Image-Based Model Updating

Philippe Simard and Frank P. Ferrie
Centre for Intelligent Machines
McGill University
Montréal (Québec), Canada, H3A 2A7
{psimard,ferrie}@cim.mcgill.ca

Abstract

This paper presents a novel image-based approach for updating the geometry of 3D models. The technique can cope with large-scale models, using a single imaging sensor to which an arbitrary motion is applied. Current approaches usually do not fully take advantage of strong prior information, often available in the form of an initial model. The approach is thus novel in that geometric anomalies are quickly detected, significantly reducing problem complexity. Hence, given a geometric model and known camera motion, the image warping can be calculated and intensity patterns can be predicted. If predictions do not match observations, the model is assumed to be incorrect. The updating is then cast as an optimization problem where differences between observations and predictions are minimized. The algorithm is tested against both synthetic and real imaging data to update a terrain model. Results show that the algorithm can automatically detect and correct geometrical problems of different types and sizes.

1 Introduction

This paper considers the problem of updating the geometry of a model using imaging data. A novel approach for both localizing and estimating geometric errors is proposed. Although the updating problem is very similar to classical shape-from-motion, it is very different in that strong *a priori* information about the scene is available initially. Despite this fact, the problem remains ill-posed as many solutions may still explain the imaging data. It is proposed that such prior information can be used to quickly localize geometric anomalies. This greatly simplifies the updating process since only small regions of the model will actually require to be updated.

Related work includes research on recursive surface reconstruction using multiple images [1,2,4,7]. Although such algorithms could be applied for model updating, they do not fully take advantage of prior information as they require reconstruction of the entire scene at each iteration. Some techniques for doing coarse-to-fine motion stereo make better use of a model. For example, Xu et al. [8] show how to use a coarse model to guide the searching (i.e. solving the correspondence problem) to get a finer model.

Hung et al.'s method [3] differs from traditional approaches by not explicitly solving the correspondence problem. Instead, it solves for the parameters that best account for the relationship between images. Unfortunately, simple parametric shapes are used as prior information about the scene structure to minimize complexity. Even though this

allows the problem to be made more tractable, it does not generalize very well for surfaces having more complex shapes.

The approach proposed in this paper is related to their work but assumes much less constrained shapes. Indeed, surfaces that can be modelled as a single-valued function (i.e. an elevation map) are specifically considered. This is possible assuming that the error in the model is not only isolated, but that it can also be detected so as to reduce the number of parameters (i.e. elevation values) that require correction.

The technique requires a single imaging device (of any type), as long as its projective geometry can be determined. It is differential and as such, requires motion of the sensor which is assumed to be known. It uses the geometric information of the model to predict the warping of sensor imagery. Given that a motion is induced, image predictions can be generated. If the predictions are different from the observations, the model is assumed to be inconsistent with reality and geometric errors can be localized. The elevation values defining the surface are then corrected using an optimization procedure in which the difference between predictions and observations is minimized. The minimization is performed using a gradient-descent strategy where the initial elevation values (provided by the model to be updated) are used as the initial starting point. The prior model serves to reduce the dimensionality of the minimization space as well as to provide an initial point for initializing the steepest descent.

The technique is thus useful for real-time applications and/or where large scale models are necessary. A growing number of applications actually require models having highly accurate geometry. Synthetic vision systems in aviation are a good example as they require a precise model for rendering artificial views of the world (using both the model and pose information). Their goal is to support navigation and situational awareness in low visibility conditions. Because the model reflects at best a nominal state of the environment, it needs to be periodically updated to ensure its consistency with reality. A sample view of such a model is shown in Figure 1a along with a corresponding picture of the real world (Figure 1b). Note that the overall geometry is quite accurate but that the mountain's shape is slightly incorrect along the occluding contour. A good model updating technique should be able to detect, localize and correct such a geometric inconsistency.
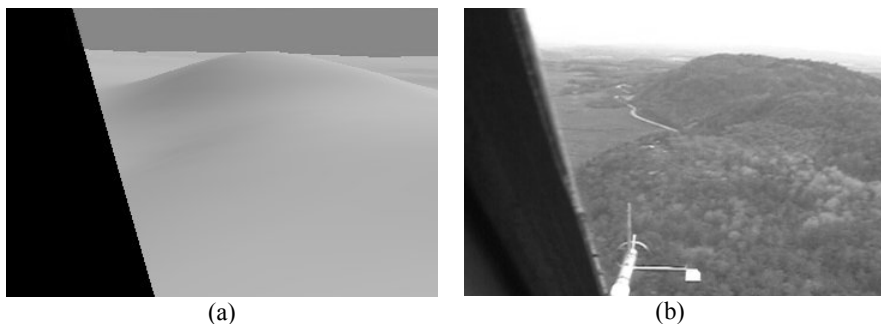

