IJCSE International Journal of Computer Sciences and Engineering Open Access
Research Paper Vol.-7, Issue-1, Jan 2019 E-ISSN: 2347-2693

Meta Analysis and Verification on Automated Image Tagging Techniques

S.Khoria

[1]CS Department, MPEC College, AKTU University, kanpur, India

Available online at: www.ijcseonline.org

Accepted: 25/Jan/2019, Published: 31/Jan/2019

Abstract— Automatic image tagging is an active topic of research in computer vision and pattern recognition. There is a huge urge in the Computer Vision community today to find ways to automatically annotate images. Machine learning techniques have facilitated image retrieval by automatically classifying and annotating images with keywords. Among them Support Vector Machines (SVMs) have been used extensively due to their generalization properties.

Keywords—Component, Formatting, Style, Styling, Insert (key words)

I. INTRODUCTION

Image annotation predicts a set of labels that are semantically similar to the given image. These labels are assigned from a predefined vocabulary set. Nowadays; with the frequent and easy access to the digital gadgets such as Digital Cameras and Mobiles etc., information in the form of digital images is increasing. Image based database is increasing leaps and bounds. To retrieve the unique image from the large image database, effective and efficient method is required. To retrieve an image from large image database is somehow very difficult task of image retrieval system. There are many methods proposed in the past to retrieve an image but still research has been going on to build an efficient method.Image can be retrieve by using visual low-level content such as shape, color, and texture or by using tags or keywords which are described by the semantic meaning of given image. To retrieve images using low-level visual features user needs to give an input as a query image and image retrieval gives set of images which are visually similar to given query image. But it is very difficult for many users to get query image each time which suffice their requirement. Content based image retrieval (CBIR) is a method which retrieves image based on low-level visual features. So to overcome problem of CBIR another method is to classify semantically all the images of the database as keywords. The entire database images are classified as a set of keywords and images can be retrieved based on these keywords. The main advantage of such method is that user can retrieve image in the same manner as they retrieve text document. One method is to manually classify all images; but it is very difficult and time consuming to classify large quantity of images manually, so some sort of automated method is required to perform this task. Automatic Image Annotation (AIA) is an automated method which maps low-level visual features for the high-level semantic features of the given image.

One method is content based image retrieval (CBIR) in which image is retrieved based on low level features like shape, color and texture. In this method user needs to apply query image in CBIR and similar images based on sample query image is retrieve by the system. But there is a semantic gap between low level visual feature and high level semantic concept that are used by the user. In manual image annotation, images are annotated manually by the user, so that images can be retrieved as easy as retrieving text document. This method is accurate but it is also inefficient because of the manual assignment of keywords to image, which is cumbersome and time consuming process. To overcome such problems of manual image annotation and to bridge the semantic gap, research in this area shifted to Automatic Image Annotation.

Section I contains the introduction of Automated image tagging , Section II contain the related work, Section III contain the some measures of methodology, Section IV contain the architecture and essential steps of ....., section V explain the ..... methodology with flow chart, Section VI describes results and discussion ....., Section VII contain the recommendation of ..... and Section VIII concludes research work with future directions).

II. RELATED WORK

Automatic Image Tagging has been active research topic in the last few years due its high impact on the Web search. To simplify the image retrieval metadata is added to images by an automatic image annotation. The general framework of automatic image annotation is shown in Figure A which

includes the following steps[15].

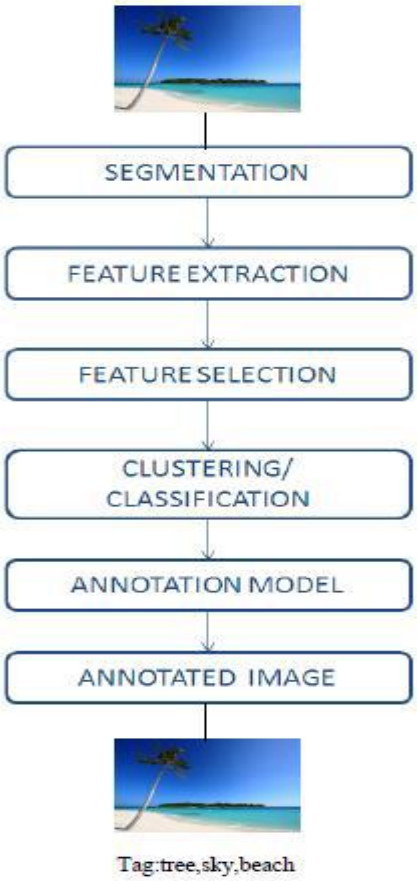


Figure A: Concept of Automatic Image Tagging

(1) Segmentation: In this step image is partitioned into group of pixels which are homogenous in nature. Image segmentation extracts visual features of image which can be merged or split in order to build objects of interest on which image analysis and interpretation can be performed.

(2) Feature Extraction: Low level visual information from segmented image is extracted using various feature descriptors like color, texture, shape. Color and texture are the most expressive to extract visual features of image.

(3) Feature Selection: It refers to the reducing high dimensional feature space to low dimensional feature space by using statistical techniques such as Principal Component Analysis and Particle Swarm Optimization Algorithm.

(4) Clustering/Classification: In this step the group of feature vectors is formed depending on efficient clustering techniques such as k means, fuzzy clustering clustering partitions the group of feature vector based on some specified common features and various similarity measures for image retrieval. The classification techniques such as k nearest neighbor, SVM, Decision Tree can also be used for grouping of feature vectors based on some predefined class label.

(5) Annotation Model: In this step annotation of testing image is done on the annotation model chosen such that labels are transferred from training to test images based on the annotation model specified used for the annotation of the image. Annotate the image based on annotation model such as probabilistic model, classification model, nonparametric approach or graph based approach. **Automatic Image Tagging Techniques**

Once images are represented with low level features using either global methods or region methods, higher level semantic can be learned from image samples. Early approach used relevance feedback to learn image semantic from humans. However, this approach has similar drawbacks to the traditional manual annotation approach. Therefore, the new trend is towards automatic image annotations. Assuming semantically labeled image samples are collected and represented with low level features, a machine learning algorithm can then be trained using the feature to semantic label matching. Once trained, the algorithm can be used to annotate new image samples. There are generally three types of AIA approaches. The first approach is the single labeling annotation using conventional classification methods. The second approach is the multi-labeling annotation which annotates an image with multiple concepts using the Bayesian methods. The third approach is the web based image annotation which uses metadata to annotate images.

