An inverse problem approach for the segmentation of snow cover in satellite images

Mihai Datcu
German Aerospace Research Establishment (DLR), German Remote Sensing Data Center (DFD), Oberpfaffenhofen, Germany
Drago Luca & Klaus Seidel
Image Science Division, Communication Technology Laboratory, ETH, Zürich, Switzerland

ABSTRACT: A new model for the correction of topographic effects in satellite images of rough terrain is described. The model simulates a synthetic image of the scene using a computer graphics approach which combines ray-tracing techniques with radiosity methods. Computation is structured on three levels: a macro level in which the image is described by the Digital Elevation Model and the light source, a meso-scale in which the model simulates the integration effect of the imaging sensor and a micro-scale which is characterized by the reflectance of the snow cover (specular and diffuse). The parameters of the model are tuned with a gradient search to fit real images acquired by the Landsat-TM sensor. The results show a better accuracy than the classical "cosine of incidence" and Minnaert models. Additionally a new technique based on maximum entropy estimation is used to determine the reflectance function of snow and compare it with the one predicted by our model.

INTRODUCTION

The radiometry of remote sensing images in rough mountainous areas is severely affected by topographic artifacts. Shadow areas, diffuse and indirect secondary lighting due to slopes covered by highly reflective snow as well as perturbant atmospheric effects cause ambiguities that appear in the interpretation of these images. These artifacts pose severe problems especially if the land-cover classification and/or image segmentation is to be performed automatically. On the other hand, manual interpretation of the images is slow and affected by human subjectivity.

Three approaches have mainly been used for the topographic correction of images in rough terrain areas. The first one uses multispectral image analysis based on the observation that in several infrared bands the images have weak dependency on the land cover and strong dependency on the topography. Image correction is then performed using band ratios or some statistical transformations based mainly on regression techniques (Naugle and Lashlee, 1992; Civco, 1989; Colby, 1991; Dozier, 1989). This method is simple to implement but it contains some heuristic assumptions and has often a poor numeric performance.

The second approach uses a radiative transfer code to obtain a deterministic description of the correction of topographic effects (Goel et al., 1991; Conese et al., 1993; Hill et al., 1995; Oren and Nayar, 1995; Richter, 1997). While this approach avoids empirical techniques, it is more difficult to implement and encounters problem in the correct description of radiance and other lighting parameters of the scene.

The approach we suggest is the third one and represents a model based approach. Our algorithm uses computer graphics techniques (ray-tracing and radiosity) to synthesize an image of the observed scene starting from the Digital Elevation Model (DEM), the position of the sun and the characteristics of the imaging sensor. But unlike other approaches, critical parameters related to the description of scene radiances are not postulated or deduced in advance but estimated on-line in an optimization loop. This loop performs a search in the parameter space to find the set of parameters that minimizes the error between the synthesized and the observed remote sensing image. In this way we avoid problems related to the off-line estimation of radiance values; we make only basic assumptions on reflectances and lighting and let the algorithm extract from the measured data the appropriate values.

THE IMAGE FORMATION MODEL

The image synthesis is based on the knowledge of the DEM, on the position and physical properties of the light source and on the scattering process for different cover types. To derive an accurate model a multiresolution approach is used that includes three
scales (Datcu and Holecz, 1993):

- At macro-scale the scene is described by the DEM and the position of the light source. These parameters are input to a computer graphics program that combines ray-tracing and radiosity techniques (Ashdown, 1994) and can render with accuracy both specular and diffuse reflections. The simulation is performed at a resolution higher than the spatial resolution of the sensor.

- At meso-scale a spatial integration is performed that models the image formation process of the sensor. The resolution of the simulated image is reduced to match the observed data. Due to the nonlinear description of the light scattering process, the result of this operation does not equal the reflection value obtained without integration. Even in a simple Lambertian assumption for each face of the low resolution DEM we have an intensity proportional to the cosine of the incidence angle:

\[ I_{\text{low}} \propto \cos(\vec{n}, \vec{I}) \]  

while for the high resolution DEM the intensity is the sum of the cosines of all the incidence angles corresponding to the same area (Figure 1):

\[ I_{\text{high}} \propto \sum_{k=1}^{N} \cos(\vec{n}_k, \vec{I}) \]  

\[ I_{\text{low}} \] is not equal to \( I_{\text{high}} \). Additionally, the interreflections between slopes that occur in snow-covered terrain cannot be neglected and influence the result of the integration.

