3D crime scene acquisition, representation and analysis

Jarek Gryz¹, Piotr Jasiobedzki², Michael Jenkin¹, Carl McDiarmid³, Michael Bondy², Ho-Kong Ng², Robert Codd-Downey¹, Sunbir Gill¹, Marcin Kwietniewski¹, Anna Topol¹, and Stephanie Wilson¹

¹ Department of Computer Science and Engineering, York University
² MDA, Space Missions, Brampton, Canada
³ Royal Canadian Mounted Police, Forensic Identification Operations Support Services

Investigating crime scenes in which chemical, biological, radiological, nuclear (CBRN) or explosive agents have been deployed poses great dangers to first responders and forensic investigators. As a result of environmental contamination the normal process of crime scene investigation may be harmful to the personnel involved. Nor is it always possible to remediate the scene prior to investigation as any prior decontamination of the scene to make it ‘safe’ for the investigators may destroy evidence essential for later investigation and criminal proceedings. Nor is it always practical for the crime scene to be fully processed while investigators wear appropriate protective gear. The difficulty associated with investigating contaminated crime scenes is difficult to overestimate. Crime scenes may be sufficiently contaminated that investigators must either don highly restrictive protective suits or investigate the site via teleoperated vehicles in order to minimize the time that personnel spend in the contaminated zone. It may also be necessary to return the scene to normal use in a rather short period of time which poses additional constraints on the investigative process.

Effective investigation of CBRN crime scenes requires the development of technologies that enable remote crime scene investigation (image capture, collection of multi-dimensional measurements and detector readings, sample collection) while marking their locations and recording observations from a properly shielded location. At the very least it is necessary to minimize personnel exposure to contaminants by reducing dramatically the time spent in the hazardous zone by developing devices and techniques that enable investigations to be conducted more effectively while wearing restrictive protective clothing. These requirements call for autonomous and semi-autonomous systems and technologies to acquire models of crime scenes and to record relevant data prior to subsequent decontamination. An integrated approach is required that can acquire surface scans of the environment and other complex sensor
Fig. 1. Traditional approaches to contaminated crime scenes. (a) A wide range of protective gear exists that can be used to protect investigators during crime scene investigation. Such suits provide gross constraints to user mobility and their ability to observe and interact with the environment. (b) For sufficiently contaminated scenes remotely operated vehicles may be used that allow the investigative team to perform the investigation from a suitably shielded vantage point. Either approach limits the abilities of crime scene investigators in fundamental ways.

data along with traditional crime scene reports. The recovered data and extracted information must be represented in formats that allow intuitive and fast access to vast amounts of heterogeneous information.

This chapter describes results from the CBRN Crime Scene Modeller (C2SM) project – a project whose goal is the development and field evaluation of technologies for real time data acquisition at CBRN crime scenes, and fast recreation of such scenes as virtual environments with access to all of the multi-model data and heterogeneous evidence associated with the scene. The C2SM project leverages recent results in 3D scene modelling to develop an experimental system that supports 3D crime scene acquisition, representation and analysis.

1 Motivation

Crime scene investigation is an essential component of modern police work. Although crime scenes are processed and suspects apprehended within 50
3D crime scene acquisition, representation and analysis

In practice processing a crime scene can take an extended period of time (days, weeks, or even months). It is a process that requires extensive and detailed analysis of the scene, recovery and analysis of physical material and the documentation of the state of the local environment. (See [4] and [14] for an in depth treatment of the process of crime scene investigation.)

