COMBINATION OF CONTENT ANALYSIS AND CONTEXT FEATURES FOR DIGITAL PHOTOGRAPH RETRIEVAL

Neil O'Hare[1], Cathal Gurrin[1], Gareth J. F. Jones[1], Alan F. Smeaton[1,2]

Centre for Digital Video Processing[1], Adaptive Information Cluster[2]
Dublin City University, Dublin 9, Ireland

Keywords: Digital Photograph management, image annotation, context, image content

Abstract

In recent years digital cameras have seen an enormous rise in popularity, leading to a huge increase in the quantity of digital photos being taken. This brings with it the challenge of organising these large collections. The MediAssist project uses date/time and GPS location for the organisation of personal collections. However, this context information is not always sufficient to support retrieval when faced with a large, shared, archive made up of photos from a number of users. We present work in this paper which retrieves photos of known objects (buildings, monuments) using both location information and content-based retrieval tools from the AceToolbox. We show that for this retrieval scenario, where a user is searching for photos of a known building or monument in a large shared collection, content-based techniques can offer a significant improvement over ranking based on context (specifically location) alone.

1 Introduction

Recent years have seen a revolution in photography with a move away from analog film towards digital technologies resulting in the accumulation of large numbers of personal digital photos. While storage devices offer ample capacity, the technology for managing digital photos has not kept pace with capture and storage advances. The MediAssist [5,9] project at the Centre for Digital Video Processing (CDVP) at Dublin City University is developing tools to enable users to efficiently search their photo archives. Our research utilises automatically generated contextual metadata for organising and searching personal photo collections. Key among this context data is location and date/time of photo capture.

This paper is organised as follows: Section 2 reviews current commercial and research tools for digital photograph management, Section 3 outlines our current demonstration system for context-only based search and describes our experimental digital photograph archive. Section 4 describes a system for combining location-based and content-based image search features, Section 5 discusses experiments and results, and Section 6 concludes the paper.

2 Existing Management Tools for Digital Photograph Archives

In this section we briefly review currently available tools for management of digital photo archives. We evaluated the functionality of 20 popular Windows-based tools and concluded that all perform the same basic management functions, managing photos using a photo album or folder metaphor and displaying thumbnails, and some providing calendar views to support organisation. We also concluded that there exists a strong reliance on users to manually annotate or categorise images, which is later used to support retrieval.

In addition to commercial products, there has also been an increasing amount of research in the area of personal photo collection management. Some systems attempt to leverage the techniques of Content-Based Image Retrieval to enhance the annotation process [12]. The MiAlbum system [15] uses a semi-automatic approach to image annotation. In addition, it is possible to exploit the 'bursty' patterns of photo capture to detect bursts of capture activity corresponding to an event such as a birthday party [3,6]. More recently, location metadata for indexing photo collections has been explored, with WWMX [13] allowing navigation of photo collections using a map-based interface, and the PhotoCompass system [8] allows for location and other contextual features to be associated with photos for later retrieval.

None of these systems combine context-based information with content-based analysis in an effective way, and this is what we address in the work presented in this paper.

3 Automatic Context Labeling of Digital Photo Archives

The basis for the MediAssist photo management system is the automated labelling of image context, thereby relieving the user of the need to manually label each image in their collection. The image context we refer to is the time and location of photo capture. The time of photo capture is stored by a digital camera when the user takes a photo and we capture location data using a separate GPS device. The integration of a camera with a GPS device provides us with a low-cost method of location-stamping digital photos by using the GPS devices tracklog capability. All the user has to do is to run an application when uploading photos that stamps the photo with the GPS location from the tracklog for the time the photo was taken [13].

