Algebraic Constraints on Moving Points in Multiple Views

Sujit Kuthirummal C. V. Jawahar P. J. Narayanan
Centre for Visual Information Technology
International Institute of Information Technology
Gachibowli, Hyderabad 500 019
{sujit@gdit., jawahar@, pjn@}iiit.net

Abstract

Multiview analysis of scenes includes the study of scene-independent constraints satisfied by a configuration of cameras for all types of scenes as well as the study of view-independent constraints satisfied by any camera on a configuration of points. In this paper, we derive new constraints involving configurations of points that move with constant velocity, with constant acceleration, and for unconstrained planar motion. We show how these constraints can be applied to problems like motion recognition, frame alignment, etc.

1. Introduction

The analysis of multiple views of the same scene has received a lot of attention recently. Many algebraic relations have been discovered between primitives in multiple views. Most relate either points or lines across views. Scene independent multiview constraints, like the Fundamental matrix, the Trilinear Tensor and the Quadrilinear Tensor that relate two, three, and four views respectively [4, 5, 7, 8, 12, 13], encapsulate the viewing parameters such as the pose and the internal parameters of the cameras producing these views. View-independent constraints on configurations of points is an active area of research currently. The relationships between projections of points moving with uniform velocity presented recently [11] fall under this category.

Multiple view situations in Computer Vision have been analyzed with two objectives: to derive scene-independent constraints relating multiple views and to derive view-independent constraints relating the multiple scene points. The second approach overlaps with the Structure From Motion (SFM), which attempts to compute the 3D information of a set of world points undergoing a rigid motion from multiple observations using a single camera. Non-rigid motion is difficult to analyze in this scheme. The case of multiple objects moving with different velocities can be considered very close to the case of non-rigid motion. The algebraic relations between the projections of multiple, linearly moving objects in a scene [11] was shown to be view-independent, under the assumption that the points move with uniform velocity. The constraints relating multiple moving objects can be classified into time-dependent and time-independent relationships, similar to the scene and view dependence mentioned above. The two view constraints on points moving with a constant velocity is another noteworthy contribution in this direction [2].

In this paper, we study the algebraic relationships between moving objects imaged from different points of view. We derive time-dependent and time-independent relationships between the velocities and accelerations of the affine projections of moving objects in Section 2. We show in Section 3 that by observing an object motion over time using stationary cameras, we can arrive at time-independent constraints on the motion. Some applications of these are outlined in Section 4. Section 5 discusses how these relationships can be extended for the general perspective projections. We present a comprehensive analysis of the multiview relations in Section 6. Section 7 presents a few concluding remarks.

2. Uniform Motion of Points

We first discuss the case of linear motion of points viewed using different cameras. The motion is characterized by uniform velocity or uniform acceleration. We derive the conditions for affine cameras in this section. The case of perspective cameras is discussed in Section 5.

2.1. Uniform Velocity

Let P be a 3D world point, moving with uniform linear velocity. Let $\mathbf{I} = \begin{bmatrix} I_x & I_y & I_z \end{bmatrix}^T$ be its initial position and $\mathbf{U} = \begin{bmatrix} U_x & U_y & U_z \end{bmatrix}^T$ be its world velocity. These vectors are represented as homogeneous vectors $\tilde{\mathbf{I}} = \begin{bmatrix} I_x & I_y & I_z & 1 \end{bmatrix}^T$ and $\tilde{\mathbf{U}} = \begin{bmatrix} U_x & U_y & U_z & 0 \end{bmatrix}^T$. Its position at any time instant t is given by

$$P = \tilde{\mathbf{I}} + \tilde{\mathbf{U}}\ t \quad (1)$$

Let an affine camera observe the motion of the point. Let $p = \begin{bmatrix} x & y & 1 \end{bmatrix}^T$ be the projection of P due to the affine camera matrix $\mathbf{M} = \begin{bmatrix} \mathbf{m_1} & m_{14} \\ \mathbf{m_2} & m_{24} \\ \mathbf{0} & 1 \end{bmatrix}$. where $\mathbf{m_i}$ represents a vector of the first three elements in the ith row of $\mathbf{M}$. Then,

