UNIVERSITY OF SOUTHAMPTON

Faculty of Physical and Applied Science

Electronics and Computer Science

ENHANCED IMAGE RETRIEVAL USING SPATIAL INFORMATION AND ONTOLOGIES

by

Zurina binti Muda

Thesis for the degree of Doctor of Philosophy

January 2012

UNIVERSITY OF SOUTHAMPTON

ABSTRACT

FACULTY OF PHYSICAL AND APPLIED SCIENCES
SCHOOL OF ELECTRONICS AND COMPUTER SCIENCE

Doctor of Philosophy

ENHANCED IMAGE RETRIEVAL USING SPATIAL INFORMATION AND ONTOLOGIES

by Zurina binti Muda

New approaches are essential to improve the inference of semantic relationships from low-level features for image annotation and retrieval. Current research on image annotation sometimes represents images in terms of regions and objects, but pays little attention to the spatial relationships between those regions or objects. Annotations are most frequently assigned at the global level, and even when assigned locally the extraction of relational descriptors is often neglected. To enrich the semantic description of the visual information, the use of spatial relationships offers one way to describe objects in an image more richly and often captures a relevant part of information in the image. In this thesis, new approaches for enhancing image annotation and retrieval by capturing spatial relationships between labelled objects in images are developed. Starting with an assumption of the availability of labelled objects, algorithms are developed for automatically extracting absolute object positional terms and relative terms describing the relative positions of objects in the image. Then, by using order of magnitude height information for objects in the domain of interest, relative distance of objects from the camera position in the real world are extracted, together with statements about nearness of objects to each other in the image and nearness in the real world. A knowledge-based representation is constructed using spatial and domain specific ontologies, and the system stores the asserted spatial statements about the images, which may then be used for image retrieval. The resulting Spatial Semantic Image System is evaluated using precision, recall and F-scores to test retrieval performance, and a small user trial is employed to compare the system's spatial assertions with those made by users. The approach is shown to be capable of handling effectively a wide range of queries requiring spatial information and the assertions made by the system are shown to be broadly in line with human perceptions.

Contents

List of Figures

List of Tables

DECLARATION OF AUTHORSHIP

I, Zurina binti Muda declare that the thesis entitled [Enhanced Image Retrieval Using Spatial Information And Ontologies] and the work presented in the thesis are both my own, and have been generated by me as the result of my own original research. I confirm that:

- this work was done wholly or mainly while in candidature for a research degree at this University;
- where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated;
- where I have consulted the published work of others, this is always clearly attributed;
- where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work;
- I have acknowledged all main sources of help;
- where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself;
- parts of this work have been published as:

1. Muda, Z. 2008. Ontological Description of Image Content Using Regions Relationships. *ESWC 2008 PhD Symposium* Tenerife, Spain.
2. Muda, Z., Lewis, P. H., Payne, T. R. & Weal, M. M. 2009. Enhanced Image Annotations Based on Spatial Information Extraction and Ontologies. *In:* IEEE International Conference On Signal & Image Processing 2009 (ICSIPA2009), Kuala Lumpur, Malaysia.: IEEE.

Signed:

Date: 31 January, 2012

Acknowledgements

First and foremost, all gratitude submitted to Allah SWT for His love and blessing bestowed on me and for making it possible for me to undertake this study and completing the thesis for the PhD.

I would like to express my utmost thank you and appreciation to my primary supervisor, Prof. Paul Lewis, for his precious ideas, consistent guidance and inspiration. My thanks and recognition also goes to my second supervisor, Dr. Terry Payne and my advisor, Dr. Mark Weal, for their ideas, constant support and vast assistance throughout my study. Their encouragements and motivations, have given me lots of opportunities to experience interesting work, to broaden my knowledge, skills and research networks via seminars and conferences, and lastly by making it possible for me to succeed and complete the PhD.

I would like to thank the Ministry of Higher Education of Malaysia and the National University of Malaysia (Universiti Kebangsaan Malaysia) for their support in funding this study.

I would also like to thank members of IAM Group, in particular, Jonathan Hare, Gian Luca and Daisy Tong for sharing their thought and invaluable assistance. And not to forget, thank you to all the participants who volunteered in the study.

Finally, I wish to express my exceptional love and appreciation to my mum and my dear family for their unconditional love and constant prayers that have given me strength and concentration in pursuing my study. And my thanks to all friends who have crossed my path during my lifetime, for their advice and encouragement.

May this study contribute and benefit my people and country.

Chapter 1

Introduction

1.1 Motivation

Rapid growth in the volume of multimedia information creates new challenges for information retrieval and sharing, and is stimulating activities on the development and application of Semantic Web technologies (Berners-Lee et al., 2001). An important component in most multimedia applications is the extraction and use of visual information. New approaches are essential to improve the inference of semantic relationships from low-level features in semantic image annotation, to improve semantic retrieval and to help to bridge the Semantic Gap in image retrieval (Hare et al., 2006).

A combination of traditional text-based and content-based approaches are still not always sufficient for dealing with the problem of image retrieval on the Web, mostly because of the problem of poor textual features. In spite of many years of research, content based image retrieval is still not an established and reliable approach and most retrieval systems rely quite heavily on text based retrieval. Unfortunately, some Web images have little or no surrounding text or associated text and often the surrounding text may be irrelevant. Sometimes the surrounding text does not describe the content of the image precisely or does not describe the image at all, which in consequence prevents the retrieval of the image through usual methods. The problem of limited collateral text needs to be solved, because if an image is embedded without any related text, textual feature extraction for the image would be impossible.

Google image search uses collateral text as a basis for searching for images and for many general image searches this can be very satisfactory, but we will see in the case study in section 1.2 that more specialised searches, for example by picture librarians, often require an ability to handle more specific queries, sometimes involving relational descriptions rarely found in collateral text. Content based image retrieval is now being used by Google Goggles for retrieval using the query by example paradigm but this does not address the types of query in our real case study where relational annotations are sometimes required.

If state of the art objects recognition could be combined with the automatic extraction of spatial relations between objects and a textual description of the relations added to the images, such a system would meet some of the needs of picture librarians. However, object recognition, although successful in limited domains, is still challenging when the number of possible objects is large, But progress in this area is being made and we make the assumption that eventually good automatic object annotation will be possible. We address specifically the problem of extracting information about spatial relations between labelled objects, making the assumption that such labelled objects are available. We use semi-automatic object labelling to achieve our starting point.

Much initial research on automatic image annotation represents images in terms of low level features, regions and sometimes object labels, but pays little attention to the spatial relationships between regions or objects. However, current annotation systems may recognise and identify a beach and an ocean in an image but fail to represent the fact that they are next to each other. The system may recognise and provide labels for an image such as *car, people, building* but fail to provide the information that the car is near and to the left of the building and the people are on the far right of the image. Spatial relationships are one way to describe regions or objects in an image and often capture more relevant parts of information in the image. Hence, to enrich the semantic description of the visual information, it is important to capture such relations. Although relatively basic, the use of spatial information in this way enriches the possibilities for semantic description of images and enhances the power and precision of queries which can be handled in automated retrieval. Therefore, a mechanism that can decompose the image into multiple regions of interest containing labelled objects and capture the spatial relationships among the objects would be a good solution, beyond basic keyword matching.

Automatic methods are highly desirable, especially methods that could automatically interpret the real semantic content of images as well as the content that determines the usefulness of them for most purposes. Manual image annotation is a tedious task and it is often difficult to provide accurate and comprehensive annotations for images. Ways to minimise human input by making the annotation process semi-automatic or fully automatic are certainly desirable. In the latter case, although automatic image annotation is an active area of research, the results often do not really satisfy the retrieval requirements and unfortunately, much initial research on image annotation has been concerned with assigning textual labels to images at the global level. Even when labels have been assigned locally to segmented regions (Tang and Lewis, 2007) or rectangular grid cells, the extraction of spatial positions and relational descriptors is often neglected.

To date, much of the research into content-based image retrieval (CBIR) has focussed on developing approaches, frameworks and systems, and a few have included research focused on spatial relationships. However, much of the research in spatial relationship extraction has been pursued without taking into consideration the benefits of integrating with an ontology. Such integration could allow image annotation of a region or object’s content to be linked with concepts defined within the selected domain ontology, generating more knowledge for the extraction and representation of the image whilst controlling or enhancing the range of vocabulary available. This would be valuable in enabling users to annotate regions or objects of images with better and more expressive meaning using spatial terminology during search queries and retrieval of required images, producing high level semantics and making semantic annotation systematically easier.

Ontologies can be seen as structured metadata for representing the semantics of data and play a crucial role for knowledge intensive applications. By using an ontology, classifying images and searching for them becomes easier and more relevant since it can ensure consistency in terminology and can help to disambiguate certain aspects of the spatial vocabulary. It can act as a knowledge base about domain objects, which can be used to increase the spatial information that can be extracted. We envisage the ontology not only holding synonyms for spatial terminology but also, for example, order of magnitude height information for certain objects which allow reasoning about their relative closeness to the camera/viewing position as will be demonstrated later. These developments not only make querying more flexible and powerful but can also lead to more accurate and precise query results (Srikanth et al., 2005).

This research proposes novel automatic approaches to the extraction of spatial information among objects in images, to improve the image annotation process and show how this, coupled with the use of domain ontologies, can expose additional knowledge as a part of knowledge extraction, leading to richer querying and retrieval facilities for image retrieval.

1.2 A Real Case Study

In an earlier research project ‘Bridging the Semantic Gap in Visual Information Retrieval’ (Enser et al., 2007), a large number of real queries submitted to picture librarians in a number of large national and international picture libraries were gathered and analysed.

At that time the researchers were not concerned with spatial information, but a reanalysis of the queries has revealed that a significant proportion involved spatial information. It demonstrated that spatial information is used by human searches in real queries in genuine search situations. Of the 96 queries we analysed, which were submitted to one library, 19 contained potential spatial terminology, i.e. about 20%. Not all the spatial terms were being used as spatial relationships, and it is interesting to see how these terms have been used. It demonstrates the complexity of language use in this area. Some examples of the use of spatial terms in this case study include the following: (possible spatial terms are in bold)

… coins **on** table….

… table **at left**…

… the moon **over** fields …

… bench **in middle** …

… benches **on left** …

… pictures **in** colour ….

… church **in** Paris …

… **in** any period …

… cloth dyers working **under** master…

These query fragments illustrate some conventional uses of spatial terminology but also underline a number of challenges for automated systems. First it was clear that queries articulated by humans are often at a semantically very high level. Also the spatial information in the query often relates to the spatial relations between objects in the 3D space of the real world, rather than the 2D plane of the image (e.g. 'coins **on** table'). In many cases they may be equivalent (*next to* or *above*) but in some cases the mapping is less obvious (*on* for example).

The queries also reveal the potential ambiguity of some terms. In 'working **under** master', the term *under* is used not as a spatial term but with respect to a hierarchy of roles and in the fragment 'in any period', the preposition *in* is used to indicate a temporal rather than a spatial location. However, our analysis demonstrates the potential value of the use of spatial information in human query formation and strengthens our view that the ability to support spatial terminology in automated image annotation and retrieval would be beneficial. The fact that spatial terminology may be used for purposes other than presenting spatial information supports our view that ontologies will be useful in helping to understand potentially ambiguous terminology during the process of searching and retrieval.

1.3 Research Aim

The aim of this research is to develop a new approach for enhancing image annotation and retrieval systems by capturing spatial relationships between labelled regions or objects in images through semi or fully automatic means, and supporting the process by incorporating such knowledge in a knowledge base. The Spatial Semantic Image System to be developed will be supported with a Spatial Relationships Ontology and a Place of Interest Ontology as the specific domain ontology. By this means, human users and software agents alike will be able to annotate, search and retrieve visual information in more effective and versatile ways.

1.4 Research Hypothesis

The use of spatial relationships in searching images with the specific requirement of relations between specific objects in images will improve the image annotation and performance of the image retrieval systems.

1.5 Research Questions

In order to test the hypothesis, research questions are formulated as follows:

1. What types of spatial relationships are required in annotating objects in images?
2. What spatial relations do humans use in describing images?
3. Which spatial relation terms will be of use in automatic image annotation and retrieval?
4. How can this information be extracted from the labelled segmented images?
5. How can the images be annotated with the spatial relationships?
6. How can the spatial relationship descriptors be developed and represented?
7. What ontologies are required to support the extraction and representation of spatial relationships?

1.6 Research Objectives

In order to achieve the research aim and fulfil the research questions, research objectives are identified as follow:

1. To study the state-of-the-art in the use of spatial relationships in image knowledge extraction and representation.
2. To identify and use existing models of co-occurrences of labels in regions or suitable existing annotation tools.
3. To study and choose spatial terminology that is commonly used by the user in describing the contents of images.
4. To design and implement algorithms for calculating spatial relationships based on existing labels and segmented regions in images.
5. To integrate the approach into a knowledge-base by including or evaluating relevant existing ontologies.
6. To demonstrate that these techniques could improve image search and retrieval.

1.7 Research Approach

Building on earlier work on automatic annotation and also on spatial information extraction, we are investigating more powerful approaches to annotating images automatically with spatial information by capturing the spatial relationships between labelled regions or objects in images and supporting the process with an enhanced ontology. The approach has four main stages:

1. Segmentation and initial labelling: although there is substantial research on the automatic annotation, reliable and widely applicable systems capable of annotating from a large vocabulary are not yet available. In order to generate labelled objects as a starting point for our work, we therefore use an open-source and semi-automatic labelling approach such as that provided by the LabelMe system (Russell et al., 2008). The availability of labelled image regions from this stage is assumed where output from this stage consists of region boundary information and labels indicating the objects represented by the regions.

2. Spatial information extraction: analysis of the regions and labels from the first stage is used to extract spatial information about the labelled objects. The information includes absolute spatial positions of objects, relative spatial positions and distance spatial position for pairs of objects.
3. Enhancements via the ontology: By reference to an appropriate ontology and reasoning where possible, additional spatial relations are inferred and a more diverse query vocabulary can be accommodated.
4. Experiments and evaluations on the spatial information and image retrieval performance are conducted to demonstrate the relevance and contribution of the research.

Therefore, given images with segmented and labelled regions or objects, our research aims to compute the spatial relationships among the regions or objects in the image, which will facilitate the process of retrieval in situations where the user needs to retrieve images with specific spatial requirements for objects in the image.

1.8 Research Scope

A preliminary survey was carried-out to try and identify spatial terminologies that are commonly used by people. The survey contributes to the scope of spatial concepts considered for the research, and algorithms for these spatial concepts have been developed. A leading object extraction or annotation tool, LabelMe, is used to provide initial input to our Spatial Semantic Image System. This data includes coordinates of labelled objects in an image and was computed by the spatial algorithms to generate their spatial information automatically based on a Spatial Relationships Ontology and a Place of Interest Ontology that has been developed.

An evaluation and survey of user evaluations on the ground truth with spatial relationships was performed to see how well the automated extraction of spatial relationships was achieved. The evaluation used real images and an image dataset taken from the LabelMe annotation tool to ensure the statistical significance of the results obtained. The dataset is a subset of everyday scenes such as city scenes or places of interest.

Generally, the Spatial Semantic Image system is developed for a wide range of uses. It is anticipated that users who could specifically benefit from the system might be those responsible for image collections, picture librarians, image retrieval specialists and those that work in the printing and publication domain who will make use of the system in order to obtain a specific image for their publication such as a newspaper, a magazine or a book.

1.9 Contributions

The research brings a number of novel contributions in image annotation and retrieval. The novel contributions are as follows:

1. An in depth investigation and comparison of annotation tools for annotating images.
2. The identification of some commonly used spatial terms by people in describing images through the use of questionnaire in a preliminary survey.
3. The design of a research framework as a base for developing an image annotation and retrieval, the Spatial Semantic Image System (Muda, 2008), see Appendix D.
4. The development and implementation of spatial relationships extraction algorithms to extract a range of different spatial relationship concepts including relative position, absolute position and distance position. Parts have been presented in IEEE International Conference On Signal & Image Processing (Muda et al., 2009), see Appendix E .
5. The development of two ontologies: the Spatial Relationships Ontology (application) and the Place of Interest Ontology (domain) to handle the expressivity of the spatial terms and concepts.
6. Demonstrations of the reliability of the system in identifying spatial terms in comparison to human manual identification.
7. An evaluation of retrieval performance showing the improvements which the system brings, particularly in terms of retrieval precision.

1.10 Thesis Structures

The structure of the remainder of this thesis is organised as follows:
Chapter 2, Literature Review, discusses the relevant literature on Semantic Web technologies, MPEG-7 standards and multimedia ontologies; and explores research on automatic image annotation and spatial relationships in images.

Chapter 3, Image Annotation Tools And Research Framework, discusses an investigation performed on existing image annotations tools, which includes Caliph & Emir, Photostuff, AKTive Media, M-OntoMat-Annotizer, and LabelMe. A comparative study is conducted and an evaluation of results is presented. From this, a research framework is developed.

Chapter 4, Choosing Spatial Terms, discusses the work done including the implementation of an online web-based survey, in order to identify spatial relationships terms that are commonly used by the user, and to select a set of specific spatial terms for further experiments and development. The results obtained and initial findings of the survey are illustrated and presented.

Chapter 5, The Development of Spatial Relationships Algorithms, describes the design and implementation of spatial relationship algorithms for relative and absolute position. An example considering objects in an image has been used to demonstrate the implementation with results and discussions.

Chapter 6, Advanced Spatial Relationships, describes the design and implementation of spatial relationship algorithms for relative distance from the camera based on a statistical analysis. A number of distance position cases involved are discussed and tested with series of real-life scenarios images. The implementation is also tested on a sample of images. Results and discussion are presented.

Chapter 7, The Spatial Relationships and Domain Ontologies, discusses the design and implementation of a Spatial Relationships Ontology as an application ontology and a Place of Interest Ontology as a domain ontology for the whole system. Some extensions to the Spatial Relationships are discussed.

Chapter 8, Integration and Evaluation of the Spatial Semantic Image System, presents the integration of the whole Spatial Semantic Image System and evaluations of the system through two major experiments: a user evaluation survey and image retrieval performance tests. Each experiment comprises methodology, results, analysis and discussion.

Chapter 9, Conclusion And Future Work concludes the research presented in all previous chapters and provides suggestions for future work.

Chapter 2

Literature Review

2.1 Introduction

This chapter discusses relevant state-of-the-art literature on Semantic Web technologies, the MPEG-7 standard, Multimedia Ontologies, automatic image annotation and spatial relationships. The literature review underpins the research described in the later chapters. Each section includes a brief discussion on how the topics discussed are directly associated to this research.

2.2 Semantic Web Technologies

The Semantic Web increases the ability to make Web resources accessible by their semantic content since information is given well-defined meaning, in a systematic standard format (Fensel et al., 2002); enabling computer-human cooperation in distributed computing environments (Uschold, 2003). The Semantic Web is an evolving extension of the World Wide Web in which Web content can be expressed not only in natural language, but also in a format that can be read and used by software agents, thus permitting them to find, share and integrate information more easily (Herman, 2001a).

The Semantic Web is comprised of a philosophy, a set of design principles, collaborative working groups and a variety of enabling technologies (Wikipedia, 2008e). If properly realised, it can assist the evolution of human knowledge as a whole (Berners-Lee et al., 2001). The Semantic Web is intended to provide machines with much better (automated) information access, based on the semantics of data and heuristics that use this metadata as intermediaries in support of humans (Fensel et al., 2002) by providing a common framework that allows data to be shared and reused across applications, enterprises and community boundaries. (Herman, 2001a, b). It is a collaborative effort led by W3C with participation from a large number of researchers and industrial partners (Hawke et al., 2011).

The semantic layer cake or stacks (Burleson, 2007) shown in Figure 2-1 illustrates the Semantic Web key enabling technologies. The Web Ontology Language describes the function of each of the Semantic Web's key enabling technologies as follows (Wikipedia, 2008e):

1. XML is classified as an extensible language because it lets everyone create their own tags, hidden labels to annotate web pages. Its primary purpose is to facilitate the sharing of structured data across different information systems, particularly via the Internet (Wikipedia, 2008c, a). XML Schema is a language for providing and restricting the structure and content of elements contained within XML documents (Wikipedia, 2008d).

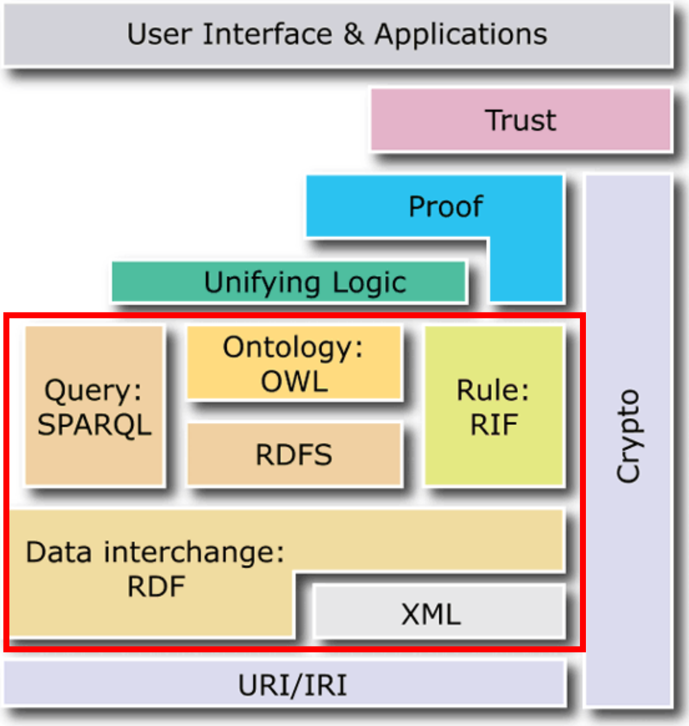


Figure 2-1 Semantic Web Layer Cake (Burleson, 2007)

2. RDF is a simple language for expressing data models, which are encoded through sets of triples. Each triple is rather like the subject, verb and object of an elementary sentence and can be written using XML tags. Subject and object are each identified by a Universal Resources Identifier. RDF Schema is a vocabulary for describing properties and classes of RDF-based resources.
3. OWL has been designed to meet the need for a Web Ontology Language (Wikipedia, 2008b). OWL adds more vocabulary for describing properties and classes, among others: relations between classes, cardinality, equality, richer typing of properties, characteristics of properties and enumerated classes.

The red line in Figure 2-1 shows parts of the layer cake covering key enabling technologies potentially relevant to this research. We may have used existing annotations or tags in XML or create new annotations in XML. The rules technology is relevant for reasoning over spatial concepts, RDF may be used to encode spatial triplestores and OWL is useful as a candidate language for a spatial ontology.

2.3 The MPEG-7 Standard

The MPEG-7 (Multimedia Content Description Interface) standard is historically important in managing and handling multimedia content such as visual, image, audio, audio-visual and video. The goal of MPEG-7 is to support the requirements for providing a rich set of standardized tools to enable the generation of multimedia descriptions which can be understood by machines as well as humans (Martínez et al., 2002). It enables fast and efficient retrieval from digital archives (pull applications) as well as filtering of streamed audiovisual broadcasts on the Internet (push applications).

The standard represents information about the content to allow searching for material that is of interest to the user, and operates in both real-time and non real-time environments (Hunter, 1999b, a). Research by Hunter (2001) and Tsinaraki et al., (2005) and projects such as HARMONY and DICEMAN (Hunter, 1999a), were carried out either to adopt or to enhance the capability of the standard. The standard potentially helps in bridging the semantic web by linking to low level descriptions with high level annotations.

Martinez, et al. (2002) point out that MPEG-7 provides the richest multimedia description tools for content management, organization, navigation and automated processing, but there are a few drawbacks to MPEG-7. Hunter (2001) has attempted to model parts of MPEG-7 in RDFS before, later integrating it with the ABC ontology model (Lagoze and Hunter, 2003). Based on the visual part of MPEG-7, an OWL DL Visual Descriptor (VDO) has been proposed by Simou, et al. (2005) for image and video analysis.

2.4 Multimedia Ontologies

Ontologies play an important role for knowledge intensive applications and can be seen as metadata descriptors that formally define terms and explicitly represent the semantics of data that it weaves together in a net, linking and communicating human knowledge and complementing it with machine processability (Ding, 2002). They aim to capture domain knowledge in a generic way and provide a commonly agreed understanding of a domain to be reused, shared and operationalized across applications and groups (Sure et al., 2002).

An ontology consists primarily of concepts and the relationships between them. A highly cited definition is:

> *"an ontology is a formal, explicit specification of a shared conceptualization. 'Conceptualization' refers to an abstract model of phenomena in the world by having identified the relevant concepts of those phenomena. 'Explicit' means that the type of concepts used, and the constraints on their use are explicitly defined. 'Formal' refers to the fact that the ontology should be machine readable. 'Shared' reflects that ontology should capture consensual knowledge accepted by the communities".*
>
> *(Gruber, 1993)*

In the field of semantic image understanding, using a multimedia ontology infrastructure is regarded to be the first step for closing the so-called, semantic gap

(Arndt et al., 2007a). A multimedia ontology has the potential to increase application interoperability, express concepts in multiple media formats, provide cross-modal relationships to support reasoning (Srikanth et al., 2005) and consuming multimedia annotations (Arndt et al., 2007b).

2.4.1 The DOLCE Ontology

The Descriptive Ontology for Linguistic and Cognitive Engineering (DOLCE) (Petridis et al., 2006) is intended to act as a starting point for comparing and elucidating the relationships with other ontologies of the WonderWeb library and for clarifying the hidden assumptions underlying existing ontologies or linguistic resources such as WordNet (Oberle et al., 2007). DOLCE is based on the fundamental distinction between endurant (i.e., objects or substances) and perdurant (i.e., events or processes) entities, with relation as participation. Spatial locations and temporal qualities encode the spatio-temporal attributes of objects or events (Oberle et al., 2007).

DOLCE is conceptually sound, and thus ideally suited for reference purposes and its features are suited for modularization and rich reference axiomatization that captures ontology design patterns such as location in space and time, dependence or part-hood. Extending or integrating DOLCE with other ontology could overcome its abstract nature and produced some advantages, because DOLCE is ideally used as a foundational ontology. It has been successfully applied in different domains, such as law, biomedicine and agriculture (Oberle et al., 2007).

2.4.2 The COMM Ontology

COMM (*a Core Ontology for Multimedia*) has been developed based on DOLCE (Oberle et al., 2007). COMM combined the advantages of extensibility and scalability of web-based solutions and solved the interoperability problem of the MPEG-7 standard for representing the metadata of all relevant multimedia objects. This ontology was designed using sound design principles (Arndt et al., 2007b), advocates the use of formal semantics, and is grounded based on an ontology development methodology, in order to describe the required multimedia semantics in terms of current semantic web languages.

The COMM ontology has satisfied the requirements of reusability, MPEG-7 compliances, extensibility, modularity and a high degree axiomatization, as described by the multimedia community for a multimedia ontology framework. The ontology modelling approach offered more possibilities for multimedia annotation than MPEG-7 since it is interoperable with existing web/domain ontologies. The evaluation of the ontology, its scalability and its adequacy in the implementation of tools are improved to be used for multimedia annotation, analysis and reasoning in large scale applications (Staab, 2007). As shown in Figure 2-2, COMM consists of:

1. Core module that contains the design patterns;
2. Modules that specialize the core module for different media types;
3. Modules that contain media independent MPEG-7 description tools such as media information or creation and production;
4. Data type module that formalizes MPEG-7 data types e.g. matrices and vectors.

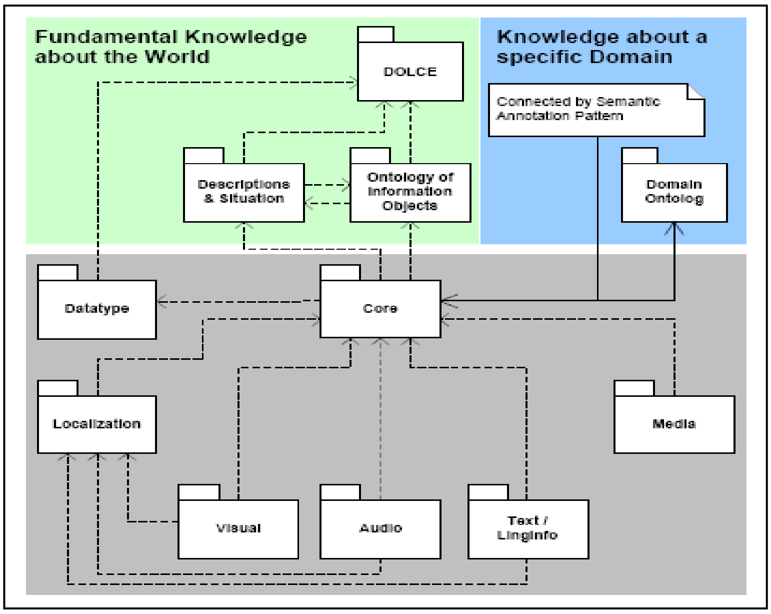

Fundamental Knowledge about the World
DOLCE
Knowledge about a specific Domain
Connected by Semantic Annotation Pattern
Descriptions & Situation
Ontology of Information Objects
Domain Ontolog
Datatype
Core
Localization
Media
Visual
Audio
Text / LingInfo

Figure 2-2 Multimedia Ontology (COMM)

(Taken from Staab (2007))

2.4.3 Ontology Engineering Environment

An ontology engineering environment is a construction tool used to develop ontologies. In recent years, research has aimed at paving the way for the construction of ontologies by ontology development environments (Duineveld et al., 2000). The existing tools included (Chebotko et al., 2005):

1. Protégé is a popular ontology construction and annotation tool developed at Stanford University that supports the Web Ontology Language through the OWL plug-in.
2. OntoEdit is an ontology editor that integrates numerous aspects of ontology engineering by combining recent methodology-based ontology development with capabilities for collaboration and inferencing comprehensively. In particular, OntoEdit focused on three main steps for ontology development (Staab et al., 2001), which covered visualising requirements specification, refinement and evaluation.

2.4.4 Discussion

As a conclusion, on top of any development of an image ontological description there may be a need for multimedia ontology to be considered. DOLCE covers some of the spatial relations which are involved with spatial locations and temporal concepts. These concepts were adopted by COMM in its spatialrel ontology. Hence, for this research, initially the DOLCE and COMM would be acceptable with reference to the immediate needs of the research, in deciding whether to use or extend the existing ontology. If extension is needed, additional concepts and descriptors for spatial relationships will need to be inserted. If a new spatial relationships ontology needs to be developed for spatial annotation and retrieval in this research, Protégé is a suitable construction tool to be employed as it is a versatile and extensible ontology editor allowing exports in a wide range of formats such as RDF, XML and OWL

2.5 Automatic Image Annotation

Automatic image annotation is the unsupervised task of adding descriptive textual terms to an image to provide direct access to the semantics. A recent review has been published. (Zhang et al., 2012). Annotation aims to add descriptive labels to regions or objects in an image or to the image as a whole in order to represent the semantic content, usually as an intermediate step to image retrieval (Smith and Chang, 1996). Automatic annotation helps to bridge the semantic gap (Ossenbruggen et al., 2001, Hare et al., 2006), by producing object labels or keyword annotations or text information which are closer to the high level semantic descriptions needed for better image retrieval (Chen et al., 2001, Sure et al., 2002).

Most content-based image retrieval (CBIR) systems have relatively poor classification performance since low-level visual features cannot easily represent the high-level semantic content of images (Srikanth et al., 2005). Hence most research in automatic image annotation focuses on inferring high-level semantic information from low-level image features such as colour, texture, shape and more recently salient regions using representations such as the SIFT (Lowe, 2004). Although Enser, et al. (2005) have pointed out some limitations of automatic image annotation, with the explosive growth of images on the Web, there has been an increasing need for tools to automatically annotate and organize image collections. As a result, research into image annotation has been an active area in recent years. The field includes research on explicit visual object classification operating typically at the local level within the image and also more global approaches, which assign labels at the global image level. One of the main limitations of automatic annotation techniques at present is the relatively small number of objects or image labels which can be assigned, i.e. the size of the vocabulary available to the annotator. In Tsai and Hung's review 0f 2008 (Tsai and Hung, 2008) the largest vocabulary size encountered was 375 but in the Pascal Challenges reported in Everingham 2010, the maximum Vocabulary Size is 20 object classes (Everingham et al., 2010). However, Nister and Strewenius (2006) have proposed an object classification scheme which they argue will scale to large numbers of objects and the intensity of research suggests that more useful vocabulary sizes will be achievable in the near future.

In image annotation and retrieval, images are represented by low level features. These basic visual features can include colour, texture and shape features or salient regions or interest regions represented by colour or texture or the SIFT feature mentioned earlier.

Most image annotation models have used the co-occurrence of image regions and words to model the association between words and images, or words and image regions (Chakravarthy et al., 2006c, Tsai and Hung, 2008, Zhang et al., 2012). Others have explored machine learning approaches to overcome some of the automatic image annotation problems. Image regions can be generated using image (Bashir and Khan, 2004) or object (Everingham et al., 2010) recognition techniques or image segmentation techniques like N-cuts and grids. With an annotated training set of images, models for image annotation can learn from the co-occurrence of words and images, or image regions. In some cases, a grid-based segmentation method is used to identify image regions and the co-occurrence of words to predict image annotations (Srikanth et al., 2005). This approach has also been used in other domain such as in medical imaging (Specovius et al., 2010).

Manual annotation of objects to generate training sets for automatic annotation is a tedious task and making this task interactive and also fun has been attempted through online computer games like the ESP game, and has been used for image-level annotation of real images (Ahn and Dabbish, 2004). The annotation process ensures that only correct labels are assigned to images. The game produced considerably large volumes of data. However, images are annotated at global-level and the data is not readily available. Another annotation game that uses the annotated images generated with the ESP game is Peekaboom (Ahn et al., 2006). The game provides objects locations and geometric labels from players' activities. The resulting collection could be used to train computer vision algorithms for a variety of tasks, including region-based level for automatic image annotation. However, since the number of annotators can be of the range of millions, there is not an objective criterion to obtain concise object localizations. Other types of game also have been developed. Curator is a class of games with a purpose for building collections to help researchers develop guidelines for collection recommender systems among other applications (Walsh and Golbeck, 2010).

Region-based image retrieval systems retrieved images based on region information (Carson et al., 1999, Ko et al., 2000, Zhou et al., 2008). Since humans are accustomed to utilizing object-level concepts (e.g. car) rather than region-level concepts (e.g., windscreen, shadow and wheel), object-based content analysis is a more reasonable approach. Therefore, it is necessary to integrate related regions into an object in order to provide a better object query environment to the user.

2.5.1 Exploiting Ontologies

The use of text ontologies as a basis for defining visual vocabulary and classification of high-level abstractions increases the number of concepts that can be utilised by an image annotation system for a given image.

Srikanth et al., (2005) used hierarchical dependencies between annotation words to generate improved visual lexicons for the translation-based approaches by exploiting ontological relationships between annotation words and demonstrated their effect on automatic image annotation and retrieval. The study used a hierarchy to capture visual similarities among different cats from an ontological resource in the WordNet, which organizes different animals in a hierarchy and placed cougar, leopard, etc. under cat. The hierarchy helped to induce the annotation words for automatic image annotation, thus supporting multimedia information retrieval (Srikanth et al., 2005).

Kallergi et al. (2009) developed search facilities across the life sciences ontology collection and implemented a new graphical ontology viewer. This tool allows for both querying and visualizing ontology terms by means of a 2D graph representation.

2.5.2 Web Image Retrieval

Cai et al., (2004) exploited visual, textual and link information to hierarchically cluster Web image search results. By exploiting link information, the inter-relationships between Web images and their textual annotations could be explored to improve Web image retrieval (Wang, 2003). As an example, a Web image retrieval system called PicASHOW (Lempel and Soffer, 2002), was based on several link analysis algorithms and could retrieve relevant images even when those are stored in files with meaningless names.

A collective classification model called relational support vector classifier was proposed based on the well-known SVM and the linkage semantic kernels (Yong-hong et al., 2005). The approach was implemented in a Web image classification prototype called ConWic (*Context-Based Web Image Classification & Clustering*) system. The ConWic system was developed to exploit visual, textual and relational information to aid classification, clustering and semantic-sensitive retrieval of Web images. The experiment on a sports Web image collection crawled from the Yahoo Sport site, showed that it achieved significant improvement in classification accuracy over SVM classifiers using visual and/or textual features (Yong-hong et al., 2005).

Bashir and Khan (2004) proposed a Web system that could populate itself by searching for keywords and subsequently retrieving images or articles. Such a system requires solid interoperability, a central ontology and semantic agent search capabilities. They presented a semi-automatic image annotation by decomposing an image into classifications of low-level or atomic concepts (for example: ball and net) by using SVM; and classification of high-level semantic concepts in a domain-specific ontology. For example, an image that contains a ball, a net and humans can be described as a basketball game by using Bayesian belief networks (Benitez and Chang, 2002, Bashir and Khan, 2004). Upon classifying the image, the system reflected the image semantics, its features, content and semantic category, as part of a semantic space. Web content distance also has been measured and utilized to address and reduced web content clustering problems where an image is associated with the textual contents of the cluster it belongs to (Alcic and Conrad, 2011).

An important effort is being carried out by Russell et al. (2008) with the purpose of creating a benchmark collections for diverse computer vision applications. The LabelME project uses a web-based online tool for segmenting and annotating images (Russell et al., 2008). Segmentations are specified and annotations defined by the user by drawing polygons around each object. This provides annotations at the local-level and has the advantage that there is a significant collection of locally annotated images which are publicly available.

Currently, Flickr is an example of a popular Web 2.0 website for online photography management applications that provides a means for photo storing, publishing, sharing and searching (Flickr, 2011). It provides an interactive environment for users to create their photo stream or album, classify images based on their interest and annotate the images by tagging title, caption, etc. It also allows other members with the same interest to be invited to provide additional information to the image with more tags, write feedback or comments. Annotations are an open tag in the form of simple words assigned at the global level, but it also provides a rectangle box to be used in identifying people in the image. However this feature is manually done and can be used by the user not just to recognise a person but also to add a note etc. For our purposes, the problem with this collection is that it has an open vocabulary, objects are normally assigned at global-level and at local-level only a specified rectangle box is allowed.

2.5.3 Discussion

Object annotation is an essential intermediate step for computers to capture and represent the objects or features contained in an image, before proceeding to capture other features such as spatial relationships. Hence this research will investigate further and explore the capabilities of current tools for extracting features such as regions or objects, and investigate the best available tools to be utilized as a starting point for generating and capturing spatial relationships between regions or objects in images. However, initially, it is quite clear from the literature review discussed above that LabelMe offers one of the best potential tools to be used. As mentioned, the LabelMe benchmark collection contains image annotations at the local-level and is publicly available. Though this tool is only semi-automatic, this is not our main concern as we are not addressing the annotation of the objects per se. It is the relations between them and their spatial positions that we are going to address. Therefore by specifically addressing the issue of annotating with spatial relations and in order to ensure accurate ground truth as a starting point for our spatial relation extraction it is seen as more reliable to use a facility such as LabelMe as it gives high accuracy compared to other image online sets as mentioned in a study done by Renn et al. (2010).

2.6 Spatial Relationships

The need for an efficient technique to store and retrieve images automatically based on their content is really essential as this will speed the storing and retrieving process while enhancing the retrieval performance. In general, an image is retrieved either by high-level semantics, which define image content at the conceptual level, or by visual characteristics, which are based on perceptual features such as colour, texture, structure, shape (Rong and Grosky, 2002, Wang, 2003, Wang et al., 2004), regions or objects in the image as in the QBIC project (Flickner et al., 1995). They can also be retrieved by their spatial relationships or by relative position of the icons (Zhou et al., 2006) or symbolic objects (Hoang et al., 2010) or as in a 2D String (Chang et al., 1987, Lee and Hsu, 1990, Lee and Chiu, 2003).

The spatial relationships are often considered to be fuzzy concepts and usually depend on human interpretation. There are two common kinds of representation of spatial relationship features:

1. Topological relationships have been applied to Geographical Information System (GIS) due to their invariance under topological transformation. Such work has been done by (Egenhofer and Franzosa, 1991).
2. Orientation, or directional relationships, concern partial and total orientation relationships among objects, describing where objects are placed relative to one another. This is more useful in image databases (Zhou et al., 2001) than topological relationships.

Zhou et al., (2001) discussed orientation relationships by focusing on transformations. The transformations consider a primary object, a reference object and a frame of reference (Hernandez, 1994). Ahmad and Grosky (2003) proposed a symbolic image representation and indexing scheme to support retrieval of domain independent spatially similar images and considered directional relations (quadrant) based on compass directions including North-East, North-West, South-East and South-West.

Hollink et al., (2004) categorised the relationships into 8 spatial relations by considering right, left, above, below, near, far, contain and next; and 9 absolute position including centre, North, South, East, West, North-East, North-West, South-East and South-West. Lee et al., (2006) present a unified representations of spatial objects for 4 topological relations and 8 directional relationships. Yuan et al., (2007) consider neighbouring relationships based on above, below, left and right.

2.6.1 Spatial Similarity-Based Retrieval

Spatial similarity-based retrieval is an important class of content-based image retrieval (Grosky and Mehrotra, 1989) and has generated a great deal of interest. The concept of similarity or approximate match is implemented to accommodate natural inconsistency during searching and retrieval. Determining similarity according to spatial relationships is generally complex and might be as difficult as semantic object or region-level in image segmentation. Many studies have developed similarity-matching algorithms to capture spatial and multiple region information in an image. For example, Zhou et al., (2001) used a similarity retrieval approach for augmented orientation spatial relationships representation to capture rotation invariant, relative distance and orientation range between symbolic objects by overcoming the ambiguity problems in other orientation representations.

Gonzalez and Reyes (2011) proposed a graph matching scheme involved colour, texture and shape features with spatial descriptors to represent topological and orientation relationships, that are obtained by means of combinatorial pyramids. A spatial similarity is measured to test the similarity between spatial features and graph matching scheme to compute the overall similarity between objects. Evaluation on COIL-100 and ETH-80 images sets proved that the combination of visual and spatial features is a promising road in order to improve the object recognition task.

- Region/Object-Based with Spatial Relationships

Tian et al., (2000) used spatial layout combined with user defined ROI (region/s of interest) (Moghaddam et al., 2001) to present the content of an image. Li et al., (2000) presented Integrated Region Matching based on spatial relationships between regions by allowing a similarity measure for regions based on image similarity comparison, while Smith and Chang (1999) decomposed the image into regions and represented those regions as strings.

Lee and Hwang (2002) proposed a domain-independent spatial similarity and annotation-based image retrieval system that decomposed the image into multiple regions of interest containing objects and allowed the user to formulate a query based on both objects of an image and their spatial relationships. The study has improved the current spatial analysis technique and the ROI representation scheme. Ko and Byun (2002) used the Hausdorff Distance to estimate spatial relationships between regions as part of their FRIP (Finding Region In the Pictures) system and named this system as Integrated FRIP (IFRIP) (Ko et al., 2000). IFRIP also incorporates relevance feedback in order to reflect the users' high-level and subjective query.

Dinesh and Guru (2011) proposed a method for recognizing partially occluded objects where corner points and their spatial relationships were used to be perceived through the application of Triangular Spatial Relationships (TSR). The perceived TSR is then used to create model object database using B-tree, an efficient multilevel indexing structure. The TSR was also used by (Hoang et al., 2010) for scene retrieval.

Wu et al. (2010) proposed an object categorization model with implicit local spatial relationship based on bag-of-words model. The model use neighbour features of one local feature as its implicit local spatial relationship integrated with its appearance feature to form two sources of information for object categorization. The algorithm is applied in Caltech-101 and Caltech-256 datasets to validate its efficiency. Yi et al. (2009) proposed a cognitive representation and Bayesian model for spatial relationship among objects to estimate the location of a robot in order to allow the robot navigated in an indoor environment. The experiment results showed that the location accuracy is improved even inaccurate sensors such as a consumer-grade camera is used.

- Symbolic Images and Quad-tree

In spatial similarity-based retrieval, to improve the efficiency of the search and retrieval process, abstract or symbolic images have been used by Chang et al., (1986), Chang et al., (1989), Gudivada (1994) and Hoang et al. (2010). Beeson et al (2010) used a symbolic descriptions for map-building, and Santosh et al. (2010) addressed the use of unified spatial relations for symbol description. The approach has an ability to express spatial relations between any numbers of components and have been used in symbol retrieval application.

Ahmad and Grosky (2003) proposed a symbolic image representation and indexing scheme to support retrieval of domain independent spatially similar images. This scheme used a Quad-tree to manage the concept of hierarchical decomposition of an image into a spatial arrangement of distinct features. While Carson et al., (1997) used a Quad-tree in their region based image querying system to obtain homogeneous clusters. The spatial positions of these regions are modelled using 2D strings and spatial relations.

- Spatial Relationship by 2D String

Similarity retrieval by using 2D Strings requires massive geometric computation and focuses on those database images that consist of icons. Chang et al. (1987) developed the concept of iconic indexing by introducing the 2D string representation of an image to present spatial relationships between symbols. Subsequently this approach has been extended to 2D-H string, 2D-PIR graph (Nabil et al., 1996), 2D-Z string (Lee and Chiu, 2003) and 2D Be-string(Wang, 2003). Based on previous research in 2D String (Lee and Hsu, 1990, Lee and Chiu, 2003), Wang (2003) proposed the 2D Be-string (two dimension begin-end boundary string) model to represent an icon by its boundaries and evaluates image similarities based on the modified ‘‘longest common subsequence” algorithm. The model solved the problems of uncertainty in query targets and/or spatial relationships, and simplifies the retrieval progress of linear transformations, including rotation and reflection of images.

- Minimal Spatial Relationships

Lee, et al. (2006) suggested the use of minimal 3D relationships in the specification of query images in the content-based retrieval of 2D images, in order to tackle the problem of costly storage (Chang et al., 1987) and image ambiguities. They proposed a unified representation of spatial relationships among image objects and a set of reduction rules to minimize these relationships based on Allen’s temporal interval algebra (Allen, 1983). This strategy requires a generalized spatial representation scheme, which handles stored spatial knowledge and computes additional spatial relationships easily by a spatial reasoning engine as well.

2.6.2 Spatial within Context Constraint

Rather than using scene context, Fan et al., (2004) and Yuan et al. (2007) represented the spatial context constraints in various graphical models by relating learning and inference algorithms. They investigated how to combine the classification performance of discriminative learning and the representation capability of graphical models in the scenario of image region annotation. The experiments were the largest scale evaluation for region annotation in supervised learning setting and could provide a useful guide for building real-world systems (Yuan et al., 2007). Other models used to exploit spatial context constraints for tasks similar to region annotation include 2D Hidden Markov Models (Li and Gray, 2000), Markov Random Fields and Conditional Random Fields (Li and Wang, 2003).

2.6.3 Dynamic Interactive Spatial Querying

Interactive similarity retrieval is used to resolve the fuzzy area involving psychological and physiological factors of individuals during the retrieval process (Ishikawa et al., 1998, Yong et al., 1998, Bartolini et al., 2001). Thus, Lee et al. (2006) proposed a dynamic similarity measure approach based on an enhanced digraph structure for interactive spatial similarity retrieval to help users navigate in an iconic image database more intuitively. The approach can be applied to any image retrieval algorithms and made use of multiple feedbacks from the users to get the hidden subjective information during the retrieval process, thus avoiding the high cost of re-computation of an interactive retrieval algorithm.

Previously, similar approaches such as FeedbackBypass (Bartolini et al., 2001) and query refinement were used to reduce duplicated computation, while Yong, et al. (1998) used an indexing structure and a dynamic measure on top of the index structure to extract information from user feedback. Compared to Mindreader (Ishikawa et al., 1998), this solution retains the objectiveness of the existing similarity index and measure, and makes use of the subjective information of the users' feedback in an objective way.

2.6.4 Discussion

Based on the previous research in spatial relationships, this research will include the representation of spatial relationships with an emphasis on orientation relationships. We are attempting not to be as rigid as the spatial similarity-based retrieval approach or within a context constraint, in order to make our approach more flexible by using regions or objects in spatial annotations.

Studies done by Hollink et al. (2004), Lee at al. (2006) and Yuan et al. (2007) are closely related and work in this report begins by building and extending this previous work. The method will identify and define concepts of spatial relationships to be considered and then proceed with the development of algorithms to compute these concepts in spatial context constraints in order to enhance the capability of the proposed tool as well as for meeting the needs and requirements of users.

2.7 Conclusion

The literature review of the topics related to this research has been discussed in this chapter with details on spatial relationships to express the state-of-the-art in the subject. The importance of the topics and their future use in this research has been discussed. From the automatic image annotation section, an investigation has been conducted to investigate current image annotation tools, which is described in detail in Chapter 3.

As for the spatial relationships, research done in the area has suggested a number of spatial terms to be considered, and an attempt has been made to look into this issue by developing an online image description survey for identifying and choosing spatial terms that are commonly used by users as discussed in Chapter 4, then developing and implementing the selected spatial relationships algorithms as explained in more detail in Chapter 5 and some extended algorithms in Chapter 6.

Chapter 3

Image Annotation Tools and Research Framework

3.1 Introduction

The use of image annotation has become significant in facilitating extraction, labelling, organizing and storing of visual information in an effort to improve image retrieval and the multimedia retrieval systems. One way to annotate an image locally is by segmenting the image manually or automatically. There is an increasing need for a tool that could annotate segmented regions or objects and provide annotation with additional knowledge such as spatial relationships supported by ontologies. As a part of the preliminary investigation, five existing tools for image annotation are discussed. Three of them: Caliph & Emir, PhotoStuff and AKTive Media, are listed in the W3C Multimedia Semantic Incubator (Burger et al., 2007), M-OntoMat-Annotizer, was developed under the AceMedia project (Akrivas et al., 2007) and LabelMe, was developed at MIT Laboratory (Russell et al., 2008).

Each of these tools has been investigated individually using a dataset of images and by a comparative study based on an evaluation framework adapted from Lewis (1995) and Duineveld et al., (2000). The study investigated image annotation features with a number of aims:

- To discover whether these image annotation tools include spatial relationships in the annotations.
- To potentially identify and select an annotation tool that annotates images locally by producing annotations of segmented regions or objects.
- To understand how the labelled regions or objects have been annotated in the tool.
- To obtain the object or region annotations for further use in this research, where spatial relationships could be augmented with spatial relationship annotations.

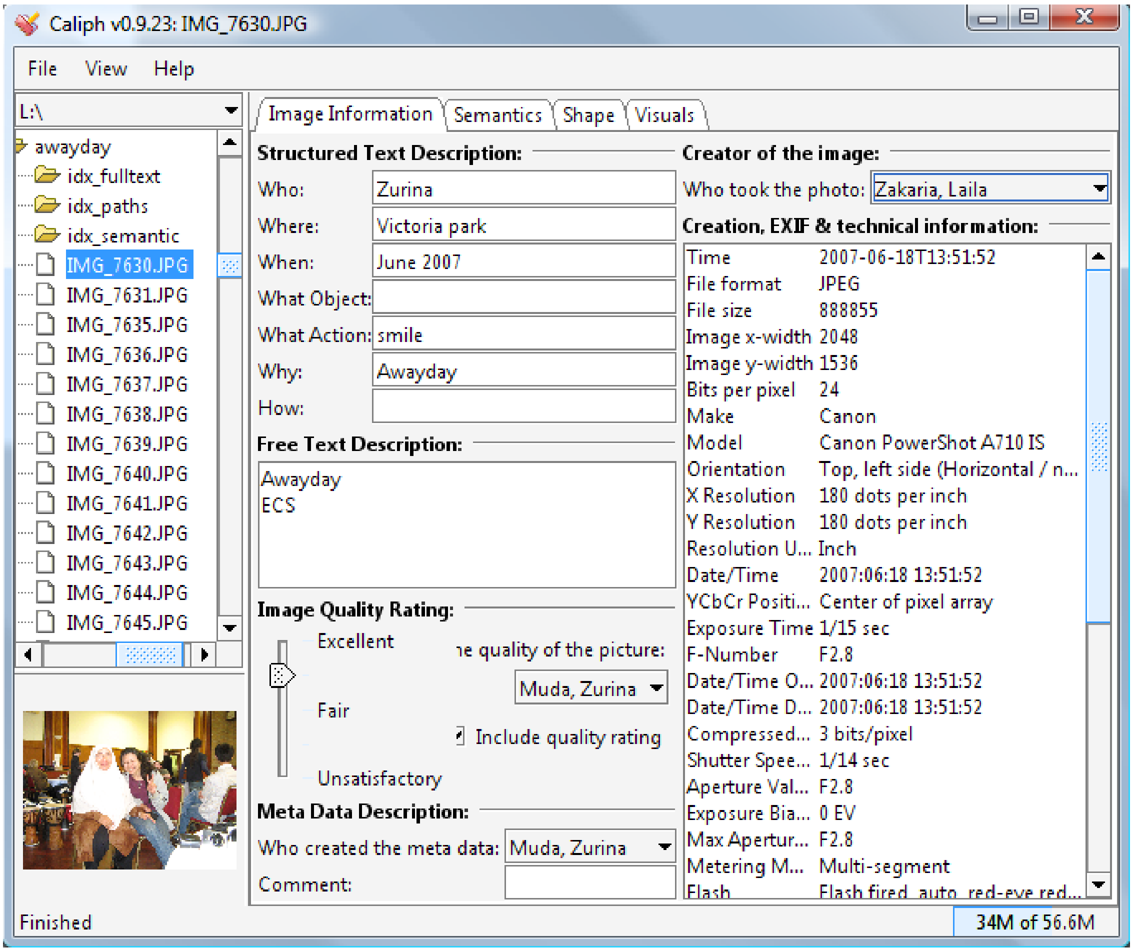
The study also evaluates the user interface components with a view to selecting one as the base technology for further experimentation, in order to provide more substantial annotations and hence help improving current image retrieval systems. From this study a comprehensive research framework is suggested and developed where spatial relationship annotation has been incorporated together with support from ontologies.

