Spatiotemporal Data Mining for Monitoring Ocean Objects

Yang Cai, Carnegie Mellon University, ycai@cmu.edu
Karl Fu, Carnegie Mellon University, xfu@cmu.edu
Daniel Chung, Carnegie Mellon University, saic@andrew.cmu.edu
Richard Stumpf, NOAA, richard.stumpf@noaa.gov
Timothy Wynne, NOAA, timothy.wynne@noaa.gov
Mitchell Tomlison, NOAA, mitchell.tomlinson@noaa.gov

Abstract

In this paper, we present our progress in spatiotemporal data mining for tracking and predicting the movement of ocean objects. We found that the spatial density filtering algorithm shows promising in preprocessing the image data. The correlation model is able to track the target based on the spatial similarity over time, when the shape is coherent. With a confined environment and enough training data, the machine learning model potentially is capable of predicting a short-term spatial dynamics. The challenge ahead is how to combine data mining models with multi-physics models.

1. Problems and Objectives

Detecting, tracking and predicting ocean objects are important to ecological and oceanographic studies as well as applications. Remote sensing database, such as SeaWiFS and MODIS, have been as means of monitoring of the spatiotemporal dynamics of ocean objects, such as harmful algal blooms (HAB) and river plumes in the presence of coastal areas. However, the current HAB computational models are limited as off-line analyses that have not been seamlessly integrated into the day-to-day field applications yet. There is a need for advanced computing techniques that could be applied to automatic detection or tracking of harmful objects, as well as the physiological status or taxonomic classification of bloom organisms, in near-shore coastal environments, as well as in the open ocean. Evaluating bloom detection techniques has a critical dependence at some level of visual analysis (Tomlinson et al., 2004). To determine chlorophyll or other cardinal property, multiple samples and parametric statistics are appropriate. For nominal properties, such as bloom type, each bloom must be treated as a single unit, regardless of the number of samples for validation. This is non-parametric problem cannot use simple pixel statistics as it requires identifying contiguous blooms.

The spatiotemporal data mining here is equivalent to object tracking and modeling, which extracts patterns from the information stream from multi-spectrum satellite images, in-situ cell counts, weather data, and qualitative and quantitative models. The problems in this study include 1) multi-resolution sensory fusion for satellite images and cell counts, 2) interaction of external forces such as wind and coastal lines with intrinsic properties such as shape and concentration, 3) multi-physics modeling that fuses biological, chemical and fluid dynamics. In this paper, we present the progress in our new project (AIST-QRS-04-3031) "Spatiotemporal Data Mining for Monitoring and Tracking Ocean Objects," sponsored by NASA-ESTO program. The objectives of the project include: 1) tracking the movement of identified targets from the satellite images, frame by frame automatically; 3) predicating conditions favorable for an anomalous event to occur where the targets have not been observed. We use SeaWiFS images from NASA, the cell counts and wind data from NOAA. We focus on US Gulf of Mexico and southeast, with examples from Cape Canaveral, Florida, both for chlorophyll and sediment (backscatter). Wind, surface temperature and coastal water circulation databases will also be used (Walker, 1996; Stumpf, et al., 1993).

2. Work-in-Progress

We have developed the particle filter and correlation based object tracking algorithms and tested for tracking the HAB objects (Cai, et al, 2005a). Spatial density filter, cellular automata are also developed to model the HAB dynamics (Cai and Joen, 2005b). The preliminary results are presented in the online links[1].

2.1 Shape clustering. We have developed a data filter and grouping algorithm that can remove noises while keeping the principal shapes. The images from the anomaly channel often contain significant sparse noises in the images. In addition, many small anomaly algae are not harmful blooms. To remove the sparse noises and small artifacts, we developed a Spatial Interaction Filter, which includes spatial density filtering (SDF) and shape grouping algorithms. SDF model clusters the data points based on the spatial density, for example, the data points that are packed within a predefined distance are classified as one group, while the data points that are sparse outside of the neighborhood distances are classified as another group. We found that it is better than Binary Morphology method because it preserves the shape better. Figure 1 shows a sample of the output. To group the 'pixels' into 'objects', we have developed a computational shape grouping algorithm, called 'snake algorithm', because it mimics the sideways movement like a snake. A snake outline deforms itself from a box to the tight fit of the outline of the targeted area. The image on the right of Figure 1 is a sample of the output. Unfortunately, the current 'snake' model is pretty slow because it uses incremental step iterations. We need to improve its speed in next phase by applying multiple resolution iterative process.

