On Hausdorff-like metrics for fuzzy sets

Laurence Boxer

Department of Computer and Information Sciences, Niagara University, Niagara University, NY 14109, USA

Received 25 November 1996; revised 3 January 1997

Abstract

In (Chaudhuri and Rosenfeld, 1996), metrics were developed for fuzzy sets \( u \) and \( v \) defined on the same support set \( S \). These metrics are based on applying the Hausdorff metric to pairs of membership level sets \( u^{-1}([t_k, 1]), v^{-1}([t_k, 1]) \), where \( t_k \in [0, 1] \) is a (possible) membership value of \( u \) or \( v \).

The metrics of (Chaudhuri and Rosenfeld, 1996) give nice comparisons of fuzzy sets by measuring the differences in their images, but suffer the undesirable restriction that fuzzy sets so compared must have the same maximum values. In the current paper, we show how the metrics of (Chaudhuri and Rosenfeld, 1996) may be modified so as to remove this restriction while preserving some of the "feel" of the metrics described in (Chaudhuri and Rosenfeld, 1996). We also present an algorithm for computing our metric that runs in \( O(mn) \) time for a 2-dimensional digital picture of \( n \) pixels on which fuzzy sets are allowed \( m \) distinct membership values. © 1997 Elsevier Science B.V.

Keywords: Fuzzy set; Hausdorff metric; Digital picture; Analysis of algorithms

1. Introduction

As the membership functions of fuzzy sets need not be continuous (nor digitally continuous in the sense of (Rosenfeld, 1986)), the measures of functional distance most often used in topology and analysis are not satisfactory for measuring the distance between fuzzy sets. Let \( Z \) be the set of integers, and consider, e.g., the fuzzy sets \( u \) and \( v \) defined on the digital interval \([a, b]_Z = \{ z \in Z | a < z <= b \} \) by

\[
\begin{align*}
u(a) & = 1, & u(z) & = 0 \quad \text{for } z \neq a, \\
u(a+1) & = 1, & v(z) & = 0 \quad \text{for } z \neq a+1.
\end{align*}
\]

Conventional distance measurements such as the sup norm,

\[
d_{sup}(u, v) = \max \{ d(u(z), v(z)) \mid z \in [a, b]_Z \}.
\]

yield a large, even maximal, value relative to the diameter of the induced metric space, since the maximum possible difference between \( u(z) \) and \( v(z) \) is attained, e.g., at \( z = a \). On the other hand, the images \( u([a, b]_Z) \) and \( v([a, b]_Z) \) are very similar with respect to separation of level sets; a distance function for fuzzy sets based on applying the Hausdorff metric to level sets therefore yields a small distance, relative to the diameter of the induced metric space, between \( u \) and \( v \). Thus, it seems appropriate to use separation of level sets as measured by the Hausdorff metric in order to measure the distance between fuzzy sets. This is the approach of (Chaudhuri and Rosenfeld, 1996).

E-mail: boxer@niagara.edu. Research partially supported by a grant from the Niagara University Research Council.

0167-8655/97/$17.00 © 1997 Elsevier Science B.V. All rights reserved.

PII S0167-8655(97)00006-8
The metrics \( d \) of (Chaudhuri and Rosenfeld, 1996) suffer the undesirable restriction that \( d(u, v) \) is only defined for fuzzy sets \( u \) and \( v \) that have the same maximum value. In this paper, we show how to modify the metrics of (Chaudhuri and Rosenfeld, 1996) to remove this restriction. We also give an algorithm for computing our metric on a 2-dimensional digital picture, with analysis of its time requirements.

2. Preliminaries

2.1. Fuzzy set

Let \( S \) be a non-empty set and let \( f : S \rightarrow [0, 1] \) be a function. The pair \((S, f)\) is called a fuzzy set or a fuzzy subset of \( S \); \( S \) is the support set of \( f \). For \( s \in S \), the value \( f(s) \) is thought of as the “degree of membership” or the “membership value” of \( s \) in \( S \). An ordinary or “crisp” subset \( T \) of \( S \) may be thought of as a fuzzy subset of \( S \) whose membership function \( f \) has image \( \{0, 1\} \), where \( T = f^{-1}(1) \). The notion of a fuzzy set has wide-ranging applications in areas such as probability, neural networks, and computer graphics. In the latter field, for example, the support set \( S \) might be a digital picture, and the value \( f(s) \) might represent the color of the pixel \( s \).

