

# Measuring the Similarity of Vector Fields Using Global Distributions

H. Quynh Dinh and Liefei Xu

Department of Computer Science  
Stevens Institute of Technology  
quynh@cs.stevens.edu, lxu1@stevens.edu

**Abstract.** Sensors such as video surveillance and weather monitoring systems record a significant amount of dynamic data which are represented by vector fields. We present a novel algorithm to measure the similarity of vector fields using global distributions that capture both vector field properties (e.g., vector orientation) and relational geometric information (e.g., the relative positions of two vectors in the field). We show that such global distributions are capable of distinguishing between vector fields of varying complexity and can be used to quantitatively compare similar fields.

**Keywords:** vector field matching, shape distribution, geometric histogram.

## 1 Introduction

Comparing vector fields is becoming increasingly relevant as sensor systems now record a significant amount of dynamic data. For example, vector fields may come from video surveillance data (via tracking and optical flow), from weather phenomena (wind speed and direction and wave currents), and from medical imaging (phase-contrast magnetic-resonance angiography captures 3D velocity fields of moving tissue [1]). In video analysis, comparing these vector fields may improve the recognition of activities, faces, and facial expressions [2,3]. For weather data, measuring the similarity between vector fields can help scientists understand how wind and wave currents have changed over time and to compare with simulations. In medical imaging, measuring the difference between vector fields enable doctors to analyze changes in tissue over time and study the variation of particular organs or tissue groups over large populations.

We present a novel algorithm to measure the similarity of vector fields using global distributions (one for each vector field) that can be quantitatively compared by computing the difference between them. Our key contribution is the development of global distributions that capture both vector field properties such as vector orientation and relational information (e.g., the relative positions of two vectors in the field). Previous work in 3D shape matching has proven that relational information is effective in shape classification, and we show that this is also true for comparing vector fields.

## 2 Related Work

Vector field analysis identifies and locates singularities (critical points where the magnitude of the vector vanishes), separatrices emanating from saddle points, and periodic

orbits. Critical points are further classified into sinks, sources, saddles, focus, and center points using the eigenvalues and eigenvectors of the Jacobian matrix. There is significant research in the area of vector field analysis and singularity classification, including [4,5,6,7,8,9]. The algorithm we present does not depend on locating critical points in vector fields because many naturally occurring vector fields do not actually contain singularities and because locating such points is computationally intensive. Instead, we collect statistics on the vector field in the form of discrete distributions, or histograms, and measure the similarity between vector fields by computing the difference, or norm, between the histograms. As results show, our approach is able to differentiate between vector fields with different configurations of singularities even though we do not explicitly locate them.

Geometric distributions (also called *shape histograms* [10] or *shape distributions* [11]) provide a way of comparing shapes based on statistical properties. These approaches record the distribution of a selected feature in a histogram for efficient storage, indexing, and comparison via norms on the histogram. A single histogram may be generated for the entire field (global histograms), or one for each surface point (local histograms) to be used as a local descriptor. Global histograms store the statistics of the entire shape and are used to compare whole objects. To compare two objects, a norm is computed between the histograms. Standard methods include the Minkowski  $L_N$  norms, the  $\chi^2$  distance, the Battacharyya distance, Earth Mover's distance, and the correlation coefficient. Geometric distributions are known for their invariance to rotation and translation of the shape and robustness to noise due to their statistical nature.

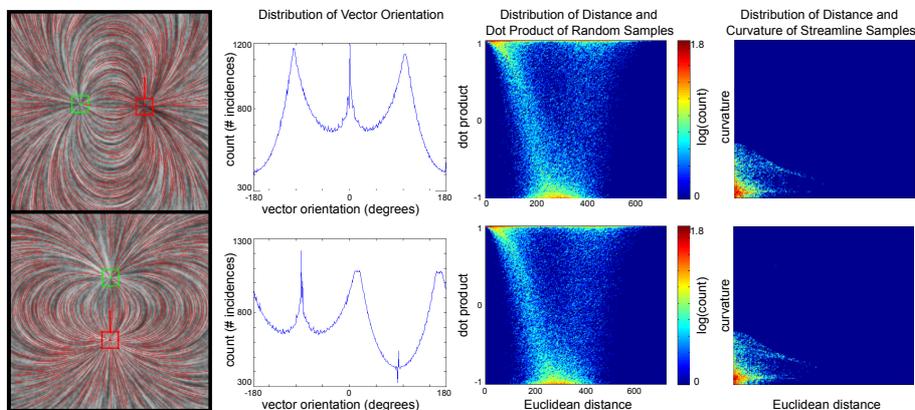
Statistics on vector fields from optical flow have been explored in [12] where the goal was to obtain a statistical consensus for use as a prior in optical flow computation, not to find a discriminating distribution. We focus on global distributions that enable us to efficiently compare two vector fields by computing the difference between a single pair of histograms.

