Spatial Video and GIS

Paul Lewis[1], Stewart Fotheringham[1], Adam Winstanley[2]

[1]National Centre for Geocomputation/[2]Department of Computer Science, National University of Ireland Maynooth, Maynooth, Co.Kildare, Ireland;
Email: {paul.lewis, stewart.fotheringham, adam.winstanley}@nuim.ie

Abstract. GIS elemental unit representations of spatial data are often defined in terms of points, lines and areas. However, another type of spatial data that is becoming frequently captured, but as yet is largely ignored in GIS, is that of video. While digital video recording is a commonly encountered medium in modern society and encompasses many forms, from simple personal camcorders through to sophisticated survey and surveillance systems, its geographical representation in a GIS has not been fully examined or realised. In the majority of cases the video footage is usually captured while the device and/or the objects being viewed are in motion. What is of particular interest is when video streams can be, or have been, associated with spatial data such as location and orientation to create geographically referenced videographic data, which, for simplicity, will be defined as spatial video. Fundamentally, the nature of video is to record space, so when spatial properties can be accurately acquired and associated with this footage, an important geographical element can be considered for integration and analysis within a GIS.

Existing spatial video systems, both commercial and research, are predominantly used in survey or LBS roles and are usually bespoke and application specific (Kim et al. 2003a; Red Hen 2005; RouteMapper 2007) . These systems do not model spatial video to any recognised standards that is generalised to be both data and platform independent. They do not support GIS integration and/or analysis from a purely spatial content perspective. A video-image/remote-sensing centric approach prevails where usage options range from simple visualisation interfaces to interactive computer vision systems. What has been largely overlooked is a spatial approach where the inherent geographical extent recorded in each video frame can be modelled and used in a geo-spatial analysis context. While this modelling approach has not been fully realised, it does exist in a GIS form based on Open Geospatial Consortium (OGC) standards, where the spatial context of video is defined in a structure called a ViewCone (Lewis et al. 2006; OGC OWS-3 2005). However, a ViewCone only defines a 2D model of the geographical extent of each frame and is restricted to a three-or-five sided polygon representation.

Thus, this paper examines the potential of modelling spatial video through the use of elemental data types within GIS; gives some examples of using this approach; describes some problems in using spatial video within GIS; and then demonstrates how these problems are being solved. This is done in three stages. Firstly, a detailed overview of spatial video in its current GIS role is provided - this is achieved through a complete introduction to the distinct elements of spatial video followed by a review of its use in both commercial and academic application areas. Secondly, a brief theoretical overview of an alternative GIS-constrained ViewCone data structure is given that defines a more flexible spatial video model for both 2D and 3D GIS analysis and visualisation. Thirdly, a selective sample of results is presented based on an implementation of this approach being applied to a constrained spatial video data source in a specific study area.

Keywords: Spatial, Video, Viewshed, Segmentation, GIS.

1. Introducing Spatial Video

In general, videography is a well-understood concept that can be defined as the process and/or set of methods and operations used to capture a sequence of moving images (Kiger 1972). It has existed for many years and is ever-present in our daily lives in numerous capture and display formats. Spatial video is a specific extension of these video formats where spatial attributes are applied to some or all of the images/frames within the captured sequence. Multiple examples of this type of spatial data exist in various formats and modes of capture, storage and use; (Livingstone et al. 1999; Rzhanov et al. 2000; Ó Riain & McCarthy 2006; Foy et al. 2006; McLoughlin et al. 2005). In general terms, spatial attributes can include any number of different descriptors that can help define a video's image/frame location, time, altitude, orientation or other spatial attribute. In particular, the spatial video data used for this study is a specific form with the following properties:

- It has been captured from a road survey vehicle, travelling along a road network route corridor.
- The video is captured approximately orthogonal to the road surface and coincident with the direction of travel.
- Each video frame is spatially indexed with a set of GPS location parameters.

Spatial video acquisition methods vary by the types of sensors used; the forms of integration and recording of the video and spatial properties; the approaches used to correct, improve and synchronise sensors' differing data capture frequencies; and the intrinsic physical properties of the platforms on which these data are acquired. However, the collection and use of spatial video is typically a linear process that involves a number of operational stages: (i) Acquisition; (ii) Processing; (iii) Storage; (iv) Distribution; (v) Analysis.

Each of these stages typically requires further subsets of heterogeneous operations that have been developed using numerous different technologies and methods to handle specific objectives or applications. A typical example of a spatial video application area is aerial videography (McCarthy 1999). This, and other similar work, is a distinct commercial and research field in its own right, with many organisations and agencies using aerial-based spatial video to acquire planar views for their respective domains. However, research into the use of spatial video in a terrestrial-based context has been minimal and largely ignored.

2. Spatial Video Challenges

The existing inherent nature of spatial video is as a bespoke data acquisition solution in many mapping, survey and environmental analysis projects. This diverse need has resulted in a number of problems and challenges when trying to define a more general GIS model for spatial video, especially with previously captured data sets. Chief amongst these problems is a broad understanding of spatial video's place as a data source for a GIS. Typically, spatial video is collected for a specific reason, used to provide a particular solution and, very likely, never used again. Thus, its place has been to satisfy

an immediate need where no further usage or applications of the data are either conceived or implemented. This has resulted in enormous amounts of spatial video being collected and then stored in various distribution archives, in many formats, to largely differing (and undocumented) levels of visual and spatial detail and quality, with no further usage.

These problems centre on there being no centralised or generalised structure to index and search this amount of video data from a GIS perspective in a similar manner to that undertaken by Web Map Services or Web Feature Services. No common structure exists that defines spatial video and the sort of GIS-based analysis that could be performed on it. The large amount of retrospective spatial video that exists has no standardised video or consistent spatial data format. The video data, in terms of the inherent resolution properties of the recording system, the perceivable sharpness of captured objects and the framing semantics have no consistent index and metadata tagging methods. For example, video-image formats do not include capture parameter information about each frame in a similar manner to an Exchangeable image file format (Exif) tag used in digital camera imagery. Also, some existing bespoke systems include video in a semantically defined form where a file directory structure contains image sequences captured in a post-acquisition process; thus, this video representation does not conform to any existing video container format specifications. The quality and accuracy of both the spatial and video data are also very variable in terms of a surveys calibration parameters. These usually have to be retrospectively inferred in order that these data can be used in a context other than that intended by a bespoke system. Also, the GPS spatial accuracy contains positioning error ranges, which, if captured in the pre-2000 selective availability era could be even larger. There are no inherently supported formatting methods for linking these spatial and video data as records of spatial orientation; such as a frame's horizontal, vertical angle or its position in relation to any spatial data sensors. These also need to be empirically deduced from the imagery or defined during a system calibration stage.

The central (and obvious) factor encompassing all these spatial video formats is that video provides a visual perspective of the environment, where a more immersive experience of a scene can be achieved through the captured motion. This aspect has always defined the application areas for which video is used in a GIS and involves any of the following objectives:

- enhancing the GIS with recorded digital imagery of the cartographic environment (McCarthy 1999; Hirose et al. 1998; Kawasaki et al. 1999).
- using the visual information to generate other spatial data sets through supervised or unsupervised visual analysis of the image content (Ó Riain & McCarthy 2006; McLoughlin et al. 2005).
- using the video's geographical content and spatial parameters to segment or semantically describe the video (Hirose et al. 1998; Nobre & Camara 2001; Navarrete & Blat 2002).

Thus, spatial video has a number of specific approaches in its application and also numerous practical uses. However, a generalised, GIS-constrained, well-formed definition is lacking that describes how any spatial video sequence can be easily

modelled, based on its geography, for easy visual playback, content analysis, sequence segmentation and indexing.

2.1. Approaches to Modelling Spatial Video

Two approaches are considered towards defining the background to these challenges: one is based on the prevailing methods of modelling spatial video in existing commercial and academic applications; the other considers a GIS-based spatial data interaction environment.

