Analysis of Keywords used in Image Understanding Tasks

Allan Hanbury

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

Abstract

In the field of computer vision, automated image annotation and object recognition are currently important research topics. It is hoped that these will lead to improved general image understanding which can be usefully applied in Content-based Image Retrieval. In this paper, an analysis of the keywords that have been used in automated image and video annotation research and evaluation campaigns is presented. The outcome of this analysis is a list of 525 keywords divided into 15 categories. Given that this list is collected from existing image annotations, it could be used to check the applicability of ontologies describing entities which are portrayable in images.

1. Introduction

The usual reason to annotate data (i.e. add metadata to it) is to simplify access to it. This is particularly important for the semantic web. The metadata added to documents or images allow for more effective searches. The problem with adding metadata manually is that it is an extremely labour-intensive and time-consuming task. In the field of computer vision, automated image annotation and object recognition are currently important research topics (Barnard et al., 2003; Carbonetto et al., 2004; Csurka et al., 2004; Li and Wang, 2003; Winn et al., 2005). This automatic generation of image metadata should allow image searches and Content-Based Image Retrieval (CBIR) to be more effective. For example, an image database could be annotated offline by running a keyword annotation algorithm. Every image containing a cup would then have the keyword "cup" associated with it. If a user wishes to find images of a specific cup in this database, he/she would select a region containing the target cup from an image. An object recognition algorithm could then categorise the selected region as a cup and a text search could be carried out to find all images in the database with an associated keyword "cup". This would significantly reduce the number of images in which it would be necessary to attempt to recognise the specific cup selected by the user.

To measure progress towards successfully carrying out this task, evaluation of algorithms which can automatically extract this sort of metadata is required. For successful evaluation of these algorithms, reliable ground truth is necessary. This ground truth should be a semantically rich description of the objects in an image (Leung and Ip, 2000). There is obviously almost no limit to how semantically rich one could make the description of an image. Indeed, for manual annotation of such documents destined to aid in online searching for them, semantic richness is an advantage. For images, one can create complex ontologies allowing the specification of objects and actions. For example, Schreiber et al. (2001) create such an ontology for annotating photographs of apes. One can specify the type of ape, how old it is and what it is doing. Nevertheless, it should be borne in mind that the automated content description and annotation algorithms being developed cannot yet be expected to perform at the same level as a human annotator. The current state-of-the-art in automated annotation tends to operate at an extremely low level — for example, there is still no algorithm that can make an error-free distinction between images of cities and images of landscapes, or which can make an error-free decision as to the presence or absence of human faces in an image.

Evaluating the abilities of current algorithms requires a rather low level of annotation. Even though different modalities of annotation exist, such as description using keywords, annotations based on ontologies and free text description, the majority of these annotations are done by assigning keywords to images. For object recognition tasks, controlled vocabularies are often used, with the vocabulary being defined by the capabilities of the object recognition algorithm used (Winn et al., 2005). In applications which aim to do a more general image labelling using a larger number of keywords, the vocabulary is often uncontrolled, as in (Li and Wang, 2003). For example, the TRECVID 2005 high-level feature detection task tested automatic detection of only 10 concepts. The IBM MARVEL Multimedia Search Engine[1] extracts only six concepts in the online image retrieval demo version[2] (face, human, indoor, outdoor, sky, nature). Carbonetto et al. (2004) use a vocabulary of at most 55 keywords. The largest number of keywords have been used by Li and Wang (2003), who assigned 433.

A good way of collecting keywords which would be useful in an ontology describing images is to analyse the vocabularies used in the ground truth of image annotation and object recognition tasks. In this way, one can find out which words are important in applications and which words correspond to objects which can be detected using current state-of-the-art image understanding algorithms. After an overview of some approaches to collecting manual image annotations (Section 2.), we analyse the annotations which have been used in image and video understanding publications and evaluation campaigns in Section 3. The list of collected keywords is at the end of the paper in Section 6.

[1] http://www.research.ibm.com/marvel

[2] http://www.alphaworks.ibm.com/tech/marvel

2. Manual annotation collection methods

The manual annotation of images is a very labour-intensive and time-consuming task. Various systems to simplify the collection of image annotations or to receive input from a large number of people have been set up.

An interesting experiment is taking place on the *Gimp-Savvy Community-Indexed Photo Archive* website[3]. This archive contains more then 27 000 free photos and images, and the users of the site are requested to annotate the images using keywords which they are free to choose (tips on choosing keywords are made available[4]). That this "free annotation by all" approach has not been totally successful can be seen by the extremely large number of "junk" keywords on the master list[5] as well as the over-annotation (assignment of too many keywords) of many of the images.

