Utilizing Spatio-Temporal Text Information for Cyclone Eye Annotation in Satellite Data

Shen-Shyang Ho and Ashit Talukder
Jet Propulsion Laboratory
California Institute of Technology
{sho, ashit.talukder}@jpl.nasa.gov

Abstract

Large amount of archived unannotated satellite data is publicly available. The automated retrieval of satellite data from these public domains based on ad-hoc user request is extremely useful to scientists for analysis and as evidence to support scientific hypotheses on weather phenomenon. One efficient approach for such data/information retrieval is to use the publicly available information about weather events (presented as text) to assist in better and more accurate identification of the relevant satellite data. Furthermore, by identifying the weather event in the satellite datasets that is available at finer temporal scales, one can fill in the significant "information gaps" in the text data.

In this paper, we describe an approach for cross-media cyclone track summarization and cyclone eye automated annotation using publicly available satellite data and cyclone information on the World Wide Web. Using a hurricane event as an example, we show (i) automated cyclone eye annotation for satellite data using spatio-temporal text information about the hurricane, and (ii) hurricane track summarization by processing and presenting the multiple media sources. Finally, we discuss some open problems and future work related to cross-media information retrieval and mining for weather events with an emphasis on tropical cyclones.

1 Introduction

An ad-hoc user query system with satellite data retrieval capability is extremely useful for scientists who use the satellite data as evidence to support scientific hypothesis on weather phenomenon. Such a system will need to access and mine publicly available multimedia sources containing text information on weather events (e.g. tropical cyclone[1], and tornado[2]) and satellite data on ftp servers (e.g. the wind field data, sea surface temperatures[3], and precipitation data[4]).

[1] http://www.nhc.noaa.gov/
[2] http://www.tornadohistoryproject.com/
[3] ftp://podaac.jpl.nasa.gov
[4] ftp://trmm.nasa.gov/

The information from each media type (text or remote satellite data), however, is incomplete when considered individually by themselves. For a tropical cyclone (or hurricane), one limitation of tracking data from the web archive is non-continuity. For example, the best track information for any tropical cyclone contains the cyclone location at every six-hour interval on the National Hurricane Center web-site. Any in-between track has to be estimated by interpolation. In Figure 1, one observes that if one were to estimate cyclone eye from the best track information at a time instance when a satellite data is available, the estimated location is obviously not the true cyclone eye location. Conversely, the raw satellite data files in the ftp servers do not have any cyclone eye annotation or meta-data that identified cyclones in the satellite data files.

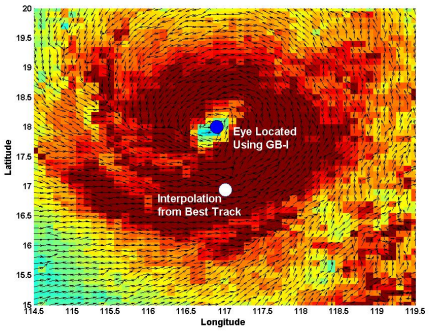


Figure 1: Estimated cyclone eye (White ○) using NHC best track is clearly not the center of the wind circulation in a QuikSCAT wind field image. Dark blue ○ is the cyclone eye located using Algorithm 2 described in Section 5.3.

The information from the two different media types can complement each other. One effective approach to automatically annotate a tropical cyclone eye in satellite data is to utilize the archived cyclone best track information to estimate a search region and then, to identify the cyclone eye from the search region using techniques previously developed in [Ho and Talukder, 2008b; 2008a]. With these additional information about the cyclone track, one can fill in the "gap" in the non-continuous track information.

In this paper, we propose an approach for cyclone track summarization and automated cyclone eye annotation which consists of three main steps: (i) identify satellite data and the regions that includes tropical cyclones, (ii) recognize and estimate the cyclone eye location, and (iii) merge and integrate the tracking data from text sources and cyclone eye estimates from satellite data.

