

On the Spacetime Geometry of Galilean Cameras

Yaser Sheikh
Robotics Institute
Carnegie Mellon University
yaser@cs.cmu.edu

Alexei Gritai
Computer Vision Laboratory
University of Central Florida
agritsay@cs.ucf.edu

Mubarak Shah
Computer Vision Laboratory
University of Central Florida
shah@cs.ucf.edu

Abstract

In this paper, a projection model is presented for cameras moving at constant velocity (which we refer to as Galilean cameras). To that end, we introduce the concept of spacetime projection and show that perspective imaging and linear pushbroom imaging are specializations of the proposed model. The epipolar geometry between two such cameras is developed and we derive the Galilean fundamental matrix. We show how six different “fundamental” matrices can be directly recovered from the Galilean fundamental matrix including the classic fundamental matrix, the Linear Pushbroom (LP) fundamental matrix and a fundamental matrix relating Epipolar Plane Images (EPIs). To estimate the parameters of this fundamental matrix and the mapping between videos in the case of planar scenes we describe linear algorithms and report experimental performance of these algorithms.

1. Introduction

A camera is normally thought of as a device that generates two dimensional images of a three dimensional world. This “projective engine” ([5]) takes a single snapshot of the world from a particular position at a particular time instant. The camera, however, is often dynamic and the output of cameras is better considered a video rather than an image. We reexamine the process of projection, not between the static world and an image plane but instead where an uncalibrated camera is moving with (unknown) constant velocity. We refer to such a camera as a *Galilean camera*¹ and model the projection of the world onto the video hyperplane. The epipolar geometry of a pair of *static* cameras has been exhaustively studied (more than two decades of research summarized by Hartley and Zisserman in [10], Faugeras and Luong in [5]), and we show that this concept can be generalized for Galilean cameras.

¹Paying homage to the observers in Galileo’s principle of relativity. Equation 4 further justifies this choice.

Our work is related to that of Wolf and Shashua in [16], where they investigate higher-dimensional mappings between k -spaces and 2-spaces that arise from different problem instances for $3 \leq k \leq 6$. However, where they provide six problem definitions describing various configurations of points moving in straight lines, we study the geometry of cameras moving with constant velocity. Bartoli in [1], Sturm in [15] and Han and Kanade in [8] all also make similar assumptions about objects moving along straight lines.

In this paper, we describe a spacetime projection model for a Galilean camera and propose a mapping function between the videos of two Galilean cameras when the scene is planar. We then present the epipolar geometry for this case and describe a normalized linear algorithm for estimating the parameters of the “fundamental” matrix relating Galilean cameras. We show how the original fundamental matrix, the LP fundamental matrix, the ortho-perspective fundamental matrix and three, as yet unknown, fundamental matrices can be directly recovered from this Galilean fundamental matrix.

The rest of this paper is organized as follows. In Section 2 we introduce the Galilean projection model used in the remainder of this paper, followed by Section 3 where we present the mapping between Galilean images when the scene is planar. A description of the relative geometry between two Galilean cameras and the resulting fundamental matrix is presented in Section 4 and specializations of this matrix to several known and unknown fundamental matrices are described in Section 5. Finally, experimental validation is presented in Section 7.

2. Galilean Projection Model

We define a *worldpoint* as $\mathbf{X} = [T \ X \ Y \ Z]^T \in \mathcal{R}^4$, on a world coordinate $\mathbf{U} = [T \ \lambda X \ \lambda Y \ \lambda Z \ \lambda]^T$ and a *videopoint* as $\mathbf{x} = [t \ u \ v]^T \in \mathcal{R}^3$ on a video coordinate system $\mathbf{u} = [t \ wu \ wv \ w]^T \in \mathcal{R}^4$. Note that an additional time dimension has been added to the usual spatial terms. When the world and camera coordinate systems are aligned, the mapping describing central projection for the spatial coordinates and orthographic projection for the temporal coordinates

dinate are,

$$(T, X, Y, Z)^T \mapsto (\alpha_t T, fX/Z + p_u, fY/Z + p_v)^T \quad (1)$$

where f is the focal length of the camera and α_t is the reciprocal of the frame-rate of the camera (causing an effect akin to time dilation) and (p_u, p_v) are the coordinates of the principal point. This can be expressed in matrix form as

$$\begin{bmatrix} T \\ X \\ Y \\ Z \end{bmatrix} \mapsto \begin{bmatrix} t \\ wu \\ wv \\ w \end{bmatrix} = \begin{bmatrix} \alpha_t & & & \\ & f & & p_u \\ & & f & p_v \\ & & & 1 \end{bmatrix} \begin{bmatrix} T \\ X \\ Y \\ Z \end{bmatrix}, \quad (2)$$

or more concisely $\mathbf{u} = \mathbf{KX}$ where \mathbf{K} is the calibration matrix. When the spatial world and camera coordinate systems are not aligned they are related by rotation and translation. The temporal coordinates are related by a translation (e.g. the world time index when camera begins recording). These transformations can be captured by a 4×4 orthogonal matrix \mathbf{Q} and a 4×5 displacement matrix \mathbf{D} , where

$$\mathbf{Q} = \begin{bmatrix} 1 & & & \\ & \mathbf{R} & & \\ & & & \\ & & & \end{bmatrix}, \mathbf{D} = \begin{bmatrix} 1 & & & -D_t \\ & 1 & & -D_x \\ & & 1 & -D_y \\ & & & 1 & -D_z \end{bmatrix}, \quad (3)$$

where $\mathbf{C} = [D_t, D_x, D_y, D_z]^T$ is the position of the camera center and \mathbf{R} is a 3×3 rotation matrix representing the orientation of the camera coordinate system. The 4×5 projection matrix relates the world and video coordinate systems, $\mathbf{u} = \mathbf{PU}$. This projection matrix can be decomposed as $\mathbf{P} = \mathbf{KQD} = \mathbf{KQ}[\mathbf{I} - \mathbf{C}]$ or simply $\mathbf{P} = \mathbf{K}[\mathbf{Q} - \mathbf{QC}]$.

