3D and Moving Target Extraction from Dynamic Pushbroom Stereo Mosaics

Zhigang Zhu*, Hao Tang, Bing Shen and George Wolberg
Department of Computer Science, The City College of New York, New York, NY 10031
*zhu@cs.ccny.cuny.edu

Abstract

In this paper, we propose a dynamic pushbroom stereo mosaic approach for representing and extracting 3D structures and independent moving targets from urban 3D scenes. Our goal is to acquire panoramic mosaic maps with motion tracking information for 3D (moving) targets using a light aerial vehicle equipped with a video camera flying over an unknown area for urban surveillance. In dynamic pushbroom stereo mosaics, independent moving targets can be easily identified in the matching process of stereo mosaics by detecting the "out-of-place" regions that violate epipolar constraints and/or give 3D anomalies. We propose a segmentation-based stereo matching approach with natural matching primitives to estimate the 3D structure of the scene, particularly the ground structures (e.g., roads) on which humans or vehicles move, and then to identify moving targets and to measure their 3D structures and movements.

1. Introduction

Mosaics have become common for representing a set of images gathered by one or more (moving) cameras. In this paper we are particularly interested in parallel-perspective mosaics with *pushbroom stereo* geometry (Chai & Shum, 2000; Zhu, et al, 2001, 2004). The "pushbroom" is borrowed from satellite pushbroom imaging (Gupta & Hartley, 1997) where a linear pushbroom camera is used. The basic idea of the pushbroom stereo is as follows. If we assume the motion of a camera is a 1D translation and the optical axis is perpendicular to the motion, then we can generate two spatio-temporal images by extracting two scanlines of pixels of each frame, one in the leading edge and the other in the trailing edge. Each mosaic image thus generated is similar to a *parallel-perspective* image captured by a linear pushbroom camera (Gupta & Hartley, 1997), which has parallel projection in the direction of the camera's motion and perspective projection in the direction perpendicular to that motion. In contrast to the common pushbroom aerial image, pushbroom stereo mosaics are obtained from two different oblique viewing angles of a single camera's field of view. Since a fixed angle between the two viewing rays is selected for generating the stereo mosaics, uniform depth resolution is achieved, which is better than with perspective stereo, or the recently developed multi-perspective stereo with circular projection (Peleg, et al, 2001; Shum & Szeliski, 1999). Pushbroom stereo mosaics can be used in applications where the motion of the camera has a dominant direction. Examples include satellite pushbroom imaging (Gupta & Hartley, 1997), airborne video surveillance (Zhu, et al, 2001, 2004), 3D reconstruction for image-based rendering (Chai & Shum, 2000), road scene representations (Zheng & Tsuji, 1992; Zhu & Hanson, 2004), under-vehicle inspection (Dickson, et al, 2002; Koschan, et al, 2004), and 3D measurements of industrial parts by an X-ray scanning system (Gupta, et al, 1994; Noble, et al, 1995), and of articles in gamma-ray cargo inspection (Zhu, et al, 2005). However, as far as we know, previous work on the aforementioned stereo panoramas (mosaics) only deals with static scenes. Most of the approaches for moving target tracking and extraction, on the other hand, are based on interframe motion analysis and expensive layer extraction (Zhou and Tao 2003; Xiao and Shah 2004; Collins, 2003).

In this paper we propose a dynamic pushbroom stereo mosaic approach for extracting independent moving targets while reconstructing 3D structures of urban scenes from aerial video. Our goal is to acquire geo-referenced mosaic maps with motion tracking information for 3D (moving) targets using a light aerial vehicle flying over an unknown 3D environment. We assume that the aerial vehicle is equipped with a video camera and orientation measurement instrumentation (e.g., GPS and INS) so that geo-locating is possible. In real world applications, the motion of cameras on air vehicles cannot usually be constrained to 1D translation. In addition, extracting one scanline from each frame of a video sequence is not sufficient to generate a uniformly dense mosaic due to large and possibly varying displacement between each pair of successive frames. Several approaches have been proposed for dealing with this real-world issue, for example, the universal mosaicing approach (Rousso, et al, 1998) and the parallel ray interpolation approach (Zhu, et al, 2004). In this paper, for easy explanation, we assume that pushbroom stereo mosaics have been generated and that the epipolar geometry obeys that of the pushbroom stereo under 1D translation. We explore the dynamic pushbroom stereo geometry to extract moving targets. The generalization of the basic principle of pushbroom stereo to more general motion (Zhu, et al, 2004) and even to circular panoramic stereo (Peleg, et al, 2001) is straightforward given that the stereo mosaics can be generated and the epipolar geometry is known.

