PHASE I NIAC FINAL REPORT

Research Grant # 07600-065

3D IMAGES FROM 2D IMAGES:

A STEP TOWARD ARTIFICIAL PERCEPTION SYSTEMS

H. JOHN CAULFIELD, PI
DISTINGUISHED RESEARCH PROFESSOR
FISK UNIVERSITY
1000 17TH AVE., N.
NASHVILLE, TN 37208

EXECUTIVE ABSTRACT

You, the reader, see a 3D world. Nature has equipped you with a brain desperate to see the 3D world so important to your survival. It contains very fast subconscious computational modules that assemble all evidence of three-dimensionality and present to your consciousness a 3D image. If you close one eye, the perception you experience does not go flat. It is quite 3D. So your brain can do something very rapidly that no computer can now do – produce 3D information from a 2D image. As there is no 3D information in a 2D image, your brain must inject that information using built-in and learned assumptions. If we are ever to produce a system that perceives the world as we do, we will need a "preconscious" module that produces a 3D image form a 2D image.

In work before writing the Phase I proposal, we showed that such a thing could be done. But we had done only simple objects at close range, and even those calculations required substantial human intervention. This was not a suitable basis for Artificial Visual Perception (AVP), but it was a good start. The goal of the Phase I effort was to determine if and by what means we might be able to accomplish this in real time (30 frames per second). Obviously, that precludes human intervention. The answer to that question we found in Phase I was a qualified "yes." We can do it in real time given multiple images of the scene from different perspectives – something that automatically happens as platforms and scenes move. The computations are still slow and difficult, but two things can save us

1. The computations can be partitioned among multiple processors and
2. We can count on Moore's law to help out in future processors.

This represents a massive advance from where we were before the Phase I effort, but we are far from having an AVP system. The 3D from 2D module will have to be automated and speeded up, other preconscious modules such as Artificial Color, optic flow, and shape classification must also be made real time. Then, somehow, these disparate reports form many preconscious modules must be bound into a single "conscious" perception. That cannot even happen in our proposed Phase II effort, but we can use Phase II to make AVP possible in a generation, say, by the year 2020. That is why we call our Phase II proposal Vision 2020.

WHERE WE STOOD BEFORE PHASE I

Before we even wrote the Phase I proposal, the PI and his colleagues at the subcontractor – Physical Optics Corporation – had shown several critical things

1. It was possible to use the almost-abandoned mathematics of catastrophe theory to analyze the information-lossy projection of a 3D scene into a 2D image;
2. Only the two stable catastrophes – fold and cusp – have any rational probability of occurring;
3. The locations, scales, and orientations of those two catastrophes provide a complete description of a 2D scene. That is, we can reconstruct the scene from those elements alone; and
4. We can reconstruct a slightly rotated scene from slightly rotated catastrophes to generate a stereo pair.

Your eye/brain is so anxious to see a 3D scene that if one of a stereo pair if greatly degraded relative to the other, you see a 3D scene with the quality of the better image. This offered hope that we could immediately start viewing some of the 2D images in NASA's massive archive in 3D. Furthermore, by controlling the rotation, we can get hypo and hyper stereo if desired. We anticipated and now anticipate with more confidence that this will serve NASA technical and public relations needs quite well. But, the ultimate system goal of Artificial Visual Perception remains the driver of our work and the aim of our Phase II proposal.

Shown in Fig. 1 is an indication of the sate of the art when we entered Phase I. The object is simple, the analysis is hard, but the result is a stereo pair. Feasibility was established.

BACKGROUND ON 3D, 2D, AND CATASTROPHES

Rigorous catastrophe theory (CT) is quite sophisticated mathematically and is a purely geometric theory, devoid of physical concepts such as illumination. Both factors weigh against it as a model of early vision. In this paper, both objections will be met. An argument will be presented that modern neural networks offer a biologically plausible way to do something very like ABC. And, classical CT will be broadened to become a photo-geometrical theory and no longer purely geometrical. We turn to that extension of CT now.

