Ontology-Assisted Object Detection: Towards the Automatic Learning with Internet

Francesco Setti[1], Dong-Seon Cheng[2], Sami Abduljalil Abdulhak[3], Roberta Ferrario[1], and Marco Cristani[3,4]

[1] ISTC–CNR, via alla Cascata 56/C, I-38123 Povo (Trento), Italy
[2] Hankuk University of Foreign Studies, Yongin, Gyeonggi-do, Corea
[3] Università degli Studi di Verona, Strada Le Grazie 15, I-37134 Verona, Italy
[4] Istituto Italiano di Tecnologia (IIT), via Morego 30, I-16163 Genova, Italy

Abstract. Automatic detection approaches depend essentially on the use of classifiers, that in turn are based on the learning of a given training set. The choice of the training data is crucial: even if this aspect is often neglected, the visual information contained in the training samples can make the difference in a detection/classification scenario. A good training set has to be sufficiently informative to capture the nature of the object under analysis, but at the same time has to be generic enough to avoid overfitting and to cope with new instances of the object of interest. In this paper we follow those approaches that pursue automatic learning from Internet data. We try to show how such training set can be made more appropriate by leveraging on semantic technologies, like lexical resources and ontologies, in the task of retrieving images from the Web through the use of a search engine. Experiments on several object classes of the CalTech101 dataset promote our idea, showing an average increment on the detection accuracy of about 8%.

Keywords: object detection, one-class SVM, machine learning, ontology, semantic search.

1 Introduction

In these last years, intriguing applications of robotics, computer vision and pattern recognition have emerged, essentially aimed at providing robots with the visual intelligence of recognizing previously unseen objects.

Humans interact with their world each day largely based on visual understanding. The human visual system is a highly capable modeling and inference device. It quickly learns the appearances for new objects that are encountered, combines weak sources of information from multiple views, attends only to the most useful regions, and integrates numerous priors.

In particular, the human visual system works in tight connection with the cognitive system. In cases of unclear perception, the visual system can be guided by memories of previous experiences that can be recalled by associative mechanisms (you don't see the shepherd very well, but you see a familiar shape close to a flock

A. Petrosino (Ed.): ICIAP 2013, Part II, LNCS 8157, pp. 191–200, 2013.

of sheep), while in cases of uncertain classification of perceived objects/scenes, useful inferences can be made on the basis of contextual information (you don't know exactly what that object is, but it moves on the surface of the sea, it should be some kind of watercraft).

Such capabilities can be hardly embedded in an autonomous agent; in fact, automatic recognition approaches depend essentially on the use of classifiers, that in turn are based on the learning of a given training set. The choice of the training data is crucial: even if this aspect is often neglected, the visual information contained in the training samples can make the difference in a detection scenario. A good training set has to be sufficiently informative to capture the nature of the object under analysis, but at the same time has to be generic enough to avoid overfitting and to cope with new instances of the object of interest.

In most of the object detection challenges, training sets are given together with the testing set, under the form of standard benchmarks. This actually hides the interesting problem of understanding which are the most informative training samples to be collected, for ensuring a fruitful learning step.

Recently, this problem has been taken into account in a cross-disciplinary competition, the Semantic Robot Vision Challenge [4]. In this competition, fully autonomous robots receive a text list of objects that they have to find. In a training phase, each robot is required to train visual classifiers for each of the objects, based on images downloaded from the Internet. This is followed by an environment exploration phase, where robots must search a realistic environment created by the organizers in order to locate instances of the object categories listed during training.

This workflow opens up to an intriguing perspective, that is, automatically learning from Internet data. If we metaphorically see the training set as playing the role of the previous visual experiences stored in the memory of a human, we easily see that we miss a part of the story: human memories are connected in various ways through associative and inferential mechanisms used to retrieve relevant and useful information. So, if the robot uses textual inputs to search the Internet, we can think of mimicking human cognitive mechanisms by providing other relevant inputs in textual format, i.e. by adding words connected to the one identifying the object that is being searched[1].

