Image annotation with Photocopain

Mischa M. Tuffield[1], Stephen Harris[1], David P. Dupplaw[1], Ajay Chakravarthy[2], Christopher Brewster[2], Nicholas Gibbins[1], Kieron O'Hara[1], Fabio Ciravegna[2], Derek Sleeman[3], Nigel R. Shadbolt[1], and Yorick Wilks[2]

[1] Department of Electronics and Computer Science, University of Southampton, United Kingdom.
[2] Department of Computer Science, University of Sheffield, United Kingdom
[3] Department of Computing Science, University of Aberdeen, United Kingdom

Abstract. Photo annotation is a resource-intensive task, yet is increasingly essential as image archives and personal photo collections grow in size. There is an inherent conflict in the process of describing and archiving personal experiences, because casual users are generally unwilling to expend large amounts of effort on creating the annotations which are required to organise their collections so that they can make best use of them. This paper describes the Photocopain system, a semi-automatic image annotation system which combines information about the context in which a photograph was captured with information from other readily available sources in order to generate outline annotations for that photograph that the user may further extend or amend.

1 Introduction

In this paper we describe the Photocopain photo annotation system, which is designed to reduce the effort of the annotation task by taking advantage of as much information as possible from free or readily available sources. We have adopted a strategy of semi-automatic annotation; while human-authored annotations are expensive, they are seen as a 'gold standard', and have considerable value by virtue of their provenance. Ultimately, our plan is to relate photographs together by developing narratives from the annotations; in this paper, we describe an architecture that can (a) integrate a number of existing tools and focus them on the annotation task, and (b) integrate as much information as possible from highly heterogeneous sources.

Photocopain integrates a variety of annotation, information extraction and integration, and user interface tools from the Advanced Knowledge Technologies project[4] (AKT), a six-year interdisciplinary collaboration working in the general area of technology and infrastructure development for the Semantic Web. AKT's aim is to extend and develop integrated technologies and services, using the Semantic Web as a unified basic framework, for acquiring, retrieving and

[4] `http://www.aktors.org/`

publishing content over the whole of the knowledge life cycle. Photocopain has also been inspired by the research programme of the *Memories for Life* network[5] (M4L) [1], which draws together a number of recent technological and scientific advances from both life sciences and computer science in the area of human and artificial memory. M4L focuses on the use of technology alongside human memory in context, to provide support for memory management.

Photographs can be viewed as externalised additions to human memory; many human memory management tasks rely on collections of photographs, even if only shoeboxes of photographic prints. Furthermore, as digital technology has dramatically increased the numbers of photographs taken (it has been estimated that up to 375 petabytes of information is created in photographic form annually), the problems associated with archiving and managing these photographs have become more pressing (describing their content, storing and retrieving them, and developing narratives to link them together). Image management is a highly labour-intensive task, and we have designed Photocopain with the intent of alleviating such burdens.

2 Related work

A number of other projects have similar aims with respect to the archiving of personal experience, most notably MyLifeBits [2], and SemanticLIFE [3] are the most noteworthy. The SemanticLIFE project uses a variety of techniques to extract metadata from personal communications, and and considers issues relating to provenance and trust; these are presented in Weippl *et al* [4]. MyLifeBits aims to organise retell personal experiences using narrative as a structuring mechanism; [5] presents valuable insight on how stories can be generated based on information such as location and time. However, neither of these systems strives to accumulate the low-cost metadata at which Photocopain is aiming.

The generation of annotations from multimedia content can be seen as an application of image classification of the type used in content-based indexing and retrieval[6]. Systems such as MAVIS2[7] and SIMPLIcity[8] use collections of image classifiers (embodied as agents, in the case of MAVIS2) to add semantic descriptions of the images' contents.

User interface issues arise when designing the annotation and the browsing stages of Photocopain. Narratives are not the only way envisaged for users to interact with Photocopain; we intend to apply the lessons learnt from faceted browsers such as mSpace [9], and from Haystack [10], to help the user navigate through the generated knowledge space. Photocopain also aims to develop a user interface that will ease the burden of annotation. Our initial investigation has suggested that avoiding the use of *deep tree-lists*, as employed by MindSwap's PhotoStuff system [11], and adopting a simple interface similar to that of `flickr.com`, will be a step in this direction.

