Similarity Measures

Simone Santini, *Member*, *IEEE*, and Ramesh Jain, *Fellow*, *IEEE*

Abstract—With complex multimedia data, we see the emergence of database systems in which the fundamental operation is *similarity assessment*. Before database issues can be addressed, it is necessary to give a definition of similarity as an operation. In this paper, we develop a similarity measure, based on fuzzy logic, that exhibits several features that match experimental findings in humans. The model is dubbed *Fuzzy Feature Contrast* (FFC) and is an extension to a more general domain of the Feature Contrast model due to Tversky. We show how the FFC model can be used to model similarity assessment from fuzzy judgment of properties, and we address the use of fuzzy measures to deal with dependencies among the properties.

Index Terms—Similarity measures, content-based retrieval, image databases, perceptual similarity, image distances.

1 Introduction

Comparing two images, or an image and a model, is the fundamental operation for many Visual Information Retrieval systems. In most systems of interest, a simple pixel-by-pixel comparison won't do: The difference that we determine must bear some correlation with the perceptual difference of the two images or with the difference between two adequate semantics associated to the two images.

Measuring meaningful image similarity is a dichotomy that rests on two elements: finding a set of features which adequately encodes the characteristics that we intend to measure and endowing the feature space with a suitable metric. Since the same feature space can be endowed with an infinity of metrics, the two problems are by no means equivalent nor does the first subsume the second.

In this paper, we consider the problem of measuring dissimilarities in feature spaces. In a number of cases, after having selected the right set of features and having characterized an image as a point in a suitable vector space, researchers make some uncritical and unwarranted assumptions about the metric of the space. Typically, the feature space is assumed to be Euclidean.

We set out to analyze alternatives to this assumption. In particular, we will analyze some similarity measures proposed in the psychological literature to model human similarity perception and will show that all of them challenge the Euclidean distance assumption in nontrivial ways.

We will consider the problem of *(dis)similarity measurement*, as opposed to *matching*. Matching and dissimilarity measurement are not seldom based on the same techniques, but they differ in emphasis and applications. Matching techniques are developed mostly for recognition of objects under several conditions of distortion [22]. Similarity measures, on the other hand, are used in applications like image databases, in which the query image is just a very partial model of the user's desires and the user looks for for images similar, according to some defined criterion, to it [1]. In *query by example*, the user selects an image, or draws a sketch, that reminds her in some way of the image she wants to retrieve. Images similar to the example according to the given criteria are retrieved and presented.

In a typical matching application, we expect a comparison to be successful for images very close to the model and unsuccessful for images different from the query. The degree of similarity of images different from the model is of no interest to us as long as it remains below a suitable acceptance threshold. On the other hand, database applications require a similarity measure that will accurately predict perceptual similarity for all images "reasonably" similar to the query.

This paper presents and analyzes various definitions of similarity measures for feature spaces. We will specifically consider the determination of similarity between images, but the measures that we present apply in more general situations. It is obviously impossible to decouple the choice of the similarity measure from the choice of features. In this paper, however, we will leave the features in the background. There is an extensive literature that deals with the choice of features for most problems of interest and to which we refer the reader [8], [9]. We are interested in finding characteristics of the distance measure that are relatively independent of the choice of the feature space.

This paper is organized as follows. Section 2 is an overview of psychological models of similarity. Section 3 introduces our Fuzzy Feature Contrast model, which is the extension of one of the psychological models from Section 2. Section 4 presents some evaluation of the model. Conclusions are drawn in Section 5.

2 Similarity Theories

In this section, we present some results on human similarity judgment introduced by psychologists and discuss merits and flaws of the various approaches. We try to put all these theories in perspective and collect them in a unified framework.

- *The authors are with the Department of Electrical and Computer Engineering, University of California, San Diego, La Jolla, CA 92093-0407. E-mail: {ssantini, jain}@ece.ucsd.edu.*

Manuscript received 12 May 1998; revised 22 Apr. 1999.
Recommended for acceptance by A. Smeulders.
For information on obtaining reprints of this article, please send e-mail to: tpami@computer.org, and reference IEEECS Log Number 107668.

The most important concept to do so is that of *geometric distance* and the related distance axioms. Theories differ in the way they deal with the properties of geometric distance and by the number and nature of distance axioms they accept or refuse. The next subsection discusses the distance axioms from the perspective of similarity measurements.

2.1 The Metric Axioms

A number of similarity measures proposed in the literature explain similarity (or, more properly, *dissimilarity*) as a distance in some suitable feature space[1] that is assumed to be a metric space.

A distinction is made between *perceived similarity*, d, and *judged similarity*, δ [2]. If A and B are the representations of the stimuli a and b in the feature space, then $d(A,B)$ is the perceptual distance between the two, while the judged distance is

$$\delta(A,B) = g[d(A,B)], \tag{1}$$

g being a suitable monotonically nondecreasing function of its argument. Note that only the judged distance δ is accessible to experimentation.

Stimuli are represented as points in a metric space, and $d(A,B)$ is the distance function of this space ([20], [24]). This model postulates that the perceptual distance d satisfies the metric axioms, the empirical validity of which has been experimentally challenged by several researchers.

The first requirement for a distance function is that

$$d(A,A) = d(B,B) \tag{2}$$

for all stimuli (constancy of self-similarity.) This hypothesis can be tested using the judged similarity since it implies $\delta(A,A) = \delta(B,B)$. The constancy of self-similarity has been refuted by Krumhansl [11].

A second axiom of the distance model is minimality:

$$d(A,B) \geq d(A,A). \tag{3}$$

Again, this hypothesis is open to experimental investigation since, due to the monotonicity of the relation between d and δ, it implies $\delta(A,B) \geq \delta(A,A)$. Tversky [25] argued that this assumption is violated in some recognition experiments.

A third axiom states that the distance between stimuli is symmetrical:

$$d(A,B) = d(B,A). \tag{4}$$

Just as in the previous cases, this axiom is subject to experimental investigation since it implies $\delta(A,B) = \delta(B,A)$. A number of investigators have attacked this assumption with direct similarity experiments [16] and observing asymmetries in confusion matrices [17]. This phenomenon has been often attributed to the different "saliency" or "goodness of form" of the stimuli. In general, the less salient stimulus is more similar to the more salient (more prototypical) than the more salient stimulus is similar to the less salient [25].

The final metric axiom is the triangle inequality:

$$d(A,B) + d(B,S_C) \geq d(A,S_C). \tag{5}$$

Epistemologically, this is the weakest axiom. The functional relation between d and δ does not guarantee that satisfaction or violation of the triangular inequality for d will translate into a similar property for δ.

The ordinal relation between distances is invariant with respect to all the transformations of the type (1) if g is monotonically increasing. A consequence of this is that the triangular inequality cannot be tested based on ordinal measurements only. It is, however, generally acknowledged that, at least for some types of stimuli, the triangular inequality does not hold [2], [26].

Tversky and Krantz [27] proved that if the distance axioms are verified and the distance is additive along straight lines in the feature space, then d is a Minkowski distance, that is, a distance of the form:

$$d_p(A,B) = \left[\sum_i (A_i - B_i)^p\right]^{\frac{1}{p}}, \tag{6}$$

where $A = \{A_1, \ldots, A_N\}$, $B = \{B_1, \ldots, B_N\}$, and $p > 0$ is a constant which characterizes the distance function.

