The 3D Recognition, Generation, Fusion, Update and Refinement (RG4) Concept.

David A. Maluf, Peter Cheeseman, Vadim N. Smelyanskyi,
Frank Kuehnel and Robin D. Morris.

NASA Ames Research Center, Mail Stop 269, Moffet Field, CA 94035
maluf@email.arc.nasa.gov

Abstract

This paper describes an active (real time) recognition strategy whereby information is inferred iteratively across several viewpoints in descent imagery. We will show how we use inverse theory within the context of parametric model generation, namely height and spectral reflection functions, to generate s model assertions. Using this strategy in an active context implies that, from every viewpoint, the proposed system must refine its hypotheses taking into account the image and the effect of uncertainties as well. The proposed system employs probabilistic solutions to the problem of iteratively merging information (images) from several viewpoints. This involves feeding the posterior distribution from all previous images as a prior for the next view. Novel approaches will be developed to accelerate the inversion search using novel statistic implementations and reducing the model complexity using foveated vision.

Foveated vision refers to imagery where the resolution varies across the image. In this paper, we allow the model to be foveated where the highest resolution region is called the foveation region. Typically, the images will have dynamic control of the location of the foveation region. For descent imagery in the Entry, Descent and Landing (EDL) process, it is possible to have more than one foveation region.

This research initiative is directed towards descent imagery in connection with NASA's Entry Descent Landing (EDL) applications. 3-D Model Recognition, Generation, Fusion, Update and Refinement (RGFUR or RG4) for height and the spectral reflection characteristics are in focus for various reasons, one of which is the prospect that their interpretation will provide for real time active vision for automated EDL.

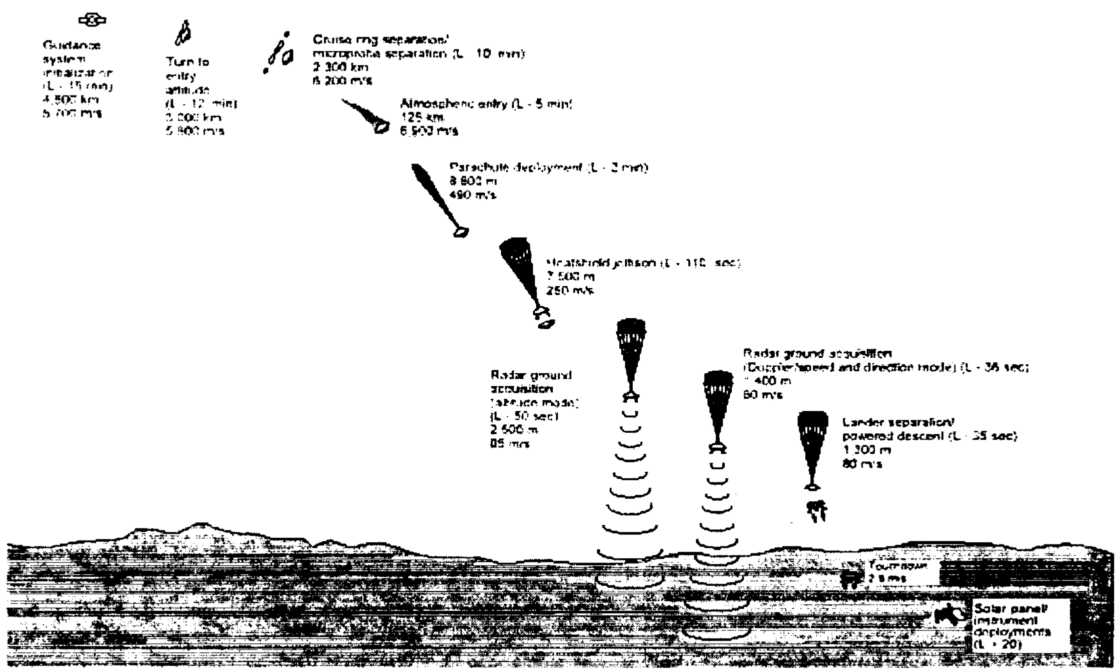
1 Introduction and Background

The period of the Entry, Descent and Landing is the missions most critical period with the highest risk factor for a potential Loss of Vehicle (LOV). Since distant missions such as Mars are constrained in payload and design, NASA must employ technology to intelligently use all available resources, optimally integrate sensor data and perform real-time decision and reason for successful Entry, Descent and Landing.

