Simplifying Motion and Structure Analysis using Planar Parallax and Image Warping

Harpreet S. Sawhney

Machine Vision Group
IBM Almaden Research Center, K54
650 Harry Road
San Jose, CA 95120

FAX: 408-927-4090
Ph.: 408-927-1799
Net: sawhney@almaden.ibm.com

Keywords: Motion parallax, Image motion analysis, Structure from multiple views, 3D Geometry.

Abstract

Robust 3D motion and structure computation and segmentation has been the subject of an enormous body of work in reconstructive vision. For linear approximations to perspective projection (weak/para perspective) [20, 24, 30, 31], and for the case of image velocities [19], elegant linear methods have been devised for robust estimation. For reconstruction under arbitrary view transformations, linear projective methods [15, 16, 28] using *point correspondences* have been suggested.

In this paper, we present a formulation for 3D motion and structure analysis using motion parallax defined with respect to an arbitrary plane in the environment. It is shown that if an image coordinate system is warped using plane projective transformation with respect to a reference view, the residual image motion is dependent only on the epipoles and has a simple relation to the 3D structure. Our computational scheme avoids point/line correspondence and is based on hierarchical estimation and image warping [10] working directly with spatio-temporal image intensities. Results on real images demonstrate how this analysis simplifies ego and multiple motion analysis, and stable scene-centered 3D reconstruction.

Submitted to the International Conference on Pattern Recognition, 1994. A preliminary version to appear in CVPR'94

Simplifying Motion and Structure Analysis using Planar Parallax and Image Warping

April 20, 1994

Keywords: Structure from multiple views, Motion parallax, 3D Geometry.

Abstract

Robust 3D motion and structure computation and segmentation has been the subject of an enormous body of work in reconstructive vision. For linear approximations to perspective projection (weak/para perspective) [20, 24, 30, 31], and for the case of image velocities [19], elegant linear methods have been devised for robust estimation. For reconstruction under arbitrary view transformations, linear projective methods [15, 16, 28] using *point correspondences* have been suggested.

In this paper, we present a formulation for 3D motion and structure analysis using motion parallax defined with respect to an arbitrary plane in the environment. It is shown that if an image coordinate system is warped using plane projective transformation with respect to a reference view, the residual image motion is dependent only on the epipoles and has a simple relation to the 3D structure. Our computational scheme avoids point/line correspondence and is based on hierarchical estimation and image warping [10] working directly with spatio-temporal image intensities. Results on real images demonstrate how this analysis simplifies ego and multiple motion analysis, and stable scene-centered 3D reconstruction.

SUMMARY

1. Which conference ?

   Computer Vision and Image Processing.

2. What is the paper about ?

   A direct method of image warping and flow computation based on planar parallax that simplifies the analysis of ego and multiple object motion, and motion and structure reconstruction.

3. What is the original contribution of this work ?

   A new derivation of image warping corresponding to a planar projective transformation, and of the residual flow after warping. Instead of using point correspondences, our method shows that warping of the image coordinate system and image intensities corresponding to a planar transformation leads to a simple form of the flow dependent only on the epipoles and the 3D geometry. It is demonstrated how this simplifies multiple motion analysis and 3D reconstruction.

4. Does the paper mainly describe an application, and should be reviewed by the applications committee ?

   No.

1 Introduction

Robust 3D motion and structure computation and segmentation has remained an important problem in motion vision. Under perspective projection, and Euclidean reconstruction with arbitrary camera viewpoints, the problem of computing rotation and translation, and the environmental structure lead to non-linear optimization problems. Many solution methods have been proposed in this situation [8, 18, 29], but the inherent ambiguities and instabilities [1, 5] associated with the problem can lead to very unreliable results in real world scenarios [6, 27]. Moreover these methods use *point correspondences*; a messy problem in its own right.

There has been a quest to invent linear methods for motion and structure analysis problems. For linear approximations to perspective projection (weak/para perspective) [20, 24, 30, 31], and for the case of image velocities [19], elegant linear methods have been devised for robust estimation. Similarly, for reconstruction under arbitrary view transformations and uncalibrated cameras, novel linear projective methods [15, 16, 28] using *point correspondences* have been suggested.

In this paper, we present a linear method for motion analysis that does not depend on point correspondences. The linearity of the method comes from using the idea of *planar parallax*, that is residual motion after the motion of an environmental plane has been compensated for. Point correspondences are avoided by formulating the analysis in terms of image warping corresponding to the image coordinate transformation for an environmental plane's motion. It is shown that if an image coordinate system is warped using plane projective transformation with respect to a reference view, the residual image motion is dependent only on the epipoles and has a simple relation to the 3D structure. The computational scheme avoids point/line correspondence and is based on hierarchical estimation and image warping [10] working directly with spatio-temporal image intensities. Results on real images demonstrate how this analysis simplifies ego and multiple motion analysis, and stable scene-centered 3D reconstruction.

Most methods for multiple motion segmentation [2, 3, 11] rely on either simple parametric models (affine/planar), or on the smoothness of flow in the image plane. No 3D motion and structure constraints are brought to bear on the process. These methods may fail to detect self motion, or may over-segment a 3D object even for simple cases of translation in depth because in such cases affine and plane projective transformations *alone* do not model image motion well. The planar parallax motion representation presented here has the potential for easy exploitation of the 3D motion and structure constraints in segmentation and reconstruction tasks. For 3D reconstruction, we show that the planar parallax method represents the 3D geometry of the scene in an intrinsic coordinate system associated with the selected environmental plane. Both Euclidean and affine reconstruction is quite simply obtained in our method.

