Eurographics/ IEEE-VGTC Symposium on Visualization (2007)
Ken Museth, Torsten Möller, and Anders Ynnerman (Editors)

Feature Identification and Extraction in Function Fields

John C. Anderson[1], Luke J. Gosink[1], Mark A. Duchaineau[2], and Kenneth I. Joy[1]

[1]Institute for Data Analysis and Visualization, Department of Computer Science, University of California, Davis
e-mail: {janderson, ljgosink, kijoy}@ucdavis.edu
[2]Center for Applied Scientific Computing, Lawrence Livermore National Laboratory
e-mail: duchaine@llnl.gov

Abstract

We present interactive techniques for identifying and extracting features in function fields. Function fields map points in n-dimensional Euclidean space to 1-dimensional scalar functions. Visual feature identification is accomplished by interactively rendering scalar distance fields, constructed by applying a function-space distance metric over the function field. Combining visual exploration with feature extraction queries, formulated as a set of function-space constraints, facilitates quantitative analysis and annotation. Numerous application domains give rise to function fields. We present results for two-dimensional hyperspectral images, and a simulated time-varying, three-dimensional air quality dataset.

Categories and Subject Descriptors (according to ACM CCS): I.3.6 [Computer Graphics]: Methodology and Techniques

1. Introduction

In scientific visualization, scalar and vector fields are well-studied data types with mature visualization techniques, including isosurfaces [LC87], slicing and volume rendering [BW01], and streamlines [Vol89]. Scalar fields, which map points to scalar values, arise in numerous applications domains, e.g. medical imaging [SFF91]. Vector fields, which map points to vector values, are often used to model flow.

In this paper we are concerned with an under-studied data type: function fields. Function fields map points p in n-dimensional Euclidean space to 1-dimensional scalar functions:

$$F : p \in \mathbb{R}^n \rightarrow f_p \in \mathscr{F}_I, \tag{1}$$

where $\mathscr{F}_I$ is the set of functions defined over a closed interval I. Datasets often store sampled functions as m-dimensional vectors associated with each point. Figure 1 illustrates the structure of two-dimensional (1(a)), and time-varying, three-dimensional (1(b)) function fields.

This paper introduces interactive techniques for identifying and extracting features in function fields. These methods do not rely on application domain-specific knowledge, and do not require expensive data preprocessing.

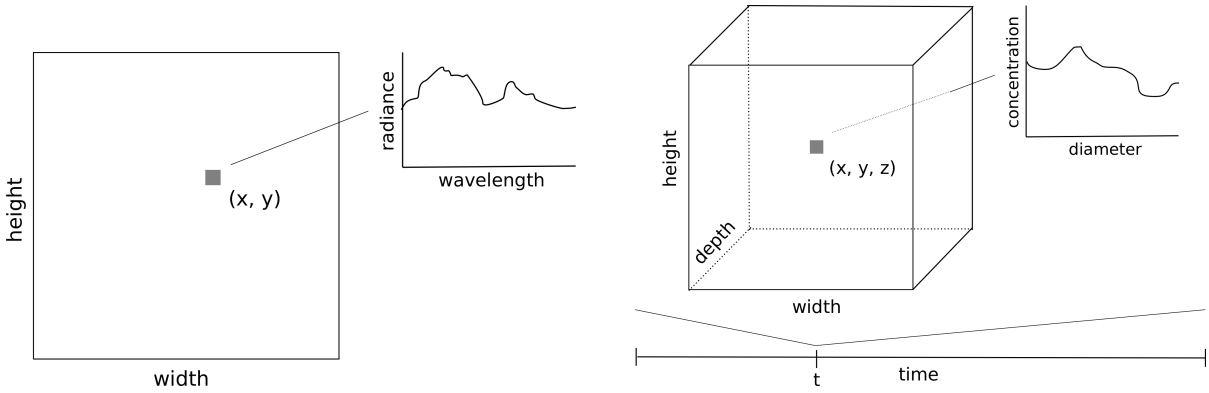
Features in function fields are spatial regions in which the 1-dimensional functions are similar. Our approach to feature identification is to interactively generate scalar distance fields by applying a function-space distance metric over the function field. These scalar fields can be rapidly rendered as images or volumes, within which users are easily able to visually identify features. Feature identification is discussed in Section 4.