$$p = \mathbf{M}\tilde{\mathbf{I}} + \mathbf{M}\tilde{\mathbf{U}}\ t. \tag{2}$$

Differentiating with respect to t, we get $\tilde{\mathbf{v}} = \mathbf{M}\tilde{\mathbf{U}}$ where $\tilde{\mathbf{v}} = [v_x, v_y, 0]^T$ is the velocity vector in the image. This implies that the velocity of a point in the image is a projection of the world velocity. The above can be expanded as

$$v_x = \mathbf{m_1}. \begin{bmatrix} U_x & U_y & U_z \end{bmatrix} \tag{3}$$
$$v_y = \mathbf{m_2}. \begin{bmatrix} U_x & U_y & U_z \end{bmatrix} \tag{4}$$

Equations 3 and 4 can be written as

$$\begin{bmatrix} U_x & U_y & U_z & v_x \end{bmatrix} \begin{bmatrix} \mathbf{m_1} & -1 \end{bmatrix}^T = 0$$

and

$$\begin{bmatrix} U_x & U_y & U_z & v_y \end{bmatrix} \begin{bmatrix} \mathbf{m_2} & -1 \end{bmatrix}^T = 0$$

If we have 4 points in the scene, P_i, $1 \leq i \leq 4$, with world velocities $\begin{bmatrix} U_{ix} & U_{iy} & U_{iz} \end{bmatrix}^T$ and image velocities $\begin{bmatrix} v_{ix} & v_{iy} \end{bmatrix}^T$, we can define a matrix $\mathbf{X_v}$ as

$$\mathbf{X_v} = \begin{bmatrix} U_{1x} & U_{1y} & U_{1z} & v_{1x} \\ U_{2x} & U_{2y} & U_{2z} & v_{2x} \\ U_{3x} & U_{3y} & U_{3z} & v_{3x} \\ U_{4x} & U_{4y} & U_{4z} & v_{4x} \end{bmatrix} \tag{5}$$

Similarly, we can define a matrix $\mathbf{Y_v}$ with the last column having the velocity vectors in y-direction. We can see that

$$\begin{aligned} \mathbf{X_v} \begin{bmatrix} \mathbf{m_1} & -1 \end{bmatrix}^T &= \mathbf{0} \quad \text{and} \\ \mathbf{Y_v} \begin{bmatrix} \mathbf{m_2} & -1 \end{bmatrix}^T &= \mathbf{0}. \end{aligned}$$

For these to have a constraint, $\mathbf{X_v}$ and $\mathbf{Y_v}$ must be rank deficient, i.e., $|\mathbf{X_v}| = |\mathbf{Y_v}| = 0$. We get the following functions of the velocities of the 3D points and the image velocities by expanding the expressions for the determinants.

$$\begin{aligned} \alpha_0 v_{1x} + \alpha_1 v_{2x} + \alpha_2 v_{3x} + \alpha_3 v_{4x} = 0 \\ \alpha_0 v_{1y} + \alpha_1 v_{2y} + \alpha_2 v_{3y} + \alpha_3 v_{4y} = 0 \end{aligned} \tag{6}$$

where α's depend only on the world velocity parameters of the 4 points. The α's are view-independent. That is, the above constraints hold for the image velocities of four points for the same α's irrespective of the pose and intrinsic parameters of the camera. They are also time-independent as time term has been eliminated. Equation 6 has four unknowns, with each view providing two equalities. Therefore, we need two views of the four points to determine all α's up to scale.

These results are similar to the Recognition Polynomials and Shape Tensors presented or discovered earlier. It was shown that polynomials to recognize a configuration of stationary points could be constructed from 2 views of 4 points under orthographic projections [1]. This was extended to recognize human gait using 2 views of 5 points under scaled-orthographic projections [2]. Time-dependent constraints involving a single view of 5 points with uniform velocity is presented in [11] for affine projection.. Our results yield view and time independent constraints involving 4 points in 2 views under the general affine projection.

