A Content-based Retrieval System for UAV-like Video and Associated Metadata

N.E. O'Connor, T. Duffy, P. Ferguson, C. Gurrin, H. Lee, D.A. Sadlier, A.F. Smeaton and K. Zhang

Centre for Digital Video Processing, Adaptive Information Cluster, Dublin City University, Dublin, Ireland

ABSTRACT

In this paper we provide an overview of a content-based retrieval (CBR) system that has been specifically designed for handling UAV video and associated meta-data. Our emphasis in designing this system is on managing large quantities of such information and providing intuitive and efficient access mechanisms to this content, rather than on analysis of the video content. The retrieval unit in our system is termed a "trip". At capture time, each trip consists of an MPEG-1 video stream and a set of time stamped GPS locations. An analysis process automatically selects and associates GPS locations with the video timeline. The indexed trip is then stored in a shared trip repository. The repository forms the backend of a MPEG-21[1] compliant Web 2.0 application for subsequent querying, browsing, annotation and video playback. The system interface allows users to search/browse across the entire archive of trips and, depending on their access rights, to annotate other users' trips with additional information. Interaction with the CBR system is via a novel interactive map-based interface. This interface supports content access by time, date, region of interest on the map, previously annotated specific locations of interest and combinations of these. To develop such a system and investigate its practical usefulness in real world scenarios, clearly a significant amount of appropriate data is required. In the absence of a large volume of UAV data with which to work, we have simulated UAV-like data using GPS tagged video content captured from moving vehicles.

Keywords: UAV, Information Management, Interaction Design

1. INTRODUCTION

When an UAV (Unmanned Arial Vehicle) captures a video stream during its flight, the resultant data is typically a time-stamped video file associated with a specific location. As the quantity of such data collected from multiple UAVs increases, managing and accessing the data becomes an obvious issue. With the potential for a huge amount of geo-temporal video data with other assocated meta-data, what will be the best way to automatically index, manage, and retrieve that best leverages the particular characteristics of this data? What kind of user-interface and visualisation techniques could best utilise this particular type of data to support efficient searching and browsing of such data? This paper contributes by describing our technical solutions to these questions.

Due to the lack of large amount of UAV data which is required for demonstrating the handling and visualisation of such data, we have collected equivalent data from land vehicle contexts. A number of GPS-enabled in-car video cameras were employed in order to generate an archive of trips by a number of users, in a period of 9 weeks in 2007. The collected data consists of video streams taken from the front window of cars on the road, time-stamped along with GPS locations at ten-second intervals.

Using this data, optimal mechanisms for automatic indexing and management of UAV data and efficient and easy-to-use user interface strategies were investigated, and a system that demonstrates our solutions was developed. The system takes in the captured video streams along with associated GPS and time information, and automatically indexes and stores for subsequent retrieval in an MPEG-21 compliant generic XML database. The front-end is a Web 2.0 interface that leverages intuitive map-based navigation and novel strategies for

Further author information: Noel E. O'Connor: E-mail: oconnorn@eeng.dcu.ie

Airborne Intelligence, Surveillance, Reconnaissance (ISR) Systems and Applications V
edited by Daniel J. Henry, Proc. of SPIE Vol. 6946, 69460I, (2008) · 0277-786X/08/$18 · doi: 10.1117/12.777527

synchronising and visualising the video stream data with geographic locations overlaid on top of the map interface which a user can access.

The paper is organised as follows. In Section 2, we briefly describe a usage scenario to provide a context and the research challenges in realising the scenario. In Section 3 we describe the overall architecture and its components of the developed system that make up the back-end processing, indexing and management of the data. In Section 4, our solution in terms of its user interface and visualisation is described in order to provide an intuitive and easy-to-use features for seamless and coherent searching, browsing, playback and annotation. Finally, Section 5 describes the possible transferrability of our solutions to true UAV video data and future development of this work.

2. USAGE SCENARIO AND DESIGN CHALLENGE

The system described in this paper supports the management of video data captured when multiple moving vehicles continuously captured their front/below view on their routes. For the purpose of implementation and data capture, we developed the system tailored for the in-car (land vehicle) video camera installed inside the front window. The users of the system have the GPS-enabled video camera installed in their cars. Whenever a user makes a trip with her car, she upload the recorded data (video and GPS data time-stamped together) to the system. Thus multiple trips from multiple users (cars) are uploaded to the system.

Once uploaded, the data is automatically analysed, visual abstractions generated, external information from web resources collected to enrich the metadata, and all this becomes ready for subsequent user access. The user can view her trip on the web-based interface (described in Section 4 in detail), review the trip with its video stream and make annotations at any point along the trip route or on the whole trip. As other users can view the trip, add comments/information/views at any point/whole trip or at the trip owner's comment, the information about the trip becomes richer and richer over time. Later on, the user can use the interface to search and browse the past trips of herself or others, play the video at any point of the trips, see other users' annotations on the trips and points, or add further annotations if wished.

