Summary of Work for Joint Research Interchanges with DARWIN Integrated Product Team 1998

Lambertus Hesselink
Stanford University
Stanford, California 94305-4035
bert@kaos.stanford.edu
Phone: 650-723-4850 Fax: 650-725-3377

March 1999

1 Introduction

The intent of Stanford University's SciVis group is to develop technologies that enabled comparative analysis and visualization techniques for simulated and experimental flow fields. These techniques would then be made available under the Joint Reasearch Interchange for potential injection into the DARWIN [1] Workspace Environment (DWE). In the past, we have focused on techniques that exploited feature based comparisons such as shock and vortex extractions. Our current research effort focuses on finding a quantitative comparison of general vector fields based on topological features. Since the method relies on topological information, grid matching and vector alignment is not needed in the comparison. This is often a problem with many data comparison techniques. In addition, since only topology based information is stored and compared for each field, there is a significant compression of information that enables large databases to be quickly searched. This report will briefly (1) describe current technologies in the area of comparison techniques, (2) will describe the theory of our new method and finally (3) summarize a few of the results.

2 Comparison Techniques

There exist a variety of comparison techniques for vector fields. These techniques basically fall into three general categories: Image, data, and feature

[1] Developmental Aeronautics Revolutionizing Wind tunnels with Intelligent systems of NASA

extraction based comparisons. In most of these cases, comparisons are made visually [1]. Image based comparisons work on the computer generated image. Often times, a numerical data set is converted into an image that simulates an experimental visualization technique (computational flow imaging). This may be easier than extracting a vector field from an image, such as Schlieren. However, visualizing a field in 3-D is quite difficult. Often times, these techniques are limited to two dimensions. In addition to side-by-side comparison of images, other techniques include image fusion, and Fourier analysis [2].

Data level comparison techniques operate directly on the raw data. An accurate comparison requires proper grid alignment which can involve problematic interpolation between two fields [3].

The last comparison category is the extraction of features. Typically features are flow specific such as vortex cores, shock surfaces, or topology. Often times there is a geometric representation of the feature and possibly a semantic representation of the system which can be compared using a pattern recognition technique [4]. This may lead to more robust comparisons. Past study in our group has focused on the geometric structure of vector fields [5]. However, this geometric structure can be visually deceiving since two vector fields may have the same underlying topological structure but are dissimilar in appearance [6]. Therefore, a quantitative measurement for comparison of vector fields is essential.

3 Description of a Vector Field

A 2-D vector field can be described as a system of two simultaneous differential equations having the following form:

$$v_x = \frac{dx}{dt} = F(x,y) \qquad (1)$$
$$v_y = \frac{dy}{dt} = G(x,y)$$

where F and G are continuous and have continuous partial derivatives in some region D.

A vector field is typically described by the number, type, and arrangement of critical points (or equilibrium points). These points are where the system is defined to be $F(x,y) = 0, G(x,y) = 0$. The number and nature of critical points will not change under continuous transformation. A critical point is said to be isolated or simple if there is an open neighborhood around it that contains no other critical points. For this report, we focus entirely on simple critical points. The global topology of the vector field is defined as the critical points and the set of their connecting streamlines. These streamlines (separatrices) divide the field into regions that are topologically equivalent to uniform flow. Hence, only the topology is needed to reconstruct the field and therefore is useful as a means of differentiating vector fields.