3.2 Caliph & Emir

Caliph & Emir, are a pair of applications that use MPEG-7 descriptors for image annotation and search of digital photos focusing on semantic metadata and content based image retrieval (Lux, 2009). Caliph & Emir were implemented using JAVA. Caliph & Emir[1] are research products developed by Know-Center and Joanneum Research at the University of Technology Graz, Austria. Figure 3-1 show the main interfaces of Caliph during annotations of “Awayday” photos. Figure 3-2 shows the Emir interface when searching for an image labelled with “Victoria Park”.

Caliph (*Common And Lightweight Interactive PHoto annotation*) was designed for supporting users in the time consuming task of annotation by allowing them to annotate digital photos manually, and extracting content based on low-level features from the image automatically. Emir (*Experimental Metadata-based Image Retrieval*) allows the retrieval of digital photos based on annotations created with Caliph (Lux et al., 2004).

[1] http://www.semanticmetadata.net.



Structured Text Description:

Who: Zurina
Where: Victoria park
When: June 2007
What Object:
What Action: smile
Why: Awayday
How:

Free Text Description:

Awayday
ECS

Image Quality Rating:

ıe quality of the picture: Muda, Zurina
☑ Include quality rating

Meta Data Description:

Who created the meta data: Muda, Zurina
Comment:

Creator of the image:

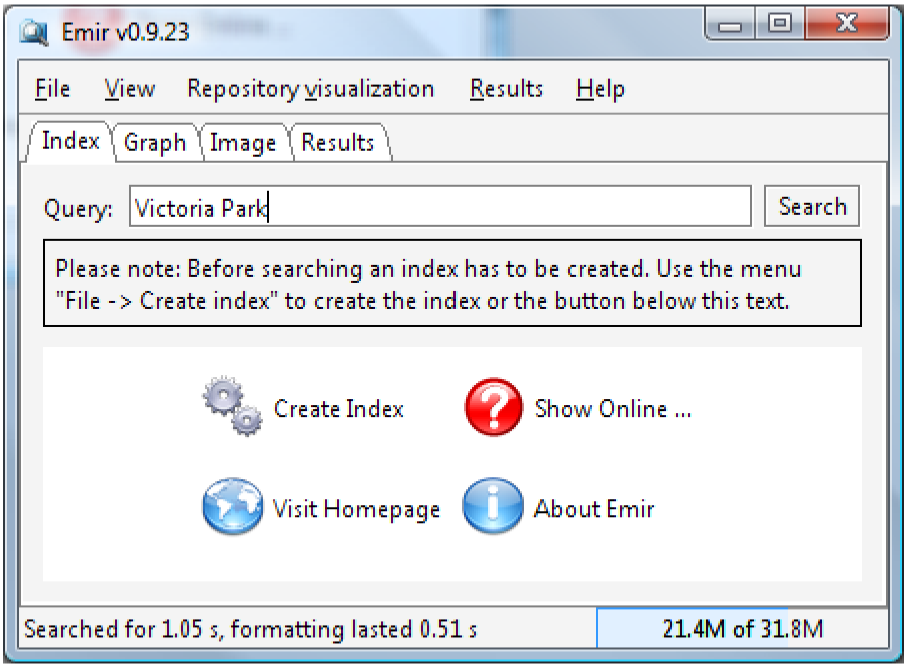
Who took the photo: Zakaria, Laila

Creation, EXIF & technical information:

```
Time            2007-06-18T13:51:52
File format     JPEG
File size       888855
Image x-width   2048
Image y-width   1536
Bits per pixel  24
Make            Canon
Model           Canon PowerShot A710 IS
Orientation     Top, left side (Horizontal / n...
X Resolution    180 dots per inch
Y Resolution    180 dots per inch
Resolution U... Inch
Date/Time       2007:06:18 13:51:52
YCbCr Positi... Center of pixel array
Exposure Time   1/15 sec
F-Number        F2.8
Date/Time O...  2007:06:18 13:51:52
Date/Time D...  2007:06:18 13:51:52
Compressed...   3 bits/pixel
Shutter Spee... 1/14 sec
Aperture Val... F2.8
Exposure Bia... 0 EV
Max Apertur...  F2.8
Metering M...   Multi-segment
Flash           Flash fired, auto, red-eye red...
```

Finished 34M of 56.6M

Figure 3-1 Caliph Annotation Interface



Query: Victoria Park

Please note: Before searching an index has to be created. Use the menu "File -> Create index" to create the index or the button below this text.

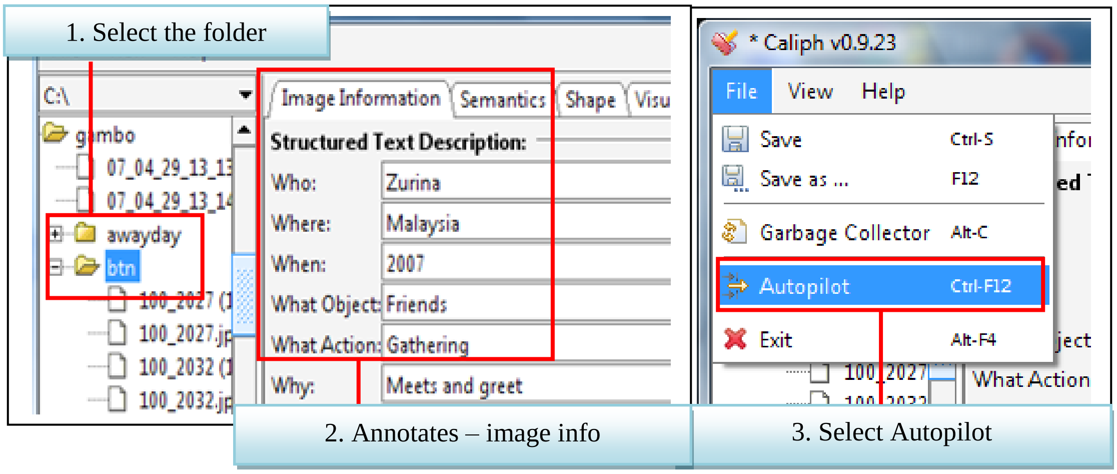
Searched for 1.05 s, formatting lasted 0.51 s 21.4M of 31.8M

Figure 3-2 Emir Querying Interface

3.2.1 Caliph

Caliph supports the creation of new MPEG-7 metadata in terms of MPEG-7 visual descriptors: ColorLayout, ScalableColour and EdgeHistogram. Annotation is manually done on the JPEG images by using free text or structured text descriptions, and one can add semantic information between those texts and rate the image quality on a scale of 1 to 5. The core element of Caliph is the semantic annotation panel that allows the user to create, define and import semantic objects like agents, places and events while maintaining a library of reusable MPEG-7 based semantic objects (Lux et al., 2003).

In making the task of annotation easier and less time consuming, Caliph provides an autopilot tool to generate common annotation for a set of images. Figure 3-3 shows the process of 1-3 on how to use the Autopilot function. By using the Autopilot, all images in the "Btn" folder are annotated with the same annotation shown in step 2, so users just need to add new information (if any) or delete information that is not relevant for a specific image in the folder independently.


1. Select the folder

2. Annotates – image info

3. Select Autopilot

Structured Text Description:

Who: Zurina

Where: Malaysia

When: 2007

What Object: Friends

What Action: Gathering

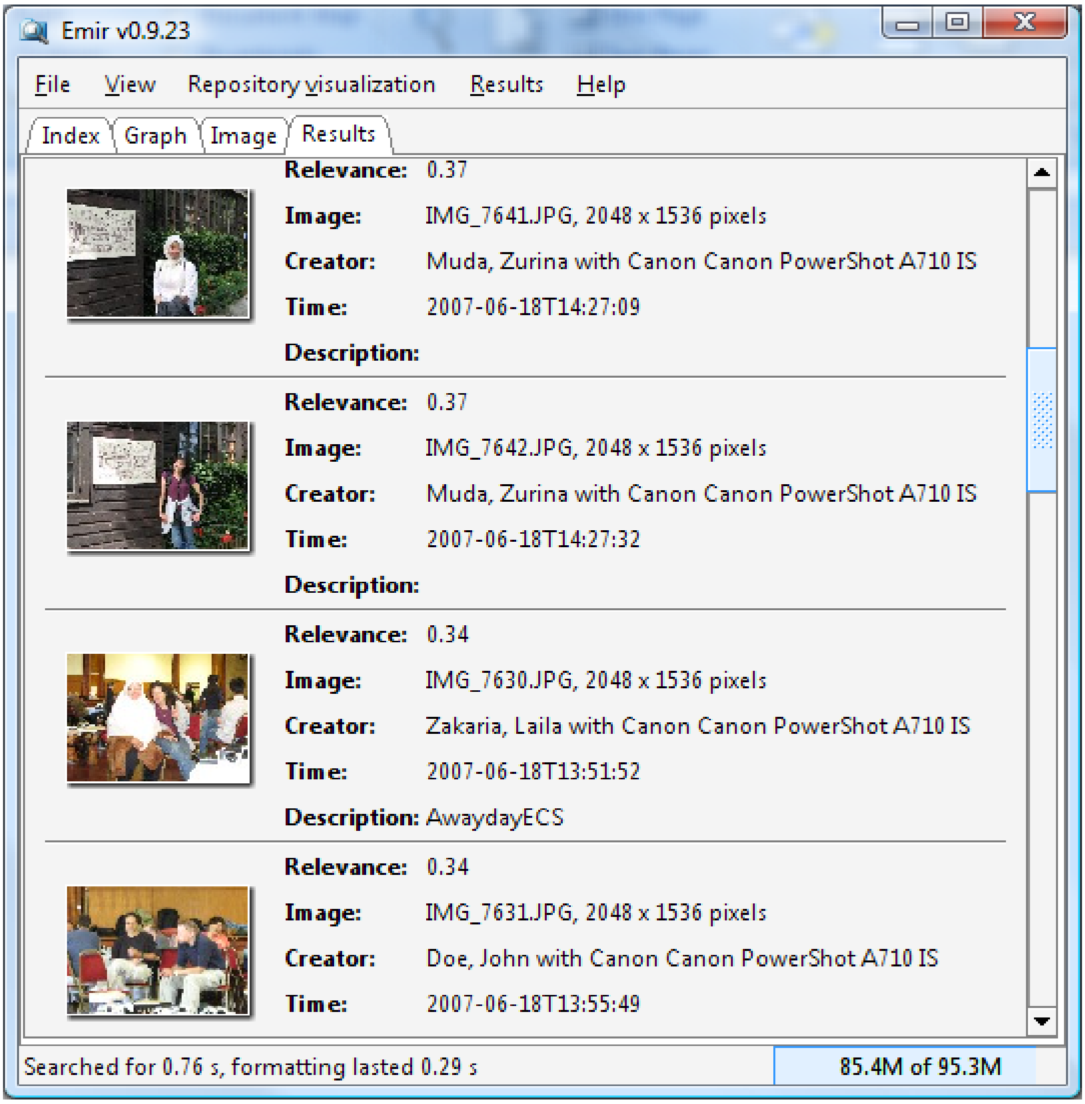
Why: Meets and greet

Figure 3-3 Flow of how to use the Autopilot in Caliph.

3.2.2 Emir

A set of photo files annotated with Caliph can be easily retrieved by using Emir. The retrieval prototype uses a file system without an index to store the descriptions, which reduces the speed of retrieval but keeps the platform independent and lightweight for easily demonstrating the software without a connection to the Internet (Lux et al., 2004). Emir allows retrieval of MPEG-7 descriptions based on keywords, simple semantic description graphs and query-by-example (QBE) by using the MPEG-7 visual descriptors: ColorLayout, ScalableColor and/or EdgeHistogram (Lux and Granitzer, 2005).

To enhance retrieval efficiency, content-based metadata is extracted and new instances of the image are created for faster visualization. When the process of searching is complete, the retrieval results are shown under the result tab. A query submitted to search for "Victoria Park" in Figure 3-2 before, has returned a list of images as a results where some of them are shown in the Emir interface in Figure 3-4.



Relevance: 0.37
Image: IMG_7641.JPG, 2048 x 1536 pixels
Creator: Muda, Zurina with Canon Canon PowerShot A710 IS
Time: 2007-06-18T14:27:09
Description:

Relevance: 0.37
Image: IMG_7642.JPG, 2048 x 1536 pixels
Creator: Muda, Zurina with Canon Canon PowerShot A710 IS
Time: 2007-06-18T14:27:32
Description:

Relevance: 0.34
Image: IMG_7630.JPG, 2048 x 1536 pixels
Creator: Zakaria, Laila with Canon Canon PowerShot A710 IS
Time: 2007-06-18T13:51:52
Description: AwaydayECS

Relevance: 0.34
Image: IMG_7631.JPG, 2048 x 1536 pixels
Creator: Doe, John with Canon Canon PowerShot A710 IS
Time: 2007-06-18T13:55:49

Searched for 0.76 s, formatting lasted 0.29 s
85.4M of 95.3M

Figure 3-4 Interface showing results for query "Victoria Park" in Emir.

In Emir the retrieval could be visualized as thumbnails in vector space based on ColorLayout, ScalableColor, EdgeHistogram or Semantic graphs by using a Repository Visualization tool.

3.3 Photostuff

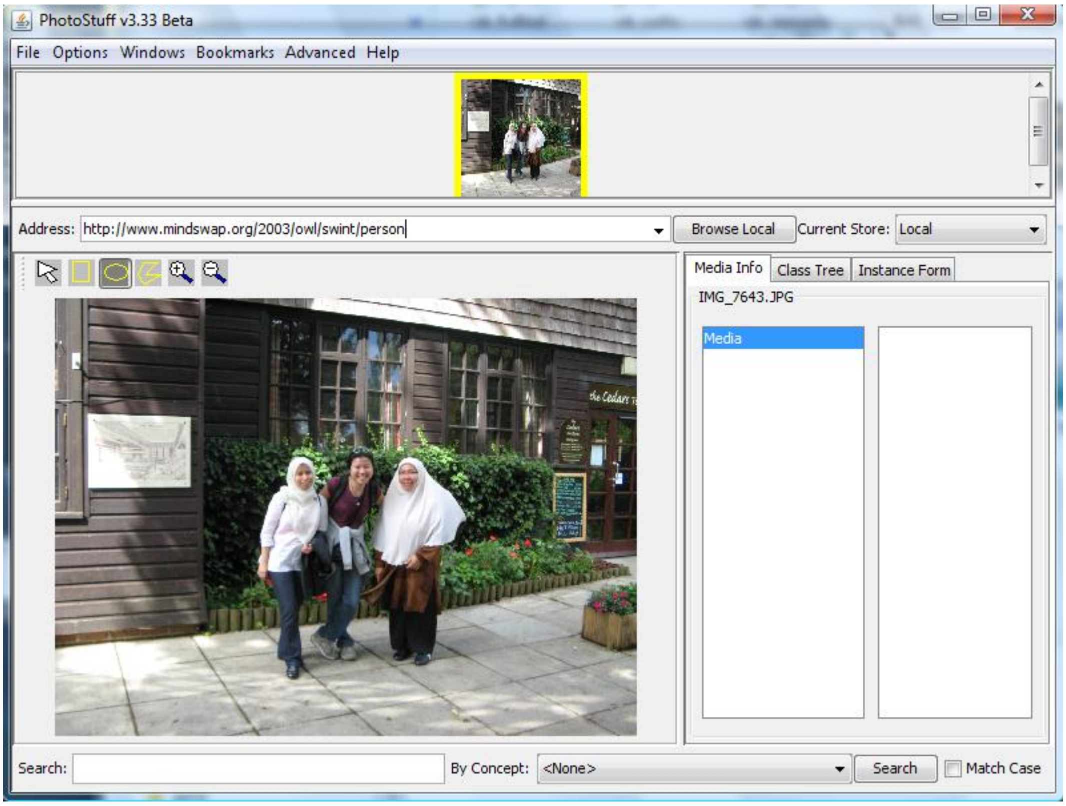
PhotoStuff is an annotation tool for digital images. It is a JAVA application, which is platform independent and open source. Photostuff was developed by Maryland Information and Network Dynamics Laboratory Semantic Web Agents Project (MINDSWAP) in the USA and was a proof of concept project.

Figure 3-5 shows the main interface of PhotoStuff. The tool annotates images using Web ontologies and exploits pre-existing embedded image metadata for automatic annotation enhancement through ontologies. The ontologies provide the expressiveness required to assert instances or classes to the contents of an image. An ability to load multiple OWL and RDFS ontologies, allows the tool to annotate an image and its regions' content with respect to a concept defined within the loaded ontologies from multiple domains (Halaschek-Wiener et al., 2005a).

The ontologies are visualized in a class tree list that can be dragged into any region or the image itself, creating a new instance of the selected class. Instances also can be loaded from any URI that refers to a RDF/XML document available on the Web. With this ability, Photostuff could also extract and used spatial ontology (if any). Figure 3-6 shows how the instances are created for the selected part of the image based on a person ontology.

Photostuff takes advantage of the existing metadata by extracting and encoding it into RDF/XML to become accessible on the Semantic Web. It is loosely coupled with a Semantic Web portal, providing image metadata management and seamless functionality, to import, perform mark-up and submit the generated annotation results. Users can annotate, share and manage their digital images on the Semantic Web

(Halaschek-Wiener et al., 2005a, Halaschek-Wiener et al., 2005b, Halaschek-Wiener et al., 2006).



Address: http://www.mindswap.org/2003/owl/swint/person

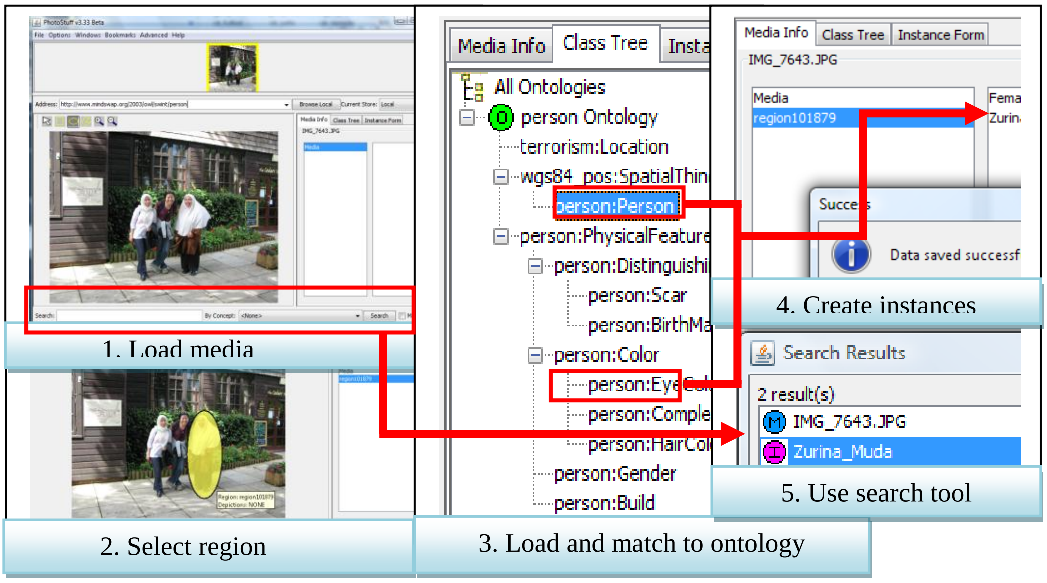
Current Store: Local

IMG_7643.JPG

Search:

By Concept: <None>

Figure 3-5 Photostuff Main Interface



1. Load media

2. Select region

3. Load and match to ontology

4. Create instances

5. Use search tool

Figure 3-6 Flow of how to annotate a part of an image in Photostuff.

3.4 AKTive Media

AKTive Media is a standalone application based on the JAVA platform, using RDF triples to represent the annotations which could be used during querying. AKTive Media[2] was developed by the Web Intelligence Technologies, Natural Language Processing Research group at the University of Sheffield, United Kingdom, and was partially funded by the AKT EPSRC and IST X-Media projects.

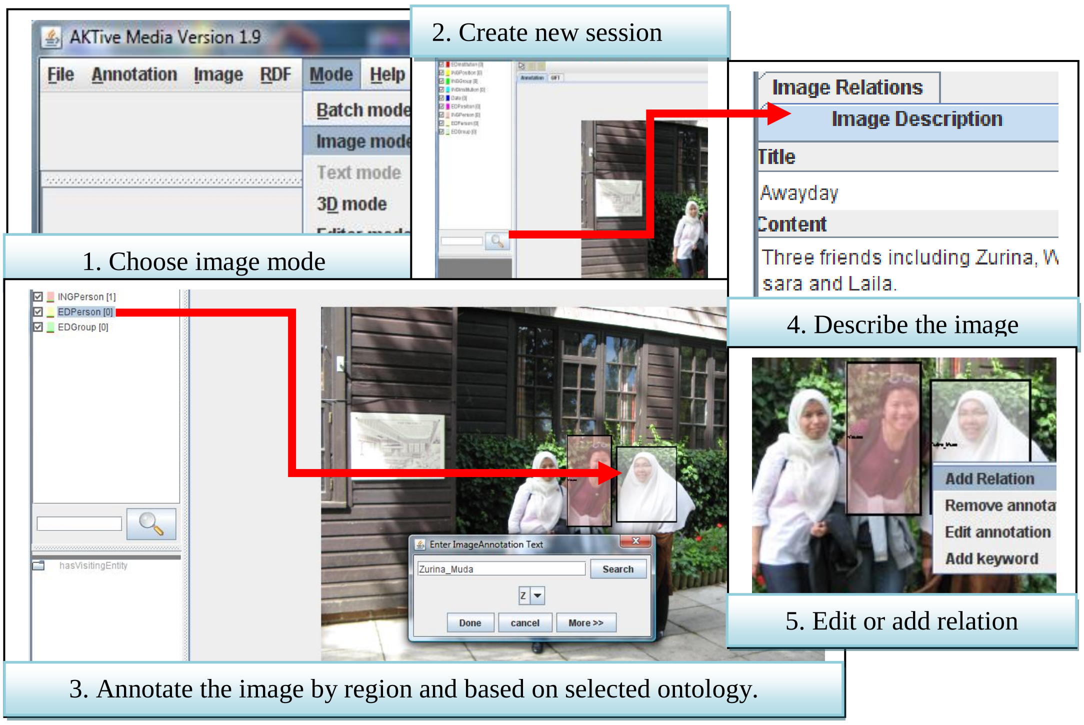
AKTive Media is a user-centric system for multimedia document enrichment. It uses Semantic Web and language technologies for acquiring, storing and reusing knowledge (Chakravarthy et al., 2006b, a). The aim is to provide a seamless interface that guides users through the annotation process by suggesting knowledge to the user in reducing the complexity of their task (Chakravarthy et al., 2006c). The user could adopt specific views of the ontology to annotate their documents without need to use the complete ontology.

The main functionalities supported are: image annotation, text annotation, cross text/image annotation and 3D functionality by supporting various types of image formats (JPG, GIF, BMP, PNG, TIFF). Currently, the 3D is not fully functioning except it has an example of a 3D object. Figure 3-7 shows how to annotate a part of an image in AKTive Media.

The whole/batch image, portions of text or images can be associated with concepts in the ontology with a point & click interface, where relational function and free-text annotations also can be added.

The AKTive Media tool is used as an interface to ease the burden of annotating the images by hand, before uploading the metadata to the user’s personal knowledge base (Chakravarthy et al., 2006a). It also actively works in the background of user applications in annotating web pages, personal memories and knowledge management (Chakravarthy et al., 2006b).

[2] http://www.dcs.shef.ac.uk/~ajay/html/cresearch.html



Title
Awayday
Content
Three friends including Zurina, W
sara and Laila.

Zurina_Muda

1. Choose image mode

2. Create new session

3. Annotate the image by region and based on selected ontology.

4. Describe the image

5. Edit or add relation

Figure 3-7 Flow of how to annotate a part of an image in AKTive Media.

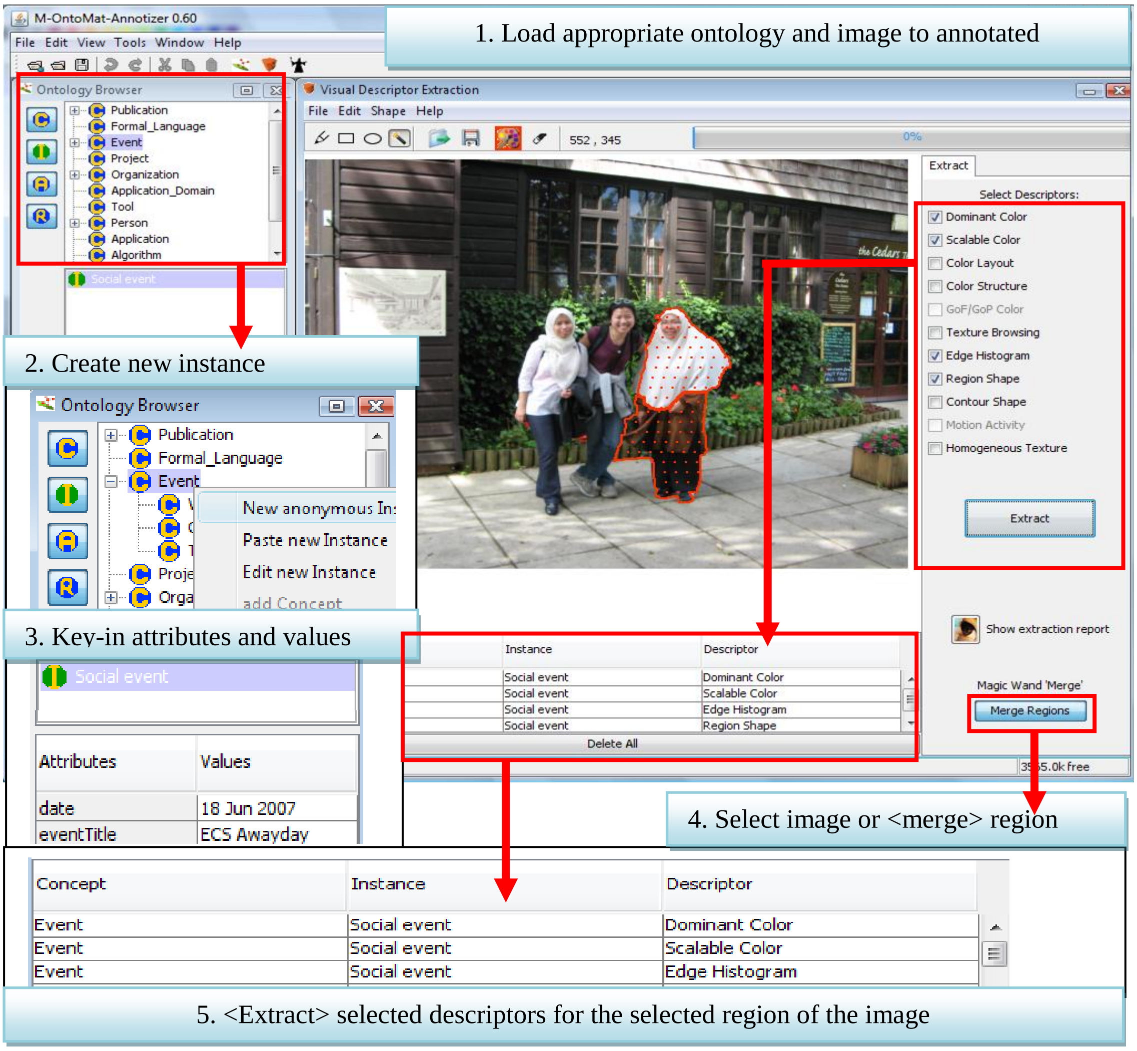
3.5 M-OntoMat-Annotizer

M-OntoMat-Annotizer, is a knowledge acquisition tool that supports the annotation of multimedia content. M-OntoMat-Annotizer (M stands for Multimedia) (Bloehdorn et al., 2005) is a tool developed by the AceMedia projects as an extension of the CREAM (*CREating Metadata for the Semantic Web*) framework (Handschuh and Staab, 2003) and OntoMat-Annotizer. The evolution included the Visual Descriptor Extraction Tool (VDE) as the core component for supporting the initialization of RDF(S) domain ontologies with low-level MPEG-7 visual descriptors. The VDE Visual Editor (see Figure 3-8) and Media Viewer present a graphical interface for loading and processing visual content, visual feature extraction and linking with domain ontology concepts.

M-OntoMat-Annotizer is a standalone application based on JAVA and implemented to exploit the ontology infrastructure and enrich the domain ontologies with multimedia descriptors (Petridis et al., 2006, Saathoff et al., 2006). It processes visual content such as image and video, and extracts MPEG-7 visual descriptors

(ISO/IEC15938-3 and FCD, 2001), called visual prototypes of ontology classes, which are stored as RDF instance (Saathoff et al., 2006). This is added to the knowledge base and can be retrieved in a flexible way during multimedia content analysis (Petridis et al., 2006), while at the same time leaves the original domain ontology unmodified.

M-OntoMat-Annotizer also supports semi-automatic segmentation of the image/frame; by allowing the user to select or draw a desired region or merge two regions by using a <Magic Wand 'Merge'> button and apply the multimedia descriptor extraction to the selected region as shown in Figure 3-8. The figure shows five steps to annotate an image and then extracted it by using MPEG-7 visual descriptor in M-OntoMat-Annotizer.



1. Load appropriate ontology and image to annotated

2. Create new instance

3. Key-in attributes and values

4. Select image or <merge> region

5. <Extract> selected descriptors for the selected region of the image

Select Descriptors:

☑ Dominant Color
☑ Scalable Color
□ Color Layout
□ Color Structure
□ GoF/GoP Color
□ Texture Browsing
☑ Edge Histogram
☑ Region Shape
□ Contour Shape
□ Motion Activity
□ Homogeneous Texture

Instance	Descriptor
Social event	Dominant Color
Social event	Scalable Color
Social event	Edge Histogram
Social event	Region Shape

Attributes	Values
date	18 Jun 2007
eventTitle	ECS Awayday

Concept	Instance	Descriptor
Event	Social event	Dominant Color
Event	Social event	Scalable Color
Event	Social event	Edge Histogram

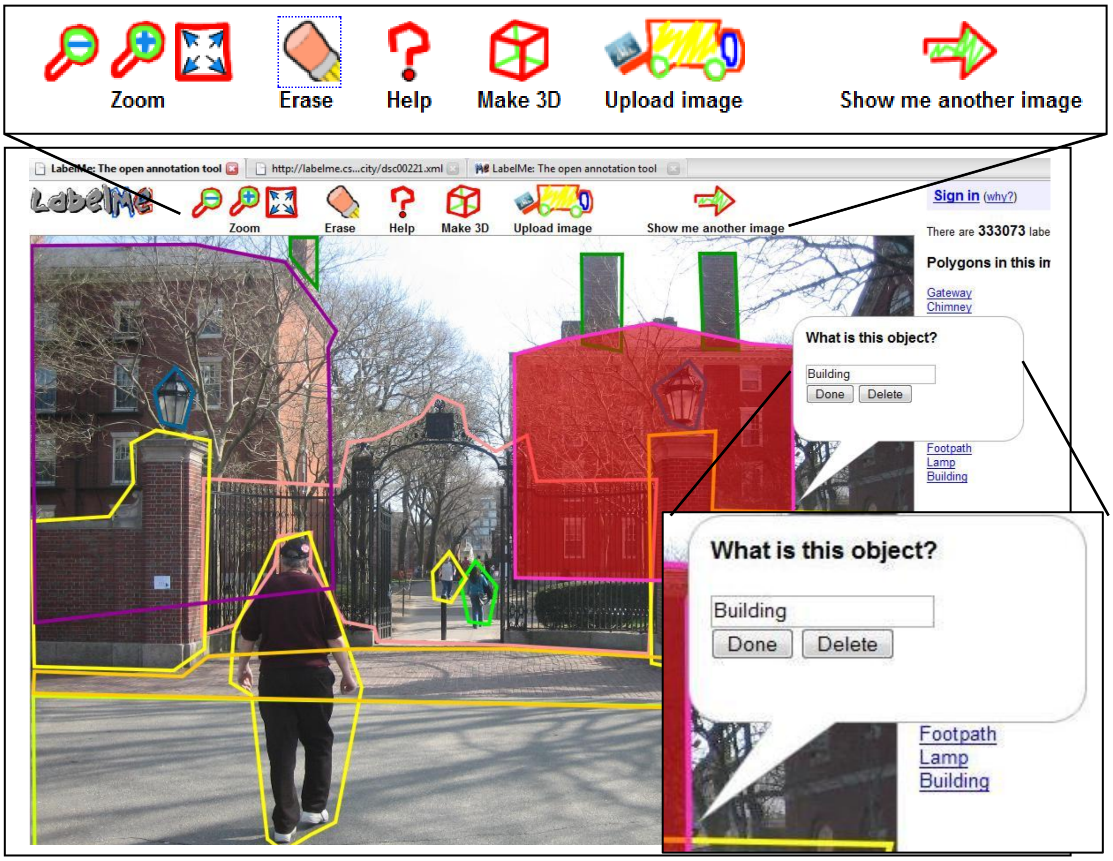
Figure 3-8 Flow of how to annotate selected region in M-OntoMat-Annotizer.

3.6 LabelMe

LabelMe is an open annotation tool that supports the annotation of image content. This web-based annotation tool is based on JAVA and allows researchers to label objects or polygons in images and share the annotations with the rest of the community. The tool was developed by Russell et al. (2008) at the Computer Science and Artificial Intelligence Laboratory in MIT. The goal of the annotation tool is to provide a drawing interface that works on many platforms, is easy to use and allows instant sharing of the collected data. The main annotation interface of the tool is shown in Figure 3-9. The web-tool's image dataset continuously grows over time (Russell et al., 2008). To date, more than 764K labelled objects annotation that have been assigned to 66589 images in LabelMe.

The tool is easy to use with straightforward point and click operations. When a user enters the LabelMe annotation page, an image is displayed. The image comes from a large image database covering a wide range of environments and object categories (Russell et al., 2008). Often the image shown has already been labelled, but the user may label a new object by clicking control points along the object's boundary and finishes by clicking on the starting control point. Upon completion, a pop-up dialog bubble will appear querying for the object name, as shown in Figure 3-9.

The user can freely type-in the object name and press enter or the *done* button to close the bubble. This label is recorded on the LabelMe server and is displayed on the presented image. The label is immediately available for download and is viewable by subsequent users who visit the same image (Russell et al., 2008). The users are free to label as many objects depicted in the image as they choose. When they are satisfied with the objects labelled in an image, they may proceed to label another image by pressing the *Show Next Image* button. The tool also enables registered users to explore, search and download the dataset of images that has been annotated. An extension with WordNet has been established where the user can view the whole annotation taxonomy that has been created by annotators who have annotated images using this tool.



There are 333073 labe

Polygons in this im

What is this object?

Building

Figure 3-9 LabelMe Screenshot with Zooming Popup Dialog Bubble and Menu.

3.7 Comparison between the Tools

In this section the tools are compared according to specific functions of the tools; types of descriptor/metadata; operating system (OS) and type of application; input and output; tool features; speed of processing; and reviews on the advantage, and disadvantage, of the tools. The comparisons are summarized in Table 3-1.

Table 3-1Based on an evaluation of these tools and with the help of the manuals and documentation, the difficulty of using these tools has been ascertained together with the amount of foreknowledge needed for the underlying knowledge representation (Duineveld et al., 2000). The result of this “difficulty of learning” study is shown in Figure 3-10. The outcome may be influenced by the help menu or documents available for the tools and how much time has been spent in exploring and using the tools.

Caliph & Emir require significant prior understanding of MPEG-7 metadata created in Caliph when using the Emir for searching. AKTive Media and M-OntoMat-Annotizer require expertise in the use of an ontology and ontological models, therefore both tools are only suitable for power-users with that particular background. AKTive Media and M-OntoMat-Annotizer are easy to use as long as the annotator is familiar with the interfaces and knows which ontology to refer to and manage to get the required information. However M-OntoMat-Anotizer is more user-friendly and easier to use as it is comprised of libraries of reusable ontologies. Photostuff is hard to learn to use because of inadequate reference documents[3]. LabelMe is the easiest tool to use for annotation because it is a straightforward point and click tool.

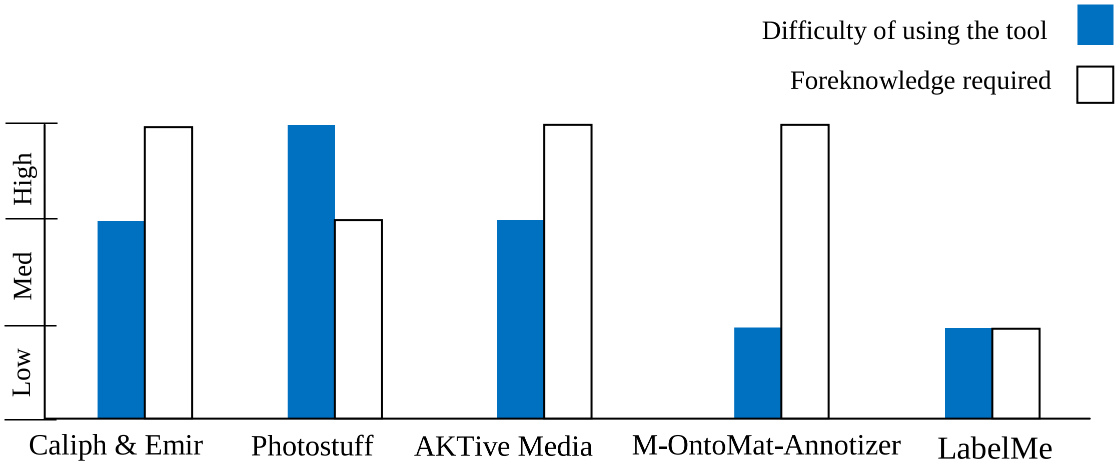


Figure 3-10 Difficulty of learning the tools.

[3] Follow-up with the author of the tool established that the development of the tool was incomplete.

Table 3-1 Comparison of image annotation tools.

Name	**Caliph and**	**Emir**	**AKTive Media**	**Photostuff**	**M-OntoMat-Annotizer**	**LabelMe**
Specific function	Manual annotation	Retrieval by text, Query by Example	Manual annotation, query system-SPARQL	Manual image annotation for Semantic Web	Manual annotation for multimedia analysis	Semi-automatic for open annotation
Level of annotation	Global		Local	Local	Local	Local
Media	Image		Text and Image	Image	Image and video	Image and video
Type of application	Standalone		Standalone	Standalone	Standalone	Web-based
Metadata	MPEG-7		RDF	RDF(S)	RDF(S) and MPEG-7	XML
Ontology representation	No ontology		RDFS, OWL, ONT, DAML.	RDFS, OWL	DAML, RDFS	No ontology
Input	JPG		Plain text, RDF, HTML; JPG, GIF, BMP, PNG, TIFF	JPG, RDF	JPG, GIF, TIFF, PNG AVI, MPEG, MOV	JPG AVI, MOV, MPG
Output	MPEG-7 (IPTC & EXIF into MPEG-7)		RDFS, OWL, DAML	RDF	RDF	XML
Speed of processing	Fast	Fast	Medium	Fast	Medium	Fast

Name	Caliph and	Emir	AKTive Media	Photostuff	M-OntoMat-Annotizer	LabelMe
Features or Functionality	• Tab: Image info, Semantics, Shape, Visual. • 3-panels • Autopilot	• Search by: textual, image description, QBE and sub-graph. • Tab: Index, graph, image, result. • Visualize result.	• Session • Corpus • Ontology • Image relation and description. • Annotation mode	• Ontology image • Media Info, Class Tree, Instance form • Media Component • Media and ontology List • Launch Bar.	• Ontology browser • VDE - Visual Descriptor Extraction • Extracted descriptions list • Image and video annotation.	• Annotation tool • Matlab Toolbox • Search box • Download datasets • WordNet • Image and video annotation
Others	• Image rating • Semantics relation	• Visual Semantics relation	• Searching node • Mode: batch, image, text, 3D and editor	• Preferences • Viewing RDF • Plug-ins	• Region Merging	• 3D pop-up • LabelMe Source
CRITICAL REVIEW						
Advantages	• Semantic relation information – agent, event and (object, place, time)	• Semantic graph for retrieval. • Visualisation of retrieval result.	• Import multiple ontologies. • Choices of mode – text, image batch and 3D.Choices of mode.	• Multiple-ontologies.	• knowledge extracted will be use for automatic semantic analysis. • Annotate and retrieval for image and video.	• Objects/Polygons Extractions • Easy search for objects and scenes • Free download of image datasets/folders • Registered user can retrieve their annotated image.
Disadvantages	• Based on text.	• Semantic relation is by default/ fix.	• Annotation complicated.	• Incomplete tools and without user manual. • Hard to use.	• Region merging only for two regions.	• Different style of annotation by different users

3.8 Evaluation Framework

A comparative study was conducted to evaluate the image description tools using an evaluation framework adapted from Lewis from (1995) and Duineveld (2000). This evaluation framework is shown in Table 3-2 and is categorized into:

1. Image annotation features components.
2. User interface components.

Table 3-2 Evaluation framework for image annotation tools.

Image Annotation Features

1. Does the tool support local image annotation?
2. Does the tool allow segmentation/region/object for image annotation?
3. Does the tool support spatial relationships?
4. Does the tool support several image formats?
5. Does the tool provide a feature for resizing image for annotation?
6. Does the tool allow group/batch annotation?
7. Does the annotation descriptions easy to understand?
8. Is there any free-text (open) for annotation?
9. Is the tool linked to ontology?
10. Does the tool provide libraries of reusable ontologies?
11. Does the tool provide libraries of reusable images?
12. Does the tool has multiple features/options for annotation?

User Interface

1. Is the tool easy to use?
2. Is there information about the term used?
3. Is it easy to find the information needed?
4. How is the speed of updating after new data inserted?
5. Does the tool's interface consistent?
6. Does the tool provide any feedback?
7. Is the meaning of the commands clear?
8. Is the menu or command function as given?
9. Are there any stability problems (crashes, hang etc)?
10. Does the tool provide assistant or Help menu?

This study was done to establish which tools might provide a good base for our automatic spatial annotation. The evaluations discounted Photostuff due to inadequate and incomplete documentation.

3.8.1 Results and Discussions

The evaluation results for the tools are shown in Table 3-3. The results allowed were either Yes (3) or No (-) or a 3-level scale of high (3), medium/reasonable (2) and low

(1) adapted from (Duineveld et al., 2000). For images descriptions components, follow-up with the developer of the tools has been established to ensure the reliability of the result.

Table 3-3 Evaluation results for image description tools.

Components	Caliph & Emir	AKTive Media	M-OntoMat Annotizer	LabelMe
Image Description Features				
1. Support local annotation.	No	Yes	Yes	Yes
2. Allow segmentation/object	-	2	3	2
3. Free-text (open) annotation.	2	1	1	3
4. Support for spatial relationships.	No	No	No	No
5. Support several image formats.	1	3	3	1
6. Group/batch annotation.	2	2	-	-
7. Annotation descriptions are understandable.	1	2	1	3
8. Linked to the ontology.	-	3	2	-
9. Libraries of reusable ontologies.	-	1	2	-
10. Libraries of reusable images.	1	1	2	3
11. Feature for resizing image.	1	3	2	2
12. Multiple features/options for annotation.	2	3	2	2
User Interface				
1. Easy to use?	2	1	2	3
2. Information of the terms used.	-	-	1	2
3. Easy to find the information needed.	3	2	1	2
4. Speed of updating new data.	3	1	1	3
5. Interface consistency.	1	2	3	3
6. Provide feedback.	3	2	2	3
7. The meaning of the commands clear.	2	2	3	3
8. The functional of menu or command.	2	1	3	3
9. System stability (crashes, hang etc)?	3	2	3	3
10. Assistant/Help menu.	2	1	2	3
Total	31	38	42	47

Scale: Yes (3)/No(-) and 3-level scale of high (3), medium/reasonable (2) and low (1)

The results in Table 3-3 show that the tools that have been discussed are involved with annotation of the whole image to some extent or level. All tools except the Caliph & Emir annotated images locally and allow the segmentation of regions or objects in the images. However, Caliph and Emir allow an input field named agent for people in the image during the annotation at the global level. Annotation based on segmented regions or objects in the image enable the user to annotate the image locally in a more specific way.

Annotation in AKTive Media and M-OntoMat-Annotizer are restricted and based on an ontology corpus loaded within the tools during the annotation. Although Caliph & Emir also has restricted the types of information that can be inserted for an image during the annotation, the tool also provides space for free text description. While in LabelMe, the annotation can be done with open text which gives users some freedom. This will enable the user to annotate their images according to their preferences, making it easier for them to refer to, retrieve or use the images later on.

None of the tools support spatial relationships. By adding spatial relationship descriptors, the annotation and knowledge of the image content becomes more expressive, specific and unique. Furthermore the process of retrieval could be done in an explicit and more powerful way where the retrieval performance would increase.

AKTive Media and M-OntoMat-Annotizer allow several image formats for annotation, while Caliph & Emir and LabelMe only allow JPG file format for annotation. JPG is a format generally used for images as it consume less space for storing providing high speed in retrieving. In addition, Caliph & Emir and AKTive Media allow batch or mode annotation which enables general annotation for a set or volume of images. This feature reduces and simplifies the annotation task.

Annotation descriptions in Caliph & Emir and M-OntoMat-Annotizer were hard to understand compared to AKTive Media. However, the annotation descriptions in LabelMe are in the form of XML and consist of detailed information about the objects in the image, with label and coordinates. This information is very easy to understand and could be used further in the research in computing spatial relationships among the objects in images.

AKTive Media and M-OntoMat-Annotizer are supported by an ontology for annotation, while Caliph & Emir are based on MPEG-7 metadata. LabelMe is not directly supported by an ontology except that it has an extension of the WordNet taxonomy attached. Therefore, in developing an annotation tool, the criterion of supporting ontologies is an essential aspect to be considered because it could help to enrich the expressiveness of the image annotation.

There was no library provided for reusable images or ontologies in Caliph & Emir. Ontology libraries were provided in both AKTive Media and M-OntoMat-Annotizer, but with limited examples of ontologies in AKTive Media compared to examples in M-OntoMat-Annotizer's library. This will make it easier to use without the need to find an ontology from another source. LabelMe contains datasets of thousands of images that are categorised into folders and can be downloaded or used online by the user. The images or folders are easy to access online or by downloading for annotation process.

AKTive Media is the only tool that provides the feature of image resizing, where a user could zoom in and out of the image for annotation. The tools also consist of various features or mode for annotation. The screen size for image annotation allocated in M-OntoMat-Annotizer is reasonable, so the function for resizing is not critical but in Caliph & Emir the space allocated is small and in need of resizing. The screen size for image annotation in LabelMe is satisfactory for annotation. LabelMe also allows the user to open an image in use in another window where the user could save it independently. Other than image annotation, currently LabelMe also supports video annotation.

In terms of interfaces, the results in Table 3-4 show that LabelMe has a very user-friendly interface and is easy to use. The interface is simple and straight forward, thus it is easy to find the information needed, while others tend to assume textual interfaces which are not user-friendly and quite hard to use unless they have online help or manuals. The annotation processing speed of Caliph & Emir and LabelMe are very fast compared to the other tools. In terms of consistency, the interface in AKTive Media and LabelMe are very consistent compared to the other tools. Interactive feedback is given by Caliph & Emir and LabelMe; sometimes by AKTive Media and M-OntoMat-Annotizer when the command cannot be used or when the user misses some steps in the annotation process. Feedback is important in helping the user to know what action to take next, to ensure they are on the right track and complete their task successfully.

3.9 Conclusion

The investigation and study presented above shows that each of the tools: Caliph & Emir, AKTive Media, Photostuff, M-OntoMat-Annotizer and LabelMe offered some special features on their own that were not offered by others. Some of the tools could be enhanced with flexible and improved open-text input and the others with ontologies. Most of the tools are also involved with manual annotation where to automate all or some of the features could enhance the capability of these tools.

Although this was a relatively rapid comparative evaluation of the particular tools, from the pros and conts, and based on the total marks given in the study shown in Table 3-3, with highest marks of 47 among other tools investigated, LabelMe has been selected to be used further in the research because it provides a substantial foundation as the base technology for further development in image annotations. LabelMe annotates the images locally thus allowing object annotation. The tool has a very user friendly interface and is easy to use. The image annotation description or output is in an understandable form and can be manipulated further. The output is in the form of x and y coordinates of the bounding box of the objects or regions in the image that has been annotated locally. These coordinates and annotations could then be used to compute and generate the spatial relationships between those regions which will make the annotations more specific and accurate and which will hopefully provide benefit by improving the image annotation and retrieval system.

In conclusion, there are many challenges to improve the existing tools to make them function semi-automatically or automatically, combining the annotation descriptors with support from an ontology, and yet, making an allowance for the annotation of spatial relationships of objects in the image content. Thus, our research framework will present and incorporate these components to be developed in the Spatial Semantic Image System.

3.9.1 The Research Framework

The research framework for the facilitating the research needed to provide ontologically based spatial annotations of image content is illustrated in

Figure 3-11. The framework consists of three main components: the Spatial Annotation Component, the Ontology Component and the Retrieval Component.

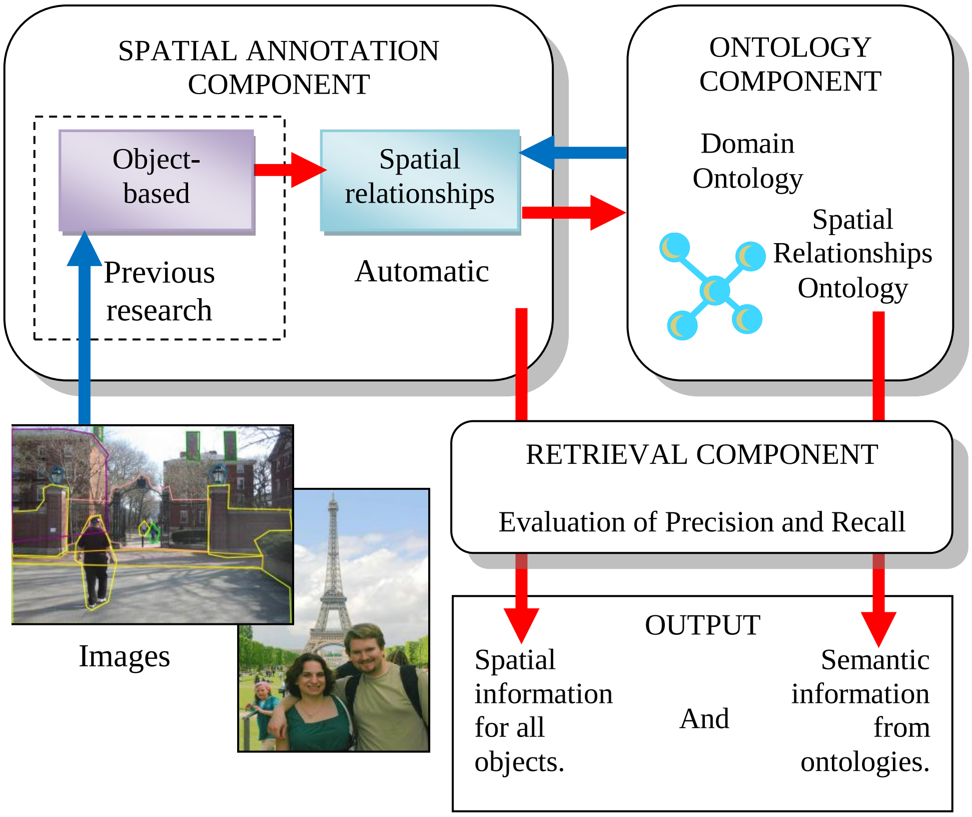

SPATIAL ANNOTATION COMPONENT
Object-based
Spatial relationships
Previous research
Automatic
ONTOLOGY COMPONENT
Domain Ontology
Spatial Relationships Ontology
Images
RETRIEVAL COMPONENT
Evaluation of Precision and Recall
OUTPUT
Spatial information for all objects.
And
Semantic information from ontologies.

Figure 3-11 The Research Framework

- **The Spatial Annotation Component**

This component should automatically extract and identify spatial information for objects in an image. It delivers statements about the absolute spatial position for single objects and spatial relationship between pairs of objects. The component will include the development and implementation of spatial relationship algorithms and spatial inferences using order of magnitude height information from the ontology. The component is intended to increase the flexibility of the input process as well as generating resourceful spatial knowledge, and flexible and more precise output automatically.

- **The Ontology Component**

This component will contain a spatial relationships ontology and a domain ontology for objects related in the image from the annotation component. The component will help to standardise and control the representation of the spatial knowledge-base to be used in describing spatial relationships between objects in images. These are necessary during the description of the image content and will be useful in supporting queries for relevant images in the retrieval component. A domain ontology has been explored to be used with the spatial relationships ontology according to the scope of the research. The ontology will also be equipped with added knowledge so that advanced spatial semantic information can be extracted as an addition to the spatial relationships information.

