Stream Handler System: An Experience of Application to Investigation of Global Tropical Cyclogenesis

D. Ermakov[a,*], A. Chernushich[a], E. Sharkov[b], Ya. Shramkov[b]

[a] Institute of Radioengineering and Electronics of RAS, Fryazino Dept., 141195, 1 Vvedenskogo sq., Fryazino, Moscow rgn., Russia – dima@ire.rssi.ru, andrey@ire216.msk.su
[b] Institute of Space Research of RAS, 117997, 84/32 Profsoyuznaya str., Moscow, Russia – e.sharkov@mail.ru, cinlun23@gmail.com

Abstract – An investigation of tropical cyclogenesis by means of remote sensing data can be viewed as a special kind of video stream analysis extended with parallel processing of ancillary data streams (generally speaking, of arbitrary nature). The use of a powerful programming platform of general purpose, oriented on data streams processing, allows interactive model testing, convenient data visualization, matching results of simulation against information from independent sources ("GLOBAL-TC" database developed in Space Research Institute of RAS) etc. The authors have tested the applicability of "Stream Handler" programming platform, developed by the "File X Ltd", Israel, for the studies of tropical cyclogenesis being performed in Space Research Institute. The principals of adaptation and application of the platform means to the task as well as methods and some results of remote data analysis are discussed in the present paper on example of processing of SSM/I data coupled with information of the "GLOBAL-TC" database.

Keywords: Cyclones, Radiometry, Technology, Vision, Databases

1. INTRODUCTION

The basis of investigation of global tropical cyclogenesis (GTC) is extensive retrospective analysis of statistics of observations of tropical cyclones (TC) and tropical disturbances (TD) coupled with synchronous with them (historical) remote data. Effectiveness of such an analysis is determined by availability of comprehensive datasets for both statistical and remote data.

Statistical information on TCs and TDs used was extracted from the "Global TC" database (Pokrovskaya, Sharkov, 2006), described in some detail later in this paper. This database stores the data of all actual worldwide observations of TCs and their initial and terminal stages, and provides the TC parameters like coordinates of the eye, pressure, maximum wind speed etc in unified manner with temporal resolution of 3 – 6 hours.

These data were coupled with the SSM/I remote data used to build global maps of atmospheric water vapour. The SSM/I remote data, though very informative for the analysis of global tropical cyclogenesis, have their intrinsic shortage: wide spatial lacunas in equatorial and subequatorial zones. To overcome this drawback some techniques of merging the data are usually implemented: either integrating the data from one satellite through several days of observation (which reduces the time resolution), or combining the data from several satellites (which introduces big uncertainties in the reference time of observation in global scale), or some combination of the two approaches. On the other hand, the analysis of "Global TC" data shows that the development of a TC from its initial stages often happens in a very short time, from half a day to several hours. Hence it is not a rare case when the critical for understanding of the TC development processes take place in a "spatial lacuna", not covered by remote observations. Obviously, it is very important to have global maps of brightness temperatures distribution at least every 6 hours or even more often.

The first authors' approach to investigation of GTC was based on a fairy common idea. The SSM/I data collected during one day and reduced to a regular grid (GRID data) were considered as an "information unit" or a "frame" on a time scale. Each frame had its unique time label (the date of observation). The records from the "Global TC" were matched against the timestamps of the frames to extract the information on reported TCs and TDs for that dates, and to impose their tracks on corresponding frames. It appeared natural to represent remote data as a video stream (of the data themselves) supplemented with a textual stream of timestamps (one label per frame) and with a parallel stream of synchronous parameters of reported TCs and TDs from the "Global TC" database.

However, a direct software implementation of this approach, a kind of "advanced video-player", though effective for data visualization and preliminary analysis, revealed significant problems. First, multichannel data were hard to be visualized and analysed altogether. Second, every change in processing algorithm (like introducing or removing any processing block or filter) required total rebuilding of the program code followed by extensive debugging and testing. Third, the run-time comparison of outputs of different algorithms was impossible.

Finally, the need was realized for use of a powerful stream-oriented processing platform of general purpose, allowing an easy adaptation to task of GTC investigation. As such, a Stream Handler (The "File X Ltd", 2007–2010) was chosen. Usage of this platform not only widened the opportunities of data visualization and primary analysis, but also stimulated implementation of processing techniques developed in the other knowledge domains (computer vision and video processing), which has already resulted in suggesting a new interpolation method for obtaining global maps of remote observations with time step of 6 hours or better and significantly diminished or totally eliminated spatial lacunas. Such global maps were built and matched against the actual data of "Global TC" database. The interpolation method, some results of primary qualitative analysis as well as the perspectives and the aims of further investigations are discussed in the present paper.

