CONTENT BASED IMAGE RETRIEVAL USING CONVOLUTIONAL NEURAL NETWORK AND EXTREME LEARNING MACHINE IN COREL DATASET

AntoAMicheal[1*], N Vadivelan[2], KBhargavi[3]

[1*,2,3]Department of Computer Science and Engineering, Teegala Krishna Reddy Engineering College, Hyderabad, Telangana, India

Abstract

The evolution of multimedia technology and rapidly increasing image collections on the Internet has attracted significant research efforts in image retrieval. Difficulties faced by text-based image retrieval motivated the researchers to develop new solutions for representation and indexing of visual information. This paper proposes a content-based image retrieval using the significant use of Convolutional Neural Network (CNN) and Extreme Learning Machine (ELM) This proposed approach extracts various features and forms as feature vectors. Apart from these extracted features, CNN is used to extract the additional features and the ELM classifies the intermediate results. The proposed approach is experimented on COREL dataset and its performance is calculated using statistical parameters such as, the precision and recall. The statistical results show that the accuracy of the proposed system is 93.58%. The experiments result shows that the proposed method outperforms the existing methods by exhibiting significant performance improvement in terms of accuracy and efficiency.

Keywords: **Content Based Image Retrieval, Corel Dataset, Convolutional Neural Network (CNN), Extreme Learning Machine (ELM), Hybrid Classification Structure.**
DOI Number:10.14704/nq.2022.20.8.NQ44408 **NeuroQuantology 2022; 20(8): 3782-3792**

Introduction

Nowadays images are broadly utilized because of its visual representation advantage. Due to the rapid advancement of computers and networks, the transmission and storage capacity of ample number of images have become possible. In older days, the image retrieval was widely required instead of text retrieval. Content-Based Image Retrieval (CBIR) is a standout amongst the best methods for getting into visual information [1]. CBIR deals with image content, such as colour, shape and structure instead of annotated text. In order to implement CBIR, the framework needs to

comprehend and interpret the content of stored images. The retrieval index should be created automatically, which provides a more visual retrieval interface to users. The fundamental idea of CBIR is to analyse image information by low level features of an image [2] and to set up feature vectors of an image as its index. CBIR has exceptionally wide and essential applications in many areas including military affairs, medical science, education, architectural design, the justice department and agriculture, etc. The advancement of CBIR exploration was clearly summarized at high level in [2, 3]. Features are the basics for CBIR, for the whole image or locally for a small group of pixels. As per the techniques utilized for CBIR, features can be grouped into low-level features and high-level features. The most practical CBIR system depends on the colour, shape, texture and other low-level features. Some researchers aim at reducing the semantic gap between visual features and the richness of human semantics [4]. With a particular ultimate goal to derive the high-level semantic features for CBIR, object ontology [5] was used to characterize high-level concepts. Supervised or unsupervised learning methods were used to associate low-level features with the query concepts. Relevance feedback was introduced into the retrieval loop for learning user's intentions and semantic templates; it was generated to support high-level image retrieval. As there is inconsistency in comprehension, the gap between semantics in the visual data for different is difficult to eliminate.

Related Work

Research on CBIR can be partitioned into two groups on the basis of the features used to retrieve the required image. Prior approaches utilized features such as shape, texture, colour and region to retrieve the required image. The current methodologies use a distinctive combination of visual features to retrieve the required image [1, 6]. The shape descriptor gives prevalent data in image retrieval because the shape is the main source through which people can perceive objects. These shape features are retrieved by two strategies, boundary-based shape feature and region-based shape feature. The boundary-based shape feature extraction technique is based on the outer boundary of an object; while the region-based shape feature technique is based on the whole region of an image. The different techniques in view of texture features have been proposed in the literature. This includes both statistical and spectral approaches. The greater part of this strategy was not able to capture the required information. Colour is the most reliable feature; it is easier to implement and robustness to background compilation. It is not influenced by image size and its orientation.

