International Journal of Multimedia and Ubiquitous Engineering
Vol.11, No.11 (2016), pp.179-188
http://dx.doi.org/10.14257/ijmue.2016.11.11.15

Ontology-Based Framework for Semantic Text and Image Retrieval Using Chord-length Shape Feature

Zohair Malki

College of Computer Science and Engineering in Yanbu, Taibah University, Saudi Arabia
Zmalki@taibahu.edu.sa

Abstract

Despite the vast research amount on the analysis and retrieval sematic images, there are still significant challenges worthy of address. This paper proposes an ontology framework to analyze and retrieve text and image based semantic search. This framework can be described in three main processes. Firstly, the Query Engine process constructs the input image query in SPARQL language. Secondly, the process of matching module is to retrieve the most affined images based on the compliance with input query. This process extracts the shape features of image's objects via chord-length features. Furthermore, the ontology manger process inserts the new relevant object's features in ontology knowledge base. Finally, the ranking module process is to classify the images which displayed in descending ordered based on matching values. Our experiment on a trained benchmark mammals shows that the proposed framework is more vigorous and yields favorable results when applying auspicious with large number of tested images without sacrificing real-time performance.

Keywords*: Semantic image retrieval, Query Engine, SPARQL, Chord-length features, Hidden Conditional Random Fields*

1. Introduction

In the last decade, the retrieval of semantic digital image is one the promising research field where many researchers investigate approaches either for image analysis [1-4] or image retrieval [5-8]. In general, the approaches of image retrievals usually based on text meta-data of keyword in which the retrieval process well done rely on image's textual description [9-11]. At present, the semantic image is retrieved using keyword-based search. Whereas there is still many problem in Yahoo and Google search engines that because the lack of storing meaningful and relationship among images in the web [12]. For the natural image domain, an ontological database is constructed, in addition to store images with their features in semantic manner. For example; when the visual image's features are assigned as input to SPARQL query, the available images with their ontological structure will retrieve. In that case, the users are not perceptible of image visual features. So, many classier algorithms applied to perform the learning of these visual features, which used as an input query image.

Other approaches are employed and used for indexing and analyzing images rely on its visual content. To interpret the tenor images, the low-level features like shape, texture and color are generally used. On an opposite side, the high level techniques are employed to recognize and retrieve particular patterns via scanning the whole image. Thus the retrieval process will be more efficient by reducing the number of iterations. In the past decade, a few researches proceed on Image Retrieval (IR) according to content similarity. Additionally, a little experimental prototype approaches and commercial products have been conducted such as Virage [13], QBIC [14], SIMPLIcity [15], Netra [16], Photobook [17], and VisualSEEK [18]. Futhermore, CBIR is referred as comprehensive surveys by

ISSN: 1975-0080 IJMUE

[19-20]. S.K. Chang and S.H. Liu propose an abstraction approach to index and retrieve images with respective to database retrieval [21]. Here, the images are indexed via its visual content like texture, color and shape. Liu *et al.* [22] conducted a survey, which display most semantic-based image retrieval techniques as a test data, as well as it perfuse various amount of portions in this field. There are many approaches that retrieve images based on either text-based or content-based. The essential difference between the two retrieval approaches of text-based and content-based is that the Human Interaction (HI) represents irreplaceable portion of latter approach. Furthermore, the humans resort to employ the text descriptors and keywords as high-level features in order to measure and interpret the similarity of images. Otherwise most features that automatically extracted via the techniques of computer vision represent low-level features as shape, texture, color, and spatial layout. In comprehensive, the forthright link is not exists between the low-level and high level in features and concepts [23].

The main contribution of used approach is to investigate an ontology framework for semantic-based image retrieval using Chord-length Shape feature of image in conjunction with conditional random field. This framework is based on four main modules: Query Engine (QE), Matching Module (MM), Ontology Manger (OM) and Ranking Module (RM). The experiment on the proposed framework favorable results when applying auspicious with large number of tested images. Our paper is organized as follows; Section 2 gives a detailed attributive of Chord-length Function and conditional random field classifier. In Section 3, the suggested approach is mentioned. After that, the experimental results are discussed in Section 4. Section 5 ends with a few concluding remarks.

