Virtual Trackball Modeling and the Exponential Map

George Stantchev*
University of Maryland

Abstract

Rotations are among the basic object transformations supported by virtually any modern 3D interactive visualization software. A common technique for rotating 3D objects via 2D mouse movement is the virtual trackball. It relies on a suitably constructed correspondence between the image plane and a sphere of proper radius (trackball) so that a displacement vector representing a mouse move induces a perceptually consistent rotation of the sphere. We propose a new virtual trackball technique based on two notions from Differential Geometry, namely the exponential map and parallel transport along geodesics on a smooth manifold. We develop a robust mathematical framework for the trackball model and we argue that the inherent correspondence between a mouse move and its induced rotation is more intuitive and coherent in comparison to other existing techniques.

Keywords: virtual trackball, 3D rotation, 2D mouse, exponential map, parallel transport, differential geometry

1 Introduction

The idea behind a virtual trackball is the following. Take a sphere S centered at the origin in $\mathbb{R}^3$. Let p be the intersection point of S with the positive z axis. Identify the image plane with the plane T_p tangent to S at p. The display window identifies with a rectangle R in T_p centered at p (see Fig. 2). A mouse move, represented by a starting point Q_s and an end point Q_e in R defines a directed line segment $\overrightarrow{Q_sQ_e}$. In a virtual trackball model, $\overrightarrow{Q_sQ_e}$ is mapped to a directed great circular arc on S, which in turn gives rise to a rotation of S. Thus, a motion of the tangent plane T_p is converted into a motion of the sphere.

A trackball model relies essentially on two steps:

1. Choose a suitable mapping from T_p to S that establishes a one-to-one correspondence between R and a subset of S, usually the upper hemisphere.

2. Use this correspondence to associate a directed line segment in R with a unique rotation of S.

Existing virtual trackball techniques have been recently surveyed be Henriksen et al. [2004]. A mapping typically used in step (1) by existing techniques is orthogonal projection along the z-axis. One model [Bell 1988] uses an orthogonal projection onto a union of the trackball sphere and a suitably matched hyperbolic surface of revolution. For step (2) all surveyed techniques, except [Chen et al. 1988], project the end points of a mouse motion vector onto the trackball using step (1) to obtain a pair of vectors in $\mathbb{R}^3$. The two vectors are normalized and associated with a rotation that carries one unit vector to the other.

*e-mail: gogo@math.umd.edu

We propose the following alternatives to the techniques surveyed in [Henriksen et al. 2004]. For projection we use the *exponential map* $\exp_p$ at p, restricted to an appropriate region in the tangent plane T_p. The exponential map, like orthogonal projection carries straight radial lines through the origin O in T_p onto geodesics through p, which for the sphere are precisely the great circles through p. However distances along each such geodesic are preserved in the following sense. A segment on a straight radial line through O is mapped by the exponential map to an arc of the same length lying on the great circle through p. This is in contrast to orthogonal projection which distorts radial distances further away from the origin in T_p.

For step (2), we use the *differential* of the exponential map to "transport" the directed line segment $\overrightarrow{Q_sQ_e}$ to a vector $\mathbf{w}$ in the tangent space T_{q_s} of S at the point $q_s = \exp_p(Q_s)$. Further, we apply the exponential map $\exp_{q_s}$ to $\mathbf{w}$ to obtain a directed great circular arc through q_s. This arc represents the rotation that we ultimately associate with the mouse move $\overrightarrow{Q_sQ_e}$. The differential of the exponential map is in general a linear isometry of Euclidean vector spaces, in our case of two copies of $\mathbb{R}^2$. In particular, the lengths of $\overrightarrow{Q_sQ_e}$ and $\mathbf{w}$ are the same, and equal to the angle of the rotation associated with $\mathbf{w}$. Moreover, if $\mathbf{w}'$ is a vector in T_{q_s} tangent to the great circular arc through q_s and p, the angle $\measuredangle(\overrightarrow{Q_sQ_e}, \overrightarrow{OQ_s})$ is equal to the angle $\measuredangle(\mathbf{w}, \mathbf{w}')$. Thus in our model a mouse move is completely "in sync" with its induced rotation both in magnitude and direction.

