Making Sense of Geometric Data

Cengiz Öztireli
ETH Zurich

Most data acquired from the real world is or can be interpreted as geometric in nature. Advanced and affordable sensors, printers, displays, and the Internet make geometric data increasingly important for many disciplines. Most geometric data comes in the form of unstructured point samples. Giving structure and meaning to this data has been one of the main challenges of computer graphics as well as other fields in the last few decades.

Vast amounts of geometric data are collected in many fields such as medical imaging, robotics, geography, seismology, architecture, and archeology, just to name a few. The data can hence represent many different structures. The datasets are massive—a conventional depth camera with a frame rate of 30 frames per second (fps) can easily generate billions of points in minutes—but the acquired data are far from perfect, with noise, outliers, and missing parts.

My PhD thesis started as an effort to turn this massive amount of data into digitally meaningful representations useful for various applications in computer graphics and beyond. We relied on the observation that the majority of geometric data in computer graphics and many other fields represent object surfaces and repetitive structures. We thus targeted the problems of reconstructing manifold surfaces, which are smooth watertight surfaces bounding objects, and stochastic point patterns that are random distributions of points with certain characteristics, from unstructured point samples.

Editor's Note

Cengiz Öztireli received a Eurographics Annual Award for Best PhD Thesis 2014.

Feature Preserving Robust Reconstructions

Reconstructing a manifold surface from points sampled on the surface is an inherently ill-posed problem in the absence of further assumptions. Thus, typically, a degree of smoothness is assumed for regularizing the problem. However, just as smooth images interfere with our perception of edges, smooth surfaces that lack sharp features and fine details are not always what we expect.

Local Fits for a Global Surface Reconstruction

One way of achieving accurate and efficient reconstructions under the influence of noise and outliers is approximating the surface using moving least squares (MLS) based approximations.[1,2] But these reconstructions inherently smooth out sharp features and fine details. In fact, MLS can also be regarded as a smoothing filter. Inspired by the bilateral filter[3] and its relation to robust statistics, the first part of my thesis deals with extending MLS with robust statistics such that we can get more accurate reconstructions with sharp features in the presence of noise and outliers.

MLS-based implicit surfaces accept points possibly with surface normals as the input and solve a local least-squares system to fit a local surface around each query point. Local kernel regression also operates with the same idea, and we indeed show that many implicit MLS surface definitions are actually local kernel regressions with various constraints.

Robust MLS Approximations

This relation then opens the door to incorporating robust statistics into the definition of MLS surfaces in ways proven to provide accurate results in statistics under the influence of outliers. For surface reconstruction, there are outlier points resulting from corrupted data as well as outliers in the normal space near sharp features.

Published by the IEEE Computer Society

Normals of a smooth patch are actually outliers for the normals of another smooth patch meeting the former at a sharp feature such as an edge or a corner. Building on this observation, we show that it is possible to extend MLS surfaces with robust statistics and formulate a nonlinear fitting that can be solved by iteratively reweighted least squares.

The result is a manifold surface definition that we call *robust implicit MLS* (RIMLS). It preserves sharp features and details under the influence of noise, outliers, and a lack of samples. It also inherits the simple and efficient nature of MLS with pure local computations and a simple mathematical definition with no special cases to handle for any types of features.

Figure 1 illustrates the effects of the spatial and normal robustness terms on the reconstruction of a curve (one-manifold in the plane).[4] The reconstruction with MLS is easily biased by even a single outlier, as the horizontal part of the reconstructed curve shows. Introducing the spatial robustness term solves this problem, but the resulting definition still lacks the sharp corner. The corner can be accurately reconstructed with the normal robustness term, and finally combining the two terms gives a sharp and unbiased reconstruction.

