Instant Scene Modeler for Crime Scene Reconstruction

Stephen Se and Piotr Jasiobedzki
MDA, Space Missions
9445 Airport Road, Brampton
Ontario, Canada L6S 4J3
{sse,pjasiobe}@mdrobotics.ca

Abstract

Crime scene reconstruction is a labor-intensive but an essential task of police investigation. Currently, the police spend a lot of time at the crime scene taking many measurements and photos, and then spend more time at the computer recreating the crime scene drawings manually. Automatic generation of photo-realistic three-dimensional calibrated models using a hand-held device is highly desirable for crime scene reconstruction as well as for many other applications.

In this paper, we present the instant Scene Modeler (iSM), a 3D imaging system that automatically creates 3D models using an off-the-shelf hand-held stereo camera. The user points the camera at a scene of interest and the system will create a photo-realistic 3D calibrated model automatically within minutes. Forensics field tests have been carried out with promising results and the advantages of 3D imaging for this application are described.

1. Introduction

Creation of photo-realistic three-dimensional (3D) calibrated models of observed scenes and objects has been an active research topic for many years. Such 3D models can be used for both visualization and measurements, and are useful for many applications including planetary rover exploration [27], forensics, mining, geology, archaeology [31], virtual reality, etc.

Object scanning and environment modeling can be regarded as two types of 3D modeling. Outside-looking-in approach is more suitable for object scanning whereas inside-looking-out approach is more suitable for environment modeling. 3D data can be obtained using various rangefinders, computed from stereo images or monocular sequences. The raw 3D measurements are then converted to representations suitable for display and manipulation.

A hand-held device is desirable in many situations as it can be used for scanning by simply moving it freely without any constraint on the motion. This allows the user to position the sensor relative to the scene to get the optimal view and full coverage. The capability of creating 3D models automatically and quickly is particularly beneficial for a number of applications ranging from forensic and traffic investigations, geology, to home and office modeling.

In this paper, we describe a novel 3D imaging system, the instant Scene Modeler (iSM). iSM creates 3D models quickly using an off-the-shelf hand-held stereo camera. The user moves the hand-held camera freely in all six degrees of freedom (dof) and the system creates a calibrated 3D model automatically within minutes. There are no tracking devices on the camera and the system recovers the camera ego-motion and integrates the 3D data using only the stereo images. iSM outputs a photo-realistic 3D surface, represented as a triangular mesh with mapped colour texture.

In Section 2, we will briefly review some previous work on 3D modeling. Afterwards, we will describe the system architecture and different modules of iSM in Section 3. Experimental results will be presented in Section 4 and we will look at how iSM can be used for the crime scene reconstruction application in Section 5. Finally, we conclude and outline some future work in Section 6.

2. Previous Work

3D modeling has been a topic of intensive research for the last few decades. Instead of providing a comprehensive literature survey, we will review some previous work on 3D modeling related to iSM in this section.

The key components of iSM include depth acquisition, view registration and model construction, therefore, we will review the various approaches for these tasks. Moreover, we will look into several other hand-held systems and compare them with iSM.

2.1. Depth Acquisition

The main approaches for depth acquisition include structured light, laser scanning and stereo. The structured light approach uses a projector to illuminate the object with patterns and recovers the 3D shape from a monocular image. It is very effective for scanning objects including the human face and body [17] but does not work well for scanning environments due to their limited range.

Blais [3] has recently reviewed the development of 3D laser imaging for the past 20 years. Autosynchronous laser

0-7695-2372-2/05/$20.00 (c) 2005 IEEE

scanners can be used for both objects and environments due to their long depth of field and high accuracy at close range.

Time-of-flight scanning laser rangefinders measure the time it takes for the light to travel to the object and back. They can acquire dense depth data, depending on the time-of-flight measurement principle, even for ranges up to 100s of metres. Continuous wave (CW) devices operate at distances up to 10s of metres with resolution of several millimetres, e.g., DeltaSphere [1] and Surphaser [29]. Pulsed laser scanners operate at distances up to 100s of metres but with lower resolution of 10-30 mm, e.g., ILRIS-3D [15] and Riegl [22]. Some of the scanning laser rangefinders can create full 360 degree scans of the environment. Laser range scanners require fast scanning mechanisms to move the beam rapidly through the scene. They have to remain stationary during data acquisition and they are typically large, heavy, and tend to be expensive.