Simple window-based correlation approaches do not work well for man-made scenes, particularly across depth boundaries and for textureless regions. An adaptive window approach (Kanade & Okutomi, 1991) has been proposed which selects at each pixel the window size that minimizes the uncertainty in disparity estimates in stereo matching. A nine window approach has also been proposed by Fusiello, et al (1997) in which the point in the right image with the smallest SSD error amongst the 9 windows and various search locations is chosen as the best estimate for the given point in the left image. Recently, color segmentation has been used as a global constraint for refining an initial depth map to get sharp depth boundaries and to obtain depth values for textureless areas (Tao, et al 2001), and for accurate layer extraction (Ke & Kanade 2002). In this paper, we provide a segmentation-based approach using *natural matching primitives* to extract 3D and motion of the targets. The segmentation-based stereo matching algorithm is proposed particularly for the dynamic pushbroom stereo geometry to facilitate 3D reconstruction and moving target extraction from 3D urban scenes. However, the proposed natural matching primitives are applicable to more general scenes and other types of stereo geometry.

The paper is organized as follows. In Section 2, we will give a mathematical framework of the dynamic pushbroom stereo, and discuss its properties for moving target extraction. In Section 3, multi-view pushbroom mosaics are proposed to estimate 3D structures of moving targets. In Section 4, our stereo matching algorithm for (3D) static and moving target extraction will be provided. Preliminary experimental results will be given in Section 5. Finally is a brief summary in Section 6.

2. Dynamic Pushbroom Stereo Mosaics

Dynamic pushbroom stereo mosaics are generated in the same way as with the static pushbroom stereo mosaics described above. Fig.1 illustrates the geometry. A 3D point *P(X,Y,Z)* on a target is first seen through the leading edge of an image frame when the camera is at location L_1. If the point P is static, we can expect to see it through the trailing edge of an image frame when the camera is at location L_2. The distance between leading and trailing edges is d_y (pixels), which denotes the constant "disparity". However, if point P moves during that time, the camera needs to be at a different location L'_2 to see this moving point through its trailing edge. For simplifying equations, we assume that the motion of the moving points between two observations (L_1 and L'_2) is a 2D motion (S_x, S_y), which indicates that the depth of the point does not change over that period of time. Therefore, the depth of the moving point can be calculated as

$$Z = F\frac{B_y - S_y}{d_y} \quad (1)$$

where F is the focal length of the camera and B_y is the distance of the two camera locations (in the y direction). Mapping this relation into stereo mosaics following the notation in Zhu, et al (2004), we have

$$Z = H(\frac{d_y + \Delta y - s_y}{d_y}) \quad (2)$$

and

$$(S_x, S_y) = (Z\frac{s_x}{F}, H\frac{s_y}{F}) = (Z\frac{\Delta x}{F}, H\frac{s_y}{F}) \quad (3)$$

where H is the depth of plane on which we want to align our stereo mosaics, (Δx, Δy) is visual motion in the stereo mosaics of the moving 3D point P, and (s_x, s_y) is the target motion represented in stereo mosaics. Obviously, we have $s_x = \Delta x$.

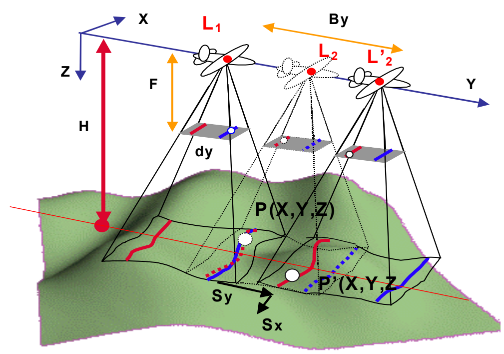

Fig. 1. Dynamic pushbroom stereo mosaics

We have the following interesting observations about the dynamic pushbroom stereo geometry for moving target extraction.

(1) *Stereo fixation*. For a static point (i.e. $S_x = S_y = 0$), the visual motion of the point with a depth H is (0,0), indicating that the stereo mosaics thus generated fixate on the plane of depth H. This fixation facilitates the detection of moving targets on that plane.

(2) *Motion accumulation*. For a moving point ($S_x \neq 0$ and/or $S_y \neq 0$), the motion between two observations accumulates over a period of time due to the large distance between the leading and trailing edges. This will increase the discrimination of slow moving objects viewed from a relatively fast moving aerial camera.

