# A Generation Method of Spatially Encoded Video Data for Geographic Information Systems

In-Hak Joo, Tae-Hyun Hwang, Kyoung-Ho Choi, and Byung-Tae Jang
Spatial Information Technology Center,
Electronics and Telecommunications Research Institute,
161 Gajeong-dong, Yuseong-Gu, Daejeon, 305-350, Korea
{ ihjoo, hth63339, khchoietri, jbt }@etri.re.kr

**Abstract:** In this paper, we present a method for generating and providing spatially encoded video data that can be effectively used by GIS applications. We collect the video data by a mobile mapping system called 4S-Van that is equipped by GPS, INS, CCD camera, and DVR system. The information about spatial object appearing in video, such as occupied region in each frame, attribute value, and geo-coordinate, are generated and encoded. We suggest methods that can generate such data for each frame in semi-automatic manner.
We adopt standard MPEG-7 metadata format for representation of the spatially encoded video data to be generally used by GIS application. The spatial and attribute information encoded to each video frame can make visual browsing between map and video possible. The generated video data can be provided and applied to various GIS applications where location and visual data are both important.
**Keywords:** Video GIS, 4S-Van, MPEG-7, Object tracking

## 1. Introduction

One of the recent trends of GIS(Geographic Information System) is to provide realistic and human-perceptible information of spatial objects by supporting media other than conventional map [1]. Many GIS applications have introduced video data because it greatly helps semantic recognition about space.

However, video data has no explicit spatial information by itself. Therefore we cannot get any information of spatial objects from video unless the additional information is included in video in whatever manner.

With such information, selection of an object on a frame of video can get the spatial and attribute information. Conversely, spatial and attribute information of an object can retrieve video frame in which the object appears and the location in the frame. For such inter-media operations, spatial information should be included in the corresponding parts of video data.

In this paper, we present a method for generating and providing spatially encoded video data that can be effectively used by GIS applications. In the following sections, we present about video data collection, object tracking from video, and coding scheme for obtained data to provide video-map interoperation.

## 2. Video Data Collection

We adopt a mobile mapping system called 4S-Van that has developed in our previous research [2], to collect video data and corresponding location. Mobile mapping system is a moving platform where many sensors are integrated that obtains 3-dimensional coordinates of spatial objects. It usually equips sensors such as GPS, INS, and cameras in a vehicle. It collects image/video data by camera at the same time when GPS and INS data is collected to determine position and attitude of the vehicle. Mobile mapping systems can construct spatial information fast and accurately from the collected image/video and position/attitude information. Fig. 1. shows main equipments and application areas of the 4S-Van.

The original role of 4S-Van is calculation of geo-coordinates of spatial objects without actual survey. We also utilize 4S-Van to collect video data that will be processed by our suggested scheme.

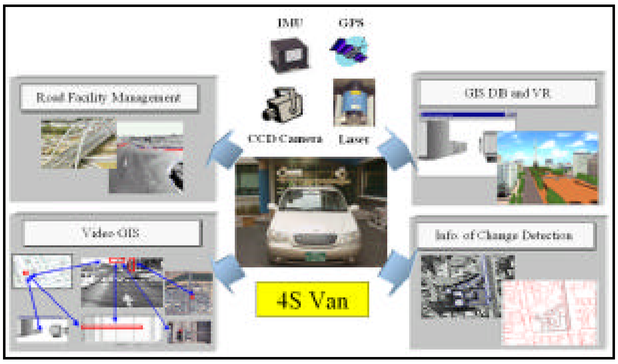


**Fig. 1. 4S-Van: equipment and application areas.**

Due to current configuration of hardware of our 4S-Van, we consider the two cases in capturing video data: (1) video is captured by video camera and DVR, frame rate is 20 frame/sec or higher, image resolution is lower, and accurate position and attitude of camera are not provided; (2) video is captured by CCD camera, frame rate is 1 frame/sec or lower, image resolution is higher, and accurate position and attitude of camera synchronized with each image are provided. We will discuss the two cases in the next section.

## 3. Object Tracking from Video Frames

For the provision of information of spatial objects on the video frame, the information should be included in the corresponding parts of video data. In detail, we need a method for extracting outline of objects in each frame, and a method for coding the obtained information into video data. Because an object appears in many frames, automatic or semi-automatic object recognition or tracking is necessary for the object information to be recorded in each video frame without time-consuming user input. In this paper, we concentrate on object tracking and do not cover automatic object recognition.

There are many object tracking methods suggested based on image processing [3]. However, most of conventional object tracking methods are inadequate for video frames acquired by camera on a running vehicle. In our work, interested objects are buildings and roadside facilities. However, the non-interested objects such as pedestrians, vehicles, and backgrounds are moving as well as interested objects in video data collected by running vehicle, which makes the object tracking more difficult. Moreover, in real scene, the objects are likely to be distinguished from background not clearly.

