

# An Extensive Survey on State of Art Approaches in Content based Image Retrieval Systems

**P. Sasikumar\*, Dr. K. Venkatachalapathy**

<sup>1</sup>Research Scholar, Dept. of Computer and Information Science, Annamalai University, Chidambaram, India

<sup>2</sup>Professor, Dept. of Computer and Information Science, Annamalai University, Chidambaram, India

Email: [mailto:ausasikumar@gmail.com](mailto:mailto:ausasikumar@gmail.com), [omsmeetha@rediffmail.com](mailto:omsmeetha@rediffmail.com)

## Abstract

Content based image retrieval (CBIR) is mainly applied for the retrieval of more relevant images from large databases based on query image (QI), provided by the client. CBIR system is operated on low-level visual features of a client input QI that makes the query formulation a difficult one and is it not suitable in generating better results. Presently, CBIR works on the strategy of automatic allocation of keywords to images which guides the clients to make QI based on the keywords to retrieve images. Image annotation is assumed as the complex issue of image classification in which images are exhibited by some low-level features as well as mapping low-level features to high-level features are carried out by applying supervised learning methodologies. This paper performs a review of different feature extraction and deep learning (DL) based CBIR models. Initially, a detailed survey of color, texture and spatial features are described. Followed by, an extensive review of DL models developed for CBIR has been provided. At last, a brief comparative analysis is made with the reviewed models under several aspects. This survey will provide a clear overview of state of art methods available for feature extraction in CBIR along with DL based CBIR models.

**Keywords:** Query Image, deep learning, CBIR, Bag of Visual Words

## 1. Introduction

The evolutions of recent techniques enhance the utilization of Internet, camera, and mobile phones. The received and shared multi-media data are increasing, and retrieving related images from a database is a difficult one. The primary requirement of any image retrieval method are searching and sorting the images which are in visual semantic relation using the query image (QI) offered by the clients. Generally, the search engines work on the network recovery of images based on text approach that need caption as input data [1]. The clients submit queries by enrolling some keywords that are corresponding with the texts which are placed in the archive. The final outcomes are produced on the basis of similarities in keywords, and it is

applied for dissimilar image content. The dissimilarity in the views of individual's awareness and physical labelling is the important cause for creating outcomes which is not relevant. It is nearly not possible to relate the idea of manually labelling previous size image archives with massive amount images. The alternate model for retrieval of images and testing applies an automated image annotation mechanism that tags images according to the image content. This model depends upon automated image annotation accuracy which identifies shape related details like texture, layout, spatial edges, and colour.

Major researches were carried out for enhancing the efficiency of automated image annotation, but the variation in visual point provides inferior impression about image retrieval (IR) process. Content based image retrieval (CBIR) is a platform that surmounts the issues since they are depend upon visual examination of data which is considered as an element of QI. The QI can be fed as an input data under the mapping of images placed inside an archive, and resemblance in visual closeness with respect to image feature vector that offers a base for the identification of image with same content. Here, low-level visual features are calculated from the query and comparing the features is validated to arrange the outcomes [2].

Query-By-Image Content (QBIC) and simplicity are the instance of IR technique depends upon the filtration of low-level visual semantics. Once the predefined models are successfully executed, CBIR as well as feature extraction models are implemented in different software's such as textile industry, remote sensing, armed forces, video realization, crime detection, and clinical image analysis. Fig. 1 gives an outline of the fundamental concept and process of IR.

The fundamental requirement for some IR classification is to explore and arrange related image from the archive with least amount of individual interface with the device. This paper discussed election of visual features for a structure which depends upon the necessities of the customer. The distinctive feature demonstration is highly required for any IR model. Moreover, the features can become highly effective and robust under the combination of low-level visual features, and maximum processing cost is necessary to accomplish better outcome. Unfortunately, the imbalanced feature selection limits the efficiency of IR process. Machine Learning (ML) model gets the feature vector as input data for training as well as testing methodologies that maximizes the performance efficiency.

