SignatureSpace: A Multidimensional, Exploratory Approach for the Analysis of Volume Data

Kith Pradhan[1], Dirk Bartz, Klaus Mueller[1]
[1] Center for Visual Computing
Stony Brook University, Stony Brook, NY, USA

WSI-2005-11
January 2005

Visual Computing for Medicine
Graphisch-Interaktive Systeme
Wilhelm-Schickard-Institut
Universität Tübingen
D-72076 Tübingen, Germany
e-mail: bartz@gris.uni-tuebingen.de
WWW: http://www.gris.uni-tuebingen.de/areas/vcm

ISSN 0946-3852

ABSTRACT

The analysis of volumetric data is a crucial part in the visualization pipeline, since it determines the features in a volume dataset and henceforth, also its rendering parameters. Unfortunately, volume analysis can also be a very tedious and difficult challenge.

To cope with this challenge, this paper describes a novel information visualization driven, explorative approach that allows users to perform an analysis in a comprehensive fashion. From the original *data volume*, a variety of auxiliary data volumes, the *signature volumes*, are computed, which are based on intensity, gradients, and various other statistical metrics. Each of these signatures (or signatures in short) is then unified into a multi-dimensional *signature space* to create a comprehensive scope for the analysis. A mosaic of visualization techniques ranging from parallel coordinates, to colormaps and opacity modulation, is available to provide insight into the structure and feature distribution of the volume dataset, and thus enables a specification of complex multi-dimensional transfer functions and segmentations.

CR Categories: I.3.3 [Picture/Image Generation]: Display algorithms; I.3.6 [Methodology and Techniques]: Graphics data structures and data types; Interaction techniques I.3.7 [Three-Dimensional Graphics and Realism]: Color, shading, shadowing, and texture; I.4.6 [Segmentation]: Edge and feature detection; Partitioning; I.4.7 [Feature Measurement]: Feature representation;

Keywords: Volume Data, Parallel Coordinates, Multi-Dimensional Visualization

1 INTRODUCTION

Volumetric data is a data concept for discrete data that is arranged on a grid structure and which can be from numerous sources. Most frequently, volumetric data is generated by 3D scanning devices (usually medical 3D scanners such as CT scanners, MRI scanners, etc.) or computed by simulations in physics, chemistry, biology, material sciences, or other engineering fields. If sufficient knowledge on the data is available, this information can be used to derive feature extraction parameters, segmentations, or the specification of transfer functions. Unfortunately, this process is not straightforward and often requires a time-consuming trial-and-error approach [18].

In most (semi-) automatic approaches to generate transfer functions, gradients of the respective voxels are computed and those with maximum gradient values (and the second derivatives of the voxel values) are considered as points of transition between two different materials [1, 12]. These approaches are basically motivated by a material boundary model in which the gradient magnitude value – usually implemented as the absolute value of the central difference at local voxels – is (almost) zero along the boundary and large across the boundary. Another effect that interferes with voxels along the boundaries is the partial volume effect in medical scanned data, which at the same time ensures a smooth gradient [9].

Bajaj et al. used next to gradients also the length and surface of a contour (isosurface), and the included volume confined by the isosurface into the contour spectrum. A super-imposed contour tree was also providing a more global view of changes in the contour spectrum [1]. In contrast to Kindlmann and Durkin [12], they did not assume a specific material boundary model, since it is more of an exploratory tool. In that sense, their approach is somewhat similar to our approach (Sections 2 and 3). Fujishiro et al. [4] use a Reeb-graph data-structure to model the topology of *critical isosurfaces* similar to the contour tree. Pekar et al. [17] also used a similar approach, where they combined gradient, Laplace-filtered voxel values (essentially using the second derivative), volume and surface area confined by the potential isosurfaces into the total gradient curve, and – by dividing the total gradient by the surface area – into the mean gradient curve.

Tenginakai et al. proposed local higher order moments and derived metrics such as skew and kurtosis to mark areas of material transition for the specification of isovalues [20, 21]. Again, the authors assumed a local material boundary model that no more than two materials interface with each other in a $w \times w \times w$ neighborhood.

He et al. took a stochastic/interaction-based approach to generate transfer functions. Piecewise function patterns were combined by a stochastic (genetic) algorithm and the results are evaluated by a human observer to guide the stochastic search process [8]. Another approach that combined interaction with a stochastic algorithm was proposed by Tzeng et al., [23] were interactive painting operations are used to mark areas of interest, or areas to be omitted from the final specification. An artificial neural-network algorithm is used as machine-learning environment to derive rules which voxels should be included. In their paper, they also consider a multidimensional classification space, where they look into the voxel value, the gradient magnitude, the neighboring voxel values, and their position in the volume dataset through the neural-network.

Most methods aiming at the (semi-) automatic transfer function generation deal with some sort of gradient magnitude analysis and often also take into account a Laplacian or the second derivative of the voxel values. However, there are situations where gradient-based methods break down, because they do not depict the material interface alone. These situations are similar to settings where the intensity alone does not specify the materials precisely enough, eg., in the tooth dataset of the transfer-function bake-off [18], or when there are problems due to noise sensitivity.

Our observation is that most classification methods can be applied to a variety of datasets from different sources, but they rarely succeed with all datasets, since they usually focus on a few metrics. In contrast, our approach takes into account a manifold of signatures – which we call signatures –, such as intensity, multiple derived (eg., gradient, second derivatives, etc), and statistical metrics (higher order moments), and derived statistical quantities (skew and kurtosis) [20]. We also inspect the directions along the gradients, providing some kind of history of the local voxels[1]. Furthermore, any meaningful signature that has not (yet) been integrated into our variety of signatures, can easily be added, since none is specifically favored.

All signatures are presented in a modified parallel coordinate display, which provides a framework to identify meaningful signatures for a specific dataset. In that sense, we combine methods from information visualization and scientific (volume) visualization into the *SignatureSpace*, a notion that has been introduced with the SimVis system of Doleisch et al. [2]. The WEAVE system is another system, published even earlier, that combines a 4D visualization of a heart simulation with parallel coordinates and scatter plots [5]. Also a somewhat related approach has been presented by Tory et al. [22], where parallel coordinates are used to represent visualization parameters, such as light parameters, viewing, and transfer functions.

