Imperial Journal of Interdisciplinary Research (IJIR)
Vol-3, Issue-2, 2017
ISSN: 2454-1362, http://www.onlinejournal.in

A Study on Automatic Image Annotation Techniques

Mrs. Riya Zakkaria P
M. Tech in Computer Science & Engineering, Thejus Engineering College, Vellarakkad

Abstract: *As applications producing visual data have increased tremendously, image retrieval became a necessity. Now visual data are common as textual data, so process of image retrieval involves searching and retrieving images from large data sets. With the advent of digital imaging and data storage, searching and indexing large image databases efficiently and effectively has become a challenging problem. Automatic annotation of digital pictures is a key technology for managing and retrieving images from large image collection. Automatic image annotation is used to predict the relevant keywords from an annotation vocabulary for a new picture. The principle objective of automatic image annotation is to help image retrieval by providing semantic keywords for search. This ability makes expansive image database administration simple. This paper presents a review of such automatic image annotation techniques which automatically assigns metadata in the form of captioning or keywords to a digital image. The different automatic image annotation algorithms in the literature are reviewed by considering their respective advantages and disadvantages and are compared based on several parameters.*

Keywords—*Image Annotation, Computer Vision, Neural Network, Latent Semantic Analysis, Genetic Algorithm*

1. Introduction

Computer vision is an interdisciplinary field that arrangements with how PCs can be made to increase abnormal state understanding from computerized pictures or recordings. People utilize their eyes and their brains to see and outwardly sense their general surroundings. .Computer vision is the science that expects to give a similar, if not better, capability to a machine or computer and it is concerned with the automatic extraction, analysis and understanding of valuable data from a single picture or a sequence of pictures. It involves the development of a theoretical and algorithmic basis to achieve automatic visual understanding. So Image processing is an important area of Computer vision. Fig 1 shows the example of computer vision.

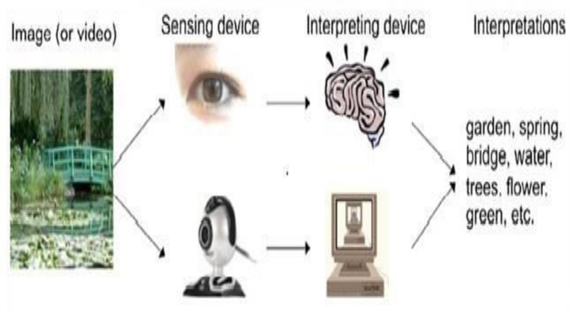

Figure 1. Example of Computer Vision

Retrieve images from large collections of digital system has become a crucial research topic. An image retrieval system is a computer system for browsing, searching and retrieving images from a large database of images. In general, IR research efforts can be divided into three types of approaches. The first approach is the traditional text based annotation (TBIR). In this approach, pictures are annotated manually by people and pictures are then retrieved in the same way as conventional text search engine. However, it is unfeasible to annotate a large amount of pictures manually. Furthermore, human annotations are normally excessively subjective and ambiguous.

The second approach focuses on content based image retrieval (CBIR), where pictures are automatically indexed and retrieved with low level content features like color, shape and texture. In CBIR different low-level visual features are extracted from every picture in the database and image retrieval is planned as looking for the best database match to the feature vector extracted from the query image. Although this process is accomplished rapidly and automatically, the outcomes are semantically significant to the query example due to the notorious semantic gap. Recent research has demonstrated that there is a significant gap between the low level visual features and semantic thoughts utilized by humans to interpret pictures. Furthermore, it is impractical for general clients to utilize a CBIR framework since clients are required to give query images.

The third approach of image retrieval is the tag based image retrieval which focuses on automatic image annotation (AIA) so that images can be retrieved in the same way as text documents. The principle thought of AIA procedures is to automatically learn semantic concept models from huge number of image samples, and utilize the concept models to label new images. When pictures are annotated with semantic labels, images can be retrieved by keywords, which is like text document retrieval. The key characteristic of AIA is that it offers keyword searching based on image content and it employs the advantages of both the text based annotation and CBIR.

