

Evaluating the Application of Semantic Inferencing Rules to Image Annotation

L. Hollink

Business Informatics
Vrije Universiteit Amsterdam
hollink@cs.vu.nl

S. Little

ITEE Department and DSTC
University of Queensland
slittle@dstc.edu.au

J. Hunter

DSTC
The University of Queensland
jane@dstc.edu.au

ABSTRACT

Semantic annotation of digital objects within large multimedia collections is a difficult and challenging task. We describe a method for semi-automatic annotation of images and apply it to and evaluate it on images of pancreatic cells. By comparing the performance of this approach in the pancreatic cell domain with previous results in the fuel cell domain, we aim to determine characteristics of a domain which indicate that the method will or will not work in that domain. We conclude by describing the types of images and domains in which we can expect satisfactory results with this approach.

Categories and Subject Descriptors:

H [Information Systems]

General Terms: Human Factors

Keywords: Image Annotation, Inferencing Rules, Evaluation.

1. INTRODUCTION

Semantic annotation of visual resources is essential to easing the discovery of the rapidly increasing quantity of digital visual content. Such descriptions enable sophisticated semantic querying of the media in terms familiar to the user's domain whilst also ensuring that the information and knowledge have a much greater chance of being discovered and exploited by services, agents and applications on the Web. Because of the quantity and complexity of visual data, manual annotation is slow, expensive and highly subjective. Despite significant advancements in the field of image analysis, the automatic generation of high-level, semantic annotations of images remains a significant challenge.

Earlier research [10] developed a semi-automatic, user-assisted approach to generating ontology-based annotations of image regions from low-level, automatically extracted features. This prototype enables experts to define rules specific to their domain, which map particular combinations of low-level visual features (colour, texture, shape, size etc.) to high-level semantic terms defined in their domain ontology. These semantic inferencing rules capture a domain-expert's understanding of how low-level features are related to ontology terms. The rules are recorded in an XML-based format and can be shared, collaboratively modified and annotated as the domain understanding shifts and changes.

To overcome the difficulty that domain experts face when developing complex rules in XML format using unfamiliar terminology, a visual interface called Rules-By-Example (RBE) was developed. For example, an oncologist labelling brain scans to enable the search and retrieval of particular types of tumours, may define the following rule: IF [(color is like *this*) AND (texture is like *this*) AND (shape is like *this*)] THEN (the object is an astrocytoma). The system assists users to construct rules with palettes of example colours, shapes defined using drawing tools, and example regions within the media collection.

The linking of low-level image data to high-level domain concepts is challenging due to what Smeulders et al [16] call the semantic gap. The question arises as to what are the conditions under which our approach can successfully establish such links and apply them to image annotation. The approach of RBE and semantic inferencing rules was previously evaluated in the domain of fuel cell microscopy [12] where a small study demonstrated promising results. The system was now extended to adapt it to the more complex domain of pancreatic cells. By comparing the characteristics of the two domains and examining the results and types of rules produced, we aim to determine the characteristics of domains, which may benefit from semantic inferencing rules and the Rules-By-Example system.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

K-CAP'05, October 2-5, 2005, Banff, Alberta, Canada.
Copyright 2005 ACM 1-59593-163-5/05/0010 ...\$5.00

In this paper we describe the system architecture and the changes made to the RBE and semantic inferencing system in order to support this new complex domain of pancreatic cell analysis. The evaluation domain, pancreatic cell images, is presented – together with the vocabularies and ontologies that were developed. We describe the rules that were defined, determine the accuracy of these rules and discuss the characteristics of domains in which this approach is likely to succeed.

2. RELATED WORK

A number of research efforts have investigated the use of automatic recognition techniques to extract low-level visual or audio features which together can be used to generate semantic descriptions of multimedia content. These include statistically-based, machine-learning methods such as those proposed by [5, 15, 22, 2] which first manually annotate sample sets and from this generate factor graphs, statistical models or other indexing techniques for the larger collection. Marques and Barman [13] integrated ontologies into a machine-learning based approach to semantically annotate images.

Overall the use of machine learning techniques to bridge the semantic gap provides a relatively powerful method for discovering complex and hidden relationships or mappings. However the ‘black-box’ method often employed can be difficult to develop and maintain because its effectiveness depends on the design and configuration of multiple variables and options. The relationships discovered between low-level media features and semantic descriptions remain hidden and are not able to be examined or manipulated by the human domain expert. In addition, extensive, detailed and specific training corpuses are required to ensure optimum performance. These cannot easily be adapted to new domains or incorporate new content or knowledge.

