

A Semantic Approach for the Indexing and Retrieval of Geo-referenced Video

Toni Navarrete, Josep Blat

Departament de Tecnologia, Universitat Pompeu Fabra
Passeig de Circumval·lació, 8. 08003 Barcelona, Spain
{toni.navarrete, josep.blat}@upf.edu
<http://www.tecn.upf.es/gti/>

Abstract. The multi-layered structure of geographic information can be used to provide a rich description for geo-referenced video sequences. Furthermore, the complex structure of video, with its temporal dimension, makes this integration challenging. In this work we define a method for indexing and retrieving geo-referenced video sequences based on the geographic content. This method is the basis for building a digital video library, and relies on a semantic framework that has been defined to represent and query the thematic information in a repository of geographic datasets. The meta-information structure used to describe video sequences is also described.

Keywords: geographic information, semantics, geo-referenced video, video indexing, video segmentation

1. Introduction

In a geo-referenced video sequence, some properties of the camera regarding its location are captured during the video recording. These properties include position (typically obtained from a GPS receiver), orientation with respect to the North (for instance captured from a digital compass), and often vertical tilt (the angle with respect to the horizon). Focal length, which determines the angle of vision, and receiver (negative or CCD in a digital camera) dimensions are also recorded. All these properties make it possible to obtain the geographic area (*area of vision* from now on) that can be seen in each frame of the video sequence. Our approach is based on describing a video sequence according to the area of vision of its frames, and according to the geographic information corresponding to this area of vision.

The geographic information used to describe videos is extracted from a repository of geographic datasets. A geographic dataset represents the real world by assigning thematic classes (for instance “forest”, “agricultural land”, “natural park” or “motorway”) to spatial elements. However, different producers structure their datasets in terms of different sets of thematic classes, which are often not precisely defined and which may be understood in different ways by different subjects. An effective

method for integrating geographic information from diverse sources has to address this semantic heterogeneity. In [1] we have defined a semantic framework to represent and integrate thematic geographic information. It comprises two main elements: an ontology defined to represent the thematic knowledge in a repository of geographic datasets; and a set of semantic services used to enable external clients to find, translate and integrate thematic information from different datasets in the repository. The ontology is expressed in OWL, while the services are based on Description Logic. An implementation has been developed in Java, using the Jena framework.

In Section 2 we provide a brief overview of different approaches using geo-referenced video. Section 3 describes our semantic-based proposal for segmenting and indexing video sequences based on thematic information, which is the basis for the definition of a digital video library. It also describes how video is modelled following a stratification-based approach. Section 4 enunciates the main types of queries in the digital library, while Section 5 formulates how meta-information describing video segments is structured. Finally, in Section 6 we discuss some possibilities for continuing this work.

2. Related Work: Using Geo-referenced Video

The use of geo-referenced video has become relatively frequent in the last years. This has been especially remarkable in organizations that maintain linear infrastructures as streets, roads and railroads. Special systems with video and location capturing devices are designed to be mounted on vehicles. This way, the vehicle can easily obtain geo-referenced videos for the whole network. Once processed and integrated in a GIS environment, these geo-referenced videos provide a rich and intuitive support for monitoring and decision making.

Several commercial systems exist to record geo-referenced videos and to add them to GIS environments through a simple post-processing procedure, which basically includes a synchronization of the video and the location data, and an interpolation to obtain a location for each frame (note that GPS receivers get measures slower than the video frame rate). Probably the most extended software tools are MediaMapper and GeoVideo, both developed by Red Hen Systems¹, which also provides different hardware systems for recording geo-referenced video that can be mounted on vehicles or on aircrafts, or that can be carried by pedestrians. These software tools enable the user to view the path of the camera on a map. They also show a cursor on that path that indicates the exact location of the camera during video playback. The user may control the video by moving this cursor on the path, as well as from a typical video control with play/stop/rewind/fast-forward buttons. An analogous commercial tool is CamNavMapper, by BlueGlen², that also covers acquisition, integration in a mapping environment and similar playback controls. A similar playback functionality is provided by ImageCat's Views³. Other simpler software tools like GeoMovie by

¹ <http://www.redhensystems.com>

² <http://www.blueglen.com/>

³ <http://www.imagecatinc.com>

Magic Instinct Software⁴ or VideoMapper⁵ provide a postproduction process that superimposes the coordinates of the camera and other meta-information on the video image.