2. Related Literature

There is no doubt that a good selection for the classification approaches assists the success and makes any system capable for real-world applications. In this paper, Semantic images are retrieved according to discriminative models like CRFs, which enforce the vigorous view-invariant task. So, this section is important in the context of understanding the motivation of doing the research and enables to investigate the novel techniques. The following two subsections briefly review the statistical chord-length function and the Conditional Random Fields classifier.

2.1. Chord-length Function

The extracted features by the shape function of chord-length are constructed by *1-D* chord-length. Formally speaking, we define the contour ∂ of a *2-D* shape as consecutive sequence of coordinates *N* points [24, 25];

$$\partial = \{z_t = (x_t, y_t) \in R^2 | t = 0, 1, \dots, N-1\} \tag{1}$$

where $z_t + N = z_t$ as ∂ is closed. The diameter *D* of the shape is given by;

$$D = max_{i,j=0}^{N-1} \|z_i - z_j\|, i \neq j \tag{2}$$

where $\|.\|$ is determined as an Euclidean distance for two points z_i and z_j. Suppose that the point $z_i \in \partial$ represents the start such that the contour ∂ be visited in anti-clockwise where it can be divided into $k > 1$ sections with equal length (*i.e.*, $\widehat{z_i s_1}, \widehat{s_1 s_2}, \dots, \widehat{s_{k-1} z_i}$). Here, s_j represents the j^{th} division point as well as $1 \leq j < k$. Thereby, the contour has *k-1* chords with following lengths;

$$L_1^{(i)}, L_2^{(i)}, \dots, L_{k-1}^{(i)} \tag{3}$$

where $L_j^{(i)}$ refers to the chord $\widehat{z_i s_j}$ length, which is determined as an Euclidean distance for the two points s_j and z_i (Figure 1). Let us now show that while the point z_i moves within the shape's contour, the chord lengths of $L_j^{(i)}$ be different as, accordingly. The presupposes of $L_j^{(i)}$ represents a function of z_i. Here this function is called the Chord-length Function (CLFs), which shortly denoted as $L_j^{(i)}$. Thus *k-1* CLFs can be obtained (*i.e.*, L_1, L_2,..., L_{k-1}). Furthermore, they are invariant to rotation and translation that because these functions are determined by dividing the contour moderately according to the moving the start point z_i, within the contour. Then, the chord length is normalized using the contour diameter D in order to be invariant in scale. Additionally, CLFs is scaled in the range [0-1] to obviously verify the key requirement of the shape descriptor. From this definition, we can determine CLFs by segmenting the contour equally. In short, we can say that it is facile to conclude that the half of CLFs, L_1, L_2,..., $L_{k/2}$ are only sufficient to characterize the shape acceptably. As a result, the local and the global features of the shape can be apprehending according to the different levels of chord-lengths. Thereby, it is viewed as a distinct competitive advantage of the CLF-based descriptor over other shape descriptors.

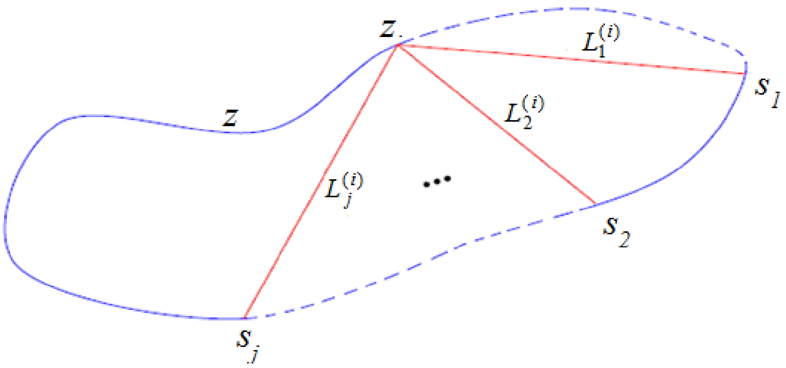

Figure 1. An Example of Clfs Obtained through the Division of a Shape Contour to a Arcs with Equal Length

