Computer-Aided Design, Vol. 27, No. 5, pp. 363–375, 1995

0010-4485/95 $10.00+0.00

Geometric modelling using rational Gaussian curves and surfaces

Ardeshir Goshtasby

A geometric modelling system based on rational Gaussian (RaG) curves and surfaces is introduced. The generation of simple geometric primitives such as lines, circles, and ellipses with RaG curves, and the generation of planes, spheres, ellipsoids, cylinders, cones, and tori with RaG surfaces are discussed. The design of freeform closed, half-closed, and open shapes using RaG surfaces is also considered. The control points of a RaG surface are not required to form a topologically rectangular grid, but, rather, they can form an arbitrary grid.

Keywords: geometric modelling, rational Gaussian curves and surfaces, shapes

'Geometric modelling' refers to a collection of techniques used to design 2D and 3D shapes, and it is usually based on a particular curve and surface formulation. For instance, Alpha_1[1] is based on B-splines, and GEOMOD[2,3] is based on nonuniform rational B-splines (NURBS). Choosing a particular formulation enables a geometric modelling system to design certain shapes easily. For instance, systems based on NURBS can reconstruct conic sections while systems based on B-splines cannot design such shapes.

This paper discusses a geometric modelling system which is based on rational Gaussian (RaG) curves and surfaces[4]. RaG curves and surfaces were originally formulated to recover structures in dense and noisy point sets. In this paper, it will be shown that RaG curves and surfaces can also be used to design 2D and 3D shapes. Formulae for RaG curves and surfaces are similar to those for NURBS except that B-spline bases are replaced by Gaussian frequency functions. This change in formulation enables the standard deviations of Gaussian functions to be used as free parameters to control the curvature values in a generated shape.

Department of Electrical Engineering and Computer Science (M/C 154), College of Engineering, University of Illinois at Chicago, 1120 Science and Engineering Offices, 851 South Morgan Street, Chicago, IL 60607-7053, USA
Paper received: 17 August 1993. Revised: 18 January 1994

In the following, the properties of RaG curves and surfaces are first reviewed, and then their applications to 2D and 3D shape design are investigated.

RAG CURVES

Formulation

A RaG curve with control points $\{\mathbf{V}_i: i=1, \ldots, n\}$ is defined by[4]

$$\mathbf{P}(u)=\sum_{i=1}^{n} \mathbf{V}_i g_i(u) \qquad u \in [0, 1] \tag{1}$$

where

$$g_i(u)=\frac{W_i G_i(u)}{\sum_{j=1}^{n} W_j G_j(u)} \tag{2}$$

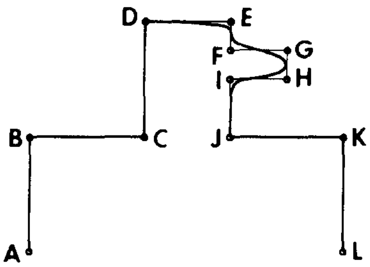
is the ith basis function of the curve, W_i is the weight of the ith control point, and

$$G_i(u)=\exp\{-(u-u_i)^2/2\sigma_i^2\} \tag{3}$$

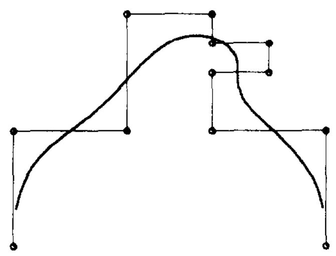
is a Gaussian function of height 1 and standard deviation σ_i centred at u_i.

Analogously to NURBS curves, the sum of the basis functions in RaG curves amounts to 1 for any value of $u \in [0, 1]$, and, therefore, RaG curves always fall inside the convex hulls of their control points. The weight of a control point shows the degree of influence of that control point on a generated curve. The larger the weight, the more the curve is pulled towards that control point. A control point with a weight of 0 will have no influence on a curve. A negative weight pushes a curve away from the corresponding control point. The parameter u_i, which is known as the ith node of the curves, is the parameter value at which the effect of the ith control point on the curve is at a maximum.

Unlike NURBS curves that have fixed basis functions, the basis functions of RaG curves can be changed by



(a)



(b)

Figure 1 Two RaG curves with same control points, nodes and weights, but different standard deviations of Gaussian functions; (a) standard deviations of all Gaussian functions are 0.03, (b) standard deviations are all 0.1
[(a) The curve gets closer to points **A**, **B**, **C**, **D**, **J**, **K** and **L** than to points **E**, **F**, **G**, **H** and **I** because the distances between the former points are larger than the distances between the latter ones. The curve near the former points is mostly influenced by one point, while the curve near the latter points is significantly influenced by two or more points. (b) Since the standard deviations of the Gaussian functions are rather large, a curve point is significantly influenced by all the control points.]

varying the standard deviations of the Gaussian functions. The standard deviation of a Gaussian function determines the degree of localness of the corresponding control point on the shape of a curve. As the value of σ_i is decreased, the ith basis function becomes more locally centred, and the effect of $\mathbf{V}_i$ on the curve becomes more local. Therefore, moving $\mathbf{V}_i$ will mostly affect the shape of the curve in its immediate neighbourhood. As σ_i is increased, the effect of $\mathbf{V}_i$ broadens, and, when a control point is moved, its effect becomes visible in a larger portion of the curve. When the standard deviations of the Gaussian functions are decreased, a RaG curve follows its control points more rigorously and produces local details in the data. Increasing the standard deviations of the Gaussian functions causes a curve to have smaller curvature values and thus it captures only the global features in a shape. This is shown in *Figure 1*. The RaG curves shown have the same control points, nodes, and weights, but have different standard deviations of the Gaussian functions. In *Figure 1a*, the curve is very close to control points **A**, **B**, **C**, **D**, **J**, **K** and **L** because the control points are relatively far from each other and the shape of the curve near these points is mostly determined by the influence of a single point. The control points **E**, **F**, **G**, **H** and **I** are relatively close to each other, and the shape of the curve near these points is determined by significant influences from two or more points.

The nodes of a parametric curve show the spacing of its basis functions in the parameter space. The nodes of a curve should be chosen such that the influence of a control point becomes maximum at the curve point closest to it. Since a curve cannot be drawn before its nodes are known, the parameter value at which a curve is closest to the corresponding control point cannot be determined. For this reason, the nodes of a curve are taken as equally spaced[5], as spaced proportionally to the chord length between control points[6], or as spaced proportionally to the square-root chord length between control points[7]. Attempts to iteratively determine the optimal spacing between the nodes of a parametric curve have also been made[8–11].

Note that the nodes of a RaG curve do not have to be in the range [0, 1]. It is only necessary that they have the same unit as that of the standard deviations of the Gaussian functions. For convenience, however, the nodes of a RaG curve are taken between 0 and 1 so that, by changing the parameter of the curve from 0 to 1, an entire curve can be traversed. Note also that, when drawing a RaG curve, the computation of each curve point requires the use of the entire set of control points. In practice, however, since Gaussian functions approach 0 exponentially, only a small number of control points are needed in the computation. For instance, assuming that the required accuracy of the computation is ε, and the standard deviations of all Gaussian functions are σ, the curve point at parameter u_0 is determined using only those control points whose nodes are in the range[4]

$$u_0-\sigma(-2\ln\varepsilon)^{1/2}\leqslant u\leqslant u_0+\sigma(-2\ln\varepsilon)^{1/2}$$

For example, when $\sigma=0.02$ and $\varepsilon=10^{-8}$, we find $u_0-0.12\leqslant u\leqslant u_0+0.12$.

Closed curves

To obtain a closed RaG curve, Equation 3 should be replaced by[4]

$$G_i(u)=\sum_{j=-\infty}^{\infty}\exp\{-[u-(u_i+j)]^2/2\sigma_i^2\} \tag{4}$$

The infinity comes from the fact that a Gaussian function extends from $-\infty$ to $+\infty$, and, when a curve is closed, it makes infinite cycles around it. Note that, since $G_i(u+1)=G_i(u)$, the basis functions of a closed RaG curve are periodic. Since a Gaussian function approaches 0 exponentially, however, it is only necessary to change j from $-\lceil\sigma(-2\ln\varepsilon)^{1/2}\rceil$ to $\lceil\sigma(-2\ln\varepsilon)^{1/2}\rceil$, where ε is the accuracy of the computation, and σ is the standard deviation of the Gaussian functions[4]. For instance, when $\sigma=0.1$ and $\varepsilon=10^{-8}$, we find j to be in the range $[-1, 1]$.

