A Computer Algorithm for Reconstructing a Scene from Two Satellite Images

Rajiv Gupta and Richard Hartley

GE Corporate R&D
1 River Road, P.O. Box 8
Schenectady, NY 12301

Abstract

A new camera model for transforming object space coordinates into image coordinates as affected by a push-broom sensor traveling in a straight line is described. It is shown that this transformation can be encoded in a 3×4 matrix M representing a non-linear Cremona transformation of object space into image space. Significantly, unlike other known models for satellite cameras, M can be estimated without recourse to iterative techniques. Experimental results demonstrate that the proposed model is quite accurate even for push-broom cameras in low-earth orbits (e.g. SPOT and LANDSAT). A 4×4 matrix Q, christened *hyperbolic essential matrix*, that contains the relative camera models in a stereo setting, is derived. The matrix Q is similar in spirit to that derived by Longuet-Higgins [**?**]. However, unlike the latter, it represents the non-linear transformation of a point $[u_1, v_1]$ in the first image to its corresponding hyperbolic epipolar curve in the second image. A linear method for computing Q from image correspondences is described. It is shown that from two images, the camera models and the corresponding 3-D points can be determined only up to an affine transformatin of 3-D object space and this is the best that can be done in the absence of any ground-truth. If 6 or more ground control points, visible in both images, are provided the stereo problem can be solved using only linear techniques.

1 Introduction

A typical system for the construction of 3-D models from stereo, push-broom imagery operates in two phases. In the first phase a set of *match points* (i.e., pixels in the two views that are the images of the same point in the real world, also referred to as tie points), are established between the two images. In the second phase, the computed match points along with ephemeris, other auxiliary information collected by the satellite, and ground control points are used to derive the orbital parameters, the orientation of the cameras, and other related parameters. With the information about the cameras known, one can analyze the disparity arising because of different elevation of various points and assign them a 3-D coordinate. This paper focuses on both of these phases the context of push-broom imagery.

In order to solve the correspondence problem to derive a set of tie-points, area-based correlation is used. Since the stereo image pair may be locally distorted with respect to each other, a 2-D projective image to image transformation, designated by M_I, is used. A key contribution of this paper is to show that the image to image transformation M_I and the hyperbolic essential matrix Q can be computed in conjunction thus making M_I respect the epipolar constraint. In other words, not only is the transformed point $M_I\tilde{u}$ close to its match point in the second image, it also lies on the epipolar curve predicted by Q.

The hyperbolic essential matrix — it predicts, for a point in the first image, the epipolar curve in the second image — can be used for accelerating the matching process. It should be emphasized that the epipolar constraint is enforced without computing the relative camera trajectories, orientation, or any other camera parameters such as scale or principal point offset in the camera model. In fact, it is shown that the transformation Q contains all the information about relative camera parameters for *completely uncalibrated linear push-broom cameras* (i.e., cameras about which nothing is known) that can be derived from a set of match points.

It shown that there exist an infinite number of camera matrices P_1 and P_2 for the two cameras that are compatible with a given Q. Thus from Q alone, it is impossible to derive the 3-D model for the scene (even upto a scale factor). Interestingly, all the admissible solutions for P_1 (and also for P_2) differ from each other by a post-multiplication by 4×4 affine transformation matrix H. In addition, the 3-D coordinates computed by one admissible solution, differ from another by an affine mapping of 3-space to 3-space affected by the same H. Thus, Q determines the 3-D points and the camera matrices up to a collineation.

The above result is somewhat surprising as the level of indeterminacy in the camera matrices is the same as that in the case of pin-hole cameras. Given the linear push-broom camera matrices actually represent a non-linear transformation, one would expect that the camera matrices and the 3-D points would be unknown by more than a collineation of space (i.e., a non-linear transformation of 3-space on 3-space). However, that is not the case.

Figure 1: Ortho-perspective projection under linear push-broom camera.

