Linear Pushbroom Cameras *

Richard I. Hartley and Rajiv Gupta

G.E. CRD, Schenectady,
NY, 12301.
Ph : (518)-387-7333
Fax : (518)-387-6845
email : hartley@crd.ge.com

Abstract

Modelling the push broom sensors commonly used in satellite imagery is quite difficult and computationally intensive due to the complicated motion of the orbiting satellite with respect to the rotating earth. In addition, the mathematical model is quite complex, involving orbital dynamics, and hence is difficult to analyze. In this paper, a simplified model of a pushbroom sensor (the *linear pushbroom* model) is introduced. It has the advantage of computational simplicity while at the same time giving very accurate results compared with the full orbiting pushbroom model.

Methods are given for solving the major standard photogrammetric problems for the linear pushbroom sensor. Simple non-iterative solutions are given for the following problems : computation of the model parameters from ground-control points; determination of relative model parameters from image correspondences between two images; scene reconstruction given image correspondences and ground-control points.

In addition, the linear pushbroom model leads to theoretical insights that will be approximately valid for the full model as well. The epipolar geometry of linear pushbroom cameras in investigated and shown to be totally different from that of a perspective camera. Nevertheless, a matrix analogous to the essential matrix of perspective cameras is shown to exist for linear pushbroom sensors. From this it is shown that a scene is determined up to an affine transformation from two views with linear pushbroom cameras.

Keywords : pushbroom sensor, satellite image, essential matrix photogrammetry, camera model

*The research described in this paper has been supported by DARPA Contract #MDA972-91-C-0053

1 Real Pushbroom Sensors

Pushbroom sensors are commonly used in satellite cameras, notably the SPOT satellite for the generation of 2-D images of the earth's surface. Fig 1 shows the idea behind a pushbroom sensor. In general terms, a pushbroom camera consists of an optical system projecting an image onto a linear array of sensors, typically a CCD array. At any time only those points are imaged that lie in the plane defined by the optical center and the line containing the sensor array. This plane will be called the instantaneous view plane or simply *view plane*. The pushbroom sensor is mounted on a moving platform, usually a satellite and as the platform moves, the view plane sweeps out a region of space. At regular intervals of time 1-dimensional images of the view plane are captured. The ensemble of these 1-dimensional images constitutes a 2-dimensional image. For a SPOT satellite, the linear array of sensors consists of 6000 pixel array of electronic sensors covering an angle of 4.2 degrees. In 9 seconds a total of 6000 line images are captured. Hence a 6000×6000 pixel image is captured in 9 seconds. Such an image covers a square with side approximately 60 Km on the ground.

The task of modelling a SPOT satellite image exactly is somewhat complex and several factors must be taken into account.

- By Kepler's Laws, the satellite is moving in an elliptical orbit with the center of the earth at one of the foci of the ellipse. The speed is not constant, but varies according to the position of the satellite in its orbit.
- The earth is rotating with respect to the plane of the satellite orbit, so the motion of the satellite with respect to the earth's surface is quite complex.
- The satellite is slowly rotating so that it is approximately fixed with respect to an orthogonal coordinate frame defined as follows : the z-axis points straight down; the x-axis lies in the plane defined by the satellite velocity vector and the z axis; the y-axis is perpendicular to the x and z axes. This coordinate frame will be called the *local orbital frame*. During one orbit, the local orbital frame undergoes a complete revolution about its y axis.
- The orientation of the satellite undergoes slight variations with respect to the local orbital frame.
- The orientation of the view plane with respect to the satellite may not be known.

Some of the parameters of the satellite motion depend on fixed physical and astronomical constants (for example, gravitational constant, mass of the earth, rotational period of the earth). Other parameters such as the major and minor axes and orientation of the satellite orbit are provided as *ephemeris* data with most images. In addition, the fluctuations of the satellite orientation with respect to the local orbital frame are provided as is also the orientation of the view plane. Nevertheless, it has proven necessary for the sake of greater accuracy to refine the ephemeris data by the use of ground-control points.

Even if the orbit of the satellite is known exactly, the task of finding the image coordinates of a point in space is relatively complex. There is no closed-form expression determining the time when the orbiting satellite will pass through a given point in its orbit (time to perigee) – it is necessary to use an approximation. Furthermore the task of determining at what time instant a given ground point will be imaged must be solved by an iterative procedure, such as Newton's method. This means that exact computation of the image produced by a pushbroom sensor is time consuming.

In this paper, a linear approximation to the pushbroom model is introduced. This new model very greatly simplifies the computations involved in working with pushbroom images. The *linear pushbroom* model is defined and discussed in section 2. In subsequent sections, many of the standard photogrammetric problems associated with parameter determination are solved for the linear pushbroom model. All the algorithms discussed are non-iterative, relatively simple, very fast, and do not rely on any extraneous information. This contrasts with parameter determination for the full pushbroom model, which is slow and requires knowledge of orbital and ephemeris parameters. In a final section, the accuracy of the linear pushbroom model is discussed, and the results of some of the algorithms described here are given. It turns out that for SPOT images of size 6000×6000 pixels covering 4.2 degrees, the linear and full models agree within less than half a pixel. This corresponds to a difference of about 6×10^{-6} radians, or about 5 metres on the ground.

Apart from allowing computational efficiency, the linear pushbroom model provides a basis for the mathematical analysis of pushbroom images. The full pushbroom model is somewhat intractable as far as analysis is concerned. On the other hand, the agreement between the full pushbroom model and the linear pushbroom model is so close that results of analyzing the linear pushbroom model will be closely applicable to the full model as well. As an example of the theoretical and practical gains achieved by studying the linear pushbroom model is Theorem 5.3 of this paper, which shows that two linear pushbroom views of a generic scene determine the scene up to an affine transformation. This has the practical consequence that affine invariants of a scene may be computed from two pushbroom views. A similar result applies to perspective views[2], as was shown in [1, 6]. It is hoped that the linear pushbroom model may provide the basis for the development of further image understanding algorithms in the same way that the pinhole camera model has given rise to a wealth of theory and algorithms.