(a) (b)

Figure 1: (a) Sample view of the model, (b) Corresponding real world view

2 Theory

2.1 Problem Setup

The notation used in this paper (inspired from [3]) will denote column vectors in boldface and transpose vectors by the superscript ^{t}. Suppose that a surface is being imaged from two different viewpoints, as shown in Figure 2. A pinhole camera model having a camera reference frame CRF_n attached to the nth position is assumed. Let P be a point in 3D space belonging to the surface and let $\mathbf{q}_n$ be its 2D image coordinates in image n. Since only two images are considered, $\mathbf{q}_1$ and $\mathbf{q}_2$ will simply be denoted by $\mathbf{s}$ and $\mathbf{u}$ respectively. Let also the coordinates of P in the world reference frame (WRF) be denoted by $\mathbf{r} = (x,y,z)^t$.
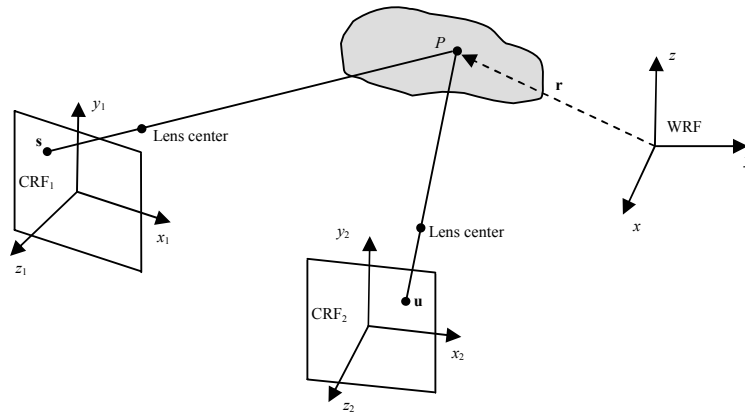

Figure 2: A surface being imaged from two different viewpoints

Given that the image of P is at coordinates $\mathbf{s}$ in image 1, its corresponding coordinates $\mathbf{u}$ in image 2 can be calculated with the following two steps:

1. *Back-projection from image 1*: find the corresponding 3D point P by back-projecting a 3D ray from the image point $\mathbf{s}$ through the camera lens center to the 3D space and then intersecting the ray with the surface.
2. *Forward-projection to image 2*: find the image point $\mathbf{u}$ by projecting the 3D point P along the ray from P, through the lens center, to the image plane of camera 2.

These two steps can be combined to define the following relationship,

$$\mathbf{u} = h(\mathbf{s}, \mathbf{b}, \mathbf{r}), \tag{1}$$

where $\mathbf{b}$ specifies the camera parameters which include the camera positions and the focal length. If the transformation of the image intensity is completely determined by the geometric transformation of the image points, the intensity of point P at $\mathbf{s}$ in image 1 and at $\mathbf{u}$ in image 2 will be identical,

$$I_1(\mathbf{s}) = I_2(\mathbf{u}) = I_2(h(\mathbf{s}, \mathbf{b}, \mathbf{r})). \tag{2}$$

2.2 Error Localization

In the context of this research, models correspond to digital elevation maps which are represented by single valued functions of the form,

$$z = f(x, y), \tag{3}$$

where z represents elevation above the *(x,y)* plane in the WRF. Referring back to Figure 1, the ultimate goal is to correct the elevation map corresponding to Figure 1a so that it is consistent with the scene shown in Figure 1b. Since errors are isolated, such erroneous points can first be localized and then adjusted by a correction procedure so as to be consistent with scene measurements. In this case, all points $\mathbf{r}$ that are inconsistent with the model (Equation 3) should be identified.

