Icon-based Visualization using Mosaic Metaphors

Thomas Nocke[1], Stefan Schlechtweg[2], Heidrun Schumann[1]
[1]University of Rostock, {nocke, schumann}@informatik.uni-rostock.de
[2]University of Magdeburg, stefans@isg.cs.uni-magdeburg.de
Germany

Abstract

This paper introduces a new approach to extend icon-based visualization methods by using a mosaic-based paradigm. We discuss, how image metaphors closely related to the application domain can be applied for icon-based representations. Therefore, we enhance visualizations by well-known Image Mosaic techniques, such as image layouts, image selection and color adaption. Furthermore, we present the results of our approach by discussing an example of a clustered real-world climate data set.

Keywords Icon-based Visualization, Mosaic-based Rendering, Climate Impact Research, Visualization Metaphors

1. Introduction

The visualization of large multi-variate data sets is still a challenging task, especially when we consider the exploration of a variety of attributes in one representation. Many techniques have been developed to visualize such multi-variate data sets (cf. [5] for an overview):

- pixel-oriented techniques represent each data value as a single pixel on the screen,
- geometric projection techniques map the n-dimensional information space to "interesting" subspaces,
- icon-based techniques display small icons representing the attribute values of the data records, and
- hierarchical and graph-based techniques display the information space as a hierarchy or a graph.

In this paper we will focus on presenting multi-variate data on maps. In this context icons are a frequently used approach. One advantage of such icons is that they combine all data values of a given data set in one primitive and, thus, enable users to recognize multi-variate correlations. However, the perceptibility of such icons quickly decreases with the number of variables being displayed. To increase the intuitive recognition of such icons, easy perceptible metaphors such as Chernoff faces [2] and Infobugs [3] can be applied.

Techniques visualizing information using intuitive metaphors have been developed for different fields. The InfoCanvas system [11] enables users to interactively select and place metaphors on an image. In [10] the simple and expressive image metaphor of "money trees" for stock data has been introduced, to give just two examples.

Especially in the domain of cartography metaphors or symbols are widely used to support the effective recognition of general information. The question we want to answer is whether metaphors can be used to communicate numerical, multi-variate data as well. In other words: Can we replace abstract icons by metaphor-based icons?

When placing such metaphor-based icons on maps, several problems similar to the generation of Image Mosaics have to be solved. In this paper we discuss how algorithms that create such Image Mosaics can be applied to enhance icon-based visualizations on maps.

The paper is organized as follows. In Section 2 we describe the background. Afterwards, we discuss our approach to mosaic-based visualization (Sec. 3). Furthermore, we describe and discuss the results of our approach on a data set from the climate impact research background (Sec. 4). Finally, we conclude with a short discussion and outline future work (Sec. 5).

2. Background

In this section, we first outline basics of icon-based visualization techniques, afterwards discuss the background of Image Mosaics and, finally, describe the background of the data set to which we have applied our approach.

2.1. Icon-based visualization techniques

Icon-based visualization techniques have been proven to be effective for the visualization of multi-variate data. They map data variables onto geometric (e.g., shape, size, orientation) and non-geometric (e.g., color and texture) visual attributes (cf. [14]). There are two general *modi operandi*:

1. techniques generating individually recognizable icons, and

2. icon-based techniques generating textures.

Well-known examples for the first class are Star-Glyphs [13] and Autoglyphs [1]. Both enable users to identify and compare certain data values. With a rising number of concurrently displayed icons, texture-based techniques from the second class are more expressive (e.g., Stick-Figure-Icons [9] and Color-Icons [6]). They support the identification of general trends and clusters in the data.

The design of suitable icons is a sensitive process. Special editors have been developed to interactively construct such icons. However, metaphor-based icons to visualize multi-variate data are – with the exception of Chernoff faces [2] – rather uncommon.

2.2. Image Mosaics

Image Mosaics have become very popular in the last few years as a form of art, which is used for advertisements, posters, etc. An Image Mosaic is created from other smaller images which together portray a larger subject. An input image is to be replaced by an output image which is constructed from several "tiles" of a size larger than a pixel and which match the tonal values of the input image (cf. [4, 7]). The creation of an Image Mosaic requires the following four major steps:

1. choose images which are to be used as mosaic tiles,
2. choose a tiling grid,
3. find an arrangement for the mosaic tiles in the grid,
4. possibly perform a color correction on the tiles to match the target image

The most interesting feature of Image Mosaics that will be used in the context of this paper is that they include two different levels of detail in one image. First, you get a global view on the portrayed subject by looking at the mosaic as a whole. Second, the individual tiles show small details depending on the selection of the tile. Typically, the tiles are selected in such a way that they portray something that is connected to the subject which is shown in the mosaic.

