A Family of Divergences between Φ-Probabilistic Sets with Application to Handshape Recognition

Juan M. León-Rojas[1], José Moreno[2], Antonio Silva[2], and Montaña Morales[3]

[1] University of Extremadura, Department of Mathematics,
Escuela Politécnica, 10071 Cáceres, Spain
jmleon@unex.es
[2] University of Extremadura, Computer Science Department,
Escuela Politécnica, 10071 Cáceres, Spain
{josemore, agua}@unex.es
[3] Ministery of Education and Culture, I. E. S. Ágora,
Computer Science Department, 10005 Cáceres, Spain
guideca@unex.es

Abstract. We introduce a family of divergences between Φ-probabilistic sets, with real supports. The supports are never unbounded to opposite sides. We start from weighted and percentiled dissimilarities between arbitrary unions of compact intervals of real numbers. As an application we model the problem of the recognition of a handshape as a metric problem between Φ-probabilistic sets. The proposed family of divergences is a suitable solution to this problem of comparing one handshape prototype, a Φ-probabilistic set, with one input handshape, a Φ-fuzzy set.

1 Introduction

The purpose of this work is to introduce some horizontal divergences between Φ-probabilistic sets, horizontal because they are defined from the dissimilarities measured between their α-cuts. The vertical point of view is revised in §2. In §3 we concentrate our efforts in the horizontal point of view. We assume the normality of the fuzzy sets to ensure the non-voidness of all their α-cuts. As the α-cuts of a normal fuzzy set with real support is a compact interval in $\mathbb{R}$, then the first we need is to measure the dissimilarity between compact intervals in $\mathbb{R}$. This is investigated in §3.1 where we motivate and propose several dissimilarities, including one, recursively defined from the dissimilarities measured between some of their subintervals. Following subsections comprise several examples of applications. The extension of the proposed divergence measures to normal but non convex sets is nearly trivial, starting from the corresponding definition of a dissimilarity measure between finite unions of compact intervals in $\mathbb{R}$(cf. §3.3). In §3.4, we apply the former horizontal divergences (cf. §3.2) between probability distributions. In §3.5, we apply the recursive schema (cf. §3.1) over Φ-fuzzy numbers. A brief discussion about a possible mixed divergence is presented in §4. Lastly, in §5, we model the problem of the recognition of a handshape as a

F.J. Ferri et al. (Eds.): SSPR&SPR 2000, LNCS 1876, pp. 777–787, 2000.

metric problem between Φ-probabilistic sets. The proposed family of divergences is a suitable solution to this problem of comparing one handshape prototype, a Φ-probabilistic sets, with one input handshape, a Φ-fuzzy set.

For later use we remember some definitions. Let $\mathcal{D}$ be a universal or reference (crisp) set. A fuzzy set A over $\mathcal{D}$ can be identified with its membership function (mf), A is always and only a function, from $\mathcal{D}$ into $[0,1]$[1]. Given $\alpha \in [0,1]$, the *α-cut* of A is the crisp set ${}^{\alpha}A = \{x : A(x) \geq \alpha\}$ and its *levels set*, $\Lambda(A) = \{\alpha \in [0,1] : A(x) = \alpha\}$. All the α-cuts of a given set A, form a decreasing sequence of nested crisp sets, i. e., $\forall \alpha_1, \alpha_2 \in [0,1]$, $\alpha_1 < \alpha_2 \Longrightarrow {}^{\alpha_1}A \supseteq {}^{\alpha_2}A$. Every fuzzy set can uniquely be represented by the family of all its α-cuts. This is usually referred to as a *decomposition* of the set [2]. In order to do that, let consider, for each $\alpha \in [0,1]$, the fuzzy set ${}_{\alpha}A = \alpha \cdot {}^{\alpha}A$, where ${}^{\alpha}A$ represents its characteristic function (viewed as an special membership function). Every fuzzy set A, is the standard fuzzy union of all the sets ${}_{\alpha}A$, variying α in $\Lambda(A)$ [3]. The *height* of A is $\mathrm{h}(A) = \sup\{A(x) : x \in \mathcal{D}\}$. The *support* of A is the set $\mathrm{supp}(A) = \{x \in \mathcal{D}\colon A(x) > 0\}$. We call *core* of A its 1-cut, $\mathrm{core}(A) = \{x \in \mathcal{D}\colon A(x) = 1\}$. A is called *normal* if $\mathrm{core}(A) \neq \emptyset$ (i. e., $\mathrm{h}(A)$=1), and *subnormal* otherwise. A is *open left* if $\lim_{x\to-\infty} A(x) = 1$ and $\lim_{x\to+\infty} A(x) = 0$; *open right* if $\lim_{x\to-\infty} A(x) = 0$ and $\lim_{x\to+\infty} A(x) = 1$; *open* if it is *open left* and *right*; and *closed* if $\lim_{x\to-\infty} A(x) = \lim_{x\to-\infty} A(x) = 0$. A is *convex* if $\forall \alpha_1, \alpha_2 \in [0,1]$, $\alpha_1 > \alpha_2 \Longrightarrow {}^{\alpha_1}A \subseteq {}^{\alpha_2}A$. A *fuzzy number* is every normal and convex fuzzy set in the real line $\mathbb{R}$. At the sintactic level, probability density functions (pdf) may be consider as fuzzy sets, exactly those which cardinality is one.