Such points are easily localized given the setup in Figure 2, the nominal surface $z = f(x,y)$ and the relative camera positions (e.g. the motion of a moving observer). For each $\mathbf{s}$ in I_1, the corresponding $\mathbf{u}$ in I_2 is calculated via Equation 1. Given the brightness constancy assumption (Equation 2), the intensity at $\mathbf{s}$ maps directly to the intensity at $\mathbf{u}$. If the model is a correct representation of the scene geometry, then the predicted intensity value at $\mathbf{u}$ should be consistent with the intensity measured at the corresponding location in I_2. Consequently, if $I_1(\mathbf{s})$ is different from $I_2(\mathbf{u})$, then $\mathbf{r}$ is assumed to be incorrect. In practice, a point is considered to be anomalous if the sum-of-absolute intensity differences in the neighborhood of $\mathbf{s}$ is smaller than a given threshold (Section 2.4 discusses on how to determine its value).

2.3 Iterative Correction

Let $\mathbf{R}$ be the vector containing all $\mathbf{r}_i$ corresponding to each position $\mathbf{s}_i$ in image 1 marked as anomalous. Following from Equation 2, the problem of correcting these anomalies can be cast in terms of minimizing the following functional,

$$e(\mathbf{R}) \equiv \sum_i \left| I_1(\mathbf{s}_i) - I_2(h(\mathbf{s}_i, \mathbf{b}, \mathbf{r}_i)) \right|, \tag{4}$$

where i corresponds to the index set of $\mathbf{R}$ and $|\cdot|$ is the absolute value. Furthermore, for the case of elevation maps, the error is assumed to be largely restricted to the z component of $\mathbf{r}$. Hence, the error function of Equation 4 becomes,

$$e(\mathbf{Z}) \equiv \sum_i \left| I_1(\mathbf{s}_i) - I_2(h(\mathbf{s}_i, \mathbf{b}, z_i)) \right|, \tag{5}$$

where the vector $\mathbf{Z}$ consists of all anomalous elevations z_i. A gradient descent strategy is used to solve for $\mathbf{Z}$ according to the following iterative scheme,

$$\mathbf{Z}_{t+1} = \mathbf{Z}_t - k \cdot \left. \frac{\partial e(\mathbf{Z})}{\partial \mathbf{Z}} \right|_{\mathbf{Z}=\mathbf{Z}_t}, \tag{6}$$

where t is the iteration index and k a positive scaling factor minimizing

$$e\left(\mathbf{Z}_t - k \cdot \left.\frac{\partial e(\mathbf{Z})}{\partial \mathbf{Z}}\right|_{\mathbf{Z}=\mathbf{Z}_t}\right). \tag{7}$$

The starting point (i.e. $\mathbf{Z}_0$) is provided by the initial elevation values of the model. The gradient can be evaluated by systematically varying the components of $\mathbf{Z}$. Unfortunately, there is no guarantee that the algorithm will converge to global minimum. In fact, the error function is not only dependent on how far off the initial model is from the actual scene, but also on the texture patterns present in the images.

2.4 Deviations from the Brightness Constancy Assumption

The brightness constancy assumption (Equation 2) is certainly useful for simplifying the problem. However, it cannot be used in general since not all real world surfaces are Lambertian. Other problems, such as sensor noise or sampling mismatches, will also limit the ability to predict the appearance of a given scene element. A simple approach to the problem is to assume that these deviations can be bounded i.e.

$$\left| I_1(\mathbf{s}) - I_2(\mathbf{u}) \right| < t, \tag{8}$$

where t is a threshold. The value of t may be determined by modelling all potential sources of faulty predictions and their relative contribution. As the derivation of a realistic model may be a hard task, a statistical approach is used instead. The basic idea is to choose a region of the model for which ground truth is available and where various typical surfaces are present. In that case, prediction errors caused by deviations from the Lambertian assumption will be isolated.

