Graphics and Image Processing
J. Foley Editor

Using Synthetic Images to Register Real Images with Surface Models

Berthold K. P. Horn and Brett L. Bachman
Massachusetts Institute of Technology

A number of image analysis tasks can benefit from registration of the image with a model of the surface being imaged. Automatic navigation using visible light or radar images requires exact alignment of such images with digital terrain models. In addition, automatic classification of terrain, using satellite imagery, requires such alignment to deal correctly with the effects of varying sun angle and surface slope. Even inspection techniques for certain industrial parts may be improved by this means.

We achieve the required alignment by matching the real image with a synthetic image obtained from a surface model and known positions of the light sources. The synthetic image intensity is calculated using the reflectance map, a convenient way of describing surface reflection as a function of surface gradient. We illustrate the technique using LANDSAT images and digital terrain models.

Key Words and Phrases: image registration, synthetic images, surface models, automatic hill shading, digital terrain models, image transformation, image matching, shaded images

CR Categories: 3.63, 3.11, 3.14, 8.2, 3.83

Permission to copy without fee all or part of this material is granted provided that the copies are not made or distributed for direct commercial advantage, the ACM copyright notice and the title of the publication and its date appear, and notice is given that copying is by permission of the Association for Computing Machinery. To copy otherwise, or to republish, requires a fee and/or specific permission.

This report describes research done at the Artificial Intelligence Laboratory of the Massachusetts Institute of Technology. Support for the laboratory's artificial intelligence research is provided in part by the Advanced Research Projects Agency of the Department of Defense under Office of Naval Research contract N00014-75-0643.

Authors' addresses: B. K. P. Horn, Artificial Intelligence Laboratory, MIT, Cambridge, MA 02139; B. L. Bachman, Department of Electrical Engineering, MIT, Cambridge, MA 02139

© 1978 ACM 0001-0782/78/1100-0914 $00.75

1. Motivation

Interesting and useful new image analysis methods may be developed if registered image intensity and surface slope information is available. Automatic change detection, for example, seems unattainable without an ability to deal with variations of appearance with changes in the sun's position. In turn, these variations can be understood only in terms of surface topography and reflectance models. Similarly, human cartographers consult both aerial photographs and topographic maps of a region to trace the paths of streams and rivers. Automatic analysis of either of these information sources alone is unlikely to lead to robust methods for performing this task.

An important application of aligned image and surface information lies in the area of automatic terrain classification. To date, no account has been taken of varying surface gradient, sun position, or the physics of light reflection in the ground cover. Classification ought to be based on measurable properties of the surface, not raw image intensities, which are only indirectly related to these properties. Classification techniques have been limited in their application to flat regions and have had to be retrained for images with different sun angles. Aligning images with surface models will permit removal of the image intensity component due to varying orientation of the surface elements.

Another application may be found in the inspection of industrial parts with complicated surfaces. Aligning images of these parts with models of their surfaces should permit one to determine defects in the surfaces which give rise to differences between real and synthesized images. It may also be possible to determine the position and orientation of a part by such techniques. This would then lead to methods which may guide a computer-controlled manipulator to retrieve one of the topmost parts in a bin full of parts. In this case, further work will be required to ascertain the effects of mutual illumination due to the proximity of parts to one another.

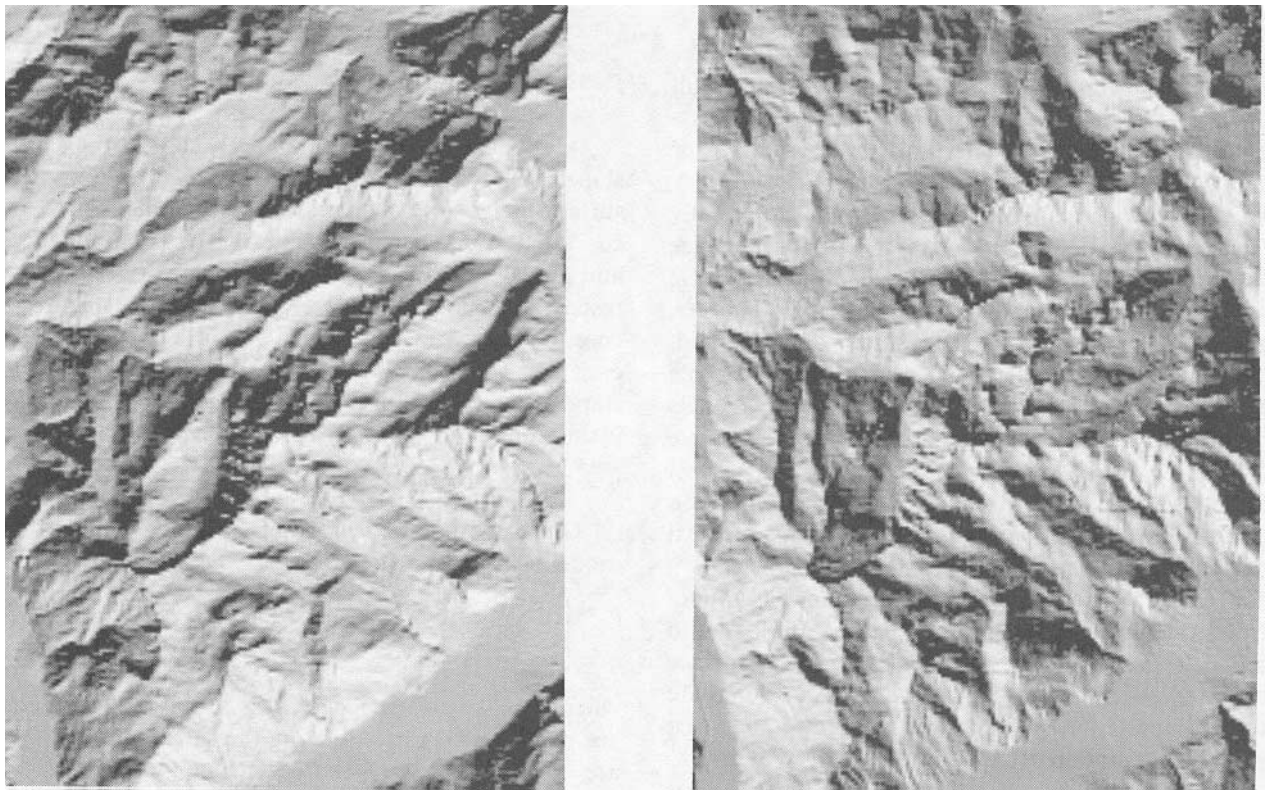
Accurate alignment of images with surface models is therefore an important prerequisite for many image understanding tasks. We describe here an automatic method of potentially high accuracy that does not depend on feature extraction or other sophisticated image analysis methods. Instead, all that is required is careful matching of the real with a synthetic image. Because this is an area-based process, it has the potential for subpixel accuracy—accuracy not easily attained with techniques dependent on alignment of *linear* features such as edges or curves. The method is here illustrated by registering LANDSAT images with digital terrain models.

