Towards Ontologies for Image Interpretation and Annotation

Hichem Bannour and Céline Hudelot
Applied Mathematics and Systems Department, Ecole Centrale Paris
92 295 Châtenay-Malabry, France
{Hichem.bannour, Celine.hudelot}@ecp.fr

Abstract

Due to the well-known semantic gap problem, a wide number of approaches have been proposed during the last decade for automatic image annotation, i.e. the textual description of images. Since these approaches are still not sufficiently efficient, a new trend is to use semantic hierarchies of concepts or ontologies to improve the image annotation process. This paper presents an overview and an analysis of the use of semantic hierarchies and ontologies to provide a deeper image understanding and a better image annotation in order to furnish retrieval facilities to users.

1 Introduction

The rapid growth of multimedia content comes with the need to effectively manage this large amount of data, i.e. providing mechanisms for indexing and retrieval that can meet users expectations. Towards this goal, semantic image analysis and interpretation has been one of the most interesting challenges during this last decade, and several attempts have addressed the so well-known *semantic gap* problem [29]. Automatic image annotation was introduced in the early 2000s, and first efforts focused on statistical learning approaches as they provide powerful and effective tools to establish associations between visual features and semantic concepts [2, 14].

Nevertheless, current approaches even if they adequately describe the visual content of images, are often limited to detect only perceptual manifestations of semantics, and then are unable to model the images semantics as it is perceived by humans. They also have many limitations when dealing with broad content image databases [16], i.e. the obtained performance varies significantly according to the considered concept number and the targeted image data sets as well. This variability may be explained by the huge intra-concept variability and wide inter-concept similarities on their visual properties that often lead to conflicting annotations. Thus, it is clear that there is a lack of coincidence between the high-level semantic concepts and the low-level features, and that semantics is not exactly/always correlated with visual appearance.

Always in the quest for models that could help to map successfully low-level features into high-level semantic concepts, some approaches make use of "contextual knowledge" by building semantic hierarchies or integrating *a priori* knowledge to improve image analysis and interpretation. Indeed, objects in the real world are always seen embedded in a specific context, and the representation of that context is essential for the analysis and the understanding of images. Contextual knowledge for image interpretation may stem from multiple sources of information, including knowledge about the expected identity, size, position and relative depth of an object within a scene. For example, topological knowledge can provide information about which objects are most likely to appear within a specific visual setting, e.g. an office typically contains a desk, a phone, and a computer, but it is unlikely that it contains a bed. Spatial information can also provide information about which locations within a visual setting are most likely to contain objects, e.g. in a beach scene, the sky is usually placed at the top, while the sea is below. Given a specific context, this kind of knowledge can help reasoning on data to improve image annotation [11, 22].

Lavrenko & al. [14] was the first to make use of contextual information (perceptual context) in image annotation. They proposed a statistical generative model which looks at the probability of associating words with image regions. They used surrounding visual context by computing a joint probability of image features over different regions in an image using a training set and using this joint probability to annotate and retrieve images. Thus in this model, the association of different regions provides context while the association of words with image regions provides meaning. However, this method does not really capture the semantics of the image, but just uses statistics on the perceptual context to improve the image description.

Therefore, it is important to make use of explicit and formal methods to represent background knowledge. This will help taking into account general and specific context of the image, and allow reasoning to improve image interpretation. Background knowledge is considered as domain knowledge. [13] underlines that among the possible representations of domain knowledge, ontologies are the most useful, and have considerable advantages, as they provide a formal framework that may contain explicit semantic definitions, which can be directly processed by a machine, and allow at the same time to derive implicit knowledge by automatic inference.

This paper explores the role of ontologies and knowledge-based approaches in the modeling and understanding of images semantics and reviews some of the current approaches to analyze how they could help narrowing the semantic gap. In Section 2, we propose a definition for image semantics and we investigate the contribution of ontologies to discover and model this semantics. Section 3 sheds light on three categories of ontology-driven approaches for image annotation and reviews some interesting proposal on each of them. The fourth Section discusses the aforementioned approaches on their advantages/drawbacks. The paper is concluded in Section 5.