Of course, the samples themselves do not provide the information regarding whether or not there should be sharp features on a surface. The sharpness of the features is thus controlled with a single user-given parameter, as shown for the reconstruction of a cube from just four samples at each face in Figure 2. The features are not limited to edges and simple corners. They can also be where more than three planes meet or where peaks and other fine details exist, as we show in Figure 3. RIMLS better preserves corners than previous MLS-based reconstruction methods such as algebraic point set surfaces (APSS).[5]

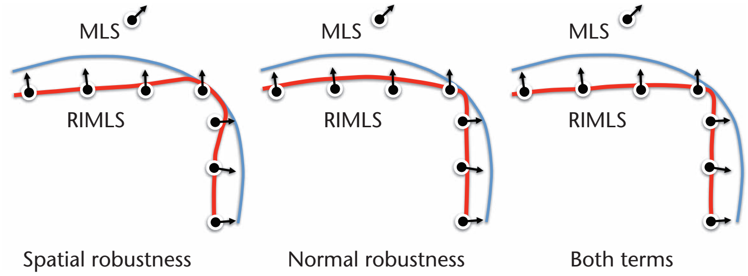

Figure 1. Reconstructed curve using robust implicit moving least squares (RIMLS). Compared with implicit MLS,[4] this process is less influenced by the outlier and preserves the sharp corner if both spatial and normal robustness terms are used.

Sampling for Accurate Reconstructions

Reconstruction is just one side of the story, however. We have experienced long running times for some of the reconstructions with RIMLS because of unnecessarily dense datasets, with millions or billions of points. On the other hand, as well investigated in signal processing, accurate reconstructions necessarily depend on adequate samplings. Without enough data points in a surface region, it is impossible to get an accurate reconstruction using MLS in that part.

In the second part of my thesis, we investigate the sampling problem: what sampling conditions are required to get the surface geometry and topology correct, while preserving sharp features and fine details and avoiding redundancy? For functions, the sampling theorem gives concise conditions to get accurate reconstructions for a regular sampling and accurate functional values. But the data we get does not represent a function. Instead, it is sampled from a manifold surface, and it is typically not regular with respect to a known metric. Hence, it is a much more difficult problem to derive sampling conditions for a particular surface definition.

Spectral Measure for Sampling

The first challenge is defining a measure that tells us the accuracy of reconstructions for a given

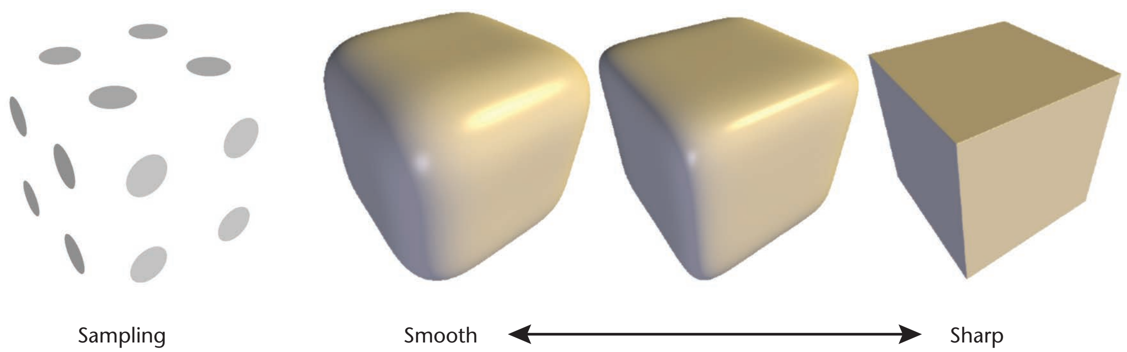

Figure 2. Reconstructed cube from just four samples at each face. The sharpness of the features is controllable by a single user-given parameter.

