RADAR IMAGE PROCESSING WITH CLUSTERS OF COMPUTERS

Alois Goller[1] and Franz Leberl

Institute for Computer Graphics and Vision
Technical University Graz
Inffeldgase 16, A-8010 Graz, Austria
Leberl@icg.tu-graz.ac.at

ABSTRACT

Some image processing algorithms such as shape-from-shading are particularly compute-intensive and time consuming. If, in addition, a data set to be processed is large, then it may make sense to perform the processing of images on multiple workstations or parallel processing systems. We have implemented shape-from-shading, stereo matching, resampling, gridding and visualization of terrain models in such a manner that they execute either on parallel machines or on clusters of work stations. We were motivated by the large image data set from NASA's Magellan mission to planet Venus, but received additional inspiration from the European Union's Center for Earth Ovservation program (CEO) and Austria's MISSION initiative for distributed processing of remote sensing images on remote workstatuions, using publicly accessible algorithms.

We show that a combination of parallel and distributed processing can successfully cope at a very high throughput with large image data sets and complex algorithms. We have developed a multi-processor approach that we denote as CDIP for Concurrent and Distributed Image Processing. Throughput for image processing tasks increases nearly linearly with the number of processors, be they on a parallel machine or arranged in a cluster of distributed workstations.

Our approach has added benefits to a user of complex image processing algorithms: the effort for code porting from one computer to another and code maintenance is reduced and the necessity for specialized parallel processing hardware to achieve very high throughput rates is eliminated.

1. INTRODUCTION

As we experience the enormous changes caused by the Internet, we also see an evergrowing flood of data, particularly visual data, specifically in the field of satellite remote sensing, an increasing need to process large data sets quickly, thereby employing increasingly complex and demanding algorithms to perform very specific tasks.

When put together, these developments inspire initiatives to provide users with tools to take advantage of large remotre sensing data sets over the Internet and to employ processing algorithms provided by remote experts and operating on remote computers.

---

[1] Currently at Chalmers University, Gothenberg, Sweden at algo@chl.chalmers.se

When processing time is excessive, then it makes sense to employ particularly organized parallel computers, or clusters of computers to speed up the creation of the result. We have thus a motivation for the study of web-based image processing in remote sensing. Specifically we are reporting on work inspired by thoughts to reprocess the 400 Gbytes of image data from NASA's Magellan mission to planet Venus. However, one can see this Megellan-challenge as a model for things to come as part of the increasing data strea from the Earth Observation System EOS, and from the European Union's Center for Earth Observation CEO.

We have developed a concept and system to employ multiple computers, some of them parallel processing machines, to process large quantities of radar image data using very specific algorithms, and doing this at a very high throughput (Goller, 1999 a and b).

Many advantages result from sucessfully implementing such technology. There is not only the obvious decrease in elapsed time before a particular computing task is completed; there also are simplifications for the user who does not need, nor desires, to always understand where the data or software actually reside, or the increased ease of creating and supporting oftentimes complex software that executes best on a particular computing environment, and needs local expert support not available at a particular user's site.

