Super-Resolved Surface Reconstruction From Multiple Images

Technical Report FIA–94–12

Peter Cheeseman
RIACS*

Bob Kanefsky
Recom Technologies

Richard Kraft
Recom Technologies

John Stutz
NASA

Robin Hanson
Recom Technologies

Artificial Intelligence Research Branch
NASA Ames Research Center, Mail Stop 269–2
Moffett Field, CA 94035, USA
Email: <last-name>@ptolemy.arc.nasa.gov
Phone: (415) 604–4946

Related information available on the World-Wide Web:
`http://ic-www.arc.nasa.gov/ic/projects/bayes-group/group/super-res/`

December 14, 1994

Abstract

This paper describes a Bayesian method for constructing a super-resolved surface model by combining information from a set of images of the given surface. We develop the theory and algorithms in detail for the 2-D reconstruction problem, appropriate for the case where all images are taken from roughly the same direction and under similar lighting conditions. We show the results of this 2-D reconstruction on Viking Martian data. These results show dramatic improvements in both spatial and gray-scale resolution. The Bayesian approach uses a neighbor correlation model as well as pixel data from the image set. Some extensions of this method are discussed, including 3-D surface reconstruction and the resolution of diffraction blurred images.

*Research Institute for Advanced Computer Science

1 Introduction

Consider the problem of how to extract as much information as possible from a set of images, all of the same scene, and of capturing this information in the form of a surface model at maximal resolution. This problem is important in many applications where maximal resolution is paramount. In this paper we focus on space-based remote imaging.

Surface reconstruction from an image set is an example of an inverse problem: if we knew exactly the shape and emittance of the surface, the illumination conditions, the camera angle, etc., we could predict what the camera would observe (the pixels) to within the measurement accuracy. This is the rendering problem addressed by computer graphics. We have the inverse problem: we are given the observed images (pixels) and must use this information to find the most probable surface that could have generated these images. Bayes's theorem provides a formal solution to inverse problems, which we apply here to the surface reconstruction problem.

Because the reconstructed surface can only be determined to within a certain maximum spatial resolution, we represent surfaces by a discrete uniform grid with the surface properties given at each grid point. For the case of a planetary surface, these surface properties could include illumination, albedo, slope, emittance at different wavelengths, etc. We will describe in detail a model using only surface emittance, and then describe how to extend this model. These properties characterize the grid point and describe how it could influence the image pixels once the camera parameters are known. This surface grid is a reconstruction and is *not* what was actually observed. For this reason we call the surface grid elements **mixels** (for **m**odel p**ixels**) to distinguish them from **pixels** which are the *observed* values. Unfortunately, in much of the vision literature, the word pixel is used interchangeably to refer to both inferred and observed values.

We are able to get super-resolved reconstructions from image sets because each pixel of each image is a new sample of some patch on the observed surface. Two images generated with *exactly* the same alignment between the camera and the surface, the same illumination conditions, etc., record the same information to within the measurement error of the camera, resulting in no net gain of information. With slightly differing alignments, however, the observed pixel values will be different, because the camera is observing slightly different patches on the surface. By relating these differences to locations on the surface, it is possible to reconstruct a model grid at a finer resolution than the observation pixilation. This technique for combining overlapping information is closely related to deconvolution (e.g. radar imaging) and computed tomography (e.g. CAT scan), and is explained in more detail in section 3. In particular, this information combining technique goes beyond the Nyquist limit for a single observed image. Fig. 1 shows schematically why subpixel resolution is possible.

We start by considering "flat" surface reconstruction. This is the best that can be achieved when the images are taken from essentially the same camera position and sun angle, but with slightly different registrations. This occurs with Landsat images, for example, where each location on Earth is imaged from essentially the same position in space. The reconstruction gives the "emittance" of the surface, which is a combination of the effects of surface albedo, illumination conditions and ground slope. We develop this theory in detail in section 2, and show its application to Viking Orbiter data of Mars. This theory includes the use of prior knowledge in the form of neighbor correlations. In section 5, we outline how to extend this approach to a 3-D surface reconstruction, where images from different directions allow us to separate variations in pixel values due to albedo from those due to ground slope.

2 2-D Surface Reconstruction

Our approach is based on Bayesian probability theory. We use a likelihood function, defined to be the probability of the observed data given a model of how the data were generated. This model of

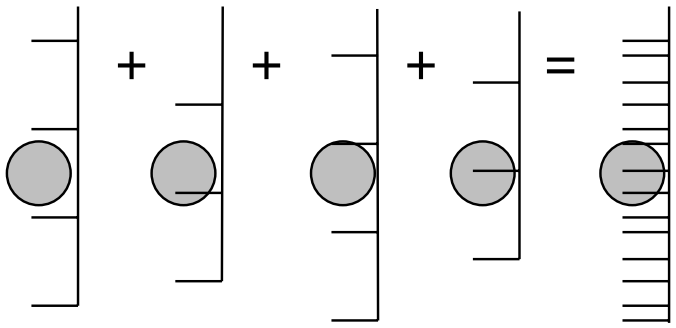


Figure 1: Because several sampling grids are used, offset randomly with respect to each other, resolution beyond the Nyquist limit of any one frame is possible.

the observation process is normally parameterized with respect to any variables that affect the process. For the current problem, these observational parameters include surface illumination, surface albedo, camera orientation, camera characteristics, optical distortions, and any data preprocessing. Computational considerations may require that some of these parameters be simplified or omitted, but doing so always entails some loss of precision.