\[ \text{Figure 1: Reflection of light on a low resolution and on a high resolution DEM. Due to the nonlinear process, the reflected intensities are not equal.} \]

- At micro-scale the facettes of the DEM surface are characterized by reflectance functions. We model the reflectance function as a linear superposition of diffuse and specular behaviour. In order to catch the interreflections between slopes we suppose that each facette may be illuminated by several light sources. The Lambertian component will thus be:

\[ I_{\text{diffuse}} = k_d C_s \sum_{j=1}^{N} I_j \cos(\vec{n}, \vec{l}_j) \]  

with \( k_d \) the diffuse reflection coefficient, \( C_s \), a coefficient related to the spectral characteristics of the surface, \( \vec{n} \) the local surface normal, \( I_j \) the intensity of the \( j \)-th source and \( \vec{l}_j \) the incidence vectors. Similarly, the specular component will be:

\[ I_{\text{specular}} = k_s C_s \sum_{j=1}^{N} I_j \cos(\vec{n}, \vec{l}_j)^m \]  

with \( m \) depending on the surface glossiness and \( \vec{l}_j \) the vector in the direction of the bisector line between the observer and the \( j \)-th light source.

As described before the model contains some parameters that have to be set like e.g. the percentage of specular reflection and of diffuse reflection, the level of the ambient light, etc. In our approach, these parameters are not preset, but are determined on-line in an optimization loop with the gradient method. The mean square error between the synthetic and the observed image is used as an objective function of the minimization. In this way we avoid the common difficulties that appear in estimating the radiances for each image pixel and let the algorithm adapt the parameters to the available measurements. The block diagram of the model is shown in Figure 2. This approach can also be interpreted in the frame of the Bayesian inference theory: the incomplete information about the data and the image formation is represented as probabilities and a measure of the goodness of the model is derived using probability theory concepts (Datcu et al., 1997).

3 RESULTS

We have been testing our method on a Landsat-5 TM image from 26 April 1992 of the Davos region in Switzerland. The data is corrected for both radiometry and geometry. It was resampled to 25 meters and 10 meters using nearest neighbour interpolation to fit the available DEMs. The 10 meter data is used for the high resolution image synthesis, the 25 meter data for the computation of the error. Figure 3 shows the geocorrected image, from which a rectangular area containing several shadows and interreflections was selected as test data (Figure 4a).
For the computer graphics simulation we used the RADIANCE 2.5 public domain software from the Lawrence Berkeley Laboratory, Berkeley, California (Ward, 1994; Larson, 1991). This software combines ray-tracing and radiosity techniques and allows an accurate representation of both specular and diffuse reflections. It offers a large range of parameters to control the quality of the simulated image and several built-in functions for surface generation, sun position computation, ambient light control, filtering, enhancements, etc. We selected as variable parameters the amount of specular reflection and the ambient light in the image. These parameters were modified by the gradient algorithm to minimize the mean square error between the simulated and the observed image.

After optimization we obtained the result in Figure 5. The simulated image renders in general with good accuracy the structure of the real image.
3.1 Evaluation of the error

In order to assess the quality of our approximation we computed the mean square error between the original TM image and synthetic images generated both with our model and with the "cosine of incidence" and Minnaert model (Itten et al., 1992). We tried to answer the following questions:

- What is the error obtained after successful minimization of the objective function? How close can our model simulate the real image?

- What is the difference between the simulation performed on a high resolution DEM followed by integration and the direct simulation on the low resolution DEM? Does the simulation on high resolution data improve the results?

- What is the performance of our model compared to other known models?

The answers to these questions are given by the images in Figure 6.

Figure 6a represents the error image computed as the module of the difference between the original image and the synthetic image generated by simulation on a high resolution DEM (i.e. between the images in Figure 5a and 5b). The mean square error obtained is 233, i.e. in the mean the pixels are affected by an error of 15.2 grey levels. The error image was inverted for a better visualization, i.e. white corresponds to zero error. Errors appear mostly in specular and shadow areas. These errors are much smaller than those obtained with other simulation methods. Note also that some of these errors are inherent to the scene: they represent rocks not covered with snow or man-made objects. E.g. the dark pixels marked with an arrow on the error image correspond to the building of the Weissfluhjoch mountain station. Other error pixels in the image were identified on the map of the area as being bare rocks. For a comparison we show in Figure 7 the map (1:25000) of the selected test area.