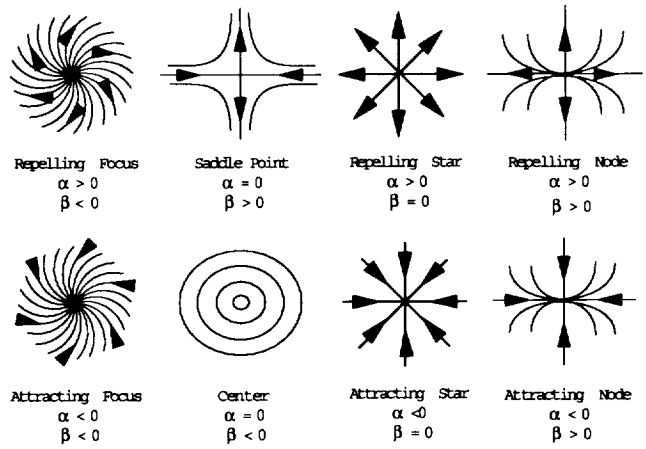


Figure 1: basic patterns for simple critical points

3.1 Classification of Simple Critical Points

The behavior of the flow about a critical point can be analyzed by investigating the streamlines in the neighborhood of the critical point. If we are sufficiently close to the critical point (say a distance dx,dy away) in most cases a first order Taylor series expansion of the velocity field is sufficient:

$$v_x(dx,dy) \approx \frac{\partial v_x}{\partial x}dx + \frac{\partial v_x}{\partial y}dy \qquad (2)$$
$$v_y(dx,dy) \approx \frac{\partial v_y}{\partial x}dx + \frac{\partial v_y}{\partial y}dy$$

Hence, the flow pattern is completely determined by the $2X2$ Jacobian matrix, $J_{ij} = \frac{\partial v_i}{\partial v_j}$ $(i,j = 1,2)$ evaluated at the critical point location. The various patterns formed in the phase-plane space can be seen by analyzing the eigenvalues of the Jacobian. The patterns are sketched in Figure 1. Notice a positive or negative real part (denoted by α) is indicative of repelling/attracting behavior. And if an eigenvalue has an imaginary part ($\beta < 0$), it indicates circulation about the point, otherwise asymptotic behavior is exhibited.

4 Vector Field Representation using Clifford Algebra

In [7] [8], Sheuermann et al. introduced Clifford algebra for vector field visualization. Clifford algebra provides a nice way to describe the relation between real and complex numbers in $2D$ space. The vector fields are defined over a

complex field in this algebra and the nonlinear vector fields are represented as multiplications of linear fields.

For the Euclidean plane we get a 4-dimensional R-algebra G_2 with the basis $1, e_1, e_2, i = e_1 e_2$ as a real vector space. Multiplication is defined as associative, bilinear and by the equations

$$\begin{aligned} 1e_j &= e_j, j = 1,2 && (3)\\ e_i e_j &= 1, j = 1,2 && (4)\\ i = e_1 e_2 &= e_2 e_1 && (5)\\ 1^2 &= 1, j = 1,2 && (6)\\ e_j^2 &= 1, j = 1,2 && (7)\\ i^2 &= -1 && (8) \end{aligned}$$

with

$$1 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \quad i = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \quad e_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \quad e_2 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$$

The usual vectors $(x, y) \in R^2$ are identified with

$$xe_1 + ye_2 \in E^2 \subset G_2 \tag{9}$$

and the complex numbers $a + bi \in C$ with

$$a1 + bi \in G_2 \tag{10}$$

4.1 Vector fields in Clifford space

A Clifford vector field is just a multivector field with values in $R^2 \subset G_2$

$$v : R^2 \to R^2 \subset G_2 \tag{11}$$

Let $z = x + iy$, $\bar{z} = x - iy$ be complex numbers in the Clifford algebra. This means

$$x = \frac{1}{2}(z + \bar{z}) \tag{12}$$

$$y = \frac{1}{2i}(z - \bar{z}) \tag{13}$$

We get

$$\begin{aligned} \vec{v}(r) &= v_1(x,y)e_1 + v_2(x,y)e_2 \\ &= v_1\left(\frac{1}{2}(z+\bar{z}), \frac{1}{2i}(z-\bar{z})\right)e_1 \\ &\quad -iv_2\left(\frac{1}{2}(z+\bar{z}), \frac{1}{2i}(z-\bar{z})\right)e_1 \\ &= E(z,\bar{z})\,e_1 \end{aligned} \tag{14}$$

