Retrieving Semantic Image Using Shape Descriptors and Latent-dynamic Conditional Random Fields

Mahmoud Elmezain[1,2], Hani M. Ibrahem[1,3]

[1]Faculty of Science and Computer Engineering, Taibah University, Yanbu, KSA
[2]Computer Science Division, Faculty of Science, Tanta University, Tanta, Egypt
[3]Mathematics & Computer Science dept., Faculty of Science, Menoufiya University
Corresponding author: M. Elmezain (mmahmoudelmezain@taibahu.edu.sa).

Abstract
This paper introduces a new approach to semantic image retrieval using shape descriptors as dispersion and moment in conjunction with discriminative model of Latent-dynamic Conditional Random Fields (LDCRFs). The target region is firstly localized via the background subtraction model. Then the features of dispersion and moments are employed to k-mean procedure to extract object's feature as second stage. After that, the learning process is carried out by LDCRFs. Finally, SPARQL language on input text or image query is to retrieve semantic image based on sequential processes of Query Engine, Matching Module and Ontology Manger. Experimental findings show that our approach can be successful retrieve images against the mammals Benchmark with rate 98.11. Such outcomes are likely to compare very positively with those accessible in the literature from other researchers.

Keywords:
Semantic image retrieval, Dispersion, Moment, SPARQL, Latent-dynamic Conditional Random Fields.

1. introduction

The image retrieval method is used to search, browse and collect pictures from an entire digital image dataset. Previously, most scientists concentrated on low-level appearance with a single function and no precision of retrieval, although the photo includes multiple visual characteristics. Previously, cooperative methods offer the concept of addressing the big gap between high-level and low-level perceptions [1, 2, 3, 4]. Texture, shape and color shape are mostly used as a low-level function in which the shape characteristics can be obtained using either boundary-based or region-based technique [5]. The boundary-based method is relied on the outer boundary; however, the region-based method relies on a high-level function to retrieve shapes in scanning the entire region. Using the mean of timely feature differences and choice [6], the gap between low-level and high-level characteristics is compacted. Semantic image retrieval is one of the entrusted research fields; such that numerous researchers study addressing techniques either to analyze image photos [7, 8, 9, 10] or to retrieve photos [11, 12, 13, 14]. It is observed that the semantic gap could be lowered by using the scene's more efficient function. This constitutes a challenge for studies on Content-Based Image Retrieval (CBIR) [15, 16]. Various machine learning methods are used to relieve this issue. In [17], A. S. N.Singha and K. Singhb analyzed photo contents in the real-time system by extracting color and shape characteristics. Their functions used the method of clustering C-Means to obtain and segment image boundaries. In addition, Fast Fourier Transform is used to afford the array features to the equivalent region. After that, a similarity matching algorithm relied on that set is functional to assess the resemblance of queried photo for retrieval. In [18], authors suggested a novel technique for determining the texture function and color feature of CBIR photos. In relation to a combination of color, texture and shape descriptors to obtain photos, they described this low-level feature using two color histogram function. In [19], the Fuzzy KNN classifier is studied for image retrieval. For dataset photos, the authors assign an original semantic state in which the assigned state to the image is altered using meaning feedback. Numerous measures are based on the similarity, hand-me-down three types of features. The authors in [20] studied CBIR's back propagation learning neural network. It is observed that the categorized images are split into background and foreground photos. The object region is given using a region-based segmentation method. In addition, the wavelet transform is to employ the texture feature for retrieving based-shape.

In [21], the authors correctly extract the features of image and then use Support Vector Machine (SVM) to efficiently obtain the required photo. In [22], a Riis photo is retrieved by using hierarchical methodology. This strategy is based on the iris dataset's pioneering indexing

Manuscript received October 5, 2024
Manuscript revised October 20, 2024
https://doi.org/10.22937/IJCSNS.2024.24.10.24

method. In addition, two dissimilar features are obtained for iris image content. The image texture is to retrieve the target photo using the indexed iris dataset where iris color is used to create photo indexing. Then the unlooked photo is filtered out by color features. Additionally, an image is filtered out without resemblance to the photo color question. This method is estimated for the raucous image of the iris. A new technique based on invariant moments algorithms, texture and color histogram is explored in [23]. They capture the image's regularity edges. Here, Gabor filter is used to obtain texture feature in relation to extracting shape feature using image color and time invariants. The significant drawback is due to low quality, which in turn makes it difficult to detect, recognize and retrieve tasks. In addition, the use of low camera resolution and environmental variables also create surveillance images of low quality.

The key contribution of this paper is to present a new approach for the retrieval of semantic images. It relies on dispersion, moment, ontology-base in conjunction with discriminative model of Latent-dynamic Conditional Random Fields (LDCRFs). The interested region is located using an adaptive Gaussian model to model the subtraction of the background. Then the features of dispersion and moments are employed to k-mean procedure to extract object's feature as second stage. After that, the learning process is carried out by LDCRFs. Finally, SPARQL language on input text or image query is to retrieve semantic image based on sequential processes of Query Engine, Matching Module and Ontology Manger. The result of our suggested approach is empirically evaluated against the Benchmark mammals. Our paper is organized as tracks; the suggested approach is discussed in three subsections for Section II. Section III performs our experimental outcomes. Section V concludes the suggested approach.