2.2. Conditional Random Fields Classifier

The discriminative Conditional Random Fields (CRFs) is considered as a model of un-directed graph, which were sophisticated to label sequential data [26]. It is being noted that there is a trade-off in weights for each feature function that because CRFs use a single exponential distribution for a given observation to model all reference states/labels (Figure 2) [27]. Formally speaking, for a given an observation sequence O, the probability of label sequence y is estimated as;

$$p(y|O,\theta) = \frac{1}{Z(O,\theta)} \cdot exp\left(\sum_{i=1}^{n} F_\theta (y_{i-1}, y_i, O, i)\right) \tag{4}$$

Such that; the parameter θ is equal to $\theta = (\lambda_1, \lambda_2, ..., \lambda_{Nf}\,; \mu_1, \mu_2..., \mu_{Ng})$, N_f refers to the transition feature function number, N_g is to the label feature function number and n represents the length of the observation sequence O. Thus, F_θ is computed as;

$$F_\theta(y_{i-1}, y_i, O, i) = \sum_f \lambda_f t_f (y_{i-1}, y_i, O, i) + \sum_g \mu_g s_g (y_i, O, i) \tag{5}$$

where $t_f(y_{i-1}, y_i, O, i)$ represents a function of transition feature at position i-1 and i. $s_g(y_i,O,i)$ is to a function of a label feature at position i. λ_f as well as μ_g are the transition weights and the function of label feature, respectively. Furthermore, $Z(O, \theta)$ represents the normalized factor such that it is computed by;

$$z(o,\theta) = \sum_y exp\left(\sum_{i=1}^{n} F_\theta (y_{i-1}, y_i, O, i)\right) \tag{6}$$

Furthermore, the classifier CRFs are based on the BFGS optimization technique, which employed to perform the training process (*i.e.*, verify optimal convergence) using gradient ascent [27].

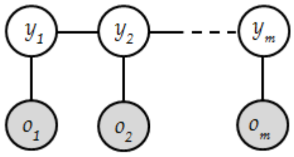

Figure 2. CRF Model in which O_j Refers to the J^{th} Corresponding Observation Value. Y_j Represents the State of O_j. h, The Gray Circles are to the Observed Variables

3. Suggested Approach

An ontology based-approach is to describe the main modules and their interaction with each other. The proposed approach contains two main modules; training and testing as in (Figure 3). Shortly, the testing processes which employed for Semantic Image Retrieval (SIR) consist of Query Engine, Matching Module, Ontology Manager and Ranking Module. The following subsections will describe the SIR approach in some details.

