Finding Point Correspondences in Motion Sequences Preserving Affine Structure

G. Sudhir*   Subhashis Banerjee†   Andrew Zisserman‡

Abstract

In this paper the problem of computing the point correspondences in 3-frames of a sequence of time-varying images of a $3D$ object undergoing non-rigid (affine) motion is addressed. It is assumed that the images are obtained through weak-perspective projections. The correspondences are established only from the analysis of the unknown affine structure of the object, without making use of any attributes of the feature points. It is shown that it is possible to establish the point correspondences uniquely (upto symmetry) in the sense that they yield a unique affine structure of the object and that the computation is possible in polynomial time. Further, two different algorithms for computing the point correspondences are presented and theoretical results regarding the correctness of the solutions are provided. Results, on real image data demonstrate that the algorithms are fast and robust.

1   Introduction

Finding correspondences between point configurations in a monocular sequence of time varying images of a 3D object undergoing arbitrary motion is a long standing problem in Computational Vision. Such correspondences are essential for motion tracking, estimation of motion parameters and determining the structure of the 3D object. In this paper we address the problems of i) establishing the point correspondences over three frames of a sequence of weak-perspective views of a single 3D object undergoing affine (non-rigid) motion, and ii) determining the affine structure of the object. Such a solution at the bootstrap stage of tracking of a $3D$ object can be used to predict the images in the subsequent frames making subsequent tracking a simple task.

The correspondence problem and the structure/motion analysis problem have typically been treated separately. On the one hand, most approaches for establishing correspondences between successive frames have relied on the principle that under small motion the image features of corresponding points are similar [1, 2, 3, 4, 5]. On the other hand, most approaches to feature-based motion analysis [6, 7, 8, 9, 10] have assumed that point correspondences between successive frames are already established and have addressed the problem of determining the structure and/or motion. In a recent paper which is significant to our approach, Koenderink and van Doorn [11] have shown that it is possible to uniquely determine the affine structure of a 3D non-rigid object from the point correspondences in two weak-perspective views. The successes of these approaches are crucially dependent on the correctness of the assumed correspondence.

*Centre for Applied Research in Electronics, Indian Institute of Technology, New Delhi 110016. Email: sudhir@care.iitd.ernet.in

†Dept. Computer Science and Engineering, Indian Institute of Technology, New Delhi 110016. Email: suban@cse.iitd.ernet.in

‡Dept. Engineering Science, University of Oxford, Oxford OX1 3PJ Email: az@robots.ox.ac.uk

In contrast, we derive the constraints for the correspondence process from the unknown structure itself. In fact, the primary aim of the present investigation is to examine to what extent the correspondences can be established only from such constraints. The use of local structures/features can only enhance the performance of our method.

In a similar approach Lee and Huang [12] have shown that, for the case of rigid-body motion and weak-perspective projections, given the correspondences of four non-coplanar (of which any three are non-collinear) points in two frames, the matches of all other points are restricted to lie along a specific straight line. They have used this constraint to give a polynomial time algorithm (exhaustive) to establish the point correspondences between two frames. However, this epipolar restriction only provides a necessary condition for the correspondences of all other points. We show that under the assumptions of affine (non-rigid) motion and weak-perspective projections, by considering the three frame correspondence problem centered on the middle frame, it becomes possible to derive additional constraints to make the correspondences unique (upto symmetry). Based on these constraints we present two alternative schemes for solving the three frame correspondence problem. The first is a fast polynomial time backtracking algorithm for the traversal of the search tree. Our second scheme is a randomized algorithm, based on the relaxation labeling framework, which invokes the constraints to solve the three frame correspondence problem. We provide theoretical results regarding the correctness of our algorithms and compare the applicability of the two algorithms under different situations.

The main assumptions of our paper are i) The motion of the object between any two time frames is an arbitrary 3D affine transformation, and ii) the views are weak-perspective projections of the 3D object (a $3D$ affine transformation followed by orthographic projection).

The rest of the paper is organized as follows. In Sec. 2 we bring out the relationship between the correspondence problem and the affine structure and show that the three frame correspondence problem has a unique solution (upto symmetry). In Sec. 3 we describe the search algorithm for solving the three frame correspondence problem and examine its time complexity. In Sec. 4 we describe our randomized algorithm and prove its correctness. In Sec. 5 we present results for some real image data.