Laser scanners, such as ILRIS-3D and Deltasphere, have been tested by some police departments for crime scene reconstruction. But the requirement of long stand-off distance between the scanner and the scene prevents the scanner from working in small enclosed regions.

Stereo imaging is a passive technique and can recover the structure of the environment by matching features detected in multiple images of the same scene. It is the most computationally intensive as the 3D data must be computed from the images. The depth data could be much noiser than the other approaches, as it relies on natural texture on the surface and the ambient lighting. Unlike the scanning rangefinders that acquire data sequentially, cameras can capture complete images in microseconds and hence they can be used as mobile sensors or operate in dynamic environments. The cost, size, mass and power requirements of stereo cameras are much lower than those of scanning rangefinders. The lack of moving parts allows cameras to be used in rugged environments.

2.2. View Registration

When multiple scans are obtained, they need to be registered together to build the 3D model. Registration can be carried out with a separate device that tracks the sensor or object position, or by matching the data sets manually (by selecting corresponding features in different scans) or automatically. The separate device can be a tracker such as Polhemus FastScan [17] that locates the sensor continuously or a turntable on which the object is placed such as Cyberware Model Shop [4].

The most common algorithm for automatic 3D data registration that uses the 3D data itself is Iterative Closest Point (ICP) algorithm [2], which iteratively minimizes the distances between the overlapping regions of two set of 3D points or surfaces. An estimate of the initial scan location is usually required to start the process.

Photogrammetry techniques (e.g., PhotoModeler [16]) can be used to create 3D models from sequences of monocular images, by manually establishing correspondences between features in different images to estimate their 3D coordinates. However, it is a time consuming process and is usually limited to simple objects with polygonal faces. Additional measurements are required to recover the scale of the model.

For vision systems, fiducials can be placed in the scene and the camera pose can be estimated by tracking these markers [12]. However, this involves changes to the environment and it is not possible for forensics applications, as crime scene should not be tampered with. The capability to track natural features in the scene to recover camera motion is much preferred.

2.3. Model Construction

Registered 3D data sets contain redundant overlapping measurements and measurement noise. They contain often too much details for efficient visualization and manipulation, and they need to be converted to other formats. One approach involves constructing geometrical models, e.g., 3D surfaces or volumes. Triangular meshes that consist of a large number of triangles are often used as they can represent complex surfaces. It is desirable to create non-redundant meshes with no overlapping faces.

The non-redundant meshes can be obtained by creating surface meshes from individual views first and then stitching them together [30]. If there is a significant overlap between the individual views, this approach is rather inefficient due to the need for repeated stitching. A volumetric approach is more efficient in such situations as the 3D points are accumulated into voxel grid structures first. Then only one triangular mesh is created for all the measurements using an iso-surface extraction algorithm, such as the marching cubes [23].

After the triangular mesh is generated, texture images are mapped to provide the photo-realism [28]. A separate camera can be calibrated with a laser scanner to provide texture images. For vision systems, the input camera images can also be used as texture. As a model requires multiple texture images, visible seams across the image boundaries may appear and the transition needs to be smoothed [5].

An alternative to surface representation is QSplat [25]. QSplat combines a multi-resolution hierarchy based on bounding spheres with a rendering system based on points. It has been used to display large geometric data sets produced in the Digital Michelangelo Project in real-time.

