Constraints on Coplanar Moving Points

Sujit Kuthirummal⋆, C.V. Jawahar, and P.J. Narayanan

Centre for Visual Information Technology
International Institute of Information Technology
Hyderabad 50019 India
pjn@iiit.net

Abstract. Configurations of dynamic points viewed by one or more cameras have not been studied much. In this paper, we present several view and time-independent constraints on different configurations of points moving on a plane. We show that 4 points with constant independent velocities or accelerations under affine projection can be characterized in a view independent manner using 2 views. Under perspective projection, 5 coplanar points under uniform linear velocity observed for 3 time instants in a single view have a view-independent characterization. The best known constraint for this case involves 6 points observed for 35 frames. Under uniform acceleration, 5 points in 5 time instants have a view-independent characterization. We also present constraints on a point undergoing arbitrary planar motion under affine projections in the Fourier domain. The constraints introduced in this paper involve fewer points or views than similar results reported in the literature and are simpler to compute in most cases. The constraints developed can be applied to many aspects of computer vision. Recognition constraints for several planar point configurations of moving points can result from them. We also show how time-alignment of views captured independently can follow from the constraints on moving point configurations.

1 Introduction

The study of view-independent constraints on the projections of a configuration of points is important for recognition of such point configurations. A number of view-independent invariants have been identified for static point configurations [1,2]. They encapsulate information about the scene independent of the cameras being used and are opposite in philosophy to the scene-independent constraints like the Fundamental Matrix [1], the multilinear tensors [3,4,5], etc. Formulating view independent constraints on the projections of dynamic point configurations is more challenging and has been studied less. Many configurations of dynamic points are possible. Points could be in general positions or could lie on a plane or on a line. The motion could be arbitrary or constrained. An interesting case is linearly moving points with independent uniform velocities or accelerations. In this paper, we derive several simple constraints on the projections of moving points and their motion parameters.

⋆ Currently with the Department of Computer Science, Columbia University

As an example of view-independent constraints on point configurations, let us consider a set of 5 world points $\mathbf{P}[i] \in R^3$, $i = 1..5$ and their images $\mathbf{p}[i]$ in homogeneous and $(x[i], y[i])$ in Cartesian coordinates viewed by an affine camera $\mathbf{M}$. Let $\mathbf{m}_i$ be the vector of the first 3 elements in the ith row of $\mathbf{M}$ and let m_{i4} be the fourth element in the ith row. Therefore, $x[i] = \mathbf{m}_1.\mathbf{P}[i] + m_{14}$ and $y[i] = \mathbf{m}_2.\mathbf{P}[i] + m_{24}$. Alternatively, $\begin{bmatrix} P[i] & 1 & x[i] \end{bmatrix}^{\mathrm{T}} \begin{bmatrix} \mathbf{m}_1 & m_{14} & -1 \end{bmatrix} = \begin{bmatrix} P[i] & 1 & y[i] \end{bmatrix}^{\mathrm{T}} \begin{bmatrix} \mathbf{m}_2 & m_{24} & -1 \end{bmatrix} = 0$. If we have at least five points, then we can form a set of equations of the form $\mathbf{C}_1\theta_1 = \mathbf{C}_2\theta_2 = 0$, where $\theta_1 = \begin{bmatrix} \mathbf{m}_1 & m_{14} \end{bmatrix}^{\mathrm{T}}$, $\theta_2 = \begin{bmatrix} \mathbf{m}_2 & m_{24} \end{bmatrix}^{\mathrm{T}}$ and each row of the measurement matrix $\mathbf{C}_1$ (or $\mathbf{C}_2$) consists of the unknown world point $\mathbf{P}[i]$, unity and the $x[i]$ (or $y[i]$) coordinate. Note that the camera parameters are factored out into vectors θ_1 and θ_1. It is obvious that $\mathbf{C}_1$ and $\mathbf{C}_2$ are rank deficient and expanding their 5×5 determinant results in constraints of the form $\sum_{i=1}^{5} \alpha_i x[i] = 0$ and $\sum_{i=1}^{5} \alpha_i y[i] = 0$ where α_i are functions of the world position of the points $\mathbf{P}_i$ and hence is the same for all views, i.e the α_i are view-independent coefficients. Note that the coefficients of $x[i]$ and $y[i]$ in the above constraint are the same αs. Thus, the total number of unknowns is 4 (up to scale). Each view gives two equations in terms of α. Therefore, we need two views of the five points to compute all the view-independent coefficients.

When the points are not in general position, the rank of $\mathbf{C}$ would be less than 4 giving rise to simple algebraic constraints. A configuration of four points on a plane yields a view-independent constraint defined over two views. Three points on a line yield a view-independent constraint that can be computed from a single view itself. For linear motion, we can arrive at view-independent algebraic constraints by factoring out the camera parameters.