2 Motion Parallax and Warping

The essential principle behind planar motion parallax is that if an image coordinate system is warped so that an environmental plane is fixated between this image and a reference image, that is the plane's image motion is nulled, then the residual image motion can be factorized into a component that depends only on the non–planar shape, and another that depends only on the epipoles (i.e. only on camera displacements and not rotation). This is called *planar motion parallax*. It is a specific instance of the well-known notion of motion parallax. For general motion parallax, it can be shown [23, 25] that if two distinct points in 3D project to the same point in an image (that is are along the same view ray), then the difference in their image displacements due to a change in the viewpoint (that is the projection, in another view, of the vector joining the two) depends only on the 3D translation (perspective) or rotation (weak perspective) between the views and the relative depth of the 3D points. However, using the general motion parallax may not be practical because finding coincident points in a view is hard; for an opaque world occluding boundaries represent such points but these may be hard to detect and computing their image motion may be hard too.

The use of planar parallax instead is practical. Many cultural and other scenes naturally contain a planar surface which can serve as a coordinate system to define the structure of the rest of the scene. For the problem of obstacle detection, Carlsson and Eklundh [4] and Enkelmann [7] used the specific constraint on image flow for a *ground plane*. The camera motion was modeled as the motion on the ground plane. In contrast, our method can use any arbitrary plane in the environment, (e.g. walls, ceilings, floor etc.) and is applicable for general rigid motion.

Figure 1 is a geometric depiction of planar parallax. Given $\mathbf{p}$ and $\mathbf{p}'$, the projections of a 3D point in two views, and given a reference plane S, if the planar motion transformation can be computed, then a virtual projection, $\mathbf{p}^{\mathbf{w}}$, corresponding to the point of intersection of the ray $\mathbf{p}'$ and S can be computed. Alternatively the primed image coordinates ($\mathbf{p}'$) can be warped to create an image of points $\mathbf{p}^{\mathbf{w}}$. Then the difference between $\mathbf{p}^{\mathbf{w}}$ and $\mathbf{p}$ in the reference view is the planar parallax motion. It is clear from the figure that these parallax vectors are all oriented towards the epipole $\mathbf{t}$ (the point of intersection of the line connecting the two camera centers, $\mathbf{OO}'$, with the reference image plane).

3 Planar Parallax under Perspective Projection

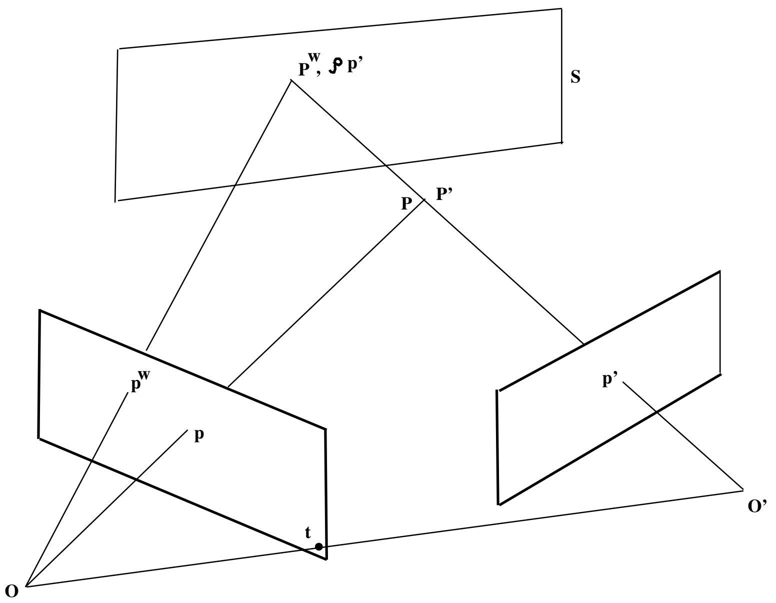

Figure 1: **Two-view Planar Parallax.**

Perspective projection is a model of projection that accounts for the pin-hole projection of a scene onto an image plane. The linear internal calibration parameters can be modeled as an affine transformation of the image coordinates. This transformation can be combined with an arbitrary rigid transformation due to camera/object motion by modeling the 3D transformation as an affine transformation. Thus, in the following formulation, the 3D transformation between any two time instants is modeled as a 3D affine transformation; a 3D rigid transformation is a special case.

A reference view and any other arbitrary view are chosen to present the motion parallax analysis. The 3D coordinate transformation between the primed coordinates, $\mathbf{P'}$, in view 2 and the reference coordinates, $\mathbf{P}$, in view 1 is written as an arbitrary 3D affine transformation:

$$\mathbf{P}' = \mathbf{A}'\mathbf{P} + \mathbf{T}'. \tag{1}$$

Let $\mathbf{N^T P} = d$ represent a plane in the reference coordinate system. Substituting this in the above equation, one can write the *plane projective transformation* as [17]:

$$\mathbf{P}' \quad \approx \quad [\mathbf{A}' + \mathbf{T}'\mathbf{N}^T/d]\mathbf{P}, \tag{2}$$

where $\approx$ denotes equality up to an unknown arbitrary scale. Note that this represents the general 8–parameter projective relationship for plane–to–plane projection.

As mentioned in the introduction, we are interested in developing a direct method to compensate for the image transformation corresponding to the above planar transformation. That is, the second image is to be registered with respect to the reference image by a warping transformation corresponding to the plane projective transformation. Thus, it is necessary to express the warping transformation for the coordinates of the second image. From equation 2, the warped projective coordinates of the second image are

$$\mathbf{P^w} \quad \approx \quad [\mathbf{A}' + \mathbf{T}'\mathbf{N}^T/d]^{-1}\mathbf{P}' \quad \approx \quad [\mathbf{I} - \mathbf{T}\mathbf{N}^T/d]^{-1}\mathbf{A}'^{-1}\mathbf{P}', \tag{3}$$

$\mathbf{T} = -\mathbf{A}'^{-1}\mathbf{T}'$ is the displacement of the second frame's origin in the reference coordinates.