From these notes, it seems like the situation for geometric models is quite desperate: Of the four basic axioms of the distance function, two are questionable, one is untenable, and the fourth is not ascertainable.

In spite of these problems, metric models are widely used in psychology, with some adjustments to account for the failure of the distance axioms.

2.1.1 The Debatable Euclidean Nature of Perception

In a very influential 1950 paper [3], Attneave investigated the perception of similarity among a group of rectangles that were allowed to change along two dimensions: area and tilt. The results were inconsistent with the Euclidean model of distance, but partial agreement was found with a *city-block* distance model of the type

$$d(A,B) = |A_1 - B_1| + |A_2 - B_2|, \tag{7}$$

where the two dimensions of the feature space represent area and tilt angle. Attneave found some discrepancy in the predictions of the model, which he attributed to nonlinearities in the feature space.

An important class of metric models was introduced by by Thurstone [23] and Shepard [21]. Shepard's model is based on *generalization* data:[2] Given a series of stimuli S_i and a corresponding series of learned responses R_i, the similarity between S_i and S_j (in the absence of any bias) is related to the probability that the stimulus S_i elicits the response associated with stimulus S_j:

$$p_{ij} = \mathbb{P}[R_j | S_i].$$

1. The (metric or otherwise) space in which the stimuli are represented is referred to using a number of different names, not necessarily equivalent, from *perceptual space* to *psychological space*. We will adhere to the generic name *feature space*.

2. The term generalization is used here in a slightly different way than in most Machine Learning papers. In ML, generalization usually means a correct inference whereby the response appropriate in a given situation is extended to cover other situations for which that response is suitable. In Shephard's papres, generalization refers to the incorrect extension of a response from the stimulus for which it was intended to other similar stimuli.

Shepard does not work directly with these quantities, but uses normalized and symmetric *generalization data*, defined as:

$$g_{ij} = \left[\frac{p_{ij}p_{ji}}{p_{ii}p_{jj}}\right]^{\frac{1}{2}}. \quad (8)$$

The model assumes that the generalization data are generated as:

$$g_{ij} = g(d(S_i, S_j)), \quad (9)$$

where g is the *generalization function* and d is a suitable *perceptual distance* between the two stimuli.

Shepard assumed that there exists, for each type of stimulus, a suitable underlying feature space such that 1) the function g is universal (it has the same form for all the types of stimuli) and 2) the function d is a metric. Note that, without the second requirement, the condition can be trivially satisfied for any monotonically decreasing function $g : \mathbb{R} \to [0, 1]$.

If we assume that the function g is monotonic, then, from the generalization data g_{ij}, it is possible to derive the ordering of the stimuli in the perceptual space with respect to any arbitrary reference. Shepard uses ordering data and nonmetric multidimensional scaling [24] to determine the lowest dimensional metric space that can explain the data. He assumes this space as the feature space for the model. There is good agreement with the experimental data if the feature space has a Minkowski metric (defined in (6)) and the generalization function is exponential:

$$g(d) = \exp(-d^{\tau}). \quad (10)$$

One important observation, at the core of Shepard's 1987 paper [21], is that, given the right feature space, the function g is universal, that is, the same exponential behavior (with different values of the parameter τ in (10)) can be found in the most diverse situations, ranging from visual stimuli to similarity of pitch in sounds.

A relevant qualitative characteristic of the model is that, as two stimuli grow apart in the feature space, the dissimilarity $1 - g(d)$ does not increase indefinitely, but it flattens out to a finite limit. A detailed discussion of of the properties of the Thurstone-Shepard model can be found in [7].

2.2 Abandoning the Distance Axioms

The distance axioms seem to provide an unnecessarily rigid system of properties for similarity measures. In particular, it seems epistemologically futile to impose on the perceptual distance d some properties—like the triangle inequality—that may fail to translate into similar properties of the judged similarity δ and are therefore beyond experimental validation. We propose the following definition regarding the epistemologically valid properties for perceptual distance functions:

Definition 1. *Let $\mathcal{D}$ be the class of monotonically increasing functions from $\mathbb{R}$ to $\mathbb{R}$. A logic predicate P over the distance functions d is an* ordinal property *if, for all $g \in \mathcal{D}$, $Pd \Rightarrow P(g \circ d)$.*

Tversky and Gati [26] identified three ordinal properties and used them to replace the metric axioms in what they call a *monotone proximity structure*. Suppose, for the sake of simplicity, that the feature space has two dimensions x and y and let $d(x_1y_1, x_2y_2)$ be the perceived distance between the stimuli $A = (x_1, y_1)$ and $B = (x_2, y_2)$. A monotone proximity structure is characterized by three properties:

Dominance:

$$d(x_1y_1, x_2y_2) > max\{d(x_1y_1, x_1y_2), d(x_1y_1, x_2y_1)\},$$

i.e., the two-dimensional dissimilarity exceeds both one-dimensional projections of that distance.

Consistency: For all x_1, x_2, x_3, x_4 and y_1, y_2, y_3, y_4

$$d(x_1y_1, x_2y_1) > d(x_3y_1, x_4y_1) \Longleftrightarrow d(x_1y_2, x_2y_2) > d(x_3y_2, x_4y_2) \quad (11)$$

and

$$d(x_1y_1, x_1y_2) > d(x_1y_3, x_1y_4) \Longleftrightarrow d(x_2y_1, x_2y_2) > d(x_2y_3, x_2y_4), \quad (12)$$

that is, the ordinal relation between dissimilarities along one dimension is independent of the other coordinate.

To introduce the third property, we give the following definition:

Definition 2. *If $d(x_1y, x_3y) > \max\{d(x_1y, x_2y), d(x_2y, x_3y)\}$, then x_2 is said to be* between *x_1 and x_3 and we write $x_1|x_2|x_3$.*

Note that, in view of consistency, "betweenness" is well defined since it is independent of the coordinate y that appears in the definition.

The third property of a monotone proximity structure is the following:

Transitivity: If $x_1|x_2|x_3$ and $x_2|x_3|x_4$, then it is $x_1|x_2|x_4$ and $x_1|x_3|x_4$.

This framework is more general than the geometric distance: While all distance measures satisfies dominance, consistency, and transitivity, not all the proximity structures satisfy the distance axioms. Dominance is a weak form of the triangle inequality that applies along the coordinate axes. Consistency ensures that certain ordinal properties related to the ordering of the features x do not change when y is changed (see [18] for details.) Transitivity ensures that the "in between" relation behaves as in the metric model, at least when moving along the axes of the feature space.

Note that, in the Euclidean model—which is isotropic—every property holds (or does not hold) for a series of collinear points irrespective of the direction of the line that joins them. In measuring the perceptual distance, the directions of the feature axes have a special status.

Most of the distance measures proposed in the literature, as well as the feature contrast model, predict that dominance consistency and transitivity hold.

