Journal of Theoretical and Applied Information Technology
15th September 2022. Vol.100. No 17

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

BAG OF VISUAL WORDS AND CNN APPROACHES FOR CONTENT-BASED IMAGE RETRIEVAL USING HOG, GCH AND ORB FEATURES

K. BHARATHI[1*, 3], Dr. M CHANDRA MOHAN[2]

[1]Associate Professor, Dept. Of MCA, Loyola Academy, Secunderabad, India
[2]Professor of CSE, Jawaharlal Nehru Technology University Hyderabad, India
[3]Research Scholar, JNTU, Hyderabad, India
*kalva.bharathi@gmail.com

ABSTRACT

Component measurements in digital photos are expanding, and identifying a specific image based on substance from a large database might be difficult at times. A content-based image retrieval (CBIR) method is suggested in this work to extract a feature vector from an image and successfully retrieve content-based pictures. This paper considers three types of image feature descriptor extraction methods: Histogram Oriented Gradients (HOG), Global Colour Histogram (GCH), and Oriented Fast and Rotated BRIEF (ORB). The image feature vectors are kept in the picture database and matched with the testing data feature vector for CBIR at the time of retrieval. In this paper, we aim to present the feature selection based on HOG, GCH and ORB methods to extract features perfectly in capturing the standard dataset CIFAR10 features. The suggested work's execution is evaluated using a Bag of Visual Words, and CNN classifiers. The proposed strategy experiments for different labels indexed elastic search procedures and all cases showed good accuracy in retrieving the correct image.

Keywords: *Information Retrieval, Query Image, Bag Of Visual Words, Elastic Search Engine, Feature Descriptors*

1. INTRODUCTION

The shared and stored multimedia data are growing, and searching or retrieving a relevant image from an archive is a challenging research problem [21–23]. Most of the search engines on the Internet retrieve the images on the basis of text-based approaches that require captions as input [24–26]. The user submits a query by entering some text or keywords that are matched with the keywords that are placed in the archive. The output is generated on the basis of matching keywords, and this process can retrieve the images that are not relevant. The difference in human visual perception and manual labelling/annotation is the main reason for generating an output that is irrelevant [27–30]. It is near impossible to apply the concept of manual labelling to existing large-size image archives that contain millions of images. The second approach for image retrieval and analysis is to apply an automatic image annotation system that can label images on the basis of image contents. The approaches based on automatic image annotation are dependent on how accurate a system is in detecting color, edges, texture, spatial layout, and shape-related information [31–33]. Significant research is being performed in this area to enhance the performance of automatic image annotation, but the difference in visual perception can mislead the retrieval process.

Basically, the similarity between the query image and the database images is used to rank the database images in decreasing order of similarity [34]. Thus, the performance of any image retrieval method depends upon the similarity computation between images. Ideally, the similarity score computation method between two images should be discriminative, robust and efficient.

We frequently utilize search engines. When we have a question, we may utilize a search engine like Google to get the best answer. The majority of query formats are text-based. However, most of the time, the text is fairly helpful in locating pertinent answers. For example, suppose you want to look for a product on the internet, like a t-shirt, but you don't know what it's called. How did you track them down? You can, however, provide a description of the clothing. The issue with utilizing descriptions is that you will receive a wide range of items. What's more, they won't be identical to the goods you're looking for, so you'll need a better technique to find them. To solve it, we may utilize the product's picture to extract its characteristics, which we can then use to find comparable items. This is referred to as content-based image retrieval. A method for obtaining suitable pictures based on a given image is known as content-based image retrieval (CBIR). An image query and an image database comprise the system. Using an action methodology, the system will begin by extracting features from all photos, whether from the query or the image database. The system will then compute similarities between the query and all photos in the database. Finally, the system will obtain any photos that are strikingly similar to the query [34].

The rest of the paper is organized as: in section2 represents the literature review and explained various methods proposed in this research area. The architecture of CBIR is described in section3. Section4 proposed the model architecture for searching a query-based image retrieval and followed by a discussion of the results. Section5 draws the conclusions of the proposed model.

