Exploiting Context Information for Image Description

Andrea Apicella, Anna Corazza, Francesco Isgrò(✉), and Giuseppe Vettigli

Università di Napoli Federico II, Napoli, Italy
francesco.isgro@unina.it

Abstract. Integrating ontological knowledge is a promising research direction to improve automatic image description. In particular, when probabilistic ontologies are available, the corresponding probabilities could be combined with the probabilities produced by a multi-class classifier applied to different parts in an image. This combination not only provides the relations existing between the different segments, but can also improve the classification accuracy. In fact, the context often gives cues suggesting the correct class of the segment. This paper discusses a possible implementation of this integration, and the first experimental results shows its effectiveness when the classifier accuracy is relatively low. For the assessment of the performance we constructed a simulated classifier which allows the a priori decision of its performance with a sufficient precision.

1 Introduction

This paper tackles the problem of recognising the content of a digital image, and being able to produce a schematic textual description. Because of the large number of images available on-line, this is a very hot research topic at the moment, as shown by the references in Sect. 2, and well performing systems using deep learning producing description in natural language have been proposed. In this work we start considering a new way to exploit context information to improve performance of classification based approaches.

When the aim is to design and implement a framework for the recognition of some of the components of a natural image, simply applying classification is not a solution as natural images classifiers only based on information extracted from the images, can be, in the most general case, error prone. The framework presented in this work aims at integrating the output of standard classifiers on different image parts with some domain knowledge, encoded in a probabilistic ontology. In fact, while standard ontologies are quite widespread as a means to manage a-priori information, they fail in the important task of dealing with real world uncertainty. Probabilistic ontologies aim at filling this gap by associating probabilities to the coded information, and provide then an adequate solution to the issue of coding the context information necessary to correctly understand the content of an image. Such information is then combined with the classifier output in order to correct possible classification errors on the basis of surrounding objects.

S. Battiato et al. (Eds.): ICIAP 2017, Part I, LNCS 10484, pp. 320–331, 2017.
https://doi.org/10.1007/978-3-319-68560-1_29

In conclusion, our aim is to improve the performance of a natural images classifier introducing in the loop knowledge coming from the real world, expressed in terms of probability of a set of spatial relations between the objects in the images. Not only the probabilistic ontology can be made available for the considered domain: it could also be built or enriched by using entities and relations extracted from a document related to the image. For example, the picture could have been extracted from a technical report or a book, where the text gives information which are related to the considered images. We wish to stress the fact that we are not thinking of a text directly commenting or describing the image, but of a text which is completed and illustrated by the image. In this case, both the classes of objects which can appear in the image and the relations connecting them could be mentioned in the text and could therefore be automatically extracted [2]. A probability can then be associated to them on the basis of the reliability of the extraction or the frequency of the item in the text.

The system we are considering, the logical scheme of which is depicted in Fig. 1, and better detailed in Sect. 3, aims at determining a set of keywords describing the content of an image and the relations existing among them. The idea is to design a system that, starting from an image, will first hypothesize the presence of some objects in the scene through a battery of image based classifiers. Considering for example the image of a building close to a water poll with some boats, it is likely that a classifier might label the reflection of the building on the water beneath the boats as a building, that is a wrong classification. We advocate that such a mis-classification can be corrected introducing the spatial relation between the boat-segment and reflected building, and the external knowledge that an image segment beneath a boat and surrounded by water is more likely to be water than a building. This world knowledge, that we plan to formalise in a probabilistic ontology [9], together with the output of the classifier, will be fed to a probabilistic model [4], in order to improve the performance of the single classifiers.

The classes associated to each segment combined by the spatial relations which can be directly extracted from an analysis of the image are eventually organized in a schematic description of its content. Relations could be further specialized by better specifing the reciprocal position of the segments. For example, the fact that a segment is in the middle, or in the upper right part of the picture, and so on.

The framework presents two main aspects of novelty. First, the use of a probabilistic ontology for a computer vision problem has, at the best of our knowledge, never been proposed before. A second element of novelty is the integration of a probabilistic model with a probabilistic ontology. A preliminary description of the general idea of the approach has been sketched in [1] in a very concise way. Here, we discuss all details and a first preliminary experimental assessment.

