Surface shape and curvature scales

Jan J Koenderink and Andrea J van Doorn

The classical surface curvature measures, such as the Gaussian *and the* mean curvature *at a point of a surface, are not very indicative of local shape. The two principal curvatures (taken as a pair) are more informative, but one would prefer a* single *shape indicator rather than a* pair *of numbers. Moreover, the shape indicator should preferably be independent of the* size *i.e. the* amount *of curvature, as distinct from the* type *of curvature. We propose two novel measures of local shape, the 'curvedness' and the 'shape index'. The curvedness is a positive number that specifies the amount of curvature, whereas the shape index is a number in the range [−1, +1] and is* scale invariant. *The shape index captures the intuitive notion of 'local shape' particularly well. The shape index can be mapped upon an intuitively natural colour scale. Two complementary shapes (like stamp and mould) map to complementary hues. The symmetrical saddle (which is very special because it is self-complementary) maps to white. When a surface is tinted according to this colour scheme, this induces an immediate perceptual segmentation of convex, concave, and hyperbolic areas. We propose it as a useful tool in graphics representation of 3D shape.*

Keywords: surface shape, curvature scales

CLASSICAL SHAPE MEASURES

Throughout this paper we deal exclusively with the curvature at a point of an (arbitrary) smooth surface. 'Smooth' means at least twice continuous differentiability in this context*.

We assume that the reader is familiar with the classical second order description of a smooth surface patch. The basic insights date from the 18th century, and are summarized elsewhere[1–5].

Utrecht Biophysics Research Institute (UBI) Buys Ballot Laboratory, PO Box 80 000, NL-3508 TA Utrecht, The Netherlands

*If the surface is given as a discrete datastructure (a triangulation, say), the theory may still apply to the smooth approximations locally fitted to the triangulation, but you will have to provide the fitting procedure: it depends on your initial data structure.

Paper received: 10 March 1991; revised paper received: 13 November 1991

When you set out to describe *shape*, the first thing to do is to factor out *position* and *attitude*. This can be done by picking an apt reference frame. For instance, if you want to describe the shape at some fiducial point on a surface, then you would put the origin of your coordinate system at the fiducial point. By referring the surface to the local coordinate system you have effectively freed yourself of the (coincidental) *position* of the object. In a similar way, you may try to find an attitude of your coordinate frame that frees you of the incidental attitude of the object. Such a coordinate frame then has to be *adapted* somehow to the shape of the surface. By having the z-axis of your (Cartesian) coordinate system point along the outward normal of the surface you get rid of two degrees of (irrelevant!) freedom. You are left with a single ambiguity: the frame can still be rotated about the normal. This remaining ambiguity can be removed by closer investigation of the local surface. You end up with the following construction: at any point of the surface you may fit an (essentially unique, *v, i.*) *principal frame*, i.e. a triad of orthonormal vectors such that the initial part of the Taylor series describing the deviations from the tangent plane is of the form:

$$z = \frac{1}{2}(\kappa_1 x^2 + \kappa_2 y^2) + O^3(x, y)$$

where the coordinates (x, y, z) are measured along the axes of the principal frame. To be exact, the fiducial surface point is taken as the origin. The coordinates (x, y) are measured along the frame vectors that together span the tangent plane, whereas the z coordinate is measured along the frame vector which is normal to the tangent plane. This is a so called 'Monge-representation' of the surface patch.

Several conventions are helpful in avoiding confusions:

- of the two possible choices of the orientation of the normal, we pick the *outward normal*;
- of the two possible orientations of the frame as a whole, we take the *right hand* one;
- for the larger part of this paper we retain the convention that $\kappa_1 \geqslant \kappa_2$.

0262–8856/92/008557–08

No generality has been sacrificed through the introduction of these constraints on the description. As a result, there are still two possible choices of the frame left, for you may still invert both of the two tangential frame vectors simultaneously. Both choices lead to the same second order terms, but differences appear in the third and higher odd order terms. (These may be used to remove the remaining ambiguity if you want.) As far as the second order is concerned, you have arrived at an intrinsic description, which is quite insensitive to the incidents of position and attitude of the object. The remaining second order structure is a geometric invariant that may be regarded as a formal definition of what is meant by the habitus, or 'local shape'. It is not completely general in the sense that we have decided on a *metrical* description, and stop at the second order. (In a way, that is our definition of 'local'.) However, this certainly covers the overwhelming majority of applications.

The scheme works at *almost every* point of a generic surface: only at certain *isolated points*, the so called *umbilical* points, do you encounter a degenerated situation. At an umbilicus you meet the condition $\kappa_1 = \kappa_2$, thus *any* pair of orthonormal vectors spanning the tangent plane may be substituted for the pair of tangential frame vectors. In the overwhelming majority of cases ('probability one') the so called *principal directions*, which are the tangential frame vectors, are uniquely defined. So called 'flat points', characterized through the condition $\kappa_2 = \kappa_2 = 0$, generically do not occur. So called 'parabolic points', characterized by the vanishing of *one* principal curvature, occur generically on *curves*, the so called 'parabolic lines'. Such curves are smooth loops that never intersect. The parabolic curves strictly separate regions of convexities $\kappa_{1,2} < 0$, concavities $(\kappa_{1,2} > 0)$, and saddle-like points (κ_1 and κ_2 of different sign). Saddle–like shapes are also known as 'hyperbolic', whereas both concave and convex shapes are also known as 'elliptic'. (Such facts can be gleaned from textbooks on the classical differential geometry of surfaces in 3-spaces[5,6].)

