Journal of Applied Security Research, 4:79–100, 2009

ISSN: 1936-1610 print / 1936-1629 online
DOI: 10.1080/19361610802210244

Routledge
Taylor & Francis Group

Mobile Sensors for Security and Surveillance

ZHIGANG ZHU, PhD
City College of New York, New York, New York, USA

A stereo mosaic representation has been developed for fusing imaging data captured by sensors (cameras) in motion. In addition to providing a wide field of view, the multiperspective mosaics with various oblique views represent occlusion regions that cannot be achieved by using stationary sensors. One or multiple stereo pairs can be formed from mosaics with different oblique viewing angles and thus can be used for 3D viewing and 3D reconstruction. This approach has been applied to a number of important security and surveillance applications, including airborne surveillance, ground vehicle navigation, under-vehicle inspection, and 3D gamma-ray cargo inspection.

KEYWORDS Video surveillance, security inspection, 3D reconstruction, video registration, push-broom imaging

INTRODUCTION

When something serious happens in a metropolitan area like New York City, imagine a technology that can fly an airplane with a video camera through the area, detect, measure, and analyze the static and dynamic objects in the area, and then reconstruct the scene into multiple three-dimensional (3D)

This work is supported by National Science Foundation (Award No. CNS-0551598), AFOSR (Award No. FA9550-08-1-0199), Army Research Office (Award No. W911NF-05-1-0011), and by funding from New York Institute for Advanced Studies, from PSC-CUNY and from Atlantic Coast Technologies, Inc. I would like to thank Dr. Rex Richardson and Dr. Victor J. Orphan at Science Applications International Corporation (SAIC) for providing gamma-ray images and the dimension data of the cargo container. The author also gives thanks to collaborators and students who have been involved in the related projects at both the City College of New York and the University of Massachusetts at Amherst.

Address correspondence to Zhigang Zhu, Department of Computer Science, City College of New York, 130th Street and Convent Avenue, New York NY 10031, USA. E-mail: zhu@cs.ccny.cuny.edu

panoramic views. This technology is being developed in the City College Visual Computing Laboratory.

Potential applications of such a technology can be significant. Currently, traffic monitoring and video surveillance is performed largely by utilizing overhead stationary cameras that are mounted at various locations. The images taken from these sources are unprocessed and sent to a control center for computer processing and/or human interpretation. Limited by the viewpoint constraints of those cameras, the images are only available for a few localized detection areas. But if the images are taken from an airborne camera, a large field of view can be covered. Furthermore, if the video data are processed such that information about the 3D static and dynamic objects in the area can be automatically detected, measured, and analyzed, it will significantly improve the capability of traffic management and video surveillance in assessing the traffic condition or activities at the scene, and then developing real-time solutions. The processing of these images can also potentially improve the visualization by presenting multiple 3D panoramic views. In fact, sensors in motion are not only found in airborne surveillance, but also in ground survey and inspection; such as driving a vehicle down the street with a video camera, scanning the under-body of a vehicle using cameras, or screening the interiors of a cargo container using X-ray or gamma-ray techniques.

However, monitoring and detection of scenes and targets through moving sensors increase the challenges in data analysis and representations. With a stationary camera, much simpler algorithms can be applied by assuming the background remains unchanged. With a moving camera, however, everything in a 3D scene is in motion due to ego-motion of the sensor. The work at the City College Visual Computing Lab in the last few years has tried to address this challenging issue. This article is a high-level summary of the novel mosaic-based approach we have developed for fusing imaging data from one or more moving cameras/sensors into a few mosaiced images, which preserve 3D information and generate a wide coverage of a scene (or an object). The article will focus on the basic concept and potential applications in security and surveillance; details of the technologies can be found in our previous technical publications that will be cited through out this article.

The proposed mosaic approach has been applied to a variety of applications, including airborne video for urban transportation planning and urban surveillance, ground mobile robot navigation, under-vehicle inspection, and gamma-ray cargo inspection (Figure 1). These applications represent very different imaging scenarios, from far-range to extreme close-range, from a single camera to an array of cameras, from visible imaging to see-through imaging. I will show that the same geometric representation can be applied to all these cases and that this representation has advantages in data compression, depth resolution, field of view and 3D perception.

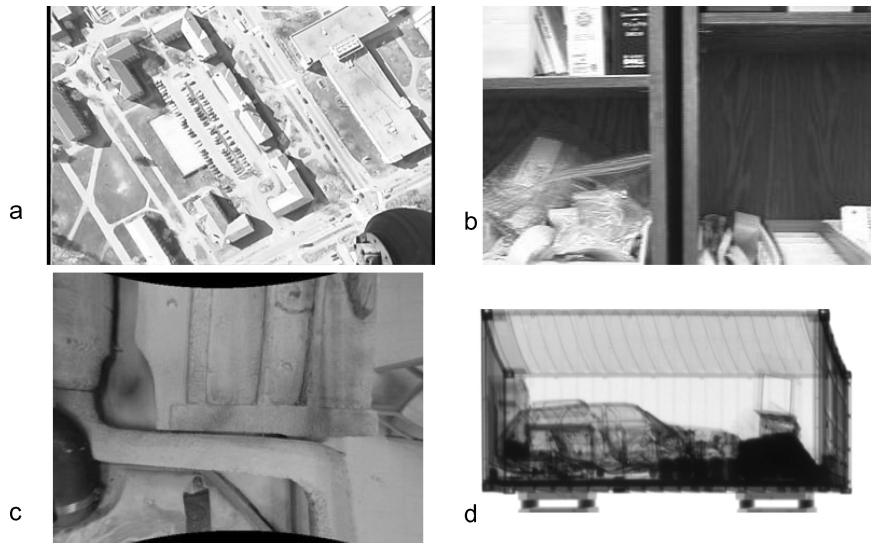

FIGURE 1 A few application examples: (a) airborne urban surveillance/transportation planning; (b) ground mobile robot; (c) under-vehicle inspection; and (d) gamma-ray cargo inspection.

STEREO VISION WITH PARALLEL PROJECTION

A normal perspective camera has a single viewpoint (i.e., nodal point), which means all the light rays pass through the common nodal point. On the other hand, in orthogonal images with parallel projections in both the x and y directions, all the rays are parallel to each other. Imagining that we have a sensor with parallel projections, we could turn the sensor to capture images, each with a different *oblique* viewing direction, including both nadir and oblique angles in both the x and y directions. Here "nadir" means that the angles in both the x and y directions are zeroes. Thus we can create multiple pairs of parallel stereo images each with a pair of different oblique viewing directions, and thus can observe surfaces occluded in a nadir view.