Once a photo has been annotated with its date/time of capture and the location at which it was captured, it is possible to derive additional contextual information, such as 'daylight status', weather or indoor/outdoor classification [5,9]. Standard astronomical algorithms calculate sunrise/sunset times for any location on any date, enabling the annotation of a light status value for each photo (daylight/darkness/dawn/dusk). In addition, given that there are about 10,500 international weather stations all over the globe which constantly log weather data, we annotate each photo with the weather data from the nearest weather station at the time the photo was taken. Finally, Indoor/outdoor classification is inferred from metadata stored by the digital camera when taking a picture, such as the ambient light levels when a picture was taken, using an approach similar in spirit to that used in [2].

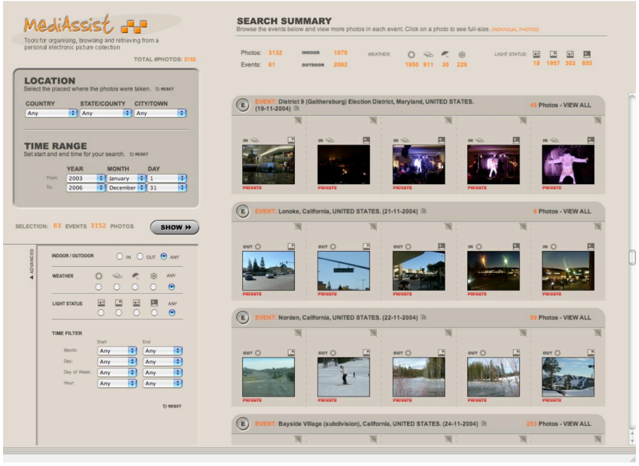
The key benefits of additionally labelling digital photos with their location are that it enables support of a number of access methodologies:

- Non-reliance on the user to manually annotate photos;
- The extended annotation of additional context information, as described;
- Search by location (county, town/city, even street);
- Search by proximity to a location or to other photos.

By using such information the browsing space (number of photos that a user has to browse) when seeking a particular photo or photos can be drastically reduced, thereby reducing the time taken to locate any given photo. We have shown by experimentation [5] with test users that the average time taken to locate a known photo using a location-aware system was just over half that needed for a time only (conventional) photo management system. In terms of query clicks, the average number of query iterations when using the location interface was also less than for time only.

3.1 The MediAssist Photo Management Tool

In order to evaluate our photo management techniques we have developed both desktop and mobile interfaces to the MediAssist tools. The mobile interface is described elsewhere [5]. The MediAssist desktop interface is shown in Figure 1.



Tools for organising, browsing and retrieving from a personal electronic picture collection

TOTAL #PHOTOS: 3152

Browse the events below and view more photos in each event. Click on a photo to see full-size. [INDIVIDUAL PHOTOS]

Photos: 3132 INDOOR 1070 WEATHER: 1950 911 30 229 LIGHT STATUS: 18 1957 302 855
Events: 61 OUTDOOR 2062

Select the placed where the photos were taken.

COUNTRY: Any STATE/COUNTY: Any CITY/TOWN: Any

Set start and end time for your search.

From: YEAR 2003 MONTH January DAY 1
To: YEAR 2006 MONTH December DAY 31

SELECTION: 63 EVENTS 3152 PHOTOS

INDOOR / OUTDOOR: ○ IN ○ OUT ● ANY

WEATHER: ● ANY

LIGHT STATUS: ● ANY

TIME FILTER — Month: Start Any, End Any; Day: Start Any, End Any; Day of Week: Start Any, End Any; Hour: Start Any, End Any

EVENT: District 9 (Gaithersburg) Election District, Maryland, UNITED STATES. (19-11-2004) — 45 Photos - VIEW ALL

EVENT: Lonoke, California, UNITED STATES. (21-11-2004) — 8 Photos - VIEW ALL

EVENT: Norden, California, UNITED STATES. (22-11-2004) — 59 Photos - VIEW ALL

EVENT: Bayside Village (subdivision), California, UNITED STATES. (24-11-2004) — 253 Photos - VIEW ALL

Fi

gure 1. **MediAssist System Interface**

The desktop interface presents the user with search options allowing the adjusting of location and time aspects of the query. More advanced search options are provided if the user required, allowing the user to specify weather (Sunny, Cloudy, Rainy, Snowy), light status (Dawn, Daylight, Dusk, Night), and Indoor/Outdoor for the photos. Advance time filters allow the user to specify particular date and time intervals corresponding to their partial recall of the temporal context of a photo-capturing event. For example the user could search for all photos taken in the evening, at the weekend, during the summer.