2.4. Hand-held Devices

Pollefeys *et al.* [18] and Nister [13] presented systems which create 3D surface models from a sequence of images taken with a hand-held video camera. The user acquires the images by moving the camera around an object or moving with the camera in the environment. The camera motion

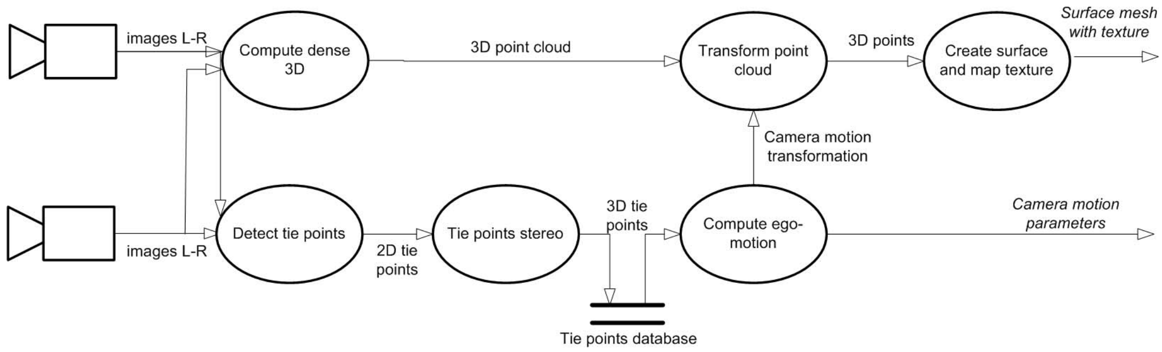


Figure 1. ***iSM system architecture.***

is recovered by matching corner features in the image sequence. Dense stereo matching is carried out between the successive frames. The input images are used as surface texture to produce photo-realistic 3D models. However, it is a scaled version of the original object as the metric size cannot be recovered from sequences of monocular images. Monocular approaches does not work well for some critical motion sequences.

Hebert [8] proposed a self-referenced sensor which consists of two cameras and a cross-hair laser light projector. Frame to frame registration is achieved using a set of fiducials projected with an additional stationary laser system. The system requires a long acquisition time as it can capture only sparse 3D data for each frame and the 3D models do not have photo-realistic appearance.

Rusinkiewicz *et al.* [24] presented a real-time 3D modeling system that allows the user to rotate an object by hand and see a continuously-updated model as the object is scanned. It consists of a 60Hz structured-light rangefinder and a real-time variant of ICP for alignment. It is limited to the outside-looking-in case and does not acquire colour.

Popescu *et al.* [19] proposed the ModelCamera, which is a low-cost hand-held scene modeling device. It consists of a digital video camera with 16 laser pointers attached to it. ModelCamera acquires 16 depth samples per frame and registers the frames using depth and colour information. The frames are merged together into a texture-mapped model. The surfaces are approximated with a few quadratics and this approach only works for smooth continuous surfaces.

iSM, described in the next section, uses stereo cameras to obtain 3D data, estimate camera motion and register successive frames together. The stereo camera approach allows free camera motion without any restrictions. The resulting models are fully calibrated (allow Euclidean measurement) and have photo-realistic appearance. The data acquisition and processing takes minutes.

3. Instant Scene Modeler

The main hardware components of iSM are a stereo camera and a computer. We currently use a colour Bumblebee stereo camera from Point Grey Research (PGR) [21] at 640x480 image resolution. It is a firewire camera and can capture up to 15 frames per second. iSM software can run on any PC with firewire interface.

Figure 1 shows the architecture of the iSM system. Images are captured by the stereo camera and dense stereo disparity is computed for each stereo pair to obtain 3D data using known camera calibration. The system does not require any external sensors for computing the camera motion as it automatically extracts and tracks natural tie points in the images. The recovered camera motion is used to integrate 3D data obtained from the sequences. The 3D data is then converted to surface meshes, which are augmented by mapping texture from the colour camera images.

3.1. Dense Stereo

The left and right images are matched to obtain dense disparity data, which is then converted to 3D depth data. We run PGR's optimized Triclops library for correlation-based dense stereo. As with other stereo algorithms, the quality (accuracy, coverage and number of outliers) of the depth data depends on the presence of texture in the images.