- **The Retrieval Component**

This component will integrate the annotation and ontology components mentioned above to facilitate retrieval enhanced with spatial information. An SQL based spatial query facility will be developed and the retrieval performance assessed in terms of precision and recall.

From the research framework, the first stage is to develop the spatial annotation component. This requires decisions about the spatial concepts and terminologies that need to be considered, identified, defined and specified based on human perspectives and this will be investigated in the following chapter.

Chapter 4

Choosing Spatial Terms

4.1 Introduction

A wide variety of spatial terminology has been used in the literature and this chapter discusses work done in order to identify and to select a set of specific spatial relationship terms to be used for further experiments and developments in this research.

The study began by looking into the previous research to make initial proposals of spatial terms that will be considered. Then, in order to identify how humans describe images using spatial terms, an online Image Description Survey has been designed and implemented to obtain image descriptions from users.

The online survey is introduced with the aim of identifying and defining common spatial terms used by users, and how they used the terms. Responses, analysis and findings of the survey are illustrated and presented. Both, the ground truth and the survey use the same images from the Corel dataset within the scope of our research domain.

4.2 Spatial Relationships Terminology

We saw in the section 2.6 that a wide range of spatial relationships has been introduced in the literature. An even wider variety of terms is used to describe those relationships. In the work which follows, we consider two main classes of spatial terms: relative terms which describe the relative positions of two objects (e.g. A is *left of* B) and absolute terms which describe the absolute position of an object (e.g. A is *at the top*). In most cases, unless otherwise stated, we consider the terms to refer to spatial positions within the image as observed by a person viewing the image as opposed to the positions within the real world. Examples of spatial terminology from the literature which initially seemed appropriate and relevant to our work are shown in Table 4-1.

Table 4-1 Related research

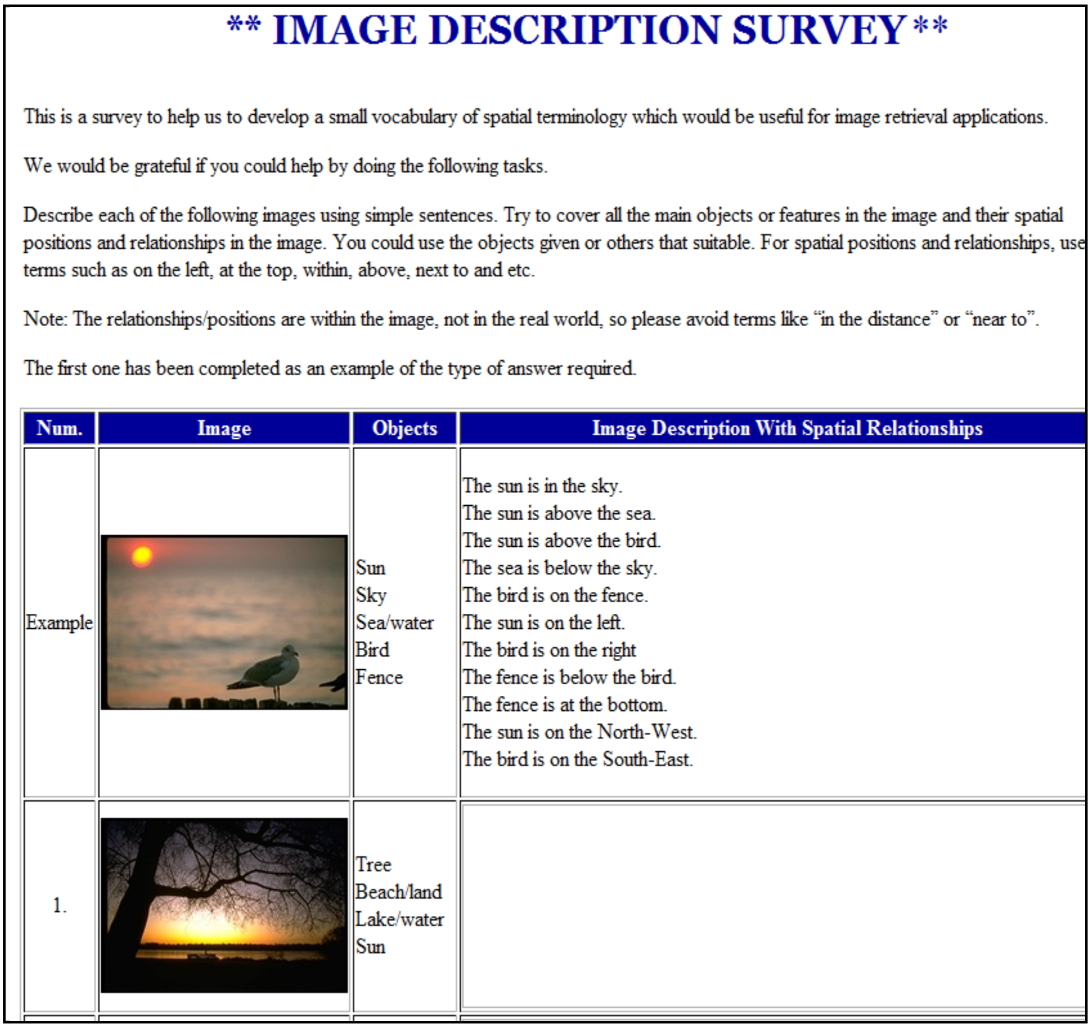
References	**Absolute Terms**	**Relative Terms**
Ahmad & Grosky (2003)	North-East, North-West, South-East, South-West.	-
Hollink, et al. (2004)	Centre, North, East, West, South, North-East, North-West, South-East, South-West.	Right, Left, Above, Below, Near, Far, Contain, Next.
Lee, et al. (2006)	Left-upper, Left-lower, Right-upper, Right-lower.	Upper, Below, Left, Right.
Yuan, et al. (2007)	-	Above, Below, Left, Right.

Before selecting a particular set of spatial terms it was deemed valuable to explore briefly how humans describe spatial relationships in images. A small online, web-based survey was therefore developed and implemented with the aim of discovering and gathering a user perspective on spatial relationships for describing images. The objectives of the online Image Description Survey are:

1. To identify spatial terms commonly used by people to describe images for image retrieval applications.
2. To identify how people use these spatial relationships in sentences describing images.
3. To identify the meaning of the spatial terms used by people from the way they used the terms in the sentences.

4.3 Development and Implementation

The survey was developed using PHP and implemented on the Web. As only some basic spatial terminology was to be identified a small number of images (ten) were selected to be evaluated by the users as the respondents of the survey. The survey could be accessed by respondents both internal and external to the university. A screen shot of the survey is shown in Figure 4-1. Respondents were asked to describe the spatial relationships and positions for the main objects in each image. The first image was been completed as an example to guide the respondent on how to complete the survey. Main objects in the images have been identified but the respondents could use them and/or include other objects in the image and could use their own terminology for the spatial terms.



** IMAGE DESCRIPTION SURVEY **

This is a survey to help us to develop a small vocabulary of spatial terminology which would be useful for image retrieval applications.

We would be grateful if you could help by doing the following tasks.

Describe each of the following images using simple sentences. Try to cover all the main objects or features in the image and their spatial positions and relationships in the image. You could use the objects given or others that suitable. For spatial positions and relationships, use terms such as on the left, at the top, within, above, next to and etc.

Note: The relationships/positions are within the image, not in the real world, so please avoid terms like "in the distance" or "near to".

The first one has been completed as an example of the type of answer required.

Num.	Image	Objects	Image Description With Spatial Relationships
Example		Sun Sky Sea/water Bird Fence	The sun is in the sky. The sun is above the sea. The sun is above the bird. The sea is below the sky. The bird is on the fence. The sun is on the left. The bird is on the right The fence is below the bird. The fence is at the bottom. The sun is on the North-West. The bird is on the South-East.
1.		Tree Beach/land Lake/water Sun	

Figure 4-1 Survey Interface

All responses were captured by a PHP Script and saved in to a data file together with the time and date of the submission. Each time a respondent submits, the data file is added with a new entry with a time and date recorded for the submission.

4.3.1 Result from the Survey

The number of respondents who filled in and submitted the survey was 15. Although a small number, it was sufficient to indicate the variety of spatial terms used by people and also those used frequently. Results from the survey were accumulated and an analysis has been done. There were 45 spatial terms used by the respondents. The spatial terms have been categorised into absolute and relative terms. Absolute terms describe the spatial position of a single object and relative terms describe the relationship between two objects.

To identify spatial terms that are suitable to be used, we considered and analysed spatial terms that have been used more than once. Hence we dropped spatial terms that were only used once and further analysis and discussion will focus on the 28 spatial terms with more than one occurrence. For these terms, the frequency with which they were used for each image from the survey is shown in Table 4-2. The table also shows the sum (∑) of the frequencies for each term across all images.

The absolute terms included are TOP, BOTTOM, LEFT, RIGHT, MIDDLE, CENTRE, FRONT/FOREGROUND, CENTRE-BOTTOM, MIDDLE-BOTTOM, MIDDLE-LEFT AND COMPASS directions. Compass directions are treated as one term and the directions used by the users include North, South, East, West, North-East, South-East and South-West but North-West has not been used. The relative terms included are ABOVE, BELOW, ON, IN, WITHIN, LEFT, NEXT, BESIDE, BY, BETWEEN, OVER, AROUND, ACROSS, UNDER, BEHIND AND SURROUND.

Table 4-2 Terms Frequency Analysis

	ABSOLUTE TERMS											**RELATIVE TERMS**																
IMAGE	**at/on TOP**	**at/on the BOTTOM**	**on the LEFT**	**on the RIGHT**	**at/on/in the MIDDLE**	**on/in the CENTRE**	**at the FRONT/FOREGROUND**	**IN THE CENTRE OF the BOTTOM (CB)**	**in the MIDDLE at the BOTTOM (MB)**	**MIDDLE-LEFT (ML)**	**COMPASS DIRECTION**	**is ABOVE**	**is BELOW**	**is ON**	**is IN the**	**is WITHIN**	**at/to LEFT of**	**NEXT to**	**BESIDE**	**BY**	**is in BETWEEN**	**OVER**	**AROUND**	**is ACROSS**	**is UNDER**	**is BEHIND**	**is/in FRONT of**	**SURROUND**
1	3	6	10	0	3	0	1	0	0	0	1	6	7	3	1	0	1	2	0	0	0	1	1	0	2	6	1	0
2	3	6	2	0	1	2	0	0	0	0	7	15	5	2	4	0	1	2	2	2	0	1	0	1	0	0	3	0
3	2	3	13	5	3	0	0	0	0	2	8	19	2	7	1	0	2	1	0	1	2	0	1	0	3	6	0	0
4	4	5	7	9	3	1	0	0	0	0	6	12	3	2	8	0	0	2	0	1	2	0	0	0	0	0	0	0
5	8	4	2	2	3	0	0	0	2	0	3	12	5	3	3	0	1	1	0	0	0	2	0	2	1	3	1	0
6	3	10	1	7	4	0	0	2	1	0	3	16	0	3	4	5	1	1	0	0	0	0	0	0	1	0	1	0
7	3	6	1	0	6	3	1	0	0	0	1	9	4	2	2	0	0	0	0	0	3	0	0	1	1	6	0	0
8	4	6	1	0	4	1	0	1	0	0	2	13	4	2	0	0	0	0	0	0	0	0	0	0	2	2	2	0
9	4	3	2	0	4	1	0	0	0	0	4	9	4	2	0	1	0	1	0	0	0	1	0	0	3	0	2	0
10	5	8	1	0	6	1	1	1	0		9	9	4	14	2	1	0	0	0	0	0	0	0	0	0	0	4	2
∑	**39**	**57**	**40**	**23**	**37**	**9**	**3**	**4**	**3**	**2**	**44**	**120**	**38**	**40**	**25**	**7**	**6**	**10**	**2**	**4**	**7**	**5**	**2**	**4**	**13**	**23**	**14**	**2**

Table 4-2 highlighted (in black) the most frequently used terms for each image. In general the table shows that the most commonly used term is ABOVE (relative) for most of the images with frequencies of 9 to 19, while the term LEFT (absolute) is commonly used for image 1 with a frequency of 10.

The total use of each spatial term shows that, the most commonly used spatial term is ABOVE (relative) with frequency of 120, followed by BOTTOM (absolute) with frequency of 57, and then LEFT (absolute) and ON (relative) with frequencies of 40.

From this result, it is apparent that most of the users use term ABOVE rather than BELOW, term LEFT rather than RIGHT though both pairs are reciprocal.

4.3.2 Image-Term Frequencies Histograms

The term frequency of each spatial relationship used by the users to describe an image is visualized in the form of histograms for better comparison. The histograms are divided into two parts by a blue line to differentiate between the absolute and relative terms.

It can be seen from the histogram that for these images the spatial term ABOVE (relative) is the most frequently used term for all images including Image 2-9 as grouped and shown in Figure 4-2, except for Image 1 where the spatial term LEFT (absolute) is the most frequently used and Image 10 where the spatial term ON (relative) is the most frequently used. Both the histogram for Image 1 and Image 10 are shown in Figure 4-3. It is worth noting again that ABOVE and BELOW, and LEFT and RIGHT are reciprocal pairs and should perhaps be considered together.

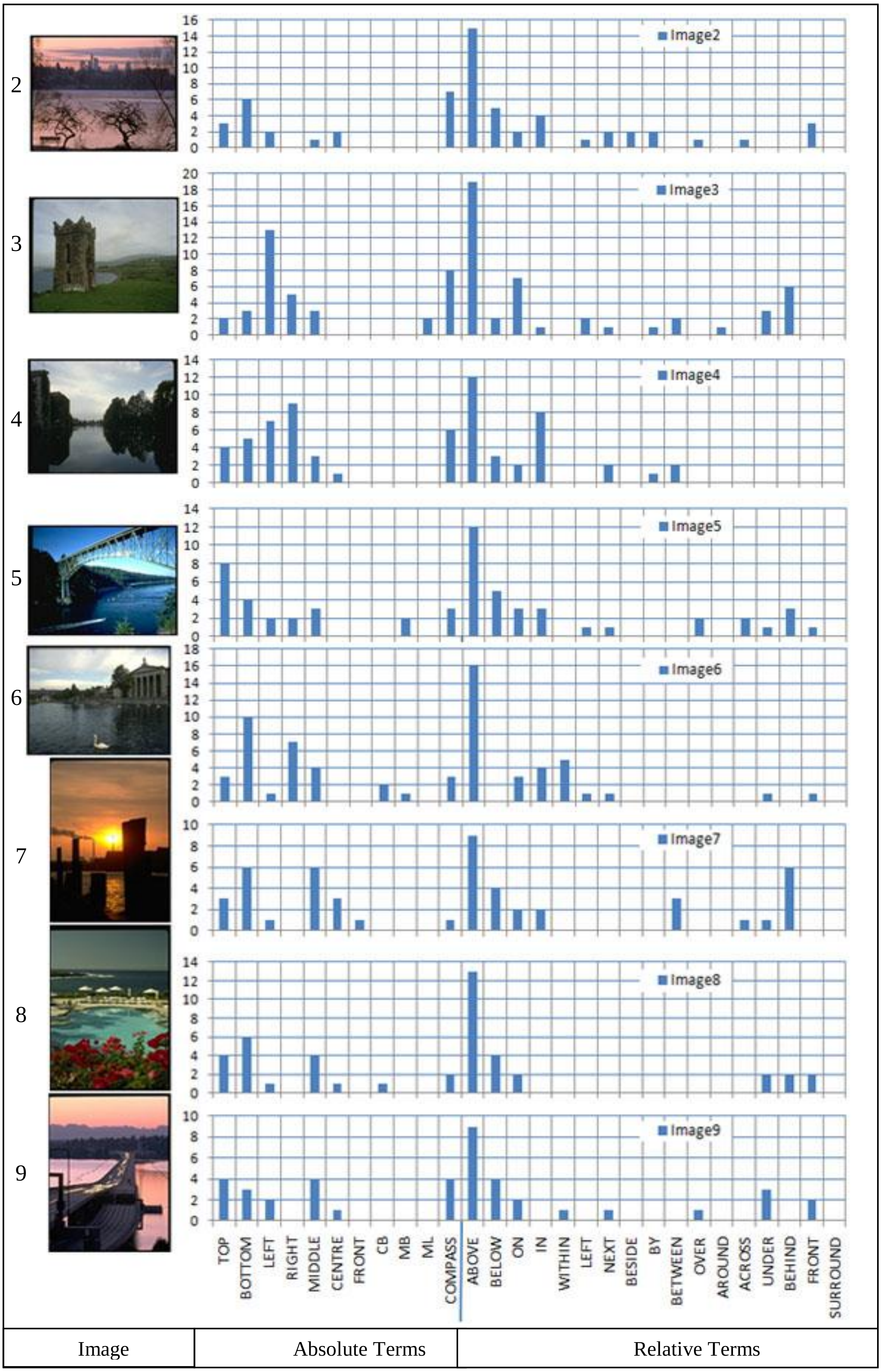


Figure 4-2 Histogram for Image 2-9

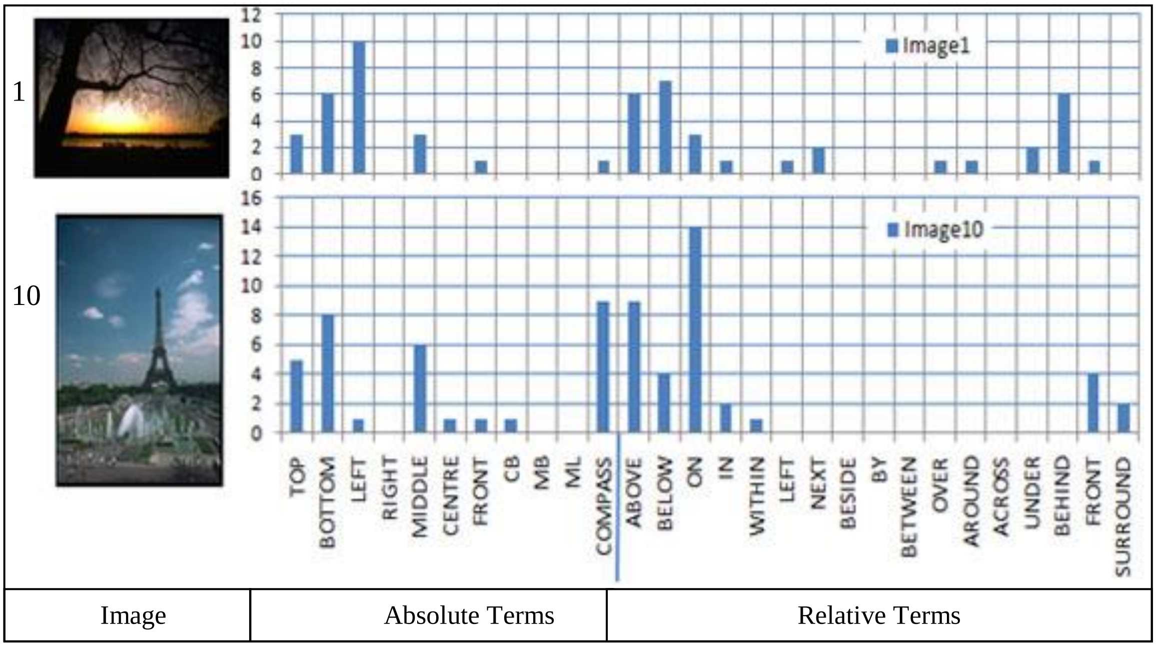


Figure 4-3 Histogram for Image 1 and 10

For both spatial relationships, Table 4-3 show the most frequently used term for absolute and relative terms in all images. Although Image 2 and Image 10 show highest histogram values for Compass directions, this is still not considered as the most frequently used as its consist 8 directions terms, where in Image 2 the 7 frequencies are a total of 1 for East, West and South-West; and 2 for North and South. In Image 10, the 9 frequencies are a total of 1 for North and 2 for South, East, West and South-West.

Table 4-3 Most frequently used term by image.

Image	Most Frequent Absolute Term	Most Frequent Relative Term
1	LEFT	BELOW
2	BOTTOM	ABOVE
3	LEFT	ABOVE
4	RIGHT	ABOVE
5	TOP	ABOVE
6	BOTTOM	ABOVE
7	BOTTOM and MIDDLE	ABOVE
8	BOTTOM	ABOVE
9	TOP and MIDDLE	ABOVE
10	MIDDLE	ON

In general, all users used the objects suggested in the survey with an emphasis on some obvious objects when describing and annotating the absolute spatial terms used for the objects in the images; for example, the tree in Image 1 and 2, where the term LEFT (absolute) is the highest commonly used in Image 1 and BOTTOM (absolute) is highest commonly used in Image 2, probably because of the dominant position of the tree in both images. And if we look closely these patterns seem to be true for all other images as well, including the tower or castle in Image 3, the bridge in Image 5 and 9, the swan in Image 6, the building in Image 7, the flowers in Image 8 and the Eiffel Tower in Image 10, which produced the result as shown in Table 4-3.

Other than referring to the most obvious object in the image, the way humans look at an image may also vary and affect their way of describing the image. As this survey uses a screen shot with scroll up-to-down, it may also cause the users to use the ABOVE term more compared to BELOW when annotating the objects in the image. Other factors may be related to ethnography, a qualitative method aimed to learn and understand cultural phenomena which may for example explain why the term LEFT is sometimes used in preference to the term RIGHT. Ethnography can describe the nature of people (user) through their writing (Philipsen, 1992), for example how they write in their native language. Europeans and some Asians write left-right, Chinese write top-down, while Arabs write right-left.

To understand and analyse all these responses is quite challenging as the results reveal some interesting facts that show the variations of users' perceptions even when looking at the same image. Some interesting notation made by the users, for example is the term BEHIND (relative), which has been used a number of times in Image 1, but sometimes it is used differently such as "The sun is behind the beach" and "The water behind the land", this might be because users consider layers when looking at the image but it is certainly a reference to the 3-D world rather than the 2-D image plane.

In other examples, such as in Image 5, user descriptions are slightly diverse. Some examples are "the steel is in the bridge" which is not related to spatial relations in the image, and "the bridge lies on the left to the right" which shows how the spatial terms are sometimes used in unusual ways by users to express their description. As for Image 6, another diverse description is, "The swan is on the centre of the bottom".

However, the object water suggested for Image 7 was never used by the users, and this might be because of other suitable annotations such as lake which they felt was more appropriate to be used. As a result, users have added more objects to the images for annotation, but these are not listed in the objects column. This happened to most of the images annotations except for Image 6. The objects included were such as horizon and branches in Image 1, bench and skyline in Image 2 etc. These objects might be more appropriate for use based on the users' perspectives and preferences.

4.3.3 Correlations Based on Terms

Here we consider correlations between terms assigned to images in order to explore whether significant relations between images can be discovered based on the spatial terms used by the respondents. The Correlation coefficient is a statistical measure that can show how strongly pairs of variables are related. The formula for correlation (r) is given as below (Trochim, 2006).

Let $\mathrm{A_{ij}}$ be the number of times the spatial terms *i* is used in image j,

$1 \leq \mathrm{i} \leq \mathrm{N}$, where N is the number of spatial terms used.

And $\mathrm{r_{jk}}$ be the correlation between image j and image k.

$$\mathrm{r_{jk}} = \frac{\mathrm{N}\sum_{i=1}^{N} \mathrm{A_{ij}}\,\mathrm{A_{ik}} - (\sum_{i=1}^{N} \mathrm{A_{ij}})(\sum_{i=1}^{N} \mathrm{A_{ik}})}{\sqrt{[\mathrm{N}\sum_{i=1}^{N} \mathrm{A_{ij}}^{2} - \left(\sum_{i=1}^{N} \mathrm{A_{ij}}\right)^{2}][\mathrm{N}\sum_{i=1}^{N} \mathrm{A_{ik}}^{2} - (\sum_{i=1}^{N} \mathrm{A_{ik}})^{2}]}}$$

From the analysis in Table 4-2, correlation between two images has been calculated and the Correlation Matrix is shown in Table 4-4. Each single value describes the degree of relationship between spatial terms used in describing the two given images. There are 45 pairs of correlation coefficient values for the 10 images. This can be calculated as follows:

Number of pairs = N(N-1)/2, where N is the number of variables.
= 10(10-1)/2
= 45 pairs.

Table 4-4 Correlation Matrix

Image	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
1	1.000									
2	0.452	1.000								
3	0.662	0.683	1.000							
4	0.464	0.715	0.741	1.000						
5	0.571	0.804	0.671	0.691	1.000					
6	0.326	0.741	0.632	0.762	0.713	1.000				
7	0.612	0.627	0.560	0.498	0.735	0.611	1.000			
8	0.581	0.852	0.691	0.597	0.874	0.795	0.837	1.000		
9	0.560	0.831	0.689	0.601	0.841	0.683	0.692	0.915	1.000	
10	0.370	0.622	0.504	0.446	0.590	0.553	0.512	0.639	0.681	1.000

The strength and significance of a correlation coefficient is measured with the strongest positive correlation being 1.0, and the closer the value is to 1.0 the stronger the correlation between the two sets of values analyzed. The following are general categories to indicate a quick way of interpreting the value:

- 0.0 to 0.2 Very weak to negligible correlation
- 0.2 to 0.4 Weak, low correlation (not very significant)
- 0.4 to 0.7 Moderate correlation
- 0.7 to 0.9 Strong, high correlation
- 0.9 to 1.0 Very strong correlation

The matrix shows that there are very strong correlations between Image 8 and Image 9; and strong/high correlations between Image 5 and Image 8 and between Image 2 and Image 8. The highest correlation value is 0.915 where there is a very high significance between Image 8 and Image 9. In Figure 4-4, for the pair of highly correlated images (Image 8-9, Image 2-8 and Image 5-8), the bar charts show the frequencies of each of the term in order to see visually where the correlation occurs.

The figure shows that there is a similarity of number and type of terms used by the users in describing both images. This might be because the location and position of the main or obvious objects referred in the spatial description for both images are quite similar, for example, the object: sky in those images is located above all other objects while most of other objects are at the bottom of the images. The composition of objects in Image 2 and Image 8 is quite comparable, though both involved different objects.

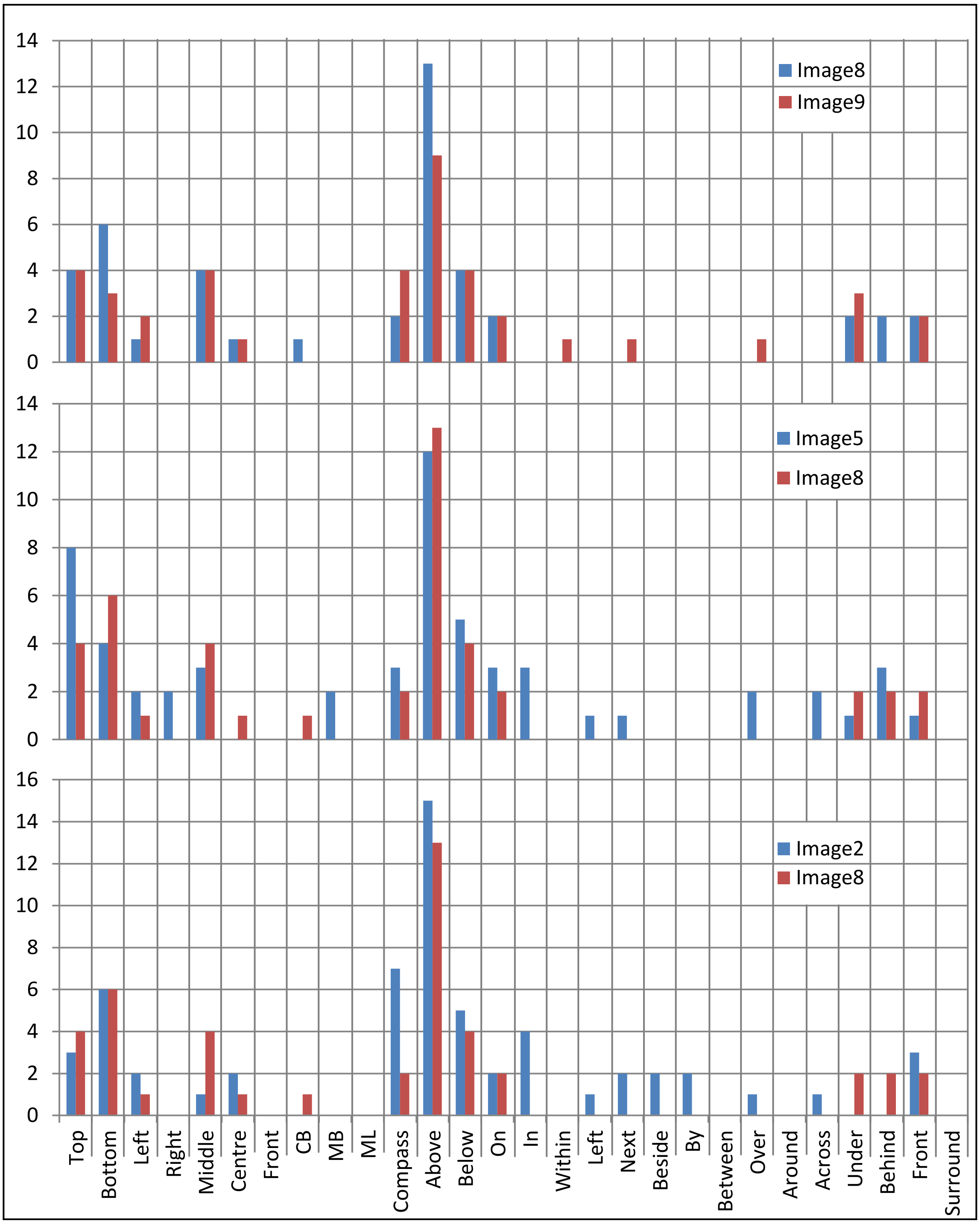


Figure 4-4 Charts for Image 8-9, Image 2-8 and Image 5-8

It also can be seen in the charts that the correlation occurs when similar terms used are TOP, BOTTOM, MIDDLE, ABOVE, BELOW, ON and FRONT.

4.4 Discussion and Conclusion

Based on previous research, spatial relationship terms have been recognised for consideration in our research. To decide which spatial terms to use, we designed a preliminary survey to identify human aspects in describing objects in images with spatial terms. The survey results showed that the most commonly used spatial terms by the respondents as prospective users in perceiving and describing those images were BOTTOM for absolute relation and ABOVE for relative relation although a wide range of other terms were also being used.

However, there might be some bias occurring during the survey. The bias may due to the ‘rule of the thirds’ applied in those images, because this is the golden rule used by professional photography and the images used in the survey are from professional image dataset. The rule states that an image should be imagined as divided into nine equal parts by two equally-spaced horizontal lines and two equally-spaced vertical lines, and that important compositional elements should be placed along these lines or their intersections (Peterson, 2003).

Also, the use of a given image with examples of sentences that contain the spatial terms also may lead users to use a certain term or limited terms or vocabularies. The survey could be improved by adding more type of images and avoiding giving any example in order to give more freedom for the users to use any spatial words they like or think of. Hence the responses would not just be limited to the words used in the given example for the first image. In this case we can reduce any bias occurring when the users write their description or annotation about spatial relations among the objects in the images.

It should also be recognised that the number of images considered and the number of users in the trial were both quite low. Larger samples would certainly improve the applicability of the conclusions.

In spite of these limitations, the survey results do show that there are a number of different ways in which each of the users used spatial terminology in describing an image. Analysing these responses is quite challenging for example to cater for the various styles of language that have been used in sentences describing the images.

From this particular analysis, it is found that 45 spatial terms have been used by the users where 28 terms have been used more than once in describing the absolute and relative positions of objects in the images. The spatial terms include compass directions.

From the results and analysis of the survey, the spatial relationship terms for absolute relation: LEFT, BOTTOM, TOP and MIDDLE; and for relative relation: ABOVE, BELOW are among the most frequently used terms, while relative relation: ON has some complicated aspects which should be measured and reconsidered. Therefore these terms and reciprocals, ABOVE-BELOW, TOP-BOTTOM and LEFT-RIGHT are selected for the initial development of this research.

Therefore, in order to enhance existing image descriptions by annotating with absolute and relative spatial terms for regions or objects, the next task is to develop algorithms to compute and extract each of the spatial terms automatically from images. This task will be presented in the next chapter.

Chapter 5

The Development of Spatial Relationship Algorithms

5.1 Introduction

We have seen in earlier chapters that useful descriptions of an image not only contain the names of important features or objects within the image but also information about their absolute and relative positions. In the previous chapter we identified a range of relative and absolute spatial terms which people use to describe images. An automatic system to extract such full descriptions of images might first begin with an automatic object recognition stage to identify and label the objects occurring in the image. Building such a system is not yet possible in the general case although much research on object recognition is being undertaken, with some success in limited domains.

In Chapter 3 some semi-automatic tools and techniques for segmenting and labelling image regions or objects were introduced and investigated. LabelMe was decided upon, not just because it obtained the highest mark in the investigation, but also because this semi-automatic tool provides a substantial foundation as the base technology, annotates images locally and the annotation description is in an understandable form and could be employed for further development in image annotations and retrieval. Then, a second stage might take the labelled object information from an image, extract the spatial positions and relationships and then assign the appropriate spatial descriptions to the image. Creating this second stage is the aim of the work in this chapter. In particular the aim is to create a small knowledge base about the image in which assertions are made about the objects in the image, their absolute positions in the image and their positions relative to each other.

The algorithms in this chapter are categorised into absolute position terms and relative terms. The relative terms will be divided into basic relative terms and also some composite relative terms. The starting point for the algorithms is the output from the semi-automatic annotation tool, LabelMe (Russell et al., 2008), which uses supervised segmentation and user interaction to produce labelled image regions which can correspond to image objects together with their names. The region data, in the form of x and y coordinates represent each point marked for the boundary of the object provided from the annotation tool, are an important part of the input to the spatial analysis system.

5.2 Extracting Spatial Information

As referred to in the research framework in

Figure 3-11 in subsection 3.9.1, the annotation component presumes that a preliminary segmentation and region annotation stage has provided relevant image regions, represented by the coordinates of pixels along their boundaries, and region labels indicating the object represented by the region. This stage has been done semi-automatically by using the LabelMe annotation tool (Russell et al., 2008). We refer to the labelled regions as objects, and extending the approach of Hollink et al., (2004), automatically extract spatial descriptors for the relative spatial relations between pairs of objects in images, and the absolute positions of individual objects within the images.

As mentioned, LabelMe provides an image as a collection of labelled objects which we describe as follows:

1. Assume that a given image$_\mathrm{i}$ ($\mathrm{I_i}$) consists of multiple labelled objects (O):

   $\mathrm{I_i} = \{\mathrm{O_1}, \mathrm{O_2} ... \mathrm{O_N}\}$
2. Each of the objects has a set of boundary coordinates (in XML from LabelMe) that will be used to compute the spatial information between the object and the

other objects in the image. For example in the image suppose we have N objects of interest, so each object is represented by:

$$\mathrm{Object}_1 = \{(x_1^1,y_1^1), (x_2^1,y_2^1),\ldots,(x_n^1,y_n^1)\}$$
$$\mathrm{Object}_2 = \{(x_1^2,y_1^2), (x_2^2,y_2^2),\ldots,(x_n^2,y_n^2)\}$$
$$:$$
$$\mathrm{Object}_N = \{(x_1^N,y_1^N), (x_2^N,y_2^N),\ldots,(x_n^N,y_n^N)\}$$

In our system the computation of spatial relationships between objects in the image proceeds as follows:

3. The output from LabelMe in XML is converted into Excel where the averages of the objects' x and y coordinates are calculated to provide an approximate value for the centroid of each object, (x_c, y_c). With LabelMe, this is the a straightforward way to calculate a representative point position for the object and in recognising the wide variety of shapes of objects this is better than using the centre of the bounding box (Hollink et al., 2004) or just choosing a single spatial location point randomly (Lee and Hwang, 2002, Lee et al., 2004, Lee et al., 2006). However, a more careful consideration of a representative point and bounding box could certainly be made. For example different object classes may benefit from different approaches. People are mainly determined by face recognition and so on but here this relatively simple and uniform approach was used in the interests of time.

4. This object's centroid and other information will be inserted into our system, where the height and width of the object's bounding box is computed for further computation.

5. In the algorithms, all relations between pairs of objects in the image are defined by computing and comparing the centroids and borders of bounding boxes of the two objects The method of using a mathematical bounding box is often applied in research in image retrieval when segmenting or annotating a region or object in images, however the validity of this may be different when computing human bounding boxes, because in some annotation tools like Flickr, humans are annotated by detecting the head and face as used in some of current cameras. This perceptual human bounding box needs more consideration and further research but must be left as future work.

6. All object positions within the image are defined using the centroids of the objects in the image.

5.3 Spatial Relationship Algorithms for Relative Position

Relative position is an orientation relationship describing where objects are placed relative to one another. In some cases, one of the objects acts as a reference to specify the position of the other objects. This relationship is sometimes referred to as directional relationships and is more useful to describe objects in an image than topological relationships (Zhou et al., 2001). The relative positions between pairs of objects in images are computed based on the object centroids and their bounding rectangles. We use the approximate *centroid* of the object rather than a centre of the bounding box used by Hollink et al. (2004) or a single spatial location point used by Lee & Hwang (2002) and Lee et al. (2004, 2006) as in some cases the use of centroid will be more meaningful, for example when dealing with a triangular pyramid shaped tower or in a more extreme case, a car with a long radio aerial.

The relative positions that we extract are *above, below, left of, right of* and the composite relations produced by integrating these basic spatial relationship which produce *above and to the left of, above and to the right of', below and to the left of* and *below and to the right of*. The width is used in the *above* and *below* concepts and the height is used in the *left of* and *right of* concepts respectively.

5.3.1 Spatial Relationships of 'Left of' and 'Right of'

The definitions for the *left of* and *right of* terms use the height (2h) of each object concerned to ensure that we only indicate an object is left or right of another if they are at approximately the same level in the image. To ensure this, we require that the difference between the y-values of the centroids should be less than half the sum of the object heights. Also we know that the relative positions *left of* and *right of* are reciprocal relations. If A is *left of* B, then conversely B is *right of* A etc. Therefore the two spatial terms can be asserted using the same rule. Using the terminology visualized in Figure 5-1 for the rules for inferring the *left of* (Hollink et al., 2004), we then inferred the reciprocal rules for the *right of* relations, defined as follows:

1. IF (($xc_1 < xc_2$) AND (($h_1 + h_2$) > $|yc_1 - yc_2|$)) THEN Object_1 is LEFT of Object_2, AND Object_2 is RIGHT of Object_1.

   OR

2. IF (($xc_1 > xc_2$) AND (($h_1 + h_2$) > $|yc_1 - yc_2|$)) THEN Object_1 is RIGHT of Object_2, AND Object_2 is LEFT of Object_1.

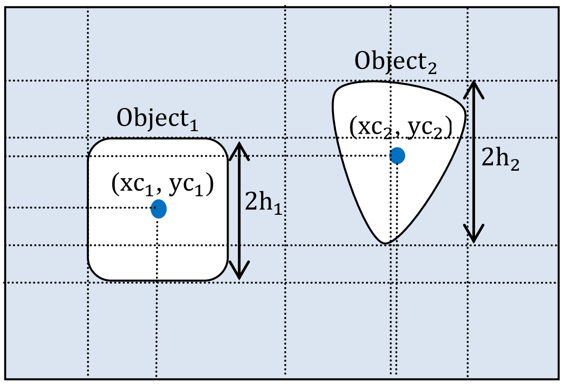


Figure 5-1 Computation of *'$Object_1$ is on the Left of $object_2$'* relation (Adapted from Hollink et al. (2004))

5.3.2 Spatial Relationships of 'Above' and 'Below'

The spatial term *above* is the highest frequency term used by the users during our initial survey in Chapter 4. In developing the algorithm for the term *above*, we also considered its reciprocal, *below*, so if A is above B, then conversely B is below A. The calculation of the relative position for the *above* and *below* terms use the width (2w) of each object involved to ensure that we only indicate an object is above or below another if they are in approximately the same left-right position. By analogy with the rules for left and right, we can define the rules for inferring *above* and *below* using the notation in Figure 5-2 as follows:

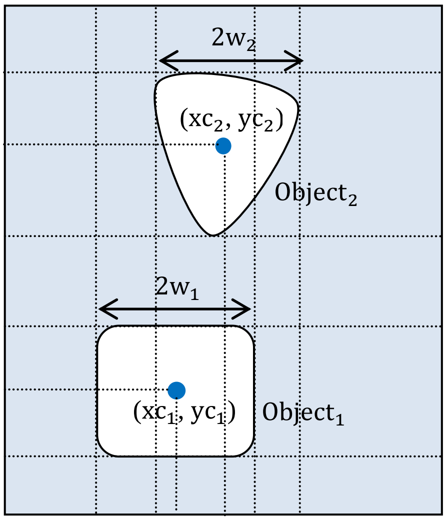


Figure 5-2 Computation of '$Object_2$ *is Above* $object_1$' relation.

The spatial concept for 'Above' and 'Below' are described as follows:

1. IF (($yc_1 > yc_2$) AND (($w_1 + w_2$) > $|xc_1 - xc_2|$)) THEN Object_1 is ABOVE Object_2, AND Object_2 is BELOW Object_1.

   OR
2. IF (($yc_1 < yc_2$) AND (($w_1 + w_2$) > $|xc_1 - xc_2|$)) THEN Object_1 is BELOW Object_2, AND Object_2 is ABOVE Object_1.

5.3.3 *Composite Spatial Relationships*

The rules in the previous section capture the relation when one object is directly to the left of another or directly above. To capture all relative positions which one object may be in with respect to another, we need composite relation positions between objects such as *above and to the left of* or *below and to the right of*. By integrating previous rules, we define rules for composite relations. An example of the composite spatial relationships is illustrated in Figure 5-3.

The rules of composite spatial relationships computations of object_1 relative to object_2 are defined as follows:

1. IF ((xc_2 - xc_1) ≥ ($w_1 + w_2$) AND ($yc_2 - yc_1$) ≥ ($h_1 + h_2$)) THEN Object_2 is ABOVE and to the RIGHT of Object_1, AND Object_1 is BELOW and to the LEFT of Object_2.

2. IF $((xc_2 - xc_1) \geq (w_1 + w_2)$ AND $(yc_1 - yc_2) \geq (h_1 + h_2))$ THEN Object_2 is BELOW and to the RIGHT of Object_1, AND Object_1 is ABOVE and to the LEFT of Object_2.
3. IF $((xc_1 - xc_2) \geq (w_1 + w_2)$ AND $(yc_1 - yc_2) \geq (h_1 + h_2))$ THEN Object_1 is ABOVE and to the RIGHT of Object_2, AND Object_2 is BELOW and to the LEFT of Object_1.
4. IF $((xc_1 - xc_2) \geq (w_1 + w_2)$ AND $(yc_2 - yc_1) \geq (h_1 + h_2))$ THEN Object_1 is BELOW and to the RIGHT of Object_2, AND Object_2 is ABOVE and to the LEFT of Object_1.

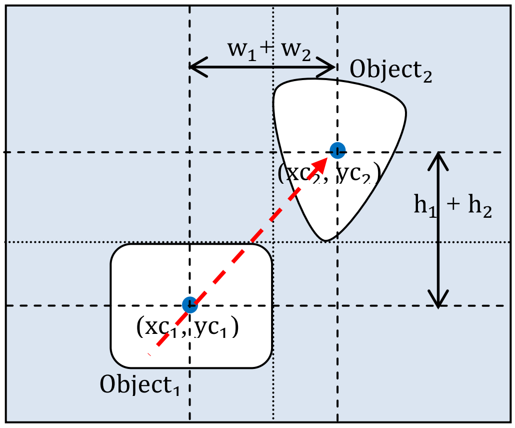


Figure 5-3 Computation of composite concept between object_1-object_2.

5.4 Spatial Relationships Algorithms for Absolute Position

In addition to the relative spatial terms between objects in the image, we also extract the absolute positions of the objects in the image. For absolute position, we use a finer grained grid than Hollink et al., (2004) and use a different notation. Hollink et al. (2004) used compass point positions defined on a 3x3 grid which is sometime seen as more suitable for geographical or topological representation.

We divide the image into a 5x5 grid defining 25 absolute spatial annotations such as object A is *in the middle of the bottom* or object B is *at the far right and at the top*. Some of these terms were used by the respondents during our initial spatial survey in Chapter 4 which emphasizes the importance of this type of spatial information. At the same time, we can cater for the more precise versions of spatial concepts like *far right* mentioned by Hollink et al., (2004) but not present in their implementation.

In addition to specifying an object's position based on a combination of the horizontal and vertical grid, these terms also allow us to specify 5 absolute horizontal positions and 5 absolute vertical positions individually such as *at the very top* or *at the far left* independently. The total of possible spatial positions is therefore 35 ie 10 for separate vertical and horizontal positions and 25 for the combined positions. All these terms for absolute position have been defined and implemented. They are computed and asserted in our knowledge base by the spatial system where appropriate.

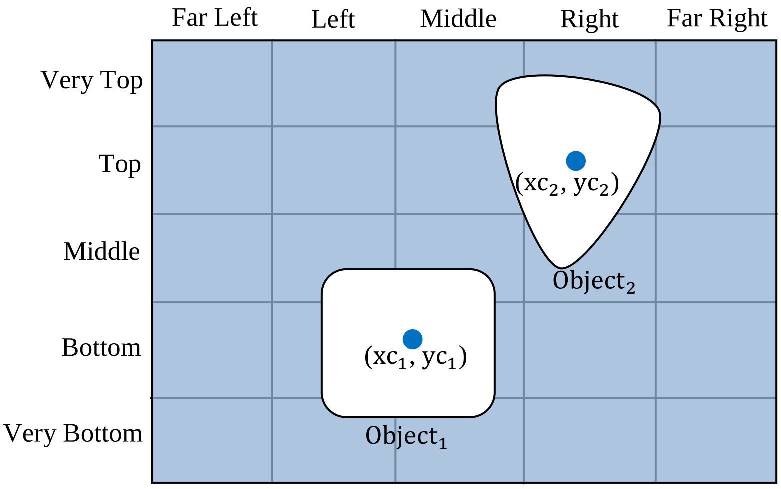


Figure 5-4 Absolute position concepts of object$_1$ and object$_2$ in the image.

5.5 The Implementation of the Algorithms

Each of the spatial information extraction rules described in section 5.3 and 5.4 has been implemented and can be applied to labelled image segmentations derived from the first stage of our framework. The implemention of the algorithms for the spatial term computation has been done using MatLab. This spatial analysis system executes all the algorithms by accepting an input from the LabelMe software of object labels and perimeter coordinates in the form of a text file along with some other important information for generating the spatial relationships for all labelled objects in the image. At this stage a sample image has been chosen to show how the algorithms work in producing the spatial relationships automatically. The output resulting from the extraction and annotation process is a series of statements providing spatial information about the objects in the image. These statements are asserted in a small knowledge base about the image.

5.5.1 *Image Annotation and Coordinates Information*

In LabelMe, objects annotated in an image are displayed in the right pane of the image shown in Figure 5-5. The coordinates of each object are captured in an XML file and can be read by clicking the XML link on the page.

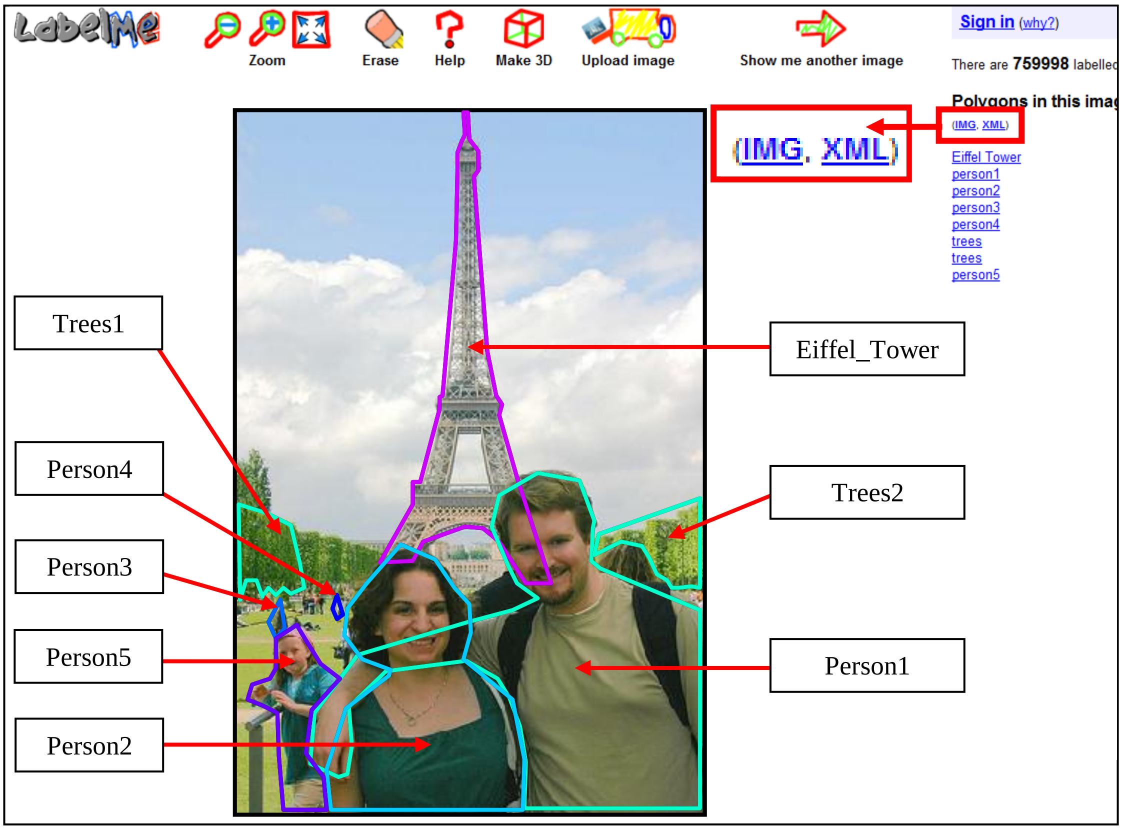


Figure 5-5 A sample of image annotated in LabelMe.

The XML file shows that, each object element is described by its name, date, id, username and the object’s boundary (x, y) coordinates points from the annotation process. An example of an annotation code for the object ‘Eiffel_Tower’ in Figure 5-5 is shown in Figure 5-6.

```
<object>
<name>Eiffel Tower</name>
<date>30-Sep-2008 00:18:46</date>
<id>0</id>
−<polygon>
<username>anonymous</username>
−<pt><x>103</x><y>322</y></pt>
−<pt><x>127</x><y>281</y></pt>
−<pt><x>127</x><y>265</y></pt>
−<pt><x>132</x><y>265</y></pt>
−<pt><x>146</x><y>213</y></pt>
−<pt><x>146</x><y>204</y></pt>
−<pt><x>148</x><y>203</y></pt>
−<pt><x>157</x><y>92</y></pt>
−<pt><x>158</x><y>43</y></pt>
−<pt><x>158</x><y>29</y></pt>
−<pt><x>162</x><y>21</y></pt>
−<pt><x>164</x><y>17</y></pt>
−<pt><x>163</x><y>1</y></pt>
−<pt><x>166</x><y>1</y></pt>
−<pt><x>167</x><y>20</y></pt>
−<pt><x>173</x><y>28</y></pt>
−<pt><x>173</x><y>42</y></pt>
−<pt><x>170</x><y>46</y></pt>
−<pt><x>180</x><y>171</y></pt>
−<pt><x>186</x><y>203</y></pt>
−<pt><x>190</x><y>210</y></pt>
−<pt><x>188</x><y>217</y></pt>
−<pt><x>201</x><y>262</y></pt>
−<pt><x>226</x><y>337</y></pt>
−<pt><x>207</x><y>337</y></pt>
−<pt><x>188</x><y>308</y></pt>
−<pt><x>177</x><y>298</y></pt>
−<pt><x>164</x><y>297</y></pt>
−<pt><x>134</x><y>307</y></pt>
−<pt><x>127</x><y>321</y></pt>
</polygon>
</object>
```

Figure 5-6 Annotation code for the object: '*Eiffel_Tower'*.

5.5.2 *An Example of the Implementation*

The implementation of the spatial extraction process is discussed further by using a sample image. As an example, the same image in Figure 5-5 that has been segmented and labelled using the semi-automatic LabelMe software (Russell et al., 2008) is used. To simplify our presentation, we only consider a subset of objects in this image.

The coordinates of the boundary pixels of the labelled objects named Eiffel_Tower, Person1, Person2, Person3, Person 4 and Tree have been extracted. The extraction includes coordinates of $x_{\max}$, $y_{\max}$, $x_{\min}$, $y_{\min}$ and all boundary x and y coordinates. The approximate centroid of each labelled object will be calculated for further spatial annotation computation. These coordinates for each object with its label serve as an input to be used in the spatial analysis system that has been developed.

The input is gathered in the form of a data matrix in a text file, so that it can be read automatically by MatLab. The data matrix for the chosen image is shown in Figure 5-7, consist of labelled region/object names, centroid of x and y, $x_{\max}$, $y_{\max}$, $x_{\min}$, $y_{\min}$ and real order of magnitude height (in metres) from (2008) of each object in the image (which is used in later computation).

