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 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 affine projections. The correspondences are established only from the analysis of the unknown 3D 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 (up to symmetry) in the sense that they yield a unique affine structure of the object and that the computation is possible in polynomial time. Two different algorithms for computing the point correspondences are presented. Results on various real image sequences, including a sequence containing independently moving objects, demonstrate the applicability of the structure based approach to motion correspondence.

[*]Dept. Computer Science, Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong.

Email: sudhir@cs.ust.hk

[†]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

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 establishing the point correspondences in a sequence of affine views of a single $3D$ object undergoing affine (non-rigid) motion, and computing its affine structure, in a single framework [14, 15]. 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 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], and track each corner individually [10] without regard to the inter-relation of the corners among themselves, i.e. structure. On the other hand, most approaches to structure from motion analysis and motion parameter estimation [7, 17, 19] have assumed that point correspondences between successive frames are already established and have addressed the problem of determining the structure and/or motion. Koenderink and van Doorn [6] have shown that it is possible to uniquely determine the affine structure of a $3D$ non-rigid object from the point correspondences in two affine views. The successes of these approaches are crucially dependent on the correctness of the assumed correspondence.

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

In a similar approach Lee and Huang [8] 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. Shashua [13] also uses this epipolar constraint and derives additional constraints from the optic flow equation to localize the matches on the epipolar. In contrast, we consider the three frame correspondence problem centered on the middle frame and derive additional constraints to make the correspondences unique (up to symmetry). We first propose a deterministic backtracking algorithm for solving the three frame correspondence problem. Though the exhaustive computation of the three frame correspondences is possible in polynomial time, the method is crucially dependent on the initial choice of the first four points. Thus, any localization or detection errors in these four points may lead to further errors in subsequent computations. In view of this, we propose a randomized algorithm based on the relaxation labeling framework, which uses a principle of random sample consensus and relies on repeated random choice of the initial points to establish the correspondences. Once the structure is initialized at the bootstrap stage of a tracking process using the 3-frame method, subsequent point correspondences can be computed using 2-frames at a time.

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. In Sec. 3 we describe the search algorithm for solving the correspondence problem and examine its time complexity. In Sec. 4 we describe our randomized algorithm, and finally, in Sec. 5, we present experimental results and analysis on some real image data.

2 The Relationship between the Affine Structure and Motion Correspondence

Consider a set of n $3D$ world points $\mathbf{X} = \{\mathbf{X_i},\ \ i = 0, \ldots, n-1\}$ in affine (non-rigid) motion and let $\mathbf{x}, \mathbf{x}', \mathbf{x}'', \ldots$ be a sequence of affine views [9] of the $3D$ point configuration. Thus, for example, $\mathbf{x_i}$ is the affine projection of the $3D$ point $\mathbf{X_i}$ in the first view. In what follows we briefly describe the notion of affine structure as introduced by Koenderink and Van Doorn [6] and examine how the motion correspondence problem is related to the unknown affine structure of a $3D$ non-rigid object undergoing affine motion.

2.1 Affine Structure from Motion

Given the image correspondences on n points in two views $\mathbf{x_i} \leftrightarrow \mathbf{x'_i}$, for $i = 0 \ldots n-1$, the $3D$ affine structure of each world point $\mathbf{X_i} = (\alpha_i, \beta_i, \gamma_i)$ can be computed in the following way [6]. The first three points (assume non-collinear) $\mathbf{x_0}, \mathbf{x_1}, \mathbf{x_2}$ define a basis for the reference plane formed by the $3D$ points $\mathbf{X_0}, \mathbf{X_1}, \mathbf{X_2}$. The (α, β) coordinates of any world point $\mathbf{X_i}$ can be found by expressing the image point $\mathbf{x_i}$ as an affine combination in this basis.

$$\begin{aligned}\mathbf{x_i} - \mathbf{x_0} &= \alpha_i(\mathbf{x_1} - \mathbf{x_0}) + \beta_i(\mathbf{x_2} - \mathbf{x_0}) &&= \alpha_i \mathbf{e_1} + \beta_i \mathbf{e_2} \\ \mathbf{x'_i} - \mathbf{x'_0} &= \alpha_i(\mathbf{x'_1} - \mathbf{x'_0}) + \beta_i(\mathbf{x'_2} - \mathbf{x'_0}) &&= \alpha_i \mathbf{e'_1} + \beta_i \mathbf{e'_2}\end{aligned} \tag{1}$$

Here, $(\mathbf{e_1}, \mathbf{e_2})$ and $(\mathbf{e'_1}, \mathbf{e'_2})$ are the affine projections of the basis vectors $(\mathbf{E_1}, \mathbf{E_2}) = (\mathbf{X_1} - \mathbf{X_0}, \mathbf{X_2} - \mathbf{X_0})$ defining the reference plane. (α_i, β_i) are also the $2D$ coordinates of $\hat{\mathbf{X}}_\mathbf{i}$, which is the projection of $\mathbf{X_i}$ onto the reference plane along the viewing direction in the first image. Given a fourth world point $\mathbf{X_3}$ which is away from the reference plane, the vector from $\mathbf{X_3}$ to the reference plane along the viewing direction in the first image can be chosen as the third basis vector $\mathbf{E_3}$ for the $3D$ affine space. The projection of the third basis vector into the first image is degenerate, reducing it to a single point. However, in the second image, the projection of the third axis vector, $\mathbf{e'_3}$, is no longer degenerate. Using the first two axes vectors, $\mathbf{e'_1}$ and $\mathbf{e'_2}$, the projection of a world point $\mathbf{X_i}$ in the second image can be predicted, as if the point $\mathbf{X_i}$ is on the reference plane defined by $\{\mathbf{E_1}, \mathbf{E_2}\}$:

$$\hat{\mathbf{x}}'_\mathbf{i} = \mathbf{x'_0} + \alpha_i \mathbf{e'_1} + \beta_i \mathbf{e'_2} \tag{2}$$

The difference between this predicted position and the observed position $\mathbf{x'_i} = \mathbf{x'_0} + \alpha_i \mathbf{e'_1} + \beta_i \mathbf{e'_2} + \gamma_i \mathbf{e'_3}$ is solely due to the γ_i component

$$\mathbf{x'_i} - \hat{\mathbf{x}}'_\mathbf{i} = \gamma_i \mathbf{e'_3} \tag{3}$$

Clearly, all these disparity vectors must be parallel and this can be used to test the validity of affine projection and affine motion.

