

NEW ONTOLOGY RETRIEVAL IMAGE METHOD IN 5K COREL IMAGES

Hossein Sahlani

Department of Information, Amin University, Tehran, Iran

ABSTRACT

Semantic annotation of images is an important research topic on both image understanding and database or web image search. Image annotation is a technique to choosing appropriate labels for images with extracting effective and hidden feature in pictures. In the feature extraction step of proposed method, we present a model, which combined effective features of visual topics (global features over an image) and regional contexts (relationship between the regions in Image and each other regions images) to automatic image annotation. In the annotation step of proposed method, we create a new ontology (base on WordNet ontology) for the semantic relationships between tags in the classification and improving semantic gap exist in the automatic image annotation. Experiments result on the 5k Corel dataset show the proposed method of image annotation in addition to reducing the complexity of the classification, increased accuracy compared to the another methods.

KEYWORDS

Automatic Image Annotation, Ontology, Statistical Models, Regional Contexts, Visual Topics.

1. INTRODUCTION

Today with the exponential growth of the Web image, huge set of digital images is available in variety collection, how to retrieval and manage them presents a significant challenge. For search in the sets you need to enter appropriate text index to access the related images. These systems have some problems: First, annotation images humanly require much time and much cost; Second, dependent on operators, the point of view of different operators is not uniform concepts in one image, so annotations of image are not concept all areas of image. This means that text-based image searches are not explicitly enough and to overcome the problem of text-based annotation systems, with gradually increasing the number of images in the databases, such as the World Wide Web were introduced content-based image retrieval systems. Content-based Image Retrieval (CBIR) methods compute relevance between images based on the visual similarity of the low-level features [2]. However, the problem is the so-called “semantic gap” between low-level features and high-level semantic concepts [4]. In these systems, use machine learning or statistical methods for classification or indexing of images that classes or indexes in each row represents a label or a tag. Features extracted from the input image give to the machine learning or another indexing system. These features are visual concepts such as color, texture and other low-level features that extracted from the images. In the image annotations system, to achieve good performance need to the number of enough extracted features. If used only one separator classifier to classify images that this may lead to low classification accuracy and complexity of the classification that in this paper used created ontology for annotation of each image. It is two main parts in the content retrieval images system: feature extraction and indexing or classification. In order to produce optimal system for each section, first must be selected the appropriate features for example texture features, colour features... Then doing feature selection algorithm that effect of select important features on the feature set that is performed with KNN

classifier and genetic algorithm and for indexing in each level we used combine statistical indexing and other classification models. Then, in section 2, has been done an overview of the other work in this area and then the proposed approach is fully described in section 3, in this section we describe how to create ontology then explain features that used and how to combine them with greater detail. In section 4 describe effect of ontology and statistical indexing system to access better result and better describe concepts and greater matching capabilities together, and compared proposed method with other methods and in the end we will conclude in section 5.

2. RELATED WORKS

Image Automatic Annotation method not based on a text labels and each images is annotated with concept and feature that extracted based on image processing from self-image. The idea is to use low-level features extracted from the image and can be achieved to their high-level semantics. Such as low-level features of image is color, texture, shape, etc. while the semantics feature or label and words that impression of the image are high-level concepts. One of the major challenges in this field is the distance between these two concepts (label and feature) and there is known as a semantic gap. In order to remove this semantic gap we using anthology based method for annotations of images. In fact, should be found the mapping between low-level features and concepts. To have this mapping should be used machine learning or another technique. The annotations of images are a classification that represents a concept or phrase for each category of the image [1]. Classification and learning process requires a series of images with different label is called the training set. Image Annotation process including selecting the appropriate training set, feature extraction from images and maps using machine learning algorithms between the low level features of image and words or labels. With regard to the above, images annotation, can be seen in the following four groups: Vector Space Models [5, 6], Classification Methods [7, 10] Graph-based Methods [11] and Statistical Models [15, 16].

1 - The Vector Space Models: The vector space model framework [5], [6] is a popular technique in information retrieval especially in text retrieval. Generally, documents are represented as vectors, each of which contains the occurrences of keywords within the document in question. The length of the vectors is equal to the vocabulary size. These approaches treat images as documents, and build visual terms which are analogous to keywords, from the image feature descriptors. Some of the methods in this category include: SvdCos method, Saliency-based semantic propagation, and cross-language latent semantic indexing based approach.

