Eurographics Conference on Visualization (EuroVis) (2015)
J. Kennedy and E. Puppo (Editors)

Clustering Moment Invariants to Identify Similarity within 2D Flow Fields

Roxana Bujack [2], Jens Kasten [2], Vijay Natarajan [3], Gerik Scheuermann [2]and Kenneth I. Joy [1]

[1] University of California Davis, USA
[2] Leipzig University, Germany
[3] Indian Institute of Science, Bangalore

Abstract

Moment invariants have proven to be a useful tool for the detection of patterns in scalar and vector fields. By their means, an interesting feature can be detected in a data set independent of its exact orientation, position, and scale. In this paper, we show that they can also be applied to explore an unknown dataset without prior determination of a query feature pattern it may possibly contain. The clustering of the high dimensional moment space reveals the major structures in the underlying flow field and gives an excellent overview for subsequent more profound exploration.

Categories and Subject Descriptors (according to ACM CCS): Image Processing and Computer Vision [I.4.7]: Feature Measurement —Moments;

1. Introduction

Scientific data arising from simulations in several science and engineering disciplines are often represented as vector fields. Ever increasing complexity of these vector fields necessitates the development of new methods for flow visualization and exploration. Direct visualization and exploration are no longer sufficient for data understanding due to the presence of visual clutter. Several researchers have advocated feature-directed approaches to address this challenge [CEH*09, NMM09]. Such approaches are effective in the context of flow visualization because the study of flow features is central to the knowledge discovery process. In the absence of a precise definition of a feature, query- or example-based exploration methods address the problem by supporting the specification of an exemplar feature followed by automatic identification of similar features or patterns within the field [BAAK*13, LCL*13, TN14, WWYM10]. The successful use of the method crucially depends on the user's ability to pose appropriate queries.

In this paper, we study the problem of automatic identification of exemplar features within an input vector field. Specifically, we propose a clustering based method to identify different types of repeating patterns both in steady and unsteady vector fields. We identify similarity within a flow field by applying a clustering algorithm to its moment invariants. That way similar regions of the field can be found without any regard to differing background flow, velocity, or orientation.

In a nutshell, the algorithm works as follows.

1. We calculate the moment invariants for every point.
2. We cluster these normalized moment vectors.
3. Color each point in correspondence with the cluster of its moments.

2. Related work

The use of moment invariants in pattern recognition started when Hu [Hu62] introduced his famous seven moment invariants in 1962. Complex moments [Tea80, AMP84] especially simplified the construction of rotational invariants because of the easy way to describe rotations using complex exponentials. Two principle approaches for the construction of an independent basis of moment invariants have been introduced. Dirilton and Newman [DN77] use normalization, Flusser [Flu00] explicitly defines the set of invariants. The step from scalar to vector valued data took some time. Building on Flusser's work, Schlemmer et al. [SHM*07, Sch11] have defined the first moment invariants for flow fields. Following the normalization approach Bujack et al. [BHSH14a] have presented a concise mathematical framework for the

construction of a complete and independent set of moment invariants for flow fields that is adaptable to vanishing moments. Both approaches have successfully been applied to flow pattern recognition tasks. For a more comprehensive introduction to moment invariants we recommend [FZS09].

Clustering algorithms have been applied to vector fields albeit for partitioning the field. Several distance measures have been proposed for comparing streamlines and other similar structures [TvW99, GPR*00, MVvW05, KNC*14]. Since these approaches are local and not invariant with respect to orientation, certain features, like vortices, are separated into many different clusters because the vectors in them point into very different directions. In contrast to that, our algorithm keeps the whole vortex in one cluster, which is more intuitive.

3. Moment Invariants for Flow Fields

In [BHSH14a], Bujack et al. constructed moment invariants for two-dimensional flow fields using complex moments and normalization. In this section, we will summarize the most important facts.

