

**THE UNIVERSITY OF MICHIGAN
COMPUTING RESEARCH LABORATORY¹**

**ESTABLISHING CORRESPONDENCE OF
NON-RIGID OBJECTS
USING SMOOTHNESS OF MOTION**

Ramesh Jain and LK. Sethi

CRL-TR-10-84

JANUARY 1984

**Room 1079, East Engineering Building
Ann Arbor, Michigan 48109
USA
Tel: (313) 763-8000**

¹Any opinions, findings, and conclusions or recommendations expressed in this publication are those of the authors.

Establishing Correspondence of Non-Rigid Objects

Using Smoothness of Motion

Ramesh Jain¹

Electrical and Computer Engineering
The University of Michigan
Ann Arbor, MI 48105

and

I. K. Sethi

Department of Computer Science
Wayne State University
Detroit, MI 48202

Abstract

Identifying the same physical object in more than one image, the object correspondence problem, is important in many applications. Most approaches for establishing correspondence in a sequence of frames depend heavily on the 2-D features and 2-D distance in the location of correspondence tokens in *two* frames. By using a sequence of frames, rather than just two or three frames, it is possible to exploit the fact that due to *inertia*, the motion of an object can not change instantaneously. We use *smoothness of motion* of an object for solving the correspondence problem for non-rigid motions. By using *path coherence*, the correspondence problem is formulated as an optimization problem. This paper presents our initial study of the efficacy of the proposed approach.

Index Terms: Correspondence, dynamic scene analysis, structure from motion, path coherence, smoothness of motion.

1. Introduction

Structure from motion has attracted significant research efforts in recently [RoA80, Tod82, TsH81, THZ82, Ull79, WeA80] from researchers working in the field of dynamic scene analysis[Jai84]. Ullman introduced the *rigidity assumption* which states that any set of elements undergoing a 2-D transformation which has a unique interpretation as a rigid body moving in space should be so interpreted. The research in human perception [Ull79] suggests that the human visual system exploits this fact. The rigidity assumption has been very influential in structure from motion research. The research in the recovery of 3-D structure from image sequences may be divided in two general classes. Some researchers have been concerned with the problem of recovery of the structure and motion using a minimal number of points in a minimal number of frames. Recently, the trajectory based recovery has attracted some attention.

1.1. Recovering Structure Using Tokens

Suppose that we apply an interest operator to consecutive frames of a sequence and extract some interesting points or tokens, such as corners, then using some method, manage to succeed in solving the correspondence problem. Ullman shows that if token correspondences have been established, it is possible to recover the 3-D location of four noncoplanar points from their 3 orthogonal projections. This gives an implicit 3-D structure of the object. He shows that if the points are noncoplanar then a unique structure can be recovered; in the case of coplanar points, the structure, to a reflection, can be recovered. For the case of perspective projections, two views of five noncoplanar points are required. The equations for the recovery are nonlinear and require iterative methods for the solution. Williams uses a heuristic method [Wil80] for the recovery. Roach and Aggarwal [RoA80] present results of their efforts to solve the resulting nonlinear equations using standard techniques available in IMSL. In their studies they observed that in the presence of noise, the minimal solution does not give correct results; significant overdetermination (2 views of 15 points, or 3 views of 8 points) may be required.

Tsai and Huang [TsH81,THZ82] introduce eight *pure parameters* for the case of a rigid planar patch undergoing general 3-D motion. Using the Lie group theory of transformations, they

show that for two given successive views the solution is unique. From computational point of view, a very attractive feature of their approach is that the parameters can be computed by solving a set of linear equations. They demonstrate that though theoretically 6 solutions are possible, practically the maximum number of solutions is two. In [THZ82] it is shown that the actual motion parameters can be estimated by computing the singular value decomposition of a 3 X 3 matrix consisting of the eight pure parameters and that the number of solutions is either one or two. This approach has been extended to rigid objects with curved surfaces in [TsH82].

