Figure 11: **RADIUS** reference image.

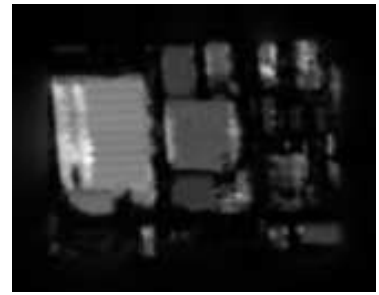Figure 12: **RADIUS** inspection image.

Figure 13: **RADIUS** magnitude of parallax flow map.
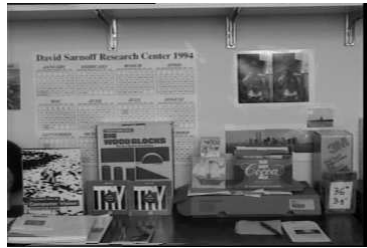

Figure 14: **Wall Frame 1**

Figure 15: **Wall Frame 2**

Figure 16: **Wall Frame 3**

Figure 17: **2D based Mosaic constructed from the 3 Wall frames.**

Figure 18: **3D corrected Mosaic constructed from the 3 Wall frames.**

side of the image is not visible. However in the forward view Figure 8 the carpet is visible in this area. This is an example of dis-occlusion.

4.2 RADIUS images

Figures 11,12 show the two original RADIUS images. These images are particularly difficult for area based matching techniques because of the large homogeneous areas. In order to register the ground plane using our direct method for estimating quadratic motion models, we provided initial masks for a few portions of the ground plane in the reference images. For these images, we have not yet obtained the ground truth data, thus we show our final result as the map of the magnitude of the parallax flow vectors (Figure 13). The parallax flow vector at any point is linearly related to the height of the point and inversely to the depth.

The parallax flow vectors in this case, were computed using the quasi-parametric flow algorithm described in Section 3. The optic flow algorithm performs very poorly on this sequence due to the presence of large areas of homogeneous patches. The quasi-parametric parallax estimation algorithm gives superior results because of the added constraint of finding matches only along an estimated parallax direction. The parallax direction is approximately parallel to the y-axis, thus computation of parallax flow along edges oriented parallel to the parallax direction is poor (aperture problem). These locations also have low confidence values. As can be noted from Figure (13), the overall parallax flow estimation results are fairly good. These results can be improved by incorporating feature based information.

4.3 3D corrected Mosaic: Wall sequence

The three original wall images Frames 1, 2 and 3 are shown in Figures 14, 15 and 16 respectively. Figure 17 shows a 2D mosaic built using only 2D affine transformations [Irani94a]. The 2D affine transformations used aligns the wall part of the images. However the objects sticking out of the wall exhibit parallax and are not registered by the affine. As a result in the 2D mosaic Figure 17, there are many ghost (duplicate) lines in the bottom half of the image. The reader's attention is drawn to the image regions corresponding to the duplicate lines in the boxes titled "TTY" and "Wooden blocks" in the left bottom and the the smearing on the book title information (e.g. Excel, Word, Getting Started) in the right bottom of Figure 17 respectively.

Figure 18 shows a 3D corrected mosaic image. In this case, using the technique described in Section 3.5, the objects sticking out of the wall are correctly positioned and no duplicate lines are visible. The 3D corrected mosaic was made by using Wall Frame 3 (Figure 16) as the final destination image. Using the parallax computed from Wall frames 1 and 2, Wall frame 2 was reprojected into the frame 3 coordinate system. This reprojected image was then merged with frame 3 to make the Mosaic image shown in Figure 18. Note in Figure 16 one can not see the boxes entitled "Wooden blocks" or "TTY". In the mosaic image, they however appear and are present in the geometrically correct locations.

References

[Bergen92] Bergen, J. R., Anandan, P., Hanna, K.J. and Hingorani, R., "Hiearchical Model-Based Motion Estimation," *Proceedings 2nd European Conference on Computer Vision-92, Springer-Verlag,* Santa Margherita Ligure, Italy, May 1992.

[Carlson90] Carlson, S. and Eklund, J., "Object detection using model based prediction and motion parallax," *Proceedings 1st European Conference on Computer Vision-90, Springer-Verlag,* 1990.

[Dutta90] Dutta, R. and M. Snyder, "Robustnesss of Correspondence-Based Structure from Motion," *Proceedings 3rd IEEE International Conference on Computer Vision,* Osaka, Japan, Dec. 1990.

[Faugeras92] Faugeras, O.D., "What can be seen in three dimensions with an uncalibrated stereo rig," *Proceedings 2nd European Conference on Computer Vision-92, Springer-Verlag,* Santa Margherita Ligure, Italy, May 1992.

[Hanna93] K. J. Hanna and N. Okamoto, "Combining stereo and motion analysis for direct estimation of scene structure", *Proceedings of the 4th International Conference on Computer Vision,* Berlin, Germany, May 1993.

[Hansen94] Hansen, M., Anandan, P., Dana, K., van der Wal, G., and Burt, P., "Real-time scene stabilization and mosaic construction", *in these proceedings.*

[Hartley93] Hartley, R. and Gupta, R., "Computing matched epipolar projections," *Proceedings IEEE Conference on Computer Vision and Pattern Recognition,* New York, 1993.

[Irani94a] Irani, M., Russo, B., and Peleg, S., "Recovery of ego-motion using image stabilization," *Proceedings IEEE Conference on Computer Vision and Pattern Recognition,* Seattle, June 1994.

