Analysis by synthesis in computational vision with application to remote sensing

R.J. Woodham,[1] E. Catanzariti,[2] and A.K. Mackworth[1]

Laboratory for Computational Vision, Department of Computer Science, University of British Columbia,
Vancouver, B.C., Canada V6T 1W5

The central problem in vision is to determine scene properties from image properties. This is difficult because the problem, formally posed, is underconstrained. Methods that infer scene properties from images make assumptions about how the world determines what we see. In remote sensing, some of these assumptions can be dealt with explicitly. Available scene knowledge, in the form of a digital terrain model and a ground cover map, is used to synthesize an image for a given date and time. The synthesis process assumes that the surface is a perfectly diffuse or "lambertian" reflector. A scene radiance equation is described based on simple models of direct solar irradiance, diffuse sky irradiance, and atmospheric path radiance. Parameters of the model are estimated from the real image. A statistical comparison of the real image and the synthetic image is used to judge how well the model represents the mapping from scene to image.

The methods presented for image synthesis are similar to those used in computer graphics. The motivation, however, is different. In graphics, the goal is to produce an effective rendering of the scene for a human viewer. Here, the goal is to predict properties of real images. In vision, one must deal with a confounding of effects due to surface shape, surface material, illumination, shadows, and atmosphere. These effects often detract from, rather than enhance, the determination of invariant scene characteristics.

Key words: computational vision, reflectance, remote sensing, computer graphics, geographic information systems, image analysis, image synthesis, surface representation.

Le problème central en vision est de déterminer des propriétés scéniques à partir de propriétés d'images. Ceci est difficile, car le problème, posé de façon formelle, n'est pas suffisamment limité. Les méthodes qui infèrent des propriétés scéniques à partir d'images font des hypothèses sur la façon dont le monde détermine ce que nous voyons. En perception à distance, certaines de ces hypothèses peuvent être traitées explicitement. Une connaissance de scène disponible, sous forme d'un modèle numérique de terrain et d'un plan de surface horizontal, est utilisé pour synthétiser une image pour une date et une heure données. Le processus de synthèse suppose que la surface est un réflecteur diffusant parfaitement, i.e., "lambertien." Une équation de la radiance de la scène est décrite en se basant sur des modèles simples de la radiance solaire directe, de la radiance diffuse du ciel, et de la radiance du parcours atmosphérique. Les paramètres du modèle sont estimés à partir de l'image réelle. Une comparaison statistique de l'image réelle et de l'image synthétique est utilisée pour estimer à quel point le modèle représente le mapping de la scène à l'image.

Les méthodes présentées pour la synthèse d'image sont similaires à celles utilisées dans les représentations graphiques par ordinateur. La motivation, néanmoins, est différente. En dessin, le but est de produire une représentation efficace de la scène pour un observateur humain. Ici, le but est de prédire les propriétés d'images réelles. En vision, on doit traiter d'un enchevêtrement d'effets dû à la forme de la surface, au matériau dont elle est constituée, à l'éclairage, aux ombres, et à l'atmosphère. Ces effets portent souvent atteinte à la détermination de caractéristiques scéniques invariantes plutôt que de mettre en valeur celles-ci.

Mots clés: vision par ordinateur, réfraction, perception à distance, images par ordinateur, systèmes d'information géographique, analyse d'image, synthèse d'image, représentation de surface.

[Traduit par la revue]

1. Introduction

Computational vision is the study of systems that produce symbolic descriptions of a world from images of that world. The computation from signal input to symbolic output is too complex to be treated as a single function. Vision requires many levels of intermediate representation. Identifying those levels and establishing the constraints that operate both within and between levels is the fundamental task of computational vision research (Woodham 1982; Mackworth 1983). In early vision, one deals with descriptions that can be computed directly from the image. Analysis by synthesis is one way that knowledge of the scene can be used in early vision.

Knowledge of the scene is expressed as a scene radiance equation. Scene knowledge includes the illumination and the intrinsic reflectance properties of the surfaces in view. Image synthesis is used to predict how specific objects will look when

viewed in a particular way. The process iterates since comparison of the real and synthetic image contributes to an emerging description of the particular scene in view. In a synthesis approach to analysis, the image domain itself becomes the unifying representation to compare the scene against what is seen. This is a simple approach. Nevertheless, it is shown to be effective when applied to remote sensing.

In remote sensing, one attempts to interpret multispectral scanner data directly in terms of ground cover. The effects of topography and atmosphere make this direct interpretation difficult. In industrial inspection, one attempts to interpret image data directly in terms of object shape. This is possible when objects have uniform optical properties (Horn 1977; Ikeuchi and Horn 1981; Woodham 1981). Separating changes in image intensity due to object shape from changes due to surface material is difficult, because trade-offs emerge that cannot be resolved locally. Nevertheless, in remote sensing, there have been attempts to extract both topographic and ground cover information from single multispectral images (Eliason *et al.* 1981; Wang *et al.* 1984). Other work uses terrain models in

an attempt to account for the dependence of scene radiance on topography and atmosphere (Smith *et al.* 1980; Justice *et al.* 1981; Shibata *et al.* 1981; Teillet *et al.* 1982; Sjoberg and Horn 1983; Woodham and Lee 1985).

Image synthesis can be thought of as a problem in computer graphics. In graphics, however, the goal is to produce an effective rendering of the scene for a human viewer. For example, image synthesis is used for hill shading in cartography (Horn 1981). But, many factors, such as shadows, skylight, and atmospheric attenuation, detract from the cartographic rendering of terrain and are not included. In addition, part of the cartographer's craft is to adjust the position of the light source locally to provide the best rendering of the terrain. Humans appear not to notice that the result is globally inconsistent and could not possibly correspond to the real lit world.

In analysis by synthesis the goal is to predict properties of real images. In remote sensing, one must deal with a confounding of effects due to surface shape, surface material, illumination, shadows, and atmosphere. These effects often hinder, rather than enhance, the determination of invariant scene characteristics.

2. Formulation

Image synthesis requires a scene radiance equation. To specify scene radiance, it is necessary to consider both geometry and radiometry.