In the following section, we discuss related work. Section 3 is devoted to the description of the different modules of the system, with a few details about the probabilistic ontology (Sect. 3.1), and to the model adopted to combine classification and ontology probabilities (Sect. 3.2). Experimental assessment is consid-

ered in Sect. 4. Some conclusions and proposals for extensions of the presented work conclude the paper.

2 Related Work

Human beings express their knowledge and communicate using natural language, and in fact they find usually easy to describe the content of images with simple and concise sentences. Because of this human skill it is not difficult for a human user, when using an image search engine, to formulate a query by means of natural language.

Due to the large amount of images available on the web, for answering to textual image queries, it will be very helpful being able to automatically describe the content of an image. However such a task is not easy at all for a machine, as it requires a visual understanding of the scene, that is almost each object in the image must be recognised, how the objects relate to each other in the scene, and in what they are involved must be understood [27]. This task is tackled in two different ways. The most classical one [10,12,13,17] tries to solve the single sub-problems separately and combines the solutions to obtain a description of an image. A different approach [6,15,27] proposes a framework that incorporates all the sub-problems in a single joint model. A method trying to merge the two main approaches has been proposed recently in [30] using a semantic attention model. The problem is, however, very far from being solved.

In the context of textual image queries, it can be enough to extract from the images a less complex description (image annotation [28]), such as a list of entities represented in the image, and information about their position and mutual spatial relation in the image. The work proposed in this document addresses this task, that is also, as mentioned above, a necessary sub-task of the more general problem of generating a description in natural language.

The use of ontologies in the context of image recognition is not new [25]. For instance, in [20] it is proposed a framework for an ontology based image retrieval for natural images, where a domain ontology was developed to model qualitative semantic image descriptions. An ontology of spatial relations, in order to guide image interpretation and the recognition of the structures it contains was proposed in [14]. In [18], low-level features describing the color, position, size and shape of segmented regions are extracted and automatically mapped to descriptors forming a simple vocabulary termed object ontology. At the best of our knowledge, a probabilistic ontology has never been used for the task of image recognition and annotation.

Contextual information have been used in image recognition for long time [19,26], and it has been already shown [3] that the use of spatial relations can decrease the response time and error rate, and that the presence of objects that have a unique interpretation improves the identification of ambiguous objects in the scene. Just to mention a few application domains, contextual information has been used for face recognition [24], medical image analysis [5], analysis of group activity [7].

In the same way the use of probabilistic models is not new in computer vision, in particular a probabilistic model combining the statistics of local appearance and position of objects was proposed already in [22] for the task of face recognition, and in [21] in an image retrieval task, showing that adding a probabilistic model in the loop would improve the recognition rate. In [32] it is proposed a probabilistic semantic model in which the visual features and the textual words are connected via a hidden layer. More recently in the context of 3D object recognition, a system that builds a probabilistic model for each object based on the distribution of its views was proposed in [29]. In [31] a weakly supervised segmentation model learning the semantic associations between sets of spatially neighbouring pixels, that is the probability of these sets to share the same semantic label. Finally [11], in the context of action recognition, presents a generative model that allows for characterising joint distributions of regions of interest, local image features, and human actions.

3 System Architecture

The proposed framework, depicted in Fig. 1, is a chain of several logical modules, each corresponding to an element of a computational pipeline. The first step is a classifier, or a set of classifiers, detecting a predefined set of interesting objects in the image, identifying then a set of segments of interest in the image.

The hypotheses formulated for each segment in the image by a statistical classifier are then fed to a probabilistic model, that has been trained off-line. The task of this module is to validate, or correct, the hypothesis formulated in the previous step, integrating the output of the classifier with the world knowledge given by a probabilistic ontology, and expressed in terms of probability of a spatial relationship between instances of two classes of image objects. The class associated to each segment, together with the relations existing between segment pairs, constitute the image description output by the system.

3.1 Probabilistic Ontology

This section discusses the construction of a fragment of Probabilistic Ontology (PO) providing the information needed by our system. We need such fragment for the experimental assessment.

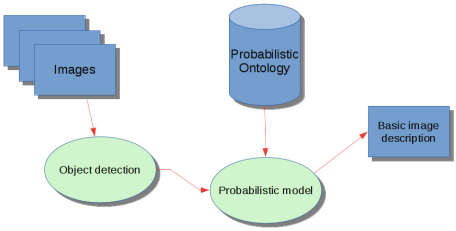

Fig. 1. Scheme of the proposed framework.

