Hillslope Topography From Unconstrained Photographs[1]

Arjun M. Heimsath[2] and Hany Farid[3]

Quantifications of Earth surface topography are essential for modeling the connections between physical and chemical processes of erosion and the shape of the landscape. Enormous investments are made in developing and testing process-based landscape evolution models. These models may never be applied to real topography because of the difficulties in obtaining high-resolution (1–2 m) topographic data in the form of digital elevation models (DEMs). Here we present a simple methodology to extract the high-resolution three-dimensional topographic surface from photographs taken with a hand-held camera with no constraints imposed on the camera positions or field survey. This technique requires only the selection of corresponding points in three or more photographs. From these corresponding points the unknown camera positions and surface topography are simultaneously estimated. We compare results from surface reconstructions estimated from high-resolution survey data from field sites in the Oregon Coast Range and northern California to verify our technique. Our most rigorous test of the algorithms presented here is from the soil-mantled hillslopes of the Santa Cruz marine terrace sequence. Results from three unconstrained photographs yield an estimated surface, with errors on the order of 1 m, that compares well with high-resolution GPS survey data and can be used as an input DEM in process-based landscape evolution modeling.

INTRODUCTION

Landscape form is the result of physical and chemical processes acting upon the surface materials of the Earth. Connections between form and process are the hallmarks of geomorphic study and, increasingly, the call for geomorphologists is to quantify the processes shaping the land surface. This direction differs from quantification of observations by seeking to derive relationships between external forces and landscape form by solving the conservation of mass equation. These mathematical representations of physical processes can be thought of as

---

[2]Department of Earth Sciences, 6105 Fairchild Hall, Dartmouth College, Hanover, New Hampshire 03755; e-mail: arjun.heimsath@dartmouth.edu
[3]Department of Computer Science, 6211 Sudikoff Lab, Dartmouth College, Hanover, New Hampshire 03755; e-mail: hany.farid@dartmouth.edu

"Transport Laws" (Dietrich and others, 2002; Dietrich and Montgomery, 1998) and are grounded in early work by Culling (1960, 1963, 1965), Kirkby (1967, 1971), and Smith and Bretherton (1972). The fundamental connections between process and form were articulated, however, far earlier by Gilbert (1909) and Davis (1892) and remain widely recognized.

Process quantification is tested and applied by numerical modeling, and the application of mathematical models to problems of landscape evolution depends on having high-resolution topographic data from real landscapes represented in the form of digital elevation models (DEMs) (e.g., Dietrich and others, 1995; Montgomery and Dietrich, 1992; Moore, O'Loughlin, and Burch, 1988). For the purposes of this paper we focus on the small catchment, or hillslope scale where the process-based model is most relevant (Dietrich and Montgomery, 1998; Montgomery and Dietrich, 1992; Zhang and Montgomery, 1994). Successful methods of obtaining and applying topographic data necessary to solve geomorphic problems have included laser-total station surveys (e.g, Heimsath and others, 1997), GPS-total station surveys (e.g., Santos and others, 2000), air-photo digitization (e.g., Dietrich and others, 1995), airborne laser altimetry surveys (e.g., Roering, Kirchner, and Dietrich, 1999), and satellite imagery (e.g., Duncan and others, 1998). Satellite imagery has the obvious problem that its present resolution cannot capture landscape form at a process-based scale. Each of these methods is extremely expensive to apply and, ironically, uncertainty at a process-based scale tends to increase with expense, thus justifying the need for a more widely available, relatively inexpensive and user-friendly approach.

Aerial photography has long been used as an efficient method of generating topographic maps and, more recently, DEMs, for geomorphic applications. The labor, expense, and skill necessary to convert stereo pairs of air photos into high-resolution 3D (three-dimensional) data is, however, costly, while the resolution of the DEMs offered by the USGS and other agencies is too coarse for process-based modeling at realistic scales (Dietrich and others, 2002; Dietrich and Montgomery, 1998; Zhang and Montgomery, 1994). Laser altimetry offers great promise to provide DEMs with high resolution (1–2 m vertical) over large areas, but is still cost-prohibitive for most researchers. Recent applications of ground-based photogrammetry have made significant advances at very high resolutions (e.g., Barker, Dixon, and Hooke, 1997; Hancock and Willgoose, 2001; Heritage and others, 1998; Lane, Chandler, and Porfiri, 2001) and at the landscape scale (e.g., Aschenwald and others, 2001), but still require high degrees of constraints upon the position of the cameras and the measurements of all scaling parameters. Reliance on application specific third-party software packages (e.g., Heritage and others, 1998; Lane, 2000) adds to the constraints on using ground-based photogrammetry. Automated digital photogrammetry (e.g., Chandler, 1999; Lane, James, Crowell, 2000; Singh, Chapman, and Atkinson, 1997; Stojic and others, 1998) is a relatively new technique that offers great promise, but also relies entirely on third-party software

and is also tightly constrained in its parameter requirements. The ideal is to have a transparent methodology, allowing clear user interface with the mathematical generation of the DEM, and no parametric constraints on the position of the camera.

Here we present a technology with important geomorphic field applications to extract high-resolution topographic data from a set of unconstrained photographs taken with a hand-held camera (there is no requirement on the camera being digital or traditional, though slides or prints need to be scanned at high resolution to create a digital image). We realized the importance of this technique from work in field settings where the expense of any of the above methods was and continues to be prohibitive. It is also a technique that enables generating data for regions of the landscape that cannot be reached to place targets of the kind used by Heritage and others (1998) and Barker, Dixon, and Hooke (1997), for examples. Importantly, the success of our methodology does not depend on any complicated field calibration or training and can be accomplished following even the most naïve field exploration. We present results to verify a technique with broad application for geomorphologists seeking to quantify the topography of diverse landforms and the application of this tool will enable much further exploration of landscape evolution models at the process-based scale.

