RANKED REPRESENTATION OF VECTOR FIELDS

Balakrishna Nakshatrala
Yodlee.com
balki@yodlee.com

David Thompson
Center for Computational Systems at the Engineering Research Center
Mississippi State University
dst@erc.msstate.edu

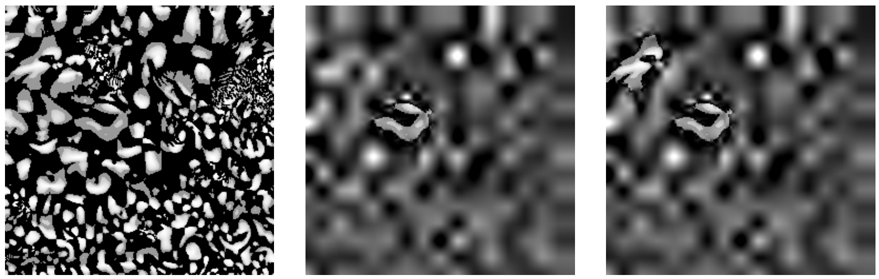
Raghu Machiraju
Department of Computer and Information Science
The Ohio State University
raghu@cis.ohio-state.edu

Abstract Browsing and visualizing large datasets is often a tedious chore. Locating features, especially in a wavelet transform domain is usually offered as a possible solution. Wavelet transforms decorrelate data and facilitate progressive access through streaming. The work reported here describes a scheme that allows the user to first visualize regions containing significant features. Various region and coefficient ranking strategies can be incorporated into this approach so that a progressively encoded bitstream can be constructed. We examine four wavelet ranking schemes and demonstrate the usefulness of the feature-based schemes for a 2D oceanographic dataset.

Keywords: Terascale visualization, vector fields, ranked representation, wavelets.

1. Introduction

In terascale visualization, locating important features in the data is one of the keys to effective data exploration. Also, for practical reasons arising from resource limitations, only parts of the dataset can be



1a. 1b. 1c.

Figure 1. 1a. Features in the original image, 1b. Feature with highest priority sent first, and 1c. Feature with next highest priority sent second.

accessed at any time. The primary motivation for this effort has its genesis in embedded visualization systems that facilitate ranked access to relevant features in a dataset as described in Machiraju et al., 2001. The operation and utility of such a system is illustrated in Figure 1. Figure 1a shows the features in a two-dimensional field. Two features are automatically selected to receive a higher priority based on a user defined criteria. Figures 1b and 1c show the reconstructed image at various stages of a progressive transmission. Initially, the background is transmitted. Then, according to a feature-based priority schedule, information is transmitted one feature at a time. Features appear according to the priority schedule and are incrementally refined over time.

The effort described herein is a summary of the work detailed in Nakshatrala, 1999 and examines various ranking strategies for regions and wavelet coefficients with special emphasis on feature based ranking strategies. A block-diagram depicting different components of the current effort along with the sections in this paper that discuss these components is shown in Figure 2. The lifting scheme is first applied to obtain the wavelet transform of the data. Vector data is treated by applying the transform to each component. Point-based feature detection methods are then used to detect features in the data at multiple scales. Techniques from multi-grid solution algorithms are used to improve gradient estimations at different scales. Segmentation of the resulting scalar field at each resolution produces a multi-scale significance map. User-specified criteria (including scale-space persistence) are used to rank the ROIs located in the significance map. Various techniques can then be used to rank the wavelet coefficients and may include feature-based methods that use the multi-scale significance map generated earlier. Finally, Section 6 describes results for a limited two-dimensional oceanographic dataset representing ocean currents in the equatorial region of

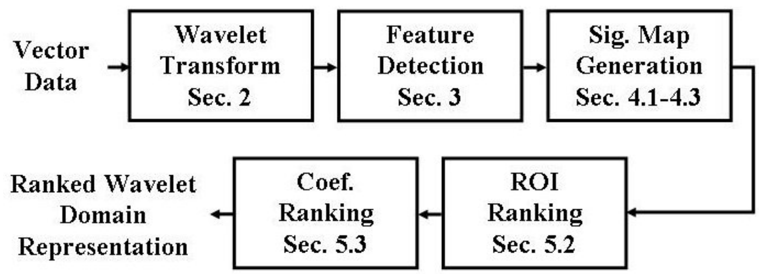


Figure 2. Generating a wavelet domain representation of a vector field

the Pacific Ocean. The dataset was generated by the Naval Layered Ocean Model (NLOM) simulation program described in Wallcraft, 1999.