Table 1. Data set statistics.

Class	# of items
Sink	371
Chair	3,604
Table	558
Computer/monitor	256+417=673
Bed	407
Flower	1,822
Total	7,435

The main drawback of ontologies when facing real world problems is related to their inability to cope with uncertain information. Due to this, in the last years much work has been devoted to the design of effective tools to attach probabilities to the information contained in ontologies, among whose the most important is probably PrOWL [8]. From the so obtained POs, it is therefore possible to obtain a priori knowledge for applications effective also in complex contexts.

As a consequence, the research area concerning POs is very active and we expect that a number of POs in different domains will be available soon. However, we need a PO in the domain of the image data set we will adopt to assess the system performance, before we can start experimentation. We therefore design and implement an ontology to use in the experiments. In particular, the schema of the ontology will contain the classes to associate to the segments and the spatial relations among them considered in our analysis. On the other hand, probabilities are estimated from the training set after segments are automatically classified and spatial relations are constructed between segment pairs. In particular, we estimate the probability that two classes are in a given relation by the frequency of such event in the data set. No smoothing have been applied. More precisely denoting with D a set of segments used to compute the probabilities, with $R = \{r_1, \ldots, r_i\}$ the set of types of relation, with C the set of segments classes, we compute the probability that $c_1 \in C$ is in relation $r \in R$ with $c_2 \in C$ as:

$$\Pr(r, c_1, c_2) = \frac{D_r(c_1, c_2)}{\sum_{c_x \in C, c_y \in C} D_r(c_x, c_y)} \tag{1}$$

where $D_r(c_x, c_y)$ is the number of times that pairs of segments in D of classes respectively c_1 and c_2 satisfy the relation r. In general, as the relations are not necessarily symmetric, we have $\Pr(r, c_1, c_2) \neq \Pr(r, c_2, c_1)$.

Since there are no tools for directly constructing a PO, we use Protégé [1] for the construction of the schema of the ontology, while we use Pronto [16] as a reasoner for POs, as it adopts the standard OWL 1.1. The import of the schema developed by Protégé into Pronto is performed by editing the corresponding XML files and adding the probabilities. An example is given in Fig. 2, where the element tagged `pronto:certainty` is added to the axiom prepared by Protégé.

```
<owl11:Axiom>
    <rdf:subject rdf:resource="URI#x"/>
    <rdf:predicate rdf:resource="&rdfs;subClassOf"/>
    <rdf:object rdf:resource="URI#y"/>
    <pronto:certainty>0.070990;0.070990</pronto:certainty>
</owl11:Axiom>
```

Fig. 2. Piece of the XML of the PO corresponding to an axiom with an associated probability.

[1] Freely available from http://protege.stanford.edu/.

Although Pronto accepts probability ranges, as we use simple values, the two extremes of the interval coincides (0.070990; 0.070990 in the example).

3.2 Combination Models

This section investigates which model to use to integrate the classifiers and the ontological knowledge.

In the task we are considering the role of POs requires providing probabilities describing the domain of interest, to be integrated with the ones associated by the classifier to each class for each input segment. The main goal of our system is the classification of the segments in the input image. We aim to exploit the relations between pairs of segments to improve this classification. More formally, every image contains a set of segments S and there are a number of possible relations R connecting segment pairs.

For each segment in the image, the classifier associates a probability distribution to the set of all possible classes C. When we consider only the classification step, we classify the segment with the most probable class: this represents our baseline, as it only considers the classifier output, without any information coming from the PO. However, we can see the output of the classifier for each segment s in the image as a random variable $c(s)$ with values in C. In the following we discuss how such random variable is integrated with the ontological probabilities.

In fact, the ontology produces, for every pair of classes $c_1, c_2 \in C$ and every possible relation $r \in R$, the probability $\Pr(r, c_1, c_2)$ that in the real world two segments of classes c_1 and c_2 respectively are in relation r: its expression is given in Eq. 1. By integrating this information with the probabilities computed by the classifier, the classification performance could improve. Moreover, the solution output by this integration is likely to be consistent with the ontological knowledge, which can be an important feature in systems where the post-processing requires a set of properties on the considered candidates. In fact, whenever a relation can not hold between two classes, the corresponding ontological probability is null, and this also lowers the probability of the corresponding couple of classes.

