

# Volume Visualization based on Statistical Transfer-Function Spaces

Martin Haidacher\*  
Institute of Computer  
Graphics and Algorithms  
Vienna University of  
Technology, Austria

Daniel Patel†  
Christian Michelsen  
Research  
Bergen, Norway

Stefan Bruckner‡  
Institute of Computer  
Graphics and Algorithms  
Vienna University of  
Technology, Austria

Armin Kanitsar§  
AGFA HealthCare  
Vienna, Austria

M. Eduard Gröller¶  
Institute of Computer  
Graphics and Algorithms  
Vienna University of  
Technology, Austria

## ABSTRACT

It is a difficult task to design transfer functions for noisy data. In traditional transfer-function spaces, data values of different materials overlap. In this paper we introduce a novel statistical transfer-function space which in the presence of noise, separates different materials in volume data sets. Our method adaptively estimates statistical properties, i.e. the mean value and the standard deviation, of the data values in the neighborhood of each sample point. These properties are used to define a transfer-function space which enables the distinction of different materials. Additionally, we present a novel approach for interacting with our new transfer-function space which enables the design of transfer functions based on statistical properties. Furthermore, we demonstrate that statistical information can be applied to enhance visual appearance in the rendering process. We compare the new method with 1D, 2D, and LH transfer functions to demonstrate its usefulness.

**Keywords:** Transfer function, statistics, classification, noisy data, shading.

**Index Terms:** I.3.7 [Computer Graphics]: Three-Dimensional Graphics and Realism—Color, shading, shadowing, and texture; I.4.10 [Image Processing and Computer Vision]: Image Representation—Volumetric

## 1 INTRODUCTION

Volume classification is a major issue in volume visualization. The goal of classification is to enhance different materials or features which are important for further analysis of the data. Transfer functions have been proven to be a powerful tool for classification. Nevertheless, in most cases it is a non-trivial task to find a good transfer function which is able to achieve the desired classification.

A transfer function is a general concept. Concrete implementations use one or more properties, derived from the data, to define a transfer-function space. A transfer function is then designed in this space. How easily different materials can be distinguished by the transfer function, depends on the data set as well as on the used properties in the definition of the transfer-function space.

Noise in the measured data is a typical problem, which complicates the classification process. The most frequently observed noise in measured data is Gaussian white noise. White noise has a mean value of zero and a symmetric variance. For different materials in a volume data set, the noise causes variations of the intensity values around an average value. Therefore, it is difficult to assign intensity values of sample points to a certain material, especially if average

values of other materials are close by. This intermixing of materials in the intensity space cannot be resolved in the transfer-function space when only local properties, such as the intensity and gradient magnitude, are used to define the transfer-function space.

In this paper we present a method which considers a larger area around each sample point to derive properties for the transfer-function space. By this, we are able to estimate the distribution of noise around the average value of a material. The statistical properties of this distribution are used to describe the material at a certain sample position. Since different materials can be distinguished by their distributions of intensity values, we are able to separate them.

For the estimation of the statistical properties, we employ an adaptive growing approach at each sample point. The extent of the growing is dependent on the local neighborhood of a sample point. The estimated properties are used to define the statistical transfer-function space. Sample points from separate materials can be seen as separate clusters in this space. We introduce special transfer-function regions which are adapted for this space to design a transfer function. Furthermore, we demonstrate how the statistical properties can be used to steer visual properties such as shading. This results in higher quality visualizations, especially for noisy data.

We use the new statistical transfer-function space to generate images for medical MR and industrial CT data. We show for these data sets, that our method classifies different materials better than other state-of-the-art transfer-function spaces.

## 2 RELATED WORK

The work presented in this paper spans several research fields. Since we are dealing with noise in the data, the analysis of noise in image processing is related to this work. The growing of regions around each sample point is similar to the scale-space analysis where data is analyzed on different scales. In this work we introduce a new transfer-function space. Therefore, most of the related work is dedicated to other transfer-function spaces.

**Image Processing.** Noise in data is a well investigated field in image processing. In this work we assume the noise in the data to be Gaussian white noise. This is specifically true for CT data sets [12, 27]. For other data sets, the Gaussian distribution is at least a good approximation of the noise distribution. In MRI, e.g., the real distribution is a Rician distribution, but for a low signal-to-noise ratio the difference to a Gaussian distribution is very small [6].

**Scale-Space Analysis.** Early works, such as Lindeberg [14], analyzed images on different scales. Over the years, different scale-spaces were investigated. The most common scale space is the linear scale space, which is generated by progressive Gaussian smoothing. In this scale-space Lindeberg [15] introduced a technique for feature detection and automatic scale selection. Due to the complexity of the scale-space generation for volume data, alternatives, such as Laplacian pyramids [5] or Wavelet transforms [20], were developed for an easier and faster representation of different scales. A method to improve the classification of features, based on the pyramid representation, was introduced by Lum *et al.* [17]. In contrast to scale-space analysis, our method uses different scales for each sample point because the growing is terminated depending on local properties of the neighborhood region.

\*e-mail:haidacher@cg.tuwien.ac.at

†e-mail:daniel@cmr.no

‡e-mail:bruckner@cg.tuwien.ac.at

§e-mail:armin.kanitsar@agfa.com

¶e-mail:groeller@cg.tuwien.ac.at

**Transfer Function Spaces.** In an early work, Levoy [13] used the data value alone to define a transfer function space. Kniss *et al.* [9] employed the data value and the gradient magnitude for the classification of different materials and borders between them. Since they only consider single data values and a very small neighborhood for the gradient magnitude, this technique is not well suited for the classification of noisy data. Hladůvka *et al.* [8] proposed curvature as an additional property for the classification. With this method special features, like ridges and valleys, could be extracted. An extension to multi-dimensional transfer functions was introduced by Roettger *et al.* [23]. The method includes spatial information in the transfer-function space. They simplified the transfer-function design-process by using the spatial information to color the transfer-function space. However, for noisy data different materials overlap in this transfer-function space.

In the work of Lum and Ma [16], a larger region is considered for the definition of the transfer-function space. Besides the data value at a sample point, a data value along the gradient direction is used as well. In data sets with sharp transitions, such as CT data, this technique can be used to highlight border areas. An extension to this method was introduced by Sereda *et al.* [26] named LH histograms. This method looks for homogeneous regions along both directions of the gradient streamline. The detected low and high data values are used to define the transfer-function space. This method provides good results for data sets with little noise. For noisy data sets, values in homogeneous regions have a high variance. Therefore, the clusters representing homogeneous regions are getting larger and overlap each other in the LH histogram space.