2.2. Hausdorff metric

The Hausdorff distance (Nadler, 1978) is a measure of how well two non-empty compact sets \( A \) and \( B \) in a metric space \( S \) resemble each other with respect to their positions. Let \( d(a, b) \) be a metric for \( S \). We abuse notation and write

\[
d(z, A) = \min \{ d(z, a) \mid a \in A \}.\]

The “nonsymmetric” or “one-way” Hausdorff measure is

\[
H^*(A, B) = \max_{a \in A} d(a, B).
\]

The Hausdorff metric \( H(A, B) \) is defined (Nadler, 1978) by

\[
H(A, B) = \max \{ H^*(A, B), H^*(B, A) \}.
\]

2.3. Fuzzy Hausdorff metrics of Chaudhuri and Rosenfeld

Let \( \mathcal{F} \) be a family of fuzzy sets defined on a non-empty metric space \( S \) such that

- \( \mathcal{F} \) has a member that is not identically 0, and
- for some finite set \( \{t_k\}_{k=1}^m \subset [0, 1], f \in \mathcal{F} \) implies \( f(S) \subset \{t_k\}_{k=1}^m \).

For example, \( S \) might be a 2-dimensional digital picture, and \( \{t_k\}_{k=1}^m \) might be the set of values representing the possible colors of pixels in \( S \).

**Theorem 2.1** (Chaudhuri and Rosenfeld, 1996).

If \( \mathcal{F} \) has the property that every member of \( \mathcal{F} \) attains a maximum value of 1, then the formula

\[
F_0(u, v) = \frac{\sum_{k=1}^m t_k H(u^{-1}([t_k, 1]), v^{-1}([t_k, 1]))}{\sum_{k=1}^m t_k}
\]

is a metric for fuzzy sets in \( \mathcal{F} \).

In the definition of \( F_0 \), the assumption that \( u \) and \( v \) have maximum value of 1 is important, because, e.g., if \( t_j = \max \{u(s) \mid s \in S\} > \max \{v(s) \mid s \in S\} \), then \( v^{-1}([t_j, 1]) \) is an empty set. To overcome this difficulty, Chaudhuri and Rosenfeld (1996) suggest modifying the fuzzy sets under discussion so that they have maximum value of 1, so that the formula of Theorem 2.1 is well-defined. This, however, raises other problems. Modifying the original fuzzy sets may radically alter their natures. Also, a metric \( \rho \) must satisfy the property that \( \rho(x, y) = 0 \) if and only if \( x = y \). The modifications of fuzzy sets \( u, v \) to \( u', v' \) with maximal values of 1 could yield \( u' \approx v' \) identically even though \( u \) and \( v \) are not identical. The latter problem is circumvented by the following.

**Theorem 2.2** (Chaudhuri and Rosenfeld, 1996).

Let \( \varepsilon > 0 \). If \( u \) and \( v \) have maximum value of 1, then the formula

\[
F_1(u, v) = \frac{\sum_{k=1}^m t_k H(u'^{-1}([t_k, 1]), v'^{-1}([t_k, 1]))}{\sum_{k=1}^m t_k} + \varepsilon \frac{\sum_{s \in S} |u(s) - v(s)|}{|S|}
\]

(where \( |S| \) is the cardinality of \( S \)) is a metric for fuzzy sets whose support set is \( S \).
3. Other fuzzy Hausdorff metrics

In this section, we define a metric that modifies $F_0$ so that it is not necessary that $u$ and $v$ take the same maximum, nor even that either have a positive maximum.