2.1.1. Video-Centred/Remote-Sensing approach:

The first approach investigates existing video-based methodologies where spatial video surveys are self-contained units of video and spatial data. This is the standard approach to handling spatial video where individual frames or groups of frames are spatially indexed in an embedded or associated file format. Various video file formats and standards exist that are capable of storing and indexing video with metadata properties; however these are not designed specifically for spatial data, resulting in bespoke solutions. These solutions have the advantage of encapsulating the spatial video data through devices and platforms that can be managed in a predefined and well-calibrated fashion. This makes for a consistent and reliable solution where output video and spatial data streams are self-contained in well-understood formats that include all the relevant survey information in a single source location.

A disadvantage of this approach, based on knowledge of existing systems, is the lack of interoperability across systems and the inconsistent spatial content representations between surveys. While different video formats could be handled in a GIS analysis system, it is each video-survey's spatial data semantics that cause significant difficulties. Typically, each survey has been captured for a specific spatial-oriented objective where the geographical information generated may only describe video features or events over isolated frame sequences (McLoughlin et al. 2008). In general, this creates a barrier to using the video's spatial data in a GIS due to potentially diverse spatial semantics. From a usage perspective this approach is remote-sensing based where the various methodologies aim to extract information from scenes based on spatial data cues and vision techniques, rather than extracting video based on a spatial operations approach.

2.1.2. GIS-centred approach:

A GIS-centred approach is based on a spatial-extent context where the object space of a video sequence is modelled and analysed in a GIS-centred manner. Effectively, a volumetric model of each video frame is generated to represent the geographical space captured in each video frame or frame sequence. The aim here is to generalise the spatial representation of video and use this approach in a GIS-based analysis context. Thus, the video can be categorised, segmented, retrieved and/or indexed based solely on the geography it contains. There are a number of possibilities for providing a solution using this approach where a semantic understanding of the spatial content of video can be defined. An advantage of this approach is its wider applicability to the multiple types of spatial video survey data and opportunities for greater content accessibility for users. The effective results of this approach are to centralise and generalise existing and future spatial video data in a generic spatial data access scenario.

The disadvantages of this approach are the work involved in the organisation and assimilation of spatial video survey formats and the different levels of spatial accuracy and video quality content contained in a single coherent GIS data source.

The prevailing methodological approach to using video in a GIS is video-centred where the creation and/or appreciation of spatial data is achieved through access to video footage. However, the prevailing direction in existing spatial data interoperability is the development of standards and systems that generalise the content (ESRI 2003). Web Feature Services and Web Map Services present spatial data storage, access and management methodologies that provide well-defined GIS interoperability profiles for a large range of spatial data. However, these standards and technologies do not currently support spatial video.

2.2. Broader Research Objectives

While spatial video is a very useful visual data source, any single bespoke application is usually restricted to a specific set of project objectives that are based on using the video's visual information to enhance the GIS. Normally the consequence of such systems is to move towards an immersive user experience of the GIS spatial content through real world imagery and/or playback sequences. In other situations, detailed spatial data can be generated on the environmental features captured in the video footage (Ó Riain & McCarthy 2006). However, the original video generally becomes redundant and unused beyond the initial project requirements, yet remains in some form of long-term storage format with no reuse approach being envisaged to extract other spatial data. However, if a semantically appropriate geographical representation can be defined for each spatial video source, and stored in a generally accessible repository, then accessing and applying other research approaches and/or ideas to the video sources can be made easier. Thus, one focus of this research is to analyse the current state-of-the-art for spatial video in GIS. Following from this, a generalised GIS-constrained spatial video data structure is proposed as a solution to modelling and accessing large volumes of spatial video survey data. This approach can then be extended to implement and demonstrate the increased flexibility, extensibility and interoperability of the new model proposed in the use and study of spatial video in a GIS role.

This objective has a number of logical restrictions that need to be handled for any particular type of scenario in which spatial video can be used. However. the one principle aim is to retrieve logical video streams or images from a spatial video data index based on well-understood GIS analysis techniques.

3. Spatial Video and GIS

Creating an interoperable relationship between the specific areas of video, spatial data and GIS geospatial analysis is not a trivial exercise. Considerable commercial and academic research effort and investment has seen many different formats and systems developed for the acquisition and use of spatial video. This section discusses some of these efforts in the context of both the commercial and academic-research application areas that use spatial video in a GIS. Spatial video is dealt with in its most general sense where the collection and post-processing methods have already defined a data set of

frames tagged with location and/or orientation information. Thus, what is considered is the use of spatially tagged video in any form in any GIS application area.

The majority of the applications of spatial video are as a visual enhancement used for improved spatial reasoning of an otherwise cartographic-orientated GIS environment. This approach has its roots in work performed by Lippman (1980) at MIT where the Aspen Movie-Maps project uses spatial video in a GIS context. This project incorporated imagery into a user-orientated information enhancement tool for computer presentation and interaction. A multitude of intermediate stages have seen development of many of the different aspects that this project introduced; from the types of mobile mapping systems used to capture the data through to the processing, storage, analysis, usage and presentation of spatial video to an end user. The current most widespread application area for spatially tagged imagery is the Internet where Amazon® released *A9.maps Block View* in 2005 (2006). However, this service has since been withdrawn and it can only be surmised that the market-dominant Google™ *StreetView* (2007) and Microsoft® *Live Maps* (2005) has contributed to its demise.

3.1. Commercial Applications

A number of commercial application areas have been developed based on spatial video usage within a GIS environment. Spatial video is used by many different types of public and commercial entities for a multitude of reasons. Examples include: government and local authorities, utilities contractors, defence and emergency services, and transportation and service companies. The uses of spatial video typically involve remote management where road network asset inventories, validation and auditing, planning and engineering assessment can be performed based on visual inspections of the environment without individuals having to be in the survey region. The systems looked at here range from high accuracy survey entities that offer dedicated, bespoke spatial video hardware and software systems, through to internet-based standards and free service offerings.

Routemapper is a marketing brand of the IBI group, a Canadian-based international consultancy company (RouteMapper 2007). *Routemapper* markets consultancy, survey and bespoke software services for the collection, analysis and use of spatial video for road and rail survey projects. In their systems, video can be controlled both temporally through normal video-player-style controls and spatially through a cartographic interface of relevant raster or vector data sets. Advanced photogrammetric techniques can be applied to the video footage to take real world measurements in both two and three dimensions.

Red Hen Systems are direct competitors who offer a complete range of spatial video collection, processing and analysis services and products (Red Hen 2005). Their software solutions, similarly, provide a standard suite of GIS and video-related controls, such as data and feature logging, temporal and spatial video searching, image and map measurement and industry standard spatial file formatting.

One underlying but significant difference between these commercial spatial video vendors is that Red Hen Systems can log and handle multiple video streams. They can edit and splice these different streams based on user-defined video and/or geographic sequences which can then be exported to a new video stream. *Routemapper* currently does not have this capability; however it can handle different types of video format as long as they are pre-processed through the validation stage where the video is frame

grabbed and spatially tagged into self-contained spatial video sequences. Red Hen Systems requires all the spatial video formats to be in a DVD format which requires separate conversion procedures.

Immersive Media® (2006) have developed a hardware and software tool set called *Immersive Video*. The video data collection hardware system is known as *Dodeca 2360* and comprises 11 camera lens embedded in a single unit with a 360° horizontal and 290° vertical coverage. This camera system captures video data at 30 frames per second that can be post-processed through an automatic mosaic application to any desired output frame rate. The most notable use of this form of spatial video is through the *StreetView* Google® *Maps* interface (2007). Immersive Media® were originally contracted by Google® to survey a number of US cities, however, they have since acquired the necessary equipment and now manage the data capture themselves. Google's® use, to-date, of this form of spatial video is very restricted in GIS terms as standard spatial operations cannot be performed. Neither the Google® *Maps* or *Earth* interfaces can query bespoke segments of *StreetView* video data based on, for example, a line or polygon spatial operation.