Figure 2 shows the parallel stereo in a 1D case, where two oblique angles β_1 and β_2 are chosen. The depth (Z) of a point P can be calculated as

$$Z = \frac{B}{\tan\beta_2 - \tan\beta_1} \tag{1}$$

where β_1 and β_2 are the angles of the two viewing directions, respectively, and B is the *adaptive baseline* between the two viewpoints. This adaptive baseline information is embedded in a pair of stereo mosaics with these two angles, and is proportional to the displacement of the corresponding image projections of the point P. The baseline is adaptive because, given

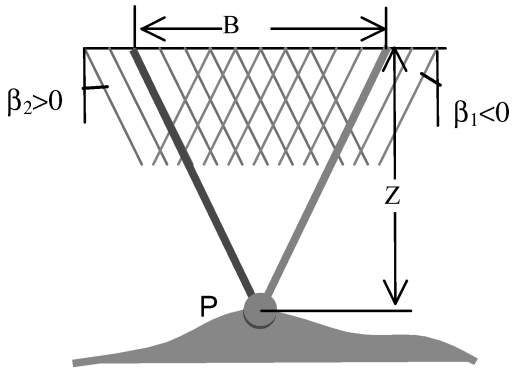

FIGURE 2 Depth from parallel stereo with multiple viewpoints: 1D case.

two angles, a point with larger depth will have a larger baseline than a point with smaller depth. It has been shown by others (Chai & Shum, 2000) and by us (Zhu et al., 2003; 2004) that parallel stereo is superior to both conventional perspective stereo and to the recently developed multiperspective stereo with concentric mosaics for 3D reconstruction (e.g., Shum & Szeliski, 1999; Peleg et al., 2001). The adaptive baseline inherent in the parallel projection geometry permits depth accuracy independent of absolute depth in theory. This result can be easily obtained from Equation 1 since depth Z is proportional to the adaptive baseline B and therefore to the recorded visual displacement of the corresponding pair in the two mosaics. In contrast, the depth error of perspective stereo and concentric stereo is proportional to the square of depth.

We can make two extensions to this 1D case of parallel stereo. First, we can select various oblique angles (instead of just two) for constructing multiple parallel projections. By doing so we can observe various degrees of occlusions and can construct stereo pairs with different depth resolution via the selection of different pairs of oblique angles.

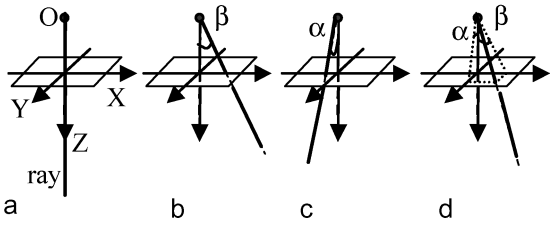

FIGURE 3 Parallel projections with two oblique angles α and β (around the x and y axes, respectively). (a) Nadir view ($\alpha = \beta = 0$); (b) β-oblique view ($\alpha = 0$, $\beta \neq 0$); (c) α-oblique view ($\alpha \neq 0$, $\beta = 0$) and (d) dual-oblique view ($\alpha \neq 0$, $\beta \neq 0$). Parallel mosaics can be formed by populating each single selected ray in both the x and y directions.

Second, the 1D parallel projection can be extended to 2D (Figure 3), with two oblique angles α and β around the x and y axes, respectively, thus obtaining a mosaiced image that has a nadir view (Figure 3a), oblique angle(s) only in one direction (Figure 3b and Figure 3c) or oblique angles in both the x and the y directions (Figure 3d).

PRACTICAL SCENARIOS AND RESEARCH ISSUES

Practical Setups

It is impractical to use a single (stationary) sensor to capture orthogonal images with full parallel projections with various oblique directions when imaging/covering a large-scale scene. However, in practice, parallel-perspective panoramic images, with parallel projection in one direction and perspective in the other, can be generated the same way as pushbroom images in satellite imaging (Gupta & Hartley, 1997), by using a 1D perspective sensor moving in the perpendicular direction of the 1D sensor array. Two such 1D sensors with two different oblique viewing angles consists of a pushbroom stereo imaging system. A real example of this geometry is gamma-ray stereo imaging for 3D cargo inspection (Zhu et al., 2005a; Zhu & Hu, 2007).

In fact, we can move one or more conventional 2D perspective cameras to form a 1D or 2D "virtual" array of cameras, to generate parallel-perspective (pushbroom) stereo or full parallel stereo. In principle, if we first assume that the optical axes of all the cameras point in the same direction (into the paper in Figure 4a), and the viewpoints of all cameras are on a single plane perpendicular to their optical axes. Then the perspective images can be organized into mosaiced images with parallel projections, each of which is generated with an oblique viewing angle, by extracting rays from the original perspective images with the same viewing direction (one ray from each image). For example, extracting a ray as shown in Figure 3a with the nadir viewing direction from each image at each camera

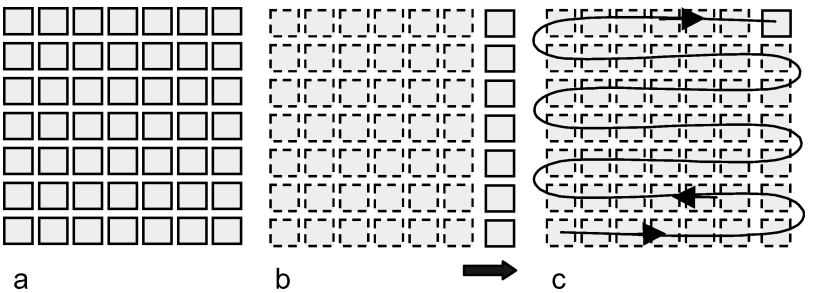

FIGURE 4 Parallel mosaics from 2D bed of cameras. (a) 2D array; (b) 1D scan array; and (c) a single scan camera.

location (in the setup of Figure 4a) will generate a parallel mosaic with nadir viewing direction. Extracting a ray as shown in Figure 3b with a β-oblique viewing direction from each image at each camera location (in the setup of Figure 4a) will generate a parallel mosaic with the β-oblique viewing direction. If the camera array is dense enough, then densely mosaiced images can be generated.

There are at least two practical ways of generating images with multiple (stereo) oblique parallel projections using existing sensors: a 1D scan of a 1D array of perspective cameras (Figure 4b), a 1D or 2D scan of a single perspective 2D-array camera (Figure 4c).

If a 1D linear array of perspective cameras is available (Figure 4b), the camera array can be "scanned" over a scene to synthesize a virtual 2D camera array. Then stereo mosaic pairs with oblique parallel projections in both directions can still be generated, given that we can accurately control or estimate the translation of the camera array. We have actually used this approach in an Under Vehicle Inspection System (UVIS)[1,2] (Dickson et al., 2002).

Even when a single camera is used, we can still generate a 2D virtual bed of cameras by moving the camera in two dimensions, along a "2D scan" path as shown in Figure 4c. This is the case for aerial video mosaics[3] (Zhu et al., 2003, 2004), where a single camera is mounted on a light aircraft flying over an area.

Issues in Video Mosaics

In real applications where parallel-projection mosaics must be generated from a video sequence (as in Figure 5), there are two challenging research issues. The first problem is camera orientation estimation (calibration). In our previous study on an aerial video application, we used external orientation instruments (i.e., GPS, INS, and a laser profiler) to ease the problem of camera orientation estimation (Zhu et al., 2005b). In the case of under-vehicle inspection using a 1D array of cameras (Dickson et al., 2002), relative relations among cameras can be obtained by an offline camera calibration procedure. However, the motion of the cameras or vehicles should be estimated through image matching. Fortunately, there exists a large body of work in pose estimation of a moving camera using bundle adjustments in the fields of computer vision and photogrammetry (e.g., Slama, 1980; Triggs et al., 2000), and even software packages, such as REALVIZ's Matchmover[4] and sba (Lourakis & Argyros, 2004), are available for this purpose. Even if the camera's motion has six degrees of freedom, as long as it has a dominant motion direction, seamless mosaics can be generated. The accuracy of pose estimation is the main issue in bundle adjustments, and is very important in producing accurate 3D reconstructions using the stereo mosaics thus generated. However, in applications of 3D rendering where accurate 3D estimation is not the main issue, an efficient image-based camera-motion