On the *Flickr*[6] photo archive, people who upload photos may also assign keywords to them. These are then used to search for images. Other users may add comments to the images. There is no standardised keyword list, so this database represents a good example of the annotation practice of amateur photographers on their own images.

An innovative approach to collecting annotations of images by keywords has been developed by Ahn and Dabbish (2004). In their ESP game[7], they aim to make the annotation of images enjoyable. Players access the ESP game server and are paired randomly. They have no way of communicating with each other. Pairs of players are shown 15 images during the game, with the aim being for both players to type in the same keyword for an image so as to advance to the next. This is an intelligent way of avoiding the problem of "junk" keywords, as the pairs of players verify the keywords. Keywords which are typed often for an image are added to a "taboo" list shown for that image, and can no longer be entered as keywords by the players. The keywords entered correspond to the whole image, although the authors have discussed implementing, for example, a "shooting game", where the players have to click on the requested object. The Peekaboom game[8] from the same research group is of this type. An image search engine based on the keywords collected from the ESP game for about 30 000 images is accessible on the web[9].

An online annotation application aimed at collecting keywords for image regions is the LabelMe tool[10]. Here the user clicks the vertices of a polygon around an object and then enters a keyword describing the object. As the vocabulary is not controlled, multiple keywords and misspelled keywords often occur, as can be seen by examining the keyword statistics on the webpage[11]. This problem is solved by a verification step by the database administrators. At present[12], there are 101 verified keywords, the majority of which are shown in Table 2. The incentive to annotate the images is that the annotator may then download the latest annotations.

[3] http://gimp-savvy.com/PHOTO-ARCHIVE/

[4] http://gimp-savvy.com/PHOTO-ARCHIVE/tips_on_indexing.html

[5] http://gimp-savvy.com/cgi-bin/masterkeys.cgi

[6] http://www.flickr.com

[7] http://www.espgame.org

[8] http://www.peekaboom.org/

[9] http://www.captcha.net/esp-search.html

[10] http://people.csail.mit.edu/brussell/research/LabelMe/intro.html

[11] 400 keywords on the 29th of July 2005.

3. Analysis of Keywords used in Annotation Experiments

In this section we analyse the keywords that have been used in image annotation, categorisation and object recognition experiments and evaluation campaigns. To begin, a brief discussion on the difference between annotation and categorisation is presented in Section 3.1. Some methods currently used for collecting manual annotations of images are listed in Section 2. We then present an analysis of the keywords that have been used in image annotation experiments. The analysis was carried out in two steps. The first step consisted of creating a list combining all the keywords used in the experiments, datasets and evaluations considered and removing the unsuitable words (Section 3.2.). The second step was the categorisation of keywords (Section 3.3.). From a practical point of view, it is useful if the keywords are sorted into categories. When one is annotating images, this simplifies the choice of a word from the keyword list — one can select the category that the image belongs to in order to reduce the choice of keywords. The result of this analysis is a list of 525 keywords assembled from various sources and divided into 15 categories.

3.1. Annotation and Categorization

There are two approaches to associating textual information with images described in the literature: *annotation* and *categorisation*. In annotation, keywords or detailed text descriptions are associated with an image, whereas in categorisation, each image is assigned to one of a number of predefined categories (Chen and Wang, 2004). This can range from more general two category classification, such as *indoor/outdoor* (Szummer and Picard, 1998) or *city/landscape* (Vailaya et al., 2001) to more specific categories such as *African people and villages*, *Dinosaurs*, *Fashion* and *Battle ships* (Chen and Wang, 2004). Categorisation can be used as an initial step in image understanding in order to guide further processing of the image. For example, in (Wang et al., 2001) a categorisation into textured/non-textured and graph/photograph classes is done as a pre-processing step. *Recognition* is concerned with the identification of particular object instances. Recognition would distinguish between images of two structurally distinct cups, while categorisation would place them in the same class (Csurka et al., 2004). Recognition also has its uses in annotation, for example in the recognition of family members in the automatic annotation of family photos.

Categorisation can be considered as annotation in which one must choose from a fixed number of keywords (the categories) and one is limited to assigning one keyword to each image. The discussion of annotation and categorisation is therefore combined in this section.

[12] 27 July 2005

3.2. Overview of Visual Keywords

We present a collection of groups of keywords which have already been used for testing automated image annotation algorithms or in automated image and video annotation evaluation campaigns.