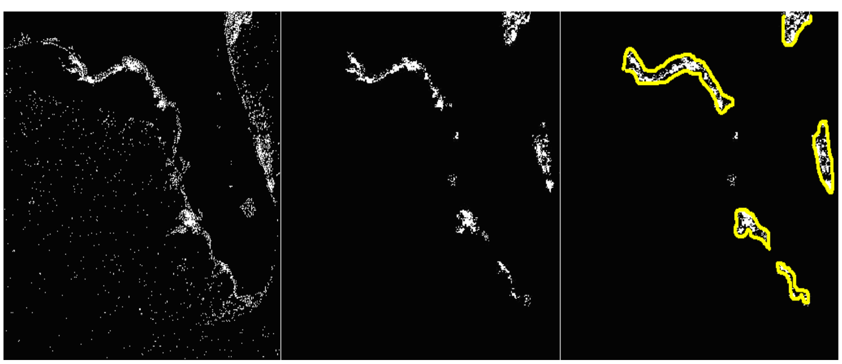

Figure 1 Original image (left), after Spatial Density Filter (middle) and after the shape grouping (right)

2.2 Tracking model. Tracking the identified HAB object is important to the HAB monitoring and analysis. We have developed a shape-based tracking algorithm that can track the predefined target in a sequence of images. The tracking algorithm is based on the Correlation Filter, which

[1] Presentations about the current NASA project AIST-QRS-04-3031:
http://www.esto.nasa.gov/conferences/estc2005/papers/a3p4.pdf
http://www.esto.nasa.gov/conferences/estc2005/Presentations/A3P4.pdf

analyzes the similarities between the previous shape and the input shapes. It uses the phase correlation of a shape to track the target. The output highlights the outline of the tracked shape:

Shape Correlation = IFFT(FFT(a). conj(FFT(b))),*

where, *a* is the test image; *b* is the reference object in the previous image to be tracked. FFT(x) represent Discrete Fast Fourier Transform IFFT(x) is Inverse Discrete Fast Fourier Transform.

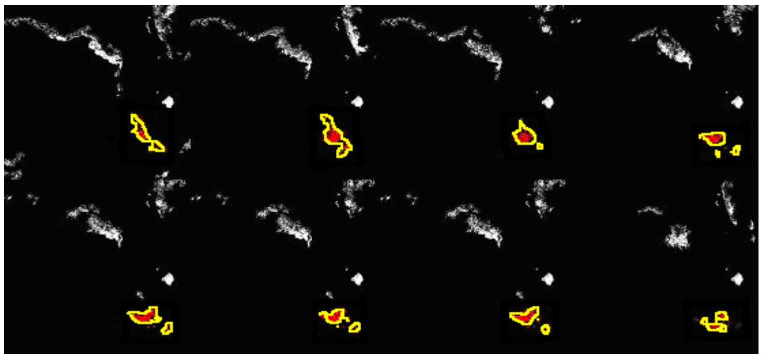

Figure 2. Tracking HAB with Correlation Filter within 4 day interval

Our initial test case shows the algorithm can track the coherent target from most of frames (79 out of total 79 frames). However, this algorithm is not yet robust in all cases. For example, it doesn't work very well when a target is split into several pieces, although is not very common in HABs. To improve the algorithm, we investigated Particle Filter algorithm, which performs better in tracking the splitting targets. However, it can only generate a box outline, rather than an accurate shape contour. See Figure 2.