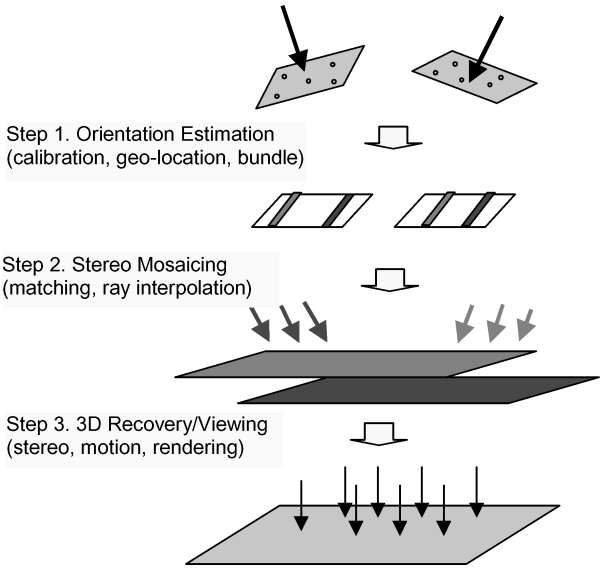

FIGURE 5 Three step approach for generating and using parallel-projection mosaics.

estimation method (Zhu et al., 2004) is used to get an approximation of the camera orientation parameters, that is, the affine transformation parameters, and then seamless mosaics can be generated with 3D perception.

The second problem is to generate dense parallel mosaics with a sparse, uneven, camera array, and for a complicated 3D scene. To solve this problem, a Parallel Ray Interpolation for Stereo Mosaics (PRISM) approach was proposed in Zhu et al. (2004). While the PRISM algorithm was originally designed to generate parallel-perspective stereo mosaics (parallel projection in one direction and perspective projection in the other), the core idea of *ray interpolation* can be used for generating a mosaic with full parallel projection at any oblique angle.

In summary, in the stereo mosaic approach for large-scale 3D scene modeling and rendering, the computation is efficiently distributed in three steps (Figure 5): (1) camera pose estimation via the external measurement units, (2) image mosaicing via ray interpolation, and (3) 3D reconstruction from a pair of stereo mosaics (Zhu et al., 2005c; Tang et al., 2006), or 3D rendering with multiview mosaics (Zhu & Hanson, 2006). In estimating camera poses (for image rectification), only sparse tie points, widely distributed, in the two images are needed for performing bundle adjustment. In generating dense parallel rays in stereo mosaics, local matches are only performed for parallel-perspective rays between small overlapping regions of successive frames. In using stereo mosaics for 3D recovery, matches are only carried out between the two final mosaics; for 3D viewing, only mosaic selection and viewing window cropping are needed. We will get into some more details in real examples provided in the next few sections.

The proposed mosaic representation has been applied to a variety of applications, including (1) airborne video for environmental monitoring and urban surveillance; (2) ground mobile robot navigation; (3) under-vehicle inspection; and (4) gamma-ray cargo inspection. These applications represent very different imaging scenarios, including far-range, medium-range, and extremely close-range imaging, from visible sensing to gamma-ray sensing. I will show that the same underlining principles (in terms of both projection geometry and stereo relations) can be applied to all four cases.

VIDEO MOSAICS FROM AERIAL VIDEO

In theory, with a camera on an airplane undergoing an ideal 1D translation and with a nadir view direction, two spatio-temporal images can be generated by extracting the two rows of pixels at the front and rear edges (slits of each frame) perpendicular to the direction of motion (Figure 6). The mosaiced images thus generated are *parallel-perspective*, with parallel projection in the direction of motion and perspective projection in the other. In addition, these mosaics are obtained from two different oblique viewing angles of a single camera's field of view, so that a stereo pair of left and right mosaics captures the inherent 3D information.

3D Mosaic Construction from Reality

We have proposed and developed a content-based 3D mosaic representation (CB3M) for long video sequences of 3D and dynamic scenes, captured by a camera mounted on an aerial mobile platform. The motion of the camera has a dominant direction of motion (as on an airplane), but six

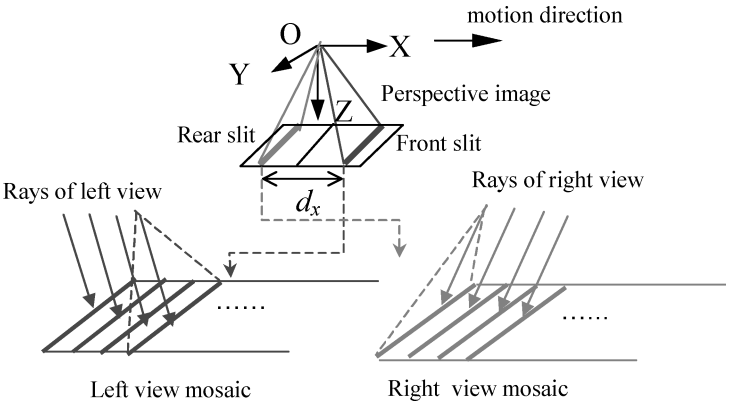

FIGURE 6 Parallel-perspective (pushbroom) stereo mosaics with a 1D camera scan path.

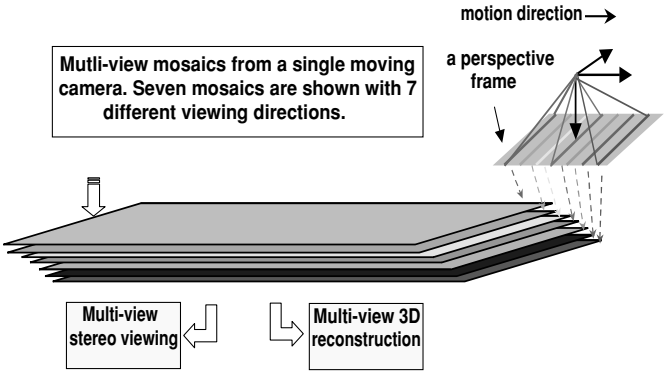

FIGURE 7 Mosaics: from many narrow FOV images to a few wide FOV mosaics.

degrees-of-freedom (DOF) motion is allowed (Zhu et al., 2004). There are two consecutive steps in constructing a CB3M representation from a video sequence on a mobile platform: stereo mosaicing and 3D/motion extraction.

In the first step, a set of parallel-perspective (pushbroom) mosaics (Zhu et al., 2004; Zhu et al., 2005c)—panoramic images combining all the video images from different viewpoints—is generated to capture both the 3D and dynamic aspects of the scene under the camera coverage. If real-time data of camera positions and orientations are available from a Geographical Position System (GPS) and an inertial navigation system (INS), the panoramic mosaics can be geo-located to the world coordinate system. This step turns thousands of images of a video sequence into a few large field-of-view (FOV) mosaics that have the same coverage as the original video sequence. Multiple wide FOV mosaics are generated from a single camera on a single flight, but the results are similar to those using multiple scan line cameras, or pushbroom cameras (Gupta & Hartley, 1997), with different oblique angles to scan through the entire scene (Figure 7). Because of the various angles of the scanning, occluded regions in one mosaic can be seen from the others. All moving objects appear in each mosaic, and by switching to different ones the dynamic aspects can also be viewed. This corresponds to the multiview stereo viewing shown in Figure 7.