A critical requirement in the investigation of any crime scene is to ensure the safety of the personnel involved in the investigation. Although this rarely appears to be an issue in television dramas, extremes of temperature and other natural conditions at the scene are a concern in any investigation. Crime scenes may occur in environments that are unsafe due to various environmental factors, and the nature of the event itself may introduce contaminates into the crime scene that can be hazardous to investigators. Such contaminates may also require extraordinary care be taken with devices that the investigative team might bring to the crime scene (e.g., cameras, notebooks, etc.), that in the most extreme cases may require that the recording devices be disposed of after use. Figure 1 illustrates some current approaches to dealing with a contaminated crime scene. Figure 1(a) shows an investigative officer wearing protective clothes. This type of gear provides limited protection but at the cost of reduced mobility, reduced view of the scene, and considerable discomfort leading to reduced time and reduced efficiency at the scene. Moreover, other officers working the scene must rely on limited communication with investigators working in the “hot zone.” Figure 1(b) shows a remotely operated vehicle that can be used in crime scene investigation. The use of a remotely operated vehicle allows investigators to process a scene from a safe vantage point but at the cost of a significant decrease in efficiency and effectiveness.

Analyzing a crime scene involves securing the scene from other uses, which for scenes that take place in public places may be difficult and at the very least may inconvenience the general public. The combination of requirements related to isolation, safety, effectiveness and efficiency comes to a head when one considers the need to conduct forensic investigation of terrorist events. Almost by definition, such events are designed to cause significant inconvenience for the general public and often take place in public locations. Such locations must be returned to general use as soon as possible after the event in order to deny the perpetrators this as a motive for the attack. Large scale terrorist attacks against major urban environments are mercifully rare but it is important to recognize the challenge posed by the response to chemical, biological, radiological, nuclear or explosive attack. Although the incidence of such attacks is low, they cannot be discounted entirely. For example, in *The Four Faces of Nuclear Terrorism* [3] the authors argue that radiological terrorist attacks are not only likely, they are inevitable given the wide availability of material and the relatively straightforward technology required in order to initiate such an attack. Conducting an effective crime scene investigation in the aftermath of such an attack requires the development of tools and technologies that allow the investigative team to investigate a contaminated scene.
Fig. 2. A sample scan obtained with the visible light scanner described in [13]. The scanner has been placed at a number of locations in the environment and the visible portion of the environment from each camera position has been recovered. These scans have been combined to provide the 3D model shown. This is a complete 3D model: It can be repositioned and rotated and viewed from different orientations.

from a safe vantage point in an effective and efficient manner. The development of technology to support such investigation is a core goal of the C2SM project.

The last ten years has seen a significant improvement in the capability of 3D scanning technology to obtain dense surface scans of large scale spaces. Laser and visible light technology now exists that can be used to capture such environments (see [1] and [11] for laser-based examples and [13] and [5] for examples that use visible light). Such sensors can be used to obtain virtual models of crime scenes but the models in and of themselves are not directly useful to an investigator. Consider the 3D model shown in Figure 2. This model, obtained at the training facility at Ontario Provincial Police College, London, Ontario, Canada, is of a simulated crime scene extending over a number of rooms. The 3D model was constructed by merging together 3D scans obtained using the iSM sensor[13] from different viewpoints in the environment. The 3D model permits virtual walkthroughs of the scene but the model lacks semantic structure that would assist in later investigation. There is no definition of the concept of ‘room’ or ‘floor’ or ‘body’. The scene is simply a collection of textured polygons. In order to provide a more structured representation it is necessary to group these textured polygons into semantically
meaningful structures. Building an effective system to permit the investigation of contaminated scenes requires the development of (i) appropriate sensor technologies to capture the 3D structure of the scene and other sensor measurements that may be necessary for investigations, (ii) techniques to segment and group these measurements into semantically meaningful structures, and (iii) the development of appropriate data representation and visualization tools to enable responders to interact with the representation in an effective manner.

2 System overview

Fundamental to the operation of the C2SM system is the ability to collect 3D surface scans of the crime scene itself. The C2SM relies on an augmented version of the Instant Scene Modeller (iSM)\cite{13} to perform this function. The C2SM sensor provides additional sensing modalities including IR, chemical
Fig. 4. The operator interface. (a) The vehicle console. This provides control of the platform itself and is used to teleoperate the device. (b) The sensor console. This provides control of the C2SM sensor.

and radiological sensing above those provided by the iSM (Figure 3). The modified iSM sensor system is normally mounted on a mobile platform and the vehicle and sensor remotely operated from a protected position (Figure 4). Alternatively, in less contaminated environments the C2SM hand-held unit can be separated from the vehicle and operated manually by a first responder in protective clothing in the hot zone supported by other staff at a safe remote location. The operator console provides control of both the vehicle (Figure 4(a)) as well as its onboard sensors (Figure 4(b)). By separating the operation of the vehicle from the sensor the sensor package can be deployed on a variety of different remotely operated vehicles.