2.2. Uniform Acceleration

We now derive relationships between points when they move with constant acceleration. Let P be a 3D world point, moving with uniform linear acceleration. Let its position at any time instant t be given by

$$P = \tilde{\mathbf{I}} + \tilde{\mathbf{U}}t + \frac{1}{2}\tilde{\mathbf{A}}t^2 \tag{7}$$

where $\tilde{\mathbf{I}}$ is the initial position of the point, $\tilde{\mathbf{U}}$ is its initial velocity and $\tilde{\mathbf{A}} = \begin{bmatrix} A_x & A_y & A_z & 1 \end{bmatrix}^T$ is its constant acceleration.

Proceeding the same way as in the previous subsection, we get a singular matrix $\mathbf{X_a}$

$$\mathbf{X_a} = \begin{bmatrix} (U_{1x} + A_{1x}t) & (U_{1y} + A_{1y}t) & (U_{1z} + A_{1z}t) & v_{1x} \\ (U_{2x} + A_{2x}t) & (U_{2y} + A_{2y}t) & (U_{2z} + A_{2z}t) & v_{2x} \\ (U_{3x} + A_{3x}t) & (U_{3y} + A_{3y}t) & (U_{3z} + A_{3z}t) & v_{3x} \\ (U_{4x} + A_{4x}t) & (U_{4y} + A_{4y}t) & (U_{4z} + A_{4z}t) & v_{4x} \end{bmatrix}$$

A similar singular matrix $\mathbf{Y_a}$ with motion parameters in y-direction also exists.

Expanding $|\mathbf{X_a}|$ and $|\mathbf{Y_a}|$, we get

$$\begin{aligned} (\beta_0 + \beta_1 t + \beta_2 t^2 + \beta_3 t^3) v_{1x} + \\ (\beta_4 + \beta_5 t + \beta_6 t^2 + \beta_7 t^3) v_{2x} + \\ (\beta_8 + \beta_9 t + \beta_{10} t^2 + \beta_{11} t^3) v_{3x} + \\ (\beta_{12} + \beta_{13} t + \beta_{14} t^2 + \beta_{15} t^3) v_{4x} &= 0 \end{aligned} \tag{8}$$

$$\begin{aligned} (\beta_0 + \beta_1 t + \beta_2 t^2 + \beta_3 t^3) v_{1y} + \\ (\beta_4 + \beta_5 t + \beta_6 t^2 + \beta_7 t^3) v_{2y} + \\ (\beta_8 + \beta_9 t + \beta_{10} t^2 + \beta_{11} t^3) v_{3y} + \\ (\beta_{12} + \beta_{13} t + \beta_{14} t^2 + \beta_{15} t^3) v_{4y} &= 0 \end{aligned} \tag{9}$$

where β's depend only on the parameters of motion of the 3D points in the world. The above relations are time-dependent and view-independent. That is, the same β's hold no matter what the pose and intrinsic parameters of the affine camera used to view them. There are 16 unknowns $(\beta_0 \ldots \beta_{15})$ in the above relation, with each time instant providing 2 equations. We, therefore, need the velocities of 4 points for 8 time instants (9 frames) for computing the

β's. Note that these β's can be computed from a single view, as opposed to the α's for the case of constant velocity, which needed two views. This is the direct result of time-dependence.

We now proceed to derive time-independent constraints for the case of constant linear acceleration in the world. For this, we differentiate the constant acceleration motion equation (Equation. 7) twice to get

$$a_x = \mathbf{m_1}. \begin{bmatrix} A_x & A_y & A_z \end{bmatrix} \quad (10)$$

$$a_y = \mathbf{m_2}. \begin{bmatrix} A_x & A_y & A_z \end{bmatrix} \quad (11)$$

where a_x and a_y are the image-accelerations of the projection of the point. We can similarly define the singular matrices $\mathbf{X_a'}$ and $\mathbf{Y_a'}$ for the four points P_i, $1 \le i \le 4$. $\mathbf{X_a'}$ is given below.

