Multifield-Graphs: An Approach to Visualizing Correlations in Multifield Scalar Data

Natascha Sauber, Holger Theisel, and Hans-Peter Seidel

Abstract—
We present an approach to visualizing correlations in 3D multifield scalar data. The core of our approach is the computation of correlation fields, which are scalar fields containing the local correlations of subsets of the multiple fields. While the visualization of the correlation fields can be done using standard 3D volume visualization techniques, their huge number makes selection and handling a challenge. We introduce the Multifield-Graph to give an overview of which multiple fields correlate and to show the strength of their correlation. This information guides the selection of informative correlation fields for visualization. We use our approach to visually analyze a number of real and synthetic multifield datasets.

Index Terms—Visualization, multifield, correlation.

1 Introduction

Most of the simulations in computer science and engineering create as output not only one, but multiple fields describing different aspects of the simulated topic. These fields, that can be scalar, vector or tensor fields, are usually defined over the same 3D domain. There are cases when it is insufficient or inconvenient to visualize the fields separately. Some fields may be redundant, while others may have an important relation to each other which is worth visualizing. This can be illustrated at an example of several scalar fields which contain local changes. Only by regarding the scalar fields separately, it is not obvious if there are locations where some or even all the fields change similarly. A method to detect these common local changes it to visualize them (see Figure 1).

There are many established approaches for visualizing a single scalar field, like slicing, isosurfacing [19] or direct volume rendering [18]. Comparisons of different fields visualized with these strategies are possible by looking at them side by side, or sequentially in time. Although this is commonly done in the visualization community, the perception of correlation between the fields is rather limited. In the first case subtle differences cannot be perceived while in the second case only pairs of fields can be simultaneously compared. It is hardly possible to obtain knowledge of how all of these fields are related. For this it is necessary to have strategies to visualize multiple fields simultaneously or to visualize the correlations between them. There are some approaches to visualizing a small number of fields at the same time, and some which can show the correlation of two fields. However, the visualization of correlations between multiple fields is still a challenge and is rarely addressed in the visualization community [14].

Our method, the Multifield-Graph, is a step in this direction. We consider a number of scalar fields defined over the same 3D domain as input and aim for the visual analysis of the correlations between these fields. This way, we are able to find features which are not contained in one field, but which can be computed only when multiple fields are taken into account. This information is stored in correlation fields. With the Multifield-Graph it is possible to get an overview of the amount of correlation which is contained in the exponentially high number of correlation fields. The Multifield-Graph can also guide the selection of promising correlation fields which can be shown with one of the standard volume visualization methods.

This paper is organized as follows: Section 2 gives an overview of previous work done in multi volume rendering and multifield visualization. The computation of correlation fields is explained in Section 3. The Multifield-Graph, our approach to deal with the huge number of correlation fields, is described in Section 4. Sections 5 shows the application of our approach at some examples.

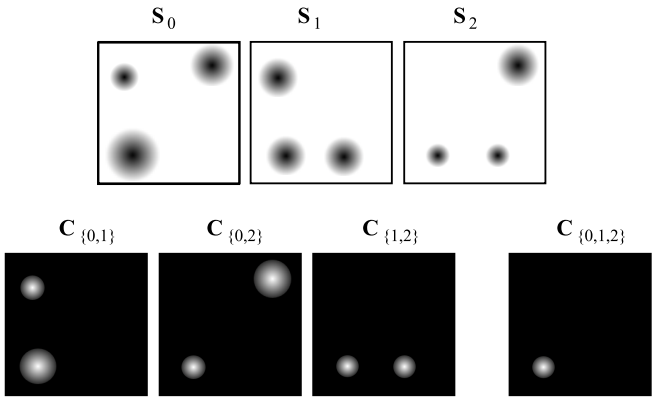

Fig. 1. This Figure shows three scalar fields $\mathbf{S}_i$ $(i \in \{0,1,2\})$, and correlation fields $\mathbf{C}_N$, which contain high values at locations, where the corresponding scalar fields $\mathbf{S}_i (i \in N)$ show similar changes.

2 Previous Work

An overview of the work which has been done in multivariate multidimensional visualization is given in [27].

A number of approaches for multifield visualization of 2D scalar fields have been introduced, which use for example partial look-through layers [15, 28] or combinations of texture elements [23, 26]. These techniques yield effective visual representations for a rather small number of fields. They tend to deliver cluttered visualizations if the number of fields is too high. Moreover, due to spatial occlusion, these methods do not extend well to 3D multifields.

There are a couple of approaches available for the simultaneous visualization of several 3D volumes. In [3] a technique is presented for combining intersecting volumetric objects. Another paper [8] introduced v-objects, which allow to efficiently render an arbitrary number

- *Natascha Sauber is with MPI Informatik Saarbrücken, E-mail: sauber@mpi-inf.mpg.de.*
- *Holger Theisel is with MPI Informatik Saarbrücken, E-mail: theisel@mpi-inf.mpg.de.*
- *Hans-Peter Seidel is with MPI Informatik Saarbrücken, E-mail: hpseidel@mpi-inf.mpg.de.*

Manuscript received 31 March 2006; accepted 1 August 2006; posted online 6 November 2006.
For information on obtaining reprints of this article, please send e-mail to: tvcg@computer.org.

of possibly intersecting objects. [21] developed a hardware based volume rendering approach to display multiple volumes simultaneously with correct overlaying. A similar approach was presented in [1]. Here a scalar and a vector field are visualized simultaneously by creating a 3D texture out of the flow field and using it to color the volume rendering of scalar field. These approaches provide a technical solution for the simultaneous visualization of multiple volumes. They deal with the occurring cluttering and occlusion by the use of clipping planes and transparency, which is effective if the number of fields is not too huge.

There are also some approaches for visualizing multiple volume data which focus more on correlations between them. A recent work [25] introduced the concepts of object and dimension correlation during projection from multidimensional-space to three-dimensional space. An approach for using multi-dimensional transfer functions for direct volume rendering was introduced in [17, 16]. In this work, the opacity of a voxel depends not only on the scalar value of one field, but on the scalar values of two fields, and optionally on the first derivative. A method similar to this but more flexible and able to handle a larger number of fields was presented in [6, 5]. Here scatterplots of every field pair can be analyzed to gain insight into global correlations of the fields. Looking at the scatterplots it is possible to specify features, like focus ranges, for every field and visualize only data points whose field values lie within these ranges. Data points within focus ranges can also be seen as correlating data. Another tool [12] was designed to explore correlations between two scalar fields. By observing one of the fields with direct volume visualization, a region of interest can be defined. There a 2D plane can show the correlations of the two fields with a colored heightfield, while the color depends on one field and the height on the other. Along a curve on this plane the correlations can be shown additionally as line plots of the two functions.

3 Correlation Fields

The goal of our method is to visualize the amount and the spatial structure of local correlations in multifield scalar data. We use the term *local correlation* between fields for regions where these fields contain similar changes in scalar value. Both the occurrence and the absence of correlations between fields may be important information.