Raw scans from the C2SM sensor are made available to other users and software modules via a communications infrastructure (see Figure 5). Scans collected using the C2SM sensor are processed and merged with traditional and non-traditional crime scene data and made available to investigators both on site and elsewhere. In order to provide a coherent and consistent interface to these representations a multimedia database system and visualization tool[9, 10] has been developed. Details of the structure of the database system and the user interface that has been developed to manipulate information within the database are described below. But before turning to these issues, we begin by examining the process of obtaining segmented environmental models.

3 Obtaining environmental models

Although the C2SM sensor can be used to obtain detailed scans of the environment, these scans are not segmented into salient features, rather the sensor produces a single 3D model capturing all visible surfaces within the environment. In order to be able to represent and reason about individual objects within these scans tools are required to enable investigators to segment scans of the complete environment into salient components.
A number of manual tools exist to edit the scans obtained by the C2SM sensor. Scans can be exported as VRML97 representations and tools have been developed to load these representations into standard editing toolkits. Although not explicitly designed to segment one object from another such editing tools can be adapted to this function. This has been found to be an extremely time consuming process, and more automatic tools are desired in order to assist in this process. 3D Lazy Snapping[6] is a technique for the assisted segmentation of 3D images based on the 2D Lazy Snapping approach developed for image segmentation[12]. In 2D Lazy Snapping the user indicates picture elements (pixels) within a 2D image to be segmented by marking sample foreground and background elements of the scene. Guided by this selection the system then selects those pixels it considers part of the foreground region and rejects those it considers part of the background. The user can then either augment the set of marked foreground and background pixels and allow the system to recompute the selection, or the user can manually adjust any of the pixels which have been mis-classified. Due to the system classifying the majority of the pixels automatically, the demands on the user are drastically reduced when compared to manual segmentation.
As in 2D Lazy Snapping the problem of segmenting the foreground object from the 3D scene is considered as a graph cut problem (see [8] for details on optimization via graph cut), with the cut dividing the foreground region from the background region. The 3D version of Lazy Snapping replaces the
2D pixel primitive of [12] with 3D primitives. Scene segmentation is based on a voxel representation of the scene. A graph is constructed with each non-empty voxel in the scene being represented by a node in the graph. Edges connect nodes in the graph when the corresponding voxels are adjacent in the scene. These voxel-voxel edges are weighted so that similarly coloured voxels have a high weight between them, while dissimilarly coloured voxels have a low weight between them. Source and sink nodes are added to the graph that are connected to all other nodes in the graph. Foreground and background colours are computed from the user’s selection of foreground and background voxels. The colour distributions of the foreground and background regions are represented via the k-means clustering algorithm[2]. For each voxel node, the minimum distance from its colour to the set of foreground and background colours is computed. When a voxel node is similar to the foreground colours the edge weight between the voxel node and the source node is high, otherwise it is low. Similarly, when the voxel node is similar to the background colours the edge weight between the voxel node and the sink node is high, otherwise it is low. After applying a min-cut/max-flow algorithm to the graph, the source-connected nodes in the resulting cut are identified as the foreground of the object and the sink-connected nodes are identified as background. If necessary the user can re-label nodes and the segmentation process repeated and final segmentation can be hand tweaked as necessary.

Figure 6 shows the process of segmenting an object from a 3D scan. The user identifies sample portions of the image that belong to the background and portions that belong to the foreground. The lazy snapping system then automatically segments the object representation into meshes belong to the foreground and background. The foreground mesh can then be inserted into the database representation of the crime scene as a separate object.