2.1. Geometry

Consider a Cartesian coordinate system defined in terms of a plane at the earth's surface with the X axis pointing east, the Y axis pointing north and the Z axis up. A terrain model can be written as a function $z = f(x, y)$, where z is surface elevation. When a terrain model is given as a discrete array of elevations on a regular grid, it is called a digital terrain model (DTM).

For vertical imagery acquired from satellites such as Landsat 1, 2, and 3 (nominal altitude 900 km), image projection is essentially orthographic. That is, surface point (x, y, z) projects to image point (x, y).

The direction of a surface normal at a terrain point can be found by taking the cross product of any two vectors lying in the tangent plane, provided they are not parallel to each other. Two such vectors are $[1, 0, p]$ and $[0, 1, q]$, where $p = \partial f(x, y)/\partial x$ is the slope in the west to east direction and $q = \partial f(x, y)/\partial y$ is the slope in the south to north direction. Their cross product is the vector $[-p, -q, 1]$. The quantity (p, q) is called the *gradient* and is used to specify direction in the earth-centered coordinate system.

Four parameters are required to specify the local geometry of the incident and the reflected rays. Often, however, one considers materials whose reflectance characteristics are invariant with respect to rotations about the surface normal. For surfaces that are isotropic in this way, only three parameters are required. Figure 1 illustrates how to define the incident- and reflected-ray geometry in terms of three angles i, e, and g. The incident angle i is the angle between the incident ray and the surface normal. The emergent angle e is the angle between the reflected ray and the surface normal. The phase angle g is the angle between the incident and reflected rays. The use of angles i, e, and g has one advantage over other possibilities. For a distant viewer and distant light source, the phase angle g is constant, independent of the surface normal.

One can determine cos i and cos e, given the surface gradient (p, q). Let the incident-ray direction have gradient (p_0, q_0). That is, the vector $[-p_0, -q_0, 1]$ points in the direction of the
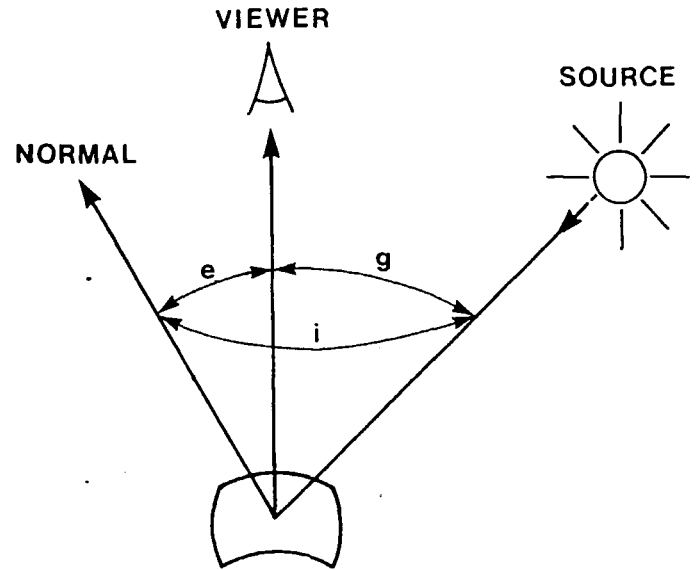

FIG. 1. The three angles i, e, and g used to specify the local geometry of the incident and the reflected ray.

light source. For nadir looking sensors, the vector $[0, 0, 1]$ points in the direction of the viewer. Expressing the cosine of the angle between two vectors as a normalized dot product of the vectors, one obtains

$$[1] \qquad \cos i = \frac{1 + pp_0 + qq_0}{\sqrt{1 + p^2 + q^2}\sqrt{1 + p_0^2 + q_0^2}}$$

$$[2] \qquad \cos e = \frac{1}{\sqrt{1 + p^2 + q^2}}$$

2.2. Radiometry

The reflectance properties of a surface material are determined by its bidirectional reflectance distribution function (BRDF). The BRDF, denoted by the symbol f_r, was introduced by Nicodemus *et al.* (1977) as a unified notation for the specification of reflectance in terms of the incident- and the reflected-beam geometry. The BRDF is an intrinsic property of a surface material. It determines how bright the surface will appear when viewed from a given direction and illuminated from another.

The BRDF allows one to determine scene radiance, L_r, for any defined incident- and reflected-beam geometry by integrating over the specified solid angles. A systematic approach to perform this integration has already been given (Horn and Sjoberg 1979). The approach has been applied to BRDF's proposed for remote sensing (Woodham and Lee 1985). Here, results are summarized for nadir viewing of perfectly diffuse surfaces under different conditions of illumination. (Perfectly diffuse surfaces are commonly referred to as lambertian surfaces (Nicodemus *et al.* 1977), a convention that will be followed throughout.)

A lambertian surface has BRDF

$$[3] \qquad f_r = \rho/\pi$$

where ρ is the bihemispherical reflectance, loosely termed *albedo*, that determines the proportion of incident light reflected by the surface. For an ideal (lossless) lambertian surface, $\rho = 1$.

When illuminated by a collimated source with irradiance E_0 measured perpendicular to the beam of light arriving from direction with gradient (p_0, q_0), one obtains, for points not in

relatively high spatial frequencies and to have sustained temporal characteristics. Conversely, the motion-detection channel was said to be sensitive to lower spatial frequencies and to have transient temporal characteristics. Since the spatial characteristics of the form channel should not change with temporal frequencies, it is natural to assume separability.

4. Evidence for inseparability

Although the use of the DOG is remarkably widespread and appears to provide a reasonable account of spatial behaviour for static stimuli, the assumption of separability is extremely restrictive. Given that the centre and the surround are computed by separate mechanisms and that it is improbable that two different neural circuits have identical temporal properties, inseparability is a natural consequence. Below, two bodies of evidence suggesting the degree of spatiotemporal inseparability are briefly reviewed (more detail can be found in Fleet (1984)). The first suggests that the centre and the surround do not have identical time courses of response, with the surround sluggish and delayed relative to the centre. The second shows inseparable amplitude and phase spectra of responses to drifting sinusoidal gratings.