Automatic Image Annotation is a standout amongst the most difficult issues in Computer Vision today. The manual annotation of images is not only expensive, but also time consuming and sometimes inaccurate as well. Therefore a computerized system that can accurately suggest annotation keywords for images after analyzing its content can actually prove to be very useful.

2. Automatic Image Annotation

In this section, we will introduce and analyze the concept of automatic image annotation. Automatic image annotation (AIA) has been studied extensively for a several years. Automatic image annotation is the technology in which a computer system automatically assigns metadata in the form of text description or keywords to a digital image[4]. This application is utilized as a part of image retrieval systems to organize and find images of interest from a database. While considering issue of under-standing picture content, it was soon found that, on a fundamental level, associating those pictures with textual descriptions was only one step ahead. This led to the formulation of a new, but closely associated problem called automatic image annotation, often referred to as auto-annotation or linguistic indexing.

Machine learning strategies have supported image retrieval by automatically classifying and annotating images with keywords. Image annotation has attracted a lot of attention due to its importance in understanding images and search areas.

2.1. Image Annotation In A Progressive Way

In this model, annotation is performed by introducing a "progressive" image annotation method [3]. Instead of annotating an image where the words are selected according to their individual probability, we try to maximize the joint probability of the annotation words. So it proposes a greedy solution to approximate the joint probability of multiple words. We first select one word w1 with largest annotation probability as the image's first annotation word. The second annotation word needs to maximize the joint probability of P{w1,w2, I}. Intuitively, if one training image has w1 as its annotation, this image will have higher probability to propagate its rest annotation words to the test image. This annotation procedure can be repeated until some predefined criterion is reached, such as fixed annotation length or a probability threshold. Each time, only one word is annotated if it has the highest joint probability with all previous annotated words. In this manner, the computation complexity increases linearly with the annotation length, avoiding the combination explosion problem. Besides, as we consider one word each time, the problem of insufficient training instances is also alleviated. The experiments show the proposed progressive annotation method can increase the number of correctly annotated words and both the average precision and recall.

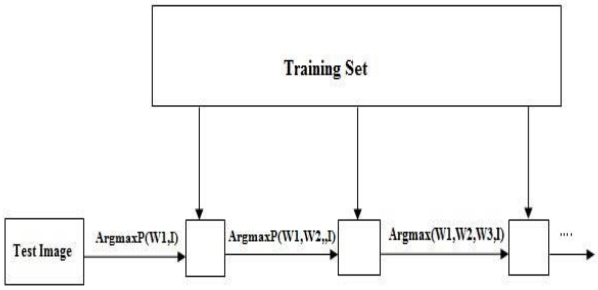

Figure 2. Progressive image annotation

If we get a set of annotation words {w1, w2...wn-1} the word wn can be selected. Therefore, we can annotate the images in a progressive manner, which requires the annotation to be performed in multiple runs. Fig2 illustrates the process of proposed method. In the first run, only the most probable word w1 is chosen, i.e. w1 = argmaxP (w1, I). In the second run, the joint probability P(w2, w1, I) instead of P(w2|I), is considered in selecting word w2. It means that w2 satisfies w2= argmaxP (wl, w2, I). Similarly, we select W3 to maximize P(W1, W2,W3,I) with fixed (Wi, Wj). The process can be repeated until the desired annotation length is reached. This progressive method can be deemed as a "greedy" solution for P(w1,W2,---,Wn, I) to avoid the combinational explosion.

2.2. A Neural Network Model for Automatic Image Annotation and Annotation Refinement

In classical probability models, there are two disadvantages. First, the prior segmentation of the images is required to construct these relevance models. That is, these relevance models are based on region-level. Unfortunately, since the development of the computer vision is limited, the image segmentation is un-ideal so that the annotation performance is unsatisfactory. Second, the latent contextual correlation among the keywords is neglected in these relevance models. To improve the mentioned problems, latent semantic analysis (LSA) based neural network (NN) annotation scheme [2] (noted as LSA-NN) is proposed in this model.