Now if the cameras are moving at constant velocity according to $\Delta\mathbf{C} = [1, \Delta D_x, \Delta D_y, \Delta D_z]^T$, we have the following series if we consider only the spatial dimensions²,

$$\begin{aligned} \hat{\mathbf{u}}(0) &= \hat{\mathbf{K}}\mathbf{R}[\mathbf{I} - \hat{\mathbf{C}}]\hat{\mathbf{U}} \\ \hat{\mathbf{u}}(1) &= \hat{\mathbf{K}}\mathbf{R}[\mathbf{I} - (\hat{\mathbf{C}} + \Delta\hat{\mathbf{C}})]\hat{\mathbf{U}} \\ &\vdots \\ \hat{\mathbf{u}}(T) &= \hat{\mathbf{K}}\mathbf{R}[\mathbf{I} - (\hat{\mathbf{C}} + T\Delta\hat{\mathbf{C}})]\hat{\mathbf{U}}. \end{aligned}$$

If we include the temporal dimension into the object vector we can rewrite these compactly as $\mathbf{u} = \mathbf{KQ}[\mathbf{G} - \mathbf{C}]\mathbf{U}$, where

$$\mathbf{G} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ -\Delta D_x & 1 & 0 & 0 \\ -\Delta D_y & 0 & 1 & 0 \\ -\Delta D_z & 0 & 0 & 1 \end{bmatrix} \quad (4)$$

²Entities with a hat denote the spatial entries of the entity, e.g. $\hat{\mathbf{U}} = [X, Y, Z, 1]^T$, $\hat{\mathbf{Q}} = \mathbf{R}$ etc.

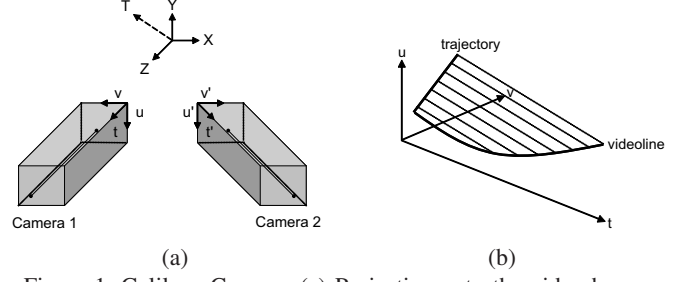


Figure 1. Galilean Cameras.(a) Projection onto the video hyperplane (b) The videoline of a point charts out a curve in spacetime.

is a Galilean transformation. We refer to the 4×5 matrix

$$\mathcal{M} = \mathbf{KQ}[\mathbf{G} - \mathbf{C}] \quad (5)$$

as the *Galilean projection matrix*. As with the spatial projection matrix, the null vector of \mathcal{M} corresponds to the spacetime location of the camera center in the world at $t = 0$. In addition, $m_{12} = m_{13} = m_{14} = 0$, where $\mathcal{M} = \{m_{ij}\}$. The video taken by a Galilean camera can therefore be properly considered a three-dimensional image projected from a four-dimensional world. It is a multiperspective (noncentral) camera in the sense described in [14] and [13], the generator being a line in 4D spacetime.

Analogous to worldlines in spacetime geometry [3], we refer to the curve charted out by successive world events from a (spatially) static point as *videolines*. It was shown by Bolles *et al.* in [2] that these curves are described hyperbolic functions on EPIs, but in the video hyperplane (assuming that the world reference frame is aligned with the camera reference frame) they follow the parametric form,

$$u(T) = p_u + \frac{-f\Delta D_x T + fX}{-\Delta D_z T + Z} \quad (6)$$

$$v(T) = p_v + \frac{-f\Delta D_y T + fY}{-\Delta D_z T + Z} \quad (7)$$

$$t(T) = \alpha_t T. \quad (8)$$

It should be noted then that straight lines in the spacetime world are not mapped to straight lines in the video hyperplane, except when the principal axis is orthogonal to the velocity vector (in which case $D_z = 0$ and Z is constant and therefore Equations 6 and 7 are linear). As a result, spatial invariants such as the cross-ratio are not preserved in spacetime.

3. Planar Geometry

In this section, we describe a transformation analogous to the planar homography relating two images of a world plane. By choosing two orthogonal basis vectors that span the scene plane as the X and Y axes of the world coordinate

system and ignoring the perpendicular Z coordinate (since all Z values will equal zero), we have,

$$\mathbf{u} = \begin{bmatrix} t \\ wu \\ wv \\ w \end{bmatrix} = \mathcal{M}'_4 \begin{bmatrix} T \\ X \\ Y \\ 1 \end{bmatrix}, \mathbf{u}' = \begin{bmatrix} t' \\ w'u' \\ w'v' \\ w' \end{bmatrix} = \mathcal{M}'_4 \begin{bmatrix} T \\ X \\ Y \\ 1 \end{bmatrix} \quad (9)$$

where \mathcal{M}'_4 and \mathcal{M}'_4 are nonsingular 4×4 matrices, constructed by removing the fourth column from \mathcal{M} and \mathcal{M}' respectively. There exists a transformation relating \mathbf{u} and \mathbf{u}' , $\mathcal{H} = \mathcal{M}'_4 \mathcal{M}_4^{-1}$ where $\mathbf{u}' = \mathcal{H}\mathbf{u}$. Considering time independently we see that,

$$t = m_{11}T + m_{14}, t' = m'_{11}T + m'_{14}, \\ \frac{t - m_{14}}{m_{11}} = \frac{t' - m'_{14}}{m'_{11}} = T,$$

from which we get the mapping $t' = h_{11}t + h_{14}$, or in other words, $h_{12} = h_{13} = 0$. As a result, we get the following functions to determine t' , u' and v' ,

$$t' = h_{11}t + h_{14}, \\ u' = \frac{h_{21}t + h_{22}u + h_{23}v + h_{24}}{h_{41}t + h_{42}x + h_{43}y + h_{44}}, \\ v' = \frac{h_{31}t + h_{32}u + h_{33}v + h_{34}}{h_{41}t + h_{42}x + h_{43}y + h_{44}},$$