2. Possible Approaches

One way to align a real image with a surface model might be through the use of a reference image obtained

Fig. 1. Early morning (9:55 G.M.T.) synthetic image.

Fig. 2. Early afternoon (13:48 G.M.T.) synthetic image.



under controlled conditions. New images could then be matched against the reference image to achieve alignment. Unfortunately, the appearance of a surface depends quite dramatically on the position of the light source (see Figures 1 and 2, for example), so that this method works only for a limited daily interval for a limited number of days each year [1]. This problem disappears when one uses synthetic images, since the position of the source can be taken into account.

A more sophisticated process would not match images directly, but first perform a feature extraction process on the real image and then match these features with those found in the reference image. One finds, however, that different features will be seen when lighting changes: for example, ridges and valleys parallel to the illumination direction tend to disappear (see Figures 1 and 2). In addition, the apparent position of a feature as well as its shape may depend somewhat on illumination. More serious may be the present feature extraction scheme's computational cost and lack of robustness.

One might next consider calculating the shape of the surface from intensities in the image [8, 9]. This, however, is computationally expensive and not likely to be very accurate in view of the variation in the nature of surface cover. A more appropriate method, estimating the local gradient using similar methods [12] and then matching these with gradients stored in the terrain model, still involves a great deal of computation.

The method chosen here depends instead on matching the real image with a synthetic image produced from the terrain model. The similarity of the two images depends in part upon how closely the assumed reflectance matches the real one. For mountainous terrain and for images taken with low sun elevations, rather simple assumptions about the reflectance properties of the surface gave very good results. Since all LANDSAT images are taken at about 9:30 local solar time, the sun elevations in this case are fairly small and image registration for all but flat terrain is straightforward.

This implies that LANDSAT images are actually not optimal for automatic terrain classification, since the intensity fluctuations due to varying surface gradients often swamp the intensity fluctuations due to variations in surface cover. An important application of our technique in fact is the removal of the intensity fluctuations due to variations in surface gradient from satellite images in order to facilitate the automatic classification of terrain. To do this, we must model the way the surface reflects light.

3. The Reflectance Map

Work on image understanding has led to a need to model the image-formation process. One aspect of this concerns the geometry of projection, that is, the relationship between the position of a point and the coordinates of its image. Less well understood is the problem

determining image intensities, which requires modeling of the way surfaces reflect light. For a particular kind of surface and a particular placement of light sources, surface reflectance can be plotted as a function of surface gradient (magnitude and direction of slope). The result is called a *reflectance map* and is usually presented as a contour map of constant reflectance in gradient space [12].

The reflectance map may be determined empirically by measuring the reflectance of a small, flat surface sample as it is oriented in many different directions while surrounded by the desired distribution of light sources. Alternatively, a test object such as a sphere or paraboloid, which contains surface elements oriented in all possible directions may be used [8, 9, 12]. Mathematical models of surfaces have also been developed in order to derive analytical expressions for surface reflectance or at least numerical values obtained by Monte Carlo simulation [6]. In related graphics work, simple phenomenological models have been used [4, 17, 2].

Since the reflectance map gives reflectance as a function of local surface gradient only, it does not take into account effects dependent on the *position* of the surface element. Two such effects which are not considered are mutual illumination of surface elements and cast shadows. Illumination of portions of a surface by neighboring surface elements when the object concerned has concavities is difficult to model and leads to global computations. Fortunately, this effect is usually small unless the surface reflectance is exceptionally high [12]. The reflectance map correctly accounts for self-shadowed surfaces, but not shadows cast by one surface element on another. Such cast shadows can however be calculated using well-known hidden-surface algorithms to predict which surface elements are not visible from the source [26, 27, 20].

One use of the reflectance map is in the determination of surface shape from intensities [8, 9] in a single image; here, however, it will be employed only in order to generate synthetic images from digital terrain models.

4. Digital Terrain Models

Work on computer-based methods for cartography, prediction of side-looking radar imagery for flight-simulators, automatic hill-shading, and machines that analyze stereo aerial photography has led to the development of digital terrain models. These models are usually in the form of an array of terrain elevations, z_{ij}, on a square or rectangular grid.

Data used for this paper's illustrations were entered into a computer after manual interpolation from a contour map and have been used previously in work on automatic hill-shading [3, 11]. It consists of an array of 175 x 240 elevations on a 100-meter grid corresponding to a 17.5 km by 24 km region of Switzerland lying between 7°1′ east to 7°15′ east and 46°8.5′ north to 46°21.5′ north.

5. The Gradient

A gradient has two components, namely the surface slope along two mutually perpendicular directions. If the surface height, z, is expressed as a function of two coordinates x and y, we define the two components, p and q of the gradient as the partial derivatives of z with respect to x and y, respectively. In particular, a Cartesian coordinate system is erected with the x-axis pointing east, the y-axis north and the z-axis up. Then, p is the slope of the surface in the west-to-east direction, while q is the slope in the south-to-north direction:

$$p = \partial z/\partial x \qquad q = \partial z/\partial y.$$

One can estimate the gradient from the digital terrain model using first differences:

$$p \simeq [z_{(i+1)j} - z_{ij}]/\Delta$$

$$q \simeq [z_{i(j+1)} - z_{ij}]/\Delta$$

where Δ is the grid-spacing. More sophisticated schemes are possible [11] for estimating the surface gradient, but are unnecessary. We assume that the imaging system is on the z-axis at a large distance from the surface, with its optical axis pointing straight down.

6. Position of the Light Sources

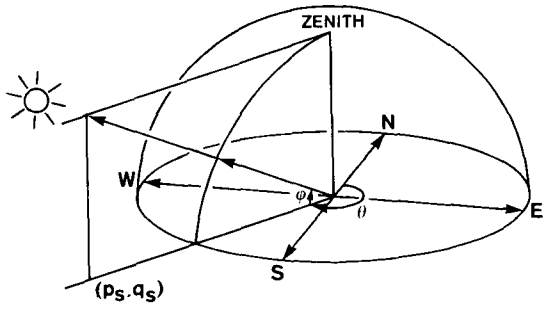
In order to be able to calculate the reflectance map, it is necessary to know the location of the light source. In our case the primary source is the sun, and its location can be determined easily by using tables intended for celestial navigation [25, 24, 7] or by straightforward computations [14, 19, 29, 10]. In either case, given the date and time, the azimuth (θ) and the elevation (ϕ) of the sun can be found. Here, azimuth is measured clockwise from north, while elevation is simply the angle between the sun and the horizon (see Figure 3). Now one can erect a unit vector at the origin of the coordinate system pointing at the light source,

$$\hat{\mathbf{n}}_s = [\sin(\theta)\cos(\phi), \cos(\theta)\cos(\phi), \sin(\phi)].$$

Since a surface element with gradient (p, q) has a normal vector $\mathbf{n} = (-p, -q, 1)$, we can identify a particular surface element that happens to be perpendicular to the direction towards the light source. Such a surface element will have a surface normal $\mathbf{n}_s = (-p_s, -q_s, 1)$, where $p_s = \sin(\theta)\cot(\phi)$ and $q_s = \cos(\theta)\cot(\phi)$. We can use the gradient (p_s, q_s) as an alternate means of specifying the position of the source (see Figure 3).