The 10 features which were tested in the TRECVID 2005 high-level feature detection task are described in Table 1. All 40 news concepts defined for TRECVID 2005 are available for download[13] (they are part of the LSCOM creation task (Hauptmann, 2004)).

Two categorisation tasks are part of the ImagEVAL[14] campaign: for the general image description task, the hierarchically organised global image categories shown in Figure 1 will be tested. There is also an object detection task, although the list of objects to be tested has not been finalised yet. The examples given are car, tree, chair, Eiffel Tower and American Flag.

The PASCAL Visual Object Classes Challenge 2005 consisted of classification and detection tasks for four objects: motorbikes, bicycles, people and cars. However, in the database collection set up as part of this challenge[15], five databases are provided with standardised ground truth object annotations. The keyword list arising from this standardisation is shown in Table 2.

As part of the EU LAVA project[16], a database consisting of 10 categories of images was made available[17]. These categories are: bikes, boats, books, cars, chairs, flowers, phones, roadsigns, shoes and soft toys.

Chen and Wang (2004) classified images into 20 categories: African people and villages, Beach, Historical buildings, Buses, Dinosaurs, Elephants, Flowers, Horses, Mountains and glaciers, Food, Dogs, Lizards, Fashion, Sunsets, Cars, Waterfalls, Antiques, Battle ships, Skiing and Deserts.

Two databases have been released by Microsoft Research in Cambridge[18]. The "Database of thousands of weakly labelled, high-res images" contains images divided into the following 23 categories: aeroplanes, cows, sheep, benches and chairs, bicycles, birds, buildings, cars, chimneys, clouds, doors, flowers, forks, knives, spoons, leaves, countryside scenes, office scenes, urban scenes, signs, trees, windows, miscellaneous. Some of these are divided into sub-classes, such as different views of cars. The "Pixel-wise labelled image database" contains 591 images in which regions are manually labelled using the following 23 labels: building, grass, tree, cow, horse, sheep, sky, mountain, aeroplane, water, face, car, bicycle, flower, sign, bird, book, chair, road, cat, dog, body, boat. The majority of the images are roughly segmented, although accurate segmentations of some of the images are available.

It is, of course, possible to greatly extend the number of categories if one is recognising specific objects, such as in the Caltech 101 category database[19] (Fei-Fei et al., 2004), which contains images of objects in the categories shown in Table 3.

If one restricts oneself to such specific categories, it is obviously possible to create many thousands. A set of 16 broader categories has been defined for the 15 200 images in the CEA-CLIC database (Moëllic et al., 2005). These are shown in Table 4.

A number of papers on automatic image or image region annotation have also been published. The following three all use parts of the Corel image database along with keywords usually extracted from the annotations accompanying the Corel images. The 55 keywords used by Carbonetto et al. (2004) are given in Table 5. Li and Wang (2003) used the largest number of keywords. They defined 600 categories of image, and to each category assigned on average 3.6 keywords. Each of the 100 images in each category was then assigned the same keywords associated with the category. For example, all images in the "Paris/France" category were assigned the keywords "Paris, European, historical building, beach, landscape, water", the images in the "Lion" category were assigned the keywords "lion, animal, wildlife, grass" and the images in the "eagle" category were assigned the keywords "wildlife, eagle, sky, bird". Barnard et al. (2003) used 323 keywords. These lists are not reproduced in this paper due to lack of space, but can be seen in (Hanbury, 2006).

3.3. Analysis of Visual Keywords

The aim of this analysis is to create a list of keywords which reflect the current interest in automated image annotation with keywords. These keywords could then serve as an initial controlled vocabulary for re-annotating the image collections used in previous experiments and for annotating new image collections.

3.3.1. Creation of a combined keyword list

The first step of the analysis consisted of creating a list combining all the keywords and categories used in the experiments, datasets and evaluations covered in Section 3.2. We then removed words which were considered to be unsuitable. These include place names, such as "Australia", "Boston" and "New Zealand", which, even for a human, are very difficult to assign to images for which one has no supplementary information. Confusing keywords, such as "history" and "north", and keywords requiring too high a level of a priori semantic information, such as "landmark" and "rare animal" were also removed. We have not yet collected statistics on how often a single keyword appears in different lists.

3.3.2. Categorisation of keywords

From a practical point of view, it is useful if the keywords are sorted into categories. When one is annotating images, this simplifies the choice of a word from the keyword list —

[13] http://www-nlpir.nist.gov/projects/tv2005/LSCOMlite_NKKCSOH.pdf

[14] http://www.imageval.org

[15] http://www.pascal-network.org/challenges/VOC/

[16] http://www.l-a-v-a.org

[17] ftp://ftp.xrce.xerox.com/pub/ftp-ipc/

[18] Downloadable here: http://www.research.microsoft.com/vision/cambridge/recognition/default.htm. Version 1 of the pixel-wise labelled image database has been ignored here, as it forms a subset of version 2.