2 Linear Push-broom Camera Model

In this section we show that if one makes the assumption that the push-broom sensor is being carried in a straight line during image acquisition stage then the camera transformation can be modeled by a 3×4 matrix P. P, referred to as the *camera matrix*, non-linearly transforms points in 3-dimensional projective space to points in 2-dimensional projective space according to the equation

$$\tilde{u} = P\tilde{x} \tag{1}$$

where $\tilde{u} = (u, wv, w)^T$ and $\tilde{x} = (x, y, z, 1)^T$. We now show how the camera matrix models, in a single formalism, the effect of translation, rotation, semi-perspectivity of the imaging geometry, non-othogonality of the velocity vector and the look plane, scaling, and cropping.

Ortho-Perspective Projection. Consider a linear array of sensors, aligned with the y-axis of the object space, traveling along x-axis, as shown in Fig. 1. Let f be the focal length of the sensor array. At $t = 1$, the sensor array is located at $(0, 0, 0)$. In every one time unit, it travels by a unit distance and acquires one row of the image. Let the u and v axes of the image be aligned with the x and y axes, the look plane be parallel to the y-z plane, and the satellite velocity vector be along the x-axis, at all times.

It is clear from the figure that under such imaging conditions, an object space point $[x, y, z, 1]$ is projected orthographically on the u along the x direction, and perspectively on v in the y-z plane as follows.

$$u = x \tag{2}$$

$$v = \frac{fy}{z}. \tag{3}$$

Thus the projection matrix, $P_{o/p}$ is given by

$$\begin{bmatrix} u \\ tv \\ t \\ 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} = 0. \tag{4}$$

Translation. The effect of translation of object space coordinates by $[x_0, y_0, z_0]$ is equivalent to multiplying $P_{o/p}$ by the translation matrix given by

$$T = \begin{bmatrix} 1 & 0 & 0 & -x_0 \\ 0 & 1 & 0 & -y_0 \\ 0 & 0 & 1 & -z_0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{5}$$

Rotation. If the object space coordinates are rotated with respect to the that used in Fig. 1, the usual rotation matrix $R = R_x R_y R_z$ can be used to bring them into alignment. The overall camera model, then, is given by $P_{o/p}RT$.

Non-orthogonal look-plane and velocity vector. If the look-plane of the sensor array is not orthogonal to the velocity vector, one can use a non-orthogonal set of axes for the 3-D points. In this non-orthogonal system, x and u, and y and v coincide, and z is parallel to the principal ray of the linear sensor, as in Fig. 1. The object space points can be mapped to this non-orthogonal R^3 by changing the basis. If $[b_{x1}, b_{x2}, b_{x3}]$, $[b_{y1}, b_{y2}, b_{y3}]$ and $[b_{z1}, b_{z2}, b_{z3}]$ are the unit vectors, in the object space coordinate system, representing the x, y and z axes of the non-orthogonal frame of reference, then the following matrix transforms and object space point into the new system.

$$B = \begin{bmatrix} b_{x1} & b_{y1} & b_{z1} & 0 \\ b_{x2} & b_{y2} & b_{z2} & 0 \\ b_{x3} & b_{y3} & b_{z3} & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{6}$$

The overall transformation matrix is given by $P = P_{o/p}BRT$.

Scaling. The image coordinates may be scaled to reflect resampling of the image to make it enlarged or shrunken. The scaling matrix S, which changes the camera matrix to $P = SP_{o/p}BRT$, is given by

$$S = \begin{bmatrix} k_u & 0 & 0 & 0 \\ 0 & k_v & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{7}$$

Variation in principal offset and image cropping. Any variation in the principal offset of the sensor (the point at which the principal ray pierces the linear array), amounts to translating the v coordinates by a fixed quantity. Similarly, image cropping translates all the image coordinates by a fixed 3-D vector, say $[u_0, v_0]$. This transformation C, which can be incorporated in the camera

matrix as $P = CSP_{o/p}BRT$, is given by

$$C = \begin{bmatrix} 0 & 0 & 0 & -u_0 \\ 0 & 0 & 0 & -v_0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{8}$$

Since P represents a 3-D to 2-D transformation, and only u, v, t are meaningful for an image, the third row of P can be dropped resulting in a 3×4 matrix.

In our approach, we avoid the actual computation of orbital or pose parameters of the cameras as far as possible. In fact the geometry of the object-space points is determined without the need for the camera parameters to be computed, though they are easily obtained if needed. This paper provides a non-iterative method for solving camera parameters given matched points and ground control points.