Eiffel_Tower	163.6	178.7	226	337	103	1	324
Person1	182.2068966	381.1034483	332	498	53	258	1.683
Person2	127.6111111	395.8888889	206	499	65	310	1.683
Person3	29.5	365	35	377	23	349	1.683
Person4	72.25	356	76	363	69	346	1.683
Tree	25.15384615	328.0769231	49	349	2	281	10
Trees2	300.6666667	321.5833333	332	341	256	277	10
Person5	38.94444444	423.0555556	74	499	8	367	1.683

Figure 5-7 Content of input from text file.

Further information in the spatial analysis includes the name of the image, the size of the image and number of objects to be considered as shown in Figure 5-8. A segment of the implementation of relative position spatial relationships is shown using pseudocode in Figure 5-8.

```
Input image name;
Input number of objects in the image;
Input image size;
Read TEXT file

FOR each object
    Calculate the width of object$_i$ ($w_i$);
    Calculate the height of object$_i$ ($h_i$);
    //Compare Object$_1$and Object$_2$ in Condition1
    IF (($xc_1 < xc_2$) AND (($h_1 + h_2$) > $|yc_1 – yc_2|$)) THEN Object$_1$ is left of Object$_2$, AND Object$_2$ is
      right of Object$_1$;
    ELSEIF (($xc_1 > xc_2$) AND (($h_1 + h_2$) > $|yc_1 – yc_2|$)) THEN Object$_1$ is right of Object$_2$, AND
      Object$_2$ is left of Object$_1$;
    ENDIF

    //Compare Object$_1$and Object$_2$ in Condition2
    IF (($yc_1 > yc_2$) AND (($w_1 + w_2$) > $|xc_1 – xc_2|$)) THEN Object$_1$ is above Object$_2$, AND Object$_2$
      is below Object$_1$;
    ELSEIF (($yc_1 < yc_2$) AND (($w_1 + w_2$) > $|xc_1 – xc_2|$)) THEN Object$_1$ is below Object$_2$, AND
      Object$_2$ is above Object$_1$;
    ENDIF

    //Compare Object$_1$and Object$_2$ in Condition3
    IF (($xc_2$ - $xc_1$) ≥ ($w_1 + w_2$) AND ($yc_2 – yc_1$ ) ≥ ($h_1 + h_2$)) THEN Object$_2$ is above and to the
      right of Object$_1$, AND Object$_1$ is below and to the left of Object$_2$;
    ELSEIF (($xc_2$ - $xc_1$) ≥ ($w_1 + w_2$) AND ($yc_1 – yc_2$ ) ≥ ($h_1 + h_2$)) THEN Object$_2$ is below and to
      the right of Object$_1$, AND Object$_1$ is above and to the left of Object$_2$;
    ELSEIF (($xc_1$ - $xc_2$) ≥ ($w_1 + w_2$) AND ($yc_1 – yc_2$ ) ≥ ($h_1 + h_2$)) THEN Object$_1$ is above and to
      the right of Object$_2$, AND Object$_2$ is below and to the left of Object$_1$;
    ELSEIF (($xc_1$ - $xc_2$) ≥ ($w_1 + w_2$) AND ($yc_2 – yc_1$ ) ≥ ($h_1 + h_2$)) THEN Object$_1$ is BELOW and
      to the right of Object$_2$, AND Object$_2$ is above and to the left of Object$_1$;
ENDFOR
```

Figure 5-8 Segment of the pseudocode for relative position.

The pseudocode shows the simplified steps undertaken to generate the spatial relationships for relative position between pairs of objects in an image automatically.

5.6 Result and Discussion

The resulting statements from the computation of the spatial information for the labelled and selected objects are generated in a text file. A part of the result for spatial relative position relationships is shown in Figure 5-9.

The result shows that the automatic spatial analysis system is working as expected and the output of the computation is a series of assertions about the relative spatial relationships between the selected labelled objects in the image.

Output segment

```
IMAGE NAME:Statue_of_Liberty_Paris_000000254
Regions Name: Eiffel_Tower, Person1, Person2, Person3, Person4, Trees1,
Trees2, Person5,

SPATIAL ANNOTATION
==================

RELATIVE POSITION:

 Eiffel_Tower is LEFT of Person1, and Person1 is RIGHT of Eiffel_Tower.
 Eiffel_Tower is ABOVE Person1, and Person1 is BELOW Eiffel_Tower.
 Eiffel_Tower is RIGHT of Person2, and Person2 is LEFT of Eiffel_Tower.
 Eiffel_Tower is ABOVE Person2, and Person2 is BELOW Eiffel_Tower.

 Eiffel_Tower is ABOVE and to the RIGHT of Person3, and Person3 is BELOW and
  to the LEFT of Eiffel_Tower.

 Eiffel_Tower is ABOVE and to the RIGHT of Person4, and Person4 is BELOW and
  to the LEFT of Eiffel_Tower.
 Eiffel_Tower is RIGHT of Trees1, and Trees1 is LEFT of Eiffel_Tower.
 Eiffel_Tower is LEFT of Trees2, and Trees2 is RIGHT of Eiffel_Tower.

 Eiffel_Tower is ABOVE and to the RIGHT of Person5, and Person5 is BELOW and
  to the LEFT of Eiffel_Tower.
 Person1 is RIGHT of Person2, and Person2 is LEFT of Person1.
 Person1 is ABOVE Person2, and Person2 is BELOW Person1.
 Person1 is RIGHT of Person3, and Person3 is LEFT of Person1.
 Person1 is RIGHT of Person4, and Person4 is LEFT of Person1.
 Person1 is BELOW Person4, and Person4 is ABOVE Person1.
 Person1 is RIGHT of Trees1, and Trees1 is LEFT of Person1.
 Person1 is BELOW Trees1, and Trees1 is ABOVE Person1.
 Person1 is LEFT of Trees2, and Trees2 is RIGHT of Person1.
 Person1 is BELOW Trees2, and Trees2 is ABOVE Person1.
 Person1 is RIGHT of Person5, and Person5 is LEFT of Person1.
 Person1 is ABOVE Person5, and Person5 is BELOW Person1.
 Person2 is RIGHT of Person3, and Person3 is LEFT of Person2.
 Person2 is RIGHT of Person4, and Person4 is LEFT of Person2.
 Person2 is BELOW Person4, and Person4 is ABOVE Person2.
 Person2 is RIGHT of Trees1, and Trees1 is LEFT of Person2.
```

```
Person2 is LEFT of Trees2, and Trees2 is RIGHT of Person2.
Person2 is RIGHT of Person5, and Person5 is LEFT of Person2.
Person2 is ABOVE Person5, and Person5 is BELOW Person2.
Person3 is LEFT of Person4, and Person4 is RIGHT of Person3.
Person3 is RIGHT of Trees1, and Trees1 is LEFT of Person3.
Person3 is BELOW Trees1, and Trees1 is ABOVE Person3.
Person3 is LEFT of Trees2, and Trees2 is RIGHT of Person3.
Person3 is LEFT of Person5, and Person5 is RIGHT of Person3.
Person3 is ABOVE Person5, and Person5 is BELOW Person3.
Person4 is RIGHT of Trees1, and Trees1 is LEFT of Person4.
Person4 is LEFT of Trees2, and Trees2 is RIGHT of Person4.
Person4 is RIGHT of Person5, and Person5 is LEFT of Person4.
Person4 is ABOVE Person5, and Person5 is BELOW Person4.
Trees1 is LEFT of Trees2, and Trees2 is RIGHT of Trees1.
Trees1 is LEFT of Person5, and Person5 is RIGHT of Trees1.
Trees1 is ABOVE Person5, and Person5 is BELOW Trees1.

Trees2 is ABOVE and to the RIGHT of Person5, and Person5 is BELOW and to the
LEFT of Trees2.
```

Figure 5-9 Sample of a part of output result for relative position.

It can be seen that many useful spatial annotations are generated. These statements have also been generated in the form of an RDF file in turtle format. The output of the RDF file will be tailored based on two ontologies that have been developed in the Ontology Component called the Spatial Relationships Ontology and the Place of Interest Ontology. This aspect of the work is described in 0.

Although the implementation of the algorithms are working successfully for the given image sample, there are many potential problems with extracting spatial relations for objects and many other aspects and alternatives which could be considered in developing the algorithms to infer the associated spatial terms. Highly unusual shapes so that objects surround other objects or appear within other objects can cause problems of interpretation and partial occlusion can mislead the algorithms. One alternative may be to consider the detail of the object border rather than the bounding box and centroid when generating the relations for example. If time had permitted it would be good to try these alternatives and choose the best one but as a proof of concept the chosen approach works well.

5.7 Conclusion

In this chapter the design and implementation of enhanced approaches to spatial information extraction using labelled segmented images has been presented. Extraction rules for both absolute and relative spatial terms have been devised and an example of their use on a sample image has been shown. The spatial analysis system has been shown to successfully produce an annotation of the spatial positions of objects and relationships between labelled objects in a sample image automatically.

In total, we extract 43 spatial terms, including 8 relative spatial terms (left, right, above, below and the four composite terms) together with 35 fine-grained absolute spatial positions. These were based on 10 individual spatial locations of horizontal and vertical positions and the 25 combined positions in the 5 by 5 grid. The ability to extract these spatial terms automatically from labelled segmented image objects and express them as assertions about the image them in a small knowledge base has been demonstrated. Although the algorithms have been demonstrated on a sample image, potential problems with objects with complex shapes and structures or with partial occlusions are recognised.

The spatial analysis system could be enhanced and expanded further to extract a wider vocabulary of spatial terms as discussed in Chapter 4 by using added information concerning the order of magnitude heights of objects. This added information about an object could be stored in and retrieved from the object's properties via the domain ontology. Later we investigate a retrieval front end to enable image queries, which include spatial information and could be made more flexible via spatial terminology held in the ontology.

In the next chapter we investigate how more complex spatial terms can be extracted by exploiting additional information about objects in the image.

Chapter 6

Advanced Spatial Relationships

6.1 Introduction

In this chapter some spatial relationships are introduced which require additional knowledge of the objects in the image. The relative sizes of objects in an image are related to their actual size in the real world and their distances from the camera. If the sizes are known, it may be possible to infer some information about the distances from the camera. For example, we may be able to infer that one object is much nearer the camera than the other or that the two objects are a similar distance from the camera. In some cases the sizes may be known accurately, for example the height of the Eiffel Tower or the height of Nelson's Column, but in other cases the sizes may only be known approximately, for example the heights of cars or adult humans. For these kinds of objects the height will be based on known or estimated average heights for the class of objects concerned. As an example, the order of magnitude heights of adult people and of trees are recorded as 1.683 metres and 10 metres respectively based on the average populations.

By recording order of magnitude heights of real world objects in the Place of Interest Ontology (see 7.4) which acts as the domain ontology for the system, we can infer more advanced spatial information for distance position for these objects based on the height information and the size of their bounding rectangles in the image.

6.2 The ‘Nearer Than’ and ‘Further Away Than’ Relationships

From simple optics it is known that the height of an object in an image depends mainly on its distance from the camera (along the optical axis) and the focal length of the lens. For an object of height *r* in the real world, if *f* is the focal length of the camera lens and *d* is the distance of the object from the camera, the height h in the image is given by:

$$h = rf / d \qquad (1)$$

This is illustrated in Figure 6-1. For arbitrary images we rarely know the focal length but it is clear from the relation above that the ratio of the heights of two objects in an image is the same as the ratio of their heights in the real world if they are the same distance from the camera.

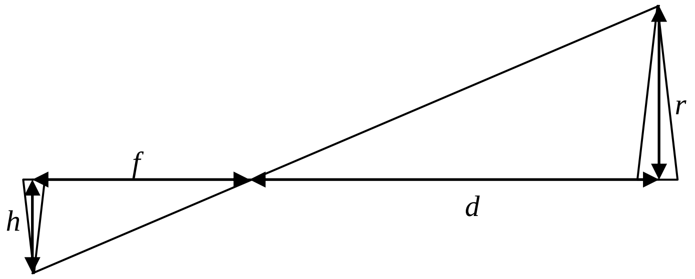


Figure 6-1 Calculation for distance from camera.

In the following sections four different cases are considered depending on the available knowledge about the heights of the objects concerned.

6.2.1 *Case 1: Two objects with accurately known heights*

Consider two objects, O_1 and O_2, which are the same distances from a camera and whose heights, r_1 and r_2 in the real world are known accurately. If their heights in the image are h_1 and h_2 respectively, then we know that

$$h_1 / h_2 = r_1 / r_2 \qquad (2)$$

We also know that if $h_1/ h_2 < r_1/r_2$ then O_1 is further away than O_2 and if $h_2/ h_1 < r_2/r_1$ then O_2 is further away than O_1.

These observations enable us to introduce a *nearer than* relation and a (reciprocal) *further away than* relation which can be extracted in cases where the heights of objects in the real world are known accurately, the objects have been suitably identified and their heights in the image extracted.

For practical purposes, in order to ensure that we only assert that an object is *further away than* or *nearer than* another to the camera, when there is a significant difference in the ratios, we introduce the following rules when object heights are known accurately:

1. If h_1/ h_2 < 0.9 r_1/r_2 then O_1 is further away than O_2
2. If h_2/ h_1 < 0.9 r_2/r_1 then O_2 is further away than O_1

6.2.2 Case 2: Two objects of Different Classes with Approximately Normally Distributed Heights

Unfortunately, many objects in the real world do not have heights which are known accurately although sets of similar objects may have heights which fall into a relatively narrow range. In such cases, knowledge of the order of magnitude heights of objects may enable us to establish whether one is nearer or further away from the camera than the other when their ratio of heights is sufficiently far from expected values. Cars for example vary in height to some extent but if one appears to be substantially larger than another, it may provide evidence that the much larger one is nearer to the camera.

In some cases we can do better than this. If we know the distribution of the ratio of heights, r_1/r_2, for objects from class 1 and class 2 then h_1/ h_2 should also belong to that distribution (from Equation 2 above) if O_1 and O_2 are at the same distance from the camera. If we can reject the hypothesis that $\mathrm{h}_1/ \mathrm{h}_2$ belongs to the distribution we can infer that they are not the same distance from the camera and one of the two objects in the image is nearer to the camera than the other.

In many cases the heights of objects of the same class may be approximately normally distributed; for example the heights of adult humans or the heights of mature oak trees. Unfortunately, the ratio (W = r_1/r_2) for such classes is usually not normally distributed and simple tests for membership of the distribution are not available. However, it was shown by Hayya et al., (1975), that, in certain circumstances, a transformation applied to the ratio, W, of two normally distributed variables, can be used to generate a variable, *z*, which is approximately normally distributed with zero mean and unit variance, N(0,1). This transformation, known as the Geary-Hinkley transformation, takes the form

$$z = (W\mu_2 - \mu_1)/(\sigma_1^2 - 2W\rho\,\sigma_1\sigma_2 + W^2\sigma_2^2)^{0.5}, \qquad (3)$$

where ρ is the correlation between the two variables and μ_1 and σ_1, μ_2 and σ_2 are the means and standard deviations of the variables in the numerator and denominator of W respectively.

In our case these are the means and standard deviations of the height distributions for object class 1 and object class 2 and typically their correlation, ρ, will be zero. Hayya et al., (1975) also show that the conditions under which the transformation holds are where the coefficient of variation (c.v.) for r_1 is greater than 0.005 and c.v. for r_2 is less than 0.39. If the Geary-Hinkley conditions are not met it may still be possible to extract the nearer than relation. If the value of r1 is very small, i.e. the c.v. is very small for object 1, regardless of the value of r2, we can make the approximation that the height of object 1 is known and so we can use Case 4. However, if the value of r1 is not less than 0.005 and r2 is greater than or equal to 0.39, both the coefficients are too large for any approximations to be made. In this case we can use the simulation approach mentioned in section 6.2.5

So if we know the mean and standard deviation of the real world height distributions of two objects under consideration in our image and assuming they meet Hayya's conditions and the heights are uncorrelated (ρ = 0 in (3) above), we can calculate z using

$$z = (W\mu_2 - \mu_1)/(\sigma_1^2 + W^2\sigma_2^2)^{0.5}. \qquad (4)$$

Then, if z is between -2 and +2 i.e. within two standard deviations of the mean, we cannot reject the hypothesis that the two objects in the image are the same distance from the camera. But if z is outside these limits we can reject the hypothesis at the

95% level and infer that one of the objects is nearer the camera than the other. Also, additionally

1. If h_1/ h_2 < μ_1/μ_2 then O_2 is nearer than O_1
2. If h_1/ h_2 > μ_1/μ_2 then O_1 is nearer than O_2

6.2.3 *Case 3: Two Objects of the Same Class with Approximately Normally Distributed Heights*

Now we consider the case where the two objects in the image are of the same class, and assume the heights of objects in that class are normally distributed and that the conditions for the Geary-Hinkley transformation are met. The case of two adult people is an example. The numerator and denominator of W are for the same distribution so $\mu_1 = \mu_2 = \mu$ and $\sigma_1 = \sigma_2 = \sigma$, and assuming the numerator and denominator are uncorrelated, equation 3 above for the transform reduces to:

$$z = \mu(W - 1)/\sigma(1 + W^2)^{0.5} \qquad (5)$$

and we can proceed as we did in case 2 but using equation 5 instead of equation 4, and where again,

1. If h_1/ h_2 < μ_1/μ_2 then O_2 is nearer than O_1
2. If h_1/ h_2> μ_1/μ_2 then O_1 is nearer than O_2

6.2.4 *Case 4: Two Objects, One from a Normally Distributed Class and the Other having Known Height*

In this case the ratio W will be normally distributed providing the object (O_1 say), which is normally distributed with mean μ_1 and standard deviation, σ_1, is placed in the numerator and the object (O_2) with known height, r_2, is placed in the denominator. W will then be from a normal distribution with mean, μ_W given by

$$\mu_W = \mu_1/ r_2 \qquad (6)$$

And standard deviation, σ_w given by

$$\sigma_W = \sigma_1/ r_2 \qquad (7)$$

If now we calculate W = $h_1/ h_2$from the image, then if W is not between $\mu_W - 2\sigma_W$ and $\mu_W + 2\sigma_W$, we can reject the hypothesis that they are at the same distance from the camera at the 95% level and,

1. If W < $\mu_W - 2\sigma_W$ then O_2 is nearer than O_1
2. If W > $\mu_W + 2\sigma_W$ then O_1 is nearer than O_2

6.2.5 Other Cases

In some situations for Case 2 above, the conditions of the Geary-Hinkley transform may not be met and in other situations the height distributions of the objects may not be normal. Providing the distributions of the object heights are known, it would be possible to use simulation to generate the distribution of the ratio of heights and identify the values of the 5% tail cut off values from the simulation. This would be much slower than the direct methods proposed above but would allow the possibility of extracting the "nearer than" relation in all situations where the distributions are known.

6.3 Extracting the Similar Distance from the Camera Relation

As mentioned in the introduction, in some circumstances, where appropriate, it may be possible to assert that two objects are at a similar distance from the camera. But this is not always the case. We begin by defining similar distance from the camera to mean that the fractional difference in their distances from the camera is less than or equal to 10%. Consider Case 1 above where the two objects have accurately known heights. In this case, from the derivation of the rules at the end of section 6.2.1, if neither the nearer than relation nor the further away than relation can be asserted, then the fractional difference in the objects' distances from the camera is less than or equal to 10% and we can assert that they are a similar distance from the camera.

For Case 2, which is the most general of the other 3 cases, both objects have heights for which only the mean and standard deviation of the class distributions are known. Fractional differences in observed height may be caused, not only by differences in distance from the camera but also by variations in the actual heights from their class means. If we are to be able to assert with some confidence that two objects are a similar distance from the camera, we need the coefficients of variation (σ / μ) of the actual heights of the object classes to be small so that any large differences between observed and mean height ratios are due to differences in distance from the camera.

Although this will be a weaker criterion than for Case 1, we arbitrarily require that the coefficient of variation for each of the two object classes should be less than 10% before we consider asserting the *similar distance from the camera* relation. If this requirement is satisfied and neither the *nearer than* relation nor the *further away than* relation *to the camera* can be asserted, then we assert that the two objects are *a similar distance from the camera.*

A rather more rigorous approach would be to say that, as we are concerned with fractional height differences we require that the standard deviation of the difference in heights should be less than say 10% of the mean height.

For two objects from the same class the standard deviation for height differences is $\sqrt{2\sigma}$ and the mean height is μ. For two objects from different classes the standard deviation is $\sqrt{\sigma_1^2 + \sigma_2^2}$ and the mean height is $½(\mu_1 + \mu_2)$. However, for our experimentation we adopt the slightly less rigorous criterion of the previous paragraph.

6.4 The ‘Near to’ Relationships

For any two objects in an image, their size in the image and the distance between them in the image can be used to extract a ‘Near to’ relation that relates to their positions in the image. However, two objects may be near to each other in the image but very far from each other in the real world. We will use the spatial terms *near in the image to* and *near in the real world to* in order to distinguish between these scenarios.

For the *near in the real world to* to be asserted, nearness in the image should be combined with the condition that the two objects are a similar distance from the camera. For the *near in the image to* relation we will use a rule based on Abella and Kender (1993). Using the same notation as that introduced in Chapter 5 section 5.3, the rule for *near in the image to* is as follows:

IF $|xc_2$ - $xc_1| \leq 1.5$ $(w_1 + w_2)$ AND $|yc_2 - yc_1| \leq 1.5$ $(h_1 + h_2))$
THEN Object_1 is NEAR IN THE IMAGE TO Object_2.

The spatial analysis system asserts *near in the real world to* relation if two objects are *near in the image to* each other and they are also at *a similar distance from the camera*. In all cases 1-4 of section 6.2, when a pair of objects is found to be a similar distance from the camera, then we could compute whether or not they are near in the real world. Therefore if the pair of objects is *near in the image* by using the previous algorithm, and they are at *a similar distance from the camera* as described in section 6.3, then we can predict that they are also *near in the real world*. Thus, the spatial relationships algorithm for the *near in the real world to* relations is as follows:

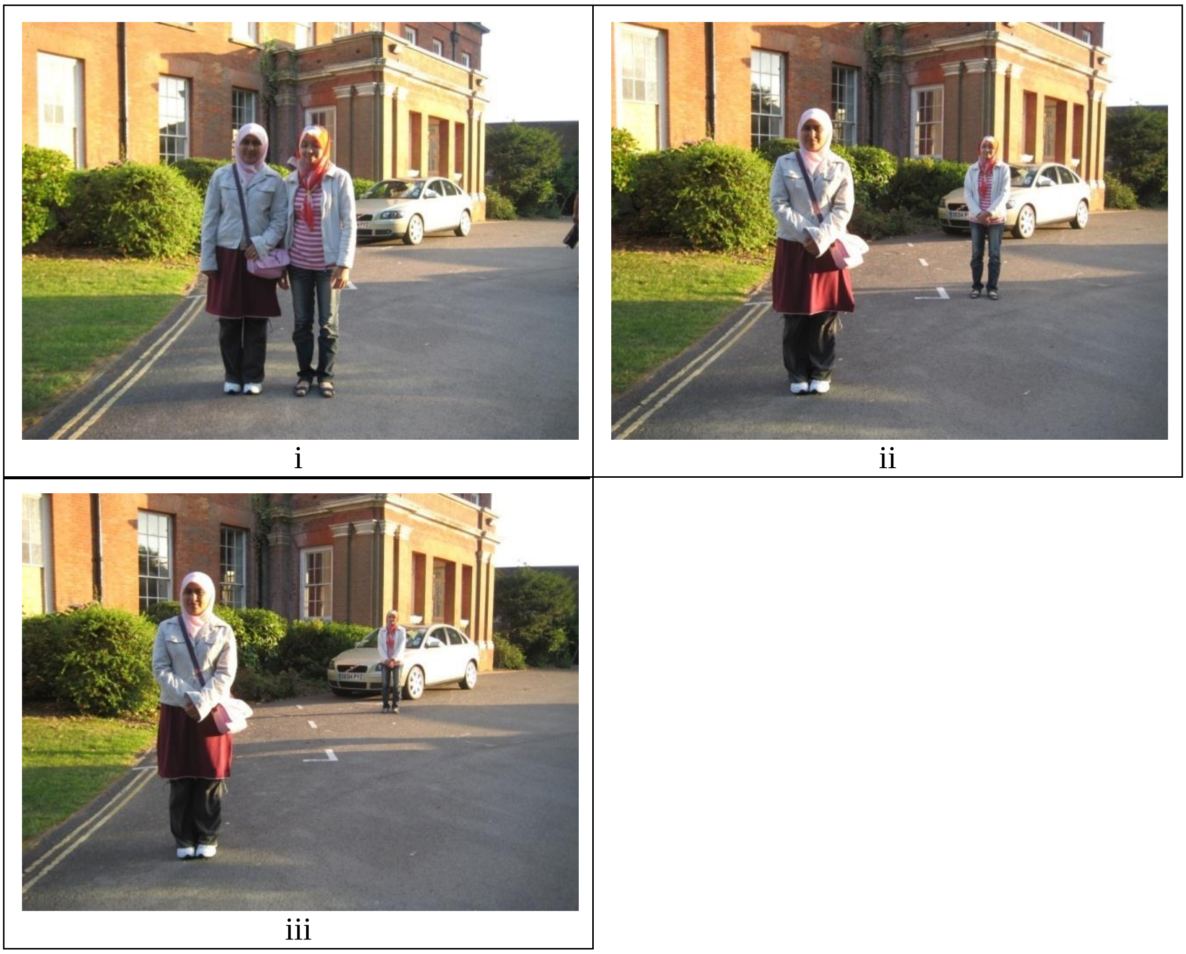
```
IF $\mathrm{Object}_1$ is NEAR IN THE IMAGE TO $\mathrm{Object}_2$ AND
IF $\mathrm{Object}_1$ and $\mathrm{Object}_2$ are a SIMILAR DISTANCE from the camera
THEN $\mathrm{Object}_1$ is NEAR IN THE REAL WORLD TO $\mathrm{Object}_2$.
```

6.5 Testing the Rules

Real-life case studies involving people with similar or different heights have been conducted to check the effectiveness of the rules introduced in this chapter. The aim of the first experiment is to test the use of the algorithms for the spatial terms: *nearer than* and *further away than*; and the spatial term: *a similar distance from the camera.* In the experiment three series of photos has been captured of two people at different positions relative to each other.

The first series involved two people with very similar heights, the second series is of two people with a small difference in height (0.04m) and the third series involves two people with a larger difference in height (0.25m). Each series of images was created with one person moving back in each image, further from the camera. The heights of both people as they appear in each of the images have been measured, together with their actual heights in the real world. The image series are described as follows.

1. Series 1: Figure 6-2-i shows a photo where two people with similar height are standing beside each other, whilst Figure 6-2-ii and iii show one person (person1) in the same position while the other person (person2) has moved further to the back.



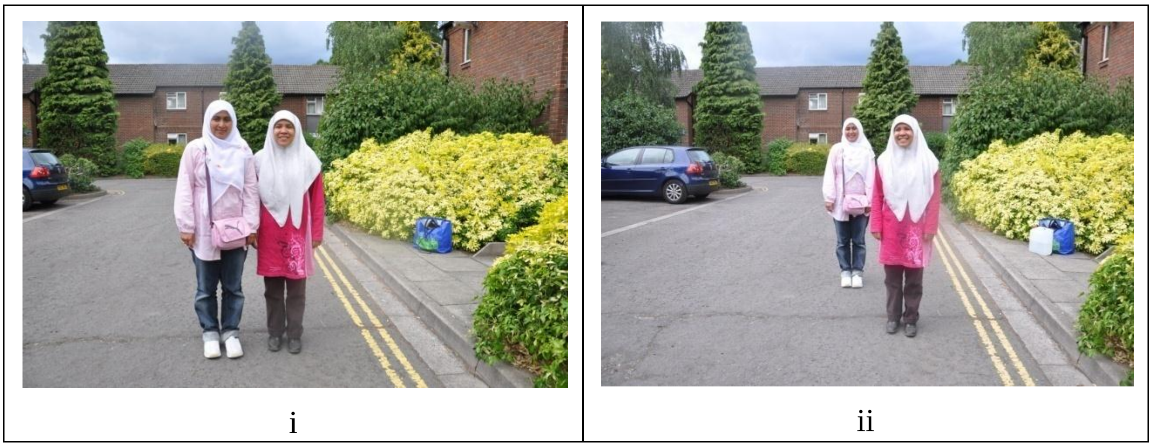
i

ii

iii

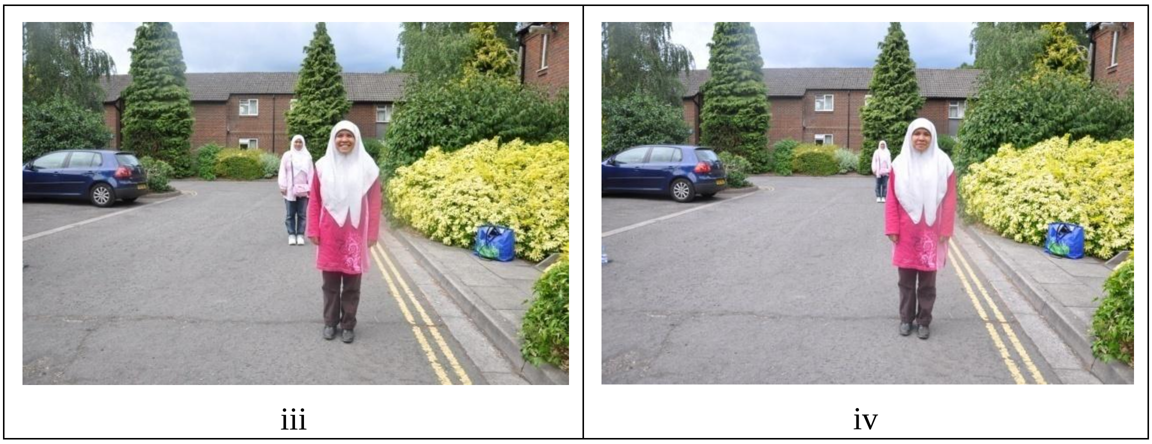
Figure 6-2 People with similar height (Person1 is on the left).

2. Series 2: Figure 6-3-i shows a photo where two people with slightly different heights are standing beside each other, whilst Figure 6-3-ii, iii and iv show one person (person1) static while the other person (person2) has moved back.



i

ii

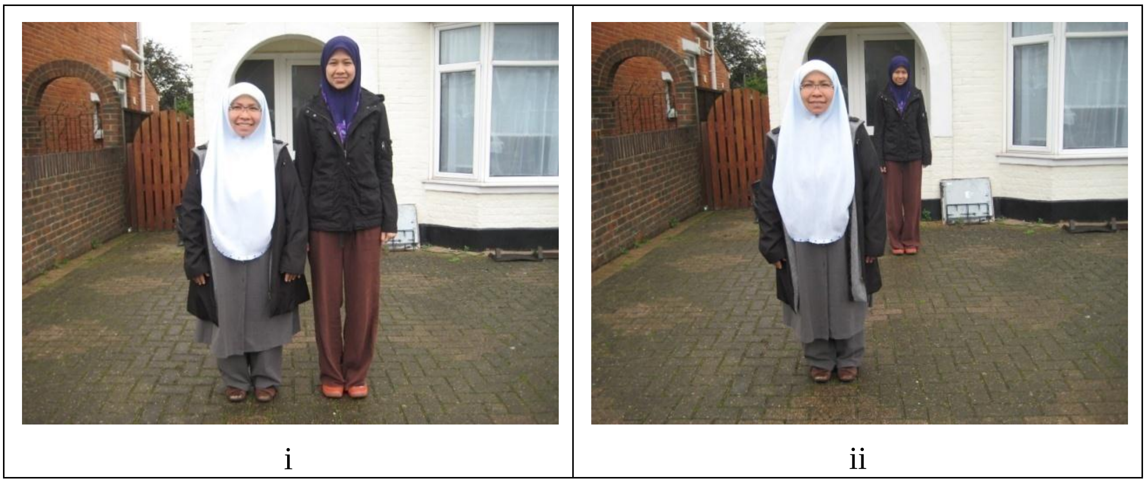


iii

iv

Figure 6-3 People with slightly different height (Person1 is on the right).

3. Series 3: Figure 6-4-i shows a photo where two people with significantly different heights are standing beside each other, whilst Figure 6-4-ii show one person (person1) in the same position stayed while another person (person2) has moved to the back.



i

ii

Figure 6-4 People with significantly different height (Person1 is on the left).

6.5.1 Results and Analysis

The height measurements for all series are shown in Table 6-1 using the same notation as used earlier in Chapter 5 (height for person$_i$ in the image is 2h$_i$ and height in the real world is M$_i$).

Table 6-1 Real-life Scenarios Experiment Results

Figure	$2h_1$	$2h_2$	M_1	M_2
Figure 6-2-i	14	14	154	154
Figure 6-2-ii	14.8	8.3	154	154
Figure 6-2-iii	14.8	5.4	154	154
Figure 6-3-i	14	13.3	154	150
Figure 6-3-ii	9.4	12.2	154	150
Figure 6-3-iii	6.1	12.2	154	150
Figure 6-3-iv	3.4	12.2	154	150
Figure 6-4-i	17	20	148	173
Figure 6-4-ii	17.2	10.6	148	173

As we know the actual heights of the people in the images we can first consider Case 1 described in subsection 6.2.1 for two objects of accurately known height. In this case if the ratio of heights in the image is not within 10% of the actual ratio of height of the objects, the system asserts that one is *nearer than* the other. If the ratio is within 10% of the actual height ratio the system asserts that they are *a similar distance from the camera* (as described in section 6.2 and section 6.3 for Case 1). The ratios of the image and of the actual heights for each of the images in the three series are shown in Table 6-2 together with the 10% limits.

Table 6-2 Analyses for Case 1

Figure	A $= h_1/h_2$	B $= M_1/M_2$	$0.9\ (M_1/M_2)$	C $=h_2/h_1$	D $= M_2/M_1$	$0.9\ (M_2/M_1)$
Figure 6-2-i	1	1	0.9	1	1	0.9
Figure 6-2-ii	1.78	1	0.9	0.5608	1	0.9
Figure 6-2-iii	2.74	1	0.9	0.3648	1	0.9
Figure 6-3-i	1.0526	1.0267	0.924	0.95	0.9740	0.8766
Figure 6-3-ii	0.7704	1.0267	0.924	1.2979	0.9740	0.8766
Figure 6-3-iii	0.5	1.0267	0.924	2	0.9740	0.8766
Figure 6-3-iv	0.2787	1.0267	0.924	3.5882	0.9740	0.8766
Figure 6-4-i	0.85	0.855	0.7695	1.1765	1.1689	1.052
Figure 6-4-ii	1.623	0.855	0.7695	0.6163	1.1689	1.052

Table 6-3 show more of the analysis and findings for Case 1. For each of the images it indicates whether the ratio of heights in the image is below the lower limit and also the resulting spatial statement asserted.

Table 6-3 Findings for Case 1

Figure	A<0.9B	C<0.9D	Spatial Statement
Figure 6-2-i	No	No	Person_1 and Person_2 a SIMILAR DISTANCE from the camera
Figure 6-2-ii	No	Yes	Person_2 is FURTHER AWAY THAN Person_1 Person_1 is NEARER THAN Person_2
Figure 6-2-iii	No	Yes	Person_1 is NEARER THAN Person_2
Figure 6-3-i	No	No	Person_1 and Person_2 a SIMILAR DISTANCE from the camera
Figure 6-3-ii	Yes	No	Person_1 is FURTHER AWAY THAN Person_2 Person_2 is NEARER THAN Person_1
Figure 6-3-iii	Yes	No	Person_1 is FURTHER AWAY THAN Person_2 Person_2 is NEARER THAN Person_1
Figure 6-3-iv	Yes	No	Person_1 is FURTHER AWAY THAN Person_2 Person_2 is NEARER THAN Person_1
Figure 6-4-i	No	No	Person_1 and Person_2 a SIMILAR DISTANCE from the camera
Figure 6-4-ii	No	Yes	Person_2 is FURTHER AWAY THAN Person_1 Person_1 is NEARER THAN Person_2

In the experiment so far, we have used our knowledge of the exact heights in the real world of the people involved. However, in general we may not know the exact heights of the objects in the image. If we do not use our knowledge of the true heights we can consider Case 3 for dealing with two objects of the same class with approximately normally distributed heights (see subsection 6.2.3).

In this case we use the mean and standard deviation of the height of the object class to calculate the Geary-Hinkley transformation, Z, of the height ratio from the image. Providing certain conditions are met, Z is normally distributed with zero mean and unit variance. The mean height for adult people is 1.68m and standard deviation is 0.11m (Statistics, 2008). For case 3 the condition for the transform to apply is that the coefficient of variation of the class height is between 0.005 and 0.39. In our case the coefficient of variation is 0.11/1.68=.065. The probabilistic threshold used in this experiment is 0.05 or 5%, where, if Z is outside the limit, we reject the hypothesis that the two people are a similar distance from the camera at the 95% level and infer that one of them is nearer to the camera than the other. The results and analyses for this case are shown in Table 6-4 and Table 6-5.

Table 6-4 Statistical Analyses Using Geary-Hinkley Transformation

Figure	**Z**	**h_1/h_2**	**μ_1/μ_2**
Figure 6-2-i	0	1	1
Figure 6-2-ii	-5.8609	1.78	1
Figure 6-2-iii	-9.1289	2.74	1
Figure 6-3-i	-0.5546	1.0526	1
Figure 6-3-ii	+2.7816	0.7704	1
Figure 6-3-iii	+6.8424	0.5	1
Figure 6-3-iv	+10.6309	0.2787	1
Figure 6-4-i	1.7487	0.85	1
Figure 6-4-ii	-4.9980	1.623	1

In Table 6-4 for each image in the three series we show the ratio of heights in the image, the ratio of class heights for the two objects (people) and the value of Z from the Geary-Hinkley transformation. In Table 6-5 for each image, the first column indicates whether Z is within the 95% limits. The next two columns indicate whether the ratio h_1/h_2 is greater than or less than μ_1/μ_2. The final column indicates the spatial statement asserted by the system. Note that for Case 3 *a similar distance* relation is only asserted if neither the *nearer than* nor *further away than* relation is asserted and the coefficient of variation (σ/μ) is less than 0.1. In our case the c. of v. is 0.065.

Table 6-5 Statistical Findings for Case 3

Figure	**Z<-2 or Z>+2**	**$h_1/h_2 > \mu_1/\mu_2$**	**$h_1/h_2 < \mu_1/\mu_2$**	**Spatial Statement**
Figure 6-2-i	No	No	No	Person_1 and Person_2a SIMILAR DISTANCE from the camera
Figure 6-2-ii	Yes	Yes	No	Person_1 is NEARER THAN Person_2
Figure 6-2-iii	Yes	Yes	No	Person_1 is NEARER THAN Person_2
Figure 6-3-i	No	Yes	No	Person_1 and Person_2a SIMILAR DISTANCE from the camera
Figure 6-3-ii	Yes	No	Yes	Person_2 is NEARER THAN is Person_1
Figure 6-3-iii	Yes	No	Yes	Person_2 is NEARER THAN is Person_1
Figure 6-3-iv	Yes	No	Yes	Person_2 is NEARER THAN is Person_1
Figure 6-4-i	No	No	Yes	Person_1 and Person_2 a SIMILAR DISTANCE from the camera
Figure 6-4-ii	Yes	Yes	No	Person_1 is NEARER THAN Person_2

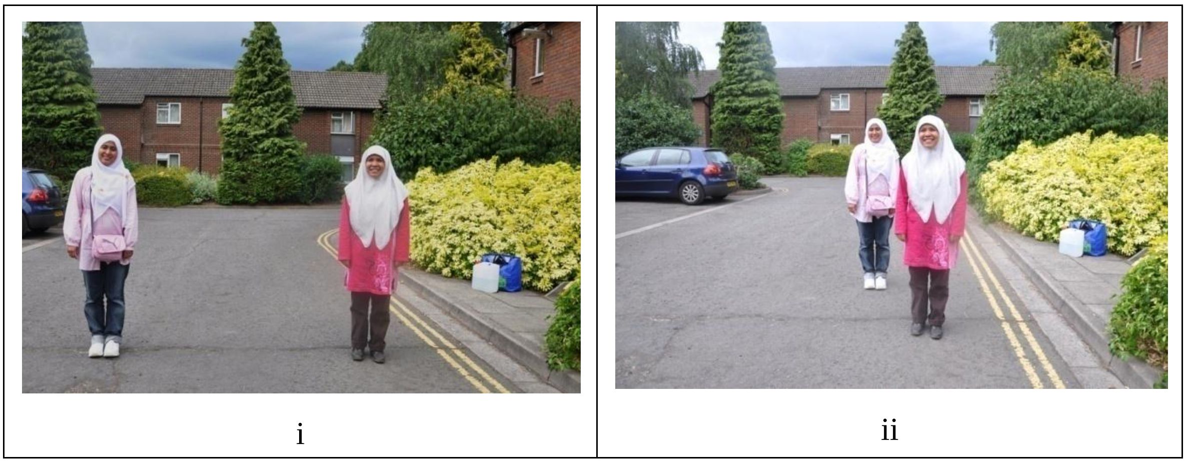
6.5.2 Discussion

By using the algorithms developed in section 6.2 for *nearer than* and *further away than* and in section 6.3 for *a similar distance from the camera*, the assertions given by the system have been calculated and shown for Case 1 (Table 6-3) when the actual heights of the people are known and Case 3 (Table 6-5) when only the mean and standard deviation of heights are known. By comparing the assertions in the tables with the images shown in the figures it can be seen that in these examples all assertions are correct.

6.5.3 Spatial Term for Near To

In section 6.4, the *near to* relation is developed and a distinction is made between *near in the image to* and *near in the real world to*. A small set of photos have been captured to study how the knowledge of height and width in the image and height information for the real world can be used to compute these spatial terms.

Figure 6-3-i has been used again in this experiment to show two persons standing side by side. They are close both in the image and the real world. Figure 6-5-i shows two people who are at the same distance from the camera but separated in the image and Figure 6-5-ii shows two people who are close in the image but at different distances from the camera.



i

ii

Figure 6-5 Nearness between two people.

The algorithms have been used to compute the nearness spatial relationships between the two people in the images. As for the analysis in the previous experiment, we consider both Case 1, where the real heights of the people are known, and case 3 where only the mean and standard deviation of the heights of adults are used. The resulting spatial statements are shown in the Table 6-6. The table shows both the statements which can be made about the images by observation and also the assertions made by applying the rules developed in section 6.4. In each case the assertions made by the system accord with observation of the images.

Table 6-6 Experiment Results and Analyses

Note: A - Person_1 is NEAR TO Person_2 in the image.
B - Person_1 is NEAR TO Person_2 in the real world.

Figure	**Observation**		**Spatial Result (Case 1)**		**Spatial Result (Case 3)**	
	A	B	A	B	A	B
Figure 6-3-i	Yes	Yes	Yes	Yes	Yes	Yes
Figure 6-5-i	-	-	-	-	-	-
Figure 6-5-ii	Yes	No	Yes	No	Yes	No

The findings in Table 6-6 show that, in these cases, the algorithms used have produced accurate spatial terms for distance position: near in the image to and near in the real world to for the given object of people in both Case 1 and Case 3.

6.6 A Further Example

The implementation for these advanced spatial relationships using order magnitude height information is presented and discussed further using a simple example. A segmented and labelled image with a subset of objects shown in Figure 6-6 has been taken from the LabelMe dataset (Russell et al., 2008). This image has been analysed and annotated with our advanced spatial relationships terms. The coordinates of the boundary pixels of the labelled objects named Eiffel_Tower, Person1, Person2, Trees1 and Trees2 have been extracted.

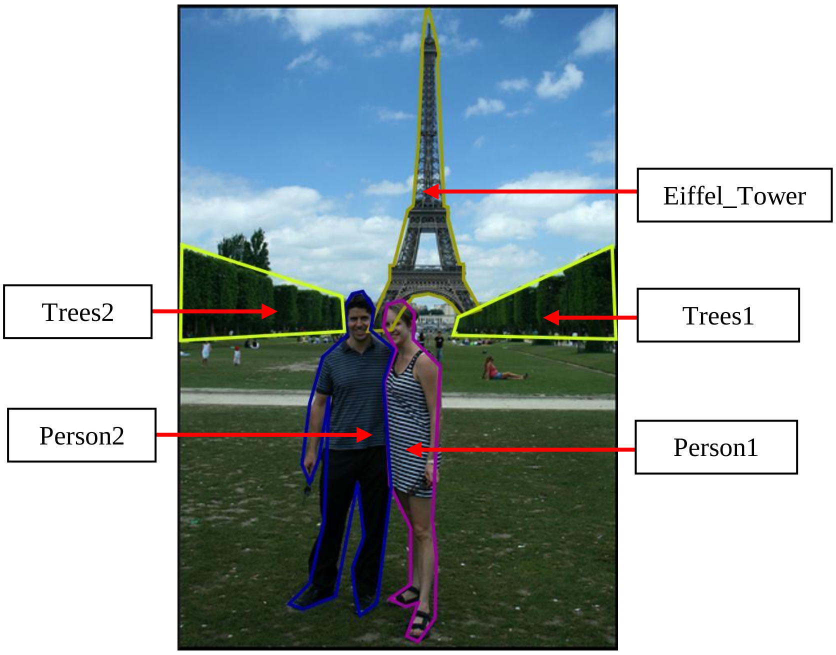


Figure 6-6 A sample of image for computation of distance relation.

The extraction rules for spatial distance relationships have been applied and the example of the output has been computed. A series of spatial statements generated from the computation of the advanced spatial relationships between those selected objects in the image is shown in Figure 6-7.

Output segment

```
SPATIAL ANNOTATION
IMAGE NAME:Eiffel_Tower_000000099
Regions Name: Eiffel_Tower, Person1, Person2, Trees1, Trees2,

MAGNITUDE OF HEIGHT
===================

DISTANCE POSITION:

 Eiffel_Tower is NEAR IN THE IMAGE TO Person1.
 Eiffel_Tower is NEAR IN THE IMAGE TO Person2.
 Eiffel_Tower is NEAR IN THE IMAGE TO Trees1.
 Person1 is NEAR IN THE IMAGE TO Person2.
 Person2 is NEAR IN THE IMAGE TO Trees2.

 Person1 is NEARER than Eiffel_Tower, and Eiffel_Tower is FURTHER AWAY
 than Person1.
 Person2 is NEARER than Eiffel_Tower, and Eiffel_Tower is FURTHER AWAY
 than Person2.
 Trees1 is NEARER than Eiffel_Tower, and Eiffel_Tower is FURTHER AWAY
 than Trees1.
 Trees2 is NEARER than Eiffel_Tower, and Eiffel_Tower is FURTHER AWAY
 than Trees2.

 Person1 and Person2 a SIMILAR DISTANCE from the camera.

 Person1 is NEAR IN THE REAL WORLD TO Person2.
```

```
Person1 is NEARER than Trees1, and Trees1 is FURTHER AWAY than
Person1.
Person1 is NEARER than Trees2, and Trees2 is FURTHER AWAY than
Person1.
Person2 is NEARER than Trees1, and Trees1 is FURTHER AWAY than
Person2.
Person2 is NEARER than Trees2, and Trees2 is FURTHER AWAY than
Person2.

Trees1 and Trees2 a SIMILAR DISTANCE from the camera.
```

Figure 6-7 The spatial statements output for distance position.

From the series of spatial statements it can be seen that 2 pairs of objects: Person1-Person2 and Trees1-Trees2 are both at a similar distances from camera. Also using the knowledge of their closeness in the image (*near in the image to*), the system can automatically determine whether the pairs are near in the real world or not. As a result, the output shows that Person1 is *near in the real world to* Person2.

This output shows that the automatic computation of the relative distance position and nearness terms for the objects related in the image is working as expected for the spatial term *nearer than*, *further away than*, *a similar distance*, *near in the image to* and *near in the real world to.*

The series of assertions about the advanced spatial relationships between the labelled objects in the image regarding their distance and closeness position has been produced. These outputs are generated in the form of an RDF file in the turtle format for subsequent use, together with the two ontologies (the Spatial Relationships Ontology and the Place of Interest Ontology) for image retrieval applications to be discussed later in Chapter 8.

6.7 Conclusion

The computation and implementation of enhanced approaches to the advanced spatial information relationships extraction has been presented. These advanced extractions used the Geary-Hinkley transformation for additional extraction rules and instance or height information in the associated domain ontology. The ontologies will be discussed further in the next chapter.

Thus, we have developed and implemented the rules to automate the distance spatial information extraction for objects in images by computing the Geary-Hinkley transformation of Z (Hayya et al., 1975) for relative order of magnitude of height information to infer 3-dimensional spatial annotations indicating distance and relative closeness of associated pairs of objects to the camera/viewer for *nearer than*, *further away than* and *similar distance from the camera* relations, and between each other for *near in the image* and *near in the real world* relations.

The development has been largely theoretical although we have implemented the extraction algorithms and provided some simple demonstrations. The approach would benefit from a more extensive evaluation using a wide range of objects for which height distributions are available. It is clear that when objects are a similar but not identical distance from the camera it will not be possible to identify which is nearest using this approach and other limitations will result from inaccuracies in available height information.

The development and implementation could be expanded by more rules to consider various additional cases which could be categorised according to the distribution of the objects involved in images and their assumed accuracies. Also if time were available it would be interesting to explore other cues to nearness to the camera such as evidence from occlusion.

In total, we can now extract 48 spatial terms, including 8 relative spatial position concepts (left, right, above, below and the composites concepts). The extractions also included 25 fine-grained absolute spatial positions based on 10 individual locations in the image and inferred 5 additional advanced spatial relationships i.e. the 3-dimensional annotations including *nearer than*, *further away than*, *a similar distance*, *near in the image to* and *near in the real world to* by using relative order of magnitude heights of objects which could be derived from the domain ontology: the Place of Interest ontology. All the extraction of spatial information annotations has been demonstrated.

In conclusion, we have proposed a new method and approach for capturing spatial information from images in order to enhance an image annotation system for more high level semantic search and retrieval.

Having completed the spatial annotation component, in the next chapter, the idea of using domain and spatial relationships ontologies to provide controlled vocabulary for the assertions during annotation and retrieval and to improve the search capabilities of the system is presented.

Chapter 7

The Spatial Relationships and Domain Ontologies

7.1 Introduction

An ontology defines a common vocabulary and provides mutual understanding for users, domain experts or software agents to annotate, communicate and share information within their domain or field. Representation of spatial relationship concepts in the form of ontology has been developed to be used for the Spatial Semantic Image System and queries for objects in images for making image retrieval become more relevant. This will contribute to a better approach to semantic knowledge extraction in images while enhancing the capability of the Semantic Web and at the same time narrowing the Semantic Gap in image retrieval.

The Spatial Relationships Ontology is developed based on Methontology development methodology by using Protégé OWL-DL to support maximum expressiveness while retaining computational completeness (Wikipedia, 2008b). The Spatial Relationships Ontology provides information about spatial relationships between objects or regions within an image. The ontology also acts as a database for a Spatial Semantic Image System to store vital information related to the objects, such as an object’s centroid (centre point) and coordinates of bounding box of the objects. The information is essential in computing and generating additional spatial relationships between the objects concerned within the images. The spatial semantic image annotation system is developed and used to annotate relations between objects and location of objects in the image that could be used for query and retrieval of relevant images. A sample of domain ontology: Place of Interest Ontology is also has been developed to represent a subset of objects in the images related and to demonstrate the functionalities of the Spatial Relationships Ontology. The ontology is going to be used in conjunction with the Spatial Relationships Ontology during the annotation and retrieval of the image.

7.2 Ontology Development Methodology

An ontology development methodology or ontological engineering provides essential steps or processes in ontology development. There is no single widely accepted methodology and each work group employs its own methodology. Many disciplines have worked towards formalising the process involved in building an ontology and develop standardized ontologies that domain experts can use to share and annotate information in their field (Noy and McGuinness, 2001). There are a number of ontology design methodologies and among them, is one called Methontology. The process of designing and developing the Spatial Relationships Ontology and Place of Interest Ontology are based on the Methontology methodology.

The Methontology which was considerably influenced by software engineering methodologies (Noy and McGuinness, 2001) and knowledge engineering methodologies (Fernández-López and Gómez-Pérez, 2002a). The Methontology has been chosen because the Spatial Relationships Ontology is intended to be a general framework in describing types of spatial relationships that are commonly used between objects in images and hence could facilitate any application which involved images about which spatial reasoning is required. Therefore, a methodology with an application independent approach is better suited for the development of this ontology since the Spatial Relationships Ontology is not targeted for a single application. Thus, the Methontology is the most appropriate approach for this case. Furthermore the Methontology is also a very mature methodology (Fernández-López and Gómez-Pérez, 2002a) since it has been used by different groups for the development of ontologies in diverse domains.

The Methontology development phase consists of specification (identifies the intended uses of the ontology); conceptualisation (consists of identifying concepts and building a conceptual model); formalisation and implementation (which transforms the conceptual model into a formal model and represent this in a formal ontology language); and maintenance (consists of updates and corrections to the ontology when necessary). The methodology also includes project management activities: planning, control, quality assurance; and support activities: knowledge acquisition, integration, evaluation, documentation and configuration management (Fernández-López et al., 1999). The development of the Spatial Relationships Ontology is discussed based on each activity in the development phase of the methodology as described in the follow section.