The input data for our approach are n scalar fields $\mathbf{S}_1,\ldots,\mathbf{S}_n$ over the same 3D domain $\mathbf{D}$. We propose to use *correlation fields* to make correlations visible which occur between a number of fields. A correlation field $\mathbf{C}_N$ with $N \subseteq \{1,\ldots,n\}$, $|N| \geq 2$, is a scalar field over $\mathbf{D}$ which describes the strength of the local correlation between the $|N|$ scalar fields $\mathbf{S}_i (i \in N)$ at each location of $\mathbf{D}$. The visualization of correlation fields is done with standard scalar 3D visualization methods.

To make the approach applicable, two problems have to be solved. Firstly a measure c for local correlation has to be chosen that is able to compute a correlation field $\mathbf{C}_N$ out of the $|N|$ scalar fields $\mathbf{S}_i (i \in N)$, such that $\mathbf{C}_N = c(\{\mathbf{S}_i : i \in N\})$. Section 3.1 gives an overview of common correlation measures and describes in detail those which we have used. Secondly, it has to be decided which correlation field should be computed and visualized, because the total number of correlation fields grows exponentially with n, and is therefore too large for exhaustive computation of all correlation fields. We introduce the Multifield-Graph, described in Section 4, to direct this decision by giving information about the amount of correlation which is present in the correlation fields.

3.1 Correlation Measure

There are several correlation measures available which are defined for two variables [11]. A very common measure, the sum of squared differences, cannot be used in our case, because the data has to have the same unit (e.g. temperature in °C). Another measure, called correlation coefficient can detect linear dependencies under the assumption of normally distributed data. The correlation ratio can detect functional dependency. Mutual information, which is based on entropy, is able to find more general dependencies. It is possible to define all of these measures locally [2, 11]. Methods which are able to detect more general correlations have the drawback that they need a larger region of input values to produce meaningful results.

It is also possible to define a correlation criteria which is based on gradient similarity [4], and which is thus inherently local. Because of the use of gradients this method also implies the spatial location of the intensity value pairs. In contrast to this, all methods mentioned in the last paragraph use only pairs of scalar values.

Since we want to compute correlation fields for n input fields, the correlation measure has to be extendable to n dimensions. The Multifield-Graph can be used with an arbitrary n-dimensional local correlation measure. We choose to use two measures: one based on gradient similarity, described in Section 3.1.1, and another one based on local correlation coefficients, described in Section 3.1.2. The differences between them are explained in Section 3.1.3. These two measures are appropriate for optimizations during the Multifield-Graph computation, described in Section 4.3.

3.1.1 Gradient Similarity

The first measure we use is a *gradients similarity* measure (GSIM). A gradient field $\mathbf{G}_i$ of a field $\mathbf{S}_i$ is defined at every position $\mathbf{x}$ as $\mathbf{G}_i = \nabla \mathbf{S}_i$, with $\mathbf{g}_i = \mathbf{G}_i(\mathbf{x})$ denoting one gradient of $\mathbf{S}_i$ at a position $\mathbf{x}$. The computation of the correlation field based on gradients, requires a similarity measure $s(\mathbf{g}_i, \mathbf{g}_j)$ for the comparison of two gradients.

In recent work [7] a local comparison measure based on gradients was proposed. This measure assigns high values to strong orthogonal gradients. Therefore the number of input fields n is limited to the number of spatial dimensions of $\mathbf{D}$, and therefore not usable for our purpose.

We decided to use a similarity measure, that assigns high values to equally strong and similarly directed gradients. This measure is extendable to an arbitrary number of fields n. The measure should be orientation independent, such that two gradients result in the same measurement in both orientation cases, $s(\mathbf{g}_i, \mathbf{g}_j) = s(\mathbf{g}_i, -\mathbf{g}_j)$. This property is useful for the Multifield-Graph computation. There, the local correlation measure is summed over the whole domain to indicate the global amount of correlation. If there existed both, positive and negative correlation, they might cancel out each other.

To make the gradient magnitudes comparable, the fields have to be normalized before the computation. We decided to scale the range of all scalar fields to $[0,1]$. Then our measure returns 0 if both gradients have a zero magnitude, the measure is 1 if both gradients have the same magnitude, and it has a small value if they have a different magnitude.

There are some vector similarity measures in literature [4, 24], but they are not independent from the orientation of the vectors. Hence we define the following similarity measure for two gradients $\mathbf{g}_i$ and $\mathbf{g}_j$:

$$
\begin{aligned}
s(\mathbf{g}_i, \mathbf{g}_j) &= (s_d(\mathbf{g}_i, \mathbf{g}_j) \cdot s_m(\mathbf{g}_i, \mathbf{g}_j))^r, \qquad (1)\\
s_d(\mathbf{g}_i, \mathbf{g}_j) &= \left(\frac{\mathbf{g}_i{}^T \mathbf{g}_j}{||\mathbf{g}_i|| \cdot ||\mathbf{g}_j||} \right)^2,\\
s_m(\mathbf{g}_i, \mathbf{g}_j) &= 4 \frac{||\mathbf{g}_i|| \cdot ||\mathbf{g}_j||}{(||\mathbf{g}_i|| + ||\mathbf{g}_j||)^2},
\end{aligned}
$$

with s_d depicting the direction similarity and s_m depicting the magnitude similarity. The exponent r regulates the sensitivity of the measure. Its usage will be explained below. This measure compares only two gradients. It is extended to n gradient vectors by taking the minimum gradient pairs similarity:

$$
\mathbf{C}_N(\mathbf{x}) = min\{s(\mathbf{g}_i, \mathbf{g}_j) : i \neq j; i, j \in N\}. \qquad (2)
$$

The resulting value means that there is no gradient pair whose gradients are less similar than this value. This serves for detecting regions in the correlation field that contain high correlations.

It should be noted that this measure does not scale linearly by the number of input fields especially between 2 and 3 fields. The measure should be minimal if the gradient directions are equally distributed in

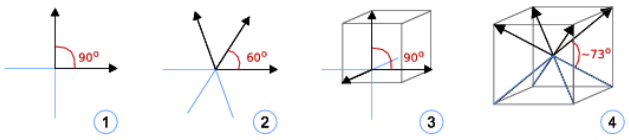

Fig. 2. Examples for couples of gradient vectors with low directional similarity. All pairs of gradients have a high α_{min} (denoted with red arcs). The angle α_{min} is the angle between the two straight lines of two gradients.

the plane or in the volume. For two gradients $\mathbf{g}_i, \mathbf{g}_j$, the correlation is minimal if the angle $\alpha_{min} = min(\sphericalangle(\mathbf{g}_i, \mathbf{g}_j), \sphericalangle(\mathbf{g}_i, -\mathbf{g}_j))$ approaches 90° (see Figure 2 (1)). In the case of more than two gradients, there are two possible arrangements which should be the least similar. In the first case is, their directions are equally distributed in the plane (see Figure 2 (2)), in the second case their directions are equally distributed in $3D$ space (see Figure 2 (3)), especially if they all have a similar α_{min} with respect to each other. In a $2D$ domain this is the case for three gradients each with an α_{min} of 60° (see Figure 2 (2)). In a $3D$ domain the same situation exists for example with four gradients having an α_{min} of about 73.2° (see Figure 2 (4)). As a consequence, the measure should be already close to zero at an enclosing angle of 60°. This is reached with equation 1 if the exponent r is sufficiently high. Throughout this paper we have chosen $r = 4$.