Our approach also make no assumptions about the internal camera parameters. In this situation, as we show in Section XYZ, it is impossible to compute the camera model from a set of match points only. We prove that a set of match points leave an ambiguity represented by a general 3-dimensional affine transform of the cameras and the object-space points. The ambiguity may be resolved by the use of ground control points, or by placing restrictions on the camera model.

2.1 Computation of P from ground control points

Let $\{\tilde{x_i} \rightarrow [u_i, v_i] | i = 1, n\}$ be the set of given ground control points. Also, let P_1, P_3 and P_3 be four dimensional row vectors representing the rows of P. Then, from Eq. 1, we have,

$$P_1 \tilde{x} = u_i, and \tag{9}$$

$$(P_2 - v_i P_3)\tilde{x} = 0. \tag{10}$$

Eqs. 9 have the form $Ax = b$, while Eqs. 10 have the form $Ax = 0$. The four unknowns in P_1 can be solved, using linear techniques, by n equations given by Eq. 9. The entries in P_2 and P_3 are indeterminate upto a multiplicative constant. Also, Eqs. 10 admits the trivial solution. Thus, by putting the additional constraint that $||[P_2|P_3]|| = 1$, where $[P_2|P_3]$ is the vector with 8 unknown entries, one can avoid the trivial solution and compute the seven independent entries of P_2 and P_3.

2.2 Recovering camera parameters from P

The camera matrix P has 11 independent entries. As shown in the previous section, P models several phenomena and its entries are a complex function of the various camera parameters. In particular, the entries in P depend on the following.

1. The linear trajectory of the camera platform (6 degrees of freedom),
2. The orientation of the look plane (3 degrees of freedom),
3. The principal point offset of the linear array of sensors and image cropping (2 degrees of freedom), and
4. unequal scale factors in two orthogonal directions in the image space. Since the unequal stretches in two directions need **not** be aligned with the image axes, then another camera parameter is needed. Thus there are 3 degrees of freedom for scale related parameters.

This this adds to 14 parameters encoded in the 11 entries in the camera matrix. Thus, in principal, from a given camera matrix one can extract 11 parameters, if any three parameters are known a priori. In practical cases, many parameters such as the focal length (magnification) of the camera, principal offset point of the linear array, and scale parameters, which are measured during pre-launch calibration stage, are likely to be known quite accurately. Hence, the other parameters can be derived if needed.

Our purpose in using general camera transforms, which can be computed from ground control points and image correspondences, is to avoid relying on any a priori knowledge of camera parameters.

3 Accuracy of the Linear Push-broom Camera Model

The linear push-broom model presented in the previous section was implemented and tested. Besides the linear push-broom model, a pin-hole camera model for frame imagery [**?**] and complete camera model that simulates the orbital geometry and the imaging characteristics of SPOT's HRV camera [**?**] were developed and implemented. All these models have been incorporated in the STEREOSYS testbed developed by Marsha Jo Hanna of SRI [**?**, **?**] and one can derive terrain elevation from a stereo pair of aerial or satellite images.

Since the standard photogrammetric bundle adjustment typical of aerial imagery cannot separate the correlation among the unknown parameters in satellite imagery, for iterative solution of the non-linear equations involved, additional constraints are required in order to obtain convergence. In SPOT camera model, the constraint equations derived from known orbital relationships governed by Kapler's laws and the auxiliary information collected by the on-board monitoring systems during image acquisition are used.

The pin-hole model and the full SPOT model have been independently validated against both synthetic data and ground truth and are known to give accurate results for frame and SPOT imagery respectively. The details of these two camera models — which will be used here for comparison — are beyond the scope of this paper.

3.1 Modeling a Push-broom Camera as a Pinhole Camera

As a first level-of approximation, one can view satellite imagery as aerial imagery from a very high platform. The following attributes, which are generally not associated with aerial imagery, make this approximation rather crude.

Partially Perspective Image. In push-broom type of imaging, the image rays captured by the detector array are restricted to a plane perpendicular to the flight path. Thus the image is perspective only in cross-flight direction; along the direction of flight, the projection is closer to being orthographic than perspective. A corollary of this fact is that the ray intersection method would not give accurate information about along-track location. In general, classical space resection techniques, by themselves, are not reliable enough for satellite imagery.