- **Single labeling annotation using binary classification**

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. The output of the classifier is the semantic concept(s) which is used for image annotation. The idea of single labeling is equivalent to collective labeling, that is, instead of labeling images individually, images are first clustered and then labeled collectively. The common machine learning tools include support vector machines (SVM), artificial neural network (ANN), and decision tree (DT). In the following, we review each of these techniques.

Image annotation using support vector machines

Support vector machine (SVM) is a supervised classifier. It has been shown with high effectiveness in high dimensional data classifications, especially when the training dataset is small. SVM can classify both linear and non-linear data due to the use of kernel mapping. The advantage of SVM over other classifiers is that it achieves optimal class boundaries by finding the maximum distance between classes. It has been successfully applied to a number of classification problems, such as text classification, object recognition and image annotation.

An SVM classifier works by finding a hyper-plane from a training set of samples to separate them. Each training

sample is represented with a feature vector and a class label. The hyper-plane is learned in such a way that it can separate the largest portion of samples of the same class from all other samples. An SVM is basically a binary classifier. However, automatic image classification and annotation needs multiclass classifier. The most common approach is to train a separate SVM for each concept with each SVM generating a probability value. During the testing phase, the decisions from all classifiers are fused to get the final class label of a test image. Figure B shows this process. The complete classifier is a two levels process. The base level consists of multiple binary classifiers and the second level fuses the decisions from the base level
classifiers.

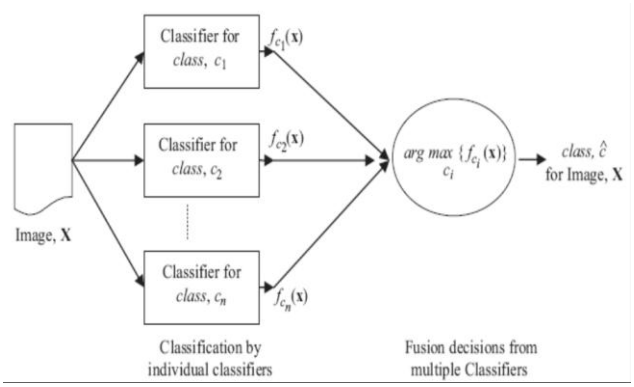
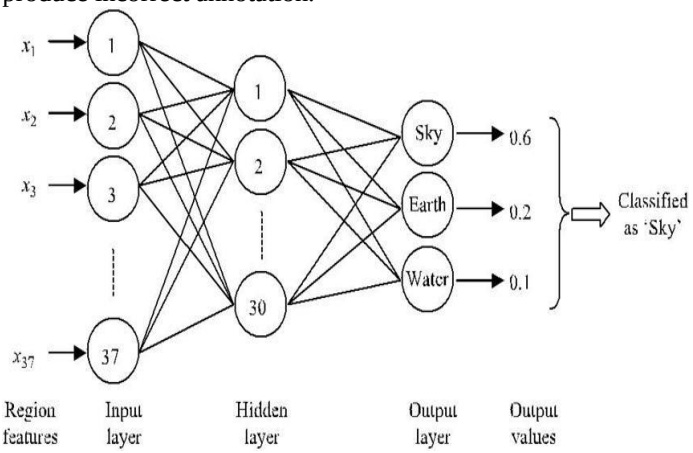


Figure B: Multiclass classifier using multiple binary SVM classifiers.

Image annotation using artificial neural network

An Artificial neural network (ANN) is a learning network that can learn from examples and can make decision for a new sample. Different from common classifiers which usually learn one class at a time, ANN can learn multiple classes at a time. An ANN consists of multiple layers of interconnected nodes, which are also known as neurons or perceptron's. Therefore, an ANN is also called multilayer perceptron (MLP). The first layer is the input layer which has neurons equal to the dimension of input sample. The number of neurons in the output layer is equal to the number of classes. This means, an ANN can learn multiple classes at a time, although single class ANN is also available. The choice of the number of hidden layers and the number of neurons at each hidden layer are open issues in ANN approaches. These numbers are usually selected empirically. The connecting edges between neurons of different layers are associated with weights. Each neuron works as a processing element and is governed by an activation function which generates output based on the weights of the connecting edges and the outputs of the neurons at the previous layers. During the training, ANN learns the edge weights so that overall learning error is minimized. While classifying a new sample, each output neuron generates a confidence measure and the class corresponding to the maximum measure indicates the decision about the sample.

An ANN can be used both for explicit classification of images, regions or pixels or implicit assignment of fuzzy decisions on images. Frate et al. use a 4-layer ANN to classify pixels of satellite images into one of the four categories: vegetation, asphalt, building, and bare soil. Based on the optimal experimental performance, they use a network of two hidden layers each consisting of 20 neurons. Kim et al. classifies images into object and non-object images using a 3-layer ANN. Instead of segmenting an image, the Centre 25% of the image is used to represent the image content. It assumes that the Centre part significantly characterizes the entire image. Because of this simplified assumption, the system cannot classify an image properly if the object appears in the other part of the image. A similar assumption is made in Park et al. about the object importance. Park et al. use segmentation algorithm to segment an image into regions and use the largest region at the Centre of the image to identify the image. The regions with similar color distribution of the central region are regarded as foreground (object) regions. The foreground regions are used to extract statistical texture features which are fed to a 3-layer ANN to classify the image into one of 30 concepts. The network consists of 49 neurons in the hidden layer. The drawback of the two approaches is that they may miss important objects from other parts of the image. For example, in a sunset/sunrise image, the sun often appears in the upper corner of the image. Furthermore, object regions may not necessarily be the largest region. In that case, the system will produce incorrect annotation.