Understanding the importance of Entry Descent and Landing is best illustrated by describing the critical phases of an Entry, Descent and Landing process for a spacecraft. It is estimated that the spacecraft's descent from the time it hits the upper atmosphere until it lands takes no more than 4 minutes and a few seconds to accomplish the final landing as in the case of the Mars Polar Lander. Enabling technologies such as active vision can continually operate and integrate the vision system to actively interpret images for enhanced model recognition which can play a crucial role in mitigating major risk factors.

We estimate that the period where on-board intelligent systems can start capturing the landing site's topographic details starts about two minutes before landing and the spacecraft is expected to be moving at about 1,000 miles per hour around 5 miles above the surface.

About 70 to 100 seconds before landing a landing radar will be activated. To this end, we anticipate to having our proposed 3-D Model Recognition, Generation, Fusion, Update and Refinement (RGFUR or RG4) to include radar readings and other sensor modalities (gyros and inertia guidance). The radar will be able to gauge the spacecraft's altitude about 40 seconds after it is turned on, at an altitude of about 1.5 miles above the surface. With a robust RG4 system, the spacecraft can rely on the on-board camera for final touch down.



Entry, descent and landing

2 Similar Work and Comparison

Johnson's work described in [10] addresses the problem of autonomous operation close to a small body. The work described in our paper differs from, and is an advance over, the work in [10] in a number of ways. In this paper we argue for a unified model of the surface of interest, with all observations aimed at building up knowledge of this model, in contrast to an approach that builds up a model piecewise and in a manner dependent on the detection of features in the images. We also propose doing absolute location relative to the entire surface model, an approach that is much more robust and accurate than location relative to a small number of landmarks. It also does not rely on the presence of explicit landmarks on the object, but instead uses the entire surface essentially as one, extended landmark. Finally, the approach we advocate gives explicit uncertainty estimates of the surface and position; the work in [10] provides uncertainty estimates by running Monte Carlo simulations. After all, a typical risk associated with the landing process is to be able to resolve the surface to the level of details and be capable of avoiding a boulder, a ditch or a crack which could result in a Loss of Vehicle (LOV).

3 Research Objectives

The ambition of this paper in active vision is to continually operate and integrate a vision system that can actively interpret images for enhanced model recognition. The proposed approach exploits super-resolution techniques [3][4] and focus of attention (foveated vision) to enable better model recognition in descent imagery.

This research initiative is directed towards descent imagery in connection with NASA's Entry Descent Landing (EDL) applications. 3-D Model Recognition, Generation, Fusion, Update and Refinement (RGFUR or RG4) for height and the spectral reflection characteristics are in focus for various reasons, one of which is the prospect that their interpretation will provide for real time active vision for automated EDL.

4 Model Recognition, Generation, Fusion, Update and Refinement (RG4) and Super-Resolution

We are investigating a Bayesian model-based approach to integrating information from multiple images of the same area into a unified model at a resolution higher than that of the contributing images (super-resolution). This model is a representation of the physical parameters describing the surface. The physical parameters we use are heights at each grid point and the surface reflectance properties at each grid point, such as albedo (for a Lambertian reflectance model) or more generally a parameterized bi-directional reflectance distribution function (BRDF). Each image is an independent sample of the area of interest, and by combining the information from these separate images, surface features smaller than the image pixel scale can be captured. Because the model is constructed at finer resolution than any image, it is possible to use it to accurately project what that surface would look like from any view point, under any lighting conditions. This projection is computed by summing the contribution from each surface patch onto each synthesized image pixel, weighted by the camera point spread function (PSF). This projection process is called rendering in computer graphics, and the realism achieved by current computer graphics indicates the viability of accurate image projection from a surface model.