A method which also uses a larger neighborhood for the classification was presented by Hadwiger *et al.* [7]. They use region growing to detect features of different sizes in industrial CT data. In a 3D transfer-function space these different features can be classified. In the work of Correa and Ma [3], a multi-scale approach is used to detect the size of features in a data set. The feature size is then used as an additional parameter for the definition of a transfer function. In both approaches the shape of a feature in the data set is the main criterion for the classification. Instead in our method, the statistical properties of materials are used for the classification. These properties are independent of object shapes.

Lundström *et al.* [18] introduced a method to classify different tissues by the local histograms in the neighborhood around a sample point. Caban and Rheingans [2] used textural properties to differentiate between materials, possibly with similar data values. These methods are able to separate materials but they use a neighborhood with a fixed size for the extraction. Thus, these approaches do not differentiate between homogeneous and inhomogeneous regions.

Laidlaw *et al.* [11] use Bayes' theorem on a small neighborhood of a voxel to classify mixed materials. Tenginakai *et al.* [24, 25] introduced a method to extract salient iso-surfaces based on statistical methods. A different classification based on statistics was introduced by Kniss *et al.* [10]. For the estimation of the statistical characteristics certain properties of the different materials have to be known. For our approach no prior knowledge of material properties is necessary. Lundström *et al.* [19] used the variance in a neighborhood of a voxel to separate materials. In comparison to our work, they used a fixed neighborhood size to estimate the variance. In our previous work (Patel *et al.* [21]) we used statistical properties to manually classify materials for differently sized neighborhood regions. In this work we extract the statistical properties for the best suited neighborhood size semi-automatically. Furthermore, we use these statistical properties to define a transfer-function space and to enhance the visual appearance of the resulting rendering.

### 3 STATISTICAL TRANSFER-FUNCTION SPACE

The idea behind the statistical transfer-function space is that materials are distinguishable according to their statistical properties.

Since the data is not segmented, we are not able to calculate the statistical properties for different materials in general. Therefore, we introduce a technique which extracts statistical properties for the neighborhood of each sample point individually. We expect that sample points from the same material get similar statistical properties. In the new transfer-function space this leads to clusters for different materials, which makes it possible to design meaningful transfer functions. In this section we describe all steps which are necessary to generate the statistical transfer-function space.

Figure 1 shows an overview of the workflow. To generate a visualization based on statistical transfer functions, different processing steps have to be applied on a volume data set. For the generation of the transfer-function space, statistical properties, i.e. the mean value and the standard deviation, are extracted first. This is done in a pre-processing step. The user defines a confidence level for this step. This confidence level is a quantity for the tolerance in the extraction step. It is further explained in Section 3.1.

The properties for each sample point are then depicted in the transfer-function space. They serve as a clue for the user to design a transfer function. The transfer function together with the statistical properties drives the successive visualization step. Additionally the statistical properties are used to enhance the shading.

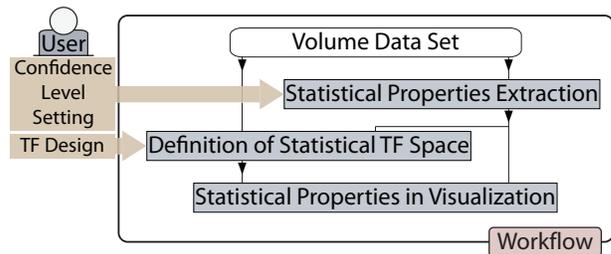


Figure 1: Statistical transfer-function workflow.

To exemplify our new method, we generated a synthetic data set of size  $128 \times 128 \times 128$ . The data set contains three different materials. In Figure 2(a) a slice through the center of the data set is shown. Material 2 in the center of the data set is a sphere, embedded between material 1 and material 3. Gaussian white noise has been added to all three materials. As mentioned before this is a realistic noise model for most data sets especially for CT and MRI. In Figure 2(b) the histograms of the materials are shown. On the horizontal axis the data values  $f(x, y, z)$  of the sample points are mapped. The vertical axis holds the frequency of occurrences  $I$  for each data value. The Gaussian distributions of all three materials have high standard deviations ( $\sigma_1 = 0.14$ ,  $\sigma_2 = 0.09$ , and  $\sigma_3 = 0.11$ ), consequently the distributions considerably overlap each other. The black line gives the frequency distribution of all three materials together.

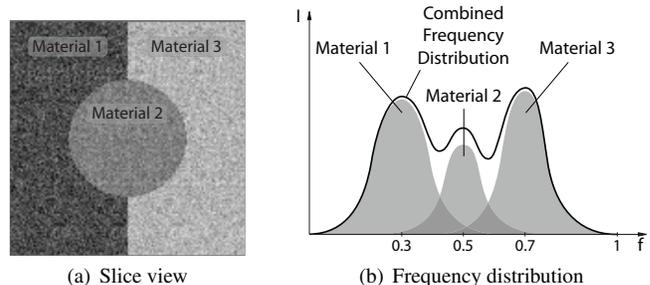


Figure 2: Slice through the synthetic data (a) and the frequency distribution of the data values (b).

The synthetic data set exemplifies noisy data. The overlapping of the distributions for different materials is a problem which often occurs in real-world data sets [26]. In the further explanation of the method, the synthetic data set is used to show the effects of the different processing steps.

### 3.1 Statistical Properties Extraction

The extraction of the statistical properties is essential for our statistical transfer-function space. For sample points within a certain material, the extracted statistical properties should be close to the statistical properties of the entire material. We achieve this by investigating a neighborhood region around each sample point. To keep the neighborhood within the same material, we introduce an adaptive growing which is dependent on the local properties.

The distribution of data values of a single material in real-world data sets, such as MRI or CT, can be approximated very well by the Gaussian white noise model. Therefore, we consider only the Gaussian distribution as basis for the calculation of statistical properties. A Gaussian distribution is described by its mean value and standard deviation. Hence, we use these two parameters as our statistical properties.