2 Linear Pushbroom Sensors

In order to simplify the pushbroom camera model to facilitate computation and to provide a basis for theoretical investigation of the pushbroom model, certain simplifying assumptions can be made, as follows.

- The platform is moving in a straight line at constant velocity with respect

[2] In the case of perspective (pinhole) images, the scene is determined up to projectivity from two views

to the world.

- The orientation of the satellite, and hence the view plane, is constant.

The first assumption is that during the time of acquisition of one image the variations in velocity of the satellite in its orbit are negligible. In addition, the motion of the earth's surface can be included in the motion of satellite, the composite motion being approximately rectilinear. The second assumption is that the rotation of the local orbital frame as well as the fluctuations of orientation with respect to this frame can be ignored. To what extent these assumptions are justified will be explored in a section 8.

We now describe our model of a pushbroom camera in mathematical terms. This simple pushbroom model will be called a *linear pushbroom camera*. The camera is modelled as a pin-hole camera moving along a linear trajectory in space with constant velocity and fixed orientation. Furthermore, the camera is constrained so that at any moment in time it images only points lying in one plane, called the view plane, passing through the center of the camera. Thus, at any moment of time, a 2-dimensional projection of the view plane onto an image line takes place. The orientation of the view plane is fixed, and it is assumed that the motion of the camera does not lie in the view plane. Consequently, the view plane sweeps out the whole of space as time varies between $-\infty$ and ∞. The image of an arbitrary point $\mathbf{x}$ in space is described by two coordinates. The first coordinate u represents the time when the point $\mathbf{x}$ is imaged (that is, lies in the view plane) and the second coordinate v represents the projection of the point on the image line.

We consider an orthogonal coordinate frame attached to the moving camera as follows. The origin of the coordinate system is the center of projection. The y axis lies in the view-plane parallel with the focal plane (in this case, the linear sensor array). The z axis lies in the view plane perpendicular to the y axis and directed so that the visible points have positive z coordinate. The x coordinate is perpendicular to the view plane such that x, y, and z axes form a right-handed coordinate frame. The ambiguity of orientation of the y axis in the above description can be resolved by requiring that the motion of the camera has a positive x component.

First of all, we consider two dimensional projection. If the coordinates of a point are $(0, y, z)$ with respect to the camera frame, then the coordinate of this point in the 1-dimensional projection will be $v = fy/z + p_v$ where f is the focal length (or magnification) of the camera and p_v is the principal point offset in the v direction. This equation may be written in the form

$$\begin{pmatrix} wv \\ w \end{pmatrix} = \begin{pmatrix} f & p_v \\ 0 & 1 \end{pmatrix} \begin{pmatrix} y \\ z \end{pmatrix} \tag{1}$$

where w is a scale factor (actually equal to z).

Now for convenience, instead of considering a stationary world and a moving camera, it will be assumed that the camera is fixed and that the world is moving. A point in space will be represented as $\mathbf{x}(t) = (x(t), y(t), z(t))^\top$ where $\mathbf{t}$ denotes time. Let the velocity vector of the points with respect to the camera frame be $-\mathbf{V} = -(V_x, V_y, V_z)^\top$. The minus sign is chosen so that the velocity of the camera with respect to the world is $\mathbf{V}$. Suppose that a moving point in space crosses the view plane at time $t_{\rm im}$ at position $(0, y_{\rm im}, z_{\rm im})^\top$. In the 2-dimensional pushbroom image, this point will be imaged at location (u, v) where $u = t_{\rm im}$ and v may be expressed using (1). This may be expressed in an equation

$$\begin{pmatrix} u \\ wv \\ w \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & f & p_v \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} t_{\rm im} \\ y_{\rm im} \\ z_{\rm im} \end{pmatrix} \tag{2}$$

Since all points are moving with the same velocity, we may write

$$\mathbf{x}(t) = \mathbf{x}_0 - t\mathbf{V} = (x_0, y_0, z_0)^\top - t(V_x, V_y, V_z) \ . \tag{3}$$

Since the view plane is the plane $x = 0$, the time t when the point $\mathbf{x}$ crosses the view plane is given by $t = x_0/V_x$. At that moment, the point will be at position

$$(0, y_{\rm im}, z_{\rm im})^\top = (0, y_0 - x_0 V_y/V_x, z_0 - x_0 V_z/V_x)^\top \ .$$

We may write this as

$$\begin{pmatrix} t_{\rm im} \\ y_{\rm im} \\ z_{\rm im} \end{pmatrix} = \begin{pmatrix} 1/V_x & 0 & 0 \\ -V_y/V_x & 1 & 0 \\ -V_z/V_x & 0 & 1 \end{pmatrix} \begin{pmatrix} x_0 \\ y_0 \\ z_0 \end{pmatrix} \tag{4}$$

Combining this with (2) gives the equation

$$\begin{pmatrix} u \\ wv \\ w \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & f & p_v \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1/V_x & 0 & 0 \\ -V_y/V_x & 1 & 0 \\ -V_z/V_x & 0 & 1 \end{pmatrix} \begin{pmatrix} x_0 \\ y_0 \\ z_0 \end{pmatrix} \tag{5}$$

Here, $(x_0, y_0, z_0)^\top$ are the coordinates of the point $\mathbf{x}$ in terms of the camera frame at time $t = 0$. Normally, however, the coordinates of a point are known not in terms of the camera-based coordinate system, but rather in terms of some fixed external orthogonal coordinate system. In particular, let the coordinates of the point in such a coordinate system be $(x, y, z)^\top$. Since both coordinate frames are orthogonal, the coordinates are related via a transformation

$$\begin{aligned} (x_0, y_0, z_0)^\top &= R\left((x, y, z)^\top - (t_x, t_y, t_z)^\top\right) \\ &= (R \mid -R\mathbf{t})(x, y, z, 1)^\top \end{aligned} \tag{6}$$

where $\mathbf{t} = (t_x, t_y, t_z)^\top$ is the location of the camera at time $t = 0$ in the external coordinate frame, and R is a rotation matrix.

