AUGMENTING LIVE AERIAL VIDEO IMAGES WITH GIS INFORMATION TO ENHANCE DECISION MAKING PROCESS DURING EMERGENCIES

Maider Laka-Iñurrategi, Julen Garcia, Kevin Alonso, Marco Quartulli

Vicomtech-IK4, Mikeletegi pasealekua 57, 20009 Donostia, Spain

{mlaka, jgarcia, kalonso, mquartulli}@vicomtech.org

Abstract

This paper presents a framework for realtime airborne imagery and video stream augmenting by direct georeferencing and GIS layer live projection. It is based on the combination of global and inertial navigation systems, GIS and 3D computer graphic techniques. A tool for enhancing visually the aerial video images with georeferentiated content contributes effectively during the decision making process at Emergency Management Centers (EMC). Furthermore, the broadcasting of the augmented video is also considered in order to inform the details of the emergency to the affected population. On the other hand, the storage of georeferenced videos allows location based video retrieval for subsequent analysis of the emergency attendance.

Keywords*: Emergency Management, GIS, Digital video, Augmented Reality, Semantic Annotation.*

1. INTRODUCTION

The use of Geographic Information Systems (GIS) within Emergency Management (EM) has enhanced the ability of practitioners to respond and manage natural and human induced hazards in a more comprehensive fashion than ever before [1]. In fact, GIS offer more efficient and fast data input, management, manipulation, geospatial analysis and output information such as cartography what is of high value in decision making tasks. Therefore, GIS effectively provides value in all four phases of the disaster management cycle: preparedness, response, recovery and mitigation (Fig. 1).

The requirements of EM tools significantly depend on the phase they are used. The response phase is highly sensitive to time, since events happen quickly and real time information is crucial for decision makers. The rest of the phases take place before and/or after the emergency occurs so the decisions are based on analytical studies of the data gathered over the time.

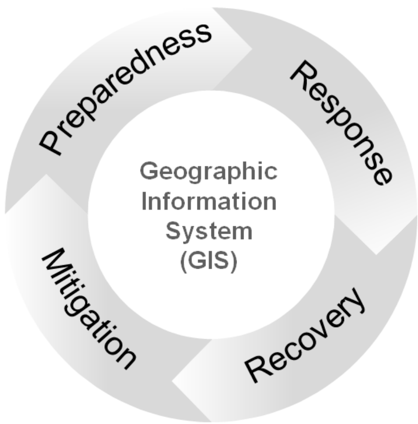


Fig. 1 Emergency Management cycle and Geographic Information System [1].

The objective of this paper is to present a system architecture and functional design to augment live video with GIS information alongside numerical and visual geo-analysis information, in a real time highly demanding environment such as Emergency Management (EM) and with reduced costs with respect to existing specialized solutions. This is done by drawing from existing developments in the fields of multimedia production, computer graphics, direct georeferencing in remote sensing and GIS management. Overlaying in real time localization-related information to the live videos can contribute to more effective search, rescue, evacuation and coordination processes in the response phase of emergency management. Moreover, storing and aggregating semantically annotated videos will allow EM centers to perform in-depth data analysis tasks required in phases as mitigation and preparedness.

2. STATE OF THE ART

The first step in an emergency situation often is to obtain as much information as possible concerning the area involved in the event: : localization, damage and residual risk extent (e.g. areas with endangered/destroyed buildings), state of the transport network (e.g. best routes to reach the event), etc. Supplying decision-makers with raw information streams on the unfolding events is a very common practice which generally yields negative results [2]. Practitioners need to assimilate all the aspects of the event in order to take fast and suitable decisions, and the interpretation of raw geographic information may last too much. In fact, successful strategies depend on the availability of accurate information presented in an appropriate and timely manner. Therefore, the technologies used in Emergency Management for successful mitigation, response, preparedness and recovery are mainly based on Geographic Information Systems.

The first step is to identify the localization of the involved area. In response phases, this might be estimated through the identification of warning calls by triangulating the cell-phone area or obtaining the GPS information from calling mobile terminals [7]. This crucial information is very sensitive to the network infrastructure that can be damaged or overloaded. Initiatives are in place (SANY, ORCHESTRA or OSIRIS European projects) supported by the Open Geospatial Consortium (OGC) to establish a dynamic management of heterogeneous Wireless Sensor Networks (WSN) for full situation awareness. All sensors would report position via the web and with metadata registered. These networks would provide more complete information to the EM Centers and would assure the information even in situations where common infrastructures would be temporarily inoperable.