Recently, we proposed two different automated knowledge sharing methodology for cyclone tracking using multiple satellite data sources [Ho and Talukder, 2008a; Talukder and Ho, 2008] without using any web archived track data. For real-time tracking and global cyclone detection in areas containing undocumented cyclones, the approach is feasible and efficient. However, for the purpose of information retrieval and mining on historical data in the Atlantic and Pacific where weather events are well documented, the proposed approach can be improved as the rich and accurate archived track information from text can augment the event detection and recognition accuracy. With the additional track information, a few steps in the cyclone detection component can be removed.

The outline of the paper is as follows. In Section 2, we review some previous work on cyclone eye location and related cross-media annotation approaches. In Section 3, we briefly review the different information and data sources that are used in this paper. In Section 4, we briefly discuss our approach. In Section 5, we describe the four main steps in our approach in detail. In Section 6, we use Hurricane Ike as an example to illustrate our approach. Finally, in Section 7, we describe future work related to cross-media information retrieval and mining for weather events, in particular, the tropical cyclone events.

2 Background

The cyclone eye locations have been determined from QuikSCAT wind field images "subjectively by observation" [Halterman and Long, 2006] with a mean distance error of 33.3 km and 21.2 km from eye location to best-track estimation for images with 25 km and 2.5 km spatial resolution, respectively. [Pao and Yeh, 2008] proposed a procedure consisting of standard image processing and morphology techniques to recognize and locate typhoon from JPEG-format infrared satellite images. [Wong *et al.*, 2008] implemented a genetic algorithm approach to automatically locate tropical cyclone eye from the radar reflectivity data at 3 km Constant Altitude Plan Position Indicator (CAPPI) and the corresponding Doppler velocity data. The approach is tested on sequences of 3 km CAPPI radar reflectivity images, with a range of 256 km and a spatial resolution of a few hundred meters, captured every 6 minutes and their corresponding Doppler velocity. The mean errors range from 0.158 degree to 0.270 degree for three different cyclones near land. [Wong and Yip, 2005] proposed cyclone eye identification using motion field analysis. Their other works include template matching using spiral equation $r = ae^{\theta \cot \alpha}$, where a and α are cyclone domain specific constants, for flow patterns [Yip and Wong, 2006] and region of interest [Wong *et al.*, 2004].

There are many related multimedia approaches for event detection and summarizations. [Medioni *et al.*, 2001] developed a system that detects and analyzes behavior of moving object from and an airborne moving platform. [Agouris and Stefanidis, 2003] proposed an efficient summarization of spatio-temporal deformable (non-rigid) event (object) by modeling the object movement and spatial extent using spatio-temporal helixes. One observes, however, that previous research on either cyclone eye identification or event detection did not include information from other modalities to improve the identification/detection accuracy.

3 Information Sources

Two main type of data sources are used in this paper, namely the archived text sources and the publicly available satellite data sources.

3.1 Text Sources

NOAA National Hurricane Center Website

The National Hurricane Center web-site contains tropical cyclone reports which include comprehensive information on tropical cyclones occurring in the North Atlantic Ocean and the Eastern Pacific Ocean[5]. In particular, it contains post-analysis six-hourly best track locations and intensities for tropical cyclones from 1958 to current year.

NOAA Hurricane Research Division Website

The Hurricane Research Division (HRD) at the Atlantic Oceanographic and Meteorological Laboratory (AOML) performs research to advance the understanding and prediction of hurricanes based on a combination of computer models, theories, and observations. In particular, HRD emphases on the use of data obtained from aircraft field campaigns. NOAA HRD data set web-site includes the aircraft cyclone eye fixes data[6].

The hurricane track available from HRD is based on their wind analysis approach assimilating and synthesizing disparate observations into a consistent wind field. The cyclone eye is the center of the tropical cyclone from this wind field. One notes that

1. Not all (North Atlantic) tropical cyclone tracks are available especially those out in the North Atlantic Ocean; and
2. The track is not the same as NHC best track.

Naval Maritime Forecast Center/Joint Typhoon Warning Center (NMFC/JTWC) Website

The Joint Typhoon Warning Center at the Naval Maritime Forecast Center provides tropical cyclone best track information for the Southern Hemisphere, Northern Indian Ocean and Western North Pacific Ocean from 1945-2008 [7].