As an indication of the increasing importance of geo-referenced video, the Open GeoSpatial Consortium has proposed a geo-video service for its OWS-3 (OGC Web Services Phase 3) interoperability initiative. The aim of this initiative is to develop “a web service for access to video data including geo-location information”. This service will provide an interface for requesting stream video, that can be controlled through play-back commands from a web service client. The service also will provide metadata in the video stream sufficient for a client to geo-locate the video.

Some examples of the use of geo-referenced video, apart from linear infrastructures management, are [2] and [3]. The former has developed a forest fire decision support system. This system has a collection of geo-referenced aerial videos, and at the moment of a fire alarm it enables firemen to watch videos of the affected area. The latter also uses geo-referenced aerial videos in a system for the validation of land-cover maps of inaccessible areas of Canada. Also relevant is [4], that has developed a system that combines geo-referenced video captured from a specially equipped vehicle called 4S-Van [5] and a 3D model of a city consisting in a 2D feature-based map with building heights and a digital elevation model. A method called VWM (Virtual World Mapping) [6, 7] enables them to link spatial segments in frames to buildings from the city model. This architecture supports visualizing geo-referenced videos enhanced with information of buildings in a GIS environment [8], as well as developing other interfaces, as the Personal navigation system [9] for portable devices. This is very related to the discipline of Augmented Reality, where geo-referenced video recorded from a camera carried by the user is processed in real-time in order to show her/him (usually through special devices) an augmented version of the image, presenting information on what s/he is seeing.

Although from a different approach, it is worth mentioning the Aspen Movie Map Project [10] developed at MIT in 1978, which is considered the first project that combined video and geographical information, and in fact is usually referred to as the birth of multimedia [11]. Using four cameras on a truck, all the straight segments of the streets of Aspen were filmed in both directions, as well as every turn (also in both directions), taking an image every three meters. The system consists of two videodiscs that enables users to “drive” through the city, deciding in each crossing which direction to follow. The user could stop in front of some of the major buildings of Aspen and walk inside. Interiors of several buildings were also filmed. A screen was used to show video, while another showed a street map of Aspen. The user could point to a spot on the map and jump directly to it, instead of finding the way through the city.

It has to be remarked that we do not consider here those approaches consisting in video clips with an overall spatial reference for the whole clip. Instead, we focus on geo-references at the level of frames or segments. A non-exhaustive list of “classical” examples of the use video clips in geographical multimedia applications are BBC Doomsday Project [12], a video-disk-based map of Great Britain where the user can

⁴ <http://www.justmagic.com>

⁵ <http://www.videomapper.com>

visualize videos and other multimedia elements from certain localities; [13], which provides a collaborative hypermedia tool for urban planning; the CD-ROM of ParcBIT [14], a hypermedia application for supporting architects to develop a plan for a technologic park; and the hypermedia application for the North Norfolk coastal management discussed in [15].

3. Semantic-based Segmentation and Indexing

We can observe that the systems discussed in the previous section can be classified in two main groups: a first one that is mainly oriented to provide visual information for geographic features in a GIS environment, namely in contexts like road management, fire decision support or land-cover maps validation; and a second one that focuses on the enhancement of the video images with information of buildings or other elements. Our proposal follows a different direction: it uses the thematic geographic information that can be extracted from a GIS or spatial database, once a given video sequence is geo-referenced, in order to segment and index the video sequence. This process of segmentation and indexing is the basis for the definition of a digital video library, and consequently our work is more related to video information retrieval than to the abovementioned systems. Our library will enable external clients to access the video collection in order to retrieve elements satisfying thematic criteria. The main type of query that the system has to support is, given a thematic class, to retrieve the video segments depicting that theme. The results of a query will be a collection of video fragments (segments) extracted from different video sources. For example, the query asking “forests” would return all the fragments of videos in the video base containing forests or related themes (subclasses of forest). We have called VideoGIS to the system that implements this type of integration of geo-referenced video and geographic information. It has to be remarked that our semantic framework for geographic information provides a formal definition for the thematic classes that describe video sequences. Furthermore, the semantic services defined in the framework will be used in the segmentation and indexing process.