Once the localization of the event is obtained, decision-makers need to understand the scope of the emergency and the secondary damages, monitor the event, plan the response, etc. For that purpose direct video captures of the events would be highly helpful. Yet, available remote sensing technology suffers for stringent requirements on data quality and for the bleeding edge novelty of sensors. Remote sensing data at video data rates is only available in very specialized situations involving high cost geographic surveillance solutions platforms e.g. remotely controlled aerial vehicles in military settings.

The interpretation of video images captured by the helicopters or other sources is not always straightforward; it is often not easy to understand which road, which side of the forest or which town is being displayed on the situation assessment screens back at the EMC. Therefore, a need to georeference the incoming video signals arises. Geo-localization of standard fixed-time media such as photographs is currently commonplace in a number of applications targeted to end users. For time-varying media, geographic tagging of videos is increasingly becoming a powerful addition to the standard set of metadata contributing to semiautomatic characterization and discovery of content.

Direct Georeferencing (DG) [3] is the process of assigning the exterior orientation and localization of each image frame or scan line directly, without the need to use traditional aerial triangulation techniques. Currently this is achieved by measuring the geographic position and orientation of the airborne imaging sensor directly using a navigation device such as Global Positioning System (GPS) and Inertial Navigation System (INS) technology. GPS provides sufficiently accurate measurements of position, while an INS computes a full position and orientation solution. When combined into a single solution, the accurate GPS position measurements are used to control the errors in the INS, effectively helping it or "aiding" it to have a higher accuracy than it could on its own. Such a system is often referred to as a GPS-Aided INS.

The real-time performances currently attainable in DG mean that continuous variations of reference frame can be coped with by effective tracking of relative motion between sensor and scene in a number of dimensions effectively allowing Direct Video Georeferencing.

The localization and orientation of the videos displayed on a situation assessment monitor at a remote location still is not enough for managers to effectively evaluate the state of an unfolding emergency event. GIS layer overlaying can considerably enhance the interpretation of the raw video. Moreover, real-time updated GIS layers containing the output of a semantic analysis of the input images might contribute to a better planning in the decision making phase.

Normally GIS layers are shown in nadir view. Overlaying GIS layers to video requires a proper transformation of the input layers in geographic coordinates according to the point of view of the camera and based on elevation information of the visible area in order to make it possible to render them in pixel coordinates. In the case of vector layers, vector to raster transformations are needed as an additional step. The result is a video stream augmented with cartography features: a map-like video stream that if produced in near-real time can be accessed by situation assessment personnel as support content for efficiently mapping and monitoring disaster situations [4]. A further application for a system based on the presented architecture is population management in a large scale emergency such as a large scale fire: disasters like the Victoria 9th February 2009 fires in Southern Australia that caused 173 fatalities in less than eight hours raise the need to have better ways to communicate and handle population that can be concentrated in urban settings or disseminated in a wide area, to avoid such losses. Communication network availability and proper message broadcasting to the general public represents a central concern.

A significant contribution is in [5]. The paper notes that in an emergency situation, the presence of a telecommunications-based (at the most basic level, a telephone-based) support service reduces stress levels, although priority must always be given to emergency services in the use of the telephone/wireless network. When essential communication services are down, coordination and activation of rescue operations in a disaster situation can become very difficult.

A system based on the architecture described in this contribution might be used to effectively prepare multimedia content to be broadcasted on television networks in order to provide the general public with timely and useful information about an unfolding event.

A further approach with respect to projecting GIS layers over a video stream in realtime is to project the video stream as a texture over a 3D model together with other selected GIS vector/raster layers. The main difference between the two approaches is the aspect of the obtained video images, since the second approach distorts the video stream and may complicate the effective visual understanding of the images.

3. SYSTEM OVERVIEW

Focused realtime remote sensing is of significant value in Emergency Management: the fastest way to reach a specific area often is a helicopter equipped with a camera that is sent to the interest point so that the command chain can take further action depending on what they see on situation assessment screens. In order to quicken decision making tasks, processing and visualization tools that contribute to fast data assimilation and comprehension are highly helpful.