2. LITERATURE REVIEW

The low-level feature extraction methods on colour, text and shape were used by most of the researchers in content-based image retrieval systems [1, 2, 3, 4, 5]. HSV colour space and shape features were used by Chaudhari and Patil to retrieve images [1]. Local binary pattern and colour moment algorithms for feature selection were proposed by Chaudhari et al. [2]. The improved query results are based on the relevance feedback system used with colour and texture explained in Janani and Palanippan [3]. The Euclidian and Manhattan distance measures are used to calculate the similarity index on texture and colour, as described in Khosla et.al. [4]. The wavelet transformation moments were proposed by Srivastava et al. [5]. To enhance image classification performance, the proposed technique integrates both global image characteristics and local region-based features [6].There are quite a few pre-trained model designs available, including Densenet [7], Resnet [8], Inception-V3 [9], MobileNet [10] and VGGnet [11].Betul AY et al. proposed a generative adversarial network for CBIR for e-commerce systems [12]. H Kavitha and M V Sudhamani [13] [14] described the edge detection and Horris corner methods used to extract images; also proposed HOG and Edge methods used for feature selection for image retrieval. MeghaAgarwal [15] used HOG and Vocabulary Tree methods for image retrieval.AnnaSaroVijendran and S. Vinod Kumar used HOG with wavelet sub-bands relevant feedback systems used for CBIR [16].The cosine distance was used in conjunction with GCH and the 2-D Haar wavelet transform to provide satisfactory resultsfor CBIR mentioned in the M.A. Tahoun et al., [17]. M.A. Tahoun et al., [18] proposed the multiple feature selection based on colour histogram for image extraction. M. Salmi and B. Boucheham [19] used HOG based Colour Coherence Vector methods to extract the query images. Chhabra. P [20] proposed ORB and SIFT features for identifying the images from the public datasets image retrievals.

3. ARCHITECTURE FOR CBIR

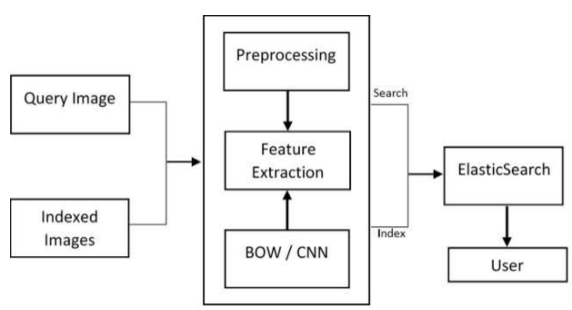

Figure1: Cbir System Architecture

The information retrieval system is also called a search engine created in Python using the Elasticsearch service. The image document and

Journal of Theoretical and Applied Information Technology
15th September 2022. Vol.100. No 17

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

queries consist of the fields like id, filename, absolute path to file, and features are found by dense vector through computer vision model. The image documents are retrieved and ranked using a feature vector. To accomplish this, the document is compared the image-query feature vector with the image-document feature vector using cosine similarity. And a value of 1.0 is added to the cosine similarity value to get a positive score. We have configured the Elasticsearch client on the localhost machine and created one node cluster with no replicas for the indexes. Here, each index is distributed in 30 shards, each shard is an occurrence of a Lucene index, that indexes and processes query for a subclass of the information in an Elasticsearch cluster. So that it is possible to gain good performance in indexing and search operations, and also to handle large amounts of data.

The mapping for the image documents that the indexes can handle, consists of content-based of documents and queries that we have described before. The only field that can be used for retrieval is "features" which is of dense vector type and the dimensions are dependent CV method. However, the size of the dense vector is always smaller than 2048 since this is the maximum size Elasticsearch can handle. To evaluate our IR system, we have used the trec_eval tool and its metrics. Specifically, the search engine is evaluated on the top 100 retrieved image documents, and we consider the mean average precision (MAP).