1.1 Related Work

Previous efforts for compressing computational datasets include predictive methods, fractal methods, vector quantization, discrete cosine transforms, and wavelets. A survey is included in Machiraju et al., 1998. Feature detection is an important component of the proposed system. In Machiraju et al., 1998, scale coherent features are detected and used to guide the ranking of the wavelet coefficients. Scalar fields such as pressure or density of a flowing media can be employed to detect certain types of features, e.g. Marcum and Gaither, 1997. Other features, such as vortices are characterized by changes in direction of a vector field. Banks and Singer, 1996 developed a vortex detection technique that exploits the kinematic and dynamic properties of a vortical flow. This approach can be contrasted with those methods based purely on the kinematic properties of the velocity field such as Helman and Hesselink, 1989, which determines critical point locations and attempts to connect them, and Sujudi and Haimes, 1995, which attempts to locate vortex core regions using the velocity gradient tensor.

2. Linear Lifting Scheme for Vectors

An intentionally simple extension of the linear-lifting scheme to perform a wavelet transform for vector data is discussed in this section. In general, the lifting scheme consists of three steps: split, predict, and update (see Sweldens, 1997 for further details). For linear lifting, the filter coefficients employed during the predict stage are $g = \{0.5, 1.0, 0.5\}$, while the coefficients for the update stage are $h = \{0.25, 0.25\}$. We use a critically sampled (i.e., decimated) transform and achieve a multidimensional transform by applying a sequence of one-dimensional wavelet transforms along each dimension in succession. For example, for a two-dimensional dataset, one-dimensional transforms are applied first along rows and then along columns. Each transform level yields four subbands corresponding to a smooth decimated representation and details along horizontal, vertical, and diagonal orientations. For vector-valued data, the wavelet transform is applied to each component of the vector in an independent fashion.

3. Detecting Regions of Swirling Flow

Berdahl and Thompson, 1993 define a derived scalar quantity called *swirl* which can be used to identify features such as vortices. The method is based on the observation that a sufficient condition for the existence of swirling motion is that the eigenvalues of the velocity gradient tensor must contain a complex conjugate pair. In this procedure, a scalar value based on the local velocity and velocity gradients is assigned to each field point. The swirl value is interpreted as the tendency for the fluid to swirl at a given point. Contiguous regions of nonzero swirl values can therefore be thought of as distinct features. Core regions of vortices are characterized by larger swirl values.

The swirl parameter τ is defined as the ratio of the time for a fluid particle to convect through the region of complex eigenvalues to the orbit time

$$\tau = \frac{t_{conv}}{t_{orbit}} = \frac{|Im\left(\lambda_{1,2}\right)|}{2\pi V_{conv}} \tag{1}$$

where $Im\left(\lambda_{1,2}\right)$ is the imaginary part of the complex conjugate pair of eigenvalues and V_{conv} is a suitable convection velocity. For small values of τ, the fluid convects too rapidly through the region of complex eigenvalues to be captured in the swirling motion. In regions of large τ, the fluid is trapped in a swirling motion. It should be noted that the swirl values used in this work are calculated as the logarithm of the result given by Equation 1.

4. Multi-Scale Significance Map Generation

In this section, we discuss the generation of a multi-scale significance map. A multi-scale significance map is obtained from multiple segmented single resolution maps. In essence, a feature pyramid is derived from this effort, marking all feature-rich spatial regions. The multi-scale map, although more expensive than a single map, allows for better discrimination of the feature preserving properties of individual wavelet coefficients. We first focus on the need for a multi-scale map. We then describe the techniques needed to generate feature-centric significance maps.