This paper proposes a framework that enhances input live images with external information stored in a GIS concerning the emergency or nearby locations. It efficiently renders vector and raster GIS layers into a realtime video image, allowing users to integrate visual information related for example to the state of the communication and transport networks, to population density and to dangerous ongoing events in a single video stream that can be broadcasted with standard means such as wireless networks or digital television links. The architecture can also be extended to foresee the development, performing realtime geographic and spatial analysis in order to complete the information shown on the screens in a command and control center. In fire event monitoring, these tools would be able to track the evolution of the area covered by the fire, notify if nearby protected regions are in serious hazard, etc. With all these tools the integration of diverse information over live images with the aim of helping the response phase of an emergency management cycle can be obtained.

Other phases of the cycle might use a system based on this architecture as well: Emergency Management Centers can also use them as information sources during the design stage of the preparedness plans. The intelligent retrieval of the stored data can be based on extended metadata including for example the localization of the video images computed during acquisition.

GIS layer projection can be carried out according to the position and orientation of the camera based on an existing digital elevation model (DEM). This projection task is really a key point in this application, yet the real time rendering and the video georeferencing processes are equally challenging. The helicopter camera moves at high velocities and therefore the processing has to take place in strictly controlled time windows.

In summary, the proposed framework considers real time geographic data acquisition and fusion for addressing the needs of intervention communities at large and for communicating with the general public via effective video production and broadcasting. In particular, the framework includes 1) tools for prior knowledge access in the form of available vector and raster geographic data archives, 2) tools for the acquisition and georeferencing of real time video streams produced on the intervention area by e.g. drones, 3) tools for the efficient assessment of current situations in semantic terms, 4) tools for the semantic integration of these different information contents referring to respectively past, present and future situations.

4. ARCHITECTURE

In this section the architecture of the proposed framework is described (Fig. 2). This system is composed by four main modules:

1. Video and Position Capture System: the system for capturing remote video images along with instantaneous camera parameters in the form of metadata required for the georeferencing stage: position, orientation and lens parameters.

2. Realtime Georeferencing and Video Analysis Module: based on the camera parameters and a digital elevation model (DEM) of the involved area, it calculates the field of view of the camera and generates the geolocalization information for georeferencing the video images. Then, it can integrate different video analysis modules to enhance the information available to the EMC.

3. Realtime GIS Layer Projection Module: required GIS layer projection is performed according to the position and orientation of the camera and based on a DEM.

4. Video Output Generator: the input video, the geolocalization information and projected GIS layers are combined by this module in an augmented reality video image stream e.g. for assessment at a remote control and command center as well as to distribute them via streaming or broadcasting to the affected population. This module also stores the generated data for future exploitation.