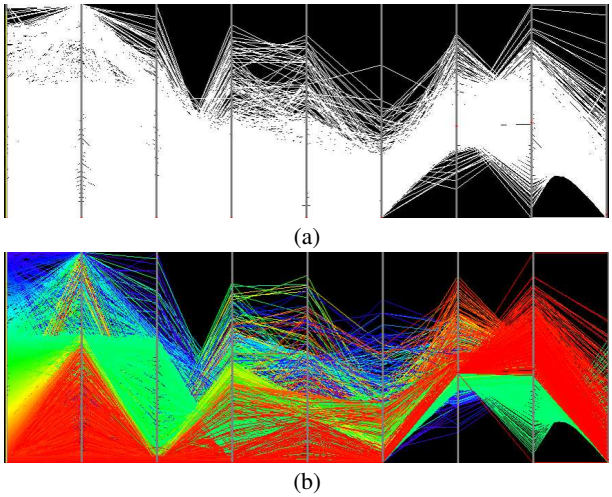
The remaining paper is organized as follows. We start by presenting our modified parallel coordinate-based display, which is in particular suited for a large number of data samples – a common drawback of parallel coordinates – in Section 2. Section 3 focuses on the different signatures that we are using for the data analysis

[1] A similar metric has also been explored in parallel by Lum and Ma [13].

and we discuss results in Section 4. Finally, we draw conclusions and discuss future work in Section 5.

2 Parallel Coordinates for Volumetric Datasets

Parallel coordinates were introduced into data visualization by Inselberg et al. [10, 11]. In contrast to traditional coordinate systems like the cartesian coordinate system, the data axes are arranged parallel to each other, enabling a representation for many more dimensions than 3D and 4D cartesian systems. Data values are plotted as polylines over the various coordinate axes, depending on their respective values. However, parallel coordinates are inflicted by two major drawbacks. First, correlations between different dimensions can only be easily detected by an observer, if the respective data axes are located next to each other. Second, while the number of dimensions can be quite high, parallel coordinates are not very well suited for a large number of data samples, since they tend to obscure the view (Fig.1a).



(a)

(b)

Figure 1: Parallel coordinates: (a) A parallel coordinate display is quickly obstructed if too many samples are visualized. (b) A colormap (on the left most intensity axis) can maintain the structure of the data.

A number of techniques have been proposed to address these problems, such as hierarchical parallel coordinates [3] and brushing functionality [14]. Most important in our view is the application of colormaps to a specific data axis (Fig.1b), a technique which is quite well established in Information Visualization [6] and respective systems, such as the XmdvTool [24], but that we could not trace back to a specific research paper.

In particular Fua et al. [3] combined the hierarchical display of data clusters with colormaps. The samples of the child dimensions in the nested cluster tree are rendered in a *variable-width opacity band*, where the opacity is decreasing the farther out of the cluster center (or the parent node in the tree) the data item is located. Miller and Wegman proposed density plots for modeling line opacity (alternatively line color saturation) to overcome display obstruction [15].

Applying Colors

Colormaps are frequently used with parallel coordinates to highlight structures in dense datasets. In our *SignatureSpace*, they can be applied to every dimension to generate a visual structure for the voxel polylines in the parallel coordinate display. Furthermore, we can limit the application of the colormap to a specific subrange of a signature dimension, effectively skipping the non-selected ranges from the colored display; they remain in their previously acquired color, which is by default white (on a black background). This range limitation for the colormapping essentially implements a brushing operation.

The question what colormap should be used is not easy to answer. In many applications a so called rainbow colormap is used [3], where the full hue range of the HSV color model is mapped to a selected color range. Unfortunately, hue is not a linearly perceived range, which can lead to interpretation problems of accordingly colored data [19]. Better are isomorphic colormaps, such as saturation or luminance maps. However, in *SignatureSpace*, we only use the colormapping to reveal how neighboring data samples evolve over the other signatures and the clear drawbacks of a rainbow/hue colormap do not really have a significant impact here. Hence, we still use the rainbow/hue colormap, but technically, any other colormap could be used as well.

In order to take more than one signature dimension into account for the coloring, we can apply a color to another, *secondary signature* dimension range. The colors of the colormap from the *primary signatures* are then blended with that color. Alternatively, other *secondary signatures* could modulate *Saturation* and *Value* in the HSV color model.

Modulating the Opacity

In order to modify our non-hierarchical version of parallel coordinates for a large number of data samples, we modulate the opacity of the voxel representation in two different ways. In the first option, *Uniform Opacity Modulation*, all data samples are assigned the same opacity value and are blended in our parallel coordinate display. Since the opacity value is controlled through a slider, the user can dynamically adjust the opacity value to the number of data samples in the volume dataset. This way, the large number of data samples (voxels of a volume dataset) – which can easily surpass millions of samples in even small datasets – does not obstruct the display (Fig.2ac).

The second option, *Weighted Opacity Modulation*, enables varying opacities of the voxels projection on all the different signature axes. If a large number of voxel polylines intersect an axis on the same value (actually on the same discrete value bucket), the opacity at that point should be higher than for a smaller number of voxel polylines. We achieve this goal by sorting all voxel values for a signature into n buckets and assign opacity values based on the number of voxels in each bucket. The opacity on the section of the polyline between two neighboring signature axes are linearly interpolated between the respective opacity values on the axes. This way, potential clusters on the signature axes are emphasized over more sparse regions (see Fig.2bd).

During the exploration of the *SignatureSpace*, we examine only one volume slice at the time, resulting in approximately 65536 interactively rendered and blended polylines at a time for our tested datasets. The more complex full volume sample exploration depends on the dataset size takes about 50 seconds. Therefore, all changes to of the overall opacity are performed interactively on slice level, and within seconds for the full volume.

