Chapter 11
Categorization

Helwig Hauser and Hamish Carr

Abstract Multifield visualization covers a range of data types that can be visualized with many different techniques. We summarize both the data types and the categories of techniques, and lay out the reasoning for dividing this Part into chapters by technique rather than by data type.

As we have seen in the previous chapter, multifield visualization covers a broad range of types of data. It is therefore possible to discuss multifield visualization according to these data types, with each type covered in a separate chapter. However, it is also possible to approach the question by considering the techniques to be applied, many of which can be applied to multiple types of multifield data. In this chapter, we therefore discuss both ways of analysing multifield visualization techniques, and why we have chosen to proceed according to technique rather than type in the subsequent chapters.

11.1 Categorization by Data Type

All multifield data shares a common attribute—that it is known or presumed that the fields are related spatially to each other. However, these relationships can arise in different ways, and this has an impact on how we analyze or visualize the data.

Broadly speaking, the individual fields in multifield data can be related in a number of ways:

1. Multi-variate data, where related properties are computed or measured,
2. Spectral data, where multiple properties are measured, but may or may not be related,

H. Hauser (✉)
University of Bergen, Bergen, Norway
e-mail: Helwig.Hauser@UiB.no

H. Carr
University of Leeds, Leeds, UK
e-mail: H.Carr@Leeds.ac.uk

C.D. Hansen et al. (eds.), *Scientific Visualization*, Mathematics and Visualization,
DOI 10.1007/978-1-4471-6497-5_11

3. Multi-run/ensemble data, where each field is a separate output of a computational or measurement process,
4. Derived fields, where new fields are generated to sharpen the understanding of existing fields,
5. Multi-scale data, where fields at different resolutions or scales are considered,
6. Other, ranging from tensor fields to time-dependent data.

We will canvass each of these types of data separately, proceeding from the types whose fields are more obviously tightly related, to those where the multifield representation is more a choice of representation than an inherent structure.

11.1.1 Multi-variate Data

Multi-variate data are common to several applications, including computational fluid dynamics (CFD), where the interaction between multiple physical quantities are modeled and computed over detailed spatial grids. In the simplest case, each location or sample in a spatial domain is assigned a coherent vector of multiple variables. The paradigm case of this type of data is CFD, where properties such as pressure and velocity are computed for each location in a grid.

Multi-variate datasets (in terms of this definition) are usually characterized by a relatively small number of variables (between two and a few dozen). Here, the visualization challenges arise from the fact that the correlation between pairs of variables is wildly heterogeneous. For example, while some variables are perfectly linearly correlated, others may be largely unrelated as, e.g., when resulting from separate solvers (say a fluid solver and a chemical reaction solver).

11.1.2 Spectral Data

Another type of multifield data is spectral data. Most commonly resulting from physical acquisition techniques (such as spectral imaging techniques), we consider datasets where data relating to different frequencies are represented as different fields. An example is spectral satellite imaging, where (concurrently) a number of images at different wavelengths are taken from the same target, resulting a multi-frequency dataset.

In comparison with multi-variate data, spectral datasets may involve much larger numbers of fields (frequencies), which leads to interesting visualization challenges. However, it is commonly the case that there is a substantial amount of coherence between all the fields. For example, the fields are often sorted in a meaningful way (usually by frequency), and responses to different frequencies tends to correlate more tightly than for example pressure and vorticity in a CFD computation.

11.1.3 Multi-run/Ensemble Data

A third type of multifield data is multi-run, multi-parameter, or ensemble data. These datasets represent multiple results from the same operation, rather than multiple related operations. Multi-run data, for example, can result from repeating a stochastical simulation a certain number of times, leading to data which can be interpreted as a statistical sample of outputs from the model. Multi-parameter or ensemble data can also result from repeating the data acquisition (simulation or measurement) while varying input parameters of either the simulated model or the measurement technique (for example regular or Monte-Carlo sampling).

Visualization of these multi-run, multi-parameter, or ensemble data usually amounts to performing a sensitivity/variability analysis of the phenomenon under consideration. In climate research, for example, the dependency of a forecast on certain model parameters can be studied. In engineering, on the other hand, the performance of a certain system component can be studied, while external driving conditions are varied.

11.1.4 Derived Fields Data

As one moves from intrinsically multifield data to data which is multifield as a result of the choice of representation, the next type to be considered is that of derived fields. In these datasets, one or more additional fields are computed directly from the known fields (as distinct from being computed at the same time as the original fields).

For example, to understand moving particles, additional descriptive quantities are often computed for each field location that—all together—explain aspects of the behaviour of the system, whether local or global.

Intrinsic to this derivation is an expectation that the derived field will depend strongly on the originating fields—thus, the derived field can either be viewed as additional information or as a reduced or simplified form of information. Even for a single scalar field, however, the opportunity of deriving fields implies that multifield visualization methods may be applicable.

11.1.5 Multi-scale Data

A further type of multifield data arises when a single field is measured at different scales or different resolutions. The effective selection of a scale can often depend on understanding the relationship between these resolutions, giving rise therefore to multifield problems. In essence, the scale axis is used to set the fields alongside each other, leading to a scale-space representation where each field represents the data at a certain scale.

Visualization questions for these data types involve the selection of an appropriate scale, or the consideration of the data through a proprietary (for example selective) reconstruction of the data (based on certain scales of interest).

11.1.6 Other Types of Multifield Data

In addition to the major types of data already listed, a multifield framework can be used as a representation for data such as tensor fields, where tensor components are interpreted as individual fields, or time-dependent data, where the time-steps are interpreted as individual fields.