2 Trackball Model Construction

We now elaborate on the construction of our trackball model and discuss some of the relevant mathematical background. We first consider a special case. Assume that a mouse move starts at the origin $O \in T_p$ and ends at Q_e. The exponential map projects $\overrightarrow{OQ_e}$ onto a directed great circular arc C_p through p of length $|OQ_e|$. In this case our trackball model agrees with all other models with respect to the axis of rotation: it is a straight line through the center of S that is perpendicular to the plane of C_p, or equivalently parallel to T_p. This coincidence reflects a common intuition about a physical trackball, namely that a (infinitesimal) rotation of the trackball initiated at a point p leaves invariant a great circular arc through p and thus its axis must be parallel to the tangent plane at p. The rotation angle is equal to the length of the arc C_p, which in our case is equal to $|OQ_e|$. This ensures the "naturality" of trackball control: the difference in length between any two mouse displacements originating at O is the same as the difference between the angles of their induced rotations.

Assume now that a mouse move starts at $Q_s \neq O$ and ends at Q_e. In this case we try to mimic the above situation by first projecting Q_s to the sphere via the exponential map at p. Let $q_s = \exp_p(Q_s)$. Note that here we implicitly identify Q_s with the vector represented by $\overrightarrow{OQ_s}$. Using q_s as a start-

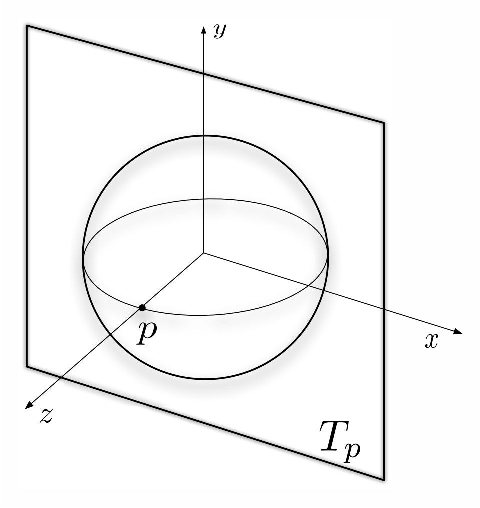

Figure 1: Trackball Model

ing point for the induced rotation we apply the method outlined for the special case above. In particular, we transport the displacement $\overrightarrow{Q_sQ_e}$ to the tangent space T_{q_s} and pretend that the mouse move actually occurred in T_{q_s} starting at the origin.

Recall that for any smooth manifolds M and N, the differential $(df)_p$ of a smooth map $f\colon M \longrightarrow N$ at $p \in M$ is a linear operator that carries the tangent space at $p \in M$ to the tangent space at $f(p) \in N$ in a way consistent with f and the local parametrizations for M and N. In our case the differential of the exponential map

$$(d\exp_p)_{Q_s}\colon T_{Q_s}(T_p) \longrightarrow T_{q_s}$$

is an isomorphism and a linear isometry between the tangent spaces of T_p and S at Q_s and $q_s = \exp_p(Q_s)$ respectively.

The directed line segment $\overrightarrow{Q_sQ_e}$ gives rise to a displacement vector $\mathbf{W} = Q_e - Q_s$ which lives in the tangent space $T_{Q_s}(T_p)$. Let

$$\mathbf{w} = (d\exp_p)_{Q_s}(\mathbf{W})$$

We use $\mathbf{w} \in T_{q_s}$ to generate a rotation of S as if the mouse move occurred in T_{q_s} with initial point at the origin. In particular we apply the exponential map at q_s to project $\mathbf{w}$ onto a directed circular arc that lies on the great circle through q_s tangent to $\mathbf{w}$. The length of this arc and hence the angle of rotation is equal to the length of $\mathbf{w}$ by the length-preserving property of the exponential map.

Let $\mathbf{W}'$ be the vector represented by the directed line segment OQ_s in T_p and let $\mathbf{W}$ be as above. By a Lemma of Gauss (compare [Carmo 1993], p. 69):

$$\left\langle (d\exp_p)_{Q_s}(\mathbf{W}), (d\exp_p)_{Q_s}(\mathbf{W}') \right\rangle = \langle \mathbf{W}, \mathbf{W}' \rangle$$

where $\langle , \rangle$ denotes the standard Euclidean scalar product. In particular the differential of $\exp_p$ is an isometry and thus preserves the angle between $\mathbf{W}'$ and $\mathbf{W}$. This property contributes to the perceptual coherence of our trackball model.

