Ontology-Based Automatic Image Annotation Exploiting Generalized Qualitative Spatial Semantics

Christos V. Smailis, Dimitris K. Iakovidis

Dept. of Informatics & Computer Technology, Technol. Educational Institute of Lamia, Greece

Abstract. Ontologies provide a formal approach to knowledge representation suitable for digital content annotation. In the context of image annotation a variety of ontology-based tools has been proposed. Most of them enable manual annotation of the images with higher level concepts whereas many of them are capable of formally representing low-level features as well. However, they either consider specific, usually quantitative, representations of the low-level features, or spatial semantics limited to 2D/3D image spaces. In this paper we propose a novel ontology-based methodology for automatic image annotation that exploits generalized qualitative spatial relations between objects, given an image domain. To represent knowledge for the spatial arrangements, we have implemented an ontology that models spatial relations in multi-dimensional vector spaces. The application of the proposed methodology is demonstrated for automatic annotation of segmented objects in chest radiographs.

1. Introduction

Knowledge authoring in the image domain was traditionally realized by manual segmentation and association of image objects to textual tags, usually arbitrarily selected. Recently, image annotation techniques based on ontologies have been proposed, enabling formal, unambiguous semantic annotation and inference. A problem arises in linking high level semantics such as concepts that are expressed in text form, with low level features of images due to their perceptual nature. This is usually referred to as semantic gap. For this purpose several annotation tools utilize ontologies in order to establish links between MPEG-7 low level feature descriptors and semantics. For example the K-Space Annotation Tool (KAT) [1] implements an ontology-based framework for the semantic annotation of images. KAT's annotation framework is based on the Core Ontology of Multi-Media (COMM) [2]. COMM models the various annotation levels and their linking (e.g. of descriptive and structural annotations), while providing MPEG-7 based structural and media descriptions of formal semantics. Similarly, PhotoStuff [3] is an ontology-based image annotation tool that expresses spatial, temporal or spatioteporal de-composition information, two internal, ontologies are used that model the different multimedia content and segment types in accordance with the MPEG-7 specifications. This provides a simple schema for linking content instances with respective low-level descriptors. A similar annotation scheme is present in M-Ontomat-Annotizer. M-Ontomat-Annotizer enables the ontol-

adfa, p. 1, 2011.

ogy-based representation of associations between domain specific concepts and their respective low-level visual descriptors. In order to formalize the linking of domain concepts with visual descriptors, M-Ontomat-Annotizer [4] employs the Visual Annotation Ontology (VAO) and the Visual Descriptor Ontology (VDO) [5], both hidden to the user. A survey of the aforementioned tools can be found at [6].

Other studies suggest bridging the semantic gap by describing images through the spatial arrangement of the included objects. Hudelot et al. [7] introduced an ontology of fuzzy 2D/3D directional and topological spatial relations that focuses on the representation of image structural knowledge instead of features such as color and texture. In [8], we presented IROn, an ontology of medical image representations theoretically extending the approach of [7] from image spaces/volumes to multidimensional spaces. However this ontology being rather tied to the medical imaging domain, contains oversimplified concept definitions of spatial relations thus providing limited expressivity.

In this paper we propose a novel methodology for automatic image annotation, based on an ontology we implemented for this purpose that builds on the modeling approach introduced in [8]. This ontology generalizes the ontological representation of spatial relations provided by IROn to any imaging domain. It has enhanced semantic expressivity by being capable of representing qualitative spatial relations not only in 2D/3D spaces, but also in multidimensional vector spaces. Furthermore, the proposed methodology is implemented within our Ratsnake annotation tool [9], enabling it to automatically annotate images. The generalized ontology of spatial relations, the proposed methodology and its application for automatic object annotation are presented in the following paragraphs.

2. Generalized Ontological Representation of Spatial Relations

Our generalized ontological model of spatial relations between objects has been implemented using the web ontology language description logics (OWL DL), which is characterized for its compactness and expressivity of description logics. The modeling approach adopted takes into account the following considerations: a) Spatial relations have their own characteristics, but at the same time act as links between different objects; b) Spatial relations should be independent of vector space dimensionality.