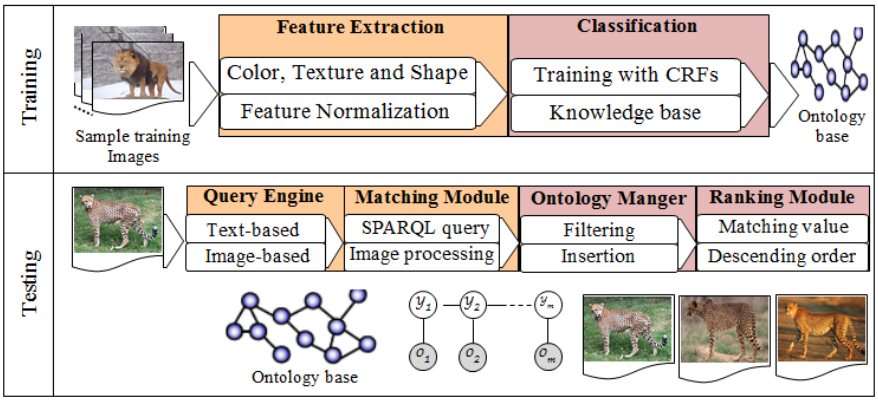

Figure 3. Suggested Approach for Text and Semantic Image Retrieval

3.1. Query Engine (QE)

In our approach, QE represents the first module and is used as a web interface from SIR. The input of this module can be expressed by two different ways for a user. Text-based method is the first way in which the user inters the specific input as a text with its information to directly perform the search process. This process is usually treated by the current search engines as Google, Alta Vista, Bing, Yahoo *etc*. The main motivation is to give possibility to users to automatically interact and learn with SIR interface. Furthermore, the text query (*e.g.*, Lion, Giraffe, Fox *etc*.) is written by the user and is directly passed to the Query module of Text-based, which is answerable for constructing the query of the input text. Here, it is being noted that all stemmer words (*i.e.*, standard stop-list) as: ―is‖, ―on‖, ―the‖, ―or‖ are deleted from input text as a first step. Then we generate the SPARQL query with all possible ―OR‖ and ―AND‖ to further process by Matching Module.

Image-based method represents the second SIR input way in which the input image contains a single object or multiple objects as well as some optional options for image description. The main motivation is to provide flexibility and a new dimension for searching. As the same way, the QE build the query by SPARQL language from the input image based on Ontology Knowledge base. Shortly, the features extraction of objects from an input image is extracted by using shape-based of Chord-length Function as previously described. Additionally, color-based feature extraction technique is employed to estimate the color pixel value of object detection. Then the texture of the object is identified using texture classification technique. As a result, these features so-called low-level features are converted into high-level ontology features. For the case of image description via the user, it is also switched into high-level ontology concepts. Thus SPARQL is ready to generate all object parameters. For example, the following query is employed to extract the semantic image by high level ontology features.

Query. **Find the image of mammals with yellow color**
SPARQL FROM: SELECT ?x?y
WHERE {?y rdfs:subClassOf: mammals. ?x: Color "yellow"}

3.2 Matching Module (MM)

The main motivation behind using matching module is to retrieve most related image rely on the same Ontology Knowledge base. Shortly, the input of this process is SPARQL query while the output of retrieved images is passed to ranking module in case of successful search. In case of failure to retrieve the relevant images from building ontology knowledge base, the matching module carries out three main processes. First, MM surfs on the internet for relevant images using query search engine as Yahoo or Bing or Google. After that the obtaining images are passed and employed to image processing module to perform their content verification (Figure 3). The image processing module checks if these images are pertinent to user query. If not, these images are verified and their entire objects are detected based on the extracted features of shape Chord-length, colored-based and texture-based techniques. Second, the extracted features (*i.e.*, low level ontology features) are switched to the high-level ontology features. Third, the query SPARQL is constructed based on the high level ontology contents in conjunction with ontology knowledge base. As a result, the retrieved relevant images are considered if they match user search query otherwise they are called non-relevant images and then discarded.

3.3 Ontology Manager (OM)

The main motivation of OM is to perform two processes filtering and insertion. After obtaining relevant images from the process of MM by surfing on the internet, the ontology manger filters these images using class, properties and instances of the ontology knowledge base (Figure 4). Then it inserts them into the ontology knowledge base. Here, the ontology (metadata) is constructed for the domain of mammal's image in which the images ontology has two components; class hierarchy and textual domain description. Thereby, the textual domain is supplementary spilt into visual text and text description. Next, the ontology knowledge base is built using the semantic description form retrieved images.