ESTIMATING SURFACE TOPOGRAPHY

Figure 1 illustrates the general problem of estimating surface topography from a collection of photographs. An arbitrary 3D scene is imaged from several distinct camera positions. A number of corresponding feature points need to be extracted from the resulting 2D (two-dimensional) images. From these point
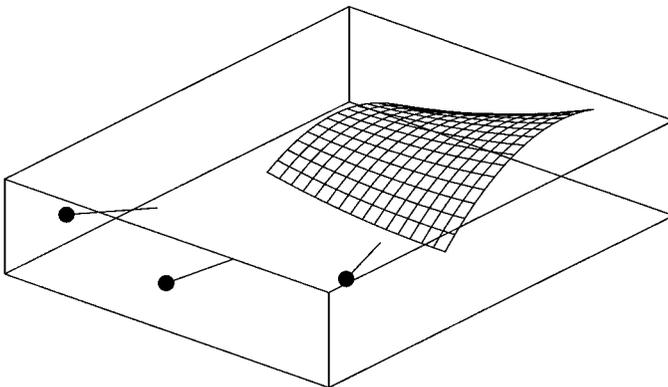
Figure 1. Given a collection of photographs from distinct cameras (solid dots) of an arbitrary scene, the camera positions and 3D scene structure need to be simultaneously estimated.
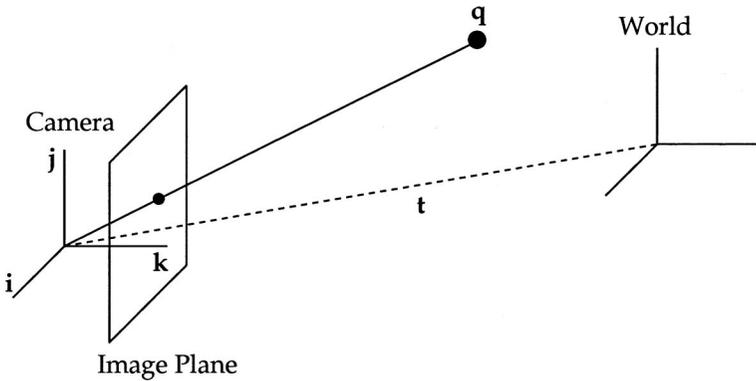
Figure 2. The perspective projection of a point $\vec{q}$ from 3D world coordinates to 2D image coordinates.

correspondences we would like to estimate the 3D structure (topography) of the imaged scene. If the position of each camera is known then this problem would be relatively straightforward. In the absence of such information, however, the problem is considerably more challenging. Within the Computer Vision community, this problem falls under the general heading of structure from motion (SFM). While SFM has received considerable attention (e.g., Boufama, Mohr, and Veillon, 1993; Faugeras, 1992, 1993; Hartley, Gupta, and Chang, 1992; Ma and others, 2000; Maybank, 1993; Poelman and Kanade, 1997; Taylor, Kriegman, and Anandan, 1991; Trigges, 1996), a successful implementation under real-world conditions still poses considerable challenges.

Factorization techniques, while not optimal, provide a simple yet effective approach to SFM (Han and Kanade, 1999; Poelman and Kanade, 1997; Tomasi and Kanade, 1992). We first review one such technique (Poelman and Kanade, 1997), and then show how new constraints unique to surface topography improve the general estimation accuracy.

Imaging Model

Under an ideal pinhole camera model the projection of a point in 3D onto a 2D image plane, Figure 2, is given by the perspective projection equations:[4]

$$x = \frac{f\,\vec{i}^{\mathrm{t}}(\vec{q} - \vec{t})}{\vec{k}^{\mathrm{t}}(\vec{q} - \vec{t})} \quad \text{and} \quad y = \frac{f\,\vec{j}^{\mathrm{t}}(\vec{q} - \vec{t})}{\vec{k}^{\mathrm{t}}(\vec{q} - \vec{t})}, \tag{1}$$

---

[4]A word on notation. Throughout, matrices will be denoted with capital letters, column vectors as $\vec{v}$ and row vectors as $\vec{v}^{\mathrm{t}}$, (where $^{\mathrm{t}}$ denotes transpose), $|\vec{v}|$ denotes vector length, and $\vec{u} \times \vec{v}$ denotes vector cross product.
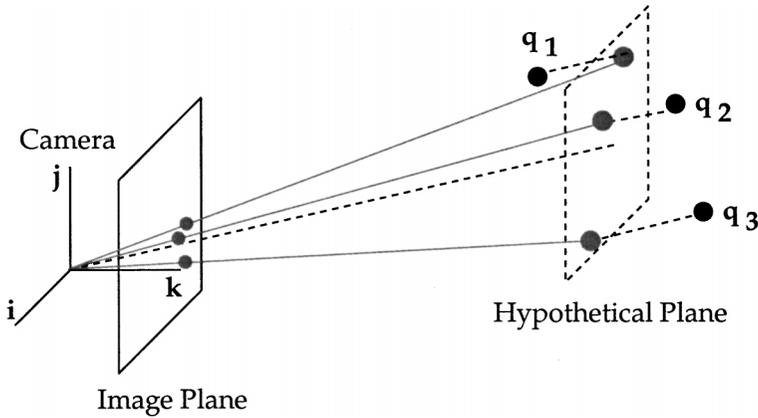
Figure 3. The paraperspective projection of three points from 3D world coordinates to 2D image coordinates.

where $\vec{i}$, $\vec{j}$, and $\vec{k}$ form the coordinate system of the camera, $\vec{t}$ is the translation between the origins of the camera and world coordinate systems, $\vec{q}$ is a point in the 3D world coordinate system, and f is the camera focal length.[5] Since the final structure estimation will only be within a scale factor, Ambiguities section, we may assume that $f = 1$. The inherent nonlinear form of these equations makes their form computationally inconvenient. Paraperspective projection is a linear approximation to perspective projection that affords a more computationally tractable solution for recovering 3D structure (Aloimonos, 1990).

Geometrically, the paraperspective projection of a 3D point involves two steps, Figure 3. The point is first projected onto a hypothetical plane parallel to the image plane. The projection is along the ray connecting the camera focal point to the center of this plane. The point is then projected according to the perspective projection model, Equation (1). Because the hypothetical plane is parallel to the image plane, this projection is a linear transformation. The paraperspective projection equations are given by

$$x = \vec{l}^{\,t}\,\vec{q} + t_x \quad \text{and} \quad y = \vec{m}^{\,t}\,\vec{q} + t_y, \tag{2}$$

where,

$$\vec{l} = \frac{\vec{i} - t_x\,\vec{k}}{-\vec{k}^{\,t}\,\vec{t}} \quad \text{and} \quad \vec{m} = \frac{\vec{j} - t_y\,\vec{k}}{-\vec{k}^{\,t}\,\vec{t}}, \tag{3}$$

[5]The focal length is the distance between the image plane and the camera center (focal point) as measured along the $\vec{k}$ axis (optical axis).

and,

$$t_x = \frac{\vec{i}^{\mathrm{t}}\,\vec{t}}{-\vec{k}^{\mathrm{t}}\,\vec{t}} \quad \text{and} \quad t_y = \frac{\vec{j}^{\mathrm{t}}\,\vec{t}}{-\vec{k}^{\mathrm{t}}\,\vec{t}}, \tag{4}$$

where, for simplicity, but without a loss of generality, it is assumed that the world coordinate system and the hypothetical plane are centered at the center of mass of the points being projected. These equations, unlike perspective projection, are linear with respect to the camera and structure parameters.