4.1 Multi-Scale Significance Map

Consider a single-scale significance map denoted by

$$S_{single} = \{s\,(i,j)\,|i,j = 0,\ldots,N-1\} \quad . \tag{2}$$

The quantity S is a scalar that indicates the presence of the desired features. This approach is valid for any feature whose presence and relative strength can be deduced using a scalar field. For flows with vortices, the swirl parameter, as defined in Equation 1, can be used as the representative scalar. The significance map delineates ROIs in the vector field of size $N \times N$ at a single resolution, usually the finest.

For large datasets, a multi-scale transform, such as a wavelet transform, is conducted to obtain a sparser representation of data by only considering coefficients which contribute to a meaningful version of the reduced data. This reduced set of coefficients can be determined in several ways. Often, coefficients with the largest magnitude are selected. We propose to select those that contribute to features at all scales. The wavelet-coefficient pyramid obtained from the L-level wavelet transform of an $N \times N$-sized vector field (with $N = 2^{L-1}$) using the lifting scheme

$$W_L = \left\{w_k\,(i,j)\,|k = 0,\ldots,L-1,\;\; i,j = 0,\ldots,2^k - 1\right\} \tag{3}$$

is arranged at multiple resolutions. Therefore, the single resolution significance map S_{single} cannot be used directly to rank the wavelet coefficients $w_k\,(i,j)$. Additionally, to implement progressive access as we described earlier, ROIs at all scales need to be ranked. We achieve this by measuring the feature strength of a ROI and its persistence across multiple scales. These attributes can only be measured if significance maps at all scales are available. This section focuses on the generation of a multi-scale significance map,

$$S_L = \left\{s_k\,(i,j)\,|k = 0,\ldots,L-1,\;\; i,j = 0,\ldots,2^k - 1\right\} \tag{4}$$

that can be used directly on the multi-scale wavelet-coefficient mask W_L to determine the significance of the wavelet coefficients in terms of their contribution to feature preservation, thereby allowing ranked access to features.

4.2 Approximate Reconstruction of Features

To generate a multi-scale significance map, it is necessary to perform a feature detection at each scale. The swirl-detection algorithm described previously requires the computation of the velocity as well as the velocity gradient at each point in the domain. We use the technique described in Machiraju et al., 2001 to approximately reconstruct the velocity and the velocity gradient using only data in the wavelet domain. This technique corrects for the effects of the linear-lifting scheme as well as truncation error and is based on techniques used in multigrid-solution algorithms (see Brandt, 1973). We note that the following discussion is conducted in terms of one-dimensional data and operations. However, since a multidimensional wavelet transform can be constructed as multiple one-dimensional transforms applied along each dimension in sequence, we apply the operations we describe below for one dimension (we consider the x direction) in a similar manner to construct an equivalent multidimensional operator.

Let the wavelet approximation obtained after an n-level wavelet transformation of a scalar quantity $u_{k,i}$ defined on the grid $x_{k,i}$ be represented by $u_{k-n,i}$. The corresponding coarser grid is $x_{k-n,i}$ and $\Delta x_{k-n} = 2^n \Delta x_k$. Here k represents the level of the wavelet transform and i indicates the spatial position.

A correction for the potentially undesirable effects of of the wavelet transform may be derived through a Taylor series analysis of the update step of the linear lifting scheme. The resulting expression is given by

$$u_{k,2^n i} = \left(1 + \frac{1}{16}\left(\sum_{k=0}^{n-1} \frac{1}{4^k}\right)\overline{\delta} + O\left(\Delta x_{k-n}^4\right)\right) u_{k-n,i} \tag{5}$$

where the operator $\overline{\delta}$ is defined by

$$\overline{\delta} = \delta^2 - \frac{1}{16}\left(\sum_{k=0}^{n-1} \frac{1}{4^k}\right)\delta^4 \tag{6}$$

and δ^2 and δ^4 are standard second and fourth difference operators. The central-difference operators used to define the velocity gradient tensor can be corrected to account for differences in truncation error between levels k and $k-n$ using

$$\frac{\delta u_{k,2^n i}}{2\Delta x_k} = \frac{\delta u_{k-n,i}}{2\Delta x_{k-n}} - \frac{1}{8}\left(\sum_{k=0}^{n-1} \frac{1}{4^k}\right)\frac{d^3 u}{dx^3}\Delta x_{k-n}^2 + O\left(\Delta x_{k-n}^4\right) \quad . \tag{7}$$

Equation 7 specifies the correction that should be added to the central-difference approximation on level $k-n$ so that it is equivalent to the central-difference approximation on level k to the order of the approximation. A standard second-order, central-difference can be used to approximate the third-derivative term.