Each video segment is indexed according to a set of thematic classes from the ontology, T_1, \dots, T_n (referred to as *indexing themes* from now on). They have to be previously selected by the user responsible of the indexing process. Each indexing theme gives rise to a different layer of meta-information of the video segment. This way, each layer of meta-information corresponds to a particular view of the thematic information of the video segment, focusing on one of the indexing themes.

It has to be noted that, although it is not compulsory, all the videos in the collection are usually indexed according to the same set of themes. However, there is no problem on indexing different videos with different focus, i.e. using different sets of indexing themes. Furthermore, T_1, \dots, T_n are usually disjoint, or at least they do not have common subclasses in the thematic ontology. But again, this is not compulsory and the user may select themes with non-empty intersection.

It is also important to note that the indexing process is driven by this set of indexing themes, and not by datasets. This way, the user selects the themes that s/he is interested in, but s/he must not know in which datasets this thematic information can

be found. The system will be responsible for finding the involved datasets and dataset values.

Our video model is based on a stratification approach [16, 17], where each selected indexing theme corresponds to a stratum or layer. We will see that the video is firstly segmented according to the thematic information, and then, each segment will be indexed according to the thematic classes appearing in its area of vision, as Figure 1 shows.

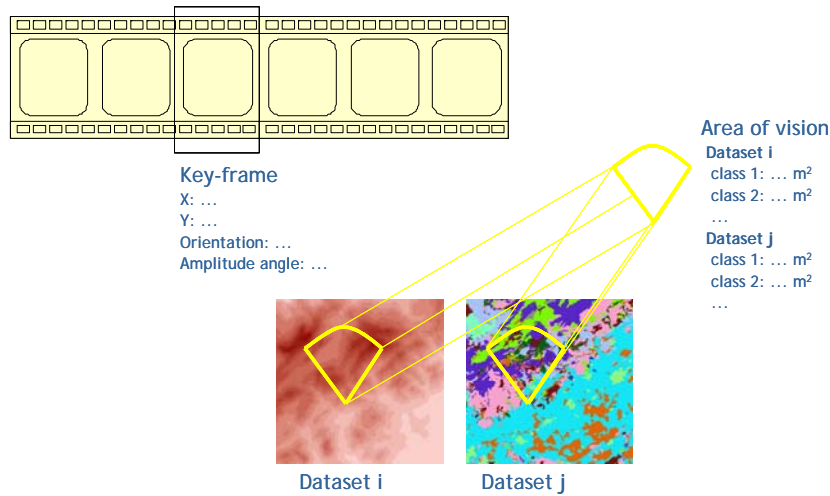


Figure 1. Overview of the process of indexing a video segment

Each frame of a geo-referenced video V has several properties related to the location and focal properties of the camera. Some of them are constant along the video, as receiver length (s) and width (w), while others may change frame-by-frame, as camera position (X,Y,Z), orientation (ρ), tilt (θ) and focal length (l). Furthermore, the angle of vision (α) of each frame can be computed from l , s and w . All these properties make it possible to obtain the area of vision A_j for a given frame f_j . We have also used a Digital Elevation Model and a viewshed analysis algorithm to determine which areas can be seen from the camera position and which are hidden by elevations, up to a certain distance d from the camera.

Segmentation is carried out according to the thematic information in the area of vision of each frame of the video. The process is equivalent to the feature-based segmentation described in [18, 19], where each segment represents a sequence of consecutive frames containing the same features, or thematic classes in this case. However, this process is done for each indexing theme, and consequently a parallel structure of n segmentations is generated. From now on, we denote by *layer* each of these different n segmentations. This way, each frame in the video belongs to at most n different segments, one for each layer or indexing theme. Note that we say “at most” since segments with no visible themes are not considered.

The segmentation algorithm firstly obtains the set of datasets where information related to the indexing theme T_i can be found in the repository. This is done through

the service *findDatasetsForTheme* defined in the semantic framework. Once the area of vision is obtained, the obtained datasets are integrated using the service *integrateDatasetsByTheme*. As a result, a new virtual dataset ds_i summarizing the information related to T_i in the source datasets will be created. This virtual dataset contains p values $\{C_{ij1}, \dots, C_{ijp}\}$, where i indicates the layer or indexing theme and j the number of frame, and where each of these values C_{ijk} is a reference to a thematic class in the ontology. If this set of values is different from the one from the previous frame, it indicates a change of segment, and consequently the previous segment is finished and a new one is created. All this process is shown in the segmentation algorithm just below, where index i is used for indexing themes and j for frames.