(3) *Epipolar constraints.* In the ideal case of 1D translation of the camera (with which we present our dynamic pushbroom stereo geometry), the correspondences of static points are along horizontal epipolar lines, i.e. $\Delta x = 0$. Therefore, for a moving target *P*, the visual motion with nonzero Δx will identify itself from the static background in the general case when the motion of the target in the *x* direction is not zero (i.e., $S_x \neq 0$). In other words, the correspondence pair of such a point will violate the epipolar line constraint for static points (i.e. $\Delta x = 0$).

(4) *3D constraints.* Even if the motion of the target happens to be in the direction of the camera's motion (i.e. the *y* direction), we can still discriminate the moving target by examining 3D anomalies. Typically, a moving target (a vehicle or a human) moves on the flat ground surface (i.e., road) over the time period when it is observed through the two edges of the video images. We can usually assume that the moving target has the same depth as its surroundings given that the distance of the camera from the ground is much larger than the height of the target. A moving target in the direction of camera movement, when treated as a static target, will show 3D anomaly - either hanging up above the road (when it moves to the opposite direction, i.e., $S_y < 0$), or hiding below the road (when it moves in the same direction, i.e., $S_y > 0$).

After a moving target has been identified, the motion parameters of the moving target can be estimated. We first estimate the depth of its surroundings and apply this depth *Z* to the target, then calculate the object motion s_y using Eq. (2) and (S_x, S_y), using Eq. (3), given the known visual motion ($\Delta x, \Delta y$) observed in the stereo mosaics.

3. Multi-View Pushbroom Mosaics

A pair of stereo mosaics (generated from the leading and trailing edges) is a compact representation for both 3D structures and target movements. However, there are two remaining issues. First, stereo matching will be difficult due to the largely separated parallel views of the stereo pair. Second, for some unusual target movements, e.g. moving too fast, changing speed or direction, we may either have two rather different images in the two mosaics (if changing speed), or see the object only once (if changing direction), or never see the object (if it maintains the same speed as the camera and thus never shows up in the second edge window).

Therefore we propose to generate multi-view mosaics (more than 2), each of them with a set of parallel rays whose viewing direction is between the leading edge to the trailing edge (Fig. 2). The multiple mosaic representation will ease the stereo correspondence problem in the same way as the multi-baseline stereo (Okutomi & Kanade, 1993). Multiple mosaics also increase the possibility to detect moving targets with unusual movements and also to distinguish the movements of the specified targets (e.g., ground vehicles) from those of trees or flags in wind. Furthermore, multiple mosaics can be used for 3D estimation of moving targets where the heights of the targets cannot be neglected. This will be detailed below.

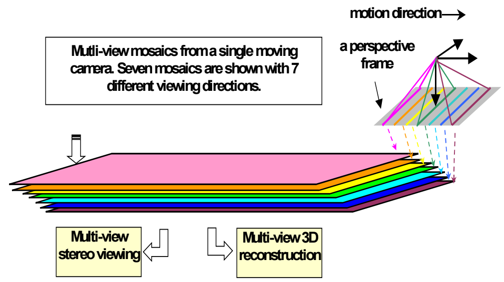

Fig. 2. Multi-view pushbroom mosaics

In order to estimate the height of a moving target from the ground, we will need to see both the bottom and the top of an object. A pair of pushbroom mosaics with one forward-looking view and the other backward-looking view exhibits obvious different occlusions; in particular, the bottom of a target (e.g., a vehicle in Fig. 3(a)) can only seen in one of the two views. However, any two of the multi-view pushbroom mosaics, both with forward-looking (or backward-looking) parallel rays, will have almost the same occlusion relation to satisfy the condition for height estimation.

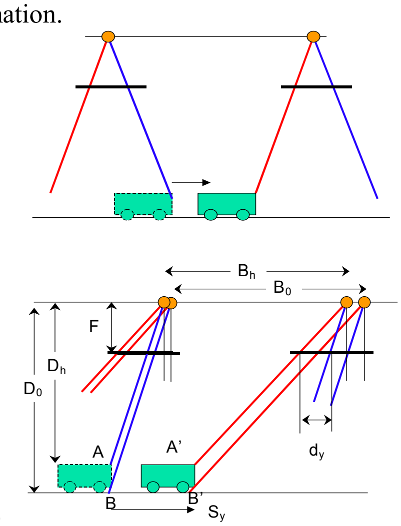

Fig. 3. Height from dynamic pushbroom stereo (a) infeasible pair; (b) feasible pair