### 3 Global Distributions for Vector Fields

In developing a global distribution for vector fields, our goal is to find the vector field property that will result in discriminating distributions that can act as *signatures* for vector fields. We now describe different global distributions we have developed, and in Section 4, show results of measuring the similarity between test vector fields of varying complexity – from few to many singularities, complex fields that contain no singularity, and real vector fields designed for painterly rendering [9] and from flow through a diesel engine cylinder [13]. We generate a single global distribution for each vector field. The distribution is a discrete histogram in which each bin covers a range of values for a given vector field property, and each bin is a count of the number of times its range was encountered in the vector field.

#### 3.1 A Simple One-dimensional Distribution

An intuitive statistic for 2D vector fields is to bin the vectors in the field according to orientation, resulting in a 1D distribution as shown in the 2nd column of Figure 1.



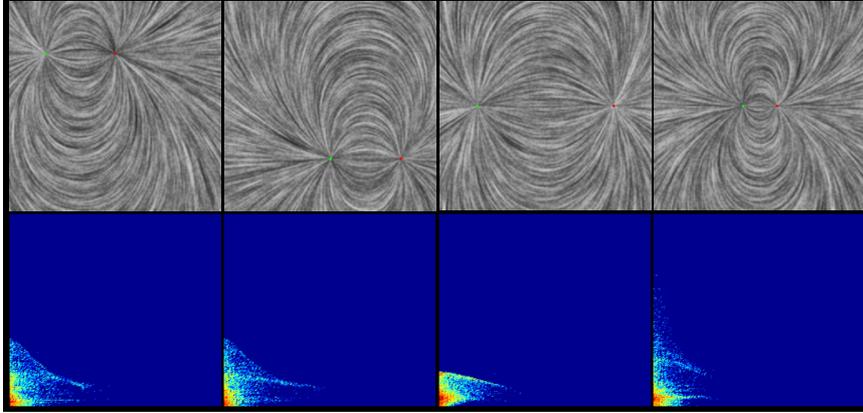
**Fig. 1.** Left to right: Two vector fields related by a rotation (green indicates a source, red a sink) overlaid with streamlines; histogram of vector orientation; histogram of distance and vector dot product between pairs of randomly selected points; and histograms of distance and curvature between points on streamlines

The 1D distributions are clearly equivalent up to a phase shift (due to the rotation of the singularities), and are thus sensitive to rotations in the vector field. To find the minimum difference between the histograms requires shifting one plot with respect to the other.

The 1D histogram also fails to account for geometric, or positional, relationships between vectors. In 3D shape matching, Osada *et al.* found that histograms which preserve positional relationships are more effective for shape matching [11]. They showed that a 1D distribution based on binning the Euclidean distance between pairs of randomly selected surface points was particularly successful in 3D shape classification. Ohbuchi *et al.* extend this concept to a 2D distribution by binning 3D surface points based on distance and the dot-product between surface normal vectors [14]. We now describe a geometry-preserving 2D distribution for vector fields that captures both statistics on vector properties (*e.g.*, vector dot product) and positional relationships between points in the vector field using the Euclidean distance.

### 3.2 Geometry-Preserving Distributions

We present a novel 2D geometry-preserving distribution for vector fields. The distributions are generated by randomly selecting pairs of points, computing the distance between them and the dot product between the vectors located at the points, and then binning the pairs in a 2D histogram based on these two computed values. The 3rd column in Figure 1 shows the resulting histograms for the vector fields on the left. A difference between the distributions can be directly computed (we use  $\chi^2$  distance) without requiring any shifts, unlike the 1D distributions based on vector orientation. Although computing the difference between 2D histograms is equivalent to shifting to find the min difference between 1D histograms, the 2D histograms are more informative and discriminating. In particular, Figure 1 shows that the 2D histograms are invariant to rotations in the vector field. The 2D geometry-preserving distributions (3rd and 4th



**Fig. 2.** Top: translation (columns 1 and 2) and scaling (columns 3 and 4) of the vector field shown in Figure 1 (top-left). Bottom: histograms of distance and curvature between points on streamlines show that the distributions are relatively invariant to translation and scaling. The  $\chi^2$  distance between the above distributions and the streamline distribution of the vector field in Figure 1 (top-left) are, from left to right: 0.077, 0.066, 0.125, 0.052 out of a max of 1.0.

columns) for the top and bottom vector fields are equivalent even though the position of the singularities are rotated by 90 degrees.

