Dartmouth College
Computer Science 44
Winter 1999

Computer Vision

Hausdorff Distance Image Comparison

Given two sets of points $A = \{a_1, \ldots, a_m\}$ and $B = \{b_1, \ldots, b_n\}$, the Hausdorff distance is defined as

$$H(A, B) = \max(h(A, B), h(B, A)) \tag{1}$$

where

$$h(A, B) = \max_{a \in A} \min_{b \in B} \|a - b\|. \tag{2}$$

The function $h(A, B)$ is called the *directed* Hausdorff 'distance' from A to B (this function is not symmetric and thus is not a true distance). It identifies the point $a \in A$ that is farthest from any point of B, and measures the distance from a to its nearest neighbor in B. Thus the Hausdorff distance, $H(A, B)$, measures the degree of mismatch between two sets, as it reflects the distance of the point of A that is farthest from any point of B and vice versa. Intuitively, if the Hausdorff distance is d, then every point of A must be within a distance d of some point of B and vice versa. It is well known that the Hausdorff distance is a metric over the set of all closed, bounded sets – it obeys the properties of identity, symmetry and triangle inequality.

One way to understand the directed Hausdorff measure is to think of it in terms of set containment. Let $B^\delta = B \oplus C_\delta$, where C_δ is a disk of radius δ and $\oplus$ is the Minkowski sum (for two point sets P and Q, $P \oplus Q = \{p + q | p \in P, q \in Q\}$). Intuitively, B^δ is the set obtained by replacing each point of B with a disk of radius δ, and taking the union of all of these disks. By definition $h(A, B) \leq \delta$ if and only if $A \subseteq B^\delta$, because in order for every point of A to be within distance δ of B it must be contained in B^δ.

The Hausdorff distance is very sensitive to even a single "outlying" point of A or B. For example, consider $A = B \cup \{x\}$, where the point x is some large distance D from any point of A. In this case $H(A, B) = D$ – it is determined solely by the point x. Thus rather than using $H(A, B)$ we use a generalization of the Hausdorff distance (which does not obey the metric properties on A and B, but does obey them on *specific subsets* of A and B). This generalized Hausdorff measure is given by taking the k-th ranked distance rather than the maximum, or largest ranked one,

$$h_k(A, B) = \underset{a \in A}{k\text{th}} \min_{b \in B} \|a - b\| \tag{3}$$

[2]These notes are adapted from a course CS 462, taught at Cornell University by Bruce Donald and Dan Huttenlocher.

where kth denotes the k-th ranked value (or equivalently the $k/m - th$ quantile of m values). For example, when $k = m$ then kth is max, and thus this measure is the same as $h(\cdot, \cdot)$. When $k = m/2$ then the median of the m individual point distances determines the overall distance. Therefore this measure generalizes the directed Hausdorff measure, by replacing the maximum with a quantile.

In terms of set containment, $h_k(A, B) \leq \delta$ if and only if there is some $A_k \subseteq A$ such that $A_k \subseteq B^\delta$, where A_k contains k points of A. Thus we can think of $h_k(A, B)$ as partitioning A into two sets, A_k which is "close to" (within δ of) B and the "outliers" $A - A_k$. This is in practice an important aspect of the generalized Hausdorff measure: it separately accounts for perturbations (by the distance δ) and for outliers (by the rank k).

In general, we are interested in using the Hausdorff distance to identify instances of some "model" bitmap M in some image bitmap I. The image may contain multiple instances of the model (or no model), the instances may be transformed in some manner, parts may be missing, and there may be other portions of the image for which there is no model. We will limit ourselves here to the case in which the model simply undergoes translation, but more complex transformations are possible in this framework. One way to phrase this problem is that we seek all translations $t \in \Re^2$ such that $h_k(M \oplus t, I) \leq \delta$. The parameter k tells us how many of the model points should be near image points in order to classify a given translation as an instance of the model (i.e., we allow $m - k$ of the m model points to be outliers). The parameter δ tells us how close each non-outlying model point must be to some image point.

There are several techniques for computing the set of all such translations. The most straightforward method is dilation and correlation. In order to find each translation such that $h_k(M \oplus t, I) \leq \delta$, we form $I^\delta = I \oplus C_\delta$ and then compute the correlation of M with I^δ. For each translation t of M with respect to I^δ, the correlation determines p, how many points of $M \oplus t$ are superimposed with I^δ (the logical **and** of $M \oplus t$ and I^δ). Any translation where $p \geq k$ is a translation where an instance was found, i.e. for which $h_k(M \oplus t, I) \leq \delta$. We refer to p/m as the Hausdorff fraction for a given translation t (at some fixed δ), and to k/m as the *threshold* fraction, as this normalizes the quantities by the number of model points. Intuitively, the Hausdorff fraction measures the percentage of $M \oplus t$ that lies near (within δ of) points of I.

The computation of $h_k(M \oplus t, I) \leq \delta$ alone does not necessarily find very good instances of M in I, rather it finds portions of I that could contain M plus some other points. For instance, a totally filled in image will match any model at any translation. Thus we often use the other direction of the generalized Hausdorff measure, $h_k(I, M \oplus t)$, to "verify" those translations where it was found that $h_k(M \oplus t, I) \leq \delta$. Rather than using the entire image, I, however, we use just the portion of the image that is covered by the translated model $M \oplus t$ (e.g., if M is represented as a bitmap, then the portion of the image that is within the extent of that bitmap). This verification, or reverse Hausdorff fraction, ensures that a given portion of the points in the image (covered by the model array) are actually near points of $M \oplus t$. It thereby rules out situations such as a totally filled in image matching any model.