Starting with the 2D retinal image with coordinates (u,v) of the 3D scene, we added a third dimension W (more properly, it should be derived from three color values R,G, and B). This 3D space is now abstract and no longer purely geometrical. We also extended the input space from (x,y,z) to the abstract 4D space (x,y,z; B), where B is brightness. We can now describe the collapse of the 4D input space into the 3D image space by ABC's nonlinear transformation (see Methods). Next we describe each physical object in terms of its own body-centered *normal* coordinates (x,h) as described in [17]. These lead to ABC singularities or catastrophes in their *canonical* or normal form. In ABC, we can

deal with only the two stable Whitney catastrophes from among the 14 catastrophes listed by Thom and Arnold Figure 1 shows how both of those catastrophes, *cusp* and *fold*, have their 3D local structure projected into the 3D retinal space.

The ABC mapping from object-space, in normal coordinates, into retina-space, has the form:

$$u = F_1(\xi, \eta) \tag{1a}$$
$$v = F_2(\xi, \eta) \tag{1b}$$
$$W(u,v) = F_3(\xi, \eta; B) \tag{1c}$$

the (1a) and (1b) are CT geometrical projections, while Eq. (1c) is a physical formula, describing photometric relations between object surface (ξ, η; B) and retina space (u,v; W). We are not the first to write it, as it is somewhat obvious. It is employed, for instance, in shape-from-shading calculations. But what we do next is new. We apply catastrophe theory in this mixed mathematical-physical hyperspace. That describes not only real dimensionality and location of the object in direction to viewer but also the effects of illumination, shading and clor changes all four together. The result is not an actual third dimension but how it appears to viewer. After a number of rigorous mathematical steps, described briefly below, we can rewrite formula (1c) as a sum of two parts:

$$W(u,v) = \underset{(I)}{M(B)} + \underset{(II)}{g(x,h)} \tag{2}$$

there *M* is the regular (Morse) form, representing standard photometric projection [5], while *g* is a new singular form, representing all object surface singularities (i.e., cusp and fold (see Fig. 1)).

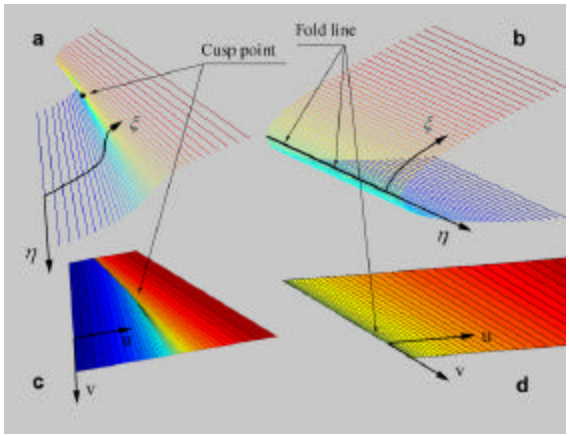

Figure 1. Illustration of Whitney's two stable catastrophes: cusp and fold; the first one described in normal coordinates as: $u = x^3 + h \cdot x$, $v = h$ and the second one by: $u = x^2$, $v = h$ We could observe that *fold* catastrophe can be identified with extrernal

boundary [8]. Both catastrophes can be presented either as 3D in (x,h) space, (a,b) or 2D intensity in (u, v) –space (c,d).

Being the generalization of Thom's geometrical lemma (21), the Formula (2) is the main result of the ABC-model. It demonstrates a surprising result that, in addition to regular photometric term (I), we also obtain the second, singular term (II) that *does not depend* on luminance-B. In order to be in agreement with vision perception that allow objects to be recognized independently of illumination, we decided to apply only term (II) (i.e., completely ignoring the regular term (I)) for image reconstruction: to our even greater surprise, we obtained quite good full scene synthesis (Fig. 2).

Fig. 2. The image on the right was reconstructed from catastrophes found in the image on the left.

A short description of the ABC computational algorithm can be done based again on Equation 2. We use two independent software programs - one for significant singular component extraction and the other for regular surface modeling. The first software program extracts highly visible, long lines of folds and cusp points. The result of this extraction is segmentation of the entire image into independent areas. Usually these areas represent a significant part of entire image. The second software program models a surface inside of the separated objects. That is, we model the jumps into different directions that represent folds as well as the smooth surfaces between them. Smooth surfaces are modeled by a least squares method and represent regular part of Equation 2. We use first and second order dependencies for the smooth surfaces.

This second program applied to results from the first a with zero-tree encoding algorithm gives compression results much better than JPEG and comparable to leading Wavelet algorithms (SPIHT). But scope of this article not description of the algorithm but similarity of this approach to processing procedure for human visual system.

