12-29-2006

ITR/IM: Enabling the Creation and Use of GeoGrids for Next Generation Geospatial Information

Peggy Agouris  
Principal Investigator; University of Maine, Orono

Mary-Kate Beard-Tisdale  
Co-Principal Investigator; University of Maine, Orono, beard@spatial.maine.edu

Chaitanya Baru  
Co-Principal Investigator; University of Maine, Orono

Sarah Nusser  
Co-Principal Investigator; University of Maine, Orono

Follow this and additional works at: https://digitalcommons.library.umaine.edu/orsp_reports

Part of the Databases and Information Systems Commons, and the Geographic Information Sciences Commons

Recommended Citation
Agouris, Peggy; Beard-Tisdale, Mary-Kate; Baru, Chaitanya; and Nusser, Sarah, "ITR/IM: Enabling the Creation and Use of GeoGrids for Next Generation Geospatial Information" (2006). University of Maine Office of Research and Sponsored Programs: Grant Reports. 134.  
https://digitalcommons.library.umaine.edu/orsp_reports/134

This Open-Access Report is brought to you for free and open access by DigitalCommons@UMaine. It has been accepted for inclusion in University of Maine Office of Research and Sponsored Programs: Grant Reports by an authorized administrator of DigitalCommons@UMaine. For more information, please contact um.library.technical.services@maine.edu.
**Senior Personnel**

<table>
<thead>
<tr>
<th>Name</th>
<th>Worked for more than 160 Hours</th>
<th>Contribution to Project</th>
</tr>
</thead>
<tbody>
<tr>
<td>Agouris, Peggy</td>
<td>Yes</td>
<td></td>
</tr>
</tbody>
</table>

**Project Participants**

<table>
<thead>
<tr>
<th>Name</th>
<th>Worked for more than 160 Hours</th>
<th>Contribution to Project</th>
</tr>
</thead>
<tbody>
<tr>
<td>Beard-Tisdale, Mary-Kate</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>Baru, Chaitanya</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>Nusser, Sarah</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>Stefanidis, Anthony</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>Ludaescher, Bertram</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>Zaslavsky, Ilya</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>Miller, Leslie</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>Stern, Hal</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>Maiti, Taps</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>Maitra, Ranjan</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Name</td>
<td>Worked for more than 160 Hours</td>
<td>Contribution to Project</td>
</tr>
<tr>
<td>-----------------------</td>
<td>-------------------------------</td>
<td>-------------------------</td>
</tr>
<tr>
<td>Gupta, Amarnath</td>
<td>No</td>
<td></td>
</tr>
<tr>
<td>Wahadj, Reza</td>
<td>No</td>
<td></td>
</tr>
<tr>
<td>Memon, Ashraf</td>
<td>No</td>
<td></td>
</tr>
<tr>
<td>Windholz, Thomas</td>
<td>Yes</td>
<td>Post-doc</td>
</tr>
<tr>
<td>Doucette, Peter</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>Croitoru, Arie</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>Bharath, Roger</td>
<td>Yes</td>
<td>Graduate Student</td>
</tr>
<tr>
<td>Bharadwaj, Rajiv</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>Liu, Fengmei</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>Kies, Courtney</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>Czuprynski, Jennifer</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Worked for more than 160 Hours: Yes
Contribution to Project:

Name: Mukhopadhyay, Pushpal

Name: Venkataraman, Vijay

Name: Schlaisich, Isolde

Name: Skalland, Ben

Name: Ding, Jing

Name: Mountrakis, Giorgos

Name: King, Joshua

Name: Manpuria, Vivek

Name: Pekurosvky, Dmitry

Name: Eickhorst, Kristin

Undergraduate Student

Technician, Programmer

Other Participant
Research Experience for Undergraduates

Organizational Partners

Nat'l Cntr fr Geographic Info & Analysis

US Geological Survey
During this project we worked together with USGS (Fort Collins) on their Invasive Species Forecasting System and more specifically, on monitoring and prediction of spreading of tamarisk in Colorado.

NASA / Goddard Space Flight Center
We worked together with NASA's Goddard Space Flight Center on the Invasive Species Forecasting System.