Thus, given the correspondences of four non-coplanar points, it is possible to assign unique affine coordinates $(\alpha_i, \beta_i, \gamma_i)$ to any other point $\mathbf{X_i}$ where the coordinates α_i and β_i are obtained by expressing $\mathbf{x_i}$ in the affine basis $\{\mathbf{e_1}, \mathbf{e_2}\}$ and γ_i is obtained from Eqn. 2 and Eqn. 3.

However, if the point configurations in the two images are related by a $2D$ affine transformation, i.e., $\mathbf{x'_i} = \Gamma \mathbf{x_i} + \epsilon$, where Γ is 2x2 matrix and ϵ is a 2-vector then the third affine coordinate

γ_i cannot be found at all. This can happen if the motion is degenerate, a cyclo-rotation or a fronto-parallel translation; or if the structure is degenerate, a planar object.

2.2 Constraints for Affine Structure Preserving Correspondence

Consider non-degenerate motions from $\mathbf{X}'$ to $\mathbf{X}$ and $\mathbf{X}$ to $\mathbf{X}''$. Suppose that the three frame correspondences for four non-coplanar points are given such that $\mathbf{x'_j} \leftrightarrow \mathbf{x_j} \leftrightarrow \mathbf{x''_j}$ for $j = 0 \ldots 3$. Then the *necessary and sufficient* conditions for $\mathbf{x'_i}$ and $\mathbf{x''_i}$ to be corresponding points of an arbitrary fifth point $\mathbf{x_i}$ (such that a unique affine structure is obtained) are

(a) $\mathbf{x'_i} - \hat{\mathbf{x'_i}}$ is parallel to $\mathbf{e'_3}$

(b) $\mathbf{x''_i} - \hat{\mathbf{x''_i}}$ is parallel to $\mathbf{e''_3}$

(c) there exists a constant γ_i such that $\mathbf{x'_i} - \hat{\mathbf{x'_i}} = \gamma_i \mathbf{e'_3}$ and $\mathbf{x''_i} - \hat{\mathbf{x''_i}} = \gamma_i \mathbf{e''_3}$

While conditions (a) and (b) specify the matching directions for points $\mathbf{x'_i}$ and $\mathbf{x''_i}$ respectively, the third condition fixes the $3D$ affine coordinate of the point $\mathbf{x_i}$ uniquely and guarantees the preservation of the affine structure. In fact the epipolar conditions of (a) and (b) above are not sufficient to uniquely establish the 2-frame correspondences. Lee and Huang [8] show that even for rigid objects, there always exists a motion to account for any extended correspondence on the epipolar, and consequently, the ambiguities on the epipolar cannot be resolved. In general, if the affine structure is unknown, it is necessary to consider the 3-frame correspondence problem and use condition (c) for establishing the correspondences uniquely. However, if the affine structure $(\alpha_i, \beta_i, \gamma_i)$ is known for any point, then 2 frames are sufficient to establish the correspondences uniquely using the above.

If either of the motions between $\mathbf{X}'$ to $\mathbf{X}$ or $\mathbf{X}$ to $\mathbf{X}''$ is degenerate, 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. Then the *necessary and sufficient* condition for $\mathbf{x'_i}$ to be the corresponding point of an arbitrary fourth point $\mathbf{x_i}$ is $\mathbf{x'_i} = \hat{\mathbf{x'_i}}$.

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 [5]. We describe the algorithm for the bootstrap stage of the tracking process where we require 3 frames and use the constraints described in the previous section to initialize the structure. We assume that, on the average, each point $\mathbf{x_i}$ has k potential matches in a region in each of the preceding and succeeding frames and m potential matches on the epipolar lines. Note that typically m is much smaller than k. We further assume that, for 3-frame correspondences, either both motions are degenerate or both motions 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 two frame problems.

Algorithm

1. Choose four points in the middle frame, $\mathbf{x_0}, \mathbf{x_1}, \mathbf{x_2}, \mathbf{x_3}$, such that any three are non-collinear. There are $O(k^4)$ possible matches of these four points in either the previous or the next frame. 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 correspondences yield the third basis vector, or $2D$ if it doesn't) is preserved for any of the $O(m^2)$ possible three frame correspondences of a fifth point $\mathbf{x_i}$. Here first check the epipolar constraint and apply the test for the invariance of the third coordinate only if the epipolar constraint is satisfied in both frames. 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 $O(m^2)$ possible ones) for each of the remaining $n-5$ points.

4. If affine structure cannot be preserved for more than, say, 20% of these points (backtrack threshold in such a case is 0.8), backtrack to step 2 and select another three frame correspondence of the four points.

5. If a satisfactory correspondence is still not found, backtrack to step 1 and select another set of 4 points.

The worst case complexity for step 2 of the above algorithm is $O(k^8m^2)$, and the complexity of step 3 is $O(nm^2)$. However, once the structure is initialized, subsequent correspondences can be established considering only two frames at a time. In such a case there are only $O(k^4)$ matches of the four points in Step 1. The correct match of the four points completely defines the match of a fifth point in Step 2. Once such a basis is found it has to be verified for the remaining $n-5$ points. Thus for 2-frame correspondences with known structure, the worst case complexity for step 2 of the above algorithm is $O(k^4)$, and the complexity of step 3 is $O(n)$. Thus once structure is initialized, subsequent tracking requires only 2-frames and is not a computationally expensive process.