The most generally perceived methodology for color feature extraction is histogram. Color histogram illustrates its distribution in an image and it involves low computational cost. The main disadvantage of color histogram is, it cannot completely consider spatial information and it is not exclusive. In spite of utilizing extracted information from an image, the majority of the CBIR frameworks yield imprecise outcomes. The semantic gap is defined as to relate the low-level features with the high-level user semantics. The relevance feedback method was used to over-come this semantic gap [7]. In content-based image retrieval framework, a distinctive feature of the queried image is explored in search for equivalent image features in the database [8]. Frequently, it is observed that there is a semantic gap between visual features and the semantic content of an image. Extracting more effective image features can decrease this semantic gap; this is a challenging task in CBIR research. Moreover,

different machine learning techniques are used to reduce this semantic gap. In [9] SVM was utilized to extract image features and it retrieves the querying image efficiently. The hierarchical methodology [10] retrieves an image, where two separate features are explored to extract the contents and texture of an image. The proposed technique is also assessed over noisy images.

In this paper, CNN and ELM are adopted for a faithful representation of images and extracts various complimentary information from an image. This information is used to retrieve images from the databases and evaluated using accuracy of the retrieved images.

Proposed System

The proposed methodology involves image database, query image, feature extraction and the extracted images. The image database used for the proposed approach is Corel dataset and it consists of different categories, namely like Africans, buses, dinosaurs, etc. The datasets were collected and fed as input to the feature extraction process. The pixel-based feature descriptors such as CNN and ELM are extracted and stored as a feature set. These extracted feature sets were compared to the existing feature sets and the efficiency of the proposed method calculated.

Figure. 1 shows the architecture of content-based image retrieval methodology using CNN and ELM. It can be seen from the figure that our network includes two stages, feature extraction and classification. The stage of feature extraction contains the convolutional layer, contrast normalization layer, and max pooling layer. The first convolutional layer consists of 96 filters, and the size of its feature map is 56 ×56 while its kernel size is 7 and the stride of the sliding window is 4. A single convolution layer is implemented after the two stages, and a full connection layer converts the feature maps into 1-D vectors which is beneficial to the classification. Finally, the ELM structure is combined with the designed CNN model, and this architecture is used to classify the age and gender tasks.

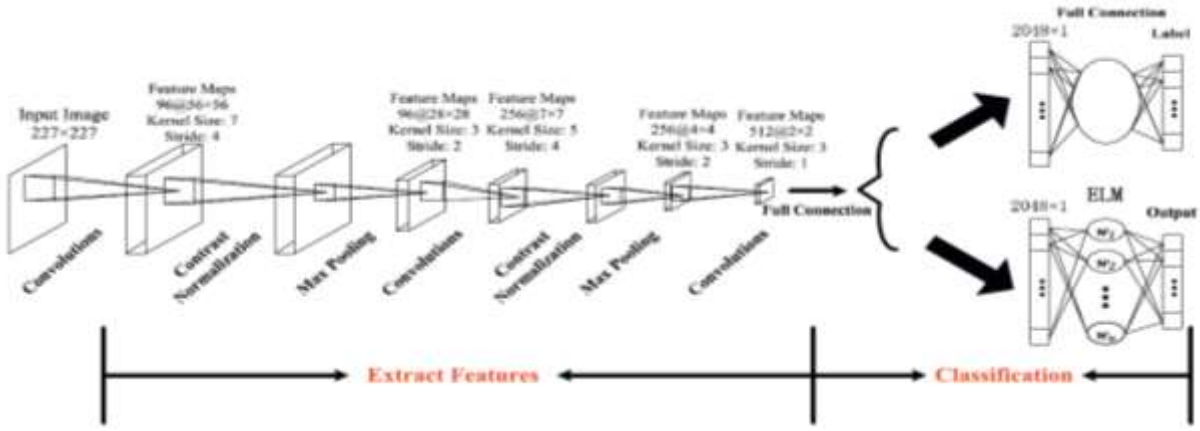

Figure 1 The architecture of Content Based Image Retrieval using CNN and ELM Model

Extreme Learning Machine Model

ELM was first proposed by Huang et al. [21 - 23, 24] which was used for the Single-Hidden-Layer Feed Forward Neural Networks (SLFNs). The input weights and hidden layer biases are randomly assigned at first, and then the training

datasets to determine the output weights that are combined. The basic structure of ELM is shown in Figure 2.

