Multi-frame Infinitesimal Motion Model for the Reconstruction of (Dynamic) Scenes with Multiple Linearly Moving Objects

Amnon Shashua and Anat Levin
School of Computer Science and Engineering,
The Hebrew University,
Jerusalem 91904, Israel
e-mail: {shashua,alevin}@cs.huji.ac.il

Abstract

We introduce new small-motion multi-frame equations applicable to the reconstruction of dynamic scenes in which points are allowed to move along straight-line paths with constant velocity. The motion equations apply to both static and dynamic points, thus prior segmentation is not necessary. We present a reconstruction algorithm of camera motion, scene structure, and point trajectories embedded into a multi-frame factorization principle which requires the minimum of 11 images and 7 points (out of which at least 3 are dynamic).

1 Introduction

In this paper we analyze the structure and motion from a video image sequence of scenes containing multiple moving points, each moving independently along some straight line path with constant velocity (for brevity, we will refer to such a scene as *dynamic*). Since the input consists of a continuous video source, we will focus on deriving a small-motion (infinitesimal) model which can treat the information arising from a dynamic scene in a uniform manner, i.e., without the need for a prior segmentation of the scene into fixed (static) and moving points, and moreover, to be able to handle scenes which consist solely of moving points (static points simply have vanishing velocity). We therefore focus on the following problem:

Given the optic-flow across an image sequence of a 3D configuration of points consisting of a mixture of static and dynamic points, including the case where all points are dynamic, recover the 3D translational and rotational components of camera motion, the 3D positions of the point configuration with respect to the first view and the 3D point trajectories (velocity vector) of the dynamic points.

The algorithm we present for doing so embeds the derived motion constraints into a factorization-based method. We show that the minimal number of views necessary for a factorization is 11 and the minimal number of points required for a recovery of camera motion, scene structure and point trajectories is 7. We also derive the solutions to reduced situations such as when the trajectories are embedded in a coplanar or collinear configuration, and when the point positions are coplanar.

1.1 Related Work

Infinitesimal motion constraints for a static 3D scene were first introduced in [6]. The first factorization-based algorithm for recovering (static) scene structure and discrete camera motion under the orthographic projection model was introduced in [8]. This was followed by a multi-body factorization method [3] for reconstructing the motion and shape of several bodies (each consisting of multiple points moving rigidly) simultaneously. Factorization-based algorithms were also derived for multi-frame problems for recovering homography matrices and for optic-flows arising from infinitesimal motion assumption of 3D scenes [5, 9].

Structure from Motion (SFM) of dynamic scenes, where each point moves independently along some trajectory, is a recent and growing topic. The topic was first introduced in [1] for 3D point configurations undergoing linear and curved motion with known projection matrices. For 2D point configurations across three views the motion constraints have a form of a $3 \times 3 \times 3$ tensor from which the appropriate homography matrices can be recovered [7]. The restriction to constant velocity (linear motion) and orthographic projection was introduced in [4] for 3D point configurations. The restriction carries a nice byproduct of embedding the orthographic motion constraints within the scene structure and point trajectories (velocities), thus giving rise to a factorization-based algorithm which requires a minimum of 5 views and 7 points.

2 Infinitesimal Motion Constraints for Dynamic 3D Scenes

Let $P_1, \cdots, P_n$ be a configuration of points, $P_i = (X_i, Y_i, z_i)$, where each point moves with constant velocity along some straight-line path $P_i + jV_i$ where $j = 0, \cdots, m$ representing a frame number. The frame number j may be replaced by a scalar α_j, $j = 0, \cdots, m$, (the α_j are fixed to all points of frame j) which represents a weaker assumption than constant velocity — nevertheless, for simplicity we will restrict ourself to constant velocity although the equations and methodology apply in the weaker case as well.

The configuration is viewed by a camera whose coordinate system at frame $j = 0$ (the reference frame) is aligned with the world coordinate system, i.e., $x_i = \frac{1}{z_i} X_i$ and $y_i = \frac{1}{z_i} Y_i$ are the image coordinates at time $j = 0$. At frame $j = 1$ the camera has undergone a small-motion displacement:

$$P_i' = (I + [w]_\times)(P_i + V_i) + t$$

where w represents the rotational component of camera motion (direction of w represents the screw-axis and $\mid w \mid$ represents the rotation (infinitesimal) angle around the axis), $[w]_x$ is the skew-symmetric matrix of vector products, thus $I + [w]_\times$ is the small-motion approximation of a rotation matrix, and t is the translational component of camera motion. In a small-motion model, $\dot{P}_i = \frac{d}{dt} P_i \approx P_i' - P_i$, thus

$$\dot{P}_i = [w]_\times P_i + (I + [w]_\times) V_i + t.$$

Let $p_i = (x_i, y_i, 1)^\top = \frac{1}{z_i} P_i$, then the optic-flow $u_i = \frac{dx_i}{dt}$ and $v_i = \frac{dy_i}{dt}$ takes the following form:

$$\begin{aligned} u_i &= \frac{d}{dt}\left(\frac{X_i}{z_i}\right) = \frac{1}{z_i}(\dot{X}_i - x_i \dot{z}_i) = \frac{1}{z_i} s_i^\top \dot{P}_i \\ &= s_i^\top [w]_x p_i + \frac{1}{z_i} s_i^\top t + \frac{1}{z_i} s_i^\top (I + [w]_x) V_i \qquad (1) \end{aligned}$$

