techniques and CBIR (content base image retrieval) applications. The main objective of this processing stage is to extract important and only relevant object from the image. The steps of the proposed image-object extraction are shown in theFig. (4).

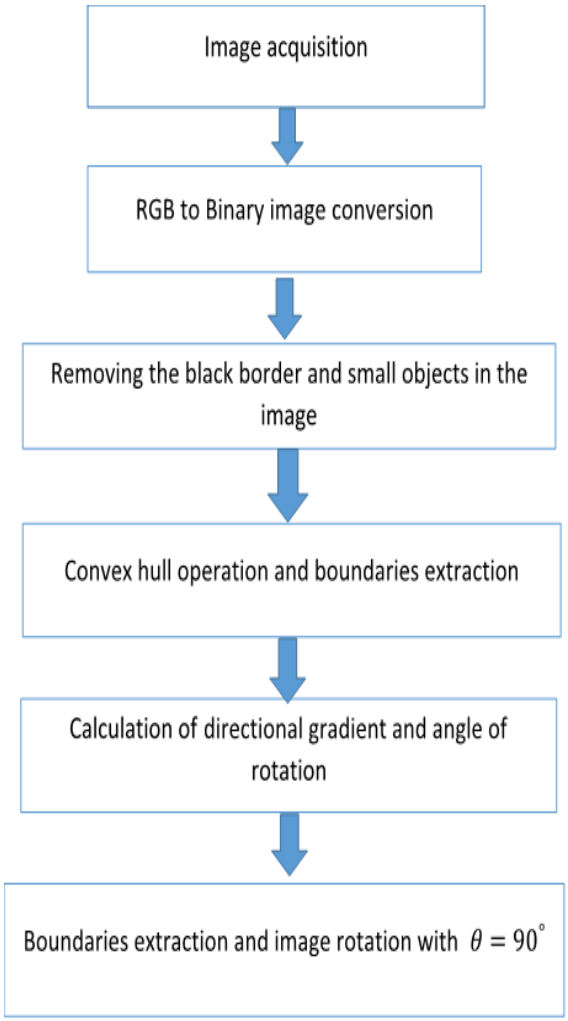


Figure (4): Image object extraction block diagram.

### C. Feature Extraction

This research is based on selection three types of features. these features basically depend on colors; most apply visual feature are based on color which considered the easy way for recognizing by human. It is also widely applied in different researches.

The suggested color features in this paper are:

- Color moment
- Color histogram (CH)
- Color coherence vector (CCV)

**a. Color Moment**

This features consist of three value, which are mean, variance and skewness. The features calculation is explained in detail in section two. this type of features are expressed in three variable $(\mu, \sigma, \delta)$ which all are stored in form of vector. The three value are calculated for each color band (red, green, blue) of RGB image, making vector of nine values. The form of final vector is expressed as ($\mu_1$ , $\mu_2$, $\mu_3$ , $\sigma_1$, $\sigma_2$, $\sigma_3$ , $\delta_1$, $\delta_2$ , $\delta_3$ ), in which calculated for each input currency note.

The reason of applying this method in this work are the simplicity of calculation, or in another meaning these three features mathematical calculation characterized by easily application and reduction in time of calculation.

b. **Color Histogram (CH)**

CH consider as color feature that applied on image that represented in HSV color space. The steps that used to find this type of feature are:

Step1: RGB image are read from the stored database, therefore converted to HSV color model.

Step2: the quantization process in the CH refers to the operation of minimizing the number of bins in CH, by making colors that have closer values to each another and put them in one bin.

In HSV color model the Hue (H) is quantized into 16 bins which represented in range [0-15], Saturation (S) is quantized into 4 bins and it buts in range equally to [0-3], and Value is quantized into 4 bins and these are in range [0-3]. This step can be described by (6).

$$QL = 16 \times H + 4 \times S + V \quad (6)$$

Where the QL is an integer value in the range between [0-255].

Step 3: The CH of image that represented in HSV color model is obtained, in which, CH is plot that represent distribution of colors in the image. Digital image of CH is describing the number of pixels that have the same color in a fixed color range that cover colors of an image.

Step 4: the final step is discarded the first and last sixth color in CH, which represent the whitest or darkest region in the image. These discarded values do not effect to the overall operation of CH.