By avoiding uncertain and time consuming iterative methods required for the solution of non linear equations, one can hope to recover the 3-D motion in realistic situations in real time. An effort [JeJ83b] to apply the approach of Tsai and Huang using the IMSL package to recover the motion parameters in a real scene shows, however, that the proposed approach is very sensitive to the location of points. It was observed that a very high precision in the location of the tokens may be required to obtain reliable results.

The feature based methods for the recovery of the structure or for the estimation of the motion parameters require two difficult steps before they are applied: the precise location of the tokens or points, and the correspondence. If we apply interest operators based on small neighborhoods, then the number of tokens extracted in a real image is very large making correspondence a difficult problem. The operators based on a large neighborhood and higher order greylevel characteristics do result in a reasonable number of tokens, reducing the complexity of the correspondence, but their location may not be precise. Even if the location is obtained at the resolution of the pixel, it appears that results obtained using the methods discussed above may not be reliable.

1.2. Trajectory Based Methods

All above methods depend on a set of points in two or three frames. If a token is traced over several frames, by solving correspondences, then one obtains the 2-D trajectory of the point. It appears that the efforts to recover the structure and motion in 3-D from the trajectories may be more reliable than those based on a set of features in a few frames [Joh76,Tod82]. A trajec-

tory may be interpolated by using curve fitting techniques to obtain a better resolution in the 2-D path.

Webb and Aggarwal [WeA82] propose the use of several monocular views for recovering the 3-D structure of moving rigid and jointed objects. They assume that a general motion of objects may be considered, over at least a short interval, as a rotation about a fixed axis and a translation. The fixed axis assumption allows recovery of the structure, under parallel projection, for as few as two points. If a point on the object is fixed then the other points trace out circles in planes normal to the axis of rotation. These circles are projected as ellipses under parallel projection. The structure of the object can be recovered to within a reflection by finding the parameters of the ellipse.

Jointed objects are composed of two or more rigid objects. Webb and Aggarwal [WeA82] present an algorithm, assuming that the feature selection and the correspondence problems have been solved and that the fixed axis assumption is satisfied for each rigid component, for identifying jointed components and for recovering their structure. The most attractive feature of this approach is that the recovery requires fitting an ellipse and hence the recovery method is not as sensitive as the methods based on 2 or 3 frames.

Recently, Sethi and Jain [SeJ83] have extended the trajectory based approach using perspective projections. They have shown that rotation of a point about a fixed axis results in an ellipse in the image plane. For the case of rotation about an axis parallel to the X or Y axis the 3-D coordinates can be recovered, to a scale, using parameters of an ellipse for one point. Their work is inspired by some experiments of Todd [Tod82], who shows that humans appear to be interpreting rigid and non-rigid rotations based on trajectories.

1.3. Correspondence Problem

In the recovery of structure from motion, the establishment of the token correspondence has been a major hurdle. Ullman [Ull79] proposed a minimal mapping approach to this problem. His approach was based on the 2-D features of the tokens and the 2-D distance between tokens. His elegant approach is computationally very attractive because of the possibility of implementation

using a network of processors.

The major problem with the approach is, however, that it fails to consider the fact that most realistic motion analysis must be performed by considering an extended region of space-time. Correspondence based on 2-D distances alone may lead to erroneous results in scenes containing several objects moving in assorted directions and in case of non-rigid motion.

Barnard and Thompson [BaT79] and Prager and Arbib [PrA83] proposed relaxation based approaches to the solution of this problem. In relaxation based methods, all tokens in the frame 2 that are within a spatial neighborhood of a token in the frame 1 are assigned a weight indicating the match strength between the token in the frame 1 and the tokens in the frame 2. It is assumed that only one of these matches is correct. The correct match is obtained by using relaxation. The relaxation process tries to find most consistent matches based on the disparities. The most consistent match implies uniform motion and hence *rigid* objects. These approaches also use 2-D displacements in two frames as a major factor in establishing the correspondence.