7.3 The Spatial Relationships Ontology

The Spatial Relationships Ontology is the core ontology for the Spatial Semantic Image System to provide knowledge and perform reasoning on the spatial relationships between objects in images.

7.3.1 Specification Phase

This is the preliminary stage in the development of an ontology, where the reasons for building the ontology and its intended uses are acknowledged. The aim of the ontology is to facilitate the query and retrieval of relevant images requested by the user with the help of spatial annotation. Thus the primary purpose of the Spatial Relationships Ontology is to answer the following competency questions:

- How to represent spatial relationships for an object in images?
- What types of spatial relationships are involved within the image?
- What kind of query could be submitted based on the Spatial Relationships Ontology and the Place of Interest Ontology?

The ontology is also intended to provide a general description framework, to facilitate any application or other ontology that needs to benefit the potential of using spatial relationships resources; such as in medical and traffic scenarios.

7.3.2 Conceptualisation Phase

In this stage, first we identified and defined the requirements for the proposed ontology (Noy and McGuinness, 2001). By referring to real queries submitted to picture librarians in a number of large national and international picture libraries (Enser et al., 2007) and the preliminary survey performed in Chapter 4, we identify significant spatial relationships concepts used in those queries and the survey, which we aim to cover in the spatial relationships ontology. Each object has an absolute position in the image and relative positions with respect to other objects in the image. The spatial relationships will refer to pairs of objects in the image and the context of the object within the image including absolute position and the relative distance from the objects to the viewer/camera.

Then we can structure the domain knowledge as a meaningful model either from scratch or reusing existing models (Corcho et al., 2007). The knowledge of the spatial relationships was structured from scratch and the composition of the relationships is visualized in the Figure 7-1.

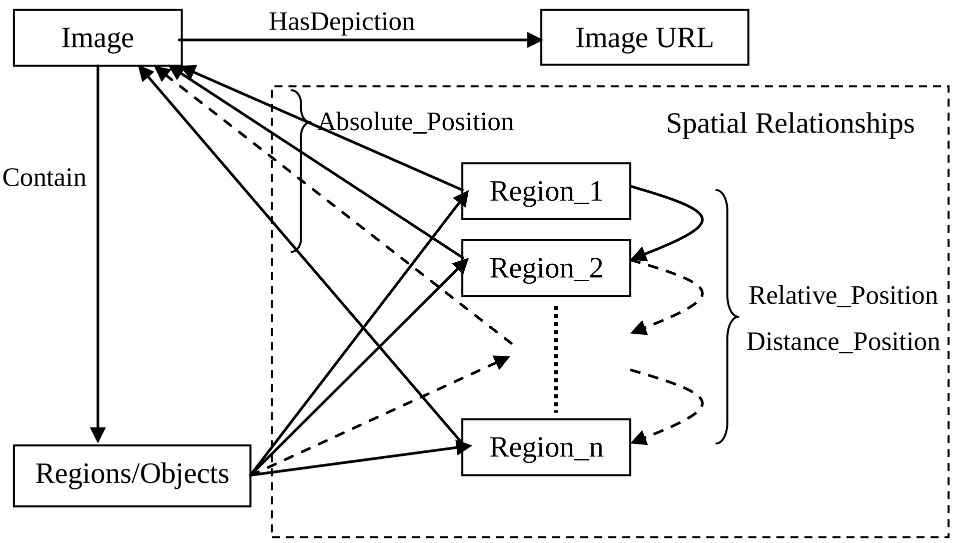


Figure 7-1 Visualisation of the relationships between objects and image.

All the spatial relationships are included in the ontology and at the same time the ontology could infer extra knowledge for the given objects according to the previous knowledge of the object. The taxonomy of the knowledge content for the spatial relationships considered in the Spatial Semantic Image System is shown in Figure 7-2. The knowledge of the spatial relationships are categorized into three main concepts which are *Relative_Position* for representing relative relationships between pairs of objects in an image, *Absolute_Position* for absolute position of the objects in an image and *Distance_Position* for representing a quantitative or relative distance between pairs of objects to the viewer (camera).

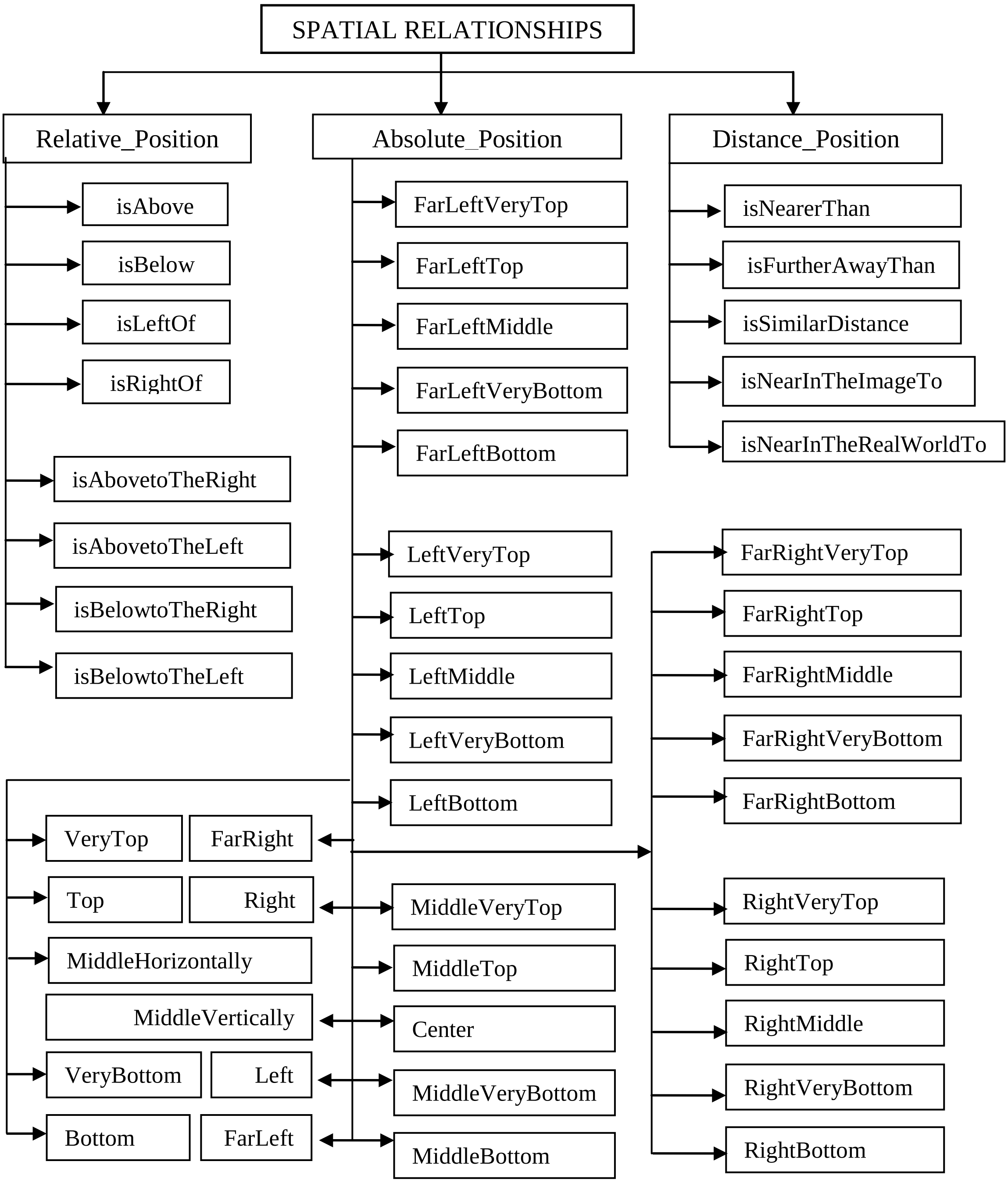


Figure 7-2 Spatial Relationships Hierarchy

These entire categories are represented as properties for class *Region*. Each category is described as follows:

1. Relative Position

Spatial terminologies that refer to relative position included in the ontology covered *isAbove*, *isBelow*, *isLeftOf* and *isRightOf* with composite relations including *isAbovetoTheRight*, *isAbovetoTheLeft*, *isBelowtoTheRight*, and *isBelowtoTheLeft*.

2. Absolute Position

   Spatial terminologies that refer to absolute position included in the ontology covered combinations of FarLeft, Left, Middle, FarRight and Right with VeryTop, Top, Middle, VeryBottom and Bottom. The Middle term in vertical and horizontal grid represent by Centre as shown in Figure 7-2. These concept consist of 35 absolute position terms.

3. Distance Position

   Spatial terminologies that refer to relative distance position included in the ontology covered *isNearerThan*, *isFurtherAwayThan*, *isSimilarDistance*, *isNearInTheImageTo* and *isNearInTheRealWorldTo*.

7.3.3 Formalisation and Implementation Phase

This formalisation stage involves transforming the conceptual model built in the previous stage and representing it as a formal-computable model, while the implementation stage builds computable models in an ontology language (Corcho et al., 2007). The Spatial Relationships ontology is developed by using the Protégé ontology editor. The ontology involved classes and properties of the related domain. A class defines a group of concepts that share some similar properties (Chebotko et al., 2009). Classes describe concepts in the domain and could consist of a superclass and subclasses. Subclasses under the same superclass are considered as siblings. A property states a relationship between concepts or from concepts to data values (Chebotko et al., 2009).

This section will describe the classes and properties involved in the development of the Spatial Relationships Ontology.

- The Spatial Relationships Classes

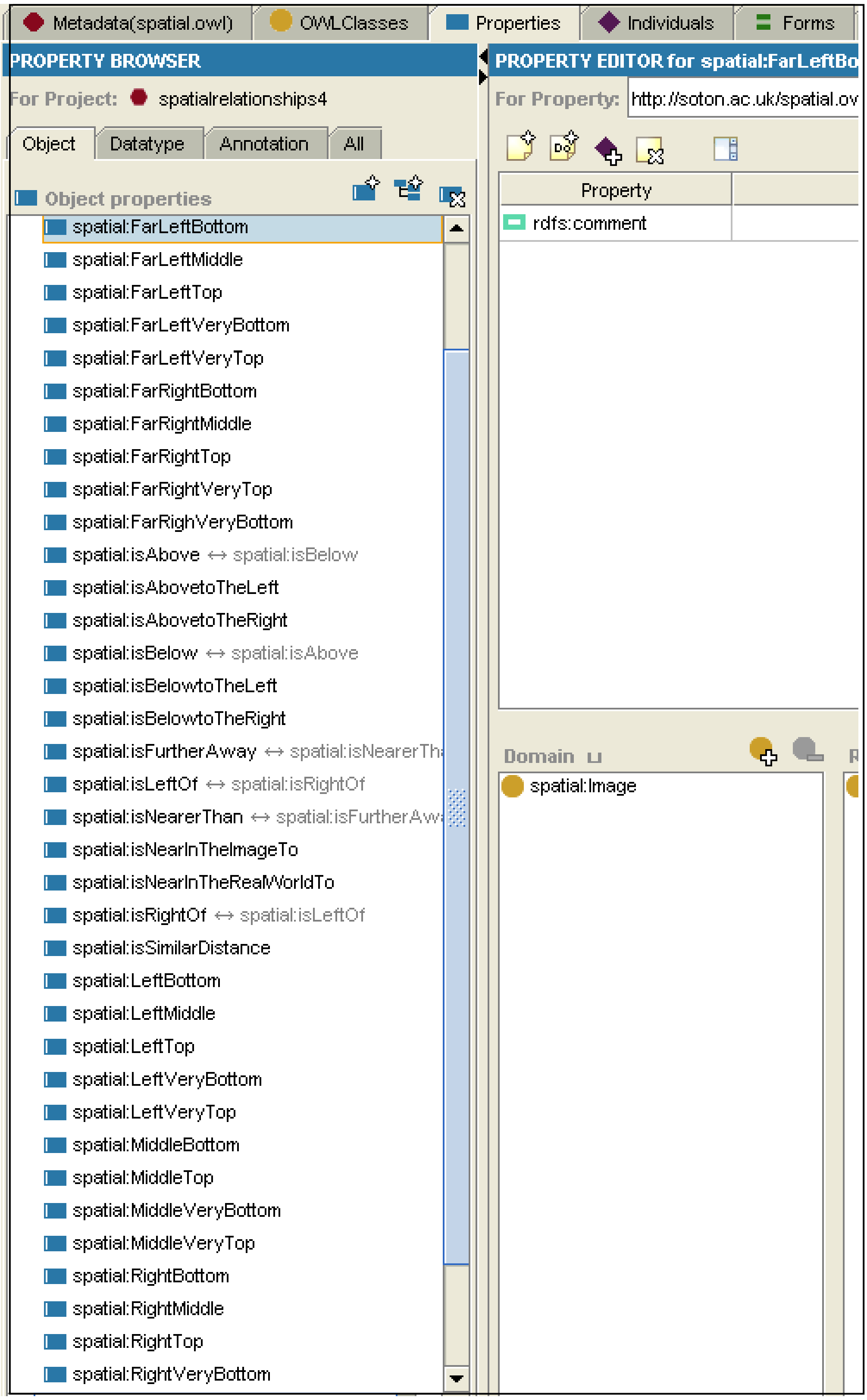
A class is sometimes called a concept and is the important content and the focus of most ontologies. Subclasses are used to represent concepts that are more specific than the superclass. The Spatial Relationships Ontology that has been developed consists of class *Image* and *Centroid*. Class *Image* also acts as superclass for subclass *Region*.

- The Spatial Relationships Properties

The spatial terminologies describing the relation between two or more objects in an image, or used to identify the position of objects in the image are considered as properties in the ontology. The spatial terminologies involved in the relative orientation, absolute position and distance relation are considered as the properties for class *Image* and class *Region*. All the properties for absolute position describe the spatial relationship between class *Region* and class *Image*, while others describe the spatial relationships within class *Region*. These properties (slots) are created and allocated under tab properties as shown in Figure 7-3. There are 48 relationships properties in this ontology. Figure 7-3 show the Protégé interface for the properties.

7.3.4 Maintenance

This stage involves activities of updating and correcting the ontology when necessary throughout the development and implementation of the Spatial Semantic Image System. The Methontology recommends a life cycle based on evolving prototypes by allowing for additions and modifications to the conceptual structure in each new version of the ontology (Fernández-López and Gómez-Pérez, 2002b). The Spatial Relationships Ontology was in fact developed in an iterative fashion, by iterating through conceptualisation and implementation stages several times before arriving at the final perceived ontology. Here the Spatial Relationships Ontology also could be used or reused by other ontologies or applications.



spatial:FarLeftBottom
spatial:FarLeftMiddle
spatial:FarLeftTop
spatial:FarLeftVeryBottom
spatial:FarLeftVeryTop
spatial:FarRightBottom
spatial:FarRightMiddle
spatial:FarRightTop
spatial:FarRightVeryTop
spatial:FarRighVeryBottom
spatial:isAbove ↔ spatial:isBelow
spatial:isAbovetoTheLeft
spatial:isAbovetoTheRight
spatial:isBelow ↔ spatial:isAbove
spatial:isBelowtoTheLeft
spatial:isBelowtoTheRight
spatial:isFurtherAway ↔ spatial:isNearerTh
spatial:isLeftOf ↔ spatial:isRightOf
spatial:isNearerThan ↔ spatial:isFurtherAw
spatial:isNearInTheImageTo
spatial:isNearInTheRealWorldTo
spatial:isRightOf ↔ spatial:isLeftOf
spatial:isSimilarDistance
spatial:LeftBottom
spatial:LeftMiddle
spatial:LeftTop
spatial:LeftVeryBottom
spatial:LeftVeryTop
spatial:MiddleBottom
spatial:MiddleTop
spatial:MiddleVeryBottom
spatial:MiddleVeryTop
spatial:RightBottom
spatial:RightMiddle
spatial:RightTop
spatial:RightVeryBottom

Figure 7-3 Spatial Relationships Properties

7.4 The Development of the Domain Ontology

A domain ontology can be reused to build others of more granularity and coverage, or can be merged with others to create new ones (Corcho et al., 2007). We have adopted DBPedia Ontology and extracted several objects to develop a simple domain ontology to be used with the Spatial Relationships Ontology in order to demonstrate the functionality of the Spatial Relationships Ontology in the Spatial Semantic Image System. The domain ontology that has been developed is a Place of Interest Ontology. The ontology was developed by using Protégé and involved a small number of classes and properties according to the objects contained and annotated within the collection of images that have been selected in the system. The taxonomy for the Place of Interest Ontology is show in Figure 7-4.

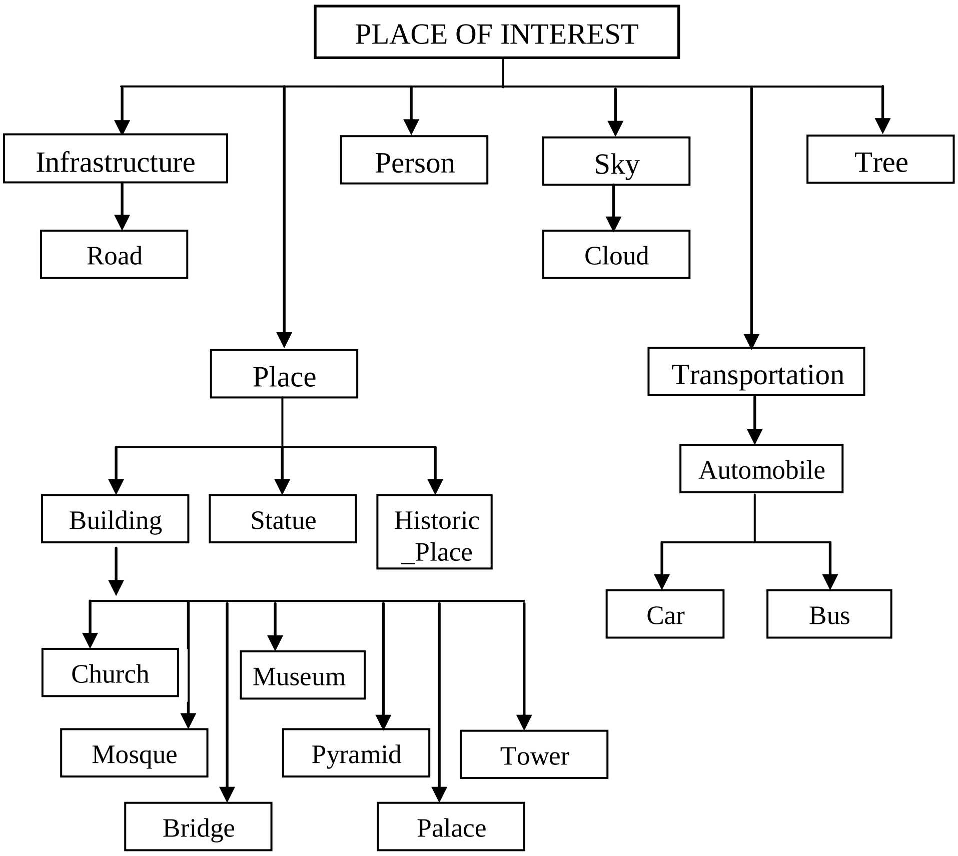


Figure 7-4 Taxonomy of the Place of Interest Ontology

The same development phases have been applied in developing the Place of Interest Ontology. The classes involved in the development of the Place of Interest Ontology are shown in the Protégé interface in Figure 7-5. All class: *Infrastructure, Person,* *Place*, *Sky*, *Transportation* and *Tree* have been described by slot *hasDepiction*: each class has depiction of image URI that contained the related instances in the class. The class *Person*, *Place*, *Transportation* and *Tree* have been described by their properties, including:

- *hasMean*: This refers to the mean height for the class. We have the mean for person, transportation and tree inserted to be used to infer the advanced spatial relationships described in Chapter 6. For a class with exact height, the exact height is inserted here, for example the height of Eiffel Tower.
- *hasStandardDeviation*: A person, a place, a transportation and a tree has standard deviation (if known) that could be used to help to infer the advanced spatial relationships. We have the standard deviation for person, transportation and tree inserted. For a class with exact height, the standard deviation is 0.

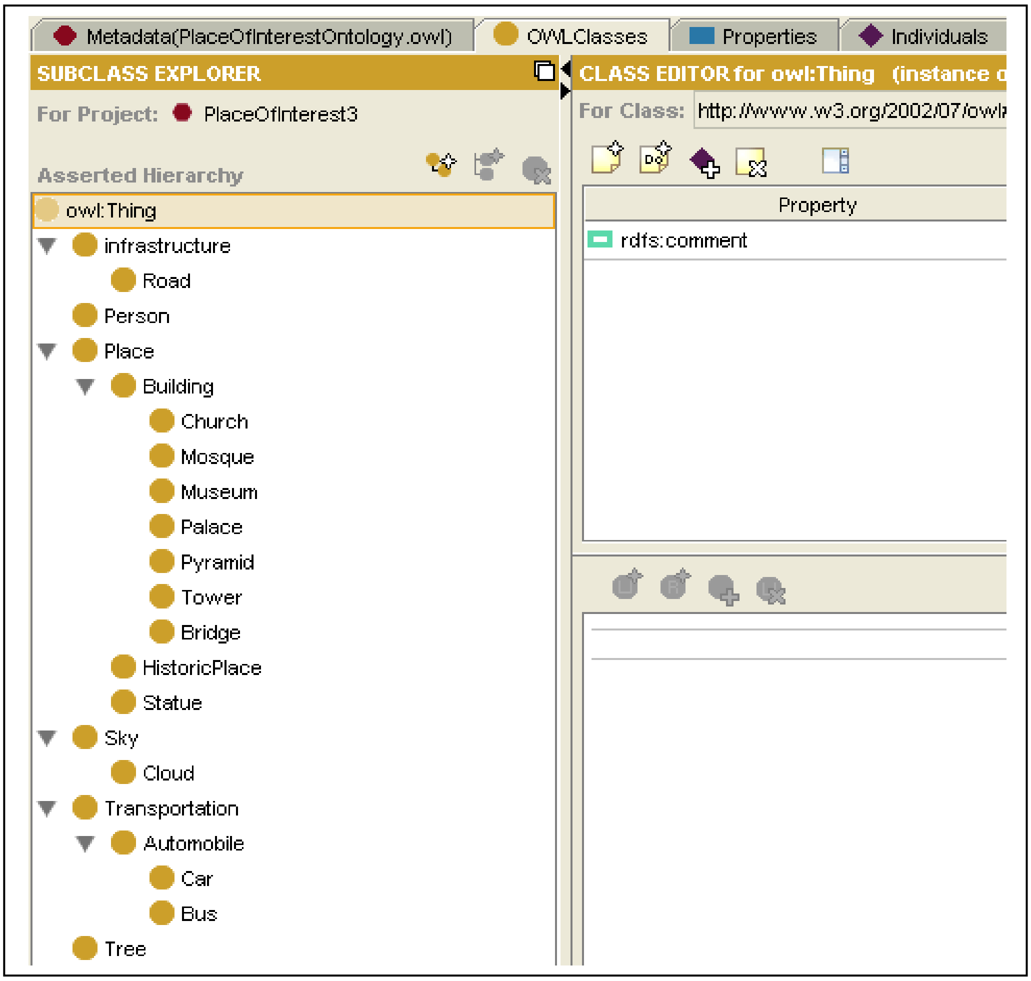


Figure 7-5 The Place of Interest Classes

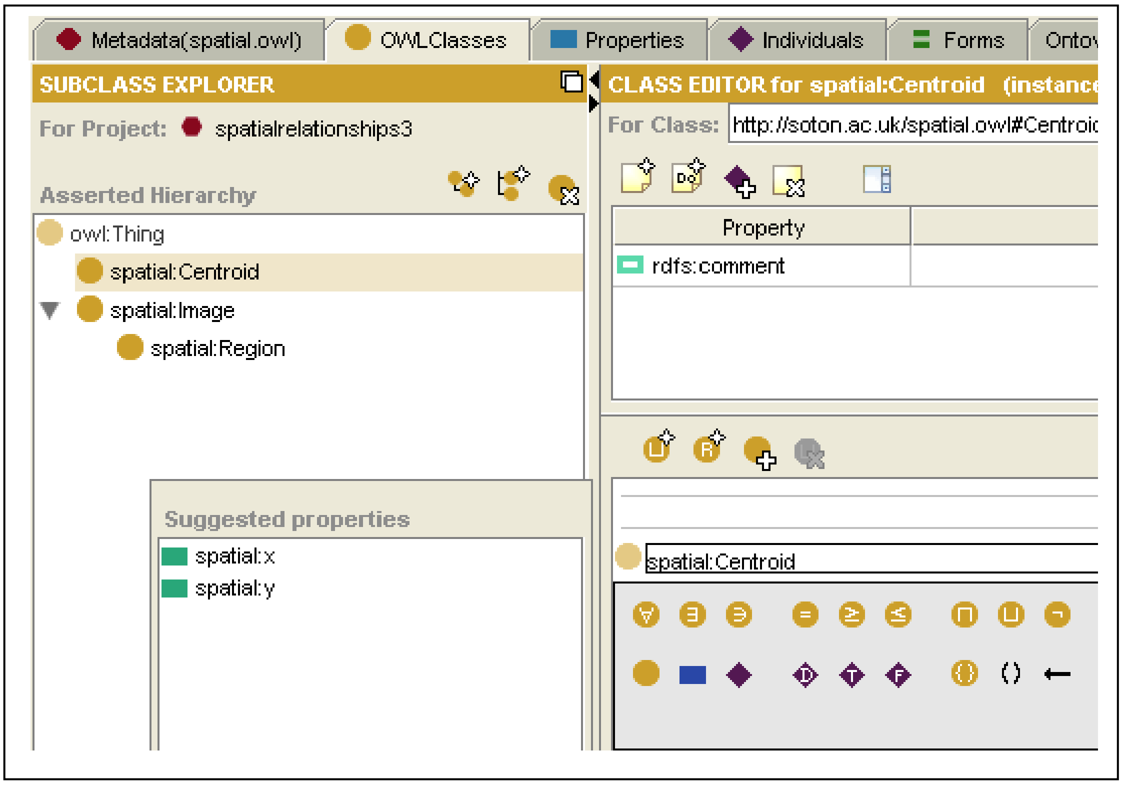
7.5 Extension to the Spatial Relationships Ontology

As mentioned before, the spatial relationships ontology that has been developed to cover data from the Spatial Semantic Image System to include a centroid coordinates and bounding box coordinates for each object and is represented as class *Centroid* with slots *x* and *y*.

The bounding box for each class Region has instances with the following slots:

- *hasXmax* and *hasXmin*: containing the coordinate x-maximum and x-minimum
- *hasYmax* and *hasYmin*: containing the coordinate y-maximum and y-minimum

All these are essential to compute the width and height of the object. As an example, Figure 7-6 shows the properties of a class *Centroid* where coordinates x and y for the associated centroid could be inserted.


For Project: spatialrelationships3

For Class: http://soton.ac.uk/spatial.owl#Centroic

Figure 7-6 Data for class *Centroid*.

By doing this, the ontology also acts as a centre of knowledge for the whole system where significant data in the Spatial Semantic Image System are stored and gathered in the same platform.

7.6 Conclusion

We have developed a relatively simple Spatial Relationships Ontology mainly as a proof of concept by including the spatial relationships introduced in earlier chapters extended with the spatial semantic image annotation system database, and Place of Interest Ontology as the domain ontology. The ontology could also be used as an application or intermediate for other domain ontologies requiring spatial relationships within the domain. However, there is much room here for improvement in the ontology which could be enhanced and expanded further with more spatial terminology to meet the requirements of future systems. The structure of these ontologies is as simple as possible but sufficiently complex to meet our particular needs as a proof of concept. Ways of extending and improving the ontology are discussed in the future work section of Chapter 9.

Having completed the Ontology Component, both ontologies will be integrated with the spatial analysis system in the Spatial Annotation Component and also the Retrieval Component. An evaluation of the integrated system using precision and recall will be carried out using the retrieval system and is presented in the next chapter to demonstrate how well the automated extraction of spatial relationships has been achieved from the integration of the annotation and ontological components.

Chapter 8

Integration and Evaluation of the Spatial Semantic Image System

8.1 Introduction

In this chapter, we first describe the integration of the spatial analysis software with the ontologies and the addition of a retrieval system to create the complete Spatial Semantic Image System (SpaSIS) in section 8.2. We then evaluate the overall performance of the system in two distinct experiments. Each experiment includes the methodology use, the results obtained and the analysis that have been performed, with details discussions and findings.

In the first experiment (section 8.3), we gathered the spatial assertions made by the system for three specific images, each containing two labelled objects. A small user study was then conducted to find how people perceived the same images by choosing from the complete list of possible statements the SpaSIS could have asserted. It was then possible to compare how similar the assertions made by the system are to the assertions selected by human viewers.

In the second experiment (section 8.4), a test set of images containing the Eiffel Tower and other objects is used to measure and evaluate the retrieval performance of the system. Precision and recall metrics together with their average F-score are used to show the retrieval performance. The results are compared with the retrieval performance for labelled images without the spatial assertions from the SpaSIS.

8.2 Integration to Create the Spatial Semantic Image System

The complete flow of the system, including the processes and tasks in the Spatial Semantic Image System, is shown in Figure 8-1. The processes are presented in two levels:

- The spatial annotation level
- The retrieval level

These two levels are to emphasize the main functions of the system and reflect the main contributions in this research.

8.2.1 The Spatial Annotation Level

As illustrated in Figure 8-1, from the LabelMe annotation tool, we have obtained inputs to our system in the form of objects labelled with the x and y coordinates of their bounding boxes. These inputs are processed by the spatial relationship algorithms in the spatial analysis system to generate spatial information in the form of assertions about the absolute and relative positions of the objects in the image. The assertions are stored in the form of a triplestore (RDF file) as a representation of the system's knowledge-base. The RDF files are also customized to the two ontologies that have been developed, the Spatial Relationships Ontology and the Place of Interest Ontology. These ontologies contain spatial relationships information and other essential knowledge about objects related in images such as the order of magnitude of the object height.

8.2.2 The Retrieval Level

The retrieval level enables users to submit queries for images in the collection. The queries may include the specification of spatial positions of objects and spatial relationships between objects. During queries over a collection of images or datasets, the content of the knowledge-base is searched and the required images are retrieved. The retrieval system uses SQL queries developed in JAVA.

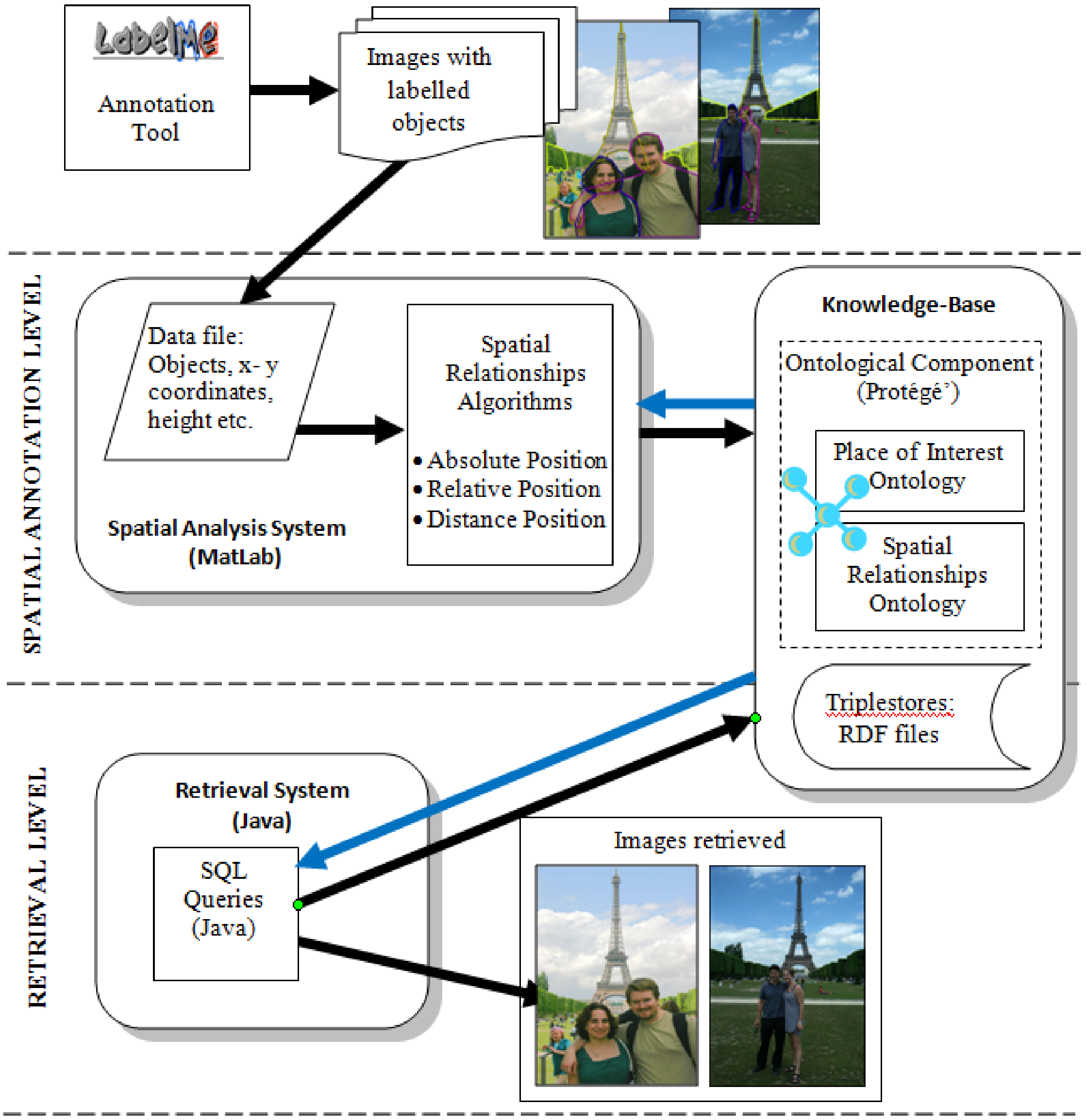

LabelMe
Annotation Tool
Images with labelled objects
SPATIAL ANNOTATION LEVEL
Data file: Objects, x- y coordinates, height etc.
Spatial Relationships Algorithms
• Absolute Position
• Relative Position
• Distance Position
Spatial Analysis System (MatLab)
Knowledge-Base
Ontological Component (Protégé)
Place of Interest Ontology
Spatial Relationships Ontology
Triplestores: RDF files
RETRIEVAL LEVEL
Retrieval System (Java)
SQL Queries (Java)
Images retrieved

Figure 8-1 Representation of the Spatial Semantics Image System

8.3 Evaluating the Spatial Assertions

In this experiment the aim was to compare the spatial assertions made by the system, for a small number of specific images, with the assertions which human subjects thought appropriate for the same images.

8.3.1 Methodology

Three images with different levels of complexity were selected from the LabelMe image collection to be used in this analysis. In each image, two main objects were labelled for which spatial relationships will be considered. Figure 8-2 shows the three images with the labelled objects intended for the analysis. The spatial assertions generated by the Spatial Semantic Image System for these three images were collected.

In a small user study each of the three images were shown to users together with a list of the 92 possible assertions (spatial statements) that the system could in principle generate about images containing two labelled objects. These three image represent different kind of spatial complexity that could trigger and point out if the system is not performs as expected. Users were asked to identify (tick) all the spatial statements which they considered to be correct for each image. A total number of 22 respondents submitted their responses. The assertions selected by the users were compared with the corresponding spatial statements generated by the SpaSIS system. The survey form for Image 1 is shown in Appendix A and the users' responses and results are shown in Appendix B.

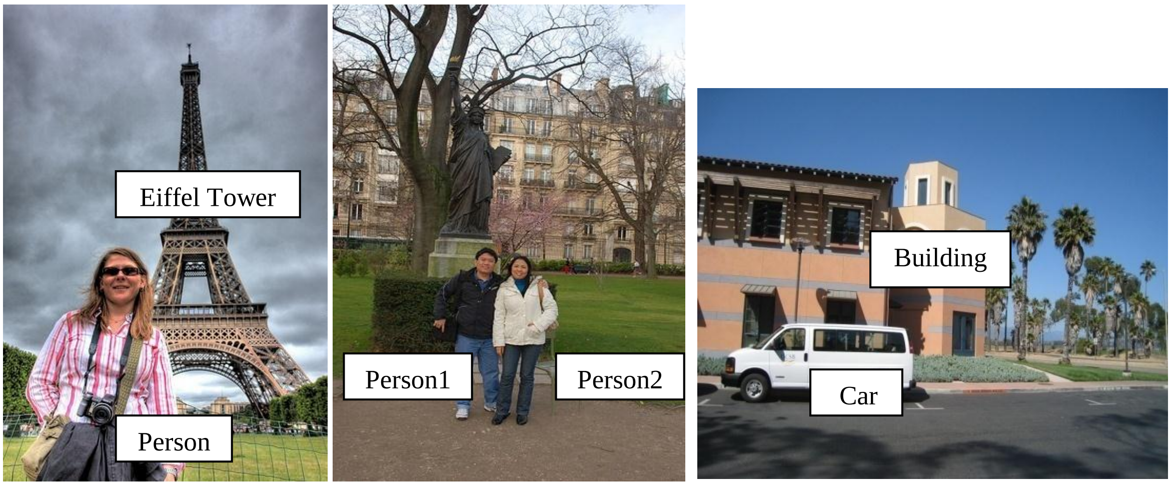


Image 1 Image 2 Image 3

Figure 8-2 Images used in the user evaluation.

8.3.2 Experimental Results

Table 8-1, Table 8-2 and Table 8-3 list the spatial statements for each of the three images which were either supported by the users, generated by the SpaSIS or both. The spatial statements that are neither selected by any users nor generated by the system are dropped from the lists in the tables. The column **Sys** contains a 1 if the spatial statement was generated by the system for that image, while column **Res** indicates the total of respondents who selected the associated spatial statement. The column **%** shows the percentage of people who selected the related spatial statement.

The table also highlights some cells with different colours to show a similarity or difference between the survey's result and the system's result. Each colour indicates the following:

1. The green colour 🟩 shows spatial statements which a significant number of survey respondents (more than 40%) selected and which was also generated by the system automatically.
2. The yellow colour 🟨 highlights a high percentage of contrast between user responses and the system i.e. where more than 40% of respondents selected the spatial statement but it was not generated by the system.
3. The blue colour 🟦 shows where not many (less than 40%) of the respondents selected the spatial statement but it has been generated by the system.

The green suggests the automatic system is working well in this case. The yellow suggests it is missing spatial relationships and the blue suggests assertions which are not selected by many users and are thus possibly invalid.

Table 8-1 Results and Analysis for Image 1

	Spatial Statements	Sys	Res	%
1	Person is left of Eiffel tower.	1	18	81.82
2	Eiffel tower is right of person.	1	18	81.82
3	Person is right of Eiffel tower.		2	9.09
4	Eiffel tower is left of person.		3	13.64
5	Person is above Eiffel tower.		1	4.55
7	Person is below Eiffel tower.	1	15	68.18
8	Eiffel tower is above person.	1	12	54.55
9	Person is below and to the right of Eiffel tower.		3	13.64
10	Person is below and to the left of Eiffel tower.		14	63.64
15	Eiffel tower is above and to the right of person.		13	59.09
16	Eiffel tower is above and to the left of person.		2	9.09
17	Person is on the far left side of the image.		5	22.73
21	Person is on the far left side and at the bottom of the image.		7	31.82
22	Person is on the far left side and at the very bottom of the image.		2	9.09
23	Person is on the left side of the image.	1	17	77.27
25	Person is on the left side and at the top of the image.		1	4.55
26	Person is on the left side and in the middle of the image.		1	4.55
27	Person is on the left side and at the bottom of the image.	1	16	72.73
28	Person is on the left side and at the very bottom of the image.		7	31.82
33	Person in the middle and at the bottom of the image.		3	13.64
48	Person is at the top of the image.		1	4.55
49	Person is in the middle of the image.		1	4.55
50	Person is at the bottom of the image.	1	15	68.18
51	Person is at the very bottom of the image.		5	22.73
64	Eiffel tower is in the middle of the image.		13	59.09
66	Eiffel tower is in the middle and at the top of the image.		2	9.09
67	Eiffel tower is in the centre of the image.		6	27.27
68	Eiffel tower in the middle and at the bottom of the image.		2	9.09
70	Eiffel tower is on the right side of the image.	1	13	59.09
71	Eiffel tower is on the right side and at the very top of the image.		1	4.55
72	Eiffel tower is on the right side and at the top of the image.		6	27.27
73	Eiffel tower is on the right side and in the middle of the image.	1	13	59.09
74	Eiffel tower is on the right side and at the bottom of the image.		2	9.09
76	Eiffel tower is on the far right side of the image.		2	9.09
79	Eiffel tower is on the far right side and in the middle of the image.		2	9.09
82	Eiffel tower is at the very top of the image.		1	4.55
83	Eiffel tower is at the top of the image.		5	22.73
84	Eiffel tower is in the middle of the image.	1	12	54.55
85	Eiffel tower is at the bottom of the image.		1	4.55
87	Person is near to Eiffel tower in the real world.		5	22.73
88	Eiffel tower is near to Person in the real world.		3	13.64
89	Person is nearer than Eiffel tower.	1	16	72.73
90	Eiffel tower is further away than person.	1	16	72.73
91	Eiffel tower is nearer than person.		2	9.09
92	Person is further away than Eiffel tower.		2	9.09

Table 8-2 Results and Analysis for Image 2

Spatial Statements		Sys	Res	%
1	Person1 is left of Person2.	1	19	86.36
2	Person2 is right of Person1.	1	20	90.91
3	Person1 is right of Person2.		3	13.64
4	Person2 is left of Person1.		2	9.09
5	Person1 is above Person2.	1	1	4.55
6	Person2 is below Person1.	1	1	4.55
12	Person1 is above and to the left of Person2.		1	4.55
13	Person2 is below and to the right of Person1.		1	4.55
20	Person1 is on the far left side and in the middle of the image.		2	9.09
21	Person1 is on the far left side and at the bottom of the image.		2	9.09
23	Person1 is on the left side of the image.		19	86.36
25	Person1 is on the left side and at the top of the image.		1	4.55
26	Person1 is on the left side and in the middle of the image.		11	50.00
27	Person1 is on the left side and at the bottom of the image.	1	11	50.00
28	Person1 is on the left side and at the very bottom of the image.		2	9.09
29	Person1 is in the middle of the image.	1	13	59.09
31	Person1 is in the middle and at the top of the image.		1	4.55
32	Person1 is in the centre of the image.		1	4.55
33	Person1 is in the middle and at the bottom of the image.	1	8	36.36
48	Person1 is at the top of the image.		1	4.55
49	Person1 is in the middle of the image.		10	45.45
50	Person1 is at the bottom of the image.	1	10	45.45
51	Person1 is at the very bottom of the image.		1	4.55
58	Person2 is on the left side of the image.		1	4.55
62	Person2 is on the left side and at the bottom of the image.		1	4.55
64	Person2 is in the middle of the image.	1	12	54.55
66	Person2 is in the middle and at the top of the image.		1	4.55
67	Person2 is in the centre of the image.		4	18.18
68	Person2 in the middle and at the bottom of the image.	1	9	40.91
70	Person2 is on the right side of the image.		16	72.73
72	Person2 is on the right side and at the top of the image.		1	4.55
73	Person2 is on the right side and in the middle of the image.		11	50.00
74	Person2 is on the right side and at the bottom of the image.	1	10	45.45
75	Person2 is on the right side and at the very bottom of the image.		1	4.55
79	Person2 is on the far right side and in the middle of the image.		2	9.09
80	Person2 is on the far right side and at the bottom of the image.		1	4.55
83	Person2 is at the top of the image.		1	4.55
84	Person2 is in the middle of the image.		10	45.45
85	Person2 is at the bottom of the image.	1	9	40.91
86	Person2 is at the very bottom of the image.		2	9.09
87	Person1 is near to Person2 in the real world.	1	21	95.45
88	Person2 is near to Person1 in the real world.	1	20	90.91

Spatial Statements		Sys	Res	%
91	Person2 is nearer than Person1.		6	27.27
92	Person1 is further away than Person2.		6	27.27

Table 8-3 Results and Analysis for Image 3

Spatial Statements		Sys	Res	%
2	Car is right of Building.		1	4.55
3	Building is right of Car.	1	6	27.27
4	Car is left of Building.	1	6	27.27
5	Building is above Car.	1	8	36.36
6	Car is below Building.	1	8	36.36
11	Building is above and to the right of Car.		6	27.27
14	Car is below and to the left of Building.		5	22.73
17	Building is on the far left side of the image.		8	36.36
18	Building is on the far left side and at the very top of the image.		1	4.55
19	Building is on the far left side and at the top of the image.		3	13.64
20	Building is on the far left side and in the middle of the image.		8	36.36
23	Building is on the left side of the image.		13	59.09
25	Building is on the left side and at the top of the image.		3	13.64
26	Building is on the left side and in the middle of the image.		12	54.55
27	Building is on the left side and at the bottom of the image.		1	4.55
29	Building is in the middle of the image.	1	12	54.55
31	Building is in the middle and at the top of the image.		3	13.64
32	Building is in the centre of the image.	1	11	50.00
33	Building is in the middle and at the bottom of the image.		1	4.55
38	Building is on the right side and in the middle of the image.		1	4.55
48	Building is at the top of the image.		3	13.64
49	Building is in the middle of the image.	1	14	63.64
50	Building is at the bottom of the image.		2	9.09
52	Car is on the far left side of the image.		8	36.36
55	Car is on the far left side and in the middle of the image.		5	22.73
56	Car is on the far left side and at the bottom of the image.		3	13.64
58	Car is on the left side of the image.	1	20	90.91
60	Car is on the left side and at the top of the image.		1	4.55
61	Car is on the left side and in the middle of the image.		7	31.82
62	Car is on the left side and at the bottom of the image.	1	10	45.45
67	Car is in the centre of the image.		1	4.55
73	Car is on the right side and in the middle of the image.		1	4.55
83	Car is at the top of the image.		1	4.55
84	Car is in the middle of the image.		6	27.27
85	Car is at the bottom of the image.	1	6	27.27
86	Car is at the very bottom of the image.		1	4.55
87	Building is near to Car in the real world.	1	14	63.64
88	Car is near to Building in the real world.	1	18	81.82

	Spatial Statements	Sys	Res	%
91	Car is nearer than Building.	1	13	59.09
92	Building is further away than Car.	1	13	59.09

8.3.3 *Analysis and Discussion*

In the following discussion, we refer to assertions with support from more than 40% of people as high support assertions whereas assertions with support from fewer than 40% of people are referred to as low support assertions. In general, the Spatial Semantic Image System asserted most of the statements which received high support from users for all three images in the survey.

Table 8-4 presents contingency tables for each image showing the assertions given high support and given low support in the survey analyses for those assertions generated and also those not generated by the spatial system. Details of the analyses and findings will be discussed further in this section.

Table 8-4 Contingency Tables for Assertions Generated or Not Generated and Given High Support or Low Support

i) **Image 1**

System	**Given by the Users**		Total
	High Support	Low Support	
Generated	12 statements	0 statements	12
Not Generated	3 statements	77 statements	80
Total	15	77	92

ii) **Image 2**

System	**Given by the Users**		Total
	High Support	Low Support	
Generated	11 statements	3 statements	14
Not Generated	6 statements	72 statements	78
Total	17	75	92

iii) **Image 3**

System	**Given by the Users**		Total
	High Support	Low Support	
Generated	9 statements	5 statements	14
Not Generated	2 statements	76 statements	78
Total	11	81	92

Table 8-4 shows that, the system generated 100% of the spatial statements in Image 1 given high support by the users, 79% (11 out of 14) of the statements in Image 2 are also given high support by the users and 64% (9 out of 14) of the statements in Image

3 are given high support by the users. It is clear that most of the spatial statements generated by the system for Image 1 to 3 had high support from the users, and thus were regarded as relevant to the images

It also can be seen in Table 8-4 that only 20% (3 out of 15) of the spatial statements given high support by the users are not generated by the system for Image 1, 35% (6 out of 17 statements) for Image 2 and 18% (2 out of 11 statements) for Image 3. This shows that only a relatively small number of spatial statements given high support by the users are not generated by the system based on the current rules.

On the other hand, the table also shows numbers of spatial statements generated by the system but given low support by the users. There were none for Image 1, but there are 3 statements for Image 2 (4%) and 5 statements for Image 3 (6%).

For Image 1, Table 8-1 shows the system generated 12 spatial statements that are related to the image, where all the 12 spatial statements have been supported by more than 12 users (55%). The spatial statements with highest support by the users are statements 1 and 2 with 18 users (82%) supporting and the lowest ones are statements 8 and 84 with 12 users (5%) supporting. These show that more than 50% of users supported the spatial statements that were generated by the system. Thus, all the spatial statements generated by the SpaSIS in the Image 1 are reliable.

For Image 2, the system generated 14 spatial statements. Table 8-2 shows all the generated spatial statements have been supported by some users with eleven spatial statements given high support by the users, while another three statements were given low support by the users. These three spatial statements are statements 5, 6 and 33 are mark as blue cells which will be discuss further in this section. The spatial statement with highest support by the user is statement 87 with 21 users (95%) supporting and the lowest ones are statements 5 and 6 with only 1 user (5%) supporting. With 11 out of 14 (79%), the analysis shows that the spatial statements generated by the system for Image 2 are reasonably reliable.

For Image 3, the system generated 14 spatial statements. Table 8-3 shows all the generated spatial statements have been supported by some users with nine spatial statements given high support by the users as shown in Table 8-3, while another five statements were given low support by the users. These five spatial statements are statements 3, 4, 5, 6 and 85 are mark as blue cells which will be discuss further in this section. The spatial statement with highest support by the user is statement 58 with 20 users (91%) supporting and the lowest ones are statements 3, 4 and 85 with only 1 user (5%) supporting. With 9 out of 14 (64%), the analysis shows that the spatial statements generated by the system for Image 3 are still reasonably reliable.

All green cells in Table 8-1 to 8-3 shows that the spatial system generated statements relating to the distance position exactly as perceived by the majority of users in the survey. For example statement 87 in Image 2: '*Person1 is near to Person2 in the real world.*' with 95% of users supporting it. The relevance and high reliability of the SpaSIS in detecting and generating spatial terms for relative distance position is obvious.

For Image 2 and 3, the blue cells show eight spatial statements generated by the system but given low support by the users. It can be seen that both images are quite complicated in their own way, which explained the differences and triggered interesting responses from the users. Even though these spatial statements received low support percentages, but there are still users who think the statements are true and agreed with the same statements generated by the system. All these observations show a variety of user perspectives which sometimes may be very inconsistent. Hence, in a way these show that the SpaSIS has the ability to produce spatial assertions that meet some distinct user requirements.

For Image 1-3, the yellow cells show eleven spatial statements that are not generated by the system but given high support by the users. However these only accounts for 3 out of 15 for Image 1 (20%), 6 out of 17 for Image 2 (35%) and 2 out of 11 for Image 3 (18%) as shown in Table 8-4, which shows the percentage of the spatial statements generated by the system are significantly high (more than 60%). Most of these spatial statements cells are related to the absolute position spatial relationships except for statements 10 and 15 for Image 1. The spatial statements 10 and 15 for Image 1 are involved with composite spatial relationships, either below to the left or *right* or *above to the left or right*. Statements 10 and 15 are not detected by the system because the current composite rules did not apply here. However, the system detected statements 1 and 7, which logically suggests the spatial statements of 10 and 15 should also be generated. This observation revealed some logical inconsistency in the current SpaSIS that will be enhanced and discussed further in the next section.

In computing the absolute positions, the centroid of each object is referred to, in order to specify and locate their positions in the image. For example in Image 2, both the main objects of persons might have been used as a baseline by the users and they may have divided the image into two parts based on that. Hence, they may perceive all other statements based on that. If we look at the image closely, both people are in the middle horizontally but they are not in the middle vertically.

For all images, the spatial statements from 47 to 51 and from 82 to 86 are referred to vertically, but this is not stated in the statements. Clearly, the middle in spatial statements 49 and 84 might be referred to as middle (vertically), which is why they are not detected by the system. This misleading information in presenting the spatial statements will be modified and mentioned in the justification.

Overall the results and analysis of the survey do show that user responses can vary significantly but, in a majority of instances the assertions made by the system are supported by a significant number of users.

8.3.4 Justification

From the discussion, we realized that there are some inconsistencies with the logic flow of our computations for relative position and composite spatial relationships, suggesting the need for some modification to the algorithms.

We explain these further based on Figure 8-3, where our initial algorithms are first considered for this situation. These are described rigorously in section 5.3.