2.3 Prediction model. We are currently developing the spatiotemporal data mining model that serves two functions: 1) to reduce the spatial data space, 2) to learn the dynamic logic from the historical data. At the current stage, we incorporate the artificial neural networks to model the spatiotemporal dynamics. The basic problem is to discover how and what extent the ocean object change of each cell *i* at time *t+1* is determined by the neighboring condition (CA assumption) or by the external forces (e.g. wind, ocean current, surface temperature, region, etc.) where holding at the previous time *t*. The neural network consists of two components: Self Organizing Maps (SOM, Kohonen, 1995) and the Supervised Neural Network (Radial Basis Function, RBF, Mathworks, 2004). SOM clusters the raw data in an image to the defined low-resolution points, which can significantly reduce the computational load in order of magnitudes. The RBF is used to learn two things: 1) the frame transition rules and 2) the correlations of the external forces. The overall structure is illustrated as Figure 3.

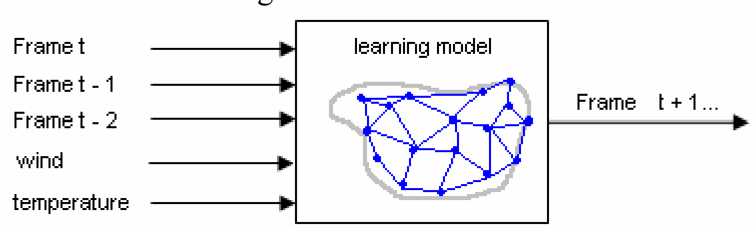

Figure 3. Spatiotemporal Learning Model

Unlike conventional methods of training frame transition rules using the Cartesian plane, frame transition rules are trained using the cylindrical plane, in which r and theta points to the same cluster in a following frame. Cylindrical coordinates provide two advantages: 1) the RBF can fit the data more smoothly without a large spread and 2) the prediction requires fewer frames in the training set.

The SOM network can reduce the data size in orders of magnitude. It is an *n*-dimensional network, where *n* is the number of inputs. The network is trained by presenting it with cluster points (the blue dots in Figure 5). We assume to have the sample data of the object at time t and t = t - 1, then the RBF learns the transition logic between the two or more images and the external variables. The output of the RBF is the prediction at the time t = t + 1, and beyond. The following is an example of our test case (Figure 4). We trained the model with four images, where the red crosses represent the location of algae blooms and the blue circle is the center of the cluster. The SOM was trained with 20 epochs with 50 clustering blue points. The target image has 5876 pixels and the clustering with the blue outline missed approximately 300 pixels, yielding a 5.1% error.

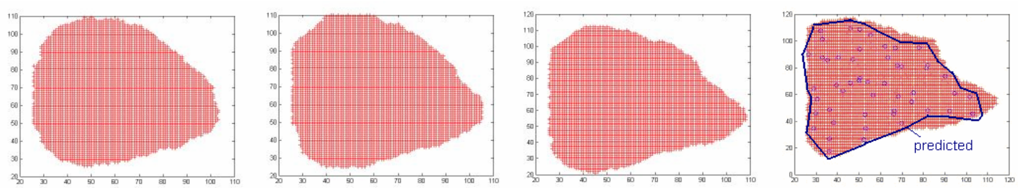

Figure 4. The first 3 images are input shapes (zero wind speed) and the last image is the predicted shape overlaid on top of the ground truth. The blue dots are the clustered data points that represent the shape.

3. Future Work

First, we would build the anomaly detection model: To date, the anomalous areas are empirically determined by the average concentration level of chlorophyll in 60 days. To give more accurate estimation, we are going to apply the statistical estimation methods, such as the spatial anomaly detection method from multiple databases. Instead of using one parameter, we would use multiple variables from multiple channels (e.g. chlorophyll and anomaly) and historical databases. We would focus on what we expect to see and then determine which regions deviate significantly from our expectations (Neil & Moore, 2005). Secondly, we would develop a decision model, for example, using Decision Tree, we can combine the scientist's heuristics and math model together. Finally, we would build the evolvable hybrid modeling interface: Our initial prediction model is based on the spatiotemporal data mining, or so-called machine-learning methods, which heavily depends on the data quality and quantity. To make a more robust detection, tracking and prediction, we need to incorporate with multi-physics models, such as the circulation models, surface temperature models, and HAB development and transport models, etc. Integration of those models has been a challenging task. At the current stage, we have a wind model and cellular automata based biological diffusion model. We expected that there will be many useful multiple physics models available in future. Our ultimate goal here is to build an interface that integrates machine-learning models with incremental multi-physics models as the system evolves.