Using the identity $[\mathbf{I} + \mathbf{u}\mathbf{v}^T]^{-1} = [\mathbf{I} - \alpha\mathbf{u}\mathbf{v}^T]$ (see [13]), where $\alpha = (1/(1 + \mathbf{v}^T\mathbf{u}))$, $(\mathbf{v}^T\mathbf{u} \neq -1)$, the above relationship can be written as the following projective transformation:

$$\mathbf{P^w} \quad \approx \quad [\mathbf{I} + \beta\mathbf{T}\mathbf{N}^T/d]\mathbf{A}'^{-1}\mathbf{P}', \qquad \beta = 1/(1 - \mathbf{N}^T\mathbf{T}/d), \qquad \mathbf{N}^T\mathbf{T}/d \neq 1, \tag{4}$$

because in general $\mathbf{T}$ (i.e. the second camera center) does not lie on the reference plane which would lead to the degenerate case of the plane projecting as a line in one image.

Equation (4) represents the warping transform applied to the second image coordinates to account for the plane projective transformation. This warping transformation will exactly register points in the second image lying on the plane with their projections in the reference frame. However, the points not lying on the plane will have some residual displacement.

Substituting equation (1) in equation (4), we get

$$\mathbf{P^w} \quad \approx \quad [I + \beta\mathbf{T}\mathbf{N}^T/d]\mathbf{A}'^{-1}(\mathbf{A}'\mathbf{P} + \mathbf{T}') \quad \approx \quad (\mathbf{P} + \gamma\mathbf{T}), \tag{5}$$

where $\gamma = -1 + \beta\mathbf{P}^T\mathbf{N}/d - \beta\mathbf{T}^T\mathbf{N}/d = (\mathbf{P}^T\mathbf{N} - d)/(-(\mathbf{T}^T\mathbf{N} - d)) = d_N/(-T_d)$, d_N is the perpendicular distance of $\mathbf{P}$ from the plane, and T_d is the distance of the translation vector $\mathbf{T}$ from the plane.

In order to see that the parallax vectors between the warped points, $\mathbf{P^w}$, and the actual points, $\mathbf{P}$, are directed towards the epipole, it is easily shown from equation (5) that

$$\mathbf{T} \cdot (\mathbf{P^w} \times \mathbf{P}) = 0. \tag{6}$$

This is a projective relationship that shows that the projection plane normals defined by all the parallax vectors lie on a great circle on the unit sphere, and the translation vector is normal to the plane of this circle. Lawn and Cipolla [9] use this structure of the motion parallax field for the *special case* of image velocities (closely spaced viewpoints) to compute the epipole. They approximate the planar flow locally as an affine transformation. However, they do not relate the parallax field to the intrinsic structure of the scene. Also, the derivation here is valid for an arbitrary view transformation of the type in equation (1), that includes the case of small displacements. We now derive the relationship between the polar parallax field and scene structure.

Given a view ray $\mathbf{p'}$ in an arbitrary view, with the knowledge of the plane projective transformation of equation (3), its warped coordinates with respect to the reference view can be computed. For points that do lie on the plane, the warping transformation leads to their real projection in the reference view. For the non–planar points, the planar motion parallax vector (the difference between the virtual planar projection and the actual projection) is given by (figure 1):

$$\mathbf{p} - \mathbf{p^w} \quad = \quad \frac{1}{P_z}\mathbf{P} - \frac{1}{P_z^w}\mathbf{P^w} \quad = \quad \frac{1}{P_z}\mathbf{P} - \frac{1}{P_z + \gamma T_z}(\mathbf{P} + \gamma \mathbf{T}) \quad = \quad (1/(1 + \frac{P_z}{d_N}/\frac{T_z}{-T_d}))(\mathbf{p} - \mathbf{t}), \tag{7}$$

where the lower case bold letters represent the respective image vectors with their z–components unity. Note that an internal homogeneous camera transformation, $(\mathbf{A_c})$, can be applied to each of the image vectors, $\mathbf{p}, \mathbf{p^w}, \mathbf{t}$, in equation (7) without changing its form. Thus, the equation is valid for an arbitrary unknown internal camera transformation.

When T_z is zero, the parallax equation becomes:

$$\mathbf{p} - \mathbf{p^w} = (-d_N/P_z)[T_x \ T_y \ 0]^T \tag{8}$$

In this case, the parallax motion vectors are all parallel, oriented towards the epipole at infinity.

For an alternative but more tedious derivation of a similar result, without using image warping, see Lee [22]. A geometric derivation of the planar parallax under perspective projection result and a similar algebraic derivation has also been recently done independently by Kumar and Anandan [21].

We have shown that the parallax vector defined with respect to an arbitrary plane is directed towards the epipole in the reference image. Thus, the parallax vector field is due only to the translational component of the 3D view transformation, as is expected of any motion parallax field. The effect of rotations on the image motion has been eliminated by choosing a warping transformation corresponding to a plane in the environment. In the warped coordinate system, the motion disparity of the plane is zero. In other words, the points on the plane have been fixated through a coordinate transformation. The residual image motion is due only to the non–planar component of the environment, and translational motion. Recall that in traditional structure from motion algorithms, decomposing the image motion into rotational and translational components is hard because of inherent ambiguities [1, 6]. This problem may be circumvented in the planar parallax approach because the rotations affect only the plane projective transformation of equation (2) and not the parallax motion.

In order to use the above result for ego-motion and scene structure analysis, a reference plane can be chosen. The second frame will be transformed with the plane's transformation towards the reference view. Then, exploiting the epipolar structure of the residual image motion, the epipole can be computed. This leads to the decomposition of the motion into the motion part and the non-planar structure part as in equations 7 and 8. A particular implementation of this scheme using direct methods is presented in the section 4.

3.1 View–Invariant Representation

Let $\eta = 1/(1+P_z/d_N \frac{T_z}{d})$ (equation 7) be the (signed) ratio of the parallax magnitude to $|(\mathbf{p}-\mathbf{t})|$, and let $\tau = 1/\eta - 1$. For simplicity we call η the *parallax magnitude*. η is a function of the distance of $\mathbf{P}$ to the reference plane, d_N, and the z–components of $\mathbf{P}$ and $\mathbf{T}$. If a point P_0 not lying on the fixated plane is chosen as a reference then for any other point P_i: $\tau_i/\tau 0 = \frac{P_{i_z}}{d_{i_N}} / \frac{P_{0_z}}{d_{0_N}}$. That is, the ratio of the non–planar parallax magnitudes are dependent only on the relative structure of the environmental points not lying on the reference plane. This ratio represents a view–independent "coordinate" of the structure of the environment that does not lie on the reference plane. Given any arbitrary viewpoint, if the new view can be warped using the transformation corresponding to the reference plane, then the relative parallax magnitude is always the τ-ratio above. Thus, fixation with respect to the reference plane not only compensates for the effects of rotations, but also provides an environment centered reference surface with respect to which the complete shape of the environment can be specified.