Kuroda and Hagiwara use four different 3-layer ANNs to hierarchically classify image regions. The numbers of neurons used in the hidden layers of the four networks are 30, 10, 20, and 20, respectively. Figure C shows how the first network classifies an image region into one of the three broad categories such as sky, water, and earth. 37 dimensional region features are fed into the input layer. Each node of the output layer corresponds to one of the classes, e.g., sky, water, and earth and produces a likelihood value. The class of the input region is determined by the maximum likelihood value. Sky and earth regions are further classified into more 21 specific classes using the other two ANNs. For

example, a sky region is classified into one of five detailed categories: blue sky, cloud, sunset, night, and light. Similarly,

Figure C: classifying a region using ANN

An earth region is classified into one of nine more specific categories. The fourth ANN does not classify any region. Instead, it associates an image with a vector of 18 dimensions: each dimension measures the degree of certain global characteristics of the image, for example, bright/dark, rural/urban, and busy/plain, etc.

The neural network has the advantage that the outputs of output layer neurons are determined by the previous layers and the connecting edges. It does not need any other parameter tuning or any assumption about the feature distribution. However, there are several essential issues with ANNs. First, the classification accuracy depends on the number of hidden layers and neurons. Second, in most ANN research works, the numbers of hidden layers and neurons are not justified. Third, the choice of appropriate activation functions for the neurons is also an issue. Fourth, the training (finding the optimal edge weights) takes long time and it can fall into local optima. Fifth, an ANN works like a black box which means that the exact relation between the input and output is not transparent.

Image annotation using Decision Tree

A decision tree (DT) is a multi-stage decision making or classification tool. Depending on the number of decisions made at each internal node of the tree, a DT can be called binary or n-ary tree. Different from other classification models whose 22 input–output relationships are difficult to describe, the input–output relationship in a DT can be expressed using human understandable rules, e.g., if–then rules.

A DT is trained using a set of labeled training samples. Samples are represented with a number of attributes. During training, a DT is built by recursively dividing the training samples into non-overlapping sets, and every time the samples are divided, the attribute used for the division is discarded. The procedure continues until all samples of a group belonging to the same class or the tree reaches its maximum depth when no attribute remains to separate them. Figure D shows the process. The tree has two types of nodes: internal and leaf node. Each internal node is associated with a decision governed by a certain attribute which divides the training samples most effectively. Each leaf node represents the outcome (class) of the majority samples that follow the path from the root of the tree to the corresponding leaf. The leaf nodes can be expressed with unique if–then–else rules. For example, the decision tree of Figure D can also be expressed using the if–then rules.

If color = Red and texture = Three edge – Leaf1
Then outcome = Red Triangle
If color = Red and texture = four edges – Leaf2
Then outcome = Red square
If color = blue and shape = square – Leaf3
Then outcome = blue square
If color = blue and shape = circular – Leaf4
Then outcome = blue circle 23

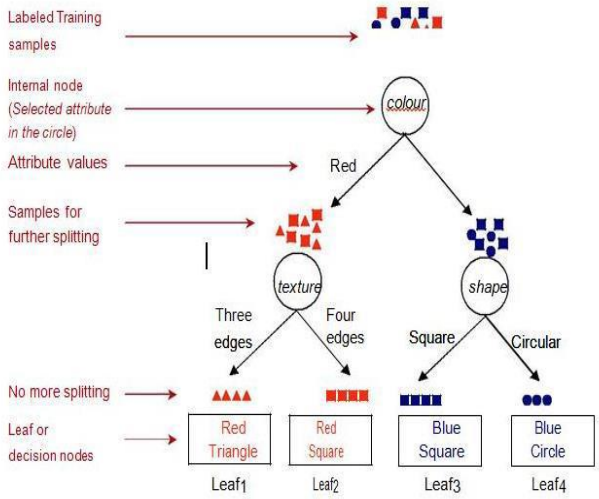


Figure D : Decision tree learning

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. Several DT algorithms are used in the literature, including ID3, C4.5 and CART. These DTs differ by the type of attributes, the attribute selection criteria, the outcome, etc. ID3 is the simplest DT algorithm that works only with discretized attributes. On the other hand C4.5 and CART can work with both discredited and continuous attributes.

- **Multi-labeling annotation using Bayesian methods**

Different from the binary classification approaches, multiple labeling methods annotate an image with multiple semantic concepts/categories. The concept of multi-labeling approach is related to the multi-instance learning, or more specifically, multi-instance multi-label (MIML) learning. In MIML, an image is represented with a bag of features or a bag of regions (multiple instances). The image is annotated with a concept label if any of the regions/instances in the bag is associated with the label. As a result, an image is annotated with multiple labels. A typical MIML is achieved using probabilistic tools such as the Bayesian methods. the Bayesian methods work by finding the posterior probability that an image belongs to any particular concept, given the observation of certain features from the image or region. This makes it 24 possible to assign an image to multiple concepts and rank images with the same concept according to the probabilities. Given a set of images { , ……., } from a set of given semantic classes { , ,……., } Bayesian models try to determine the posterior probability from the conditional probabilities and the priors. Suppose, an image I is represented by the feature vector x. Given prior probabilities P ( ) and conditional probability densities P (x|), the

probability of an unknown image I belonging to class is determined by:

$$P(c_i|X) = \frac{p(x|c_i)p(c_i)}{P(x)}$$

From above equation it can be seen that a Bayesian framework essentially has four components: one output component P (c_i |X) and three input components: P (c_i), and P(x). Because the distribution p(x) is usually uniform for all classes, the class of image I can be decided using the maximizing a posterior (MAP) criterion. There are generally two types of approaches to model the conditional probabilities; first one is non-parametric approach and other is parametric approach.

Non-parametric approach

In this approach, the conditional probabilities are calculated without any prior assumption about the distribution of the image features. Rather, the actual feature distribution is learned from the features of the training samples using certain statistics. In practice, the image features are first quantized into clusters using a certain clustering algorithm. Next, the continuous features are replaced by the cluster centroids. This process discretizes the image feature space.

The complete annotation process of this approach is shown in Figure E Given a new image, its features are extracted and compared with cluster Centre's. The closest clusters Centre's are selected. The conditional probability models corresponding to the selected cluster Centre's are then used to calculate the posterior probabilities. The MAP criterion is then used to annotate the new image.