Thus, a nonsingular 4×4 matrix \mathcal{H} relates the spacetime coordinates of two videos captured by Galilean cameras observing a planar scene.

Definition 3.1 (Planar Galilean Mapping) A planar Galilean mapping is a linear transformation of $\mathbf{u} = [t \ u \ v \ 1]^T$, representable as a nonsingular 4×4 matrix \mathcal{H} ,

$$\begin{bmatrix} t' \\ w'u' \\ w'v' \\ w' \end{bmatrix} = \begin{bmatrix} h_{11} & 0 & 0 & h_{14} \\ h_{21} & h_{22} & h_{23} & h_{24} \\ h_{31} & h_{32} & h_{33} & h_{34} \\ h_{41} & h_{42} & h_{43} & h_{44} \end{bmatrix} \begin{bmatrix} t \\ u \\ v \\ 1 \end{bmatrix}. \quad (10)$$

This matrix \mathcal{H} is an inhomogeneous matrix that can be divided into an inhomogeneous part, i.e. the first row, \mathbf{h}^1 and a homogeneous part, i.e. the second, third and fourth rows, \mathbf{h}^2 , \mathbf{h}^3 and \mathbf{h}^4 . Unlike the planar homography, this mapping does not form a group or in other words the product of two planar Galilean mappings is not, in general, a planar Galilean mapping.

To estimate the parameters of this mapping, the homogeneous and inhomogeneous parts can be computed separately. The Direct Linear Transformation Algorithm (see [10]) can be used to estimate the homogeneous part since,

$$\begin{bmatrix} u'_i \\ v'_i \\ w'_i \end{bmatrix} \times \begin{bmatrix} \mathbf{u}_i^T \mathbf{h}^2 \\ \mathbf{u}_i^T \mathbf{h}^3 \\ \mathbf{u}_i^T \mathbf{h}^4 \end{bmatrix} = \mathbf{0}. \quad (11)$$

An over-determined homogeneous system of equations can be constructed as,

$$\begin{bmatrix} \mathbf{0}^T & -w'_i \mathbf{u}_i^T & v'_i \mathbf{u}_i^T \\ w'_i \mathbf{u}_i^T & \mathbf{0}^T & -u'_i \mathbf{u}_i^T \\ -v'_i \mathbf{u}_i^T & u'_i \mathbf{u}_i^T & \mathbf{0}^T \end{bmatrix} \begin{bmatrix} \mathbf{h}^2 \\ \mathbf{h}^3 \\ \mathbf{h}^4 \end{bmatrix} = \mathbf{0}, \quad (12)$$

and the solution can be found using SVD (see Section 4.1 in [10] for further details). For the inhomogeneous part, the following linear system of equations can be solved using least squares,

$$[\mathbf{u}_i]^T [h_{11} \ h_{13}]^T = t'_i. \quad (13)$$

4. Two View Geometry

Consider a pair of Galilean cameras that move in different directions at different velocities³. The coordinates of the corresponding projections in first and second camera are $\mathbf{u} = (t, uw, vw, w)^T$ and $\mathbf{u}' = (t', u'w', v'w', w')^T$ respectively. The imaged coordinates in the two cameras are $\mathbf{u} = \mathcal{M}\mathbf{U}$ and $\mathbf{u}' = \mathcal{M}'\mathbf{U}$. This pair of equations may be rewritten as

$$\mathcal{A}\mathbf{g} = \mathbf{0} \quad (14)$$

where,

$$\mathcal{A} = \begin{bmatrix} m_{11} & 0 & 0 & 0 & m_{15} - t & 0 & 0 \\ m_{21} & m_{22} & m_{23} & m_{24} & m_{25} & u & 0 \\ m_{31} & m_{32} & m_{33} & m_{34} & m_{35} & v & 0 \\ m_{42} & m_{42} & m_{43} & m_{44} & m_{45} & 1 & 0 \\ m'_{11} & 0 & 0 & 0 & m'_{15} - t' & 0 & 0 \\ m'_{21} & m'_{22} & m'_{23} & m'_{24} & m'_{25} & 0 & u' \\ m'_{31} & m'_{32} & m'_{33} & m'_{34} & m'_{35} & 0 & v' \\ m'_{41} & m'_{42} & m'_{43} & m'_{44} & m'_{45} & 0 & 1 \end{bmatrix}, \quad (15)$$

m_{ij} are the elements of \mathcal{M} and $\mathbf{g} = [T, X, Y, Z, 1, -w, -w']^T$. Since \mathcal{A} in the homogeneous system of Equation 14 is a 8×7 matrix, it must have a rank of at most six for a solution to exist. As a result, any 7×7 minor must have a zero determinant. There are eight different ways to choose the 7×7 minor to solve the system, but only two interesting variations. The first selection uses both rows containing the temporal indices (t, t') and five rows containing the spatial indices (u, v, u', v') and the second selection uses one row containing the temporal indices and six rows containing the spatial indices. As in [7], $\det(\mathcal{A}_i) = 0$ will produce the fundamental polynomial that has interaction terms (between u, v, t, u', v' and t' but no squared terms. Hence, there exists a 6×6 matrix called the Galilean fundamental matrix,

$$(t'u', t'v', t', u', v', 1)\mathbf{\Gamma}(tu, tv, t, u, v, 1)^T = 0. \quad (16)$$

³By the principal of relativity we can assume one of the cameras to be stationary, but to maintain a symmetric formulation between both cameras we do not make that assumption here.

