

Chapter One: Introduction

This dissertation addresses an important problem in the visualization of scientific data: Given a family of single-valued functions $F_k(x,y)$ in two dimensions, all sampled on a regular, finite, two-dimensional grid, i,j , devise visualizations that allow each function to be visually discriminated and studied separately and at the same time studied for comparisons and correlations with other functions in the family.

We look at data graphically for many reasons: to get an intuitive feel for values and trends, to understand extent and relative proportion. We display data to inform a curious but non-expert audience, to change the mind of the viewer. An image has the power to communicate what numbers cannot. Why? Because numbers are difficult to grasp and internalize in a way that we fully understand their meaning.

Figure 1.1 shows a set of four single-valued functions from the Intergovernmental Panel on Climate Change [IPCC, 2003]. They are ground-frost frequency (days), precipitation (millimeters/day), wet-day frequency (days), and vapor pressure (hPa). The function values represent the average climate conditions for the month of February from 1981 through 1990. Each variable is originally a continuous function, sampled once at each weather station and resampled onto a regular grid, i,j . The measured values are discrete values with limited precision; this is true of all data stored digitally.

I developed a technique called *Data-Driven Spots (DDS)*, where each function $F_k(x,y)$ is sampled, by a random collection of circular Gaussians with a uniform standard deviation, rather than presented in full. Human visual perception estimates the function's values where it is not represented in the sample. These sampled functions, called *layers*, are presented overlaid. Each function is sampled and displayed in some regions where the others are not, thus each can be discriminated separately; since all are shown simultaneously, correlations can also be estimated.

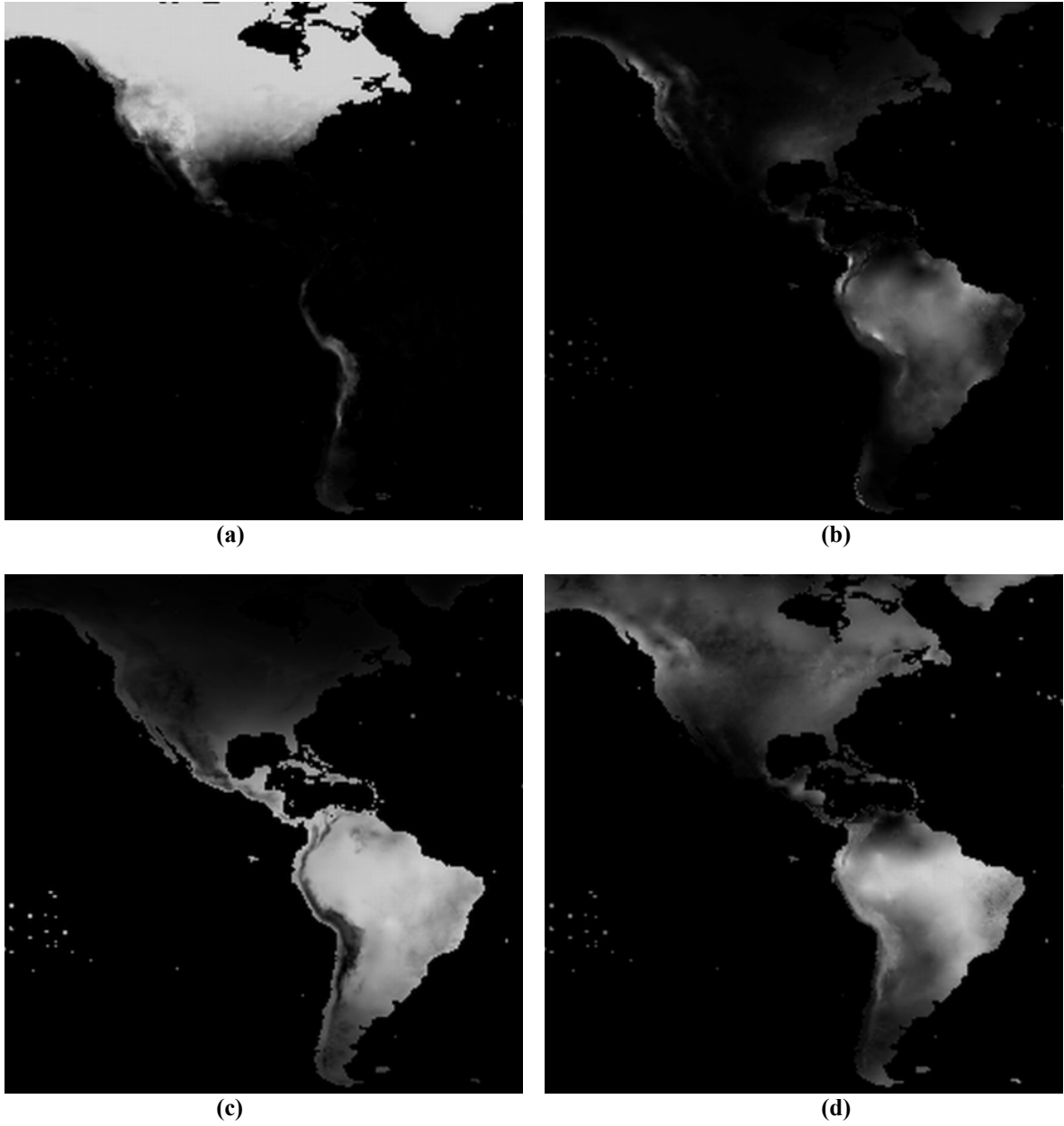


Figure 1.1: Average climate conditions for the month of February, 1981-1990. Ground-frost frequency (days) is shown in (a), precipitation (millimeters/day) in (b), vapor pressure (hPA) in (c), and wet-day frequency (days) in (d). Data publicly available from the Intergovernmental Panel on Climate Change [IPCC, 2003].

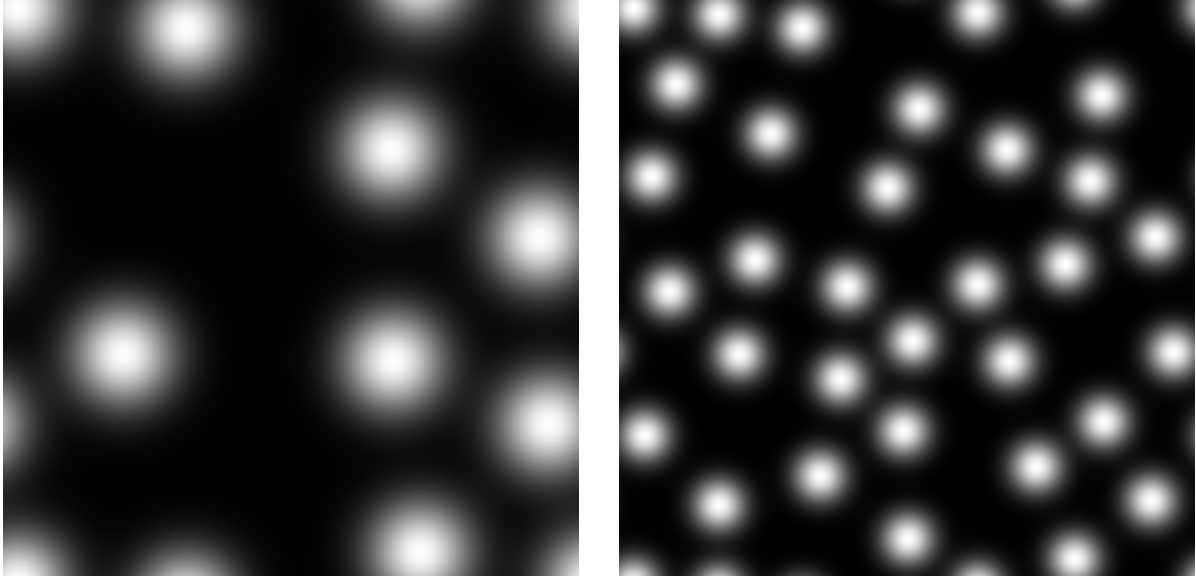


Figure 1.2: Gaussian sampling arrays.

Each function $F_k(i,j)$ is sampled spatially by a two-dimensional array of Gaussians, where each Gaussian in the sampling array has the same standard deviation. There is a unique Gaussian sampling function, $G_k(i,j)$, for each $F_k(i,j)$ in the function family. Examples of Gaussian arrays are shown in Figure 1.2.

Gaussians are randomly placed in the array with the center points located at integral i,j points on the grid and constrained to maintain a minimal distance from one another. Because each Gaussian function is non-zero to infinity, they overlap, but only in the tails of the functions where the values are low. For explanatory convenience, I define a spot to be the circular area around the center of a Gaussian, with a radius of 2.5 standard deviations.

I devised two techniques, each with a static and dynamic variant, for visualizing the layers. In the *alpha-blending* technique, each layer is distinguished by hue and/or the standard deviation of the Gaussians. Data values are multiplied by the Gaussian at each i,j grid point; the result determines the opacity of that layer at that point. Blended with a neutral gray background, each layer has color saturation at i,j proportional to the modulated data value. Since opacities, not saturations, are displayed, lower layers are mostly visible between the spots on upper layers, and changing the order of layers changes the final visualization.

In the *bump-mapped* technique, bumps are visualized by the standard computer graphics technique of bump-mapping, in which off-view-axis illumination and perturbed surface normals are combined to yield an appearance of bumps and/or indents, even though the surface geometry is not modeled. Data values are multiplied by the Gaussian bump perturbation map to produce a data-modulated bump map where low values produce shallower-appearing bumps and high values produce taller-appearing bumps. The layers are distinguished by the standard deviation of the Gaussians. Order of layers doesn't affect the final visualization, since the bump maps are additively combined.

Data-Driven Spots Alpha-Blending

Alpha-blending is a standard computer graphics technique for displaying semi-transparent surfaces [Haeberli and Akeley, 1990]. Alpha represents opacity, and the blending operation combines the colors of two semitransparent surfaces. An alpha value of 1.0 corresponds to complete opacity, and an alpha value of 0.0 corresponds to complete transparency.

The array of alpha values for each layer, $\alpha_k(i,j)$, is the product of the function $F_k(i,j)$ multiplied by the Gaussian sampling array $G_k(i,j)$, for each grid point, i,j . The alpha value is then used to blend the color for the layer, C_k , onto the visualization image. Each layer is blended in succession onto the visualization image; the result is stored in an accumulation image. The blending process for layer k is shown in equation 1.1.

$$Accum_k(i,j) = (1 - \alpha_k(i,j)) * Accum_{k-1}(i,j) + \alpha_k(i,j) * C_k \quad 1.1$$

Because the blending factors are not symmetric, the order layers are added to the image effects the final image color; thus it is important that the user try different orders to produce the most revealing final image.

I refer to the product image of the function $F_k(i,j)$ multiplied by the Gaussian sampling array $G_k(i,j)$, as a variable-modulated Gaussian array. Data nearer the center of a Gaussian is sampled by a value that is near one and is represented more faithfully; data further out is sampled with a low value and contributes little to the final image. An example of a variable-modulated spot array for the ground-frost variable is shown in Figure 1.3.