Figure 6b represents the error image between the synthetic image (simulation on a low resolution DEM) and the original TM image. All error images are inverted, white represents zero error.
in a gradient loop, but the simulation is performed directly on the 25m elevation data and no integration is performed. The mean square error is in this case 334, i.e. in the mean the pixels are affected by an error of 18.3 gray levels. This higher error is visible especially in the shadow area in the central-eastern part of the image. Although the simulation of a higher resolution image is more costly in terms of computer time and power, it proves to be justified since it leads to a lower rendering error.

Figure 7: Map of the selected test area, 1:25000.

In order to compare the accuracy of our model with that of other algorithms we performed simulations of the test area with the "cosine of incidence" model and the Minnaert model. The mean square error of the "cosine of incidence" model is 1569, the Minnaert model achieves lower errors in the range 650-1500 depending on the value of the Minnaert constant. Both these mean square errors are substantially higher than that of 233 achieved with our algorithm. The error images for the "cosine of incidence" model and the Minnaert model with constant $k = 0.5$ are shown in Figures 6c and d respectively. Especially shadow areas show a high level of error, the models do not manage to render accurately the pixel intensity in these zones.

An overview of the results of our experiments is presented in Table 3.1. This overview shows a comparison of the mean square errors (MSE) obtained with several models for our test area. The smallest error is obtained with our model using a simulation on the high resolution DEM.

3.2 Validation of the model from reflectance data

Our model assumes at micro-scale that the reflectance of snow is a linear superposition of a diffuse and a specular component. This assumption is new: the "cosine of incidence" model uses pure Lambertian behaviour, other models like the Minnaert model or the C-correction algorithm simulate the non-Lambertian component by a correction factor (Itten et al., 1992). It makes sense to try to validate the assumptions of these models by comparing the reflectance they predict with the real (measured or computed) reflectance function for snow. This comparison would represent an additional argument for the selection of an appropriate simulation algorithm.

In the case of an isotropic snow cover, the reflectance function is a function of three variables: $\phi_1$, the angle of the view direction with the plane normal $n$, $\phi_2$, the angle of the illumination direction with $n$ and $\theta$, the angle made by the projections of these two directions on the base plane (Figure 8a). Since a function of three variables cannot be visualized easily, the usual representations of the reflectance function are its projections on one or two dimensions, keeping two, respectively one of the three variables at a fixed value. E.g. if $\phi_1$ is kept at a constant value, and $\phi_2$ and $\theta$ vary from $0 \ldots 180^\circ$, $0 \ldots 360^\circ$ respectively, our model predicts a reflectance function with a shape similar to the one in Figure 8b. This shape is a superposition of the hemisphere that corresponds to the diffuse component and a "reflection lobe" that corresponds to the specular component (Dateu and Holecz, 1993). We will show that at least for a given range of values $\phi_1$ the real (measured) reflectance of snow conforms to this pattern which supports the use of our model.

The main difficulty of this task is the fact that the reflectance function of snow is not known in the general case and is difficult to determine both experimentally and theoretically due to the complex process of snow metamorphism (Gude, 1997). In a period of only a few days snow may change its properties and structure dramatically which makes any retroactive attempt to determine its reflectance practically impossible. Still the reflectance information can be recorded on satellite images. In the following we will describe a new method which allows the estimation of the reflectance function from remote sensing data. This method uses only a few samples of this function extracted from our Landsat-TM image then performs a generalization according to the principle of maximum entropy. The result is an estimate of the reflectance function which confirms the assumptions we included in our model.

a) Description of the method

The classical experiment for determining the reflectance function of snow consists in positioning a light source and a light measuring device (camera) above a plane covered with snow and performing measurements for different angles $\phi_1$, $\phi_2$ and $\theta$. An equivalent result can be obtained if the positions of the light source and the camera are fixed, but the plane covered with snow varies its inclination such as to cover the same range of angles. We apply this idea to a Landsat-TM image of a rough mountainous area for which the positions of the sun and sensor are fixed (and known) and the rough terrain provides a wide range of local inclinations. The intensity of the pixels in the Landsat image are samples of the reflectance functions for
Table 1: Mean Square Error between original and synthetic images

<table>
<thead>
<tr>
<th>Model</th>
<th>High Res.</th>
<th>Low Res.</th>
<th>Cos of incidence</th>
<th>Minnaert k = 0.8</th>
<th>Minnaert k = 0.5</th>
<th>Minnaert k = 0.2</th>
</tr>
</thead>
<tbody>
<tr>
<td>MSE</td>
<td>233</td>
<td>334</td>
<td>1569</td>
<td>1241</td>
<td>855</td>
<td>655</td>
</tr>
</tbody>
</table>

Figure 8: a) The reflectance function is a function of the three variables $\phi_1$, $\phi_2$, and $\theta$. b) Shape of the reflectance function predicted by our model for $\phi_2$ constant: superposition of diffuse and specular component.

angles $\phi_1$, $\phi_2$ and $\theta$ that can be computed from the sun and sensor position and the local inclination of the DEM.