[Irani94b] Irani, M., Hsu, S., and Anandan, P., "Video compression using mosaic representations", *David Sarnoff Research Center Internal Report,* Aug, 1994.

[Oliensis91] Oliensis, J. and J. I. Thomas, "Incorporating motion error in multi–frame structure from motion," *Proceedings IEEE Workshop on Visual Motion,* Princeton, N.J., Oct. 1991.

[Shashua94a] Shashua, A. and Navab, N., "Relative Affine Structure: Theory and Application to 3D reconstruction from perspective views," *Proceedings IEEE Conference on Computer Vision and Pattern Recognition,* Seattle, June 1994.

[Shashua94b] Shashua, A. and Toelg, S., "The quadric reference surface: Applications in registering views of complex 3D objects," *Proceedings Third European Conference on Computer Vision, Springer Verlag,* Vol. II, Stockholm, Sweden, May 1994.

[Sawhney94] Sawhney, H.S., "3D Geometry from planar parallax," *Proceedings IEEE Conference on Computer Vision and Pattern Recognition,* Seattle, June 1994.

[Szeliski94] Szeliski, R., "Recovering 3D shape and motion from image streams using nonlinear least squares," *Journal of Visual Communication and Image Representation,* Vol. 5, No. 1, March, pp. 10-28, 1994.

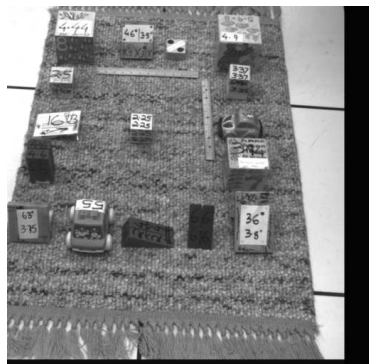[Wolberg90] Wolberg, G., "Digital Image Warping," *IEEE Computer Society Press Monograph,* ISBN 0-8186-8944-7, 1990.

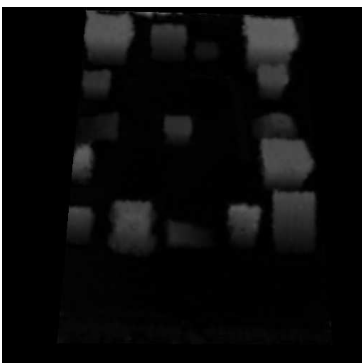Figure 5: **Oblique side aerial view: inspection image, Frame 3**
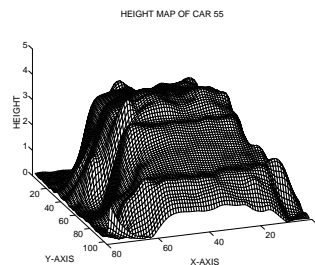

Figure 6: **Oblique side view height map.**
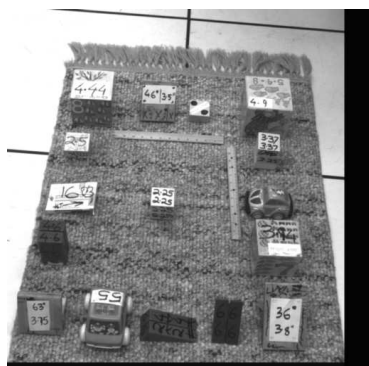

Figure 7: **Oblique aerial view: Wire frame drawing of car 55.**

Figure 8: **Oblique forward aerial view: reference image, Frame 1.**
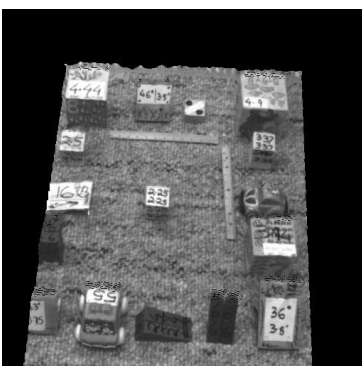

Figure 9: **New view, synthetic image of Oblique forward Frame 1.**
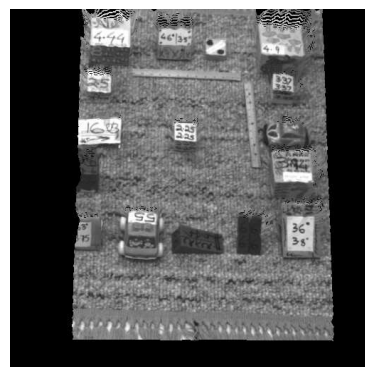

Figure 10: **New view, synthetic nadir image.**

height of a large majority of the points within 0.2 - 0.3 inches. The best structure from motion algorithms [Oliensis91] would estimate depth of points typically to an average accuracy of 3 % or so. which is the same as about 2 inches. The recovery of heights from this data would have about the same accuracy of 2 inches for camera height of 69 inches. Therefore, our results are an order of magnitude better in accuracy. Figure 7 shows a wire-frame reconstruction of the height map of the car labeled 55[2] on the image. The height map shown in Figure (6) was used to construct the wire-frame drawing. The two bumps seen on top of the car in the wire-frame drawing are physically there in the car, they correspond to two semi-transparent flashing signal sirens.