2 Image semantics

Semantics is the study of meaning. In language, it is also the study of the meaning or the interpretation of a word, sentence, or other language form. With respect to image retrieval, it can be defined as the meaning sought by the user in the image content. For example, a user would like to go beyond a query like "I seek an image that visually looks like this one" or "an image that contains a car", and would prefer to be able to ask query such as "find me an image that contains a spectacular jump of a killer whale" or "a figure which describes global warming" or "find me cars on a race track". Referring to the theory of meaning, image semantics can also be defined as the meaning sought by the user in the image content with a particular context. For example a same image can be interpreted as an image containing: "an airplane" or a "military aircraft" or the "F-34C carrier variant", depending on users background knowledge.

Therefore, it is clear that the information sought by users in these large databases is located at the semantic level, and then it is necessary to have at one's disposal effective means to extract, analyze, and provide a semantic description of image. In [16], the authors identified five categories of techniques that provide a semantic description of images. Among them, we focus in this paper on the use of explicit representations of background knowledge, i.e. Ontology-driven approaches. These approaches provide a powerful framework to model contextual knowledge and allow reasoning on this knowledge for image analysis and interpretation. Indeed, image semantics is context-sensitive, i.e. semantics is not an intrinsic property captured during the image acquisition process, but an emergent property of the interaction of the user and the image database [25]. Thus, modeling contextual knowledge is an essential task for image understanding and analysis. Contextual knowledge can help recognition of object within scenes by providing predictions about objects that are most likely to appear in a specific setting, i.e. topological information, along with the locations that are most likely to contain objects in the scene, i.e. spatial information. Ontologically-driven approaches are widely accepted now as very appropriate to model and take contexts into account. Thereby, unlike other techniques that allow a semantic description of images, ontologically-driven approaches are best suited for image retrieval systems as they model the semantics of images through relationships that help reasoning about it and understanding its meaning. Furthermore an appropriate ontology can make explicit the relationships between the labels and concepts. Other approaches just provide a latent correlation between the low-level features and the tags.

As well, semantic hierarchies (also known as hierarchies of concepts) are a particular case of ontologies as they are collections of classes ordered by the transitive closure of explicitly declared subclass or subtype relations. Being $\mathcal{A}$ a subclass of $\mathcal{B}$, captures the fact that the state and the behavior of the elements of $\mathcal{A}$ are coherent with the intended meaning of $\mathcal{B}$, while disregarding the additional features and functionalities that characterize the subclass [17]. Semantic hierarchies are being used in image retrieval as a framework for hierarchical image classification, and then to provide a multi-level image annotation.

3 Ontology-driven approaches for image annotation

Recent advances in the Semantic Web technology have motivated several works in the field of images retrieval. As a result, a considerable number of multimedia ontologies have been proposed [12, 27, 4], which attempt to define a standard for description of low-level multimedia content. Other domain ontologies have been proposed to allow semantic interpretation and reasoning over the extracted descriptions [1, 22, 23].

Ontologies are a formal, explicit specification of a shared conceptualization [8]. "Formal" reflects that ontology is machine-readable and allows reasoning about its content from the human and the machine side. "Explicit" means that the type of concepts used, and the constraints on their use are explicitly defined. "Shared" refers to the common knowledge embodied in ontology. "Conceptualization" refers to the model obtained by abstracting some phenomena existing in the real world by identifying the relevant concepts of those phenomena. Thus, ontologies allow capturing the relevant knowledge of a domain, provide a common understanding of this domain knowledge, determine acknowledged vocabulary of this domain, and give the explicit definition of the vocabulary(terms) and the relations between these vocabularies in formal models at different levels.

The uses of ontologies in the image retrieval field usually target the following goals:

1. A unified description of low level features: where ontologies are used to provide standard of description of low-level features - e.g. [4].

2. Visual description ontology: where ontologies are used to represent the different types of relations among image features such as edges, lines and region - e.g. [31]. Typically the use of ontologies comes during the image analysis process, and target to optimize or to argue on this task.

3. Knowledge description: ontologies are used to model the concepts (objects) and relations among them. Typically, these are all approaches that use reasoning on concepts or on contextual information, i.e. after the image analysis process. These approaches most often tackle the problem of image interpretation - e.g. [11].

4. Semantic mapping: ontologies are used to help the mapping between the visual level and the semantic level. As for example the use of semantic hierarchies to reduce the semantic gap, e.g. [7].

