3DNSITE: A networked interactive 3D visualization system to support decision process and training in crisis management

paper_1008

Abstract

We report about the 3DNSITE system, a client-server 3D visualization tool developed to stream and visualize tridimensional hybrid data, in order to simplify location recognition for crisis managers and first responders during emergency operations, both for simulation and training. In this peculiar context is very important to share the same data in a web-based environment, accessing this with portable devices using simple andfast interfaces, since the specific end-users are not necessary skilled with virtual reality and 3D objects interaction. The dataset itself is not static as well as in generical purpose web based 3D browsers but dynamic, changing during the evolution of the real or simulated crisis. According with these constraints we propose a system which enhances already presented state-of-the-art methods achieving an excellent scalability on portable devices, in terms of performance and scalability. The proposed system scales over different devices according to their hardware resources and the available network bandwidth, exploiting a state-of-the-art multi-resolution representation for the 3D model and a multi-level cache system, employed to web access both the images and the hierarchical 3D model structure. 3DNSITE is integrated in a more complex and articulated training and decision framework for emergency operations, using a novel aided interface designed for touchscreen devices to easily navigate between 3D aligned photographs. Unlike common popular 3D maps browsers the free 3D navigation through the whole 3D model is always supported, assuring an interactive representation of the scenario also when no photo is available.

CR Categories: I.3.2 [Computer Graphics]: Graphics Systems—Distributed/network graphics I.3.6 [Computer Graphics]: Methodology and Techniques—Interaction techniques I.3.7 [Computer Graphics]: Three-dimensional graphics and realism—Virtual reality I.3.8 [Computer Graphics]: Applications

Keywords: Virtual reality, 3D interaction, Input and interaction technologies, Visualization

1 Introduction

Modern societies have experienced a spate of catastrophic events in recent years. Urbanised areas are especially vulnerable to the onset of crises and disasters, the combination of dense population concentrations (36.6M in Tokyo, 10.5M in Paris [UN: 2009]) and complex architectural environments makes it very hard to anticipate, prepare for and manage the impact of natural, industrial or man-made incidents. The recent crises have demonstrated the inherent difficulties that urban safety and crisis managers face when a large-scale disaster threatens an urban environment. In this everchanging environment, it is hard to design proper emergency plans, to train security organisations and effectively handle crisis management procedures. Therefore, it is essential for public authorities to better plan and train organisations and crisis managers, and new approaches and technologies are researched and developed to serve these critical needs. In these approaches one of the options that is

Figure 1: ***Crisis management and simulation setup.*** *Up left: a control room virtual whiteboard supporting multitouch controls. Up right: a tablet employed during a training session. Bottom left: Firebrigates truck hosting a field command post. Bottom right: On truck field command post control room.*

researched is the helping of remote navigation/location awareness in complex environments through the exploitation of 3D or near-3D data. Two particular datatypes are of particular importance to the security domain: the extremely massive point clouds and georeferenced(three-dimensionalized) photographs, since they can be acquired very rapidly and provide both measurable and visually recognizable description of a site. The basic idea behind the approach is that 3D data of some form (shapes of buildings/environments) is already widely available, and will become more common in the future due to the improvement and reduction in cost of 3D acquisition technologies. With these information, that can go from complete 3D reconstruction of sites (e.g. as those acquired by aerial or terrestrial 3D laser scanners to 3D calibrated photographs) is possible to present to users an easy to understand depiction of a natural/built environment. All this information can be used in applications to create navigation tools that will complement traditional 2D maps in a number of tasks. In this context we present the 3DNSITE system, a specific client-server visualization tool already integrated inside a wider and more articulated training and decision framework for emergency operations. The focus of the 3DNSITE system is to handle, store, stream and visualize tridimensional hybrid data (eg. point clouds, meshes,embedded 3D aligned photographs),in order to simplify location recognition for first responders, crisis managers and trainers. Within this peculiar environment there are several issues and research challenges. The first requirement is browsing and sharing the same hybrid scenario between agents deployed on the field and control rooms, compatibly with the diversified hardware and network resources (see Fig. 1). Infact except for the

Figure 2: ***Live navigation screenshots using the 3DNSITE viewer.*** *Interactive 3D navigation over the gas storage site of Geomethane in Manosque (France). This site has been used as training emergency scenario on the framework involving 3DNSITE (the 3DNSITE system is in charge to store, stream(server side) and visualize(client side) tridimensional hybrid data) involving the Operational Center of the Fire and Rescue Services of Alpes de Hautes, the local Gendarmerie and many real crisis managers and first responders. The company is an underground hydrocarbon storage site with a 7.5 million m3 storage capacity. It was created in 1969 and is classified as a SEVESO (high-risk production site). Due to this classification has been an excellent test case for the all system.*