Due to the mixed and temporal nature of the data (video stream with associated GPS location and annotations at any point along the way), two major challenges are faced:

1. Automatically indexing/managing the data so that it can be retrieved from useful access points (where the term *access point* here is related to the fundamental unit of retrieval corresponding to a "trip" – see section 4 for more details);
2. Providing effective user access for searching, browsing, playback, and annotation while ensuring the high usability in terms of efficiency, ease of learning and ease of use.

In the following section, we describe how we answered the first challenge, and in Section 4 we describe solutions to the second.

3. SYSTEM DESCRIPTION

The CBR system we describe in this paper is a web interface to a search and browse system for visual UAV trip data. Figure 1 shows the overall architecture of the system, with key components (and inter-relationships) displayed. We will describe briefly the functionality and construction of each of these key components, including the metadata stored in the system for each user and trip.

3.1. A Trip as the Unit of Retrieval

The data we are dealing with is multiple instances of temporally annotated video stream that we call a "trip", with the following characteristics:

- A trip is associated with a continuous video stream along its route;

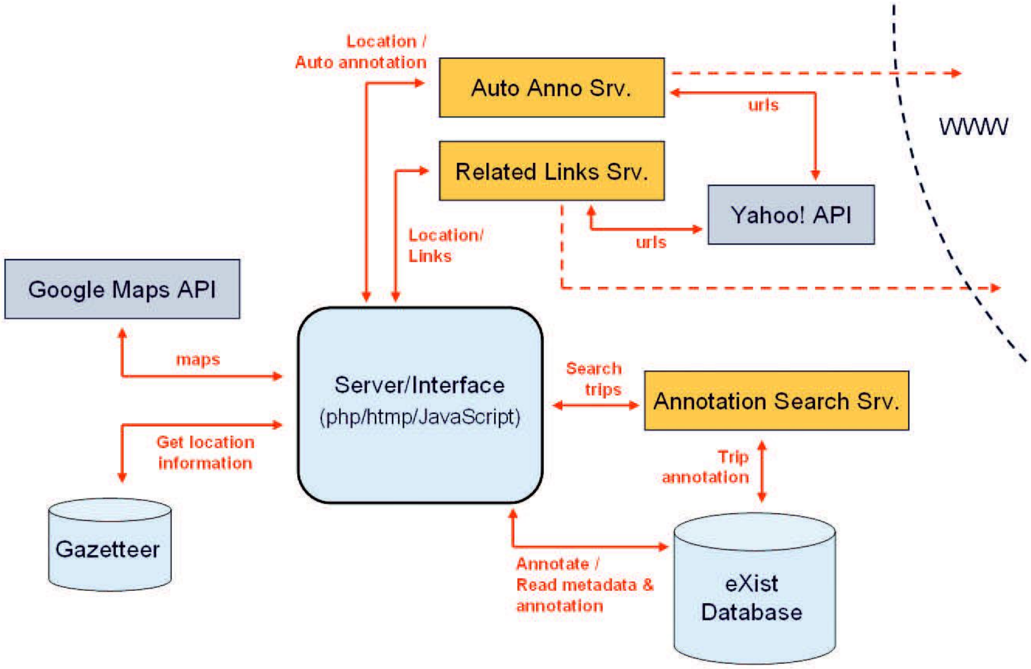


Figure 1. System Architecture

- A trip has a number of GPS points along its route (every 10 seconds in the case of our system);
- Each GPS point has a keyframe image extracted from the video stream at that point;
- Each GPS point has automatically-derived annotation (including location names, nearby landmarks, web links to related sites, etc.);
- Some GPS points have text-based, manual user annotation, and possibly a thread of such annotation (i.e. user annotation on another user annotation).

At time of writing, there are seventy six trips uploaded in the CBR system. For these trips, the average trip length was 28 minutes, with an average of 170 points per trip. In total this resulted in 35 Hours of video with storage space requirements of 12 GB .

3.2. Trip Metadata

As mentioned, every trip/route is stored as a MPEG-21 Digital Item, which in turn is composed of two parts:

- Digital Identifier - a URN for the digital item.
- Resource - refers to the MPEG-7[2] file which contains the trip annotations, video file reference, key frame references and other trip details, including; a reference to the MPEG-1[3]video file, details of each point on the trip, and a reference to the animated keyframe used to visually represent the trip.

Recall that a trip is composed of many points, and each point on a trip contains metadata for:

- Reference to the keyframe image for this point
- GPS coordinates
- Time details
- Both Automatic (see below) and Manual annotations for this point.

3.3. User Metadata

In our application scenario, we envisage multiple users accessing the content, albeit with different access rights and privilidges. Thus, in addition to the typically stored user metadata, every user with digital items (termed a *trip owner*) in the system has a license file limiting access rights to that user's digital items. This license file defines access rights to:

- Modify (digital item owner)
- Annotate (annotators)
- Print (all)

3.4. MPEG-21 Support

MPEG-21 support is a key feature of this system and has two main roles:

- Each trip/route is represented as a Digital Item in our XML Document Store (see Section 3.5 below).
- Access rights to the Digital Items is implemented as defined in MPEG-21. The requires license files are also stored in the XML database.