1. Assume object_1 has a centroid at O
2. Also assume:
   - The height of object_1 is $2\mathrm{h}_1$ and height of Object_2 is $2\mathrm{h}_2$
   - The width of object_1 is $2\mathrm{w}_1$ and width of Object_2 is $2\mathrm{w}_2$

- Let length of CD be $2(h_1 + h_2)$ and length of AB is $2(w_1 + w_2)$

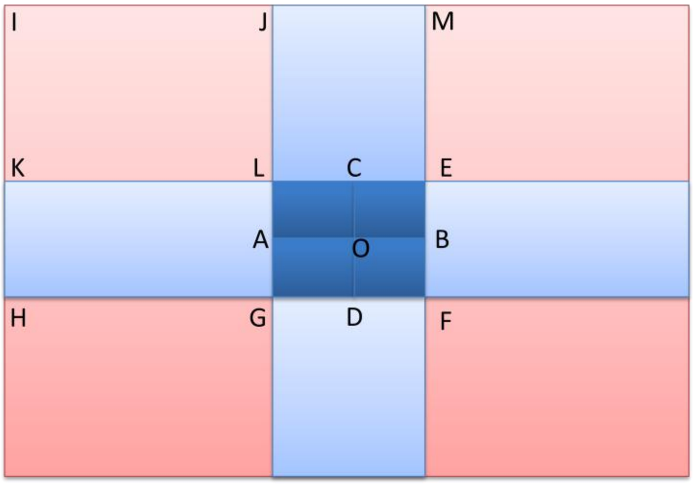


Figure 8-3 Representation of current spatial algorithms.

3. Thus using the original rules from section 5.3:
   - Object_2 is left of object_1 is asserted if centroid of Object_2 is in rectangle CDHK
   - Object_2 is above object_1 is asserted if centroid of Object_2 is in rectangle ABMJ
   - Object_2 is above and left of object_1 is asserted if centroid of Object_2 is in rectangle IJKL
4. Note if the centroid of Object2 is in LCAO only the following are asserted:
   - Object_2 is left of object_1
   - Object_2 is above object_1
5. In IJKL only the following is asserted:
   - Object_2 is above and left of object_1

Some better, more consistent rules satisfying the basic rules of logic are now proposed consisting two approaches, as follows:

New Rules (A)

- Object_2 is left of object_1 when the x coordinate of centroid of object_2 is less than x_{min} of object_1 AND the x coordinate of centroid of object_1 is greater than x_{max} of object_2.

New Rules (B)

- Object_2 is left of object_1 when the x_{max} of object_2 is less than x_{min} of object_1.

- Object_2 is below object_1 when the y coordinate of the centroid of object_2 is less than $\mathrm{y}_{\mathrm{min}}$ of object_1 AND the y coordinate of the centroid of object_1 is greater than $\mathrm{y}_{\mathrm{max}}$ of object_2.

- Object_2 is below object_1 when the $\mathrm{y}_{\mathrm{max}}$ of object_2 is less than $\mathrm{y}_{\mathrm{min}}$of object_1.

Also

- Object_1 is right of object_2 when object_2 is left of object_1.
- Object_1 is above object_2 when object_2 is below object_1.

And we can ‘AND’ any asserted statements ie we don’t need extra rules for composite spatial relationships such as *below to the left or right* etc.

The two new approaches (New Rules A and B) have been developed and implemented. Both approaches together with our old approach are tested in the next image retrieval experiments where retrieval performance is measured based on the queries submitted. As for an absolute position: middle, a precise statement was added to the current system to differentiate between middle (horizontally) and middle (vertically) so that the new modified SpaSIS system detects and generates the information required by the users. All rules will use the same new modified absolute position spatial relationships statements.

8.4 Image Retrieval Performance

The Spatial Semantic Image System has annotated 100 images with spatial relationships based on the modified spatial algorithms. RDF files for each image have been generated to hold the relationships knowledge generated according to the ontologies developed. To retrieve the images a spatial retrieval mechanism has been developed based on the structured query language (SQL) using JAVA. SQL is a standard language used to communicate with a relational database to perform database queries and manipulations (Stair and Reynolds, 2001). The base command, SELECT, is accompanied by many options and clauses used to compose queries from simple to complex, from vague to more specific ones (Plew and Stephens, 2002).

In general, the SpaSIS retrieval system used SQL queries in the form of a SELECT–FROM–WHERE clause for the selection of objects with the spatial relationship terms available in the system. The SQL queries will call relevant images by referring to a folder consisting of RDF files for images containing the spatial ontological annotation statements created earlier in the spatial analysis system and the ontological components. All the images can be accessed by using the URL given in its RDF file.

8.4.1 Retrieval Performance Techniques: Precision and Recall

Precision and recall are widely used for statistical evaluation in information retrieval. Precision can be seen as a measure of exactness or fidelity of a retrieval, whereas recall is a measure of its completeness. In an image retrieval scenario, precision measures the ability of a system to present only relevant images and is defined as the number of relevant images returned by a search divided by the total number of images retrieved by that search. Recall measures the ability of a system to present all relevant images and is defined as the number of relevant images returned by a search divided by the total number of existing relevant images that should be retrieved. The information retrieval metrics formulas are given below:

$$\boxed{\text{Precision} = \frac{\text{number of relevant items retrieved}}{\text{total number of items retrieved}}}$$

$$\boxed{\text{Recall} = \frac{\text{number of relevant items retrieved}}{\text{number of relevant items in collection}}}$$

In statistics, the F-score (also F-measure or F_1 score) is a measure of a test's accuracy. The F-score is the harmonic mean of precision and recall and can be interpreted as a weighted average of the precision and recall, where F score reaches its best value at 1 and worst score at 0. The F score is often used in the field of information retrieval for measuring search, document classification, and query classification performance (Wikipedia, 2011). The formula for F-score is given as:

$$\boxed{\text{F-score} = 2\,.\ \frac{\text{Precision x recall}}{\text{Precision + recall}}}$$

8.4.2 Methodology

The objective of this experiment is to measure and evaluate the image retrieval performed by the Spatial Semantic Image System. The information retrieval metrics of precision and recall are used to measure image retrieval performance. Queries are formed to retrieve relevant images from a collection of images. We use a collection of 100 images downloaded from LabelMe as our initial dataset. Our domain is places of interest, and in this research we are focusing on one of the world's main attractions, which is the Eiffel Tower in Paris. These 100 images contained the Eiffel Tower and other objects such as person, tree, car, bridge, building etc. Some images were labelled just with the Eiffel Tower while others were also labelled with other object as well. The labelling was done by LabelMe users.

The 100 images were processed through the Spatial Semantic Image system to create sets of assertions covering the labelled objects in the images which could be used in the retrieval process. A number of queries have been submitted to LabelMe and the Spatial Semantic Image System for the same dataset. The queries cover queries for absolute position, distance position and relative position. These queries is structured in such a way, from simple to more complicated one, where in the relative position, three different set of rules are tested to be used in retrieving those images. The queries are listed as follows:

1. Find me an image of the Eiffel Tower in the centre of the image.
2. Find me an image of a person near in the image to a person.
3. Find me an image of a person near in the real world to a person.
4. Find me an image of a person who is nearer the camera than the Eiffel Tower.
5. Find me an image of a person who is nearer the camera than a person.
6. Find an image of a person who is right of the Eiffel Tower.
7. Find an image of a person who is below the Eiffel Tower.
8. Find me an image of a person who is below and to the right of the Eiffel Tower.
9. Find me an image of a tree which is left of the Eiffel Tower.
10. Find me an image of a tree which is right of the Eiffel Tower.

Query 1 cover queries for absolute position; Query 2-1 cover for distance position; and Query 6-10 cover for relative position. For the relative position queries, the Spatial System is categorised into 3 different rule sets which are generated based on the description in section 8.3.4, and including Old Rules (OR), New Rules A (NRA) and New Rules B (NRB). OR is based on the original algorithm rules discussed in section 5.3, and NRA and NRB are the new rules justified in 8.3.4.

In the SpaSIS system, the whole query is submitted in a single query/search which consists of one/two objects with a spatial relationship term. However, the search tool in LabelMe only allows for object or scene searching. Queries for two objects for example query 1, 'Eiffel Tower' and 'Person' is submitted separately in LabelMe or by using MatLab code without involving spatial relationships terms. The total images retrieved for both queries are counted as the LabelMe retrieval for the calculation of precision and recall. Then from the precision and recall, the F_1 score is then determined showing the average of the precision and recall in order to present the queries performance.

The ground truth for retrieval performance uses the fact that the objects are manually labelled in the LabelMe system and the correctness of the spatial relations was assessed manually through individual inspection by the author.

8.4.3 Results and Analysis

1. Query 1: Find me an image of *the Eiffel Tower in the centre of the image*.

This query is submitted to evaluate the retrieval performance for absolute position. Table 8-5 shows the results for the query. In LabelMe a query for the Eiffel Tower is submitted which will result in all the images in the collection being retrieved i.e. 100 images including 61 of the relevant images. The number of images retrieved with the SpaSIS system is 56, where 43 images are relevant.

Table 8-5 Retrieval performance for Query 1

Item	LabelMe	SpaSIS
Num of relevant image retrieved	61	47
Total Num of image retrieved	100	57
Precision	0.61	0.82
Num of relevant image in the dataset	61	61
Recall	1	0.78
F-score	0.76	0.80

The number of relevant image in the dataset that matched the Query 1 is 61 images. Table 8-5 shows the precision for Query 1 in the SpaSIS (0.82) is considerably better than for LabelMe (0.61). But, conversely the recall for LabelMe (1) is better than the SpaSIS (0.78). This is because the LabelMe retrieved all images containing the *Eiffel Tower* including all the relevant images. However, The F-score in the SpaSIS (0.8) is better than for LabelMe (0.76).

2. Query 2: Find me an image of *a person near in the image to a person*.

This query is submitted to evaluate the retrieval for distance position using the spatial term *near in the image to*. In the retrieval, LabelMe returns 34 images of person, while SpaSIS returns 12 images. Table 8-6 shows the results for this query.

Table 8-6 Retrieval performance for Query 2

Item	LabelMe	SpaSIS
Num of relevant image retrieved	15	11
Total Num of image retrieved	34	12
Precision	0.44	0.92
Num of relevant image in the dataset	15	15
Recall	1	0.73
F-score	0.61	0.81

Table 8-6 shows the number of relevant images in the dataset that matched Query 2 is 15 images where the precision in the SpaSIS (0.92) is better than LabelMe (0.44). Conversely the recall for LabelMe (1) is better than the SpaSIS (0.73). As before this is because LabelMe retrieved all images containing the *Person* including all the relevant images regardless of whether they are near to each other or not. However the F-score for the SpaSIS (0.81) is better than LabelMe (0.61).

3. Query 3: Find me an image of *a person near in the real world to a person*.

This query is submitted to evaluate the retrieval for distance position for the spatial term *near in the real world to* which is computed using order of magnitude of the objects height. The spatial relationships for *near in the real world to* is generated automatically when the two rules *near in the image to* and *a similar distance to* were satisfied.

Table 8-7 Retrieval performance for Query 3

Item	LabelMe	SpaSIS
Num of relevant image retrieved	13	8
Total Num of image retrieved	34	9
Precision	0.41	0.89
Num of relevant image in dataset	13	13
Recall	1	0.62
F-score	0.58	0.73

In the retrieval, LabelMe returns the 34 images of person, while SpaSIS returns 9 images. Table 8-7 shows the results for Query 3. The number of relevant image in the dataset that matched Query 3 is 13 images and the precision in the SpaSIS (0.89) is better than LabelMe (0.41). Conversely as usual the recall for LabelMe (1) is better than the SpaSIS (0.62). However the average of precision and recall shown in F-score concludes that the SpaSIS (0.73) is better than LabelMe (0.58).

4. Query 4: Find me an image of *a person who is nearer the camera than the Eiffel Tower*.

Query 4 is performed to evaluate the retrieval in distance position for spatial term *nearer the camera than* which also reflects the spatial term *further away from the camera than*. The spatial relationships for *nearer the camera than* is generated based on the use of orders of magnitude of the object heights. In LabelMe, the query for person returns 34 images and query for the Eiffel Tower returns 100 images, while the Spatial System returns 34 images for the single whole query as shown in Table 8-8.

Table 8-8 shows the number of relevant image in the dataset that matched the Query 4 is 34 with precision and recall for both LabelMe and the SpaSIS is 1. This results in a value of 1 in F-score for both systems as well. To further investigate the retrieval for this spatial term, Query 5 is submitted.

Table 8-8 Retrieval performance for Query 4

Item	LabelMe	SpaSIS
Num of relevant image retrieved	34	34
Total Num of image retrieved	34	34
Precision	1	1
Num of relevant image in the dataset	34	34
Recall	1	1
F-score	1	1

5. Query 5: Find me an image of *a person who is nearer the camera than a person*.

Query 5 is submitted also to investigate further the retrieval in distance position for spatial terms *nearer the camera than* and *further away from the camera than*. In LabelMe, Query 5 returns 34 images of person, while SpaSIS returns 13 images as shown in Table 8-9.

Table 8-9 Retrieval performance for Query 5

Item	LabelMe	SpaSIS
Num of relevant image retrieved	11	11
Total Num of image retrieved	34	13
Precision	0.32	0.85
Num of relevant image in the dataset	11	11
Recall	1	1
F-score	0.48	0.92

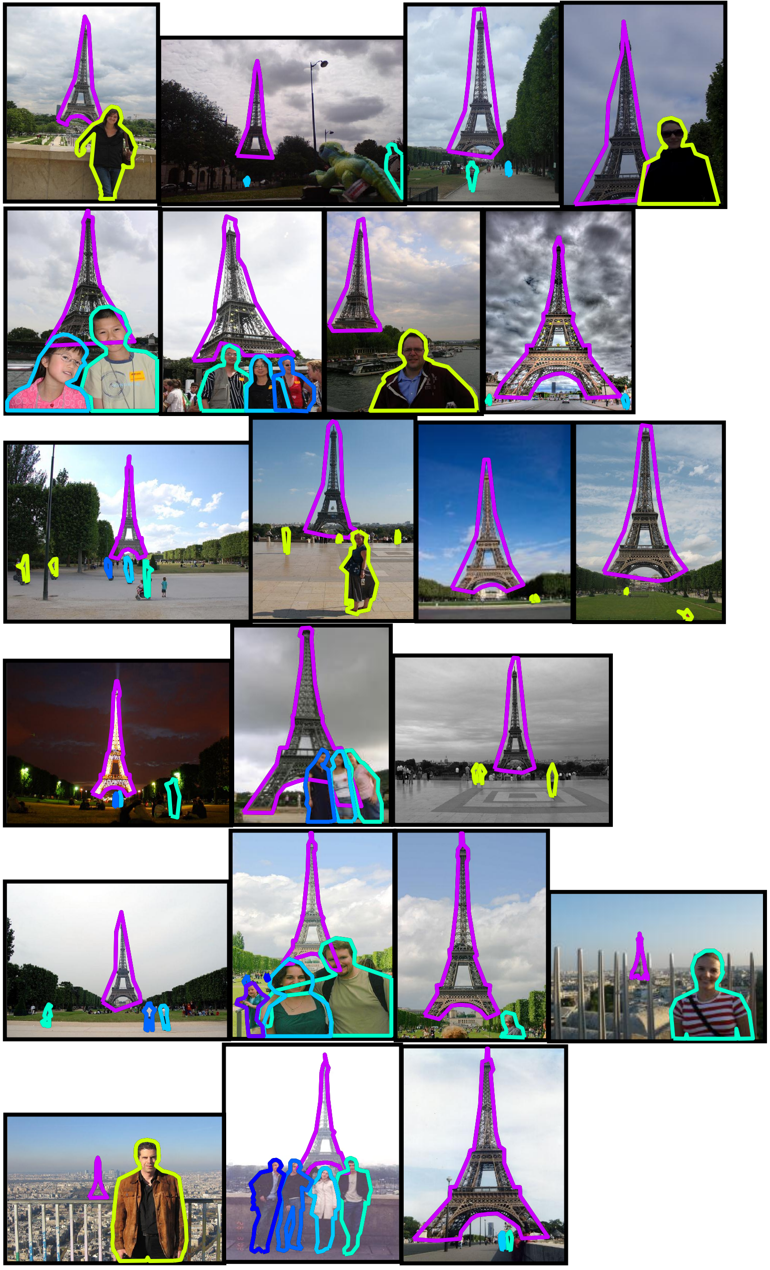
Table 8-9 show the number of relevant image in the dataset that matched the Query 5 is 11. The precision for Query 5 in the SpaSIS (0.85) is better than LabelMe (0.32) and the recall value for both are the same (1).

With higher precision and the same recall, these demonstrate that the SpaSIS system is more powerful in retrieving the images with more relevance and higher reliability. This is presented in the value of F-score where the SpaSIS (0.92) outperforms LabelMe (0.48). The results show that the SpaSIS produce a significant improvement in retrieval performance when distance position spatial relations are involved in the query.

6. Query 6: Find an image of *a person who is right of the Eiffel Tower.*

This query is submitted to assess the retrieval in relative position for the spatial term *is left of* (with the reciprocal spatial term *is right of*). In LabelMe a query for object Person matched 34 images and query for the Eiffel Tower matched 100 images. As for the SpaSIS system, each query has been submitted in one single query. The number of relevant images that matched Query 6 is 22 images are shown in Table 8-10. A bigger size of images retrieved in LabelMe (34) is attached in Appendix C.

Table 8-10 Relevant images for Query 6

Image Name	Relevant Images
1. Eiffel_Tower_000000006 2. Eiffel_Tower_000000024 3. Eiffel_Tower_000000026 4. Eiffel_Tower_000000063 5. Eiffel_Tower_000000074 6. Eiffel_Tower_000000075 7. Eiffel_Tower_000000089 8. Eiffel_Tower_000000165 9. Eiffel_Tower_000000434 10. Eiffel_Tower_000000652 11. Eiffel_Tower_000000656 12. Eiffel_Tower_000000680 13. Eiffel_Tower_000000725 14. Eiffel_Tower_000000951 15. Eiffel_Tower_000000957 16. Eiffel_Tower_000000968 17. Statue_of_Liberty_Paris_000000254 18. Statue_of_Liberty_Paris_000000256 19. Top_of_Eiffel_Tower_000000019 20. Top_of_Eiffel_Tower_000000125 21. torre_eiffel_07_09_altavista 22. torre_eiffel_14_18_altavista	

Results in Table 8-11 show three different results for three different rules: Old Rules (OR), New Rules A (NRA) and New Rules B (NRB). Query 6 returns 22 relevant images in LabelMe, 20 relevant images with SpaSIS Old Rules, 17 with SpaSIS New Rules A and 20 with New Rules B. The results show how the rules affect the retrieval performance in the SpaSIS system.

Table 8-11 Retrieval performance for Query 6

Item	**LabelMe**	**SpaSIS**		
		OR	NRA	NRB
Num of relevant image retrieved	22	20	17	20
Total Num of image retrieved	34	22	17	20
Precision	0.65	0.91	1	1
Num of relevant image in the dataset	22	22	22	22
Recall	1	0.91	0.77	0.91
F-score	0.79	0.91	0.87	0.95

From the Table 8-11, highest precision for Query 6 is retrieved by the SpaSIS based on New Rules A and B with a value of 1, and the lowest value is 0.65 by LabelMe. The precision for the Old Rules is 0.91 where 2 images retrieved are not relevant. However the highest recall is retrieved by LabelMe with a value of 1 as usual followed by SpaSIS with the Old Rules and the New Rules B. The results show how the rules can affect the retrieval performance.

For SpaSIS, Table 8-11 shows that the precision with the Old Rules and the New Rules B are 100% with a value of 1, which are better than the New Rules A. However the Old Rules give the same recall as New Rules B (0.91) but higher compared to the New Rules A (0.77). The New Rules B gives the highest F-score among all.

7. Query 7: Find an image of *a person who is below the Eiffel Tower.*

This query is also submitted to assess the retrieval in other relative position for the spatial term *is belo*w (with the reciprocal spatial term *is above)*. Results in Table 8-12 show the number of relevant images that matched Query 7 is 32 images. Query 7 returns 32 relevant images in LabelMe, 28 relevant images with SpaSIS Old Rules, 26 with SpaSIS New Rules A and 31 with New Rules B.

Table 8-12 shows that the highest precision for Query 7 is retrieved by the SpaSIS based on New Rules and Old Rules with a value of 1, and the lowest value is by LabelMe (0.65). However the highest recall is retrieved by LabelMe (1) followed by the SpaSIS with New Rules B (0.97). In this case, the LabelMe retrieved all images labelled with person regardless of position relative to the Eiffel Tower, so the probability it retrieved a person on the right of the Eiffel Tower is higher compare to the SpaSIS, because the system only retrieved images that contain the Eiffel Tower and person with spatial: *is below*.

Table 8-12 Retrieval performance for Query 7

Item	**LabelMe**	**SpaSIS**		
		OR	NRA	NRB
Num of relevant image retrieved	32	28	26	31
Total Num of image retrieved	34	28	26	32
Precision	0.65	1	1	0.97
Num of relevant image in the dataset	32	32	32	32
Recall	1	0.88	0.81	0.97
F-score	0.97	0.94	0.90	0.97

For the SpaSIS, Table 8-12 shows that the precision in the Old Rules and New Rules A are 100% (1) which are a better than the New Rules B (0.97), but this is not really significantly different. However the New Rules B gives a better recall than others while it’s F-score is the highest and is equal to F-score in LabelMe.

8. **Query 8: Find me an image of *a person who is below and to the right of the Eiffel Tower.***

This query is submitted to assess the retrieval in composite relation, which is a combination of two relative positions in previous queries (Query 6 and 7). The retrieval results for this query are given in Table 8-13.

The number of relevant images in the dataset that matched Query 8 is 21 images. Query 8 returns 21 relevant images in LabelMe, 0 images in Old Rules, 11 in New Rules A and 18 in New Rules B.

The highest precision for Query 8 is retrieved by the SpaSIS based on New Rules A and B with a value of 1, and the lowest value is 0 by the Old Rules. However the highest recall is retrieved by LabelMe with a value of 1 (as usual) followed by the SpaSIS in New Rules B with a value of 0.86.

Table 8-13 Retrieval performance for Query 8

Item	**LabelMe**	**SpaSIS**		
		OR	NRA	NRB
Num of relevant image retrieved	21	0	11	18
Total Num of image retrieved	34	0	11	18
Precision	0.62	0	1	1
Num of relevant image in the dataset	21	21	21	21
Recall	1	0	0.52	0.86
F-score	0.77	0	0.68	0.92

For SpaSIS, Table 8-13 shows that that the precision in for New Rules A and B are 100% with a value of 1, which are better than the Old Rules. However LabelMe gives the highest recall and the New Rules B give the highest F-score among all.

9. Query 9: Find me an image of *a tree which is left of the Eiffel Tower.*

This query is submitted to assess the retrieval for other relative positions. Table 8-14 shows the results for Query 9. The number of relevant images in the dataset that matched the Query 9 is 24 images. Query 9 returns 24 relevant images in LabelMe, 19 images with Old Rules, 15 with New Rules A and 18 with New Rules B.

Table 8-14 Retrieval performance for Query 9

Item	**LabelMe**	**SpaSIS**		
		OR	NRA	NRB
Num of relevant image retrieved	24	19	15	18
Total Num of image retrieved	27	19	15	18
Precision	0.89	1	1	1
Num of relevant image in the dataset	24	24	24	24
Recall	1	0.79	0.63	0.75
F-score	0.94	0.88	0.77	0.86

From the Table 8-14, the highest precision for Query 9 is retrieved by all rules in the SpaSIS with a value of 1, and the lowest value is 0.89 by LabelMe. However the highest recall is for LabelMe with a value of 1 as usual, and the lowest one by the New Rules A (0.63).

For SpaSIS, Table 8-14 shows that the precision in all rules is 1, while the recall with the Old Rules (0.79) is the highest compared to both New Rules A and B. The LabelMe gives the highest F-score among all, and the Old Rules in the SpaSIS gives better F-score than other rules. This is one of the queries where LabelMe give the highest F-score. This might be happening because the images in the dataset contain the labelled object of tree which coincidentally is on the left of the Eiffel Tower. Thus we submitted our next query: Query 10 to see how the retrieval performance might change.

10. Query 10: Find me an image of *a tree which is right of the Eiffel Tower*.

The query is submitted to assess retrieval with the relative position opposite to that for Query 9. Table 8-15 shows the results for Query 9 where the number of relevant image in the dataset is 24 images. The Query 9 returns 24 relevant images in LabelMe, 23 images in Old Rules, 17 in New Rules A and 20 in New Rules B.

Table 8-15 shows the highest precision for Query 10 is retrieved by all rules in the SpaSIS with a value of 1, and the lowest value is 0.89 by LabelMe. However the highest recall is retrieved by LabelMe with a value of 1 as usual.

Table 8-15 Retrieval performance for Query 10

Item	**LabelMe**	**SpaSIS**		
		OR	NRA	NRB
Num of relevant image retrieved	24	23	17	20
Total Num of image retrieved	27	23	17	20
Precision	0.89	1	1	1
Num of relevant image in the dataset	24	24	24	24
Recall	1	0.96	0.71	0.83
F-score	0.94	0.98	0.83	0.91

As for the SpaSIS, Table 8-15 shows that the precision in all rules is 1, while the recall in the Old Rules (0.96) is the highest compared to both New Rules A (0.71) and the New Rules B (0.83). The Old Rules (0.98) gives the highest F-score among all. Again this shows our SpaSIS System performs better than the LabelMe system alone.

8.4.4 Discussion

The results for the retrieval performance are accumulated as shown in Table 8-16. All precision values for the Semantic Spatial Image System are higher than for LabelMe except for the Old Rules in Query 8 which will be discussed further in relation to the rules used. Six of the queries gave precision values in SpaSIS of 1 for query 4, 6, 7, 8, 9 and 10, while the precision for others are more than 0.8. This demonstrates that the SpaSIS presents a better way of retrieving relevant images compared to LabelMe. Also with spatial relationships as a part of the query, the query becomes more precise and contributes to a better or even 100% retrieval performance.

Table 8-16 Retrieval performance for all queries

Query	System/Rules		Precision	Recall	F-score
1	LabelMe		0.60	1	0.71
	Spatial System		0.82	0.78	0.80
2	LabelMe		0.44	1	0.61
	Spatial System		0.92	0.73	0.81
3	LabelMe		0.41	1	0.58
	Spatial System		0.89	0.62	0.73
4	LabelMe		1	1	1
	Spatial System		1	1	1
5	LabelMe		0.32	1	0.48
	Spatial System		0.85	1	0.92
6	LabelMe		0.65	1	0.79
	Spatial System	OR	0.91	0.91	0.91
		NRA	1	0.77	0.87
		NRB	1	0.91	0.95
7	LabelMe		0.65	1	0.97
	Spatial System	OR	1	0.88	0.94
		NRA	1	0.81	0.90
		NRB	0.97	0.97	0.97
8	LabelMe		0.62	1	0.77
	Spatial System	OR	0	0	0
		NRA	1	0.52	0.68
		NRB	1	0.86	0.92
9	LabelMe		0.89	1	0.94
	Spatial System	OR	1	0.79	0.88
		NRA	1	0.63	0.77
		NRB	1	0.75	0.86
10	LabelMe		0.89	1	0.94
	Spatial System	OR	1	0.96	0.98
		NRA	1	0.71	0.83
		NRB	1	0.83	0.91

All the recall values with LabelMe are 1 and always better than the SpaSIS system but this reflects the way the data set was created with restricted objects in the images. However the recall in the SpaSIS system is increased when the query becomes more specific. This we can see from Query 6 compared to Query 3, which demonstrates that the use of spatial relationships as a part of the query enhances the retrieval performance.

Overall the average of retrieval performance F-scores shows that the SpaSIS system outperforms the LabelMe except for Query 9. However to justify this, Query 10 was submitted and our system outperformed the LabelMe system as expected.

Both systems perform 100% in Query 4 and 97% in Query 6. As mentioned before, the recall in LabelMe is always better than the SpaSIS because it's retrieved all the images related including the relevance. So in both Query 4 and 6, its recall has given a major contribution in the F-score.

The use of different rules in the SpaSIS system affected the image retrieval which produced a different recall and precision, thus contributed to a different F-score. The retrieval performance for Query 6-10, demonstrated some significant differences in the result for each rule used in the SpaSIS. The blue colour in Table 8-16 highlights the highest F-score among the rules. The New Rules B outperforms other rules by giving the highest F-scores in Query 6-8, while the Old Rules outperforms the others in Query 9-10.

Hence, if given images with all objects annotated, the Spatial Semantic Image System will return better results in the precision and recall, F-score and outperform the LabelMe in image retrieval performance. For relative position, the best rules in the Spatial System are the New Rules B but this might not always be true, because other factors such as the request/query submitted, the dataset involved, the ground truth etc, might also influence the retrieval performance.

8.5 Conclusion

This chapter discussed the integration of the spatial analysis system, the ontologies and the retrieval system to create the complete Spatial Semantic Image System (SpaSIS). The chapter also presented two distinct experiments to demonstrate the annotation and retrieval performance of the system.

In the first experiment, evaluation on the spatial assertions made by the SpaSIS has been discussed by comparing the results and findings from the user evaluation. The outcome is discussed and some justification has been presented to cater for user variation in perspective and to suggest more powerful algorithms. The study also shows varieties of human interpretation and perspectives, which are sometime inconsistent and may affect the outcome of the retrieval, but still the relevance of our research is confirmed, and we have presented a number of novel contributions.

In the second experiment, the image retrieval performance evaluation provided key evidence that the Spatial Semantic Image System can enhance current image retrieval systems by providing a better annotation with spatial relationships and contributing to an improved and better quality of retrieval performance.

In conclusion, the experimental results and findings demonstrate the feasibility and contribution of the research to enhancing current image annotation and retrieval. For significant uptake of the approach, the availability of robust automatic object annotation at the local level within the image is required. The extensive research in this area is, however, encouraging.

Chapter 9

Conclusions and Future Work

9.1 Introduction

Research in image annotation shows the importance of tools for managing the flow of visual information on the Web in order to satisfy the diversity of users' requirement in bridging the Semantic Gap. The aim of this research has been to develop a new approach for enhancing image annotation and retrieval systems by capturing spatial relationships between labelled regions or objects in images automatically, and supporting the process with ontologies.

The Spatial Semantic Image System (SpaSIS) has been successfully implemented and provides extended annotation features offering users a more comprehensive way to retrieve images. The SpaSIS consists of a proof of concept spatial analysis and retrieval system developed at the annotation and retrieval level. A number of basic evaluation experiments have been conducted to demonstrate the relevance of the research and its novel contributions in image annotation and retrieval.

This chapter reflects on the achieved objectives, the main findings and major novel work. It includes a look towards future work.

9.2 Conclusions

A study was conducted to identify and present the state-of-the-art in the use of spatial relationships in image knowledge extraction and retrieval. This study revealed many opportunities in finding and searching new ways to improve current image retrieval with spatial relationships annotation and with support from ontologies, the image annotation and retrieval may be done in a more specific and systematic way.

We have identified and used a suitable existing annotation tool named LabelMe as a base technology to build a test bed for further experimentation. LabelMe is used for recognizing and annotating objects in images to produce object coordinates as an input to the research. A modest investigation has been reported to study five currently available image annotation tools with detailed evaluation results and discussion. We have designed a research framework and methodology which was used during the development and implementation of the SpaSIS.

A preliminary online questionnaire survey was implemented to understand how people used spatial terminologies in describing spatial relationships among objects in images and to identify spatial terms that are commonly used by them. Although this was a small study and would benefit from a larger sample size with more diverse images, the results from show that there are many spatial terms used by these users and considerable variety in their use. These findings were analysed and discussed. Based on the spatial term frequencies we identified a group of regularly used spatial relationships to be considered in our research.

Based on these findings, spatial algorithms for the selected spatial terminologies were designed, developed and implemented to compute spatial relationships based on existing labels and segmented objects in images. The developments cover relative positions and absolute position leading to 43 spatial terms. The implementation was also demonstrated where these spatial terms were automatically generated.

Extended algorithms for more advanced spatial relationships were also developed and implemented to cover more expressivity in spatial relationships in terms of the 3D environment. Using the Geary-Hinkley transformation (Hayya et al., 1975) we showed that it is possible to extract the *nearer than* and *further away than* relation when object height information is available and certain conditions on the object height distributions are met. Algorithms for the relations *similar distance from the camera*, *near in the image to* and *near in the real world to* were also proposed, developed and tested.

The output of the Semantic Spatial Image System is stored in a knowledge-base. In order to store and extend the knowledge of the spatial annotations and to handle the expressivity of the spatial terms used, we first explored some existing ontologies, but decided to develop our own simple prototypes to test our system. The relation extraction system is integrated with a knowledge-base from the two prototype ontologies including the Spatial Relationships ontology and a Place of Interest ontology acting as the domain ontology.

The domain ontology contains the order of magnitude of heights of objects for further use in spatial relationship computation. By reasoning with the mean and standard deviation of the object’s height in the domain ontology the more advanced spatial terms were obtained.

To demonstrate that our techniques improved image annotation and retrieval, evaluations based on two experiments have been conducted. A survey has been conducted to compare the spatial assertions made by people and those made by the system. Results and findings of this experiment show the potential of our approach. A retrieval performance study has also been performed to demonstrate the quality of precision, recall and F-score of the SpaSIS.

In conclusion, the aim of the research to develop a new approach for enhancing image annotation and retrieval systems by capturing spatial relationships between labelled regions or objects in images, and supporting the process with ontologies has been achieved. The hypothesis is satisfied that the use of spatial relationships in searching images with specific requirements for spatial information has improved the image annotation and increased the performance of the retrieval system. It should be emphasised that this is a prototype system and there is much scope for improvement as discussed further in this chapter under future work. However, the research has created a new method to enhance image annotation for better search and retrieval of images by contributing a constructive spatial semantic approach helping to bridge the Semantic Gap in image annotation and retrieval.

9.3 Novel Contributions

This section lists the three major aspects of the research that have the most novel contributions.

The spatial relationships algorithms for the absolute position, relative position and distance position have been designed, developed and implemented to compute the spatial relationships based on existing labelled objects in images. The development and implementation of all the spatial relationships algorithms demonstrates how spatial relationships concepts can be extracted automatically.

With the implementation of the algorithms, an image’s spatial information is automatically generated by the system and the spatial statements asserted in a knowledge base. The output is in the form of RDF files and consists of information on objects in the image and spatial relationships information between the objects based on two prototype ontologies: the Spatial Relationships Ontology and the Place of Interest Ontology. The knowledge-base representation is essential during retrieval and has been used during the information retrieval evaluation experiments.

In order to extend the expressivity of the spatial annotation and to control the spatial terms used, the system is also integrated with knowledge-bases from the two ontologies that have been developed: the Spatial Relationships Ontology and Place of Interest Ontology. The Place of Interest Ontology acts as a domain ontology and enables the use of the order of magnitude heights of objects. The spatial relationships ontology controls the spatial terminology and provides scope for more flexible query formulation in the future.

9.4 Limitations

While we have made substantial contributions described above, there are still some limitations, as discussed further.

Although the evaluation method proposed in subsection 8.4 in Chapter 8 provides information about the retrieval performance for the spatial algorithms developed in Chapter 5 and Chapter 6, the method of evaluation could be enhanced by using more robust evaluation techniques for concept detection such as those in the NIST TrecVid benchmark suggested by Voorhees (2001) or in other ways such as the method offered by the IAPR-TC12 benchmark (Escalante et al., 2010).

It is also realised that, with the research targeted to the specific domain of images of places of interest, and in the time available we have only been able to use, test and evaluate a small number of carefully selected and relevant images in LabelMe. To evaluate our approach rigorously or for multiple evaluation scales, we would need to use thousands of randomly selected images from a larger number of images or collections such as ImageCLEF (Clough et al., 2007) and Pascal VOC Challenge (Everingham et al., 2010).

Although the evaluation method proposed in this research only shows a comparison between search tool with spatial query and the one without the spatial query facility, a comparison to a similar system might show some further insights and could give better recommendations of how to improve the system.

The scalability of automatic semantic annotation needs further investigation (i.e. improving annotation times and resolving co-reference or ambiguity issue). We rely on people’s annotation in the LabelMe image dataset which is sometime inconsistent, instead of generating our own annotation. Thus quality, correctness and completeness of the annotated image may limit our access to the right image in the collection.

With a trusted source of statistical information about the height of people we have used this information to apply the advanced of spatial relationships to people. If time permitted and we had access to other relevant height information for objects, we could investigate whether these methods could be applied to other object. This is interesting aspect to discover for future research.

9.5 Future Work

Whilst the research has covered much ground, there is still a great deal that could be improved and enhanced further with the following suggestions for future work.

9.5.1 3D Spatial Relationships Algorithms

The computations of the spatial algorithms in the Semantic Spatial Image System are mostly concerned with spatial terms that are limited to the two dimensions in the image plane with some extension to three dimensions for the relative distance position in the more advance spatial relationships algorithms. The system could be enhanced to incorporate more expressivity of spatial information in the 3D environment. For example Lee et al. (2004) had presented the use of a spatial location algebra for 3-D image scenes limited to a number of spatial terms. More 3D spatial terms could be implemented including spatial terms found in the preliminary survey. Some examples of the spatial terms involved are: on, behind, within and around.

To implement this enhancement, the 3D spatial algorithms should take into consideration the criteria of transition (scaling, moving and rotating) of an object in the image, the degree of the perspective view of the object, the environment or scene involved etc. Work on 3-D scene analysis in computer vision will contribute here in the future. By doing this, the system could offer more options and facilities in a more specific way for the user during annotation and retrieval.

At the same time, the spatial relationships algorithms developed with a capability of identifying the relation between objects and their position in an image can be further explored, implemented and adapted in a different application area and more specific domains such as medical and Geographical Information Systems.

9.5.2 Enhancing the Spatial Relationships Ontology

The spatial ontology presented in Chapter 7 is a very basic ontology developed as a proof of concept in order to show how such an ontology could be used in the spatial semantic image system. Although the Spatial Relationships Ontology can capture reciprocal relations crudely, the identification of reciprocal relations and the use of reasoning to infer the reciprocals would be a more intuitive approach. Another improvement would be to structure the knowledge more hierarchically where related terms could be grouped together with one spatial term as a superclass and others in the same group as its subclass. For example, the term FarLeft could be a subclass to the term Left etc. Also the 25 absolute positions could be inferred from conjunctions between any of the five row and five column positions. A similar approach could be taken to the composite concepts such as above left and above right etc.

The introduction of ontologies also offers scope for handling synonyms and for reasoning over the OWL ontologies in a variety of ways. This could be achieved using rules language such SWRL in Protégé(Horrocks et al., 2004), SPIN in TopBraid (Knublauch, 2011) or an Ontological Logic Programming by Sensoy et al. (2011) where the rules related to the spatial terms are computed and could be done in the same platform.

In order to support enhancements for 3D, the Spatial Relationships Ontology could be expanded to include more classes and properties of the new 3D spatial terminologies for spatial relationships. Added knowledge to compute the 3D spatial relationships might also be needed where this information could be retrieved from the domain ontology. Hence the Place of Interest Ontology could be enhanced by including more classes and properties of objects in broader domains. On the other hand, linking to a bigger domain of knowledgebase or ontology such as DBPedia or Geonames would provide other advantages where more objects with the order of magnitude of heights could be retrieved and thereby enhance the functionality of the Spatial Relationships Ontology.

9.5.3 Integration with Other Domain Ontology

Currently, we have integrated the Spatial Relationships Ontology (application ontology) with the Place of Interest Ontology (domain ontology) that contained limited classes and properties to demonstrate the function of the application ontology. The Spatial Relationships Ontology could be used together with other appropriate ontologies such as medical or transportation.

9.5.4 Improving the Retrieval System

In order to improve the retrieval performance of the system, the development of a user semantic relevance feedback might be a good way to involve users in the retrieval process. This feedback could enable the user to respond by ranking the retrieved images. The relevance feedback might also allow the user to select one of the retrieved images, which seems more appropriate to the user to conduct more searches. This will benefit the user as well as the system for future enhancement to improve the quality of retrieval.

At the same time, a visual interface with a point and click representation for the retrieval system would make the process of querying easier and more elegant to use. The interface could be a controlled retrieval interface with 2 or 3 fields specified for object/s and spatial relationships or an open query where a user may input the object/s they are searching for.

9.5.5 Enhancing LabelMe with Spatial Annotation

As we know, the Semantic Spatial Image System gathered an input from LabelMe. LabelMe is an open-source web application. Therefore, the capability and potential that the system has is compatible and could be integrated with LabelMe. By doing this the computation of spatial relationships between objects in the image could be done simultaneously when a user annotates an image in LabelMe. This could be done by adding another tab for spatial annotation that would be generated automatically when the user labelled the object. Hence, LabelMe could offer more facilities and outputs to the user instead of just object annotation, but also spatial annotations between those objects.

9.6 Concluding Remarks

The Spatial Semantic Image System facilitates more specific annotations with better and more effective retrievals within the image annotation and retrieval domains. The key contributions of the system such as the framework, the algorithms, the knowledge-base and the ontologies themselves, provide effective approaches and techniques for annotating objects in the image with spatial relationships information for specific retrievals.

The research methods and findings may be used as a good basis for further investigation and research into using spatial relationships for images within other areas or domains. It is hoped that these findings and contributions will add to the body of knowledge in the area of image annotation and retrieval, while at the same time contributing some new findings to human knowledge as a whole.

References

Abella, A. & Kender, J. R. 1993. Qualitative describing objects using spatial prepositions. *In:* Proceedings of IEEE Workshop on Qualitative Vision. 33-37.

Ahmad, I. & Grosky, W. I. 2003. Indexing and retrieval of images by spatial constraints. *Journal of Visual Communication and Image Representation,* 14**,** 291-320.

Ahn, L. v. & Dabbish, L. 2004. Labeling images with a computer game. *Proceedings of the SIGCHI conference on Human factors in computing systems.* Vienna, Austria: ACM.

Ahn, L. v., Liu, R. & Blum, M. 2006. Peekaboom: a game for locating objects in images. *Proceedings of the SIGCHI conference on Human Factors in computing systems.* Montr\&\#233;al, Qu\&\#233;bec, Canada: ACM.

Akrivas, G., Berrani, S.-A., Douze, M., Heinecke, J., O'Connor, N., Papadopoulos, G. T., Saathoff, C. & Waddington, S. 2007. aceMedia: Knowledge-based Semantic Annotation and Retrieval of Multimedia Content. *1st Workshop on Multimedia Annotation and Retrieval enabled by Shared Ontologies (MAReSO) in SAMT2007* Genova, Italy.

Alcic, S. & Conrad, S. 2011. A Clustering-based Approach to Web Image Context Extraction *MMEDIA 2011 : The Third International Conferences on Advances in Multimedia.* Budapest, Hungary.

Allen, J. F. 1983. Maintaining knowledge about temporal intervals. *Commun. ACM,* 26**,** 832-843.

Arndt, R., Troncy, R. e., Staab, S. & Hardman, L. 2007a. Adding Formal Semantics to MPEG-7: Designing a Well-Founded Multimedia Ontology for the Web. Berlin Heidelberg: Department of Computer Science, Univ. Koblenz-Landau.

Arndt, R., Troncy, R. e., Staab, S., Hardman, L. & Vacura, M. 2007b. COMM: Designing a Well-Founded Multimedia Ontology for the Web. *LNCS Proc. of ISWC/ASWC 2007.* Berlin Heidelberg: Springer-Verlag

Bartolini, I., Ciaccia, P. & Waas, F. 2001. FeedbackBypass: A New Approach to Interactive Similarity Query Processing. *Proceedings of the 27th International Conference on Very Large Data Bases.* Morgan Kaufmann Publishers Inc.

Bashir, A. & Khan, L. 2004. A Framework for Image Annotation Using Semantic Web. *International Workshop in Mining for and from Semantic Web (MSW2004).* Seattle, USA.

Beeson, P., Modayil, J. & Kuipers, B. 2010. Factoring the Mapping Problem: Mobile Robot Map-building in the Hybrid Spatial Semantic Hierarchy. *The International Journal of Robotics Research,* 29 4428-459.

Benitez, A. B. & Chang, S.-F. 2002. Multimedia Knowledge Integration, Summarization and Evaluation. *Int. Workshop on Multimedia Data Mining.* Edmonton, Canada.

Berners-Lee, T., Hendler, J. & Lassila, O. 2001. The Semantic Web. *Scientific American.*

Bloehdorn, S., Petridis, K., Saathoff, C., Simou, N., Tzouvaras, V., Avrithis, Y., Handschuh, S., Kompatsiaris, Y., Staab, S. & Strintzis, M. G. 2005. Semantic Annotation of Images and Videos for Multimedia Analysis. *In:* Euzenat, A. G.-P. a. J. (ed.) *Proc. of the Second European Semantic Web Conference, ESWC 2005.* Heraklion, Crete, Greece.

Burger, T., Popolizio, P. & Troncy, R. 2007. *Multimedia Semantics: Overview of Relevant Tools and Resources* [Online]. W3C Multimedia Semantic Incubator Group. [Accessed 2007].

Burleson, C. 2007. Introduction to the Semantic Web Vision and Technologies - Part 1 - Overview. [Accessed November 2007].

Cai, D., He, X., Li, Z., Ma, W.-Y. & Wen, J.-R. 2004. Hierarchical clustering of WWW image search results using visual, textual and link information. *Proceedings of the 12th annual ACM international conference on Multimedia.* New York, NY, USA: ACM.

Carson, C., Belongie, S., Greenspan, H. & Malik, J. 1997. Region-Based Image Querying. *Proceedings of the 1997 Workshop on Content-Based Access of Image and Video Libraries (CBAIVL '97).* IEEE Computer Society.

Carson, C., Thomas, M., Belongie, S., Hellerstein, J. M. & Malik, J. 1999. Blobworld: a system for region-based image indexing and retrieval. *In:* Third International Conference on Visual Information Systems. Springer, 509-516.

Chakravarthy, A., Ciravegna, F. & Lanfranchi, V. 2006a. AKTiveMedia: Cross-media Document Annotation and Enrichment. *In:* Proc. of the 15th International Semantic Web Conference (ISWC2006).

Chakravarthy, A., Ciravegna, F. & Lanfranchi, V. 2006b. Cross-media document annotation and enrichment. *In:* Proceedings of the 1st Semantic Authoring and Annotation Workshop (SAAW2006).

Chakravarthy, A., Lanfranchi, V. & Ciravegna, F. 2006c. Requirements for multimedia document enrichment. *In:* Proceedings of the 15th international conference on World Wide Web, Edinburgh, Scotland. ACM.

Chang, S. K., Shi, Q. Y. & Yan, C. W. 1986. Iconic indexing by 2D strings. . *In:* IEEE Computer Society Workshop on Visual Languages, Dallas, Texas.

Chang, S. K., Shi, Q. Y. & Yan, C. W. 1987. Iconic indexing by 2-D strings. *IEEE Trans. Pattern Anal. Mach. Intell.,* 9**,** 413-428.

Chang, S. K., Jungert, E. & Li, Y. 1989. Representation and retrieval of symbolic pictures using generalized 2D string. *In:* SPIE Proceedings on Visual Communications and Image Processing, Philadelphia. 1360-1372.

Chebotko, A., Lu, S., Fotouhi, F. & Aristar, A. 2005. An Ontology-Based Multimedia Annotator for the Semantic Web of Language Engineering. *International Journal on Semantic Web & Information Systems, 2005,* 1**,** 50-67.

Chebotko, A., Lu, S., Fotouhi, F. & Aristar, A. 2009. An Ontology-Based Multimedia Annotator for the Semantic Web of Language Engineering. *DBLP Journal*.

Chen, Z., Wenyin, L., Zhang, F., Li, M. & Zhang, H. 2001. Web mining for Web image retrieval. *Journal of the American Society for Information Science and Technology,* 52**,** 831-839.

Clough, P., Grubinger, M., Deselaers, T., Hanbury, A. & MÄuller, H. 2007. Overview of the ImageCLEF 2007 Photographic Retrieval Task. *In Advances in*

Multilingual and Multimodal Information Retrieval, 8th Workshop of the Cross-Language Evaluation Forum, CLEF 2007. Budapest, Hungary: Springer.

Corcho, Ó., Fernández-López, M. & Gómez-Pérez, A. 2007. Ontological Engineering: What are Ontologies and How Can We Build Them? *Semantic Web Services.* Nueva York Premier Reference Source.

Dinesh, R. & Guru, D. S. 2011. Concept Of Triangular Spatial Relationship And B-Tree For Partially Occluded Object Recognition: An Efficient And Robust Approach. *International Journal of Image and Graphics (IJIG),* 10 423-448.

Ding, Y. 2002. Ontology: The enabler for the Semantic Web.

Duineveld, A. J., Stoter, R., R.Weiden, M., Kenepa, B. & Benjamin, V. R. 2000. WonderTools? A comparative study of ontological engineering tools. *Int. J. Human-Computer Studies,* 52**,** 1111-1133.

Egenhofer, M. J. & Franzosa, R. 1991. Point-set topological spatial relations *International Journal of Geographical Information Systems,* 5**,** 161-174.

Enser, P. G. B., Sandom, C. J. & Lewis, P. H. 2005. Automatic annotation of images from the practitioner perspective. *LNCS Proceedings of the 6th ACM international conference on Image and video retrieval (CIVR'05).* Springer Berlin / Heidelberg.

Enser, P. G. B., Sandom, C. J., Hare, J. S. & Lewis, P. H. 2007. Facing the reality of semantic image retrieval *Journal of Documentation,* 63 (4) 465-481.

Escalante, H. J., Carlos A. Hernandez, Jesus A. Gonzalez, A. Lopez-Lopez, Manuel Montes, Eduardo F. Morales, Sucar, L. E. & Villasenor, L. 2010. The Segmented and Annotated IAPR-TC12 Benchmark. *Computer Vision and Image Understanding,* 114.

Everingham, M., Gool, L. V., Williams, C. K. I., JohnWinn & Zisserman, A. 2010. The PASCAL Visual Object Classes (VOC) Challenge. *Int J Comput Vision***,** 303-338.

Fan, J., Gao, Y. & Luo, H. 2004. Multi-level annotation of natural scenes using dominant image components and semantic concepts. *Proceedings of the 12th annual ACM international conference on Multimedia.* New York, NY, USA: ACM.

Fensel, D., Bussler, C., Ding, Y., Kartseva, V., Klein, M., Korotkiy, M., Omelayenko, B. & Siebes, R. 2002. Semantic Web Application Areas. *Oracle White Paper*.

Fernández-López, M., Gómez-Pérez, A., Sierra, J. P. & Sierra, A. P. 1999. Building a chemical ontology using Methontology and the Ontology Design Environment. *Intelligent Systems and their Applications, IEEE,* 14**,** 37-46.

Fernández-López, M. & Gómez-Pérez, A. 2002a. Overview and analysis of methodologies for building ontologies. *The Knowledge Engineering Review,* 17**,** 129-156.

Fernández-López, M. & Gómez-Pérez, A. 2002b. A survey on methodologies for developing, maintaining, integrating, evaluating and reengineering ontologies. *. Deliverable 1.4.*

Flickner, M., Sawhney, H., Niblack, W., Ashley, J., Huang, Q., Dom, B., Gorkani, M., Hafner, J., Lee, D., Petkovic, D., Steele, D. & Yanker, P. 1995. Query by Image and Video Content: The QBIC System. *Computer,* 28**,** 23-32.

Flickr. 2011. *Flickr from Yahoo* [Online]. Available: http://www.flickr.com/ [Accessed].

Grosky, W. I. & Mehrotra, R. 1989. Guest Editors' Introduction: Image Database Management. *Computer,* 22**,** 7-8.

Gruber, T. R. 1993. A translation approach to portable ontology specifications. *Knowl. Acquis.,* 5**,** 199-220.

Gudivada, V. N. 1994. TESSA: an image testbed for evaluating 2-D spatial similarity algorithms. *SIGIR Forum,* 28**,** 17-36.

Halaschek-Wiener, C., Golbeck, J., Schain, A., Grove, M., Parsia, B. & Hendler, J. 2005a. PhotoStuff - An Image Annotation Tool for the Semantic Web. *In:* Gil, Y., Motta, E., Benjamins, V. R. & Musen, M. A. (eds.) *Proc. of the 4th International Semantic Web Conference (ISWC2005).* Galway, Ireland.

Halaschek-Wiener, C., Schain, A., Golbeck, J., Grove, M., Parsia, B. & Hendler, J. 2005b. A flexible approach for managing digital images on the semantic web. *In:* 5th International Workshop on Knowledge Markup and Semantic Annotation, Galway, Ireland. 49-58.

Halaschek-Wiener, C., Golbeck, J., Schain, A., Grove, M., Parsia, B. & Hendler, J. A. 2006. Annotation and Provenance Tracking in Semantic Web Photo Libraries. *In:* International provenance and annotation workshop (IPAW2006) 82-89.

Handschuh, S. & Staab, S. 2003. CREAM: CREAting metadata for the Semantic Web. *Comput. Netw.,* 42**,** 579-598.

Hare, J. S., Lewis, P. H., Enser, P. G. B. & Sandom, C. J. 2006. Mind the Gap: Another look at the problem of the semantic gap in image retrieval. *In:* Multimedia Content Analysis, Management and Retrieval January San Jose, California, USA., 17-19.

Hawke, S., Herman, I. & Prud'hommeaux, E. 2011. W3C Semantic Web Activity. Available: http://www.w3.org/2001/sw/#spec [Accessed 07 November].

Hayya, J., Armstrong, D. & Gressis, N. 1975. A Note on the Ratio of Two Normally Distributed Variables. *Management Science* 21 (11)**,** 1338-1341.

Herman, I. 2001a. W3C Semantic Web FAQ. *W3C Semantic Web* [Online]. Available: http://www.w3.org/2001/sw/SW-FAQ [Accessed January 2008].

Herman, I. 2001b. Specification. *W3C Semantic Web Activity* [Online]. Available: http://www.w3.org/2001/sw/#spec.