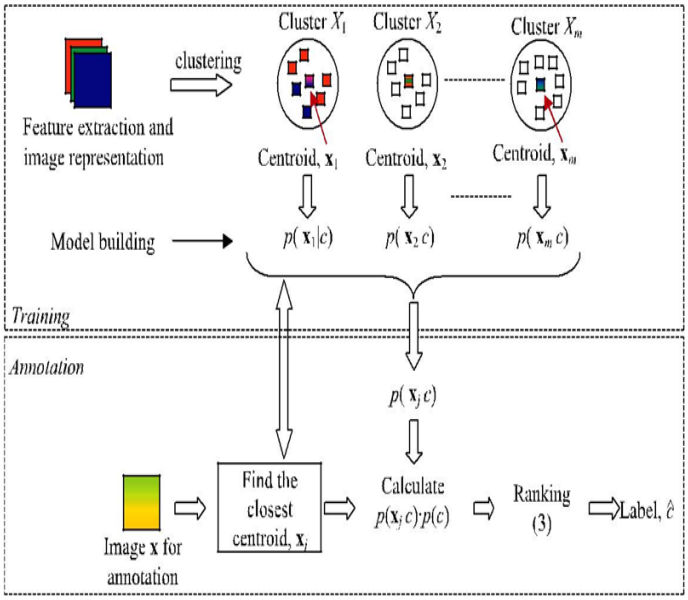


Figure E: the general Bayesian Annotation Model

Parametric approach

In this approach, the feature space is assumed to follow a certain type of known continuous distribution. Therefore, the conditional probability P (x|c) is modeled using this feature distribution. The general process is similar to that shown in Figure E Features or regions are first clustered and quantized; the conditional probability model is then built for each cluster (or blob).

Image annotation incorporating metadata

The WWW is a rich source of both imagery and text information. Web images often come with text descriptions, URL, HTML code, etc. The web information can be used for image annotation and retrieval. A number of techniques have been developed for annotating web images, most of them integrating both metadata and visual features for accurate image annotation. Therefore, these methods can be called hybrid methods. Cai et al. proposed a two level annotation and clustering mechanism: textual clustering for semantic annotation and visual clustering for reorganization of images within each semantic category. Images from web pages are first represented using three types of features: textual features (derived from surrounding text), link graph (derived from three complex hyperlink matrices) and visual features (derived from color moments on local Fourier transform). The textual features and link graph are used to cluster images into semantic category which is equivalent to annotation. 26 However, images within each of the semantic categories may not be perceptually similar. Therefore, they apply a second level of clustering on each of the semantic categories to reorganize the images into clusters based on visual features. The major issue with this method is that the textual features especially the link graph features are not reliable, as shown in existing image search engines.

As annotations from text description can be noisy, these annotations need refinement. This is especially needed for web based image annotation, because each image is usually annotated with multiple words which may not be related to each other. In a refinement stage, it preserves the annotations which are strongly correlated and rejects those which are not so strongly related to each other. Jin et al. use WordNet for annotation refinement. WordNet is an online lexicon where more than 150 K words are hierarchically organized. The words in WordNet maintain 'is a kind of' or 'is a part of' relationships which are used to find similarity between words. After getting the annotations of an image using any existing method, Jin et al use WordNet to calculate the total similarity of each word to other words of the annotation set. If the similarity is below certain threshold, it is discarded. The principle is that the words, which are very similar they first cluster a short listed web images based on their textual and visual similarity. The available words of each cluster are ranked based on text ranking technique. The rankings of words from all the clusters are fused together to generate a final ranking. Existing annotation approaches can use this ranking to decide which concepts should be learned because the top ranked words in the list are easier to learn than the bottom ones. The problem is that the ranking needs to be learned every time the database changes [1].

III. METHODOLOGY

The sequence of processing steps in automatic image tagging systems, namely image pre-processing, Feature Extraction and Region Segmentation, Artificial neural network. Automatic Image Tagging system is mainly divides into two parts: the first part is training a classifier network with use of a training image database, and the second part is testing a network for sample query images. These two parts are described here.

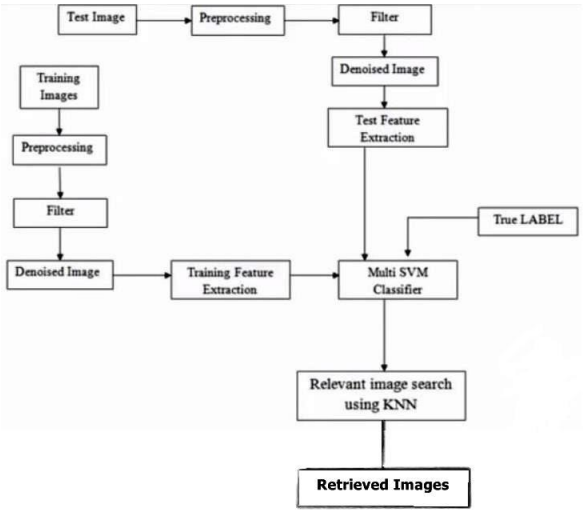


Figure F : Flow Diagram of Automatic Image Tagging System

The Proposed Approch

The procedure that we followed consists of doing some preprocessing on the image database to extract a dictionary of keywords we wished to use. The next step involved extracting the relevant features from the image and then use these extracted features to segment the images into regions. Once this has been done the last step is to train a set of neural networks to learn the association between the segmented regions of an image and the annotation keywords for that image which are present in the dictionary that was obtained during the preprocessing step. Figure G summarizes the entire training approach [11].

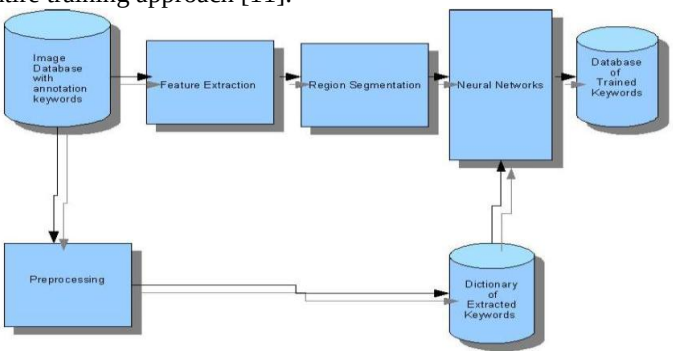


Figure G: The Training Process

COREL Dataset

The COREL dataset is a very commonly used dataset for the purpose of image tagging algorithms having been evaluated on this set of images. The dataset contains 5,000 images where 100 images on the same keywords. Each image in the dataset contains between 1 and 5 keywords, with a total of 374 keywords in the vocabulary, 371 words of which are present in the training dataset. If we include only those words that exist in the testing set then the vocabulary size is reduced further to 260 words. Most authors appear to split the dataset into 4500 training images, and 500 test images [17].