Note that the information loss is not uniform across all parts of the scene. The scale factor has influenced the reconstructed image. The *dome* in Figure 2b is well preserved, while the background *leaves* are less well described. This is shown in more detail in Figure 3, where those specific scene portions are shown at a high DR. To give some sense of the quality of reproduction, we measure the *Peak-to-Signal-to-Noise-Ratio* or PSNR for those same two scene portions and various *a* or DR values, as in Figure 4. Objects become clearer as we attend closer to them. Remember that we did not define any object

a priori. The *dome* arose as an object in Figure 2 fully unsupervised. The *leaves* emerge as distinct objects as we decrease *a*. This is precisely the way human vision seems to work.

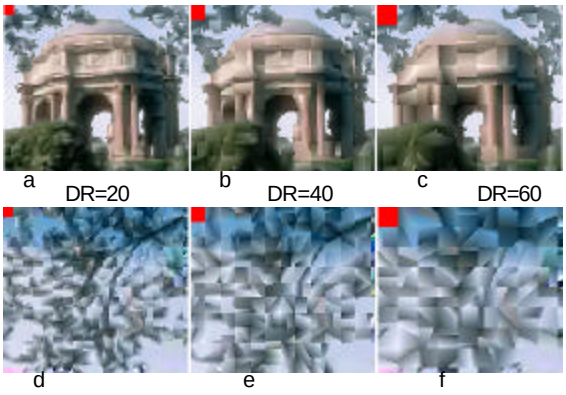

Figure 3. Similar to Figure 2b, with characteristic fragments of the scene, such as the *dome* and the *leaves* for DR-values: 20, 40, 60, and *a* –values: 0.016; 0.023; 0.031 respectively. We could observe that image quality of the *dome* is significantly better than that of the *leaves* for corresponding *a*-value (compare: (a) with (d), or (b) with (e)). This means that the ABC *non-linear* filtering provides an automatic segmentation process, by extracting an "object of interest" (here: the *dome*) by analogy to good paintings.

Impressionist painters achieved the effect of backgrounds with "too-high *a*" routinely. Unlike nature that is somewhat fractal with new information greeting every increase in resolution, impressionist paintings look realistic only from a particular distance.

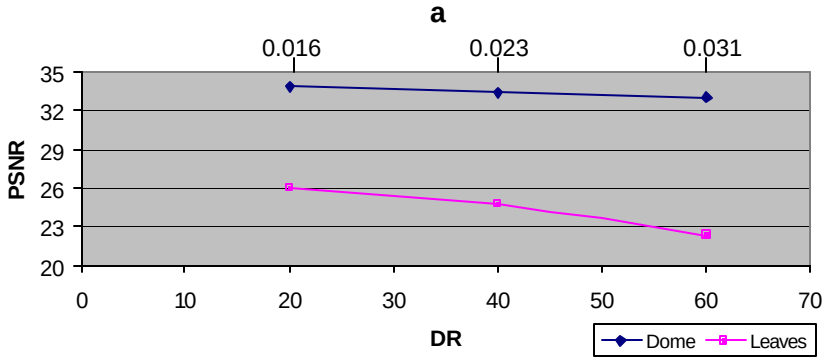

Figure 4. Peak-Signal-to-Noise-Ratio (PSNR) as a function of *Data Reduction* (DR), or Catastrophe Resolving Element's *a*-value (proportional to DR value), for two

characteristic fragments of the scene (Figure 1), as in Figures 3a, 3b, 3c (the *dome)*, and Figures 3d, 3e, 3f, (the *leaves*). We see that PSNR-values for the *dome* are always significantly higher than the corresponding PSNR-values for the *leaves*. The PSNR-values are defined as: $PSNR = 10\log_{10}\left\{\frac{255^2}{\frac{1}{N \cdot M}\sum_{i=1}^{N}\sum_{j=1}^{M}[d_{ij} - f_{ij}]^2}\right\}$, where d_{ij} is the pixel gray value (averaged over color), for data-reduced image, and f is its corresponding value for the original image, summarized over *N by M*-number of all pixels of the frame.

The most common way of producing 3D visualization is stereo display. Each eye is given a different view of the scene and the viewer's visual processing system causes him to perceive the 3D scene that would have to have been present to have caused those disparate views. With ABC, we can rotate the local catastrophes and thus create a view of the object from a different direction. The original and the rotated image (both derived from the original 2D image) constitute a synthetic stereo pair that allows a viewer to see the scene in 3D. Figure 5 illustrates the results of such a rotation. Obviously, we have assumed relatively small rotations that keep each catastrophe in both views. That often happens. Our method simply does not apply if the rotation and scene are such that our assumption breaks down.