As described in Section 3.5, the parallax map computed after registering the ground plane is an affine invariant. We used the technique described in Section 3.5 to make synthetic images of the scene from novel view-points (see Figures 9 and 10). The user specifies these new views by giving six or more point correspondences between the source image and destination (or new) image. We then use the pre-computed parallax map and the correspondences to *transfer* by forward warping [Wolberg90] the source image to the destina-

tion image. Figures 9 and 10 were created by using the parallax map computed from the Oblique_side sequence frames 1 and 3 (Figures 4 & 5) and frame 1 was used as the source image. The synthetic image Figure 9 corresponds to frame 1 of the Oblique_forward sequence (Figure 8). We computed optic flow between the original frame 1 of the forward sequence and the synthetic frame 1. The average length of the flow vectors was 0.1 or 0.2 pixels. This signifies that we were able to predict the Oblique_forward sequence image 1 to an average accuracy of about 0.2 pixels using the two images from the Oblique_side sequence. Figure 10 shows an approximate nadir view synthesized from the two Oblique_side sequence images. The reader can notice that the perspective is much less in the nadir view and the vertical sides of the blocks are foreshortened much more. Finally, in both Figures 9 and 10 one can see finger-print like distortions on the upper side of the blocks. These distortions are actually dis-occluded areas. In the original images of the Oblique_side sequence, these areas are not visible. Therefore, ideally, in the synthetic views these areas should be totally void and appear black. However because of some leakage at occluding contours because of our area- based registration techniques, they appear as finger print patterns. Note in Figures 4 and 5, the carpet between the blocks 444 and 225 on the upper left

---

[2]The slant of the front window of the car is 55 degrees

for reconstruction of the video sequence at the receiver station.

To construct a 3D corrected mosaic we need at least three not completely overlapping images of a scene. We now describe the steps involved in building a 3D corrected mosaic using three image frames. The first step is to register the first two images to build a parallax map in the second frame's coordinate system. With this parallax map, we compute the 11 pose parameters which register the second image with the third image. The second image is then reprojected by forward warping to create a synthetic image taken from the third view-point. This synthetic image however contains image regions common to the first two images but not present in the third image. The final step to obtain the mosaic is to merge the synthetic third image with the actual third image. An example of this technique can be seen in the results section 4. Finally, to estimate the 11 pose parameters in the mosaic case, we do not use point correspondences. Rather, we directly register the second image with the first image using the estimated parallax map as an input. We again minimize equation (15) using equation (18) and this time only estimate the 11 pose parameters.

4 Results

4.1 Oblique block images

The oblique aerial images were taken in the laboratory by mounting a camera on a tripod and simulating an aerial fly path. In the first image pair (Oblique_forward sequence, Figure 8) the camera is moving forward approximately 4 inches in the y-z plane of the camera. In the second image pair (Oblique_side sequence, Figures 4,5) the camera is moving sideways approximately 4 inches, parallel to the x-axis of the image. The camera was tilted at approximately 35 degrees with respect to gravity and was 69 inches above the ground. The height of the blocks ranged from 1 inch to 4.9 inches. In Figure (4), the numbers written on top of the rectangular flat-top blocks are the height of the blocks in inches. Similarly for the triangular blocks, the slant of the block (degrees) was written on top of the block.

The ground plane in the images were registered and the heights computed using the algorithms described in the previous sections. To compute the height images, we used the height of 3 known points. In Table 1, the blocks from which the three points were selected are marked by the symbol "*". The height image for the Oblique_side pair can be seen in Figure 6, the results for the Oblique_forward pair are very similar. . The bright areas in the images correspond to areas with greater height. There is a a lot of noise around the borders of the images. This corresponds to border areas where there was no information available in one or the other image. These regions are detected by a low confidence value obtained from the optic flow estimation algorithm [Bergen92] and therefore do not pose a problem. Table 1 lists quantitative results in recovering heights. For each flat top in the image, the actual height of the block is provided and the average estimated height and the standard
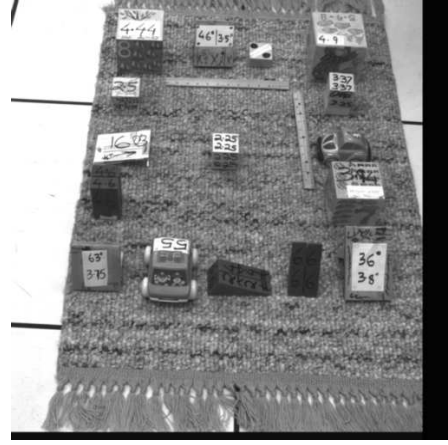

Figure 4: **Oblique side aerial view: reference image, Frame 1.**

Table 1: **Estimation of heights from oblique aerial views.**

BLO-CK	TRUE HT.	AVG. HT.	STD. DEV.	NO. PTS	CONF. THR.
OBLIQUE FORWARD SEQUENCE					
	inches	inches	inches		
*444	4.44	4.41	0.15	1794	0.0
49	4.90	4.62	0.31	2852	0.0
49	4.90	4.70	0.08	324	50.0
337	3.37	3.24	0.09	675	0.0
*394	3.94	3.89	0.20	2900	0.0
*CAR55	3.68	3.66	0.18	840	0.0
25	2.50	2.43	0.12	667	0.0
225	2.25	2.38	0.08	728	0.0
OBLIQUE SIDE SEQUENCE					
	inches	inches	inches		
*444	4.44	4.34	0.14	1386	0.0
*49	4.90	4.62	0.76	2360	0.0
49	4.90	4.86	0.04	233	50.0
337	3.37	3.33	0.10	550	0.0
*394	3.94	3.84	0.27	2365	0.0
CAR55	3.68	3.45	0.25	624	0.0
25	2.50	2.26	0.23	513	0.0
225	2.25	2.29	0.12	588	0.0