We associate the following log-linear probability to the two classes associated to each context $x = (s_1, s_2, r : r(s_1, s_2))$ built around the relation type r connecting segments s_1 and s_2:

$$\Pr(c_1, c_2|x) = \frac{e^{v_{c_1} f_C(s_1,c_1)+v_{c_2} f_C(s_2,c_2)+v_{r,c_1,c_2} f_{PO}(r(s_1,s_2),c_1,c_2)}}{Z_{x,c_1,c_2}} \tag{2}$$

where $f_C(s, c) = \Pr(c(s) = c)$ and $f_{PO}(r, c_1, c_2) = \Pr(r(c_1, c_2))$, while Z_{x,c_1,c_2} is a normalisation factor depending on x and on the classes assigned to the two segments. Note that the features $f_C(\cdot)$ are produced by the classifier, while $f_{PO}(\cdot)$ depends on the probabilistic ontology. In conclusion, we consider two families of parameters: *class parameters* v_c for each class c and *relation parameters* v_{r,c_1,c_2} for each type of relation r and pair of classes (c_1, c_2). All in all, there are $|C|$ class parameters and $|R||C|^2$ *relation parameters.*

The parameters are estimated during the training, which maximises the likelihood of the training set. For this optimisation, we use the *Toolkit for Advanced Optimisation* (TAO) library, which implements a variety of optimisation algorithms for several classes of problems (unconstrained, bound-constrained, and PDE-constrained minimisation, nonlinear least-squares, and complementarity). In our work we focus on unconstrained minimisation methods which are used to minimise a function of many variables without any constraints on the variables. The method that we have used is *Limited Memory Variable Metric*, it is a *quasi-Newton* optimisation solver and it solves the Newton step using an approximation factor which is composed using the *BFGS* update formula.

Once we have estimated all the parameters $V = \{v_c, v_{r,c_i,c_j}\}$ with $c, c_i, c_j \in C$ and $r \in R$, we aim to assign the correct class to each segment in the input image. To do so, we consider two different models: in the former, to which we refer as M1, we assign to the classes in a given context a score which is equal to the $\Pr(c_1, c_2|x)$ as given by Eq. 2, while in the latter, M2, the score is given by its logarithm. In fact, when adopting, as in our case, a log-linear expression, only considering exponents is much more efficient than directly summing probabilities. We therefore obtain the following expressions for the scores sc_1 and sc_2 respectively corresponding to $M1$ and $M2$.

$$sc_1(c_1, c_2|x) = \Pr(c_1, c_2|x) = \frac{e^{v_{c_1} f_C(s_1,c_1) + v_{c_2} f_C(s_2,c_2) + v_{r,c_1,c_2} f_{PO}(r(s_1,s_2),c_1,c_2)}}{Z_{x,c_1,c_2}}$$

$$\begin{aligned} sc_2(c_1, c_2|x) = \log \Pr(c_1, c_2|x) = v_{c_1} f_C(s_1, c_1) + v_{c_2} f_C(s_2, c_2) \\ + v_{r,c_1,c_2} f_{PO}(r(s_1, s_2), c_1, c_2) - \log Z_{x,c_1,c_2} \end{aligned} \quad (3)$$

For each context x, we then compute the score that a given class c is associated to one segment, by summing the scores that every class is associated to each segment and that the relation assumes any of all possible relation types. We then associate to the first segment the class which maximises such a score in all segment pairs including it:

$$SC(c|s) = \max_{s_2 : \exists r, r(s_1,s_2)} \sum_{c_2 \in C} \sum_{r \in R} sc(c, c_2|(s_1, s_2, r : r(s_1, s_2))). \quad (4)$$

In this expression, sc stays for sc_1 or sc_2 depending on the adopted model. Note that since all relation types we consider are symmetrical, for every context $x = (s_1, s_2, r : r(s_1, s_2))$ also the symmetrical one $x' = (s_2, s_1, r(s_2, s_1))$ is defined, and therefore we can express the score as considering only the first of the two cases. However, when asymmetrical relations are also considered, the expressions can be easily generalised.

Finally, we assign to each segment the class which maximises the score of the class given the segment:

$$c^*(s) = \arg\max_{c \in C} SC(c|s) \quad (5)$$

To complete the textual description, the relations existing between segment pairs and used for determining the contexts defined above are added.

