SHOEBOX: A DIGITAL PHOTO MANAGEMENT SYSTEM

Timothy J. Mills, David Pye, David Sinclair, and Kenneth R. Wood

AT&T Laboratories Cambridge,
24a Trumpington Street, Cambridge CB2 1QA, England

shoebox@uk.research.att.com

ABSTRACT

This paper reports recent work at AT&T to develop a system for the management of personal digital photograph collections. Shoebox, the resulting software package, provides a range of browsing and searching facilities, utilising spoken annotations and image content to enable both semantically similar and visually similar images to be retrieved. We report on the design of the system, the construction of a test collection, and the evaluation of its searching facilities. The results show that audio annotation is an effective means of retrieval for photographs, which significantly out-performs image content-based techniques.

Keywords: speech indexing, image retrieval, test collection development for MMIR.

1. INTRODUCTION

The DART (Digital Asset Retrieval Technology) project [1] at AT&T Laboratories Cambridge is concerned with management of digital media such as text and hypertext documents, images, audio and video recordings. DART aims to provide the means to index, annotate, navigate and retrieve from diverse collections of these assets. The project was motivated in part by our successful collaboration with Cambridge University in the Video Mail Retrieval project [2,3] but it has evolved to encompass a much broader range of issues in multimedia information retrieval.

In developing the core DART technologies, we have built an application to help manage personal photograph collections. With the increasing popularity of digital cameras, the cost of producing large numbers of photographs has been dramatically reduced. We therefore believe that a tool to manage such collections of photographs will be essential. The approach taken exploits both image content and text derived from spoken annotations. The user-interface is crucial to the success of the application and has been designed with speed, flexibility and visual appeal in mind.

The human effort required to annotate a photograph is often attacked as a non-scalable element. However, people are quite often willing to spend time talking about their own photographs. Thus by performing speech recognition on spoken annotations, we generate a source of semantic information about the picture content which is amenable to text retrieval. In fact, some digital cameras are already equipped with digital audio recording facilities.

To extract visual information, we use image segmentation in which each image is divided into regions based on colour and texture properties. The properties of each region are described by feature vectors which are indexed to allow quick content-based image retrieval.

A personal collection of some 600 annotated photos has been obtained. We report the retrieval effectiveness of the system with this collection, evaluating annotation-based retrieval and a number of image-content based methods.

2. RELATED WORK

There are now a number of production-quality commercial and freeware packages available for managing digital photos, including CompuPic [4], ThumbsPlus [5] and various packages bundled with digital cameras. Some of these provide limited visual content-based indexing, but we have seen none that uses speech recognition to provide searchable audio annotations, and none has been subjected to any formal evaluation.

Relevant research-led systems include Show&Tell [6,7] and FotoFile [8]. Show&Tell also focuses on personal image collections and utilises audio annotations in addition to visual-based methods. An annotation is made when a photograph is taken. In contrast, our software encourages annotations to be made when loading images into the system. Thus the annotations can be made under acoustically good conditions, alleviating the reported problems with speech recognition accuracy. FotoFile allows users to manage more general multimedia objects, e.g. video as well as still photos, but again focuses on personal collections. Although it does not employ audio annotations, it does place great emphasis on making annotation easy for users. It also uses various types of visual content-based indexing including face detection and recognition.

Most visual-based image retrieval techniques start with histograms of pixel properties, whereas our primary technique begins by segmenting an image into coherent regions. Blobworld [9] takes a similar approach to ours, although with an emphasis on a few salient regions rather than a complete segmentation of the entire image. An interesting hybrid approach is presented in [10].

There have been a number of projects investigating the suitability of automatic speech recognition transcripts for information retrieval including, for example, [3] and [11]. Our use of ASR transcripts differs from this work in that they are used as a means to retrieve the associated photographs, not the annotations themselves.

3. OVERVIEW OF SHOEBOX

Shoebox aims to provide a comprehensive set of facilities for the management of a digital photograph collection, including photograph acquisition, browsing, searching and publishing.

