Dense Shape and Motion from Region Correspondences by Factorization

Jianbo Ma and Narendra Ahuja
Beckman Institute and Department of Electrical and Computer Engineering
University of Illinois at Urbana-Champaign
Urbana, IL, 61801

Abstract

In this paper, we propose an algorithm for estimating dense shape and motion of dynamic piecewise planar scenes from region correspondences using factorization. Region correspondences are used since they are easier to establish and more reliable than either line or point correspondences. The image measurements required are the centroid and area for each region. Singular value decomposition is employed to find the basis of range space of the motion, shape, and surface normal matrices. By imposing model constraints, motion, shape, and surface normal can be recovered only from region correspondences.

1 Introduction

Recovering 3-D shape and motion from image sequences is an important task in computer vision. The existing methods can be classified into two categories: the feature based and the optical flow based. The accuracy of structure and motion estimates depends on the accuracy of feature correspondences and optical flow. Most of the existing structure from motion algorithms use point or line features. This paper uses regions to estimate image motion and thereby 3D structure and motion. The use of regions is expected to reduce the sensitivity to noise, thus improving the estimates.

Several efforts have been made to solve the problem of structure from motion using region information. Sull and Ahuja [11] proposed a region-based method to estimate the 3-D structure and motion of textured piecewise planar surfaces. They assume that each plane has at least four regions, and use Hough transform to estimate the second order image flow and to perform region matching. A closed form solution is then obtained by solving the flow field equations. Negahdaripour and Lee [10] estimate first order optical flow. They prove that in order to recover the motion and 3-D planar structure, at least two image regions with distinct first order flow structures have to be identified. Using the coefficients of the flows for such regions, a closed form solution for the camera motion and the orientation of the surface normals of the regions are found. Lee and Cooper [5] approximate a 3-D surface by planar patches. Region matching between images is done by using affine invariants. Having recovered the region matches, they give a closed form solution for motion and structure of the regions. Also, Schweitzer and Krishnan [6] approximate image motion induced by camera or object motion by an affine coordinate transformation and extract 3-D information directly from the affine parameters. Their algorithm uses expressions involving six affine parameters computed for each image patch. All these methods have to estimate parameters relating regions in motion. In this paper, we introduce a region-based factorization method for estimating dense shape and motion of textured, piecewise planar scenes directly from region correspondences.

Factorization for motion and structure estimation has been used in two related contexts under orthographic assumptions [1, 4, 2]. Singular value decomposition (SVD) technique is the key step for this family of algorithms. One purpose of SVD is to find the basis for range space of parameter matrices, and the other purpose is to reduce measurement noise by directly exploring data model constraints (rank theorem). Debrunner and Ahuja [1, 2] propose an algorithm based on factorization by assuming that motion is constant over a short period, which gives closed-form expressions of shape and motion for multiple objects motion. Tomasi and Kanade [4] factorize the feature point locations in image stream into the object shape and camera motion. A family of structure from motion algorithms using long sequences based on factorization has been developed [4, 3, 8]. These algorithms use models ranging from orthographic projection model to affine camera model and point correspondences to line correspondences. In this paper, we propose an

algorithm that uses region features assumed to lie on piecewise planar surfaces imaged under orthographic projection.

A multiscale segmentation of all image frames is performed and the regions are matched [12, 9]. The centroids and areas of regions are recorded in the relative position matrix and the area matrix. These two matrices carry information about motion and structure, which can be extracted by singular value decomposition followed by a normalization procedure. The key idea is that by including region area measurements into the motion and structure normalization step, both the 3-D coordinates of region centroids and the corresponding planar patch normals can be computed, which means dense shape and motion can be recovered.

This paper is organized as follows. We first analyze the data model - the relationship between the parameters to be estimated and the measurements. Then, an algorithm based on orthographic projection is presented followed by dense shape and motion solutions under weak-perspective and para-perspective models. Finally, the algorithm is tested using synthetic data and real images.