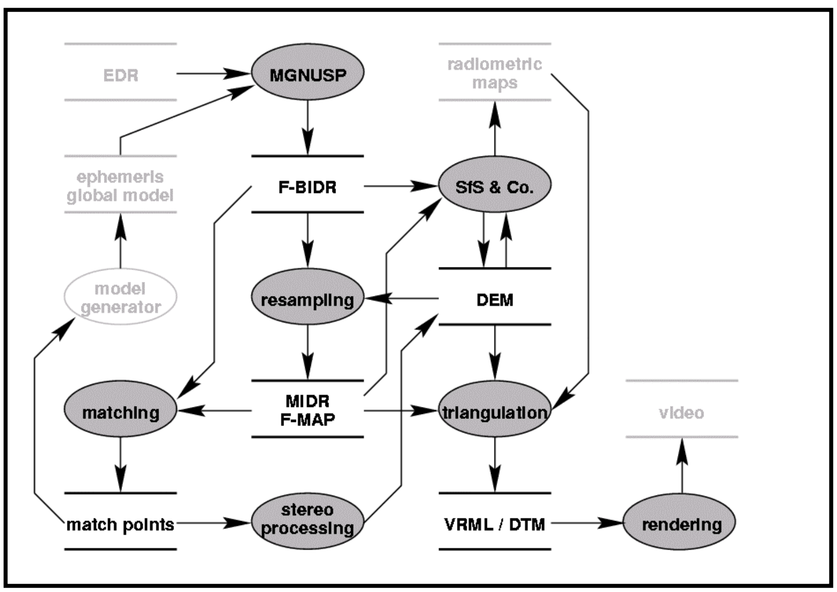


Figure 1: Key radar image processing algorithms for Magellan processing, beginning with signal processing into Full-Resolution Basic Image Data Records F-BIDRs, then producing DEMs and Mosaiced Image Data Records (MIDRs) and full resolution maps F-MAPs.

We show that the use of remote computers, either in clusters or specific parallel architectures, and remotely residing software and large data sets can successfully be combined into a system with linearly increasing throughput as the number of processors increases. It may take 2 hours to obtain a certain Digital Elevation Model (DEM) from overlapping radar images on one computer, and we may see this reduced to a mere 10 minutes using multiple computers.

2. ALGORITHMS

Processing of radar images begins with the image formation from the raw signals. Particualry time consuming radargrammetric algorithms are then shape-from shading (SfS, Thomas et al., 1991), stereo matching (Hensley and Shaffer, 1994), gridding of elevation models, resampling for ortho photos and terrain visualization and rendering. Our work has addressed all of these algorithms (Goller, 1999a), but focussed on shape from shading and stereo matching. Figure 1 illustrates the data flow through a system for processing radar images, in particular those from NASA's Magellan mission. Separately marked in Figure 2 are typical iterative data flows connected with the extraction of digital terrain models. A sequence of stereo matching and shape-from-shading produces the most detailed representation of the terrain. Figure 3 shows an example of stereo matching and refinement by the exploitation of shading variations.

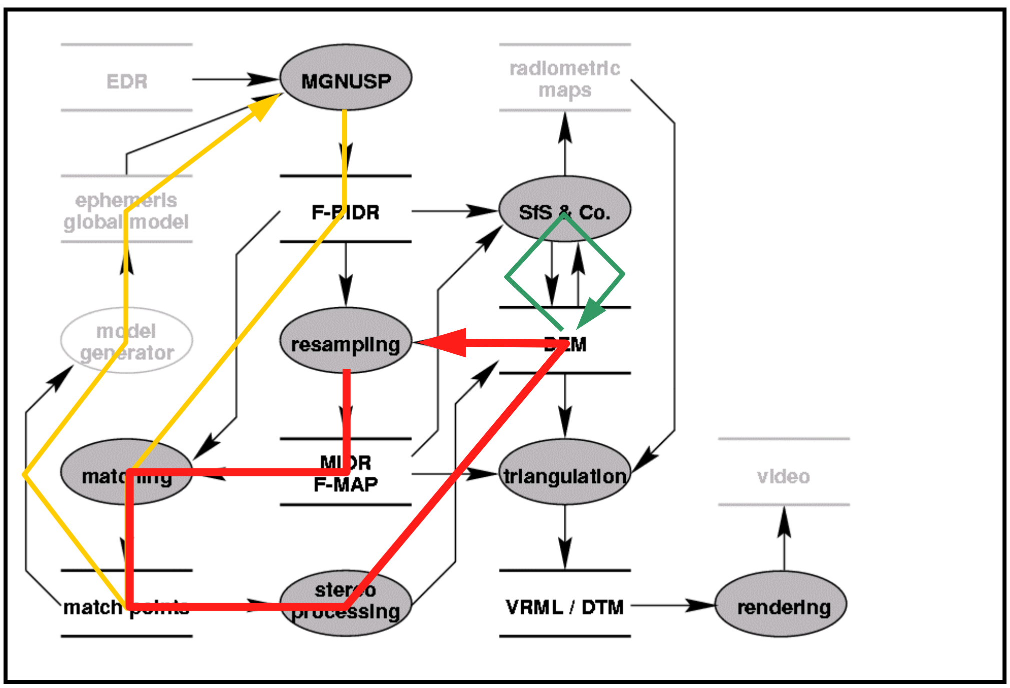


Figure 2: Adding to Figure 1 some iterative processing sequences. At the left is the sequernce that produces improved F-BIDRs after a DEM has been created. In the center is the sequence to refine the DEM by improving the input images ny a geometric resampling. At the right is the iterative use of shape-from-shading.