The CBIR system architecture is depicted in the above figure1. Generally, the content-based image retrieval system consists of two phases data insertion and data querying. In the first phase, the features are extracted and stored in the image database manually. Where in the second phase of query processing, the user inputs the query in the form of an image pattern by applying a metric like Euclidian distance to check the similarity between the query image and image database. It ranks in the image databases images based on the similarity index, most similar image is retrieved.

3.1. Bag of Visual Words Model Architecture

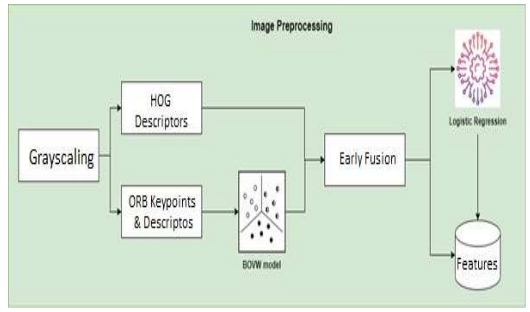

Figure2: Model Architecture

The Bag of Visual words is taken from the bag of words of NLP. In the Bag of Words model, we count the number of times each word appears in a text, utilize the frequency of each word to determine the document's keywords (features) and create a frequency histogram from it. We consider a document to be a collection of words.

1. The Bag of Visual Words (BOVW) represents a picture as a collection of characteristics. Key points and descriptions are examples of features.
   a. Key points are unique in that they remain constant no matter what modification we apply to a picture (rotate, reduce, expand, etc.).
   b. A descriptor is a summary of the main point.
2. We generate visual language by using key points and descriptors, and then we quantize the picture attributes.

By doing so, we were able to properly express pictures as a frequency histogram of visual characteristics. Later, we can accomplish numerous tasks, including categorization, retrieval, and more, with the help of visual vocabulary.

The following 4 figures from figure3.1 to 3.4 explains the visual descriptors of the Bag-of-visual-words.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

