A VISUAL ONTOLOGY QUERY INTERFACE FOR CONTENT-BASED IMAGE RETRIEVAL

D. N. F. AWANG ISKANDAR, JAMES A. THOM AND S. M. M. TAHAGHOGHI
School of Computer Science and Information Technology, RMIT University, Melbourne, Australia
{dayang.awgiskandar, james.thom, seyed.tahaghoghi}@rmit.edu.au

Abstract. Various querying techniques have been developed for content-based image retrieval. We propose a *Visual Ontology Query Interface* for querying an OWL ontology built using content-based image retrieval techniques. With the query interface, users are able to formulate various ontology queries without having to know SPARQL, an ontology query language proposed by The World Wide Web Consortium.

1. Introduction

The search for images has been explored using text and image content. In content-based image retrieval (CBIR), image content is frequently represented through image features. Commonly used features include colour, texture, or shape descriptors for objects found within an image.

The main problem in CBIR is the semantic gap, defined as the lack of correlation between the semantic categories that a user is looking for and the low-level features that CBIR systems offer. To bridge the semantic gap, we need to capture the information need by allowing the user to express their query to the CBIR system; images in the collection that have low-level features similar to the query examples are returned in response to the query. This paradigm is known as query by example (QBE), where the query is expressed as one or more example images (Smeulders, Worring, Santini, Gupta and Jain, 2000). Presenting a whole image as the query limits the ability of users to express their information needs. A study by Enser (1993) underscores the need of expressing region or regions of interest as the query into CBIR systems.

In this paper, we propose and explore the use of a query interface that automatically translates the visual query into an ontology query language. The query interface enables a user to select the region of interest as the query example, allowing the CBIR system to retrieve images that contain regions similar to those specified in the query. Instead of accessing the image region low-level features directly, we use an image ontology as part of a top down approach towards bridging the semantic gap in CBIR (Hare, Lewis, Enser and Sandom, 2006; Enser, 1993).

The remainder of this paper is organised as follows. In Section 2, we describe existing research on CBIR and ontology. In Section 3, we explain the structure and content of the ontology. In Section 4, we demonstrate the query interface and processing. We conclude in Section 5 with a discussion of our findings and suggestions for future work.

2. Related Work

Most effort to minimize the semantic gap has focused on automatic image annotation (Barnard and Forsyth, 2001; Jelmini and Marchand-Maillet, 2003; Troncy et al., 2007) where images are annotated by using keywords or described formally using an ontology (Hyvönen, Saarela, Styrman and Viljanen, 2003; Schreiber, Dubbeldam, Wielemaker and Wielinga, 2001; Troncy et al., 2007). An ontology is used to define concepts, properties and the relationships among the concepts (Brewster et al., 2004).

Web ontology languages have been proposed as part of research related to the Semantic Web. RDF, RDF Schema, OIL and DAML+OIL are among the earliest web ontology languages, while OWL is the most prominent ontology language (Escórcio and Cardosa, 2007) and the current W3C recommendation.[1]

Ontology query languages allow expressions to be written that can be evaluated against an ontology. The queries can be used by knowledge management applications as a basis for inference actions. Existing ontology query languages include OntoQL, SPARQL, DQL (previous version of DAML+OIL), SeRQL, TRIPLE, RDQL, N3, and Versa. The **S**PARQL **P**rotocol **a**nd **R**DF **Q**uery **L**anguage (SPARQL), is an ontology protocol[2] and query language[3] that has been adopted by W3C as the means to access and query ontologies built using RDF. Currently, SPARQL has been extended to support OWL ontologies.

[1] http://www.w3.org/TR/owl-guide

[2] http://www.w3.org/TR/rdf-sparql-protocol

[3] http://www.w3.org/TR/rdf-sparql-query

Town (2006) used ontologies to relate semantic descriptors to their parametric representations for visual image processing. He shows that using ontologies leads to an effective computational and representational mechanism. Town also proposed an ontological query language, OQUEL.