According to this approach, spatial relations are modeled as concepts instead of properties (this reification has been also used in previous studies [7],[8]). To ensure independency from space dimensionality, the spatial relations can only be defined between 1D projections of a reference and a target object, across a certain axis. An axis may participate in the definition of one or more multidimensional spaces. Currently two types of spatial relations have been included in our ontology, namely directional and topological. Each spatial relation can also be linked to its inverse. Directional relations are categorized into positive and negative ones, whereas topological relations are divided into eight main categories that are based on the ones used by region connection calculus 8 (RCC 8). The rest of this section describes the concepts included in our ontology in detail:

● An Object refers to the set of objects that are associated through spatial relations between each other. In description logics syntax [10] this is expressed as:

$\text{Object} \sqsubseteq \top$

● In order to refer to the objects that are used as a reference in the spatial relations, the concept ReferenceObject has been defined:

$\text{ReferenceObject} \equiv \text{Object} \sqcap$
$\exists\ \text{reference.SpatialRelation} \sqcap \geq 1\ \text{reference}$

● TargetObject refers to the objects that are used as targets in Spatial Relations:

$\text{TargetObject} \equiv \text{Object} \sqcap$
$\exists\ \text{target.SpatialRelation} \sqcap \geq 1\ \text{target}$

The concepts ReferenceObject and TargetObject overlap each other and are subsumed by Object.
● The concept NumericValue, enables the representation of numbers as instances of this concept:

$\text{NumericValue} \sqsubseteq \top$

This is needed in order to represent distinct numeric values regardless of their actual value and to overcome the inability of OWL DL to express numeric datatype properties that can be used for reasoning tasks.
● The concept VectorSpace represents a multi-dimensional vector space. A vector space may be defined by many axes that can also belong to other vector spaces as well:

$\text{VectorSpace} \sqsubseteq (\exists\ \text{definedBy.Axis}) \sqcap (\forall\ \text{definedBy.Axis})$
$\sqcap (\geq 1\ \text{definedBy})$

● The Axis concept represents an axis that may define one or more vector spaces at the same time:

$\text{Axis} \sqsubseteq (\exists\ \text{defines.VectorSpace}) \sqcap (\forall\ \text{defines.VectorSpace}) \sqcap (\geq 1$
$\text{defines})$

● SpatialRelation refers to the set of spatial relations that are defined according to a reference object and a target object across an Axis:

$\text{SpatialRelation} \sqsubseteq (\exists\ \text{reference.Object}) \sqcap (\exists\ \text{target.Object}) \sqcap (\exists$
$\text{hasAxis.Axis}) \sqcap (\forall\ \text{reference.Object}) \sqcap (\forall\ \text{target.Object}) \sqcap (\forall$
$\text{hasAxis.Axis}) \sqcap (= 1\ \text{reference}) \sqcap (= 1\ \text{target}) \sqcap (= 1\ \text{hasAxis})$

● The SpatialRelation concept subsumes the concept DirectionalRelation that refers to the set of relations implying direction across an axis. A NumericValue indicating the number of intermediate objects (or their absence if this value represents zero) between the projections of two objects on this axis is required. This way one can uniquely describe the relative position of the target objects in a vector space using a reference object and multiple directional relations.

```
DirectionalRelation ⊑ SpatialRelation ⊓
(∃ numberOfIntermediateObjects.NumericValue) ⊓
(= 1 numberOfIntermediateObjects)
```

DirectionalRelation subsumes the following disjoint concepts: PositiveDirectionalRelation, NegativeDirectionalRelation.

● SpatialRelation also subsumes the concept TopologicalRelation which represents basic relations based on RCC 8. Each topological relation is defined along an axis.

```
TopologicalRelation ⊑ SpatialRelation
```

TopologicalRelation subsumes the following disjoint concepts: Equal, ExternalConnection, Non-TangentialProperPart, Non-TangentialProperPartInverse, PartialOverlap, TangentialProperPart, TangentialProperPartInverse, and Disconnected.