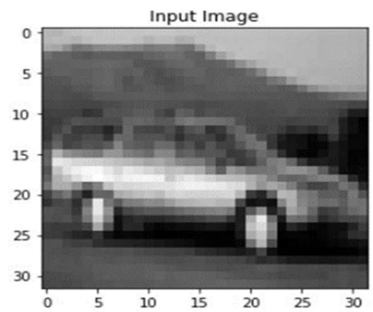

Figure3.1: Input image

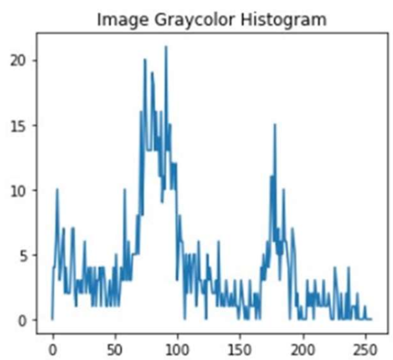

Figure3.2: Grayscale histogram

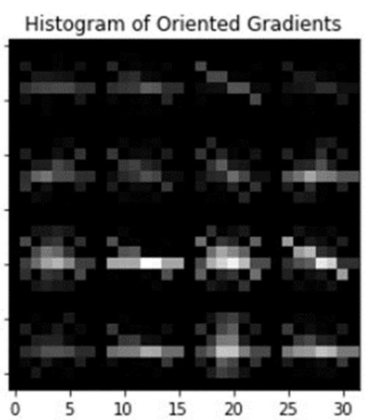

Figure3.3: Gradients

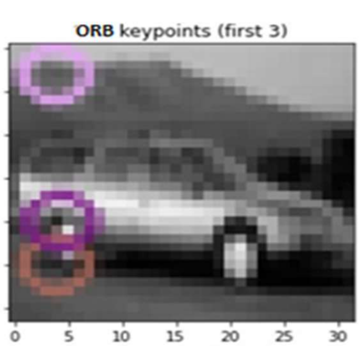

Figure3.4: Descriptors

We'll just generate visual vocabularies using the ORB key points and descriptors because GCH and HOG features are already in histogram form. Clustering models are used to build visual vocabularies. The centroid of each cluster is therefore regarded as a visual word, and the visual vocabulary is comprised of all centroids. The visual vocabulary will be "universal" if the training set is sufficiently representative. Finally, a vector quantize takes a feature vector and maps it to the index of the visual words in the visual lexicon that is closest to it.To identify the visual vocabulary, we used mini-batch k-means (with a batch size of 64 photos). In addition, because the CIFAR-10 pictures are tiny, we evaluated vocabulary sizes ranging from 100 to 1000 visual words (32x32 pixels). Here, the visual words may not be considered when the size of the vocabulary is too small and if it is too large then it becomes overfitting for training.

We'll use the Davies-Bouldin Index measure to assess our Bag of Visual Words model.The score is defined as the average similarity measure of each cluster to its most comparable cluster, where similarity is defined as the ratio of within-cluster to between-cluster distances. As a result, clusters that are further apart and less scattered will score higher. Lower numbers suggest more effective grouping.According to the findings, a vocabulary of 400 visual words is adequate.

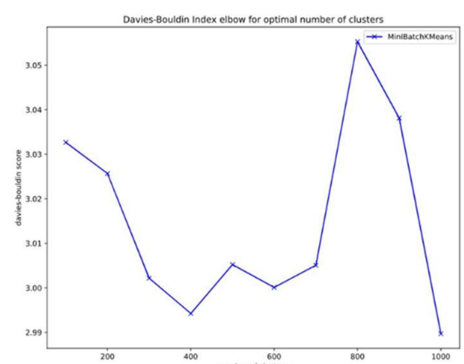

Figure4: Number of Clusters

Having extracted the visual vocabulary and quantized the vectors of ORB descriptors, we are ready to decide what features are we going to feed our classification models and the search engine.We have used early fusion to combine the GCH descriptors, HOG descriptors and ORB descriptors after quantization before we feed them to our classification model. The combinations are the following:

- HOG and GCH descriptors
- HOG and ORB descriptors
- HOG, GCH and ORB descriptors

To test only combinations that include HOG descriptors because the Histogram of Oriented Gradients is a good feature for machine learning applications.The models we have tested are the Gaussian Naïve Bayes,Logistic Regression,k Nearest Neighbors,Linear Support Vector Machines, and Random Forest.Moreover, we have tuned their hyperparameters exhaustively using grid-search and 5-fold cross-validation. The primary metric used to access our models is accuracy since CIFAR-10 is a well-balanced dataset. Also, we have considered precision, since the models will be used for the image retrieval task.

Table1: Machine Learning models comparison

Accuracy	HOG + GCH		HOG + ORB		HOG + GCH + ORB	
	Train	Test	Train	Test	Train	Test
Naïve Bayes	40	39	42	38	43	39
Logistic Regression	57	54	62	59	62	56
kNN	98	27	99	29	99	26
Random Forest	100	55	100	56	100	54
Linear SVM	55	48	56	54	62	57

Table2: Machine Learning models comparison

Precision	HOG + GCH		HOG + ORB		HOG + GCH + ORB	
	Train	Test	Train	Test	Train	Test
Naïve Bayes	37	38	39	38	41	40
Logistic Regression	53	56	67	60	62	55
kNN	99	23	100	23	99	28
Random Forest	98	58	99	52	94	59
Linear SVM	55	50	59	57	63	55

As the results are shown in table1 and table2, Logistic Regression utilizing HOG and ORB features is the top model since it accomplishes the highest accuracy and precision score and simultaneously doesn't overfit.