[5] `http://www.memoriesforlife.org/`

3 Annotation Sources

We have examined potential sources of information that can be used to produce annotations, with a particular emphasis placed on ambient contextual information which can be applied to photographs with minimal effort on the part of the user. At this stage, we deliberately considered the annotation task in abstract terms, making the conscious decision to examine the inputs to the annotation task, rather than the structure of the task itself, for two main reasons:

First, with this research, we intended to examine mechanisms and architectures for acquiring and organising cheaply available information; we consider information to be cheap if its acquisition does not require a great deal of human intervention. Our initial research milestone was to discover which information sources *are* cheap, as opposed to appearing cheap; easily acquired information may have hidden costs associated with its representation, or its integration with other information.

Secondly, it is clear that a user carrying out the image annotation task has a number of different priorities. They will be interested in different aspects of the photograph depending on the specific task they are performing: annotating a photo library for the press; organising a personal collection of photographs; or selecting CCTV stills during a security operation. Rather than developing a special-purpose system that only supports a specific task, we have developed a flexible system architecture that allows for images to be annotated by combining multiple information sources. However, it may be that the readily-available information that we are interested in is more useful for the task on which we are concentrating, namely that of annotating personal photo collections, than others, highlighting the importance of a flexibile reuseable architecture.

The information sources that we have identified for the task of annotating personal photo collections include the following:

Camera metadata.
Exchangeable Image File (EXIF) [12] metadata record camera parameters at the point that the photograph was taken. These parameters include: aperture setting; focal length of the lens; exposure time; time of photo; flash information; camera orientation (portrait/landscape); and focal distance. We can also derive other information from these metadata, such as the average scene brightness of an image. There may be further image metadata in the NAA Information Interchange Model (IPTC) [13] fields; this information is not typically captured in the camera, but is added subsequently to image capture.

Global positioning data.
GPS data can be recorded live in EXIF or IPTC, or alternatively GPS track logs can be used to determine location accurately. This is primarily of use when the camera is used outdoors.

Calendar data.
Calendaring software produces machine-readable diaries and calendars are available in a number of formats, including iCal [14]. These can be used

to provide a personal context for a photograph, by recording where the user planned to be when the image was taken. Although iCal files provide information about the timing of events, the descriptions of the events are in free text; we perform simple named entity extraction on the location string to identify place names with which the image is then annotated.

Image analysis.
Image analysis techniques such as low resolution colour analysis or local image sharpness measurements can also be used to infer information about the subjects of a photo. We describe the feature extraction techniques that we have used more fully in Section 4.3.

Community tagging.
We use the community photograph site `flickr.com` as a source of training data for the Photocopain system. Flickr users may associate images with a number of free text tags (e.g. 'Les Carr', 'WWW2006', 'Edinburgh'); we use the photographs associated with certain tags as a training set for our image analysis algorithms.

Given our aim of supporting personal photo collections, we selected a handful of flickr's most popular tags to be our initial content-based annotations (vocabulary). These include: landscape, cityscape, portrait, groupphoto, architecture, seascape, and flower. The decision to use this dataset has ensured that any proposed annotations are grounded within flickr's shared conceptualisation of these terms. For example, if Photocopain proposes an annotation "landscape", what it actually means is "this image is similiar to images tagged landscape by the flickr community" as apposed to the developer's understanding of the word. A caveat worth mentioning arises due to our reliance on the EXIF data, which means that we only ever use a subset of flickr's images, i.e. the ones with sufficient embedded metadata.

Another incentive of using flickr was to facilitate our initial goal of developing a generic architecture to support image annotation. The very nature of our task requires personalisation; for example a user may not be interested in taking landscape photos, and may wish to have a classifier based on another flickr tag such as "car".

Although these information sources are quite basic, and will not allow the creation of sophisticated annotations such as "Yet another shot of Briony falling into the pool", they can be combined to derive other relevant information. For example, the focal length, focal distance and sensor size could be used to calculate the size of an object at the focal point, which might enable us to differentiate between a leaf and a tree, even though both are in the same location and have a similar colour. The aim of our system is not to create annotations per se, but to generate hints which guide the user as they create the annotations.

As a further example, consider how colour information and an edge detection algorithm can be used in combination with the technical information available from the camera and other devices. If we know that the picture was taken with a wide angle lens, the GPS log shows that the photo was taken in a city, and the picture is largely grey with lots of coherent edges, we may be able to infer that the picture is of a building or a piece of architecture, or even a cityscape if it contains lots of blue at the top of the picture.