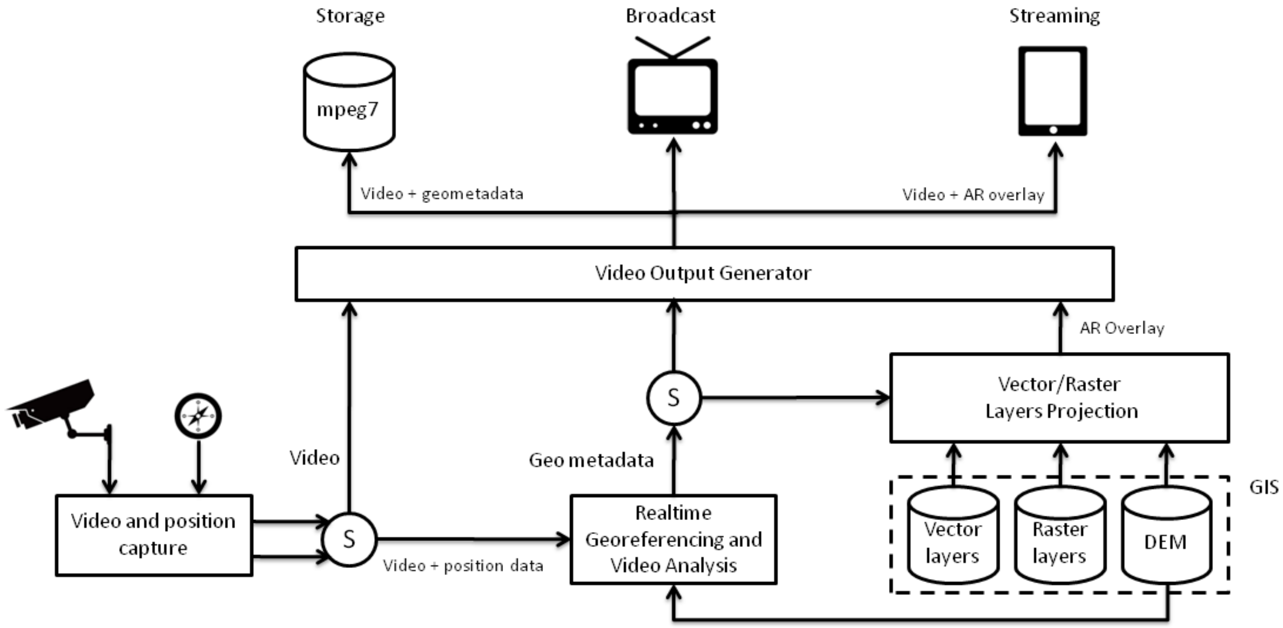


Fig. 2 Architecture of the proposed framework for augmenting and georeferencing aerial video imagery in realtime with GIS data.

1. Video and Position Capture System

The proposed architecture includes a video camera equipped with a GPS-Aided INS that can be used for Direct Georeferencing of aerial video imagery and is typically comprised of four main components: an Inertial Measurement Unit or IMU, a dual frequency low-noise GPS receiver, a real-time computer system and a post-processing software suite. The heart of the system is the Integrated Inertial Navigation software that is implemented both in real-time on the computer system and in postmission using a module of the post-processing software suite. The software implements the signal processing algorithms that blend the GPS measurements with the inertial navigation solution. As a result it is produced a position and orientation solution that retains the dynamic accuracy of the inertial navigation solution but has the absolute accuracy of the GPS. Solutions based on other GNSS systems such as Galileo are also possible.

The inertial navigator solves Newton's equations of motion on the rotating Earth by integrating acceleration and angular rates sensed by the IMU. In order to do this, the inertial navigator must first be initialized with known position and velocity from the GPS, and aligned with respect to the true vertical and true North. Once aligned the inertial navigator has established a local-level mathematical frame of reference called the navigation frame, whose heading is known with respect to North, and to which the orientation of the IMU is known, as shown in Fig. 3.

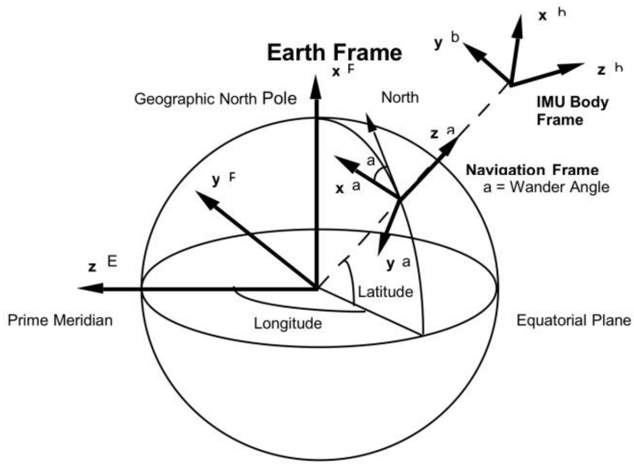


Fig. 3 Frames of the references used in Inertial Navigation System (INS).

After removing the rotation rate of the Earth (computed as a function of position), the navigator integrates the incremental angles from the IMU to continuously compute the change in orientation of the IMU with respect to the navigation frame. It then uses the orientations to resolve the incremental velocities from the accelerometers into the local-level navigation frame, which integrates to compute the position change of the navigation frame over the Earth.

In order to accurately compute the ground coordinates of a point using Direct Georeferencing, a number of requirements need to be met. First, the IMU needs to be rigidly attached directly to the camera to ensure there is not flexure between the camera perspective center and the sensing center of the IMU. Then, the physical misalignments of IMU with respect to the Camera need to be calibrated. The lever-arm offsets from the camera perspective center to the IMU and to the GPS antenna need to be calibrated and the exact time of image exposure needs to be recorded by the GPS-Aided INS so that the navigation parameters at that time can be computed. The position and orientation from the GPS-Aided INS navigation data need to be transformed to the earth observing (EO) trace used by most photogrammetric plotters. This usually involves a transformation from the geographic frame to a local level user frame, which means the appropriate transformation parameters must be known or computed. Finally, the camera interior geometry (principal point, lens distortion, focal length) must be well calibrated and stable.