3.2. Visual Embedding Model Architecture

- The feature extraction is performed by the deep learning models. Deep neural networks learn high-level characteristics in the hidden layers. The picture is first processed using a series of convolutional layers. The network learns expanding and relatively complicated characteristics in its layers through convolutional layers. The modified picture information is then sent through the fully linked layers and converted into a classification or prediction.
- The extracted features from the deep learning models we have utilized are the visual embedding. The bundle of attributes from the final fully connected layer (before to a loss layer) attached to a CNN is known as visual embedding. The dimension of the embedding vector space is determined by the size of the completely linked layer.
- The visual embedding is learned by jointly training the feature extractor with the embedding layer and the classifier. We register a forward hook to our deep learning model so that we can access this vector space (features) and utilize them for retrieval later.
- Finally, since Elasticsearch can only handle vectors with a maximum size of 2048, we apply dimensionality reduction with Principal Component Analysis (PCA) and reduce the size of the visual embedding to 2000 features.

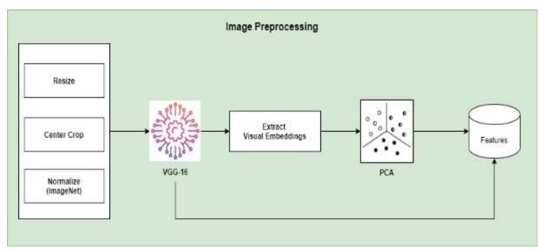

Figure5: Visual Embedding Model Architecture

3.3. Deep Learning Model Architecture

That the final feature vector for each image will be a fusion of the visual embedding and the predicted class one-hot vector. Before passing the image data to the deep learning models (batch size 64), apply some transformations: resizing, centre cropping and normalization using mean and standard deviation from ImageNet.The models we have tested are Neural Network, Convolutional

Neural Network (custom), VGG-16, Inception v1, and ResNet-50. Combine the deep learning models and the visual embedding layer for training on the classification task. Thus, we choose cross-entropy loss as the cost function. Moreover, we make use of Adam optimizer and ReduceLROnPlateau scheduler (reduces learning rate when accuracy has stopped improving). Finally, we apply additional regularization by optimizing the weight decay. Considering the pre-trained models, we have decided to use transfer learning. We believe that the features the pre-trained models have acquired, during training on ImageNet, can provide better results than training those networks from scratch. That is because the features learned in the first convolutional layers are quite generic. Since CIFAR-10 data are similar to ImageNet, we use the pre-trained models as feature extractors, meaning that we freeze all the layers of the pre-trained model and only train the classifier part. Considering hyper parameter tuning, we have used Ray tune which is a popular framework for scalable hyper parameter tuning. During tuning, we utilize early stopping with ASHA Scheduler (a better version of Hyper Band scheduler) for eliminating bad trials based on validation accuracy.

- The grace period is selected in such a way that we won’t fall into local optima.
- Not only do we tune learning rate and weight decay, but also, we tune the number of layers, nodes and dropout probability in the cases of the Neural Network and the custom CNN.

4. RESULTS AND DISCUSSION

These experiments are built with Keras, a Python-based neural network toolkit that can run on top of either TensorFlow or Theano. It is used as a personal computer and has the specifications: Intel® CoreTM i5-4310M CPU @ 2.70GHz, 4, 8 Gb RAM, and 64-bit.

CIFAR-10 is a well-known computer-vision dataset for object recognition. It is a subset of the 80 million small pictures collection, comprising 60,000 32x32 colour images having one of ten item classes, each with 6000 photos. The CIFAR-10 images data corpus was downloaded from the TensorFlow data repository. These images are categorized into airplane, automobile, bird, cat, deer, dog, frog, horse, ship and truck. All the images are preprocessed, loaded with a batch size of 32 and considered as 50000 images data for training and the remaining 10000 data for testing. The Sequential model is composed of three convolution blocks, each of which has a max-pooling layer. A ReLU activation function turns on a fully connected layer with 128 units on top. This model has not been altered in any way. From the drop-down menu, choose the Adam optimizer and the Sparse Categorical Cross entropy loss function.

Table3: Comparison Of Deep Learning Models

Deep Learning Models	Accuracy		Precision	
	Train	Test	Train	Test
NN	67%	55%	67%	55%
CNN	78%	67%	78%	67%
VGG16	89%	84%	90%	85%
Inception v1	89%	85%	89%	85%
ResNet-50	84%	81%	84%	82%

As the results are shown in the above table3, VGG-16 is the top model since it accomplishes the highest accuracy and precision score and simultaneously doesn’t overfit.