Image Pre-processing

After initial preprocessing on the coral image database, it was discovered that all the images in the database could be annotated with 374 unique keywords. Our initial approach was to determine appropriate semantic concepts that could be constructed from this dictionary of keywords. The idea was to then create profiling models for each of these concepts. After considering various approaches, finally a word cooccurrence matrix approach to determine concepts was implemented. A few of the examples of the concepts that were extracted for this database are city, mountain, sky, sun, water, clouds, tree, lake, sea, beach, boats, people, branch, leaf, grass, palm, horizon, hills, waves, birds to name a few. But once the concepts were successfully extracted it was difficult to map images to concepts because it is difficult to heuristically determine as to which image can be mapped to which concept, based on the annotation keywords of the images. To overcome this difficulty we finally created a dictionary of keywords based on their frequency of occurrence in the image database. Pre-processing is a common name for operations with images at the lowest level of abstraction- both input and output are intensity images. These iconic images are usually of the same kind as the original data captured by the sensor, with an intensity image usually represented by a matrix or matrice; of image function values (brightnesses).The aim of pre-processing is an improvement of the image data that suppresses unwanted distortions or enhances some image features important for further processing, although geometric transformations of images (e.g., rotation, scaling, translation) are also classified as pre-processing methods here since similar techniques are used. Four categories of image pre-processing methods according to the size of the pixel neighborhood that is used for the calculation of a new pixel brightness:

- pixel brightness transformations,
- geometric transformations,
- pre-processing methods that use a local neighborhood of the processed pixel, and
- image restoration that requires knowledge about the entire image.

Feature Extraction

In the second step, features are extracted out of the image regions which are choosen from the first step. features can be global or local. When the region is choosen to be the whole

image, features are global, describing the whole image. When the region is choosen to be a partition, segment or salient region, features are local, describing individual parts of the image. Features can also be categorised as being general or domain-specific. General features include commonly used features such as color, shape and texture. However, for special applications such as fingerprint recognition and lipreading, general features are not applicable, so domain-specific features have to be developed.Theoretically, the combination of different kinds of features will produce a more robust image description. In the following, some different image features used for image auto-annotation are described.

Color

Color is perhaps the most popular choice of visual features. It can be expressed in many different kinds of color-spaces, such as the most widely used RGB space. RGB representation is in wide-spread use mostly because it describes an image in its literal colour properties. The color histogram, which can be calculated both globally and locally, is one of the most widely used colour descriptors. It is calculated by discretising the colour space first and then counting the number of occurrence of each discretised colour in the image. Because histograms are accumulated over the whole image or region, with no information about locations, they are obviously invariant to translation and rotation of objects.

Shape

Shape features descibe the silhouettes of objects, so it requires the objects to be segmented out firstly. As we have mentioned, automatic segmentation techniques are still immature nowadays, the effectiveness of using shape descriptors in image auto-annotation applications is limited. However, shape plays an important role in some narrow domains such as trademark retrieval and recognition.

Texture

The definition of texture is vague, as Tuceryan and Jain (1993) said ”we recognize texture when we see it but it is very difficult to define”. They listed some example definitions attempted by different researchers. We understand texture as homogeneous visual patterns in images that manifest some kind of coherence or periodicity, such as wallpaper and bricks.

This is the first step in the training procedure once the preprocessing is complete. In image classification and retrieval, images are represented using low level features. Because an image is an unstructured array of pixels, the first step in semantic understanding is to extract efficient and effective visual features from these pixels. Appropriate feature representation significantly improves the performance of the semantic learning techniques.While both global and region based image representations are used in the existing image retrieval techniques, the trend is towards using region based features. Region based feature extraction needs prior image segmentation while global features are directly calculated from the whole image.

The process of feature extraction follows from the procedure used in Wang,Li et al, in their work on the development of the SIMPLIcity system. This is also similar to the feature extraction procedure implemented by Li and Wang in. The process of feature extraction involves the extraction of a 6 element vector for each block of the image after having divided the entire image into regions of blocks of size 4X4. Three elements of this vector are used to represent the color information in a 4X4 block and the other three elements are used to represent texture information over the 4X4 block. The feature extraction for the color information is done in the L,u,v space. This involves in converting the image from the default RGB color space to the L,u,v color space. The first element of the vector consists of the average value of the L component over a 4X4 pixel region of the image, the second element consists of the average value of the u component over a 4X4 pixel region of the image and the third element similarly consists of the average value of the v component over a 4X4 pixel region of the image. For the extraction of texture a single level Daubechies' 4 wavelet transform is performed,on each 4X4 block to obtain four frequency bands each of size 2X2 corresponding to LL,LH,HL and the HH bands [11].

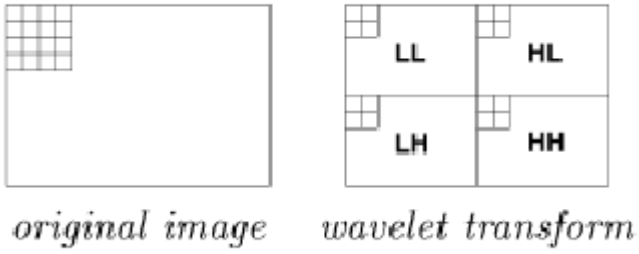


Figure H:Wavelet Transform

The three texture feature elements were obtained by computing the squared average of the wavelet coefficients of the HL,LH and HH bands. Thus each 4X4 block can be completely represented by a six element vector.

Region Segmentation

The second step in our training approach is to cluster the image into a set of 16 clusters that describe the image. We used the K-means clustering algorithm to perform the task of clustering to limit our output clusters to 16 clusters as we mentioned.

The idea of using 16 clusters for each image is to unify the representation of all images to have the same number of attributes to prepare a uniquely sized input vector for the next step in the training approach, which is the neural network part. We added the spatial features (x, y coordinates) to make the clustering more robust and thus we have ended with 8 features of which three color features, three texture features and two spatial features.