In work on automatic hill-shading, for example, one uses $p_s = -0.707$ and $q_s = 0.707$ to agree with standard cartographic conventions which require that the light

Fig. 3. Definition of azimuth and elevation of the sun.



source be in the northwest at 45° elevation [$\theta = (7/4)\pi$, $\phi = \pi/4$) [11].

Not all light reflected from the surface comes directly from the sun; some of it is scattered in the atmosphere. One could add a small component to the reflectance map to account for this and rather simple models of how much light a surface element captures from the general sky illumination would do. This was not done for the examples here since the effect is very small in the near infrared, as demonstrated by the very dark appearance of shadowed surface elements in bands 6 and 7 of LANDSAT images.

7. Reflectance as a Function of the Gradient

Reflectance of a surface can be expressed as a function of the incident angle (i), the emittance angle (e), and the phase angle (g) (see Figure 4). We use a simple, idealized reflectance model for the surface material,

$$\Phi_1(i, e, g) = \rho \cos(i).$$

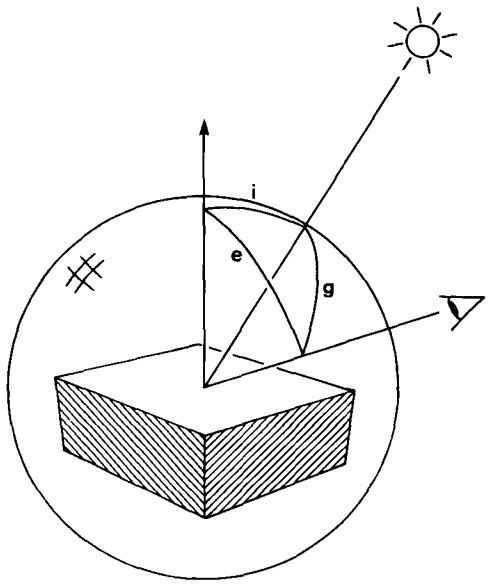
This reflectance function models a surface which, as a perfect diffuser, appears equally bright from all viewing directions. Here, ρ is an "albedo" factor and the cosine of the incident angle simply accounts for the foreshortening of the surface element as seen from the source.[1] It is not necessary for the reflectance to be a function of the incident angle only, in fact more sophisticated models of surface reflectance are possible [12], but are unnecessary for this application.

It is more convenient to express the reflectance as a function of the gradient (p, q). This is straightforward, since the phase angle g is constant [12]. The incident angle is the angle between the local normal ($-p$, $-q$, 1) and the direction to the light source ($-p_s$, $-q_s$, 1). The cosine of this angle can then be found by taking the dot-product of the corresponding unit vectors,

$$\cos(i) = \frac{(1 + p_s p + q_s q)}{\sqrt{1 + p_s^2 + q_s^2}\sqrt{1 + p^2 + q^2}}.$$

[1] "Albedo," for purposes of this paper, will simply be the ratio of reflectance of the surface to that of a perfectly diffuse surface, also called a Lambertian reflector.

Fig. 4. The geometry of light reflection from a surface element is governed by the incident angle, i, the emittance angle, e, and the phase angle, g.



Finally,

$$\Phi_1(p, q) = \frac{\rho(1 + p_s p + q_s q)}{\sqrt{1 + p_s^2 + q_s^2}\sqrt{1 + p^2 + q^2}}.$$

Another reflectance function, similar to that of materials in the maria of the moon and rocky planets [9, 6], is a little easier to calculate.

$$\Phi_2(p, q) = \rho \cos(i)/\cos(e) = \frac{\rho(1 + p_s p + q_s q)}{\sqrt{1 + p_s^2 + q_s^2}}.$$

This reflectance function models a surface which reflects equal amounts of light in all directions. For small slopes and low sun elevations, it is very much like the first one, since then $(1 + p^2 + q^2)$ will be near unity. Both functions were tried and both produce good alignment—in fact, it is difficult to distinguish synthetic images produced using these two reflectance functions.

8. Synthetic Images

Given the projection equations that relate points on the objects to images of said points, and given a terrain model allowing calculation of surface gradient, it is possible to predict how an image would appear under given illuminating conditions, provided the reflectance map is available. We assume simple orthographic projection here as appropriate for a distant spacecraft looking vertically down with a narrow angle of view. Perspective projection would require several changes in the algorithm. There would no longer be a simple relationship between points in the synthetic image and points in the surface model, for example, and some of the techniques used in computer graphics would be useful [4, 17,

Fig. 5a. Reflectance map used in the synthesis of Figure 1. The curves shown are contours of constant $\Phi_1(p,q)$ for $\rho = 1$.

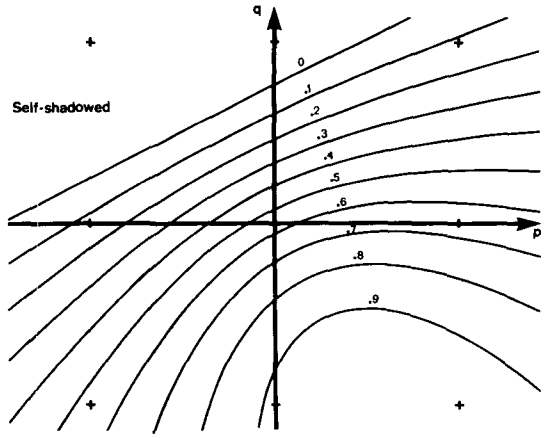


Fig. 5b. Reflectance map used in the synthesis of Figure 2.

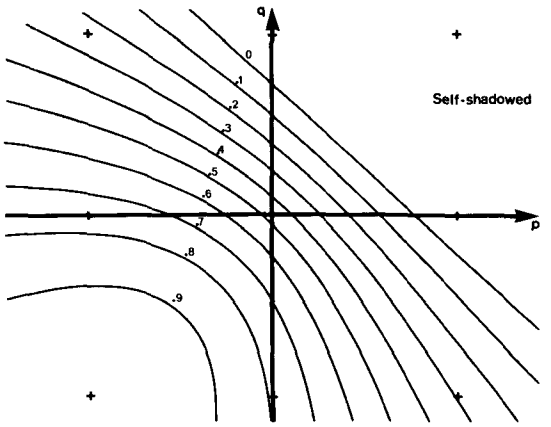


Fig. 6a. Alternate reflectance map, which could have been used in place of the one shown in Figure 5a. The curves shown are contours of constant $\Phi_2(p,q)$ for $\rho = 1$.