Background Noise Removal

In contrast to the typical database query orientation of information visualization oriented data, typical volumetric data on regular grids have data samples for every grid position. Consequently, many of these data samples do not represent interesting information. Very often, they are only background noise or simply represent *empty* regions. Nevertheless, these irrelevant data samples can block the

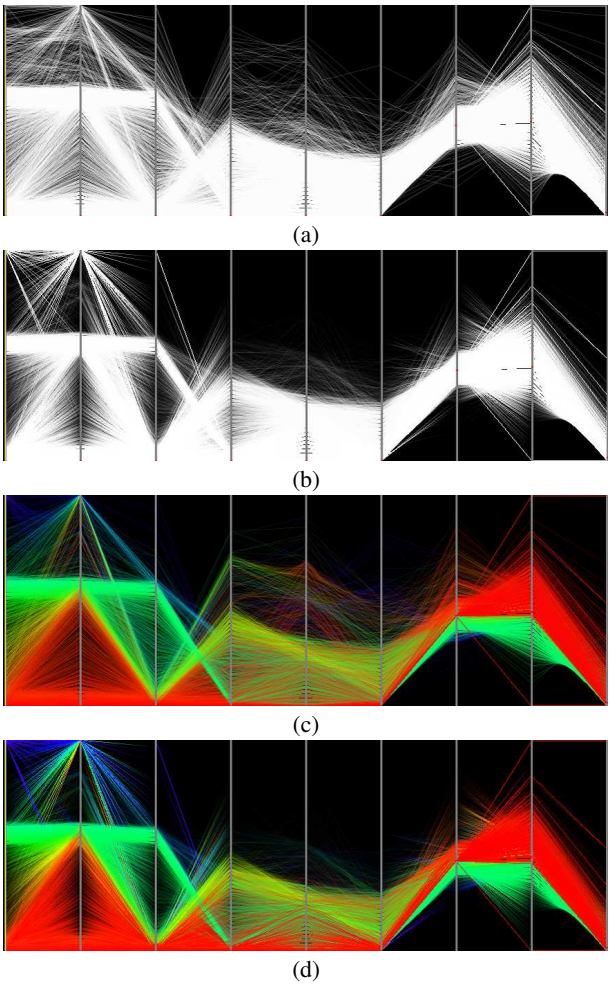


Figure 2: Unfolding clusters with opacity modulation: (a) shows an all-white coloring with Uniform Opacity Modulation (UOM), and (b) shows the same data with Weighted Opacity Modulation (WOM). UOM and WOM with an applied intensity colormap is shown in (c) and (d) respectively.

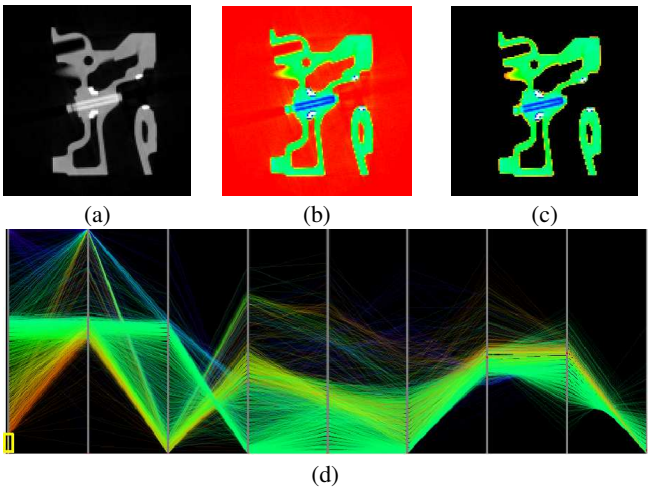


Figure 3: Background noise removal in a slice from the *engine* dataset: (a) Original (cropped) slice; (b) Background noise is also included in the voxel selection. (c) Background noise voxels are removed depending on how they belong to a certain value range in a specific signature (intensity in this example). The different colors mark the different materials according to the intensity colormap in (d). The remaining orange/yellow voxels are due to partial volume effects. (d) shows the *SignatureSpace* where the lower (red) intensity parts (yellow frame) are removed (compared to Fig. 2c)

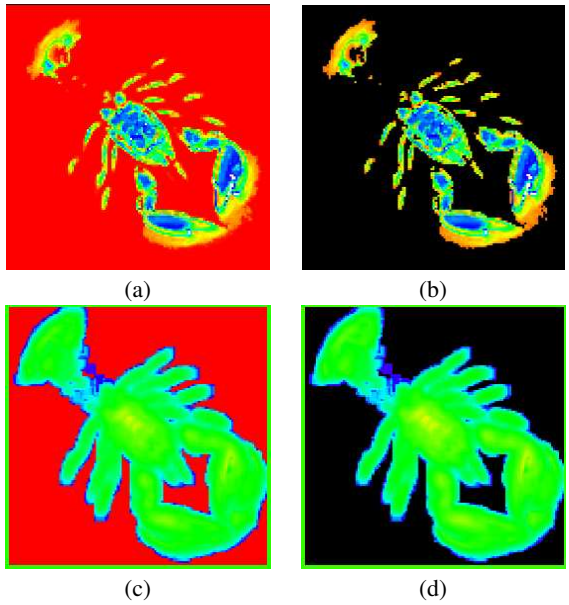


Figure 4: Background noise removal in a slice from the *lobster* dataset: (a) Dataset with colormapped intensity signature; (b) Background noise voxels are removed based on intensity. (c) Dataset with colormapped skew signature; (d) Background noise voxels are removed based on skew.

view on the actual data (Fig.3b). To compensate for this effect, we provided a removal functionality that removes data samples from the rendering process. After specifying a value range in a signature dimension, the respective voxels with a polyline intersecting that value range are removed (Fig.3c). Typically, this operation is performed in the voxel intensity signature, which represents the original voxel values (Fig.3d). However, it can also be applied to any other signature that exposes irrelevant voxels. The *lobster* dataset (Fig.4a) for example shows that the background noise removal operation can be easier achieved, without a potential loss of voxels that are subject to the partial volume effect or are right on the boundary of non-noise voxels. Figure 4 compares the background noise removal effects based on the intensity signature (Fig. 4ab) and noise removal based on the skew signature (Fig. 4cd). While the intensity based noise removal generates a much tighter result, it can lead to problems if the *lobster* needs to be segmented, since parts of the shell might be gone as well.