However, the 2D mosaic representation is still a 2D array of image points, lacking the representation of object contents, such as buildings, roads, and vehicles and other facilities. Therefore, in the second step, a segmentation-based stereo matching algorithm (Zhu et al., 2005c; Tang et al., 2006) is applied to extract parametric representations of the color, structure and motion of the dynamic and/or 3D objects in an urban scene, and to create a content-based 3D mosaic (CB3M) representation (Zhu & Tang, 2006). CB3M is a highly compressed visual representation for very long video sequences of dynamic 3D scenes. In the CB3M representation, the panoramic mosaics are segmented into planar regions, which are the primitives for

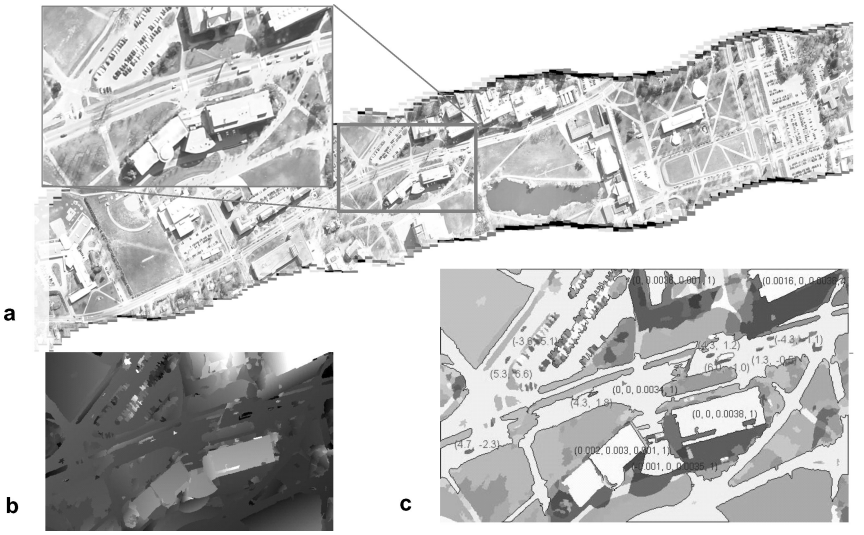

FIGURE 8 Content-based 3D mosaic representation of an aerial video sequence: (a) a pair of stereo mosaics from the total nine mosaics and a close-up window; (b) the height map of the objects inside that window; (c) the CB3M representation with some of the regions labeled by their boundaries and plane parameters (a, b, c, d), and the detected moving targets marked by their boundaries and motion vectors (s_x, s_y).

content representation. Each region is represented by its mean color, region boundary (i.e., object contour), plane equation in 3D space, in the form of

$$aX + bY + cZ = d \tag{2}$$

from which its orientation and height can be derived, and motion direction and speed in the form of a 2D motion vector of speed (s_x, s_y), if it is a dynamic object. Relations of each region with its neighbors are also built for further object representations (such as buildings, road networks) and target recognition. This second step is depicted in Figure 7 as "multi-view 3D reconstruction."

Figure 8 shows an example of CB3M from a real video sequence when the airplane was about 300 meters above the ground. Figure 8a shows a pair of stereo mosaics (embedded in the red/green-blue channels of a color picture if viewed in the online color version) that are used to extract 3D information—similar to the stereo vision of humans, but with an extended field of view (FOV). A close-up window is marked in the stereo mosaics, which includes various 3D structures and moving objects (vehicles). Figure 8b is a "height" map of the scene in the close-up window generated using the proposed method; the brighter the pixel is, the higher the object is. Note that the sharp depth boundaries are obtained for the buildings with different heights and varying roof shapes.

The moving objects are shown by their contours and velocities (s_x, s_y). The CB3M representation (of the small portion in 8a) is shown in Figure 8c, with the mean color, the object contour, plane parameters (a, b, c, d), and a motion vector (if in motion) for each region. For this example, the original image sequence has 1000 frames of 640 * 480 color images. With its CB3M representation, a compression ratio of more than 10,000:1 is achieved. The high compression ratio is important when multiple large field-of-view mosaics need to be created and stored.

3D Mosaic Visualization

In the previous section, we mainly discussed how to generate internal computerized models of urban traffic scenes, that is, the CB3M representations, from video sequences. In this section, we will describe our algorithms for visualizing scenes using the Content-Based 3D Mosaic (CB3M) representation. Note that the modeling and the rendering of the CB3M representations, which are usually called image-based modeling and rendering, are two different aspects of the problem: (1) video analysis (from raw video data to computerized models) using computer vision techniques, (2) and video resynthesis (from computerized models to visual presentations to humans) using computer graphics and visualization techniques.

Mosaics with various oblique angles represent scenes from the corresponding viewing angles (Figure 7). Visualization algorithms that we are investigating correspond to two levels of utilizations of such representations: 2D-mosaic-based visualization, and CB3M-based visualization. First, a human can perceive a 3D scene from a pair of mosaics with different oblique angles (e.g., using polarized glasses, or even with a pair of red/blue glasses) without any explicit 3D recovery. This feature leads to a very efficient multiview stereo viewing approach (Zhu & Hanson, 2006): the translation and rotation of the virtual camera for a virtual fly-through are implemented by simply sliding a window across the mosaic(s) and switching between different mosaic pairs. This level of visualization gives users a quick and compelling way to control the viewing of dynamic 3D scenes with a large field of view. A mosaic-based fly-through demo may be found at the author's website,[5] which uses nine oblique mosaics generated from a real video sequence of the UMass campus. This result shows *motion parallax, occlusion, and also moving objects* in multiple parallel-perspective mosaics.

However, the 3D and dynamic contents are still interpreted by the users rather than an automated computer vision system. Therefore, in the second level, we will study how to produce real-time visualizations of an urban, dynamic scene using its content-based 3D mosaic (CB3M) representation. In the CB3M representation, each patch (region) is represented by its boundary (in a 2D mosaic), its color, its plane parameters, and its motion parameters (if in motion). Therefore, the 3D and dynamic aspects can be rendered

into user-selected views to better reconstruct the scene. We will also study the augmented reality techniques to show virtual objects—such as region boundaries, plane parameters, region classifications, vehicle motion information, and traffic statistics. Figure 8c shows some preliminary results of this idea. With the CB3M representation, users' choices of "viewing modes" for visualization could be bird-eye *panoramic views* of the entire scene ("global views"), *3D views* of particular areas from various viewing directions ("local views"), and global and local views with *annotated* facility and traffic information ("annotated views"). The users can also compare views of different physical sites and/or viewing modes.

Advantages of 3D Mosaics

In conclusion, the CB3M representation has the potential to provide the following means for security, surveillance and transportation applications.

1. The entire image sequence of a scene from a fly-through is transformed in real time into a few large FOV **panoramic mosaics**. This provides a synopsis of the scene with all the static and dynamic objects in a single view (Figure 8a). A simple graphic user interface (GUI) can also be developed to perform a virtual fly-through of the area, with controls of both viewing locations and orientations of the scene.
2. The **3D contents** of the CB3M representation provide three-dimensional measurements of objects in the scene. In Figure 8c, the boundary, the color and the planar parameters (*a, b, c, d*) of each region in the form of $aX + bY + cZ = d$ are shown in blue. Because each object (e.g., a building) has been represented into 3D planar regions and their relations, further object recognition and higher level feature extraction are made possible. In other words, recognition of the characteristics of the buildings (e.g., shapes, numbers of doors and windows) and the road network (e.g., roads and vehicles) is made possible.
3. The **motion contents** of the CB3M representation provide dynamic measurements of moving targets in the scene. In Figure 8c, the motion parameters (s_x, s_y) of each moving object (vehicle) are marked in red, representing the translational speed (in 2D) of the object. This not only provide information about the vehicle's direction and speed, but also the traffic situation of a road segment, because each road "region" can also be extracted based on its 3D information and shape.
4. The CB3M representation is **highly compressed**. Usually a compression ratio of thousands to ten thousands can be achieved. This saves space when a lot of data for a large area needs to be archived.