3 Mathematical Details

In this section we provide more mathematical details and explicit formulas for the trackball construction outlined in the previous paragraphs. First, we give a general description of the exponential map. Let $\mathbf{v}$ be a non-zero vector in the tangent space T_p at a point p of a smooth manifold M. There exists (locally) a unique geodesic line γ_v on M, whose tangent vector at p is equal to $\mathbf{v}$, i.e.

$$\gamma_v(0) = p, \qquad \left.\frac{d}{dt}\right|_{t=0} \gamma_v(t) = \mathbf{v} \tag{1}$$

The exponential map at p is the function $\exp_p\colon T_p \to M$, such that

$$\exp_p(\mathbf{v}) = \gamma_v(1)$$

When $\mathbf{v} = \mathbf{0}$, we set $\exp_p(\mathbf{0}) = p$ for all $t \in \mathbb{R}$. The point p is called *the base point* of $\exp_p$. In the case of the sphere, the geodesic γ_v is defined globally and is just the great circle through p that is tangent to $\mathbf{v}$ and parametrized so that (1) holds.

For a compact surface, such as the sphere, the exponential map can never be one-to-one globally, but theory guarantees the existence of a neighborhood of p (called *normal*), for which the map will be one-to-one (and hence a diffeomorphism). For a sphere of radius r, and centered at the origin in $\mathbb{R}^3$, such a neighborhood can be easily computed: if $\mathbf{u}$ is a unit vector in T_p, and $\mathbf{p}$ is the unit vector corresponding to the point p, both viewed as vectors in $\mathbb{R}^3$, the map $\exp_p$ wraps the radial line along $\mathbf{u}$ around the great circle through p to which $\mathbf{u}$ is tangent, i.e.

$$\exp_p(t\mathbf{u}) = r[\cos(t/r)\mathbf{p} + \sin(t/r)\mathbf{u}] \tag{2}$$

It follows that the exponential map is one-to-one for $0 \leq t < \pi r$. In our context, the parameter t represents distance from the display window's center point in screen coordinates. Thus to ensure the injectivity of the exponential map it suffices to take r such that

$$r > \frac{d}{2\pi}$$

where d is the length of the window's diagonal. We use, however, a larger radius for the following reason. Mouse displacement vectors whose starting point is mapped to a point below the equatorial circle do not, in general, induce perceptually intuitive rotations. This observation has been confirmed in the context of other trackball models as well [Henriksen et al. 2004]. One possibility to ensure that no points in the window are mapped by $\exp_p$ to the lower hemisphere is to choose:

$$r = \frac{d}{\pi}$$

In this case the four corners of the window are mapped to the equatorial circle of S and all other points are sent to the upper hemisphere. This however would make it impossible to generate certain rotations, including for instance, a rotation around the z axis, since the only way to induce one would be to start at a point Q_s that maps to the equator and move in direction perpendicular to OQ_s.

As a compromise, we have chosen for our model:

$$r = \frac{w}{\pi}$$

where w is display window's width. In this case the inscribed disk D of radius $w/2$ is mapped onto the upper hemisphere, which guarantees that the full set of rotations of S can be generated. However, the complement of D, i.e. union of the four corners outside of D, is mapped into the lower hemisphere, which as mentioned before, is not desirable. To prevent this, we modify our projection map so that it coincides with $\exp_p$ in D and is equal to the restriction of $\exp_p$ to the boundary ∂D, outside of D. Thus points in the complement of D are mapped always to the equatorial circle of S. This is not unreasonable since from a user's point of view, a mouse move is unlikely to be entirely contained outside of D. Even in case that happens, the induced rotation will be along the equatorial circle, which is close to what is intuitively anticipated.

We now get back to the explicit formula for the exponential map. In the special case when p is on the positive z-axis, $\mathbf{u}$ is a unit vector parallel to the xy-plane, and formula (2) reduces to the following:

$$\exp_p(t\mathbf{u}) = \begin{pmatrix} r\sin(t/r)\mathbf{u}_x \\ r\sin(t/r)\mathbf{u}_y \\ r\cos(t/r) \end{pmatrix} \tag{3}$$

where $\mathbf{u}_x$ and $\mathbf{u}_y$ are respectively the x- and the y-coordinates of $\mathbf{u}$.

Next, we focus on $d\exp_p$. Using the analytic definitions directly to compute an explicit formula for $d\exp_p$ is cumbersome and less intuitive, so we resort to a more geometric argument instead. It is based on the observation that in this context $d\exp_p$ induces *parallel transport* along geodesics through p. In particular, let $\mathbf{V} \in T_p$ be a vector tangent to S at p and let $\mathbf{W}$ be a vector in the tangent space $T_{\mathbf{V}}T_p$. Using the fact that $\mathbb{R}^2 \cong T_p \cong T_{\mathbf{V}}T_p$ and the inherent parallelism of R^2, we consider $\mathbf{W}$ as an element of T_p (simply translate $\mathbf{W}$ to the origin of T_p). Parallel transport along the geodesic $\gamma(t)$ tangent to $\mathbf{V}$ at p gives rise to a vector field $\mathbf{W}(t)$ along γ, such that $\mathbf{W}(0) = \mathbf{W}$ and $\langle \mathbf{W}(t), \gamma'(t)\rangle = \text{const.}$. In particular

$$\big\langle \mathbf{W}(0), \gamma'(0)\big\rangle = \langle \mathbf{W}, \mathbf{V}\rangle = \big\langle \mathbf{W}(1), \gamma'(1)\big\rangle \tag{4}$$