This annotation scheme is composed of three parts. First, LSA is introduced to reveal the latent contextual correlation among the keywords. Second, NN is obtained for characterizing the hidden association between the visual content of the image and the textual keyword with the labeled training images. Third, given a test image, the learnt NN is able to effectively provide the keywords to be annotated. The experimental results show that the high annotation accuracy can be achieved at image-level.

LSA is such a popular algorithm, which is extensively used in text analysis. LSA is also applied to mining the latent semantic relations in the textual keyword space. The detailed algorithm is shown as follows:

2.2.1. LSA Algorithm

Step 1: Stack N_t keyword vectors to form a keyword matrix M_w ε $R^{Nk \times Nt}$ where the row corresponds to the keywords and the column corresponds to the images.

Step 2: Discompose the keyword matrix Mw via singular value decomposition (SVD).
$Mw = MuMsM^{T}{}_{v}$ where $M_{U\,\varepsilon} R^{Nk \times Nu}$, M_S ε $R^{Nu \times Nu}$, M_V ε R^{NuXNt}

Step 3: Transform the keyword vector $Wt = (W_1,W_2,\ldots,W_{Nk)}{}^{T}$ into the contextual keyword vector $V_t = (V_1,V_2,\ldots,V_{Nu})^T$ by: $V_t = W_t M_u$.

Here $T=\{I_1,I_2,\ldots,I_{Nt}\}$ be the set of the labeled images, where Nt is the total number of the labeled images, and $K=\{k_1,k_2,\ldots,k_{Nk}\}$ be the set of labeled keywords where Nk denotes the total number of the keywords. Especially, for the t-th labeled image It.

The latent contextual correlation among keywords is exploited via the transformed matrix MU called as the latent space. Compared the keyword vector Wt with the contextual keyword vector Vt, we can find the fact of NU << NK. It means that the dimension of the vector is also reduced greatly. Thus, the computational cost of NN will be decreased largely.

2.3. Efficient Automatic Image Annotation Using Optimized Weighted Complementary Feature Fusion Using GA

In conventional methods, same weights are taken for each feature and this will raise the feature dimensionality and certain image features will override the feature that the user is truly intrigued on a specific picture. So, to conquer these issues, in this paper, weights of different features should be assigned appropriately using Genetic algorithm (GA) similar to human perception which gives an optimized feature vector of each image. Genetic algorithm is commonly used for finding fittest solutions exploring through large solution domain that may not discover in a lifetime. GAs keep running on set of solution points at a time rather than performing one at a time, so the issue of being stacked on local minima is reduced.

This model aims at automatic image annotation using the optimized weights obtained through the Genetic Algorithm [1]. By coding of variables, GAs can discretize the solution space even though the function is continuous. As weights of different descriptor features are assigned appropriately, certain image features in the query image where traditional algorithms had failed due to multiple local optima's and where objective functions is not smooth becomes more accurate. There is a lot of scope for enhancing the proposed system including making the genetic algorithm much more interactive to the user by giving the user freedom for selecting the suitable images.

This method works in three phases: Training phase, Fusion Phase and Annotation phase. Training phase proceeds in two steps. First the Histogram of Oriented Gradients (HOG), Speed-Up Robust Features (SURF), and Hue Saturation Value (HSV) color features are acquired from training set of images. Secondly, average feature descriptor representing each class is obtained. In fusing phase , weights of various features are identified by using Genetic algorithm for each class of images. In the annotation phase ,first the features of each image is extracted and synthesized as in training phase, then these images are classified based on the similarity measure and finally these images have been annotated according to the model of each class obtained in the training phase and the weights of features obtained during the fusing phase.

3. Comparative study

This section compares each automatic image annotation techniques discussed in the literature review based on several parameters. Table 1 show the evaluation summary of different automatic image annotation methods. It shows the data that can be handles by the automatic image annotation techniques. Table 2 summarizes the advantages and disadvantages of different type of annotation methods. The variations of common parameters give the best technique with better performance.