These uses generally contribute to improve the image retrieval by incorporating background knowledge. The integrated knowledge can help in various spots of the image retrieval process: going from image analysis, to annotation enrichment, passing through the mapping of visual features into semantic concepts, assigning a meaning to tags and disambiguation. However, since the major challenge with the semantic gap is to provide effective tools to help mapping between low-level features and semantic concepts, semantic mapping is almost getting all the lights and it has been one of the most active issue. In fact, while usual methods are just limited to providing a latent correlation between semantic and visual space, ontology-driven approaches can make explicit this relationship. And therefore they provide an effective way to map low-level features into semantic concepts by building rules that supply semantic association between features and concepts, while maintaining a semantic structure to this process and allowing reasoning to check the consistency of this mapping. Results in this area keep growing.

Explicit background knowledge has been introduced under three different shapes for annotating images: as heavy-weight ontologies, as semantic hierarchies (light-weight ontologies) and by the use of a formal languages: i.e. description logics. In the following we review the use of explicit background knowledge under these different forms, and we discuss their role to narrow the semantic gap by providing a good understanding and a more efficient image annotation.

3.1 Heavy-weight ontologies (HWO)

Heavy-weight ontologies are a fully described ontologies, including concept definitions and relations. Those ontologies make intensive use of axioms to model knowledge and restrict domain semantics. HWO, tailored to the image semantics understanding, have been used to attach meaning to the produced annotations and to help extracting, querying, analyzing and interpreting these annotations.

For example, in M-OntoMat-Annotizer [24], low level MPEG-7 visual descriptions are linked to conventional Semantic Web ontologies and annotations. M-OntoMat-Annotizer is used in order to construct ontologies that include prototypical instances of high-level domain concepts together with a formal specification of corresponding visual descriptors. Thus, it formalizes the interrelationship of high- and low-level multimedia concept descriptions allowing for new kinds of multimedia content analysis and reasoning.

A multi-ontology based multimedia annotation model is proposed in [6]. A domain independent multimedia ontology is integrated with multiple domain ontologies in an effort to provide multiple domain specific views of multimedia content. Thus, accessing multimedia content can less be subjective to users' background knowledge and their need of information. [9] suggested to use an ontology as an extra level in between the search query and keywords. So when performing a concept-based search, the search engine automatically performs inference to find all narrower concepts of the query concept. However, this method focuses on the query understanding and does not take into account image content. [19] proposed an image retrieval methodology where low-level features are extracted from image regions, and mapped automatically to intermediate level descriptors called "object ontology", which is used for the definition of high-level concepts. Nevertheless, the suggested "object ontology" is purely visual as it defines a simple vocabulary to describe perceptual manifestation of semantics (objects or regions).

It is therefore essential to provide tools that allow reasoning, not only on image content or on its textual description (concepts), but on both of them. This will help reducing the semantic gap and allow a deeper understanding of image content, and so an accurate image annotation.

3.2 Light-weight ontologies: semantic hierarchies

Light-weight ontologies are partially described ontologies, which typically use only "is-a" relationships. They are a subclass of HWO. Examples of light-weight ontologies are: taxonomies, thesauri and semantic hierarchies. Many approaches use ontologies as hierarchies of concepts to provide, a multi-level of abstraction, image annotation.

In [18], the authors proposed a semantic hierarchy classifier based on WordNet to integrate prior knowledge about inter-class relationships into the visual appearance learning. In the same spirit, [5] proposed "ImageNet", a large-scale ontology of images built upon the backbone of the WordNet structure. ImageNet aims at populating the majority of the 80,000 synsets of WordNet with an average of 500-1000 clean and full resolution images. Images are selected manually. As well, LSCOM [21] aims to design a taxonomy with a coverage of around 1 000 concepts for broadcast news video retrieval. The aforementioned approaches can be qualified as language-based hierarchy, as those hierarchies are built upon textual information extracted from Wordnet. While these hierarchies are useful to provide a meaningful structure (organization) for images, they ignore visual information which is an important part of image semantics. Notice that the objective from constructing these ontologies is to provide a reasoning framework to image classification. As classification is based on visual features, it is important to take visual information into account while building semantic hierarchies.

Acknowledging the above problem, visual hierarchies have been explored recently [28, 31]. Visual hierarchies are hierarchies constructed upon visual features. [31] proposed an image parsing to text description (I2T) framework, which generates text descriptions for images and videos. I2T is mainly based on an And-or Graph for visual knowledge representation. This graph provides a graphical model serving as prior knowledge to represent diverse visual patterns and to provide top-down hypotheses during the image parsing. However the building of this graph is handmade. [28] propose to group visual objects using a multi-layer hierarchy tree that is based on common visual elements. This is achieved by adapting, to the visual domain, the generative Hierarchical Latent Dirichlet Allocation (hLDA) model. Unluckily, visual hierarchies are hard to interpret, and seem to be not adequate for reasoning. Since the building of these hierarchies is based on visual features similarity, they just provide a visual taxonomy. So, it is clear that a more meaningful image hierarchy should be built upon both semantic and visual information.