Point Correspondences

Consider now, a collection of p points, $\vec{q}_i$, seen from $n \geq 3$ distinct cameras. Denote the *ith* point in image j as the coordinate pair $(x_j(i)\ y_j(i))$. A collection of such points may be obtained from digital images either by hand, or through an automatic extraction procedure (e.g., Lucas and Kanade, 1981). Selecting points by hand involves choosing objects (e.g., stones, bare patches, leaves) identifiable at a pixel scale across all three images. The selected points are packed into a single $2n \times p$ measurement matrix:

$$W = \begin{pmatrix} x_1(1) & \cdots & x_1(p) \\ \vdots & \ddots & \vdots \\ x_n(1) & \cdots & x_n(p) \\ y_1(1) & \cdots & y_1(p) \\ \vdots & \ddots & \vdots \\ y_n(1) & \cdots & y_n(p) \end{pmatrix}. \tag{5}$$

Under a model of paraperspective projection, Equations (2)–(4), the measurement matrix is of the form

$$W = C\,S + T, \tag{6}$$

where the columns of the $3 \times p$ shape matrix S contain the 3D coordinates of the points $\vec{q}_i$. The $2n \times 3$ matrix C and the $2n \times p$ matrix T embody the camera

positions and are given by

$$
C = \begin{pmatrix} \vec{l}_1^{\,t} \\ \vdots \\ \vec{l}_n^{\,t} \\ \vec{m}_1^{\,t} \\ \vdots \\ \vec{m}_n^{\,t} \end{pmatrix} \quad \text{and} \quad T = \begin{pmatrix} t_{x1} \\ \vdots \\ t_{xn} \\ t_{y1} \\ \vdots \\ t_{yn} \end{pmatrix} (1 \cdots 1). \tag{7}
$$

Given a measurement matrix W, and known camera positions (matrices C and T), it would be trivial to solve Equation (6) for the desired 3D structure matrix S. In the absence of such knowledge, however, the problem is considerably more challenging. The problem is made more tractable, however, by observing that since W is a product of a $2n \times 3$ and $3 \times p$ matrix, it will be rank deficient, with a rank of at most 3. In the next section, this rank deficiency is exploited to simultaneously estimate the camera position and scene structure.

Camera and Structure Estimation

Given corresponding 2D points from three or more images, our goal is to determine the 3D coordinates of these points. These 2D coordinates form the measurement matrix, Equation (5). As per our model, Equation (6), the translation matrix T, Equation (7), can be estimated directly as

$$
t_{x_j} = \frac{1}{p} \sum_{i=1}^{p} x_j(i) \quad \text{and} \quad t_{y_j} = \frac{1}{p} \sum_{i=1}^{p} y_j(i), \tag{8}
$$

for $j \in [1, n]$. The translation portion, T, of Equation (6) can be subtracted from the measurement matrix, W, by subtracting t_{x_j} from row j, and t_{y_j} from row $n + j$. The 3D camera position, C, and scene structure, S, will be simultaneously estimated from this "zero-meaned" matrix, W_z.

The matrix W_z is first decomposed according to the singular value decomposition (SVD) as

$$
W_z = U \, D \, V, \tag{9}
$$

where U and V are orthonormal matrices and D is a diagonal matrix. Since the measurement matrix is, in the absence of noise, at most rank 3, we can expect there

to be at most three nonzero diagonal elements in the matrix D. As such, these three matrices can be further decomposed as

$$U = (\,U_1 \mid U_2\,)$$
$$D = \begin{pmatrix} D_1 & 0 \\ 0 & D_2 \end{pmatrix}$$
$$V = \begin{pmatrix} V_1 \\ - \\ V_2 \end{pmatrix}, \tag{10}$$

where the matrices of interest U_1, D_1, and V_1 are of size $2n \times 3$, 3×3, and $3 \times p$, respectively. And where, by the rank deficiency of W_z, D_2 is a zero matrix, and hence, $W_z = U_1 \, D_1 \, V_1$. As per our model, Equation (6), the estimated 3D camera positions and scene structure are given by

$$\tilde{C} = U_1 \sqrt{D_1} \quad \text{and} \quad \tilde{S} = \sqrt{D_1} \, V_1, \tag{11}$$

where the square root is applied to the individual diagonal elements of the matrix D_1. This decomposition is unfortunately not unique since for any invertible matrix M, $W_z = C \, S = (C \, M)\,(M^{-1} \, S)$. In other words we have recovered the 3D camera position and scene structure but only within a linear transformation. What remains then is to impose additional constraints in order to determine the form of the linear transformation M.

Metric Constraints

Ideally, metric constraints would be placed on the camera matrix C by insisting that the estimated coordinate system, $\vec{i}$, $\vec{j}$, $\vec{k}$, of each camera are unit length and orthogonal to one another. Such a constraint, unfortunately, leads to a nonlinear minimization. As a compromise we ask first that the vectors $\vec{i}$ and $\vec{j}$ simply have the same magnitude. From Equation (3), the vector length constraint yields the following relationship:

$$\frac{|\vec{l}_j|^2}{1 + t_{x\,j}^2} = \frac{|\vec{m}_j|^2}{1 + t_{y\,j}^2} \quad \left[= \frac{1}{(\vec{k}_j^{\,t} \, \vec{t}_j)^2} \right], \tag{12}$$

for $j \in [1, n]$. The orthogonality constraint then yields the following:

$$\vec{l}_j^t \, \vec{m}_j = \frac{t_{xj} \, t_{yj}}{(\vec{k}_j^t \, \vec{t}_j)^2}$$

$$= t_{xj} t_{yj} \frac{1}{2} \left(\frac{|\vec{l}_j|^2}{1 + t_{xj}^2} + \frac{|\vec{m}_j|^2}{1 + t_{yj}^2} \right). \tag{13}$$

Over n images, these constraints provide $2n$ constraints on the desired matrix M.

While these constraints are nonlinear in the matrix M, they are linear in the symmetric matrix $Q = M^t \, M$. As such, Equations (12)–(13) form an overconstrained system of linear equations in the six unique elements of the symmetric matrix Q, and are solved using standard least-squares. From Q, the desired matrix M is estimated by decomposing according to the SVD ($Q = U \, D \, V$), from which $M = U \, \sqrt{D}$. The final camera position and scene structure are then simply: $\tilde{C} \, M$ and $M^{-1} \tilde{S}$. The columns of $M^{-1} \tilde{S}$ contain the estimated 3D coordinates of each point $\vec{q}_i$.

Camera Parameters

It is relatively straightforward to show that each camera coordinate system, $\vec{i}$, $\vec{j}$, $\vec{k}$, can be estimated from the previously estimated camera matrix, $\tilde{C} \, M$, (i.e., $\vec{l}$ and $\vec{m}$, Equation (3)), as follows:

$$\vec{k} = G^{-1} \, \vec{h}, \quad \vec{i} = \vec{l}' \times \vec{k}, \quad \text{and} \quad \vec{j} = \vec{m}' \times \vec{k}, \tag{14}$$

where,

$$G = \begin{pmatrix} (\vec{l}' \times \vec{m}')^t \\ \vec{l}'^t \\ \vec{m}'^t \end{pmatrix} \quad \text{and} \quad \vec{h} = \begin{pmatrix} 1 \\ -t_x \\ -t_y \end{pmatrix}, \tag{15}$$

and

$$\vec{l}' = \frac{\vec{l} \, \sqrt{1 + t_x^2}}{|\vec{l}|} \quad \text{and} \quad \vec{m}' = \frac{\vec{m} \, \sqrt{1 + t_y^2}}{|\vec{m}|}. \tag{16}$$