3.2. SIFT Extraction

We have chosen to use a high level set of natural visual features called Scale Invariant Feature Transform (SIFT) as the tie points to compute the camera motion. SIFT was developed by Lowe [11] for image feature generation in object recognition applications. The features are invariant to image translation, scaling, rotation, and partially invariant to illumination changes and affine or 3D projection. These characteristics make them suitable as landmarks for robust matching when the cameras are moving around in an environment. Such natural landmarks are observed from different angles, distances or under different illumination.

Previous approaches to feature detection, such as the widely used Harris corner detector [7], are sensitive to the

(a) (b)

Figure 2. *(a) Stereo matched SIFT features. (b) SIFT features that are matched to the database.*

scale of an image and therefore are less suitable for building feature databases that can be matched from a range of camera positions. A comparison between Harris corners and SIFT features is shown in Table 1.

The SIFT features are detected by identifying repeatable points in a pyramid of scaled images. Feature locations are identified by detecting maxima and minima in the Difference-Of-Gaussian pyramid. A subpixel location, scale and orientation are associated with each SIFT feature. In order to achieve high specificity, a local feature vector [11] is formed by measuring the local image gradients at a number of orientations in coordinates relative to the location, scale and orientation of the feature.

The local and multi-scale nature of the features makes them insensitive to noise, clutter and occlusion, while the detailed local image properties represented by the features make them highly selective for matching to large databases.

3.3. Camera Ego-motion Estimation

With known stereo camera geometry, the SIFT features in the left and right images are matched using the following criteria: epipolar constraint, disparity constraint, orientation constraint, scale constraint, local feature vector constraint and unique match constraint [26]. The subpixel disparity for each matched feature is also computed. The stereo matched SIFT features in a lab scene are shown in Figure 2(a), where the length of the line is proportional to the disparity. We can see that all the matches are consistent and correct. Typically, we obtain hundreds of SIFT 3D features.

Subsequently we can compute the 3D position (X, Y, Z) of each stereo matched SIFT feature, using the following equations:

$$X = \frac{(u - u_0)I}{d}; \qquad Y = \frac{(v_0 - v)I}{d}; \qquad Z = \frac{fI}{d}$$

where (u, v, d) are the SIFT image location and disparity, (u_0, v_0) are the image centre coordinates, I is the baseline distance and f is the focal length.

For our hand-held camera, we recover the 6 dof camera ego-motion when the camera moves freely in 3D. We employ a Simultaneous Localization And Mapping (SLAM) approach that uses the SIFT features to localize and simultaneously build a database map [26]. Instead of frame to frame matching, we match the SIFT features at each frame with the database to reduce error accumulation. Olson *et al.* [14] reported a 27.7% reduction in rover navigation error when multi-frame tracking is used, rather than considering each pair of frames separately.

The local image vector is used to match the highly distinctive SIFT features with very few false matches. We can then find the camera movement that would bring each projected SIFT feature into the best alignment with its matching feature. A weighted least squares procedure is carried out taking into account the feature uncertainty. Features with large least squares errors are discarded as outliers.

Figure 2(b) shows the SIFT features that are matched to the database. The line connecting the previous position to the current position is analogous to optical flow and we can see that all the matches are consistent and correct.

3.4. Mesh Creation

As the camera moves around, dense 3D data is obtained relative to the camera position at each frame. All data sets must be transformed to one reference coordinate system before they can be combined together. We have chosen to use the initial camera pose as the reference and all 3D data sets are transformed to this coordinate system using the camera pose estimated for each data set.

Using all 3D points obtained from the stereo processing is not efficient as there is a lot of redundant measurements, and the data may contain noise and missing regions (due to incorrect matches or lack of texture). Representing 3D data as a triangular mesh reduces the amount of data when multiple sets of 3D points are combined. Furthermore, creating surface meshes fills up small holes and eliminates outliers, resulting in smoother and more realistic reconstructions.

To generate triangular meshes as 3D models, we employ a voxel-based method [23], which accumulates 3D points with their associated normals. It creates a mesh using all the 3D points, fills up holes and works well for data with significant overlap. The 3D data is accumulated into voxels at each frame. Outliers are filtered out using their local orientation and by selecting the threshold of range measurements required per voxel for a valid mesh vertex. It takes a few seconds to construct the triangular mesh at the end, which is dependent on the data size and the voxel resolution.

