Continuous Shape Transformation and Metrics on Shapes

Ernest Davis
Courant Institute
New York, New York
DAVISE@CS.NYU.EDU

April 27, 2000

Abstract

A natural approach to defining continuous change of shape is in terms of a metric that measures the difference between two regions. We consider four such metrics over regions: the Hausdorff distance, the dual-Hausdorff distance, the area of the symmetric difference, and the optimal-homeomorphism metric. Each of these gives a different criterion for continuous change. We establish qualitative properties of all of these; in particular, the continuity of basic functions such as union, intersection, set difference, area, distance, and the boundary function; the transition graph between RCC relations (Randell, Cui, and Cohn, 1992). We discuss the physical significance of these different criteria.

We also show that the history-based definition of continuity proposed by Muller (1998) is equivalent to continuity with respect to the Hausdorff distance. An examination of the difference between the transition rules that we have found for the Hausdorff distance and the transition theorems that Muller derives leads to the conclusion that Muller's analysis of state transitions is not adequate. We propose an alternative characterization of transitions in Muller's first-order language over histories.

1 Introduction

Many physical processes — biological growth, movement of a string, object, inflation of a balloon, bending of a rod, evaporation of a puddle, and so on — involve the continuous change in the shape of an object. The knowledge that the shape is a continuous function of time is an important spatio-temporal constraint in qualitative reasoning about the process.

Previous studies in the AI literature of continuous change of shape have mostly proceeded by postulating desired properties. Randell, Cui, and Cohn (1992) provide constraints on the transitions between the topological (RCC-8) relations that can occur in continuous shape change; for example, if object A is disconnected from object B at time $t1$ and A overlaps B at time $t2 > t1$ then A must be externally connected to B at some time in between. Muller (1998) gives a first-order characterization of continuous change in a topological language of space and time; we will discuss this in detail below (section 6.)

A mathematically more conventional approach is to define a metric over the space of regions and then use the standard epsilon-delta definition of continuity. Such an approach has two advantages. First, standard theorems about continuous functions come for free; for example, that the composition of two continuous functions is continuous. Second, by providing a semantic grounding for continuity, it allows one to prove the correctness of a transition graph like Randell, Cui, and Cohn's, or of an axiomatic characterization like Muller's rather than just positing them.

However, there is no standard metric over spatial regions. Rather, a number of different metrics suitable for different purposes are defined in the literature. For example, Mumford (1991) surveys six different metrics that have been proposed as similarity measures in computer vision. What functions are continuous, and therefore what properties are held by all continuous functions, depend on the metric used. As we shall discuss below, different metric functions on shapes are associated with different physical scenarios, with different methods of obtaining shape information, and with different shape representations.

The purpose of this paper is to describe the qualitative properties of different metrics over regions. The properties that we will consider are:

- The continuity or discontinuity of a number of basic functions: union, set-difference, intersection, area, distance, diameter, in-radius, smoothed circumference, the convex-hull function, and the projection function (section 4).
- The transition graph between binary topological relations (section 5).

Furthermore, we show that the history-based definition of continuity proposed by Muller (1998) is equivalent to continuity with respect to the Hausdorff distance (section 6). An examination of the difference between the transition rules that we have found for the Hausdorff distance and the transition theorems that Muller derives leads to the conclusion that Muller's analysis of state transitions is not adequate. We propose an alternative characterization of transitions in Muller's first-order language over histories.

Determining the continuity properties of different metrics is also relevant to the problem of spatial computing with shape tolerances (e.g. Requicha, 1983; Joskowicz, Sacks, and Srinivasan, 1997; Davis, 1999). If we determine, for instance, that the area of a region is continuous with respect to the metric $\mathrm{d}_A(\mathbf{P}, \mathbf{Q})$ that means that if we have an approximation $\mathbf{Q}$ of shape $\mathbf{P}$ and $\mathrm{d}_A(\mathbf{P}, \mathbf{Q})$ is sufficiently small, then we are justified in using area($\mathbf{Q}$) as an estimate of area($\mathbf{P}$). If no such continuity property holds, then the estimate of the area may be very far off.

The issue has a similar importance in computer vision. A key step in many computer vision system is to match a region found in an image against a model or against a region found in a different image. The criterion of matching can often be posed in terms of closeness relative to some metric on regions. If so, then continuity properties can be important in various ways. For example, if the area function is continuous relative to the matching criterion, then a necessary condition for closeness in the metric is closeness in area, and so a wide discrepancy in area can be used to prune out invalid matches. If the area function is discontinuous, then such pruning is not legitimate.

Throughout this paper, we will take the space of regions to be all normal[1] bounded regions in Euclidean space. Somewhat surprisingly, for the purposes of this paper it makes very little difference whether we require all regions to be connected, or whether we allow disconnected regions. Also, the dimensionality of the space makes very little difference. For convenience of writing and of constructing diagrams, we will mostly speak of regions in the plane, but essentially everything applies, with obvious changes, to regions in three-space or higher dimensions. (One-dimensional space, of course, is rather different.)

Section 2 will present the various metrics on shapes that we consider in this paper. Since our chief intended application is physical reasoning, we will focus on those metrics where the constraint of continuous change has a natural physical interpretation. Section 3 discusses some basic concepts of topology that are used later in the paper. This paper adduces a large number of theorems; however, almost all of them are extremely easy. The few proofs that present any difficulty are

[1] A region is normal if it is equal to the closure of its interior

discussed in appendix A. Appendix B presents the details of the expression of transition rules in Muller's first-order language over histories.

2 Metrics over regions

In this section, we define four different metrics that measure the difference between regions $\mathbf{P}$ and $\mathbf{Q}$.

Definition 1: The *Hausdorff* distance from $\mathbf{A}$ to $\mathbf{B}$ is defined as the maximum of either the maximal distance from a point $\mathbf{p} \in \mathbf{A}$ to $\mathbf{B}$ or the maximal distance from a point $\mathbf{q} \in \mathbf{B}$ to $\mathbf{A}$.

$$\mathrm{d}_H(\mathbf{A}, \mathbf{B}) = \max(\sup_{q \in A} \inf_{p \in B} \mathrm{d}(\mathbf{p}, \mathbf{q}), \sup_{p \in B} \inf_{q \in A} \mathrm{d}(\mathbf{p}, \mathbf{q}))$$

We denote the closure of the complement of region $\mathbf{R}$ as $\mathbf{R}^c$.

Definition 2: The *dual-Hausdorff* distance from $\mathbf{A}$ to $\mathbf{B}$, denoted "$\mathrm{d}_{Hd}(\mathbf{A}, \mathbf{B})$", is the maximum of the Hausdorff distance between $\mathbf{A}$ and $\mathbf{B}$ and the Hausdorff distance between their complements.

$$\mathrm{d}_{Hd}(\mathbf{A}, \mathbf{B}) = \max(\mathrm{d}_H(\mathbf{A}, \mathbf{B}), \mathrm{d}_H(\mathbf{A}^c, \mathbf{B}^c))$$

Example 1: Let $\mathbf{A}$ be the square with vertices $\langle 1, 1\rangle, \langle -1, 1\rangle\langle -1, -1\rangle, \langle 1, -1\rangle$ and let $\mathbf{B}$ be the circle centered at the origin of radius 1.2 (Figure 1). Then the distance from the point $\mathbf{a1} = \langle 1, 1\rangle$ in $\mathbf{A}$ to the closest point $\mathbf{b1} = \langle 0.6\sqrt{2}, 0.6\sqrt{2}\rangle$ in $\mathbf{B}$ is $\sqrt{2} - 1.2 \approx 0.214$. Moreover, this is the greatest distance from any point in $\mathbf{A}$ to the closest point in $\mathbf{B}$. The distance from the point $\mathbf{b2} = \langle 1.2, 0\rangle$ in $\mathbf{B}$ to the nearest point $\mathbf{a2} = \langle 1, 0\rangle$ in $\mathbf{A}$ is 0.2. Moreover, this is the greatest distance from any point in $\mathbf{B}$ to the nearest point in $\mathbf{A}$. Therefore, the Hausdorff distance between $\mathbf{A}$ and $\mathbf{B}$, $\mathrm{d}_H(\mathbf{A}, \mathbf{B}) = \max(0.214, 0.2) = 0.214$.

The Hausdorff distance between $\mathbf{A}^c$ and $\mathbf{B}^c$ is computed as follows: The distance from the point $\mathbf{b1} = \langle 0.6\sqrt{2}, 0.6\sqrt{2}\rangle$ in $\mathbf{B}^c$ to the closest point $\mathbf{a3} = \langle 1, 0.6\sqrt{2}\rangle$ in $\mathbf{A}^c$ is equal to $1 - 0.6\sqrt{2} = 0.151$. Moreover, this is the greatest distance from any point in $\mathbf{B}^c$ to $\mathbf{A}^c$. The distance from the point $\mathbf{a2} = \langle 1, 0\rangle$ in $\mathbf{A}^c$ to the point $\mathbf{b2} = \langle 1.2, 0\rangle$ in $\mathbf{B}^c$ is 0.2. Moreover, this is the greatest distance from any point in $\mathbf{A}^c$ to $\mathbf{B}^c$. Thus, the Hausdorff distance from $\mathbf{A}^c$ to $\mathbf{B}^c$, $\mathrm{d}_H(\mathbf{A}^c, \mathbf{B}^c) = \max(0.151, 0.2) = 0.2$.

The dual-Hausdorff distance from $\mathbf{A}$ to $\mathbf{B}$, $\mathrm{d}_{Hd}(\mathbf{A}, \mathbf{B}) = \max(\mathrm{d}_H(\mathbf{A}, \mathbf{B}), \mathrm{d}_H(\mathbf{A}^c, \mathbf{B}^c)) = 0.214$.

Example 2: Figure 2 illustrates the difference between the Hausdorff distance and the dual Hausdorff distance. In the figure on the left, let $\mathbf{P}$ be the square and let $\mathbf{Q}$ be the union of all the small circles. Then the Hausdorff distance between $\mathbf{P}$ and $\mathbf{Q}$, $\mathrm{d}_H(\mathbf{P}, \mathbf{Q})$ is equal to half the distance between two consecutive circles on a diagonal. The midpoint of any such diagonal is the point in $\mathbf{P}$ that is furthest from $\mathbf{Q}$. However, the dual Hausdorff distance between $\mathbf{P}$ and $\mathbf{Q}$, $\mathrm{d}_{Hd}(\mathbf{P}, \mathbf{Q})$ is equal to half the width of the large square; the center of the square is in $\mathbf{Q}^c$ but no point in $\mathbf{P}^c$ is closer than the midpoint of sides of the square.

The right-hand part of figure 2 shows that the same kind of thing can happen even if regions are required to be connected. Let $\mathbf{P}$ be the large rectangle, and let $\mathbf{Q}$ be the thin snaky region. Then the Hausdorff distance between them, $\mathrm{d}_H(\mathbf{P}, \mathbf{Q})$, is half the distance between vertical columns of the snake; every point in $\mathbf{P}$ is at most that distance from the nearest column of $\mathbf{Q}$. However, the dual-Hausdorff distance is again half the height of $\mathbf{P}$.

These examples, with variants, will serve as examples for almost all the differences between properties of the Hausdorff distance and the dual-Hausdorff distance mentioned in this paper.