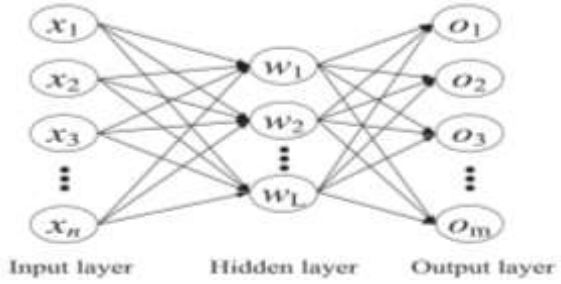

Figure.2. The Structure of ELM

ELM is not only widely used to process binary classification [25–28], but also used for multi-classification due to its good properties. As CNNs show excellent performance on extracting feature from the input images, which can reflect the important character attributes of the input images. Therefore, we can integrate the advantages of CNNs and ELM based on the analysis above, which means CNNs extract features from the input images while ELM classify the input feature vectors.

Convolutional Layer

In the convolutional layer, convolutions are performed between the previous layer and a series of filters, extract features from the input feature maps [29, 30]. After that, the outputs of the convolutions will add an additive bias and an element-wise non- linear activation function is applied on the front results. ReLU function is used as the nonlinear function in the experiment.

In general, η_{ij}^{mn} denotes the value of a unit at position *(m, n)* in the j^{th} feature map in the i^{th} layer and it can be expressed as Eq. (1):

$$\eta_{ij}^{mn} = \sigma\left(b_{ij} + \sum_{\delta} \sum_{p=0}^{p_i-1} \sum_{q=0}^{q_i-1} w_{ij\delta}^{pq} \eta_{(i-1)\delta}^{(m+p)(n+q)} \right) \quad (1)$$

Where b_{ij} represents the bias of this feature map while δ indexes over the set of the feature maps in the *(i −1)* th layer which are connected to this convolutional layer. $w_{ij\delta}^{pq}$ denotes the value at the position (p,q) of the kernel which is connected to the k^{th} feature map and the height and width of the filter kernel are P_i and Q_i.

Contrast Normalization Layer

The goal of the local contrast normalization layer is not only to enhance the local competitions between one neuron and its neighbours, but also to force features of different feature maps in the same spatial location to be computed [30]. In order to achieve the target, two normalization operations, i.e., subtractive and divisive, are performed.

Max Pooling Layer

The purpose of pooling strategy is to transform the joint feature representation into a novel, more useful one which keeps crucial information while discards irrelevant details. Each feature map in the sub sampling layer is getting by max pooling operations which are carried out on the corresponding feature map in convolutional layers [29]. Eq. (2) is the value of a unit at position *(m,n)* in the j^{th} feature map in the i^{th} layer or sub sampling layer after max pooling operation:

$$\eta_{ij}^{mn} = \max\left\{\eta_{(i-1)j}^{mn}, \eta_{(i-1)}^{(m+1)(n+1)} j, \ldots, \eta_{(i-1)j}^{(m+P_i)(n+Q_i)}\right\} \quad (2)$$

The max pooling operation generates position invariance over larger local regions and down samples the input feature maps. In this time, the numbers of feature maps in the sub sampling layer are 96 while the size of the filter is 3 and the stride of the sliding window is 2. The aim of max pooling action is to detect the maximum response of the generated feature maps while reduces the resolution of the feature map. Moreover, the pooling operation also offers built-in invariance to small shifts and distortions.

ELM Classification Layer

After the convolution and sub sampling operations, ELM is used to classify the 1-D vectors which are converted from feature maps. The ELM updates the output weights while input weights and hidden-layer biases are randomly set, thus we will randomly generate the input parameters and calculate the output weights during the training stage [26]. The whole process without iteration operation improves the neural network generalization ability. Figure 3 shows the output (containing 2048 ×1 dimensionality) of full-connection layer is the input of ELM while the numbers of hidden nodes are variables. The connection between ELM and convolutional network is a critical process and we can see from Fig. 3 that our input of ELM is the output of the full connection layer whose preceding layer is a convolutional layer. Forward-propagation and back-propagation operations are the principal parts in the architecture.

Process of our CNN+ELM

The steps are summarized as follows:

Step 1: Tune the parameters of CNN during the training stage when the connection between convolutional layers and output labels is full connection layers.

Step 2: Compute the hidden layer weights and cache the intermediate β matrices, meanwhile verify the accuracy of fine- tuned network.

Step 3: Stop the training process and calculate the average of β.

Step 4: Classify the unknown dataset using the architecture.

In order to fine tune the network, the structure is trained for more than 10K iterations. This process is performed to tune the parameters of CNN and makes it own the ability of extracting discriminative features.