Table 1. Evaluation summary of different automatic image annotation methods

Method	Using Algorithm	Data base	Performance
Image Annotation In a Progressive Way	Greedy	Corel dataset	Low
A Novel Image Annotation Scheme Based On Neural Network	LSA-NN	Corel dataset	High
Efficient Automatic Image Annotation using Genetic Algorithm	Genetic Algorithm	Corel dataset	High

Table 2. Contrast of different annotation methods

Method	Advantages	Disadvantages
Image Annotation In a Progressive Way	Fast, Model free, Multiple labelling	Annotation results are highly affected by segmentation, Large number of parameters , chance of ambiguities
A Novel Image Annotation Scheme Based On Neural Network	Multiclass outputs, Less ambiguities, robust to noisy data, suitable for complex problem	Expensive training and complex task
Efficient Automatic Image Annotation using Genetic Algorithm	As the number of iterations Increases in genetic algorithm, weights obtained becomes much more accurate and yields better result	Expensive and complex aspects, time taken for convergence, Incomprehensi ble solutions

4. Conclusion

The major research field in image processing is to develop efficient image retrieval methods to retrieve the relevant images from the large image dataset. Automatic annotation of digital pictures is a key innovation for managing and retrieving images from huge image collection. In this paper review on various approaches of automatic image annotation and its comparison is discussed. With rapid developments of digital technology, retrieval of visual images is necessary for various purposes. From the comparative study it is clear that different algorithm can be chosen to perform image annotation. In practice, each algorithm can be useful based on its applications and properties. If we can combine different methods together the accuracy and efficiency of the result may get improved. Thus to enhance the image retrieval results various approaches of automatic image annotation can be applied depending on the image dataset and irrelevant images will be reduced in the image retrieval. The performance of the annotation model can be verified by annotation based image retrieval and applying ranking on the image and filtering noisy keywords assigned to the image and calculating the precision and recall measure for the image. In future work, we aim to develop and implement an efficient automatic image annotation method to overcome.

5. References

[1] Ajimi Ameer.,SreeKumar,K..,"Efficient Automatic Image Annotation using Optimized weighted Complementary Feature Fusion using Genetic Algorithm", *Second International Symposium on Computer Vision and the Internet*,2015

[2] Yufeng Zhao.,Yao Zhao., Zhenfeng Zhu., Jeng Shyang Pan., "A Novel Image Annotation Scheme Based on Neural Network", *Intelligent Systems Design and Applications, ISDA '08. Eighth International Conference on*, vol.3, pp.644-647, 2008.

[3] Bin Wang., Zhiwei Li., Nenghai Yu., Mingjing Li., "Image Annotation in a ProgressiveWay", *IEEE International Conference on Multimedia and Expo*,vol., no., pp.811-814,2007.

[4] Zhang,D., Md.Islam,M., Lu,G., "A review on automatic image annotation techniques", *ELSEVIER Pattern Recognition*,no.45,PP.346-362,2012.

[5] Duygulu,P., Barnard,K..,de Freitas,N.,Forsyth,D., "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, 2002.

[6] Jeon, J., Lavrenko,V., Manmatha,R., "Automatic Image Annotation and Retrieval using Cross-Media Relevance

Models", *26th Annual Int. ACM SIGIR Conference*, Toronto, Canada,2003.

[7] Chapelle,O., Haffner,P.,Vapnik,V.N.,"Support vector machines for histogram based image classification",*IEEE Transactions on Neural Networks*, Vol. 10, pp.10551064, 1999.

[8] Town, C.P., Sinclair,D., "Content-Based image retrieval using semantic visual categories", *Technical Report,No.MV01-211,Society for Manufacturing Engineers*, 2001.

[9] Frate,F.D.,Pacifici,F.,Schiavon,G.,Solimini,C.,"Use of neural networks for automatic classification from high resolution images", *IEEE Transactionson Geoscience and Remote Sensing*,Vol. 4, pp. 800809, April 2007.

[10] Shaohua Wan.,"Image Annotation Using the Simple Decision Tree*",Management of e-Commerce and e-Government (ICMeCG), Fifth International Conference on* ,pp.141-146,2011.