2 Vertical Approach

As a general procedure, we can measure the divergence between two real bounded functions f and g, from $\mathcal{D}$ into $[0,1]$, measuring the local dissimilarities at each point x of the common domain $\mathcal{D}$, $\delta(f(x), g(x))$, and then define a divergence $D(f,g)$. To define this D, from the local dissimilarities, is to find a way of aggregating all those local informations. Three ways seem to be obvious: optimistically, $D_{\inf}(f,g) = \inf\{\delta(f(x), g(x)) : x \in \mathcal{D}\}$, pessimistically, $D_{\sup}(f,g) = \sup\{\delta(f(x), g(x)) : x \in \mathcal{D}\}$, or an averaged one, a kind of mean between those extremal cases, $D_{\inf}(f,g) \leq D_{\mathrm{av}}(f,g) \leq D_{\sup}(f,g)$.

This $D_{\mathrm{av}}(f,g)$ could be defined from a φ-mean format [4]. Given a set of n values $X = \{x_1, \ldots, x_n\}$, if φ is a continuous and monotonic function in $[\inf X, \sup X]$, the φ-mean of $x_1, \ldots, x_n$ is defined as the value M_φ such that,

$$\varphi(M_\varphi) = \sum_{i=1}^{n} c_i \varphi(x_i) \tag{1}$$

Obviously, $\inf X \leq M_\varphi \leq \sup X$, and M_φ, actually from $[0,1]^{|\mathcal{D}|}$ into $[0,1]$, is a monotonic increasing function in all its arguments. If $c_i = 1/|\mathcal{D}|$, for all i,

then we are looking for the solution $M_\varphi\{\}$ of

$$\varphi\left(M_\varphi\{\delta(f(x),g(x)) : x \in \mathcal{D}\}\right) = \frac{1}{|\mathcal{D}|}\sum_{x\in\mathcal{D}} \varphi\left(\delta(f(x),g(x))\right) \tag{2}$$

If $\varphi(x) = x^\beta$, M_φ is known as a *generalized mean*, that we denote σ_β,

$$\left(\sigma_\beta\left\{\delta\left(f(x),g(x)\right) : x \in \mathcal{D}\right\}\right)^\beta = \frac{1}{|\mathcal{D}|}\sum_{x\in\mathcal{D}} \left(\delta(f(x),g(x))\right)^\beta \tag{3}$$

and such that, in the countable infinite case, we should be in the space of sumable series of power β, and if $\beta \to \infty$, $\sigma_\beta = \sup X$. If $\delta(x,y) = |x-y|$ we recognize the standardized family of discrete Minkowski distances, for $1 \leq p < \infty$, with $d_\infty(f,g) = \lim_{p\to\infty} d_p(f,g) = \sup\{|f(x)-g(x)| : x \in \mathcal{D}\}$. If $\mathcal{D}= [a,b]$ and f and g are continuous in $\mathcal{D}$, then we can 'add up' all the local dissimilarities,

$$(d_p(f,g))^p = \int_a^b |f(x)-g(x)|^p\, dx \tag{4}$$

where if $p = 1$, the integral represents the total area between the curves.

3 Horizontal Approach

Dubois and Prade [5] proposed a fuzzy-valued comparison index between fuzzy sets,

$$IC(A,B) = \int \alpha / IC({}^\alpha A, {}^\alpha B) \tag{5}$$

defined from the values of the comparison index acting on their α-cuts. This index assumes that the comparison between two fuzzy sets at high membership degrees, needs to weight more than that with lower ones. The proposed divergences can take into account this fact. We assume the normality of the fuzzy sets to ensure the non-voidness of all their α-cuts. As the α-cuts of a normal fuzzy set with real support is an interval, then the first we need is to measure the dissimilarity between intervals.