A user creates a Shoebox into which photographs (or other images) may be placed, sourced from a digital camera, scanner, the filesystem, web pages or the Windows clipboard. These can be grouped by the user into "rolls" of arbitrary size.

Both rolls and photographs can be annotated with either text or audio. Speech recognition is applied to audio annotations to generate a text transcript. This text is used in addition to photograph and roll titles as a source of keywords for text indexing, as described in Section 5.1.

A selection of image segmentation schemes may be applied to the images to generate indexing terms for content-based image searching, as detailed in Section 5.2.

Functions are supplied to correct basic colour properties, rotate images, and send the photographs to any image manipulation tools for further editing. Photographs may be printed, sent via email, or published as a set of web pages together with their accompanying annotations. These web pages may be downloaded back into Shoebox, enabling simple exchange of annotated photos across the Internet.

With the exception of the image and audio files themselves, Shoebox stores all its data and metadata in an object-oriented database which was designed specifically to support information retrieval [12]. It makes use of an inference network retrieval model based on that of [13].

4. BROWSING

The most basic and important feature of any photo-management software is its support for browsing photographs. Shoebox provides several ways of browsing the photograph collection.

For very quick browsing of this index, we make use of "tooltips" i.e. a window that appears next to the cursor when it hovers over an item (Figure 1). Roll tooltips display the source of the images (e.g. digital camera, filesystem etc.). Photograph tooltips display the time and date on which the photograph was taken, a thumbnail of the photograph, and indicate whether the photograph has been segmented or annotated. Annotation tooltips display the name of the annotator and the date and time at which the annotation was made. Additionally, for text and speech-recognised audio annotations, up to five (compound) nouns from the annotation are displayed. These nouns are selected as follows. A Brill tagger [14] is used to determine noun phrases in the transcript. These are scored heuristically so as to favour high-frequency proper nouns, and the top five are selected. This aims to pick out any names of people or places which are often contained in the annotations. For audio annotations, the audio is played while the user hovers over the annotation.

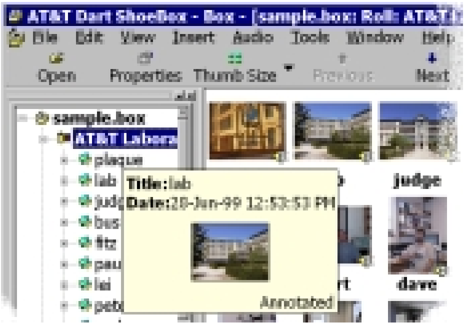


Figure 1: Roll view with image tooltips.

The timeline view provides chronological navigation of the photographs (Figure 2). This index groups photos by date, either as a conventional calendar view or in date clusters, where photos taken on consecutive days are grouped together by month and year. This enables easy selection of photos within a date range.

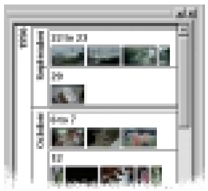


Figure 2: Timeline view.

The third browsing index is a content-based view (Figure 3). We create a hierarchy of compound nouns extracted from the annotation text or speech transcripts using the techniques described in [15]. This "topic view" gives a textual overview of the contents of a photo collection. Selecting a term from the index retrieves all photos containing that term.

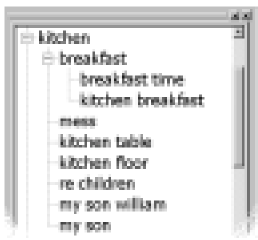


Figure 3: Topic view.

Screenshots depicting most of the features of Shoebox can be found on the Shoebox web pages which are linked to from [1].

5. SEARCHING

There are two primary methods of searching for photographs in Shoebox. The first uses conventional full-text indexing of roll and photograph titles, text annotations and automatic speech recognition transcripts of audio annotations. The second method uses whole-image and region-based image retrieval to provide searching by visual similarity.