Finally, putting this together with (5) leads to

$$\begin{aligned} \begin{pmatrix} u \\ wv \\ w \end{pmatrix} &= \begin{pmatrix} 1 & 0 & 0 \\ 0 & f & p_v \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1/V_x & 0 & 0 \\ -V_y/V_x & 1 & 0 \\ -V_z/V_x & 0 & 1 \end{pmatrix} (R \mid -R\mathbf{t}) \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} \\ &= M(x, y, z, 1)^\top \end{aligned} \tag{7}$$

Equation (7) should be compared with the basic equation describing pinhole, or perspective cameras, namely $(wu, wv, w)^\top = M(x, y, z, 1)^\top$ where $(x, y, z)^\top$ are the coordinates of a world point, $(u, v)^\top$ are the coordinates of the corresponding image point and w is a scale factor. It may be seen that a linear pushbroom image may be thought of as a projective image in one direction (the v direction) and an orthographic image in the other direction (the u direction).

3 Parameter Retrieval

It may be seen that the last two rows of matrix M may be multiplied by a constant without affecting the relationship between world point coordinates (x, y, z) and image coordinates (u, v) expressed by (7). This means that the 3×4 matrix M contains only 11 degrees of freedom. On the other hand, it may be verified that the formation of a linear pushbroom image is also described by 11 parameters, namely the position (3) and orientation (3) of the camera at time t=0, the velocity of the camera (3) and the focal length and v-offset (2). It will next be shown how the linear pushbroom parameters may be computed given the matrix M. This comes down to finding a factorization of M of the kind given in (7). The corresponding problem for pinhole cameras has been solved by Ganapathy ([3]) and Strat ([14]).

First of all we determine the position of the camera at time $t = 0$, referred to subsequently as the initial position of the camera. Multiplying out the product (7) it may be seen that M is of the form $(K \mid -K\mathbf{t})$ for a non-singular 3×3 matrix K. Therefore, it is easy to solve for $\mathbf{t}$ by solving the linear equations $K\mathbf{t} = -\mathbf{c}_4$ where $\mathbf{c}_4$ is the last column of M, and K is the left-hand 3×3 block.

Next, we consider the matrix K. According to (7), and bearing in mind that the two bottom rows of K may be multiplied by a constant factor k, matrix K is of the form

$$K = \begin{pmatrix} 1/V_x & 0 & 0 \\ -k(fV_y/V_x + p_v V_z/V_x) & kf & kp_v \\ -kV_z/V_x & 0 & k \end{pmatrix} R \ . \tag{8}$$

where R is a rotation matrix. In order to find this factorization, we may multiply K on the right by a sequence of rotation matrices to reduce it to the form of the left hand factor in (8). The necessary rotations will be successive Givens rotations about the z, y and x axes with angles chosen to eliminate the (1,2), (1,3) and (3,2) entries of K. In this way, we find a factorization of K as a product $K = LR$ where R is a rotation matrix and L is a matrix having zeros in the required positions. It is not hard to verify that such a factorization is unique. Equating L with the left hand matrix in (8) it is seen that the parameters f, p_v, V_x, V_y and V_z may easily be read from the matrix L. In summary

Proposition 3.1. *The 11 parameters of a linear pushbroom camera are uniquely determined and may be computed from the* 3×4 *camera matrix.*

4 Determination of the Camera Matrix

In this section it will be shown how a linear pushbroom camera matrix may be computed given a set of ground control points. The method is an adaptation of the method of Roberts or Sutherland ([15]) used for the pinhole cameras. In particular, denoting by $\mathbf{m_1}^\top$, $\mathbf{m_2}^\top$ and $\mathbf{m_3}^\top$ the three rows of the matrix M and $\mathbf{x} = (x, y, z, 1)^\top$ a ground control point, (7) may be written in the form of three equations

$$u = \mathbf{m_1}^\top \mathbf{x}$$

$$
\begin{aligned}
wv &= \mathbf{m_2}^\top \mathbf{x} \\
w &= \mathbf{m_3}^\top \mathbf{x} \ .
\end{aligned} \tag{9}
$$

The unknown factor w can be eliminated leading to two equations

$$
\begin{aligned}
u &= \mathbf{m_1}^\top \mathbf{x} \\
v\mathbf{m_3}^\top \mathbf{x} &= \mathbf{m_2}^\top \mathbf{x}
\end{aligned} \tag{10}
$$

Supposing that the world coordinates (x, y, z) and image coordinates (u, v) are known, equations (10) are a set of linear equations in the unknown entries of the matrix M. Given sufficient ground control points we can solve for the matrix M. Note that the entries in the row $m_1{}^\top$ rely only on the u coordinates of the ground control points. Given four ground control points, we can solve for the first row of M. Similarly, the second and third rows of M depend only on the v coordinates of the matrix. Given five ground control points we can solve for the second and third rows of M up to the undetermined factor. With more ground control points, linear least squares solutions methods can be used to determine the best solution.

5 Relative Camera Model Determination

The problem of determining the relative camera placement of two or more pinhole cameras and consequent determination of pinhole cameras has been extensively considered. Most relevant to the present paper is the work of Longuet-Higgins ([8]) who defined the so-called essential matrix Q, which may be determined from eight or more correspondence points between two images by linear techniques. Other non-linear techniques for determining Q, more stable in the presence of noise, have been published ([19, 18, 9, 16]). Those techniques relate especially to so called "calibrated cameras", for which the internal parameters are known. A paper that deals with the determination of the essential matrix for uncalibrated cameras is [2]. As for the determination of the world coordinates of points see from two pinhole cameras, it has been shown ([1, 7]) that for uncalibrated cameras the position of world points is determined up to an unknown projective transform by their images in two separate views. A similar result for linear pushbroom cameras will be shown here, except that the world points are determined up to affine transformation, rather that projective transformation. The proof depends on a definition of a matrix Q for linear pushbroom cameras analogous to the essential matrix for pinhole cameras.