Hernandez, D. 1994. *Qualitative Representation of Spatial Knowledge*, Springer-Verlag New York, Inc.

Hoang, N. V., Gouet-Brunet, V., Rukoz, M. & Manouvrier, M. 2010. Embedding spatial information into image content description for scene retrieval. *Pattern Recogn.,* 43**,** 3013-3024.

Hollink, L., Nguyen, G., Schreiber, G., Wielemaker, J. & Wielinga, B. 2004. Adding spatial semantics to image annotations. *In:* 4th International Workshop on Knowledge Markup and Semantic Annotation at ISWC'04. 31-40.

Hunter, J. 1999a. MPEG-7: Behind the Scenes. Corporation for National Research Initiatives.

Hunter, J. 1999b. MPEG-7: Behind the Scenes. D-Lib Magazine Online.

Hunter, J. 2001. *Adding Multimedia to the Semantic Web - Building an MPEG-7 Ontology*. *In:* First International Semantic Web Working Symposium (SWWS'01) California, USA. 261-281.

Ishikawa, Y., Subramanya, R. & Faloutsos, C. 1998. MindReader: Querying Databases Through Multiple Examples. *Proceedings of the 24rd International Conference on Very Large Data Bases.* Morgan Kaufmann Publishers Inc.

ISO/IEC15938-3 & FCD 2001. Information Technology - *Multimedia Content Description Interface - Part 3:* Description Definition Language. Singapore

Kallergi, A., Bei, Y. & Verbeek, F. J. 2009. The Ontology Viewer: Facilitating Image Annotation with Ontology Terms in the CSIDx Imaging Database. *Workshop* *on Visual Interfaces to the Social and the Semantic Web (VISSW2009), IUI2009,*. Florida, USA.

Ko, B., Lee, H.-S. & Byun, H. 2000. Region-Based Image Retrieval System Using Efficient Feature Description. *Proceedings of the International Conference on Pattern Recognition - Volume 4.* IEEE Computer Society.

Ko, B. & Byun, H. 2002. Multiple Regions and Their Spatial Relationship-Based Image Retrieval. *Proceedings of the International Conference on Image and Video Retrieval.* Springer-Verlag.

Lagoze, C. & Hunter, J. 2003. The ABC Ontology and Model. *Journal of Digital Information,* 2**,** 9-11.

Lee, A. J. T. & Chiu, H.-P. 2003. 2D Z-string: a new spatial knowledge representation for image databases. *Pattern Recogn. Lett.,* 24**,** 3015-3026.

Lee, S.-C., Hwang, E. & Han, J.-G. 2006. Efficient Image Retrieval Based on Minimal Spatial Relationships. *Journal of Information Science and Engineering,* 22(2)**,** 461-473.

Lee, S.-Y. & Hsu, F.-J. 1990. 2D C-string: a new spatial knowledge representation for image database systems. *Pattern Recogn.,* 23**,** 1077-1087.

Lee, S. & Hwang, E. 2002. Spatial Similarity and Annotation-based Image Retrieval System. *Proceedings of the Fourth IEEE International Symposium on Multimedia Software Engineering.* IEEE Computer Society.

Lee, S., Hwang, E. & Lee, Y. 2004. Using 3D Spatial Relationships for Image Retrieval by XML Annotation. *LNCS Proc. Computational Science and Its Applications - ICCSA 2004.* Springer Berlin / Heidelberg.

Lempel, R. & Soffer, A. 2002. PicASHOW: pictorial authority search by hyperlinks on the web. *ACM Trans. Inf. Syst.,* 20**,** 1-24.

Lewis, J. R. 1995. IBM computer usability satisfaction questionnaires: psychometric evaluation and instructions for use. *Int. J. Hum.-Comput. Interact.,* 7**,** 57-78.

Li, J. & Gray, R. M. 2000. *Image Segmentation and Compression Using Hidden Markov Models*, Kluwer Academic Publishers.

Li, J., Wang, J. Z. & Wiederhold, G. 2000. IRM: integrated region matching for image retrieval. *Proceedings of the eighth ACM international conference on Multimedia.* Marina del Rey, California, United States: ACM.

Li, J. & Wang, J. Z. 2003. Automatic Linguistic Indexing of Pictures by a Statistical Modeling Approach. *IEEE Trans. Pattern Anal. Mach. Intell.,* 25**,** 1075-1088.

Lowe, D. G. 2004. Distinctive Image Features from Scale-Invariant Keypoints. *International Journal of Computer Vision,* 60(2)**,** 91-110.

Lux, M., Becker, J. & Krottmaier, H. 2003. Caliph & Emir: Semantic Annotation and Retrieval in Personal Digital Photo Library. . *In:* Proceedings of CAiSE'03. 15th Conference on Advanced Information Systems Engineering 16-20 June. Velden, Austria. 85-89.

Lux, M., Klieber, W. & Granitzer, M. 2004. Caliph & Emir: Semantics in Multimedia Retrieval and Annotation. *In:* 19th International CODATA Conference, November. Berlin, Germany.

Lux, M. & Granitzer, M. 2005. Retrieval of MPEG-7 based Semantic Descriptions. *BTW Workshop.* Germany Know Centre and The University of Karlsruhe.

Lux, M. 2009. Caliph & Emir: MPEG-7 photo annotation and retrieval *In:* Proceedings of the seventeen ACM international conference on Multimedia Beijing, China. 925-926.

Martínez, J. M., Koenen, R. & Pereira, F. 2002. MPEG-7: The Generic Multimedia Content Description Standard, Part 1. *IEEE Multimedia***,** 78-87.

Moghaddam, B., Biermann, H. & Margaritis, D. 2001. Regions-of-Interest and Spatial Layout for Content-Based Image Retrieval. *Multimedia Tools Appl.,* 14**,** 201-210.

Morales-González, A. & García-Reyes, E. 2011. Simple object recognition based on spatial relations and visual features represented using irregular pyramids. *Multimedia Tools and Applications***,** 1-23.

Muda, Z. 2008. Ontological Description of Image Content Using Regions Relationships. *ESWC 2008 PhD Symposium* Tenerife, Spain.

Muda, Z., Lewis, P. H., Payne, T. R. & Weal, M. M. 2009. Enhanced Image Annotations Based on Spatial Information Extraction and Ontologies. *In:* IEEE International Conference On Signal & Image Processing 2009 (ICSIPA2009), Kuala Lumpur, Malaysia.: IEEE.

Nabil, M., Ngu, A. H. H. & Shepherd, J. 1996. Picture Similarity Retrieval Using the 2D Projection Interval Representation. *IEEE Trans. on Knowl. and Data Eng.,* 8**,** 533-539.

Nister, D. & Stewenius, H. 2006. Scalable Recognition with a Vocabulary Tree. *Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Volume 2.* IEEE Computer Society.

Noy, N. F. & McGuinness, D. L. 2001. Ontology Development 101: A Guide to creating your first Ontology. *Stanford Knowledge Systems Laboratory Technical Report KSL-01-05 and Stanford Medical Informatics Technical Report SMI-2001-0880* Canada Stanford University.

Oberle, D., Ankolekar, A., Hitzler, P., Cimiano, P., Sintek, M., Kiesel, M., Mougouie, B., Baumann, S., Vembu, S. & Romanelli, M. 2007. *DOLCE ergo SUMO: On foundational and domain models in the SmartWeb Integrated Ontology (SWIntO). Journal of Web Semantics,* 5**,** 156-174.

Ossenbruggen, J. v., Geurts, J., Cornelissen, F., Hardman, L. & Rutledge, L. 2001. *Toward Second and Third Generation Web-Based Multimedia*. *In:* The Tenth International Conference on theWorld WideWeb (WWW10) Hong Kong. ACM 479-488.

Peterson, B. F. 2003. *Learning to see creatively: Design, Color and Composition in Photography,* New York, Amphoto Press.

Petridis, K., Saathoff, C., Anastasopoulos, D., Kompatsiaris, Y. & Staab, S. 2006. M-OntoMat-Annotizer: Image Annotation Linking Ontologies and Multimedia Low-Level Features *In:* 10th Intnl. Conf. on Knowledge Based, Intelligent Information and Engineering Systems. 633-640.

Philipsen, G. 1992. *Speaking Culturally: Explorations in Social Communication,* Albany, New York: State University of New York Press.

Plew, R. & Stephens, R. 2002. *Sams Teach Yourself SQL in 24 Hours,* Indiana, Pearsons Education.

Renn, M., van Beusekom, J., Keysers, D. & Breuel, T. 2010. Automatic Image Tagging Using Community-Driven Online Image Databases

Adaptive Multimedia Retrieval. Identifying, Summarizing, and Recommending Image and Music. *In:* Detyniecki, M., Leiner, U. & Nürnberger, A. (eds.). Springer Berlin / Heidelberg.

Rong, Z. & Grosky, W. I. 2002. Narrowing the semantic gap - improved text-based web document retrieval using visual features. *Multimedia, IEEE Transactions on,* 4**,** 189-200.

Russell, B. C., Torralba, A., Murphy, K. P. & Freeman, W. T. 2008. LabelMe: A Database and Web-Based Tool for Image Annotation. *Int. J. Comput. Vision,* 77**,** 157-173.

Saathoff, C., Petridis, K., Anastasopoulos, D., Kompatsiaris, Y. & Staab, S. 2006. M-OntoMat-Annotizer: Linking Ontologies with Multimedia Low-Level Features for Automatic Image Annotation *the 3rd European Semantic Web Conference (ESWC 2006)* Budva, Montenegro.

Santosh, K. C., Wendlingy, L. & Lamiroy, B. 2010. Using Spatial Relations for Graphical Symbol Description. *2010 International Conference on Pattern Recognition.*

Simou, N., Saathoff, C., Dasiopoulou, S., Spyrou, E., Voisine, N., Tzouvaras, V., Kompatsiaris, I., Avrithis, Y. S. & Staab, S. 2005. An Ontology Infrastructure for Multimedia Reasoning. *VLBV*, 51-60.

Smith, J. R. & Chang, S.-F. 1996. Visual SEEK: a fully automated content based image query system. *In:* Proc. of ACM Multimedia. 87-98.

Smith, J. R. & Chang, S.-F. 1999. Integrated spatial and feature image query. *Multimedia Syst.*, 7**,** 129-140.

Specovius, S., Siewert, R., Doege, J., Schnapauff, D., Denecke, T. & Krefting, D. 2010. Grid based evaluation of a liver segmentation method for contrast enhanced abdominal MRI. *Healthgrid Applications and Core Technologies - Proceedings of HealthGrid 2010,* 159**,** 159-170.

Srikanth, M., Varner, J., Bowden, M. & Moldovan, D. 2005. Exploiting ontologies for automatic image annotation. *In:* Proceedings of the 28th annual international ACM SIGIR conference on Research and development in information retrieval, Salvador, Brazil. ACM, 552-558.

Staab, S., Studer, R., Schnurr, H. P. & Sure, Y. 2001. Knowledge processes and ontologies. *Intelligent Systems, IEEE*, 16**,** 26-34.

Staab, S. 2007. Multimedia Ontology. *Summer School in Multimedia Semantics (SSMS2007).* Glasgow University of *Karlsruhe*.

Stair, R. M. & Reynolds, G. W. 2001. *Principles of Information Systems,* USA, Thomson Learning.

Statistics, N. 2008. Health Survey for England 2007 Latest trends. Available: http://www.ic.nhs.uk/statistics-and-data-collections/health-and-lifestyles-related-surveys/health-survey-for-england/health-survey-for-england-2007-latest-trends-%5Bns%5D.

Sure, Y., Erdmann, M., Angele, J., Staab, S., Studer, R. & Wenke, D. 2002. OntoEdit: Collaborative ontology development for the semantic web. *LNCS Proc. of the 1st Int. Semantic Web Conf. (ISWC 2002)* Sardinia, Italy: Springer.

Tang, J. & Lewis, P. H. 2007. An image based feature space and mapping for linking regions and words. *In:* Ranchordas, A., Arajo, H. & Vitri, J., eds. VISAPP 2007: Proceedings of the Second International Conference on Computer Vision Theory and Applications, , March 8-11. Barcelona, Spain,. INSTICC - Institute for Systems and Technologies of Information, Control and Communication, 29-35.

Tian, Q., Wu, Y. & Huang, T. S. 2000. Combine User Defined Region-of-Interest and Spatial Layout for Image Retrieval. *In:* Proc. IEEE 2000 International Conference on Image Processing (ICIP'2000), Sept. Vancouver, Canada. 746-749.

Trochim, M. K. W. 2006. Research Methods Knowledge Base: Correlation. [Accessed February 2009].

Tsai, C.-F. & Hung, C. 2008. Automatically Annotating Images with Keywords: A Review of Image Annotation Systems *Recent Patents on Computer Science,* 1**,** 55-68.

Tsinaraki, C., Polydoros, P., Kazasis, F. & Christodoulakis, S. 2005. Ontology-Based Semantic Indexing for MPEG-7 and TV-Anytime Audiovisual Content. *Multimedia Tools Appl.,* 26**,** 299-325.

Uschold, M. 2003. Where are the Semantics in the Semantic Web? *AI Magazine.*

Voorhees, E. 2001. The Philosophy of Information Retrieval Evaluation. *Proceeding of the Second Workshop of the Cross-Language Evaluation Forum (CLEF2001).*

Walsh, G. & Golbeck, J. 2010. Curator: a game with a purpose for collection recommendation. *Proceedings of the 28th international conference on Human factors in computing systems.* Atlanta, Georgia, USA: ACM.

Wang, X.-J., Ma, W.-Y., Xue, G.-R. & Li, X. 2004. Multi-model similarity propagation and its application for web image retrieval. *Proceedings of the 12th annual ACM international conference on Multimedia.* New York, NY, USA: ACM.

Wang, Y.-H. 2003. Image indexing and similarity retrieval based on spatial relationship model. *Inf. Sci. Inf. Comput. Sci.,* 154**,** 39-58.

Wikipedia. 2008a. XML. Available: http://en.wikipedia.org/wiki/XML.

Wikipedia. 2008b. Web Ontology Language. Available: http://en.wikipedia.org/wiki/Web_Ontology_Language.

Wikipedia. 2008c. Mark-up Language. Available: http://en.wikipedia.org/wiki/Markup_language.

Wikipedia. 2008d. XML Schema. Available: http://en.wikipedia.org/wiki/XML_Schema.

Wikipedia. 2008e. Semantic Web. 2008. Available: http://en.wikipedia.org/wiki/Semantic_Web [Accessed 2008].

Wikipedia 2011. F1 score.

Wu, L., Luo, S. & Sun, W. 2010. A Novel Object Categorization Model with Implicit Local Spatial Relationship

Advances in Neural Networks - ISNN 2010. *In:* Zhang, L., Lu, B.-L. & Kwok, J. (eds.). Springer Berlin / Heidelberg.

Yi, C., Suh, I., Lim, G., Jeong, S. & Choi, B.-U. 2009. Cognitive Representation and Bayeisan Model of Spatial Object Contexts for Robot Localization

Advances in Neuro-Information Processing. *In:* Köppen, M., Kasabov, N. & Coghill, G. (eds.). Springer Berlin / Heidelberg.

Yong-hong, T., Tie-jun, H. & Wen, G. 2005. Exploiting Multi-Context Analysis in Semantic Image Classification. *Journal of Zheijiang University Science,* 6A(11)**,** 1268-1283.

Yong, R., Huang, T. S., Ortega, M. & Mehrotra, S. 1998. Relevance feedback: a power tool for interactive content-based image retrieval. *Circuits and Systems for Video Technology, IEEE Transactions on,* 8**,** 644-655.

Yuan, J., Li, J. & Zhang, B. 2007. Exploiting spatial context constraints for automatic image region annotation. *Proceedings of the 15th international conference on Multimedia.* Augsburg, Germany: ACM.

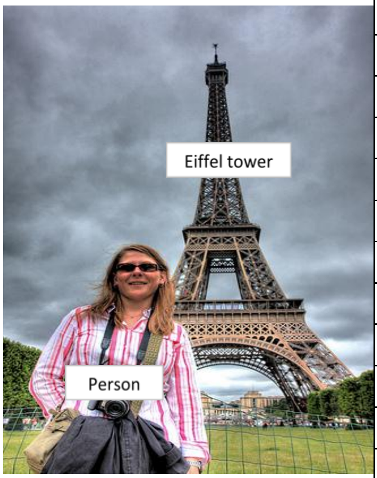
Zhang, D., Islam, M. M. & Lu, G. 2012. A Review on Automatic Image Annotation Techniques. *Pattern Recognition,* 45**,** 346-362.

Zhou, X. M., Ang, C. H. & Ling, T. W. 2001. Image retrieval based on object's orientation spatial relationship. *Pattern Recogn. Lett.,* 22**,** 469-477.

Zhou, X. M., Ang, C. H. & Ling, T. W. 2006. Dynamic interactive spatial similarity retrieval in iconic image databases using enhanced digraph. *Proceedings of the 2006 ACM symposium on Applied computing.* Dijon, France: ACM.

Zhou, Y.-M., Wang, J.-K. & Yang, A.-M. 2008. A Method of Region-based Calculating Image Similarity for RBIR System. *In:* Proceedings of the 9th International Conference for Young Computer Scientists, ICYCS 2008, November 18-21. Zhang Jia Jie, Hunan, China,. IEEE Computer Society, 814-819.

Appendix A The User Evaluation Survey

IMAGE 1		SPATIAL STATEMENTS	TICK (X)
	1	Person is left of Eiffel tower.	
	2	Eiffel tower is right of person.	
	3	Person is right of Eiffel tower.	
	4	Eiffel tower is left of person.	
	5	Person is above Eiffel tower.	
	6	Eiffel tower is below person.	
	7	Person is below Eiffel tower.	
	8	Eiffel tower is above person.	
	9	Person is below and to the right of Eiffel tower.	
	10	Person is below and to the left of Eiffel tower.	
	11	Person is above and to the right of Eiffel tower.	
	12	Person is above and to the left of Eiffel tower.	
	13	Eiffel tower is below and to the right of person.	
	14	Eiffel tower is below and to the left of person.	
	15	Eiffel tower is above and to the right of person.	
	16	Eiffel tower is above and to the left of person.	
	17	Person is on the far left side of the image.	
	18	Person is on the far left side and at the very top of the image.	
	19	Person is on the far left side and at the top of the image.	
	20	Person is on the far left side and in the middle of the image.	
	21	Person is on the far left side and at the bottom of the image.	
	22	Person is on the far left side and at the very bottom of the image.	
	23	Person is on the left side of the image.	
	24	Person is on the left side and at the very top of the image.	
	25	Person is on the left side and at the top of the image.	
	26	Person is on the left side and in the middle of the image.	
	27	Person is on the left side and at the bottom of the image.	
	28	Person is on the left side and at the very bottom of the image.	
	29	Person is in the middle of the image.	
	30	Person is in the middle and at the very top of the image.	
	31	Person is in the middle and at the top of the image.	
	32	Person is in the centre of the image.	
	33	Person in the middle and at the bottom of the image.	
	34	Person in the middle and at the very bottom of the image.	
	35	Person is on the right side of the image.	
	36	Person is on the right side and at the very top of the image.	
	37	Person is on the right side and at the top of the image.	
	38	Person is on the right side and in the middle of the image.	
	39	Person is on the right side and at the bottom of the image.	

40	Person is on the right side and at the very bottom of the image.	
41	Person is on the far right side of the image.	
	SPATIAL STATEMENTS	**TICK (X)**
42	Person is on the far right side and at the very top of the image.	
43	Person is on the far right side and at the top of the image.	
44	Person is on the far right side and in the middle of the image.	
45	Person is on the far right side and at the bottom of the image.	
46	Person is on the far right side and at the very bottom of the image.	
47	Person is at the very top of the image.	
48	Person is at the top of the image.	
49	Person is in the middle of the image.	
50	Person is at the bottom of the image.	
51	Person is at the very bottom of the image.	
52	Eiffel tower is on the far left side of the image.	
53	Eiffel tower is on the far left side and at the very top of the image.	
54	Eiffel tower is on the far left side and at the top of the image.	
55	Eiffel tower is on the far left side and in the middle of the image.	
56	Eiffel tower is on the far left side and at the bottom of the image.	
57	Eiffel tower is on the far left side and at the very bottom of the image.	
58	Eiffel tower is on the left side of the image.	
59	Eiffel tower is on the left side and at the very top of the image.	
60	Eiffel tower is on the left side and at the top of the image.	
61	Eiffel tower is on the left side and in the middle of the image.	
62	Eiffel tower is on the left side and at the bottom of the image.	
63	Eiffel tower is on the left side and at the very bottom of the image.	
64	Eiffel tower is in the middle of the image.	
65	Eiffel tower is in the middle and at the very top of the image.	
66	Eiffel tower is in the middle and at the top of the image.	
67	Eiffel tower is in the centre of the image.	
68	Eiffel tower in the middle and at the bottom of the image.	
69	Eiffel tower in the middle and at the very bottom of the image.	
70	Eiffel tower is on the right side of the image.	
71	Eiffel tower is on the right side and at the very top of the image.	
72	Eiffel tower is on the right side and at the top of the image.	
73	Eiffel tower is on the right side and in the middle of the image.	
74	Eiffel tower is on the right side and at the bottom of the image.	
75	Eiffel tower is on the right side and at the very bottom of the image.	
76	Eiffel tower is on the far right side of the image.	
77	Eiffel tower is on the far right side and at the very top of the image.	
78	Eiffel tower is on the far right side and at the top of the image.	
79	Eiffel tower is on the far right side and in the middle of the	

		image.	
	80	Eiffel tower is on the far right side and at the bottom of the image.	
IMAGE 1		**SPATIAL STATEMENTS**	**TICK (X)**
	81	Eiffel tower is on the far right side and at the very bottom of the image.	
	82	Eiffel tower is at the very top of the image.	
	83	Eiffel tower is at the top of the image.	
	84	Eiffel tower is in the middle of the image.	
	85	Eiffel tower is at the bottom of the image.	
	86	Eiffel tower is at the very bottom of the image.	
	87	Person is near or next to Eiffel tower in real world.	
	88	Eiffel tower is near or next to Person in real world.	
	89	Person is nearer than Eiffel tower.	
	90	Eiffel tower is further away than person.	
	91	Eiffel tower is nearer than person.	
	92	Person is further away than Eiffel tower.	

Thank You!

Appendix B The User Evaluation Survey Results

IMAGE 1 SPATIAL STATEMENTS		SYS	R1	R2	R3	R4	R5	R6	R7	R8	R9	R10	R11	R12	R13	R14	R15	R16	R17	R18	R19	R20	R21	R22	RES	%
1	Person is left of Eiffel tower.	1	1	1		1	1	1	1	1	1	1	1	1	1	1	1			1	1		1	1	18	81.82
2	Eiffel tower is right of person.	1	1	1		1	1	1	1	1	1	1	1	1	1	1				1	1	1	1	1	18	81.82
3	Person is right of Eiffel tower.																	1				1			2	9.09
4	Eiffel tower is left of person.																	1	1			1			3	13.64
5	Person is above Eiffel tower.																							1	1	4.55
6	Eiffel tower is below person.																								0	0.00
7	Person is below Eiffel tower.	1	1			1	1	1	1		1	1	1	1	1	1			1		1	1	1		15	68.18
8	Eiffel tower is above person.	1		1		1	1		1			1	1		1	1			1		1	1	1		12	54.55
9	Person is below and to the right of Eiffel tower.		1							1												1			3	13.64
10	Person is below and to the left of Eiffel tower.			1		1	1	1	1		1	1	1	1	1	1					1		1	1	14	63.64
11	Person is above and to the right of Eiffel tower.																								0	0
12	Person is above and to the left of Eiffel tower.																								0	0
13	Eiffel tower is below and to the right of person.																								0	0

SPATIAL STATEMENTS		SYS	R1	R2	R3	R4	R5	R6	R7	R8	R9	R10	R11	R12	R13	R14	R15	R16	R17	R18	R19	R20	R21	R22	RES	%
14	Eiffel tower is below and to the left of person.																								0	0
15	Eiffel tower is above and to the right of person.		1			1	1		1	1	1	1		1	1	1					1	1	1		13	59.09
16	Eiffel tower is above and to the left of person.												1						1						2	9.09
17	Person is on the far left side of the image.							1			1			1					1			1			5	22.73
18	Person is on the far left side and at the very top of the image.																								0	0
19	Person is on the far left side and at the top of the image.																								0	0
20	Person is on the far left side and in the middle of the image.																								0	0
21	Person is on the far left side and at the bottom of the image.			1				1		1	1			1								1		1	7	31.82
22	Person is on the far left side and at the very bottom of the image.										1			1											2	9.09
23	Person is on the left side of the image.	1		1	1	1	1			1	1	1	1	1	1		1	1		1	1	1	1	1	17	77.27
24	Person is on the left side and at the very top of the image.																								0	0

SPATIAL STATEMENTS		SYS	R1	R2	R3	R4	R5	R6	R7	R8	R9	R10	R11	R12	R13	R14	R15	R16	R17	R18	R19	R20	R21	R22	RES	%
25	Person is on the left side and at the top of the image.																	1							1	4.55
26	Person is on the left side and in the middle of the image.															1									1	4.55
27	Person is on the left side and at the bottom of the image.	1	1	1	1		1		1	1	1		1	1	1	1	1			1	1	1	1		16	72.73
28	Person is on the left side and at the very bottom of the image.				1	1					1		1	1						1		1			7	31.82
29	Person is in the middle of the image.																								0	0
30	Person is in the middle and at the very top of the image.																								0	0
31	Person is in the middle and at the top of the image.																								0	0
32	Person is in the centre of the image.																								0	0
33	Person in the middle and at the bottom of the image.											1				1						1			3	13.64
34	Person in the middle and at the very bottom of the image.																								0	0
35	Person is on the right side of the image.																								0	0

SPATIAL STATEMENTS		SYS	R1	R2	R3	R4	R5	R6	R7	R8	R9	R10	R11	R12	R13	R14	R15	R16	R17	R18	R19	R20	R21	R22	RES	%
36	Person is on the right side and at the very top of the image.																								0	0
37	Person is on the right side and at the top of the image.																								0	0
38	Person is on the right side and in the middle of the image.																								0	0
39	Person is on the right side and at the bottom of the image.																								0	0
40	Person is on the right side and at the very bottom of the image.																								0	0
41	Person is on the far right side of the image.																								0	0
42	Person is on the far right side and at the very top of the image.																								0	0
43	Person is on the far right side and at the top of the image.																								0	0
44	Person is on the far right side and in the middle of the image.																								0	0
45	Person is on the far right side and at the bottom of the image.																								0	0
46	Person is on the far right side and at the very																								0	0

	bottom of the image.																									
SPATIAL STATEMENTS		**SYS**	**R1**	**R2**	**R3**	**R4**	**R5**	**R6**	**R7**	**R8**	**R9**	**R10**	**R11**	**R12**	**R13**	**R14**	**R15**	**R16**	**R17**	**R18**	**R19**	**R20**	**R21**	**R22**	**RES**	**%**
47	Person is at the very top of the image.																								0	00
48	Person is at the top of the image.																	1							1	4.55
49	Person is in the middle of the image.										1														1	4.55
50	Person is at the bottom of the image.	1		1	1	1	1		1	1		1	1	1	1	1				1	1	1	1		15	68.18
51	Person is at the very bottom of the image.				1	1							1	1								1			5	22.73
52	Eiffel tower is on the far left side of the image.																								0	0
53	Eiffel tower is on the far left side and at the very top of the image.																								0	0
54	Eiffel tower is on the far left side and at the top of the image.																								0	0
55	Eiffel tower is on the far left side and in the middle of the image.																								0	0
56	Eiffel tower is on the far left side and at the bottom of the image.																								0	0
57	Eiffel tower is on the far left side and at the very bottom of the image.																								0	0
58	Eiffel tower is on the left side of the image.																								0	0

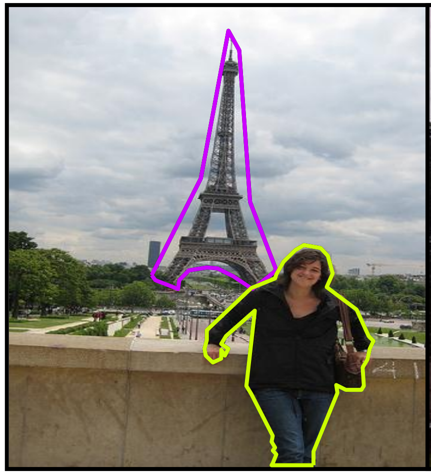
SPATIAL STATEMENTS		SYS	R1	R2	R3	R4	R5	R6	R7	R8	R9	R10	R11	R12	R13	R14	R15	R16	R17	R18	R19	R20	R21	R22	RES	%
59	Eiffel tower is on the left side and at the very top of the image.																								0	0
60	Eiffel tower is on the left side and at the top of the image.																								0	0
61	Eiffel tower is on the left side and in the middle of the image.																								0	0
62	Eiffel tower is on the left side and at the bottom of the image.																								0	0
63	Eiffel tower is on the left side and at the very bottom of the image.																								0	0
64	Eiffel tower is in the middle of the image.		1	1	1		1		1		1		1			1		1		1	1	1	1		13	59.09
65	Eiffel tower is in the middle and at the very top of the image.																								0	0
66	Eiffel tower is in the middle and at the top of the image.				1								1												2	9.09
67	Eiffel tower is in the centre of the image.			1	1						1				1			1				1			6	27.27
68	Eiffel tower in the middle and at the bottom of the image.				1															1					2	9.09
69	Eiffel tower in the middle and at the very bottom of the image.																								0	0

SPATIAL STATEMENTS		SYS	R1	R2	R3	R4	R5	R6	R7	R8	R9	R10	R11	R12	R13	R14	R15	R16	R17	R18	R19	R20	R21	R22	RES	%
70	Eiffel tower is on the right side of the image.	1			1	1					1		1	1	1	1	1			1	1	1	1	1	13	59.09
71	Eiffel tower is on the right side and at the very top of the image.					1																			1	4.55
72	Eiffel tower is on the right side and at the top of the image.				1	1				1			1	1						1					6	27.27
73	Eiffel tower is on the right side and in the middle of the image.	1			1				1		1	1	1	1	1	1		1		1	1	1	1		13	59.09
74	Eiffel tower is on the right side and at the bottom of the image.				1															1					2	9.09
75	Eiffel tower is on the right side and at the very bottom of the image.																								0	0
76	Eiffel tower is on the far right side of the image.																1					1			2	9.09
77	Eiffel tower is on the far right side and at the very top of the image.																								0	0
78	Eiffel tower is on the far right side and at the top of the image.																								0	0
79	Eiffel tower is on the far right side and in the middle of the image.										1											1			2	9.09
80	Eiffel tower is on the far right side and at the bottom of the image.																								0	0

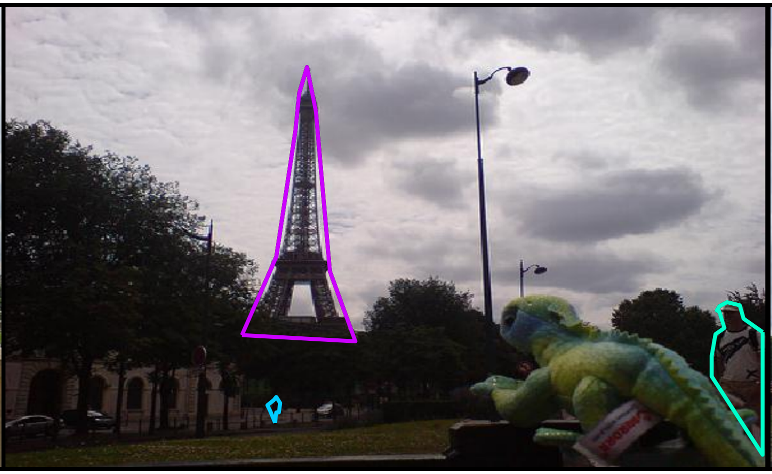
SPATIAL STATEMENTS		SYS	R1	R2	R3	R4	R5	R6	R7	R8	R9	R10	R11	R12	R13	R14	R15	R16	R17	R18	R19	R20	R21	R22	RES	%
81	Eiffel tower is on the far right side and at the very bottom of the image.																								0	0
82	Eiffel tower is at the very top of the image.					1																			1	4.55
83	Eiffel tower is at the top of the image.				1	1							1	1						1					5	22.73
84	Eiffel tower is in the middle of the image.	1		1	1		1	1	1				1		1			1		1	1	1	1		12	54.55
85	Eiffel tower is at the bottom of the image.				1																				1	4.55
86	Eiffel tower is at the very bottom of the image.																								0	0
87	Person is near to Eiffel tower in real world.			1			1								1			1					1		5	22.73
88	Eiffel tower is near to Person in real world.						1								1								1		3	13.64
89	Person is nearer than Eiffel tower.	1	1	1		1		1	1		1	1	1		1	1			1	1	1	1	1	1	16	72.73
90	Eiffel tower is further away than person.	1	1	1		1		1	1		1	1	1		1	1			1	1	1	1	1	1	16	72.73
91	Eiffel tower is nearer than person.										1								1						2	9.09
92	Person is further away than Eiffel tower.										1								1						2	9.09

Appendix C The Retrieval Images

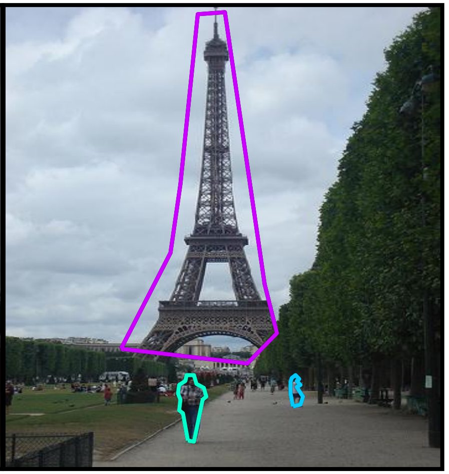
22 of images contained Person that is relevant to Query 6: *a person is right of the Eiffel Tower.*



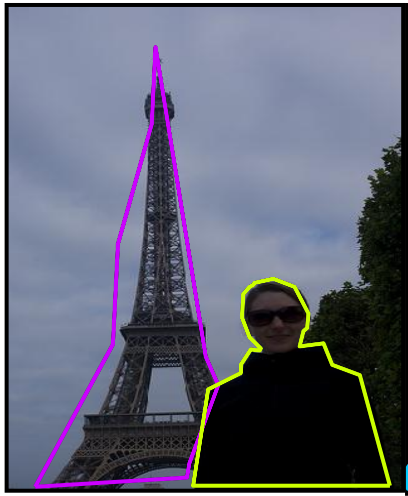
Eiffel_Tower_000000006



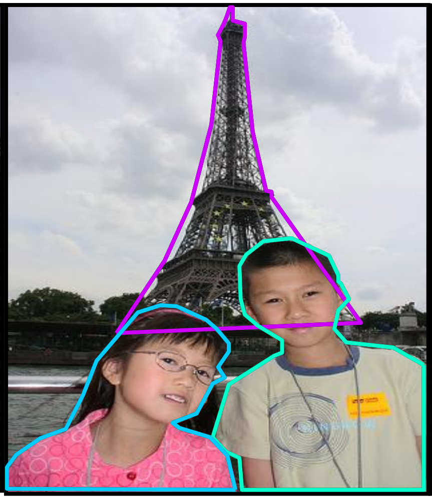
Eiffel_Tower_000000024



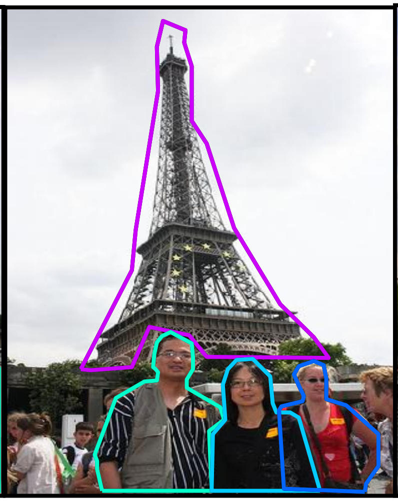
Eiffel_Tower_000000026



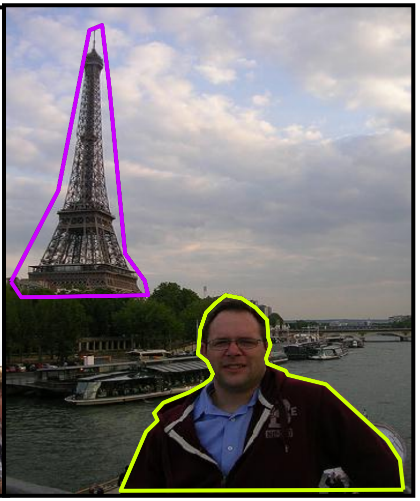
Eiffel_Tower_000000063



Eiffel_Tower_000000074

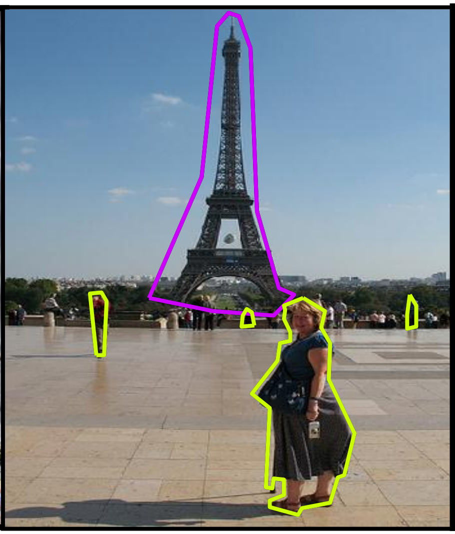


Eiffel_Tower_000000075

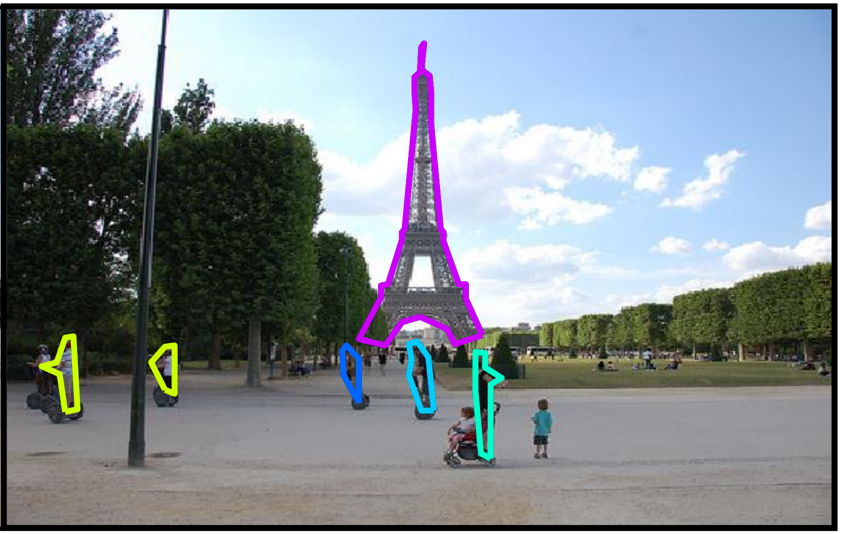


Eiffel_Tower_000000089

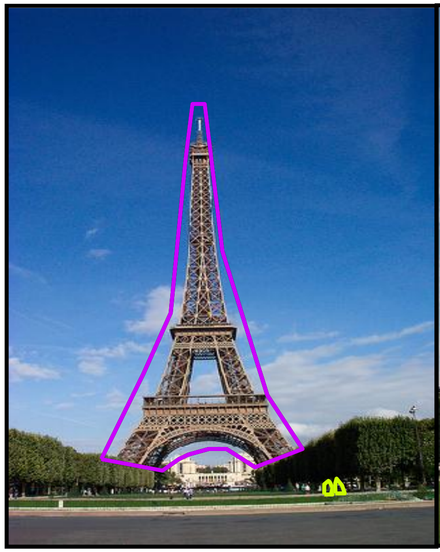
Eiffel_Tower_000000165



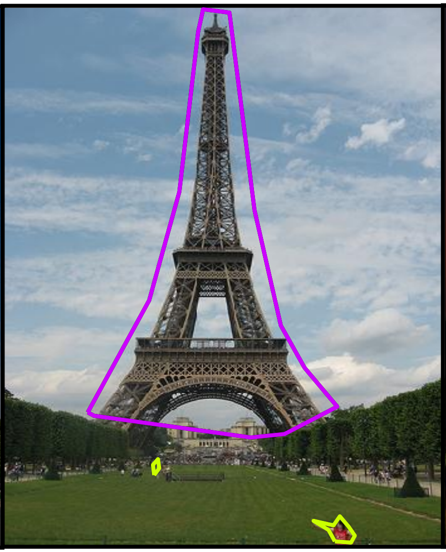
Eiffel_Tower_000000434



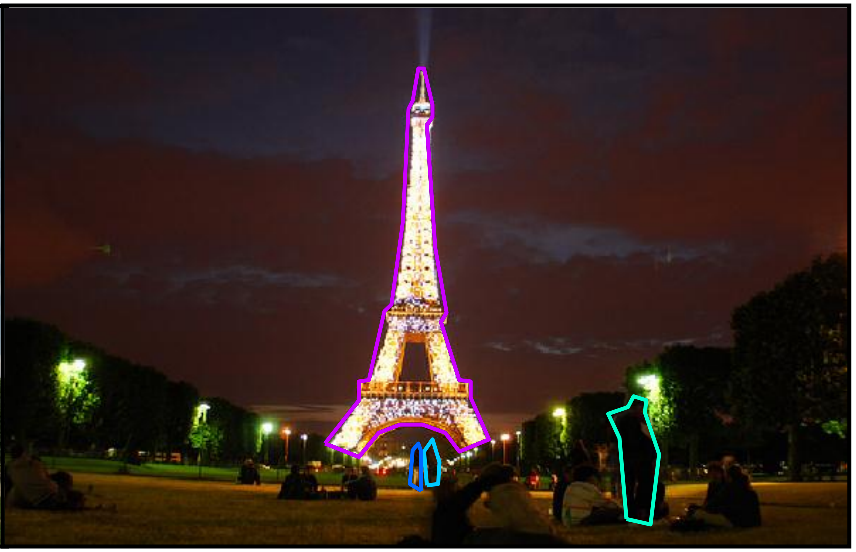
Eiffel_Tower_000000652



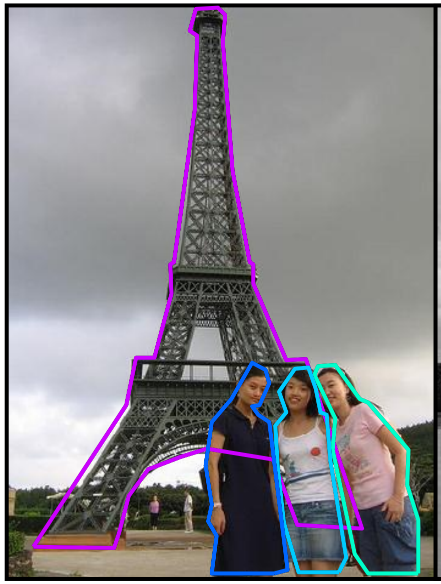
Eiffel_Tower_000000656



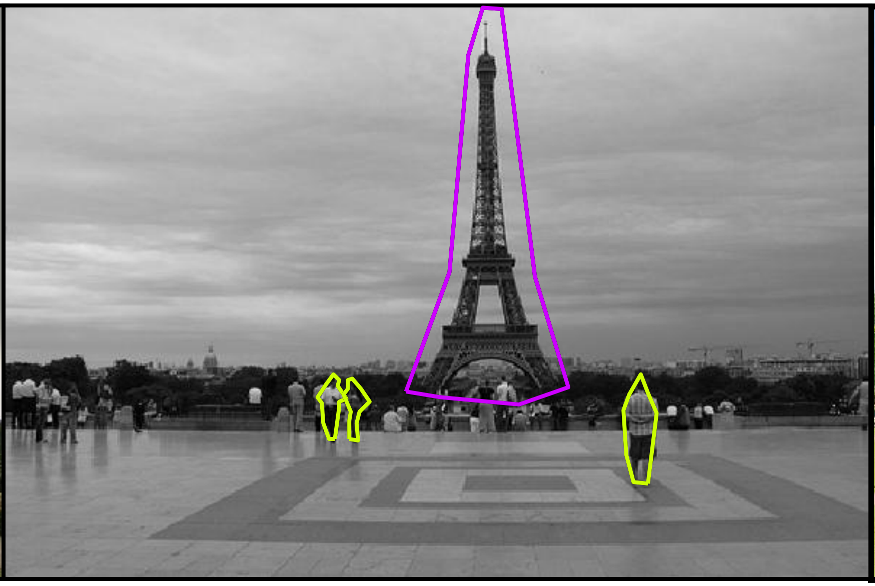
Eiffel_Tower_000000680



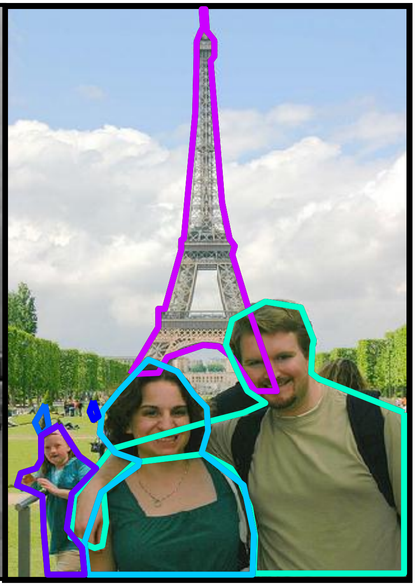
Eiffel_Tower_000000725



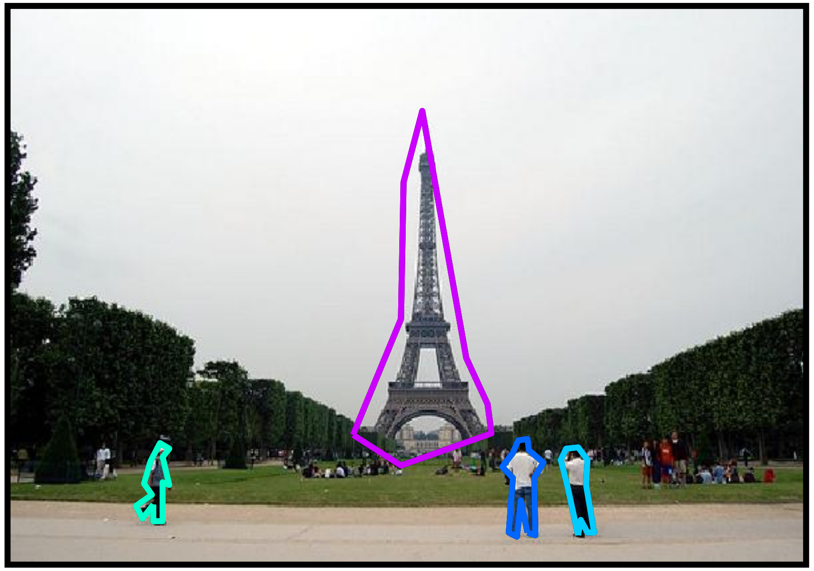
Eiffel_Tower_000000951



Eiffel_Tower_000000957

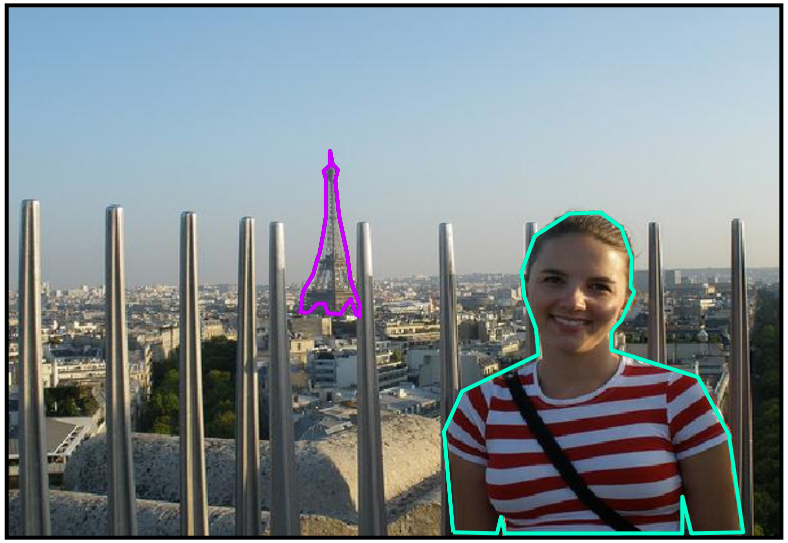


Statue_of_Liberty_Paris_000000254

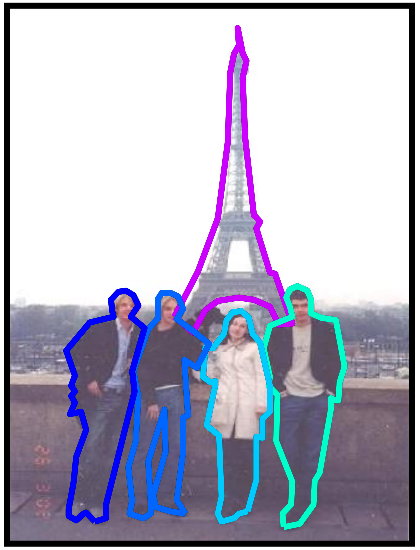


Eiffel_Tower_ 000000968

Statue_of_Liberty_Paris_ 000000256

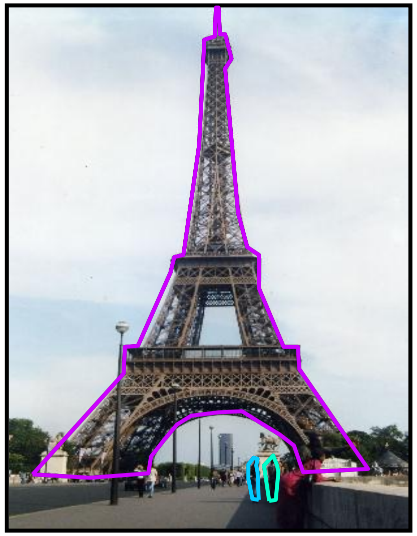


Top_of_Eiffel_Tower_
000000019



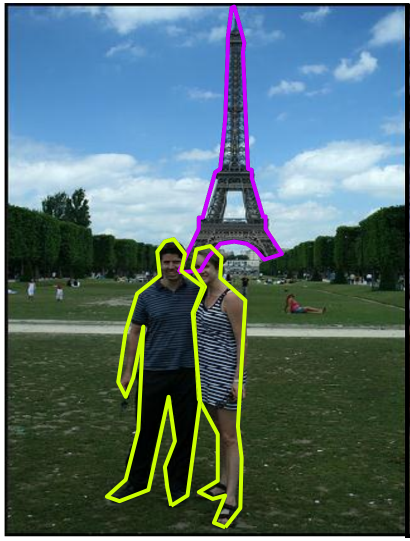
torre_eiffel_07_
09_altavista

Top_of_Eiffel_Tower_
000000125

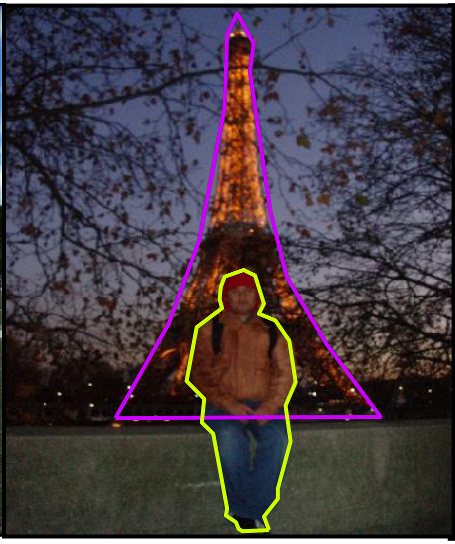


torre_eiffel_14_
18 altavista

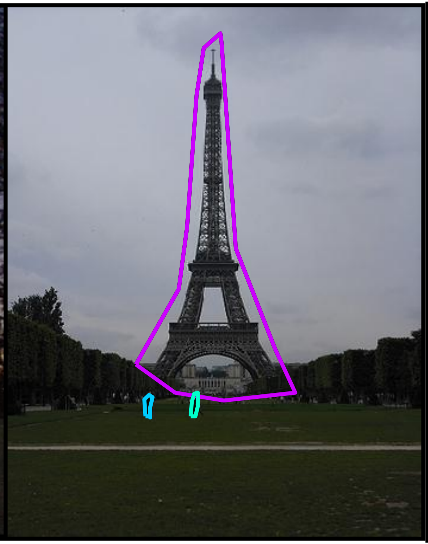
12 of images contained Person that is not relevant to Query 6