In this paper, we follow the trend of learning from Internet; we think that automatic sampling of Web images for learning is feasible with the massive growth of user-tagged images on social media websites and this can be extremely fruitful if assisted by the knowledge that lexical resources (such as WordNet [3]) and ontologies (such as DOLCE [6]) can offer, in order to automatically associate

[1] An alternative way to address such issue would be that of mimicking the associative mechanism by searching through images already structured in a network, like that of the dataset that is being built within the ImageNet initiative (`http://www.image-net.org/`), where images are grouped based on their tags, organized according to the WordNet structure. Nonetheless, our aim in the paper is that of retrieving a good training set of images from the Internet, not from a well-organized dataset. For this reason, we won't enter in a discussion of the results of ImageNet.

related concepts to the search. For this reason, we propose a method based on WordNet and an ontological analysis working together to generate a set of words connected to the object we are interested to detect. These new words, together with the original ones, are used as keywords for Google search in order to create an enriched training set. SIFT descriptors and bag-of-words classification framework are then used to train a single class SVM classifier.

On one side, the Google image search engine allows to specify heterogeneous expressions in natural language, providing highly relevant images as output. On the other hand, lexical resources and ontologies allow to exploit semantic relations between concepts expressed by target words in the search of visual information.

Our approach exploits standard local object descriptors, such as SIFT, collected and quantized as bags of words. Almost all the works on this topic exploit the same kind of features. Given a class of objects of interest and the words used to denote them, lexical resources tell us *which* are the words semantically connected to them and ontologies tell us *why* they are connected, by referring to the relations holding between the objects denoted by such words (the connection between the words/concepts is used to explain *how* the denoted objects are related). Thus, what we can expect by the employment of lexical resources is a broader coverage (more relevant results can be obtained by adding connected words to the search), while the employment of ontologies should enhance precision, as words that are connected in a way that is not interesting for the task should be automatically pruned away by the constraints that are formally expressed in the ontology.

The Caltech101 object recognition dataset has been used for the tests, consistently giving improvements of performances in terms of detection accuracy on all the classes where our approach can be applied (44), for an average of about 8%. We decided to keep the classifier as simple as possible, for this reason we are confident that the improvements in performances have to be attributed to our image selection strategy.

The rest of the paper is organized as follows: Sec. 2 presents the state of the art and related works; Sec. 3 sketches the main intuitions and motivations behind the introduction of lexical resources and ontologies in the loop; Sec. 4 illustrates the method adopted in the present study, while Sec. 5 displays the results of the experiments and, finally, Sec. 6 concludes the paper and draws the lines for future directions of research.

2 State of the Art

Few and recent are the approaches that can be related to our methodology, mostly based on incremental learning. In [16], an incremental version of Support Vector Machines is used to acquire visual categories. In the context of human-robot interaction, some recent approaches also explore the combination of incremental learning and interaction with teachers to ground vocabulary about physical objects [10,5,13].

More similar to our scheme, one common and straightforward way is to directly use the top-ranked images from Web search for classifier learning [12,14]. On the other hand, [8] uses lexical sources to help an image retrieval system.

Curious George [7] is a robot developed for the Semantic Robot Vision Challenge (SRVC) competition. Its object recognition component uses images retrieved from Computer Vision databases in the Web to build several classifiers based on SIFT features, shape and deformable parts models, which are subsequently combined for object recognition.

In [15] the authors propose an unsupervised learning algorithm for visual categories using potential images obtained from the Web and being inspired by [1]. The idea is to produce several images by translating the category name into different languages and crawling the Web for images using those translated terms. The negative class is collected from random images obtained from different categories.

A multi-modal approach using text, metadata and visual features is used in [11]. Candidate images are obtained by a text-based Web search querying on the object identifier. The task is then to prune irrelevant images away and re-rank the remainder. Re-ranking happens by using a Bayes posterior estimator trained on the text surrounding the image and meta data features; after that, the top-ranked images are improved by eliminating drawings and symbolic images using an SVM classifier previously trained with a hand-labeled dataset.

3 Prospective Exploitation of Lexical Resources and Ontologies

With the aim of improving (both in terms of coverage and precision) a text based search of tagged images in the Internet, in this section we will try to show why an approach that integrates lexical resources and ontologies is needed.