Segmentation algorithm

```

for each  $T_i \in \{T_1, \dots, T_n\}$  do
   $DsSet_i = findDatasetsForTheme(T_i)$ 
  //where  $DsSet_i$  is a set of  $m$  datasets  $\{ds_{i1}, \dots, ds_{im}\}$ 
  for each frame  $f_j$  in  $V$  do
    for each  $ds_{iu} \in DsSet_i$  do
       $dsA_{iuj} = \text{region of } ds_{iu} \text{ contained in } A_j$ 
    end for
     $ds_{ij} = integrateDatasetsByTheme(dsA_{i1j}, \dots, dsA_{imj}, T_i)$ 
    // where  $ds_{ij}$  has  $p$  values  $\{C_{ij1}, \dots, C_{ijp}\}$ 
    if  $\{C_{ij1}, \dots, C_{ijp}\} \neq \{C_{i(j-1)1}, \dots, C_{i(j-1)p}\}$  then
      if  $j \neq 0$  then
        finish segment at layer  $i$ 
      end if
      start new segment at layer  $i$ 
    end if
  end for
  finish segment at layer  $i$ 
end for

```

Each segment obtained through this algorithm has now to be indexed according to the set of thematic classes in its corresponding area of vision. The following tuple represents an index entry for a given segment at layer i :

$\langle V, T, sf, ef, C, area \rangle$

where V is the video (identified through its URI); T is one of the indexing themes; sf_i and ef_i are respectively the starting and ending frame of the segment being indexed, which belongs to the layer related to T ; C is one of the visible thematic classes at the segment; and $area$ is the average of the areas of the spatial extent of C at every frame in the segment. Note that the set of visible thematic classes at a given layer may be empty. In this case, the index entry would not be created.

The following algorithm includes both segmentation and indexing. We use index i for indexing themes, index j for frames, and index k for visible thematic classes in a frame or segment. The variable $areaC_{ik}$ is used to store the sum of the areas of the spatial extent of C_{ijk} at every frame j along the current segment at layer i , while the

variable sf_i stores the frame where the current segment at layer i started. Note that segments with an empty set of visible themes are not indexed.

Segmentation and indexing algorithm

```

for each  $T_i \in \{T_1, \dots, T_n\}$  do
   $DsSet_i = findDatasetsForTheme(T_i)$ 
  for each frame  $f_j$  in  $V$  do
    for each  $ds_{iu} \in DsSet_i$  do
       $dsA_{iuj} =$  region of  $ds_{iu}$  contained in  $A_j$ 
    end for
     $ds_{ij} = integrateDatasetsByTheme(dsA_{i1j}, \dots, dsA_{imj}, T_i)$ 
    // where  $ds_{ij}$  has  $p$  values  $\{C_{ij1}, \dots, C_{ijp}\}$ 
    if  $\{C_{ij1}, \dots, C_{ijp}\} \neq \{C_{i(j-1)1}, \dots, C_{i(j-1)q}\}$  then
      if  $j \neq 0$  then // finish segment at layer  $i$ :
        for each  $C_{i(j-1)k} \in \{C_{i(j-1)1}, \dots, C_{i(j-1)q}\}$  do
          add index entry
             $\langle V, T_i, sf_i, j-1, C_{i(j-1)k}, areaC_{ik}/(j-sf_i) \rangle$ 
        end for
      end if
      // start new segment at layer  $i$ :
       $sf_i = j$ 
      for each  $C_{ijk} \in \{C_{ij1}, \dots, C_{ijp}\}$  do
         $areaC_{ik} = |e(C_{ijk}, ds_{ij})|$ 
      end for
    else // continue segment at layer  $i$ :
      for each  $C_{ijk} \in \{C_{ij1}, \dots, C_{ijp}\}$  do
         $areaC_{ik} = areaC_{ik} + |e(C_{ijk}, ds_{ij})|$ 
      end for
    end if
  end for
  //finish segment at layer  $i$ :
  for each  $C_{ijk} \in \{C_{ij1}, \dots, C_{ijp}\}$  do
    add index entry
       $\langle V, T_i, sf_i, j, C_{ijk}, areaC_{ik}/(j-sf_i+1) \rangle$ 
  end for
end for

```

Since two consecutive frames usually have the same thematic information, a simple optimization can be done to this algorithm by only checking sets of themes each F frames. F is a parameter that the user may modify, and its default value is the same as the property fps . This indicates that, by default, a comparison is made for each second of video. If the comparison between frames f_a and f_{a+F} results in different sets of themes, f_a will be then compared with $f_{a+F/2}$. The distance between frames being compared is successively divided by two until the exact frame for the end of segment is found.