In the absence of any localization and detection errors of corner detectors, the preservation of affine structure in the three frame correspondences of five points almost always indicates a correct correspondence and the average time complexity of the algorithm to find one structure preserving 3-frame correspondence is much less than $O(k^8m^2 + nm^2)$. Note that this is a pessimistic bound because many of the $O(k^8)$ four point mappings are usually not one-to-one and can be discarded without any further checks. Further, m, the average number of matching candidates on the epipolar for each point is usually quite small. This was adequately supported by our simulations on synthetic data.

However, the method is crucially dependent on the initial choice of the first four points. In presence of noise, which is inevitable in real images, any localization and detection errors of the first four points lead to errors in subsequent computations. For establishing the three frame correspondences it is imperative that the four basis points be detected correctly in the 3 frames. Further, their localization errors should also be small to ensure the preservation of affine structure for majority of the remaining points within acceptable tolerances. Such errors lead to back-tracking to step 1 a large number of times to choose different sets of four points. Since there are $O(n^4)$ ways of choosing the first four points, the process becomes computationally expensive. Experimental results in Sec. 5 demonstrate this point. In view of this, we present a randomized algorithm in the next section which is driven by repeated random choices of four basis points.

4 A Relaxation Algorithm for the Correspondence Problem

In this section, we present a randomized algorithm for solving the 3-frame correspondence problem such that a unique affine structure can be obtained. The algorithm is not biased to the initial choice of the basis and it establishes the correspondences using a random sample consensus principle by repeated random choice of the basis points. We pose the correspondence problem in the framework of Relaxation Labeling (RL) [4] and solve it using a cooperating team of Learning Automata (LA) [16]. 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 $\mathbf{x} = \{\mathbf{x_i}, \quad i = 0 \ldots n-1\}$. For the first process we identify the label sets as $\mathbf{x}' = \{\mathbf{x'_j}, \quad j = 0 \ldots n-1\} \;\; \forall i$, and for the second process we identify the label sets as $\mathbf{x}'' = \{\mathbf{x''_j}, \quad j = 0 \ldots n-1\} \;\; \forall i$. We assume, for convenience, that the number of points are the same in all the three frames. However, the algorithm is valid even if they are different. The weighted label assignments for the two processes, at any instant k, are given as

$$
\begin{aligned}
\underline{p1}^{(k)} &= \left\{ p1_i^{(k)}(\lambda_j^{\mathbf{x}'}),\; \lambda_j^{\mathbf{x}'} \in \mathbf{x}',\; \sum_{j=1}^{n} p1_i^{(k)}(\lambda_j^{\mathbf{x}'}) \;=\; 1 \right\}_{i=1,2,\ldots,n} \\
\underline{p2}^{(k)} &= \left\{ p2_i^{(k)}(\lambda_j^{\mathbf{x}''}),\; \lambda_j^{\mathbf{x}''} \in \mathbf{x}'',\; \sum_{j=1}^{n} p2_i^{(k)}(\lambda_j^{\mathbf{x}''}) \;=\; 1 \right\}_{i=1,2,\ldots,n}
\end{aligned}
\tag{4}
$$

where, $p1_i^{(k)}(\lambda_j^{\mathbf{x}'})$ and $p2_i^{(k)}(\lambda_j^{\mathbf{x}''})$ are the probabilities, at instant k, of choosing the labels $\lambda_j^{\mathbf{x}'}$ and $\lambda_j^{\mathbf{x}''}$ for object (point) $\mathbf{x_i}$ from the sets $\mathbf{x}'$ and $\mathbf{x}''$ by the two processes respectively, and n is the number of labels in each set. Thus, for each $\mathbf{x_i}$, the labels are chosen from the sets $\mathbf{x}'$ and $\mathbf{x}''$ independently.

We solve the relaxation labeling problem using a cooperative team of Learning Automata [16]. 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 (point) $\mathbf{x_i}$ and identify its action set as label set $\mathbf{x}'$ and/or $\mathbf{x}''$. Let $p_i^{(k)} = (p_i^{(k)}(\lambda_1), \ldots, p_i^{(k)}(\lambda_n))$ be the action probability distribution of the automaton $\mathbf{x_i}$ (Object and Automaton used synonymously) at instant k. Here, $p_i^{(k)}(\lambda_j)$ is the probability of $\mathbf{x_i}$ choosing the action (label) λ_j at instant k.

The automata team functions as follows. Each automaton $\mathbf{x_i}$ starts with an initial probability distribution $p_i^{(0)}$ over its label set. At each instant k, each of the automata $\mathbf{x_i}$ 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 $\mathbf{x_i}$. Let β_{iq} denote the response to $\mathbf{x_i}$ when it chooses the action λ_q. $\beta_{iq} = 0$ or 1 (penalty/reward) depending on whether the constraints are violated or satisfied.

Using this response β_{iq}, each automaton $\mathbf{x_i}$ updates its action probability distribution using the L_{R-I} (Linear Reward-Inaction) [16] learning algorithm described by the following equation.

$$\begin{aligned} p_i^{k+1}(\lambda_q) &= p_i^k(\lambda_q) + a[1 - p_i^k(\lambda_q)]\beta_{iq}^k \\ p_i^{k+1}(\lambda_r) &= p_i^k(\lambda_r) - a\beta_{iq}^k p_i^k(\lambda_r) \quad \forall r \neq q \end{aligned} \tag{5}$$

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

In relaxation labeling terminology, a labeling $\underline{p}$ is said to be

(i) *unambiguous* if all components of $\underline{p}$ are unit vectors where for every object the probability corresponding to any *one* label is 1 and the probabilities corresponding to the other labels are 0, and,

(ii) *consistent* if for every $\mathbf{x_i}$, the expected value of the response is maximized by the chosen label; i.e., $E[\beta_{iq}] \geq E[\beta_{is}]$ for any $s \neq q$.