5.1 Text and Speech Indexing

While some digital cameras provide audio recording facilities, during our project the quality was not sufficient for adequate speech recognition performance. In addition to the notably poor audio quality of such in-built microphones, another problem was noted – users had a tendency to speak in an informal conversational manner with little attempt to facilitate automatic transcription. This has long been noted in the speech recognition community as a major source of recognition errors. Audio annotations were therefore provided using a close-talking headset microphone.

Initially, the Entropic Truetalk Transcriber [16] was used as a speech recognition engine. This is a relatively sophisticated speech recognition package capable of state-of-the-art performance on dictated, broadcast and even conversational speech, and it is for this engine that we report retrieval results. We constructed a trigram language model (using a 60k vocabulary) from 50 million words of the British National Corpus of spoken and written British English, along with appropriate acoustic models for British English. The out-of-vocabulary rate for the test-set used was 3.12%, which is a little lower than might be expected given the unpredictable and personal nature of a photo collection. It was considered unnecessary therefore to complement word-based transcriptions with open vocabulary recognition techniques such as phone lattices [3]. The benefits of such an approach would probably be outweighed by speed and scalability issues.

Since most of the data is relatively cleanly dictated into a good quality microphone, it was thought that a mass-market dictation package was worth considering. Subsequently, a version of the software compatible with the Microsoft Speech API (SAPI 4.0) has been developed. SAPI itself comes complete with a free Microsoft Speech Recognition engine for U.S. English dictated speech. After a sufficiently long speaker enrollment procedure, this proved to perform adequately for our needs. With a loss of some degree of accuracy, SAPI provides us with a lighter, more portable version of our application.

While our software allows users to correct transcriptions or simply add their own textual annotations, this feature was not used in the work we present here.

The transcripts were stemmed with a Porter stemmer [17] and indexed with a conventional inverted-file index with position information to allow phrase searching.

When the user enters a text query, the results are presented in a side panel. As the user types the query, any stopwords entered are highlighted in green, and any words not found in the index are marked in red.

5.2. Image Indexing

While annotation provides an obvious benefit for searching a collection of photographs, it does require some degree of user effort. By using speech input, we have tried to keep this effort to a minimum. As an alternative, we also provide visual searching via image content analysis. While this topic has generated a great deal of research interest in recent years, the difficulty of the general task has meant that a good solution has yet to be achieved in all but the most limited of domains.

Image searching is beset by a number of problems:

- What properties are suitable for image comparison?
- What distance measures correspond to human perception?
- How do we perform comparisons quickly?
- How do we formulate queries?

In the literature, both global and local image properties have been investigated. Global image properties are derived from the image as a whole, while local properties are derived from regions within the image.

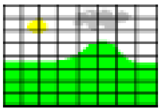
Whether local or global, the properties extracted typically include colour, texture, spatial location and shape properties, and there are many ways to characterise these properties. As more properties are included, it becomes easier to differentiate between the images or regions described, but the resulting feature vectors increase in dimensionality, making efficient indexing difficult.

There are many ways to compare feature vectors. The choice of distance function should reflect human perception, so that similar vectors correspond to perceptually similar images or image regions. Our work and others [9] have shown that non-metric distance functions can yield better results than metric distances such as the Euclidean distance. However, use of non-metric distances precludes use of multidimensional indexing structures such as the M-Tree [18]. This makes high speed retrieval somewhat difficult to achieve.

Query formulation is another tricky problem. Often the user must provide a starting image and request that the system finds similar pictures. If local image properties are used, the user may be able to select some regions from the starting image. Some other systems expect the user to draw a picture of what they are looking for. Unless the image collection is very simple, it seems rather optimistic to expect the user to manage anything but the crudest of input.

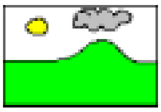
Recent work has shown that region-based indexing can perform as well as or better than global colour histograms [9]. In this work, region segmentation was used as an approximation to identifying objects within the image. We have considered similar segmentation schemes, but also consider regions obtained by simply dividing the image into a grid and computing image properties for these regions. We hope to discover whether the computational expense of a sophisticated segmentation scheme is justified by improved retrieval effectiveness.

