Feature Based Visualization of Geophysical Data

Qing Yang and Bahram Parvin
Computing Sciences
Lawrence Berkeley National Laboratory
Berkeley, CA 94720

Abstract

Our goal is to develop a feature based framework for data mining and forecasting from geophysical data fields. These data may be generated from either numerical simulation models or space based platforms. This paper focuses on pertinent features from sea surface temperature (SST) fields that are observed with the AVHRR satellite. Our contribution consist of three components: (1) A method for tracking feature velocities from from fluid motion with incompressibility constraint, (2) a method for localizing singular events such as vortices and saddle points from underlying feature velocities, and (3) application of our protocol to 12 years of high resolution real data to reveal novel seasonal and inter-annual trends based on computed events.

1 Introduction

The current generation of environmental satellites and numerical simulation models generate a massive amount of spatio-temporal data. These data represent different geophysical fields (such as sea surface temperature, wind stress, precipitation) at different spatial and temporal resolutions. For example, the NASA MODIS sensor could potentially generate 200 Gigabytes of data per day, where 50-100 Gigabytes of it are ocean products. Similarly, a 20 year ocean circulation model, with $O(5-10)$ Km resolution, at 20 minute time steps produces about 4 Terabytes of prognostic data. Such a vast amount of data poses a serious performance, management, and comprehension problem. Furthermore, existing comparative analyses of climate codes are based on aggregate measures such as average temperature and raindrops, and they ignore higher-level feature activities. Particular features of interest include vortices and saddle points. Ocean vortices are an important component of global circulation because they are an efficient transport and mixing mechanism. Similarly, atmospheric vortices are important for meteorological studies, e.g., their frequency, strength, duration, and tracks. If vortices can be detected and tracked, pertinent questions can be asked about other sources of data.

This paper focuses on a feature-based representation of spatio-temporal sea surface temperature (SST) data for providing an intelligent summary of the underlying images and to facilitate correlation studies between different types of events (in different fields) for hypothesis testing, comparative analysis of numerical simulation and observational data, and efficient visualization of meaningful information. An example of SST data is shown in Figure 1. This is achieved by computing the feature velocities from the fluid motion and then computing pertinent features from the computed vector field. We have applied our methods to 12 years of sea surface temperature data. The data set has a spatial resolution of $9-18$ Km that is recorded by the AVHRR sensor once every two days. Specifically, we present

- A novel formulation of flow field computation that incorporates an incompressibility constraint for tracking fluid motion. The algorithm is then implemented through a multigrid representation to reduce the computational complexity.
- A robust approach for detecting and tracking vortices and saddle points from the vector field representation. Vortices are important in two ways. First, they are an important component of the global circulations. Thus, their duration, direction, and localization are valuable to climatic studies. Second, they are singular events that provide a compact representation of the spatio-temporal images. Thus, they can be used as indices for representation and quick access. Additionally, we show that our approach is applicable to other geophysical fields as well.
- An interesting climatic result indicating that the number of singular events (in SST data) is seasonally correlated and that the vortices have a preferred localization. This is observed by accumulating features over a 12 year period and representing their occurrences as a probability density function.

*This work is supported by the Director, Office of Energy Science Research, Office of Computation and Technology Research, Mathematical, Information, and Computational Sciences Division of the U. S. Department of Energy under Contract No. DE-AC03-76SF00098 with the University of California. The publication number is LBNL-45326. E-mail: qyang,parvin@portnoy.lbl.gov

Organization of paper is as follows. Section 2 reviews previous models of flow field computation for fluid motion. Section 3 outlines the details of our approach that includes flow computation and feature extraction from SST data. Section 4 shows an application of feature based analysis for climatic studies. Section 5 concludes the paper.