The LA algorithm for relaxation labeling converges to an unambiguous and consistent labeling if there exists one. The complete convergence result is available in [16, 11].

Algorithm

Start with initial labeling such that

$p1_i^{(0)}$ and $p2_i^{(0)}$ are uniform distributions over the label sets

Set $k = 0$.

1. Selection of labels

   for each $\mathbf{x_i}$ $(i = 1, 2, \ldots, n)$ do (simultaneously)

   choose a label $\lambda_i^{\mathbf{x}'}$ according to $p1_i^{(k)}$

   choose a label $\lambda_i^{\mathbf{x}''}$ according to $p2_i^{(k)}$

2. Generation of responses

   for each $\mathbf{x_i}$ do

   randomly (uniform) select four other object points $\mathbf{x_0}, \mathbf{x_1}, \mathbf{x_2}, \mathbf{x_3}$ such that $\mathbf{x_0}, \mathbf{x_1}, \mathbf{x_2}$ are non-collinear

   Let the current labels for $\mathbf{x_i}, \mathbf{x_0}, \mathbf{x_1}, \mathbf{x_2}, \mathbf{x_3}$ in sets $\mathbf{x}'$ and $\mathbf{x}''$ be $\mathbf{x'_i}, \mathbf{x'_0}, \mathbf{x'_1}, \mathbf{x'_2}, \mathbf{x'_3}$ and $\mathbf{x''_i}, \mathbf{x''_0}, \mathbf{x''_1}, \mathbf{x''_2}, \mathbf{x''_3}$ respectively

   Let $V_1 = \mathbf{x'_i} - \hat{\mathbf{x}}'_{\mathbf{i}}, \qquad V_2 = \mathbf{x'_3} - \hat{\mathbf{x}}'_{\mathbf{3}} = \mathbf{e'_3}$

   $V_3 = \mathbf{x''_i} - \hat{\mathbf{x}}''_{\mathbf{i}}, \qquad V_4 = \mathbf{x''_3} - \hat{\mathbf{x}}''_{\mathbf{3}} = \mathbf{e''_3}$

   (see Eqn. 2)

   Case

   a) $V_1 = V_2 = V_3 = V_4 = 0$: (degenerate case)

   $$\beta_{i\lambda_i^{\mathbf{x}'}} \longleftarrow \beta_{i\lambda_i^{\mathbf{x}''}} \longleftarrow 1$$

   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)

   $$\beta_{i\lambda_i^{\mathbf{x}'}} \longleftarrow \beta_{i\lambda_i^{\mathbf{x}''}} \longleftarrow 1$$

   else

   $$\beta_{i\lambda_i^{\mathbf{x}'}} \longleftarrow \beta_{i\lambda_i^{\mathbf{x}''}} \longleftarrow 0$$

   end.

3. Update $\underline{p1}^{(k)}$ and $\underline{p2}^{(k)}$ through the L_{R-I} scheme described in Eqn. 5

4. $k = k + 1$

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

In the above algorithm we have assumed that either both the motions (from $\mathbf{X}'$ to $\mathbf{X}$ and $\mathbf{X}$ to $\mathbf{X}''$) are degenerate or both are non-degenerate.

We analyze the correctness of the above algorithm with the following observations:

Let $\underline{p1}$ and $\underline{p2}$ be the unambiguous labelings obtained by the above algorithm. Then,

1. If the 3-frame correspondence denoted by $\underline{p1}$ and $\underline{p2}$ preserves a unique affine structure then $\underline{p1}$ and $\underline{p2}$ are consistent.

2. Suppose $\underline{p1}$ and $\underline{p2}$ are consistent labelings. Then in the 3-frame correspondence denoted by $\underline{p1}$ and $\underline{p2}$ an object point $\mathbf{x_i}$ can have incorrect labels only if there exists $(k-1)$ other object points such that these $(k-1)$ points together with $\mathbf{x_i}$ preserve some affine structure and $k \geq 5$, $k > m$, where m is the number of object points with correct labels.

The proof of the above is straightforward [14] and is omitted here.

According to the second part of the above, an object point $\mathbf{x_i}$ can have incorrect labels only if there exist other object points, some of these with incorrect labels, so that these points together preserve some affine structure distinct from the correct one, and more over, the number of points which preserve the spurious affine structure has to be more than the number of points which preserve the correct affine structure. Though such a situation is possible for symmetrical object, it is extremely unlikely otherwise, since all the constraints of Sec. 2.2 will have to be satisfied by the spurious affine structure. This is adequately supported by the experimental results presented in the next section.

5 Experiments and Results

In this section, we present some experimental results on examples of i) non-degenerate motion, ii) degenerate motion, and, finally, iii) a scene with multiple objects moving independently. In each case, the images were obtained from a sufficient distance to ensure that the affine camera approximation is valid.

For each of our examples, we use the Plessey corner detector [3] to detect the corner points. To minimize the number of spurious corners, we use a large σ of the order of 2.5 to 3.5 for Gaussian smoothing. As a consequence, though we obtain satisfactory detection, the localization of the corners are poor, often resulting in a displacement of up to 2 to 3 pixels.

Since the affine camera is only an approximation to real imaging situations, the constraints described in Sec. 2.2 can never be satisfied exactly. Thus, we decide that the vectors $\mathbf{x}'_\mathbf{i} - \hat{\mathbf{x}}'_\mathbf{i}$ and $\mathbf{e}'_\mathbf{3}$ are parallel (epipolar check) if their normalized dot product is close to 1. We use a threshold T_1 for this check. We use the same procedure for checking the non-collinearity of the 3 basis points. We also assume that a vector is the zero vector if its distance from the origin is less than a threshold T_2, and use this to determine the preservation of affine structure. In all our experiments we choose T_1 to be 0.9 and T_2 to be 3 pixels.