2   The Relationship between the Affine Structure and Motion Correspondence

In this section we examine how the motion correspondence problem is related to the unknown affine structure of a $3D$ non-rigid object undergoing affine motion.

Let $A = (A_1, A_2, \ldots, A_n)$, $B = (B_1, B_2, \ldots, B_n)$ and $C = (C_1, C_2, \ldots, C_n)$ be three n point configurations of a $3D$ object in affine motion. Thus these configurations are affine transformations of each other. Let $\bar{A} = (\bar{A}_1, \bar{A}_2, \ldots, \bar{A}_n)$, $\bar{B} = (\bar{B}_1, \bar{B}_2, \ldots, \bar{B}_n)$, $\bar{C} = (\bar{C}_1, \bar{C}_2, \ldots, \bar{C}_n)$ be three orthographic projections of A, B and C respectively. Thus, $\bar{A}$, $\bar{B}$ and $\bar{C}$ are weak perspective views of the object in affine motion. Without loss of generality, we assume that $\bar{A}$ is the middle frame and $\bar{B}$, $\bar{A}$ and $\bar{C}$ is the sequence of time-varying images under consideration, and address the problem of simultaneously establishing the correspondences between $\bar{A}$ to $\bar{B}$ and $\bar{A}$ to $\bar{C}$.

In what follows, we derive the constraints for establishing the correspondences from the unknown affine structure of the object. To this end, we first bring out the fact that, under general motion, it is possible to derive the affine structure of the object, uniquely, from the point correspondences between any two views, say $\bar{A}$ and $\bar{B}$ [11].

2.1   The Affine Structure from Motion Theorem

Note, that two sets of 2D point configurations, say $\bar{A}$ and $\bar{B}$, are similar (equivalent upto uniform scaling and rotation) if the affine motion between A and B is a case of fronto-parallel translation or a rotation about an axis parallel to the viewing direction. We call

such motions degenerate. In this subsection, we describe the affine structure from motion theorem considering only non-degenerate motions. The treatment follows [11].

Let the point correspondences in $\bar{A}$ and $\bar{B}$ be given and let $\bar{A}_1$, $\bar{A}_2$ and $\bar{A}_3$ be three non-collinear points. Since they are projections of the world points A_1, A_2 and A_3 respectively, we can assign the $3D$ affine coordinates $(0,0,0),(1,0,0)$ and $(0,1,0)$ to A_1, A_2 and A_3, and this assignment completely fixes the $2D$ affine basis for the plane formed by A_1, A_2 and A_3. If $f(.)$ is the $2D$ affine transformation induced between the images $\bar{A}$ and $\bar{B}$ by the correspondence of the first three points (i.e., $f(\bar{A}_i) = \bar{B}_i$ for $i = 1, 2, 3$), then, the vector $\bar{B}_4 - f(\bar{A}_4)$ along the image plane can be taken as the third basis vector for the $3D$ object. In fact, it is easy to show [11] that repeating the above operation with a fifth point A_k, we obtain a vector $\bar{B}_k - f(\bar{A}_k)$, which must be parallel to $\bar{B}_4 - f(\bar{A}_4)$. Thus, it is possible to assign unique affine coordinates (α, β, γ) to any other world point A_k where the coordinates α and β are obtained by expressing $\bar{A}_k$ as an affine combination of $\bar{A}_1$, $\bar{A}_2$ and $\bar{A}_3$, and γ is given by the ratio of the lengths of the parallel vectors $\bar{B}_k - f(\bar{A}_k)$ and $\bar{B}_4 - f(\bar{A}_4)$.

2.2 The Constraints for 3-frame, Affine Structure Preserving Correspondence

Our objective is to simultaneously obtain the point correspondences between set $\bar{A}$ to $\bar{B}$ and set $\bar{A}$ to $\bar{C}$ such that the affine structure of the 3D object is preserved. In other words, the affine structure derived from the correspondences in $\bar{A}$ to $\bar{C}$ must be identical to that derived from $\bar{A}$ to $\bar{B}$. This observation and the affine structure from motion result described in the previous sub-section gives us the following theorem.