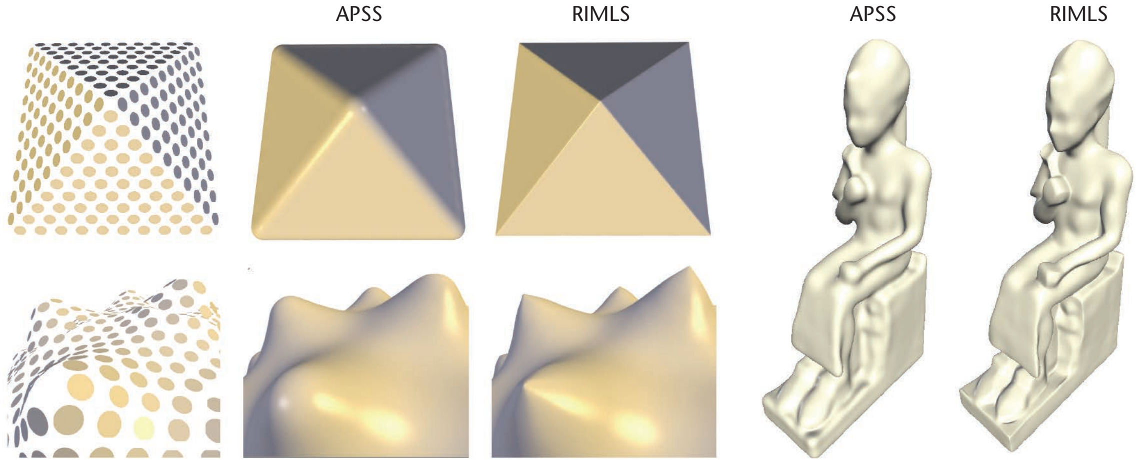

Figure 3. Feature preservation. Corners where more than three planes meet or where peaks and other fine details exist are better preserved with RIMLS than with previous MLS-based reconstruction methods such as algebraic point set surfaces (APSS).[5]

sampling of a surface. Such a measure should only depend on the intrinsic structure of the surface and should be independent of the actual representation.

The Laplace-Beltrami operator and its spectrum have been proven to accurately represent manifolds both in theory and practice for many problems. In fact, it is one of the most fruitful tools in geometric processing. Its spectrum contains a vast amount of information about the surface and satisfies all the required properties we seek. The Laplace-Beltrami operator is a generalization of the Laplace operator. It acts on the functions living on the manifold surface, instead of those defined in an Euclidean space. Intuitively, the eigenvalues correspond to frequencies, and the eigenfunctions correspond to oscillating functions with those frequencies, reminiscent of cosines and sines in the Euclidean space.

We thus define the contribution of a point to the surface definition as the change it makes to the Laplace-Beltrami spectrum of the reconstructed surface. Conceptually, we compute the Laplace-Beltrami spectrum before and after adding or removing one or more points. If the change is not significant, we say the added or removed points are not important for an accurate reconstruction.

Computing the Measure

However, getting from the reconstructed surface definition to the spectrum is challenging. First of all, we need to define which method we use to reconstruct the surface because sampling conditions are meaningful only when the reconstruction method is defined. We use our RIMLS as the basis reconstruction method.

In principle, for a greedy sampling algorithm, we could mesh the surface reconstructed with RIMLS before and after adding a point to the point set defining the surface and then compute the change in the spectrum. But this involves computing the eigen-decompositions of global matrices at each step, which is too expensive. Furthermore, to get accurate samplings, we also need to optimize with respect to this measure. Taking derivatives and designing optimization procedures directly for the pure measure is also difficult.

Instead, we utilize the relation between the Laplace-Beltrami operator and surface's heat kernel and its approximations with kernels in the embedding Euclidean space. The heat kernel defines how heat is transferred between two points on the surface. For the Euclidean space, this is just an isotropic Gaussian. For general manifolds, it can have different forms depending on where we are on the manifold surface. The heat kernel's relations to the Laplace-Beltrami operator and Gaussian kernels have been studied extensively in the machine learning and harmonic analysis literature.[6,7] Utilizing these relations, our derivations finally result in a local measure based on the kernel definition we use to compute RIMLS. This measure involves inverting a local kernel matrix constructed using the neighboring sample points for each query point and is actually equivalent to the distance of the query point to the subspace defined by the neighboring points in the kernel's feature space.