4 A multimedia database for crime scene representation

Crime scene data poses several challenges for its storage and processing by a database system.

- **The size of the data.** The iSM sensor can collect as much as 500GB of data per hour and it may take many hours to capture a large crime scene. The 3D model rendered from the raw data is an order of magnitude larger, thus, a model of a single crime scene can be in terabytes.
- **The complex relationships between database objects.** Most of the objects are related through a hierarchical ordering, but even more complex relationships are present. An underlying principle in the representation is the hierarchical nature of data collected at crime scenes. In general, entities are associated with larger structures: They may contain one another (e.g., the gun is in the drawer) or the process of data collection enforces specific relationships between entities (e.g., the coroner’s report is associated with
Fig. 7. Hierarchical scenes. (a) Shows an overview of a simulated scene. This is a view of the 3d model which can be repositioned by the user as shown in (b). Each object in the scene can be manipulated in a variety of different ways. (c) Shows a subscene within the primary scene.

the body of the victim). Another property of crime scene data, especially the data associated with 3D scene acquisition and other sensor systems, is that the data may be available at a variety of different resolutions. The use of different resolutions may be related to performance issues (e.g., it is more efficient to render the 3D representation of the crime scene with fewer polygons), or it may be necessary to view the crime scene with certain objects removed or obscured (e.g., to protect the identity of the victim).

- **The large variety and inefficiency of functions and tools for interacting with 3D objects.** A typical database system provides few distinct ways of representing the same piece of data. For 3D data, there is a wide spectrum of such representations. A user may retrieve an object from a database and modify it in various ways (for example, by taking a 2D snapshot, changing resolution, removing certain objects from it, etc.). These operations may be time consuming and it may be necessary to store these user-generated objects for future use.

- **The importance of the user interface.** The client interface must provide the user with the ability to easily interact with and modify the data. From the database system perspective, the hardest task is to clearly describe the relationships between objects (for example, that one object is a 2d snapshot of the scene or a low resolution version of it). There is no formal language like SQL to describe precisely to the user how the objects are related to each other. As the number of such objects becomes larger the task of representing efficiently and clearly their relationships becomes a challenge.

Given a set of segmented and labelled scans and information related to more traditional crime scene data a mechanism is required that provides a uniform method for access. Within the C2SM project this uniform access is
Fig. 8. A real environment within the viewer. Shown here is a simulated crime scene modelled within an abandoned warehouse. The robot’s path is indicated by small spheres. Sensor measurements are indicated by larger spheres and associated directional markers. Two views are shown.

provided through a multimedia database system that associates with each item in the database a 3D frame that represents where that item was found with respect to the scene. Using a database system to represent the crime scene data provides dual functionality: it provides persistent storage of the raw data acquired from the crime scene and access and manipulation methods to support analysis of the scene data for investigators. The first function is important for providing evidence in a court of law. Here, the data must be represented in an unaltered state so that the court can view the evidence as it was when it was recorded. The second function is important to aid the investigators by providing the data they require in a quick and easy to understand manner.
Fig. 9. Interacting with the viewer. (a) Sensor readings and scene annotations are represented in the 3D view as small geometric primitives. Selecting a primitive brings up a menu showing possible interactions with that sensor reading or annotation. Here the user has selected a high resolution still image that was taken from a particular point in the scene. (b) Selecting the ‘View Image’ option from the menu of choices for this object brings up a high resolution still image of the scene taken from this vantage point.

In order to provide some overall structure to the data associated with a crime scene the database utilizes a hierarchical representation. At the root level there is the entire crime scene. The children of the root level may include such things as a body or a table in the room. The children of the table may
include a box resting on top of it. The children of the box may include a pen inside the box, and so on.

