

Shape and Geometric Features-based Semantic Image Retrieval Using Multi-class Support Vector Machine

Zohair Malki

College of Computer Science and Engineering in Yanbu, Taibah University, Saudi Arabia

Zmalki@taibahu.edu.sa

Abstract

In this paper, a new approach to retrieve semantic images based on shape and geometric features of image in conjunction with multi-class support vector machine is proposed. Zernike moment as shape feature is to verify the invariance of objects for silhouette image. In addition, a set of geometrical features is to explore the objects shape using two features of rectangularity and circularity. Then the extracted features are normalized and employed for multi-class support vector machine either for learning or retrieving processes. The retrieving process relies on three main tasks which namely Query Engine, Matching Module and Ontology Manger, respectively. Query Engine is to build the input text or image query using SPARQL language. The matching module extracts the shape and geometric features of image's objects and employ them to Ontology Manger which in turn inserts them in ontology knowledge base. Benchmark mammals have been conducted to empirically conclude the outcome of proposed approach. Our experiment on text and image retrieval yields efficient results to problematic phenomena than previously reported.

Keywords: Zernike moment, Multi-class support vector machine, Query Engine, SPARQL.

1. Introduction

Image retrieval is a method for searching, browsing and retrieving images from a big digital images database. In the last decade, researchers focused on low-level vision in which the retrieval accuracy is still out the user expectation. So, interactive techniques appeared in current years to alleviate the large gap between low-level and high-level concepts [1, 2, 3, 4]. To interpret the meaning images, the low-level features such as color, texture and shape are mostly used. In contrast, the high-level techniques are active to retrieve patterns by scanning whole image. The semantic digital image retrieval is one the talented exploration field where several researchers consider approaches either for image analysis [5, 6, 7, 8] or image retrieval [9, 10, 11, 12]. Most techniques of image retrieval usually used text meta-data that relied on the description oh image's textual [13, 14, 15]. Currently, the retrieved of semantic images depends on keyword-based search. But there are silent numerous problems in Google search engines due to the absence of storing key and relationship among web images [16]. In natural image domain, the images with their features are stored in semantic manner to construct an ontological data set. So, the accessible images with their ontological structure will retrieve when the pictorial image's features are allocated as input to SPARQL query. To learn these pictorial features, various classifier techniques are used and then retrieve the images with by features as input query image. A few researches progress on image retrieval with respect to content similarity. So, a slight experimental prototype techniques have been conducted as QBIC [17], Photobook [18], Netra [19], Virage [20], SIMPLICITY [21] and Visual SEEK [22].

Furthermore, Content-based image retrieval (CBIR) is stated as all-inclusive surveys by [23, 24]. S.K. Chang and S.H. Liu presented an approach to index and retrieve images with

respect to proposed database retrieval [25]. The images are indexed using their pictorial content such as color, texture and shape. There are several methods that retrieve images relied on either content-based or text-based. The indispensable difference between text-based and content-based is that the Human Interaction (HI) characterizes exceptional portion of final approach. Besides, the human's engagement the text keywords and descriptors as high-level features to interpret and measure images similarity. If not most features which routinely take out via the computer vision techniques characterize low-level features as color, texture, shape and spatial layout. In wide-ranging, the upfront link is not occurring between the low-level and high level in concepts and features and [26].

The main contribution is to propose a new approach to retrieve semantic images based on shape and geometric features of image in conjunction with multi-class support vector machine is proposed. A set of geometrical features is to explore the objects shape using two features of rectangularity and circularity. The retrieving process relies on Query Engine, Matching Module and Ontology Manger. Query Engine is to build the input text or image query using SPARQL language. The matching module employs the shape and geometric features of image's objects Ontology Manger constructs ontology knowledge base. Our experiment on text and image retrieval of benchmark mammals yields efficient results to problematic phenomena when applying fortunate with large number of tested images. Our paper is organized as follows; Section 2 gives a detailed attributive of shape and geometric features, and support vector machine. In Section 3, the suggested approach is detailed. After that, the experimental results are debated in Section 4. Section 5 concludes the proposed approach.