3.1 Dissimilarity Measures between Intervals of Real Numbers

Let $I = [i_0, i_1]$ and $J = [j_0, j_1]$ two compact intervals of real numbers. One bijection that transform I into J is $f(x) = ((j_0 - j_1)x + i_0 j_1 - i_1 j_0)/(i_0 - i_1)$, Observe that f is a composition of an homothety and a translation, with ratios $(j_0 - j_1)/(i_0 - i_1)$ and $(i_0 j_1 - i_1 j_0)/(i_0 - i_1)$, respectively. Without any loss of generality, we can assume that I and J are subintervals of $[0,1]$. This is because, given a concrete working environment $\mathcal{W}$, there exist a real number k, such that *all* the possible intervals to compare, are subintervals of $[-k,k]$. Thus we

can stantardize by range, $1/\sup_{\mathcal{W}}\{\delta(i_r, j_s)\}$, with $r, s \in \{0, 1\}$, or this k can be generalized to be an inaccessible real number (actual infinity), a real number too large to be described in any floating point number system to be implemented on a computer. Similar sign inaccessible real numbers are incomparable since we cannot know which real number each of them represents. Then, without any loss of generality, we can *only* work with subintervals of $[0, 1]$, because by f, we transform $[-k, k]$ into $[0, 1]$.

α-Percentage Points If $X = [x_0, x_1]$ is a compact interval of real numbers, we define the *α-percentage point* in X, as $x_{(\alpha)} = (1 - \alpha)x_0 + \alpha x_1$, with $\alpha \in [0, 1]$. Although the percentage points may be unequally spaced, we assume, first, that for all $\alpha \in \mathbb{Z}_N^*$, $x_{(\alpha+1)} - x_{(\alpha)}$ is the same value, that is, the percentage points are equally spaced. More general situations will be treated further.

One can think to compare I and J 'to the right', measuring the distance between i_0 and j_0, but then, the dissimilarity will not be sensible to different right queues, al-though it should be the right proceeding if the interval space is $\{[x, 1) : x \in (0, 1)\}$ or if it is lateralized, i. e., if it is $\{[a, x) : x \in (a, 1)\}$, with fixed a, measuring a certain 'adecuation towards a'. The situation is similar if comparing 'to the left', between i_1 and j_1. A first and usual way to mitigate those disadvantages, is to use the following extension to intervals, of the family of Minkowski standarized metrics,

$$\bar{d}_p^{(2)}(I, J) = \left[\frac{1}{2}\left(|i_0 - j_0|^p + |i_1 - j_1|^p\right)\right]^{1/p} \tag{6}$$

for $1 \leq p < \infty$, and $\bar{d}_\infty^{(2)}(I, J) = \max\{|i_0 - j_0|, |i_1 - j_1|\}$. Given I and J, if $|i_0 - j_0| = |i_1 - j_1|$, i. e., if the interval transformer f, defined above, is a translation, then $\bar{d}_p^{(2)}(I, J) = |i_0 - j_0|$, for all $1 \leq p \leq \infty$. There exists three situations, completely different, for which $\bar{d}_p^{(2)}(I, J)$ should be the same. First, if the interval transformer f, defined above, is 'only' a translation; second, if f is a contraction, and third, if f is a lateral dilation, being I and J concentric, in the second case. As the metrics $\bar{d}_p^{(2)}$, only consider local dissimilarities at the endpoints, they assign the same distance value to all the pairs of intervals with equal endpoint-difference. But this does not correspond to which our natural intuition should ever say. For example, if I represents a range request, and J and K are such that $J \subset I$ and $K \not\subset I$, then $d(I, J)$ should must be lower than $d(I, K)$. This is the case if $I = [0, 3]$, $J = [1, 2]$ and $K = [1, 4]$, but $\bar{d}_p^{(2)}(I, J) = \bar{d}_p^{(2)}(I, K) = 1$. A first idea to correct it, is to consider also, the dissimilarities between the mean points of the intervals,

$$\bar{d}_p^{(3)}(I, J) = \left[\frac{1}{3}\left(|i_0 - j_0|^p + |i_{1/2} - j_{1/2}|^p + |i_1 - j_1|^p\right)\right]^{1/p} \tag{7}$$

One obvious reason is that, as I and J are concentrics ($i_{1/2} = j_{1/2}$), we divide by a greater quantity, and then $\bar{d}_p^{(3)}(I, J)$ is lower. For the example

above, $\bar{d}_1^{(3)}(I,J) = 2/3 < \bar{d}_1^{(3)}(I,K) = \bar{d}_2^{(3)}(J,K) = 1$ y $\bar{d}_2^{(3)}(I,J) = 0.8615 < \bar{d}_2^{(3)}(I,K) = 1 < \bar{d}_2^{(3)}(J,K) = 1.291$.