2. Proposed Approach

Our suggested approach includes three main phases; preprocessing, shape descriptor, and ontology basis. The subsequent subsections explain these phases as shown in Figure 1.

A. PREPROCESSING

Background subtraction method is an acceptably used to segment required patterns in a sequence scene. Every new image is contrasted to the constructed background model. It is worth saying that Gaussian Mixture Model (GMM) is a case of bigger brand for density models, which have various tasks through united components [24].

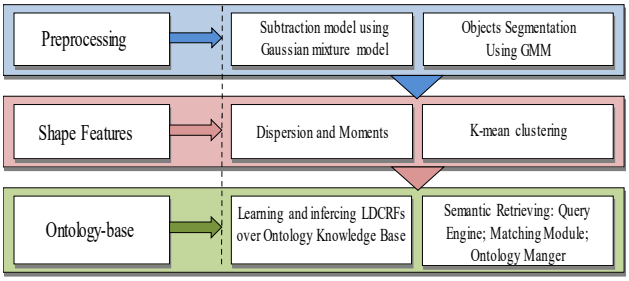

FIGURE 1. Semantic Image Retrieval Approach.

Properly speaking, let X_t be a pixel in the current image, as well as K is a distribution number. Then each pixel can be separately classified using Gaussian combination K as in Eq. 1.

$$p(X_t) = \sum_{i=1}^{K} \omega_{i,t} \cdot \eta(X_t; \mu_{i,t}; \Sigma_{i,t}) \tag{1}$$

Such that η is Gaussian's likelihood density function. The meaning, covariance and estimation of the prior likelihood (i.e. weighting function) of the i^{th} element, respectively, shall be $\mu_{i,t}$, $\Sigma_{i,t}$ and $\omega_{i,t}$. Furthermore, a constructive algorithm automatically chooses the the number of components using the requirements for maximizing a likelihood function [24]. Then the background is optimized based on minimizing error function E as in Eq. 2.

$$E = -\sum_{n=1}^{N} ln\Big(\sum_{i=1}^{K} \eta(X; \mu_i; \Sigma_i) \cdot \omega_i\Big) \tag{2}$$

Such that, N is to the number of data points X_n. The background distribution with lowest variance always runs on top using a threshold $\gamma = 0.5$. Consequently, all X pixels with no components are candidates for representation in the foreground. In addition, the EM technique is a Maximum Likelihood (ML) superior circumstance technique [25], so that the mixture model parameters fit the best for specified dataset in ML context. In Figure 2, using background subtraction method, the tiger pattern in the left image frame is segmented regarding the correct image source. The reader can refer to [26] for extra information.

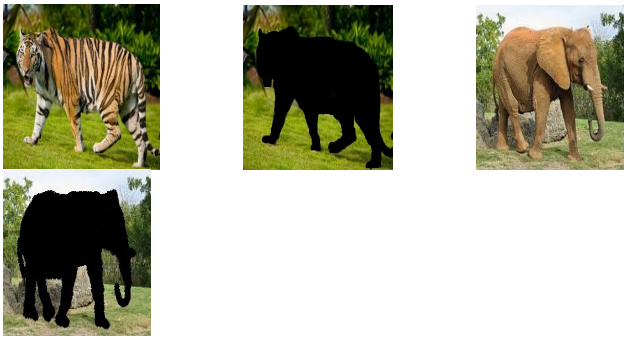

FIGURE 2. Outcome segmentation of tiger and Proboscidea based on the technique of background subtraction.

B. SHAPE FEATURES: DISPERSION AND MOMENTS

So far, we've focused on perimeter or shape descriptions. The natural counterpart is to define the area or the region using the description of shape's region. There are two primary challengers here that vary in motivation: fundamental regional descriptors that characterize the features of region's geometric, whereas moments focus on the region's density. First, though, we shall look at dispersion to describe regions. Moreover, Dispersion (irregularity) was evaluated as the proportion of foremost chord length to region [27]. An easy kind of this metric may be described as irregularity as shown in Eq. 3;

$$I(S) = \frac{\pi \max\left((x_i - \bar{x})^2 + (y_i - \bar{y})^2\right)}{A(S)} \tag{3}$$

where $(\bar{x}, \bar{y})$ is to the coordinates of the mass centre of region, which defined using the scalar measurements of the region geometric characteristics. The easiest assets are provided by size or region. In particular, the area of the region on a plane is described as by Eq. 4;

$$A(S) = \int_x \int_y I(x, y) \mathrm{d}y \mathrm{d}x \tag{4}$$

Such that $I(x, y)$=1 is in pixel shape, (x, y) is in S shape, and 0 has not been. Note that the numerator defines the maximum circle area of the region. This metric therefore describes the density of the region. As the ratio of the maximum to the minimum radius, an alternative dispersion measure can also be provided. This is an alternative irregularity form as in Eq. 5;

$$IR(S) = \frac{\max\left(\sqrt{(x_i - \bar{x})^2 + (y_i - \bar{y})^2}\right)}{\min\left(\sqrt{(x_i - \bar{x})^2 + (y_i - \bar{y})^2}\right)} \tag{5}$$

This measure describes the proportion between the maximum circle radius of the region and the base circle of the region. Thus, the measure will enhance as the region spreads. Thus, the irregularity of a circle is unity, IR (circle)=1; the irregularity of a curve is greater IR (square)= $\sqrt{2}$. As such, the measure increases on uneven forms, while compactness measurement decreases. Again, the measure is a metric of form for optimal shapes regardless of size. In fact, perimeter measurements will vary with rotation owing to the nature of discrete images and are more likely to be affected by noise than measurements of region (because measurements of region have inherent average features). Clearly, the measure of compactness and dispersion for the circle is near to unity. The compactness decreases as the ellipse dispersion increases. The condensed region has the lowest metric compactness and the largest values of dispersion. Clearly, these measurements can be used to characterize areas of distinct forms and thus discriminate between them.