Methods for linking visual thesauri or ontologies to multimedia have been developed by Hoogs et al [7] and Tansley [18]. These utilise the relationships described by the ontology or thesaurus to enable more complex, semantic queries across collections of annotated media and to infer new information. However the difficulty of forming the relationships between the media and the ontological terms still remains.

Rules have long been used as a means of capturing expert knowledge [4, 17] and extracting information from data. If recorded in an open and transparent fashion, they are able to clarify understanding of a domain’s paradigm and act as a catalyst for discussion and exchange. The Semantic Web initiative [21] includes a layer for logic, reasoning and rules processing. The standards (XML, RDF, OWL, RuleML, SWRL) and processing tools (CWM, Mandarax, JESS) that are emerging to supply this layer are intended to provide open, interoperable formats for the exchange, discussion and

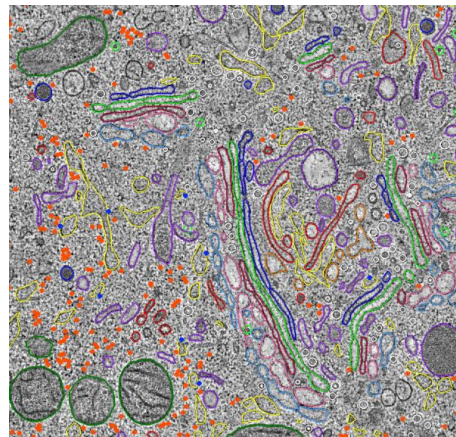


Figure 1: Segmented image of a pancreatic cell

application of data, ontologies and rules. For example, Hatala and Richards [6] have used ontologies in combination with rules to improve metadata for learning objects by suggesting relevant values.

We believe that our approach overcomes some of the limitations in existing image annotation approaches (such as the difficulty of determining the distinguishing features and adapting to different domains). The resulting semantic inferencing rules are a form of knowledge in themselves and can be discussed, annotated, shared and applied as the user directs. We do this through the complementary use of semantic web technologies and an interface which allows domain experts to intuitively and interactively develop and define semantic inferencing rules in an interoperable, machine-understandable and shareable format.

3. THE DOMAIN OF PANCREAS CELLS

At the Institute for Molecular Bioscience of the University of Queensland, the Visible Cell project aims to heighten the understanding of processes in mammalian cells [11]. One of the goals of this project is to create a 3D image of a cell. To this end, pancreatic cells were cut into 400-nm-thick slices and each slice was studied by electron tomography [14]. Molecular biologists segmented and annotated the images of these slices by drawing lines around each cellular component (Fig. 1).

480 of these images were combined into a single, high-resolution three-dimensional reconstruction of a $3.1 \times 3.2 \times 1.2 \mu\text{m}^3$ area in a pancreatic cell. Figure 2 illustrates the spatial layout of a number of important cellular components including: the golgi apparatus, endoplasmic reticulum, mitochondria, ribosomes, and different types of vesicles. A new high-throughput microscope will soon be employed that is capable of producing even larger numbers of images, heightening the need for automatic segmentation and annotation.



Figure 2: 3D image of a pancreatic cell

Segmented 2D images were used as input to our RBE system. Each segmented region depicts a cellular component. The components that are visible in a cell all have distinguishing combinations of textures, colors, sizes, and other features. We used the Matlab Image Processing Toolbox to extract these features. The toolbox provides a number of built-in functions for feature extraction and custom routines for extracting additional features can be constructed relatively easily.

The pancreatic images differ in a number of ways from the images in the previous domain of fuel cell microscopy. In the pancreas domain a larger number of classes has to be identified, there is less visual distinctiveness between objects of different classes and less visual uniformity of objects of the same class. Because of this, the rules in the pancreas domain are more complex than in the fuel cell domain.

4. VOCABULARIES

In order to produce semantic annotations, we use existing vocabularies to represent the domain of pancreatic cells, the visual image features and the semantic inferring rules.

