3-DIMENSIONAL VOLUME REPRESENTATION FOR GEOSPATIAL DATA IN VOXEL MODELS

Franklin Tanner, Contractor
Daniel Crispell, Contractor
Raymond Isbell, Science and Technology Program Manager
National Geospatial-Intelligence Agency
Springfield, VA 22150
franklin.r.tanner.ctr@nga.mil
dancrispell@gmail.com
raymond.m.isbell@nga.mil

ABSTRACT

Extracting useful geospatial data from imagery is a fundamental challenge that has seen significant growth over the years as technology advances have been brought to bear on the problem. An important component of this problem addresses how the data should be represented to ensure the information content is accurately captured, preserved, and conveyed to consumers. Much of the information contained in the imagery is redundant and should be transformed so that only the essential information is retained and stored, allowing the redundant data to be discarded. An efficient mechanism for achieving this goal is the 3D Voxel model. Essentially, it takes as input the two dimensional information contained in a set of images and transforms it into a 3D volumetric representation of the imaged region of interest. There are immediate and enduring benefits to this approach. First, there will be a massive reduction in data storage requirements, particularly for large sets of overlapping imagery. Another benefit stems from improvement in visualization. Terrain, objects, and features appear more intuitive and natural in a 3D Voxel model, and therefore, convey information easily. There are many others. This paper addresses the progress being made by a number of US Federal Government organizations in developing Voxel technology, and discusses the technology's application to a number of military and civilian applications.

KEYWORDS: 3D, modeling, visualization, data reduction, compression, voxels, scene reconstruction

BACKGROUND AND MOTIVATION

Data Explosion

The U.S. government faces an enormous challenge in dealing with data overload. Rapid expansion of Unmanned Aerial Vehicles (UAVs) as the Intelligence, Surveillance, and Reconnaissance (ISR) collection platform of choice has exacerbated the information overload problem. Early UAV platforms hosted sensors capable of collecting less than one *mega-pixel* per frame and less than 10 frames per second. The newest generation of UAV sensors such as DARPA's Autonomous Real-Time Ground Ubiquitous Surveillance Imaging System (ARGUS-IS) platform can collect up to 1.8 *giga-pixels per frame at 10 frames per second.* A single ARGU-IS deployed sensor can collect 38 terabytes per day. The cost of developing, operating, and maintaining a 90 day archive is approximately $10k per terabyte. A ninety day archive then would cost ~$34 million dollars plus recapitalization costs every 3 to 5 yrs. The required bandwidth to support the dissemination of this volume of data drives-up the cost even further. These costs are prohibitive, and thus a solution is needed to reduce the amount of data collected, stored, and disseminated.

There are two approaches that can address the problem. One uses traditional (or perhaps even new) image compression techniques that can be either lossless or lossy. Image quality issues have essentially killed the lossy image compression approach because the image quality falls off dramatically the more aggressive the compression algorithm becomes. On the other hand, lossless techniques do not offer sufficient compression ratios to mitigate the data volume problem. Clearly, a new approach to reducing the volume of data collected in ISR operations is needed if success in irregular warfare is to result. This paper presents a new approach that may solve the problem based on Probabilistic Volumetric Models (PVRs).

Redundant Information

The magnitude of the ISR imagery data collected presents challenges to the entire PED pipeline. A particularly important aspect of this is how data should be represented to ensure the information content is accurately preserved and conveyed to consumers for exploitation and for forensic research. For many ISR applications, a majority of the physical scene does not change compared to previous collections.

Figure 1 contains several images taken from a helicopter flying over the Rhode Island capitol building. A majority of the physical scene does not change over time. Aside from the vehicles moving on the road, there is very little scene variation from image to image.

Figure 1. Image frames of Rhode Island Capitol building from video sequence (imagery courtesy Brown University).

This leads to the question, "If a majority of the physical scene does not change, then why collect, process, maintain, and disseminate the pixels with redundant content?" If there is a model of the scene independent of the current image being collected, could a majority of that image be discarded as redundant? Clearly, the answer is yes. The question remaining is "Can 3D models answer this challenge?" The aforementioned PVRs provide one solution for generating 3D scene models that can be used to identify and reject redundant imagery data being collected.