Understanding of photographic genres is certainly relevant here. For instance, portraits are typically taken at focal lengths of around 100mm in 35mm equivalent length, with wide aperture (f/4 or above), fill flash (reduced power), and the subject of the photo, a head/shoulders, will be of a standard size. The background at top left and top right will be out of focus (low sharpness), and the photo will have a portrait aspect ratio. These are fairly standard settings which are used by the majority of photographers; they are also the settings that an automatic camera set to 'portrait' will usually use. In the latter case, the photograph may record the automatic setting of the camera, in which case the annotation system may not need to search for the individual settings of the camera.

4 Photocopain

The Photocopain system has a service-based architecture, as shown in Figure 1, and has been designed so that new annotation or classification services may be added on an ad-hoc basis. The connections between system components have been implemented as simple Web Services (HTTP or SOAP) wherever possible, and interactions with the central RDF triplestore make use of the SPARQL RDF query language[15]. We now describe the components of the Photocopain system, and illustrate their interaction with a simple workflow.

4.1 RDF Triplestore

At the heart of the system is the SPARQL-compliant [15] RDF triplestore 3store [16], which is used to store the image annotations, both candidate and user-vetted. The key role of the triplestore is to act as a persistent store for the system, and to mediate the interactions between the other system components.

4.2 Data Sources

The inputs provided to Photocopain include photographs, calendars and geographical positioning data. Photographs are retrieved from the web servers to which they have been uploaded by the photographers. The EXIF image metadata is extracted and converted into RDF using the W3C EXIF schema [17]. This basic metadata is then augmented with other calculated values from the image metadata (light levels, for example), and a collection of feature extraction algorithms are run over the images (see Section 4.3). These metadata and features are then asserted into the triplestore. Information about the user's schedule is taken from a calendar server (typically in iCal format), and converted into RDF-iCal before being asserted into the triplestore. At present, we have made the assumption that this temporal information is limited to the user's calendar,

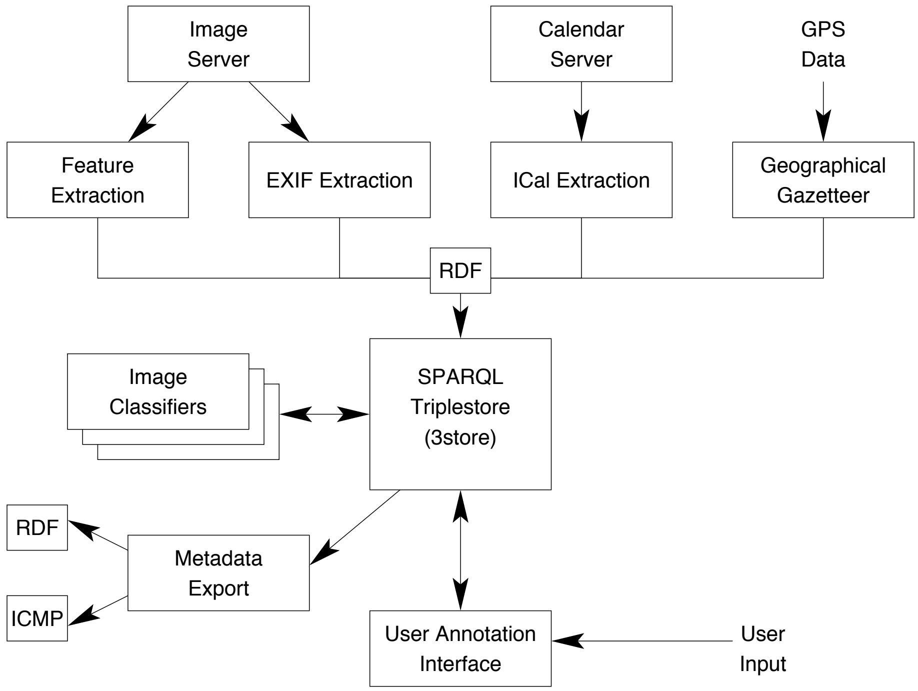

Fig. 1. The Photocopain Architecture

but there may be other sources which could prove valuable (for example, news or weather feeds).