2. Related Literature

The success of retrieving semantic images due to a good selection of classifier technique in addition to features extraction which carry out the vigorous view-invariant task. The motivation behind this section is to understand and investigate the novel techniques that capable for real-time application. The following two subsections briefly review the shape and geometric feature as well as multi-class support vector machine classifier.

2.1 Shape Feature: Zernike Moments

Zernike moments are the orthogonal set of rotation-invariant moments and verify the invariance of objects for silhouette image [27]. The motivation behind using Zernike moment is to carry out the translation and scale invariance using moment normalization. Mathematically, the complex Zernike moment (Z_{pq}) of repetition q and order p for the intensity function $f(\rho, \theta)$ of exact image is;

$$Z_{pq} = \frac{p+1}{\lambda_N} \sum_{x=0}^{N-1} \sum_{y=0}^{N-1} f\left(\frac{x}{a} + \bar{x}, \frac{y}{a} + \bar{y}, t\right) R_{pq}(\rho) e^{-jq\theta} \quad 1$$

where, the normalization factor is represented by λ_N , P is to a positive integer. q represents either negative integer or positive with respect to the constraints condition; $p - |q| = \text{even}$, $|q| \leq p$. With respect to the translation and scale factors, the function f is normalized using the of silhouette image centroid (\bar{x}, \bar{y}) and the scaled factor a . $R_{pq}(\rho)$ is a radial polynomial [27]. Thus, the invariant features of Zernike moment (i.e., invariant moment Hu) to scale, rotate and translate the shape is achieved by $G_z = [z_{00}, z_{11}, z_{22}]$. Here, the percentage error of feature invariants is experimentally smaller than 0.5%.

2.2 Geometric Features

The object shape is geometrical explored with standard specification as circle (i.e., circularity: *Cir*) and rectangle features (i.e., rectangularity: *Rect*) as in Eq. 2. These features can vary from one object to another to discriminate the objects in correct manner.

$$F_{geo} = (Cir, Rect)^T \quad 2$$

2.2.1 Circularity

The shape of the object close to a circle is so-called Circularity. The motivation behind using this feature is to provide circularity as one. In addition, the circularity is ranged from 1 to infinity. Shortly, the circularity is denoted by the symbol *Cir* as follow;

$$Cir = \frac{Perimeter^2}{4\pi \times Area} \quad 3$$

In which; the *Perimeter* represents the object's contour while the *Area* refers to the total number of pixel's object.

2.2.2 Rectangularity

The shape of the object close to a rectangle is so-called Rectangularity. We compute the orientation of target object with respect to the central moments of all contour points. The length *l* and width *w* is calculated by obtaining the difference between the largest and the smallest orientation in a rotation. Thus, *Rect* is equal to 1 and is ranged from 0.5 to infinity (Eq. 4);

$$Rect = \frac{Area}{l \times w} \quad 4$$

Such that, *Area* is to the total pixels of object's contour. Thus, the scalability problem is appearing because of geometrical features infinity in range. The problem is alleviated via normalizing the feature vector as follows;

$$Cir_{norm} = \frac{Cir - minCir}{maxCir - minCir}, Rect_{norm} = \frac{Rect - minRect}{maxRect - minRect} \quad 5$$

where *maxCir* and *minCir* are the maximum and minimum circularity of the object region respectively. It is being noted that, the notation *Cir_{norm}* is the same of rectangularity. Thus, the shape and geometric feature is integrated by the vector *F_{img}* as follow;

$$F_{img} = (z_{00}, z_{11}, z_{22}, Cir_{norm}, Rect_{norm}) \quad 6$$

2.3 Support Vector Machine

There are countless problems to train multiclass, which need a moral selection for the classifier to correctly carry out the retrieving process. Whereas, the extracted features from one frame is assign to on class. The motivation behind using Support Vector Machine

(SVMs) is to have a highly accurate paradigm in addition to its capabilities for outstanding generalization. Moreover, SVMs can deal with structure principle of risk minimization as well as alleviating the problem of data over-fitting exists in neural network [28]. Furthermore, SVMs structure is easily carried out dichotomic classes conspicuously at higher-dimensional space. The ability of SVMs is to build a maximal separating hyper-plane by maximizing the distance between two corresponding hyper-planes as in Fig. 1.