deviation of the heights of the points on that block are also listed. The number of points over which the average was computed and the confidence threshold used to select the points is also listed. For the results listed in the table, we selected all points belonging to a block by using a confidence threshold of 0. In the case of BLOCK49 the standard deviation of the obtained heights is comparatively large: 0.31 for the oblique_forward sequence and 0.76 for the oblique_side sequence. This large variation is due to regions of uniform intensity in the BLOCK49. If we select points on BLOCK49 with confidence value greater than 50, (see Table 1) that the standard deviation drops considerably. We are able to estimate the

of 8 equations for the other 6 unknowns: normal vector $\vec{N}$ and the rotation vector $\vec{\Omega}$. Since the translation used in equation (17) is T_2 while the translation we compute from the parallax flow vectors is T_1, we must invert the quadratic transformation defined by the parameters $p_1..p_8$ (or directly estimate the inverse quadratic transformation by inter-changing the two images during estimation.) We determine translation only up to a scaling factor and therefore need the height of at least one point to determine the heights of all other points in absolute coordinates.

Finally, for determining the vector $\vec{K}$ when focal length is unknown but image center is known we need the height of two points. This is a combination of the two previous cases. Since we do not know the focal length, we cannot use all eight of the quadratic parameters given in equation (17). However, the linear parameters $p_1..p_4$ do not depend on the focal length and we can use the equations pertaining to these parameters. On inspecting these equations, we note that for the case T_z equal to zero, we cannot determine the normal component N_z. However we can determine N_x and N_y up to a scaling factor. This is also true for the case when $T_z \neq 0$ since we do not know focal length, the translation vector, we recover is a scaled version of the vector $[fT_x \; fT_y \; T_z]$. Therefore, whether T_z is zero or not, we are able to determine only one component of the normal $\vec{N}$ and subsequently the vector $\vec{K}$. We use the height of at least two points and equation (23) to determine the vector $\vec{K}$ completely.

3.4 Multi-frame height recovery

In this section, we present a multi-frame extension to the two frame height estimation techniques. Our multi-frame technique is a batch method which uses one reference frame and one reference plane. Each frame in the multi-frame sequence is registered to the reference frame so that the reference plane is aligned. We then compute the residual parallax displacement vectors between every pair of frames. Hanna et.al. [Hanna93] have extended their quasi-parametric registration technique to operate in this multi-frame batch mode. Finally, we infer a height map from this sequence of estimated residual parallax displacement vectors.

Equation (22) can be rewritten in the form:

$$\frac{I_i}{S_i} = d_P(K_1 x + K_2 y + K_3) \qquad (24)$$

In the above equation, we note that the left hand side terms I_i and S_i each vary from frame to frame, however the right hand side is constant over the entire sequence. Therefore the ratio of I_i and S_i is an invariant across frames. It is similar to the relative affine structure measurement discussed by [Shashua94a] and [Sawhney94].

For a set of N inspection frames given the height of 3 points, we obtain a set of $3N$ linear equations and there are $(N + 3)$ unknowns : (N S_i terms and the vector $\vec{K}$). The equations are linear because we estimate $1/S_i$ instead of S_i. We can solve this system of equations to determine $N + 3$ unknowns and then use equation (24) to estimate the height at all other points. This method

corresponds to the case when we do not focal length or image center. We can similarly extend the techniques for the cases when center and/or focal length are known.

We experimented with this technique and observed that the performance depended chiefly only on the largest baseline present in the multiframe motion. The performance of the two frame technique was comparable when the frames with the largest baseline were used to do the computation. Therefore in the results section, we only present results for the two frame case.

3.5 View extrapolation and 3D corrected mosaic reconstruction

In this subsection, we briefly describe our technique for new view generation (the *transfer* problem in photogrammetry) and 3D corrected panoramic mosaic reconstruction. In the previous section the invariance of the quantity $\frac{I_i}{S_i}$ was used to combine image information from multiple views in order to obtain a single accurate height measurement at each pixel. The same property can be exploited to extrapolate from a given set of views to a new view for which the camera position and orientation (or equivalent information, such as 2D point correspondences) is known. Alternatively, the total motion vector at each point between a reference image and any desired camera view can be expressed as a combination of the plane and parallax components given in Equations (18), (17) and (19). Then, given a reference image and the parallax map $\gamma(x, y)$ the problem of extrapolating a new view reduces to that of pose estimation i.e. finding the 11 unknown motion parameters $(p_1, \ldots, p_8)$ and (T_{2x}, T_{2y}, T_{2z}). This information can be specified either in the form of at least 6 corresponding points or in the form of 3D translation and rotation parameters. The new view is generated by forward warping [Wolberg90] the source image using the estimated 11 pose parameters of the new frame and the prior computed parallax field.

One application of the view extrapolation process is the construction of image mosaics based on multiple views that are "corrected" for parallax. Recently there has been considerable interest in image mosaics that provide panoramic views of the scene constructed from multiple images. The mosaic image is constructed by aligning a number of 2D views of the same scene to each other and assembling them into a single image. For a detailed example of mosaic construction, see [Hansen94]. The interest in mosaics is partly motivated by applications of the mosaic representations to a number of traditional problems in visualization and video exploitation and compression [Irani94b]. These current techniques, however, rely on 2D planar surface registration which is adequate in situations when the 3D parallax effects in the scene are small. However, when applied to more complex scenes, the mosaics that are generated contain undesirable visual artifacts, and their compression efficiency is reduced. Using the parallax information, geometrically correct mosaic images can be reconstructed. The mosaic images and the associated parallax fields (signifying 3D information) comprise an efficient compression of the video sequence data. These can be transmitted along with the 11 associated pose parameters for each frame

center. The reader is referred to [Bergen92] for further details about this registration technique.