We illustrate our approach to creating gradient correlation fields with an example. Figure 3 (first row) shows slices through four simple 3D scalar fields, where the scalar value is coded in the gray value. The colored arrows show the gradients of the scalar fields. The second row shows slices of our computed correlation fields. In regions with high correlation the gradients have similar direction and magnitude. The third row shows slices of the same correlation fields, but computed with local correlation coefficient (Section 3.1.2).

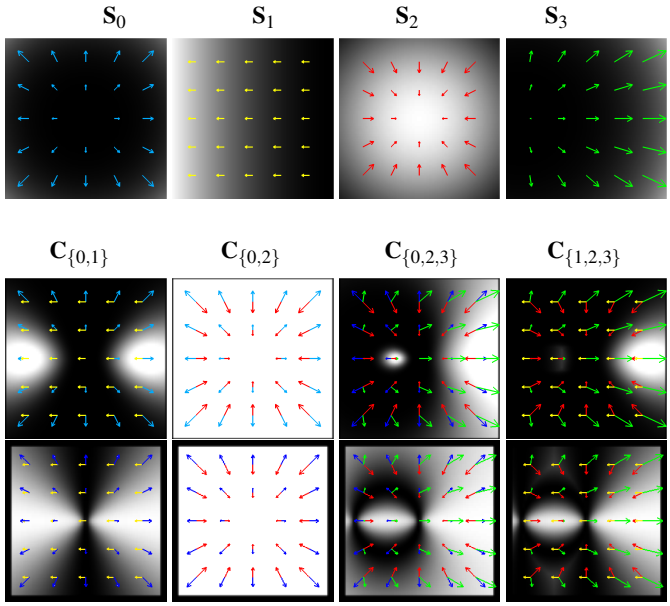

Fig. 3. First row: slices through 4 scalar fields and their gradients. Correlation fields for different combinations of the 4 scalar fields as slices computed with GSIM (second row) and computed with LCC (third row). The visible differences between the two measures result from two reasons. Firstly, GSIM drops faster to zero than LCC, which results from exponent r of Equation 1, and secondly GSIM is sensitive to differences in gradient magnitude in contrast to LCC which is not.

3.1.2 Local Correlation Coefficient

The second correlation measure called LCC is based on a local correlation coefficient, as proposed in [2]. It detects linear dependencies between the values of two fields within a local region. A similar approach was also used in pattern recognition [20]. There, a volume correlation filter is used to find a certain pattern within a volume.

The correlation coefficient is 1 in the case of an increasing linear relationship and -1 in the case of a decreasing linear relationship. It is 0 if there is no linear dependency. It is computed using a window function $G_{\mathbf{x}}$, which is centered around $\mathbf{x}$ and normalized such that $\int_{\mathbf{p}\in\mathbf{D}} G_{\mathbf{x}}(\mathbf{p})d\mathbf{p} = 1$. We define the mean $\bar{\mathbf{S}}_i$, the covariance cov_{ij}, the standard deviation σ_i, and the correlation coefficient $corr_{ij}$, locally at a location $\mathbf{x}$ in the following way:

$$\begin{aligned}
\bar{\mathbf{S}}_i(\mathbf{x}) &= \int_{\mathbf{p}\in\mathbf{D}} G_{\mathbf{x}}(\mathbf{p}) \cdot \mathbf{S}_i(\mathbf{p}) d\mathbf{p}, \\
cov_{ij}(\mathbf{x}) &= \int_{\mathbf{p}\in\mathbf{D}} G_{\mathbf{x}}(\mathbf{p}) \cdot (\mathbf{S}_i(\mathbf{p}) - \bar{\mathbf{S}}_i(\mathbf{x})) \cdot (\mathbf{S}_j(\mathbf{p}) - \bar{\mathbf{S}}_j(\mathbf{x})) d\mathbf{p}, \\
\sigma_i(\mathbf{x}) &= \sqrt{cov_{ii}(\mathbf{x})}, \\
corr_{ij}(\mathbf{x}) &= \frac{cov_{ij}(\mathbf{x})}{\sigma_i(\mathbf{x}) \cdot \sigma_j(\mathbf{x})}.
\end{aligned} \tag{3}$$

Throughout this paper we will use a box filter as window function $G_{\mathbf{x}}$. To extend this measure for a comparison of n fields, the minimum of the absolute correlation value of pair fields is used:

$$\mathbf{C}_N(\mathbf{x}) = min\{|corr_{ij}(\mathbf{x})| : i \neq j; i, j \in N\}. \tag{4}$$

We use the absolute value of the correlation coefficient, for the same reasons that we use an orientation independent gradient similarity measure, namely to avoid a canceling out effect between positive and negative correlation during Multifield-Graph computation. We illustrate this measure in the same example as the GSIM measure (see Figure 3 third row)

3.1.3 Comparison of measures

The measures described in the last two sections produce similar correlations in most cases. The measures differ in three aspects: in the dependence on scaling of value ranges, how fast the correlation measure drops to zero, and in the size of the area of influence.

One aspect where the measures differ is the sensitivity to the scaling of the data value range. The LCC is independent of scaling. GSIM assigns high values only if the gradient magnitude of the fields are equally strong. This can be seen while comparing the correlation field $\mathbf{C}_{\{0,1\}}$ in Figure 3 at the horizontal line in the center of the field where all gradients are parallel. Another difference between this measures is, that GSIM drops faster to zero than LCC. This difference depends on the coefficient r of Equation 1 and can also be seen in the correlation fields of Figure 3.

The other main difference is the size of the area of influence. The gradient is defined at a point, contrary to the local correlation coefficient which is defined over a region. The size of the region of influence leads to a blurring effect. The LCC measure assigns high values also to the neighborhood of strong single gradients (see Figure 4 (1)). For the same reasons LCC can detect similarities between slightly shifted gradients, between a smooth gradient and a smooth gradient overlayed with a higher frequency and between a smooth and a single strong gradient (see Figure 4 (2), (3), (4)).

For detecting extensive changes LCC is more suited while for subtle local changes GSIM is to prefer.

4 The Multifield-Graph

Let $\mathscr{C} := \{\mathbf{C}_N : N \subseteq \{1,\ldots,n\}\}$ be the set of all possible correlation fields. Since the number of all fields $|\mathscr{C}|$ grows exponentially with n, it is not practical to compute and examine each of these fields, and it is not known in advance which correlation fields will contain

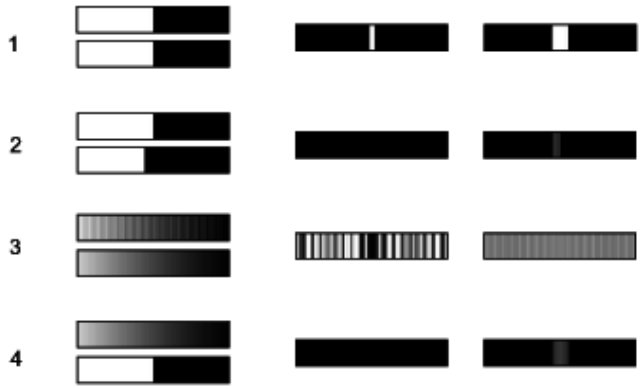

Fig. 4. Four pairs of functions (left column) are compared with GSIM (middle column) and LCC (right column).

meaningful information. To deal with this problem we introduce the Multifield-Graph. The Multifield-Graph has two functions. Firstly, it gains information about the amount of correlation contained in each correlation field. We call this information in the following *correlation overview information*. The nodes of the graph are displayed as icons with correlation overview information graphically encoded. Secondly, the Multifield-Graph allows the focusing on the display of certain nodes. This is useful, because it is not possible to perceive the information of all of them at once, because of their exponential number. The different focus strategies that can be used to reduce the displayed nodes are described in Section 4.2. The optimized computation of the Multifield-Graph are described in Section 4.3.