3.5. Texture Mapping

Photo-realistic appearance of the reconstructed scene is created by mapping camera images as texture. Such surfaces are more visually appealing and easier to interpret as they provide additional surface details. Colour images from the stereo camera are used for texture mapping. In general, other modalities such as Infra-red (IR) images can be used

	Harris corners	**SIFT features**
Algorithm complexity	Easy to detect	Complex detection algorithm
Localization accuracy	Sub-pixel	Sub-pixel
Scales	Single or multiple scales	Multi-scale representation
Description	Image windows	Specific local image feature vector
Matching method	Correlation between image windows	Distance between feature vectors
Validation	Required, as many mismatches	Not required, as unique matches

Table 1. ***Comparison between Harris corners and SIFT features.***

(a) *(b)*

Figure 3. ***(a) iSM hand-held stereo camera. (b) iSM laptop carried by the user on a harness.***

for texture mapping, as long as the IR camera has been calibrated with the stereo vision system.

As each triangle may be observed in multiple images, the algorithm selects the best texture image for each triangle. A texture image is considered to be better if it is captured when the camera is facing the triangle directly. If the camera is looking at the triangle at an angle, then its quality is lower due to the lower and non-uniform resolution caused by perspective distortion. To find the best texture, the algorithm analyses all the images and selects the one that gives the largest area upon 2D projection according to the camera pose.

Moreover, we need to take into account any occlusion. For example, if there is an object in front of the triangle, then the image captured when the camera is facing the triangle directly should not be used, as the texture will be for the object in front. The algorithm then selects another texture image that gives the second largest area upon projection.

The texture images have slightly darker peripheral than the centre due to vignetting caused by the camera lens. Flat field correction [6] is applied to the images to reduce the visible seams across the texture images in the 3D models.

4. Experimental Results

Figure 3(a) shows the iSM hand-held stereo camera. We use a laptop PC with a Pentium IV 2.4GHz processor and 1GB RAM. As laptops only provide firewire interface without power, an external hub with power supply is required. For portability, a battery is used to supply power to the hub and camera; the laptop is carried by the user on a harness, as shown in Figure 3(b). This setup facilitates field tests in different environments.

Figure 4. ***Some input images from the house sequence.***

In one of the experiments, we modeled a facade of a house. The camera was moved freely pointing at different portions of the house and about 30 seconds of images were captured. Figure 4 shows some hand-held images from the house sequence. Then, the system processed these images automatically and created a photo-realistic 3D model in around 5 minutes.

The output 3D model is stored in the VRML (Virtual Reality Modeling Language) format and can be visualized using any of the 3D viewers widely available on the web. The user can navigate in the 3D model, and view it from any direction and distance.

Figure 5(a) shows the triangular mesh of the 3D model without the texture. Figure 5(b) shows the same view of the resulting 3D model with texture overlaid, while Figure 5(c) and (d) show two other views of the 3D model. We can see that iSM can reconstruct the overall 3D model by integrating all the input images, each of which captured with a limited field of view.

More advanced visualization and user interaction is provided in the iSM visualization GUI (Graphical User Interface). In addition to standard viewing, the user can perform

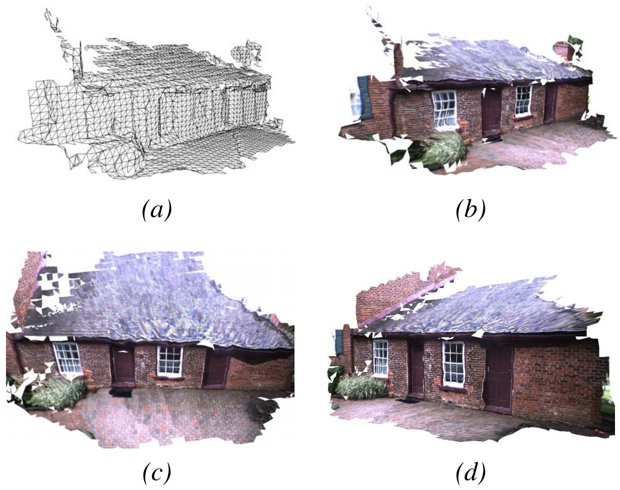


Figure 5. ***Different views of the house 3D model. (a) Wireframe 3D model without texture. (b) 3D model with texture for the same view as (a). (c) and (d) show two other views of the 3D model.***