2. DATA OVERVIEW

In present work the two (independent) data sources were used. The information on actually reported TCs and TDs was extracted from the "Global TC" database. The SSM/I remote data were used to build global maps of atmospheric water

* Corresponding author

vapour. For the aims of preliminary analysis the data of year 2005 were selected and the attention was mainly attracted to the time interval of August, 23 through September, 3 (Katrina TC).

2.1 The "Global TC" database

The "Global TC" database was developed in Space Research Institute of RAS (IKI RAS) for investigations into tropical cyclogenesis and climatology tasks. The database is routinely populated with actual data of globally reported TCs and TDs and currently contains information for years 1983 – 2010.

The database consists of tables of two kinds: chronology and evolution ones. Chronology tables serve primarily for the purpose of fast search and referencing the cyclones. A sample record of a chronology table of the "Global TC" database is given in Table 1 (some fields are omitted to save space).

Table 1. Chronology record for Katrina TC in "Global TC"

N	Number	Name	Lat	Long	Dates
60	0511	KATRINA	25.0	-73.0	08/23-09/03

Evolution tables contain records describing individual TC, its track, stage, and some physical parameters like pressure in the eye, maximum wind speed etc. Sample records for Katrina TC are given in Table 2 (some fields are omitted to save space).

Table 2. Katrina TC track stored in "Global TC" (fragment)

N	Stage	Date	Time	Lat	Long	Pres	Wind
…	…	…	…	…	…	…	…
11	T	08/25	21	26.1	-79.9	985	33
12	T	08/26	3	25.5	-80.7	984	33
13	T	08/26	9	25.3	-81.5	987	33
14	T	08/26	15	25.1	-82.2	981	33
15	T	08/26	16	25.1	-82.2	971	44
16	T	08/26	21	24.8	-82.9	965	44
…	…	…	…	…	…	…	…

With the information from these tables locating active TCs for the given date and time was a simple two-step procedure: first, from chronology tables all active TCs were enumerated; second, coordinates of active TCs for the given time were extracted from corresponding evolution tables along with other necessary parameters. Comprehensive information on the "Global TC" database can be found in (Pokrovskaya, Sharkov, 2006).

2.2 Remote microwave data

In the present work the SSM/I remote data collected in the year 2005 by satellites F13, F14, and F15 were used. Based on these data in GRID format (absolute brightness temperatures on a regular grid) global maps of atmospheric water vapour were built by formula suggested in (Ruprecht, 1996):

$$W = 131.95 - 39.50 \ln(280 - T_{22V}) + 12.49 \ln(280 - T_{37V}) \quad (1)$$

where T_{22V}, T_{37V} are brightness temperatures measured at 22 GHz and 37 GHz, vertical polarization correspondingly.

For each water vapour map (one per day) a corresponding timestamp was generated in textual form which contained the date and the conditional time 00:00:00 (midnight) of observation. The sequence of all water vapour maps formed a video stream, while the sequence of their timestamps formed a parallel textual stream, and these streams were saved to an AVI file altogether. This file (along with the data of the "Global TC" database) was the input data for further processing.

It must be emphasized that reducing multichannel SSM/I data to a single video stream was performed for the aims of preliminary qualitative analysis and better presentation of results, and was not a restriction of the Stream Handler.

3. STREAM HANDLER OVERVIEW

The Stream Handler system is a powerful programming platform for stream-oriented distributed processing developed by the "File X Ltd", Israel. In terms of high-level programming every processing algorithm is represented by a graph (see Figure 1). Nodes of a graph (big rectangles) are processing units called gadgets; each gadget is associated with specific function (tuneable filter) transforming input data into its output.

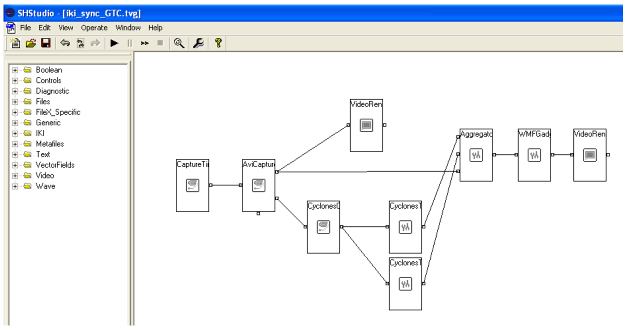

Figure 1. Stream Handler Studio

To make gadgets exchange data a developer uses wires (lines in Figure 1) which connect output and input pins (small squares) of selected gadgets. Special kind of gadgets (capture gadgets) is used to read input data from files or devices. Other terminal gadgets (render gadgets) perform output data visualization or/and saving to a file. Stream Handler provides an extensive set of libraries for reading, saving and processing data in various widely used formats. New gadgets and data formats can be included into the system by means of lower-level programming.