4.1 Definition and construction

Let $\mathscr{G} := (\mathscr{V}, \mathscr{E})$ be the Multifield-Graph where $\mathscr{V}$ denotes the set of all nodes and $\mathscr{E}$ the set of all edges. Every node $V_{N_i} \in \mathscr{V}$ of the graph corresponds to a correlation field $\mathbf{C}_{N_i}$. Two nodes V_{N_i} and V_{N_j} are connected with an edge $E_{ij} \in \mathscr{E}$ if $|N_i| = |N_j| - 1$ and $N_i \subset N_j$. The nodes are displayed as squares, the edges as lines connecting the nodes. All nodes with the same $|N|$ are displayed in the same row.

To provide insight into the correlation contained in one correlation field, some correlation overview information is computed. For this, an arbitrary characteristic of the correlation field can be used, depending on the interpretation aim of the application. Possible characteristics usable as correlation overview information could be the distribution of the correlation values, the size of connected correlating areas, certain spatial structures of correlating areas or frequency of changes in the correlation field.

We choose to employ two other characteristics which are intuitive and straightforward. They are both based on $\mathbf{D}_{N,\theta} = \{\mathbf{x} \in \mathbf{D} : \mathbf{C}_N(\mathbf{x}) \geq \theta\}$ denoting the subset of $\mathbf{D}$ in which $\mathbf{C}_N$ exceeds a certain threshold θ. One characteristic is the ratio between the volumes of $\mathbf{D}_{N,\theta}$ and $\mathbf{D}$:

$$rc_N = \frac{vol(\mathbf{D}_{N,\theta})}{vol(\mathbf{D})}. \tag{5}$$

The other characteristic is the average of the values in $\mathbf{D}_{N,\theta}$

$$ac_N = \frac{\int_{\mathbf{x} \in \mathbf{D}_{N,\theta}} \mathbf{C}_N(\mathbf{x}) d\mathbf{x}}{vol(\mathbf{D}_{N,\theta})}. \tag{6}$$

These characteristics provide useful information if θ is sufficiently high. Then rc_N indicates the volume of highly correlating areas, namely areas with correlation above θ. ac_N suggests whether the field contains a reasonable amount of correlation even higher than θ.

The value of the two characteristics is displayed as a colored disc, where the size of the disc depends on rc_N and the color on ac_N. Red depicts high correlation and green depicts low correlation. A large disc means many correlating values, a small disc the opposite.

To illustrate this explanation, two example Multifield-Graphs are shown in Figure 5. The graphs show the correlation between the four correlation fields of Figure 3, whereas the left one is computed with the GSIM measure, the right one with the LCC measure. To ease the interpretation of the colors, a bar that displays the color coding of the ac_N value is displayed next to each graph. A large amount of correlation between field $\mathbf{S}_0$ and $\mathbf{S}_2$ can be observed as well as a rather small amount of correlation between the fields $\mathbf{S}_0$ and $\mathbf{S}_1$. Furthermore a reasonable amount of correlation can be detected between the scalar fields ($\mathbf{S}_0$, $\mathbf{S}_2$, $\mathbf{S}_3$). In order to see where these correlations occur, the correlation field $\mathbf{C}_{\{0,2,3\}}$ can be visualized (see Figure 3). The two graphs look similar, but differ in scaling of ac_N and rc_N, because the GSIM measure assigns lower values than the LCC measure. To see the differences between the two measures in detail, one must analyze the correlation fields.

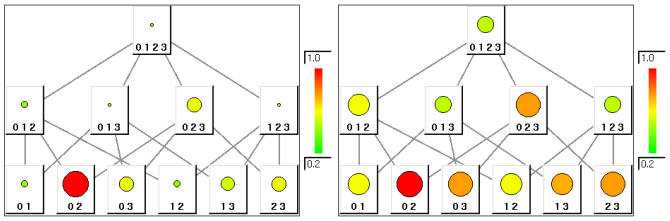

Fig. 5. The Figure shows Multifield-Graphs of the synthetic scalar fields of Figure 3. The left one is computed with GSIM, the right one is computed with LCC, both with threshold $\theta = 0.2$.

4.2 Graph focus

For a small number of scalar fields, for instance $n = 6$, the whole graph can be displayed, giving an overview of all existing correlations and relations between the fields (see Figure 6 (top)). As n gets larger, it is no longer possible to display all nodes at once because of their exponential number. We use different focus strategies for reducing the number of displayed nodes to ease the detection of meaningful correlation fields: firstly to show only nodes whose contained correlation overview information fulfills some constraints, secondly to show only nodes whose set of identities of original fields N fulfills some constraints. One possibility of following the first strategy is to display only nodes with $rc_N > \eta$ and with $ac_N > \xi$, where η and ξ are certain thresholds (see Figure 6 (bottom)). This is useful for detecting high correlations. One application of the second strategy is to show only nodes V_{N_j} with $N_j \subseteq N_i$. This is useful for showing relations in one subset N_i of the fields. In Figure 6 such relations are made visible by darkening the node icons and edges of all subset fields of $\mathbf{C}_{\{1,3,4,5\}}$. This strategy can also be used to display only nodes with $k_1 > |N| > k_2$, where $k_1, k_2 \in \{2 \ldots n\}$.

Another focus strategy, spatial focusing, acts on the domain $\mathbf{D}$ instead of the graph display. Hereby the Multifield-Graph is computed only for a subdomain $\mathbf{D}_s \subseteq \mathbf{D}$ of the domain $\mathbf{D}$. These spatially restricted Multifield-Graphs are also called local Multifield-Graphs. Local Multifield-Graphs may differ from each other as well as from the Multifield-Graph of the whole domain. For small subdomains $\mathbf{D}_s$ the computation of local Multifield-Graphs is fast enough to allow an interactive browsing through the domain $\mathbf{D}$, to explore local changes of the correlations between all fields at once.

4.3 Optimization

The Multifield-Graph is designed to allow the exploration of the space of correlation fields. The time critical aspects are the construction of the Multifield-Graph, primarily the computation of the correlation overview information, and the computation of the correlation fields which should be visualized.

