

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.

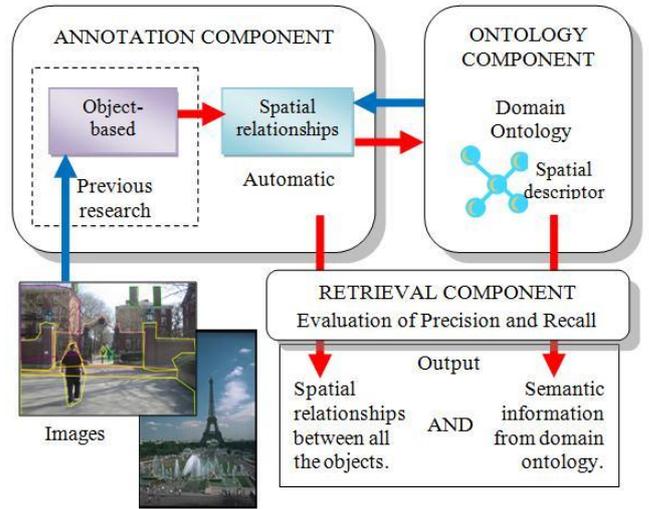


Fig. 1. The research framework

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): $I_i = \{O_1, O_2 \dots O_n\}$

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.

$$\text{Object}_1 = \{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}$$

$$\text{Object}_2 = \{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}$$

⋮

$$\text{Object}_N = \{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}$$

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 $((xc_1 < xc_2) \text{ AND } ((h_1 + h_2) > |yc_1 - yc_2|))$ THEN Object₁ is on the LEFT of Object₂ [22], AND Object₂ is on the RIGHT of Object₁.
- IF $((xc_1 > xc_2) \text{ AND } ((h_1 + h_2) > |yc_1 - yc_2|))$ THEN Object₁ is on the RIGHT of Object₂, AND Object₂ is on the LEFT of Object₁.

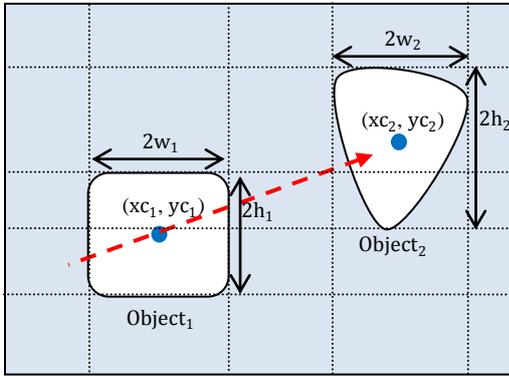


Fig. 2. Computation of ‘object₂ is on the Right of object₁’ relation

Similarly, the rules for inferring ‘above’ and ‘below’ relations are defined as follows:

- IF $((yc_1 > yc_2) \text{ AND } ((w_1 + w_2) > |xc_1 - xc_2|))$ THEN Object₁ is ABOVE of Object₂, AND Object₂ is BELOW of Object₁.
- IF $((yc_1 < yc_2) \text{ AND } ((w_1 + w_2) > |xc_1 - xc_2|))$ THEN Object₁ is BELOW of Object₂, AND Object₂ is ABOVE of Object₁.

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 $((xc_2 - xc_1) \geq (w_1 + w_2) \text{ AND } (yc_2 - yc_1) \geq (h_1 + h_2))$ THEN Object₂ is ABOVE and to the RIGHT of Object₁, AND Object₁ is BELOW and to the LEFT of Object₂.
- IF $((xc_2 - xc_1) \geq (w_1 + w_2) \text{ AND } (yc_1 - yc_2) \geq (h_1 + h_2))$ THEN Object₂ is BELOW and to the RIGHT of Object₁, AND Object₁ is ABOVE and to the LEFT of Object₂.

- IF $((xc_1 - xc_2) \geq (w_1 + w_2) \text{ AND } (yc_1 - yc_2) \geq (h_1 + h_2))$ THEN Object₁ is ABOVE and to the RIGHT of Object₂, AND Object₂ is BELOW and to the LEFT of Object₁.
- IF $((xc_1 - xc_2) \geq (w_1 + w_2) \text{ AND } (yc_2 - yc_1) \geq (h_1 + h_2))$ THEN Object₁ is BELOW and to the RIGHT of Object₂, AND Object₂ is ABOVE and to the LEFT of Object₁.

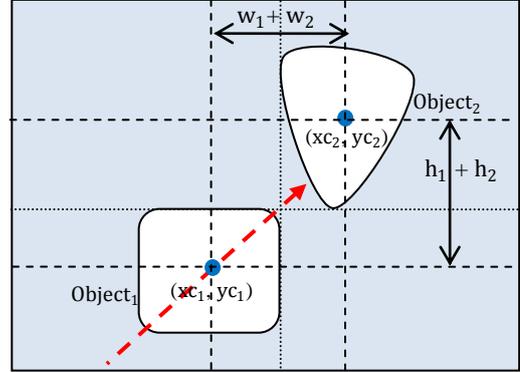


Fig. 3. Computation of ‘object₂ is Above and to the Right of object₁’ relation

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’.

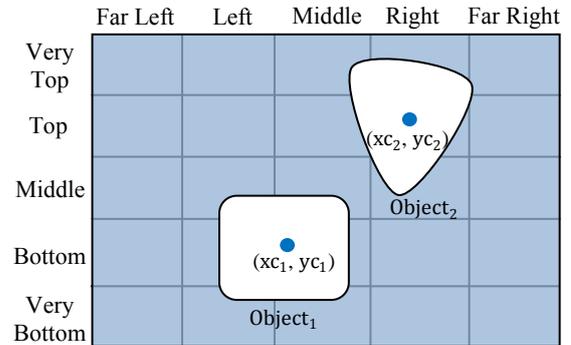


Fig. 4. Absolute position concepts

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:

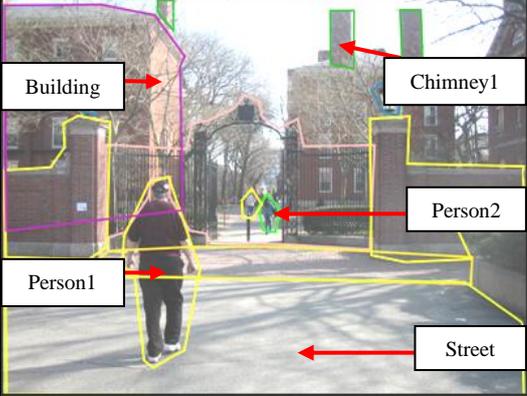
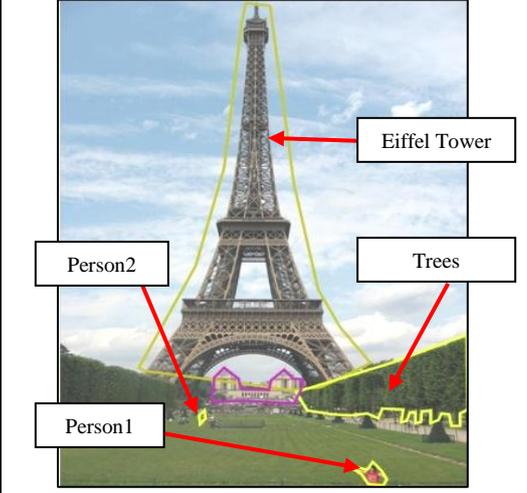
- IF $2h_i/2h_j > 3 * M_i/M_j$ THEN object_i is nearer (the viewer) than object_j AND object_j is further away than object_i.

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
	<p>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.</p>
	<p>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.</p>

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.