From the estimated camera coordinate system, it is also straightforward to compute the translation vector as follows:

$$\vec{t} = \begin{pmatrix} \vec{i}^{\,\mathrm{t}} \\ \vec{j}^{\,\mathrm{t}} \\ -\vec{k}^{\,\mathrm{t}} \end{pmatrix}^{-1} \begin{pmatrix} t_x\, t_z \\ t_y\, t_z \\ t_z \end{pmatrix}, \tag{17}$$

where,

$$t_z = \sqrt{\frac{1}{2}\left(\frac{1+t_x^2}{|\vec{l}|^2} + \frac{1+t_y^2}{|\vec{m}|^2}\right)}. \tag{18}$$

Estimation Refinement

As a final refinement we perform a nonlinear minimization on each of the camera positions, $(\vec{i},\ \vec{j},\ \vec{k},\ \vec{t})_j$, and scene structure, $\vec{q}_i$. Each 3D point is projected, under perspective projection, Equation (1), through each of the estimated camera positions and compared to the measured 2D positions. The error metric, E, takes the form

$$E = \sum_{j=1}^{p}\sum_{i=1}^{n}\left(\left[x_j(i) - \frac{\vec{i}_j^{\,\mathrm{t}}\,(\vec{q}_i - \vec{t}_j)}{\vec{k}_j^{\,\mathrm{t}}\,(\vec{q}_i - \vec{t}_j)}\right]^2 + \left[y_j(i) - \frac{\vec{j}_j^{\,\mathrm{t}}\,(\vec{q}_i - \vec{t}_j)}{\vec{k}_j^{\,\mathrm{t}}\,(\vec{q}_i - \vec{t}_j)}\right]^2\right), \tag{19}$$

and is minimized using standard gradient descent minimization. This minimization is initialized with the results of the estimate under the paraperspective imaging model. This minimization is performed iteratively, where on each iteration each camera position is separately minimized and then the position of each 3D point is separately minimized. This entire process is repeated until the difference in error between successive iterations is below a specified threshold.

Smoothness Constraint

In the most general case, one is reluctant to introduce explicit constraints on the structure to be estimated. In the case of estimating surface topography, however, it is reasonable to impose a smoothness constraint on the final estimated structure. This is similar to "removing sinks" in most landscape evolution models. We impose a smoothness constraint by encouraging the estimated structure to be locally piecewise planar. This is accomplished by performing a gradient descent minimization on the estimated 3D elevation that minimizes the magnitude of the

second derivative in elevation averaged across the entire estimated structure. This minimization is incorporated in a sequential fashion to the minimization described in the previous section.

This constraint, which typically would not be added to a general purpose SFM algorithm, has the advantage of better conditioning the numerical stability of the surface topography estimation. It does have the slight disadvantage of potentially dulling sharp peaks in the topography. Since the smoothness constraint is imposed over a relatively small area this dulling effect should not, however, be particularly severe.

Ambiguities

There are several inherent ambiguities in the estimated scene structure. The first is that the structure can only be estimated within an arbitrary scale factor and rigid-body rotation. The scale ambiguity can be resolved with explicit knowledge of the distance between any two points in the scene, or size of any object in the scene (e.g., a rock or bush known to be 1 m wide), while the rigid-body ambiguity can be resolved from the 3D position of three or more scene points. The second is a sign ambiguity that arises from the factoring of the final transformation matrix $Q = M^t\,M$. This ambiguity manifests itself in that the same structure reflected about any axis will yield identical results. Visual cues (e.g., a ridge crest is higher than the valley bottom) in the image can be used to resolve this ambiguity without the need for a field survey.

Estimation Results

Shown along the left portion of Figure 4 are three synthetically generated images of a virtual 10 cm unit cube placed at a distance of 250 cm from three virtual cameras with effective focal lengths of 3.5 cm, and rotated by −10, 0, or 10 deg about vertical, and translated horizontally by 5, 0, or −5 cm. Shown in the lower right portion of Figure 4 is the true (filled circle) and estimated (open circle) structure of the cube. The slight errors in the reconstruction are due most likely to the inherent limitations of the approximate paraperspective imaging model. Since the structure is estimated only within a scale factor and arbitrary rotation, the estimated structure is scaled and rotated to bring it into correspondence with the true structure (previous section).

RESULTS

Shown in Figures 5 and 6 are noses from two landscapes previously surveyed with laser total stations as part of work first reported in Heimsath and others (2001) and Heimsath and others (1997), respectively. Both soil mantled, convex-up noses
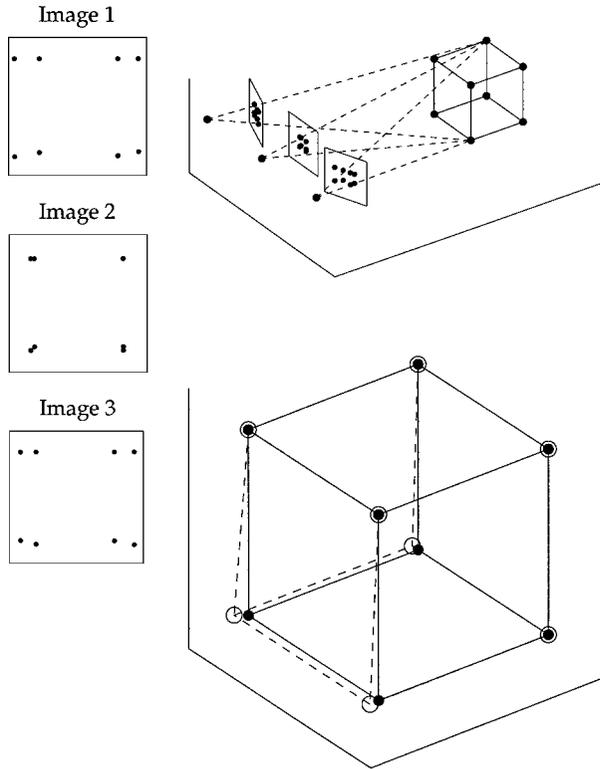
Figure 4. Structure estimation from three images. Shown in the top right is the virtual imaging model (not to scale). Shown to the left are the three images from which the 3D structure is estimated. Shown below is the true and estimated structure. The solid lines and filled dots represent the actual structure, and the dashed lines and open circles represent the estimated structure.