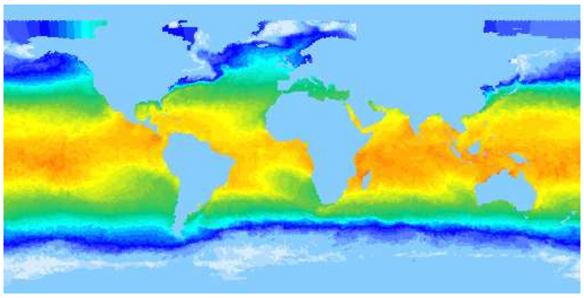


Figure 1: An example of SST image (Date: 1-2-1998).

2 Previous Work

Measurement and analysis of feature velocities is often referred to as computation of optical flow in the imaging literature. Review and enhancement of these techniques can be found in [3, 7]. However, tracking fluid motion poses additional constraints that have been addressed by only a few papers. In this context, we are not interested in tracking compressible fluid that focuses on clouds and images obtained from turbulent flows [9, 12]. These systems pose an affine model over the time varying imagery and ignore the inherent incompressibility constraint.

Cohen and Herlin [4] proposed a non-quadratic regularization technique for solving the optical flow constraint equation and applied it to oceanographic images. Their approach is applicable to irregularly spaced images with missing data. The regularization problem was solved by finite difference methods with finer tessellation near the motion boundary. Their method does not incorporate any constraints imposed by the fluid motion. Amini [1] extended the Horn-Schunck equation to include fluid X-ray images of contrast velocity in arteries. This was expressed in terms of zero divergence of flow field to *simplify* the solution. A major drawback of his approach is that the corresponding partial differential equations are of a higher order.

Raw oceanic data consist of satellite swath data and artifacts introduced by clouds. Modern oceanic processes perform cloud removal and time-space interpolation to produce dense data for daily, regional, and global fields. These data are then calibrated with in-situ measurements to produce a dense temperature map at each point on the oceanic grid. Dense oceanic data is the starting point for the proposed analysis that should satisfy the fluid incompressibility constraint. Additionally, because the data resides in the spherical coordinate system, special treatment is needed. Our formulation operates on dense data, does not require solution of higher order PDEs, and has been implemented in a multigrid framework for better computational efficiency.

With respect to feature extraction from vector field, previous approaches rely on eigenvalue analysis from the local Jacobian [10, 11]. Our experience indicates that feature extraction is highly dependent on the smoothness of flow computation. If the flow field is regularized (as it is in our case) then events of interest can be easily localized. We will present an elegant approach for detection of vortices and saddle points that has proved to work effectively in our data sets.

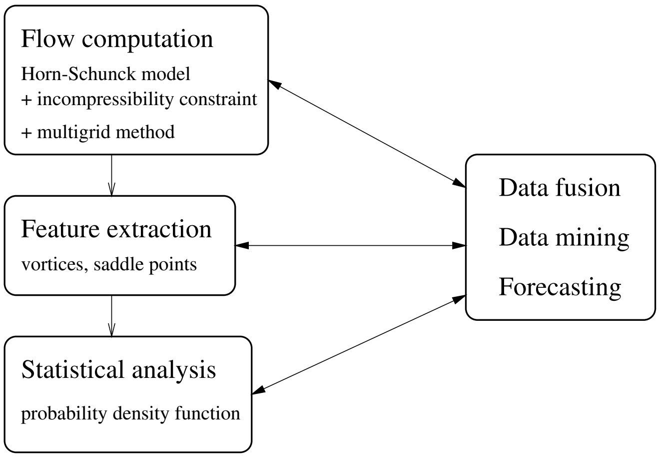


Figure 2: Block diagram of our SST system

3 Approach

Our ultimate goal is to represent each geophysical field with a feature set to explore data fusion, data mining, and forecasting. The lay-out is shown in Figure 2, and each step is summarized below.

1. *Computation of feature velocities:* The classic Horn-Schunck model is integrated with the incompressibility constraint (i.e. zero-divergence condition) and implemented using multigrid finite-difference method. The feature velocities are computed normal to the isothermals.