The Oxford Metrics Group (OMG 1984) is an umbrella company for a group of specialist technology companies that supply motion capture and visual geometry systems and services to a variety of commercial and governmental organisations. Of particular interest is the *2D3* product's ability to generate 3D relief maps from UAV-captured aerial motion imagery. However, in the context of this paper their work specialises in computer-vision-based techniques that expand spatial video utilisation through better visualisation as opposed to the GIS approach discussed later.

A publicity article from the GEOINT 2009 conference by Francica (2009) also highlights a number of commercial/military orientated video applications that use these sources of spatial data in GIS. They include a mapping platform called *DataMaster*, developed by Boeing, where standard video search and cataloguing options are enhanced to include imagery visualisation in 3D environments. A *Full Motion Video Asset Management Engine (FMVAME)* is also showcased by Harris Corporation. They have concentrated on solutions to indexing video with spatial data and compressing the results for faster real-time transmission. Extending from this, and in partnership with a number of other US organisations, a new military orientated project called *'Valiant Angel'* (Lawlor 2008) is aimed at supporting the integration, exploitation and dissemination of video data collected from UAV's for better informed decisions.

While a number of other vendors of spatial video solutions exist, it is difficult to determine their core technologies and structures; as in the case of Intergraph's® *GeoMedia*, (in section 3.3), and Google's® *StreetView*, these complete systems are provided through incorporation of existing technologies from other vendors.

3.2. Academic Applications

In this section academic contributions to the development and use of spatial video are considered with particular concentration on data structuring and GIS interfacing. Spatial video data collection is not a concern here as the multiple methods and techniques of capturing video and spatial data are incidental to the methods of indexing, searching or using them in a GIS. While a large amount of literature exists for the multiple mobile platform methods that have been developed for the collection of spatial video, along with

multiple algorithms and techniques for post and real time video frame-to-spatial data indexing, no significant amount of work has been completed that considers a broader theoretical or practical GIS context for the spatial video data generated. No single piece of academic work identifies spatial video as the data source from which a generalised data structure or set of spatial operations can be defined.

Three significant points of view should be considered in relation to the literature relevant to this study:

1) The methods of indexing and storing video with spatial data.
2) The theoretical data structure model for spatial video in a GIS, particularly any three dimensional forms.
3) The use of spatial video concepts in GIS-based operational queries.

Research on these topics overlaps in many instances although some research is self-contained and only relates to GIS modelling or video frame spatial data indexing.

O'Connor et al. (2008) have implemented one specific example of a methodology for the storage, indexing and retrieval of video based on spatial metadata. They highlight a system where the MPEG7 (2006) and MPEG21 (Bormans & Hill 2002) video file multimedia metadata standards implementations are used to provide a complete and extensible video frame indexing system. By using these standards, not only can spatial data be associated with each frame, but multiple types of metadata can extend the searchable functionality of the video streams that are defined.

They also develop a user interface to query a spatial video database. However, only where a GPS tag has been recorded is the video key-frame indexed, therefore it is the video images that form the indexing control for return-of-video sequences. Also, the spatial queries involving region-based operations only return all the video key-frame images inside the region, as defined by the GPS location of where the image was taken. Based on the spatial video data structure discussed in this paper, and redefining some spatial operations semantics, this type of spatial operation should have a more precise meaning. Because of the GIS-constrained approach it has enabled the system to determine the difference between a video frame that was captured within a region, but does not visualise it, from one that visualises a region but may not have been captured within it.

Nobre et al. (2001) are among the first to introduce the notion of a geographical space being captured in each video frame image where a GIS data structure can be used to model this space. In this case a decision support system is developed for retrieval of video sequences based on user interest spatial queries. This system is heavily dependent on manual user calibration based on visual image analysis. Captured video is geo-referenced based on GPS data, followed by equal division of the line that the video traverses to represent the points where each frame is located. Each frame can then be queried and manually geo-referenced to determine the view frustum object space in a similar data structure format as that shown in figure 1. This is based on manual adjustment of key images that are calibrated based on visual inspection of real world object projections onto the image plane. Using this methodology an accurate measurement of the camera frame object space can be achieved based on arbitrary calibration.

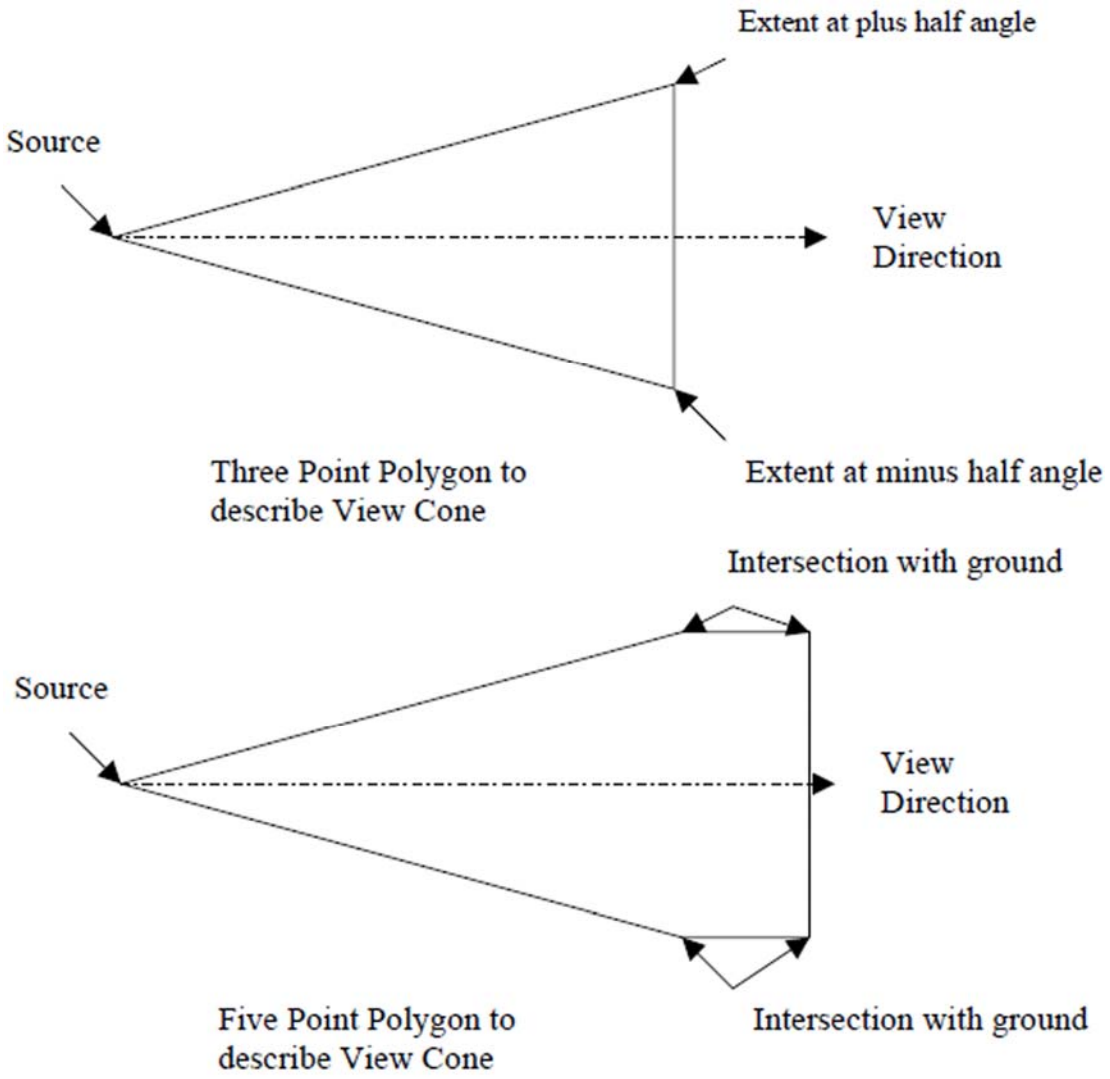

Figure 1. These two polygon specifications are from the OGC Geo-Video Service ViewCone specifications document and define the only spatial extent profiles implemented in the standard.