In addition to providing a natural mechanism for representing the property that some objects are contained within others, the use of a hierarchical representation also provides a natural mechanism for representing the fact that certain objects in a crime scene are treated as separate ‘scenes’ in terms of processing. For example, a weapon or body may have significant information associated with it that may be collected at a different site (e.g., at the crime lab or coroner’s lab) or at a different time. A hierarchical representation provides a straightforward mechanism for incorporating this data within the overall representation. This relationship is illustrated in Figure 7. Figure 7(a) shows an overview of a simulated scene, and a closeup of part of the scene is shown in Figure 7(b). Within the scene the desk was identified as being of particular importance and was encoded as a subsence. Selection of the subsence focuses the viewer to reflect only objects contained in that subsence (here the desk and its contents).

An object in the hierarchy may have one of a range of different views. These views may represent the object at different resolutions (so as to obtain display efficiencies when manipulating the dataset) or views may intentionally obscure certain features of the scene (so as to obtain sanitized views of the database for public consumption). One view of any object is its ‘raw’ view that provides an unaltered view for criminal proceedings.

Objects within the hierarchy can be associated with other primitive entities or views. These other views correspond to different classes of evidence that may be identified with a crime scene, including

- ‘Traditional’ objects such as specific objects described via standard paper forms.
- Multimedia objects such as video or audio clips that describe an object located at a specific location.
- Textured 3D meshes that describe the surfaces of objects and locations within the crime scene.
- Spatial distributions such as levels of radiation, temperature or gas concentrations.

Figure 9 illustrates how this association acts in practice. Figure 9(a) shows a close up view of a portion of a crime scene which includes a number of objects with associated views. For example, the small green spheres are associated with points along the motion path of the vehicle. The small blue semi-transparent cube corresponds to a position in the scene from which the robot collected a high resolution image. Clicking on the cube (as shown in Figure 9(a)) brings up a menu of possible interactions with the cube, one of which is to view the high resolution image captured at this point (Figure 9(b)).

All entities in the database have an associated renderable view. For 3D scans, the scan itself provides that view. For other objects the renderable view is a three dimensional graphical token that can be rendered in the scene.
Fig. 10. XML representation of an object in the database. Objects encapsulate sufficient information to display and manipulate with them. They also encode interactions that are possible with the object.

The renderable view of each object is maintained as a node within a global 3D scene graph (see [15]). Labelled nodes within the scene graph correspond to entities in the database. The process of displaying (enabling) specific entities within the scene are encoded as operations on this global scene graph.

The system and the database must be able to unobtrusively exchange information about the objects, their representations, their relationship with other child/parent objects, and available object manipulation actions. The eXtensible Markup Language (XML) is well suited for sharing this information and decouples the database interface and the user interface implementations. XML is used to encode the set of possible interactions the user can make with respect to a specific object in the database. Figure 10 shows a sample encoding for the “Body of a victim” from the demo scene from Figure 7. The media tags describe the various views of this object. This includes providing sufficient information for the renderer to manipulate the scene graph to show the appropriate view. The report tags describe external information associated with this entity, here a set of three different PDF documents that can be viewed. Finally the info tag provides a natural language description of this entity.

The database uses a standard relational database model to represent information about the scene. Every object in the scene is identified with a unique label and one table in the database maintains a mapping between this label and its corresponding XML description file and the nodes within the scene graph that are used to render the object. The user interface allows users to query the database using standard SQL queries and only displays objects in the current scene that much the current query.

5 Summary

This chapter provided an overview of the CBRN Crime Scene Modeller (C2SM) project – a project whose goal is the development and field evaluation
of technologies for data acquisition at CBRN crime scenes, and fast recreation of such scenes as virtual environments with access to all multi-modal data and heterogeneous evidence collected at the scene.

Crime scene investigation requires the collection and analysis of vast amounts of heterogeneous data. Within the C2SM project a centralized multimedia database system is used to provide the necessary set of access and storage primitives to access the data. All objects in the scene are associated with a visual token (typically a polygonal mesh) with a specific pose in 3D and this embedding within space is a central aspect of the storage structure.

Acknowledgments

This work was supported by MDA and Chemical, Biological Radiological-Nuclear and Explosives Research and Technology Initiative (CRTI Project 05-0122TD). The financial support NSERC to M. Jenkin and J. Gryz is gratefully acknowledged.

References