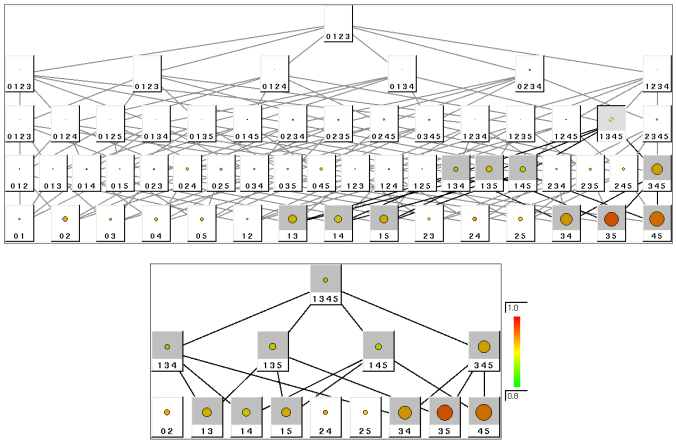

Fig. 6. First row: the whole Multifield-Graph built out of 6 scalar fields. Second row: same Multifield-Graph focused on nodes with $rc_N > 0.18$. In both graphs $\mathbf{C}_{\{1,3,4,5\}}$ and all corresponding subset fields are emphasized.

For the computation of a correlation field $\mathbf{C}_N$ the correlation measure c is evaluated at every point $\mathbf{x} \in \mathbf{D}$. For computing the correlation overview information of the Multifield-Graph, all fields $\mathbf{C}_N$ at every point $\mathbf{x}$ have to be computed every time the threshold θ changes. Since it is not possible to store all correlation fields, they must be computed on the fly.

There are some strategies to shorten the computation time. The simplest strategy is to reduce the number of points used for the computation of the correlation overview information. The reduction is done by subsampling. This is maintainable because the Multi-Graph only gives an overview anyway.

Another optimization is possible because of the decomposability property of the correlation measures we use. Each correlation field $\mathbf{C}_N$ can be constructed out of the contained pair correlation fields $\mathbf{C}_{N_k}$ with $|N_k| = 2$ and $N_k \subseteq N$ (see Equations 2 and 4). If the pair correlation fields fit into main memory, they can be precomputed and stored. And they can be reused for the computation of fields with $|N| > 2$. Then the computation time of the Multifield-Graph is independent from the computation time of a correlation field. This is especially advantageous if the LCC measure with a big window size is used.

5 Application

We used our Multifield-Graph approach for analyzing an analytic and a simulation dataset (Section 5.1 and 5.2).

5.1 ABC-flow features

The first multifield dataset we analyzed is a collection of 8 scalar fields describing characteristic properties of a 3D time-dependent flow. In fact, we consider the so-called ABC (Arnold-Beltrami-Childress) flow field

$$\mathbf{v}(\mathbf{x},t) = \begin{pmatrix} (A+(1-e^{-t/10})\sin(2\pi t))\sin z + C\cos y \\ B\sin x + (A+(1-e^{-t/10})\sin(2\pi t))\cos z \\ C\sin y + B\cos x \end{pmatrix} \quad (7)$$

which has recently attracted attention in the fluid dynamics community because it describes an unsteady solution of Euler's equation [10]. We consider the following scalar fields with resolution of 64^3. They are derived from $\mathbf{v}$ for $A = \sqrt{3}, B = \sqrt{2}, C = 1$ at the time $t = 0$ within the domain $[0, 2\pi]^3$:

- $\mathbf{S}_0$: λ_2 - a measure for vortical structures in the flow [13].
- $\mathbf{S}_1$: $\|\mathbf{v}\|$: the velocity magnitude of the flow
- $\mathbf{S}_2$: the Lyapunov exponent, a measure of the stableness of a flow [9]
- $\mathbf{S}_3, \mathbf{S}_4, \mathbf{S}_5$: x-, y-, z-component of the average flow direction of a path line over a integration time of 2π.
- $\mathbf{S}_6$: average particle velocity magnitude of the path line integrated over a 2π time.
- $\mathbf{S}_7$: distance between start and end point of a path line integration over a 2π time.

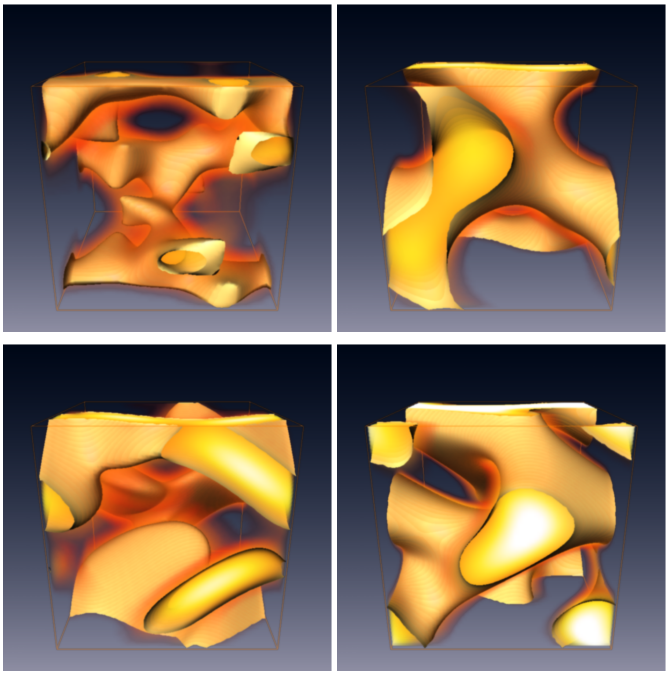

Fig. 7. Some scalar feature fields of the ABC-flow, $\mathbf{S}_0$ (top left), $\mathbf{S}_1$ (top right), $\mathbf{S}_2$ (bottom left) and $\mathbf{S}_7$ (bottom right).

The spatial structure of some of the ABC-flow-feature fields is illustrated in Figure 7. Figure 8 shows the Multifield-Graph of the data set. It is computed with a threshold $\theta = 0.8$ and focused on nodes with an average correlation $ac_N > 0.87$ and ratio of correlating values $rc_N > 0.11$, which means that 11 per cent of the volume have a correlation value higher than 0.8. We used the local correlation coefficient as correlation measure (see Section 3.1.2), with a window size of 7x7x7 samples.

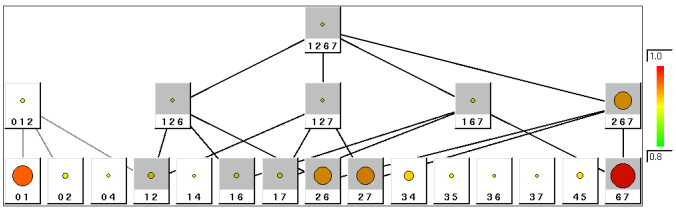

Fig. 8. Multifield-Graph of the ABC-flow-feature multifield computed with LCC and $\theta = 0.8$, focused on nodes with $ac_N > 0.87$ and $rcN > 0.11$

Figure 8 shows that there are rather strong correlations between $\mathbf{S}_0$ and $\mathbf{S}_1$. This is clearly a characteristic of the ABC flow since in general λ_2 and the magnitude of a vector field are uncorrelated because λ_2 is Galilean invariant. The visualization of $\mathbf{C}_{\{0,1\}}$ (Figure 9) shows dented-sphere-shaped areas of weaker correlation between $\mathbf{S}_0$ and $\mathbf{S}_1$. Figure 8 also shows a rather low correlation between $\mathbf{S}_1$ and $\mathbf{S}_6$. However, an analysis of $\mathbf{C}_{\{1,6\}}$ (Figure 9) reveals a clear structure in the

domain. There is even a relevant correlation between the 4 fields $\mathbf{S}_1$, $\mathbf{S}_2$, $\mathbf{S}_6$, $\mathbf{S}_7$ (see Figure 9). The strongest correlation was detected between $\mathbf{S}_6$ and $\mathbf{S}_7$: all sample points reported a correlation above the threshold. All correlation fields of Figure 9 were visualized with Isosurfaces using an isovalue of 0.8.