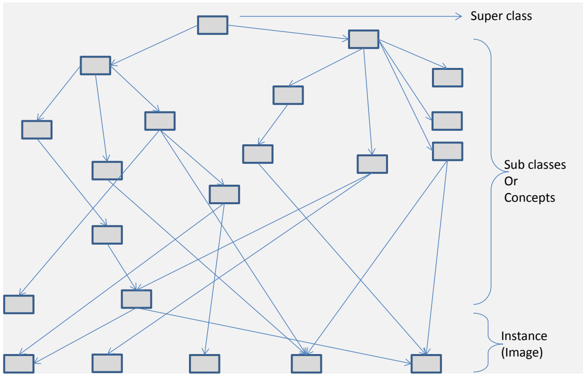

Figure 4. Partial Structure of Ontology Knowledge Base Containing Class and Instances

3.4. Ranking Module (RM)

Based on the resultant image with user query using OM module, the ranking module performs the image ranking based on the matching value. Here, the matching value is computed as a summation for the matched ontology features and the user query reference. As a result, the RM module sorts the resultant images in descending order based on matching value. After sorting, the system will showed the higher ranking images as a user request.

4. Experimental Results

We decided to construct realistic mammal's dataset in which the ontology knowledge base includes on images of 25 various mammals of 50 pictures for a mammal. Figure 5 illustrates some partial images of mammal's dataset.

Figure 5. Sample Training from Mammal's Dataset

Our proposed approach to retrieve Semantic Image Retrieval is trained or tested on this dataset in addition to surf on the Internet for testing occasionally. To give an unbiased estimation for the CRFs classifier, the dataset was splitted into two thirds; one for training and one third for testing. The sample of training images is completely different than testing data, where these processes is performing on *Intel(R) CPU 3.4 GHz Core(TM) i7 PC with 8 GB of RAM*. In our work, the parameters of CRFs are learned by BFGS optimization technique with 200 iterations in conjunction with gradient ascent to converge. Here, it being noted that, the training process to build ontology contents is more costly on a standard desktop PC with mentioned specification above. Contrariwise the testing is fasting because of using forward score to retrieve the label with the highest likelihood. The system interface to retrieve relevant semantic image is design in Matlab language.

Our experiment was provided promising results when applying auspicious with large number of tested images. The proposed system has the ability to retrieve images either by their contents or text name. In Figure 6, the user enters the browse of giraffe image as input to our system so-called Semantic Image Retrieval (SIR). After that, the Query Engine builds its queries and carries out on ontology knowledge base. If the response is positive (*i.e.*, knowledge base contains the resultant images), the processes of filtering, ontology updating and web image search are ignored. Furthermore, the Ranking Module ranks the resultant images and displays the higher ten images in descending order due to the SIR system specification. If the response is negative the searching are in the web image taking in consideration the filtering and updating ontology contents processes.

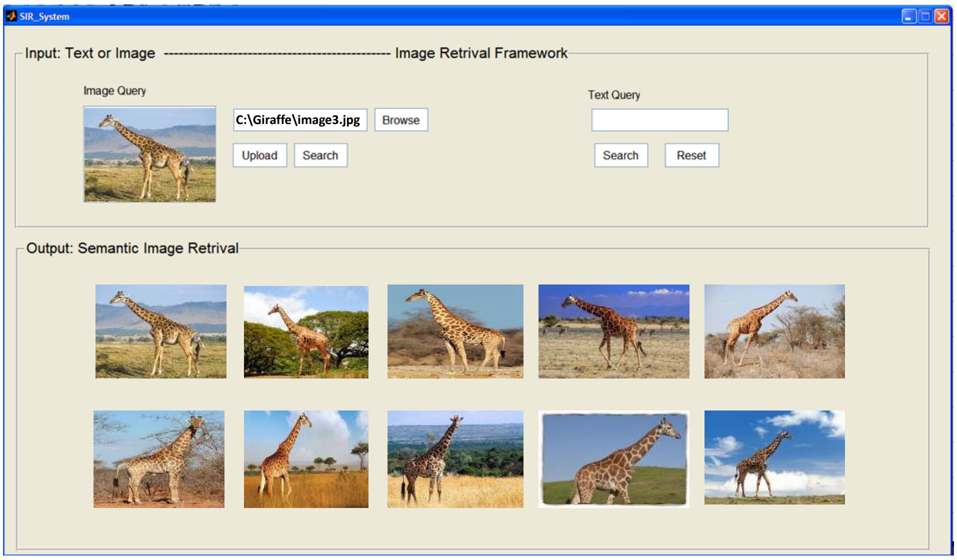

Figure 6. Retrieving the Relevant Giraffe Images Based on Ontology Knowledge Base