One of the object tracking method that can be applied to images with background moving together is block matching algorithm, which analyzes characteristics of pixel value to track objects. Though it can be used for such image sets, it cannot match object if the object moves rapidly and search area is not large enough.

We adopt the block matching algorithms as basic object tracking method, and support methods for reduce search area by exploiting camera parameters, which is the main issue about this algorithm. Because the obtainable information and image characteristics are quite different for two cases mentioned in the previous secion, we devise solution for two cases separately.

**1) Video frames with 20 frame/sec or higher frame rate**

For this type of video frames, accurate position and attitude of camera are not provided, while the object in each frame moves very slowly that make object tracking easy.

We assume that 4S-Van moves straight in almost constant velocity. Note that, in actual, the assumption holds in relatively short period of a video sequence. Further, we assume that the user designate the region information of an object for the first frame in which the object appears. Then the object tracking algorithm is applied to the following frames. It means that fully automatic object recognition is not considered in our work.

Because object moves slightly in adjacent frames, the search area for block matching algorithm is relatively small. Further, because the moving direction and distance are empirically predictable with the assumption above, the search area can become smaller. In our brief experiments, in most cases, we can easily find same object in the next adjacent frame. For unconsidered case such as rapid turn at crossroad, the method may not work, which is remained further study for improvement.

**2) Image sets with 1 frame/sec or lower frame rate**

When image sets are collected with 1 frame/sec, the object moves more rapidly from an image to next image. Because search area should be larger, the object tracking becomes more difficult and more likely to fail.

Instead, we can use the exact position and attitude of camera corresponding to each image. With the camera parameters, we can calculate 3-dimensional world coordinate of a spatial object from the pixel coordinate corresponding to the object. Again, the 3-dimensional world coordinate of the spatial object can be converted to pixel coordinate of next adjacent image, as long as the camera parameters provided for stereo imges. These two kinds of conversion method are essential functions of 4S-Van software. The algorithm is as follows.

```
In the image an object first appears
    Select an object and designate outline of the object, rep-
    resented as set of pixels
    For each pixel constituting the outline
        Convert it to world coordinate
For next images, repeat
    For each world coordinate constituting the outline
        Convert it to pixel coordinate
        Determine search area near the pixel (with an empiri-
        cal threshold)
        Use block matching method to find corresponding
        pixel
While the object disappears
```

With the conversion methods applied, though not exact to pixel, we can predict the position on which the object appears in any consequent images. It means the search area for block matching method can be reduced. This method also has room for improvement. Main issue about these methods is more accurate prediction, which comes from accurate 4S-Van data.

## 4. Data Encoding

To provide the spatially encoded video data to various applications, the data are encoded by standardized format. We adopt XML-based MPEG-7 format [4] and design a schema by adding user-defined elements to include necessary information. Fig. 2. is a block diagram for the generation and encoding of MPEG-7 metadata using the suggested method.

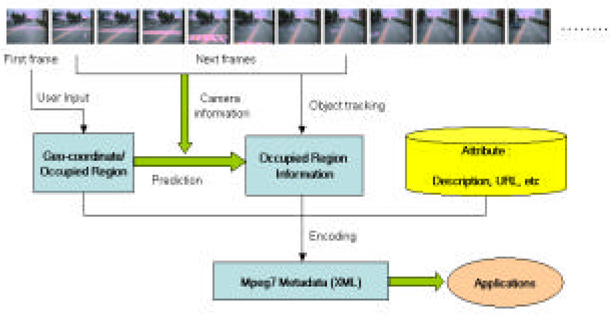


**Fig. 2. Block diagram for the generation and encoding of MPEG-7 metadata.**

We design contents of metadata (Fig. 3.) that can be divided into two parts: (1) video header that contains basic information about video as a whole, such as date, time, creator; (2) frame information that contains object information such as geo-coordinates, occupied region, keywords, and related URL for each frame.

```
For a video sequence
  General Infomation
    Version, Identifier, Creator, CreationLocation, In-
    strument, Rights
  Media Acquisition Information
    Date, Time, Weather, CoordinateSystem, MediaName, Me-
    diaID, FreeTextAnnotation
For each frame
  Annotation for Parameters
    MediaTime, GeoCoordinates, FrameKeyword, Keyword
  Annotation for Crossroads
    CrossRoadID, CrossFrameNo
  Annotation for Object
    ObjectID, ObjectName, Region, PointListNumber, PosX,
    PosY, URL, ObjectKeyword, FreeTextAnnotation
```