For Shoebox, we have limited ourselves to computationally fast methods. We have experimented with a number of segmentation schemes, ranging from the extremely simple to more elaborate methods.



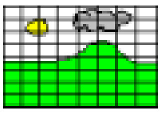
Grid segmentation

This scheme simply divides the image into 64 regions in an 8 by 8 grid.



Voronoi segmentation

Colour-based segmentation [19] on an image thumbnail.



Voronoi grid segmentation

The output of the Voronoi segmenter is further segmented using an 8 by 8 grid.

We have compared two indexing schemes for the extracted feature vectors. The first uses an M-Tree with a Euclidean distance measure. The second uses a method akin to conventional text indexing, in which the feature space is quantised coarsely, and the corresponding values are used as 'words' in an inverted index. Although it is likely to suffer from problems due to the quantisation, the use of a B-Tree rather than an M-Tree makes retrieval as fast as conventional text-retrieval.

In addition to the familiar "Find Similar" query paradigm, we allow users to select regions from an image as the target of a search. An outline of the regions generated by our segmentation algorithm is placed over the image, and regions can be selected, perhaps to isolate just one object from within a photograph. This enables more directed queries than image searches using global image properties. The system can highlight those regions in the resulting images that match the query regions, giving the user some insight into why the images were retrieved.

6. EVALUATION

We have evaluated our image searching methods in the context of a personal photograph collection. While the image retrieval methods rely on low-level features and aim to find visually similar images, we believe that if image searching is to be at all useful to a user, it must retrieve semantically similar images, and so this is our measure of success or failure. Our image retrieval results should therefore be considered in the light of these rather stringent relevance criteria.

6.1 Test Collection

Our starting point for evaluation of Shoebox was to obtain a collection of annotated photographs. A personal collection of some 500 photographs taken between 1996 and 1998 were scanned from APS film. This was augmented with photos taken with the photographer's digital camera. The total collection size was 575 images. Each photograph had an associated timestamp, which in the case of APS photographs was stored on a magnetic stripe on the film. Photographs were grouped together according to the film (or memory card) from which they were taken.

The photographer annotated the images resulting in 671 spoken annotations. The Shoebox software enables annotations to be made on single photographs or groups of photographs, and allows multiple annotations per image. In typical use, an annotation is made on an entire film roll (for example, "these photographs were taken on my holiday in France") as well as one or more individual annotations to describe each picture.

It is worth mentioning that the annotator is a fluent though non-native English speaker. The speaking style is semi-spontaneous; although largely planned speech, it is often littered with spontaneous remarks. These factors combine with an out-of-vocabulary (OOV) rate of 3.12% to explain a relatively high word error rate of 28.4%. Of the OOV words, 87% are proper nouns which include the names of family members and (non-British) place names. These are probably the sorts of words which might feature heavily in queries.

In addition to automatic speech recognition, the annotations were transcribed manually.

6.2 Text and Speech Retrieval

The retrieval task was to locate photographs by searching their associated transcribed annotations. For the retrieval experiments reported in this paper, we used a collection of 25 requests generated by two users, neither of whom was the photographer. Each request consisted of a natural language statement describing what would be considered as relevant, together with a set of keywords used as input to the search engine. For example, photographs relevant to the keyword query "Greg, Magda" had to show Greg or Magda. Pictures taken at Greg and Magda's wedding not containing either of these people were considered non-relevant.

The average number of keywords per request was 2.0. A relevance assessment subset was formed for each request by taking the union of the set of photographs retrieved by using ASR transcripts and the set of photographs retrieved when using manually transcribed annotations. The OOV rate for the queries was 2% and the queries had an average of 14.1 highly relevant photographs.

The results in Table 1 show text retrieval performance. Precision at ranked list cutoffs of 5, 10, 15 and 20 documents, and standard TREC average precision is given. The average precision of retrieval via ASR relative to manual transcripts is 71.9%. We recognise that with a small test collection specific figures are neither reliable nor significant.