The parallax vectors and the direction of translation are simultaneously estimated using the quasi-parametric technique described in [Hanna93].The quasi parametric technique is generally more accurate than using optic flow, but requires an initial estimate for translation. If needed, an initial estimate of the translation direction can be obtained by using the optical flow obtained by using the technique also described in [Bergen92].

3.2 Simultaneous registration

For the simultaneous approach, we express the total motion vector of a point as the sum of the motion vector due to the planar surface (u_p, v_p) (as represented above) and the residual parallax motion vector (u_r, v_r). This residual vector can be represented as:

$$
\begin{aligned}
(u, v) &= (u_p, v_p) + (u_r, v_r) & (18) \\
u_r(\mathbf{x}) &= \gamma(fT_{2x} - xT_{2z}) \\
v_r(\mathbf{x}) &= \gamma(fT_{2y} - yT_{2z}) & (19)
\end{aligned}
$$

where $\gamma = H/P_z T_\perp$, and as in Section 2, H is the perpendicular distance of the point of interest from the plane and P_z is its depth. In this representation, however, the third quantity $T_\perp$ is the perpendicular distance from the center of the *first* camera to the plane. The above expressions for (u, v) from equation (18) are substituted into equation 15 to obtain the complete objective function. We use a variation of the quasi-parametric ego motion estimation algorithm described in [Hanna93] to simultaneously estimate the 8 planar surface parameters $(p_1 \ldots p_8)$, the three translational motion components (T_{2x}, T_{2y}, T_{2z}) and the parallax magnitude field γ. This approach does not require the presence of an actual 3D plane in the scene. The planar registration parameters obtained correspond to a virtual 3D plane which gives rise to the smallest parallax field (the average plane of the 3D scene). Alternatively, we can set the parameters so that average frontal plane is obtained.

3.3 Height recovery based on two frames

The parallax flow vectors vary directly with height and inversely with depth. To factor out the depth and get the height alone, we use a characteristic property of aerial view images. For aerial view images, the depth of the plane is typically much greater than the height of objects on the ground. In nadir aerial images a weak perspective projection approximation can be used, since the depth of all points is approximately the same. Whereas, in an oblique view, there can be considerable depth variation across the image. However, for any single point in an oblique aerial image, the depth of that image point is essentially the same as the depth of a possibly virtual 3D point obtained by extending the line of sight ray and intersecting it with the ground plane. Therefore, we can factor out the depth in the parallax equations (4,5) by estimating the equation of the ground plane (we actually only need to estimate the normal of the plane up to a scale factor).

We use the magnitude of the displacement vectors to infer the magnitude of the height and the direction of the flow vectors to infer the sign of the height. For the sequential approach, the magnitude of the displacement vector $\gamma_2 = \sqrt{(x_w - x_1)^2 + (y_w - y_1)^2}$ for the case where the translation is parallel to the image plane is given by the equation:

$$
\gamma_2 = \frac{fH}{P_z T_{2\perp}} \sqrt{T_{1x}^2 + T_{1y}^2} \tag{20}
$$

The magnitude of the displacement vector for the case where $T_{1z} \neq 0$ is given by:

$$
\gamma_2 = \frac{T_{1z}H}{P_z T_{2\perp}} \gamma_F \tag{21}
$$

where $\gamma_F = \sqrt{(x_w - x_F)^2 + (y_w - y_F)^2}$ is the distance of the point (x_w, y_w) from the FOE.

In the case of the simultaneous approach, these equations have to be modified by substituting the components of T_2 for the corresponding ones of T_1, wherever they occur.

Noting that $N_2^T \vec{p_1} = \frac{f}{P_z}$ and using the aerial view property equations (21) and 20 can be written in the form:

$$
I = S_2 H N_2^T \vec{p_1} \tag{22}
$$

where $I = \frac{\gamma_2}{\gamma_F}$ is an image based measurement. S is a proportionality factor which depends solely on the translation vector T_1 and the distance $T_{2\perp}$. The above equation can be rewritten as:

$$
H = \frac{I}{(K_1 x + K_2 y + K_3)} \tag{23}
$$

where $\vec{K}$ is an unknown vector and its components are given by $K_1 = SN_{2x}$, $K_2 = SN_{2x}$ and $K_3 = fSN_{2z}$ The height H of any image point can be computed using equation (23). In the case of the simultaneous approach, these equations have to be modified by substituting γ for γ_2, T_2 for T_1, and $T_\perp$ for $T_{2\perp}$.

Intrinsic camera parameters such as focal length and image center can be calibrated for using standard techniques. However, there are many applications where it is not possible to do this calibration nor obtain these intrinsic parameters apriori. We therefore develop three different ways to estimate the vector $\vec{K}$ based on whether or not we know the focal length and image center.

If the focal length and center are both unknown, height of at least three points are known, these can be used together with equation (23) to linearly estimate the vector $\vec{K}$. We do not need to know the image center, because the unknown offset simply loads the third component of the vector $\vec{K}$ in Equation 23. We then use this vector $\vec{K}$, again with equation (23) to determine the height at any other point, again in a linear fashion. The experimental results presented in section 4 use this case.