main control room setup the users usually operate on the field using portable devices, clearly having strong limitations in terms of 3D capabilities, memory, storage and network bandwidth. Sometimes the control rooms themselves are located close the crisis site (eg. firebrigates truck), having limited resources also in terms of space, electric power supplies,etc. (see Fig. 1 bottom). Focusing the attention on the visualization side of the problem, though many solution are presented in literature distinctly for 3D model and images, the combined visualization of 3D models and embedded images needs specific solutions, both in terms of hardware and network resources scalability and in terms of user interface. Experience with crisis context actors (eg. police, fire departments, medical services) shows that often they are not used to virtual reality or 3D objects interaction and they prefer to interact with fast and simple aided interfaces than full free 3D navigators. According with these constraints we propose several solutions, both adapting already presented state-of-the-art methods and enhancing them, in order to enable efficent sharing, streaming and browsing of massive 3D models with embedded three-dimensionalized photographs. In order to achieve scalability over network and hardware resources we propose a combination of a priority-based,multi-level cache system with a multiresolution dynamic hierarchical representation of the 3D model [Cignoni et al. 2005] [Gobbetti and Marton 2004]. This view-dependent adaptive approach for the 3D model assures best performance in host-to-graphics and network communications, and completes the scenario depiction with a good approximation when no real image is available from some point of view, without need to synthesize artificial images. The benefits of these multiresolution representations are clearly visible in the results (see 2), mainly for the portable devices, where this support is mandatory to obtain interactivity (30fps vs. 1fps). Contextually a novel aided interface is presented to easily navigate between 3D aligned photographs as well as their viewports. The navigation method is designed for touchscreen portable devices (eg. tablets) or control room electronic whiteboards.This method is articulated in two step: an out-of-core pre-processing phase where starting from the 3D model and the related aligned images all the viewports are arranged and stored on a server, according with the image-to-image semantic distances and linear orderings, and a second step where run-time the client accesses the dataset via HTTP protocol, using the pre-computed informations and the touchscreen input to predict and address the user navigation. One the option researched in the emergency context is the opportunity to udate the scenario according to the real or simulated evolution of the crisis, to achieve this a method to automatically insert and align new photos coming from the field has been researched and implemented. Using images extra info (eg. exif data) or GPS/AGPS records taken during the data acquisition campaign, a geo spatial reference frame is calculated and stored both for 3D model and the 3D embedded images, this whole data can be considered as a kind of documentation of the site. 3DNSITE is integrated in a more complex and articulated training and decision framework for emergency operations. The performances and possibilities of the system are demostrated on several real scenarios employed in this main framework (see Fig. 2).

2 Related work

Even if there is a growing need both for decision support and crisis simulation systems (see [Boin 2009]), the crisis literature has paid little attention to the use of IT-based simulation tools (see [Dugdale J. and N. 2010]) until recent years. Nowadays the increasingly complex nature of crisis management asks for the support of Virtual Reality technologies, especially for the simulation of complex crisis and contingency scenarios that would be difficult to recreate and validate in real conditions. In the decision support and crisis management field only few methods of training have been proposed and partially employed, anyway they still haven't been fully translated into working software [Van De Walle et al. 2009] [Palen et al. 2007] [Lanfranchi and Ireson 2009]. Considering that crisis managers in control rooms are often away from the crisis location and do not see images of the disaster, as well as agents deployed on the site for the first time need to quickly orient themseves, having a clear picture of the problem and share the data in a collaborative network is an important part of the decision support and increases the efficiency of the actions [Carver and Turoff 2007]. Systems as 3DNSITE try to match those requirements improving navigation/location awareness through the exploitation of massive point clouds and three-dimensionalized photographs . Several solutions have been presented in literature about mobile maps browsing, 3D mobile navigation as well as photobrowsing, only a few existing browsers support the navigation of mixed 2D and 3D datasets [Snavely et al. 2006], [Vincent 2007], [Snavely et al. 2008] [Kopf et al. 2010] [Goesele et al. 2010]. 3DNSITE uses a navigation paradygm is similar to those adopted by Google StreetView [Vincent 2007] and Movie-Maps [Lippman 1980] and where the scene is visualized from predefined points of view, allows the smooth transition between aligned photos like Phototourism [Snavely et al. 2006]. To present the photographs embedded in the 3D world during the navigation, we adopt a projective texturing approach [Segal et al. 1992], which consists in rendering the 3D geometry and accessing, for each