Figure 2 shows two closed RaG curves with the same control points, weights, and nodes, but different standard deviations of the Gaussian functions. As can be observed, by increasing the standard deviations of the Gaussian functions, a curve with smaller curvature values is obtained. The left tip of *Figure 2a*, which looks like a kink, is not actually a kink, but a highly curved point, as is evident in the magnified circular window. Note that, since Gaussian functions are infinitely differentiable everywhere, closed as well as open RaG curves are infinitely differentiable everywhere.

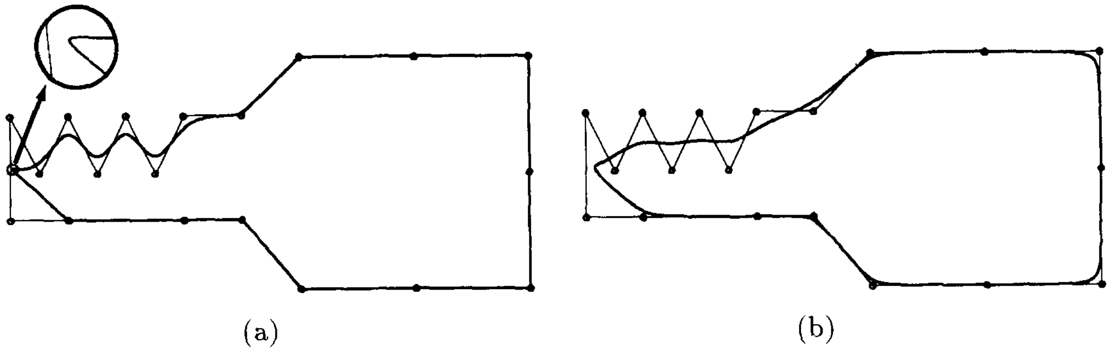


Figure 2 Two closed RaG curves obtained from same control points and weights, but different standard deviations of Gaussian functions [(a) The standard deviations of Gaussian functions are 0.02, (b) the standard deviations are 0.03. (a) The left tip of the curve is not a kink, but rather a highly curved point as shown in the magnified circular window.]

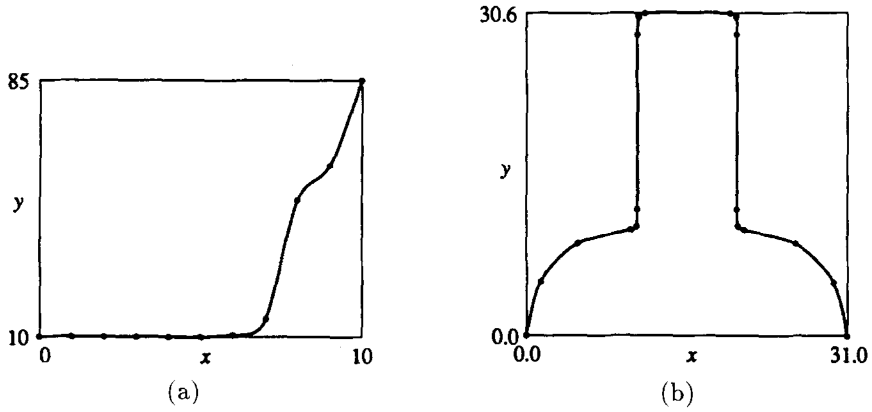


Figure 3 RaG curves interpolating data; (a) data of Akima[13], with standard deviations of Gaussian functions of 0.05, (b) data of Lee[7], with standard deviations of Gaussian functions of 0.03

Interpolating curves

In computer-aided design, it is sometimes necessary to design a curve that interpolates a set of points. To obtain an interpolating curve, the control points of the curve have to be determined by requiring the curve to pass through given points. Therefore, using interpolating points $\{\mathbf{P}_k \equiv \mathbf{P}(u_k)\colon k=1,\ldots,n\}$, control points $\{\mathbf{V}_i\colon i=1,\ldots,n\}$ should be determined by an inverse process. As with approximating curves, the weights, standard deviations of the Gaussian functions, and nodes of an interpolating curve must be provided before it can be drawn. Methods for the estimation of the nodes of a parametric curve have been reviewed by Farin[12]. If no information about the weights of a curve is given, they may be given the same value (e.g. 1). The standard deviations of the Gaussian functions are selected such that desired curvature values are obtained for a shape. For instance, to obtain a shape with smaller curvatures, larger standard deviations of the Gaussian functions should be used.

For a curve to interpolate points $\{\mathbf{P}(u_k)\colon k=1,\ldots,n\}$ we have to determine $\{\mathbf{V}_i\colon i=1,\ldots,n\}$ such that

$$\mathbf{P}_k = \frac{\sum_{i=1}^{n} \mathbf{V}_i W_i \exp\{-(u_k-u_i)^2/2\sigma_i^2\}}{\sum_{j=1}^{n} W_j \exp\{-(u_k-u_j)^2/2\sigma_j^2\}} \qquad k=1,\ldots,n \tag{5}$$

Figure 3 shows RaG curves interpolating the data of Akima[13] and Lee[7] with equal weights and uniform parameter spacing being used. Note that, in spite of uniformly spaced nodes, monotonicity has been preserved in the data of Akima, and no loops or overshoots have been obtained in the data of Lee. This can be attributed to the standard deviations of the Gaussian functions which can be freely varied to achieve desired curvatures in a shape. In this example, the standard deviations of the Gaussian functions were interactively chosen to generate these curves. In *Figure 3a*, monotonicity has been preserved, because it was possible to select appropriately small standard deviations. If large standard deviations are chosen, interpolating RaG curves may also generate undesirable fluctuations.

Because of the nature of RaG functions, the matrix of coefficients obtained in Equation 5 is diagonally dominant. As the standard deviations of the Gaussian functions decrease, the elements of the matrix of coefficients become more distinct. As the standard deviations of the Gaussian functions increase, the off-diagonal elements in the matrix of coefficients become more similar to the diagonal elements, making the system of equations ill conditioned. Since arbitrarily small standard deviations of the Gaussian functions can always be chosen, interpolation by RaG curves can always be guaranteed to be well conditioned. The standard deviations of the Gaussian functions may be chosen such

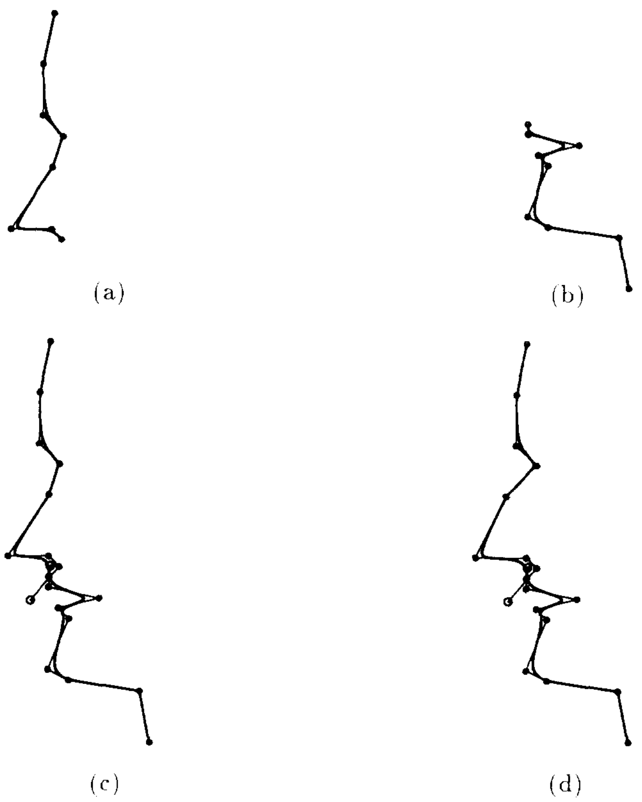


Figure 4 Composite curves; (a), (b) two separately designed curve segments, (c) smoothly joining the two curve segments by satisfying Equations 10, (d) modifying upper curve segment in *Figure 4c* using Equations 11 without disturbing lower segment

that the matrix of coefficients has a condition number below a prescribed value. This capability is not shared by NURBS curves, since they have fixed basis functions.