We derive several view-independent constraints on the projections of a dynamic scene in this paper. They are independent of the camera parameters. Some of these constraints are time-dependent while others are time-independent. The computational requirements of these constraints depend on the configuration and on its dependence on time. We also derive constraints on points with arbitrary planar motion under affine projection. These are computed from a Fourier domain representation of the trajectory. The constraints derived here find applications in recognition of dynamic point configurations in multiple views, time-alignment between views, etc.

2   Points with Linear Motion

We first consider the case of uniform linear motion. When a point moves in the world with uniform linear velocity or acceleration, its projections in various views move in a parameterizable manner. The view-independent relationships between projections of points moving with uniform velocity presented recently [6] fall under this category. The two view constraints on points moving with uniform velocity [7] is another contribution in this direction. In this section, we study the projection of points moving in a linear fashion imaged under affine and projective camera models. Let $\mathbf{P}$ be a 3D world point, moving with uniform

linear polynomial motion. Its position at any time instant t is given by

$$\mathbf{P}_t = \begin{bmatrix} \mathbf{I} \\ 1 \end{bmatrix} + \begin{bmatrix} \mathbf{Q}_1 \\ 0 \end{bmatrix} t + \begin{bmatrix} \mathbf{Q}_2 \\ 0 \end{bmatrix} t^2 + \ldots + \begin{bmatrix} \mathbf{Q}_n \\ 0 \end{bmatrix} t^n \tag{1}$$

where $\mathbf{I}$ is the initial position and $\mathbf{Q}_i$ are 3-vectors. Let $\mathbf{p}_t^l = \begin{bmatrix} x_t^l & y_t^l & 1 \end{bmatrix}^\mathrm{T}$ be the projection of $\mathbf{P}$ in view l at time t due to a camera characterized by the camera matrix $\mathbf{M}^l$.

2.1   Uniform Linear Motion under Affine Projection

When the camera is affine, we can differentiate the projection $\mathbf{p}_t^l$ with respect to t to get the velocity

$$\tilde{\mathbf{v}}_t^l = \mathbf{M}^l \sum_{i=1}^{n} \begin{bmatrix} \mathbf{Q}_i \\ 0 \end{bmatrix} t^{(i-1)}. \tag{2}$$

If the point moves with uniform velocity $\mathbf{U}$ in the world, the image velocity can be written as $\tilde{\mathbf{v}}^l = \mathbf{M}^l [\mathbf{U} \ 0]^\mathrm{T}$. Thus, the projected point moves with uniform velocity that is the projection of the world velocity. If the point moves with uniform linear acceleration, its image velocity is given by $\tilde{\mathbf{v}}^l = \mathbf{M}^l([\mathbf{U} \ 0]^\mathrm{T} + [\mathbf{A} \ 0]^\mathrm{T} t)$ and its image acceleration is given by $\tilde{\mathbf{a}}^l = \mathbf{M}^l \mathbf{A}$, where $\mathbf{A}$ is the world acceleration of the point. This implies that the projection of a point moving with uniform linear acceleration in the world has uniform linear acceleration [8]. Such simple parameterization is not available for the general projective camera.

2.2   Uniform Linear Motion under Perspective Projection

We consider the image motion of points undergoing uniform linear motion in the world. Since the point $\mathbf{P}_t$ projects to $\mathbf{p}_t^l = \mathbf{M}^l \mathbf{P}_t$, we can write x_t and y_t as

$$x_t = \frac{\sum_{i=0}^{n} \psi_i t^i}{\sum_{i=0}^{n} \phi_i t^i} \quad \text{and} \quad y_t = \frac{\sum_{i=0}^{n} \chi_i t^i}{\sum_{i=0}^{n} \phi_i t^i} \tag{3}$$

where ψ_i, ϕ_i, and χ_i are functions of $\mathbf{I}$, $\mathbf{Q}_i$, and $\mathbf{M}^l$ and hence constant for a point in a particular view. We can parameterize the projection of the point at time t with $3n + 2$ unknowns up to scale. These parameters can be computed from $\lceil (1.5n + 1) \rceil$ time instants since each time instant provides two equations.