We have made several such simplifications in the work described in this section. The first is that we model the "surface" as a plane lacking curvature and local relief — i.e., as a grid only of emittance values. Thus, the value of each mixel is simply a scalar — its emittance. The second is the substitution of a simple transformation (affine or quadratic) for the projective observation geometry and for any optical and electronic distortions of the camera system. A third lies in using a preprocessing step to deal with telemetry noise. For Mars images, we ignore ambient light contribution from a diffuse background and atmospheric attenuation, as they are negligible in the data sets we use.

In our approach, we begin by constructing a likelihood function that gives the probability of each pixel value, given the imaged surface and observation conditions. We take the likelihood of the entire image to be just the product of likelihoods of each pixel. This means we are assuming that the measurement error of a pixel is conditionally independent of the value of its neighbors. This conditional independence assumption is symbolically represented as:

$$\begin{aligned} & P\left[\text{all pixel values} \mid \text{observation params, mixels}\right] \\ & = \prod_p P\left[(\text{pixel(p)} = \Phi_p \mid \text{observation params, mixels}\right] . \end{aligned} \tag{1}$$

Here, pixel(p) is a location of a pixel on some image in the image set, and Φ_p is an observed energy value[1]. What we read off the camera is the radiant energy received by each pixel[9]. Note the split of parameters into two sets: observation parameters and mixels. We will explain the significance of this split below.

We assume that the probability of an observed pixel value is normally distributed, so that the likelihood of each pixel is given approximately by:

$$P\left[\text{pixel(p)} = \Phi_p \mid \text{observed params, surface model}\right] \approx N[\Phi_p \mid \hat{\Phi}_p, \sigma]\Delta\Phi_p . \tag{2}$$

Here $N[x \mid \mu, \sigma]$ is the standard normal (or Gaussian) distribution of x given a mean μ and standard deviation σ. The $\Delta\Phi_p$ term is the observed minimum gray-scale difference. The standard deviation σ

[1]This is not physically correct, as the camera outputs joules. However, one can multiply the flux Φ by the exposure time and the pixel size to obtain joules Q. We will stick to flux values to keep Eqn.(1) camera independent.

of the observed pixels from their expected values is assumed to be the same for all pixels in an image. This deviation results from measurement error (especially quantization error[2]) and model errors of various kinds (e.g. slight mis-registration). If these many sources of error are largely independent, the central limit theorem leads us to expect the resulting error distribution to be close to normal. Experimental data confirm this expectation, as is discussed in section 3.2.The normal approximation in this case assumes that $\sigma >> \Delta\Phi_p$. This distribution is just the trapezoid approximation to the integral of a normal density over the interval from Φ_p to $\Phi_p + \Delta\Phi_p$.

In Eqn. (2), the term $\hat{\Phi}_p$ represents the expected radiant energy[3]value for pixel p and is a function of the observation parameters and surface model. The parameters used in determining $\hat{\Phi}_p$, as used in likelihood Eqn. (2), are:

1. **Mixel Values:** This is the model of the reconstructed 2-D surface represented by an "emittance" value at each grid point (mixel);

2. **Registration Parameters:** These geometric parameters define how a pixel image maps onto the reconstructed mixel grid. Here, we use affine and quadratic transformations to define a 2-D (camera) to 2-D (mixel grid) function;

3. **Point Spread Function (PSF):** This function defines how points on the surface (mixels) contribute to the observed pixels through the camera optics, including any distortions produced by camera readout;

4. **Camera Shading:** These parameters are necessary for cameras, such as a vidicon[4], with a nonuniform readout gain across the image plane. These parameters define a scaling factor that varies depending on where on the image plane a particular pixel falls.

The contribution of these parameters to $\hat{\Phi}_p$ is shown diagrammatically in Fig. 2. Given values for the mixels and the parameters relating mixels to pixels, it is possible to calculate the expected value of a given pixel, $\hat{\Phi}_p$, by summing the contribution of each mixel, as weighted by the PSF. This is explained in detail in the next section. This pixel prediction process is just the "forward" graphics problem, shown in Fig. 2.

In a maximum likelihood (ML) approach, the goal is to find the set of parameter values that maximizes Eqn. (1)—in particular, the ML estimates of the mixels is a way of reconstructing an unknown surface from the images. Note that finding the ML mixel values is a way of solving the "inverse" graphics problem (i.e. finding a model from data) given the likelihood (i.e. probability of the data given a model). When the resolution chosen for the mixel grid is *overdetermined* by the corresponding pixel values, the ML approach is reasonable. The mixels are overdetermined by the pixels when there is no value for the mixels which can exactly predict all the pixel values. The overdetermined situation means that the mixel grid is at a coarser spatial resolution than is otherwise achievable. If the ML approach is tried at too fine a resolution, the mixel values are *underconstrained*—i.e., there are many mixel grids that would predict the pixel values exactly, and there is no principled way of choosing among them.

The Bayesian approach used here is similar to the ML approach, but it uses additional (prior) knowledge in the form of expectations about correlations among neighboring mixels. This additional knowledge in the Bayesian maximum a posterior (MAP) estimate allows any scale mixel grid. If too coarse a mixel grid is used (i.e. the mixels are overdetermined by the pixels), then the neighbor correlations have little effect, and the MAP estimate is essentially the same as the ML estimate. However, if a very fine mixel grid is used (i.e. the mixels are underdetermined), then the effect of the neighbor correlations competes with the fit to the data to give a reasonable compromise result that

[2]The quantization of continous emittance values into integers

[3]See discussion below in section 3.0 for terminology.