Surprisingly, these approaches to the solution of the correspondence problem try to solve the problem using tokens in two frames only. In fact, most research in dynamic scene analysis, or time varying image analysis, has been concerned with only 2 or 3 frames, *not with dynamic scenes*. It is not clear, why most researchers in this field have been concerned with recovery of information considering minimal number of points in minimal number of frames, rather than recovery of reliable information (see Brady's comment on Todd's paper in [Bra83]). It appears that for real world problems, like many other problems in dynamic scene analysis, the correspondence problem may be solved more reliably using more than two frames. Due to the self-imposed limitation of working with only two frames, the rigidity assumption was essential for solving the problem. In two frames the most powerful constraint is certainly the spatial uniformity of displacements. The rigidity assumption, with orthographic projections, is very appealing because it guarantees the uniformity and is domain-independent. If we consider more than two frames in establishing correspondence, however, we may admit non-rigid objects in our analysis. This can be done using *path coherence* and assuming *smoothness of motion*. In this paper we introduce the

notion of *path coherence* and use it for solving the correspondence problem as an optimization problem.

2. Path Coherence and Smoothness of Motion

We exploit the fact that due to inertia, the motion of a physical entity can not change instantaneously. If a frame sequence is acquired at a rate such that no dramatic changes take place between two consecutive frames, then for most physical objects no abrupt change in the motion can be observed. Thus a very reasonable assumption for the analysis of real world dynamic scenes is: *the motion of an object at any time instant can not change abruptly*. This is *path coherence*.

The projection of a smooth 3-D trajectory will be a smooth 2-D trajectory in both orthographic and perspective projections. In a scene several objects may be undergoing random motions. The motion of each object, however, will be smooth. Thus, if a dynamic scene is sampled at a rate fast enough to capture all significant events in the scene, then the observed motion of all objects will be smooth trajectories. As suggested by Todd [Tod82], more reliable motion perception may be achieved by considering the primitive units of motion perception to be the global trajectories in an optic array defined over an extended region of space-time. Based on the above facts, we propose a hypothesis for the analysis of motion in an extended space-time region. This hypothesis, called *smoothness of motion*, may be stated as follows:

If a set of elements undergoing a 2-D transformation has a unique interpretation as a set of elements following smooth 2-D trajectories, then it should be so interpreted.

Clearly, two frames are not enough to apply smoothness of motion. We require at least three frames to compute the change in the motion. For complex motion, more frames will be required. For a demonstration of the generality of the smoothness of motion, let us consider the projection of two points in three frames, as shown in Figure 1. The points are moving in different directions. The three points under consideration are at the crossover of the trajectories of the points. The rigidity assumption can not be used in this situation. Moreover, if distances in two

frames are considered the basis in finding trajectory of the points then one will obtain the abrupt change in the direction of motion from the second frame to the third frame. The path coherence will allow the determination of the correct correspondences over three frames and will result in the correct trajectories, as shown in Figure 2.

Figure 1

In this figure we show the trajectories of two points. We are concerned with the three points on the crossover of the trajectories. The points in the first, second, and third frames are \square , \times , and \triangle , respectively.

