International Journal of Scientific Research in Computer Science and Engineering
Vol.5, Issue.4, pp.16-21, August (2017)

Research Paper

E-ISSN: 2320-7639

New Non-Parametric Model for Automatic Annotations of Images in Annotation Based Image Retrieval

Kanakam Siva Ram Prasad[1*]

[1*]Department of IT, Sasi Institute of Technology and Engineering, JNTU Kakinada, Tadepalligudem, India

[*]Corresponding Author: ksrprasad@sasi.ac.in, Tel.: +91 7893928147

Available online at: www.isroset.org

Received 20th Jun 2017, Revised 04th Jul 2017, Accepted 12th Aug 2017, Online 30th Aug 2017

Abstract—In this paper we propose an automatic approach to annotating and retrieving images based on a training set of images. We assume that regions in an image can be described using a small vocabulary of blobs. Blobs are generated from image features using clustering. Given a training set of images with annotations, the annotation process implemented in our system is based on CMRM. Using a set of annotated images the system learns the joint distribution of the blobs and concepts in this paper show those probabilistic models which allow predicting the probability of generating a word given the blobs in an image. This may be used to automatically annotate and retrieve images given a word as a query. We show that relevance models. Allow us to derive these probabilities in a natural way. Experiments show that the annotation performance of this cross-media relevance model is almost six times as than a model based on word-blob co-occurrence model and twice as good as a state of the art model derived from machine translation.

Keywords—*Automatic Image Annotation, Content Based Image Retrieval, Semantic Gap, Annotation Based Image Retrieval, relevance model*

I. INTRODUCTION

Annotation based image retrieval systems are an attempt to incorporate the more efficient semantic content into both text based quires and image captions. ABIR has to be supported due to two causes. First, CBIR has more critical problems of content understanding. Second, the above problems in ABIR may be mitigated due to the negative effects. Hence, in the near future it is necessary for an automatic image annotation (AIA) system to be integrated with current ABIR systems. The tagging is done automatically using content analysis and the retrieval is done using ABIR. The automatic annotation method used in earlier ABIR system is *Translation Model* a substantial improvement on the Co-occurrence Model assumes that image annotation can be viewed as the task of translating from a vocabulary of blobs to a vocabulary of words. Given a set of annotated training images, they show how one can use one of the classical machine translation models to annotate a test set of images. Isolated pixels or even regions in an image are often hard to interpret. It is the context in which an image region is placed that gives it meaning.

AIA is situated on the frontier of different fields*: image analysis, machine learning, media understanding and information retrieval*. Usually image analysis is based on feature vectors and the training of annotation concepts is based on machine learning techniques. Automatic annotation of new images is possible only after the learning phase is completed. General object recognition and scene understanding techniques are used to extract the semantics from data. This is an extremely hard task because AIA systems have to detect at least a few hundred objects at the same time from a large image database.

AIA is a challenge that has been identified as one of the hot-topics in the new age of image retrieval. Image annotation is a difficult task for two main reasons: *Semantic gap problem*– it is hard to extract semantically meaningful entities using just low level image features. Low-level features can be easily extracted from images but they are not completely descriptive for image content. High-level semantic information is meaningful and effective for image retrieval. *Lack of correspondence* between the keywords and image regions in the training data. The semantic gap is due to at least two main problems: First, *Semantic extraction problem* - how to extract the semantic regions from image data? Current object recognition techniques do not cover completely this problem. And second is *Semantic interpretation problem* – is represented by complexity, ambiguity and subjectivity in user interpretation. Representing the content of the image using image features and then performing non textual queries like color and texture is not an easy task for users. They prefer instead textual queries and this request can be satisfied using automatic annotation.