Table4: Calculations Of BOVW And VGG-16

	Accuracy		Precision		Mean Avg Precision (MAP)
	Train	Test	Train	Test	
BOVW	60%	56%	59%	56%	0.0019
VGG-16	89%	85%	89%	85%	0.0088

From the table4, according to [15], [19-20] the best model is VGG-16 with mean average precision. As it classifies the best on the CIFAR dataset. We believe that the choice of training the models in the classification setting has prevented us from gaining the best results. The cross-entropy loss's goal is to categorize features into predetermined classes, hence its performance is low when compared to losses that incorporate similarity (and dissimilarity) restrictions in the embedding space during training. The following figure6.1 explain model accuracy during training and testing per iteration whereas, figure6.2 shows loss function in training and testing.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

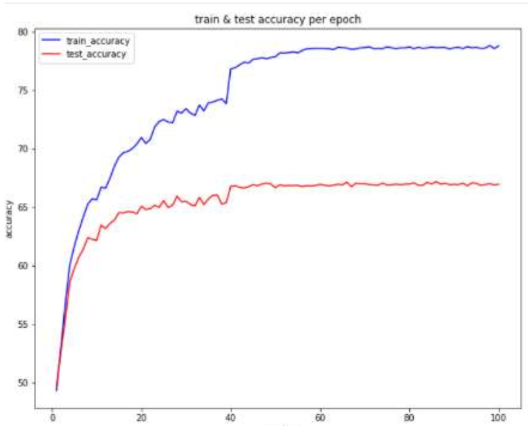

Figure6.1: Model Accuracy

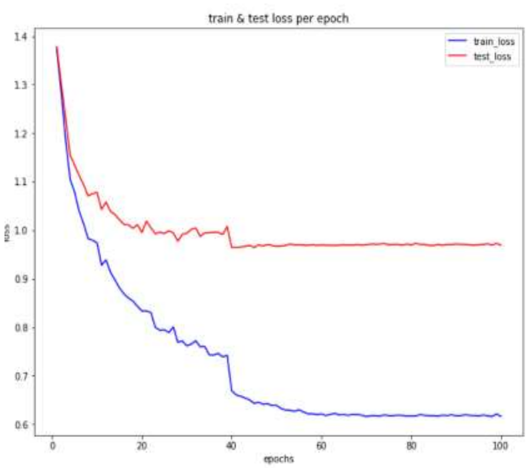

Figure6.2: Model loss

Machine Learning Vs Deep Learning

As expected, the VGG-16 model has surpassed the Bag of Visual Words model. This is because the VGG-16 model (trained on ImageNet and fine-tuned on CIFAR-10) has learned complex features that are representative of the data and thus it can distinguish better the CIFAR-10 classes and classify them more accurately images. In contrast, the BOVW model is dependent on feature engineering· the better the handcrafted features are the more predictive power the machine learning model will have. Some of the produced features in the BOVW model are local features, meaning that they contain specific information about the images they were extracted from. Because of that, only a small proportion of extracted features is representative of the data and thus our machine learning model cannot generalize.

Search engine

Similarly, as before, the IR system dependent on VGG-16 gives the best MAP results. Since VGG-16 can distinguish better the CIFAR-10 classes, our IR system will be more accurate (higher precision). So, during retrieval, there are more true positives (relevant images with respect to the query) leading to the set of retrieved documents. One could say that the sorting by the similarity function (cosine in our case) is better.

A better alternative would be to use:

- Contrastive loss, which maximizes the training aim by encouraging all comparable class instances to move infinitesimally closer to one other in the output embedding space, while driving examples from other classes to move far away.
- Triplet loss, which encourages data points from the same class to be closer to each other than to a data point from another class by taking both positive and negative pair distances into account at the same location.