Composite curves

The formulations given in the preceding sections were for a single curve segment. In design, however, it is often necessary to construct a complex shape that is composed of many curve segments, each designed separately. Suppose that two curve segments as shown in *Figures 4a* and *b* are given, and there is a need to smoothly join them. Let the endpoint of a curve segment which should be joined to another segment be denoted the *free end*, and the other endpoint be the *fixed end*. Since it is possible that the curve segments are already connected to other segments at their fixed ends, care must be taken to ensure that, when the two curve segments are joined, the positions and tangents of the fixed ends of the curves do not change. To ensure this, the new positions of two control points at the fixed end of each segment are determined. In addition, a new control point is introduced at the free end of each curve segment to ensure that they join smoothly. Therefore, to smoothly join two curve segments without disturbing the segments that are connected to them, the positions of three control points should be determined from each segment.

Assuming curve segments

$$\mathbf{P1}(u)=\sum_{i=1}^{m}\mathbf{V1}_i g1_i(u) \tag{6}$$

and

$$\mathbf{P2}(u)=\sum_{i=1}^{n}\mathbf{V2}_i g2_i(u) \tag{7}$$

are given, and these curve segments after being smoothly joined are represented by

$$\mathbf{P3}(u)=\sum_{i=1}^{m+1}\mathbf{V3}_i g3_i(u) \tag{8}$$

and

$$\mathbf{P4}(u)=\sum_{i=0}^{n}\mathbf{V4}_i g4_i(u) \tag{9}$$

we determine three control points from each curve segment, namely $\mathbf{V3}_1$, $\mathbf{V3}_2$, $\mathbf{V3}_{m+1}$ and $\mathbf{V4}_0$, $\mathbf{V4}_{n-1}$, $\mathbf{V4}_n$, by solving

$$\begin{aligned}
\mathbf{P3}(0)&=\mathbf{P1}(0)\\
\mathbf{P3}'(0)&=\mathbf{P1}'(0)\\
\mathbf{P3}(1)&=\mathbf{P4}(0)\\
\mathbf{P3}'(1)&=\mathbf{P4}'(0)\\
\mathbf{P4}(1)&=\mathbf{P2}(1)\\
\mathbf{P4}'(1)&=\mathbf{P2}'(1)
\end{aligned} \tag{10}$$

In Equations 10, $\mathbf{V1}_i=\mathbf{V3}_i$ for $i=3,\ldots,m$, and $\mathbf{V2}_i=\mathbf{V4}_i$ for $i=1,\ldots,n-2$. *Figure 4c* shows a composite curve obtained in this manner using the curve segments of *Figures 4a* and *b*.

When editing a curve segment in a composite formulation, care must be taken to ensure that the positions and tangents of the endpoints of the segment do not change. For this to happen, two control points at each end of the segment are allowed to move. Assuming that a curve segment before and after modification is given by $\mathbf{P}(u)$ and $\mathbf{Q}(u)$, the new positions of two control points from each end of the segment are determined by solving

$$\begin{aligned}
\mathbf{P}(0)&=\mathbf{Q}(0)\\
\mathbf{P}(1)&=\mathbf{Q}(1)\\
\mathbf{P}'(0)&=\mathbf{Q}'(0)\\
\mathbf{P}'(1)&=\mathbf{Q}'(1)
\end{aligned} \tag{11}$$

Figure 4d shows modification of the upper segment in *Figure 4c* without disturbing the lower segment using this formulation.

The above formulations provide tools for the design of freeform curves. Certain shapes, however, have well known components, such as lines, circles and ellipses, which have to be reproduced exactly. One of the reasons why NURBS curves are widely used in geometric modelling is that they can reproduce conic sections exactly.

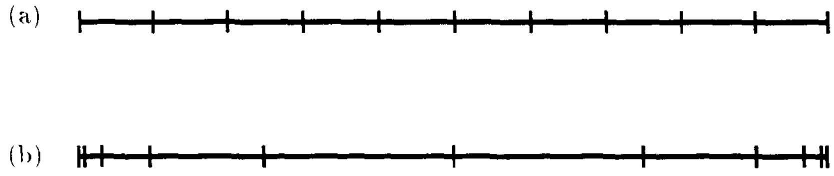


Figure 5 Interpolating RaG lines; (a) standard deviations 1.0, (b) standard deviations 0.3
[The markings show line points obtained with equal increments of 0.1 in u.]

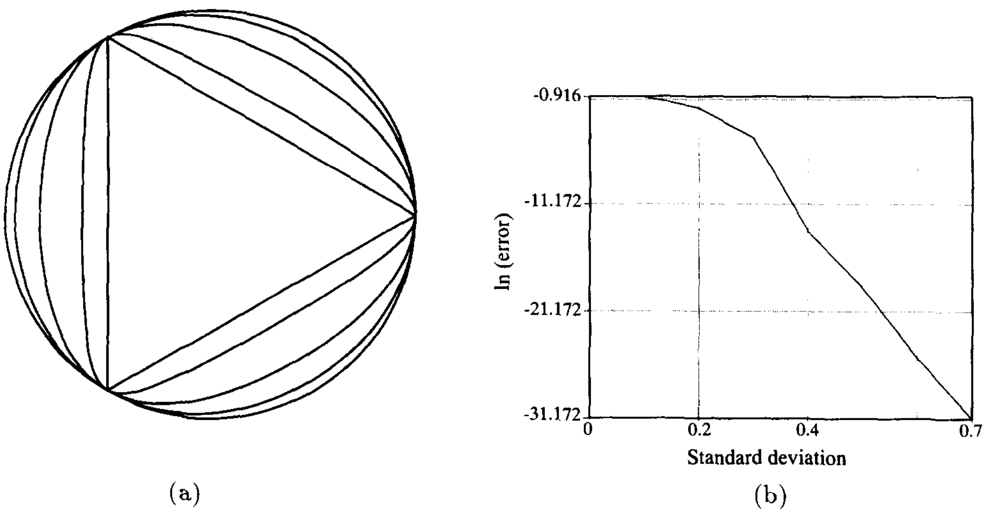


Figure 6 Circle; (a) closed RaG curves interpolating vertices of equilateral triangle, (b) maximum distance between RaG curves and circle as function of standard deviations of Gaussian functions
[(a) The standard deviations of Gaussian functions are 0.02, 0.15, 0.2, 0.25 and 0.5. Owing to the quantization process involved in plotting the curves, when the standard deviations are 0.02, the curve and the triangle cannot be distinguished from each other, and, when the standard deviations are 0.5, the curve falls exactly on the interpolating circle. (b) The radius of the interpolating circle is 0.8.]

Although RaG curves cannot reproduce conic sections exactly, in the following, it will be shown that they can reproduce them within prescribed accuracies.

Lines

A RaG curve interpolating/approximating two points $\mathbf{V}_1$ and $\mathbf{V}_2$,

$$\mathbf{P}(u)=\mathbf{V}_1 g_1(u)+\mathbf{V}_2 g_2(u) \tag{12}$$

will result in a line because $g_1(u)+g_2(u)=1$.

Assuming that u represents time and $\mathbf{P}(u)$ represents the position of a point after travelling for u seconds along a line from a start point, we see that the distance travelled is a function of the standard deviations of the Gaussian functions. As the standard deviations of the Gaussian functions increase, the distance travelled in a unit time (or the travelling speed) approaches a constant value. As the standard deviations of the Gaussian functions decrease, the speed increases towards the centre of the line and decreases towards the endpoints of the line. *Figure 5* shows the density of points or distances travelled with constant increments of u when different standard deviations of the Gaussian functions are used. This property is useful when using a linear trajectory to animate the motion of an object that has to accelerate when departing from a start point, and decelerate when arriving at an endpoint.