There are many annotation models proposed and split in two categories: (1) *Parametric models*: Co-occurrence Model, Translation Model, Correlation Latent Dirichlet Allocation. (2) *Non-parametric models*: Cross Media Relevance Model (CMRM), Continuous Cross-Media Relevance Model (CRM), Multiple Bernoulli Relevance Model (MBRM), Coherent Language Model (CLM).One approach to automatically annotating images is to look at the probability of associating words with image regions. Used a *Co-occurrence Model* in which they looked at the co-occurrence of words with image regions created using a regular grid. Problems using machine learning approaches are examined and proposed to describe images using a vocabulary of blobs. Each image is generated by using a certain number of these blobs. *Query expansion* is a standard technique for reducing ambiguity in information retrieval. One approach to doing this is to perform an initial query and then expand queries using terms from the top relevant documents. This increases the performance substantially. In the image context, tigers are more often associated with grass, water, trees or sky and less often with objects like cars or computers. *Relevance-based language models* were introduced to allow query expansion to be performed in a more formal manner. These models have been successfully used for both ad-hoc retrieval and cross-language retrieval. In this model every image may be described using a small vocabulary of blobs. Using training set of annotated images, we learn the joint distribution of blobs and words which we call a *cross-media relevance model (CMRM)* for images. There are two ways this model can be used. In the first case, which corresponds to document based expansion, the blobs corresponding to each test image are used to generate words and associated probabilities from the joint distribution of blobs and words. Each test image can, therefore, be annotated with a vector of probabilities for all the words in the vocabulary. This is called the *probabilistic annotation-based cross media relevance model (PACMRM)*. Given a query word, this model can be used to rank the images using a language modeling approach. While this model is useful for ranked retrieval, it is less useful for people to look at.

Fixed length annotations can be generated by using the words (without their probabilities) to annotate the images. This model is called the *fixed annotation-based cross-media relevance model (FACMRM)*. FACMRM is not useful for ranked retrieval (since there are no probabilities associated with the annotations) but is easy for people to use when the number of annotations is small. In the second case, which corresponds to query expansion, the query word(s) is used to generate a set of blob probabilities from the joint distribution of blobs and words. This vector of blob probabilities is compared with the vector of blobs for each test image using Kullback-Liebler (KL) divergence and the resulting KL distance is used to rank the images. This model is called the *direct-retrieval cross-media relevance model (DRCMRM)*. *Cross-media relevance models* are not translation models in the sense of translating words to blobs. Instead, these models take advantage of the joint distribution of words and blobs.

In our model, we assign words to entire images and not to specific blobs because the blob vocabulary can give rise to many errors. Our annotation-based model performs much better than either the Co-occurrence Model or the Translation Model on the same dataset. FACMRM has a much higher recall than the Translation Model. Both models perform substantially better than the Co-occurrence Model. PACMRM and DRCMRM cannot be directly compared to the other systems since the Translation Model and co occurrence model have not been used for ranked retrieval.

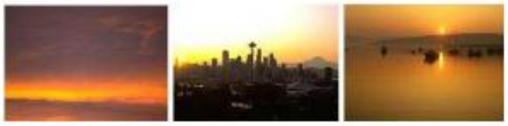

Figure 1: Images automatically annotated as “sunset” (FACMRM) but not manually annotated as “sunset”. The color of sunset may not show up clearly in black and white versions of this figure

Figure 1 illustrates the power of the relevance model. The figure shows three images (from the test set) which were annotated with “sunset” by FACMRM. Although the three are clearly pictures of sunset (the last picture shows both a sun and a sunset), the word “sunset” was missing from the manual annotations. In these cases, the model allows us to catch errors in manual annotation.

II. RELATED WORK

Content Based Image Retrieval CBIR systems search images using low level features such as color, texture, shape, spatial layout etc. which can be automatically extracted and used to index images. Humans tend to associate images with keywords rather than query image. The initial requirement of CBIR systems is to provide query similar image to the retrieval system. The CBIR systems fail to meet user expectations because those systems are unable to index images according to the high level features (keywords, text descriptors etc) as perceived by the user. The main challenge in the CBIR is the two gaps namely **semantic gap** and **sensory gap.**

The basis of Content-based Image Retrieval is to extract and index some *visual features* of the images. There are general features (e.g., color, texture, shape, etc.) and domain-specific features (e.g., objects contained in the image). Domain-specific feature extraction can vary with the application domain and is based on pattern recognition. One drawback of current CBIR systems is that they are based on basic image features that capture low-level characteristics such as color, textures or shape. This approach fails to capture the high-level patterns corresponding to the semantic content of the image; this may produce poor results depending on the type of images the system deals with.