Scatter Plots

Parallel coordinates usually show a projection of the data distribution on a dimension axis. Although the density of polyline intersections (or their opacities) at certain values of that axis gives cues how the data is distributed, often one would like to see a histogram of the

data, possibly together with another signature dimension. Similar to many other systems (e.g., [7]), our SignatureSpace also provides this kind of scatter plots of the intensity (first signature dimension) and any other signature to allow a more detailed view on the data distribution.

3 Analysis of Volume Data Using the SignatureSpace

Our modified parallel coordinates display consists of nine different signatures as described below. Almost all of the individual signatures have been used previously in the various approaches for the analysis of volumetric datasets or for the (semi-)automatic design of transfer functions. Only two signatures have been added by the authors of this paper, namely the up and down history, where we follow the local gradient one unit step forward and backward to examine the voxel intensity at the new position. This way, the signatures provide some notion of the materials on the other side of the boundary[2].

Currently, the *SignatureSpace* consists of the following signature dimensions:

- Intensity: the actual voxel value in the volume dataset (*data volume*)
- Up history of the voxel: we follow the gradient for one unit step and take the next voxel value as up history.
- Down history of the voxel: similar to the up history, but this time, we follow the negative gradient direction.
- Gradient magnitude: computed by the central difference at the voxels [1, 12].
- Second order moment[3], which is computed by

$$m_2 = \frac{1}{N}\sum_{j=0}^{N-1}(x_j - M)^2 \tag{1}$$

  with mean

$$M = \frac{1}{N}\sum_{j=0}^{N-1} x_j \tag{2}$$

  where N is the number of voxels and x_j is the respective voxel value [20].
- Second derivative: the central difference at the voxel position within the gradient field [12, 17].
- Third order moment[3], which is computed by

$$m_3 = \frac{1}{N}\sum_{j=0}^{N-1}(x_j - M)^3 \tag{3}$$

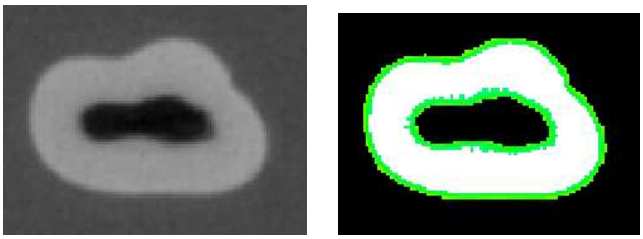
  with M, N, x_j as in m_2 [20].
- Skew: a derived statistical value computed from higher order moments m_2, mean M, and voxel value x_j [16].

$$skew = \frac{1}{(N-1)m_2^{3/2}}\sum_{j=0}^{N-1}(x_j - M)^3 \tag{4}$$

- Kurtosis: also a derived statistical value from the higher order moments m_2, mean M, and voxel value x_j [16].

$$kurtosis = \frac{1}{(N-1)m_2^2}\sum_{j=0}^{N-1}(x_j - M)^4 - 3 \tag{5}$$

These signatures span the *SignatureSpace*, which we use for the analysis of volume data. However, additional signatures can be integrated, if one should decide that they convey useful information. Note that the different signatures are in different data ranges, thus there is only little range correlation between some of the signatures. Currently, we map the whole data range onto the respective coordinate axis. However, a sub range could be selected, which is either computed or selected manually.



(a) (b)

Figure 5: Slice from the *tooth* dataset, an industrial CT scan of a human tooth, which was part of the transfer-function bake-off. (a) shows an original cut slice from the dataset, and (b) shows voxels that are subject to the *partial volume effect* are rendered in green, while the actual feature voxel are rendered in white.

In the next paragraphs, we discuss the advantages of the *SignatureSpace* with several datasets, which are described in Table 1. These datasets represent various features with attenuating boundary features and typical datasets taken from daily life practice.

The traditional specification of transfer functions was only based on the intensity histogram of the voxels. Depending on the source and nature of a volume dataset, different materials could be identified by peaks and valleys in the histogram, if the intensity range of these materials were separable. The famous *engine* dataset – which is a CT scan of an engine block (see Table 1 and Fig. 3) – is an example for these cases. Also CT scans of a patient body provide easy differentiations between tissue, bone, and air. Even more interesting is the fact that a colormap applied to the intensity signature will also expose voxels that are subject to the notorious partial volume effect. With this effect, the sampling rate at volume reconstruction cannot correctly represent the material transition from a high intensity to a significantly lower one, and results in "voxel halo" of lower intensity around the high intensity materials (Fig. 5).

In case of an overlap, these materials cannot be differentiated by intensities alone. Furthermore, if a set of voxels of one material are adjacent to two different materials, eg., one with higher voxel values and one with lower voxel values, the traditional transfer functions based on a standard histogram cannot identify all material interfaces between those materials. This differentiation required the gradient magnitude – since a local maximum in the gradient magnitude exposed one of the respective material interfaces – or a zero crossing in the second derivative [12]. An example of this situation can be seen in the *tooth* dataset (Fig. 5), and in the *cones* datasets (see Table 1 and Fig. 6). In the latter dataset, we see three different quarter cones of different intensity are put into each other, creating three different material boundaries (background/low, low/middle, middle/high; high/background is not present here), similar to the

[2] Lum and Ma also used the gradient direction for the specification of lighting transfer function [13].

[3] Please note, that we do not use the local higher order moments described in [20], but the "regular" higher order moments which consider the whole voxel context as neighborhood. Therefore, N represents all voxels in the current context, which can either be the current data slice, or the whole volume.