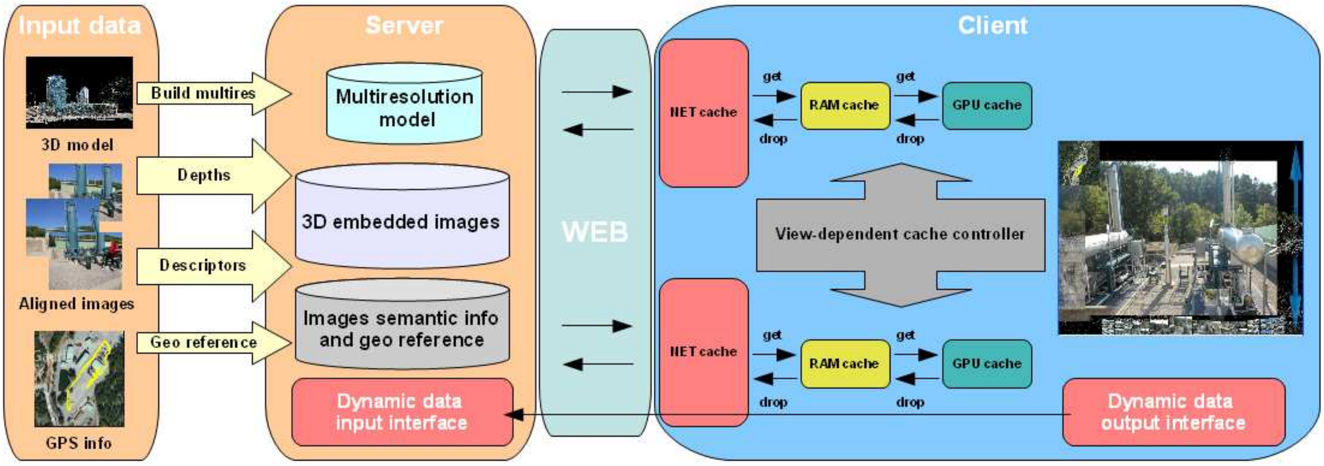

Figure 3: ***Architecture overview.*** *Starting from images already aligned with a 3D model and GPS data we perform an out-of-core preprocessing phase where the original 3D model is processed and stored in a multiresolution structure; images depths, descriptors, semantic distances and a geo spatial reference frame are computed. These informations are stored in a metadata index file and used both for 3D environment aided navigation and to set the images cache priorities according with the view paramaters. The 3D model and the geographical reference frame are considered as a kind of "skeleton" of the scenario, where pre-existent images and new images are 3D embedded. From the index file containing the list of repositories the clients access the data by HTTP protocol. The rendering and the caches are updated according with the current viewport parameters, using 3 cache levels: HTTP, RAM and GPU, at the same time the pre-computed informations are exploited to predict and address the user navigation. At run-time one or more images can by inserted in the existent dataset, acquired directly by a mobile client or collected in the main control room by various sources.*

produced pixel,at the color in the selected RGB image according to its projected position on screen. Several projective approaches have been considered effective at this purpose [Pintus et al. 2011a], the experience with real crisis context end-users has suggested an approach where the images are projected preserving their original viewport, positioning the user in scene exactly in the photographer location. Unlike common 3D photo browsers [Snavely et al. 2006] [Vincent 2007] 3DNSITE enables the user to perform free-point-of-view browsing at interactive frame-rates, exploiting the presence of dense point clouds or, more in general,large detailed 3D models. Note that a brute force approach for rendering these models is not feasible, as well as the loss of detail produced by the simplification needed to reduce their size is usually not acceptable. Level-of-detail (LOD) or multi-resolution data structures allow to encode multiple representations of the same shape, and support rendering at interactive pace by selecting a representation that fits the rendering budget and the current view-point. In order to achieve an efficient rendering and data streaming 3DNSITE adopts state-of-the-art multi-resolution techniques [Gobbetti and Marton 2004], [Cignoni et al. 2004], [Cignoni et al. 2005] which are flexible enough to support the array of 3D data kinds we are interested in and lends itself to an efficient memory management caching system.

3 Context and system overview

3.1 Context

The 3DNSITE system is developed for first responders, crisis managers and trainers. Therefore the potential users for the system can be divided in those three categories: First responders are operational units from (for example) the Police, fire department and medical services that operate on the field during a crisis situation. Crisis managers are people on the strategic level that come together when the crisis or incident becomes larger and more decisions need to be made and tasks need to be carried out. Trainers are those who organize training for first responders and crisis managers. The first category could operate on the field with the support of portable devices, such as touchscreen tablets or similar (see Fig. 1 up right), the other two categories could have access both the portable devices as well as large touchscreen displays, using them as a kind of whiteboard (see Fig. 1 up left). Typical users are not skilled with virtual reality or 3D objects interaction and usually they prefer to interact with fast and simple aided interfaces than full free 3D navigation throught the model. The experience with crisis agents and managers highlights the requirement to easily browse the set of images and contextually to have a view of the images embedded in the 3D world, navigating between views without experience "feel lost" sensation. Unlike common popular 3D maps and browsers the free 3D navigation through the point cloud is always supported, in order to guarantee a representation of the scenario also when no photo is available. This feature combined with state-of-the-art multi-resolution representations of the dataset is important key-strength of the system compared with other general purpose systems. The same view-dependent multiresolution system assures interactive frame-rates also when the dataset are massive and complex and an efficent support to streaming and scaling over limited portable devices. This view-dependent/output sensitive philosophy is applied to the image cache controller with the aid of the precomputed image descriptors. In the typical scenario of a crisis evolution or simulation the operators from the control rooms and the agents from the field must share the same data, causing the system must naturally runs in a web-based environment. Moreover the daset itself is dynamic and would be interfaced with institutional or military information systems. As far as 3DNSITE is concerned, the data involved are geotagged images, 3D models and any other content that can be represented as a geographical position or direction.

3.2 System architecture

The system overview is illustrated in fig. 3. The first step is an out-of-core pre-preprocessing of the input data. The original 3D model is processed and stored in a hierarchical multiresolution structure, using this multiresolution model and the related aligned images the viewports are arranged and stored on the server, according to the image-to-image semantic distances and their linear orderings. These informations will be used both for 3D environment aided navigation and to set the images cache priorities according with the view paramaters. Using images extra info (eg. exif data) or GPS/AGPS tracks recorded during the data acquisition campaign a geo spatial reference frame is calculated and stored for 3D model and the 3D embedded images. The 3D model and the geographical reference frame are considered as a kind of ”skeleton” of the scenario, where pre-existent images and new images will 3D embedded. The server contains the two basic repositories for the 3D model and for the images and a metadata index file with the all the precomputed viewport informations and geographical references. The clients access them using the HTTP protocol, starting from the index file which contains the list of repository urls. A priority-based,multi-level cache system is employed to access the high resolution images, their thumbnails and the 3D model multiresolution structure. The rendering and the caches are updated according with the current viewport parameters, using 3 cache levels: HTTP, RAM and GPU; at the same time the pre-computed informations are exploited to predict and address the user navigation. At run-time one or more images can by inserted in the existent dataset, acquired directly by a mobile client or collected in the main control room by various sources. In both cases the 3DNSITE system processed them and regenerating the metadata file and uploading the new one to the server. This operation can be performed locally on the server or remotely (eg. the client itself can send to the server the new image and updated metadata file).