With the help of graphical user interface of Stream Handler Studio a developer creates and runs a processing algorithm by inserting and connecting gadgets and defining data sources (files or devices). Graphical interface allows visualization of intermediate and final results of processing, tuning the processing filters and editing the graph while processing is running. More technical documentation on Stream Handler system can be found at (The "File X Ltd", 2007–2010).

4. APPLICATIONS

To overcome (at least partly) known shortage of remote data, spatial lacunas and poor temporal resolution, a new method of data interpolation was suggested. In present paper details of program realization are avoided while discussed are the logic of the approach, results achieved and the ways to advance them.

4.1 Data assimilation

As described in subsection 2.2 the remote data were stored in AVI files readable by Stream Handler AVI capture gadget. This gadget read video and textual streams of a file in parallel frame by frame. Then video frames could be processed by means of standard and/or specially developed filter gadgets, while text frames (timestamps) were used as input data for other specially developed capture gadget which searched through the "Global

TC" database (see subsection 2.1) and sent records of active TCs to its output. Eventually, the active TCs were mapped over corresponding video frames to visualize on screen.

4.2 Data analysis

Once the initial maps of water vapour had lacunas and a 1 day temporal resolution, large parts of tracks of TCs were not covered by remote observations. The first stage of data analysis aimed to find the best approach to data preparation for future interpolation, and the optimal method of interpolation itself.

Basically, the two principal opportunities were considered: either using the data of only one satellite to eliminate spatial lacunas by processing several sequential frames altogether (an analogue to removing defects on a film of poor quality), or using the data from many satellites to build composite frames without spatial lacunas and to interpolate data in time domain to increase the temporal resolution (an analogue to increasing video frame rate, e.g., from 16 to 25 frames per second).

The first approach (using data of one satellite) seemed to be more attractive, but was found to be not feasible due to the big sizes and relatively slow motion of the lacunas: 3 to 4 sequential frames were required to cover every grid point at least once, i.e. 3 to 4 days of observations. During such a time interval the water vapour distributions change significantly, so that interpolation technique can not perform correctly.

The second approach (using data of many satellites altogether) had a serious disadvantage introducing uncertainties in time of observation. Using a range of Stream Handler gadgets, many variants of merging the data were tested (like averaging, alteration etc) and finally one of the best approaches was found to be as follows. All the data of one satellite (e.g. F13) collected in ascending mode was taken as a basis (reference data). Empty spaces were then complemented with the data of other satellites (one by one) also collected in ascending mode. Only if some lacunas remained after this procedure the data collected on descending modes were used to eliminate them. This algorithm, though giving the most "smooth" maps, obviously didn't solve the problem of "non-uniform" time scale. In terms of video stream processing the resulting maps had a complex frame scan. A method to partly overcome this disadvantage was suggested, however, on the first stage of interpolation these composite frames were considered as ordinary video frames.

4.3 Advanced techniques: data interpolation

The idea of temporal interpolation is based on the assumption that sequential frames of a video stream fix the stages of some continuous process. In this case, calculating the vector field of apparent shifts of all elements of one frame which lead to its transformation into the next frame, one may reconstruct (in linear approximation) all intermediate stages of the process by dividing this vector field by some scalar factor and applying estimated shifts to the elements of the first frame.

Stream Handler contains a gadget which performs a linear motion approximation described. This gadget was applied to video stream consisted of composite maps of atmospheric water vapour (see subsection 4.2), and the scalar factor was set to 8. Hence, while the initial stream had one day (24 hours) temporal resolution, the interpolated stream had a 3 hour temporal resolution. The unlimited increase in temporal resolution didn't make sense due to nonlinearity of real physical processes in atmosphere; 3 hour resolution was set as upper limit for two reasons. The first reason was better matching against the "Global TC" data with its temporal resolution 3 – 6 hours. The second reason was implementation of a "local time adjustment" algorithm (see Figure 2).