If focal length and center are known, we can infer the normal of the plane by using equation (17). This equation relates the quadratic registration parameters to the translation, rotation and normal of the plane, but the translation direction is computed during the quasi-parametric residual estimation. The translation direction together with equation (17) gives us a linear set
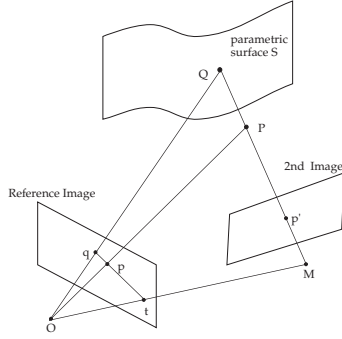
Figure 1: Residual parallax field after surface alignment is epipolar

Figure 2: Residual parallax magnitude when parametric surface is aligned

Figure 3: Residual parallax magnitude when plane is aligned

After some algebraic manipulation, it can be shown that the parallax vector is,

$$\delta\mathbf{u} = \mathbf{q} - \mathbf{p} = \frac{T_z(Q_z - P_z)}{Q_z(P_z - T_z)}(\mathbf{p} - \mathbf{t_1}) \qquad (10)$$

(b). For the case when parametric surface is a plane, refer to Figure 3. In this case, instead of the substitution given above in Equation 7, we can use

$$\frac{|\vec{Q'}|}{|\vec{P'}|} = \frac{T_{2\perp}}{T_{2\perp} - H} \qquad (11)$$

where H and $T_{2\perp}$ are the perpendicular distances from the plane to the points P and M respectively. Going through the same steps as in the case of the derivation of the residual parallax with respect to a parametric surface, but using Equation 11 everywhere instead of Equation 7, we obtain the result

$$\delta\mathbf{u} = \frac{HT_z}{T_{2\perp}P_z - HT_z}(\mathbf{p} - \mathbf{t_1}) \qquad (12)$$

Upon some further algebraic manipulation, we can easily obtain

$$\delta\mathbf{u} = \mathbf{q} - \mathbf{p} = \frac{HT_z}{T_{2\perp}P_z}(\mathbf{q} - \mathbf{t_1}) \qquad (13)$$

We provide this last version of the parallax result because it is algebraically simpler than the previous one. For the same reason, this was the form that was used in our height interpretation algorithms described in Section 4.

Note that the above derivation implicitly assumes that $T_z \neq 0$. (Specifically, this assumption was made during the derivation of Equation 9).

From Equation 13, it is clear that if $T_z = 0$, the parallax vector is given by:

$$\delta\mathbf{u} = -\frac{fH}{T_{2\perp}P_z}\vec{T_1} \qquad (14)$$

3 Registration and Interpretation

We have developed two techniques for recovering the planar and parallax motions. The first technique takes a *sequential registration* approach, in which the plane is first registered using a 8 parameter quadratic transformation. The residua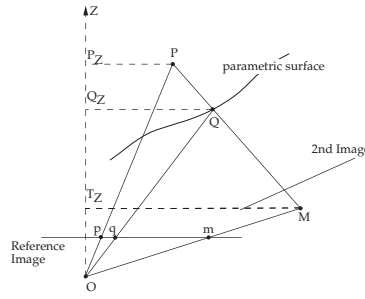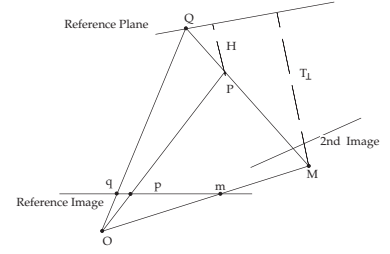l parallax is then estimated as a separate step. The second technique simultaneously estimates the planar and parallax motion components, and is hence referred to as a *simultaneous registration*.

3.1 Sequential registration

To register a plane, we use the hierarchical direct registration technique described in [Bergen92] with a planar surface flow field model. This technique first constructs a Laplacian pyramid from each of the two input images, and then estimates the motion parameters in a coarse-fine manner. Within each level the Sum of squared difference (SSD) measure integrated over user selected regions of interest is used as a match measure. This measure is minimized with respect to the quadratic flow field parameters. The SSD error measure for estimating the flow field within a region is:

$$E(\{\mathbf{u}\}) = \sum_{\mathbf{x}} (I(\mathbf{x},t) - I(\mathbf{x} - \mathbf{u}(\mathbf{x}), t-1))^2 \qquad (15)$$

where $\mathbf{x} = (x,y)$ denotes the spatial image position of a point, I the (Laplacian pyramid) image intensity and $\mathbf{u}(\mathbf{x}) = (u(x,y), v(x,y))$ denotes the image velocity at that point, and the sum is computed over all the points within the region and $\{\mathbf{u}\}$ is used to denote the entire flow field within that region. The motion field of a planar surface can be represented as:

$$
\begin{aligned}
u_p(\mathbf{x}) &= p_1 x + p_2 y + p_5 + p_7 x^2 + p_8 xy \\
v_p(\mathbf{x}) &= p_3 x + p_4 y + p_6 + p_7 xy + p_8 y^2
\end{aligned} \qquad (16)
$$

where

$$
\begin{bmatrix} p_1 \\ p_2 \\ p_3 \\ p_4 \\ p_5 \\ p_6 \\ p_7 \\ p_8 \end{bmatrix} = \begin{bmatrix} N_{2x}T_{2x} - N_{2z}T_{2z} \\ N_{2y}T_{2x} - \Omega_z \\ N_{2x}T_{2y} + \Omega_z \\ N_{2y}T_{2y} - N_{2z}T_{2z} \\ f(N_{2z}T_{2x} + \Omega_y) \\ f(N_{2z}T_{2y} - \Omega_x) \\ \frac{1}{f}(\Omega_y - N_{2x}T_{2z}) \\ \frac{1}{f}(-\Omega_x - N_{2y}T_{2z}) \end{bmatrix} \qquad (17)
$$