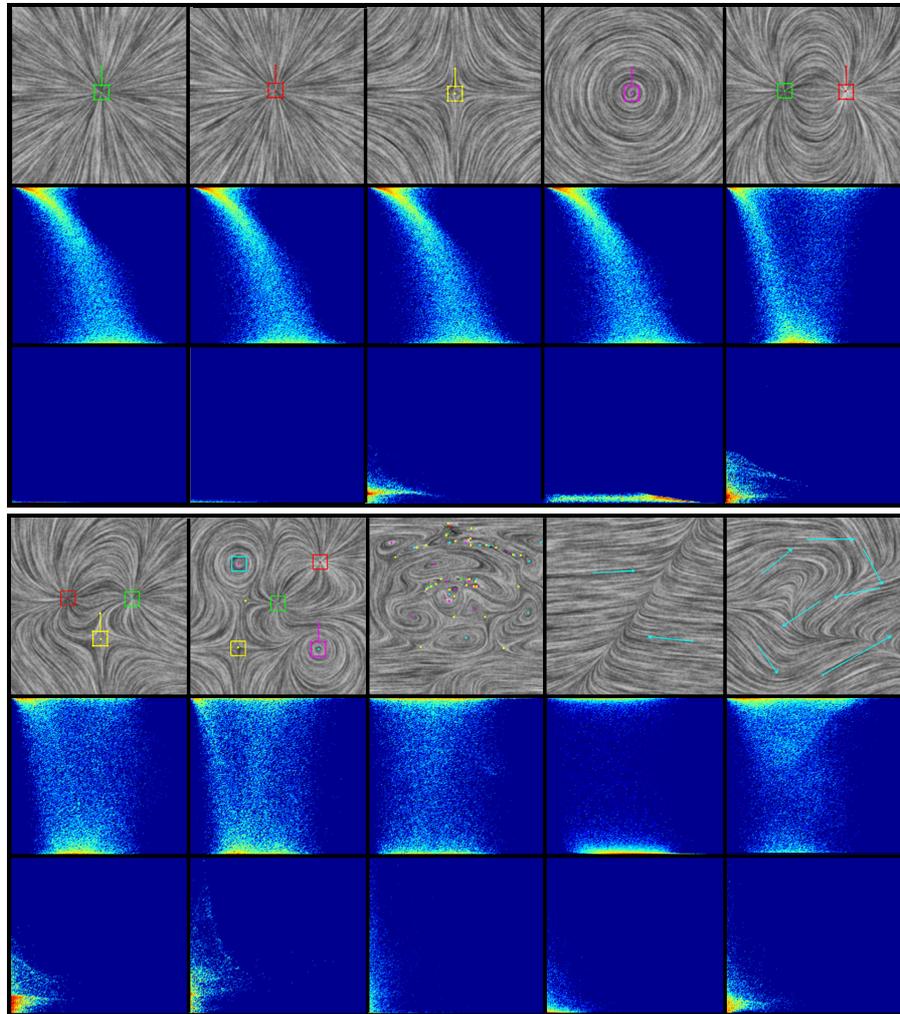
Note that we do not compute the distance and vector dot-product between every pair of points in the domain because it is computationally expensive (for  $512 \times 512$  vector fields, there are  $512^4$  pairs). Instead, we use a large number of samples – 10,000 pairs of points are randomly selected. Figure 4 (left) shows the  $\chi^2$  distance between the vector field distributions based on the distance and dot-product between randomly selected pairs of points for 14 different data sets (10 of which are shown in Figure 3).