4.1 A vocabulary for the pancreas domain

Our domain involves semantic descriptions of cellular components. The Medical Subject Headings thesaurus¹ (MeSH) is used for indexing and searching biomedical and health related documents. Since MeSH contains a hierarchy of cellular components we decided to reuse this large existing thesaurus, rather than build our own. We used the version of Van Assem et al. [20], who trans-

¹<http://www.nlm.nih.gov/mesh/meshhome.html>

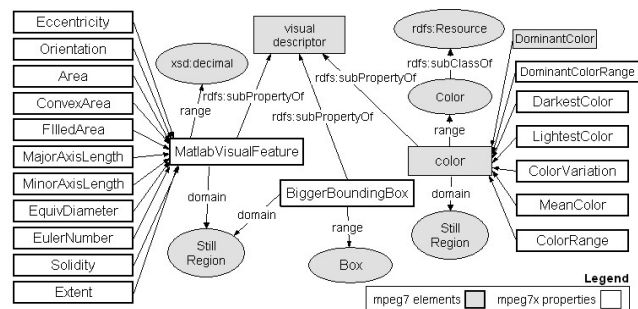


Figure 3: MPEG-7 Visual Descriptor ‘Color’.

lated MeSH from the native format to RDF. In meetings with molecular biology experts, we established the terms that they use to describe the cellular components in this collection of pancreatic cell images, and the mappings between these terms and MeSH concepts. Direct mappings were not always possible due to the fact that the experts use functional descriptions such as “vesicle carrying cargo”, or visual descriptions such as “small tubular vesicles”. In these cases we extended the MeSH thesaurus with subclasses of the class “vesicle”.

4.2 Multimedia Ontology

The MPEG-7 ontology is used as a vocabulary for the visual image features. MPEG-7 is a standard for describing multimedia content published by the Moving Picture Experts Group (MPEG) [19]. The MPEG-7 OWL ontology as published by Hunter [9] includes low-level visual properties such as color, shape and motion.

In various meetings with the domain experts, we asked them to identify those visual characteristics that distinguish different cellular components. We used the Matlab Image Processing Toolbox to extract these visual characteristics for each region. However, MPEG-7 does not contain classes that describe many of these detailed visual features. Therefore, we extended MPEG-7 with 14 Matlab built-in image features (e.g., area, eccentricity) and 10 custom image features that we constructed with the Matlab Image Processing Toolbox (e.g., color range, mean color) (see Fig. 3). All Matlab image features were represented as subproperties of the existing MPEG-7 property ‘visualDescriptor’. The custom image features all concerned color characteristics and were added as subproperties of `mpeg7:color`.

Many of the Matlab visual concepts are too specific to be comprehensible to anyone but image analysis specialists. We cannot expect biologists to understand what ‘eccentricity’ is, or what values the property ‘density’ may take. Instead of these specialist terms, the molecular biologists use more commonly known terms like ‘long’, ‘round’, and ‘close’ to describe visual characteristics. We defined rules to translate low-level terms such as ‘eccentricity’ and ‘density’ to more familiar commonly-

used intermediate-level terms like ‘long’ and ‘close’. Abe-lla and Kender researched such links between low-level features and commonsense terms [1]. They found, for example, that humans consider two objects to be close to each other if the bounding boxes of two objects multiplied by 1.6 overlap. We used this to construct the following rule: IF (the bigger bounding box of region1 overlaps with the bigger bounding box of region2) THEN (region1 is close to region2) Additional intermediate-level terms we defined in this way include: *long*, *solid*, *irregular*, *round*, *dense* and *touching*. These were added to the MPEG-7 ontology in the same way as the low-level Matlab image features, as sub-properties of existing MPEG-7 properties.

4.3 Rule Language

We use RuleML to represent both the rules that relate low-level visual features to intermediate-level terms, and the semantic inferencing rules that experts defined to annotate cellular components. RuleML aims to provide a shareable, XML-based rule markup language for rule storage, interchange, retrieval and firing/application [3]. Using this format ensures that our rules are machine readable and interoperable with existing tools and standards. A proposal for a new rule language that combines RuleML and OWL has now been submitted to the W3C as the Semantic Web Rule Language (SWRL) [8]. We plan to upgrade to SWRL once it stabilises and tool support increases, since it is more expressive than RuleML. The biggest improvement from our point of view is the inclusion of numerous built-in relations. For readability, we use the SWRL informal syntax rather than the lengthy XML syntax of RuleML, in the rule examples throughout this document.