$$\mathbf{X_a'} = \begin{bmatrix} A_{1x} & A_{1y} & A_{1z} & a_{1x} \\ A_{2x} & A_{2y} & A_{2z} & a_{2x} \\ A_{3x} & A_{3y} & A_{3z} & a_{3x} \\ A_{4x} & A_{4y} & A_{4z} & a_{4x} \end{bmatrix} \quad (12)$$

Expanding the determinants of $\mathbf{X_a'}$ and $\mathbf{Y_a'}$, we get

$$\begin{aligned} \gamma_0 a_{1x} + \gamma_1 a_{2x} + \gamma_2 a_{3x} + \gamma_3 a_{4x} = 0 \\ \gamma_0 a_{1y} + \gamma_1 a_{2y} + \gamma_2 a_{3y} + \gamma_3 a_{4y} = 0, \end{aligned} \quad (13)$$

where γ's are functions of world accelerations parameters of the 4 points only. The γ's are view and time independent. The system of Equations 13 has four unknowns and we need 2 views of the 4 points to determine all the γ's.

Levin et al. [11] derive constraints for motion constrained to elliptic paths. It is the first time that view or time independent constraints for points moving with constant linear acceleration have been derived.

3. Modelling Trajectory as a Contour

Assumptions of linear motion with constant velocity or acceleration are valid in many situations. Can we arrive at such constraints for general non-rigid motion of points in space? That would be most beneficial. We derive algebraic constraints for such a situation in this section.

Our approach is based on the following observation. A moving 3D point traces out a contour or curve in space over time. This contour gets mapped to a contour in the image. If we have multiple views of the same object motion, their consistency reduces to the matching of corresponding shapes. Any contour-matching approach can be used for this step. If the 3D motion of the point is restricted to a plane, the problem reduces to planar contour recognition.

The matching constraints using vector Fourier representation presented in [10] can yield algebraic constraints on points moving arbitrarily, as long as the motion is planar. The analysis of the motion of the object can be carried out by studying the contours traced out in the various views. Many surveillance applications involve studying the motion of an object (like vehicles on the ground) from cameras that are far away from them (at top of tall buildings or on satellites). The trajectory of the objects is restricted to a plane and the cameras are affine in practice in this case. We demonstrate how the necessary conditions for matching a planar contour across multiple views [10] can yield a view and time independent constraint for arbitrary planar motion. We first consider the case of linear motion before considering arbitrary non-rigid motion.

3.1. Linear Motion

Linear motion includes motion along a world straight line with no restrictions on the velocity, acceleration, etc. In this case, the trajectories of the points will be straight lines. Constraints that hold for matching lines in multiple views can be used on each moving point independently. For example, the trifocal tensor can relate lines in three views. If a world line is imaged as l^1, l^2, and l^3 in three views, they are related by a trilinear constraint [6, 7, 12].

$$l^1 = (l^2)^T \mathcal{T} l^3 \quad (14)$$

where $\mathcal{T}$ is a suitable tensor.

3.2. Arbitrary Motion

Let **O** be a set of N points on the planar trajectory of a point and let $\mathbf{P}_l$ be its images using an affine camera in view $\mathbf{V}_l$, where l is the view index. Let $(u^l[i], v^l[i])$ be the image coordinates of points on the contour traced out in view $\mathbf{V}_l$. We represent this contour using a sequence of vectors of complex numbers as given below.

$$\mathbf{x}^l[i] = \begin{bmatrix} u^l[i] + j0 \\ v^l[i] + j0 \end{bmatrix}$$

Under affine projection, the image-to-image homography between a pair of views is affine also. Thus, the corresponding points of the contour in view l are related to the points in the reference view 0 by the relation

$$\mathbf{x}^l[i] = \mathbf{A}_l \mathbf{x}^0[i] + \mathbf{b_l}, \;\; 0 \le i < N \quad (15)$$

where $\mathbf{A_l}$ is an arbitrary 2×2 matrix. Taking the Fourier transform of Equation 15, we get