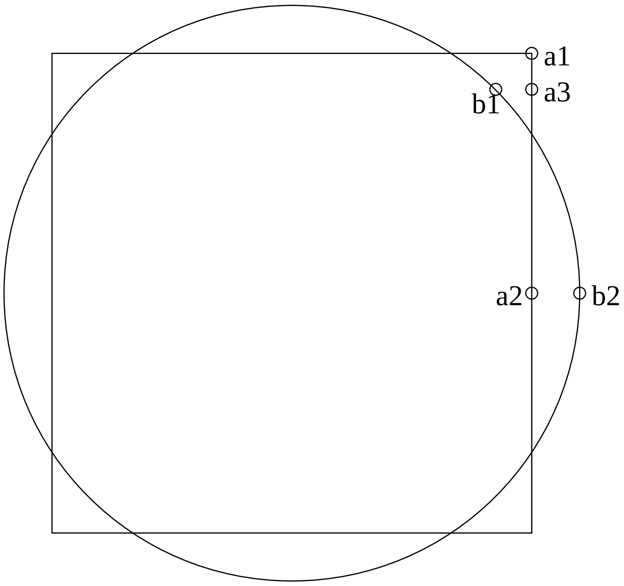


Figure 1: Hausdorff and dual-Hausdorff distances: Example 1

The Hausdorff distance is well known in the literature. The dual-Hausdorff distance was introduced in (Davis, 1999).

The constraint that shapes change continuously relative to the Hausdorff distance corresponds to physical scenarios such as the region occupied by a quantity of gas in a vacuum. The Hausdorff distance between the regions occupied by the gas at time T1 and T2 corresponds to the maximum distance between the position of a molecule at T1 and its position at T2 (more precisely, the Hausdorff distance is the minimal possible value of the latter, over all possible ways of rearranging the molecules between the two scenarios). Therefore, the physical constraint that each molecule moves continuously corresponds to the constraint that the region occupied by the gas changes continuously relative to the Hausdorff distance.

Similarly, the constraint that shapes change continuously relative to the dual-Hausdorff distance corresponds to physical scenarios such as the region occupied by bubbles of gas inside a liquid. Since

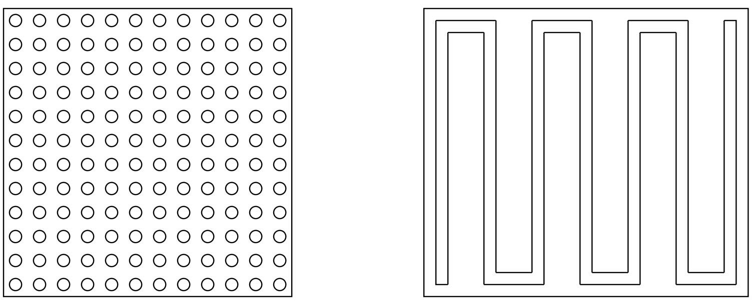


Figure 2: Hausdorff and dual-Hausdorff distances

both the molecules of the gas and the molecules of the liquid must move continuously, both the region occupied by the gas and the region occupied by the liquid must change continuously in the Hausdorff distance. Since the region occupied by the liquid is the complement of the region occupied by the gas, the region occupied by the gas changes continuously in the dual-Hausdorff distance.

Definition 3: The metric $\mathrm{d}_A(\mathbf{P}, \mathbf{Q})$ is defined as the area of the symmetric difference $(\mathbf{P} - \mathbf{Q}) \cup (\mathbf{Q} - \mathbf{P})$. For example, in figure 1, the symmetric difference between **A** and **B** consists of the four corners of the square that lie outside the circle together with the four "sides" of the circle that lie outside the square. The area of this region is 0.92.

The following physical scenario gives rise to continuous change in the metric $\mathrm{d}_A(\mathbf{P}, \mathbf{Q})$: Imagine that it has been raining over an uneven parking lot, so that the lot is now full of puddles. The rain now stops and the puddles gradually evaporate. Let $\mathbf{R}(t)$ be the region occupied by all the puddles at time t (either the two- or the three-dimensional region.) Then $\mathbf{R}(t)$ changes continuously relative to the metric $\mathrm{d}_A(\mathbf{P}, \mathbf{Q})$.

The metric $\mathrm{d}_A(\mathbf{P}, \mathbf{Q})$ also corresponds to the following method of evaluating the difference between two regions: Fix a standard large region **O** containing all the regions of interest, and sample points at random within **O**. For a fixed sample size, the number of points which differ on **P** and **Q** — that is, either lie in **P** and not in **Q** or vice versa — is proportional to $\mathrm{d}_A(\mathbf{P}, \mathbf{Q})$.

The metric $\mathrm{d}_A(\mathbf{P}, \mathbf{Q})$ is very easy to compute in a bit vector representation; it is simply the number of pixels in **P** XOR **Q**.

Definition 4: The *optimal homeomorphism* metric between **P** and **Q**, denoted $\mathrm{d}_O(\mathbf{P}, \mathbf{Q})$, is defined as follows: If σ is a homeomorphism[2] from the plane to itself, we define $c(\sigma)$, the cost of σ, to be the least upper bound of $\mathrm{d}(\mathbf{x}, \sigma(\mathbf{x}))$ over all points $\mathbf{x}$ in the plane. Then $\mathrm{d}_O(\mathbf{P}, \mathbf{Q})$ is defined as the minimum value of $c(\sigma)$ over all homeomorphisms σ over the plane such that $\sigma(\mathbf{P}) = \mathbf{Q}$.

Physically, continuous motion with respect to the metric d_O corresponds to the following scenario: Draw the region **P** on a large, transparent rubber sheet, and draw **Q** on a table. Now consider methods for continuous deforming the sheet without tearing or folding it so that **P** lies on top of **Q**. The "best" such method is to be considered the method that moves the points in the sheet as little as possible.

Figure 3 illustrates the difference between the dual-Hausdorff distance and the optimal-homeomorphism distance. Let **P** be the inner rectangle, and let **Q** be the outer figure, consisiting of a rectangle and a peninsula. The dual-Hausdorff distance between regions **P** and **Q** is equal to the distance from **a** to **b**. Every point in **P** is within $\mathrm{d}(\mathbf{a}, \mathbf{b})$ of a point in **Q** and vice versa, and every point in $\mathbf{P}^c$ is within $\mathrm{d}(\mathbf{a}, \mathbf{b})$ of a point in $\mathbf{Q}^c$, and vice versa. On the other hand, the optimal homeomorphism distance from **P** to **Q** is equal to the distance from **m** to **b**. The optimal homeomorphism associates the whole "peninsula" of **Q** with a small neighborhood of the point **m** in **P**, and the whole "inlet" of $\mathbf{Q}^c$ with a small neighborhood of **m** in $\mathbf{P}^c$.

The optimal-homeomorphism distance has two serious defects as a metric. The first is that if **P** and **Q** are not homeomorphic then $\mathrm{d}_O(\mathbf{P}, \mathbf{Q}) = \infty$, so it gives no measure of greater and lesser similarity among non-homeomorphic pairs of regions. The second is that it is not at all clear how it can be computed in general.

Theorem 1: The four functions d_H, d_{Hd}, d_A, and d_O are all metrics over the space of bounded regular regions.

The proof is straightforward.

[2] A homeomorphism is a continuous one-to-one function whose inverse is also continuous.

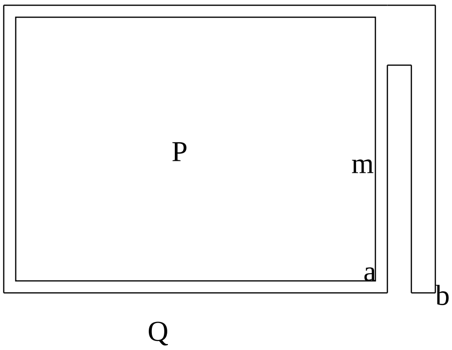


Figure 3: Dual-Hausdorff distance and optimal homeomorphism distance

3 Comparative topologies

We now discuss the topologies that these four metrics induce over the space of regular regions. As this space is rather abstract and likely to be unfamiliar, it will be helpful to review some basic definitions from point-set topology:[3]

Definition 5: A *topology* over a space $\mathcal{S}$ is a collection $\mathcal{O}$ of subsets of $\mathcal{S}$ with the following four properties:

- $\mathcal{S}$ is an element of $\mathcal{O}$.
- The empty set is an element of $\mathcal{O}$.
- If $\mathcal{T}$ is a subcollection of $\mathcal{O}$, then the union of the sets in $\mathcal{T}$ is an element of $\mathcal{O}$.
- If P and Q are elements of $\mathcal{O}$ then $P \cap Q$ is an element of $\mathcal{O}$.

The elements of $\mathcal{O}$ are called the *open sets* in the topology. Set P is said to be *closed* in the topology if $\mathcal{S} - P$ is open.

Definition 6: Let P, Q be subsets of $\mathcal{S}$. The *interior* of P is the union of all open subsets of P. The *closure* of P is the intersection of all closed supersets of P. P is said to be *dense* in Q if Q is a subset of the closure of P.

Definition 7: Let μ is a metric over a space $\mathcal{S}$, let $x \in \mathcal{S}$ and let $\epsilon > 0$. The *open ball* of radius ϵ around x, denoted $\mathrm{B}_\mu(x, \epsilon)$ is the set of all points in $\mathcal{S}$ within ϵ of x.

$$\mathrm{B}_\mu(x, \epsilon) = \{y \in \mathcal{S} \mid \mu(y, x) < \epsilon\}$$

If μ is a metric over a space $\mathcal{S}$, then the topology associated with μ is defined as follows: A set O is open relative to μ if, for every point $x \in S$ there exists an $\epsilon > 0$ such that $\mathrm{B}_\mu(x, \epsilon) \subset O$.

Given two different topologies $\mathcal{O}$ and $\mathcal{U}$ over the same set $\mathcal{S}$, we say that $\mathcal{O}$ is *finer* than $\mathcal{U}$ if every open set in $\mathcal{O}$ is also open in $\mathcal{U}$. We say that $\mathcal{O}$ is *strictly finer* than $\mathcal{U}$ if $\mathcal{O}$ is finer than $\mathcal{U}$ but $\mathcal{U}$ is not finer than $\mathcal{O}$.

Figure 4 shows the comparative fineness of the topologies generated by our four metrics. That is, d_A and d_H are incomparable; they are both coarser than d_{Hd} which is coarser than d_O.

[3] We have already used these terms in connection with the topology of Euclidean space; however, that is a much simpler and more familiar context.

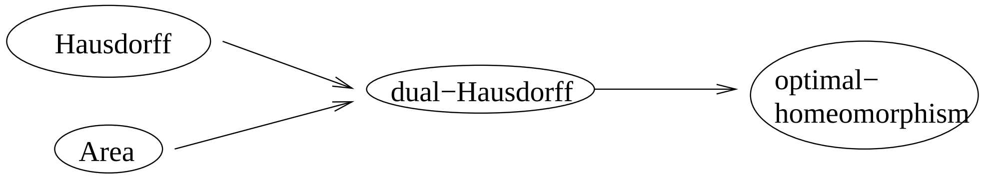


Figure 4: Comparative fineness of metrics

The following well-known theorems will be useful:

Lemma 2: Let μ and η be two different metrics over the same space $\mathcal{S}$. The topology generated by μ is finer than the topology generated by η if and only if the following condition holds: For any point $x \in \mathcal{S}$ and any infinite sequence $y_1, y_2 \ldots$ in $\mathcal{S}$, if the sequence of values $\mu(y_1, x), \mu(y_2, x) \ldots$ converges to 0, then the sequence $\eta(y_1, x), \eta(y_2, x) \ldots$ also converges to 0.