[4]A vidicon camera is an obsolete electron beam readout camera, such as used in the Viking Orbiter images shown in this paper.

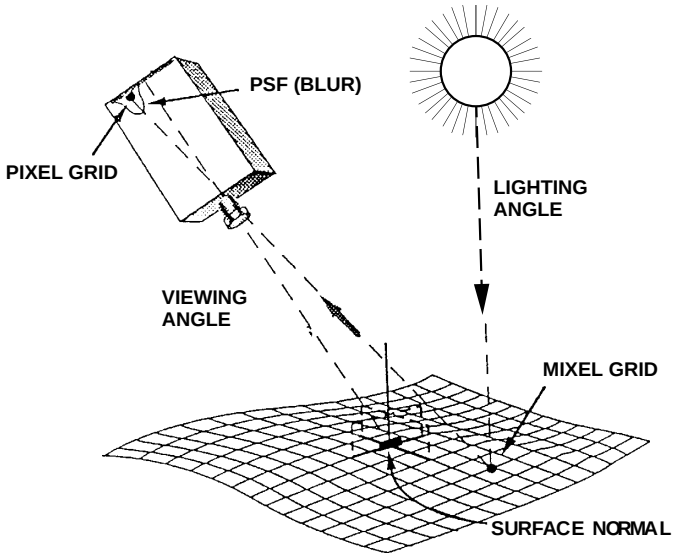


Figure 2: Parameters Relating Pixels to Mixels

uses all the information. The optimal mixel resolution is near the borderline between overdetermined and underdetermined, where the neighbor correlation information begins to suppress the affects of the noise in the data. Thus the prior term acts much in the same manner as a "regularization" term in related approaches.

3 MAP Reconstruction

Given pixel data and the parameters that specify the imaging model, we want to jointly estimate the mixel grid values together with other auxiliary model parameters. In a Bayesian approach one seeks a combination of all these parameters which has maximum posterior (MAP) probability, which is the same (up to a normalization factor) as seeking a maximum joint probability:

$$\begin{aligned} \text{Joint Probability} &= \text{Likelihood} \times \text{Prior Probability} \\ P[\text{Mixels, Pixels, Params}] &= P[\text{Pixels} \mid \text{Mixels, Params}] \times P[\text{Mixels} \mid \text{Params}] \times P[\text{Params}], \end{aligned}$$

where "Mixels" refers to the set of all mixel values, "Pixels" refers to the set of all pixels in all images, and "Params" refers to the auxiliary observational parameters (registration parameters, PSF, etc.) listed above.

Repeating Eqn. (2), the likelihood term is:

$$P[\text{Pixels} \mid \text{Mixels, Params}] = \prod_p N[\Phi_p \mid \hat{\Phi}_p, \sigma_p] \Delta\Phi_p . \tag{3}$$

We now specify the mean for each pixel $\hat{\Phi}_p$ to be a linear combination of the emittances from mixels projected near the pixel location:

$$\hat{\Phi}_p = \sum_i \omega_{ip} m_i . \tag{4}$$

Here m_i is the emittance of the ith mixel and ω_{ip} is the mixel-pixel weight defined by the PSF and registration information. For images that are not significantly diffraction blurred, the radiant energy

at a point in the image plane is a sum of contributing emittance values (not amplitudes) from the generating surface, as shown in equation (4).

So far, we have used the *physical* terminology appropriate for describing the quantities radiating from the mixel grid (emittance) and captured by the camera (radiant energy). While this convention is admittedly arbitrary, (we could, for example, equip the camera model with an exposure term E(Q) that turns radiant energy values Q to flux values ϕ) it is more precise than the alternate computer graphics convention of labeling a host of quantities with "intensity", regardless of its being a light source, a CRT raster, a planetary surface, etc. While in the present 2-D superresolution case these distinctions may not seem useful, such precision is helpful when the model is extended to 3-D surface reconstruction. It is in this context of anticipated extention that the terminology in this paper is chosen.

3.1 Prior

The prior probability term $P[\text{Mixels} \mid \text{Params}]$ is the distinctly Bayesian contribution, and it embodies one's beliefs *before* seeing the data about the kinds of scenes or landscapes one might observe. The simple prior used in this paper describes how mixel intensities m_i relate to each other. The remaining model parameters — the point spread function coefficients, optical angles, etc.—are highly over-determined by the data, so we can reasonably neglect the priors $P[\text{Params}]$ on these parameters.

To gain insight into the appropriate prior over the mixel intensities, we analyzed Viking Orbiter imagery. This prior can be thought of as a means of preferring a given solution when many solutions fit the data equally. We choose a prior that makes a reconstruction more likely if its mixel values are highly correlated with their neighbors (i.e. there is emittance continuity). A simple, probabilistic model of continuity would be to estimate the value of a mixel m_i by a weighted sum of its neighbors:

$$\hat{m_i} = \sum_j \alpha_{ij} m_j \,. \tag{5}$$

Here, the α_{ij} are weights: mixel m_j contributes α_{ij} to mixel m_i. While this form is fairly general, we choose to start with a particularly simple relationship where $\alpha_{ii} = 0$, $\alpha_{ij} = \alpha_{j\,i} = 1/4$ if $|i - j| = 1$, and $\alpha_{ij} = 0$ otherwise. This just means that a mixel is directly correlated only with its four cardinal neighbors. Because the neighbors are correlated with *their* neighbors, etc., this also indirectly implies long range correlation.