For the extraction of the statistical properties, we iteratively grow a spherical neighborhood by increasing the radius by one voxel in each step. We compare for each growing step if the newly grown hull still belongs to the same material. Figure 3 shows a cross section of such a neighborhood. In the following explanation we use two different notations for the statistical properties. The mean value  $\mu_r$  and the standard deviation  $\sigma_r$  for a certain radius  $r$ , are the estimations for the statistical properties of all points within a sphere of radius  $r$ .  $\hat{\mu}_r$  and  $\hat{\sigma}_r$  are the statistical properties of the points in the outer hull of the sphere (see Figure 3).

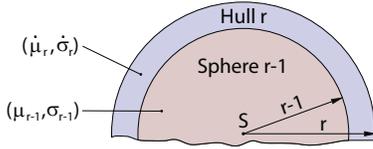


Figure 3: The standard deviation and mean for the sphere and the outer hull.

In each growing step several calculations are done to decide if the growing should be terminated or not. Figure 4 shows the processing steps for each growing step. The goal of these steps is to detect whether the sphere grows into another material. In such a case the loop is terminated.

As initial parameters for the loop, the mean value and the standard deviation of a sphere with a radius of one are used. The data values of points in the sphere may not be normally distributed. This can happen when the sphere intersects two materials. Such a situation should be detected in an early stage so growing can be terminated. Therefore, we apply a normal-distribution test, as will be described later in this section, to check if the points in the initial sphere are normally distributed. Only if this test is passed, the loop is started with the initial statistical properties. Otherwise  $\mu_1$  and  $\sigma_1$  of the initial sphere are used as statistical properties for the actual sample point.

In the loop depicted in Figure 4, the first step is the calculation of the statistical properties  $\hat{\mu}_r$  and  $\hat{\sigma}_r$  for points in the hull at a radius  $r$ . In the next step it is tested if the distribution in the hull is normally distributed. If this is the case, the properties are compared with  $\mu_{r-1}$  and  $\sigma_{r-1}$ . In the case the statistical properties are similar, the statistical properties of the hull are merged with the statistical properties of the sphere  $r-1$ . If a sample point lies in the center of a large homogeneous area, the loop is terminated when the maximum radius  $r_{max}$  is reached.

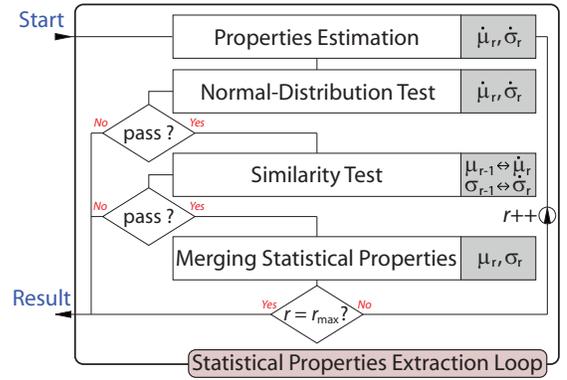


Figure 4: Calculation loop for the extraction of the statistical properties.

For the extraction of the statistical properties, a confidence level  $\omega$  has to be set. This confidence level expresses the general confidence in the distribution of data values in a data set with respect to the general noise level. It can be set differently to adapt the model for various data types such as MRI or CT. In the following part of this section, the usage of  $\omega$  and the other processing steps are described in more detail.

#### Properties Estimation

In each successive cycle of the extraction loop, the statistical properties of a larger region are considered. Since the statistical properties for sphere  $r-1$  are already known, we are interested in the statistical properties  $\hat{\mu}_r$  and  $\hat{\sigma}_r$  of the additional points in the hull  $r$  (see Figure 3).

As the distribution is considered to be Gaussian, we estimate the mean value and the standard deviation for the points in the hull  $r$  [4]. The mean value  $\hat{\mu}_r$  is the average and  $\hat{\sigma}_r$  is the biased standard deviation of all points:

$$\hat{\mu}_r = \frac{1}{N_r} \sum_{i=1}^{N_r} f_i, \quad \hat{\sigma}_r = \sqrt{\frac{1}{N_r} \sum_{i=1}^{N_r} (f_i - \hat{\mu}_r)^2} \quad (1)$$

$N_r$  is the number of points included in the hull  $r$ .  $f_i$  denotes the data value of a sample point in the hull. With these estimations for the mean value and standard deviation we expect to get values which are close to the real statistical properties of the material in the hull.

#### Normal-Distribution Test

Before we apply a similarity test with the derived properties of the hull and the statistical properties of the inner sphere, it must be ensured that the distribution is normally distributed. This is necessary because the similarity test is based on normal distributions. When two materials are intersecting the hull, the distribution is not normally distributed. In such a case the distribution would have two peaks. With the normal-distribution test we want to detect such situations and terminate the loop.

In statistics, several normality tests exist. We chose the Jarque-Bera test (JB). This test uses the third-order moment, i.e. the skewness  $\hat{S}_r$ , and the fourth-order moment, i.e. the kurtosis  $\hat{K}_r$ , of the points in the hull at radius  $r$  for the calculation of JB:

$$JB = \frac{N_r}{6} \left( \hat{S}_r^2 + \frac{(\hat{K}_r - 3)^2}{4} \right) \quad (2)$$

The benefit of using this test is that it is not necessary to sort the point values of the distribution. Therefore, it can be efficiently implemented on the graphics hardware.

The parameter  $JB$  is used to test for the null hypothesis with a test level  $1 - \alpha$  of 99.9%. This results in a test value of 13.82 according to statistical lookup-tables:

$$H_0 : JB < 13.82_{(1-\alpha=0.999)} \quad (3)$$

With the high test level, only distributions with a high divergence to a normal distribution are declined by the null hypothesis. This is necessary, due to the low number of samples which are used for the test. Only distributions which are very different from a normal distribution will fail the test.

If the null hypothesis is declined, the loop is terminated and  $\mu_{r-1}$  and  $\sigma_{r-1}$  of sphere  $r - 1$  are taken as statistical properties for the sample point  $S$ . If the test passes the null hypothesis, we continue with the similarity test.

### Similarity Test

In this processing step, we measure the similarity between the statistical properties in the hull  $r$  and the statistical properties of the sphere  $r - 1$ . The goal is to detect whether the hull is still part of the same material as the sphere  $r - 1$ .