Training Stage using Hybrid Structure

The training stage not only tunes the parameters of convolutional layer, but also achieves the corresponding hidden layer weights of ELM. The feed-forward process of the architecture is as same as a plain CNN for every 1000 iterations, ELM layers, instead of full connection layers, will be invoked and corresponding hidden layer weights are calculated [29]. At the same time, intermediate results β matrices are stored in the memory for final average results. When ELM classifier works and the whole iterations continue, the system will adopt stochastic- tic gradient descent to tune the relevant parameters of the entire convolutional networks. During process of back propagation, the operations between convolutional layer and sub sampling layer or sub sampling layer and convolutional layer are

eISSN1303-5150

as same as a single convolution neural network. After that, the local gradient is computed in the full connection layer. Compared with a plain CNN, the proposed architecture transforms the feature maps into 1-D vectors in the process of forward propagation, so it is just needed to transform the local gradient in the input layer of ELM to convolutional layer.

Evaluation Metrics

The performance of the proposed CBIR framework is measured in three aspects, namely, efficiency, effectiveness and computational complexity. The effectiveness of a framework is related to the retrieval accuracy of the framework and is measured using the equation (5):

$$Accuracy = \frac{R_i}{T_i} \quad - \quad (5)$$

Where *Ri* is the number of relevant retrieved images, *T* is the total number of relevant images in the image database, and *Ti* is the total number of all retrieved images. The proposed system's effectiveness is measured in terms of Average Recognition Rate (ARR). This is defined as the percentage of retrieved images in top matches that belongs to the same class as a query image.

The image database used for the proposed approach is Corel dataset and it consists of different categories, namely like Africans, buses, dinosaurs, etc. The datasets

Results and Discussions

The tests are carried out on COREL photo database. The COREL database contains more than 5000 pictures organized in categories. Each category has about 100 images. There are 50 categories and the corresponding database is composed of 5000 images. The results presented here are five categories directly extracted from the 50 categories. Figure 3 shows that the images are taken from different categories in Corel database, such as African, buses, dinosaur, monuments and beaches, etc. The dinosaur image retrieved using the proposed methodology is shown in Figure 4. Table 1 shows the comparison of proposed method 's precision with the other existing methods. The accuracy of the proposed image retrieval method achieves on average 80.06%, wherein Lin et al [12] achieves 73.3%, Elalami [13] achieves 73.96%, Poursistani et al. [14] achieves 72.34%, Guo et al. [15] achieves 76.5%, Subra et al [16] achieves 75.26, Walia et al. [17] achieves 66.2%, Irtaza et al. [18] achieves 73.00%, Elalami [19] achieves 75.8% and Zeng et al. [20] achieves 79.5%. The statistical results show that the accuracy of the proposed method is more than the existing methods, which implies that the proposed method retrieves more relevant images than the existing methods. The comparison of accuracy between the proposed and the other existing methods is shown in Figure 5.

Table 1. Comparison of proposed method precision with other method's precision metric

Method/ Images	Lin et al. [12]	Elalami [13]	Poursistani et al. [14]	Guo et al. [15]	Subra et al. [16]	Walia et al. [17]	Irtaza et al. [18]	Elalami [19]	Zeng et al. [20]	Anto et al. [31]	**Proposed**
Africa	68.30	70.30	70.20	84.70	69.75	51.00	65.00	72.60	72.50	73.30	**75.23**
Buses	88.80	87.60	76.30	85.30	89.65	78.00	85.00	89.10	89.20	90.05	**92.59**
Dinosaurs	99.25	98.70	100.00	99.30	98.70	100.00	93.00	99.30	100.00	99.35	**99.68**
Monuments	56.15	57.10	70.80	67.80	63.95	58.00	62.00	58.70	70.60	73.25	**78.64**
Beaches	54.00	56.10	44.40	45.40	54.25	44.00	60.00	59.30	65.20	64.35	**68.89**