If we compare this to icon-based visualization techniques, the icons representing the data can be regarded as the image tiles and the display of the whole data set as the mosaic itself. The challenges are now similar to the four steps described above:

1. find iconic representations of the individual data items,
2. find a layout for the icons depending on the underlaying data set or chose a layout,
3. place the icons according to the chosen layout,
4. possibly perform a color correction to fulfill the goals regarding a global overview of the data.

We want to discuss this strategy considering as example a concrete application scenario.

2.3. Application: a maize harvest data set

A challenge for climate researchers is to develop methods that support the evaluation of climate model results and to reproduce extreme conditions. Important instruments in this context are cluster analysis and visualization techniques (see [8]). Here especially, the visual exploration of the multi-variate features of spatially-scattered clusters can benefit from Image Mosaic algorithms.

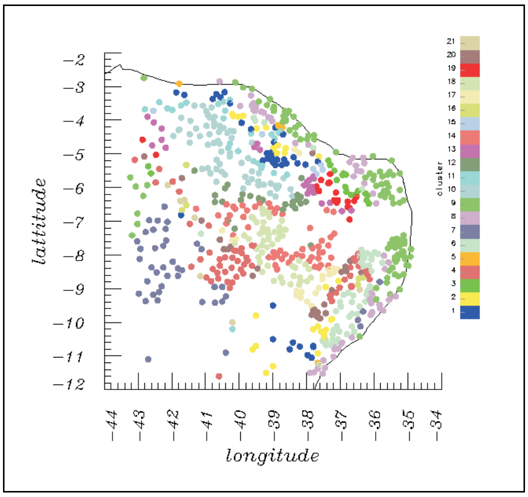

Figure 1. Visualization of clusters representing the risk of a drought for maize cultivation during the year 1983 in the semi-arid Northeast of Brazil based on regional climate model results.

The data set that we will be using as example throughout the paper represents a severe drought that occurred during the year 1983 in the semi-arid northeast of Brazil. To generate the data set, idealized criteria, identified by climate researchers, were assumed based on certain total precipitation thresholds. Doing so, it is possible to characterize the risk of potential total yield loss for maize as one of the major agricultural crops of this region. These criteria have been used to derive six parameters expressing the positive differences between the individual precipitation thresholds and the actual rainfall. For these, a cluster analysis on the 865 measurement stations has been performed (see Fig. 1).

In this figure, similar observation stations – referring to the mentioned parameters – are aggregated into the same cluster, whereas dissimilar stations belong to different clusters. Unfortunately, such an image of circles representing the cluster ID of each observation station does not allow to get an overall overview about cluster features as well as to explore all the parameters constituting the cluster in one image. To achieve a better expressiveness and effectiveness of the image, the following tasks have to be supported:

1. the cognition of general cluster information, e.g., the identification of extreme cultivation conditions,
2. the exploration of individual parameters and of parameter dependencies, and
3. the identification and localization of measurement stations.

This can be achieved using metaphor-based icons.

3. An approach for mosaic image visualization

In the following we want to discuss, how the tasks from Section 2.3 can be supported using the Image Mosaic approach outlined in Section 2.2. The first step is to find an adequate iconic metaphor for the individual measurement stations.

3.1. Designing the mosaic pattern

The data set consists of six parameters and a cluster ID at each 2D position, altogether seven values. Since the cluster ID is of more general character and of higher importance for a general overview, we need to encode it separately. Considering the meaning of the parameters in an agricultural context, a metaphor based on a simplified maize cob is very straightforward and intuitive (see Fig. 2). Low risks for the maize harvest are represented by a thick, yellow cob (Fig. 2 left), whereas high risks are represented by a thin, brown cob. Other image metaphors, for instance, displaying clouds, plants and the sun are imaginable. We have decided against it, since the maize cob is well-perceptible even in an image with a high number of measurement stations.