3.2 Affine Reconstruction

If the internal camera parameters are known, and the 3D transformation between views is a rigid transformation (that is, the matrix $\mathbf{A}$ is a rotation matrix $\mathbf{R}$), then the reference plane can be reconstructed in a Euclidean frame attached to the reference view, and subsequently the whole scene can be reconstructed. The plane can be reconstructed in two ways: (i) by solving for the translation from the epipolar constraint of equation (7), and then solving for the rotation and the plane parameters from equation (2), or (ii) by solving for the plane and motion parameters directly from equation (2) [12]. The latter case may be unstable because it relies on higher order information (more than affine) in the image displacements; these generally are unreliable for commonly used small field-of-view cameras [1]. After solving for the plane, by choosing the ratio P_{0_z}/d_{0_N} for a reference non-planar point to be unity, all the other points can be reconstructed using their respective ratios P_{i_z}/d_{i_N} and their view rays $\mathbf{p}$. In particular, say for a point, $P_z/d_N = \alpha$, then since $\mathbf{P} = \lambda \mathbf{p}$, the two constraints define an intersection of the view ray with a plane. This intersection defines λ uniquely. If the reference plane is given by $\mathbf{P}^T\mathbf{N} = d$, then

$$\mathbf{P} = ((\alpha d)/\mathbf{p}^T(\alpha \mathbf{N} - \mathbf{z}))\mathbf{p}, \tag{9}$$

where $\mathbf{z}$ is the unit vector along the optical axis in the reference view.

However, when the internal camera parameters are unknown, and euclidean reconstruction is not required, then the reconstructed $\mathbf{P}$ of equation (9) represents the 3D geometry of the scene up to an arbitrary 3D affine transformation. To see this, assume that three points on the reference plane, and a fourth reference point not on the plane have been chosen arbitrarily and specified a set of 3D coordinates (say the standard affine basis). The coordinates of these four points are related to their true 3D coordinates (in some coordinate system) through a 12-parameter 3D affine transformation. This is left unspecified in the reconstruction.

Let three points on the reference plane and a non-planar reference point, ($\mathbf{P_0}$), be given some arbitrary 3D coordinates. Assume that these coordinates define the scene points in the coordinate system of the reference view. Thus, these coordinates are related to their true world coordinates through a transformation

$$\mathbf{P} = \mathbf{A}\mathbf{P_w} + \mathbf{T}. \tag{10}$$

Note that the internal camera transformation, $\mathbf{A_c}$, relating the ideal pin-hole model image coordinates to the measured image coordinates has been absorbed in the 3D affine transformation. The planar points define a plane $\mathbf{P}^T\mathbf{N} = d$. With $\mathbf{P_0}$ and the plane thus defined, the ratio P_{0_z}/d_{0_N} is fixed.

For any other non-planar point (fifth and more), say, the ratio P_z/d_N is α (as computed by the planar image warping and residual motion computation algorithm to be illustrated in the results section). Then, as in equation (9),

$$\mathbf{P} = ((\alpha d)/\mathbf{p}^T(\alpha\mathrm{N} - \mathbf{z}))\mathbf{p}, \tag{11}$$

defines the 3D $\mathbf{P}$. However, in this case, the 3D geometry can be specified only up to an unknown affine transformation that brings the arbitrarily selected four reference points into registration with the known corresponding points in the scene. Therefore, all the scenes related through a 3D affine transformation are indistinguishable in this approach. This is similar to the affine and projectively invariant reconstruction methods [15, 14, 28].

Note that in the reconstruction above, no explicit reconstruction of the 3D motion is required. Also, any arbitrary view, when warped for the reference plane with respect to the reference view, will lead to the same 3D reconstruction in a canonical coordinate system defined by the four chosen points.

3.3 Interpretation of the Parallax Magnitude

In particular, the three reference points on the plane can be chosen to be the points on the reference image plane. Thus, this image plane becomes the reference plane. Therefore, the plane normal $\mathbf{N} = \mathbf{z}$ and $d = 1$. Given these and α as above,

$$\mathbf{P} \;=\; ((\alpha d)/\mathbf{p}^T(\alpha\mathrm{N} - \mathbf{z}))\mathbf{p} \;=\; (1/(1 - \frac{1}{\alpha}))\mathbf{p}. \tag{12}$$

Recall that the magnitude of the parallax vector from equation 7 is proportional to $(1/(1 + \frac{P_z}{d_N}/\frac{T_z}{-T_d}))$. T_z/T_d can be conveniently set to unity to fix the overall scale. Then, the magnitude becomes $(1/(1 - P_z/d_N))$ which is the same as $(1/(1 - 1/\alpha))$ in equation 12 because $\alpha = P_z/d_N$. Therefore, in the coordinates of the reference image, the quantity $(1/(1 + \frac{P_z}{d_N}/\frac{T_z}{-T_d}))$ of equation 7 is directly *the depth of the corresponding point*. Of course, the structure reconstruction is valid up to an arbitrary 3D affine transformation as shown above. For an exact relationship between this interpretation and Shashua's [28] projective depth result see [26].

Figure 2: Two frames of a traffic sequence.