Through the properties-estimation step, we get the statistical properties of the hull. We have calculated the statistical properties of the sphere  $r - 1$  in the preceding loop cycle. Since all the values are estimations and the numbers of points which are involved in the estimation is rather low, we use a variant of the student's t-test for the similarity test. This test is best suited for Gaussian-distributed populations with small sample sizes [4].

In our case, we have two independent samples which were used for the estimations. Since the mean values as well as the standard deviations of both estimations can vary, we use a generalized form of the student's t-test also known as the Welch's t-test [28]. As primary parameter for the similarity test, a  $t_r$  parameter is calculated:

$$t_r = \frac{\mu_{r-1} - \hat{\mu}_r}{\sqrt{\frac{\sigma_{r-1}^2}{N_{r-1}-1} + \frac{\hat{\sigma}_r^2}{N_r-1}}} \quad (4)$$

The  $t_r$  parameter is dependent on the mean values, the standard deviations, and sample sizes of both distributions. Additionally, a degree-of-freedom  $\delta_r$  has to be calculated:

$$\delta_r = \frac{\left(\frac{\sigma_{r-1}^2}{N_{r-1}-1} + \frac{\hat{\sigma}_r^2}{N_r-1}\right)^2}{\frac{\sigma_{r-1}^4}{(N_{r-1}-1)^3} + \frac{\hat{\sigma}_r^4}{(N_r-1)^3}} \quad (5)$$

The degree-of-freedom  $\delta_r$  is only dependent on the standard deviations and the sample sizes, but not on the mean values. The  $\delta_r$  value together with the confidence level  $\omega$  are used as parameters to get a reference  $t_\omega(\delta_r)$  value in a t-test lookup-table. This value is used to test for the null hypothesis  $H_0$ :

$$H_0 : |t_r| < t_\omega(\delta_r) \quad (6)$$

If the null hypothesis is true, it is assumed that both Gaussian distributions are the same. If the null hypothesis is declined then both distributions are expected to be different with a probability of  $1 - \omega$ . Therefore, a small confidence level  $\omega$  results in a high probability that both distributions are not similar if the null hypothesis is declined. On the other hand, the reference  $t_\omega(\delta_r)$  value for a small  $\omega$  is high which makes the similarity test less selective.

As in the step earlier  $\mu_{r-1}$  and  $\sigma_{r-1}$  of sphere  $r - 1$  are taken as statistical properties if the test is failed. Otherwise we continue with the next growing step.

### Merging Statistical Properties

If the statistical properties have passed the normal-distribution test and the similarity test, we assume that the material in the outer hull still is the same as in the sphere  $r - 1$ . Therefore, the statistical properties of both areas can be merged together.

This step results in a new  $\mu_r$  and  $\sigma_r$ . These statistical properties represent the distribution of all points in the sphere  $r$ . The merged statistical properties are used in the successive cycle of the loop to do the similarity test with the next larger hull:  $r + 1$ .

The loop is terminated when the normal-distribution test or the similarity test fails or when the maximum radius is reached. In the first two cases we store  $\mu_{r-1}$  and  $\sigma_{r-1}$  as statistical properties  $\mu$  and  $\sigma$  for the sample point  $S$ . In the third case we take the statistical properties  $\mu_r$  and  $\sigma_r$  after the merging step.

Additionally, we store the radius  $r_{break}$  at which the loop is terminated. The closer  $r_{break}$  is to  $r_{max}$  the more significant the statistical properties are at this point, because the population of points for the estimation is larger. In the next section,  $r_{break}$  is used to highlight statistical properties with a higher significance.

Figure 5 shows the statistical properties  $\mu$  and  $\sigma$  for the synthetic data set at different confidence levels  $\omega$ . For a low  $\omega$  of 0.1%, the similarity test is more easily passed. Therefore, the sphere grows larger and results in more consistent values for  $\mu$  and  $\sigma$ . For the transfer-function space this means that the clusters of materials are smaller. If the confidence level  $\omega$  is high, e.g. 30%, the loop has a higher probability of being terminated. Since in this case a smaller area is used to estimate the statistical properties, the result for  $\mu$  and  $\sigma$  is less smooth compared to a low  $\omega$  but details, such as borders, are preserved better.

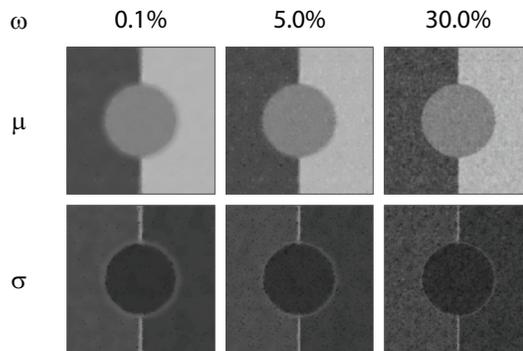


Figure 5: Statistical properties ( $\mu$  and  $\sigma$ ) of the synthetic data for three different confidence levels  $\omega = \{0.1\%, 5.0\%, 30.0\%\}$ . The brighter a point is, the higher is its value.

Congruent to the characteristic in this example,  $\omega$  should be chosen according to the type of data. If a modality is very noisy and the distribution of points does not exactly follow a normal distribution, the confidence value should be chosen rather low. Therefore, some details get lost but the clusters for different materials in the transfer-function space are smaller. MRI is an example of such data. In such a case,  $\omega$  should be set to a low value, such as 0.1%. For less noisy data types with a Gaussian-like distribution of the data values, e.g. CT,  $\omega$  can be set to a higher value, such as 20%. Then details are better preserved and the clusters are small enough due to the low noise level. The synthetic data set is rather noisy but the distribution of the points follows exactly a normal distribution. Therefore, a confidence level of 5% was chosen. The values for the confidence

levels were found through experiments, where we tried to achieve the best balance between noise reduction and detail preservation. This has to be done only once for a certain type of data source. In Section 5 we demonstrate that our presented values work well for different MRI and CT data sets.

### 3.2 Definition of Statistical Transfer-Function Space

By extracting the statistical properties we get a mean value  $\mu$  and a standard deviation  $\sigma$  for each sample point. In the next step we use this information for the design of a transfer function. First we have to define a transfer-function space, which is used for the design of a transfer function. The intent of this statistical transfer-function space is to separate different materials in the presence of noise.