Theorem 2.1 *Consider non-degenerate motions from B to A and A to C. Suppose that the three frame correspondences for four non-coplanar points are given such that* $(\bar{B}_1, \bar{B}_2, \bar{B}_3, \bar{B}_4)$ —— $(\bar{A}_1, \bar{A}_2, \bar{A}_3, \bar{A}_4)$ —— $(\bar{C}_1, \bar{C}_2, \bar{C}_3, \bar{C}_4)$. *Let* $f()$ *and* $g()$ *be the 2D affine transformations induced by the correspondences of the first three points, i.e.,* $f(\bar{A}_i) = \bar{B}_i$ *and* $g(\bar{A}_i) = \bar{C}_i$ *for* $i = 1, 2, 3$. *Then the necessary and sufficient conditions for* $\bar{B}_k$ *and* $\bar{C}_k$ *to be corresponding points of an arbitrary fifth point* $\bar{A}_k$ *(such that a unique affine structure is obtained) are*

(a) $\bar{B}_k - f(\bar{A}_k)$ *is parallel to* $\bar{B}_4 - f(\bar{A}_4)$

(b) $\bar{C}_k - g(\bar{A}_k)$ *is parallel to* $\bar{C}_4 - g(\bar{A}_4)$

(c) $\frac{\bar{B}_k - f(\bar{A}_k)}{\bar{B}_4 - f(\bar{A}_4)} = \frac{\bar{C}_k - g(\bar{A}_k)}{\bar{C}_4 - g(\bar{A}_4)}$

While conditions (a) and (b) in the above theorem specify the matching directions for points $\bar{B}_k$ and $\bar{C}_k$ respectively, the third condition fixes the $3D$ affine coordinate of the point A_k uniquely and guarantees the preservation of the affine structure. Any three frame correspondence which satisfies the above three conditions for all possible choice of 5 points yields a unique affine structure for the object. Thus, these conditions can be viewed as constraints for the correspondence process. Lee and Huang [12] consider the 2-frame correspondence problem and give the condition (a) in the above theorem as a necessary condition. In fact, it becomes necessary to consider the 3-frame correspondence problem to obtain the sufficient condition also. However, if either of the motions between B to A or A to C is degenerate, the third basis vector cannot be found at all, and, consequently, the constraints derived from the above theorem cannot be applied. In such a case the $3D$ affine structure of the object cannot be determined. The correspondence problem then reduces to a $2D$ affine point matching problem between two frames. The resulting constraint is then given by the following theorem.

Theorem 2.2 *Consider degenerate motion from B to A. Suppose the two frame correspondences for 3 non-collinear points are given such that* $(\bar{B}_1, \bar{B}_2, \bar{B}_3)$ —— $(\bar{A}_1, \bar{A}_2, \bar{A}_3)$. *Let* $f()$ *be the 2D affine transformation induced by the correspondences of these three*

points,i.e., $f(\bar{A}_i) = \bar{B}_i$ for $i = 1, 2, 3$. Then the necessary and sufficient condition for $\bar{B}_k$ to be the corresponding point of an arbitrary fourth point $\bar{A}_k$ is $f(\bar{A}_k) = \bar{B}_k$.

It is important to note that while the constraints derived from the above theorems guarantee a unique affine structure ($2D$ or $3D$), they by no means guarantee that there is a unique way of establishing the 3-frame correspondence. If the object has some inherent symmetry then more than one correspondence can satisfy the constraints derived from the above theorems. However, all such correspondences will yield a unique affine structure and will be considered correct for our purpose.

In what follows we present our algorithms for establishing the three frame correspondence. We assume that either both the motions (from B to A and A to C) are degenerate or both are non-degenerate. If only one of them is non-degenerate, then no structure preserving constraints can be invoked and the correspondence problem will have to be solved as two independent 2-frame problems.

3 The Search Algorithm for the Correspondence Problem

In what follows we describe the search algorithm for obtaining a solution to the correspondence problem and examine its time complexity. The algorithm is similar to the alignment method in object recognition [13]. We assume that each point $\bar{A}_i$ has k potential matches in each of B and C on the average.

Algorithm

1. Choose four points in $\bar{A}$ such that any three are non-collinear. There are $O(k^4)$ possible matches of these four points in either B or C. Thus there are $O(k^8)$ possible three frame correspondences of these four points. Each of these correspondences define the affine basis for each of the two 2-frame correspondences.

2. Check each one of these $O(k^8)$ four point correspondences to see whether the affine structure (either $3D$ if the four point correspondence yields the third basis vector, or $2D$ if it doesn't) is preserved for any of the $O(k^2)$ possible three frame correspondences of a pre-determined fifth point $\bar{A}_i \in \bar{A}$. Go to Step 3 when such a four point correspondence is found.