Figure 3 shows the well-groundedness of the above continuity preference. First, the central peak at 0 shows that mixels are in fact correlated with the four cardinal neighbors. The moments of this distribution suggests that a multivariate normal form for the prior,

$$P[\text{Mixels} \mid \text{Params}] = N[m_i \mid \overline{m_i}, \Sigma_{ij}] \prod_i dm_i. \tag{6}$$

is appropriate. Here, the matrix Σ_{ij} collects the α_{ij} dependencies. The $\overline{m_i}$ in Eqn. (6) represents putative "mean" values for a mixel at position i. The model we implement only uses a constant value $\overline{m_i} = \overline{m}$, the mean value of all mixels. Possible extentions could make use of varying means $\overline{m_i}$ to capture "trends" in an image.

Appearances may be deceptive, however. Figure 4 shows the same calculation as in Figure 3; although the shape looks fairly Gaussian, examining the moments of the distribution reveals that there is much more energy in the tails than is the case for a normal model. This is because individual pixels in Earth imagery can differ substantially from their neighbors, *e.g.* a road traversing ground.

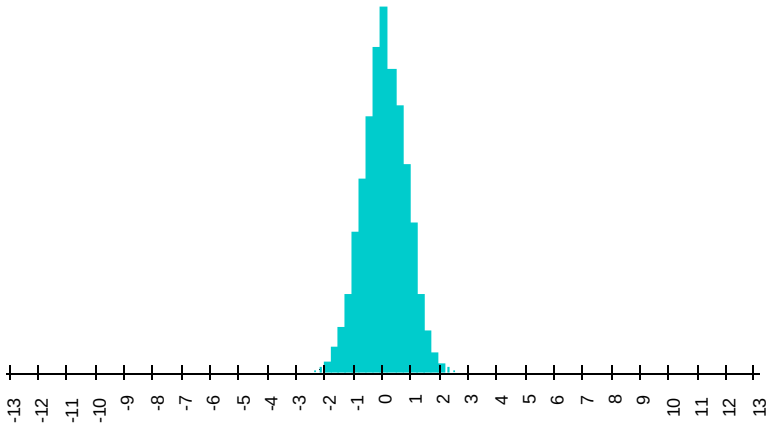


Figure 3: Distribution of $m_i - \hat{m}_i$ for a Viking Orbiter image of Mars. Horizontal axis is in 7-bit Data Numbers.

3.2 MAP Equation

Now that we are equipped with the prior

$$P[\text{Mixels} \mid \text{Params}] = N[m_i \mid \overline{m}, \Sigma_{ij}] \prod_i \Delta d m_i \tag{7}$$

we combine it with our likelihood

$$P[\text{Pixels} \mid \text{Mixels, Params}] = \prod_p N[\Phi_p \mid \hat{\Phi}_p, \sigma] \Delta \Phi_p . \tag{8}$$

Keeping in mind Eqn. (4) that

$$\hat{\Phi}_p = \sum_i \omega_{ip} m_i ,$$

we can rewrite the likelihood as

$$\prod_p N[\Phi_p \mid m_i, \omega_{ip}, \sigma] \Delta \Phi_p . \tag{9}$$

The MAP solution seeks to maximize the product of Eqns. (7) and (9), i.e. the "joint" distribution. Since they are both multivariate normal distributions, the joint is as well. Thus,

$$\begin{aligned} MAP &= Max(N[m_i \mid \overline{m}, \Sigma_{ij}] * \prod_p N[\Phi_p \mid \hat{\Phi}_p, \sigma]) \\ &= Max(N[m_i \mid \overline{m}, A_{ij}]) \end{aligned} \tag{10}$$

where

$$A_{ij} = \Sigma_{ij} + \frac{1}{\sigma^2} \sum_p \omega_{ip} \omega_{jp} . \tag{11}$$

The matrix A_{ij} in Eqn. (11) is calculated by standard methods in completing the square for multivariate distributions[10]. The peak of the distribution in Eqn. (10) are the m_i that satisfy

$$\sum_i A_{ij} (m_i - \overline{m}) = \frac{1}{\sigma^2} \sum_p \omega_{ip} (\Phi_p - \overline{m}). \tag{12}$$

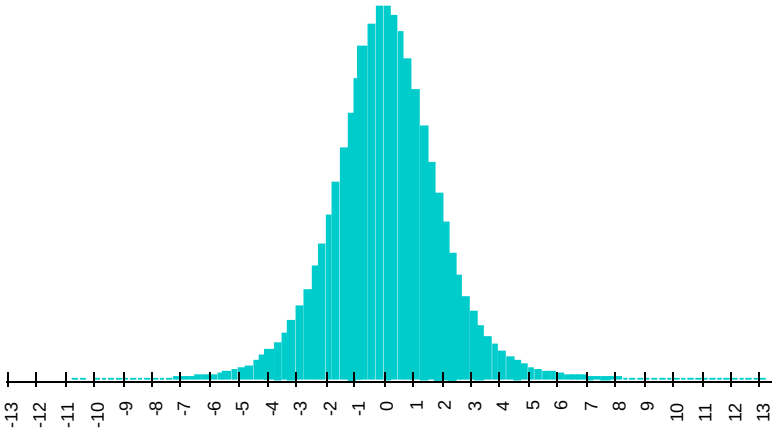


Figure 4: Distribution of $m_i - \hat{m}_i$ for a Landsat image of Kansas. Horizontal axis is in 7-bit Data Numbers.

We can thus find the maximum posterior mixel grid m_i given the auxiliary parameters by simply solving this linear equation. The method we use to actually compute the maximum posterior mixel grid is discussed in the next section.