2 - Statistical models: The basic idea of this method is to estimate the tags associated with an image, in these methods, the features extracted from the images of the different areas and then the attention to the frequency of features the entire or segmented image are labeled, These methods are widely used in textual retrieval [3]. The co-occurrence model proposed by Mori et al [16] is perhaps one of the first attempts at automatic image annotation. They first divide images into rectangular tiles of the same size and calculate a feature descriptor of color and texture for each tile. All the descriptors are clustered into a number of groups, each of which is represented by the centered. Each tile inherits the whole set of labels from the original image. Then, they estimate the probability of a keyword W related to a cluster C by the co-occurrence of the keyword and the image tiles within the cluster. In [15] proposed a machine translation model for automatic image annotation. They argued that region based image annotation is more interesting because global annotation does not give information on which part of the image is related to which keyword. In their point of view, the process of attaching keywords to image regions is similar to the translation of one form of representation image regions. Indeed, their machine translation model applies one of the classical statistical machine translation models to translate from the set of keywords of an image to the set of blobs that generated by clustering the image features. In [8],

proposed Cross Media Relevance Model (CMRM) for automatic image annotation. CMRM benefits the joint distribution of keywords and blobs. Each image describe by visual features (blobs) as well as textual keywords. In CMRM the probability of observing a set of blobs with a set of keywords estimated. In [9], proposed a Continues-space Relevance Model (CRM) that improves CMRM by using continuous probability density functions to estimate the probability of observing a region given an image. In [10], modified the above model and proposed Multiple Bernoulli Relevance Model (MBRM) such that the probability of observing labels given an image was modeled as a multiple-Bernoulli distribution. In MBRM images simply divided into rectangular tiles instead of applying automatic segmentation algorithms. There exist two problems in above algorithms. First, most existing algorithms have taken one of two approaches, using regional or global features exclusively. Second, conventional approaches consider each word separately without the textual context relations. As a result the textual context relations among annotation words have been ignored. By textual context we refer to co-occurrence relationship among words. Wang et al. [11] proposed a model for image annotation that combines global, regional, and contextual features by an extended CMRM. They incorporate the three kinds of information to describe image semantics to annotate images by estimating their joint probability.

3 - Classification or Annotation using Keywords Methods: in this approach exist a number of predefined labels and cluster that train images is clustered in given groups when given an image without any label for labeling, classification algorithms identify the components of image and assign it to the corresponding label. Each image can be annotated use a list of associated keywords. Annotation labels can include custom words or include a list of predefined words (dictionaries). Method [11] and [12] classify these images based on extracted features from them. In method [12] in addition to sift features extracted from different areas of the images, spatial feature extract from each image too that led to have 1416 features for each of the areas of the images. In this method, first sift feature extracted from images then these features are combined together that the total sift features extracted for each of the areas of the images are 1024. Then spatial features between the areas of the images is extracted too that the total number of them are 392 so for the areas of the images this method or technique yields a total of 1416 features for each of the areas of images. In method [12], is not used on the local sift features and used features basis of texture and color of images and other features that generally used in image classification. In [3] used the support vector machine to classify features extracted from the images. In [4] was used a simple spatial vector model that directly classify images from a comparison of the features histogram. In [5], was used combination of two support vector machine classification (first for a global of images features and another for a locally features of image areas) to automatically annotate images.

4 - Graph-based or Ontology-based Methods: ontology is a fundamentally involves subjects (entities) and relationships between them. Add a hierarchical structure to the library of words leads to the production of a classifier that named Ontology that can be solves the problem and ambiguity of concept words [9, 10, and 6]. In the method [9] using a combination of several ontology (that self-created these hierarchical structure of ontology) to classify images and the results of each ontology by initialization of weight are combined and used Part II that explain in [10] is division of features that by using machine learning techniques classify the images that are in a class and the images are not in the class. (This is done separately for each of the layers) and thus train the system how to classify the images.

Among the image annotation models classification approaches are the oldest in this domain. But statistical models and graph based approaches have more chance. So we proposed approach in this paper follows combining of statistical based models and ontology-based methods.

3. ONTOLOGY

We do not know how the humans brain extract concepts and low-level features from objects, this is reason of different between machine and human perception and this different is called the semantic gap. There are patterns in the human brain that with combining low-level features describes concepts and meaning and eliminates the semantic gap.

In retrieval system, each image is known as a feature vector. At the beginning are computed feature vectors for all images of database and stored in the feature database. Image retrieval systems, sort of images feature vectors that to minimize the time required for finding similar images in an image retrieval system and to maximize efficiency system. Several techniques are used for indexing images that can be pointed to hierarchical structures so to find similar images using a top-down search, image feature vector compare with ontology branches and select the best branch with the shorter distance and the process will continue to achieve the ultimate branches or leaves thus can be considered a multi-index images system that the index helps to determine concepts completely. The relationships between the vocabularies of ontology also used to create the image tag that is considered to be a more appropriate label for the image.