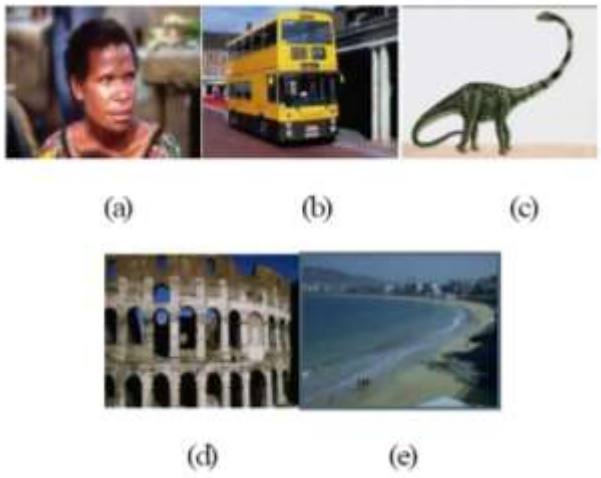

Figure 3 Image Datasets used from Corel database a) African dataset b) Buses c) Dinosaur d) Monuments and e) Beaches

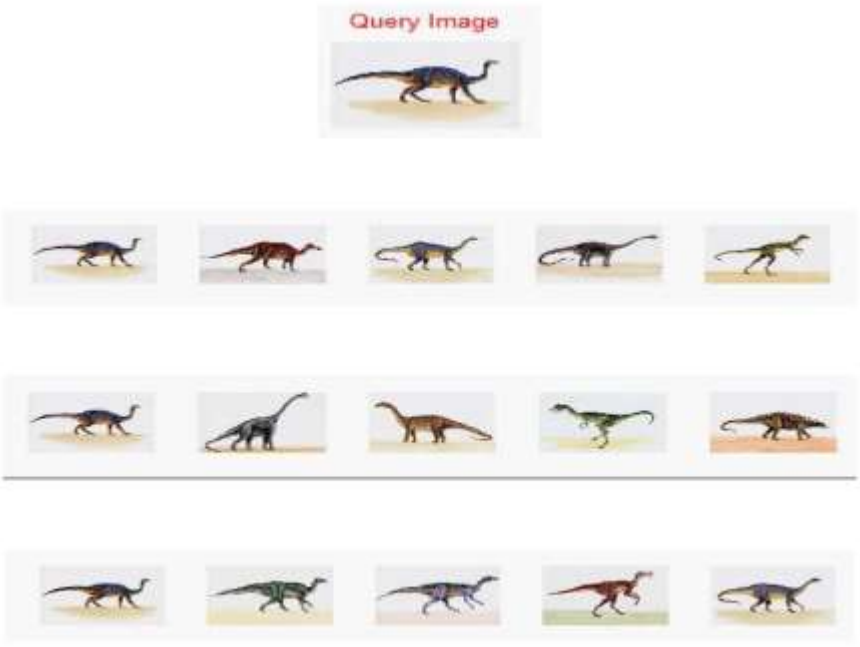

Figure 4 Results of proposed Content Based Image Retrieval on Dinosaur dataset using CNN and ELM