5.1 Definition of the Essential Matrix

Consider a point $\mathbf{x} = (x, y, z)^\top$ in space as viewed by two linear pushbroom cameras with camera matrices M and M'. Let the images of the two points be $\mathbf{u} = (u, v)^\top$ and $\mathbf{u'} = (u', v')^\top$. This gives a pair of equations

$$
\begin{aligned}
(u, wv, w)^\top &= M(x, y, z, 1)^\top \\
(u', w'v', w')^\top &= M'(x, y, z, 1)^\top
\end{aligned} \tag{11}
$$

This pair of equations may be written in a different form as

$$\begin{pmatrix} m_{11} & m_{12} & m_{13} & m_{14}-u & 0 & 0 \\ m_{21} & m_{22} & m_{23} & m_{24} & v & 0 \\ m_{31} & m_{32} & m_{33} & m_{34} & 1 & 0 \\ m'_{11} & m'_{12} & m'_{13} & m'_{14}-u' & 0 & 0 \\ m'_{21} & m'_{22} & m'_{23} & m'_{24} & 0 & v' \\ m'_{31} & m'_{32} & m'_{33} & m'_{34} & 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \\ 1 \\ w \\ w' \end{pmatrix} = 0 \qquad (12)$$

The 6×6 matrix in (12) will be denoted $A(M, M')$. Considered as a set of linear equations in the variables x, y, z, w and w' and constant 1, this is a set of six homogeneous equations in six unknowns (imagining 1 to be an unknown). If this system is to have a solution, then $\det A(M, M') = 0$. This condition gives rise to a cubic equation $p(u, v, u', v') = 0$ where the coefficients of p are determined by the entries of M and M'. The polynomial p will be called the *essential polynomial* corresponding to the two cameras. Because of the particular form of p, there exists a 4×4 matrix Q such that

$$(u', u'v', v', 1)Q(u, uv, v, 1) = 0 \qquad (13)$$

The matrix Q will be called the *essential matrix* corresponding to the linear pushbroom camera pair $\{M, M'\}$. Matrix Q is just a convenient way to display the coefficients of the essential polynomial. Since the entries of Q depend only on the two camera matrices, M and M', equation (13) must be satisfied by any pair of corresponding image points (u, v) and (u', v').

It is seen that if either M or M' is replaced by an equivalent matrix by multiplying the last two rows by a constant c, then the effect is to multiply $\det A(M, M')$, and hence the fundamental polynomial p and matrix Q by the same constant c (not c^2 as may appear at first sight). Consequently, two essential polynomial or matrices that differ by a constant non-zero factor will be considered equivalent. The same basic proof method used above may be used to prove the existence of the essential matrix for pinhole cameras.

A closer examination of the matrix $A(M, M')$ in (12) reveals that $p = \det A(M, M')$ contains no terms in uu', uvu', $uu'v'$ or $uvu'v'$. In other words, the top left hand 2×2 submatrix of Q is zero. Since Q is defined only up to a constant factor, it contains no more than 11 degrees of freedom. Given a set of 11 or more image-to-image correspondences the matrix Q can be determined by the solution of a set of linear equations just as with pinhole cameras.

5.2 Extraction of Relative Cameras from Q

Longuet-Higgins ([8]) showed that for calibrated cameras the relative position and orientation of the two cameras may be deduced from the essential matrix. This result was extended to uncalibrated cameras in [7] where it was shown that if M_1 and M'_1 are one pair of cameras corresponding to an essential matrix Q and if M_2 and M'_2 are another pair corresponding to the same essential matrix, then there is a 4×4 matrix H such that $M_1 = M_2H$ and $M'_1 = M'_2H$. This result will be shown to hold for linear pushbroom cameras with the restriction that H must be a matrix representing an affine transformation, that is, the last row of H is $(0, 0, 0, 1)$.

First of all, it will be shown that M and M' may be multiplied by an arbitrary affine transformation matrix without changing the essential matrix. Let H be a 4×4 affine transformtion matrix and let $\hat{H}$ be the 6×6 matrix

$$\hat{H} = \begin{pmatrix} H & 0 \\ 0 & I \end{pmatrix}$$

where I is the 2×2 identity matrix. If A is the matrix in (12) it may be verified with a little work that $A(M, M')\hat{H} = A(MH, M'H)$, where the assumption that the last row of H is $(0, 0, 0, 1)$ is necessary. Therefore, $\det A(MH, M'H) = \det A(M, M') \det H$ and so the fundamental polynomials corresponding to pairs $\{M, M'\}$ and $\{MH, MH'\}$ differ by a constant factor and so are equivalent.

Next we will consider to what extent the two camera matrices M and M' can be determined from the essential matrix. As has just been demonstrated, they may be multiplied by an arbitrary 4×4 affine matrix H. Therefore, we may choose to set the matrix M' to a particularly simple form $(I \mid 0)$ where I is an identity matrix, by multiplication of both M and M' by the affine matrix $\begin{pmatrix} M'^{-1} & \mathbf{t} \\ 0 & 1 \end{pmatrix}$. It will be seen that with the assumption that $M' = (I \mid 0)$, the other matrix M is almost uniquely determined by the essential matrix.

