Bump-Mapping

Sampling Variables with the Gaussian Sampling Array -- Bump-Mapping

As with DDS alpha-blending, the Gaussian sampling array is used to spatially sample the variables before display. However, unlike for DDS alpha-blending, the Gaussian sampling array must be converted into a bump-mapped representation before the sampling can occur. To convert the Gaussian sampling array to a bump-map representation, each pixel in the Gaussian sampling array is treated as a height value, and the normal vectors are calculated directly from the height field. Once the normal vectors for the Gaussian height field are calculated, they are stored as normal perturbation vectors – the difference between a normal vector of (0, 0, 1) and the normal corresponding to the bump map.

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 2.4). 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.

$$N_x(i,j) = F_k(i,j) * d_{xk}(i,j)$$
$$N_y(i,j) = F_k(i,j) * d_{yk}(i,j)$$
$$N_z(i,j) = 1.0 \qquad 2.4$$

High data values produce taller-looking bumps, whereas low data values produce shallower-looking bumps. Multiple variable-modulated normal arrays can be added together to produce multiple fields of bumps on a surface. Examples of single and multiple DDS bump-maps are shown in Figures 2.21 and 2.22. The order the layers are added does not matter. Smaller bumps always appear as if they lie on top of larger bumps, larger bumps never overwhelm or cover the smaller bumps. The normal arrays can also be applied to an arbitrary 3D surface, such as a topographic map.