2 Data Model

Suppose an object surface consists of R planar patches whose correspondences are known across image frames. Denote region r in f-th frame $RE_{r,f}$, where $r = 1, \cdots, R; f = 1, \cdots, F$. Due to the piecewise planar nature of the scene, the dense shape can be represented by (1) the centroid of each region $\overrightarrow{C_r}$, which is a 3-D vector in the world coordinate system, and (2) the unit normal vector $\overrightarrow{n_r}$. The objective then is to recover $\overrightarrow{C_r}$ and $\overrightarrow{n_r}$ for each region.

The recovery is based on the locations and areas of image regions, since under rigidity constraint, the relative positions as well as the areas of the corresponding 3D planar patches are invariant to motion. Two measurement matrices, called *relative position matrix* and *region area matrix* are formed, and motion and structure are recovered from them by singular value decomposition.

2.1 Relative Position Matrix

The relative region positions are represented by the relative position vectors, defined to be $\overrightarrow{P}_r = \overrightarrow{C}_r - \overrightarrow{C}_{ref}, r = 1, \cdots, R$, which gives the position of region r relative to some reference region (Figure 1). The reference region RE_{ref} can be arbitrarily chosen from the R regions.

Under orthographic projection, a 3-D vector (or point) $\overrightarrow{S_p}$ in world coordinate system is related to the

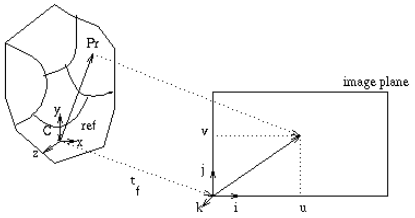


Figure 1: Relative Position Vectors

image plane pixel (u_{fp}, v_{fp}) by

$$u_{fp} = \overrightarrow{i}_f^T(\overrightarrow{S}_p - \overrightarrow{t}_f), v_{fp} = \overrightarrow{j}_f^T(\overrightarrow{S}_p - \overrightarrow{t}_f) \quad (1)$$

where $\overrightarrow{t}_f = (a_f, b_f, c_f)^T$is the vector from the world origin to the origin of image frame f, and $\overrightarrow{i}_f$ and $\overrightarrow{j}_f$ are the direction vectors along x and y axes in the image plane.

By moving the world origin to the centroid of the reference region, we obtain the relationship between the relative position vector $\overrightarrow{P}_r$ and its projection on the image plane $\tilde{u}_{fr}$ as

$$\tilde{u}_{fr} = \overrightarrow{i}_f^T \overrightarrow{P}_r, \tilde{v}_{fr} = \overrightarrow{j}_f^T \overrightarrow{P}_r \quad (2)$$

We measure the projections of the relative position vectors to form a measurement matrix $\tilde{W}$ as

$$\tilde{W} = RP \quad (3)$$

where $R = \left[\begin{array}{cccccccc} \overrightarrow{i_1} & \overrightarrow{i_2} & \cdots & \overrightarrow{i_F} & \overrightarrow{j_1} & \overrightarrow{j_2} & \cdots & \overrightarrow{j_F} \end{array} \right]^T$ is a $2F \times 3$ matrix representing motion (rotation), and $P = \left[\begin{array}{ccc} \overrightarrow{P_1} & \cdots & \overrightarrow{P_R} \end{array} \right]$ is a $3 \times (R-1)$ relative position matrix. Obviously, $\tilde{W}$ is at most of rank three due to the fact that it is the product of a $2F \times 3$ matrix and a $3 \times (R-1)$ matrix.

2.2 Region Area Matrix

Let S_W be the actual area of a planar patch, S_I be the area projected onto the image plane, and $\overrightarrow{n}$ be the unit normal vector of the planar patch. Under orthographic projection model, let $\overrightarrow{N_r}$ be the weighted normal vector of patch r, i.e. $\overrightarrow{N_r} = S_W \overrightarrow{n_r}$. Given R region correspondences in F frames, denote the area of region r in frame f as $a_{fr} = \overrightarrow{k_f}^T \overrightarrow{N_r}$, where $r = 1, \cdots, R$ and $f = 1, \cdots, F$. Form an $F \times R$ measurement matrix A as

$$A = \left[\begin{array}{cccc} a_{11} & a_{12} & \cdots & a_{1R} \\ a_{21} & a_{22} & \cdots & a_{2R} \\ \vdots & \vdots & \ddots & \vdots \\ a_{F1} & a_{F2} & \cdots & a_{FR} \end{array} \right] \quad (4)$$