Transcription:		**Manual**	**ASR**
Prec.	5 docs	0.8160	0.7727
	10 docs	0.6800	0.6136
	15 docs	0.5920	0.5303
	20 docs	0.5240	0.4568
Average precision		0.9443	0.6788

Table 1: Retrieval precision.

On the whole, annotations have proved to be a useful means of retrieving photographs, despite the inaccuracies introduced by speech recognition. Degradation in retrieval performance due to speech recognition errors is roughly comparable to that reported in [2]. However, during our experiments, some problems were evident. While roll annotations were often used to label related groups of images (e.g. "These photos were taken while on holiday in Poland"), they were sometimes misused. In particular, for those photographs which had been scanned from APS negatives it was frequently the case that the last few photos belonged to a different context. Thus the roll annotation did not apply to the last few photos, and while the annotation was relevant to the query, the corresponding photograph was a false hit during retrieval. With the use of digital cameras, which may hold a large number of images on a single memory card, it may be the case that users will make an effort to group logically related photographs into rolls.

While annotations usually proved to be good descriptions of the photographs, sometimes the rambling nature of the spoken annotations would produce something like "This whole roll is about Japan. Well… no it's not actually about Japan…". The photographer had not thought to delete the annotation and start again!

6.3 Image Retrieval

The aims of our image retrieval experiments were to compare the effects on retrieval performance of the different segmentation and indexing schemes, and to compare retrieval using selected image regions with retrieval using the entire image.

To compare the effects of the segmentation schemes independently of the effects of feature vectors, it is necessary to choose a feature vector which can be computed for regions output from any of our segmentation schemes. We have chosen a very simple feature vector, which represents the average colour in HSV colour space, together with variances in each of the colour channels as a coarse measure of texture, and the size of the region relative to the image size.

On average, the Voronoi, grid and Voronoi-grid segmentations produced 26, 64 and 167 regions per image. For each segmentation scheme, the two indexing techniques described in Section 5.2 were used. We established a baseline for the retrieval performance as follows. Randomly generated feature vectors were assigned to images and retrieval performance was computed using the test query set. The hope was that all our image retrieval techniques would out-perform this baseline.

We also compared the image content-based retrieval techniques to retrieval by date, in which the retrieved set is those photograph taken on the same day. Finally, conventional text retrieval was used, where the query was generated automatically from the query image by taking the (unweighted) set of all words appearing in ASR transcripts of the associated annotations.

A set of 25 images were selected at random from the collection. A brief description such as "child in a green dress" or "wedding scene" was given to each query image, which represented what the searcher was looking for. These descriptions were used to determine whether results matched. A relevance assessment subset was formed for each request by taking the union of the relevant photographs from the top 25 retrieved by each method. The queries had an average of 21.0 relevant photographs.

The results in Table 2 show retrieval performance of the various methods. Photograph retrieval by image content was disappointing. While all methods performed better than the 'random' benchmark, none came close to simply retrieving photographs taken on the same day. The annotation-based retrieval method performed best of all.

In all cases, M-Tree indexing out-performed B-Tree indexing, showing that the retrieval was not robust to problems incurred by quantisation.

Surprisingly, image segmentation seemed to produce little or no benefit over a simple grid. We have identified two reasons for this.

Indexing method	Segmentation	Average precision
Full-text	N/A	0.44
Date	N/A	0.38
M-Tree	Voronoi grid	0.29
	Grid	0.28
	Voronoi	0.24
B-Tree	grid	0.23
	Voronoi grid	0.23
	Voronoi	0.20
Random	N/A	0.13

Table 2: Average retrieval precision for 25 whole-image queries.

Firstly, unlike the grid segmenter, Voronoi segmentation tries to isolate perceptual image regions. However coarsely, the grid feature vector models the co-occurrence of adjacent perceptual image regions at their boundaries.

Secondly, the Voronoi-segmented image represents each region with a single feature vector. In contrast, the grid segmenter models this with parameters proportional to the area of the region. This makes partial matches more likely; *part* of a large region might be match another image.