Moments depict the layout of a form (the arrangement of its pixels), a bit like mixing region, compactness, irregularity and descriptions of higher order together. Hu [28] acquired a seven-time collection invariating translation, orientation and scale. The equations are calculated from second- and third-order moments. Under Maitra's photo comparison, Hu invariants are extended to be invariant [29]. Later, in particular affine transformation, Flusser and Suk [30] acquired the invariant moment that is invariant. The equations of Hu-Moments are defined as shown from Eq. 6 to Eq. 12;

$$M_1 = \mu_{20} + \mu_{02} \tag{6}$$

$$M_2 = (\mu_{20} - \mu_{02})^2 + 4\mu_{11}{}^2 \tag{7}$$

$$M_3 = (\mu_{30} - 3\mu_{12})^2 + (3\mu_{21} - \mu_{03})^2 \tag{8}$$

$$M_4 = (\mu_{30} + \mu_{12})^2 + (\mu_{21} + \mu_{03})^2 \tag{9}$$

$$\begin{aligned} M_5 = {} & (\mu_{30} - 3\mu_{12})(\mu_{30} + \mu_{12})[(\mu_{30} + \mu_{12})^2 - 3(\mu_{21} + \mu_{03})^2] \\ & + (3\mu_{21} - \mu_{03})(\mu_{21} + \mu_{03})[3(\mu_{30} + \mu_{12})^2 - (\mu_{21} + \mu_{03})^2] \end{aligned} \tag{10}$$

$$M_6 = (\mu_{20} - \mu_{02})[(\mu_{30} + \mu_{12})^2 - (\mu_{21} + \mu_{03})^2] + 4\mu_{11}(\mu_{30} + \mu_{12})(\mu_{21} + \mu_{03}) \tag{11}$$

$$\begin{aligned} M_7 = {} & (3\mu_{12} - \mu_{03})(\mu_{30} + \mu_{12})[(\mu_{30} + \mu_{12})^2 - 3(\mu_{21} + \mu_{03})^2] \\ & + (3\mu_{12} - \mu_{30})(\mu_{21} + \mu_{03})[3(\mu_{30} + \mu_{12})^2 - (\mu_{21} + \mu_{03})^2] \end{aligned} \tag{12}$$

Hu-Moments are derived from a seven-moment set. From second and third order moments, these seven moments are acquired. However, moments of zero and first order will not be used in this stage. The first six Hu-Moments are invariant to reflect and the seventh moment alters the shape of the goal [31]. The following set of characteristics are composed of statistical vectors (Eq. 13);

$$M_{stat} = (M_1, M_2, M_3, M_4, M_5, M_6, M_7)^T \tag{13}$$

Here, h_1 is to first Hu-Moment. In addition, the notation for all other characteristics in this collection is similar. Then Hu-Moments are normalized using next equation:

$$M_i = \frac{M_i - minM_{all}}{maxM_{all} - minM_{all}} \tag{14}$$

where M_i refers to i^{th} Hu-Moment feature. $minM_{all}$ as well $maxM_{all}$ represent the minimum and the maximum values with respect to all classes set, respectively.

After obtaining the shape's features, k-mean procedure is employed to quantize these features. The primary motive behind using k-mean is the Euclidean distance between all cluster points and cluster center point. Here, each point is assigned with one initialized cluster at first. Then the middle point of each is recalculated based on the mean points of the competent cluster. Therefore, until convergence, the procedures are repeated. In other words, K-means algorithm depends on two main phases: assignment phase and update phase. Each point is categorized into a particular cluster for the assignment phase, which has the nearest mean. Consequently, in update phase, the obtained mean point is considered to every cluster based on their instance. EM algorithm calculates the highest probable estimate iteratively for the unknown blend parameters during two phases [32]. Hence, in a motivating trait, the later phases are well ready to realize k-means.

C. ONTOLOGY-BASE

No one can disagree that the achievement of retrieving semantic image relies on an appropriate choice of classifier method that performs the powerful view-invariant process. The purpose of this work is to tackle and perform the retrieval process using the LDCRFs classifier capable of real-time apps. The consequent two subsections are currently evaluating learning procedures and semantic retrieval.

