A Study of Vocabularies for Image Annotation

Allan Hanbury*

Pattern Recognition and Image Processing group (PRIP),
Institute of Computer Aided Automation, Vienna University of Technology,
Favoritenstraße 9/1832, A-1040 Vienna, Austria
hanbury@prip.tuwien.ac.at

Abstract. In order to evaluate image annotation and object categorisation algorithms, ground truth in the form of a set of images correctly annotated with text describing each image is required. Statistics on the WordNet categories of keywords collected from recent automated image annotation and object categorisation publications and evaluation campaigns are presented. These statistics provide a snapshot of keywords used to train and test current image annotation systems as well as information on the usefulness of WordNet for categorising them.

1 Introduction

Automated image annotation and object categorisation are currently important research topics in the field of computer vision [1,2,3,4]. To measure progress towards successfully carrying out this task, evaluation of algorithms which automatically extract this sort of metadata is required. For successful evaluation of these algorithms, reliable ground truth is necessary. This ground truth is usually in the form of a manual or computer-assisted annotation of images by keywords.

For *automated image annotation*, the aim is to automatically assign suitable keywords to describe images or regions of images based on image features [1,2,3]. *Object Categorisation* is concerned with the identification of particular objects [4,5]. Object categorisation can be seen as annotation of an image by keywords describing the objects present. Due to the nature of the algorithms applied, automated image annotation techniques have in general annotated the images using a selection from a larger vocabulary of keywords than object categorisation.

In this paper we provide an overview of the keywords that have been used for the annotation of images for evaluation purposes. We present statistics on the distribution of the WordNet categories of keywords from datasets used in automated image annotation and object categorisation research and evaluation campaigns. This provides a snapshot of keywords used to train and test current image annotation systems. The list of keywords created in this work could be used as the start of a more comprehensive vocabulary for the annotation of images, similarly to the way in which the vocabulary in [6] was begun.

* This work was partially supported by the European Union Network of Excellence MUSCLE (FP6-507752).

B. Falcidieno et al. (Eds.): SAMT 2007, LNCS 4816, pp. 284–287, 2007.

Table 1. The sources of the keywords. The left column gives the source and references, and the right column gives the number of keywords obtained from the source.

Source	# Keywords
PASCAL VOC Challenge 2005 databases [5]	101
EU LAVA Project [7]	10
Chen and Wang [8]	20
Microsoft Research Cambridge Databases [4]	35
Fei-Fei et al. [9]	101
Carbonetto et al. [2]	55
Li and Wang [3]	433
Barnard et al. [1]	323
University of Washington Ground Truth Image Database	392

2 Analysis of Visual Keywords

We created a list combining all the keywords used in the papers and datasets listed in Table 1. These sources correspond to datasets which have been made available on-line. Each keyword was entered into the list only once, even if it occurred in more than one source list. Nouns in plural forms were converted to singular form. Keywords not present in WordNet were excluded. This led to a combined list containing 792 keywords.

We categorised these keywords using WordNet [10] categories from a higher level. Choosing a suitable level for all classes of words proved to be difficult, as the different branches of the WordNet hierarchy have different depths. For example taking the top level of the WordNet hierarchy (the word "entity") as level 0, the word "goalpost" is found at level 11, while the word "waterfall" is at level 4. We therefore decided to use a mixture of categories from levels 3 and 4 of the hierarchy, manually chosen to minimize the number of categories (e.g., each keyword should if possible not form its own category, as would happen if we chose level 4 for the keywords describing bodies of water) while preventing the creation of categories containing too many keywords. The "food" category occurs in both levels 3 and 4 in different branches of the hierarchy, referring to "solid food" and "nutrients". To prevent the creation of two very similar categories, the members of both categories have been fused. For proper nouns, we used the "instance of" relation to place them in the hierarchy. Verbs and adjectives were placed in their own categories. Some words occur in more than one branch of the WordNet hierarchy. In this case, we manually chose the "most visual" of the branches, or included more than one branch if the corresponding senses were applicable to images. The distribution of the number of keywords per category is shown in Table 2. The full categorised keyword list is available for download[1].

The *artefact* category contains the most keywords, followed by the *living thing* category, where the latter includes humans, plants and animals. This reflects the presence of objects and animals in both the Corel dataset and the datasets

[1] `http://muscle.prip.tuwien.ac.at/keywords_with_wordnet_categories.txt`

Table 2. The number (#, column 2) of keywords occurring in each category. The level ℓ at which each category occurs in the WordNet hierarchy is shown in the 3rd column. The rightmost 3 columns contain the names of the levels above the chosen category.