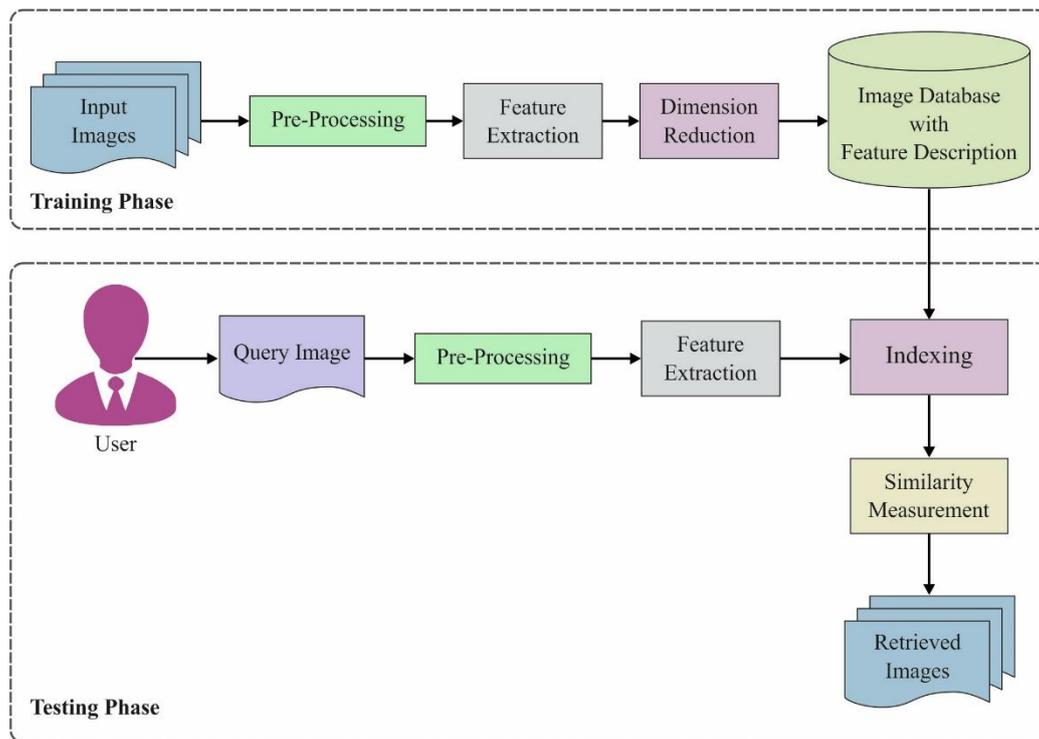


Fig. 1. General block diagram of CBIR

Recently, the IR process is highly dependent on Deep Neural Networks (DNN) which is capable of providing optimal results at expensive platform. This paper aims to give a compact outline of complex objectives in CBIR as well as feature depiction. The presented models are defined briefly by showing the major contributions and future objectives of CBIR and feature extraction are also described.

Image annotation is assumed as the complex issue of image classification in which images are exhibited by some low-level features as well as mapping low-level features to high-level features are carried out by applying supervised learning methodologies. This paper performs a review of different feature extraction and deep learning (DL) based CBIR models. Initially, a detailed survey of color, texture and spatial features are described. Followed by, an extensive review of DL models developed for CBIR has been provided. At last, a brief comparative analysis is made with the reviewed models under several aspects.

## 2. Background information: Query Formulation

In CBIR, the client illustrates his or her imagined intent to few concrete visual queries. A feature of the query has an important fact on the retrieval outcomes. The better as well as particular query can suitably diminish the retrieval complexity and cause satisfactory retrieval

outcomes. Usually, there are different types of query structure, namely query by sketch, image, color map, and so on. Several query methods produce considerably distinguishing outcomes. Subsequent, it is discuss all of individuals representative query structures. A common query structure is query-by-example images. i.e., a client has an instance image and is similar in the retrieving further or optimal images on the similar or same semantic. For example, the image holder can need to ensure the image is utilized in few websites with no approval; cybercrime require verifying a terrorists logos shown in webpages for antiterrorist activities. For eliminating the cause of the backdrop, bounding boxes can be particular in the instance image to constrain the targeted area for query. As the instance images are purpose without small human contribution, it can be suitable to create quantitative examination depends on it in order to lead plan of the equivalent techniques. So, query-by-example is an extremely search query structure method in the study on CBIR.

Also query-by-example, a client can also illustrate the target using sketch maps. Thus, the QI is a contour image. Because sketch is further near to semantic illustration, it inclined to assist retrieve intention outcomes in clients' thoughts from the semantic viewpoint. Primary studies based on sketch oriented CBIR are restricted for exploring to superior artworks, namely clip art and easy pattern. Because an edge is treated as an important element in the sketch based retrieval process and is used in several searching engines. But, there are 2 non trivial problems on sketch based query. Initially, while a few easy models, namely fish, sun and flower, may be simply taken as easy shapes, mostly, it can be complex for a client to rapidly sketch out what he needs to explore. Furthermore, as the images in the dataset are generally ordinary images, it requires designing exceptional techniques for converting it into sketch maps dependable with client target. Other query structure is color map. The client is permitted to identify the spatially distributing the colors is provided grid similar palette for creating a color map that is utilized as query to IR with same colors in the comparative areas of the image. By grainy shape embedding, the color map dependent query is simply involving client communication for improving the retrieval outcomes although is restricted with potential models to be signified. Also color modifications is common in image acquisition which poses diverse issues in the color features.