The combined segmentation scheme (Voronoi grid) was intended to address the second issue, while retaining the isolation of perceptual regions. Our results show that there is little improvement in retrieval performance over grid segmentation for the Voronoi-grid scheme, despite a 38% increase in the number of feature vectors. There is certainly not enough improvement to justify the extra computational cost.

We may argue that we are doing the Voronoi segmentation an injustice by using such simple image features. From a Voronoi segmented image we might try to extract shape properties, which would be impossible from the uniformly rectangular regions produced by the grid segmentation scheme. However, it is unlikely that shape would be a useful property for retrieval when dealing with the complex images found in photographs, as shape varies with (three dimensional) orientation. Shape properties would also be corrupted by under-segmentation, in which regions as perceived by a human are merged with their neighbours, and over-segmentation, in which a single region is segmented as more than one region.

One important feature that we have elected not to use is the spatial position of image regions. Use of position has its pros and cons. Absolute position is sometimes useful (e.g. sky is usually at the top of a picture), but a simple translation of an object within an image should not render it irretrievable. Relative position appears to be more useful (e.g. the constituent regions of a person should appear more-or-less together in retrieved images) but mirroring an image should again not cause it to be considered dissimilar. For these reasons, we have not yet included spatial position in our feature vectors.

Indexing method	Segmentation	Search Type	Av. Prec.
Date	N/A	N/A	0.35
Full-text	N/A	N/A	0.33
M-Tree	Grid	Partial	0.29
	Voronoi grid		0.28
	Voronoi		0.26
	Grid	Whole	0.30
	Voronoi grid		0.30
	Voronoi		0.27
Random	N/A	N/A	0.09

Table 3: Average retrieval precision for 6 image queries.

Of the 25 query images, six contained an object or objects which were the target of the search (e.g. “child in a green dress”) while the rest did not (e.g. “wedding scene”). These six images were then used as partial-image queries, in which only the regions making up the object of the search were selected. These queries had an average of 21.3 relevant photographs. Table 3 presents the average precision results over the six queries. It compares three partial-image M-Tree indexed segmentations methods with date, full-text, random and the corresponding whole-image searches for the six queries. Again, it must be noted that, given the very small number of queries the figures presented can be neither reliable nor significant.

The same general trends are true of partial-image queries as for whole-image ones. Voronoi-grid and simple grid segmentations perform almost equally well, and both do better than the Voronoi segmentation scheme alone. However it is striking that *better* results are achieved by selecting the whole image than by selecting what the searcher considered to be the relevant parts of the image. This is easy to explain. We already know that simply selecting images taken at around the same time as the query image out-performs all our image content-based methods. This is because these photos are usually taken of the same thing, or in the same place. Just as the object of the photograph does not change, likewise the background setting remains the same. This means that the whole-image query has more data on which to match than the partial-image query. While it is possible to imagine situations in which this is not so, it appears to be the case in the personal photograph collection which we have used, and we presume it also to be true in other such collections. This contrasts sharply with the stock photo collections usually used to illustrate image retrieval, in which the photo collection might contain a hundred images of polar bears in different settings. Indeed, partial-image searching has been shown to perform better than whole-image searching in this situation [9].

For the management of personal photograph collections, retrieval both by date and annotations out-performed visual-based retrieval. We conclude that visual-based retrieval tools may not be especially valuable.

8. FURTHER WORK

We are currently conducting user trials with Shoebox. Around a dozen members of our laboratory have been supplied with digital cameras and the Shoebox software. The software logs user operations, thus allowing us to determine which facilities are actually used.

It is quite possible that users may not be willing to annotate images and may never even wish to perform a search. We are therefore looking at contexts other than personal photo collections in which to try Shoebox. The Hamilton-Kerr Institute, one of the biggest art restoration centers in the U.K., has a collection of cross-sections of paint samples taken from fine art. These cross-sections can be interpreted by an expert to judge similarity of technique and materials between artists and paintings. An initial investigation has revealed that the sort of image queries possible with Shoebox would prove to be useful in this context, and we are now designing a study to test this hypothesis.