4 Methods and tools

4.1 Server side pre-processing

We assume that the photographs are calibrated with the associated 3D model, such as those for example created with Structure from Motion pipelines [Snavely et al. 2006], [Pintus et al. 2011b] [Tuite et al. 2011] [Wan et al. 2012]. These hybrid models can be exploited for all operation that require location recognition, they can augment dense 3D models improving visualization from specific points of views, or replace them when possible in applications that do not require precise measurements. The first step is to transform the original 3D model (usually a point cloud) in a multiresolution structure, in order to handle datasets which exceed the capacity of the client GPU RAM and efficiently scale over portable devices with limited bandwidth resources. The construction process creates a hierarchy over the samples of the datasets, simply by reordering and clustering them into point clouds of approximately constant size arranged in a binary tree. In other words, the final multiresolution model has exactly the same points of the input model, but grouped into patches and organized in a level of detail representation. The root of the level of detail tree represents the entire model with a single cloud. These patches at different resolution can be assembled in different combinations to produce the full model. At run-time, selective refinement queries based on projected error estimation and regions of interest are performed on the multiresolution hierarchy to rapidly produce view-dependent continuous model representations by combining precomputed patches. The benefits of this approach are that the workload required for a unit refinement/coarsening step is amortized on a large number of point primitives, and the small point clusters can be optimized off-line for best performance in host-to-graphics and network communications. This hierarchical data structure is split in a index tree and a point cloud (or triangles) repository. The access to this repository is made through an output-sensitive/view-dependent controlled cache system. The same view-dependent philosophy combined with the precomputed image descriptors is employed to control the accesses to the images cache system (see 4.2). In a second phase the hi-res images are compressed in to JPEG format, then for each image the minimal image-space depth of its content is precomputed and stored. To determine this quantity a depth buffer of the 3D model is rendered from the image viewpoint, as defined in the camera calibration. An image ordering and distances are also precomputed as weel as an abstract descriptor is associated to each image. This descriptor, used to estimate good orderings and the semantic distances among images, is a weighted average of time-of-shot, image shot position, image shot orientation, color distribution and spatial-color-layout. The descriptors are exploited run-time both to drive the images cache priorities and to aid the user interaction. High-res images and all the precomputed metadata are stored in their proper repository ready for compression and streaming over HTTP. In addition, using images extra info (eg. exif data) or GPS/AGPS records taken during the data acquisition campaign, a geo spatial reference frame is calculated and stored for 3D model and the 3D embedded images (see fig. 4). The GPS records are compared with the shot positions given by the input data and the geographical spatial reference frame is calculated using a $RANSAC$ method based on [Capel 2005], [Chum et al. 2003], [Torr and Zisserman 2000] and [Fischler and Bolles 1981]. The method achieves good results and an acceptable precision for this specific application.

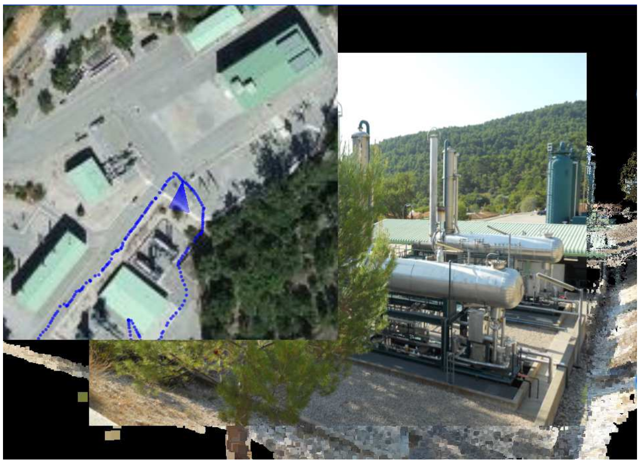

Figure 4: ***GPS track recorded at Geomethane site*** *All the dataset has been georeferenced comparing the GPS record of the agent who taked the pictures with the shot positions extimated by the structure-from-motion pipeline*

4.2 Data distribution

3DNSITE employs a priority-based, multi-level cache system which strives to optimize data access and the allocation of network and hardware resources. The cache pipeline is costitued by three levels: HTTP, RAM and GPU, with an additional fourth DISK level between HTTP and RAM provided when required. Such cache system is required to manage thousands of items and frequent priority updates and locking, as well as the synchronization of different threads. Each cache level operates in its own thread allowing for blocking operations on files and sockets greatly simplifying the implementation. To the 3D data blocks is assigned a priority determined by the multiresolution 3D subsystem, which takes in account visibility, distance, resolution, and so on. The aligned images are loaded from the network through the cache system according with a priority rule. Given a fixed size budget of RAM memory and GPU memory (depending of the hardware resources) the client starts to load the images giving the highest priority to the current camera viewport, and giving decreasing priority as the semantic difference with the current image increase (fig. 5). The minimum amount of cache instances is two, one for the 3D data blocks or one for the images (see fig. 3), though more caches can be instanced for further image collections (eg. further cache instances can be instanced for the thumbnails or temporally different image sets). Given the size of the dataset, compression