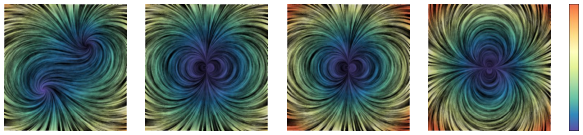
The moments of a function are its projections with respect to a function space basis. We handle functions with compact support over $\mathbb{R}^2 \simeq \mathbb{C}$ and use complex moments

$$c_{p,q} = \int_{\mathbb{C}} z^p \bar{z}^q f(z)\,\mathrm{d}z, \tag{1}$$

which are the projections to the standard complex monomials $z^p\bar{z}^q$. Calculated up to a given order $p+q=n \in \mathbb{N}$, they form a feature vector $c = (c_{0,0},...,c_{0,n})^T \in \mathbb{C}^{(n+1)(n+2)/2}$ in a high dimensional feature space. The coordinates in this space describe the function. That means functions can be compared using the distance in the moment space.

The user is free to choose any combination of the following flow field transformations with respect to which he wants the moments to be invariant. The **Outer Translation** by $t \in \mathbb{C}$, $f'(z) = f(z)+t$, can be interpreted as a distortion of the pattern by some background flow or as a moving frame of reference. The **Outer Scaling** by the factor $s \in \mathbb{R}^+$, $f'(z) = sf(z)$, changes the velocity of the flow. The **Total Rotation** by $\alpha \in [-\pi,\pi)$ adequately describes the behavior of a flow field when its orientation is changed, $f'(z) = e^{i\alpha} f(e^{-i\alpha}z)$. The **Total Reflection** is closely related to total rotation. An arbitrary reflection can be produced from two total rotations and a total reflection along the real axis. Since we cover the rotation separately, it is sufficient to consider the total reflection along the real axis, $f'(z) = \overline{(f(\bar{z}))}$.

Moment invariants can be constructed from moments via normalization, which works as follows. In order to achieve an invariant description of a field, a standard position is defined and the field is transformed into this standard position. An illustration of the normalization process of an example flow field is shown in Figure 1. Please note that the figures are for illustrative purpose only. In the implementation, the function itself is not put into standard position, but only its moments. Certain moments are set to predefined values. They will take the same values for any field. All the remaining moments can be used as independent discriminators.



The example. Background. Velocity. Rotation.

Figure 1: Normalization example, color represents speed.

Suppose the complex function $f : \mathbb{C} \to \mathbb{C}$ with support on the circular area $A = B_r(0)$ is transformed by outer scaling by $s \in \mathbb{R}^+$, outer translation by $t \in \mathbb{C}$, and total rotation by $\alpha \in [-\pi,\pi)$ and let

$$f'(z) = se^{i\alpha}\left(f\left(e^{-i\alpha}z\right)+t\right). \tag{2}$$

Then, the complex moments $c'_{p,q}$ of f' satisfy

$$c'_{p,q} = se^{i\alpha(p-q+1)}\left(c_{p,q} + t\int_A z^p\bar{z}^q\,\mathrm{d}z\right). \tag{3}$$

Under total reflection, the moments are complex conjugated. For indices $p_0, q_0 \in \mathbb{N}$ with $c_{p_0,q_0} \neq 0$, we can calculate the transformation parameters that put the function in the standard position defined by $c_{0,0} = 0, c_{p_0,q_0} = 1$. They suffice

$$t_0 = -\frac{c_{0,0}}{\int_A \mathrm{d}z}, s_0 = \frac{1}{|c_{p_0,q_0}|}, \alpha_0^j = \frac{2j\pi - \arg(c_{p_0,q_0})}{p_0 - q_0 + 1}, \tag{4}$$

with $j = 1,...,|p_0 - q_0 + 1|$ and transform the moments into moment invariants using (3). A detailed description of the procedure can be found in [BHSH14b].

4. Adaption of the Normalization

The idea of clustering the moment feature space presents new challenges for the normalization.

