Bridging the Semantic Gap in Multimedia Information Retrieval

Top-down and Bottom-up Approaches

Jonathon S. Hare[1], Patrick A.S. Sinclair[1], Paul H. Lewis[1], Kirk Martinez[1], Peter G. B. Enser[2], and Christine J. Sandom[2]

[1] Intelligence, Agents, Multimedia Group,
School of Electronics and Computer Science,
University of Southampton,
Southampton, SO17 1BJ, UK
{jsh2, pass, phl, km}@ecs.soton.ac.uk
[2] School of Computing, Mathematical and Information Sciences,
University of Brighton,
Brighton, BN2 4GJ, UK
{p.g.b.enser, c.sandom}@bton.ac.uk

Abstract. Semantic representation of multimedia information is vital for enabling the kind of multimedia search capabilities that professional searchers require. Manual annotation is often not possible because of the shear scale of the multimedia information that needs indexing. This paper explores the ways in which we are using both top-down, ontologically driven approaches and bottom-up, automatic-annotation approaches to provide retrieval facilities to users. We also discuss many of the current techniques that we are investigating to combine these top-down and bottom-up approaches.

1 Introduction

The hallmark of a good retrieval system is its ability to respond to a user's queries and present results in a desired fashion. In the past there has been a tendency for research to focus on content-based retrieval techniques, ignoring the issues of users. In spite of this, some investigators have attempted to characterise image queries, providing insights in retrieval system design [1, 2, 3, 4] and highlighting the problem of what has become known as the *semantic gap*.

In the survey of content-based image retrieval by Smeulders et al. [5], the semantic gap is described as;

> ...the lack of coincidence between the information that one can extract from the visual data and the interpretation that the same data have for a user in a given situation.

At the end of the survey the authors conclude that:

> A critical point in the advancement of content-based retrieval is the semantic gap, where the meaning of an image is rarely self-evident. The aim of content-based retrieval systems must be to provide maximum support in bridging the semantic gap between the simplicity of available visual features and the richness of the user semantics.

Techniques for attempting to bridge the semantic gap in multimedia retrieval have mostly used an *auto-annotation* approach, in which keyword annotations are applied to unlabelled images. Enser et al. [6] discuss some short-comings of auto-annotation due to their lack of *richness* when compared to real image annotations in archival collections. Enser et al. [6] goes on to suggest that perhaps a way forward is to combine shareable ontologies to make explicit the relationships between the keyword labels and concepts they represent (e.g [7, 8, 9]). Zhao and Grosky [10] proposed an approach to bridging the semantic gap using Latent Semantic Indexing (see also [11, 12]).

This paper describes our experiences in building multimedia retrieval systems, and how through various techniques we are attempting to bridge the semantic gap to improve retrieval quality for end users. Our approaches to attacking this gap have been twofold; In the first section of the paper we describe some attempts to automatically learn and apply the semantics of multimedia objects from the bottom-up. In the second section, we describe our work in trying to bridge the gap from the top down, by using structured knowledge representations in the form of ontologies.

The third section of the paper discusses our current work and ideas for combining these bottom-up and top-down approaches to further our goal of improved retrieval effectiveness.

2 Bottom-up approaches

Current bottom-up approaches to generating semantics for multimedia entities generally all fall into the same information pipeline. In general, this information pipeline consists of a number of processing stages, between the raw media and the semantics, as illustrated in Figure 1.

Currently, work on the automatic annotation of media has mostly concentrated on the processing stages between the raw media and the labelled scene. Of course, not all techniques follow the information pipeline shown in Figure 1 exactly. For example, a number of auto-annotation techniques directly associate descriptors with labels, without any concept of objects. The next subsection gives an overview of existing auto-annotation techniques, and the following subsection describes an alternative technique which we have been developing that avoids some of the problems associated with the various auto-annotation approaches.

2.1 Automatic Annotation

The first attempt at automatic annotation was perhaps the work of Mori et al. [13], which applied a co-occurrence model to keywords and low-level features

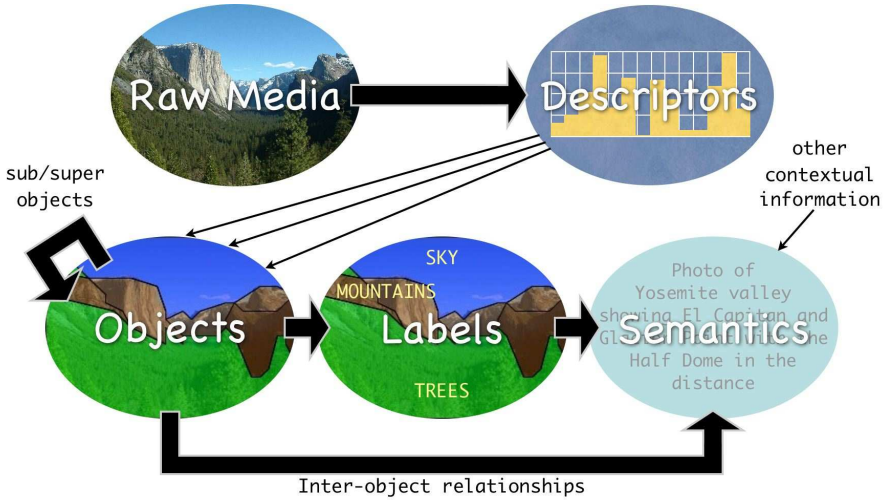

Fig. 1. A generalised information pipeline from raw media to semantics.