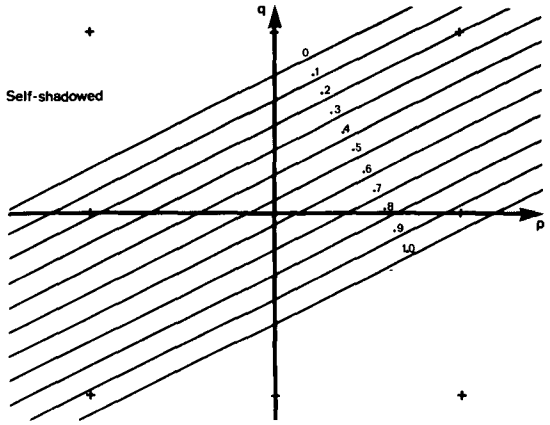
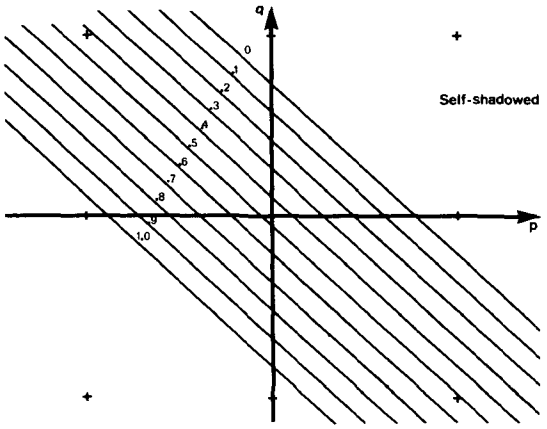


Fig. 6b. Alternate reflectance map that could have been used in synthesis of Figure 2.



2]. In the case of LANDSAT images, however, the departure from orthographic projection is very small and only in one direction due to the special nature of the scanning device. (This distortion along scan lines is easy to deal with if a surface model is available.)

The process of producing the synthetic image is simple. An estimate of the gradient is made for each point in the digital terrain model by considering neighboring elevations. The gradient's components, p and q, are then used to look up or calculate the expected reflectance. An appropriate intensity is placed in the image at the point determined by the projection equation. All computations are simple and local, and the work grows linearly with the number of picture cells in the synthetic image.

Sample synthetic images are shown in Figures 1 and 2. The two images are of the same region with differences in assumed location of the light source. In Figure 1 the sun is at an elevation of 34° and azimuth of 153°, corresponding to its true position at 9:55 G.M.T., 1972/Oct./9, while for Figure 2 it was at an elevation of 28° and an azimuth of 223°, corresponding to its position at 13:48 G.M.T. later on the same day. The corresponding reflectance maps are shown in Figure 5.

Reflectance maps for the simpler reflectance function $\Phi_2(p, q)$ under the same circumstances are shown in Figure 6. Note that near the origin there is very little difference between $\Phi_1(p, q)$ and $\Phi_2(p, q)$. Since most surface elements in this terrain model have slopes less than $1/\sqrt{2}$, synthetic images produced using these two reflectance maps are similar.

Since the elevation data are typically rather coarsely quantized as a result of the fixed contour interval on the original topographic map, p and q usually take on only a few discrete values. In this case, it is convenient to establish a lookup table for the reflectance map by simply precalculating the reflectance for these values. Models with arbitrarily complex reflectance functions can then be easily accommodated as can reflectance functions determined experimentally and known only for a discrete set of surface orientations.

Since the real image was somewhat smoothed in the process of being reproduced and digitized, we found it advantageous to perform a similar smoothing operation of the synthetic images so that the resolution of the two approximately matched. Alignment of real and synthetic images was, however, not dependent on this refinement.

9. Relationship to Work in Computer Graphics

Considerable progress has been made by researchers in computer graphics in their effort to produce synthetic images of structures defined by mathematical or numerical models [4, 17, 2]. One difference between their methods and those needed to generate the synthetic images used here arises from the relative simplicity of the satellite imaging situation. Here there are no hidden surfaces and thus no need for hidden-surface elimination [26, 27, 20]. Because of the near-orthographic projection, the computation is simple and does not require interpolation or successive refinement of surface patches. The resolution of the synthetic images can be easily matched to the resolution of the available surface model thus avoiding problems due to undersampling or aliasing.

The introduction of the reflectance map [12] permits important conceptual and computational advances. The computation of the local image intensity can proceed by table lookup no matter how complicated the reflectance function is. The reflectance map also provides a clear and easily interpreted visualization of the reflectance properties of a surface.

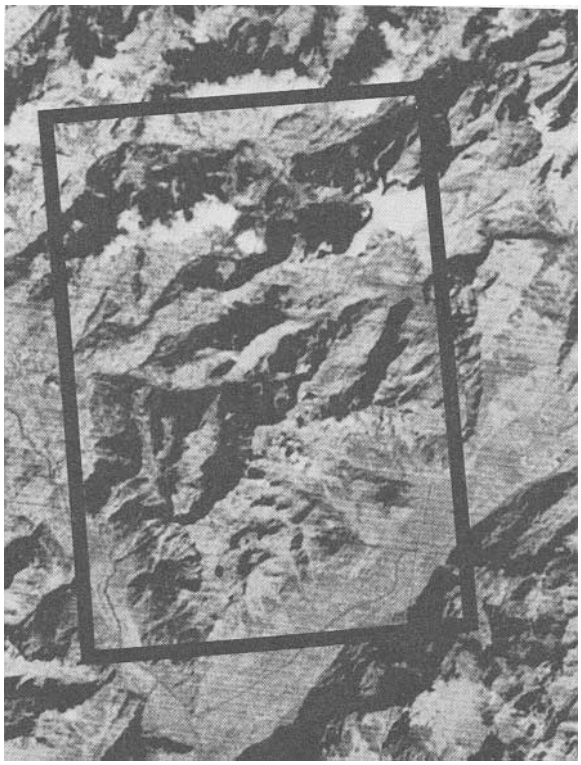
Finally, the ultimate purpose of image generation in the two situations is different. In one case, the images are intended to appear pleasing to a human observer. Here however they are to be matched against real images. This requires careful attention to the illumination model and the reflectance properties of *real* surfaces.