4 Experimental Assessment

This section describes and discusses the quantitative assessment of the performance of the proposed approach.

For this first experimental assessment of the combination model proposed we chose a data-set where interesting objects have been manually segmented and labelled, so to have a reliable ground-truth for estimating the performance of our model. The data set chosen is the *MIT-Indoor* including $1,700$ manually segmented images These pictures are taken in indoor surroundings, including kitchens, bedrooms, libraries, gymns and so on. Whenever an actual system based on the proposed approach is implemented, the best available solution for the segmentation will be included. We randomly divided the data in three parts: two of them, containing each the 30% of the data, are used to train the PO and the combination model respectively, while the remaining 40% of the data are used to assess the system performance. Note that in our view it is important that the data used to train the PO and the combination models are different, as in actual domains they usually have different origins.

The system performance is evaluated in terms of classification accuracy, i.e. the rate of segments which have been correctly classified. In particular, we considered six classes obtained by clustering the data set ones and then taking the six with a larger number of items: the adopted classes and the number of times they occur in the data set are reported in Table 1. Furthermore, we considered three relation types corresponding to the relative position of two segments in an image: *near*, *very near* and *intersecting*. Clearly, all three the relations are symmetrical.

The role of the classifier in our system is to produce a probability distribution on the set of classes for every input segment. The literature on object recognition is very rich [23]. The risk in choosing one approach or the other is that the final results would depend on this choice and its influence can not be distinguished by the one of the combination model. We therefore decided to substitute the actual classification with a random simulation able to produce any given performance. In this way, it is possible to describe the dependence of the system performance on the classification accuracy. All in all, we therefore need a method to simulate the behaviour of a multi-class classifier with an assigned accuracy a.

For this goal, we use the strategy described by the pseudo-code in Fig. 3. Given a segment, we randomly choose a score in $[0, 1]$ by the function $U(0, 1)$ for each class in the class set C. We then assign, with a probability given by the desired accuracy a, the maximum score to the gold class, while the other scores are randomly assigned to the remaining classes. The scores are finally normalised to obtain a probability distribution. As the classifier assigns to each segment the maximum probability class, we have that this corresponds to the right choice in the a percentage of cases, resulting in the desired accuracy. The use of a simulated classifier is not novel (see, for instance, [33]).

As we aim to assess the improvement we can obtain by introducing the ontological knowledge, we compare the system performance with a baseline consisting

```
maxClassProb←0.0;
BestClass← ∅;
for CurrentClass ∈ ClassSet do
    NewClassProb ∼ U(0, 1);
    ClassProb[CurrentClass]← newClassProb;
    if ClassProb[CurrentClass] > MaxClassProb then
        MaxClassProbValue← ClassProb[CurrentClass];
        BestClass ← CurrentClass;
    else
        if TossingACoin == Head then
            RandomClass ← CurrentClass;
        end if
    end if
end for
Accuracy ∼ U(0, 1);
Gold ← GoldClass(Segment);
if Accuracy < DesiredAccuracy then
    Swap(ClassProb[BestClass],ClassProb[Gold]);
else
    swap(ClassProb[RandomClass],ClassProb[Gold]);
end if
normalize(ClassProb);
```

Fig. 3. Pseudo-code of the simulated classifier.

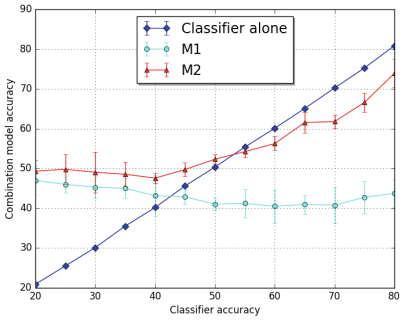

Fig. 4. Performance of the two systems compared with the baseline. Error bars give the 95% confidence intervals.

in the (simulated) classifier alone. The two approaches discussed in Sect. 3.2 are applied to combine the PO into the system: M1 and M2.

4.1 Results and Discussion

The system accuracy of the approaches proposed in this paper are depicted in Fig. 4 and compared with the accuracy of the statistical classifier applied alone.