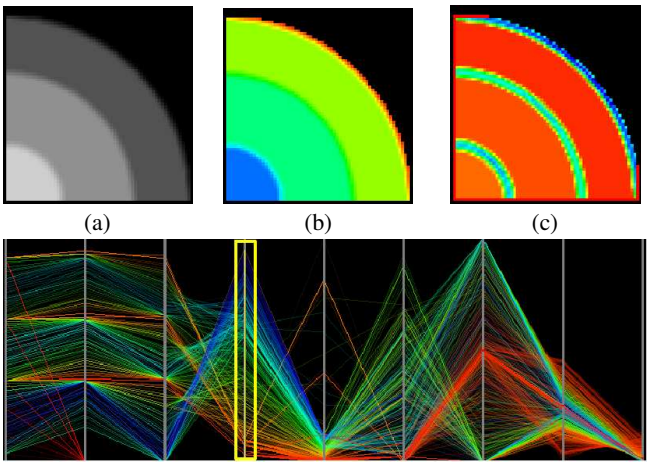


Figure 6: Slice from the *Cones* dataset: (a) shows the original slice, (b) shows a colormapped slice based on the intensity signature,(c) shows a colormapped slice based on the gradient signature, and (d) shows the SignatureSpace with a colormap based on the (yellow framed axis) gradient signature.

cylinders dataset in [12]. The *cones* dataset was generated by a volume modeling system.

Some datasets, however, do not expose a specific isosurface that could be specified. This frequently happens if a dataset represents a probability density or other discretized distribution functions. In these situations, a distinct gradient cannot be found and higher order functions need to be used. Figures 7 and 8 show examples of such an analysis, where the gradient colormap does not expose any specifically useful information. The more suitable third order moment colormap can be used to generate a transfer function.

Another situation is demonstrated with the *Plexiglas phantom* dataset (see Table 1 and Fig. 9). This dataset scanned by a rotational C-arm scanner and it is aimed at measuring reconstruction quality based on different scanning setups. It consists of three connected cavities with a Plexiglas boundary, filled with a (exactly measured quantity of) contrast agent. Due to reconstruction artifacts of the scanning devices, the reconstructed volume dataset contains an attenuating region in the center of the cavities, in particular of the main cavity. Regular intensity or gradient-based analysis will interpret the attenuated region as a second boundary, which of course is only based on the data artifact and is not a real feature (Fig. 9b). Note that we used two different signatures to classify the dataset; we used the intensity signature for background noise removal and the third order moment for the colored classification.

A quite similar effect is shown in Figure 10, where the colormapped second order moment signature generates significantly more homogenous results than the colormap based on the gradient signature. Therefore, the second order moment signature enables an easier classification or segmentation task of the different materials. Here again, we used two different signatures for removal (intensity) and colored classification (second order moment).

With higher analysis methods such as second order moments and in particular the third order moment, shown in Figure 9c, this area is correctly interpreted as homogeneous region, although the artifacts in the volume dataset are generated by the attenuation.

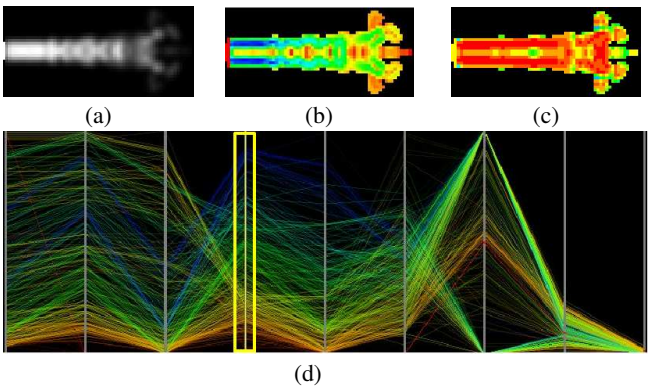


Figure 7: Colormapped slice from the *fuel* dataset, which was generated by a simulation of a fuel injection into a combustion chamber. (a) shows the original (cropped) slice, (b) shows a colormapped slice based on the gradient signature, which exposes the attenuation artifacts, and (c) shows a colormapped slice based on the skew. The SignatureSpace with a gradient colormap (yellow framed signature axis) is shown in (d).

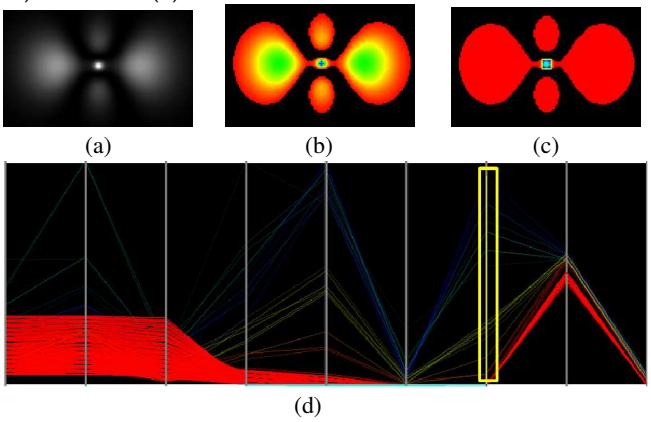


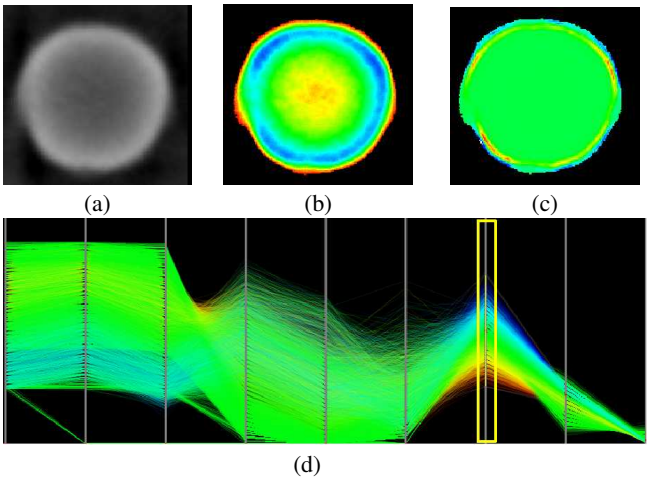
Figure 8: Colormapped slice from the *hydrogen* dataset, which was generated by a simulation of spatial probability of the electron in a hydrogen atom, residing in a strong magnetic field. (a) shows the original (cropped and brightened) slice, (b) shows a colormapped slice based on the intensity signature, and (c) shows a colormapped slice based on the 3rd order moment. The SignatureSpace with a 3rd order moment colormap (yellow framed signature axis) is shown in (d).