We can parameterize the moving point as the intersection of the line of motion of the projection and lines perpendicular to it at various time instants. If $(b, -a, d)$ is the line of motion of the projection in the image over time, the line perpendicular to it can be written as $\mathbf{l}(t) = (a, b, c(t))$. Since a and b are constants, only $c(t)$ (a measure of the distance of the line from the origin) varies with time. Only two of the three parameters a, b, d are independent as a line is defined up to scale. Since $\mathbf{p}_t$ lies on $\mathbf{l}(t)$, we have $\mathbf{l}(t)^\mathrm{T} \mathbf{p}_t = 0$. Replacing $\mathbf{p}_t$ with $\mathbf{M}\mathbf{P}_t$ and expanding, we get

$$\sum_{i=0}^{n} \mu_i \, t^i + c(t) \, (\sum_{i=0}^{n} \eta_i \, t^i) = 0 \tag{4}$$

where μ_i and η_i are functions of $\mathbf{I}$, $\mathbf{Q}_i$, $\mathbf{M}^l$, a, and b, and are constant for a point in a view. The term $c(t)$ can be parameterized using $(2n + 1)$ unknowns up to scale. Since μ's and η's are functions of ϕ's, ψ's and χ's, no new information is gained by this parameterization. However, the time-dependent part of the motion can be parameterized using fewer parameters, by factoring the time-independent parts out. The point at time t can be obtained by taking the cross product of the lines $(b, -a, d)$ and $(a, b, c(t))$. This representation of the position of the projection has fewer *essential* unknowns than the parameterization of Equation 3.

Uniform Linear Velocity: If $n = 1$ Equation 4 becomes $\mu_0 + \mu_1 t + c(t)(\eta_0 + t) = 0$ with $\eta_1 = 1$. The parameterization will have 3 unknowns up to scale and needs 3 time instants to compute them. The parameters have been partitioned into time-dependent and time-independent parts. The line of motion $(b, -a, d)$ (2 unknowns up to scale) can be computed from projections at any two time instants. Together, the time-dependent and time-independent aspects make up the 5 degrees of freedom associated with the system.

Uniform Linear Acceleration: The simple parameterization gives x_t and y_t as ratios of two polynomials in t of degree 2, with 8 unknowns and needs measurements at 4 time instants to compute them. The new parameterization has only 5 unknowns and can be determined from 5 time instants. The polynomial constraint is given by $\mu_0 + \mu_1 t + \mu_2 t^2 + c(t)(\eta_0 + \eta_1 t + t^2) = 0$ with $\eta_2 = 1$.

2.3 General Linear Motion

Under general linear motion, the trajectories of the points will be straight lines and constraints on matching lines in multiple views are satisfied by each moving point independently. If a world line is imaged by projective cameras as l^1, l^2, and l^3, the projections are related by a trilinear constraint [3,4,9] as

$$l^1 = (l^2)^{\mathrm{T}} \mathcal{T} l^3 \tag{5}$$

where $\mathcal{T}$ is a suitable tensor. This gives a constraint on the trajectories of points undergoing general linear motion. Nothing more can be said about them since no more information is available other than the linearity of their trajectories.

3 Motion Analysis in Fourier Domain

If we have a number of moving points, their collective properties can be exploited in addition to the motion constraints. Properties of collections can be captured in the Fourier domain. We consider a configuration of a large number of points moving with independent uniform linear velocities in this section. We also explore Fourier domain representation of a point undergoing arbitrary co-planar motion.

3.1 Multiple Linearly Moving Points

Recognition of deformable shapes has been studied and applied to tracking of non-rigid objects when the deformation between two consecutive frames is

small [10,11], in the context of handwriting recognition [12,13], and for contour extraction and modeling [14]. Some approaches suggest learning a deformable model from examples, while some use deformable templates and ascertain a match by determining how much a template has to be deformed to get the test shape. These techniques do not assume any specific structure in the deformation. Our work on the other hand attempts to develop a sound theoretical model when the deformation has a particular structure.

Let $\mathbf{P}[i]$ be the sequence of N points moving with independent uniform linear velocities $\mathbf{V}[i]$ like points on the envelope of an evolving planar boundary. Let the projection of $\mathbf{P}[i]$ in view l at time t be $\mathbf{p}_t^l[i]$. A homography maps points in one view to points in the other [1]. If the homography is affine

$$\mathbf{p}_t^l[i] = \mathbf{A}^l \mathbf{p}_t^0[i] + \mathbf{b}^l, \quad 0 \leq i < N \tag{6}$$

where $\mathbf{A}^l$ is the upper 2×2 minor of the homography and $\mathbf{b}^l$ is taken from its third column. An unknown shift λ_l aligns the points between views 0 and l. Taking the Fourier transform of Equation 6 and ignoring the frequency term corresponding to $k = 0$, we get

$$\bar{\mathbf{P}}_t^l[k] = \mathbf{A}^l \bar{\mathbf{P}}_t^0[k] e^{j2\pi\lambda_l k/N}, \quad 0 < k < N \tag{7}$$

