Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 3 (2011) 414–419

Procedia Computer Science

www.elsevier.com/locate/procedia

WCIT-2010

Adding semantics to the reliable object annotated image databases

Irfanullah[a*], Nida Aslam[a,b], Jonathan Loo[a], Roohullah[c], Martin Loomes[a]

[a]*School of Engineering & Information Sciences, Middlesex University, The Burroughs London, NW4 4BT,UK*
[b]*Department of Information Technology, Kohat University of Science & Technology, Kohat, 26000, Pakistan*
[c]*Departmentof Computer & Information Sciences,Universiti Teknologi Petronas, Bandar Seri Iskandar, Malysia*

Abstract

Semantically enriched multimedia information is crucial for equipping the kind of multimedia search potentials that professional searchers need. But the semantic interpretation of multimedia is obsolete without some mechanism for understanding semantic content that is not explicitly available. Manual annotation is the only source to overwhelming this, which is not only time consuming and costly but also lacks semantic enrichment in terms of concept diversity and concept enrichability. In this paper, we present semantically enhanced information extraction model that calculate the semantic intensity (SI) of each object in the image and then enhance the tagged concept with the assistance of lexical and conceptual knowledgebases .i.e. WordNet and ConceptNet. Noises, redundant and unusual words are then filtered out by means of various techniques like semantic similarity, stopwords and words unification. The experiment has been carried out on the LabelMe datasets. Results demonstrate the substantial improvement in terms of concept diversity, concept enrichment and retrieval performance.

Selection and/or peer-review under responsibility of the Guest Editor.

*Keywords:*Semantic Intensity; Multimedia Annotation; Semantic Gap; Knowledgebase.

1. Introduction

The betteruse of digital technologies for processing, distribution and production of multimedia within the past decade has raised the need to store valuable information within pictorial form. A major problem in dealing with large corpus of images databasesis the efficiency of retrieval. One of the main issues in succeeding such efficiency is the design of anappropriate indexing scheme. In order to be able to efficiently retrieve required information from large corpus of images two major approaches have been meanwhile established. They mainly differ in the way a query is articulated. Content-based Image Retrieval approaches try to find images semantically similar to a given query image example by equating them on a low-level basis. The requirement of an initial query image, however, incapacitates these approaches in any retrieval scheme, since the accessibility of such an image would most likely already solve the retrieval task. Moreover, comprehensive investigations on CBIR systems show that there is a gap between image visual features and high-level semantic concepts. Therefore, in Annotation-based Image Retrieval an image collection is searched based on a textual description of the depicted content. While this advent is best-suited in situations where the desired pictorial information can be effectively illustrated by means of keywords, it demands for interpretation of the depicted contents into a textual representation (annotation), which is either done manually or by automatic means. Each of them has their own pros and cons. In many situations, we want to find the images related to a specific concept, i.e. *"Park"* or we want to find the keywords that best describe the contents of an unseen image [15]. Sometime the annotator (manual or automatic) goes wrong to express the semantics accurately, whilst sometimes the user query words quite different to the ones used in the annotation describing the same

* Irfanullah, Tel: (+44) 07916860919
E-mail Address: pir.irfan@hotmail.com

1877-0509
doi:10.1016/j.procs.2010.12.069

semantics. That means, there is a gap exist between users query space and an image representation space. This leads to the lower precisions and recalls of queries. The user may get an overwhelming but large percent of irrelevant images in the result sets. In fact, this is a tough problem in multimedia retrieval systems.

In this paper, we propose novel techniques suitable for enhancing and refining the annotation of the images.Initially, the annotation of the images is analysed to prune the noisy keywords and SI is then calculated for each object, which is the dominancy factor of the object in the image. The objects having SI value below a certain threshold are discarded. The redundant objects are combined to one instance by adding their SI values. Enhancing and refinement for annotation with the help of lexical and conceptual knowledgebases are then applied to achieve concept diversity. The refining and validation processes are executed in two stages,semantic similarity is calculated among the original and generated keywords and discard the keywords with a value less than a certain threshold value. Semantic distance is then calculated among the generated keywords to additional purify the enhance annotations.The experiment has been carried out on the LabelMe datasets. Results demonstrate the substantial improvement in terms of concept diversity, concept enrichment and retrieval performance.