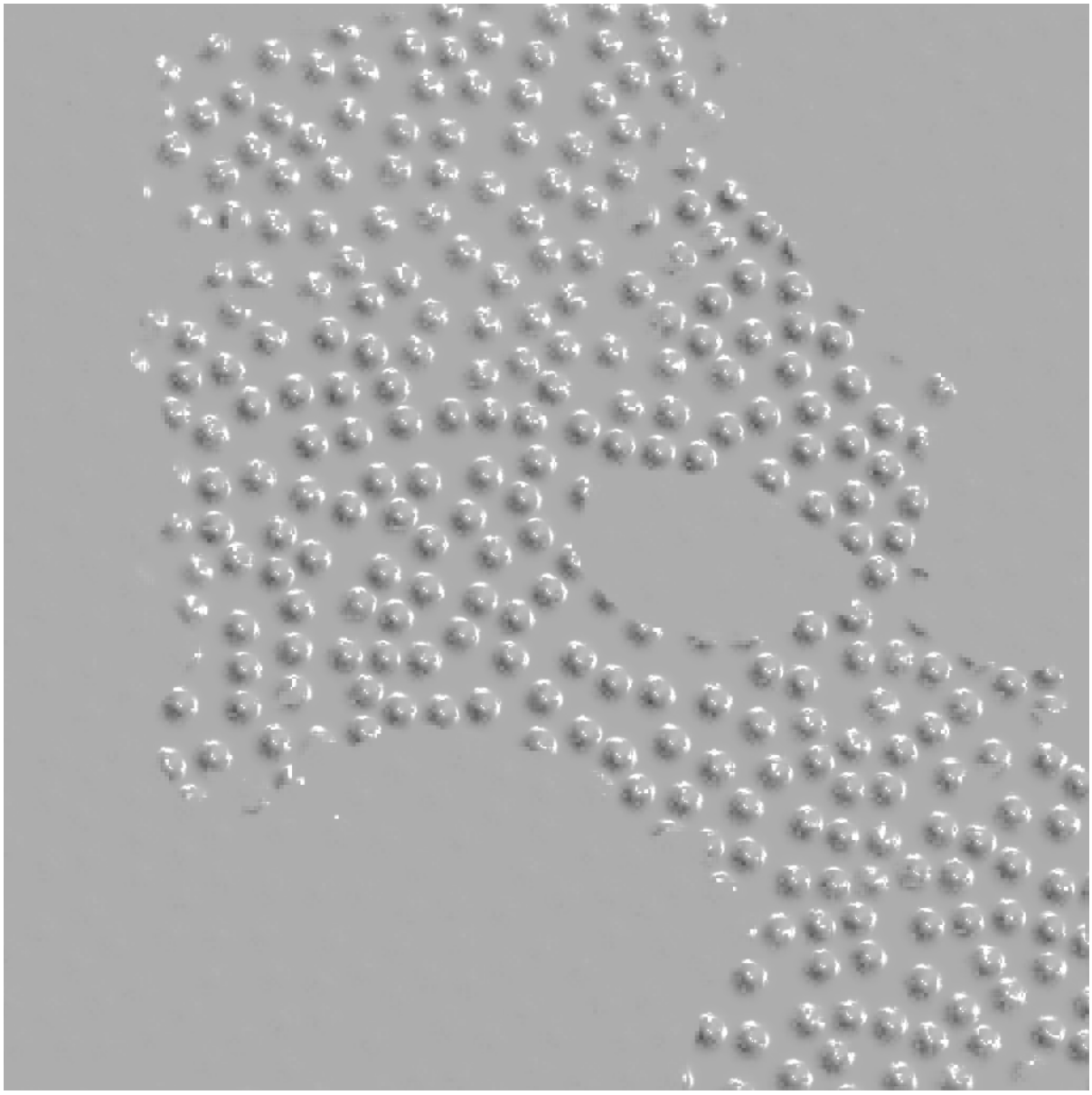

Figure 2.21: DDS bump-mapping showing one layer of the SEM data set

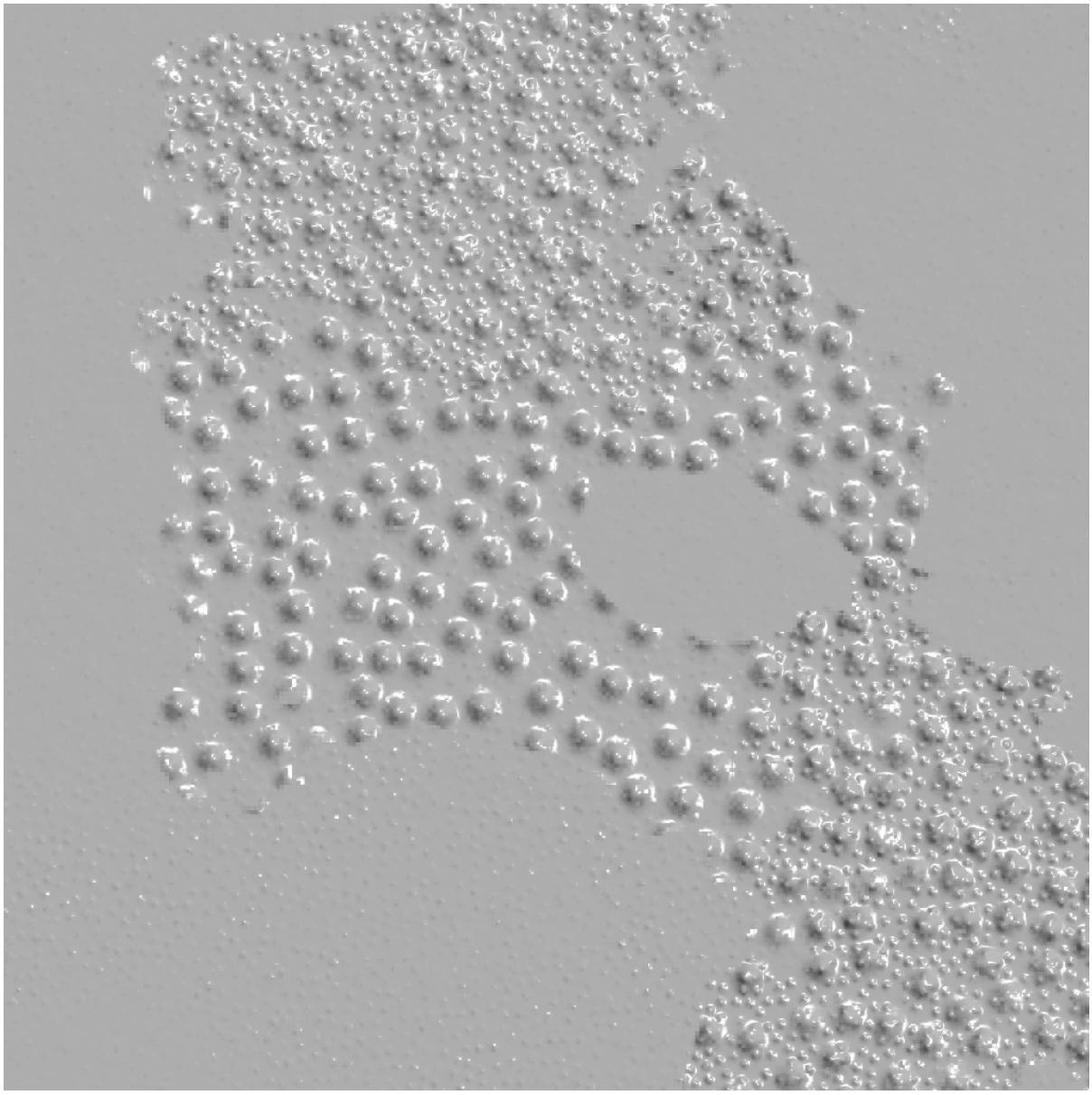

Figure 2.22: DDS bump-mapping showing two layers of the SEM data set. Sulfur is shown in larger bumps and iron is shown in smaller bumps.

As for DDS alpha-blending, each variable-modulated normal array is derived from a Gaussian sampling array where the standard deviations of the bumps are uniform across the layer. The standard deviation must be different for each variable displayed with DDS bump-mapping, so the viewer can differentiate among variable layers. Unlike alpha-blending where both color and standard deviation are available to differentiate among layers, bump-mapping has only the standard deviation.

Displaying Bump-Mapped Layers

Once the normals are generated for each variable-modulated bump-map, the surface must be rendered for display. The visual appearance of a bump-map is highly dependent on the lighting model used to render the surface. If the light appears to come from above and to the side of the viewer, the bumps will be seen as raised, or protruding, from the surface. If the light appears to originate from below the viewer, then the bump-map is seen as indenting the surface.

The study of how we understand bump maps fits into the area of perception called *Shape from Shading*. Ramachandran [1988] found that depending on where the viewer mentally positions the light, in an image with both bumps and dimples that are defined exclusively by shading information, dimples can be seen as bumps and vise-versa. In one image he displayed two rows of discs that can be seen as either convex or concave. Once one row is seen as convex, the other is seen as concave. Ramachandran claims that it is very difficult to see both rows as simultaneously convex. He believes that this indicates that the brain cannot accept multiple contradictory light sources at a time. He also shows that the visual system tends to automatically assume the illumination comes from above. This assumption of light position intuitively makes sense, as that is the direction from which everything in the natural, outdoor, world is lit. Ramachandran also found that for an image with concave and convex-appearing shapes intermixed it is possible to mentally group all the convex elements together to form a cluster that is clearly separate from the background. This implies that depth, as elicited from shading, is a feature that allows perceptual grouping.

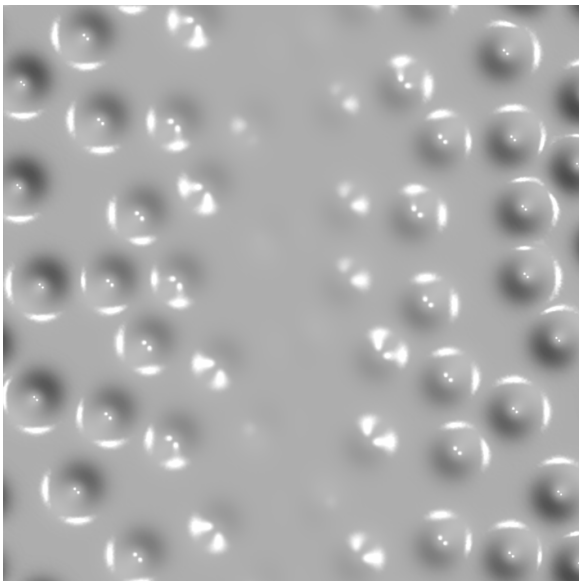

Figure 2.23: DDS bump-mapping showing a gradient from -1 to 1. Negative values appear as dimples that indent the surface.

