Image Retrieval Based on Edge Histogram Descriptor of MPEG-7

Bilal Ahmed[1] & Musa M. Ameen[2] & Mohammed S. Anwar[3]

[1]Information Technology Department, Tishk International University, Erbil, Iraq
[2&3]Computer Engineering Department, Tishk International University, Erbil, Iraq
Correspondence: Bilal Ahmed, Tishk International University, Erbil, Iraq.
Email: bilal.ahmed@tiu.edu.iq

Doi: 10.23918/eajse.v6i2p189

Abstract: A major research area in computer vision is content-based image retrieval. MPEG-7 sets up a list of descriptions of the structured image content. We examine the weakness and lack of retrieval approaches based on global characteristics in this study by incorporating the commonly used feature descriptors of MPEG-7. In the meantime, to satisfy user requirements for assessing spatial information similarities, an image retrieval approach based on texture region features for MPEG-7 is recommended. Retrieval tests show the validity and efficiency of our approach. This paper also defines our approach to color quantization, extraction, and matching processes of features and so on in depth.

Keywords: MPEG-7, CBIR, Edge Histogram Descriptor

1. Introduction

The retrieval of content-based images is characterized as retrieving related images from image databases by assessing the similarities of image content (CBIR). To measure image similarity in current CBIR systems, visual characteristics, such as texture, shape, spatial relationship, and color (gray scale), are often introduced. It is necessary to extract the features of each image to speed up the process of extracting content-based images from an image database, and then use these image features to create an image database index. So, an important topic for discussion is how to characterize an image's features. Provided a sample image, a feature matching procedure is performed to scan for images identical to the sample image from a broad image database.

MPEG-7 is a format that has been provided by MPEG (Moving Picture Experts Group). The MPEG-7 format, officially known as the "Multimedia Content Description Interface," offers a rich range of structured multimedia content description tools and facilitates content-based multimedia management.

The color histogram is invariable for translation and rotation, one of the most common color measures, and thus easy to determine. But spatial color data cannot be interpreted and varies with intensity changes and color distortion (Chow, 2006). Some enhanced color histogram methods have been proposed to fix this problem (Pass, 1997; Huang, 1997), but they still lack the fuzziness of colors usually represented by the color histogram. For retrieving content-based images, the texture function is also useful.

Received: September 4, 2020
Accepted: November 26, 2020
Ahmed, B., & Ameen, M.M., & Anwar, M.S. (2021). Image Retrieval Based on Edge Histogram Descriptor of MPEG-7. *Eurasian Journal of Science and Engineering*, 6(2), 189-199

Sim et al. (2004) suggested an algorithm for texture retrieval, first computing the power spectrum of an initial texture image, and then using the power spectrum of the normalized image and modified Zernike moments to retrieve image texture features, this algorithm retains insensitivity to some form of combination of localization, scaling, and rotation, minimizing computational difficulty compared to ease.

All gradient vectors generated from sub-images of LL, LH and HL are used to improve the texture-discriminating ability of this method to define the texture function by wavelet decomposing an input (original) image, as defined in (Huang et al., 2003). Most of the CBIR systems currently available are based on various variants of the global image characteristics that are simple to formulate and rotation and translation invariant, but these global image characteristics are unable to determine semantic similarity in terms of spatial distribution, which may lead to poor retrieval performance.

A CBIR medical imaging survey was conducted by (Akgül et al., 2011). The writers explored CBIR's existing state of the art medical imaging techniques. They came up with CBIR's latest problems and prospects in the process of medical diagnostics. The authors sought to center the researcher's attention on organizational challenges in medical CBIR and recommended some methods to resolve them. Using the color moment and Gabor texture function, the new CBIR technique was proposed by (Huang et al., 2010). To obtain the color moment, they transform it to an HSV image from RGB, and calculate the three moments for each color space by obtaining the equalized histogram of the three components of the HSV.

The purpose and significance of this study is to include an overview of MPEG-7 visual descriptors and act as a reference or a catalogue so that reader, researchers, scholars, and other users to get familiar with the content-based image retrieval using MPEG-7. In addition, the mechanism of computing using MPEG-7, and EHD are mentioned.