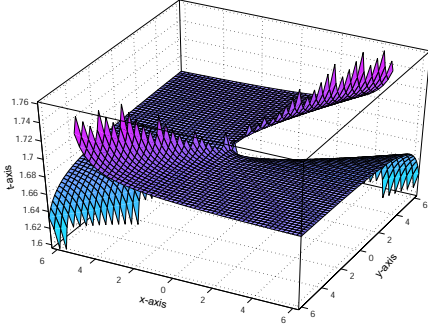


Figure 2. Epipolar Surface. The videopoint in the second video corresponding to a space-time point in the first video must lie on this surface.

However, in all the eight variations (of different minors), nine interaction terms do not exist all, i.e., out of a total of 36 possible interaction terms only 27 appear.

Definition 4.1 (Galilean Fundamental Matrix) If \mathbf{u} and \mathbf{u}' are videopoints corresponding to the same worldline under two Galilean cameras, there exists a 6×6 matrix $\mathbf{\Gamma}$ referred to as the Galilean Fundamental Matrix such that,

$$\begin{pmatrix} t'u' \\ t'v' \\ t' \\ u' \\ v' \\ 1 \end{pmatrix}^T \begin{pmatrix} 0 & 0 & 0 & f_1 & f_2 & f_3 \\ 0 & 0 & 0 & f_4 & f_5 & f_6 \\ 0 & 0 & 0 & f_7 & f_8 & f_9 \\ f_{10} & f_{11} & f_{12} & f_{13} & f_{14} & f_{15} \\ f_{16} & f_{17} & f_{18} & f_{19} & f_{20} & f_{21} \\ f_{22} & f_{23} & f_{24} & f_{25} & f_{26} & f_{27} \end{pmatrix} \begin{pmatrix} tu \\ tv \\ t \\ u \\ v \\ 1 \end{pmatrix} = 0.$$

$\mathbf{\Gamma}$ can be written more compactly as,

$$\mathbf{\Gamma} = \begin{pmatrix} \mathbf{0} & \Delta\mathbf{F}' \\ \Delta\mathbf{F} & \mathbf{F}_{00} \end{pmatrix}, \quad (17)$$

where \mathbf{F}_{00} is the fundamental matrix between the image in the first video at time $t = 0$ and the image in the second video at time $t' = 0$, and $(\Delta\mathbf{F}, \Delta\mathbf{F}')$ are matrices that capture information about the velocity of each camera as will be seen presently.

4.1. Epipolar Geometry

Unless there is zero motion, no epipoles (single image points of the opposite camera center) in the usual sense exist. In general, epipolar lines (or curves) in the usual sense do not exist either, instead there are epipolar surfaces in one camera corresponding to a point in the other camera. These surfaces are defined by setting a spatiotemporal point in one camera, i.e. (t', u', v') and applying the Galilean fundamental matrix. The surface is defined by the

$$s_1tu + s_2tv + s_3t + s_4u + s_5v + s_6 = 0$$

where $\mathbf{s} = [s_1, \dots, s_6]$ is computed as

$$\mathbf{s} = [t'u', t'v', t', u', v', 1]\mathbf{\Gamma}.$$

An example of this surface is plotted in Figure 2. It can be seen that this surface is ruled, since the intersection with each time plane is a line (corresponding to the classic epipolar line of that image).

5. Specializations

The work by Feldman *et al.* in [6] on Crossed-Slits projection, considers the algebraic constraints between different slices of the spacetime volume. We show that several such specializations can be directly derived from the constraints described in this paper. The different specializations are shown in Figure 3(a), (b) and (c) for the original fundamental matrix, the orthoperspective fundamental matrix and the linear pushbroom fundamental matrix respectively. Proceeding similarly it is straightforward to recover the fundamental matrices for the configurations in Figure 3(d), (e) and (f).

5.1. Between Perspective Images

The classic fundamental matrix between two uncalibrated perspective images was derived independently by Faugeras in [5] and Hartley in [9]. For corresponding points, this singular 3×3 matrix satisfies the constraint $[u', v', 1]\mathbf{F}[u, v, 1]^T = 0$. This matrix can be directly recovered from the Galilean fundamental matrix. For the (t, t') pair image we can recover the fundamental matrix $\mathbf{F}_{tt'}$ by partially collapsing $\mathbf{\Gamma}$ and plugging in the values of (t, t') . Thus $\mathbf{F}_{tt'} =$

$$\begin{pmatrix} f_1t' + f_{10}t + f_{13} & f_2t' + f_{11}t + f_{14} & f_3t' + f_{12}t + f_{15} \\ f_4t' + f_{16}t + f_{19} & f_5t' + f_{17}t + f_{20} & f_6t' + f_{18}t + f_{21} \\ f_7t' + f_{22}t + f_{25} & f_8t' + f_{23}t + f_{26} & f_9t' + f_{24}t + f_{27} \end{pmatrix},$$

or simply

$$\mathbf{F}_{tt'} = \Delta\mathbf{F}t + \Delta\mathbf{F}'t' + \mathbf{F}_{00}. \quad (18)$$

Thus, $\Delta\mathbf{F} + \mathbf{F}_{00}$ is the fundamental matrix between the image in the first video at time $t = 0$ and the second video at time $t' = 1$ and $\Delta\mathbf{F}' + \mathbf{F}_{00}$ is the fundamental matrix between the image in the first video at time $t = 1$, and the second video at time $t' = 0$. We can infer the following property from Equation 18.