The measurement matrix can be expressed in matrix form as

$$A = KN \tag{5}$$

where,

$$\begin{aligned} K &= \left[\begin{array}{cccc} \overrightarrow{k_1} & \overrightarrow{k_2} & \cdots & \overrightarrow{k_F} \end{array} \right]^T \\ N &= \left[\begin{array}{cccc} \overrightarrow{N_1} & \overrightarrow{N_2} & \cdots & \overrightarrow{N_R} \end{array} \right] \\ &= \left[\begin{array}{cccc} \overrightarrow{n_1} & \overrightarrow{n_2} & \cdots & \overrightarrow{n_R} \end{array} \right] \left[\begin{array}{cccc} a_1 & 0 & \cdots & 0 \\ 0 & a_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & a_F \end{array} \right] \end{aligned}$$

Clearly, the rank of matrix A is also at most three.

3 Shape and Motion Estimation

From the above analysis, we have two measurement matrices: the relative position matrix and the region area matrix. These two measurement matrices are of at most rank three from the data model. We perform singular value decomposition on these two matrices and find their best rank three approximates respectively. One purpose for this is performing denoising of the measurement matrices according to the data model. Because the inherent data model can only produce rank three measurement matrices, if the measurement matrices have rank larger than three, it must be caused by the measurement noise. So, the rank three approximation can be considered as noise filtering in the least squares sense. Another purpose of singular value decomposition is to find the basis of the motion subspace, the shape subspace, and the surface normal subspace, after which we can use the metric constraints to solve for motion matrix, shape matrix and surface normal matrix.

3.1 Relative Position Matrix Factorization

From the relative position matrix equation (3), $\tilde{W} = RP$, we compute the singular value decomposition of matrix $\tilde{W}$ [4], $\tilde{W} = O_1 \Sigma O_2$, such that $O_1^T O_1 = O_2^T O_2 = O_2 O_2^T = I$, where I is the $(R-1) \times (R-1)$ identity matrix. Σ is a diagonal matrix whose diagonal entries are the singular values. If the singular values are sorted in non-decreasing order, the first three columns of O_1 denoted as O_1', the first 3×3 submatrix of Σ denoted as Σ', and the first 3 rows of O_2 denoted as O_2' will form the best rank three approximation to $\tilde{W}$ as

$$\tilde{W} = O_1' \Sigma' O_2' \tag{6}$$

If we define $\hat{R} = O_1'$ and $\hat{P} = \Sigma' O_2'$, then we can write $\tilde{W} = \hat{R}\hat{P}$ $\hat{R}$ and $\hat{P}$ are with respect to a linear transform Q_1 of the true solution R and P, that is

$$R = \hat{R} Q_1, S = Q_1^{-1} \hat{P} \tag{7}$$

where, Q_1 is an invertible 3×3 matrix. Q_1 is determined later by using the metric constraints in the normalization step.

3.2 Region Area Matrix Factorization

Referring to equation (5), $A = KN$, we compute the SVD of matrix A, and find best rank three approximation of A as

$$A = O_1' \Sigma' O_2' = \hat{K}\hat{N} \tag{8}$$

where $\hat{K} = O_1'$ and $\hat{N} = \Sigma' O_2'$. The true solution matrix K and the weighted surface normal matrix N are related to $\hat{K}$ and $\hat{N}$ by a linear transform Q_2, i.e.

$$K = \hat{K} Q_2, N = Q_2^{-1} \hat{N} \tag{9}$$

where, Q_2 is an invertible 3×3 matrix and can be solved for in a normalization step.

3.3 Normalization: Solve for Q_1 and Q_2

From the above two factorization steps, we have the affine motion and structure up to the liner transforms Q_1 and Q_2. In order to calculate the true solution, Q_1 and Q_2 have to be calculated by imposing metric constraints [4].