$$\mathbf{X}^l[k] = \mathbf{A_l} \mathbf{X}^0[k] + \mathbf{b_l}\delta[0], \;\; 0 \le k < N$$

which can be rewritten as (ignoring the DC component k = 0)

$$\mathbf{X}^l[k] = \mathbf{A_l} \mathbf{X}^0[k], \;\; 0 < k < N \quad (16)$$

where $\mathbf{X}^l = \left[\mathbf{U}^l, \ \ \mathbf{V}^l\right]^T$; $\mathbf{U}^l$ and $\mathbf{V}^l$ are Fourier transform sequences of u^l and v^l respectively. We can now define the following measure κ for the points on the contour in the view l as

$$\kappa^l[k] = (\mathbf{X}^l[k])^{*T} \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \bar{\mathbf{X}}^l[k], \ \ 0 < k < N \tag{17}$$

(* denotes the complex conjugate). It can be shown that [10]

$$\kappa^l[k] = |\mathbf{A_l}| \ \kappa^0[k], \ \ 0 < k < N. \tag{18}$$

where, $|\mathbf{A_l}|$ denotes the determinant of $\mathbf{A_l}$. The κ values defined by Equation 17, which can be computed independently for each view from the Fourier transform of the contour points, identify the contour formed by the motion, upto a scale factor. Consider the following $M \times (N-1)$ matrix for M views of the planar contour

$$\Theta = \begin{bmatrix} \kappa^0[1] & \kappa^0[2] & \cdots & \kappa^0[N-1] \\ \kappa^1[1] & \kappa^1[2] & \cdots & \kappa^1[N-1] \\ \cdots & \cdots & \cdots & \cdots \\ \kappa^{M-1}[1] & \cdots & \cdots & \kappa^{M-1}[N-1] \end{bmatrix}. \tag{19}$$

It can be seen from Equation 18 that rank$(\Theta) = 1$. See reference [10] for a detailed discussion on rank constraints for shape matching.

Equation 19 gives a constraint on measures that can be computed from each view independently. The trajectory of a point moving in a plane can be tracked over time to generate the contour in each view. This constraint is view and time independent. There are no restrictions on the number of views or frames per se. In practice, however, the Fourier transform will be reliable only if the curve has sufficient length. The motion of the point is arbitrary. If a number of points can be tracked independently, each contour will yield a different constraint, all of which have to be satisfied simultaneously. It is clear that non-rigid motion is also covered by these constraints.

4. Applications

In this section, we show the applications of view independent constraints for a number of common problems. The goal of this effort is not to solve these problems in the best possible way. Rather, we hint at the diverse applications to which these constraints can contribute positively.

4.1. Frame Alignment

Suppose we have two video sequences $\mathbf{A}$ and $\mathbf{B}$ of a world motion taken by two affine cameras. We want to align them so that the same time instant can be identified. A few solutions to this problem have appeared in the literature [3, 9].

Uniform Acceleration Motion: If we can identify points moving with constant acceleration, Equations 8 and 9 would hold for both views for the same β's. The time t can be replaced with the frame number. From the image velocities of the projections of 4 points in 8 frames in view $\mathbf{A}$, β's that characterize the point configuration can be computed. We want to identify the corresponding frame k in view $\mathbf{A}$ for the frame j in view $\mathbf{B}$. The image velocities of the projections of the four points in view $\mathbf{B}$ at time instant j are (v_{ixj}, v_{iyj}), $1 \leq i \leq 4$. Therefore, the shift is $(k-j)$. We have

$$\begin{aligned} &(\ \beta_0 + \beta_1 k + \beta_2 k^2 + \beta_3 k^3 \ )v_{1xj} + \\ &(\ \beta_4 + \beta_5 k + \beta_6 k^2 + \beta_7 k^3 \ \ )v_{2xj} + \\ &(\beta_8 + \beta_9 k + \beta_{10} k^2 + \beta_{11} k^3)v_{3xj} + \\ &(\ \beta_{12} + \beta_{13} k + \beta_{14} k^2 + \beta_{15} t^3)v_{4xj} \quad = \quad 0 \end{aligned} \tag{20}$$