Theorem 1. (Fundamental Boost Matrix) The matrices $(\Delta\mathbf{F}, \Delta\mathbf{F}')$ are rank-2 matrices. As a result, the rank of $\mathbf{\Gamma}$ is at most 5.

Proof. At $t = 0$ and $t' = 0$, a fundamental matrix \mathbf{F}_{00} can be decomposed into $\hat{\mathbf{K}}'^{-T}[-\mathbf{R}\hat{\mathbf{C}}]_{\times}\mathbf{R}\hat{\mathbf{K}}^{-1}$. At $t = 1$, if the second camera is displaced by $\Delta\hat{\mathbf{C}}$ then the fundamental matrix becomes,

$$\begin{aligned} & \hat{\mathbf{K}}'^{-T}[-\mathbf{R}(\hat{\mathbf{C}} + \Delta\hat{\mathbf{C}})]_{\times}\mathbf{R}\hat{\mathbf{K}}^{-1} \\ &= \hat{\mathbf{K}}'^{-T}[-\mathbf{R}\hat{\mathbf{C}}]_{\times}\mathbf{R}\hat{\mathbf{K}}^{-1} - \hat{\mathbf{K}}'^{-T}[\mathbf{R}\Delta\hat{\mathbf{C}}]_{\times}\mathbf{R}\hat{\mathbf{K}}^{-1}. \end{aligned}$$

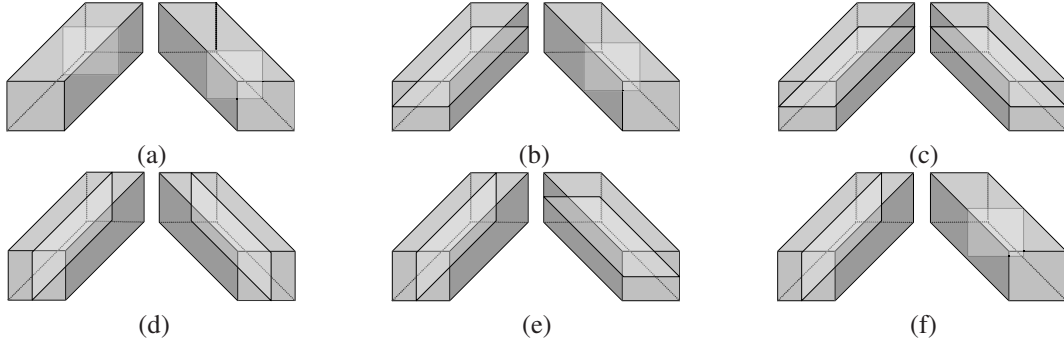


Figure 3. Specializations. “Fundamental” matrices can be recovered between (a) a pair of perspective images, (b) an EPI and a perspective image, (c) a pair of EPIs, (d) a pair of LP images, (e) a LP and an EPI, (f) a LP and a perspective images

From Equation 18,

$$\begin{aligned} &= \mathbf{F}_{00} + \Delta\mathbf{F} \\ &= \mathbf{F}_{10}. \end{aligned}$$

Since $[\Delta\hat{\mathbf{C}}]_{\times}$ is a skew-symmetric matrix, it follows that $\Delta\mathbf{F}$ is a rank 2 matrix. The left 6×3 submatrix or the upper 3×6 submatrix of $\mathbf{\Gamma}$ are both therefore rank 2 matrices, and $\mathbf{\Gamma}$ has a rank of at most 5. \square

5.2. Between Linear Pushbroom Images

If the camera motion satisfies the conditions described in [7], linear pushbroom images can be recovered from a horizontal slice of the video volume. Between two such images, the LP fundamental matrix was derived by Gupta and Hartley in [7]. The relationship captured by this matrix is expressed as $(t'v', t', v', 1)\mathbf{F}_{uu'}(tv, t, v, 1)^T = 0$. This 4×4 matrix too can be directly derived from $\mathbf{\Gamma}$. Thus, given (u, u') , we can compute $\mathbf{F}_{uu'} =$

$$\begin{pmatrix} 0 & 0 \\ 0 & 0 \\ f_{17} & f_{18} + f_{16}u \\ f_{11}u' + f_{23} & f_{22}u + f_{24} + (f_{10}u + f_{12})u' \\ f_5 & f_4u + f_6 \\ f_{2u'} + f_8 & f_7u + f_9 + (f_1u + f_3)u' \\ f_{20} & f_{19}u + f_{21} \\ f_{14}u' + f_{26} & f_{25}u + f_{27} + (f_{13}u + f_{15})u' \end{pmatrix}. \quad (19)$$

It can be observed that the structure of the matrix is the same as the one derived in [7].

5.3. Between Epipolar Plane Images

Epipolar plane images were defined by Bolles *et al.* in [2] as the collection of epipolar lines that correspond to one epipolar plane in the world. We can recover the fundamental matrix between two EPIs. In this case it has a form similar

to the LP fundamental matrix,

$$\begin{pmatrix} 0 & 0 \\ 0 & 0 \\ f_{10} & f_{12} + f_{11}v \\ f_{16}v' + f_{22} & f_{23}v + f_{24} + (f_{17}v + f_{18})v' \\ f_1 & f_2v + f_3 \\ f_{2v'} + f_7 & f_8v + f_9 + (f_5v + f_6)v' \\ f_{13} & f_{14}v + f_{15} \\ f_{19}v' + f_{25} & f_{26}v + f_{27} + (f_{20}v + f_{21})v' \end{pmatrix}. \quad (20)$$