2. *Extraction of events:* Singular events corresponding to vortices and saddle points are extracted from the underlying vector field. These events provide a representation of raw data that is not only information preserving, but also provides the basis for hypothesis testing. We will also show that these techniques are applicable to other geophysical fields as well.

3. *Probabilistic analysis:* Singular events are accumulated for seasonal and inter-annual trend anal-

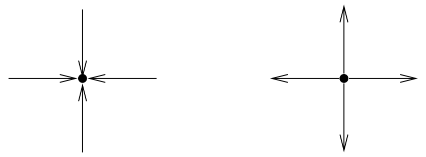


Figure 3: Critical points with non-zero divergence: (left) a sink; (right) a source.

ysis. We will show that even with simple techniques, new insights can be gained.

4. *Data fusion, mining, and forecasting:* This is an on-going work with limited results at the present point.

3.1 Computation of Feature Velocities

Let $I(x, y, t)$ be the image at time t, with (u, v) as the velocity vector at each point. The flow field equation with brightness constancy assumption is given by:

$$\frac{dI(x,y,t)}{dt} = I_x u + I_y v + I_t = 0 \qquad (1)$$

where the subscripts x, y, and t represent the partial derivatives. Horn and Schunck [6] constrained the problem by incorporating local smoothness in the flow. This is given by:

$$\begin{aligned}(u^*, v^*) &= \arg\min E := \int\int (I_x u + I_y v + I_t)^2 \\ &\quad +\alpha(|\nabla u|^2 + |\nabla v|^2) dxdy\end{aligned} \qquad (2)$$

where α is the weighting factor of the smoothness term, and the equation is evaluated in the spherical coordinate system. The velocity vector due to (incompressible) fluid motion has to have zero divergence at each point, i.e. $u_x + v_y = -w_z$ Under weak constraints, the above equation can be expressed as $u_x + v_y = 0$. In this context, incompressibility is enforced along the temperature-gradient (normal to iso-thermals). A vector field with zero-divergence does not contain sinks and sources. A counter-example is shown in Fig. 3.

This constraint is expressed as a penalty term in the energy functional:

$$\begin{aligned}(u^*_\beta, v^*_\beta) &= \arg\min E := \tfrac{1}{2}\int\int [(I_x u + I_y v + I_t)^2 \\ &+ \alpha(|\nabla u|^2 + |\nabla v|^2) + \beta(u_x + v_y)^2] dxdy\end{aligned} \qquad (3)$$

If $\beta \to +\infty$, then (u^*_β, v^*_β) converges to the solution of Eq.2 subjected to $u_x + v_y = 0$.

The Euler-Lagrange equations of (3) are

$$\begin{cases} -I_x(I_x u + I_y v + I_t) + (\alpha+\beta)u_{xx} + \alpha u_{yy} + \beta v_{xy} = 0 \\ -I_y(I_x u + I_y v + I_t) + \alpha v_{xx} + (\alpha+\beta)v_{yy} + \beta u_{xy} = 0 \end{cases} \qquad (4)$$

The above PDE are solved with finite difference methods. Recall that 2nd-order derivatives of u's (and similarly v's) are given by [8]:

$$\begin{aligned} u_{xx}|_{j,i} &= \tfrac{u_{j+h,i} + u_{j-h,i} - 2u_{j,i}}{h^2} \\ u_{yy}|_{j,i} &= \tfrac{u_{j,i+h} + u_{j,i-h} - 2u_{j,i}}{h^2} \\ u_{xy}|_{j,i} &= \tfrac{u_{j+h,i+h} + u_{j-h,i-h} - u_{j+h,i-h} - u_{j-h,i+h}}{4h^2} \end{aligned} \qquad (5)$$