The resulting procedure to compute approximations to the velocity and velocity gradients using only the transformed data is as follows:

- Compute an approximate reconstruction to the velocity components at the current level using Equation 5 applied for each direction.
- Compute the gradients using these reconstructed values using Equation 7.
- Use these reconstructed values and gradient approximations to compute the swirl.

Since the above methods is point-based it is necessary to create regions. A simple segmentation procedure is applied at all scales to derive various ROIs.

4.3 Segmentation

The segmentation process involves identifying different swirl regions in the significance map and assigning each a label to facilitate later access. Here, a feature is defined as a set of connected points that have non-zero swirl and are surrounded by a zero-swirl region. A simple two-pass scan-line algorithm similar to those used in image processing and described in Jain, 1989 is used for segmentation of the significance map.

5. Wavelet Coefficient Ranking Strategies

This section describes ranking strategies that are suitable for wavelet-based compression techniques. Ranking is a two-step process. The ROIs are ranked first to allow access to regions of significant data. Then, coefficients of each ROI are ranked. To compare ranking strategies, an error analysis of the reconstructed vector field is required. This embodies the rate-distortion approach. Thus, at a given rate the smallest distortion or error should result in the best representation. We first describe the error metrics employed to quantify these results. We then describe the ROI ranking scheme and four different wavelet coefficient ranking schemes.

5.1 Error Metrics

The most commonly used error metric for scalar data is the Mean Square Error (MSE). Let the original and reconstructed signals for a scalar field ω be represented by

$$\begin{aligned} \omega_{orig} &= \{\omega\,(i,j)\,|i,j = 0,\ldots,N-1\} \\ \omega_{recon} &= \{\tilde{\omega}\,(i,j)\,|i,j = 0,\ldots,N-1\}. \end{aligned} \tag{8}$$

The MSE of the difference between the two fields is given by

$$MSE\left(\omega_{orig}, \omega_{recon}\right) = \frac{1}{N \times N} \sum_{j=0}^{N-1} \sum_{i=0}^{N-1} \left(\delta\omega_{i,j}\right)^2 \tag{9}$$

where

$$\delta\omega_{i,j} = \tilde{\omega}_{i,j} - \omega_{i,j} \tag{10}$$

measures the difference between the two fields.

In the case of a vector field, there are errors in direction as well as magnitude. A simple error measure for the magnitude error can be defined using Equation 9 with the magnitude of the difference of the reconstructed and original velocity vectors,

$$\delta\omega_{i,j} = |\tilde{v}_{i,j} - v_{i,j}| \tag{11}$$

$\tilde{v}_{i,j}$ and $v_{i,j}$, respectively. Similarly, a metric for measuring the direction error can be based on the angle between the original and the reconstructed vector using

$$\delta\omega_{i,j} = \cos^{-1}\left(\frac{\tilde{v}_{i,j} \cdot v_{i,j}}{|\tilde{v}_{i,j}||v_{i,j}|}\right). \tag{12}$$

An error metric based on the swirl parameter (see Equation 1) can also be defined by directly substituting the swirl field $\tau_{i,j}$ and $\tilde{\tau}_{i,j}$ into Equations 9 and 10.