These features, particularly location, allow for efficient searching through archives of personal photographs [5]. The user can take advantage of their recollection of the time and place where a photo was taken to quickly form complex queries and find a particular photo or group of photos. For example a user can ask for all photos taken in Dublin, during the summer and at the weekend. Other details about the temporal context of the photo, such as the year or the day of the week, can also be specified, but can also be left unconstrained if the user has no memory of them.

By utilising time context of photo capture alone, photos can be grouped into logical events (as shown in Figure 1). These events exploit the bursty nature of photo capture, in which a user will take many photos for a given event, and then perhaps none for a while before another burst of photos at another event [3,6].

We have been using this photo management system within our research group and have collected over 11,000 photos taken by 16 users, an average of over 700 photos per user. These photos have been taken in 28 different countries representing 475 different locations within these countries. These photos comprise our test collection for the experiments outlined in this paper.

4 Integrating Content-Based Ranking with Location-Based Filtering

The type of context-based searching described above has been shown to be very useful when searching one's own personal collection [5], However, when searching across large archives of photos from many different users we are presented with a whole new set of challenges. For example, given a photo of the Chrysler building in Manhattan that I have taken, can I find more images of this building taken by other users? While this may not be a typical user request today, the increasing use of on-line photo management sites such as Flickr [4] and Phlog.net [10], coupled with new generations of high picture-quality camera-phones with integrated GPS functionality[1], makes it increasingly likely that such requirements will be commonplace in the near future. The obvious technique is to locate other photos taken at a similar location and, for a small collection of photos, a user may browse for and locate the desired photos. However when searching over a large shared library of personal photos, maybe millions, location ranking alone cannot provide an adequate solution.

The hypothesis presented in this paper is that the integration of photo-content analysis would improve the results of searching for similar photos of buildings or monuments across other peoples' shared archives of location stamped digital photos compared to searching using location information alone or content analysis alone.

Our proposed solution is to filter the collection based on location, and then to rank the photos based on content-based similarity to the user's own seed image, placing images that are more likely to match the query towards the top of the returned list. We use the AceToolbox [1] to create the content-based rankings, using the automatically extracted feature descriptors described below:

- Local Colour Descriptor (Colour Layout - CLD) is a compact and resolution-invariant regionalised representation of colour in an image. The feature extraction process consists of four parts; first, the image is partitioned in 64 (8x8) blocks; second, the representative colour of each block is determined by using the average colour in each block; third, a DCT transform is applied to these three (one for each of the colour components) tiny image icons of size 8x8, resulting in three sets of 64 coefficients; last a few low-frequency coefficients are selected using zigzag-scanning and nonlinearly quantized to form the CLD. The use of this descriptor supports matching photos that appear visually similar based on the colours occurring in the photos and the regions of the photos in which these colours occur.

- Edge Histogram Descriptor (EHD) is designed to capture the spatial distribution of edges in an image. This operates by dividing the image into 4x4 subimages (16 non-overlapping blocks) and then edges are categorized into 5 types (0°, 45°, 90°, 135° and "nondirectional") in each block. The output is a 5 bins histogram for each block, giving a total of 5x16 = 80 histogram bins. The use of this descriptor supports matching photos that appear visually similar based on the regionalised occurrences of edges in the photos.

- Homogenous Texture Descriptor (HDT) describes directionality, coarseness, and regularity of patterns in images. It is computed by first filtering the image with a bank of orientation and scale sensitive (Gabor) filters, and then computing the mean and standard deviation of the filtered outputs in the frequency domain. In this work we only use the mean values to compute the similarity between the images. The use of this descriptor supports matching photos that appear visually similar based on the textures that occur in the photos, such as grass and foliage texture in photos of the countryside.