This property of RaG lines carries over to RaG curves. When the standard deviations of the Gaussian functions are small, a curve closely follows its control points, and generates high curvature values. To accurately reproduce high curvatures in a shape, small segments should be drawn at high curvature values. Since, in RaG curves, high curvatures are obtained only when small standard deviations of the Gaussian functions are used, and small standard deviations automatically produce small segments at high curvature values, a constant increment in u can faithfully generate a curve that contains low as well as high curvature values.

Circles

A closed RaG curve interpolating the vertices of an equilateral triangle with uniform parameter spacing, equal weights, and equal standard deviations of the Gaussian functions approaches a circle as the standard deviations of the Gaussian functions increase. This is because the curvatures of the curve at different values of u approach the same value as the standard deviations of the Gaussian functions increase. In *Figure 6a*, RaG curves interpolating the vertices of an equilateral triangle are shown. *Figure 6b* shows the maximum distance between an actual circle and the RaG curve that interpolates three

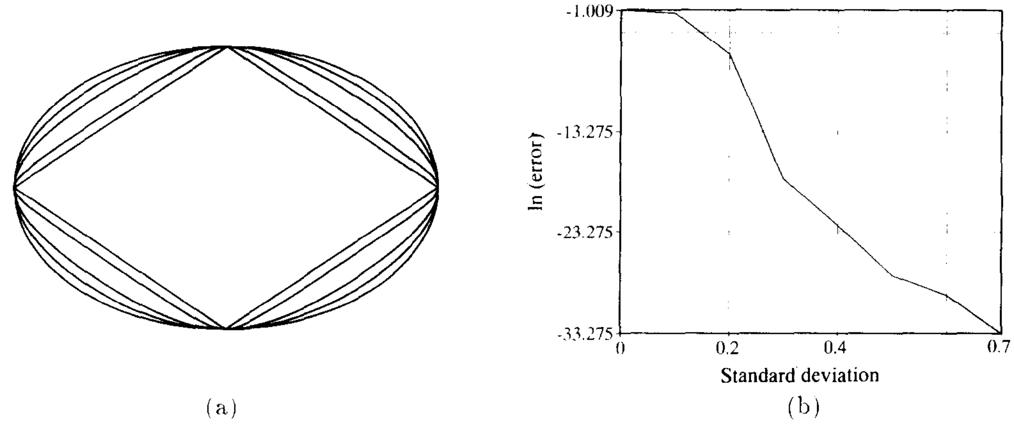


Figure 7 Ellipse; (a) closed RaG curves interpolating vertices of rhombus with standard deviations of 0.02, 0.1, 0.13, 0.15 and 0.5, (b) maximum distance between RaG curves and ellipse as function of standard deviations of Gaussian functions
[The major and minor diameters of the ellipse are 1 and 2/3, respectively.]

equally spaced points on the circle as a function of the standard deviations of the Gaussian functions. The maximum distance between the circle and an interpolating RaG curve is determined by finding the distances of points on the curve to the centre of the circle at increments of 0.001 in u, subtracting them from the radius of the circle, and taking the maximum value.

Note that RaG curves define a circle with only three points while NURBS curves require six points to define a circle[14]. By examining *Figure 6b*, it can be observed that, although RaG curves cannot reproduce circles exactly, they can reproduce them with prescribed accuracies.

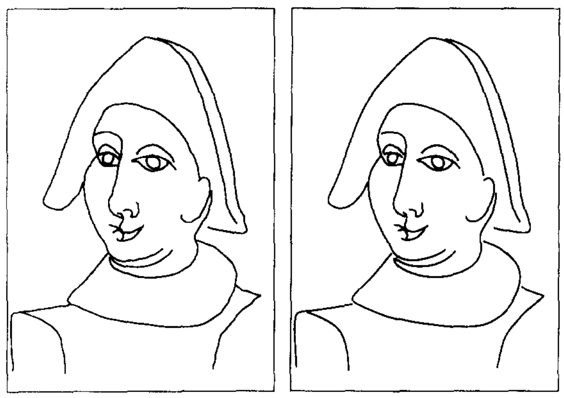
Ellipses

A closed RaG curve interpolating the vertices of a rhombus with equal parameter spacing and equal weights approaches an ellipse as the standard deviations of the Gaussian functions increase (see *Figure 7a*). The error between a true ellipse and its approximating RaG curve as a function of the standard deviations of the Gaussian functions is shown in *Figure 7b*. As the standard deviations of the Gaussian functions increase, any error in the approximation decreases.

Note that, not only can RaG curves generate circles and ellipses with prescribed accuracies, but they can also generate shapes which fall between a triangle and its interpolating circle and a rhombus and its interpolating ellipse.

Complex shapes

With RaG curves, a designer can move the positions of the control points and change their weights just as with NURBS curves. In addition, with RaG curves, a designer can change the standard deviations of the Gaussian functions to make a curve accurately reproduce details in some areas of a shape while smoothing details in other areas. With RaG curves, a designer can fit a single curve segment to a large number of points, thus reducing the number of segments needed to design a complex shape. *Figures 8* and *9* and *Colour Plates 1* and *2* show



(a) (b)

Figure 8 Reproduction of first drawing by Picasso with RaG curves; (a) standard deviations of 0.02, (b) standard deviations of 0.04

applications of RaG curves to the reproduction of two famous drawings by Picasso. The curves were generated by tracing the contours in the original drawings by a digitizer device and fitting RaG curves to the points obtained.

RAG SURFACES

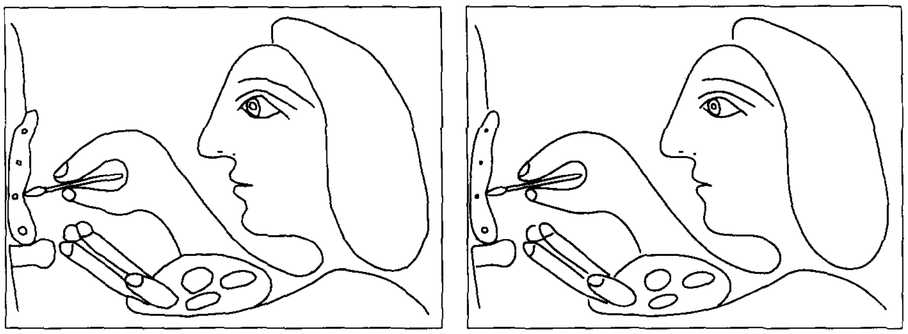
Formulation

A RaG surface approximating a polyhedron with vertices $\{\mathbf{V}_i: i=1, \ldots, n\}$ is defined by[4]

$$\mathbf{P}(u, v)=\sum_{i=1}^{n} \mathbf{V}_i g_i(u, v) \qquad u\in[0, 1] \tag{13}$$

where

$$g_i(u, v)=\frac{W_i g_i(u, v)}{\sum_{j=1}^{n} W_j G_j(u, v)} \tag{14}$$



(a) (b)

Figure 9 Reproduction of second drawing by Picasso using RaG curves; (a) standard deviations of 0.02, (b) standard deviations of 0.04

Table 1 Control points and nodes used for surfaces in *Colour Plate 3*

i	x_i	y_i	z_i	u_i	v_i
1	100	100	100	0.0	0.0
2	100	200	100	0.0	1.0
3	250	150	100	0.5	0.5
4	200	120	50	1.0	0.0
5	200	170	50	1.0	1.0

is the ith basis function of the surface, and

$$G_i(u, v) = \exp\{-[(u-u_i)^2 + (v-v_i)^2]/2\sigma_i^2\} \tag{15}$$

is a 2D Gaussian function of height 1. In this formula, (u_i, v_i) is the ith node of the surface, and it shows the position of the ith Gaussian function in the parameter space. σ_i is the standard deviation of the ith Gaussian function. The larger the value of σ_i, the broader the effect of the ith control point on a generated surface, and the smaller the value of σ_i, the more local the effect of the ith control point on a surface.