[19] http://www.vision.caltech.edu/Image_Datasets/Caltech101/Caltech101.html

Keywords	Segment contains video of ...
People walking/running	more than one person walking or running
Explosion or fire	an explosion or fire
Map	a map
US flag	a US flag
Building exterior	the exterior of a building
Waterscape/waterfront	a waterscape or waterfront
Mountain	a mountain or mountain range with slope(s) visible
Prisoner	a captive person, e.g., imprisoned, behind bars, in jail, in handcuffs, etc.
Sports	any sport in action
Car	an automobile

Table 1: The 10 features which were tested in the TRECVID 2005 high-level feature detection task.

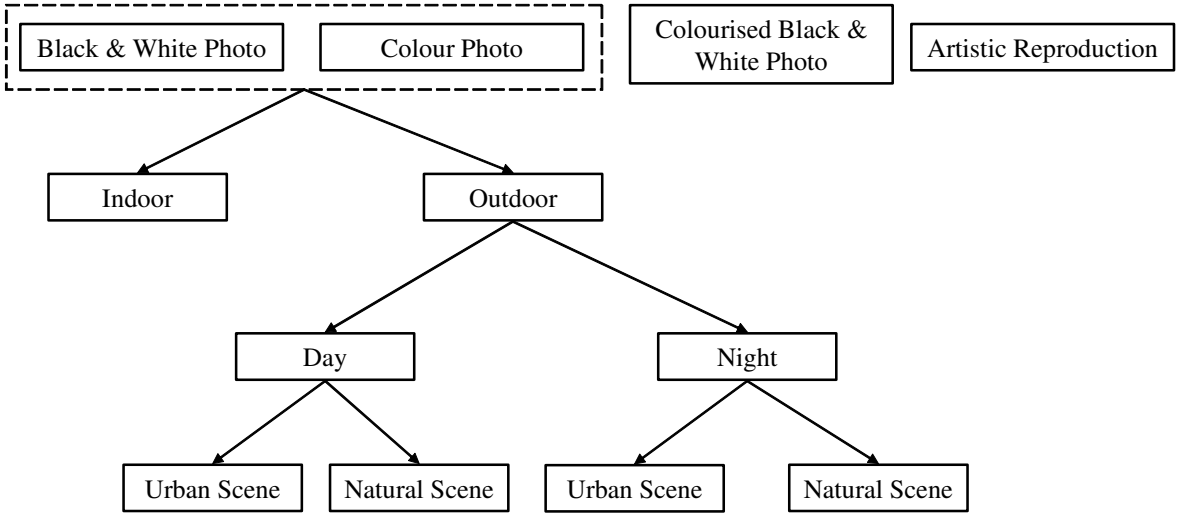

Figure 1: The hierarchy of keywords used in the global image characteristics task of ImagEVAL.

one can select the category that the image belongs to in order to reduce the choice of keywords. The 16 categories of the CEA-CLIC database (Moëllic et al., 2005), with some minor changes, turn out to be well-suited to grouping the combined list of keywords.The changes are:

- the fusion of the "Architecture" and "City" categories to form an "Architecture / City" category. This was done as it is often difficult for an annotator to decide between these two categories.
- the addition of an "Abstract / Global" category to contains words such as "female" and "exterior".
- the removal of the "Mathematics" category, which has no members in the list of keywords collected.
- the removal of the "linguistic" category, as this is an image category and not a keyword category.
- the addition of the "Anatomy and Medicine" category, which at present includes one keyword, but can be expanded later.

The list of categories and their descriptions are in Table 6. We assigned each of the keywords in the combined list to at least one category. A few keywords were assigned to two categories, for example, "grass" appears in the "Texture" and "Nature and Landscapes" categories. A table showing the keywords assigned to each category is given in Section 6. A histogram of the number of keywords per category is shown in Figure 2.

One can see from this histogram that the categories "Objects", "Nature and Landscapes" and "Zoology" contain the most keywords, which could be an indicator that these categories have received the most attention in past research on automated image annotation and categorisation. This could be because of the image databases used — the Corel databases, for example, appear to contain a high proportion of natural and animal images. The man-made objects appear to be more prevalent in the databases designed for object categorisation experiments.