Other Collaborators or Contacts
There are a few organizations that, although not official partners in this project, have contributed significantly to our progress, through discussions and collaborations with key project personnel.
These organizations are:
- US Forest Service
- Bureau of Labor Statistics
- National Geospatial-Intelligence Agency (formerly National Imagery and Mapping Agency)
- Army TEC
- Milcord LLC

Activities and Findings

Research and Education Activities:
The GeoGrid project addressed the development of concepts, algorithms, and system architectures to enable users on a grid to query, analyze, and contribute to multivariate, quality-aware geospatial information. The project research agenda was arranged into three complementary thrust areas that represented the key issues behind our vision. They were:
- the establishment of a statistical framework for assessing geospatial data quality;
- the development of novel mediation approaches for quality-aware spatiotemporal information; and
- the development of space- and accuracy-aware adaptive systems and agents to improve the organization of information within GeoGrid and to evaluate information contributions to it.

In terms of research area 1, the focus of our research during year 1 was to develop statistical descriptors for expressing the quality of geospatial data as a single source or as a new data source created via computational manipulations. The approach involved developing an understanding of relational data models and corresponding operations, statistical summaries for spatial data quality, and traditional concepts applied to quality assessment for geospatial data.

In year 2, our primary focus has been to consider methods for estimating accuracy summaries that can be used in computer-based queries of discrete-valued thematic raster data. The computing goals were to incorporate quality measures for geospatial data sources into the query process, and to generate new accuracy summaries for newly constructed data products that represent the response to a data query. Example queries include: (1) generating a new map that represents a sub-region of the original map, and (2) generating a new map by collapsing one or more discrete thematic categories.

In year 3 and the no-cost extension year, we worked on accuracy summaries estimation methods. Our research is being illustrated with the thematic land cover and field ground truth data from Iowa’s GAP Analysis Project (GAP) land cover map.

Estimators for standard overall and user's accuracies under these two transformations have been derived when ground truth data or a sample from a benchmark source is available. In addition, we have developed more complex pixel-based representations of accuracy using covariates associated with the heterogeneity of classification surrounding a pixel. A naïve estimator for a pixel-level probability of correct classification has been derived last year. This year, estimators that account for spatial correlation were explored and several models evaluated that varied in
During the no-cost extension years, we worked heavily on the comparison of spatiotemporal trajectories using pose normalization to recognize video feeds. We chained intelligently to match appropriate environmental conditions at each node and then using the chained algorithm to preprocess the video feeds at network nodes to remove any localized variations. We have extended the scope of metadata to define processes that can video feeds from individual sensors, by employing process modeling techniques that sense environmental conditions and allow preprocessing in the form of images, since information can be communicated via the visual channel in a highly effective and efficient manner. We focused on ways to effectively communicate geospatial data quality to users. We decided to visualize quality information sources. In year 3 we focused on two aspects. The first dealt with the identification of information gaps in the available heterogeneous geospatial datasets and the representation of quality parameters. The second aspect of our research dealt with capturing and storing domain expertise to create an environment for the automatic, on-demand selection of the appropriate analysis tools for any type of geospatial information related task.

In addition to the above, during year 2, we worked heavily on the issue of similarity learning in spatiotemporal distributed databases. During year 3, we continued working on and refining our information retrieval approach to adjust query returns based on similarity profiles that express more accurately user anticipation of results.

Our neuro-fuzzy-based system attempted to express similarity within one-dimensional, quantitative attributes. Our training was performed in multiple stages and was based on a training dataset provided by the user. Our novel neuro-fuzzy system outperformed the currently used distance-based nearest neighbor methods. It did so by design because it recognized and supported distance dependent preference, while simultaneously offering advanced modeling capabilities. Our system also exhibited high robustness as demonstrated through statistical simulations. This was partially due to the ability of the algorithm to adjust its complexity as the complexity of the similarity preference increases.

In year 2 we also worked on heterogeneous data quality assessment, conflation, visualization, and communication in distributed geospatial information sources. In year 3 we focused on ways to effectively communicate geospatial data quality to users. We decided to visualize quality in the form of images, since information can be communicated via the visual channel in a highly effective and efficient manner. We used 3D perspective representations and color hue and saturation to convey different quality attributes.

In year 3 we also focused our on-going research on metadata to include the newest channel of distributed geospatial data collections, namely video feeds from individual sensors, by employing process modeling techniques that sense environmental conditions and allow preprocessing of the video feeds at network nodes to remove any localized variations. We have extended the scope of metadata to define processes that can be chained intelligently to match appropriate environmental conditions at each node and then using the chained algorithm to preprocess the video feeds.

During the no-cost extension years, we worked heavily on the comparison of spatiotemporal trajectories using pose normalization to recognize how spatial correlation was treated, including logistic regression, generalized estimating equations, and autologistic regression. Results indicate that it can be difficult to fit appropriate spatial covariance models for discrete data.