where $\bar{\mathbf{P}}_t^l = \begin{bmatrix} \mathbf{X}_t^l & \mathbf{Y}_t^l \end{bmatrix}^{\mathrm{T}}$; $\mathbf{X}_t^l$ and $\mathbf{Y}_t^l$ are the Fourier transforms of the sequences x_t^l and y_t^l respectively. A point moving with uniform velocity in the world moves with uniform velocity in an affine view (Section 2.1). The projection at any time t is given by $\mathbf{p}_t^l[i] = \mathbf{p}_0^l[i] + \mathbf{v}^l[i]t$, $\quad 0 \leq i < N$ where $\mathbf{v}^l[i] = [v_x^l[i]\ v_y^l[i]]^{\mathrm{T}}$ is the velocity vector in the image. Taking the Fourier Transform of both sides, we get

$$\bar{\mathbf{P}}_t^l[k] = \bar{\mathbf{P}}_0^l[k] + \bar{\mathbf{V}}^l[k]\ t \tag{8}$$

where $\bar{\mathbf{V}}^l$ is the Fourier Transform of the sequence $\mathbf{v}^l$. We define a sequence measure κ on $\bar{\mathbf{P}}_t^l$ as

$$\kappa_t^l[k] = \bar{\mathbf{P}}_t^l[k]^{*T} \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \bar{\mathbf{P}}_t^l[k], \quad 0 < k < N \tag{9}$$

Using Equations 7 and 8, it can be shown that

$$\kappa_t^l[k] = |\mathbf{A}^l|\ (\alpha_1[k] + \alpha_2[k]\ t + \alpha_3[k]\ t^2) \tag{10}$$

where α's are functions of measurements $\mathbf{p}^0$ and $\mathbf{v}^0$ (or their Fourier domain representation) made only in the reference view. The κ sequence and hence the α's are pure imaginary and can be computed by observing 2 frames in the reference view to determine position ($\mathbf{p}_0^0$) values and velocity ($\mathbf{v}^0$) values. No time synchronization or point correspondence is required between views as the shift term λ_l gets eliminated in κ. In the reference view, $\kappa_t^0[k] = \alpha_1[k] + \alpha_2[k]\ t + \alpha_3[k]\ t^2$. The sequence measure $\kappa_t^l[k]$ is thus view-independent but time-dependent.

3.2 Arbitrary Motion of a Point

Are there any view-independent constraints on a point undergoing arbitrary planar motion? The image of the point in any view will trace out a contour (closed or open) over time, which is the projection of its world trajectory. The problem reduces to the analysis of planar contours and view-independent constraints for planar contours will be applicable to the moving point. We now present the contour constraints presented in [15] to characterize arbitrary planar motion of a point under affine imaging conditions.

Let $\mathbf{P}[i]$ be the sequence of N points on the closed planar trajectory of a point and let $(x^l[i], y^l[i])$ be its images in view l. (The index i is a measure of time in this case as the point is at different locations at different times.) Assuming that the views are related by an affine homography, the points on the contour in view l are related to corresponding points on the contour in the reference view 0 as

$$\mathbf{p}^l[i] = \mathbf{A}^l \mathbf{p}^0[i] + \mathbf{b}^l, \ \ 0 \le i < N \tag{11}$$

where $\mathbf{A}^l$ and b^l are as in Equation 6. The time alignment information across views is not typically available. Taking the Fourier transform of Equation 11 and discarding the DC term, we get $\bar{\mathbf{P}}^l[k] = \mathbf{A}^l \bar{\mathbf{P}}^0[k] e^{j2\pi\lambda_l k/N}, \ \ 0 < k < N$ where λ_l is the time alignment parameter and $\bar{\mathbf{P}}^l$ the Fourier transform as in Equation 7. We can define a time-independent sequence measure κ^l similar to the one given in Equation 9. We can easily see

$$\kappa^l[k] = |A^l| \ \kappa^0[k]. \tag{12}$$

Thus, $\kappa[k]$ is a relative view-invariant sequence for the point having arbitrary motion. It can be computed in any view by tracking the point over time to construct the contour $\mathbf{p}[i]$.

4   Applications of View-Independent Constraints

We describe how the parameterizations and constraints developed in this paper can be applied to the problems of recognition and time-alignment.

4.1 Configuration of 4 Points under Affine Projection

In Section 2.1 we had parameterized the velocity and acceleration of the projection of a point moving with uniform linear polynomial motion. We now use those parameterizations to derive view independent constraints on configurations of 4 points moving with independent uniform linear motion parameters.