Otherwise relevant images can be semantically retrieved using their text as an input to SIR system. As in Figure 7, the Query Engine builds the corresponding query of the input text "Primate" using the SIR system. Thereby, the relevant images will be displayed and ranked related to the knowledge base and Ranking Module.

To evaluate the proposed system, Precision and Recall which commonly used metrics in international relations world are considered. Here precision measures how accurately the system retrieves the relevant images (*i.e.*, true positive) from all images of ontology contents (Eq. 7). It is being noted that precision does not provide all actual information about the system performance because of not considering all images that retrieved. Whereas, Recall is retrieving the relevant images from the total number of related images which should have been retrieved (Eq. 8). In similar, Recall has not considered the retrieving of unrelated imaged (false negative).

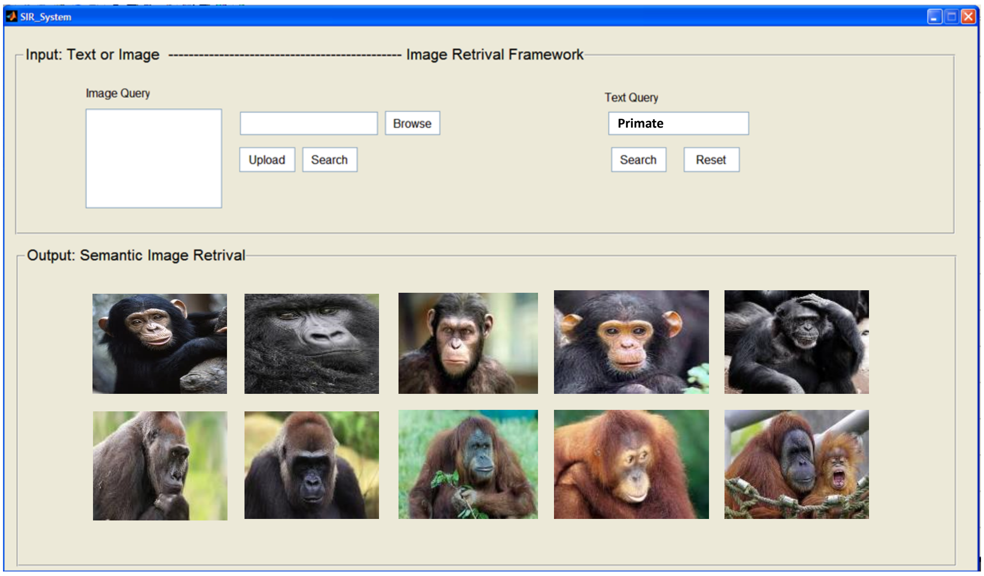

Figure 7. Retrieving Relevant Images of Primate by Text Name as Input

$$precision = \frac{true\ positive}{true\ positive + false\ positve} \tag{7}$$

$$Recall = \frac{true\ positive}{true\ positive + false\ negative} \tag{8}$$

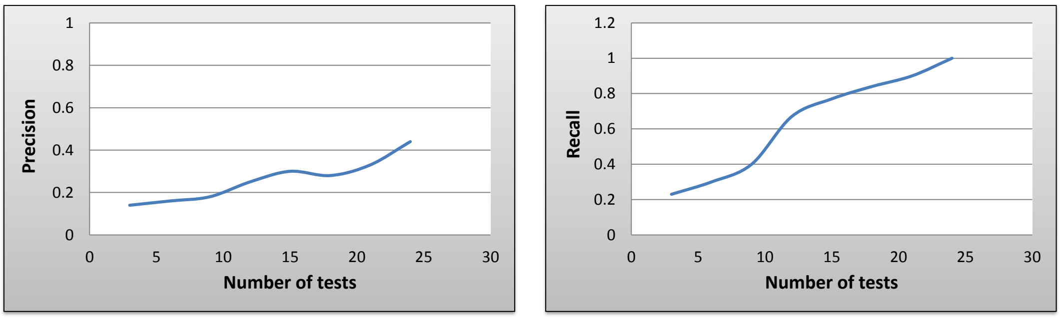

Figure 8. Precision and Recall Matrices for Semantic Image Retrieval System

As a result, the recall as well as average precision on the proposed semantic image retrieval system is shown in Figure 8. Here, x-axis refers to the number of tests with 415 images of 25 collection classes, while y-axis represents the average precision and recall that estimated via Eq. 7 and Eq. 8 respectively. It is being noted that the value of recall/precision tale in range of over 0.95 and 0.45 respectively. Additionally, the higher values of recall/precision be in range of (0.6 over 1.0)/(0.2 over 0.45) respectively. Thus the system has the ability to retrieve relevant images with from various collections with 93.5% retrieval accuracy. The time complexity of CRFs classifier represent $O(NL^2)$ where N refers to the number of input features and L represents the number of labels such that each label is assigned for a class image. In offline mode, the space complexity to carry out our experiment is nearly the same of time complexity. As for good election for features, CRFs initialization parameters and a high accuracy of object segmentation, the high accuracy rate of relevant images accuracy was achieved.