In addition to visual exploration, the extraction of features is important for quantitative analysis and annotation (Section 5). Users construct queries as a set of function-space constraints against the field's functions. We demonstrate a number of queries to extract features such as golf courses, water, and areas of high pollution.

2. Datasets

Function fields arise in many application domains. We demonstrate our methods on two distinct function field datasets: hyperspectral images from the domain of remote sensing, and a particulate pollution dataset used in air quality research.

Hyperspectral imaging systems are used in remote sensing for a broad range of applications, including environ-



(a) Hyperspectral Image (b) CRPAQS Dataset

Figure 1: *Graphical overview of the two function field datasets used in this paper. In (a), we show the data layout of hyperspectral images. Hyperspectral images are spatially two-dimensional with pixels that are sampled functions of radiance (or reflectance) versus wavelength. In (b), we show the data layout for the California Regional Particulate Air Quality Study (CRPAQS) dataset. Each cell in this time-varying, three-dimensional function field contains a sampled function of particle concentration versus diameter.*

mental studies and military preparation. The primary benefit of using a hyperspectral imagery system is that each pixel contains data for multiple spectral channels (instead of only grayscale or RGB), thus allowing more in-depth image analysis. The Airborne Visible InfraRed Imaging Spectrometer (AVIRIS) [VGC*93] is aircraft-mounted and acquires calibrated 614x512 images of up-welling spectral radiance. In AVIRIS images, each pixel consists of 224 radiance (or reflectance) samples over visible and short-wave infrared wavelengths. We use two images in this paper, each approximately 270 megabytes: one of Moffett Field and the San Francisco Bay, and another of an area approximately 18 kilometers to the east of Moffett Field.

The second dataset is an air quality simulation from the California Regional Particulate Air Quality Study (CRPAQS). This study is concerned with particulate pollution throughout the San Joaquin Valley, California, U.S.A. The size of aerosol particles is an important factor in their toxicity; thus, each cell contains a sampled function of particle concentration versus diameter. The full dataset contains nine particle types, however we only consider SO_4. The dataset is organized as a 5-dimensional rectilinear grid: a time dimension (25 timesteps), three dimensions spatially (185x185x15), and a particle diameter dimension (9 bins). Despite the low spatial resolution, this dataset is approximately 450 megabytes.

3. Related Work

Function fields in visualization have been under-studied compared to scalar and vector fields. Direct rendering of function fields is difficult. Two-dimensional datasets can be treated as 3D volumes or 2D animations [ESG97]. Three-dimensional datasets become 4D volumes or a time-series of 3D volumes [HAF*96]. Using animation to visualize the extra dimension in function fields is difficult when the dataset itself is time-varying.

Fields with functions representing statistical distributions have been referred to as "distribution fields". Kao et al. [KLDP02, KKL*05] use parametric statistics and shape descriptors to give an overall impression of two-dimensional distribution fields. Luo et al. [LKDP03] transform distribution fields into scalar fields using unary operators upon individual distributions. For example, a three-dimensional distribution field could be visualized by volume rendering the standard deviations of each distribution in the field.

For hyperspectral imagery, each pixel may be colored by integrating the radiance versus wavelength functions with color matching functions. Examples of color matching functions include CIE XYZ, which models the wavelength-dependent response of the human eye [WS00], and spectrally weighted envelopes of Jacobson and Gupta [JG05]. Color matching functions, while useful for hyperspectral imagery, do not generalize well to function fields from other application domains.

Hyperspectral imagery may also be analyzed using a library of measured spectral signatures for various natural and man-made materials. Linear spectral unmixing [SD93] estimates the ratios of materials present in each pixel. Of course, spectral unmixing is domain-specific since it relies upon the spectral behavior of mixed-material regions, as well as the availability of a spectral library.