Lexical resources, like WordNet, are built as networks, whose nodes are connected by lexical and semantic relations; each node in WordNet is a *synset* (set of synonymous words, expressing the same concept), and some of the relations connecting synsets are *hyponymy* (linking a more generic concept to more specific ones) and its opposite relation, *hyperonymy* (linking a more specific concept to more general ones), *meronymy* (linking concepts denoting parts with concepts denoting the corresponding whole), and so on. Intuitively, given a word, a search engine could point to it, retrieve all other words in the synset and include them in the search and then expand the search by associating to the original word its neighbors in the network. Even though often presented as an ontology, WordNet was not originally meant to be used as such, so it disregards some distinctions that are ontologically relevant. Take for instance the "part of" relation: WordNet meronymy includes the proper part of relation (like in "a finger is part of a hand)", the constitution relation (like in "milk is part of cappuccino"), and the membership relation (like in "a sheep is part of a flock") and does not distinguish between them. Nonetheless, such distinction may be relevant (probably adding "milk" to "cappuccino" will produce more false positives) and most of all, it

becomes relevant when the robot has to search real objects in the environment, as for instance, simplifying a lot, it will look for an object located within the boundaries of another object when instructed with the "proper part of" information and an object among other similar objects if instructed with the "member of" information. Such distinctions are, on the other hand, deeply analyzed and expressed with formal axioms by top-level ontologies, like DOLCE [6].

To sum up, we could say that lexical resources and ontologies tackle two different "semantic layers" and could be used in an integrated way to exploit their complementary strengths, as already pointed out in [9], that also mentions a distinction that is particularly meaningful to our purposes:

> *Lexical resources are conceptually very dense but they do not have a dense network of constraints. On the other hand, ontologies, specially top-level ones, are not densely populated but offer a dense network of constraints for their concepts.* [9, 187]

Being densely populated, WordNet promises to enhance coverage results, while ontologies, thanks to their being densely constrained, should make results more precise.

For the sake of this paper, we won't go into details concerning the choice of the ontology to be used, as it would imply entering a lively debate in the ontology community. In the case of an automatic agent that has to learn (for example, a robot), we could imagine that a fairly small and simple ontology would suffice. Our mentioning DOLCE is due to the fact that, methodologically, and in view of applying in the future such method on open world scenarios, even these micro domain-relative ontologies should be built following the principles and as specifications of foundational or top-level ontologies, so that they are as much well-founded and as much interoperable as possible with other ontologies.

Thus, our idea is, once we have a description of the environment in which the automatic agent should move, we build a small ontology of it based on DOLCE, trying to express as many relations between the fundamental entities of the domain, so to have a possibly scarcely populated, but rather densely constrained, ontology.

4 Proposed Method

In this work we propose a unique framework meant to automatically generate a training set for object detection. Given the object we want to detect in an images' testing set, the simplest way to use the Internet to generate the training set is to use the Google image search engine by searching the name of the object [12,14]. This strategy has two main drawbacks: first, several keywords are ambiguous. Words can have different meanings and refer to different objects, depending on the context (e.g. *mouse* can refer both to the animal and to the computer device); in this case, Google answers with images of both subjects, without any distinction (see Fig. 1). Thus, the training set will be formed by images of two different (and semantically mixed) classes. On the other hand, when an object can have very

Fig. 1. Example of the first 8 images retrieved by Google image search engine for the keywords "mouse" and "mouse+animal"

different characteristics, the top-ranked images are usually referring to the most common configuration, creating a training set only representative of a specific subset of the testing set (e.g. *face* can be a frontal face or a face profile, but almost all images given by Google are relative to frontal faces). Moreover, in some cases the name of a class is also the name of a very famous person or character (e.g. *bush* is both a small tree and a very popular surname). Second, in order to obtain a good enough number of training images, you need to use also low-ranked images, increasing the probability to take into account not representative (or wrong) images.

The strategy we are proposing in this paper is based on a prior knowledge of the environment, formally given by an ontology (used to represent knowledge of the primitive entities belonging to the environment and their properties and relations) and an electronic lexical database of English terms, organized as a thesaurus – WordNet (used to represent terminological and lexical knowledge). Since both the entities can be uploaded in an autonomous agent, we exploit this hypothesis, and given a set of images for testing and the name of the class to detect, we design the following learning scheme:

1. ask WordNet for a set of words connected to the class name (synonym, hyperonym, hyponym and meronym);
2. use ontological analysis to select a subset of connected words that are related to visual specifications of the original word; i.e. ignore the words that do not provide any difference in the appearance of the object (e.g. the material of which an object is built is not relevant to detect its appearance).
3. search with Google Image the first N images with each of the selected keywords (together with the class name)
4. train a bag-of-words model by using these images as training set; in particular:
   - for each image in the training set, extract dense SIFT descriptors, discretize them referring to a universal codebook and build the normalized histogram of words;
   - train a Single Class Support Vector Machine classifier;
   - for each image in the testing set, extract dense SIFT descriptors, discretize them referring to the universal codebook and build the normalized histogram of words;
   - apply the previously trained SVM classifier onto these image descriptors.