3. With the three frame correspondence of four points found in Step 2 attempt to determine an affine structure preserving three frame correspondence (from the $O(k^2)$ possible ones) for each of the remaining $n - 5$ points. If affine structure cannot be preserved for more than 20% of these points, backtrack to step 2 and select another three frame correspondence of the four points.

The worst case complexity for Step 2 in the above algorithm is $O(k^{10})$, and the complexity of Step 3 is $O(nk^2)$. Thus worst case complexity, when backtracking is necessary for each of the three frame correspondences of four points, is $O(nk^{10})$. Note that in such a case exhaustive enumeration occurs and with this time complexity it is possible to find all possible structure preserving three frame correspondences. However, the preservation of the affine structure in the three frame correspondence of five points almost always indicates a correct correspondence, and the average time complexity of the algorithm to find one structure preserving correspondence is only $O(k^{10} + nk^2)$. Note that this is a pessimistic bound because many of the $O(k^8)$ four point mapping are usually not one-to-one and can be discarded without any further checks.

The algorithm presented above is fast and reliable as verified by our results presented in Sec. 5. However it relies heavily on the initial choice of four points in $\bar{A}$ (Step 1). This may pose problems in practice in the case of spurious or missing points. Considering different initial choice of four points in $\bar{A}$ increases the time complexity considerably. In view of this, we present a randomized algorithm in the next section which is driven by repeated random choices of the four basis points.

4 A Relaxation Algorithm for the Correspondence Problem

In this section, we present an algorithm for solving the 3-frame correspondence problem of finding a mapping $M : \bar{A} \longrightarrow B \times C$ such that a unique affine structure can be obtained. We pose the correspondence problem in the framework of Relaxation Labeling (RL) [14] and solve it using a cooperating team of Learning Automata (LA) [15]. To pose the 3-frame correspondence problem in the framework of relaxation labeling, we consider two coupled relaxation processes. For both the processes we identify the object set O as $\bar{A} = (\bar{A}_1, \bar{A}_2, \ldots, \bar{A}_n)$. For the first process we identify the label sets as $\Lambda^i = B \; \forall i$, and for the second process we identify the label sets as $\Lambda^i = \bar{C} \; \forall i$. The weighted label assignments for the two processes are given as

$$\underline{p1} = \left\{ p1_i(\lambda_i^B), \; \lambda_i^B \in B, \; \sum_{j=1}^{n} p1_i(\lambda_j^B) = 1 \right\}_{i=1,2,\ldots,n}$$

$$\underline{p2} = \left\{ p2_i(\lambda_i^C), \; \lambda_i^C \in C, \; \sum_{j=1}^{n} p2_i(\lambda_j^C) = 1 \right\}_{i=1,2,\ldots,n} \tag{1}$$

where, $p1_i(\lambda_i^B)$ and $p2_i(\lambda_i^C)$ are the probabilities of choosing the labels λ_i^B and λ_i^C for object $\bar{A}_i$ from the sets B and C by the two processes respectively. Thus, for each $\bar{A}_i$, the labels are chosen from the sets B and C independently.

We solve the relaxation labeling problem using a cooperative team of Learning Automata [15]. A Learning Automaton can be considered as a simple decision making device. It has a set of actions and the decision it has to make is choosing an action. The automaton functions in a random environment which supplies a random reaction for every choice made by the automaton. The automaton learns to choose that action which has the highest expected value of environmental reaction. For this, the automaton maintains a probability distribution over the set of actions and at each instant it chooses an action at random depending on the probability distribution. Then, using the random reaction supplied by the environment, it updates the action probability distribution through a Learning Algorithm.

For solving the labeling problem using a team of Learning Automata, we associate a Learning Automaton with each object O_i and identify its action set as the label set Λ^i. Let $p_i^{(k)} = (p_i^{(k)}(\lambda_1^i), \ldots, p_i^{(k)}(\lambda_{d_i}^i))$ be the action probability distribution of the automaton O_i (Object and Automaton used synonymously) at instant k. Here, $p_i^{(k)}(\lambda)$ is the probability of O_i choosing the action (label) λ at instant k.

The automata team functions as follows. Each automaton O_i starts with a initial probability distribution $p_i^{(0)}$ over its label set. At each instant k, each of the automata O_i chooses a label at random depending on $\underline{p}^{(k)}$. This results in a tentative labeling of the objects by the team. For this choice, a reaction which depends on the constraints is supplied to each automaton O_i. Let β_{iq} denote the response to O_i when it chooses the action λ_q^i. $\beta_{iq} = 0$ or 1 (penalty/reward) depending on whether the constraints are violated or satisfied.