A certain technique for automated image classification is implemented after feature vectors are obtained using MPEG-7 descriptors to get an optimized result from the query image. First, we include a description of the extraction process of EHD features in the rest of the paper. The image classification system, with a thorough explanation of tuning parameters, is then described. Finally, the methods of study and test results are presented and discussed.

This study paper is divided into five major section, Section I is related to the introduction part to the content-based image retrieval and MPEG-7. CBIR is mentioned in detail in section II. Edge Histogram Description and functionalities of the descriptor is discussed in Section III. Experimental results discussed in Section IV. Section V presenting conclusion part.

2. Content Based Image Retrieval

Content-based image retrieval (aka Query by Image Content (QBIC)) refers to the method of extracting expected images according to the content of the image query from large image databases. CBIR is a topic that receives significant interest from the scientific population interested in research on computer vision. The purpose of automated image classification for a given input image is to assign it to the image group that is most comparable in terms of visual content.

CBIR varies from conventional search, which is based on metadata such as image-related keywords, tags, or explanations. Because of the set of preset parameters, the keywords often restrict the scope of queries. While the CBIR allows for the ability to scan for content that does not have predefined requirements. The visual similarity decision between images depends on image descriptors which, to decide the degree of their similarity, should be able to classify the images. Feature vectors are derived because of the descriptor algorithm, giving a numerical representation of the image content. Such a descriptor will generate high-variance characteristics and a respectable distribution over the group structure. Furthermore, at different stages of image quality and resolution, the extraction method of the description should be stable.

The average efficiency of the classification scheme is ultimately influenced by the option of the descriptor. The MPEG-7 specification defines various graphic descriptors for still images to provide ang get standardized video and image content descriptors.

Three classes of color descriptors may be differentiated based on the image properties used to derive them: shape descriptors, texture, color descriptors, and motion descriptors. Although MPEG-7 indicates 11 distinct descriptors of images (5 colors, 3 textures and 3 shapes), most of them are highly dependent on each other (Eidenberger, 2004). Therefore, for good image classification, correct collection of only a few independent descriptors is necessary. Statistical study of image recovery MPEG-7 descriptors in Eidenberger (2004) has shown that Color Structure, Dominant Color, Texture Browsing, and Edge Histogram Descriptor are the best descriptors for mixing. The others rely heavily on these ones.

These are the several descriptors that are part of the MPEG-7 family. They are listed as.

- Color descriptors
  Which include: Scalable color, Dominant color, color temperature, color layout, GofGop color, color structure, and Illumination invariant color.
- Texture Descriptors
  Which include: Homogeneous texture, texture browsing, and edge histogram descriptor.
- Shape Desciptors
  Which include: Contour shape, shape 3D, region shape, and perceptual 3D shape.
- Motion Desciptors
  Which include: Motion trajectory, camera motion, motion activity, and parametric motion.
  Other descriptors include: Face identity descriptor, localization descriptor, video signature, and image signature.

3. Edge Histogram Descriptor (EHD)

The Edge Histogram Descriptor (EHD) describes the spatial distribution of 5 edge orientation forms in subimages called local image partitions (Manjunath et al., 2001). Texture, like paint, is a strong low-level descriptor for applications to scan and download images. Edge distribution is a strong texture signature that is useful for matching picture to image. MPEG-7 offers multiple descriptors to capture texture functionality, and one of the most common descriptors is EHD, which is helpful when texture properties are not homogenous in the underlying region. The Edge Histogram Descriptor collects the edge spatial distribution, much in the same spirit as the Color Layout Descriptor (CLD). EHD edges of capture, which are divided roughly into five groups: Vertical, Horizontal, Diagonal in 45 degrees, Diagonal in 135 degrees, and Isotropic (Non-orientation specific).

The input image is sub divided into 4x4(16) non overlapping blocks. Each extracted block is further divided into 2x2 blocks for capturing local edge orientation. If image dimensions are not completely divisible by 4 then resize image so that it become 4x4 image in size as shown in Fig. 1. An edge histogram can be used to represent the local-edge distribution for each of the 4x4=16 sub-images. The MPEG-7 norm decomposes the input (original) image into 4x4=16 sub--images to solve this kind of problem and represents each subimage with a 5-bin uniform edge histogram (one for each edge orientation) (Cvetković & Nikolić, 2011). Edges in the sub-images are extracted to produce the edge histogram and divided into five groups based on the orientation. Since there are 16 sub-images, it produces a histogram with total of 16x5 = 80 bins. Because of its high precision and invariance to resolution, EHD is one of the most used descriptors (Batko, et al., 2010).