Lemma 3: Let μ and η be two metrics over the same space $\mathcal{S}$ and suppose that the topology generated by μ is finer than the topology generated by η. Then

- Let $f(t)$ be a function from the real line to $\mathcal{S}$. If f is continuous relative to μ then f is also continuous relative to η.
- Let $g(x)$ be a function from $\mathcal{S}$ to the real line. If g is continuous relative to η then g is also continuous relative to μ.

Lemma 3 above enables us to use figure 4 to carry over continuity results from one metric to another. For example, if we show that the distance function is continuous relative to d_H, then it follows immediately that it is continuous relative to d_{Hd} and d_O.

Finally, it will be useful to distinguish two particular kinds of discontinuity:

Definition 8: A function f over space $\mathcal{S}$ is *discontinuous everywhere* if it is discontinuous at every point of $\mathcal{S}$. For example, the function "diameter(**R**)", defined as the maximum distance between any two points in **R**, is discontinuous everywhere with respect to the metric d_A; for any region **R**, there are regions, consisting of **R** plus small region at a considerable distance away, that differ from **R** by an arbitrarily small amount in terms of the area of the symmetric difference, but have a diameter which is arbitrarily larger.

In the more familiar venue of functions from $\Re^k$ to $\Re^m$, everywhere discontinuous functions are generally considered pathological. However, as we shall see, in the context of functions over regions, they are entirely standard.

Definition 9: A function f over space $\mathcal{S}$ is *continuous almost everywhere*[4] if it is continuous over a dense, open set in $\mathcal{S}$.

If a function is not continuous, not almost everywhere continuous, and not everywhere discontinuous, we will say that it is "sometimes" discontinuous.

We will use the notations "Int(**R**)" and "Bd(**R**)" to mean the interior and boundary of region **R**.

[4] Note that this definition does not require that $\mathcal{S}$ be a measure space. It is not easy to define a reasonable measure over the space of regions. If $\mathcal{S}$ is a measure space, then this condition is sufficient, though not necessary, for the condition, "continuous except over a set of measure zero."

P DC **Q**	$\mathbf{P} \cap \mathbf{Q}$ is empty.
P EC **Q**	Int($\mathbf{P}$) $\cap$ Int($\mathbf{Q}$) is empty; Bd($\mathbf{P}$) $\cap$ Bd($\mathbf{Q}$) is non-empty.
P OV **Q**	Int($\mathbf{P}$) $\cap$ Int($\mathbf{Q}$) is non-empty; Int($\mathbf{P}$) $-$ Int($\mathbf{Q}$) is non-empty; and Int($\mathbf{Q}$) $-$ Int($\mathbf{P}$) is non-empty.
P EQ **Q**	$\mathbf{P} = \mathbf{Q}$.
P TPP **Q**	$\mathbf{P} \subset \mathbf{Q}$; $\mathbf{P} \neq \mathbf{Q}$; and Bd($\mathbf{P}$) $\cap$ Bd($\mathbf{Q}$) is non-empty.
P NTPP **Q**	$\mathbf{P} \subset \mathbf{Q}$; and Bd($\mathbf{P}$) $\cap$ Bd($\mathbf{Q}$) is empty.

Table 1: RCC-8 relations

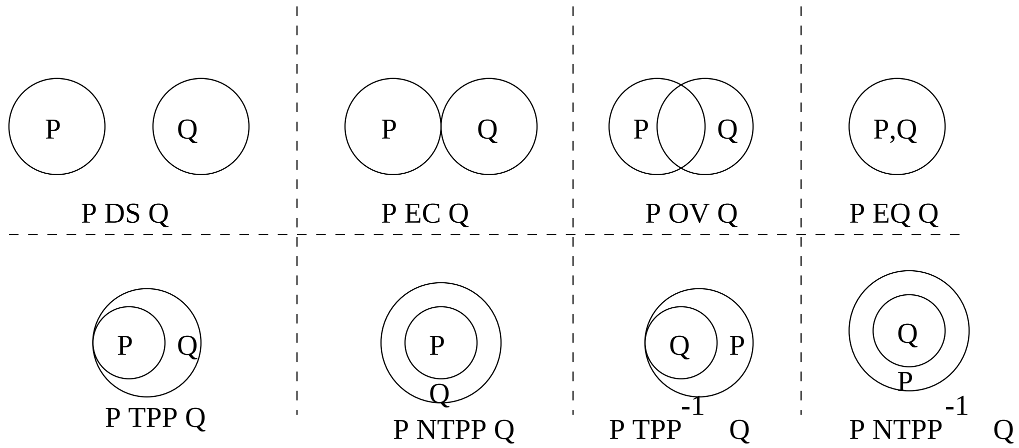


Figure 5: The RCC-8 relations

Finally, as we shall make repeated use in this paper of the RCC-8 relations (Randell, Cui, and Cohn, 1992), we define them in table 1 and illustrate them in figure 5.

4 The continuity of some basic functions

We now proceed to determine the continuity or discontinuity of some basic spatial functions under the various metrics above.

4.1 Some changes that must be continuous and some that must be discontinuous

There are certain types of change that, it seems reasonable to say, should be always be considered as continuous. In particular, a gradual translation, rotation, or change of scale that is continuous in the controlling parameter, should be considered as a continuous change by *any* reasonable theory of continuous change of shape. This indeed holds in all of the four metrics above.

A natural generalization of this condition is the following rule:

Rule 1: Let $L(t)$ be a continuous function from time t to the space of non-singular linear transformations. Let $\mathbf{P}$ be a bounded, regular region. Then the function of t, apply$(L(t), \mathbf{P})$ is a continuous function from time to regions.

This includes continuous translations, rotations, and skewings (affine transformations).

At the other extreme, it seems clear that a change that involves an entire circle or other open region appearing out of nowhere or disappearing into nowhere must be considered discontinuous. At least, if this is not discontinuous, it is hard to see what would be discontinuous.

Rule 2: Let $\mathbf{O}$ be an open region, and let $L(t)$ be a function from time t to the space of regions. If $L(t)$ is disjoint from $\mathbf{O}$ for all $t < 0$ and $L(t)$ contains $\mathbf{O}$ for all $t > 0$, then L is discontinuous at $t = 0$.

It is easily seen that rules 1 and 2 holds in all four of our metrics. (This is not as trivial an observation as it seems. For instance, if unbounded regions are admitted, then rotations and scalings are discontinuous in all our metrics, and translations are discountinuous relative to the metric d_A.)

4.2 Measures

We now consider the continuity of a number of real-valued functions over the space of regions.

The function area$(\mathbf{R})$ is continuous under the metrics d_A, d_{Hd} and d_O. It is everywhere discontinuous under the metric d_H.

Figure 2 illustrates the discontinuity of the area function under d_H. Clearly, by making the dots smaller and smaller and closer and closer together, or making the snake narrower and narrower and its bands closer and closer together, the Hausdorff distance between $\mathbf{P}$ and $\mathbf{Q}$ may be made arbitrarily small, even while the area of $\mathbf{Q}$ approaches zero.

The functions distance$(\mathbf{P}, \mathbf{Q})$ and diameter$(\mathbf{P})$ are continuous under the metrics d_H, d_{Hd} and d_O. They are everywhere discontinuous under the metric d_A.

The radius of the largest inscribed circle in $\mathbf{P}$ is continuous under the metrics d_{Hd} and d_O. It is everywhere discontinous under the metrics d_A and d_H.

The circumference of $\mathbf{P}$ is not continuous under any of these metrics, since it can always be made arbirtrarily long by adding sufficiently many, arbitrarily small, notches on the boundary. However, there is a smoothed version of the circumference that is continuous under the metric d_O.

Definition 10: Let ϕ be a simple curve in the plane, and let $\Delta > 0$ be a distance. Let $A(\phi, \Delta)$ be the set of all paths ψ such that $\mathrm{d}_O(\phi, \psi) \leq \Delta$. We define the Δ-smoothed length of ϕ to be the greatest lower bound over the arc-length of paths in $A(\phi, \Delta)$.

$$\mathrm{smooth}(\phi, \Delta) = \inf_{\psi \in A(\phi,\Delta)} \mathrm{length}(\psi).$$

(One might ask, why use the metric d_O in definition 10 rather than d_H? The answer is that the corresponding function defined with d_H is not at all well behaved. It is not continuous under Δ and it is not continuous even under uniform expansion of the curve ϕ.)

Definition 11: The Δ-*smoothed circumference* of region $\mathbf{P}$ is the sum of the Δ-smoothed lengths of the boundaries of $\mathbf{P}$.

Theorem 4: Let $\mathbf{P}$ range over the space of bounded, regular regions, with finitely many, piecewise smooth boundaries. For $\Delta > 0$, the Δ-smoothed circumference of $\mathbf{P}$ is a continuous function of $\mathbf{P}$ under the metric d_O. It is everywhere discontinuous under the metrics d_A, d_H, or d_{Hd}.

A similar result applies to the length of paths within the region.

Definition 12: Let **R** be a connected normal region and let $\Delta > 0$ be a distance. For any two points **x** and **y** in **R**, define the Δ-smoothed path-distance from **x** to **y** through **R** to be the greatest lower bound of the Δ-smoothed length of ϕ over all curves ϕ such that $\mathbf{x}, \mathbf{y} \in \phi \subset \mathbf{R}$. Define the Δ-*smoothed path diameter* to be the maximum over all $\mathbf{x}, \mathbf{y}$ of the Δ-smoothed path-distance from **x** to **y** through **R**.

Theorem 5: For $\Delta > 0$, the Δ-smoothed path diameter of **P** is a continuous function of **P** under the metric d_O. It is everywhere discontinuous under the metrics d_A, d_H, or d_{Hd}.

4.3 Functions from regions to regions

Definition 13: Let **Z** be any point set. The partial function "Norm(**Z**)", called the non-null normalization of **Z**, is defined as the closure of the interior of **Z**, if this is non-empty; else it is undefined. The partial function NormInt(**P**, **Q**) is defined as Norm($\mathbf{P} \cap \mathbf{Q}$). The partial function NormDiff(**P**, **Q**) is defined as Norm($\mathbf{P} - \mathbf{Q}$).

The union function $\mathbf{P} \cup \mathbf{Q}$ is continuous with respect to d_A and d_H. The functions NormInt(**P**, **Q**) and NormDiff(**P**, **Q**) are continuous with respect to d_A and discontinuous with respect to d_H everywhere in their domain. All three functions are almost everywhere continuous with respect to d_{Hd}. Specifically, $\mathbf{P} \cup \mathbf{Q}$ is discontinuous with respect to d_{Hd} at a pair of regions **P** and **Q** just if there is a point $\mathbf{x} \in \text{Bd}(\mathbf{P}) \cap \text{Bd}(\mathbf{Q}) \cap \text{Int}(\mathbf{P} \cup \mathbf{Q})$. NormInt(**P**, **Q**) is discontinuous at **P**, **Q** just if there exists a point $\mathbf{x} \in \text{Bd}(\mathbf{P}) \cap \text{Bd}(\mathbf{Q})$ that is not in NormInt(**P**, **Q**). NormDiff(**P**, **Q**) is discontinuous at **P**, **Q** just if there exists a point $\mathbf{x} \in \text{Bd}(\mathbf{P}) \cap \text{Bd}(\mathbf{Q})$ that is not in NormDiff(**P**, **Q**). (Figure 6). All three functions are sometimes discontinuous with respect to d_O; specifically, they are discontinuous unless the boundaries of **P** and **Q** are disjoint.