The control points of a RaG surface are not required to form a regular grid, and an arbitrary grid of points is sufficient to generate a surface. *Colour Plate 3* shows RaG surfaces generated from five control points. The surfaces have the same control points and nodes, given in *Table 1*, but they have different standard deviations of the Gaussian functions. The nodes of a RaG surface do not have to be in the range $[0, 1] \times [0, 1]$. It is only necessary that they have the same unit as that of the standard deviations of the Gaussian functions. For convenience, the nodes of a RaG surface are varied in $[0, 1] \times [0, 1]$ so that, by changing u and v from 0 to 1, the surface can be drawn in its entirety.

Half-closed surfaces

In design, there is often a need to construct a surface that is smoothly closed from one side but open from the other. For a half-closed surface, Equation 15 should be replaced by[4]

$$G_i(u, v) = \sum_{j=-\infty}^{\infty} \exp(-\{[(u-(u_i+j)]^2 + (v-v_i)^2\}/2\sigma_i^2) \tag{16}$$

Table 2 Control points and nodes used for funnel in *Colour Plate 4*

i	x_i	y_i	z_i	u_i	v_i
1	100	100	100	0.00	0.0
2	100	200	100	0.25	0.0
3	200	200	100	0.50	0.0
4	200	100	100	0.75	0.0
5	145	145	300	0.00	1.0
6	145	155	300	0.25	1.0
7	155	155	300	0.50	1.0
8	155	145	300	0.75	1.0
9	140	140	200	0.00	0.5
10	140	160	200	0.25	0.5
11	160	160	200	0.50	0.5
12	160	140	200	0.75	0.5

The infinity in Equation 16 is due to the fact that a 2D Gaussian function extends to infinity in all directions, and it makes infinite cycles along the closed side of a surface. In practice, however, since a Gaussian function approaches zero exponentially, it is only necessary to vary j between $-\lceil\sigma(-2\ln\varepsilon)^{1/2}\rceil$ and $\lceil\sigma(-2\ln\varepsilon)^{1/2}\rceil$. For example, when $\sigma=0.1$ and $\varepsilon=10^{-8}$, we find j to be in $[-1, 1]$.

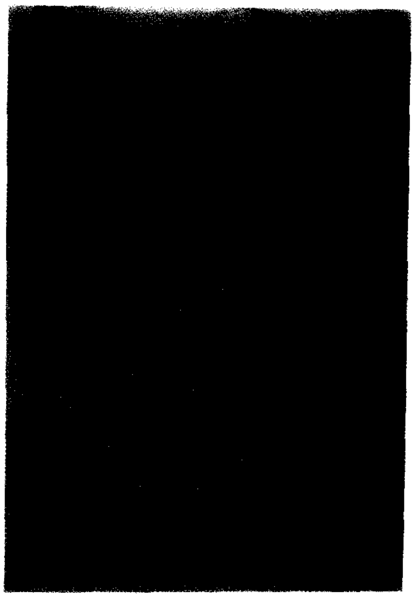
A funnel designed with a half-closed surface is shown in *Colour Plate 4*. The coordinates of the control points and the nodes of the surface are shown in *Table 2*. The weights at all the points are the same, and the standard deviations of all the Gaussian functions are 0.3.

Closed surfaces

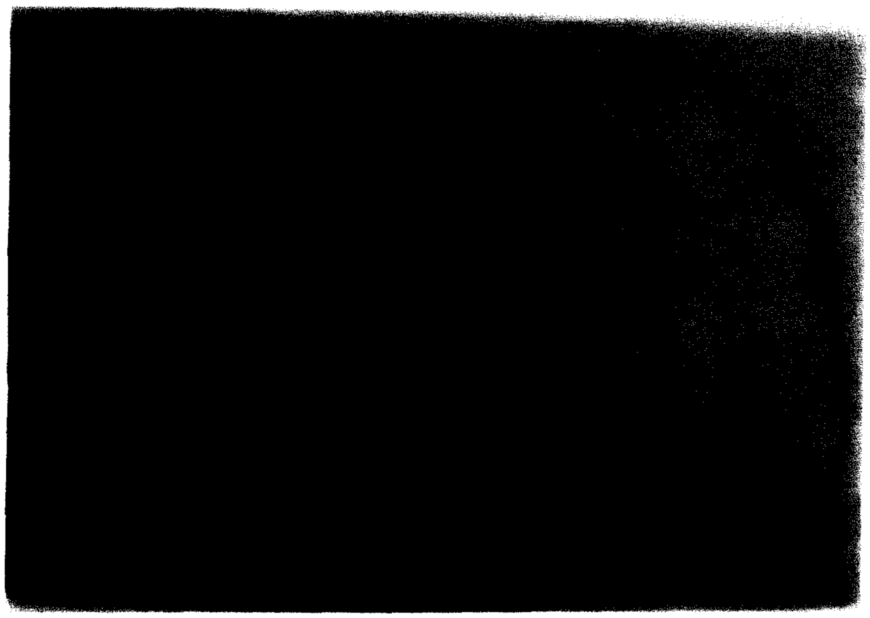
In a closed surface, a Gaussian function makes infinite cycles along both u and v. Therefore, Equation 15 should be replaced by[4]

$$G_i(u, v) = \sum_{j=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} \exp(-\{[u-(u_i+j)]^2 + [v-(v_i+k)]^2\}/2\sigma_i^2) \tag{17}$$

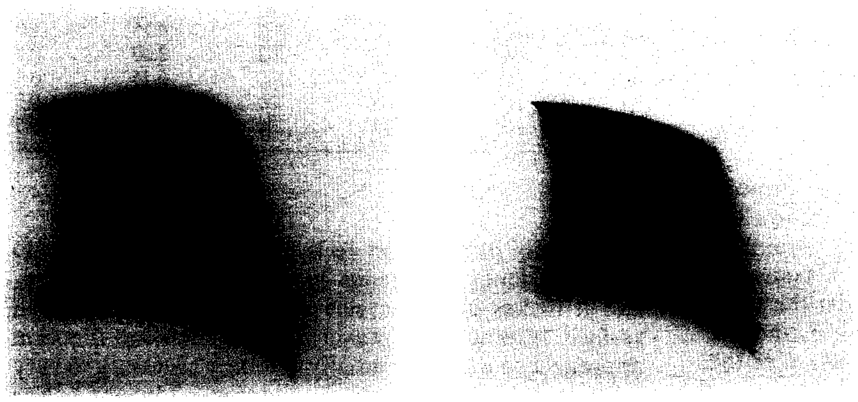
Again, in practice, j and k are small numbers between $-\lceil\sigma(-2\ln\varepsilon)^{1/2}\rceil$ and $\lceil\sigma(-2\ln\varepsilon)^{1/2}\rceil$. A closed surface generated using the control points and nodes given in *Table 3* is shown in *Colour Plate 5*. Equal weights were used, and the standard deviations of all the Gaussian functions are 0.3.



Colour Plate 1 First drawing by Picasso

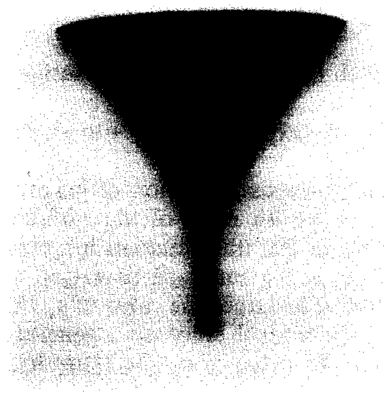


Colour Plate 2 Second drawing by Picasso

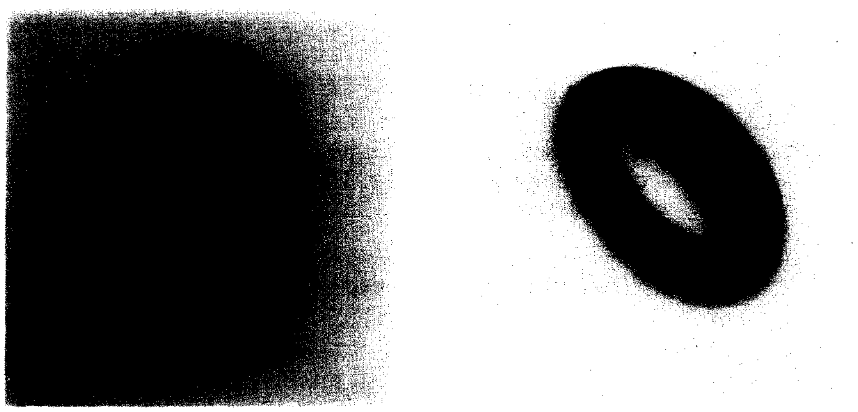


a b

Colour Plate 3 RAG surfaces
[The surfaces were obtained using the control points and nodes given in *Table 1*. Equal weights were used. (a) The standard deviations of all the Gaussian functions were 0.3, (b) the standard deviations were all 0.4.]