For the transfer-function space we use the original data value  $f$  of each sample point together with the mean value  $\mu$  and the standard deviation  $\sigma$ . We follow the convention of using the horizontal axis of a transfer-function space for the data value. The horizontal axis is, however, also used to depict the mean  $\mu$ . In the new transfer-function space the data value  $f$  on the horizontal axis is considered as starting point of a line segment. The statistical properties  $\mu$  and  $\sigma$  for each sample point define a second point, where the standard deviation is given on the vertical axis.

In Figure 6 the transfer-function space is shown. On the left side the properties of a sample point  $S$  are drawn as a line segment in this space.

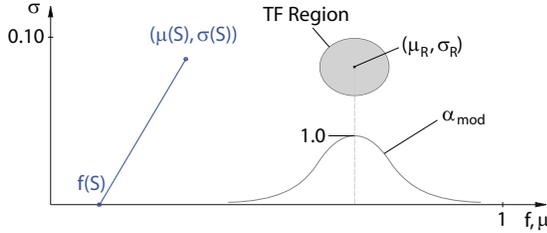


Figure 6: Schematic representation of the statistical transfer-function space. On the left side the properties of a sample point are shown as a line segment. On the right side a transfer-function region in the new space can be seen.

Since the clusters for each material are expected to show up around a certain mean value and standard deviation, an elliptical area is used to define a transfer-function region. On the right side of Figure 6 an example of such a region is shown. The center of the region is at  $(\mu_R, \sigma_R)$ . A sample point is only classified by this region when its statistical properties  $(\mu, \sigma)$  lie within the elliptical area.

The optical properties for each region are represented with a color  $c_R$  and an opacity  $\alpha_R$ . The position of the transfer-function region also defines the corresponding Gaussian distribution of data values along the horizontal axis. The opacity of sample points is manipulated according to this distribution:

$$\alpha_{mod} = \alpha_R * e^{-\frac{1}{2} \left( \frac{f - \mu_R}{\sigma_R} \right)^2} \quad (7)$$

$\alpha_R$  is the opacity defined for the respective transfer-function region.  $f$  is the data value of the sample point and  $(\mu_R, \sigma_R)$  is the center point of the region. Equation 7 is represented as a curve along the horizontal axis in Figure 6, below the transfer-function region. The opacity for data values  $f$  close to the center  $\mu_R$  of the region are modified less in contrast to the opacity of data values  $f$ , which are more different to  $\mu_R$ . By this modification, we reduce the influence of sample points which are less likely to be part of the desired distribution.

Figure 7 shows the sample points of the synthetic data set in the statistical transfer-function space. In Figure 7(a) each sample point is represented as a line segment. Since the line segments of points with high  $\sigma$  values might occlude points with low  $\sigma$  values, an alternative representation is shown in Figure 7(b). In this representation only a single dot at  $(\mu, \sigma)$  is drawn for each sample point. In both representations three large clusters for the different materials and three smaller clusters for the borders can be seen. The border between material 1 and 3 actually consists of two small clusters. This results from the fact that the mean values of material 1 and 3 are very different. Therefore, it makes a difference if the origin sample point of a neighborhood-region lies in material 1 or 3, because more points from one or the other material are then used for the estimation of the mean value.

If the difference between the mean values of two materials is smaller, e.g. between material 1 and 2, then only one cluster shows up in the transfer-function space.

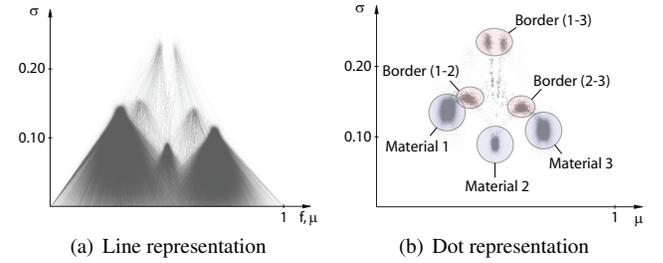


Figure 7: Line and dot representation in the statistical transfer-function space of the synthetic data.

To enhance points with higher significance, we weight the opacity of line segments or dots by the termination radius  $r_{break}$  of the calculation loop. Thus, statistically well defined interior regions are emphasized.

### 3.3 Statistical Properties in Visualization

In addition to the definition of a transfer function we use the statistical properties as input for the shading process. In volume visualization gradient-based techniques are most common for shading. They are computational less expensive and faster as gradient-free shading techniques. For noisy data the gradient-based techniques have the penalty that the noise in the data deteriorates the estimation of the gradient and, therefore the shading [1]. Especially in homogeneous regions, where the gradient magnitude is rather low, noise has a high impact on the estimation of the gradient direction.

To reduce the influence of noise on the gradient, we use the mean values of the sample points to estimate the gradient direction. The mean values are smoother than the original data values and, therefore, the gradient direction in nearly homogeneous regions is estimated better in comparison to the widely used central-difference method. With this gradient we calculate a color  $c_{shaded}$  based on Phong shading [22].

For the observer, shading is important for the perception of surfaces. To avoid visual clutter resulted when shading all parts of the volume, we apply shading only for border regions. For this purpose the standard deviation  $\sigma$  can be used.  $\sigma$  is higher in border regions and it is less affected by noise than the gradient magnitude. Therefore,  $\sigma$  is used to interpolate between the shaded color  $c_{shaded}$  and the unshaded color  $c_{unshaded}$ :

$$c = (1 - \sigma)c_{unshaded} + \sigma c_{shaded} \quad (8)$$

The lower  $\sigma$  is, the less shading is applied. This leads to a visualization where the border areas are shaded more as opposed to

the rest. The synthetic data with traditional shading in Figure 8(a) is shown in contrast to the shading based on statistical properties in Figure 8(b). With statistical shading the influence of noise is clearly reduced without modifying the data itself from filtering or other techniques.

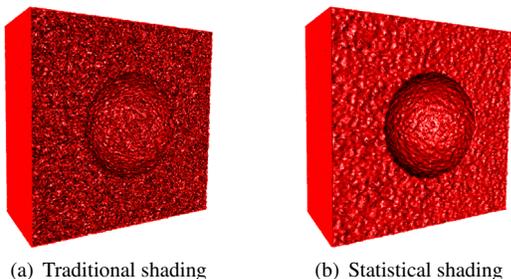


Figure 8: Visual difference between traditional shading and statistical shading. With statistical shading the influence of noise is reduced.