Using similar derivation for $v_i = \frac{dy_i}{dt}$ we have:

$$v_i = r_i^\top [w]_x p_i + \frac{1}{z_i} r_i^\top t + \frac{1}{z_i} r_i^\top (I + [w]_x) V_i \qquad (2)$$

where $s_i = (1, 0, -x_i)$ and $r_i = (0, 1, -y_i)$. The optic-flow equations 1 and 2 are the motion constraints bringing together image measurements u_i, v_i, camera motion w, t, scene depth (structure) z_i, and the velocities V_i — all but the image measurements are unknown.

Note that there is a global translational ambiguity in determining t and V_i: $V_i' = V_i + q$ for some arbitrary q, is compensated by $t' = t - (I + [w]_\times) q$ while leaving the flow vectors (u_i, v_i) unchanged. Therefore, in the solution of structure (z_i), motion (t, w) and velocity (V_i) the translational component of the motion can be interchanged with a global shift q of the point velocities $V_1, ..., V_n$. A single known static point (V_i is known to be zero) can resolve the ambiguity, but that is for later.

3 Factorization

Because the optic flow equations are bilinear in the unknowns it is a simple matter to write down the estimation problem as a factorization algorithm, as follows. Consider $m + 1$ frames, $j = 0, \cdots, m$ and let w_j, t_j be the camera motion from the reference frame to frame j and let u_{ij}, v_{ij} be the optic flow of point i between the reference frame and frame j. Note that m should not be “too large” otherwise the infinitesimal assumption would breakdown in practical settings (see later about real-image experiments). Because

$$s^\top [w]_x p = s^\top (w \times p) = w^\top (s \times p)$$

we have the following relation:

$$\begin{pmatrix} u_{ij} \\ v_{ij} \end{pmatrix} = \begin{bmatrix} s_i \times p_i & \frac{1}{z_i} s_i^\top & \frac{1}{z_i} s_i^\top V_i & \frac{1}{z_i}(s_i \times V_i) \\ r_i \times p_i & \frac{1}{z_i} r^\top & \frac{1}{z_i} r_i^\top V_i & \frac{1}{z_i}(r_i \times V_i) \end{bmatrix} \begin{pmatrix} w_j \\ t_j \\ j \\ jw_j \end{pmatrix}$$

Grouping all the image measurements together we obtain the following matrix equation:

$$W = \begin{bmatrix} U \\ - \\ V \end{bmatrix} = \begin{bmatrix} S_x \\ - \\ S_y \end{bmatrix} M = SM, \qquad (3)$$

where $U = (u_{ij})$ is the $n \times m$ matrix whose entries are u_{ij} and $V = (v_{ij})$. The matrix M is $10 \times m$ where the j’th column is the vector $(w, t, j, jw)^\top$, S_x, S_y are $n \times 10$ matrices, where the i’th row of S_x consists of:

$$[s_i \times p_i,\ \frac{1}{z_i} s_i^\top,\ \frac{1}{z_i} s_i^\top V_i,\ \frac{1}{z_i}(s_i \times V_i)],$$

and the i’th row of S_y consists of:

$$[r_i \times p_i,\ \frac{1}{z_i} r_i^\top,\ \frac{1}{z_i} r_i^\top V_i,\ \frac{1}{z_i}(r_i \times V_i)].$$

Thus the rank of the measurement matrix W is bounded from above by 10, therefore using SVD we can find two matrices $K_{2n \times 10}$,$L_{10 \times m}$ such that $W = KL$. The shape matrix S and the motion matrix M can be determined by $S = KA$ and $M = A^{-1} L$ for some 10×10 matrix A. The unknown matrix A must satisfy “structure” constraints determined by the way the matrices S and M are built.

3.1 Solving for A

We Notice that when trying to recover A, to satisfy $S = KA$, we obtain 3 separate linear systems, regarding 3 different groups of columns:

$$\begin{aligned} S_{1-3} &= KA_{1-3} \\ S_{4-6} &= KA_{4-6} \\ S_{7-10} &= KA_{7-10} \end{aligned}$$

where S_{i-j} denotes the sub-matrix of S consisting of columns i to j (inclusive). The first 3 columns of S are known, so each tracked point gives 6 equations, and we have 30 unknowns for columns $1-3$ of A, therefore we need at least 5 points for a unique solution.