Principal Component Analysis (PCA) [Jol02] is an ubiquitous dimension reduction technique. For a set of vectors in m-dimensional space, PCA identifies a set of ordered, or-

thonormal basis vectors ("components"). Transforming the data vectors into a space spanned by the first k, $k < m$, of these basis vectors yields a dimension reduction that maximally preserves variance. PCA has been used to display hyperspectral imagery by associating components with color channels to produce color images [TKDO03].

Multidimensional scaling (MDS) [CC00] is often used to embed high-dimensional, non-spatial data samples in a low-dimensional metric space. The goal of the embedding is to create an analogue between similarity in the original data space and distance in the new space (i.e., similar samples are close, dissimilar samples are distant). Once MDS has been performed, the low-dimensional space may be visualized to study the similarity structure of the original data. Spatial data is ill-suited to MDS visualization since the original spatial layout of the data samples is lost.

In this paper, we introduce interactive techniques for identifying and extracting features in function fields. We do not assume that datasets can be explored using a static color mapping or via a single statistic. Instead, users interactively direct the generation of scalar distance fields, within which features may be visually identified. In addition to a distance-based visualization approach, we describe how features may be extracted using queries. The combination of visual exploration and feature extraction creates a powerful framework for quantitative analysis and annotation.

4. Feature Identification

In this paper, we consider function field "features" to be spatial regions in which the 1-dimensional functions are similar. This is a common definition of features in unsupervised classification (see [JMF99]), and proves useful in many application domains.

Take a body of water in a hyperspectral image. Pixels of water will form regions, and their spectral signatures will be largely similar (often interpreted by our eyes as greenish-blue). The same definition of features applies in the CRPAQS dataset. Pollution is created in cities, factories, etc., and transported by winds or other diffusive factors. Thus, it is reasonable to expect functional similarity in particulate pollution features.

Our approach is to produce visualizations that support the feature identification process by highlighting similarities and differences within function fields.

We begin by defining a distance metric D, representing the *dissimilarity* between two 1-dimensional functions, a and b, each with m samples. Given a weight $w_i \in [0,1]$ for each sample, $i = 1,\ldots,m$, we use a weighted Euclidean distance metric:

$$D(a,b) = \left(\sum_{i=1}^{m} w_i \, (a_i - b_i)^2 \right)^{\frac{1}{2}}. \qquad (2)$$

The sample weights are modifiable, providing users with a degree of control over the distance metric.

From an n-dimensional function field F, the above metric is used to generate an n-dimensional scalar field S. The scalar value at point p is the distance in function-space between f_p and the 1-dimensional function f_c at a user-positioned probe $c \in \mathbb{R}^n$:

$$S : p \in \mathbb{R}^n \rightarrow D(f_p, f_c). \qquad (3)$$

The probe is directly positioned in two dimensions using a mouse. In three dimensions, the probe is positioned using a "full space cursor" [NDRO87].

Prior to visualization, the field S is normalized such that its values are in the range $[0,1]$. Traditional rendering techniques may be applied to S since it is a scalar, rather than function, field: two-dimensional function fields become images, while three-dimensional fields may be volume rendered.

We use weighted Euclidean distance due to its efficient evaluation, despite the availability of distance metrics tailored to sampled functions (e.g., Earth Mover's Distance [RTG98], and Chang's spectral distance metrics [Cha00]). This efficiency allows users to reposition the probe, modify the sample weights, and generate new distance field renderings interactively for moderately-sized datasets.