Table 1 summarizes typical performance and throughput times for some of the algorithms. It is to be noted that in the case of Magellan data, a systematic process will involve multiple images each of a size in the 100 Mbyte range, while some of the perfomance assessments are based on patches of 1k x 1k pixels.

MGNUSP SAR Image Formation	6 hours per F-BIDR
Resampling	a few minutes
Image Matching	several tens of minutes
Stereo Processing	a few minutes
Shape-from-Shading	1 hour
DEM Surface Triangulation	some minutes
Rendering and Visualization	from seconds to hours

Table 1: Typical image processing algorithms and their perfomance, typically on a single processor computer, in this case often an SGI Indy.

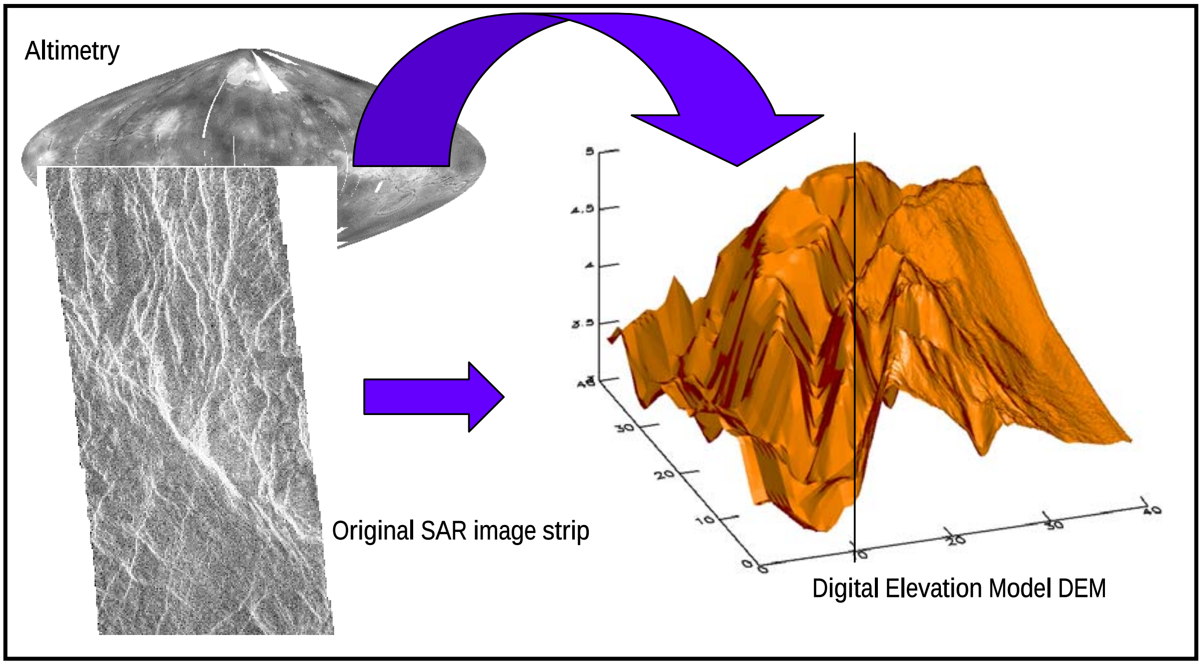


Figure 3: Concept of Digital Elevation Modeling (DEM) from Magellan radar images, using stereo matching and shape-from shading. The final DEM evolves from a preliminary altimeter data set, then improves via stereo matching into a refined DEM (left half of the 3D diagram), followed by shape-from-shading to produce the final result (right half of 3D diagram).

3. PARALLELIZATION

Data decomposition is the obvious approach for parallel processing or processing on multiple computers. Figure 4 illustrates the concept that is applicable if processing for an image location needs only pixels in a limited Region-of-Interest ROI. Patches in the decomposition should overlap in cases where along the edge of a patch irregularities occur: in matching, the edges cannot be processed, and in SfS, the edge of a patch displays so-called “ringing” effects. Overlaps permit one to eliminate such problem areas.