Using this response β_{iq}, each automaton O_i updates its action probability distribution using the L_{R-I} (Linear Reward-Inaction) [15] learning algorithm described by the following equation.

$$p_i^{k+1}(\lambda_q^i) = p_i^k(\lambda_q^i) + a[1 - p_i^k(\lambda_q^i)]\beta_{iq}^k$$

$$p_i^{k+1}(\lambda_r^i) = p_i^k(\lambda_r^i) - a\beta_{iq}^k p_i^k(\lambda_r^i) \quad \forall r \neq q \tag{2}$$

Here $0 < a < 1$ is a constant.

A convergence result for the above algorithm is given in [15]. In what follows we describe the LA algorithm for establishing the three frame correspondence. **Algorithm**

Start with initial labeling $\underline{p}^{(0)}$ such that
$$p1_i^{(0)} \text{ and } p2_i^{(0)} \text{ are uniform distributionss over the label sets}$$
Set k = 0.

1. Selection of labels

for each $\bar{A}_i (i = 1, 2, \ldots, n)$ do (simultaneously)
choose a label $\lambda_i^B \in \Lambda^B$ according to $p1_i^{(k)}$
choose a label $\lambda_i^C \in \Lambda^C$ according to $p2_i^{(k)}$

2. Generation of responses

for each $\bar{A}_i$ do
randomly (uniform) select four other object points $\bar{A}_j \bar{A}_k \bar{A}_l \bar{A}_m$
such that $i \neq j \neq k \neq l \neq m$ and $\bar{A}_i \bar{A}_j \bar{A}_k$ are non-collinear

Let $f()$ be the 2D affine transformation that maps
$\bar{A}_i \bar{A}_j \bar{A}_k$ to $\lambda_i^B \lambda_j^B \lambda_k^B$

Let $g()$ be the 2D affine transformation that maps
$\bar{A}_i \bar{A}_j \bar{A}_k$ to $\lambda_i^C \lambda_j^C \lambda_k^C$

Let $V_1 \longrightarrow \lambda_l^B - f(\bar{A}_l)$, $\quad V_2 \longrightarrow \lambda_m^B - f(\bar{A}_m)$
$\quad\quad V_3 \longrightarrow \lambda_l^C - g(\bar{A}_l)$, $\quad V_4 \longrightarrow \lambda_m^C - g(\bar{A}_m)$

Case

a) $V_1 = V_2 = V_3 = V_4 = 0$: (degenerate case)
$$\beta_{i\lambda_i^B} \longrightarrow \beta_{i\lambda_i^C} \longrightarrow 1$$

b) $V_1 \parallel V_2$, $V_3 \parallel V_4$ and $\frac{V_2}{V_1} = \frac{V_4}{V_3}$: (non-degenerate case)
$$\beta_{i\lambda_i^B} \longrightarrow \beta_{i\lambda_i^C} \longrightarrow 1$$

else
$$\beta_{i\lambda_i^B} \longrightarrow \beta_{i\lambda_i^C} \longrightarrow 0$$

end.

3. Update $\underline{p1}^{(k)}$ and $\underline{p2}^{(k)}$ through the L_{R-I} scheme described in Eqn. 2

4. $k = k + 1$
if $k > T$ (a pre-determined number) stop else GO TO 1.

It can be shown [16] that an object point $\bar{A}_i$ can have incorrect labels λ_i^B and λ_i^C (($\lambda_i^B \neq \bar{B}_i$) or ($\lambda_i^C \neq \bar{C}_i$)) only if there exists $(k-1)$ other object points such that these $(k-1)$ points together with $\bar{A}_i$ preserve some affine structure (distinct from the correct one) and $k \geq 5$, $k > m$, where m is the number of object points with correct labels.