The exceeding query structures are suitable to utilize for input but might until be complex for expressing the client's semantic objective. For alleviating this issue, structure the query with text words in some particular outline in the image plain is developed. A structured object query is also search by a latent ranking SVM method. It can be noticeable, in the above query systems

obtained by most predefining work; the query gets the structure of single image that can be insufficient to reflect client intension in some conditions. When offered with many probe images as query, a few novel schemes are supposed to collaboratively signify the query or fuse the retrieval outcomes of all single probes. So, it can be interesting to investigate the process of retrieving videos using a sequence of frame as query.

In CBIR, the key issue is the way of proficiently computing the comparison among images. Because the visual objects or scenes can endure several alterations or transformations, it can be impossible to directly relate images in pixel-by-pixel basis. Generally, visual features are removed from images and afterwards changed into a fixed size vector for image demonstration. Assuming the conflict among large databases and necessity to efficiently attain query responses, it can be essential to set the visual features to make possible the subsequent indexing as well as image similarity.

### **3. Review of CBIR Models**

#### **3.1. Existing works on Color Features based CBIR models**

Color is the main significant low-level visual features in the eye can distinguish among visual based upon colour. The colour feature is fixed and firmly influenced by the image rotation, scale, and translation. During the utilization of Dominant Color Descriptor (DCD), general color details of image are interchanged with minimum quantity of colours. DCD is proved as main MPEG-7 colour descriptors and applies a powerful, small, and instinctive arrangement for reporting the indicative color distribution and feature.

[3], proposed conventional method can recover image related to its label and annotation which unable to satisfy the client needs; thus, this paper concentrated on a new image retrieval process which is dependent to its content. The planned techniques use an undersized image descriptor that is unreliable based upon content by a 2 phase clustering method. COIL-100 image records are utilized for the research. Simulation results got from the research confirmed that proposed model is good.

According to [4] retrieving colour of the picture is depends on its content, which is begins from the combination of colour and consistency features. It gives an effectual and unpredictable evaluation of how premature person could made visual contents. The colour consolidation and texture features provides dynamic set of colour IR methods. The outcome captured from the research expose that the proposed technique recovered pictures gives further accuracy

compared to former conventional technique However; the feature dimension is small and needs a prominent calculation charge. A pair-wise evaluation for the two low-level features is utilized for determining relationship measures.

Many researches performed an investigation on the overall function of invariant descriptor. The moment functions are composed of orthogonality as well as rotation invariant. PZM verified that it must have high dynamic to noisy image throughout the Zernike moments. [5] offered a novel technique to receive entire pseudo-Zernike moment invariants groups. The connection among pseudo-Zernike moments of actual image and shape, however difference scales and orientations of images are developed initially. A total group of scales and rotations invariant got through this connection. This model showed that it has high efficiency in identifying pattern about former approaches.

[6] suggested a novel technique for cataloguing photographs on the basis of features obtained from Error Diffusion Block Truncation Coding (EDBTC). In order to initialize image feature descriptor, 2 color quantizers and bit-map images with vector quantization (VQ) are employed on the basis of EDTBC. For imposing the similarities among the QI, 2 features like Color Histogram Feature (CHF) and Bit Pattern Histogram Feature (BHF) has been utilized. These 2 modules are obtained from VQ-indexed color quantizer as well as VQ-indexed bitmap image, correspondingly. The distance measured from CHF and BHF extract the possibilities from images. Simulation outcomes exhibits that the planned method works proper than previous BTC oriented image indexing and former existing IR method. The EDTBC has high-quality of image compression and indexing image for CBIR.

[7] applied a Multiscale Distance Coherence Vector (MDCV) for CBIR. The reason at the back is that the various shapes might have parallel descriptor and DCV is not applicable in reducing the entire artefacts. The newly developed model applies the Gaussian functions for maximizing the image contour curve. The proposed model is fixed for diverse process such as scaling transformation, rotation, and translation.

### **3.2. Existing works on Texture Features based CBIR models**

In order to determine the presented method (MSD) for IR, [8] performs the experiments on Corel datasets as it is not composed of definite datasets for CBIR. Corel-5000 and Corel-10000 were employed with maximum number of images, followed by HSV, RGB, and Lab color space for estimating the IR function. These 2 datasets have maximum retrieval and recall rates

for MSD under the application of various color quantization levels whereas texture based quantization levels are estimated and provides optimal HSV and Lab color space, however, it is inferior on RGB color space. For obtaining best outcome for memory space, retrieval precision, and efficiency, maximum color as well as quantization levels were utilized in MSD whereas minimum numbers are used for IR task. The maximum retrieval as well as recall proportions of MSD are related with alternate approaches like Gabor MTH on Corel datasets are applied for IR task, where the final outcome of MSD exhibits that, it performs quite-well than compared technologies.