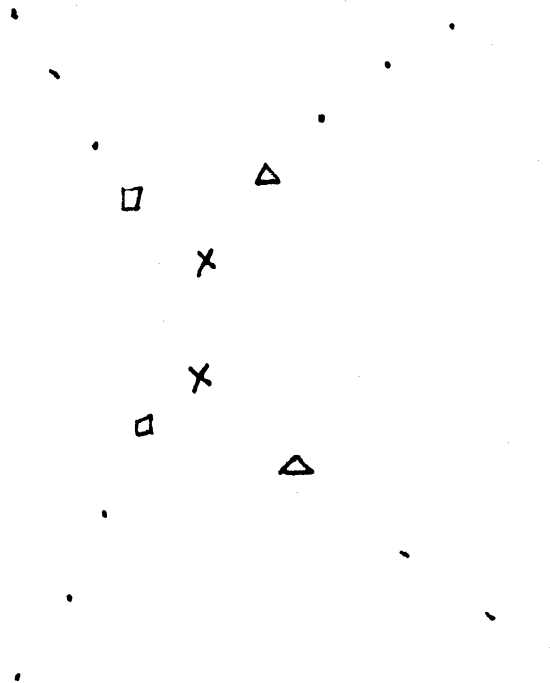
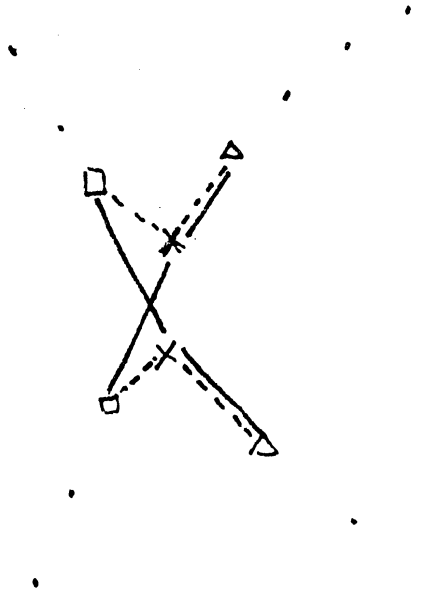


Figure 2

The correspondence established using 2-D distance and the smoothness of motion are shown using broken and full lines, respectively.



If we are considering a trajectory, then for a physical entity the instantaneous change in the motion may be obtained by comparing the motion vectors from the previous frame and the current frame. The path coherence requires that there should not be abrupt changes in the motion vectors obtained from consecutive frames.

Let us consider a trajectory, T_i , for a point P_i in a image sequence. The coordinates of the point are given by a vector X_i in a frame. The coordinate in the k^{th} frame for this point is denoted by $X_{i,k}$. Let us represent a trajectory T_i as

$$T_i = \langle X_{i_1}, X_{i_2}, \dots, X_{i_n} \rangle$$

where $X_{i,k}$ is the point in the k^{th} frame participating in the i^{th} trajectory. A set of points in the k^{th} frame will be denoted by X^k .

Now let us consider the deviation d_i^k in the path of the point in the k^{th} frame. We define a measure of deviation as

$$d_i^k = \Psi(X_{i,k-1}, \vec{X}_{i,k}, X_{i,k}, \vec{X}_{i,k+1})$$

We may define the deviation for the trajectory as

$$D_i = \sum_{k=2}^{n-1} d_i^k$$

Note that for the translational motion, the difference in the direction will be a suitable measure of deviation; in more complex cases one may need more rigorous measures. For translational motion and ideal images, D_i will be zero.

3. Finding Trajectories

In a general but noise free case, all elements may be moving in assorted directions. We can, however, hope to recover elements in each frame by using a feature detector. The problem now is to find trajectories of m points in n frames. We know that

1. An element in a frame can only belong to one trajectory,
2. There should be m trajectories, each containing n points,
3. For each trajectory the deviation should be minimal.

We extend the formulation used by Ullman [Ull79] to solve the correspondence problem. Now we define an n -dimensional array, C , to represent m valid trajectories. This array will have

$$C(i_1, i_2, \dots, i_n) = 1$$

for a valid i^{th} trajectory. All other entries will be zero. Thus for a set of trajectories under consideration, m out of m^n entries in the array C will be 1, rest will be 0.