The unknown variables include the 18 matrix components in Q_1 and Q_2. Using the metric constraints

$$\overrightarrow{i}_f^T Q_1 Q_1^T \overrightarrow{i}_f = \overrightarrow{j}_f^T Q_1 Q_1^T \overrightarrow{j}_f = \overrightarrow{k}_f^T Q_2 Q_2^T \overrightarrow{k}_f = 1, \tag{10}$$

$$\overrightarrow{i}_f^T Q_1 Q_1^T \overrightarrow{j}_f = \overrightarrow{i}_f^T Q_1 Q_2^T \overrightarrow{k}_f = \overrightarrow{j}_f^T Q_1 Q_2^T \overrightarrow{k}_f = 0, \tag{11}$$

this non-linear data fitting problem is solved by the Levenberg-Marquardt method.

4 Outline of the Algorithm

Based on the development in the previous sections, we now have an algorithm for dense shape and motion estimation:

1. Compute the SVD of $\tilde{W} = O_1 \Sigma O_2$ and find best rank three approximate of $\tilde{W} = \hat{R}\hat{P}$, where $\hat{R} = O_1'$ and $\hat{P} = \Sigma' O_2'$.

2. Compute the SVD of $A = O_1 \Sigma O_2$ and find best rank three approximate $A = \widehat{K}\widehat{N}$, where $\widehat{K} = O_1'$ and$\widehat{N} = \Sigma' O_2'$.

3. Compute the matrices Q_1, Q_2 by imposing the metric constraints.

4. Compute the rotation matrix R and K as $R = \widehat{R}Q_1$ and $K = \widehat{K}Q_2$. Compute the relative position vector matrix P as $P = Q_1^{-1}\widehat{P}$. Compute the weighted surface normal matrix as $N = Q_2^{-1}\widehat{N}$.

5. If desired, align the first camera reference system with the world reference system by forming the products RR_0, $R_0^T P$, KR_0, and $R_0^T N$, where the orthonormal matrix $R_0 = [\overrightarrow{i_1}\,\overrightarrow{j_1}\,\overrightarrow{k_1}]$ rotates the first camera reference system into the identity matrix.

6. Normalize the weighted surface normal vectors in matrix N to get the unit surface normals. i.e. unit surface normals, $\overrightarrow{n_r} = \overrightarrow{N_r}/\|\overrightarrow{N_r}\|$, and region area $a_r = \|\overrightarrow{N_r}\|$, for $r = 1, 2, \cdots, R$.

7. Calculate shape from region centroid 3-D coordinates $\overrightarrow{P_r}$ and surface normal $\overrightarrow{n_r}$, since in region r, the 3D coordinates for every pixel $\overrightarrow{P}$ satisfies equation:$(\overrightarrow{P} - \overrightarrow{P_r})^T \overrightarrow{n_r} = 0$.

4.1 Alternative Method for Motion, Shape and Surface Normal

In the algorithm described above, a non-linear data fitting procedure is invoked to calculate Q_1 and Q_2. Instead, a linear procedure can be used if the relative position matrix is more accurate than the region area matrix. We can first calculate shape and motion from the relative position matrix using Tomasi and Kanade's algorithm [4, 13].

From the metric constraints, i.e.

$$\overrightarrow{i}_f^T \overrightarrow{i_f} = \overrightarrow{j_f}^T \overrightarrow{j_f} = 1, \overrightarrow{i_f}^T \overrightarrow{j_f} = 0 \tag{12}$$

we obtain the system of $3 \times F$ over-determined equations for Q_1. Instead of solving for Q_1, if let $L = Q_1 Q_1^T$, the $3 \times F$ equations become

$$\hat{\overrightarrow{i_f}}^T L \hat{\overrightarrow{i_f}} = \hat{\overrightarrow{j_f}}^T L \hat{\overrightarrow{j_f}} = 1, \hat{\overrightarrow{i_f}}^T L \hat{\overrightarrow{j_f}} = 0 \tag{13}$$

where $L \in R^{3\times 3}$ is a symmetric matrix and $\hat{i_f}$ and $\hat{j_f}$ are the rows of $\widehat{R}$. By denoting $\hat{\overrightarrow{i_f}}^T = [\overrightarrow{i_{f1}}, \overrightarrow{i_{f2}}, \overrightarrow{i_{f3}}]$, $\hat{\overrightarrow{j_f}} = [\overrightarrow{j_{f1}}, \overrightarrow{j_{f2}}, \overrightarrow{j_{f3}}]$, and