#### 4 IMPLEMENTATION

The steps which are described in Section 3 can be divided into two parts of computation. The first part is the extraction of statistical properties. This is a pre-processing step which has to be recalculated only when the confidence level  $\omega$  is changed. The other part is the visualization with the usage of the statistical transfer function. This part is done in real-time.

Since the extraction of the statistical properties is a highly parallel process, we use graphics hardware. Nevertheless, the extraction is expensive because many lookups have to be performed to estimate the statistical properties.

The maximum number of lookups per sample point is dependent on the maximum radius  $r_{max}$  of the calculation loop. We decided to set the maximum radius to a value of 6. With this radius, a maximum of 925 lookups are done per sample point. This leads to the following 99% confidence limits:

$$\mu = \bar{\mu} \pm (0.0424\sigma) \quad (9)$$

This means that the estimated  $\mu$  is, with a confidence of 99%, not farther away from the real mean value  $\bar{\mu}$  than 0.0424 times  $\sigma$ . Even for a large  $\sigma$  the interval for  $\mu$  is small enough for our purpose of estimating the statistical properties.

In addition to the maximum radius  $r_{max}$ , the confidence level  $\omega$  influences the speed of the calculation. The lower  $\omega$ , the less likely it is that the calculation loop is terminated. Therefore, more lookups have to be done for the estimation of the statistical properties. In Figure 9 the termination level at different radii is shown. The graph shows the percentage of all sample points for which the calculation loop is already terminated at a certain radius. For the synthetic data set different confidence levels are used. It can be seen that the termination level at the maximum radius of 6 is much lower for an  $\omega$  of 0.1% in comparison to an  $\omega$  of 30%. For the low  $\omega$  the loop is only terminated for points close to the border. A high  $\omega$  causes terminations also for large variations in homogeneous regions. As shown in Figure 5, the visual difference of various  $\omega$  values can be seen in the smoothness.

Apart from the synthetic data set, Figure 9 also shows the termination level for two real-world data sets with according confidence levels for each type of data. In the case of CT data, we get a high termination level even at a low radius. This results from the fact that areas with zero variation, e.g. air, do not pass the initial normal-distribution test. For MRI data, the curve is more linear, due to the fact that all parts contain at least some noise.

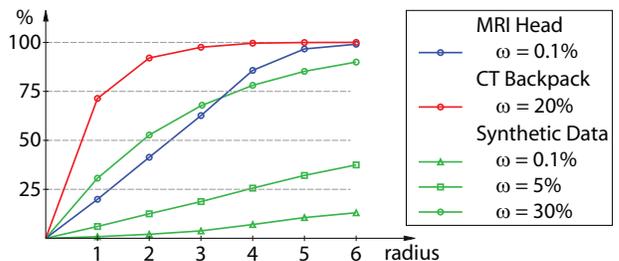


Figure 9: Termination level of the calculation loop for different data sets and confidence levels  $\omega$ .

Table 1 shows the time measurements for the estimation of the statistical properties for the data sets of Figure 9 on a GeForce GTX 260. The different settings for the level  $\omega$  have only a small influence on the calculation time. This results from the fact that most time is used to initialize the graphics hardware for the calculation.

MRI Head	CT Backpack	Synthetic Data		
256×256×128	256×256×186	128×128×128		
$\omega=0.1\%$	$\omega=20\%$	$\omega=0.1\%$	$\omega=5\%$	$\omega=30\%$
2.386 <sub>sec</sub>	2.330 <sub>sec</sub>	0.654 <sub>sec</sub>	0.649 <sub>sec</sub>	0.646 <sub>sec</sub>

Table 1: Time measurement for the estimation of statistical properties for different data sets.

After the extraction of the statistical properties, we use raycasting for the visualization. The statistical properties  $\mu$  and  $\sigma$  are stored in additional channels of the volume. To classify a sample point with the designed transfer function, the parameters of all transfer-function regions are handed over to the graphics card. There it is tested if the statistical properties of a sample point lie within a transfer-function region. If so, the color and opacity is assigned to the sample point as described in Section 3.2. This classification can be implemented efficiently on the graphics card. We get interactive rendering rates for data sets of size  $256 \times 256 \times 256$  on a GeForce 8800 GT graphics card.

## 5 RESULTS AND DISCUSSION

In this section we show some results generated with our new method and compare them with other techniques.

### 5.1 Synthetic Data Set

For the explanation of the method we have introduced a synthetic data set, as shown in Figure 2. The three different materials in this data set are rather noisy. Therefore, it is difficult to separate the materials in common transfer-function spaces. In Figure 10(a) the 2D transfer-function space with axis  $f$  and  $|f'|$  and in Figure 10(b) the LH histogram-space were used to classify the different materials. With the 2D transfer function we were not able to classify all points correctly, because of the density overlapping. Especially at the border between material 1 and material 3 (blue and yellow) points are classified as material 2 (red). In the LH histogram-space it is easier to separate the different materials and the border but transitions are very ragged.

In Figure 10(c) the result of our method is shown for an  $\omega$  of 5%. Since the different materials have different statistical properties they can be clearly seen as clusters in the statistical transfer-function space. For the synthetic data, smoothing techniques would be able to reduce the cluster sizes in the 2D transfer-function space and in the LH histogram-space. However, the smoothing only clusters the data values. Our approach uses the standard deviation for

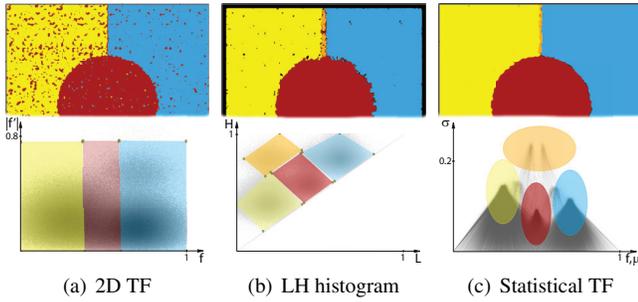


Figure 10: Classification results of the materials in the synthetic data set with different methods.

the classification as well. Therefore, we are better able to classify different materials especially if they differ by their standard deviation, such as border regions.