of rectangular image regions. The current techniques for auto-annotation generally fall into two categories; those that first segment images into regions, or 'blobs' and those that take a more scene-orientated approach, using global information. The segmentation approach has recently been pursued by a number of researchers. Duygulu et al. [14] proposed a method by which a machine translation model was applied to translate between keyword annotations and a discrete vocabulary of clustered 'blobs'. The data-set proposed by Duygulu et al. [14] has become a popular benchmark of annotation systems in the literature. Jeon et al. [15] improved on the results of Duygulu et al. [14] by recasting the problem as cross-lingual information retrieval and applying the Cross-Media Relevance Model (CMRM) to the annotation task. Jeon et al. [15] also showed that better (ranked) retrieval results could be obtained by using probabilistic annotation, rather than *hard* annotation. Lavrenko et al. [16] used the Continuous-space Relevance Model (CRM) to build continuous probability density functions to describe the process of generating blob features. The CRM model was shown to outperform the CMRM model significantly. Metzler and Manmatha [17] propose an inference network approach to link regions and their annotations; unseen images can be annotated by propagating belief through the network to the nodes representing keywords. The models by Monay and Gatica-Perez [18], Feng et al. [19] and Jeon et al. [20] use rectangular regions rather than blobs. Monay and Gatica-Perez [18] investigate Latent Space models of annotation using Latent Semantic Analysis and Probabilistic Latent Semantic Analysis, Feng et al. [19] use a multiple Bernoulli distribution to model the relationship between the blocks and keywords, whilst Jeon et al. [20] use a machine translation approach based on Maximum Entropy. Blei and Jordan [21] describe an extension to Latent Dirichlet Allocation [22] which assumes a mixture of latent factors

is used to generate keywords and blob features. This approach is extended to multi-modal data in the article by Barnard et al. [23].

Oliva and Torralba [24, 25] explored a scene oriented approach to annotation in which they showed that basic scene annotations, such as 'buildings' and 'street' could be applied using relevant low-level global filters. Yavlinsky et al. [26] explored the possibility of using simple global features together with robust non-parametric density estimation using the technique of kernel smoothing. The results shown by Yavlinsky et al. [26] were comparable with the inference network [17] and CRM [16]. Notably, Yavlinsky et al. showed that the Corel data-set proposed by Duygulu et al. [14] could be annotated remarkably well by just using global colour information.

Most of the auto-annotation approaches described above perform annotations in a *hard* manner; that is, they explicitly apply some number of annotations to an image. A *hard* auto-annotator can cause problems in retrieval because it may inadvertently annotate with a similar, but wrong label; for example, labelling an image of a horse with "foal". Jeon *et al* [15] first noted that this was the case when they compared the retrieval results from a fixed-length hard annotator with a probabilistic annotator. Duygulu *et al* [14] attempt to get around this problem by creating clusters of keywords with similar meaning.

2.2 Semantic spaces

We are currently investigating a different approach to auto-annotation [27, 28]; Instead of applying *hard* annotations, we have developed an approach in which annotation is performed implicitly in a *soft* manner. The premise behind our approach is simple; a semantic-space of documents (images) and terms (keywords) is created using a linear algebraic technique. Similar documents and/or terms within this semantic-space share similar positions within the space. For example, given sufficient training data, this allows a search for "horse" to return images of both horses and foals because the terms "horse" and "foal" share similar locations within the semantic space.

Building a semantic-space: Using linear algebra to associate images and terms Latent Semantic Indexing is a technique originally developed for textual information retrieval. Berry *et al* [29] described how Latent Semantic Indexing can be used for cross-language retrieval because it ignores both syntax and explicit semantics in the documents being indexed. In particular, Berry *et al* cite the work of Landauer and Littman [30] who demonstrate a system based on LSI for performing text searching on a set of French and English documents where the queries could be in either French or English (or conceivably both), and the system would return documents in both languages which corresponded to the query. The work of Landauer and Littman negates the need for explicit translations of all the English documents into French; instead, the system was trained on a set of English documents and versions of the documents translated into French, and through a process called 'folding-in, the remaining English