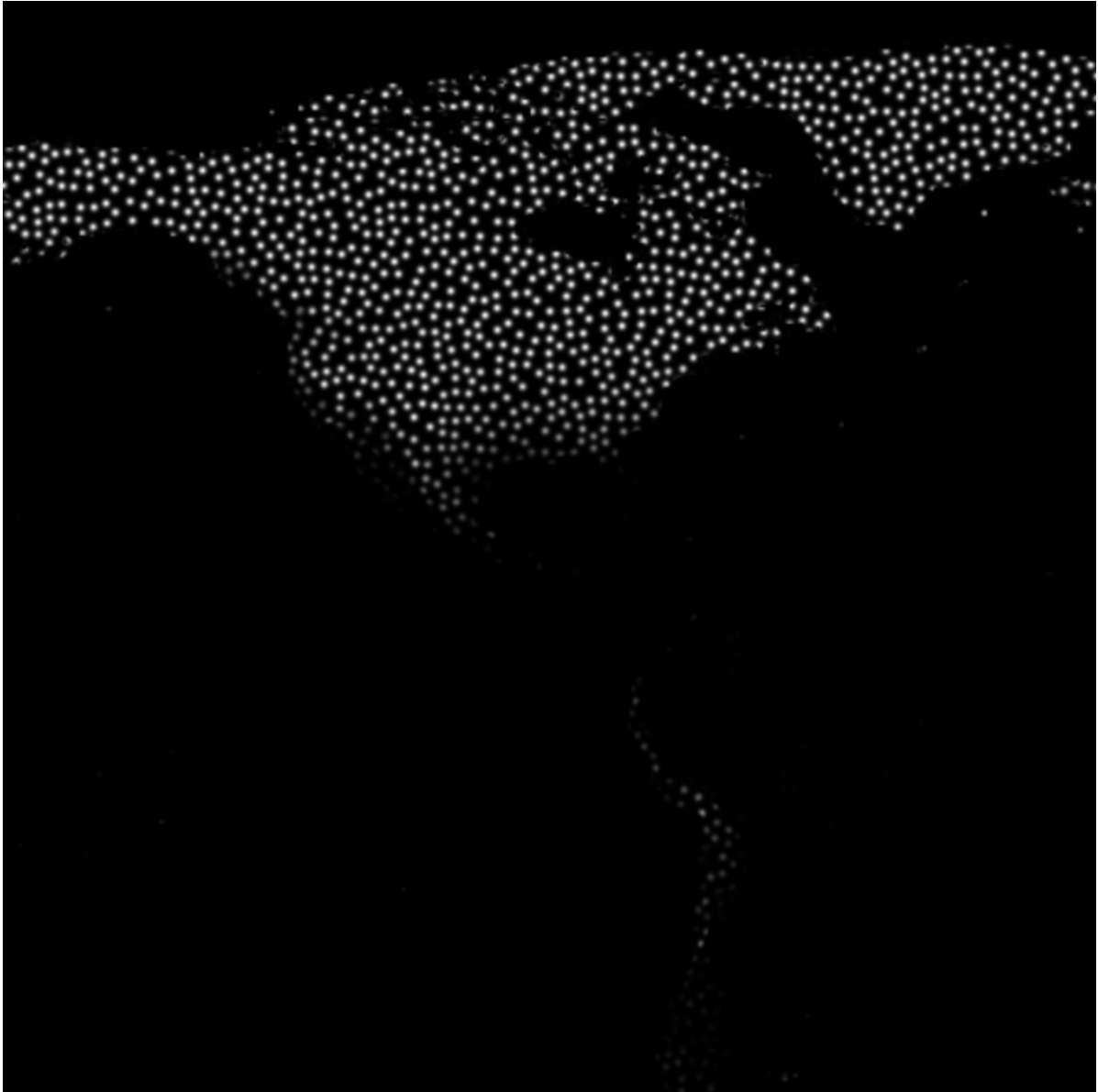


Figure 1.3: Variable-modulated Gaussian array for ground-frost frequency (days) averaged for the months of February during the years 1981-1990.

When the variable-modulated Gaussian array is displayed with DDS alpha-blending, high data values near the center of a Gaussian spot translate to high opacity, and low data values near the center of a Gaussian translate to low opacity. In between spots there is near-zero opacity, or full transparency.

If one layer is displayed over a gray background, the net result is an image of a gray surface covered with colored spots. Within any given spot the data can be seen in the level of transparency or the amount of color over the gray background. A discontinuity at a data boundary can be seen if it falls within a spot.

Figure 1.4 shows the four data functions from Figure 1.1 combined in one image. Each function is displayed with a separate visual layer. Figure 1.5 shows the same data: ground-frost frequency, precipitation, wet-day frequency, and vapor pressure, measured during the month of August during the years 1981-1990.

Visual obstruction of one variable by another may occur wherever the Gaussian spots used to sample the two variables overlap near their center points. In this case, if one (or both) variable has a low value and produces a transparent region in the layer, no visual obstruction occurs. However, if both variables have a high value at the sample location, then only the variable displayed by the uppermost layer is seen. One solution to this problem is to decrease the Gaussian standard deviation as layers are applied, so smaller spots are stacked on top of larger ones. The human mind interpolates the continuation of the larger spot even when partially obscured by a smaller spot on top – this is based on the perceptual phenomenon of *good continuation* [Hoffman, 1998], discussed more in Chapter Three.

It remains an interesting question how many distinguishable layers can be displayed with DDS alpha-blending using different colors and Gaussian standard deviations for each layer. The experimental validation of DDS began to explore this question and showed that up to six distinguishable layers could be displayed with color. The experiment did not seek to find the maximum number of layers that could be displayed with DDS alpha-blending, however.

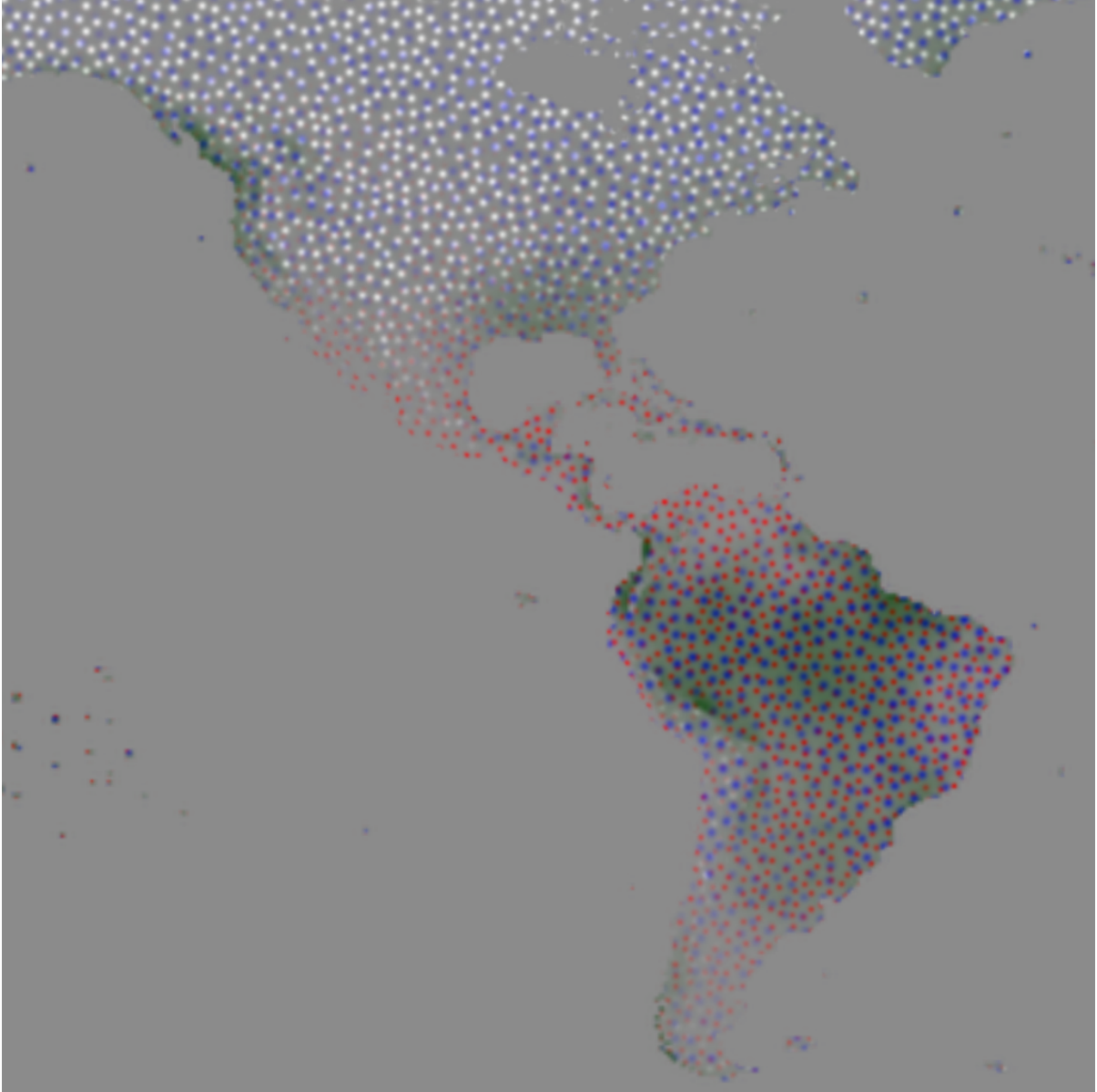


Figure 1.4: Average climate conditions for the month of February during the years 1981-1990. Ground-frost frequency (days) is shown in white, precipitation (millimeters/day) in blue, vapor pressure (hPA) in red, and wet-day frequency (days) in green.

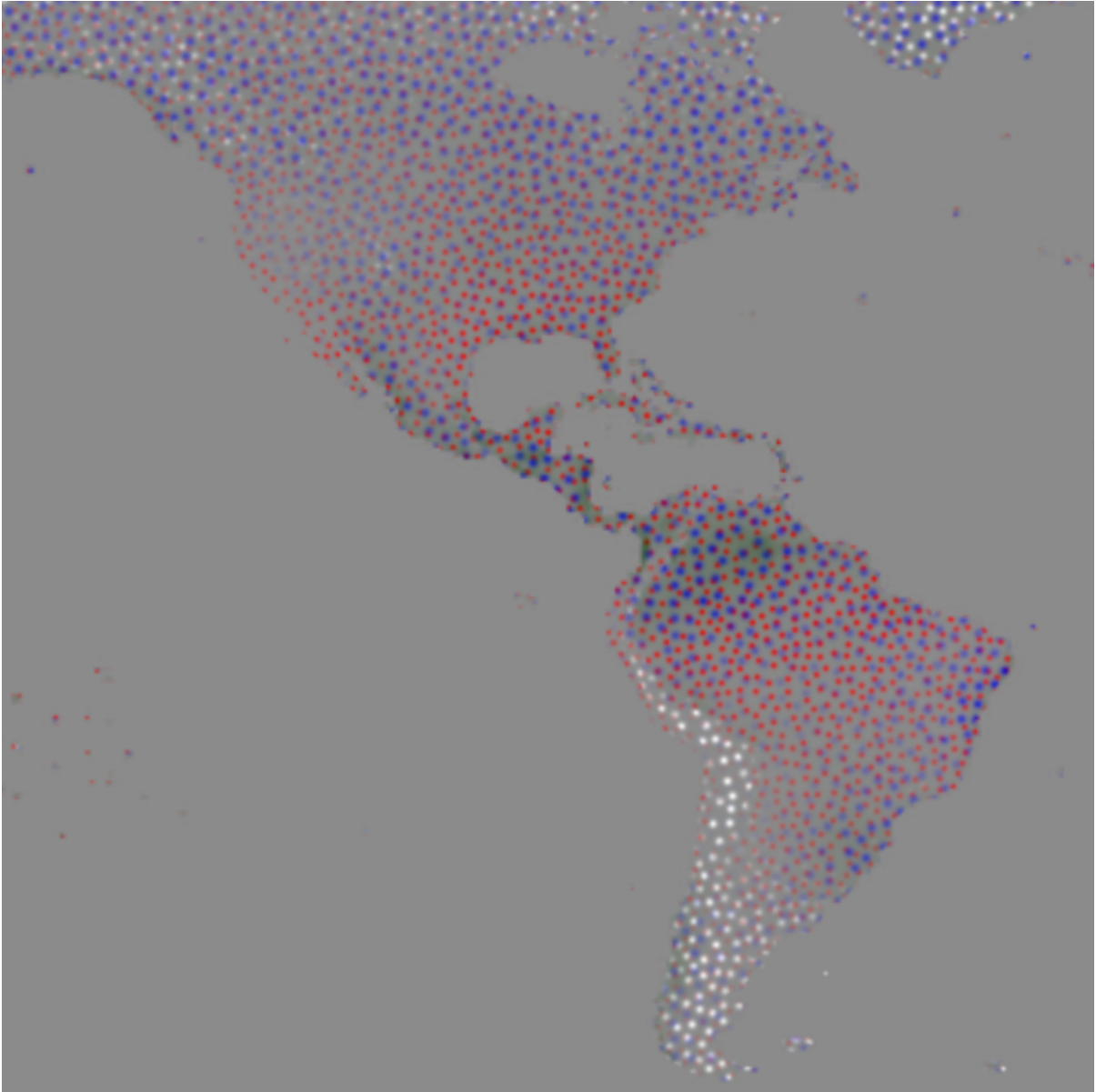


Figure 1.5: Average climate conditions for the month of August during the years 1981-1990. Ground-frost frequency (days) is shown in white, precipitation (millimeters/day) in blue, vapor pressure (hPA) in red, and wet-day frequency (days) in green.