Another consistent finding from the *Shape from Shading* literature is the pre-attentive pop-out of a concave dimple among convex bump distractors. The pre-attentive sensitivity is asymmetric: there is little to no-pop out for a convex bump among concave dimple distractors. Braun [1993] found that this pop-out is a parallel process that is independent of visual attention and the number of bumps in the image. Sun and Perona [1997] found that pop-out for convex over concave dominates shape cues from stereo information, and they claim that even though stereo information is important in early vision, like shape from shading, shape from stereo depends on our assumption of an overhead light source and that convex is more salient than concave. Convex bump maps may be more natural, being lit from above, and therefore easier for the viewer to understand.

DDS bump-mapping can display negative data values as dimples, which indent the surface, zero values as a flat surface, and positive values as bumps (Figure 2.23).

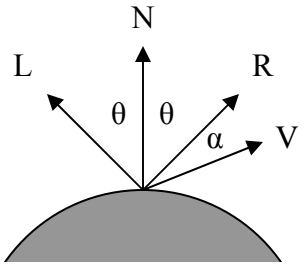

Figure 2.24: The light direction L is reflected about the normal vector N, producing the reflection vector R. DDS uses Phong illumination [1975], which models specular highlights as brightest when the view direction V is in the direction of R. The intensity of the highlights falls off with increasing angle between the two vectors, represented by α. Image based on [Foley, van Dam, Feiner, and Hughes, 1990].

Light Equation

All DDS bump-mapped layers are displayed with a lighting model that has two light sources, one from above the viewer and one to the right of the viewer. DDS uses the Phong illumination model [1975] to illuminate the bump-mapped surface. In the Phong model, the intensity of light, *I*, coming from a point on the surface is a function of how non-directional (ambient) light, and directional light reflect from the surface. Directional light is modeled with both diffuse and specular components. The diffuse reflection is a function of the angle between the directional light source and the normal vector. The specular reflection is a function of the angle between the directional light, reflected about the normal vector, and the view vector. The specular reflection is therefore view-dependent and will change with different viewing angles, whereas the diffuse reflection is view-independent and is constant for all view angles.

The diagram in Figure 2.24 illustrates the relationship between the direction to the light source vector *L*, the normal vector *N*, and the view direction vector *V*. In the diagram the vector *R* is the reflection of *L* about *N*. θ is the angle between vectors *L* and *N* and α (not to be confused with the α in alpha-blending) is the angle between the reflection vector *R*, and the view vector *V*.

Reflection coefficient	Object color (RGB)	Light Color (RGB)	Angle	Specular highlight
ka: ambient		I_a: ambient		
kd: diffuse	O_d: diffuse	I_p: point intensity	θ = angle between light and normal vector	
ks: specular	O_s: specular		α = angle between reflection vector and viewer	$n = 50$

Table 2.2: Parameters in the lighting equation from [Foley, van Dam, Feiner, and Hughes, 1990].

Surface material properties are also included in the light equation and interact with the illumination. Both diffuse and specular reflections are scaled by corresponding diffuse and specular reflection coefficients. In addition to the reflection coefficients, the equation models the color of the light and the diffuse and specular color components of the surface. The full light equation, taken from [Foley, van Dam, Feiner, and Hughes, 1990], is given below. Table 2.2 lists the parameters from the equation.

$$I = I_a k_a O_d + Ip[k_d O_d \cos(\theta) + k_s O_s \cos^n(\alpha)] \qquad 2.5$$

In the early implementation of DDS on the PixelFlow graphics machine [Eyles et al., 1997], every parameter in the lighting equation was under user interactive control, and each parameter could be used to display a different spatial variable. Figure 2.25 shows an example where two spatial variables from an SEM scan were used as input parameters into a bidirectional reflectance distribution function (BRDF) lighting model based on Cook and Torrance [1982].

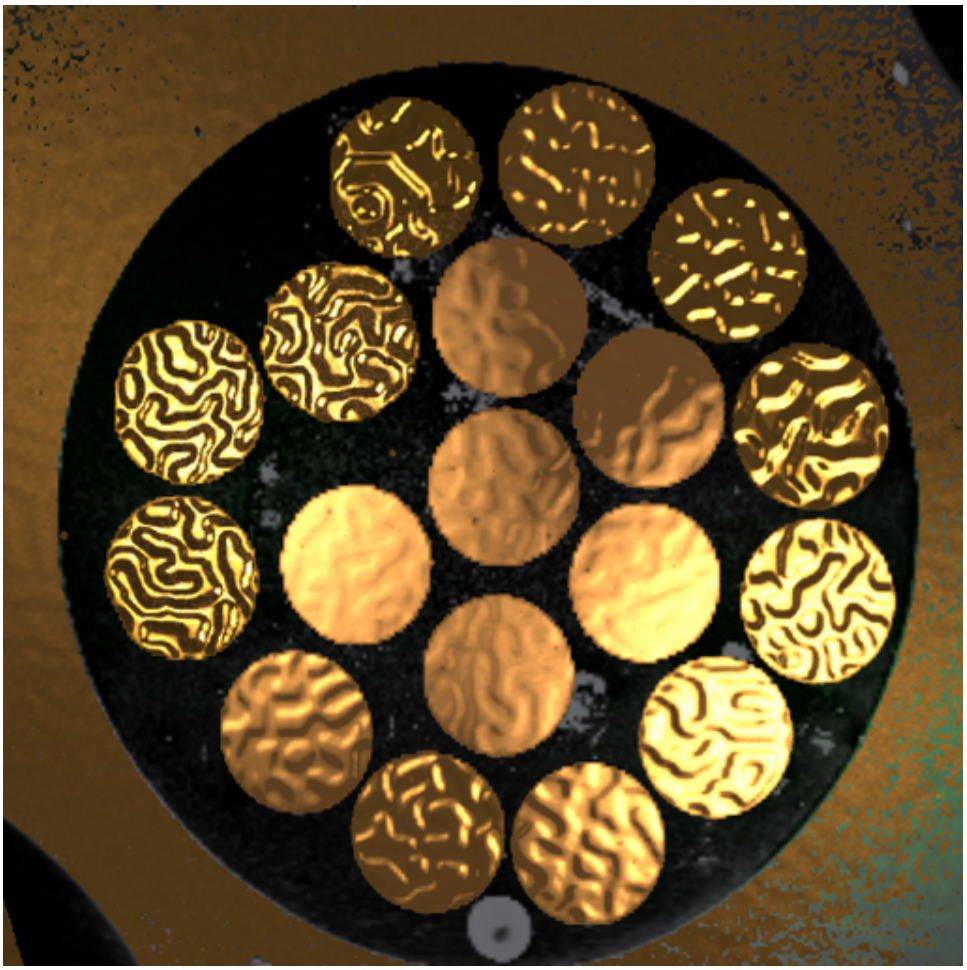

Figure 2.25: A BRDF visualization of a cross-section of a wire cable with 18 wires. The reflectance parameters of copper and gold are used to display the different amounts of each element present in the wires. The reaction-diffusion stripe height redundantly presents the amount of gold.