3. Ontology-based Automatic Image Annotation

Spatial relations are often more reliable descriptors than other object properties in images of static contexts [7]. For example, in chest radiographs the texture of a lung may vary depending on the subject's pathology, whereas its relative position with respect to the spinal cord will remain approximately the same. The methodology presented in this section exploits the semantic description of the spatial arrangement of objects to automatically annotate the objects, within an image or a sequence of images of the same domain, by ontological reasoning. This methodology has been implemented within our Ratsnake annotation tool and enables automatic image annotation within a specified image domain. The whole process is divided in two phases: a training and a labeling phase.

3.1 Training Phase

During the training phase, users must load an image in the annotation tool and annotate objects of interest in the image either manually or by using the semi-automatic image annotation framework proposed in [9]. Each object must be assigned a new textual label or a semantic concept from a domain ontology loaded in the annotation tool. After that, users must specify the image domain, by either submitting a new textual label or by providing a representative concept from one of the loaded ontologies. Then, a new ontology, using the spatial ontology presented in Section 2, is automatically generated to describe the knowledge of the spatial arrangement of objects in the specified domain. This ontology has two parts; a fixed part which holds fundamental concepts regarding the image domain and the segmented image objects, and a dynamically generated part which holds the spatial relations between these objects.

The fixed part of the concept hierarchy in the automatically generated ontology consists of three main classes:

● CoreElements is the superset of all the other classes in the automatically generated ontology.
● Image, in the training phase, represents the training image from which the domain knowledge is extracted.

```
Image ⊑ CoreElements
```

● SpatialObject subsumes automatically generated concepts that represent the set of the manually annotated objects on the training image.

```
SpatialObject ⊑ CoreElements
```

● ImageDomain represents the image domain specified by the user. Each image domain comprises a set of annotation types that should be contained in images of that domain.

```
ImageDomain ⊑ CoreElements
```

In the dynamically generated part user specified image domains are asserted in the ontology as subclasses of the ImageDomain class. The types of annotated image objects are asserted as classes which inherit both the SpatialObject class as well as a subclass of image domain that represents a user specified domain. The instances of the segmented objects in the training image are asserted as individuals of the class that represents the annotation type.

In order to extract the spatial relations between the segmented objects we consider that each object is represented by its center of gravity (CoG). Of course alternative representations of objects could be considered as well. The 1D projection of every segmented image object is spatially related to the 1D projection of a reference object, across each axis of the 2D image plane, using individuals of the subclasses of the SpatialRelation class, defined in the proposed spatial ontology. These include the directional PositiveDirectionalRelation, NegativeDirectionalRelation that can be used to express orientation on an axis and the topological Equal that can be used to assert that the projections of the two objects are located at the same position on an axis. An arbitrarily selected object, that is common for all images of the domain, can be considered as a reference. For images of static context such as the chest radiographs, this reference can be defined as the image center. After this step, the classes that represent annotation types obtain certain restrictions based on the spatial arrangement of the objects. These restrictions define how each instance of a certain annotation type can be related to the reference of that domain thus making the classification of the segmented objects possible.

3.2 Labeling Phase

In the beginning of the labeling phase, images are loaded and segmented in the annotation tool. All segmented ROIs are initially unlabeled. Next, the domain of the images must be specified by the user. Once again spatial relations between the ROIs are extracted with the method used during the training phase. The individuals representing the unlabeled segmented objects are asserted as instances of SpatialObject.

In order to infer the class of the individuals that represent the segmented objects, the restrictions defined in these classes can be exploited by a reasoner, such as FACT++ or Pellet. When instance classification is completed by the reasoner the names of these classes can be assigned as labels to each segmented object. In the following section, both the training and the labeling phases are demonstrated.

4. Automatic Annotation of Objects in Chest Radiographs

We consider the use case of automatic annotation of objects in segmented chest radiographs. For the purposes of our study, we have considered chest radiographs from the publicly available Database of the Japanese Society of Radiological Technology (JSRT) [11]. For each of these images, the ground truth segmented areas from [12] are used. Each image consists of the following objects: heart, left lung, right lung, left clavicle and right clavicle as shown in Fig.2. During the training phase, a random image is selected and its contents are manually annotated by linking them to concepts of the Foundational Model of Anatomy (FMA) [13]. The automatically generated ontology is populated by creating a class for the domain of chest radiographs that subsumes all the classes that represent the types of the segmented objects.