Now we can formulate the trajectory finding method as an optimization problem. According to the hypothesis of the smoothness of motion, we want to select those trajectories from the set of all possible trajectories that maximize the smoothness of the trajectories. Thus we want to select that set which minimizes the function $E(i_1, i_2, \dots, i_n)$, given by

$$E(i_1, i_2, \dots, i_n) = \sum_{X_1 \in X^1} \dots \sum_{X_n \in X^n} \left(\sum_{k=2}^{n-1} \Psi(X_{k-1} \vec{X}_k, X_k \vec{X}_{k+1}) \right) * C(i_1, i_2, \dots, i_n)$$

Subject to the conditions

$$\sum_{X_{i_1} \in X^1} \alpha(i_1, i_2, \dots, i_n) = 1$$

$$\sum_{X_{i_n} \in X^n} \alpha(i_1, i_2, \dots, i_n) = 1$$

and

$$V(i_1, i_2, \dots, i_n) \geq 0 \text{ for } X_{i_1} \in X^1, \dots, X_{i_n} \in X^n.$$

The set of trajectories that minimizes this function should be interpreted as the correct set describing the motion of m points in n frames. By solving this minimization problem we are solving the correspondence problem using n frames. This problem appears to be similar to the *assignment* problem. The additional complexity arising due to n frames may be handled using dynamic programming. We are investigating computational methods to solve this problem.

4. Experiments

We studied the efficacy of the approach considering synthetic data. The motion of the points was considered random. The measure of deviation used was one minus the dot product of the vectors for the consecutive frame pairs, thus

$$d_i = 1 - X_i^{*T} X_{i+1}^*$$

In the current experiments with synthetic frames, we generated several 3-D trajectories for points moving in assorted directions. Due to computational requirement, in the first study, our experiments have been limited to studying the use of path coherence. We considered only 3 frames at a time and tried to establish correspondence over 3 frames by minimizing the function

$$\sum_{k=1}^n d_k$$

We performed several experiments by considering real and discrete coordinate values in the image plane for the perspective projection of points. In Figure 3, we show a composite frame showing 4 points in 10 frames. The trajectories obtained by considering 3 points at a time are

shown in Figure 4. In this case all correct trajectories were obtained. In some cases, wrong trajectories were obtained. Such a situation is shown in Figure 5. As can be seen, in this case the simple deviation function used is not sufficient due to curvilinear and intersecting motion of points. This function considers only change in the direction of motion; it does not consider the magnitude of the motion vectors. By using better deviation function and smoothness of motion, not just 3 frames, we may be able to establish correct correspondence.

We are in the process of implementing these ideas. We are also studying computational methods for efficient implementation of this approach.

Figure 3 A composite frame shows 4 points in 10 frames. The points start near the center of the image and move outwards.

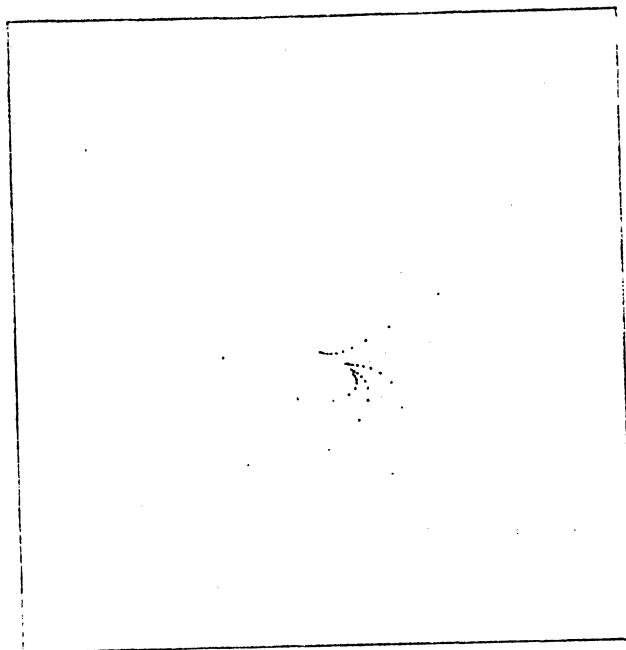


Figure 4 The trajectories established using path coherence are shown in this figure. All trajectories were correctly determined.