The coefficients $\kappa_{1,2}$ are known as the *principal curvatures*. They are the decisive parameters that fully describe local surface shape up to the second order *modulo* a rigid movement.

The classical shape measures[3,5] are:

- the Gaussian curvature, which numerically equals the *product* of the principal curvatures. It also equals the determinant of the Hessian at the origin of $z(x, y)$ in our representation: this reveals its invariant nature. It thus equals the square of the *geometrical* average of the principal curvatures;
- the mean curvature, which numerically equals the *arithmetic* average of the two principal curvatures. It also equals the trace of the Hessian at the origin, which again reveals the invariant nature of this definition.

It is perhaps not superfluous to remark here that the simple interpretation in terms of the Hessian (i.e. the symmetrical matrix of mixed second order partial derivatives of $z(x, y)$ with respect to x and y) is *only* valid in representations where the magnitude of the gradient of z vanishes. (That is $(\partial z/\partial x)^2 + (\partial z/\partial y)^2 = 0$.) Otherwise, the expressions for the Gaussian and mean curvatures are more complicated, and contain both first and second partial derivatives of $z(x, y)$ with respect to x and y.

This second order structure suffices to determine the curvature of *arbitrary planar sections* of the surface through the fiducial point. This involves Meusnier's and Euler's theorems[1,4,5]. Euler's theorem expresses the curvature of *normal sections* in terms of the principal curvatures and the angle of the section with respect to the principal directions. Meusnier's theorem expresses the curvature of *skew sections* in terms of the obliquety.

These measures possess a coordinate independent *geometrical* meaning. Of the many possible interpretations we merely mention that the Gaussian curvature (**K**) specifies the 'spread of normals', i.e. the (oriented!) solid angle filled by all normals at an infinitesimal patch divided by the surface area of the patch. Likewise, the mean curvature (**H**) measures the spread of normals for the points of infinitesimal arcs, divided by the arc length, and averaged over all surface directions. These geometrical interpretations enable you to identify the Gaussian curvature as the continuous equivalent of the *vertex curvature* of a polyhedral vertex[2], whereas the mean curvature, likewise, is the continuous equivalent of the (average) polyhedral *edge curvature*.

The Gaussian curvature has a special property that accounts for the fact that most of the mathematical literature is almost completely dedicated to closer study of this curvature measure: it can be defined through metrical operations that are completely confined to the surface itself[2]. This remarkable fact thus says that the Gaussian curvature is *intrinsic* to the surface, and does not depend on the embedding in 3D space. For instance, bending a surface without stretching conserves the Gaussian curvature, although this action definitely changes the 'shape' in the commonsense of the word. In this paper, any operation other than an isometry or similarity (change in size) is considered to destroy the shape. We will not be concerned with such 'deformations'. This removes the special significance of the Gaussian curvature.

Neither the Gaussian nor the mean curvature, by themselves, capture the intuitive notion of 'local shape' very well. You need *both* to be able to construct the initial second order part of the Taylor series. If you want an idea of what the shape is, the best option is to compute the principal curvatures from the Gaussian and mean curvatures, and take it from there. (Clearly, you have $\kappa_{1,2}^2 - 2\mathbf{H}\kappa_{1,2} + \mathbf{K} = 0$.)

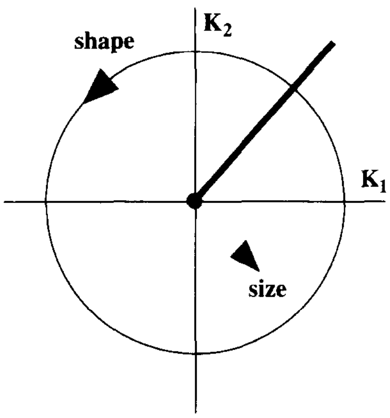
The classical curvature measures are very convenient if you happen to be involved in work of a mathematical nature (e.g. in intrinsic differential geometry, in the investigation of global properties of manifolds, etc.). They are less immediately useful in the case of the parametrization of *extrinsic shape*, however.

CURVEDNESS AND SHAPE INDEX

Most people apparently accept the following statement as a self-evident truth:

> *All spheres are of the same* shape *although they may differ in size*

Figure 1. (κ_1, κ_2)-plane. On a half ray at the origin you find shapes that are scaled copies *of each other. Such similar surfaces are said to be of the same shape. Then 'shape' can be measured as a direction of a halfray in the (κ_1, κ_2)-plane. Every point is on a unique halfray, thus can be ascribed a unique shape,* except *the origin. But the origin represents the flat points, which – indeed – are of indeterminate shape. The distance from the origin merely reflects a size change*



Such a usage of the word 'shape' implies a *scaling invariance*: intuitively, all local approximations for which the *ratio* of the principal curvatures is equal are of the same shape.