4. Conclusion

We analyse the keywords which have been used to annotate images in a number of image retrieval publications and evaluation campaigns. A significant contribution is the creation of a combined keyword list based on these keywords. From this analysis one can see that the main automated annotation effort has been directed at images of everyday objects; nature and landscapes; and animals (zoology). As

aeroplaneSide	apple	background	bicycle	bicycleSide
bookshelf	bookshelfFrontal	bookshelfPart	bookshelfSide	bookshelfWhole
bottle	building	buildingPart	buildingRegion	buildingWhole
can	car	carFrontal	carPart	carRear
carSide	cd	chair	chairPart	chairWhole
coffeemachine	coffeemachinePart	coffeemachineWhole	cog	cow
cowSide	cpu	desk	deskFrontal	deskPark
deskPart	deskWhole	donotenterSign	door	doorFrontal
doorSide	face	filecabinet	firehydrant	freezer
frontalWindow	head	keyboard	keyboardPart	keyboardRotated
light	motorbike	motorbikeSide	mouse	mousepad
mug	onewaySign	paperCup	parkingMeter	person
personSitting	personStanding	personWalking	poster	posterClutter
pot	printer	projector	roadRegion	screen
screenFrontal	screenPart	screenWhole	shelves	sink
sky	skyRegion	sofa	sofaPart	sofaWhole
speaker	steps	stopSign	street	streetSign
streetlight	tableLamp	telephone	torso	trafficlight
trafficlightSide	trash	trashWhole	tree	treePart
treeRegion	treeWhole	walksideRegion	wallClock	watercooler
window				

Table 2: The keywords in the PASCAL Object Recognition Database Collection (the prefix "PAS" has been removed from each keyword).

Faces	Faces easy	Leopards	Motorbikes	accordion	airplanes
anchor	ant	barrel	bass	beaver	binocular
bonsai	brain	brontosaurus	buddha	butterfly	camera
cannon	car side	ceiling fan	cellphone	chair	chandelier
cougar body	cougar face	crab	crayfish	crocodile	crocodile head
cup	dalmatian	dollar bill	dolphin	dragonfly	electric guitar
elephant	emu	euphonium	ewer	ferry	flamingo
flamingo head	garfield	gerenuk	gramophone	grand piano	hawksbill
headphone	hedgehog	helicopter	ibis	inline skate	joshua tree
kangaroo	ketch	lamp	laptop	llama	lobster
lotus	mandolin	mayfly	menorah	metronome	minaret
nautilus	octopus	okapi	pagoda	panda	pigeon
pizza	platypus	pyramid	revolver	rhino	rooster
saxophone	schooner	scissors	scorpion	seahorse	snoopy
soccer ball	stapler	starfish	stegosaurus	stop sign	strawberry
sunflower	tick	trilobite	umbrella	watch	water lilly
wheelchair	wildcat	windsor chair	wrench	yin yang	

Table 3: The 101 categories used by Fei-Fei et al. (Fei-Fei et al., 2004).

these keywords were extracted from annotations of existing image datasets, they should be well-suited to a more precise re-annotation of these same datasets. For the same reason, they are also suited to verify the applicability of newly developed image ontologies intended to represent portrayable entities and objects.

A disadvantage is that while the keywords in this list certainly correspond well to the images used in image annotation experiments so far, there is no guarantee that these images are representative of all possible electronic images. It would therefore be useful to compare this collection of keywords to an ontology constructed in a more rigorous way, such as the ontology of portrayable objects based on WordNet (Zinger et al., 2005). This should provide a useful link between possible portrayable objects and those that are often found in images, or that are of interest to image understanding researchers.

5. References

Luis von Ahn and Laura Dabbish. 2004. Labeling images with a computer game. In *Proc. ACM CHI*, pages 319–326.

Kobus Barnard, Pinar Duygulu, Nando de Freitas, David Forsyth, David Blei, and Michael I. Jordan. 2003. Matching words and pictures. *Journal of Machine Learning Research*, 3:1107–1135.

Category	Description
Food	Images of food, and meals.
Architecture	Images of architecture, architectural details, castles, churches, Asian temples.
Arts	Paintings, sculptures, stained glass, engravings.
Botanic	Various plants, trees, flowers.
Linguistic	Images containing text areas.
Mathematics	Fractals.
Music	Images of musical instruments.
Objects	Images representing everyday objects such as coins, scissors, etc.
Nature & Landscapes	Landscapes, valley, hills, deserts, etc.
Society	Images with people.
Sports & Games	Stadiums, items from games and sports.
Symbols	Iconic symbols, roadsigns, national flags (real and synthetic images)
Technical	Images involving transportation, robotics, computer science.
Textures	Rock, sky, grass, wall, sand, etc.
City	Buildings, roads, streets, etc.
Zoology	Images of animals (mammals, reptiles, bird, fish).