In the above equation, (T_{2x}, T_{2y}, T_{2z}) denotes the translation vector between the cameras, $(\Omega_x, \Omega_y, \Omega_z)$ denotes the angular-velocity vector, and (N_{2x}, N_{2y}, N_{2z}) denotes the normal vector to the planar surface from the camera

frame information to obtain accurate heights, and to extrapolate new views from a given set of views (i.e., in photogrammetric terms, to achieve "transfer"). We use the view extrapolation process to construct a panoramic mosaic image by combining multiple views that is accurate in terms of 3D positions of surfaces. In section 4, we present real data height estimation results, which are an order of magnitude better than the results obtained using the best traditional (perspective projection) structure for motion algorithms [Oliensis91]. We also present results for new view generation or view extrapolation and 3D corrected panoramic mosaic reconstructions.

2   Parametric surfaces

Consider two camera views, one denoted as the "reference" camera, and the other the "inspection" camera. A 3D point $\vec{P}_1$ in the reference camera coordinate system gets mapped to the 3D point $\vec{P}_2$ in the inspection camera coordinate system by a rigid body transformation:

$$\vec{P}_2 = R(\vec{P}_1) + \vec{T}_2 = R(\vec{P}_1 - \vec{T}_1) \qquad (1)$$

The mapping can be represented by a rotation (R) followed by a translation ($\vec{T}_2$) or by a translation ($\vec{T}_1$) followed by a rotation (R). With perspective projection, the image coordinates (x,y) of a projected point P are given by the vector $\vec{p}$:

$$\vec{p} = \begin{bmatrix} x \\ y \\ f \end{bmatrix} = \frac{f}{P_z} P \qquad (2)$$

where f is the focal length.

Theorem

I. Given two views of a scene (possibly from two distinct uncalibrated cameras), if the image motion corresponding to an arbitrary parametric surface is compensated (by applying an appropriate 2D parametric transformation to one of the images) then the residual parallax displacement field on the reference image plane is an epipolar field.

II (a). Let P be a point not on the surface that is registered, and let p be its image in the reference view (see Figure 2). Let T_1 denote the baseline vector between the cameras and Q be the point where the ray connecting P to the the second camera center intersects the surface. Then the residual parallax displacement $\delta\mathbf{u}$ at image location $\mathbf{p}$ can be shown to be

$$\delta\mathbf{u} = \mathbf{q} - \mathbf{p} = \frac{T_z(Q_z - P_z)}{Q_z(P_z - T_z)}(\mathbf{p} - \mathbf{t_1}) \qquad (3)$$

where P_z and Q_z denote the depths of points P and Q, T_z is the z component of translation vector T_1 (assumed non-zero), and $\mathbf{t_1}$ denotes the epipole corresponding to T_1.

II (b). If the surface that is aligned is a plane, then the residual parallax displacement simplifies in the case of $T_z \neq 0$ to:

$$\delta\mathbf{u} = \frac{HT_z(\mathbf{p} - \mathbf{t_1})}{T_{2\perp}P_z - HT_z} = \frac{HT_z}{T_{2\perp}P_z}(\mathbf{q} - \mathbf{t_1}) \qquad (4)$$

and in the case of $T_z = 0$ to:

$$\delta\mathbf{u} = -\frac{fH}{T_{2\perp}P_z}\vec{T}_1 \qquad (5)$$

where H is the perpendicular distance from the point P to the reference plane, $T_{2\perp}$ is the perpendicular distance between the second camera center M and the reference plane (see Figure 3).

Proof: Part I Referring to Figure 1, let S denote the surface of interest, P an environmental point not on S, and O and M the two camera centers. The image of P on the reference view is p. Let the ray MP intersect the surface S at Q. The warping process would warp p', the image of P on the second image to q, the image of Q on the reference image, since the transformation applied aligns all points on the surface S. Therefore, the residual parallax vector is pq, which is the image of the line PQ. It is immediately obvious from the figure that pq lies on the plane OMP, which is the epipolar plane passing through p. Since the above argument is true for any point P, the parallax displacement field is an epipolar field.

It should be noted that this theorem (and the proof) does not deal with the problem of actually determining the transformation required to align the parametric surface. Nor is it concerned about the existence of such a transformation–in fact, the existence is presumed by the theorem. These issues, however, are significant for the practical application of the theorem. Although a complete discussion of the existence and the estimation of the transformation for aligning general parametric surfaces is outside the scope of this paper, the following observations can be made. First, the case of a planar reference surface has been solved by us and others (e.g., [Sawhney94]) and is also handled in this paper. The case of a general quadratic surface is addressed in [Shashua94b]. In the most general case, we briefly note that a numerical solution can be used based on a minimal number of point correspondences and surface interpolation.