3.5. Databases

To store trip, user and access control data, an eXist XML database was employed, which allowed us to natively store MPEG-21 data for each trip. In addition a SQL Server relational database was employed to act as a gazateer lookup server, converting GPS points into placenames, from a large gazeteer of placenames. The system was programmed using a number of languages (JAVA for search servers, PERL and PHP/Javascript for the interface). A number of JAVA servers were employed to support the search and linkage Functionality.

3.6. Upload

Uploading of new trips was done by the user after a trip was complete. Initial automatic processing of video content and GPS logs indexed the trip data into the eXist database, generating automatically the metadata for the trip and the points on the trip automatically.

3.7. External Web Services

The two external web services employed were for mapping and external content linkage. To support the mapping functionality, we choose the Google MAPS API, which provided all the mapping functionality we required. The external content search and linkage was provided by the Yahoo! Search API. It should be noted that the system relies on these services for "added value" above and beyond basic functionality. However, the core system itself is self-contained and can operate without these services. Furthermore, the modular design and the use of APIs ensures that services could be swapped in or replaced in the future.

3.8. Annotation Search Server

The annotation search server was a conventional text search engine which operated over the manual and automatic annotations (see Section 3.9 below) of trip data. It operated as a TCP Server and runs as a background process, indexing and searching over the MPEG-21 data from the eXist database, along with any automatic or manual trip and point annotations. The search server employed the BM25 text retrieval model and included a stopword removal phase.[4] BM25 is a implementation of a probabilistic model of Information Retrieval. BM25 is one of the most effective statistically-based approaches and can be implemented efficiently using inverted files and as such was suitable for our deployment.

3.9. Related Links Server

The related links server would, for any given location, execute a WWW query to a well-known search engine (Yahoo!) using that search engine's API. The related links server would return three high quality links for any location that a trip passes through. These links are presented to the user as a source of additional information.

3.10. Auto Annotation Server

The automatic annotation server operates in a similar way to the Related links server in that automatic annotations are mined from WWW content for each landmark location on a trip however the automatic annotations are based on locating Wikipedia content for each location. The Wikipedia content is downloaded and indexed by the annotation search server and are searchable.

4. ACCESS - SEARCHING, BROWSING, PLAYBACK AND ANNOTATION

One of the main contributions of this paper is the ways in which the system provides access to the UVA-like video data and its associated metadata. In order to design the interaction mechanisms and visualisation techniques optimally, it is important to understand and utilise the specific characteristics of the data to be interfaced with the users. The data we are dealing with is the multiple instances of temporally annotated video stream that we call a ”trip”, whose characteristics are discussed in Section 3.1.

In this section we describe the overall interaction style, the visual representation of a trip, various access points for searching, browsing, playback and annotation interaction that the system provides. Using a Web 2.0 as its platform, the system is accessible on a conventional web browser. In order to maximise the map-based navigation, the system has been developed with the recommended screen resolution of 1920x1200 or higher, although smaller screens do accommodate its interface.

4.1. Map-based Interaction

Considering the trips to be visualised are geographically bound entities (i.e. a single thread of multiple GPS points make up a unique instance of a trip), map-based interface has been chosen for its overall interface. At the outset, having a navigable and zoomable interactive map as the main interface provides minimal learning for the users as the map interface is now already commonly used feature on the web with popular API-based map services such as Yahoo! Maps[*], Microsoft Maps[†] and Google Maps[‡]. Having an excellent mapping of the physical world onto a computer screen, a map-based interface naturally lends itself to intuitive and easy-to-use with high perceived affordance. Applications of these maps include in-car satellite navigation, GIS (Geographic Information System) interface, personal/social photo browsing,[5,6] travel arrangement[7] and many more.

The system uses Google Maps for its map navigation interface. In the designed interface, once a user logs in to the system a map occupies the whole screen with the trip information appearing/disappearing overlaid on top of the map whenever useful (see Figure 2). A user can navigate the map with the usual dragging action with cursor and zoom-in/out, as provided by Google Maps API. Overlaid on top of the map is the visualisation of the ten most recent trips the user uploaded, explained in the next subsection.