5.2 Ranking ROIs

We employ a scale-coherent method to rank features. This method assigns a feature that persists across scales a higher rank than a feature that persists across a smaller number of scales. To quantify the persistence of a feature across scales, it is necessary to track a feature across scales. Simple tracking is conducted by checking for the intersection of features in one scale with the features in other scales. If the intersection area is greater than a certain threshold, it is likely that they are the same features at different scales because they exist at the same spatial location. The algorithm also allows for the possibility that features are created or destroyed. We should note that all efforts to describe the evolution of features are, at best, empirical in nature. However, the approach employed here appears sufficient to track features in oceanographic data. Details of the algorithm are given in Nakshatrala, 1999. At the end of the tracking process, features that exist from coarsest scale onwards get a higher rank compared to those that appeared only at finer scales.

5.3 Prioritizing the Wavelet Coefficients

We now use the significance map to prioritize the wavelet coefficients $w_k(i, j)$ from each ROI. The priority assignment strategies considered here fall into two categories: those that aim at minimizing the overall MSE without using any feature information and are based on the magnitude of the wavelet coefficients and those that use application specific information about features in the dataset. These methods use the feature-based parameter, swirl, to rank or assign priority to the coefficients from the wavelet pyramid.

- Ranking Based on Magnitude - Scheme 1: The most direct method is to order the coefficients by their magnitude (in decreasing order). This ranking scheme does not use any knowledge of the features.
- Ranking Based on Scale and Magnitude - Scheme 2: With Ranking Scheme 1, a problem arises because it is based solely on the magnitude of the wavelet coefficients and does not accord any importance to scale. A wavelet coefficient at finer scale represents higher frequencies and has a small region of influence whereas a coefficient at lower scale represents lower frequencies and has a broader region of influence. Hence, Scheme 2 was developed to take into consideration both scale and magnitude in ranking the coefficients.
- Ranking Based on Scale and Swirl - Scheme 3: With a slight modification to Ranking Scheme 2, a new feature-based ranking can be developed. This new ranking uses the knowledge of features to rank the wavelet coefficients. The idea is to first order the coefficients by their scale and at each scale, rank the coefficients in the regions of interest higher than other coefficients. This is achieved using the multi-scale significance map described in Section 4.
- Ranking Based on Feature and Scale - Scheme 4: The final ranking scheme provides ranked access to features in a dataset, which is the objective of this work. This scheme is similar to Scheme 3 except that instead of progressively embedding all the features simultaneously, they are sent sequentially. Thus, the coefficient stream consists of features arranged in a ranked fashion. The coefficients corresponding to each feature are ranked by scale and then at each scale, ranked by their significance value. This scheme uses additional information to rank features obtained using the algorithms described in Section 4, along with the significance map showing the regions of interest.

6. Results

The four ranking schemes are compared using the error metrics based on MSE of magnitude, direction, and swirl. The results are tabulated in Table 1, Table 2, and Table 3, respectively. The results for all metrics indicate that the smallest error occurs for Ranking Scheme 2. This can be explained as follows. In general, the significance of a wavelet coefficient in terms of reconstruction error decreases from the coarsest to the finest scales because the lower frequencies are more important than the high frequencies. In essence, large-scale features are defined by the lower frequencies. Therefore, the ranking based on scale-and-magnitude is the best of the four schemes considered here for applications desiring the smallest overall distortion.

On the other hand, from a feature preservation perspective, ranking Schemes 3 and 4 are more promising. Scheme 3 does well in preserving the core of the swirl regions even at very low data rates indicating its usefulness in visualization for fluid flow applications. With reference to this dataset with densely packed features, the performance of Ranking Scheme 4 is found to be very close to that of Ranking Scheme 3. It is noteworthy, however, that Scheme 4 out-performs Scheme 3 at low bit rates in terms of feature preservation. Figure 3 shows the swirl field computed from a reconstructed velocity field using the ranking scheme based on feature and scale (Ranking Scheme 4) for different data rates. It is evident that schemes of this type can produce visually acceptable images even for low data rate reconstructions ($\leq$10%).

7. Conclusions

This paper explored feature-based wavelet representations of vector fields arising from flow simulations. The desired end-result was a progressively embedded wavelet coefficient stream that facilitated visualization of significant features first. A macroscopic derived feature, namely swirl, was used to locate features in the wavelet domain. A significance map was used to delineate the presence of features and hence create regions of interest (ROIs). Scale- and space-coherent, feature-based ranking schemes were then used to rank these regions. The wavelet coefficients in the ROIs were later assigned priority in four different ways. It was determined that ranking schemes based on feature presence out-performed those based purely on coefficient magnitude in terms of feature preservation at low bit rates. On-going and future efforts include the development of feature-preserving wavelet transforms, detection of other multi-scale region-based feature detection algorithms for features besides vortices, and robust region ranking algorithms.