In this paper we used created ontology based on WordNet and ImageNet ontology [14] to consider the relationship between the concepts to assign appropriate label to images. This ontology is fully described in continue. This ontology is used topological, behavioral, semantic and lexical relationships in the data.

3.1. Create Ontology

There are specific image databases for training and testing of image annotation systems, for example using of images retrieval database such as Corel, ImageNet or ImageClef database and other existing databases. However, in this paper must be use of databases that have a hierarchical structure (ontology) or create hierarchical structure (ontology) for used database, that we used Corel database, that hasn't hierarchical structure in the labels so we create this in the labels with WordNet ontology.

To do this, the database was divided into 6 main categories (based on ontology WordNet, ImageClef) that include: natural, manmade, human, animals, food and other items. Hierarchical structure of the database must be done with sufficient accuracy otherwise creates problems in the classification.

To create the ontology we used of 3 persons idea, in most cases these ideas were similar but in some cases that there was a difference of opinion, majority decision was made for the labels and in some cases causes the label to be placed in "other" classes.

Each picture is labeled from the root of the ontology to reach a leaf in hierarchical structure. For example, an image of apple can be labeled as follows:

Entity → Landscape-Nature → Vegetation → Fruit → Apple

Ontology pictures, is shown as follow. To enhance the clarity, the parts of ontology shown separately.

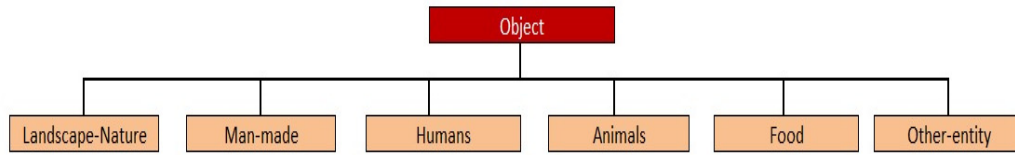


Figure 1: Ontology of entity objects

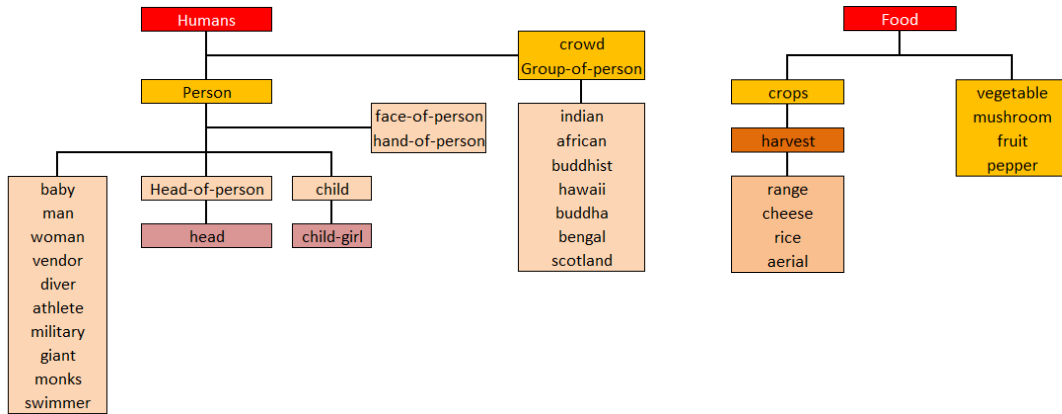


Figure 2: Ontology of humans and foods

Each area of the image should be like to the dictionary words and describe with these words, thus used fuzzy weighted number. In which each regions of images related with different weights to all the words of the dictionary thus, there are no restrictions to connection of the regions of image, to dictionary words. Crisp weighting is relating each area of the pictures to just a visually word and can be demonstrated in Fig 3. Clearly, this method is much better than the case that the feature vector is just belong to the word “h”, because exist a few distance between the feature vector and the word “I” and it is good that feature vector belong to both the center words with appropriate weight. To clarify the difference between these two methods, an example of histograms for a sample image with the hard and fuzzy weighting method is presented. Fig 3 shows an example that areas of an image belong to the dictionary words with hard weighting method. In order to easily, only four words are used as a dictionary word and the image is divided into four rectangular areas. As showed in figure 3, each of the regions of the image, belong to a dictionary word. Each areas of image belong to a dictionary word that histogram of this is shown in the middle column of Fig 3. Histogram of the image is corresponds to the sum of the number of occurrences of each word or object in the image.