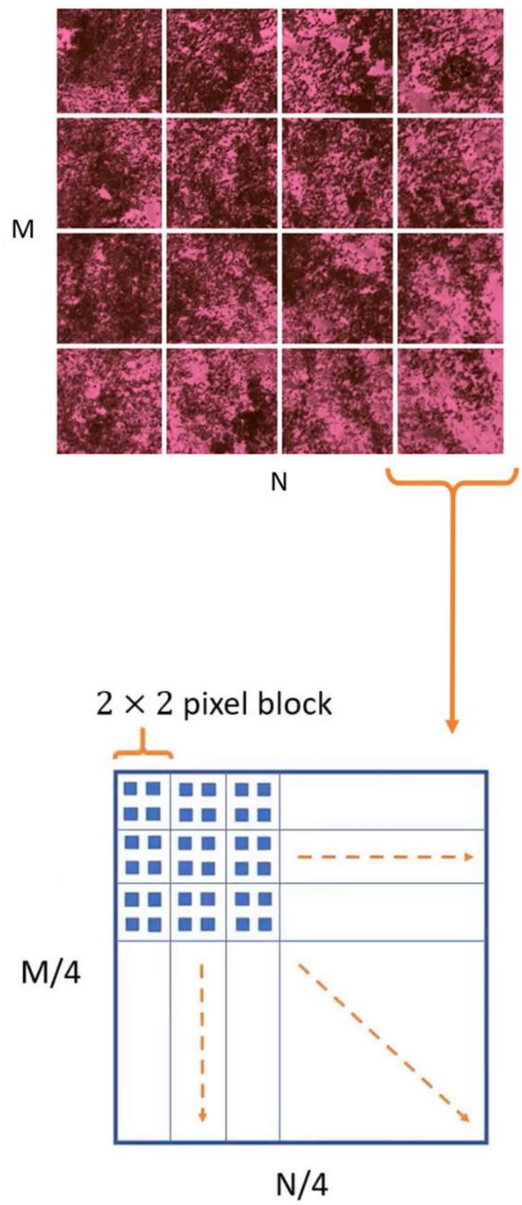


Figure 1: Input image block distribution of 4x4(16)

EHD extraction begins with an edge orientation measurement using the five edge operator masks shown in Figure 2. Edge orientations are not computed directly on image pixels to allow resolution invariance. Instead, the image is first sub divided into predefined block numbers (4x4) and each block is sampled down to 2x2 pixels by basic average sampling.

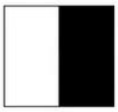
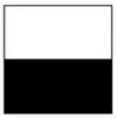
Edge Type	Visual Representation	Operator Mask
Vertical Edge		$\begin{bmatrix} 1 & -1 \\ 1 & -1 \end{bmatrix}$
Horizontal Edge		$\begin{bmatrix} 1 & 1 \\ -1 & -1 \end{bmatrix}$
Diagonal (45^0)		$\begin{bmatrix} \sqrt{2} & 0 \\ 0 & -\sqrt{2} \end{bmatrix}$
Diagonal (135^0)		$\begin{bmatrix} 0 & \sqrt{2} \\ -\sqrt{2} & 0 \end{bmatrix}$
Non-Orientation Type Edge		$\begin{bmatrix} 2 & -2 \\ -2 & 2 \end{bmatrix}$

Figure 2: Five masks of EHD

Magnitudes are then determined on 2x2 samples for five edge mask responses, and the highest value is selected as the dominant edge orientation. The edge histogram should be computed after the dominant edge orientation for each image block has been determined (4x4 blocks). If an image is to be represented by just one global edge orientation histogram, it is possible to use the same edge orientation histogram for two images with entirely different edge distributions (Cvetković & Nikolić, 2011). Each operator is applied on 2x2 block as in Equation 1.