4.2. Visualising a Trip

The designed visual representation of a trip is illustrated in Figure 3. As multiple trips can be overlaid on a map at a given time, a trip is assigned a unique colour on the map for easy visual discrimination among different trips. Manually annotated points by a user are represented as a point with circle around it in order to easily differentiate from a point without user annotation, as the user-annotated points are those worth focusing and browsing. Each GPS point is numbered in order to provide a temporal cue as to which direction the trip progressed, and visible when the user moves her mouse cursor over a point. Clicking on any point on a trip will slide in a semi-transparent panel from the right edge of the screen on top of the map, displaying information about that point (detailed below). The maximum number of trips that can be displayed on a map at a given

[*] http://maps.yahoo.com/

[†] http://maps.live.com/

[‡] http://maps.google.com/

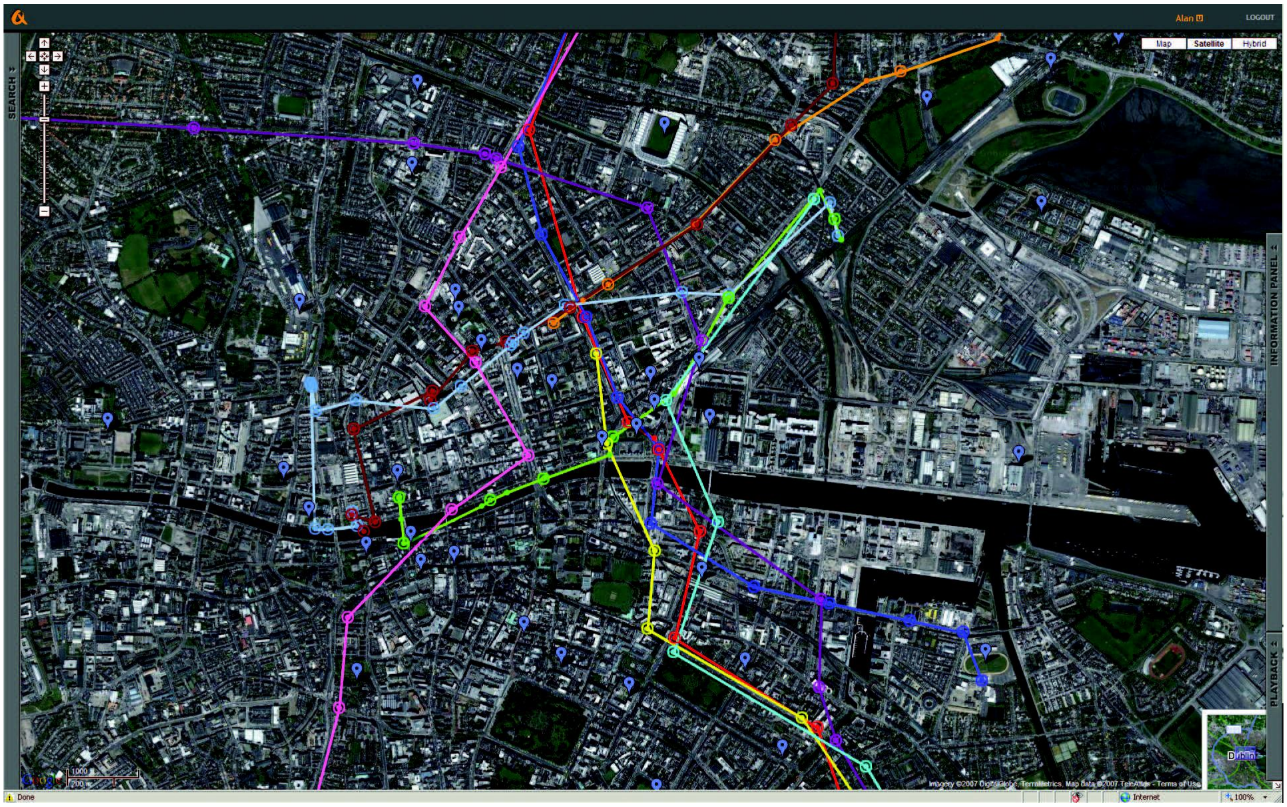

Alan
LOGOUT
Map
Satellite
Hybrid
SEARCH
INFORMATION PANEL
PLAYBACK
Dublin
Imagery ©2007 DigitalGlobe, Terrametrics, Map data ©2007 Tele Atlas - Terms of Use
Done
Internet
100%

Figure 2. Initial interface: map occupies the whole screen

time is ten for the following two reasons: 1) as the number of trips increases on the screen, potentially more and more trip routes overlap with one another and the colour-coding becomes less discriminative, thus become visually less meaningful and more difficult to get the sense of each trip; 2) displaying a large number of points takes time to load and thus becomes less and less acceptable for the user. By showing the groups of ten trips each time (ten most recent trips when browsing, ten best matched trips when searching), we circumvent these two problems. If the display/refresh rate becomes better in the future, the latter issue will be less of a problem and we will consider increasing the maximum number of trips on a screen.