Numerous color images have been gathered from natural scenarios like disasters, clients, and textures to process the IR function according to the textures. Basically, the retrieving results are composed of smoothness, regularity, distribution, and coarseness along with color data. The accuracy comparison over the presented method (colorCM) and gray-level co-occurrence matrix (GLCM) model offers the final outcome for the determination of projected approach. Due to the adequate properties present in color co-occurrence matrix, it is highly effective when compared with GLCM. For CBIR, Corel, COIL, and Caltech-101 datasets with numerous images classes for corresponding datasets have been employed. The mean accuracy as well as recall rates accomplished by projected approach is achieved by unifying the Neural Network (NN) with bandlet transform and the state-of-the-art retrieving modules. The experimental outcome depicts that the study provided in performs highly effective when compared with respect to mean accuracy and recall rate.

Using the Corel image gallery and images for classical IR, [9] carried out the process for exhibiting the efficiency of projected approach that depends upon SVM structure. This image gallery is classified as, Corel A and Corel B. Initially, Corel A is composed of 1000 images which is further categorized into 10 classes while Corel B contains 9900 images. Followed by, mean accuracy and recall rates are accomplished by the presented model with remarkable IR models. Diverse amount of attained images has been applied for executing the retrieval ability of SVM which implements better outcome. Therefore, the results and comparison implies that the projected method accomplishes optimal outcome which is highly reliable for IR task. [10] computed the performance on Brodatz and Vistex datasets for CBIR with gray-scale as well as color images for predefined datasets. Distance from QI and a dataset image should be estimated, and the image with least distance is obtained and calculates the precision as well as recall rates and finally, compared the simulation outcome with one another.

### 3.3. Existing works on Spatial Features based CBIR models

The image spatial features are based on the position of objects inside 2D image space. Bag of Visual Words (BoVW) is defined as well-known models which eliminates image spatial layout at the time of showing image histograms. Spatial Pyramid Matching (SPM) is meant to be familiar approach which is capable of capturing image spatial parameters; however it is delicate for rotations and scaling. [11] employed an approach for encoding respective spatial data to show the histogram of BoVW method. It is stimulated by calculating the overall geometric correlation from the sets of identical VW that corresponds to the core part of an image. Here, 5 databases are applied to assess the process of presented model which depends upon the corresponding spatial data.

[12] deployed Hybrid Geometric Spatial IR (HGSIR) under the application of image classification relied approaches. To show the better experimental outcome of the presented model, 5 datasets has been applied. The simulation outcome implies that the study performs optimal over existing methodologies with respect to accurate image classification. Alternatively, a new framework to show the images with spatial data to reversed index of BoVW. The spatial details are fixed by processing the overall spatial inclination of VW in gyration invariant. The arithmetical relation of VW has been determined. It is accomplished by processing by orthogonal vector from single point of identical VW. The histogram of VW is processed on the basis of orthogonal vector size which offers the data regarding corresponding location of linear VW. In order to estimate the provided approaches, 4 datasets have been utilized.

[13] projected 2 models for computing the image representation. The base of such methods are histogram of triangles which embeds the spatial data for reversed index of BoF depiction. A single image is classified as 2 or 4 triangles that is applied separately to calculate the histograms of triangles for 2 levels. It applies 2 datasets to compute the newly developed model. Experimental results depicts that the projected method outperforms in IR process. [14] employed a method under the application of symbol analysis which has been processed with the help of scale-invariant feature transform BoVW. For adding the spatial data to BoVW, it applies circular tiling while changing the angles histograms in previous technique (proposed by Rahat) for making rotation invariant, since it is not rotation invariant in advance. Next, the adjusted angles are combined with circular tiling's that obtains enhanced classification value and minimizes computation difficulty.

[15] presented a global as well as local relative spatial distribution of VW across an image termed as soft pair-wise spatial angle-distance histogram. The primary objective of this model is to offer effective representation of relative spatial details and computes the classifications, thus researchers ended up that the developed approach outperforms and enhance the working efficiency.

### **3.4. Existing works on Deep Learning based CBIR models**

Based on the QI provided by the client to the neural network (NN) model, the classifier has the ability of classifying the content with respect to different features. This section deals with the use of DL models, which confine all data, performs learning the content by the separation of the features to the bottom. The database manages a single data centre which comprises finite important features [3]. DL models exhibits its better results and extracts the data effectively. The DL model retrieves the content based on the representation of features and similarity determination. The process of finding digital image from huge database is frequently needed and so CBIR called as query-based IR (QBIR) is utilized. Several methods are employed to solve the problems like scale-invariant transformation and vector of locally collected descriptor. The accomplished simulation outcomes have immense efficiency in Deep Convolutional Neural Network (DCNN), a novel frequency-inverse document frequency (TF-IDF) is applied as vector descriptor with weighted convolutional text frequency on CNN. Then, the learned file of convolution neuron model in convolutional layers was applied for deducting VW. Here, degree of the visual pattern is offered by assuming all filters as TF portion while 3 methods are considered to be IDF portion as projected [16].