VideoGIS is a system defined in work by Navarrete et al. (2002) where spatial video indexing is based on geographical content segmentation. An XML spatial-data schematic, a high-level process algorithm, and an indexing structure are described, while the spatial implementations are based on standardised OGC GIS data structures. However, no detail is provided as to the automatic creation or usage of these data structures in a spatial video query context. Both this piece of work and Nobre's systems touch on some of the core concepts in the development of a spatial video GIS query data model in this paper. They introduce both the concept of modelling video frame object space as a geographical extent and using GIS data structures for this purpose; however they do not elaborate on these points nor on their possible uses in a GIS analysis context.

A number of papers from the Electronics and Telecommunications Research Institute (ETRI) in Korea detail a *VideoGIS* system called *GeoVideo*. The literature

defines an incremental development of, firstly, a systems specification for a spatial video system (Kim, Kim, Lee, Park, et al. 2003a); secondly, a data structure for metadata tagging of the video (Hwang et al. 2003); and finally, an implementation of the concepts in a mobile location based service (Qiang et al. 2004; Kim et al. 2003b). In these papers they present a final product called *MediaGIS* where a fully implemented and complete system from the point of data collection through to data distribution to end users is detailed.

The metadata spatial storage mechanisms use the MPEG7 data structures for video frame annotation which includes the spatial variables. Work also performed at ETRI specifically developed an implementation of this data structure based on MPEG standards (Joung et al. 2003). Centralised 3D databases form the backbone of the spatial queries that return the relevant image or spatial video sequence. Upon a successful user query, a viewing frustum is generated based on the pre-processed spatial and orientation data that indexes each image and where the perspective projections for transfer from 2D image space and 3D object space are calibrated in relation to existing 3D city models. This study does not assume availability of such rich data sources and as such only assumes and improves the viewing frustum based on empirical testing and modelling.

In a similar context, work in Milosavljević et al. (2010) augments a GIS with stationary video imagery. This work has leveraged the power of integrating existing spatial data and video surveillance systems. They have developed a real-time imagery retrieval and camera control system based on a spatial context interaction with the GIS system. Spatial query operations are two fold in that information from the cameras will define the GIS spatial data view or the GIS spatial can be used to control where the camera views.

Chen et al. (2009) have recently published an LBS-context spatial video system that integrates videos with map driving directions. A user interface has been developed that dynamically generates a video sequence for a driving route that can be played back. In many ways this system presents nothing that has not been achieved in a number of the commercial solutions mentioned in the previous section, however they are one of the first to target it for research purposes, and in a driving directions context. A twenty-user case study was completed to test the proof-of-concept and results suggest an improved immersive experience for users planning a journey, especially over the traditional process of map reading. Considerations in user interface design have also benefitted from implementations that consider other research findings into wayfinding and route memories. This involves the system storing, highlighting and expanding imagery critical to the focal points of a route, such as junctions and landmarks.

In Hirose et al. (1998) an interactive system of video imagery navigation has been completed based on a multi-view spatial video data collection, processing and query system. This work has since generated a number of extension projects that use multiple cameras to capture spatial video and which can be played back in dynamic video players where the viewpoint control is only limited by the degrees of freedom in the video capture system. In Hirose, a 360° horizontal viewing system is possible based on eight cameras calibrated with positional and orientation sensors. In further work such as in Pintaric et al. (2000) the same result is achieved with an array of digital sensors instead of individual cameras. Ultimately, this sort of work has lead to the *Immersive Video* systems highlighted in the previous section.

Highlighted in Cho (2007) is the ability to define accurately a camera's location based on the calculation of its viewing frustum. This work constructs 3D imagery from 2D camera pictures without the spatial location of the camera being known. This location can be determined by solving a number of well-documented systems of equations in computer vision (Hartley & Zisserman 2003). In this work, this is achieved based on at least six reference points that relate 3D Lidar data to 2D image points. Such a calibration can define the viewing frustum parameters to back calculate the camera location. Conceptually, the work presented in this paper has reversed this process as the camera location is known and the camera parameters can be assumed to an acceptable error range. However, Lidar data, if available, could be fused with the procedures presented here to produce more accurate viewing frustums for each video frame.

3.3. Hybrid Developments

This section describes the only existing methodological approach, through open standards, of video modelling in a GIS-constrained way. The Open Geospatial Consortium (OGC) is an international collaborative organisation that collates input on a broad range of geospatial issues from more than 360 organisations that include government, private and public sectors. In 2005 the OGC Web Services phase three (OWS-3) initiative defined a number of working areas that included a set of software profiles for the development and enhancement of a Geo-Decision Support Service (GeoDSS). Directly related to this study is one particular GeoDSS subtask: the implementation of a Geo-Video Service that can standardise access to video that includes geo-location information (OGC OWS-3 2005).

This service is still only in draft document stage (Lewis et al. 2006) but is very comprehensive none-the-less. Importantly, the set of concepts and specifications contained in this document form the basis for the adapted implementations that are developed in this study. The core similarity is the Geo-Video Service ViewCone concept. This is a two dimensional geometric shape that defines the viewable geographic extent or spatial extent bounding box of each frame of video within a spatial video file and is shown in figure 1. It is computed based on calibrated camera parameters and recorded spatial variables.

The OGC Geo-Video service work was undertaken by Intergraph® Corporation in 2005/2006 and extended into a full motion video analysis software add-on to their *GeoMedia* commercial product line (Intergraph 2008). This has been achieved through various collaborations of technology with EchoStrom's (2003) video georeferencing and ingesting software.

This technology forms a departure from the traditional role of separated spatial video viewers that augment the normal GIS interface. However, while this approach provides a new and more immersive video experience in a GIS, its application in this context is still only developing. It is also only defined for aerial video where the process of ortho-correcting and warping each video frame onto a relevant ellipsoidal plane is simpler, in relative terms, to the implementation of the same process in an oblique situation, i.e. terrestrially collected spatial video. Many more problems exist in moving these methodologies into a more oblique context as severe pixilation is but one of many problems that develop from warping video frames in a plane orthogonal to an ellipsoid's surface.

Importantly, this software defines a methodology where "video can be queried with all other geo-referenced enterprise content, allowing it to be easily located and utilized at a later date" (Wagner 2009). This context is discussed later in terms of the broad GIS constrained approach used in this paper to model spatial video in a GIS.

4. Towards a Solution to Generic Spatial Video Use

In this section, the limited OGC ViewCone model (Lewis et al. 2006) is restructured to a data structure called a Viewpoint. This approach provides a generic solution to modelling video frame object space by generalising an open spatial-coverage form through a more flexible and easily extensible format than that provided through the ViewCone method. The Viewpoint concept is based on combining the theories behind existing methodologies that model viewable regions, which include isovist, viewshed and frustum forms (Worboys & Duckham 2003; Benedikt 1979; Turner et al. 2001). Its form is based on a 3D viewshed approach that has been defined where a viewing frustum represents each frame's geographical space; figure 2 shows the principle elements that define the basic structure. It is a more flexible form as it can be altered and/or updated to improve the modelling accuracy of terrestrial spatial video. In the terrestrial case numerous situations require this flexible modelling construct as occlusions can randomly enter and leave the video's viewing space as well as geographical restrictions affecting both the depth-of-field and field-of-view of the camera. For the examples and figures presented in this paper, a base case maximal Viewpoint structure is defined; however the accuracy of the Viewpoint structure can be defined to many more levels of complexity dependent on the methods employed to construct these extent regions.