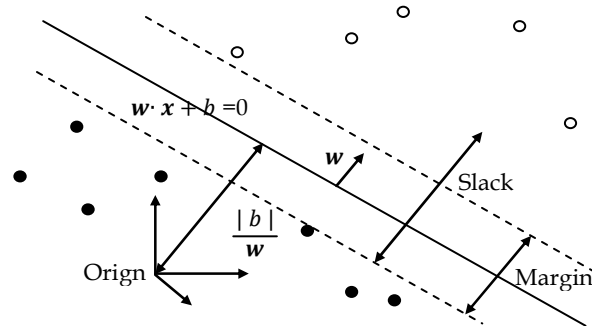


Fig. 1. Sideline of SVMs hyper-plane

From the hyper-plane, we can supplementary be performing good separation relying on largest distance, which margin a low-slung classifier generalization error.

Mathematically, assume that the learning dataset is represented as $D = \{(x_i, y_i) \mid x_i \in R_d, y_i \in \{-1, +1\}\}$. Vapnik *et. al* [28] performed the learning problem by permitting some examples that be debased in their margin limitations. In so doing, slack variables are considered to verbalize the latent violations and to a drawback parameter for blocking the margin violations. The function which linearly used to learn SVMs is stated as;

$$f(x) = \text{sign}(w \cdot x + b) \quad 7$$

where w is a weight vector, x is to input sample and b represents a used threshold. To characterize the minimum distance between hyper-plane and support vectors, we maximized the margin of hyper-plane, which separated via the trained SVMs learner. We can formulate SVMs as follows;

$$\gamma = \frac{2}{\|w\|} \quad 8$$

Such that γ refers to hyper-plane margin. Fig. 1 shows the maximization margin of hyper plane while the input data is linearly separated and mapped to high dimension domain using SVMs (Fig. 2).

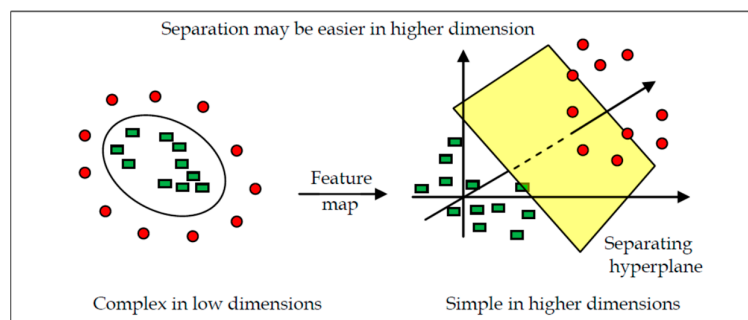


Fig. 2. Mapping task from complex low to simple high dimensions.

The mapping does not impact for the time of learning process since it has dot product and kernel trick. When the number of extracted features is so considerable, the SVMs classifier is a good classifier and vigorous to dimensionality profanity. One advantage of SVM is to realize regression optimization throughout the learning and the testing operations. As well, the SVM structure can be developed using margin, kernel type and duality characteristics domain. But, there are various problems as local minima and non-linear, which can discriminately distinguish between classes and correctly separate them using SVMs. The decision of SVMs relies on relevance score to determine the highest score among all classes. We can grade the retrieved image using fuzzy membership to determine relevance score that used as a measure to build a graded ground truth [28]. The ground truth is constructed in binary label without manual energy. The fuzzy score is either negative or positive class posterior probability. It is based on the function of fitting sigmoid in which it is obtained after calculating the decision value of $f(x_s)$ for each x_s learning samples (Eq. 9).

$$\sigma_s = \frac{1}{1 + \exp(af(x_s) + b)} \quad 9$$

It is being noted that the two parameters a and b are adapted depend on the learning dataset. Also, the learning sample which either negative or positive is sorted due to the fuzzy relevance scores. Here, multi-SVM classifier is considered to correctly treat with multiclass problem where the classes are associated to their increasing relevance score. The motivation behind using multi-SVM algorithm is to decrease the time complexity because of constructing binary SVMs. Thus, each binary classification can be relocated to multi-class easily using relevance scores σ_s .

