A Comparative Study On Automatic Annotation Of Visual Concepts By Web Image Mining

Khasim Syed [1], Dr. S.V.N Srinivasu [2]
[1]Dept of CSE
[1]Research Scholar, Rayalaseema University, Kurnool, (AP) – India.
[2]Professor & Principal, Dept of CSE
[2]IITS, Markapur, (AP) – India.

Abstract- *The ability to search the content of document images is essential for the usability and popularity. Towards the goal of large scale annotation, we presented a novel framework for annotation of visual concepts using web image mining. It applies a retrieval based approach for recognition of web images. Using existing techniques, the annotation time for large collections is very high, while the annotation performance degrades with increase in number of keywords. In the current work, annotation was performed by matching the cancroids of clusters between keywords and test visual words. One possible extension could be to match a huge set of clusters is based on graph matching algorithms. The statistical results are obtained using Corel dataset which contains 10,908 different images. The average precision and recall values are 0.8883 and 0.7125. A comparison was made between proposed system and a number of current automatic annotation methods. The performance of the framework over document image collections was found to be satisfactory and this approach is shown to be scalable to large multimedia collections.*

Keywords- Automatic annotation, Visual word, Reverse annotation, web mining, Corel, Graph matching.

I. INTRODUCTION

The state-of-the art in visual object retrieval from huge databases allows for seeking millions of images on the object level. [26] Online image repositories similar to Flickr include hundreds of millions of images and are expanding rapidly. Along with that desires for supporting indexing, searching and browsing is becoming more and more pressing. In addition, users expect that the search system agree to text queries and retrieve relevant outcomes in interactive times. Automatic annotation is a well-designed option to explicit identification in images. In the past years of images retrieval, images were annotated manually. Since manual attempt was costly, this was affordable for military and medical domains. [27] Content based image retrieval (CBIR) technique retrieve images associated to the query image (QI) from massive databases. The characteristic sets derived by the current CBIR systems are limited. This limit of the system is effectiveness. Since the beginning of economical imaging devices, the number of digital images and videos has grown exponentially. Large collections of images and videos are now available and shared online. Efficient retrieval from such type of large collections of multimedia data is becoming a vital issue.

In our research work, we utilize the web image mining framework. Unlike past annotation, where keywords are recognized for a given image, but in this frame work, the relevant images are detected for every keyword. [10] This changes the annotation issue from classification to verification. This enhances the annotation performance as well as reduces the annotation consuming time. This framework is initially designed for state, where the number of images to be annotated is much greater than the number of keywords to be annotated. As the architecture of web is rising extremely, huge databases are requiring supporting it. With the development of different applications such as Google Earth, Teleradiology etc., thus a huge collection of images & videos are to be stored & shared in online. [9]The retrieval from such big databases is becoming complex day to day operations. To cut the time complexity for Image retrieval and performing accurate clustering in Image mines is extremely significant problem. To overcome such type of problems, we projected annotation of medical documents using web image mining, which helps in a superior extent so that man-hours of Internet users will be, reduces.

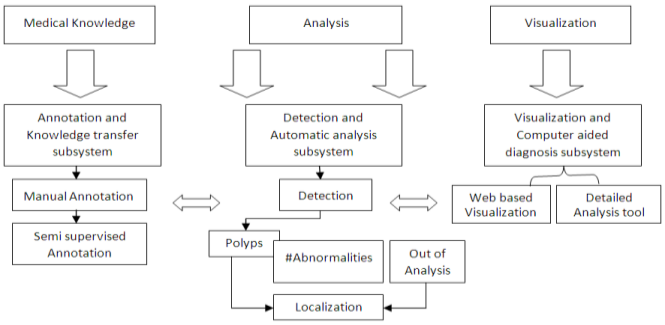


Figure 1. Automatic Image annotation of medical documents

This paper presents a comparative study related to the automatic annotation of visual concepts and annotation based web image mining. The rest of the paper is organized as follows: Section II reviews annotation of visual concepts.

Section III briefly reviews automatic image annotation methods and Section IV describes evaluation of image retrieval system. Section V including comparison of different approaches of automatic image annotation. Section IV concludes the paper.