Figure 2.25 shows the cross section of a wire bundle with 18 wires, scanned on an SEM microscope. The data contains information about the relative amounts of gold and copper in the wires – this information is displayed with a BRDF shading model of copper and gold reflectance functions. The circular-cross sections of the wires in the center of the image look more like copper and the cross sections on the outer ring look more like gold. Black regions show the space between wires. Also displayed is a reaction-diffusion stripe pattern, which duplicates the gold data values and appears taller in areas where gold dominates (outer wires) and shallower where copper dominates (center wires). In the image the light source is positioned directly above the surface, instead of positioned at infinity. This causes the appearance of color gradient within the wires. The gradient is due to the lighting model, not the underlying data values, and in later work we changed the lighting model, placing the light at infinity, to prevent this effect.

Applying variable data to the input parameters of the light equation is similar in result to applying the variable data to the input parameters of a texture-generation technique, such as the reaction-diffusion or spot noise techniques, discussed earlier in this Chapter. Both techniques produce areas in the image that are visibly distinct, yet neither technique allow the viewer to visually discriminate, or parse, the variables on an individual basis. Again, this is similar in result to the glyph techniques, which I describe more in Chapter Four.

Shape Analysis of Bump-Mapped Layers

I wanted to understand how well bump-mapping displays data, specifically boundary information. I used a shape analysis technique that analyzes changes in the normal vectors across the surface. The shape analysis of DDS bump-mapping was designed to answer three questions: first, are sharp changes in the normal vectors, such that occur along spatial boundaries in the data, visible to the shape analysis? Second, are the boundaries still visible when multiple layers of bumps are present, where one bump-mapped boundary overlaps another. Third, how are levels in the data seen through the shape analysis?

Two metrics are used in the shape analysis of surfaces in R^3, the shape index S and the curvedness measure C; both are described in Koenderink's book, *Solid Shape* [1990]. The shape index S is a measure of the type of shape, and ranges from –1 to + 1, from concave in two-dimensions (holds water), through saddle-shaped to convex in two-dimensions (does not hold water). The C metric is a measure of how much shape – it measures the curvedness and ranges from -infinity to +infinity, from flat (no curvature) to needle-sharp, singular, points (high curvature).

Koenderink describes the shape index as a measure of *which* shape and the curvedness is a measure of *how much* shape; he likens the shape index and curvedness to hue and brightness, in that shape, like hue, categorizes and curvedness is similar to levels of intensity. The shape index is described as a qualitative measure of surface characteristics, whereas curvedness is a quantitative measure.

What aspects of a 3D object are important in determining how we perceive the surface – what defines shape perceptually? If we perceive shapes of 3D objects based on metrics similar or somehow equivalent to the *S* and *C* metrics, and if the *S* and *C* metrics can find boundary information in a data-modulated bump map, then perhaps we can see the boundary information in a bump-map display as well.

S and C Shape Metrics

Whether a point on a surface is 2D concave, saddle, or 2D convex is determined by the values of the two principal curvatures at that point. The two principal directions of a surface are 1) the direction where the surface has the highest positive curvature and 2) the direction in which the surface has the highest negative curvature. The body of a cylinder provides a good explanation for the two principal curvatures: for an ant walking in the direction running down the length of the cylinder, the surface appears flat and the curvature is zero, whereas in the direction around the circumference of the cylinder, the curvature is greatest. These two directions are the principal directions for the cylinder. For smooth surfaces the two principal directions are perpendicular from one another.

To calculate both *S* and *C*, the two principal curvatures are calculated first. Both principal curvatures are calculated by looking at how the normal changes for a given direction of change on the surface – this direction is referred to as the *walking* direction. Whether the normal falls forward along a particular walking direction, or whether it falls backward gives a measure of whether the surface is convex of concave at that point.

The amount of tip in the normal for a given walking direction from point A to point B is calculated by subtracting the normal at B from the normal at A. The dot product of the difference vector and the walking direction is calculated, giving a measure of the change in normal in the walking direction. Figure 2.26 shows an illustrative diagram of this process. The principal curvatures at the point A are the maximum and minimum values produced over every walking direction from point A.

Although more efficient methods for calculating the principal curvatures exist, they can be used only if the surface is smoothly varying at the point of calculation. Because DDS bump-mapping modulates the normal with the data variable, which may be very noisy, the data-modulated normal vectors may not be smoothly varying.