show the characteristic form of hillslopes shaped by diffusion-like sediment transport processed as first articulated by Gilbert (1909) and Davis (1892), helping to direct the work of others cited above. Despite their similar forms, the landscapes are located in different climatic and tectonic environments as described in Heimsath and others (2001), Montgomery and others (1997), and Roering Kirchner, and Dietrich (1999) for the Oregon Coast Range site, Figure 5, and Dietrich and others (1995), Heimsath and others (1997, 1999), and Montgomery and Dietrich (1988), for the northern California Coast Range site, Figure 6. Differences in morphology are evident here, and we explore the process-based significance of these morphologic differences elsewhere (Heimsath and Farid, unpublished data). With results presented here we focus on the nature of the photogrammetric reconstructions.
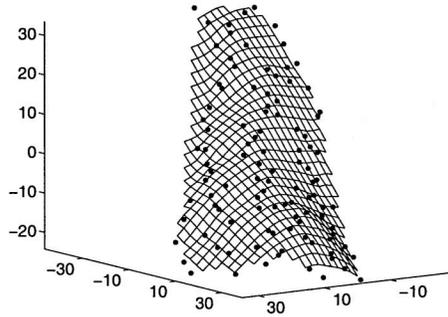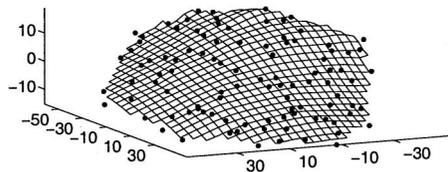
Real
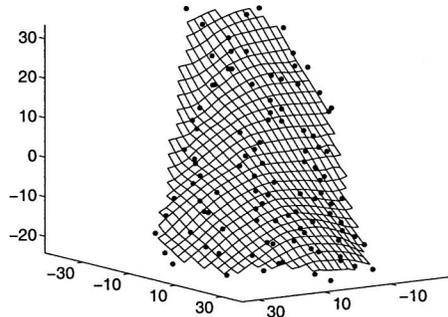
Estimated

Figure 5. Photograph shows the ridge, Coos3, reported in Heimsath and others (2001) and adjacent unchanneled hollow, visible because of clear-cut forestry in the Oregon Coast Range. In the middle panel is a fitted surface to 100 points (black dots) from the original laser total station survey. This surface compares extremely well to data from laser altimetry as shown in Heimsath and others (2001), and also compares well with the estimated surface shown in the lower panel. Units are in meters with arbitrary values.
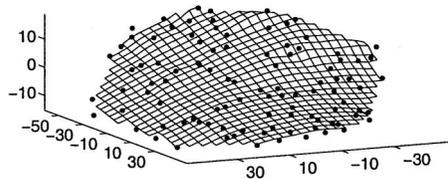
Figure 6. Photograph shows the more subdued topography of northern California showing Nose 4 used in Heimsath and others (1997, 1999). In the middle panel is a fitted surface to 100 points (black dots) from the original laser total station survey. This surface compares well to data from the air-photo-based DEM of Dietrich and others (1995) as discussed in Heimsath and others (1999), and compares very well to the estimated surface shown in the lower panel. Units are in meters with arbitrary values.

An area nearby the ridge used in Figure 5 is shown in Figure 7 and is chosen to test our methods over changes in form from ridge to valley (note the changes in curvature from convex-up to concave-up and back again, corresponding to ridge-valley-ridge, in photograph and data). Data used for this part of the field
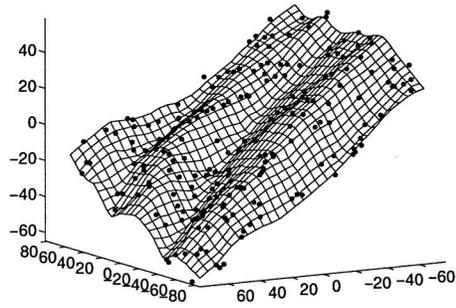
Real
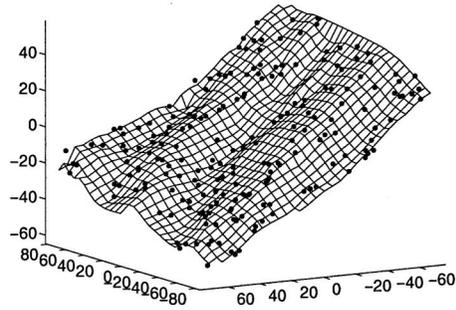
Estimated

Figure 7. Photograph shows a ridge-valley sequence near the ridge of Figure 5. We extracted 200 points from the high-resolution laser altimetry data referenced above to construct the surface in the middle panel. This significantly more complex surface compares quite favorably to the estimated surface mapped in the lower panel. Units are in meters with arbitrary values.

area are from the laser altimetry data used for some of the calculations reported by
Roering, Kirchner, and Dietrich (1999) and to generate the field map of Heimsath
and others (2001). We extracted 200 points randomly from the roughly 2 m scale
original data points to plot the "Real" topography on the middle panel.

Also shown in Figures 5, 6, and 7 are the estimated surfaces. The mean
error in the absolute value of the difference in elevation between the real and
estimated surfaces are 1.01, 1.18, and 1.99 m with a standard deviation of 0.82,
0.81, and 1.62 m, respectively. This error is computed from the fitted surfaces
on an identical sampling lattice. The errors in local slope (first derviative of el-
evation) are 11.9, 13.4, and 14.9%, respectively. In these examples, only a sin-
gle photograph was available, so we simply projected the known 3D points onto
three virtual cameras following a model of perspective projection, Equation (1).
The resulting "images" yielded the necessary 2D point correspondences, from
which the unknown camera positions and surface structure were simultaneously
estimated.

The landscape shown in Figure 8 is from the Santa Cruz marine terrace
sequence studied extensively across disciplines (e.g., Anderson, Densmore, and
Ellis, 1999; Perg, Anderson, and Finkel, 2001; Rosenbloom and Anderson, 1994).
We use this landscape as a well-constrained test of our methodology that will
have further application when comparing landscape development across terraces of
known ages (Heimsath and Farid, unpublished data). Here we use 100 hand selected
points shown in the three photographs of the same nose to simultaneously estimate
the unknown camera positions and surface topography. The bottom two surfaces
show the real (from GPS total station survey) and estimated surface topographies
of the nose shown in the photographs. The mean error in elevation between the
real and estimated surfaces is 0.90 m with a standard deviation of 0.75 m.

In summary, the steps from photographs to surface topography are as follows:

1. Photograph the target surface from three distinct viewpoints (the views
   should be separated by at least a few meters). As a general rule we rec-
   ommend moving left and right of a central viewing position by at least
   a few meters and rotating the camera about the vertical axis by approxi-
   mately 10–30 deg. If photographing with a traditional camera, digitize the
   photographs at a minimum of 600 dpi.
2. In each photograph select between 50 and 100 feature points (e.g., a point
   on a boulder, a leaf on the ground, the base of a shrub, etc.). The resulting
   point positions form the measurement matrix of Equation (5). Our software
   has a simple interface that allows users to manually extract and store the
   pixel location of each point.
3. Follow the series of steps outlined in the previous section in order to
   determine the surface topography. Source code for these computations are
   freely available upon request.