There are several reasons that make color histogram rarely to apply in different researches. one of these that CH compute 256 value for each image, which make it difficult to store or retrieves these value specially when have large database. The second reason, the CH compute the number of pixels in each color band and don't care to another related information in the image, in which make CH with the same plot for two image (only pixels in each color without considered related information). However, the CH apply in this work with some modification in order to get the best advantage from this method

### c. Color Coherence Vector (CCV)

CCV is powerful color features, that is apply on color image. The CCV function required serval input to complete its operation. The first input is color image (or RGB image) which is the base for this feature. Second important input is the percentage of the image size to consider a component's pixels are coherent. The third and it's the last important input in this process (number of Colors), which represents the number of different colors in the Color Coherence Vector (default = 27 colors). While the final output from this is CCV feature values, which is matrix represent the input image in CCV feature.

This computed matrix contain both coherent ($\alpha$) and incoherent ($\beta$) values, moreover this matrix would display double number of vector (vector number equal to color number). For more arrangement in display this matrix it should perform reshape operation in order to display coherent and incoherent ($\alpha \; and \; \beta$) value of the same index as pairs (in one vector) as explained in section four. This method represents an improvement from CH feature. However, considered related information to each pixel and don't depend only on colors in its calculations.

### d. Features Minimization

Feature selection process applied in a different application in which remove of all irrelevant features, or remove features that don't effect to overall results. The features selection process involved four steps, these steps established as: generation procedures (random type), evaluation function (Euclidean distance), stopping criterion and validation procedures. Features minimization are based on the idea of features selection.

Because of the large data, the last two features type (CH and CCV), should apply technique to minimizes these features vectors to become easy to process, store and retrieve in less time.

In this work, features selection process pass through four stages, and the techniques for each stage are selected accurately in order to get correct performance after check by Euclidean distance. Features minimization process for each feature is explained as following:

1. MVS features minimization
   Figure (5) show the performance accuracy verses the number of selected features in the MVS type. this plot high evaluation function performance rate appears when the number of features is seven only and result is 97.86, while low evaluation function performance rate appears when applying only one features and the rate closed up to 60%. Therefore, the selected feature

vector number of MVS are $[(\mu_1, \mu_2, \mu_3, \sigma_1, \sigma_2, \delta_1, \delta_2)]$

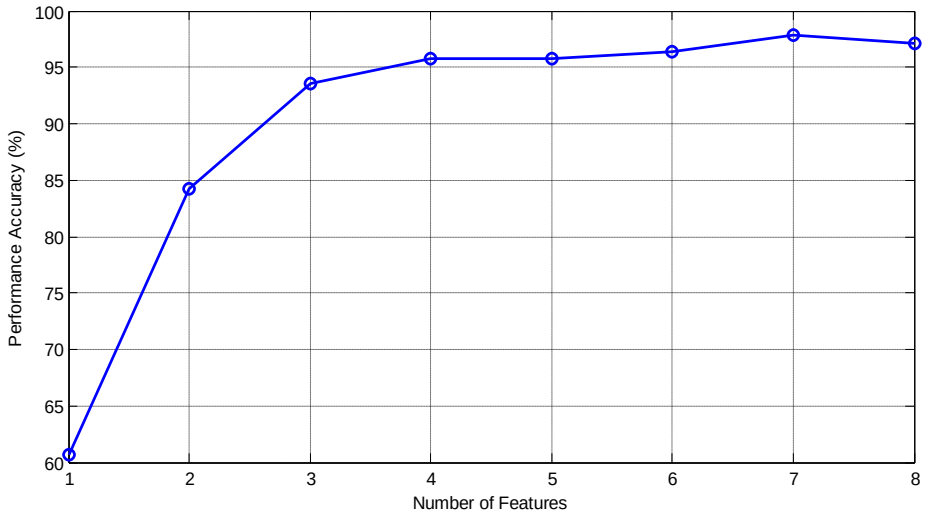


Figure (5): The evaluation function performance with different number MVS feature

2. CH features minimization

The CH calculate 256 feature for each image, this number of features are so large and it is necessary to be minimized. Fig. (6) shows the measure of evaluation function performance rate with different feature selection of CH. low rate shown when selected features in range [228 to 237] and the performance rate is less than 90%. While the high rate appears when feature number equally to 55 feature with a performance rate for these are higher than 98%. The selected features vector number is 55 feature in which make the performance higher than 98%, these feathers number are