Fast Sampling Algorithms

We utilize the resulting measure in fast and out-of-core sampling algorithms. The algorithms ensure that each sample point contributes maximally and equally to the surface, which leads to a considerable reduction in the number of points needed to represent the surface. The resulting samplings adapt to the surface definition via the kernels used such

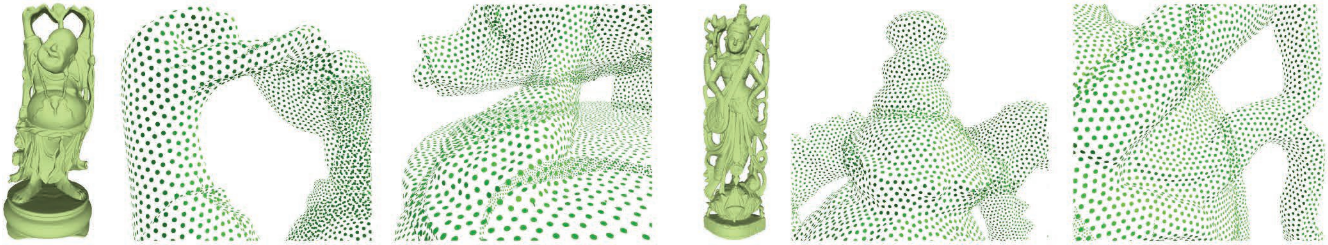

Figure 4. Evenly distributed samples. Because each sample contributes equally and maximally to the surface, we get distributions with blue noise characteristics on the surfaces.

that we get accurate reconstructions while avoiding redundancy, when used together with RIMLS. The samples are also distributed evenly on the surface, as we show in Figure 4, and the sampling adapts to the features depending on the kernel definition (Figure 5). Hence, we can also mesh the points directly without computing the implicit surface for accurate reconstructions, as in Figure 6.

Figure 5. Sampling and kernel definition. By changing the surface and hence the kernel definition, we can also get adaptive samplings (right) to better preserve surface features.

Understanding and Reproducing Patterns in Nature

The point samples do not always represent deterministic structures such as surfaces, as we have assumed so far. Many real-world data actually come from repeated structures and patterns. It is a well-studied fact that nature is fundamentally repetitive at various scales. Examples range from surface textures to the distribution of trees in a forest or the dynamic locations of people in a crowd. These distributions seem random but also exhibit certain characteristics.

We need tools to analyze and synthesize such general point distributions with various characteristics. However, most works in computer graphics only focus on understanding the so-called *blue noise distributions*, where the points have a certain distance between them and are otherwise randomly distributed, similar to the distributions we studied for surface sampling. There is a general need for a deeper theoretical understanding of point patterns.

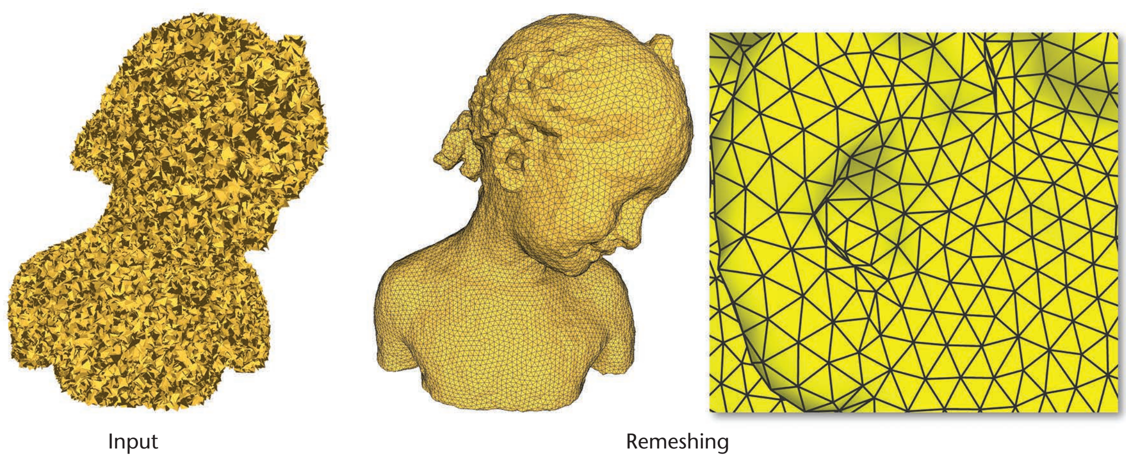

Figure 6. Generated samplings. The points can be meshed directly without computing the implicit surface.