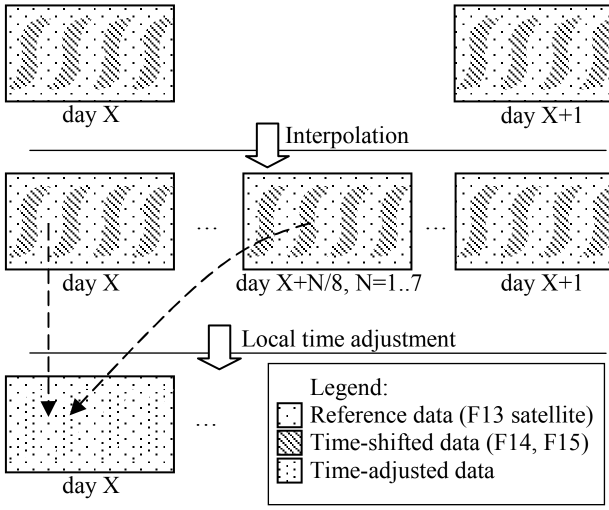

Figure 2. Temporal interpolation and local time adjustment of global maps of atmospheric water vapour

As it was mentioned before, the initial maps of water vapour had temporal discontinuities due to merging data of several satellites. Visually, it resulted in sharp "tears" and displacements of large objects and their edges on both initial and interpolated video frames. These discontinuities were of order of several hours and, hence, were irremovable in the data stream with one day resolution. However, the temporal interpolation made it possible to partly reduce or eliminate them. Indeed, if some region on initial map was filled with the data collected, e.g., 9 hours earlier than the data in the neighbouring background, then the synchronous data for that region could be found on the interpolated frame with 9 hour shift (see Figure 2).

Hence, an algorithm of "local time adjustment" for interpolated video stream was implemented as follows. On every video frame the regions of time shift (collocated with spatial lacunas in corresponding reference data from F13 satellite) were found. The data in these regions for every frame were replaced with the data of the same regions on other, "matching" frame. A "matching" frame was iteratively set to be the next frame, the next after next one etc, as well as the previous one etc. A matching rule which provided the smoothest (visually estimated) results in terms of continuity of objects' edges was considered to be optimal for local time adjustment.

As a result of procedures described above new interpolated and time-adjusted video streams were created which contained atmospheric water vapour maps of global coverage with effective temporal resolution of about 3 hours. To make some estimation of the quality of the resulting stream the "Global TC" data were mapped over interpolated frames. Once TCs had characteristic signatures on water vapour maps forming a circle of high concentration values, an indication of adequate interpolation was a good collocation of the centres of these circular objects on the video frames with the coordinates of reported TCs. Some results of this analysis are discussed below.

5. DISCUSSION OF RESULTS

An example of temporal interpolation of atmospheric water vapour maps is present in Figure 3. Each frame is marked with its timestamp in its lower-right corner. The actually reported TCs and TDs are indicated by small squares labelled with corresponding TC name (if any). Narrow dark “scratches” on some frames are the remainders of spatial lacunas not totally covered by satellite data because only ascending swaths were used in this example.