To help discriminate among the different models, Tversky and Gati proposed a fourth ordinal axiom that they call the *corner inequality*. If $x_1|x_2|x_3$ and $y_1|y_2|y_3$, the corner inequality holds if

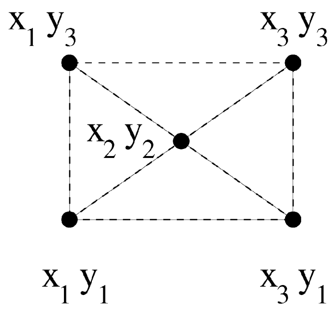

Fig. 1. The corner inequality: The "corner" path $x_1y_1 \to x_3y_1 \to x_3y_3$ is longer than that path $x_1y_1 \to x_2y_2 \to x_3y_3$ whenever x_2y_2 is inside the rectangle.

$$d(x_1y_1, x_3y_1) > d(x_1y_1, x_2y_2)$$

and

$$d(x_3y_1, x_3y_3) > d(x_2y_2, x_3y_3) \tag{13}$$

or

$$d(x_1y_1, x_3y_1) > d(x_2y_2, x_3y_3)$$

and

$$d(x_3y_1, x_3y_3) > d(x_1y_1, x_2y_2). \tag{14}$$

From Fig. 1, it is easy to see that the corner equality holds if the "corner" path from x_1y_1 to x_3y_3 is longer than the diagonal path. Minkowski metrics satisfy the corner inequality, so observed violations of the corner inequality would falsify models based on Minkowski metrics. Tversky and Gati present evidence that, under certain conditions, experiments show violations of the corner inequality, thus seemingly invalidating most geometric models of similarity.

2.2.1 Set-Theoretic Similarity

In a 1977 paper [25], Tversky proposed his famous *feature contrast model*. Instead of considering stimuli as points in a metric space, Tversky characterized them as sets of binary features. In other words, a stimulus a is characterized by the set A of features that the stimulus *possesses*. Equivalently, a feature set is the set of logic predicates which are true for the stimulus in question. Let a, b be two stimuli, A, B the respective sets of features, and $s(a, b)$ a measure of the similarity between a and b. Tversky's theory is based on the following assumptions:

Matching: $s(a, b) = F(A \cap B, A - B, B - A)$.

Monotonicity: $s(a, b) > s(a, c)$ whenever $A \cap C \subseteq A \cap B$, $A - B \subseteq A - C$, $B - A \subseteq C - A$.

A function that satisfies matching and monotonicity is called a *matching function*. Let the expression $F(X, Y, Z)$ be defined whenever there are A, B such that $X = A \cap B$, $Y = A - B$, $Z = B - A$. Define $V \approx W$ if there exist X, Y, Z such that one or more of the following holds:

$$\begin{aligned} F(V, Y, Z) &= F(W, Y, Z) \\ F(X, V, Z) &= F(X, W, Z) \\ F(X, Y, V) &= F(X, Y, W). \end{aligned} \tag{15}$$

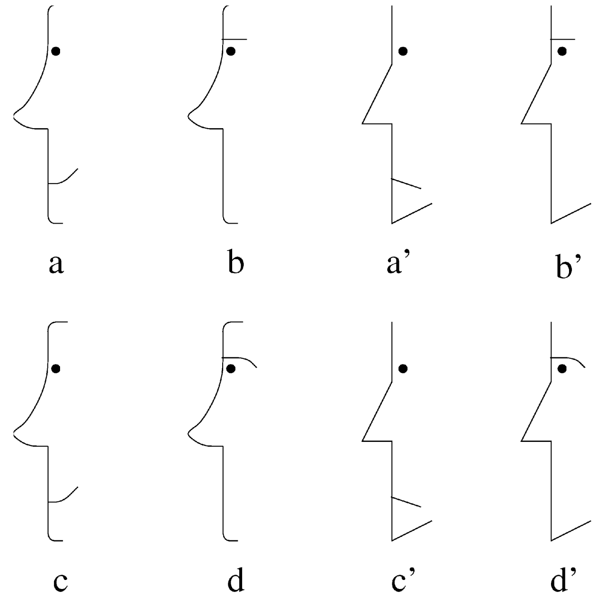

Fig. 2. An example of independence. If a and b are considered more similar than a' and b', then c and d will appear more similar than c' and d'.

The pairs of stimuli (a, b) and (c, d) are said to agree on one (two, three) components whenever one (resp. two, three) of the following hold:

$$\begin{aligned} (A \cap B) &\approx (C \cap D) \\ (A - B) &\approx (C - D) \\ (B - A) &\approx (D - C). \end{aligned} \tag{16}$$

Based on these definitions, Tversky postulates a third property of the similarity measure:

Independence: Suppose the pairs (a, b) and (c, d), as well as the pairs (a', b') and (c', d'), agree on the same two components, while the pairs (a, b) and (a', b'), as well as (c, d) and (c', d'), agree on the remaining (third) component. Then:

$$s(a, b) \geq s(a', b') \Longleftrightarrow s(c, d) \geq s(c', d'). \tag{17}$$

We refer to [25] for details. An example of independence is in Fig. 2. In this case, the independence property states that if (a, b) are "closer" than (c, d), then (a', b') are "closer" than (c', d'). This hypothesis, with some caveat about the selection of features, can be checked experimentally.

The main result of Tversky's paper is the following *representation theorem*:

Theorem 1. *Let s be a similarity function for which matching, monotonicity and independence hold. Then, there are a similarity function S and a nonnegative function f and two constants $\alpha, \beta \geq 0$ such that, for all stimuli a, b, c, d:*

- $S(a, b) \geq S(c, d) \Longleftrightarrow s(a, b) \geq s(c, d)$,
- $S(a, b) = f(A \cap B) - \alpha f(A - B) - \beta f(B - A)$.

This result implies that any similarity ordering that satisfies matching, monotonicity, and independence can be obtained using a linear combination (contrast) of a function of the common features $(A \cap B)$ and of the distinctive features ($A - B$ and $B - A$.) This representation is called the *contrast model*.

This model can account for violation of all the geometric distance axioms. In particular, $S(a, b)$ is asymmetric if $\alpha \neq \beta$. If $S(a, b)$ is the answer to the question "how is a similar to b?" then, when making the comparison, subjects focus more on the features of a (the subject) than on those of b (the referent.) This corresponds to the use of Tversky's measure with $\alpha > \beta$: In this case, the model predicts

$$S(a, b) > S(b, a) \text{ whenever } f(A) < f(B). \tag{18}$$

This implies that the direction of the asymmetry is determined by the relative "salience" of the stimuli: If b is more salient than a, then a is more similar to b than vice versa. In other words, the variant is more similar to the prototype than the prototype to the variant, a phenomenon that Tversky confirmed experimentally. In addition, the feature contrast model accounts for violation of the corner inequality.

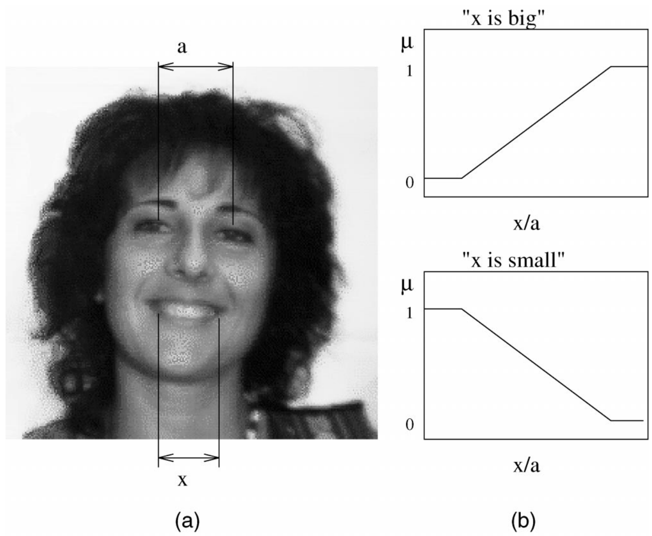

Fig. 3. Determination of the truth value of the predicates "the mouth is wide" and "the mouth is narrow." The width of the mouth x is measured and normalized with respect to the distance between the eyes a. Then, two membership functions are used to determine the truth value of the two predicates.