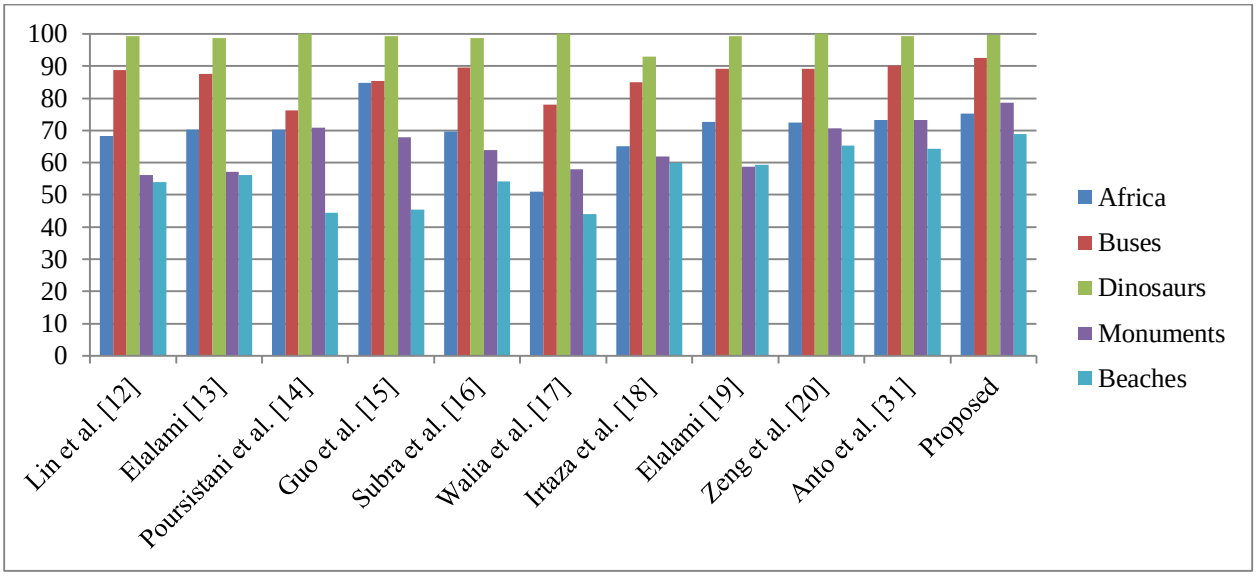

Figure 5 Comparison of proposed approach accuracy with the other method's accuracy for different semantic classes in the COREL database.

The observation from statistics shows that accuracy of the proposed method is higher than the existing methods and this indicates that the proposed method retrieves more relevant images from the image database.

Conclusion

This study proposes a content-based image retrieval using CNN and ELM feature descriptors. There is a significant difference between the proposed image retrieval performance and other existing method's performance. The study is experimented on COREL dataset and its retrieval performance is evaluated using accuracy. This research discloses the effectiveness of ELM feature descriptors and suggests that the proposed model can be useful in retrieving relevant images. The statistical results show that the accuracy of the proposed approach is 80.06%. The comparative study shows that the proposed content-based image retrieval method proves its feature extraction capability than the other image retrieval methods. The future enhancement will be a design of improving the system and utilizing the same to predict crime prevention, medical diagnosis, intellectual property and textile industry.

Conflicts of Interest

The authors have no conflicts of interest to declare.

References

[1] Muneesawang P, Guan L. An Interactive Approach for CBIR Using a Network of Radial Basis Functions, IEEE Transactions on Multimedia, 2004; 6 (5): 703–716.

[2] Datta R, Joshi D, Li D, Wang JZ. Image Retrieval: Ideas, Influences, and Trends of the New Age, ACM Computing Surveys, 2008; 40 (2):1-60.

[3] Smeulders AW, Worring M, Santini S, Gupta A, Jain R. Content-Based Image Retrieval at the End of the Early Years, IEEE Transactions on Pattern Analysis and Machine Intelligence, 2000; 22(12):1349-1380.

[4] Lu Y, Hu C, Zhu X, Zhang H, Yang Q. A Unified Framework for Semantics and Feature Based Relevance Feedback in Image Retrieval Systems, Proc. Eighth ACM international conference on Multimedia, 2000: 31-37.

[5] Mezaris V, Kompatsiaris I, Strintzis MG. An Ontology Approach to Object-Based Image Retrieval, *Proc. International Conference on Image Processing*, 2003, 2:511-514.

[6] Chun YD, Kim NC, Jang IH. Content- Based Image Retrieval Using Multi-resolution Color and Texture Features, IEEE Transactions on Multimedia, 2008; 10(6): 1073-1084.

[7] Bugatti PH, Traina C, Traina AJM. Improving Content-Based Retrieval of Medical Images through Dynamic Distance with Relevance Feedback, *Proc. 24th International Symposium on Computer-Based Medical Systems (CBMS)*, 2011: 1–6.

*e*ISSN1303-5150

[8] Iqbal K, Odetayo MO, James A. Content-Based Image Retrieval Approach for Biometric Security using Colour, Texture and Shape Features Controlled by Fuzzy Heuristics, Journal of Computer and System Sciences, 2012; 78(4): 1258-1277.

[9] Royal M, Chang R, Qi X. Learning from Relevance Feedback Sessions Using a k Nearest-Neighbor-Based Semantic Repository, *Proc. IEEE International Conference on Multimedia and Expo (ICME07)*, 2007: 1994–1997.

[10] Rao Y, Mundur P, Yesha Y. Fuzzy SVM Ensembles for Relevance Feedback in Image Retrieval, LNCS, 2006; 4071: 350- 359.

[11] Huang J, Kumar SR, Mitra M, Zhu W, Zabih R. Image Indexing Using Color Correlograms, *Proc. Conference on Computer Vision and Pattern Recognition*, 1997: 762– 768.

[12] Lin CH, Chen RT, Chan YK. A Smart Content-Based Image Retrieval System Based on Color and Texture Feature, Image and Vision Computing, 2009; 27(6): 658-665.