A typical example of a rule in our domain is the rule below which is applied to recognise mature granules in an image. The rule:

$$\begin{aligned} &mpeg7 : StillRegion(region) \wedge \\ &mpeg7x : Dense(region) \wedge \\ &mpeg7 : DominantColor(region, col) \wedge \\ &swrlb : lessThan(col, 100) \\ &\rightarrow mpeg7 : Depicts(region, mesh : MatureGranule) \end{aligned}$$

states that a region depicts a mature granule if it is dense and its DominantColor value is less than 100. The prefixes *mpeg7*, *mpeg7x* and *swrlb* indicate terms from the MPEG-7 ontology, our extensions to the MPEG-7 ontology and SWRL built-ins, respectively.

5. SEMANTIC INFERENCING AND THE RULES-BY-EXAMPLE INTERFACE

As the previous section demonstrates, semantic inferencing rules to relate low-level image features to high-level semantic annotations can be complex, require understanding of specific media feature definitions and are formatted in XML. In contrast, domain experts generally have only a basic understanding of rule structure

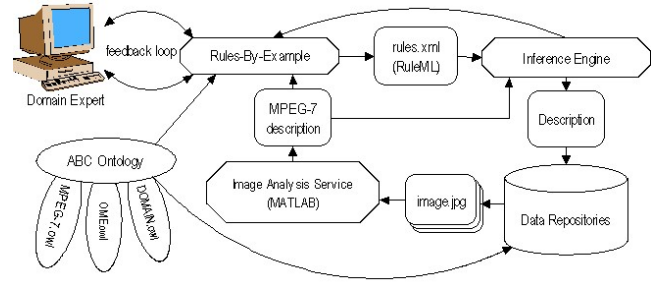


Figure 4: Architecture for the RBE System

and its application, often have limited knowledge of media terminology and rarely enjoy constructing XML syntax. To overcome this the Rules-By-Example system exploits: domain ontologies and multimedia ontologies; intermediate rules to describe media features in terms more commonly used by domain experts (long, dark etc.); and example-based definitions of media features.

6. PREVIOUS WORK

Figure 4 illustrates the components of the Rules-By-Example (RBE) system. The application utilises the MPEG-7 descriptions and ontology and the domain ontology to build an interface which incorporates familiar, semantic terms. MPEG-7 based descriptions of the regions within the image are loaded into the RBE application and the user is able to specify values for visual features from palettes of sample colors and textures, drawing tools or by specifying regions from sample images selected from the data repository. This method of defining the rules, “by example”, is more intuitive and reduces the prior knowledge and understanding required to build semantic inferencing rules. For example, the user is able to drag-n-drop colour selections as opposed to entering RGB definitions. More information about its implementation can be found in [12].

As the user is constructing a rule, the system is able to evaluate it against a sample set of labeled images and present the number of currently matching regions. This simple feedback enables the user to determine when a rule may be most accurate or when a rule has become too specific. Once the user is satisfied with a rule, it is saved in RuleML format, augmented with MathML where necessary to describe mathematical relationships and saved to an XML database. This can then be made available to collaborators over the Web for discussion, re-use and refinement as a result of application to their own collections of images. A complete set of metadata describing the rule, including evaluation results and the data set used in development, is also recorded ensuring the provenance of the process is well documented.

Overall the Rules-By-Example interface allows the user to quickly develop, apply and refine highly complex rules while reducing the need to understand low-level

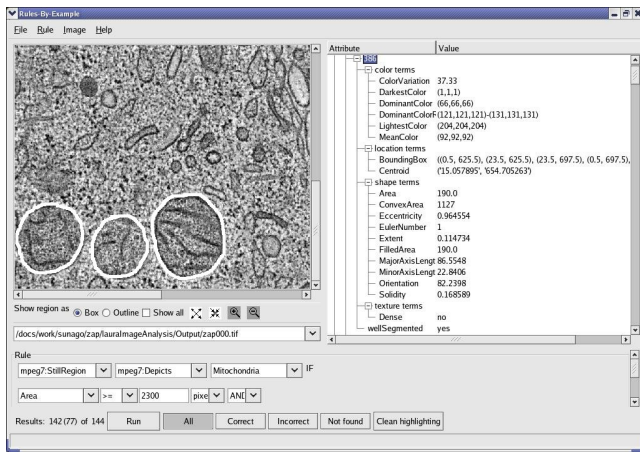


Figure 5: Screenshot of the RBE System

MPEG-7 terms or values. In addition, it enables users to direct the system so that it focuses on the objects, regions or distinguishing features of highest priority or interest – in contrast to traditional approaches which require pre-selection of important features. The rules themselves are recorded in a flexible, interoperable format and together with their provenance metadata can be easily distributed, discussed and modified.