And a similar relation in v_y values. These can be written as

$$\begin{aligned} \eta_1(k)v_{1xj} + \eta_2(k)v_{2xj} + \eta_3(k)v_{3xj} + \eta_4(k)v_{4xj} = 0 \\ \eta_1(k)v_{1yj} + \eta_2(k)v_{2yj} + \eta_3(k)v_{3yj} + \eta_4(k)v_{4yj} = 0 \end{aligned} \tag{21}$$

where $\eta_i(k) = (\beta_{i*4} + \beta_{i*4+1}k + \beta_{i*4+2}k^2 + \beta_{i*4+3}k^3)$, $0 \leq i \leq 3$. We can solve for k using a linear least squares solution technique by minimizing the sum of squares of the error functions

$$\begin{aligned} f_1(k) = \eta_1(k)v_{1xj} + \eta_2(k)v_{2xj} + \eta_3(k)v_{3xj} + \eta_4(k)v_{4xj} \\ f_2(k) = \eta_1(k)v_{1yj} + \eta_2(k)v_{2yj} + \eta_3(k)v_{3yj} + \eta_4(k)v_{4yj}. \end{aligned}$$

Alternately, we can solve for the roots of a cubic polynomial of the form

$$\gamma_0 k^3 + \gamma_1 k^2 + \gamma_2 k + \gamma_3 = 0 \tag{22}$$

where $\gamma_0 = (\beta_3 v_{1xj} + \beta_7 v_{2xj} + \beta_{11} v_{3xj} + \beta_{15} v_{4xj}), \gamma_1 = (\beta_2 v_{1xj} + \beta_6 v_{2xj} + \beta_{10} v_{3xj} + \beta_{14} v_{4xj}), \gamma_2 = (\beta_1 v_{1xj} + \beta_5 v_{2xj} + \beta_9 v_{3xj} + \beta_{13} v_{4xj}),$ and $\gamma_3 = (\beta_0 v_{1xj} + \beta_4 v_{2xj} + \beta_8 v_{3xj} + \beta_{12} v_{4xj})$.

General Planar Motion If points undergoing general planar motion can be tracked across time, the trajectory of each forms a contour which is viewed by both video cameras. Both videos see the same contour, but the starting points are different. The problem reduces to contour matching under affine homography and unknown shift. Solutions for this situation using a measure similar to κ has been presented in [10]. The new measure is

$$\begin{aligned} \kappa'_p(l)[k] \quad &= \quad (\mathbf{X}^l[k])^{*T} \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \mathbf{X}^l[p] \\ &= \quad |\mathbf{A}_l| \ \kappa'_p(0)[k] \ e^{-j2\pi\lambda_l(k-p)/N} \end{aligned} \tag{23}$$

where p is a constant (typically 1 or 2) and λ_l is the unknown shift in view l compared to the reference view 0. The ratio $\frac{\kappa'_p(l)[k]}{\kappa'_p(0)[k]}$ will be a complex sinusoid. The shift λ_l can be recovered from the inverse Fourier transform of this ratio. For more details, see [10].

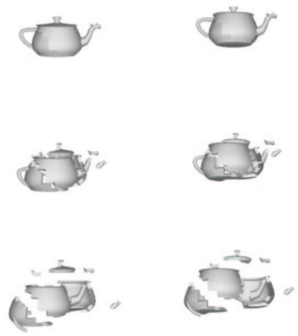
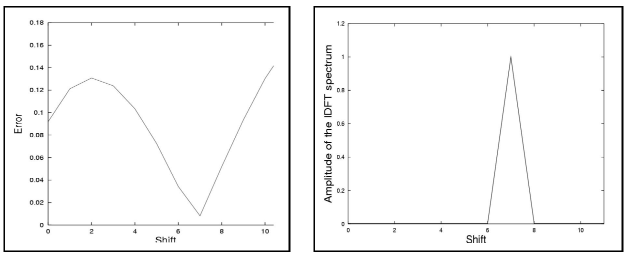


Figure 1: Two image sequences of an exploding pot



(a) (b)