The Message Passing Interface (MPI, 1996) was used to organize the parallel execution of the algorithms. Ptaches for SfS comprise 128 x 128 pixels, decomposition is thus an ideal paradigm for parallelization, with a reduction of the time complexity from $O(n^2 log(n))$ to $O(n^2)$.

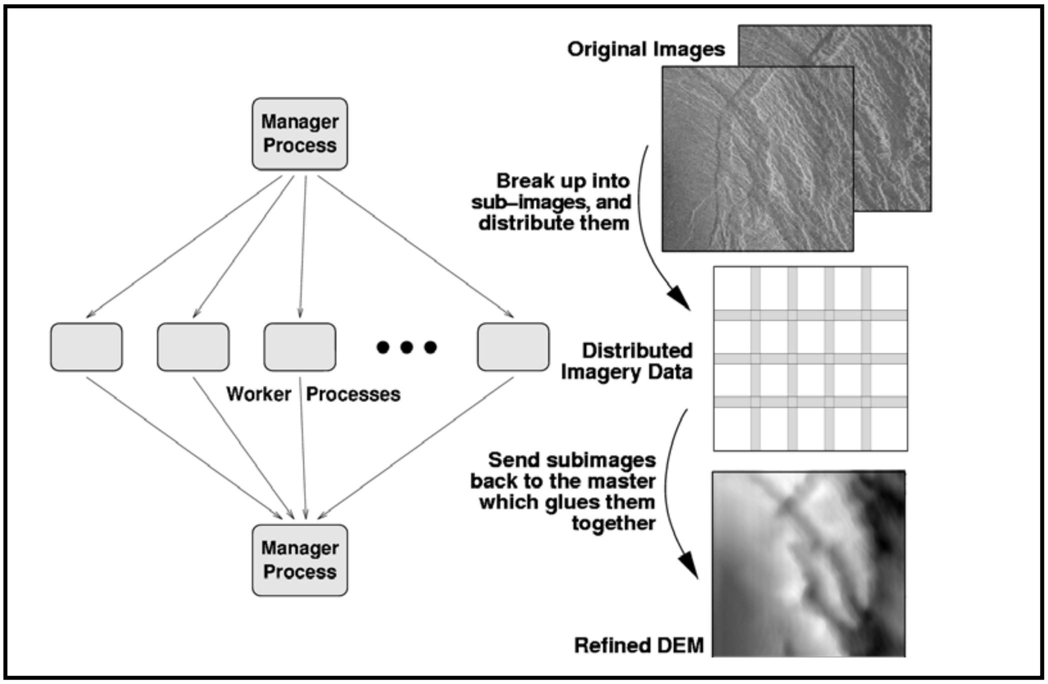


Figure 4: Parallelization by segmenting the data sets into patches and processing patches in parallel (data decomposition).

Managing the decomposed data sets is based on a "Manager/Worker" or "Task Farm" approach. The manager is one processor among the many to load and store all dat, preparing the patches and communicating them to other processors.

4. CDIP ARCHITECTURE

The Internet changes everything, and offers numerous new ways of dealing with computing resources, expert knowledge, algorithms and data sets. Adding Java and NetSolve opens an array of new options in using specialists and their image processing knowledge, as well as data sets located at remote sites, at one's own desk top. That is the model for the Center for Earth Observation CEO of the Europen Union's Joint Research Center.

Netsolve (Casanova and Dongarra, 1997) makes reosuces available to any user anywhere in a network. The basic buidling blocks of an approach we denote as CDIP (Concurrent and Distributed Image Processing) are shown in Figure 5. There are a front end, broker and a back end. Figure 6 illustrates the graphical user interface of the front end using the Magellan example.