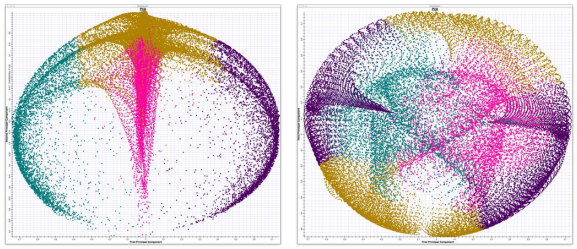
In order to solve the pattern detection task, Bujack et al. [BHSH14b] used the moment with highest magnitude in the pattern as the normalizer c_{p_0,q_0}. Since we do not have a pattern that suggests the normalizing moment, we choose the dominant moment throughout the whole field $p_0, q_0 = \mathrm{argmax}_{p+q<n}(\sum_{z\in\mathbb{C}} c_{p,q}(z)^2)$. Even though this one is the best choice, the original normalization process will lead to problems at the positions where this moment vanishes.

First, normalization with respect to scaling causes huge entries, which pull the centroids of the clusters far away from the reasonable areas in the feature space. We solved this problem by normalizing the whole moment vector in the feature space to have unit Euclidean length $c'_{p,q} = c_{p,q}/\sqrt{\sum_{p+q<n} c_{p,q}(z)^2}$, if it was not numerically zero in the first place.

Second, normalization with respect to rotation relies on the argument of c_{p_0,q_0}, which is unstable for vanishing magnitude. This problem is taken care of using a stabilizer that weighs the overall length of the moment vector with the magnitude of the normalizing moment $c'_{p,q} = c'_{p,q}\sqrt{1-(1-|c'_{p_0,q_0}|)^2}$ allowing a smooth transition when it passes through zero. This strategy will equalize the positions of a vanishing normalizer. An alternative that maintains a discrimination in these areas is to extend the moment vector by the normalized moments with respect to different normalizers.

5. Clustering

The normalized moment vectors live in the high dimensional unit sphere in the feature space. They tend to accumulate along the coordinate axes, especially for smaller radii, which emphasize lower order moments. The principle component analysis (PCA) of the feature space reveals that principle axes are often correlated to coordinate axes. An illustration of the PCA of the feature space of our case study from Section 6 can be found in Figure 2.



Radius is one. Radius is two.

Figure 2: PCA of the feature space of the vortex street from Figure 3, color represents the cluster affiliation for $k = 4$.

To perform the clustering in the feature space, we choose the well known k-means algorithm [Mac67]. For a given number $k \in \mathbb{N}$, it works as follows. Initially, k cluster centroids are randomly placed in the data set. Then iteratively, two steps are performed until the cluster formation does no longer change.

1. Each point is assigned to its closest cluster centroid.
2. The new centroids are the mean of all assigned points.

We use k-means because it is fast, partitions the feature space into clusters of similar diameter, and does not take varying density into account. This is useful in our application because the further two points are away from each other in the feature space, the more their respective vector field patterns differ in shape. In contrast to that, the density does not resemble any similarity or dissimilarity of the patterns, but only represents how frequently this pattern appears throughout the data set. Therefore, a decomposition into evenly distributed, equally sized clusters represents a good partitioning with respect to the similarity of the patterns.

One issue of k-means is the distribution of the initial cluster centroids because it influences the number of iterations and the outcome of the algorithm. We place the first centroid into the point representing the normalizing moment and keep adding the data points with the largest distance to all previously selected centroids, similar to k-means++. That guarantees an even and deterministic initial distribution, which is in accordance with the nature of the feature space and k-means itself.

When it comes to the distance measure, we have to be careful. The number of equally valid standard positions in the normalization process depends on the rotational symmetry of the normalizer. To treat that, we equip k-means with the metric

$$dist(c,c') = \min_{j=1,\dots,2|p_0-q_0+1|} \sqrt{\sum_{p+q\le n} (c^j_{p,q} - c'^j_{p,q})^2} \quad (5)$$

introduced in [BHSH14a]. The coloring in the right of Figure 2 illustrates how the purple (or yellow) cluster actually combines the moment vectors that only differ because they were mapped to different but equivalent standard positions during normalization.