Substituting them into Eq. (4), we have

$$\begin{cases} (I_x^2 + \frac{4\alpha+2\beta}{h^2})u_{j,i} + I_x I_y v_{j,i} = \Delta_1 \\ I_x I_y u_{j,i} + (I_y^2 + \frac{4\alpha+2\beta}{h^2})v_{j,i} = \Delta_2 \end{cases} \qquad (6)$$

where

$$\begin{aligned} \Delta_1 &= -I_x I_t + \tfrac{\alpha+\beta}{h^2}(u_{j+h,i} + u_{j-h,i}) + \tfrac{\alpha}{h^2}(u_{j,i+h} + \\ & u_{j,i-h}) + \tfrac{\beta}{4h^2}(v_{j+h,i+h} + v_{j-h,i-h} - v_{j+h,i-h} - v_{j-h,i+h}) \\ \Delta_2 &= -I_y I_t + \tfrac{\alpha}{h^2}(v_{j+h,i} + v_{j-h,i}) + \tfrac{\alpha+\beta}{h^2}(v_{j,i+h} + \\ & v_{j,i-h}) + \tfrac{\frac{\beta}{h^2}}{4}(u_{j+h,i+h} + u_{j-h,i-h} - u_{j+h,i-h} - u_{j-h,i+h}) \end{aligned}$$

$u_{j,i}$ and $v_{j,i}$ can be represented by Δ_1 and Δ_2:

$$\begin{cases} u_{j,i} = \frac{1}{D}[(I_y^2 + \frac{4\alpha+2\beta}{h^2})\Delta_1 - IxIy\Delta_2] \\ v_{j,i} = \frac{1}{D}[-IxIy\Delta_1 + (I_x^2 + \frac{4\alpha+2\beta}{h^2})\Delta_2] \end{cases} \qquad (7)$$

where $D = \frac{4\alpha+2\beta}{h^2}(Ix^2 + I_y^2) + (\frac{4\alpha+2\beta}{h^2})^2$. Then an iterative method can be described as

$$\begin{cases} u_{j,i}^{(n+1)} = \frac{1}{D}[(I_y^2 + \frac{4\alpha+2\beta}{h^2})\Delta_1^{(n)} - IxIy\Delta_2^{(n)}] \\ v_{j,i}^{(n+1)} = \frac{1}{D}[-IxIy\Delta_1^{(n)} + (I_x^2 + \frac{4\alpha+2\beta}{h^2})\Delta_2^{(n)}] \end{cases} \qquad (8)$$

where $(u_{j,i}^{(n)}, v_{j,i}^{(n)})$ is the velocity field at the n-th step. The process converges when (8) approaches (6), which means that an optimal solution has been found. The convergence of the iteration defined by (8) is guaranteed, but due to space limitation, the proof is not given here. Note that if $\beta = 0$ then this formulation is identical to the original Horn-Schunck model.

3.1.1 Multigrid Approach

The solution to the above equations is expressed through a multigrid technique to reduce the computational complexity. Let h be the window size where finite differences are computed. Then by setting $h = 2^K, 2^{K-1}, \cdots, 1$ successively, we can propagate from coarse to fine grid through simple linear interpolation and refinement. Linear interpolation is used because motion is locally smooth. It simply sets an initial condition to refine the solution at the next level. This multigrid method can be described as follows:

- STEP 0. $h = 2^K$
- STEP 1. Repeat the following algorithm until it converges:

for $(i = 0; i < M; i = i + h)$

for $(j = 0; j < N; j = j + h)$ let

$$\begin{cases} u_{j,i} = \frac{1}{D}[(I_y^2 + \frac{4\alpha+2\beta}{h^2})\Delta_1^{(n)} - IxIy\Delta_2^{(n)}] \\ v_{j,i} = \frac{1}{D}[-IxIy\Delta_1^{(n)} + (I_x^2 + \frac{4\alpha+2\beta}{h^2})\Delta_2^{(n)}] \end{cases}$$

where $M \times N$ is the image size.