For the sake of completeness, we considered a very wide range of accuracies for the simulated classifier: from 20% up to 80%, even if in actual conditions, the values of classifiers accuracy is more likely under $50-60\%$. However, in any case, we see that the M2 outperforms the M1, whose performance even deteriorates when the classifier accuracy improves. A possible explication for this behaviour could be that too much confidence is given to the a priori PO score with respect to the actual input data evidence.

On the other hand, the M2 improves on the simple classifier when the latter performance are inferior than about 55%, that is in realistic experimental conditions. We can observe how performance of this model are much better than the classifier alone when the latter performance are worse than 30%, and this can be the case when the task is not too easy. Even for classifiers obtaining an accuracy between 30% and 55%, the adoption of an approach integrating PO knowledge is advantageous.

Last, but not least, we observe that even when M2 performs worse than the classifier alone, its accuracy improves with the classifier accuracy, so that the two curves are approximately parallel. This could suggest that a better ontology design, resulting in a better PO, could help the system to overcome the performance obtained by the classifier alone.

5 Conclusions and Future Work

In this paper, we proposed and experimentally evaluated two different probabilistic models to integrate the probabilities derived from a probabilistic ontology with the ones produced by a statistical classifier. One of the two proved to perform in an acceptable way and could be used in an actual system.

For the sake of obtaining a clear view of the integration module performance, we tried to minimise the effect of the other modules. Therefore, we started from images which had been manually segmented and simulated a classifier in such a way that its accuracy could be controlled. As a future work, we plan to assess the performance of the proposed approach when coupled with state-of-the-art modules.

A fragment of a probabilistic ontology has been built by using three relations which could be automatically recognised in the input images, while the corresponding probabilities have been estimated from their frequencies. When more sophisticated ontologies will be available, containing information from large data sets, we expect the integration to give even better results.

Acknowledgments. We are grateful to M. Benerecetti e P. A. Bonatti for useful discussions about the most effective ways to represent knowledge. The research presented in this paper was partially supported by the national project CHIS - Cultural Heritage Information System and Perception, the national project Perception, Performativity and Cognitive Sciences (PRIN Bando 2015, 2015TM24JS).

References

1. Apicella, A., Corazza, A., Isgrò, F., Vettigli, G.: Integrating a priori probabilistic knowledge into classification for image description. In: Proceedings of the 26th IEEE WETICE Conference (2017)
2. Bach, N., Badaskar, S.: A review of relation extraction. Carnegie Mellon University, Language Technologies Institute (2007)
3. Bar, M., Ullman, S.: Spatial context in recognition. Perception **25**(3), 343–352 (1996)
4. Bishop, C.M.: Pattern Recognition and Machine Learning. Springer, New York (2006)
5. Bloch, I., Colliot, O., Camara, O., Graud, T.: Fusion of spatial relationships for guiding recognition, example of brain structure recognition in 3D MRI. Pattern Recognit. Lett. **26**(4), 449–457 (2005)
6. Chen, X., Zitnick, C.L.: Mind's eye: a recurrent visual representation for image caption generation. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 2422–2431, June 2015
7. Choi, W., Shahid, K., Savarese, S.: Learning context for collective activity recognition. CVPR **2011**, 3273–3280 (2011)
8. Da Costa, P.C.G.: Bayesian semantics for the semantic web. Ph.D. thesis, George Mason University, Fairfax, VA, USA, aAI3179141 (2005)
9. Ding, Z., Peng, Y.: A probabilistic extension to ontology language owl. In: Proceedings of the 37th Annual Hawaii International Conference on System Sciences (HICSS 2004), Track 4, vol. 4, p. 40111.1 (2004)