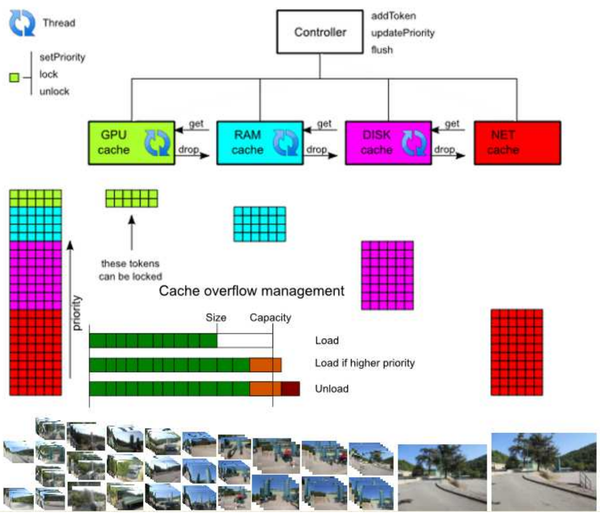

Figure 5: ***Multilevel priority based cache system*** *The aligned images are loaded from the network through the cache system according with a priority rule. Given a fixed size budget of RAM memory and GPU memory (depending of the hardware resources) the client starts to load the images giving the highest priority to the current camera viewport, and giving decreasing priority as the semantic difference with the current image increase.*

techniques are also necessary. In a cache system, decompression can be considered as a part of the process of loading an item from a lower to an higher cache level. Using a compression schema means trading some loading speed for storage space and bandwidth. 3D data blocks are stored remotely already in the form of VBOs (vertex buffer objects). Compression/decompression of 3D models have also been considered, but available systems do not deliver enough deflating speed for our purposes. For the images a JPEG compression schema has been adopted. Though the JPEG decompression cannot easily implemented on GPU level without specific hardware requisites, this system assures an advantageous compression ratio and a wide compatibility with different devices.

4.3 Client side 3D navigation

In our current implementation we use a HTTP/1.1 persistent connection approach and optionally employ HTTP pipelining for the 3D data blocks, the same technique has been exploited successfully in [Bettio et al. 2007] [Gobbetti and Marton 2004] for the same purpose. The combination of these two techniques improve bandwidth usage and reduce network latency, while keeping the protocol simple from API point-of-view, since clients benefit from an underlying connection-based implementation hidden under a reliable

Figure 6: ***Highlight of the next image candidate*** *Live navigation in the Geomethane scenario. Touching a point in bottom left corner of the screen the most rappresentative image for that spatial place is highlighted. With a double click the camera automatically moves to the point of view from which the candidate image has been taken.*

connectionless interface. The pipelining approach allows multiple HTTP requests to be written together out to the socket connecting client and server without waiting for the corresponding responses. The client then waits for the responses to arrive in the order in which they were requested. Since each 3D data block consists of several thousands of points or triangles already precomputed in the preprocessing step, assembling at rendering time the view-dependent representation is extremely fast and results in very low CPU load. Each block is optimized,cached in the GPU through the multi-level cache system, and rendered with a single CPU call for maximum performance. The rendering algorithm selects the best representation according to the rendering budget and the availability of the blocks, thus guaranteeing a minimum frame rate. The resulting technique has the following properties: it is fully adaptive and is able to retain all the original topological and geometrical detail, even for massive datasets; it is not limited to meshes of a particular topological genus or with a particular subdivision connectivity and it preserves geometric continuity of variable resolution representations at no run-time cost; it is strongly GPU-bound and is over one order of magnitude faster than existing adaptive tessellation solutions on current PC platforms, since its patch-based structure successfully exploits on-board caching, cache coherent stripification, compressed out of core representation and speculative prefetching for efficient rendering on commodity graphics platforms with limited main memory; it enables high quality simplified representations to be constructed with a distributed out of core simplification algorithm. As already said in 4.2 the aligned images are loaded from the network through a cache system. The cache controller is output-sensitive, deciding the load priorities according with current viewport, image descriptor differences and hardware capabilities. The requestes are stopped when the memory budget is filled and restart if the user changes his point of view and consequently the current viewport/image. To present the photographs embedded and projected in the 3D world during the navigation, we adopt a projective texturing approach, with an effect equivalent to cast the image as a slide from a virtual projector into the 3D scene, simulating the viewport from which the photograph has been originally taken the photo. This solution has been choosen after have considered different projection methods, meeting the end-users requirements which prefer to see the real images as they have been taken without any projection artifact. The image rectangle is defined as the section of the view frustum pyramid of the correspondign shot, cut at distance

D from the camera and roughly corresponding to the precomputed minumum depth of the objects featured in the image. When the view-position discrepancy increases the texture projection is progressively disabled, and the scenario depiction is provided by the underlying 3D model. Thanks to the presence of an efficent multiresolution structure the application can exploits massive and high detailed models at this purpose, providing a good and useful representation of the site also when no photograph is available from the current point of view. One the main important feature required in such as context is a fast and simple interface to navigate in the 3D environment. Considering the users skills and the time critical situations this interface must avoid "feel lost" perceptions, and must keep the focus on the real images taken from the field. Starting from an initial $currentimage$ the interface system locates the camera on the related image viewport, from this position the user could starts his navigation interacting with the touchscreen. Following the touch of the user over the screen, the system highlights (see fig. 6) the possible nearest images from this position according with the metric defined at the preprocessing time (see 4.1), or if the user is touching inside the current viewport itself the camera can rotate according with the movement, or zoom in the current image if a double-click is catched. If the double-click is catched outsides the current view the camera moves to the touched highlighted viewport, which becomes the new current one. The image descriptor is defined by the metric (1),