Table 1: Classification of Critical Points using α β values

α	β	Type	α	β	Type
$=0$	<0	Center	—	>0	Saddle $\|\beta\| > \|\alpha\|$
>0	<0	Repelling Focus	<0	<0	Attracting Focus
>0	$=0$	Repelling Star	<0	$=0$	Attracting Star
>0	>0	Repelling Node $\|\beta\| < \|\alpha\|$	<0	>0	Attracting Node $\|\beta\| < \|\alpha\|$

Generally, a linear vector field can easily be shown as:

$$\begin{aligned} \vec{v}(r) &= E(z,\bar{z})e_1 \\ &= (az + b\bar{z} + c)e_1 \end{aligned} \tag{15}$$

where $a, b, c \subset C$.

Let $E : C^2 \to C$ be the polynomial so that $\vec{v} = E(z,\bar{z})e_1$. Let $F_k : C^2 \to C, k = 1, \ldots, n$ be the irreducible components of E, so that $E(z,\bar{z}) = \prod_{k=1}^{n} F_k$, then an arbitrary polynomial vector field with isolated critical points can be expressed as:

$$\begin{aligned} \vec{v}(r) &= E(z,\bar{z})e_1 \\ &= \prod_{k=1}^{n}(a_k z + b_k \bar{z} + c_k)e_1 \end{aligned} \tag{16}$$

where z_k is the unique zero of $a_k z + b_k \bar{z} + c_k$.

5 α-β Space and its Use as a Metric

For a linear vector field $\vec{v} = (az + b\bar{z} + c)e_1$, let $a = a_1 + a_2 i$ and $b = b_1 + b_2 i$. Eigenvalues of the Jacobian around its critical point z_0 are $\lambda_1 = b_1 + \sqrt{|a|^2 - b_2^2}$ and $\lambda_2 = b_1 - \sqrt{|a|^2 - b_2^2}$.

Let $\alpha = b_1$ and $\beta = sign(|a|^2 - b_2^2)\sqrt{||a|^2 - b_2^2|}$, criteria for basic patterns of simple critical points are:

Selection of α and β as shown in Figure 1 and delineated below can be mapped to a_1, b_1, a_2, b_2 to yield any desired field:

Notice our definition of saddle is more relaxed than shown in Figure 1. The values of α and β determine the type of critical point but it is not sufficient to be used as a metric to differentiate between two types of critical points. So we introduce a new α-β space where the 8 simple critical points are mapped onto the α,β axes at their respective (α, β) points. Vectors in this space obey all the rules defined for a regular 2-D Euclidean space. All points in this space are normalized as follows:

$$\alpha' = \frac{\alpha}{\sqrt{\alpha^2+\beta^2}} \qquad \beta' = \frac{\beta}{\sqrt{\alpha^2+\beta^2}} \tag{17}$$

It is shown in [9] that the actual values of α and β do not determine the portrait of the critical point only the ratio between them. Hence, this normalization

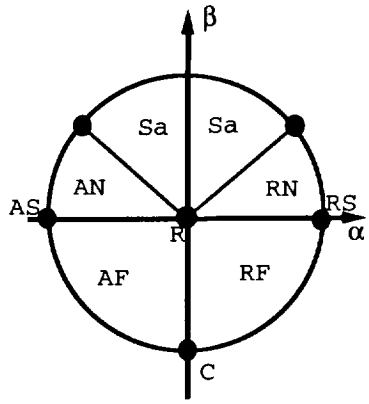


Figure 2: Basic patterns for critical points in α-β space;C for center, RN for node, AN for attracting node, RF for repelling focus, AF for attracting focus, St for star, Sa for Saddle and R for regular point

maps all points onto a unit circle (Figure 2) and thereby provides a means of relatively quantifying the difference between various points. Also note that a regular vector field with no critical points, $\vec{v} = const \cdot e_1$ has $\alpha = 0$ and $\beta = 0$ and sits at the origin of the unit circle. For the remainder of the report, α and β values will be assumed normalized.