4. Queries

The main query that has to be implemented in this digital library is the retrieval of the video segments that depict a given thematic class T from the ontology. The implementation of this query consists in obtaining those video segments with one index entry containing a theme C subclass of T in its fifth element.

$$\langle \dots, \dots, \dots, \dots, C, \dots \rangle$$
$$C \sqsubseteq T$$

In case of several segments to be returned, they are also ordered according to a simple ranking algorithm that prioritizes those index entries that maximize the ratio between $areaC_{ik}$ and the average area of vision of the segment.

Another significant type of query is related to the playback of a whole source video sequence. For this video (first element of the index), and for each indexing theme or layer (second element), its corresponding index entries are retrieved and ordered according to the starting frame (third element). This produces a metadata structure of n layers that will be attached to the video in order for clients be able to interpret the related thematic information, as it will be described in the next section.

Finally, it is worth clarifying that this thematic-based approach does not exclude other information being used in parallel segmentations and indexes. For instance, another layer could be used to include manual annotations, which would determine a separate segmentation layer with its own indexing structure permitting non-thematic types of queries.

5. Video Meta-information

The video segments that are returned as the result of a query have to be provided with meta-information describing its thematic content. This way, clients can be developed to take profit of this meta-information. Spatial information will only be processed at the time of indexing the video sequence, and consequently, clients do not need to have access to geographic datasets and, since they do not have to deal with spatial data, can be more generic.

A video segment comprises only a layer of meta-information, containing its indexing theme and a list of visible themes (related to the indexing theme) with their spatial extent area. The segment meta-information also includes the camera properties of a representative frame, which will enable clients to find spatial relations between video segments according to their camera location. The following XML fragment exemplifies how the meta-information of a video segment is structured.


```

<VideoSegment video="..." indexingTheme="Ti"
  startFrame="s" endFrame="e" averageVisibleArea="...">
  <CameraPropertiesAtFrame frameNumber="...">
    <CameraPosition>
      <X>X</X>
      <Y>Y</Y>
      <Z>Z</Z>
    </CameraPosition>
    <CameraOrientation> $\rho$ </CameraOrientation>
    <CameraTilt> $\theta$ </cameraTilt>
    <CameraAngleOfVision> $\alpha$ </CameraAngleOfVision>
  </CameraPropertiesAtFrame>

  <VisibleTheme theme="C1e1"
    spatialExtentArea="area11" />
  <VisibleTheme theme="C1e2"
    spatialExtentArea="area12" />
  ...
  <VisibleTheme theme="C1ep"
    spatialExtentArea="area1p" />
</VideoSegment >

```

It has also to be noted that translations of the thematic classes to classes in a vocabulary understandable by the client may be needed. This would be done using one of the services of the semantic framework, *translateThemeToDataset*.

In the case of visualizing a complete video, the *Video* element contains a list of *VideoSegment* elements, as the following fragment of XML code exemplifies:

```

<Video uri="..." fps="fps">
  <VideoSegment video="..." indexingTheme="T1"
    startFrame="s11" endFrame="e11"
    averageVisibleArea="...">
    ...
  </VideoSegment >
  <VideoSegment video="..." indexingTheme="T1"
    startFrame="s12" endFrame="e12"
    averageVisibleArea="...">
    ...
  </VideoSegment >
  ...
  <VideoSegment video="..." indexingTheme="T2"
    startFrame="s21" endFrame="e21"
    averageVisibleArea="...">
    ...
  </VideoSegment >
  ...
</Video>

```

The schema for the XML documents appearing in this section can be found in [1], where an analysis on how this meta-information can be represented according to some metadata standards is also provided. This will enable more generic clients to be developed. In the specific case of MPEG-7 [20], the meta-information schema is

based on 1) defining a semantic world for each video segment through the *semantic* element, 2) representing thematic classes through the *semanticBase* element, and 3) defining a graph of semantic relations between video segments and its corresponding thematic classes.