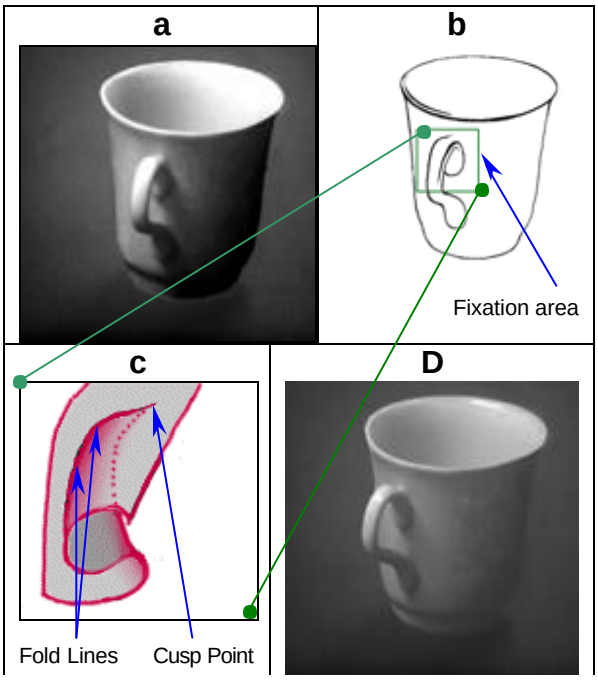

Figure 5. Demonstration of 3D nature of catastrophes on the basis of a simple photographic object, a *cup*, including: (a) data-reduced cup (DR=20:1); (b) cup's fixation area: an ear; (c) detailed illustration of catastrophes as 3D objects; a *cusp* area is shown only from one side; the second hidden side is shown as a broken line; (d) using 3D profile of both catastrophes, a cup, coded as in Figure 2a, has been automatically rotated by 2°, thus, demonstrating local 3D dimensionality of a monoscopic image in catastrophic representation (Figure 2a).

The presented analytic modeling of image analysis/synthesis demonstrates that high data reduction is possible, with only two primary elements – catastrophes. The primary elements are basic building blocks for any scene, any image, and any object. A question arises if such modeling can, in principle, be a basis for analysis of visual perception. In this context, we can observe that the ABC algorithm reduces membership of primary elements to an absolute minimum, an optimum situation from an informational point of view (a membership with only single primary element is rather an unrealistic scenario). Moreover, within the ABC system, entire 3D image geometry is reduced into the 2D intensity retina pattern, and such mapping is locally *isomorphic* (or even *homeomorphic*). This means that stereopsis would be a rather local phenomenon, well observable at larger distances.

Figure 1. Illustration of Whitney's two stable catastrophes: cusp and fold; the first one described in normal coordinates as: $u = x^3 + h \cdot x, \quad v = h$ and the second one by: $u = x^2, \quad v = h$ We could observe that *fold* catastrophe can be identified with extrernal boundary [8]. Both catastrophes can be presented either as 3D in (x, h) space, (a,b) or 2D intensity in (u, v) –space (c,d).

Attneave pointed out that there is much redundancy in natural images and suggested that the subjective prominence of borders provides an example of a psychological mechanism that takes advantage of this fact: you can represent an object more economically by signalling transitions between object and non-object because these are the unexpected, and therefore information-bearing, parts of the image. He illustrated with his famous picture of a sleeping cat that the same rule applies to the orientation of boundaries, for the picture was produced simply by connecting the major transition points in the direction of the border that outlines it. This is precisely the reconstruction process ABC represents. It locates the discontinuities (catastrophes) along with their orientation, scale, and size and uses that information to reconstruct the image. Thus, we argue, Attneave's work shows that humans use a process much like the ABC process discussed here.