4 Results using Direct Method for Parallax

Bergen et al. [10] presented a multi-resolution iterative method for computing parametric and non-parametric image motion between two frames. We adopt their method for the implementation of the planar parallax idea presented above. The essential idea behind the direct method is to model image motion as

$$I_2(\mathbf{p}) = I_1(\mathbf{p} - \mathbf{u}(\mathbf{p};\mathbf{a})) \tag{13}$$

between images I_1 and I_2; $\mathbf{p}$ is the 2D vector of image coordinates, and $\mathbf{u}(\mathbf{p};\mathbf{a})$ is the displacement vector at $\mathbf{p}$ described using a parameter vector $\mathbf{a}$. For instance, $\mathbf{a}$ is an eight-dimensional vector for a plane projective transformation. In order to compute motions of varying magnitudes, the images are represented at multiple scales using Gaussian or Laplacian pyramids. Starting at the coarsest level, given some initial estimate of the displacement vectors, iterative refinement of the displacement vectors is performed by successive warping of one image towards the reference image and recomputation of the new incremental displacement vectors. Given an estimate $\mathbf{u}_i$, an incremental update $\delta\mathbf{u}$ is computed by minimizing the quadratic error measure

$$E(\delta\mathbf{u}) = \sum_{\mathbf{p}} (\delta I(\mathbf{p}) + \nabla I_2 . \delta\mathbf{u}(\mathbf{p}))^2 \tag{14}$$

where

$$\delta I(\mathbf{p}) = I_2(\mathbf{p}) - I_1(\mathbf{p} - \mathbf{u}_i). \tag{15}$$

Note that I_2 is considered the reference image. $I_1(\mathbf{p}-\mathbf{u}_i)$ is the warped image one, warped corresponding to the current estimate of the displacement vectors. We now illustrate how the planar parallax approach using the direct method simplifies analysis of (i) multiple motions, and (ii) 3D geometry.

Traffic Scene : Multiple Motions

Figure 2 shows two 480×512 frames of a traffic sequence in which the camera moved sideways and there are independently moving objects too (trucks and other vehicles). The camera motion is unknown. In order to give a sense of the motion, figure 4 shows the image flow computed using the non-parametric direct method with a displacement vector for every point. The flow at every point $\mathbf{p}$ is computed by solving for a single $\mathbf{u}$ (equation 14) in a small neighborhood within the coarse to fine estimate and warp framework [10]. The flow[1] shows the motion due to the camera of the static environment and that of the moving objects. There is no clear separation between these various components.

For the purposes of demonstration here, the region of the road was selected in frame 2, the reference frame, as the reference plane region, and a warping transformation was computed using the direct method. A method like Irani et al.'s [11] could be used for selecting the plane automatically. Our multi-resolution direct method computes the planar transformation and performs image warping together iteratively from coarse to fine scales. Note that only the bounding box of the planar region in frame 2 were specified without any point correspondences. The direct estimation process automatically brought the corresponding regions in registration while solving for the planar transformation. We note here that a six-parameter affine transformation resulted in better registration result than an eight-parameter planar transformation. Figure 3 shows frame 1 warped according to the transformation computed[2].

Finally, figure 5 shows the displacement vectors between the reference frame, frame 2, and the affine warped frame 1. In this figure the qualitative structure and motion of the scene is quite apparent. The flow in the region of the road is mostly zero. The flow corresponding to the region of the trees in the background is mostly parallel and is due to the translational component of the sideways camera motion (FOE out of the image plane). For the two trucks moving in the foreground, the displacement vectors are both due to their independent motion and the translational motion of the camera. Since their independent motion is mostly translational, these flow vectors point towards a focus of expansion. Therefore, the structure of the displacement field after compensating for the planar warp is considerably simpler than the original field.

It is to be noted here that most methods for multiple motion segmentation [2, 3, 11] rely on either simple parametric models (affine/planar), or on the smoothness of flow in the image plane. No 3D motion and structure constraints are brought to bear on the process. These methods may fail to detect self motion, or may over-segment an object even for simple cases of translation in depth as is the case with the moving trucks in our example. The above planar parallax representation allows for easy exploitation of the 3D motion and structure constraints in the segmentation and reconstruction task. We are in the process of using this representation for object and structure segmentation to identify objects and events of interest in a video sequence.

Box Scene : Shape Reconstruction

The second set of results is on images of a rotating box; two frames are shown in figure 6. The box was rotated about 4^o around an axis through the centers of the top and bottom faces with the rest of the scene and camera stationary.

The left face of the box in the second frame (reference frame) was specified as the region of the reference plane, and the first frame was registered with respect to the second using direct automatic

---

[1] In all the flow figures, the scale is chosen by the display program and is not the same as of the original images.

[2] In both this and the next example, the comparison between the reference frame and the warped frame is not very illustrative when shown as hard copy images, but the fixated planar region and the residual motion is a very compelling display when shown as a movie.

planar transformation warping as described in the previous section. Again, in this case, it was found that an affine transformation (6-parameter) was sufficient for the planar registration. The output of this process is shown as a warped image (frame 1 warped), and as the difference between the warped frame 1 and the reference frame (frame 2) in figure 7. In order to highlight the simplified representation obtained after planar warping, we show the "raw" flow between frames 1 and 2, and the residual flow after warping between frame 1 warped and frame 2 in figure 8. Clearly, in the region of the box, the original flow is rotational but after planar warping, the residual flow is mostly along the horizontal axis because in this case, the effective translation is mostly along the x axis.

Furthermore, from equation 8, in this case of purely x/y translation, the flow magnitude is directly the depth in the coordinates of the reference image up to an unknown 3D affine transformation. Since all of the box that is not on the reference plane is on one side of it, if we plot the magnitude of flow as a function of the image plane xy-coordinate system, then this will represent the intrinsic structure of the box up to an arbitrary 3D affine transformation. The intrinsic shape estimate is shown as a surface plot in figures 9 and 10. The viewpoint has been chosen to make the computed shape fairly explicit. (All the surface plots use some arbitrary scale and coordinate system specific to the plotting programs.) Note that in the regions corresponding to the background, the flow is arbitrary because the background was stationary and only the box was moving; but the planar warping was applied to the whole image which leads to some arbitrary flow for the background. The surface plots clearly show that the qualitative estimates of the planar facets of the box and the overall shape have been recovered fairly well.