Table 1. Performance of ranking schemes measured by magnitude error

% coefficients	Scheme 1	Scheme 2	Scheme 3	Scheme 4
1	8.84353	5.04709	5.32602	5.32602
5	2.32089	2.60399	2.65862	2.65499
10	1.13143	0.945083	2.16447	2.36633
50	0.0730064	0.0721421	0.366211	0.369339
90	0.00233092	0.002299	0.00329009	0.107419

Table 2. Performance of ranking schemes measured by direction error

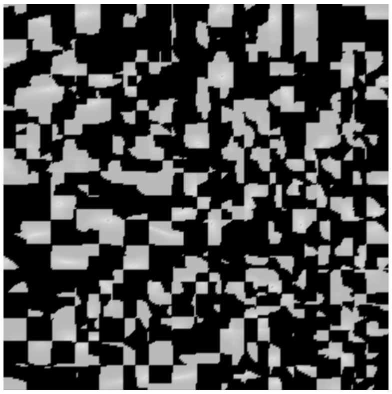
% coefficients	Scheme 1	Scheme 2	Scheme 3	Scheme 4
1	0.655156	0.229031	0.240577	0.240577
5	0.152475	0.0785144	0.0825083	0.0821206
10	0.0727711	0.0435664	0.0600747	0.0666317
50	0.00618106	0.00561497	0.011194	0.0110675
90	0.0011587	0.00115214	0.00124038	0.00345433

Table 3. Performance of ranking schemes measured by swirl error

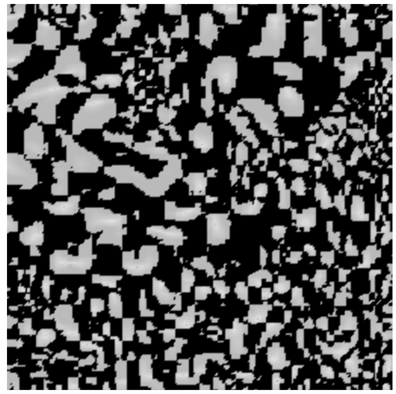
% coefficients	Scheme 1	Scheme 2	Scheme 3	Scheme 4
1	0.00356073	0.00343956	0.00343946	0.00343946
5	0.00294123	0.00251154	0.00258059	0.00258031
10	0.00240242	0.00212799	0.00233285	0.00232595
50	0.000860466	0.000802647	0.00120138	0.00111869
90	0.000355893	0.000353267	0.000376555	0.000588404

8. Acknowledgements

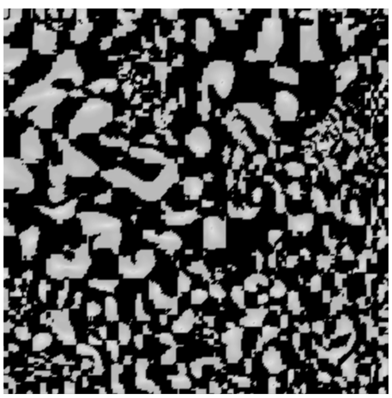
This work was supported, in part, by an NSF CAREER Award (ACI-9734483), an NSF Large Data and Scientific Software Visualization Program Grant (ACI-9982344), and the Mississippi State University Engineering Research Center (MSU-ERC). Thanks also go to Profs. Bharat Soni and James Fowler (MSU-ERC) and Dr. William Schroeder (KitWare, Inc. and Rensselaer Polytechnic Institute) for suggestions and comments.