This intuitive notion of shape is apparently based on a *polar coordinate system* in the (κ_1, κ_2) parameter plane. The direction encodes the shape, whereas the distance from the origin encodes the size[7]. To exploit this idea we temporarily drop the conventional ordering $\kappa_1 \geqslant \kappa_2$ and consider all possible second order forms.

The following properties of the polar representation are especially noteworthy:

- the origin represents the uncurved patch, i.e. the *planar point* (see Figure 1). At a planar point the direction (shape) is indeterminate, as it should be: various infinitesimal perturbations may lead to *qualitatively* different shapes, thus 'the shape' at a planar point is really *indeterminate*;
- all points of a half-ray from the origin represent the *same* shape, although they differ in *size* (see Figure 1);
- diametrically opposite points at an equal distance from the origin represent surfaces that fit together snugly, as a 'stamp' and its corresponding 'mould', thus the second order shapes are also 'opposite', they are called 'complementary' in this paper (see Figure 2);
- points at an equal distance from the origin, related through a reflection at the axis $\kappa_1 = \kappa_2$, are

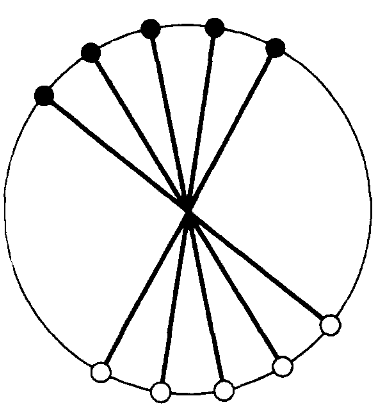
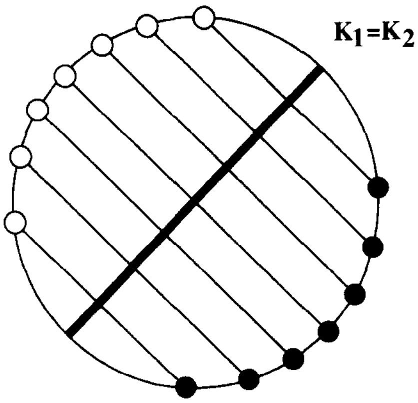


Figure 2. Antipodal points on a circle with centre at the origin are related in a special way: they are 'complementary' i.e. related as a mould and its stamp. They can be fitted into each other

Figure 3. Two points on a circle with centre at the origin and related by a reflection on the $\kappa_1 - \kappa_2 = 0$ axis are congruent. *A right angle turn about the z-axis will bring them in superposition. Notice that shapes* (a, $-$a) *and* ($-$a, a) *are both complementary and congruent. You may superimpose them in two different ways: flipping the height (reflection in* xy*-plane) or a right angle turn about the z-axis*



congruent. You may superimpose them after a right angle turn about the z-axis (see Figure 3). Since all points related through a permutation of κ_1 and κ_2 represent equal shapes, you need only consider *half* of the directions in the parameter plane. That is why we previously introduced the convention $\kappa_1 \geqslant \kappa_2$;
- the direction $\kappa_1 = -\kappa_2$ is special in the sense that points on this line at an equal distance from the origin are simultaneously *complementary* and *congruent*. These shapes 'are their own moulds', thus they do not come in distinct positive/negative pairs as all other shapes do. Inverting the surface ($z(x, y) \rightarrow -z(x, y)$) is (for the second order) equivalent to a ninety degree turn about the z-axis;
- if you disregard the fact that the surface is an oriented interface, dividing the material from the enveloping air, you may *identify* stamps and moulds. This is the usual course taken by authors of treatises on differential geometry. However, it is immediately obvious that this makes very little sense in most transactions with the material world, since a cup is for most practical purposes quite different from a cap, e.g. only the cup will hold tea.

From these considerations it follows that you should not parametrize 'shape' by the (two-sided) *rays* of the (κ_1, κ_2)-plane. This would provide the shape space with the topology of $\mathbf{P}^1$, that is the projective line. In that case you would not distinguish inside from outside. Nor are the half-rays at the origin to be used. These would lead to a topology as that of the unit circle $\mathbf{S}^1$. In that case the same shape would appear twice in shape space. Rather, the space of shapes has the topology of $\mathbf{D}^1$, that is the 1-dimensional disc, or 'linear segment' in this case. This is most clearly brought out through the following definition of a *shape index*.

$$s = \frac{2}{\pi} \arctan \frac{\kappa_2 + \kappa_1}{\kappa_2 - \kappa_1} \qquad (\kappa_1 \geqslant \kappa_2)$$

All patches, except for the planar patch which has an indeterminate shape index as should be, map on the segment $s \in [-1, +1]$. This representation has many intuitively 'natural' properties:

- two shapes for which the shape index differs merely by sign represent complementary pairs that will fit together as 'stamp' and 'mould' when suitably scaled;
- the shape for which the shape index vanishes – and consequently has indeterminate sign – represents the objects which are *congruent to their own moulds*;
- convexities and concavities find their places on opposite sides of the shape scale. These basic shapes are separated by those shapes which are neither convex nor concave, that are the saddle-like objects. The transitional shapes that divide the convexities/concavities from the saddle-shapes are the cylindrical ridge and the cylindrical rut;
- if shapes are drawn at random from an isotropic distribution in the (κ_1, κ_2)-plane, then the shape index scale will be uniformly covered, thus the scale is 'well tempered'. This follows immediately from the fact that the shape index s is directly proportional to the *angle* with the $\kappa_1 + \kappa_2 = 0$-axis.