2. Related State of the Art

Almost all of the existing image annotation work can be classified into two categories. Firstly, classification approaches, where each keyword (concept) is considered as a unique class of the classifier, the SVM [2, 4, 6], Gaussian Mixture Hierarchical Model [1, 4], Bayes Point Machine [5] and so on are few examples of them. Secondly, taking advantage of the statistical models for image annotation i.e. Duygulu et al [7] strived to map keywords to individual image objects. Pan et al. [9] have proposed various methods to discover correlations between image features and keywords. Nikhil Garg et al. [8] used a co-occurrence model, and reduce the noisy keywords from the annotated images of flicker and coral datasets. Based on translation model, F. Kang et al [12] propose two modified translation models. Jeon et al [10] introduce cross-media relevance model (CMRM), Lavrenko et al [11] propose continuous relevance model for the image annotation. However, in all of this work annotation contains many noisy keywords and there is no attempt to extend this “limit” of automatic image annotation problem.

Research in text mining area manages to build considerable commonsense knowledgebases. Commonsense is recognized as the information and facts that are expected to be normally known by ordinary people. WordNet, Cyc and ConceptNet are considered to be the widest commonsense knowledgebases currently in use. In multimedia annotation domain, these knowledgebases have recently received more attention for solving annotation issues, by finding related concepts. Altadmir, Ahmed et al. [13] put forward a framework for video annotation enhancing and validation using WordNet and ConceptNet. They enhance the existing annotation by adjoining synonym set with each term and then validate each term using ConceptNet *“capableOf”, “usedFor”* and *“locationAt”*. Yohan, Khan et al. [14] bring up the innovative approach using semantic similarity measure among annotated keywords.

3. Proposed Model:

Let $L = \{t_1, t_2, \ldots, t_n\} = \sum_{i=1}^{n} t_i$ be the list of the label tag per image and $C = \{x_1, x_2, \ldots, x_n\} = \bigcup_{j=1}^{m} x_j$ is the corpus of images dataset representing list of the annotated images, where $x_j$represent individual image. By combining both of the equations, the corpus become

$$C = \bigcup_{j=1}^{m} \left(\sum_{i=1}^{n} t_i\right)_j \quad (1)$$

In the following sections, the proposed model is described in details.

3.1. Annotation Purification

The LabelMe online annotation tool provideseasy to use environment for the user to annotate objects with the user define tag as a result problem like redundancy, irrelevant and unusual keywords are continuously generated during the annotations process. For example, for the street imagesthe words like building, people and trees are common to be redundant. The irrelevant words like *“az0003”, “ghkdf65we”, “oi45nelfds”*need to be discarded straight away, while the unusual words like *“personsitting”, “caroccluded”, “personwalking”* require unification process. Let $L' = \{t_1, t_2, \ldots, t_n\} = \sum_{i=1}^{n} t_i'$ represent the purified list of the labels tag with the image, then equation (1) become

$$C' = \cup_{j=1}^{m}(\sum_{i=1}^{n} t_i')_j \quad (2)$$

3.2. Semantic Intensity Calculation

The online tool for LabelMe datasets to annotate objects with the user defines tag, object edges are represented in the form of polygon in the annotated dataset as shown in figure 1. Area of the irregular polygon is $A_{poly} = \frac{1}{2}\sum_{i=0}^{n-1}(x_i y_{i+1} - x_{i+1} y_i)$.The SI for the given object can be calculated as $SI = \frac{A_{poly}}{I_s}$, where $I_s = h * w$, represents size of the image. Now the equation (2) becomes

$$C' = \cup_{j=1}^{m}(\sum_{i=1}^{n}(t', SI))_j \quad (3)$$

The annotation purification processes are then exercised to remove the labels with SI value less than a specified threshold.

3.3. Annotation Enhancement

The annotations in the purified corpus are then enhanced for indexing and retrieval purposes. ConceptNet and WordNet are selected to be utilized in this work for several reasons, like both of them are used in wide domain for annotation and retrieval, having natural language form with semantic relational structure. The ConceptNet nodes mainly address everyday life and have the ability to connect objects and their events, while WordNet nodes mainly on formal taxonomies and support single words. *"Synsets"* support is available with WordNet while *"Conceptset"* support is built for the ConceptNet based on the different relationship exists.