6. Future Work

At the moment of writing this paper, we have conducted some initial tests of the digital video library. The segmentation and indexing algorithms have been implemented as a plug-in of a GIS. In these tests, for simplicity reasons, we have used virtual video sequences obtained from a 3D environment. To do so, we have developed a tool that uses a 3D digital elevation model (with the orto-photo as texture) as a virtual world. Once the user draws a path on a map, the camera is moved in the 3D world according to the path, generating both a video sequence and its geo-reference information.

Apart from a deeper evaluation with “real” videos, our plans for future work are mainly focused on exploring how the results of a query can be better presented in a more usable way.

Storyboard is the most widely used metaphor for showing the results of searches. Following this metaphor, a video is represented by means of a set of temporarily ordered shot’s key-frames that are simultaneously presented on screen as thumbnails. This type of interface avoids downloading big amounts of video data and provides a quick general view of the video. In the case of queries returning several video segments, each one is represented by means of a thumbnail of one of its frames. Storyboards are used in the majority of video retrieval systems. For instance, [21] describes this type of interface in Informedia, and [22] in the Físchlár Digital Video Library. These interfaces also provide more elements as a relevance indicator or a title. However, storyboards may be too long and thus, become useless. Furthermore, they do not provide relations between the returned segments.

To partially avoid these problems, another type of interface was developed, called video collage. It aims at visualizing news video segments that results from a query, together with their respective contexts in Informedia [23, 24]. A video collage contains a rectangular panel where news video segments are placed by means of thumbnails, as well as text lists showing the “who,” “what,” “when,” and “where” information of the selected video segments. The panel may be organized either spatially or temporarily. In the spatial distribution, segments are placed on a map according to the location names that they contain. In the temporary distribution, video segments are organized according to two axes: the vertical axe is the relevance of the news video segment in respect to the query, while the horizontal axe corresponds to the date when the news were broadcasted. The user may change the granularity of the panel in both cases in order to properly visualize crowded spatial regions or time periods.

For our system, we would like to explore a different approach that goes beyond a retrieval system aiming at a hypermedia presentation generator. This way, given a thematic query, the system returns not just a list of resulting video segments, but a

hypermedia presentation dynamically generated. This presentation contains two main parts: on the one hand, a video sequence dynamically generated from a selection among the resulting video segments, and on the other hand, a textual information describing the thematic information attached to video segments and suggesting links to other information or segments. We call *video-itinerary* to this dynamic presentation, since the result of searching a certain theme is a kind of visual itinerary along the locations where this theme is present. How video-itineraries are dynamically generated and how they can be navigated to build a hypervideo network are the two key issues here.

Regarding the first issue, the generation of the video-itinerary, the most relevant research topic is how the composition of segments can provide a narrative thread to the final sequence. The two main semiotic elements of films, *mise-en-scène* (what can be seen in a shot) and montage (how shots are sequenced, also called editing) influence this process. In our case, *mise-en-scène* is described through the thematic information associated to segments, but the way they are selected and temporarily organized may provide extra narrative elements to the final sequence. Montage treatises contain different techniques that have to be considered for giving continuity to sequences and providing them with the desired narrative (for instance temporal or spatial change). Although techniques based on sound and on characters (for instance in a dialogue sequence) cannot be applied in our context, other techniques could be considered. Particularly relevant for our context is documentary montage theory. In our case, montage may be influenced by different factors, including the relevance of the searched theme in the obtained video segments, or the location of the camera in video segments in order to group them according to spatial criteria providing the feeling of an itinerary. In this context, the user may also be interested in specifying an expected duration for this video and a type of rhythm, factors that would also have influence on the selection and composition of segments. Detection of zoom operations or other typical camera movements could also be considered.

The other relevant research issue is navigation, which is determined according to two main axes: thematic information and spatial location. This way, links can be automatically created from a segment to other segments containing related geographic themes or to segments that belong to the same spatial area. Other spatial relations between segments could also determine links, like segments closer to a certain distance or even those with a location that can be reached following a certain direction (North, South, East or West) from the current segment. The result is a hypervideo network that is dynamically built.