For later reference, we also list here the following:

- the extreme ends of the scale represent the 'umbilical points'. These look locally like either the outside ($s = 1$) or the inside ($s = -1$) of a spherical surface (a cap or cup). Notice that arbitrary small perturbations of these shapes invariably have the effect of moving the shape towards the cylindrical. Hence $s = \pm 1$ are true *endpoints* of the shape scale;
- the cylindrical shapes are at $s = 0.5$ (ridge) and $s = -0.5$ (rut). For $-0.5 < s < 0$ you have saddle-ruts, for $0 < s < 0.5$ you have saddle-ridges. For $0.5 < |s| < 1$ the shapes are ellipsoidal, tending towards the spherical when $|s|$ approaches unity, to the cylindrical when $|s|$ approaches 0.5.
- the range $-1 < s < -0.5$ represents the concavities (concave ruts, or trough-shapes), the range $0.5 < s < 1$ represents the convexities (convex ridges, or dome-shapes), whereas the range $-0.5 < s < 0.5$ represents the saddle-like shapes (saddle ruts and ridges, with the symmetrical saddle at $s = 0$);
- generically, the umbilics ($|s| = 1$) occur only at *isolated points* on the surface. The parabolic points ($|s| = 0.5$) occur on *curves*, of which there are two distinct types ($s = \pm 0.5$). These curves are smooth, closed loops on compact, smooth surfaces. They can be nested and juxtaposed, but (generically) never intersect. The symmetrical saddles ($s = 0$) also generically occur on curves (they are like the 'zero crossings of the Laplacian' so popular in image processing). The dome and trough shapes occur over areas that never touch each other. They are necessarily separated by saddle-like areas.

In addition to the shape index, it makes sense to define an always positive number c, henceforth referred to as the 'curvedness', to specify the amount, or 'intensity' of the surface curvature. Although several alternative definitions would serve, we define the curvedness as the distance from the origin in the (κ_1, κ_2)-plane. We propose:

$$c = \sqrt{\frac{\kappa_1^2 + \kappa_2^2}{2}}$$

The scaling is such that the curvedness equals the absolute value of the reciprocal radius in the case of spheres. This makes it easy to develop an intuitive appreciation for the magnitude of the curvedness. Unit curvedness is ascribed to the unit sphere and unit saddle ($|\kappa_{1,2}| = 1$), whereas the unit cylinder ($\kappa_1 = 1$, $\kappa_2 = 0$) has only curvedness $1/\sqrt{2}$. The curvedness is inversely proportional with the *size* of the object.

Whereas the shape index scale is quite independent of the choice of a unit of length, the curvedness scale is not. Curvedness has the dimension of reciprocal length. In practice one has to point out some fiducial sphere as the 'unit sphere' to fix the curvedness scale.

The curvedness has some obvious properties that makes it a desirable curvature measure in an intuitive sense:

- the curvedness vanishes *only* at the planar points. Recall that both the Gaussian and mean curvature also vanish at the planar points. The Gaussian curvature vanishes also on the *parabolic curves*, however, whereas the mean curvature also vanishes on the loci where the surface is locally *minimal* (that is, $\kappa_1 = -\kappa_2$). But both at the parabolic points, and at the minimal points, the surface has a decidedly 'curved look' to most observers. This tends to be a perennial source of confusion to pupils of differential geometry classes. For a generic surface the curvedness vanishes *nowhere*.
- the curvedness scales inversely with size;
- the curvedness is trivially coordinate independent, i.e. it has true geometrical significance.

Taken as a pair, the curvedness and the shape index specify the local second order geometry, except for rigid motions, or congruences. Considered in isolation, the shape index does specify the local second order geometry up to a scaling factor, or similarity.

In some cases the logarithm of the ratio of the curvedness to the curvedness of the fiducial sphere will be the most convenient parameter, especially if the range of curvednesses encountered in the application is very large. Intuitively, the curvedness *dissimilarity* between a sphere and another sphere of double the diameter of the first one is independent of the absolute size. Thus a logarithmic transformation is quite natural. The shape index and this logarithmic measure may be considered as parameters of a *conformal image* of the polar coordinates in the curvatures-plane. Thus relations between similar shapes of similar sizes are as in the $\kappa_1\kappa_2$-plane. We do not use such a representation in this paper, however. The discussion will mainly focus on the shape index.

From our own psychophysical studies in shape recognition and discrimination (using such various cues as shading, motion parallax or binocular disparity) we find that a measure like the 'shape index' is much better suited than the classical shape measures if the intention is to pinpoint visually meaningful local shape features. Even naive observers easily pick out shape-index extrema and are able to score broad classes of local shapes on this basis. Broad categories that can be used almost immediately are: the *trough*, the *saddle rut*, and *ridge*, and the *dome* shapes; where you take the *spherical cup*, the *cylindrical rut* and *ridge*, and the *spherical cap* as the 'anchor-points' of the scale.