1. Latent-dynamic Conditional Random Fields

Classification is the last phase of the work in the suggested scheme. retrieving images by text of image frames in our suggested approach assigns them to the corresponding classes. Throughout this stage, LDCRFs are used as a classifier (Fig. 3). LDCRFs are undirected graphical models designed for sequential data labeling [33]. Each label (state) represents a mammal. For certain characteristics, LDCRFs use a single exponential distribution to model all reference states, so there is a weight trade-off for this feature. For each observation sequence, the sub-structure variables $h = \{h_1, h_2, ..., h_m\}$ are assumed where the variables are not found in the learning data sets and thus form hidden variables set in the model.

The likelihood of label sequence y specified features sequence x is considered as in Eq. 15;

$$p(y|x,\theta) = \sum_{h:\forall h_i \in H_{yi}} p(h|x,\theta) \tag{15}$$

where, for the label y_j, h_i signifies a member to set H_{yi} of probable hidden states.

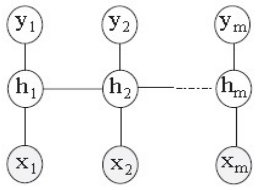

FIGURE 3. LDCRFs models where x_j relates to the respective j^{th} feature value, h_j is a concealed state allocated to x_j. y_j is to x_j state, in which the observed variables are characterized by gray lines.

$$p(h|x,\theta) = \frac{1}{Z(x,\theta)} \exp\left(\sum_{i=1}^{n} F_\theta(h_{i-1}, h_i, x, i)\right) \tag{16}$$

Here, parameter $\theta = (\lambda_1, \lambda_2, ..., \lambda_{Nf}; \mu_1, \mu_2, ..., \mu_{Ng})$, Nf is to transition function while Ng refers to the state function for the extracted features. Here, n is considered as a length for the features sequence x. Mathematically, F_θ is defined as well as in Eq. 17;

$$F_\theta(h_{i-1}, h_i, x, i) = \sum_f \lambda_f t_f(h_{i-1}, h_i, x, i) + \sum_g \mu_g s_g(h_i, x, i) \tag{17}$$

Here, t_f (h_{i-1}, h_i, x, i) represents the transition function at location $(i\text{-}1)$ and i. $Sg(h_i, x, i)$ is to state function at location i. Whereas, the weighted features of transition and state functions are represented by λ_f and μg, respectively. $Z(x,\theta)$ is referred to as normalized variable and calculated with respect to Eq. 18;

$$Z(x,\theta) = \sum_h \exp\left(\sum_{i=1}^{n} F_\theta(h_{i-1}, h_i, x, i)\right) \tag{18}$$

a) LDCRFs Model Parameter Learning

LDCRFs Model Parameter Learning: with respect to the training data $D = \{(x(j), y(j))\}^{Td}$, the parameter $\theta = (\lambda_1, \lambda_2, ..., \lambda_{Nf}; \mu_1, \mu_2, ..., \mu_{Ng})$, is computed, such that, $x(j)$ is a sequence of training features, $y(j)$ is a sequence of the corresponding state. T_d is the amount of sequences for learning. Eq. 19 calculates the objective function to know the parameter that maximizes the training data's log-likelihood.

$$L(\theta) = \sum_{j=1}^{T_d} \log\, p(y^{(j)}|x^{(j)}, \theta) \\ = \sum_{j=1}^{T_d} \Big(\sum_{i=1}^{n} F_\theta(h_{i-1}^{(j)}, h_i^{(j)}, x^{(j)}, i) - \log Z(x^{(j)}, \theta) \Big) \tag{19}$$

It is observed that maximization of probability can be carried out using a method of gradient ascension to converge with 300 iteration with the BFGS optimization technique [33] (Eq. 20).

$$\frac{\partial L(\theta)}{\partial \theta} = \sum_{j=1}^{T_d} \Big(\sum_{i=1}^{n} \frac{\partial F_\theta(h_{i-1}^{(j)}, h_i^{(j)}, x^{(j)}, i)}{\partial \theta} - \\ \sum_{\infty} p(h|x^{(j)}) \sum_{i=1}^{n} \frac{\partial F_\theta(h_{i-1}, h_i, x^{(j)}, i)}{\partial \theta} \Big) \tag{20}$$

b) LDCRFs Model Inferencing (Retrieving)

LDCRFs Model inferencing (retrieving): In order to calculate the likelihood $p(y|x, \theta)$ of state sequence y for given a feature sequence x, the matrices set is calculated as in Eq. 21. Special dummy start h_0 and dummy stop h_{n+1} hidden state is introduced to simplify some terms. Suppose $p(h|x, \theta)$ is provided via Eq.16. We define $M_i(x)$ of $|S \times S|$ matrix for every position i in the features sequence as shown in Eq. 21;

$$M_i(h', h|x) = \exp\big(F_\theta(h', h, x, i)\big) \tag{21}$$

where $S = h_1, h_2, ..., h_l$ refers to hidden states set for our proposed approach. l is to hidden states number for mammal label. At time i, $\hat{h}$ and h are denoting the states' S. Using those parameters, the conditional probability to state sequence y is calculated as in Eq. 22;

$$p(y|x, \theta) = \frac{\prod_i^{n+1} M_i(h_{i-1}, h_i|x)}{Z(x, \theta)} \tag{22}$$

Therefore, $Z(x, \theta)$ is normalized to characterizes the product entry of those matrices (Eq. 23).

$$Z(x, \theta) = \Big(\prod_{i=1}^{n+1} M_i(i) \Big)_{start,stop} \tag{23}$$

2. Semantic Retrieving

Regarding the proposed system of semantic image retrieval, the semantic image is retrieved either using image-based or text-based. Figure 4 summarizes the primary procedures that perform the retrieval outcome. The road map of these procedures will be detailed below.