If we were to choose triplet loss as our loss function, we should also consider an appropriate method for mining informative points in order to improve training convergence and computational complexity. The popular sampling approaches are batch all, batch hard, batch weighted and batch sample may be used to improve the accuracy.

5. CONCLUSION

In this paper, an image search engine system for real images is proposed and implemented to searchsimilar images to a query image from an image database. First, we segment out the main object inan image and then extract its features.In general, a classifier trained by the object-based feature vectors is more precise than that trained bythe feature vectors from a whole image. Next, relevant features are selected from the original featureset for facilitating image classification, using the BOVW and CNN with the VGG-16 model has given good accuracy and precision.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

REFERENCES

[1] R. Chaudhari and A. M. Patil, "Content-based image retrieval using colour and shape features," International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering, Vol. 1, No. 5, pp. 386-392, 2012.

[2] R. Choudhary, N. Raina, N. Chaudhary, R. Chauhan, and R. H. Goudar, "An integrated approach to content-based image retrieval," Proc. International Conference on Advances in Computing, Communications and Informatics, pp. 2404-2410, 2014.

[3] MS. R. Janani and S. K. Palaniappan, "An improved CBIR method using colour and texture properties with relevance feedback," International Journal of Innovative Research in Computer and Communication Engineering, Vol. 2, Special No. 1, pp. 47-56, 2014.

[4] G. Khosla, N. Rajpal, and J. Singh, "Evaluation of Euclidean and Manhanttan metrics in content based image retrieval system," International Journal of Engineering Research and Applications, Vol. 4, No. 9, pp. 43-49, 2014.

[5] P. Srivastava, O. Prakash, and A. Khare, "Content-based image retrieval using moments of wavelet transform," Proc. the International Conference on Control, Automation and Information Sciences, pp. 159-164, 2014.

[6] W. S. Chow and M. K. M. Rahman, "A new image classification technique using tree-structured regional features," Neurocomputing, Vol. 70, No. 4-6, pp. 1040-1050, 2007.

[7] Iandola, Forrest, et al. "Densenet: Implementing efficient convnet descriptor pyramids." arXiv preprint arXiv:1404.1869 (2014).

[8] Szegedy, Christian, et al. "Inception-v4, inception-resnet and the impact of residual connections on learning." Thirty-First AAAI Conference on Artificial Intelligence. 2017.

[9] Szegedy, Christian, et al. "Rethinking the inception architecture for computer vision." Proceedings of the IEEE conference on computer vision and pattern recognition. 2016.

[10] Sandler, Mark, et al. "Mobilenetv2: Inverted residuals and linear bottlenecks." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018.

[11] Simonyan, Karen, and Andrew Zisserman. "Very deep convolutional networks for large-scale image recognition." arXiv preprint arXiv:1409.1556 (2014).

[12] Betul AY et al., A Content-based Image Retrieval Adversarial Networks, International Conference on Deep Learning and Machine Learning in Emerging Applications (Deep-ML), pp. 44-49, 2019.DOI 10.1109/Deep-ML.2019.00017.

[13] H Kavitha and M V Sudhamani. Image Retrieval based on object recognition using the Harris corner and edge detection Technique, International Conference on Communication, VLSI & Signal Processing, pp. 181-184, 2013.

[14] H Kavitha and M V Sudhamani, Image Retrieval using Hog and EdgeFeatures, International Conference on Emerging Computation and Information Technologies, Elsevier, pp. 51-59, Nov. 2013.

[15] Agarwal, Megha, HOG Feature and Vocabulary Tree for Content Based Image Retrieval. International Journal of Signal and Imaging Systems Engineering. 3. 246, 2010. 10.1504/IJSISE.2010.038020.

[16] Anna SaroVijendran and S. Vinod Kumar, A New Content Based Image Retrieval System by HOG of Wavelet Sub Bands International Journal of Engineering Research & Technology (IJERT), Vol:3 Iss:3, 2014, pp. 1476-1480, ISSN: 2278-0181.