In columns 4-6 each point contributes 5 equations (after eliminating the one unknown $\frac{1}{z_i}$), therefore we need at least 6 points. Those columns can be recovered only up to a global scale. The 5 constraints per point are as follows:

$$\begin{aligned} K_xA_5 &= 0 && (4) \\ K_yA_4 &= 0 && (5) \\ K_xA_4 &= K_yA_5 && (6) \\ (K_xA_6)_i &= x_i(K_xA_4)_i && (7) \\ (K_yA_6)_i &= y_i(K_yA_5)_i && (8) \end{aligned}$$

where K_x, K_y are the upper and lower $n \times m$ sub-matrices of K and A_j denotes the j'th columns of A.

So far, the equation counting was straightforward. However, the determination of 40 unknowns A_{7-10} is more subtle. Here we expect the global translational ambiguity to have an effect, for example. For columns 7-10, each point contributes 5 equations (after eliminating the 3 unknown $\frac{1}{z_i}V_i$ from the 8 measurements) and there are 40 unknowns. The 5 constraints are:

$$\begin{aligned} x_iK_xA_4 &= K_xA_2 && (9) \\ y_iK_yA_4 &= K_yA_3 && (10) \\ K_y(y_iA_2 + A_1) &= (y_i^2+1)K_xA_4 && (11) \\ K_x(x_iA_3 - A_1) &= (x_i^2+1)K_xA_4 && (12) \\ (x_iK_y + y_iK_x)A_4 &= K_xA_3 + K_yA_2 && (13) \end{aligned}$$

The question is what is the rank of the estimation matrix for the 40 unknowns A_{7-10}? The rank is 35 as shown next.

Claim 1 *The rank of the estimation matrix for the 40 unknowns of A_{7-10} is bounded by 35.*

Proof: Is it sufficient to consider $U = K_xAA^{-1}M$ where we wish to find A that satisfies $S_x = K_xA$. Ambiguity arises if we can replace V_i with V_i' such that $S_x(V_i') = S_xB$ for some matrix B. The number of free variables in B will determine the number of degrees of freedom in determining A. Let

$$V_i' = aV_i + q + bz_ip_i$$

where the vector q and the scalars a, b are free variables. By substitution we find that entries $7-10$ the i'th row of $S_x(V_i')$ has the following form:

$$a\frac{1}{z_i}s_i^\top V_i + \frac{1}{z_i}s_i^\top q,\ a\frac{1}{z_i}(s_i \times V_i) + \frac{1}{z_i}(s_i \times q) + b(s_i \times p_i)$$

Recall that $s_i^\top p_i = 0$ (which is why this term dropped out in the 7'th entry). We have that B_{7-10} has the following form:

$$B_{7-10} = \begin{bmatrix} 0 & & & \\ 0 & & bI & \\ 0 & & & \\ q & & [q]_\times & \\ a & 0 & 0 & 0 \\ 0 & & & \\ 0 & & aI & \\ 0 & & & \end{bmatrix}$$

Since B_{7-10} contains 5 free variables, the rank of the estimation matrix for the 40 unknowns in A_{7-10} cannot exceed 35. □

An immediate conclusion is that the minimal number of points for the recovery of A is 7. The free parameters a, b can be resolved by noticing that the j'th column of $B^{-1}M$ has the following form:

$$B^{-1}M_j = \begin{pmatrix} w_j - \frac{a}{b}w_j \\ t_j - \frac{1}{a}(q + [q]_\times w_j) \\ \frac{1}{a}j \\ \frac{1}{a}jw_j \end{pmatrix}$$

Consider for example the column $j = 1$. Denote by $\alpha_1, \cdots, \alpha_{10}$ the entries of the column. Let $u = (\alpha_1, \alpha_2, \alpha_3)^\top$ and $v = (\alpha_8, \alpha_9, \alpha_{10})^\top$. Then, $a = \frac{1}{\alpha_7}$ and $au - v = bu$ from which we can recover b.

To summarize, we perform the following steps:

1. Given the optic-flow matrix W (having at least 11 views), perform SVD to obtain $W = KL$. In this process one can reduce measurement error by enforcing the singular values from the 11'th position and upwards to vanish.

2. Recover A up some arbitrary element of the 5-dimensional null space (in recovering A_{7-10}. Let $S' = KA$ and $M' = A^{-1}L$.

3. Recover a, b from M'. Construct the matrix B with the entries $1/a, 1/b$ in the proper places. Then, $S = S'B$ and $M = B^{-1}M'$. The recovered S, M are the structure and motion and velocities up to the global translation/velocity ambiguity.

We have therefore shown that we can recover the complete SFM and point trajectories uniquely (up to the intrinsic ambiguity of global shift of velocities) with a factorization-based algorithm which requires the minimum of 11 views and 7 points.

