Innovative Information Technology for Space Weather Research

March 3, 2004

ABSTRACT

Solar activity is closely related to the near earth environment – summarized descriptively as space weather. Changes in space weather have adverse effect on many aspects of life and systems on earth and in space, such as safety of spacecrafts and astronauts, interruption of communication and navigation syst4ems, damage to power grids and trans-continental pipelines due to grounding electric currents caused by geomagnetic storms, and hazardous conditions to polar routed flights. Real-time, high-quality data and data processing would be a key element to forecast space weather promptly and accurately. Recently, we obtained a funding from US National Science Foundation to apply innovative technology for space weather prediction.

(1) We use the technologies of image processing and pattern recognition, such as image morphology segmentation, Support Vector Machines (SVMs), and neural networks to detect and characterize three important solar activities in real-time: filament eruptions, flares, and emerging flux regions (EFRs). Combining the real time detection with the recent statistical study on the relationship among filament eruptions, flares, EFRs, coronal mass ejections (CMEs), and geomagnetic storms, we are establishing real time report of solar events and automatic forecasting of earth directed CMEs and subsequent geomagnetic storms.

(2) We combine state-of-art parallel computing techniques with phase diversity speckle imaging techniques, to yield near real-time diffraction limited images with a cadence of approximately 10 sec. We utilize the multiplicity of parallel paradigms to optimize the calculation of phase diverse speckle imaging to improve calculation speed. With such data, we can monitor flare producing active regions continuously and carry out targeted studies of the evolution and flows in flare producing active regions.

(3) We are developing Web based software tools to post our processed data, events and forecasting in real time, and to be integrated with current solar activity and space weather prediction Web pages at BBSO. This will also be a part of Virtual Solar Observatory (VSO) being developed by the solar physics community. We are also developing a high speed server and advanced user interface and data visualization, so the real-time data and forecasting can be available to the community efficiently.

A. Introduction

The sun is the source of space weather – the disturbances whose radiative, field and particle energy directly impacts earth. Solar activity changes the radiative and particle output of the sun, producing corresponding changes in the near-earth space environment, as well as at the earth's surface. The most dramatic events on the sun, insofar as space weather effects are concerned, are solar flares and coronal mass ejections (CMEs).

A flare is defined as a sudden, rapid, and intense variation in brightness. A solar flare occurs when magnetic energy that has built up in the solar atmosphere is suddenly released. Radiation is emitted across virtually the entire electromagnetic spectrum, from radio waves at the long wavelength end, through optical emission to X-rays and gamma rays at the short wavelength end. The amount of energy released is the equivalent of millions of 100-megaton hydrogen bombs exploding at the same time! Flares are often associated with CMEs.

A CME is due to a large scale rearrangement of the solar magnetic field which can induce a geo-effective solar wind structure. Each CME may carry away a mass of up to 10^{13} kg and release up to 10^{25} J of energy from coronal magnetic fields (Harrison, 1994, 1995). Following their discovery, CMEs were soon found to be correlated with the occurrence of geomagnetic storms. CMEs are easily observed at the solar limb, where they are seen against a dark background. However, the earth-directed, disk CMEs are much harder to detect – although they are the ones that have the most important geomagnetic effects. Therefore, detecting the early manifestation of CMEs on the solar disk is essential to understanding earth-directed CMEs themselves. Establishing the correlation among the occurrence of CMEs and some other solar surface phenomena, such as filament eruptions and flares, is a long-standing problem in solar physics. After firm relationships are established, the correlated solar surface phenomena can provide an early warning of the occurrence of earth-directed CMEs and geo-effectiveness of CMEs. A key challenge to move this research area forward is to have real-time, automatic and accurate detection and characterization of solar events, and to use the results for real time space weather forecasts.

Space weather forecast faces another challenge: obtaining and utilizing real-time diffraction limited images. An important goal in ground-based astronomy is to improve the angular resolution that can be achieved. The angular resolution is nearly always limited by phase aberrations introduced by atmospheric turbulence. For the special case of ground-based solar astronomy, the spatial resolution is typically limited to about $0.5''$ for short-exposure images ($\leq$20 ms) and to about $1.0''$ for long-exposure images (about 1 s). $1''$ is about 730 km on the sun. Many basic processes on the Sun, however, take place at scales below $1.0''$. The photon mean free path in the lower photosphere corresponds to about $0.1''$ at disk center. Magnetic structures may occur on even smaller scales. For example, magnetic elements outside sunspots have typical diameters smaller than $0.2''$ (Keller, 1995). Despite their small size, these small structures are believed to play an important role in large-scale phenomena such as the solar magnetic dynamo or triggering of solar flares and CMEs. To understand a variety of solar phenomena, it is, therefore, indispensable to reach a spatial resolution well below the seeing limit and possibly approaching the diffraction limit of existing and future large solar telescopes. A number of sophisticated techniques have been conceived to combat the deleterious effects of atmospheric turbulence in astronomical imaging in general. Among these are speckle imaging, phase-diversity and phase-diverse speckle imaging. Phase-diversity speckle imaging techniques are particularly attractive for solar astronomy because (1) they require relatively simple and inexpensive instrumentation, (2) they perform well with relatively few images in high-signal regimes, (3) they lead to a joint estimation of the object and the wavefront, and (4) they obviate the need for complicated calibration.