The convex-hull function is continuous with respect to d_H, d_{Hd} and d_O but not with respect to d_A.

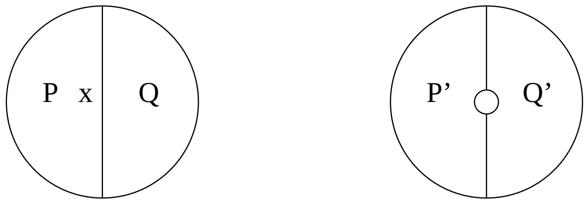
Projection functions, from 3-space to the plane or from the plane to the line, are continuous with respect to d_H. They are everywhere discontinuous with respect to d_A and d_O. They are almost everywhere continuous with respect to d_{Hd}. Specifically, let Π be a projection from space $\mathcal{S}$ to space $\mathcal{T}$ and let **Q** be a region in $\mathcal{S}$. Then Π is discontinuous at **Q** relative to d_{Hd} just if there exists a line **L** in $\mathcal{S}$ such that $\Pi(\mathbf{L})$ is a single point in $\text{Int}(\Pi(\mathbf{Q}))$ and **L** is disjoint from Int(**Q**) (Figure 7).

Table 2 summarizes the results from this section.

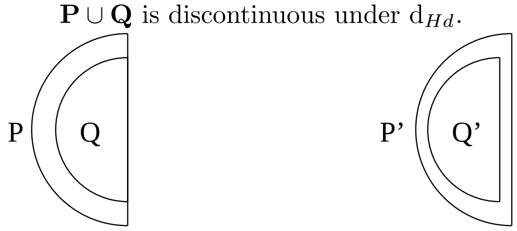
5 Transition networks

Randell, Cui, and Cohn (1992) presented the transition network shown in figure 8 for the RCC-8 relations. Two relations X and Y in this network are connected by an arc if it is possible for two continuous region-valued fluents $\mathbf{P}(t)$ and $\mathbf{Q}(t)$ to transition from the relation X to the relation Y without going through any intermediate relations. Galton (1993) observes, in effect, that any definition of continuous motion satisfying rule 1 must include all these transitions. Similar transition networks have been developed for other systems of binary topological relations [Cite].

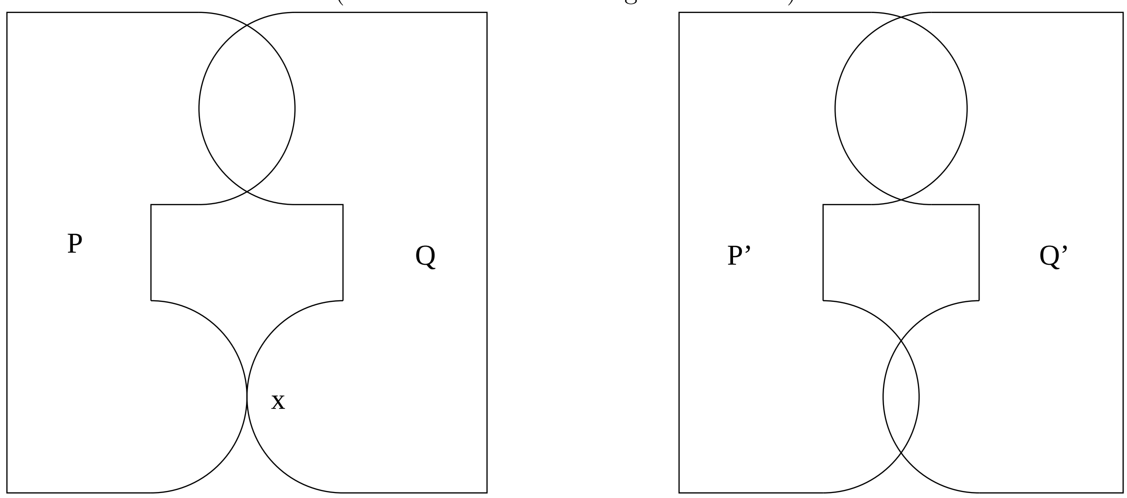
Galton (1995) extends the notation of figure 8 by changing the undirected edges to directed arcs. An arc from relation X to Y means that there are continuous functions $\mathbf{f}(t)$ and $\mathbf{g}(t)$ from time to the space of regions such that $\mathbf{f}(0)$ and $\mathbf{g}(0)$ are related by relation X and $\mathbf{f}(t)$ and $\mathbf{g}(t)$ are related by relation Y for $t > 0$. That is, two regions that are in relation X at one time can immediately change to relation Y. Galton's terminology is that relation X "dominates" relation Y (perhaps not



$\mathbf{P} \cup \mathbf{Q}$ is discontinuous under d_{Hd}.



NormDiff($\mathbf{P}, \mathbf{Q}$) is discontinuous under d_H.
(Note: P is the entire large semi-circle.)



NormInt($\mathbf{P}, \mathbf{Q}$) is discontinuous under d_H.

Figure 6: Discontinuities of Boolean functions

	d_A	d_H	d_{Hd}	d_O
area	Cont.	Discont.	Cont.	Cont.
distance, diameter	Discont.	Cont.	Cont.	Cont.
in-radius	Discont.	Discont.	Cont.	Cont.
Δ-smoothed circumference, path diameter	Discont.	Discont.	Discont.	Cont.
Union, NormInt, NormDist	Cont	Discont.	a.e. Cont	Sometimes
convex hull	Discont.	Cont.	Cont.	Cont.
projection	Discont.	Cont.	a.e. Cont.	Discont.

Table 2: Continuity of some basic functions

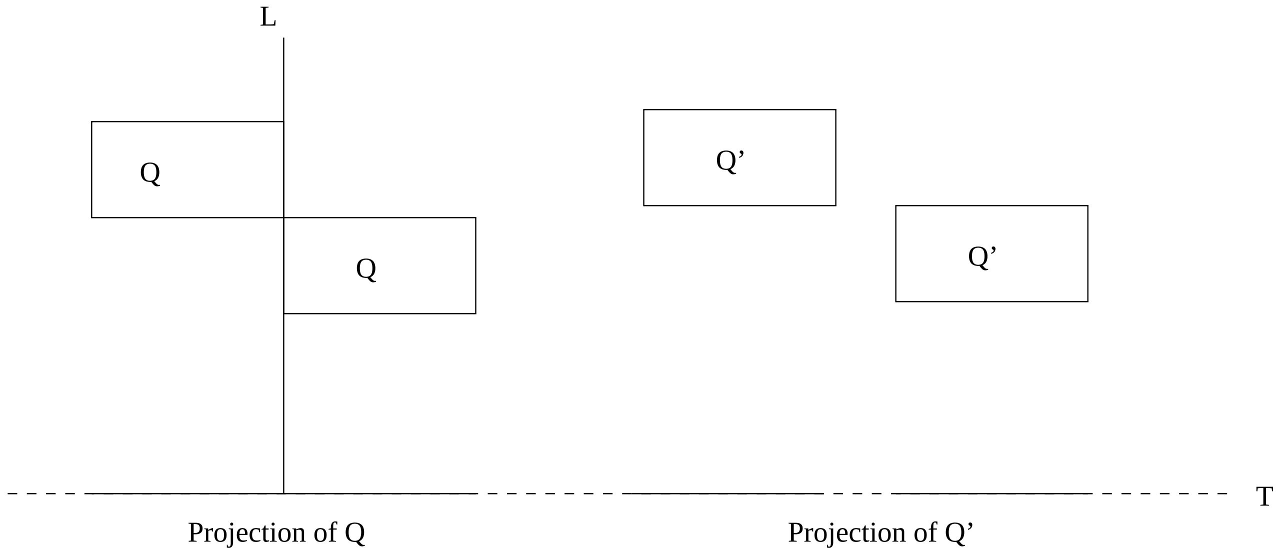


Figure 7: Discontinuity of projection function

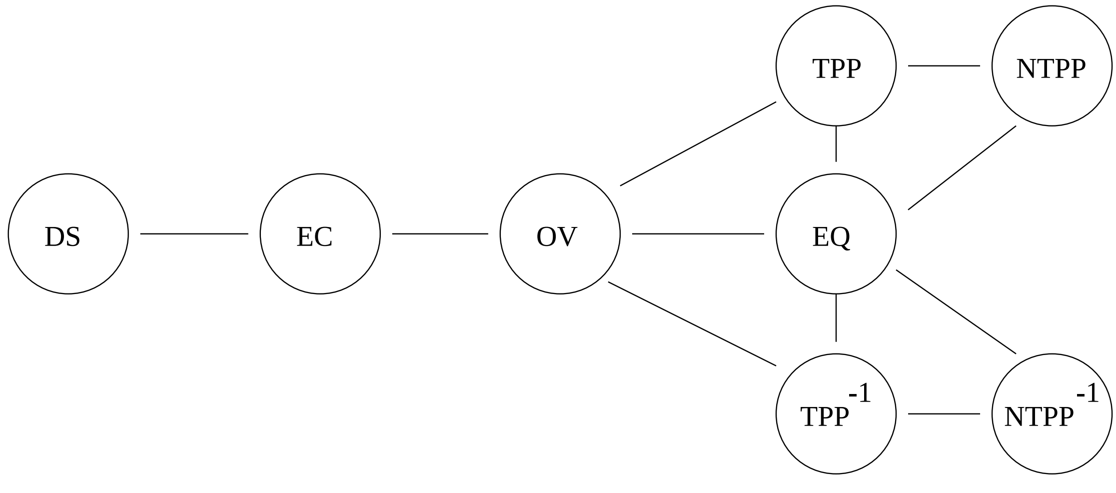


Figure 8: Undirected transition graph from (Randell, Cui, and Cohn, 1992)

the most suggestive of terms). Such a distinction between properties that can change immediately and those that require a finite duration to change correspond to the QSIM rule of "ϵ-transition." (Kuipers, 1986)

An alternative, more abstract, interpretation[5] of these arcs is as follows: Let $\mathcal{S}$ be the space of normal regions in the plane. A binary relation X can be considered as a subset of $\mathcal{S} \times \mathcal{S}$. There is an arc from relation X to Y if X intersects the boundary of Y within $\mathcal{S} \times \mathcal{S}$.

The transition networks for the RCC-8 relations for the topologies d_A, d_H, d_{Hd}, and d_O are displayed in Figures 9, 10, 11, and 11, respectively. (Topologies d_{Hd} and d_O have the same transitions.) The last of these is the one given by Galton (1995).

Some observations about these transition graphs:

In figure 7, the significance of the arrow from the dashed circle on the right to the dashed circle on the left is that *every* relation on the right can undergo a transition to *any* relation on the left. That is, there should be a arrow from each of the five states on the right to each of the three states on the left; however, we have summarized these in terms of the dashed circles, in order to simplify the diagram.