## 5.2 Real-World Data Sets

In the real world, noise is typically present in measured data sets. The amount of noise varies between acquisition techniques. In MRI data sets, the noise level is rather high. Therefore, it is especially difficult to classify different materials in such data sets. A common problem is the visualization of the brain in an MRI scan of the head. Figure 11 shows different results of this task for different classification techniques. Below each rendering result, the settings of the transfer function for each of these spaces are shown. In comparison to the 1D (Figure 11(a)) and 2D transfer function (Figure 11(b)) as well as the LH histogram-space (Figure 11(c)), we can better separate the brain from other tissues with our method (Figure 11(d)). In the statistical transfer-function space it is also easier to design a transfer function because the brain tissue has different statistical properties than other tissues in this data set and is more tightly clustered. As can be seen in Figure 11(d) the cluster is rather large due to the different matters in the brain but it is distinguishable from other clusters of other tissues. In comparison to this, in the 2D transfer-function space and in the LH histogram-space no cluster for the brain tissue can be seen. Additionally, Figure 11 shows that the shading based on statistical properties (Figure 11(d)) results in a smoother surface in comparison to normal gradient-based shading (Figure 11(a)-(c)).

The ability of the new method to separate different materials by their statistical properties can be used for many applications. An example is the detection of certain materials in CT scans for security checks. Figure 12 shows the scan of a backpack containing three different fluids. With a 2D transfer function, as shown in Figure 12(a), the different fluids cannot be classified without classifying also other parts of the data set. It is also hard to detect the fluids in the transfer-function space because they do not show up as clusters. In comparison, our method can clearly classify the fluids, as shown in Figure 12(b). It can also be seen that the fluids show up as clusters with very low standard deviation in the transfer-function space. This makes it much easier to define a transfer function. Furthermore, the shading with our method is slightly smoother.

Figure 13 shows a result of an MRI scan. The data set contains a tumor inside the brain. The statistical properties of the tumor are actually different from the rest of the brain which is captured in our transfer-function space. This can be seen in Figure 13. Since the tumor is rather small, only a few sample points show up in the area of the classification region of the tumor (red region). However, with other methods, such as 1D, 2D, and LH transfer functions, we were not able to clearly separate the tumor from the brain.

The results show that the new method can be used for various data sets and different tasks. The main reason for this is the con-

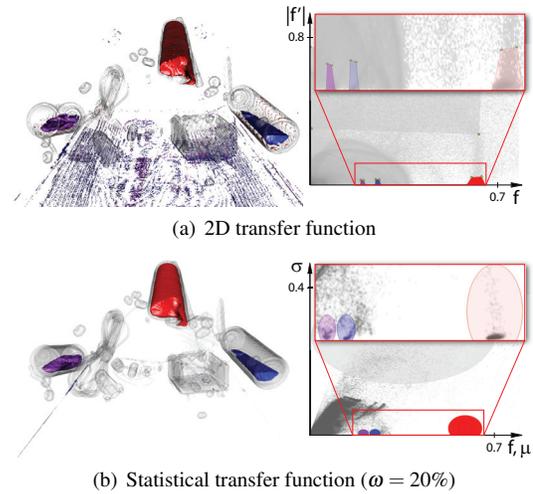


Figure 12: Detection of different fluids in a CT scan of a backpack.

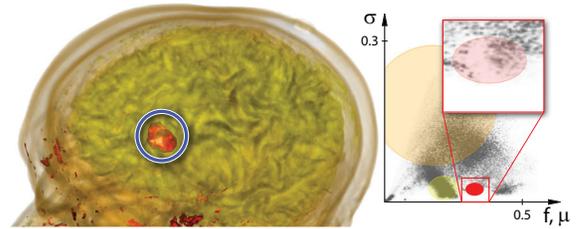


Figure 13: Detection of a brain tumor. For the extraction of the statistical properties  $\omega$  was set to 0.1%.

fidence level  $\omega$ , which can be set according to the type of data. For MRI, e.g., where the material distributions slightly differ from a Gaussian distribution, we set  $\omega$  to a low level in comparison to CT data. We are able to classify different materials even if the data type is different. This is not as easily possible in common other transfer-function spaces.

There are some limitations of the new technique. One drawback is the rather high memory consumption because for each sample point two different statistical properties have to be stored together with the data value. Thus the data size is tripled. For large data sets this could exceed the memory of a graphics card. Another penalty can occur for noise distributions very different from Gaussian white noise. In such cases the test methods have to be adapted to the given frequency distribution in the data sources. For this work we concentrated on measured data, where the distributions of data values are similar to a Gaussian distribution.

Although the confidence level is the only parameter which has to be set by the user, we want to do further research to automatically define this parameter. For this task we want to use the termination level in our future work. With the termination level at different radii it is probably possible to detect if  $\omega$  is either too high or too low for a given data set. Furthermore, we want to develop algorithms for the automatic detection of clusters in the statistical transfer-function space. By using the exit radius  $r_{break}$  of the calculation loop, we should be able to automatically find significant clusters for different materials. This additional step should accelerate the design process for transfer functions because an initial setting can be provided.

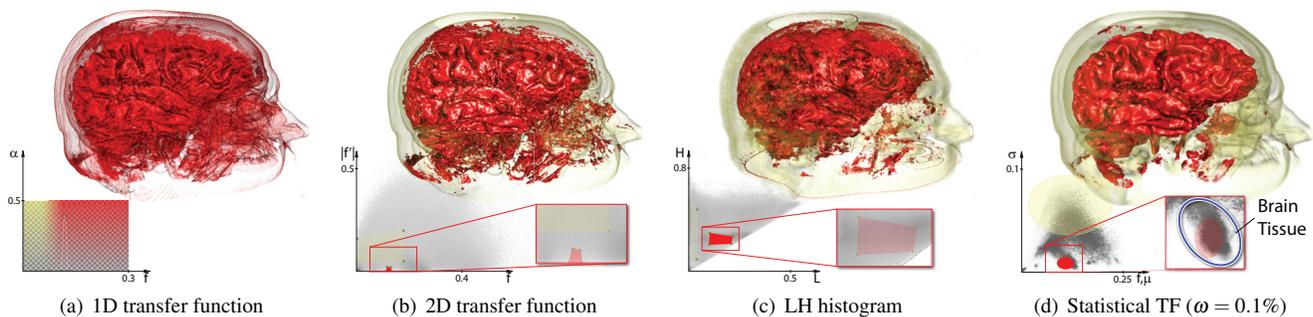


Figure 11: Comparison of the statistical transfer-function space with the 1D and 2D transfer function space as well as the LH histogram-space. The task for the generation of the results was to classify the brain in the different spaces.