Under the assumption that $M' = (I \mid 0)$, the essential matrix may be computed explicitly in terms of the entries of M. Using Mathematica([10]) or by hand it may be computed that.

$$Q = (q_{ij}) = \begin{pmatrix} 0 & 0 & m_{11}m_{33} - m_{13}m_{31} & m_{13}m_{21} - m_{11}m_{23} \\ 0 & 0 & m_{11}m_{32} - m_{12}m_{31} & m_{12}m_{21} - m_{11}m_{22} \\ m_{22} & -m_{32} & m_{14}m_{32} - m_{12}m_{34} & m_{12}m_{24} - m_{14}m_{22} \\ m_{23} & -m_{33} & m_{14}m_{33} - m_{13}m_{34} & m_{13}m_{24} - m_{14}m_{23} \end{pmatrix} \quad (14)$$

Given the entries q_{ij} of Q the question is whether it is possible to retrieve the values of the entries m_{ij}. This involves the solution of a set of 12 equations in the 12 unknown values m_{ij}. The four entries m_{22}, m_{23}, m_{32} and m_{33} may be immediately obtained from the bottom left hand block of Q. In particular,

$$\begin{aligned} m_{22} &= q_{31} \\ m_{23} &= q_{41} \\ m_{32} &= -q_{32} \\ m_{33} &= -q_{42} \end{aligned} \quad (15)$$

Retrieval of the remaining entries is more tricky but may be accomplished as follows. The four non-zero entries in the first row can be rewritten in the following form (using (15) to substitute for m_{22}, m_{23}, m_{32} and m_{33}).

$$\begin{pmatrix} -q_{42} & 0 & -m_{13} & -q_{13} \\ -q_{41} & m_{13} & 0 & -q_{14} \\ -q_{32} & 0 & -m_{12} & -q_{23} \\ -q_{31} & m_{12} & 0 & -q_{24} \end{pmatrix} \begin{pmatrix} m_{11} \\ m_{21} \\ m_{31} \\ 1 \end{pmatrix} = 0 \ . \quad (16)$$

Similarly, the bottom right hand 2×2 block gives a set of equations

$$\begin{pmatrix} -q_{42} & 0 & -m_{13} & -q_{43} \\ -q_{41} & m_{13} & 0 & -q_{44} \\ -q_{32} & 0 & -m_{12} & -q_{33} \\ -q_{31} & m_{12} & 0 & -q_{34} \end{pmatrix} \begin{pmatrix} m_{14} \\ m_{24} \\ m_{34} \\ 1 \end{pmatrix} = 0 \ . \quad (17)$$

Immediately it can be seen that if we have a solution m_{ij}, then a new solution may be obtained by multiplying m_{12} and m_{13} by any non-zero constant c and dividing m_{21}, m_{31}, m_{24} and m_{34} by the same constant c. In other words, unless $m_{13} = 0$, which may easily be checked, we may assume that $m_{13} = 1$. From the assumption of a solution to (16) and (17) may be deduced that 4×4 matrices in (16) and (17) must both have zero determinant. With $m_{13} = 1$, each of (16) and (17) gives a quadratic equation in m_{12}. In order for a solution to exist for the sought matrix M, these two quadratics must have a common root. This condition is a necessary condition for a matrix to be an essential matrix. Rearranging the matrices slightly, writing λ instead of m_{12} and expressing the existence of a common root in terms of the resultant leads to the following statement.

Theorem 5.2. *If a matrix 4×4 matrix $Q = (q_{ij})$ is an essential matrix, then*

1. $q_{11} = q_{12} = q_{21} = q_{22} = 0$
2. *the resultant of the polynomials*

$$\det \begin{pmatrix} \lambda & 0 & q_{31} & q_{24} \\ 0 & \lambda & q_{32} & q_{23} \\ 1 & 0 & q_{41} & q_{14} \\ 0 & 1 & q_{42} & q_{13} \end{pmatrix} \tag{18}$$

   and

$$\det \begin{pmatrix} \lambda & 0 & q_{31} & q_{34} \\ 0 & \lambda & q_{32} & q_{33} \\ 1 & 0 & q_{41} & q_{44} \\ 0 & 1 & q_{42} & q_{43} \end{pmatrix} \tag{19}$$

   vanishes.
3. *The discriminants of the polynomials (18) and (19) are both non-negative.*

If the two quadratics have a common root, then this common root will be the value of m_{12}. The linear equations (16) may then be solved for m_{11}, m_{21} and m_{31}. Similarly, equations (17) may be solved for m_{14}, m_{24} and m_{34}. Unless $q_{31}q_{42} - q_{41}q_{32}$ vanishes, the first three columns of the matrices (18) and (19) will be linearly independent and the solutions for the m_{ij} will exist and be unique.

To recapitulate, if m_{12} is a common root of the two quadratic polynomials (18) and (19), m_{13} is chosen to equal 1, and $q_{31}q_{42} - q_{41}q_{32} \neq 0$ then the matrix $M = (m_{ij})$ may be uniquely determined by the solution of a set of linear equations. Relaxing the condition $m_{13} = 1$, leads to a family of solutions of the form

$$\begin{pmatrix} m_{11} & m_{12}c & m_{13}c & m_{14} \\ m_{21}/c & m_{22} & m_{23} & m_{24}/c \\ m_{31}/c & m_{32} & m_{33} & m_{34}/c \end{pmatrix} \tag{20}$$

However, up to multiplication by the diagonal affine matrix $\mathrm{diag}(1, 1/c, 1/c, 1)$ all such matrices are equivalent. Furthermore, the matrix $M' = (I \mid 0)$ is mapped unto an equivalent matrix by multiplication by $\mathrm{diag}(1, 1/c, 1/c, 1)$. This shows that once m_{12} is determined, the matrix pair $\{M, M'\}$ may be computed uniquely up to affine equivalence.

Finally, we consider the possibility that the equations (18) and (19) have two common roots. This can only occur if the coefficients of Q satisfy certain restrictive identities that may be deduced from (18) and (19). This allows us to state

Theorem 5.3. *Given a 4×4 matrix Q satisfying the conditions of Proposition 5.2, the pair of camera matrices $\{M, M'\}$ corresponding to Q is uniquely determined up to affine equivalence, unless Q lies in a lower dimensional critical set.*

5.3 More about the Critical Set

It is not the purpose here to undertake a complete investigation of the critical set. As previously stated, conditions under which there are two common roots to (18) and (19) leading to two distinct solutions for M may be deduced from the form of (18) and (19). This investigation will give a condition in terms of the entries of Q. More enlightening would be a conditions in terms of the entries of the matrix M for the solution to be ambiguous. This will be investigated next.