We use two separate location-based approaches to filtering the collection prior to the content-based ranking. For the first approach all photos outside a certain radius of the query image are removed from the collection, leaving a subset of nearby images to be ranked. For the second approach all photos within the same city/town as the query image are included in the ranking. The city/town information is taken from the USGS [14] gazetteer.

In our experiments, we compare the results of using these content–based rankings with ranking based on the geographic distance from the query image. In addition, for combined approaches the retrieval scores of the different features were combined using the traditional CombSUM fusion method. This approach works by first linearly normalising each feature's scores between 0 and 1 and then combining these by summing (or equivalently averaging) respective document scores across features [11]. As yet, we have not investigated other IR fusion methods for this search task but it is noteworthy that the CombSUM fusion method performed very well compared to weight variants and Borda fusion methods for combining visual features for the TRECVid search tasks [7].

5 Preliminary Experiments and Results

We were interested in determining if the integration of photo-content analysis would improve the results of searching for similar photos of buildings or monuments across shared archives of location-stamped digital photos, compared to searching using location information alone or content analysis alone.

For preliminary evaluation we selected 6 query images, each corresponding to a known building or monument. These are shown in Appendix A. For each topic, the collection included relevant photos taken either by more than one user or during multiple photo-taking sessions by one user (i.e. separated by weeks or months), and these provided the ground truth for our evaluation.

[1] An example of a GPS enabled cameraphone is DoCoMo's F505iGPS camera phone.

	Colour Layout	Edge Histogram	Homogenous Texture	Location	CET	CETL
Location (200m)	0.584	0.356	0.5613	0.344	0.614	0.566
Location (500m)	0.589	0.355	0.5353	0.359	0.608	0.578
Location (1 km)	0.556	0.345	0.5269	0.387	0.631	0.615
Location (2 km)	0.546	0.341	0.5132	0.389	0.623	0.638
Location (5 km)	0.514	0.316	0.4933	0.389	0.590	0.60
Location (same city)	0.546	0.339	0.5215	0.358	0.572	0.562
All Photos	0.139	0.165	0.0788	0.389	0.260	0.367

***Table 1.* Mean Average Precision over all 6 queries. Rows represent method used to filter the data prior to ranking. Columns represent the ranking method used. CET – Colour Layout, Edge Histogram and Homogenous Texture combined. CETL – Colour Layout, Edge Histogram, Homogenous Texture and Location combined**

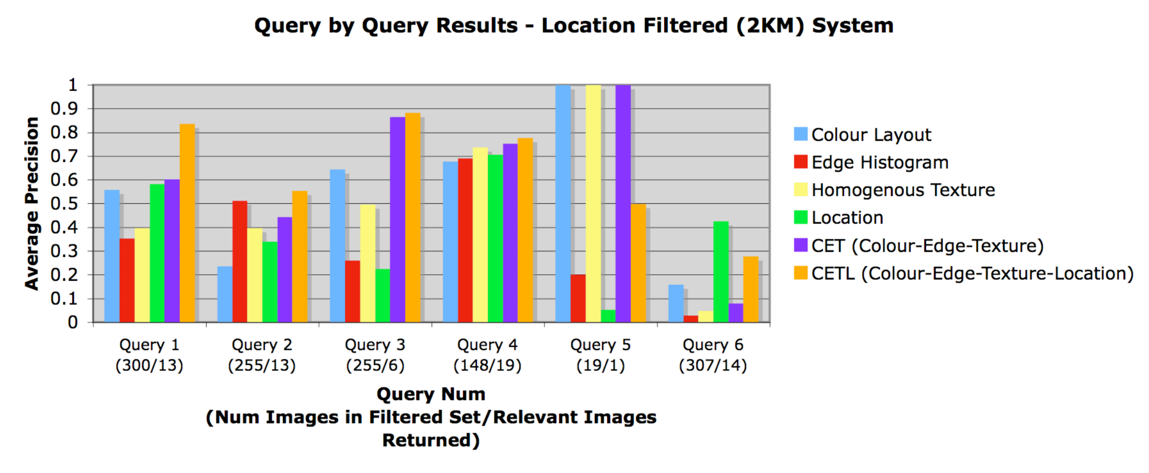


***Figure 2.* Average Precision for each ranking method for each query, using the location based filtered set, with all images further than 2KM from the query image deleted.**