[ 1 9 26 27 33 40 41
44 52 53 54 57 58 59 62
63 64 73 82 87 88 90
95 97 98 99 100 103 104
112 113 119 123 124 128 129
133 136 144 152 165 174 177
188 190 195 204210221222224 232
233 238 244]

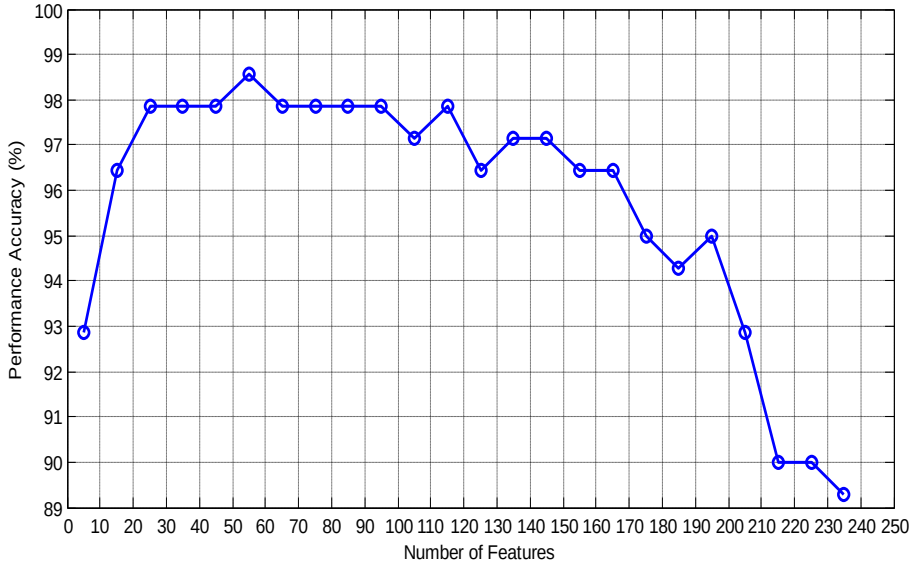


Figure (6): The evaluation function performance with different number CH feature

3. CCV features minimization

Figure (7) shows the evaluation function performance accuracy for CCV features type. high performance rate was observed when the number of features is five only and result is equally to 100%, while low performance rate appears when applying only two features and the rate is equally to 88%. The selected features vector number are $[\beta_4 \beta_{15} \alpha_{17} \alpha_{18} \beta_{24}]$ which give 100% performance rate.

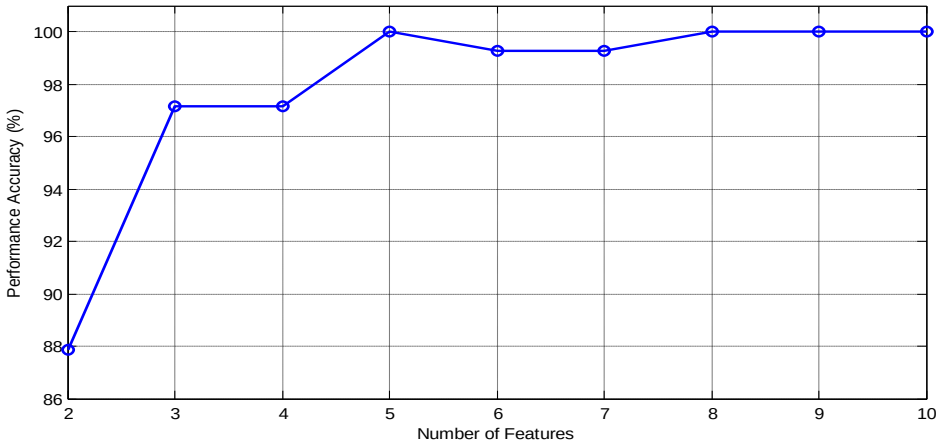


Figure (7): The evaluation function performance with different number CCV feature

## VII. Conclusion

In this paper color features extraction for Iraqi dinar(IQD) based on color moment (MVS), color histogram (CH) and color coherent vector (CCV) are explained. These features are extracted and tested for different denominations for Iraqi dinar.

After features extraction process, the features minimization should have performed in order to store only important and relevant vectors that highly effect on classification process. After minimization process the number of important vectors for MVS features is only seven, for CH features is only fifty-five vectors and for CCV features is only five relevant features vectors. The simulation result for the MVS was equally to 97.86%. While for CH feature 98.7% and the result for CCV features is equally to 100%.Finally, the best applied color feature used for Iraqi paper currency identification is CCV feature.