$$L = \begin{bmatrix} l_1 & l_2 & l_3 \\ l_2 & l_4 & l_5 \\ l_3 & l_5 & l_6 \end{bmatrix} \tag{14}$$

the metric constraints can be rewritten as

$$G\overrightarrow{l} = \overrightarrow{c} \tag{15}$$

where $G \in R^{3F\times 6}$, $\overrightarrow{l} \in R^6$, and $c \in R^{3F}$ are defined by

$$G = \begin{bmatrix} g^T(\overrightarrow{i_1}, \overrightarrow{i_1}) \\ \vdots \\ g^T(\overrightarrow{i_f}, \overrightarrow{i_f}) \\ g^T(\overrightarrow{j_1}, \overrightarrow{j_1}) \\ \vdots \\ g^T(\overrightarrow{j_f}, \overrightarrow{j_f}) \\ g^T(\overrightarrow{i_1}, \overrightarrow{j_1}) \\ \vdots \\ g^T(\overrightarrow{i_f}, \overrightarrow{j_f}) \end{bmatrix}$$

$$\overrightarrow{l} = [l_1, \cdots, l_6]^T, \overrightarrow{c} = [1, \cdots, 1, 0, \cdots, 0]^T$$

$$g^T(a_f, b_f) = [a_{f1}b_{f1}, a_{f1}b_{f2} + a_{f2}b_{f1}, a_{f1}b_{f3} + a_{f3}b_{f1}, a_{f2}b_{f2}, a_{f2}b_{f3} + a_{f3}b_{f2}, a_{f3}b_{f3}]. \tag{16}$$

The minimum norm least squares solution of the system is given by the pseudo-inverse method, i.e.

$$\overrightarrow{l} = G^\dagger \overrightarrow{c}. \tag{17}$$

where $G^\dagger$ is the Moore-Penrose pseudo-inverse of matrix G. When G has full column rank, $G^\dagger = (G^T G)^{-1} G^T$. The vector $\overrightarrow{l}$ determines the symmetric matrix L, whose eigendecomposition gives Q_1.

After motion R and the shape S are calculated, the K matrix can be determined by $\overrightarrow{k_f} = \overrightarrow{i_f} \times \overrightarrow{j_f}$. Then, from equation (5) $A = KN$, the surface normal can be determined as $N = K^\dagger A$, where $K^\dagger$ is the Moore-Penrose pseudo-inverse of matrix K. After normalization, we get both the true region areas and the surface normals.

4.2 Shape and Motion Under Weak-perspective and Para-Perspective Models

4.2.1 Weak-perspective

Under weak-perspective projection, the relationship between weighted surface normal $\overrightarrow{N}$ and Z-axis direction vector $\overrightarrow{k}$ is

$$\overrightarrow{k}^T \overrightarrow{N}/d^2 = \alpha \tag{18}$$

where d is the scaling factor of weak-perspective model and α is the region area in image plane.

If we define $\overrightarrow{i_f}' = \overrightarrow{i_f}/d_f$ and $\overrightarrow{j_f}' = \overrightarrow{j_f}/d_f$, absorb the scaling factor into the weighted surface normal and the Z-axis direction $\overrightarrow{k}$, *i.e.*, $\overrightarrow{N}' = 1/d\overrightarrow{N}$ and $\overrightarrow{k}' = 1/d\,\overrightarrow{k}$, the measurement matrices $\widetilde{W}$ and A and the

factorization step are the same as in the orthographic case. Because we do not know the value of depth d_f, the metric constrains in the normalization step cannot be the same as in orthographic case [3]. Instead we can impose the constrains

$$\|\overrightarrow{i_f}'\|^2 = \|\overrightarrow{j_f}'\|^2 = \|\overrightarrow{k_f}'\|^2 \tag{19}$$

We can still use the orthogonality constraints

$$\overrightarrow{i_f}'^T \overrightarrow{j_f}' = \overrightarrow{i_f}'^T \overrightarrow{k_f}' = \overrightarrow{j_f}'^T \overrightarrow{k_f}' = 0 \tag{20}$$

In order to avoid trivial solutions, we impose the constraint $\|\overrightarrow{i_f}'\| = 1$.