$$EO_{type} = \left|\sum_{k=0}^{3} a_k . d_k\right| \quad [1]$$

Where $[a_k] = \begin{bmatrix} a_0 & a_1 \\ a_2 & a_3 \end{bmatrix}$ represents 2x2 image sub block and $[d_k] = \begin{bmatrix} d_0 & d_1 \\ d_2 & d_3 \end{bmatrix}$ represents the edge detector. For each block of $\left(\frac{M}{4} \times \frac{N}{4}\right)$, five point bin is initialized as Bin=[V, H, D45, D135, NOE] = [0, 0, 0, 0, 0]; which represents Vertical Edge Orientation (V), Horizontal Edge Orientation (H), Diagonal Edge Orientation at 45 degrees (D45), Diagonal Edge Orientation at 135 degrees (D135), and Non-edge Orientation (NOE).

Maximum of these five values is compared with a threshold value (T) to find dominant edge orientation as $(EO_{dominant}) = \max(EO_v, EO_h, EO_{d45}, EO_{d135}, EO_{noe}) > T$. The $EO_{dominant}$ will be equal to anyone of these 5 orientations (which is maximum among them). The count of the corresponding Bin Point is increased by one and this process will be repeated for all 2x2 sub blocks in one $\left(\frac{M}{4} \times \frac{N}{4}\right)$ image block. For each $\left(\frac{M}{4} \times \frac{N}{4}\right)$ image block, we get the complete Bin as explained above. This Bin can be expressed as, $Bin[1] = [b_0, b_1, b_2, b_3, b_4]$, and this operation will be repeated for all sixteen $\left(\frac{M}{4} \times \frac{N}{4}\right)$ image blocks to get all 16 bins.

There are 80 bins in the full edge orientation histogram, generated by concatenating 16 sub-image histograms, each comprising 5 bins as shown in Figure 3.

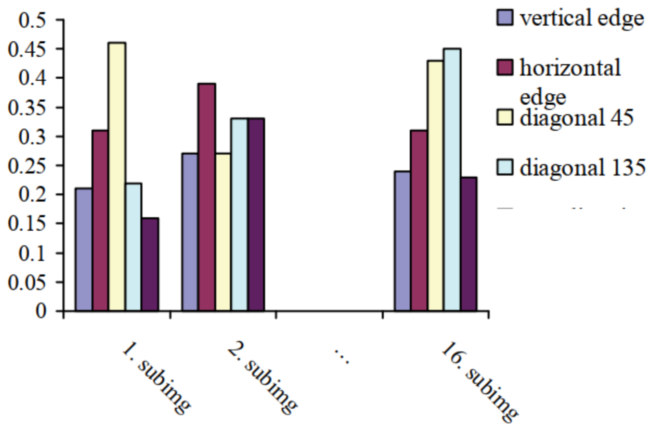


Figure 3: Sample of Edge Histogram Descriptor (EHD)

Finally, feature vector can be represented as:

$$\mathbf{f}^{\mathbf{EHD}} = (f_1^{EHD}, f_2^{EHD}, f_3^{EHD}, \ldots, f_{79}^{EHD}, f_{80}^{EHD}) \quad [2]$$

A global bin can also find by taking mean (column wise) of this bin's matrix as $GlobalBin = mean(AllBins)$. If this $GlobalBin$ is combined with all calculated Bins, the length of the EHD vector will be 85 and EHD vector will be represented as.

$$EHD = [Bin[1], Bin[2], \ldots, Bin[16], GlobalBin] \quad [3]$$

4. Experimental Results

We used the so-called COREL1000 image dataset (Wang, 2001), which is commonly used to validate image classification systems, for the purpose of testing the process. It consists of a total of 1000 pictures divided into 10 categories: citizens of Africa, beaches, houses, dinosaurs, buses, elephants, horses, trees, food, and mountains. The 100 photos in each group which makes total of 1000. Samples of images are presented in Figure 4.

It should be noticed that images within the group show broad differences in the colour and artifacts they contain, rendering the research database like requests for real-world applications. We also evaluated image classification output using different EHD feature vectors for two images, as well as using a collection of images accessible in the dataset. Using our implementation of part of the MPEG-7 specification in MATLAB, EHD features have been extracted. Tests were replicated over numerous randomly chosen COREL1000 dataset training and test image subsets.