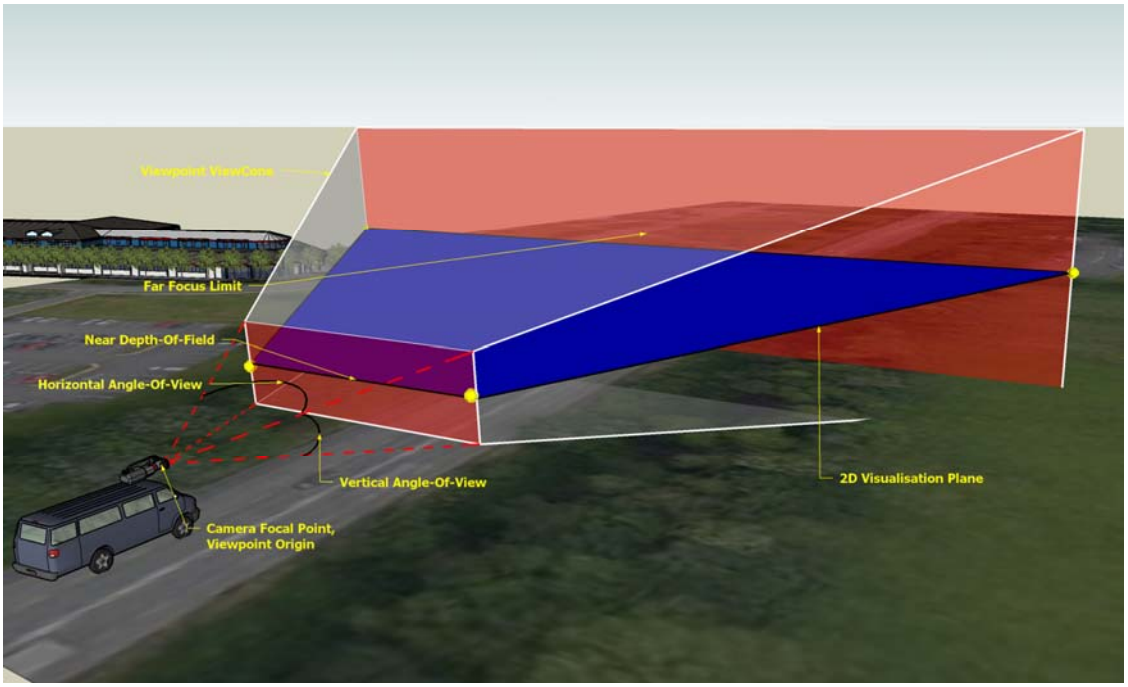

Figure 2. 3D spatial video single-frame Viewpoint representation. This image also highlights the 2D visualisation polygon as the central bisecting plane which is used in the section 5 examples.

For this paper each Viewpoint has been constructed to approximate an image's geographical extent by using a minimal set of spatial and camera parameters. In this case the spatial elements have been derived from GPS where the latitude, longitude, altitude and heading can be determined; however, this particular source requires a bespoke semi-supervised process to minimise the various errors inherent to it and maximise the video-frame synchronisation accuracy. This process defines the Viewpoint origin, shown in figure 2, and its ViewCone-element spatial orientation. Generating the geographical extent space involves extrapolating from this origin all the points that will define the six polygons which form the polyhedral ViewCone shape shown in figure 2. These points can be calculated by solving the geodetic forward algorithm as defined in Vincenty (1975). To achieve, this the video stream angle-of-view and depth-of-field are calculated based on the camera's focal-length, aperture and circle-of-confusion. These can be acquired either from the known calibration data that may accompany a particular survey or by a fitting approach that empirically determines these variables from a supervised viewing of the video on a ground truth backdrop.

This methodology has been applied to 46 minutes of oblique terrestrial spatial video. Approximately 75,000 Viewpoints were generated in 3D and stored in a *PostGIS* spatial database. To demonstrate how spatial video can be used in a GIS, numerous spatial video geospatial analysis operations have been performed on the 2D central dissecting plane of each Viewpoint, shown in figure 2. The 2D central dissecting plane is, at its simplest, a four point polygon feature calculated as the area enclosed by half the height of both the near depth-of-field and the far-focus-limit planes, as shown in figure 2. These planes are the vertical bounds of the ViewCone space based on the orientation of this video data-set. The following sections describe how this model is used in a selective set of geospatial analysis operations.

5. Video Geospatial Analysis

In this section, examples of GIS-based analysis of spatial video are presented. They are selectively chosen from many possible approaches to using GIS to analyse, segment or categorise video and are based on queries involving the plan-view 2D central dividing plane taken of each Viewpoint, as shown in figure 2. Significantly, in this case the Viewpoint data structure defines the maximum possible geographical coverage regardless of occlusions or other restrictions that may exist in the image space. Common to these examples is the use of non-video spatial data in the query context role where point, line and polygon units, relevant to the survey area, are used to provide an experimental platform that demonstrates how different GIS operations and spatial data sources can be used with spatial video. Each example is operationally independent of the video but directly related to its index for playback. Figure 3 shows an overview of the study area and the tracks of each survey route.

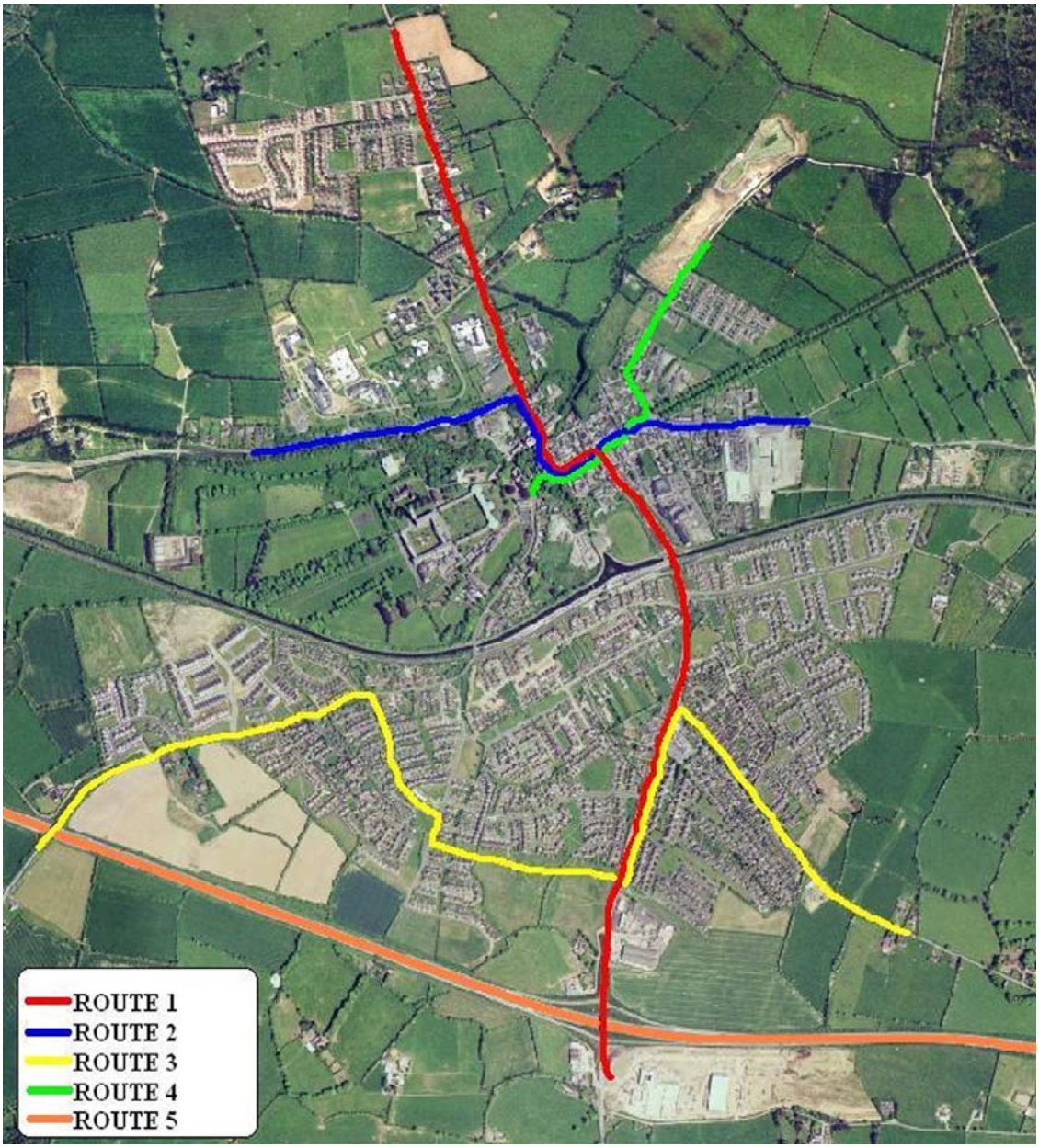

Figure 3. Overview orthophoto of the study area with five, independently collected, spatial video survey routes overlaid. Each route defines a point-to-point return survey, i.e. video and spatial data has been captured in both directions. A PostGIS database has been populated from these surveys using a GIS-Centred spatial approach that defines a Viewpoint data structure representation of the video footage.