Hyvönen et al. (2003) implemented the Ontagator web-based system to retrieve images using an ontology. An image query is posed by choosing the subject category followed by an image example. Therefore, in the search process, users view the ontology and select the class of interest. The system will return all images related to the class. These images are then presented to the user. Liu et al. (2004) also implemented a web-based system in which the search for the matching images is done by processing a text-based query.

All the above-mentioned approaches use information about the whole image in the ontology, whereas Mezaris et al. (2003; 2004) propose an approach for region-based image retrieval using an object ontology and relevance feedback. Their approach utilises an unsupervised segmentation method for dividing the images into regions that are later indexed. The object ontology is used to represent the low-level features and act as an object relation identifier. For example the shape features are predefined in the ontology as slightly oblong, moderately oblong, very oblong. The ontology is simply a vocabulary listing and not represented in any ontology language. The query is formulated in text using the predefined keywords. Regions that match the query keywords are presented to the user. The user can give feedback on the retrieved images, the system learns the relevant answers using a Support Vector Machine (SVM). The Constraint Similarity Measure (CSM) is used to filter out the unrelated images (Mezaris et al., 2004).

3. The Image Ontology

To reduce the problem of object segmentation, we test our approach on a domain where regions are easily separated: a collection of comic strips. In this domain, objects and characters comprise multiple regions of approximately uniform colour.

We have created an image collection that consists of comic strip panels from the Goats comic.[4] These include extended information that describes every panel. This description assists us in performing relevance judgements on our retrieval results.

[4] http://www.goats.com

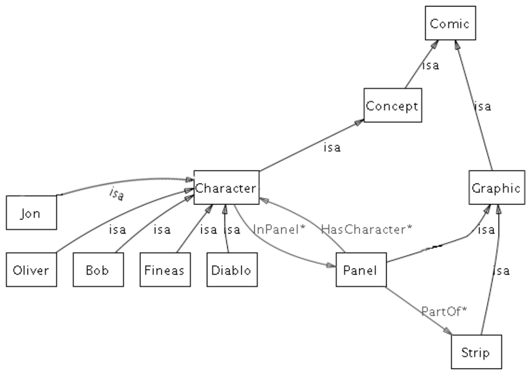

Figure 1. Visual graph of the image ontology. This figure is best viewed in colour.

The collection consists of 452 coloured strips, each containing one to five panels. Dividing the strips into panels gives us a total of 1440 panels. We tested the retrieval effectiveness using 1115 regions extracted from 202 panels. From this point onwards, we refer to individual panels as images.

The objects in the comics have relatively consistent size and orientation, guiding our choice of the following region-based and contour-based shape features: the region area; the mean grey level value of the pixels in the region; the compactness (also known as circularity ratio or thinness ratio) of the region; and shape boundary of the region. Compactness reflects how circular the shape is. A compactness value of one indicates that the shape is a circle and zero compactness signifies a narrow shape with infinite width. Shape boundary is represented by the keypoints surrounding the shape.

We adopted the equal-weight linear combination technique from our previous work (Awang Iskandar et al., 2008) to recognise and label five concepts representing the main characters in the Goats comic strips — Bob (an alien), Diablo (a chicken), Fineas (a fish), Jon (a person) and Oliver (a chick). This technique was compared with classification using machine learning algorithms in combining shape features, and we found that an equal-weight linear combination of shape features is simpler and at least as effective as using a machine learning algorithm (Awang Iskandar et al., 2008).