There will be ambiguous solutions to the problem of estimating the matrix M if the polynomials (18) and (19) have two common roots. Suppose that the matrix Q is of the form given in (14). Then we may compute the two quadratic polynomials from (18) and (19). The results[3] are

$$\begin{aligned} p_1(\lambda) &= (m_{13}\lambda - m_{12})(m_{22}m_{31} - m_{21}m_{32} - \lambda(m_{23}m_{31} - m_{21}m_{33})) \\ p_2(\lambda) &= (m_{13}\lambda - m_{12})(m_{22}m_{34} - m_{24}m_{32} - \lambda(m_{23}m_{34} - m_{24}m_{33})) \end{aligned}$$

As expected, $p_1(\lambda)$ and $p_2(\lambda)$ have a common root $\lambda = m_{12}/m_{13}$. The second root of p_1 and p_2 is the same if and only if two linear polynomials $(m_{22}m_{31} - m_{21}m_{32} - \lambda(m_{23}m_{31} - m_{21}m_{33}))$ and $(m_{22}m_{34} - m_{24}m_{32} - \lambda(m_{23}m_{34} - m_{24}m_{33}))$ have the same root. This is so if and only if

$$(m_{21}m_{34} - m_{24}m_{31})(m_{22}m_{33} - m_{23}m_{32}) = 0 \qquad (21)$$

Since the right hand side of this expression is a product of two factors, there are two separate conditions under which an ambiguous solution exists. The first condition $(m_{21}m_{34} - m_{24}m_{31}) = 0$ corresponds geometrically to the situation where the trajectories of the two cameras meet in space. This may be seen as follows. A point $\mathbf{x} = (x, y, z)^\top$ lies on the trajectory of the centre of projection of a camera with matrix M if and only if $M(x, y, z, 1)^\top = (u, 0, 0)^\top$, for under these circumstances the v coordinate of the image is undefined. In particular, the points that lie on the trajectory of the camera M' with matrix $(I \mid 0)$ are of the form $(x, 0, 0)^\top$. Such a point will also lie on the trajectory of the camera with matrix M if and only if $xm_{21} + m_{24} = xm_{31} + m_{34} = 0$ for some x – that is, if and only if $m_{21}m_{34} - m_{24}m_{31} = 0$.

The geometrical meaning of the other condition has not been determined so far.[4]

[3] These computations were carried out using Mathematica ([10])

[4] The condition seems to be that trajectory of the second camera lies parallel to the view plane of the other camera, but this needs to be checked.

6 Scene Reconstruction

Once two camera matrices have been determined, the position of the points $\mathbf{x}_i$ in space may be determined by solving (12). This will determine the position of the points in space up to an affine transformation of space.

In the case where both point matches between images and ground-control points are given, the scene may be reconstructed by using the matched points to determine the scene up to affine transformation, and then using the ground-control points to determine the absolute placement of the scene. If the ground control points are visible in both images, then it is easy to find the correct affine transformation. This is done by determining the position of the ground control points in the reconstructed image, and then determining the 3-D affine transformation that will take these points on to the absolute ground-control locations.

If ground-control points are available that are visible in one image only, it is still possible to use them to determine the absolute location of the reconstructed point set. A method for doing this is given in [7] and will not be repeated here.

7 Computation of the Essential Matrix

The essential matrix may be computed from image correspondences in much the same way as Longuet-Higgins computes the perspective essential matrix ([8]). Given 11 or more point-to-point correspondences between a pair of linear pushbroom images, equation (13) can be used to solve for the 12 non-zero entries of Q, up to multiplication by an unknown scale. Unfortunately, in the presence of noise, the solution found in this way for Q will not satisfy the second condition of (5.2) exactly. Consequently, when solving for the matrix M, one will find that the two polynomials (18) and (19) do not have a common root. Various strategies are possible at this stage.

One strategy is as follows. Consider each of the two roots m_{12} of (18) and with each such value of m_{12} proceed as follows : Substitute each such m_{12} in turn into the equation (17). giving a set of four equations in three unknowns; solve (17) to find the least-squares solution for m_{14}, m_{24} and m_{34}. Finally accept the root of (18) that leads to the best least-squares solution. One could do this the other way round as well starting by considering the roots of (19) and accepting the best of the four solutions found. A different strategy is to choose m_{12} to be the number that is closest to being a root of each of (18) and (19). This is the algorithm that we have implemented, with good results so far.

To obtain the best results, however, it is probably necessary to take the conditions of Proposition 5.2 into account explicitly and compute an essential matrix satisfying these conditions using explicit assumptions about the source of error to formulate a cost function to be minimized. This has been shown to be the best approach for perspective cameras ([9, 16]).