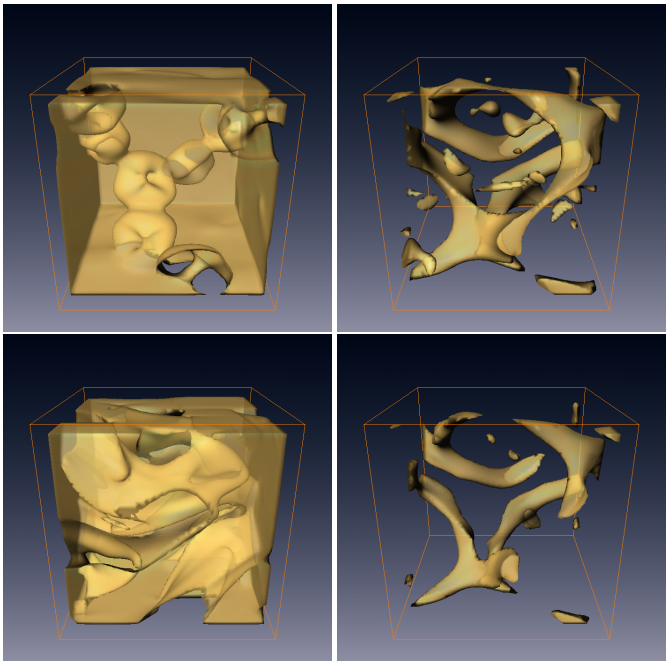

Fig. 9. Correlation fields $\mathbf{C}_{\{0,1\}}$, $\mathbf{C}_{\{1,6\}}$, $\mathbf{C}_{\{2,7\}}$, $\mathbf{C}_{\{1,2,6,7\}}$ computed with local correlation coefficient, showing correlations between scalar feature fields of the ABC-flow.

Since the GSIM measure is more sensitive to subtle differences than the LCC measure, we used GSIM to analyze correlations between $\mathbf{S}_6$ and $\mathbf{S}_7$. This gives tube-like structures of areas of less similarity of the gradients, as shown in Figure 10 (left). In contrast the GSIM measure, the LCC indicates all values to be above the threshold. The closeup shown in Figure 10 (right) shows that outside the tube the gradients of $\mathbf{S}_6$ and $\mathbf{S}_7$ are rather parallel, while their direction differs inside.

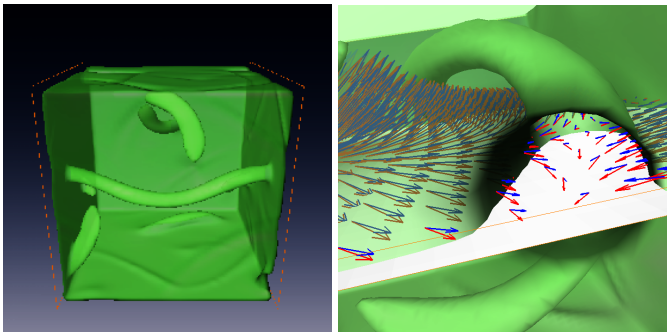

Fig. 10. Correlation field $\mathbf{C}_{\{6,7\}}$ computed with GSIM (left) and a closeup of this field (right) with gradients of $\mathbf{S}_6$ shown blue and gradients of $\mathbf{S}_7$ shown red. The similarity of gradients within the surface and the dissimilarity outside the surface, within the tube structure, are visible.

5.2 Hurricane

The second multifield dataset we analyzed is a simulation of the hurricane Isabel which took place in September of 2003 in the west Atlantic region. The simulation was done by the National Center for Atmospheric Research in the United States. The data consists of several time-varying scalar and vector fields. For our analysis we used the first time step of the following scalar fields:

- $\mathbf{S}_0$: cloud: total cloud moisture mixing ratio
- $\mathbf{S}_1$: wind velocity
- $\mathbf{S}_2$: precipitation: total precipitation mixing ratio
- $\mathbf{S}_3$: pressure
- $\mathbf{S}_4$: vapour
- $\mathbf{S}_5$: temperature

The data is defined over a 500x500x100 grid, which is equivalent to the physical scale of 2139km (east-west) x 2004km (north-south) x 19.8km. For our analysis we subsampled the data to 100x100x25 for the LCC analysis and 250x250x50 for the GSIM analysis. The visualizations of this section are all done with direct volume rendering, because these technique is suited for showing smoothly varying fields. The used transfer function assigns opacity to values with a correlation higher than 0.9.

To get an overview of the spatial structure of the analyzed fields, they are shown in Figure 11 using a direct volume rendering approach.

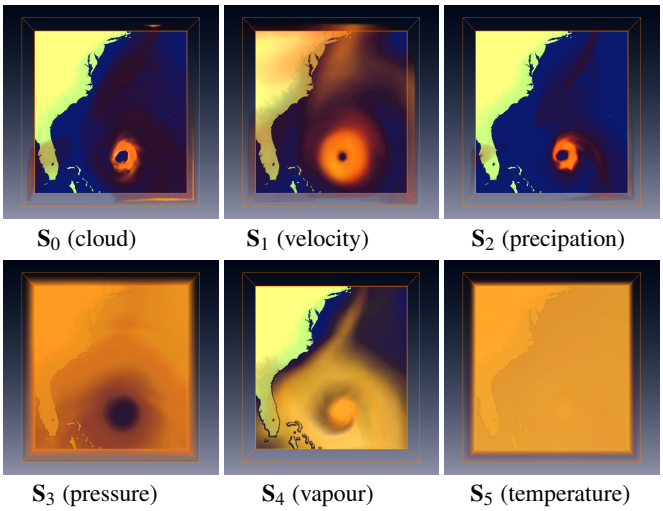

Fig. 11. Multiple Hurricane scalar fields.

We computed the Multifield-Graph of these fields using the LCC measure with a box filter of size of 5x5x5 and a threshold $\theta = 0.8$. The view on the graph is focused at nodes with $rc_N > 0.1$ and can be seen in Figure 12. It depicts strong correlation between $\mathbf{S}_3$ and $\mathbf{S}_5$, which can be seen in detail in the visualization of the corresponding correlation field $\mathbf{C}_{\{3,5\}}$ in Figure 13 (top left). These two scalar fields, containing pressure and temperature data, are highly correlated in the area not directly affected by the hurricane and in the eye of the storm itself, whereas low correlation can be detected in the highly swirling area around the center of the storm. The correlation fields of the field triples $\{3,4,5\}$ and $\{0,2,5\}$ contain also a reasonable amount of correlation, as visible in Figures 13 (top right and bottom left). Relevant correlations occur between a quadruple of scalar fields $\{1,3,4,5\}$ (see Figure 13 bottom right). Remarkably, the highest correlation in this set of fields is not found within the hurricane, but over the continent.