From the above constraints, we compute Q_1 and Q_2. Other steps are the same as for orthographic case. We can also first solve for Q_1 and then compute the surface normal.

4.2.2 Para-Perspective

Under the para-perspective projection model, the area in the image plane can be expressed as [7]

$$S_I = 1/d^2 S_W (\overrightarrow{r}^T \overrightarrow{n})/(\overrightarrow{r}^T \overrightarrow{k}) \tag{21}$$

where, $\overrightarrow{r}$ is the projection direction vector. This equation is non-linear and could not facilitate factorization of the region area matrix. However, by para-perspective factorization method [3], $\overrightarrow{k_f}$, d_f, and $\overrightarrow{r_f}$ can be solved. From the above equation, we get

$$\overrightarrow{r_f}^T \overrightarrow{n_f} S_W = d^2 S_I \overrightarrow{r_f}^T \overrightarrow{k_f} \tag{22}$$

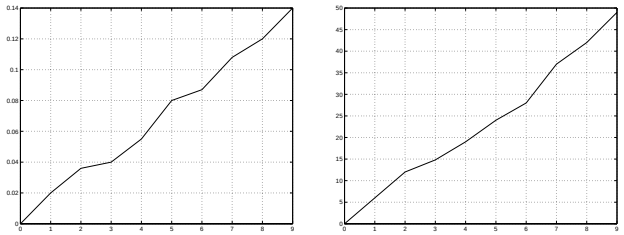
By measuring the region area in many frames and including the computed parameters, we can form a least-squares approximation of the surface normal $\overrightarrow{n}$. Thus the shape and motion can also be computed by forming the region area measurements.

5 Experiments

5.1 Synthetic Data

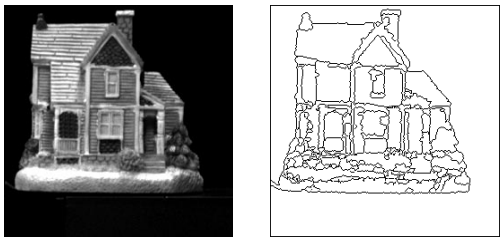
The synthetic data used were created by simulating 200 planar patches in random 3D motion. 200 3D coordinates, 200 3D unit directional vectors, and 200 positive numbers are generated randomly to simulate the centroids of the planar patches, the associated surface normals and the region areas, respectively. 50 random $\overrightarrow{i}$, $\overrightarrow{j}$, and $\overrightarrow{k}$'s are generated to represent 50 frames of random camera motion. The measurement matrices are then created by orthographic projection and perturbed by addicative Gaussian noise, which simulate the region tracking error and image noise.

Figure 2(a)(b) show the sum of squared estimation error of motion and shape as a function of the variance



(a)Motion Error (b)Shape Error

Figure 2: Error Vs. Image Noise



(a) (b)

Figure 3: First frame and region segmentation

of the Gaussian noise (in pixels). We can see that the proposed method can obtain good estimates over a long sequence.

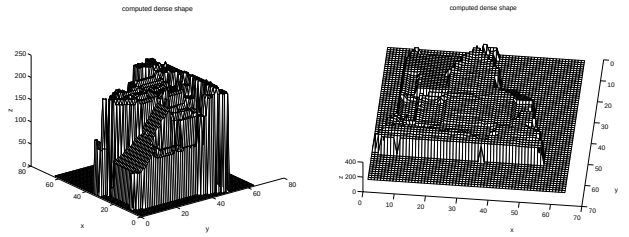
5.2 Real Sequence

Figure 3(a) shows the first frame of a sequence in which a model house is placed on a rotating platform with angular velocity of 1 degree per frame. 10 frames are used for this experiment.

Region detection and region matching is done by the algorithm proposed by Tabb and Ahuja [9, 12]. The algorithm gives the region trajectories over F frames together with the 2-D region motion approximated by an affine model. Figure 3(b) shows the regions obtained for matching by the multiscale segmentation algorithm.