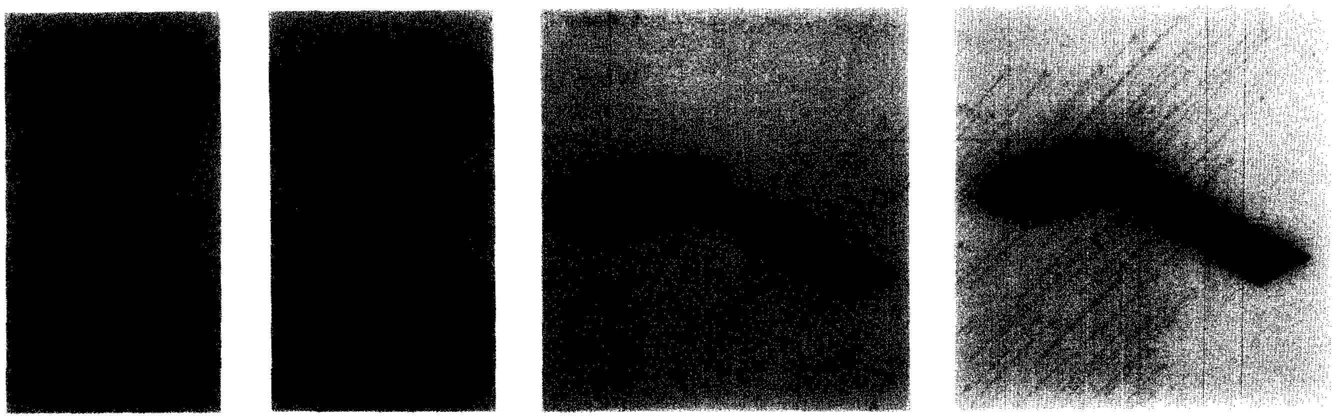


Colour Plate 4 Funnel designed using half-closed RAG surface
[The control points and nodes used are given in *Table 2*. The weights at all the points were the same, and the standard deviations of all the Gaussian functions were 0.3.]



a b

Colour Plate 5 Closed surfaces; (a) ring obtained by approximating closed RaG surface, (b) torus obtained if surface is made to interpolate the points
[(a) All the weights were the same, and the standard deviations of all the Gaussian functions were 0.3. The points used are given in *Table 3*.]



a b c d

Colour Plate 6 Blending and composite surfaces; (a) two disjoint surfaces, (b) surface obtained by smoothly joining the surfaces using Equation 19, (c) two very close surfaces, (d) composite surface obtained by deforming these surfaces to join them smoothly using Equation 20
[The standard deviations of all the Gaussian functions used in *Colour Plate 6* were 0.15.]

Table 3 Points used for ring and torus in *Colour Plate 5*

i	x_i	y_i	z_i	u_i	v_i
1	−1.15	−0.67	0.00	0.00	0.00
2	−0.72	−0.42	−0.29	0.00	0.33
3	−0.72	−0.42	0.29	0.00	0.67
4	1.15	−0.67	0.00	0.33	0.00
5	0.72	−0.42	−0.29	0.33	0.33
6	0.72	−0.42	0.29	0.33	0.67
7	0.00	1.33	0.00	0.67	0.00
8	0.00	0.83	−0.29	0.67	0.33
9	0.00	0.83	0.29	0.67	0.67

Note that closed, half-closed and open RaG surfaces are infinitely differentiable everywhere and in all directions. This property is inherited from Gaussian functions.

Blending and composite surfaces

In geometric modelling, there is often a need to smoothly connect two disjoint surfaces. Suppose that surfaces $\mathbf{P}_1(u_1, v)$ and $\mathbf{P}_2(u_2, v)$ are given and we want to obtain a new surface $\mathbf{P}(u, v)$ that represents both of the surfaces and the blending surface that smoothly joins the two. Let the blending surface be denoted by $\mathbf{P}_3(u_3, v)$. For a constant value of v, the curve that smoothly connects point $\mathbf{P}_1(1, v)$ on the right border of surface $\mathbf{P}_1(u_1, v)$ to point $\mathbf{P}_2(0, v)$ on the left border of surface $\mathbf{P}_2(u_2, v)$ lies on the blending surface. Therefore,

$$\mathbf{P}_3(u_3, v)=\mathbf{P}_1(u_3+1, v)g_1(u_3)+\mathbf{P}_2(u_3-1, v)g_2(u_3) \quad (18)$$

$g_1(u_3)$ and $g_2(u_3)$ are defined by Equation 2 with $n=2$. Assuming that surfaces $\mathbf{P}_1(u_1, v)$ and $\mathbf{P}_2(u_2, v)$ are generated by varying u_1 and u_2 from 0 to 1, the u component of the nodes of surface $\mathbf{P}_2(u_2, v)$ are shifted in the parameter space by two units, and the surface that is composed of the two given surfaces and the blending surface is defined by

$$\mathbf{P}(u, v)=\mathbf{P}_1(u, v)g_1(u-1)+\mathbf{P}_2(u, v)g_2(u-1) \quad (19)$$

Now, by varying u from 0 to 1, a surface that resembles $\mathbf{P}_1$ is obtained, by varying u from 1 to 2 a surface that corresponds to $\mathbf{P}_3$ is obtained, and by varying u from 2 to 3 a surface that resembles $\mathbf{P}_2$ is obtained.

The shapes of the surfaces near the borders that join together will change somewhat to provide a smooth transition from one surface to the next. The smaller the standard deviations of the Gaussian functions in g_1 and g_2, the smaller the change obtained. *Colour Plate 6b* shows an example of a blending surface joining the two surfaces shown in *Colour Plate 6a.*

If the given surfaces are very close or intersecting, as shown in *Colour Plate 6c*, a blending surface is not required to smoothly join the surfaces. The borders of the two surfaces can be appropriately deformed to join smoothly. This deformation can be achieved by formulating the composite surface as

$$\mathbf{P}(u, v)=\mathbf{P}_1(u, v)+\mathbf{P}_2(u, v) \quad (20)$$

Here, the u component of the nodes of surface $\mathbf{P}_2$ should be shifted by one unit in the parameter space to allow Equation 20 to generate the entire composite surface by varying u from 0 to 2. *Colour Plate 6d* shows the composite surface obtained from Equation 20 applied to the surfaces shown in *Colour Plate 6c*. Note that the surface defined by Equation 19 requires that u be varied from 0 to 3, while the surface defined by Equation 20 requires that u be varied from 0 to 2. Also, note that surfaces defined by Equations 19 and 20 are infinitely differentiable everywhere.