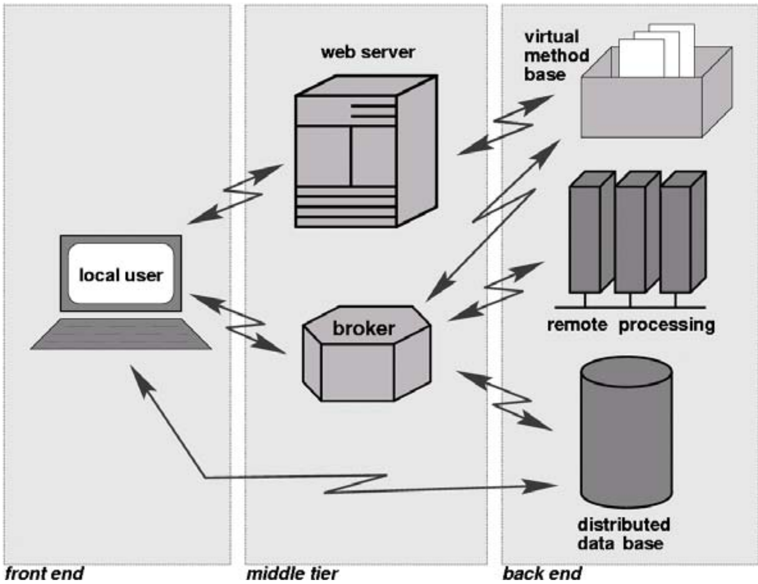


Figure 5: Basic architecture for a Concurrent and Distributed Image Processing scheme.

Figure 6: Example of a simple graphical user Interface for CDIP, using Magellan data.

While the fronm end is a Java applet running within a web browser, the back end consists of the computing resources such as a cluster of work stations or any other type of powerful computing arrangement.

The broker is the critical and intelligent segment of CDIP. It may be a simple data base or a sophisticated decision support system.

The entire configuration addresses inter-operability, queuing for a unified access to a variety of paltforms, method data bases with various image processing functions, scheduling via meta information and interfacing with the user.

5. COMPARISONS

We were able to employ the following computers:

SGI Power Challenge with 20 nodes,
Intel Paragon with 121 and with 1134 nodes
MEIKO CS2 with 121 and 1134 nodes,
Cray T3D with 512 nodes,
CoW (Cluster of Workstations)

Bringing up the various algorithms on these different machines illustrated the importnace to being machine independent. Table 2 describes the various ways in which the result of paralleization can be assessed. "Speedup" S(p) is the most commonly used parameter to describe the effect of parallelization, and it is computed as a ratio of throughput T(1) using 1 processor and T(p) using p processors.

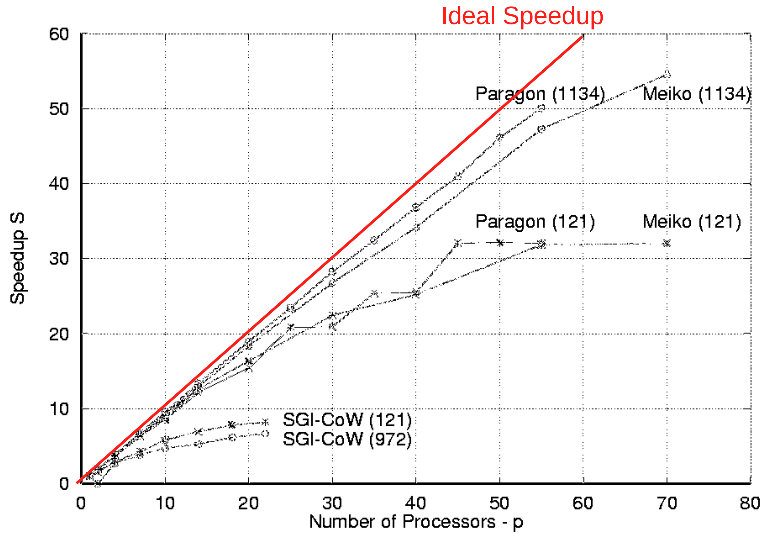
Wall-clock time T, CPU time T_{CPU},
Speedup $S = T(1) / T(p)$,
Efficiency $e = S / p$,
Efficacy $\eta = S^2 / p$
Scaleup $\gamma = T(1,d) / T(p, p*d)$

Table 2: Evaluation parameters for the usefulness of parallel computing, for example assessing the efficiency "e" as a measure of the increase in speed-up as more processors get used.