Speckle imaging and phase diverse speckle imaging apply complex numerical algorithms for data analysis. Usually, only a few data sets are analyzed per year, and only at the end of this process does one discover whether something really interesting has been captured. This is due to the long computation time, so raw images have to be saved to be processed off-line, and a hard disk is filled up very quickly. The time-lag between the observation and the data analysis is far too long for any sort of rapid response, such as would be needed for space weather warnings and flare forecasting. Furthermore, the time lag renders the whole scientific enterprise less efficient than it needs to be considering today's computer technologies. The power of parallel computing has just started to be exploited in solar physics. Supercomputers have been used for numerical calculations in astrophysics, however they cannot be used for real-time data processing because the observational data cannot be transferred to the supercomputer centers in real-time.

Parallel processing of solar data will literally provide a new window through which we can observe the Sun in exquisite detail and study the evolution of granulation and sunspots in the active regions leading to major flares. The underlying data processing algorithms are understood, but the complexity is such that only parallel computing enables us to visualize and interpret large data sets effectively.

B. Automatic Detection and Real-Time Monitoring of Solar Activity

B.1 Introduction and Current Status

Recently, we have gained significant experience in the detection of sudden disappearance of solar filaments. This was the center part of the Ph.D. theses of J. Gao (2002) and the thesis work of Ming Qu. We combine thresholding and region growing methods which were based on segmentation principle (Beveridge, 1989) in the detection.

(Fill in the results by Frank, Ming and David on flare detection et al.)

We advance our studies in the following two areas:

(1) We are developing new detection methods to detect filament eruptions, flares and Emerging Flux Regions (EFRs), and compare the effectiveness of different methods.

These three kinds of features have different morphologies in solar chromosphere. Filaments are elongated dark features, flares are rapid brightenings and EFRs have dark fibril structure in the center of the growing region (arch filaments) and bright patches at footpoints. These three activities have different time scales: impulsive flares occur in on time scales of minutes; filament eruptions may be associated with flares, and the process lasts between minutes and hours; an EFR may take hours to a day to be fully developed. We will develop pattern recognition methods to detect all three kinds of dynamic phenomenon. Examples of existing solar feature include statistical pattern recognition method developed by Turmon et al. (2002) to label faculae and sunspots; neural network detection of flares by Fernandez Borda et al. (2002); and Machine Learning systems reviewed by Mjolsness and Decoste (2001). However, we found that detection of dynamic solar activities is still far from being mature.

(2) We characterize the detected solar features. There are many properties of solar features worthy monitoring and we can not include them all. We will concentrate on the following features which are most relevant to the space weather research:

The scientific goal is to develop new automatic procedures to detect disappearing filaments, flares and EFRs. The tasks of segmentation and classification are challenging since it is very difficult to take into account the complexity involved in the human classification process.

Mjolsness and Decoste (2001) have highlighted the enormous potential of machine learning (ML) for the future of science. In the context of solar physics application of ML is only just beginning. Here we outline our plans to use several ML techniques to achieve our goal of on-line (real-time) monitoring of dynamic phenomena on the solar surface. Not only do we wish to detect eruptive events like CMEs when they happen, but also we would like to be able to predict their likely onset based on spatial-temporal surface indicators prior to such events.

We concentrate on the following tasks.

Features Classification Using Neural Network Approaches.

RBF and SVM Approaches

C. Real-Time Image Reconstruction with Parallel Computer Cluster

C.1 Introduction – Two Different Methods

Seeing is the primary obstacle to obtaining high resolution time-series solar observations from the ground. Turbulence in the atmosphere along the line of sight randomly distorts the wavefronts. The result is blurring and geometrical distortions in the collected images.