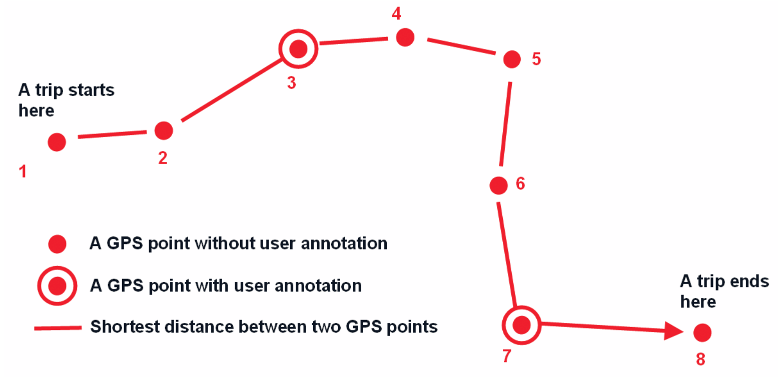

A trip starts here
1
2
3
4
5
6
7
8
A trip ends here
A GPS point without user annotation
A GPS point with user annotation
Shortest distance between two GPS points

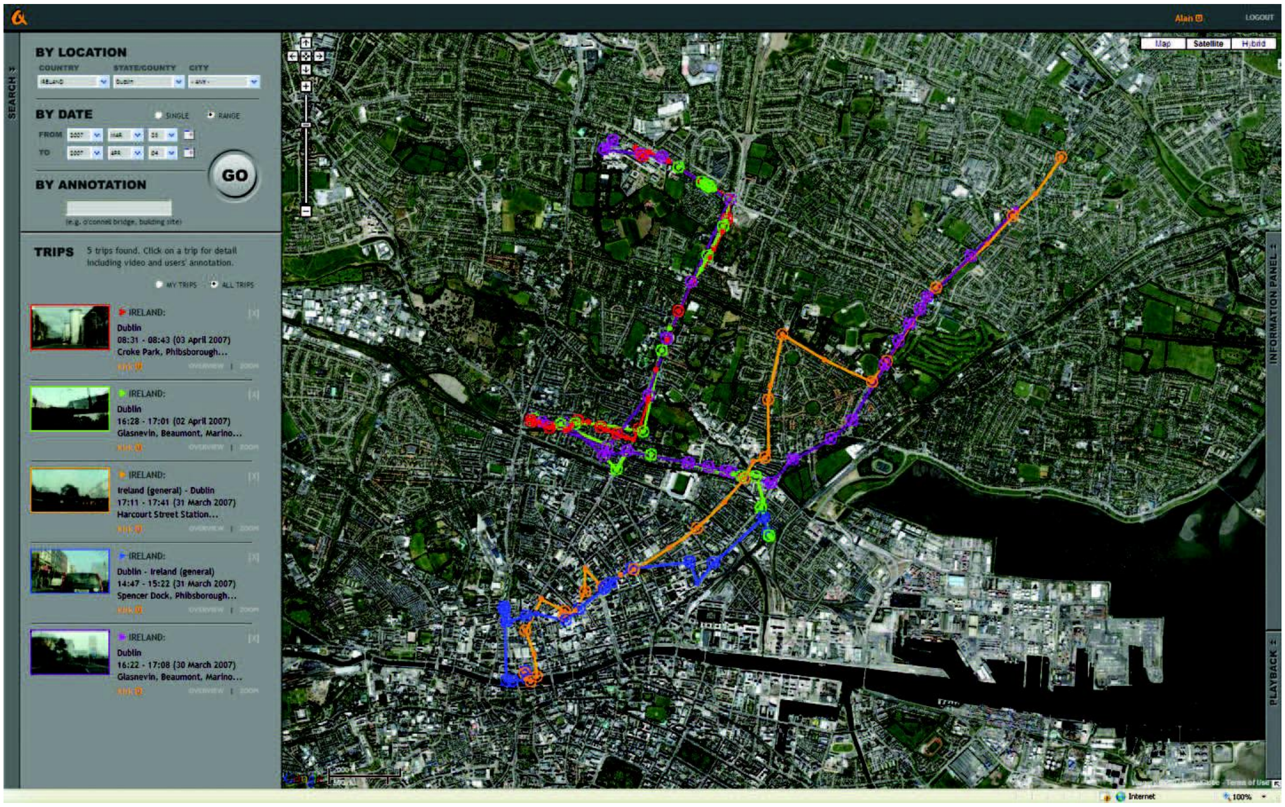
Figure 3. Visual representation of a trip

4.3. Searching and Browsing

While navigating the map with the trips overlaid provides a useful and intuitive mechanism for browsing the data in itself, the ability to search for specific trips provides a more powerful feature for the user to specify her specific query by one or combinations of the following:

- Date
- Location name (city, county, country, etc.)
- Region of interest on the map
- Annotation (both automatic and manual)

On the left edge of the map interface, there is a thin vertical search bar (see Figure 2). Clicking on this bar will slide in the search panel on the left side of the screen (see Figure 4) on top of the map.



BY LOCATION

COUNTRY: IRELAND

STATE/COUNTY: Dublin

CITY: - ANY -

BY DATE ○ SINGLE ● RANGE

FROM 2007 MAR 20

TO 2007 APR 04

BY ANNOTATION

(e.g. o'connell bridge, building site)

TRIPS 5 trips found. Click on a trip for detail including video and users' annotation.

○ MY TRIPS ● ALL TRIPS

IRELAND:
Dublin
08:31 - 08:43 (03 April 2007)
Croke Park, Phibsborough...

IRELAND:
Dublin
16:28 - 17:01 (02 April 2007)
Glasnevin, Beaumont, Marino...