Real

Estimated
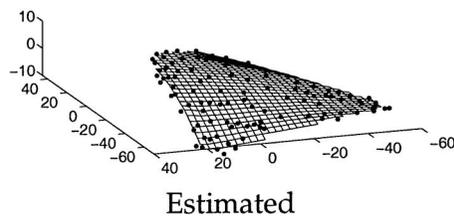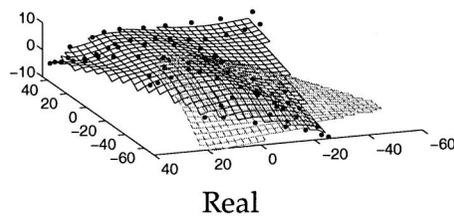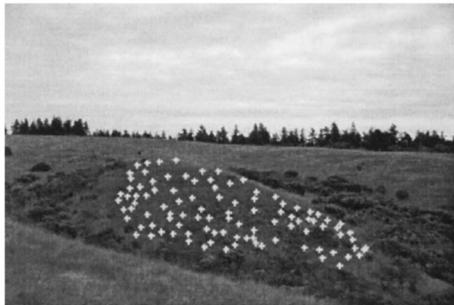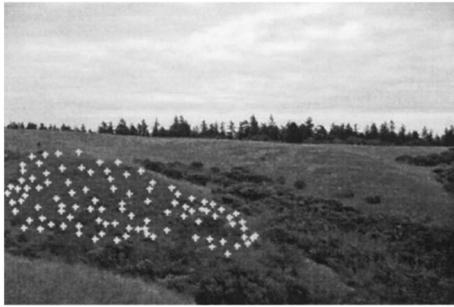
DISCUSSION

Computational Limitations

Factorization techniques as described here afford a simple and yet effective approach for the recovery of surface topography from photographs taken with uncalibrated and unknown camera positions. This technique requires the selection of corresponding points (on the order of 100) from three or more photographs, from which the camera positions and surface topography are simultaneously estimated. Computationally, this technique begins with a paraperspective approximation to the geometry of image formation. This approximation affords a closed-form analytic solution for surface topography, and is further refined through successive nonlinear minimizations that assume a more realistic imaging model, and imposes an overall smoothness constraint on the recovered structure. These nonlinear minimizations yield a more accurate and stable estimate.

There are, of course, a number of different computational approaches from which to choose. We have adopted this particular technique because in our experience other more sophisticated techniques appear to be very sensitive to even slight (subpixel) errors in point correspondences. Whereas, the proposed technique requires only a relatively coarse point selection process. There are still, nevertheless, some limitations. As can be seen in Figures 5–8, there is a consistent flattening of the estimated structure. This is due most likely to the initial paraperspective approximation that assumes that the points being imaged lie on a fronto-parallel plane. This is a fundamental limitation and its effect can be minimized by photographing from a vantage point that is in line with the surface normal of the overall topography. Even with this limitation, we find that surface topography can be approximated with, on average, an error of 1–2 in elevation.

Related Photogrammetric Methods

Since Icarus attempted flight to get a better look at the Earth's surface, humans have been developing more and more efficient ways to map large areas of the landscape. Recent developments in photogrammetry have led to significant

←——

Figure 8. Photographs show nose from Blackrock terrace (Terrace 4) on the Santa Cruz, California marine terrace sequence as reported in Rosenbloom and Anderson (1994). Crosses show 100 hand selected corresponding points. These points are used to simultaneously estimate the camera positions and surface topography. Second from the bottom is a fitted surface from 76 survey points obtained with GPS. This surface compares quite favorably with the estimated surface shown below. Units are in meters with arbitrary values.

improvements over the painstaking methods of digitizing aerial photographs. These methods range from the landscape (e.g., Aschenwald and others, 2001) to the experimental scale (e.g., Hancock and Willgoose, 2001). Studies at the intermediate scales, of interest when considering process-based landscape evolution, appear to be at the reach scale ($\approx$10 m directed at fluvial erosion studies (e.g., Barker, Dixon, and Hooke, 1997; Heritage and others, 1998), rather than at hillslope (>100 m length scale) or even first or second-order watershed scales (>1 ha). Debris flow mapping by Coe, Glancy, and Whitney (1997), for example, relied on DEMs built from aerial stereo-photographs. The other fundamental limitation of the methods presented in these and other studies is the reliance on third-party software. The procedure we present is an entirely stand-alone process that enables close control on the nature of the surface reconstruction. While there is indeed great appeal to have automated programs extract a DEM from landscape-scale photographs, we do not believe such a procedure exists in a way that is affordable to most researchers.

Aschenwald and others (2001) present a study with interests most similar to the ones that could be addressed by the technique we developed here. Their geo-rectification of terrestrial, high-oblique photographs for input into a GIS matches our interest in generating high-resolution topographic data. The DEM used in their study was, however, extracted separately, from existing contour lines rather from the photographic image. Their orthorectified photo images are then draped upon the existing DEM and precisely located by using 15 ground control points (over about 3 km^2) and used to compare preanalyzed time-series images. Methods used are not transparent, however, because of the use of software packages. This study presents a combination of disparate methods used previously to explore different questions. In contrast, we present algorithms to generate 3D coordinates from three or more corresponding points whose 2D coordinates are their locations on the 2D photographic image. Like the study of Aschenwald and others (2001), no ground placement of specific points is required, but unlike their study, our methods actually generate the DEM that represents the landscape. Resolution of the DEM depends, at this point, on how many corresponding points are user selected from between the photographs. We are currently exploring algorithms that will select points automatically, similar to the technique fulfilled by software used by, for example, Heritage and others (1998) and reviewed in Lane (2000).

High-resolution extraction of DEMs from photographs at a field (Barker, Dixon, and Hooke, 1997; Heritage and others, 1998; Lane, 2000; Westaway, Lane, and Hicks, 2000) or experimental (Chandler, 1999; Hancock and Willgoose, 2001; Lane, Chandler, and Porfiri, 2001) scale is a technique currently reliant upon third-party software, but with applications similar to ours. The bottom line for workers such as these is the generation of a DEM. While the accuracies reported by Barker, Dixon, and Hooke (1997), Hancock and Willgoose (2001), Heritage and others

(1998), and Lane, James, and Crowell (2000) are admirable, the constraints needed are unreasonable for the field applications that our method is directed toward. For example, Hancock and Willgoose (2001) and Lane, Chandler, and Porfiri (2001) use digital cameras placed precisely above their experimental landscape and are able to extract DEMs of the evolving features at a millimeter resolution, but they were measuring an area less than 2.5 m^2. At least eight precisely located control points were also used to calibrate the photogrammetry, a task that would be unreasonable in remote field settings. Both Barker, Dixon, and Hooke (1997) and Heritage and others (1998), for examples, apply photogrammetry to the fluvial environment and use their DEMs to determine morphological change. Control points were surveyed in precisely and used in both studies to calibrate the extracted DEM. Additionally, the camera positions were precisely located in relation to the studied areas. Both these requirements, and their reliance on application specific third-party software, make their procedures unwieldy for our applications. Our implementation relies on the popular and general-purpose numerical analysis package MatLab.