10. The Real Image

The image used for this paper's illustrations is a portion of a LANDSAT [1] image acquired about 9:55 G.M.T. 1972/October/9 (ERTS-1 1078-09555). Channel 6 (near infrared, 0.7μ to 0.8μ) was used, although all four channels appear suitable, with channel 4 (green, 0.5μ to 0.6μ) being most sensitive to moisture in the air column above the surface, and channel 7 (infrared, 0.8μ to 1.1μ) best able to penetrate thin layers of clouds and even snow. An enlargement of a transparency made from the original satellite image is shown in Figure 7. This should be compared with the synthetic image, generated from the digital terrain model, shown in Figure 1.

Note that the "footprint" of a LANDSAT picture cell (that is, the imaging systems instantaneous field of view) is about 79 x 79 meters [1], quite compatible with the resolution of the terrain model, 100 x 100 meters. The digitized image used was actually of somewhat lower resolution, however, due to limitations of the optics and electron-optics of our scanning system. Fortunately, alignment of images with terrain models is possible even with low quality image data. Further application of the aligned image and surface model information in such tasks as terrain classification however will require use of the raw sensor data, which is available on magnetic tape [1].

Fig. 7. Enlargement of the transparency containing the real image used in the alignment experiments. The region covered by the digital terrain model is shown outlined.



11. Transformation Parameters

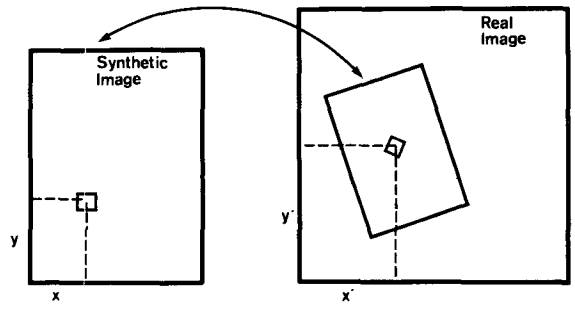
Before we can match the synthetic and the real image, we must determine the nature of the transformation between them. If the real image is truly an orthographic projection obtained by looking straight down, it is possible to describe this transformation as a combination of a translation, a rotation, and a scale change. If we use x and y to designate points in the synthetic image and x' and y' for points in the real image, we may write:

$$\begin{vmatrix} x' - x'_0 \\ y' - y'_0 \end{vmatrix} = s \begin{vmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{vmatrix} \begin{vmatrix} x - x_0 \\ y - y_0 \end{vmatrix} + \begin{vmatrix} \Delta x \\ \Delta y \end{vmatrix}$$

where Δx and Δy are the shifts in x' and y', respectively, θ is the angle of rotation, and s is the scale factor. Rotation and scaling take place relative to the centers (x_0, y_0) and (x'_0, y'_0) of the two images in order to better decouple the effects of rotation and scaling from translations. That is, the average shift in x' and y' induced by a change in rotation angle or scale is zero.

In our case, the available terrain model restricts the size of the synthetic image. The area over which matching of the two will be performed is thus always fixed by the border of the synthetic image. The geometry of the coordinate transformation is illustrated in Figure 8.

Fig. 8. Coordinate transformation from synthetic image to real image.



12. Choice of Similarity Measure

In order to determine the best set of transformation parameters (Δx, Δy, s, θ) one must be able to measure how closely the images match for a particular choice of parameter values. Let S_{ij} be the intensity of the synthetic image at the ith picture cell across in the jth row from the bottom of the image, and define R_{ij} similarly for the real image. Because of the nature of the coordinate transformation, we cannot expect that the point in the real image corresponding to the point (i, j) in the synthetic image will fall precisely on one of the picture cells. Consequently, S_{ij} will have to be compared with $R(x', y')$, which is interpolated from the array of real image intensities. Here (x', y') is obtained from (i, j) by the transformation described in the previous section.

One measure of difference between the two images may be obtained by summing the absolute values of differences over the whole array. Alternately, one might sum the squares of the differences:

$$\sum_{i=1}^{n}\sum_{j=1}^{m} \{S_{ij} - R(x', y')\}^2.$$

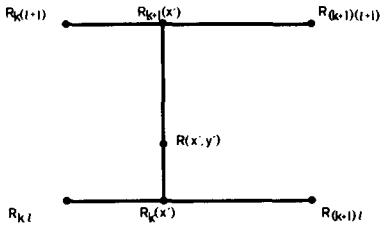
This measure will be *minimal* for exact alignment of the images. Expanding the square, one decomposes this result into three terms, the first being the sum of S_{ij}^2, the last the sum of $R^2(x', y')$. The first is constant, since we always use the full synthetic image; the last varies slowly as different regions of the real image are covered. The sum of $S_{ij}R(x', y')$ is interesting since this term varies most rapidly with changes in the transformation. In fact, a very useful measure of the similarity of the two images is the correlation:

$$\sum_{i=1}^{n}\sum_{j=1}^{m} S_{ij}R(x', y').$$

This measure will be *maximal* when the images are properly aligned. It has the advantage of being relatively insensitive to constant multiplying factors. These may arise in the real image due to changes in the adjustment of the optical or electronic systems.

Note that image intensity is the product of a constant factor which depends on the details of the imaging system (such as the lens opening and the focal length),

Fig. 9. Simple interpolation scheme applied to the real image array.



the intensity of the illumination striking the surface, and the reflectance of the surface. We assume all but the last factor is constant and thus speak interchangeably of changes in surface reflectance and changes in image intensities.

13. Interpolation Scheme

The real image intensity at the point (x', y') has to be estimated from the array of known image intensities. If we let $k = \lfloor x' \rfloor$, and $l = \lfloor y' \rfloor$ be the integer parts of x' and y', then $R(x', y')$ can be estimated from R_{kl}, $R_{(k+1)l}$, $R_{k(l+1)}$, and $R_{(k+1)(l+1)}$ by linear interpolation (see Figure 9).

$$R_l(x') = (k + 1 - x')R_{kl} + (x' - k)R_{(k+1)l}$$

$$R_{(l+1)}(x') = (k + 1 - x')R_{k(l+1)} + (x' - k)R_{(k+1)(l+1)}$$

$$R(x', y') = (l + 1 - y')R_l(x') + (y' - l)R_{(l+1)}(x').$$

The answer is independent of the order of interpolation and, in fact, corresponds to the result obtained by fitting a polynomial of the form $(a + bx' + cy' + dx'y')$ to the values at the four indicated points. Alignment was found not to be impaired, however, when nearest neighbor interpolation was used instead. This may be a result of the smoothing of the real image as previously described.

14. Choice of Normalization Method

High output may result as the transformation is changed simply because the region of the real image used happens to have a high average gray level. Spurious background slopes and false maxima may then result if the raw correlation is used. For this and other reasons, it is convenient to normalize. One approach essentially amounts to dividing each of the two images by its standard deviation; alternately, one can divide the raw correlation by

$$\sqrt{\sum_{i=1}^{n}\sum_{j=1}^{m} S_{ij}^2} \times \sqrt{\sum_{i=1}^{n}\sum_{j=1}^{m} R^2(x', y')}.$$

One additional advantage of this approach is that a perfect match of the two images now corresponds to a normalized correlation of one. An alternate method uses a normalization factor that is slightly easier to compute and which has certain advantages if the standard devia-

Fig. 10a. Variation of similarity measure with translation in x direction.