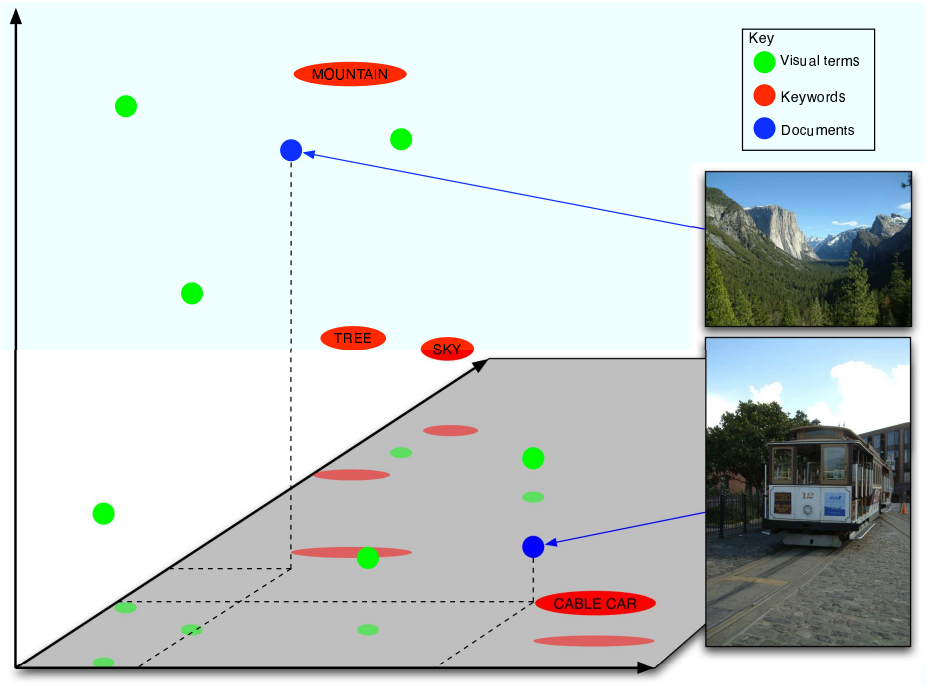

Fig. 2. An illustration of a semantic space

documents were indexed without the need for explicit translations. This idea has become known as *Cross-Language Latent Semantic Indexing* (CL-LSI).

Monay and Gatica-Perez[18] attempted to use straight LSI (without 'folding-in') with simple cross-domain vectors for auto-annotation. They first created a training matrix of cross-domain vectors and applied LSI. By querying the left-hand subspace they were able to rank an un-annotated query document against each annotation term in order to assess likely annotations to apply to the image. Our approach, described below, is different because we do not explicitly annotate images, but rather just place them in a semantic-space which can be queried by keyword.

Our idea is based on a generalisation of CL-LSI. In general any document (be it text, image, or even video) can be described by a series of observations made about its content. We refer to each of these observations as terms. In order to create a semantic-space for searching images, we first create a 'training' matrix of terms and documents that describe observations about a set of annotated training images; these observations consist of low-level descriptors and observations of which keywords occur in each of the images. This training term-document matrix then has LSI applied to it. The final stage in building the semantic-space is to 'fold-in' the corpus of un-annotated images, using purely the visual observations. The result of this process is two matrices; one representing the coordinates of the terms in the semantic space, and the other representing the coordinates of documents in the space. Similarity of terms and documents can be assessed by calculating the angle between the respective coordinate vectors.

Experiments with our semantic-space approach [27, 28] to multimedia retrieval have been rather promising, especially considering the simplicity of the technique. The semantic space has been shown to outperform Duygulu et al.'s [14] machine translation technique, with colour histogram image features alone.

3 Top-down approaches to finding media

There is an increasing interest and research on the use of ontologies and semantic web tools with multimedia collections. Early work on semantically describing images using ontologies as a tool for annotating and searching images more intelligently was described by Schreiber et al [31]. Several authors have described efforts to move the MPEG-7 description of multimedia information closer to ontology languages such as RDF and OWL [32, 33].

The aim of using ontologies to describe multimedia resources is to provide well-structured information to improve the accuracy of retrieval. Semantic web technologies also facilitate the integration of heterogeneous information sources and formats. Well-structured information is crucial for providing advanced browsing and visualisation facilities, as opposed to more traditional query-based systems. This is demonstrated in the development of semantic web based interface frameworks, such as mSpace [34].

There are several approaches to semantically annotating multimedia. The aceMedia project [35] is developing a knowledge infrastructure for multimedia

analysis, which incorporates a visual description ontology and a multimedia structure ontology. They have also developed the M-OntoMat-Annotizer tool that allows users to manually annotate multimedia items with semantic information.

Recently there has been strong interest in photo annotation, in particular dealing with semantically annotating and sharing personal photographic collections. Some projects are investigating the combination of the context in which a photograph was captured with information from other readily available sources in order to generate outline annotations. For example, by capturing the time and location when a photo is taken, it can be correlated with local weather reports and even sunrise and sunset times. Other approaches are going further by examining the context around the user, using sources such as calendars and the users social network.