$$\frac{\sum_1^n x_i f_i}{\sum_1^n f_i} \tag{1}$$

where the features f_i are: $shot$ position,shotdirection,timeof$shot$, $color\ distribution$, $spatial\ color\ layout$, $depth$, and the weights x_i are defined by the user at pre-processing time, according with the scenario requirements (see 5). Considering the current image descriptor all the others shots are ordered by increasing difference from this one and are rendered back-to-front in a dedicate not visible OpenGl buffer. To identify the different shots in the 3D space we draw in this buffer the projected viewports of the images as oriented quads, encoding the related id in the colors. Moving the touch on the screen outside the current view results in probing this hidden frame buffer, returning the id of the touched viewport throught its color. Since several viewports could share the same pixel positions on the screen, the back-to-front method assure that the returned id is the nearest to the current one between the concurrent viewports in those pixels.

5 Implementation and results

The $3DNSITE$ system has been released under Windows and Linux platforms and developed with C++, OpenGL and the Nokia Qt toolkit. The framework includes two modules: a pre-processor sub-system called $3Dnsite\ Generator$ and a client module called $3Dnsite\ Viewer$.

5.1 Pre-processor module

The $3Dnsite\ Generator$ prepares the datasets before they are stored in a server, offering an user interface in order to tune the data building according with the users requirements. The 3D models and the related aligned images can come from different meanings. Since active 3D range scanning technology is becoming a diffused resource, it is becoming quite common for some industries routinely doing laser scanning to monitor their factories (eg. power plants,complex pipelines,etc.) as well as the photographs are a fast and easy way to document a site. On the other hand also cheaper passive methods are becoming quite popular. Thanks recent Computer Vision advances in Structure from Motion (SfM),

Figure 7: ***Navigation with a tablet in a training dataset*** *One the main important feature required in such as context is a fast and simple interface to navigate in the 3D environment. Considering the users skills and the time critical situations this interface must avoid "feel lost" perceptions, and must keep the focus on the real images taken from the field.*

Dataset	3D model Size MSamples	3D model Patches	3D model Time	Images Size Mpixels	Images Count	Images Time
Geomethane	7.5	463	2m22s	12	300	5m20s
Training building	0.5	30	10s	12	60	41s

Table 1: ***Dataset pre-processing stats.*** *The pre-processor transformed the original 3D model in a multiresolution structure, in order to handle datasets which exceed the capacity of the client GPU RAM and to enable network streaming. The statistics show as the time to build a new dataset of this type is very short, allowing the trainer to quickly create new scenarios.*

it is possible to extract more affordable, dense 3D samplings from large image datasets, even those composed of heterogeneous pictures shot under uncontrolled conditions. Given a dense set of photographs, SfM algorithms produce medium quality colored point clouds, often enough detailed for all operations that require location recognition. The way in which the 3D model is acquired also determines which techniques can be used to calibrate and align the images on the 3D model (eg. [Pintus et al. 2011b]). When the images are used to compute a point-cloud 3D model, they are already calibrated and aligned to that model. The datasets presented here are acquired by a SfM pipeline, starting from a set of images and with the support of a GPS tracker for the spatial reference. To process them we have used a commodity desktop PC with an Intel Core2 Q6600 2.6GHz and 2GB ram. The first dataset presented in this paragraph is the gas storage site of Geomethane in Manosque (France). This site has been used as training emergency scenario involving the Operational Center of the Fire and Rescue Services of Alpes de Hautes,the local Gendarmerie and many real crisis managers and first responders (see 5.2). The 3D point cloud has 7.5Million samples and has been obtained from 300 12Mpixel photographs. As already discussed the pre-processor transformed the original 3D model in a multiresolution structure, in order to handle datasets which exceed the capacity of the client GPU RAM and to enable network streaming (see tab. 1). The images metadata has been processed using an average depth for each image as reference plane. Different image depth calculations (avaliable options are also minimum and maximum depth) result in a different images projection at rendering, as well as different features weights result in a different processing time and a different behavior of the user interface (see eq. 1). The 3Dnsite Generator enables the user to set

User	Device	hardware performances Multires fps	No multires fps	Connection	RAM MB	GPU MB	Network peak MB/s	startup s	average KB/s
Manager 1	whiteboard	172	16	ethernet	1024	256	11	0	700
Manager 2	laptop	70	9	wireless N	512	256	600	0	600
Manager 3	netbook	35	3	ethernet	512	256	3.6	3	400
Manager 4	netbook	27	3	ethernet	512	128	3.4	3	300
Agent 1	netbook	26	2	wireless N	512	128	0.600	6	220
Agent 2	tablet	26	1	wireless N	1024	256	0.593	6	470
Agent 3	tablet	24	1	wireless N	1024	128	0.535	8	260
Agent 4	tablet	23	1	wireless N	512	128	0.587	12	400
Agent 5	tablet	19	1	3G-HSDPA	1024	128	0.220	22	125
Agent 6	tablet	8	1	EDGE-GPRS	1024	128	0.050	34	36