3. Proposed Approach

The proposed approach includes on two major processes; learning and retrieving as in Fig. 3. In learning process the shape and geometric feature and normalized and employed to SVMs to construct ontology knowledge base of mammal's dataset. Retrieving process performs Semantic Image Retrieval (SIR) based on the processes of Query Engine, Matching Module and Ontology Manger.

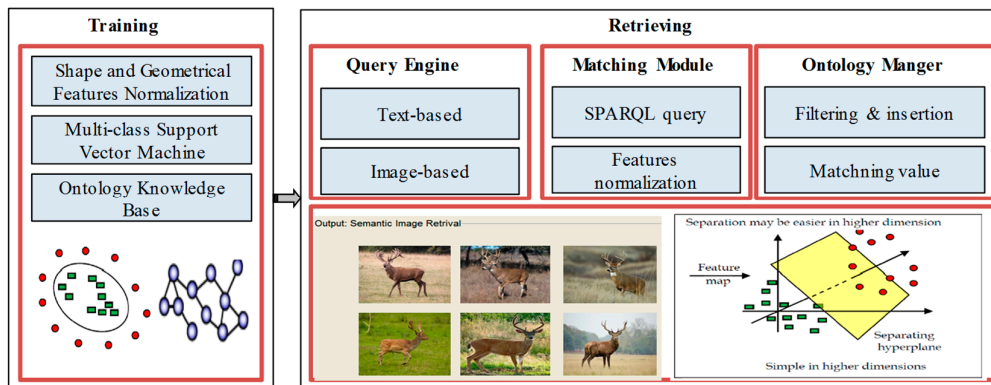


Fig. 3. Text and Semantic Image Retrieval Approach.

3.1 Query Engine

From Fig. 3, the first process of retrieving semantic image is called Query Engine (QE). Two different ways can be used for input; text-based or image-based methods. In text-based

method, the user used the system interface SIR to perform the search process by a text input. This process is typically frozen by current search engines as Yahoo, Alta Vista Google, Bing, etc. The chief impetus is to provide likelihood to users to inevitably learn and interact with SIR interface. For example, the text query as Giraffe, Deer, Ox, Tiger etc. is entered via the user. Then it is directly going to QE as Text-based, which is blamable to build the input text query. The second SIR input is named image-based method, which holds objects and some optional options for description. The key impetus is to run a new dimension for flexible searching. The QE construct the query using SPARQL language related to the input image of Ontology Knowledge base. Here, the object features are extracted using Zernike moment, circularity and rectangularity, which switched to high-level ontology features. After that, SPARQL generates the object parameters. Shortly speaking, the high-level ontology features is extracted semantic image as;

Query. Find the image of mammals with $(z_{00}, z_{11}, z_{22}, Cir_{norm}, Rect_{norm})$
SPARQL FROM: SELECT?x?y
WHERE {?y rdfs:subClassOf:mammals.? x: z₀₀, z₁₁, z₂₂, Cir_{norm}, Rect_{norm}

3.2 Matching Module

The second process of retrieving semantic image is named Matching Module (MM) as in Fig. 3. Based on ontology knowledge base, the input of this process is SPARQL query which obtained by QE while the output is to Ontology Manger process in case of successful search. The matching process carries out three major processes in case of failure to retrieve the relevant images. First, MM surfs for relevant images using query search engine as Google and then the gaining images are conceded to processing module to do their content verification (Fig. 3). Furthermore, the obtained images are checked against relevant to user query. If not they are verified based on the features of Zernike moment, circularity and rectangularity, which switched to high-level ontology features. Finally, the query SPARQL is built relied on the high-level ontology and ontology knowledge base. Thus, we considered the retrieved relevant images when they match the user search query; else they are discarded.

3.3 Ontology Manager

The final process of retrieving approach in Ontology Manger (OM), which carries out the tasks of filtering, insertion and ranking. OM process filter the obtained relevant images using properties, class and instance of ontology knowledge base. After that it inserts image resultant in ontology knowledge base, which is constructed using semantic description for retrieved images. Finally, the ranking task runs the image ranking relied on matching value, which calculated by a summation of matched ontology features and user query reference. To obtain the results, the resultant images are sorted in descending order based on matching value and the higher ranking is selected as a user request. For further specifics, the person who reads can state to [27].