Many Perspective Centers. In push-broom imagery, each line is imaged independently. This implies that there are numerous, highly correlated, perspective centers associated with each image. In a SPOT image, for example, there are 6000 perspective centers per image. Hence there are 6×6000 parameters as compared to 6 parameters for frame photography.

Terrain Height to Altitude Ratio. Even for the satellites in low earth orbits, the terrain height to altitude ratio for satellite imagery is about 80 times smaller than that for aerial photography. Put another way, in satellite imagery, the terrain appears to be relatively flat. Thus the parallax arising due to difference in heights and view angles in stereo pairs is not as pronounced. This implies that the parameters have to be known considerably more accurately if meaningful information is to be extracted from satellite imagery.

Field of View Angle. Generally the size of the detector array capturing the image is much smaller than its distance from the perspective center. For example, SPOT's HRV cameras use a CCD array that is only 7.8cm long while the focal length of the camera is 1.082m. This implies that the field-of-view (FOV) angle is generally very small (only about 4.2 degrees for SPOT). This narrow FOV contributes to the instability of the solution.

Parameter Correlation. Scan line imaging suffers from high degree of parameter correlation. For example, any small rotation of the linear array around an axis perpendicular to flight path (i.e., along the length of the array), can be compensated by changing the position of the satellite along the orbital path. The maximum likelihood least-squares estimation model must be sufficiently constrained to take care of this problem. This also underscores the need and importance of linear solution for the problem.

The above distinguishing features of satellite imagery imply that the standard perspective camera model will perform rather poorly for push-broom imagery. Our experiments reveal that this is indeed the case.

3.2 Experimental Results

Three experiments were conducted to measure the accuracy of the linear push-broom model. A stereo pair of SPOT images, 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 [**?**]) 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.

Two performance metrics are 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.

To get a lower-bound for the acceptable accuracy, in the first experiment SPOT imagery was treated as imagery acquired using a frame camera and a pin-hole camera model was computed for each image in the stereo pair. The linear push-broom must perform substantially better than a pin-hole camera, both in camera model accuracy, as well as the accuracy of the derived terrain, for it to be acceptable.

In the second and the third experiment the same SPOT imagery was used to compute the linear push-broom models and the complex SPOT models. Clearly, form the linear push-broom model one cannot expect the accuracy of the detailed model that, unlike the linear push-broom, accounts for the eccentricity of the orbit, rotation of the earth, and satellite's attitude drift from one image row to the other. The complex model, in way, provides an upper bound for the accuracy of the linear push-broom 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 $[x, y, z, 1]$ is

	Pin-hole Model	Linear Push-broom Model	Detailed SPOT Model
No ground control pts.	25	25	25
No match pts in model computation	511	100	100
No 3-D terrain pionts generated	68,131	68,131	68,131
RMS model error	11.13 pixels	0.80 pixels	0.73 pixels
Accuracy of generated terrain	380.79m	35.67m	11.10m

Table 1: A comparison of the three camera models.

Figure 2: Terrain elevation derived from a pair of SPOT images assuming pin-hole camera model.

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.

An attempt model a pair of 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 is a poor approximation for push-broom 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 bit misleading as generated terrain is more accurate than the advertised 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! Figs. 2, 3 and 4 show the terrain generated by the pinhole, linear pushbroom, and the complex SPOT models, respectively. Fig. 4 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.

It is interesting to note the distortion introduced in the terrain shown in Fig. 2. Because the images, which are only partially perspective, have been regarded as fully perspective images, the error in the relief is nearly 0 along the center of the image. The error monotonically increases, in opposite directions, as one moves away from the center along the orthographic axis. The error profile — the difference between the terrain generated by the pinhole model and the complex model — is shown in Fig. 5. No such systematic error is discernible in the error profile for the linear push-broom model shown in Fig. 6.

Figure 3: Terrain elevation derived for the same image pair assuming the linear push-broom camera model.

Figure 4: Terrain elevation derived using a complex camera model that simulates the orbital mechanics and SPOT's HRV cameras.