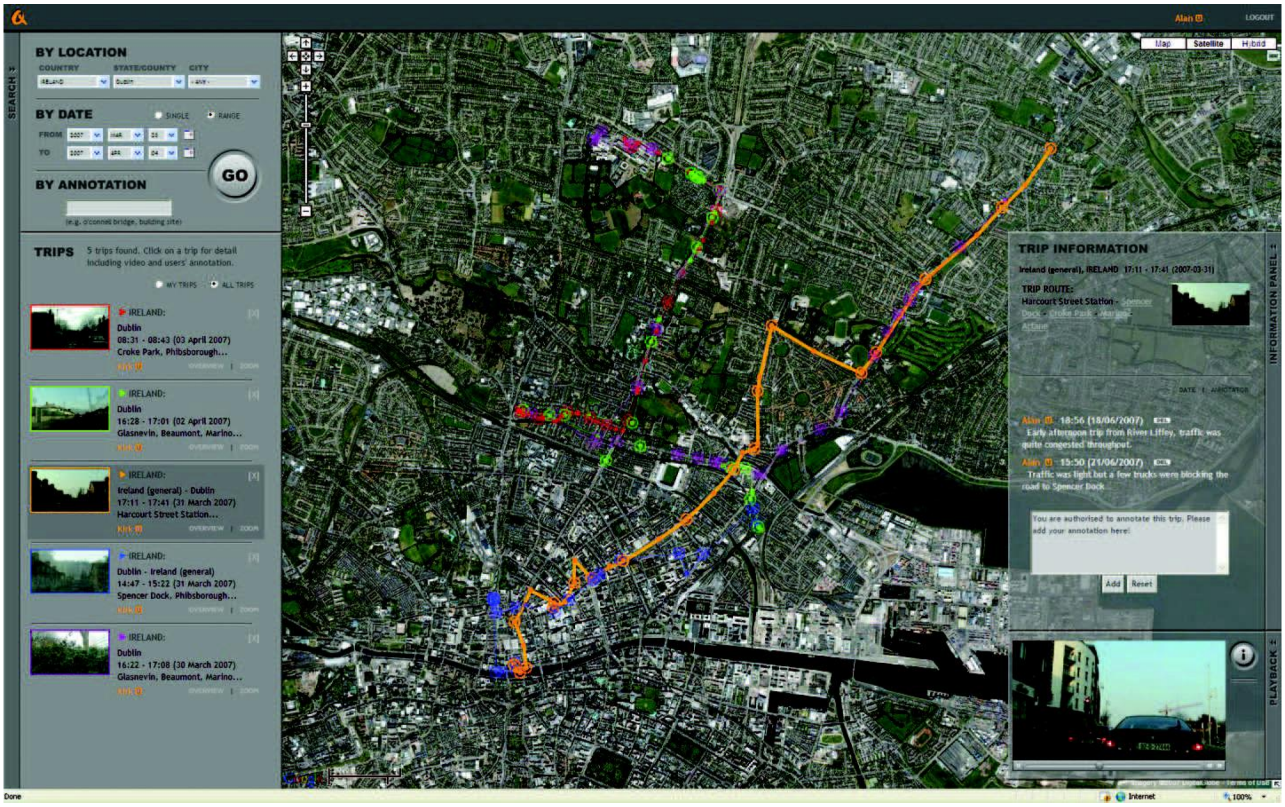
IRELAND:
Ireland (general) - Dublin
17:11 - 17:41 (31 March 2007)
Harcourt Street Station...

IRELAND:
Dublin - Ireland (general)
14:47 - 15:22 (31 March 2007)
Spencer Dock, Phibsborough...

IRELAND:
Dublin
16:22 - 17:08 (30 March 2007)
Glasnevin, Beaumont, Marino...

Figure 4. Search panel and search result

On the upper part of the search panel, the user can specify her query by either selecting country/city/county names from the drop-down boxes, by specifying a date/range of dates, or by typing in text to be matched against automatic/manual annotation text. In Figure 4, the user is searching for trips that happened in Dublin, Ireland, between 20 March and 4 April 2007. Clicking on the GO button triggers retrieval based on the combinations of the specified query, and the result is displayed at the lower part of the search panel (in Figure 4 five trips were retrieved). Each trip in the retrieval result is represented by an animated keyframe slide-showing keyframes extracted from the GPS points of that trip, along with location, date/time/annotation snippets. Along with the 5 entries on the panel, the visual representation of the corresponding trips are overlaid on the map. The user can now select one of the trips, either from the trip list on the search panel or directly a trip route on the map, upon which a semi-transparent panel slides in from the right edge of the screen displaying information about that trip including date/time/annotation, as well as the actual video clip playing from the start of the trip (Figure 5). The panel can be slided out at anytime by clicking on the thin bar on the right edge of the screen where the panel appeared from.



Country: IRELAND; State/County: Dublin; City: - ANY -

From: 2007 MAR 25; To: 2007 APR 04

5 trips found. Click on a trip for detail including video and users' annotation.