2. Realtime Georeferencing and Video Analysis Module

A Real Time Georeferencing and Video Analysis Module needs to be implemented to compute the field of view of the moving camera platform and the geolocalization of the video images. This last process is based on the camera video images, on the associated metadata and on a digital elevation model (DEM) of the region of interest.

The georeferencing task is based in the calculation of the camera field of view. The field of view, in advance FOV, is a very common term in the optical instrumentation industry. The FOV represents the part of the world visible through the camera. Therefore, objects outside the FOV are not recorded by the camera. For a lens used to form a rectangular frame, three fields of view are needed to describe exactly the region that the camera is capturing; the horizontal FOV, the vertical FOV and the diagonal FOV.

Once the FOV is known the next step of the Georeferencing Module is to calculate the latitude, longitude and elevation in the central and edges points of the video image. Knowing the position of the real camera and its FOV, the geographic values of the images are extracted from the digital elevation model of the involved area using 3D world ray tracing techniques.

Apart from georeferencing task, the module is designed to be able to integrate different kind of video analysis elements that would enhance the semantic annotation capabilities and therefore the posterior richness of the semantic searches (e.g., fire front segmentation).

On one hand, the FOV calculation and the camera position and orientation are key input parameters in the Vector/Raster Layer Projection Module. On the other hand, the values of latitude, longitude and elevation will be used by Video

Output Generator Module along with the other metadata, obtained from visual analysis elements, to add location related annotations to the video stream.

3. Realtime GIS Layer Projection Module

Realtime GIS layer projection is the module that takes the raw GIS data and processes to assemble/render a camera's viewpoint matching overlay. This computationally intense task can be done in realtime due to 3D Scene Graph used beneath. The 3D Scene Graph should be based on OpenGL that uses the graphic card among other performance increasing features to reduce the processing time (such as OpenSceneGraph).

The camera-consistent projection of the GIS layers can be carried out according to the real position and orientation of the input camera obtained in the "Realtime Georeferencing Module" and based on a digital elevation model (DEM). A 3D terrain model can be loaded as a prebuilt object or generated on the fly using elevation and orthophoto layers of an interest area.

Once the DEM is loaded in the 3D Scene Graph, a virtual camera matching the position and orientation of the real one can be defined. Other values described in or generated from the real video stream metadata such as the field of view of the camera are also needed to define the properties of the virtual camera. Then the desired vector and raster layers can be projected into the terrain model to compose the AR overlay. This vector and raster feature layers are gathered from a GIS database: raster layers are treated as textures while vector layers are converted to polygons allowing their manipulation as 3D objects.

A projection procedure can adapt views of these features to the elevation of the terrain according to the defined position and orientation of the camera. Terrain caused occlusions are also taken into account (Fig. 4). The GIS layers can be rendered by themselves as images with transparencies in order to overlay them to the real camera images. The rendering of the vector layers can be carried out after the generation of a graphic card-friendly triangulated 3D geometry and in case of the raster tiles these can be transferred to OpenGL texture units.

As a result, a virtual viewpoint matching representation can be generated where the desired feature vector and raster layers can be composed in realtime.

4. Video Output Generator

1. *Data output and realtime cartography*

The main product of this framework is the video stream with some vector/raster layers overlaid. These GIS layers are selected by the EMC users and projected over the video according to the view point and position of the real camera. This system manages to cartography in real time the video images captured form the remote sensing provider in order to enhance the information received at control and command.

Apart from the layers stored in GIS archive, realtime generated layers may also be shown over the video. For instance, in a fire monitoring event, parallel image segmentation, object tracking and classification and even evolution simulation tasks would be helpful to trace the unfolding event. This information could also be rendered as an additional GIS layer over the video.