[15] presented a method based on visual features and tags to automatically build a "semantivisual" image hierarchy. They constructed an image hierarchy that encodes a general-to-specific image relationship, and which is used as a knowledge ontology to perform image classification and annotation tasks more accurately. [7] proposed a hierarchical concept learning approach for image annotation, by incorporating concept ontology and multi-task learning to exploit the strong inter-concept correlations. Furthermore, they proposed an algorithm called "product of mixture-experts" in order to model the contextual relationships between image concepts and several patterns of the relevant salient objects, with which they co-appear.

Hierarchies of concepts have shown to be very useful to fill the semantic gap. However, most works has focused on the problem of hierarchies building, or used them as a hierarchical classification framework, but only few approaches have used them as a source of implicit knowledge. Nevertheless it should be more appropriate to use them as a source of explicit knowledge, to exploit the strong inter-classes (concepts) correlation and also the hierarchical relationships that can help reasoning about the adequate treatment or decision to take.

3.3 Formal semantics: description logics for image interpretation

While ontologies often play a passive taxonomic role, some approaches consider ontologies as an active inference framework for computer vision. These approaches focus on the use of formal semantics, i.e. explicit knowledge. They are based on inference and make use of background knowledge to semantic image analysis and/or interpretation, ensuring thus the acquisition of interpretations that match human cognition. To make inference, these approaches are based on a sets of objects (concepts) and relationships (roles) between them. Appropriate statements (axioms) capture the conditions that need to be met by the "reasonable" states (interpretations) of the domain. Image interpretation here can be formalized as: 1) deduction: where interpretation is an instantiation of formal knowledge consistent with evidence about the real-world domain [11, 10], or as 2) abduction: where interpretation is an instantiation of formal knowledge which allows to deduce the evidence [30, 23].

For example, spatial relations among objects and regions appear also crucial in the concept detection process. In [11], an ontology of spatial relations is proposed to facilitate image interpretation. [20] propose an approach based on a visual thesaurus and visual context to improve concept detection. The authors introduce local (topological and unified) context in the analysis, to refine the confidence values of regions before taking decision. The work presented by [30] proposes an iterative process where low-level detections (induction) are compared with high-level models to derive new hypotheses (deduction). These can in turn guide the search for evidence to confirm or reject the hypotheses on the basis of expectations defined over the lower level features. In [26], a knowledge-assisted analysis architecture is proposed to perform the refinement of an initial set of over-segmented regions. They also used a fuzzy reasoning engine for the extraction of additional implicit knowledge and the improvement of region-based classification by incorporating spatial relations and neighborhood information.

The use of formal logics appears to be very helpful since it allows refining the predictions on image analysis and interpretation by the use of reasoning on background knowledge. For more information about the use of Description Logics in image interpretation, see [22, 3].

4 Discussion

So far, we have seen that several works using ontologies have been proposed recently to improve the automatic annotation of images, and knowledge based approaches have been widely explored as they seem to be more efficient to model image semantics. These works have tackled several issues, including image analysis, image annotation, and image understanding. As aforementioned, we can distinguish that some approaches are using knowledge implicitly (i.e. without reasoning), as for example [14] who used a joint probability to model contextual information to improve image annotation, while other approaches are using knowledge explicitly to reason for either image analysis or the improvement of image interpretation.

Image annotation is a difficult task because of the uncertainty introduced by statistical learning algorithms, the problem of scaling, and depends on the accuracy of the ground truth of the learning dataset. It often consists in the prediction of the appropriate tags, given a set of learning images and a target image. The use of explicit knowledge can help model, reduce, or even remove this uncertainty by supplying formal frameworks to argue about the coherence of data and/or extracted information from images. Image interpretation is also a difficult task, since there is no explicit link between image features and semantics. Image interpretation is the assignment of meanings to images, thus to understand the semantics behind the image it is often required to consider contextual knowledge, as aforementioned. Towards this objective, explicit knowledge was modeled and used in the image annotation field under three different ways: as heavy-weight ontology, as a concepts hierarchy, or by the use of description logics.