To fully solve problems in image congestion and visual obstruction caused by overlapping layers, the final image displayed must be both interactive and dynamic. By *interactive* I mean that the image is generated by a person who interacts with the system to find a good mapping of variables to Gaussian arrays. The DDS software system provides multiple display parameters for each visual layer that can be interactively manipulated by the viewer: the hue and size of the spots, and animation parameters such as the direction and speed of motion. The viewer may also manipulate the order in which the layers are applied. A *dynamic* display is animated by translating the Gaussian sampling array of a layer as a texture unit so the spots move across the background in unison while maintaining constant relative placement. A moving layer is seen as visually separate from static layers. Because each layer moves as a unit, visual separation among multiple moving layers can be achieved by assigning each a different speed and/or direction of motion.

Another benefit of animation is that the moving Gaussians sample different points in the variable's spatial extent over time, so all data values are eventually sampled and displayed. The viewer is able to form a complete mental representation of the data while watching the animation. The final visualization image, the accumulation image, and the Gaussian sampling array are all functions of time, t , while the data function remains constant (equations 1.2 and 1.3).

$$\alpha_{tk}(i,j) = F_k(i,j) * G_{tk}(i,j), \quad 1.2$$

$$Accum_{tk}(i,j) = (1 - \alpha_{tk}(i,j)) * Accum_{t,k-1}(i,j) + \alpha_{tk}(i,j) * C_k \quad 1.3$$

Data-Driven Spots Bump-Mapping

Bump-mapping, first developed by Blinn [Blinn, 1978], is now a standard computer graphics technique for adding the illusion of detailed surface geometry without actually altering the surface representation. Bump-mapping modifies the surface normals, causing a modification of the shading, and the perception of a bumpy surface. How the bump-mapped surface looks depends on the surface geometry and surface normals, the illumination model used to light the surface, the location of the light source, and the location of the viewer.

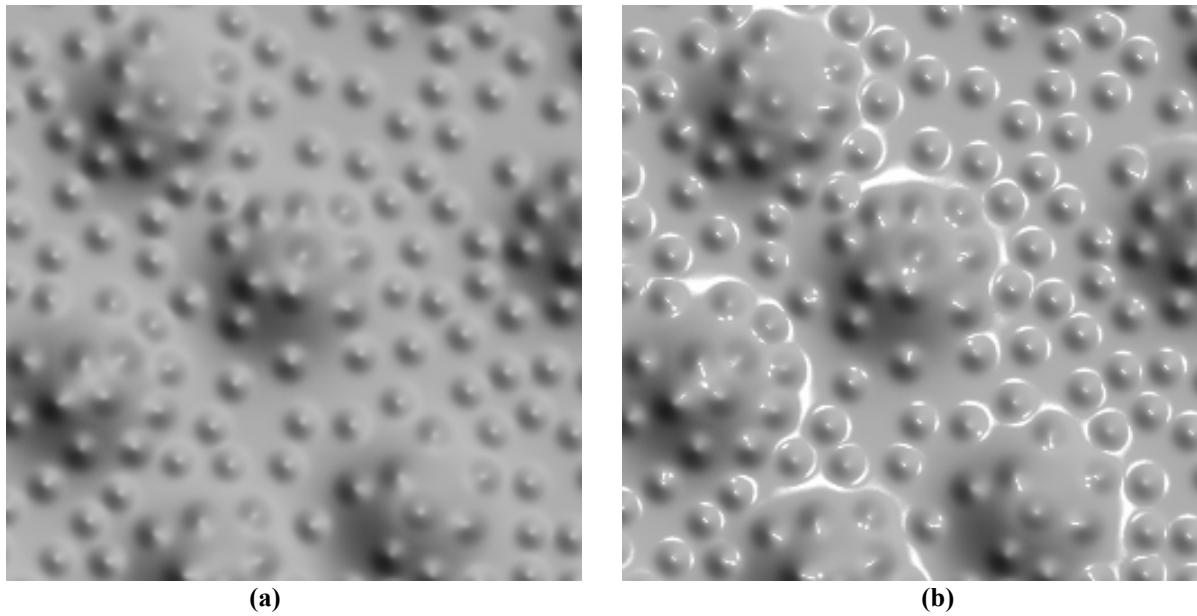


Figure 1.6: Large and small bumps shown with diffuse only reflection (a) and with both diffuse and specular reflection (b). The surface is lit by two different light sources: one from above and another from the right of the viewer.

DDS uses Phong illumination [1975], which models both diffuse (scattered) reflection and specular (directional) reflection of light from a surface. Diffuse reflection is view-independent and models a random scattering of light from matte surface materials such as clay or chalk. Specular reflection is view-dependent and models bright highlights from smooth surfaces such as plastic or polished metal. Specular reflection produces bright highlights that can be seen when the viewer looks in the direction of the reflected light, thus it depends on the angle the light is reflected about the surface normal and on the view direction. Figure 1.6 shows an example of two bump-maps on the same surface, one with diffuse-only reflection and the other with both diffuse and specular reflection. There are two light sources modeled in the images, one from above the viewer and another from the right of the viewer.

Diffuse and specular reflection models both use the surface normals in the lighting equation – this means either can be used to illuminate a bump-mapped surface. When the surface normals are used in the lighting calculation, changes in surface shading can be produced by changing the direction of the normal – without changing the underlying surface geometry. Bump-mapping varies the surface normal of a smooth surface to produce shading that would occur if the surface were bumpy – the normals are true to the bumps, but the geometry is not. Because the geometry does not match the normals, bump-mapping produces several inconsistencies in the surface appearance. For example a bump-mapped surface will look incorrectly smooth when seen in silhouette. In DDS the surface is always oriented perpendicular to the view direction, with the lights positioned slightly above and to the right of the view direction; thus the illusion of the bump-mapped surface is seen with minimal artifact.

The Gaussian sampling array, $G_k(i,j)$, is processed to produce a normal-perturbation map. The normal-perturbation map specifies $dx(i,j)$, $dy(i,j)$, values that are added to the normal of a flat surface to produce the appearance of height-varying Gaussian bumps across the surface. To display the data, $F_k(i,j)$, with bump-mapping, the normal perturbation map is multiplied by the function $F_k(i,j)$, at each grid point i,j , producing a variable-modulated normal, $N(i,j)$. (equation 1.4). The normal, $N(i,j)$, is then re-normalized to unit length. A visualization image with multiple bump-maps is made by adding the variable-modulated perturbation maps together.

$$\begin{aligned}
 N_x(i,j) &= F_k(i,j) * d_{xk}(i,j) \\
 N_y(i,j) &= F_k(i,j) * d_{yk}(i,j) \\
 N_z(i,j) &= 1.0
 \end{aligned}
 \tag{1.4}$$

When the Gaussian spot array is displayed with DDS bump-mapping it appears as a three-dimensional surface with raised bumps protruding from the surface. High data values produce taller-appearing bumps and low data values produce shallower-appearing bumps. The DDS visualization technique can display multiple variable-modulated bump-maps on one surface; they appear as bumps on top of bumps. The size, i.e. standard deviation, of the Gaussians must be different for each layer so the viewer can distinguish among bump-mapped variables. The order in which the bump layers are added to the surface does not matter. Figure 1.7 shows a bump-mapped image of two of the climate variables: precipitation with smaller bumps and wet-day frequency with larger bumps.

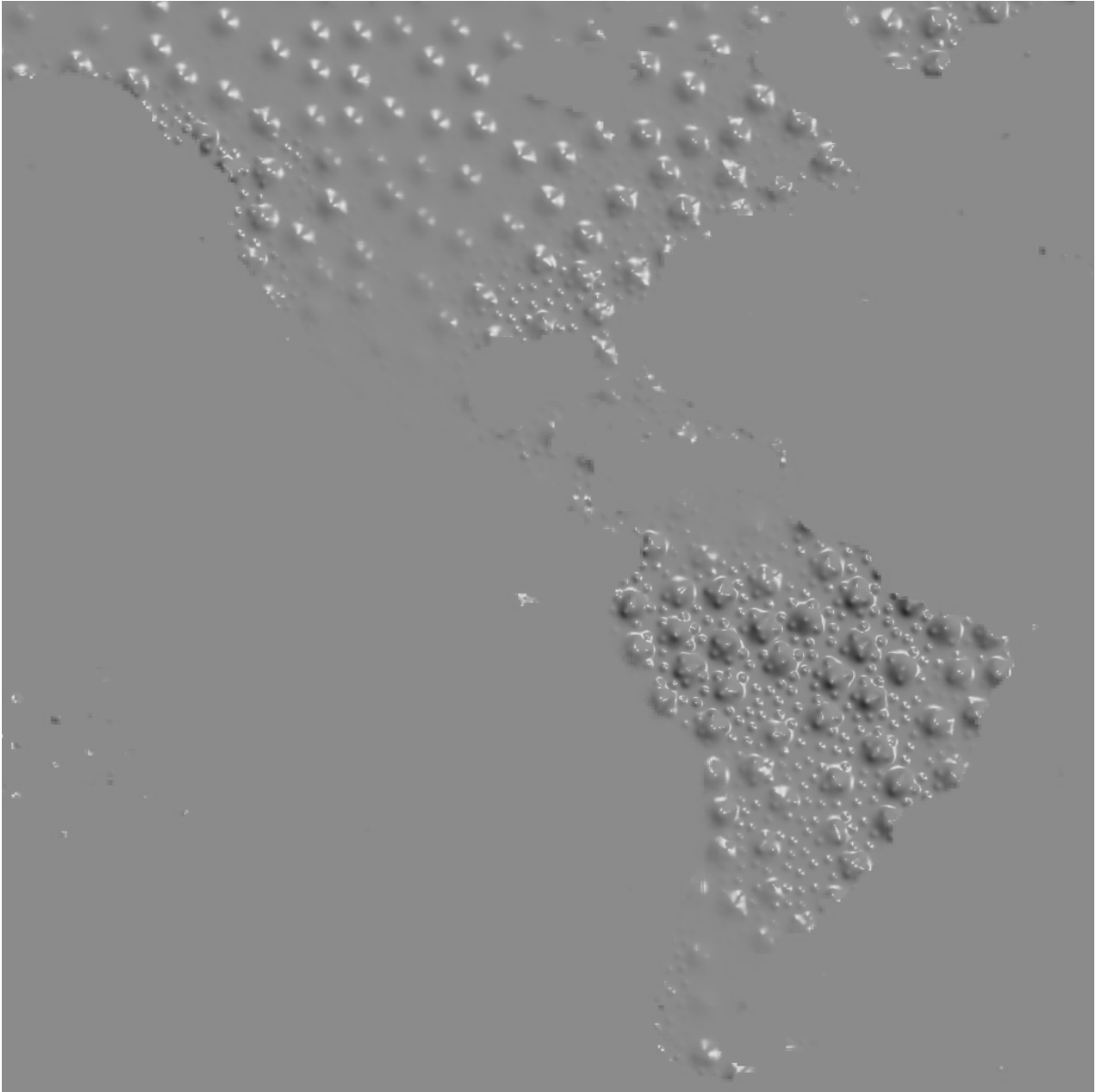


Figure 1.7: Bump-mapping display of average climate conditions for the month of February during the years 1981-1990. Precipitation (millimeters/day) is shown in small bumps and wet-day frequency (days) in large bumps.