Computing t first requires that imaging data has to be acquired from different viewpoints and if possible, for different lighting conditions. Predictions for the appearance of different surface elements then have to be generated and their deviations from their observed appearance have to be measured. These deviations are finally averaged to produce a realistic threshold value.

3 Experimental Results

The application field examined in this paper is the one of synthetic vision in aviation. The particular system considered, called the Enhanced and Synthetic Vision System (ESVS), is to be used on helicopters for Search and Rescue operations [5]. It includes an on-board image generator to render a synthetic image of local terrain (determined from a model consisting of a digital elevation map and cultural data), registered through the aircraft navigation system. A helmet mounted display enables the pilot to localize the position and attitude of the aircraft in the local environment. Situational awareness is sufficient to allow the pilot to navigate en-route by recognizing landmarks that would otherwise be hidden by poor visibility.

To achieve this goal, such a system requires geometrically correct models of the areas over which the aircraft will be flown. As a consequence, models have to be updated regularly to make sure they correspond to reality. Given this context, the

algorithm was used to update a model consisting of a digital elevation map (DEM) provided by the Canadian Government. The DEM has grid posts every 93m in latitude and 65m in longitude and elevation errors of up to 25m.

3.1 Artificial Images

The algorithm was first tested against synthetically generated data to determine its general performance in addition of studying its convergence properties. A checkerboard textured flat terrain (having the same resolution as the real DEM) was used in the experiments. The resolution of the surface was kept constant for the experimentation (i.e. no extra elevation points were added when updating the model) and interpolation between adjacent points was done using triangles.

Known elevation errors were introduced in two adjacent elevation points (of -25m and 25m respectively). The surface was imaged from the top looking straight down at an altitude of 300m and a lateral displacement of 30m was induced. Two checkerboard textures of different scale were used to see the performance of the algorithm subject to different imaging data (one having the same grid size as the DEM and the other using a grid size about 5 times smaller). The algorithm was first used to verify the terrain geometry and the two erroneous elevation points (out of the 133 visible elevation points) were correctly identified. The anomaly detection algorithm will indeed work given that the geometric error in the model generates a difference in intensity between the prediction and the observation. A closest point approach was used to deal with the difference in resolution between the image and the terrain (the image resolution was much higher than that of the terrain) when marking invalid model points.

In order to evaluate the shape of the error function (Equation 5), the values of the two elevation points were systematically varied from -100m to 100m. The resulting plots for both textures in consideration are shown in Figure 3. Observe that both functions have a global minimum corresponding to the required corrections (-25m for point #1 and 25m for point #2). However, the error function displayed in Figure 3b contains many local minima. A regular texture having a high frequency pattern (relative to the DEM grid size) causes many sets of parameter values to explain the observed data. These results thus clearly demonstrate that the error function is mainly dependent on the texture patterns present in the imaging data. They also reveal that the accuracy of the solution depends on where the gradient descent is started, or in other words, on how large the model errors are initially.

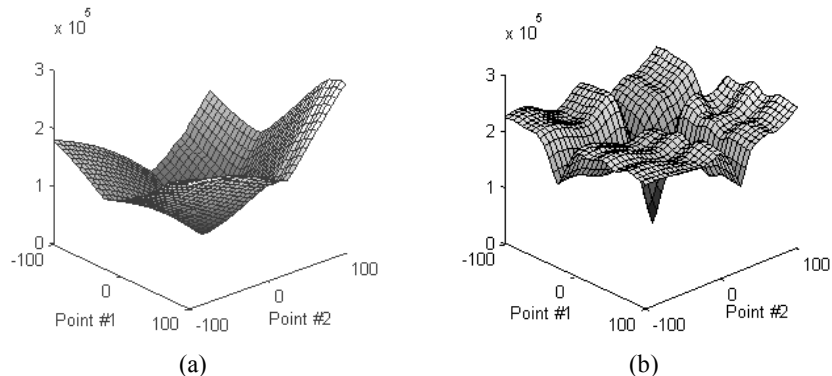(a)   (b)