4.1. Longer surround time course

Figure 1 shows a typical response pattern produced by a bipolar cell or an X-type ganglion cell when a circular disc is turned on and then off in the centre of its receptive field. Of particular interest is the "transient peak" apparent at the onset and offset of the stimulus. This transient peak is widely attributed to the surround having a longer time course of response relative to the centre, with the initial rise time indicative of the centre response, and the slower decay back to the sustained response indicative of the surround rise time and antagonism (Rodieck 1965; Kaneko 1970; Werblin 1974). In general, there is considerable electrophysiological evidence for a centre-surround latency difference at the mammalian ganglion-cell and bipolar-cell levels (there are similar results for lower animals). The longer surround time course has been attributed to a combination of fixed delay and slower rise time for the surround (Rodieck and Stone 1965; Richter and Ullman 1982; Fleet 1984; Fleet et al. 1985).

4.2. Inseparable amplitude and phase spectra

Recent physiological experiments have shown that spatial-frequency tuning does indeed change significantly with temporal frequency. In fact, the amplitude spectrum is spheroidal and radially band-pass at moderate mean light levels. Figure 2 shows data reported by Derrington and Lennie (1982) for an X cell and a Y cell stimulated with drifting sinusoidal gratings. In each case the spatial-frequency sensitivity profile is shown for a variety of temporal frequencies. Notice that spatial tuning changes from roughly band-pass at low temporal frequencies to low-pass at higher temporal frequencies. Enroth-Cugell et al. (1983) obtained similar results for X cells as well as for a subgroup of the W-type ganglion-cell population. Dawis et al. (1984) showed inseparability in both phase and amplitude components of responses, confirming the earlier results.

The spatiotemporal inseparability exhibited by these units is pronounced. Physiological evidence shows that the centre and surround do not have identical temporal properties. This is supported by sine-wave grating studies which report changes in spatial-frequency tuning with temporal frequency. Therefore, the DOG model is not valid for time-varying images. Although inseparability has been observed and included in several recent
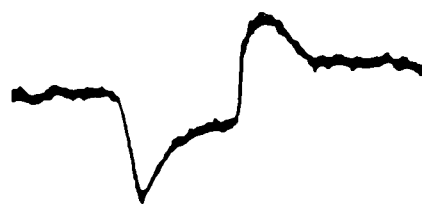

FIG. 1. The response profile of a goldfish bipolar cell, for a stimulus in which a disc is turned on and then off, exhibits a faster rise time than decay time. The amplitude of response is plotted on the vertical axis and time on the horizontal axis (from Kaneko (1970)).

modelling studies (Derrington and Lennie 1982; Richter and Ullman 1982; Enroth-Cugell et al. 1983; Dawis et al. 1984), its functional significance has not been fully explored in the literature and is discussed further below.

5. Spatiotemporal centre-surround (CS) model

To accommodate the inseparable behaviour discussed above, the conventional DOG model has been extended to include physiologically relevant temporal mechanisms. The spatiotemporal centre-surround (CS) model, first proposed by Richter and Ullman (1982) and further examined by Fleet (1984) and Fleet et al. (1985), includes exponential low-pass temporal filters for the centre and the surround. In addition, the surround is delayed relative to the centre and is assumed to have a longer rise time. The impulse-response function (receptive field) for the inseparable CS model is given by

$$[4] \qquad CS(x, t) = K(t; \lambda_c)G(x; \sigma_c) - K(t - d; \lambda_s)G(x; \sigma_s)$$

where $G(x, \sigma)$ is the 2-d spatial Gaussian [2], $d \geq 0$ is the surround delay, and

$$[5] \qquad K(t; \lambda) = \begin{cases} \lambda e^{-\lambda t} & \text{for } t \geq 0, \lambda > 0 \\ 0 & \text{for } t < 0 \end{cases}$$

is the temporal exponential with rise time λ^{-1}. Essentially, the three functionally significant parameters that affect the form of the model are (1) the ratio of the spatial standard deviations, σ_s/σ_c; (2) the ratio of the time constants, $\lambda_s^{-1}/\lambda_c^{-1}$; and (3) the surround delay, d. The standard deviation of the centre, σ_c, and the time constant of the centre, λ_c^{-1}, are free to be set according to the spatial and temporal frequencies the system is required to discern and are restricted in a discrete environment only by spatial and temporal sampling rates.

6. Response behaviour of the CS filter

The behaviour of the CS model has been analyzed in two ways. First, we consider the response patterns produced by the CS filter [4] in response to a variety of spatiotemporal inputs. This is an extension of the analysis done by Richter and Ullman (1982). Secondly, we apply Fourier analysis and consider the form of the amplitude spectrum with parameter values of physiological interest. In particular we find that frequency analysis allows us to compare the qualitative behaviour of the CS model with physiological observations and contrast the CS model with the DOG model.

6.1. Response patterns

Figure 3a shows the superposition of response patterns produced by a mudpuppy bipolar cell when presented (individually) with different-sized discs flashed on and then off in the centre of its receptive field. The smaller disc (giving the lowest steady state) is approximately the size of the central excitatory

FIG. 2. The digital terrain model (DTM) of the study site. Here, brightness is directly proportional to elevation.
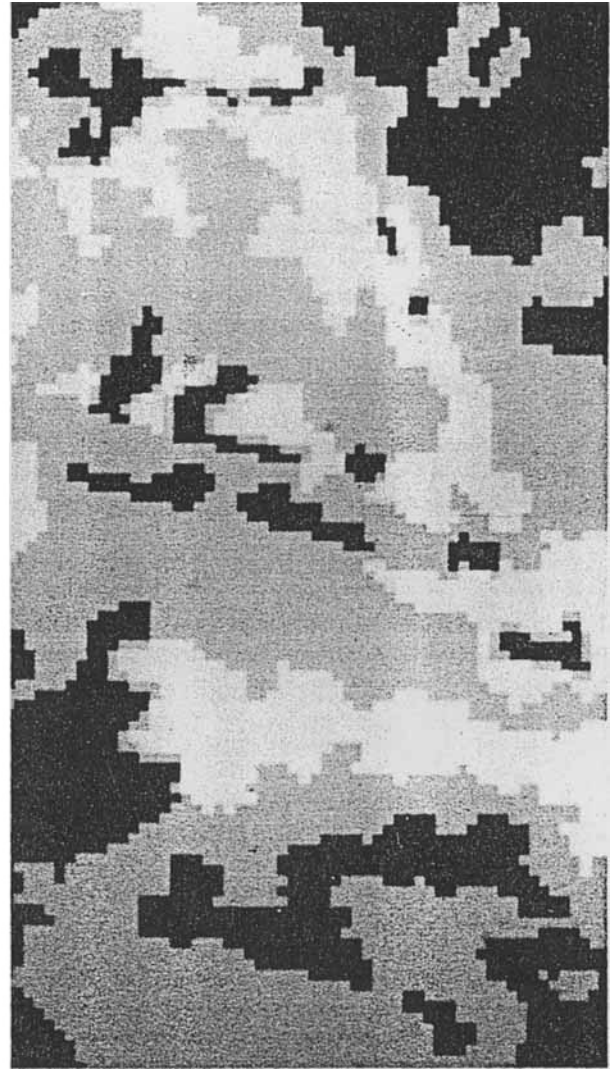