3.2 Satellite Sources

QuikSCAT Wind Field Data Source

The QuikSCAT (Quick Scatterometer) satellite carries a specialized microwave radar that measures near ocean surface

[5] http://www.nhc.noaa.gov/pastall.shtml

[6] http://www.aoml.noaa.gov/hrd/data_sub/hurr.html

[7] http://metocph.nmci.navy.mil/jtwc/best_tracks/

wind speed and direction under all weather and cloud conditions [Lungu and et. al., 2006]. The satellite orbits the Earth with a 1800 km wide measurement swath on the Earth surface. It takes about 101 minutes to complete one full polar revolution. The scatterometer is able to provide measurements on a particular region twice per day.

There are 5 data products available to the public. In this paper, we are interested in the Level 2B data which consists of rows of ocean wind vectors in 25km and 12.5km wind vector cells (WVC). There are 25 fields in the data structure for the Level 2B data. We are, however, only interested in the latitude, $l_a \in [-90.00, 90.00]$, longitude, $l_o \in [0.00, 359.99]$, and the most likely wind speed, $ws \in [0.00, 50.00]$ and direction, $wd \in [0.00, 359.00]$.

TRMM-adjusted Merged Precipitation Images

The Tropical Rainfall Measurement Mission (TRMM) satellite carries five remote sensing instruments onboard, namely: Precipitation Radar (PR), TRMM Microwave Imager (TMI), Visible Infrared Scanner (VIRS), Clouds and Earth Radiant Energy Sensor (CERES), and Lightning Imaging Sensor (LIS). The main objective of the TRMM satellite is to monitor and study tropical rainfall. TRMM satellite orbits between 35 degrees north and 35 degrees south of the equator. It takes measurements between 50 degrees north and 50 degrees south of the equator.

In this paper, we use the TRMM-adjusted merged precipitation data (3B42 data). This data is produced using the combined instrument rain calibration algorithm with an optimal combination of 2B31 data (vertical hydrometeor profiles using PR radar and TMI data), (Level) 2A-12 data (vertical hydrometeor profiles at each pixel from TMI data), SSMI (Special Sensor Microwave/Imager), AMSR (Advanced Microwave Scanning Radiometer on board the Advanced Earth Observing Satellite-II (ADEOS-II)) and AMSU (Advanced Microwave Sounding Unit on NOAA geostationary satellites) precipitation estimates. The 3B42 data quantifies global rainfall every 3 hours with each pixel representing a square region of dimension 0.25 deg.

4 Proposed Approach Overview

In this section, we provide an overview of our proposed approach for cyclone track summarization and automated cyclone eye annotation for satellite images (see Figure 2). The proposed cross-media approach consists of five components, namely:

1. **Track data retrieval:** Based on user request, the track information is retrieved from the United States National Oceanic and Atmospheric Administration (NOAA) National Hurricane Center (NHC) and also the Hurricane Research Division (HRD), if available.
2. **Search region estimation:** Based on the retrieved track information, search regions are estimated based on the spatio-temporal information from the track information. The search region estimate ensures reduces the computational cost and better annotation accuracy since the search is now carried out on a smaller region.

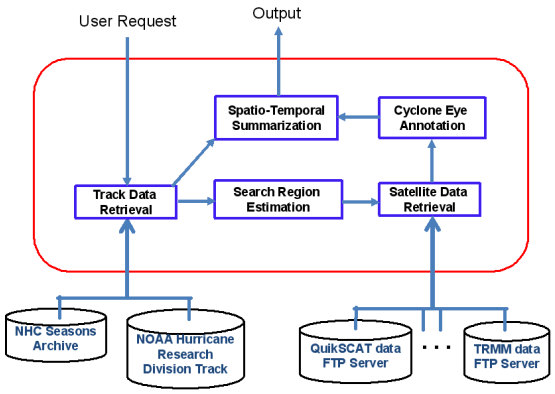


Figure 2: An Overview of the Proposed Cross-Media Approach for Cyclone Track Summarization and Automated Satellite Image Annotation.