Taking each of the six query images, we first filtered the set of images by location, either all photos within a certain distance (measured in metres) of the query image, or all photos taken in the same city/town as the query image. As a baseline we also ran all of the ranking algorithms on the unfiltered set of 11,203 images. Three of the ranking approaches are content-based approaches using image descriptors taken from the AceToolbox (Colour Layout, Edge Histogram and Homogenous Texture). The fourth ranking approach is based on the spatial distance between the capture locations of the images, which can be easily calculated using latitude and longitude co-ordinates.

It is worth noting that because we use a separate GPS device and camera and only store the location information at one minute intervals, the location data could not be considered 100% accurate and as a result some photos may be given the same location as the preceding photos. Also, GPS technology requires line of sight to satellites, so in indoor situations or in heavily built-up areas the signal can be lost for some time, again compromising the accuracy of the location information, at which point our software assumes the last known good location. While it is possible that integrated photo capture and GPS devices may help to solve the former issue, the latter issue will not be solved by integrating the two devices and we believe that the GPS data we currently gather is accurate with respect to currently available technologies. Using the cell ID from a mobile phone network provider will not solve these accuracy problems either, as this type of location information is not highly precise.

We ran each of the ranking approaches on each of the filtered sets to simultaneously evaluate the ranking methods and the filtering approaches. The results of these experiments are summarised in Table 1. It is very encouraging that all of the content-based descriptors outperform the location-based ranking, though we are aware that the number of query images used was small and additional experiments will be run to provide a more statistically reliable result.

As expected, the results of all three content-based ranking schemes are poor on the unfiltered collection (‘All Photos’ in Table 1), reflecting the fact that content-based image retrieval techniques struggle to discover semantic similarity between large collections of unrelated images. On the other hand, the combined approaches on the unfiltered collection do give a significant improvement.