Finally, we determine positioning information for the photograph by comparing tracks logs with the time the photograph was taken. We currently use GPS and WiFi access points for position information, but we envisage that we could also use GSM phone services to determine the location of the user at the time the photograph was taken. We use a gazetteer with an ontology of places to provide additional geographical information (e.g. city, country, timezone, wilderness), which is linked with other data from the user and asserted into the triplestore.

4.3 Image Feature Extraction and Classifiers

Much of the current work on semantic annotation of multimedia data attempts to solve the computer vision problem. Our guiding principle with Photocopain has been simplicity; we study the use of simple, well-understood techniques with additional metadata. The system is extensible so that new feature extraction algorithms can be included to extend its breadth of recognition, or its classification accuracy. For prototyping purposes we have implemented four feature extraction techniques from the literature to see how they perform with the extra contextual information that we are able to provide.

A simple CIELab colour-map is used alongside other feature extraction methods for increasing the quality of a match. The colour map lays a 10x10 grid over the image and takes the average colour in the perceptually-uniform CIELab colour-space for each grid element. This spatial-colour map can be used to qualify classifications (such as providing more evidence towards a 'landscape' annotation due to the top of the image being the colour of sky, for example).

A face detection module provides an estimate of the number of faces that appear within an image. This uses the Hue, Intensity, Texture (HIT) Map[18] technique to detect face-like areas. The method performs a combination of thresholded hue, and intensity channels to obtain a 'skin' image. Clearly this is quite a simplification of the description of a face, but it works to identify skin regions well. After connected-component labelling and segmentation, regions are filtered based on shape and size allowing the module to return only those images which are likely to be faces. However, this process often leads to false positives, so combining this with the EXIF data from the camera lens, an approximate physical size for the possible face region can also be calcuated. As faces can only ever be certain sizes, this provides more evidence for identifying whether a region within the image is a face. Combined with other EXIF-based metadata it is possible to annotate images as being portraits, or group photos.

The artifical vs. natural feature extractor uses classification of the edge direction coherence vector[19] to classify image content, based on the assumption that artificial structures tend to be created with straight-edges, whereas natural structures do not. A Canny edge-detector is used to filter the image and short edges are binned into a histogram. An edge-tracing algorithm is used to find those edges that are coherent, and these are also binned into another histogram. The two histograms make up the edge direction coherence vector, which can be classified against, or is able to provide a measure of "artificial-ness". In our early prototype we exclusively classified an image as artifical or natural, however, it became clear that an image can be both artificial and natural simultaneously (for example a photo of a cityscape behind a park). Creating two classifiers, rather than one, allows images to be annotated with both labels if necessary.

A focus-map performs Fourier transforms over the image in a grid. These transforms are filtered to obtain a measure of high and low frequency components within each grid element. The measures are then linearly combined at each location to provide an overall map for the image, of where focus lies. In images that have a shallow depth of field, and therefore contain unfocussed regions, the map is able to provide both evidence of this, and also information on where the feature extractions should be targetted, if possible. Combined with other techniques, this extraction may be able to provide more evidence towards portrait annotations (photos of which tend to have shallow depth of fields at a medium focal length), or landscape annotations (photos of which tend to have a large depth of field, focussed near the hyperfocal distance).