3. Experimental Results

To motivate our work, we decided to build data set for representative mammals. The ontology knowledge base of this data set contains various images of 50 frames for a mammal. We have 25 various mammals as Lion, Horse, Deer, etc. Fig. 4 depicts nearby limited images from mammal's dataset. To provide neutral estimation for SVMs classifier, we divided the dataset into two thirds for learning and one third for testing processes. Take in account, the sample for learning process is completely unlike testing process.



Fig. 4. Sample learning images of mammal's database.

To retrieve relevant semantic image either by text or image, the system interface (SIR) is projected in Matlab language (Fig. 5). The SIR system provides good promising results to retrieve semantic images with tested the input images or input texts. In Fig. 5, the user carries out the SIR interface by browsing the horse image as input. Then, the QE constructs the image's query in conjunction with ontology knowledge base. When the response of SIR is positive, the search in web images is overlooked and the processes of filtering and ontology well done. The OM ranks the semantic resultant images and shows high ten images in descending order. Uncertainty the SIR response is negative, the searching takes place in web images and then the filtering process and updating ontology contents process are carried out to provide the result.

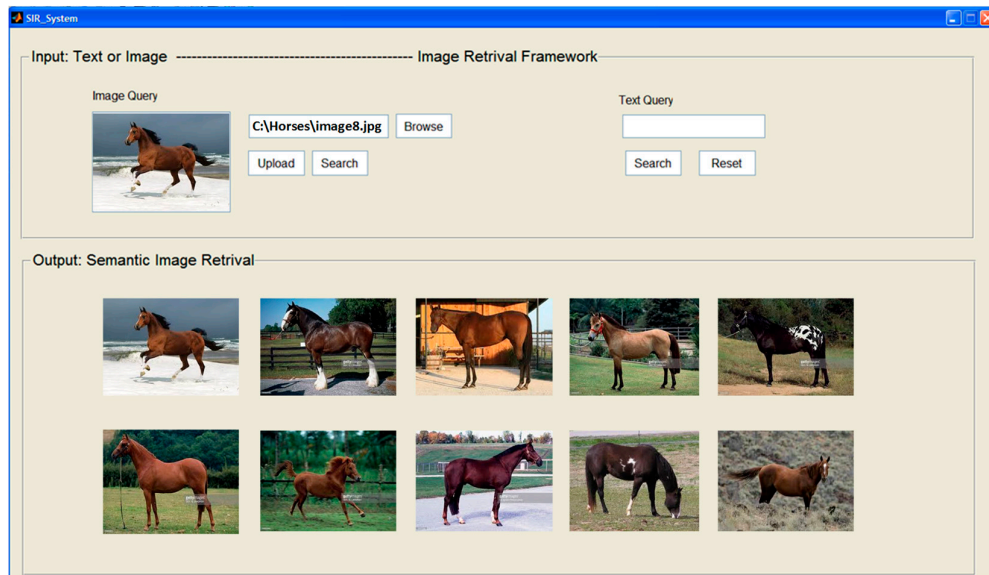


Figure 5. Retrieving the relevant horse images relied on ontology base.

Additionally, we can retrieve the relevant images semantically using the text as input. In Fig. 6, the user carries out the SIR interface by writing 'Deer' as text in the text query bottom of SIR system interface. Then, the QE constructs the corresponding query of 'Deer'. Thus, the relevant Deer images will be shown and then marked due to ontology knowledge base and OM. We use Precision and Recall to evaluate our proposed system. They use metrics in International Relations World (IRW). From all ontology contents images, Precision measure the system performance to retrieve the relevant images (Eq.10). It is also called true positive. Formally speaking, the Precision does not afford all concrete information around the system performance since it is not reflecting all images, which retrieved. Moreover, in Eq. 11, Recall is to retrieve the relevant images via the total number of associated images that have been retrieved. It is being noted that, the Recall is named false negative in which it is not considering the retrieving distinct images. In Fig.7, the Recall in addition to the average Precision of our proposed system is displayed, in which

x-axis is to the 415 images tests of 25 gathering classes, and y-axis refers to the Recall and the average Precision which computed using Eq. 10 and Eq. 11, respectively.