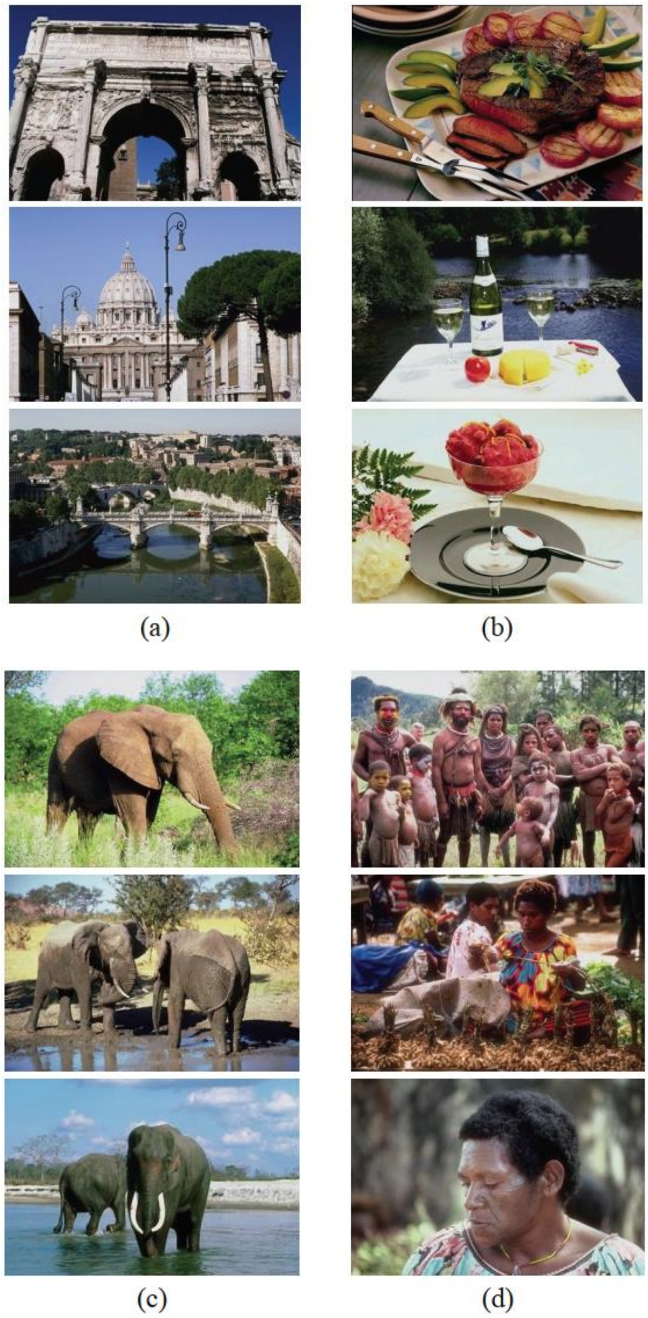


Figure 4: Samples of a few different category of dataset images: (a) Buildings (b) Food (c) Animals (d) Africa people (Wang, 2001)

The proposed method first applies feature extraction on image database using Edge Histogram Descriptor (EHD) of MPEG-7, then all the extracted features will be saved in features database. Later when a query image is received then same feature extraction algorithm will be applied on input image. The result of feature extraction will be checked against feature database using similarity check. Then the nearest result will be shown as an output to the user as shown in Figure 5.

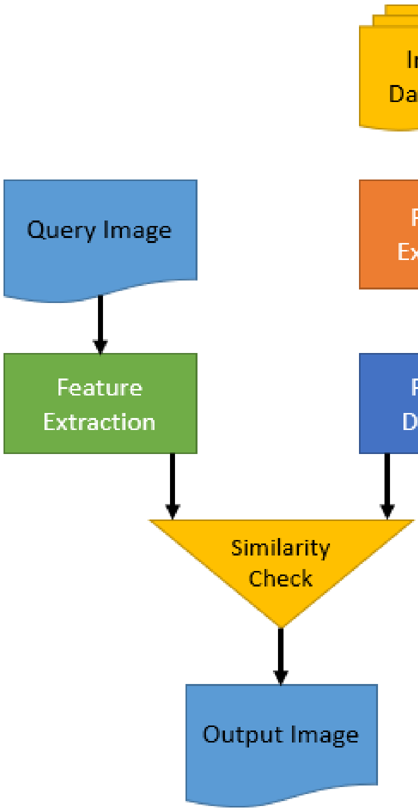


Figure 5: Samples of a few different category of dataset images: (a) Buildings (b) Food (c) Animals (d) Africa people (Wang, 2001)

Figure 6 shows the similarity of two images along with the global bin histogram. The bins indicate the histogram similarity for vertical, horizontal, diagonal 45, diagonal 135, and noe respectively. Table 1. shows distance value difference between two images available in Figure 6.