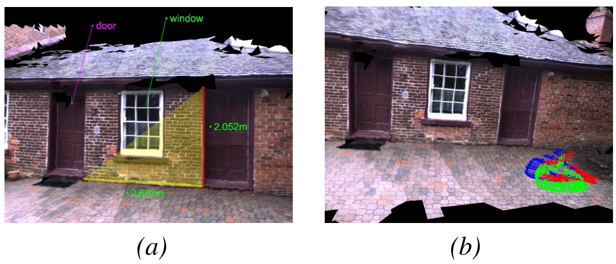


Figure 6. ***iSM visualization GUI. (a) Annotation and measurement on the 3D model. (b) The recovered camera trajectory.***

measurements (such as distance, angle, area) on the 3D model and can annotate the model, as shown in Figure 6(a). The camera trajectory recovered from the ego-motion estimation can be visualized, as shown in Figure 6(b). Moreover, the GUI has other features such as movie creation using trajectories defined with keyframes, keyframe interpolation and export of 3D models into DXF format.

Dense stereo depth resolution is given by:

$$\Delta Z = \frac{\Delta d}{fI} Z^2$$

The depth resolution decreases quadratically with the distance and improves as the baseline or the image resolution increases. Figure 7 shows the depth resolution of our Bumblebee stereo camera at various distances, with 12cm baseline at 640x480 image resolution and 0.1 pixel disparity resolution (Δd).

Experimental results show that for scenes within 3m from the camera, our 3D models have an accuracy within 2cm,

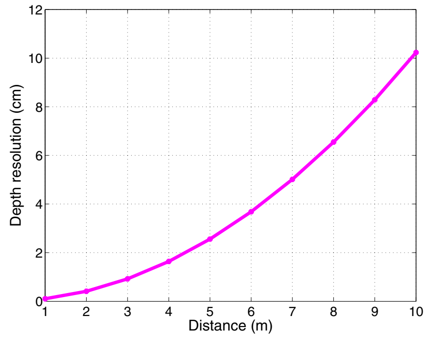


Figure 7. ***Stereo camera depth resolution at various distances.***

including dense stereo uncertainty and camera motion estimation error. Stereo camera with wider baseline and higher resolution can be used for larger environments to achieve better accuracy.

On the other hand, dense stereo could result in sparse depth data and hence poor coverage in the 3D model if the environment lacks features.

5. Forensic Applications

While the original motivation for this work is for planetary rovers, there are many terrestrial applications that can make use of photo-realistic 3D calibrated models. In this section, we present some experimental results from field tests carried out in the forensics domain and describe how 3D imaging is more advantageous than 2D imaging in this application.

Crime scene reconstruction is one of the most important aspect of police and forensic investigation. Currently, the police spends a lot of time at each crime scene taking many photos and performing manual measurements.

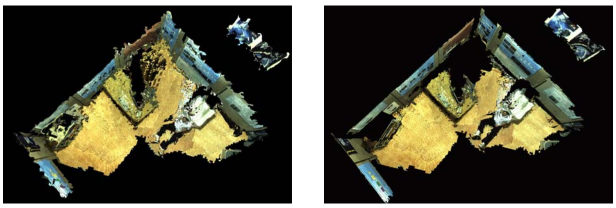
With iSM, they can create a 3D model of the crime scene quickly without much disturbance to the crime scene. Unlike traditional 2D imaging, measurements can be performed on the 3D model. As our 3D model is fully calibrated, there is no need to measure a reference object in the scene to scale the model. The police can also perform additional measurements they may have missed using the 3D model after the crime scene is released.