Equation 2 can be written as $v^l_x(t) = \mathbf{m}_1 \sum_i \mathbf{Q}_i \ t^{i-1}$ and $v^l_y(t) = \mathbf{m}_2 \sum_i \mathbf{Q}_i \ t^{i-1}$. Rearranging terms, we get

$$\left[(\sum_{j=1}^{n} \mathbf{Q}_j t^{j-1})^{\mathrm{T}} \ v^l_x(t) \right] [\mathbf{m}_1 \ -1]^{\mathrm{T}} = \left[(\sum_{j=1}^{n} \mathbf{Q}_j t^{j-1})^{\mathrm{T}} \ v^l_y(t) \right] [\mathbf{m}_2 \ -1]^{\mathrm{T}} = 0$$

where $[\mathbf{m}_i \; m_{i4}]$ is the i'th row of $\mathbf{M}^l$, and v_x^l and v_y^l are the x and y components of the point's velocity in view l. If we have at least four points, then we can form a set of equations of the form $\mathbf{C}\theta = 0$, where each row of the measurement matrix $\mathbf{C}$ consists of the unknown world point motion parameters $\mathbf{Q}_i$, and the velocities along the x or y coordinate. $\mathbf{C}$ is a rank deficient matrix with a maximum rank of 3. Equating its 4×4 determinant to 0 results in the following linear constraints.

$$\zeta_0 v_{1x}^l + \zeta_1 v_{2x}^l + \zeta_2 v_{3x}^l + \zeta_3 v_{4x}^l = \zeta_0 v_{1y}^l + \zeta_1 v_{2y}^l + \zeta_2 v_{3y}^l + \zeta_3 v_{4y}^l = 0 \qquad (13)$$

where ζ_i is a polynomial of order $3(n-1)$ in the time-parameter t. ζ_i's are view-independent as each has $3n-2$ terms that are functions of the world motion parameters. The total number of view independent parameters is $4(3n-2)-1 = 12n-9$ up to scale. Each time instant provides 2 equations in the unknowns; we need $(6n-4)$ measurements of the velocities in one or more frames in one or more views to compute the ζ_i values.

Uniform Linear Velocity: When the points move with independent uniform linear velocities, $n = 1$ and we get linear view and time independent constraints on the velocities of the projections. These constraints have 3 view independent coefficients, computing which needs the measurement of the velocities of the four points in 2 views.

These results are better than the Recognition Polynomials and Shape Tensors presented earlier [7,6]. A view independent representation of a configuration of stationary points could be constructed from 2 views of 4 points under orthographic projections [16]. This was extended to recognize human gait using 2 views of 5 points under scaled-orthographic projections [7]. Time-dependent constraints involving a single view of 5 points with uniform velocity is presented in [6] for affine projection. Our results yield view and time independent constraints involving 4 points in 2 views under general affine projection – which is a significant improvement.

Uniform Linear Acceleration: When the points in the configuration move with independent uniform linear accelerations, $n = 2$, giving us linear, view-independent, time-dependent constraints on the velocities of the projections. These constraints have 15 view independent coefficients computing which needs measuring the velocities of the four points at a total of 8 time instants in one or more views.

Proceeding in a similar manner and factoring out the camera parameters as above, we can formulate linear time and view independent constraints on the accelerations of the projections, which have the same form and computational requirements as the constraints on the velocities of the projections of points moving with uniform linear velocities.

4.2   Configurations of 5 Points under Projective Cameras

Invariants provide us with the ability to come up with representations of the features in a scene that do not depend on the view, and can prove to be extremely handy when processing information from multiple views. For instance, to recognize a configuration of five coplanar points from any view of the same, we can compute the cross ratio of areas of the projections of the five points, which

would be the same no matter which view we compute it in [2]. The cross-ratio of the areas of five points $\mathbf{x}[1]$, $\mathbf{x}[2]$, $\mathbf{x}[3]$, $\mathbf{x}[4]$, and $\mathbf{x}[5]$, no three of which are collinear, is defined as

$$cr(\mathbf{x}[1], \mathbf{x}[2], \mathbf{x}[3], \mathbf{x}[4], \mathbf{x}[5]) = \frac{\triangle_{\mathbf{x}[1]\mathbf{x}[2]\mathbf{x}[5]} \cdot \triangle_{\mathbf{x}[3]\mathbf{x}[4]\mathbf{x}[5]}}{\triangle_{\mathbf{x}[1]\mathbf{x}[3]\mathbf{x}[5]} \cdot \triangle_{\mathbf{x}[2]\mathbf{x}[4]\mathbf{x}[5]}} \tag{14}$$

where $\triangle_{\mathbf{x}[i]\mathbf{x}[j]\mathbf{x}[k]}$ is the area of the triangle formed by points $\mathbf{x}[i], \mathbf{x}[j], \mathbf{x}[k]$. This is for a static configuration of points or for snapshots of the scene taken at the same time. We now extend this to dynamic scenes where points move with uniform linear velocities or accelerations to arrive at time varying invariants for such configurations. Due to the novel parameterization for projective cameras described in the previous section, the number of unknowns needed to compute the time-varying invariants are fewer when compared to a naive parameterization approach.