$$precision = \frac{true\ positive}{true\ positive + false\ positive} \quad (10)$$

$$Recall = \frac{true\ positive}{true\ positive + false\ negative} \quad (11)$$

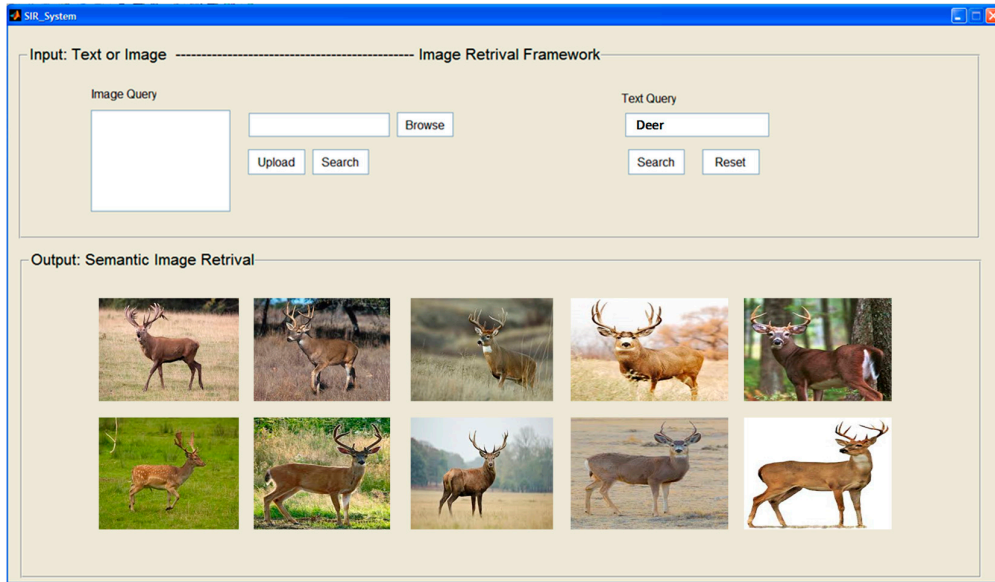


Figure 6. Retrieving relevant Deer images by the input 'Deer' in text query bottom.

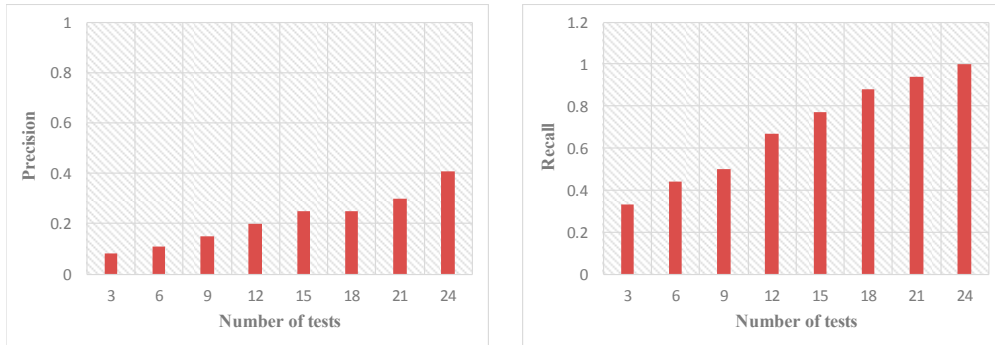


Figure 7. Precision and Recall matrices of SIR system.

From fig. 7, it is observed that the value of Precision/Recall is in the range of concluded 0.94/0.41. Furthermore, the higher value of Precision/Recall lies in the range of (0.08 over 0.41)/(0.33 over 1.0), respectively. The system can retrieve semantic images with 96.33 % retrieval accuracy.