Figure 1 illustrates the ontology structure and Figure 2 is a snippet of the OWL representation. We divided the ontology structure into two general classes – Concept and Graphic. The class Concept has a subclass Character that further contains subclasses that represent the characters – Bob, Diablo, Fineas, Oliver and Jon. The Graphic subclasses are Strip and Panel.

```
<?xml version="1.0"?>
    <rdf:RDF
    xmlns="http://dayang.cs.rmit.edu.au/~dayang/ComicOntology#"
    xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
    xmlns:xsd="http://www.w3.org/2001/XMLSchema#"
    xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
    xmlns:owl="http://www.w3.org/2002/07/owl#"
    xml:base="http://dayang.cs.rmit.edu.au/~dayang/ComicOntology">
    <owl:Ontology rdf:about="Goats Comic"/>
    <owl:Class rdf:ID="Concept"><rdfs:subClassOf><owl:Class rdf:ID="Comic"/>
    </rdfs:subClassOf></owl:Class>
    <owl:Class rdf:ID="Graphic"><rdfs:subClassOf rdf:resource="#Comic"/></owl:Class>
    <owl:Class rdf:ID="Strip"><rdfs:subClassOf rdf:resource="#Graphic"/></owl:Class>
    <owl:Class rdf:ID="Panel"><rdfs:subClassOf rdf:resource="#Graphic"/></owl:Class>
    <owl:Class rdf:ID="Character"><rdfs:subClassOf rdf:resource="#Concept"/></owl:Class>
    <owl:ObjectProperty rdf:ID="HasCharacter"/>
    <owl:ObjectProperty rdf:ID="PartOf"/>
    <owl:ObjectProperty rdf:ID="InPanel"/>
    <owl:DatatypeProperty rdf:ID="name"/>
    <owl:DatatypeProperty rdf:ID="image_Ref"/>
    <Panel rdf:ID="goats050713.png-a">
    <HasCharacter><Diablo rdf:ID="Diablo_goats050713.png-a">
    <CentroidX rdf:datatype="http://www.w3.org/2001/XMLSchema#unsignedInt">43</CentroidX>
    <CentroidY rdf:datatype="http://www.w3.org/2001/XMLSchema#unsignedInt">123</CentroidY>
    <InPanel rdf:resource="#goats050713.png-a"/>
    <name rdf:datatype="http://www.w3.org/2001/XMLSchema#string">Diablo</name> </Diablo>
    </HasCharacter><HasCharacter><Fineas rdf:ID="Fineas_goats050713.png-a">
    <CentroidX rdf:datatype="http://www.w3.org/2001/XMLSchema#unsignedInt">165</CentroidX>
    <CentroidY rdf:datatype="http://www.w3.org/2001/XMLSchema#unsignedInt">227</CentroidY>
    <InPanel rdf:resource="#goats050713.png-a"/>
    <name rdf:datatype="http://www.w3.org/2001/XMLSchema#string">Fineas</name> </Fineas>
    </HasCharacter><HasCharacter><Oliver rdf:ID="Oliver_goats050713.png-a">
    <CentroidX rdf:datatype="http://www.w3.org/2001/XMLSchema#unsignedInt">192</CentroidX>
    <CentroidY rdf:datatype="http://www.w3.org/2001/XMLSchema#unsignedInt">112</CentroidY>
    <InPanel rdf:resource="#goats050713.png-a"/>
    <name rdf:datatype="http://www.w3.org/2001/XMLSchema#string">Oliver</name>
    </Oliver></HasCharacter>
    &lt;image_Ref rdf:datatype="http://www.w3.org/2001/XMLSchema#string"&gt;
      goats050713.png-a&lt;/image_Ref&gt;<PartOf><Strip rdf:ID="goats050713">
    <name rdf:datatype="http://www.w3.org/2001/XMLSchema#string">goats050713</name>
    </Strip></PartOf></Panel>
    ...
</rdf:RDF>
```


Figure 2. Ontology in OWL.

We adopted this ontology structure so the graphic elements of the image collection are separated from the semantic concepts. Classes in an ontology are connected using relationships. The class Character has an InPanel relationship with the class Panel. The InPanel relationship has an inverse relationship of HasCharacter. The class Panel has a relationship of Partof with the class Strip. All other concepts are connected with an is-a relationship. The ontology contains a total of 3680 instances.