3 Fuzzy Set-Theoretic Measures

Tversky's experiments showed that the feature-contrast model has a number of desirable properties; most noticeably, it explains violation of symmetry and of the corner equality.

One serious problem for the adoption of the feature-contrast model in visual information systems is its characterization of features. In Tversky's theory, each stimulus is characterized by the presence or absence of features. This convention forces Tversky to adopt complex mechanisms for the representation of numerical quantities. For instance, positive quantities—such as a length—are discretized into a sequence l_i and represented as a collection of feature sets such that if $l_1 < l_2 \cdots < l_n$, then $A_1 \subset A_2 \cdots \subset A_n$. Quantities that can be either positive or negative are represented by even more complex constructions.

In computer vision, the assumption of binary features would thus leave us with the problem of evaluating logic predicates based on some continuous and noisy measurements, yielding brittle and unreliable features.

In the next subsection, we introduce the use of fuzzy predicates in the feature contrast model. The use of fuzzy logic will allow us to extend Tversky's results to situations in which modeling by enumeration of features is impossible or problematic.

Not all the stimuli influence similarity perception according to the same mechanism [2]. Tversky's feature contrast model applies to a particular type of features: those that can be expressed as *predicates* over the stimuli domain. In this section, we will consider only this type of features. A unification of all types of stimuli in a geometric framework can be found in [19].

3.1 Fuzzy Features Contrast Model

Consider a typical task in computer vision: assessing the similarity between faces. A face is characterized by a number of features of different types but, for the following discussion, we will only consider *geometric features* like the size of the mouth, the shape of the chin, and so on.

A predicate like *the mouth of this person is wide* can be modeled as a fuzzy predicate whose truth is based on the measurement of the width of the mouth. For instance, we can measure the width of the mouth x in Fig. 3a and use two *truth functions* (see below) like those in Fig. 3b to determine the truth value of the predicates "the mouth is wide" and "the mouth is narrow."

In general, we have an image I and a number of measurements ϕ_i on the image. We want to use these measurements to assess the truth of *n fuzzy predicates*. Some care must be taken to define the truth value of a fuzzy predicate. We use the following definition:

Definition 3. *Let Ω be a set and $\phi : \Omega \to \mathbb{R}^m$ a set of measurements on the elements of Ω. Let $P\omega$ be a predicate about the element $\omega \in \Omega$. The* truth *of the predicate $P\omega$ is*

$$T(P\omega) = \mu(\phi(\omega))$$

with $\mu : \mathbb{R}^m \to [0, 1]$.

In the example above, for instance, we say that the truth value of the predicate "The mouth of X is wide" depends on measurements of the face (viz. the measurement of the mouth width).

From the measurements ϕ, we derive the truth values of p fuzzy predicates, and collect them into a vector:

$$\mu(\phi) = \{\mu_1(\phi), \ldots \mu_p(\phi)\}. \tag{19}$$

We call $\mu(\phi)$ the (fuzzy) set of *true predicates* on the measurements ϕ. The set is fuzzy in that a predicate P_j belongs to $\mu(\phi)$ to the extent $\mu_j(\phi)$.

In order to apply the feature contrast model to the fuzzy sets $\mu(\phi)$ and $\mu(\psi)$ of the predicates true for the measurements ϕ and ψ, we need to choose a suitable salience function f and compute the fuzzy sets $\mu(\phi) \cap \mu(\psi)$, $\mu(\phi) - \mu(\psi)$, and $\mu(\psi) - \mu(\phi)$.

We assume that the saliency of the fuzzy set $\mu = \{\mu_1 \ldots \mu_p\}$ is given by its cardinality:

$$f(\mu) = \sum_{i=1}^{p} \mu_i. \tag{20}$$

The intersection of the sets $\mu(\phi)$ and $\mu(\psi)$ is defined in the tradition way:

$$\mu_\cap(\phi, \psi) = \{\min\{\mu_1(\phi), \mu_1(\psi)\}, \dots \min\{\mu_p(\phi), \mu_p(\psi)\}, \}. \tag{21}$$

The difference between two fuzzy sets A and B is traditionally defined as $\max\{\mu_A, 1 - \mu_B\}$. This definition, however, leads to some undesired effects [18] that can be avoided by requiring that the relation $\forall A(A - A = \emptyset)$ continue to hold in our fuzzy domain. A possible definition that makes the relation true is:

$$\mu_-(\phi, \psi) = \{\max\{\mu_1(\phi) - \mu_1(\psi), 0\}, \dots \max\{\mu_p(\phi) - \mu_p(\psi, 0\}, \}. \tag{22}$$

With these definitions, we can write Tversky's similarity function between two fuzzy sets $\mu(\phi)$ and $\mu(\psi)$ corresponding to measurements made on two images as:

$$S(\phi, \psi) = \sum_{i=1}^{p} \max\{\mu_i(\phi), \mu_i(\psi) - \alpha \sum_{i=1}^{p} \max\{\mu_i(\phi) - \mu_i(\psi), 0\} - \beta \sum_{i=1}^{p} \max\{\mu_i(\phi) - \mu_i(\psi), 0\}. \tag{23}$$

The Tversky dissimilarity is defined as

$$D(\phi, \psi) = p - S(\phi, \psi). \tag{24}$$

We refer to the model defined by (23) and (24) as the *Fuzzy Features Contrast* (FFC) model.

It is easy to see that the fuzzy feature contrast model can be asymmetric (if $\alpha \neq \beta$.) It is also easy to find an example of violation of the corner inequality. Consider Fig. 1 with $x_1 = y_1 = 0$, $x_2 = y_2 = \frac{1}{4}$, and $x_3 = y_3 = 1$. Let the membership function in the FFC model be

$$\mu(x) = \begin{cases} 0 & \text{if } x < 0 \\ x & \text{if } 0 \leq x \leq 1 \\ 1 & \text{if } x > 1 \end{cases} \tag{25}$$

Then, we have:

$$\begin{aligned} d(x_1 y_1, x_3 y_1) &= d((0,0),(1,0)) = \beta \\ d(x_1 y_1, x_2 y_2) &= d((0,0),(\frac{1}{4},\frac{1}{4})) = \frac{\beta}{2} \\ d(x_3 y_1, x_3 y_3) &= d((1,0),(1,1)) = \beta - 1 \\ d(x_2 y_2, x_3 y_3) &= d((\frac{1}{4},\frac{1}{4}),(1,1)) = \frac{3\beta - 1}{2}. \end{aligned}$$

Condition (13) is violated if $\beta > \frac{\beta}{2}$ and $\beta - 1 < \frac{3\beta - 1}{2}$, while condition (14) is violated if $\beta > \frac{3\beta-1}{2}$ and $\beta - 1 < \frac{\beta}{2}$. Thus, the corner inequality is violated for $\beta < 1$.