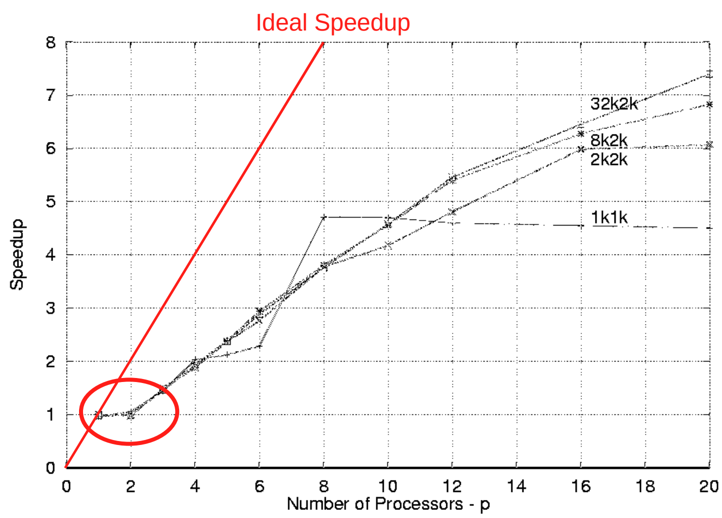
Figure 7 presents the throughput increases achieved with various systems, addressing stereo matching and shape-from-shading. Experiments were perfomed with image patches of 1k x 1k pixles, 2k x 3k pixels and larger patches.

Matching:
On a Cray T3D, we found the speed up for *matching* for example to be in the range of about 3 when 4 processors are being used, but is less effective as more processors get added. The efficacy is at 2, the efficiency at less than 0.5. All three parameters get smaller as more processors get added, since data transfer speeds limit what can be accomplished: the 10 Mbit Ethernet connection to the disk was the limiting factor.

On the Power Challenge and other machines, those limits were less prevalent.



Shape-from-Shading



Matching

Figure 7: Speed-up using mutiple processing nodes, either on a parallel processing computer, or by using mutiple work stations in a cluster, based on image patches of 1k x 1k pixels.

Shape-from-Shading:

For SfS the time is used to process a single patch was at 120 seconds per patch using a single processor on the fast computers such as a Paragon or Meiko, and 500 seconds on an Indy. The speed-up is maintained linearly as more processors are added, until the number of processors exceeds about ¼ of the number of batches and load imbalancing starts to become visible.

Figure 3 described an example of a shape-from-shading computation for a Magellan area. Instead of waiting an hour to obtain a result on a single processor work station for a 2k x

4k DEM, the approach used here reduces the wait to 12 minutes using a cluster of work stations on a computer network.

6. OUTLOOK

Internet-based advantages in remote sensing image processing have become a hotly researched subject in major organizations. One of the difficulties of accepting remote sensing lies in the complex analysis methods, the need to have expert knowledge about these methods, the many locations at which data are stored, and the need for the individual user to have the results quickly and at one's own desk top computer. This cannot be accomplished without the Internet and an ability to process data remotely and quickly.

CDIP has been developed to learn how to support the user with remotely located data and remotely available image processing methods running on specialized computers, thereby exploiting the possibility of parallel execution of algorithms. We found that with Java and Netsolve, we could indeed achieve an ease of access and a gain in efficiency and throughput in remote sensing image processing almost linear with the increase of processors, using Magellan data of planet Venus as the data set to experiment on. Key algorithms were shape-from-shading and stereo-matching, two methods that are to be intimately connected in SAR image analysis to produce detailed and accurate Digital Elevation Models.

LITERATURE

Casanova H., J. Dongarra (1997) *NetSolve: A Network Server for Solving Computational Science Problems*. Int'l J. of Supercomputer Applications and High Performance Computing, Vol. 11, No. 3, pp. 212-223.

Goller A. (1999a) *Parallel and Distributed Processing of Large Image Data Sets*. Doctoral Diss., Technical Univ. Graz, Austria.

Goller A. (1999b) *Parallel Strategies for Large SAR Image Data Sets in a Distributed Environment.* Computing, Vol 62, No.4, pp 277-291.

Hensley S., S. Shaffer (1994) *Automatic DEM Generation Using Magellan Stereo Data*. IGARSS 94, Pasadena CA., pp 1470-1472.

MPI (1996) *The Message Passing Interface – Standards Specification V1.1*.

Thomas J., W. Kober, F. Leberl (1991) *Multiple Image SAR Shape-from-Shading*. Photogrammetric Engineering and Remote Sensing. Vol. 57, No.1