An application of spatial focusing is shown in Figure 14. Some correlation fields, like $\mathbf{C}_{\{4,5\}}$, contain several layers of highly correlating values. We compute the local Multifield-Graphs for each of these layers with a threshold $\theta = 0.95$ to see if the field correlations differ between these layers. Figure 14 (top) shows the three different layers and their corresponding local multifield graphs all focused at nodes

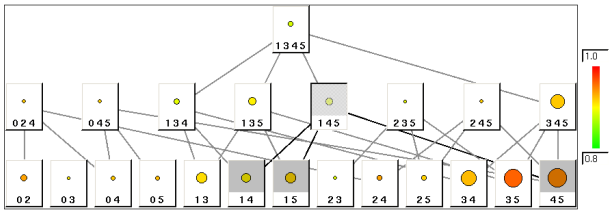

Fig. 12. Multifield-Graph computed with LCC and $\theta = 0.8$ from 6 scalar fields from the hurricane dataset. The graph is focused on nodes with $rc_N > 0.11$.

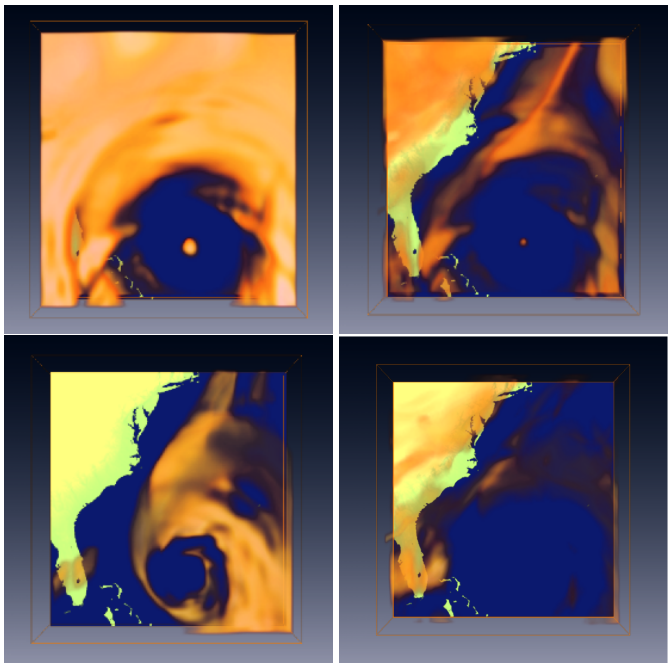

Fig. 13. Correlation fields of the hurricane multifield computed with LCC: $\mathbf{C}_{\{3,5\}}$ (top left), $\mathbf{C}_{\{3,4,5\}}$ (top right), $\mathbf{C}_{\{0,2,4\}}$ (bottom left) and $\mathbf{C}_{\{1,3,4,5\}}$ (bottom right).

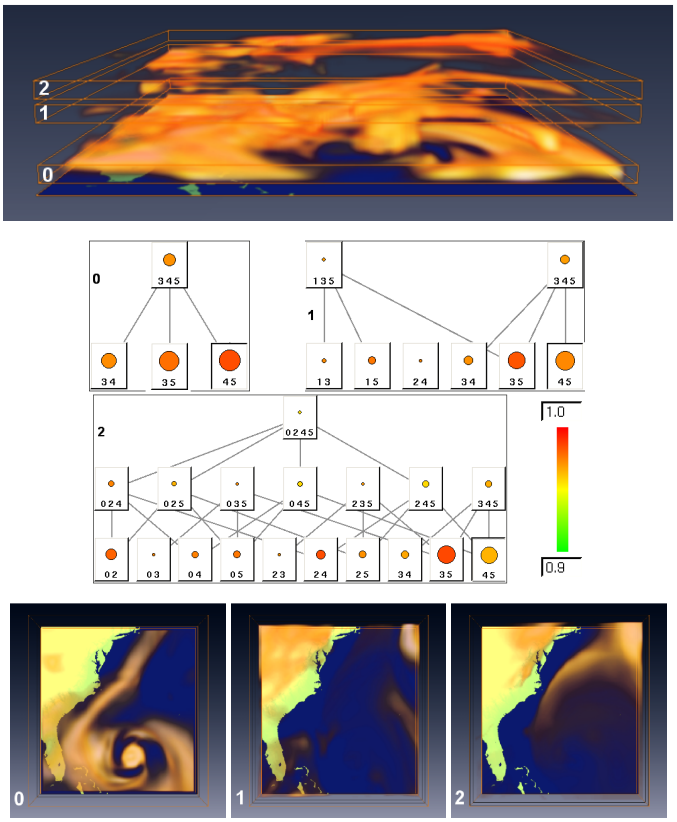

Fig. 14. In the visualization of the correlation field $\mathbf{C}_{\{4,5\}}$ (top) several layers of high correlation are visible. For three of these regions $(\mathbf{0},\mathbf{1},\mathbf{2})$, a local Multifield-Graph is computed, using LCC and $\theta = 0.95$ (second and third row). The color coding of the Multifield-Graph is shown in third row right. All graphs are focused on nodes with $rc_N = 0.11$. In the bottom row, the regions $(\mathbf{0},\mathbf{1},\mathbf{2})$ of $\mathbf{C}_{\{4,5\}}$ are visualized separately.

with $rc_N > 0.11$ (Figure 14 second and third row). It is visible that in layer 0 and 1 only some fields correlate and in layer 2 correlations occur between several field tuples. The correlation fields of these layers can be visualized separately, see in Figure 14 (bottom). Here we can see that in the lower layers there is a stronger correlation between vapour and temperature around the center of the hurricane, whereas only low correlation is present in the upper layers.

The computation time for the LCC measure with a window of 7x7x7 is rather high. The preprocessing step, the computation of the pair correlation fields, took around 20 minutes. The Multifield-Graph computation out of the pair correlation fields takes only about 3 seconds and is thus interactive. The computation was performed at a current PC with an AMD Athlon 3500 processor.

We also performed an analysis with the GSIM measure. Here the preprocessing step is much faster. It takes about one minute to do the preprocessing for the six fields with a resolution of 250x250x50. Two of the resulting correlation fields $\mathbf{C}_{\{2,4\}}$ and $\mathbf{C}_{\{4,5\}}$ are shown in Figure 15.

6 Conclusion

As described in this paper we made the following contributions:

- We presented a new approach to visualizing the correlations between different scalar fields in a multifield dataset. The approach is based on the calculation and visualization of correlation fields, which encode the correlation of certain subsets of the original scalar fields.
- We introduced the Multifield-Graph to deal with the huge number and complexity of correlation fields. It can be used for getting an overview of the amount of correlations in all possible correlation fields as well as to interactively focus at different interpretation aims like showing only information of highly correlating fields.

We used our approach on real and synthetic data sets. Our approach is flexible in the sense that it can be used with arbitrary correlation measures and interpretation goals.

In the future, we intend to test our approach with other correlation measures like mutual information, which are able to indicate more general dependencies. We are looking for optimization strategies to enable the analysis of a larger number of fields. We also aim to develop methods for comparing vector fields with scalar fields and for the comparison of vector fields with each other.