A property similar to the representation theorem can be proven for the fuzzy case. Let $y \in \mathbb{R}^n$, and

$$y \xrightarrow{i} x = (y_1, \dots, y_{i-1}, x, y_{i+1}, \dots, y_n). \tag{26}$$

Then, the following theorem holds:

Theorem 2. *Let $F : \mathbb{R}^3 \to \mathbb{R}$ be an analytic function such that the following properties hold:*

1. *$F(x, y, z)$ is monotonically nondecreasing in x and monotonically nonincreasing in y and z. The partial derivatives of F are nonzero almost everywhere.*
2. *For all $i = 1, 2, 3$, if $\forall j \neq i, s_j = s'_j, t_j = t'_j$, $s_i = t_i$, $s'_i = t'_i$, then $F(s) \leq F(t) \Leftrightarrow F(s') \leq F(t')$.*
3. *For all s, the sets $\{t | F(s) \geq F(t)\}$ and $\{t | F(t) \geq F(s)\}$ are closed in the product topology of $\mathbb{R}^3$.*
4. $$\begin{aligned} F(y \xrightarrow{i} x_1) - F(y \xrightarrow{i} x_2) &= F(y \xrightarrow{i} z_1) - F(y \xrightarrow{i} z_2) \\ &\Leftrightarrow \forall j \exists u : \\ &F(u \xrightarrow{i} x_1) - F(u \xrightarrow{i} x_2) \\ &= F(u \xrightarrow{i} z_1) - F(u \xrightarrow{i} z_2). \end{aligned}$$

Then, there are functions G and $f : \mathbb{R} \to \mathbb{R}$ such that

$$G(x, y, z) = \alpha f(x) + \beta f(y) + \gamma f(z) \tag{27}$$

and $F(x, y, z) \geq F(x', y', z') \Leftrightarrow G(x, y, z) \geq G(x', y', z')$.

Proof (sketch). By Theorem 3 in [5], continuity and conditions 1-3 guarantee that F can be written as

$$F(x, y, z) = V(f_1(x) + f_2(y) + f_3(z)) \tag{28}$$

with V monotonically increasing. Because of monotonicity, V is irrelevant for ordinal properties, and F can be replaced by

$$\tilde{F}(x, y, z) = f_1(x) + f_2(y) + f_3(z). \tag{29}$$

Property 4 (which is analogous to Tverky's [25]) implies that, for $i \neq j$,

$$f_i(x_1) - f_i(x_2) = f_i(y_1) - f_i(y_2) \Leftrightarrow f_j(x_1) - f_j(x_2) = f_j(y_1) - f_j(y_2). \tag{30}$$

By the monotonicity properties of F, the derivatives of f_i and f_j have either the same sign or opposite signs for all the values for which they are nonzero. Assume, without loss of generality, that they have the same sign. Also, since the derivatives are zero almost everywhere, then, for almost all x_1 and almost all y_1, it is possible to find x_2 and y_2 for which (30) holds. Considering two sequences $y_n \to y_1$ and $x_n \to x_1$, (30) implies, in the limit of $n \to \infty$,

$$\frac{f'_i(x_1)}{f'_i(y_1)} = \frac{f'_j(x_1)}{f'_j(y_1)} \tag{31}$$

almost everywhere. By fixing y_1, this implies that $f'_i(x) = \alpha_i f'_j(x)$ almost everywhere, so, by continuity, $f_i(x) = \alpha_i f_j(x) + c_i$ (note that the form (29) implies that, if condition 4 holds for one u, it holds for all u). The constants c_i can be collected together and eliminated, since they are irrelevant for ordering. □

The complete proof of this theorem can be found in [18].

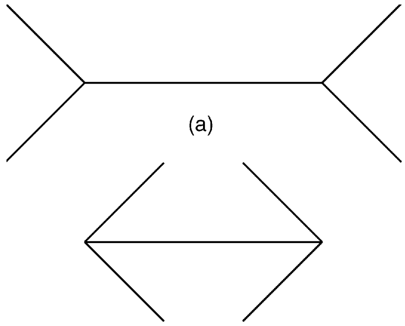

Fig. 4. A proof that the truth of a fuzzy predicate can depend on measures of quantities different from the subject of the predicate: In this case, the truth of the predicate "the line is long" must be different in the two cases since the predicate "line A is longer than line B" has a truth value different from zero. Yet, the length of the two lines is the same. Therefore, the truth of the predicate depends on other measures.

3.2 Feature Dependencies

Our translation of Tversky's measure suffers from a serious drawback: It considers all the features as independent. For instance, in our model, the truth of the statement "the mouth is wide" depends only on the width of the mouth and not on the other measurements. This independence property is easily proven to be false for human perception. For instance, in the famous visual illusion of Fig. 4, the line (a) appears longer than the line (b), although measurement reveals that the two have the same length. This has important consequences for our fuzzy definition.

Let us assume that the the truth of the predicate "line A is longer than line B" is given by a fuzzy inference rule like

If line A is long and line B is short, then line A is longer than line B.

We will use the following fuzzy implication rule: If we have two predicates "X is A," with a truth value $\mu_A(X)$ and a predicate "Y is B," with a truth value $\mu_B(Y)$, then the truth value of the implication "if X is A then Y is B" is given by:

$$\mu_{A\Rightarrow B}(X,Y) = \max\{\mu_A(X), \mu_B(Y)\}. \tag{32}$$

Let μ' be the truth value of the predicate "line A is longer than line B," and μ_A, μ_B be the truth values of the predicates "line A is long" and "line B is long," respectively. We have:

$$\mu_{\Rightarrow} = \max\{\min\{\mu_A, 1-\mu_B\}, 1-\mu'\}. \tag{33}$$

Since the predicate "line A is longer than line B" is perceived as true, we have $\mu' > 1/2$, and $1-\mu' < 1/2$. Moreover, the implication is valid, therefore, $\mu_{\Rightarrow} > 1/2$. This implies

$$\min\{\mu_A, 1-\mu_B\} > \frac{1}{2}. \tag{34}$$

This relation must be true for all the values of μ_A. In particular, the effect is strong when the line A is not judged neither "long" nor "short," that is, when $\mu_A = 1/2$. In this case, for the inequality to be true, we must have $\mu_B < 1/2$, that is, *line B is perceived as shorter than line A*.

This fact cannot be explained if the arguments of μ_A and μ_B are simply the length of the respective lines. But, since the length of a line can be judged when the line is presented in isolation, the values μ_A and μ_B must be completely determined by the length of the respective lines.

We assume that the truth of each predicate is not affected by the truth of other predicates, but the way the predicates interact is: If two predicates tend to be true together, they reinforce each other. This model applies to the following situation: Imagine you know the length of the segment (a) in Fig. 4 (possibly its length relative to the whole figure); then, you can express a judgment on whether the predicate "segment (a) is long" is true. This judgment does not depend on the other features on the image and, if x is the length of the segment, it is has truth value $\mu_a(x_i)$.