The MIAKT project [36] allowed access to patient information, including multimedia data, in such a way that it supports knowledge management across the various modalities that exist within the breast cancer screening programme. The MIAKT system enabled the annotation of images with ontologically controlled terms, sometimes derived automatically from content-based image descriptors extracted by image-processing routines from regions of interest delineated by medical experts. The aim of the project was to demonstrate enhanced support at the semantic level for decision making which needs to draw on low level features and their descriptions as well as the related case notes. It also provides a platform for reasoning about new cases on the basis of the semantically integrated set of (multimedia) case histories.

3.1 Case Study: Cultural heritage multimedia collections

In the Sculpteur and eCHASE projects, we have been working with a variety of European cultural heritage institutions, including museums, galleries, picture libraries and television broadcasters. These institutions systematically create detailed and rich documentation of their collections, with multimedia playing a key role. Museums and galleries use multimedia representations of items in their collections, such as paintings, sculptures and cultural artefacts. On the other hand, picture libraries and broadcasters maintain multimedia collections covering a wide range of subjects, from historical events and figures to celebrity portraits and material supporting news stories.

These multimedia collections are supported by rich documentation, i.e. metadata covering the context around the multimedia items. This includes details on the people, events and objects that are depicted and/or related to the object. However, there challenges to making this material accessible to users. Each institution uses its own metadata formats and the information is locked away in legacy content management systems. Often this information is in an unstructured form, such as documents, reports, articles and so on. When the data is structured, for example in a relational database, a large proportion of the rich information is handled in free text fields, such as captions or descriptions.

We have been investigating ontological approaches to exploit this information to provide searching and browsing across the different multimedia collections in a uniform way. As a common model we use the CIDOC Conceptual Reference Model (CIDOC CRM), a core ontology for describing the semantics of schema and data structure elements used in museum object documentation. We have mapped each institutions metadata schema to the CIDOC CRM and exposed it (see [7]) using a Z39.50 Search and Retrieve Web Service (SRW) [37]. The SRW allows queries and results to be expressed in terms of the CIDOC CRM. It also allows us to transform the query results into RDF, providing a mechanism for harvesting semantically described metadata from relational databases.

In our experience, it is not sufficient to perform the mapping of the metadata schema: the data instances used in different museum and gallery legacy systems also need to be rationalised and harmonised. This is partly a data cleaning issue to do with misspellings, syntactic differences and poorly structured data. However, part of the problem is also the need for a consensus of agreement on common semantics in the cultural heritage domain for people, places, events, terminology and so on. In the eCHASE project we are developing services and tools for processing and cleaning the metadata provided by our content partners. These services can be orchestrated and enacted through the Taverna workbench [38], a web services-based workflow system.

Several metadata cleaning and processing services have been developed. These include a service to transform different date formats, including free text descriptions, into a consistent date format so that precise time-based queries can be performed across the different collections. Place descriptions are mapped to a common gazetteer, allowing places to be related properly to their larger regions or countries. Different approaches to the alignment of the different keywording and classification schemes used by the different institutions are being studied. Much of the rich information is held in free text descriptions such as image captions, so the use of information extraction tools such as GATE [39] is being investigated.

3.2 Case Study: Poodle Images

Within our work in the "Bridging the Semantic Gap in Image Retrieval" project, we have been investigating how the use of test-bed ontologies can meet the needs of real image searchers in limited domains [28, 40, 6]. As part of this investigation, we have collaborated with picture librarians at the Kennel Club picture library and collected sample queries, images and meta-data. In order to investigate the potential of ontology-driven search, we created a thesaurus of the image meta-data, and modelled this using the SKOS [41] ontology. A search engine, with a 'Google' style interface, that allows searchers to search either by keyword or by concept was created in order to allow us to experiment with ontology driven search. When performing a concept-based search, the search engine automatically performs inferencing to find all narrower concepts of the query concept.