Example 5.1 Rigid non-degenerate motion

Our first example is of a sequence of a toy jeep undergoing non-degenerate motion. The jeep was moved without moving the wheels to ensure rigidity. In Fig. 1 we show three frames of the sequence, the detected corners and the affine flow vectors $(\mathbf{x}'_\mathbf{i} - \hat{\mathbf{x}}'_\mathbf{i})$ for $(i \geq 3)$ computed using the relaxation algorithm for 3-frame correspondence. For visual clarity, we repeat the middle frame twice. We also show the bases formed by the first three points in each frame and the axes of rotation. The axes of rotation are computed after making the necessary corrections for cyclo-rotation and overall scaling [6]. The estimated cyclo-rotations were -0.8 degrees and 2.5 degrees between the second to first and second to third frames respectively. The corresponding overall scalings were estimated to be 1.1 and 1.02.

In Fig. 2 we show the affine flow vectors and the axes of rotation computed from a different basis using identical point correspondences. Note that the estimated direction of the affine flow vectors are grossly wrong and the affine structure is also not preserved using this basis. This is due to the localization errors of the 3 basis points. Hence, the structure based motion analysis becomes unreliable when the computations are carried out with respect to a single basis. A mechanism for selecting different bases combinations is imperative for the success of the structure based motion analysis. Such a mechanism is built-in to the relaxation algorithm due to repeated random choice of the basis, resulting in robustness with respect to the localization and detection errors. On the other hand, for the search algorithm, it is necessary to test a large number of choices for the initial five points till a satisfactory structure preserving correspondence is obtained. Consequently, on the average, the search algorithm is the slower of the two. We observe similar effects in our other examples also.

In Fig. 3 we show the 2-frame correspondence results (Optical flow [6]) obtained after the structure is initialized. For comparison, we also show the correspondences obtained by feature tracking using cross correlation of intensity values over 5x5 windows placed on the corners.

Example 5.2 Non-rigid, non-degenerate motion

In Fig. 4 we show 3-frames of a sequence of a face in nod motion, the detected corners, the affine flow vectors computed using the relaxation algorithm, the first 3 points of the affine bases, and the axes of rotation. Note that the affine flow vectors, which are projections of the rotation are mostly vertical.

Example 5.3 Rigid degenerate motion

Our third example is of a toy van undergoing pure translation. In Fig. 5 we present the 3-frame correspondences obtained using the relaxation algorithm. The search algorithm also gives similar results. This case demonstrates that the correspondence algorithms can automatically take care of degenerate motion situations. Since the motion is degenerate, the affine flow is essentially zero.

Example 5.4 Multiple rigid motions

Our final example is of a sequence of a scene consisting of three toy vehicles undergoing independent degenerate and non-degenerate motions. Note that in this situation there is no single affine transformation that can account for the motions and consequently, the constraints described in Sec. 2.2 do not hold. However, the relaxation algorithm which is based on the preservation of affine structure over repeated random choices of only 5 points can still be applied. Accidental choices of 5 point correspondences in the same object can still fetch rewards, and consequently the algorithm can find independent clusters of affine structure preserving correspondences. Though, in such a situation, considerably more number of iterations are required. In Fig. 6 we show the 3-frame correspondences computed using the relaxation algorithm. The search algorithm which attempts to find a correspondence which preserves a single affine structure cannot directly be applied to this situation. However, it can also be suitably modified to deal with multiple rigid motions. See [18] for a related approach.

In all the above experiments, the relaxation algorithm took about 1 minute on a HP-9000/735 workstation. The number of iterations required for convergence was about 2000 (with the parameter a being 0.05) for all examples except for the one with multiple objects in which case the number of iterations required were 10000 (with the parameter a being 0.02).

The performance of the search algorithm for initializing structure varied according to the localization errors in the initial bases points. In case of low localization errors it took a few seconds but the run time increased considerably when the localization errors were high. In some cases, for high backtrack threshold of 0.9, it failed to give correspondence results altogether. Once the structure was initialized, the 2-frame correspondence algorithm took only a few seconds to establish correspondence in subsequent frames.

6 Conclusion

We have presented two algorithms for computing the 3-frame point correspondences based on constraints derived from the unknown affine structure of the object. Our first algorithm is based on a deterministic search technique while the second is based on a stochastic relaxation framework. Although the two algorithms have been presented as though completely disparate, the use of a robust estimator [12] can combine the strength of both.

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 and the affine multiple views geometry. 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 to achieve faster convergence. Thus the use of local features can supplement our method. The methods presented in this paper can be used to quickly validate the results obtained by any correspondence algorithm that uses local features.

The basic computation in both our algorithms is the verification of the preservation of affine structure for a correspondence of 5 points. The relaxation algorithm has the advantage that in the relaxation process, the preservation of the affine structure is verified by random choices of the 5 points resulting in robust estimation of structure and correspondence. Since different affine bases are chosen randomly, the algorithm is less susceptible to spurious/missing points and the localization errors of the corner detectors. In fact, as our experiments clearly demonstrate, the success of such a structure based motion analysis is crucially dependent on a mechanism to switch bases to account for the localization/detection errors [10]. It is imperative that any algorithm based on such constraints must have this feature. Such a mechanism can also be built into the search algorithm using a robust estimator such as the random sample consensus (RANSAC). Note that the complexity of search algorithm indicated in this paper is the worst case, and on the average the seacrh algorithm performs much better.

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 [9].

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] W.E.L. Grimson. Computational Experiments with a Feature Based Stereo Algorithm. *IEEE Trans. on Pattern Anal. and Machine Intell.*, 7:17–34, 1985.