IRELAND:
Dublin
08:31 - 08:43 (03 April 2007)
Croke Park, Phibsborough...
Kirk

IRELAND:
Dublin
16:28 - 17:01 (02 April 2007)
Glasnevin, Beaumont, Marino...
Kirk

IRELAND:
Ireland (general) - Dublin
17:11 - 17:41 (31 March 2007)
Harcourt Street Station...
Kirk

IRELAND:
Dublin - Ireland (general)
14:47 - 15:22 (31 March 2007)
Spencer Dock, Phibsborough...
Kirk

IRELAND:
Dublin
16:22 - 17:08 (30 March 2007)
Glasnevin, Beaumont, Marino...
Kirk

TRIP INFORMATION

Ireland (general), IRELAND 17:11 - 17:41 (2007-03-31)

TRIP ROUTE:
Harcourt Street Station - Spencer Dock - Croke Park - Marino - Artane

Alan 18:56 (18/06/2007)
Early afternoon trip from River Liffey, traffic was quite congested throughout.

Alan 15:50 (21/06/2007)
Traffic was light but a few trucks were blocking the road to Spencer Dock

You are authorised to annotate this trip. Please add your annotation here!

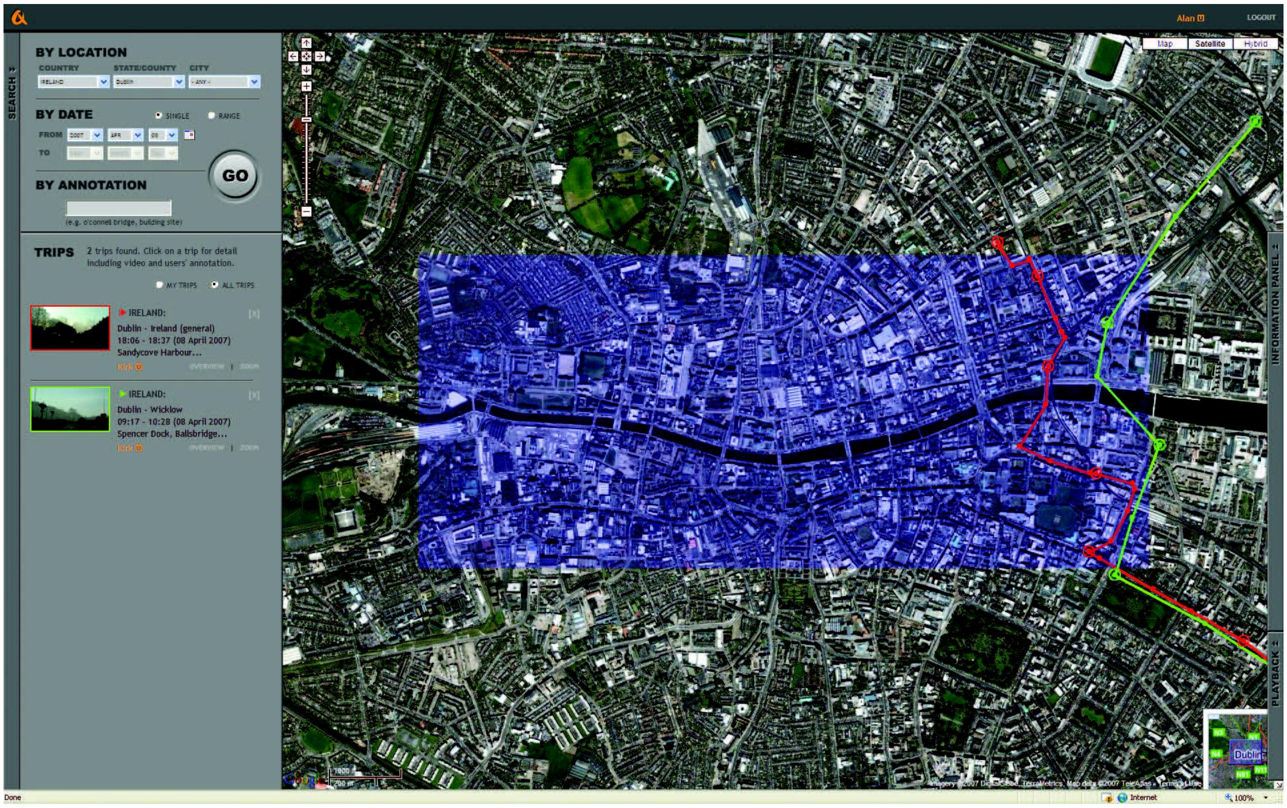
Figure 5. Trip information displayed on the sliding panel on the right side of the screen

As can be seen in Figure 5, the selected trip (in orange) is highlighted on the map and other four trips de-highlighted. On the trip information panel just slided in from right side, the user can see automatic annotation text (including city name, nearby-landmarks, and links to web resources), and all users' manual annotation text about the selected trip. The user can add her own manual annotation by typing in the text box and clicking on ADD button. Similarly, the user can further select a particular point on the trip (by clicking on the point on the trip route) to see information specifically on that point on the panel and video stream playing from that point onwards.