Figure 3: Error function for (a) Coarse checkerboard textured terrain (b) Fine checkerboard textured terrain

Another test was performed by adding random errors to the flat terrain. These ranged from -25m to 25m (uniformly distributed) over a 4x4 cluster of adjacent elevation points. The gradient descent algorithm was run against ten different combinations of elevation errors for the two different checkerboard textures. Figure 4 uses solid lines to display the average elevation errors (and associated standard deviation) for the first ten iterations. In addition, cross marks are used to show the individual error data points (16 points x 10 tests = 160 points total), providing a better overview of the distribution. Observe that given the coarse checkerboard textured surface, the mean error was decreased from about 13m to 0.5m. Also notice that, although the finer texture provides many local minima to the error function, the algorithm still managed to decrease the error to less than 3 meters for 90% of points.
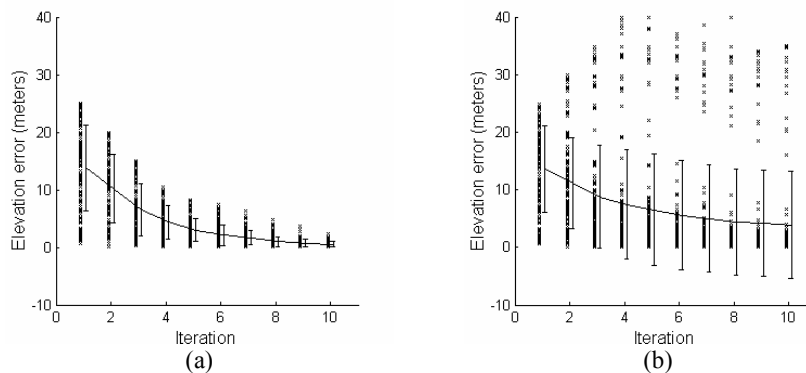

Figure 4: Elevation error versus iterations for (a) Coarse checkerboard textured terrain (b) Fine checkerboard textured terrain

3.2 Real Images

A standard digital video camera having a field-of-view of 45^ox30^o and a resolution of 720x480 was mounted inside a Bell 205 helicopter (Figure 5). The pilot flew at relatively low altitudes over a forested terrain in the Gatineau Hills region near Ottawa (Canada) according to a predetermined flight path. A differential GPS and inertial sensors (with 0.5m accuracy and less than 0.5^o rotation errors respectively) were used to log the position/attitude information of the aircraft.

Figure 5: (a) Helicopter used for data collection, (b) Camera attached to window frame

To register the model with the imaging data, the correspondence between the camera reference frame and the world reference frame (in which the model is expressed) had to

be established. The intrinsic parameters of the camera were determined by a standard calibration procedure. The location and orientation of the camera relative to the DGPS receiver and the attitude sensors were also measured. As the position calculated by the GPS was expressed in the same coordinate system as the DEM (i.e. latitude/longitude/altitude), no additional processing was necessary to register them. Initial alignment in attitude was performed manually. The yaw component was aligned by plotting the positional data given by the DGPS and by comparing the angle of the flight path with the measured yaw. Alignment in roll and pitch was performed using the horizon (present in both the synthetic and sensor imagery) as a landmark. Recent ESVS flight trials demonstrate that the use of such high precision navigation systems coupled with initial alignment procedures are sufficient to register the model with the imaging data [6]. Although the data was processed offline to update the model geometry, the required computations could be done in real-time onboard using a standard PC and optimized software.