Proof: Part II (a). Referring to Figure 2, let $\vec{P}$ denote the vector OP, $\vec{Q}$ denote the vector OQ, and $\vec{T}$ denote the vector OM. Also, let $\vec{P}'$ and $\vec{Q}'$ denote the vectors MP and MQ.

From Figure 2 we see that

$$\vec{Q} = \vec{T} + \vec{P}'\frac{|\vec{Q}'|}{|\vec{P}'|} \qquad (6)$$

Using the fact that $\vec{P}' = \vec{P} - \vec{T}$ and taking the z component of both sides of the above equation, we get

$$\frac{|\vec{Q}'|}{|\vec{P}'|} = \frac{Q_z - T_z}{P_z - T_z} \qquad (7)$$

Substituting this back into Equation 6,

$$\vec{Q} = \vec{T}_1 + (\vec{P} - \vec{T}_1)\frac{Q_z - T_z}{P_z - T_z}. \qquad (8)$$

Since $\mathbf{q} = f\vec{Q}/Q_z$, $\mathbf{p} = f\vec{P}/P_z$, and (assuming $T_z \neq 0$) $\mathbf{t_1} = f\vec{T}_1/T_z$ we get

$$\begin{aligned} \mathbf{q} &= \frac{f}{Q_z}\left(\vec{T}_1 + \frac{Q_z - T_z}{P_z - T_z}(\vec{P} - \vec{T}_1)\right) \\ &= \left(\frac{T_z}{Q_z} - \frac{T_z(Q_z - T_z)}{P_z - T_z}\right)\mathbf{t_1} + \frac{P_z(Q_z - T_z)}{Q_z(P_z - T_z)}\mathbf{p} \end{aligned} \quad (9)$$

Shape recovery from multiple views: a parallax based approach

R. Kumar   P. Anandan   K. Hanna[*]

David Sarnoff Research Center,
CN5300, Princeton, NJ-08530, U.S.A.

Abstract

Given two arbitrary views of a scene under central projection, if the motion of points on a parametric surface is compensated, the residual parallax displacement field on the reference image is an epipolar field. The parallax magnitude at a point, after suitable scaling, is an affine invariant; if the surface aligned is a plane, it is directly proportional to the height of the point from the plane and inversely proportional to its depth from the camera. We exploit the above result to infer 3D height information from *oblique* aerial 2D images. We use direct methods to register the aerial images, develop methods to infer height information under the following three conditions: (i) focal length and image center are both known, (ii) only the focal length is known, and (iii) both are unknown. We use the invariance property of the scaled parallax magnitudes to combine multiple frame information to obtain accurate heights, and to extrapolate new views from a given set of views (i.e., in photogrammetric terms, to achieve "transfer"). We use the view extrapolation process to construct a panoramic mosaic image by combining multiple views that is accurate in terms of 3D positions of surfaces.

1 Introduction

Traditional methods in motion analysis have expressed the image motion of rigid bodies as a sum of two image displacement fields: a rotation field and a epipolar (translation) field. A majority of these methods have treated the problems of motion measurement and estimation as independent problems. This approach has been plagued with problems in getting robust and accurate estimates [Dutta90]. Errors in the estimate of the relative orientation (especially the rotation component e.g. when the motion is parallel to the image plane) between the cameras severely affects the depth estimates.

In this paper, we develop an alternative approach which is based on decomposing the motion field into the image motion of a parametric surface and a residual parallax field. The motion of the surface can be expressed as a parametric motion field and is estimated using a direct technique [Bergen92]. The direct approach provides a more accurate alignment of the surface than using pure "bottom-up" flow-fields. The residual parallax field is an epipolar field (see Section 2) and is quasi-parametric; it is also estimated using a direct method. The parallax magnitudes, when normalized to cancel out a scale factor that depends on the magnitude of camera translation are relative affine invariants[1] [Shashua94a] and [Sawhney94]. From the parallax maps, a projective 3D reconstruction [Faugeras92] of the scene can be made. This projective 3D reconstruction is related to the euclidean 3D construction by a collineation (4×4) matrix.

Our work is related to the recent work using projective geometry of [Hartley93, Sawhney94, Shashua94a, Szeliski94] and motion stabilization of [Irani94a]. It also extends the work of [Carlson90] in obstacle detection by removing the requirement placed by their technique that the camera translation should be parallel to the reference plane. It differs from previous work in several ways: First, our derivation of the surface+parallax decomposition of image motion is not restricted only to planar surfaces. Second, we use "direct" estimation methods for aligning the plane and estimating residual parallax, thereby avoiding many of the problems introduced by bottom-up optical flow estimation. Finally, our approach does not require the explicit estimation of rotation and translation to estimate structure. Instead, the rotational component of motion is cancelled as a part of the surface registration process. This has the advantage that the instabilities introduced during the computation of rotation matrices from the noisy flow fields is removed. Our approach improves both the motion measurement or correspondence and 3D motion interpretation stages of dynamic image analysis.

We exploit the above results to infer 3D height information from *oblique* aerial 2D images under the following three conditions: (i) focal length and image center are known, (ii) focal length is unknown but image center is known and (iii) focal length and image center are both unknown. We achieve this by applying direct registration methods to align the planar reference surfaces, and recover and interpret the parallax relative to that surface (see Section 3). We use the invariance property of the scaled parallax magnitudes to combine multiple

---

[*]Readers may contact first author at Email: kumar@sarnoff.com

[1]Relative affine invariants are invariants to an affine deformation of the intrinsic camera parameters (focal length, image center etc.) and are relative to a reference plane in the scene.