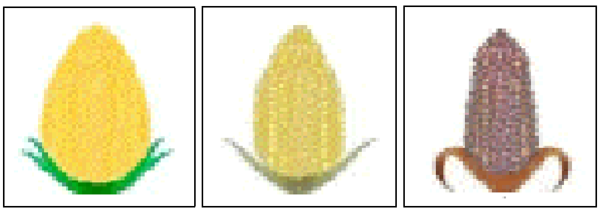

Figure 2. The three base icons displaying maize conditions: good (left), middle (center) and bad (right) conditions

The metaphoric images from Fig. 2 can be arranged and chosen analogous to mosaic images, yellow cobs in "good" regions and brown cobs for "bad" regions. However, keeping in mind that we want to analyze six parameters, we need to extend this simple metaphor. A straightforward approach is to draw six of these maize cobs (one for each parameter). Unfortunately, due to the requirement that the small images should remain identifiable to retain their metaphoric meaning, the resulting rectangular icon gets quite large, producing overlapping problems in regions with high point density. Thus, we subdivide the maize cob image into six regions, and construct the resulting image for a parameter combination based on six image parts of Fig. 2. For instance, high values of parameter p_1 lead to a thick, yellow head on the left cob side. Fig. 3 illustrates this procedure. Doing so, we avoid to enlarge the image and the user gets all information at a glance.

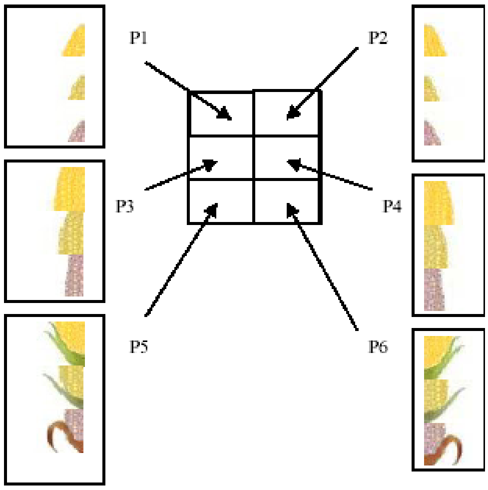

Figure 3. Construction of a metaphor-based icon, representing six parameters

Furthermore, using a certain background color, these icons simply allow to encode the cluster membership as well (see Fig. 4). Doing so, the resulting mosaic image gets a well-perceptible "coat of paint", that encodes this important information.

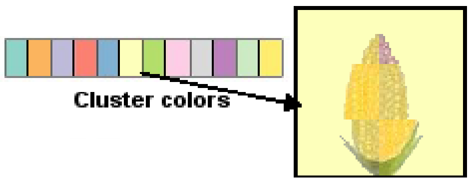

Figure 4. Metaphor-based icon representing six parameters with the background color identifying the cluster

3.2. The layout problem

After the mosaic tiles have been designed we have to find a layout for the tiles and place them according to the chosen layout (steps 2 and 3 from Section 2.2). Three basic grid layout types have been introduced [12]:

- scattered layouts,
- regular layouts and
- multi-resolution layouts.

Fig. 5 illustrates these layout types. Each of them has certain advantages and disadvantages for the visualization of our data set.

The *scattered layout* for Image Mosaics is a random process, in general used as an artistic feature. This procedure is – in general – not suitable for a visualization of numerical data, since the tile images (or icons) in visual representations – in contrast to mosaic images – can not be placed to arbitrary positions. In the visualization context we need to keep the important relation between the visual representation of the measured data and the position of the measurement stations. Based on these considerations, there are two possible scattered layout mechanisms:

- using a straightforward layout placing icons of constant size at the position of the measurement stations, and
- placing icons into given or calculated (e.g., by a Voronoi tesselation) regions, adjusting their position and size according to the region's extent.

We use the first mechanism, placing images at the measurement station positions only (see Fig. 5 left). One advantage of this layout is that it directly represents the underlying distribution of 2D locations and avoids possibly misleading interpretations. A disadvantage, however, is the possible overlapping of icons in regions with high station density.

To overcome this drawback, regular and multi-resolution layouts can be applied. *Regular layouts* (Fig. 5 center) show in each grid cell an aggregated information over all measurement stations (see Sec. 3.3). Three kinds of regular mosaic layouts are commonly used:

- rectangular grid (standard regular grid),
- hexagonal grid, and
- brick wall layout.