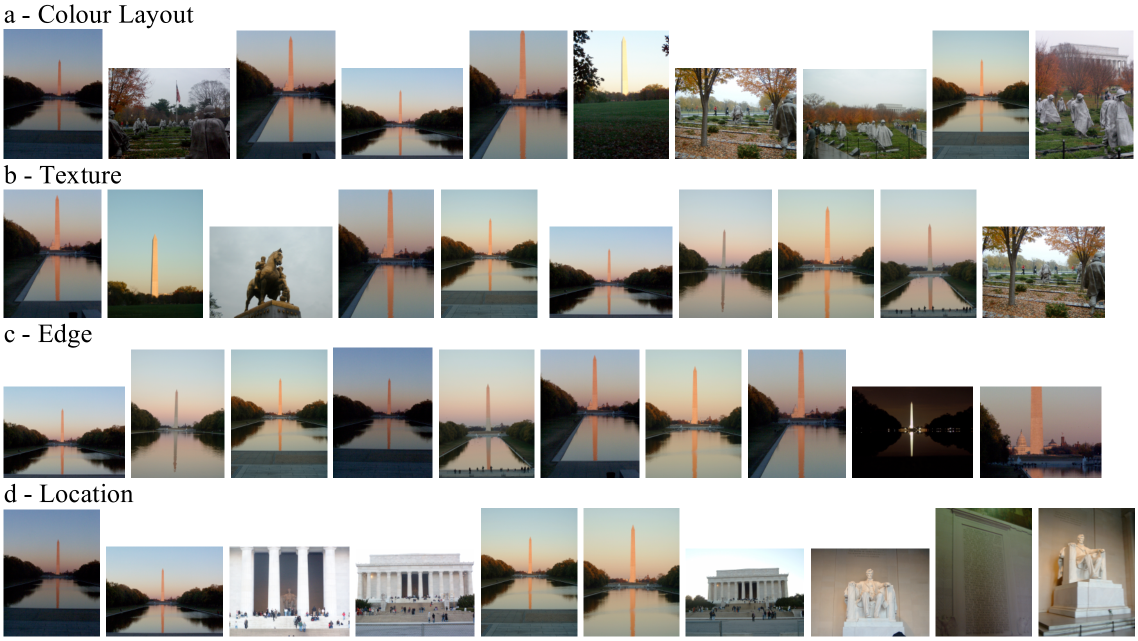


***Figure 3.* Top 10 rank lists for each ranking method, using 500 metre location-based filtering.**

By using the knowledge that relevant images must be geographically near to the query image and filtering the result set accordingly we have shown that we can achieve significant improvements in performance. Both Edge Histogram and Colour Layout outperform Location ranking once all photos further than 5km away are removed. The fact that MAP for location ranking peaks at the 5km filter indicates that, for this dataset, no relevant images were captured further than 5km from the query image. Filtering by photos taken in the same city is not as effective for this dataset as filtering based on distance from the query image.

We can see that the combined runs, CET (Colour Layout, Edge Histogram and Homogenous Texture) and CETL (Colour Layout, Edge Histogram, Homogenous Texture and Location), using the CombSUM approach, work particularly well and for all of the filtered sets the combined runs achieve an improvement over the best performing individual feature.

Figure 2 shows the query by query results from the filtered set (photos within 5KM of the query image) for all features and for the combined approaches. We can see that the location ranking is outperformed by both combined approaches for all queries but one. It is particularly encouraging that the CET combined run outperforms location consistently since this runs relies solely on content-based features and does not use location information in its ranking, but only for filtering. Also, for four of the queries, at least one of the individual content-based approaches outperforms location.

Figure 3 shows the top 10 results for each of the ranking method on Query 2, the Washington Memorial in Washington D.C., using the location-based filtering of the collection (photos within 500M of the query image). It can be seen that the location ranking returns some images (those at rank 3,4 and 7) which were taken from the same place but with the camera facing in the opposite direction. Also, the true location of the images at rank 8, 9 and 10 is over 100 metres from the location of the query image but the coarse accuracy of the location data means that these images make it into the top ten of the ranking.

6 Conclusions and Future Work

We have shown that for searches for known objects with a fixed location, such as buildings and monuments, it is possible to combine contextual information (i.e. location) with content-based image retrieval techniques in order by return a much higher quality ranked list of images than is possible with location information alone. As stated, this experiment is small in scale and it would be good to conduct some larger scale experiments.

Also, we will extend the approach to incorporate annotation into our system. This will allow text-based queries to initiate content-based searches: if a user labels a number of photos as the Chrysler building, for example, text based query can rank a filtered set of images based on similarity to photos with this label. This way a user could perform content-based searches without needing a seed image.

Acknowledgements

The MediAssist project is supported by Enterprise Ireland under Grant No CFTD-03-216. We are grateful to the *AceMedia* project for use of the AceMedia toolkit. This work is partly supported by Science Foundation Ireland under grant number 03/IN.3/I361.

References

[1] The AceMedia project, available at http://www.acemedia.org

[2] Boutell, M., Luo, J. "Beyond Pixels: Exploiting Camera Metadata for Photo Classification." *Pattern Recognition* 38(6): 935-946 (2005)

[3] Cooper, M., Foote, J. and Girgensohn, A. "Automatically organizing digital photographs using time and content." In *Proc. of the IEEE Intl. Conf. on Image Processing (ICIP 2003)* (Barcelona, Spain, Sept, 2003).