As these transitions are not obvious and somewhat counter-intuitive, we will give an example of the transition from EQ to DC; the other transitions are analogous. The example is similar to the right hand of figure 2. For any $t \in (0, 1))$, define the curve $\phi(t) = \{\langle u, \sin(u/t)\rangle | u \in [0, 1]\}$. Thus, $\phi(t)$ is a sine curve that ranges in the y-direction between -1 and 1, and in the x-direction between 0 and 1, with wavelength $2\pi t$. Let $\psi(t)$ be parallel to $\phi(t)$ but shifted down a distance t; that is, $\psi(t) = \{\langle u, \sin(u/t) - t\rangle | u \in [0, 1]\}$. Let $\delta(t)$ be the mininum distance between $\phi(t)$ and $\psi(t)$. Let $\mathbf{f}(t)$ be the region of points within distance $\delta/3$ of $\phi(t)$ and let $\mathbf{g}(t)$ be the region of points within distance $\delta/3$ of $\psi(t)$. Finally, let $\mathbf{f}(0) = \mathbf{g}(0)$ be the rectangle $[0, 1] \times [-1, 1]$. Then it is immediate that $\mathbf{f}(t)$ DC $\mathbf{g}(t)$ for all $t > 0$; that $\mathbf{f}(0) = \mathbf{g}(0)$; and that both $\mathbf{f}$ and $\mathbf{g}$ are continuous relative to the metric d_H.

The reader may object that an example such as the previous one is obviously just a mathematical pathology, without real-world significance. Consider, however, the following sequence of examples:

First, consider a large marching band spread uniformly (either in a regular pattern or randomly) in a rectangle. One says that the shape of the band is the rectangle.

Second, imagine that the marching band imitates, as far as possible, the region function $\mathbf{f}(t)$ above, in backwards time. That is, the band starts out forming a thick S shape with two bends, and then gradually adds more and more bends spaced closer and closer, the curve gradually becoming thinner and thinner. Eventually, the band is uniformly spread through the rectangle. Thus, for a while, the band has a snake shape; later, its shape is the rectangle; and there is no other kind of shape that it occupies at any intermediate time.

Third, imagine that there are two such bands, that start in S shapes that are initially well separated, but the S's move closer together as they develop. Then, as long as the bands are viewed as forming S's, their shapes are disjoint; once they are viewed as forming rectangles, then they each form the same rectangle. Thus, the shapes of the bands transition from DC to EQ with no intermediate relations.

Finally, rather than think about two marching bands composed of people, which is a rather unimportant case, think about two gasses composed of molecules. The two gasses start out in separate swirls, that grow thinner and thinner, and denser and denser, until each of the gasses fills

[5]Strictly speaking, this interpretation is logically weaker than the one in the previous paragraph; that is, there could be relations that satisfy the property here but do not satisfy the property of the previous paragraph. However, this distinction does not arise with the relations and topologies that we are considering.

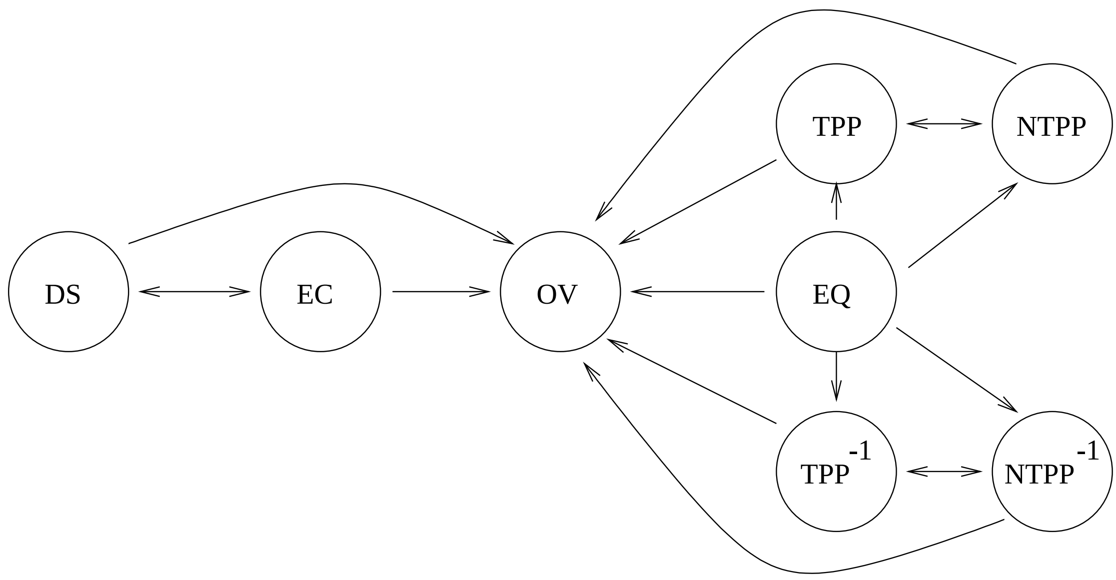


Figure 9: Transition graph for d_A

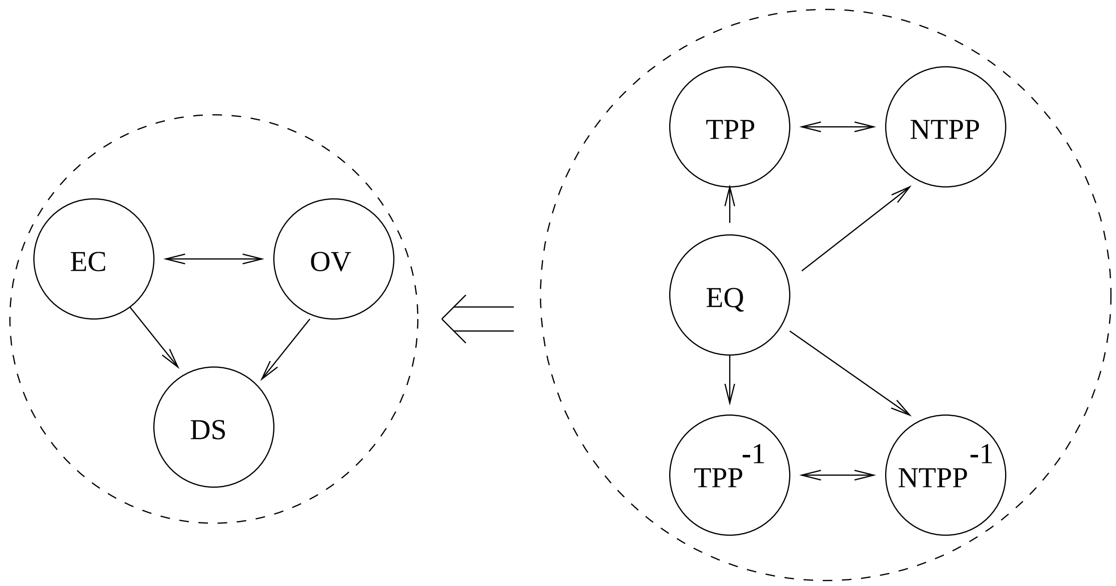


Figure 10: Transition graph for d_H

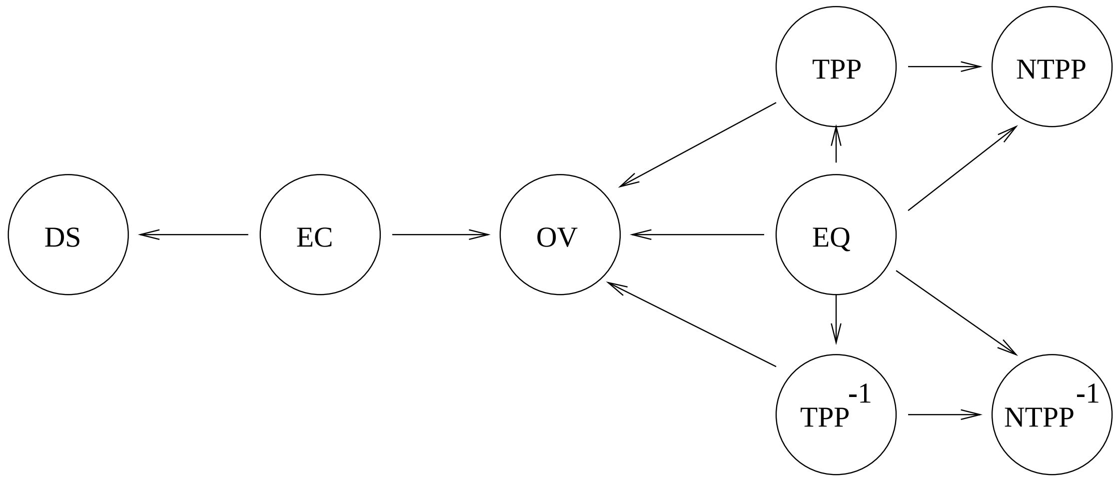


Figure 11: Transition graph for d_{Hd} and d_O

the volume.

Now, this is certainly not a decisive argument that the transition from DC to EQ is a useful one. For one thing, it could be argued that the volume at the end is filled by the mixture of the two gasses and not by either gas individually. For another, gasses do not behave like marching bands, and in reality will certainly start to mix, and thus occupy overlapping regions, before they occupy coextensional regions. But the example does, I think, suggest that the possibility of a EQ to DC cannot be dismissed quite as readily as one might at first suppose, and that it may, in fact, depend in a subtle way on the idealization of a discrete collection filling a space.

6 Muller's theory

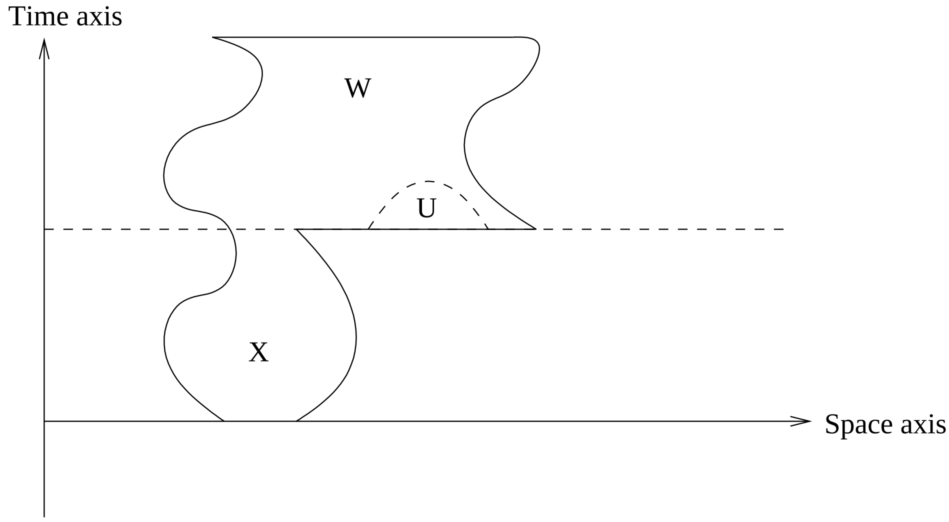
In an important and elegant paper, Philippe Muller (1998) develops a theory of motion based on the geometry of four-dimensional regions of space-time, called "histories". This approach was suggested in Hayes' "Naive Physics Manifesto" (1979), but Muller's paper was the first to explore it seriously. The paper is relevant to the analysis here, because it proposes a definition of continuity and purports to derive transition relations from that definition.