Figure 4(a) and Figure 4(b) show the computed dense shape from two different view directions. Figure 5 shows the image warped on the computed shape.

We can see that reasonably good scene structure has been reconstructed. Due to the fact that some regions are not planar in nature, e.g. the base of the house and the bushes in front of the house, and some regions are not detected consistently from frame to frame, e.g. an adjacent small region may merge with the region, errors in the computed dense shape are noticeable, e.g. boundary continuity between adjacent regions may be violated. Further refinement of the dense shape can be done by either an interpolation or an optimization procedure.



(a) (b)

Figure 4: computed dense shape

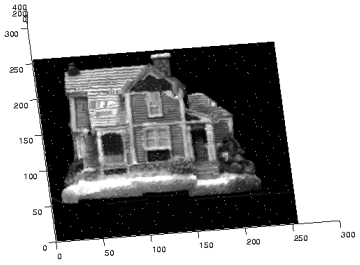


Figure 5: Warp the image onto the computed dense shape

6 Conclusions

We have described a factorization based approach for recovering shape and motion of piecewise planar objects using region measurements of long video sequences. The surface normal of each region can be recovered from region area measurements which enables recovery of dense shape and motion. Compared with existing factorization methods which use constraints along x and y directions, we introduce the constraints in the depth direction. The advantage of the proposed method is that only region correspondences are required. The region correspondences can be more reliably established compared to either point or line correspondences. This method can be naturally integrated with multiscale image segmentation to yield a multiscale shape from motion.

Acknowledgments

The support of the U.S. Army Research Laboratory under cooperative agreement $DAAL01-96-2-0003$ is gratefully acknowledged.

References

[1] C.Debrunner and N.Ahuja. A direct data approximation based motion estimation algorithm. *Proc. 10th Int'l Conf. Pattern Recognition*, pages 384–389, June 1990.

[2] C.Debrunner and N.Ahuja. Segmentation and factorization-based motion and structure estimation for long image sequences. *IEEE Trans.on PAMI*, 20(2):206–211, February 1998.

[3] C.J.Poelman and T.Kanade. A paraperspective factorization method for shape and motion recovery. *IEEE Trans.on PAMI*, 19(3):206–218, March 1997.

[4] C.Tomasi and T.Kanade. Shape and motion from image streams under orthography: A factorization method. *International Journal of Computer Vision*, pages 137–154, September 1992.

[5] C.Y.Lee and D.B.Cooper. Structure from motion: a region based approach using affine transformations and moment invariants. In *Proceedings, IEEE Intl. Conf. on Robotics and Automation*, pages 120–127, 1993.

[6] H.Schweitzer and R.Krishnan. Structure from multiple 2d affine correspondences without camera calibration. In *Proceedings, CVPR'96*, pages 258–262, 1996.

[7] J.Y.Aloimonos. Perspective approximations. *Image and Vision Computing*, 8(3):177–192, August 1990.

[8] L.Quan and T.Kanade. Affine structure from line correspondences with uncalibrated affine cameras. *IEEE Trans.on PAMI*, 19(8):834–845, August 1997.

[9] M.Tabb and N.Ahuja. Multiscale image segmentation by integrated edge and region detection. *IEEE Trans.on Image Processing*, 6(5):642–655, May 1997.

[10] S.Negahdaripoour and S.Lee. Motion recovery from image sequences using first order flow information. In *Proceedings, IEEE Workshop on Visual Motion*, pages 132–139, 1991.

[11] S.Sull and N.Ahuja. Segmentation, matching and estimation of structure and motion of textured piecewise planar surfaces. In *Proceedings, IEEE Workshop on Visual Motion*, pages 274–279, 1991.

[12] M.Tabb. *Multiscale structure detection and its application to image segmentation and motion analysis.* PhD thesis, ECE Dept, University of Illinois at Urbana-Champaign, 1995.

[13] T.Morita and T.Kanade. A sequential factorization method for recovering shape and motion from image streams. *IEEE Trans.on PAMI*, 19(8):858–867, August 1997.