5 Conclusions

A new technique for simplifying the structure of image flow for multiple motion and scene-centered structure analysis has been demonstrated in this work. The formulation of motion parallax with respect to a reference plane in terms of warped image coordinates allows a natural use of direct methods for analyzing image motion. It is also shown how affine and Euclidean structure is quite simply represented in the parallax flow. This work represents steps towards building a set of application tools for (semi) automatic annotation and analysis of motion videos for indexing and retrieval.

References

[1] G. Adiv. Inherent ambiguities in recovering 3D information from a noisy flow field. *IEEE PAMI*, 11(5):477–489, 1989.

[2] M. J. Black and P. Anandan. Robust dynamic motion estimation over time. In *CVPR*, pages 296–302, 1991.

[3] P. Bouthemy and E. Francois. Motion segm. and qual. dynamic scene analysis from an image seq. *IJCV*, 10(2):157–182, 1993.

[4] S. Carlsson and J. Eklundh. Obj. det. using model based pred. and motion parallax. In *1st ECCV*, pages 297–306, 1990.

[5] K. Daniilidis and H. H. Nagel. Analytical results on error sensitivity of motion estimation from two views. In *1st ECCV*, pages 199–208, 1990.

[6] K. Daniilidis and H. H. Nagel. The coupling of rotation and translation in motion estimation of planar surfaces. In *CVPR*, pages 188–193, 1993.

[7] W. Enkelmann. Obst. detection by evaluation of opt. flow fields from image seqs. *IVC*, 9(3):160–168, 1991.

[8] A. Azarbayejani et al. Recursive estimation of structure and motion using relative orientation constraints. In *CVPR*, pages 294–299, 1993.

[9] J. Lawn et al. Epipole estimation using affine motion parallax. Technical Report CUED/F-INFENG/TR 138, Camb. Univ. Engg. Dept., 1993.

[10] J. R. Bergen et al. Hierarchical model–based motion estimation. In *2nd ECCV*, pages 237–252, 1992.

[11] M. Irani et al. Detecting and tracking multiple moving objects using temporal integration. In *ECCV*, pages 282–287, 1992.

[12] O. D. Faugeras et al. Let us suppose the world is piece-wise planar. In *3rd Intl. Symp. on Rob. Res.*, 1987.

[13] R. Hartley et al. Computing matched epipolar projections. In *CVPR*, pages 549–555, 1993.

[14] R. Mohr et al. Relative 3D reconstruction using multiple uncalibrated images. In *CVPR*, pages 543–548, 1993.

[15] O. D. Faugeras. What can be seen in three dimensiona with an uncalibrated stereo rig ? In *2nd ECCV*, pages 563–578, 1992.

[16] R. I. Hartley. Estimation of relative camera positions for uncalibrated cameras. In *2nd ECCV*, pages 579–587, 1992.

[17] J. C. Hay. Optical motions and space perception: An extension of Gibson's analysis. *Psych. Rev.*, 73:550–565, 1966.

[18] B. K. P. Horn. Relative orientation. *IJCV*, 4(1):59–78, 1990.

[19] A. D. Jepson and D. J. Heeger. Linear subspace methods for recovering translational direction. Technical Report RBCV–TR–92–40, University of Toronto, 1992.

[20] J. J. Koenderink and A. J. van Doorn. Affine structure from motion. *JOSA A*, 81:377–385, 1991.

[21] Rakesh Kumar and P. Anandan. Personal Communication.

[22] Chia-Hoang Lee. Structure and motion from two perspective views via planar patch. In *ICCV*, pages 158–164, 1988.

[23] H. C. Longuet-Higgins and K. Prazdny. The interpretation of a moving retinal image. In *Proc. Royal Society of London B*, pages 385–397, 1980.

[24] C. J. Poelman and T. Kanade. A paraperspective factorization method for shape and motion recovery. Technical Report CMU–CS–92–208, Carnegie Mellon University, 1992.

[25] J. H. Rieger and D. T. Lawton. Processing differential image motion. *JOSA A*, 2(2):354–360, 1985.

[26] H. S. Sawhney. 3D geometry from planar parallax. Technical Report RJ 9650 (84090) CS, IBM, 1993.

[27] H. S. Sawhney and A. R. Hanson. Comparative results of some motion algorithms on real image sequences. In *DARPA IUW*, 1990.

[28] A. Shashua. Projective depth: A geometric invariant for 3D reconstruction from two perspective/orthographic views and for visual recognition. In *ICCV*, pages 583–590, 1993.

[29] R. Szeliski and S. B. Kang. Recovering 3D shape and motion from image streams using non–linear least squares. In *CVPR*, pages 752–753, 1993.

[30] C. Tomasi and T. Kanade. Shape and motion from image streams under orthography: A factorization method. *IJCV*, 9(2):137–154, 1992.

[31] D. Weinshall. Model based invariants for 3D vision. *IJCV*, 10(1):27–42, 1993.

Figure 3: **Affine warped frame 1.**

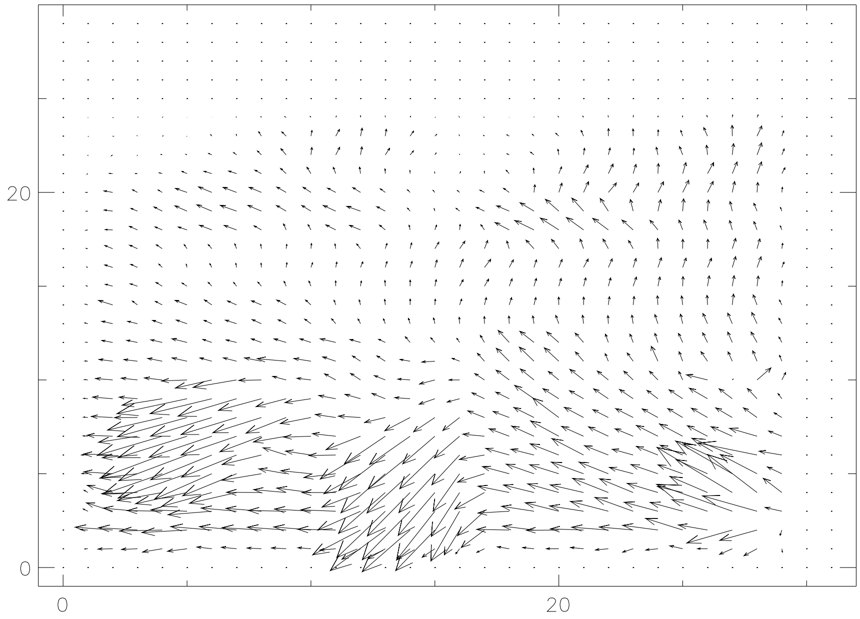

Figure 4: **Displacement field between frames 1 and 2.** Subsampled 30×32 from 480×512 frames.