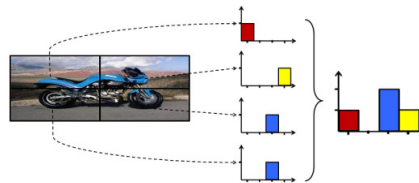


Figure 3: Crisp labeling

In Fig 4 shown a simple example of fuzzy weighting method. In the middle of it is shown weights belong each of the four regions of the image to the each of the vocabulary words. Unlike the previous case, each sub-image is belonging to every dictionary word. But with different weights,

for example, the first histogram that describe the top left regions of the image, shows the first word dictionary is the best word for describe this region, in addition the second and fourth words in the dictionary are describe this region too and should be considered appropriate weights for their in the histogram.

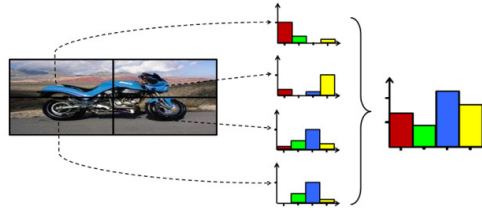


Figure 1: Fuzzy labeling

Carefully at Fig 3 shows that ignored similarity in the hard weighting method that regions in picture only belong to the closest word in the dictionary and with fixed weight. Using of ontologies allow to find semantic relation between the labels. In the present study semantic relation is to find the degree of similarity tags. To determine this, can be used two approaches: soft (fuzzy) scoring and hard (crisp) scoring that describe above. In continue is explained how to obtain the weight of each of the words to describe areas of the image. In the beginning create a matrix with the number of labels size (according to the data base $374 * 374$) and initial value one in the main diagonal and zero for other (in the crisp scoring only diagonal matrix elements are one and other are zero). According to the distance elements ontology increased the values in this matrix to 1. This amount is determined according to the number of common tag in ontology.

For example, detecting of tiger and cat is very difficult and in crisp scoring if label has detected cat just cat label give 1 score, But the soft scoring in addition to the cat, tiger which has a little distance with cat in the ontology, give 0.8 score and element (cat, tiger) in matrix is 0.8. Both cat and tiger are in the 5 th level in the ontology and have four common labels or parents so the score of element (cat, tiger) in matrix is 0.8. Other score will also be calculated as follows:

Consider the visual words ontology including the k word as $V = \{W_1, W_2... W_k\}$ and input image consists of M locally visual words that descriptor by the symbol P_j where j values from 1 to M varies. Carefully in this equation realize that the closest word in the ontology to visual word W_i would receive a weight equal to one and other words base on their distance from the visual word W_i give weights smaller than one. For example, Figure 9.

4. PROPOSED METHOD

The training part and testing part of the image annotation proposed method is shown in Fig (5). To generate automatic Image annotations system with the proper performance should be correctly and carefully chose the various components of the system. In Fig. 5 shown the proposed method consists of several components, which include:

- 1- Select indexed images database to train
- 2- Create ontology
- 3- Images Segmentation: the images in the database have segmented by the automatic segmentation algorithms such as JSEG or ISATOOL or method with intervention of human, so test images are being segmented firstly.
- 4- Feature extraction: should be extracted visual features of image for classification, to access better result and distinguishing images from together features must be selected and extracted appropriately.

5- Bag of words: when the number of samples in the categories to be much using of bag of words can be useful.

6- Features Selection: Due to the large number of features extracted from images use feature selection techniques to reduce the complexity and may even lead to improved results.

7- Annotation or Labelling the Images: to create best label for test image we use statistical method in local features and KNN classifier in global features and combined them to find probability of each label.