3.2 Static Scene

A particular case of the above is when the scene is static, i.e., V_i vanishes for all points. This brings us back to the rank 6 observation of [5] of the optic-flow matrix W, whereas here we can extend this further into a SFM algorithm. Specifically, we have in this case

$$\left(\begin{array}{c} u_{ij} \\ v_{ij} \end{array} \right) = \left[\begin{array}{cc} s_i \times p_i & \frac{1}{z_i} s_i^\top \\ r_i \times p_i & \frac{1}{z_i} r^\top \end{array} \right] \left(\begin{array}{c} w_j \\ t_j \end{array} \right)$$

which when stacked together we obtain the form of eqn. 3 where S_x, S_y are $n \times 6$ matrices and M is a $6 \times m$ matrix. Thus the rank of the $2n \times m$ measurement matrix W is bounded by 6. Let $W = KL$ provided by an SVD decomposition of W, then we seek a 6×6 matrix A such that $S = KA$ and $M = A^{-1}L$. The constraints on A follow the constraints on A_{1-3} and A_{4-6} discussed in the previous section. We need therefore at least 6 points and 7 frames in order to use a factorization principle to recover structure and motion from an image sequence of a static scene.

4 Reduced Configurations

The situation we described so far was *general* in the sense that the point positions $P_1, ..., P_n$ *and* the velocities $V_1, ..., V_n$ live in a 3D space. Some practical situations arise, for example, when all the velocities are coplanar ($\dim \mathrm{Span}\{V_i\} = 2$) or along parallel lines ($\dim \mathrm{Span}\{V_i\} = 1$), or when the point configuration $P_1, ..., P_n$ is coplanar at the starting stage ($P_i + jV_i$ may not be coplanar), or any combination of the above.

We call these situations collectively as *reduced configurations* unlike degenerate, because in fact as we will show there are no degeneracies — we can achieve the full reconstruction (up to the global velocity shift ambiguity) as in the general case — but there are additional subtleties that require special handling.

4.1 Coplanar Trajectories

Assume $\dim \mathrm{Span}\{V_i\} = 2$, i.e., there exists $n = [n_1, n_2, n_3]^\top$ and l such that every velocity V_i in the scene satisfies $n^\top V_i = l$. This constraint affects the number of degrees of freedom of A_{7-10} which instead of having 5 degrees of freedom will have now 6. We have the following claim:

Claim 2 *When* $\dim \mathrm{Span}\{V_i\} = 2$*, the rank of the estimation matrix for the 40 unknowns of* A_{7-10} *is bounded by 34.*

Proof: As in Claim 1, ambiguity in the estimation of A from the equations $S_x = K_x A$ arises if we can replace V_i with V_i' such that $S_x(V_i') = S_x B$ for some matrix B. The number of free variables in B will determine the number of degrees of freedom in determining A. Let

$$V_i' = aV_i + q + bz_ip_i + f(V_i \times n)$$

where the vector q and the scalars a, b, f are free variables. Recall the following identities:

$$\begin{array}{rcl} a \times (b \times c) & = & (a^\top c)b - (a^\top b)c \\ (a \times b) \times c & = & (a^\top c)b - (c^\top b)a, \end{array}$$

which are used for establishing the following identity:

$$\begin{array}{rcl} s_i & \times & (V_i \times n) = (s_i^\top n)V_i - (s_i^\top V_i)n \\ & = & (s_i^\top n)V_i - (s_i^\top V_i)n - (n^\top V_i)s_i + (n^\top V_i)s_i \\ & = & (s_i^\top n)V_i - (s_i^\top V_i)n - (n^\top V_i)s_i + ls_i \\ & = & (s_i \times V_i) \times n - (s_i^\top n)V_i + ls_i \end{array} \quad (14)$$

By substitution we find the entry 7 of the i'th row of $S_x(V')$ becomes:

$$a\frac{1}{z_i}s_i^\top V_i + \frac{1}{z_i}s_i^\top q + f\frac{1}{z_i}s_i^\top(V_i \times n)$$

and entries $8 - 10$ become:

$$\begin{array}{l} a\frac{1}{z_i}(s_i \times V_i) + \frac{1}{z_i}(s_i \times q) + b(s_i \times p_i) \\ +f\frac{1}{z_i}((s_i \times V_i) \times n - (s_i^\top V_i)n + ls_i) \end{array}$$

We have that B_{7-10} has the following form:

$$B_{7-10} = \left[\begin{array}{cc} 0 & \\ 0 & bI \\ 0 & \\ q & [q]_\times + flI \\ a & -fn^\top \\ fn & aI + f[n]_\times \end{array} \right]$$

Since B_{7-10} contains 6 free variables (vector q and scalars a, b, f), the rank of the estimation matrix for the 40 unknowns in A_{7-10} cannot exceed 34. $\Box$

In order to solve for the free parameters we do the following. Since the entire model can be recovered up to a global translation of the velocities V_i we can translate the coordinate system such that one arbitrarily chosen point, say p_1, is static i.e. $v_1 = 0$, and the trajectory plane passes through the origin, i.e., $l = 0$. To see why this is so, note that by setting $a = 1$, $b = f = 0$, and $q = -V_1$ we obtain a matrix B_o such that SB_o has all the velocities v_i replaced by $v_i - V_1$ and hence the entries $7-10$ of two of its rows vanish. It is therefore possible to add those 8 constraints on the vanishing entries of the system KA for solving for A_{7-10}. The rank of the estimation matrix for A_{7-10} becomes 37 (instead of 34) as described below.