CBIR technologies have shown a lot of limitations regarding lack of the support of high level semantic knowledge and the fact of being far away from the human query perception. Although the user seeks the semantic similarity, the database can only provide the mathematical similarity by means of data processing. An emerging new and possibly more challenging field is arising which is automatic concept recognition from the visual features of image. There is what is called the semantic gap. Shortly, it can be defined as the gap between the human vision and the results of the CBIR systems. Many solutions were proposed to reduce the semantic gap such as: (1) Incorporating the query concepts with the low level features by using the machine query learning tools. (2) Using objects ontology to define high level concepts.

(3) Generating semantic templates to support high level Information Retrieval. (4) Introducing Relevance Feedback (RF) into retrieval process for continuous learning of user intention. (5) Making use of visual contents and textual information.

2.1 Annotation-Based Image Retrieval

Image annotation, the task of associating text to the semantic content of images, is a good way to reduce the semantic gap and can be used as an intermediate step to image retrieval. It enables users to retrieve images by text queries and often provides semantically better results than content-based image retrieval. In recent years, it is observed that image annotation has attracted more and more research interests. When images are retrieved using these annotations, such retrieval is known as annotation-based image retrieval (ABIR).

The ABIR technique primarily relies on the textual information associated with an image to complete the search and retrieval process. Using the game of cricket as the domain, we describe a benchmarking study that evaluates the effectiveness of three popular search engines in executing image-based searches. Second, we present details of an empirical study aimed at quantifying the impact of inter-human variability of the annotations on the effectiveness of search engines. Both these efforts are aimed at better understanding the challenges with image search and retrieval methods that purely rely on ad hoc annotations provided by the humans.

In some scenarios most of the times desired pictorial information can be efficiently described by means of keywords. The process of assigning a set of keywords (or text) to an image is called as annotation. Image Annotation systems attempt to reduce the semantic gap. The task of automatically assigning semantic labels to images is known as Automatic Image Annotation (AIA). Automatic image annotation is also known as auto-annotation or linguistic indexing.

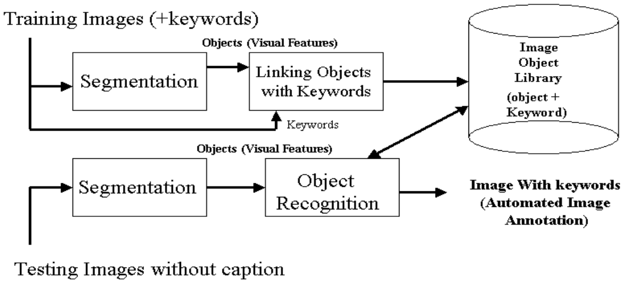
Training Images (+keywords)
Objects (Visual Features)
Segmentation
Linking Objects with Keywords
Keywords
Image Object Library (object + Keyword)
Objects (Visual Features)
Segmentation
Object Recognition
Image With keywords (Automated Image Annotation)
Testing Images without caption

Fig3: Annotation process

Major steps in this frame work are

- Segmentation into regions
- Clustering to construct blob-tokens
- Analyze correspondence between key words and blob-tokens
- Auto Annotation

2.2 Cross-media relevance models

Cross-media relevance models (CMRM): Assume that images may be described from small vocabulary of blobs. From a training set of annotated images, learn the joint distribution of blobs and words. And Allow query expansión Standard technique forreducing ambiguity in information retrieval. Perform initial query and expand by using terms from the top relevant documents. Example in image context: tigers more oftenassociated with grass, water, tres hanwith carsor computers.