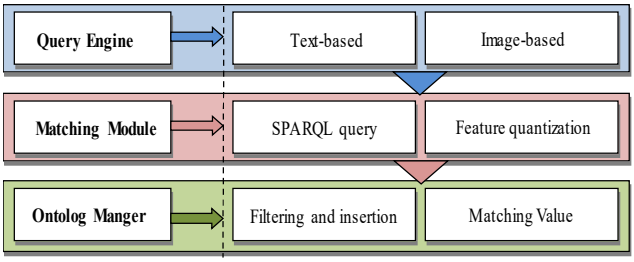

FIGURE 4. Retrieving semantic images with image or text query-based road map.

The semantic image is obtained in our suggested system either through text-based or image-based query, where the **Query Engine (QE)** is accountable for these two distinct techniques. The search process is begun for the text-based technique when the user enters the text entry in our sematic image retrieve system's dialog box. Bing, Alta Vista Google and Yahoo are the search engine.

The main motive is to give users the chance to learn and experiment with the interface sematic image retrieve. For example, a user enters text query such as "Lion," "Tiger," and "Giraffe," etc. Then the text-based of QE directly constructs the input text request. The sematic image retrieve scheme will retain object with some description of elective alternatives for the second image-based input method. In short, QE constructs a Query using SPARQL language for input image relied on knowledge base of ontology. Here the quantized high-level features ontology of dispersion and moment is prepared to SPARQL, which gets the parameters of an object to begin flexible search. Rightfully talking, the query for the quantized features of dispersion and moment for given input image, incomes the formula;

Query. Find the image of mammals with ($F_{dispersion\&moment}$)

SPARQL FROM: SELECT?K?L
WHERE {?L rdfs:subClassOf:mammals.?
K: $F_{dispersion\&moment}$

From Figure 4., **Matching Module (MM)** is the second process and retrieves semantic image with respect to QE Query in a good access scenario. But the MM conducts three primary processes in the case of a failure to retrieve appropriate image. At the beginning, MM surfing in query search engine like Google for the appropriate images. After that, the processing module approves the obtained photos to organize their verification of content. In addition, the acquired images are screened against user request appropriate (Fig. 4). If this has not occurred, it is proven that they depend on the dispersion and moments feature that have transformed to feature of high-level ontology. SPAEQL is finally made up of high-level ontology based on ontology. As a result, when matching a

user search query, we are well-thought-out for retrieving appropriate images; otherwise they are cast-off. With respect to Figure. 4, The end process is **Ontology Manger (OM)**, which carries out three primary filtering, insertion and ranking functions. First, the appropriate acquired images are filtered using the instance features of the ontology knowledge base. The second process is to construct the semantic description of retrieved photos and inserts the images that result in the knowledge base of ontology. Then ranking is the third task, which relied on the threshold value. This value is estimated using a user query reference summation and the corresponding ontology features function. The outcome is achieved by sorting the resulting photos regarding threshold value in descendant order. Then the acquired photos with greater ranking are closely selected as a request for the user.

3. Experiments Discussion

We used the mammal's dataset in our work, which is constructed by *Z. Malki* [34]. He has provided us the dataset to compare our suggested approach with him. The comparison is therefore fair everywhere the same conditions for the dataset are available. This knowledge base of ontology involves 25 different mammals. Every mammal has 50 distinct frames. This means the dataset includes 1250 frames for multiple mammals such as, Zebra, Proboscidea, Horse, Dear, Ox, Giraffe, Lion, etc. (Fig. 5). We categorical to split the ontology knowledge base into two-thirds for learning LDCRFs classifier and one-third for retrieving. It is observed that the sample of the training data is a completely divergent testing (i.e. retrieving) data.

In addition, we construct our system using Matlab language in conjunction some function in C++ language to retrieve the appropriate semantic photo either by text or photo as depicted in in Figure 6 and the Figure 7. For example, the user carries out our designate system and then browses the Proboscidea image as an input (Figure 7).

FIGURE 5. Mammal's dataset with respect to some sample image's frame.

As a consequence, the QE constructs the request of the image with regard to the knowledge base of ontology. In case of a positive reaction, the MM will retrieve the appropriate pictures. Therefore, the ontology and filtering procedures are well performed. Then OM starts to rank the resulting pictures and retrieves the top ten semantic pictures in descendant order. If the reaction is negative, the search process will occur in Web pictures. After that, in relation to updating the ontology content, the filtering method is performed to retrieve the outcome. Another situation is that the relevant semantic photos can be retrieved using text query as input in in our system interface (Figure 7). For example, when the user writes the text "Deer," QE will constructs the corresponding "Deer" query and then MM will retrieve the relevant images. After that we manage the processes of filtering and ranking regarding to knowledge base of ontology as well as OM.