The experimental validation of DDS showed that, in the presence of distractors, viewers had more difficulty discriminating among multiple bump-mapped layers than among multiple alpha-blended layers. The pilot experiment tested up to four bump-mapped layers in combination with up to five alpha-blended layers, for a total of nine layers in an image, where both alpha-blended and bump-mapped layers were specified alternately as targets while the other layers served alternately as distractors. Task performance for the bump-mapped targets showed that the viewers confused two different bump-mapped layers that had similar sized bumps – the two smallest sizes. The main experiment reduced the number of bump-mapped layers in the images from four to three, and increased the differences in bump standard deviations. Viewers in the second experiment confused the bump layers less, but there was still some evidence that multiple bump layers diminished performance. The question of how many bump layers can be displayed at once has not been fully answered. I believe that an expert user, someone who is very familiar with the DDS display technique, could discriminate up to four or five different bump mapped layers in the presence of alpha-blended layers. However, the difficulty of discrimination and therefore the likelihood of mistaking one layer for another increase significantly with the number of different bump-mapped layers in an image.

Combining bump-mapped layers with alpha-blended layers means that the viewer must adapt him or herself to two different display methods. When multiple layers of information are displayed in a single image, the degree of layer consistency affects the number of layers that can be discriminated by increasing or decreasing the viewer's cognitive load. Displaying multiple layers with different techniques increases the mental load on the viewer by requiring him or her to learn and remember the meaning of each display technique. By keeping the display techniques uniform across layers, the viewer must only learn one mapping, which frees him or her to process more data. By reducing the number of display techniques that the viewer must know, it is possible to increase the amount of data he or she can process and understand; Tufte discusses this benefit of a uniform display in his book [Tufte, 1983].

Experimental Validation of Data-Driven Spots

I designed pilot and main human-factors experimental evaluations of the Data-Driven Spot display technique. The studies sought to determine whether multiple variables are visually discriminable when displayed together in a single image. Chapter Three presents the pilot and main experiments fully; the main experiment is outlined briefly below. Three questions examined in the study are:

Research Questions

Question One: Does overlay provide a performance benefit for an overlap estimation task and an intersection-sketch task? How do participants perform with overlaid multi-layer DDS images compared to single-layer DDS images shown side-by-side?

Question Two: Where is the benefit crossover? If presenting targets overlaid in a single image allows for more accurate performance for an overlap estimation task and an intersection-sketch task, then is there a point where the multitude of additional layers causes enough visual interference that the task is more accurately performed looking at the targets side-by-side?

Question Three: Are DDS alpha-blended layers more visually salient than DDS bump-mapped layers in a multi-layer visualization? How do the display parameters of individual target DDS layers influence their visual salience in the presence of distractors?

Methods

The test images are multi-layer DDS images of binary functions, $F_k(i,j)$, of coherent shapes, such as squares, circles, triangles, ellipses and rectangles. Each $F_k(i,j)$ is displayed on a separate visual layer, sampled with either DDS alpha-blended spots or DDS bump-mapping. An example of a trial image is shown in Figure 1.8.

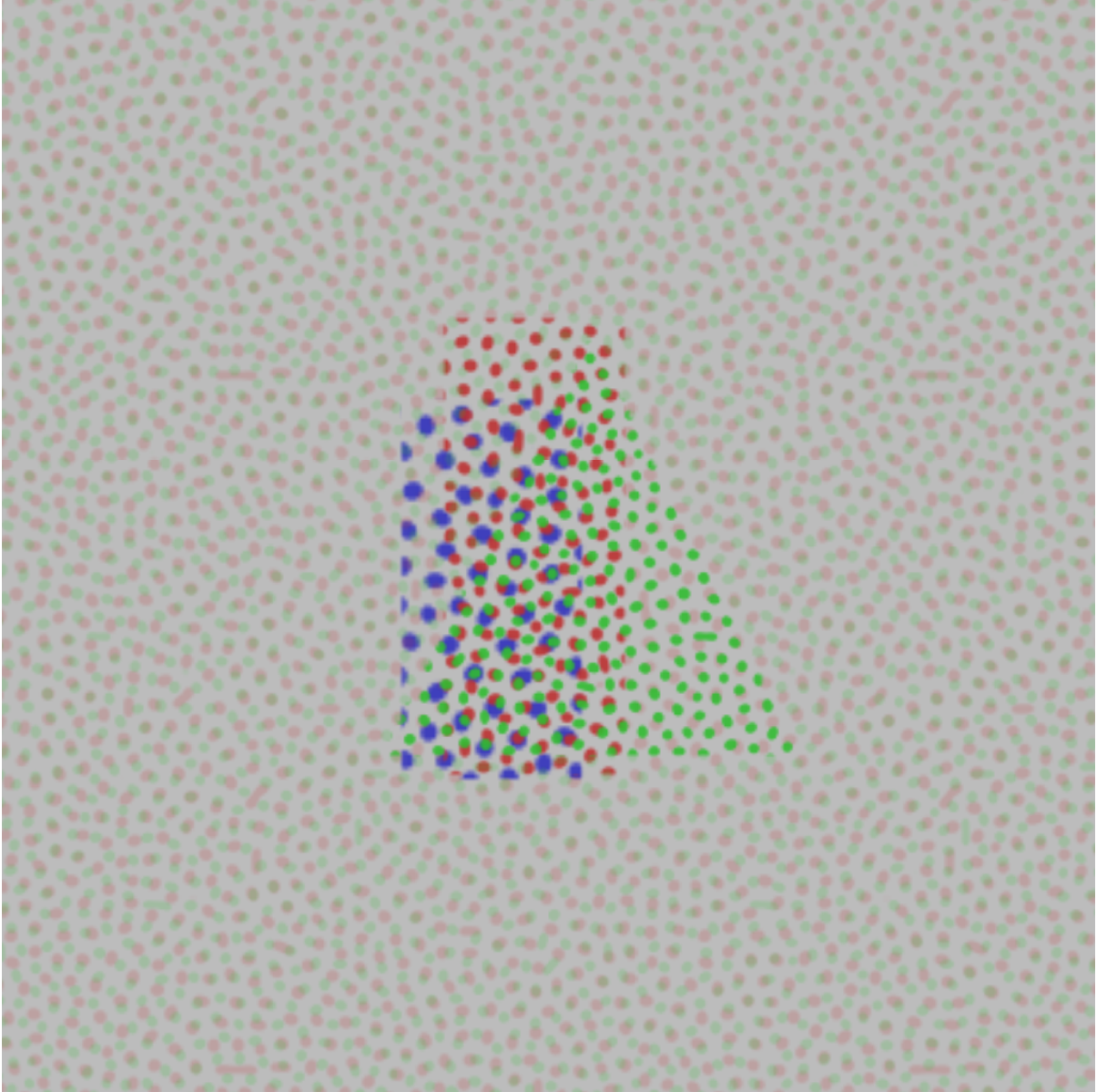


Figure 1.8: An example trial image from the study. The two target layers are the vertical rectangle shown in red alpha-blended spots and triangle shown in green alpha-blended spots. The rectangle in blue spots serves as a distractor layer.

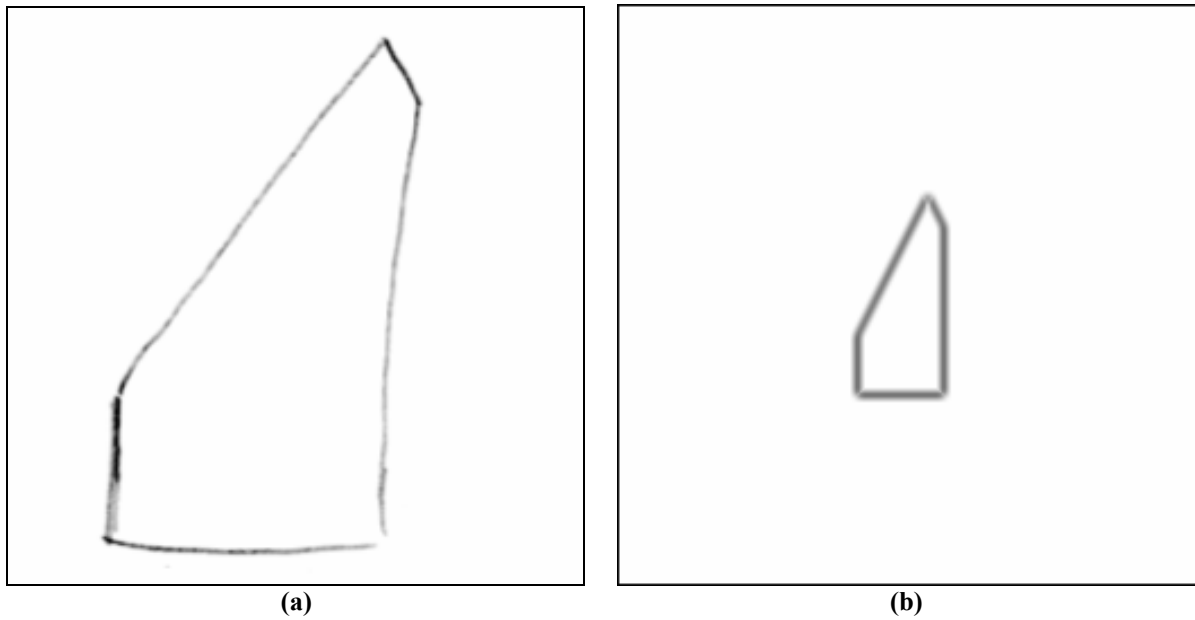
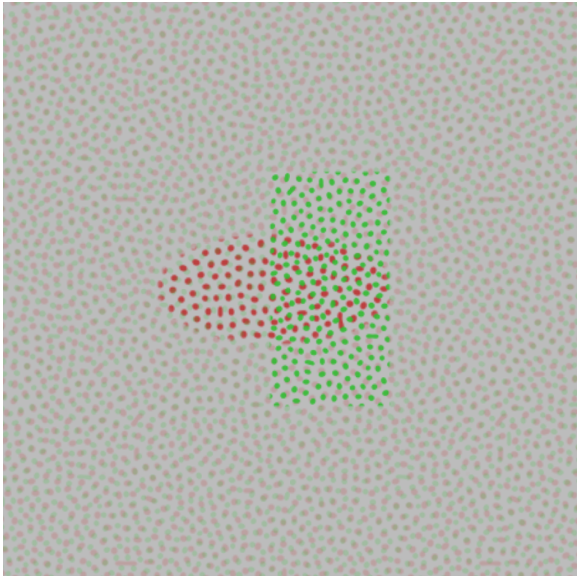


Figure 1.9: An example intersection sketch for the trial image in Figure 1.8 is shown in (a). The correct intersection is in (b).