5.1 Point-In-View

In this example a point-in-view spatial query operation is performed on the Viewpoints database where video footage is retrieved that has captured a visualisation of the point(s) of interest. This involved a set of control points being created in a GIS application to generate a query data set. This data set is then used in an SQL spatial

operation to retrieve all video frames that capture these points. The results should form linearly logical sets of video sequences that represent all frames that the points appear in. Mostly, this is achieved through standard SQL control statements where each Viewpoint construct represents the query search space. This effectively resulted in a point-in-polygon operation where each point was searched for all the Viewpoint polygons it is contained within.

The results of this operation, which include ground truth empirical testing, showed that the approach taken provided accurate results, an example of which is shown in figure 4. However, various considerations need to be defined as the Viewpoint search space is not necessarily an accurate representation of the video frames' geographical extent. This is because a maximal ViewCone has been calculated that does not take into account any occlusions in the physical space but assumes the image is capturing a expansive planer region. In the case of the query results that relate to figure 4, testing has highlighted that the video frames captured to the left of the point-in-view contain occlusions. When reviewing the result frames' physical space, another building and some large trees obscure the building represented by the point-in-view query, thus they are not optimally representative views. Section 5.4 discusses a solution to this.

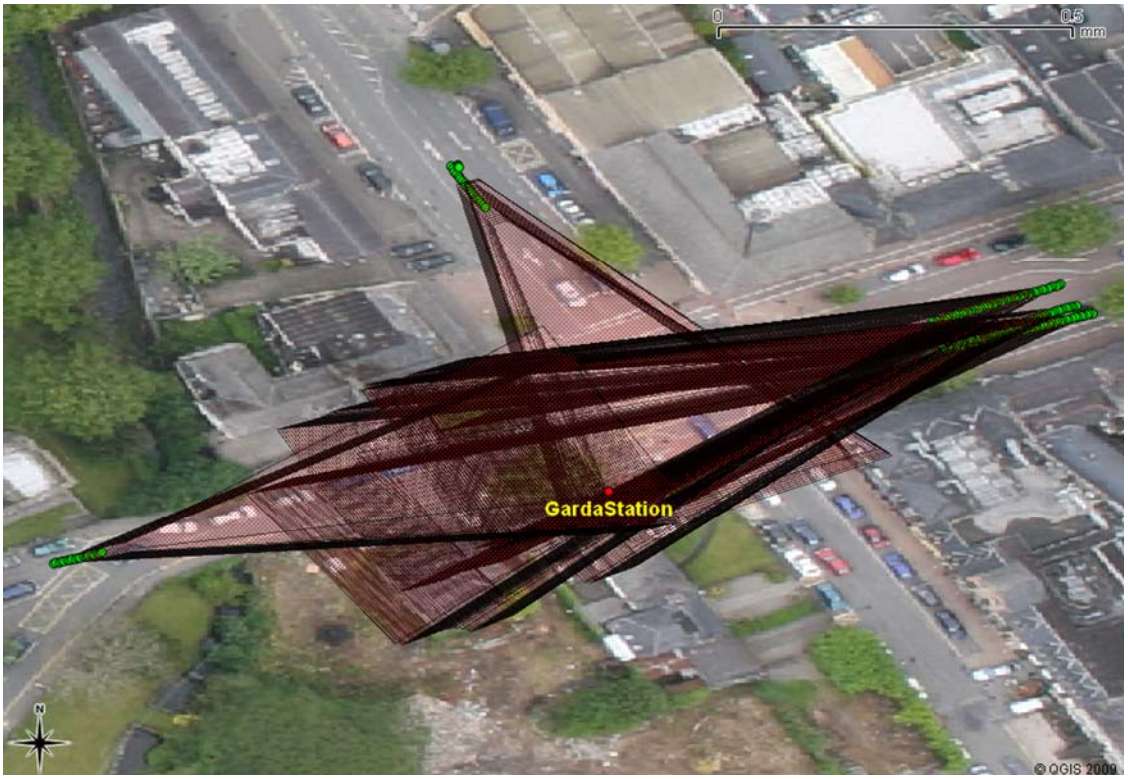

Figure 4. Sample result visualisation of a Point-In-View operation. The (red) point, labelled 'GardaStation', is the point-of-interest of which all video footage is requested. The results show each Viewpoint that contains the query point; each video frame capture location (Viewpoint origin) is highlighted as a green point while the geographical extent polygon is highlighted as the transparent red polygons. Each green Viewpoint capture location indexes a video frame for visual retrieval and playback.

Dependent on the search space context, a maximal Viewpoint becomes useful because the point-in-view's location represents a building centroid rather than a polygon based footprint. If each Viewpoint polygon accurately represented the space it captured, i.e. the ViewCone polygon's boundary ended at the building facia, then another approach would be required to formulate this spatial operation as no Viewpoint would be returned. Thus, this operation would need to consider a number of other spatial data sets to help define a higher level of accuracy. In another sense this approach to implementing the point-in-view operation enables the spatial video to be segmented more accurately for those frames that do not contain the query object.

5.2 Polygon-In-View

In this example an Irish census-district polygon data set is used. From this set a single polygon forms the search space query where relevant spatial video sequences are returned that record the polygon's geographical space. Normally this query could be likened to a point-in-polygon search; however there is an important semantic difference. Here the query question is being constrained to define video that captures space in a polygon as opposed to video captured in the polygon, probably a more reasonable approach given the specific data type that video is, i.e. a visual record of space. Thus, the context is: what is in a video frame and what is not in a video frame, i.e. video that records space in the query region can be captured within, on or outside the polygon boundaries. Thus, frame-capture location points cannot fully satisfy the point-in-polygon approach as frames captured inside the search polygon may not be recording region space. Also, frames captured outside the search polygon may record space that should be included. Therefore, the Viewpoint geographical viewshed extent can be used to control whether the video frame is relevant to the search space or not.

Figure 5 highlights the results of one approach to this spatial operation; in this case a Viewpoint spatial intersection was implemented where at least 60% of each ViewCone's area is contained within the query polygon, regardless of the capture location point being contained within it. Thus, all video footage captured within, on or outside the search polygon has been excluded from the results set if 60% or less of its viewshed geographical extent area is not contained within the frame. Obviously this specific approach, and the intersection percentage bound, could be easily altered dependent on the desired results. However, the main point being highlighted is that the Viewpoint modelling approach provides the extra spatial detail to fully define the query question in relation to the visual nature of video and the space it captures.

Figure 5. Viewpoint controlled polygon-in-view operation where a query polygon is the search space. The resulting Viewpoints that capture a visualisation of the polygon are shown as the green points with each associated geographical extent polygon of the area captured overlaid as the red transparent ViewCone.

5.3 Thematic Coverage

Based on the same polygon data set used in the previous example, this example investigates a content analysis approach to spatial video segmentation. The query polygons have been assigned with land use categorisations which have then been spatially intersected with the Viewpoints database. Many possible approaches to acquiring results can be achieved from this operation; in this case the objective is to define both the total area captured by land use and the percentage of total video available. Table 1 presents the results where the percentage of video footage and the geographical area are defined based on each land-use type. In the case of the area captured results, these represent the union of all the Viewpoint polygons by region and the calculation of the union's area, not the summed total area of each Viewpoint's ViewCone. Also, the percentage of spatial video is representative of only a portion of the total database as only a small subsection of the polygon data set was used in the query operation.