4. The Visual Query Interface and Query Processing

The Visual Ontology Query Interface (VOQI) is a simple query interface that supports querying up to three combinations of regions. The regions of interest that belong to the characters are presented as icons. There are two types of query: a *character query*, to find the desired characters in an image, and a *spatial location query*, which allows the user to pose a query that retrieves characters, based on their spatial location. To combine the characters, we use several *combination operators*:

- `AND` to perform a logical conjunction on two expressions. This operator is to retrieve images containing more than one character.
- `OR` to perform a logical disjunction on two expressions. We follow the *exclusive OR* operator (XOR) to retrieve images containing either one of the characters specified in the query.
- `NOT` to perform logical negation on an expression. We specified this operator as `NOT WITH` that returns images without the specified character.
- `RIGHT OF` to perform the "right" positional expression; and
- `LEFT OF` to perform the "left" positional expression.

When a query contains multiple combination operators, the system evaluates and resolves the query according to the *order of precedence*. This means that the `NOT` operator will be evaluated first, then `AND`, followed by `OR`. In the case of a spatial location query, we evaluate the combination from left to right.

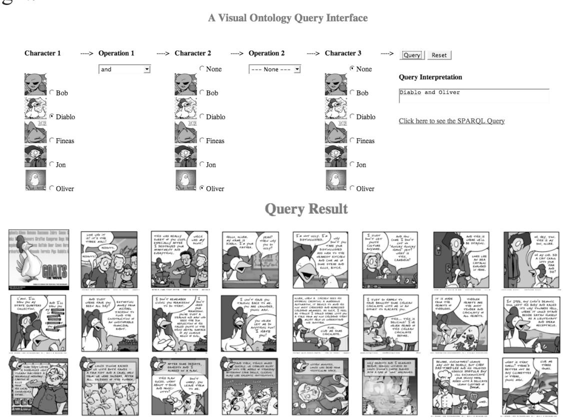

Figure 3. Querying two characters with the `AND` operator.

A sample query "Find Oliver `AND` Diablo" is depicted in Figure 3. Upon submitting the query, the VOQI *query translator engine* interprets the query into SPARQL, which is specified in Figure 4. The SPARQL query is then issued to the *ontology query engine*. A set of images matching the query is returned and displayed to the user.

Interpreting the SPARQL query, there is a SELECT clause that identifies the criteria to appear in the query results; a WHERE clause that specifies the criteria for selecting results from the database; and a FILTER clause that

```
PREFIX comic: <http://dayang.cs.rmit.edu.au/~dayang/comicontology#>
SELECT ?Panel
FROM
<http://dayang.cs.rmit.edu.au/~dayang/comicontology.owl>
WHERE {
?x comic:name ?CharacterName1;
comic:InPanel ?Panel.
?y comic:name ?CharacterName2;
comic:InPanel ?Panel.
FILTER regex( ?CharacterName1, "^Diablo", "i")
FILTER regex( ?CharacterName2, "^Oliver", "i")}
```

Figure 4. The SPARQL Query converted from the VOQI interface.

restricts the results according the expression. For this particular query, we are searching for panels that contain both the characters "Diablo" and "Oliver".

5. Conclusion and Future Work

In this paper we have described an ontology engineered using CBIR techniques to recognize objects that are in the images. The ontology is represented using OWL, a portable and machine independent ontology language. We have shown that synergy between ontology and image annotations is possible, and this method can reduce the gap between image features and high-level semantics by providing the relationships between objects in the image.

We also built a visual interface to ease the ontology querying process where there is no need for the user to master the SPARQL query language. This interface can easily be adapted to mobile devices as it is a simple input form.

We plan to design a more interactive query interface where the regions of interest (depicted by the character's icon) can be dragged and dropped onto a canvas. The query could then be built based on the position of the icon on the canvas.