6. Case Study

In order to demonstrate the outcome of our algorithm, we use a 2D computational fluid dynamic simulation of a von Kármán vortex street shown in Figure 3. We visualize the allocation of the regions in the field to the clusters by dyeing each point in the color assigned to its cluster. We use a rainbow color map with constant intensity and brightness. It guarantees a maximal variety of hues and prevents the emphasis of certain clusters that are randomly assigned a color of higher intensity. It further gives us the opportunity to visualize the immediate behavior of the field with removed mean flow using line integral convolution (LIC) encoded in the brightness of the colors.

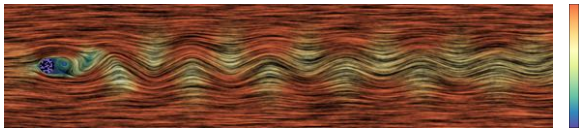
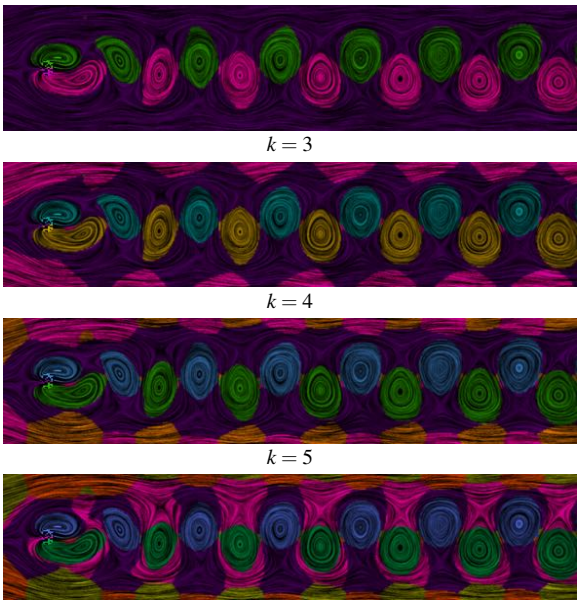


Figure 3: The case study is a von Kármán vortex street. The color map represents the speed of the flow.

The method has some degrees of freedom and we will briefly discuss their influence to the result of the clustering.

The greatest influence on the resulting image is the chosen number of clusters k. The result of the clustering for maximal order $n = 2$, integration radius $r = 1$, and varying k can be found in Figure 4. The repeating structures of the von Kármán vortex street are perfectly represented by the clustering for any k. Further, the clustering by the moments is able to

discriminate flow behavior that is not possible by judging the LIC image. For example, the difference between the orientation of the vortices or the separation of the pink and the purple clusters in the image with $k = 6$. The flow looks similar on first sight, but indeed there is no rotation that can identify the two. They differ by a reflection, with respect to which we did not choose to normalize in this example.



$k = 3$ $k = 4$ $k = 5$ $k = 6$

Figure 4: Influence of k on the clustering. Radius $r = 1$.

The second biggest influence on the output of the algorithm is the radius of the integration area $r \in \mathbb{R}$. This is the scale of the patterns that are compared. Larger integration areas increase the influence of the higher order moments. While the clustering is dominated by the vortices with radius one, the main impact for radius two is made by the monkey saddles. That leads to a stronger decomposition of the area above and below the vortices in Figure 5. The size of the features a user is interested in will vary strongly from application to application. Our algorithm permits the user to cluster all scales simultaneously in one feature space. That way, patterns that occur on different scales are comprehended into the same cluster and colored in the same color throughout all scales. The user can browse through the different scales and follow the development of the clusters. An example of that is shown the supplementary video.