Interpolating surfaces

For a RaG surface to interpolate a set of scattered points, its control points have to be determined by solving a system of equations. Because of the shapes of RaG basis functions, the matrix of coefficients obtained in this interpolation will be diagonally dominant. As the standard deviations of the Gaussian functions decrease, the diagonal elements become more dominant and distinct, guaranteeing a solution. As the standard

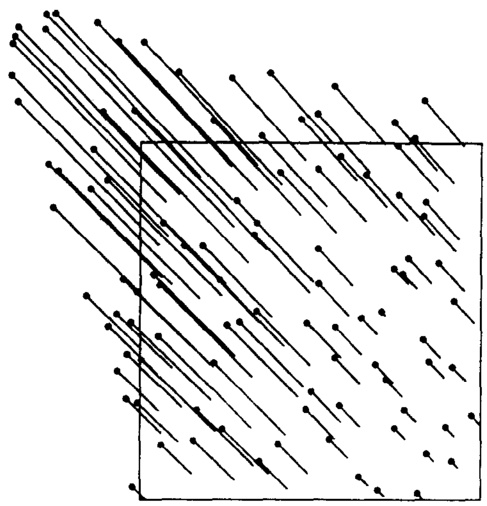


Figure 10 Scattered points due to Franke[15]

deviations of the Gaussian functions increase, the off-diagonal elements become similar to the diagonal elements, making the system of equations ill conditioned. To ensure successful interpolation when a large number of highly varying points are given, very small standard deviations should be used. The existence of standard deviations of Gaussian functions as free parameters enables a designer to generate surfaces with different curvature values interpolating the same set of points.

A RaG surface interpolating the data of Franke[15] (see *Figure 10*) is shown in *Colour Plate 7*. Note that the surface is well behaved, and contains no fluctuations. The weights at all the points are the same, the standard deviations of all the Gaussian functions are 0.1, and the nodes of the surface are proportional to the positions of data points in the xy plane.

Not only can RaG surfaces generate freeform closed, half-closed and open surfaces, but they can also approximate well known surface primitives such as planes, cylinders, cones, spheres, ellipsoids and tori with prescribed accuracies, as demonstrated below.

Planes

A RaG surface interpolating/approximating three non-collinear points $\mathbf{V}_1$, $\mathbf{V}_2$ and $\mathbf{V}_3$,

$$\mathbf{P}(u, v) = \sum_{i=1}^{3} \mathbf{V}_i g_i(u, v) \tag{21}$$

results in a plane, because $\sum_{i=1}^{3} g_i(u, v) = 1$.

Cylinders

A half-closed RaG surface interpolating the vertices of two equal and parallel equilateral triangles approaches a cylinder as the standard deviations of the Gaussian functions increase. This is because, for a constant value of u, a line is traced, while, for a constant value of v, a circle is traced on the surface. *Colour Plate 8a* shows a cylinder obtained in this manner.

Cones

When the area of one of the triangles used to define a cylinder approaches zero, a cylinder converges to a cone, as shown in *Colour Plate 8b*.

Spheres

A sphere is obtained by rotating a half circle about the line that connects its two endpoints. Although a circle can be generated with three points, a half circle requires four or more points for reconstruction. To generate a sphere, therefore, a half-closed surface is interpolated to the vertices of four or more equally spaced equilateral parallel triangles. As the sizes of the upper and lower triangles decrease, the cylinder is transformed to a sphere. A sphere obtained in this manner will have a hole at the top and a hole at the bottom, but the sizes of the holes can be arbitrarily reduced by decreasing the sizes of the upper and lower triangles. For the cylinder to converge to a sphere, the vertices of the triangles should be on the sphere. A sphere obtained in this manner is shown in *Colour Plate 9a*.

Ellipsoids

If, instead of triangles as in *Colour Plate 9a*, rhombi are used, an ellipsoid will be obtained as shown in *Colour Plate 9b*. There is a discontinuity at the top and the bottom of the ellipsoid, but the sizes of the discontinuities can be arbitrarily reduced by decreasing the sizes of the upper and lower rhombi. If the triangles used to design a sphere are not equally spaced, an egg-shaped surface will be obtained as shown in *Colour Plate 9c*.

Tori

A closed RaG surface interpolating the vertices of three equilateral triangles that are centred at a larger equilateral triangle converges to a torus as the standard deviations of the Gaussian functions increase, as shown in *Colour Plate 5b*. This is because, for constant values of u and v, circles are drawn on the surface obtained.

Complex shapes

The RaG formulation allows the fitting of a single patch to a large number of points. *Figure 11* shows an array of range data obtained from a person's face in wireframe form. In *Figure 11*, every fifth line in the array is drawn. Fitting an open RaG surface to the points results in the surface shown in *Colour Plate 10*. In this example, the nodes of the surface are proportional to the (x, y) coordinates of the points, all the weights are the same, and the standard deviations of all the Gaussian functions are 0.03.

Another example of RaG surfaces representing complex shapes is shown in *Figure 12* and *Colour Plate 11*. *Figure 12* depicts points from the left and right ventricles of a human heart obtained from a volumetric magnetic resonance image by a feature selector. Two half-closed

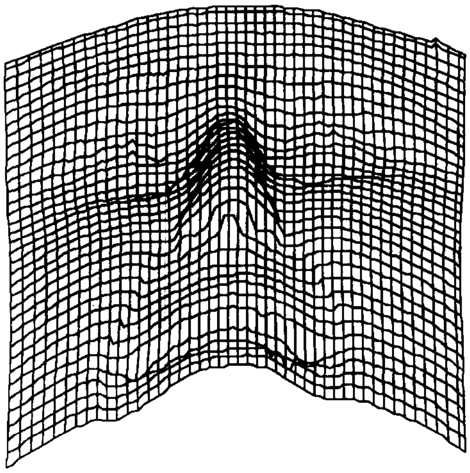


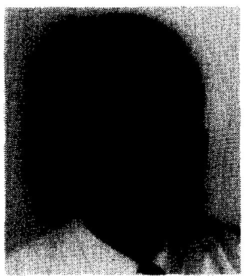
Figure 11 Array of data points shown in wireframe form [Only every fifth line in the wireframe is shown.]

Figure 12 Nonuniformly spaced points on left and right ventricles of human heart obtained by feature selector from volumetric magnetic resonance image

surfaces were fitted to the points to obtain the surfaces shown in *Colour Plate 11*. The surfaces fill in areas where data are missing and smooth noisy details on the ventricles. In this example, v_is were selected along the axis of a half-closed surface, and u_is were selected angularly about the axis. All the weights are the same, and the standard deviations of all the Gaussian functions are 0.03.

CONCLUSIONS

A geometric modelling system based on RaG curves and surfaces has been described. RaG curves and surfaces can approximate well known geometric primitives with prescribed accuracies and can generate complex freeform shapes. A geometric modelling system based on the RaG formulation has the unique ability (which is not shared by the NURBS formulation) to design closed, half-closed, and open shapes from irregularly spaced points. Another advantage of the RaG formulation over the NURBS formulation is that RaG basis functions can be varied by changing the standard deviations of the Gaussian functions, allowing different shapes to be designed while approximating/interpolating the same set of points. This capability is not shared by the NURBS formulation, since it has fixed basis functions. The disadvantage of the RaG formulation in comparison with the NURBS formulation is that it is global, and the computation of a curve or a surface point theoretically depends on all its control points (although, in practice, a portion of them only is sufficient). In contrast, the NURBS formulation is local, and the computation of a curve or a surface point requires the use of only a limited number of the control points. The computation time needed to find a RaG curve point or a surface point varies with the number of the control points and the standard deviations of the Gaussian functions, while the computation time needed to find a NURBS curve point or a surface point is fixed.



Dr Ardeshir Goshtasby is a graduate of the University of Tokyo, Japan, the University of Kentucky, USA, and Michigan State University, USA. He is on the faculty of the Department of Electrical Engineering and Computer Science, University of Illinois at Chicago, USA. His research has been centred on the registration of multimodal and multidimensional images, and curve and surface design. His contributions have been the formulation of image warping as an approximation problem, the discovery of new properties of B-spline curves and surfaces by viewing them as digital filters, and the introduction of a new class of parametric curves and surfaces known as rational Gaussian curves and surfaces, which he has used to solve various computer vision and geometric modelling problems. His current research interests are in the representation and analysis of volumetric images, and the recovery of structure in unorganized multidimensional data.