4 Reconstruction Algorithm

We now concern ourselves with how to solve Eqn. (12). The fundamental "catch" is that *if* we knew the true values of the various parameters (PSF, registration, etc.), we could solve Eq. (12) exactly. However, to estimate the necessary parameters to high accuracy, we would have to know the true mixel grid! Our way around this dilemma is to iterate between two processes: use a current estimate of the mixel grid to (re-)estimate the parameters; use these new parameter estimates to re-calculate a better mixel grid, and so on. To start this process, we need either a nominal mixel grid or nominal set of parameters. Our data includes nominal values for parameters such as the camera location, the sun angle, etc., so we choose these as starting values. We begin our bootstrap process by estimating the registration parameters.

4.1 Registration

The registration parameters define the correspondence between points on the image plane and those on the modeled surface. The parameters can be thought of as coefficients to a function that projects surface points to the image plane; such a function depends on the planetary curvature, the imaging system's optics, and the camera's location and orientation relative to the surface. In the case of vidicon cameras, there is additional image distortion due to the read-out process that cannot easily be distinguished from the above geometric effects. Thus the projection function in principle varies for each point pair. The registration problem is to estimate the projection parameters for each image to the mixel grid that captures all of these components of the projection function.

As stated above, the strategy to estimate registration parameters first involves constructing a first guess at the mixel grid. To do this, we pick at random one of the pixel images and interpolate its pixel values onto a grid at the desired mixel resolution. Using this interpolated mixel grid as a reference image, we search for accurate relative registration parameters that map each image optimally onto the reference grid. But the quantization in the reference grid causes pixel-sized "jumps" in registration

values. This in turn creates a hazardous search process! To avoid these jumps, we smooth the reference image with a Gaussian-like filter.

In theory, we could find a MAP estimate of the registration (or any other) parameters; instead, we seek a simpler ML estimate and ignore priors on the parameters. This is because of the large ratio of information (pixels) to the number of parameters that need be estimated. If we assume an independent Gaussian likelihood for each pixel relative to its projected value from the reference mixel grid, as in Eqn. (3), then finding the ML estimate of the registration parameters reduces to finding the registration with the smallest sum of squared pixel differences from their projected values (i.e., a minimum squared error). In other words, the optimal registration parameters for an image gives the minimum squared error when the mixel values projected through the PSF are compared to the corresponding pixels.

There is one difficulty: moving features of an image "off the edge" of the reference mixel grid during registration. Clearly, image pixels not matched with anything shouldn't contribute to the total error; however, pushing hard-to-register features out of the picture is a false minimum! In our data we had available a larger image that contained the entire image set as subimages, and avoided the "edge" problem by processing the larger image as the reference.

Optimal registration parameters were determined by the Simplex algorithm [1], which searches for a minimum of the squared error by systematically varying the registration parameters, and then calculating the squared error for each such registration. (However, we assume that the error is a *smooth* function of the registration parameters. This is the reason for the Gaussian filter referred to above.) The algorithm stops when successive squared error values of the trial registrations are indistinguishable. We found that unless the registration search starts relatively close to the true registration (i.e., one has good nominal information), the search can get trapped in local minima. There are more efficient search algorithms than the Simplex algorithm, but they are not generally as robust. Note that standard methods for accurate relative image registration required locating "features" common to both images and finding a global mapping for all features to their counterparts in the other image [2]. The method described here uses *all* the information in both images, and this is part of the reason for the very high (subpixel) accuracy achieved by the method described here. However, feature based methods may be a good way of obtaining a close initial registration, when nominal registrations are not available or too inaccurate.

The affine transformation set is the parameter space in which the registration search is executed, and is sufficient for accurate registration provided that the principle nonlinear camera effects[3] are not severe. When these effects interfere, we have extended to a quadratic family of transformations.

4.2 PSF and Other Parameters

The point spread function (PSF) describes how the light energy from a point on the external surface is distributed over the image plane. The spreading of the surface point energy is usually due to the optical system's diffraction and aberration pattern. Typically, the PSF diameter is significantly smaller than the pixel dimensions, so that the images are not diffraction limited. With the scanning electron beam detector used in a vidicon, the PSF can be extended to model the diffuse *readout* spot as well. Since the PSF is a function of the imaging system, it does not depend on the particular image. In practice, the PSF can vary across the image plane, and with time. We have not attempted to model this variation, and work with an average PSF derived from the instrument's bench calibration [3].

"Shading" is the characteristic smooth variation in detector sensitivity across the image plane in vidicon tubes, equivalent to the variation of individual cell sensitivities in array detectors. The likelihood model must take shading into account, and can be learned from the data, given a rough idea of the registration: since all images contain the same subregion under similar lighting and viewing angles, any systematic differences in their appearance must be due to shading. We assume the shading function is a second order polynomial function of pixel position, and currently search

for coefficients which make the subregions have the most similar mean intensities.

Defects in the optical system or on the image plane generate blemishes — e.g., dust particles and scratches — common to all images from that camera. A blemish map is used to identify suspect pixels. Rather than interpolating the missing values as is common practice [4], we ignore these pixels, so that the corresponding mixels may be influenced only by the other frames. Also, since spacecraft that are many light-minutes away cannot be asked to retransmit corrupted data packets, they do not implement a reliable transport protocol, and some pixels have incorrect values. Usually no more than two bits are affected; our preprocessor uses this to help detect corrupted pixels. In principle it could use it to recover the correct value, but this would make little practical difference in our case. We simply ignore all suspected corrupt pixels, as well as missing pixels and reseaux marks[5].