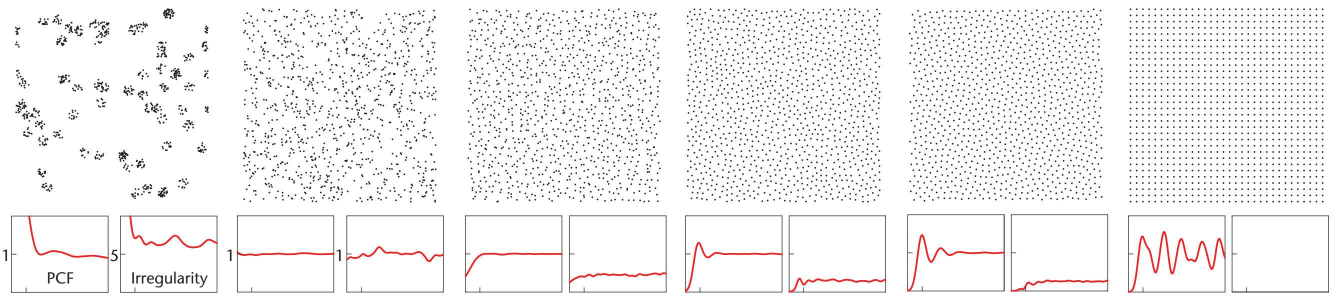

Figure 7. Example pair correlation functions (PCFs). The PCF and irregularity measures provide an intuitive and descriptive characterization of general point distributions.

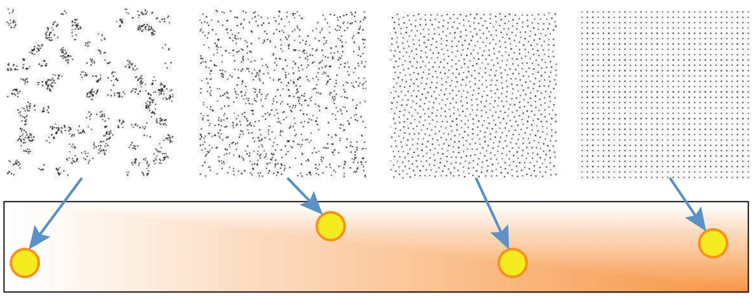

Figure 8. Relations among point distributions. We show that point distributions can be mapped to a space where many distributions live in an essentially two-dimensional subspace.

Point Processes as a Theoretical Basis

In other domains such as physics and spatial statistics, more complex patterns have been heavily investigated, under the discipline called *point processes*.[8] Each distribution can be an instance of an underlying point process, or conversely, a point process can generate infinitely many different point distributions with common characteristics. To understand a distribution's characteristics, we thus need to reconstruct its underlying point process.

In the last part of my thesis, we propose learning a point distribution's underlying point process and using it to analyze the pattern the distribution represents as well as to synthesize new point distributions with the same characteristics.

As the basis of our methods, we utilize first- and second-order product density measures and, in particular, the *pair correlation function* (PCF), which intuitively measures the probability of having a pair of points at particular locations in space, assuming that the point pattern is translation and rotation invariant. The PCF only depends on the distance between two locations in space, and hence it is a one-dimensional function, regardless of the space the points live in.

Analyzing and Relating Point Distributions

We propose estimating a smoothed version of the PCF of a point process from a point distribution(s). Each given distribution can be regarded as a sample generated by an underlying point process, and hence it is normally difficult to compute the characteristics of a point process from a single distribution. Fortunately, the invariance assumption significantly reduces the space of allowable PCFs because the distance between each pair of points can be regarded as a sample from the probability distribution of the distances that the PCF measures.

Figure 7 plots example PCFs extracted from distributions with different characteristics, along with an associated irregularity measure we derived from the PCFs. Intuitively, this measure quantifies how irregular the point distribution is at different distances. As you can observe from the plots, the PCF is actually a normalized measure of how many points are present at certain distances.