We have performed some experiments to validate our idea, and in the following section we will show some results.

5 Experiments

We present here some preliminary results obtained by using the proposed algorithm on the Caltech101 dataset [2]. For these preliminary experiments we took into account only the *hyponymy* relation, thus only those words with at least one hyponym can be processed, meaning 44 classes over 101. For each class, we manually selected the additional words related to visual specifications.[2] Then, we employed these additional words to provide keywords for the Google image search, together with the class name (e.g. for the class *helicopter* the following keywords have been used: 'helicopter', 'helicopter+cargo', 'helicopter+shuttle', 'helicopter+skyhook'). Table 1 shows all the classes we processed and the additional words used in the experiments.

After that, we performed a set of experiments in order to evaluate the contribution of this enriched set of words. For each class, the testing set is composed by the full set of test images in Caltech101 dataset. As universal codebook we used the one provided by the ImageNet Large Scale Visual Recognition Challenge 2011 (ILSVRC2011) development kit, which allows a discretization of SIFT features computed over millions of images into 1000 visual words.

Each training set is formed by 100 images automatically taken from the Internet. We performed experiments by changing the composition of the training set by means of a parameter R, varying this number from 0 to 1 with step 0.1; in practice, R represents the ratio of images taken by using also the additional words as keyword, equally split within all words. Thus, $R = 0$ means that only the class name is used for the search, while $R = 1$ only the enriched combinations <class name + additional words> are used.

A One-class SVM classifier has been trained on each training set's composition and then it has been applied to the testing set.

As performance measure we use the detection accuracy, defined as the percentage of correctly detected images over all the testing set.

Figure 2 shows the average detection accuracy over 44 classes, and the histogram of the best performing composition of the training set for all the classes.

We mostly improve the performances of the detection process, both in terms of average accuracy (about 8%) and in terms of the best performing training set: 23 object classes over 44 are best performing with the enriched words ($R = 1$), and 37 are best performing with $R \geq 0.7$. One single class ('pizza') is best performing with $R = 0$, this is probably due to the extremely low number of hyponyms for

[2] In order to automatize this task, an ontology discriminating visible and non visible objects, properties and relations should be built. Moreover, in the mentioned contest, the task consisted in finding objects of some class, but we could also be interested in finding objects with some specific property, or objects involved in some particular event. The more the task is complicated, the more the ontology could be of help in finding the most appropriate terms to associate in the search.