Table. 1 Comparison with our previous work

Method	Feature type	Classifier	Retrieving rate
Our method	Shape and geometric	Multi-SVMs	96.33 %
Previous method [27]	CLF and color features	CRFs	93.50 %

To measure the effectiveness of our proposed system, the gained results have been compared with my previous work as illustrated in Table 1. By the comparison, our system using Multi-class SVM achieves competitively with [27] and its result equated favorably. In addition, we compared our system with have used similar dataset and experimental setups. So, the comparison is significant.

4. Summary and Conclusion

In this paper, a new approach to retrieve semantic images based on multi-class support vector machine is proposed. The shape feature of Zernike moment is to verify the invariance of objects for silhouette image. Additionally, a set of geometrical features explores the objects shape using two features of rectangularity and circularity. After that the extracted features are normalized to multi-class SVM either for learning or retrieving processes. The retrieving process is based on three key tasks which so-called; Query Engine, Matching Module and Ontology Manger, respectively. Benchmark mammals of 25 different mammals of 50 images had been conducted to empirically conclude the outcome SIR system. Our experiment on text and image retrieval yields efficient results with 96.33% retrieval accuracy than previously reported.

References

- [1] T. Wang and W. Wang, "Research on New Multi-Feature Large-Scale Image Retrieval Algorithm based on Semantic Parsing and Modified Kernel Clustering Method", *International Journal of Security and Its Applications* Vol. 10, No. 1 (2016), pp.139-154.
- [2] J. varghese, T. B. Mary, "High level semantic image retrieval algorithm for Corel database", *IRF International Conference*, (2016), pp. 1-5.
- [3] J. Almazán, J. Revaud, and D. Larlus, "Deep Image Retrieval: Learning Global Representations for Image Search", *Computer Vision – ECCV*, Volume 9910 of the series *Lecture Notes in Computer Science*, (2016), pp 241-257.
- [4] M. Zand, S. Doraisamy, A. Abdul Halin, "Ontology-Based Semantic Image Segmentation Using Mixture Models and Multiple CRFs", *IEEE Transactions on Image Processing*, Vol. 25, No. 7, (2016), pp. 3223-3248.
- [5] M.S. Meharban and Dr.S. Priya, "A Review on Image Retrieval Techniques," *Bonfring International Journal of Advances in Image Processing*, Vol. 6, No. 2, (2016), pp. 7-10.
- [6] Y. Kleiman, G. Goldberg, Y. Amsterdamer, D. Cohen-Or, "Toward semantic image similarity from crowdsourced clustering," *The Visual Computer*, Volume 32, Issue 6, (2016), pp 1045–1055.
- [7] K. Kumar, Z. ul-abidin, J. Ping Li, and R. A. Shaikh, "Content Based Image Retrieval Using Gray Scale Weighted Average Method," *International Journal of Advanced Computer Science and Applications*, Vol. 7, No. 1, (2016), pp. 1-6.
- [8] T. Dacheng, W. Dianhui, and M. Fionn, "Machine Learning in Intelligent Image Processing", In *Signal Processing*, Vol. 93 (6), (2013), pp. 1399-1400.
- [9] Z. Zeng, "A Novel Local Structure Descriptor for Color Image Retrieval," *Information Journal*, Vol. 7, No. 1, (2016), pp. 1-14.
- [10] P. P. Mane, N. G. Bawane, "An effective technique for the content based image retrieval to reduce the semantic gap based on an optimal classifier technique," *Pattern Recognition and Image Analysis*, Volume 26, Issue 3, (2016), pp 597–607.
- [11] X. Wang, H. Yang, Y. Li, W. Li, and J. Chen, "A New Svm-Based Active Feedback Scheme for Image Retrieval", In *Engineering Applications of Artificial Intelligence*, Vol. 37, (2015), pp. 43-53