6.1 Application to Pancreatic Cell Images

Since the prototype RBE interface was developed for the fuel cell domain, a few changes were required to apply it to the new pancreatic cell domain. Firstly the fuel cell ontology was replaced by the MeSH ontology. Secondly, changes were made to the data access module to connect to the new database and schema used for the pancreatic cell images. The most significant change was required due to the greater complexity of the pancreatic cell images. This required the definition of more intermediate rules than were required by the simpler fuel cell images. Changes were made to the way in which the available media properties were loaded into the interface and displayed with the image’s metadata. This involved grouping mpeg7x properties based on their root in the MPEG-7 ontology (color, shape etc). Figure 5 shows a screenshot of the RBE interface, highlighting regions in the pancreas image which match the current rule.

The application and processing of the rules remains a bottleneck for the system. Due to the limitations of RuleML in defining mathematical relationships and the current limitations of the inferencing engine used in this prototype (Mandarax), pre-processing of data by Python scripts is required to apply relationships such as *less than*, *equal to*, *greater than* etc. Processing of the rules involved a combination of Python scripts, complex MySQL queries, pre-application of intermediate rules and processing through the Mandarax inferencing engine. This series of steps severely limited the accuracy

and responsiveness of the dynamic feedback component of the RBE interface and may have had a negative impact on rules developed by the domain expert.

In conclusion, the migration and application of the semantic inferencing and RBE system to the new domain of pancreatic cells was relatively painless. A primary goal of the RBE system was to maintain as much domain independence as possible by enabling different domain ontologies to be easily plugged in. It was a simple process to replace the existing fuel cell ontology with the MeSH ontology to provide the semantic terms for the rules. The increased complexity of the cell images required extensions to provide greater support for intermediate rules. The issue of processing speed that limits the dynamic response is a critical one. Consequently a number of alternative tools and standards, such as SWRL translation through XSLT to CLIPS (JESS) or Prolog are being investigated to overcome this.

7. EVALUATION

7.1 Setup

Thirty pancreatic cell images, all slices from one cell, were used to evaluate the system. The regions had previously been manually segmented by domain experts into 7580 regions in total. A domain expert (molecular biologist) used the Rules-By-Example interface to create rules for the annotation of cellular components occurring within the images. Application of the rules generated annotations for each segmented region in the images. The correctness of the annotations was determined by comparing them with manual annotations that had been specified at the time of segmentation.

The biologists’ feedback on the RBE interface was on-the-whole positive. They noted, however, that the system would only be helpful if combined with accurate automatic segmentation. A project on automatic segmentation of pancreatic images is currently underway.

7.2 Rules formulated by a domain expert

Five rules were formulated by the domain expert to identify five classes of cellular components: golgi stack; endoplasmic reticulum; mitochondrion; ribosome; and mature granule. The *Golgi Apparatus* consists of a series (a ‘stack’) of long thin components alongside each other. Looking at the components separately would not distinguish them from components of other types, but when two or more of them occur close to each other, it is a very strong indication that they are part of a *golgi stack*. Hence the rule to recognise a golgi stack is: a region depicts a *Golgi Apparatus* if its eccentricity is greater than 0.98 and it is close to a region with an eccentricity greater than 0.98.

```
mpeg7 : StillRegion(region) ∧
mpeg7x : eccentricity(region, ecc) ∧
swrlb : greaterThan(ecc, 0.98 ∧
```

```

mpeg7x : close(region, region_y) ∧
mpeg7x : eccentricity(region_y, ecc_y) ∧
swrlb : greaterThan(ecc_y), 0.98 ∧
→ mpeg7 : Depicts(region, mesh : GolgiApparatus)

```

The *endoplasmatic reticulum* (ER) is an irregularly shaped component that can be distinguished by the fact that ribosomes are attached to it. However, the images in our collection are slices of a 3-dimensional cell and the connection with a ribosome is not necessarily visible in every slice. Therefore, the rule for ER is: a region depicts an *Endoplasmic Reticulum* if the region is touching a ribosome or the region is adjacent in the z-plane to a region touching a ribosome. We used Python scripts to calculate if two regions were ‘adjacent in the z-plane’.

```

mpeg7 : StillRegion(region) ∧
(touching(region, region_a) ∧
mpeg7 : Depicts(region_a, mesh : Ribosome)) ∨
(z_adjacent(region, region_b) ∧
(touching(region_b, region_c) ∧
mpeg7 : Depicts(region_c, mesh : Ribosome))
→ mpeg7 : Depicts(region, mesh : ER)