5 Results

In this section, we present results for some real test data. However, to apply the algorithms presented above in practical situations, we make several modifications. The epipolar constraint for matching can never be strictly satisfied because the weak-perspective approximation is always violated by real images. Consequently, to determine whether vectors $f(\bar{A}_4) - B_4$ and $f(\bar{A}_k) - B_k$ are parallel, we test whether the perpendicular projection of the point $f(\bar{A}_k) - B_k$ on to the direction given by $f(\bar{A}_4) - B_4$ is less than a
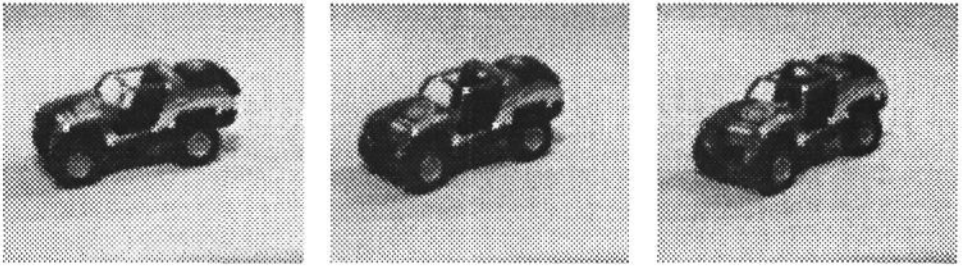
Figure 1: The three frames of a toy jeep undergoing non-degenerate motion. The bright crosses are the corners considered for the 3-frame correspondence
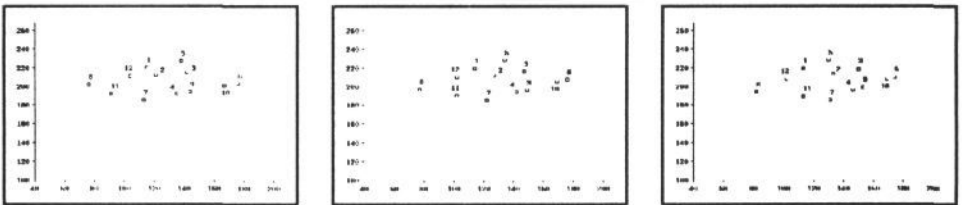

Figure 2: The numbered corners in the three frames for the toy jeep undergoing non-degenerate motion

threshold T_1. We also assume that a vector $f(\bar{A}_k) - \hat{B}_k$ is the zero vector if its distance from the origin is less than same threshold T_1. Further, we restrict the label sets for the point $\bar{A}_i$ to only those points in B and C which fall within a fixed radius (say R) of $\bar{A}_i$. This restriction not only speeds up the algorithms significantly, but also limits the number of possible correspondences by breaking the symmetry. In all our experiments we assume such restrictions. However, note that our algorithms are perfectly valid even without this restriction.

Example 5.1 Our first example is with real image data of a toy jeep undergoing non-degenerate motion. The images were grabbed from sufficient distance to ensure that the weak perspective approximation is not severely violated. The three images are given in Fig. 1 with the corners considered for the 3-frame correspondence marked as bright crosses. These points were selected from the output of a state of the art corner detection algorithm [17]. These corners with their number labels are shown in Fig. 2. The label set for each point in the middle frame was restricted to those falling within a 30 pixel radius in the other two frames. With this restriction, the typical size of each label set was 3 to 5. A threshold of 5 pixels was used to detect the parallelism and a threshold of 0.2 was used to check the equality of the third coordinates. The LA algorithm was run for 300 iterations which took about 1.5 minutes on a SUN spark station. The search algorithm took about 15 secs on the average with different rearrangements of the input list. The results are given in Table 1.

Example 5.2 Our second example is with real image data of a toy jeep undergoing degenerate motion. The three images are given in Fig. 3 with the corners considered for the 3-frame correspondence marked as bright crosses. These corners with their number labels are shown in Fig. 4. The various parameters chosen for this example were identical to the previous one. Correct convergence was obtained for all the object points and the results are given in Table 2.

Table 1: Results on real perspective images of a toy jeep undergoing nondegenerate motion

Points in Middle Frame	Label Set (First Frame)	Label Set (Third Frame)	Search Final Label Pairs	Relaxation Final Label Pairs
1	{1,2,5,12}	{1,2,12}	(1,1)	(1,1)
2	{1,2,5}	{1,2,3,5}	(2,2)	(2,2)
3	{2,3,9}	{3,5}	(3,3)	(3,3)
4	{4,7,9}	{4,9}	(4,4)	(4,4)
5	{2,3,5}	{3,5}	(5,5)	(5,5)
6	{6,10}	{6,10}	(6,6)	(6,6)
7	{7,11}	{4,7}	(7,7)	(7,7)
8	{8}	{8,11}	(8,8)	(8,8)
9	{4,9}	{3,4,9,10}	(9,9)	(9,9)
10	{6,9,10}	{6,10}	(10,10)	(10,10)
11	{11,12}	{7,11}	(11,11)	(11,11)
12	{1,12}	{1,2,12}	(12,12)	(12,12)