Ontologies (and concepts hierarchies) have a considerable advantages as they provide a formal framework that may contain explicit semantic definitions which can be processed by a machine, and allow at the same time to derive implicit knowledge by automatic inference. However, the proposed approaches in the literature did not use all the reasoning ability provided by ontologies, and generally stayed limited to define standards for image descriptors or used them to reason about the annotations. The boldest attempts have been those providing methods for the automatic building of concept hierarchies.

Three types of hierarchies have recently been explored in computer vision: 1) language-based hierarchy, 2) visual hierarchy, 3) semantic hierarchy: based on both semantic and visual features. The use of semantic hierarchies, which are based on visual and semantic information, is more convenient as it cares about perceptual and conceptual semantics. Indeed conceptual semantics may not correspond to image semantics, and then having a "semantic-visual" hierarchy could help reasoning on both images and concepts.

However, building and using concept hierarchies for image analysis constrains the reasoning to the inheritance relationships, i.e. "is-a" relationship. It should therefore enrich the types of relationships used to reason about images, including composition relationships, spatial, topological, etc. to benefit from the strong reasoning power on contextual knowledge. Building "*semantic network*" for image analysis instead of semantic hierarchies, is maybe a good direction to narrow the semantic gap and to improve image semantics modeling.

On the other side, Description Logics based approaches have successfully managed to exploit some interesting reasoning properties in the context of high-level image interpretation. Given the high expressivity and well-defined inference services coming with them, Description Logics became widely adopted as formalism for conceptual modeling in formal image interpretation frameworks. Deductive and abductive reasoning were introduced as inference standards, where for deductive reasoning: if Σ is a logical theory and α a set of facts, through deduction is verified whether φ is logically entailed, that is whether $\Sigma, \alpha \models \varphi$. For abductive reasoning: given Σ and φ abduction consists in finding "explanations" α so that the entailment $\Sigma, \alpha \models \varphi$ is true. For example, let's say Σ is (visual or contextual) background knowledge regarding concept "car", α is (visual or contextual) information extracted from an image containing a "car" and φ is an instance of concept "car", deduction is then, given $\Sigma, \alpha \models \varphi$.

The final challenge in the field of image retrieval system with high-level semantics is maybe to bring these different communities together, to integrate different knowledge sources, as well as the use of inference to provide a multi-stage reasoning framework for image interpretation. Contributions of different knowledge sources can provide a good reasoning platform, by encoding different background knowledge, to image understanding. Using inference can lead to have a powerful formalism for conceptual modeling of image semantics.

5 Conclusion

This paper provides a comprehensive survey of recent works towards the use of semantic hierarchies and ontologies for the image annotation purpose. We discussed a number of techniques and approaches which tackled image annotation and interpretation in order to attempt to narrow the "semantic gap". The paper highlights the importance of reasoning and contextual knowledge in the image understanding process, emphasizes the limitations of current approaches and tries to provide solutions that can overcome these limitations.

Several approaches have been developed for image annotation based on ontologies and semantic hierarchies. Some others approaches used formal logics to provide a reasoning framework for image interpretation. Our survey showed that to implement a complete image retrieval system with high-level semantics, it is required to integrate different knowledge sources, as well the use of inference to provide a multi-stage reasoning framework for image interpretation.

References

[1] A. D. Bagdanov, M. Bertini, A. D. Bimbo, G. Serra, and C. Torniai. Semantic annotation and retrieval of video events using multimedia ontologies. In *ICSC*, pages 713–720, 2007.

[2] K. Barnard, P. Duygulu, D. Forsyth, N. de Freitas, D. M. Blei, and M. I. Jordan. Matching words and pictures. *JMLR*, 3:1107–1135, 2003.

[3] S. Dasiopoulou and I. Kompatsiaris. Trends and issues in description logics frameworks for image interpretation. In *Artificial Intelligence: Theories, Models and Applications*, volume 6040, pages 61–70. 2010.

[4] S. Dasiopoulou, V. Tzouvaras, I. Kompatsiaris, and M. G. Strintzis. Enquiring mpeg-7 based multimedia ontologies. *Multimedia Tools Appl.*, 46:331–370, January 2010.

[5] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei. Imagenet: A large-scale hierarchical image database. In *CVPR*, 2009.

[6] A. Dong and H. Li. Multi-ontology based multimedia annotation for domain-specific information retrieval. In *SUTC*, pages 158–165, 2006.

[7] J. Fan, Y. Gao, and H. Luo. Integrating concept ontology and multitask learning to achieve more effective classifier training for multilevel image annotation. *IEEE Trans. IP*, 17(3):407–426, 2008.