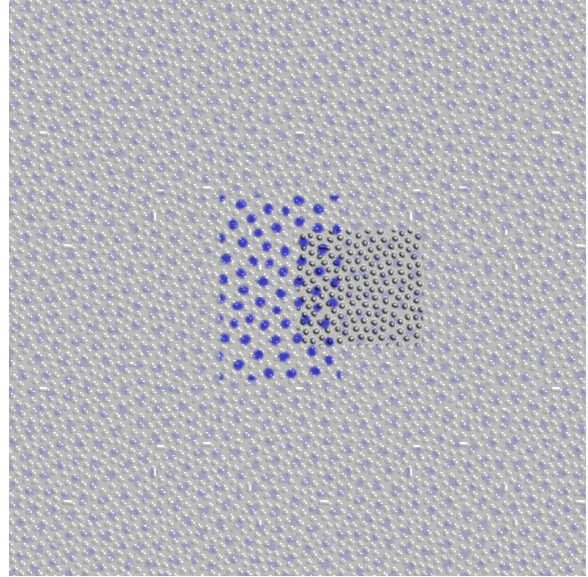
During an experimental trial, participants are asked to focus on two of the layers: target A and target B. Participants perform two tasks. First, they give a numerical estimate of the percentage of target A that is overlapped by target B. Second, they sketch the intersection region of targets A and B. There was no time limit for either task. An intersection sketch and the correct intersection are shown in Figure 1.9.

Two independent variables are systematically manipulated during the experimental trials. The first independent variable that is manipulated across trials is how the data values of the two target shapes are displayed. The first method for displaying the data is DDS alpha-blending, where different layers are distinguished by hue and by standard deviation of the Gaussians. The second method for displaying the data is DDS bump-mapping, where layers are distinguished by the sizes of the bumps.

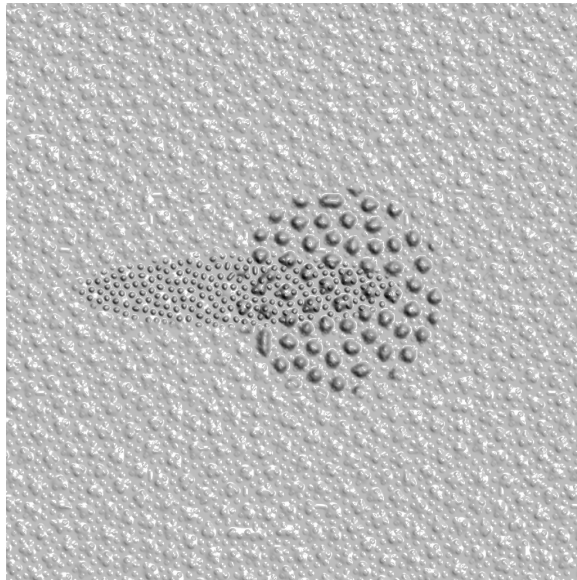
The *Target Display Type* condition has three categories. In the first category both target layers are displayed with DDS alpha-blending, distinguished by the hue and sizes of the spots. In the second category both target layers are displayed with DDS bump-mapping, distinguished by the sizes of the bumps. In the third category one target layer is displayed with DDS alpha-blending and the other target layer is displayed with DDS bump-mapping. These categories are referred to as *Color-Color*, *Bump-Bump*, and *Color-Bump*; examples are shown in Figure 1.10.



(a)



(b)



(c)

Figure 1.10: Examples of the different *Target Display Types*. *Color-Color* in (a), *Color-Bump* in (b), and *Bump-Bump* in (c).

The second independent variable manipulated during the experimental trials is the *Display Condition*; it has 10 categories. For one category of *Display Condition* target A is displayed in one DDS layer with the intersection region of target A and target B displayed in a second DDS layer. Both DDS layers are overlaid in a single image. In this category of *Display Condition*, target B is not displayed, nor are any other shapes. This category is referred to as *Intersection-View*.

The second category of *Display Condition* presents target A and target B in two separate DDS layers, not overlaid in a single image, but in two images shown side by side. This category of *Display Condition* is referred to as *Side-by-side-View*.

The remaining eight categories of *Display Condition* present target A and target B in two separate DDS layers, overlaid in one image. In these categories, the image contain between zero and seven extra shapes, each in a separate DDS layer. All shapes are overlaid in the test image. These categories are referred to as *C0-C7* respectively. The term *distractor* is used for the non-target layer in categories *C0-C7*. Figures 1.11 through 1.14 show examples of *Display Condition* levels *Side-by-side view*, *Intersection view*, *C0*, *C1*, *C3*, *C5*, and *C7*.

The target layers are displayed with up to seven distractor layers, representing images with up to nine single-valued binary functions. The results of the experimental validation of the DDS data visualization technique lead to the following:

Thesis Statement

People can discriminate up to nine two-dimensional single-valued binary functions, $F_k(i,j)$, of coherent shapes displayed with Data-Driven Spots.

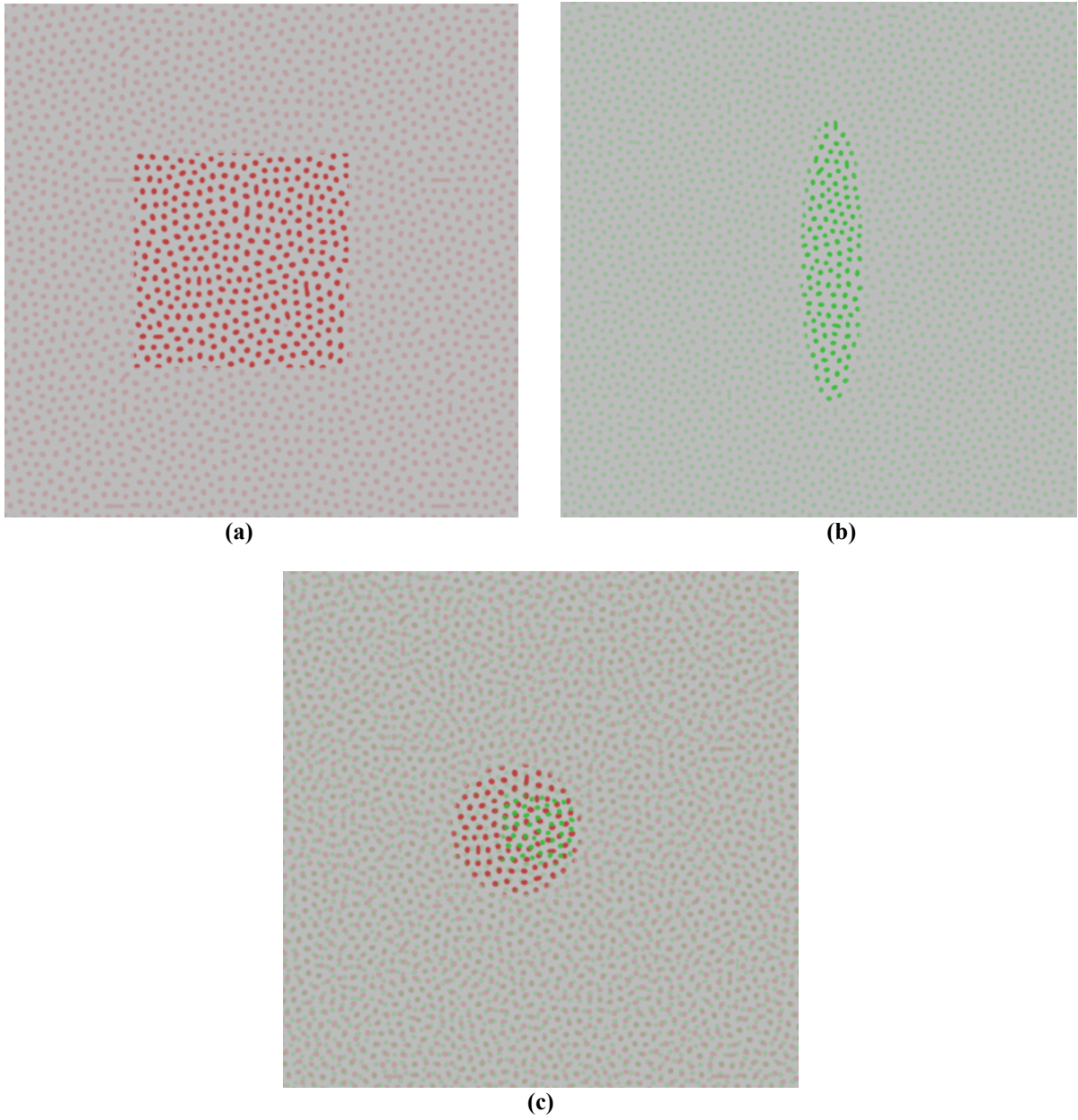


Figure 1.11: Examples of the different *Display Conditions*. *Side-by-side view* in (a) and (b), and *Intersection view* in (c).

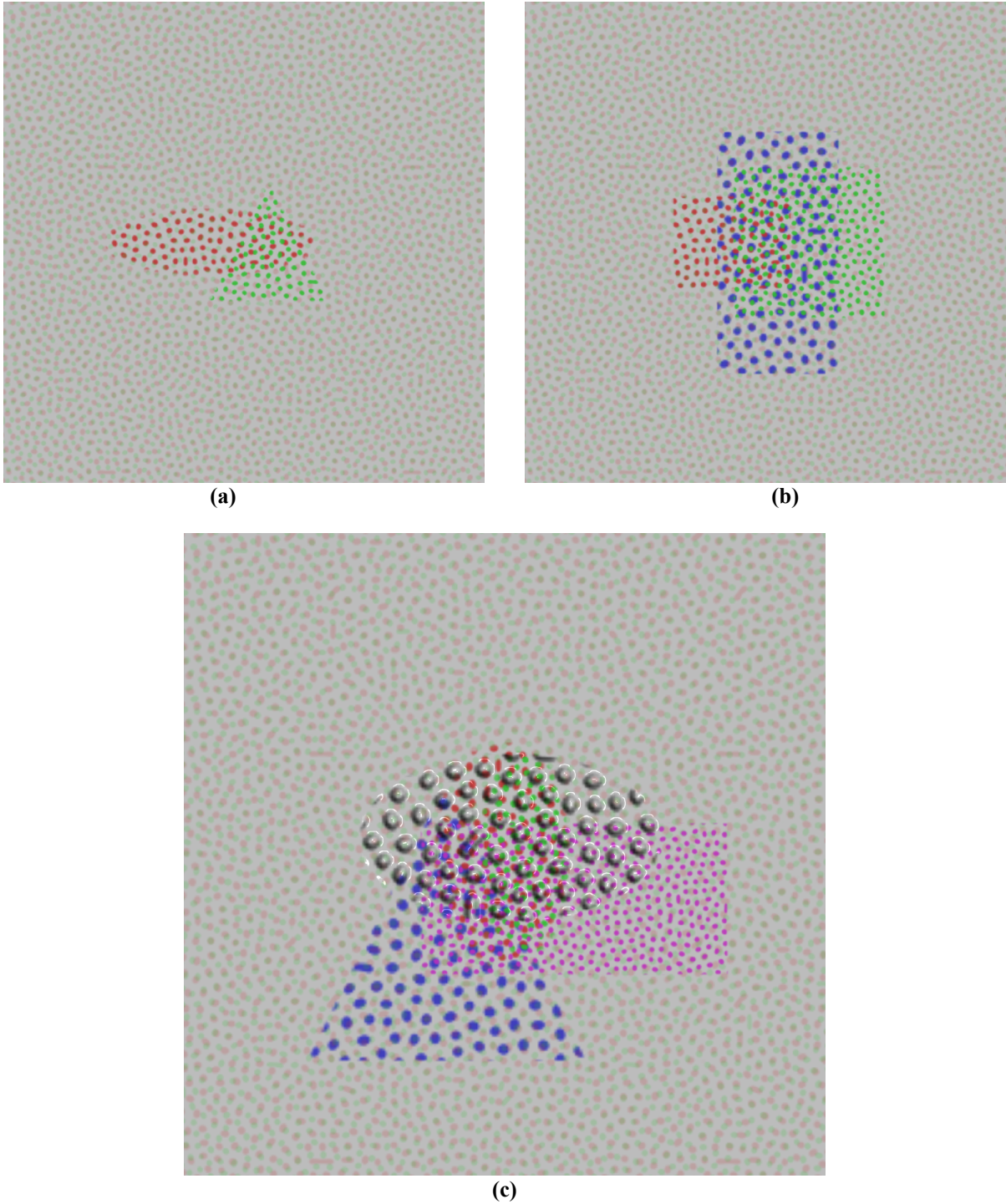


Figure 1.12: Examples of the different *Display Conditions*. *C0* in (a) with no distractors, targets are a red horizontal oval and a green triangle. *C1* in (b), targets are a red rectangle and a green square; the blue spot rectangle is the distractor. *C3* in (c), targets are a red vertical oval and a green vertical oval; the blue spot triangle, purple spot rectangle, and large bump oval are distractors.