[3] C. Harris and M. Stephens. A Combined Corner and Edge Detector. In *Proc. 4th Alvey Vision Conference*, pages 153–158, 1988.

[4] 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.

[5] D. P. Huttenlocher and S. Ullman. Object Recognition using Allignment. In *Proc. Intl. Conference on Computer Vision*, 1987.

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

[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] 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.

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

[10] I.D. Reid and D.W. Murray. Tracking Foveated Corner Clusters Using Affine Structure. In *Proc. Intl. Conference on Computer Vision*, 1993.

[11] P.S. Sastry and M.A.L. Thathachar. Analysis of Stochastic Automata Algorithm for Relaxation Labeling. *IEEE Trans. on Pattern Anal. and Machine Intell.*, 16:538–543, 1994.

[12] L.S. Shapiro and J.M. Brady. Rejecting Outliers and Estimating Errors in an Orthogonal Regression Framework. *Phil. Tran. Royal Soc. Lond. A*, 350:407–439, 1995.

[13] A. Shashua. Correspondence and Affine Shape from two Orthographic Views: Motion and Recognition. Technical report, A.I.Memo No. 1327, MIT, 1991.

[14] G. Sudhir. *Cooperative Algorithms for Stereo and Motion Analysis.* PhD thesis, Center for Applied Research in Electronics, Indian Institute of Technology, New Delhi, India, 1993.

[15] G. Sudhir, S. Banerjee, and A. Zisserman. Finding Point Correspondences in Motion Sequences Preserving Affine Structure. In *Proc. British Machine Vision Conf.*, 1993.

[16] 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.

[17] C. Tomasi and T. Kanade. Shape and Motion from Image Streams under Orthography: a Factorization Method. *Intl. J. Computer Vision*, 9:137–154, 1992.

[18] P.H.S. Torr, A. Zisserman, and D.W. Murray. Motion Clustering Using the Trilinear Constraint over Three Views. In *Proc. of Workshop on Geometrical Modeling and Invariants for Computer Vision, Xidian Univ. Press*, 1995.

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

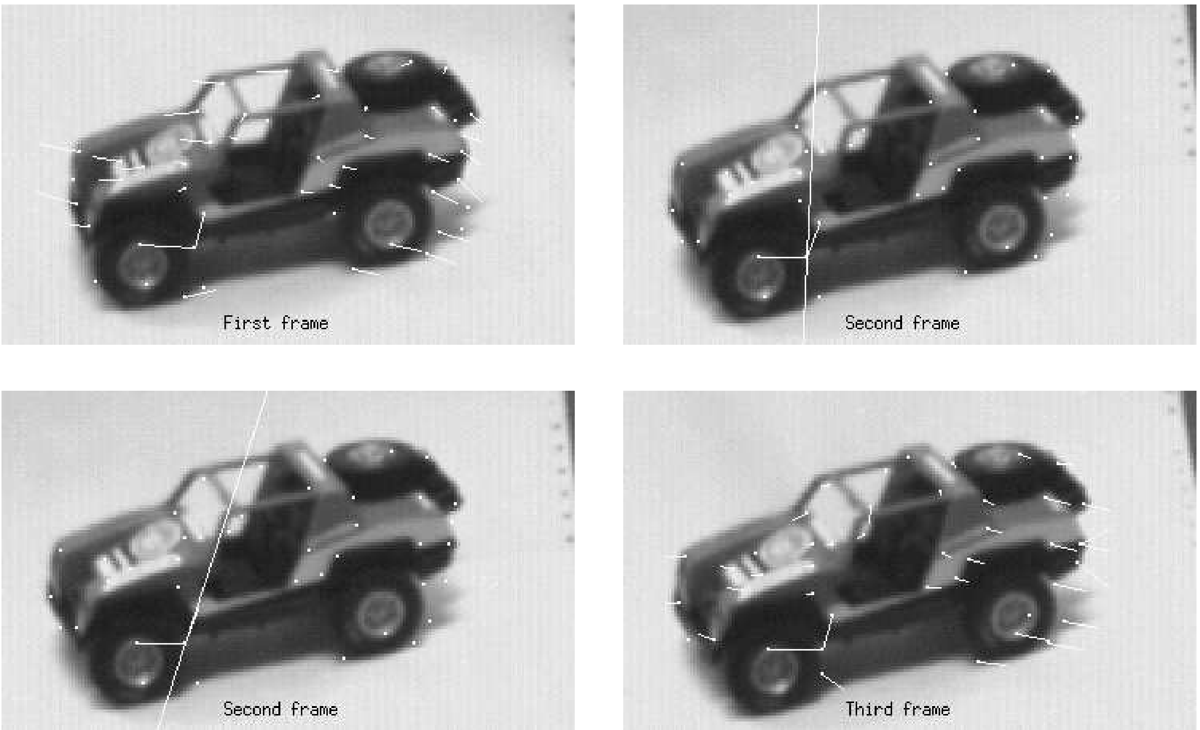

Figure 1: The affine flow obtained from the correspondence results of the relaxation algorithm. The 2D basis used for computation of affine coordinates and the estimated axes of rotation are also shown.