But by definition $\gamma(1) = \exp_p(\mathbf{V})$, and

$$\gamma'(1) = (d\exp_p)_{\mathbf{V}}(\mathbf{V})$$

and hence by Gauss' Lemma, and Equation (4)

$$\big\langle \mathbf{W}(1), (d\exp_p)_{\mathbf{V}}(\mathbf{V})\big\rangle = \big\langle (d\exp_p)_{\mathbf{V}}(\mathbf{W}), (d\exp_p)_{\mathbf{V}}(\mathbf{V})\big\rangle \tag{5}$$

Since the tangent spaces of S are two dimensional, Equation (5) implies that $(d\exp_p)_{\mathbf{V}}(\mathbf{W})$ is collinear with $\mathbf{W}(1)$, i.e. equal up to a scalar. Furthermore, since parallel transport is an orientation preserving isometry of tangent spaces along γ, that scalar must be 1:

$$\mathbf{W}(1) = (d\exp_p)_{\mathbf{V}}(\mathbf{W}) \tag{6}$$

Thus, one way to compute $(d\exp_p)_{\mathbf{V}}(\mathbf{W})$ is to:

a. choose an orthonormal frame $\mathbf{F} = \{\mathbf{v}, \mathbf{v}^{\perp}\}$ in T_p, such that $\mathbf{v}$ is a unit vector collinear with $\mathbf{V}$ and $\mathbf{v}^{\perp}$ is a unit vector perpendicular to $\mathbf{v}$.

b. express $\mathbf{W}$ in the frame $\mathbf{F}$

c. parallel transport $\mathbf{F}$ along the geodesic $\gamma(t)$ into the tangent space at $\exp_p(\mathbf{V})$; the image of $\mathbf{F}$ is a frame $\overline{\mathbf{F}} = \{\overline{\mathbf{v}}, \overline{\mathbf{v}}^{\perp}\}$ in $T_{\exp_p(\mathbf{V})}$ such that $\overline{\mathbf{v}} = \gamma'(1)$ and $\overline{\mathbf{v}}^{\perp}$ is the orthogonal unit vector that completes $\overline{\mathbf{F}}$ to a frame that has the same orientation as $\mathbf{F}$.

d. compute $\mathbf{W}(1)$ in terms of $\overline{\mathbf{F}}$ using the expression for $\mathbf{W}$ in $\mathbf{F}$.

To facilitate computations in steps (b) through (d) we make use of the following observation. The standard embedding of S in $\mathbb{R}^3$ allows us to treat vectors in the tangent spaces of S as vectors in $\mathbb{R}^3$. For instance extending vectors in T_p to $\mathbb{R}^3$ amounts to adding a third coordinate equal to 0. We use the same notation for the extensions of $\mathbf{F}$ and $\overline{\mathbf{F}}$ to $\mathbb{R}^3$. Augment each frame with a third unit vector that is orthogonal to both vectors in it to obtain a new right-oriented orthonormal frame, again denoted, by abuse of notation, with the same symbol. There is an orthogonal 3×3 matrix A, such that:

$$\overline{\mathbf{F}} = A \cdot \mathbf{F}$$

Therefore $\mathbf{W}$ and $\mathbf{W}(1)$ are related by A accordingly:

$$\mathbf{W}(1) = A \cdot \mathbf{W}$$

To compute A, we arrange $\mathbf{F}$ and $\overline{\mathbf{F}}$ in matrix form by listing their elements as column vectors. Let F and $\overline{F}$ be the corresponding matrices. Then:

$$A = \overline{F} \cdot F^{-1} = \overline{F} \cdot F^t \tag{7}$$

where F^t denotes the transpose of F and the second equality follows from the orthogonality of F.