Uniform Linear Velocity: If the points lie on a plane during the motion, the various views of the point configuration are related by a projective homography [1]. To express the configuration in a view-independent manner, we use an invariant to projective transformations of 2D [2]. Given the projections of a configuration of five coplanar points, which are in general position in the image, i.e., no three are collinear, we can define an invariant like the cross ratio of areas (Equation 14). The cross ratio of areas of the parametric representations of the projections of five points having independent uniform velocities is the ratio of two polynomials of degree 6 in the time parameter t.

$$I_v^l(t) = \frac{N_v^l(t)}{D_v^l(t)} = \frac{\sum_{i=0}^{6} \gamma^l t^i}{\sum_{i=0}^{6} \delta^l t^i}$$

where $I_v^l(t)$ is the invariant computed in view l at time t and γ^l and δ^l terms are functions of the parameters used to represent the points in view l. The number of essential unknowns in this expression is only 15 (3 for each point) and measurements made in only three time instants in each view are required to determine this time varying invariant. This is a significant theoretical advancement over the formulation presented in Levin et al. [6] that requires the projections of 6 points having coplanar independent uniform linear velocities, has 35 unknowns, computing which need 34 time instants.

To recognize a configuration, we need to determine whether the invariants computed in all the views are identical or not. This implies that

$$I_v^0(t) = I_v^l(t) \Rightarrow \frac{N_v^0(t)}{D_v^0(t)} = \frac{N_v^l(t)}{D_v^l(t)} \Rightarrow N_v^0(t) * D_v^l(t) = N_v^l(t) * D_v^0(t)$$

Therefore, for a configuration of 5 points moving with uniform linear velocities, the ratio of the coefficients of t^i in $N_v^0(t) * D_v^l(t)$ and $N_v^l(t) * D_v^0(t)$ should be 1 for $0 \le i \le 12$. This necessary constraint for recognition, however, holds only when time-alignment across views is known. For recognition, we can also make use of the additional necessary constraint that there should exist a unique homography that maps the lines of motion of the projection in the test and reference views.

Uniform Linear Acceleration: When all 5 points of a configuration moving with independent linear accelerations lie on the same plane always, we can define a time varying invariant for the configuration similar to the one above. The time varying invariant obtained on computing the cross ratio of the areas of the parametric projections of the configuration is the ratio of two polynomials of order 12 in the time parameter t and has only 25 unknowns (5 for each point) determining which need measurements made at 5 time instants.

$$I_a^l(t) = \frac{N_a^l(t)}{D_a^l(t)} = \frac{\sum_{i=0}^{12} \sigma^l t^i}{\sum_{i=0}^{12} \tau^l t^i}$$

where $I_a^l(t)$ is the invariant computed in view l at time t and the σ^l and τ^l terms are functions of the parameters used for representing the points in view l.

As in the case of uniform linear velocity, the value of the invariant computed in all the views have to be the same, which implies that the ratio of the coefficients of t^i in $N_a^0(t) * D_a^l(t)$ and $N_a^l(t) * D_a^0(t)$ should be 1 for $0 \leq i \leq 24$. Like in the case of uniform velocity, for recognition of the configuration, we can make use of the additional necessary constraint that there should exist a homography that relates the lines of motion in the two views.

4.3   Recognition Constraints in Fourier Domain

In Section 3.1 we had modeled configurations of many points having independent uniform linear velocities and their motion in the Fourier domain. In this subsection, we use those models to derive constraints for recognizing such configurations.

Configuration at the Same Time in Multiple Views: It has been shown in Equation 10 that

$$\kappa_t^l[k] = |\mathbf{A}^l| \kappa_t^0[k] \tag{15}$$

Equation 15 provides a recognition mechanism for such a case. Given M views, we can compute a $M \times (N-1)$ measurement matrix $\mathbf{C}_1$ constructed by stacking the κ_t^l measures for the various views, one row for each view. Since the various rows are scaled versions of each other, the rank of $\mathbf{C}_1$ would be 1. Therefore a necessary algebraic recognition constraint is $rank(\mathbf{C}_1) = 1$.