4. Summary and Conclusion

This paper proposed an ontology framework to retrieve text and image based on a semantic search. The shape features of image's objects are extracted via chord-length features as well as color and texture of objects are considered too. The Query Engine process constructs the input image query in SPARQL language. After that, the process of matching module retrieves the most affined images relied on the compliance with input query. Furthermore, the ontology manger process inserts the new relevant object's features in ontology knowledge base and ranks the resultant images in descending ordered rely on matching values. Our experiments on a realistic mammal's dataset of 25 various mammals of 50 pictures for a mammal were carried out. It yields favorable results when applying auspicious with the considerable number of tested images of 93.5% retrieval accuracy.

References

[1] J. Shotton, J. Winn, C. Rother, and A. Criminisi, "Texton Boost for Image Understanding: Multi-Class Object Recognition and Segmentation by Jointly Modeling Texture, Layout, and Context", International Journal of Computer Vision, vol. 81, no. 1, **(2009)**.

[2] J. Shotton, A. Blake, and R. Chipolla, "Contour-Based Learning for Object Detection", In Proceeding of International Conference on Computer Vision, **(2005)**.

[3] V. Petert, and T. Kristof, "Adaptive and Optimal Difference Operators in Image Processing", Pattern Recognition, vol. 42, no. 10, **(2009)**, pp. 2317-2326.

[4] T. Dacheng, W. Dianhui, and M. Fionn, "Machine Learning in Intelligent Image Processing", In Signal Processing, vol. 93, no. 6, (**2013**), pp. 1399-1400.

[5] T. Quack, U. Monich, L. Thiele, and B. S. Manjunath, "Cortina: A System for Large-Scale, Content-Based Web Image Retrieval", In Proceedings of ACM international conference on Multimedia, **(2004)**.

[6] N. Alajlan, M. S. Kamel, and G. H. Freeman, "Geometry-Based Image Retrieval in Binary Image Databases", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 30, no. 6, **(2008)**, pp. 11003-11013.

[7] 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

[8] 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. no. 257, **(2014)**, pp. 1574-1587.

[9] Y. Liu, D. Zhang, G. Lu, and W.-Y. Ma, "A Survey of Content-Based Image Retrieval with High-Level Semantics", In Pattern Recognition, vol. 40, no. 11, **(2007)**, pp. 262-282.

[10] 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, no. 12, **(2011)**, pp. 7067-7075.