Adaptive optics (AO) can fix some of the problem, but the high bandwidth correction needed allows only partial correction. Also AO is not so effective for large fields of view because of anisoplanatism. Large fields of view are required for space weather applications. The solution is post processing of image data in subfields smaller than the anisoplanatic angle, which for many applications is on the order of $\lesssim 5''$. The problem is to separate the object and the aberrations, two unknown quantities, in the images in the presence of noise, a third unknown. The exposure time for each frame must be short enough that the evolving atmosphere can be regarded as frozen during the exposure. The data collection model is shown in Fig. 1. Two different approaches have been developed and used for larger solar data sets: Speckle Interferometry and Phase Diversity.

Speckle imaging is a relatively mature technique for obtaining fine-resolution images in the presence of atmospheric turbulence. This technique requires the collection of many short-exposure images of a static object. Clever processing of this short-exposure time series affords the restoration of fine-resolution information that would be irretrievable if a single, long-exposure image were collected (Dainty, 1984). Speckle imaging requires the collection of 100 or more images

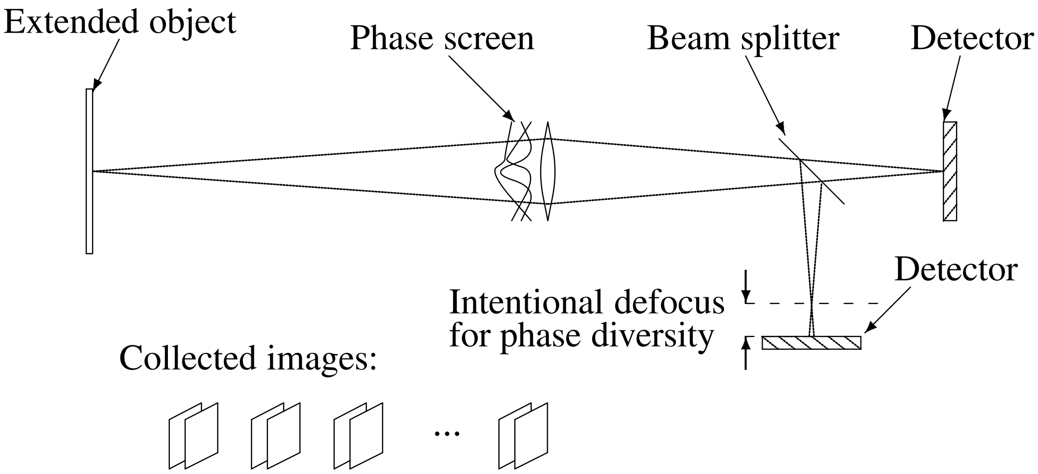

Figure 1: Typical data collection scheme, illustrating the data collection model. Only the straight through path is used for speckle interferometry. Short exposure images are collected for each of multiple atmospheric realization. For phase diversity methods, the conventional short-exposure images are supplemented with additional, simultaneous images that are intentionally defocused.

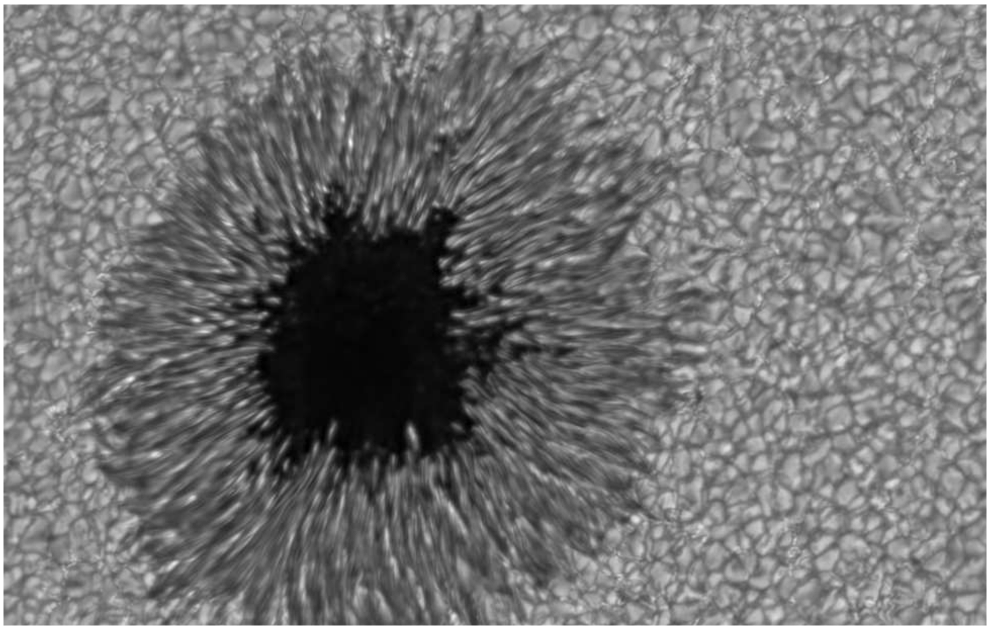

Figure 2: Speckle reconstruction of active region NOAA 8673 observed with the 1k $\times$ 1k pixel, 12-bit SMD CCD camera on 27 August 1999. The FOV is $76'' \times 40''$, the exposure time is 4 ms, and the Fried-parameter amounts to 12.6 cm.

