Using Text Search for Personal Photo Collections with the MediAssist System

Neil O'Hare[1], Cathal Gurrin[1], Gareth J.F. Jones[1], Hyowon Lee[1], Noel E. O'Connor[1,2], Alan F. Smeaton[1,2]
[1]Centre for Digital Video Processing, Dublin City University, Ireland
[2]Adaptive Information Cluster, Dublin City University, Ireland
nohare@computing.dcu.ie

ABSTRACT

The MediAssist system enables organisation and searching of personal digital photo collections based on contextual information, content-based analysis and semi-automatic annotation. One mode of user interaction uses automatically extracted features to create text surrogates for photos, which enables text search of photo collections without manual annotation. Our evaluation shows that this text search facility is effective for known-item search.

1. INTRODUCTION

The MediAssist [6] project at the Centre for Digital Video Processing (CDVP) has been developing tools for efficient searching of personal photo archives using contextual metadata (e.g. time, location), content-based analysis (e.g. face detection) and semi-automatic annotation. Location metadata for indexing photo collections has previously been explored, e.g. [5][8]. The PhotoCompass system [5] allows for location and other contextual features to be associated with photos for later retrieval. Ahern et al [1] use context to recommend recipients for sharing photos taken with a context-aware phone, although their system does not support retrieval. MediAssist makes significant contributions beyond this work: it provides tools for text-based searching using the results of automatic analysis, which is the major contribution this paper; and it incorporates tools for semi-automatic person annotation using a combination of content and context analysis to suggest people's names.

2. PHOTO INDEXING

The MediAssist research photo archive contains over 17,000 photos, all of which have time (from the EXIF header) and location (using techniques described in [8]) information, and all of which are indexed using the methods described below.

2.1 Context-Based Photo Indexing

- Time of capture is augmented to include a number of time-based fields: year, month, day of month, day of week and hour, enabling queries such as, for example, 'all photos taken in the summer, at the weekend' [6].
- Latitude/longidude co-ordinates are converted into placenames using a publicly available gazetteer [6].
- Standard astronomical algorithms calculate the light status of all images (i.e. day/night/dusk/dawn) [6].
- Each photo is annotated with weather data from the nearest weather station when the photo was taken [6].
- Indoor/outdoor classification is inferred from digital camera metadata, such as the ambient light levels [6].
- We also segment personal photo collections into 'Events' collections by detecting large temporal gaps between consecutively captured photos, similar to [3].

2.2 Content-Based Photo Indexing

- Faces present in images are detected [2].
- Having detected faces, the corresponding body-patch (which models the clothes worn by an individual during an event) and face recognition features are extracted to suggest identities for these faces [2].
- Large building objects in images are detected [6].

2.3 Semi-Automatic Annotation Tools

In the MediAssist interface, icons give feedback about the automatically annotated features described above. Clicking on these icons (for a single image or a group of images) corrects these annotations if the automatic annotation is incorrect. Undetected faces can be added and false faces removed, while body patch and face recognition features suggest names for detected faces. These facilities allow users to refine the automatic annotations and add people names [6].

3. TEXT INDEXING

Text-search is widely deployed in information retrieval search engines, such as Google, and is intuitive and popular with users. To provide text-based search in the MediAssist system, the automatic context and content-based features described in Section 2 are mined to construct text surrogates for all photos, creating a textual representation of each feature associated with a photo. An example text

SAC '07, March 11-15, 2007, Soeul, KOREA

surrogate could be 'dublin ireland september saturday weekend afternoon person alan'. The text search engine uses the standard and proven BM25 text retrieval model [7]. The system interface presents a simple text search box to allow for the quick and easy formulation of text queries.

4. EVALUATION OF TEXT SEARCH

In this investigation we explore the accuracy of our simple text-based search interface for searching personal photos. We use a known-item search scenario: given a document, can we retrieve that document from a large collection? We assume that the user is looking for a known photo about which they remember some of the metadata features. We used a subset of 1,000 photos from the MediAssist collection to tune the BM25 parameters. A separate set of 1,000 photos was used as test queries to evaluate the system, in both cases searching the entire 17,000+ collection.