3.1 Strategies for Using the *SignatureSpace*

The *SignatureSpace* provides techniques for an exploratory data analysis approach for volumetric data. While some of the individual signatures do assume a data model, others do not. Consequently, we can use the *SignatureSpace* without having specific knowledge about the data to be analyzed.

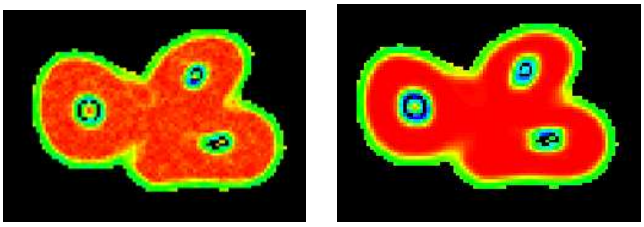
Nevertheless, there are different characteristics of the *SignatureSpace* that can be used to examine specific signature dimensions for the various datasets.

- Clustering: As we know from information visualization, clusters expose similarities or correlation between data samples.
- Color distribution: The color distribution shows how the data samples move through the different signature dimensions, and thus expose possible correlations. The color distribution is also partially compensating the problems arising due to the



(a) (b) (c) (d)

Figure 9: Colormapped slice from *Plexiglas phantom* dataset. (a) shows the original (cropped) slice, (b) shows a colormapped slice based on the intensity signature, which exposes the attenuation artifacts, and (c) shows a colormapped slice based on the higher 3rd moment that depicts the interior successfully. The SignatureSpace with a 3rd order moment colormap (yellow framed signature axis) is shown in (d).



(a) (b)

Figure 10: Slice from the *tooth* dataset, an industrial CT scan of a human tooth, which was part of the transfer-function bake-off. (a) shows slice with colormap of gradient signature, while (b) shows slice with colormap of 2nd order moment signature.

different possible sequences of dimensions in the parallel coordinates.

- Sample distribution: If data samples are represented well in one signature dimension, they are not mapped into a very small data range of the respective signature axis.

In short, we are looking for signatures, where the data covers a significant part of the data range, and possibly has a reasonable sample cluster. We completely ignore signatures where the data is projected into a very small range, like in the kurtosis (last signature axis) in Figures 7 and 8.

The examples we have shown so far, use a variety of different signatures to extract a meaningful representation. Datasets that have a good material differentiation, with no intensity (voxel value) overlap, can be easily analyzed using intensity and gradient signatures, whereas the gradient signature is only necessary, if one material has boundaries to more than one other material (eg., datasets *cones*, *tooth*, and *engine* in Figures 6, 5, and 3).

However, if the boundary is less distinct, or might even suggest a misunderstanding of the data features (eg., *Plexiglas phantom*), intensity and gradient-based analysis will not lead to a successful

Table 1: Dataset overview: The table shows the respective dataset resolutions, the total number of samples, the actually not-removed number of samples, and the used pre-processing time to compute the signatures on a Centrino Laptop running at 1700MHz, with 1GB of main memory. Some datasets have been cropped ([c]) and some datasets have been downsampled ([d]) to save memory space and (pre-)processing time.

Datasets Resolution	# Samples	# Relevant Samples	Pre-Processing Time (s)
Plexiglas Phantom[c] 162x172x214	5.963M	2,717K	565s
Tooth[c] 137x113x155	2.400M	426K	222
MRI Head[d] 127x127x 80	1.290M	502K	150
Lobster[d] 128x128x128	2.097M	311K	297
Engine[d] 128x128x128	2.097M	172K	258
Hydrogen Atom 128x128x128	2.097M	60K	257
Cones 64x64x64	262K	133K	26
Fuel 64x64x64	262K	8K	33

solution. In this case, the third higher order moment was able to grasp the true data characteristics (Fig. 9).

Another example are the *fuel* and *hydrogen* datasets, generated by physical simulations (Figs. 7 and 8). Both datasets represent densities, where a specific boundary – except for a general outside boundary – cannot be found. Our *SignatureSpace* provides a framework that allows for an analysis that is not focusing on boundaries alone. In particular the third order moment signature for the *hydrogen* datasets allows for the identification of the different features in the dataset, as shown in Figure 8c, where lower density regions are marked in red, while the hot spot in the center is identified as a different feature. These marked regions can later be used as segmentation mask in typical segmentation environments.

In our last example, we demonstrate how the *SignatureSpace* can be used to classify different materials in *MRI head*, a downsampled T1-weighted MRI dataset. For the classification, we use again two signatures; the intensity signature to identify the background noise voxels (red in Figure 11b) and the second order moment signature for the identification of the skin (blue) and of the boundaries of the main brain fluid filled cavities (green). The respective *SignatureSpace* is shown in Figure 11c. Note the different colormap bands in the first (intensity) and fifth (second order moment) signatures from the left. Alternatively, the background noise voxels could have been identified by the down history signature (at the very bottom of the axis), or by the skew signature (also at the very bottom of the axis). The correlation between these signature dimensions becomes obvious in our modified parallel coordinate approach (Fig. 11c).

The process of removing background samples and finding an appropriate signature for the classification (or a combination of signatures) takes usually less than a minute. If one (or more) signatures are brushed to select a specific set of data samples, the overall time increased in the presented examples to less than five minutes[4].