10. Elliott, D., Keller, F.: Image description using visual dependency representations. EMNLP **13**, 1292–1302 (2013)
11. Eweiwi, A., Cheema, M.S., Bauckhage, C.: Action recognition in still images by learning spatial interest regions from videos. Pattern Recognit. Lett. **51**, 8–15 (2015)
12. Fang, H., Gupta, S., Iandola, F., Srivastava, R.K., Deng, L., Dollar, P., Gao, J., He, X., Mitchell, M., Platt, J.C., Lawrence Zitnick, C., Zweig, G.: From captions to visual concepts and back. In: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1473–1482, June 2015
13. Farhadi, A., Hejrati, M., Sadeghi, M.A., Young, P., Rashtchian, C., Hockenmaier, J., Forsyth, D.: Every picture tells a story: generating sentences from images. In: Daniilidis, K., Maragos, P., Paragios, N. (eds.) ECCV 2010. LNCS, vol. 6314, pp. 15–29. Springer, Heidelberg (2010). doi:10.1007/978-3-642-15561-1_2
14. Hudelot, C., Atif, J., Bloch, I.: Fuzzy spatial relation ontology for image interpretation. Fuzzy Sets Syst. **159**(15), 1929–1951 (2008)
15. Karpathy, A., Fei-Fei, L.: Deep visual-semantic alignments for generating image descriptions. IEEE Trans. Pattern Anal. Mach. Intell. **39**(4), 664–676 (2017)
16. Bobillo, F., Costa, P.C.G., d'Amato, C., Fanizzi, N., Laskey, K.B., Laskey, K.J., Lukasiewicz, T., Nickles, M., Pool, M. (eds.): Uncertainty Reasoning for the Semantic Web II: International Workshops URSW 2008-2010 Held at ISWC and UniDL 2010 Held at Floc, Revised Selected Papers. Springer, Heidelberg (2013)
17. Kulkarni, G., Premraj, V., Dhar, S., Li, S., Choi, Y., Berg, A.C., Berg, T.L.: Baby talk: understanding and generating simple image descriptions. In: Proceedings of the 2011 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2011, pp. 1601–1608 (2011)
18. Mezaris, V., Kompatsiaris, I., Strintzis, M.G.: An ontology approach to object-based image retrieval. In: Proceedings of 2003 International Conference on Image Processing, ICIP 2003, vol. 2, pp. 511–514, September 2003
19. Oliva, A., Torralba, A.: The role of context in object recognition. Trends Cogn. Sci. **11**(12), 520–527 (2007)
20. Sarwar, S., Qayyum, Z.U., Majeed, S.: Ontology based image retrieval framework using qualitative semantic image descriptions. Proced. Comput. Sci. **22**, 285–294 (2013)
21. Schmid, C.: A structured probabilistic model for recognition. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol. 2, p. 490 (1999)
22. Schneiderman, H., Kanade, T.: Probabilistic modeling of local appearance and spatial relationships for object recognition. In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, p. 45. IEEE Computer Society (1998)
23. Szeliski, R.: Computer Vision: Algorithms and Applications, 1st edn. Springer, New York Inc. (2010)
24. Tanaka, J.W., Sengco, J.A.: Features and their configuration in face recognition. Mem. Cogn. **25**(5), 583–592 (1997)
25. Tousch, A.M., Herbin, S., Audibert, J.Y.: Semantic hierarchies for image annotation: a survey. Pattern Recogn. **45**(1), 333–345 (2012)
26. Toussaint, G.: The use of context in pattern recognition. Pattern Recogn. **10**(3), 189–204 (1978)
27. Vinyals, O., Toshev, A., Bengio, S., Erhan, D.: Show and tell: a neural image caption generator. In: IEEE Conference on Computer Vision and Pattern Recognition, CVPR, pp. 3156–3164 (2015)

28. Wang, C., Blei, D.M., Fei-Fei, L.: Simultaneous image classification and annotation. In: CVPR. pp. 1903–1910. IEEE Computer Society (2009)
29. Wang, M., Gao, Y., Lu, K., Rui, Y.: View-based discriminative probabilistic modeling for 3D object retrieval and recognition. Trans. Image Proc. **22**(4), 1395–1407 (2013)
30. You, Q., Jin, H., Wang, Z., Fang, C., Luo, J.: Image captioning with semantic attention. In: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 4651–4659, June 2016
31. Zhang, L., Yang, Y., Gao, Y., Yu, Y., Wang, C., Li, X.: A probabilistic associative model for segmenting weakly supervised images. IEEE Trans. Image Proc. **23**(9), 4150–4159 (2014)
32. Zhang, R., Zhang, Z., Li, M., Ma, W.Y., Zhang, H.J.: A probabilistic semantic model for image annotation and multimodal image retrieval. In: Tenth IEEE International Conference on Computer Vision, ICCV 2005, vol. 1, pp. 846–851, October 2005
33. Zouari, H., Heutte, L., Lecourtier, Y.: Simulating classifier ensembles of fixed diversity for studying plurality voting performance. In: Proceedings of the 17th International Conference on Pattern Recognition, ICPR 2004, vol. 1, pp. 232–235, August 2004