Fig. 3(b) illustrates the case of a pair of backward-looking pushbroom stereo mosaics. Point *A* and *B* are two points on a target (vehicle), one on the top and the other on the bottom. Both of them are first seen in the mosaic with parallel rays of a smaller slanting angle, and then seen in the mosaic with parallel rays of a larger slanting angle. The distance between the two different rays within an image frame is still defined as d_y. The visual motion in the y direction is Δy_h and Δy_0, respectively, and can be measured in the stereo pair. Between the two parallel views, let us assume the motion of the target is S_y in 3D space and s_y in the mosaiced images. Then the depths of the points on the top and on the bottom are

$$Z_h = F\frac{B_h - S_y}{d_y} = H(\frac{d_y + \Delta y_h - s_y}{d_y}) \quad (4)$$

and

$$Z_0 = F\frac{B_0 - S_y}{d_y} = H(\frac{d_y + \Delta y_0 - s_y}{d_y}) \quad (5)$$

respectively. Depth Z_0 of the bottom point could be obtained from the surroundings (ground) of the target. Then, the object motion s_y (and therefore S_y) can be calculated using Eq. (5). Finally, the depth of the point on the top, Z_h, can be estimated using Eq. (4), given the known visual motion of that point, Δy_h, and its independent motion component s_y obtained from the bottom point *B*.

Multiple pushbroom mosaics can also be used for image-based rendering with stereo viewing in which the translation across the area is simply a shift of a pair of mosaics, and the change of viewing directions is simply a switch between two consecutive pairs of mosaics. A video clip of a mosaic-based virtual fly-through is show at our web site (Zhu 2002), where "motion parallax" of the 3D buildings and moving vehicles are amplified. In the video clip, a pair red/blue glasses will help the viewer to perceive better 3D, but due to the virtual fly-through with motion parallax, the 3D effects are also obvious without using glasses. In the next section, we will discuss a new method to extract both of the 3D buildings and moving targets from these stereo mosaics.

4. Natural Primitive Stereo Matching

Dynamic pushbroom stereo mosaics provide several advantages for 3D reconstruction and moving target extraction. The stereo mosaics are aligned on a dominant plane (e.g., the ground). All the static objects obey the epipolar geometry, i.e. along the epipolar lines of pushbroom stereo. An independent moving object, on the other hand, either violates the epipolar geometry if the motion is not in the direction of sensor motion or at least exhibits 3D anomaly - hanging above the road or hiding below the road even if motion happens to be in the same direction of the sensor motion. With all these geometric constraints in mind, we propose a segmentation-based approach to integrate the estimation of 3D structure of an urban scene and the extraction of independent moving objects from a pair of dynamic pushbroom stereo mosaics. The approach starts with one of the mosaics, for example, the left mosaic, by segmenting it into homogeneous color regions that are treated as planar patches. We apply the mean-shift-based approach proposed by Comanicu & Meer (2002) for color segmentation. Then the stereo matching is performed based on these patches, between two original color mosaics. The basic idea is to only match those pixels that belong to each region (patch) between two images in order to both produce sharp depth boundaries for man-made targets and to facilitate the searching and discrimination of the moving targets (each covered by one or more homogeneous color patches). The proposed algorithm has the following five steps.

Step 1. *Matching primitive selection*. After segmenting the left image using the mean-shift method, homogenous color patches and then the natural matching primitives are extracted.

Step 2. *Epipolar test*. Using pushbroom epipolar geometry in stereo matching, static objects will find correct matches but moving objects will be outliers without correct "matches".

Step 3. *3D anomaly test*. After ground surface fitting (and road detection), moving objects in the same motion direction will exhibit wrong 3D characteristics (hanging above roads or hiding below roads).

Step 4. *Motion extraction.* Search matches for outliers (which could be moving objects) with a 2D and larger search range, or along the road directions (if available).

Step 5. *3D estimation*. Using the dynamic pushbroom stereo proposed in Section 3, the 3D structures and motion of moving objects could be derived.

In the following two subsections, we will detail two important issues in the segmentation-based stereo matching approach: natural matching primitive selection and an integrated analysis of 3D structure and motion for both static and moving targets.

4.1. Natural matching primitives

In this paper, we will use color segmentation to obtain natural matching primitives for both 3D reconstruction and moving target extraction. The selection and matching of the natural matching primitives includes the following five sub-steps.

(1) *Segmentation and Interest point extraction*. The left mosaic is segmented into homogeneous color regions using the mean-shift approach (Comanicu & Meer 2002). We assume that each homogeneous color region (patch) is planar in 3D. However, each planar surface in 3D may be divided into several color

patches. Then the boundary of each region is traced as a close curve. All the neighborhood regions are also connected with the region in processing for further use. Finally we use a line fitting approach to extract interest points along the region's boundary. The boundary is first fitted with connected straight-line segments using an iterative curve splitting approach. The connecting points between line segments are defined as interest points.