3. **Satellite data retrieval and preprocessing:** After satellite data is retrieved, search region estimates and the temporal information from the satellite data files are used to constrain the region that is processed to locate the cyclone. Irrelevant satellite data that do not coincide with the search region are discarded. The search region estimates are then used for subseting the satellite data. Ungridded data will be gridded.
4. **Cyclone eye annotation**. The cyclone eye annotation algorithm is satellite data type dependent. Each satellite data feature requires a different algorithm for cyclone detection [Ho and Talukder, 2008a].
5. **Spatio-temporal summarization.** The cyclone track summarization consists of merging and integrating tracking information from the text archives and cyclone eye annotation information from different satellite data sources.

5 Methodology and Algorithms

In this paper, we focus on (i) search region estimation, (ii) satellite data retrieval, (iii) cyclone detection and eye annotation, and (iv) the information integration from the different media sources.

5.1 Search Region Estimation

Using the tracking information

$$I = \{(x_1, y_1, t_1), \cdots, (x_i, y_i, t_i), \cdots, (x_n, y_n, t_n)\} \quad (1)$$

from the text archive where x_i is the longitude, y_i is the latitude, and t_i is the coordinated universal time (UTC), we construct a search region $R = [(x_s - \epsilon, y_s - \epsilon), (x_s + \epsilon, y_s + \epsilon)]$ (which specifies the top right corner and the bottom left corner of the search region) for satellite data available at t^* such that

$$(x_s, y_s) = (x_i, y_i) + t^* \frac{[(x_{i+1}, y_{i+1}) - (x_i, y_i)]}{(t_{i+1} - t_i)} \quad (2)$$

and

$$\epsilon = \max[abs(y_{i+1} - y_i), abs(x_{i+1} - x_i)]. \quad (3)$$

where $abs(\cdot)$ returns the absolute value and $t_i \leq t^* \leq t_{i+1}$. This corresponds to using interpolated tracks from the text archives to constrain the regions for remote satellite data processing.

The estimated search region is used in two ways:

1. To identify the satellite data that may contain the weather event; and
2. To provide a focus region where an annotation algorithm is applied.

5.2 Satellite Data Retrieval and Preprocessing

Data from different satellite sources are stored in different public servers. One needs to retrieve the relevant data from these distributed servers to a local system for preprocessing. Satellite data files from various sources within the cyclone track time interval $[t_1, t_n]$ are downloaded into a local machine. If a file contains data within an estimated search region, it will be subseted and gridded, else it is deleted from the temporary storage (see Algorithm 1).

After data retrieval, data preprocessing is done differently for different satellite data. In this paper, the QuikSCAT data files are subseted based on the search region estimates. Then, nearest neighbor interpolation is used for image gridding on the subseted region. For the TRMM-adjusted merged precipitation data, no gridding is necessary since it is gridded data. One needs only to subset the region based on the search region estimate.

Input: Cyclone track time interval, $[t_1, t_n]$
Output: $\mathcal{S}$, a set of gridded subseted satellite data (from various sources)
Download satellite data files from various sources with timestamp between t_1 and t_n to a local machine;
foreach *data file,* f **do**
  Compute Search Region estimate, R;
  if f *does not contain* R*;*
  then
    Remove f from the local machine;
  else
    Subset f based on R;
    if f_s*, the subseted data, is NOT gridded* **then**
      grid f_s;
    end
  end
end

Algorithm 1: Data Retrieval and Preprocessing.

5.3 Cyclone Eye Annotation

QuikSCAT Wind Field

Our previous work [Ho and Talukder, 2008a] involved unconstrained searches for undocumented cyclones that required segmentation, classification, and tracking outputs for cyclone detection over the entire globe. A few of the preprocessing steps are redundant when the very likely region (search region estimate) that the cyclone exists is estimated from the archived tracking information. Again, one notes that this holds true only when one is mining historical satellite data for documented cyclones, which is valid in populated regions in the North Atlantic and Pacific regions(cyclones may not be documented in sparsely populated regions and in open oceans).