Finally, our somewhat negative conclusions about the utility of visual-based photo retrieval must be taken with a grain of salt. There is still a great deal of work left to be done in this area. We intend to conduct further experiments with variations on our current techniques, and are also beginning to investigate machine learning techniques for classifying image regions. The latter will allow automatic derivation of some higher-level semantic image properties while avoiding the intractability of general scene understanding.

REFERENCES

1. *DART Project*, AT&T Laboratories Cambridge, http://www.uk.research.att.com/dart/.

2. G. J. F. Jones, J. T. Foote, K. Sparck Jones and S. J. Young.. *Retrieving Spoken Documents by Combining Multiple Index Sources*. Proceedings of the ACM SIGIR Conference, Zurich, Switzerland, 1996.

3. S. J. Young, M. G. Brown, J. T. Foote, G. J. F. Jones and K. Spärck Jones. *Acoustic Indexing for Multimedia Retrieval and Browsing*. Proceedings of the International. Conference on Acoustics, Speech and Signal Processing ICASSP 97, 1:199–202. Munich, April 1997.

4. *CompuPic*. Photodex Corporation. http://www.compupic.com.

5. *ThumbsPlus*. Cerious Software Inc. http://www.thumbsplus.com.

6. *Show&Tell: A Multimedia System for Semi-Automated Image Annotation and Content-Based Retrieval*. http://www.cedar.buffalo.edu/ShowTell/.

7. R.K. Srihari et al. *Multimedia Indexing and Retrieval of Voice-Annotated Consumer Photos.* Proceedings of the Multimedia Indexing and Retrieval Workshop, SIGIR'99, pp 1-16, 1999.

8. A. Kuchinsky et al. *FotoFile: A Consumer Multimedia Organization and Retrieval System*. In Proceedings of the CHI 99 conference on Human factors in computing systems, pp 496-503, 1999.

9. C. Carson, M. Thomas, S. Belongie, J. M. Hellerstein and J. Malik. *Blobworld: A system for region-based image indexing and retrieval.* Third International Conference on Visual Information Systems, June 1999.

10. N. R. Howe. *Percentile Blobs for Image Similarity*. Proceedings of the IEEE Workshop on Content-based Access of Image and Video Libraries, pp 78-83, June 1998.

11. A. G. Hauptmann and M. J. Witbrock. *Informedia: News-on-Demand Multimedia Information Acquisition and Retrieval*. In Intelligent Multimedia Information Retrieval, Mark T. Maybury, Ed. AAAI Press, pps. 213-239, 1997.

12. T. J. Mills, K. Moody and K. Rodden. *Cobra: a new approach to IR system design*. Proceedings of RIAO'97, Montreal.

13. H. Turtle. *Inference Networks for Document Retrieval* Ph.D. dissertation. CIIR, University of Massachusetts, 1991.

14. E. Brill. *Some Advances in Transformation-Based Part of Speech Tagging*. Proceedings of the Twelth National Conference on Artificial Intelligence, 1994.

15. M. Sanderson and B. Croft. *Deriving concept hierarchies from text*. In Proceedings of the 22nd International Conference on Research and Development in Information Retrieval (SIGIR'99). August, 1999.

16. V. Valtcho, D. Kershaw, and J. Odell. *The Truetalk Transcriber Book*, Version 1.0, Entropic Ltd. http://www.entropic.com.

17. M. F. Porter. *An algorithm for suffix stripping.* Program, volume 14 no. 3 pages 130-137, July 1980.

18. P. Ciaccia, M. Patella, and P. Zezula. *M-tree: An Efficient Access Method for Similarity Search in Metric Spaces*. Proceedings of VLDB, 1997.

19. D. Sinclair. *Voronoi seeded colour image segmentation.* Technical Report 1999.3, AT&T Laboratories Cambridge, 1999.