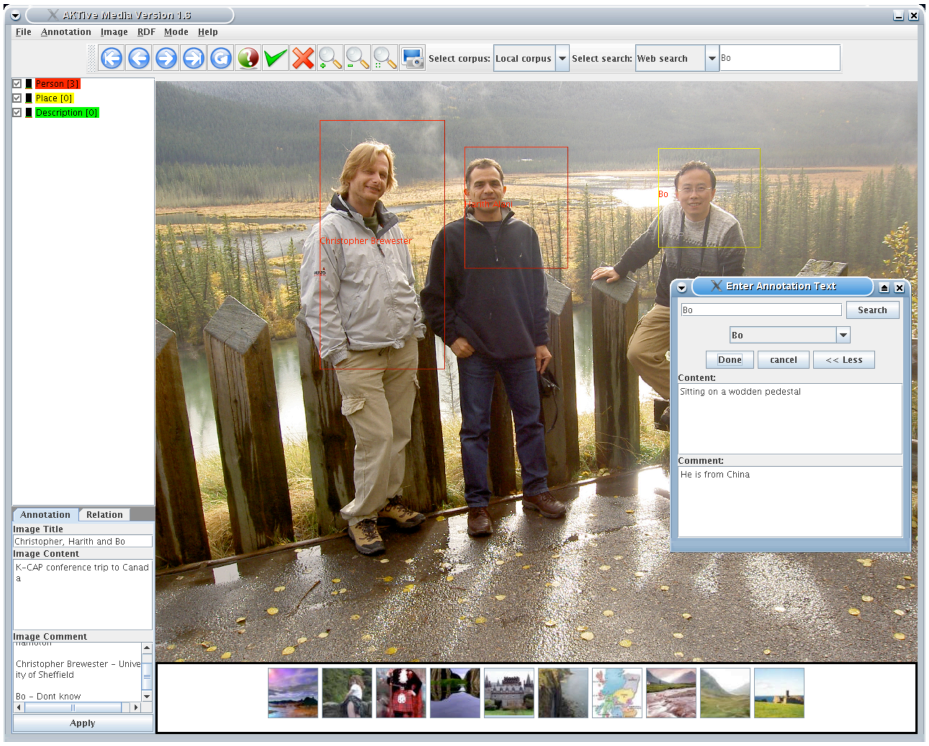

Fig. 2. The AKTive Media Annotation Tool

4.4 User Annotation Interface

In the Photocopain system, we treat the annotations produced by the image classifiers and the context-based annotators as candidate annotations that require vetting by a user. We use the AKTive Media system[20], shown in Figure 2, to allow users to expand and elaborate on (and where necessary, to correct) the automatically generated candidate annotations.

AKTive Media is a stand-alone, user-centred, multimedia document annotation system which we use as the user interface for Photocopain. At a basic level, AKTive Media allows manual annotation of images and image regions. There are two different modes of operation: annotation of a single image, or batch annotation of a set of images. Annotations in AKTive Media can be created either as arbitrary free-text labels, or with respect to an appropriate domain ontology. Ontologies may be selected and loaded into the pane on the left of the user interface, and AKTive Media can then assist the user in selecting the correct terms with which to describe the image.

The candidate annotations proposed by the other modules in Photocopain are retrieved from the RDF triplestore and made available to the user for inspection and confirmation. Some candidate annotations undergo further processing; if a caption exists for the whole image, this is analysed using natural language processing techniques in order to generate candidate annotations for regions of the image. For example, if a free-text caption for an image contains a number of named entities, and the user has identified certain regions of the image for further annotation, then there is a likelihood that the image regions may correspond to named entities in the image caption. Finally, AKTive Media provides a metadata search facility for annotated images, either by free-text keywords, or by ontology terms.

In addition to being submitted to the central RDF triplestore, image annotations are also cached within the application, which allows access to previous annotations made during a particular session. AKTive Media can also retrieve annotations which have been made about an image by other users, which the user can use to inform their own annotations. Although our pilot application is the management of personal photo collections, this demonstrates how the Photocopain system might be made to support a community rather than a single user.

4.5 Metadata Export

Once a set of annotations have been generated for an image, they may be exported as RDF. These metadata may be sent back to the image's originating web server, or to a web gallery; the RDF is used to add annotations to the HTML pages containing the image, allowing other systems to search and process the images. Alternatively, the metadata may be stored within the image itself by being translated into ICMP format. This is a more appropriate approach if the image is to be submitted to a service which is better placed to use within-image metadata, such as an image library or archive.

One of our long term goals for this work is the creation of narrative structures which present photographs in context, in order to better archive personal experiences. The annotations form a foundation for these structures, and enable chronological and narrative based queries over the data such as: "Where did we go in the last week of our holiday in the Peloponnese?" or "Who pulled me out of the crevasse I fell into in the mountains?" or even "What happened next?". This form of meta-level query is a valuable way of unlocking memories; we use the narrative structure to facilitate the most effective presentation of the data to the user whether that be as a slide show, themed collection (such as an online album), a structured tree [21], or a knowledge exploration interface such as that used in the Flamenco project [22].

4.6 Workflow

The annotation workflow for the Photocopain system is shown in Figure 3. Broadly speaking, there are two parallel workflows shown in this figure: one concentrates on annotations which can be gleaned from the content of the image, while the other creates annotations based on its context.