Table 4: The 16 categories in the CEA-CLIC image database and their descriptions (Moëllic et al., 2005).

airplane	astronaut	atm	bear	beluga	bill	bird
boat	building	cheetah	church	cloud	coin	coral
cow	crab	dolphin	earth	elephant	fish	flag
flowers	fox	goat	grass	ground	hand	horse
house	lion	log	map	mountain	mountains	person
pilot	polarbear	rabbit	road	rock	sand	sheep
shuttle	sky	snow	space	tiger	tracks	train
trees	trunk	water	whale	wolf	zebra	

Table 5: The 55 keywords used by Carbonetto et al. (Carbonetto et al., 2004).

#	Category	Description
0	Abstract / Global	Words which describe the whole image or which are applicable to more than one class of objects.
1	Food	Food and meals.
2	Architecture / City	Architecture, architectural details, castles, churches, Asian temples, buildings, roads, streets, etc.
3	Arts	Paintings, sculptures, stained glass, engravings.
4	Botanic	Plants, trees, flowers.
5	Objects	Everyday objects such as coins, scissors, etc.
6	Nature & Landscapes	Landscapes, valley, hills, deserts, etc.
7	Society	People, groups of people, activities undertaken by society (celebrations, parades, war, etc.).
8	Sports & Games	Stadiums, items from games and sports.
9	Symbols	Iconic symbols, roadsigns, national flags
10	Technical	Transportation, robotics, computer science.
11	Textures	Words which describe a texture.
12	Zoology	Animals (mammals, reptiles, birds, fish).
13	Anatomy and Medicine	Biological organs, anatomical diagrams, etc.
14	Music	Musical instruments.

Table 6: The 15 categories of the combined keyword list and their descriptions. The first column contains a category number.

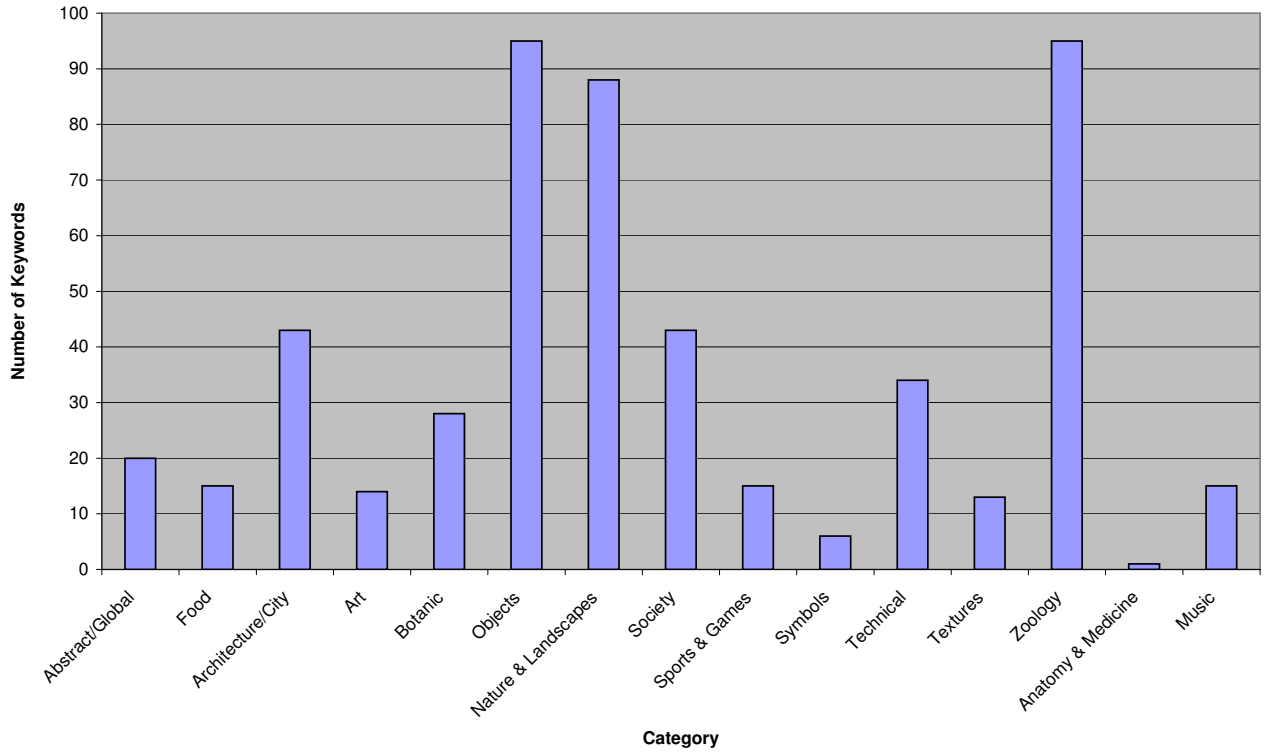

Figure 2: The number of keywords in each category.