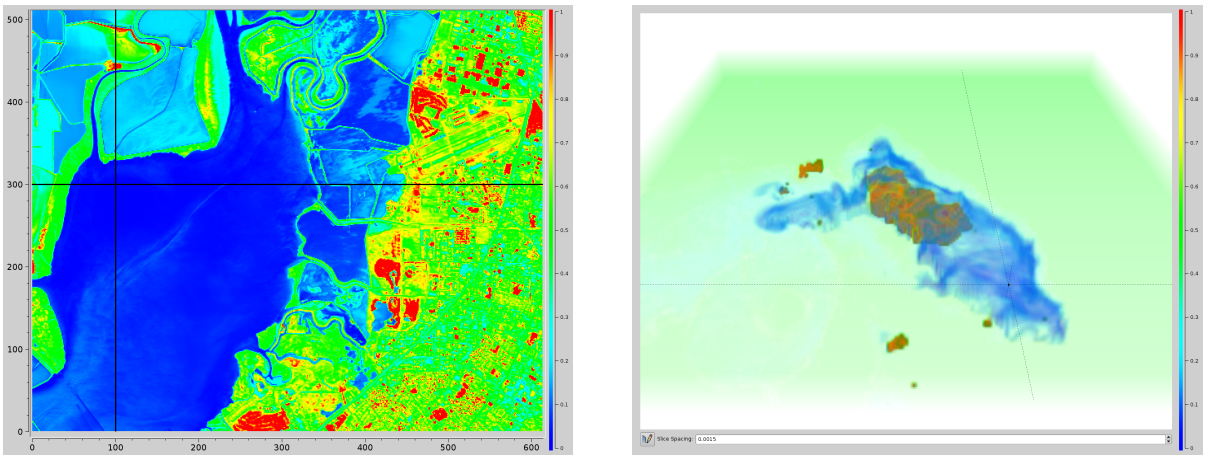
Figure 2 shows distance field renderings for feature identification within two- and three-dimensional function fields. In 2(a), the user has positioned the probe over water in a hyperspectral image. As expected, other water pixels have low distance values (blue) since their functions are "close" in function-space to the probe function. In 2(b), the distance field of the CRPAQS dataset shows both low (blue) and high (red) distance regions located over the central San Joaquin Valley, California, U.S.A. In the high distance region, medium-sized SO_4 particles have higher concentration than at the probe.

4.1. Collisions

Generating and rendering distance fields enables visual feature identification. "Collisions" can occur when dissimilar functions map to similar distances with respect to the probe. Consider Figure 2(a); features with functions mapping to high distances (red) include certain types of buildings and golf courses. The real similarity between these functions is that they are *dissimilar* to water.

Collisions are not unique to our mapping from functions to scalar values; for example, they also occur in [TKDO03, JG05, LKDP03] as described in Section 3. Unlike other projections, however, our method allows collisions to be resolved easily, and interactively, by the user.

Collisions may be resolved by changing the probe location, and thus the function to which all other functions are



(a) Hyperspectral Image (b) CRPAQS Dataset

Figure 2: *Distance field renderings for feature identification within two- and three-dimensional function fields. In (a), we show a distance field for a hyperspectral image. The probe is positioned over water. As expected, other water pixels have low distance values (blue) due to functional similarity. Features with functions mapping to high distances (red) include certain types of buildings and golf courses. In (b), we show a distance field of the CRPAQS dataset. The dark blue and red regions are located over the central San Joaquin Valley, California, U.S.A.*

compared. Alternately, the user may modify the distance metric by changing the sample weights w_i, $i = 1,\ldots,m$. In addition to resolving collisions, sample weights may be modified for illustrative visualizations of function fields – it is simple to emphasize similarities or play down differences between functions (and *vice versa*).

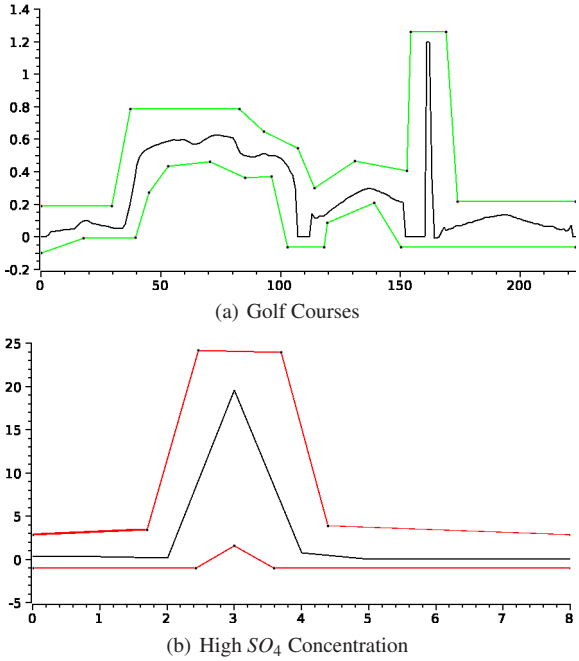
5. Feature Extraction

Combining visual exploration with feature extraction opens the door to performing quantitative analysis, such as calculating the size of a body of water or determining how long a pollution source remains active. It also makes it simple to annotate function fields with overlays.