- STEP 2. Linear interpolation.

for $(i = 0; i < M; i++)\{$

$i_0 = [i/h] \cdot h, r_i = (i - i_0)/h;$

for $(j = 0; j < N; j++)$ $\{$

$j_0 = [j/h] \cdot h, r_j = (j - j_0)/h;$

$$\begin{aligned} u_{j,i} = &\ (1 - r_i)(1 - r_j)u_{j_0,i_0} + r_i(1 - r_j)u_{j_0,i_0+h} + \\ &\ (1 - r_i)r_j u_{j_0+h,i_0} + r_i r_j u_{j_0+h,i_0+h} \\ v_{j,i} = &\ (1 - r_i)(1 - r_j)v_{j_0,i_0} + r_i(1 - r_j)v_{j_0,i_0+h} + \\ &\ (1 - r_i)r_j v_{j_0+h,i_0} + r_i r_j v_{j_0+h,i_0+h} \end{aligned}$$

$\}$ $\}$

- STEP 3. if $h = 1$, stop; else $h = h/2$, goto STEP 1.

Note that in STEP 1, all the pixels are updated asynchronously and not simultaneously. Why is the asynchronous iteration faster than the synchronous? An intuitive explanation is that the pixels that have been updated will be used to update their neighborhood. Even though this idea is rather simple, it can speed up the convergence greatly. In fact, it is a deterministic version of the classic Gibbs sampler [5].

3.2 Evaluation of Feature Velocities

Both α and β of the regularization parameters are important for accurate measurement of feature velocities. We have compared our measurements with ground truth data (obtained from surface drifters) and verified that our measurements are consistent. For example, from Nov. 5, 1986 to Nov 18, 1997, Aoki et al. [2] indicated a motion field of 3-5cm/sec to the west from 35° to 40° N in the region 30° to 40°N and 140° E to 170° W. Our measurements in the same area indicates a motion of 3.7 cm/sec.

3.3 Detection of Vortices and Saddle Points

Singularities in the flow field can provide a compact abstraction in the velocity field. This issue has been addressed in literature [10], where an algorithm based on the analysis of local Jacobian was proposed. Here, we propose an alternative method for robust localization of events and their subsequent tracking.

Let $F = (u, v)$ be a vector field and J be a Jordan curve with no critical point on it. The index of J is defined by

$$\text{Index}(J) = \frac{1}{2\pi} \oint_J \frac{udv - vdu}{u^2 + v^2}$$

Because it is not easy to localize the critical points, we compute the index over the entire field. At each point P, we choose a small circle J_P around P and compute Index(J_P). The flow field (u, v) can then be classified according to:

1. The index of a vortex is equal to +1 (the classification of singular points in a vector field is given in [10]), and
2. The index of a saddle point is equal to −1.

There is no node in the vector field because of the zero-divergence constraint, but singularities do occur. Recall that a point (x, y) is singular iff

$$\begin{cases} u(x, y) = 0 \\ v(x, y) = 0 \end{cases} \tag{9}$$

However, using the above equation to localize singularities leads to computational instability. A better approach is to exploit the inherent local minimum velocity volume to simplify the problem. Thus,

- Step 1: Find all local minima of the velocity field:

  $S = \{$ all the local minima of $\sqrt{u(x, y)^2 + v(x, y)^2}\}$

- Step 2: $\forall(x, y) \in S$:

  (1) Let $R = \frac{1}{2}\max\{r|S \bigcap J^r_{(x,y)} = \phi\}$ where $J^r_{(x,y)}$ is the circle centered at (x, y) with radius r.
  (2) Compute the index of $J^R_{(x,y)}$, that is,

  $$\text{Index}(J^R_{(x,y)}) = \frac{1}{2\pi} \oint_{J^R_{(x,y)}} \frac{udv - vdu}{u^2 + v^2}$$

  (3) If Index($J^R_{(x,y)}$) = 0, then (x, y) is not singular; if Index($J^R_{(x,y)}$) = 1, then (x, y) is a vortex; if Index($J^R_{(x,y)}$) = −1, then (x, y) is a saddle point.