For visualization purposes rectangular as well as hexagonal grids are suitable for different icon shapes. However, a brick wall layout is a more artistic instrument and seems to bring no new benefit for the visualization. We use the rectangular grid, which is especially suited for the rectangular icon we have designed. Moreover, we have applied a region merging of neighboring grid cells with the same (mean) cluster to larger blocks (2×2, 3×3, $\dots$), and display these regions with a larger icon. This allows to identify large homogeneous regions quickly.

A regular grid layout enables users to get a quick overview of the data set and to explore the overall distribution of values and clusters. A drawback of such layouts is that the original parameter and cluster values are aggregated and, hence, no longer explicitly visible. As a solution to this problem, the grid resolution can be changed in a detail-on-demand manner. A low grid resolution supports a fast overview of distributions in the data with large mosaic icons, and a high grid resolution allows to get details of certain regions if required with smaller icons. Furthermore, to keep an overview of the point distribution in a certain region, small dots representing the measurement stations can be faded in (see Fig. 6).

As a third layout technique, a *multi-resolution layout* (see Fig. 5 right) can be applied. Here, icons are placed in a region, if there is exactly one cluster in this region. If not, the region is subdivided. Then, parameter values of similar stations only from the same cluster are aggregated and, thus, we avoid to merge regions with inhomogeneous properties. We used a quadtree-based multi-resolution algorithm.

This layout allows to gain a deeper insight into local cluster distributions, keeping all clusters visible. Now the question is how to visualize the metaphor-based icons in grid cells with varying sizes. On the one hand, the icons can be visualized in constant size independently of the grid cell size. On the other hand, the icon size can be scaled to the available space of the grid cell. The decision between these possibilities is background dependent. For instance, if large homogeneous regions should be perceived quickly, the second technique is appropriate (see Fig. 5 right).

A quadtree bases on a successive subdivision of rectangles and, by doing so, approximates the underlying regions. Actually, region boundaries have irregular shape. Thus, the quadtree subdivision can cause the extension of icons into "undefined" regions. To avoid drawing mosaic tiles into these regions (in this case to the Atlantic ocean (Fig. 5 right)), clipping has to be integrated into the quadtree traversal algorithm. Altogether, a quadtree-based multi-resolution layout strongly adulterates the region boundaries (Fig. 5 right), whereas a scattered layout only minimally changes the region shape representation (Fig. 5 left).

3.3. Icon placement

After the layout step has been performed we have to decide which icons to place on the map (step 3 from Section 2.2). This decision is strongly layout-dependent, especially owing to different region merging strategies (see Sec. 3.2).

In the case of a scattered layout, we simply construct the icons based on the individual station parameter values. If,

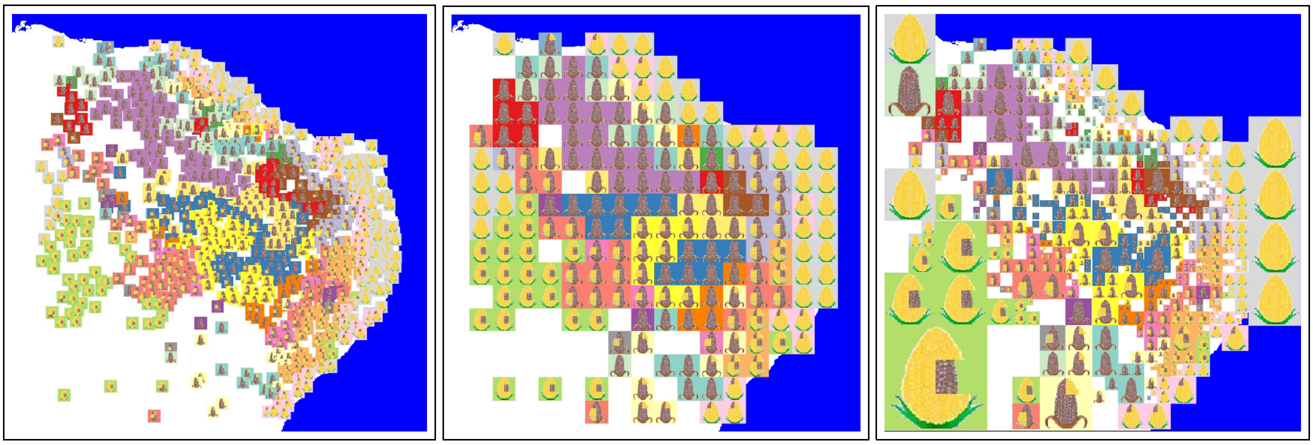

Figure 5. Three layout types of Image Mosaics: scattered(left), regular (center) and quadtree-based multi-resolution (right) layout