Each class that represents an annotation type obtains restrictions that define how individuals of that class should be related in space to the center of the image across each axis of the 2D plane. For example the left lung is positioned higher in the Y-axis than the center of the image and is on the left of it across the X-axis as shown in Fig.1. Therefore the concept Left_Lung should have the following restrictions which are automatically generated during the training phase:

```
Left_lung≡∃ target ((∋numberOfIntermediateObjects
{Value-0}) ⊓ (∃reference ReferenceObject) ⊓ (∃ hasAxis {X-
Axis}) ⊓ NegativeDirectionalRelationship) ⊓ ∃ target ((∋number-
OfIntermediateObjects {Value-0}) ⊓ (∃reference ReferenceObject)
⊓ (∃hasAxis {Y-Axis}) ⊓ PositiveDirectionalRelationship)
```

The rest of the annotation type classes obtain similar restrictions. Thus, spatial knowledge for the domain is collected. An example of the produced class hierarchy is illustrated in Fig. 2.

During the labeling phase chest radiograph images are segmented while their domain is explicitly defined. Individuals of the type SpatialObject representing the instances of the unclassified segmented objects are then created. Each of them is affiliated to an individual representing the center of the image that belongs to the chest-radiograph domain using instances of the spatial relations defined in our spatial ontology. For example, an individual representing a positive directional relation between the center of an image and a left lung across the axis X is automatically asserted as:

```
NegativeDirectionalRelationX0_ImageCenter-Left_lung:
```

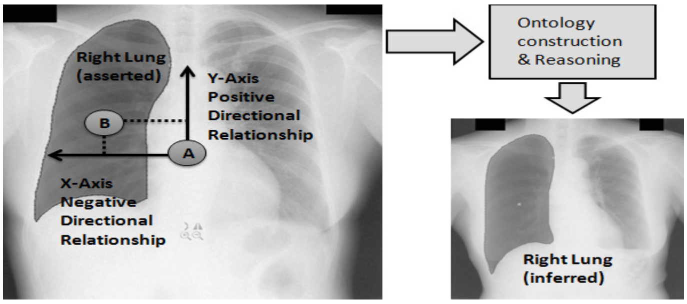

Fig. 1. In the training phase spatial knowledge is extracted from a representative training image. During the labeling phase it is used to infer the annotation types of the segmented objects.

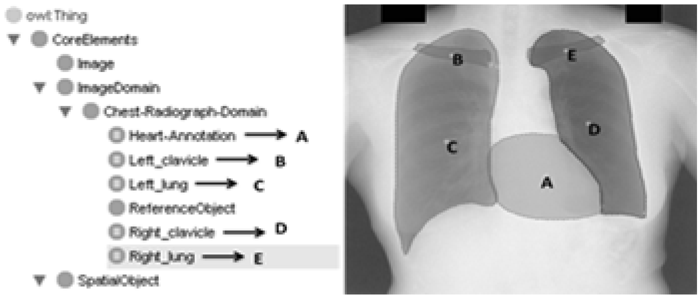

Fig. 2. An example class hierarchy of the automatically generated ontology. Each unlabeled segmented object is assigned to a certain annotation type during the labeling phase.

```
∃reference.{ImageCenter-Individual} ⊓ ∃target.{LeftLung-
Individual} ⊓ ∃hasAxis.{ X-Axis-Individual} ⊓ ∃numberOfInterme-
diateObjects.{Value-0}
```

After this step a reasoner can infer the type of each segmented object thus making possible the labeling of the unclassified objects in each image. In our experiment every segmented object in all of the images of JSRT were successfully labeled.