Acknowledgements

We would like to thank Tino Weinkauf from ZIB Berlin for his constant support. Also thanks to Kaleigh Smith and Hitoshi Yamauchi for their help. Hurricane Isabel data produced by the Weather Research and Forecast (WRF) model, courtesy of NCAR and the U.S. National Science Foundation (NSF). All visualizations in this paper have been created using AMIRA – a system for advanced visual data analysis [22] (see `http://amira.zib.de/`).

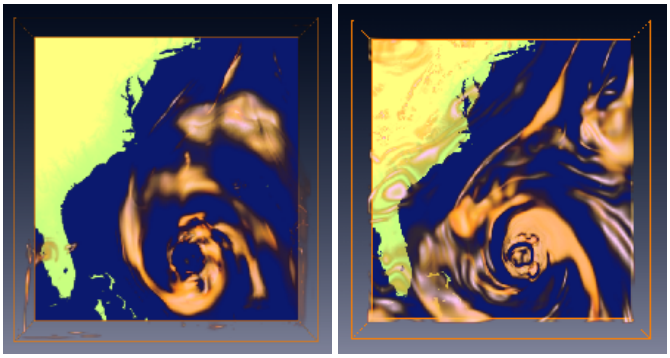

Fig. 15. Correlation fields of the hurricane multifield computed with GSIM: $\mathbf{C}_{\{2,4\}}$ (left), $\mathbf{C}_{\{4,5\}}$ (right).

REFERENCES

[1] O. A. A. Helgeland. Visualization of vector fields using seed lic and volume rendering. *IEEE Transactions in Visualization and Computer Graphics*, 10(6):673–682, 2004.

[2] P. Cachier and X. Pennec. Non-rigid registration by gradient descent on a gaussian-windowed similarity measure using convolutions. In *Proceedings of the IEEE Workshop on Mathematical Methods in Biomedical Image Analysis (MMBIA '00)*, pages 182–189, 2000.

[3] W. Cai and G. Sakas. Data intermixing and multi-volume rendering. In *Computer Graphics Forum (Eurographics '99)*, volume 18(3), pages 359–368, 1999.

[4] A. Crouzil, L. Massip-Pailhes, and S. Castan. A new correlation criterion based on gradient fields similarity. In *1996 International Conference on Pattern Recognition (ICPR '96)*, pages 632–636, 1996.

[5] H. Doleisch, M. Gasser, and H. Hauser. Interactive feature specification for focus+context visualization of complex simulation data. In *Proceedings on the symposium on Data visualisation 2003 (VISSYM '03)*, pages 239–248, 2003.

[6] H. Doleisch, M. Mayer, M. Gasser, P. Priesching, and H. Hauser. Interactive feature specification for simulation data on time-varying grids. In *Simulation and Visualization (SimVis '05)*, pages 291–304, 2005.

[7] H. Edelsbrunner, J. Harer, V. Natarajan, and V. Pascucci. Local and global comparison of continuous functions. In *Proceedings conference on Visualization '04 (VIS '04)*, pages 275–280, 2004.

[8] S. Grimm, S. Bruckner, A. Kanitsar, and E. Gröller. Flexible direct multi-volume rendering in interactive scenes. In *Vision, Modeling, and Visualization (VMV '04)*, pages 386–379, 2004.

[9] G. Haller. Distinguished material surfaces and coherent structures in three-dimensional fluid flows. *Physica D Nonlinear Phenomena*, 149:248–277, 2001.

[10] G. Haller. An objective definition of a vortex. *Journal of Fluid Mechanics*, 525:1–26, 2005.

[11] G. Hermosillo, C. Chefd'hotel, and O. Faugeras. Variational methods for multimodal image matching. *International Journal of Computer Vision*, 50(3):329–343, 2002.

[12] D. Jen, P. Parente, J. Robbins, C. Weigle, R. M. T. II, A. Burette, and R. Weinberg. Imagesurfer: A tool for visualizing correlations between two volume scalar fields. In *Proceedings on the conference on Visualization '04 (VIS '04)*, pages 529–536, 2004.

[13] J. Jeong and F. Hussain. On the identification of a vortex. *Journal of Fluid Mechanics*, 285:69–94, 1995.

[14] C. Johnson. Top scientific visualization research problems. *IEEE Computer Graphics and Applications*, 24(4):13–17, 2004.

[15] R. M. Kirby, H. Marmanis, and D. H. Laidlaw. Visualizing multivalued data from 2d incompressible flows using concepts from painting. In *Proceedings of the 10th IEEE Visualization 1999 Conference (VIS '99)*, pages 333–340, 1999.

[16] J. Kniss, C. Hansen, M. Grenier, and T. Robinson. Volume rendering multivariate data to visualize meteorological simulations: a case study. In *Proceedings of the symposium on Data Visualisation 2002 (VISSYM '02)*, pages 189–194, 2002.

[17] J. Kniss, G. Kindlmann, and C. Hansen. Multidimensional transfer functions for interactive volume rendering. *IEEE Transactions in Visualization and Computer Graphics*, 8(3):270–285, 2002.

[18] M. Levoy. Display of surfaces from volume data. *IEEE Computer Graphics and Application*, 8(3):29–37, 1988.

[19] W. E. Lorensen and H. E. Cline. Marching cubes: A high resolution 3d surface construction algorithm. In *Proceedings of the 14th annual conference on Computer graphics and interactive techniques (SIGGRAPH '87)*, pages 163–169, 1987.

[20] A. Mahalanobis, B. V. Kumar, and A. J. V. Nevel. Volume correlation filters for recognizing patterns in 3d data. In *Proceedings of SPIE*, volume 4471, pages 51–58, 2001.

[21] F. Rössler, E. Tejada, T. Fangmeier, T. Ertl, and M. Knauff. GPU-based Multi-Volume Rendering for the Visualization of Functional Brain Images. In *Procceedings of SimVis 2006*, pages 305–318, 2006.

[22] D. Stalling, M. Westerhoff, and H.-C. Hege. Amira: A highly interactive system for visual data analysis. *The Visualization Handbook*, pages 749–767, 2005.

[23] R. Taylor. Visualizing multiple fields on the same surface. *IEEE Computer Graphics and Applications*, 22(3):6–10, 2002.

[24] A. Telea and J. J. van Wijk. Simplified representation of vector fields. In *Proceedings of the conference on Visualization '99 (VIS '99)*, pages 35–42, 1999.

[25] S. T. Teoh and K.-L. Ma. Hifocon: Object and dimensional coherence and correlation in multidimensional visualization. In *International Symposium on Visual Computing (ISVC '05)*, pages 235–242, 2005.

[26] T. Urness, V. Interrante, I. Marusic, E. Longmire, and B. Ganapathisubramani. Effectively visualizing multi-valued flow data using color and texture. In *Proceedings on the 14th IEEE Visualization 2003 (VIS '03)*, pages 115–121, 2003.

[27] P. C. Wong and R. D. Bergeron. 30 years of multidimensional multivariate visualization. In *Scientific Visualization, Overviews, Methodologies, and Techniques*, pages 3–33, 1997.

[28] P. C. Wong, H. Foote, D. L. Kao, R. Leung, and J. Thomas. Multivariate visualization with data fusion. *Information Visualization*, 1(3/4):182–193, 2002.