The essence of super-resolution in RG4 is to use Bayesian inference to invert the image rendering process. That is, in rendering, the surface and its reflectance properties are assumed known, as is the location and properties of the camera and the lighting source (typically the sun), and this information is used to generate an image under those conditions. In the Bayesian model-based inference process, the rendering process is reversed. That is, given the images, we find the most likely surface that would have generated them.

The model would consist of a discretized grid covering the area of interest, where each grid point stores the geophysical parameters of the corresponding ground location. These parameters mainly include elevation and reflectance spectral characteristics. This model is chosen so that what the camera is expected to see can be projected from the model. Model update consists of comparing the expected pixel values with the observed, and changing the model to better fit the data (including previous data). This update will be accomplished by computationally efficient Bayesian inference that inverts the image rendering process as used in computer graphics. The search for the most likely surface will be performed by a novel type of gradient descent, where the gradient is computed analytically.

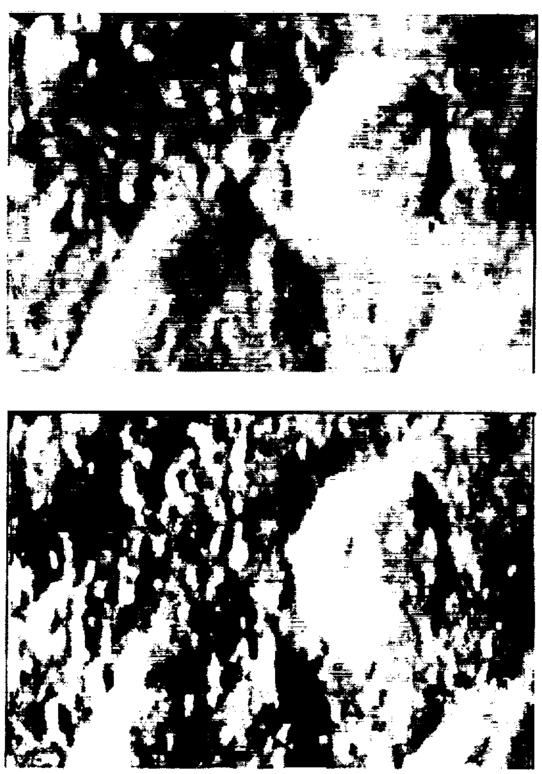


Figure 1: Top image is one of the two images taken from Clementine imagery to super-resolve the image on the bottom. With two images only (similar to the one on the left), the right image contains more detailed features (The right image is inferred from 3D model).

NASA has developed this process of model-based inversion over the last few years, starting from the simple 2-D models, and working up to the full 3-D surface reconstruction problem [3][4]. We are now able to super-resolve the heights and albedos of the true surface from multiple images, where the images can be taken from any viewpoint and under any lighting conditions. On artificial images generated from the model, we are able to reconstruct the surface to essentially the noise level of the data.

4.1 Research

Super-resolution is a very useful product for the Entry, Descent, and Landing process where the resolved model is beyond what can be extracted using the best available image. The main reason for developing the super-resolution capability is to allow the integration of information from different images without the problem of aliasing and mismatched pixel grids. Super-resolution solves this problem because any pixel maps onto many ground points, so that intensity of any pixel can be accurately computed by summing up the corresponding ground points. In fact, the surface model becomes the repository of the pixel's information, so that a system does not need to have multiple images persistent in its memory, but rather a model. EDL processes and post processes will thus interact with the surface model, and can view it from any direction or under any lighting conditions, including viewpoints that were not originally available!

In implementing this research, we extend 3-D super-resolution algorithms to solve a number of technical problems that arise in this application. In particular we will find workable solutions to the following problems using the approach outlined below.