**Reference:**

[1] A. P. and S. G.S, "Image Retrieval from Database using Different Image Features", in International Conference on Computational Systems for Health & Sustainability, Bangalore, 2015.

[2] F. A. and M. R. K.," A New Color Feature Extraction Method Based on Quad Histogram", 3rd International Conference on Environmental Science and Information Application Technology, Hong Kong, 2011.

[3] N. Chauhan and M. Goyani, " Enhanced Multistage Content Based Image Retrieval", International Journal of Computer Science and Mobile Computing, Vol. 2, No. 5, pg.175 – 179, 2013.

[4] S. Das, D. Rudrapal and R. Sarkar, "An Efficient Method for Content Based Image Retrieval Using Color Moment and Texture Descriptors," in International Conference on Computing, Communication and Information Technology (ICCCIT 2012), Chennai, 2012.

[5] H. H., P. K. Bhuravarjula and V. V. Kumar, "A Novel Image Retrieval Method using color Moments," *International Journal of Electronics and Computer Science Engineering*, Vol. 1, No. 4, pp. 2432-2438, 2012.

[6] S. R. Kodituwakku and S. Selvarajah, "Comparison of Color Features for Image Retrieval," Indian Journal of Computer Science and Engineering, Vol. 1, No. 3, pp. 207-211, 2010.

[7] A. Singla and M. Garg, "CBIR Approach Based on Combined HSV, Auto Correlogram, Color Moments and Gabor Wavelet," *International Journal of Engineering and Computer Science*, Vol. 3, No. 10, pp. 9007-9012, 2014.

[8] A. Thakur and A. Dhole, "Object Recognition from Image Using Grid Based Color Moments Feature Extraction Method," International Journal of Research in Engineering and Technology, Vol. 2, No. 3, pp. 333 -336, 2013.

[9] M. Mosbah and B. Boucheham, "The Influence of the Color Model on the Performance of a CBIR System Based on Color Moments," Journal of Communication and Computer, Vol. 11, No. 3, pp. 266-273, 2014.

[10] M.Dey, B. Raman and M. Verma, "A Novel Color- and Texture-Based Image Retrieval Technique Using Multi-

Resolution Local Extrema Peak Valley Pattern and RGB Color Histogram," *Pattern Analysis & Applications*, Vol. 19, No. 4, p. 1159–1179, 2016.

[11] A. R. Kumar and D. Saravanan, "Content Based Image Retrieval Using Color Histogram," *International Journal of Computer Science and Information Technologies*, Vol. 4, No. 2, pp. 242 - 245, 2013.

[12] K. Meskaldji, S. Boucherkha and S. Chikhi, "Color Quantization and its Impact on Color Histogram," 1st International Conference on Networked Digital Technologies, Ostrava, Czech Republic, 2009.

[13] P. Kaur, S. Thakral and M. Singh, "Color Based Image Retrieval System," IOSR Journal of Computer Engineering, Vol. 1, No. 5, pp. 1-5, 2012.

[14] M. Mustikasari, S. Madenda, E. Prasetyo, D. Kerami and S. Harmanto, "Content Based Image Retrieval Using Local Color Histogram," International Journal of Engineering Research, Vol. 3, No. 8, pp. 507-511, 2014.

[15] A. Al-Hamami and H. Al-Rashdan, "Improving the Effectiveness of the Color Coherence Vector," The International Arab Journal of Information Technology, Vol. 7, No. 3, pp. 324 - 332, 2010.

[16] K. Roy and J. Mukherjee, "Image Similarity Measure using Color Histogram, Color Coherence Vector, and Sobel Method," International Journal of Science and Research (IJSR), Vol. 2, No. 1, pp. 538 - 543, 2013.

[17] G. Pass and R. Zabih, "Histogram Refinement for Content-based Image Retrieval," in Application of Comuter Vision, Sarasota, FL, USA, 2002.

[18] G. Pass, R. Zabih and J. Miller, "Comparing Images using Color Coherence Vectors," Proceedings of the 4rth ACM International Conference on Multimedia, Boston, 1996.

[19] M. Dash and H. Liu, "Feature Selection for Classification," Intelligent Data Analysis,Vol. 1, No. 4, pp. 131-156, 1997.