Muller's language is a first-order language over the universe of histories. Muller's paper admits both regions that are regular and those that are "regular and open"; i.e. the interior of regular regions. I will modify this here to include only regular regions; this restriction does not affect the issues under discussion here, and it simplifies the presentation.

The language contains three primitive binary relations:[6]

Cxy – Regions x and y are connected; that is, they share at least one point.
$x < y$ — Region x strictly precedes region y temporally.
$x \gtrless y$ — The temporal projections of regions x and y share at least one instant.

[6] In this section I will follow Muller's notational conventions: variables are lower-case; predicates are either prefix or infix; atomic formulas are strings of symbols without further punctuation.



W is the whole history in the solid boundary. X is the portion of W below the dotted line.
U is the semi-circle on the left, also a subregion of W.

Figure 12: Muller's definition of discontinuity

Other relations between regions are defined in terms of non-recursive first-order formulas over these.

Muller proposes the following definition (D4.2) for continuity: Region w represents a continuous function from time to space if it is satisfies the following:

$$\text{D4.2: CONTINU}w \triangleq \text{CON}_t w \wedge \forall_x \forall_u \; ((\text{TS}xw) \wedge x \lessgtr u \wedge \text{P}uw) \Rightarrow \text{C}xu.$$

Here "$CON_t w$" means that the temporal projection of w is a connected time interval. "Puw" means that u is a subregion of w. "TSxw" means that x is a "time-slice" of w. This predicate is a little tricky. For a normal regions w and x, it asserts that x is the normalization[7] of the restriction of w to a time period i, where i is a regular subset of the time-line.

Figure 12 shows how a discontinuous function of time fails to satisfy definition D4.2. Note that x does not contain the line l.

It can be shown that the graph of a function from time to regions satisfies definition D4.2 if and only if it is continuous with respect to the Hausdorff metric. More precisely, we can state the following theorem:

Theorem 6: Let w be a bounded normal history whose temporal projection is a connected time interval I, and let $w(t)$ be the cross-section of w at time $t \in I$. Then w satisfies Muller's definition D4.2 iff $w(t)$ is continuous in the Hausdorff distance.

The proof is given in section 8.3 of appendix A.

6.1 Transitions in mereotopology

Theorem 6 above leads to a conflict between Muller's analysis of transitions and our own. Muller claims to show that it follows from his definition that the only transitions possible are those shown

[7] The normalization of a region r is the closure of the interior of r.

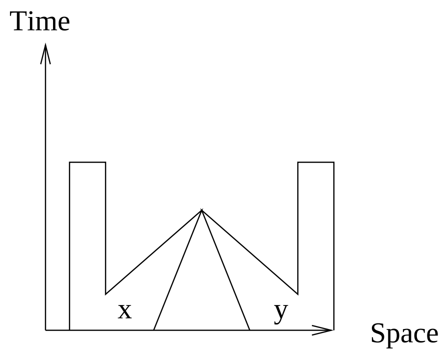


Figure 13: The normalized cross-sections of x and y are always disconnected

in figure 8. We have shown, on the contrary, that functions continuous in the Hausdorff distance can execute any of the transitions in figure 10. Indeed, note that the transition from EQ to DS described in section 5 satisfies Muller's of continuity. Where, then, is the difference between Muller's account of the transitions and ours?

The resolution is that Muller is using a different, and, we believe, flawed formal definition of a "transition". That is, the formal theorems Th 4.3 – 4.6 that Muller proves are, indeed, true, but his interpretation of these theorems as expressing transition relations is incorrect.

Consider, for example, the form of Th. 4.3:

$$\text{TS}xw \wedge \text{P}yw \wedge \text{DC}xz \wedge \text{OV}_{sp}yz \wedge x \lesseqgtr y \Rightarrow \neg\text{CONTINU}w$$

Muller claims that this theorem expresses the impossiblity of a transition from DC to OV by a continuous function w.

The relation $\text{OV}_{sp}yz$ is intended to mean that y and z overlap spatially in every temporal slice of their common domain. Muller's formal definition is

$$\text{OV}_{sp}xy \triangleq \text{OV}xy \wedge x \subseteq_t x \cdot y$$

An analogous spatial relation is defined for each of the RCC relations.

There are a couple of peculiarities with the above form of Thm. 4.3. First, it is not at all obvious why this axiom should express the non-existence of a transition relation. Second, the relations DC and OV occupy non-symmetric logical positions in this formula. (Muller is not trying to identify the "directionality" of transitions, in the sense of section 5.)

Furthermore, the spatial RCC relations are defined *ad hoc*; it is not clear what precisely Muller intends and therefore not clear whether his formal definitions achieve his intentions. For example, the above definition of OV_{sp} has at least three apparent formal flaws. First, it is asymmetric in x and y. Second, Muller does not define the notation $x \cdot y$ and it is not clear whether he means the intersection or the normalized intersection. Third, the condition $\text{OV}xy$ is redundant if $x \cdot y$ means the normalized intersection, and does not accomplish anything reasonable if $x \cdot y$ means the intersection. Another example: Muller claims that DC_{sp} is equivalent to DC; however, figure 13 shows two histories x and y that are not DC but whose normalized spatial cross sections are always disconnected.

A more systematic approach to representing transitions can be accomplished along the following lines: Let $x(t)$ and $y(t)$ be two functions from time intervals to normal spatial regions, and let R and P be two spatial relations. We say that x and y transition from R to P if, for some t_1, t_2 in the domains of both x and y, if $\mathrm{R}(x(t_1), y(t_1))$, $\mathrm{P}(x(t_2), y(t_2))$ and for all $t \in [t_1, t_2]$ either $\mathrm{R}(x(t), y(t))$ or $\mathrm{P}(x(t), y(t))$.

Now, turning to the language of histories, let x be a normal history whose temporal projection is a connected interval i. Define the corresponding function $x(t)$ to be the normalization of the cross-section of x at t, which we will abbreviate $\mathrm{Nx}(x, t)$. (This is not, of course, a function in the formal language, as it maps to a purely spatial region; it is just for our informal discussion.) x is then equal to the closure of the set $\{\langle t, p \rangle \mid p \in x(t)\}$. We can then define two normal histories x and y as transitioning transition from R to P if $\mathrm{Nx}(x, t)$ and $\mathrm{Nx}(y, t)$ transition from R to P as defined above. Our task, then, is to give a definition of the concept "The normalized cross-sections of x and y at time t have spatial relation R," in Muller's language of histories. This is done in appendix B.

It can reasonably be objected to this analysis that, though it observes the letter of the mereotopological enterprise, it violates the spirit, as it achieves its ends by using the very great expressive power of first-order logic over histories to, in effect, define time instants and spatial points. Certainly, proving the correctness of rules that state the non-existence of transitions, or worse, those that state the existence of transitions, from plausible mereotopological axioms, would seem to be daunting if not hopeless in this expression of these rules. One might hope, therefore, that a more natural mereotopological expression of transition rules could be found that could indeed be proved in a mereotopological theory. There is reason to doubt, however, that a such a characterization could be found that would be entirely satisfying, as, ultimately, the whole notion of these transition runs somewhat counter to the meretopological point of view. For example, there are histories that go from TPP to TPP^{-1} passing through EQ for an *instant* of time, so a rule that blocks a direct transition from TPP to TPP^{-1} must, so to speak, detect this spatial condition that holds at a single instant. There is something very non-mereotopological about this.

7 An Open Problem

One goal that we have been unable to achieve is to give an explanation, grounded in the topologies of the space of regular regions, of the following observation (A.G. Cohn, personal communication): All the entries in the composition table of the RCC relations correspond to connected subsets of the transition network. It is true that each of the RCC relations constitutes a connected set in the space $\mathcal{S} \times \mathcal{S}$ where $\mathcal{S}$ is the space of regular regions under any of our topologies, except d_O, but it is not true in general that composing of two relations, each of which is connected, gives a new relation that is also connected. It seems as though there should be a fundamental reason that this property holds over the composition tables, but I have not found it.

8 Appendix A: Proofs

This paper adduces many theorems;[8] however, most of these are entirely straightforward. The only ones of any difficulty are the relation of the dual-Hausdorff metric to the area; the relation of the optimal-homeomorphism metric to the circumference and interal diameter; and the fact that Muller's definition is equvalent to continuity in the Hausdorff metric. These are proven in this appendix.

[8] For example, for every pair of nodes in figures 9 through 11 there is a theorem stating either that there is or that there is not an arc between them.

8.1 The dual-Hausdorff distance and the area

Definition 14: Fix a coordinate system in the plane. Let $\delta > 0$. A *δ-grid* square is a square $[N\delta, (N+1)\delta] \times [M\delta, (M+1)\delta]$ where N and M are integers.

Definition 15: Fix a coordinate system in the plane. Let $\mathbf{R}$ be a bounded regular region, and let $\delta > 0$. The *δ-grid approximation of* $\mathbf{R}$, denoted "$\mathrm{Gr}(\mathbf{R}, \delta)$" is the union of all δ-grid square contained in $\mathbf{R}$.

The following is a basic property of area (indeed, it can be taken as the definition of area):

Lemma 7: Let $\mathbf{R}$ be a bounded regular region. Then for any $\epsilon > 0$ there exists an $\eta > 0$ such that for all positive $\delta < \eta$, $\mathrm{area}(\mathbf{R}) - \mathrm{area}(\mathrm{Gr}(\mathbf{R}, \delta)) < \epsilon$.

Lemma 8: Let $\mathbf{R}$ be a bounded regular region, and let $\epsilon > 0$. Then there exists $\mu > 0$ such that, for any region $\mathbf{S}$, if $\mathrm{d}_H(\mathbf{R}^c, \mathbf{S}^c) < \mu$, then $\mathrm{area}(\mathbf{R} - \mathbf{S}) < \epsilon$.

Proof: Using lemma 7, choose δ so that $\mathrm{area}(\mathbf{R}) - \mathrm{area}(\mathrm{Gr}(\mathbf{R}, \delta)) < \epsilon/2$. Let $K = \mathrm{area}(\mathrm{Gr}(\mathbf{R}, \delta))/\ \delta^2$, the number of δ-grid squares in $\mathrm{Gr}(\mathbf{R}, \delta)$. Let $\mu = \epsilon/8K\delta$. Choose $\mathbf{S}$ so that $\mathrm{d}_H(\mathbf{R}^c, \mathbf{S}^c) < \mu$.