Features in function fields are spatial regions in which the 1-dimensional functions are similar. While it is tempting to use one of the distance fields generated during exploration for feature extraction, user interaction might be needed to resolve collisions. In order to make feature extraction robust and reusable across multiple datasets we perform extraction in function-space.

We define a feature query as a set of constraints over the closed interval I. For a dataset with m samples per function, these constraints take the form of minimum-maximum intervals Q_i for each sample, $i = 1,\ldots,m$. A point p in the dataset with 1-dimensional function f_p, is part of the feature if and only if $f_{p_i} \in Q_i$ for $i = 1,\ldots,m$.

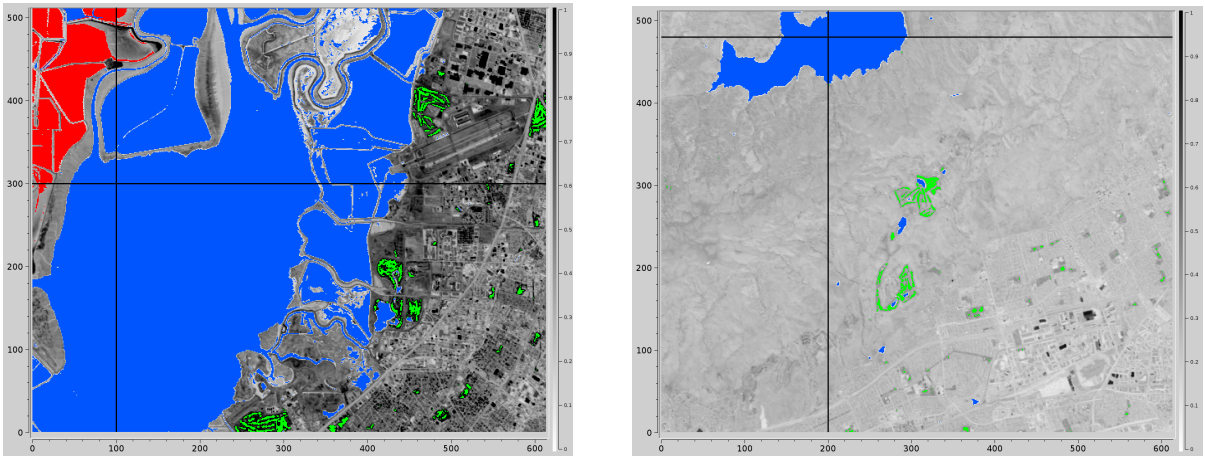
In our system, users first explore function fields using distance field renderings. Once a feature has been identified,



(a) Golf Courses

(b) High SO_4 Concentration

Figure 3: *In (a), we show the function-space constraints (green) used to extract golf courses from hyperspectral images (Figure 4). In (b), we show the constraints (red) used to extract regions from the CRPAQS dataset in which medium-sized SO_4 particles have high concentration (Figure 5). In both, the black curves are functions that satisfy the feature queries.*



(a) (b)

Figure 4: *Hyperspectral images annotated with overlays produced by three queries: golf courses as shown in Figure 3(a) (green), water (blue), and evaporation ponds containing brine shrimp (red). The feature queries were constructed by a user exploring the hyperspectral image of Moffett Field and the San Francisco Bay in (a). The image in (b) shows an area approximately 18 kilometers to the east of Moffett Field; golf courses and water were extracted using the pre-constructed queries without modification.*

the user is able to sketch a pair of curves that define the feature query's minimum and maximum constraints. For example, Figure 3(a) shows the constraint curves, in green, used to extract golf courses from hyperspectral images; the black curve plots the function of a pixel from a golf course. Figure 4 shows two hyperspectral images annotated with overlays produced by three queries: golf courses (green), water (blue), and evaporation ponds containing brine shrimp (red).