```

The rules for mitochondria and ribosomes are shown below. Section 4.3 describes the rule for mature granules.

```

mpeg7 : StillRegion(region) ∧
mpeg7x : shape(region, mpeg7x : solid) ∧
mpeg7x : eccentricity(region, eccentricity) ∧
swrlb : greaterThan(eccentricity, 0.9) ∧
mpeg7x : area(region, area) ∧
swrlb : greaterThan(area, 2300) ∧
mpeg7 : DominantColor(region, dominantColor) ∧
swrlb : greaterThanOrEqual(dominantColor, 105)
→ mpeg7 : Depicts(region, mesh : Mitochondrion)

```

```

mpeg7 : StillRegion(region) ∧
mpeg7x : area(region, area) ∧
swrlb : lessThan(area, 80pixels) ∧
Dense(region) ∧
mpeg7x : solidity(region, so) ∧
swrlb : greaterThan(area, 0.95) ∧
→ mpeg7 : Depicts(region, mesh : Ribosome)

```

7.3 Retrieval Results

Table 1 shows the results of applying the rules, with the number of relevant regions in the collection (Rel), the number of regions retrieved (Ret) and the number of relevant regions retrieved (RetRel). Precision is defined as RetRel/Ret, recall as RetRel/Rel. Results are shown for retrieval of the golgi stack (Go), the endoplasmatic reticulum (ER), mature granules (MG), mitochondria (Mi), ribosomes (Ri) and the mean of all regions.

The results demonstrate a number of things. Firstly,

	Go	ER	MG	Mi	Ri	Mean
Rel	2486	1463	221	105	1125	
Ret	185	153	1098	346	824	
RetRel	126	37	10	11	824	
Prec. (%)	68.11	24.18	0.91	3.18	100	39.28
Recall (%)	5.07	2.53	4.52	10.48	73.24	19.17

Table 1: Rules by domain expert

	Go	ER	MG	Mi	Ri	Mean
Rel	2486	1463	221	105	1125	
Ret	487	119	965	183	844	
RetRel	365	36	10	12	844	
Prec. (%)	74.95	30.25	1.04	6.56	100.00	42.56
Recall (%)	14.68	2.46	4.52	11.43	75.02	21.62

Table 2: Ordered rules

precision is higher than recall. Given that the aim is to construct a 3-dimensional image from a stack of images, this is a good thing: an incorrectly annotated region would cause more problems than a missing region that can be detected by neighbouring slices in the image stack. Secondly, the precision of the rule for ribosomes is 100%. This can be explained by the way segmentation was performed: a predefined circle with perfect shape and fixed size was used to manually segment the regions depicting ribosomes. This makes it very easy to extract ribosomes based on size and shape. Recall is less than 100% because regions can overlap, which breaks the perfect circular shape. Finally, mature granules score poorly on both recall and precision. We expect this is due to the large variations in shape that mature granules tend to display.

The order in which rules are executed affects the results, since an annotation given by one rule can not be overwritten by the next rule. Table 1 depicts the mean results of all possible orders of execution. To improve the overall results, we sorted the rules so that the most accurate and reliable rules were executed first, while the more general, imprecise rules were executed last. We found a slight improvement in both recall and precision (Table 2). The optimal order is Ri, Go, Mi, ER, MG.

One of the benefits of using rules is that they can be easily refined and improved. For example, we can tweak recall and precision by either increasing or decreasing the thresholds in the rules. When *better recall* is more desirable than *high precision*, one can modify the rules so that more regions are included. Using our knowledge of images features, we were able to make slight modifications to the domain expert’s rules that improved recall and precision (Table 3). We added a rule to extract tubular vesicles. Like ribosomes, they are easily recognisable due to the perfect circular shape of the segmentation. By executing this accurate and reliable rule first, we improved the results further.