4.3 Initial Composite

Once the above methods are used to find good initial estimates of the basic parameters (PSF, registration parameters etc.), we next construct a composite mixel grid using information from *all* the pixel images. We construct the value of a composite mixel by calculating the "votes" from every pixel that could affect it from any frame, as weighted through a *compositing kernel*. These "votes" are accumulated to give a total mixel value

$$m_i = \frac{\sum_p \omega_{ip} \mathbf{\Phi}_p}{\sum_p \omega_{ip}}$$

for each mixel, replacing the values of the reference grid. Clearly, those pixels that are nearest the projected position of a mixel have the strongest vote for that mixel. The compositing kernel functions algorithmically like a PSF, but needn't be the same function. For narrow kernels, the pixel-mixel "voting" is almost 1-to-1, but for diffuse kernels, each mixel value is the weighted combination of information from many pixels, leading to a "blurred" composite. In fact, if a small kernel is used that accurately models the actual PSF, and the noise content of the imagery is relatively small, this becomes a quick method for producing a super-resolved image.

4.4 Iterative Improvement

The composite is used as a starting point in a search for the MAP estimate of Eqn. (12). We use a standard iterative method (Jacobi's method) to solve the matrix equation. The Jacobi method solves an equation of the form $Ax = b$ by triangular decomposition

$$A = L + D + U$$

and updating

$$D \cdot x^{(r)} = -(L + U) \cdot x^{(r-1)} + b$$

which, if $\Delta x = x^{(r)} - x^{(r-1)}$ can be rewritten

$$\Delta x = D^{-1}(b - A \cdot x^{(r-1)}) \tag{13}$$

This can be shown to be equivalent to the method of "substitution", a useful fact for extensions of the model. To implement Eqn. (13), note that

$$D = \frac{1}{s^2}(1 + \sum_j \alpha_{ij}^2) + \frac{1}{\sigma^2} \sum_p \omega_{ip}^2 \tag{14}$$

[5]These are permanent marks on the camera faceplate used for calibrating the optics in the Vidicon camera[3].

is the denominator of Eqn. (13). Combining Eqn. (14) with Eqn. (13) for the numerator, one obtains the following iterative mixel re-estimation formula.

$$\Delta m_i = \lambda \frac{\frac{s^2}{\sigma^2} \sum_p \omega_{ip} (\mathbf{\Phi}_p - \hat{\mathbf{\Phi}}_p) - (m_i - \hat{m}_i) + \sum_j \alpha_{ij} (m_j - \hat{m}_j)}{\frac{s^2}{\sigma^2} \sum_p \omega_{ip}^2 + 1 + \sum_j \alpha_{ij}^2} \tag{15}$$

The results of applying this iterative formula to initial composite mixel grids is shown in Figs. 4.6 12); a noticeable sharpening of the composite is demonstrated. When the mixel grid resolution is too coarse, the mixels are overdetermined by the pixels, so the MAP mixel estimate is essentially the same as the ML estimate. In Eqn. (15), this means that the first (data) term in the numerator dominates the other two (mixel neighbor correlation) terms. When the mixel grid resolution is large enough (underconstrained by the pixels), the two terms in the numerator balance each other—i.e. the data term tries to force the mixels to exactly agree with the data, while the mixel neighbor term tries to make all mixels look like their neighbors ("smoothing"). It is the tension between these two effects that leads to plausible images, even when the mixels are underconstrained by the data. This is the case where the prior term acts as a regularization term.

In Eqn. (15), all the necessary parameters (s, σ, and the registration, PSF, etc. parameters that go into ω_{ip}) are assumed known. The λ parameter regulates the amount that any mixel can change, and is there purely for purposes of numerical stability. Some of these parameters, such as the PSF, are often well known ahead of time. Other parameters, such as the registration, can be initially estimated from an interpolated version of a single image. Since we find a much more probable mixel grid as a result of compositing and iteration, we can then re-estimate these parameters, and even repeat this convergence cycle. Fortunately, this re-estimation is not needed in practice more than twice. The reason for this is that parameters, such as the registration parameters, are typically estimated from thousands of pixels in the interpolated initial mixel grid, and so are already very accurate.

The ratio s^2/σ^2 of mixel to pixel deviation is more difficult estimate, as the most probable value can be many orders of magnitude different from what one estimates from a composite. We initially intended to re-estimate these parameters *during* the iterative convergence cycle from the residual error in each new mixel grid. What we did not realize that in some cases this dynamic re-estimation would result in *diverging* from the correct answer. So now when prior information is not enough to set these parameters, we must resort to an explicit search. We take a small but hopely representative patch of an image and seek parameters values which maximize our quality measure, the determinant of the matrix A_{ij}.

4.5 Complexity

One may ask why an iterative method, like the one above, was chosen over an algebraic computation of Eqn. (12). Essentially, the former method has a lower computational complexity. To keep the comparison clear we will stick to the "just constrained" case, which is described as follows. If we have f frames, with p pixels per frame, the number of mixels N is set to the total number of pixels P:

$$N = P = fp$$

It is "just constrained" because the number of data P is equal to the number of parameters N. Let k denote the radius of the point spread function in mixels. Then each iteration step of Eqn. (15) is of complexity $\mathcal{O}(k^2)P$, whereas the complexity of an algebraic solution of Eqn. (12) is $\mathcal{O}(k^4)P$. As a point of comparison, the compositing routine is of order $\mathcal{O}(k)P$, suggesting that finding an optimal compositing routine would be a good strategy for obtaining quick (if improvable) results.