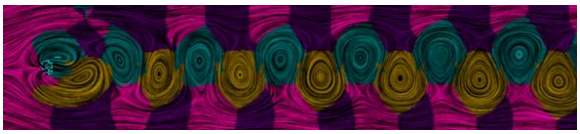
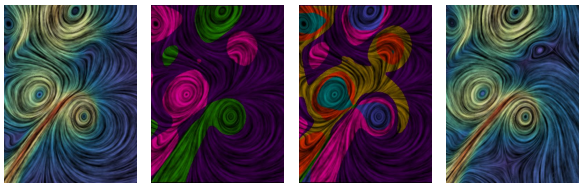


Figure 5: Influence of the radius, $r = 2, k = 4$.

The normalization with respect to translation results in Lagrangian invariance also in time-dependent vector fields. Our von Kármán vortex street is a time-dependent data set. It consists of 32 time slices that periodically form a continuous flow. Our algorithm is capable of comprehending the moments of each time slice into one big feature space. The result is a coherent clustering for time dependent data. The same patterns are assigned the same color independent of the time slice they are in. Please find the result in the video in the supplementary material. The feature space and the resulting clustering also depend on the maximal chosen moment order $n \in \mathbb{N}$ and the number of time steps and scales included into the feature space. For the figures in this section, we used the moments of only one time step and one scale. For the video, we included all time steps and many scales $r = 0.5, 0.6, ..., 2$ respectively. The maximal moment order is always $n = 2$.



Velocity field. $k = 3$. $k = 7$. Different FOR.

Figure 6: The swirling jet and its clustering.

We also study the results of applying the clustering algorithm on a different data set. The flow field in Figure 6 is the result of the simulation of a fast stream entering a liquid at rest. The dominant features are the vortices. We consider especially interesting how the patterns that show a strongly curved behavior but no vortex cores are allocated to the clusters of the vortices, too. The reason is that they are vortices in a different frame of reference, as can be seen in the right image. The moments are Galilean invariant and therefore reveal what a Lagrangian view would reveal.

7. Conclusion

In this paper, we presented a method to segment a vector field by clustering the moment invariants of the field. The moment invariants represent the field in a high dimensional feature space that transfers similar flow patterns into measurable spatial relations. Using normalization, similar flow structures can be elaborated even though they differ in velocity, frame of reference, or orientation. That way, we were able to extract flow features in a vector field without any predefined description in an automatic fashion.

The main drawback of the method is the great number of degrees of freedom. Initialization, integration radius, maximal order, and treatment of the case of a vanishing normalizer change the outcome of the method. In future work, we will investigate these parameters in more detail, compare different clustering algorithms, and approach 3D flow fields.

Acknowledgements

We thank Dr. Mutschke from the TU Dresden for providing the von Kármán vortex street dataset and Prof. Kollmann from the University of California Davis for producing the swirling jet dataset. We would further like to thank the FAnToM development group from the Leipzig University for providing the environment for the visualization of the presented work. This work was partially supported by the European Social Fund (Application No. 100098251).

References

[AMP84] ABU-MOSTAFA Y. S., PSALTIS D.: Recognitive Aspects of Moment Invariants. *IEEE Transactions on Pattern Analysis and Machine Intelligence* (1984), 698–706. 1

[BAAK*13] BEYER J., AL-AWAMI A., KASTHURI N., LICHTMAN J. W., PFISTER H., HADWIGER M.: Connectomeexplorer: Query-guided visual analysis of large volumetric neuroscience data. *IEEE Transactions on Visualization and Computer Graphics 19*, 12 (2013), 2868–2877. 1

[BHSH14a] BUJACK R., HOTZ I., SCHEUERMANN G., HITZER E.: Moment Invariants for 2D Flow Fields via Normalization. In *IEEE Pacific Visualization Symposium (PacificVis)* (2014). 1, 2, 3

[BHSH14b] BUJACK R., HOTZ I., SCHEUERMANN G., HITZER E.: Moment Invariants for 2D Flow Fields via Normalization in Detail. *IEEE Transactions on Visualization and Computer Graphics PP*, 99 (2014), 1–1. 2