Peter Carbonetto, Nando de Freitas, and Kobus Barnard. 2004. A statistical model for general contextual object recognition. In *Proceedings of the ECCV 2004, Part I*, pages 350–362.

Yixin Chen and James Z. Wang. 2004. Image categorization by learning and reasoning with regions. *Journal of Machine Learning Research*, 5:913–939.

Gabriella Csurka, Christopher R. Dance, Lixin Fan, Jutta Willamowski, and Cedric Bray. 2004. Visual categorization with bags of keypoints. In *Workshop on Statistical Learning in Computer Vision (at ECCV)*.

L. Fei-Fei, R. Fergus, and P. Perona. 2004. Learning generative visual models from few training examples an incremental bayesian approach tested on 101 object categories. In *Proceedings of the Workshop on Generative-Model Based Vision*, June.

Allan Hanbury. 2006. Review of image annotation for the evaluation of computer vision algorithms. Technical Report PRIP-TR-102, PRIP, TU Wien, January.

Alexander G. Hauptmann. 2004. Towards a large scale concept ontology for broadcast video. In *Proceedings of the Third Intl. Conf on Image and Video Retrieval*, pages 674–675.

Clement H. C. Leung and Horace Ho-Shing Ip. 2000. Benchmarking for content-based visual information search. In *Proceedings of the 4th International Conference on Advances in Visual Information Systems*, pages 442–456.

Jia Li and James Z. Wang. 2003. Automatic linguistic indexing of pictures by a statistical modeling approach. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 25(9):1075–1088.

Pierre-Alain Moëllic, Patrick Hède, Gregory Grefenstette, and Christophe Millet. 2005. Evaluating content based image retrieval techniques with the one million images clic testbed. In *Proceedings of the Second World Enformatika Congress, WEC'05*, pages 171–174.

A. Th. (Guus) Schreiber, Barbara Dubbeldam, Jan Wielemaker, and Bob Wielinga. 2001. Ontology-based photo annotation. *IEEE Intelligent Systems*, 16(3):66–74.

M. Szummer and R. W. Picard. 1998. Indoor-outdoor image classification. In *Proc. IEEE International Workshop on Content-based Access of Image and Video Databases*, pages 42–51.

A. Vailaya, M. A. T. Figueiredo, A. K. Jain, and H.-J. Zhang. 2001. Image classification for content-based indexing. *IEEE Transactions on Image Processing*, 10(1):117–130.

James Z. Wang, Jia Li, and Gio Wiederhold. 2001. SIMPLIcity: Semantics-sensitive integrated matching for picture libraries. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(9):947–963.

J. Winn, A. Criminisi, and T. Minka. 2005. Object categorization by learned universal visual dictionary. In *Proceedings of the International Conference on Computer Vision(ICCV)*.

S. Zinger, C. Millet, B. Mathieu, G. Grefenstette, P. Hède, and P.-A. Moëllic. 2005. Extracting an ontology of portrayable objects from WordNet. In *Proceedings of the MUSCLE/ImageCLEF Workshop on Image and Video Retrieval Evaluation*, pages 17–23, Vienna, Austria, September.

6. Combined Keyword List

The following table lists the combined keyword list. It is a simple two-level hierarchy, with 15 headings at the top level (in bold). Note that some words are repeated under more than one heading.