(2) *Natural window definition*. Each interest point *P(x,y)* of a region **R** in consideration will be used as the center of an m×m rectangular window in the left mosaic. Only those points that are within the window, inside the regions, or on the boundary will be used for matching (Fig. 4) in order to keep sharp depth boundaries. The window is defined as a *natural matching window* and the set of pixels involved in matching is called a *natural matching primitive*. To facilitate the computation of correlation for stereo matching, we define a region mask ***M*** of size m×m centered at that interest point such that

$$M(i,j) = \begin{cases} 1, \text{if } (x+i, y+j) \in \mathbf{R} \\ 0, \text{otherwise} \end{cases} \qquad (6)$$

We changed the size m of the natural window depending on the sizes of the regions. In our experiments, we use m = 23 for large regions (with diameter >=23) and m = 15 for small regions. We also want to include a few more pixels (1-2) around the region boundary (but not belonging to the region) so that we have sufficient image features to match. Therefore, a dilation operation will be applied to the mask ***M*** to generate a region mask covering pixels across the depth boundary. Fig. 4 shows four such windows for the four interest points for the top region of the box. Note the yellow-shaded portions within each rectangular window, indicating that the pixels for stereo matching cover the depth boundaries.

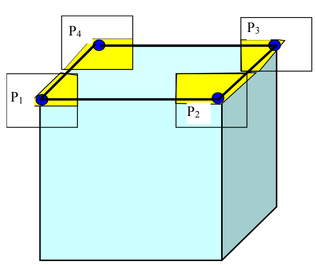

Fig. 4. Natural matching primitive

(3) *Natural window-based correlation*. Let the left mosaic and right mosaics be denoted as $\boldsymbol{I_1}$ and $\boldsymbol{I_2}$, respectively. The weighted cross-correlation, based on the natural window centered at the point *P(x,y)* in the left mosaic, is defined as

$$C(\Delta x, \Delta y) = \frac{\sum_{i,j} M(i,j) I_1(x+i, y+j) I_2(x+i+\Delta x, y+j+\Delta y)}{\sum_{i,j} M(i,j)} \qquad (6)$$

Note that we still carry out correlation between two color images but only on those interest points on each region boundary, and only with those pixels within the region and on the boundaries. This equation works both for static objects when the searching of correspondences is along epipolar lines of the pushbroom stereo and also for moving targets when the searching should be in 2D and with a larger search range. In the real implementation, we first perform matches with epipolar constraints of the pushbroom stereo, and those without good matches with be treated as "outliers" for further examination to see whether or not they are moving objects. Fig. 5 shows a real example of natural-window-based stereo matching result for a static object (top of a building). The 19 selected interest points and their correspondences are marked on the boundaries in the left and right images, respectively. One mismatch and a small error in match are also indicated on images.

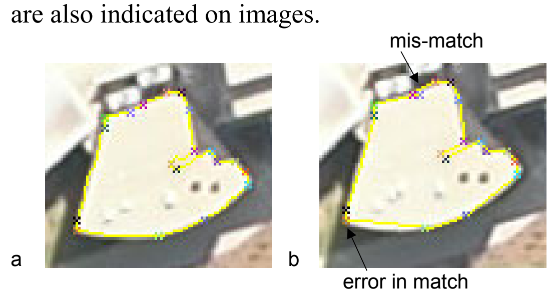

Fig. 5. An example of region matching results. The matches are marked as "X", with corresponding colors.

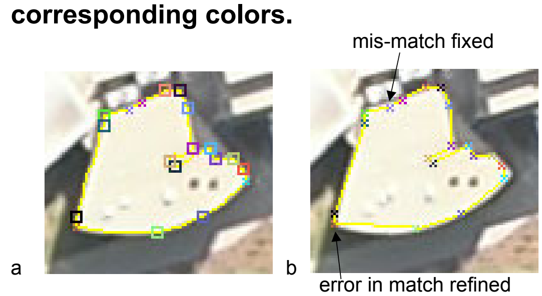

Fig. 6. An example of surface fitting results

4. 2. Surface fitting and motion estimation

Assuming that each homogeneous color region is planar in 3D, we fit a 3D plane for each region after we obtain the 3D coordinates of the interest points of the region using the pushbroom stereo geometry (assuming that it is static, i.e., $s_y = 0$ in Eq. (2)). Seed points (≥3) are selected for plane fitting based on their correlation values. Then, the 3D coordinates of all the interest points are refined by constraining them on the fitted plane. Then, using the 3D plane information the region in the left mosaic can be warped to the right image to evaluate the matching and the fitting. Fig. 6 shows the results of fitting and back-projection of the fitted region onto the right image. The 15 seed interest points (out of 19) used for planar fitting are indicated on the left image as squares. Both the mismatch and the small error in the initial match are fixed.