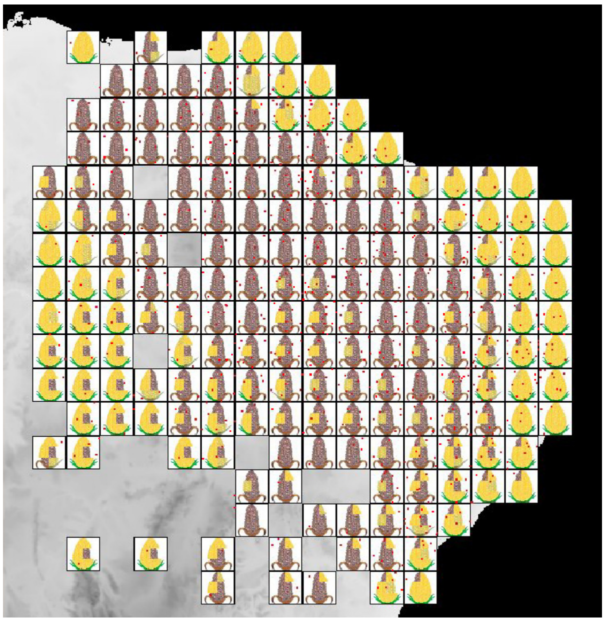

Figure 6. Mosaic image in regular layout with stations faded in (red dots)

on the other hand, the icon should represent a region possibly consisting of more than one station, the user can decide which icon to display:

- a "typical" station in the region,
- an icon based on the mean values for all the stations in the region,
- a cluster representative of the
  - most common cluster (for a regular layout), or
  - only cluster (for a multi-resolution layout)

  in the region.

3.4. Color correction

Finally, we have to parameterize the metaphoric icons, first of all applying a color correction (step 4 from Section 2.2).

First of all, a general consideration: when designing the icons (see Sec. 3.1) we have chosen colors for the parameter categories for each of the six icon parts. This process can be seen as a kind of color correction with only three colors. To generalize this process, any number of categories and referring image parts can be produced using a color interpolation of the "key images" (in the sense of color correction). Since we – in addition to the colors – use different shapes, we have to interpolate the shapes as well. Further research needs to be done to evaluate if – in dependency of the applied metaphors – interpolation of colors and shapes is a suitable procedure.

A further kind of color correction for the metaphoric icon is the adaption of the background color representing the cluster membership. For the maize icon, two modi support different exploration tasks. To explore spatial cluster distributions, the background of the icon can be completely filled out. On the other hand, the cluster membership can be mapped to the surrounding rectangle border colors only, keeping the rest of the background transparent. This allows to correlate the parameter values with other features on the map, e.g. a height field or other geographic information.

4. Implementation and results

The presented techniques have been implemented in a prototypical system with a variety of interaction possibilities. The modular system can be easily adapted to other 2D clustered multi-variate data sets and other image metaphors. Fig. 7 shows the GUI of the system.