Figure 2: Alignment of sequences (a) Searching over the range of possible shifts, (b) Shift determination in Fourier Domain. (See text for more details)

Experimental Results To determine the validity of this technique for frame alignment, tests were conducted on a number of scenes. A simulated explosion of the teapot, with the various fragments flying off in different directions with different but constant acceleration, was used for experimentation. Figure 1 shows a few frames from the image sequences of the exploding tea pot from two different view points. Tests were carried out to evaluate the applicability of this technique for frame alignment, by varying the starting point of the second video sequence. In all the cases the proper shift value was recovered. In the first experiment, the two sequences were misaligned by seven frames. The fit-error graph obtained on searching over the possible range of shifts, for aligning the sequences is shown in Figure 2(a). Figure 2 (b) shows the amplitude spectrum of the IDFT of the ratio $\frac{\kappa'_p(l)[k]}{\kappa'_p(0)[k]}$ computed from two views of general planar motion, misaligned by seven frames. It can be seen that there is a distinct peak at the proper shift value.

4.2. Recognition

Recognition of motion using single view constraints was reported earlier [11]. Single view constraints can be used to index into motions of a person performing a sitting movement, for instance. The same philosophy can be used for recognition using our constraints, which need less points.

It is also possible to recognize objects undergoing nonrigid motion using these constraints, as long as parts have uniform velocity or uniform acceleration. Suppose we can track N sets of 4 points on the deforming body in all the views. From 8 frames of each view and each set of points, a set of β's can be computed. The β's for a set of points computed from any view will satisfy the constraints expressed in Equation 21 for that set of points, for all views. This should be true for all N sets of points, for the body to be the same in all the views. This recognition strategy was tested on views of the exploding tea pot scene. A sets of points on the tea pot were tracked across the explosion in all the views. It was found that the β values computed for each set of points from one view were valid for the same set of points in all views. The coefficients computed were highly similar for a specific motion and different for different motion. These coefficients may be used for gait recognition or similar applications.

4.3. View Consistency

These constraints can be used for checking the consistency of a system of views involving motion in a multiview environment. This is easy to see as the constraints we derived are view independent and should come out to be the same in all views. If we can track N sets of 4 points each for many time instant in each view, we can compute N sets of β values from the reference view. The same β values will hold for all views for each set of points at every time instant. Given the image velocities in the other views and the β values computed from the reference view, the matching time instants can be computed for each view, as in Subsection 4.1. A necessary condition for all the views to belong to the same motion configuration is that all computed N time instants that align the views be congruent.

5. Perspective Cameras

We take a look at how the above relationships can be extended to perspective cameras. Suppose that we have a world point P that moves as per Equation 1. Let this point be viewed by a perspective camera represented by the camera matrix $\mathbf{M}$. The projection of P will trace a line in the image. Let us represent this line l by $\begin{bmatrix} a & b & c \end{bmatrix}^T$. Let p be the projection of P due to $\mathbf{M}$ at some time instant t.

$$l^T p = 0$$

$$\begin{bmatrix} a & b & c \end{bmatrix} \mathbf{M} \begin{bmatrix} (I_x + U_x t) & (I_y + U_y t) & (I_z + U_z t) & 1 \end{bmatrix}^T = 0$$

Differentiating w.r.t. t, we get

$$(a\mathbf{m_1} + b\mathbf{m_2} + c\mathbf{m_3}) . \begin{bmatrix} U_x & U_y & U_z \end{bmatrix} = 0$$

Type	Conditions	Invariant	Source
Stationary	Multiple	Scene	Many
Uniform V	5 pts, 8 frames	View	Levin et al.
Uniform V	4 pts, 2 views	View, Time	This paper
Uniform A	4 pts, 9 frames	View	This paper
Uniform A	4 pts, 2 views	View, Time	This paper
Uniform ω	6 pts	View, Time	Levin et al.
Planar	A few, 1 view	View, Time	This paper

Table 1: Summary of the multiview constraints

Proceeding in a similar manner as before to eliminate the projection terms m_{ij}, we can derive a view-independent relation for uniform velocity motion under perspective projection. However, the number of unknowns are very high and hence the number of views or time instants needed to solve for them runs into a large number. Extension to constant acceleration motion also has similar properties under perspective projection. Levin et al. sketch the extension of their view-independent constraints to perspective cameras. Their extension required 6 points in 49 time instants under perspective projection, compared to 5 points in 8 time instants under affine projection. Similar ideas must be applied to reduce the view requirements to practical levels. We are actively pursuing this currently.