Now let $\mathbf{Q}$ be any δ-grid square in $\mathrm{Gr}(\mathbf{R}, \delta)$ and let $\mathbf{x}$ be a point in $\mathbf{Q}$ that is more than μ from the edges of $\mathbf{Q}$. Then, since $\mathbf{Q} \subset \mathbf{R}$, $\mathrm{d}(\mathbf{x}, \mathbf{R}^c) \geq \mathrm{d}(\mathbf{x}, \mathbf{Q}^c) > \mu$, so $\mathbf{x} \notin \mathbf{S}^c$. That is, $\mathbf{S}$ contains the entire square $\mathbf{Q}$ except for the strips within μ of the edges. The area of $\mathbf{Q} \cap \mathbf{S}$ is therefore at least $(\delta - 2\mu)^2 > \delta^2 - 4\mu\delta$. Summing over all the K δ-grid squares, we derive that $\mathrm{area}(\mathbf{S} \cap \mathrm{Gr}(\mathbf{R}, \delta)) > K(\delta^2 - 4\mu\delta) = \mathrm{area}(\mathrm{Gr}(\mathbf{R}, \delta)) - \epsilon/2$. Therefore $\mathrm{area}(\mathbf{R} - \mathbf{S}) \leq \mathrm{area}(\mathbf{R} - (\mathbf{S} \cap \mathrm{Gr}(\mathbf{R}, \delta))) = \mathrm{area}(\mathbf{R}) - \mathrm{area}(\mathbf{S} \cap \mathrm{Gr}(\mathbf{R}, \delta))) \leq \mathrm{area}(\mathbf{R}) - (\mathrm{area}(\mathrm{Gr}(\mathbf{R}, \delta)) - \epsilon/2) < \epsilon$. ∎

Theorem 9: Let $\mathbf{R}$ be a bounded regular region, and let $\epsilon > 0$. Then there exists $\mu > 0$ such that, for any region $\mathbf{S}$, if $\mathrm{d}_{Hd}(\mathbf{R}, \mathbf{S}) < \mu$, then $\mathrm{d}_A(\mathbf{R}, \mathbf{S}) < \epsilon$.

Proof: Let $\mathbf{O}$ be a bounded regular region that contains all points within distance 1 of $\mathbf{R}$. Let $\mathbf{Q}$ be the closure of $\mathbf{O} - \mathbf{R}$. By lemma 8, there exists μ_1 such that, for any regular region $\mathbf{S}$. $\mathrm{d}_H(\mathbf{R}^c, \mathbf{S}^c) < \mu_1$, then $\mathrm{area}(\mathbf{R} - \mathbf{S}) < \epsilon/2$. Also by lemma 8, there exists μ_2 such that, for any regular region $\mathbf{T}$, if $\mathrm{d}_H(\mathbf{Q}^c, \mathbf{T}^c) < \mu_2$, then $\mathrm{area}(\mathbf{Q} - \mathbf{T}) < \epsilon/2$. Let $\mu = \min(\mu_1, \mu_2, 1)$. Let $\mathbf{S}$ be any region such that $\mathrm{d}_{Hd}(\mathbf{R}, \mathbf{S}) < \mu$. Clearly $\mathbf{S} \subset \mathbf{O}$. Let $\mathbf{T} = \mathbf{O} - \mathbf{S}$. Then $\mathrm{d}_H(\mathbf{R}^c, \mathbf{S}^c) < \mu \leq \mu_1$, and $\mathrm{d}_H(\mathbf{Q}^c, \mathbf{T}^c) = \mathrm{d}_H(\mathbf{R}, \mathbf{S}) < \mu \leq \mu_2$. Therefore $\mathrm{d}_A(\mathbf{R}, \mathbf{S}) = \mathrm{area}(\mathbf{R} - \mathbf{S}) + \mathrm{area}(\mathbf{S} - \mathbf{R}) = \mathrm{area}(\mathbf{R} - \mathbf{S}) + \mathrm{area}(\mathbf{Q} - \mathbf{T}) < \epsilon$.

Corollary 10: The area of a region is continuous under the metric d_{Hd}.

Corollary 11: The metric d_{Hd} defines a topology finer than that defined by d_A.

8.2 The optimal-homeomorphism metric and smoothed path lengths

Definition 16: For the purposes of this section, a *path* in the plane is a continuously differentiable function from [0,1] to the plane. If $\phi(t)$ and $\psi(t)$ are paths, then $\mathrm{d}_O(\phi, \psi)$ is defined to be $\max_{t \in [0,1]} \mathrm{d}(\phi(t), \psi(t))$.

We repeat definition 10:

Definition 10: Let ϕ be a simple curve in the plane, and let $\Delta > 0$ be a distance. Let $A(\phi, \Delta)$ be the set of all paths ψ such that $\mathrm{d}_O(\phi, \psi) \leq \Delta$. We define the Δ-smoothed length of ϕ to be the greatest lower bound over the arc-length of paths in $A(\phi, \Delta)$.

$$\mathrm{smooth}(\phi, \Delta) = \inf_{\psi \in A(\phi,\Delta)} \mathrm{length}(\psi).$$

Lemma 12: Let ϕ and ψ be paths and let p be a value between 0 and 1. Define the path $\theta(t) = p\phi(t) + (1-p)\psi(t)$. Then length$(\theta) \leq p \cdot$ length$(\phi) + (1-p) \cdot$ length(ψ).

Proof:

$$\text{length}(\theta) = \int_0^1 \mid \dot{\theta}(t) \mid dt = \int_0^1 \mid p\dot{\phi}(t) + (1-p)\dot{\psi}(t) \mid dt \leq$$

$$\int_0^1 \mid p\dot{\phi}(t) \mid + \mid (1-p)\dot{\psi}(t) \mid dt = p \cdot \text{length}(\phi) + (1-p) \cdot \text{length}(\psi)$$

Lemma 13: Let ϕ be a path, and let $\alpha > \beta > 0$. Then

$$\text{smooth}(\phi, \beta) \geq \text{smooth}(\phi, \alpha) \geq \text{smooth}(\phi, \beta) - \frac{\alpha - \beta}{\beta}(\text{length}(\phi) - \text{smooth}(\phi, \beta))$$

Proof: The first inequality is trivial, since $A(\phi, \alpha) \supset A(\phi, \beta)$, and so the infimum is being taken over a smaller set.

Let ψ be a path in $A(\phi, \alpha)$ such that length$(\psi) <$ length(ϕ). Define the path $\theta(t)$ to be the linear sum $\theta(t) = (\beta/\alpha)\psi(t) + (1 - \beta/\alpha)\phi(t)$. Then $\text{d}(\phi(t), \theta(t)) = (\beta/\alpha)\ \text{d}(\phi(t), \psi(t)) \leq \beta$, so θ is in $A(\phi, \beta)$.

By lemma 12

$$\text{length}(\theta) \leq \frac{\beta}{\alpha}\text{length}(\psi) + (1 - \frac{\beta}{\alpha})\text{length}(\phi)$$

But length$(\theta(t)) \geq$ smooth(ϕ, β) and smooth(ϕ, α) is the infimum of length(ψ) for all such ψ. Hence, smooth$(\phi, \beta) \leq (\beta/\alpha)$ smooth$(\phi, \alpha) + (1 - \beta/\alpha)$ length(ϕ). An algebraic transformation gives the second inequality in the statement of the lemma. ∎

Theorem 14: For any path ϕ and any $\Delta > 0$ and $\epsilon > 0$ there exists $\delta > 0$ such that, if ψ is a path and $\text{d}_O(\phi, \psi) < \delta$, then $\mid \text{smooth}(\phi, \Delta) - \text{smooth}(\psi, \Delta) \mid < \epsilon$.

Proof: Let $\delta < \Delta$ and let ψ be any path such that $\text{d}_O(\phi, \psi) < \delta$. Then
$A(\phi, \Delta - \delta) \subset A(\psi, \Delta) \subset A(\phi, \Delta + \delta)$.
Therefore, smooth$(\phi, \Delta - \delta) \geq$ smooth$(\psi, \Delta) \geq$ smooth$(\phi, \Delta + \delta)$.
But by lemma 13,

$$\text{smooth}(\phi, \Delta) \geq \text{smooth}(\phi, \Delta - \delta) - \frac{\delta}{\Delta - \delta}(\text{length}(\phi) - \text{smooth}(\phi, \Delta - \delta))$$

so

$$\text{smooth}(\phi, \Delta - \delta) \leq \text{smooth}(\phi, \Delta) + \frac{\delta}{\Delta - \delta}\text{length}(\phi)$$

Also,

$$\text{smooth}(\phi, \Delta + \delta) \geq \text{smooth}(\phi, \Delta) - \frac{\delta}{\Delta}(\text{length}(\phi) - \text{smooth}(\phi, \Delta))$$

Hence, if we choose $\delta < \min(\Delta/2, \Delta\epsilon/2\text{length}(\phi))$, then the conclusion of the theorem is satisfied.

Corollary 15: For any fixed $\Delta > 0$, the Δ-smoothed circumference and the Δ-smoothed path diameter of a region are continuous functions with respect to the metric d_O.

Corollary 15 is the interesting part of theorems 4 and 5 (section 4.2).

8.3 Muller's definition and continuity in the Hausdorff metric

In this section, we prove that Muller's definition of continuity (D4.2) is equivalent to continuity in the Hausdorff metric.

We repeat the definition:

$$\text{D4.2: CONTINU}w \triangleq \text{CON}_t w \wedge \forall_x \forall_u \ ((\text{TS}xw) \wedge x \lessgtr u \wedge \text{P}uw) \Rightarrow \text{C}xu.$$

Theorem 16: Let w be a bounded normal history whose temporal projection is a connected time interval I, and let $w(t)$ be the cross-section of w at time $t \in I$. Then w satisfies Muller's definition D4.2 iff $w(t)$ is continuous in the Hausdorff distance.

Proof: Suppose that w does not satisfy D4.2. Then there exists a time slice x of w and a normal subset u of w such that x and u meet in time but x is not connected to u. Let t_0 be a time instant within the temporal projections of both x and u. Let $x(t)$ $u(t)$ be the cross-sections of histories x and u at time t. Note that, for every t, $x(t)$, $w(t)$, and $u(t)$ are closed regions of space, though not necessarily regular regions of space.

Since x is a normal history, there is a time interval i containing t_0 in the time projection of x. Assume, without loss of generality, that i precedes t_0. By definition of a time slice, for any t in the interior of i, $w(t) = x(t)$. Let p be any spatial point in $u(t_0) \subset w(t_0)$. Then, since u is disconnected from x, p is not in $x(t_0)$, so the spatio-temporal point $\langle t_0, p\rangle$ is not in x. Since x is closed , there must exist an $\epsilon > 0$ and a subinterval of i, (t_1, t_0), such that for all $t \in (t_1, t_0)$ $\text{d}(p, x(t)) > \epsilon$. But for $t \in (t_1, t_0)$, $\text{d}_H(w(t_0), w(t)) \geq \text{d}(p, w(t)) = \text{d}(p, x(t)) > \epsilon$. Hence $w(t)$ is discontinuous with respect to d_H at t_0.

Conversely, suppose that $w(t)$ is discontinuous with respect to d_H at time t_0. Then there exists an $\epsilon > 0$ and a sequence of times $t_1, t_2 \ldots$ converging to t_0 such that $\text{d}_H(w(t_i), w(t_0) > \epsilon$ for all t_i. There must be an infinite subsequence of these that lies on one side of t_0, and the case where this is above t_0 is symmetric with the case where it is below. Therefore, we can assume without loss of generality that the t_i converge to t_0 from below. For each such t_i, either there exists a point $p_i \in w(t_i)$ such that $\text{d}(p_i, w(t_0)) > \epsilon$ or there exists a point $q_i \in w(t_0)$ such that $\text{d}(q_i, w(t_i)) > \epsilon$. At least one of the sequences p_i, q_i must be infinite.

Suppose that the sequence p_i is infinite. Since w is bounded, the p_i must all lie in some bounded region of space. Hence, they must have a cluster point p. Now there is a problem: since w is closed the spatio-temporal point $\langle t_0, p\rangle$ must lie in w, but $\text{d}(p, w(t_0))$ must be greater than ϵ, which is a contradiction.