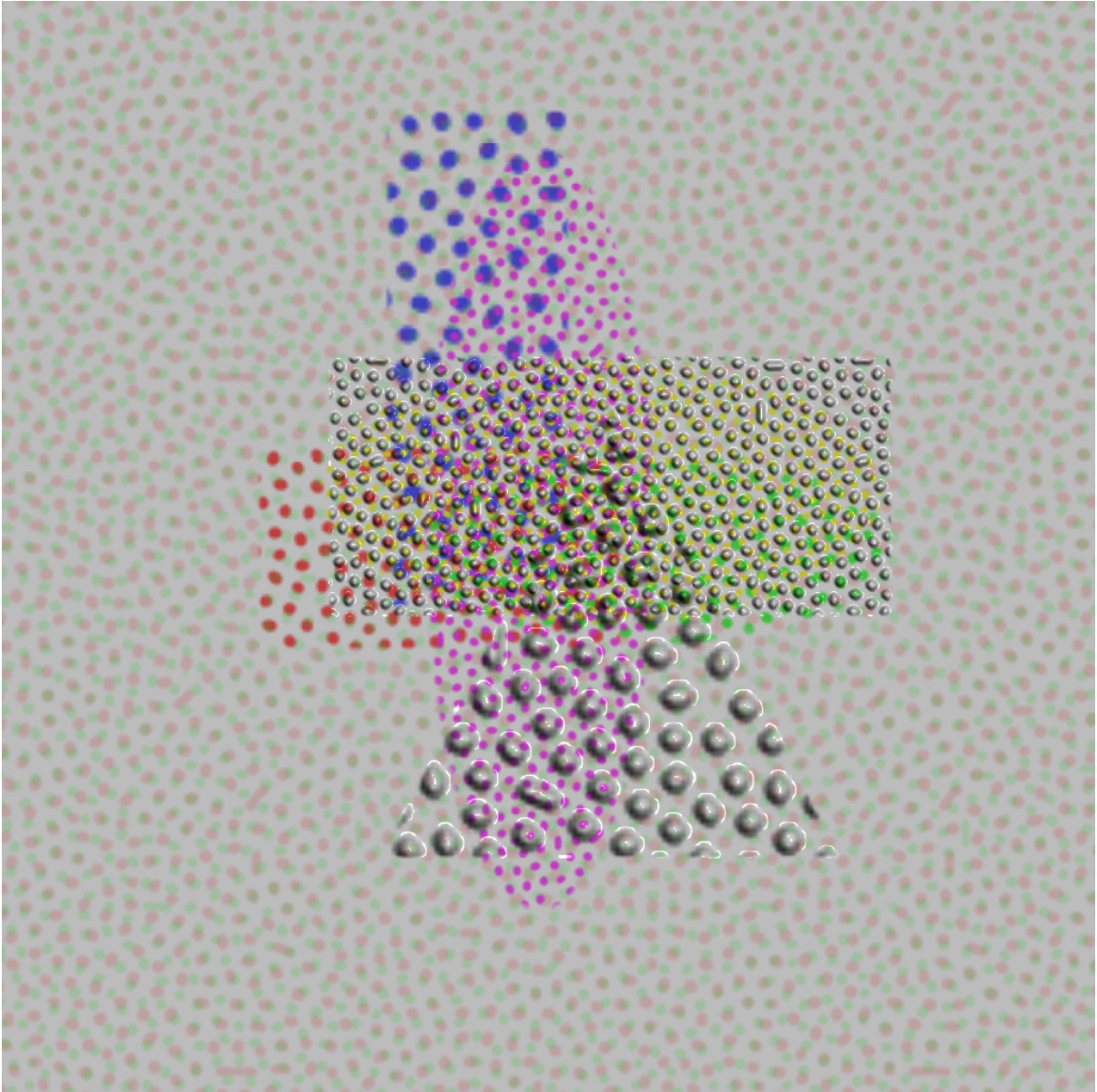


Figure 1.13: Examples of *Display Condition C5*. Targets are a red spot horizontal rectangle and a green spot horizontal oval. Distractors are a blue spot rectangle, a purple spot oval, a large bump triangle, medium bump rectangle, and a yellow spot oval.

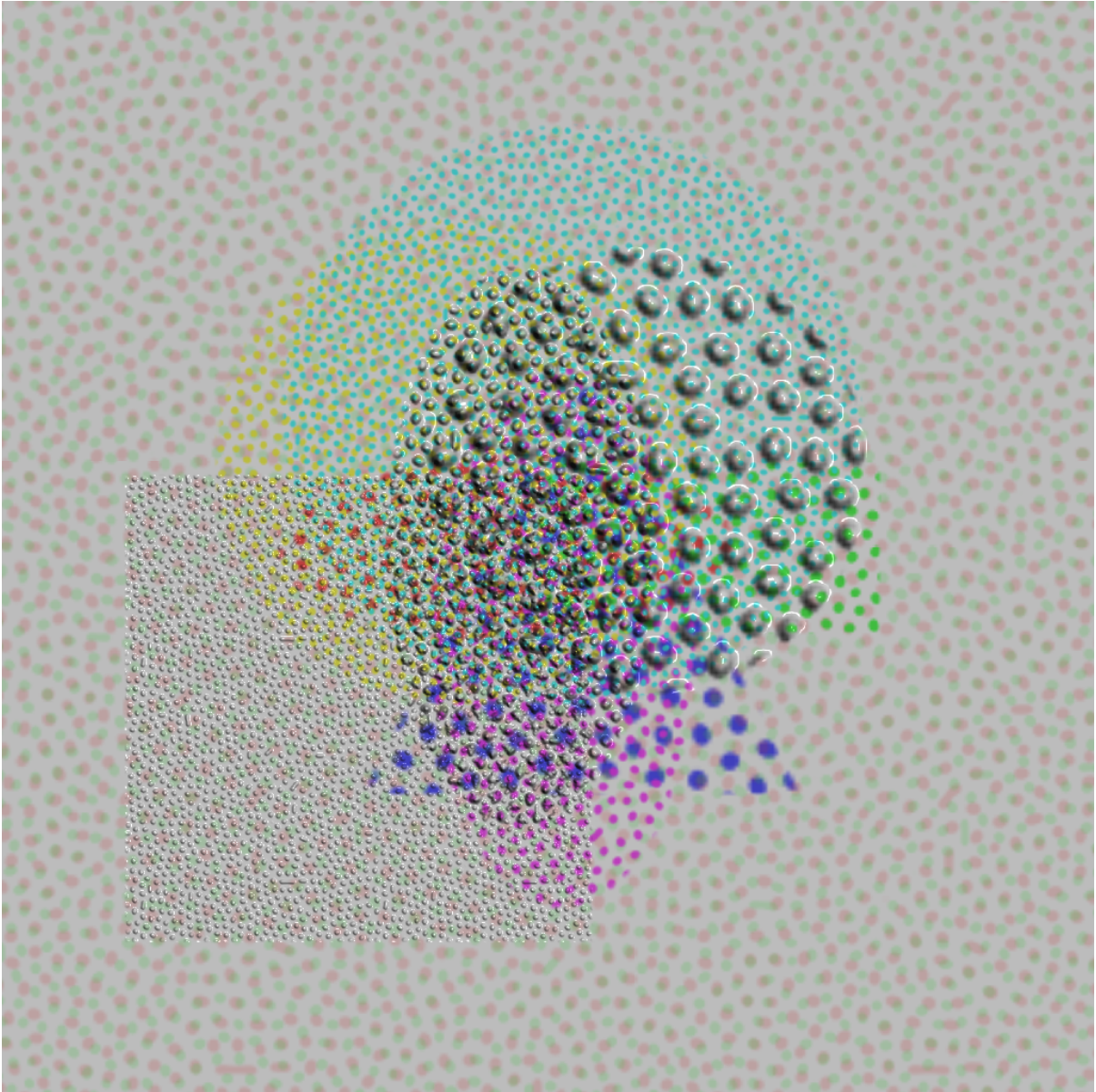


Figure 1.14: Examples of *Display Condition C7*. Targets are a red spot horizontal oval and a green spot horizontal rectangle. Distractors are a blue spot triangle, a purple spot oval, a large bump circle, medium bump oval, and a yellow spot circle, small bump square, and a cyan spot circle.

Even in images with this much overlap and congestion, people performed twice as well as for images in the *Side-by-side view*, which *never* had distractors.

Experimental Results

Two performance measures are calculated for each category of *Display Condition* for each of the three sessions. The first metric represents the error in overlap estimate, compared to the correct percentage of overlap. The error is calculated as the absolute value of the difference. The second metric represents the sketch score. The sketches are compared to the correct computer-derived shapes, all sketches are graded on a three-point scale of wrong: -1, mostly correct: 0, exact: 1. Sketches are judged to be exact if they match the desired shape in form; positioning and scale of the shape relative to the image boundary is not considered.

Each participant completes three experimental sessions with *Target Display Type* varying across sessions. During a session participants complete five different trials for each of the 10 categories of *Display Condition*. This design produces 50 trials per participant for each session and 150 trials for each participant overall. A total of 30 people participated in the experiment, giving 4500 sketches and 4500 overlap estimates.

Question One: How do participants perform with multi-layer DDS images compared to single-layer DDS images shown side-by-side?

Performance for the *Side-by-Side* is significantly worse than for *C7* ($p < 0.0001$). Error for the *Side-by-side* view is twice that of *C7* for the overlap estimation task. *Target Display Type* is not a significant factor. Both the overlap estimation task and the sketch task produced the same results in the analysis. Figure 1.15 shows a plot comparing performance for the *Color-Color* group conditions *C0-C7* with the *Side-by-side* view. The *Side-by-side* view never had distractors, however, in the graph, performance for the *Side-by-side* view is drawn across the conditions *C0-C7* to aid comparisons.