K-means clustering Algorithm

The k-Means is an algorithm to cluster objects, or data points, into k partitions, or clusters. The clusters are discoverd through an refinement process that updates the position of clusters iteratively. During each iteration, all the training points are assigned to the closest cluster based on the distance to the cluster centroid. Then, the centroid of each cluster is updated by the new cluster centroid which is calculated as the centroid of all the points that belong to it. The process is repeated until the points no longer switch clusters, or after a pre-defined number of iterations. Decisions on the value of k and the starting cluster centroids are essential to the performance of k-Means. A common choice of the initial centroids is to choose k sample points at random and use them as the centroids.

There are always K clusters. There is always at least one item in each cluster. The clusters are non-hierarchical and they do not overlap. Every member of a cluster is closer to its cluster than any other cluster because closeness does not always involve the „centre" of clusters. Kmeans clustering in particular when using heuristics such as Lloyd's algorithm is rather easy to implement and apply even on large data sets. As such, it has been successfully used in various topics, ranging from market segmentation, computer vision and astronomy to agriculture. It often is used as a pre-processing step for other algorithms, for example to find a starting configuration. In statistics and data mining, k-means clustering is a method of cluster analysis which aims to partition n observations into k clusters in which each observation belongs to the cluster with the nearest mean.

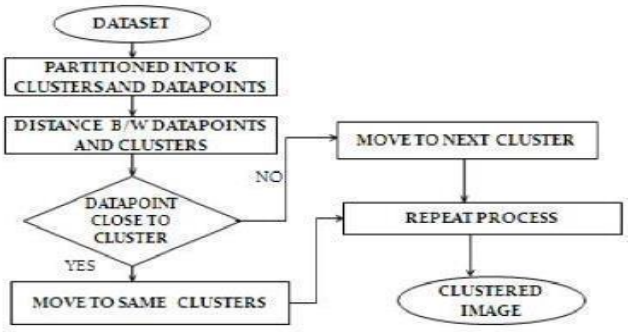


Figure I: Flow Chart for K-Means Algorithm

K-Means algorithm is an unsupervised clustering algorithm that classifies the input data points into multiple classes based on their inherent distance from each other. The algorithm assumes that the data features form a vector space and tries to find natural clustering in them .The points are clustered around centroids $\mu_i \forall_i =$ 1……k which are obtained by minimizing the objective

$$V = \sum_{i=1}^{k} \sum_{x_j \in S_i} (x_j - \mu_i)^2$$

where there are k clusters , i = 1,2,……, k and is the centroid or mean point of all the points $x_j \in S_i$. The algorithm takes a 2 dimensional image as input. Various steps in the algorithm are as follows:

1. Compute the intensity distribution(also called the histogram) of the intensities.
2. Initialize the centroids with k random intensities.
3. Repeat the following steps until the cluster labels of the image does not change anymore.
4. Cluster the points based on distance of their intensities from the centroid intensities.

$$C^{(i)} := \arg\min_j \left\| x^{(i)} - \mu_j \right\|^2$$

5. Compute the new centroid for each of the clusters.

$$\mu_i = \frac{\sum_{i=1}^{m} 1\{c_{(i)}=j\} x^{(i)}}{\sum_{i=1}^{m} 1\{c_{(i)}=j\}}$$

where k is a parameter of the algorithm (the number of clusters to be found), i iterates over the all the intensities, j iterates over all the centroids and are the centroid intensities.[13]

Artificial Neural Networks

ANNs are information processing system inspired by the ability of the human brain to learn from observation. ANN is a learning network which consists of multiple layers of interconnected nodes, which are also known as neurons or perceptions. An ANN can learn from example and make decision for a new sample. Different from other common classifiers which usually learn one class at a time, ANN can learn multiple classes at a time. The last step in our training approach is to build and train a neural network to learn relationship between image segments and the annotation keywords.

We chose to use neural networks to do the learning because, firstly, it typically takes a very little amount of time for simulating of the network, and using a neural network also allows us to learn more than one concept at the same time, thereby not requiring us to learn each concept alone. In our case since we have a dictionary of 374 words, it would be extremely time consuming to learn all such concepts individually. The usage of the Neural Network toolbox in MATLAB makes it extremely convenient, to make a neural network network learn such associations between keywords and image segments and also eliminates the time spent in debugging and testing unreliable code. The last and the most important reason made us use a neural network is the fact that it can be dealt with it as a 'black-box' which actually does the learning.

We processed more than 100 images of the training set successfully and for each one of these images we performed the feature extraction and region segmentation steps that were mentioned. Each image is thus described as an input-output vector pair. The input to the neural network is a vector obtained as the result of the performing region segmentation using K-means, so we have 374 attributes vector to describe the input of the image to the neural network, and for the target output we used a 1748 element vector, so that a 1 in a particular position would indicate the presence of an annotation keyword for that image, corresponding to an entry

of that particular keyword in the dictionary that was obtained during the preprocessing step. Therefore the input to the neural net is a 374*100 matrix representing the entire training set and a corresponding 1748*100 matrix representing the output. In the final approach we used feed forward networks with one hidden layer consisting of neurons and we used backpropagation algorithm to train the networkRelevant details should be given including experimental design and the technique (s) used along with appropriate statistical methods used clearly along with the year of experimentation (field and laboratory).

IV. RESULTS AND DISCUSSION

EVALUATION

The performance of annotation system is calculated by using precision and recall. Precision and Recall values in annotation system are evaluated for each word and the mean of all words are considered as the performance of the system. Accordingly,

$$\text{Precision} = \frac{\text{no of correct annotation label}}{\text{total annotated label}}$$

$$\text{Recall} = \frac{\text{no of correct annotated label}}{\text{total label in testset}}$$

Once the test images are labeled by auto-annotation systems, annotation qualities need to be assessed for performance comparisons between different systems. A number of evaluation metrics have been used by researchers, some of which are introduced here [15].