FIG. 3. The ground cover map of the study site. Old growth (class 1) is shown as light gray. Second growth (class 2) is shown as white. Recent logging (class 3) is shown as dark gray. A small amount of water (class 4) also is present and is shown as black.

Island, British Columbia, centred at geographic coordinates W 123:50:00, N 48:35:30. Elevation data along ridges and channels was manually digitized from the 1 : 50 000 Canadian National Topographic System (NTS) map sheet 92 B/12 (Shawnigan Lake). The ridge and channel structure was represented initially using a triangulated irregular network (TIN) (Peucker *et al.* 1978). Two grid representations were produced from the TIN. A 263 × 346 DTM (60 m grid) was used for the precise geometric rectification of Landsat images. A smaller scale 131 × 173 DTM (120 m grid) was generated for the work described below. Grid coordinates in the DTM correspond to the Universal Transverse Mercator (UTM) map projection used in Canadian NTS maps. The DTM is shown in Fig. 2. Within the study site, elevations vary from 150 to 850 m above sea level.

A forest cover map for a 7.0 × 13.0 km^2 area within the test site was available from previous studies (Catanzariti and Mackworth 1978; Starr and Mackworth 1978). The seven original ground cover classes were reduced to four broad classes: old growth (class 1), second growth (class 2), recent logging (class 3), and water (class 4). Ground cover classes were also represented in grid form. When positions in the forest

cover map are given in UTM map coordinates, no additional rectification of map to image or DTM is required. A 58 × 108 (120 m grid) grid representation was used to combine Landsat, DTM, and forest cover data. The ground cover map is shown in Fig. 3. Within the study site, 53.9% of the area is old growth, 19.8% is second growth, 26.0% is recent logging, and 0.3% is water.

3.2. The Landsat images

Landsat-1 was the first in a series of remote sensing satellites launched by NASA beginning in 1972. The operational sensor on Landsat-1 was a multispectral scanner (MSS) with four spectral bands. Band 4 (0.5–0.6 μm) is in the visible green. Band 5 (0.6–0.7 μm) is in the visible red. Band 6 (0.7–0.8 μm) and band 7 (0.8–1.1 μm) are in the near infrared. Sensor outputs are digitized on board the satellite and transmitted to ground receiving stations for further processing and distribution. Landsat MSS bands can be displayed in a variety of ways. A standard false colour composite is produced by displaying band 4 as blue, band 5 as green, and band 7 as red. This corre-
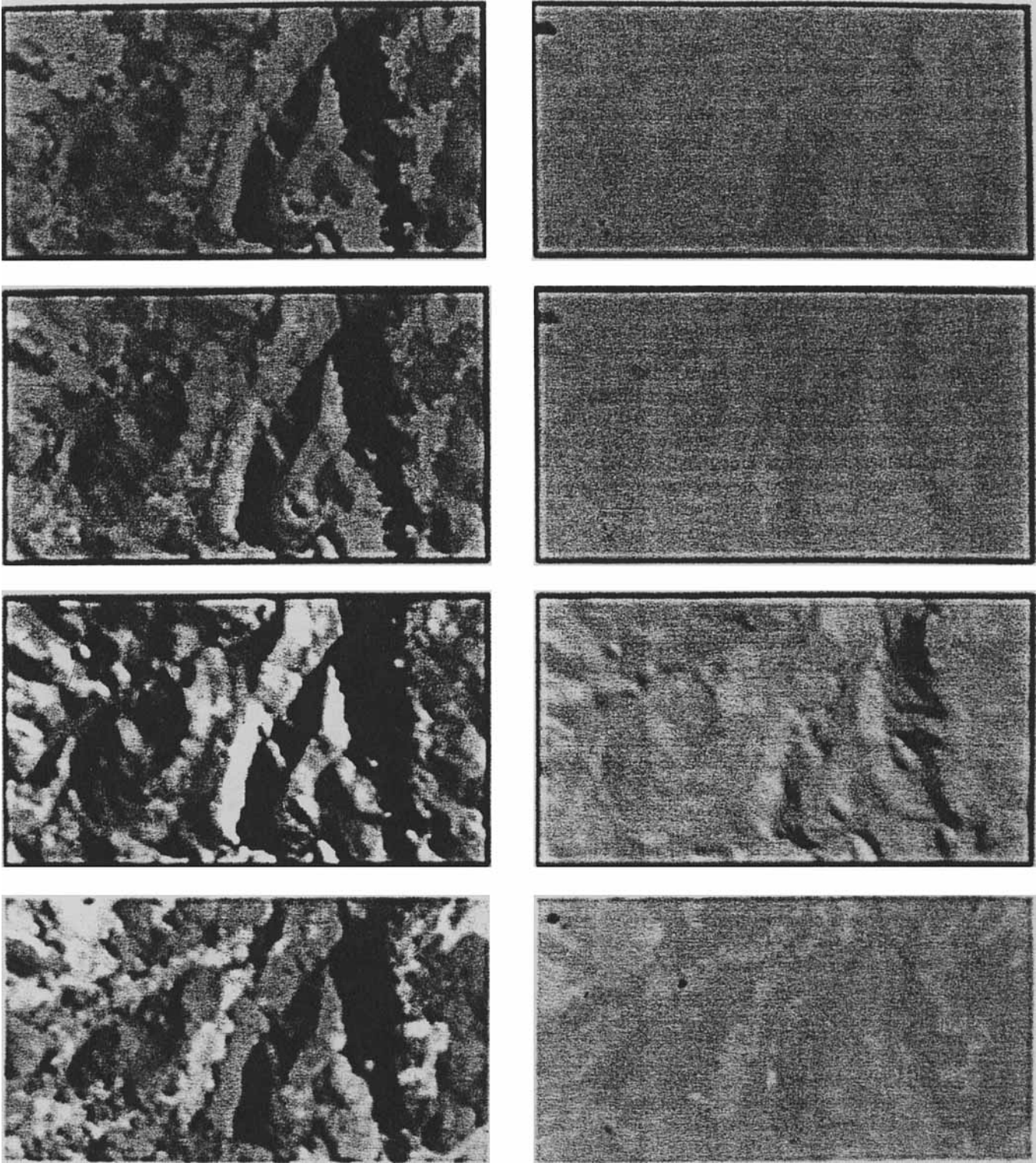
FIG. 4. The eight false colour composite images. The top row is for January 8, 1973. The bottom row is for August 12, 1973. The left-most column shows the real Landsat MSS data. The remaining three columns from left to right show the synthetic images generated by Methods B, E, and H described in the text.