This approach is different from the traditional 3D modeling methods from satellite images, which are usually very time-consuming. Mosaic

approaches (Irani et al., 1996; Leung & Chen, 2000)—the creation of panoramic images—have been proposed for representation and compression of video sequences, but most of the work is with panning (i.e., rotating) cameras instead of moving (i.e., translating) cameras as in the cases of airborne traffic monitoring. Some work has been done in 3D reconstruction of panoramic mosaics (Li et al., 2004; Sun & Peleg, 2004) but usually the results are 3D depth maps of static scenes instead of high-level 3D representations for both static and dynamic targets that can be readily used in surveillance and transportation applications.

NYC: Another Example

We have also generated mosaics from a video sequence of a NYC HD (high-definition) aerial video dataset (vol. 2) we ordered online.[6] The video clip, NYC125H2, has about 25 seconds, or 758 frames of high-definition progressive video (1920 * 1080). Rooftops and city streets are seen as the camera looks ahead and down in a close flight just over One Penn Plaza and beyond in New York City. Yellow taxicabs make up a noticeable percentage of the vehicles traveling the grid of streets in this district of mostly lower-rising buildings, with a few scattered high-rise buildings. You may view the low-resolution version of the video following the link we have provided earlier. Our main task is to recover the full 3D model of the area automatically; an area with cluttered buildings of various heights, from less than ten to more than a hundred meters in height. Figure 9 shows one of the four multiview mosaics generated and used for 3D reconstruction and moving

FIGURE 9 A 4816 (W) × 2016 (H) mosaic from a 758-frame high-resolution NYC video sequence.

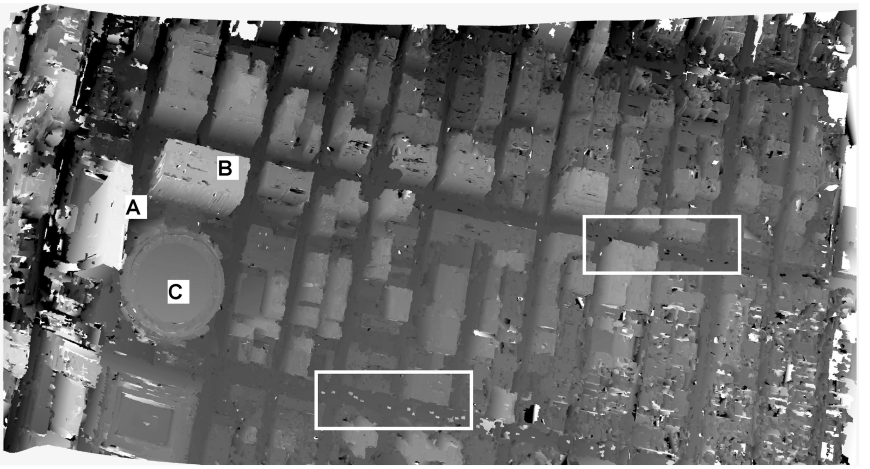

FIGURE 10 Height map from multi-view (4) mosaics.

target detection. The mosaic that is shown here has been turned 90 degrees, therefore the camera moves in the direction from the left to the right in the mosaic. The size of the mosaic is 4816 (W) × 2016 (H).

Figure 10 shows the 3D reconstruction results of the NYC video data, represented in the leftmost mosaic—the reference mosaic. The figure shows the height map generated from multiview mosaics. Due to the lack of flight and camera parameters, we roughly estimate the main parameters of the camera (i.e., the height of the flight and the camera's focal length) from some known buildings. However, this gives us a good indication of how well we can obtain the 3D structure of this very complex scene. For example, the average heights of the three buildings at One Penn Plaza (marked as A, B, and C in Figure 16c) are 105.32 m, 48.83 m, and 19.93 m, respectively. Our approach handles scenes with dramatically varying depths. Readers may visually check the heights of those buildings with GoogleEarth.

The moving objects (vehicles) create "outliers" in the height map, as can be clearly seen on the height map (the brighter the color is, the higher the object is). For example, on the one-way road indicated in the first window in Figure 9, vehicles moved from the right to the left in the figure, therefore, their estimated heights are much higher than the ground if assumed static. On the other hand, on the one-way road indicated in the second window in Figure 9, vehicles moved from the left to the right in the figure, therefore, their estimated heights are much lower than the ground if assumed static.

After further applying the knowledge of road directions that are obtained from a dominant plane clustering procedure, moving targets are searched and extracted. In Figure 11, all of the *moving* targets (vehicles) are extracted, except the three circled in the figure. These three vehicles are merged with the road in color segmentation. Other vehicles that are not detected were

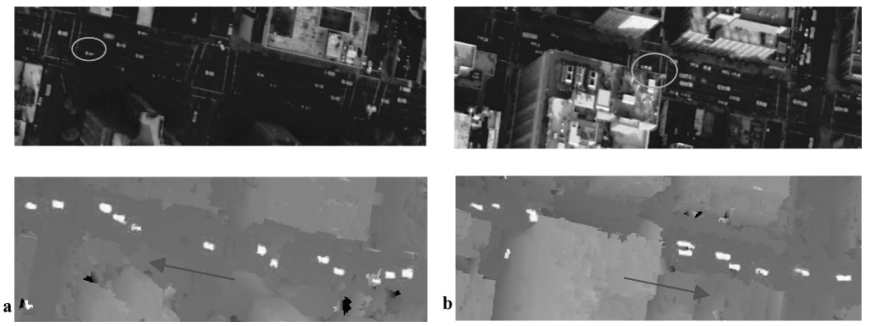

FIGURE 11 Moving target detection using the road direction constraint. In the figure (a) and (b) are the corresponding color images and height maps of the 1st (down-left) and 2nd (up-right) windows in Figure 15, with the detected moving targets painted in white. The two circles show the three moving targets that are not detected. The arrows indicate the directions of the roads along which the moving targets are searched.

stationary; most of them are on the orthogonal roads with red traffic signals on for stop, and a few were parked on these two one-way roads.

VIDEO MOSAICS FOR MOBILE ROBOT APPLICATION

The same approach has also been applied to ground mobile robot applications where the range of the roadside scenes to the camera on a mobile robot is from tens of feet (indoor) to hundreds of feet (outdoor). The road-side parallel-perspective stereo mosaics can be used for human–robot interaction and landmark localization in robot navigation. Figure 12 shows three parallel-perspective mosaics from a 517-frame video sequence captured from a mobile robot viewing a group of bookshelves and cabinets at close range as the robot moves from one end to the other in a laboratory.

For this example, 11 mosaics are generated. A panoramic depth map was constructed from the mosaics (Figure 12c). The video clip of a virtual walk-through using these 11 mosaics can be found at the author's website.[7] Figure 13 shows a few snapshots extracted from the "rendered" video clip at two camera locations, one viewing the connection between two book shelves (the 1st row), and the other viewing one end of a cabinet (the 2nd row). In this example, the 3D and occlusion effects are dramatic.