Let us assume that the error is in the specification of the image coordinates $(u_i, v_i)^\top$ and $(u'_i, v'_i)^\top$ in the two images, and that the errors in coordinates are independent Gaussian variables. For simplicity we assume that the variances of individual pixels are the same, though this assumption is not necessary. In this case, the correct minimization problem is to find a matrix Q and coordinates $(\hat{u}_i, \hat{v}_i)^\top$ and $(\hat{u}'_i, \hat{v}'_i)^\top$ such that Q is of the form specified by Proposition 5.2, the epipolar constraint $(\hat{u}'_i, \hat{u}'_i\hat{v}'_i, \hat{v}'_i, 1)^\top Q(\hat{u}_i, \hat{u}_i\hat{v}_i, \hat{v}_i, 1) = 0$ is satisfied for all i and the difference $\sum_i (u_i - \hat{u}_i)^2 + (v_i - \hat{v}_i)^2 + (u'_i - \hat{u}'_i)^2 + (v'_i - \hat{v}'_i)^2$ is minimized. This problem can be solved using a standard photogrammetric resection approach. Instead of solving for Q, we solve for the camera matrix M and the world point locations $\mathbf{x}_i$. We assume without loss of generality that $M' = (I \mid 0)$. Given an estimate for M and each $\hat{\mathbf{x}}_i$, image coordinates $(\hat{u}_i, \hat{v}_i)^\top$ and $(\hat{u}'_i, \hat{v}'_i)^\top$ are computed from the basic formula (7) and the cost function to be minimized is the squared pixel error. An initial estimate of the camera matrix M may be computed using the straight-forward linear approach given above. By non-linear least squared iteration a final estimate for M and each $\mathbf{x}_i$ is found. The essential matrix Q may be computed from the final estimate of M. Although this method has not been tested on pushbroom cameras, it has proven successful with perspective cameras.

The problem with the above method is that a large non-linear problem must be solved. A comparison of this method and a different, fast and almost optimal method that uses the correlation matrix of Q may form the subject of another paper.

The question of numerical stability is important in implementing algorithms using the linear pushbroom model. In particular, it is easy to encounter situations in which the determination of the linear pushbroom model parameters is very badly conditioned. In particular, if a set of ground-control points lie in a plane or are very close to being planar, then it is easily seen (just as with perspective cameras) that the determination of the model parameters is ambiguous. We have developed techniques (not described here) for handling some cases of instability, but care is still necessary. The algorithms described in this paper can not be used in cases where the object set lies in a plane.

8 Experimental Results

Several experiments were conducted to measure the accuracy of the linear pushbroom model.

In the first experiment, the accuracy of the linear pushbroom model was compared with the full model. In order to make this comparison, a full SPOT model was used. The details of this model are given in [4]. The full model takes account of orbital dynamics, provided ephemeris data and attitude drift data to model the imaging process as accurately as possible. A different model is discussed in [17].

Using the full pushbroom model parametrized to an actual orbit and ephemeris data, and an artificial terrain model, a set of ground to image correspondences were computed, one such ground control point being computed every 120 pixels. This gave a 51×51 grid of ground-control points covering approximately 6000×6000 pixels. Next, these ground control points were used to instantiate the linear pushbroom model using the algorithm of section 4. In this experiment, the locations of ground points were fixed for both the full and linear pushbroom models. The difference was measured between the corresponding image points as computed by each of the models. The absolute value of error as it varies across the image is shown in Fig 2. The maximum error was less than 0.4 pixels with an RMS error of 0.16 pixels. As can be seen, for a complete SPOT image, the error incurred by using the linear pushbroom model is less than half a pixel, and much less over most of the image.

To test whether a perspective camera model could do as well, the same set of ground control points were modelled using a perspective camera model. The result was an RMS error of 16.8 pixels with a maximum pixel error of over 45 pixels. Fig 3 shows the error distribution across the image.

Next, the linear pushbroom model was compared with the full model on a pair of real images with matched points computed using a stereo matching algorithm. A stereo pair of SPOT images of the Malibu region, centered approximately at 34 deg 5 min north, and 118 deg 32 min west (images with (J, K) = (541, 281) and (541, 281) in SPOT's grid reference system [13]) were used. We estimated the camera models for these two images using a set of 25 ground control points, visible in both images, picked form USGS maps and several automatically generated image to image correspondences found using STEREOSYS ([5])

Two performance metrics were computed. The accuracy with which the camera model maps the ground points to their corresponding image points is important. The RMS difference between the known image coordinates and the image coordinates computed using the derived camera models was measured. An application-specific metric, viz. the accuracy of the terrain elevation model generated from a stereo pair, was also measured.

Once again, the data was modelled using a perspective camera model, a linear pushbroom model and a full pushbroom model.

In order to make the results directly comparable, the same ground control points and image to image correspondences were used for camera model computations in all three experiments. (The number of tie or match points in computation of the pin-hole camera is an exception where 511 tie-points, instead of 100, were provided in an attempt to boost its accuracy.) In addition, the terrain model was also generated using the same set of match points.

The results of these three experiments are tabulated in Table 1. The first and the second row list the number of ground control points and the number of ties points used in the camera model computation. The third row gives the number of match points for which a point on the terrain was generated. The camera model accuracy, i.e., accuracy with which a ground point $(x, y, z)^\top$ is mapped into its corresponding image point, listed in the fourth row. Finally, the RMS difference between the generated terrain and the ground truth (DMA DTED data) is given in the fifth row.

The attempt to model SPOT's HRV cameras by perspective cameras yielded camera models with a combined accuracy of about 11 pixels. This is a large error because for a high platform such as a satellite, even a single pixel error can translate into a discrepancy of tens of meter along the horizontal and vertical dimensions (the exact amount depends on the pixel resolution and the look angles). This is reflected in the accuracy of the generated terrain which is as much as 380 meters off, on the average. Thus, as expected, a pin-hole camera

	Pin-hole Model	Linear Push-broom Model	Full SPOT Model
Num. gc pts	25	25	25
Num. match pts	511	100	100
Num. terrain points	68,131	68,131	68,131
RMS error	11.13 pixels	0.80 pixels	0.73 pixels
Terrain accuracy	380.79m	35.67m	11.10m
Time	∼5 sec.	∼5 sec.	> 20 min.

Table 1: A comparison of the three camera models.

is a poor approximation for pushbroom camera. The linear push, on the other hand, is quite competitive with the detailed model, both in terms of camera model accuracy, as well as the accuracy of the generated terrain.