Modified Minkowski Dissimilarities In general, we reinforce the former idea, with the extra information coming from all the local dissimilarities between the corresponding α-percentage points. A first family of modified Minkowski dissimilarities, is defined as

$$d_{(p)}(I,J) = \left(\int_0^1 \left| \Delta_0^{I,J} + \alpha \left(\Delta_1^{I,J} - \Delta_0^{I,J} \right) \right|^p d\alpha \right)^{\frac{1}{p}} \tag{8}$$

where $1 \leq p < \infty$ and $\Delta_k^{I,J} = i_k - j_k$ $(k \in \{0,1\})$. The case $p = 2$ is widely studied by Bertoluzza, Corral and Salas [6]. If $p = \infty$,

$$d_{(\infty)}(I,J) = \sup_{\alpha \in [0,1]} \{|i_{(\alpha)} - j_{(\alpha)}|\}. \tag{9}$$

With respect to the former observations, if f is a translation, the dissimilarities are the same, but in the other two cases, the dissimilarities $d_{(p)}(I,J)$ are monotonic increasing with respect to p, towards $d_{(\infty)}(I,J)$.

Weighted Dissimilarities A weighted dissimilarity between the intervals could also be defined from a φ-mean format (cf. eqn. (1)). In the finite case we define it as the solution ${}^{\varphi}d_{(p)}^{\omega}(I,J)$ of

$$\varphi \left({}^{\varphi}d_{(p)}^{\omega}(I,J) \right) = \sum_{\alpha=1}^{N} \omega(\alpha) \varphi \left(\left| i_{(\alpha)} - j_{(\alpha)} \right| \right) \tag{10}$$

with $\sum_{\alpha=1}^{N} \omega(\alpha) = 1$. If $\varphi(x) = x^p$, another modified Minkowski dissimilarity, that if $1 \leq p < \infty$,

$$d_{(p)}^{\omega}(I,J) = \left(\sum_{\alpha=1}^{N} \omega(\alpha) \left| i_{(\alpha)} - j_{(\alpha)} \right|^p \right)^{\frac{1}{p}} \tag{11}$$

In the continuous case, given I and J, if ω is a normalized Lebesgue weight measure on $([0,1],\mathcal{B}([0,1]))$, and if $\varphi \left(\left| \Delta_0^{I,J} + \alpha \left(\Delta_1^{I,J} - \Delta_0^{I,J} \right) \right| \right)$ is ω-integrable, we define a dissimilarity between I and J as the solution $d_{\omega}^{\varphi}(I,J)$ of the integral equation

$$\varphi \left(d_{\omega}^{\varphi}(I,J) \right) = \int_0^1 \varphi \left(\left| \Delta_0^{I,J} + \alpha \left(\Delta_1^{I,J} - \Delta_0^{I,J} \right) \right| \right) d\omega(\alpha) \tag{12}$$

with $\Delta_k^{I,J} = i_k - j_k$ $(k \in \{0,1\})$, provided the integral exists and φ is continuous and monotonic at its domain. This last expresion depends on the definition of

ω, and there no exists a general solution to the integral proposed. If $\xi(\alpha)$ is the Radon-Nykodim derivate of the measure ω, that is $\xi(\alpha) = d\omega(\alpha)/d\alpha$, then

$$\varphi\left(d_{\omega}^{\varphi}(I,J)\right) = \int_0^1 \xi(\alpha)\varphi\left(\left|\Delta_0^{I,J} + \alpha\left(\Delta_1^{I,J} - \Delta_0^{I,J}\right)\right|\right) d\alpha \tag{13}$$

(α, β)-Percentage Intervals If $X = [x_1, x_2]$ is a compact interval of real numbers, we define the (α, β)-*percentage interval* of X as $I_{(\alpha,\beta)} = \left[x_{(\alpha)}, x_{(1-\beta)}\right]$. Then we consider the following free-form approximation of ξ in $[0, 1]$,

$$\xi(\alpha) = \sum_{k=1}^{M} c_k \chi_{I_k}(\alpha) \tag{14}$$

where $\Gamma_{[0,1]} = \{I_1, \ldots, I_M\}$ is a finite collection of disjoint (α, β)-percentage subintervals of $[0, 1]$, assuming that there are r and s, such that $1 \leq r, s \leq M$ and $i_{r,0} = 0$ and $i_{s,1} = 1$. χ_I is the characteristic function of I, whose area, with respect to a measure μ, is $\mu(I)$, for every interval that contains I. The special case $\chi_{[a,a]}$ is $\delta(x - a)$, the Dirac δ-distribution. Observe that it is necessary a normalization

$$\sum_{k=1}^{M} c_k \mu(I_k) = 1 \tag{15}$$

to include the Minkowski metris d_p as special cases.