[4] Flickr, http://www.flickr.com/

[5] Gurrin, C., Jones, G., Lee, H., O'Hare, N., Smeaton, A.F., and Murphy, N. "Mobile Access to Personal Digital Photograph Archives." MobileHCI 2005, Salzburg, Austria, 19-22 Sept 2005.

[6] Graham, A., Garcia-Molina, H., Paepcke, A. and Winograd, T." Time as Essence for Photo Browsing Through Personal Digital Libraries." ACM Joint Conference on Digital Libraries. July, 2002

[7] McDonald, K., Smeaton, A.F. "A Comparison of Score, Rank and Probability-based Fusion Methods for Video Shot Retrieval." *Proceedings of the Fourth International Conference on Image and Video Retrieval (CIVR-2005), 2005.*

[8] Naaman, M., Harada, S., Wang, Q., Garcia-Molina, H. and Paepcke, A. "Context data in geo-referenced digital photo collections." In *Proc. of 12th ACM Conf. on Multimedia (MM'04)* (New York, NY, Oct., 2004).

[9] O'Hare, N., Gurrin, C., Lee, H., Murphy, N., Smeaton, A.F., Jones, G.J.F. "My Digital Photos: Where and When?" To Appear in *ACM Multimedia '05*, Singapore. (2005)

[10] Phlog.net http://www.phlog.net/

[11] Fox, E.A. and Shaw, J.A. "Combination of multiple searches." *Proceedings of the 2nd Text Retrieval Conference (TREC-2),* NIST Special Publications, 1994.

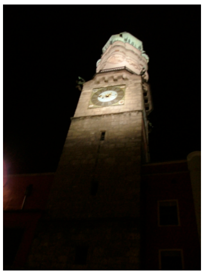
[12] Smeulders, A., Worring, M., Santini, S., Gupta, A. and Jain, A. "Content-based image retrieval at the end of the early years." *IEEE Trans. Pattern Analysis and Machine Intelligence, 22*, 12, 1349--1380, 2000.

[13] Toyama, K., Logan, R., Roseway, A., and Anandan, P. "Geographic location tags in images." ACM Multimedia 2003, New York, October 2003.

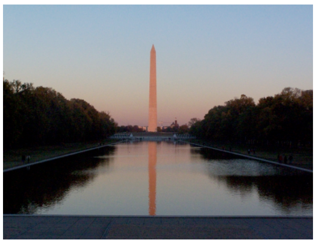
[14] USGS (U.S. Geological Survey). Website: http://www.usgs.gov/ Last visited Dec 2004.

[15] Wenyin L, Sun Y, Zhang H, "MiAlbum-A System for Home Photo Management Using the Semi-Automatic Image Annotation Approach," ACM Multimedia, 2000.

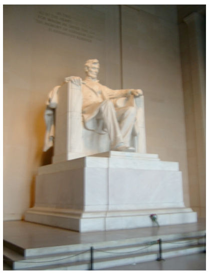
Appendix A. Query Images.



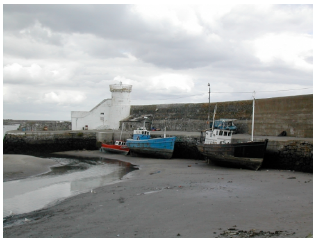
Query 1. Clock Tower, Innsbruck, Austria



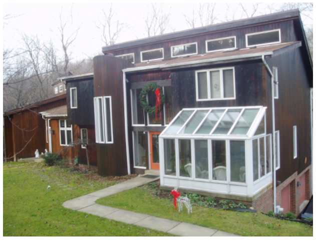
Query 2. Washington Memorial. Washington DC, USA



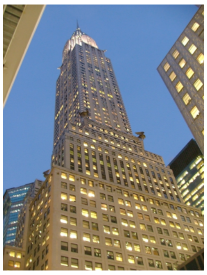
Query 3. Lincoln Memorial, Washington DC, USA



Query 4. Balbriggan Pier, Co. Dublin, Ireland



Query 5. Private Residence



Query 6. Chrysler Building, Manhattan, New York, USA Washington DC, USA