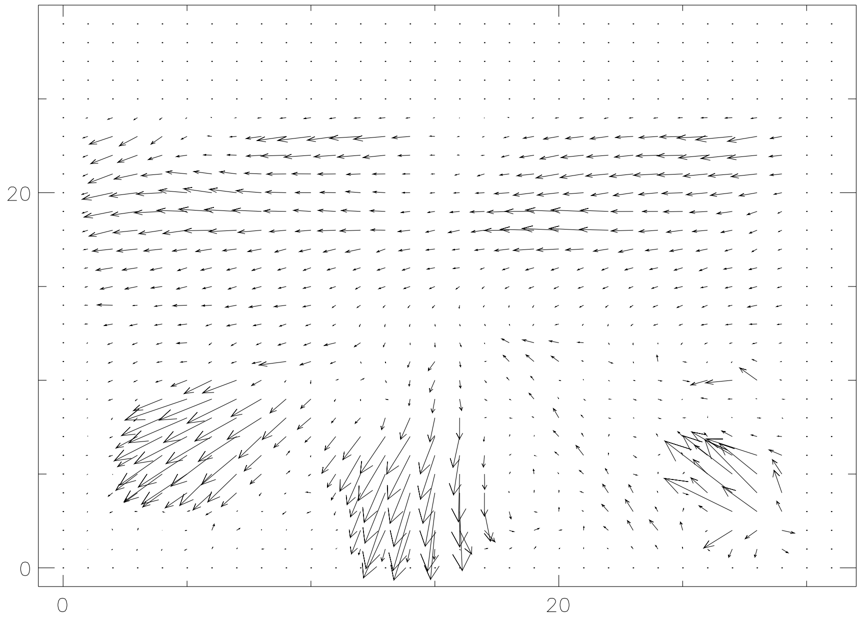

Figure 5: **Displacement field between affine warped frame 1 and 2.** Subsampled 30×32.

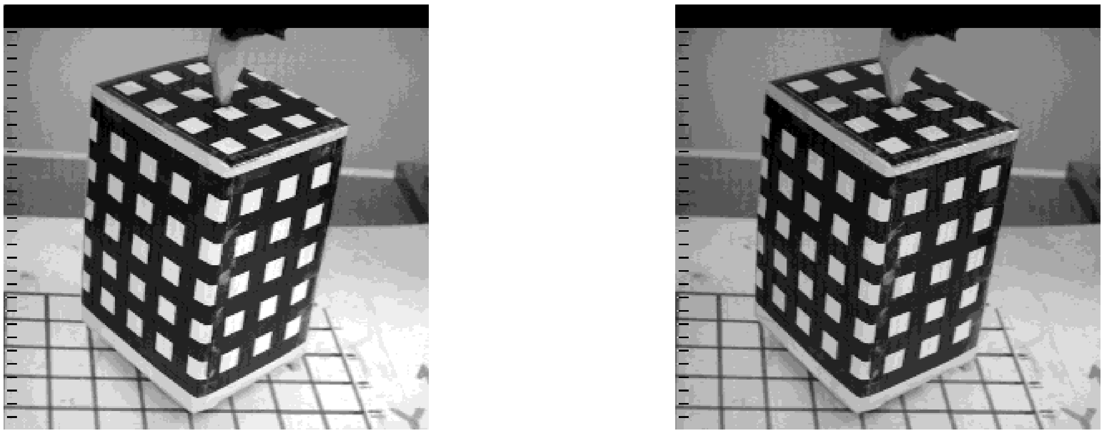

Figure 6: Two frames of a box sequence.

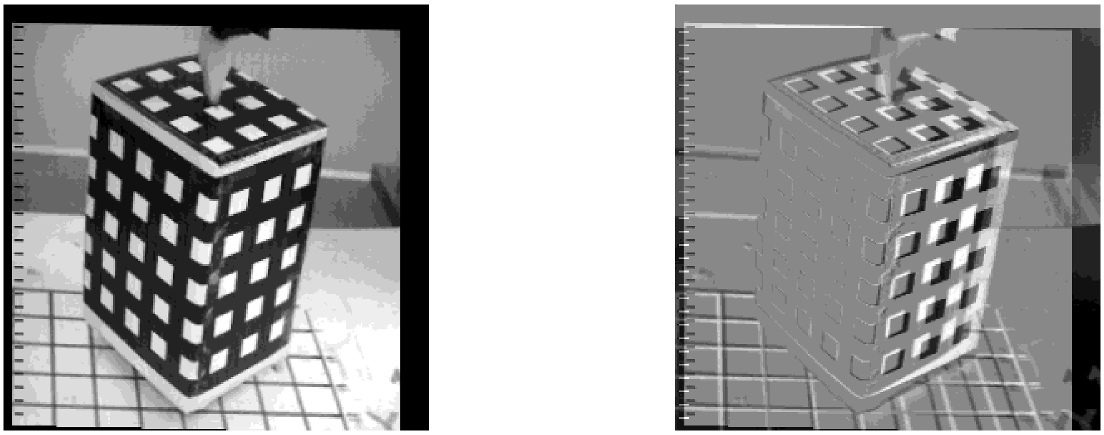

Figure 7: Frame 1 warped and difference between frame 1 warped and frame 2.