II. ANNOTATION OF VISUAL CONCEPTS

The visual concept identification and annotation task is a multi-label classification challenge. It aims at the automatic annotation of a huge number of customer images with multiple annotations. The job can be solved by following three dissimilar approaches: i. Multi-modal approaches that consider visual information and/or Flickr user tags and/or EXIF data. ii. Automatic annotation with Flickr user tags. iii. Automatic annotation with visual data only. In all three cases the participants are asked to annotate the images of the test set with a predefined set of keywords (the concepts). This defined set of keywords allows for an automatic evaluation and comparison of the different approaches. The target of Image retrieval is, given a set of database images and input images, to discover all images in the database that are alike to the input images. Since the notion of visual similarity has an extremely wide meaning, the human perception of similarity is notoriously not easy to capture. Two images can be similar because they feature the same painting. Ravi finds that her cat looks similar to Raj's cat, while Ram thinks that cats in general are similar to tigers. If we think of similarity of two images as a binary function, which decides whether or not a given image pair is similar, then these dissimilar notions of similarity can be seen as allowing different numbers of degrees of freedom. Two images of the similar painting, for instance, only vary by the screening angle and potentially some lighting differences, while two images of the identical cat, additionally vary by the articulation of the cat, which adds a tremendous amount of degrees of freedom. State of the art Image Retrieval currently only deals with the former situation where there are only a few degrees of freedom. Since it was initially proposed by [19], the Visual Word based method to Image retrieval has been a success shining tale in Computer Vision. The following Figure 2 shows a frame work for annotation of visual concepts**.**

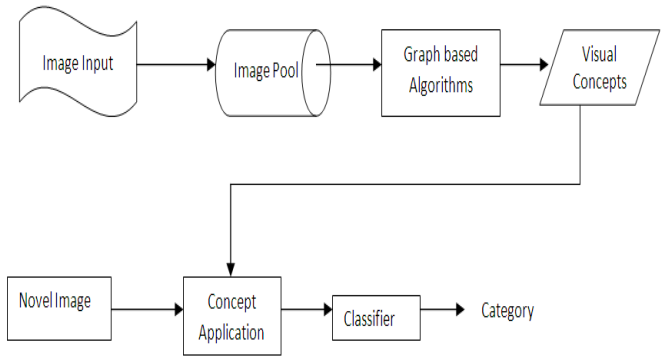


Figure 2. A framework for annotation of visual concepts

A. Visual Information

We utilize the Regularized Least Squares Classifier (RLSC) [21] as a binary classifier to notice a concept in an image. The outcome of the classifier can be used for image annotation; however this measure is not normalized therefore not fit to merge dissimilar features. The outcome of the classifier must be transformed to a probability. We acclimatize the method proposed in [20] to the RLSC. Assuming w is a Bernoulli random variable where the outcome can be one of two concepts, the probability p(z|a) can be obtained using the output of the classifier f (a) and a sigmoid function [20],

$$= e^{\frac{\|a_j - a\|^2}{2\sigma^2}} \quad (1)$$

In [20] several approaches to estimate the A and B parameters are discussed. Presently, we set them manually but in the future they will be estimated. Given the training set S_m = {(a_i, b_i) m_i=1} where labels $b_i \in \{-1, 1\}$ and a_i is a vector of image features, the decision boundary between the two classes (e.g., Indoor and Outdoor) is obtained by the differentiate function,

$$f(a) = \sum_{i=1}^{m} c_1 K(a_i, a) \quad (2)$$

Where K (a_i, a) is the Gaussian Kernel K (a_i, a) $= e^{\frac{\|a_j - a\|^2}{2\sigma^2}}$, m is the number of training points and c = [c_1, ..., c_m]T, is a vector of coefficients estimated by Least Squares [21],

$(m^{\varphi} J + K)c = b.$ (3)

Where J is the identity matrix, K is a square positive definite matrix with the elements $K_{i,j}$ = K (a_i, a_j), b is a vector with coordinates b_i and φ is a regularization parameter. To select the optimal values for σ and γ the cross-validation method is used. A point x with f (a) ≤ 0, is classified in the negative class (b = −1), and a point with f (a) > 0 is classified in the positive class (b = 1). If multiple features are used different classifiers are obtained.

III. AUTOMATIC ANNOTATION METHODS

The issue of automatic image tagging is closely related of image understanding and image classification. The term automatic annotation is used to illustrate the methods based on semantic concepts. Different approaches have been stated in order to annotate pictures with keywords describing their content. [30] Classify them using the following categories: (1) manual annotation, (2) collaborative annotation, (3) annotation with recognized words using ASR (Automatic Speech Recognition) tools, (4) annotation using an entertainment application, (5) semi-automatic and (6) automatic annotation. The table 1 describes features of annotation techniques:

Image annotation has received important attention in the research community over the past few years. Automatic image annotation assigns keywords to the image based on low level features automatically. [29]Automatic Image annotation can be classified into four approaches (as shown in figure 3): Probabilistic Modelling, Classification, and Graph Based and Parametric approach

Table 1. Summarizes the Characteristics of annotation methods

Annotation methods	Features			
	Effort	Accuracy	Input	Output
Manual Annotation	High	High	Text	Keywords
Semi- Automatic	Medium	Medium	Images	Contents
Entertainment	Low	High	Text	Keywords
ASR Tools	Medium	Medium	Audio	Keywords
Automatic	Low	Low	Images	Contents