We use the Precision and Recall scheme based on metrics in International Relations World to evaluate the assessment of our suggested approach for the retrieval of semantic photos. True positive (i.e. Precision,) measures the efficiency of the scheme to retrieve appropriate images (Eq. 24). Here, all truthful data does not provide the efficiency of the scheme because not all the semantic images collected are excellent. On the other side, false negative (i.e., recall) retrieves the appropriate semantic images using the complete number of related pictures that have been recovered (Eq.25). The separate pictures are not well-thought-out. Precision and recall are thus calculated using Eq. 24 and Eq. 25, such that we tested 25 classes of 415 different pictures. It is observed that the findings acquired yield effective outcomes. For Figure 8, *x*-axis represents the number of tests and *y*-axis refers to either Precision or Recall. The value of Precision/ Recall from

the Figure 8 is within the range of 0.35/0.99. Precision / Recall's greater value is in the range of (0.08 over 0.35)/(0.37 over 1.0), respectively. With 98.11 percent accuracy, the proposed approach can retrieve semantic images.

$$precision = \frac{true\ positive}{false\ positve + true\ positive} \quad (24)$$

$$Recall = \frac{true\ positive}{false\ negative + true\ positive} \quad (25)$$

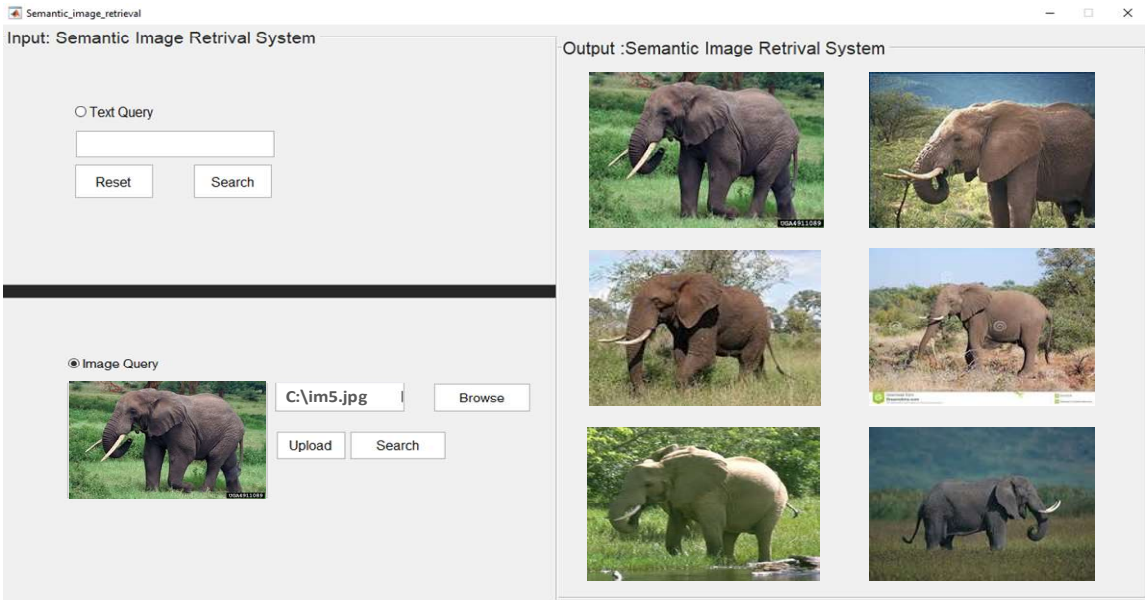

FIGURE 6. Image query to retrieve Proboscidea photos from the ontology knowledge base.

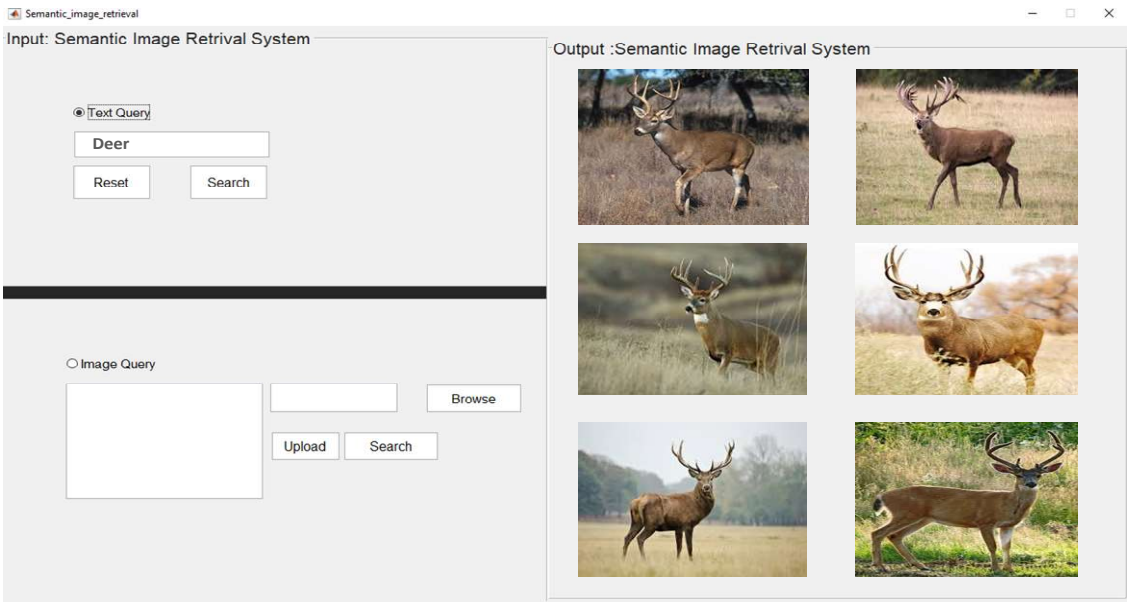

FIGURE 7. Text query to retrieve Dear photos from the ontology knowledge base.