Geomorphic Applications

Application of process-based geomorphic transport laws towards understanding how landscape form changes with time depends on having high-resolution topographic data from real landscapes. Process-based modeling typically estimates how landscapes evolve under different climatic, tectonic, and anthropogenic influences (e.g., Anderson, Densmore, and Ellis, 1999; Dietrich and others, 1995; Montgomery and Dietrich, 1992; Moore, O'Loughlin, and Burch, 1988). The limitations on the above methods of generating high-resolution DEMs motivated this study and our low-cost methodology to extract topographic data is applicable to landscapes across a wide range of environmental conditions. The only stipulation is that the landscape can be photographed (i.e., dense vegetative cover presents a visual barrier that even airborne laser altimetry, with its ability to penetrate through some vegetative cover, has difficulty overcoming) that at least three frames capture the same set of points, and that there is some estimate of scale between all of the pictures. When these conditions are met, our methodology can yield high resolution DEMs at scales dependent only on the scale captured by the photographs. The example chosen for field verification here provides the ideal first-cut test of the methodology.

Landscape development shown by the ridge-hollow topography (Fig. 8) at this emergent marine terrace site is constrained by the time since the terrace emerged from Pacific Ocean. Extreme differences in terrace ages estimated by different studies (most recently examined by (Perg, Anderson, and Finkel, 2001) can therefore be tested with a simple landscape evolution model coupled with detailed topographic data extracted nonintrusively by methods we report here. These

data could, for example, be used with the terrace emergence modeling effort of Anderson, Densmore, and Ellis (1999), combined with a process-based landscape evolution model (Dietrich and others, 1995) to more accurately define the uplift history of the Santa Cruz terrace sequences. Specifically, differences between Perg, Anderson, and Finkel's determination of the emergence ages of the terraces (Perg, Anderson, and Finkel, 2001) and those by Bradley and Addicott (1968) and Bradley and Griggs (1976) are up to an order of magnitude. Physical parameters that help constrain the processes shaping these landscapes are relatively (compared to other landscapes) well known for the region (e.g., Heimsath and others, 1997; Rosenbloom and Anderson, 1994) and can therefore be used to test which estimate of terrace age is more appropriate. The only thing missing is the topographic data, which is where the methodology we present here becomes relevant. This field area is especially good for testing such methodology as the landscape is easily accessible (Fig. 8) and can therefore be thoroughly surveyed as shown here. Such an example is one of numerous that will benefit from this methodology.

CONCLUSIONS

Making connections between landscape form and the geomorphic processes responsible for shaping that form has been of interest to geomorphologists for over 100 years. Increasingly, our understanding of how landscape surfaces are shaped under different geomorphic processes is improved by numerically modeling real landscapes represented by DEMs. Typical methods for extracting DEMs from remotely sensed imagery are expensive and can require large investments of time by skilled workers. The method we present here is ideally suited for use on landscapes that are difficult to reach, or when field time and resources are limited. With relatively little time behind the computer and only a few photographs, this methodology enables the extraction of accurate high-resolution topographic data. We find it especially appealing that the algorithms are transparent to the field scientist and that the results can be readily compared to the photographed surface.

In the photographic process there is an inherent loss of information, namely 3D structure. It is possible to estimate this 3D information from several photographs taken from two or more calibrated cameras (i.e., cameras with known positions relative to one another). Surprisingly, this information can also be estimated from three or more photographs with unknown camera positions. Building on work from the structure from motion literature, we have presented a simple and effective computational technique for estimating 3D structure from three or more photographs. This process is particularly attractive as it imposes no constraints on the scene being imaged or on the camera's positions. Mathematically, this technique involves standard tools from Linear Algebra and Numerical Analysis available in most standard mathematical software packages (e.g., MatLab, Mathematica, or Maple).

Our implementation employs MatLab: the complete source code is freely available upon request.

ACKNOWLEDGMENTS

REFERENCES

Aloimonos, J. Y., 1990, Perspective approximations: Image Vision Computing, v. 8, no. 3, p. 177–192.

Anderson, R. S., Densmore, A. L., and Ellis, M. A., 1999, The generation and degradation of marine terraces: Basin Res., v. 11, no. 1, p. 7–19.

Aschenwald, J., Leichter, K., Tasserand, E., and Tappeiner, U., 2001, Spatio-temporal landscape analysis in mountainous terrain by means of small format photography: A methodological approach: IEEE Trans. Geosci. Remote Sensing, v. 39, no. 4, p. 885–893.

Barker, R., Dixon, L., and Hooke, J., 1997, Use of terrestrial photogrammetry for monitoring and measuring bank erosion: Earth Surf. Processes Landforms, v. 22, no. 13, p. 1217–1227.

Boufama, B., Mohr, R., and Veillon, F., 1993, Euclidean constraints for uncalibrated reconstruction, *in* International conference on computer vision: IEEE Computer Society Press, Los Alamitos, CA, p. 466–470.

Bradley, W. C., and Addicott, W. O., 1968, Age of first marine terrace near santa cruz, California: Geol. Soc. Am. Bull., v. 79, p. 1203–1210.

Bradley, W. C., and Griggs, G. B., 1976, Form, genesis, and deformation of central californa wave-cut platforms. Geol. Soc. Am. Bull., v. 87, p. 433–449.

Chandler, J., 1999, Effective application of automated digital photogrammetry for geomorphological research: Earth Surf. Processes Landforms, v. 24, no. 1, p. 51–63.

Coe, J. A., Glancy, P. A., and Whitney, J. W., 1997, Volumetric analysis and hydrologic characterization of a modern debris flow near yucca mountain, nevada: Geomorphology, v. 20, no. 1–2, p. 11–28.

Culling, W. E. H., 1960, Analytical theory of erosion. J. Geol., v. 68, p. 336–344.

Culling, W. E. H., 1963, Soil creep and the development of hillside slopes: J. Geol., v. 71, no. 2, p. 127–161.

Culling, W. E. H., 1965, Theory of erosion on soil-covered slopes: J. Geol., v. 73, p. 230–254.

Davis, W. M., 1892, The convex profile of badland divides: Science, v. 20, p. 245.

Dietrich, W. E., Bellugi, D., Heimsath, A. M., Roering, J. J., Sklar, L., and Stock, J. D., 2002, Prediction in geomorphology [Monograph]: American Geophysical Union, Washington, DC.

Dietrich, W. E., and Montgomery, D. R., 1998, Hillslopes, channels, and landscape scale: Cambridge University Press, Cambridge, England.

Dietrich, W. E., Reiss, R., Hsu, M.-L., and Montgomery, D. R., 1995, A process-based model for colluvial soil depth and shallow landsliding using digital elevation data: Hydrol. Processes, v. 9, p. 383–400.

Duncan, C. C., Klein, A. J., Masek, J. G., and Isacks, B. L., 1998, Comparison of late pleistocene and modern glacier extents in central Nepal based on digital elevation data and satellite imagery: Quaternary Res., v. 49, no. 3, p. 241–254.

Faugeras, O., 1992, What can be seen in three dimensions with an uncalibrated stereo rig? *in* European conference on computer vision: Springer-Verlag, New York, New York, p. 563–578.