In year 2, we have also considered the impact of positional inaccuracy on thematic accuracy as part of these investigations. We have also looked at the issue of accuracy in relational databases. We have developed models of accuracy for alphanumeric data in the relational environment. We have interacted with project team members on the relational accuracy models.

In years 3 and 4, we incorporated the results from the modeling research into a sampling application that incorporated error measures into the sampling scheme. Also, we worked on alternative formulations for quantifying the accuracy level of a record or a data element, which we continued working on during the no-cost extension years using probability and statistical theory. The relational algebra has also been redefined with respect to the accuracy model.

The focus of our work on the second research area during the first years of the project was to provide an understanding of relational data models and corresponding operations, statistical summaries for spatial data quality, and traditional concepts applied to quality assessment for geospatial data. In particular, we were interested in query algebras that can deal with error information in an integrated fashion.

We have performed research on query rewriting for various types of queries (including queries with group-by), for situations where spatial metadata contain accuracy measures, and designed a Web service for error propagation. The web service works by matching queries with pre-defined error propagation templates, based on data types and function signatures used in a query. We also added accuracy descriptions to metadata XML files generated by ArcCatalog, developed a parser for consuming such error descriptions served by ArcIMS metadata server, and developed a first draft of an error propagation web service.

During year 3, we developed a MultiView ArcIMS viewer prototype. The system allows to bring together and overlay services from multiple ArcIMS servers. This was a platform for visual inspection of discrepancies, errors, and relative completeness of data served from several services. We also extended the error propagation web service to work on top of PostgreSQL/PostGIS/Mapserver sources. As part of the system, we developed procedures for registering/loading/mapping user-supplied spatial data using the PostgreSQL/PostGIS/Mapserver bundle. In year 3 and the no-cost extension year we have been using this setup for running Monte Carlo simulations of error propagation on DEMs, examining various scenarios and potential applications.

With respect to the third research area, during year 1 we started working on two aspects. The first dealt with the identification of information gaps in the available heterogeneous geospatial datasets and the representation of quality parameters. The second aspect of our research dealt with capturing and storing domain expertise to create an environment for the automatic, on-demand selection of the appropriate analysis tools for any type of geospatial information related task.

In the above, during year 2, we worked heavily on the issue of similarity learning in spatiotemporal distributed databases. During year 3, we continued working on and refining our information retrieval approach to adjust query returns based on similarity profiles that express more accurately user anticipation of results.

Our neuro-fuzzy-based system attempted to express similarity within one-dimensional, quantitative attributes. Our training was performed in multiple stages and was based on a training dataset provided by the user. Our novel neuro-fuzzy system outperformed the currently used distance-based nearest neighbor methods. It did so by design because it recognized and supported distance dependent preference, while simultaneously offering advanced modeling capabilities. Our system also exhibited high robustness as demonstrated through statistical simulations. This was partially due to the ability of the algorithm to adjust its complexity as the complexity of the similarity preference increases.

In year 2 we also worked on heterogeneous data quality assessment, conflation, visualization, and communication in distributed geospatial information sources. In year 3 we focused on ways to effectively communicate geospatial data quality to users. We decided to visualize quality in the form of images, since information can be communicated via the visual channel in a highly effective and efficient manner. We used 3D perspective representations and color hue and saturation to convey different quality attributes.

In year 3 we also focused our on-going research on metadata to include the newest channel of distributed geospatial data collections, namely video feeds from individual sensors, by employing process modeling techniques that sense environmental conditions and allow preprocessing of the video feeds at network nodes to remove any localized variations. We have extended the scope of metadata to define processes that can be chained intelligently to match appropriate environmental conditions at each node and then using the chained algorithm to preprocess the video feeds.

During the no-cost extension years, we worked heavily on the comparison of spatiotemporal trajectories using pose normalization to recognize how spatial correlation was treated, including logistic regression, generalized estimating equations, and autologistic regression. Results indicate that it can be difficult to fit appropriate spatial covariance models for discrete data.
identical spatiotemporal trajectories. To achieve this goal we developed a novel approach which is invariant to translation, rotation and scaling. We also considered issues relevant to sensor networks, which are emerging as the next generation geospatial information collection paradigm. More specifically, we investigated the topics of mobility and information aggregation. In addition, during the last no-cost extension year of the project we worked on issues related to multi-object and multi-sensor environments, aiming at the development of sensor deployment optimization solutions.

Details about our major research findings can be found in our publications.