Figure 2: Top image is one of the twelve synthetic images of Silicon Valley area used to super-resolve the second image. With twelve images only, the right image contains crisp detailed features. The bottom plot is the surface inferred from the images (not shown is the albedo field).

Shape from Motion

A main objective of RG4 is to achieve a surface inference in "real time". To that extend we obtain a fast shape from motion alogrithm which can feed itself as prior knowledge. Standard "shape from motion" algorithms [1] maintain the assumption of constant surface reflectance properties and are not extendable in nature to super-resolution. We plan to use our new shape-from-motion technique to "bootstrap" a super-resolution inference for natural surface formation where varying albedo properties and shadows are correctly accounted for.

Multi-Spectral Integration

EDL on-board instruments have multiple spectral bands will have different coverages, i..e. different widths and ranges. Our approach to solving the problem posed by integrating this heterogeneous information is to consider the model's surface by a wavelength dependent reflectance. That is, instead of a single number to represent the (Lambertian) surface reflectance for a particular band, we will represent the reflectance as a "smooth" function of wavelength, where the function is represented by a small number of coefficients that are estimated from the data. This function can then be integrated with each band spectral response function (a property of the instrument) to get the expected reflectance for that band.

Super-Resolution

One of the major achievements in this research is the method to achieve a recursive linear minimization as part of the desired inference for three-dimensional surface reconstruction to the extent that the resolution of inferred surface mesh is higher than the spatial resolution of input images. This technique also allows images to be super-resolved in both two or three dimensions (according to the nature of the data).

Accelerated Search

In statistical inference scheme, the solution for the gradient step in linear minimization for large sparse linear systems for which direct methods such as Conjugate Gradient is expensive in terms of both time and storage cost. For the class of descent imagery problem of using Bayesian inference for 3-D model parameter estimation, we plan to use a novel iterative technique which solves the problem of search minimization efficiently in terms of storage and memory cost. This novel technique takes root in a recent discovery for a model prior which reduces the covariance matrix complexity from a quadratic to a linear representation. As a result, the amount data will be relatively linear to the size of the model which is essential especially in a scarce computing environment.

Foveated Vision

We also support a Foveated Vision capability with variable resolution--that is, the surface triangles may be very small in some areas (super-resolved) and very coarse in other areas (under-resolved). The primary value of foveated vision is in the model reconstruction where high resolution information is transmitted in the regions of the image that are selected as important. On the other hand, low resolution information is processed at a second stage under contraints (e.g. time and computing resources). Foveated vision is crutial in descent imagery and will enable control in the resolution of pixel/model relationships.

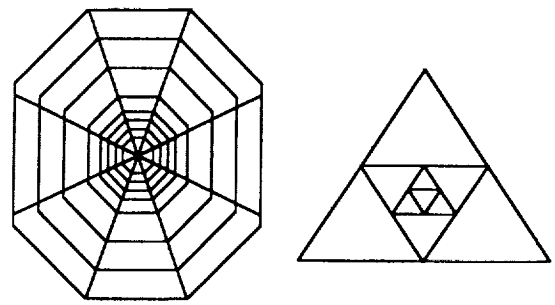


Figure 3: Left image is our planned 'spider-web' type mesh with a foveated center (not necessary centered in the middle). Right image is a typical non-uniform grids.

We extend 3-D surface models to foveated models using traditional triangulated surface, but distribution of the heights would no longer be tied to a uniform grid but to Foveated model (Figure 3.a). This extension is not difficult in principle, but the changed representation affects triangle indexing, and so affects efficiency.

4.2 Active Recognition: Concepts and Technical Aspects

The key idea behind active recognition in a sequential recognition strategy is that of improving interpretation by accumulating evidence in real time. The important aspect in the Entry Descent and Landing recognition problem is to compute on-line a 3D model from sensory data linked to the different sensor hardware which support the different phases in descent process (e.g. different cameras, FOV, RADAR, LADAR, altimeters, gyros etc.).