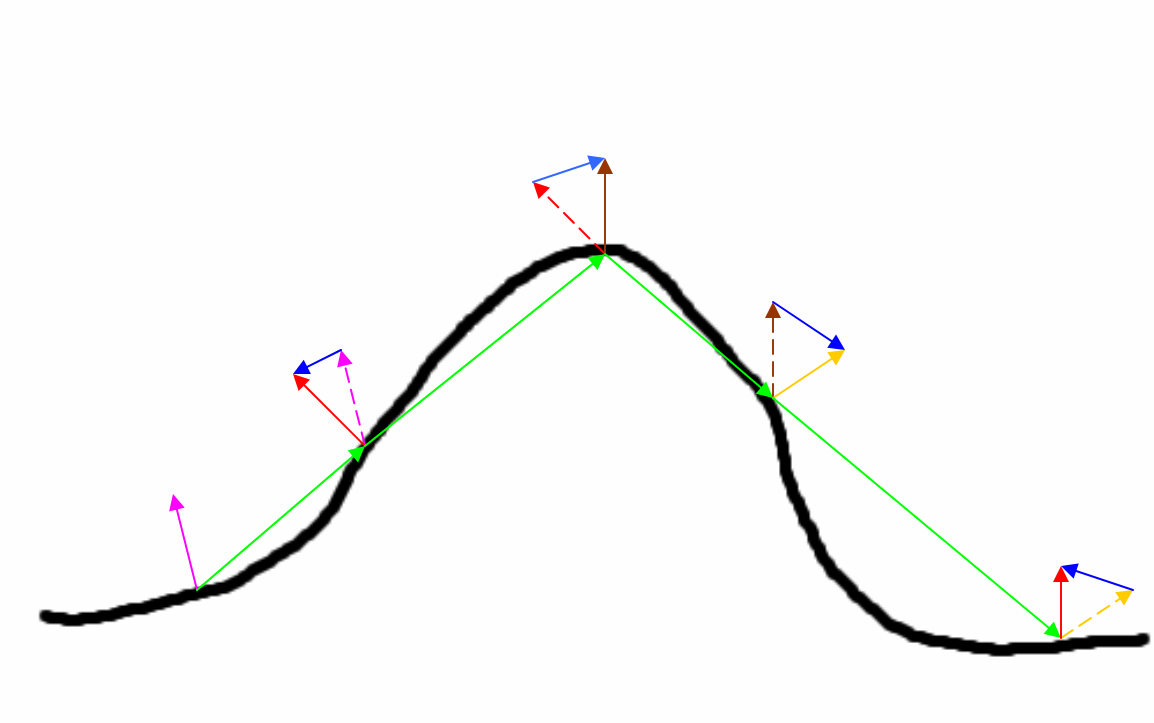

Figure 2.26: This diagram illustrates how the change in the normal along a given walking direction is calculated. The walking-direction moves from left to right along the curved line, which represents a slice of a surface. Starting with the purple normal vector, the walking direction moves up-slope to the red vector. The difference between the old normal (translated to the new point and displayed with a dashed line) and the new normal is shown as a blue vector. The dot product of the difference vector and the walking direction is calculated, giving a negative number – this reflects the fact that the normal fell 'backward' when walking upslope. The example continues, from the red normal vector to the brown, from brown to yellow and from yellow to red and the right-most point on the surface.

Once the principal curvatures, κ_1 and κ_2, are found, S and C are calculated for each point on the surface using the following equations:

$$S = (2.0 / \pi) * arctangent((\kappa_1 + \kappa_2) / (\kappa_1 - \kappa_2)) \qquad -1 <= S <= 1 \qquad 2.6$$

$$C = (2.0 / \pi) * log (sqrt((\kappa_1^2 + \kappa_2^2) / 2.0)) \qquad -\infty <= C <= +\infty \qquad 2.7$$

Figure 2.27 shows images representing S and C metrics calculated for a single Gaussian bump. Two color maps are used in Figure 2.27 as well as in other examples in this section. The first color scale, which is used for the S metric, was chosen to show both positive and negative values, with a clear zero-point. The color map shows convex regions as cyan, saddle regions as gray, points of minimal-saddle (flat) as black, and concave areas as purple. As there are not concave regions on a single Gaussian, there is no purple in the images.

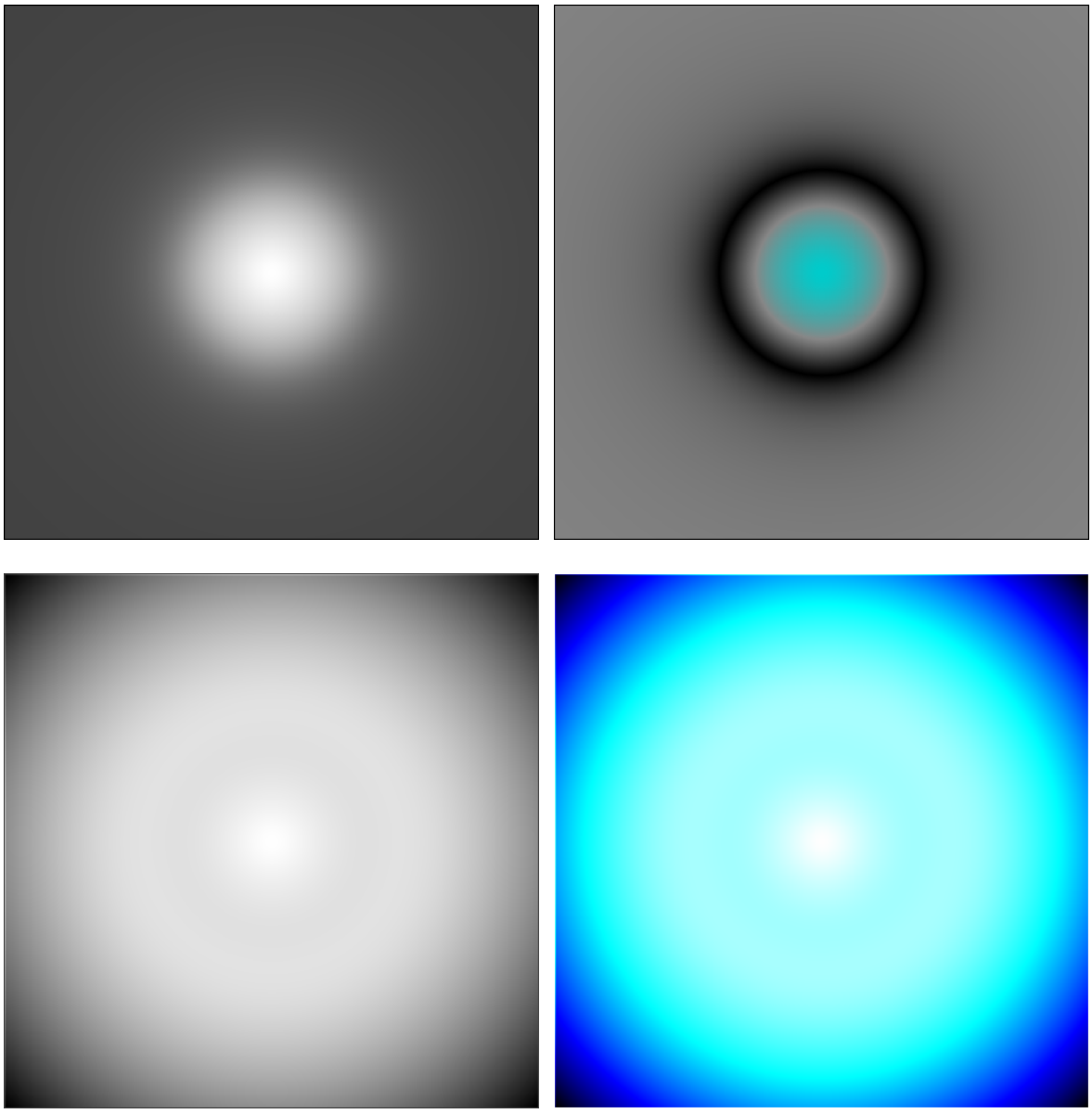

Figure 2.27: The *S* and *C* metrics for a single Gaussian spot. Images in the top row show the *S* metric, shown on the left in grayscale and on the right with a color map that indicates convex with cyan, hyperbolic regions with gray, minimal saddle regions with black and concave regions with purple (although there are no concave regions with one Gaussian). Images in the second row show the *C* metric, with grayscale image on the left and a heated body color map that transitions from black to blue to cyan and to white. High values of *C* map are white and low values are black, and cyan though blue indicates higher to lower values.