category	#	ℓ	level 1	level 2	level 3
artefact	264	4	physical entity	object	whole
living thing	153	3	physical entity	object	
location	114	3	physical entity	object	
event	40	4	abstract entity	abstraction	psychological feature
adjective	34				
food	22	3	physical entity	substance	
		4	physical entity	substance	solid
natural object	21	4	physical entity	object	whole
geological formation	18	3	physical entity	object	
cognition	14	4	abstract entity	abstraction	psychological feature
body of water	14	3	physical entity	thing	
material	12	3	physical entity	substance	
body part	11	4	physical entity	thing	part
fundamental quantity	10	4	abstract entity	abstraction	measure
natural phenomenon	10	4	physical entity	process	phenomenon
land	10	3	physical entity	object	
attribute	9	3	abstract entity	abstraction	
group	8	3	abstract entity	abstraction	
communication	8	3	abstract entity	abstraction	
gas	5	4	physical entity	substance	fluid
solid	5	3	physical entity	substance	
relation	4	3	abstract entity	abstraction	
drug	2	4	physical entity	causal agent	agent
suspension	1	4	physical entity	substance	mixture
chemical process	1	4	physical entity	process	natural process
verb	1				
bodily process	1	4	physical entity	process	organic process

collected for object categorisation tasks. The *location* keywords also contain a large number of proper nouns, such as "Mexico" and "British Columbia", which are present in the Corel annotations. It is less likely that these proper nouns can be successfully associated automatically with images. The *event* category contains keywords such as "war", "parade" and various types of sport.

The categories containing few keywords show which branches of the WordNet hierarchy are sparsely populated: the only *bodily process* is "dining", the only *chemical process* is "fire" and the only *suspension* is "steam". These sparsely populated branches could also be seen as a demonstration that the WordNet hierarchy is not ideally suited to intuitive categorisation of concepts found in images. An example of this is that "sky" is classified under *gas*. The reason that there are 5 keywords in this category is that the Washington Database contains the combined keywords "clear sky", "cloudy sky", "overcast sky", etc. These common keywords forming part of sparsely populated branches should also be considered when WordNet is used as the basis for a vocabulary, as in [11].

The advantage of using WordNet categories is that it is straightforward to examine categories obtained from other levels of the hierarchy. A solution to the problem of some of the sparsely populated categories would be to use the level 2 category *substance*. This would fuse the categories *food*, *material*, *gas*, *solid* and *suspension* into a single category containing 45 keywords. However, "sky" would now fall into the rather non-intuitive *substance* category.

3 Conclusion

We analyse the keywords that have been used to annotate images in a number of publications and evaluation campaigns. These keywords are placed into categories obtained from higher levels of the WordNet hierarchy. From this analysis one can see that the main automated annotation effort has been directed at images of everyday objects and of living things. This categorisation also reveals some disadvantages of using the WordNet hierarchy to create intuitive categories for an image annotation vocabulary. The investigation of a more intuitive categorisation of keywords for image annotation is an interesting topic to pursue.

References

1. Barnard, K., Duygulu, P., de Freitas, N., Forsyth, D., Blei, D., Jordan, M.I.: Matching words and pictures. Journal of Machine Learning Research 3, 1107–1135 (2003)
2. Carbonetto, P., de Freitas, N., Barnard, K.: A statistical model for general contextual object recognition. In: Pajdla, T., Matas, J(G.) (eds.) ECCV 2004. LNCS, vol. 3021, pp. 350–362. Springer, Heidelberg (2004)
3. Li, J., Wang, J.Z.: Automatic linguistic indexing of pictures by a statistical modeling approach. IEEE Trans. PAMI 25(9), 1075–1088 (2003)
4. Winn, J., Criminisi, A., Minka, T.: Object categorization by learned universal visual dictionary. In: Proc. ICCV, pp. 1800–1807 (2005)
5. Everingham, M., et al.: The 2005 PASCAL visual object classes challenge. In: Selected Proceedings of the First PASCAL Challenges Workshop (2006)
6. Jörgenson, C., Jörgenson, P.: Testing a vocabulary for image indexing and ground truthing. In: Proc. Internet Imaging III, pp. 207–215 (2002)
7. Perronnin, F., Dance, C., Csurka, G., Bressan, M.: Adapted vocabularies for generic visual categorization. In: Leonardis, A., Bischof, H., Pinz, A. (eds.) ECCV 2006. LNCS, vol. 3954, pp. 464–475. Springer, Heidelberg (2006)
8. Chen, Y., Wang, J.Z.: Image categorization by learning and reasoning with regions. Journal of Machine Learning Research 5, 913–939 (2004)
9. Fei-Fei, L., Fergus, R., Perona, P.: Learning generative visual models from few training examples an incremental bayesian approach tested on 101 object categories. In: Proc. Workshop on Generative-Model Based Vision (June 2004)
10. Miller, G.A., Beckwith, R., Fellbaum, C., Gross, D., Miller, K.: Introduction to WordNet: An on-line lexical database. International Journal of Lexicography 3(4), 235–244 (1990)
11. Zinger, S., Millet, C., Mathieu, B., Grefenstette, G., Hède, P., Moëllic, P.A.: Extracting an ontology of portrayable objects from WordNet. In: Proc. MUSCLE/ImageCLEF Workshop on Image & Video Retrieval Evaluation, pp. 17–23 (2005)