One important characteristic of vortex is its "size" and a simple solution is proposed and implemented. If a point a (x, y) is a vortex, then its size $R^*(x, y)$ can be defined as:

$$R^*(x, y) = \max\{R|\text{Index}(J^R_{(x,y)}) = 1\}$$

that is to say, $R^*(x, y)$ is the largest R such that the index of $J^R_{(x,y)}$ continues to be 1.

Fig. 4 shows the feature velocities (date: the 200th day of the year 1992) corresponding to a pair of consecutive images. Each vortex is marked by red blocks. A circle drawn around each vortex indicates its size. The direction of feature velocities are shown with yellow arrows. The magnitude of these velocity vectors are shown with the underlying intensity distribution.

4 Feature based climatic studies

We have applied the above technique to 12 years of SST data at 18 Km resolution. The data are dense with temporal resolution that is 2 days apart. A database is then created for seasonal and inter-annual studies. Each singular event is represented with a number of attributes, which can be classified into the following categories:

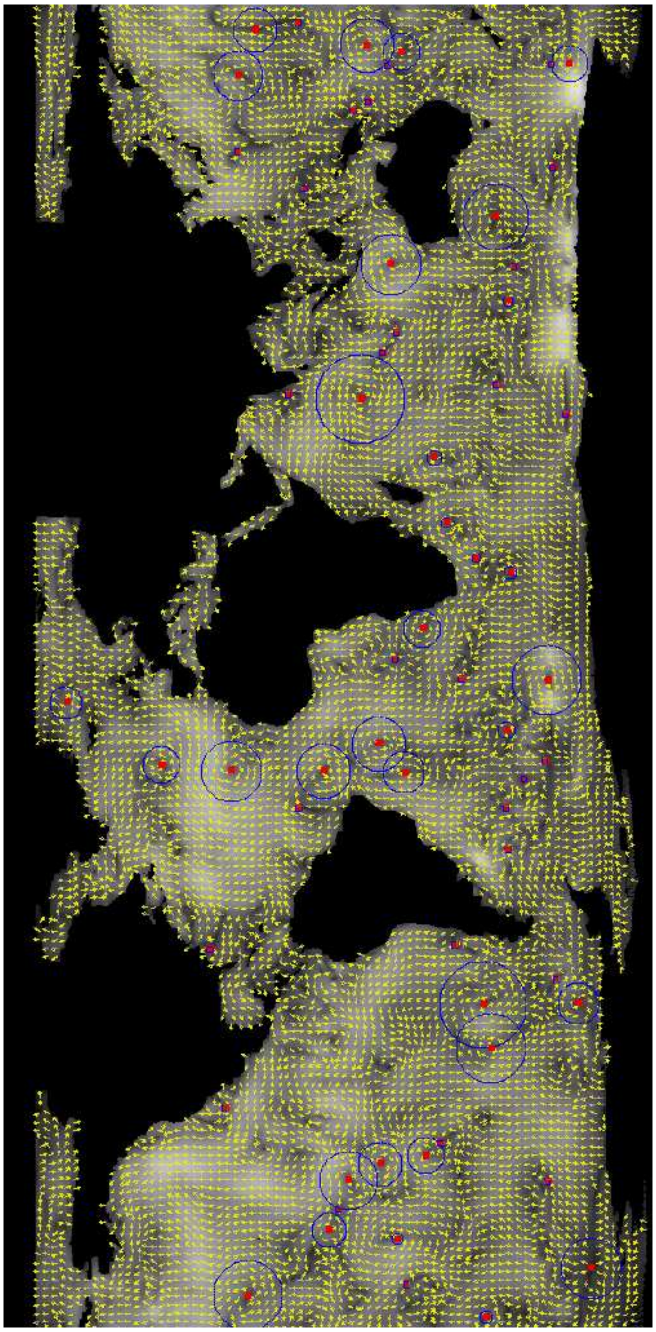


Figure 4: Feature velocities and computed vortices: Feature directions are shown with yellow arrows, and their corresponding magnitudes are shown with the underlying brightness map. Vortices are marked with a red block, and their size is encoded with a blue circle. This result is generated from a pair of consecutive images around day 200 from 1992.