COLOUR SCALE FOR SHAPE CATEGORIES

In many cases in which a fast and certain judgement of local shape is important, one would like to be able to 'colour' the surface (on a graphics display) with the local shape. In our applications (medical image processing and presentation) one is, for instance, interested in deviations from the normal surface shape of the left cardiac ventricle wall, even as a function of time. If the observer has to rely on 'shape from shading', etc. in a photographic realistic rendering of the object, this tends to be a very hard task. However, if one 'paints' the surface with the exact local shape the task is much alleviated and, in fact, almost trivial. In this section we propose a design for such a natural colour scale.

One would like to have the following properties realized in such a scale:

- the three global classes of convexities, concavities and saddle shapes should perceptually segregate immediately ('preattentively'). Thus they should be assigned hues that naturally group internally, whereas the groups should bear little affinity to each other;
- easily recognizable and nameable shapes should be assigned easily recognizable and nameable[8,9] hues. The spontaneously nameable shapes are cup, cap, ridge, rut and saddle. The spontaneously nameable hues are red, green, blue, yellow and white;
- the extremes of the scale (convex and concave umbilicals) should stand out clearly;
- complementary shapes (shape indices differ only in sign) should be represented by complementary hues;
- the shape index scale should map on a continuous curve in colour space (local neighbourhood relations are to be preserved).

These requirements are rather restrictive. The first and last requirements might even seem to be incompatible. However, colour space has a natural structure[10] that allows one to meet all requirements. There is only little leeway in the construction of a colour scale. Here is one way of reasoning that leads to an embedding of shape space in colour space that meets all requirements: we start by noting that the fourth requirement fully determines the hue to be assigned to the symmetrical saddle. Since this shape has no complementary it should be assigned an achromatic colour (white or grey).

Because the elliptic and hyperbolic shapes are both mathematically and perceptually quite distinct, we would like to map them on equally distinct families of hues. In colour space one distinguishes the red–green family from the yellow–blue family. This insight was introduced in colour science by Hering[11], and has indeed many and well confirmed roots in physiology and psychology.

Since red is the most salient colour we (arbitrarily) assign it to the convex umbilical. Then the concave umbilics should be assigned green. (Actually more of a blue–green. However, precise colour naming[12] is not an important issue here.) Thus spherical shells will be tinted red on the outside, green on the inside. Concavities in a globally convex region will stand out very distinctly.

To assign colours to the other shapes we need to assign colours to the concave and convex cylindrical shapes (parabolic shapes). The other hues then follow by continuity.

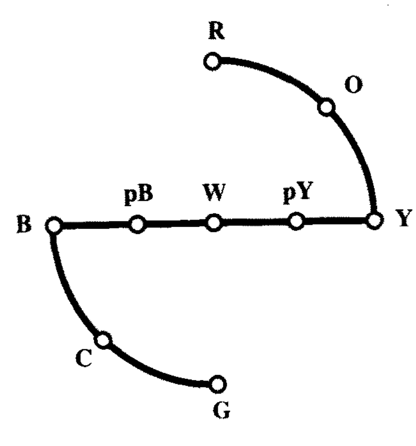
Since the symmetrical saddle has to be assigned white, the saddles must be mapped on a curve that crosses the colour triangle through the white point. If we let the endpoints of this curve be on the perimeter (spectral locus), the elliptical shapes will all be of maximum saturation, and the distinction between hyperbolical and elliptical shapes will be evident.

The elliptical and hyperbolical shapes appear perceptually as almost orthogonal spaces. We may reflect this fact in the colour scale by using the red–green subspace for the elliptical and the yellow–blue subspace for the hyperbolical shapes. This is an apt choice, because the yellow–blue space indeed contains white in the middle. Since yellow is the centre of the red–green subspace we assign it to the convex cylinder. Then the convexities map upon the range red–yellow, which is half of the red–green subspace. The concavities then map automatically on the complementary range green–blue. (Blue is the concave cylinder.) Thus the elliptical shapes map upon two disjoint ranges, that are connected by the saddles, just as they should.

In this way the various relations between the shapes are nicely reflected by the quite similar relations (of complementarity and continuity in both space and natural subspaces) between the various colours. In Figure 4 the embedding of the shape index scale in colour space is indicated in a schematic fashion. There is remarkably little freedom for other than minor changes to this scale. The ony obvious major change that would meet the requirements is an inversion of the scale.

The *curvedness* would naturally be represented by the *intensity*. For the intensity scale is a halfline just like

Figure 4. Colour scale for shape mapped on the colour circle. Letters designate colours: **R***: red;* **O***: orange;* **Y***: yellow;* **pY***: pale yellow;* **W***: white;* **pB***: pale blue;* **B***: blue;* **C***: cyan;* **G***: green. Notice complementary pairs* **R–G**, **O–C**, **Y–B**, **pY–pB**, *and the self-complementarity of* **W***)*



the curvedness scale $c \in [0, \infty]$, moreover, at the origin both colour (hue) and shape (shape index) become undefined. With decreasing intensity (curvedness), all hues (shapes) become black (planar): thus black is a 'non-hue' like planar is a 'non-shape'.