Defining features as a set of constraints in function-space makes queries reusable across multiple datasets. The queries used to extract features in Figure 4 were constructed by a user exploring the hyperspectral image of Moffett Field and the San Francisco Bay in 4(a). The image in 4(b) shows an area approximately 18 kilometers to the east of Moffett Field; golf courses and water were extracted using the pre-constructed queries without modification.

Feature queries work on datasets of arbitrary spatial dimension, and upon time-varying datasets. Figure 3(b) shows a simple query that can be used to extract regions from the CRPAQS dataset in which medium-sized SO_4 particles have high concentration. In time-varying datasets it is often possible to reuse a query across multiple timesteps. Figure 5 shows distance field renderings and the regions extracted by the aforementioned query for timesteps 0, 5, 10, 15, and 20. For clarity we only show ground layer images from the three-dimensional results.

6. Implementation & Performance

The datasets used in this paper were previously described in Section 2. The first dataset contains multiple AVIRIS hyperspectral images of Moffett Field and the San Francisco Bay area. The second function field dataset is an air quality simulation from the California Regional Particulate Air Quality Study (CRPAQS).

We have tested our methods on a 2.6 Ghz Mobile Pentium 4-M laptop with 1.0 Gb RAM and a nVidia GeForce 4200 Go graphics card. Figure 6 shows part of our software system. The upper plot shows the probe function in black, and the constraint curves in red defining the query that extracts evaporation ponds containing brine shrimp from hyperspectral images. The lower plot shows the sample weights curve. The minimum-maximum constraint curves and sample weights curve are modifiable by the user; control points can be added, removed, and manipulated. The right side of the interface provides more controls for feature queries.

Since our approach to feature identification is user-driven, moving the probe location, changing sample weights, and generating the resulting distance fields must be interactive. We have used Single Instruction, Multiple Data (SIMD) instructions (e.g., SSE2 for Intel processors) to vectorize the code for the generation of distance fields (Equations 2 and 3). Table 1 shows timing results for distance field generation. For hyperspectral images, distance fields are generated at the rate of approximately 6 per second; for the CRPAQS dataset, approximately 25 per second. Generating distance

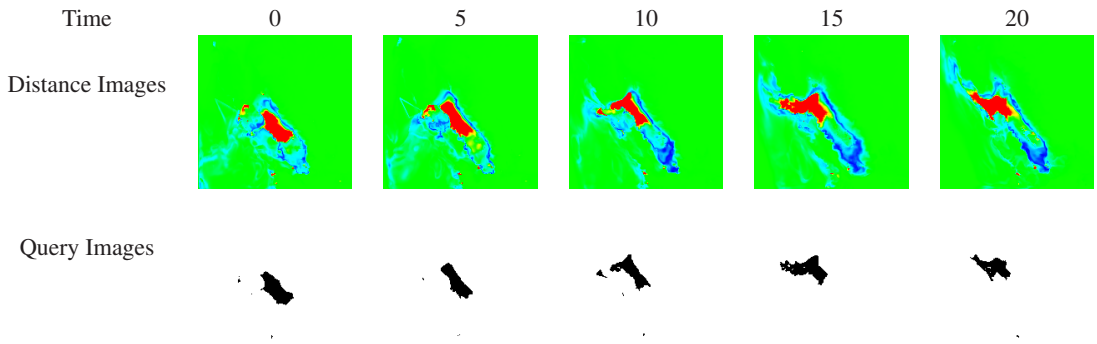


Figure 5: *Distance field renderings generated from the CRPAQS dataset, and the results of using the feature query shown in Figure 3(b) to extract regions in which medium-sized SO_4 particles have high concentration. For clarity we only show ground layer images from the three-dimensional results.*

fields for all 25 timesteps in the CRPAQS dataset can be performed in about 1 second.