### 3.3 Distributions over Streamlines

The 2D distribution we describe above can distinguish between simple and complex vector fields (*e.g.*, those with one singularity versus many). However, it cannot distinguish between different complex fields or different simple fields. For example, the plot of Figure 4 (left) shows that simple fields 1 through 4 are equally different from each other. The same is true of complex fields 5 through 12 and 14 in the plot.

To compute a distribution that further distinguishes between vector fields, we implicitly record additional structural information by extracting streamlines from the vector field (as shown in Figure 1). Streamlines are formed by randomly seeding points in the field and allowing them to move through the field guided by the vectors. The paths taken by the points form streamlines. We uniformly distribute streamlines by preventing them from growing into regions where streamline density is high, similar to [16].

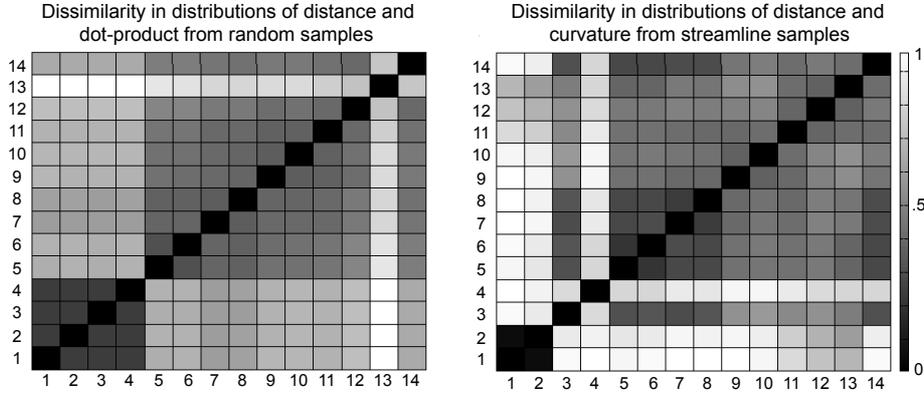
Streamlines enhance the structure present in vector fields and provide a framework for recording positional relationships between vectors. Instead of randomly selecting pairs of points from across the entire vector field as in the previous section, we randomly



**Fig. 3.** Top row: vector fields corresponding to 1, 2, 3, 4, 5, 8, 9, 12, 13, 14 in plots of Figure 4. Middle row: distance and dot-product distributions from random samples over entire field. Bottom row: distance and curvature distributions from streamline samples. Due to space limitations, we have omitted four vector fields which are rotated versions of some of the above. They can be found at our website [15].

select pairs of points from the *same* streamline. The intuition is that points on the same streamline have a stronger geometric relationship than randomly selected pairs of points which may be very far apart. The resulting distribution now only collects data between strongly related points that are along the same path in the vector field.

With streamlines available, we can record curvature as the second property (rather than vector dot product) in the 2D histogram. Given two randomly selected points on



**Fig. 4.**  $\chi^2$  distance between 2D distributions of 14 different vector fields. Darker (closer to 0) means more similar. Vector fields are shown in Figure 3.

the same streamline, we compute the mean of the two points' discrete curvature and bin them based on distance and mean curvature. For a point  $(x, y)$  on the streamline, the discrete curvature  $k(x, y)$  is:

$$k(x, y) = \frac{|x'y'' - y'x''|}{(x'^2 + y'^2)^{3/2}} \quad (1)$$

In the equation above,  $x'$  and  $x''$  are the discrete first and second derivatives of  $x$ , respectively (and similarly for  $y$ ). Each streamline is an ordered list of uniformly distributed points, making it easy to compute the discrete derivatives via finite differencing. The last column of Figure 1 are distributions of distance and curvature between points sampled on the same streamline. Figure 3 (bottom rows) are distributions of distance and curvature for the 10 vector fields shown above them. The color coding in Figure 1 is used for all distributions.

In Figure 1, we showed the rotational invariance of our geometry-preserving distributions. We further test the translational and scale invariance by moving the pair of sink and source singularities and by scaling them to be farther apart and closer together. In comparing the distributions of Figure 2 to that of Figure 1 (4th column, top), we can see that the distributions are invariant to translation (1st two columns) and partially invariant to scaling (last two columns). When the vector field is scaled-up, the singularities are farther apart and cover a larger region of the field. The original vector field is essentially cropped, and results in a distribution that, when compared to the original, measures a difference. Scale invariance does hold when the vector field is scaled down.