Natural as the intensity dimension appears, it also appears to be the case that human observers find it hard to use. The reason is that, for example, a brown is not perceived as an orange of low intensity, or an olive as a dark green, but these colours appear as distinct hues[11,13]. We have found that it works much better to keep the shape and size dimensions *separate*. The curvedness is best appreciated perceptually in a monochrome image with intensity representing curvedness, the shape is best appreciated in a colour display with maximum intensity hues. An added reason is the different *uses* of such displays: in the curvedness representation one looks for a smooth distribution, in the shape index representation one looks for a *segmentation* of the surface according to shape categories. The suggested representations are optimally adapted to such uses.

Because one typically aims at a segmentation of a surface according to local shape, it makes sense to design a discrete scale with a limited number of shape categories. The simplest example is a three point scale: concave, saddle, convex. We have found that our observers easily (and reproducibly) use a finer scale. The categories shown in Table 1 are rather easy to remember and use.

This is a nine-point scale on which subjects make hardly any mistakes when both principal curvatures are in an appropriate range (neither almost planar, nor too curved; of course, subjects cannot handle almost planar patches or needle tips). The scale is graphically represented in Figure 5. We show small inset figures of shapes taken from the midpoints of the bins to give a feeling for the habitus of the local surface from the various categories. These examples have equal curvedness.

In the rightmost column of Table 1 we have indicated the RGB-values that will yield the right hue on most colour devices. Excellent results are obtained on two bit per colour RGB–devices. On one bit per colour RGB–devices, the scale **R–Y–W–B–G** yields a natural and useful five–point scale. Thus software can be written that will adapt to the available hardware in a graceful manner. The red–yellow range is composed of long wavelength radiation, whereas the blue–green range is composed of short wavelength radiation. This is an approximation of Ostwald's[13] longwave and shortwave 'optimal colours'.

Table 1. Shape categories

Mnemonic	*Index range*	*Colour name*	*(R, G, B)*
Sperhical cup	$s \in [-1, -7/8)$	Green	(0, 1, 0)
Trough	$s \in [-7/8, -5/8)$	Cyan	(0, 1, ½)
Rut	$s \in [-5/8, -3/8)$	Blue	(0, 1, 1)
Saddle rut	$s \in [-3/8, -1/8)$	Pale Blue	(½, 1, 1)
Saddle	$s \in [-1/8, +1/8)$	White	(1, 1, 1)
Saddle ridge	$s \in [+1/8, +3/8)$	Pale Yellow	(1, 1, ½)
Ridge	$s \in [+3/8, +5/8)$	Yellow	(1, 1, 0)
Dome	$s \in [+5/8, +7/8)$	Orange	(1, ½, 0)
Spherical cap	$s \in [+7/8, +1]$	Red	(1, 0, 0)

One point one immediately notices is that the bins at the extremes of the scale are of only half the width of the other bins. Thus it might seem that this scale is uneven. This is partly true: if shapes are drawn 'at random' (e.g. (κ_1, κ_2) from an isotropic normal distribution with mean (0, 0)), then the bins at the extremes receive only half as many items as the other bins. From another point of view, the endpoints of the scale ($s = \pm 1$) are not at the *edges* but at the *centres* of the bins at the extremes. This is immediately obvious if you draw the bins in the $\kappa_1\kappa_2$-plane: in this representation all the bins subtend equal angular sectors, thus the scale is 'equally tempered' on the full circle. In the fully circular representation the bins on the inside of the s-scale appear *twice*, in two different orientations. For the shapes at the ends of the scale the orientation is very ill defined. Because the inner bins occur twice they collect twice as many samples.

When you use a scale with an even number of bins you circumvent this problem. However, we prefer to have the 'anchor points' of the shape scale (the easily remembered and recognized cases) at the centres of the bins. For the bins at the extreme ends of the scale the 'centres' happen to be at the outer edges of these bins. The fact that the bins at the end are half as large as the other ones is of minor importance in practice. The really important fact is that the categories span equal ranges of shape variation (angles in the $\kappa_1\kappa_2$-plane).

EXAMPLES

To illustrate the use and the convenience of the scale proposed in this paper we present a few examples. These examples have been picked to illustrate various

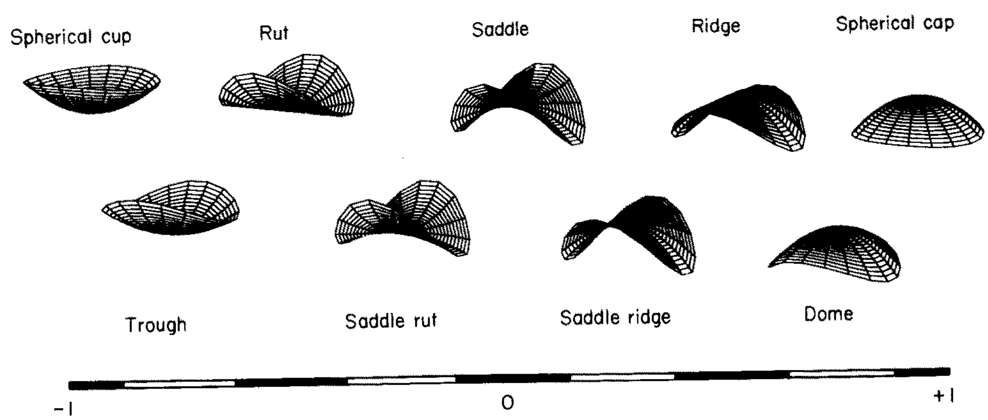


Figure 5. Illustration of shape index scale divided into nine categories: spherical cup, trough, rut, saddle rut, saddle, saddle ridge, ridge, dome and spherical cap. The inset figures are shapes of identical curvedness, drawn from the centres of the categories. They are all placed in the same attitude, with the outward normal pointing upwards

aspects of the new curvature measure especially well. That is why we have opted for synthetic data, rather than real data.