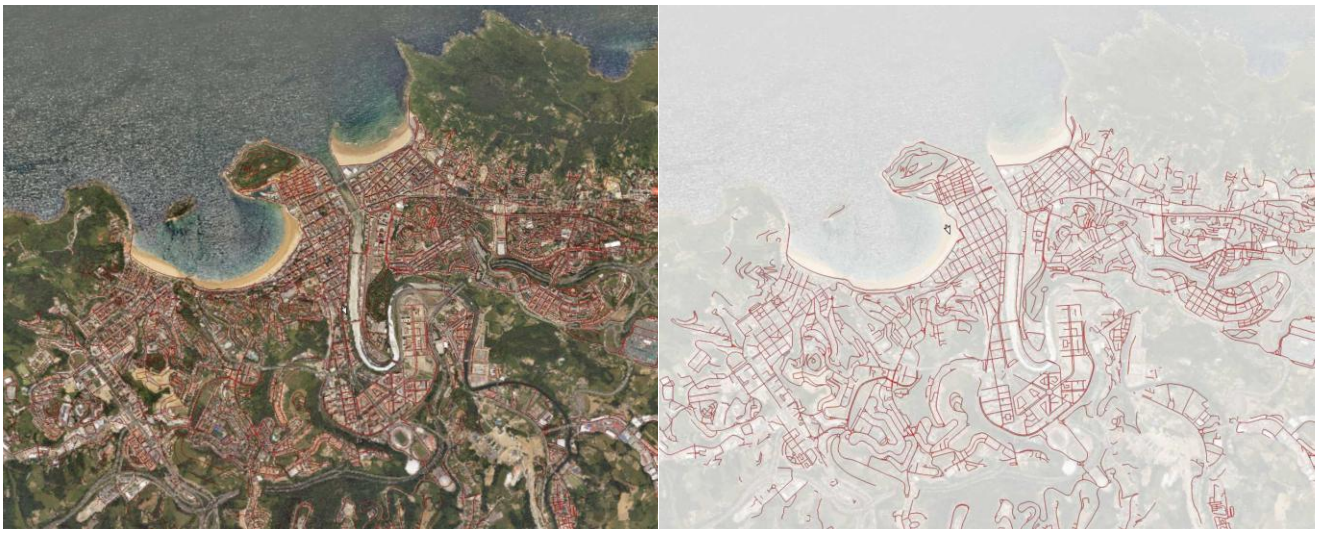
2. *Streaming and broadcasting*

There are different desired outputs that combine input video, geometadata and AR overlays, in order to fit to some specific uses.

Inside the EMC the main video wall should display a big resolution signal composed by the video and the AR overlay. Technically this can be achieved setting up a multidisplay setup where this big resolution image is shown through diverse projectors.

In the other hand, a video and overlay composed signal can be used for broadcasting purposes, like TV news or even streaming widgets for news websites. This output signal needs to fulfill the broadcasting standard, that's why the video output generator sends the video flow through a SDI adapter card.

Moreover, a feed orientated to mobile devices can be served to remotely monitor the situation and take decisions based on realtime updated information. This enables a team that works in the area to evaluate the situation on the spot and based on external experts opinion that also have access to this information.



(a) (b)

(c) (d)

Fig. 4 Ortophoto with vector overlay (a), and vector layer (b). Vector layer projection over the DEM according to the example camera parameters (c). Vector layer projected output image with alpha channel for real video overlaying (d).

3. Storage and semantic retrieval

Taking as input Real Time Georeferencing Module´s generated location metadata and video camera´s native metadata, Video Output Generation Module could be used to create a MPEG-7 video integrating the video and the generated metadata.

Multimedia Content Description Interface, or MPEG-7, provides standardized core technologies allowing the description of audiovisual data content in multimedia environments [6]. The goal of the MPEG-7 standard is to allow interoperable searching, indexing, filtering and access of audio-visual (AV) content by enabling interoperability among devices and applications that deal with AV content description.

MPEG-7 and specifically Description Definition Language (DDL) allows the definitions of different Descriptors and Descriptions Schemes according to the domain, which offers a high grade of flexibility in the modeling of annotations in the emergency scope. Using these capabilities, a Georeference Descriptor Scheme could be used to store the calculated georeferenced metadata. Alongside the designed Georeference Descriptor Scheme, MPEG-7 Visual Motion Descriptors could be used to store camera motion metadata (Fig. 5).

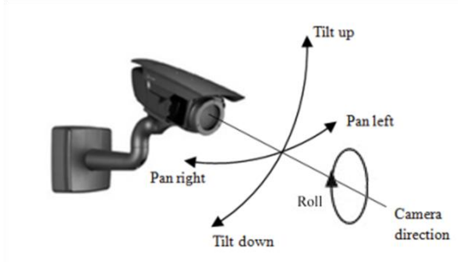


Fig. 5 Camera motion parameters.

Storing the generated metadata would allow future semantic content searches, enabling the efficient access to the key scenes that would help in mitigation and preparedness management stages.