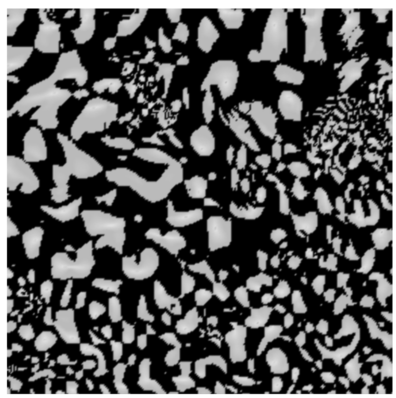
3a. 1% of coefficients



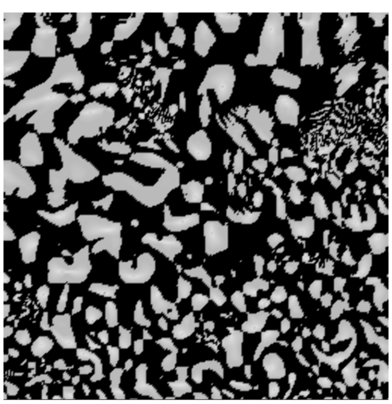
3b. 5% of coefficients



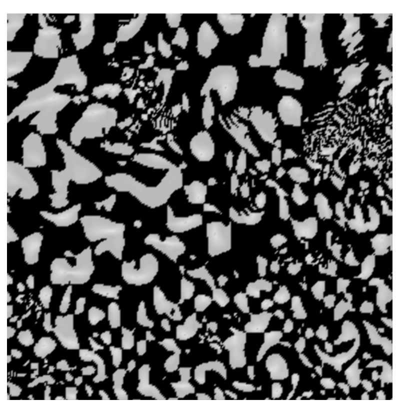
3c. 10% of coefficients



3d. 30% of coefficients



3e. 50% of coefficients



3f. 90% of coefficients

Figure 3. Swirl field computed from velocity field reconstructed at different data rates using Ranking Scheme 4

References

Banks, D. C. and Singer, B. A. (1996). Vortex Tubes in Turbulent Flows: Identification, Representation, Reconstruction. In *Proc. Visualization '96*, pages 132–139.

Berdahl, C. H. and Thompson, D. S. (1993). Eduction of Swirling Structure using the Velocity Gradient Tensor. *AIAA J.*, 31(1):97–103.

Brandt, A. (1973). Multi-Level Adaptive Technique (MLAT) for Fast Numerical Solution to Boundary Value Problems. In *Proc. Third Int. Conf. Num. Methods. Fluid Mech.*, pages 82–89. Springer-Verlag.

Helman, J. and Hesselink, L. (1989). Representation and Display of Vector Field Topology in Fluid Flow Datasets. *IEEE Computer*, 22:27–36.

Jain, A. K. (1989). *Fundamentals of Digital Image Processing*. Prentice Hall.

Machiraju, R., Fowler, J., Thompson, D., Schroeder, W., and Soni, B. K. (2001). EVITA - Efficient Visualization and Interrogation of Tera-Scale Data. In *Data Mining for Scientific and Engineering Applications*, pages 257–279. Kluwer Academic Publishers.

Machiraju, R., Zhu, Z., Fry, B., and Moorhead, R. (1998). Structure Significant Representation of Computational Field Simulation Datasets. *IEEE Trans. Vis. Graphics*, 4(2):117–129.

Marcum, D. L. and Gaither, K. P. (1997). Solution Adaptive Unstructured Grid Generation using Pseudo-Pattern Recognition Techniques. *AIAA Paper 97-1860, AIAA 13th Computational Fluid Dynamics Conference*, Snowmass Village, CO.

Nakshatrala, B. (1999). Feature-Based Embedded Representation of Vector Fields. Master's thesis, Mississippi State University.

Sujudi, D. and Haimes, R. (1995). Identification of Swirling Flow in 3D Vector Fields. *AIAA Paper 95-1715, AIAA 12th Computational Fluid Dynamics Conference*, San Diego, CA.

Sweldens, W. (1997). The Lifting Scheme: A Construction of Second Generation Wavelets. *J. Math. Anal.*, 2(29):511–546.

Wallcraft, A. J. (1999). The Naval Layered Ocean Model Users Guide. Technical Report 35, Naval Oceanographic and Atmospheric Research Laboratory.