	Go	ER	MG	Mi	Ri	TV	Mean
Rel	2486	1463	221	105	1125	148	
Ret	4689	48	578	83	855	148	
RetRel	2182	35	207	75	834	148	
Prec.	46.53	75.92	35.81	90.36	97.54	100	73.85
Recall	87.77	2.39	93.67	71.43	74.13	100	71.57

Table 3: Rules by the authors

7.4 A comparison to the fuel cell domain

Both recall and precision were significantly higher for fuel cell images than for the pancreatic cell images. The differences between the two domains lie in a number of factors. Firstly, the pancreas domain consists of a higher number of distinct object categories. We did not find, however, that this had a negative effect. Adding one more category (tubular vesicles) even improved the results. Secondly, the variations in appearance within categories of pancreatic cellular components was greater than in the fuel cell domain. The golgi stack and the ER, for example, can be seen in many shapes and sizes. This causes the rules to be more complex. Retrieving the golgi involves utilising the shape and position of neighbouring regions, while the rule to retrieve the ER includes examining and recognising regions in related images. Precision and recall of the golgi rule are less than that of the results in the fuel cell domain, but still reasonable. The rule to retrieve the ER proved too complex for our system at this stage, due in part to limitations in the segmentation of ER regions and touching ribosomes. The maximum recall we could achieve for ER was 2.4% which is unacceptable.

Regions in the pancreatic cell images that have similar visual characteristics to the regions in the fuel cell images, namely the mitochondria, ribosomes and tubular vesicles, achieve similar, satisfactory results.

8. DISCUSSION AND CONCLUSION

In this paper, we describe an interface for annotation of regions in images based on user-formulated semantic inferencing rules. We measured the performance in the domain of pancreatic cell images. We see possibilities for improvement in this domain by making more use of the 3-dimensional nature of the images and the volumetric and spatial relationships between image regions. Considering that one cellular component appears in several image slices, one can complete gaps in annotations of unidentifiable regions by examining the annotations of regions in the surrounding slices. Exploiting the third (depth) dimension could also be valuable in other domains, such as brain scans or mammograms.

In the pancreatic cell domain, the precision of the inferencing rules varied significantly. Applying the rules in order of decreasing precision showed an improvement in the results. We are planning to extend the RBE interface so that domain experts can easily manipulate

the order in which rules are executed. We are also investigating possible system upgrades that provide more direct feedback on the effect of the rules. However, this could increase the risk of over-fitting the rules to the data, and make them less applicable to other or new image collections in the same domain.

Another option to improve the ease of rule formulation is a fuzzy representation. Concepts like the size of a region or the irregularity of a shape are not easily expressed with precise boundaries, and using fuzzy logic might provide a more natural representation. A hybrid approach which incorporates machine-learning techniques to optimise values based on user-defined combinations and value ranges may also be useful.

The aim of this study was to determine domain characteristics that suit the semantic inferencing and Rules-By-Example approach. One of the significant findings was that knowledge of multimedia and image analysis terms is both a prerequisite and impediment to obtaining good results. We sought to overcome this barrier by providing intuitive graphical tools to formulate the rules, and by defining intermediate-level terms for building rules. We still found, however, that the results of applying rules defined by domain experts were significantly less than results of rules defined by the authors. This disparity may be reduced in time as the domain experts' use of the system improves. Other possible solutions include: (1) training of domain experts to make more them familiar with multimedia terms and techniques, and (2) enabling image analysis experts and domain experts to formulate rules collaboratively.

Our research determined that the segmentation step is important for the quality of the annotations. Ribosomes and tubular vesicles were annotated almost perfectly thanks to their simple shape and the level of segmentation. Clatharin coated vesicles, on the other hand, were a class of cellular components for which the domain experts considered it infeasible to define a rule. The clatharin coat of these vesicles appears as a cloud on the outside of the segmented region, and can therefore not be described by our image analysis techniques.

Our research indicated that the system performs better if the domain's scope is relatively narrow and consists of well-understood concepts that are widely agreed upon, so that subjectivity is minimised. Classes need to display small visual variance and be clearly visibly distinguishable from other classes. Given these characteristics, we expect this approach to be valuable for semantically annotating images in other medical domains, botanical domains (i.e., plant identification), or for analysis of remote surveillance satellite images.

Adapting the system to a new domain proved to be relatively easy. Most of the time and effort went into the

process of selecting and extending existing vocabularies to make them suitable for the specific domain. The RDF and OWL versions of MeSH and MPEG-7 provided us with easily extensible, interoperable, conceptual frameworks. RuleML was not sufficiently expressive for the types of rules required in a complex domain such as pancreatic cell images. The lack of built-in relations and limited integration with RDF/OWL meant that we had to use Python scripts and SQL queries to pre-process the data. This adversely affected the interactivity and responsiveness of the system. We expect that by using SWRL (which combines RuleML and OWL), this problem will be largely overcome.