## 4 Results and Conclusions

We now present results of using our geometry-preserving distributions to measure similarity between a set of 14 test vector fields and on the flow through a diesel engine cylinder, a real-world dataset.

#### 4.1 Measuring Similarity on Test Vector Fields

We use Zhang *et al.*'s vector field design tool [9] to generate a test dataset of 14 vector fields of varying complexity – from few to many singularities (1, 2, 3, 5, and >5), complex fields that contain no singularity, and two real vector fields designed for painterly rendering [9] (fields 11 and 12). Figure 3 shows 10 of the vector fields. Source singularities are green; sinks are red; saddles are yellow; centers are pink (attracting) and cyan (repelling); and regular vectors are cyan arrows. Squares surrounding singularities are control knobs for rotation. All vector fields are 512x512, and histograms 200x200 in size. The  $\chi^2$  distance between N-bin normalized histograms,  $f$  and  $g$ , is:

$$\chi^2 : D(f, g) = \frac{1}{2} \sum_{i=1}^N \frac{(f[i] - g[i])^2}{f[i] + g[i]} \quad (2)$$

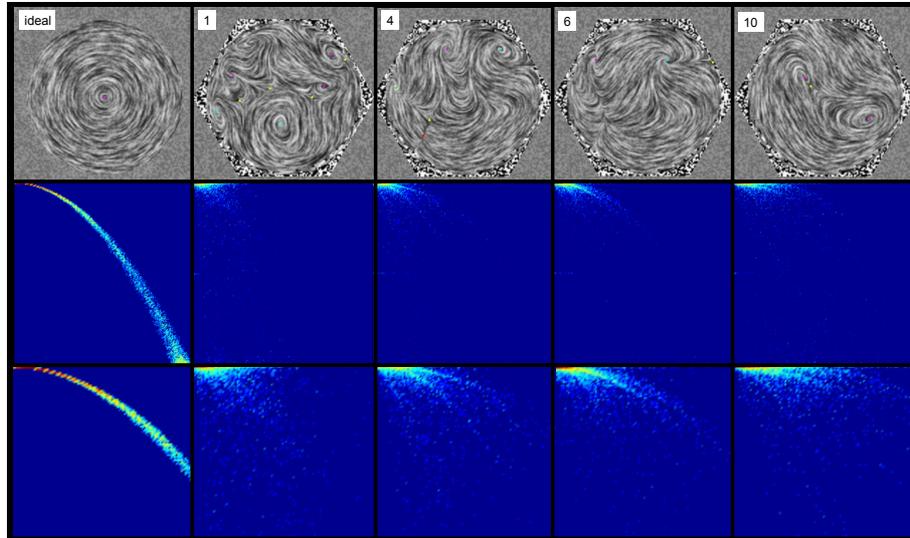
Results are plotted in Figure 4 for all 14 vector fields (darker means more similar). The plots show that streamline distributions (right plot) provide a finer measure of similarity (or dissimilarity) between the vector fields than our first geometry-preserving distribution (left plot) which samples the entire field and captures broad differences between them. This also bears out in the distributions themselves (Figure 3) which reveal streamline distributions to be more distinct than those generated from randomly sampling the entire vector field. We found the computed differences between streamline distributions of distance and dot-product to be similar to results using curvature.

Note that vector fields 1 and 2 are essentially identical – 1 contains a sink in the center, whereas 2 contains a source. The only difference between the vector fields is the vector direction (inward for a sink, outward for a source). The streamline distributions show that these two fields are very similar, and the  $\chi^2$  distance is close to 0. Intuitively, all samples on the same streamline with have a dot-product of 1.0 and a curvature near 0 for both vector fields. Randomly sampling from the entire field does not preserve this relationship, and the computed  $\chi^2$  distance is greater.

#### 4.2 Measuring Similarity on Engine Cylinder Slices

We tested our algorithm on a real-world dataset – the simulated flow within the combustion chamber of a diesel engine. Engineers typically slice through the cylinder along its length to analyze the simulation results, resulting in the circular flow patterns seen in Figure 5. The ideal flow pattern is a spiral helix extending the length of the cylinder. This motion results in an optimal mixing of air and fuel leading to a more efficient combustion process [13]. The 2D flow pattern obtained from slicing the ideal helix is a swirl motion (Figure 5 left). By comparing the slices of the vector field along the cylinder to the ideal swirl, we can provide a quantitative analysis of how close the flow through a designed cylinder is to ideal, and order the slices based on similarity to the ideal. This ordering provides engineers with information on where to improve their design.