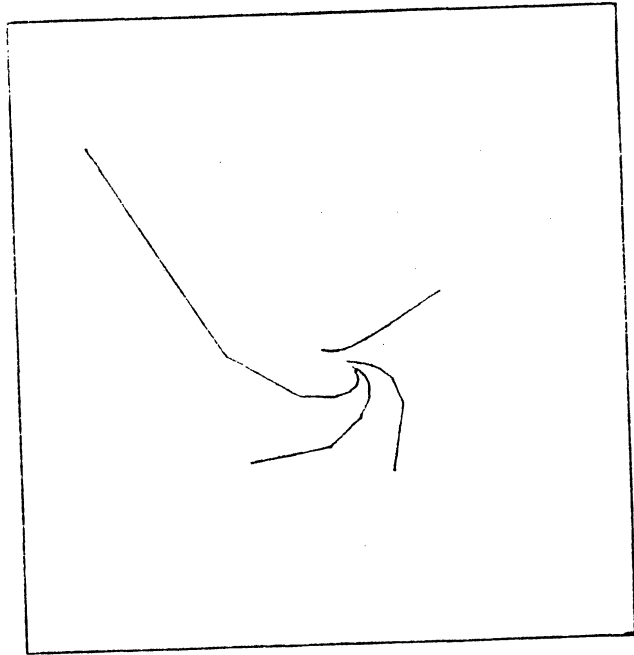
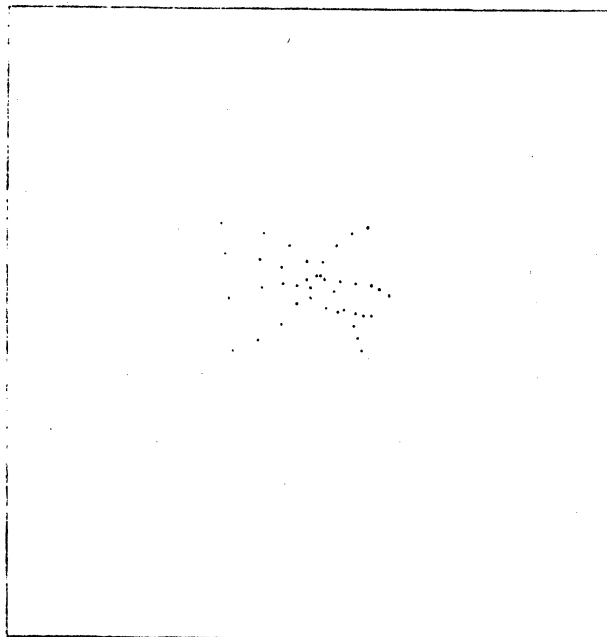


Figure 5 In this case due to the intersecting trajectories, path coherence based on the dot product of motion vectors is not enough. Smoothness of motion with path coherence function using rate of change of direction and the amount of displacement, may result in correct correspondence.



5. Refinements

The proposed minimization approach for solving the correspondence problem requires a minimum of 3 frames. The computational cost for m points in n frames increases exponentially with n . We are studying fast methods for optimization of this problem. In this paper, we do not address the computational methods used in the optimization.

Note that in this approach, we did not use any information about the pictorial features of the token. The tokens in a real world scene will be extracted using a feature detector. The strength and nature of the features may be used to eliminate poor matches from further consideration. The use of features will result in a significant reduction in the computational cost. Moreover, in the presence of noise, features will make this approach robust. We are studying the use of features in this approach.

In the proposed implementation of the smoothness of motion approach, we did not consider the spatial location and the 2-D distances. Clearly, by considering these in the optimization criterion, a more robust function may be defined.