[11] S. Murala and Q. M. J. Wu, "Expert Content-Based Image Retrieval System Using Robust Local Patterns", Journal of Visual Communication and Image Representation, vol. 25, no. 6, **(2014)**, pp. 1324-1334.

[12] 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, no. 6, **(2011)**, pp. 32 - 37, **(2011)**.

[13] A. Gupta, and R. Jain, "Visual Information Retrieval", In Communication of ACM, vol. 40, no. 5, **(1997)**, pp. 70-79.

[14] C. Faloutsos, R. Barber, M. Flickner, J. Hafner, W. Niblack, D. Petkovic, and W. Equitz, "Efficient and Effective Querying by Image Content", Journal of Intelligent System, vol. 3, **(1994)**, pp. 231-262.

[15] J. Z. Wang, J. Li, and G. Wiederhold, "Simplicity: Semantics-Sensitive Integrated Matching for Picture Libraries", In IEEE Trans. of Pattern Analysis and Machine Intelligent, vol. 23, no. 9, **(2001)**, pp. 947-963.

[16] W. Y. Ma, and B. Manjunath, "Netra: A Toolbox for Navigating Large Image Databases", In Proceedings of the IEEE International Conference on Image Processing, **(1997)**, pp. 568-571.

[17] A. Pentland, R.W. Picard and S. Scaroff, "Photo book: Content-Based Manipulation for Image Databases", International journal of Computer Vision, vol. 18, no. 3, **(1996)**, pp. 233-254.

[18] J. R. Smith, and S. F. Chang, "Visual Seek: A Fully Automatic Content- Based Query System", Proceedings of the Fourth ACM International Conference on Multimedia, **(1996)**, pp. 87-98.

[19] F. Long, H. J. Zhang and D. D. Feng, "Fundamentals of content-based image retrieval", In D. Feng (Ed.), Multimedia Information Retrieval and Management, Springer, Berlin, **(2003)**.

[20] Y. Rui, T.S. Huang and S.-F. Chang, "Image Retrieval: Current Techniques, Promising Directions, and Open Issues", Journal of Visual Communication Image Representation, vol. 10, no. 4, **(1999)**, pp. 39-62.

[21] S. K. Chang and S. H. Liu, "Picture Indexing and Abstraction Techniques for Pictorial Databases", IEEE Trans. on Pattern Analysis. Mach. Intelligence, vol. 6, no. 4, **(1984)**, pp. 475-483.
[22] Y. Liu, D. Zhang, G. Lu, and W. Ma, "A Survey of Content-Based Image Retrieval with High-Level Semantics", Journal of Pattern Recognition, vol. 40, no. 1, **(2007)**, pp. 262-282.
[23] 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.
[24] S. Sadek, A. Al-Hamadi, B. Michaelis, and U. Sayed, "An SVM Approach for Activity Recognition Based on Chord-Length Function Shape Features", IEEE Conf. on ICIP, **(2012)**, pp. 765-7680.
[25] C. Shi, and B. Wang, "Shape Retrieval Using Statistical Chord-Length Features", Advances in Image and Video Technology, First Pacific Rim Symposium, **(2006)**, pp. 403-410.
[26] J. Lafferty, A. McCallum, and F. Pereira, "Conditional Random Fields: Probabilistic Models for Segmenting and Labeling Sequence Data", Conf. on ICML, **(2001)**, pp. 282-289.
[27] H. Yang, S. Sclaroff, and S. Lee, "Sign Language Spotting With a Threshold Model Based on Conditional Random Fields", IEEE Trans. on PAMI, vol. 31, no. 7, **(2009)**, pp. 1264-1277.

Author

Zohair Malki, Working at the Department of Information Science, The collage of Computer Science and Engineering in Yanbu, Taibah University, Saudi Arabia. Currently he is the dean of the collage of Computer Science and Engineering in Yanbu, his research interests include, E-learning and education techniques; using ICT in education, computer technology, and computer software and algorithms.