Configuration at Different Times in Multiple Views: The problem of recognizing the contour when we have its views at different time instants is a more challenging problem. Let us assume that in the reference view (0), we are able to track the points in two frames (identify points in a view across time) and hence able to identify all αs. Now given the configuration observed in any other view at any time t, we can recognize it to be the same as the one observed in the reference view. Observe that Equation 10 states that κ_t^l is a linear combination of the vectors α_i, the time t being a component of the linear combination coefficients. Given M views, we can construct a $(M+2) \times (N-1)$ measurement matrix $\mathbf{C}_2$ whose first three rows contain the vectors α_i, $i = 1, 2, 3$. The κ_t^l computed in the various views (except the reference view) then contribute one row each to $\mathbf{C}_2$. Note that the time instants at which κ is computed in a view need not be

the same in all views. Since, every row constructed from κ_t^l can be expressed as a linear combination of the first 3 rows, a necessary algebraic recognition constraint is $rank(\mathbf{C}_2) = 3$. This technique does not need correspondence across views and assumes tracking only in the reference view.

Recognizing Arbitrary Point Motion: In Section 3.2, we modeled the motion of a point moving on a closed arbitrary planar trajectory in the world as a contour and mapped the problem of its analysis to contour analysis. We evaluate the κ measure for the Fourier domain representation of the contour in view l. It can be shown that [15]

$$\kappa^l[k] = |\mathbf{A}^l| \; \kappa^0[k], \;\; 0 < k < N. \tag{16}$$

The κ values can be computed independently for each view from the Fourier domain. The κ sequence is invariant up to scale and can recognize the contour formed by the motion. Given M views of the motion, we can construct a $M \times (N-1)$ matrix $\mathbf{C}_a$, the ith row of $\mathbf{C}_a$ consisting of the κ values computed in the ith view. It can be seen from Equation 16 that rank of $\mathbf{C}_a$ is 1. This constraint is view-independent as the κ can be computed independently in each view. There are no restrictions on the number of frames in which the motion is observed. In practice the Fourier transform will be reliable only if the curve has sufficient length. If a number of points can be tracked independently, each contour will yield a different constraint, all of which have to be satisfied simultaneously. The above result hints that there can exist a number of algebraic constraints on the trajectory traced out by the projections of a moving point in a view.

4.4 Time Alignment

The recognition constraints presented here do not need time alignment information across views. We can determine time alignment using these constraints as we show next. This time alignment can then be used to align frames of synchronized videos captured from multiple viewpoints. We consider the problem wherein we have to time align two image sequences $\mathbf{A}$ and $\mathbf{B}$ of the same world motion. To do this, we need to determine the shift λ that when applied to $\mathbf{B}$ would ensure that the kth image in each sequence is a snap shot of the world at the k time instant.

Point Configurations of 5 Points: We can use the invariants described in Section 4.2 to time align views $\mathbf{A}$ and $\mathbf{B}$ of a configuration of 5 points moving with independent uniform linear velocities or accelerations. Let time t in view $\mathbf{A}$ be the time instant with reference to which we want to align view $\mathbf{B}$. In view $\mathbf{A}$, we compute the value of the invariant for the point configuration at time t and in view $\mathbf{B}$, we compute the parameters of the time varying invariant (for uniform velocity or acceleration as the case may be). We then perform a search over the range of possible values of λ seeking that shift at which the invariants computed at times t in $\mathbf{A}$ and $(t + \lambda)$ in $\mathbf{B}$ are identical.

Point Configurations of Many Points: The techniques for recognizing a deforming contour presented in Section 4.3 do not depend on the time instant at which the κ values are computed in a view. In fact, they can be used to determine the time

parameter. Let κ_τ^l be the κ sequence computed at time τ in view **B**. Normalizing $\kappa_\tau^l[k]$ (Equation 10) with respect to a fixed frequency (say p) gives

$$\frac{\kappa_\tau^l[k]}{\kappa_\tau^l[p]} = \frac{\alpha_1[k] + \alpha_2[k]t + \alpha_3[k]t^2}{\alpha_1[p] + \alpha_2[p]t + \alpha_3[p]t^2} \tag{17}$$

The αs can be computed in the reference view **A**, if we are able to track points in it for at least two frames. Equation 17 is a quadratic in time t, solving for which, we can find the time instant (frame number) in **A** corresponding to the time instant τ in **B**. The value of λ in this case is given by $(t - \tau)$.

Arbitrary Motion: In section 4.3 we have described how we can use the κ measure to recognize the projections of the closed planar trajectory of a point undergoing arbitrary motion. We can modify the definition of κ to define a new measure κ'

$$\kappa'^l[k] = \bar{\mathbf{P}}^l[k]^{*T} \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \bar{\mathbf{P}}^l[p] \quad = \kappa'^0[k]e^{-j2\pi\lambda_l(k-p)/N}, \quad 0 < k < N \tag{18}$$

where p is a constant (typically 1 or 2). The ratio of $\frac{\kappa'^l}{\kappa'^0}$ will be a complex sinusoid. The inverse Fourier transform of this quotient series would show a peak at λ. Thus, by looking for a peak in the inverse Fourier transform spectrum of the quotient series, we can determine time alignment information.