The nine-point scale introduced in the previous section is illustrated in Plate 1 (see p. 565). The various relations between the colours (yellow–blue and red–green subspaces, complementaries, achromatic point) are very clear, and the reader is invited to verify them.

Triaxial ellipsoid

The triaxial ellipsoid is one of the classical shapes from geometry and mathematical physics. Its main features are the locations of the principal axes and associated planes of (mirror) symmetry, and the four 'circular points'. The circular points are special in the sense that any plane that is parallel to a tangent plane at a circular point and that intersects the ellipsoid does so with a *circular* cross section. This is most remarkable, because a plane drawn at random is almost sure to intersect the ellipsoid with an *elliptical* cross section. At the circular points the surface is umbilic (shape index unity).

One would intuitively expect to encounter these major features in a plot of the classical curvature measures (Gaussian and mean curvature). Such is not the case, however. Only some vague indication of the symmetry planes is apparent.

The shape index very clearly reveals the circular points (see Figure 6). They appear as local extrema of the shape index. Other extrema and saddlepoints of the shape index occur on the endpoints of the principal axes, just as in the cases of the Gaussian and mean curvature representations. The circular points are not revealed by plots of either the Gaussian, the mean, or either of the principal curvatures.

In a colour representation most of the ellipsoid is orange, changing towards orange–yellow on the major equator, and with saturated red blobs at the circular points.

Torus

The case of the torus of revolution is of interest because it is one of the few well known classical shapes with a rather large hyperbolic patch. The torus has a pair of tangent planes that touch it along circles. These circles are parabolic curves, and divide the surface of the torus into two regions: a saddle-like region and a convex region. In the example (see Plate 2, p. 565) the proportions have been picked in such a way that the surface on the inner equator is locally of a symmetrical saddle shape (thus maps on white).

At the parabolic points the surface is locally of the shape of a convex cylinder, with the cylinder axis along the parabolic curve. Thus these circles map on yellow. At the outer equator the surface is locally dome shaped, and thus maps on orange to orange–red.

In this case, plots of the Gaussian curvature give very similar information as the shape index.

Local convexity

The Gaussian of revolution is a convenient model for a local convexity on a globally planar surface (see Plate 3, p. 565).

Because of the circular symmetry, the top of the Gaussian hill is an umbilical, and because it is convex it maps to red. At the inflections of the generating curve the surface has a (circular) parabolic curve. The local cylinder axes lie in planes through the symmetry axis, thus the surface is locally of a ridge shape and maps to yellow. The hyperbolic region at larger distances from the centre maps on pale yellowish to white.

This pattern is characteristic for any local protrusion on a globally much less curved surface.

Local concavity

The inverted Gaussian of revolution is a convenient model for a local concavity on a globally planar surface (see Plate 4, p. 565).

Because of the circular symmetry, the very bottom of the Gaussian pit is an umbilical, and because it is concave it maps to green. At the inflections of the generating curve the surface has a (circular) parabolic curve. The local cylinder axes lie in planes through the symmetry axis, thus the surface is locally of a rut shape and maps to blue. The hyperbolic region at larger distances from the centre maps on pale blue to white.

This pattern is characteristic for any local pit in a globally much less curved surface. Notice how corresponding points in Plates 3 and 4 are complementary coloured.

Two types of parabolic curve

In the previous two examples we demonstrated two different types of parabolic curves ('yellow' and 'blue' ones). It is one of the advantages of the colour scale that these two types (like convexities or concavities) are