Table 1. Aggregated results of the Viewpoints database thematic coverage operation. The video content is determined to contain various areas of thematic geographical content based on the polygon data sets land use assignments but also a percentage of the total video held in the database.

Coverage Type	**Total Area Captured m²**	**% of total Spatial Video**
Residential	40,489.03m²	5.0

Urban	65,683.27m²	8.1
Educational	38,803.22m²	4.8
Rural	40,365.74m²	5.0

While this case presents aggregated analysis results, they are all still linked to video frame indexes. Thus, the sequences of spatial video represented by the 5% residential land use type are easily acquired for playback. Also, where a Viewpoint's ViewCone crosses polygon boundaries, the area calculation is apportioned based on the boundary intersections with the ViewCone. As mentioned, these results could be composed from many alternative perspectives that include temporal, spatial and visual questions. As an example, time could be an important constraint as video records space over a discrete temporal period land use may change over time. Thus, the spatial operation query would require a temporal bound as only certain video footage will be an accurate visual representation of a particular land use classification at a particular time.

5.4 Viewpoint Accuracy

So far this paper has presented a Viewpoint data structure that models spatial video through a modified version of the existing OGC ViewCone approach. In both models a basic calculation of the geographical extent of each video frame is formulated, based on the camera's parameters and spatial-data. The extrapolated ViewCone in each case defines a maximal extent with no allowance for physical space occlusions. In this section a spatial data approach is highlighted through a single example that generates a more accurate Viewpoint model of the space captured in each frame using a buildings footprint data set. This operation intersects each Viewpoint with the buildings footprint data set to calculate a more accurate viewable region, which can then be updated to the database. Figure 6 shows the results of this operation where the original ViewCones have been recalculated to a more accurate representation of the space recorded in each video image.

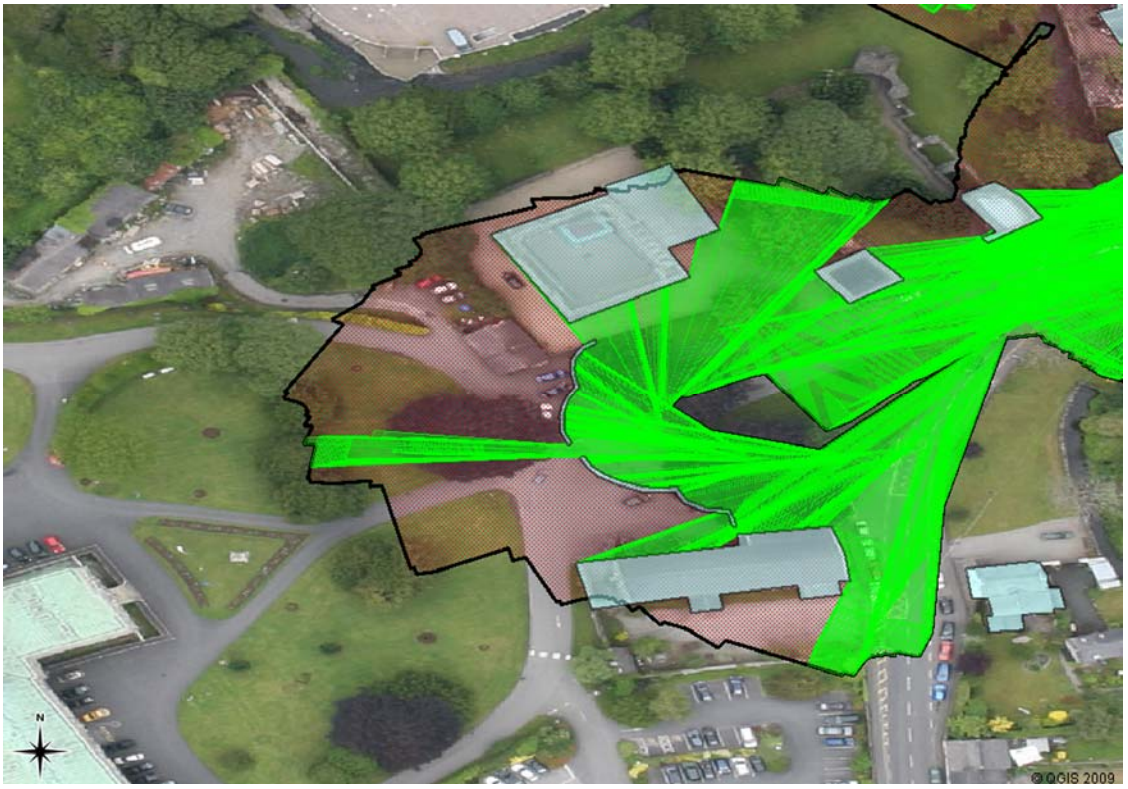

Figure 6. Shown here are the results of an intersection operation between the original geographical extent of each video frame's ViewCone and a buildings data set (light blue polygons) that represent image object-space occlusions. For easier viewing, a spatial union has been performed on all the original ViewCone geographical extents and is shown as the transparent red area with a heavy black outline. The operation results are shown as the overlaid bright green polygons where each one is a more accurate representation of the image object-space.

This process has many possible extensions using many other sources and forms of non-video spatial data. It presents a novel GIS-constrained approach to generating an accurate viewshed of a video frame's object space where the visual characteristics of the video are ignored. This is in contrast to existing remote sensing techniques where the visual characteristics of the video are central to photogrammetric processes for the extraction of spatial features. Research is ongoing using Lidar data and other 3D models of the environment to generate more accurate intersection data sets. However, depending on the expected results from a Viewpoint-based spatial operation, the data structure needs to be capable of either dynamic generation or storage in different forms. This has been mentioned in section 5.1 where the point-in-view search would not have been as effective unless a maximum ViewCone had been used.

6. Conclusions

Modelling video in GIS has been largely ignored for numerous reasons. Data volumes involved in collecting and storing video footage, lack of methods to accurately spatially reference video, inaccessibility and unsuitability of video data access over low-bandwidth networks have all hindered the large scale use of spatial video to-date.

However, with improving network technologies and cheaper more accurate collection and storage systems these issues are becoming less of a problem. This is evidenced in the academic research and expansion of commercial systems development as highlighted in the state-of-the-art review presented in this paper. Thus, it has been shown that spatial video is playing a far greater role in GIS with many diverse research directions yet to be undertaken. The many different levels of research into video and GIS range from the higher-level semantic understanding of GIS-Video through to the more applied and basic standardisations of sensor data integration that currently do not exist. Towards dealing with these diverse challenges, this paper has defined a broader context for discussing and defining the role of spatial video and GIS. Some inherent problems are presented and solutions introduced that work through examples that include how video should be spatially referenced, how best to model video in a GIS, what can video be used for in GIS and how can a GIS be used to analyse/segment video. The later sections of this paper tackle these last three points, where a more complete modelling approach is presented in the form of a Viewpoint, followed by a number of geospatial analysis approaches that both describe and segment spatial video. While the Viewpoint form builds on the current standard methodological approach to GIS-Video integration, (Lewis 2006), much work remains to be undertaken in how GIS geospatial functionalities apply to video data. It is likely that spatial video will play an increasingly important role in GIS operations extending the uses for the usual data type classification of point, line, areas and volumes. Consequently, this paper describes and encourages greater discussion on how standard GIS operations can function with spatial video.

Acknowledgements

This work has been completed at the National Centre for Geocomputation at NUI Maynooth. It was made possible through funding from the Strategic Research Cluster grant (07/SRC/I1168) awarded by Science Foundation Ireland under the National Development Plan. Elements were also completed through funding from the Embark IRCSET postgraduate research grant (RS/2005/331) and the NUIM John and Pat Hume Research Scholarship.

References

A9 Maps, 2006. Business shop front spatial image location and direction system. *Internet Source*. Available at: http://maps.a9.com/?