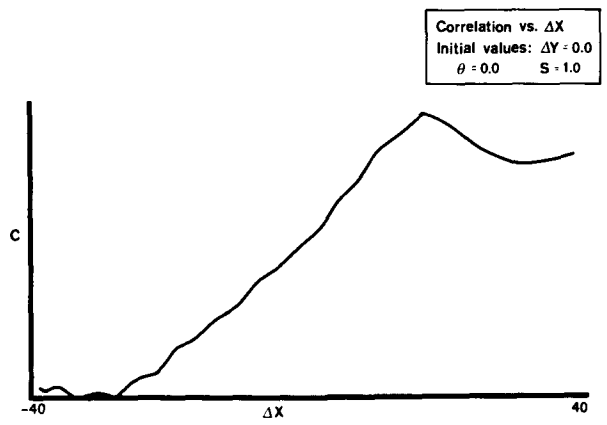


Fig. 10b. Variation of similarity measure with translation in y direction.

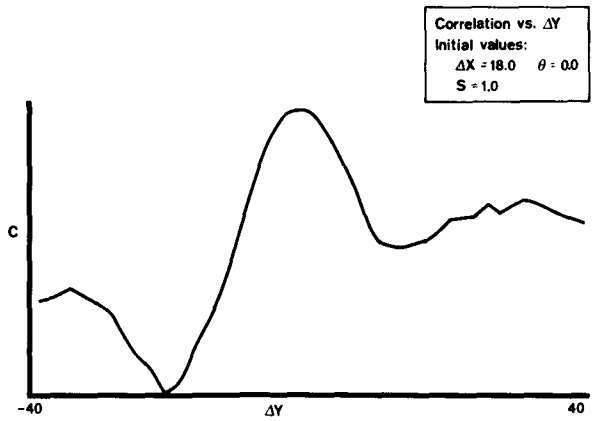


Fig. 10c. Variation of similarity measure with rotation.

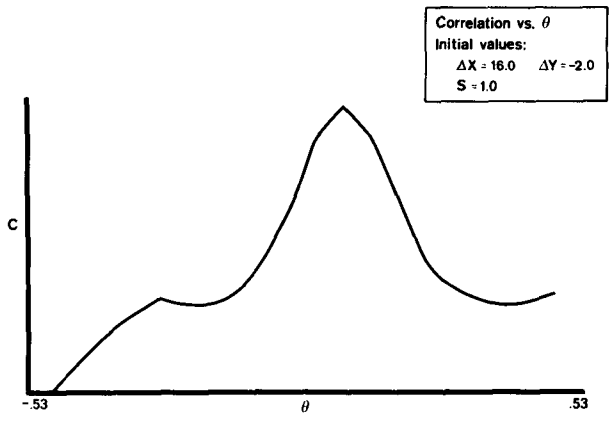
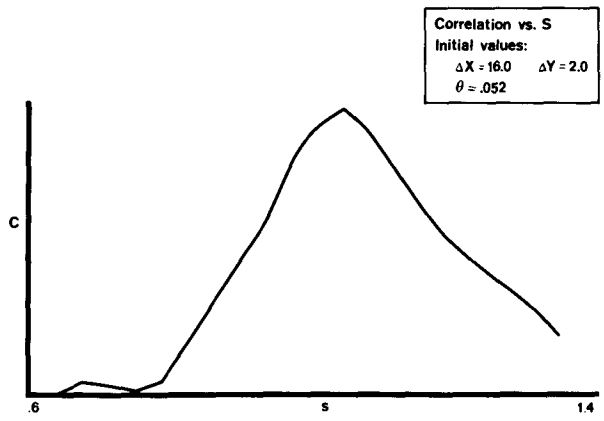


Fig. 10d. Variation of similarity measure with scale changes.



tions of the two images are similar. Instead of using the geometric mean, Hans Moravec proposes the arithmetic mean [13]:

$$\left[\sum_{i=1}^{n}\sum_{j=1}^{m} S_{ij}^2 + \sum_{i=1}^{n}\sum_{j=1}^{m} R^2(x', y')\right] \Big/ 2.$$

The first term need not be recomputed, since the full synthetic image is always used. Since we found the alignment procedure insensitive to the choice of normalization method, we used the second in our illustrations.

15. Locating the Best Match

Now that we have shown how to calculate a good similarity measure, we must describe an efficient method for finding the best possible transformation parameters. Exhaustive search is clearly out of the question. Fortunately, the similarity measure allows the use of standard hill-climbing techniques. This is because it tends to vary smoothly with changes in parameters and often is monotonic (at least for small ranges of the parameters).

When images are not seriously misaligned, profiles of the similarity measure usually are unimodal with a well-defined peak when plotted against one of the four parameters of the transformation (see Figure 10). It is possible to optimize each parameter in turn, using simple search techniques in one dimension. The process can then be iterated. A few passes of this process typically produce convergence. (More sophisticated schemes could reduce the amount of computation, but were not explored.)

When the images are initially *not* reasonably aligned, more care has to be taken to avoid being trapped by local maxima. Solving this problem using a more extensive search leads to prohibitively lengthy computations. A way of reducing the cost of comparing images had to be found.

16. Using Reduced Images

One way to reduce the computation is to use only subimages or "windows" extracted from the original images. This is useful for fine matching, but is not satisfactory here because of the lack of global context.

Alternately, one might use sampled images obtained by picking one image intensity to represent a small block

of image intensities. This is satisfactory as long as the original images are smoothed and do not have any high resolution features. If this is not the case, aliasing due to undersampling will produce images of poor quality unsuitable for comparisons.

One solution to this dilemma is to low-pass filter the images before sampling. A simple approximation to this process uses averages of small blocks of image intensities. The easiest method involves making one image intensity in the reduced image equal to the average of a 2×2 block of intensities in the original image. This technique can be applied repeatedly to produce ever smaller images and has been used in a number of other applications [13, 21].

The results of the application of this reduction process to real and synthetic images can be seen in Figure 11. First, the most highly reduced image is used to get coarse alignment. In this case, extensive search in the parameter space is permissible, since the number of picture cells in the images to be matched is very small. This coarse alignment is then refined using the next larger reduced images (with four times as many picture cells). Finally, the full resolution images are used directly to fine tune the alignment. False local maxima are, fortunately, much rarer with the highly reduced pictures, thus further speeding the search process. It is as if the high resolution features are the ones leading to false local maxima.

Quick convergence was obtained when translation was optimized before rotation and scale change. In each case the best values found so far for each parameter were used while searching for an optimum value in another parameter.