In summary, we have demonstrated that catastrophe-based ABC is a possible model for vision perception, since we could not find any contradiction with equivalent neurobiological results, while the following features of visual cortex seem to agree with the ABC system: foveation search with singular fixation points; local matching

corresponding parts of two stereoscopic retina images; local 3D features of monoscopic images; 2D structure of visual cortex [22]; high data reduction [26]; modular and modestly–parallel visual cortex architecture [27]; highly-nonlinear and hierarchic feature extraction [26-28] leading to neural net models [31-32]; highly effective pattern recognition, highly-independent on illumination, shadowing, color, orientation and scale; and finally, excellent image quality reconstruction (synthesis) from highly-disperse singular elements.

It should be noted that a total number of singularities have been determined by only two factors: scale, represented by *a*-value, and their possible types (here are only two: *cusp* and *fold*). The remaining part of algorithmic procedure, including the reconstruction (synthesis) of a scene, has been done automatically, by algorithmic computing.

TECHNICAL ACCOMPLISHMENTS IN PHASE I

Task 1 Review of Shape Recovery Methods and Implementations

Within this task a comprehensive review of shape from shade and photoinclinimetric methods presently employed will be compiled with an emphasis on the utility to spaceborne and planet based imagery including the contributions of catastrophe theory to the problem. A broad range of differing approaches to the problem of the analysis of terrain and material since Rindflesch's seminal work on lunar profiles in 1961, the unique properties of many planetary surfaces and the predictable illumination patterns has led to the particular utility of these methods to planetary sciences. A synthesis of these methods and a reevaluation of the specific algorithms operating with respect to the recent emergence of easily implemented massive parallelism on the scale of individual ASIC chips and compact processor clusters. In addition the utility of the catastrophe based methods for quantification of boundaries and refinement of boundary conditions will be evaluated. In the same manner an evaluation of the methods for shape recovery from texture gradients will be evaluated in the same manner.

While the basic problem of extracting the gradient and integrating the resultant 3D information has been clearly defined in terms of a first order PDF in two dimensions a broad range of differing methods for the solution of this superficially simple problem have been applied in an attempt to obtain stable and consistent solutions. It is now possible for the significance of various methods to be evaluated as they relate to a consolidated vision system for probes and analysis.

Complexities associated with the difficulties associated with the determination of initial conditions and the presence of noise, compression artifacts, and quantization errors, stable and accurate convergence to a consistent value is often quite challenging. As a result a broad variety of methods have been attempted within the field. These various methods will be categorized and tested for the present application. In addition the implications of recent findings with regard to catastrophe theory prompts a need for a

reevaluation of the various methods employed in the light of the improved nature of the boundary conditions this method provides. Finally necessity of employing compression due to bandwidth limitations within the system it will be necessary to evaluate these methods in the presence of the various common forms of compression.

A comprehensive analysis of the methods for shape recovery by the use of stochastic analysis of image texture in a manner complementary to that of the shade to shape algorithm in order to allow for analysis of broken ground which is too finely textured as to allow for the recovery of shape from shading. Once again the specific conditions of the application will be applied to test the various methods reviewed. In addition, the effects of compression will be reviewed for each method considered.

Finally the methods will be reviewed in conjunction with various computer vision and image understanding methods. By viewing shape recovery within the broader context of the full range of computer vision tools it

As expected, we found that only the two stable catastrophes (cusp and fold) were necessary to sort out 3D information.

We selected a representative dataset for analysis and processing. In order to obtain a clear means of calibrating our measurement process we selected data from the Mars digital image mosaic (MDIM) with registered Mars orbital laser altimeter (MOLA) data. By employing these datasets it is possible to confirm the results of the our elevation estimates, as well as modeling us to model our lighting model.

Our initial dataset is the northernmost volcanic peak of the *Tharsis Montes* mountain chain on mars (near *Olympus Mons*), this area was selected since it predominantly consists of one major peak, and fills a 256x256 image space at 1/32 of a degree resolution. Figure 1 shows the location of the image within a topographic map of Mars. The specific image was extracted from the MDIM 2.0 dataset by the “Plot-A-Planet” website of the Planetary Data system (PDS). The specific information on the image is shown in Table 1.