A multiple point with a set of α's and β's corresponds to a set of points in the $\alpha - \beta$ space. For example, $\vec{v} = \bar{z}^2 e_1$ is a dipole which has two $(1, 0)$ point in $\alpha - \beta$ space; and $\vec{v} = (z - (2 + 2i)\bar{z} + c_1)(z + (2 + 2i)\bar{z} + c_2)e_1$ has one point at $(-\frac{2}{\sqrt{7}}, -\frac{\sqrt{3}}{\sqrt{7}})$ and another point at $(\frac{2}{\sqrt{7}}, \frac{\sqrt{3}}{\sqrt{7}})$ in $\alpha - \beta$ space.

6 Earth Mover's Distance

6.1 EMD analysis

The Earth Mover's Distance is first introduced in [10] [11] for content-based image retrieval in a large data base. It is used to compute the minimal amount of work that must be performed to transform one feature distribution into the other. Feature distribution in [10] [11] are the color and texture signatures of an image.

After careful study, we found that the EMD concept can be used to compute the differences between vector fields. Here, the feature distribution is redefined as the characteristics of a vector field.

Definition 1 (feature distribution) *A feature distribution for a vector field*

is the set of α and β values associated with the vector field's critical points:

$$\{(\alpha_1, \beta_1), (\alpha_2, \beta_2), \ldots, (\alpha_n, \beta_n)\}.$$

Definition 2 (Energy) *The energy for a vector field is:*

$$Energy = \sqrt{\sum_{i=1}^{n} (\alpha_i^2 + \beta_i^2)},$$

where n is the total number of critical points in this field.

This energy here is a quantity that characterizes the critical points of a vector field. It is different from the physical energy. The concept "work" is used to measure the energy differences between two vector fields or the amount of energy used to transform one vector field into the other.

Definition 3 (Work) *For two vector fields with feature distributions*

$$\{(\alpha_1, \beta_1), (\alpha_2, \beta_2), \ldots, (\alpha_n, \beta_n)\}$$

and

$$\{(\alpha'_1, \beta'_1), (\alpha'_2, \beta'_2), \ldots, (\alpha'_n, \beta'_n)\}.$$

The amount of work necessary for transforming one vector field into the other is defined as: $Work = \sqrt{\sum_{i=1}^{n} ((\alpha_i - \alpha'_i)^2 + (\beta_i - \beta'_i)^2)}$.

Intuitively, given two feature distributions, one distribution can be seen as a set of discrete point-objects with a certain amount of mass of earth spread in space, the other as a collection of holes in the same space. The work measures the least amount of energy needed to fill the holes with earth and is called the Earth Mover's Distance (EMD). Computing the EMD is based on a solution to the old *transportation problem* from linear optimization [12]. This is a bipartite network flow problem which can be formalized as the following linear programming problem: Let I be a set of suppliers, J a set of consumers, c_{ij} the cost to ship a unit of supply from $i \in I$ to $j \in J$

$$c_{ij} = \sqrt{(\alpha_i - \alpha_j)^2 + (\beta_i - \beta_j)^2}$$

and it is the same as the Euclidean distance $d_{ij} = \|\vec{v}_i - \vec{v}_j\|$ in α-β space. A critical point either exists as a whole or does not exist, it can not be split. In this case, the transportation problem has the property that the optimal flow f_{ij} can only be 0 or 1 [13]. We want to seek a set of f_{ij} that minimizes the overall cost:

$$EMD(\mathbf{x}, \mathbf{y}) = min \sum_{i \in I} \sum_{j \in J} c_{ij} f_{ij} \tag{18}$$

subject to the following constraints:

$$f_{ij} \geq 0 \quad i \in I,\ j \in J \tag{19}$$

$$\sum_{i \in I} f_{ij} = y_i, \quad j \in J \tag{20}$$

$$\sum_{j \in J} f_{ij} = x_i, \quad i \in I \tag{21}$$

$$\sum_{j \in J} y_j = \sum_{i \in I} x_i \tag{22}$$

Where x_i is the total supply of supplier i and y_j is the total capacity of consumer j. Constraint (19) allows shipping of supplies from a supplier to a consumer and not vice versa. Constraint (20) forces the consumers to fill up all of their capacities and constraint (21) limits the supply that a supplier can send as a total amount. Constraint (22) is a feasibility condition that ensures that the total demand equals the total supply, in other words, the distributions have the same overall mass and the EMD is a true metric [10].

It is likely that a set of vector fields will not have the same number of distributions. In order to satisfy constraint (22), we can create regular points to make the supply equal the demand without changing the vector fields. For example, if the supplier field contains 3 critical points

$$\vec{v} = \prod_{i=1}^{3} (a_i z + b_i \bar{z} + c_i) e_1 \tag{23}$$

and the consumer field contains 5 critical points

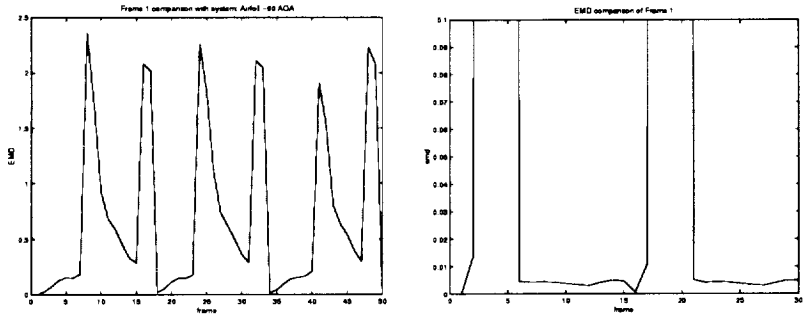
$$\vec{v'} = \prod_{j=1}^{5} (a'_i z + b'_i \bar{z} + c'_i) e_1 \tag{24}$$

The supplier side has two fewer points in the $\alpha - \beta$ space. Now let

$$\vec{v} = \prod_{i=1}^{3} (a_i z + b_i \bar{z} + c_i) \cdot 1 \cdot 1 e_1 \tag{25}$$

and the vector field remains unchanged. However, now we have two more regular points corresponding to 1 with $\alpha = 0$ and $\beta = 0$, and both the supplier and the consumer have 5 points in their feature distributions. All the conditions are satisfied, and we are ready to compute the EMD for these two fields and find out the dissimilarity between them.

In order to evaluate the meaningfulness of our new metric, we use *Multidimensional Scaling*(MDS) [14] [15] to embed the vector fields in a two-dimensional Euclidean space so that distances in the embedding are as close as possible to the true EMDs between vector fields. The MDS is introduced in the next section.



a b

Figure 3: EMD of the flow over an a) airfoil b)circular cylinder.

7 Display of EMDs for a Large Set of Vector Fields

The above discussions are for comparison of a pair of vector fields. If there exist a large set of vector fields and we want to compare their topologies, it is necessary to display them in a more meaningful way than a 1D list sorted by their EMDs. Yossi Rubner et al. have used Multidimensional Scaling Method (MDS) [10, 11] to display a set of images on a 2D map. Given n objects in a high dimension, the MDS method computes a configuration in a lower dimension space such that the distance between every pair of objects in this low dimension space best matches the real distance in the high dimension. Inspired by their work, we compute the EMDs between every pair of vector fields and position the vector fields on a 2D map such that the distances between the vector fields match their EMD values as close as possible.