Precision and Recall

The most popular metrics for comparing different information retrieval systems precision and recall are also widely adopted for evaluating the effectiveness of auto-annotation approaches. In the information retrieval community, precision of a query is defined as the ratio of the number of relevant documents that are returned by the system to the total number of documents returned, and recall is defined as the ratio of the number of relevant documents returned to the total number of relevant documents in the database. There are two versions, per-image based and per-word based in automatic image annotation approaches to evaluate its effectiveness.

Per-image Precision and Recall

Per-image precision and recall are calculated on the basis of a single test image. For each test image, precision is defined as the ratio of the number of words that are correctly predicted to the total number of words predicted, and recall is the ratio of the number of words that are correctly predicted to the number of words in the ground- truth or manual annotations. Mathematically, they are calculated as follows.

Per Image Recall = $\frac{r}{n}$

Where

r: the number of correctly predicted words
n: the number of manual labels in the test image
w : the number of wrongly predicted words

Per-image precision and recall values are averaged over the whole set of test images to generate the mean per-image precision and recall.

Per-word Precision and Recall

Duygulu et al[16] used mean per-word precision and recall to evaluate their annotation effectiveness. Per-word precision and recall are calculated on the basis of each keyword in the vocabulary. Specifically, precision is defined as the number of images correctly predicted with a given word, divided by the total number of images predicted with this word. Recall is defined as the number of images correctly predicted with a given word, divided by the total number of images having this word in its ground-truth or manual annotations. The values of precision and recall are averaged over the words in the vocabulary to generate the mean per-word precision and recall.

Keyword Number with Recall>0

Duygulu et al [16] also used the keyword number with recall>0 to show the diversity of correct words that can be predicted by the automatic image annotation.

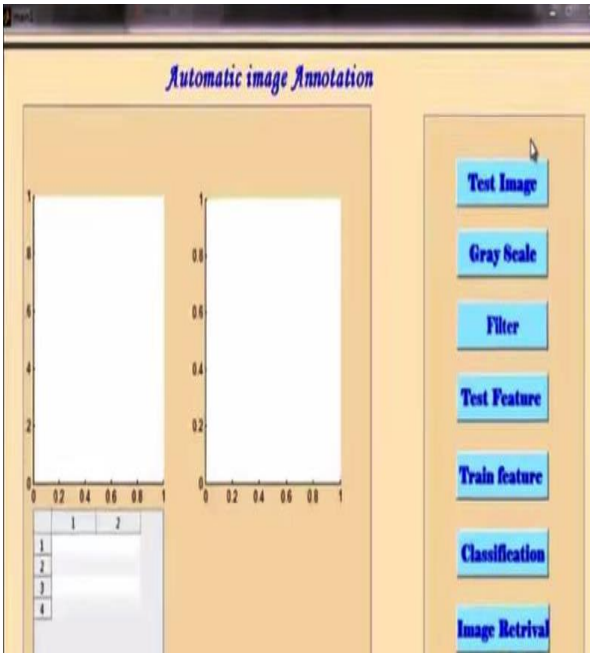
V. CONCLUSION AND FUTURE SCOPE

Conclusions

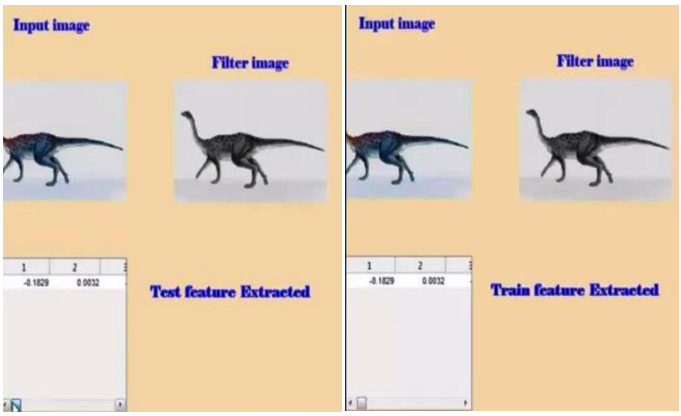
This thesis has introduced a number of approaches and techniques for automatic image tagging, which is a subject that is receiving rapidly increasing attentions in recent years. One of the most important aims of automatic image tagging is to bridge the semantic gap, which is considered as a vital problem existing in the traditional content based image retrieval systems.

Automatic Image Tagging or Automatic Image Annotation (AIA) gives better performance in image retrieval system. In image retrieval system, image can be easily retrieve using keywords if images are annotated by appropriate keywords. In AIA method image is automatically classify by keywords using classification network. In this thesis, we focused on a neural network model as a classification network of AIA system. Neural Network gives better performance when image is classifies by more than one class i.e. keywords. We can easily train neural network than other classification network and once a neural network is trained we can easily annotate sample images. It is not necessary that AIA system gives satisfactory result each time, so various annotation refinement algorithms proposed to refine the result of classification network.

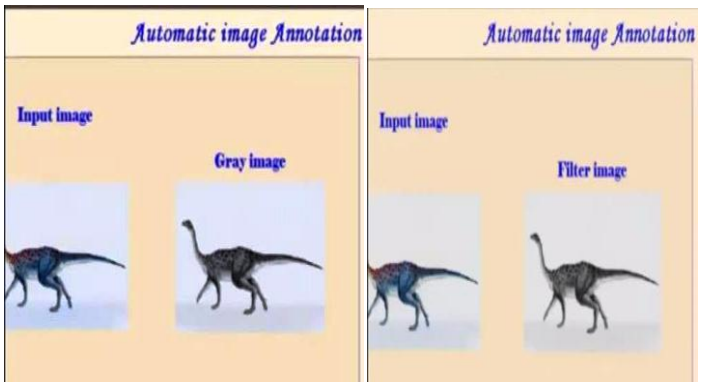
Experimental Results:



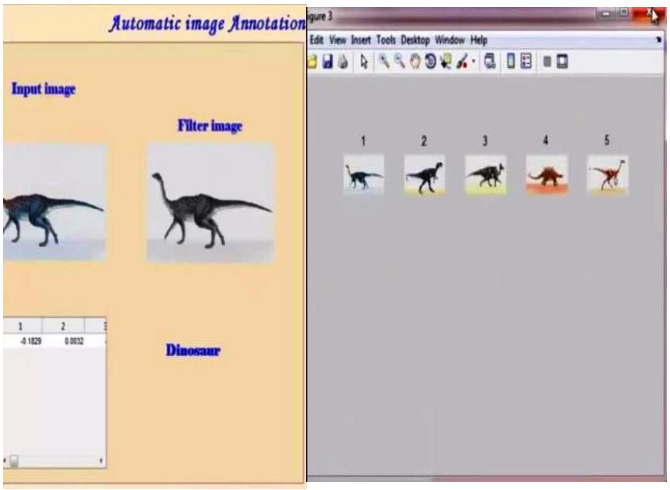
1/24/2019 Automatic Image Tagging 1



3/3/2010 Automatic Image Tagging 23



1/24/2019 Automatic Image Tagging 22



1/24/2019 Automatic Image Tagging 24

Future Scope

Automatic Image Tagging is a very challenging research area. There are several major issues in Automatic image tagging research.