**Fig. 3. Contents of video metadata description.**

Fig. 4. shows part of example of MPEG-7 metadata. The most important information among the metadata is <region> element that contains data of object's occupied region in pixel coordinate. It can be created object tracking method shown above, and plays an essential role in video-map two-way search. Another important element is <ObjectKeyword>, which enables user to search objects appearing in video by keyword. <URL> element has URL related to building, by which the applications can directly connect Web page of company, bank, or restaurant located at the building. <Crossroad> element is for future extension of representation of crossroad model. Note that, any redundant element between frames, for example <ObjectName>, is included only in a key frame and described as a link for other frames.


```
<?xml version="1.0" encoding="euc-kr" ?>
<Mpeg7>
<DescriptionMetadata>
    <Version>0.95</Version>
    <PublicIdentifier />
    <PrivateIdentifier>VideoGISDescriptionExample</PrivateIdentifier>
    <Creator>Tae-Hyun Hwang</Creator>
    <CreationLocation>Daejon,305-350, KOREA</CreationLocation>
    <Instrument>
      <Tool>MPEG-7-GEN ver. 0.5</Tool>
    </Instrument>
    <Rights />
</DescriptionMetadata>
<ContentDescription>
<VideoGIS>
<AcquisitionInformation>
     <DateAndTime>2003-04-20T13:20:25+09:00</DateAndTime>
     <Weather />
     <CoordinateSystem>longitude and latitude</CoordinateSystem>
     <MediaName>etri.mpg</MediaName>
     <MediaID>r001234</MediaID>
     <FreeTextAnnotation> any comment </FreeTextAnnotation>
</AcquisitionInformation>
<TemporalDecomposition>
<VideoGISSegment>
    <ParametersAnnotation>
    <MediaTime>
      <MediaTimePoint> 000001 </MediaTimePoint>
    </MediaTime>
    <GeoCoordinates>
      <CoordX>131.1254</CoordX>
      <CoordY>34.1254</CoordY>
    </GeoCoordinates>
    <FrameKeyword>
      <Keyword> crossroad </Keyword>
    </FrameKeyword>
    </ParametersAnnotation>
    <CrossroadsAnnotation>
    <Crossroads>
      <CrossRoadID />
      <CrossFrameNo />
    </Crossroads>
    </CrossroadsAnnotation>
    <ObjectAnnotation>
    <Objects relatedFrame="none" relatedOjbetID="none">
      <ObjectID>obj001234</ObjectID>
      <ObjectName> ETRI main building </ObjectName>
    <Region>
      <PointListNumber>4</PointListNumber>
      <PointList>
      <PosX>125</PosX> <PosY>254</PosY> <PosX>125</PosX> <PosY>54</PosY>
      <PosX>25</PosX> <PosY>54</PosY> <PosX>25</PosX> <PosY>254</PosY>
      </PointList>
    </Region>
    <URL>http://www.etri.re.kr</URL>
    <ObjectKeyword>
      <Keyword>5+ story building</Keyword>
      <Keyword>research orgarnization</Keyword>
    </ObjectKeyword>
    <FreeTextAnnotation> any comment </FreeTextAnnotation>
    </Objects>
    </ObjectAnnotation>
</VideoGISSegment>
</TemporalDecomposition>
</VideoGIS>
</ContentDescription>
</Mpeg7>
```


**Fig. 4. Example of MPEG-7 metadata (part).**

# 5. Conclusions

In this paper, we present a method for generating and providing spatially encoded video data that can be effectively used by GIS applications. The information about spatial object appearing in video, such as occupied region in each frame, attribute value, and geo-coordinate, are generated and encoded. We suggest methods that can generate such data for each frame in semi-automatic manner, by adopting object tracking method.

To provide generally usable video data, we adopt standard MPEG-7 metadata format for representation and encoding of video data. With the information encoded to each video frame, selecting a spatial object in any frame can get the location and attribute of the object. Conversely, specifying a spatial object can retrieve a video frame where it appears.

The video and metadata generated by the suggested method can be effectively used by GIS application where location and visual data are both important, for example, tourist information services. The suggested technology will be also used for applications based on Web or mobile environments

# References


[1] Jae-Jun Yoo, In-Hak Joo, Kwang-Woo Nam, and Jong-Hoon Lee, "The Design and Implementation of Video Geographic Information System," Proc. 29th KISS Fall Conference, Korea, Oct 2002, pp. 274-276.

[2] A study on Application for 4S-Van, Seung-Yong Lee, Eun-Young Han, and Byoung-Woo Oh, Proc. ISRS(International Symposium on Remote Sensing) 2001, Nov 2001, PP.124-127.

[3] F.Marques and C. Molina, "Object tracking for content-based functionalities," in SPIE Vis. Commun. Image Processing, vol 3024, San Jose, CA, Feb. 1997, pp. 190-199.

[4] ISO/IEC JTC 1/SC 29/WG 11: "Information Technology-Multimedia Content Description Interface- Part 5: Multimedia Description Schemes," MPEG document N4005, Singapore, March, 2001.