Note that an iterative approach could be applied here to refine the matches after the initial surface fitting by using the evaluation of the warping from the left to the right mosaics and also by using the occlusion constraints from neighborhood regions, which have been obtained in region tracing step in Section 4.1. For example, the selection of the seed points for surface fitting can be refined by removing those points that could be on occluding boundaries after we check the initial 3D surface relations and adding some other points that have reliable matches after image warping evaluation. This is still our ongoing work.

Neighborhood regions can also be merged into a single plane if they share the same planar equation, with some error tolerant range. After region refinement and merging, large and (near) horizontal ground regions can be easily identified. It is also possible to analyze the shape of the ground regions to estimate road directions in which vehicles move. For those smaller neighborhood regions that happen to move in the same direction as the camera, will have large depth differences from the surrounding ground regions when treated as static objects. This 3D anomaly can be used to identify those regions as moving objects. By assuming that their depths are the same as the surroundings, their motion parameters can be estimated using Eq. (2). For those "outliers" that do not find matches in the first pass (along epipolar lines), searching for matches can be performed along possible road directions (if obtained from the surrounding ground regions), or simply performed in a much larger 2D searching range.

5. Experimental Results

We have performed preliminary experiments for stereo matching and moving object extraction on pushbroom stereo mosaics from real video sequences. Fig. 7(a) shows a pair of stereo mosaics from a video sequence that was taken when the airplane was about 300 meters above the ground. The "constant disparity" for this pair of pushbroom mosaic is 72 (i.e. $d_y = 72$ in Eq. (2)). Figs. 7(b) –(i) show the results of a small window of the 8Kx1K stereo mosaics. In Figs. 7(b) and (c), the dynamic pushbroom stereo pair has both stationary buildings and ground vehicles moving in different directions. Figs. 7(d) and (e) show the segmentation result of the left image in Fig. 7(b), where the color label image is shown in Fig. 7(d) and the region boundaries are shown in Fig. 7(e). Note that a planar 3D region may be segmented into several color patches. Figs. 7(f) and (g) show the matching results for some typical regions: ground planes in Fig. 7(f), buildings and moving vehicles in Fig. 7(g). The region boundaries in the left image and the corresponding matches in the right image are drawn in blue and red, respectively, both on the left image. As is obvious in Fig. 7(a), the stereo mosaics do not exhibit the epipolar geometry of exact horizontal epipolar lines. However, the correspondence points of static objects are within a very small range of variations in the vertical coordinates between the two mosaics. Therefore, the automatic searching for correspondences of static objects is along the actual epipolar curves determined by stereo mosaicing (Zhu, et al, 2001, 2004). The correspondences of the moving vehicles are searched along several predefined road directions (as marked in Fig. 7(g)), with a search range of ±20 pixels in the direction of road and ±6 pixels perpendicular to the road direction. Those road directions could be obtained by analyzing the ground plane regions as shown in Fig. 7(f), which is our ongoing work. From the region matching results it is clear that the ground plane regions are almost aligned (in Fig. 7(f)) and the points on the tops of the buildings move along their epipolar curves, almost in the horizontal direction in Fig. 7(g). The moving vehicles, on the other hand, exhibit much larger motion magnitudes and obvious different motion directions. Fig. 7(h) shows the depth map of the static regions. Note that many regions, particularly those on top of each building are correctly merged, and the depth boundaries are clearly sharp and accurate. Fig. 7(i) shows the matched moving targets marked on the left mosaiced image, in blue and red respectively.

6. Conclusions and Discussions

In this paper we present a new approach to extract both 3D structure and independent moving targets from long video sequences. The principles of dynamic pushbroom stereo mosaics are presented, which shows that the new geometry has advantages, in both moving object extraction and of 3D estimation, in terms of panoramic field of view, adaptive baseline system, independent motion accumulation, and parallel-perspective epipolar and 3D constraints for discriminating moving targets. The idea of a multi-view pushbroom mosaic is proposed to show the potential to estimate 3D structures of moving targets and to analyze different motion patterns. The multi-view pushbroom mosaics also provide an effective way for image-based rendering without any 3D reconstruction.