Feature queries also evaluate rapidly in our system, thus allowing users to interactively change function-space constraints. Table 1 shows timing results and coverage for query evaluation. In all function field datasets, multiple queries may be evaluated per second. Coverage, the percentage of total cells returned by a query, is an example of quantitative analysis facilitated by feature extraction.

7. Discussion

Function fields are a relatively new data type in scientific visualization. We have presented methods for identifying and extracting features in function fields. Our methods are interactive, and have been useful for exploring, annotating, and performing quantitative analysis on function fields from multiple application domains.

As with any research, however, current limitations motivate future work. Some particularly salient topics include:

- Features are assumed to be spatial regions of similar functions. This definition is general, and applicable to many application domains, however it is not universal. It is easy to imagine applications in which features are not regions, but points, surfaces, or something else entirely.
- No optimizations have been presented for maintaining interactivity while working with large function fields.
- The construction of feature queries involves some trial and error. As a first approximation, users typically form the constraint curves into a rough envelope around an exemplar function. Further adjustments to the constraint curves allow the user to fine-tune the query.

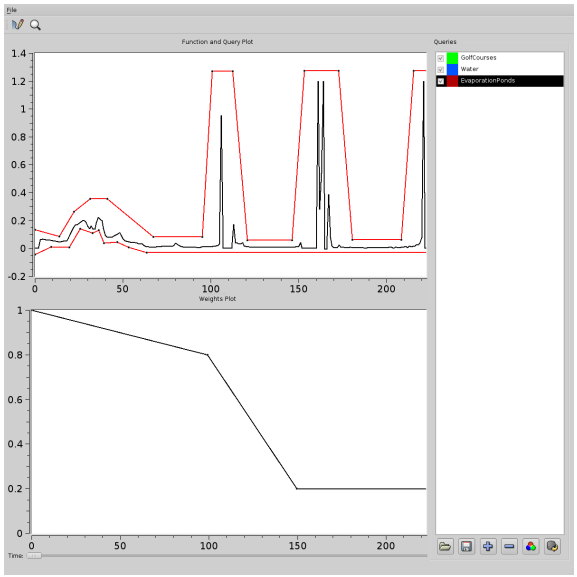


Figure 6: *Part of our software system. The upper plot shows the probe function in black, and the constraint curves in red defining the query that extracts evaporation ponds containing brine shrimp from hyperspectral images. The lower plot shows the sample weights curve. The minimum-maximum constraint curves and sample weights curve are modifiable by the user; control points can be added, removed, and manipulated. The right side of the interface provides more controls for feature queries.*

Dataset	Distance Field Generation (ms)	Query	Execution (ms)	Coverage
Hyperspectral Image	165	Water	320	46.11 %
		Golf Courses	218	0.82 %
		Evaporation Ponds	257	3.66 %
CRPAQS	40	High SO_4 Concentration	23	0.85 %

Table 1: *Timings and coverage. For hyperspectral images, distance fields are generated at the rate of approximately 6 per second; for the CRPAQS dataset, approximately 25 per second. Regions extracted by the feature queries are shown in Figures 4(a) and 5.*

Acknowledgments

This work was supported by Lawrence Livermore National Laboratory under a Student Employee Graduate Research Fellowship, and by Lawrence Berkeley National Laboratory. We would like to thank colleagues in the Visualization and Graphics Group of the Institute for Data Analysis and Visualization (IDAV) at UC Davis for their support during the course of this work.

AVIRIS hyperspectral images are available from http://aviris.jpl.nasa.gov/; the authors would like to thank the NASA Jet Propulsion Laboratory for making select datasets freely available. We would also like to thank Anthony S. Wexler from the UC Davis Air Quality Research Center for providing the CRPAQS dataset.

References

[BW01] BRODLIE K., WOOD J.: Recent advances in volume visualization. *Computer Graphics Forum 20*, 2 (2001), 125–148.

[CC00] COX T. F., COX M. A. A.: *Multidimensional Scaling*, second ed. Chapman & Hall/CRC, Sept. 2000.