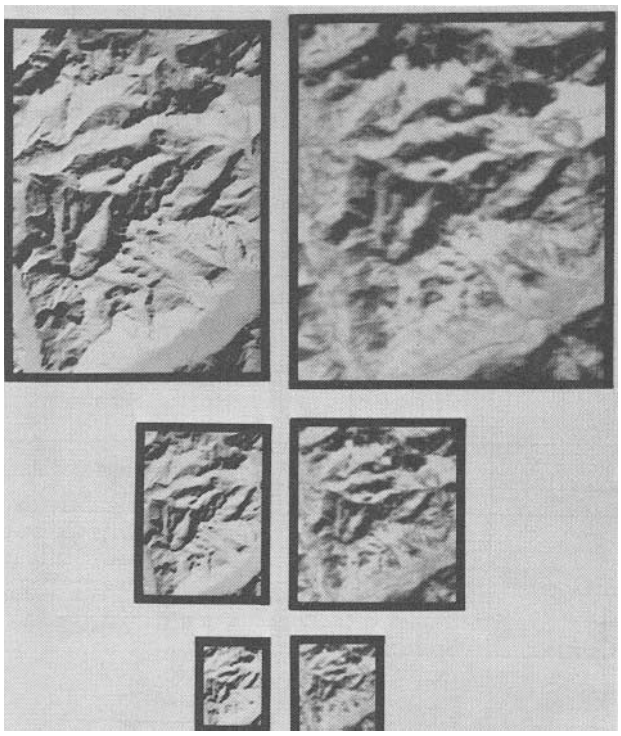
17. Results of Registration Experiments

We matched the real and synthetic images using the similarity measure and search technique just described. We tried several combinations of implementation details, and in all cases achieved alignment which corresponded to a high value of the normalized correlation, and which was very close to that determined manually. For the images shown here, the normalized correlation coefficient reaches 0.92 for optimum alignment, and the match is such that no features are more than two picture cells from the expected place, with almost all closer than one. (The major errors in position appear to be due to perspective distortion, with which the process is not now designed to cope.) The accuracy with which translation, rotation, and scaling were determined can be estimated from the above statement.

Overall, the process appears quite robust, even with degraded data. Details of interpolation, normalization, search technique, and even the reflectance map did not matter a great deal.

Having stated that *alignment* can be accurately achieved, we may now ask how similar the real and synthetic images are. There are a number of uninformative numerical ways of answering this question. Graphic illustrations, such as images of the differences between the real and synthetic image, are more easily understood. For example, we plot real image intensity versus synthetic image intensity in Figure 12. Although one might expect a straight line of slope one, the scattergram shows clusters of points, some near the expected line, some not.

Fig. 11. Successive reduction by factors of two applied to both the synthetic (left) and the real (right) image.



The clusters of points indicated by the arrow labeled A (Figure 12) corresponds chiefly to image points showing cloud or snow cover, with intensity sufficient to saturate the image digitizer. Here the real image intensity exceeds the synthetic image intensity. Arrow B indicates the cluster of points which corresponds to shadowed points. Those near the vertical axis and to its left come from self-shadowed surface elements, while those to the right are regions lying inside shadows cast by other portions of the surface. These cast shadows are not now simulated in the synthetic image. Here the synthetic image is brighter than the real image. Finally, the cluster of points indicated by arrow C arises from the valley floor, which covers a fairly large area and has essentially zero gradient. As a result, the synthetic image has constant intensity here, while the real image shows both darker features (such as the river) and brighter ones (such as those due to the cities and vegetation cover). Most of the ground cover in the valley appears to have higher albedo than the bare rock which is exposed in the higher regions, as suggested by the position of this cluster *above* the line of slope one.

If we were to remove these three clusters of points,

Fig. 12a. Scattergram of real image intensities versus synthetic image intensities based on $\Phi_1(p,q)$.

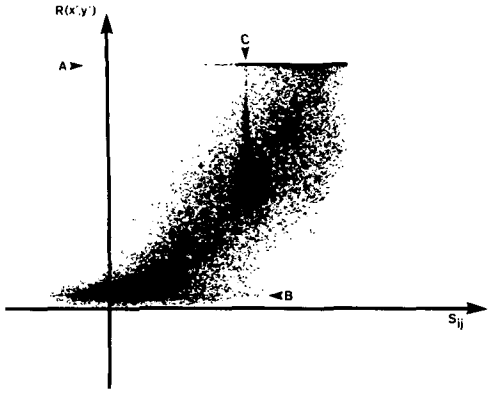
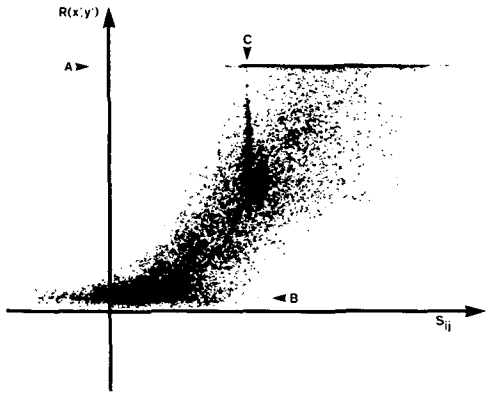


Fig. 12b. Scattergram of real image intensities versus synthetic image intensities based on $\Phi_2(p,q)$.



the remainder would form one elongated cluster with major axis at about 45°. This shows that, while there may not be an accurate point-by-point equality of intensities, there is a high correlation between intensities in the real and synthetic images.

Note, by the way, that no quantization of intensity is apparent in these scattergrams. This is a result of the smoothing applied to the synthetic image and the interpolation used on the real image. Without smoothing, the synthetic image has fairly coarse quantization levels because of the coarse quantization of elevations as indicated earlier. Without interpolation, the real image, too, has fairly coarse quantization due to the image digitization procedure.

Finally, note that we achieve our goal of obtaining accurate alignment. Detailed *matching* of synthetic and real image intensity is a new problem which can be approached now that the problem of image *registration* has been solved.

18. The Influence of Sun Elevation

Aerial or satellite photographs obtained when the sun is low in the sky show the surface topography most clearly. In this case, the scene radiance is the major factor in determining surface reflectance. Ridges and valleys stand out in stark relief, and one gets an immediate impression of the shape of portions of the surface. Conversely, variations in surface cover tend to be most important when the sun is high in the sky. Photographs obtained under such conditions are difficult to align with a topographic map—at least for a beginner.