VIDEO MOSAICS FOR UNDER-VEHICLE INSPECTION

As one of the real applications of full parallel stereo mosaics, an approximate version of mosaics with full parallel projections has been generated from a virtual bed of 2D camera arrays by driving a car over a 1D array of cameras in

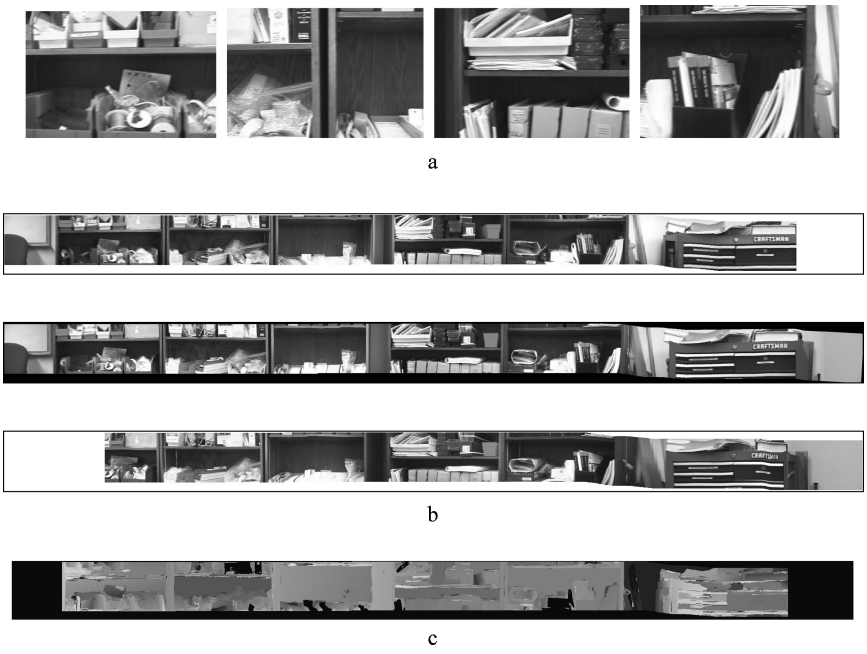

FIGURE 12 Ground video application. (a) A few frames from a 517-frame sequence of image size 320*240. (b) Ground video mosaics: a left view, the center view, and a right view are shown. Each mosaic is 4160 * 288. (c) The depth map that can be used for robot navigation and target detection.

an under-vehicle inspection system (UVIS)[1] (Dickson et al., 2002). UVIS is a system designed for security checkpoints such as those at borders, embassies, large sporting events, and so on. It is an example of generating mosaics from very short-range video; a 2D virtual array of camera is necessary for full coverage of the vehicle undercarriage.

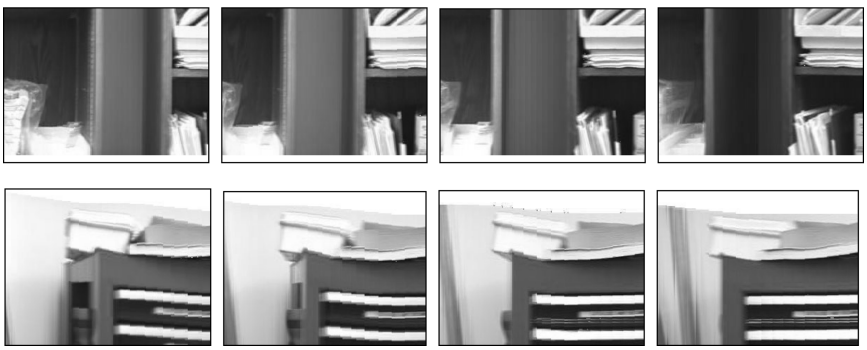

FIGURE 13 Mosaic-based walk-through: stereoscopic snapshots (in the online color version, 3D effect may be viewed with red/blue glasses).

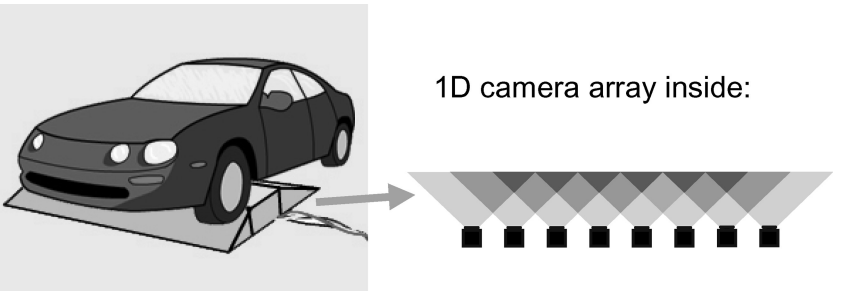

FIGURE 14 Conceptual 1D camera array for under-vehicle inspection.

Figure 14 illustrates the system setup where an array of cameras is housed in a platform. When a car drives over the platform, several mosaics with different oblique angles of the underside of a car are created. The mosaics can then be viewed by an inspector to thoroughly examine the underside of the vehicle from different angles. Figure 15 shows such a mosaic covering the full under-body of a vehicle, generated from a 1D array of 13 cameras spaced 3 inches apart traveling down the length of the vehicle taking pictures every 3 inches. This is equivalent to a stationary 1D array of cameras and a moving vehicle. The 1D array of 13 cameras are simulated by laterally shifting the real experimental set-up of 4 side-by-side cameras spaced 3 inch apart.

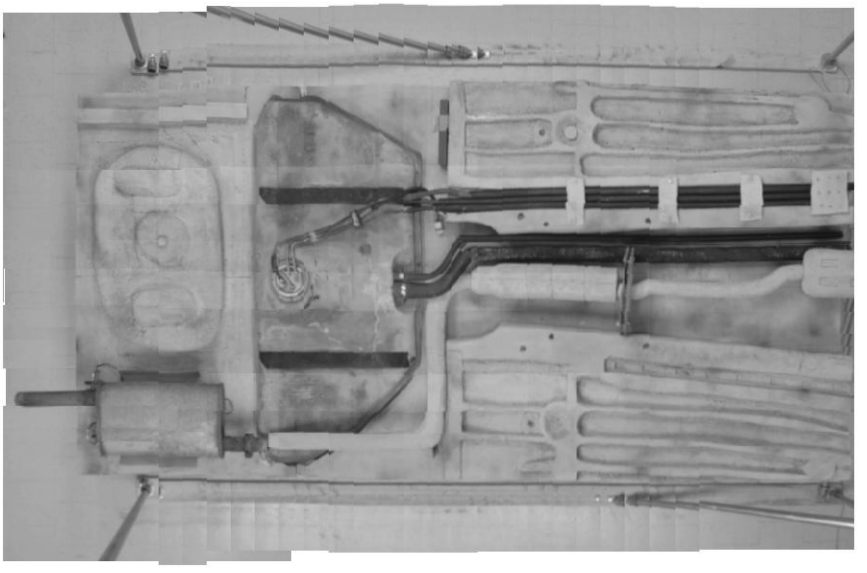

FIGURE 15 2D parallel mosaic from "13 cameras" spaced 3 inches apart traveling down the length of the vehicle.