The second color map is used to display the *C* metric, which ranges in value from completely flat (zero curvature) to highly curved singular points. The color transitions are based on a blackbody color scale where flat points are black and as the curvature increases the color transitions from black through blue to cyan and finally to white.

Figure 2.28 shows two different fields of Gaussian bumps at different sizes with the corresponding *S* and *C* metrics. A single Gaussian bump has no concave regions. It is only by adding multiple Gaussians together that concave regions occur between the bumps – this is clearly shown in the image of the *S* metric. The image of the *C* metric shows that the concave areas are relatively flat compared to the tops of the Gaussians. Both images correspond well to how we interpret the heights of the Gaussian bumps in the top row images.

Two overlapping fields of bumps are shown in Figure 2.29. The image at the top is the rendered bump map; the images on the bottom are the *S* and *C* metrics. The *S* metric shows the pattern of large and small spots, clearly. The *C* metric does not distinguish well between the different spots on the other hand.

A DDS bump-map of the SEM backscattering variable is shown in Figure 2.30, with the corresponding *S* metric, shown with a gray scale color map, presented in Figure 2.31. A gray scale color map was chosen for this example because it highlighted the boundary information detected by the *S* metric best. Notice how the boundary in the SEM data from high to low values is clearly outlined in white in the images of the *S* metric.

When data is used to modulate the apparent height of a bump map, different values in the data can be seen. The SEM data has sharp boundaries where on one side of the boundary the data is high and on the other side the data is low. Sharp boundaries in the data cause the bump map to rapidly change from highly-curvy to flat and from convex to minimal-saddle, so we would expect both shape metrics to show boundary information in the data. However the results are best with the *S* metric for the example shown.

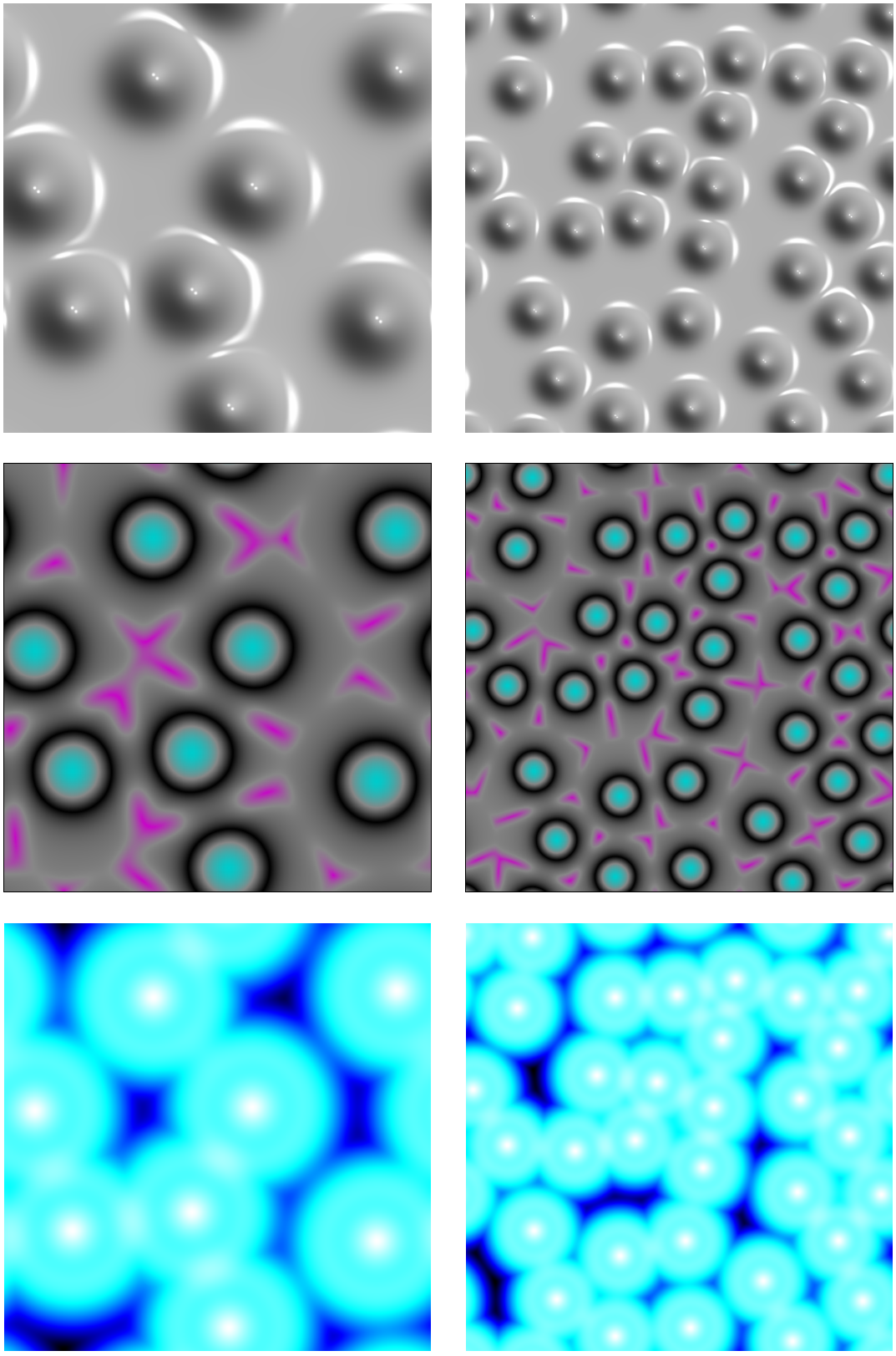

Figure 2.28: Two fields of Gaussian bumps with the corresponding *S* and *C* metrics, shown in columns. Notice how the presence of overlapping Gaussians creates concave regions.