Table 1. Object classes and additional words used in the experiments

Class name	Additional words	best R
airplane	airliner, amphibian, amphibian aircraft, attack aircraft, biplane, fighter, bomber, fighter aircraft, hanger queen, hydroplane, monoplane, multiengine airplane, multiengine plane, propeller plan, reconnaissance plane, seaplane, tanker plan	1
anchor	granel, mooring, mushroom, sheet, waist	0.5
ant	army, bulldog carpenter, driver, fire, Formica rufa, legionary, little black, Monomorium minimum, slave, wood	0.8
barrel	beer, beer keg butt, hogshead, keg, pickle, shook, tun, wine, wine cask	1
bass	basso continuo, continuo, figured, ground, thorough	1
beaver	Castor canadensis, Castor fiber, New World, Old World	1
bonsai	ming tree	0.8
brain	ego, noddle, subconscious, subconscious mind, tabula rasa, unconscious, unconscious mind	1
butterfly	danaid, lycaenid, pierid, ringlet, sulfur, sulphur	1
camera	box, box Kodak, candid, digital, flash, Polaroid, Polaroid land, portrait, reflex	1
cannon	basilisk, culverin, harpoon gun	1
chair	amrchair, barber, chaise, chaise longue, dabed, Eames, feeding, fighting, folding, garden, highchair, lawn, rocker, rocking, side, straight, swivel, throne, wheelchair	1
crab	Alaska, Alaska king, Alaskan king, Cancer irroratus, Cancer magister, Dungeness, fiddler, Jonah, king, Menippe mercenaria, Paralithodes camtschatic, pea, rock, spider, stone, swimming	0.9
crayfish	American, ecrevisse, Old World	0.4
crocodile	African, Asian, Crocodylus niloticus, Crocodylus porosus, Nile	1
cup	beaker, chalice, coffee, clylix, Dixie, drinking, globlet, grace, kylix, measuring, moustache, mustaches, paper, syphus, teacup	0.9
dolphin	Corphaena, equisetis, Corphaena Hippurus	0.6
elphant	African, Elephas maximus, gomphothere, Indian, Loxodonta african, mammoth, rogue	0.9
face	counenance, physiognomy, phiz, smiler, visage, mug	1
gramophone	victrola	1
headphone	receiver, telephone receiver	1
hedgehog	Old world porcupine, New World procupine	0.4
helicopter	Cargo, shuttle, skyhook	1
ibis	Wood, wood stroke, Ibis, sacred, Threskiornis aethiopica	0.9
kangaroo	Gaint, brush, wallaby, hyspiprymnodon moschatus, great grey, kangaroo rat, Macropus giganteus, musk, rat	1
lamp	Calcium light,candle, discharge, electric, flash, flash bulb, gas, hurricane, kerosine, lantern, neon, oil, neon inductionn, neon tube, photoflash, rear, rear light, spirit, spot, spotlight, storm, storm lantern, street, streetlight, tail, taillight, taper, tornado lantern, wax light	1
leopard	leopardess, panther	0.9
llama	Domestic, alpaca, gaunaco, lama guanicoe, lama pacos, lama peruana	1
lobster	American, European, Langoustine, lobster tail, Maine, Northern, Norwegian, scampo	1
motorbikes	moped	1
pigeon	Columba livia, Columba palumbus, cushat, domestic, dove, Ectopistes migratorius, passenger, pouter, ringdove, rock, rock dove, squab, wood	0.8
pizza	anchovy, cheese, pepperoni, sausage, Sicilian	0
revolver	colt	0.9
rhino	Certotherium simum, diceros simus, Indian ceros, Indian rhinoceros, indianceros, Rhinocero antiquitatis, Rhinoceros unicornis, white rhinoceros, woolly rhinoceros	1
schooner	sharpshooter	1
scissors	slipper, shears, snuffers	0.9
seahorse	Atlantic walrus, Odobenus divergens, odobenus rosmarus, pacific walrus	0.2
strawberry	beach, chilean, cultivated, Fragaria ananassa, Fragaria chiloensis, Fragaria vesca, Fragaria virginiana, garden, scarlet, virginia, wild, wood	0.7
sunflower	common, giant, girasol, Helianthis angustifolius, helianthus annuus, helianthus guganteus, helianthus laetiflorus, Indian potato, Jerusalem artichoke, mirasol, prairie, showy, swamp, tall	1
tick	ticktock, tictac, tocktact	0.9
watch	Analog, digital, hunter, hunting, pocket, pendulum, wrist, wristwatch	0.7
wheelchair	bath chair, motorized	1
wildcat	Cougar, European wildcat, eyra, felis bengalensis, flis chaus, flis concolor, felis ocreata, flis pardalis, felis serval, felis silvestris, felis tigrina, felis wiedi,felis yagouaroundi, jaguarondi, jaguarondi, jaguarondi cat, jungle+cat, kaffir cat, kaffircat, leopard cat, cat+margay, margay cat, margay, ocelot, painter, panther, panther cat, puma, sand cat, serval, tiger cat	0.7
wrench	sprain	0.4

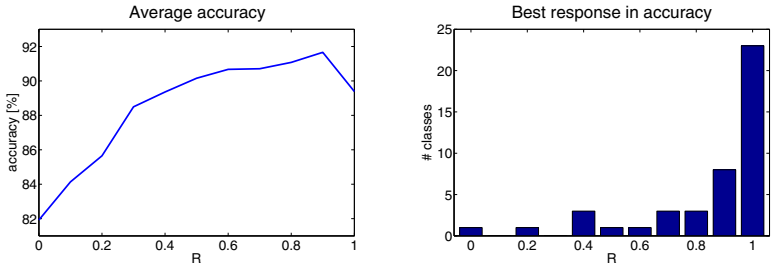

Fig. 2. Summary of results over 44 classes of the Caltech101 dataset. Left, the average detection accuracy over all the classes. Right, for each class the highest accuracy value (and the related value of R) has been detected and the histogram of best performing ratio R is shown.

this class (only 4) and the very good performances with the original keyword (accuracy about 98%).

6 Conclusions and Future Directions

In this paper we have tried to address the problem of automatic learning from Internet, in which autonomous agents have to learn the visual aspect of an object, starting simply from a single text identifier. For simplicity, we employ Google as Internet crawler.