The algorithm was first tested against a flyby scenario near the Ottawa airport. In that region, the model only consisted of a flat terrain (Figure 6a). The camera images before and after a combination of a forward and right motion of about 100m in each direction are shown in Figure 6b and 6c respectively (note that the left portion of the image is obscured by the window mounting and that the white object in the lower/center region is the air data boom attached to the aircraft). The corresponding predicted image (generated using Equation 1 and Equation 2) and difference map are shown in Figure 6d and Figure 6e. Note that locations where the algorithm has detected intensity differences between predictions and observations are shown in light tones.
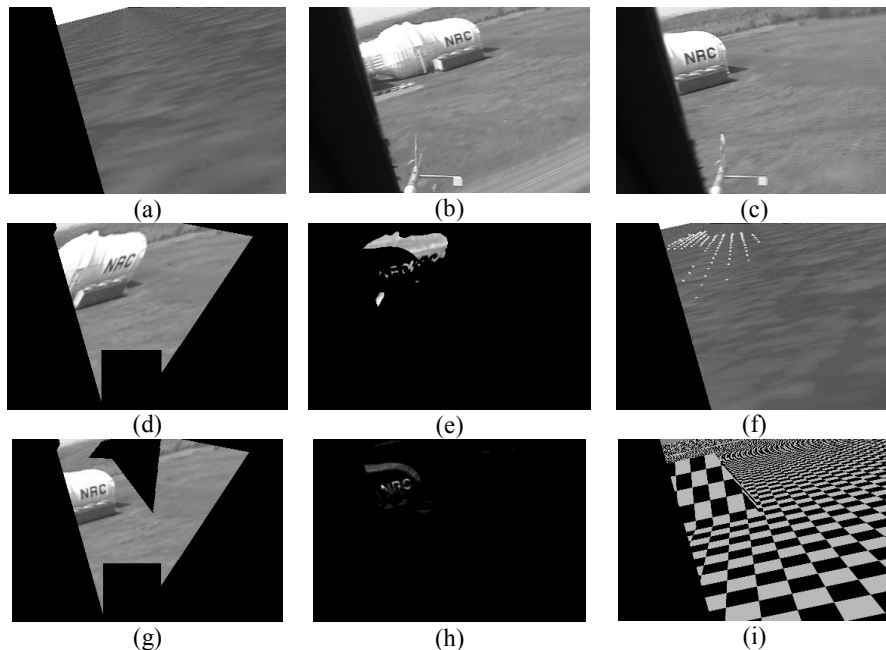

Figure 6: (a) Model view (b) Camera image (c) Camera image after forward/right motion (d) Predicted image (e) Difference map (f) Incorrect model points (g) Updated prediction (h) Updated difference map (i) Updated model

Model locations identified as incorrect by the algorithm are shown in white tones in Figure 6f. Notice that the building was indeed correctly identified as missing from the model. The model was then updated and the resulting updated prediction, updated difference map and updated model are shown in Figure 6g, 6h and 6i respectively. Note that a pyramid shaped structure is now present in the model, which roughly corresponds to the missing structure (in position and size). Given the relatively low resolution of the model, the solution represents the best approximation to the missing cylindrical structure.

A second test was performed during an approach to a mountain, which in the model (Figure 7a) has slightly incorrect geometry. The camera image before and after a forward motion of 250m are shown in Figure 7b and 7c respectively. The corresponding predicted image is presented in Figure 7d. Observe that the general appearance of the scene is correct. However, the difference map (Figure 7e) reveals that the mountain appearance is largely correct but that the right part of the ridge contains some errors (Figure 7e, 7f, 7h and 7i are close-up views of the region indicated by the dashed box in Figure 7c). Model locations identified as erroneous are shown in dark tones in Figure 7f. The minimization algorithm was then used to update the model and the resulting updated prediction, updated difference map and corrected model are shown in Figure 7g, 7h and 7i. The improvement to the geometry can be observed by comparing Figure 7f and 7i. In fact, this also shows up in the difference map since the number of detected anomalies has been reduced.
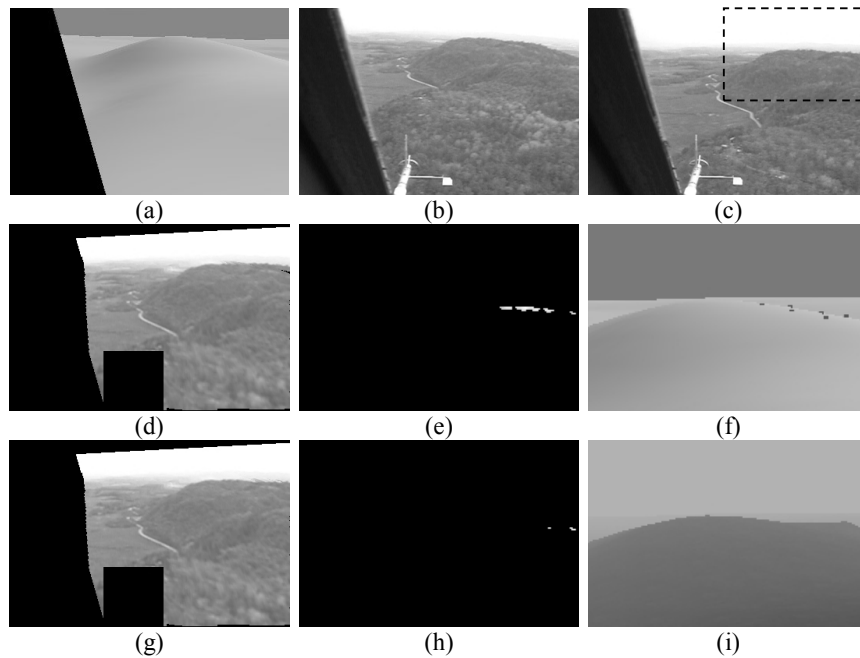