8 Application: Flow over an Airfoil and Cylinder

Rogers and Kwak computed the flow past a 2-D airfoil at $-90°$ angle of attack [16]. The model was of interest since the flow of the wake of an XV-15 Tilt Rotor aircraft degraded the lifting capability during hover. An incompressible, time accurate, Navier-Stokes code with artificial compressibility at a Reynolds number of 200 was used to compute the flow over a NACA 64A223M airfoil. Fifty frames were computed. During this time the flow entered into a periodic vortex shedding cycle. Earth mover's distance was computed over the 50 frames. The plot in Figure 3a depicts the EMD comparison of frame 1 with the remaining 49 frames. At frame 1, the EMD is zero, which is expected since the work required to convert a frame into itself is zero. The periodic nature is apparent. We see a repetition approximately every 17 frames. Also we see

a sudden EMD rise when comparing frame 1 with frame 8 indicating a significant topological feature difference. Frame 1 contains three critical points: an attracting/repelling focus and a saddle. Frame 8 contains 5 critical points: two saddles, an attracting/repelling focus pair and a node. Since the flow is incompressible, the velocity divergence, $\nabla \cdot \vec{v}$, is expected to be zero everywhere in the flow. Hence only saddles and centers are to be extracted. However, it is common due to numerical computation for α to not exactly be zero, however, we should not expect to find a node. Upon closer examination of the data, the velocity divergence for certain frames is not zero near the tips of the foil where nodes are being extracted. We believe that the flow solver may not have fully converged and therefore we see this sudden jump of discontinuity in the field. The LIC images of frames 1, 34, and 50 are depicted in Figure 4 have very similar earth mover's distance and as can be seen look nearly identical. Frame 8 differs from the others due to its variation in topology (formation of nodes) and is apparent in the figure.

We contrast the flow over an airfoil with the flow over a circular cylinder simulated by Rogers and Kwak under the same flow conditions [16]. In this case, the flow is divergence free and the EMD values are quite similar. Thirty frames were computed capturing a complete cycle of vortex shedding. As can be seen from the plot in figure 3b, Frames 1 and 16 have nearly identical EMD values leading one to believe the period to be every 15 frames. Due to the symmetry of the flow, this is not far from the truth. In fact the flow produces a mirror image of itself every 15 frames as it sheds the alternate vortex and hence leads to the same topology. Figure 5b depicts the alternate vortex being shed to the image found in figure 5a. Furthermore, we see from figure 3b an increase in EMD value for frame 3. This increase is due to the dissipation of the saddle-center pair as it moves down stream (figure 5c). The EMD drops by frame 6 as the next saddle-center pair is shed (figure 5d). By frame 16, the saddle-center has moved down stream such that the α, β values are nearly identical to frame 1. Two frames later the saddle-center dissipate and the cycle repeats.

9 Discussion

We have demonstrated the effectiveness of topology based feature comparisons for vector fields. The use of a quantitative measure between fields provides the means for fast automated comparisons as well as an indepth study of flow fields as demonstrated with the time history data. For the airfoil data, we have shown the effectiveness of the method as a diagnostic tool. The clear EMD difference provides an immediate alert into calculation problems for particular frames. For the cylindrical data, the periodic nature of the flow was revealed. The EMD difference also provides insight into the evolution of the flow field.

We currently are researching techniques to extend this method to three dimensional vector fields and eventually to tensor fields.