so that the ensemble average over the class of all possible realizations may be approximated by an arithmetic average over a finite number of realizations. Speckle-imaging methods have been successfully adapted to the solar-imaging problem (Keller and von der Lühe, 1992; de Boer and Kneer, 1992; von der Lühe, 1993, 1994). The parallel operation for speckle imaging has been implemented at BBSO (Denker et al., 2001).

Phase Diversity (PD) methods, utilize a model of the image formation process to constrain the estimation of the aberrations and the common object in two image channels with a known difference in the pupil phase, usually a focus shift. An image collected in this second optical channel will contain the effects of the unknown phase aberrations but will also be influenced by the intentional defocus, which adds a known quadratic phase.

Our research objective is to create, in the terminology of `www.darpa.mil/baa/baa02-09.htm`, a High Productivity Computing System (HPCS) for space weather research that will use state-of-art parallel computing methodologies and tools to optimize phase diverse speckle codes, and to obtain large size diffraction limited data in real time.

Given the high volume of data and processing times, that involve image preprocessing, FFT operations, iterative methods that themselves involve sparse matrix computations and FFT operations, parallelism provides probably the only choice to reduce running and response times to achieve real-time or close to real-time performance. In general, the use of parallelism can improve performance in two ways:

1. better *scalability* of the massive number of images is possible, since larger images or increasing number of images can be handled within the same response time (response time remains the same, large problem sizes can be solved within the same amount of time),
2. a significant reduction in *response time* for a given problem size can be achieved (problem size remains the same, response time decreases).

In this research effort we mainly concentrate on achieving objective 2; objective 1, however, may be dealt with if increasing the number of images processed could substantially improve performance or the reliability of the predictions.

The nature of the problems that need to be solved and the use of task parallelism, does not necessitate at least initially the use of dynamic load balancing approaches to replace the static approaches that will be employed. High availability will be maintained by keeping track of the state of the nodes of the cluster; this information will be collected and used by the load-balancer responsible for task allocation.

The end-result of this research effort will be the use of our experiences in parallel algorithm design and implementation in building a High productivity computing system for space weather research.

D. Distribution and Application of Data and Results for Space Weather Research

The system discussed above will be a powerful tool for many areas in high resolution solar physics research. As the scientific objective is for real space weather, we will concentrate on a targeted study: diffraction limited atmospheric structure of flaring active regions.

Our ultimate goal is to provide high quality data and space weather forecasting in real time. BBSO now has established a web server, and our automatic solar activity detection and real-time diffraction solar images will be important additions to the system.

In our current on-line system, every hour, images from our Global Hα Network are transferred to the web-based BBSO Active Region Monitor (ARM, `www.bbso.njit.edu/arm`), which also includes the most recent full-disk EUV, soft X-ray, continuum and magnetogram data from the Solar and Heliospheric Observatory (SoHO) and YOHKOH (until December 2001). ARM also includes a variety of active region properties from the National Oceanic and Atmospheric Administration's (NOAA) Space Environment Center (SEC), such as up-to-date active region positions and flare identifications. Furthermore, we have developed a Flare Prediction System (FPS) which estimates the probability for each active region to produce C-, M-, or X-class flares based on nearly eight years of NOAA data from cycle 22. This, in addition to BBSO's daily solar activity reports, has proven to be a useful resource for activity forecasting (Gallagher et al., 2002).

The new Web pages will include real time report of all the filaments (and their disappearances), flares and EFRs. The size, location and two ribbon separation speed of flares, and information of disappearing filaments will be posted in real time. The real-time diffraction limited active region images will be also posted on the web, along with flows derived in real time by local correlation tracking and feature tracking.

Figure 3 shows the structure of new contents of space weather forecasting based on results from our research, and the link among different components.

Virtual Solar Observatory (VSO) is a community effort, which attempts to distribute ground-based and space data to users uniformly and efficiently (Hill, 2001). We actively collaborate with NSO group to include our real-time data and forecasting as a part of VSO. The VSO effort is lead by NSO.