sponds to the dyes used in colour infrared aerial film and produces a product that is familiar to photointerpreters.

Two Landsat MSS images were chosen to correspond closely in time to the date of preparation of the ground truth map. The first, a winter image, was acquired January 8, 1973 (Landsat frame-id 1169–18373). The second, a summer image, was acquired August 12, 1973 (Landsat frame-id 1385–18365). Landsat image line and column coordinates are not given in an earth-based reference frame. Rectification of the two Landsat images to the DTM was required. Rectification was performed automatically using the method described in Little (1982). The two false colour composite images, rectified to UTM coordinates, are shown as the left-most column of Fig. 4.

3.3. Ancillary data

Given latitude, longitude, date, and time, the position of the sun is determined by a computer program based on the method of Horn (1978). The exact timing of overflight for each Landsat MSS image is contained in the ancillary data recorded with each scan line. The January 8, 1973 image was acquired at 18:37 GMT (10:37 am PST). At that time, the sun had elevation $15.4°$ above the horizon and azimuth $154.9°$ measured clockwise from north. The August 12, 1973 image was acquired at 18:38 GMT (11:38 am PDT) with the sun at elevation $50.1°$ and azimuth $138.9°$. The study area is small enough that sun elevation and azimuth can be considered constant throughout.

Six one bit-per-pixel mask files are generated for use in subsequent analysis. One mask file is generated for each of the four ground cover classes. A fifth mask file is generated, using the gradient (p, q), for flat terrain. (Flat terrain is excluded when estimating parameters related to slope and aspect.) Finally, a sixth mask file is generated for points that lie in shadow. Approximately 12% of the January 8, 1973 image consists of pixels shadowed from the sun. For the August 12, 1973 image, the sun was sufficiently high in the sky that no image points were in shadow.

3.4. Synthetic image examples

The values of A, B, C and ρ_k, $k = 1, ..., 4$, required for image synthesis can either be hypothesized from assumptions about the world or estimated from the real image. Eight different methods for determining these parameters are described. Each method has been applied separately to each date and to each Landsat band. The results obtained are summarized in Table 1. The synthetic images corresponding to three of the methods are shown in Fig. 4.

The first five methods do not consider sky illumination explicitly. That is, parameter B of [9] is set to zero and scene radiance is given as the sum of two terms

[13] $L_r(i, e, g) = A\rho_k \cos i + C$

(Once again, the first term, due to solar irradiance, is zero for points in shadow.)

Method A considers ρ_k to be constant. Method A accounts only for the dependence of scene radiance on topography and the direct solar beam. This serves as a useful benchmark. The correlation coefficients for Method A, given in Table 1, are typical of those reported elsewhere for similar terrain and ground cover (Horn and Bachman 1978; Teillet et al. 1982; Woodham and Lee 1985).

Two qualitative observations can be made. First, method A did better for the January image than for the August image. This is mainly due to the lower sun elevation in January, $15.4°$ compared with $50.1°$. As is often noted in guidelines for aerial

photography, a low sun angle accentuates topographic variation, whereas a high sun angle is recommended to better delineate ground cover. Second, for each date, method A did better in band 7 than in band 4 or 5. The relative contribution of skylight and path radiance is greater in the shorter wavelength bands. Since method A accounts only for the direct solar beam, it is expected to be better in band 7 where atmospheric effects are minimized.

But ground cover is significant. Computing correlation coefficients on a per class basis supports the hypothesis that the correlation coefficients are different for each class. For example, method A, applied to band 7 of the January image, has an overall correlation coefficient of 0.68. On an individual class basis, the correlation coefficients are: old growth 0.77, second growth 0.75, and recent logging 0.48. (Water is excluded since lakes are flat. Estimating scene radiance as a function of the gradient (p, q) is meaningless for horizontal surfaces.)

Again, these results are consistent with intuition. Old growth is the most homogeneous cover class and thus would be most likely to have a single albedo throughout. Second growth is next. Recently logged areas, however, have different ground cover depending on circumstances and do not really represent a homogeneous cover type. It would not be as likely that a single albedo could account for as high a proportion of the variance for this class.

Method B estimates values for $A\rho_k$ for each ground cover class without regard to path radiance. A linear regression is performed using [13] subject to the constraint $C = 0$. The resulting synthetic images are shown in the second column of Fig. 4. Although some improvement in the overall correlation coefficients is noted, it is clear from a visual comparison of the real and synthetic images that too much variation is being attributed to topography and not enough to ground cover.

Method C first estimates the additive term C in [13] by linear regression without regard to ground cover. Then, method C estimates values for $A\rho_k$ for each ground cover class by linear regression, subject to the constraint that $L_r = C$ when $\cos i = 0$. The results show improvement, especially for the August image.