By presenting great IR approach with an improved results, the methods links the TF-IDF with CNN analyses for visual contents. Developers have performed several researches on 4 IR database and the test results demonstrate the subsistence of a method. For handling the large-scale, [17] projected a hashing method for extracting the features from image and learn its binary representation. Here, developers modelled a pair-wise matrix and an objective function using DL platform which learn its image binary representations. The research is carried out on 1000 of histopathology image to show the accuracy of the proposed model.

[18] planned unsupervised VH technique named as Semantics based Visual Hashing (SAVH). It model applies 2 elements namely, off-line learning and on-line learning. In former one, the image pixels are converted into arithmetical vector representation by removing the optical as well as consistency feature. Afterwards, maximizing the text in the visual graph using topic

hypergraph, followed by hash code of an image is skilled for conserving the association of image among the semantic and image, and, the hash function code is created inside linear aggressive method. Such characteristics are applied for enhancing the state of CBIR. In computer vision application, CNN utilizations have exposed an extraordinary efficiency, particularly in CBIR methods. Mostly, CNN techniques obtain the feature in the preceding layer under the application of single CNN and seamless quantization method and its demerits restrict the application of convolutional layer to identify local image pattern. Hence, the newly developed model is recognized as bilinear CNN-based model. This process applied 2 parallel CNN model for separating the feature with no advanced details of semantics. The feature is removed for activating the convolutional layer instead of reducing low-dimensional feature. The model is tested and offers extreme intimacy. It is highly applicable for reducing the image representation to the greater extent as it applies diverse quantized levels for feature extraction, thus it is effective in IR function, searching time as well as storage cost. Later, the bi-linear CRB-CNN is incredibly performance by learning a difficult image which has various semantic. Only minimum duration is required for eliminating the feature and undersized disk is essential to characterize and save the image. At last, dedicated tanning is used with no influence of metadata, annotations, tags which is embedded with CRB-CNN to process the feature extraction in CBIR. Moreover, this model is used on the large-scale database to perform IR task and depicts the retrieving function.

In order to compute an effective image searching, hashing function is a well-known model which gathers massive client concentration in CBIR. It provides identical binary code for the same image context that maps high-dimensional visual details to low-dimensional information. It has been considered that the semantic labels are exhibited using multiple latent layer attributes whereas classification is based on these parameters. The primary aim of SSDH is that, the unification of retrieval and classification modules into single mechanism. SSDH is reliable for large-scale search, and the extended version of DNN for classification, SSDH is elegant and simply accessible.

Remarkable image examinations as well as discrimination of visual details are considered to be open research issues. Numerous studies were presented under the application of combined views using graph-based approach. It is highly complex than existing models which determines the boundaries of image and search resemblance of consistency limits. The researchers have developed a Multi-View Label Sharing (MVLS) approach for open research issues and it is highly complex to reach the better association. For classifying the visual data and its

representation, optimization is carried out and it combines the transformation and classification attributes for transformation matrix learning as well as classifier training. The final outcome attained from MVLS is composed of 6 views and 9 views which are computed in future. The simulation outcomes are related with a number of state of the art researches and outcomes exhibits the efficiency of the planned MVLS model.

To understand the image and object classification, approaches like CNN and local feature have displayed superior outcomes under various platforms. The application of CNN approach is yet a difficult for precise categorization while handling minimum training data and labels. The smooth constraints are applied to maintain the semantic gap; however the performance of CNN is degraded as it is composed of minimum sized training set. Developers in projected an MV framework with few label and view consistency (MVFL-VC). The discriminative efficiency of the learned factor is increased via unlabeled training image. For estimating the proposed model, experiment is carried out in various databases. The projected MVFL-VC technique could be utilized with alternate image classification as well as representation approaches. The analysis and simulation outcome shows the efficiency of newly developed method.

In [21], a new model for classifying produce images has been developed. The presented model uses Eigenvector based fusion technique, ANN and SVM. This model is validated on PI100 database and attained a detection rate of 87.2%.

Extracting domain space knowledge is applicable for reducing the semantic gaps. The researchers have established multi-view semantics representation (MVSR) that is mainly applied for semantics representation and visual analysis. The presented model classifies the images according to the semantics and visual similarities. There are 2 visual similarities for training samples which offers reliable and identical awareness are capable for handling diverse classification models and views. This approach is based relied on MVSR and it is highly discriminative when compared with alternate modules as the semantic data is applied for upcoming iteration from every perception as well as from unique set of images. This model can be determined by using publicly available dataset and the attained results showcases the efficiency of MVSR. Hence, the final outcome demonstrates that MVSR enhanced the classification function with respect to precision for image sets with maximum visual differences. A brief explanation of different feature extraction and DL based CBIR models is showcased in Table 1.