We next interpret the PCF as the mean of a distribution specific to a point process in a functional space. In this space, it is possible to establish relations among distributions. We empirically show the interesting property that many distributions in this space live fundamentally in a two-dimensional subspace, which we illustrate in Figure 8. At one end, we have the clustering distributions, which are irregular. As we approach the other end, the distributions become more regular, with the regular grid as the rightmost point in this space. Hence, regularity plays a fundamental role in distinguishing and relating many of the point distributions we encounter.

General Synthesis Algorithm

Finally, we would like to synthesize new distributions with characteristics extracted from other distributions. This means that the resulting synthesis algorithm can mimic any previous algorithm proposed for generating sampling patterns, given a sample distribution generated by that algorithm.

We again use the PCF for this purpose because it is widely accepted to uniquely describe almost all practical isotropic distributions in physics and statistics.[8] Given an example point distribution, the algorithm finds an output distribution with a PCF matching that of the example distribution.

We show some example results obtained with this minimization in Figure 9. For each case, although point locations in the example and output distributions differ, they exhibit the same visual pattern. The synthesis algorithm is also general such that example and output distributions can contain different numbers of points, live in different metric spaces, and belong to multiple classes.

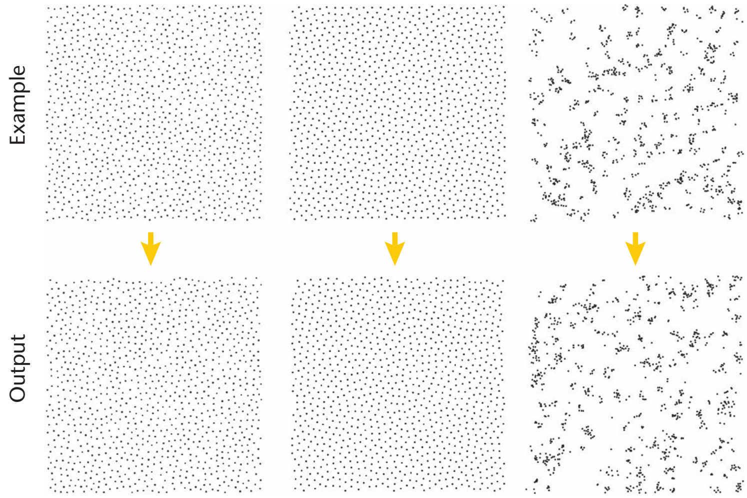

Figure 9. A general synthesis algorithm. We propose a pattern reconstruction algorithm that can mimic any previous point distribution generation algorithms given an example distribution.

Post-thesis Developments

We have been pleased that the ideas and techniques proposed in this thesis have helped us and other researchers with several theoretical and practical developments in computer graphics as well as in other fields. (See https://graphics.ethz.ch/~cengizo/ for more details.)

We strongly believe in making the code and implementations of research works available to the public for the advancement of computer science. We thus integrated RIMLS into the MeshLab software (http://meshlab.sourceforge.net) and provide our sampling codes via our website as well as on demand. This has proven to be useful indeed, as researchers from many fields have been able to use our methods for their own purposes and to test the weaknesses and limits for particular applications.

Applications

In geometry processing, many works have investigated our methods and utilized them for different applications. Our surface sampling and reconstruction methods have been applied to objects with difficult sharp features, multiple parts, or structural constraints. They have been utilized in rendering, filtering geometry, feature extraction, remeshing, alignment of range scans, surface deformations, volume visualization, texturing, model detection in scenes, free viewpoint video, and data fusion from depth and color cameras.

Our algorithms have also been used as base tools for research problems in related fields. In robotics, accurate 3D reconstructions of the scenes the robots live in are becoming increasingly common. Surface reconstruction is thus emerging as an important problem. Several works have experimented with our methods for this problem because of their accuracy and efficiency.

In medical applications, researchers are experimenting with new 3D scanning technologies to make the procedures and visualizations convenient for the doctors and patients while keeping the costs low. Our reconstruction method has been used in several applications in the medical field such as 3D orthodontic dental models and surgical planning.

Our techniques have also been utilized in some works on understanding the motions of cells, visualizations for astrophysics, and preserving cultural heritage.