- [12] M. Zhang, K. Zhang, Q.Feng, J. Wang, J. Kong, and Y. Lu “A Novel Image Retrieval Method Based on Hybrid Information Descriptors”, Journal of Visual Communication and Image Representation, Vol. 25 (7), (2014), pp.1574-1587.
- [13] N. Ali, KB. Bajwa, R. Sablatnig, SA. Chatzichristofis, Z. Iqbal, and M. Rashid,” A Novel Image Retrieval Based on Visual Words Integration of SIFT and SURF,” PLoS ONE 11(6): e0157428. , (2016), pp. 1-20.
- [14] A. Al Azemi, S. Nefti, U. Manzoor, and Y.Rezgui,“Building a Bilingual Bio-Ontology Platform for Knowledge Discovery”, International Journal of Innovative Computing, Information and Control, Vol. 7(12), (2011), pp. 7067-7075.
- [15] S.Murala, Q.M. Jonathan Wu, “Expert Content-Based Image Retrieval System Using Robust Local Patterns”, Journal of Visual Communication and Image Representation, Vol. 25(6), (2014), pp. 1324-1334.
- [16] B. Ramamurthy and K.R. Chandran, “Content based Image Retrieval for Medical Images using Canny Edge Detection Algorithm”, International Journal of Computer Applications, Vol. 17(6), (2011), pp. 32 - 37, 2011.
- [17] S. Shirazi , N. A. Khan , A. I. Umar , M. I. Razzak , S. Naz and B. AlHaqbani,” Content-Based Image Retrieval Using Texture Color Shape and Region,” International Journal of Advanced Computer Science and Applications, Vol. 7, No. 1, (2016), pp. 418-426.
- [18] N. Rastogi and D. Chaudhary ,” Enhanced Semantic Image Retrieval using Feature Extraction and KNN Techniques,” International Journal of Computer Applications, Vol. 136, No.13, (2016), pp. 23-28.
- [19] T. Jasperlin, and D. Gnanadurai,” Proportional Analysis on Texture Image Retrieval Using Double Density 2D DWT and Double Density Dual Tree DWT, “International Journal of Applied Engineering Research 10, no. 1, (2015), pp. 37- 48.
- [20] F. Radenovic, G. Tolias and O. Chum,” CNN Image Retrieval Learns From BoW: Unsupervised Fine-Tuning With Hard Examples,” . In: ECCV, (2016), pp 3-21.
- [21] S. Zeng, R. Huang, H. Wang and Z. Kang,” Retrieval Using Spatiograms of Colors Quantized by GaussianMixture Models,” Neurocomputing, Vol. 171, (2016) , pp. 673–684.
- [22] N. Ali, KB. Bajwa, R. Sablatnig and Z. Mehmood,” Image Retrieval by Addition of Spatial Information Based on Histograms of Triangular Regions,” Computers & Electrical Engineering, Vol. 54, (2016), pp. 539–550
- [23] F. Long, H.J. Zhang, D.D. Feng, “Fundamentals of content-based image retrieval”, In D. Feng (Ed.), Multimedia Information Retrieval and Management, Springer, Berlin, (2003).
- [24] Y. Rui, T.S. Huang, S.-F. Chang, “Image Retrieval: Current Techniques, Promising Directions, and Open Issues”, Journal of Visual Communication Image Representation, Vol. 10 (4), (1999), pp. 39–62.
- [25] S.K. Chang, S.H. Liu, “Picture Indexing and Abstraction Techniques for Pictorial Databases”, IEEE Trans. on Pattern Analysis. Mach. Intelligence, Vol. 6 (4), (1984), pp. 475–483.
- [26] I.K. Sethi, I.L. Coman, “Mining Association Rules Between Low-Level Image Features and High-Level Concepts”, Proceedings of the SPIE Data Mining and Knowledge Discovery, Vol. 3, (2001), pp. 279–290.
- [27] M. Ahmad, and S.-W. Lee, “Human Action Recognition Using Shape and CLG-motion Flow from Multi-view Image Sequences,” Journal of Pattern Recognition, Vol. 41, (2008), pp.2237-2252.
- [28] V. N. Vapnik, “An Overview of Statistical Learning Theory”, IEEE Transactions On Neural Networks, Vol. 10, No. 5, (1999), pp. 988- 999.
- [29] Zohair Malki , “ Ontology-Based Framework for Semantic Text and Image Retrieval Using Chord-length Shape Feature”, International Journal of Multimedia and Ubiquitous Engineering, Vol.11, No.11 (2016), pp.179-188.



© 2017 by the authors; licensee *Preprints*, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons by Attribution (CC-BY) license (<http://creativecommons.org/licenses/by/4.0/>).