## 6 CONCLUSION

In this paper we introduced a novel transfer-function space, based on statistical properties. With these statistical properties from the neighborhood of each sample point we are able to better separate different materials in comparison to other widely used transfer-function techniques. Even for noisy data sets the materials are still distinguishable by their statistical properties.

Additionally, we use the statistical properties in the shading calculations. Through this approach, the influence of noise on the shading is reduced.

In our experiments, the novel transfer-function space has proven to produce better results for different tasks than other common transfer-function techniques. Therefore, we believe the statistical transfer-function space can be used for classifying different materials in a volume data set. Through the user-specified confidence level it can be employed for data sets from various modalities.

## ACKNOWLEDGEMENTS

The work presented in this paper has been done in the scope of the SmartCT (FFG, no. 818108/16647) project and the ScaleVS (WWTF, no. ICT08-040) project.

## REFERENCES

- [1] S. Bruckner and M. E. Gröller. Instant volume visualization using maximum intensity difference accumulation. *Computer Graphics Forum*, 28(3):775–782, 2009.
- [2] J. J. Caban and P. Rheingans. Texture-based transfer functions for direct volume rendering. *IEEE TVCG*, 14(6):1364–1371, 2008.
- [3] C. Correa and K.-L. Ma. Size-based transfer functions: A new volume exploration technique. *IEEE TVCG*, 14(6):1380–1387, 2008.
- [4] B. S. Everitt. *The Cambridge Dictionary of Statistics*. Cambridge University Press, 1998.
- [5] M. H. Ghavamnia and X. D. Yang. Direct rendering of Laplacian pyramid compressed volume data. In *VIS '95: Proceedings of the IEEE Visualization '95*, pages 192–199, 1995.
- [6] H. Gudbjartsson and S. Patz. The Rician distribution of noisy MRI data. *Magnetic Resonance in Medicine*, 34(6):910–914, 1995.
- [7] M. Hadwiger, L. Fritz, C. Rezk-Salama, T. Höllt, G. Geier, and T. Pabel. Interactive volume exploration for feature detection and quantification in industrial CT data. *IEEE TVCG*, 14(6):1507–1514, 2008.
- [8] J. Hladůvka, A. König, and M. E. Gröller. Curvature-based transfer functions for direct volume rendering. In *Spring Conference on Computer Graphics 2000 (SCCG 2000)*, volume 16, pages 58–65, 2000.
- [9] J. Kniss, G. Kindlmann, and C. Hansen. Multidimensional transfer functions for interactive volume rendering. *IEEE TVCG*, 8(3):270–285, 2002.
- [10] J. Kniss, R. V. Uitert, A. Stephens, G.-S. Li, T. Tasdizen, and C. Hansen. Statistically quantitative volume visualization. In *VIS '05: Proceedings of the IEEE Visualization '05*, pages 287–294, 2005.
- [11] D. H. Laidlaw, K. W. Fleischer, and A. H. Barr. Partial-volume bayesian classification of material mixtures in MR volume data using voxel histograms. *IEEE Transactions on Medical Imaging*, 17(1):74–86, 1998.
- [12] T. Lei and W. Sewchand. Statistical approach to X-ray CT imaging and its applications in image analysis part I: Statistical analysis of X-ray CT imaging. *IEEE Transactions on Medical Imaging*, 11(2):53–61, 1992.
- [13] M. Levoy. Display of surfaces from volume data. *IEEE Computer Graphics and Applications*, 8(3):29–37, 1988.
- [14] T. Lindeberg. Scale-space for discrete signals. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12:234–254, 1990.
- [15] T. Lindeberg. Feature detection with automatic scale selection. *International Journal of Computer Vision*, 30:79–116, 1998.
- [16] E. B. Lum and K.-L. Ma. Lighting transfer functions using gradient aligned sampling. In *VIS '04: Proceedings of the IEEE Visualization '04*, pages 289–296, 2004.
- [17] E. B. Lum, J. Shearer, and K.-L. Ma. Interactive multi-scale exploration for volume classification. *The Visual Computer*, 22(9–11):622–630, 2006.
- [18] C. Lundström, P. Ljung, and A. Ynnerman. Local histograms for design of transfer functions in direct volume rendering. *IEEE TVCG*, 12(6):1570–1579, 2006.
- [19] C. Lundström, P. Ljung, and A. Ynnerman. Multi-dimensional transfer function design using sorted histograms. In *Proceedings Eurographics/IEEE Workshop on Volume Graphics 2006*, pages 1–8, 2006.
- [20] S. Muraki. Volume data and wavelet transforms. *IEEE Computer Graphics and Applications*, 13(4):50–56, 1993.
- [21] D. Patel, M. Haidacher, J.-P. Balabanian, and M. E. Gröller. Moment curves. In *Proceedings of the IEEE Pacific Visualization Symposium 2009*, pages 201–208, 2009.
- [22] B. T. Phong. Illumination for computer generated pictures. *Communications of the ACM*, 18(6):311–317, 1975.
- [23] S. Roettger, M. Bauer, and M. Stamminger. Spatialized transfer functions. In *Proceedings of Eurographics/IEEE VGTC Symposium on Visualization*, pages 271–278, 2005.
- [24] S. Tenginakai, J. Lee, and R. Machiraju. Salient iso-surface detection with model-independent statistical signatures. In *VIS '01: Proceedings of the IEEE Visualization '01*, pages 231–238, 2001.
- [25] S. Tenginakai and R. Machiraju. Statistical computation of salient iso-values. In *VISSYM '02: Proceedings of the symposium on Data Visualisation 2002*, pages 19–24, 2002.
- [26] P. Šereda, A. V. Bartolů, I. W. Serlie, and F. A. Gerritsen. Visualization of boundaries in volumetric data sets using LH histograms. *IEEE TVCG*, 12(2):208–218, 2006.
- [27] J. Wang, T. Li, H. Lu, and Z. Liang. Penalized weighted least-squares approach to sinogram noise reduction and image reconstruction for low-dose X-ray computed tomography. *IEEE Transactions on Medical Imaging*, 25(10):1272–1283, 2006.
- [28] B. L. Welch. The generalization of "student's" problem when several different population variances are involved. *Biometrika*, 34(1–2):28–35, 1947.