In terms of educational activities, throughout the duration of this project we have been integrating our research findings in the relevant courses taught by the PIs as well as various selected study courses. During the no-cost extension years we have developed new regular course offerings that are based in part on our work for this project.

**Findings:**

During the last year of the project we were able to proceed with the above stated research goals and work successfully on the above outlined research activities. In all of the above areas we have arrived at solid scientific conclusions presented in refereed publications and have developed efficient solution environments that correspond to our project objectives. A small sample of our findings includes the following:

We have spent time surveying cross-disciplinary literature to familiarize the project team with concepts in the statistical, computational and geospatial sciences required to address project objectives. Selected readings were covered from basic texts on geographic information systems, probabilistic and relational databases, and statistical models for accuracy assessment of land cover maps. We have also spent time on a survey of data quality and error modeling approaches. We worked on extensions to the relational model and query algebra to deal with error quantities. We formalized our approaches and examined the effect on the relational algebra, in particular aggregation operations.

We have considered the case for a single geospatial data source with a categorical attribute. We developed statistical descriptions as well as visual representations of the accuracy (and other quality parameters) of land cover maps and appropriate database structures to support queries that take into account quality summaries.

Also, a notable development during the past years is related to Web-based federated spatial databases. These databases usually do not support accuracy-aware query processing despite growing availability of individual data quality descriptions. We introduced a mechanism for making data quality information operational through the process of accuracy-based query rewriting. A number of keywords are added to SQL to express accuracy constraints in a spatial query. Depending on the specification of error in underlying data sets, a query containing accuracy constraints is rewritten as a single or a set of crisp spatial queries, i.e. those that do not require further accuracy-based disambiguation. The rewriting uses binding of GIS spatial operations to error models via a collection of error propagation templates. This mechanism is implemented as a Web service that works in conjunction with a spatial mediator to aid in rewriting spatial queries in the presence of accuracy metadata. In a Web environment, where users may need to choose among and access different sources with different data quality characteristics, for solving tasks with different accuracy requirements, having a formal error propagation web service is important. As a result, we developed our previously mentioned error propagation web service software.

During year 3 and the no-cost extension year, and in addition to the activities outlined above, GeoGrid error description templates were included as part of registration of spatial data sources in the GEONgrid, which is part of the GEON project, in which some of the researchers of this project are also participating. We also developed and customized our spatial metadata explorer to index GEON and GeoGrid datasets.

In addition to the above, and on a more general level, one of our broader research accomplishment in this project, which is in compliance with our overall project goals, is that expert and non-expert users of distributed geospatial data can review at a glance which data are available and suitable for their tasks, and also form an intuitive understanding on how relevant their query results are as compared to their query objectives.

By addressing spatiotemporal issues as well, we managed to provide our users with tools which can not only depict and track in space and time geospatial objects and phenomena of interest but which can also be used in predicting future relevance, evolution patterns or unexpected behaviors in them. As a result, the collection of novel concepts, algorithms and tools developed in this project enables users to have increased accessibility and make better use of distributed geospatial information for both casual and scientific applications.

**Training and Development:**

Several graduate research assistants and post-doctoral researchers at all 3 partner sites are involved in this project. As a result, they were able to become exposed to interdisciplinary research topics. It should be noted that during their involvement in this project, supported students have repeatedly received awards by the American Society for Photogrammetry and Remote Sensing, and NASA. All supported students are directly involved in training through the development
of research concepts and novel algorithms and tools for their theses. Some of these students have already graduated and moved on to faculty or post-doctoral positions.

Senior researchers also benefited from the progress of this collaborative project that involves scientists with expertise in geospatial information systems, computer science, and statistics.

The teaching skills of the involved faculty were also enhanced by their work on developing new courses based in part on the activities of this project, as well as their continuous effort to incorporate in their classroom teaching research findings from the complementary yet diverse fields participating in this project.

**Outreach Activities:**

Numerous presentations based on our project findings were given to various audiences ranging from interested federal and state agencies (e.g., Army, NGA, Maine DOT, US Senate Lobbyists) to high school students. We have to note our project group's continued participation in the Dept. of Spatial Information Engineering's 'Spatial Horizons' high school outreach program, as well as University of Maine's 'Expanding your Horizons' program that is geared towards female high school students in an effort to increase their participation in science and engineering.