[Cha00] CHANG C.-I.: An information-theoretic approach to spectral variability, similarity, and discrimination for hyperspectral image analysis. *IEEE Transactions on Information Theory 46*, 5 (2000), 1927–1932.

[ESG97] EHLSCHLAEGER C. R., SHORTRIDGE A. M., GOODCHILD M. F.: Visualizing spatial data uncertainty using animation. *Computational Geosciences 23*, 4 (1997), 387–395.

[HAF*96] HIBBARD W. L., ANDERSON J., FOSTER I., PAUL B. E., JACOB R., SCHAFER C., TYREE M. K.: Exploring coupled atmosphere-ocean models using Vis5D. *The International Journal of Supercomputer Applications and High Performance Computing 10*, 2/3 (Summer/Fall 1996), 211–222.

[JG05] JACOBSON N., GUPTA M.: Design goals and solutions for display of hyperspectral images. In *IEEE Image Processing* (Sept. 2005), vol. 2, pp. 622–625.

[JMF99] JAIN A. K., MURTY M. N., FLYNN P. J.: Data clustering: a review. *ACM Comput. Surv. 31*, 3 (1999), 264–323.

[Jol02] JOLLIFFE I. T.: *Principal Component Analysis*, second ed. Springer, Oct. 2002.

[KKL*05] KAO D., KRAMER M., LOVE A., DUNGAN J., PANG A.: Visualizing distributions from multi-return lidar data to understand forest structure. *Cartographic Journal, Special Issue on GeoVisualization 42*, 1 (June 2005), 1–14.

[KLDP02] KAO D., LUO A., DUNGAN J. L., PANG A.: Visualizing spatially varying distribution data. *Information Visualization* (2002), 219–226.

[LC87] LORENSEN W. E., CLINE H. E.: Marching cubes: A high resolution 3D surface construction algorithm. In *Proc. SIGGRAPH* (1987), pp. 163–169.

[LKDP03] LUO A., KAO D., DUNGAN J., PANG A.: Visualizing spatial distribution data sets. In *Proceedings of the Symposium on Data Visualisation* (Grenoble, France, 2003), G.-P. Bonneau, S. Hahmann C. D. H., (Ed.), Eurographics Association, pp. 29–38.

[NDRO87] NIELSON G. M., DAN R. OLSEN J.: Direct manipulation techniques for 3d objects using 2d locator devices. In *SI3D '86: Proceedings of the 1986 Workshop on Interactive 3D Graphics* (New York, NY, USA, 1987), ACM Press, pp. 175–182.

[RTG98] RUBNER Y., TOMASI C., GUIBAS L. J.: A metric for distributions with applications to image databases. In *ICCV '98: Proceedings of the Sixth International Conference on Computer Vision* (Washington, DC, USA, 1998), IEEE Computer Society, pp. 59–66.

[SD93] SETTLE J., DRAKE N.: Linear mixing and the estimation of ground cover proportions. *International Journal of Remote Sensing 14*, 6 (1993), 1159–1177.

[SFF91] STYTZ M. R., FRIEDER G., FRIEDER O.: Three-dimensional medical imaging: algorithms and computer systems. *ACM Comput. Surv. 23*, 4 (1991), 421–499.

[TKDO03] TYO J. S., KONSOLAKIS A., DIERSEN D. I., OLSEN R. C.: Principal-components-based display strategy for spectral imagery. *IEEE Trans. Geoscience and Remote Sensing 41*, 3 (Mar. 2003), 708–718.

[VGC*93] VANE G., GREEN R., CHRIEN T., ENMARK H., HANSEN E., PORTER W.: The airborne visible infrared imaging spectrometer. In *Remote Sens. Environ.* (1993), vol. 44, pp. 127–143.

[Vol89] VOLPE G.: *Streamlines and Streamribbons in Aerodynamics*. Tech. Rep. 89-0140, AIAA 27th Aerospace Science Meeting, Jan. 1989.

[WS00] WYSZECKI G., STILES W. S.: *Color Science: Concepts and Methods, Quantitative Data and Formulae*, second ed. Wiley, 2000.