Note that we have considered $\mathbf{p}[i]$ and $\mathbf{v}[i]$ to be independent. In an application, one would expect them to be correlated and consequently the signals representing the sequences of positions and velocities would be smooth. As a result, the higher frequencies in their Fourier representation would be negligible and hence we can work with fewer frequencies in these cases.

5   Discussion

In this paper, we presented several constraints on the projections of coplanar points in motion. Linear motion with uniform velocity or acceleration and arbi-

Table 1. Motion: Summary of the multiview constraints on moving points in general position (unless stated otherwise). (CC = Coplanar Configuration)

Type	Camera	Conditions	Time Invariant	Source
Uniform V	Affine	5 pts, 8 frames	No	Levin et al.
Uniform V	Affine	4 pts, 2 views	Yes	Authors
Uniform V	Projective	6 pts, 49 frames, 1 view	No	Levin et al.
Uniform V	Projective	6 pts, 35 frames, 1 view, (CC)	No	Levin et al.
Uniform V	Projective	5 pts, 3 frames, 1 view, (CC)	No	Authors
Uniform A	Affine	4 pts, 9 frames	No	Authors
Uniform A	Affine	4 pts, 2 views	Yes	Authors
Uniform A	Projective	5 pts, 5 frames, 1 view, (CC)	No	Authors
Uniform ω	Projective	6 pts	Yes	Levin et al.
Uniform V	Affine	Many pts (CC)	No	Authors
Arbitrary	Affine	1 pt, Planar closed trajectory	No	Authors

trary planar motion were considered. Table 1 summarizes the constraints available on moving points. Our constraints have fewer computational requirements than published results. We showed how these constraints translate into recognition constraints. We also presented methods to compute the time-alignment between views from image structure only. These can form the basis of recognition applications like human identification using motion characteristics, tracking moving points for ballistic applications, detecting inconsistent video sequences of a dynamic scene based on geometric inconsistency, etc.

References

1. Hartley, R., Zisserman, A.: Multiple View Geometry in Computer Vision. Cambridge University Press (2000)
2. Mundy, J., Zisserman, A.: Geometric Invariances in Computer Vision. MIT Press (1992)
3. Hartley, R.: Lines and points in three views and the trifocal tensor. International Journal of Computer Vision **22** (1997) 125–140
4. Shashua, A.: Trilinear tensor: The Fundamental Construct of Multiple-view Geometry and its applications. AFPAC (1997) 190–206
5. Triggs, B.: Matching Constraints and the Joint Image. International Conference on Computer Vision (1995) 338–343
6. Levin, A., Wolf, L., Shashua, A.: Time-varying Shape Tensors for Scenes with Multiple Moving Points. IEEE Conference on Computer Vision and Pattern Recognition (2001) I:623–630
7. Carlsson, S.: Recognizing walking people. European Conference on Computer Vision (June 2000) 472–486
8. Kuthirummal, S., Jawahar, C.V., Narayanan, P.J.: Algebraic constraints on moving points in multiple views. Indian Conference on Computer Vision, Graphics and Image Processing (2002)
9. Hartley, R.: Lines and points in three views: An integrated approach. Proc. ARPA Image Understanding Workshop (1994)
10. Cohen, I., Ayache, N., Sulger, P.: Tracking points on deformable objects using curvature information. European Conference on Computer Vision (1992) 458–466
11. Kass, M., Witkin, A., Terzopoulos, D.: Snakes:active contour models. International Journal on Computer Vision **1(4)** (1988) 321–331
12. Tsukumo, J.: Handprinted kanji character recognition based on flexible template matching. International Conference on Pattern Recognition (1992) 483–486
13. Basri, R., Costa, L., Geiger, D., Jacobs, D.: Determining the similarity of deformable shapes. IEEE Workshop on Physics-Based Modelling in Computer Vision (1995) 135–143
14. Lai, K., R, C.: Deformable contours:modeling and extraction. IEEE Conference on Computer Vision and Pattern Recognition (1994) 601–608
15. Kuthirummal, S., Jawahar, C.V., Narayanan, P.J.: Planar Shape Recognition across Multiple Views. International Conference on Pattern Recognition (2002) 456–459
16. Bennet, B.M., Hoffman, D.D., Prakash, C.: Recognition Polynomials. Journal of the Optical Society of America **10** (1993) 759–764