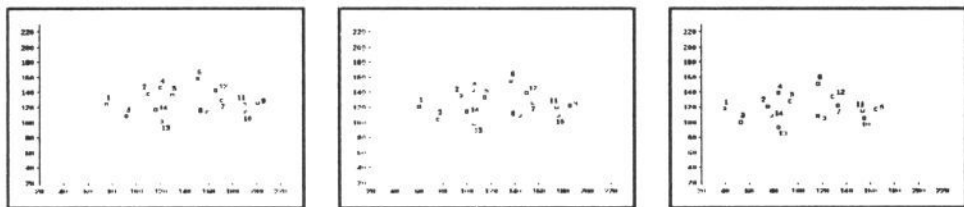**Figure 3**: The three frames for the toy jeep undergoing degenerate motion

Figure 4: The numbered corners in the three frames for the toy jeep undergoing degenerate motion

Table 2: Results on real perspective images of a toy jeep undergoing degenerate motion

Points in Middle Frame	Label Set (First Frame)	Label Set (Third Frame)	Search Final Label Pairs	Relaxation Final Label Pairs
1	{1,2,3,14}	{1}	(1,1)	(1,1)
2	{2,4,5,6}	{1,2,3,4,14}	(2,2)	(2,2)
3	{2,3,13,14}	{1,3}	(3,3)	(3,3)
4	{4,5,6,10}	{2,4,5,14}	(4,4)	(4,4)
5	{5,6,7,8,10}	{2,4,5,14}	(5,5)	(5,5)
6	{6,10}	{4,5,6}	(6,6)	(6,6)
7	{7,9,10,11,12}	{5,7,8,12}	(7,7)	(7,7)
8	{7,8,9,10,11,12}	{8,13,14}	(8,8)	(8,8)
9	{9}	{7,8,9,10,11,12}	(9,9)	(9,9)
10	{9,10,11}	{7,8,10,11}	(10,10)	(10,10)
11	{9,10,11}	{7,8,9,10,11,12}	(11,11)	(11,11)
12	{7,9,11,12}	{5,6,7,8,12}	(12,12)	(12,12)
13	{8,13}	{3,13,14}	(13,13)	(13,13)
14	{4,5,8,13,14}	{1,2,3,13,14}	(14,14)	(14,14)

6   Conclusion

We have addressed the problem of computing the point correspondences in 3-frames of a sequence of time-varying images of a $3D$ non-rigid object undergoing affine motion. We have assumed that the images are obtained using weak-perspective projections. We have shown that it is possible to compute the 3-frame point correspondences uniquely (upto symmetry) using only the constraints derived from the unknown affine structure of the object (see also [18, 19]). Further, this computation is possible in polynomial time. We have presented two algorithms for computing the 3-frame correspondences. Our first algorithm is based on a deterministic search technique while the second is based on a stochastic relaxation framework. We have also presented results on some real image data.

The primary aim of this paper has been to examine to what extent the point correspondences can be established only from an analysis based on the unknown affine structure of the object. Hence we have not made use of any local attributes of the feature points. However, the similarity of local features can be used to enhance the performance of both our algorithms. In the search algorithm, the point correspondences based on local features can be used to determine the four initial points and the sequence in which the search can be conducted. In the relaxation algorithm similarity of local features can be used to bias the initial probability distributions for the coupled relaxation processes towards the correct matches. Thus the use of local features can increase the speed of computation for both our algorithms.

The basic computation in both our algorithms is the verification of the preservation of affine structure for a three frame correspondence of five points. In the Search algorithm this operation has to be performed $O(k^{10} + nk^2)$ times. In practice, however, the Search algorithm works much faster than this because many of the $O(k^8)$ four point mappings are not one-to-one and can be discarded straight away. Further, it is seldom necessary to check all of these four point mappings, and the correct one can be found much faster. In the Relaxation Algorithm this operation has to be performed $O(nT)$ times (where T is the number of iterations). But the relaxation algorithm requires the expensive floating point computations for the generation of random samples and the update of probability vectors and is the slower of the two. However, the relaxation algorithm is inherently parallel and can be implemented very efficiently on an array processor. Further, the Relaxation algorithm has the advantage that in the relaxation process the preservation