What is the sun elevation for which these two effects are about equally important? Finding this value will allow us to separate the imaging situations into two classes: those which are more suited for determining topography and those which are more conducive to terrain classification success. We will use a simple model of surface reflectance. Suppose that the surface has materials varying in albedo between ρ_1 and ρ_2. Next, suppose that the surface slopes are all less than or equal to $\tan(e)$. The incident angles will vary between $e - (90° - \phi)$ and $e + (90° - \phi)$, where ϕ is the elevation of the sun. If we use the same simple reflectance function employed before, we find that for the two influences on reflectance to be just equal:

$$\rho_1 \cos(e + 90° - \phi) = \rho_2 \cos(e - 90° + \phi).$$

Expanding the cosine and rearranging this equation leads to:

$$\tan(\phi) = \left| \frac{\rho_1 + \rho_2}{\rho_1 - \rho_2} \right| \tan(e).$$

When, for example, the surface materials have reflectances covering a range of two to one and the sun elevation is 35°, then regions with surface slopes above approximately 0.23 ($e \simeq 13°$) will have image intensities affected more by surface gradient than by surface cover. Conversely, flatter surfaces will result in images more affected by variations in surface cover than by the area's topography.

One conclusion is that alignment of images with terrain models is feasible *without* detailed knowledge of the surface materials if the sun elevation is small and the surface slopes are high.

19. Summary and Conclusions

We have seen that real images can be aligned with surface models using synthetic images as an intermediary. This process works well despite many factors which contribute to differences between the real and synthetic images. The computations, while lengthy, are straightforward, and reduced images have been used to speed up the search for the best set of transformation parameters.

Several applications of aligned images and surface information have been presented. More can be found; for problems in a different domain, see [15], for example. Aside from change detection, passive navigation, photo-interpretation, and inspection of industrial parts, perhaps

the most important application lies in the area of terrain classification.

So far, no account has been taken of the effect of varying surface gradient, sun position, and reflective properties of ground cover. Recently, some interest has arisen in an understanding of how surface layers reflect light [18, 28, 16] and how this understanding might aid the interpretation of satellite imagery [23, 22, 5].

It is imperative that interpretation of image information be guided by an understanding of the imaging process. This, in turn, can be achieved if one understands how light is reflected from various surfaces and how this might be affected by such factors as light source position, moisture content, and the point in the growth cycle of vegetation.

Acknowledgments. We would like to thank Kurt Brassel, Thomas Peucker, N. Balasubramanian, George E. Lukes, and Harry Barrow for generously supplying digital terrain models. We would also like to express our appreciation to Blenda Horn, Eva Kampits, Karen Prendergast, Patrick Winston, Robert Woodham, and the editor, who provided numerous helpful suggestions and improved the presentation of the results.

Received Septemper 1977; revised March 1978

References

1. Bernstein, R. Digital image processing of earth observation sensor data. *IBM J. Res. and Develop. 20* (Jan. 1976), 40–57.
2. Blinn, J.F., and Newell, M.E. Texture and reflection in computer generated images. *Comm. ACM 19*, 10 (Oct. 1976), 542–547.
3. Brassel, K. *Modelle und Versuche zur automatischen Schraglichtschattierung,* Buchdruckerei E. Brassel, Klosters, Switzerland, 1973.
4. Catmull, E.A. Computer display of curved surfaces. *Proc. Conf. Comptr. Graphics, Pattern Recog. and Data Struct.* May 1975, pp. 11–17 (available from ACM, New York).
5. Duggin, M. J. Likely effects of solar elevation on the quantification of changes in vegetation with maturity using sequential LANDSAT images. *Appl. Optics 16* (March 1977), 521–523.
6. Hapke, B. An improved theoretical lunar photometric function. *Astronomical J. 71* (June 1966), 333–339.
7. Explanatory Supplement to the Astronomical Ephemeris and the American Ephemeris and Nautical Almanac. Her Majesty's Stationery Office, London, England, 1961.
8. Horn, B.K.P. Shape from shading: A method for obtaining the shape of a smooth opaque object from one view. MAC-TR-79, M.I.T., Cambridge, Mass., Nov. 1970.
9. Horn, B.K.P. Obtaining shape from shading information. In *The Psychology of Computer Vision,* P. H. Winston, Ed., McGraw-Hill, New York, 1975, pp. 115–155.
10. Horn, B.K.P. Celestial navigation suite for a programmable calculator. Unpublished 1976.
11. Horn, B.K.P. Automatic hill-shading using the reflectance map. Unpublished (1976).
12. Horn, B.K.P. Understanding image intensity. *Artif. Intell. 8* (April 1977), 201–231.
13. Moravec, H.P. Towards automatic visual obstacle avoidance. *Proc. Fifth Int. Conf. on Artif. Intell.,* Cambridge, Mass., 1977, 584.
14. Newcomb, S. *A Compendium of Spherical Astronomy.* Macmillan, New York, 1895, 1906.
15. Nitzan, D., Brain, A.E., and Duda, R.O. The measurement and use of registered reflectance and range data in scene analysis. *Proc. IEEE 65* (Feb. 1977), 206–220.
16. Oetking, P. Photometric studies of diffusely reflecting surfaces with applications to the brightness of the moon. *J. Geophysical Res. 71* (May 1966), 2505.
17. Phong, B.-T. Illumination for computer generated images. *Comm. ACM 18*, 6 (June 1975), 311–317.
18. Simmons, E.L. Particle model theory of diffuse reflectance: Effect of non-uniform particle size. *Appl. Optics 15* (1976), 603–604.
19. Smart, W.M. *Textbook on Spherical Astronomy.* Cambridge U. Press, Cambridge, England, 1931, 1965.
20. Sutherland, I.E., Sproull, R.F., and Schumacker, R.A. A characterization of ten hidden surface algorithms. *Comptng. Surveys 6*, 1 (March 1974), 1–55.
21. Tanimoto, S.L. Pictorial feature distortion in a pyramid. *Computer Graphics and Image Processing 5* (Sept. 1976), 333–352.
22. Tucker, C.J. Asymptotic nature of grass canopy spectral reflectance. *Appl. Optics 16* (May 1977), 1151–1156.
23. Tucker, C.J., and Garret, M.W. Leaf optical system modeled as a stochastic process. *Appl. Optics 16* (March 1977), 635–642.
24. The American Ephemeris and Nautical Almanac for the year 1972. U.S. Govt. Printing Office, Washington, D.C.
25. The Nautical Almanac for the year 1972. U.S. Govt. Printing Office, Washington, D.C.
26. Warnock, J.E. A hidden-line algorithm for half-tone picture representation. Rep. TR 4-15, Dept. Comptr. Sci., U. of Utah, Salt Lake City, Utah, 1969.
27. Watkins, G.S. A real-time visible surface algorithm. Tech. Rep. UTEC-CSC-70-101, Dept. Comptr. Sci., U. of Utah, Salt Lake City, Utah, 1970.
28. Wendlandt, W.W.M., and Hecht, H.G. *Reflectance Spectroscopy.* Interscience Publishers, New York, 1966.
29. Woolard, E.W., and Clemance, G.M. *Spherical Astronomy.* Cambridge U. Press, Cambridge, England, 1966.