5.4. Between a Pushbroom and a Perspective Image

Recently in [11], Khan *et al.* derived the 4×3 perspective-orthoperspective fundamental matrix between a pushbroom image and a perspective image. The relationship captured by this matrix is expressed as $(t'v', t', v', 1)\mathbf{F}_{u't}(u, v, 1)^T = 0$. This matrix can also be directly derived from $\mathbf{\Gamma}$. Thus, given (t, u') , we can compute $\mathbf{F}_{tu'} =$

$$\begin{pmatrix} f_4 & f_5 \\ f_1u' + f_7 & f_2u' + f_8 \\ f_{16}t + f_{19} & f_{17}t + f_{20} \\ f_{10}tu' + f_{22}t + f_{13}u' + f_{25} & f_{11}tu' + f_{23}t + f_{14}u' + f_{26} \\ f_6 & f_3u' + f_9 \\ f_{18}t + f_{21} & f_{12}tu' + f_{24}t + f_{15}u' + f_{27} \end{pmatrix},$$

or simply $\begin{pmatrix} \phi\Delta\mathbf{F}' \\ \phi t\Delta\mathbf{F} + \mathbf{F}_{00} \end{pmatrix}$, where $\phi = \begin{pmatrix} 0 & 1 & 0 \\ u' & 0 & 1 \end{pmatrix}$. Similarly, it is straightforward to recover $\mathbf{F}_{vu'}$ the fundamental matrices between an EPI and an LP image, and $\mathbf{F}_{tv'}$ between an EPI and a perspective image.

6. Normalized Linear Algorithm

A linear algorithm can be used to estimate the parameters of $\mathbf{\Gamma}$. Equation 17 can be rewritten as the homogeneous system $A\gamma = 0$ where $\gamma = [f_1, \dots, f_{27}]^T$ is a 27-vector, constructed from the non-zero elements of $\mathbf{\Gamma}$ and A is a

Objective

Given $n \geq 26$ matches from corresponding videolines, estimate the Galilean fundamental matrix Γ such that $\mathbf{p}'^T \Gamma \mathbf{p} = 0$.

Algorithm

1. **Normalization:** Normalize the coordinates through an appropriate scaling and translation.
2. **Linear Solution:** Perform singular value decomposition on A and determine Γ by selecting the singular vector corresponding to the smallest singular value of A and reconstructing a 6×6 matrix.
3. **Rank Constraint:** Set the smallest singular value of Γ to zero, enforce the rank 5 constraint.
4. **Denormalization:** Denormalize Γ according to the original scaling and translation.

Figure 4. A linear algorithm for estimating Γ

matrix constructed from spatiotemporal coordinates of corresponding points. If noiseless points correspond exactly to each other the rank of A is 26 and the null-vector of A corresponds to an estimate of γ . In the presence measurement noise, the 27th singular value will be non zero. In that case, the singular vector corresponding to the smallest singular value of A can be used as an estimate of γ . The rank constraints on Γ can be enforced post facto by setting the smallest singular values of Γ to zero and reconstructing the matrix. Of course, as with other linear algorithms of this sort, to obtain good estimates it is important to appropriately normalize the data (see Section 4.4 of [10]). Lastly, to obtain a meaningful solution, A has to have a rank of more than 26. It is emphasized that to ensure the rank of A is greater than 17, correspondences from different videopoints of the same worldpoint must be used. For instance, if videolines of a static world point in the scene of length n are associated in both cameras, there are n^2 rows that should be added to A .

7. Experimentation

An experiment was conducted where two cameras were placed on a moving walkway 8 feet apart, looking at different angles and moving at approximately 2 miles per hour in the same direction. A pair of 1000 frame sequences were recorded at a resolution of 240×360 by two SONY HDV cameras (images were downsampled) and 22 videopoints were tracked across 6 frames in each of the two videos (frames 971 to 976 in both sequences). The motion of the cameras was not perpendicular to the optical axis of either camera. At 30 fps, the distance traversed by both cameras during this period was about 98 feet, and the distance traveled in between successive frames was approximately 1.2 inches. Three slices of this video are shown in

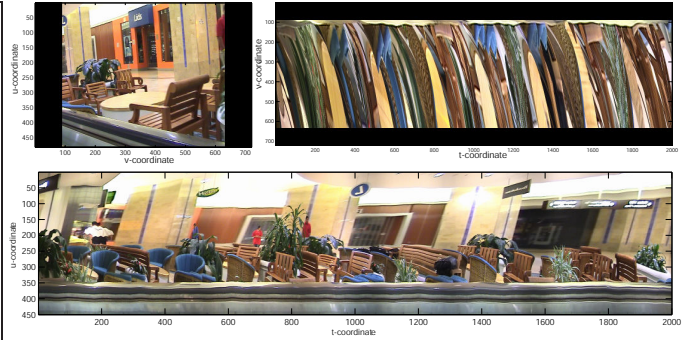


Figure 5. Three slices of a video taken from a Galilean camera.

Figure 5 (the results in this paper are best seen in color). These points were used to estimate the Galilean Fundamental matrix using the linear algorithm presented in this paper. To evaluate this estimate, different time slices of the video were analyzed using the different specializations of the Galilean Fundamental matrix. A set of 6 points (different from the ones used during computation) were selected in both frames and the points and the epipolar lines of their correspondences between two perspective images are plotted in Figure 7. Despite the fact that the videopoints correspondences were taken towards the end of the sequence and the frames in this figure were taken towards the beginning (frame 1 from video 1 and frame 20 from video 2) the fundamental matrix recovered is accurate. This demonstrates that Γ may be used to predicting fundamental matrices in the future without having point correspondences for those frames. The epipolar curves induced by points in frames 960, 980 and 1000 onto the pushbroom image in video 1 generated from column 250 are shown in Figure 8. As the corresponding frames move, the asymptote translates in the LP image translates too. Figure 7 shows frame 980 with the points used to plot the curves in Figures 8 and 7. The corresponding epipolar lines from the EPI are missing since it is difficult to find point correspondences on EPIs. Nine points lay on a plane and were used to compute the planar Galilean mapping shown in Figure 6. The figure shows the first sequence where the yellow boxes show the positions of the nine points in that sequence and the black points indicate the positions of the points in Sequence 2 (a) before and (b) after warping.