However, when the whole image is perceived, the length of the segment is perceived differently depending on the presence or absence of other features (like the existence of outwardly pointing diagonal segments.) We postulate that, although the truth of the predicate "the horizontal line is long" is still the same, the *measure* of the set of true features is changed because of the interaction between different predicates.

The latter model can be defined mathematically by replacing the function f in the definition of the fuzzy feature contrast similarity with a fuzzy integral defined over a suitable fuzzy measure. We use a *Choquet Integral*, and a fuzzy measure that models the interaction between the different predicates [15].

Definition 4. *Let $X = \{x_1, \ldots x_n\}$ be a finite universe. A* fuzzy measure *m is a set function $m : \mathcal{P}(X) \to [0,1]$ such that $m(\emptyset) = 0$, $m(X) = 1$, and ,for all subsets A and B of X, $A \subseteq B \Rightarrow m(A) \leq m(B)$, where $\mathcal{P}(X)$ indicates the power set of X, that is, the set of all subsets of X.*

Definition 5. *Let m be a fuzzy measure on X. The* discrete Choquet integral *of a function $f : X \to \mathbb{R}^+$ with respect to m is defined as:*

$$\oint (f(x_1), \ldots, f(x_n))m = \sum_{i=1}^{n} \big(f(x_{(i)}) - f(x_{(i-1)})\big) m(A_{(i)}), \tag{35}$$

where the notation $\cdot_{(i)}$ means that the indices have been permutated so that

$$0 \leq f(x_{(1)}) \leq f(x_{(2)}) \cdots \leq f(x_{(n)}) \leq 1, \tag{36}$$

$f(x_{(0)})$ is defined to be 0, and

$$A_{(i)} = \{x_{(i)}, \ldots, x_{(n)}\}. \tag{37}$$

Let X be the universe of fuzzy predicates $x_i = \mu_i(\phi)$, where ϕ is a measurement vector and μ_i the truth function for the ith predicate. Also, let f be the identity function $f(x) = x$. Let us suppose, for the ease of notation, that the predicates are already ordered so that

$$0 \leq \mu_1(\phi) \leq \mu_2(\phi) \cdots \leq \mu_n(\phi) \leq 1, \tag{38}$$

and let us define the dummy predicate μ_0 that is always false, i.e., $\mu_0(\phi) = 0$.

Lemma 1. *The fuzzy cardinality of the set of true predicates is equal to n times the Choquet integral of the identity function* *when m is additive and equidistributed (i.e., sets of the same cardinality have equal measure).*

Proof. Since $m(X) = 1$ and m is equidistributed, we have $m(\{\mu_i\}) = 1/n$. Moreover, because of additivity, we have

$$m(A_i) = \sum_{j=i}^{n} m(\{\mu_i\}) = \frac{(n-i+1)}{n}, \tag{39}$$

therefore, the Choquet integral can be written as:

$$\oint(\mu_1, \ldots \mu_n)m = \sum_{i=1}^{n}(\mu_i - \mu_{i-1})\frac{(n-i+1)}{n} = \frac{1}{n}\sum_{i=0}^{n}\mu_i - \mu_0, \tag{40}$$

which, since $\mu_0 = 0$ by definition, is the desired result.□

Thus, when the measure is additive and equidistributed, the Cocquet integral reduces to the cardinality of the fuzzy set, which is the saliency function we used in (23). To see how we can use a nonadditive measure to model dependence between predicates, suppose that all the predicates are independent except for μ_{n-1} and μ_n. Assume that the fact that μ_n is true increases the possibility that μ_{n-1} be also true. Referring to Fig. 4, the two predicates might be:

P_1: "The diagonal lines point strongly outward."

P_2: "The horizontal line is long."

What is the effect of this dependency on the fuzzy measure? Since the perception of the outwardly pointing diagonal lines increases the perception of the length of the line, the predicate P_2 is, in a sense, "more true" due to the truth of P_1. In terms of the fuzzy measure, we can say that it is:

$$\begin{aligned} m(\{x_{n-1}, x_n\}) &= m(\{x_{n-1}\}) + m(\{x_n\}) \\ &\quad + \gamma_{n-1,n} m(\{x_{n-1}\}) m(\{x_n\}), \end{aligned} \tag{41}$$

where $\gamma_{n-1,n} \geq 0$ is a coefficient that models the dependence between the two predicates. Consider an equidistributed measure:

$$m(\{x_i\}) = \frac{1}{n} \tag{42}$$

with the dependency between x_{n-1} and x_n yielding:

$$m(\{x_{n-1}, x_n\}) = \frac{2 + \frac{\gamma_{n-1,n}}{n}}{n} = \frac{2 + \tilde{\gamma}}{n}. \tag{43}$$

Also, suppose that all the other measures are additive, that is

$$m(\{x_{i_1}, \ldots, x_{i_p}\}) = \sum_{j=1}^{p} m(\{x_{i_j}\}) = \frac{p}{n} \tag{44}$$

if either x_{n-1} or x_n do not belong to $\{x_{i_1}, \ldots, x_{i_p}\}$, and

$$\begin{aligned} m(\{x_{i_1}, \ldots, x_{i_p}\}) &= \sum_{j=1}^{p} m(\{x_{i_j}\}) + \gamma_{n-1,n} m(\{x_{n-1}\}) m(\{x_n\}) \\ &= \frac{p}{n} + \tilde{\gamma} \end{aligned} \tag{45}$$

if they do.

When we compute the Choquet integral, we order the predicates by their truth value. Suppose that the value $\mu(x_{n-1})$ is the hth in the ordering, and that $\mu(x_n)$ is the kth ($x_{(h)} = x_{n-1}$, and $x_{(k)} = x_n$) with, say, $k > h$. In this case, in the Choquet integral, there will be $n - k$ subsets that contain both x_{n-1} and x_n, therefore:

$$\oint(\mu_1, \ldots \mu_n)m = \frac{1}{n}\sum_{i \neq n-1}^{n} \mu_i + \frac{(n-k)}{n}\tilde{\gamma}\mu_n. \tag{46}$$

In the following, we will assume a fuzzy measure of the form:

$$m(\{x_{i_1}, \ldots, x_{i_p}\}) = \sum_{j=1}^{p} m(\{x_{i_j}\}) + \gamma_{i_1,\cdots,i_p}\prod_{j=1}^{p} m(\{x_{i_j}\}). \tag{47}$$

The 2^n constants $\gamma_{i_1,\cdots,i_p}$, $i_1 < i_2 \cdots < i_p$ uniquely characterize the measure, and must be determined experimentally.