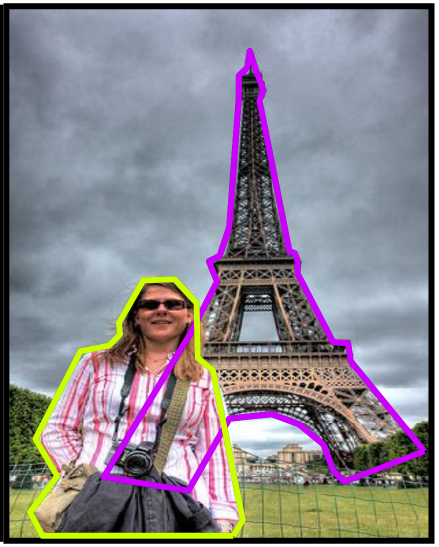
Eiffel_Tower_000000099



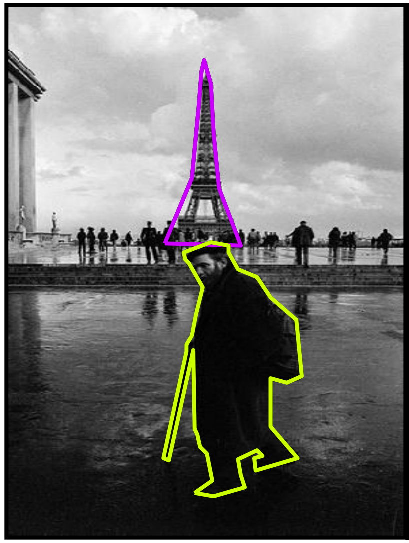
Eiffel_Tower_000000107



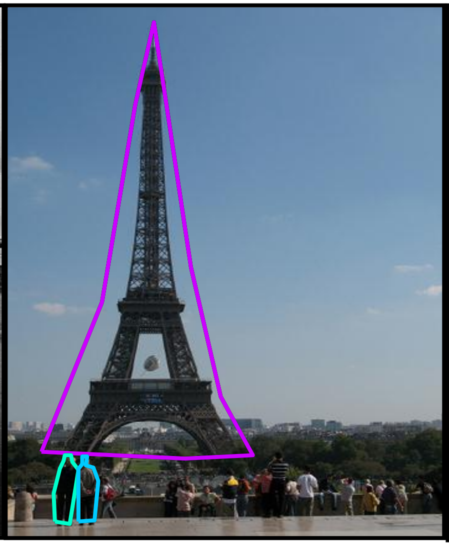
Eiffel_Tower_000000148



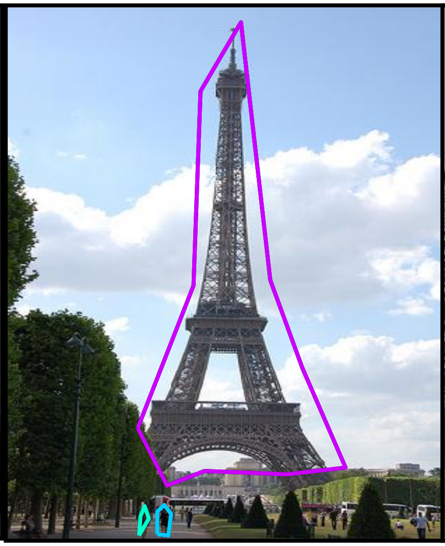
Eiffel_Tower_000000173



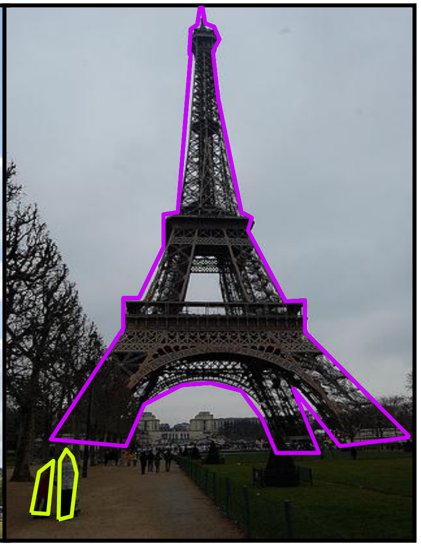
Eiffel_Tower_000000178



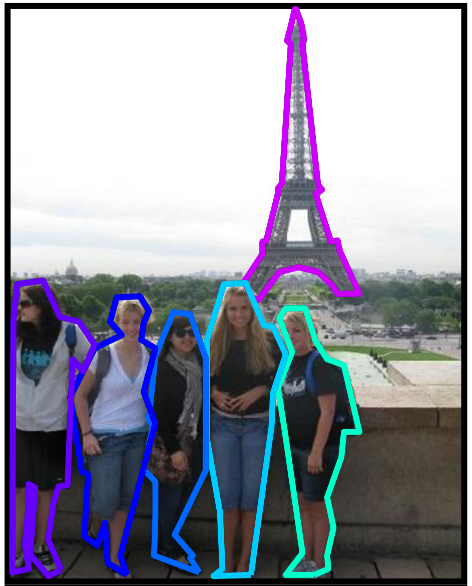
Eiffel_Tower_000000383



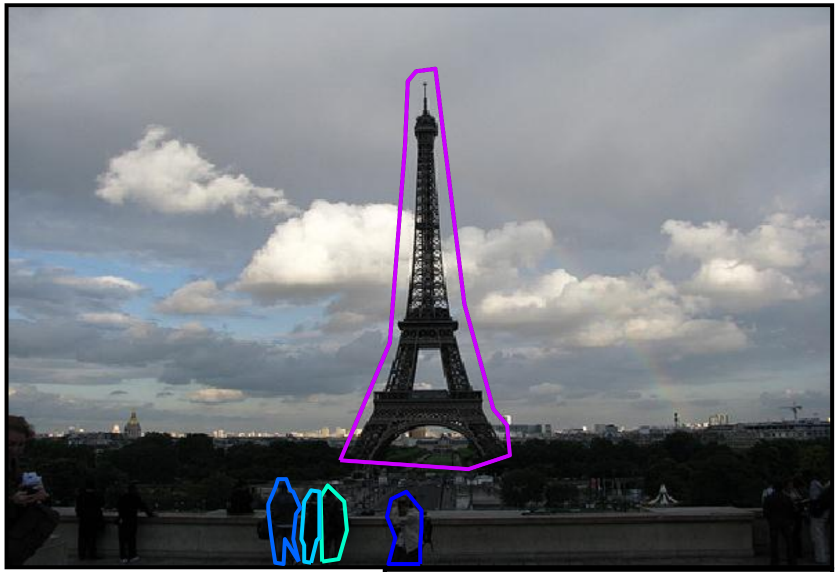
Eiffel_Tower_000000658



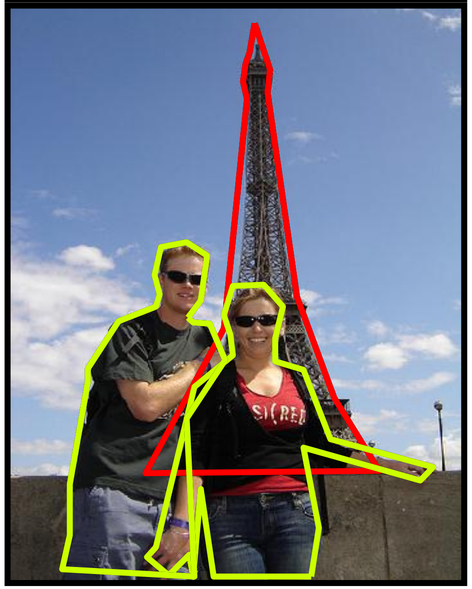
Eiffel_Tower_000000689



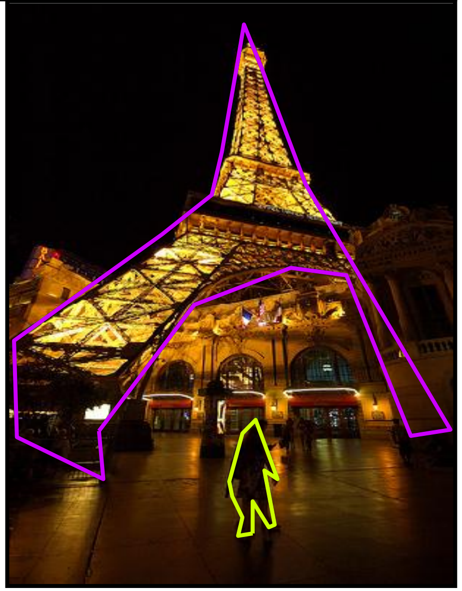
Eiffel_Tower _000000696



Eiffel_Tower_ 000000701



Eiffel_Tower_ 000000986



Eiffel_Tower _000000990

Appendix D Paper Published in ESWC2008

Ontological Description of Image Content Using Regions Relationships

Zurina Muda

School of Electronics and Computer Science
University of Southampton, United Kingdom
{zm06r@ecs.soton.ac.uk}

Extended Abstract

Keywords: Spatial relationships, image annotation and ontology.

1 Research Problem And Aim

Rapid growth in the volume of multimedia information creates new challenges for information retrieval and sharing, and thus anticipates the emergence of the Semantic Web [2, 3]. The principal component in most of multimedia applications is the use of visual information and new approaches are essential to improve the inferring of semantic relationships from low-level features for semantic image annotation and retrieval. Much initial research on image annotation represents images in terms of colours, texture, blobs and regions, but pays little attention to the spatial relationships between regions or objects. Annotations are most frequently assigned at the global level [17] and even when assigned locally the extraction of relational descriptors is often neglected. However, current annotation system might recognise and identify a beach and an ocean in an image but fail to represent the fact that they are next to each other. Therefore, to enrich the semantic description of the visual information, it is important to capture such relations.

The aim of this research is an attempt to develop a new approach or technique for enhancing annotation systems, either through automatic or semi-automatic means, by capturing the spatial relationships between labelled regions or objects in images and incorporating such knowledge in a knowledge base such as an ontology. By this means, human users and software agents alike will be able to search, retrieve and analyze visual information in more powerful ways.

2 Related Work

Ontologies play an important role for knowledge intensive applications to enable content-based access, interoperability and communication across the Web. These ontologies become the backbone for enabling the Semantic Web [20]. The number of multimedia ontologies available is still rather small, and well-designed ontologies that fulfill the requirements [5] of reusability, MPEG-7 compliance, extensibility, modularity and interoperability are rare [18]. The COMM ontology which is under development elsewhere and is based on DOLCE ontology as a foundational ontology is of particular relevance.

A pure combination of traditional text-based and content-based approaches is not sufficient for dealing with the problem of image retrieval on the Web, mostly because of the problem of its text based orientation. Some Web images have irrelevant, few or even no surrounding texts. Thus, the problem of limited collateral text for the annotation of images needs to be solved. Besides, manual image annotation is a tedious task and often it is difficult to make accurate annotations on images. There are many annotation tools available but human input is still needed to supervise the process. So, there should be a way to minimize the human input by making the annotation process semi or fully automatic. In the latter case, although there is much research on automatic image annotation, the results often do not really satisfy the retrieval requirements because of the flexibility and variety of user needs.

To date, many contend-based image retrieval research systems, frameworks and approaches have been reported. Li et. al[14] presented Integrated Region Matching, a similarity measure for region based image similarity comparison. Ko & Byun[8] used Hausdorff Distance to estimate spatial relationships between regions in their Integrated Finding Region In the Pictures (IFRIP) as extension to their previous FRIP [9]. Laaksonen et. al[10] proposed a context-adaptive analysis of image content, by using automatic image segmentation. Lee at. al[12] proposed a new domain-independent spatial similarity and annotation-based image retrieval system. Zhou et. al [21] proposed an approach for computing the orientation spatial similarity between two symbolic objects in an image. Wang [18] proposed a new spatial-relationship representation model called two dimension begin-end boundary string (2D Be-string), based on previous research in 2D String [11]. Ahmad & Grosky [2] proposed a symbolic image representation and indexing scheme to support retrieval of domain independent spatially similar images.

However, all the research in spatial relationships has been pursued independently without taking into consideration the problems of integrating them with an ontology. Such integration would be valuable in producing high level semantics by making semantic annotation systematically easier and more meaningful. In doing so, existing ontologies such as DOLCE and COMM will be evaluated to identify both their relevance and effectiveness in achieving the research aim.

3 Contributions And Evaluation

As part of a preliminary experiment, a comparative analysis of three existing annotation tools has been carried out: Caliph & Emir [15], AKTive Media [6], and M-OntoMat-Annotizer [16]. Each of these tools has been explored individually by using a group of images and a comparative study based on an evaluation framework adapted from Lewis[13] and Duineveld[7] has been performed and results obtained. The comparative study investigated image description features (including annotation) and user interface components to find out the capabilities of existing image descriptions tools and to establish whether the spatial relationships are included and, if so, what the relationships might be. For image description components, follow-up with the developer of the tools has been established to ensure the reliability of the result.

The study shows that, each of the tools offered some special features compared to others and all tools were involved with manual annotations of the whole image. In addition M-OntoMat-Annotizer and AKTive Media allowed segmentation and annotation of the selected regions in images. Caliph & Emir and AKTive Media support some relations but not spatial relationships. Neither of these tools considered the specific locations of objects nor regions in the image for annotation or retrieval.

Based on the study and the previous research, currently, several existing annotation or description tools enable automatic segmentation by grouping multiple regions together and use manual annotation to annotate those regions. By adding the locator description where spatial relationships are considered, the knowledge of the image content becomes more specific and retrieval could be more efficient and performed in an explicit way.

This research will use existing automatic segmentation algorithm when available and manual combining of regions into composite regions for recognised objects. These will be manually annotated in the first instance together with spatial relationships between the objects. From there, an automatic annotation of spatial relationships among the objects in the image plane could be developed based on various available approaches by integrating directional and topological representation of spatial relationships. The process is simplified as illustrated in Fig.2.

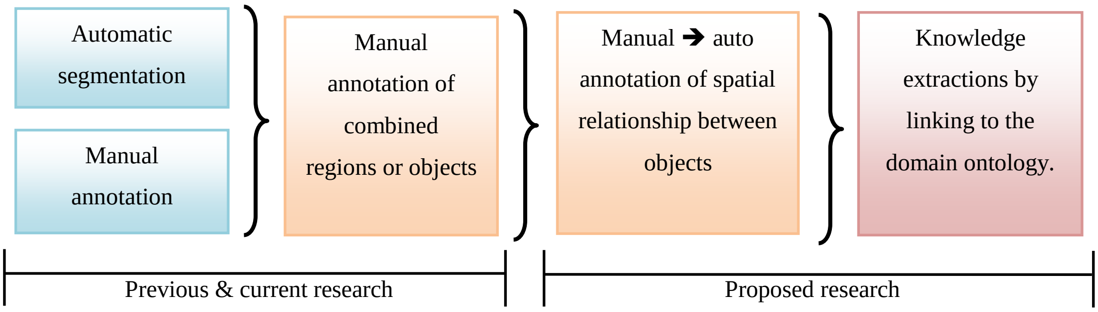

Automatic segmentation
Manual annotation
Manual annotation of combined regions or objects
Manual ➔ auto annotation of spatial relationship between objects
Knowledge extractions by linking to the domain ontology.
Previous & current research
Proposed research

Fig.2. Research outline.

Therefore the expected contribution will be a new approach or technique to automate spatial relationships extraction between the composite regions or objects in images and linking the knowledge to an extended multimedia ontology. The approach or technique should be reliable in order to counter the uncertainty of matching images with the real world cases. For example, this is how it would works when given an image of a beach:

Existing tool would provide the annotation of regions of the image corresponding to: the beach, the ocean, the sky and the coconut tree objects are recognised.

Our approach then identify that: a. The coconut tree is within the beach; b. The beach is next to the ocean; c. The ocean is below the sky.

By reasoning over appropriate domain ontology, and exploiting the entailed spatial relationships, we would be able to infer that if the beach is in Hawaii, then the ocean must be the Pacific Ocean.

For the time being, the domain of the research would be a subset of everyday scenes such as city scenes or places of interest, but later other domains such as medical domain, may be considered to test the generality of the approach. Evaluation on ground truth with spatial relationships in term of precision and recall test will be made to see how well the automated extraction of spatial relationships has been achieved. The evaluation will use sufficient images such as Corel dataset to ensure statistical significance of the result obtained.

4 Work Plan

In order to accomplish the aim, the research plan is assigned into two levels – a macro plan using a Gantt chart for general activities and corresponding timelines, and micro plan using a K-chart [1] for the specific planning and execution of research. The research framework is illustrated in Fig. 3 and consists of:

Annotation component – automatically extracts and identifiers spatial relationships between multiple segmented regions or objects.

Ontological component – logics and reasoning of the extended existing multimedia ontology specifically in terms of spatial descriptors and locators.

Retrieval component – image retrieval mechanisms based on spatial relationships to evaluate the functionality and effectiveness of the approach.

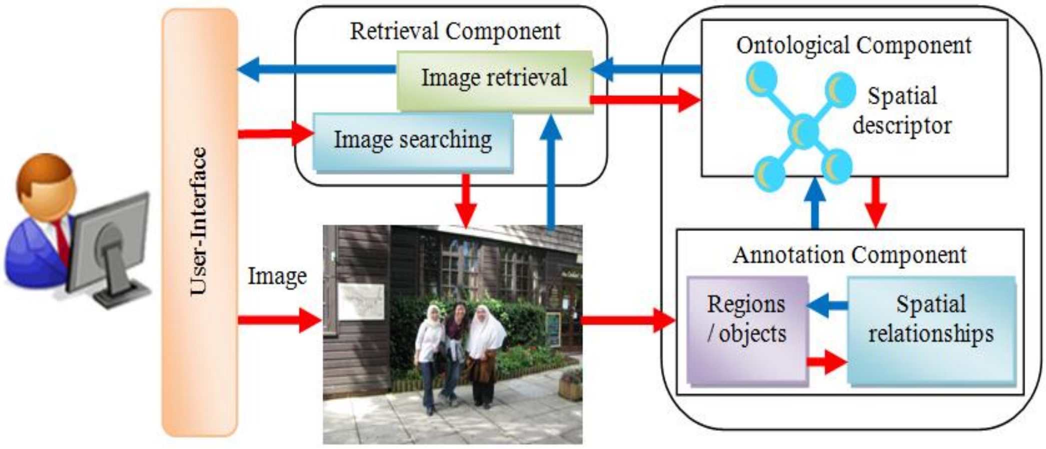


Fig.3. Research framework.

So far, the literature reviews and some preliminary experiment have been performed. However further practical works in the research and development phase is now being carried out. As a conclusion, it is hoped that this research will generate a constructive semantics approach in enabling the Semantic Web as well as bridging the Semantic Gap in image retrieval, while at the same time contributing new finding to human knowledge as a whole.

References

1. Abdullah, M. K., Mohd Suradi, N. R., Jamaluddin, N, Mokhtar, A.S., Abu Talib, A. R. & Zainuddin, M. F.: K-Chart: A Tool for Research Planning and Monitoring. J. of Quality Management And Analysis, vol 2(1), 123-130 (2006.)
2. Ahmad, I. & Grosky, W. I.: Indexing and Retrieval of Images by Spatial Constraints. J. of Visual Communication and Image Representation, vol. 14(3), Elsevier, 291-320, (2003)
3. Berners-Lee, T., Hendler, J. & Lassila, O.: The semantic web. Scientific American, (2001)
4. Berners-Lee, T., Fischetti, M. & Francisco, H.: Weaving the web: The original design and ultimate destiny of the World Wide Web by its Inventor (1999)

5. Bloehdorn, S. Petridis, K. Saathoff, C. Simou, N. Tzouvaras, V. Avrithis, Y. Handschuh, S. Kompatsiaris, I. Staab, S. & Strintzis, M.G.: Semantic Annotation of Images and Videos for Multimedia Analysis. In Proc. of the 2nd ESWC2005, (2005)
6. Chakravarthy, A., Ciravegna, F. & Lanfranchi, V.: Cross-media document annotation and enrichment. In: Proc of the 1st SAAW2006, (2006)
7. Duineveld, A. J., Stoter, R., Weiden, M. R., Kenepa, B. & Benjamins, V.R.: WonderTools? A comparative study of ontological engineering tools. In: Proc. of the 12th Workshop on Knowledge Acquisition, Modeling and Management. Alberta, Canada, October (1999)
8. Ko, B. & Byun, H.: Multiple Regions and Their Spatial Relationship-Based Image Retrieval. In: Proc.of the international CIVR2002. Lew, et al. (Eds). LNCS, Vol. 2383. Springer-Verlag, London, 81-90(2002)
9. Ko, B. C., Lee, H. S. & Byun, H.: Region-based Image Retrieval System Using Efficient Feature Description. In: Proc. of the 15th ICPR2000, vol. 4, 283-286. Spain, Sept., (2000)
10. Laaksonen, J., Koskela, M & Oja, E.: PicSOM-Self-organizing image retrieval with MPEG-7 content descriptions. IEEE Trans. on Neural Networks, 13(4), 841–853 (2002)
11. Lee, S.C., Hwang E.J & Lee, Y.K.: Using 3D Spatial Relationships for Image Retrieval by XML Annotation. ICCSA2004, LNCS 3046, 838–848 (2004)
12. Lee, S. & Hwang, E.: Spatial Similarity and Annotation-Based Image Retrieval System. IEEE 4th Inter. Sym. on Multimedia Software Engineering, Newport Beach, CA (2002)
13. Lewis, J.R.: IBM Computer Usability Satisfaction Questionnaires: Psychometric Evaluation and Instruction for Use. Inter. J. of HCI, 7(1), 57-78 (1995)
14. Li, J., Wang, J. Z., Wiederhold G.: IRM: Integrated Region Matching for Image Retrieval. ACM Multimedia, pp. 147-156, (2002)
15. Lux, M. Becker, J. & Krottmaier, H. Calph & Emir: Semantic Annotation and Retrieval in Personal Digital Photo Libraried. In: Proc. of 15th CAiSE'03. pp. 85-89, Austria (2003)
16. Saathoff, C., Petridis, K., Anastasopoulos, D., Timmermann, N., Kompatsiaris, I. & Staab, S.: M-OntoMat-Annotizer: Linking Ontologies with Multimedia Low-Level Features for Automatic Image Annotation. In: Posters of the 3rd ESWC 2006, Montenegro, (2006)
17. Srikanth, M., Varner, J., Bowden, M. & Moldovan, D.: Exploiting Ontologies for Automatic Image Annotation. In: Proc. of the 28th Annual Inter. ACM SIGIR Conference on Research and Development in Information Retrieval, Salvador, Brazil, 552-558 (2005)
18. Staab, S.: Multimedia Ontology. Summer School in Multimedia Semantics (SSMS2007), Glasgow, (2007)
19. Wang, Y. H.: Image Indexing and Similarity Retrieval Based on Spatial Relationship Model. Inf. Sci. Comput. Sci. 154, 1-2, Elsevier, New York, USA, pp. 39-58, Aug. (2003)
20. Ying D.: Ontology: The enabler for the Semantic Web, http://citeseer.ist.psu.edu/601004.html (2002)
21. Zhou X.M., Ang C. H. & Ling T. W.: Image Retrieval based on object's orientation spatial relationship. Pattern Recognition Letters 22. Elsevier Science, 469-477 (2001)

Appendix E Paper Published in ICSIPA2009

Enhanced Image Annotations Based on Spatial Information Extraction and Ontologies

Zurina Muda [#1], Paul H. Lewis[#2], Terry R. Payne [*3], Mark J. Weal [#4]

[#] *School of Electronic & Computer Science, University of Southampton*
United Kingdom
[1] zm06r@ecs.soton.ac.uk
[2] phl@ecs.soton.ac.uk
[4] mjw@ecs.soton.ac.uk
[*] *Department of Computer Science, University of Liverpool*
United Kingdom
[3] T.R.Payne@liverpool.ac.uk

Abstract—**Current research on image annotation often represents images in terms of labelled regions or objects, but pays little attention to the spatial positions or relationships between those regions or objects. To be effective, general purpose image retrieval systems require images with comprehensive annotations describing fully the content of the image. Much research is being done on automatic image annotation schemes but few authors address the issue of spatial annotations directly. This paper begins with a brief analysis of real picture queries to librarians showing how spatial terms are used to formulate queries. The paper is then concerned with the development of an enhanced automatic image annotation system, which extracts spatial information about objects in the image. The approach uses region boundaries and region labels to generate annotations describing absolute object positions and also relative positions between pairs of objects. A domain ontology and spatial information ontology are also used to extract more complex information about the relative closeness of objects to the viewer.**

I. Introduction

Rapid growth in the volume of multimedia information creates new challenges for information retrieval and sharing, and is stimulating activities on the development and application of Semantic Web technologies [1]. An important element in many multimedia applications is the extraction and use of visual information, and new approaches are needed to improve the extraction and inference of semantic relationships from low-level features in order to improve semantic retrieval and bridge the Semantic Gap [2].

A. Motivation

Combinations of traditional text-based and content-based approaches are still not sufficient for dealing with the problem of effective image retrieval on the Web, mainly because of the problem of poor textual annotations. Many Web images have irrelevant, little or even no surrounding or associated text. Sometimes the surrounding text does not describe the content of the image precisely or unhelpfully, does not describe the image at all. Automatic image annotation is an active area of research, but unfortunately, much initial research on image annotation has been concerned with assigning textual labels to images at the global level. Even when labels have been assigned locally to segmented regions or rectangular grid cells, little attention has been paid to the spatial relationships between regions or objects [3]. In this paper we are not only concerned with annotations which label objects individually but also annotations which indicate both relative and absolute spatial information about the objects. Current annotation systems may provide labels for an image such as *car, people, building* but fail to provide the information that the car is near and to the left of the building and the people are on the far right of the image. Although relatively basic, the use of spatial information in this way enriches the possibilities for semantic description of the images and enhances the power and precision of queries which can be handled in automated retrieval.

Manual image annotation is a tedious task and it is often difficult to provide accurate and comprehensive annotations for images. Ways to minimise the human input by making the annotation process semi-automatic or fully automatic are certainly desirable.

In this paper we present some novel automatic approaches to the extraction of spatial information to improve the annotation process and show briefly how this, coupled with the use of related ontologies, can lead to richer querying and retrieval facilities. Currently, much of the research on spatial relation extraction is pursued without integrating with an ontology. Using an ontology can ensure consistency in terminology and can help to disambiguate certain aspects of spatial vocabulary. It can act as a knowledge base about domain objects which can be used for increasing the spatial information that can be extracted. We envisage the ontology not only holding synonyms for spatial terminology but also, for example, order of magnitude height information for certain objects which allows reasoning about their relative closeness to the camera/viewing position. These developments not only make querying more flexible and powerful but can also lead to more accurate and precise query results [4].

B. Aim and Approach

Building on earlier work on automatic annotation and also on spatial information extraction, we are investigating more powerful approaches to annotating images automatically with spatial information by capturing the spatial relationships between labelled regions or objects in images and supporting the process with an enhanced ontology. By this means, human users and software agents alike will be able to search, retrieve and analyse visual information in more versatile ways.

The approach has three main stages:

- Segmentation and initial labelling: an automatic annotator such as the approach we have described earlier [5] or a semi-automatic labelling approach such as that provided by the LabelMe system [6], is used to provide region or object level annotations. The output from this stage consists of region boundary information and labels indicating the objects represented by the regions.
- Basic spatial information extraction: analysis of the regions and labels from the first stage is used to extract basic spatial information about the labelled objects. The information includes absolute spatial positions of objects and relative spatial positions for pairs of objects.
- Enhancements via the ontology: By reference to an appropriate ontology and reasoning where possible, additional spatial relations are inferred and diverse query vocabulary is accommodated.

This paper is concerned with the second and third stages where spatial information is extracted from the image regions and also additional information inferred using the ontology. The availability of labelled image regions from the first stage is assumed.

In the next section we discuss previous and related work on spatial information extraction from images and in section III we present a short analysis of the use of spatial descriptions in real queries submitted to picture librarians. In section IV the research framework and approach to spatial information extraction is developed. Section V shows results from a real example and section VI presents conclusions and future work.

II. RELATED WORKS

To date, much of the research into Content-Based Image Retrieval has focussed on non-textual representation of the spatial information. Some typical approaches include abstract or symbolic images that were used in [7]-[9] based on work initially done by Tanimoto in 1976 [10]. Ahmad & Grosky [11] proposed a symbolic image representation and indexing scheme to support retrieval of domain independent, spatially similar images, whereas Tian, et al. [12] used spatial layout combined with user defined region(s) of interest [13] to present the content of an image. Lee & Hwang [14] proposed a domain-independent spatial similarity and annotation-based image retrieval system that decomposed the image into multiple regions of interest containing objects and allowed the user to formulate a query based on both objects of an image and their spatial relationships. Ko & Byun [15] used the Hausdorff Distance to estimate spatial relationships between regions as part of their FRIP (Finding Region In the Pictures) [16] system and named this system as Integrated FRIP (IFRIP). Li, et al. [17] presented Integrated Region Matching based on spatial relationships between regions by allowing a similarity measure for regions based on image similarity comparison, while Smith & Chang [18] decomposed the image into regions and represented those regions as strings. Similarity retrieval by using 2D Strings requires massive geometric computation and focuses on those database images that consist of icons. Chang et al. [8] introduced the 2D string representation of an image to present spatial relationships between symbols, while Wang [19] proposed the 2D Be-string (two dimension begin-end boundary string) model based on [8] and [20] to represent an icon by its boundaries and evaluates image similarities based on the modified ‘‘longest common subsequence” algorithm [21].

All the research mentioned above was based on the content similarity of the images, where two or more images were compared based on the spatial similarity of iconic objects in the image and do not refer to the semantic knowledge of the image content directly.

More focused and relevant research on spatial relationships has been done by Hollink et al. [22], Lee et al. [23] and Yuan et al. [24]. In particular Hollink et al. [22] extracted eight spatial relations (right, left, above, below, near, far, contains, next) and nine absolute positions essentially on a 3x3 grid (labelled centre, north, south, east, west, north-east, north-west, south-east and south-west). Lee et al. [23] presented unified representations of spatial objects for both topological and directional relationships and considered 8 directional and 4 topological relations, and Yuan et al. [24] considered neighbouring relationships (on, above, below, left, right).

Based on the previous research in spatial information extraction, this research includes absolute and relative information, building particularly on the work of Hollink et al. [22] but extending it both in the granularity of the absolute positions, the extraction of combined relations (like above and to the left of) and through the use of object properties in the ontology to infer more complex spatial relations.

III. A REAL CASE STUDY

In an earlier research project ‘Bridging the Semantic Gap in Visual Information Retrieval’ [25] with the University of Brighton, we gathered and analysed a large number of real queries submitted to picture librarians in a number of large national and international picture libraries.

At that time we were not concerned with spatial information but a re-analysis of the queries has revealed that a significant proportion involved spatial information. It demonstrated that spatial information is used in real queries.

Of the 96 queries we analysed, which were submitted to one library, 19 contained spatial terminology, i.e. about 20%. Fragmentary examples include the following: (spatial terms are in bold)

… coins **on** table….

… table **at left**…
… cloth dyers working **under** master…
… the moon **over** fields …
… pictures **in** colour ….
… bench **in** middle …
… benches **on left** …
… church **in** Paris …
… **in** any period …

These query fragments illustrate some conventional uses of spatial terminology but also underline a number of challenges for automated systems. First it was clear that queries articulated by humans are often at a semantically very high level. Also the spatial information in the query often relates to the spatial relations between objects in the 3-D space of the real world, rather than the 2-D plane of the image (eg 'over the field'). In many cases they may be equivalent ('next to' or 'above') but in some cases the mapping is less obvious ('on' for example).

The queries also reveal the potential ambiguity of some terms. In 'working under master', the term 'under' is used not as a spatial term but with respect to a hierarchy of roles and in the fragment 'in any period', the preposition 'in' is used to indicate a temporal rather than a spatial location.

However, our analysis demonstrates the value and use of spatial information in human query formation and strengthens our view that the ability to support spatial terminology in automated image annotation and retrieval would be beneficial.

The fact that spatial terminology may be used for purposes other than presenting spatial information supports our view that ontologies will be useful in helping to understand potentially ambiguous terminology during the process of searching and retrieval.

IV. The Research Framework

The research framework for the development of the annotation system is illustrated in Fig. 1. The framework consists of three main components, which include:

- The Annotation Component
  This component automatically extracts and identifies spatial information. It delivers statements about the absolute spatial position for single objects and spatial relationship between pairs of objects.
- The Ontology Component
  This component contains a spatial relationships ontology and domain object information together with logic and reasoning facilities. The component uses ontological reasoning to identify the correct spatial terminology to be used in describing spatial relationships and attempts to resolve ambiguous meanings used in the query or description of the image content as mentioned in the real case study earlier.
- The Retrieval Component
  This component integrates with both the annotation and ontology components mentioned above to facilitate retrieval enhanced with spatial information.

Here we concentrate mainly on the annotation component and to a certain extent on the ontology component by focusing on the development and implementation of the spatial relationship algorithms and the spatial inferences using order of magnitude height information from the ontology.

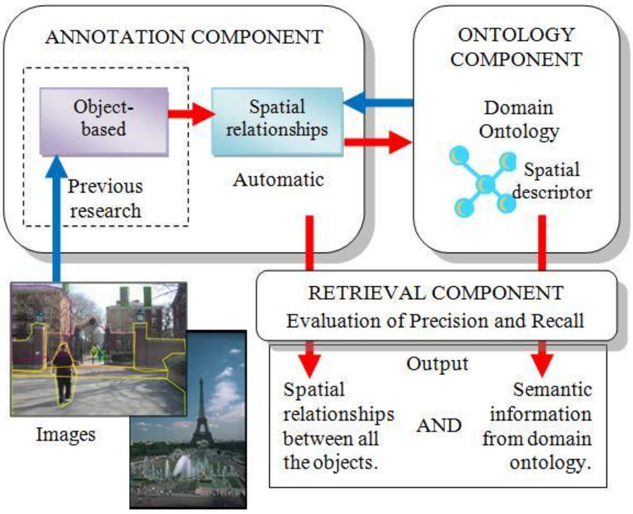


Fig. 1. The research framework

A. *Extracting Basic Spatial Information*

The annotation component in the framework assumes that a preliminary segmentation and region annotation stage has provided image regions, represented by the coordinates of pixels along their boundary, and region labels indicating the object represented by the region. This stage may be automatic as described in [5] or semi-automatic, for example by using the LabelMe software [6].

We refer to the labelled regions as objects and, extending the approach of Hollink et al. [22], automatically extract spatial descriptors for the absolute positions of individual objects and the relative spatial relations between all pairs of objects.

By considering all directions from each object in an image, spatial information between an object and the other objects can be computed. The computation of spatial relationships between objects in an image is described as follows:

Assume that a given image $_i$(I_i) consists of multiple labelled objects (O): $I_i = \{O_1, O_2 ... O_n\}$

Each of the objects has a set of coordinates that will be used to compute the spatial information between the object and the other objects in the image.

$\text{Object}_1 = \{(x_1,y_1), (x_2,y_2),…, (x_n,y_n)\}$
$\text{Object}_2 = \{(x_1,y_1), (x_2,y_2),…, (x_n,y_n)\}$
:
$\text{Object}_N = \{(x_1,y_1), (x_2,y_2),…, (x_n,y_n)\}$

The averages of the objects' x and y coordinates are calculated to give the centre of gravity (C) of each object in the image, represented as (x_c, y_c). All relations between objects are defined by computing and comparing the centres of gravity and borders of bounding boxes of two relative objects.

We use the centre of gravity to represent the "centroid" by contrast with the centre of the bounding box used by Hollink et al. [22], as in some cases it will be more meaningful, for example when dealing with a pyramid or in a more extreme case, a car with a long radio aerial.

The relative positions between pairs of objects are then computed based on these centroids and the bounding rectangles. The basic relations we extract are 'left of', 'right of', 'above' and 'below'. The height is used in the 'left of' and 'right of' concepts and the width is used in the 'above' and 'below' concepts to ensure that we only indicate an object is left or right of another if they are at approximately the same level in the image and similarly we only say an object is above or below another if they are in approximately the same left-right position. Left-right and above-below are of course reciprocal relations so if A is above B, B is below A etc. The rules for inferring 'left of' and 'right of' relations are defined as follows, and illustrated in Fig. 2.

- IF $((xc_1 < xc_2)$ AND $((h_1 + h_2) > |yc_1 - yc_2|))$ THEN $Object_1$ is on the LEFT of $Object_2$ [22] AND $Object_2$ is on the RIGHT of $Object_1$.
- IF $((xc_1 > xc_2)$ AND $((h_1 + h_2) > |yc_1 - yc_2|))$ THEN $Object_1$ is on the RIGHT of $Object_2$, AND $Object_2$ is on the LEFT of $Object_1$.

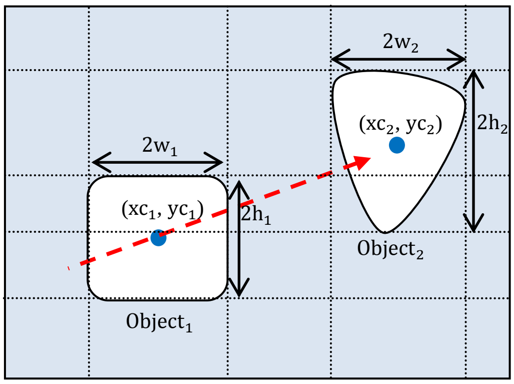


Fig. 2. Computation of 'object$_2$ is on the Right of object$_1$' relation

Similarly, the rules for inferring 'above' and 'below' relations are defined as follows:

- IF $((yc_1 > yc_2)$ AND $((w_1 + w_2) > |xc_1 - xc_2|))$ THEN $Object_1$ is ABOVE of $Object_2$, AND $Object_2$ is BELOW of $Object_1$.
- IF $((yc_1 < yc_2)$ AND $((w_1 + w_2) > |xc_1 - xc_2|))$ THEN $Object_1$ is BELOW of $Object_2$, AND $Object_2$ is ABOVE of $Object_1$.

By integrating these rules, we define rules for composite relations (eg 'above and to the right' etc) as follows and the example is illustrated in Fig. 3.

- IF $((xc_2 - xc_1) \geq (w_1 + w_2)$ AND $(yc_2 - yc_1) \geq (h_1 + h_2))$ THEN $Object_2$ is ABOVE and to the RIGHT of $Object_1$, AND $Object_1$ is BELOW and to the LEFT of $Object_2$.
- IF $((xc_2 - xc_1) \geq (w_1 + w_2)$ AND $(yc_1 - yc_2) \geq (h_1 + h_2))$ THEN $Object_2$ is BELOW and to the RIGHT of $Object_1$, AND $Object_1$ is ABOVE and to the LEFT of $Object_2$.
- IF $((xc_1 - xc_2) \geq (w_1 + w_2)$ AND $(yc_1 - yc_2) \geq (h_1 + h_2))$ THEN $Object_1$ is ABOVE and to the RIGHT of $Object_2$, AND $Object_2$ is BELOW and to the LEFT of $Object_1$.
- IF $((xc_1 - xc_2) \geq (w_1 + w_2)$ AND $(yc_2 - yc_1) \geq (h_1 + h_2))$ THEN $Object_1$ is BELOW and to the RIGHT of $Object_2$, AND $Object_2$ is ABOVE and to the LEFT of $Object_1$.

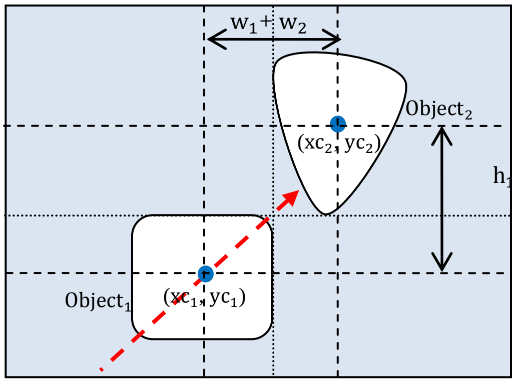


Fig. 3. Computation of 'object$_2$ is Above and to the Right of object$_1$' relation

In addition to the spatial relationships between objects in the image, we also extract the absolute positions of the objects in the image. For absolute position, we use a finer grained grid than [22] and use a different notation. Hollink et al. [22] used compass point positions defined on a 3x3 grid which is more suitable for geographical or topological representation. We divide the image into a 5x5 grid defining 25 absolute position annotations as shown in Fig. 4. This facilitates such absolute spatial annotations as 'at the far right at the top' or 'in the middle of the bottom'.

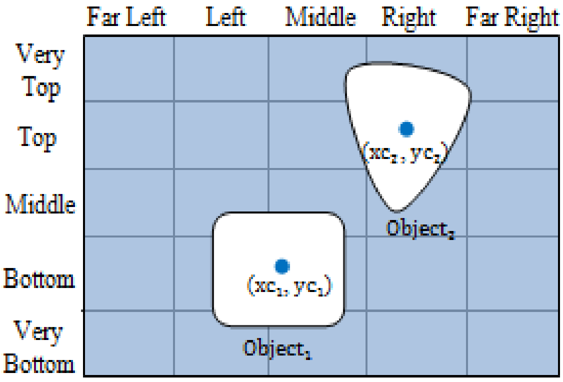


Fig. 4. Absolute position concepts

B. Using the Ontology

By recording order of magnitude height information with objects in the domain ontology we can infer additional spatial information using the heights of bounding rectangles. As an example, the order of magnitude heights of person and buildings are recorded as 2 metres and 10 metres respectively.

Then if the order of magnitude height for $object_n$ is M_n, as a simple heuristic we could infer that if $object_i$ is much nearer to the camera position (or the viewer) than $object_j$, then $2h_i/2h_j$ will be significantly greater than M_i/M_j.

We introduce a general heuristic:

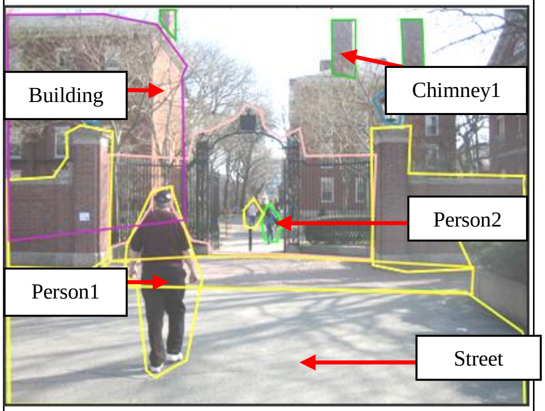
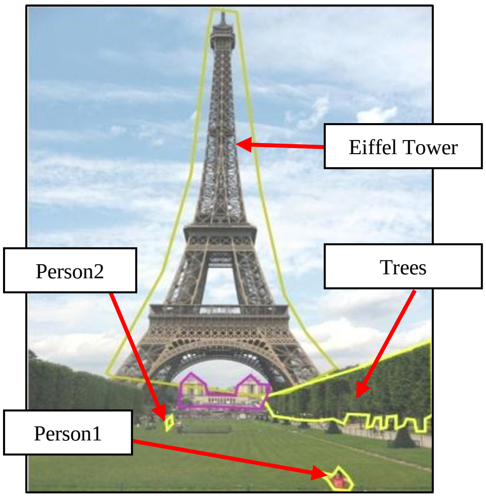
- IF $2h_i/2h_j > 3^* M_i/M_j$ THEN $object_i$ is nearer (the viewer) than $object_j$ AND $object_j$ is further away than $object_i$.The ontology has many other uses in the processing of spatial annotations, as hinted at earlier, but these will be the subject of a separate paper.

V. The Implementation And A Real Example

Each of the spatial information extraction rules described above has been implemented and can be applied to labelled image segmentations derived from the first stage of our framework. As an example, two images shown in Table I have been segmented and labelled using the semi-automatic LabelMe software [6]. To simplify our presentation, we only consider a subset of objects in those images. The coordinates of the boundary pixels of the labelled objects have been extracted and the output from the extraction and annotation process is a series of statements providing spatial information about the objects in each image.

Table I

Sample of Images and Results

Sample of Images	Spatial Annotation Statements
	Building on the LEFT of Chimney1, and Chimney1 on the RIGHT of Building. Building is ABOVE Street, and Street is BELOW Building. Building on the LEFT of Person1, and Person1 on the RIGHT of Building. Building is ABOVE Person1, and Person1 is BELOW Building. Building on the LEFT of Person2, and Person2 on the RIGHT of Building. Chimney1 is ABOVE Street, and Street is BELOW Chimney1. Chimney1 is ABOVE and to the RIGHT of Person1. Person1 is BELOW and to the LEFT of Chimney1. Chimney1 is ABOVE and to the RIGHT of Person2. Person2 is BELOW and to the LEFT of Chimney1. Street on the RIGHT of Person1, and Person1 on the LEFT of Street. Street is BELOW Person1, and Person1 is ABOVE Street. Street is BELOW Person2, and Person2 is ABOVE Street. Person1 on the LEFT of Person2, and Person2 on the RIGHT of Person1. Building is on the LEFT side and at the TOP of the image. Chimney1 is on the RIGHT side and at the VERY TOP of the image. Street is in the MIDDLE and at the VERY BOTTOM of the image. Person1 is on the LEFT side and at the BOTTOM of the image. Person2 is in the centre of the image. Person1 is NEARER than Building, and Building is FURTHER AWAY than Person1. Person1 is NEARER than Person2, and Person2 is FURTHER AWAY than Person1.
	Eiffel Tower on the LEFT of Trees, and Trees on the RIGHT of Eiffel Tower. Eiffel Tower is ABOVE Trees, and Trees is BELOW Eiffel Tower. Eiffel Tower is ABOVE Person1, and Person1 is BELOW Eiffel Tower. Eiffel Tower on the RIGHT of Person2, and Person2 on the LEFT of Eiffel Tower. Eiffel Tower is ABOVE Person2, and Person2 is BELOW Eiffel Tower. Eiffel Tower on the RIGHT of Building, and Building on the LEFT of Eiffel Tower. Eiffel Tower is ABOVE Building, and Building is BELOW Eiffel Tower. Trees are ABOVE Person1, and Person1 is BELOW Trees. Trees on the RIGHT of Person2, and Person2 on the LEFT of Trees. Trees on the RIGHT of Building, and Building on the LEFT of Trees. Person1 is BELOW and to the RIGHT of Person2. Person2 is ABOVE and to the LEFT of Person1. Person1 is BELOW and to the RIGHT of Building. Building is ABOVE and to the LEFT of Person1. Eiffel Tower is on the LEFT side and at the BOTTOM of the image. Person1 is NEARER than Eiffel Tower, and Eiffel Tower is FURTHER AWAY than Person1. Person2 is NEARER than Eiffel Tower, and Eiffel Tower is FURTHER AWAY than Person2. Eiffel Tower is NEARER than Building, and Building is FURTHER AWAY than Eiffel Tower. Person1 is NEARER than Person2, and Person2 is FURTHER AWAY than Person1. Person1 is NEARER than Building, and Building is FURTHER AWAY than Person1. Person2 is NEARER than Building, and Building is FURTHER AWAY than Person2.

The annotation statements extracted for the selected labelled objects in the images are shown in Table I. It can be seen that many useful annotations are generated including relative, absolute and 3-dimensional annotations.

These preliminary results show that the automatic annotator is working as expected, although some annotations illustrate areas where additional heuristics are required. However, the implementation is an on-going process and is being enhanced to improve the flexibility and reliability of the approach.

VI. Conclusion And Future Work

We have presented the design and implementation of enhanced approaches to spatial information extraction using labelled segmented images, extraction rules and ontology based object information. We have developed and implemented rules to automate relative and absolute spatial information extraction for objects in images. We also considered a general heuristic for relative order of magnitude height information to infer 3-dimensional annotations indicating relative closeness of objects to the viewer.

In total, we extract 35 spatial information concepts, including 8 spatial relationships concepts (left, right, above, below and the composites concepts). The system also extracts 25 fine-grained absolute spatial positions in the image and can infer 2 additional 3-dimensional annotation including ‘nearer than’ and ‘further away than’ relations by using relative order of magnitude height of objects from the ontology. The extraction of spatial information annotations has been demonstrated.

The spatial annotation extraction system will be enhanced and expanded further to include a wider vocabulary of spatial terms and to use other information on the domain objects via the ontology and knowledge base.

In the near future a retrieval front end will be implemented to enable image queries, which can include spatial information and which are made more flexible via the spatial terminology in the ontology. In conclusion, we have proposed a new method and approach for capturing spatial information from images in order to enhance an image annotation system for more high level semantic search and retrieval.

Acknowledgement

Special thanks to the Government of Malaysia and the Universiti Kebangsaan Malaysia for funding Zurina Muda with a research studentship for doctorate programme at the University of Southampton. Paul Lewis is grateful for support through the EU funded project Living Knowledge (Grant Agreement: 231126).

References

[1] T. Berners-Lee, J. Hendler, and O. Lassila, "The Semantic Web," *Scientific American,* pp. 34-43, 2001.

[2] J. S. Hare, P. H. Lewis, P. G. B. Enser, and C. J. Sandom, "Mind the Gap: Another look at the problem of the semantic gap in image retrieval," *in Multimedia Content Analysis, Management and Retrieval*, 2006, p. 17.

[3] Z. Muda, "Ontological Description of Image Content Using Regions Relationships," *in ESWC 2008 PhD Symposium*, 2008, p. 46.

[4] M. Srikanth, J. Varner, M. Bowden, and D. Moldovan, "Exploiting ontologies for automatic image annotation," *in Proc. of the 28th annual international ACM SIGIR conference on Research and development in information retrieval,* 2005, p. 552.

[5] J. Tang and P. H. Lewis, "An image based feature space and mapping for linking regions and words," *in VISAPP 2007: Proc. of the Second International Conference on Computer Vision Theory and Applications*, 2007, p. 29.

[6] B. C. Russell, A. Torralba, K. P. Murphy, and W. T. Freeman, "LabelMe: A Database and Web-Based Tool for Image Annotation," *Int. J. Comput. Vision*, vol. 77, pp. 157-173, 2008.

[7] S. K. Chang, E. Jungert, and Y. Li, "Representation and retrieval of symbolic pictures using generalized 2D string," *in SPIE Proc. on Visual Communications and Image Processing,* 1989, p. 1360.

[8] S. K. Chang, Q. Y. Shi, and C. W. Yan, "Iconic indexing by 2-D strings," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 9, pp. 413-428, 1987.

[9] S. K. Chang, Q. Y. Shi, and C. W. Yan, "Iconic indexing by 2D strings," *IEEE Computer Society Workshop on Visual Languages,* 1986.

[10] X. M. Zhou, C. H. Ang, and T. W. Ling, "Image retrieval based on object's orientation spatial relationship," *Pattern Recogn. Lett.,* vol. 22, pp. 469-477, 2001.

[11] I. Ahmad and W. I. Grosky, "Indexing and retrieval of images by spatial constraints," *Journal of Visual Communication and Image Representation,* vol. 14, pp. 291-320, 2003.

[12] Q. Tian, Y. Wu, and T. S. Huang, "Combine User Defined Region-of-Interest and Spatial Layout for Image Retrieval," in *Proc. IEEE 2000 International Conference on Image Processing (ICIP’2000),* 2000, p. 746-749.

[13] B. Moghaddam, H. Biermann, and D. Margaritis, "Regions-of-Interest and Spatial Layout for Content-Based Image Retrieval," *Multimedia Tools Appl.,* vol. 14, pp. 201-210, 2001.

[14] S. Lee and E. Hwang, "Spatial Similarity and Annotation-based Image Retrieval System," *in Proc. of the Fourth IEEE International Symposium on Multimedia Software Engineering,* 2002, p. 33.

[15] B. Ko and H. Byun, Multiple Regions and Their Spatial Relationship-Based Image Retrieval, ser. Lecture Notes in Computer Science Berlin, Germany: Springer-Verlag, 2002, vol.2383.

[16] B. Ko, H.-S. Lee, and H. Byun, "Region-Based Image Retrieval System Using Efficient Feature Description," *in Proc. of the International Conference on Pattern Recognition - Volume 4*, 2000, p.4283.

[17] J. Li, J. Z. Wang, and G. Wiederhold, "IRM: integrated region matching for image retrieval," *in Proc. of the eighth ACM international conference on Multimedia*, 2000, p. 147.

[18] J. R. Smith and S.-F. Chang, "Integrated spatial and feature image query," *Multimedia Syst.,* vol. 7, pp. 129-140, 1999.

[19] Y.-H. Wang, "Image indexing and similarity retrieval based on spatial relationship model," *Inf. Sci. Inf. Comput. Sci.,* vol. 154, pp. 39-58, 2003.

[20] A. J. T. Lee and H.-P. Chiu, "2D Z-string: a new spatial knowledge representation for image databases," *Pattern Recogn. Lett.,* vol. 24, pp. 3015-3026, 2003.

[21] X.-J. Wang, W.-Y. Ma, G.-R. Xue, and X. Li, "Multi-model similarity propagation and its application for web image retrieval," *in Proc. of the 12th annual ACM international conference on Multimedia,* 2004, p. 944.

[22] L. Hollink, G. Nguyen, G. Schreiber, J. Wielemaker, and B. Wielinga, "Adding spatial semantics to image annotations," *in 4th International Workshop on Knowledge Markup and Semantic Annotation at ISWC’04,* 2004, p. 31.

[23] S.-C. Lee, E. Hwang, and J.-G. Han, "Efficient Image Retrieval Based on Minimal Spatial Relationships," *Journal of Information Science and Engineering,* vol. 22(2), pp. 461-473, March 2006.

[24] J. Yuan, J. Li, and B. Zhang, "Exploiting spatial context constraints for automatic image region annotation," *in Proc. of the 15th international conference on Multimedia: ACM,* 2007, p. 595.

[25] P. G. B. Enser, C. J. Sandom, J. S. Hare, and P. H. Lewis, "Facing the reality of semantic image retrieval " *Journal of Documentation*, vol. 63(4), pp. 465-481, 2007.