References

1. Navarrete, T. Semantic Integration of Thematic Geographic Information in a Multimedia Context. PhD Thesis. Departament de Tecnologia, Universitat Pompeu Fabra. Barcelona. 2006
2. Nobre, E.M.N. and A.S. Câmara. Spatial Video: exploring space using multiple digital videos. in 6th Eurographics Workshop on Multimedia. 2001. Manchester, UK: Springer-Verlag.

3. Wulder, M., J. White, and S. McDonald, Truth in the Air. Geolocated Video Validates Satellite Land Cover, in GeoWorld. 2005.
4. Yoo, J., et al. Vehicular Image Based Geographic Information System for Telematics Environments. Integrating Map World into Real World. in IEEE International Geoscience and Remote Sensing Symposium. 2005: IEEE.
5. Lee, S., et al., 4SVan: A Prototype Mobile Mapping System for GIS. Korean Journal of Remote Sensing, 2003. 19(1).
6. Kim, S.-S. and J.-H. Park. Geographic hypermedia using search space transformation. in International Conference on Pattern Recognition. 2004.
7. Kim, S.-S., K.-H. Kim, and K.-O. Kim, Web-Based Media GIS Architecture Using the Virtual World Mapping Technique. Korean Journal of Remote Sensing, 2003. 19(1): p. 71-80.
8. Kim, S.-S., et al. A unified visualization framework for spatial and temporal analysis in 4D GIS. in IEEE International Geoscience and Remote Sensing Symposium. 2003: IEEE.
9. Hwang, T.-H., et al. MPEG-7 metadata for video-based GIS applications. in IEEE International Geoscience and Remote Sensing Symposium. 2003: IEEE.
10. Lippman, A. Movie maps: An application of the optical videodisc to computer graphics. in Computer Graphics (SIGGRAPH). 1980.
11. Negroponte, N., Being digital. 1995, p. 65-67. London, UK: Hodder and Stoughton.
12. Openshaw, S., C. Wymer, and M. Charlton, A geographical information and mapping system for the BBC Domesday optical disks. Transactions of the Institute of British Geographers, 1986. 11: p. 296-304.
13. Shiffer, M.J., Towards a Collaborative Planning System. Environment and Planning B: Planning and Design, 1992. 19: p. 709-722.
14. Blat, J., et al., Designing multimedia GIS for territorial planning: the ParcBIT case. Environment and Planning B: Planning and Design, 1995. 22: p. 665-678.
15. Raper, J., Multidimensional Geographic Information Science. 2001, London, UK: Taylor & Francis.
16. Aguierre Smith, T.S. If you could see what I mean. MS Thesis. Massachusetts Institute of Technology. Cambridge, Massachusetts, USA. 1992
17. Davenport, G., T.S. Aguierre Smith, and N. Pincever, Cinematic primitives for multimedia. IEEE Computer Graphics & Applications, 1991.
18. Navarrete, T. and J. Blat. VideoGIS: Segmenting and indexing video based on geographic information. in 5th AGILE Conference on Geographic Information Science. 2002. Palma de Mallorca, Spain.
19. Navarrete, T. VideoGIS: Combining Video and Geographical Information. DEA (Master) thesis. Departament de Tecnologia, Universitat Pompeu Fabra. Barcelona. 2001
20. ISO, Multimedia Content Description Interface. 2000, International Organization for Standardization JTC1/SC29/WG11.
21. Christel, M. Evaluation and User Studies with Respect to Video Summarization and Browsing. in Multimedia Content Analysis, Management and Retrieval 2006, part of the IS&T/SPIE Symposium on Electronic Imaging 2006. 2006. San Jose, California, USA.

22. Smeaton, A.F. and H. Lee, Designing the User Interface for the Físchlár Digital Video Library. *Journal of Digital Information*, 2002. 2(4).
23. Ng, T.D., et al. Collages as Dynamic Summaries of Mined Video Content for Intelligent Multimedia Knowledge Management. in *AAAI Spring Symposium Series on Intelligent Multimedia Knowledge Management*. 2003. Palo Alto, California, USA.
24. Christel, M., et al. Collages as Dynamic Summaries for News Video. in *ACM Multimedia*. 2002. Juan-les-Pins, France.