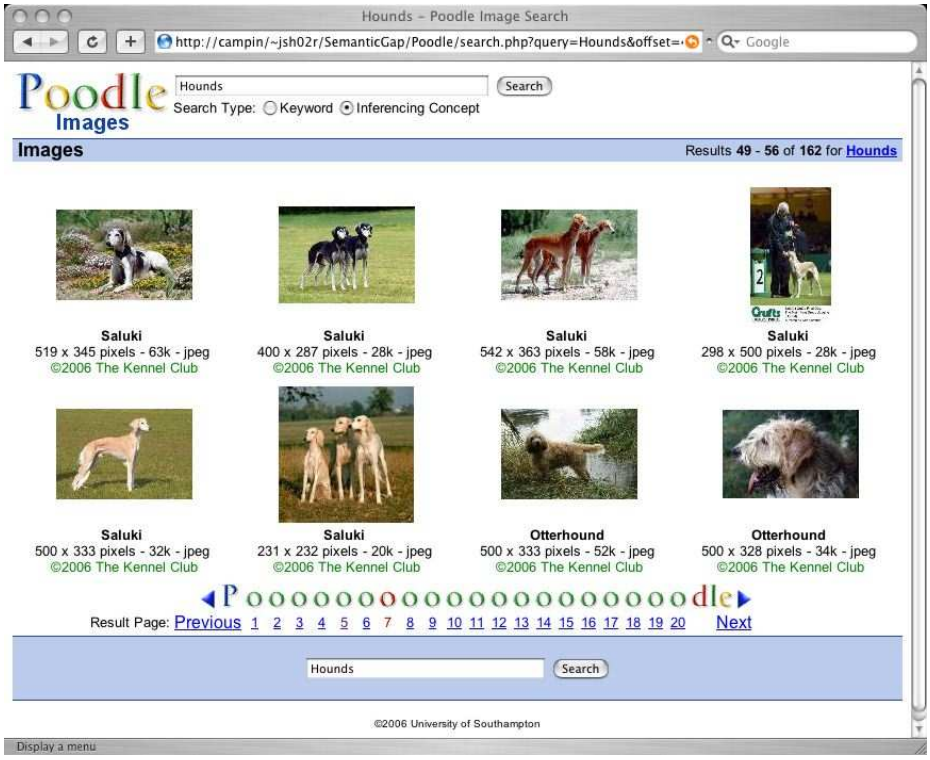

Fig. 3. The 'Poodle' image search engine.

This inferencing expands the usefulness of the search engine; For example, consider searching for images with the term 'hound'. When performing a keyword search, the search engine will find all of the images labelled with 'hound' or any term containing 'hound', such as 'bloodhound', 'greyhound' or 'basset hound'. In total, from our sample collection of 2704 dog-related images, the keyword search for 'hound' returns 95 images. When performing a concept-based search for the concept representing a 'hound', the results returned are somewhat different, with the search engine returning a total of 162 images. 95 of these images are the same as those found with the keyword search, but the remainder are of images depicting the concept representing a 'hound' that were not explicitly annotated as being a type of 'hound' in the keywords. For example, images of Beagles, Basenji's and Podengo's appear in the results set. These breeds of dog are all part of the hound class.

4 Discussion

Thus far in the paper, we have discussed both bottom-up and top-down approaches to finding multimedia documents. In this section we reiterate some of

the associated issues, and discuss some possible solutions through the integration and convergence of the two approaches.

As previously mentioned, *hard* auto-annotation techniques can cause problems due to mislabelling when used for search. By mediating the auto-annotation with a top-down ontological approach, it may be possible to improve retrieval performance in the presence of imperfect annotations. There are many possible ways of achieving this. For example, the simplest method would to be to put the ontology as an extra level in between the search query and the keywords, as in the "Poodle Images" example in Section 3.2.

At the other extreme, another approach would be to actually train the auto-annotator to annotate the images with concepts rather than keywords. This could lead to intriguing possibilities; For example, the annotator could make use of the ontology to maintain consistency of annotations (i.e. it could help avoid unlikely combinations of annotations, such as "Elephant" and "Ocean"). Annotations could also be made *safer* — for example, the auto-annotator could be biased so that it annotates with more generic concepts, unless the likelihood of a specific concept exceeded a certain threshold. This would avoid the "horse"/"foal" mislabelling described earlier, as the images would be labelled with "horse" unless the image features suggest that the image depicts a foal with a very high confidence. This idea is illustrated in Figure 4.

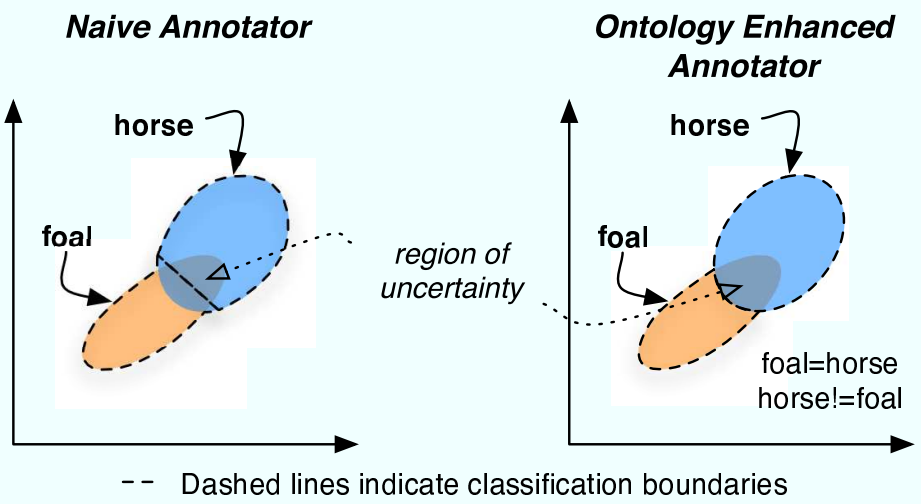

Fig. 4. Improving annotation using ontology provided knowledge of relationships between concepts.