[17] M. A. Tahoun, K. A. Nagaty, T. I. El-Arief and M. A-Megeed, "A robust content-based image retrieval system using multiple features representations," Proceedings IEEE Networking, Sensing and Control, 2005, pp. 116-122, doi: 10.1109/ICNSC.2005.1461172.

[18] M. A. Tahounet al., Robust Content-Based Image Retrieval System using Multiple

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Features Representations, Proceedings IEEE, 2005,0-7803-8812-7/05.

[19] M. Salmi and B. Boucheham, Content based image retrieval based on Cell Color Coherence Vector (Cell-CCV), International Symposium ISKO-Maghreb: Concepts and Tools for knowledge Management (ISKO-Maghreb), 2014, pp. 1-5, doi: 10.1109/ISKO-Maghreb.2014.7033451.

[20] Chhabra, P., Garg, N.K. & Kumar, M. Content-based image retrieval system using ORB and SIFT features. Neural Comput&Applic 32, 2725–2733 (2020). https://doi.org/10.1007/s00521-018-3677-9.

[21] D. Zhang, M. M. Islam, and G. Lu, "A review on automatic image annotation techniques," Pattern Recognition, vol. 45, no. 1, pp. 346–362, 2012.

[22] Y. Liu, D. Zhang, G. Lu, and W.-Y. Ma, "A survey of content-based image retrieval with high-level semantics," Pattern Recognition, vol. 40, no. 1, pp. 262–282, 2007.

[23] T. Khalil, M. U. Akram, H. Raja, A. Jameel, and I. Basit, "Detection of glaucoma using cup to disc ratio from spectral domain optical coherence tomography images," IEEE Access, vol. 6, pp. 4560–4576, 2018.

[24] S. Yang, L. Li, S. Wang, W. Zhang, Q. Huang, and Q. Tian, "SkeletonNet: a hybrid network with a skeleton-embedding process for multi-view image representation learning," IEEE Transactions on Multimedia, vol. 1, no. 1, 2019.

[25] W. Zhao, L. Yan, and Y. Zhang, "Geometric-constrained multi-view image matching method based on semi-global optimization," Geo-Spatial Information Science, vol. 21, no. 2, pp. 115–126, 2018.

[26] W. Zhou, H. Li, and Q. Tian, "Recent advance in contentbased image retrieval: a literature survey," 2017, https://arxiv.org/abs/1706.06064.

[27] A. Amelio, "A new axiomatic methodology for the image similarity," Applied Soft Computing, vol. 81, p. 105474, 2019.

[28] C. Celik and H. S. Bilge, "Content based image retrieval with sparse representations and local feature descriptors: a comparative study," Pattern Recognition, vol. 68, pp. 1–13, 2017.

[29] T. Khalil, M. UsmanAkram, S. Khalid, and A. Jameel, "Improved automated detection of glaucoma from fundus image using hybrid structural and textural features," IET Image Processing, vol. 11, no. 9, pp. 693–700, 2017.

[30] L. Amelio, R. Jankovi´c, and A. Amelio, "A new dissimilarity measure for clustering with application to dermoscopic images," in Proceedings of the 2018 9th International Conference on Information, Intelligence, Systems and Applications (IISA), pp. 1–8, IEEE, Zakynthos, Greece, July 2018.

[31] S. Susan, P. Agrawal, M. Mittal, and S. Bansal, "New shape descriptor in the context of edge continuity," CAAI Transactions on Intelligence Technology, vol. 4, no. 2, pp. 101–109, 2019.

[32] L. Piras and G. Giacinto, "Information fusion in content based image retrieval: a comprehensive overview," Information Fusion, vol. 37, pp. 50–60, 2017.

[33] L. Amelio and A. Amelio, "Classification methods in image analysis with a special focus on medical analytics," in Machine Learning Paradigms, pp. 31–69, Springer, Basel, Switzerland, 2019.

[34] AfshanLatif et al. Content-Based Image Retrieval and Feature Extraction: A Comprehensive Review, Hindawi, Mathematical Problems in Engineering, 21 pages, August 2019, https://doi.org/10.1155/2019/9658350.