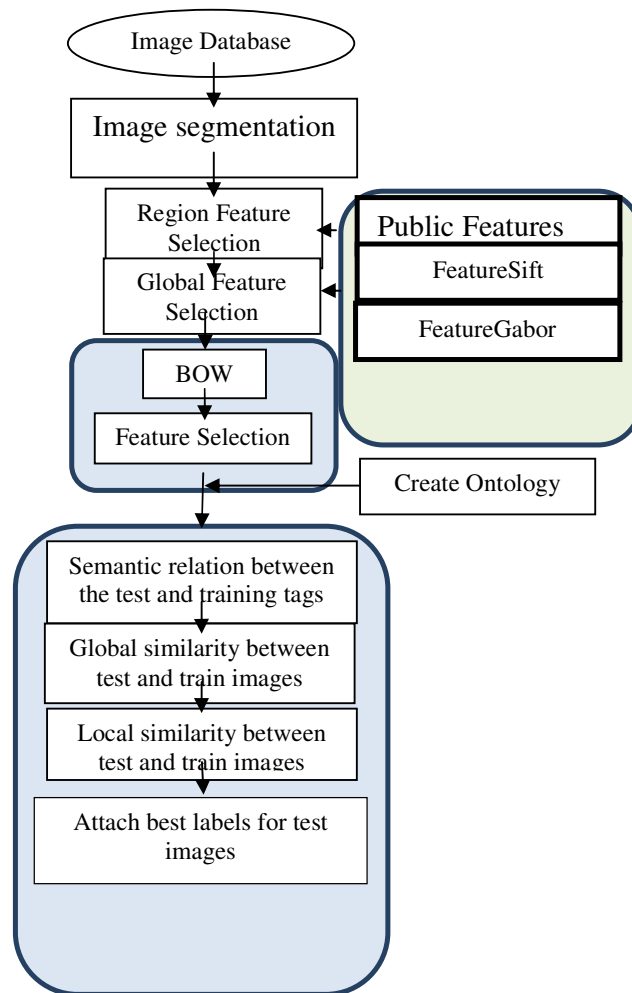


Figure 2: Block diagram of the proposed method

4.1. Image Segmentation

As mentioned in the previous section is used the Corel5k database for classification that in this database images are segmented to different object. In this database is used of JSEG algorithm [15] for the images segmentation. After segmentation of the image using different masks in image

to separated various components of image (For each segmented image create a mask that size of magnitude and latitude of the mask is same as image and initial value of this mask is zero and then the value of the place of the segment is one when complete the mask then with point multiple of mask and original image access to the image that is black for places does not belong to the segment and other points marked piece is the segmented image and ready for the next step). In Fig 6, is showed the image of database with isolated segmentation parts of image (every pieces of image is obtained with point multiple of masks value in the original image that is segmented with JSEG algorithm and human help).



Figure 3: Example of image segmentation

4.2 Feature Extraction

The most important part of artificial intelligence projects that have the training and testing parts that are related to the stage of feature extraction and classification. After entry initial data or pictures and doing pre-processing stage such as segmentation then considered certain methods of image feature extraction that are comparable. Output of these methods is numbers or vectors that describe the characteristics of the image, such as colour histogram and etc. Feature Extraction in this project has several parts:

4.2.1 General Features

The first part (General Features) is related to the features that exist on the site of ECCV_2002 conference for this database. Each image segment is represented by 36 features. Since each image has a different number of segments. Note that the features are redundant. Note also that the RGB and L*a*b features were duplicated to increase their weight for a specific experiment. The 36 features are:

- area, x, y, boundary/area, convexity, moment-of-inertia (6)
- ave RGB (6)
- RGB stdev (3)
- ave L*a*b (6)
- lab stdev (3)
- mean oriented energy, 30 degree increments (12)

4.2.2 Local Feature Extraction

The second part is the local features extraction of images, in addition of global features extraction to improve the accuracy of classification, use of local features extraction (according to various

references of image retrieval) like sift from whole of image is suggested¹. Local features of the images are showing more detail. The main steps related to sift algorithm to extract local features as follows:

1. Find the maximum and minimum in scale space: in this stage, search for local maximum and minimum in the all of scales that uses a Gaussian filter to detect desired points that are independent of scaling and rotation.
2. Focus on key points: in this stage, identify scale and location of the key points.
3. Determination orientation: determine one or more orientation for each key point based on local gradient.
4. Describe the key points: to describe the key points, local gradient image around the key point in the selected scale is calculated. The extracted features are independent of the orientation, scale and location.

4.2.3 Global Feature Extraction

In addition of the local features that described in above, some of the features can be extracted from the whole of the images. Global features can be located in general features classes but the global Features are extracted from the whole of image and general features are extracted from the separated segments of images. One of the major features in this property is Gabor feature.

4.2.4 Combine Features

With regard to the independence of the features, can be individually investigated the effects of any features to detect similarity between test and train images.

4.3 Bag of words

When the number of elements in the class is too high, the usage of bag of words is useful and important, result of using bag of words are savings in time and memory, for example, when the number of elements in each class is 10,000 and each of the elements have 150 features a matrix of $150 * 10,000$ for elements of this class should be considered but when using bag of words (as the representative of the 10,000 element are 100 center element, for recognition of 100 center we must using clustering within the 10,000 element and the number centers of these clusters are considered 100) dimensions of this matrix are $150 * 100$. Thus the feature vectors that are more similar to each other that are diagnosed in a cluster and have same treated. For each cluster, center is calculated and added to the dictionary as a visual word. In fig 12, is shown the concept of visual bag of words.

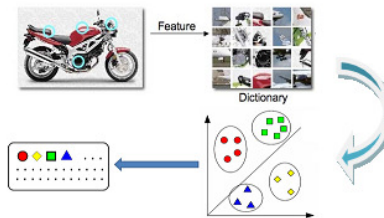


Figure 7: Create visual Bag of Word

The algorithm used in clustering algorithm is k mean, According to Fig 7, it is observed that the number of centers is significant and can increase the detection accuracy.