Claim 3 *By setting an arbitrary point, say p_1, as static, the additional 8 constraints that arise from it shift the coordinate system such that the trajectory plane passes through the origin ($l = 0$) and the rank of the estimation matrix for the 40 unknowns of A_{7-10} is bounded by 37.*

Proof: Ambiguity in the estimation of A arises if we can replace V_i with V_i', such that $S(V_i') = SB$. The number of free variables in B will determine the number of degrees of freedom in determining A. The 8 additional constraints enforce $V_1' = 0$, therefore:

$$0 = V_1' = aV_1 + q + bz_1p_1 + f(V_1 \times n),$$

where $V_1 = 0$ as well. Thus, q is completely determined by $q = -bz_1p_1$, and because

$$0 = s_1 \times (V_1 \times n) = (s_i \times V_1) \times n - (s_1^\top n)V_1 + ls_1$$

we have also that $l = 0$. Therefore, only 3 degrees of freedom remain in B which are a, b, f. □

We can recover the scalars a, b and the vector fn as follows. Recall that via SVD we have the factorization $W = KL$ and in turn $U = K_xL$ and $V = K_yL$. We recover A where A_{7-10} is determined by choosing an arbitrary solution from the 3-dimensional null space. Let $S' = KA$ and $M' = A^{-1}L$ where $S' = SB$ where B_{7-10} (where we have free variables) as above. Thus, $W = S'M' = SBB^{-1}M$ and $M = BM'$. Consider the j'th column of M' and let the entries in the resulting vector be denoted as (m_1, m_2, λ, m_3) where m_i are vectors, 3 components each, and λ is a scalar. If B is chosen correctly then the j'th column of BM' should have the form (w_j, t_j, j, jw_j). We have:

$$BM_j' = \begin{pmatrix} m_1 + bm_3 \\ m_2 + \lambda q + [q]_x m_3 \\ a\lambda - fn^\top m_3 \\ fn \cdot m_3 + am_3 + f[n]_\times m_3 \end{pmatrix},$$

where $x \cdot y$ denotes the vector (x_1y_1, x_2y_2, x_3y_3). This provides the following linear constraints on a, b, fn:

$$\begin{aligned} j(m_1 + bm_3) &= fn \cdot m_3 + am_3 + f[n]_\times m_3 \\ a\lambda - fn^\top m_3 &= j \end{aligned}$$

Thus, using several columns of M' we can solve for a, b, fn, and given b, we can recover q (recall $q = -bz_1p_1$). Note that since $l = 0$, then a, b, fn, q completely determine B. In Summary, we have shown that we can recover S, M up to the global shift by $q = -V_1$ when the velocities span a 2D space.

4.2 Coplanar Points at Starting Moment

Assume $\dim \mathrm{Span}\{P_i\} = 2$, i.e., there exists $n = [n_1, n_2, n_3]^\top$ such that every point P_i in the scene satisfies $n^\top P_i = 1$, or $n^\top p_i = \frac{1}{z_i}$. Just like in the coplanar trajectories case, we have 6 degrees of freedom in recovering A_{7-10}:

Claim 4 *When $\dim \mathrm{Span}\{P_i\} = 2$, the rank of the estimation matrix for the 40 unknowns of A_{7-10} is bounded by 34.*

Proof: Following the proof of Claim 2, let

$$V_i' = aV_i + q + bz_ip_i + fz_i(p_i \times n)$$

where the vector q and the scalars a, b, f are free variables. We can establish the following identity:

$$s_i \times (p_i \times n) = (s_i \times p_i) \times n + s_i$$

by following the derivation of eqn. 14. By substitution we find the entry 7 of the i'th row of $S_x(V')$ becomes:

$$a\frac{1}{z_i}s_i^\top V_i + \frac{1}{z_i}s_i^\top q + fn^\top(s_i \times p_i)$$

and entries $8 - 10$ become:

$$\begin{aligned} &a\frac{1}{z_i}(s_i \times V_i) + \frac{1}{z_i}(s_i \times q) + b(s_i \times p_i) \\ &+f(s_i \times p_i) \times n + f\frac{1}{z_i}s_i \end{aligned}$$

We have that B_{7-10} has the following form:

$$B_{7-10} = \left[\begin{array}{cccc} fn & & bI + f[n]_\times & \\ q & & [q]_\times + fI & \\ a & 0 & 0 & 0 \\ 0 & & & \\ 0 & & aI & \\ 0 & & & \end{array}\right]$$

Since B_{7-10} contains 6 free variables (vector q and scalars a, b, f), the rank of the estimation matrix for the 40 unknowns in A_{7-10} cannot exceed 34. □