6. Conclusion

This paper presents some of our current ideas, under active study, on establishing correspondence in a sequence of frames containing non-rigid motion. The proposed method uses path coherence to establish correspondence for smooth arbitrary motion of several objects. Our early experiments, though performed using synthetic data and very *simple* deviation function, are very encouraging. We are actively studying several aspects of this approach.

A very important aspect of the proposed approach is that it considers information available in a dynamic scene, rather than trying to determine correspondence using only two frames. By considering extended space-time region, we will be able to extend dynamic scene analysis to more complex motions of non-rigid objects.

REFERENCES

[BaT79]

Barnard, S.T. and W.B. Thompson, "Disparity analysis of images," *IEEE Trans. on PAMI*, vol. PAMI-2, 1980, pp. 333-340.

[Bra83]

Brady, M., "Parallelism in Vision," *Artificial Intelligence*, Vol. 21, 1983, pp.271-283.

[HoS81]

Horn, B. K. P. and B. G. Schunck, "Determining optical flow," *Artificial Intelligence* vol. 17, 1981, pp.185-203.

[Jai83a]

Jain, R., "Direct computation of the focus of expansion," *IEEE Trans. on PAMI*, vol. PAMI-5, pp.58-64, 1983.

[Jai84]

Jain, R., "Dynamic Scene Analysis", in *Progress in Pattern Recognition, Vol. 2, Ed. A. Rosenfeld and L. N. Kanal*, North Holland, 1984.

[JeJ83a]

Jerian, C. P. and R. Jain, "Determining motion parameters for scenes with translation and rotation", *Proc. Workshop on Motion: Representation and Control*, April 4-6, Toronto, 1983.

[Joh76]

Johansson, G., "Spatio-temporal differentiation and integration in visual motion perception", *Psych. Research*, vol. 38, pp.379-383, 1976.

[PrA83]

Prager, J.M., and M. A. Arbib, "Computing the Optic Flow: The MATCH Algorithm and Prediction," *Computer Vision, Graphics, and Image Processing*, col. 24, 1983, pp. 271-304.

[RoA80]

Roach, J.W. and J.K. Aggarwal, "Determining the movement of objects from a sequence of images," *IEEE Trans. on PAMI*, vol. PAMI-2, No.6, Nov. 1980, pp. 554-562.

[SeJ83]

Sethi, I. K. and R. Jain, "Determining three dimensional structure of rotating objects using a sequence of monocular views", *Technical Report, Dept. of Computer Science, Wayne State University, Detroit, 1983.*

[Tod82]

Todd, J. T., "Visual information about rigid and nonrigid motion: A geometric analysis", *J. Experimental Psychology: Human Perception and Performance*, vol. *, pp.238-252, 1982.

[TsH81]

Tsai, R. Y. and T. S. Huang, "Estimaing three-dimensional motion parameters of a rigid planar patch", *Proc of PRIP*, pp. 94-97, 1981.

[TsH82]

Tsai, R. Y. and T. S. Huang, "Uniqueness and estimation of three-dimensional motion parameters or rigid objects with curved surfaces," *Proc. IEEE Conf. Pattern Recog. and Image Processing*, 1982, pp. 112-118.

[THZ82]

Tsai, R. Y., T. S. Huang, and W.L. Zhu, "Estimating three-dimensional motion parameters of a rigid planar patch, II: Singular value decomposition", *IEEE Trans. Acoustics, Speech and Signal Processing*, vol. ASSP-30, pp.525-534.

[Ull79]

Ullman, S., *The interpretation of visual motion* Cambridge, Mass, MIT Press, 1979

[WeA82]

Webb, J. A. and J. K. Aggarwal, "Structure from motion of rigid and jointed objects," *Artificial Intelligence*, vol. 19, pp.107-130, 1982.

[Wil80]

Williams, T.D., "Depth from Motion in Real World Scenes", *IEEE Trans. PAMI*, PAMI-2, 1980, pp. 511-516.