4.4 Features Selection

One of the important features of the proposed method in this paper is that combine features extracted from the image and prevention of climbing feature vector length. For this purpose in each stage of annotations, after image segmentation chosen best set of features with features selection algorithm.

These reduce the computational complexity of the program and sometimes enhance the accuracy of the proposed algorithm.

In this paper feature selection algorithm uses of genetic algorithm technique, in this algorithm fitness function is measure mean of F1. The results of applying this algorithm is selected 15 important features from 36 general features, 100 important features from 640 (according to BoW, 5 bag extracted from each image and each bag has 128 features) local (Sift) Features and 35 important features from 512 global (Gabor) features. Thus select features that are important at each stage of classification.

4.5 Classification

Annotation or classification is doing in two steps:

- The first step is classification base on generally, locally and globally features separately.
- The second step is combining the result of different classification in first step.

4.5.1 Local classification

Suppose we are given a collection C of un-annotated images. Each image $I \in C$ is represented by a discrete set of blob numbers, $I = \{b_1 \dots b_m\}$. In this section we develop a formal model that allows us to answer the following question:

Given an un-annotated image $I \in C$, how can we automatically assign meaningful keywords to that image?

We assume there exists a training collection T , of annotated images, where each image $J \in T$ has a dual representation in terms of both words and blobs:

$J = \{b_1 \dots b_m; w_1 \dots w_n\}$. Here $\{b_1 \dots b_m\}$ represents the blobs corresponding to regions of the image and $\{w_1 \dots w_n\}$ represents the words in the image caption. The number of blobs and words in each image (m and n) may be different from image to image. We only assume that a set of keywords $\{w_1 \dots w_n\}$ is related to the set of objects represented by blobs $\{b_1 \dots b_m\}$.

Suppose we are given an un-annotated image $I \in C$. We have the blob representation of that image $I = \{b_1 \dots b_m\}$, and want to automatically select a set of words $\{w_1 \dots w_n\}$ that accurately reflects the content of the image.

That is, we need to estimate the probability $P(w/I)$ for every word w in the vocabulary. The probability of drawing the word w is best approximated by the conditional probability of observing w given that we previously observed $b_1 \dots b_m$ as a random sample from the same distribution:

$$P(w|I) \approx P(w|b_1 \dots b_m)$$

We cannot use the prevalent maximum-likelihood estimator for that probability because the image representation $b_1 \dots b_k$ does not contain any words. However, we can use the training set T of annotated images to estimate the joint probability of observing the word w and the blobs $b_1 \dots b_m$ in the same image, and then marginalizing the distribution with respect to w . The joint distribution can be computed as the expectation over the images J in the training set:

$$P(w|b_1, \dots, b_m) = \sum_{J \in T} P(J) P(w, b_1, \dots, b_m | J)$$

We assume that the events of observing w and $b_1 \dots b_m$ are mutually independent once we pick the image J . This assumption follows directly from our earlier decision to model each image as an urn containing both words and blobs. Since the events are independent, we can rewrite above equation as follows:

$$p(w, b_1, \dots, b_m) = \sum_{J \in T} P(J) P(w|J) \prod_{i=1}^m P(b_i|J)$$

The prior probabilities $P(J)$ can be kept uniform over all images in T . Since the images J in the training set contain both words and blobs. Specially, the probability of drawing the word w or a blob b from the model of image J is given by:

$$p(w|J) = \frac{\text{num}(w, J)}{\text{num}(J)}$$

$$p(b|J) = \frac{\text{num}(b, J)}{\text{num}(J)}$$

Here, $\text{num}(w, J)$ denotes the actual number of times the word w occurs in the caption of image J (usually 0 or 1, since the same word is rarely used multiple times in a caption). Similarly, $\text{num}(b, J)$ reflects the actual number of times some region of the image J is labelled with blob b , $\text{num}(J)$ stands for the aggregate count of all words and blobs occurring in image J .

4.5.2 Global Classification

In this part of the proposed method are compared features of the test image with features of training images and test image labeled with the labels of annotated training images. In Local classification images used regional features and annotated a set of areas, in other words, use of areas independently. If the Global classification method considered the relationship between these areas. In general, the distribution of visual words cannot be independent of the distribution areas, then areas of distribution and distribution of visual words are dependent.