Year	*Day 1-90*	*91-181*	*182-273*	*Day 274-365*
1986	5194	3991	5126	4425
1987	4970	3881	5382	4455
1988	5119	3991	5241	4854
1989	5051	4303	5060	4230
1990	4933	4183	5298	4470
1991	5166	4133	5542	4870
1992	5254	4273	5446	4617
1993	4988	4043	5247	4171
1994	5057	3942	5071	3380
1995	5109	4212	5089	4419
1996	5283	4090	5210	4847
1997	5422	4280	5278	4519
1998	5095	4074	5133	4534

Table 1: Computed number of events (vortices and saddle points) in each season and each year of SST data.

1. *Temporal attributes*: that include year, season, and date.
2. *Singularity types*: that include event types, e.g. vortex or saddle point.
3. *Spatial attributes*: that include location, size of vortices, their direction, and speed.
4. *Physical attributes at singular locations*: that include temperature, wind stress, and precipitation. Often these measurements (wind stress and precipitation) are not available at the same spatial and temporal resolution. Therefore, at a specific location, these quantities are interpolated from neighboring values.

4.1 Trend analysis in feature database

The proposed database allows both seasonal and inter-annual trend analysis. Furthermore, dynamic behavior of vortices can be tracked for their preferred birth location and duration. Table 1 shows the number of events (vortices and saddle points) for each year and each season. This table indicates (almost) equal number of vortices and saddle points in each year. Furthermore, the number of events in each season (in both northern and southern hemisphere) is highly correlated. The number of events is higher in the first 90 days than the second 90 days. It rises again during the third 90 days and then drops during the fourth 90 days. This seasonal periodic effect has not been observed in the past and it is currently being investigated.

The long term global climatic study is performed by constructing a probability density distribution for seasonal and inter-annual periods. The density distribution is simply the probability of occurrences of vortices and saddle points at a specific spatial location. An immediate conclusion is that PDF is not uniform and that events have a preferred localization.

One particular observation indicates preferred inter-annual localization of events in the sub-polar region and Gulf stream.

5 Conclusions

In this paper, we formulated flow computation in oceanographic images as a constrained variational problem with incompressibility constraint and developed a multigrid approach for its efficient implementation. We then computed events of interest, e.g. vortices and saddle points, from regularized vector fields. The technique was applied to 12 years of high resolution data (from AVHRR spacecraft) where we showed that (a) the number of events in each season is highly correlated, and (b) events have a preferred localization in 40° South and and Gulf stream. This has been the first global climatic study of the events in SST data, which has been made possible by utilizing techniques that are evolved from the field of computer vision.

Acknowledgments: The authors deeply appreciate Prof. Arthur Mariano of the University of Miami for providing data and valuable discussions. We also thank Mr. Wes Bethel for his help in data visualization.

References

[1] A.A. Amini. A scalar function formulation for optical flow: Application to x-ray imaging. In *IEEE Workshop on BioMedical Imaging*, pages A:117–124, 1994.

[2] S. Aoki, S. Imawaki, and K. Ichikawa. Baroclinic disturbances propagating westward in the kuroshio extension region as seen by a satellite altimeter and radiometers. *Geophysical Research-Oceans*, 100:839–855, 1995.

[3] J.L. Barron, D.J. Fleet, and S.S. Beauchemin. Performance of optical flow techniques. In *CVPR92*, pages 236–242, 1992.

[4] I. Cohen and I. Herlin. A motion computation and interpretation framework for oceanographic satellite images. In *SCV95*, pages 13–18, 1995.