One difficulty with using regression in this context is that the solution always passes through the data points. Thus, it can be the case that real intensities occur that are less than the estimate of the path radiance term C. This violates the world model in the sense that C is assumed to be an additive term applied to all measurements. Woodham and Lee (1985) estimate path radiance as a function of elevation by a constrained optimization technique that fits the solution curve under the data rather than through it. When C does not vary with elevation, this is equivalent to setting C equal to the minimum intensity recorded.

Method D estimates values for $A\rho_k$ for each ground cover class by linear regression, subject to the constraint that C equals the minimum intensity value recorded. The results, as expected, are slightly worse than for method C, but the model has been forced to conform to assumptions made about the world.

Method E considers the possibility that C also depends on the ground cover class. Values of $A\rho_k$ and C_k are estimated for each ground cover class by linear regression. The resulting synthetic images are shown in the third column of Fig. 4. Method E is the best of the first five, although the differences between methods C, D, and E are slight.

The last three methods include the sky as a hemispherical uniform source. There are several ways to estimate the parameters A, B, and C of [9]. Once a path radiance correction

TABLE 1. Table of values of the correlation coefficient between bands 4, 5, and 7 of the two Landsat MSS images, acquired January 8, 1973 and August 12, 1973, respectively, and the corresponding synthetic images generated by the eight methods described in the text

	January 8, 1973			August 12, 1973		
	4	5	7	4	5	7
DTM (Method A)	0.33	0.42	0.68	0.23	0.20	0.46
DTM + ground cover (Method B)	0.40	0.48	0.71	0.30	0.40	0.45
DTM + ground cover + path radiance (Method C)	0.51	0.55	0.71	0.55	0.56	0.53
DTM + ground cover + path radiance (Method D)	0.46	0.52	0.70	0.54	0.55	0.48
DTM + ground cover + path radiance (Method E)	0.58	0.61	0.73	0.56	0.55	0.48
DTM + ground cover + sky (Method F)	0.48	0.56	0.73	0.32	0.43	0.44
DTM + ground cover + path radiance + sky (Method G)	0.58	0.41	0.74	0.53	0.55	0.45
DTM + ground cover + path radiance + sky (Method H)	0.59	0.63	0.76	0.54	0.56	0.41

has been applied, local measurements across shadow boundaries have been used to estimate the relative contributions of sun and sky (Sjoberg and Horn 1983; Woodham and Lee 1985). This has not proven very successful in practice. The reason is not entirely clear. One possibility is that points near shadow boundaries are a poor choice, because they correspond to points where a significant fraction of the sky is occluded by neighbouring terrain. The fraction of the sky seen would then not be a local function of the gradient (p, q) alone.

Method F estimates values for $A\rho_k$ and $B\rho_k$ for each cover class without regard to path radiance. Multiple linear regression is performed subject to the constraint that $C = 0$. Method G sets C equal to the minimum intensity value recorded, as did method D, and then estimates $A\rho_k$ and $B\rho_k$ for each cover class. Method H lets C depend on the ground cover class, as did method E, and estimates values for $A\rho_k$, $B\rho_k$, and C_k for each cover class. The synthetic images produced by method H are shown in the right-most column of Fig. 4.

4. Discussion

The atmosphere and adjacent targets further complicate the scene radiance equation for remote sensing as illustrated in Fig. 5. The atmosphere has optical thickness. This causes attenuation of the direct solar beam before it reaches the target. The radiance reflected from the target is also attenuated by the atmosphere before it reaches the sensor. Both the upward and downward transmission of the atmosphere depend on elevation and other factors and thus vary spatially.

Skylight includes radiation from the sun scattered by the atmosphere to the target, radiation reflected directly to the target from adjacent targets, and radiation reflected from adjacent targets that is scattered by the atmosphere back to the target. The component of sky radiance due to adjacent targets is small in areas of low albedo, but may become significant for areas of high albedo and in rugged terrain.

Path radiance includes radiation scattered to the sensor from the direct solar beam and radiation scattered to the sensor from
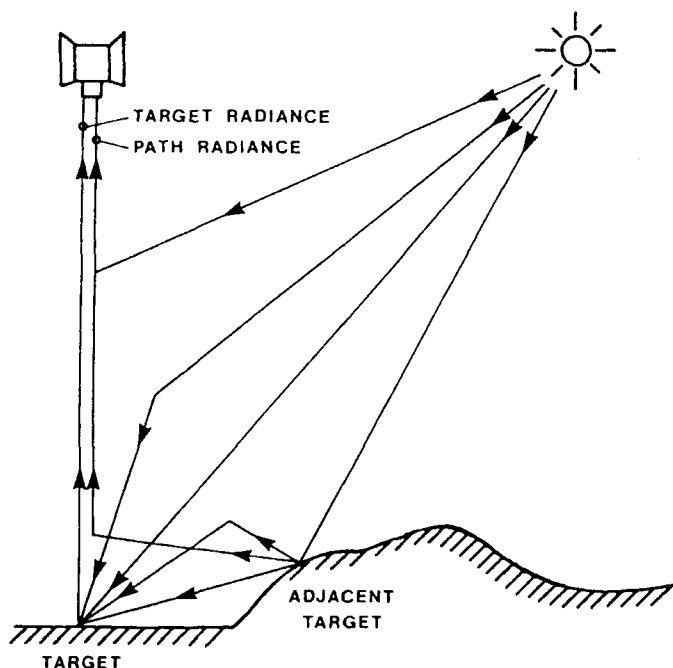

FIG. 5. Components of scene radiance. A target receives both direct solar and diffuse sky radiance. The sky component includes scattered solar radiation and radiation from adjacent targets that is reflected directly or scattered back to the target. The sensor receives radiance that includes an additional path component not originating from the target. Path radiance includes radiation scattered to the sensor from the solar beam and from radiation reflected from adjacent targets.

light reflected by adjacent target areas. The component of path radiance due to adjacent targets is small in areas of low albedo, but may become significant as the albedo of the ground increases (Otterman et al. 1980). Thus, adjacent targets can increase both path radiance and skylight at the target. These two effects are difficult to separate (Dozier and Frew 1981).