If 3D models can identify and reject redundant imagery data, how would it work? Once a 3D model of a scene has been constructed over an Area of Interest (AOI), the model can be rendered from any arbitrary viewpoint by projecting the model into an arbitrary camera frame. In essence, the model provides a *prediction* of how a scene should appear. A simple example is shown in Figure 2. The onboard model is able to predict how the scene should appear. Given a newly collected video frame, onboard processing can extract the movers and other relevant changes and transmit only that small amount of information to the ground (the red blobs in the "Movers" image).

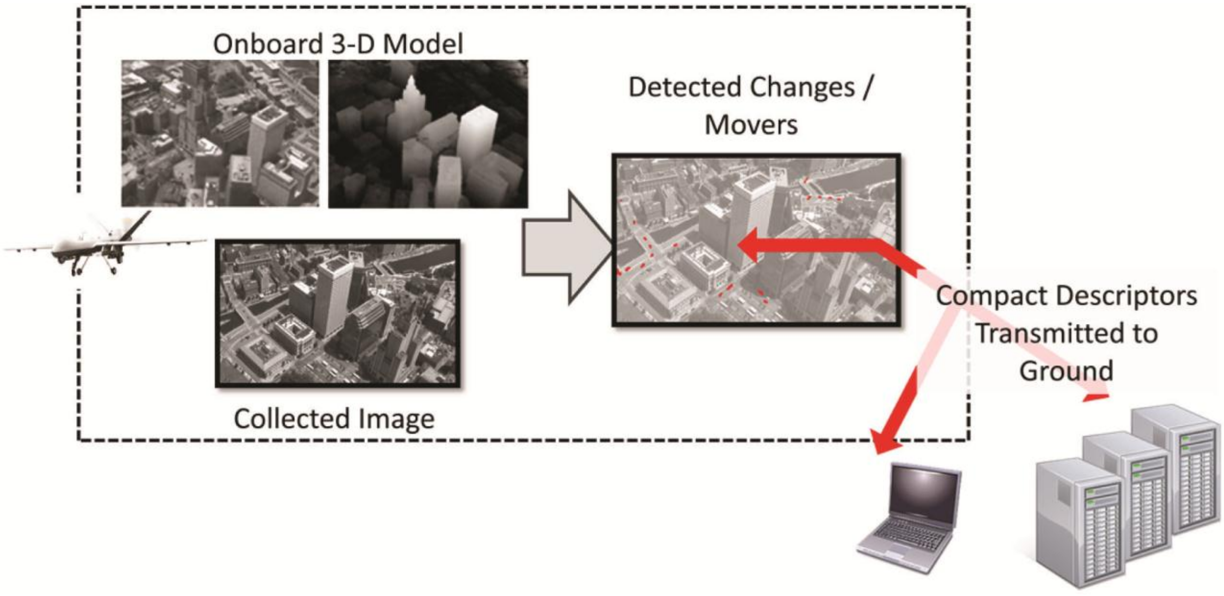


Figure 2. PVR model can predict the scene appearance and extract the relevant information for transmission to the ground.

CONCEPT OF OPERATIONS

A PVR could be employed at various points in the PED pipeline. The concept is illustrated in

Figure 3. The process begins with the creation of a PVR at the factory. The factory is generic concept that represents the origination body of the PVR. PVR creation involves the backprojection of image data from multiple viewpoints into a 3-D grid of voxels. The photoconsistency constraint is used to determine which voxels exist in "empty" space, and which likely exist on visible surfaces in the scene. The photoconsistency constraint simply states that an "occupied" voxel will project to consistent pixel intensities in all views in which it is visible. Conversely, the projections of "empty" voxels are more random in nature. As the reconstruction process progresses, visibility and occlusion relationships are learned and taken into account. Each voxel in a completed PVR model contains a representation of the likelihood that the voxel contains a surface, as well as a probabilistic representation of the surface appearance. For a detailed description of PVR creation, the reader is directed to (Crispell D. M., 2011). We propose using a high-resolution (30cm or better), hi-accuracy base map or similar alternative to initialize and geo-register a PVR and serve as the source for developing the initial baseline model with resolution equivalent to the base map. This approach is taken to ensure the fidelity of the models is acceptable in lieu of imagery to the user community. Note, the base map required for model initialization may be more complete than a traditional base map. For example, PVR construction requires oblique imagery data along with intensity information that may not be present in a traditional base map.