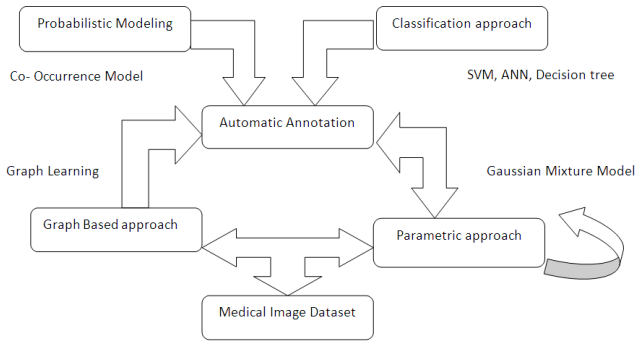


Figure 3. Classification of Automatic annotation Methods

A. CLASSIFICATION BASED APPROACH

In this approach, low level features are extracted from image content, and the features are fed directly into a conventional binary classifier which gives a yes or no vote. [23] The general machine learning tools include Artificial Neural Network (ANN), Decision Tree (DT), and Support Vector Machine (SVM). The mechanism of SVM classifier works by finding a hyper plane from a training set of samples to divide them. Feature vector and class tag is linked with every training sample. An SVM is fundamentally a binary classifier. The output of the classifier is the semantic concept which is used for image annotation. The aim is to describe a hyper plane which segment the set of samples such that all the points with the same label are on the similar side of the hyper plane. [16] Use the above mentioned fundamental framework to train 14 SVM classifiers for 14 images matched concepts. Images are represented with HSV histogram. [7] To train an SVM for a particular concept, training images belonging to that idea are regarded as positive instance while the others are regarded as negative instance. Therefore, every trained classifier can be regarded as one vs. all classifier. [13]During testing, every classifier produces a probabilistic decision. The class with utmost probability is chosen as the topic of the test image. An Artificial neural network (ANN) is a learning network that can learn from examples and can make decision for a novel sample. An ANN consists of multiple layers of interconnected nodes, which are also known as neurons or perceptions. The initial layer is the input layer which has neurons equal to the dimension of input sample. The number of neurons in the outcome layer is equivalent to the number of classes. The performance consequences of automatic image annotation is extremely affected by the segmentation consequences so to avoid prior segmentation Zhao et al proposed an approach called, hidden semantic analysis (LSA) based neural network (NN) annotation scheme. The annotation scheme is comprises of three parts. First, LSA is introduced to disclose the latent contextual correlation among the keywords. Second, with the tagged training images, Neural Network is obtained for characterizing the hidden linking between the visual content of the image and the textual keyword. Third, given a test image, the learnt Neural Network is able to effectively provide the keywords to be annotated.

A decision tree is multi level decision making approach. [15] Depending on the number of decisions made at every internal node of the tree, a DT can be called binary or n-ary tree. During training, a DT is built by recursively separating the training samples into non-overlapping sets, and each time the samples are separated, the attribute used for the division is discarded. The process continues until all samples of a cluster belonging to the similar class or the tree reaches its maximum depth when no attribute remains to divide them. Wan et al proposed an SDT algorithm a greedy algorithm, top-down recursive construct. In it the automatic annotation procedure is same to the image classification procedure. An image can contain a number of regions; every region has different semantic content and different visual features (color,

texture). Initially, image can be separated into different regions which have a sure same visual features, extracted image visual data of every region, structured the training dataset, and then the system use the assembly algorithm to classify the training set. Each class has a matching class tag, the class tag can be keywords, and at last the system attains automatic image annotation. [16]In this method, it uses the Simple Decision Tree (SDT) classification algorithm, to get better the decision tree algorithm, which compute model by the heuristic search of model space for rapid decision tree algorithm. To label a new sample, the tree is traversed from the root node to a leaf node using the attribute value of the new sample. The decision of the sample is the outcome of the leaf node where the sample reaches. Decision Tree algorithm is simple to interpret and recognize and can study with little number of samples. It is also hearty for incomplete and noisy data. The advantage of this type of method is that the retrieval is competent as there is no require to do image indexing and expensive online matching as in other IR approaches. The drawback of this kind of approach is that it does not consider the fact that many images belong to multiple categories.