A general solution requires a solution to the radiative transfer

problem for the ambient radiation field (Turner and Spencer 1972). This further couples all atmospheric effects together and makes it difficult to treat them separately. A complete treatment is beyond the current state of the art. It is useful, however, to summarize the assumptions that have been made in the model presented here.

1. The atmosphere is assumed to be an optically thin, horizontally homogeneous layer; and the scene is assumed to have a narrow range of elevations. This allows atmospheric parameters to be absorbed into the constants A, B, and C of [9].

2. Radiation arising from adjacent targets, including clouds, is not considered.

3. The sensor views the target from directly overhead.

4. The sky is assumed to be a uniform hemispherical source.

5. Ground cover is assumed to be lambertian with BRDF $f_r = \rho/\pi$, where ρ is the albedo of the surface material.

Assumptions 1 and 2 are the most restrictive in that they allow scene radiance to be defined as a local function of the gradient (p, q). Adding a dependence of atmospheric parameters on elevation is a simple extension, since scene radiance remains a locally computable function. This extension is warranted when there is significant relief change in the scene (Woodham 1980; Sjoberg and Horn 1983; Woodham and Lee 1985). A full generalization of assumptions 1 and 2 would necessitate analysis over extended spatial contexts. Assumption 3 simplifies the coordinate transformations required, but is otherwise not restrictive. The results can easily be extended to off-nadir sensors. Assumptions 4 and 5 are straightforward to relax should better models become available. A scene radiance equation can be derived for any distribution of sky radiance and for any given BRDF (Horn and Sjoberg 1979).

The assumption that natural surfaces behave like lambertian reflectors has been questioned in the remote sensing literature (Smith et al. 1980; Justice et al. 1981; Teillet et al. 1982). One approach is to assume scene radiance is given by [13], where A and C are constants and ρ is the albedo. Constants A and C are estimated by a linear regression of brightness and cos i, assuming some average value for ρ. An albedo map is then produced by solving [13] locally for ρ. That is,

$$[14] \qquad \rho = \frac{L_r - C}{A \cos i}$$

This approach has been found to "overcorrect" in areas of steep terrain, particularly in the shorter wavelength bands. This subjective observation is supported by a more objective measure. That is, correction based on this method does not lead to improved accuracy in spectral classification.

In retrospect, this result should not be surprising. Sky radiance is significant and cannot be ignored, particularly in the shorter wavelength bands. Steep terrain tends also to correspond to terrain approaching the grazing angle of incident solar radiation. Here, sky radiance dominates and any method that does not explicitly take it into account will necessarily overcorrect. Sky radiance varies with slope and aspect, and cannot be absorbed into the constant C. Thus, this method to estimate albedo would fail with the same symptoms, even if the surface were lambertian.

The lambertian assumption has been retained here for three reasons: (1) The lambertian assumption leads to a simple phenomenological model for scene radiance that is easy to deal with computationally; (2) the lambertian model is the only BRDF that allows an intrinsic albedo to be computed from an image as a scalar, independent of viewer position; and (3) when

atmospheric parameters are explicitly modeled, sensitivity analysis shows that the scene radiance [9] is influenced more by errors in the estimate of optical depth than by potential deviations from the assumption of ideal lambertian reflectance (Sjoberg and Horn 1983).

Of course, this is not to say the natural surfaces are lambertian. It is sufficient to point out that the rejection of the lambertian assumption is premature. More work needs to be done to deal with sky radiance, path radiance, and adjacent targets before a definitive conclusion can be given.

The simple model presented here has, in fact, performed well for the Shawnigan Lake study site. For example, method H, applied to the January image, accounts for a significant component of the total variance observed in the real image. Specifically, the model accounts for 35% of the total variance in band 4, 40% in band 5, and 58% in band 7.

The remaining variance combines components due to (1) errors in sensor calibration; (2) errors in the DTM and ground cover knowledge base; (3) systematic errors in the model; and (4) natural variability in the world not accounted for in the model. It is assumed that the transfer characteristics of the sensor are known, so that scene radiance can be determined from image irradiance. This is a reasonable assumption, in general, although there were some difficulties in calibrating the Landsat 1 multispectral scanner. One would expect some variance due to errors in the DTM and in the ground cover map. Careful scrutiny of the two Landsat images, for example, reveals that additional logging activity has occurred between January 8, 1973 and August 12, 1973. Thus, the ground cover map is inaccurate for at least one of the two dates. One application of the ideas presented here is to note image regions that deviate from the values predicted by the synthetic image. This can be the basis for automatic map verification and update. Systematic errors in the model arise to the degree that the real world violates any of the assumptions listed above. It remains an open question to determine whether the models that now exist are sufficiently robust to be useful in practice. Finally, it is not likely that a single albedo term ρ_k is adequate to model each class in a ground cover map, even if the earth's surface were locally lambertian. There is bound to be variability in each cover class not accounted for by a single scalar ρ_k. The simple model presented literally attempts to account for each pixel in the image. Synthetic images can be augmented by adding noise components appropriate to each cover class. While these augmented synthetic images may look more realistic, the comparison between real and synthetic image, as measured by correlation, would, of course, degrade.

5. Conclusions

To the extent that the laws of physical optics are adequately represented, a scene radiance equation must, of necessity, be correct. It determines the image as a function of the scene. But, vision is the inverse problem. The task in vision is to determine the scene as a function of the image. Existence, uniqueness, and stability of the solution to the inverse problem cannot be assured without additional constraint.

The application to remote sensing described here assumes surface material to be lambertian and derives scene radiance for both collimated and hemispherical uniform source illumination. Path radiance is also considered. The model presented is simple and no doubt the real world is more complex. Nevertheless, the fundamental difficulty has been clearly demonstrated. The problem of determining surface properties from image proper-

ties is underconstrained. Equation [9] gives scene radiance as a function of local scene properties (p, q) and ρ_k and constants A, B, and C. Clearly, there are different combinations of (p, q) and ρ_k that can give rise to the same scene radiance. Further confounding occurs if parameters A, B, and C also vary spatially.