Benedikt, M.L., 1979. To take hold of space: isovists and isovist fields. *Environment and Planning B*, 6, 47-65.

Bormans, J. & Hill, K., 2002. MPEG-21 Overview v.5. *Internet Source*. Available at: http://www.chiariglione.org/mpeg/standards/mpeg-21/mpeg-21.htm.

Chen, B. et al., 2009. Integrated Videos and Maps for Driving Directions. In *User Interface Science and Technology (Proceedings of UIST 2009*. Victoria, BC.

Cho, P., 2007. 3D Organization of 2D Urban Imagery. In *Applied Imagery Pattern Recognition Workshop - 2007*. pp. 3-8.

ESRI, 2003. Spatial Data Standards and GIS Interoperability. *New York*, (January). Available at: http://www.esri.com/library/whitepapers/pdfs/spatial-data-standards.pdf.

EchoStorm, 2003. Video Data Management and Integration Technologies. *Internet Source*, 2009. Available at: http://www.echostorm.net/.

Foy, S. et al., 2006. Inertial Navigation Sensor and GPS integrated for Mobile mapping. In *GISRUK 2006*. NUIM, Maynooth, Ireland: GIS Research UK and Ireland.

Francica, J., 2009. GEOINT 2009: From Intel to Imaging, But Full Motion Video Stole the Show. *Internet Source*. Available at: http://www.directionsmag.com/printer.php?article_id=3319.

Google, 2007. Google Street View. *Internet Source*, 2008. Available at: http://maps.google.com.

Hartley, R. & Zisserman, A., 2003. *Multiple View Geometry in Computer Vision* 2 ed., Cambridge University Press.

Hirose, M., Watanabe, S. & Endo, T., 1998. Generation of wide-range virtual spaces using photographic images. In Atlanta, GA, USA: IEEE, pp. 234-241.

Hwang, T. et al., 2003. MPEG-7 metadata for video-based GIS applications. In *IGARSS 2003. 2003 IEEE International Geoscience and Remote Sensing Symposium. Proceedings (IEEE Cat. No.03CH37477)*. IEEE, pp. 3641-3643.

Immersive Media, 2006. Immersive Video GIS Viewer. *Internet Source*, 2009. Available at: http://www.immersivemedia.com/#102.

Intergraph, 2008. GeoMedia Motion Video Analyst. *Internet Source*. Available at: http://www.intergraph.com/sgi/downloads.aspx?assetid=DFI-US-0018A-ENG.

Joung, Y., Kim, K. & Kang, K., 2003. The Development of MPEG-7 Interface over MPEG-4. In *IEEE International Conference on Consumer Electronics, 2003.* IEEE, pp. 276-277.

Kawasaki, H. et al., 1999. Automatic modeling of a 3D city map from real-world video. In *ACM international conference on Multimedia*. Orlando, Florida, United States: ACM Press, pp. 11-18.

Kiger, B., 1972. Videography, What Does It All Mean? , 9. Available at: http://www.experimentaltvcenter.org/history/pdf/kigercinematography_2728.pdf.

Kim, K. et al., 2003a. GEOVIDEO: The Video Geographic Information System as a first step toward MEDIAGIS. In *ASPRS 2003*. Anchorage, Alaska: American Society for Photogrammetry and Remote Sensing, pp. 1-6.

Kim, K. et al., 2003b. The interactive geographic video. In *IGARSS 2003. 2003 IEEE International Geoscience and Remote Sensing Symposium. Proceedings (IEEE Cat. No.03CH37477)*. Ieee, pp. 59-61.

Lawlor, M., 2008. Valiant Angel, Innovative Technologies Help Create Joint Solutions at the Speed of War. *Internet Source*. Available at: http://www.afcea.org/signal/articles/anmviewer.asp?a=1719&print=yes.

Lewis, J., Grindstaff, G. & Whitaker, S., 2006. Open Geospatial Consortium Geo-Video Web Service. , 1-36. Available at: http://portal.opengeospatial.org/files/?artifact_id=12899.

Lippman, A., 1980. Movie-Maps: An Application of the Optical Videodisc to Computer Graphics. *ACM SIGGRAPH Computer Graphics*, 14(3), 32-42.

Livingstone, D., Raper, J. & McCarthy, T., 1999. Integrating aerial videography and digital photography with terrain modelling: an application for coastal geomorphology. *Geomorphology*, 29(1-2), 77-92.
MPEG7, 2006. MPEG-7 Overview. *Internet Source*, 2008. Available at: http://www.chiariglione.org/mpeg/standards/mpeg-7/mpeg-7.htm.
McCarthy, T., 1999. Integrating aerial video with GIS, PhD. Thesis.
McLoughlin, S.D. et al., 2005. Classification of road sign type using mobile stereo vision. *Proceedings of SPIE*, 5823, 133-142.
McLoughlin, S.D. et al., 2008. Mobile mapping for the automated analysis of road signage and delineation. *Intelligent Transport Systems, IET*, 2(1), 61-73.
Microsoft, 2005. Live Search Maps. *Internet Source*, 2008. Available at: http://maps.live.com.
Milosavljević, A., Dimitrijević, A. & Rančić, D., 2010. GIS augmented video surveillance. *International Journal of Geographical Information Science*, In Press.
Navarrete, T. & Blat, J., 2002. VideoGIS: Segmenting and indexing video based on geographic information. In *5th Agile Conference on Geographic Information Science*. Palma, Balearic Islands, Spain, pp. 1-7.
Nobre, E.M. & Camara, A.S., 2001. Spatial Video, Exploring Space Using Multiple Digital Videos. In *6th Eurographics Workshop on Multimedia*. Lisbon: Environmental Systems Analysis Group, New University of Lisbon.
O'Connor, N.E. et al., 2008. A content-based retrieval system for UAV-like video and associated metadata. *Proceedings of SPIE*, 6946, 69460I-69460I-10.
OGC OWS-3, 2005. Open Geospatial Consortium, Open Web Services phase 3. *Internet Source*, 2008. Available at: www.opengeospatial.org/projects/initiatives/ows-3.
OMG, 1984. Motion Capture and Visual Geometry Systems and Services Supplier. *Oxford Media Group*, 2009. Available at: http://www.omg3d.com/html/company.html.
Pintaric, T., Neumann, U. & Rizzo, A., 2000. *Immersive Panoramic Video*, Los Angeles, California, USA. Available at: http://graphics.usc.edu/cgit/publications/publications.php.
Qiang, L. et al., 2004. A scalable VideoGIS system for GPS-guided vehicles. *Signal Processing: Image Communication*, 20(3), 205-218.
Red Hen, 2005. Spatial Video hardware and software website. *Internet Source*, 2005. Available at: http://www.redhensystems.com.
RouteMapper, 2007. Spatial Video solutions for transportation sector. *Internet Source*. Available at: www.routemapper.net.
Rzhanov, Y., Linnett, L.M. & Forbes, R., 2000. Underwater video mosaicing for seabed mapping. In *Image Processing, 2000.* Vancouver, BC, Canada: Image Processing, 2000, pp. 224-227.
Turner, A. et al., 2001. From isovists to visibility graphs: a methodology for the analysis of architectural space. *Environment and Planning B: Planning and Design*, 28(1), 103-121.
Vincenty, T., 1975. Direct and Inverse Solutions of Geodesics on the ellipsoid with application of nested equations. *Survey Review*, 23(176), 88-93.

Wagner, R., 2009. Intergraph Introduces Motion Video Exploitation Solution for Increased Real-Time Awareness and Actionable Intelligence. *Internet Source*. Available at: http://www.intergraph.com/assets/pressreleases/2009/10-19-2009.aspx.

Worboys, M. & Duckham, M., 2003. Uncertain Viewsheds. In *GIS A Computing Perspective*. CRC Press, p. 355.

Ó Riain, G. & McCarthy, T., 2006. Innovative Data Capture and Presentation Techniques in support of the EU Environmental Noise Directive. Available at: http://erc.epa.ie/safer/resource?id=6d6b6300-d163-102c-9c91-0a68ec663af0.