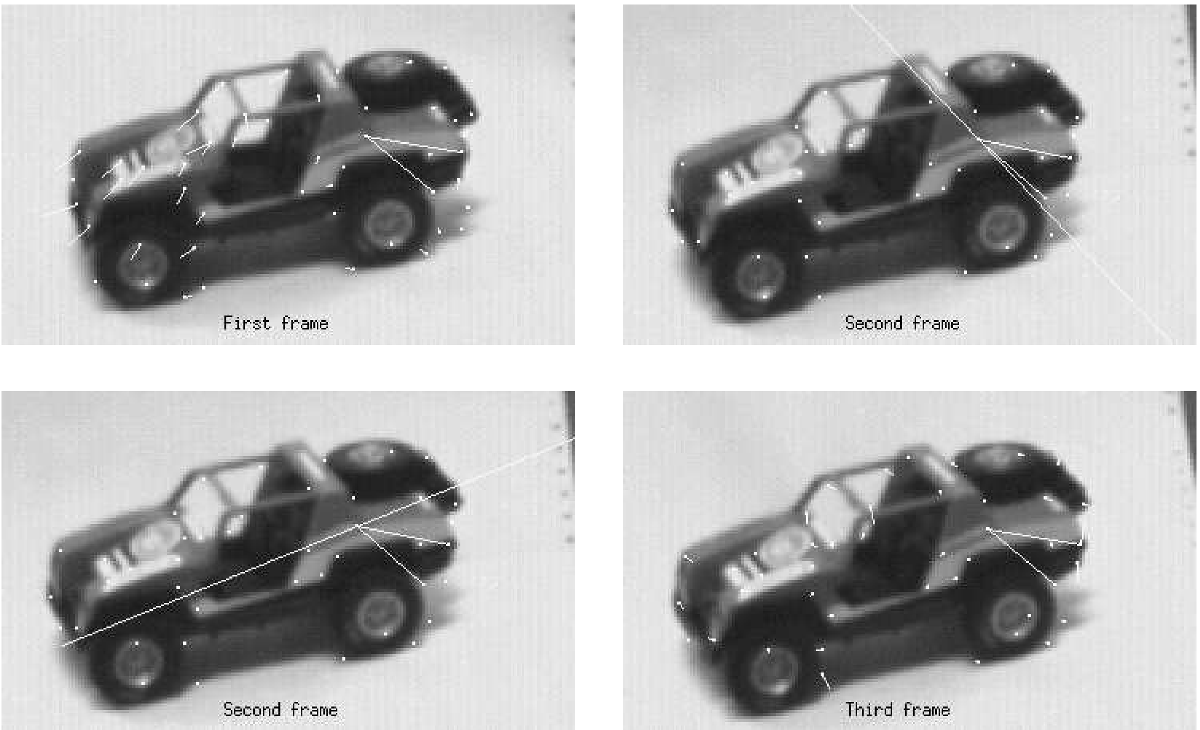

Figure 2: The affine flow obtained from another 2D basis. The estimated direction of affine flow vectors, and hence the axis of rotation, are grossly wrong due to the localization errors of the corners forming the 2D basis.

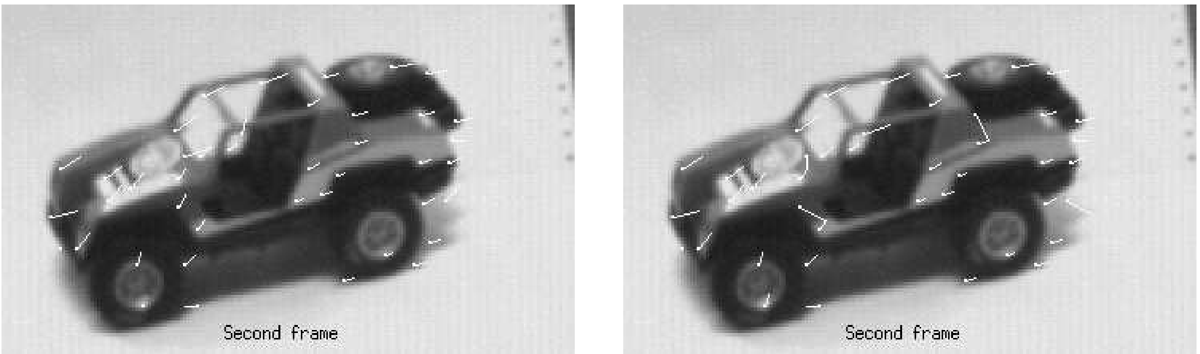

Figure 3: 2-frame correspondence results (Optical flow). Left: result obtained using our method. Right: result obtained using cross-correlation

Figure 4: The affine flow obtained from the correspondence results of the relaxation algorithm. The 2D basis used for computation of affine coordinates and the estimated axes of rotation are also shown.