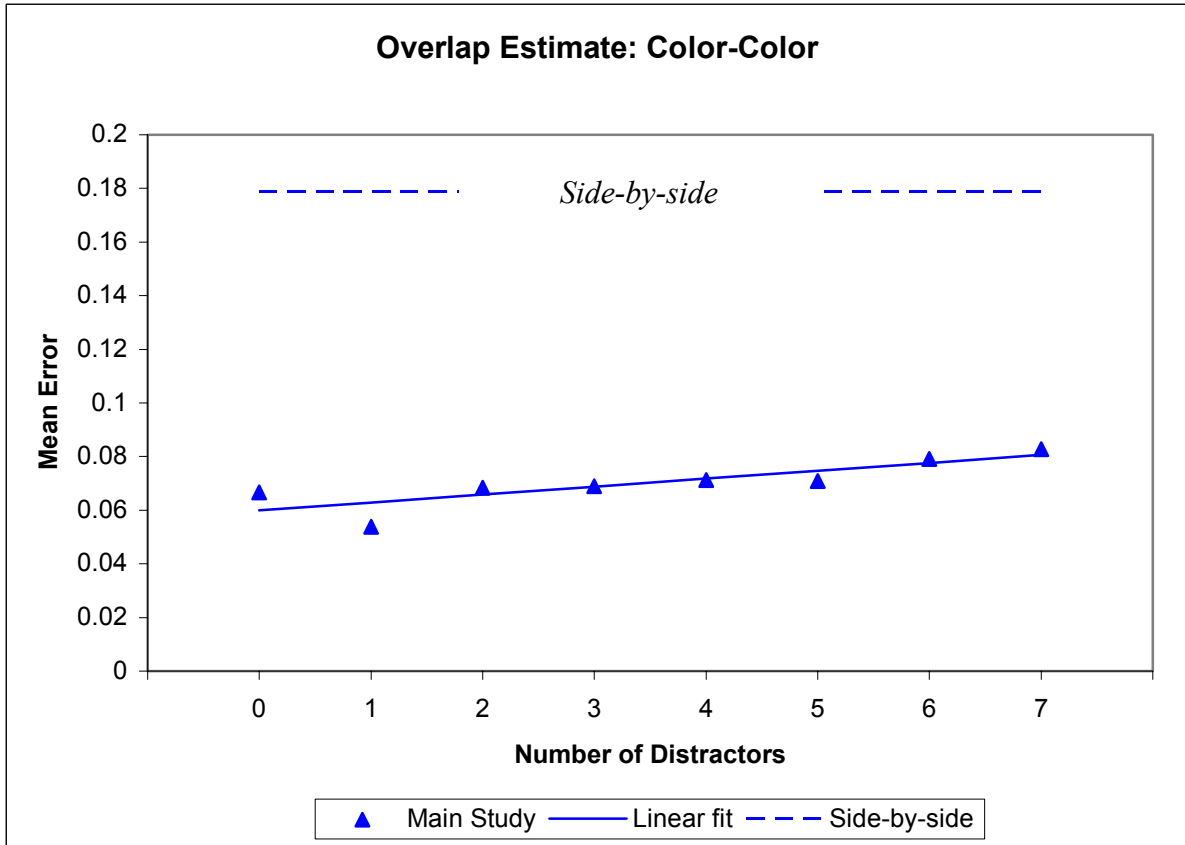


Figure 1.15: Performance for the *Color-Color* group: conditions *C0-C7* are shown in blue triangles; the solid blue line is the result of the linear regression analysis which gives both intercept and slope. The *Side-by-side* condition is shown in the horizontal dashed line at the top of the graph. The *Side-by-side* view never had distractors, however, in the graph, performance for the *Side-by-side* view is drawn across the conditions *C0-C7* to aid comparisons. Error for the *side-by-side* view was significantly worse than for all the overlapping views.

Question Two: Is there a point where the multitude of additional layers causes enough visual interference that the task is more accurately performed looking at the targets side-by-side?

Performance for the overlap conditions is always better than for the *Side-by-side* view, for up to nine layers tested in the experiment. A test for zero slope finds significant non-zero slopes for all three conditions: *Color-Color* slope is 0.003 ($p = 0.0028$); *Bump-Bump* slope is 0.004 ($p = 0.0008$); *Color-Bump* slope is 0.007 ($p = 0.0003$). The difference in performance for zero distractors and seven distractors predicted by this analysis is only 2 percentage points for the *Color-Color* group, 2.86 percentage points for the *Bump-Bump* group, and only slightly more: 5 percentage points for the *Color-Bump* group.

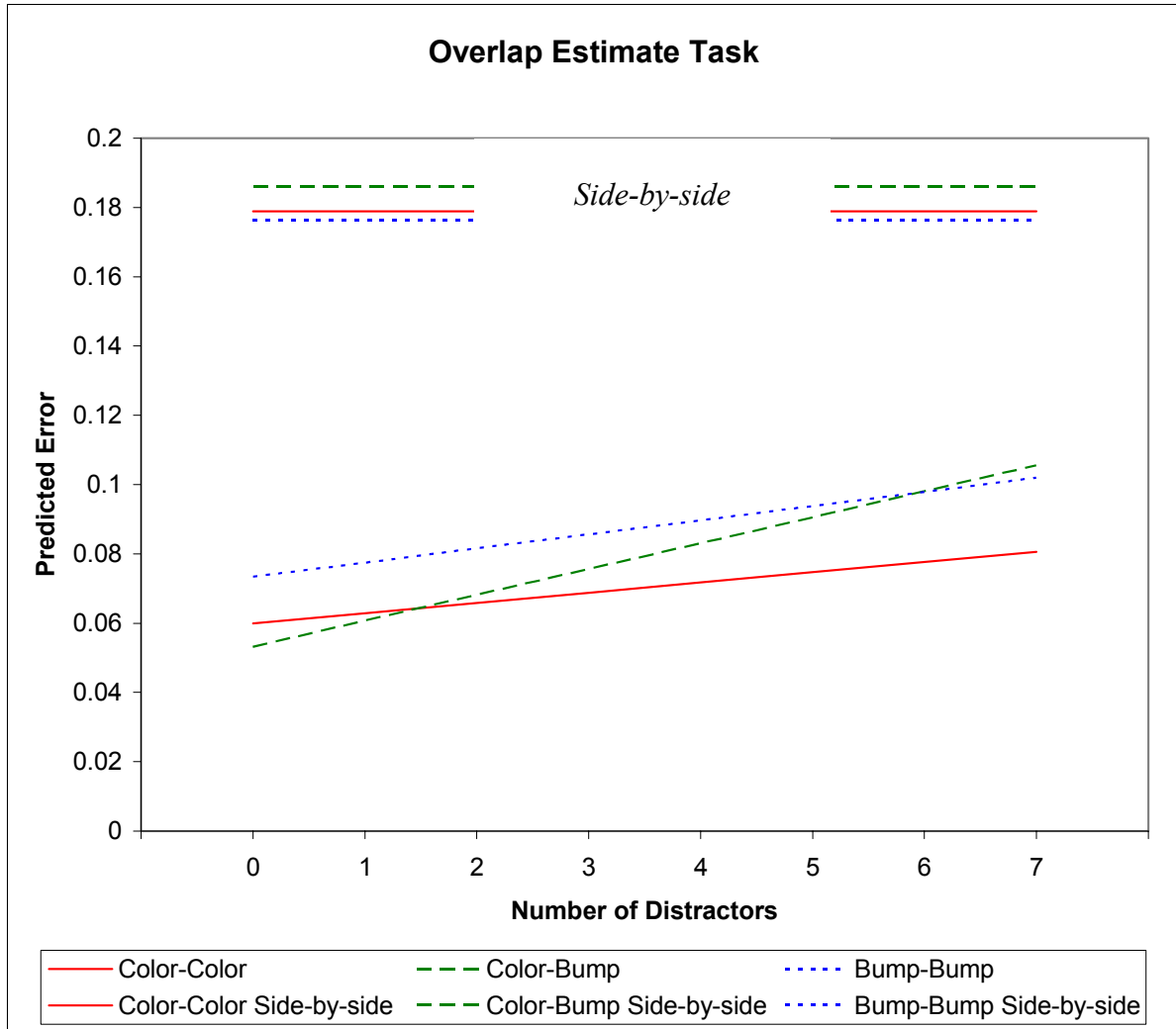


Figure 1.16: Linear fits for the *Color-Color*, *Color-Bump*, and *Bump-Bump* groups. The statistical analysis found a significant non-zero slope for each group. The *Side-by-side* condition is shown in the horizontal lines at the top of the graph. The *Side-by-side* view never had distractors, however, in the graph, performance for the *Side-by-side* view is drawn across the conditions C0-C7 to aid comparisons. Error for the *side-by-side* view was

How many distractors could be handled using each of the techniques? This experiment does not answer that question. As a speculation, linear extrapolation would put the crossing point at 40 distractors for the *Color-Color* group, 25 distractors for the *Bump-Bump* group, and 18 distractors for the *Color-Bump* group. Figure 1.16 shows the predicted performance graphs for the three groups compared to the average *Side-by-side* view performances.

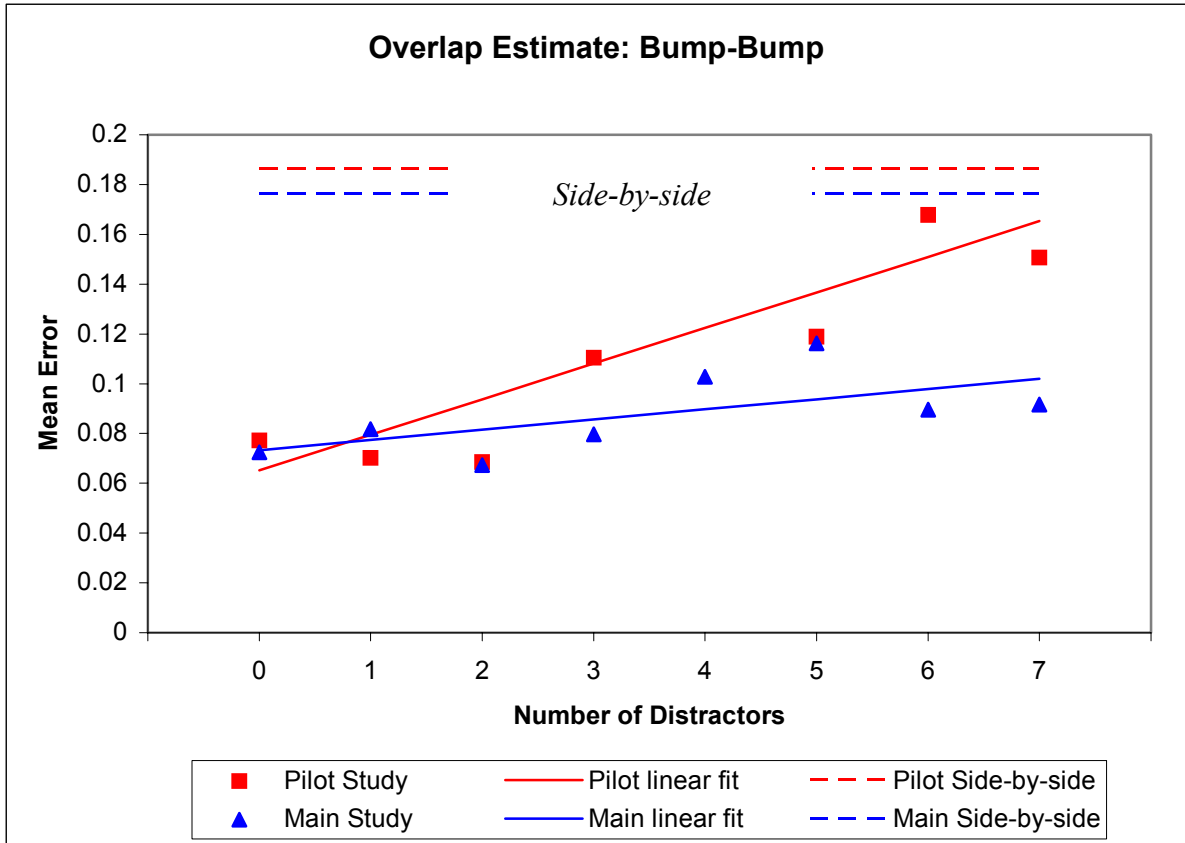


Figure 1.17: Comparison of the *Bump-Bump* groups for the pilot and main studies. Here the distractor effect of bumps on top of other bumps is the most evident. For the four bump case (solid red line), performance degrades, approaching the performance for the *Side-by-side* view, while the three bump case remains significantly below the *Side-by-side* view performance. The *Side-by-side* view never had distractors, however, in the graph, performance for the *Side-by-side* view is drawn across the conditions C0-C7 to aid comparisons.