**Table 1** A summary of different feature extraction and DL approaches for CBIR

<b>Author</b>	<b>Application</b>	<b>Method</b>	<b>Feature</b>	<b>Dataset</b>	<b>Performance measure</b>
Duanmu [3]	IR	Moment invariant	Color	COIL-100	-
Wang et al. [4]	CBIR	Color, texture features	Color	Corel	61% of accuracy
Zhang et al. [5]	Object identification	Complete set of pseudo-Zernike invariants	Color	COIL-100	-
Guo et al. [6]	CBIR	Error diffusion, block truncation coding	Color	Corel	80% of accuracy
Jiexian et al. [7]	CBIR	Multi-scale distance coherence vector model	Color	MPEG-7 image database	97% of accuracy
Liu et al. [8]	IR	MSD	Texture	Corel databases	precision and recall Corel 5k: 0.56, 0.067 Corel10k: 0.41, 0.055
Lasmar and Berthoumieu [19]	Texture image retrieval	GC-MGG and GCMWbl	Texture	Vistex, Brodatz, ALOT	-
Fadaei et al. [9]	CBIR	LDRP	Texture	Brodatz and Vistex	Precision: 0.808, 0.9191
Karpathy and Fei-Fei [20]	Generation of descriptions of image regions	CNN and multimodal RNN	deep-learning	Flickr8K, Flickr30 K and MSCOCO	Encouraging result

Kondylidis et al. [16]	CBIR	CNN based tf-idf	deep-learning	INRIA Holidays, Oxford 5k, Paris 6k, UK Bench	Improved result
Shi et al. [17]	Retrieving and classifying histopathology images	PDRH technique	deep-learning	Medical images	Accuracy: 0.975
Oyewole and Olugbara [21]	Classifying produce images	Eigen color fusion with ML	Fusion	PI100 corpora	Accuracy: 0.87