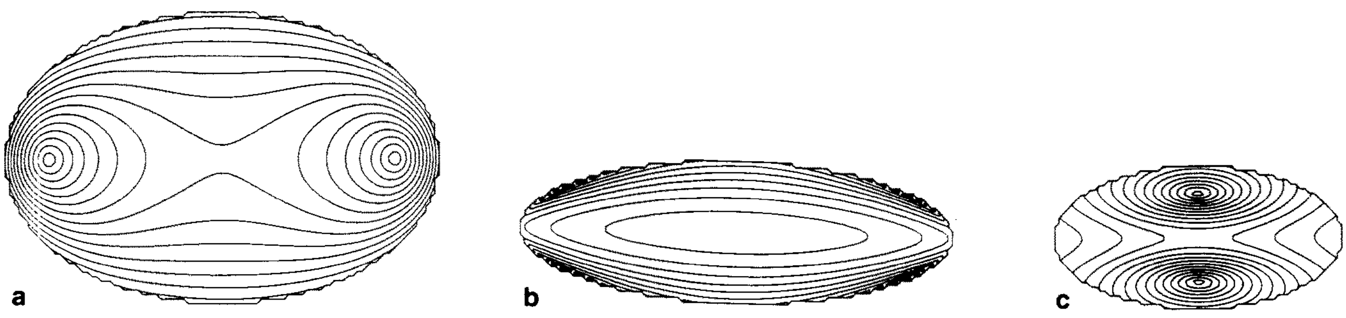


Figure 6. Curves of equal shape index *on the triaxial ellipsoid in orthogonal projection on the XY, XZ, and YZ planes. We use the American style 'third angle projection'. Notice the striking rendering of the umbilicals on the flattish top of the ellipsoid, and the extremum on the equator where the local shape verges on the cylindrical*

clearly distinguished. This is not the case for either the Gaussian or the mean curvature.

In the example of this section we show a surface patch where a concavity and a convexity almost meet (see Plate 5, p. 565; of course, they *have* to be separated by a hyperbolic region).

The example is a perturbed 'handkerchief surface' (René Thom's *mouchoir plié en quatre*[14]). In the usual representations one tends to miss the crucial point that the two branches of parabolic curve are of *opposite* types. This fact is sufficient to see immediately that there has to be a curve of symmetrical saddles in the hyperbolic area (because yellow can only change to blue via white), and that no perturbation be imagined that would make the elliptical regions merge.

CONCLUSION

We have introduced a pair of novel measures of local surface shape: the 'shape index' and the 'curvedness'. The measures describe the second order structure of the surface in the neighbourhood of any one of its points. The only points where this breaks down would be planar points, but these do not occur on generic surfaces (i.e. an infinitesimal perturbation is sufficient to remove them).

The information contained in the shape index and curvedness are formally equivalent to either the two principal curvatures, or to the Gaussian and mean curvature taken as a pair. The major advantage of our proposal is that, unlike principal curvatures or Gaussian and mean curvature, *size* and *shape* are decoupled: the shape index specifies shape, quite independently of size, whereas the curvedness specifies the size. Other advantages are that the shape index does not depend on the (arbitrary) assignment of principal direction (e.g. interchanging the two principal curvatures leaves the shape index invariant), and that the salient curves defined by the Gaussian and the mean curvature (parabolic curves and minimal curves respectively) are reflected in curves of equal shape index (± 0.5 and 0, respectively). Moreover, the parabolic curves (zero Gaussian curvature) are neatly distinguished as two different types by the shape index. Thus the shape index really carries all the relevant shape information, a *single* number suffices instead of two.

An additional advantage is that the shape index scale is easily adapted to a categorical scale, and that the categories correspond to intuitively distinct shapes. We have exploited this fact in the design of a colour scale that enables a fast assessment of shape variations over the surface of an object.

The main thrust of this paper then is the design of a greatly improved user interface in the display of smooth objects for the task of detecting shape variations.

ACKNOWLEDGEMENTS

This research is part of the SPIN project, '3D Computer Vision', supported by the Dutch Ministry of Economic Affairs.

REFERENCES

1 **Euler, L** 'Recherches sur la courbure des surfaces', *Mém. Ac. Berlin*, Vol 16 (1760, appeared 1776)
2 **Gauss, C F** 'Disquisitiones generales circa superficies curvas', *Ostwalds Klassiker der exakten Wissenschaften,* 5 (1827)
3 **Hilbert, D and Cohn-Vossen, S** *Anschauliche Geometrie*, Dover, New York (1944)
4 **Meusnier, J B** 'Mémoire sur la courbure des surfaces', *Mém. Sav. étr.*, Vol 10 (1785, read 1776)
5 **Spivak, M** *Differential Geometry (Vol III)*, Publish or Perish, Inc., Berkeley, CA (1975)
6 **Koenderink, J J** *Solid Shape*, MIT Press, Cambridge, MA (1990)
7 **Koenderink, J J and van Doorn, A J** 'Invariant properties of the motion parallax field due to the movement of rigid bodies relative to an observer', *Optica Acta*, Vol 22 (1975) pp 773–791
8 **Dimmick, F L and Hubbard, M R** 'The spectral location of psychologically unique yellow, green and blue', *Am. J. Psychol.*, Vol 52 (1939) pp 242–254
9 **Dimmick, F L and Hubbard, M R** 'The spectral components of psychologically unique red', *Am. J. Psychol.*, Vol 52 (1939) pp 348–353
10 **Bouma, P J** *Physical Aspects of Colour*, N.V. Philips Gloeilampenfabrieken, Eindhoven (1945)
11 **Hering, E** *Grundzüge der Lehre vom Lichtsinn*, Julius Springer, Berlin (1920)
12 **Kelly, K L and Judd, D B** *Color, universal language and dictionary of names*, National Bureau of Standards, USA, Spec. Publ. 440 (1976)
13 **Ostwald, W** 'Grundsätzliches zur messende Farbenlehre', *S.-B. preuss. Akad. Wiss. (Physik.-math. Kl.)*, Vol 30 (1937) p 414
14 **Thom, R** *Stabilité structurelle et morphogénèse*, Benjamin, New York (1972)