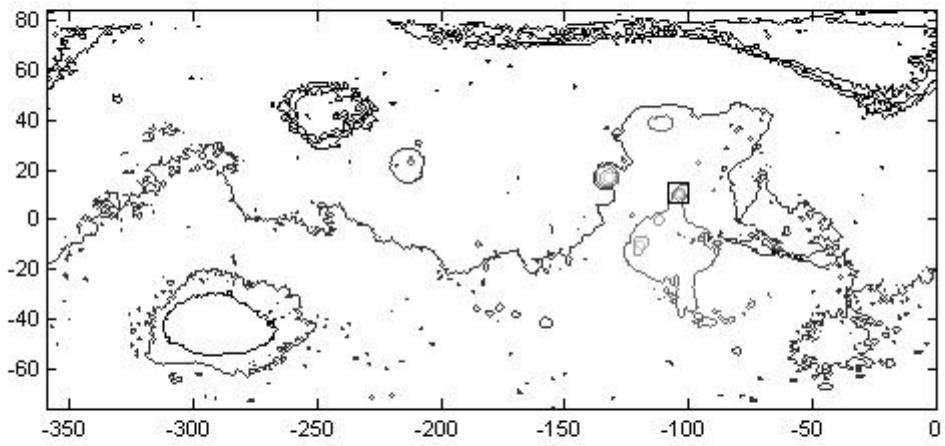

Figure 6. These are the raw data we used to compute views from various directions.

During this period we made considerable progress toward automatic extraction of catastrophes as suggested below. The critical operation is finding the points at which the surface normals are themselves normal to our direction of observation. We can compute this, given those data, from any direction of view in principle. That would allow us to reconstruct a view from that angle. Most but not strictly all spots that have normals normal to the view directions are, in fact, catastrophes.

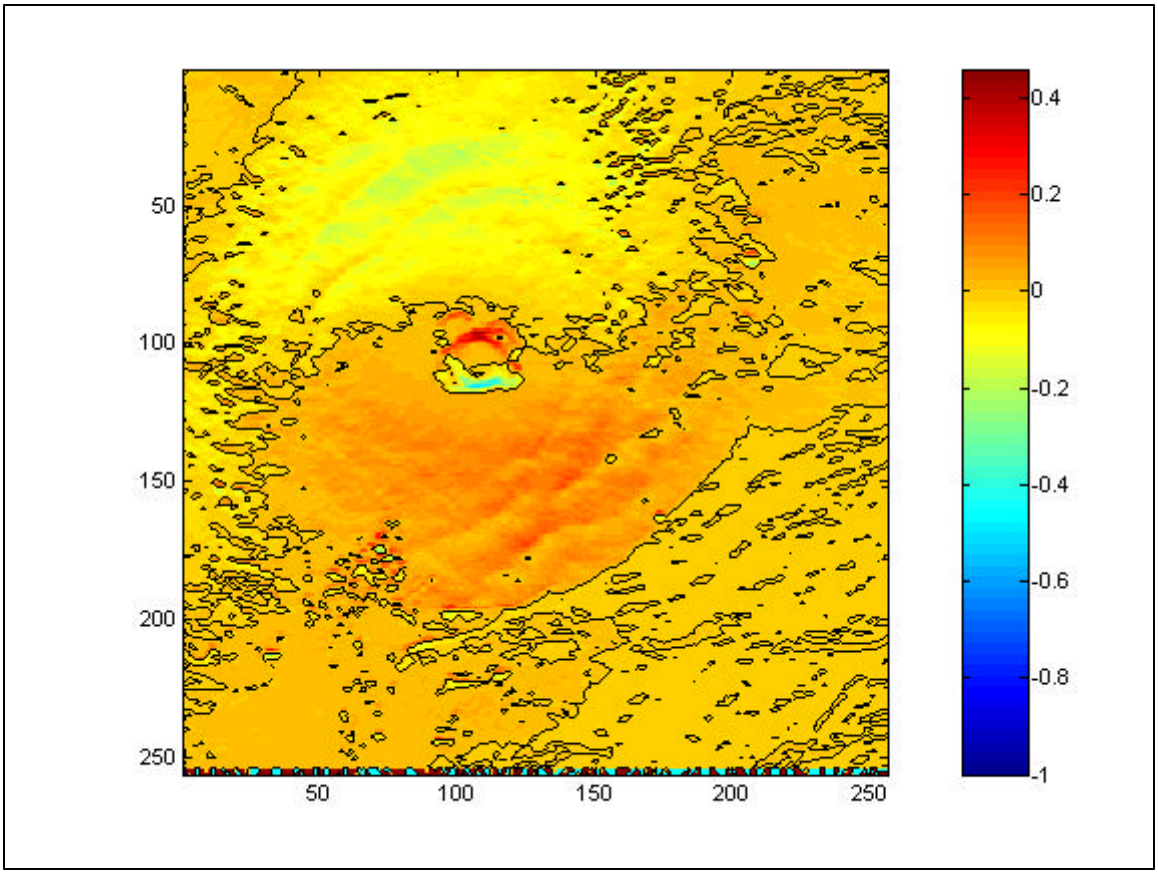

Figure 7. Mapping of dot product of normals of MOLA data with vector directed at [0,-1,0] color indicates value of dot product for each normal and black line indicates points at which the value is zero (i.e. a catastrophe).