Thus, we measure the dissimilarity between I and J, from the dissimilarities between some of their (α, β)-percentage subintervals, proceeding with these in a similar manner, and son on. We have then, a recursive schema

$$\varphi_{y+1}\left(d_{\omega_{y+1}}^{\varphi_{y+1}}\left(S_{y+1,k}^{I}, S_{y+1,k}^{J}\right)\right) = \sum_{r=1}^{N_y} v_r^{y+1} \varphi_{y+1}\left(d_{\omega_y}^{\varphi_y}\left(S_{y,r}^{I}, S_{y,r}^{J}\right)\right) \tag{16}$$

for k, from 1 to N_{y+1}, assuming that $\sum_{r=1}^{N_y} v_r^{y+1} = 1$. The end of the recursion is reached in $P + 1$ steps, assumed that the computation of $d_{\omega_P}^{\varphi_P}\left(S_{P,r}^{I}, S_{P,r}^{J}\right)$, for all $r = 1, \ldots, N_P$, is defined from its α-percentage points, i. e., from $I_{(\alpha,1-\alpha)}$, as in eqn. (10).The functions φ and φ_y are continuous and monotonic at the corresponding intervals.

For example, we can modify eqn. (6) in the following way: let be $I = [i_0, i_1]$, $J = [j_0, j_1]$, $S_{1,0}^{I} = [i_0, i_0 + \varepsilon_{I,0}]$, $S_{1,1}^{I} = [i_1 - \varepsilon_{I,1}, i_1]$, $S_{1,0}^{J} = [j_0, j_0 + \varepsilon_{J,0}]$, $S_{1,1}^{J} = [j_1 - \varepsilon_{J,1}, j_1]$, then a possible dissimilarity is defined as

$$\overline{d}_p^{(2)}(I,J) = \left[\frac{1}{2}\left(\left(\overline{d}_p^{(2)}\left(S_{1,0}^{I}, S_{1,0}^{J}\right)\right)^p + \left(\overline{d}_p^{(2)}\left(S_{1,1}^{I}, S_{1,1}^{J}\right)\right)^p\right)\right]^{\frac{1}{p}} \tag{17}$$

Observe that when we consider a finite collection $\Gamma_I = \{I_1, \ldots, I_M\}$ of (α, β)-percentage subintervals of an interval $I = [i_0, i_1]$, we assume that there exist r and s, such that $1 \leq r, s \leq M$ and $i_{r,0} = i_0$ and $i_{s,1} = i_1$.

3.2 The Horizontal Divergence

In order to define a divergence between two sets A and B, we have only to aggregate the local dissimilarities obtained in the former sections. Using a φ-mean format (cf. eqn. (1)), if A and B are (closed) fuzzy numbers and $d_\alpha(A,B)$ is a dissimilarity measure between the α-cuts of A and B, we propose as divergence the possible solution $D_{\varphi,\phi}(A,B)$ of

$$\varphi\left(D_{\varphi,\phi}(A,B)\right) = \int_0^1 \varphi\left(d_\alpha(A,B)\right) d\phi(\alpha) \tag{18}$$

where ϕ is a Lebesgue measurable function, $\varphi\left(d_\alpha(A,B)\right)$ is ϕ-integrable and the integral above exists.

3.3 Horizontal Divergence between 'Climbing' Fuzzy Sets

An instance of non-convex but normal fuzzy sets is that we call *climbing fuzzy set*, a normal fuzzy set such that *all* the reference points where it reaches a local maximum belongs to its core. Because of the normality, all the α-cuts of A are non void. If A has real support, each of its α-cuts is a finite union of disjoint compact intervals of real numbers, $X{=}I_1 \cup I_2 \cup ... \cup I_n{=}[i_{1,0}, i_{1,1}] \cup [i_{2,0}, i_{2,1}] \cup ... \cup [i_{n,0}, i_{n,1}]$. We can easily extend the dissimilarities defined above to this case. For example, the definition of the *α-percentage point* in X is $x_{(\alpha)} = i_{k,0} + \alpha L - S_{k-1}$, where $S_{k-1}/L \leq \alpha \leq S_k/L$, and $L_h = i_{h,1} - i_{h,0}$, $S_0 = 0$, $S_k = L_1 + L_2 + ... + L_k$ and $L = S_n$.