Table 2: ***Client rendering stats.*** *All the devices have been set to 1200x800 screen resolution using the Geomethane dataset to compare performances. The average bandwidth is intenteded during the navigation; at the application startup there is a one-shot bandwidth peak depending on memory budget and images size. The support of the 3D multiresolution system increase in importance going on portable devices, where no interaction is possible without it. The 2 last columns in hardware performances show the memory budget both for RAM and GPU caches: for local applications these values can be set almost to the full memory capacity, this is not valid for the network application, where an high GPU budget increases the average bandwidth required to fill the last cache stage.*

up 6 different feature weights in a scale from 0 to 1: shot position, shot direction, time of shot, color distribution,spatial color layout and image depth. In the Geomethane dataset the dominant values are shot position (0.8), shot direction (0.4) and image depth (0.3). The other features have been considered as with values less than 0.1. While the Geomethane scenario can be considered a kind documentation of the site, the second typology of data proposed in 1 is a pure training scenario. Infact one important requirement for these systems is to train agents on undiscovered sites, where they have to orient themseves and to quickly find strategical locations. The pre-processing statistics show as the time to build a new dataset of this type is very short, allowing the trainer to quickly create new scenarios. Due to the spatially limited area the dominant values for the metadata creation are the shot direction (0.9), the shot position (0.2) and the image depth (0.4). Notably these training dataset are tipically small places but characterized by high-res photos and they have anyway a large amounth of data to stream.

5.2 Client module

The client setup tested has been developed essentially for two user profiles, a manager profile and an agent profile. The manager profile follows the operations from the operational headquarters using an electronic whiteboard connected to a commodity desktop (fig. 1 up left), or from the field command post using a laptop (fig. 1 bottom left), employing 3DNSITE to plane detailed operations and give orders to the agents. Instead the agent profile has to perform specific tasks in several strategic locations, assuming he has never been in that place before. Many users from these two profiles have tested the system together with real trainers, driving the development of the navigation metaphor, finding it confortable and intuitive. Both profiles have found the interface useful related to the specific requirements, as well as the agents have been discovered themseves familiar with the application for its similarity with popular web based 3D maps navigators. The hardware employed for the test has been: commodity desktop PC with an Intel Core2 E6600 2.4GHz, 2GB ram and an nvidia GeForce GTX560 connected with an 40 inches multitouch whiteboard, Alienware M17xR3 laptop, with an Intel Core i7 Processor 2630QM, 6GB ram and nvidia GeForce GTX460M, Compag Mini netbook with Intel Atom and nvidia ION, Acer Iconia 500 tablet with AMD Fusion C-60 and Radeon HD6290. All the devices have been set to 1200x800 screen resolution using the Geomethane dataset to compare performances. As it is possible to see in tab. 2 the support of the 3D multiresolution system increase in importance going on portable devices, where no interaction is possible without it. At the same time the multilevel cache scales well over the hardware and the network bandwidth, assuring the interactivity of the application also with 3G mobile bandwidth, especially considering the size and detail of the images (eg. 12Mpixels). The 2 last columns in the hardware performances show the memory budget both for RAM and GPU caches: For local applications these values can be set almost to the full memory capacity, this is not valid for the network application, where a high GPU budget increase the average bandwidth required to fill the last cache stage (increasing the performance as well). The most important difference between devices (and their available network bandwidth) is not in terms of interactivity, which is well supported by the scalability of the system, but in the time needed to load the cache at the application startup. In this phase the network bandwidth has the peak illustrated in tab. 2 in the first column of $Network\ bandwidth$; while in desktop and laptop cases the startup time of the application is almost istantaneous, the time needed increase with the reduction of hardware and network capabilities. The network setup on the Geomethane site was a WiFi-N provided by a mobile antenna from the Fire Brigate command post. Altought is becoming quite common to set mobile wireless connection in this context, we've perfomed several tests using 3G and EDGE connections, indeed experiencing a loss of performances but obtaining a good interactivity navigating as well.

6 Conclusions and future work

We have presented a web-based system to interactively navigate in complex 3D environment during the evolution/simulation of a crisis. This tool,called 3DNSITE, has been developed to share,stream and visualize tridimensional hybrid data integrated in a larger and complex training and decision framework for emergency operations. Tested inside this framework 3DNSITE has demonstrated to match some important crisis context users requirements, in particular achieving scalability over limited network and hardware resources with a good interactivity. Although the single methods are already included in computer graphics literature, the combination and the enhancement of them to match the peculiar web-based environment requirements has led to very good results, in term of performance and crisis context user benefits. The scalability over network and hardware resources is achieved by the combination of a priority-based,multi-level cache system with a multiresolution dynamic hierarchical representation of the 3D model. This output-sensitive approach for the 3D model assures best performance in host-to-graphics and network communications, and complete the scenario depiction with a good approximation when no real image is available from some point of view, without need to synthesize artificial images. Despite the models employed for the tests can't be considered huge massive models in terms of computer graphics, they are big enough to limit the interaction in remote operations with limited hardware and network resources, as show in 5.2. The strong attitude to network scalability of the multiresolution method proposed is also promising for future enhancements and applications, since the 3D models are continously increasing in size and complexity. The same view-dependent/output sensitive philosophy is applied to the high resolution image cache controller with the aid of the precomputed image descriptors, supporting a fast and efficent navigation interface. The opportunity to udate the scenario according to the real or simulated evolution of the crisis is now supported within the limit of few images, the current work is focalized to obtain a better dynamic and progressive dataset update, both the images and the 3D model. **Acknowledgments.** Removed for blind reviewing.