Once created, the model is then disseminated to sensing platforms, remote ground stations, and tactical users' handheld devices. At this point, the factory and all of the sensing and exploitation platforms contain the same volumetric model. The volumetric model is stored at the factory as the standard representation of the AOI.

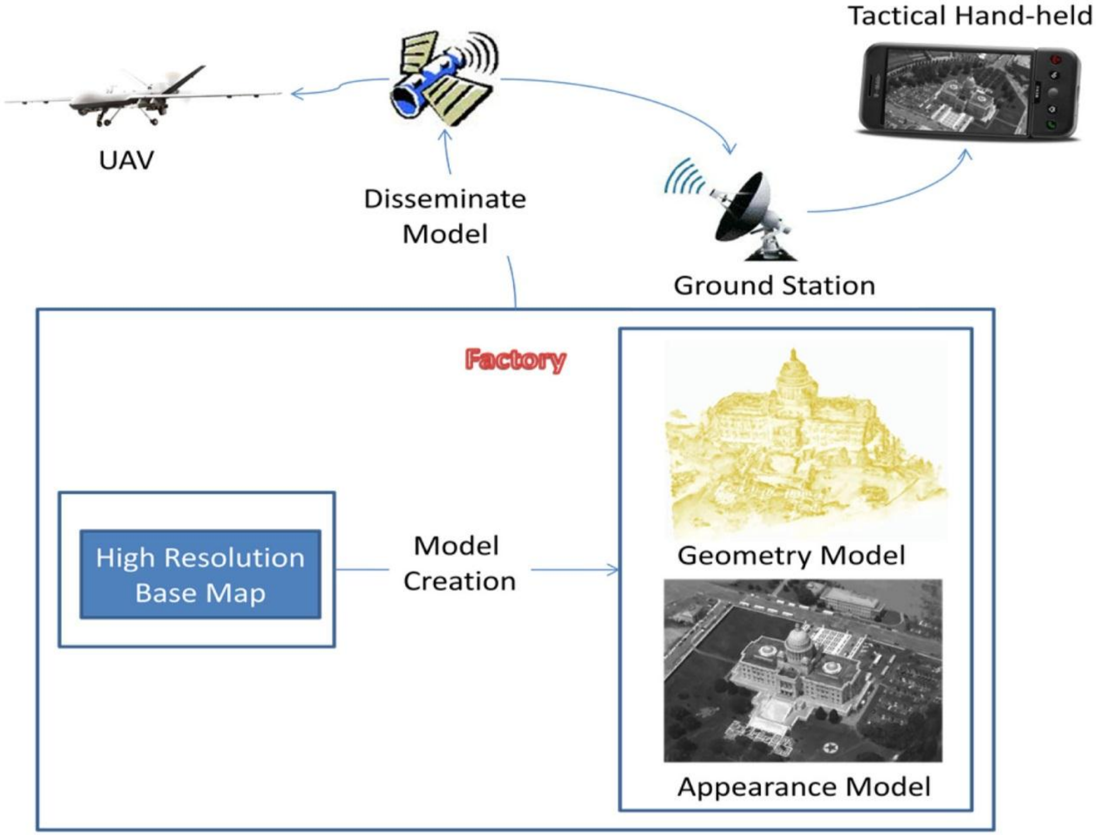


Figure 3. Model Creation at Factory.

Alternatively, for a long dwell UAV, the model could be built on-board using one pass of data and transmitted to the ground. For simplicity, we describe the CONOP where the model is built at a factory on the ground using a hi-resolution base map as described above.

Now, the UAV platform begins collecting new imagery. Since the camera pose information is well known, using the model we can predict the appearance of the scene. Based on this prediction, we can determine what is different in the scene relative to the model. This is the cornerstone of the data reduction effort; instead of sending entire frames of data to the ground or as compressed H.264 data stream, the volumetric model only sends *unexpected* pixels (i.e. new information) to the ground along with the camera pose information. We are calling this technique model-based data reduction (MBDR).