Acknowledgements

This research was undertaken with a scholarship provided by the Malaysian Ministry of Higher Education. We thank Jonathan Rosenberg, author of the Goats Comic for permission to publish the comic images.

References

Awang Iskandar, D. N. F., Thom, J. A. and Tahaghoghi, S. M. M.: 2008, Content-based image retrieval using image regions as query examples, *In Prooceedings of the Nineteenth Australasian Database Conference (ADC 2008)*, Vol. 75 of CRPIT, ACS, Wollongong, Australia.

Barnard, K. and Forsyth, D.: 2001, Learning the semantics of words and pictures, *In Proceedings of the 8th International Conference on Computer Vision*, pp. 408–415.

Brewster, C., O'Hara, K., Fuller, S., Wilks, Y., Franconi, E., Musen, M. A., Ellman, J. and Shum, S. B.: 2004, Knowledge representation with ontologies: The present and future, *IEEE Intelligent Systems*, pp. 72–81.

Enser, P. G. B.: 1993, Query analysis in a visual information retrieval context*, Journal of Document and Text Management* 1(1): 25–52.

Escórcio, A. L. N. and Cardosa, J.: 2007, Editing tools for ontology creation, *Semantic Web Services: Theory, Tools and Applications*, Information science reference, chapter 4, pp. 71–95.

Gandon, F.: 2002, *Ontology engineering: a survey and a return on experience*, Technical Report 4396, INRIA.

Hare, J. S., Lewis, P. H., Enser, P. G. B. and Sandom, C. J.: 2006, Mind the gap: Another look at the problem of the semantic gap in image retrieval, *In Proceedings of Multimedia Content Analysis, Management and Retrieval 2006, SPIE*, Vol. 6073, pp. 607309–1.

Hyvönen, E., Saarela, S., Styrman, A. and Viljanen, K.: 2003, Ontology-based image retrieval, *In Proceedings of the 12th International World Wide Web Conference* (WWW2003). poster papers.

Hyvönen, E., Saarela, S. and Viljanen, K.: 2003, Ontogator: Combining view and ontology-based search with semantic browsing, *In Proceedings of XML Finland 2003*. Paper presented at the international SEPIA Conference, Helsinki, Sept. 18-20, 2003.

Jelmini, C. and Marchand-Maillet, S.: 2003, Deva: an extensible ontology-based annotation model for visual document collections, *SPIE*, Vol. 5018.

Liu, S., Chia, L.-T. and Chan, S.: 2004, Ontology for nature-scene image retrieval., In CoopIS/DOA/ODBASE (2), Vol. 3291 of *Lecture Notes in Computer Science*, Springer, pp. 1050–1061.

Mezaris, V., Kompatsiaris, I. and Strintzis, M. G.: 2003, An ontology approach to object-based image retrieval., *In Proceedings of the IEEE International Conference on Image Processing (ICIP '2003)*, pp. 511–514.

Mezaris, V., Kompatsiaris, I. and Strintzis, M. G., 2004, Region-based image retrieval using an object ontology and relevance feedback, *EURASIP Journal on Applied Signal Processing* (6): 886–901.

Schreiber, A. T. G., Dubbeldam, B., Wielemaker, J. and Wielinga, B.: 2001, Ontology-based photo annotation, *IEEE Intelligent Systems* **16**(3): 66–74.

Smeulders, A. W. M., Worring, M., Santini, S., Gupta, A. and Jain, R.: 2000, Content-based image retrieval at the end of the early years*, IEEE Transactions on Pattern Analysis and Machine Intelligence* **22**(12): 1349–1380.

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

Troncy, R., van Ossenbruggen, J., Pan, J. Z. and Stamou, G.: 2007, *Image annotation on the semantic web,* W3C incubator group report.
URL: http://www.w3.org/2005/Incubator/mmsem/XGR-image-annotation