Our workshop on Next Generation Geospatial Information took place in Boston on October 19-21, 2003. In addition to NSF's support for this workshop, we have also managed to secure an endorsement by the International Society for Photogrammetry and Remote Sensing. The post-workshop proceedings of this refereed event have just been published as a book by Taylor and Francis. More details about this successful workshop can be found at http://dipa.spatial.maine.edu/NG2I03/CD_Contents/index.html

In addition, through the involvement of some of the PIs in two IGERT projects at the University of Maine, we were able to take advantage of more systematic and established K-12 outreach activities to expose high-school students to exciting scientific issues related to this project.

Based on work related to (and funded by) this project, one the PIs (Anthony Stefanidis) has co-organized two successful national workshops on GeoSensor Networks (2003 and 2006). Both workshops resulted in the publication of two high-quality reference books, co-edited by Dr. Stefanidis, including collections of refereed papers.

**Journal Publications**


Books or Other One-time Publications

Collection: San Diego Supercomputer Center Technical Reports
Bibliography: San Diego Supercomputer Center Technical Reports

Editor(s): C. McGlone, E. Mikhail, J. Bethel
Bibliography: John Wiley Publishers

A. Stefanidis and S. Nittel, "GeoSensor Networks", (2004). Book, Published
Editor(s): A. Stefanidis and S. Nittel
Bibliography: CRC Press

S. Nittel and A. Stefanidis, "GeoSensor Networks and Virtual GeoReality", (2004). Book Chapter, Published
Editor(s): A. Stefanidis and S. Nittel
Collection: GeoSensor Networks
Bibliography: CRC Press

Editor(s): P. Agouris, A. Croitoru
Collection: ISPRS Book Series
Bibliography: Taylor and Francis

Editor(s): Y. Manolopoulos, A. Papadopoulos, M. Vassilakopoulos
Collection: Spatial Databases: Technologies, Techniques, and Trends
Bibliography: IDEA Group Publishing, Hershey, PA

Editor(s): Pragya Agarwal, Andre Skupin
Collection: Self-Organising Maps: Applications in Geographic Information Science
Bibliography: John Wiley Publishers

Contributions within Discipline:
The results of our work in this project supported the extension of the concept of the computational grid to facilitate ubiquitous access, interaction, and contributions of quality-aware next generation geospatial information. By developing novel query processes as well as quality and similarity metrics and models, the project contributed to enabling the integration and use of large collections of disperse information of varying quality and accuracy. This supported the evolution of a novel geocomputational paradigm, moving away from current standards-driven approaches to an inclusive, adaptive system, with example potential applications in mobile computing, bioinformatics, and geographic information systems.

Our work demonstrated that the above described expected contributions are feasible and can be accomplished by our project team within the framework of this project. As an example, our research has lead to new methods of representing accuracy measures using statistical approaches. We hope that our work in this project can fundamentally change the way in which geospatial data are processed so that quality of information is taken into account.
Contributions to Other Disciplines:
As previously mentioned, algorithms and frameworks developed under this project aim to enable the integration of geospatial information that varies in terms of accuracy and resolution. This fosters a novel data sharing paradigm, broadening the established geospatial data producer and user communities.

Therefore, in addition to the above mentioned expected contributions in all fields that utilize and/or rely on geospatial information (e.g., biodiversity and ecology), our research contributes towards improving the ability of citizens and scientists to obtain and use data whose quality has been quantified (or estimated in some manner), as well as visualized and communicated effectively to expert and non-expert users alike.

Contributions to Human Resource Development:
There were several graduate students as well as post-doctoral researchers who acquired unique interdisciplinary scientific training through their involvement in this project. These young scientists are expected in turn to make significant contributions to society through the experience they gained working on the GeoGrid project. Some of our students have already graduated and moved on to faculty, post-doctoral or industry positions.

Contributions to Resources for Research and Education:
This project has enabled the University of Maine to sustain and expand its self-funded Digital Image Processing and Analysis Laboratory, and as a result, provide equipment and resources to the local community for training, education, research and outreach activities.

Contributions Beyond Science and Engineering:
As previously mentioned, our intention in the GeoGrid project was to facilitate the establishment of a loose federation of geospatial information sources with corresponding services. GeoGrid was designed in anticipation of a reality where users and producers, processes and products, requirements and expectations, all vary well beyond today's standards-driven world.

New scientific and societal trends have create a new geocomputational paradigm, where users in a variety of environments (ranging from office to home to the field), through a variety of computational modalities (ranging from workstations to hand-held devices), can access, analyze, and/or contribute geospatial information of varying quality. Our work in this project fostered this new geocomputational paradigm, and enabled reaching not only scientists in a variety of fields, but also non-expert users (e.g. firefighters, the police, homeowners).

Categories for which nothing is reported:
Any Web/Internet Site