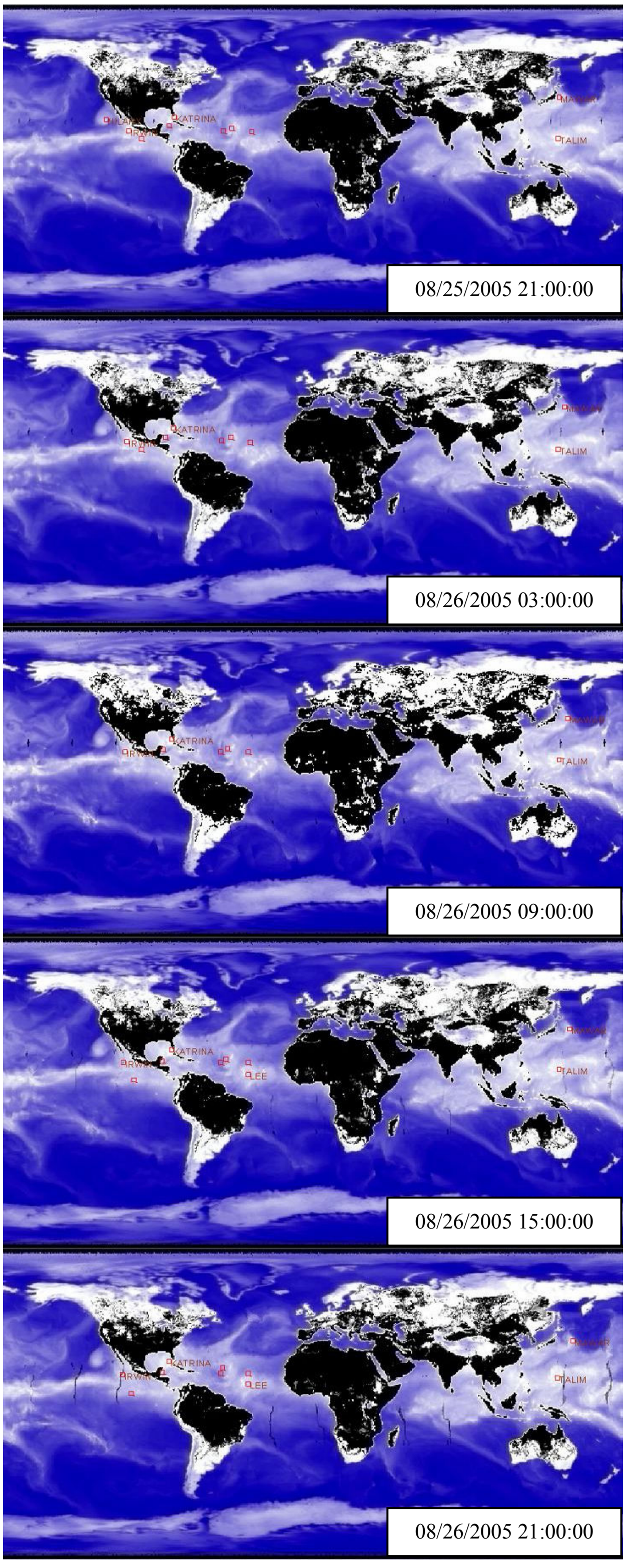

Figure 3. Atmospheric water vapour global maps with 6 hour temporal resolution (linear interpolation)

As it can be seen the results of interpolation look adequate when matched against actual positions of reported TCs. However, in a long run many cases of lags and/or leads of actual data can be observed. This is mainly due to the significant nonlinearity of real TC motion. Obviously, nonlinearity must be included into the interpolating algorithm to better reflect the real physical processes in atmosphere. One way to do so is to take into account the dependency of estimated vector field of motions on time. Indeed, this vector field (which describes the transformation of a reference frame into next one) evolves for every pair of reference frames, but in current software realization it is considered constant between two reference frames, i.e. “frozen” for every 24 hour interval. A more physically sensible approach is to smooth the field of motions by interpolating it with better temporal resolution, probably equal to that of the resulting (interpolated) video stream.

Other issue to be discussed is the artefacts (also noticeable in Figure 3) of merging the data of several satellites. For many reasons a simple approach of “local time adjustment” can not completely overcome this problem. Maybe the most promising approach is to use the satellite data in SWATH format (original scans) which provides precise knowledge of observation time for every pixel on a frame along with better spatial resolution. The main challenge here is to advance standard processing algorithms to perform on irregular grid formed by sensor scans.

6. CONCLUSION

A method of interpolation of remote data was suggested in order to overcome their known shortage: spatial lacunas and poor temporal resolution. This method was based on the suggested concept of remote data as a video stream with complex frame scan. The preliminary analysis showed that the method gave adequate results in linear approximation, though introduction of nonlinearity could be a significant step to advance its performance. A complicated and hard work on the software implementation and on testing and analyzing the algorithm by processing large amounts of data of different kind became feasible only by means of a powerful programming platform for stream-oriented distributed processing.

REFERENCES

I.V. Pokrovskaya, E.A. Sharkov, “Tropical Cyclones and Tropical Disturbances of the World Ocean: Chronology and Evolution. Version 3.1 (1983–2005)”, M: Poligraph servis, 2006, 728 p.p.

E. Ruprecht, “Atmospheric water vapour and cloud water: an overview”, Adv. Space Res., vol 18, no 7, p.p. (7)5–(7)16, 1996.

The “File X Ltd”, 2007–2010. “Stream Handler project”, http://www.thefilex.com/SH/

ACKNOLEDGEMENTS

The SSM/I data were provided by the Global Hydrology Resource Center (GHRC) at the Global Hydrology and Climate Center, Huntsville, Alabama, US (http://ghrc.nsstc.nasa.gov/).

The authors are grateful to the “File X Ltd”, Israel for fruitful cooperation.

The work was partly supported by the RFBR grant N 09-05-01019-a.