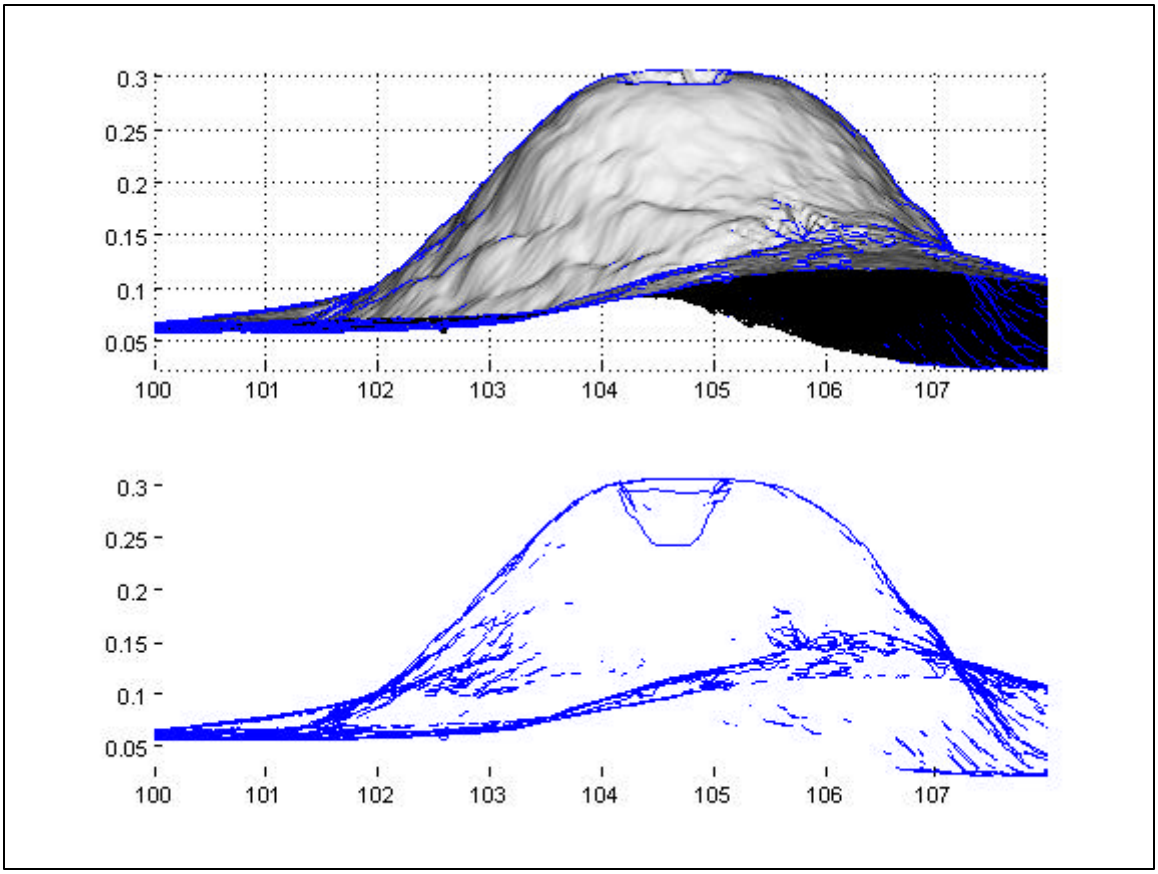

Figure 9. Upper picture: Plot of MOLA elevations (vertical dimensions exaggerated) as an orthogonal projection from the [0,-1,0] direction, catastrophes shown in blue. Lower picture: Catastrophes shown alone including those occluded from view in upper picture.