The Semantic Space approach to finding unannotated images does not suffer the drawbacks of the *hard* auto-annotators in terms of the mislabelling of images. The semantic space approach can be seen as an alternative to ontology-driven

search. In fact, a semantic space is essentially a view of an ontology with all of the relationships removed — the semantic space only preserves the similarity relationships of concepts in terms of their spatial relation. Whilst the semantic space doesn't model semantic structure to the same degree as an ontology, it does have advantages in terms of its relative ease of creation compared to the labour costs involved in creating a full ontology.

Although the semantic space can be seen as an alternative to an ontology, this does not mean that the technologies can't be combined. In fact the use of an ontology in the training stage of building a semantic space could lead to a much better search performance. One of the problems of training a semantic space with image features and keywords is that the similarity of keywords is poorly modelled. This can lead to interesting effects, such as unrelated keywords being close to each other in the space due to some shared visual feature. For example, in our test implementation using global colour features, the keywords "sun", "flower" and "petals" all occur close to each other in the semantic space due to the colour yellow occurring in the images depicting each of the terms. One possible approach to avoid this problem would be to use an ontology to build pseudo-documents that model the dependent and independent keyword terms. These pseudo-documents could then be incorporated into the training stage, and the end result would be that the separation of the unrelated keywords within the space would be increased.

There are many issues with the keywording schemes currently employed by professional picture archives, in particular when applying auto annotation systems. Generally keywords are assigned to images by professional cataloguers, who must choose a set of terms that they feel will match the widest possible set of relevant queries. To overcome the limited indexing and searching mechanisms for the archives' systems, these terms range from the very general to the very specific and often overlap, for example to cover synonyms and related terms. For instance, when describing a "foal", the following terms might be applied: "foal", "horse", "vertebrate", "animal" and so on. In addition, there might be archive-specific keywording conventions, such as describing the image format, or whether an image is colour or black and white.

As described earlier the use of an ontology for annotations, where images are tagged with concepts rather than simple keywords, may overcome some of these issues. Using the most specific concept in the training stage would result in a more streamlined semantic space, and therefore more consistent annotations. Query expansion could then be used to retrieve images using broader query terms.

However, there are significant challenges in converting existing keyword-based image archives to concept-based annotations, such as dealing with word disambiguation. We are developing metadata cleaning and transforming tools in the eCHASE project, using resources such as Wordnet [42]. It will be interesting to compare the results of auto annotation on the raw keyword-based metadata and the concept-based metadata after it has been transformed.

The use of an ontology defines and structures different types of information about each image. Beyond keywording and other classifications schemes, this might include dates, places, photographer name, artist names for paintings and so on. Depending on the application or scenario, it may be sensible to train or build semantic spaces using different combinations of these features, or perhaps apply weightings to certain types of information. This would not be possible using a flat keywording scheme where each annotation is treated in the same way.

5 Conclusions

This paper has described both bottom-up and top-down approaches to annotating media in order to attempt to bridge the semantic gap for the purposes of multimedia information retrieval. The paper has outlined the ways in which we are currently using both approaches in building our current retrieval systems. However, we believe that best strategy for improving retrieval performance will ultimately be provided by combining these approaches.

The paper discussed a number of ideas and techniques which we are currently developing that combine both bottom-up and top-down knowledge. Our future work consists of building upon these ideas and testing them in real world scenarios with real users.

Previous experience with real multimedia searchers has shown that their queries are very rarely simple (e.g. 'find me an image depicting a sunset') and are usually much more involved. Part of our future work, especially regarding the semantic space technique, will be to look at how these more complex queries can be answered.

Acknowledgements

The 'Bridging the semantic gap in information retrieval' project is funded by the Arts and Humanities Research Council (MRG-AN6770/APN17429), whose support is gratefully acknowledged. We are grateful to the Kennel Club, David Dalton, Farlap Photography, Marc Henrie and Keri Lummis for allowing the use of their images.

eCHASE is a €3.5M eContent project that started in January 2005 and runs for 2 years. eCHASE is co-funded by the European Commission, DG Information Society, under the contract EDC 11262, within the "European digital content for the global networks" programme. The eCHASE consortium includes Istituto Geografico De Agostini S.p.A., IT Innovation and IAM within Electronics and Computer Science at the University of Southampton, Fratelli Alinari, Giunti Interactive Lab S.r.l., Hewlett Packard, Österreichischer Rundfunk, System Simulation Ltd and Getty Images.