The samples obtained in this way are spread very sparsely in the 3-dimensional space \{\phi_1, \phi_2, \theta\}. E.g. for our test area of 90 x 50 pixels, a simple triangulation yields about 9000 points. If we want to describe the reflectance function with one point for each degree value of $\phi_1$, $\phi_2$ and $\theta$, we need 1.45 million points. A method is thus needed to interpolate or estimate from the available points (0.6% of the total) the rest of the samples. This is done using the maximum entropy principle.

b) The maximum entropy principle

The principle of maximum entropy (Jaynes, 1957) was first introduced as a general method of inference about unknown probability densities subject to a set of constraints, given a prior estimate of those probability densities. The early rationales were based mainly on intuitive arguments or on the properties of entropy as an information measure, but later approaches (Shore and Johnson, 1980) showed that the maximum entropy formalism emerges as a consequence of requiring that the methods of inductive inference be self-consistent, thus providing a more general framework by making no reference to information measures.

In terms of the inference of the probability densities, suppose one has a system which has a set of states $S = \{s_1\}$, having the unknown probabilities $p(s_i)$. Additional information about the probabilities of the states is also available in term of constraints on the distribution $p$: values of some expectations, or other information. These expectations constraint the space of the solutions to a certain domain (preferable convex), but usually there remains an infinite set of distributions that are not ruled out by the constraints. The principle of maximum entropy states that of all distributions $p$ that satisfy the constraints, one should choose the one with the largest entropy $H(p)$. $H(p)$ is given by:

$$H(p) = - \sum_i p(s_i) \log(p(s_i))$$

In our application we understand the reflectance function as a probability distribution: each of its samples is a probability measure for the scattering of radiation on the direction given by the independent variables $\phi_1$, $\phi_2$, and $\theta$. The reflection function is thus a 3-dimensional histogram from which we want to estimate the underlying probability density. Due to the course of dimensionality, learning such a probability distribution from observations requires very large data sets. The maximum entropy method allows the estimation process to be based on only a few samples. In our case the training set contains 0.6% of the data.

c) Estimated reflection function

Figure 9 shows the result of the maximum entropy estimate for the reflectance function. The function is plotted for two particular values of $\phi_1$: $\phi_1 = 11^\circ$ and $\phi_2 = 20^\circ$. The independent variables are $\theta$ ($0 < \theta < 180^\circ$) and $\phi_2$ ($0 < \phi_2 < 90^\circ$). The shape of the function looks very similar to the predicted one (Figure 8b). However, this happens only if $\phi_1$ is in the range $[5^\circ ... 25^\circ]$. Outside this interval the estimated function has a pure diffuse appearance. Further work is thus necessary to determine if this is an artifact due to the to reparation of the training data in parameter space or is intrinsic to the reflectance of the snow in the given frequency range for our test area.
Figure 9: Samples of the reflectance function determined with our method. a) $\phi_1 = 11^\circ$, b) $\phi_2 = 20^\circ$. Note the similarity with the predicted shape.

4 CONCLUSIONS

This paper describes a new model for the correction of topographic effects in satellite images of snow covered rough terrain. The model simulates a synthetic image using computer graphic techniques and relies on a multiresolution approach. Its parameters are estimated in an optimization loop which minimizes the mean square error between the synthetic and the real image.

The results of the performed experiments confirm the validity of the approach. The error obtained with the suggested method is much smaller than the error obtained with other models and also smaller than the error obtained if no multiresolution approach is used. The model is additionally validated by comparing the reflectance values it predicts with the ones measured from the satellite image. To this purpose a new method for the derivation of the reflectance function from remote sensing data is developed. This method is based on the maximum entropy principle which allows the estimation of a distribution function from a very sparse training set.

5 ACKNOWLEDGEMENT

We would like to acknowledge the Remote Sensing Laboratories of the University of Zürich for providing us the data on which this investigation is based.

References