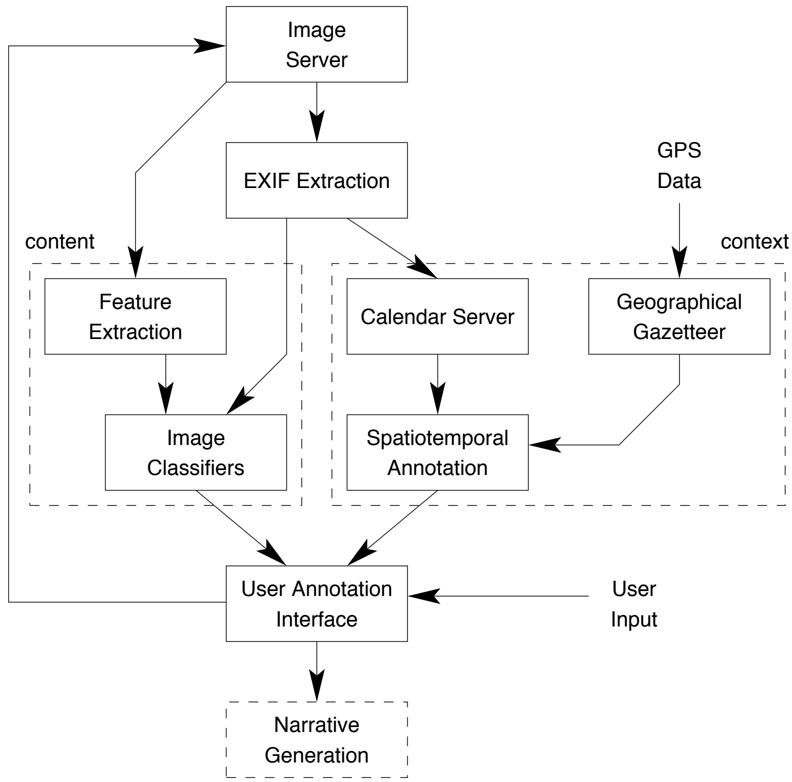

Fig. 3. Photocopain Annotation Workflow

The content-based annotation workflow, shown on the left, contains an EXIF extraction service which retrieves the camera metadata from the image, and a set of feature extraction services which analyse the content of the image. These metadata and features are passed to the image classifiers, which perform simple classification on the images (e.g. natural vs. artificial, indoors vs. outdoors, and so on). Finally, these classifications are passed as candidate annotations to the user annotation interface.

The context-based workflow is executed at the same time, and uses the EXIF metadata as the basis for its annotation. Time and date information are used to identify contemporary events in the user's calendar, and are correlated with GPS track data to identify a geographical locus within which the image was taken. This geographical information is expressed as lat-long coordinates, and is converted to user-friendly labels (the name of a town or relevant geographical feature) using a gazetteer of placenames and locations. This temporal and spatial information is then passed to the user annotation interface.

In the AKTive Media user annotation interface, the user may accept any of the candidate annotations, with or without editing, and may also create new annotations of their own. Finally, the annotations created in the user annotation interface may then be exported as RDF back to the RDF triplestore, to the originating image server, or passed to further services.

5 Evaluation

Our initial evaluation has focused on the performance of our novel combinations of context and content based feature classifiers. In both cases, we use a training set of 150 images, with 75 instances representing each classification, and a test set of 30 images.

For the classification of images as indoor or outdoor, we have found that a combination of the EXIF data and the CIELab colour map data without any principle component analysis performed the best, yielding only 4 errors in the 30 tests, giving us a precision of around 87%. Our classification of natural and artificial environments yielded its best results with a combination of the edge direction coherence vector, the CIELab colour map and the nearest neighbour clustering algorithm, giving 3 errors in the 30 test, a precision of 90%.

We are currently creating more automatic classifiers (flowers, seascape, cars), and changing the decision process from a binary one (i.e. either indoor or outdoor) to unary one containing certainity factors, such as indoor 90%, outdoor 5%, artificial 70%, flower 2% and so on, and will revisit the evaluation of our classifiers when this is completed. We also intend to perform a more holistic evaluation of Photocopain by using it to annotate the photographic record of a small-scale event such as a workshop where the parameters of time and space can be mapped relatively easily onto a standard calendar of events. We will test our modules incrementally, as our datasets grow in size

6 Future Work

Once annotations giving a spatiotemporal context for a photograph have been established, it becomes possible to use narrative structures to organise image collections and exploit this context. Our current research is concentrating on the construction of narratives and stories to allow users to make better sense of the continuous stream of digital data they are acquiring[23]. These narratives could link a collection of photos taken when on holiday in a foreign country, or a collection of photos taken at conferences during the previous year. Assuming a story is the presentation of a succession of events based around a given topic, we aim to generate such narratives based on the resources available.