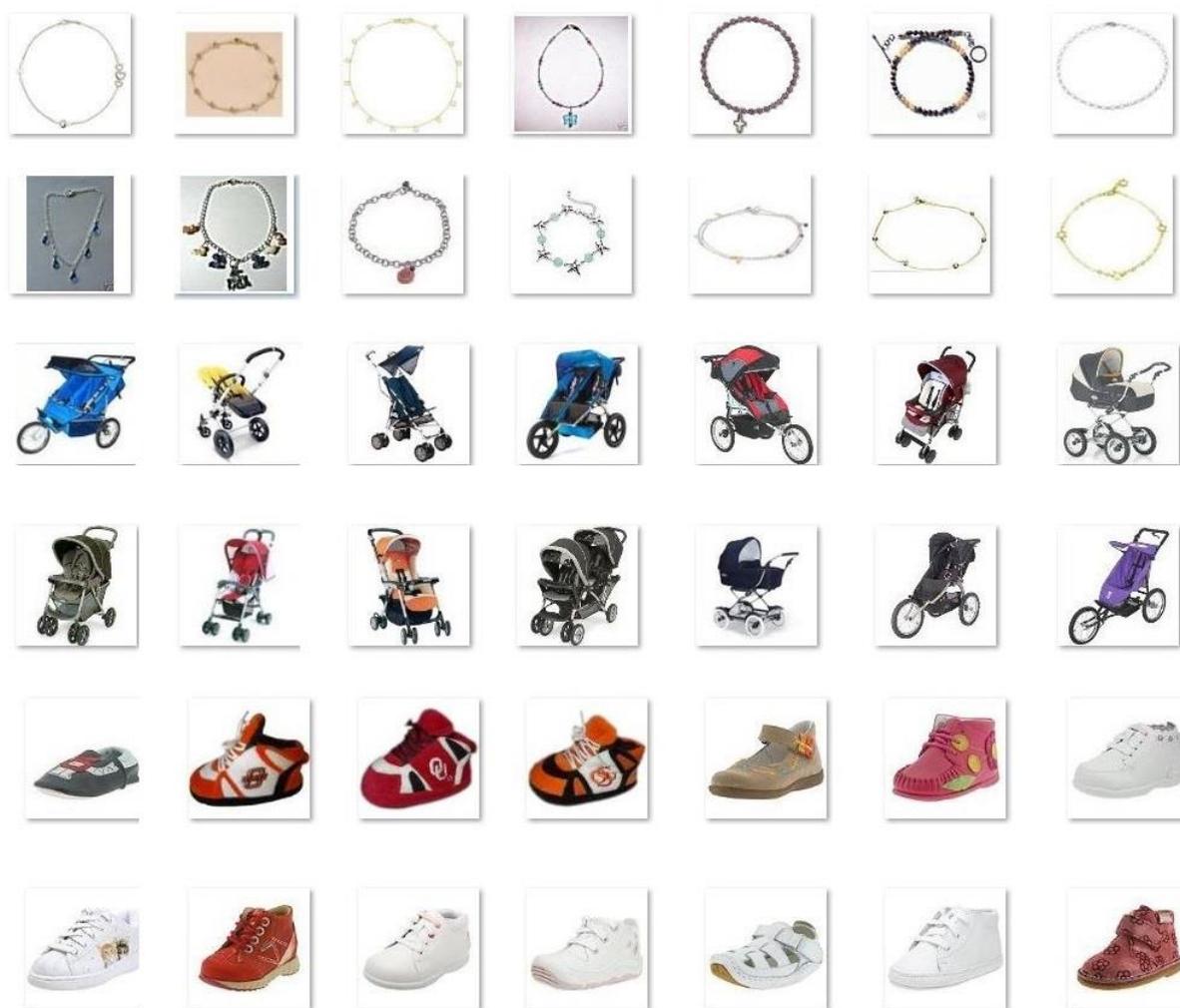
#### 4. Discussion

The dataset commonly used to validate the performance of the CBIR model is Corel10 dataset. Sample test images from Corel10K dataset is shown in Fig. 2. It comprises 100 classes and totally 10,000 images from diverse contents such as sunset, beach, flower, building, car, horses, mountains, fish, food, door, etc. Every class label comprises 100 images of size  $192 \times 128$  or  $128 \times 192$  in the JPEG format.



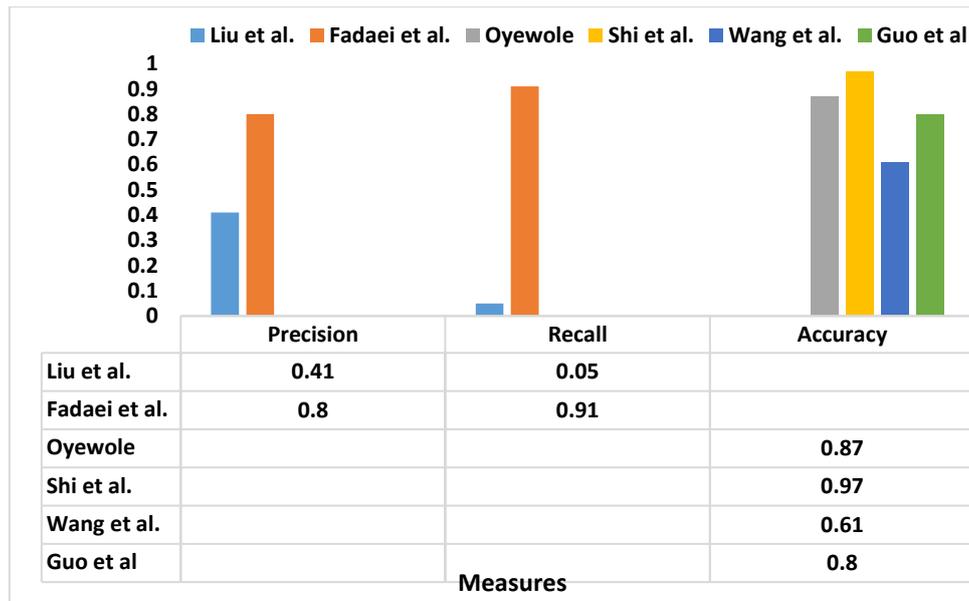
**Fig. 2.** Sample test images from Corel10K dataset [22]

Another dataset used for the validation of the CBIR is CI100 dataset. The corpus contains 10000 low resolution color images of e-commerce products, comes with a total of 100 classes. Every image comprises the dominant object in comparatively stable form, which is precisely the way product images generally look on e-commerce websites like Amazon.com, Jumia.com, eBay, Web.com and Volusion. Sample test images from PI100 dataset is shown in Fig. 3.



**Fig. 3.** Sample images from PI100 dataset [21]

Fig. 4 provides a comparative study of different methods in terms of precision, recall and accuracy. On measuring the results in terms of precision and recall, it is indicated that the method developed by Liu et al. has attained ineffective retrieval performance with the minimum precision and recall of 0.41 and 0.8. Besides, the model devised by Fadaei et al. has attained higher precision and recall of 0.05 and 0.91 respectively. Similarly, on assessing the retrieval performance in terms of accuracy, the method by Wang et al. has failed to exhibit better results by attaining a minimal accuracy of 0.61. In the same way, the model by Guo et al. has attained somewhat higher accuracy of 0.8 whereas even higher accuracy of 0.87 has been offered by Oyewole model.



**Fig. 4.** Comparative analysis of different methods interns under several measures

Finally, a maximum retrieval performance is provided by Shi et al. by achieving an accuracy of 0.97. These experimental analysis indicated that the performance of the retrieval process is yet to be enhanced in the future works.

#### 4. Conclusion

Due to the exponential increase in the digital media, it is essential to improve the process of searching and retrieving images from the database. The major issue posed in this task in the time needed to perform retrieval process and maximum retrieval rate. Based on the QI posed by the client, the CBIR model will retrieve the images from the database. This paper has performed a comprehensive review of different feature extraction and DL based CBIR models. At the beginning, a detailed survey of color, texture and spatial features are described. Followed by, an extensive review of DL models developed for CBIR has been provided. At last, a brief comparative analysis is made with the reviewed models under several aspects. An experimental analysis indicated the need of designing new CBIR models to attain maximum retrieval rate.