[4]The longest classification time required the MRI head (Fig. 11). It involved an intensity colormap for background voxel removal, and a color

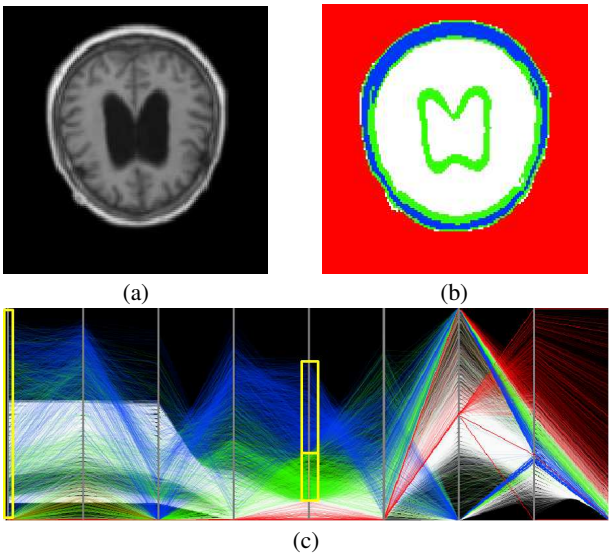


Figure 11: Colored slice from the *MRI head* dataset. (a) shows an original (cropped) slice, (b) shows a classification based on the intensity and 2nd order moment signatures. The *SignatureSpace* with a color band setting based on the intensity (red background; 1st framed signature axis) and 2nd order moment signatures (blue and green; 2nd yellow framed signature axis) is shown in (c).

Please note that the objective of this paper is not to propose higher order moments or their derivates as driving signatures for the data analysis, but to provide a holistic approach to data analysis, which keeps all available signatures in the focus of the analysis.

3.2 Memory Usage

The major problem of the *SignatureSpace* is its large memory consumption. Every signature requires a full volume sized data array to store the signature values for each voxel. To differentiate the various signatures from the original voxel volume, we call the additional volumes *signature volumes* and the original voxel volume *data volume*. Besides the large number of data and signature volumes, the memory consumption is also driven by the float values that are required by many of the *signature volumes*.

There are several possible solutions for the memory problem. Depending on the required accuracy and data range, the float data of some of the signatures could be downsampled into a single-byte integer representation. Another possibility of saving memory comes from the observation that the analysis is often based only on a subset of signatures. Hence, we can save the memory of the signatures that are not required. Finally, suitable compression schemes can be used to reduce the required memory, but at the same time, this would reduce the display and interaction time, since the respective signature volumes need to be decoded first.

Currently, we do not use any of the suggested solutions for the memory problem, but a downsampling of the dataset dimensions for large datasets (see Table 1). While there is a loss of accuracy involved, we believe that this can be neglected for the purpose of data analysis.

band brushing of the 2nd order movement signature.

4 Results

In this section, we briefly present some rendering results, which have been generated based on some of the signature volumes. In Figure 12ab, we show two snapshots from the isosurface rendering of the *Engine* dataset. Please note how the gradient isosurface is bigger and more blurred (Fig. 12b), than the isosurface from the intensity signature (Fig. 12a). A similar observation can be done for the MRI head, which is rendered using the 2nd order moment (Fig. 12cd). These blurring effects are due to the inherent low-pass filtering of the gradient and 2nd order moment volumes. However, in a semi-transparent volume rendered image, this blurring would be less apparent.

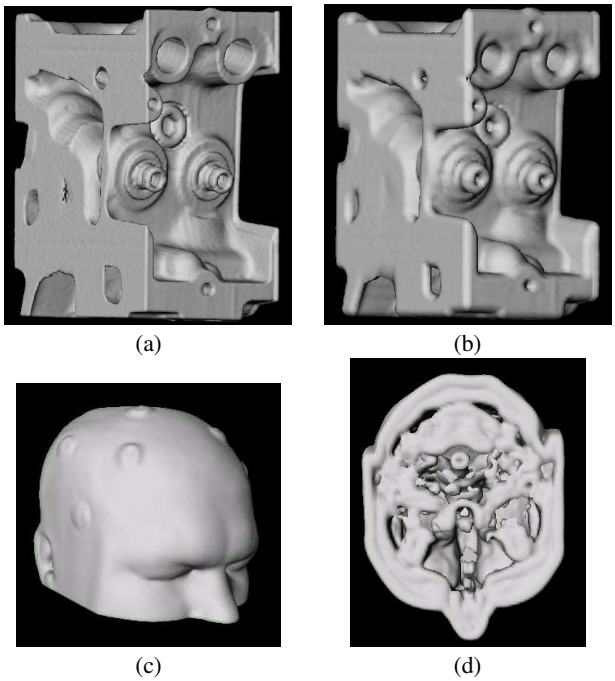


Figure 12: Isosurface rendering of *Engine* dataset. (a) shows the isosurface from the intensity signature, (b) shows the isosurface from the gradient signature. (c) and (d) show two snapshots from the skin isosurface of the 2nd order moment signature volume. Note that the bumps in the surface are fiducial markers attached to the head for an operation.

Table 1 gives an overview of the datasets used in our experiments. It also contains the time needed by a Centrino laptop, equipped with 1GB of memory and one CPU running at 1.7GHz. Please note that these timings are required only once to compute the *SignatureSpace*. Afterwards, it is stored to disk and can be used at any time.

SignatureSpace offers options to render the polylines of the parallel coordinate display based on the current volume slice, or based on the whole volume. Slice-based rendering is achieved at interactive framerates ($>$ 5fps), while full volume rendering of the *SignatureSpace* took up to 30s for the largest volume (Plexiglas phantom).

5 Conclusion and Future Work

In this paper, we introduced the *SignatureSpace*, a novel approach for the analysis of volumetric datasets by adopting techniques from information and scientific visualization. In this approach, several

different metrics derived from the original volume datasets are presented in a parallel coordinate system, which has been improved to be able to handle volume datasets, since they contain orders of magnitude more data samples than typical information visualization datasets. In contrast to many other approaches for the analysis of volume datasets, we do not favor a specific data or boundary model; all signatures are presented equally to the user, who can quickly determine the most useful signature dimension for her/his analysis goal. The powerful combination of signatures enables a data-driven multi-metric classification strategy.

Since the *SignatureSpace* provides analysis only in the data or sample space – no spatial relationships are considered, it hence is a classification approach that provides support for the design of multi-dimensional transfer functions. However, the presented classification strategies can be used as a pre-process for any segmentation task.

In the future, we would like to explore further the exploration of multi-modal datasets with the *SignatureSpace*, where each modality (or data attribute values) is presented next to each other. This way the similarities throughout the different modalities or attributes can be observed in one display.