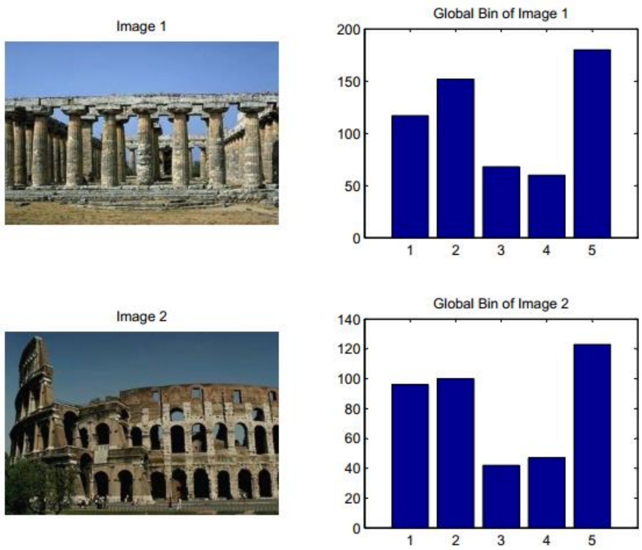


Figure 6: Global bin of two sample images along with histogram

Table 1: Distance difference between two images

Type	Value
L2 Distance	783.1437
L1 Distance	4780

The difference between two images with vertical and horizontal texture is tested using two sample images as shown in Figure 7. The bins 1 and 2 indicate the difference of vertical and horizontal texture, respectively.

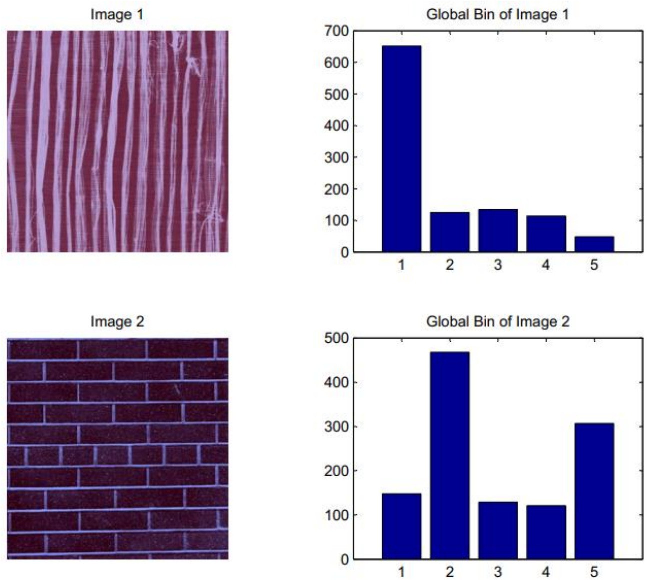


Figure 7: Global bin of two sample images along with H and V texture

Another test has been conducted to retrieve best related images that are similar visually to query image. The proposed method compares the input image with all images available in the dataset. Then it will sort the top 5 images from the result based on the difference in total bins as shown in Figure 8 and 9.