3.4 Horizontal Divergence between Probability Distributions

As a *probability distribution* is a special case of *open right* fuzzy set, with height 1, we can define an horizontal divergence between two probability distributions P and Q, from the former dissimilarity measure between their α-cuts. If $\arg P(\alpha)$ and $\arg Q(\alpha)$ are real numbers, then the α-cuts are ${}^\alpha P = [\arg P(\alpha), \infty)$ and ${}^\alpha Q = [\arg Q(\alpha), \infty)$. If we note $M_{\arg}(P,Q;\alpha) = \max\left(\arg P(\alpha), \arg Q(\alpha)\right)$ and $m_{\arg}(P,Q;\alpha) = \min\left(\arg P(\alpha), \arg Q(\alpha)\right)$, then the horizontal divergence between P and Q at level α is $d_\alpha(P,Q) = M_{\arg}(P,Q;\alpha) - m_{\arg}(P,Q;\alpha)$. If $\arg P(\alpha)$ is an interval of real numbers, then ${}^\alpha P = [\min \arg P(\alpha), \infty)$, and so in order to compute $d_\alpha(P,Q)$ it is necessary to use a dissimilarity between intervals of real numbers.

A divergence can be defined in a similar way to compare two *open left* fuzzy sets. In the case of and *open* fuzzy set, we assume that its support is not $\mathbb{R}$ (any α-cut is non void), and then ${}^\alpha A = (-\infty, \min \arg A(\alpha)] \cup [\max \arg A(\alpha), +\infty)$.

3.5 Horizontal Divergence between Φ-Fuzzy Numbers

Another example of interest refers to Φ-fuzzy sets. In order to express a greater uncertainty, Sambuc [7] proposed the concept of *Φ-fuzzy set* (or *interval-valued*

fuzzy set), with its membership function, $A : \mathcal{D} \to 2^{[0,1]}$ is defined as $A(x) = [\underline{A}(x), \overline{A}(x)]$, an interval defined by a lower and an upper membership function. We say that a Φ-fuzzy set A is a Φ-fuzzy number if $\underline{A}$ is a fuzzy number. If instead of single intervals, we consider a finite union of them, it is known as P-fuzzy set [8]. For example, denoting $f_{(0)}(x) = \min f(x)$ and $f_{(1)}(x) = \max f(x)$, if A is a Φ-fuzzy number, then its α-cut is a union of three compact intervals ${}^{\alpha}A = [\arg_{(0)} \overline{A}(x), \arg_{(0)} \underline{A}(x)] \cup [\arg_{(0)} \underline{A}(x), \arg_{(1)} \underline{A}(x)] \cup [\arg_{(1)} \underline{A}(x), \arg_{(1)} \overline{A}(x)]$. We can use the former recursive schema to define a divergence between two Φ-fuzzy numbers, weighting more the middle subinterval $[\arg_{(0)} \underline{A}(x), \arg_{(1)} \underline{A}(x)]$ (the least uncertain one) than the ending intervals.

4 Mixed Approach

In the general problem of comparing two fuzzy sets, several situations can be considered. If they have a common support, or the intersection between both supports is large, perhaps the most natural procedure is to measure the dissimilarity between them vertically, but if they have disjoint supports, doing it horizontally seems to be more natural. If the intersection of the supports is not so large, perhaps we have to think in a mixed approach. In such cases, we must aggregate the horizontal and the vertical divergences. For example, if we are interesting in compare two Φ-fuzzy numbers, we can do it whether horizontally or whether vertically, using some interval dissimilarity. Observe that any vertically approach only consider intervals in $[0, 1]$.

5 Application to Handshape Recognition

Our objective here is to model the problem of the recognition of a handshape as a metric problem between Φ-probabilistic sets, i. e., a Φ-fuzzy set such that $\underline{A}$ and $\overline{A}$ are probability functions [9], as part of a wider study of Spanish Sign Language [10]. The data adquisition is made with two mechanical devices, a sensored glove CyberGlove^{TM} (which measures flexion and abduction angles, thumb rotation, palm arch, wrist pitch and wrist yaw) and a 3D sensor Polhemus Isotrack^{TM}.