As we have already pointed out, the colormaps applied to the different signature dimensions can also be used to specify (multi-dimensional) transfer functions. We would like to explore that option further, in particular the possibility of automatically specifying suitable transfer functions.

Finally, the memory problem needs to be addressed in order to analyze truly large volumetric datasets. Right now, we only use dimension downsampling to handle medium and large datasets.

ACKNOWLEDGMENTS

The *fuel* and *hydrogen* datasets are courtesy of the Priority Program SFB 382 of the German Research Foundation (DFG) and were downloaded from http://www.volvis.org. Also the original versions of the *lobster* and *engine* datasets can be found there. The *Plexiglas phantom* dataset is courtesy of Özlem Gürvit of the University Hospital Frankfurt, the MRI head of the University Hospital Tübingen, and the tooth dataset is courtesy of General Electric Corporate Research.

This research has been supported, in part, by NSF CAREER grant ACI-0093157, by the Department of Neurosurgery of the University Hospital Tübingen, and by DFG project VIRTUE of the focus program 1124. We would like to thank Urs Kanus for proof reading.

REFERENCES

[1] C. Bajaj, V. Pascucci, and D. Schikore. The Contour Spectrum. In *Proc. of IEEE Visualization*, pages 167–174, 1997.

[2] H. Doleisch, M. Gasser, and H. Hauser. Interactive Feature Specification for Focus+Context Visualization of Complex Simulation Data. In *Proc. of EG/IEEE VGTC Symposium on Visualization*, pages 239–248, 2003.

[3] Y. Fua, M. Ward, and E. Rundensteiner. Hierarchical Parallel Coordinates for Visualizing Large Multivariate Data Sets. In *Proc. of IEEE Visualization*, pages 43–50, 1999.

[4] I. Fujishiro, T. Azuma, and Y. Takeshima. Automating Transfer Function Design for Comprehensible Volume Rendering Based on 3D Field Topology Analysis. In *Proc. of IEEE Visualization*, pages 467–470, 1999.

[5] D. Gresh, B. Rogowitz, R. Winslow, D. Scollan, and C. Yung. WEAVE: A System for Visually Linking 3-D and Statistical Visualization, Applied to Cardiac Simulation and Measurement Data. In *Proc. of IEEE Visualization*, pages 489–492, 2000.

[6] G. Grinstein, D. Keim, and M. Ward. Information Visualization, Visual Data Mining, and Its Application to Drug Design. In *IEEE Visualization, tutorial 1*, 2002.

[7] H. Hauser, F. Ledermann, and H. Doleisch. Angular Brushing of Extended Parallel Coordinates. In *Proc. of IEEE Symposium on Information Visualization*, 2002.

[8] T. He, L. Hong, A. Kaufman, and H. Pfister. Generation of Transfer Functions with Stochastic Search Techniques. In *Proc. of IEEE Visualization*, pages 227–234, 1996.

[9] H. Höhne and R. Bernstein. Shading 3D-Images from CT using Gray-level Gradients. In *IEEE Transactions on Medical Imaging*, volume MI-5, pages 45–47, 1986.

[10] A. Inselberg. The Plane with Parallel Coordinates. *The Visual Computer*, 1:69–92, 1985.

[11] A. Inselberg and B. Dimsdale. Parallel Coordinates: A Tool for Visualizing Multidimensional Geometry. In *Proc. of IEEE Visualization*, pages 361–378, 1990.

[12] G. Kindlmann and J. Durkin. Semi-Automatic Generation of Transfer Functions for Direct Volume Rendering. In *Proc. of Symposium on Volume Visualization*, pages 79–86, 1998.

[13] E. Lum and K. Ma. Lighting Transfer Functions for Direct Volume Rendering. In *Proc. of IEEE Visualization*, pages –, 2004.

[14] A. Martin and M. Ward. High Dimensional Brushing for Interactive Exploration of Multivariate Data. In *Proc. of IEEE Visualization*, pages 271–278, 1995.

[15] J. Miller and E. Wegman. Construction of Line Densities for Parallel Coordinate Plots. In *Computing and Graphics in Statistics*, pages 107–124. Springer-Verlag, Heidelberg/New York, 1991.

[16] NIST. NIST/SEMATECH e-Handbook of Statistical Methods. http://www.itl.nist.gov/div898/handbook/, accessed April 2004, 2004.

[17] V. Pekar, R. Wiemker, and D. Hempel. Fast Detection of Meaningful Isosurfaces for Volume Data Visualization. In *Proc. of IEEE Visualization*, pages 223–230, 2001.

[18] H. Pfister, B. Lorensen, C. Bajaj, G. Kindlmann, W. Schroeder, L. Avila, R. Machiraju, and J. Lee. The Transfer Function Bake-off. *IEEE Computer Graphics and Applications*, 21(3):16–22, 2001.

[19] B. Rogowitz, L. Treinish, and S. Bryson. How Not To Lie With Visualization. *Computers in Physics*, 10(3):268–274, 1996.

[20] S. Tenginakai, J. Lee, and R. Machiraju. Salient Iso-Surface Detection with Model-Independent Statistical Signatures. In *Proc. of IEEE Visualization*, pages 231–238, 2001.

[21] S. Tenginakai and R. Machiraju. Statistical Computation of Salient Iso-Values. In *Proc. of EG/IEEE VGTC Symposium on Visualization*, pages 19–24, 2002.

[22] M. Tory, S. Potts, and T. Möller. A Parallel Coordinates Style Interface for Exploratory Volume Visualization (Poster). In *IEEE Transactions on Visualization and Computer Graphics*, pages 71–80, 2005.

[23] F. Tzeng, E. Lum, and K. Ma. A Novel Interface for Higher-Dimensional Classification of Volume Data. In *Proc. of IEEE Visualization*, pages 505–512, 2003.

[24] M. Ward. XmdvTool: Integrating Multiple Methods for Visualizing Multivariate Data. In *Proc. of IEEE Visualization*, pages 326–333, 1994.