Abstract / Global				
background	black	black_and_white	blue	color
exterior	female	fractal	green	group
indoor	interior	male	nature	orange
outdoor	pattern	red	shadow	yellow

Food				
apple	cuisine	dessert	drink	feast
food	fruit	grapes	herb_spice	orange
pizza	pumpkin	strawberry	vegetable	wine

Architecture / City				
arch	architecture	building	castle	chimney
church	city	college	column	courtyard
dock	fountain	harbor	historical_building	hotel
house	hut	industry	kitchen	market
minaret	monument	mosque	museum	office
pagoda	palace	park	pillar	restaurant
roof	ruin	shop	skyline	stairs
statue	street	studio	temple	tower
town	village	window		

Art Objects				
art	carving	decoration	design	drawing
graffiti	mosaic	mural	painting	photo
poster	sculpture	statue	still_life	

Botanic				
apple	bonsai	botany	branch	bush
cactus	flower	foliage	fungus	grapes
leaf	lichen	log	moss	mushroom
orchid	palm	perenial	petal	plant
pumpkin	rose	seed	strawberry	sunflower
tree	tulip	water_lily		

Objects (man-made everyday)				
anchor	antique	atm	balloon	barbecue
barrel	bath	bead	bench	bicycle
binoculars	book	bookshelf	bottle	camera
can	candy	card	cd	cellphone
chair	clock	cloth	coffee_machine	cog
coin	cup	currency	decoration	desk
dish	dogsled	doll	door	dress

Easter_egg	fabric	fan	fence	file_cabinet
fire_hydrant	firearm	firework	flag	floor
freezer	furniture	glass	gun	hat
headphones	horn	jewelry	keyboard	lamp
light	map	marble	mask	medicine
money	mousepad	mug	paper	paper_cup
parking_meter	pill	pot	printer	projector
relic	scissors	screen	shelves	shoe
sink	sofa	speaker	sponge	stamp
stapler	table	telephone	textile	tool
toy	traffic_light	trash	umbrella	wall
watch	watercooler	wheelchair	wood	wrench

Nature and Landscapes				
agriculture	autumn	barnyard	bay	beach
canyon	cave	cliff	cloud	coast
coral	crop	crystal	dawn	desert
dune	dusk	earth	farm	field
flowerbed	forest	frost	frozen	garden
gem	glacier	grass	ground	hill
ice	iceberg	island	lake	landscape
maritime	meadow	mountain	night	ocean
pastoral	path	peak	plain	planet
polar	pyramid	rapids	reef	reflection
river	road	rock	ruin	runway
rural	sail	sand	shell	shore
shrine	sky	smoke	snow	space
spring	star	steam	stone	sub_sea
summer	sun	sunset	surf	tree
tropical	tundra	valley	vegetation	vineyard
volcano	wall	water	waterfall	wave
wind	winter	woodland		

Society				
astronaut	baby	ballet	barbecue	battle
builder	business	child	Christmas	costume
couple	diver	face	fashion	festival
fight	glamour	graffiti	guard	hand
head	holiday	home	hunter	leisure
man	model	occupation	parade	person
pilot	pomp_and_pageantry	religion	royal	sacred
science	travel	tribal	war	woman
work	worship	youth		

Sports and Games				
fitness	football	game	golf	kungfu
play	polo	race	rafting	recreation
rodeo	ski	sport	tennis	wind_surfer

Symbols

public_sign	road_sign	sign_do_not_enter	sign_stop	sign_oneway
sign_yield				

Technical				
aeroplane	aviation	balloon	battle_ship	boat
bridge	bus	cannon	canoe	car
communication	engine	ferry	helicopter	highway
jet	lighthouse	locomotive	machine	military
molecule	motorcycle	pathology	railroad	road
runway	sailboat	ship	space_shuttle	street
tallship	train	transportation	vehicle	

Textures				
fabric	fire	glass	grass	ground
ice	marble	sand	skin	stone
textile	texture	wood		

Zoology				
anemone	angelfish	animal	ant	antelope
antlers	bear	beaver	beetle	bird
bobcat	bull	butterfly	camel	caribou
cat	caterpillar	cheetah	coral	cougar
cow	coyote	crab	crayfish	crocodile
cub	deer	dinosaur	dog	dolphin
dragonfly	eagle	elephant	elk	feline
fish	flamingo	foal	fowl	fox
giraffe	goat	hawk	hedgehog	herd
hippopotamus	horn	horse	iguana	insect
jaguar	kangaroo	kitten	leopard	lion
lizard	llama	lobster	lynx	mammal
moth	mouse	nest	ocean_animal	octopus
owl	panda	penguin	pet	pigeon
polar_bear	predator	primate	rabbit	reptile
rhinoceros	rodent	rooster	scorpion	seahorse
seal	sheep	skin	snake	sponge
squirrel	starfish	tiger	turtle	whale
wildcat	wildlife	wolf	young_animal	zebra

Anatomy and Medicine				
brain				

Musical Instruments				
accordion	cello	double_bass	electric_guitar	guitar
horn	mandolin	piano	piano_grand	saxophone
trombone	trumpet	tuba	viola	violin