Faugeras, O., 1993, Three-dimensional computer vision—a geometric viewpoint: MIT Press, Cambridge, MA.

Gilbert, G. K., 1909, The convexity of hilltops: J. Geol., v. 17, no. 4, p. 344–350.

Han, M., and Kanade, T., 1999, Perspective factorization methods for Euclidean reconstruction, Technical report CMU-RI-TR-99-22: The Robotics Institute, Carnegie Mellon University, Pittsburgh, PA.

Hancock, G., and Willgoose, G., 2001, The production of digital elevation models for experimental model landscapes. Earth Surf. Processes Landforms, v. 26, no. 5, p. 475–490.

Hartley, R., Gupta, R., and Chang, T., 1992, Stereo from uncalibrated cameras, Conference on computer vision and pattern recognition: IEEE Computer Society Press, Los Alamitos, CA, p. 761–764.

Heimsath, A. M., Dietrich, W. E., Nishiizumi, K., and Finkel, R. C., 1997, The soil production function and landscape equilibrium: Nature, v. 388, p. 358–361.

Heimsath, A. M., Dietrich, W. E., Nishiizumi, K., and Finkel, R. C., 1999, Cosmogenic nuclides, topography, and the spatial variation of soil depth: Geomorphology, v. 27, no. 1–2, p. 151–172.

Heimsath, A. M., Dietrich, W. E., Nishiizumi, K., and Finkel, R. C., 2001, Stochastic processes of soil produc- tion and transport: Erosion rates, topographic variation, and cosmogenic nuclides in the oregon coast range. Earth Surf. Processes Landforms, v. 26, p. 531–552.

Heritage, G. L., Fuller, I. C., Charlton, M. E., Brewer, P. A., and Passmore, D. P., 1998, Cdw photogrammetry of low relief fluvial features: Accuracy and implications for reach-scale sediment budgeting. Earth Surf. Processes Landforms, v. 23, no. 13, p. 1219–1233.

Kirkby, M. J., 1967, Measurement and theory of soil creep: J. Geol., v. 75, no. 4, p. 359–378.

Kirkby, M. J., 1971, Hillslope process-response models based on the continuity equation: Inst. Br. Geogr. Spec. Publ., v. 3, p. 15–30.

Lane, S. N., 2000, The measurement of river channel morphology using digital photogrammetr: Photogram. Rec., v. 16, no. 96, p. 937–957.

Lane, S. N., Chandler, J. H., and Porfiri, K., 2001, Monitoring river channel and flume surfaces with digital photogrammetry: J. Hydraul. Eng.-asce, v. 127, no. 10, p. 871–877.

Lane, S. N., James, T. D., and Crowell, M. D., 2000, Application of digital photogrammetry to complex topography for geomorphological research: Photogram. Rec., v. 16, no. 95, p. 793–821.

Lucas, B. D., and Kanade, T., 1981, An iterative image registration technique with an application to stereo vision, *in* 7th international joint conference on artificial intelligence: Springer-Verlag, New York, New York, p. 674–679.

Ma, Y., Soatto, S., Kosecka, J., and Sastry, S., 2000, Euclidean reconstruction and reprojection up to sub-groups: Int. J. Comput. Vision, v. 38, no. 3, p. 219–229.

Maybank, S., 1993, Theory of reconstruction from image motion: Springer, Berline.

Montgomery, D. R., and Dietrich, W. E., 1988, Where do channels begin? Nature, v. 336, no. 6196, p. 232–234.

Montgomery, D. R., and Dietrich, W. E., 1992, Channel initiation and the problem of landscape scale: Science, v. 255, no. 5046, p. 826–830.

Montgomery, D. R., Dietrich, W. E., Torres, R., Anderson, S. P., Heffner, J. T., and Logue, K., 1997, Hydrologic response of a steep, unchanneled valley to natural and applied rainfall: Water Resour. Res., v. 33, no. 1, p. 91–109.

Moore, I. D., O'Loughlin, E. M., and Burch, G. J., 1988, A contour-based topographic model for hydroloigcal and ecological applications: Earth Surf. Processes Landforms, v. 13, p. 305–320.

Perg, L. A., Anderson, R. S., and Finkel, R. C., 2001, Use of a new ^{10}BE and ^{26}AL inventory method to date marine terraces, Santa Cruz, California, USA: Geology, v. 29, no. 10, p. 879–882.

Poelman, C. J., and Kanade, T., 1997, A paraperspective factorization method for shape and motion recov ery: IEEE Trans. Pattern Anal. Mach. Intelligence, v. 19, no. 3, p. 206–218.

Roering, J. J., Kirchner, J. W., and Dietrich, W. E., 1999, Evidence for non-linear, diffusive sediment transport on hillslopes and implications for landscape morphology: Water Resour. Res., v. 35, no. 3, p. 853–870.

Rosenbloom, N. A., and Anderson, R. S., 1994, Hillslope and channel evolution in a marine terraced landscape, Santa Cruz, California: J. Geophys. Res., v. 99, no. B7, p. 14013–014029.

Santos, M. L. M., Guenat, C., Bouzelboudjen, M., and Golay, F., 2000, Three-dimensional gis cartography applied to the study of the spatial variation of soil horizons in a swiss floodplain: Geoderma, v. 97, no. 3–4, p. 351–366.

Singh, R., Chapman, D. P., and Atkinson, K. B., 1997, Digital photogrammetry for automatic close range measurements of textureless and featureless objects: Photogram. Rec., v. 15, no. 89, p. 691–702.

Smith, T. R., and Bretherton, F. P., 1972, Stability and the conservation of mass in drainage basin evolution: Water Resour. Res., v. 8, p. 1506–1529.

Stojic, M., Chandler, J., Ashmore, P., and Luce, J., 1998, The assessment of sediment transport rates by automated digital photogrammetry: Photogram. Eng. Remote Sensing, v. 64, no. 5, p. 387–395.

Taylor, C., Kriegman, D., and Anandan, P., 1991, Structure and motion estimation from multiple images, *in* IEEE workshop on Visual Motion: IEEE Computer Society Press, Los Alamitos, CA, p. 242–248.

Tomasi, C., and Kanade, T., 1992, Shape and motion from image streams under orthography—a factorization method: Int. J. Comput. Vision, v. 9, no. 2, p. 137–154.

Triggs, B., 1996, Factorization methods for projective structure and motion, *in* Computer vision and pattern recognition: IEEE Computer Society Press, Los Alamitos, CA, p. 845–851.

Westaway, R. M., Lane, S. N., and Hicks, D. M., 2000, The development of an automated correction procedure for digital photogrammetry for the study of wide, shallow, gravel-bed rivers: Earth Surf. Processes Landforms, v. 25, no. 2, p. 209–226.

Zhang, W. H., and Montgomery, D. R., 1994, Digital elevation model grid size, landscape representation, and hydrologic simulations: Water Resour. Res. v. 30, no. 4, p. 1019–1028.