We consider the Annotation Enhancement (aE) task, where the system extends the existing annotation for each image from $x \in X$ in the purified corpus C' by using lexical and conceptual knowledge bases, which extends the equation (3)

$$C' = \cup_{j=1}^{m}(\sum_{i=1}^{n}(t', aE, SI))_j \quad (4)$$

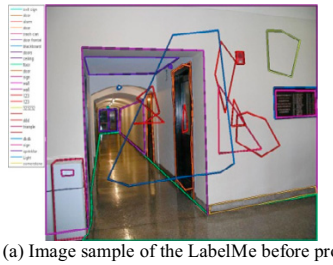

Figure 1. (a) Image sample of the LabelMe before processing

(b) Image with purified annotation, where number of instance for each object are represented in parenthesis.

3.4. Annotation Refinement and Validation

The expanded form of the lexical and conceptual knowledge bases comes up with too many keywords. Some of them are irrelevantthat decrease the precision of the query. For the betterprecision, we have to remove these noisy keywords. For refinement, we applied semantic similarity among the original and each of the generated keywords and discard the keywords that fails to achieve the threshold. Semantic distances among the generated keywords for each label are then calculated to further validate the enhanced annotation. After the annotation refinement and validation, the equation (4) becomes

$$C' = \cup_{j=1}^{m}(\sum_{i=1}^{n}(t', aE', SI))_j \quad (5)$$

4. Experimental Results

Experiments were performed on the LabelMe datasets on some of the category from the LabelMe 31.8 GB datasets that contains total of 181, 932 images with 56946 annotated images, 352475 annotated objects and total of 12126 classes. The results were evaluated using Enrichment Ration,Retrieval Degree, and Concept Diversity.

4.1. Enrichment Ratio

Tagging ratio, which is the average number of labels per image and enhancement ratio, which is the percentage of tagging ratio increase after enhancing and refinement annotation, formulas are explained in equation

$$T = \frac{\sum_{i=1}^{n}(C_i)}{N} \tag{6}$$

Where, T is the tagging ration and C_i is the number of Concepts tag with the image respectively.

$$E = \frac{T_1}{T_2} \tag{7}$$

Where E is the Enrichment Ratio for the T_1 and T_2which is the tagging ration before and after concept enhancement respectively. As tagging ratio has risen from 6.19 tags per image in the dataset to 13.54 tags after annotation enhancement and refinement, whilst enrichment ratio has achieved a considerable degree about 219%. There is although 2.90 unusual tags per images were removed or corrected by unification module. Figure 2 depicts the ratio of initial tags to the resulted of enhancement and refinement tags.

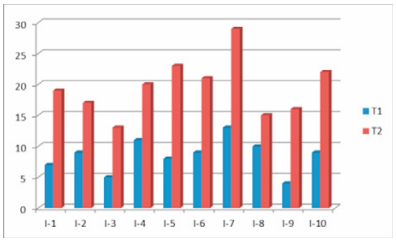

Figure 2. (a) Tagging Ratio of the original and enhance annotation for the 10 sample images. (b) Enrichment Ratio of the LabelMe Dataset

4.2. Retrieval Degree

Retrieval degree is the number of correct images retrieved with a simple object based query. In figure 3, the retrieval degrees of a different object based queries are shown, that depicts object based query on original datasets, enhancement & refine. Using our proposed framework, the retrieval degree has been increased up to noticeable level.

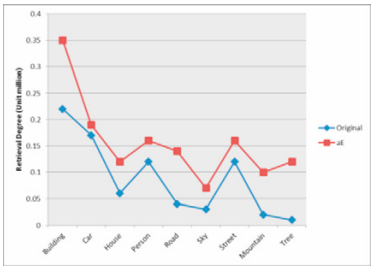

Figure 3. Retrieval Degree of different objects based query.

Figure 4. Concept Diversity

4.3. Concept Diversity

The Concept Diversity of annotations expresses the different topics or objects name exist in the dataset. It has been raised in a noticeable degree also from 12126 different tags to 14324. This diversity achieves 118% increase in the topic indexed. Figure 4 demonstrate this increasing of all differentiated tags.

These results exhibits that searching and retrieval for images over enhanced annotation outperforms searching and retrieval using the original labels. In adding to that, annotation enhanced by the proposed model surpasses both these enhanced by WordNet and ConceptNet combinely, in terms of concept diversity, labels enrichment ability, and most importantly retrieval performance.