In order to maintain our objective of building a High Productivity Computing System for space weather research, we shall also develop a web-based interface to the parallel computing module previously described. The results of the data processing will become available through this interface. Open-source software tools will be used for the development of this interface and the access of the raw and the processed data. A Linux-based Apache web-server will be utilized. Web-based forms in HTML will be processed by Perl scripts and if needed C programs to generate and return the required responses by accessing a MySQL database server/engine.

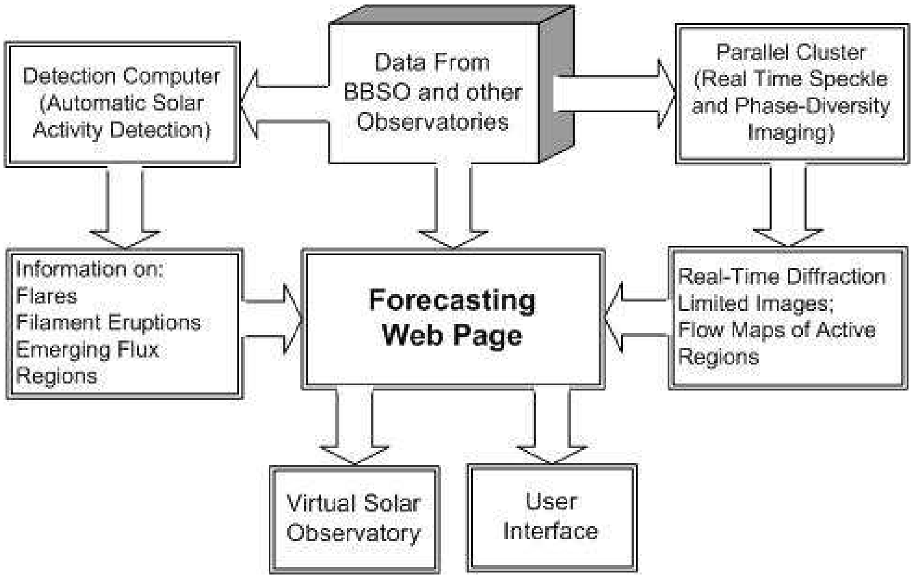

Figure 3: A block diagram showing the interconnection of different research components.

References

Beveridge, R.: 1989, Segmenting images using localized histograms and region merging, *International Journal of Computer Vision* **2**, 311–347.

Dainty, J. C.: 1984, *in* J. C. Dainty (ed.), *Laser Speckle and Related Phenomena*, Vol. 9 of *Topics in Applied Physics*, Springer, p. 255.

de Boer, C. R. and Kneer, F.: 1992, Speckle observations of abnormal solar granulation, *Astronomy & Astrophysics* **264**, L24–L26.

Denker, C., Yang, G. and Wang, H.: 2001, Near real-time image reconstruction, *Solar Physics* **202**, 63–70.

Fernandez Borda, R. A., Mininni, P. O., Mandrini, C. H., Gomez, D. O., Bauer, O. H. and Marta, G.: 2002, Automatic solar flare deteciton using neural network techniques, *Solar Physics* **206**, 347–357.

Gallagher, P., Moon, Y. and Wang, H.: 2002, Active-region monitoring and flare forecasting - i. data processing and first results, *Solar Physics* **209**, 171–183.

Gao, J.: 2002, *Development of an Automatic Filament Disappearance Detection System*, PhD thesis, New Jersey Institute of Technology.

Harrison, R. A.: 1994, A statistical study of the coronal mass ejection phenomenon, *Advances in Space Research* **14**, 23–28.

Harrison, R. A.: 1995, The nature of solar flares associated with coronal mass ejection, *Astronomy & Astrophysics* **304**, 585–595.

Hill, F.: 2001, The virtual solar observatory - status and plans, *AGU Spring Meeting*.

Keller, C. U.: 1995, Properties of solar magnetic fields from speckle polarimetry, *Reviews in*

Modern Astronomy **8**, 27–60.

Keller, C. U. and von der Lühe, O.: 1992, Solar speckle polarimetry, *Astronomy & Astrophysics* **261**, 321–328.

Mjolsness, E. and Decoste, D.: 2001, Machine learning for science: State of the art and future prospects, *Science* **293**, 2051–2055.

Turmon, M., Pap, J. M. and Mukhtar, S.: 2002, Statistical pattern recognition for labeling solar active regions: Application to soho/mdi imagery, *Astrophysical Journal* **568**, 396–407.

von der Lühe: 1993, Speckle imaging of solar small structure: I. Methods, *Astronomy & Astrophysics* **268**, 374–390.

von der Lühe: 1994, Speckle imaging of solar small structure: II. Study of small scale structure in active regions, *Astronomy & Astrophysics* **281**(3), 889–910.