Acknowledgement

This study is supported by NASA ESTO grant AIST-QRS-04-3031. We are indebted to our collaborators in NOAA, The authors appreciate the comments and suggestions from Dr. Karen Meo, Dr. Kai-Dee Chu, Dr. Steven Smith, Dr. Gene Carl Feldman and Dr. James Acker from NASA. Also, many thanks to Professors Christos Faloulus and Mel Siegel of Carnegie Mellon University for their input.

References

1. Cai, Y. and S. Chung, R. Stumpf, T. Wynne, M. Tomlinson, et al, Spatial Interaction Model for Monitoring Harmful Algae Blooms, proceedings of NASA ESTC-05 Conference, Washington DC, 2005
2. Cai, Y. and Y. Joen, Cellular Metaphor Augmented Spatiotemporal Data, International Conference of Multimodal Interfaces, Trento, Italy, 2005
3. Cai, Y. (editor), Ambient Intelligence for Scientific Discovery, Lecture Notes in Artificial Intelligence, LNAI 3345, Springer, Feb. 2005
4. Framinan, M.B. and O.B. Brown, 1996. Study of the Rio de la Plata turbidity front, Part I: spatial and temporal distribution. Continental Shelf Research, 16(10): 1259-1282.
5. Tomlinson, M.C., R.P. Stumpf, V. Ransibrahmanakul, E.W. Truby, G.J. Kirkpatrick, B.A. Pederson, G.A. Vargo, C. A. Heil., 2004. Evaluation of the use of SeaWiFS imagery for detecting Karenia brevis harmful algal blooms in the eastern Gulf of Mexico. Remote Sensing of Environment, v. 91, pp. 293-303.
6. Salisbury, J.E., J.W. Campbell, L.D. Meeker, C. Vorosmarty, 2001. Ocean color and river data reveal fluvial influence in coastal waters. EOS, Transactions, American Geophysical Union, 82(20): 221, 226, 227.
7. Stumpf, R.P., 1988, Sediment transport in Chesapeake Bay during floods: analysis using satellite and surface observations: Journal of Coastal Research, v. 4 p. 1-15.
8. Stumpf, R.P. and P. Goldschmidt, 1992, Remote sensing of suspended sediment discharge in the western Gulf of Maine, April 1987 100-year flood. Journal of Coastal Research, v. 8, p. 218-225.
9. Viola, P..*Alignment by Maximization of Mutual Information.* PhD thesis, Massachusetts Institute of Technology. 1995
10. Kalman, R.E. A new approach to linear filtering and prediction problems, Transactions of the ASME Journal of Basic Engineering, Vol. 82, Series D., page 35-45, 1960
11. Menzies, T. and Y. Hu, Data Mining For Very Busy People. IEEE Computer, October 2003, pgs. 18-25.
12. Neill, D., Moore, A.W. Sabhnani, and Daniel, K. Detection of emerging space-time clusters. *Proceedings of the 11th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 218-227, 2005.
13. Daniel B. Neill and Andrew W. Moore. Anomalous spatial cluster detection. *Proceedings of the KDD 2005 Workshop on Data Mining Methods for Anomaly Detection*, 2005.
14. Maheshkumar R. Sabhnani, Daniel B. Neill, Andrew W. Moore, Fu-Chiang Tsui, Michael M. Wagner, and Jeremy U. Espino. Detecting anomalous patterns in pharmacy retail data. *Proceedings of the KDD 2005 Workshop on Data Mining Methods for Anomaly Detection*, 2005.
15. Kulldorff,M. A spatial scan statistics. Communications in Statistics: Theory and Methods, 26(6), 1481-1496, 1997
16. Wolfram, S. A New Kind of Science, Published by Wolfram Media (2002), ISBN 1-57955-008-8
17. Mathworks, MATLAB Manual, 2004