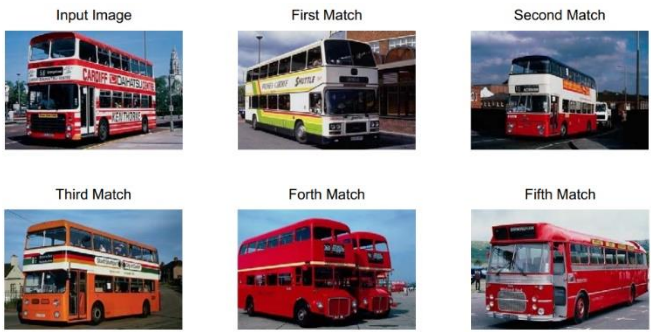


Figure 8: Sample of best matches for query image

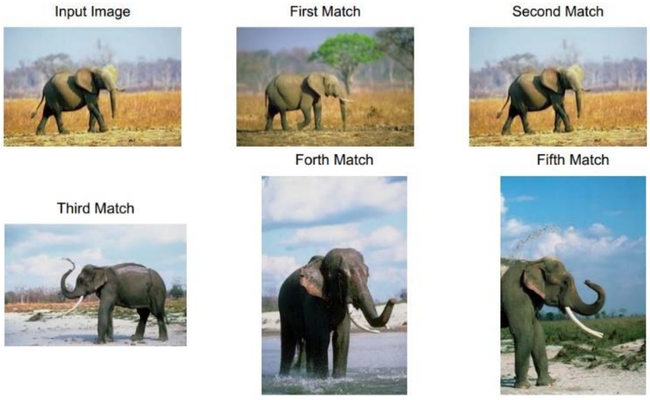


Figure 9: Sample of best matches for query image

5. Conclusion

CBIR refers to process of retrieving expected images from large image databases according to the contents of the query image. In this research paper we presented results of our research in image retrieval for visually similar images from image dataset. One of the popular visual descriptors of MPEG-7 was used for feature vector extraction. Extracted texture feature vector served as input and it will be checked against feature database. The texture descriptors of MPEG-7 are added to regions that increase the usability of our solution. We plan to study the output of other forms of MPEG-7 descriptors in the future. Future studies will also concentrate on extending the methods listed to problems of image classification for various datasets.

References

Akgül, C. B., Rubin, D. L., Napel, S., Beaulieu, C. F., Greenspan, H., & Acar, B. (2011). Content-based image retrieval in radiology: current status and future directions. *Journal of Digital Imaging*, 24(2), 208-222.

Batko, M., Falchi, F., Lucchese, C., Novak, D., Perego, R., Rabitti, F., ... & Zezula, P. (2010). Building a web-scale image similarity search system. *Multimedia Tools and Applications*, 47(3), 599-629.

Chow, T. W., Rahman, M. K. M., & Wu, S. (2006). Content-based image retrieval by using tree-structured features and multi-layer self-organizing map. *Pattern Analysis and Applications*, 9(1), 1-20.

Cvetković, S. S., & Nikolić, S. V. (2011). Merged MPEG-7 visual descriptors for image classification. In 2011 10th International Conference on Telecommunication in Modern Satellite Cable and Broadcasting Services (TELSIKS) (Vol. 1, pp. 345-348). IEEE.

Eidenberger, H. (2004). Statistical analysis of content-based MPEG-7 descriptors for image retrieval. *Multimedia Systems*, 10(2), 84-97.

Huang, J., Kumar, S. R., Mitra, M., Zhu, W. J., & Zabih, R. (1997, June). Image indexing using color correlograms. In Proceedings of IEEE computer society conference on Computer Vision and Pattern Recognition (pp. 762-768). IEEE.

Huang, P. W., & Dai, S. K. (2003). Image retrieval by texture similarity. *Pattern Recognition*, 36(3), 665-679.

Huang, Z. C., Chan, P. P., Ng, W. W., & Yeung, D. S. (2010). Content-based image retrieval using color moment and Gabor texture feature. In 2010 International conference on machine learning and cybernetics (Vol. 2, pp. 719-724). IEEE.

Manjunath, B. S., Ohm, J. R., Vasudevan, V. V., & Yamada, A. (2001). Color and texture descriptors. *IEEE Transactions on Circuits and Systems for Video Technology*, 11(6), 703-715.

Pass, G., Zabih, R., & Miller, J. (1997). Comparing images using color coherence vectors. In Proceedings of the fourth ACM international conference on Multimedia (pp. 65-73).

Sim, D. G., Kim, H. K., & Park, R. H. (2004). Invariant texture retrieval using modified Zernike moments. *Image and Vision Computing*, 22(4), 331-342.

Wang, J. Z., Li, J., & Wiederhold, G. (2001). SIMPLIcity: Semantics-sensitive integrated matching for picture libraries. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(9), 947-963.