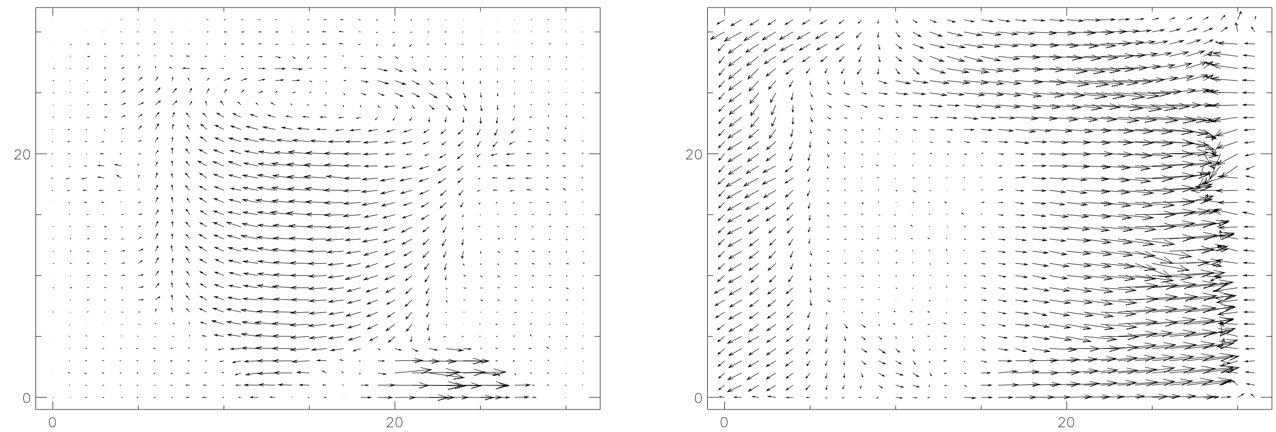

Figure 8: Flow between original frames 1 and 2, and between warped frame 1 and frame2.

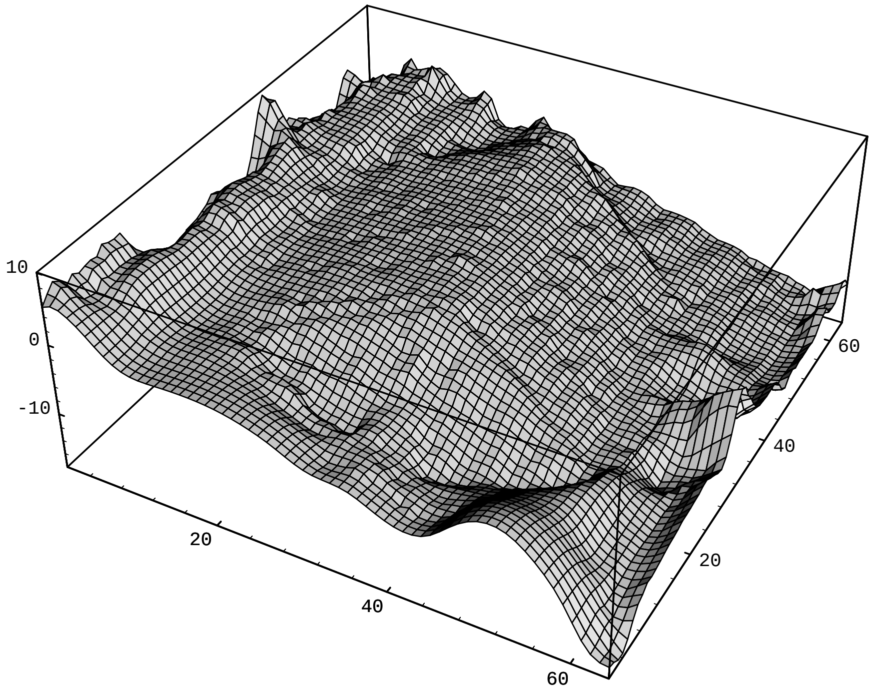

Figure 9: **Grided surface plot of the box.**

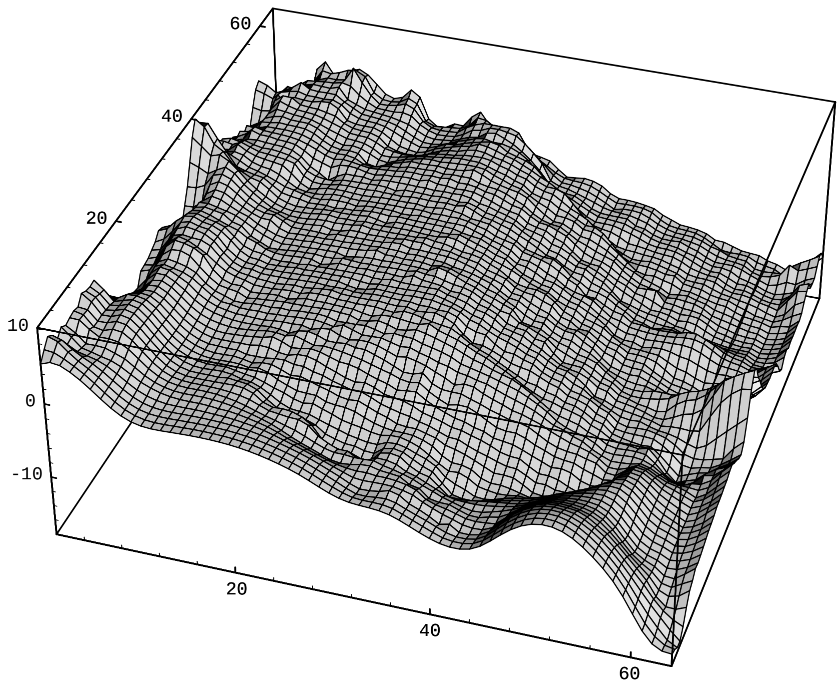

Figure 10: **Grided surface plot of the box.**