5. Conclusion

In this paper, a novel image annotation enhancement and refinement model is presented, that take advantages of the lexical and conceptual knowledgebases. The noises in the annotated data of the LabelMe are removed and then SI for each of the object is calculated that help in further purification of the annotation by discarding the object names with low SI values. The redundancy is control to one instance per image by adding their SI values. Enhancements are applied by taking synset and conceptset from WordNet and ConceptNet. The refining and validation process are then exercise in two stages, firstly semantic similarity are applied between the original and generated keywords and discard the keywords with a value less than a certain threshold value. Semantic distance are calculated among the generated keywords to further purify the enhance annotations. The experiment has been carried out on the LabelMe datasets. Results demonstrate the substantial improvement in terms of concept diversity, concept enrichment and retrieval performance.

References

1. GVS Raj Kumar et al, “Image Segmentation by Using Finite Bivariate Doubly Truncated Gaussian Mixture”, International Journal of Engineering Science and Technology, Vol. 2(4), pp: 694-703, 2010
2. Ch.Srinivasa Rao, S.Srinivas Kumar, B.Chandra Mohan, "Content Based Image Retrieval Using Exact Legendre Moments and Support Vector Machine", The International Journal of Multimedia and its Application, Vol.2, No.2, 2010, pp: 69-79
3. Y. Gao, J. Fan, H. Luo, X. Xue, & R. Jain. “Automatic image annotation by incorporating feature hierarchy and boosting to scale up SVM classifiers”. In Proceedings of the 14th Annual ACM International Conference on Multimedia, Santa Barbara, CA, USA, 2006.
4. Carneiro, G., & Vasconcelos, Formulating semantic image annotation s a supervised learning problem. In Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition 2005.
5. Chang, E. Kingshy, G. Sychay, G. & Wu, G. CBSA: content-based soft annotation for multimodal image retrieval using Bayes point machines. IEEE Trans on CSVT, 13(1), 26–28.2003
6. Stefan Rueping, "SVM Classifier Estimation from Group Probabilities". In Proceedings of the 27th International Conference on Machine Learning, Haifa, Israel, 2010
7. Duygulu, P., Barnard, K., De Freitas, N., & Forsyth, D. “*Object recognition as machine translation: learning a lexicon for a fixed image vocabulary”.* In Proceedings of the 7th European Conference on Computer Vision (ECCV) Part IV, Copenhagen, Denmark, 97–112. 2002.
8. Nikhil Garg et al. "Tagging and retrieving images with co-occurrence models: from corel to flickr", Proceedings of the First ACM workshop on Large-scale multimedia retrieval and mining Beijing, China, 2009.
9. Pan, J. Y., Yang, H. J., Faloutsos, C., & Duygulu, P. Automatic multimedia cross-modal correlation discovery. In Proceedings of the 10th ACM SIGKDD Conference KDD 2004. Seattle, WA, pp: 653–658. 2004
10. Jeon, J., Lavrenko, V., & Manmatha, R. Automatic image annotation and retrieval using cross-media relevance models. Proceedings of the 26th Annual International ACM SIGIR Conference, Toronto, Canada, pp: 119–126. 2003.
11. Lavrenko, V. Feng, S. L., & Manmatha. Statistical models for automatic video annotation and retrieval. International Conference on Acoustics, Speech and Signal Processing, (ICASSP) Montreal, QC, Canada, pp: 17–21. 2004.
12. Kang, F., Jin, R., & Chai, J. Y. Regularizing translation models for better automatic image annotation. In Proceedings of the 13th Conference on Information and Knowledge Management, Washington D. C., USA, 8-13, pp: 350-359. 2004.

13. Altadmri, Amjad and Ahmed, “Video databases annotation enhancing using commonsense knowledge bases for indexing and retrieval”. In proceeding of the 13th International Conference on Artificial Intelligence and Soft Computing.,Palma de Mallorca, Spain. 2009
14. Yohan Jin, Latifur Khan and B. Prabhakaran,"Knowledge Based Image Annotation Refinement", Journal of Signal Processing Systems, Volume 58, Number 3, pp: 387-406, 2009
15. Duygulu, P., Barnard, K., De Freitas, N., & Forsyth, D. Object recognition as machine translation: learning a lexicon for a fixed image vocabulary. In Proceedings of the 7th European Conference on Computer Vision (ECCV) Part IV, Copenhagen, Denmark, 97–112. 2002
16. LabelMe statistics on July 8th, 2010