To this end, we propose a two-layered approach to allow maximum automation and full flexibility. The first layer is an automatically generated time line, based on the timestamps of the images, and using a default chronological granularity generated from the calendar spread of the data. At this level, user intervention is limited to altering the granularity of the time segments should the user feel this necessary. We combine this with location informaion, which provides useful cues for generating stories to refresh existing memories [5].

The second layer organises photos into a narrative structure based on educated guesses made from the metadata, allowing users to correct or adjust the narrative; it is a core research issue in this project to determine the appropriate balance between user intervention and automatically generated narrative structures. In essence, the annotated images function as instances in an underlying narrative ontology (or narrative script), which will allow for sequential organisation of the instance events into a cognitively coherent narrative structure [24, 25]. These scripts contain stereotypical narrative constructs [26] (analogous to the stereotypical interactions described by Schank [27]) from which the user specifies sequential and causal relations between images, and have been influenced by the definition of narrative concepts in the OntoMedia ontology [28].

7 Conclusion

In this paper, we have presented the Photocopain system, which integrates a number of existing and well-understood technologies that address the task of photo annotation, both those which rely on the content of the images, and those which examine their context.

Organising non-textual artifacts such as photographs in such a way as to facilitate retrieval is a difficult problem; the usual retrieval strategy of exploiting keywords is of course unavailable without annotation. However, annotation requires significant effort on the part of the user, and so we have focussed on ways of automating or semi-automating the image annotation task. Photocopain takes an unusual route of using information that is readily and, in many cases, cheaply available. This information is extremely heterogeneous, and the value of much of it is low (certainly compared to manually-generated annotations). Although the benefits may be small, the low costs of acquiring it may mean that some of these sources are still worth exploiting. We argue that it is possible to process and integrate this sort of information, and that basic processing might yield further information. Such information, and inferences over it, could be used to help suggest annotations or annotative strategies for users, particularly in sharply circumscribed domains, and in relatively routine annotation tasks.

It is clear that much work has still to be done, and that the value of the various information sources needs to be evaluated in genuine task contexts. Nevertheless, given the cost of annotation, experiments on the value of found information is producing interesting results.

8 Acknowledgements

This work is supported under the Advanced Knowledge Technologies (AKT) Interdisciplinary Research Collaboration (IRC), which is sponsored by the UK Engineering and Physical Sciences Research Council under grant number GR/N15764/01.