It is clearly understood that the image resolution in the early stage does not guarantee enough information either for quantitative or for qualitative model recognition. But acquiring uncertainties serves to condition prior expectations about the model and establishes a quantitative representation.

Practically, a meaningful qualitative recognition for a 3D-model reconstruction can be achieved after only a few sequences of images have been collected. To achieve a quantitative recognition the 3D model recognition is optimally obtained by computing the probability of the 3D model given the image sequences or

$$p(h,\rho \mid I_1 \ldots I_n)$$

Where h_i and ρ_i are the parameters of a height field and an albedo field.

At different stages along the descent process, image sequences with small frame-to-frame camera motion can be treated actively to provide an early 3D model. This real time behavior leverages from small motions, which minimizes the correspondence problem between successive images and the knowledge of the camera trajectory. However, this sacrifices depth resolution of the small baseline between consecutive image pairs [9]. Solution to this problem is trivially sought through a probabilistic incremental integration (e.g. Kalman Filter). In this particular active recognition, we will employ a matching and extraction technique which takes advantage of the lateral motion of the camera and transforms the search problem to a one-dimensional search problem (search is limited to foveated region).

Shape from shading-derived techniques provides gradient vector fields of the surface $\nabla h(r)$ and can be readily obtained in "real-time" from a single image source under very simplifying assumptions. Our approach is to reconstruct the height field $h(r)$ without the knowledge of the boundary conditions, which are directly obtained by the other sensor modalities and; in particular, the radar readings at a later stage. With single radar readings (initial condition), the height field $h(r)$ is readily reconstructed.

5 Recursive Super-resolution

Our current and existing super-resolution system can address many problems: the images may be of differing resolutions (e.g. multiple concurent cameras); the surface albedo is not assumed constant; the density of the model is user and data driven.

The model that we are trying to infer is defined to be the topology and reflectance properties of the surface being observed. For simplicity we define the surface over a grid of points, and currently define a height value, h_i and an albedo value, ρ_i at each grid point. Bayes' theorem then states that to infer values for the heights and albedos from the image data, we use the expression

$$p(h,\rho \mid I_1 \ldots I_n) = p(I_1 \ldots I_n \mid h,\rho)\, p(h,\rho),$$

which states that the *posterior* distribution of the heights and albedos is proportional to the *likelihood* – the probability of observing the image data, I, given the current values of the heights and albedos – multiplied by the *prior* distribution over the model.

To the extent of super-resolution, we make the assumption that the likelihood is due to zero mean Gaussian errors between the observed images, I, and the images synthesized from the model,

$$p(I_1 \ldots I_n \mid h,\rho) \propto \prod_i \operatorname{Exp}[-\tfrac{1}{2}(\frac{I_i - \hat{I}(h,\rho)_i}{\sigma_e^2})^2]$$

$\hat{I}(h,\rho)$, resulting in the likelihood being

where the product is taken over all pixels in all the images in the data set. The prior used is based on penalizing the curvature of the surface. It is a penalty encouraging a continuity in the inferred surface.

Because the likelihood is a function of the images synthesized from the model, it is clearly a non-linear function of the heights and albedos, and this makes optimizing the posterior distribution difficult. However, we have found that an optimal solution to the nonlinear function can be obtained by a novel Conjugate Gradient (CG) search.

We expand $\hat{I}(h,\rho)$ about the current estimate, h_0, ρ_0, and replace it by

$$\hat{I}(h_0,\rho_0) + \mathbf{D}\begin{bmatrix} h - h_0 \\ \rho - \rho_0 \end{bmatrix}$$

where $\mathbf{D}$ is the matrix of derivatives evaluated at h_0, ρ_0

$$\mathbf{D}_{i,j} = \frac{\partial \text{pixel}_i}{\partial \text{height(or albedo)}_j}.$$

The minimization of the log-posterior then becomes the minimization of a quadratic form, and can be performed using the conjugate gradient method. This minimization finds the minimum of the local linear approximation. At the minimum, we recompute $\hat{I}(h,\rho)$ and $\mathbf{D}$ and minimize the log-posterior iteratively.