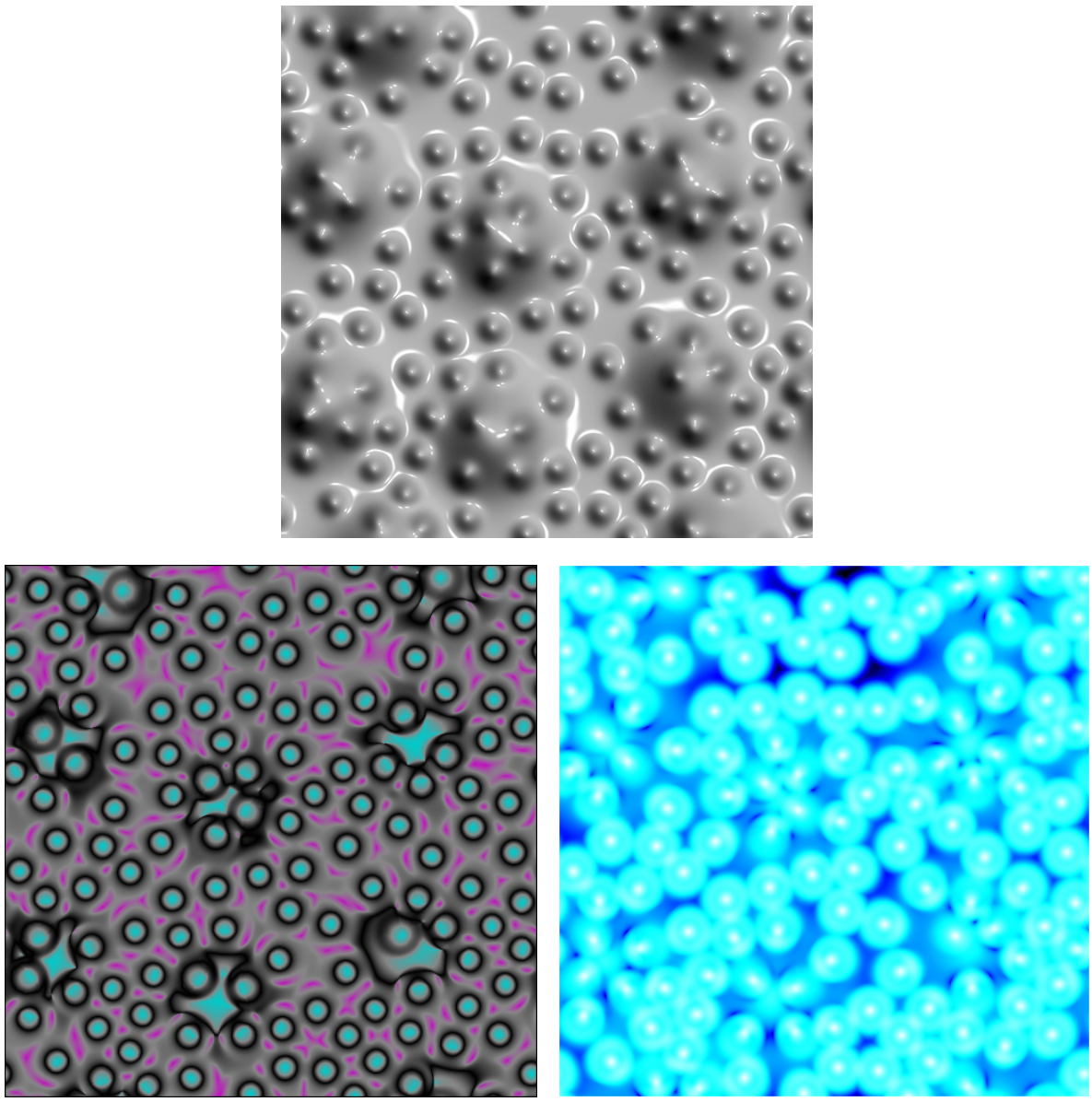

Figure 2.29: A surface with two different size bump maps, with the shape index *S* shown on the lower left and the *C* metric on the lower right.