In order to determine the free variables a, b and vector fn we do the following. Let $W = KAA^{-1}L = S'M'$ where K, L determined by SVD of W and A is determined by choosing an arbitrary solution from the 6-dimensional null space. Denote the j'th column of M' as (m_1, m_2, λ, m_3), then BM'_j has the form:

$$BM'_j = \begin{pmatrix} m_1 + \lambda f n + b m_3 + f[n]_\times m_3 \\ m_2 + \lambda q + [q]_x m_3 + f m_3 \\ a\lambda \\ a m_3 \end{pmatrix},$$

This provides the following linear constraints on a, b, fn:

$$\begin{aligned} a\lambda &= j \\ \frac{1}{j} a m_3 &= m_1 + \lambda f n + b m_3 + f[n]_\times m_3 \end{aligned}$$

from which we can solve for a, b, fn (via several columns of M'). In order to separate f and n we can either shift the velocities by $q = V_1$ as was done in the previous section, or alternatively, recover z_i (up to a global scale) from A because the depth values are not affected by the ambiguity in A_{7-10}. Once z_i are recovered one can solve for n, and then solve for a, b, f from the equations above. The recovery of z_i is also useful for purposes of *distinguishing* between the rank 34 caused by coplanar trajectories and the rank 34 caused by coplanar points. Once a, b, f, n are recovered, one can substitute their values into B above (up to a global shift q). In Summary, we have shown that we can recover S, M up to the global shift q when the points P_i span a 2D space.

4.3 Parallel Trajectories

Another case of interest, as it may often occur in practical situations, is the case where $\dim \operatorname{Span}\{V_i\} = 1$, i.e., all the straight-line trajectories are parallel to each other. As in Section 4.1, we can assume that some arbitrary point, say p_1, is static, i.e., $V_1 = 0$. The implies that the all the (parallel) line trajectories pass through the origin, i.e., $V_i = \gamma_i V$ (and $\gamma_1 = 0$) for some fixed vector V. The motion constraints for this special case are:

$$\begin{aligned} u_i &= s_i^\top [w]_x p_i + \frac{1}{z_i} s_i^\top t + \frac{\gamma_i}{z_i} s_i^\top (I + [w]_x) V \\ v_i &= r_i^\top [w]_x p_i + \frac{1}{z_i} r_i^\top t + \frac{\gamma_i}{z_i} r_i^\top (I + [w]_x) V \end{aligned}$$

And the matrices S, M have the following form:

$$M = \begin{bmatrix} w_1 & \cdots & w_m \\ t_1 & \cdots & t_m \\ (I - [w_1]_x)V & \cdots & (I - [w_m]_x)mV \end{bmatrix}_{9 \times m}$$

And:

$$S = \begin{bmatrix} s_1 \times p_1 & \frac{1}{z_1} s_1 & \frac{\gamma_1}{z_1} s_1 \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ s_n \times p_n & \frac{1}{z_n} s_n & \frac{\gamma_n}{z_n} s_n \\ r_1 \times p_1 & \frac{1}{z_1} r_1 & \frac{\gamma_1}{z_1} r_1 \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ r_n \times p_n & \frac{1}{z_n} r_n & \frac{\gamma_n}{z_n} r_n \end{bmatrix}_{2n \times 9}$$

Note that the ranks of S and M are bounded by 9, so also $rank(W) \leq 9$. Since depth z_i and the velocities $\gamma_i V$ are defined only up to scale, we have two degrees of freedom embodied into the matrix B_0 of the form:

$$B_0 = \begin{bmatrix} I_{3\times 3} & 0 & 0 \\ 0 & \alpha I_{3\times 3} & 0 \\ 0 & 0 & \beta I_{3\times 3} \end{bmatrix}_{9 \times 9}$$

such that $S' = SB_0$ and $M' = B_0^{-1} M$ are indistinguishable from the original pair S, M. Using SVD we find $K_{2n\times 9}$, and $L_{9\times m}$, such that there is a matrix $A_{9\times 9}$ that satisfies: $S = KA$, $M = A^{-1}L$. We write estimation matrices for A_{4-6} and A_{7-9}. Each point contributes 5 linear constraints on A_{4-6} and A_{7-9}, those constraints have the same form as in equations 4 - 8. Since we assume that the first point is static, the last 3 columns in the first rows of S_x, S_y must be zero, thus providing 6 additional constraints to the estimation matrix for A_{7-9}.

Claim 5 *Using only constraints on the shape matrix, A can be recovered only up to 3 degrees of freedom*

Proof: Ambiguity in the solution for A_{4-6} and A_{7-9} arises if we can replace z_i with z'_i and γ_i with γ'_i such that $S_x(z'_i, \gamma'_i) = S_x B$ for some matrix B. The number of free variables in B will determine the number of degrees of freedom in determining A. One can easily verify that B has the following form:

$$B = \begin{bmatrix} I_{3\times 3} & 0 & 0 \\ 0 & a I_{3\times 3} & 0 \\ 0 & b I_{3\times 3} & c I_{3\times 3} \end{bmatrix}_{9 \times 9}$$

containing three free variables — two for the columns A_{4-6} and one for A_{7-9}). Note that because $\gamma_1 = 0$ we were allowed only to scale columns A_{7-9}. $\square$