8. Discussion

In this work, we present a spacetime projection model for cameras moving at constant velocities. In practice, the assumption of constant velocity is often reasonable for short durations of time, especially when the camera is mounted on a robot, elevator or on a vehicle such as an aircraft, train, car or a spacecraft. An important application for the ideas described in this paper is for prediction of relative camera

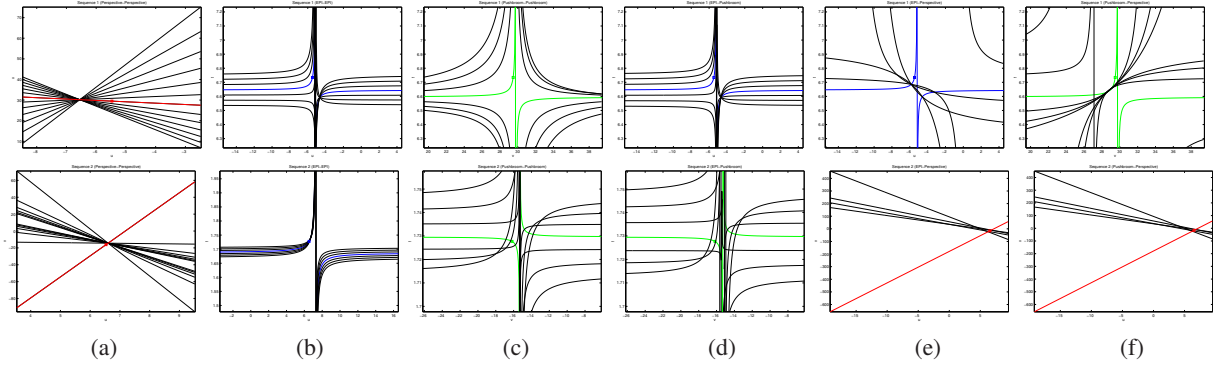


Figure 11. Corresponding points and their epipolar curves. (a) Between two perspective images (b) Between two LP images (c) Between two EPIs (d) Between a LP image and an EPI (e) Between a LP image and a perspective (f) Between an EPI and a perspective.

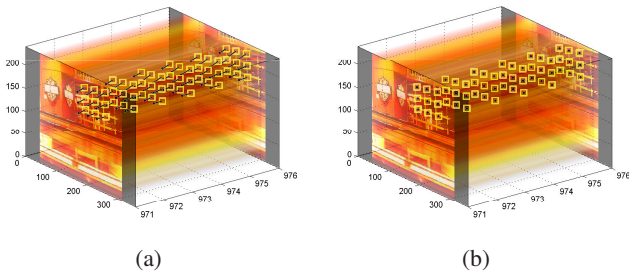


Figure 6. Videopoints mapped using the planar Galilean mapping. Yellow squares indicate the position of the points in Sequence 1, black points indicate the position of (a) corresponding points in Sequence 2 and (b) corresponding points after warping.

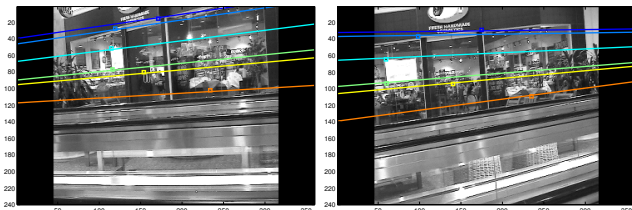


Figure 7. Recovering the fundamental matrix between frame 1 in sequence 1 and frame 20 in sequence 2 from the Galilean fundamental matrix. The Galilean fundamental matrix was computed from videopoints in six frames (971 to 976 in both sequences).

position. When cameras move, the degree of overlap between their fields of view usually changes and when the fields of view become disjoint, estimation of relative camera position becomes impossible. However, if the motion of the cameras follow some structured motion (like constant velocity) the ideas presented here can be used to predict the fundamental matrix relating views even when their fields of view are disjoint. We investigate the relative geometry relating a pair of such cameras in planar and general scenes. We show how three known fundamental matrices are specializations of this matrix and could be readily recovered from the proposed fundamental matrix, providing a unifying

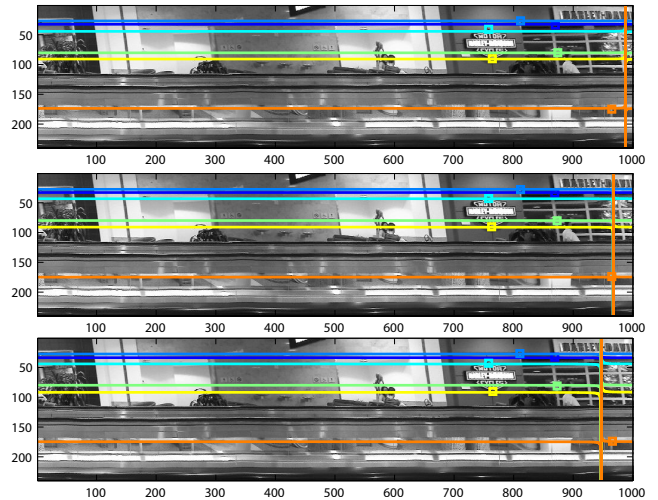


Figure 8. The epipolar curves induced by frames 960, 980 and 1000 of camera 2 on the pushbroom images of camera 1 corresponding to column 250.

link between the classic fundamental matrix and the LP fundamental matrix. In addition we describe three new fundamental matrices that can also be recovered. In the future, we intend to investigate the application of different motion models, such as a constant acceleration model, and study the relationships between three or more Galilean cameras.