On the ground, all of the necessary data is present to perform a completely lossless reconstruction of the image observed by the UAV. The volumetric model uses the camera pose information to predict what the image should look like then the change image is added to the predicted image to reconstruct the *exact* frame that the sensing platform collected. This significantly reduces the amount of communication bandwidth required to display the information collected by the UAV.

In some cases, however, the reconstruction of the original video frames may not be desired. Indeed, the CONOPS includes the fact that the PVRs will be disseminated to end users at all echelons. Using the PVR, we can transmit vector data or shape files instead of pixel data. This data can then be overlaid on the PVR to visualize the geolocation of movers or other sources of change in the model. Model updates can also be disseminated so that the end users can quickly update their models based on the latest collections. This is particularly important to tactical users that may have limited communication bandwidth. By transmitting shape files instead of imagery, we can enable a bandwidth disadvantaged field user with a hand held device to receive important tactical information in near real-time.

Limitations

Despite recent advances in volumetric modeling technologies, there are still numerous challenges remaining. High quality camera models that can provide precise camera pose information are still mostly unavailable. Both the capitol and the downtown sequences had no camera model information available hence we were forced to use Snavely, et.al.'s "Bundler" software to solve for the camera pose information (Snavely, 2006). In order to produce high quality volumetric models, camera pose information should be accurate to the resolution of the voxel data. For example, if the volume defined has 1 meter voxels, the camera model parameters must be sufficiently accurate to provide backprojection within1 meter. Even Bundler does not consistently provide this level of accuracy. This manifests itself in high frequency noise around edges of buildings where extremely small registration errors cause changes to be flagged that are not true changes.

State of the art PVR methods have been ported to high performance GPUs to improve performance. However, building large models is still computationally intensive and has large memory requirements. Processing volumes on an ARGUS-IS scale in terms of both geospatial resolution and at the required frame rate is not feasible using the current generation of PVR software.

Model Lifecycle Management

The lifecycle management of a volumetric model is still an open question. For example, when should a new model be generated? Should only a single model be created and utilized or should the model be continuously updated and refined with more information?

PVR modeling technology is a background modeling technique. It can be thought of as representing the average state of the scene over time. Seasonal variation in imagery collected over long periods of time will cause some individual voxel probability distributions to be flatter. In other words, given a scene that contains a tree with leaves and the same scene with the same tree without leaves, the model will not be able to predict with high confidence the state of the tree (i.e. with or without leaves). This begs the question, should different models be used for different times of year? Likewise, shadows are present in different places depending on sun position. How should the model handle illumination changes over the day? A bidirectional reflectance distribution function (BRDF) or other physics based modeling approach may help alleviate this issue. However, those approaches remain beyond the current state of the art.

Other Considerations

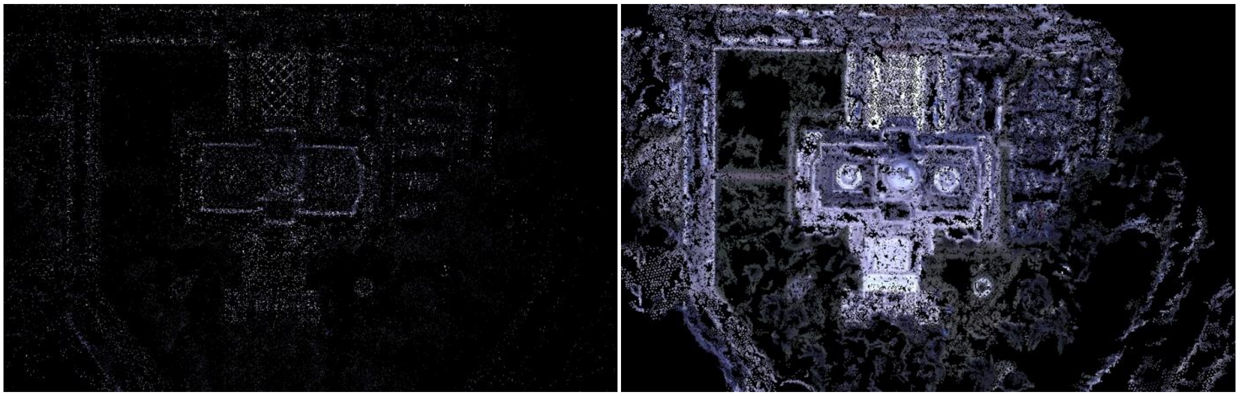
Some of the obvious questions to be answered include, "How detailed must the model be in order to be operationally useful?" Higher resolution models require greater computational resources, and result in larger file sizes which themselves pose challenges in terms of storage and dissemination. The tradeoff between model fidelity verses processing and PVR storage requirements is still an open question.