The γ parameters must let the measure satisfy the three requirements of Definition 4. In particular, the measure of a set must be greater or equal the measure of all its subsets. Let us consider, without loss of generality, the two sets $A = \{x_1, \ldots x_p\}$ and $B = \{x_1, \ldots, x_k\}$, with $k > p$. Then, we have:

$$\begin{aligned} m(A) &= \sum_{i=1}^{p} m(\{x_i\}) + \gamma_A \prod_{i=1}^{p} m(\{x_i\}) \\ m(B) &= \sum_{i=1}^{k} m(\{x_i\}) + \gamma_B \prod_{i=1}^{k} m(\{x_i\}). \end{aligned} \tag{48}$$

By definition of fuzzy measure, it must be $m(B) \geq m(A)$ and, therefore,

$$\begin{aligned} m(B) - m(A) = &\sum_{i=p+1}^{k} m(\{x_i\}) + \gamma_B \prod_{i=1}^{k} m(\{x_i\}) \\ &- \gamma_A \prod_{i=1}^{p} m(\{x_i\}) \\ &\geq 0. \end{aligned} \tag{49}$$

From this relation, it is possible to derive the relation between γ_A and γ_B:

$$\gamma_A \leq \frac{\sum_{i=p+1}^{k} m(\{x_i\})}{\prod_{i=1}^{p} m(\{x_i\})} + \gamma_B \prod_{i=p+1}^{k} m(\{x_i\}). \tag{50}$$

In the case of an equidistributed measure ($m(\{x_i\}) = 1/n$), we have:

$$\gamma_A \leq n^p(k-p) + \frac{\gamma_B}{n^{k-p}} \tag{51}$$

whenever $A \subseteq B$.

Given a fuzzy measure m that takes into account the dependence among features, we can define the Tversky similarity as:

$$\begin{aligned} S(\phi, \psi) = &\oint \min\{\mu_i(\phi) + \mu_i(\psi) - 1, 0\} m - \alpha \\ &\oint \max\{\mu_i(\phi) - \mu_i(\psi), 0\} m - \beta \\ &\oint \max\{\mu_i(\phi) - \mu_i(\psi), 0\} m. \end{aligned} \tag{52}$$

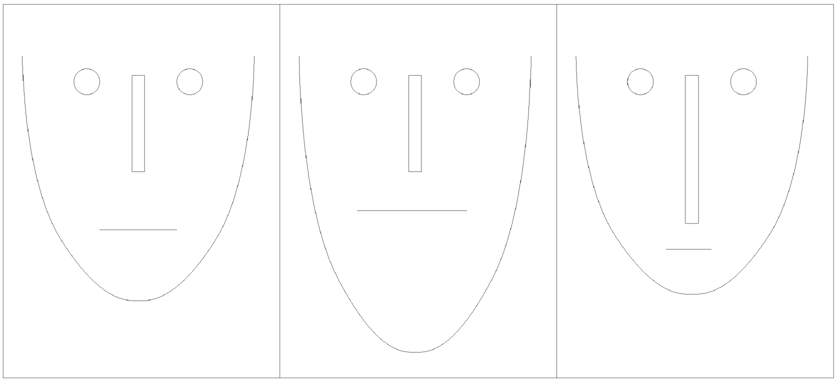

Fig. 5. Three face sketeches used in our face similarity experiment.

Both this measure and (23) reduce to the usual Tversky similarity if the features are binary and the measure is additive and equidistributed.

4 EXAMPLES

In this section, we present a comparison between some of the similarity measures introduced so far. We will consider the Euclidean Distance, the Attneave city-block distance, the Thurstone-Shepard model, and the FFC model.

4.1 Similarity of Faces

In this experiment, we use the similarity measures to characterize the similarity between face-like stimuli. Similarity of faces is a complex issue that depends on a number of factors, like the color and the shape of the hair, the texture of the skin, the geometry of the face components, and so on. In this experiment, we have chosen a simplified approach and we will determine similarity based only on geometric measurements. The features are computed on simple image sketches like those in Fig. 5. Our set consisted of 10 such sketches. The reason for using these sketches, rather than full face images, is the poverty of our feature set. Face images contain very important clues that are not characterized by our geometric features (hair and skin color, etc.) These features tend to bias the human judgment of faces, so it is impossible to compare the result of human judgment with those of geometric features in these conditions. Since we are evaluating similarity measures and not features and since the geometric features that we use are powerful enough to characterize the face sketches that we use, we believe that, in this case, the simplification is epistemologically justified.

4.1.1 Distance Measures

The geometric measurements we derive from a face image are described in Fig. 6. All the measurements are normalized, dividing them by the distance between the eyes. These measurements provide support for the five predicates of Table 1 (see also [4] for the rationale behind this choice.) The predicates can be collected in a set of features, and used to compute Tversky similarity. The FFC similarity model uses the truth value of the predicates, while metric distances are based on the geometric measurements.

4.1.2 Method

The experiment was organized as follows: We selected four subjects with no knowledge of our activity in similarity measures. Each subject was asked to rank nine of the sketches (like those in Fig. 5) for similarity with respect to the 10th (the "query" sketch.) The query sketch was chosen at random and each subject was asked to give a total of three rankings with respect to three different query sketches. Each subject was also asked to divide the ranked images in three groups: The first group consisted of faces judged "very similar" to the query, the second group consisited of faces judged "not very similar" to the query, and the third of "completely different" faces. The reason for this classification will be clear in the following. Whenever possible (for two subjects out of four), the subject was asked to repeat the experiment with the same query sketches after two weeks to check for stability.

The ordering given by any subject was compared with the orders obtained on the same sketch by the Euclidean distance, the Attneave distance, the Thurstone-Shepard distance, and two versions of the FFC distance: one without feature interaction and one with feature interaction. We compared the orderings using the *weighted displacement* measures proposed in [6].

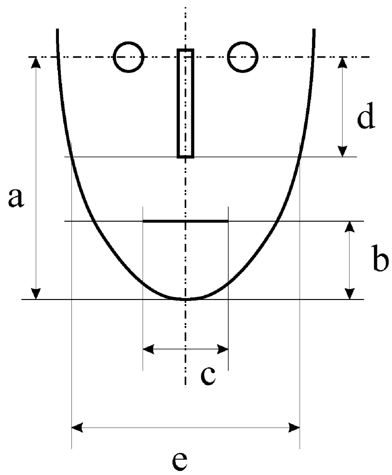

Fig. 6. These five measures are taken from a face image to provide support to the fuzzy predicates used for the similarity assessment.

TABLE 1
Predicates Used for Similarity Evaluation and Measured Quantities that Support Their Truth

Predicate	Supporting quantity
Long face	a
Long chin	b
Wide mouth	c
Long nose	d
Large face	e

All these measures are normalized with respect to the distance between their eyes.

Assume that we have a query q which operates on a database of n images. We consider the ordering given by the human subject as the "ground truth." Let $L_t = \{I_1, \dots I_n\}$ be this ordering. In addition, we have a measure of relevance $0 \leq S(I, q) \leq 1$ such that, for the real order,

$$\forall i \quad S(I_i, q) \geq S(I_{i+1}, q). \tag{53}$$

In our case, we use the categorization given by the subject as a relevance measure and set $S(I_i, q) = 0.8$ for images "very similar" to the query, $S(I_i, q) = 0.5$ for images "not very similar" to the query, and $S(I_i, q) = 0.05$ for images "completely different."

Because of imperfections, the database is not giving us the ordering O_t, but an order $L_d = \{I_{\pi_1}, \dots, I_{\pi_n}\}$, where $\pi_1, \dots \pi_n$ is a permutation of $1, \dots n$. The displacement of I_i in O_d is defined as $d_q(I_i) = |i - \pi_i|$. The *relative weighted displacement* of L_d is defined as

$$W_q = \frac{\sum_i S(I_i, q) d_q(I_i)}{\Omega}, \tag{54}$$

where $\Omega = \lfloor \frac{n^2}{2} \rfloor$ is a normalization factor. W_q is zero if $L_d = L_t$, and $W_q \leq 1$ for all L_d and S.