In view of the follow equation, it can be concluded that coefficient α show the similarity of test image I with train images. That in the proposed method reflects the distribution of words in the train that defines the test image. It is easily understood that the more like a test image to a train image, similar distribution of visual word in them, in other words the amount of distribution of visual word train image J or coefficient α should be bigger.

$$p(w|I) = \sum_{J \in T} P(w|J)\alpha_j$$

4.5.3 Combine Classifications

In this part of the proposed method, the results of the above combine together that suggested the follows equation:

Combine classification = A* Local classification + B*Global classification

Local classification = (Matrix Ontology)^C * P (w|I)

Global classification = (Matrix Ontology)^C * P (w|I)

As can be seen in the above, variables or classifications combine together by using the coefficients A, B. While using the created ontology in classifications and for control amount of ontology effect used a coefficient (c).

To obtain the variables A, B, C used genetic algorithm. F1 is fitness evaluation and the range of A, B is from 0 to 100 and in multiples of 5 and parameter C considered in the range of 5 to 25. The results of 500 time's genetic algorithm as following values:

A = 45

B = 10

C = 18

5. RESULTS

To review the results, existing database divided in to training and testing sets then extracting relevant features and then tagged classes or labels to the components of the images, so it is possible (depending on the number of components of the picture) each image has number labels.

5.1. Image database

We have used a set of Corel images consists of 5000 images to evaluate the proposed model. The image database consists of 50 main topics and each class includes 100 images. COREL-5K was first collected by Duygulu et al. [17] and it has been extensively used by other researchers, so it has been known as a standard dataset in image annotation researches. Each image is also associated with 1-5 keywords. Therefore, there are 374 keywords in the dataset. We partition the whole data set into a training set and a test set, to review the results, existing database divided into training (90 percent) and testing (10 percent) sets, which means 4500 training images and 500 test images. For the region features, we use the JSEG algorithm [23] to segment each image into 1 to 11 regions. Image regions with area less than 1/25 of the whole image are discarded. In average there are five image regions per image. Each image region is represented by a 36-D feature vector. For the dense grid, we sample 13×13 pixels image patches without overlapping. The average number of image patches per image is around 550. The image regions are clustered into 500 image blobs.

5.2 Experimental Results

It should be noted that to evaluate the performance of different methods used the measure of accuracy to compute the rate of correct diagnosis and misdiagnosis for each of the images.

The most common comparison criteria can be used in image annotation is measuring Precision, Recall, F1 and NZR, which are calculated using the following formula:

$$\text{Precision}(w) = \frac{r}{n}$$

$$\text{Recall}(W) = \frac{r}{N}$$

$$\text{F-Measure} = \frac{2 \times \text{Precision} \times \text{Recall}}{(\text{Precision} + \text{Recall})}$$

NZR = None Zero Recall

The above variables are defined as follows:

- r: number of **common words** annotated manually or automatically
- n: total number of words **annotated automatically** for each image
- N: Number of words **annotated manually** for each image

The number of words by Recall greater than zero is shown the number of correct words that predicted by an automatic annotation method. When Recall of a keyword is greater than zero that one or more is predicted correctly, otherwise the Recall won't greater than zero.

The results of the experiments show the effectiveness of the proposed approach. Figure 8, shows some of images with manually and automatically labels with proposed method.

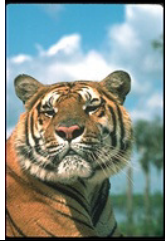



| | | | | |
|--|--|--|---|--|
| Image s |  |  |  |  |
| Huma n labels | Sky, head, cat, tiger | Flowers, stems, petals | sky, jet, plan | Sun, sea |
| Propo sed metho d label | Tiger, cat, sky, tree, elouds | Flower, sky, stems, grass, plants | Sky, plan, jet, sea, bird | Sun, sea, bird, sky, beach |

Figure 4: some of images with manually and automatically labels with proposed method

5.3 Compared with other methods

For comparison, the proposed method was compared with existing methods. In table 1 is shown results of NZR and average of precision, recall and F1 for different methods.