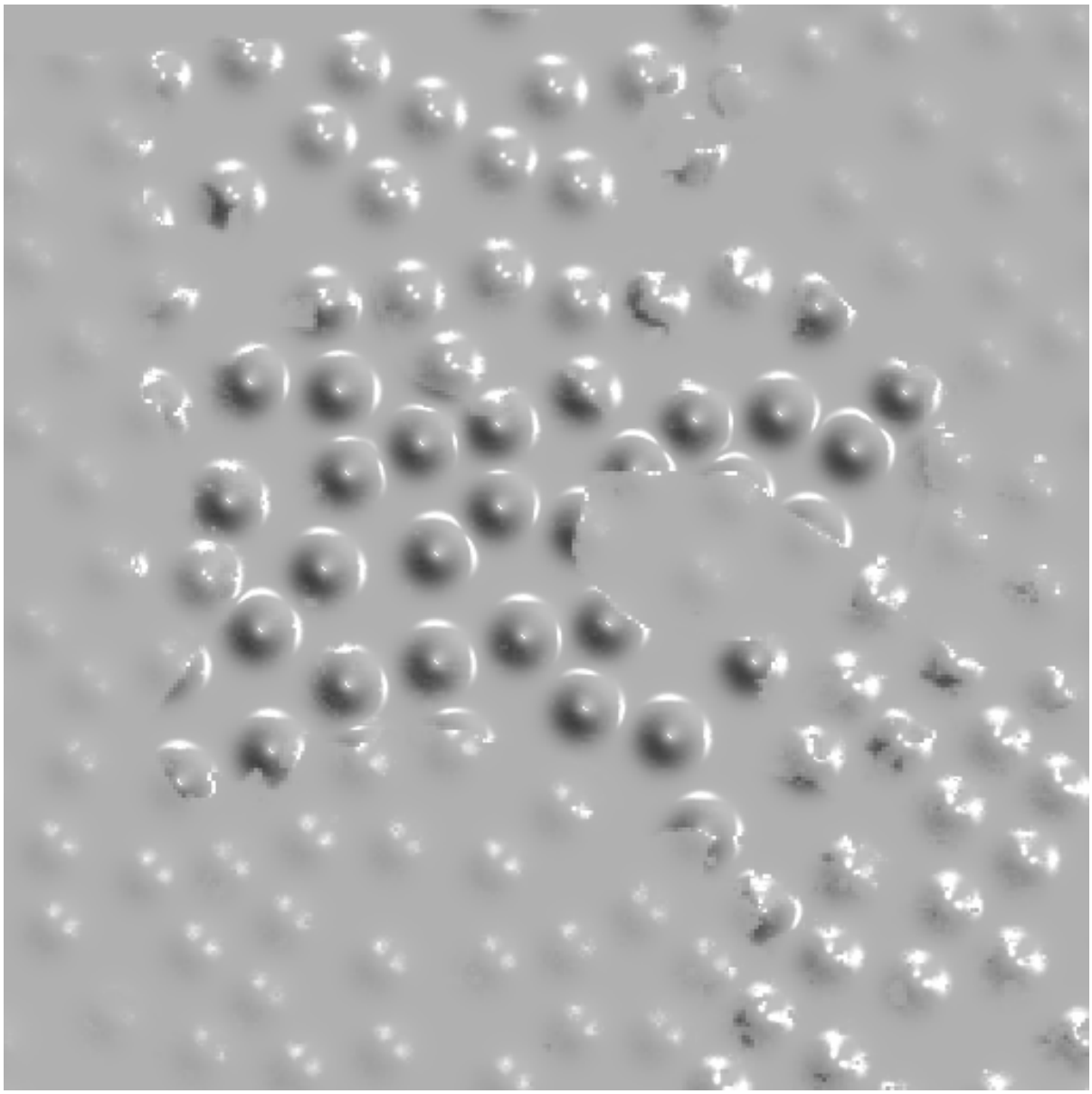

Figure 2.30: The SEM back scattering data layer shown with DDS bump-mapping. Different levels in the data create visibly distinct regions; at least five different values are visible in the height of the bumps.

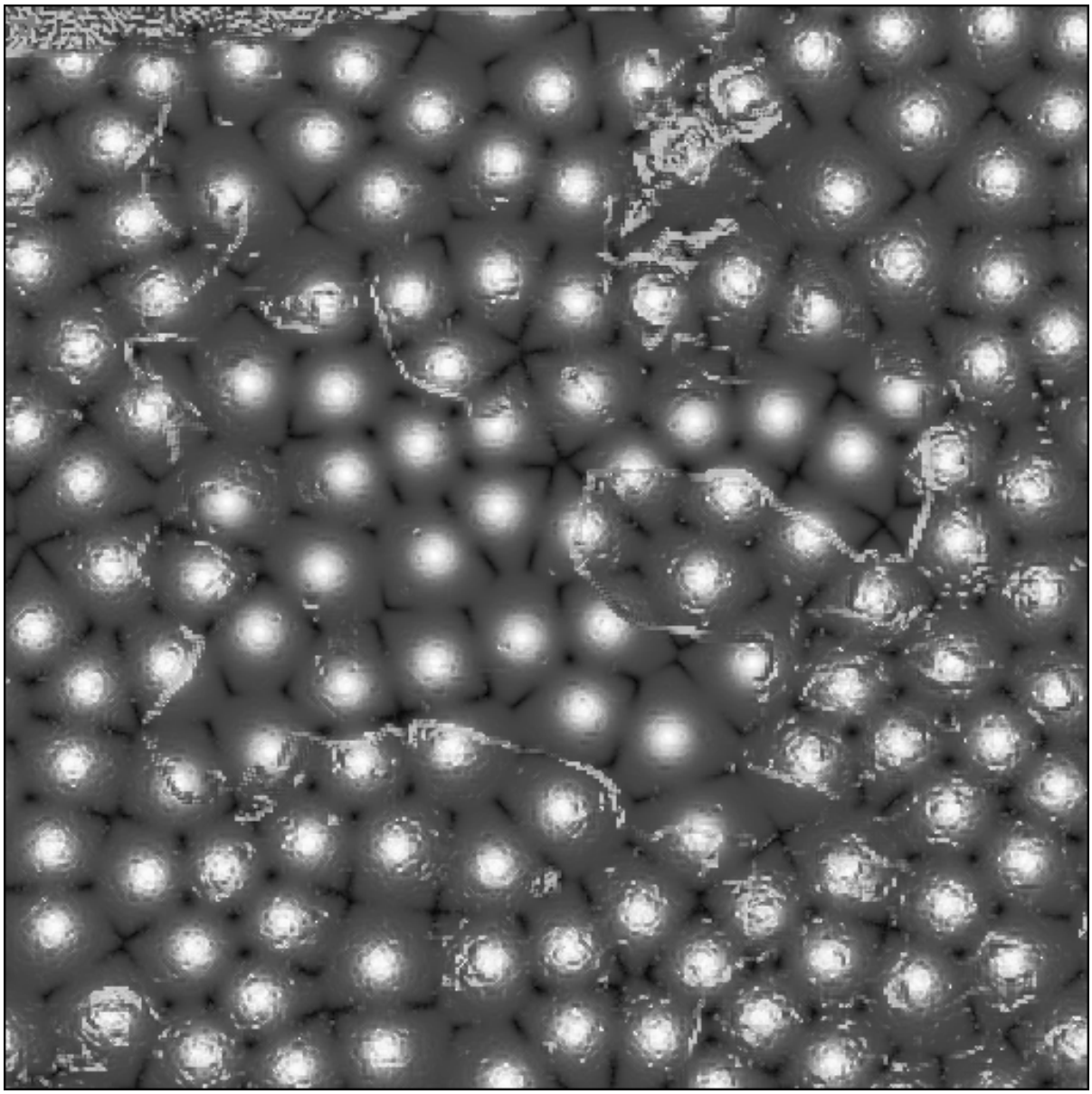

Figure 2.31: The image is a grayscale image of the *S* metric calculated on the data-modulated bump map in Figure 2.30. The boundaries in the data are visibly outlined in white. Value information is lost as both tall and shallow bumps are convex and show as white regions in the *S* image.