

# Knowledge Discovery from Very Large Scientific Databases : “Digital Typhoon” Project as a Case Study



Research Associate,  
Research Center for Testbeds and Prototyping

## Asanobu Kitamoto

He earned B. Sc. from Electronic Engineering, Faculty of Engineering, University of Tokyo in 1992, and Ph.D. from Electronic Engineering, Division of Engineering, University of Tokyo in 1997. From 1997 to 2000, he served as Research associate at National Center for Science Information Systems. After 2000, he is Research associate at National Institute of Informatics. His research interests include visual informatics, data mining, and pattern recognition.

In many disciplines of science, such as bioinformatics and astronomy, gaining more and more popularity is the new style of scientific research in which very large scientific database infrastructure serves as a key component not only for the archiving of comprehensive and consistent observations but also for computationally demanding data analysis/mining processes. Our project, “Digital Typhoon” is one example of such style of research, being started with building the comprehensive archive of typhoon image database infrastructure, from which we try to discover useful knowledge of the typhoon with statistical evidence based on the huge amount of observation data.

The infrastructure, which we name the “typhoon image collection,” archives about 44,000 high-quality images processed from satellite images of the meteorological satellite “Himawari (GMS)” that represent about 260 typhoon sequences in the Northern and Southern Western Pacific basin since the 1995 typhoon season. Early this year, this typhoon image collection started its operation, with automatically collecting new typhoon images as soon as a new typhoon is formed.

The next step is our primary challenge, an image data mining, and for this purpose we put together various ideas studied in related research areas such as data representation, database search engines, query languages, data browsing, and data visualization. Our final goal is to represent quantitative properties of typhoon cloud patterns and discover knowledge and rules that are relevant for solving important real-world problems such as typhoon

analysis and prediction.

In addition, the typhoon image collection is a unique data collection in its own right. It consists of many image sequences of real-world complex systems, so it can serve as a practical benchmark collection not only for image processing and meteorology, but also for a wide range of scientific research. To promote the dissemination of the typhoon image collection, we started to host a dedicated web site for our project on which we provide up-to-date typhoon images and related information and research papers for both experts and non-experts.

- URL : <http://www.digital-typhoon.org/>

K-means clustering of typhoon cloud patterns in the Western North Pacific basin to show the diversity of cloud patterns.