REFERENCES

1 Riesenfeld, R F, Cohen, E, Fish, R, Thomas, S, Cobb, E, Barsky, B, Schweitzer, D and Lane, J 'Using the Oslo algorithm as a basis for CAD/CAM geometric modeling' *Proc. Nat. Conf. NCGA* Fairfax, VA, USA (1981) pp 345–356

2 Klosterman, A L, Ard, R H and Klahs, J W 'Geometric modeling speeds design of mechanical assemblies' *Comput. Tech. Rev.* (1982) pp 103–111

3 Tiller, W 'Rational B-splines for curve and surface representation' *Comput. Graph. & Applic.* Vol 3 (1983) pp 61–69

4 Goshtasby, A 'Design and recovery of 2-D and 3-D shapes using rational Gaussian curves and surfaces' *Int. J. Comput. Vision* Vol 10 No 3 (1993) pp 233–256

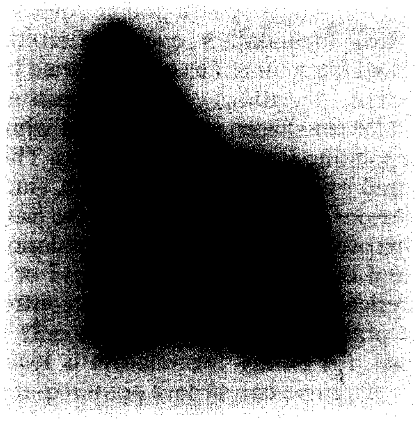
5 Faux, I D and Pratt, M J *Computational Geometry for Design and Manufacture* Ellis Horwood, UK (1979)

6 Epstein, M 'On the influence of parametrization in parametric interpolation' *Numer. Anal.* Vol 17 (1980) pp 238–246

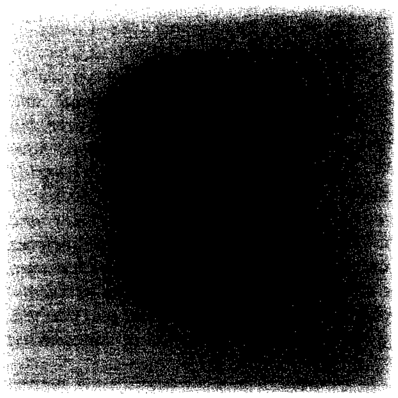
7 Lee, E T Y 'Choosing nodes in parametric curve interpolation' *Comput.-Aided Des.* Vol 21 (1989) pp 363–370

8 Cohen, E and O'Dell, C L 'A data dependent parametrization for spline approximation' *in* Lyche and Schumaker (Eds.) *Mathematical Methods in Computer Aided Geometric Design* (1989) pp 155–166

9 Marin, S P 'An approach to data parametrization in parametric



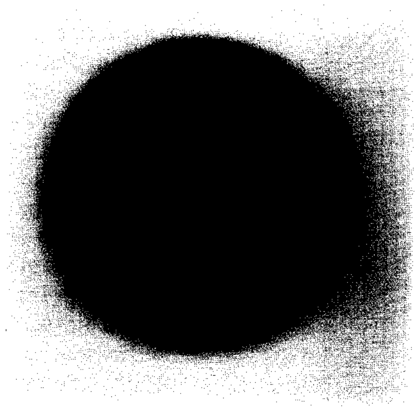
Colour Plate 7 RaG surface interpolating points in *Figure 10*
[The nodes were proportional to the (x, y) coordinates of the points, the weights were the same, and the standard deviations of all the Gaussian functions were 0.1.]



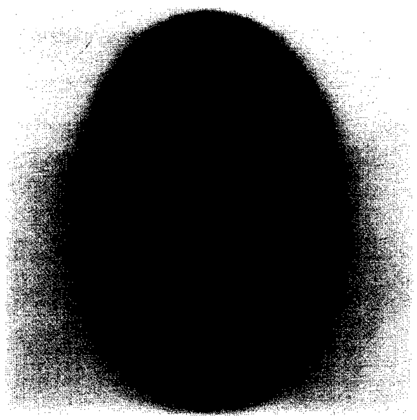
a

b

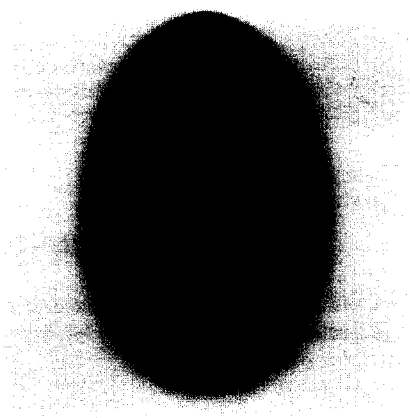
Colour Plate 8 Cylinder; (a) cylinder obtained by interpolating half-closed RaG surface to vertices of two equal and parallel equilateral triangles, (b) when area of one triangle approaches zero, cylinder is transformed to cone



a

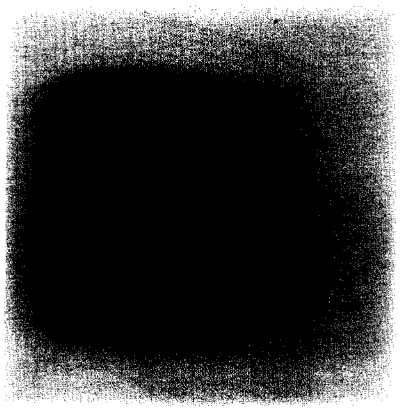


b

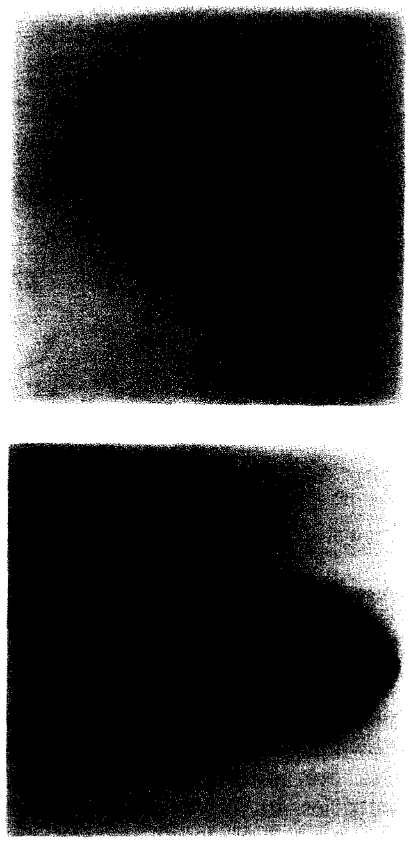


c

Colour Plate 9 Interpolating half-closed RaG surface to vertices of five equal and parallel equilateral triangles results in a cylinder; (a) when areas of upper and lower triangles approach zero, a sphere is obtained, (b) cylinder converges to ellipsoid if, instead of triangles, rhombi are used, (c) egg-shaped object is obtained if the triangles used in *Colour Plate 9a* are not equally spaced



Colour Plate 10 Approximation of wireframe shown in *Figure 11* by a RaG surface
[Each grid point was used as a control point.]



Colour Plate 11 Views of two half-closed RaG surfaces fitting points shown in *Figure 12*

cubic spline interpolation problems' *J. Approx. Theor.* Vol 41 (1984) pp 64–86
10 Hoschek, J 'Intrinsic parametrization for approximation' *Comput. Aided Geom. Des.* Vol 5 (1988) pp 27–31
11 Sarkar, B and Menq, C-H 'Parameter optimization in approximating curves and surfaces to measurement data' *Comput. Aided Geom. Des.* Vol 8 (1988) pp 268–290
12 Farin, G *Curves and Surfaces for Computer Aided Geometric Design* Academic Press, USA (1988) pp 109–117
13 Akima, A 'A new method of interpolation and smooth curve curve fitting based on local procedures' *J. ACM* Vol 17 No 4 (1970) pp 589–602
14 Piegl, L 'On NURBS: a survey' *Comput. Graph. & Applic.* Vol 11 No 1 (1991) pp 55–71
15 Franke, R 'Scattered data interpolation: tests of some methods' *Math. Comput.* Vol 10 (1982) pp 181–220