We let $S'$ be a non-empty set disjoint from $S$. It is therefore natural to let $u'$ and $v'$ be extensions of $u$ and $v$, respectively, to $S \cup S'$, such that $u'(x) = v'(x) = 0$ for all $x \in S'$; in this way, points of $S'$ have minimal membership in the fuzzy sets defined by the functions $u'$ and $v'$. (For example, if $S$ is the set of pixels of a digital picture, $S'$ might be a border for the picture.) Now, we may modify the formula for $F_0$ by replacing intervals of the form $[t_k, 1]$ by intervals of the form $[0, t_k]$. We also replace the weight factors $t_k$ by weights $1 + t_k$ to ensure that all weights are strictly positive.

In a sense, the change of intervals from $[t_k, 1]$ to $[0, t_k]$ is a change of emphasis from degree of membership to degree of nonmembership; however, if we consider that nonmembership for the fuzzy set whose membership function is $u'$ is the same as membership for the fuzzy set whose membership function is $1 - u'$ (which may be regarded as analogous to a photographic negative), this does not seem to be a radical change of view. We have the following.

**Theorem 3.1.** Let $\mathcal{F}$ be a family of fuzzy sets defined on a non-empty metric space $S$ such that
\begin{itemize}
    \item for some finite set $\{t_k\}_{k=1}^m \subset [0, 1]$, $f \in \mathcal{F}$ implies $f(S) \subset \{t_k\}_{k=1}^m$; and
    \item for every $f \in \mathcal{F}$ and every $t_k$, $f^{-1}([0, t_k])$ is a compact subset of $S$.
\end{itemize}

Then the formula

$$
F_2(u, v) = \frac{\sum_{k=1}^m (1 + t_k) H(u^{-1}([0, t_k]), v^{-1}([0, t_k])))}{\sum_{k=1}^m (1 + t_k)}
$$

is a metric for members of $\mathcal{F}$.

**Proof.** Observe the formula $F_2(u, v)$ is defined, as the sets $u^{-1}([0, t_k])$ and $v^{-1}([0, t_k])$ contain $S'$, hence are non-empty compact subsets of $S$.

1. We must show that $F_2(u, v) \geq 0$ and that $F_2(u, v) = 0$ if and only if $u$ and $v$ are identically equal. It is clear that $F_2(u, v) \geq 0$ for all members of $\mathcal{F}$ and that if $u$ and $v$ are identically equal, then $F_2(u, v) = 0$. If $u$ and $v$ are not identical, there is a $k$ such that $u^{-1}([0, t_k]) \neq v^{-1}([0, t_k])$, and it follows that $F_2(u, v) > 0$.

2. We must show that $F_2(u, v) = F_2(v, u)$ for all $u, v$ in $\mathcal{F}$. This follows from the fact that $H$ is a metric.

3. We must show the triangle inequality,

$$
F_2(u, w) \leq F_2(u, v) + F_2(v, w),
$$

holds for all $u, v, w$ in $\mathcal{F}$. This follows from the fact that $H$ satisfies a triangle inequality.

Thus, $F_2$ is a metric. \( \square \)

In (Chaudhuri and Rosenfeld, 1996), an analog of the metric $F_1$ was developed.

**Theorem 3.2** (Chaudhuri and Rosenfeld, 1996).

Suppose $\mathcal{F}$ is a family of fuzzy sets with support set $S$, where $S$ is an uncountable compact metric space, and every member of $\mathcal{F}$ is a continuous function attaining a maximum value of 1. Then the formula

$$
F_3(u, v) = \int_0^1 \mu H(u^{-1}([\mu, 1]), v^{-1}([\mu, 1])) \, d\mu + \epsilon \frac{\int_S |u(s) - v(s)| \, ds}{\int_S ds}
$$

defines a metric on $\mathcal{F}$.

In a fashion similar to our development of the metric $F_2$ from $F_0$, we define, as above, extensions $u'$ and $v'$ of $u$ and $v$, respectively, and obtain the following result, in which we do not need to assume a fixed maximum value for members of $\mathcal{F}$. The proof is omitted, as it is a straightforward modification of the proof given in (Chaudhuri and Rosenfeld, 1996) of Theorem 3.2.