Input: Gridded subseted QuikSCAT Data with m pixels from Algorithm 1.
Output: E, Vortex (Eye) Center
foreach *Pixel* i **do**
  Compute the normal vector $\hat{n}_i$ to the direction vector $\hat{d}_i$;
  Calculate which 8-neighbors $\hat{n}_i$ is pointing;
  Update the neighbor count N_k of the pixel k $\hat{n}_i$ is pointing;
  Update l_k, list of neighbor pixels, pointing at k
end
$MaxNeighbor = \max_{1 \leq k \leq m} N_k$.
$VC = \{i \mid N_i > MaxNeighbor - 1\}$;
foreach $j \in VC$ **do**
  $root \leftarrow j$;
  $Count[j] = TreeSize(root, l_{root})$
end
$E = \arg\max_{j \in VC} Count[j]$;

Algorithm 2: QuikSCAT Cyclone Eye Annotation Algorithm.

The cyclone eye annotation algorithm is shown in Algorithm 2. Intuitively, Algorithm 2 works as follows.

1. An arc (directed edge) is computed for each pixel based on the normal vector of the wind direction at the pixel. The arc points to one of the eight neighbors for each pixel.
2. The likely eye locations are those that have many neighbors pointing at them. These pixels are in the VC set.
3. Compute the size of all the spanning trees using the pixels in VC as root nodes for spanning trees.
4. The root node that grows the largest spanning tree is the cyclone eye.

TRMM Precipitation Data

Our previous work [Ho and Talukder, 2008a] involved TRMM daa processing over a constrained search region based on Kalman filter estimates from the previous observation which generates larger uncertainty due to the large time interval (maximum of 3 hours time lapse) from the previous satellite image. The subseted TRMM precipitation data is likely to be more focused and accurate since it is derived from the best track information. As a result, the image segmentation and centroid computation (see Algorithm 3) for cyclone eye annotation is expected to have a higher accuracy since the constrained search region is closer to the true cyclone location.

Input: Subseted TRMM precipitation data, P, from Algorithm 1.
Output: E, Vortex (Eye) Center
$R = Segmentation(P)$;
$E = Centroid(R)$;

Algorithm 3: TRMM Cyclone Eye Annotation Algorithm.

5.4 Information Merging and Integration

After the cyclone eyes information, I_T and I_Q, are located from the TRMM and QuikSCAT satellite data respectively, one merges I_T and I_Q, together with the archived best track information and aircraft eye fixes information.

During merging, if the time instance is identical, we choose the one that has the highest confidence. The confidence level for information from different sources using our approach is shown in Table 1, from the highest (1) to the lowest (4). One notes the archived text information has higher confidence level than the satellite data as these text information are analyzed and compiled by experts and scientists. In this paper, we assume the aircraft eye fixes have the highest confidence as they are physically measured by flying an aircraft over the hurricane even though scientists usually prefer the NHC best track information as it contains complete information (i.e., all tropical cyclone best tracks are available).

Confidence Level	Data Sources
1	NOAA HRD aircraft eye fixes
2	NOAA NHC best track
3	QuikSCAT wind field
4	TRMM precipitation rate

Table 1: Confidence level for different data sources.

6 Example: Hurricane Ike 2008

Using Hurricane Ike that occurred in the North Atlantic Ocean in 2008 as an example, we present some results for

1. satellite image auto-annotation of hurricane eye using techniques.
2. hurricane track summarization using our proposed cross-media approach.

Figure 3 shows integrated track for Hurricane Ike based on NHC best track, NOAA aircraft eye fixes, and satellite tracks derived from TRMM and QuikSCAT images using a Kalman filter-based tracking methodology in [Ho and Talukder, 2008a].

Figure 4 shows the hurricane track summarization based on our proposed approach. The NHC archive best track information and NOAA Hurricane Division aircraft eye fixes are used to estimate a search region for each satellite image for hurricane eye auto-annotation.

By comparing Figure 3 and Figure 4, we observe a much smoother track summarization when the archived aircraft and best track information is used to assist the automated annotation step. The high quality (confidence) track information improves the annotation accuracy and results in a smoother track summarization.

7 Discussions and Future Work

This paper does not attempt to propose any ground breaking idea on cross-media mining and information access. However, it highlights the usefulness of cross-media information as a powerful tool to solve practical problems in the Earth/climate science research. Our future work on utilizing cross-media information include the following.