4 Hyperbolic Essential Matrix

The essential matrix Q corresponding to the matching points was defined by Longuet-Higgins [**?**] to be a 3×3 matrix defined by the relation $\tilde{u}_i'^T Q \tilde{u}_i = 0$ for all i. In addition, it was shown in [**?**], and generalized in [**?**] for any arbitrary camera pairs, that Q can be factorized as RS, where R is a rotation matrix and S is a skew-symmetric matrix. In other words, for any stereo pair of images, one can find q_{ij} such that for all match point pairs $[u_k, v_k, 1]$ and $[u'_k, v'_k, 1]$,

$$\begin{bmatrix} u'_k & v'_k & 1 \end{bmatrix} \begin{bmatrix} q_{11} & q_{12} & q_{13} \\ q_{21} & q_{22} & q_{23} \\ q_{31} & q_{32} & q_{33} \end{bmatrix} \begin{bmatrix} u_k \\ v_k \\ 1 \end{bmatrix} = 0. \tag{11}$$

We show that a similar matrix exists for linear push-broom cameras. The following theorem establishes the existence of the hyperbolic essential matrix Q and its form.

THEOREM:

An alternative derivation of Q is as follows. Rewriting Eqs. (**??**) and (**??**), we get:

$$\begin{bmatrix} m_{11} & m_{12} & m_{13} & m_{14} - u & 0 & 0 \\ m_{21} & m_{22} & m_{23} & m_{24} & -v & 0 \\ m_{31} & m_{32} & m_{33} & m_{34} & -1 & 0 \\ n_{11} & n_{12} & n_{13} & n_{14} - u' & 0 & 0 \\ n_{21} & n_{22} & n_{23} & n_{24} & & v' \\ n_{31} & n_{32} & n_{33} & n_{34} & 0 & -1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ 1 \\ w_1 \\ w_2 \end{bmatrix} = 0. \tag{12}$$

The determinant of the above matrix must be 0.... □

4.1 Computation of Q

In order to compute the 12 unknown q_{ij} entries with the help of match points, Eq. (**??**) can be posed as a (possibly overconstrained) system of equations, one equation for each pair of match

Figure 5: Difference between terrain from the pinhole model and the ground the complex model.

Figure 6: Difference between terrain from linear the push-broom model and the complex model.

points. This system of equations has the form $Bx = 0$ where x is the 12 dimensional vector of q_{ij}s. Thus x (equivalently, Q) can be computed using linear, non-iterative techniques, if the constraint $||x|| = 1$ is imposed to avoid the trivial solution.

5 Relative Placement of Cameras and Points

For frame cameras with known internal calibration the essential matrix Q can be separated into a product $RS(T)$ [**?**], where R is a rotation matrix and $S(T)$ is a skew symmetric matrix. It is also possible to accomplish such a factorization for completely uncalibrated cameras [**?**].

5.1 Decomposition of the Hyperbolic Essential Matrix

If the first camera matrix is $[I|0]$ ...

5.2 A Fundamental Result

Theorem 1 *P_1 and P_2 are unknown upto a collineation of space.*

6 Methodology of Stereo Extraction

With a sufficient number of match points, the analysis of their relative disparities to compute the 3-D point locations can proceed as follows.

1. Using the procedure outlined in Section XYZ, compute the transformation Q which contains the relative information about the cameras. Unfortunately, as shown in Lemma XYZ, from a given set of correspondences between the two images, and the Q-matrix, it is *impossible* to determine uniquely all the camera parameters and the position of points in object-space that are compatible with the given data.

2. Using the procedure outlined in Section XYZ, decompose the Q matrix and derive a pair of camera matrices, P_1 and P_2, that are compatible with Q. The camera transformations P_1 and P_2 are obviously not unique.

3. By virtue of Theorem 1, the actual solution (i.e., the 3-D location of points and the camera transformations matrices) that are compatible with the given set of match points are related

to P_1 and P_2 via a 3-dimensional affine transformation H. Since both P_1 and P_2 are off by an unknown affine transformation H, ground control points are used, first to compute H, and then to compute the absolute 3-D location of the points.

7 Conclusion

Acknowledgment

Experimental validation of much of the research presented in this paper would not have been possible without the STEREOSYS program developed by Marsha Jo Hanna at SRI. The authors thank her for use of the program and sharing the source code with them. They also wish to thank Pat Taylor for converting STEREOSYS to C++ and interfacing it to a general-purpose image class hierarchy.