The solution we have proposed in order to collect a more relevant training set of images is that of refining the Internet search by associating, to the name of the searched object, other words that are semantically related to it. In our experiments we have firstly collected images only by typing the name of the object, and we have evaluated the results in a detection scenario. Then, we have performed the same experiment by adding words that are related to the name of the object through WordNet, with an ontological check, at the moment performed manually, but that could possibly be automatized with the addition of an axiomatic ontology, obtaining definitely better results.

Obviously, the level of increase in precision strongly depends on the ontology to be built: a richly axiomatized ontology allows to express a lot more constraints, but can be more difficult to use for automatic reasoning tasks, as it may incur in computability problems.

Another thing to be taken seriously into consideration is to include, among the ontology properties, those strongly related to the visual appearance of objects, like their qualities, or their geometrical properties. Even for this reason, having ontologies that are well-founded (in this case on mereogeometries and mereotopologies, as well as on theories of qualities) can be of great help. A formal ontology that relies on a serious mereogeometrical analysis, should be able to express, for instance, the fact that a certain object is symmetrical if seen on the front, but asymmetrical if seen on the side, and so on.

This last remark is interesting not only with respect to the training phase, for the selection of images in the training set, but also to the exploration phase, in which the robot may make inferences on what it is actually seeing.

Acknowledgements. F. Setti and R. Ferrario are supported by the VisCoSo project grant, financed by the Autonomous Province of Trento through the "Team 2011" funding programme. D.S. Cheng is supported by the Hankuk University of Foreign Studies Research Fund of 2013.

References

1. Bunescu, R., Mooney, R.: Multiple instance learning for sparse positive bags. In: ICML 2007, pp. 105–112. ACM, New York (2007)
2. Fei-Fei, L., Fergus, R., Perona, P.: Learning generative visual models from few training examples: an incremental bayesian approach tested on 101 object categories. Comput. Vis. Image Underst. 106(1), 59–70 (2007)
3. Fellbaum, C.: Wordnet and wordnets. In: Brown, K. (ed.) Encyclopedia of Language and Linguistics, pp. 665–670. Elsevier, Oxford (2005)
4. Helmer, S., Meger, D., Viswanathan, P., McCann, S., Dockrey, M., Fazli, P., Southey, T., Muja, M., Joya, M., Little, J., Lowe, D., Mackworth, A.: Semantic robot vision challenge: Current state and future directions. CoRR (2009)
5. Lopes, L.S., Chauhan, A.: How many words can my robot learn? an approach and experiments with one-class learning. Interaction Studies 8, 53–81 (2007)
6. Masolo, C., Borgo, S., Gangemi, A., Guarino, N., Oltramari, A.: Wonderweb deliverable d18. Tech. rep., CNR (2003)
7. Meger, D., Muja, M., Helmer, S., Gupta, A., Gamroth, C., Hoffman, T., Baumann, M., Southey, T., Fazli, P., Wohlkinger, W., Viswanathan, P., Little, J., Lowe, D., Orwell, J.: Curious george: An integrated visual search platform. In: CRV, pp. 107–114 (2010)
8. Popescu, A., Millet, C., Moëllic, P.A.: Ontology driven content based image retrieval. In: CIVR, pp. 387–394 (2007)
9. Prévot, L., Borgo, S., Oltramari, A.: Ontology and the lexicon: a multi-disciplinary perspective (introduction). Studies in Natural Language Processing, pp. 185–200. Cambridge University Press (April 2010)
10. Roy, D., Pentland, A.: Learning words from sights and sounds: A computational model (1999), `http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.34.9295`
11. Schroff, F., Criminisi, A., Zisserman, A.: Harvesting image databases from the web. In: ICCV (2007)
12. Setz, A.T., Snoek, C.G.M.: Can social tagged images aid concept-based video search? In: ICME (2009)
13. Steels, L., Kaplan, F.: Aibo's first words: The social learning of language and meaning. Evol. of Communication 4(1), 3–32 (2001)
14. Ulges, A., Schulze, C., Koch, M., Breuel, T.: Learning automatic concept detectors from online video. Comput. Vis. Image Underst. 114(4), 429–438 (2010)
15. Vijayanarasimhan, S., Grauman, K.: Keywords to visual categories: Multiple-instance learning for weakly supervised object categorization. In: CVPR (2008)
16. Yeh, T., Darrell, T.: Dynamic visual category learning. In: CVPR (2008)