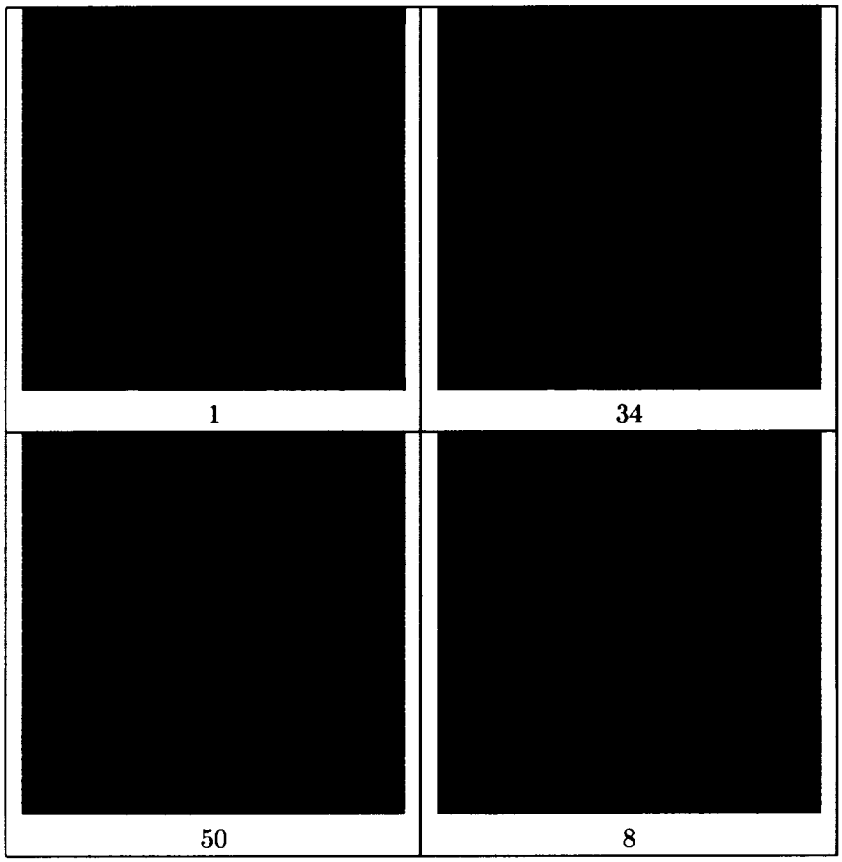


Figure 4: Topologically similar frames 1, 34 and 50 of flow about an airfoil. Frame 8 is topologically dissimilar.

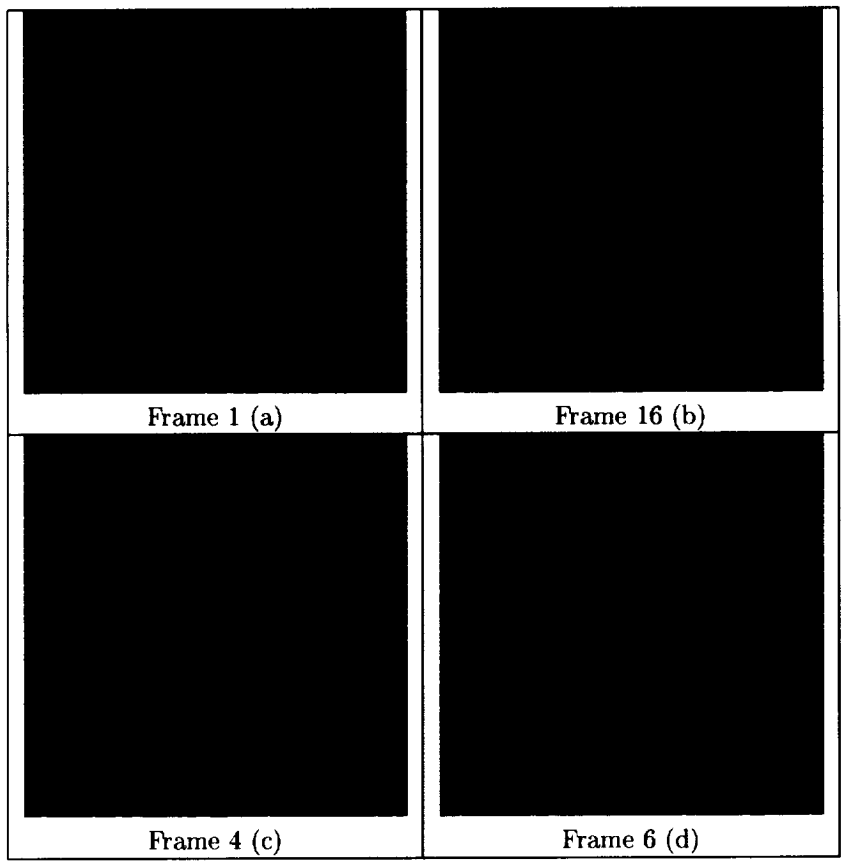


Figure 5: Topologically similar frames 1 and 16. Frame 4 depicts the down stream dissipation of saddle-center pair producing a larger EMD. Frame 6 depicts formation of new saddle-center pair at the bottom of the cylinder.