Figure 7: (a) Model view (b) Camera image (c) Camera after forward motion (d) Predicted image (e) Difference map (f) Incorrect model points (g) Updated prediction (h) Updated difference map (i) Updated model

4 Conclusions

A differential approach was presented to update a 3D model using a single imaging device. It differs from traditional approaches by quickly identifying geometric anomalies. This significantly reduces the complexity of the problem, making the technique applicable to real-time applications and large scale models. The approach was tested against synthetically generated data to update a surface having known errors. The algorithm was shown to be quite robust to different imaging conditions in both detecting and correcting geometric anomalies. In addition, a standard camera mounted on a helicopter was used to correct a model of a terrain. Results demonstrated that errors can be automatically identified and fixed. In particular, a missing man-made structure as well as terrain anomalies were correctly handled by the technique. Future work involves the development of a statistical technique to better cope with deviations from the brightness constancy assumption.

Acknowledgments

References

[1] Azarbayejani, A., Horowitz, B. and Pentland, A., "Recursive Estimation of Structure and Motion using Relative Orientation Constraints", *Conference on Computer Vision and Pattern Recognition*, 1993, pp. 294-299.

[2] Beardsley, P.A., Zisserman, A. and Murray, W., "Sequential Updating of Projective and Affine Structure from Motion", *International Journal of Computer Vision*, 1997, pp. 235-259.

[3] Hung, Y.-P., Cooper, D., Cernuschi-Frias, B., "Asymptotic Bayesian Surface Estimation Using an Image Sequence", *Int. Journal of Computer Vision*, vol. 6:2, 1991, pp. 106-132.

[4] Kumar, R., Sawhney, H.S. and Hanson, A.R., "3D Model Acquisition from Monocular Image Sequences", *Conference on Computer Vision and Pattern Recognition*, 1992, pp. 209-215.

[5] Kruk, R.V., Link, N.K., MacKay, W.J. and Jennings, S., "Enhanced and Synthetic Vision System for Helicopter Search and Rescue Mission Support", *American Helicopter Society 55th Annual Forum*, Montreal, Quebec (Canada), 1999.

[6] Link, N.K., Kruk, R.V., MacKay, D., Jennings, S., Craig, G., "Hybrid Enhanced and Synthetic Vision System Architecture for Rotorcraft Operations", *SPIE Conference on Enhanced and Synthetic Vision*, 2002.

[7] Matthies, L., Kanade T. and Szeliski R., "Kalman Filter-based Algorithms for Estimating Depth from Image Sequences", *International Journal of Computer Vision*, 3, 1989, pp. 209-238.

[8] Xu, G., Tsuji, S. and Asada, M., "A Motion Stereo Method Based on Coarse-to-Fine Control Strategy", *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. PAMI-9, no. 2, 1987, pp. 332-336.