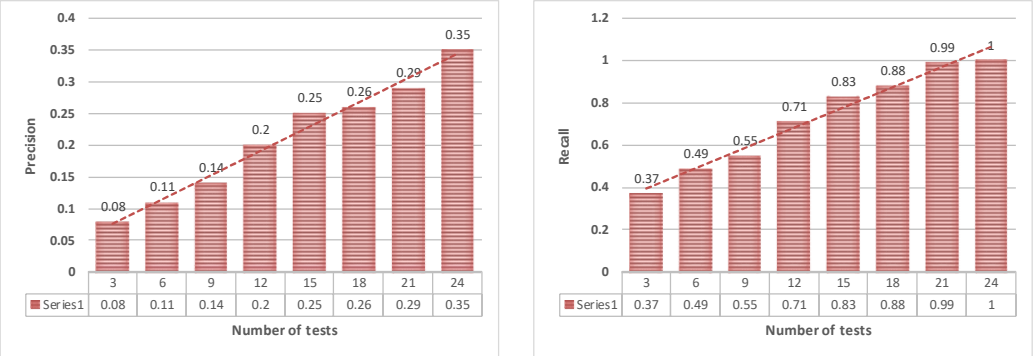

FIGURE 8. Outcome of the Precision and Recall metrices for our approach.

TABLE I
A COMPARISON BETWEEN OUR PROPOSED APPROACH WITH HAVE USED SIMILAR EXPERIMENTAL SETUPS AND MAMMALS' DATASET.

Method	*Classifier*	*Feature type*	*Retrieving rate*
Our Method	LDCRFs	Dispersion and moment features	98.11%
Zohair Malki [34]	CRF	CLF and color features	93.50 %
Zohair Malki [35]	Multi-SVM	Shape and geometric	96.33 %
M. Elmezain [36]	HMM	Symmetry features	97.76 %

We create a comparison between our suggested approach using comparable experimental setups and datasets to attain the principle of fairness. To evaluate the efficiency of the approach, the outcomes acquired were compared with [34, 35, 36] (Table I). In [34], the author used for retrieving images, Chord length shape features and color feature in conjunction with Conditional Random Filed (CRF). The drawback is too complicated because of the elevated dimensionality of the color element. Nevertheless in [35], for the classifier of multi-support vector machine, the author used two different types of features like shape and geometric features. Zernike moments are the shape feature, but the geometric features are circularity and rectangularity. The drawback here is to the element, which depends on the position of the image and is not efficient for natural objects. In [36], *M.Elmezain* used Hidden Markov Model (HMM) classifier over symmetry features to retrieve semantic images either by text or image query. The drawback is to many unstructured parameters of HMM. In addition, it limited by their first order Markov property as well cannot convey dependencies between concealed states.

5. Conclusion

This work presented a new approach for retrieving semantic images either by text or image queries. This approach is based on shape descriptors as dispersion and moment in conjunction with a discriminative model of LDCRFs. The target region firstly is localized via the background subtraction model. Then k-mean procedure quantized the extracted features, which carried out by dispersion and moments. LDCRFs with the main processes of Query Engine, Matching Module and Ontology Manger are employed to retrieve relevant images from building ontology dataset or from Web images. Experimental results demonstrated that, our approach successfully retrieved images against the mammals Benchmark of 25 different mammals with rate 98.11% precision. Additionally, a comparison between our proposed approach with have used similar experimental setups and dataset had done. Such outcomes are likely to compare very positively with those accessible in the literature from other researchers.

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, pp.139-154, 2016.

[2] J. varghese, T. B. Mary, "High level semantic image retrieval algorithm for Corel database", IRF International Conference, pp. 1-5, 2016.

[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, pp 241-257, 2016.

[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, pp. 3223-3248, 2016.

[5] T. S. B.S. Manjunath, P. Salembier, "Introduction to mpeg-7: Multimedia content description interface," Wiley, Chichester, ISBN: 978-0-471-48678-7, 2002.

[6] S. S. E.Rashedi, H.Nezamabadi-pour, “A simultaneous feature adapta-tion and feature selection method for content-based image retrieval systems,” Knowledge-Based Systems, Volume 39, pp. 85-94, 2013.

[7] 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, pp. 7-10, 2016.

[8] Y. Kleiman, G. Goldberg, Y. Amsterdamer, D. Cohen-Or, “ Toward semantic image similarity from crowdsourced clustering,” The Visual Computer, Volume 32, Issue 6, pp 1045–1055, 2016.

[9] 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, pp. 1-6, 2016.

[10] T. Dacheng, W. Dianhui, and M. Fionn, “Machine Learning in Intelligent Image Processing”, In Signal Processing, Vol. 93 (6), pp. 1399-1400, 2013.