of the affine structure is verified by random choices of five points. Since different affine bases are chosen randomly, the algorithm is less susceptible to spurious/missing points. If for a particular object point the corresponding point in any one of the other views is missing, then the automaton corresponding to the object point will not get any support and consequently will not converge. Correct convergence for the subset of points which are present in all the three views can still be obtained. In contrast, in the search algorithm the four basis points remain fixed, and, consequently, it is less robust to spurious/missing points. We have obtained encouraging results in our experiments with spurious/missing points [16] which we will present in a forthcoming paper. Current work extends the ideas to scenes with multiple objects. The simultaneous solution of the correspondence problem and the structure from motion problem presented in this paper can be considered as a first step towards affine invariant model based recognition of $3D$ objects [18].

References

[1] S.T. Barnard and W.B. Thompson., Disparity Analysis of Images., *IEEE Trans. on Pattern Anal. and Machine Intell.*, 2:333–340, 1980.

[2] C. Wang, H. Sun, S. Yada, and A. Rosenfeld., Some Experiments in Relaxation Image Matching using Corner Features., *Pattern Recognition*, 16:167–182, 1983.

[3] H.J. Lee and H.C. Deng., Three-Frame Corner Matching and Moving Object Extraction in a Sequence of Images., *Computer Vision Graphics and Image Processing*, 52:210–238, 1990.

[4] G.G. Pieroni and M.F. Costabile., A Method for Detecting Correspondences in a Sequence of Modifying Shapes., *Pattern Recognition Letters*, 6:403–412, 1985.

[5] N.S. Robin and M. Zuhua., Computing Correspondences in a Sequence of Non-Rigid Shapes., *Pattern Recognition*, 25:901–911, 1992.

[6] S. Ullman., *The Interpretation of Visual Motion.*, MIT Press, Cambridge, USA, 1979.

[7] C.H. Lee and T. Huang., Motion and Structure from Orthographic Projections., *IEEE Trans. on Pattern Anal. and Machine Intell.*, 11:536–540, 1989.

[8] X. Zhuang, T.S. Huang, and R.M. Haralick. A Simple Procedure to Solve Motion and Structure from Three Orthographic Views., *IEEE Trans. on Robotics and Automation*, 4:236–239, 1988.

[9] X Hu and N. Ahuja., Motion Estimation under Orthographic Projection. *IEEE Trans. on Robotics and Automation*, 7:848–853, 1991.

[10] O.D. Faugeras and S.J. Maybank. Motion from Point Matches: Multiplicity of Solutions., In *IEEE Workshop on Motion*, pages 248–255, 1989.

[11] J.J. Koenderink and A.J. van Doorn., Affine Structure from Motion., *J. of Opt. Soc. of Am. Series A*, 8:377–385, 1991.

[12] C.H. Lee and T. Huang., Finding Point Correspondences and Determining Motion of a Rigid Object from Two Weak Perspective Views., *Computer Vision Graphics and Image Processing*, 52:309–327, 1990.

[13] Y. Lamdan, J.T. Schwartz, and H.J. Wolfson., On Recognition of 3D Objects from 2D Images., In *Proc. ICRA*, pages 1407–1413, 1988.

[14] R.A. Hummel and S.W. Zucker., On the Foundations of Relaxation Labelling Processes., *IEEE Trans. on Pattern Anal. and Machine Intell.*, 5:267–286, 1983.

[15] M.A.L. Thathachar and P.S. Sastry., Relaxation Labelling by a Team of Learning Automata., *IEEE Trans. on Pattern Anal. and Machine Intell.*, 6:256–268, 1986.

[16] G. Sudhir, S. Banerjee, K.K. Biswas, and R. Bahl. A Relaxation Algorithm for Point Correspondences Preserving Affine Structure., Technical report, Dept. Computer Science and Engg., IIT, Delhi, 1993.

[17] H. Wang and M. Brady., Corner Detection with Sub-pixel Accuracy., Technical report, University of Oxford, 1992.

[18] J.L. Mundy and A. Zisserman., *Geometric Invariance in Computer Vision.*, MIT Press, 1992.

[19] S. Demey, A. Zisserman, and P. Beardsley., Affine and Projective Structure from Motion., In *British Machine Vision Conf.*, pages 49–58, 1992.