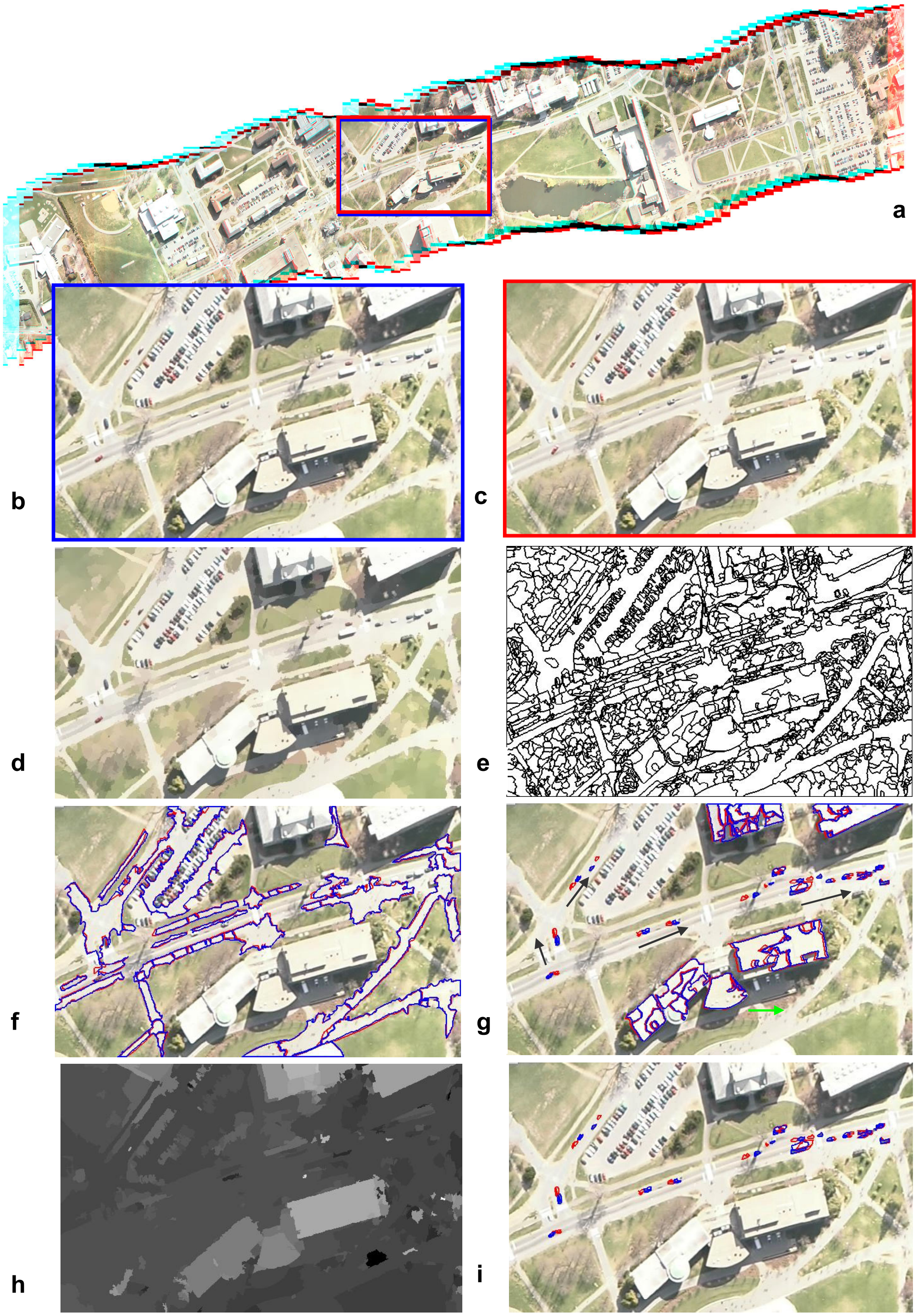

Fig. 7. Structure and motion from dynamic pushbroom stereo mosaics. (a) Stereo mosaics; (b)- (i) results for a portion of the stereo mosaics: (b) and (c) are the left and right mosaics; (d) and (e) are left color labels and boundaries; (f) and (g) are matches of ground regions and non-ground regions marked on the left mosaic, with blue and red respectively; (h) depth map of the static regions; (I) moving targets marked on the left mosaic (motion: blue to red). Not all the matches are shown for easy visualization.

Based on the properties of the dynamic pushbroom stereo mosaics, we propose a new segmentation-based stereo matching approach for both 3D reconstruction and moving target extraction from a pair of dynamic pushbroom stereo mosaics. A simple yet effective natural matching primitive selection method is provided. This method is effective for stereo matching of man-made scenes, particularly when both 3D facilities and moving targets need to be extracted. We discussed the natural-primitive-based matching approach in the scenario of parallel-perspective pushbroom stereo geometry, but apparently the method is also applicable to other types of stereo geometry such as perspective stereo, full parallel stereo, and circular projection panoramic stereo.