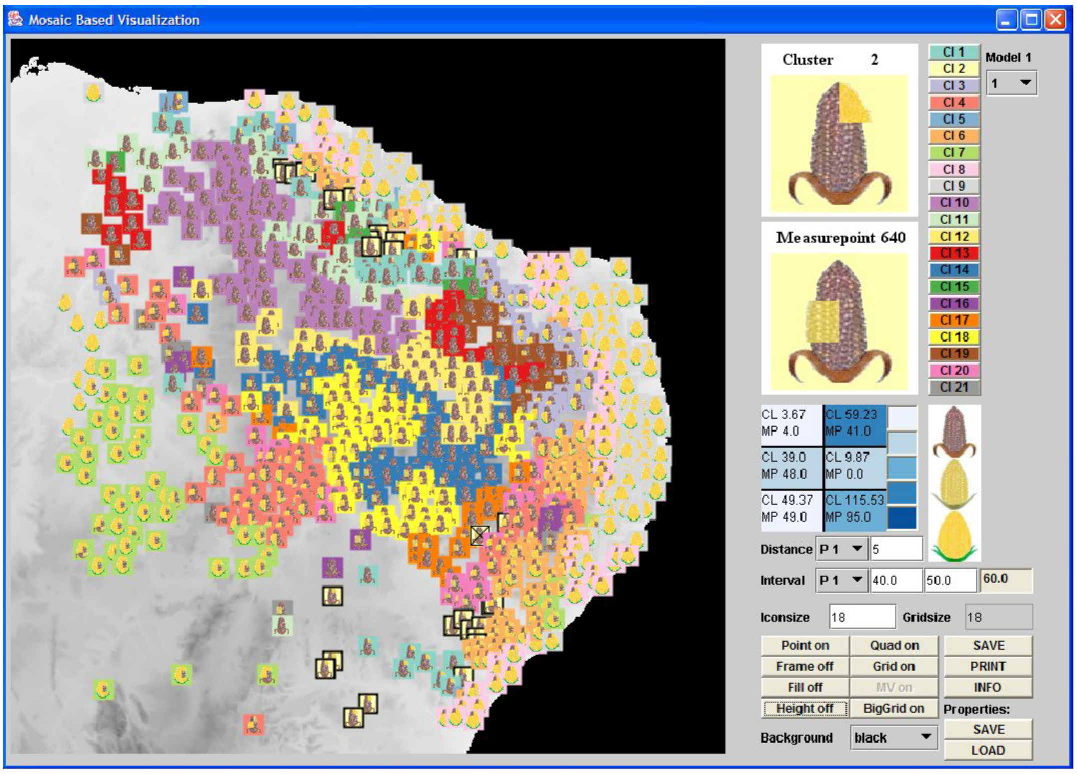

Figure 7. The mosaic framework: Image Mosaic map (left) and parameter control and legends (right); a certain measurement station and all the stations of the same cluster have been selected

The user interface consists of a visualization window (Fig. 7 left), and a GUI area showing legends and parameters (Fig. 7 right). To obtain detailed information about clusters and their properties, a zoomed icon displays the currently selected cluster and its mean values in the top left part of the GUI area. To gain details about measurement stations, the zoomed icon for the currently selected measurement station is drawn below. Alongside to these two icons, the legend for the cluster colors is presented. On the left hand side of this legend the user can change the cluster "model" and, thus, can compare various clusterings calculated for the data set. This allows a statistical evaluation of the deployed proximity measures and cluster algorithms.

Below the icon for the currently selected station, the numerical values of the selected cluster's mean parameters and the selected station parameters are represented. Additionally – to enable users to evaluate the homogeneity of a certain cluster – the differences are displayed using white-blue color coding.

The image icon was designed on the base of thresholds distinguishing between good and bad maize harvest conditions. These thresholds can be interactively adapted and, thus, trigger the mosaic tiles on the map. The base image icons for good and bad conditions (Fig. 2) are depicted on the right hand side of the numerical values of the currently selected measurement stations and the currently selected cluster and their differences, below the cluster color legend.

Moreover, a wide variety of other parameters can be set. This includes grid size, icon size, background color, layout style, icon calculation and others. Furthermore, to reproduce successful parameterizations, a certain parameter combination can be stored and reloaded.

Altogether, we developed a system for metaphor-based Image Mosaics that supports a variety of exploration tasks (see tasks from 2.3). First, general cluster information can be easily perceived. Especially the multi-resolution layout gives a fast overview of local cluster distributions, not merging clusters together (see Fig. 5 right). Extreme conditions and the referring clusters can be easily recognized. Second, individual parameters, their spatial distributions and their dependencies can be easily explored. In this context, the regular layout (see Fig. 5 center and Fig. 6) is suited to get a fast overview of the spatial distribution of the parameter values. In the costal regions and in the south eastern parts, very good conditions for the maize harvest are predicted. On the other hand, in a large area from the north-west to the south-east, the risk for a potential total yield loss is high. Third, to identify and explore individual measurement stations together with their clusters, the scattered layout is especially suited (see Fig. 5 left and Fig. 7). The user can interactively focus on a certain measurement station, study its parameter values and compare it with the cluster mean values and with other measurement stations of this cluster.