[8] T. R. Gruber. Toward principles for the design of ontologies used for knowledge sharing. *International Journal of Human-Computer Studies*, 43(5-6):907–928, 1995.

[9] J. S. Hare, P. A. S. Sinclair, P. H. Lewis, K. Martinez, P. G. Enser, and C. J. Sandom. Bridging the semantic gap in multimedia information retrieval: Top-down and bottom-up approaches. In *ESWC*, 2006.

[10] J. Hartz and B. Neumann. Learning a knowledge base of ontological concepts for high-level scene interpretation. In *ICMLA*, pages 436–443, 2007.

[11] C. Hudelot, J. Atif, and I. Bloch. Fuzzy spatial relation ontology for image interpretation. *Fuzzy Sets Systems*, 159(15):1929–1951, 2008.

[12] J. Hunter. Adding multimedia to the semantic web - building an mpeg-7 ontology. In *SWWS*, pages 261–281, 2001.

[13] Y. Kompatsiaris and P. Hobson. *Semantic Multimedia and Ontologies: Theory and Applications*. Springer, 2008.

[14] V. Lavrenko, R. Manmatha, and J. Jeon. A model for learning the semantics of pictures. In *NIPS*. MIT Press, 2003.

[15] L.-J. Li, C. Wang, Y. Lim, D. Blei, and L. Fei-Fei. Building and using a semantivisual image hierarchy. In *CVPR*, 2010.

[16] Y. Liu, D. Zhang, G. Lu, and W.-Y. Ma. A survey of content-based image retrieval with high-level semantics. *Pattern Recognition*, 40(1):262 – 282, 2007.

[17] F. Logozzo and A. Cortesi. Semantic hierarchy refactoring by abstract interpretation. In *VMCAI*, pages 313–331, 2006.

[18] M. Marszalek and C. Schmid. Semantic hierarchies for visual object recognition. In *CVPR*, 2007.

[19] V. Mezaris, I. Kompatsiaris, and M. G. Strintzis. An ontology approach to object-based image retrieval. In *ICIP*, pages 511–514, 2003.

[20] P. Mylonas, E. Spyrou, Y. Avrithis, and S. Kollias. Using visual context and region semantics for high-level concept detection. *IEEE Trans. MultiMedia*, 11(2):229–243, 2009.

[21] M. Naphade, J. R. Smith, J. Tesic, S.-F. Chang, W. Hsu, L. Kennedy, A. Hauptmann, and J. Curtis. Large-scale concept ontology for multimedia. *IEEE MultiMedia*, 13:86–91, July 2006.

[22] B. Neumann and R. Möller. On scene interpretation with description logics. *Image Vision Computing*, 26(1):82–101, 2008.

[23] S. E. Peraldi, A. Kaya, S. Melzer, R. Möller, and M. Wessel. Multimedia interpretation as abduction. In *DL*, 2007.

[24] K. Petridis, D. Anastasopoulos, C. Saathoff, Y. Kompatsiaris, and S. Staab. Montomat-annotizer: Image annotation linking ontologies and multimedia low-level features. In *KES*, 2006.

[25] S. Santini, A. Gupta, and R. Jain. Emergent semantics through interaction in image databases. *KDE*, 13(3):337–351, 2001.

[26] N. Simou, T. Athanasiadis, G. Stoilos, and S. Kollias. Image indexing and retrieval using expressive fuzzy description logics. *Signal, Image and Video Processing*, 2(4):321–335, 2008.

[27] N. Simou, V. Tzouvaras, Y. Avrithis, G. Stamou, and S. Kollias. A visual descriptor ontology for multimedia reasoning. In *WIAMIS*, 2005.

[28] J. Sivic, B. C. Russell, A. Zisserman, W. T. Freeman, and A. A. Efros. Unsupervised discovery of visual object class hierarchies. In *CVPR*, 2008.

[29] A. W. M. Smeulders, S. Member, M. Worring, S. Santini, A. Gupta, and R. Jain. Content-based image retrieval at the end of the early years. *IEEE PAMI*, 22:1349–1380, 2000.

[30] C. Town. Ontological inference for image and video analysis. *Machine Vision and Applications*, 17(2):94–115, 2006.

[31] B. Yao, X. Yang, L. Lin, M. W. Lee, and S. C. Zhu. I2t: Image parsing to text description. In *Proc. of IEEE*, 2009.