In future work, we intend to investigate the use of this model for the analysis of dynamic scenes. Since nonrigid objects can be treated as rigid objects in spacetime the ideas presented in this paper can be used to recover dynamic or nonrigid structure of scenes.

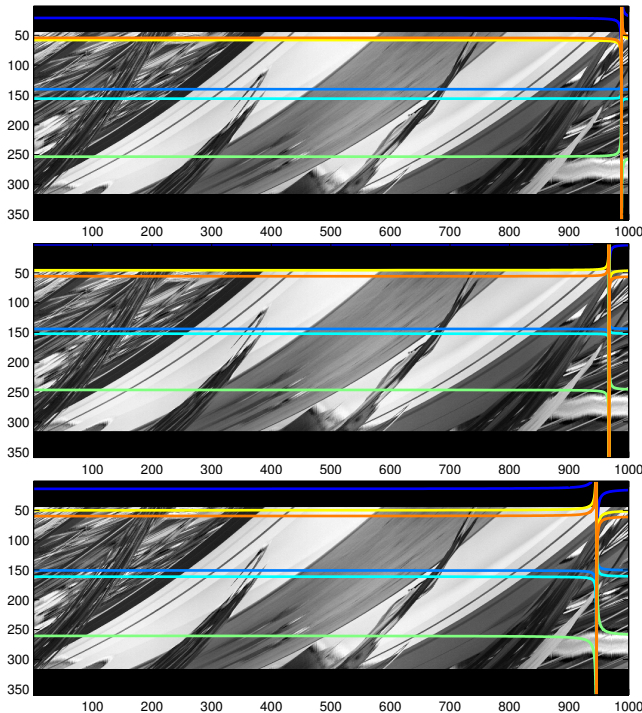


Figure 9. The epipolar curves induced by frames 960, 980 and 1000 of camera 2 on the EPIs of camera 1 corresponding to row 100.

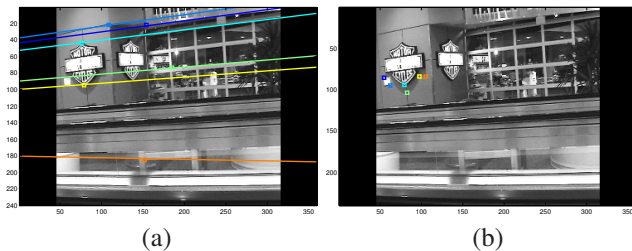


Figure 10. Frame 980 with the points inducing the epipolar curves in (a) Figure 8 and (b) Figure 7.

Acknowledgements

The authors thank Takeo Kanade for his useful comments and suggestions. This research was funded by the U.S. Government VACE program.

References

- [1] A. Bartoli, “The Geometry of Dynamic Scenes - On Coplanar and Convergent Linear Motions Embedded in 3D Static Scenes”, *Computer Vision and Image Understanding*, 2003.
- [2] R. Bolles, H. Baker and D. Marimont, “Epipolar-plane Image Analysis: An Approach to Determining Structure from Motion”, *International Journal of Computer Vision*, 1987.
- [3] J. Callahan, “The Geometry of Spacetime”, *Springer-Verlag*, 2000.

- [4] O. Faugeras, “What can be seen in three dimensions with an uncalibrated stereo rig?”, *European Conference on Computer Vision*, 1992.
- [5] O. Faugeras and Q.-T. Luong, “The Geometry of Multiple Images”, *MIT Press*, 2001.
- [6] D. Feldman, T. Pajdla and D. Weinshall, “On the Epipolar Geometry of the Crossed-Slits Projection”, *IEEE International Conference on Computer Vision*, 2003.
- [7] R. Gupta and R. Hartley, “Linear Pushbroom Cameras”, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1997.
- [8] M. Han and T. Kanade, “Reconstruction of a scene with multiple linearly moving objects”, *IEEE International Conference on Computer Vision and Pattern Recognition*, 2000.
- [9] R. Hartley, “Estimation of Relative Camera Positions for Uncalibrated Cameras”, *European Conference on Computer Vision*, 1992.
- [10] R. Hartley and A. Zisserman, “Multiple View Geometry in Computer Vision”, *Cambridge University Press*, 2000.
- [11] S. Khan, F. Rafi and M. Shah, “Where was the Picture Taken: Image Localization in Route Panoramas using Epipolar Geometry”, *International Conference on Multimedia and Expo*, 2006.
- [12] H. Longuet-Higgins, “A Computer Algorithm for Reconstructing a Scene From Two Projections,” *Nature*, 1981.
- [13] T. Pajdla, ‘Stereo with Oblique Cameras’, *International Journal of Computer Vision*, 2002.
- [14] S. Seitz, “The Space of All Stereo Images”, *Proceedings of the IEEE International Conference on Computer Vision*, 2001.
- [15] P. Sturm, “Structure and Motion for Dynamic Scenes - the case of points moving in planes”, *European Conference on Computer Vision*, 2002.
- [16] L. Wolf and A. Shashua, “On Projection Matrices $\mathcal{P}^k \rightarrow \mathcal{P}^2$, $k = 3, \dots, 6$, and their Application in Computer Vision”, *International Journal of Computer Vision*, 2002.
- [17] Z. Zhang, “Determining the Epipolar Geometry and its Uncertainty: A Review”, *International Journal of Computer Vision*, 1998.