Table 1: Compare the proposed approach with other methods

| Methods | precision | recall | F1 | NZR |
|--|-----------|--------|------|-----|
| PLSA-FUSION-2010-[18] | 0.19 | 0.22 | 0.2 | 112 |
| TMIML-2010-[8] | 0.23 | 0.27 | 0.25 | 130 |
| JEC-2010-[1] | 0.27 | 0.32 | 0.29 | 139 |
| Lasso-2010-[1] | 0.24 | 0.29 | 0.26 | 127 |
| GMM-2011-[19] | 0.15 | 0.19 | 0.17 | 93 |
| GM-PLSA-2011-[17] | 0.25 | 0.26 | 0.25 | 125 |
| CKSM-2011-[20] | 0.29 | 0.35 | 0.31 | 147 |
| HDGM-2012-[9] | 0.29 | 0.3 | 0.3 | 146 |
| CVM-DP-2012-[21] | 0.35 | 0.25 | 0.29 | - |
| feature fusion -2014-[2] | 0.27 | 0.33 | 0.3 | 141 |
| DWML-kNN -2012-[11] | 0.28 | 0.33 | 0.3 | 133 |
| IAGA -2014-[12] | 0.3 | 0.327 | 0.31 | 132 |
| Deljooi-2015 – [16] | 0.3 | 0.31 | 0.3 | 134 |
| Proposed method without ontology | 0.3 | 0.32 | 0.3 | 130 |
| Proposed method only with Local Classification | 0.31 | 0.34 | 0.32 | 135 |
| Proposed method only with Global Classification | 0.31 | 0.34 | 0.32 | 129 |
| Proposed method | 0.32 | 0.35 | 0.33 | 139 |

As Table 1 shows, the accuracy of the proposed method compared to similar approaches and show the proposed method without the ontology has lower accuracy than other proposed methods, the original proposed method have a better performance.

In this paper used an algorithm that combine ontology model and extended CMRM to combine low-level features and high-level features image to display concepts. It has good flexibility and can easily be extended to larger areas. The ontology consists of concepts and associated structures that provides better results. The hierarchical structure eliminated the semantic gap.

6. CONCLUSIONS

In this paper, we have proposed a method for AIA (automatic image annotation) which is improved the result combine local and global classifications. Instead of using the direct distribution of regions in CMRM, we utilize the distribution of topics between regions. Moreover regional contexts are more accurate than textual contexts or the relationship between keywords in CMRM. The proposed approach combines the hierarchical structure and regional contexts and visual topics for automatic image annotation, as for the dependence between these two kinds of information in the image; we use the ontology approach based on WSM (Weighted Sum Method) to integrate them. The proposed method is tested on a 5000 Corel data set and the results show that utilization of regional contexts or considering the relationship between regions improves the performance significantly and its incorporation with the visual topics leads to the best performance.

REFERENCES

- [1] Einarsson S. H, (2004), "Data structures for intermediate search results in the Eff2 image retrieval system," Reykjavík University, technical report.
- [2] A. Tousch, S. Herbin, J. Audlibert, (2012) "Semantic Hierarchies for Image Annotation: A Survey," Elsevier Ltd., Pattern Recognition, vl.40, pp. 333-345.
- [3] Long F, Zhang H and Dagan Feng D, (2003), "Fundamentals of content-based image retrieval, in Multimedia Information Retrieval and Management Technological Fundamentals and Applications," Springer-Verlag, pp. 1-26.
- [4] D. Zhang, Md.M. Islam, G. Lu, (2011) "A review on automatic image annotation techniques," Elsevier Ltd., Pattern Recognition, vol. 45.
- [5] Li X, Chen S, Shyu M and Furht B, (2002), "An Effective Content-Based Visual Image Retrieval System," in 26th IEEE Computer Society International Computer Software and Applications Conference (COMPSAC), Oxford, pp. 914- 919.
- [6] Rui Y., Huang Th. S. and Chang Sh., (1999), "Image Retrieval: Current Techniques, Promising Directions, and Open Issues," Journal of Visual Communication and Image Representation, vol. 10, pp. 39–62.
- [8] S. J. Hwang, K. Grauman, and F. Sha. (2014), Semantic kernel forests from multiple taxonomies. Neural Information Processing Systems.
- [9] S. J. Hwang, K. Grauman, and F. Sha. (2014), "Analogy-preserving semantic embedding for visual object categorization". International Conference on Machine Learning, pages 639–647.
- [10] Mohammad Mehdi Farhangi, Mohsen Soryani, Mahmood Fathy, (2012): Improvement the Bag of Words Image Representation Using Spatial Information. ACITY (2): 681-690.
- [11] Guang-Hai Liu, Lei Zhang, Ying-Kun Hou, Zuo-Yong Li, Jing-Yu Yang, (2010) "Image retrieval based on multi-texton histogram", Pattern Recognition Journals of science direct, Volume 43, Issue 7, July, Pages 2380–2389.
- [12] HengameDeljooi, Ahmad R. Eskandari, (2014) "A Novel Semantic Statistical Model for Automatic Image Annotation Using the Relationship between the Regions Based on Multi-Criteria Decision Making", International Journal of Electrical and Computer Engineering (IJECE), Vol. 4, No. 1, February, pp. 37-51.
- [13] HengameDeljooi, SomayyeJafaraliJassbi,(2015) "A Multi Criteria Decision Making Based Approach for Semantic Image Annotation", International Journal of Computer-Aided Technologies (IJCAx) Vol.2, No.1.