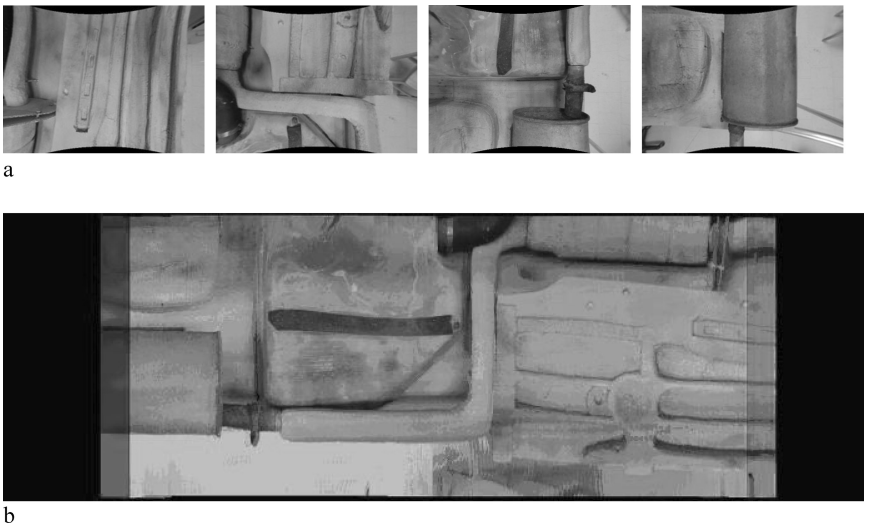

FIGURE 16 Under-vehicle inspection: (a) four frames from a 130-frame video sequence with image size 611 × 447; (b) one of the stereo mosaic pair (in the online color version, 3D effect may be viewed with red/blue glasses).

Figure 16 shows a pair in red-blue channels of the five mosaics, each with different oblique views, generated from a 130-frame video sequence (sample video frames are shown in Figure 16a). Different "occluded" regions under a pipe in the center can be observed by switching to different mosaics used in mosaic-based rendering (Zhu & Hanson, 2006). A PPT demo of these five oblique parallel views of the mosaics can be found at the author's website.[2] More results on 2D parallel-projection mosaics can be found at the UMass UVIS site.[1]

In the case of the 1D camera array, the fixed cameras were pre-calibrated and the geometric and photometric distortions of these wide FOV cameras were corrected. However, challenges remain because (1) the distance between cameras are large compared to the very short viewing distances to the bottom of the car; and (2) without the assistance of GPS/INS for pose estimation, we need to determine the car's motion by other means, for example, tracking line features on the car.

GAMMA-RAY PUSHBROOM STEREO FOR CARGO INSPECTION

The system diagram of the gamma-ray cargo inspection system (Orphan et al., 2002) is shown in Figure 17. A 1D detector array of 256 NaI-PMT probes, counts the gamma-ray photons passing through the vehicle/cargo under inspection from a gamma-ray point source. Either the vehicle/cargo or the gamma-ray system (the source and the detector) moves in a straight line in order to obtain a 2D scan of gamma-ray images.

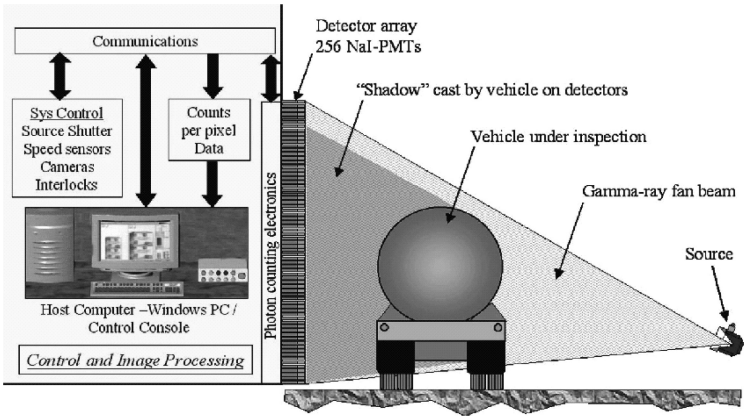

FIGURE 17 Linear pushbroom sensor model of a gamma-ray cargo inspection system (courtesy SAIC, San Diego, CA, USA).

A dual-scanning system is a *linear pushbroom stereovision system*. It can be constructed with two approaches: two linear pushbroom scanning sensors with different scanning angles, or a single scanning sensor to scan the same cargo twice with two different scanning directions. The first approach can be used to detect moving targets inside a cargo container. Figure 18 shows two real gamma-ray images, with different scanning angles—10 and 20 degrees, respectively. Each image has a size of 621 × 256 pixels, that is, 621 scans of the 256-pixel linear images.

We have proposed a practical approach for 3D measurements in gamma-ray (or X-ray) cargo inspection (Zhu et al., 2005; Zhu & Hu, 2007). Two

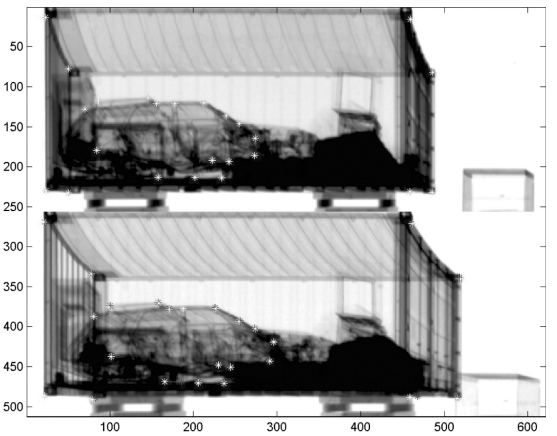

FIGURE 18 Real gamma-ray images from two different scanning angles—ten and twenty degrees (original images courtesy SAIC, San Diego, CA, USA). Each image has a size of 621 × 256 pixels, that is, 621 scans of the 256-pixel linear images. This figure also shows the matching of two sets of points in the two images in white asterisks.

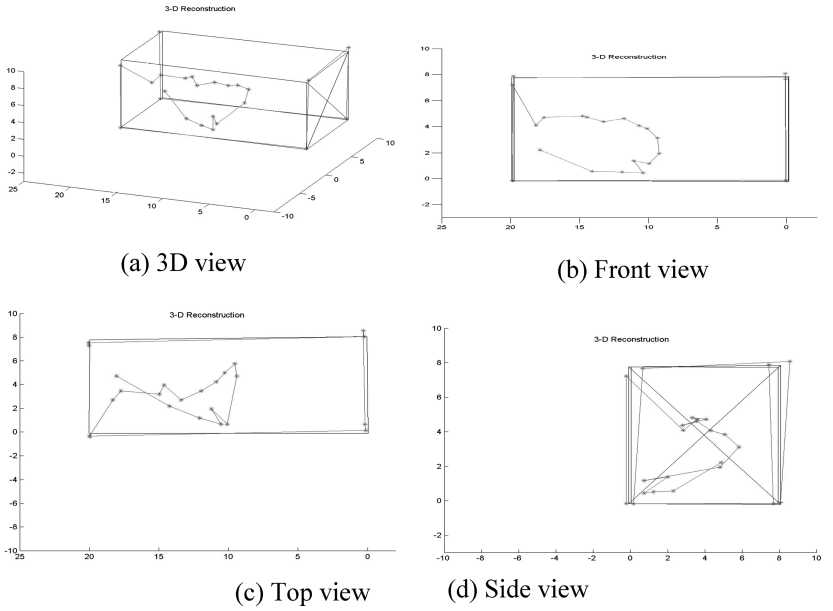

FIGURE 19 3D measurements and visualization of objects inside the cargo container. The darker rectangular frames show the "ground truth" data of the cargo container. The lighter lines (red in the online version, with asterisks) show the 3D estimates of the contours along the selected points in Figure 18, from automated stereo matches, for the cargo container and an object (car) inside.

important steps are carried out: sensor calibration and stereo matching. Thanks to the constraints of the real scanning system, we model the system by using a linear pushbroom model with only one rotation angle. This greatly simplifies the calibration procedure and increases the robustness of the parameter estimation. A fast and automated stereo matching algorithm based on the free-form deformable registration approach (Zhu & Hu, 2007) is proposed to obtain 3D measurements of objects inside the cargo. Figure 19 shows the 3D measurements and visualization of objects inside the cargo container. The darker rectangular frames show the "ground truth" data of cargo container. The lighter lines (red in the online color version, with asterisks) show the 3D estimates of the contours along the selected points in Figure 18, from automated stereo matches, for the cargo container and an object (car) inside. With both the automatic matching procedure and the interactive 3D visualization procedure, I hope that this 3D measurement approach, for cargo inspection, can be put into practical use.