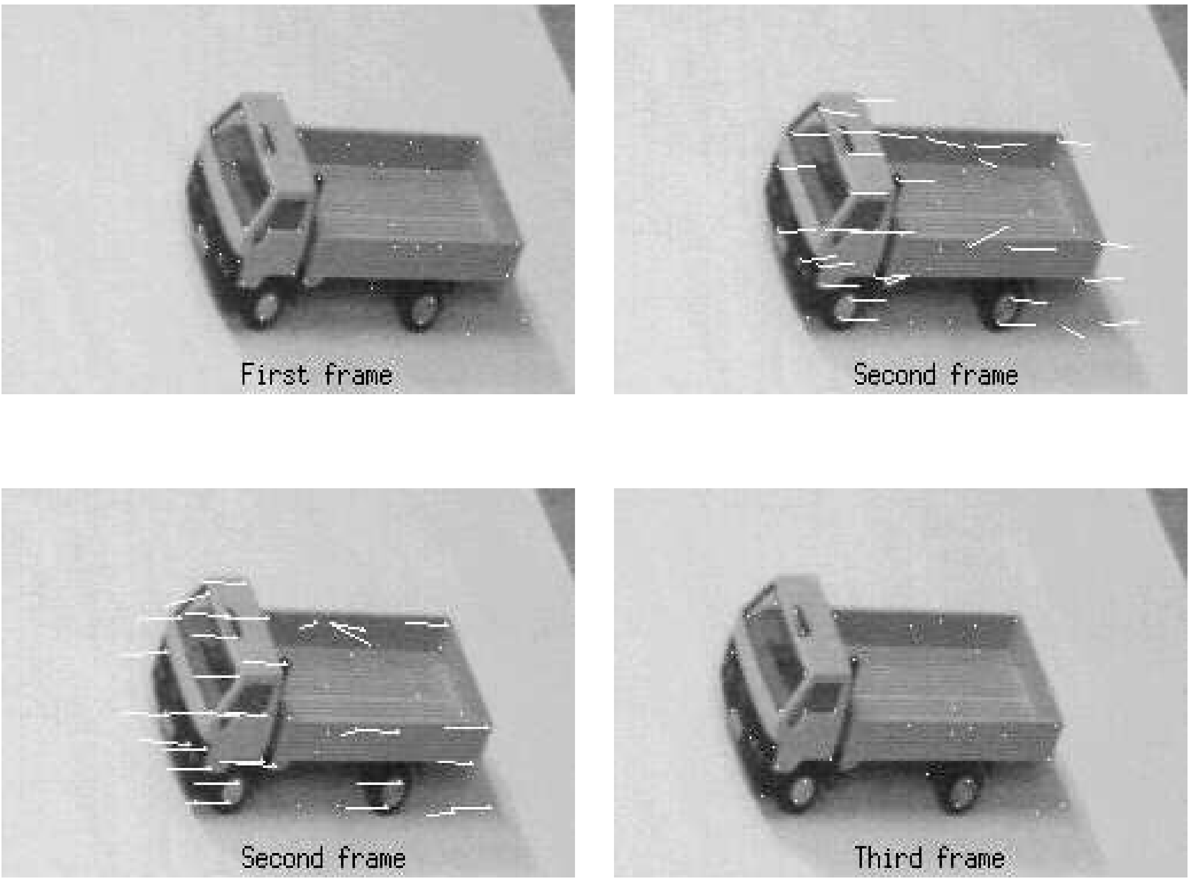

Figure 5: 3-frame correspondence results (Optical flow) for degenerate motion.

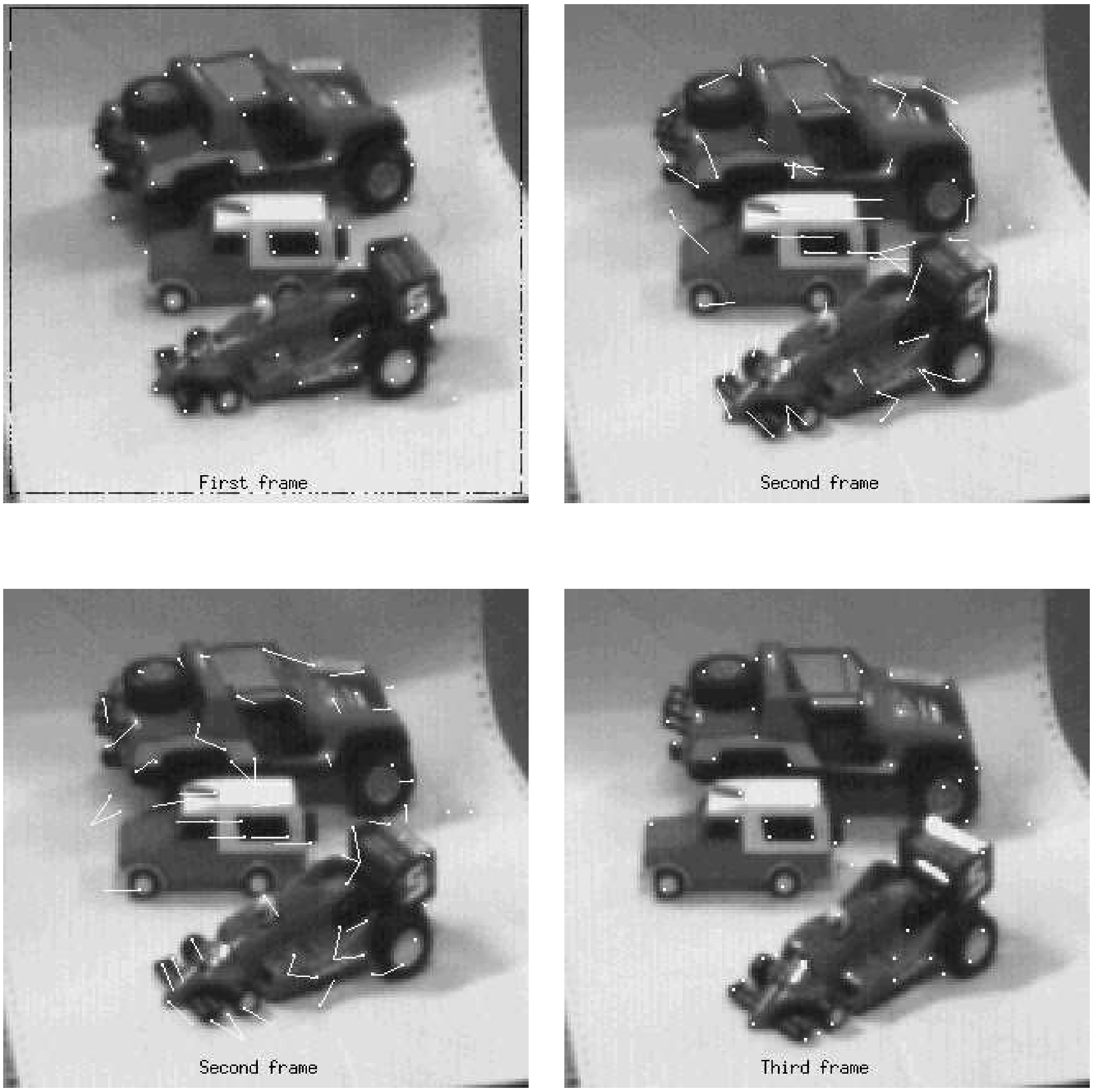

Figure 6: The Optical flow obtained from the correspondence results of the relaxation algorithm