Suppose that the sequence q_i is infinite. Again, these must have a cluster point q. Since $w(t_0)$ is compact, $q \in w(t_0)$. By renumbering, restrict the sequence q_i so that $\text{d}(q, q_i) < \epsilon/2$ for all i. Then $\text{d}(q, w_i) > \epsilon/2$ for all i. Let x be the time slice of w over $[0, t_0]$, and let u be a normal subset of w contaning $\langle t_0, q\rangle$ of diameter less than $\epsilon/2$. Then the temporal projection of x and u share the point t_0, but x and u are not connected, so by definition D4.2 w is not continuous. ∎

9 Appendix B: Expressing spatial relations between cross-sections of histories

In this section we show how to assert in Muller's language that a given RCC spatial relation holds between every temporal cross-section of two histories.

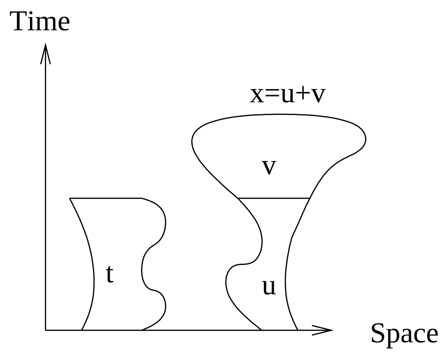


Figure 14: The relation SPLIT$txuv$

I. We will identify a time instant t with any history that ends at t. The following temporal predicates will be useful:

$$\text{MEET}xy \triangleq x \lessgtr y \wedge \neg x\sigma y \wedge \forall_{wz}(\text{P}wx \wedge \text{P}zy) \Rightarrow w < z \vee w \lessgtr z$$

The notation $x\sigma y$ means that the interiors of the time-projections of x and y overlap. Muller gives the following definition:

$$x\sigma y \triangleq \exists_z z \subseteq_t x \wedge z \subseteq_t y$$

Muller also gives a defintion of MEET, but his requires the use of open histories, which we have excluded.

$$\text{SAME-END}xy \triangleq \exists_z \text{MEET}xz \wedge \text{MEET}yz$$

(x and y end at the same time.)

II. If time t is in the interior of the temporal projection of history x, then x can be split into two histories u and v such that u ends at the end of t and v begins at the beginning of t, using the following formula.

$$\text{SPLIT}txuv \triangleq x = u + v \wedge \text{SAME-END}tu \wedge \text{MEET}tv$$

Note that the cross-section of x at t is equal to the union of the cross-sections of u and v, and likewise for the normalized cross-sections (Figure 14).

III. A history a "touches" history u from above if u meets a and u is externally connected to a. History a touches history v from below if a meets v and a is externally connected to v.

$$\text{TOUCHX1}au \triangleq \text{MEET}ua \wedge \text{EC}ua$$

$$\text{TOUCHX2}av \triangleq \text{MEET}av \wedge \text{EC}av$$

IV. If history u meets history v, then they have the same meeting face if every history a that touches u is externally connected to a history b that touches v and vice versa. (Figure 15)

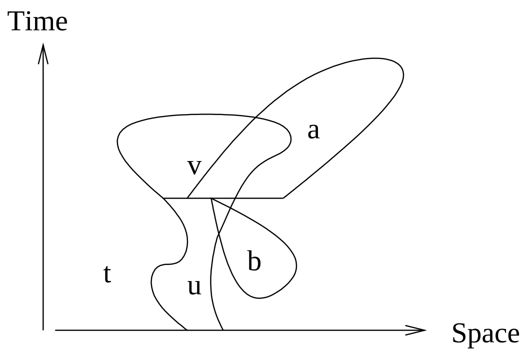


Figure 15: The relation SAME-FACEuv

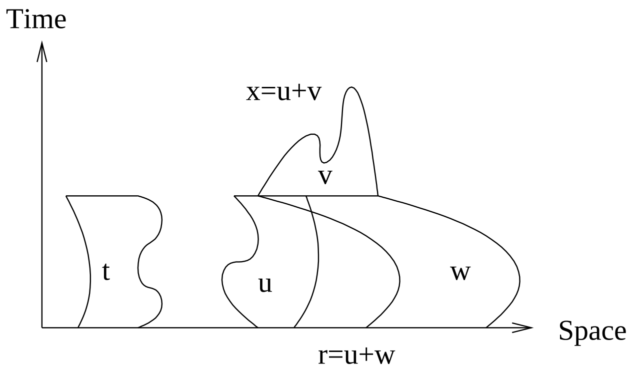


Figure 16: The relation REFLECTxtr

$$\begin{aligned}&\text{SAME-FACE}uv \triangleq\\&\text{MEET}uv \land [[\forall_a \text{TOUCHX1}au \Rightarrow \exists_b \text{TOUCHX2}bv \land \text{EC}ab]\land\\&\qquad\qquad [\forall_b \text{TOUCHX2}bv \Rightarrow \exists_a \text{TOUCHX1}au \land \text{EC}ab]]\end{aligned}$$

V. We can now simplify the analysis of the cross section of history x at time t by creating a new history r that consists of: the union of [the part of x before t] with [a "reflection" of the part of x after t]. (Figure 16.)

$$\text{REFLECT}xtr \triangleq \exists_{u,v,w} \text{SPLIT}xtuv \land \text{SAME-FACE}vw \land r = u + w$$

VI. History a touches the boundary of the end of history u if a has the same end as u, and no matter how thin you slice a, the later half is EC to u (Figure 17).

$$\text{TOUCHBD}au \triangleq \text{SAME-END}au \land \forall_s \text{SPLIT}sabc \Rightarrow \text{EC}uc$$

VI. History a touches the interior of the end of history u if a touches u but is disconnected from any history that touches the boundary of the end of u (Figure 18.)

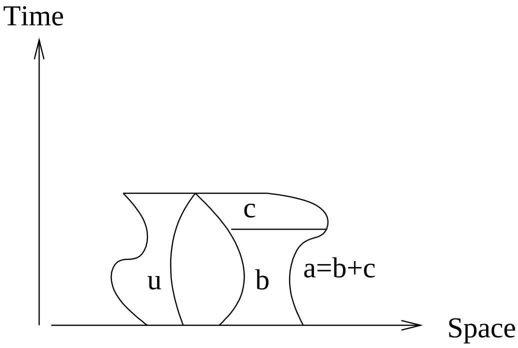


Figure 17: The relation TOUCHBDau

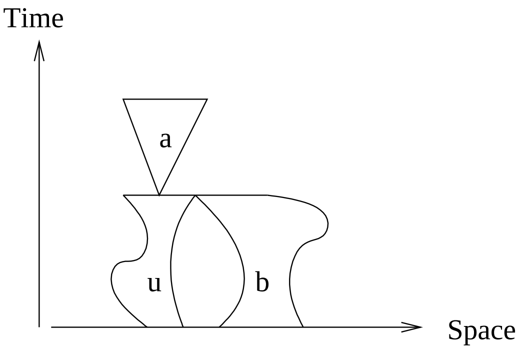


Figure 18: The relation TOUCHINTau

$$\text{TOUCHINT}au \triangleq \text{TOUCH1}au \wedge \forall_b \text{TOUCHBD}bu \Rightarrow \text{DC}ba$$

VII. The normalized cross-section of history x at time t is a subset of the normalized cross-section of history y at time t if the following holds: Let q and r be reflections of x and y at t. Then any region that touches the interior of q also touches the interior of r. (We will subscript c to RCC relations to indicate spatial relations that hold at an instantaneous cross-section.)

$$\text{PP}_c txy \triangleq \exists_{q,r} \text{REFLECT}xtq \wedge \text{REFLECT}ytr \wedge \forall_a \text{TOUCHINT}aq \Rightarrow \text{TOUCHINT}ar$$

VII. The boundary of normalized cross-section of history x at time t is disjoint from the boundary of the normalized cross-section of history y at time t if the following holds: Let q and r be reflections of x and y at t. Let a and b be histories that touch the boundaries of the ends of q and r. Then there exist subhistories c in a and d in b such that c touches the boundary of the end of q, d touches the boundary of the end of r, but c and d are disconnected.

$\text{DISJ-BD}txy \triangleq$
$\exists_{q,r} \text{REFLECT}xtq \wedge \text{REFLECT}ytr \wedge$

$\forall_{a,b}$[TOUCH-BD$aq \wedge$ TOUCH-BDbr]$\Rightarrow$
$\exists_{c,d}$PP$ca \wedge$ PP$db \wedge$ TOUCH-BD$cq \wedge$ TOUCH-BD$da \wedge$ DCcd.

VIII. The remaining spatial relations over normalized cross-sections of x and y can be defined from PP$_c$ and DISJ-BD$_c$ along the lines of table 1.

TPP$_c txy \triangleq$ PP$_c txy \wedge \neg$DISJ-BDtxy.
NTPP$_c xy \triangleq$ PP$_c txy \wedge$ DISJ-BDtxy.
DS$_c txy \triangleq \neg\exists_wPP_c twx \wedge$ PP$_c twy$.
(The cross-sections of x and y at t have no common interior points.)
OV$_c txy \triangleq \quad \exists_{a,b,c}PP_c tax \wedge$ PP$_c tay \wedge$ PP$_c tbx \wedge$ PP$_c tcy \wedge$ DS$_c tbx \wedge$ DS$_c tcy$.
EC$_c txy \triangleq$ DS$_c txy \wedge \neg$DISJ-BDtxy.
DC$_c txy \triangleq$ DS$_c txy \wedge$ DISJ-BDtxy.
EQ$_c txy \triangleq$ PP$_c txy \wedge$ PP$_c tyx$.

References

E. Davis (1999). Shape Approximation in Kinematic Systems. In preparation.

A. P. Galton (1993). Towards an integrated logic of space, time, and motion. *Proceedings IJCAI-93*, pages 1550 - 1555.

A.P. Galton (1995). Towards a qualitative theory of motion. In W. Kuhn and A. Frank (eds.) *Spatial Information Theory: A theoretical basis for GIS — Proceedings of COSIT-95* Lectures in Computer Science #988, Springer-Verlag, Berlin, pages 377-396.

P. Hayes (1979). The Naive Physics Manifesto. In D. Michie. (ed.) *Expert Systems in the Microelectronic Age.* Edinburgh: Edinburgh University Press.

P. Hayes (1985). Naive Physics 1: Ontology for Liquids. In J. Hobbs and R. Moore. (eds.) *Formal Theories of the Commonsense World.* Norwood, N.J.: Ablex Pubs.

L. Joskowicz, E. Sacks, and V. Srinivasan (1997). Kinematic Tolerance Analysis. *Computer-Aided Design,* Vol. 29 No. 2.

D. Mumford (1991). Mathematical Theories of Shape: Do they model perception? In *Geometric Methods in Computer Vision.* Society of Photo-Optical Instrumentation Engineers, Vol. 1570.

P. Muller (1998). A Qualitative Theory of Motion Based on Spatio-Temporal Primitives. *Sixth International Conference on Principles of Knowledge Representation and Reasoning,* Trento, Italy, pages 131-141.

D. A. Randell, Z. Cui, and A. G. Cohn (1992). A Spatial Logic Based on Regions and Connection. *Third International Conference on Principles of Knowledge Representation and Reasoning,* Boston, Mass. pages 165-176.

A. A. G. Requicha (1983). Towards a Theory of Geometric Tolerancing, *The International Journal of Robotics Research,* vol 2, no. 4, pages 45-60.