The queries were created automatically using a subset of the terms found in the text surrogate for that image. Some features are guaranteed to be correct (time, location, semi-automatic person identity). The following features are not guaranteed to be accurate: indoor/outdoor, weather, number of people, and presence of buildings. If a term is found in the surrogate it accurately describes that document for all except the potentially incorrect features listed above. In our experiments we use subsets of the terms to create queries, allowing us to explore the retrieval utility of different feature subsets, e.g. the time-based terms, the 100% accurate annotation terms. For each feature subset we run the 1,000 known item test queries, and calculate the Mean Rank and the Mean Reciprocal Rank. Mean Reciprocal Rank is the favoured evaluation measure for known item search because it is not sensitive to the number of document, nor is it severely influenced by target documents retrieved at large ranks [4]. We also use Mean Rank because it is more intuitively understandable.

The results are shown in Table 1. Using the complete document surrogate as the query gives a mean rank of 12.8, showing that the terms used for indexing have powerful discriminatory power over our collection. Time alone performs very well, showing the power of this feature, although a user is unlikely to construct such a detailed query describing all possible temporal attributes of the image. Combining location with Month or Light Status gives a good improvement over location alone, showing the power of these features, and illustrating that combining location with simple time features can be very useful. Adding people to the location search gives a good improvement, but not so strong as one would expect, because only a subset of the images have people annotations, so this feature does not affect all queries. However, those queries that do have people annotations see a significant improvement in retrieval accuracy.

Query Terms Used	Mean Rank	Mean Reciprocal Rank
Complete Text Surrogate	12.85	0.41
All Location Features	77.59	0.19
Country + Continent	419.0	0.0004
State + City	77.59	0.19
All Time Features	20.23	0.25
Location + Time	18.68	0.28
Location + People	62.62	0.22
Location + Light Status	71.05	0.20
Location + Month	69.48	0.18
Location + Year	67.46	0.19
Year + Month + Light + Day Of Week	116.15	0.08
Indoor/Outdoor + Building + Weather	650.0	0.0007

Table 1: Mean Rank and Mean Reciprocal Rank for Known Item Search Using various document term subsets as the Query.

5. CONCLUSIONS AND FUTURE WORK

We have presented and evaluated of a simple, user-friendly, text-based search interface which uses automatically extracted features supplemented with semi-automatic annotation. Our results clearly show the strength of indexing by time and location in particular. For our future work we plan to conduct user evaluations of this system to test real users ability to formulate text queries and retrieve photos. We also plan to evaluate efficiency and usablity of the semi-automatic person annotation feature using real users.

Acknowledgments

The MediAssist project is supported by Enterprise Ireland under Grant No CFTD-03-216. This work is partly supported by Science Foundation Ireland under Grant No 03/IN.3/I361

6. REFERENCES

[1] S. Ahern, S. King, and M. Davis. MMM2: mobile media metadata for photo sharing. In *ACM Multimedia*, pages 267–268, Singapore, November 2005.

[2] S. Cooray, N. O'Connor, C. Gurrin, G. Jones, N. O'Hare, and A. F. Smeaton. Identifying person re-occurrences for personal photo management applications. In *VIE*, pages 144–149, Bangalore, India, September 2006.

[3] A. Graham, H. Garcia-Molina, A. Paepcke, and T. Winograd. Time as essence for photo browsing through personal digital libraries. In *ACM Joint Conference on Digital Libraries*, pages 326–335, Portland, USA, July 2002.

[4] P. B. Kantor and E. M. Voorhees. Report on the trec-5 confusion track. In *TREC-5*, pages 65–74, Gaithersburg, Maryland, USA, November 1996.

[5] M. Naaman, S. Harada, Q. Wang, H. Garcia-Molina, and A. Paepcke. Context data in geo-referenced digital photo collections. In *ACM Multimedia*, pages 196–203, New York, USA, October 2004.

[6] N. O'Hare, H. Lee, S. Cooray, C. Gurrin, G. Jones, J. Malobabic, N. O'Connor, A. F. Smeaton, and B. Uscilowski. Mediassist: Using content-based analysis and context to manage personal photo collections. In *CIVR*, pages 529–532, Tempe, Arizona, USA, 2006.

[7] S. E. Robertson, S. Walker, , S. Jones, M. M. Hancock-Beaulieu, and M. Gatford. Okapi at TREC-3. In *TREC-3*, pages 109–126, Gaithersburg, Maryland, USA, 1995.

[8] K. Toyama, R. Logan, and A. Roseway. Geographic location tags on digital images. In *ACM MultiMedia*, pages 156–166, New York, NY, USA, November 2003.