B. PROBABILISTIC MODELLING APPROACH

In this approach annotation of image is done by estimating the joint probability of an image with a set of words. [11] Make use of the statistical machine translation model and applied the EM () algorithm to teach a maximum likelihood connection of words to image regions using a bi-lingual corpus. The pre-processed COREL data-set made available by [11] has become a broadly used and popular benchmark of annotation systems in the literature. The major essence of the algorithm is that it utilizes the probability table to probable correspondences and using it to process the estimate of the probability table. It annotates the image by partitioning the segments into blobs and finding the relationship of words and blobs by electing the words with uppermost probability. [14] Proposed a probabilistic generative model which is based on Bernoulli process to generate words and kernel density estimate to generate image features. It simultaneously learns the joint probabilities of associating words with image features using a training set of images with keywords and then generates multiple probabilistic annotations for each image. [8]This approach uses multiple grid segmentation and feature extraction is done using color and texture characteristics of image. Each training image has many annotations .This approach focuses on the presence or absence of words in the annotation rather than its importance. It does not rely on clustering and models continuous features.

[31] Proposed a Co-occurrence Model which is based on the co-occurrence of words with image regions created using a regular grid. The annotation process began by partitioning images into rectangular tiles of the same size. Then, for each tile, a feature descriptor which was fusion of color and texture is calculated. All the descriptors were then clustered into a number of groups which is represented by the centroid. Each tile inherited the whole set of labels from the original image. Then, the estimation of the probability of a label related to a cluster by the co-occurrence of the label and the image tiles within the cluster is done. Wang et al proposed progressive model to approximate the shared probability of words in for a given an image, the word with uppermost probability is primary annotated. Then, the successive words are annotated by incorporating the information of formerly annotated words. In this model combined probability of words is computed on basis of greedy algorithm.

C. PARAMETRIC APPROACH

In this approach, the feature space is assumed to follow a certain type of recognized continuous distribution. The conditional probability p(x|c) is modeled using multivariate Gaussian distribution where x and c are mean and concept label associated with feature vector. Both Li and Wang and [18] learn the conditional probability models concept by concept and then use the models to annotate unknown images. [17] First break down training images in each concept into regions which are represented using LUV colors and wavelet texture features. They then cluster regions into clusters which they call prototypes. For each prototype, a Gaussian model is learned. Finally, a Gaussian Mixture Model (GMM) is built for each concept by averaging the Gaussian model so find individual prototypes within the concept .To annotate an unknown image, its region features are extracted and the posterior probability of the image belonging to a concept is computed based on the concept GMM model. The drawback of this method is that parameter estimation for the Gaussian models is complex. [18] Do not segment images into regions. Instead, they assume that image features follow certain Gaussian distributions and directly learn a GMM for each training image within a model using expectation maximization (EM) algorithm. This is equivalent to a simultaneous segmentation and model learning process. They build the concept GMM by averaging the individual GMMs contained by the concept. In the annotation stage, a GMM is learned for the unknown image and the GMM is then matched with every one concept model. The concepts with the optimum match are selected as the annotations for the unknown image.

D. GRAPH BASED APPROACH

Projected an approach for AIA [12] using image based graph learning and word based graph learning. For a given the annotated training set and the visual features of all the images, the image-based graph learning aims to propagate labels from the annotated images to the un-annotated images by their visual similarities. The labeling matrix is another essential component during the graph learning. The image-based graph learning only focuses on the visual similarities among images, while the word correlations are not analyzed. Two words with high co-occurrence in the training set will lead to high probability to annotate certain image mutually, such as cloud and sky, water, and fish. Therefore, the word co-occurrence becomes an informative representation of the word correlation. To better capture the complex distribution of the image data, the Nearest Spanning Chain-based method was proposed to construct the image-based graph. The word-based graph learning was performed by exploring three kinds of word correlations. One is the word co occurrence in the training set, and the other two are derived from the web context.

IV. RESULTS AND DISCUSSION

A. Image dataset

The experiments in this learning employed the Corel dataset which contains 10,908 dissimilar images with every image in the size of 256*384 or 384*256. As such, the outcomes were reported utilizing the ten semantic sets with every comprising of 100 images. These datasets are in the groups of Food, Buses, Elephants, Mountains, Beach, Buildings, Flowers, Africa, Horses and Dinosaurs. These groups were used in reporting the results owing to the fact that the majority of the outstanding researches, for instance, [1-6, 28] employed these groups in demonstrating the effectiveness of their methods of CBIR.