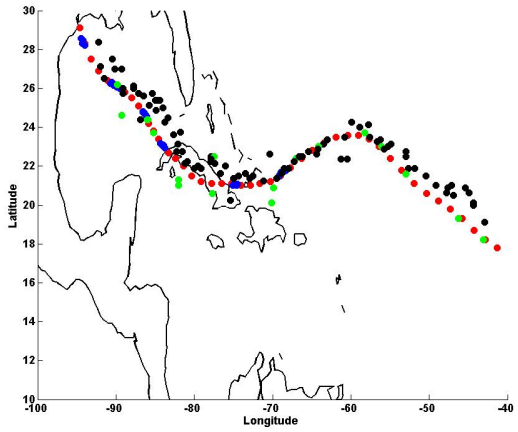


Figure 3: Eye location summarization for Hurricane Ike. Eyes located in TRMM and QuikSCAT images without using NHC and NOAA information. Red ○: NHC best track. Blue ○: NOAA aircraft eye fixes. Black ○: Eye location in TRMM image. Green ○: Eye location in QuikSCAT satellite image.

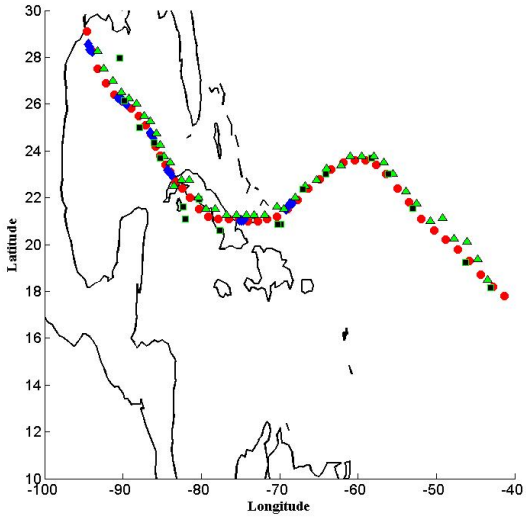


Figure 4: Eye location summarization for Hurricane Ike. Eyes located in TRMM and QuikSCAT images using NHC and NOAA information. Red ○: NHC best track. Blue ○: NOAA aircraft eye fixes. Green triangle: Eye location in TRMM image. Black square: Eye location in QuikSCAT image.

1. **Feature learning via cross-media annotation**

   The TRMM satellite image is known to be weak indicator for cyclone as precipitation does not implied cyclone [Ho and Talukder, 2008a]. From Figure 3, one observes that the capability to locate cyclone eyes us-

ing Algorithm 3 and Kalman filter-based tracking is very weak as located cyclone eyes are further away from the NHC best track. From Figure 5, one clearly sees from the TRMM images that the centroid of the segmented region is not the cyclone eye. Hence, Algorithm 3 needs to be improved (Improved transfer learning solutions for multi-source cyclone classification have been discussed in [Talukder and Ho, 2008]).

Using the best track information available every six hours, one can accurately annotate and segment the TRMM satellite images at 0600, 1200, 1800, and 0000. With this accurately annotated and segmented images, one can gather a set of positive training examples which can be used either for training or to select more discriminating features for learning.

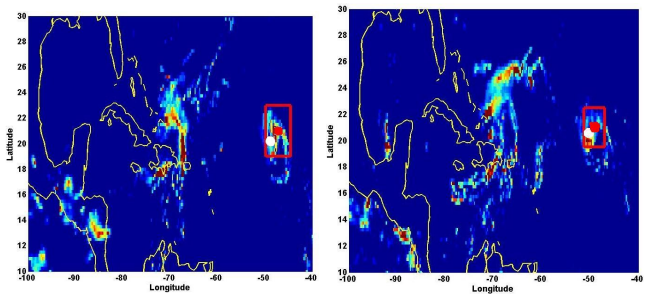


Figure 5: TRMM satellite image of Hurricane Ike on Sep 3 0600 (Left) and 1200 (Right). White ○ : Cyclone Eye Estimates using NHC best track. Red ○: Eye location using Algorithm 3 and Kalman filter-based tracking.

2. **Improve cyclone eye estimation for satellite images via cross-media spatio-temporal knowledge**.