The last entry on the fifth row (the 11.10m accuracy for the terrain generated by the complex model) is a little misleading since generated terrain is more accurate than the claimed accuracy of the ground-truth it is being compared with. This figure is a statement about the accuracy of the ground-truth, instead of the other way around. Figures 4 and 5 show the terrain generated by the perspective and the full SPOT models, respectively. Fig 5 can be regarded as the ground truth. In all these figures, the Pacific Ocean has been independently set to have an elevation of 0. Also, since the area covered is rather large (about 60km×60km), the terrain relief has been considerably exaggerated compared to the horizontal dimensions.

9 Epipolar Geometry

One of the most striking differences between linear pushbroom and perspective cameras is the epipolar geometry. First of all there are no epipoles in the familiar manner of perspective cameras, since the two pushbroom cameras are moving with respect to each other. Neither is it true that epipolar lines are straight lines. Consider a pair of matched point $(u, v)^\top$ and $(u', v')^\top$ in two images. According to equation (13) these points satisfy $(u', u'v', v', 1)^\top Q(u, uv, v, 1) = 0$. Now, fixing $(u, v)^\top$ and enquiring for the locus of all possible matched points $(u', v')^\top$, and writing $(\alpha, \beta, \gamma, \delta)^\top = Q(u, uv, v, 1)^\top$, we see that $\alpha u' + \beta u'v' + \gamma v' + \delta = 0$. This is the equation of a hyperbola – epipolar loci are hyperbolas for linear pushbroom cameras. The epipolar locus of a point is the projection in the second image of a straight line emanating from the instantaneous centre of projection of the first camera. In general, therefore, lines in space map to hyperbolas in the image. Only one of the two branches of the hyperbola will be visible in the image. The other branch will lie behind the camera.

Fig 6 shows the images of a set of lines in space as taken with a linear pushbroom camera. The curvature of the lines is exaggerated by the wide field of view.

References

[1] Faugeras, O., "What can be seen in three dimensions with an uncalibrated stereo rig?", Proc. of ECCV-92, G. Sandini Ed., LNCS-Series Vol. 588, Springer- Verlag, 1992, pp. 563 – 578.

[2] O. D. Faugeras, Q.-T Luong and S. J. Maybank, *"Camera Self-Calibration: Theory and Experiments,"* Proc. of ECCV-92, G. Sandini Ed., LNCS-Series Vol. 588, Springer- Verlag, 1992, pp. 321 – 334.

[3] S. Ganapathy, *" Decomposition of Transformation matrices for Robot Vision,"* IEEE Robotics Conference, Atla Ga. 1984., also in *"Pattern Recognition Letters,* vol. 2, pp. 410-412, Dec 1989.

[4] R. Gupta, R. Hartley, *"A camera model for space-borne, pushbroom imaging systems,"*, Technical Report, GE Corporate R&D, 1 River Road, Schenectady, NY 12301, 1992.

[5] M.J. Hannah, *"Bootstrap stereo,"* Proc. Image Understanding Workshop, College Park, MD, April 1980, pp. 201–208.

[6] R. Hartley, "Invariants of 3D point sets", submitted for publication.

[7] R. Hartley, R. Gupta and T. Chang, "Stereo from Uncalibrated Cameras", Proceedings Computer Vision and Pattern Recognition conference (CVPR-92), 1992.

[8] H.C. Longuet-Higgins, *"A computer algorithm for reconstructing a scene from two projections,"* Nature, Vol. 293, 10 Sept. 1981.

[9] B.K.P. Horn, *"Relative Orientation Revisited"*, Journal of the Optical Society of America, A Vol 8, number 10, pp. 1630-1638, Oct 91.

[10] Wolfram, S., Mathematica, "A System for Doing Mathematics by Computer," Addison-Wesley, Redwood City, California, 1988.

[11] Charles A. Rothwell, Andrew Zisserman, David A. Forsyth and Joseph L. Mundy, "Canonical Frames for Planar Object Recognition", Proc. of ECCV-92, G. Sandini Ed., LNCS-Series Vol. 588, Springer- Verlag, 1992, pp. 757-772.

[12] Gunnar Sparr, "Depth Computations from Polyhedral Images", Proc. of ECCV-92, G. Sandini Ed., LNCS-Series Vol. 588, Springer- Verlag, 1992, pp. 378 – 386.

[13] *"SPOT User's Handbook,"* Vol 1, 2 and 3, SPOT Image Corporation.

[14] T.M. Strat, *" Recovering the camera parameters from a transformation matrix,"* Proc. of DARPA Image Understanding Workshop, New Orleans, LA, pp. 264–271, 1984. reprinted in *Readings in Computer Vision,* pp 93 – 100, ed. M. Fishler and O. Firschein, Morgan Kaufmann Publishers, 1987.

[15] I.E. Sutherland, *"Three dimensional data input by tablet,"* Proceedings of IEEE, Vol. 62, No. 4, pp. 453–461, April 1974.

[16] Minas E. Spetsakis and Yiannis Aloimonos, *"Optimal Visual Motion Estimation,"* IEEE Trans. Patt. Anal. Machine Intell., Vol PAMI-14, No. 9, September 1992, pp 959 – 964.

[17] Ashley P. Tam, *"Terrain elevation extraction from digital SPOT satellite imagery,"* Masters Thesis, Dept. of Surveying Engineering, Calgary, Alberta, July 1990.

[18] R. Y. Tsai and T. S. Huang, *"Uniqueness and estimation of three dimensional motion parameters of rigid objects with curved surfaces"*, IEEE Trans. Patt. Anal. Machine Intell., vol. PAMI-6, pp. 13-27, 1984.

[19] J. Weng, T. S. Huang and N. Ahuja, *" Motion and Structure from two perspective views: Algorithms, error analysis and error estimation, "* IEEE Trans. on PAMI, vol 11, no. 5, pp. 451–476, May 1989.

Figure 1. Pushbroom satellite

Figure 2. Error profile for linear pushbroom model

Figure 3. Error profile for perspective model

Figure 4. Terrain reconstructed from perspective model

Figure 5. Terrain reconstructed from full model

Figure 6. Epipolar lines