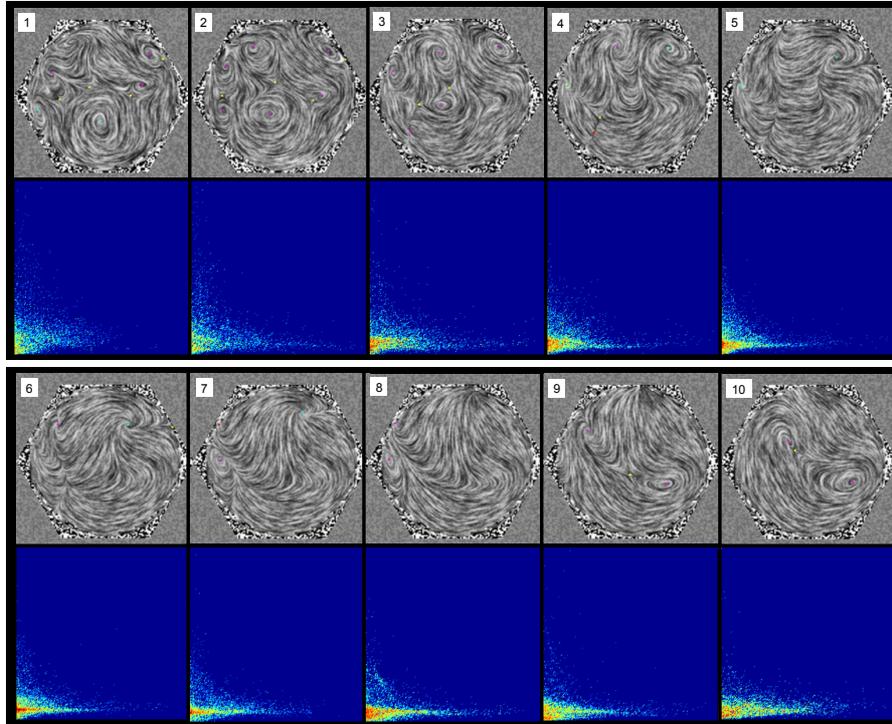
We computed the similarity of the ideal flow to 10 slices through the cylinder using streamline distributions and computed the  $\chi^2$  distance between them. In Figure 5,



**Fig. 5.** Left to right: ideal flow in a slice of the engine cylinder, and slices 1, 4, 6, and 10 from the simulated flow in a designed combustion chamber. 1 and 10 are near the cylinder entrance and exit, respectively, and are where the most turbulence occurs. Slice 6 is closest to ideal. Top: planar vector fields from cylinder data. Middle: streamline distributions of distance and dot-product for each slice. Bottom: close-up of top-left corner of the histograms show how they differ. Note that the data is on a hexagonal grid, leading to the hexagonal boundaries around the circular slices.

we show 4 representative slices and their streamline distributions of distance and dot-product. The numbering of the slices corresponds to their positions along the length of the cylinder from where the fluid enters at slice 1 to where it exits at slice 10. For completeness, we show all 10 slices and distance and curvature distributions in Figure 6.

The simulation data is embedded in a hexagonal grid, and hence, are essentially aligned. We can directly compare vectors of a slice to the ideal slice in a point-wise manner at each grid point using the dot-product. To do so, we sum up 1 minus the normalized vector dot-product for each grid point pair, so that similar vectors will contribute 0, while vectors pointing in opposite directions will contribute 2 towards the sum. Similar slices will have a sum closer to 0. For a consistent comparison between this direct approach using the dot-product and our streamline approach, we compute the streamline distributions of distance and dot-product. The results of both methods are similar and consistent with the engineers' expectations. Slices close to the top of the cylinder where the fluid enters and slices near the bottom where the fluid exits tend to be most turbulent and far from ideal. Slices in the center where the flow approaches a stable steady-state is closer to ideal. Using streamline distributions based on dot-product, the order of the 10 slices from most to least similar to ideal is: 6, 7, 5, 8, 9, 10, 4, 3, 2, 1. Using the sum of 1 minus normalized dot-product between corresponding pairs, the order is: 6, 7, 5, 8, 1, 4, 3, 2, 9, 10. The point-wise comparison is, of course, more efficient than computing and comparing the streamline distributions.