In first tests, climate researchers received new insights into the data set that could not be achieved by earlier techniques (see Fig. 1). For instance, the mentioned spatial distribution of regions with high risk of a drought for maize cultivation can be separated very well from areas with a low risk, and the single risk causes (represented by the parameters) can be identified and compared directly. Thus, climate researchers feel that our approach has high potential for the estimation of extreme climate conditions in general. A further achievement is that the approach supports to acquire information at all levels of detail, from an overview of the whole data set (using a regular layout), via regional distributions (e.g., using multi-resolution layouts) to detailed information about station parameters and cluster properties. "Typical" stations, regions and clusters as well as outliers can be examined.

5. Conclusions and future work

In this paper, we have investigated the applicability of mosaic-based methods to enhance icon-based visualizations. Based on the example of a practical agricultural data set, we introduced a new image metaphor and discussed how methods well-known from Image Mosaics, such as image layouts and color adaption, can be used for visualization purposes. Based on these considerations, we described our modular mosaic-based visualization system and presented first results for the application background.

Finally, further research has to be done to explore the application, interpretation and reusability of our approach. First tests with users gave promising insights, however, we have to further test the system with other data sets and suitable metaphors. Especially, this includes to compare the usability of metaphors with non-metaphoric icon techniques. Moreover, we want to extend spatial cluster exploration. This includes to enable users to compare two clusterings in one image and to visualize cluster variances in addition to the currently displayed mean values. Furthermore, we plan to apply focus+context techniques in our framework and to evaluate their use to metaphoric icons on maps.

Acknowledgements

The authors thank Mario Baalcke for implementing the Mosaic Visualization System. Furthermore, we thank our partners Uwe Böhm and Michael Flechsig from the Potsdam Institute of Climate Impact Research for providing the data set and for helpful discussions regarding its mosaic-based visualization.

References

[1] J. Beddow. Shape Coding of Multidimensional Data on a Microcomputer Display. In *Proceedings of the IEEE Visualization conference*, pages 238–246, 1990.

[2] H. Chernoff. The Use of Faces to Represent Points in k-Dimensional Space Graphically. *Journal of American Statistical Association*, 68:361–368, 1973.

[3] M. C. Chuah and S. G. Eick. Information Rich Glyphs for Software Management Data. *IEEE Computer Graphics and Applications*, 18(4):24–29, July/August 1998.

[4] Adam Finkelstein and Marisa Range. Image mosaics. Technical Report TR-574-98, Princeton University, Computer Science Department, March 1998.

[5] D. A. Keim and H.-P. Kriegel. Visualization Techniques for Mining Large Databases: A Comparison. *IEEE Transactions on Knowledge and Data Engineering*, 8(6):923–938, December 1996.

[6] H. Levkowitz. Color Icons: Merging Color and Texture Perception for Integrated Visualization of Multiple Parameters. In *Proceedings of the IEEE Visualization conference*, pages 164–170. IEEE Press, 1991.

[7] T. McKenna and G. R. Arce. New image mosaic structures. Technical report, Department of Electrical and Computer Engineering, University of Delaware, 2000.

[8] T. Nocke, H. Schumann, and U. Boehm. Methods for the Visualization of Clustered Climate Data. *Computational Statistics*, 19(1):75–94, 2004.

[9] R.M. Pickett and G. G. Grinstein. Iconographics Displays for Visualizing Multidimensional Data. In *Proceedings of the IEEE Conference on Systems, Man, and Cybernetics*, pages 514–519, 1988.

[10] X. Shen and P. Eades. Using MoneyTree to Represent Financial Data. In *Proceedings of IEEE Information Visualization, IV'04, London*, pages 285–289, 2004.

[11] J. Stasko, T. Miller, C. Plaue, Z. Pousman, and O. Ullah. Personalized Peripheral Information Awareness through Information Art. In *Proceedings of the International Conference on Ubiquitous Computing*, pages 18–35, September 2004.

[12] T. Strothotte and S. Schlechtweg. *Non-Photorealistic Computer Graphics: Modeling, Rendering, and Animation*. Morgan Kaufmann, 2002.

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

[14] M.O. Ward. A taxonomy of glyph placement strategies for multidimensional data visualization. *Palgrave Information Visualization*, 1:194–210, 2002.