#### References

- [1] 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.

- [2] 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.
- [3] X. Duanmu, "Image retrieval using color moment invariant," in *Proceedings of the 2010 Seventh International Conference on Information Technology: New Generations (ITNG)*, pp. 200–203, IEEE, Las Vegas, NV, USA, April 2010.
- [4] X.-Y. Wang, B.-B. Zhang, and H.-Y. Yang, "Content-based image retrieval by integrating color and texture features," *Multimedia Tools and Applications*, vol. 68, no. 3, pp. 545–569, 2014.
- [5] H. Zhang, Z. Dong, and H. Shu, "Object recognition by a complete set of pseudo-Zernike moment invariants," in *Proceedings of the 2010 IEEE International Conference on Acoustics Speech and Signal Processing (ICASSP)*, pp. 930–933, IEEE, Dallas, TX, USA, March 2010.
- [6] J. M. Guo, H. Prasetyo, and J. H. Chen, "Content-based image retrieval using error diffusion block truncation coding features," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 25, no. 3, pp. 466–481, 2015.
- [7] Z. Jiexian, L. Xiupeng, and F. Yu, "Multiscale distance coherence vector algorithm for content-based image retrieval," *@e Scientific World Journal*, vol. 2014, Article ID 615973, 13 pages, 2014.
- [8] G.-H. Liu, Z.-Y. Li, L. Zhang, and Y. Xu, "Image retrieval based on micro-structure descriptor," *Pattern Recognition*, vol. 44, no. 9, pp. 2123–2133, 2011.
- [9] A. Irtaza and M. A. Jaffar, "Categorical image retrieval through genetically optimized support vector machines (GOSVM) and hybrid texture features," *Signal, Image and Video Processing*, vol. 9, no. 7, pp. 1503–1519, 2015.
- [10] S. Fadaei, R. Amirfattahi, and M. R. Ahmadzadeh, "Local derivative radial patterns: a new texture descriptor for content-based image retrieval," *Signal Processing*, vol. 137, pp. 274–286, 2017.
- [11] B. Zafar, R. Ashraf, N. Ali et al., "A novel discriminating and relative global spatial image representation with applications in CBIR," *Applied Sciences*, vol. 8, no. 11, p. 2242, 2018.
- [12] N. Ali, B. Zafar, F. Riaz et al., "A hybrid geometric spatial image representation for scene classification," *PLoS One*, vol. 13, no. 9, Article ID e0203339, 2018.
- [13] N. Ali, K. B. Bajwa, R. Sablatnig, and Z. Mehmood, "Image retrieval by addition of spatial information based on histograms of triangular regions," *Computers & Electrical Engineering*, vol. 54, pp. 539–550, 2016.

- [14] H. Anwar, S. Zambanini, and M. Kampel, “A rotation-invariant bag of visual words model for symbols based ancient coin classification,” in Proceedings of the 2014 IEEE International Conference on Image Processing (ICIP), pp. 5257–5261, IEEE, Paris, France, October 2014.
- [15] R. Khan, C. Barat, D. Muselet, and C. Ducottet, “Spatial histograms of soft pairwise similar patches to improve the bag-of-visual-words model,” *Computer Vision and Image Understanding*, vol. 132, pp. 102–112, 2015.
- [16] N. Kondylidis, M. Tzelepi, and A. Tefas, “Exploiting tf-idf in deep convolutional neural networks for content based image retrieval,” *Multimedia Tools and Applications*, vol. 77, no. 23, pp. 30729–30748, 2018.
- [17] X. Shi, M. Sapkota, F. Xing, F. Liu, L. Cui, and L. Yang, “Pairwise based deep ranking hashing for histopathology image classification and retrieval,” *Pattern Recognition*, vol. 81, pp. 14–22, 2018.
- [18] L. Zhu, J. Shen, L. Xie, and Z. Cheng, “Unsupervised visual hashing with semantic assistant for content-based image retrieval,” *IEEE Transactions on Knowledge and Data Engineering*, vol. 29, no. 2, pp. 472–486, 2017.
- [19] N.-E. Lasmar and Y. Berthoumieu, “Gaussian copula multivariate modeling for texture image retrieval using wavelet transforms,” *IEEE Transactions on Image Processing*, vol. 23, no. 5, pp. 2246–2261, 2014.
- [20] A. Karpathy and L. Fei-Fei, “Deep visual-semantic alignments for generating image descriptions,” in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 3128–3137, Boston, MA, USA, June 2015.
- [21] Oyewole, S.A. and Olugbara, O.O., 2018. Product image classification using Eigen Colour feature with ensemble machine learning. *Egyptian Informatics Journal*, 19(2), pp.83-100.
- [22] <http://www.ci.gxnu.edu.cn/cbir/Dataset.aspx>