In computational vision, there are two strategies to follow. One strategy seeks additional constraint from *a priori* restrictions on the scene. This has led to progress in industrial applications where the environment can be controlled and where the visual task is often simple and well defined. It is also leading to progress in remote sensing when additional knowledge of the scene is available in the form of a DTM and a ground cover map. The second strategy imposes additional constraint on the perceiver, independent of the scene. Constraints on the perceiver may result in a unique solution, even though the physical optics of the problem does not. Of course, in such circumstances the solution will occasionally be incorrect. Nevertheless, it seems that this second strategy must apply in human vision.

A scene radiance equation is a useful tool because it establishes how the world determines what we see. It provides a theory of the problem of vision and helps to make computational vision a theoretical science as well as an experimental one. This paper has demonstrated the use of image synthesis as a tool for image analysis. Image synthesis has been viewed as a domain mapping. Synthesis maps scene knowledge into a common representation, namely the image, to facilitate the analysis of what is seen.

Acknowledgements

CATANZARITI, E., and MACKWORTH, A. K. 1978. Forests and pyramids: using image hierarchies to understand Landsat images. Proceedings Canadian Symposium on Remote Sensing, Victoria, BC, pp. 284–291.

DOZIER, J., and FREW, J. 1981. Atmospheric corrections to satellite radiometric data over rugged terrain. Remote Sensing of Environment, 11, pp. 191–205.

ELIASON, P. T., SODERBLOM, L. A., and CHAVEZ, P. S., JR. 1981. Extraction of topographic and spectral albedo information from multispectral images. Photogrammetric Engineering and Remote Sensing, 48, pp. 1571–1579.

HAYS, W. L., and WINKLER, R. L. 1970. Statistics: probability, inference and decision. Vol. II. Holt, Rinehart and Winston, New York, NY.

HORN, B. K. P. 1977. Understanding image intensities. Artificial Intelligence, 8, pp. 201–231.

———— 1978. The position of the sun. Artificial Intelligence Laboratory, MIT, Cambridge, MA, Report AI-WP-162.

———— 1981. Hill-shading and the reflectance map. Proceedings Institute of Electrical and Electronics Engineers, 69, pp. 14–47.

HORN, B. K. P., and BACHMAN, B. L. 1978. Using synthetic images to register real images with surface models. Communications of the Association for Computing Machinery, 21, pp. 914–924.

HORN, B. K. P., and SJOBERG, R. W. 1979. Calculating the reflectance map. Applied Optics, 18, pp. 1770–1779.

IKEUCHI, K., and HORN, B. K. P. 1981. Numerical shape from shading and occluding boundaries. Artificial Intelligence, 17, pp. 141–184.

JUSTICE, C. O., WHARTON, S. W., and HOLBEN, B. N. 1981. Application of digital terrain data to quantify and reduce the topographic effect on Landsat data. International Journal of Remote Sensing, 2, pp. 213–230.

LITTLE, J. J. 1982. Automatic registration of Landsat MSS images to digital elevation models. Proceedings Workshop on Computer Vision: Representation and Control, Rindge, NH, pp. 178–184.

MACKWORTH, A. K. 1983. Constraints, descriptions and domain mappings in computational vision. In Physical and Biological Processing of Images. Edited by O. J. Braddick and A. C. Sleigh. Springer-Verlag, New York, pp. 33–40.

NICODEMUS, F. E., RICHMOND, J. C., HSIA, J. J., GINSBURG, I. W., and LIMPERIS, T. 1977. Geometrical considerations and nomenclature for reflectance. NBS Monograph 160, National Bureau of Standards, Washington, DC.

OTTERMAN, J., UNGAR, S., KAUFMAN, Y., and PODOLAK, M. 1980. Atmospheric effects on radiometric imaging from satellite under low optical thickness conditions. Remote Sensing of Environment, 9, pp. 115–129.

PEUCKER, T. K., FOWLER, R. J., LITTLE, J. J., and MARK, D. M. 1978. The triangulated irregular network. Proceedings Digital Terrain Models Symposium, St. Louis, MO, pp. 516–532.

SHIBATA, T., FREI, W., and SUTTON, M. 1981. Digital correction of solar illumination and viewing angle artifacts in remotely sensed images. Proceedings Machine Processing of Remotely Sensed Data Symposium, West Lafayette, IN, pp. 169–177.

SJOBERG, R. W., and HORN, B. K. P. 1983. Atmospheric effects in satellite imaging of mountainous terrain. Applied Optics, 22, pp. 1702–1716.

SMITH, J. A., LIN, T. L., and RANSON, K. J. 1980. The Lambertian assumption and Landsat data. Photogrammetric Engineering and Remote Sensing, 46, pp. 1183–1189.

STARR, D., and MACKWORTH, A. K. 1978. Exploiting spectral, spatial and semantic constraints in the segmentation of Landsat images. Canadian Journal of Remote Sensing, 4, pp. 101–107.

TEILLET, P. M., GUINDON, B., and GOODENOUGH, D. G. 1982. On the slope-aspect correction of multispectral scanner data. Canadian Journal of Remote Sensing, 8, pp. 84–106.

TURNER, R. E., and SPENCER, M. M. 1972. Atmospheric model for correction of spacecraft data. Proceedings International Symposium on Remote Sensing of Environment, Ann Arbor, MI, pp. 895–934.

WANG, S., HARALICK, R. M., and CAMPBELL, J. 1984. Relative elevation determination from Landsat imagery. Photogrammetria, 39, pp. 193–215.

WOODHAM, R. J. 1980. Using digital terrain data to model image formation in remote sensing. Proceedings Society of Photo–Optical Instrumentation Engineers, 238, pp. 361–369.

———— 1981. Analysing images of curved surfaces. Artificial Intelligence, 17, pp. 117–140.

———— 1982. Aspects of computational vision. Proceedings Canadian Society for Computational Studies of Intelligence Conference, Saskatoon, Sask., pp. 237–241.

WOODHAM, R. J., and LEE, T. K. 1985. Photometric method for radiometric correction of multispectral data. Canadian Journal of Remote Sensing (submitted). Available as TR-84-14, Department of Computer Science, University of British Columbia, Vancouver, B.C.