Task 2 Development of a multiluminant shape recovery system

The inherent difficulties associated with the use of simple monoscopic shape recovery discussed previously has led to the invention of a novel and highly innovative method of shape recovery the multiluminant method. The multiluminant method is an active imaging method that employs illumination by an array of modulated LEDs in conjunction with either a CCD or CMOS imaging device. By rapidly cycling between differing positions and wavelengths of the LED array it is possible to process out both ambiguities in shading and albdeo. By incorporating optical notch filters it is possible to operate the proposed system even under intense sunlight without interference. The basic principle is to obtain a series of images of the scene that we wish to perform shape recovery upon which differ only with regard to the angle and wavelength of the source employed. Images illuminated from a specific angle but differing angle are first processed to determine the differences between the intensities of the image between the images, the variation of brightness relative to the wavelength are employed to extract variations in albdeo and texture which are generally wavelength dependent as opposed to shading which is as a rule is independent of wavelength. Once this has been done, the consolidated image from each illumination angle is then compared. The scene under consideration is identical in each case but the shading differs. The variations in shading

and shadowing may then be used to resolve the ambiguities generally encountered within standard monoscopic shape from shading algorithms. In particular due to the presence of multiple shading conditions from known sources it is possible to unambiguously determine the gradient of the imagery when three or more illumination points are known. Once this has been done both ambiguities in shading and aldebo are eliminated and a reliable mapping of the slope of each surface has been computed. This technique differs from stereophotographic methods in that a single fixed camera (and thus a single fixed viewpoint) is employed. The method differs with regard to the use of structured light in that there is no need for a projector or reticule

It overcomes a difficulty encountered earlier in Phase I wherein we found that the information in a single image was insufficient to produce a reliable 3D image for many NASA purposes. Instead, we decided that (at least in some cases), we could control the images from a single camera by using multiple illuminants. This is a powerful new concept that should be especially valuable in one of our planned long-term applications – planetary exploration. See below for how we plan to insert this now into NASA's long term plans.

NONTECHNICAL ACCOMPLISHMENTS IN PHASE II

If Vision 2020 – our long-term goal of providing NASA a useful Artificial Visual Perception system by the year 2020 – is to succeed, it will need far more time and money than NIAC can provide. We must recruit interest in it both within and without NASA from researchers who can add the components we cannot and funding agencies that can provide the time and money. We believe that only NIAC is a place where long-term goals are taken seriously. The other NASA and non-NASA agencies need something tat can be done quickly. Every component of Vision 2020 must stand on its own and be independently valuable. Not surprisingly, that is the way nature built your visual perception system. It evolved many components independently and bound their results together. We were faced with the problem of getting NASA and other agencies interested in 3D from 2D.

The strategy we chose was to bundle this advance with several others to persuade NASA that a new era of optical telescopes is beginning. Since Galileo, optical telescopes have become bigger, optically better, located in better places (e.g. space), and recorded electronically rather than by eye. Those were evolutionary, small changes. Can we do something both radically different from anything Galileo ever dreamed and better in some measurable way than any extension of current telescopes will ever be? We thought so and thought that 3D from 2D was one approach. Accordingly, we put together a workshop to explore that concept. Ultimately, it will result in a widely read article. IEEE Spectrum is considering it at this writing. We offer a summary of the workshop conclusions here. It does appear that there is interest within NASA. But with a new administrator and no budget, no one will commit anything at this writing. On the other hand, Dr. Ravindra Athale at DARPA has expressed great interest and is working with the workshop attendees on fleshing out a program.

Of perhaps greater importance, the PI has gotten himself included in a NASA sponsored, NASA backed "Planetary Exploration Study" that aims at advising NASA as to how to explore Mars one quarter into this century. Some key NASA participants are

NASA Ames Research Center
Dr. Geoffrey Briggs
Dr. Butler Hine

NASA Johnson Space Center
Douglas Cooke
Brenda Ward

The concepts in our Phase II proposal that arose from our Phase I work are now being widely discussed within NASA and seem likely to impact a report NASA will take seriously in planning planetary missions. Therefore, we will be in good position to insinuate this work into future NASA activities and to insure that its impact will outlast the seed money put in by NIAC.

CONCLUSIONS

During Phase I we made some significant progress toward our long-term goals. In particular, we

1. Showed that 3D from 2D could be done under realistic circumstances in real time and
2. Found two extremely promising ways to continue this part of Vision 2020.

Those accomplishments are the basis for a Phase II proposal that will take Vision 2020 far enough to assure its continuation by other and its eventual success.