We call the observed sensor values, the *clues*. In order to make easier the exposition, let assume that the unique sensor devices are those of flexion and that the angles are normalized into $[0, 1]$, where 1 means the quality 'completely flexed'. Then each clue refers to this quality. Thus, our reference set is the finite set of predicates $\mathcal{D} = \{s_1 \equiv MPJ(\mathrm{t}),\ s_2 \equiv IJ(\mathrm{t}),\ s_3 \equiv MPJ(\mathrm{i}),\ s_4 \equiv PIJ(\mathrm{i}),\ s_5 \equiv MPJ(\mathrm{m}),\ s_6 \equiv PIJ(\mathrm{m}),\ s_7 \equiv MPJ(\mathrm{r}),\ s_8 \equiv PIJ(\mathrm{r}),\ s_9 \equiv MPJ(\mathrm{p}),\ s_{10} \equiv PIJ(\mathrm{p})\}$, where the constants are t=thumb, i=index, m=middle, r=ring, and p=pinkie, and the functions are MPJ=MetacarpoPhalangeal Joint, IJ=Interphalangeal Joint, and P=Proximal. The predicates mean, for instance, $s_7 \equiv MPJ(\mathrm{r}) \equiv$ 'ring's MPJ angle is completely flexed'. For example, a näive approximation to the ASL (American Sign Language) handshape 'i' (pinkie extended and the rest completely flexed) is the fuzzy set, in Zadeh's notation $\sum_{i=1}^{10} \alpha_i / s_i$,

where α_9 and α_{10} are approximately 0, and the others are approximately 1. Observe that because of the anatomical configuration of the human hand and the characteristics of 'i' to be recognized, although flexed, it is enough that the thumb is close to the index and the thumb-tip is below the knuckle line.

In general, the handshapes of a sign language have not to be so precise. Fingers do not have to achieve an exact position but an interval of possible positions, for example, signing 'w' in ASL, index, middle and ring are extended, but the angle between the possible directions (viewing the fingers as vectors from knuckles) has not to be exactly zero: differences of about 10-15 or even 20 degrees are oftenly interpreted as 'w'. Thus, a fuzzy set seems to be a suitable representation of the uncertainty for a handshape.

Assume that, in general, we have three finite sets $\mathcal{Q}$, $\mathcal{S}$ and $\mathcal{E}$, of handshapes, sensor devices and experts signers, respectively. Given a handshape q to be learned by the system, for every expert $e \in \mathcal{E}$, q is defined as a set of $|\mathcal{S}|$ probability distributions $D_{q,e} = \{P_{q,e,s} : s \in \mathcal{S}\}$. Once performed q by all of the experts, q is defined as an overset $\{D_{q,e} : e \in \mathcal{E}\}$ of subsets of probability distributions. Observe that if we suppose a finite range for each sensor device (a quantization of $[0, 1]$), the distributions in $D_{q,e}$ are discretes.

Observe also that given an expert e, and a sensor device s, we have a probability mass function $p_{e,s}$. We have assigned, heuristically, to each expert e, a confidence level $\alpha_e \in [0, 1]$, so the $\alpha_e \cdot 100\%$ more probable executions of q are the only accepted (we call the significance α_e-cut of $p_{e,s}$, and we denote it ${}^{\alpha_e}p_{e,s}$, the set that comprises all of them, given e and s).Thus, it seems reasonable to estimate the handshape prototype signed by a given expert e, as the Φ-fuzzy set

$$\hat{q}_e = \sum_{s \in \mathcal{S}} [\underline{s}(e), \overline{s}(e)]/s \tag{19}$$

where the sample range of the sensor device s, given that the expert e has signed the handshape q, is estimated as $[\underline{s}(e), \overline{s}(e)]$, with $\underline{s}(e) = \min\{\underline{s}_i : i = 1, \ldots, n(e, q) \wedge \underline{s}_i \in {}^{\alpha_e}p_{e,s}\}$ and $\overline{s}(e) = \max\{\overline{s}_i : i = 1, \ldots, n(e, q) \wedge \overline{s}_i \in {}^{\alpha_e}p_{e,s}\}$ and $n(e, q)$ is the number of times that e has signed q (the sample size for e and q). Thus, given a set of experts $\mathcal{E}$, once performed all the handshapes by all of them, we have $|\mathcal{Q}|$ *classes of handshapes*, all of them with $|\mathcal{E}|$ elements, the former Φ-fuzzy sets $\hat{q}_e$.

Each execution of a handshape by a signer (the input pattern to the system) is a plain fuzzy set of observed values $\{o(s) : s \in \mathcal{S}\}$ from the sensor devices, although really, because of possible errors in the measure process, we are only sure that the true value belongs to some interval $[o(s) - \Delta_s, o(s) + \Delta_s]$, where Δ_s is an estimation of the error associated to s. Then, each input pattern could be represented as a Φ-fuzzy set.