Acknowledgements

Thanks to Brad Marsh and Adam Costin from the Institute of Molecular Biology for their feedback and the images. This work has been funded in part by DSTC through the Australian Federal Government's CRC Programme. The first author is funded by the IOP program of the Dutch ministry of Economic Affairs.

9. REFERENCES

- [1] A. Abella and J.R. Kender. From images to sentences via spatial relations. In *Proc. of the W. on Integr. of Image and Speech Understanding, 1999*.
- [2] B. Adams. Where does computational media aesthetics fit? *Multimedia, IEEE*, 10(2):18–27, 2003.
- [3] H. Boley, S. Tabet, and G. Wagner. Design rationale of RuleML: A markup language for semantic web rules. In *Semantic Web Working Symposium, 2001*.
- [4] B.G. Buchanan and E.H. Shortliffe, editors. *Rule-based expert systems: the MYCIN experiments of the Stanford Heuristic Programming Project*. Addison-Wesley, 1984.
- [5] S. F. Chang, W. Chen, and H. Sundaram. Semantic Visual Templates: linking visual features to semantics. In *IEEE International Conference on Image Processing.*, Chicago, 1998.
- [6] M. Hatala and G. Richards. Value-added Metatagging: Ontology and Rule based Methods for Smarter Metadata. In *Proc. of Rules and Rule Markup Languages for the Semantic Web, 2003*.
- [7] A. Hoogs, J. Rittscher, G. Stein, and J. Schmiederer. Video content annotation using visual analysis and a large semantic knowledgebase. In *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition, 2003*.
- [8] I. Horrocks, P.F. Patel-Schneider, H. Boley, S. Tabet and B. Groszof, and M. Dean. Swrl: A semantic web rule language combining owl and ruleml. W3c submission, W3C, May 2004.
- [9] J. Hunter. Adding Multimedia to the Semantic Web - Building an MPEG-7 Ontology. In *Int. Semantic Web Working Symposium, July 2001*.
- [10] J. Hunter, J. Drennan, and S. Little. Realizing the Hydrogen Economy through Semantic Web Technologies. *IEEE Intelligent Systems Journal - Special Issue on eScience*, 19(1):40–47, 2004.
- [11] Institute for Molecular Bioscience. The Visible Cell Project. See <http://www.imb.uq.edu.au>
- [12] S. Little and J. Hunter. Rules-By-Example - a Novel Approach to Semantic Indexing and Querying of Images. In *Proc. of ISWC, 2004*.
- [13] O. Marques and N. Barman. Semi-automatic Semantic Annotation of Images Using Machine Learning Techniques. In *Proc. of ISWC, 2003*.
- [14] B.J. Marsh, D.N. Mastronarde, K.F. Buttle, K.E. Howell, and J.R. McIntosh. Organellar relationships in the golgi region of the pancreatic beta cell line, hit-t15, visualized by high resolution electron tomography. *Proceedings of the National Academy of Sciences of the United States of America*, 98(5):2399–2406, January 2001.
- [15] T.S. Naphade, M.R. Huang. Detecting semantic concepts using context and audiovisual features. In *Proc. of the Workshop on Detection and Recognition of Events in Video, 2001*.
- [16] A.W.M. Smeulders, M. Worring, S. Santini, A. Gupta, and R. Jain. Content-based image retrieval at the end of the early years. *Pattern Analysis and Machine Intelligence*, 22(12), 2000.
- [17] John F. Sowa. *Knowledge Representation: Logical, Philosophical and Computational Foundations*. Brooks/Cole, 2000.
- [18] Robert Hugh Tansley. *The Multimedia Thesaurus: Adding a Semantic Layer to Multimedia Information*. phd, Uni. of Southampton, 2000.
- [19] ISO/IEC 15938-5 FDIS Information Technology. MPEG-7 Multimedia Content Description Interface - Part 5: Multimedia Description Schemes, 2001.
- [20] M. van Assem, M.R. Menken, A.Th. Schreiber, J. Wielemaker, and B. Wielinga. A method for converting thesauri to RDF/OWL. In *Proc. of the Third Int. Semantic Web Conference, 2004*.
- [21] W3C. Semantic Web Activity. See <http://www.w3.org/2001/sw/>, May 2005.
- [22] R. Zhao and W.I. Groszky. Negotiating The Semantic Gap: From Feature Maps to Semantic Landscapes. *Pattern Recog.*, 35(3):51–58, 2002.