References

1. O'Hara, K., Shadbolt, N., Hitch, G., Hall, W., Beagrie, N.: Memories for life: a review of the science and technology. Journal of the Royal Society Interface (2006) In press.
2. Gemmell, J., Lueder, R., Bell, G.: The MyLifeBits lifetime store. In: Proceedings of ACM SIGMM 2003 Workshop on Experiential Telepresence (ETP 2003), Berkeley, CA (2003)
3. Admed, M., Hoang, H.H., Karim, M.S., Khusro, S., Lanzenberger, M., Latif, K., Michlmayr, E., Mustofa, K., Hguyen, H.T., Rauber, A., Schatten, A., Tho, M.N., Tjoa, A.M.: 'SemanticLIFE' - a framework for managing information of a human lifetime. In: In Proceedings of the 6th International Conference on Information Integration and Web-based Applications and Services, Jakarta, Indonesia (2004)
4. Weippl, E., Schatten, A., Karim, S., Tjoa, A.M.: SemanticLIFE collaboration: Security requirements and solutions - security aspects of semantic knowledge management. In: PAKM. (2004) 365–377
5. Gemmell, J., Aris, A., Lueder, R.: Telling stories with MyLifeBits. ICME 2005 **8** (2005) 6–9
6. Smeulders, A., Worring, M., Santini, S., Gupta, A., Jain, R.: Content-based image retrieval at the end of the early years. IEEE Trans. on Pattern Analysis and Machine Intelligence **22**(12) (2000) 1349–1380
7. Joyce, D., Lewis, P., Tansley, R., Dobie, M., Hall, W.: Semiotics and agents for integrating and navigating through multimedia representations. Proc. Storage and Retrieval for Media Databases **3972** (2000) 120–131
8. Wang, J., Li, J., Wiederhold, G.: Simplicity: Semantics-sensitive integrated matching for picture libraries. IEEE Transactions on Pattern Analysis and Machine Intelligence **23**(9) (2001) 947–963
9. schraefel, m. c., Karam, M., Zhao, S.: mSpace: interaction design for user-determined, adaptable domain exploration in hypermedia. In Bra, P.D., ed.: Proceedings of AH 2003: Workshop on Adaptive Hypermedia and Adaptive Web Based Systems., Nottingham, UK, Springer (2003) 217–235
10. Huynh, D., Karger, D., Quan, D.: Haystack: A platform for creating, organising, and visualising information using RDF. In: Semantic Web Workshop, The Eleventh World Wide Web Conference (WWW2002). (2002)
11. Halaschek-Wiener, C., Schain, A., Grove, M., Parsia, B., Hendler, J.: Management of digital images on the Semantic Web. In: Proceedings of the International Semantic Web Conference (ISWC2005), Galway, Ireland (2005)
12. EXIF: Exchangeable image file format for digital still cameras: EXIF version 2.2. Technical report, Japan Electronics and Information Technology Industries Association (2002)
13. IPTC: IPTC - NAA information interchange model version 4. Technical report, Comit International des Tlcommunications de Presse, Version No. 4 (1999)
14. Dawson, F., Stenerson, D.: Internet calendaring and scheduling core object specification. Request For Comments 2445, Internet Engineering Task Force (1998)
15. World Wide Web Consortium: SPARQL query language for RDF, working draft. Technical report, World Wide Web Consortium (2005)
16. Harris, S., Gibbins, N.: 3store: Efficient bulk RDF storage. In: PSSS'03, Sanibel, Fl (2003) 1–15
17. Schema, E.W.R.: Exif vocabulary workspace - rdf schema. Technical report, W3C (2003)

18. Mateos, G.G., Chicote, C.V.: A unified approach to face detection, segmentation and location using HIT maps. In: In Symposium Nacional de Re-conocimiento de Formas y Anlisis de Imgenes, Benicasim, Castelln (2001)
19. Vailaya, A., Jain, A., Zhang, H.: On image classification: City images vs. landscapes. In: Pattern Recognition. (1998) 1921–1935
20. Chakravarthy, A., Lanfranchi, V., Petrelli, D., Ciravegna, F.: Requirements for multimedia document enrichment. In: Poster Proceedings of the Fifteenth World Wide Web Conference (WWW2006). (2006) To appear.
21. Spinelli, G.: A multimedia system to collect, manage and search in narrative productions. In Bullinger, H.J., Ziegler, J., eds.: HCI (1), Lawrence Erlbaum (1999) 451–455
22. Yee, K.P., Swearingen, K., Li, K., Hearst, M.: Faceted metadata for image search and browsing. In: Proceedings of the conference on Human factors in computing systems, ACM Press (2003) 401–408
23. Zarri, G.P.: A knowledge engineering approach to deal with 'narrative' multimedia documents. In Bjørner, D., Broy, M., Zamulin, A.V., eds.: Ershov Memorial Conference. Volume 2244 of Lecture Notes in Computer Science., Springer (2001) 363–377
24. Furtado, A.L.: Narratives and temporal databases: An interdisciplinary perspective. In Chen, P.P., Akoka, J., Kangassalo, H., Thalheim, B., eds.: Conceptual Modeling. Volume 1565 of Lecture Notes in Computer Science., Springer (1997) 73–86
25. Hobbs, R.L.: Using XML to support military decision-making. In Wood, L., ed.: XML Conference, Philadelphia, PA (2003) 7-12 December 2003.
26. Uhlír, J., Falc, M.: Annotating narratives using ontologies and conceptual graphs. In: DEXA Workshops, IEEE Computer Society (2004) 84–88
27. Schank, R.C., Abelson, R.P.: Scripts, Plans, Goals and Understanding: an Inquiry into Human Knowledge Structures. L. Erlbaum (1977)
28. Jewell, M.O., Lawrence, F., Tuffield, M.M., Nixon, M.S., Prügel-Bennett, A., Shadbolt, N.R., Millard, D.E., schraefel, m. c.: OntoMedia: An ontology for the representation of heterogeneous media. In: Proceedings of MultiMedia Information Retrieval (MMIR) Workshop at SIGIR, SIGIR (2005)