References

[1] Enser, P.G.B.: Pictorial information retrieval. Journal of Documentation **51** (1995) 126–170

[2] Armitage, L.H., Enser, P.G.B.: Analysis of user need in image archives. Journal of Information Sciences **23** (1997) 287–299

[3] Ornager, S.: Image retrieval: Theoretical and empirical user studies on accessing information in images. In: Proceedings of the 60th American Society of Information Retrieval Annual Meeting. Volume 34. (1997) 202–211

[4] Hollink, L., Schreiber, A.T., Wielinga, B.J., Worring, M.: Classification of user image descriptions. Int. J. Hum.-Comput. Stud. **61** (2004) 601–626

[5] Smeulders, A.W.M., Worring, M., Santini, S., Gupta, A., Jain, R.: Content-based image retrieval at the end of the early years. IEEE Trans. Pattern Anal. Mach. Intell. **22** (2000) 1349–1380

[6] Enser, P.G.B., Sandom, C.J., Lewis, P.H.: Automatic annotation of images from the practitioner perspective. In Leow, W.K., Lew, M.S., Chua, T.S., Ma, W.Y., Chaisorn, L., Bakker, E.M., eds.: CIVR. Volume 3568 of Lecture Notes in Computer Science., Singapore, Springer (2005) 497–506

[7] Addis, M., Boniface, M., Goodall, S., Grimwood, P., Kim, S., Lewis, P., Martinez, K., Steveson, A.: SCULPTEUR: Towards a New Paradigm for Multimedia Museum Information Handling. In: International Semantic Web Conference (ISWC 2003), Florida, USA (2003) 582–596

[8] Goodall, S., Lewis, P.H., Martinez, K., Sinclair, P., Addis, M., Lahanier, C., Stevenson, J.: Knowledge-based exploration of multimedia museum collections. In: Proceedings of European Workshop on the Integration of Knowledge, Semantics and Digital Media Technology (EWIMT), London, U.K. (2004)

[9] Hu, B., Dasmahapatra, S., Lewis, P., Shadbolt, N.: Ontology-based medical image annotation with description logics. In: Proceedings of The 15th IEEE International Conference on Tools with Artificial Intelligence, IEEE Computer Society Press (2003) 77–82

[10] Zhao, R., Grosky, W.I.: From features to semantics: Some preliminary results. In: IEEE International Conference on Multimedia and Expo (II). (2000) 679–682

[11] Grosky, W.I., Zhao, R.: Negotiating the semantic gap: From feature maps to semantic landscapes. Lecture Notes in Computer Science **2234** (2001) 33–??

[12] Cascia, M.L., Sethi, S., Sclaroff, S.: Combining textual and visual cues for content-based image retrieval on the world wide web. In: CBAIVL '98: Proceedings of the IEEE Workshop on Content - Based Access of Image and Video Libraries, Washington, DC, USA, IEEE Computer Society (1998) 24

[13] Mori, Y., Takahashi, H., Oka, R.: Image-to-word transformation based on dividing and vector quantizing images with words. In: Proceedings of the First International Workshop on Multimedia Intelligent Storage and Retrieval Management (MISRM'99). (1999)

[14] Duygulu, P., Barnard, K., de Freitas, J.F.G., Forsyth, D.A.: Object recognition as machine translation: Learning a lexicon for a fixed image vocabulary. In: ECCV '02: Proceedings of the 7th European Conference on Computer Vision-Part IV, London, UK, Springer-Verlag (2002) 97–112

[15] Jeon, J., Lavrenko, V., Manmatha, R.: Automatic image annotation and retrieval using cross-media relevance models. In: SIGIR '03: Proceedings of the 26th annual international ACM SIGIR conference on Research and development in informaion retrieval, New York, NY, USA, ACM Press (2003) 119–126