2.3 Documentbasedexpansion

PACMRM (probabilistic annotation CMRM): Blobs corresponding to each test image are used to generate words and associated probabilities. Each test generates a vector of probabilities for every word in vocabulary. FACMRM (fixe dannotation-based CMRM) Use top N words from PACMRM to annotate images.

2.4 Querybasedexpansion

DRCMRM (direct-retrieval CMRM): Query words used to generate a set of blob probabilities. Vector of blob probabilities compared with vector from test imageusing Kullback-Lieber divergence and resulting KL distance**.** Segmentation of images into regions yields fragile and erroneous results.

Normalized-cuts are used instead:

- 33 features extracted from images.
- K (=500) clustering algorithm used to cluster regions based on features. Vocabulary of 500 blobs.

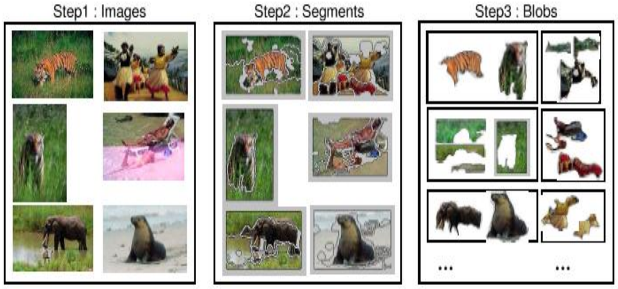

Fig4: images to blobs

III. METHODOLOGY

The annotation process implemented in our system is based on CMRM. Using a set of annotated images the system learns the joint distribution of the blobs and concepts. The blobs are clusters of image regions obtained using the K-means algorithm. Having the set of blobs each image from the test set is represented using a discrete sequence of blobs identifiers. The distribution is used to generate a set of concepts for a new image. Each new image is segmented using a original segmentation algorithm, which integrates pixels into a grid-graph. The usage of the hexagonal structure improves the time complexity of the used methods and the quality of the segmentation results. The meaningful keywords assigned by the annotation system to each new image are retrieved from an ontology created in an original manner starting from the information provided by The concepts and the relationships between them in the ontology are inferred from the concepts list, from the ontology's paths and from the existing relationships between regions.

3.1 Segmentation

For image segmentation, Used a original and efficient segmentation algorithm based on color and geometric features of an image. The efficiency of this algorithm concerns two main aspects: (a) Minimizing the running time – a hexagonal structure based on the image pixels is constructed and used in color and syntactic based segmentation. (b) Using a method for segmentation of color images based on spanning trees and both color and syntactic features of regions.

A similar approach is used in where image segmentation is produced by creating a forest of minimum spanning trees of the connected components of the associated weighted graph of the image. A particularity of this approach is the basic usage of the hexagonal structure instead of color pixels. In this way the hexagonal structure can be represented as a grid-graph G = (V, E) where each hexagon h in the structure has a corresponding vertex v∈V, as presented in Fig.5.

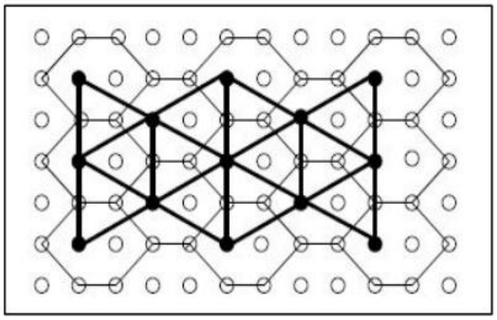

Fig5: grid graph constructed on hexagonal structure of an image

Each hexagon has six neighbours and each neighbourhood connection is represented by an edge in the set E of the graph. To each hexagon two important attributes are associated: the dominant color and the coordinates of the gravity centre. For determining these attributes were used eight pixels: the six pixels of the hexagon frontier, and two interior pixels of the hexagon. Image segmentation is realized in two distinct steps:

(1) *a pre-segmentation step* – only color information is used to determine an initial segmentation. A color based region model is used to obtain a forest of maximum spanning trees based on a modified form of the Kruskal's algorithm. For each region of the input image it is obtained a maximal spanning tree. The evidence for a boundary between two adjacent regions is based on the difference between the internal contrast and the external contrast between the regions

(2) *a syntactic-based segmentation* – color and geometric properties of regions are used. It is used a new graph which has a vertex for each connected component determined by the color-based segmentation algorithm. The region model contains in addition some geometric properties of regions such as the area of the region and the region boundary. A forest of minimum spanning trees is obtained using a modified form of the Boruvka's algorithm. Each minimum spanning tree represents a region determined by the segmentation algorithm.

The annotation process contains several steps as follows ***Obtaining the ontology*** the information provided by the dataset is processed by the Importer module. The concepts associated with images and their hierarchical structure is identified. ***Obtaining the clusters*** we have used K-means algorithm to quantize the feature vectors obtained from the training set and to generate blobs. After the quantization, each image in the training set was represented as a set of blobs identifiers. For each blob it is computed a median feature vector and a list of concepts that were assigned to the test images that have that blob in their representation. ***Image***

segmentation the segmentation algorithm described in above sections used to obtain a list of regions from each new image. ***Automated image annotation*** this task is performed according with the steps involved by the Annotate Image method presented above. The entire annotation process is summarized in following figure.

3.2 Annotation model

Given a training set of images with annotations this model allows predicting the probability of generating a word given the blobs in an image. A test image I is annotated by estimating the joint probability of a keyword w and a set of blobs

$$P(w ,b1 ,..., bm) = \sum_{J\in T} P(J)\, P(w ,b1 ,..., bm \,|J)$$

For the annotation process the following assumptions are made: this is given a collection C of un-annotated images and each image I from C to can be represented by a discrete set of blobs $I = \{b_1 ,...,b_m\}$. There exists a training collection T, of annotated images, where each image J from T has a dual representation in terms of both words and blobs $J = \{w_1 ...w_m ;b_1...b_m \}$ where P(J) is kept uniform over all images in T. The number of blobs m and words in each image (m and n) may be different from image to image and no underlying one to one correspondence is assumed between the set of blobs and the set of words; it is assumed that the set of blobs is related to the set of words. $\boldsymbol{P(w,b_1...,b_m|J)}$ represents the joint probability of keyword w and the set of blobs $\boldsymbol{\{b_1 ...,b_m \}}$ conditioned on training image J. In CMRM it is assumed that, given image J, the events of observing a particular keyword w and any of the blobs $\boldsymbol{\{b_1 ,...,b_m\}}$ are mutually independent, so that the joint probability can be factorized into individual conditional probabilities. This means that$\boldsymbol{P(W ,b_1 ,...,b_m |J)}$can be written as:

$$P(W ,b1 ,..., bm|J) = \prod_{i=1}^{m} P(bi|J) ....(2)$$

$$P(W|J) = (1-\alpha j)\frac{\#(w,J)}{|J|} + \alpha j\frac{\#(w,T)}{|T|} .... (3)$$

$$P(b|J) = (1-\beta j)\frac{\#(b,J)}{|J|} + \beta j\frac{\#(b,T)}{|T|} ...... (4)$$

Where

(a) P(b|J) , P(w|J) denote the probabilities of selecting the word w, the blob b from the model of the image J. (b) #(w, J) denotes the actual number of times the word w occurs in the caption of image J. (c) #(w, T) is the total number of times w occurs in all captions in the training set T. (d) #(b, J) reflects the actual number of times some region of the image J is labelled with blob b. (e) #(b, T) is the cumulative number of occurrences of blob b in the training set. (f) |J| stands for the count of all words and blobs occurring in image J. (g) |T| denotes the total size of the training set. (h) The prior probabilities P(J) can be kept uniform over all images in T. The smoothing parameters α and β determine the degree of interpolation between the maximum likelihood estimates and the background probabilities for the words and the blobs respectively.