Today, most 3D graphics systems rely on polygonal meshes or 3D points to represent data. For example, the Web3D Consortium has defined the VRML standard and its successor the ISO ratified Extensible 3D Graphics (X3D) standard. At its most detailed level, X3D uses vector primitives such as NURBS and point based notation to represent 3D content (Web3D Consortium 2011). However, an open, volumetric, 3D data representation standard is still not available. While a PVR model can be output as VRML or as a polygonal mesh, these data formats do not include a way to represent probability distributions within each voxel. Even if a standard did exist, there is limited tool support to read and visualize this type of data.

Developing an open 3D standard was the focus of a recent DARPA program called Geometric Representations Integrated Data-spaces (GRID). GRID attempted to create a standard 3D data format similar to what MP3 is to music. After some initial investment into developing a standard, DARPA stopped funding the effort. The ISR community is currently without an open standard for 3D volumetric data.

Alternative Automated 3D Reconstruction Technologies

PVR is one technology that can automatically reconstruct 3D geometry. For example, Snavely, et.al's Bundle Adjustment software is able to recover 3D geometry and create a sparse point cloud reconstruction of the scene (Snavely, 2006). Likewise, Furukawa, et.al's Patch-based Multi-View Stereo (PVMS) algorithm is able to perform a dense reconstruction (Furukawa, 2009). Figure 4 below illustrates sparse and dense 3D reconstructions of the Rhode Island Capitol building on the left and right respectively. Lawrence Livermore National Laboratory's Persistics program has the capability to reconstruct 3D geometry using a dense correspondence method (Heller, 2011). PVR methods offers some advantages and disadvantages over these approaches. However, contrasting these approaches is beyond the scope of this paper and will be the subject of a future investigation.



(a) (b)

Figure 4. (a) Sparse 3D Reconstruction using (Snavely, 2006) and (b) dense reconstruction using (Furukawa, 2009). Both reconstructions produced using VisualSfM software (Wu, 2011).

SUMMARY

The massive volume of ISR data being collected now and in the future requires a solution for the processing, exploitation, and dissemination of that data. In this paper, we present a concept for model-based data reduction based on probabilistic volumetric representations (PVR) of a scene's geometry and appearance. By having a reference model that can predict a majority of a scene's content, redundant information can be identified and eliminated greatly reducing the amount of data that is disseminated and stored.

REFERENCES

Crispell, D., Mundy, J., Taubin, G., 2011. A Variable-Resolution Probabilstic Three-Dimensional Model for Change Detection, *IEEE Transactions on Geoscience and Remote Sensing*

Furukawa, Y., Ponce, J., 2009. Accurate, Dense, and Robust Multi-View Stereoposis, *IEEE Transactions on Pattern Analysis and Machine Intelligence*.

Heller, A., 2011. From Video to Knowledge. *Science and Technology Review*, pp. 4-11.

Snavely, N., Seitz, S., Szeliski, R., 2006. Photo Tourism: Exploring image collections in 3D. *ACM Transactions on Graphics (Proceedings of SIGGRAPH 2006)*, 25(3): 835:846

Wu, C., 2011.VisualSfM. http://www.cs.washington.edu/homes/ccwu/vsfm/ (accessed October 18, 2011).

PUBLIC RELEASE

This paper was cleared for public release under NGA case 12-107.