4.6 Results

Fig. 4.6 gives results for a U.S. postage stamp digitized at low resolution by a scanner, and for Viking Orbiter images of Mars [5]. The Viking reconstruction uses a series of 24 vidicon images of Mars;

the data are from a high spacecraft altitude, with frames of very similar sun and camera angles. From a larger image, which also includes the edge of the polar cap, we extracted a rather underexposed 128×128 pixel regions containing the same four prominent craters. These regions represent the same area to within a few pixels. The images were preprocessed using the techniques described in section 4.2. Vidicon blemishes and telemetry noise were mapped and subsequently ignored; the shading response was modeled; image registration used only affine (as opposed to quadratic) transforms. Restoration was done at a 1:4 mixel scale (1:16 area ratio), making the restoration slightly overconstrained. We leave it to the reader to judge the restoration's quality.

5 Extensions

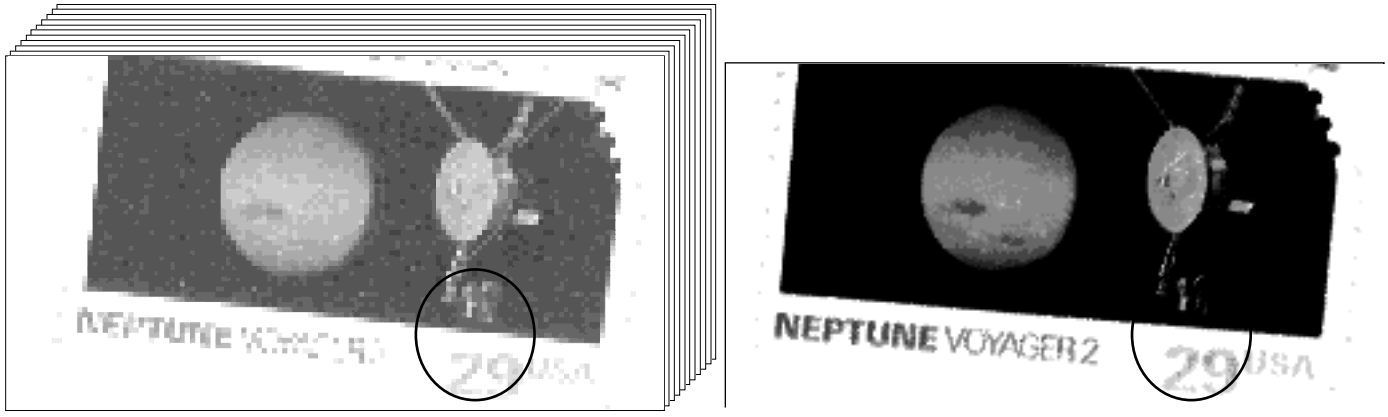
In the above surface reconstruction, we gave mixels a single scalar emittance value. In some applications, color information is available; e.g., Landsat/TM records seven spectral bands in each exposure, and Viking Orbiter and Voyager took gray-scale pictures through various color filters. The 2-D reconstruction described above can be used on each spectral band separately, to get super-resolved surfaces for each band. However, this approach ignores the fact that the surface features are often very similar across bands. A mixel having emittance values in each band could ensure even higher resolution if mixels are correlated not only with neighbors in a given band, but across bands as well.

In the above we combined the effect of albedo and ground effects into a single emittance value, which is appropriate if all the images are taken from essentially the same direction under the same illumination conditions. However, for most of Voyager and Viking data, there are many views of the same surface taken from different directions with different illumination. The theory described above can in principle be extended to handle this case as well. It requires the mixels to have albedo and height values; the registration process is similar, but with more parameters. We have derived the ML equations for the surface model assuming all of the lighting differences in the images are due to either slope or albedo, and not to shadows or occlusions. The effects due to slope and albedo can be distinguished because the effects of parallax vary independently of effects due to surface albedo. Note that representing the surface emittance by a single scalar (albedo) is an approximation that assumes Lambertian scattering. Many real surfaces are not Lambertian. Using bi-directional reflectance parameters, including a specular reflectance component, would give a more accurate surface model. The priors on the surface may involve properties such as continuity, smoothness, and texture. Additionally, we would need to model effects such as atmospheric attenuation, clouds, and the camera "hot-spot" for Earth observation data.

6 Relation to Other Work

The research reported in this paper was mainly motivated by attempts to integrate information from Landsat images taken on different passes. The difficulty here is that such images did not exactly overlay each other, so pixel-to-pixel comparison is not possible. A standard approach to this problem is "rubber-sheeting", which attempts to fit one image grid to another (reference) grid by resampling the first image onto the reference grid. Reference grid points are mapped, through an appropriate transform, onto the new image, and new grid elements are computed by taking an area weighted average of the overlain image pixels. The resulting resampled grid is perfectly aligned with the reference grid. The technique is extensively used to rectify and rotate Landsat and similar images to fit the geographical survey grid.