   The archived cyclone text information provides prior knowledge for cyclone eye annotation in the satellite images. An open problem is how this prior knowledge can be used to improve eye estimation when the cyclone eye is not in the QuikSCAT swath image (see Figure 6 (Left)) or when it is near/on land in the QuikSCAT swath image (see Figure 6 (Right)).

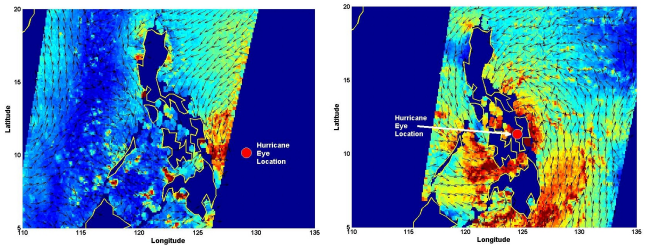


Figure 6: Typhoon Fengshen occurring in the Western Pacific Ocean in 2008. (Left) Cyclone partially occluded with its eye not in the swath image. (Right) The cyclone eye is near/on land.

Acknowledgments

This work was carried out at the Jet Propulsion Laboratory, California Institute of Technology with funding from the NASA Applied Information Systems Research (AISR) Program. The author is supported by the NASA Postdoctoral Program (NPP) administered by Oak Ridge Associated Universities (ORAU) through a contract with NASA.

References

[Agouris and Stefanidis, 2003] Peggy Agouris and Anthony Stefanidis. Efficient summarization of spatiotemporal events. *Commun. ACM*, 46(1):65–66, 2003.

[Halterman and Long, 2006] R. R. Halterman and D. G. Long. A comparison of hurricane eye determination using standard and ultra-high resolution quikscat winds. In *IEEE International Geoscience and Remote Sensing Symposium*, pages 4134–4137, 2006.

[Ho and Talukder, 2008a] Shen-Shyang Ho and Ashit Talukder. Automated cyclone discovery and tracking using knowledge sharing in multiple heterogeneous satellite data. In Ying Li, Bing Liu, and Sunita Sarawagi, editors, *KDD*, pages 928–936. ACM, 2008.

[Ho and Talukder, 2008b] Shen-Shyang Ho and Ashit Talukder. Automated cyclone identification from remote quikscat satellite data. *IEEE Aerospace Conference*, 2008.

[Lungu and et. al., 2006] Ted Lungu and et. al. *QuikSCAT Science Data Product User's Manual*. 2006.

[Medioni *et al.*, 2001] Gérard G. Medioni, Isaac Cohen, François Brémond, Somboon Hongeng, and Ramakant Nevatia. Event detection and analysis from video streams. *IEEE Trans. Pattern Anal. Mach. Intell.*, 23(8):873–889, 2001.

[Pao and Yeh, 2008] Tsang-Long Pao and Jun-Heng Yeh. Typhoon locating and reconstruction from the infrared satellite cloud image. *Journal of Multimedia*, 3(2):45–51, 2008.

[Talukder and Ho, 2008] Ashit Talukder and Shen-Shyang Ho. Classification from disparate multiple streaming data sources. In *NIPS-08 workshop, Learning from Multiple Sources*, 2008.

[Wong and Yip, 2005] Ka Yan Wong and Chi Lap Yip. An intelligent tropical cyclone eye fix system using motion field analysis. In *ICTAI*, pages 652–656. IEEE Computer Society, 2005.

[Wong *et al.*, 2004] Ka Yan Wong, Chi Lap Yip, Ping Wah Li, and Wai Wan Tsang. Automatic template matching method for tropical cyclone eye fix. In *ICPR (3)*, pages 650–653, 2004.

[Wong *et al.*, 2008] Ka Yan Wong, Chi Lap Yip, and Ping Wah Li. Automatic tropical cyclone eye fix using genetic algorithm. *Expert Syst. Appl.*, 34(1):643–656, 2008.

[Yip and Wong, 2006] Chi Lap Yip and Ka Yan Wong. Identifying centers of circulating and spiraling flow patterns. In *ICPR (1)*, pages 769–772. IEEE Computer Society, 2006.