At this point, the recorded patterns and the input one are represented by Φ-fuzzy sets. The problem of the classification of the input can be solved by classical thecniques as 'nearest neighbours', evaluating dissimilarities between the input and all the pattern in a class, for the different classes. We can use, for example, the former recursive proposal as in §3.5.

5.1 Handshapes as Φ-Probabilistic Sets

For each class of recorded patterns (Φ-fuzzy sets) we define a prototype as a Φ-probabilistic set. Given a probabilistic space $(\Omega, \mathcal{B}, P)$ and $(\Omega_c, \mathcal{B}_c)$ a space of characteristics, we can define a *Φ-probabilistic set* A by a pair of mappings $\underline{A}, \overline{A} : \mathcal{D} \times \Omega \to \Omega_c$, where $\underline{A}, \overline{A}(x, .)$ are measurables in $(\mathcal{B}, \mathcal{B}_c)$, for all $x \in \mathcal{D}$.

Given q, we define a prototype as the Φ-probabilistic set $\underline{q}, \overline{q} : \mathcal{S} \times \Omega \to \Omega_c$, with $\Omega = \{\{[\underline{s}(e), \overline{s}(e)] : s \in \mathcal{S}\} : e \in \mathcal{E}\}$ and $\Omega_c = [0, 1]$. Thus, the problem of the recognition of a handshape is modeled as a metric problem between Φ-probabilistic sets. The family of divergences defined above is a suitable solution to the problem of comparing one handshape prototype, a Φ-probabilistic set, with one input handshape, a Φ-fuzzy set. A Φ-fuzzy set is a Φ-probabilistic set such that all lower and upper probabilities are Dirac deltas. Given two Φ-probabilistic sets A and B, we can calculate a divergence between them in several ways, for example, we can compute

$$\mathcal{O}_{|\mathcal{S}|} \left\{ \mathcal{O}_2 \left(D(\underline{A}(s), \underline{B}(s)), D(\overline{A}(s), \overline{B}(s)) \right) : s \in \mathcal{S} \right\} \tag{20}$$

where D is a divergence between probability distributions and $\mathcal{O}_{|\mathcal{S}|}$ and $\mathcal{O}_2$ are aggregation operators of arities $|\mathcal{S}|$ and 2.

An alternative is based upon the definition of expected set of a Φ-probabilistic set. Given a Φ-probabilistic set A, we can average over Ω, obtaining its lower and upper mean-value membership functions. The *expected set* (a Φ-fuzzy set) of A is defined as

$$E(A) = \left[\underline{E(A)}(s), \overline{E(A)}(s) \right] = \left[\int_\Omega \underline{A}(s, \omega) dP(\omega), \int_\Omega \overline{A}(s, \omega) dP(\omega) \right] \tag{21}$$

Given two Φ-probabilistic sets A and B, we can demote them to their expected Φ-fuzzy sets and compute

$$\mathcal{O}_{|\mathcal{S}|} \left\{ D(E(A)(s), E(B)(s)) : s \in \mathcal{S} \right\} \tag{22}$$

where D is a dissimilarity between intervals of real numbers.

References

1. Bezdek, J.: Fuzzy models. What are they, and why? IEEE Transactions on Fuzzy Systems **1**(1) (1993) 1-6
2. Zadeh, L. A.: Similarity relations and fuzzy orderings. Information Sciences **3**(2) (1971) 177-200
3. Klir, G. J., Yuan, B.: Fuzzy Sets and Fuzzy Logic. Theory and Applications. Prentice Hall PTR, Upper Saddle River, New Jersey (1995)
4. Calot, G.: Cours de Statistique Descriptive. Dunod, Paris (1965)
5. Dubois, D., Prade, H.: A unifying view of comparison indices in a fuzzy set-theoretic framework. In: Yager, R. R. (ed.): Fuzzy Set and Possibility Theory. Recent Developments. Pergamon Press, New York (1982)

6. Bertoluzza, C., Corral, N., Salas, A.: On a new class of distances between fuzzy numbers. Mathware & Soft Computing **2** (1995) 71-84
7. Sambuc, R. Fonctions Φ-Floues. Application à l'Aide au Diagnostic en Pathologie Thyroïdienne. PhD Thesis, Faculty of Medicine of Marseille (1975)
8. Kaufmann, A.: Introduction to the Theory of Fuzzy Subsets. Academic Press, New York (1975)
9. Kaufmann, A.: Les Expertons. Hermes, Paris (1987)
10. Moreno J., León-Rojas, J. M., Silva, A.: Sistema de traducción automática del Lenguaje de Signos Español al Español oral. Novática **136** (1998) 60-64