Conclusions

This paper made two contributions. First, it presented a generalized ontology that can describe spatial relations in multidimensional spaces. Then we proposed an automatic annotation methodology that uses the aforementioned ontology to automatically annotate objects that belong to images of the same domain. Advantages of the proposed methodology over conventional approaches include: a) It is non-parametric, b) It does not require any feature normalization since it utilizes relative object descriptions, c) It offers a general framework for bridging the semantic gap, that is particularly suitable for images of static context. Comparative advantages of the proposed approaches over the previous ones, such as M-Ontomat Annotizer [4], include exploitation of the spatial semantics to classify objects and capability to describe spatial relations within feature spaces. Future research directions include the application of the proposed annotation methodology for the automatic annotation of clusters of feature vectors in multidimensional spaces for data mining.

Acknowledgment

This work was realized in the framework of the project DebugIT, co-funded by the EC FP7 grant agreement n° FP7–217139. The EC, DG INFSO, is not liable for any use that may be made of the information in this document. It reflects solely the views of the authors.

References

1. Saathoff, C., Schenk, S., Scherp, A.: Kat: the k-space annotation tool. In: International Conference on Semantic and Digital Media Technologies, Germany (2008)
2. Arndt, R., Troncy, R., Staab, S., Hardman, L., Vacura, M.,: COMM: Designing a Well-Founded Multimedia Ontology for the Web. In 6th Int. Semantic Web Conf., Busan (2007)
3. Halaschek-Wiener, C., Golbeck, J., Schain, A., Grove, M., Parsia, B., Hendler, J.A.: PhotoStuff — An Image Annotation Tool for the Semantic Web. In: 4th International Semantic Web Conference Posters, Galway (2005)
4. Petridis, K., Anastasopoulos, D., Saathoff, C., Timmermann, N., Kompatsiaris, I., Staab, S.: M-OntoMat-Annotizer: Image Annotation. Linking Ontologies and Multimedia Low-Level Feature. In: 10th International Conference on Knowledge-Based & Intelligent Information & Engineering Systems, Bournemouth (2006)
5. Simou, N., Tzouvaras, V., Avrithis, Y., Stamou, G., Kollias, S: A visual descriptor ontology for multimedia reasoning. In: Workshop on Image Analysis for Multimedia Interactive Services, Montreux (2005)
6. Dasiopoulou, S., Giannakidou E., Litos G., Malasioti P., Kompatsiaris Y.: A Survey of Semantic Image and Video Annotation Tools. In G. Paliouras, C.D. Spyropoulos, G. Tsatsaronis (eds.). LNCS, Vol. 6050, pp. 196-239, Springer, Berlin (2011)
7. Hudelot, C., Atif, J., Bloch , I.: Fuzzy Spatial Relation Ontology for Image Interpretation. Fuzzy Sets and Systems, 159, 1929-1951 (2008)
8. Iakovidis D.K., Schober D., Boeker M., Schulz S.: An Ontology of Image Representations for Medical Image Mining. In 9th International Conference on Information Technology and Applications in Biomedicine, Larnaca (2009)
9. Iakovidis D.K., Smailis C.V.: Efficient Semantically-Aware Annotation of Images. In International Conference of Imaging Systems and Tech, pp. 146-149, Penang (2011).
10. Baader F., Calvanese D., McGuinness D., Nardi D., Patel-Schneider P.: The Description Logic Handbook: Theory, Impl. and Appl. Cambridge University Press, Cambridge, 2003.
11. Shiraishi, J., et al.: Development of a Digital Image Database for Chest Radiographs with and without a Lung Nodule: Receiver Operating Characteristic Analysis of Radiologists Detection of Pulmonary Nodules, Am. J. Roentgenol., vol. 174, pp. 71–74, 2000.
12. Van Ginneken, B., Stegmann, M.B., Loog, M., Segmentation of Anatomical Structures in Chest Radiographs using Supervised Methods: a Comparative Study on a Public Database. Medical Image Analysis, 10, 19-40 (2006)
13. Golbreich, C., Zhang, S., Bodenreider, O.,: The foundational model of anatomy in OWL: Experience and perspectives. Journal of Web Semantics, Web Semantics: Science, Services and Agents on the World Wide Web, 4, 181-195 (2006)