Exploring New Ideas

Apart from algorithms, the ideas on applying robust methods to MLS, quantifying the change in a manifold using the Laplace-Beltrami spectrum, and analysis and synthesis of distributions with theoretical tools have also been utilized and extended in several ways.

In geometric processing, robust averaging and the idea of a joint space of point positions and attributes have been applied to various interpolation problems. In machine learning and data mining, the spectral measure we propose has been used for manifold learning, classification, and anomaly detection. Several works on sampling have investigated the synthesis of point distributions with general characteristics and how the PCF can be further utilized to analyze distributions in the context of reconstruction and integration of functions, rendering, and discrete and continuous texture synthesis.

Prospects for the Future

In terms of applications, we expect to see an increasing interest in 3D reconstructions as a result of the availability of affordable scanning devices as well as wider adoption of the technology in various fields. We believe our sampling and reconstruction algorithms will continue to be useful for practical systems on object surface generation and scene building. We have recently demonstrated an example of utilizing dynamic data from multiple depth and color cameras for spatiotemporally smooth and accurate reconstructions.[9]

We have also seen increased utilization of robust and feature-preserving techniques for interpolation problems arising in many problems in computer graphics. However, the relations between MLS approximations, local kernel regression, robust methods, sparse methods, and other approximation techniques are still to be explored and exploited further.

Measures based on the Laplace-Beltrami spectrum connect with kernel methods, manifold learning, spectral graph theory, data mining, and geometry processing. Such measures have also been applied to nonmanifold geometries. We hope that our ideas and insights will be useful in further explorations of nonmanifolds and more complex data.

Sampling is an important problem in rendering, image processing, and texture synthesis, where the reconstructed or integrated function should have no noticeable noise and regularity artifacts. So far, density of points has mostly been considered for improving the error and perceptual quality, excluding low discrepancy and blue noise distributions. The theoretical tools we present can be used to explore how correlations among point locations, along with the density, can be adapted to the functions to be represented.

The analysis and synthesis methods for patterns can be extended using marked and space-time point processes, which can be useful in example-based physically based simulations, crowd simulations, and geometry textures.

We hope that the ideas, methods, and algorithms proposed will continue to be useful and are looking forward to new explorations.

References

1. M. Alexa et al., "Computing and Rendering Point Set Surfaces," *IEEE Trans. Visualization and Computer Graphics*, vol. 9, no. 1, 2003, pp. 3–15.
2. D. Levin, "Mesh-Independent Surface Interpolation," *Geometric Modeling for Scientific Visualization*, G. Brunnett et al., eds., Springer, 2004, pp. 37–49.
3. C. Tomasi and R. Manduchi, "Bilateral Filtering for Gray and Color Images," *Proc. 6th Int'l Conf. Computer Vision* (ICCV), 1998, pp. 839–846.
4. K. Ravikrishna, "Provably Good Moving Least Squares," *Proc. ACM-SIAM Symp. Discrete Algorithms*, 2005, pp. 1008–1018.
5. G. Guennebaud and M. Gross, "Algebraic Point Set Surfaces," *ACM Trans. Graphics*, vol. 26, no. 3, 2007, article no. 23.
6. R.R. Coifman, and S. Lafon, "Diffusion Maps," *Applied and Computational Harmonic Analysis*, vol. 21, no. 1, 2006, pp. 5–30.
7. M. Belkin and P. Niyogi, "Convergence of Laplacian Eigenmaps," *Advances in Neural Information Processing Systems 19*, B. Schölkopf, J. Platt, and T. Hofmann, eds., MIT Press, 2006, pp. 129–136.
8. J. Illian et al., *Statistical Analysis and Modelling of Spatial Point Patterns*, John Wiley and Sons, 2008.
9. C. Kuster et al., "Spatio-Temporal Geometry Fusion for Multiple Hybrid Cameras Using Moving Least Squares Surfaces," *Computer Graphics Forum*, vol. 33, no. 2, 2014, pp. 1–10.

Cengiz Öztireli *is a senior research associate in the Computer Graphics Laboratory at ETH Zurich. Contact him at cengizo@inf.ethz.ch.*

Contact department editor Jim Foley at foley@cc.gatech.edu.