[11] Z. Zeng,” A Novel Local Structure Descriptor for Color Image Retrieval,” Information Journal, Vol. 7, No. 1, pp. 1-14, 2016.

[12] 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, pp 597–607, 2016.

[13] 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, pp. 43-53, 2015.

[14] 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), pp.1574-1587, 2014.

[15] 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.

[16] 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), pp. 39–62, 1999.

[17] A. S. N.Singha, K. Singhb, “A novel approach for content based image retrieval,” Procedia Technology, vol. 4, pp. 245-250, 2012.

[18] L. L. Z. F. J. Yuea, Zhenbo Li, “Content-based image retrieval using color and texture fused features,” Mathematical and Computer Modelling, Vol. 54, pp. 1121-1127, 2011.

[19] E. K. H.N.pour, “Concept learning by fuzzy k-nn classification and relevance feedback for efficient image retrieval,” Expert Systems with Applications, vol. 36, pp. 5948-5954, 2009.

[20] S. K. K. S. B. Park, J.W. Lee, “Content-based image classification using a neural network,” Pattern Recognition Letters, vol. 25, pp. 287-300, 2004.

[21] Y. Y. Y. Rao, P. Mundur, “Fuzzy svm ensembles for relevance feedback in image retrieval,” LNCS, vol. 4071, pp. 350-359, 2006.

[22] P. G. U. Jayaraman, S.Prakash, “An efficient color and texture based iris image retrieval technique,” Expert Systems with Applications, Vol. 39, pp. 4915-4926, 2012.

[23] A. J. K. Iqbal, M. O. Odetayo, “Content-based image retrieval approach for biometric security using colour, texture and shape features controlledby fuzzy heuristics,” Journal of Computer and System Sciences, vol. 78, pp. 1258-1277, 2012.

[24] S. J. McKenna, Y. Raja, S. Gong, “Tracking colour objects using adaptive mixture models,” Journal of Image and Vision Computing 17, (**1999**), pp. 225–231.

[25] M. H. Yang and N. Ahuja, Gaussian Mixture Model of Human Skin Color and Its Applications in Image and Video Databases, In SPIE/EI&T Storage and Retrieval for Image and Video Databases, pp. 458-466, 1999.

[26] M. Elmezain, Adaptive Foreground with Cast Shadow Segmentation Using Gaussian Mixture Models and Invariant Color Features, International Journal of Engineering Science and Innovative Technology (IJESIT), pp. 438–445, 2013.

[27] Y. Q. Chen, M. S. Nixon and D. W. Thomas, Texture Classification using Statistical Geometric Features, Pattern Recog., 28(4), pp. 537–552, 1995.

[28] M. Hu, Visual Pattern Recognition by Moment Invariants. In IRE Transaction on Information Theory, Vol. 8, No. 2, pp. 179-187, ISSN 0096-1000, 1962.

[29] S. Maitra, Moment Invariants. In Proceeding of the IEEE, Vol. 67, pp. 697-699, 1979.

[30] J. Flusser, T. Suk, Pattern Recognition by Affine Moment Invariants. In Journal of Pattern Recognition, Vol. 26, No. 1, pp. 167-174, 1993.

[31] J. Davis, G. Bradski, Real-time Motion Template Gradients using Intel CVLib. In Proceeding of IEEE ICCV Workshop on Framerate Vision, pp. 1-20, 1999.

[32] T. F. E. Wikipedia, http://en.wikipedia.org/wiki/K-means clustering.

[33] M. Elmezain, A. Al-Hamadi, LDCRFs – Based Hand Gesture Recognition", IEEE International Conference on SMC, pp.2670-2675, 2012.

[34] Z. 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 pp.179-188, 2016.

[35] Z. Malki, “Shape and Geometric Features-based Semantic Image Retrieval Using Multi-class Support Vector Machine”, Journal of Theoretical and Applied Information Technology, Vol.95, No 20, pp.5535-5543, 2017.

[36] M. Elmezain, Shape Symmetry-based Semantic Image Retrieval Using Hidden Markov Model, Journal of Theoretical and Applied Information Technology, Vol. 96, pp. 3172-3181,2018.

MAHMOUD ELMEZAIN was born in Egypt. Between1997 and 2004 he worked as demonstrator in Dept. of Statistic and Computer Science, Tanta University, Egypt. He received his master’s degree in computer science from Helwan University, Egypt in 2004. He received PhD Degree in Computer Science from Institute for Electronics, Signal Processing and Communication sat Otto-von-Guericke-University of Magdeburg, Germany. His work focuses on image processing, pattern recognition, human-computer interaction and action recognition. Dr.-Ing. Elmezain is the author of more than 55 articles in peer-reviewed international journals and conferences.

HANI M. IBRAHIM, Lecturer of Computer Science, Department. of Mathematics, Faculty of Science, Menoufyia University, Egypt. In 2008, he obtained his Ph.D. degree in Computer Science, Department of Mathematics, Faculty of Science, Menoufyia University, Egypt. In 2004, he obtained his master’s degree in computer science, Department of Mathematics, Faculty of Science, Menoufyia University, Egypt. His research interests include Image Processing, Pattern Recognition, Biometric, Neural Networks, Artificial Intelligence, and Social Networks.