B. Evaluation of Retrieval System

When users tag an image the game module calculates the score of the player move, using a formula that includes the trust level in the player, the probability of a tag set an image (obtained by the automatic algorithm outcome) and the feedback given by all previous users. [24]If this outcome is a strong value, it is considered a correct annotation, and the user score and trust level are increased, increasing indirectly the feedback provided by all users. Given a set of pictures $F = \{J_1, ..., J_N\}$ ($F \subset D_{img}$) and a set of concepts $U_{sc} = \{w_1, ...,wN_{con}\}$ ($U_{sc} \subset V_{con}$), the score obtained by matching the concept w in the image J is computed by,

$$S_{total}(J, w, n, m) = D_{group}(m) + \left[1 - D_{group}(m)\right] S_{new}(J, w, n) \quad (4)$$

Where n represents the number of correct annotations provided by the user, m is the number of times the concept w was annotated in image J, S_{new} is function that appraise the annotation using the semantic theory and the trust in the player (see (4)) and $D_{group}(m)$ define the cluster trust achieved by the correctness of the similar annotation provided by new users,

$$D_{group}(m) = 1 - e^{\left(\frac{m}{kg}\right)} \quad (5)$$

The exponential parameter k_g is estimated in order to get a cluster trust near the maximum value after m annotations. We considered that three players providing the same annotation (m = 3) means a high cluster trust and for this reason k_g is obtained assuming this state. The ESP GAME [22] validates an annotation with two players. With this equation, when m = 2, the score is not the maximum but is a value that accept the system to classify the annotation as correct. When a concept w is annotated for the first time in an image I the score is computed by,

$$S_{new}(J, w, n) = D_{player}\left[(n) + C_{player}(n)\right] p(w / j) \quad (6)$$

Where p(w|J) is the probability obtain by the automatic method (semantic concepts) and C_{player} is the trust of the system in the player that notify the quality of past annotations provided by the player,

$$D_{player}(n) = \begin{cases} k_p(n), n < k_{moves} \\ k_{conf}(n), n \geq k_{moves} \end{cases}$$

Where K_{moves} is a constant with the number of superior moves to reach to the player trust maximum value k_{conf} and k_p is a constant that is used to increment the player trust.

The number of right moves n increases when the cluster trust is diverse from zero and the score is greater than a defined threshold. It decreases when the score is over another threshold. These thresholds were obtained analytically. When the group trust is zero this means the score is attained using only the semantic concepts and the player faith. In these cases, it is hard to identify the accuracy of the annotation.

C. Image annotation algorithm

An annotation on an image $J \in C_{img}$ of a concept w_i acceptance to a vocabulary $U_{con} = \{w_1, w_2, ..., w_k\}$, is clear as, $B(J, w_i)$. Given a set $F \subset C_{img}$ with N_l images and a set of $U_{sc} \subset U_{con}$ with N_{con} concepts, the automatic approach is defined by the following steps:

1. The subsets L and U_{sc} are given in the interface.
2. The user chosen single image $Jl \in F$ and a concept $wk \in Usc$;
3. The user form an annotation, $B_i(J_l, w_k)$;
4. The score is calculated using the automatic models $p(w_k|J_l)$, the faith of the game in the player and the feedback given by all past users;
5. For every concepts $wk \in U_{sc}$, if the $|\{B_1, B_2, ..., B_{NA}\}| > Nupd$ for a concept wk, then the training set is modernized and the model for the concept wk is computed again;
6. Go to 2.

A semantic model is trained once more when the number of dissimilar right annotations with the concept is over N_{upd}. An annotation is considered accurate when it is performed by at least two users. As a outcome of this algorithm, a set of annotations $B = \{B_1, B_2, ..., B_{Ntotal}\}$ is attained and the semantic concepts of the set U_{con} are estimated with a huge training set. If two diverse players given the similar wrong annotation the algorithm fails and this can boost the number of failures of the correlated concept but this is not a usual circumstances. Both subsets, F and U_{sc}, used in each stage of the game are chosen in the automatic annotation block. Therefore, the learning procedure is driven by the automatic model.

Precision refers to a compute of the capacity of the system in retrieving just the images that are similar to the input image. Meanwhile, the Recall rate called the optimistic rate or sensitivity, gauges the capacity of CBIR systems in retrieving the image that are same to the QIs. For the elaboration of the outcomes, calculation was finished to precision and recall according to the number of input images (from the test dataset) and the retrieved similar images from the corel image database.

$$\text{Recall} = \frac{\text{number of similar image retrieved}}{\text{total number of similar images in the database}} \quad (8)$$

$$\text{Precision} = \frac{\text{number of similar image retrieved}}{\text{total number of images retrieved}} \quad (9)$$