**Theorem 3.3.** Suppose $\mathcal{F}$ is a family $\mathcal{F}$ of fuzzy sets with support set $S$, where $S \cup S'$ is an uncountable compact metric space, and every member of $S$ is a continuous function. Then the formula

$$
F_4(u, v) = \int_0^1 \mu H(u^{-1}([\mu]), v^{-1}([\mu])) \, d\mu
$$

defines a metric on $\mathcal{F}$. 

4. Algorithm and analysis

In this section, we give an algorithm for computing $F_2(u, v)$, where $u, v$ are fuzzy sets defined on a digital picture $S$ of cardinality $n$. We assume there are positive integers $c, r$ such that $cr = n$ and

\[ S = \{(x, y) \mid x, y \in \mathbb{Z}, 1 \leq x \leq c, 1 \leq y \leq r\} \]

Let

\[ S' = \{(0, c + 1) \times [0, r + 1] \} \]

\[ \cup ([1, c] \times \{0, r + 1\}) \] .

We may think of $S'$ as a border for the digital picture $S$. We assume the set $\{t_k\}_{k=1}^m$ of possible membership values of members of $\mathcal{F}$ is known. We also assume that for each $s \in S \cup S'$, $u'(s)$ and $v'(s)$ may be computed in $\Theta(1)$ time. We have the following.

Theorem 4.1. Suppose the support set $S$ is a digital picture of cardinality $n$ and $S'$ is the border of $S$ as described above. Then $F_2(u, v)$ can be computed in $\Theta(mn)$ time.

Proof. We give the following algorithm.
1. Compute the denominator, $\sum_{k=1}^m (1 + t_k)$, of $F_2(u, v)$. This takes $\Theta(m)$ time.
2. Compute the numerator,

\[ \sum_{k=1}^m (1 + t_k) H(u'^{-1}([0, t_k]), v'^{-1}([0, t_k])) \]

of $F_2(u, v)$. For each $k$, the algorithm of (Shonkwiler, 1989) may be used to compute

\[ H(u'^{-1}([0, t_k]), v'^{-1}([0, t_k])) \]

in $\Theta(n)$ time (this algorithm is presented in (Shonkwiler, 1989) for square pictures, but it trivially modifies to pictures of $r$ rows and $c$ columns, $rc = n$, in $\Theta(n)$ running time). Thus, this step takes $\Theta(mn)$ time.
3. Now compute $F_2(u, v)$ from its numerator and denominator in $\Theta(1)$ time.

Thus, our algorithm takes $\Theta(mn)$ time. □

5. Further remarks

We have modified the metric $F_0$ of (Chaudhuri and Rosenfeld, 1996) to obtain the metric $F_2$. $F_2$ seems to have the same “feel” as $F_0$, but has the important advantages that it does not require the fuzzy sets to which it is applied to have either
- the same maximum values, or even
- a positive maximum value.

On the other hand, the metric $F_0$ has an important advantage over $F_4$: for crisp subsets $u$ and $v$ defined on $S$, $F_0$ coincides with the Hausdorff metric, while $F_2$ does not.

It is easy to see that by replacing the weights $(1 + t_k)$ in $F_2$ by other positive constants $t'_k$, the resulting formula is still a metric. We have chosen to use weights $(1 + t_k)$ in order to preserve the emphasis in $F_0$ on the degree of membership. We did not use weights $t_k$ since we want a positive contribution to the metric for the case $t_k = 0$; otherwise, the resulting formula would give a “distance” of 0 for non-identical crisp sets, hence would not be a metric.

We have given an algorithm for computation of $F_2(u, v)$ that requires $\Theta(mn)$ time on 2-dimensional digital pictures of $n$ pixels each when there is a maximum of $m$ different membership levels for the fuzzy sets $u$ and $v$. This algorithm is easily modified to compute $F_0(u, v)$ or $F_1(u, v)$ in $\Theta(mn)$ time.

We thank an anonymous referee for pointing out an error in the first draft submitted, and for other suggestions that improved the presentation of our results.

References