**Fig. 6.** 10 slices of the simulated flow within a designed combustion chamber and histograms of distance and curvature between randomly selected pairs of points on the same streamline. Note that the distribution for the ideal slice (not shown) is a horizontal line near the bottom.

Although we are not taking advantage of the rotational invariance of our distributions (since the cylinder slices are aligned), we have found that the positional, or contextual, information implicitly stored in the distributions result in a better ordering of slice similarity than the point-wise comparison. In the point-wise comparison, slice 1 falls in the middle of the ordering based on similarity to ideal even though it is the slice with the most visual turbulence, and slices 2 and 3 are closer to ideal than slices 9 and 10 near the exit even though they appear more turbulent than 9 and 10.

## 5 Future Work

We have shown that geometry-preserving distributions – in particular, streamline distributions – enable us to quantitatively compare complex vector fields and distinguish between them. They are invariant to rotation and translation of the vector field, and have limited invariance to scaling. In future work, we will apply our algorithms to vector fields generated from optical flow and dense tracking of video.

## Acknowledgements

We thank Robert S. Laramee of Swansea University for the original diesel engine data, and Eugene Zhang and Lin Zhongzang of Oregon State for extracting the planar slices from the data, for the vector field design tool, and for the vector fields from painterly rendering. This work was funded in part by NSF CreativeIT Award# IIS-0742440.

## References

1. Tovar, M.: Vector-field classification in magnetic-resonance angiography. In: Proc. of AMIA Symposium, pp. 926–930 (1998)
2. Pamudurthy, S., Guan, E., Mueller, K., Rafailovich, M.: Dynamic approach for face recognition using digital image skin correlation. Audio- and Video-based Biometric Person Authentication (2005)
3. Lu, X., Jain, A.: Deformation modeling for robust 3d face matching. In: Proc. of CVPR (2006)
4. Helman, J., Hesselink, L.: Surface representations of two- and three- dimensional fluid flow topology. In: Proc. of IEEE Visualization, pp. 6–13 (1990)
5. Scheuermann, G., Krüger, H., Menzel, M., Rockwood, A.: Visualizing nonlinear vector field topology. IEEE Transactions on Visualization and Computer Graphics (TVCG) 4(2), 109–116 (1998)
6. Tricoche, X., Scheuermann, G., Hagen, H.: Continuous topology simplification of planar vector fields. In: Proc. of IEEE Visualization, pp. 159–166 (2001)
7. Tricoche, X., Scheuermann, G., Hagen, H.: Topology-based visualization of time-dependent 2d vector fields. In: Data Visualization, Proc. of IEEE TVCG Symposium on Visualization, pp. 117–126 (2001)
8. Ebling, J., Scheuermann, G.: Clifford convolution and pattern matching on vector fields. In: Proc. of IEEE Visualization, pp. 193–200 (2003)
9. Zhang, E., Mischaikow, K., Turk, G.: Vector field design on surfaces. ACM Transactions on Graphics (TOG) 25(4), 1294–1326 (2006)
10. Ankerst, M., Kastenmuller, G., Kriegel, H., Seidl, T.: 3d shape histograms for similarity search and classification in spatial databases. In: Proc. of 6th International Symposium on Spatial Databases (1999)
11. Osada, R., Funkhouser, T., Chazelle, B., Dobkin, D.: Shape distributions. ACM Trans. on Graphics 21(4), 807–832 (2002)
12. Roth, S., Black, M.: On the spatial statistics of optical flow. Int'l. J. of Computer Vision 74, 33–50 (2007)
13. Laramee, R., Weiskopf, D., Schneider, J., Hauser, H.: Investigating swirl and tumble flow with a comparison of visualization techniques. In: Proc. of IEEE Visualization, pp. 51–58 (2004)
14. Ohbuchi, R., Minamitani, T., Takei, T.: Shape-similarity search of 3d models by using enhanced shape functions. In: Proc. Theory and Practice of Computer Graphics, pp. 97–104 (2003)
15. [http://www.cs.stevens.edu/~quynh/vfield\\_global.html](http://www.cs.stevens.edu/~quynh/vfield_global.html)
16. Verma, V., Kao, D., Pang, A.: A flow-guided streamline seeding strategy. In: Proc. of IEEE Visualization, pp. 163–170 (2000)