It now remains to give explicit formulas for F and $\overline{F}$ in terms of the data associated with a mouse move. Assume, as above, that we are given an initial point Q_s and a terminal point Q_e in the tangent plane T_p and that p is on the positive z-axis. Let O be the origin in T_p. We have the following recipe:

a. let $\mathbf{V} = (\mathbf{V}_x, \mathbf{V}_y, 0)$ and $\mathbf{W} = (\mathbf{W}_x, \mathbf{W}_y, 0)$ be unit vectors representing respectively $\overrightarrow{OQ_s}$ and $\overrightarrow{Q_sQ_e}$ in $\mathbb{R}^3$. Let $t = \|\overrightarrow{OQ_s}\|$ and let $s = \|\overrightarrow{Q_sQ_e}\|$.

b. the orthonormal frame $\mathbf{F}$ in $\mathbb{R}^3$ is then:

$$\mathbf{F} = \{\mathbf{e}_z, \mathbf{V}, \mathbf{e}_z \times \mathbf{V}\}$$

where $\mathbf{e}_z$ is the unit vector along the z-axis, and $\times$ denotes the standard vector product of $\mathbb{R}^3$. Consequently, in terms of the standard basis of $\mathbb{R}^3$ the matrix F is given by

$$F = \begin{pmatrix} 0 & \mathbf{V}_x & -\mathbf{V}_y \\ 0 & \mathbf{V}_y & \mathbf{V}_x \\ 1 & 0 & 0 \end{pmatrix} \tag{8}$$

c. find the elements of the extended orthonormal frame $\overline{\mathbf{F}}$ at $\exp_p(Q_s)$:

$$\overline{\mathbf{F}} = \{\exp_p(\mathbf{V}), \mathbf{u}, \exp_p(\mathbf{V}) \times \mathbf{u}\}$$

where $\mathbf{u}$ is the unit tangent vector of $\gamma(t)$ at $\exp_p(\mathbf{V})$. Using formula (3), the coordinate expression for $\overline{F}$ is readily computed:

$$\overline{F} = \begin{pmatrix} \sin(t/r)\mathbf{V}_x & \cos(t/r)\mathbf{V}_x & -\mathbf{V}_y \\ \sin(t/r)\mathbf{V}_y & \cos(t/r)\mathbf{V}_y & \mathbf{V}_x \\ \cos(t/r) & -\sin(t/r) & 0 \end{pmatrix} \tag{9}$$

Combining equations (7), (8), and (9) we obtain an explicit formula for A.

Finally, we compute the axis of the rotation induced by $A \cdot \mathbf{W} = (d\exp_p)_{\mathbf{V}}(\mathbf{W})$. To do that it suffices to take the cross product of $A \cdot \mathbf{W}$ and $\exp_p(\mathbf{V})$, since these two vectors span the plane containing the great circle tangent to $A \cdot \mathbf{W}$ at $\exp_p(\mathbf{V})$. From the discussion above, this great circle contains the image of $(d\exp_p)_{\mathbf{V}}(\mathbf{W})$ under the exponential map based at $\exp_p(\mathbf{V})$ and therefore is left invariant under the induced rotation. Thus the axis of rotation is collinear with the vector:

$$(A \cdot \mathbf{W}) \times (\sin(t/r)\mathbf{V}_x, \sin(t/r)\mathbf{V}_y, \cos(t/r))^t$$

The angle of rotation is equal to the length of $\mathbf{W}$, by virtue of the exponential map definition.

4 Conclusion

We have proposed a new virtual trackball technique that is based on a different mathematical framework compared to existing techniques. In particular we use exponential map to project points from the image plane to the virtual trackball. Also, we transport the vector associated with a mouse move along trackball geodesics in order to mimic the basic case of inducing rotation when the initial point of the move is at the origin. We believe that since the exponential map and parallel transport preserve distances and angles in a sense rigorously defined above, rotations computed with our technique are more natural and intuitive for the user. It is our intention to assess the validity of this statement with a robust performance evaluation study based on current HCI methodology.

References

BELL, G., 1988. Bell's trackball. http://www.dispersoid.net/code/trackball.c.

CARMO, M. P. 1993. *Riemannian Geometry*. Birkhäuser Boston.

CHEN, M., MOUNTFORD, S., AND SELLEN, A. 1988. A study in interactive 3-d rotation using 2-d control devices. *Computer Graphics 22*, 4 (August), 121–129.

HENRIKSEN, K., SPORRING, J., AND HORNBÆK, K. 2004. Virtual trackballs revisited. *IEEE Transactions on Visualization and Computer Graphics 10*, 2, 206–216.

SHOEMAKE, K. 1992. Arcball: A user interface for specifying three-dimensional orientation using a mouse. In *Proc. Graphics Interface '92*, 151–156.