Representing such data in a multifield form allows the use of existing visualization methods such as coordinated multiple views with linking and brushing, or focus+context visualization. Of course, an additional challenge is generated by the fact that an important semantic aspect of the data (that the fields actually make up a tensor or a time series) is possibly lost (or cannot be exploited).

11.1.7 Summary

If we look at the various types of multifield data, we see that nearly all of the types require similar tasks to be performed, and in particular require the detection or visualization of correlations between the fields. As a result, many of the techniques applicable to one type will tend to be applicable to other types, and a categorization by data type risks the repetitive discussion of the same techniques. We therefore consider in the next section the techniques that are applicable to multifield visualization, then return to the question of which approach to adopt.

11.2 Categorization by Visualization Approach

As we have seen above, one way to categorize multifield visualization is to focus on the type of data. A second way to categorize is to observe that many techniques cut across all of the types of data as discussed above. The advantage of this characterization is that it gives a principled context in which to discuss not only those techniques that have already been reported, but also in which to discuss classes of techniques that could be introduced in the future. A second advantage of this approach is that we can extrapolate more readily from techniques known to work for single-field data, whether scalar, vector or tensor.

Broadly speaking, we can observe that the visualization of single-field data relies on mapping the data to properties of the human visual system, on providing the user interactive tools for isolating regions of the data, and on the detection of significant features in the data. In addition to these basic categories, existing multifield visualizations often rely on the mathematical habit of reducing complex problems to simpler problems with known solutions. In the context of multifield visualization, this usually means computing a single scalar or vector field based on the input data, then visualizing that single field.

This therefore leaves four broad categories of approaches to multifield visualization, in approximate order of difficulty:

1. Visual Channel Mapping
2. Derived Fields
3. Interactive Exploration
4. Feature Detection and Analysis

Each of these will be covered in a separate section, but we start with a high-level overview of these methods first.

11.2.1 Visual Channel Mapping

For single fields or for multifields with small numbers of variables, the first set of approaches, including much of the work published to date, involves mapping data properties to visual properties. So, for example, one dependent variable may be mapped to the red channel, a second to the green channel, and a third to the blue channel. Alternately, one channel could map to hue, a second to saturation, and a third to brightness.[1] Visual channels that can be exploited this way are not, however, restricted to colour alone—as we will see in Chap. 12, texture and geometric shape are also used to represent data properties.

A core problem with visual channel mapping is that the human visual system has a limit on how many different visual channels can be perceived at once. Moreover, the amount of precision in the visual system limits the qualitative conclusions that can be drawn. However, due to the simplicity and straightforwardness of visual channel mapping, it often forms the basis for the methods to be developed in subsequent chapters.

11.2.2 Derived Fields

Once the visual channel limitations are realized, the next set of methods relies on reducing the number of visual channels by combining elements of multiple data variables in a single channel. This is usually done by computing some summative property that encapsulates a relationship between the variables, thus reducing the

[1] We note that both of these mappings are poor choices visually, but they are the easiest illustrations of the principle.

multifield to a single field, which is then visualized directly using existing techniques. These methods may include measures of complexity or correlation between variables, or derived properties such as vorticity.

While these methods can be very effective, they tend to work best at detecting relationships that are known or suspected. This is because the choice of the summative property is usually guided by a sense that a particular aspect of the data set is significant. Moreover, they presume that the phenomenon being studied is uniform throughout the domain: thus, if two properties are weakly correlated in one area, but strongly correlated elsewhere, these methods will be less successful.

11.2.3 Interactive Exploration

A third category of visualization techniques relies on the experience and intuition of the user, by providing an interactive tool for exploring the data. Inevitably, this relies on visual channel mapping and derived fields to give the user sufficient insight to identify features, and increasingly, on feature detection as well.

Interactive exploration can operate by manipulation of the visual channel mapping, by the provision of geometric tools to identify regions of the data, by selection of paradigm points or regions as seeds for similarity measures, by combination of properties through logical rules, or by reference to abstract descriptions or secondary visualizations.

While often the most effective approach, interactive exploration starts breaking down with larger data sets, as does direct visualization itself, as the volume of data outstrips the humans visual and cognitive capacity to understand the data.

11.2.4 Feature Detection and Analysis

All three categories described so far share a common difficulty: that, as the amount of data increases, less and less of it can be presented to the user. In short, the question is not “how can we visualize the data”, but “what subset of the data can we visualize”. As a result, visualization techniques have increasingly relied on abstract definitions of features, either specific to a domain, specific to a type of data, or common to multiple domains and data types. These features are detected computationally and presented to the user either as the answer to a question, or as the seeds to an interactive exploration.

Philosophically, these methods shade off into the disciplines of image analysis, computer vision and data analysis, all of which share a common interest in detecting features in masses of data. However, one set of methods which is distinctive in visualization is the reliance on formal mathematics such as topology to extract abstract features either for further analysis or for direct visualization.

11.2.5 Summary

Clearly, a breakdown according to techniques runs a similar risk of repetition to the risk observed for a breakdown according to types. However, as we can see from the discussion above, a breakdown according to techniques has more obvious and clearer demarcations, in addition to providing a roadmap for as yet unidentified techniques.

11.3 Conclusion

In short, while we can categorize multifield visualization either by the type of data or by the type of technique, we have chosen the latter for two reasons. First, similar techniques are visible across all types of data, and it is therefore easier to consider a single technique and, if necessary, its application to different types. Second, it allows us to extrapolate future techniques out from the accumulated experience of working with single-field data.