From the Bayesian perspective, the rubber-sheeting approach makes little sense, because the new averaged "pixels" are neither actual observations nor a surface model. Worse, the averaging process destroys information—it is impossible to recover the original image from the rubber-sheeted image. This information loss makes pixel-by-pixel comparison very dubious. The super-resolved surface

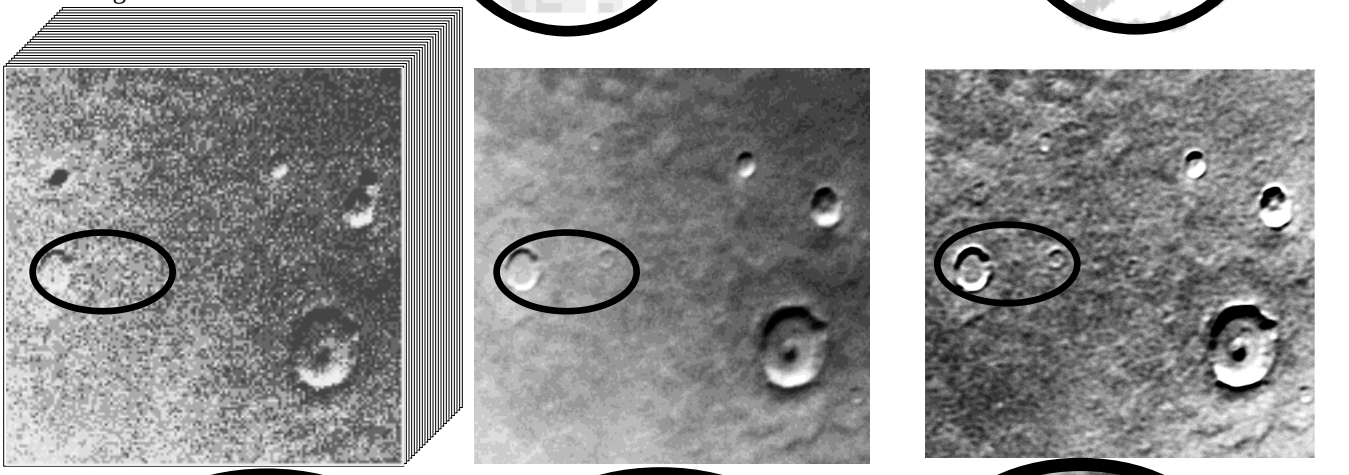


postage stamp scanned 63 times on an Apple scanner as an 8-bit grayscale image

scanned at 72 pixels / inch

reconstructed at 576 mixels / inch

extracted from Viking Orbiter images 217S42–88



742 meters / pixel

raw

composite

186 meters / mixel

iterated

186 meters / mixel

modeling described in this paper does allow the integration and comparison of information from many images through the accumulated super-resolution surface model.

A related approach to the Bayesian 3-D surface reconstruction described above is called "Shape from Shading" [6]. This approach integrates observed surface intensity gradients from a single image to give a 3-D elevation model of the generating surface. It assumes a constant albedo, known illumination conditions, and surface continuity. Shape from shading can be extended to multiple images [7], and the result is greater detail in the elevation map because each grid point contains information from multiple images. However, the constant albedo assumption is a strong limitation on the ability to extract information from multiple images.

A Bayesian approach very similar to ours is described in [8]. This approach does surface reconstruction using images from different viewpoints, and a neighbor correlation prior with a Gaussian noise model. The surface is represented by planar patches joined to form a curved surface. Unlike our work, these authors assume smooth large scale surfaces that can be represented by large parameterized "surface patches". Because these patches are estimated from many pixels from many images, the parameters that describe them are accurately determined, and so the overall surface is accurately estimated. In our approach we achieve super-resolution, and there is no aggregation of surface mixels into large scale patches. Although our goals and assumptions are significantly different we use the same basic Bayesian approach.

A related area of study is in combining images from video [11]. Here, the registration of images passes from the discrete to the continuous, and thus the techniques of "optical flow" are used.

Acknowledgements

We gratefully acknowledge the fruitful contributions of Chris Wallace and Wray Buntine.

References

[1] W. Press, B. Flannery, S. Teukolsky, W. Vetterling,*Numerical Recipes in C*, Cambridge University Press, 1988.

[2] R. W. Gaskell. *Digital Identification of Cartographic Control Points*, Photogrammetric Engineering and Remote Sensing, Vol. 54, No. 6, Part 1, June 1988, pp. 723-727.

[3] M. Benesh, and T. Thorpe. *Viking Orbiter 1975 visual imaging subsystem calibration report*, JPL Document 611–125, Jet Propulsion Laboratory, Pasadena, Ca., 1976.

[4] E. Eliason et. al. *Adaptive box filters for removal of random noise from digital images*, Photogrammetric Engineering and Remote Sensing, 56, 453-456, 1990.

[5] M. Carr et. al. *Archive of Digital Images from NASA's Viking Orbiter 1 and 2 Missions*, Planetary Data System, National Space Science Data Center, CD-ROM: USA_NASA_PDS_VO_1003, Frames VO217S42-88.

[6] B.K.P. Horn and M.J. Brooks.*Shape from Shading*, MIT Press, Cambridge, Massachusetts, 1989

[7] J. Thomas, W. Kober, and F. Leberl.*Multiple Image SAR Shape-from-Shading*, Photogrammetric Engineering and Remote Sensing, Vol. 57, No. 1, Jan. 1991, pp. 51-59.

[8] Y. P. Hung and D. B. Cooper.*Maximum a posteriori probability 3D surface reconstruction using multiple intensity images directly*, SPIE Vol. 1260 Sensing and Reconstruction of Three-Dimensional Objects and Scenes (1990), pp. 36-48.

[9] C. Elachi.*Introduction to the Physics and Techniques of Remote Sensing*, Wiley and Sons, New York, 1987

[10] K. Mardia, J. Kent, and J. Bibby. *Multivariate Analysis*, Academic Press, London, 1979.

[11] S. Mann and R. Picard. *Virtual Bellows: Constructing High Quality Stills from Video*, Proc. ICIP, Austin, Texas, Nov. 1994.