The preliminary experimental results are very promising. Ongoing and future work includes: (1) extension of the framework to other panoramic stereo geometry (Peleg, et al 2001; Zhu et al 2004; Chai & Shum, 2000); (2) ground surface analysis and road direction estimation; (3) fully automated algorithms for 3D reconstruction and moving target extraction from multi-view pushbroom mosaics; and (4) dynamic stereo mosaic generation under more general motion.

7. Acknowledgements

This work is supported by the Air Force Research Lab under the RASER program (Award No. FA8650-05-1-1853) and an AFRL/HECB Grant (No. F33615-03-1-63-83), and by New York Institute for Advanced Studies. Some data were captured in the UMass Computer Vision Lab under a NSF grant (No EIA-9726401). The U.S. Government is authorized to reproduce and distribute reprints for Governmental purposes notwithstanding any copyright notation thereon. However, the views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of the U.S. Government. Thanks are also given to Mr. Robert Hill at the City College of New York for proofreading the manuscript.

8. References

Chai, J. and H –Y. Shum, 2000. Parallel projections for stereo reconstruction. In *Proc. CVPR'00*: II 493-500.

Collins, R., 2003. Mean-shift blob tracking through scale space, In *Proc.CVPR'03*.

Comanicu, D. and P. Meer, 2002. Mean shift: a robust approach toward feature space analysis. *IEEE Trans. PAMI*, May 2002

Dickson, P., J. Li, Z. Zhu, A. Hanson,, E. Riseman, H. Sabrin, H. Schultz and G. Whitten, 2002. Mosaic generation for under-vehicle inspection. *IEEE Workshop on Applications of Computer Vision*, Orlando, Florida, Dec 3-4, 2002

Fusiello, A., V. Roberto and E. Trucco, 1997. Efficient stereo with multiple windowing, In *IEEE CVPR*: 858-863

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

Kanade, T. and M. Okutomi, 1991. A Stereo Matching Algorithm with an Adaptive Window: Theory and Experiment, In *Proc. IEEE ICRA'91*, Vol. 2: 1088-1095

Ke, Q. and T. Kanade, 2001. A subspace approach to layer extraction, In *Proc. IEEE CVPR'01*.

Koschan, A., D. Page, J.-C. Ng, M. Abidi, D. Gorsich, and G. Gerhart, 2004. SAFER under vehicle inspection through video mosaic building," *International Journal of Industrial Robot*, September 2004, 31(5): 435-442

Noble, A., R. Hartley, J. Mundy and J. Farley. X-Ray Metrology for Quality Assurance, In *Proc IEEE ICRA'94*, vol 2, pp 1113-1119

Okutomi M. and T. Kanade,1993. A multiple-baseline stereo," *IEEE Trans. PAMI*, vol. 15, no. 4, pp. 353-363.

Peleg S, Ben-Ezra M and Pritch Y., 2001. Omnistereo: panoramic stereo imaging, *IEEE Trans. PAMI*, 23(3): 279-290.

B. Rousso, S. Peleg, I. Finci and A. Rav-Acha, 1998. Universal mosaicing using pipe projection, In *Proc. ICCV'98:* 945-952.

Shum, H.-Y. and Szeliski, R., 1999. Stereo reconstruction from multiperspective panoramas. In *Proc. ICCV'99*: 14-21

Tao, H., H. S. Sawhney and R. Kumar, 2001. A global matching framework for stereo computation, In *Proc. ICCV'01*

Xiao, J. and M. Shah, 2004. Motion layer extraction in the presence of occlusion using graph cut, In *Proc. CVPR'04*

Zhou, Y. and H. Tao, 2003. A background layer model for object tracking through occlusion," In *Proc. ICCV'03*: 1079-1085.

Zhu, Z. , E. M. Riseman and A. R. Hanson, 2001. Parallel-perspective stereo mosaics. In *Proc. ICCV'01*, vol I: 345-352.

Zhu, Z., 2002. Mosaic-based 3D Scene Rendering. http://www-cs.engr.ccny.cuny.edu/~zhu/CampusVirtualFly.avi

Zhu, Z. and A. R. Hanson, 2004. LAMP: 3D Layered, Adaptive-resolution and Multi-perspective Panorama - a New Scene Representation, *Computer Vision and Image Understanding*, 96(3), Dec 2004, pp 294-326.

Z. Zhu, E. M. Riseman, A. R. Hanson, 2004. Generalized Parallel-Perspective Stereo Mosaics from Airborne Videos, *IEEE Trans PAMI*, 26(2), Feb 2004, pp 226-237

Zhu, Z., L. Zhao, J. Lei, 2005. 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, 2005, San Diego, CA, USA