6. Discussion on Multiview Relations

We summarize the different multiview constraints and their characteristics in this section. We are dealing with the class of view-independent constraints that hold good from any view for a configuration of points. Table 1 summarizes the different relations reported in literature.

Constraints on stationary points are essentially the multilinear relations encoded in the Fundamental Matrix, Trilinear Tensor, etc. These are scene-independent. The number of points required to recover these depends on the number of degrees of freedom. Since the points are stationary, time has no relevance.

Levin et al. [11] give view-independent, time-dependent constraints involving 5 points and 8 time instants. In this paper, we gave new view and time independent constraints involving 4 points in 2 views for this case. These work for affine cameras. Both of these are extensible to the perspective camera, but considerably more points or time instants will be required. We also showed how this problem can be reduced to that of stationary lines using the trajectory of the moving points.

This case hasn't been addressed earlier in literature. We derived new view-independent, time-dependent set of constraints for affine cameras that require 4 points and 8 time instants. We also gave time and view independent constraints involving 4 points in 2 views.

Levin et al. derived view-independent constraints for the case of constant angular velocity along an elliptic path.

This problem has not been addressed earlier. We derived constraints by reducing this problem to that of planar contour matching. Our constraints are view and time independent and require a minimum of 1 point for a sufficient number of time instants to evolve a reasonable contour.

7. Conclusions and Future Work

We derived several novel view-independent constraints that characterize the configuration of points moving under affine projection in this paper. The motion models under which such constraints are valid include constant velocity motion, constant acceleration motion, and general planar motion. We outlined the application of these constraints for frame alignment, view consistency, and recognition. Future extensions include general non-planar motion of points and general perspective cameras for projection.

References

[1] B. M. Bennet, D. D. Hoffman, and C. Prakash. Recognition Polynomials. *Journal of the Optical Society of America*, 10:759–764, 1993.

[2] S. Carlsson. Recognizing walking people. *European Conference on Computer Vision*, June 2000.

[3] Y. Caspi and M. Irani. Alignment of non-overlapping sequences. *ICCV*, 2:76–83, 2001.

[4] O. Faugeras. *Three Dimensional Computer Vision*. MIT Press, 1992.

[5] O. Faugeras and Q. Luong. *The Geometry of Multiple Images*. MIT Press, 2001.

[6] R. Hartley. Lines and points in three views: An integrated approach. *Proc. ARPA Image Understanding Workshop*, 1994.

[7] R. Hartley. Lines and points in three views and the trifocal tensor. *IJCV*, 22:125–140, 1997.

[8] R. Hartley and A. Zisserman. *Multiple View Gemoetry in Computer Vision*. Cambridge University Press, 2000.

[9] S. Kuthirumal, C. V. Jawahar, and P. J. Narayanan. Video frame alignment in multiple views. *International Conference on Image Processing*, 2002.

[10] S. Kuthirummal, C. V. Jawahar, and P. J. Narayanan. Planar Shape Recognition across Multiple Views. *International Conference on Pattern Recognition*, 2002.

[11] A. Levin, L. Wolf, and A. Shashua. Time-varying Shape Tensors for Scenes with Multiple Moving Points. *IEEE Conference on Computer Vision and Pattern Recognition*, 2001.

[12] A. Shashua. Trilinear tensor: The Fundamental Construct of Multiple-view Geometry and its applications. *Int. Worskshop on AFPAC*, 1997.

[13] B. Triggs. Matching Constraints and the Joint Image. *International Conference on Computer Vision*, pages 338–343, 1995.