References

[1] F. H. Post and J. J. van Wijk, "Visual representations of vector fields: recent developments and research directions," in *Scientific Visualization* (L. Rosenblum, R. Earnshaw, and et al., eds.), pp. 181–195, Academic Press, Heidelberg, 1994.

[2] H.-G. Pagendarm and F. H. Post, "Comparative visualization-approaches and examples," in *Visualization in Scientific Computing* (M. Göbel, H. Müller, and B. Urban, eds.), pp. 95–108, Springer-Verlag, 1995.

[3] Q. Shen, A. Pang, and S. Uselton, "Data level comparison of wind tunnel and computational fluid dynamics data," in *Proc. IEEE Visualization '97*, pp. 67–74, CS Press, Los Alamitos, CA., 1997.

[4] L. Hesselink, "Digital image processing in flow visualization," *Annual Review of Fluid Mechanics*, vol. 20, pp. 421–485, 1988.

[5] J. Helman and L. Hesselink, "Representation and display of vector field topology in fluid flow data sets," *IEEE Computer*, vol. 22, pp. 27–36, Aug 1989. Also appears in *Visualization in Scientific Computing*, G. M. Nielson & B. Shriver, eds. Companion videotape available from IEEE Computer Society Press.

[6] Y. Lavin, R. Batra, and L. Hesselink, "Feature comparisons of vector fields using earth mover's distance," in *Proc. IEEE/ACM Visualization '98*, pp. 413–415, North Carolina, October 1998.

[7] G.Sheuermann, H.Krüger, M.Menzel, and A.Rockwood, "Visualization of higher order singularities in vector fields," in *Proc. IEEE Visualization '97*, pp. 67–74, CS Press, Los Alamitos, CA., 1997.

[8] G.Sheuermann, H.Kruger, M.Menzel, and A.Rockwood, "Visualizing non-linear vector field topology," *Computer Graphics and Its Applications, Special Edition*, 1998.

[9] Y. Lavin, *Topology based visualization for vector and tensor fields.* PhD thesis, Stanford University, 1998.

[10] Y. Rubner, C. Tomasi, and L. J. Guibas, "A metric for distributions with applications to image databases," in *Proc. IEEE Internations Conferences on Computer Vision*, 1998.

[11] Y. Rubner, L. J. Guibas, and C. Tomasi, "The earth mover's distance, multi-dimensional scaling, and color-based image retrieval," in *Proceedings of the ARPA Image Understanding Workshop*, 1997.

[12] G. B. Dantzig, "Application of the simplex method to a transportation problem.," *Activity Analysis of Production and Allocation*, pp. 359–373, 1951.

[13] G. B. Dantzig, *Linear programming.* Springer series in operations research, Springer, 1997.

[14] J. Kruskal, "Multi-dimensional scaling by optimizing goodness-of-fit to a non-metric hypothesis.," *Psychometrika*, vol. 29, pp. 1–27, 1964.

[15] R. Shepard, "The analysis of proximities: Multidimensional scaling with an unknown distance function, i and ii.," *Psychometrika*, vol. 29, pp. 125–140,219–246, 1962.

[16] S. Rogers and D. Kwak, "An upwind differencing scheme for the time-accurate incompressible navier-stokes equations," in *Proceedings of the AIAA 6th Applied Aerodynamics Conference*, pp. 492–502, American Institute of Aeronautics and Astronautics, 1988.