Eqs. (8) And (9) comprise the calculation of the precision and recall for the query image [6]. Graph-based matching techniques are used for problem solving, learning, and discovery. Where a comprehensive search is not practical, heuristic methods are used to speed up the procedure of finding a satisfactory solution. This approach includes using a rule of thumb, an educated imagine, an intuitive judgment, or common sense. In additional precise terms, heuristics are strategies using readily accessible, though loosely applicable, information to control problem solving in human beings and machines. In computer science, mathematical optimization, artificial intelligence and a heuristic is a technique designed for solving a problem more speedily when classic methods are too slow, or for finding an approximate solution when classic methods fail to discover several exact ones, but they do not assurance that the finest will be found, therefore they may be considered as around and not precise algorithms. These algorithms frequently discover a solution close to the most excellent one and they identify it fast and easily. Sometimes these algorithms can be accurate, that is they actually discover the finest solution, but the algorithm is still called heuristic until this best solution is proven to be the best. Comparisons were made between the proposed system and a number of current automatic annotation methods [1-6, 38]. This allows the measurement of the usability of the proposed method. The motivation for this selection to compare with these methods is that the outcomes of these methods were reported via a general denomination of ten semantic sets where an individual set contains 100 images of Corel dataset. As such, it is possible to compare clear outcomes using the reported results. This makes the performance comparison achievable. The comparison of the average precision for each group of the proposed system with other comparative systems can be referred in Table 3. As evidenced by the outcomes, the proposed system demonstrates sounder performance with respect to precision in comparison to other systems. Comparison of the average recall rates for all clusters of the proposed system with the similar comparative systems is shown in Table 3. The recall results of the proposed system achieved the best recall rates.

As can be infer, the above comparison outcomes reveal the capacity of the proposed system in generating improved precision and recall rates. Its performance also supersedes other state-of the art methods [1-6, 28] particularly with respect to precision and recall rates. In specific, the average precision and recall rates obtained were 0.8883 and 0.7125 respectively. This is factored by the fact that the authors in [1-6, 28] created the systems of CBIR that extract a restricted number of feature sets. This restricts retrieval in terms of efficiency and competence. On the other hand, the system proposed in this study extracted robust and extensive

set of features. The meta-heuristic techniques were employed for optimizing the precision of the retrieved images. The addition of the ILS algorithm with the GA has raised the quality of solution via the increase of the fitness number. This has helped in the development of the exploitation procedure when the searching process is being conducted. Clearly, the experimental outcomes are demonstrating the capacity of the meta-heuristic techniques in assisting the retrieval of the great amount of the relevant images to the query image. The following Table 2 and Table 3 describe the Comparison of average precision results and average recall results of different authors using different methods. Such as image retrieval using interactive genetic algorithm (IRIGA), a bandelet transform based image representation technique, Curvelet-based image retrieval scheme, motif co-occurrence matrix (MCM), Three techniques(3D color histogram and the Gabor filter algorithm, genetic algorithm, preliminary and deeply reduction for extracting technique), Bandelets transform based image representation technique with SVM [25].

Class	Madhavi et al. (2016)	Asharf et al. (2015)	Rao et al. (2011)	Youssef (2012)	Lin et al. (2009)	Jhawar et al. (2004)	ElAlami (2011)	Ashraf et al. (2016)	Khasim et al. (2017)
Buses	0.846	0.95	0.89	0.92	0.88	0.74	0.87	0.9	0.96
Mountains	0.811	0.75	0.51	0.74	0.52	0.29	0.53	0.7	0.82
Beach	0.892	0.70	0.53	0.64	0.54	0.39	0.56	0.75	0.90
Elephents	0.727	0.80	0.57	0.78	0.65	0.30	0.67	0.9	0.83
Food	0.871	0.75	0.69	0.81	0.73	0.36	0.74	0.8	0.87
Flowers	0.917	0.95	0.89	0.95	0.89	0.85	0.91	0.8	0.96
Africa	0.828	0.95	0.56	0.64	0.68	0.45	0.70	0.8	0.83
Horses	0.951	0.95	0.78	0.95	0.80	0.56	0.83	0.9	0.96
Dinosaurs	0.828	0.95	0.98	0.99	0.99	0.91	0.97	1	0.99
Buildings	0.632	0.95	0.61	0.70	0.54	0.37	0.57	0.75	0.75
Average	**0.830**	**0.820**	**0.701**	**0.812**	**0.722**	**0.522**	**0.735**	**0.83**	**0.8883**

Table 2. A Comparison of average precision results

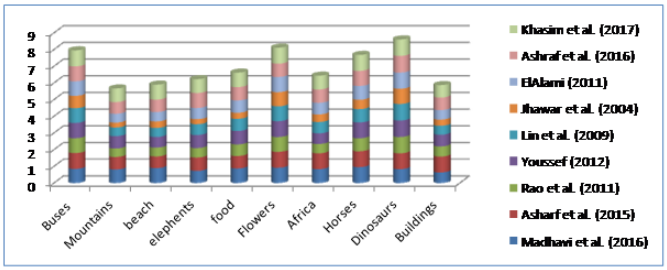
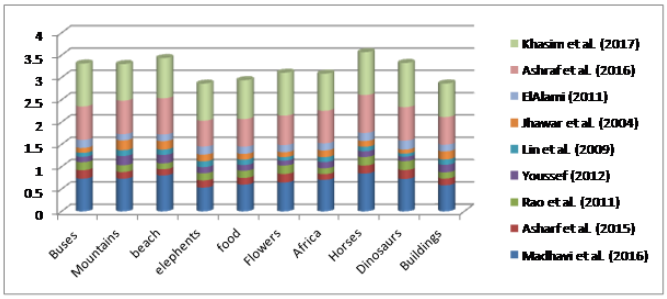