One can solve for the scalars a, b and the vector V up to a global scale as follows. Choose a solution A from the 3-parameter solution space (A_{4-6} from a 2-parameter space, and A_{7-9} up to global scale). Set $S' = KA$ and $M' = A^{-1}L$ and we are searching for B, whose structure is described above, such that $S' = SB$, $M' = B^{-1}M$. Denote M'_j, the j'th column of M', by $(m_1, m_2, m_3)^\top$. Then,

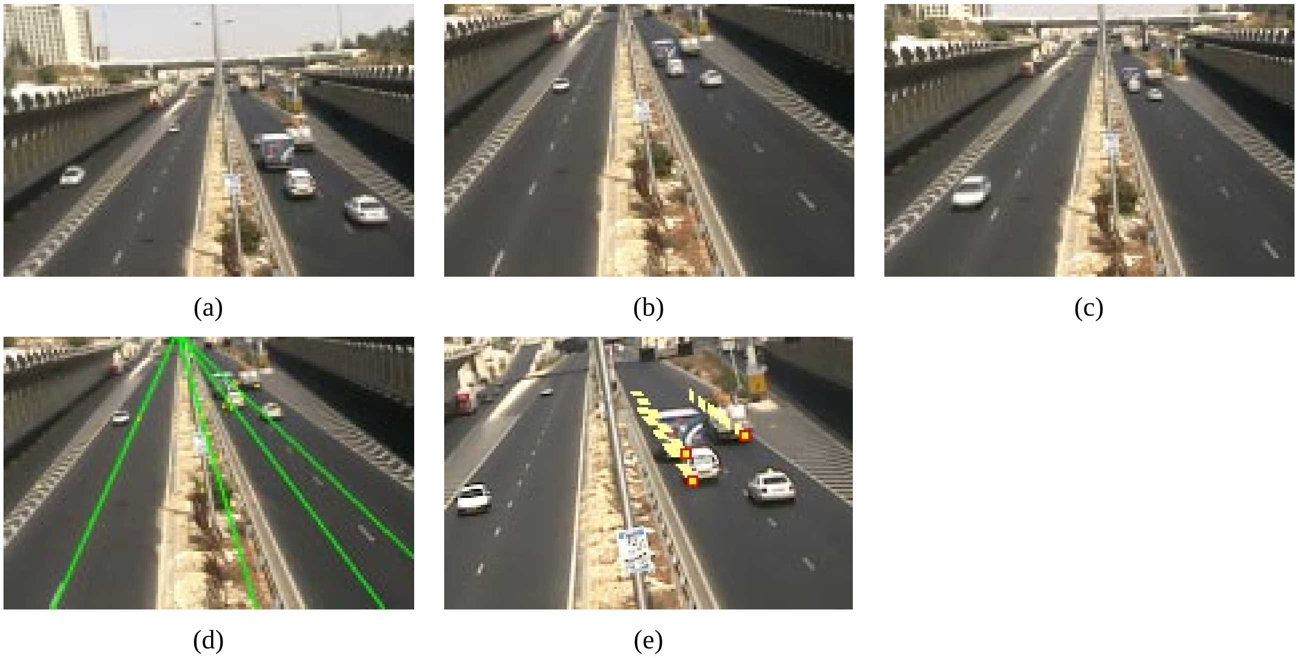

Figure 1. (a),(b),(c) 1st, 40th and 80th images of the traffic scene. (d) The tracked points and the road direction projected over the reference frame. (e) Predicted locations of 3 of the tracked points.

$BM'_j = (m_1, am_2, bm_2 + cm_3)^\top$ which should have the form of $M_j = (w_j, t_j, jV - jw \times V)^\top$. We have therefore the constraint equation per j'th column of M':

$$jV - jm_1 \times V = bm_2 + cm_3$$

which provides 3 homogeneous equations per column for the unknowns b, c and vector V. The scalar a is set arbitrarily, thus we have shown that we can recover S, M up to the intrinsic ambiguity (matrix B_0) of global scale of z_i and global scale of $\gamma_i V$. Note that the method does not depend on whether the point configuration P_i lie on a 2D or 3D space.

5 Experiments

We have conducted a number of experiments of which two will be described below. The first experiment involves the *parallel trajectories* configuration (Section 4.3) taken by a moving camera viewing a typical highway traffic scene with moving vehicles. The second experiment involves the *coplanar trajectories* (Section 4.1) configuration on a semi-synthetic setup. It is semi-synthetic in the sense that the scene is constructed by computer graphics rendering whereas the image measurements are taken with the same image-processing tools of point tracking as with the real-image sequence.

Fig. 1(a-c) displays three images of a sequence of 80 frames of a highway traffic situation. The sequence was taken by a hand-held video camera from a bridge over the road. The constant velocity assumption in such a scene is fairly reasonable and because the vehicles were moving on a straight section of the road, the parallel trajectory model is also reasonable.