[16] Lavrenko, V., Manmatha, R., Jeon, J.: A model for learning the semantics of pictures. In Thrun, S., Saul, L., Schölkopf, B., eds.: Advances in Neural Information Processing Systems 16. MIT Press, Cambridge, MA (2004)
[17] Metzler, D., Manmatha, R.: An inference network approach to image retrieval. In Enser, P.G.B., Kompatsiaris, Y., O'Connor, N.E., Smeaton, A.F., Smeulders, A.W.M., eds.: CIVR. Volume 3115 of Lecture Notes in Computer Science., Springer (2004) 42–50
[18] Monay, F., Gatica-Perez, D.: On image auto-annotation with latent space models. In: MULTIMEDIA '03: Proceedings of the eleventh ACM international conference on Multimedia, ACM Press (2003) 275–278
[19] Feng, S.L., Manmatha, R., Lavrenko, V.: Multiple bernoulli relevance models for image and video annotation. In: CVPR (2). (2004) 1002–1009
[20] Jeon, J., Manmatha, R.: Using maximum entropy for automatic image annotation. In Enser, P.G.B., Kompatsiaris, Y., O'Connor, N.E., Smeaton, A.F., Smeulders, A.W.M., eds.: CIVR. Volume 3115 of Lecture Notes in Computer Science., Springer (2004) 24–32
[21] Blei, D.M., Jordan, M.I.: Modeling annotated data. In: SIGIR '03: Proceedings of the 26th annual international ACM SIGIR conference on Research and development in informaion retrieval, New York, NY, USA, ACM Press (2003) 127–134
[22] Blei, D.M., Ng, A.Y., Jordan, M.I.: Latent dirichlet allocation. J. Mach. Learn. Res. **3** (2003) 993–1022
[23] Barnard, K., Duygulu, P., Forsyth, D., de Freitas, N., Blei, D.M., Jordan, M.I.: Matching words and pictures. J. Mach. Learn. Res. **3** (2003) 1107–1135
[24] Oliva, A., Torralba, A.: Modeling the shape of the scene: A holistic representation of the spatial envelope. Int. J. Comput. Vision **42** (2001) 145–175
[25] Oliva, A., Torralba, A.B.: Scene-centered description from spatial envelope properties. In: BMCV '02: Proceedings of the Second International Workshop on Biologically Motivated Computer Vision, London, UK, Springer-Verlag (2002) 263–272
[26] Yavlinsky, A., Schofield, E., Rüger, S.: Automated Image Annotation Using Global Features and Robust Nonparametric Density Estimation. In Leow, W.K., Lew, M.S., Chua, T.S., Ma, W.Y., Chaisorn, L., Bakker, E.M., eds.: Image and Video Retrieval. Volume 3568 of LNCS., Singapore, Springer (2005) 507–517
[27] Hare, J.S.: Saliency for Image Description and Retrieval. PhD thesis, University of Southampton (2005)
[28] Hare, J.S., Lewis, P.H., Enser, P.G.B., Sandom, C.J.: Mind the gap. In Chang, E.Y., Hanjalic, A., Sebe, N., eds.: Multimedia Content Analysis, Management, and Retrieval 2006. Volume 6073., San Jose, California, USA, SPIE (2006) 607309–1–607309–12
[29] Berry, M.W., Dumais, S.T., O'Brien, G.W.: Using linear algebra for intelligent information retrieval. Technical Report UT-CS-94-270, University of Tennessee (1994)
[30] Landauer, T.K., Littman, M.L.: Fully automatic cross-language document retrieval using latent semantic indexing. In: Proceedings of the Sixth Annual Conference of the UW Centre for the New Oxford English Dictionary and Text Research, Waterloo, Ontario, Canada (1990) 31–38
[31] Schreiber, A.T.G., Dubbeldam, B., Wielemaker, J., Wielinga, B.: Ontology-based photo annotation. IEEE Intelligent Systems **16** (2001) 66–74
[32] Hunter, J.: Adding multimedia to the semantic web: Building an mpeg-7 ontology. In Cruz, I.F., Decker, S., Euzenat, J., McGuinness, D.L., eds.: SWWS. (2001) 261–283

[33] Tsinaraki, C., Polydoros, P., Moumoutzis, N., Christodoulakis, S.: Coupling owl with mpeg-7 and tv-anytime for domain-specific multimedia information integration and retrieval. In: Proceedings of RIAO 2004, Avignon, France (2004)
[34] schraefel, m., Karam, M., Zhao, S.: mSpace: interaction design for user-determined, adaptable domain exploration in hypermedia. In De Bra, P., ed.: Proceedings of AH 2003: Workshop on Adaptive Hypermedia and Adaptive Web Based Systems, Nottingham, UK (2003) 217–235
[35] Kompatsiaris, I., Avrithis, Y., Hobson, P., Strinzis, M.: Integrating knowledge, semantics and content for user-centred intelligent media services: the acemedia project. In: Proceedings of Workshop on Image Analysis for Multimedia Interactive Services (WIAMIS '04), Lisboa, Portugal (2004)
[36] Dupplaw, D., Dasmahapatra, S., Hu, B., Lewis, P.H., Shadbolt, N.: Multimedia Distributed Knowledge Management in MIAKT. In Handshuh, S., Declerck, T., eds.: Knowledge Markup and Semantic Annotation, 3rd International Semantic Web Conference, Hiroshima, Japan (2004) 81–90
[37] Library of Congress: Srw. `http://www.loc.gov/standards/sru/srw/index.html` (2005)
[38] MyGRID Project: Taverna. `http://taverna.sourceforge.net` (2006)
[39] Cunningham, H., Maynard, D., Bontcheva, K., Tablan, V.: GATE: A framework and graphical development environment for robust NLP tools and applications. In: Proceedings of the 40th Anniversary Meeting of the Association for Computational Linguistics. (2002)
[40] Enser, P.G.B., Sandom, C.J., Lewis, P.H.: Surveying the reality of semantic image retrieval. In Bres, S., Laurini, R., eds.: VISUAL 2005. Volume 3736 of LNCS., Amsterdam, Netherlands, Springer (2005) 177–188
[41] World Wide Web Consortium: SKOS Core Guide `http://www.w3.org/TR/swbp-skos-core-guide/`. Technical report, A. Miles and D. Brickley, eds. (2005)
[42] Princeton University: WordNet: a lexical database for the English language. `http://wordnet.princeton.edu` (2006)