IV. Results and Discussion

Two standard measures that are used for analyzing the performance from the annotation perspective are ***Accuracy:*** The accuracy of the auto-annotated test images is measured as the percentage of correctly annotated concepts and for a given test image J⊂T' is defined as

$$accuracy = \frac{r}{|J|}$$

Where variable r represents the number of correctly predicted concepts in J. The disadvantage of this measure is represented by the fact that it does not take into account for the number of wrong predicted concepts with respect to the vocabulary size |W|. ***Normalized score (NS).*** It is extended directly from accuracy and penalizes the wrong predictions. This measure is defined as

$$NS = \frac{r}{|W_J|} - \frac{r}{|W| - |W_J|}$$

Where variable r' denotes the number of wrong predicted concepts in J.

V. Conclusion And Future Scope

The paper describes the extension of an image annotation model that can be used for annotating natural images. The CMRM annotation model has proved to be very efficient by several studies. This model learns the joint probability of concepts and blobs. Two important factors for the annotation process we have used a segmentation algorithm based on a hexagonal structure which was proved to satisfy both requirements: a better quality and a smaller running time. Each new image was annotated with concepts taken from an ontology created starting from the information provided by the benchmark: the hierarchical organization of the vocabulary and the spatial relationships between regions. The experimental results have proved that our proposed modified model produces better results that the initial model.

In the future it is intended to evaluate the modified version from the semantic base image retrieval point of view, using the two methods provided by CMRM: Annotation-based Retrieval Mode and Direct Retrieval Model

References

[1] *Models J. Jeon, V. Lavrenko and R. Manmatha, Automatic Image Annotation and Retrieval using CrossMediaRelevance*, Center for Intelligent Information Retrieval Computer Science department University of Massachusetts Amherst, MA 01003.

[2] RitendraDattaJia Li James Z. Wang, *Content-Based Image Retrieval - Approaches and Trends of the New Age*, The Pennsylvania State University, University Park, PA 16802, USA datta@cse.psu.edu, jiali@stat.psu.edu, jwang@ist.psu.edu

[3] .K. Barnard and D. Forsyth , *Learning the semantics of words and pictures, In International Conference on Computer Vision*, Vol.2, pages 408-415, 2001.

[4]. D. Blei, Michael, and M. I. Jordan, *Modeling annotated data. To appear in the Proceedings of the 26th annual international ACM SIGIR conference.*

[5] K. Barnard, P. Duygulu, N. de Freitas, D. Forsyth, D. Blei, and M. I. Jordan., Matching *words and pictures.Journal of Machine Learning Research*, 3:1107{1135, 2003}.

[6] HimaliChaudhari et al, *A Survey on Automatic Annotation and Annotation Based Image Retrieval.* (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 5 (2) , 2014, 1368-1371.

[7] R.Ramya. M.E, *An Efficient Query Mining Framework Using Spatial Hidden Markov Models for Automatic Annotation of Images.* International Journal of Computer Trends and Technology (IJCTT) – volume 11 number 1 – May 2014 ISSN: 2231-2803

[8] A. Agarwal, S.S. Bhadouria "*An Evaluation of Dominant Color descriptor and Wavelet Transform on YCbCr Color Space for CBIR"* Dept. of Computer Science and Engineeing, NITM, Gwalior, India.International Journal of Scientific Research in Computer Science and Engineering, Vol.5, Issue.2, pp.56-62, April (2017)

Author Profile

Mr. K Siva ram prasad pursed *Bachelor of Computer Application* from Acharya Nagarjuna University of Andhra Pradesh 2004 and *Master of Computer Applications* from Acharya nagarjuna University in year 2007. And *Master of Technology* (computer science and engineering) From JNTU Kakinada in the year 2013. and currently working as Assistant Professor in Department of Information Technology, since 2008. He is a member of IEANG He has published 3 international journals.His main research work focuses on Network Security, Big Data Analytics, Data Mining and Computational Intelligence based education.