Another way of querying is by specifying a region on the map by dragging the mouse cursor on it. As shown in Figure 6, the user specified in the search panel Dublin, Ireland as location, 8 April 2007 for the date, then specified a rectangular region around the River Liffey by pointing top-left corner and dragging it down to bottom-right corner, which gives visual feedback as semi-transparent blue rectangle on the map. The system retrieved two trips that satisfy all these conditions and are displayed as two trips on the map (red and green trip routes) that pass through the specified rectangular region on the map.

4.4. User Annotation

While automatic annotation on trip level and point level happens off-line when the trip is uploaded to the system, the user can manually add text annotation on the trip as a whole or at a specific (GPS) point on the trip, during interaction time. Annotation adds value to one's own data as well as increases informativeness for other users collectively, its usefulness widely exploited in the "commenting" features of many popular Web 2.0 services today



BY LOCATION

COUNTRY: IRELAND
STATE/COUNTY: Dublin
CITY: - ANY -

BY DATE ● SINGLE ○ RANGE

FROM 2007 APR 08

BY ANNOTATION

(e.g. o'connell bridge, building site)

TRIPS 2 trips found. Click on a trip for detail including video and users' annotation.

○ MY TRIPS ● ALL TRIPS

IRELAND:
Dublin - Ireland (general)
18:06 - 18:37 (08 April 2007)
Sandycove Harbour...
Kirk

IRELAND:
Dublin - Wicklow
09:17 - 10:28 (08 April 2007)
Spencer Dock, Ballsbridge...
Kirk

Figure 6. Specifying a region directly on the map for searching

in a variety of domains (blogging, photo sharing, product review/recommendation, etc.). In the context of UAV applications, landmarks or any objects worth attention for a user can be located and annotated which can be viewed by other users and additional annotations could be added by more knowledgeable users.

While playing the video stream of a trip, the user can click on the 'i' button beside the player panel, which will pause the playback and the annotation text box opens up on the information panel above (see right side of Figure 5). In this way, at any point throughout the trip, the user can annotate where something interesting has been captured in the video (e.g. a building, road, etc.). On the information panel (right side of Figure 5), each user annotation entry identifies the user who made that annotation. If an annotation was made by the current user, she can edit/delete it by clicking on the EDIT button beside the annotation entry. Search panel, information panel and playback panel can be all individually slided in and out by clicking on the thin bar on the edge of each panel, giving the user control over what she wants to see and how a large proportion on the interface should be occupied by map and by panels.

5. CONCLUSION

In this paper we described an automatic indexing and management solution for data composed of video sequences stamped with geographic location and associated annotation at arbitrary points in the video, common to UAV-captured data. We developed an integrated system that supports intuitive and easy-to-use searching, browsing, playback and annotation for this type of data accessible on a Web 2.0 interface. While informal user testing showed that the user's activity of playing the video to locate a particular object of interest (such as a particular building, a bridge, or a traffic light) is well-supported and effective and allows a smooth transition to and from navigation of the map, the true UAV data will have video contents taken from ariel view whose contents/perspectives result in slightly different impact to the users who want to locate a specific object, artefact or geographic landmarks from the video playback. However, its continuous streams taken by multiple vehicles and associated geographic locations and multi-point annotations by multiple users make up the major characteristics of the data and we believe our solutions can be adopted to manage true UAV data. Future work includes obtaining real UAV sample data and incorporating it into the system in order to further tune its interface parameters more specifically tailored for the data, user evaluation to fix any usability problems introduced and to obtain specific points of improvement on the interface.

ACKNOWLEDGMENTS

The research and development reported in this paper was funded by QinetiQ. This work was carried out as part of the Electronic Systems Domain of the MoD Research Programme.

REFERENCES

1. H. Kosch, *Distributed Multimedia Database Technologies Supported by MPEG-7 and MPEG-21*, CRC Press, 2004.
2. B. S. Manjunath, P. Salembier, and T. Sikora, *Introduction to MPEG-7: Multimedia Content Description Interface*, John Wiley and Sons, 2002.
3. T. Sikora, "MPEG digital video coding standards," in *Chapter 2, Compressed Video Over Networks*, pp. 35–78, 2000.
4. S. E. Robertson, S. Walker, S. Jones, M. M. Hancock-Beaulieu, and M. Gatford, "Okapi at TREC-3," in *Proceedings of TREC-3*, D. K.Harman, ed., pp. 109–126, 1995.
5. K. Toyama, R. Logan, A. Roseway, and P. Anandan, "Geographic location tags on digital images," in *Proceedings of the 11th ACM Interactional Conference on Multimedia 2003*, pp. 156–166, 2003.
6. "Flickr - explore worldmap (http://www.flickr.com/map/)," 2008.
7. "Yahoo! travel (http://travel.yahoo.com/)," 2008.