[5] S. Geman and D. Geman. Stochastic relaxation, gibbs distribution and bayesian restoration of images. *PAMI*, 6:721–741, 1984.

[6] B. K. P. Horn and B. G. Schunck. Determining optical flow. *Artificial Intelligence*, 17:185–203, 1981.

[7] S.H. Lai and B.C. Vemuri. Robust and efficient algorithms for optical flow computation. In *SCV95*, pages 455–460, 1995.

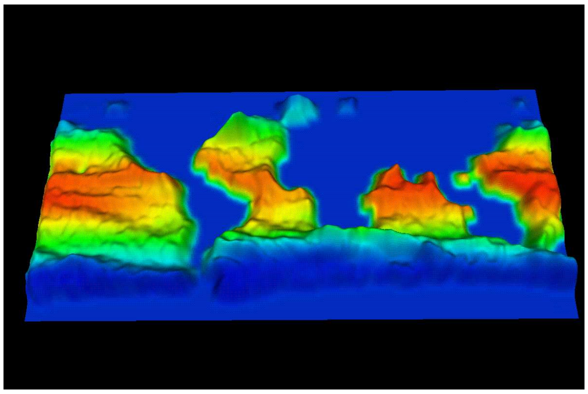
[8] R.J. LeVeque. Finite difference methods for differential equations. *Math Note 585-6*, pages 225–270, 1998.

[9] M. Maurizot, P. Bouthemy, and B. Delyon. 2d fluid motion analysis from a single image. In *CVPR'98*, pages 184–189, Santa Barbara, California, 1998.

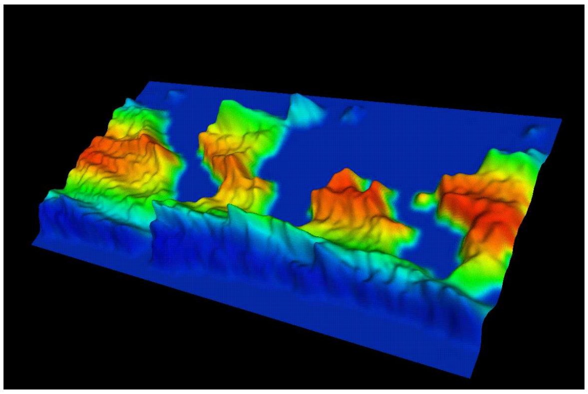
[10] A. R. Rao and R. C. Jain. Computerized flow field analysis: Oriented texture fields. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pages 693–709, 1992.

[11] Y.L. Xiong and S.A. Shafer. Dense structure from a dense optical flow sequence. In *Intl. Symposium on Computer Vision*, pages 1–6, Coral Gables, Florida, 1995.

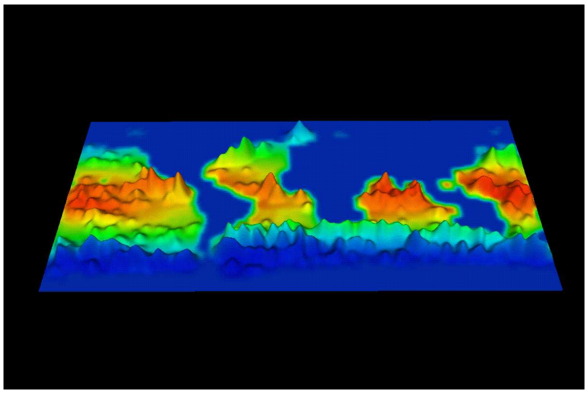
[12] L. Zhou, C. Kambhamettu, and D. Goldgof. Extracting nonrigid motion and 3d structure of hurricanes from satellite image sequences without correspondences. In *CVPR'99*, pages 281–285, Fort Collins, Colorado, 1999.



(a)



(b)



(c)

Figure 5: Frequency of occurrence of vortices and saddle points, shown at 1° resolution, from 1986 to 1998: (a,b) Two views of all events; (c) local maxima of PDF over 12 years.