Question Three: Are DDS alpha-blended layers more visually salient than DDS bump-mapped layers in a multi-layer visualization?

The answer is *yes*. DDS alpha-blended layers are easier to see in a multi-layer image than DDS bump-mapped layers. This is most readily seen when the results for the pilot and main studies are compared, see Figures 1.17 and 1.18. The pilot and main studies test different numbers of DDS alpha-blended and DDS bump-mapped distractors. In the pilot study up to four layers are displayed with DDS bump-mapping, and this causes a significant interference with the participants' ability to see the bump-mapped targets. When the number of layers displayed with bump-mapping is reduced to three, performance dramatically improves – this is immediately obvious when looking at the graph (Figure 1.17).

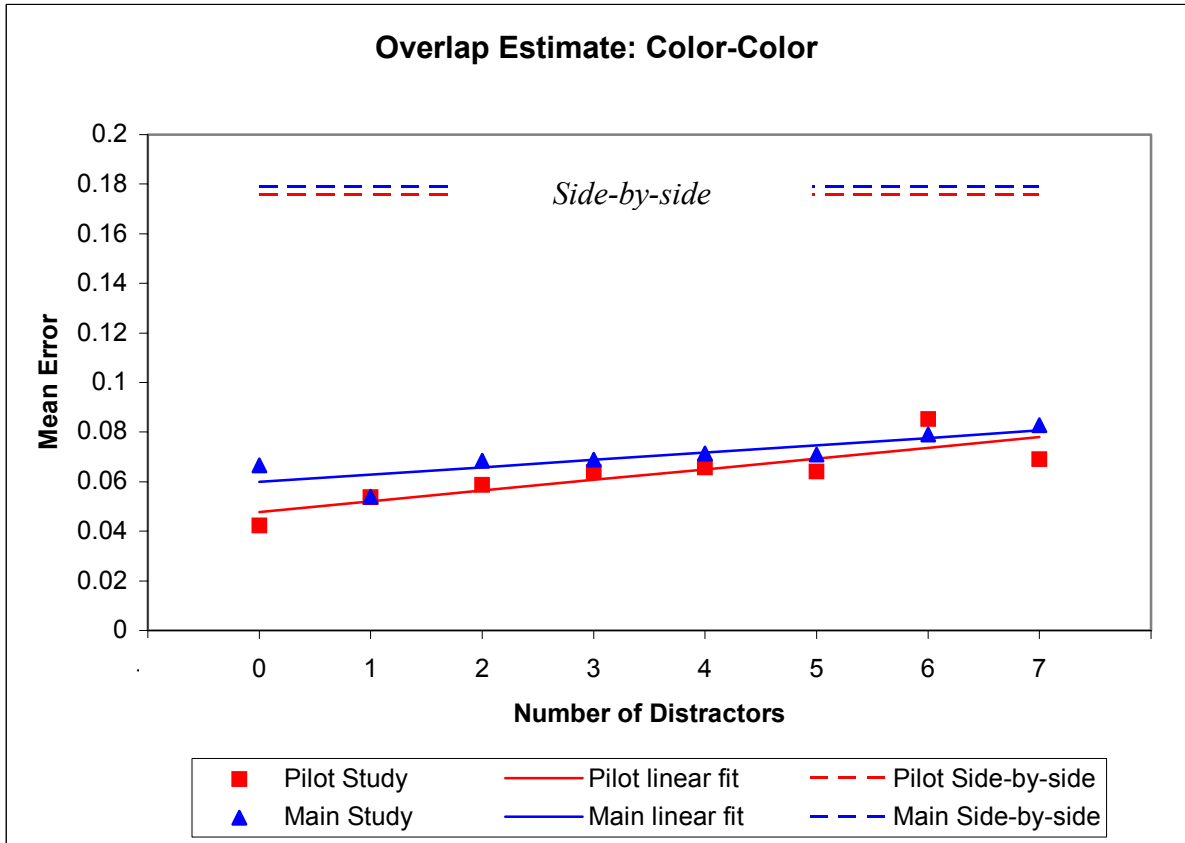


Figure 1.18: Comparison of the *Color-Color* groups for the pilot and main studies. The *Side-by-side* view never had distractors, however, in the graph, performance for the *Side-by-side* view is drawn across the conditions *C0-C7* to aid comparisons. The *Color-Color* results for both studies are identical.

Performance for the *Color-Color* groups remains remarkably consistent between the pilot and main studies, despite several differences in experimental design and the choice of colors and standard deviations used of the alpha-blended layers. This consistency across studies strengthens and validates the result that DDS alpha-blended layers are visually salient in the presence of distractors (Figure 1.18).

Observations

The experimental analysis supports the thesis statement: *People can discriminate up to nine two-dimensional single-valued binary functions, $F_k(i,j)$, of coherent shapes displayed with Data-Driven Spots.* The following list, while not proven by experimental analysis, is supported by many observations of people creating and viewing DDS visualizations.

1. *Animation of a DDS layer dramatically increases the visual salience of that layer* – the viewer’s attention is immediately drawn to the moving spots and the underlying data variable. DDS layers that are low in value, and therefore displayed with faint spots, become much easier to see. I have also noticed that small, spatially isolated areas that may go unnoticed become more salient with animation: the flickering on/off of color as the areas are alternately sampled and not sampled is quite noticeable.

Animation increases the visual salience of boundaries in the data. The moving Gaussians sample different points in the variable’s spatial extent over time, so all data values are eventually sampled and displayed. I believe that the viewer has a complete and accurate mental representation of the boundary while watching the animation.

2. *People can discriminate up to nine two-dimensional single-valued non-binary functions, $F_k(i,j)$, displayed with Data-Driven Spots.* DDS can display a variety of types of 2D single-valued functions, such as scanning electron microscope (SEM), climate, and census data. Many examples are included throughout this dissertation. It is also interesting to note that people can see photographs and perceptual illusions such as the Necker cube, subjective contours, and the face-vase illusion when displayed with DDS, in spite of the spatial sampling. Human perception of visual illusions provides insight into how our visual system works; that we still see the illusions when sampled with DDS indicates that the sampling does not impair our perception of important information in the data.

3. *DDS visualizations enable the viewer to investigate spatial correlations among variables.* DDS is meant as a qualitative tool for scientific exploration during the early stages of investigation when a scientist is searching for regions of interest or visual patterns in the data – areas where he might focus his analysis further. Because multiple spatial variables are layered simultaneously in one image, a scientist can compare how the variables change together across a region and can possibly create new insights into how the variables interact.

4. *Data in each DDS layer is sampled by a 2D array of Gaussian spots to enable layering of multiple variables in a single image with minimal visual obstruction of the data.* The Gaussian spots act as windows through which the data can be seen. When a boundary in the data falls across a spot the spot is cut, and the boundary is clearly visible. Sampling with spots conserves boundary information in the data by creating the perception that the boundary continues between spots through subjective contour illusions. To prevent confusing a boundary in the data with a feature of the sample shape, sampling must be done with a *spot* shape instead of a shape with second-order edge discontinuities, such as a square or triangle – the underlying sample unit cannot contain cusps in its outline.

The smooth edge of the Gaussian also appears to enhance the perception of sharp boundaries in the data, when they cross the Gaussian spots. If the spots were truncated at 2.5 standard deviations, the sharp edge of the Gaussian spot could interfere with the viewer's perception of boundaries in the data.

I have also observed that sampling with Gaussian arrays is better than sampling with reaction-diffusion derived textures. The reaction-diffusion texture generation techniques create spots with non-uniform shapes and non-uniform intensities. The Gaussian arrays are created such that each Gaussian is identical in shape and intensity.

5. *User exploration to find a good mapping of spot density, size, and color, layer order, and animation speed and direction for each variable displayed is important for the successful use of DDS as a display technique.* Creating a visualization that shows the data with minimal visual interference among layers takes a few minutes of exploration. This exploration is not without benefit, however, as interacting with the data visualization increases the user's understanding and knowledge of the data.

Users should decrease the spot standard deviation with successive layers, so that smaller spots are layered on top of larger spots. Likewise, users should decrease the spot sample density with each layer so that lower layers are sampled more densely and upper layers are sampled more sparsely, allowing the underneath layers to be better seen. Lighter colors, such as yellow, cyan and light green, are easier to see when used for the upper layers, whereas darker colors, such as dark blue and green, purple and red can be used for lower layers.

It is important that the user try the variables in different layer positions to find a visualization with acceptable amounts of overlap. Likewise, the user should try different spot densities, colors, and sizes for each variable as the data may be easier to see when displayed with a different hue or size of spot. Animation should also be used during the investigation to test the visibility of low data values and high-frequency boundaries in the static image.

Most importantly, each variable layer should have unique color and size of spots, so that each variable is visually distinct. I believe DDS layers are visually distinct because they are limited in the number of features each can display – one color, one size of spots, and one data variable per layer.

6. DDS can effectively display two-dimensional single-valued functions, $F_k(i,j)$, on a 3D terrain. Examples of the climate data displayed on a topographic map are shown in Chapters Two and Five. I also believe that DDS can be extended to effectively display three-dimensional single-valued functions, $F_k(x,y,z)$, by replacing the 2D Gaussian spots with 3D Gaussian-like blobs.

Chapter Summaries

Chapter Two presents the Data-Driven Spots algorithm and implementation. Generating Gaussian sampling arrays is described, followed by a discussion of other texture techniques. Both DDS alpha-blending and DDS bump-mapping are presented in full detail. The benefits of animating DDS layers, which produces powerful results, is also described. Finally the Chapter concludes with a visualization gallery of different types of data, such as scanning electron microscope, climate, census, and habitat data.

Chapter Three starts with an overview of the experimental evaluation of DDS. A literature review of related work in perceptual psychology follows. The chapter is then divided into the pilot and the main study sections, each of which describes the experimental design, examples, and analysis. A comparison and discussion of the results concludes the Chapter.

Chapter Four provides a review of related work in multidimensional data visualization. A taxonomy of separable and integrated displays is presented, and the techniques are compared. Several images from the related work are presented to illustrate the display methods.

Chapter Five outlines directions for future work, starting with examples of animated DDS layers. Four broad lines of inquiry are discussed: 1) Investigating the powerful impact of animated DDS layers. 2) Investigating further the cognitive and perceptual issues surrounding DDS. How do people interact with the DDS visualization system and how does manipulating the different parameter mappings impact understanding? 2) Extending the DDS visualization technique to higher-dimensional data, such as volume visualization or visualization of data on a three-dimensional surface, such as a globe with elevation mapped to surface height. 3) Augmenting the DDS Visualization Toolkit with knowledge-based systems to aid users in creating visualizations of their data with DDS.