Using the KLT point tracking package [2] we tracked 28 points on static objects and on different moving cars, moving at different speeds and along both lane directions. The vector V representing the direction of vehicle travel was computed as described in Section 4.3 and projected onto the reference image (first image of the sequence) displayed in Fig. 1d. The projected line appears accurately aligned with the road direction including the pitch angle (lines meet at the true horizon). We then recovered the velocity magnitude γ_i of each point and made use of it for predicting the position of the vehicle in subsequent frames (assuming a constant velocity motion).

We also recovered the private velocity (which is a scale of the common velocity) for each of the tracked points. This enables us to predict the location of the point in another x frames (i.e. after x time intervals). In figure 1(e) we projected the expected movement of 3 of the points using the frequency of 6 time intervals. We can see that the truck on the right moved a little slower than the white van on its

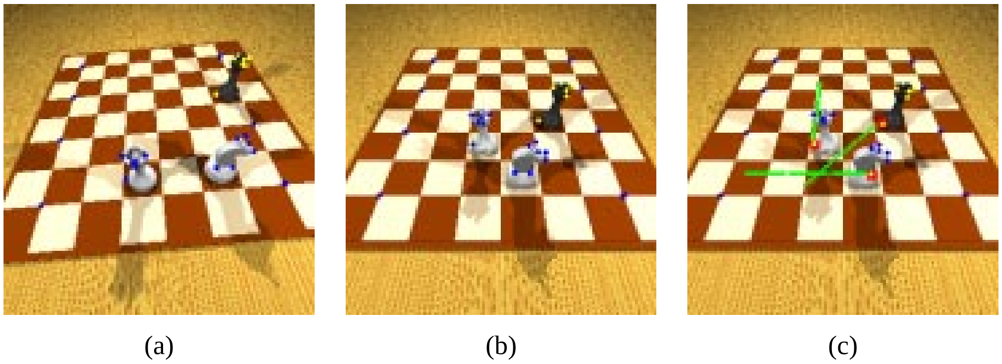

Figure 2. (a),(b) 2 images from the chess sequence. (d) 3 of the tracked points and their velocities projected onto the reference frame.

left. Fig. 1e displays the predicted position of three points — the predicted positions remain on the vehicle thus indicating good accuracy in the recovery of the model (motion, structure and velocities).

Fig. 2(a-c) displays three images, out of a sequence of 60 frames, of a coplanar trajectories configuration of chess pieces in motion (points span a 3D space while velocities span a 2D space). The recovered velocities V_i were recovered using the method in Section 4.1 and projected on the reference image (Fig. 2d). The projected lines appear to trace accurately the straight line motion of the chess pieces.

In both experiments fairly long sequences were taken covering image displacements of dozens of pixels in some cases. The effect of discrete motion (instead of infinitesimal) on the final results was minimal, thus suggesting robustness of the approach.

6 Summary

We have introduced the small-motion equations for handling multiple linearly moving points under constant velocity and a factorization-based algorithm for extracting the parameters of scene structure, camera motion and point trajectories from the image-flow measurements. Our method covered a majority of situations of interest starting from the general 3D point configuration and 3D line trajectories, to combinations of 2D point configuration and 2D trajectories up to 1D trajectories. In all those cases we have shown that the information can be recovered (in a robust manner as seen from the experiments) uniquely (up to an intrinsic shift ambiguity).

References

[1] S. Avidan and A. Shashua. Trajectory Triangulation: 3D Reconstruction of Moving Points from a Monocular Image Sequence. *IEEE Transaction on Pattern Analysis and Machine Intelligence*(PAMI), Vol. 22(4),pp.348–357,2000.

[2] S. Birchfield, Author. An implementation of the Kanade-Lucas-Tomasi feature tracker Version 1.1.5 http://www-leland.stanford.edu/group/OTL, (650) 723-0651.

[3] J. Costeira and T. Kanade. A Multibody Factorization Method for Independent Moving Objects. *1998 International Journal on Computer Vision*, Kluwer, Vol. 29, No. 3, September, 1998.

[4] M. Han and T. Kanade. Reconstruction of a Scene with Multiple Linearly Moving Objects.In *Proc. of Computer Vision and Pattern Recognition*, June, 2000.

[5] M. Irani. Multi-Frame Optical Flow Estimation Using Subspace Constraints. *IEEE International Conference on Computer Vision (ICCV)*, Corfu, September 1999.

[6] H.C. Longuett-Higgins and K. Prazdny. The interpretation of a moving retinal image. *Proceedings of the Royal Society of London B*, 208:385-397,(1980).

[7] A. Shashua and L. Wolf. Homography Tensors: On Algebraic Entities That Represent Three Views of Static or Moving Planar Points. In *Proc. of the European Conference on Computer Vision (ECCV)*, June 2000, Dublin, Ireland.

[8] C. Tomasi and T. Kanade Shape and Motion from Image Streams under Orthography: a Factorization Method *IJCV*, 9(2):137-154,1992.

[9] . L. Zelnik-Manor and M. Irani. Multi-Frame Alignment of Planes. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 1999.