[CEH*09] CHEN M., EBERT D., HAGEN H., LARAMEE R. S., VAN LIERE R., MA K.-L., RIBARSKY W., SCHEUERMANN G., SILVER D.: Data, information, and knowledge in visualization. *Computer Graphics and Applications 29*, 1 (2009), 12–19. 1

[DN77] DIRILTEN H., NEWMAN T.: Pattern Matching Under Affine Transformations. *IEEE Transactions on Computers 26*, 3 (1977), 314–317. 1

[Flu00] FLUSSER J.: On the independence of rotation moment invariants. *Pattern Recognition 33*, 9 (2000), 1405–1410. 1

[FZS09] FLUSSER J., ZITOVA B., SUK T.: *Moments and Moment Invariants in Pattern Recognition*. Wiley, 2009. 2

[GPR*00] GARCKE H., PREUÃ§ER T., RUMPF M., TELEA A., WEIKARD U., WIJK J.: A continuous clustering method for vector fields. In *In Proceedings of Visualization âĂŹ00* (2000), IEEE Computer Society Press, pp. 351–358. 2

[Hu62] HU M.-K.: Visual pattern recognition by moment invariants. *IRE Transactions on Information Theory 8*, 2 (1962), 179–187. 1

[KNC*14] KAISER E., NOACK B. R., CORDIER L., SPOHN A., SEGOND M., ABEL M., DAVILLER G., ÖSTH J., KRAJANOVIC S., K. NIVEN R.: Cluster-based reduced-order modelling of a mixing layer. *Journal of Fluid Mechanics 754* (2014), 365–414. 2

[LCL*13] LU K., CHAUDHURI A., LEE T.-Y., SHEN H.-W., WONG P. C.: Exploring vector fields with distribution-based streamline analysis. In *IEEE Pacific Visualization Symposium (PacificVis)* (2013), pp. 257–264. 1

[Mac67] MACQUEEN J.: Some methods for classification and analysis of multivariate observations. In *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Statistics* (Berkeley, Calif., 1967), University of California Press, pp. 281–297. 3

[MVvW05] MOBERTS B., VILANOVA A., VAN WIJK J. J.: Evaluation of fiber clustering methods for diffusion tensor imaging. In *IEEE Visualization* (2005), pp. 65–72. 2

[NMM09] NAM J. E., MAURER M., MUELLER K.: A high-dimensional feature clustering approach to support knowledge-assisted visualization. *Computers & Graphics 33*, 5 (2009), 607–615. 1

[Sch11] SCHLEMMER M.: *Pattern Recognition for Feature Based and Comparative Visualization*. PhD thesis, Universität Kaiserslautern, Germany, 2011. 1

[SHM*07] SCHLEMMER M., HERINGER M., MORR F., HOTZ I., HERING-BERTRAM M., GARTH C., KOLLMANN W., HAMANN B., HAGEN H.: Moment Invariants for the Analysis of 2D Flow Fields. *IEEE Transactions on Visualization and Computer Graphics 13*, 6 (2007), 1743–1750. 1

[Tea80] TEAGUE M. R.: Image analysis via the general theory of moments∗. *Journal of the Optical Society of America 70*, 8 (1980), 920–930. 1

[TN14] THOMAS D. M., NATARAJAN V.: Multiscale symmetry detection in scalar fields by clustering contours. *IEEE Transactions on Visualization and Computer Graphics 20*, 12 (2014), 2427–2436. 1

[TvW99] TELEA A., VAN WIJK J. J.: Simplified representation of vector fields. In *Proceedings of the Conference on Visualization '99: Celebrating Ten Years* (Los Alamitos, CA, USA, 1999), VIS '99, IEEE Computer Society Press, pp. 35–42. 2

[WWYM10] WEI J., WANG C., YU H., MA K.-L.: A sketch-based interface for classifying and visualizing vector fields. In *IEEE Pacific Visualization Symposium (PacificVis)* (2010), IEEE, pp. 129–136. 1