References

BETTIO, F., GOBBETTI, E., MARTON, F., AND PINTORE, G. 2007. High-quality networked terrain rendering from compressed bitstreams. In *Proc. ACM Web3D International Symposium*, New York, NY, USA, ACM Press, 37–44.

BOIN, A. 2009. The new world of crises and crisis management: Implications for policymaking and research. *Review of Policy Research 26*, 4, 367–377.

CAPEL, D. 2005. An effective bail-out test for ransac consensus scoring. In *BMVC British Machine Vision Conf*, 629–638.

CARVER, L., AND TUROFF, M. 2007. Human-computer interaction: the human and computer as a team in emergency management information systems. *Commun. ACM 50*, 3 (Mar.), 33–38.

CHUM, O., MATAS, J., AND KITTLER, J. 2003. Locally optimized ransac. In *DAGM-Symposium*, 236–243.

CIGNONI, P., GANOVELLI, F., GOBBETTI, E., MARTON, F., PONCHIO, F., AND SCOPIGNO, R. 2004. Adaptive tetrapuzzles: efficient out-of-core construction and visualization of gigantic multiresolution polygonal models. *ACM Trans. Graph. 23*, 3, 796–803.

CIGNONI, P., GANOVELLI, F., GOBBETTI, E., MARTON, F., PONCHIO, F., AND SCOPIGNO, R. 2005. Batched multi triangulation. In *Proc. IEEE Visualization*, 207–214.

DUGDALE J., BELLAMINE-BEN SAOUD N., P. B., AND N., P. 2010. Simulation and emergency management. *Information Systems for Emergency Management*.

FISCHLER, M. A., AND BOLLES, R. C. 1981. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Commun. ACM 24*, 6 (June), 381–395.

GOBBETTI, E., AND MARTON, F. 2004. Layered point clouds: a simple and efficient multiresolution structure for distributing and rendering gigantic point-sampled models. *Computers & Graphics 28*, 6 (Dec.), 815–826.

GOESELE, M., ACKERMANN, J., FUHRMANN, S., HAUBOLD, C., KLOWSKY, R., STEEDLY, D., AND SZELISKI, R. 2010. Ambient point clouds for view interpolation. In *ACM SIGGRAPH 2010 papers*, ACM, New York, NY, USA, SIGGRAPH '10, 95:1–95:6.

KOPF, J., CHEN, B., SZELISKI, R., AND COHEN, M. 2010. Street slide: browsing street level imagery. In *ACM SIGGRAPH 2010 papers*, ACM, New York, NY, USA, SIGGRAPH '10, 96:1–96:8.

LANFRANCHI, V., AND IRESON, N. 2009. User requirements for a collective intelligence emergency response system. In *Proceedings of the 23rd British HCI Group Annual Conference on People and Computers: Celebrating People and Technology*, British Computer Society, Swinton, UK, UK, BCS-HCI '09, 198–203.

LIPPMAN, A. 1980. Movie-maps: An application of the optical videodisc to computer graphics. *SIGGRAPH Comput. Graph. 14* (July), 32–42.

PALEN, L., HILTZ, S. R., AND LIU, S. B. 2007. Online forums supporting grassroots participation in emergency preparedness and response. *Commun. ACM 50*, 3 (Mar.), 54–58.

PINTUS, R., GOBBETTI, E., AND CALLIERI, M. 2011. Fast low-memory seamless photo blending on massive point clouds using a streaming framework. *ACM Journal on Computing and Cultural Heritage*. To appear.

PINTUS, R., GOBBETTI, E., AND COMBET, R. 2011. Fast and robust semi-automatic registration of photographs to 3d geometry. In *The 12th International Symposium on Virtual Reality, Archaeology and Cultural Heritage*.

SEGAL, M., KOROBKIN, C., VAN WIDENFELT, R., FORAN, J., AND HAEBERLI, P. 1992. Fast shadows and lighting effects using texture mapping. In *SIGGRAPH'92*, 249–252.

SNAVELY, N., SEITZ, S. M., AND SZELISKI, R. 2006. Photo tourism: exploring photo collections in 3d. In *ACM SIGGRAPH 2006 Papers*, ACM, New York, NY, USA, SIGGRAPH '06, 835–846.

SNAVELY, N., GARG, R., SEITZ, S. M., AND SZELISKI, R. 2008. Finding paths through the world's photos. In *ACM SIGGRAPH 2008 papers*, ACM, New York, NY, USA, SIGGRAPH '08, 15:1–15:11.

TORR, P. H. S., AND ZISSERMAN, A. 2000. MLESAC: A new robust estimator with application to estimating image geometry. *Computer Vision and Image Understanding 78*, 138–156.

TUITE, K., SNAVELY, N., HSIAO, D.-Y., TABING, N., AND POPOVIC, Z. 2011. Photocity: training experts at large-scale image acquisition through a competitive game. In *Proceedings of the 2011 annual conference on Human factors in computing systems*, ACM, New York, NY, USA, CHI '11, 1383–1392.

2009. United nations (u.n.) population division world urbanization prospects (the 2009 revision).

VAN DE WALLE, B., TUROFF, M., AND HILTZ, S. R. 2009. *Information Systems for Emergency Management*. M.E.Sharpe.

VINCENT, L. 2007. Taking online maps down to street level. *Computer 40* (December), 118–120.

WAN, G., SNAVELY, N., COHEN-OR, D., ZHENG, Q., CHEN, B., AND LI, S. 2012. Sorting unorganized photo sets for urban reconstruction. *Graph. Models 74*, 1 (Jan.), 14–28.