The 3D model can be shown to other officers who have not been to the crime scene. Apart from helping the police investigation, the 3D model can potentially be shown in court so that the judge and the jury can understand the crime scene better.

Figure 8 shows selected images obtained with the handheld camera at a mock crime scene set up in our lab, while the camera operator was moving around in the scene. The image sequence was captured approximately within 1 minute and the processing took around 10 minutes. Figure 9 shows the 3D model generated from the input sequence.

Figure 8. ***Some input images from the mock crime scene sequence.***

Figure 9. ***3D model of the mock crime scene.***



(a) *(b)*

Figure 10. ***(a) Bird's eye view of the original mock crime scene 3D model. (b) Mock crime scene 3D model after plane fitting.***

Movies for the input image sequence and the output 3D model are available at [10].

For indoor man-made environments, there are a lot of planar objects such as walls and floor. As there is uncertainty associated with our depth data obtained from dense stereo, these surfaces may not be exactly planar but appear bumpy. We have developed a plane finding routine using Hough Transform [9] to find the dominant planes in the 3D model automatically. Afterwards, plane fitting is applied to each dominant plane to replace the original surface, as a post-processing step. Figure 10 shows the bird's eye view of the 3D model before and after the plane fitting. This can be further optimized by replacing the planar regions with fewer triangles without degrading the model quality too much.

With the 3D models, the police can carry out ballistic analysis to investigate bullet trajectories. They can then hypothesize how the shooting took place. Moreover, blood splatter analysis information can be integrated with the system. The 3D models and these analysis can be used in various animation packages to help the police theorize how the crime happened.

Other forensic applications include traffic accident scene reconstruction to theorize how the accident happened. Our system can also be used for 3D site modeling and facility inspection. For example, a site can be scanned before and after an event to detect changes.

6. Conclusion

In this paper, we have presented a novel 3D modeling system, the instant Scene Modeler (iSM). iSM uses a hand-held stereo camera for recording images and a laptop for acquisition and processing. It creates photo-realistic 3D calibrated models of environments automatically (no user interaction) within minutes. Promising experimental results in various indoor and outdoor environments were shown. We have also discussed application and benefits of using iSM for forensic crime scene modeling.

From discussions with various police departments, while our 3D models are visually appealing and fast to create, the resolution and accuracy of the system need to be increased to meet the requirements of forensics applications. The current 3D models could be used to help understand the crime scenes and to perform some rough measurements only, while the precise measurements still need to be collected with distometer.

The quality and accuracy of the 3D model depend on the camera motion estimation and the dense stereo matching. Camera ego-motion accuracy is affected by the number of SIFT features in the environment and their distribution. Ego-motion error decreases when the number of SIFT features increases and when features are more widely spread out. However, long-term drifts in the motion estimation may occur for long sequences.

For future work, external sensors such as an orientation sensor can be incorporated to augment the visual motion estimation. Backward correction technique [20] can be applied to improve multiview registration and loop closure. Better dense stereo algorithms can be used to provide denser, more accurate depth data and thereby producing better 3D models.

Acknowledgements

We would like to thank Bovaird House and Andrew Roth for the field tests.

References

[1] 3rdTech. http://www.3rdtech.com.

[2] P. Besl and N. McKay. A method for registration of 3-d shapes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14(2):239–256, 1992.

[3] F. Blais. Review of 20 years of range sensor development. *Journal of Electronic Imaging*, 13(1):231–240, January 2004.

[4] Cyberware. http://www.cyberware.com.

[5] P.E. Debevec, C.J. Taylor, and J. Malik. Modeling and rendering architecture from photographs: A hybrid geometry- and image-based approach. In *Proceedings of SIGGRAPH'96*, New Orleans, Louisiana, August 1996.

[6] N. Goldstein. User's guide for geobox3. Technical Report RG-MA-51-0338, MDA, Richmond, B.C., Canada, 2003.

[7] C.J. Harris and M. Stephens. A combined corner and edge detector. In *Proceedings of 4th Alvey Vision Conference*, pages 147–151, Manchester, 1988.