CONCLUSIONS

This article presents a pushbroom stereo mosaic approach for 3D reconstruction and visualization when one or more sensors (cameras) are in motion to

cover a large field of view of a scene or object. The proposed representation provides wide FOV, preserves extensive 3D information, and represents occlusions. This representation can be used as both an advanced video interface for surveillance and security or a pre-processing step for 3D reconstruction for these applications.

Several practical applications have been investigated, where parallel-perspective or full parallel projection mosaics can be generated. Related research issues are discussed in generating and using the parallel mosaics. In particular, the article presented a general ray interpolation approach for parallel-projection mosaic generation, and discussed some practical issues in generating the mosaics. A mosaic-based 3D rendering method, almost without any computation, allows for very effective 3D rendering of various complicated visual scenes, from forestry scenes to urban scenes, and from very far-range to extreme close-range. A Content-Based 3D Mosaic representation is also discussed to further extract both 3D and moving targets from the mosaics. Experimental results were given for four important applications—aerial video surveillance, ground mobile robot navigation, under vehicle inspection, and gamma-ray cargo inspection.

NOTES

1. http://vis-www.cs.umass.edu/projects/uvis/index.html
2. http://www-cs.engr.ccny.cuny.edu/~zhu/mosaic4uvis.html
3. http://www-cs.engr.ccny.cuny.edu/~zhu/StereoMosaic.html
4. http://www.realviz.com/products/mpro/index.php
5. http://www-cs.engr.ccny.cuny.edu/~zhu/CampusVirtualFly.avi
6. http://www.artbeats.com/prod/browse.php
7. http://www-cs.engr.ccny.cuny.edu/~zhu/Multiview/indoor1Render.avi

REFERENCES

Chai, J., & H.-Y. Shum (2000). Parallel projections for stereo reconstruction. *CVPR'00*: *II* 493–500.

Dickson, P., J. Li, Z. Zhu, A. R. Hanson, E. M. Riseman, H. Sabrin, H. Schultz, & G. Whitten (2002). Mosaic generation for under-vehicle inspection. *WACV'02*: 251–256.

Gupta, R., & R. Hartley (1997). Linear pushbroom cameras. *IEEE Trans PAMI, 19*(9), Sep.: 963–975.

Irani, M., P. Anandan, J. Bergen, R. Kumar, & S. Hsu (1996). Mosaic representations of video sequences and their applications. *Signal Processing: Image Communication*, vol. *8*, no. 4, May.

Leung, W. H., & T. Chen (2000). Compression with mosaic prediction for image-based rendering applications, *IEEE Intl. Conf. Multimedia & Expo.*, New York, July.

Li, Y., H.-Y. Shum, C.-K. Tang, & R. Szeliski (2004). Stereo reconstruction from multiperspective panoramas. *IEEE Trans. on PAMI, 26*(1): pp 45–62.

Lourakis, M. I. A., & A. A. Argyros (2004). The design and implementation of a generic sparse bundle adjustment software package based on the Levenberg-Marquardt algorithm, *ICS/FORTH Technical Report No. 340*, Foundation for Research and Technology—Hellas, Heraklion, Crete, Greece, August (http://www.ics.forth.gr/~lourakis/sba/).

Orphan, V. J., R. Richardson, & D. W. Bowlin (2002). VACIS™—a safe, reliable and cost- effective cargo inspection technology. *Port Technology International*, pp. 61–65. www.porttechnology.org/journals/ed16/section02.shtml.

Peleg, S., M. Ben-Ezra, & Y. Pritch (2001). OmniStereo: panoramic stereo imaging. *IEEE Trans. PAMI, March*: 279–290.

Shum, H.-Y., & R. Szeliski (1999). Stereo reconstruction from multiperspective panoramas. *ICCV'99*: 14–21.

Slama, C. C. (1980). Manual of Photogrammetry. Fourth Edition, *American Society of Photogrammetry*.

Sun, C., & S. Peleg (2004). Fast Panoramic Stereo Matching using Cylindrical Maximum Surfaces. *IEEE Trans. SMC, Part B*, 34, Feb.: 760–765.

Tang, H., Z. Zhu, G. Wolberg, & J. R. Layne (2006). Dynamic 3D urban scene modeling using multiple pushbroom mosaics, the *Third International Symposium on 3D Data Processing, Visualization and Transmission (3DPVT 2006)*, University of North Carolina, Chapel Hill, USA, June 14–16.

Triggs, B., P. McLauchlan, R. Hartley, & A. Fitzgibbon (2000). Bundle adjustment—a modern synthesis. In *Vision Algorithms: Theory and Practice*, Lecture Notes in Computer Science, vol 1883, pp. 298–372, eds. B. Triggs, A. Zisserman and R. Szeliski, Springer-Verlag.

Zhu, Z., A. R. Hanson, H. Schultz, & E. M. Riseman (2003). Generation and error characteristics of parallel-perspective stereo mosaics from real video. In *Video Registration*, M. Shah and R. Kumar (Eds.), Kluwer: 72–105.

Zhu, Z., E. M. Riseman, & A. Hanson (2004). Generalized parallel-perspective stereo mosaics from airborne videos. *IEEE Trans. PAMI, 26*(2), Feb: 226–237.

Zhu, Z., L. Zhao, & J. Lei (2005a). 3D Measurements in Cargo Inspection with a Gamma-Ray Linear Pushbroom Stereo System. *IEEE Workshop on Advanced 3D Imaging for Safety and Security*, June 25, San Diego, CA, USA.

Zhu, Z., E. M. Riseman, A. R. Hanson, & H. Schultz (2005b). An efficient method for geo-referenced video mosaicing for environmental monitoring. *Machine Vision Applications Journal*, Springer-Verlag, 16(4): 203–216.

Zhu, Z., H. Tang, B. Shen, & G. Wolberg (2005c). 3D and moving target extraction from dynamic pushbroom stereo mosaics. *IEEE Workshop on Advanced 3D Imaging for Safety and Security*, June 25, San Diego, CA, USA.

Zhu, Z., & A. R. Hanson (2006). Mosaic-based 3D scene representation and rendering, *Signal Processing: Image Communication*, *Special Issue on Representation of Still and Dynamic Scenes*, Elsevier, vol 21, no 6, Oct: 739–754. doi: 10.1016/j.image.2006.08.002.

Zhu, Z., & H. Tang (2006). Content-Based Dynamic 3D Mosaics. *IEEE Workshop on Three-Dimensional Cinematography (3DCINE'06)*, June 22, New York City (in conjunction with CVPR).

Zhu, Z., & Y.-C. Hu (2007). Stereo Matching and 3D Visualization for Gamma-Ray Cargo Inspection. *Proceedings of the Eighth IEEE Workshop on Applications of Computer Vision,* Feb 21st–22nd, Austin, Texas, USA.