5. CONCLUSIONS

In this paper the architecture for a complete framework integrating diverse technologies for developing new Emergency Management tools has been described. It considers navigation and INS technologies for camera parameter capture in realtime; computer graphics techniques for video image geolocalization, analysis and GIS layer live projections; and multimedia descriptors for location aware video coding and storage. A system based on the proposed architecture might be used to georeference the video images captured from a remote sensing camera and enhance them by overlaying GIS layers in realtime by virtual reality techniques.

Very specialized solutions exist for airborne imagery and video stream analysis that many emergency management centers cannot currently afford. The proposed architecture based on multimedia management tools, georeferencing algorithms, virtual reality techniques and GIS tools aim at providing EMCs with a solution for receiving and managing realtime imagery with limited resources.

Furthermore, software based open architectures as the one proposed in this paper can facilitate the adaptation of the system to the specific requirements of each user. In this case, the features of the camera, the video analysis required and the final publication needs could be accommodated by the proposed architecture.

REFERENCES

[1] T. Nyerges, H. Couclelis y R. McMaster, The Sage Handbook of GIS and Society, SAGE Publications Ltd, 2011.

[2] L. Montoya, «Geo-data acquisition through mobile GIS and digital video: an urban disaster management perspective.,» 2003.

[3] J. Hutton y J. Hutton, «10 Years of Direct Georeferencing For Airborne Photogrammetry,» *Photogrammetric Week,* 2005.

[4] W. S. S., H. J., K. B., N. H. M., W. S. S. y W. Z. H., «A Real-time Photogrammetric Mapping System,» *ISPRS*, vol. XXXV, pp. 61-65, 2004.

[5] M. P. Acinas, «Management of information and messages to the population in emergency, evacuation and simulation exercise situations,» *SEMES Journal of the Spanish Society of Emergency Medicine,* vol. 19, pp. 88-95, 2007.

[6] *MPEG-7 (ISO/IEC 15938-1:2002).*

[7] N. Marchetti, Telecommunications in Disaster Areas, River Publishers, 2011.

BIOGRAPHY OF AUTHORS

Maider Laka-Iñurrategi received her diploma degree in Telecommunication Engineering from the University of Navarra (2005). She is a staff researcher member of the Vicomtech-IK4 since 2005 in the field of Digital Television, Augmented Reality and GIS. She collaborated with the University of the Basque Country in the Computer Architecture and Technology as a teacher (2008). Nowadays she is studying a Master Degree in Geophysics and Meteorology. Her research interests include GIS, geo-visualizations, virtual reality and environmental information systems.

Julen Garcia is with the Department of Digital Tv & Multimedia Services, Vicomtech-IK4 GraphicsMedia.net, Spain. He received his Telecommunication Engineering degree in 2008 from Mondragon Unibertsitatea of Arrasate-Mondragón, Spain. His research interests are in real-time 3D graphics and low-cost professional TV broadcasts productions. Since 2008 October he is working at Vicomtech-IK4, where he designs and develops projects of the Digital Television and Multimedia Services area.

Kevin Alonso Gonzalez studied Technical Telecommunication Engineering in the specialty of Communication Systems (2002-2005) at Engineering Technical High School of Bilbao. He studied Telecommunication Engineering (2005-2007) at the Engineering Technical High School of Bilbao. Since July of 2008 he has worked in Vicomtech-IK4 as staff researcher in the field of Digital Television and Multimedia Services, participating in the European Research Project RUSHES. During 2009 and 2010 he has combined his research work with the study of a Master Degree on Space Science and Technology. His research interests include Semantic Web technologies and applications for multimedia asset analysis and management.

Marco Quartulli received the laurea degree in physics from the University of Bari, Italy, in 1997 and a PhD in EE and CS from the University of Siegen, Germany, in 2005. He worked from 1997 to 2010 on remote sensing ground segment engineering, image analysis, archives and mining for Advanced Computer Systems, Italy. From 2000 to 2003, he was with the Image Analysis Group at the Remote Sensing Technology Institute of the German Aerospace Center (DLR) in Germany. Since 2010, he has joined the Digital Television department of Vicomtech-IK4 in Spain. His research interests include multimedia mining and asset management, multiple acquisition high-resolution optical and radar remote sensing, scene understanding, data fusion and Bayesian modeling.