Table 3. A Comparison of average recall results

Class	Madhavi et al. (2016)	Asharf et al. (2015)	Youssef (2012)	Lin et al. (2009)	Jhawar et al. (2004)	ElAlami (2011)	Ashraf et al. (2016)	Khasim et al. (2017)
Buses	0.733	0.19	0.18	0.12	0.09	0.11	0.18	0.75
Mountains	0.732	0.15	0.15	0.21	0.13	0.22	0.14	0.75
beach	0.805	0.14	0.13	0.19	0.12	0.19	0.15	0.81
elephents	0.533	0.16	0.16	0.14	0.13	0.15	0.18	0.58
food	0.600	0.15	0.16	0.13	0.12	0.13	0.16	0.62
Flowers	0.647	0.19	0.19	0.11	0.08	0.11	0.16	0.66
Africa	0.706	0.13	0.13	0.14	0.11	0.15	0.16	0.73
Horses	0.848	0.18	0.19	0.13	0.10	0.13	0.18	0.85
Dinosaurs	0.726	0.20	0.20	0.10	0.07	0.09	0.2	0.75
Buildings	0.585	0.15	0.14	0.17	0.12	0.18	0.15	0.62
Average	**0.691**	**0.164**	**0.163**	**0.144**	**0.107**	**0.146**	**0.166**	**0.7125**



V. CONCLUSION

The state-of-the art in visual object retrieval from huge databases allows for searching millions of images on the object level. The ability to search the content of document images is essential for the usability and popularity. Towards the goal of large scale annotation, we presented a novel framework for annotation of visual concepts using web image mining. It applies a retrieval based approach for recognition of web images. Using existing techniques, the annotation time for large collections is very high, while the annotation performance degrades with increase in number of keywords. In the present work, annotation was performed by matching the centroids of clusters between keywords and test visual words. One possible extension could be to match a large set of clusters is based on graph matching algorithms. The statistical results are obtained using Corel dataset which contains 10,908 different images and these datasets are in the groups of Food, Buses, Elephants, Mountains, Beach, Buildings, Flowers, Africa, Horses and Dinosaurs. The average precision and recall values are 0.8883 and 0.7125. A comparison was made between proposed system and a number of current automatic annotation methods. The performance of the framework over document image collections was found to be satisfactory and this approach is shown to be scalable to large multimedia collections.

REFERENCES

[1] Ashraf, R., Bashir, K., Irtaza, A., Mahmood, M.T., 2015. Content based image retrieval using embedded neural networks with bandletized regions. Entropy 17, 3552–3580.

[2] Ashraf, R., Bashir, K., Mahmood, T., 2016. Content-based image retrieval by exploring bandletized regions through support vector machines. J. Inform. Sci. Eng. 32, 245–269.

[3] Madhavi, K.V., Tamilkodi, R., Sudha, K.J., 2016. An innovative method for retrieving relevant images by getting the top-ranked images first using interactive genetic algorithm. Proc. Comput. Sci. 79, 254–261.