[8] P. Hebert. A self-referenced hand-held range sensor. In *Proceedings of Third International Conference on 3D Digital Imaging and Modeling (3DIM 2001)*, pages 5–12, Quebec City, Canada, May 2001.

[9] P.V.C. Hough. Method and means of recognizing complex patterns, December 1962. U.S. Patent 306965418.

[10] Instant Scene Modeler (iSM). MDA Space Missions. http://www.mdrobotics.ca/ism_frame.html.

[11] D.G. Lowe. Object recognition from local scale-invariant features. In *Proceedings of the Seventh International Conference on Computer Vision (ICCV'99)*, pages 1150–1157, Kerkyra, Greece, September 1999.

[12] U. Neumann and Y. Cho. A self-tracking augmented reality system. In *Proceedings of ACM International Symposium on Virtual Reality and Applications*, pages 109–115, July 1996.

[13] D. Nister. Automatic passive recovery of 3d from images and video. In *Proceedings of Second International Symposium on 3D Data Processing, Visualization & Transmission (3DPVT04)*, Thessaloniki, Greece, September 2004.

[14] C.F. Olson, L.H. Matthies, M. Schoppers, and M.W. Maimone. Robust stereo ego-motion for long distance navigation. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Volume 2*, pages 453–458, South Carolina, June 2000.

[15] Optech. http://www.optech.ca.

[16] PhotoModeler. http://www.photomodeler.com.

[17] Polhemus. http://www.polhemus.com.

[18] M. Pollefeys, R. Koch, M. Vergauwen, and L. Van Gool. Hand-held acquisition of 3d models with a video camera. In *Proceedings of Second International Conference on 3D Digital Imaging and Modeling (3DIM'99)*, pages 14–23, Ottawa, Canada, October 1999.

[19] V. Popescu, E. Sacks, and G. Bahmutov. The modelcamera: a hand-held device for interactive modeling. In *Proceedings of Fourth International Conference on 3D Digital Imaging and Modeling (3DIM 2003)*, pages 285–292, Banff, Canada, October 2003.

[20] K. Pulli. Multiview registration for large data sets. In *Proceedings of Second International Conference on 3D Digital Imaging and Modeling (3DIM'99)*, Ottawa, Canada, October 1999.

[21] Point Grey Research. http://www.ptgrey.com.

[22] Riegl. http://www.riegl.com.

[23] G. Roth and E. Wibowo. An efficient volumetric method for building closed triangular meshes from 3-d image and point data. In *Proceedings of Graphics Interface (GI)*, pages 173–180, Kelowna, B.C., Canada, 1997.

[24] S. Rusinkiewicz, O. Hall-Holt, and M. Levoy. Real-time 3d model acquisition. In *Proceedings of SIGGRAPH'02*, San Antonio, Texas, July 2002.

[25] S. Rusinkiewicz and M. Levoy. Qsplat: A multiresolution point rendering system for large meshes. In *Proceedings of SIGGRAPH'00*, New Orleans, Louisiana, July 2000.

[26] S. Se, D. Lowe, and J. Little. Mobile robot localization and mapping with uncertainty using scale-invariant visual landmarks. *International Journal of Robotics Research*, 21(8):735–758, August 2002.

[27] S. Se, H. Ng, P. Jasiobedzki, and T. Moyung. Vision based modeling and localization for planetary exploration rovers. In *Proceedings of International Astronautical Congress (IAC)*, Vancouver, Canada, October 2004.

[28] M. Soucy, G. Godin, and M. Rioux. A texture-mapping approach for the compression of colored 3d triangulations. *The Visual Computer*, 12:503–514, 1996.

[29] Surphaser. http://www.surphaser.com.

[30] G. Turk and M. Levoy. Zippered polygon meshes from range images. In *Proceedings of SIGGRAPH'94*, pages 311–318, Orlando, Florida, July 1994.

[31] L. Van Gool, M. Pollefeys, M. Proesmans, and A. Zalesny. Modelling sagalassos: creation of a 3d archaeological virtual site. In A.K. Bowman and J.M. Brady, editors, *Images and Artefacts of the Ancient World*. Oxford University Press, 2005.