[13] Elalami ME. A Novel Image Retrieval Model Based on the Most Relevant Features, Knowledge-Based Systems, 2011;24(1): 23–32.

[14] Poursistani P, Nezamabadi-pour H, Moghadam RA, Saeed M. Image Indexing and Retrieval on JPEG Compressed Domain Based on Vector Quantization, Mathematical and Computer Modelling, 2013; 57(5–6):1005–1017.

[15] Guo JM., Prasetyo H, Su HS. Image Indexing using the Color and Bit Pattern Feature Fusion, Journal of Visual International Journal Of Current Engineering And Scientific Research (IJCESR), 2017; 4(8): 1360–1379.

[16] Subrahmanyam M, Wu QMJ, Maheshwari RP, Balasubramanian R. Modified Color Motif Co-Occurrence Matrix for Image Indexing and Retrieval, Computers and Electrical Engineering, 2013; 39(3): 762- 774.

[17] Walia E, Pal A. Fusion Framework for Effective Color Image Retrieval, Journal of Visual Communication and Image Representation, 2014; 25(6): 1335–1348.

[18] Irtaza A, Jaffar MA, Aleisa E, Choi TS. Embedding Neural Networks for Semantic Association in Content Based Image Retrieval, Multimedia Tools and Applications, 2014; 72(2): 1911–1931.

[19] ElAlami ME. A New Matching Strategy for Content Based Image Retrieval System, Applied Soft Computing, 2014; 14: 407–418.

[20] Zeng S, Huang R, Wang H, Kang Z. Image Retrieval Using Spatiograms of Colors Quantized by Gaussian Mixture Models', Neurocomputing, 2016; 171(1): 673–684.

[21] G.B. Huang, Q.Y. Zhu, C.K. Siew, Extreme learning machine: theory and applications, Neurocomputing,2006; 70(1–3): 489–501.

[22] F.S. Khan, J. van de Weijer, R.M. Anwer, M. Felsberg, C. Gatta, Semantic pyramids for

*e*ISSN1303-5150

gender and action recognition, IEEE Transactions on Image Processing, 2014; 23(8): 3633–3645
doi: 10.1109/TIP.2014.2331759 .

[23] H. Guang-Bin, C. Lei, S. Chee-Kheong, Universal approximation using incremental constructive feedforward networks with random hidden nodes., IEEE Transactions on Neural Networks, 2006; 17(4): 879–892.

[24] M. Duan, K. Li, X. Liao, K. Li, A parallel multiclassification algorithm for big data using an extreme learning machine, IEEE Transactions on Neural Networks Learning System, 2017; 99:1–15 doi: 10.1109/TNNLS.2017.2654357.

[25] B. Zuo , G.B. Huang , D. Wang , W. Han , M.B. Westover , Sparse extreme learning machine for classification, IEEE Trans. Cybern, 2014; 44(10): 1858–1870.

[26] G.-B. Huang, H. Zhou, X. Ding, R. Zhang, Extreme learning machine for regression and multiclass classification, IEEE Trans. Syst. Man Cybern. Part B Cybern, 2012; 42(2): 513–529
doi: 10.1109/TSMCB.2011.2168604.

[27] Y. Yang , Q.M. Wu , Y. Wang , K.M. Zeeshan , X. Lin , X. Yuan , Data partition learning with multiple extreme learning machines, IEEE Trans. Cybern, 2014; 45(6): 1463–1475.

[28] J. Luo , C.M. Vong , P.K. Wong , Sparse Bayesian extreme learning machine for multi-classification., IEEE Trans. Neural Netw. Learn. Syst, 2014; 25(4): 836–843.

[29] J. Shuiwang, Y. Ming, Y. Kai , 3d convolutional neural networks for human action recognition, IEEE Trans. Pattern Anal. Mach. Intell.2013; 35(1): 221–231.

[30] Z. Dong, Y. Wu, M. Pei, Y. Jia , Vehicle type classification using a semi-supervised convolutional neural network, IEEE Trans. Intell. Transp. Syst, 2015; 16(4): 1–10.

[31]Anto.A.Micheal, Pradeepthi.K.V, Quantitative Analysis of Content Based Image Retrieval Using HOG, LBP and Gabor Feature Descriptors in Corel Dataset, International Journal of Current Engineering and Scientific Research,2017; 4(8): 57-64.