[4] ElAlami, M.E., 2011. A novel image retrieval model based on the most relevant features. Knowl.-Based Syst. 24, 23–32.
[5] Jhanwar, N., Chaudhuri, S., Seetharaman, G., Zavidovique, B., 2004. Content based image retrieval using motif cooccurrence matrix. Image Vis. Comput. 22, 1211– 1220.
[6] Rao, M.B., Rao, B.P., Govardhan, A., 2011. CTDCIRS: content based image retrieval system based on dominant color and texture features. Int. J. Comput. Appl. 18, 40–46.
[7] M.Sangeetha,K.Anandakumar and A.Bharathi," Automatic Image Annotation and Retrieval: A Survey", Volume: 03 Issue: 04 | Apr-2016. www.irjet.net.
[8] D. Zhang, Md. M. Islam, G. Lu, 2012. "A review on automatic image annotation techniques", Pattern Recognition, vol. 45, no. 1, pp. 346–362.
[9] Y. Poornima and P.S. Hiremath,"Survey on Content Based Image Retreival System and Gap Analysis for Visual Art Image Retreival System", IJCSI International Journal of Computer Science Issues, Vol. 10, Issue 3, No 1, May 2013 ISSN (Print): 1694-0814 | ISSN (Online): 1694-0784.
[10] Zhiwu Lu, Horace H. S. and Qizhen He," Context-based multi-label image annotation", Proceeding CIVR '09 Proceedings of the ACM International Conference on Image and Video Retrieval Article No. 30, Santorini, Fira, Greece — July 08 - 10, 2009 .
[11] P. Duygulu, K. Barnard, N. de Freitas, D. Forsyth, 2002. "Object recognition as machine translation: learning a lexicon for a fixed image vocabulary", In Seventh European Conference on Computer Vision (ECCV), Vol. 4, pp. 97-112.
[12] J. Liu, M. Li, Q. Liu, H. Lu, and S. Ma, 2009."Image annotation via graph learning," Pattern Recogn., 42(2), ,pp. 218-228.
[13] Manjunath, B.S.,Ohm, J.-R., Vasudevan, V.V.; Yamada, A.,2001. "Color and texture descriptors," Circuits and Systems for Video Technology, IEEE Transactions on , vol.11, no.6, pp.703,715.
[14] Yufeng Zhao,Yao Zhao, Zhenfeng Zhu; Jeng-Shyang Pan, 2008."A Novel Image Annotation Scheme Based on Neural Network," Intelligent Systems Design and Applications, ISDA '08. Eighth International Conference on, vol.3, no., pp.644, 647.
[15] Shaohua Wan, 2011."Image Annotation Using the Simple Decision Tree," Management of e-Commerce and e-Government (ICMeCG), 2011 Fifth International Conference on , vol., no., pp.141,146.
[16] Chapelle, O, Heffner, P., Vapnik, V.N., 1999. "Support vector machines for histogram-based image classification," IEEE Transactions on Neural Networks, vol.10, no.5, pp.1055, 1064.
[17] J. Li, J.Z. Wang, Real-time computerized annotation of pictures, IEEE PAMI 30 (6) (2008) 985–1002.
[18] G. Carneiro, A.B. Chan, P.J. Moreno, N. Vasconcelos, Supervised learning of semantic classes for image annotation and retrieval, IEEE PAMI 29 (3) (2007) 394–410.
[19] Josef Sivic and Andrew Zisserman," Video Google: A Text Retrieval Approach to Object Matching in Videos", Proceedings of the Ninth IEEE International Conference on Computer Vision (ICCV 2003) 2-Volume Set.
[20] Platt J (1999) Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods. In: Advances in large margin classifiers. MIT Press, pp 61–74.
[21] Poggio T, Smale S (2003) The mathematics of learning: dealing with data. In: Notice of American Mathematical Society, pp 537–544.
[22] von Ahn L, Dabbish L (2004) Labeling images with a computer game. In: Proceedings of the SIGCHI conference on human factors in computing systems CHI '04, pp 319–326.
[23] D. Zhang, Md. M. Islam, G. Lu, 2012. "A review on automatic image annotation techniques", Pattern Recognition, vol. 45, no. 1, pp. 346–362.
[24] Jiebo L, Boutell M, Brown C (2006) Pictures are not taken in a vacuum—an overview of exploiting context for semantic scene content understanding. IEEE Signal Process Mag 22:101– 114.
[25] Platt J (1999) Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods. In: Advances in large margin classifiers. MIT Press, pp 61–74.
[26] Marcel Worring, Ton van Rijn and Ork de Rooij, "Browsing visual collections using graphs", Proceeding MIR '07 Proceedings of the international workshop on Workshop on multimedia information retrieval, Pages 307-312, Augsburg, Bavaria, Germany,September 24 - 29, 2007.
[27] Prof. Vikram M Kakade and Ishwar A. Keche," Review on Content Based Image Retrieval (CBIR) Technique", International Journal Of Engineering And Computer Science ISSN: 2319-7242 Volume 6 Issue 3 March 2017, Page No. 20414-20416.
[28] Lin, C.-H., Chen, R.-T., Chan, Y.-K., 2009. A smart content-based image retrieval system based on color and texture feature. Image Vis. Comput. 27, 658–665.
[29] Dhatri Pandya and Prof. Bhumika Shah,"Comparative Study on Automatic Image Annotation", International Journal of Emerging Technology and Advanced Engineering Website: www.ijetae.com (ISSN 2250-2459,

ISO 9001:2008 Certified Journal, Volume 4, Issue 3, March 2014).

[30] Rui Jesus, Arnaldo J, Abrantes and Nuno Correia," Methods for automatic and assisted image annotation", Multimed Tools Appl (2011) 55:7–26 DOI 10.1007/s11042-010-0586-z.

[31] Jesus R, Dias R, Frias R, Abrantes A, Correia N (2008) Memoria mobile: sharing pictures of a point of interest. In: Proceedings of the working conference on advanced visual interfaces (AVI '08). ACM, New York.