5.1 The RG4 system: embedding stronger prior

For an Entry and Descent real-time process, strong prior about the surface model is highly desirable and therefore we plan to extend the super-resolution technique to include the shading information ∇h_s. Bayes' theorem then states that to infer values for the heights and albedos from the image data as well from the slopes, we use the expression

$$p(h,\rho \mid I_1 \ldots I_n) = p(I_1 \ldots I_n \mid h,\rho)\, p(h,\rho)\, p(\nabla h_s - \nabla h)\, p(h_s - h)$$

Here, it shall be remarked that h_S and ∇h_s are independent prior information obtained separately (i.e. shape from motion and image to surface gradient mappings). Furthermore, h_S will be obtained directly from a fast shape from motion method. Using the form of the prior in the previous equation makes it feasible to account for uncertainties in the independent measurements of h_S and ∇h_s. In addition, we plan to use the prior h_S to integrate the radar and other altimeter readings whenever they become available. We therefore "bootstrap'" the inference of the actual height field and albedos. Potentially, this leaves us with the advantage of rewriting the Bayesian inference process on the deviation (fluctuation) between the prior and height field rather than the height field itself, thus the parameters will be

$$\delta h = h - h_S,$$

and are believed to be small, such that a fast convergence of the inference process can be guaranteed.

6 Final Remarks

An operational software system based on this proposed demonstration system would use images to update the surface model as soon as they are received. The Bayesian approach gives a solution to the problem of how much the prior model should be believed when the new data disagrees with the prior model. Not only does this allow model update when there is conflicting information, but it can also serve as a change detection warning system. This is possible because the model projects expected values. If measurements are many standard deviations from expectations, then it is a signal for likely change.

Another planned operation of the RG4 system is to use the constructed model as a topographical map after the landing phase. The super-resolved model can be employed to focus the desired exploration phase of the mission. Models constructed from altitudes will provide a much wider scope in the landing site topography.

References

1. B. Horn and M. Brooks. Shape from Shading. MIT Press, 1989
2. Z. Zhang, R. Deriche, O. Faugeras and Q.T. Luong. A robust technique for matching two uncalibrated images through the recovery of the epipolar geometry. Technical report No 2273, INRIA, Sophia Antipolis, 1994
3. Morris, R. D., Cheeseman, P., Smelyanskiy, V. N., Maluf, D. A., A Bayesian Approach to High Resolution 3D Surface Reconstruction from Multiple Images,Proceedings of IEEE Signal Processing workshop on Higher-Order Statistics, June, 1999, pp. 140-143.
4. Smelyanskiy, V. N., Cheeseman, P., Maluf, D. A., Morris, R. D., Bayesian Super-Resolved Surface Reconstruction, Computer Vision and Pattern Recognition, 1999.
5. J.D. Foley, A van Dam, S.K. Feiner and J.F. Huges, Computer Graphics, principles and practice. Second Edition, Addison-Wesley, 1990
6. M. Armstrong, A. Zisserman and R. Hartley. "Self-Calibration from image triplets". In Proceedings of the European Conference on Computer Vision, pp 3-16, 1996
7. Q. Zheng and R. Chellappa. Estimation of Illuminant Direction, Albedo and Shape from Shading. IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol 13, No 7, July 1991.
8. L. Matthies, T. Kanade, R. Szeliski. Kalman Filter-based Algorithms for Estimating Depth from Image Sequences, International Journal of Computer Vision, Vol. 3, pp. 209-236, 1986.
9. Y. Ohta, T. Kanade, Stereo by Intra- and Inter- Scanline search Using Dynamic Programming, IEEE Trans. PAMI, Vol. 7 pp. 139-154.
10. A.E. Johnson, Y. Cheng and L.H. Matthies, Machine Vision for Autonomous Small Body Navigation, In Proceedings of IEEE Aerospace 2000 Conference.