The **first** issue is **high dimensional feature analysis**. Currently The **first** issue is **high dimensional feature analysis**. Currently, all existing features have limitations of describing images and none of existing features is powerful enough to represent the large variety of images in nature. Common practice is to combine several types of features to represent as many images as possible. However, the

processing and analysing of high dimensional image features is a very complex issue. Due to the 'curse of dimensionality', the perfor-mance of classifiers degrades dramatically when feature dimen- sion is too high.Therefore, features need to be further mined to select the right number of features and right features for annotation. The recent advance in sub space research offers promising solution in this regard.

The **second** issue is **how to build an effective annotation model**. Most existing AIA models are learned from low level image features. However, due to the 'combinatorial explosion' of required image to build an annotation model, the number of sample images is not large enough to train an accurate model. Therefore, textual information or metadata should be employed to improve annotation accuracy. However, very often, metadata is either not accurate or not adequate. How to integrate both low level visual information and high level textual information into a coherent annotation model is a challenging issue.

The **third** issue is **that currently annotation and ranking are done online simultaneously in the multiple labeling annotation approaches**. This is not efficient for image retrieval. The alternative is to do the annotation offline as in the single labeling approach and separate the ranking from annotation, that is, images are first annotated with a concept/category and ranking is also done offline after annotation. Once the images are annotated and ranked offline, retrieval is instant. The **fourth** issue is **how to rank images within each of the categories resulted from the single labeling techniques, so as to improve retrieval accuracy**. Since images within each category show certain distribution pattern, a Gaussian mixture model followed by an MAP ranking offers a practical solution. 41 The **fifth** issue is **the lack of standard vocabulary and taxonomy for annotation**. At this moment, arbitrary vocabularies are used in AIA literature. It is not known how images should be categorized. A hierarchical modeling of image semantics is needed to categories images properly. A hierarchical taxonomy not only standardizes the annotation vocabulary but also allows step by step annotation which is more practical.

Finally, there is no commonly acceptable image database for AIA training and evaluation. All AIA methods require a large number of pre-labeled image samples for training the model. At this moment, different AIA methods use different image datasets for training and evaluation, making it difficult to compare the performance. The database issue is closely related to the taxonomy issue. If a standard taxonomy of image semantics is available, a standard database can also be created accordingly.

All these issues point to future research directions in Automatic Image Tagging area

REFERENCES

[1] Dengsheng Zhang, Md. Monirul Islam, Guojun Lu. "A review on automatic image annotation techniques" .(2012)

[2] Nasullah Khalid Alham, Maozhen Li , Yang Liu, Suhel Hammoud . "A MapReduce-based distributed SVM algorithm for automatic image annotation" .

[3] Minmin Chen, Alice Zheng, and Kilian Q. Weinberger, "Fast Image Tagging".

[4] A.W.M. Smeulders, M. Worring, S. Santini, A. Gupta, and R. Jain . "Content-based image retrieval at the end of the early years . Pattern Analysis and Machine Intelligence, IEEE".

[5] X. Qi and Y. Han. "Incorporating multiple svms for automatic image annotation . Pattern Recognition", 40(2):728–741, February 2007.

[6] A. Yavlinsky, E. Schofield, and S. Rger . "Automated image annotation using global features and robust nonparametric density estimation". In International Conference on Image and Video Retrieval, pages 507–517. Springer, 2005.

[7] O. Chapelle, P. Haffner, and V. N. Vapnik. "Support vector machines for histogram-based image classification". Neural Networks, IEEE Transactions on, 10(5):1055–1064, 1999.

[8] V. Lavrenko, R. Manmatha, and J. Jeon. "A model for learning the semantics of pictures". In in NIPS. MIT Press, 2003.

[9] Ying Liua,, Dengsheng Zhanga, Guojun Lua, Wei-Ying Ma, "A survey of content-based image retrieval with high-level semantics".

[10] Tanveer J. Siddiqui , "Bridging the Semantic Gap".

[11] Aanchan K Mohan and Marwan A.Torki, "Automatic Image Annotation using Neural Networks".

[12]Alpesh Dabhi, Bhavesh Prajapati , "A Neural Network Model for Automatic Image Annotation and Annotation Refinement": A survey 2014 IJEDR | Volume 2, Issue 1 43

[13]Suman Tatiraju, Avi Mehta , "Image Segmentation using k-means clustering, EM and Normalized Cuts"

[14]Sarthak panda, "Color Image Segmentation Using K-means Clustering and Thresholding Technique" (march 2015)

[15] Dhatri Pandya1, Prof. Bhumika Shah, "Comparative Study on Automatic Image Annotation" (ISSN 2250-2459, ISO 9001:2008 Certified Journal, Volume 4, Issue 3, March 2014)

[16] 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.

[17] Reena Pagare and Anita Shinde , "A study on Image Annotation Techniques", International Journal of Computer Applications Volume 37-No6. January 2012.

[18] Lei Wu Member,IEEE Rong Jin, Anil K. Jain, Fellow, IEEE , "Tag Completion for Image Retrieval".

[19] Dongping Tian , "Support vector machine for Automatic Image Annotation" International Journal of Hybrid Information Technology Vol.8 No.11(2015).

[20]Dataset : http://www.ci.gxnu.edu.cn/cbir/Dataset.aspx

[21]Attributes: http://sci2s.ugr.es/keel/dataset/data/multilabel/corel5k-names.txt