The results relative to the first subject were used to adjust the parameters of the distances. For the Thurstone-Shepard model, the best results were obtained when the underlying Minkowski metric had exponent $p = 2$. Since this coincides with the Euclidean distance, we decided not to optimize the Thurston-Shepard model, but to use $p = 0.3$ as a contrast to the other metric models. For the FFC models, the best results were obtained with $\alpha = 1$, $\beta = 6$ (see (23)). We also introduced an interaction between the features "long face" and "large mouth" with $\tilde{\gamma} = 0.5$ (see (46)).

4.1.3 Results

The results relative to the other three subjects were used for comparison. For every ranking provided by a subject, the ordering relative to the same query sketch was obtained using each of the five similarity measures and the weighted displacement was computed. The results were then averaged. Table 2 shows the average and the variance for the five similarity measures. In order to establish whether the differences are significant, we performed an analysis of the variance, with an hypothesis acceptance level $\alpha = 0.05$.[3]

3. Given the null hypothesis "all the measures provide the same result," $\alpha = 0.05$ means that we are accepting a 5 percent chance of rejecting the null hypothesis when it is in fact true. A 5 percent level is the norm in psychology and behavioral sciences.

For the whole ensemble, we obtained $F = 14.292$ [10], which leads to the conclusion that the differences are indeed significant. In order to establish which differences are significant, we computed the F ratio for each pair of distances. The results are shown in Table 3. The difference between two measures should be considered significant if the F value at the intersection of the respective row and column is greater than 4.75 (for the determination of this value, see [10].)

The $\hat{\omega}^2$ measure (the fraction of the variance due to actual differences among the measures) gives the results in Table 4. The quantity $\hat{\omega}^2$ measures the fraction of the variance that is due to actual differences in the experimental conditions, rather than random variations between the subjects. Most of the values are around 0.5 or greater, indicating a strong dependence of the variance on actual differences between the similarity measures. The results of the comparison between the two feature contrast measures is not as strong as the difference between these and the other measures, although a value $\hat{\omega}^2 = 0.09$ still indicates a significant effect.

This experiment is, of course, not conclusive and it represents only a first step in the evaluation of the similarity measures for several reasons. First, due to a number of constraints, it was possible only to check two of our subjects for stability. Since for both the subjects the ordering was found stable (weighted displacement less than 0.02), we extrapolated to the other subjects. More importantly, we didn't accurately determine the influence of the parameters on the evaluation, although partial results seem to indicate that the performance is relatively stable in the presence of changes. On the other hand, the relatively small number of subjects is not a serious problem in this case since, due to the high value of $\hat{\omega}^2$, the sensitivity of the experiment is around 0.8, which is considered an acceptable value [10].

4.2 Similarity of Textures

In this section, we consider the determination of similarity of texture images. Texture identification is an important problem in computer vision which has received considerable attention (see, for instance, [13], [14].) In this experiment, we are concerned with texture similarity: Given a texture sample, find *similar* samples in a database.

We used 100 images from the MIT VisTex texture database [28]. The database contains images extracted from different classes of textures, like bark, bricks, fabric, flowers, and so on. Textures were characterized using the Gabor features introduced in [14]. These features work on gray-level images, so color was disregarded for the whole experiment (e.g., human subjects were shown gray-level versions of the texture images.) Also, based on the results of the previous experiment, we tested only the Euclidean and the Fuzzy Feature Contrast metrics.

4.2.1 Distance Measurement

Manjunath and Ma's features ([14]) are collected in a vector of 60 elements. If we measure the Euclidean distance between two raw vectors, we could encounter scale problems: Features that have an inherently larger scale would be predominant. This is especially a problem for the Euclidean distance, since FFC normalized all the features in [0, 1] via the membership function. In order to

TABLE 2
Average (μ) and Variance (σ^2) of the Weighted Displacement for the Five Measures Considered

	A	E	TS	FFC1	FFC2
μ	0.084	0.078	0.111	0.051	0.049
σ^2	0.055	0.041	0.043	0.066	0.059

A: Attneave. E: Euclid. TS: Thurstone-Shepard. FFC1: Fuzzy Feature contrast without feature interaction. FFC2: Fuzzy Feature contrast with feature interaction.

provide a more objective comparison, we tried two types of Euclidean measure: normalized and not normalized. Let $x_i = \{x_{i1}, \dots, x_{im}\}$, $1 \le i \le N$, $m = 60$, the feature vector of the ith image. We compute the componentwise averages

$$\bar{x}_j = \frac{1}{N}\sum_{i=1}^{N} x_{ij} \tag{55}$$

and the componentwise standard deviation

$$\sigma_j = \left[\frac{1}{N}\sum_{i=1}^{N}(x_{ij} - \bar{x}_j)^2\right]^{\frac{1}{2}}. \tag{56}$$

With these definitions, the scaled Euclidean distance is defined as:

$$\begin{aligned} d_{E'}(x_i, x_k) &= \left[\sum_j \left(\frac{x_{ij} - \bar{x}_j}{\sigma_j} - \frac{x_{kj} - \bar{x}_j}{\sigma_j}\right)^2\right]^{\frac{1}{2}} \\ &= \left[\sum_j \left(\frac{x_{ij} - x_{kj}}{\sigma_j}\right)^2\right]^{\frac{1}{2}}. \end{aligned} \tag{57}$$

Experimentally, the two distances gave similar results, the scaled distance being slightly better than the unscaled. In the rest of this section, we will only consider the scaled Euclidean distance (from now on, we will just call it the Euclidean distance for the sake of brevity.) The distance measure for FFC is given by (24), with membership function

$$\mu(x_{ij}) = \left(1 + \exp\left(-\frac{x_{ij} - \bar{x}_j}{\sigma_j}\right)\right)^{-1}. \tag{58}$$

4.2.2 Method

Due to the substantially larger size of the database in this example, it was impractical to use the same method as in the previous example. While it is feasible to ask a subject to order 10 sketches with respect to a stimulus, it is unfeasible to ask to rank 100 texture images. Therefore, we followed a different procedure. For a given experiment, we selected one query image, x_q, ordered the database using both the Euclidean and the FFC measures, and, for each measure, collected the 10 images closer to the query. Let A_E and A_T be the sets of the 10 images closer to the query using the Euclidean and the FFC measures, respectively. We then considered the set $A = A_E \cup A_T$ of the images returned by either of the queries. In our case, this set contained between 12 and 16 images, depending on the number of images common to the two queries.

The set A was presented to our subjects and they were asked to rank the images with respect to the query. We then took the first 10 images ranked by the subjects and compared them with the ordering obtained by the two similarity measures using the same measure as in the previous experiment.

Note that, with this technique, it is impossible to provide an absolute measure of the performance of a certain similarity measure with respect to human performance because our subjects don't see the whole database. There might be images in the database that a person would judge very similar to the query but, if both our distance measures miss them, the subject will never see them. The only results that this technique can give is a measurement of the relative performace of two similarity measures.

Fig. 7 shows a sample experiment. The first row contains the top 10 images returned by the FFC distance. The second row contains the top 10 images returned by the Euclidean distance. All images contained in the first two rows were shown to one of our subjects and she was asked to rank them. The results are shown in the third row of Fig. 7.

4.2.3 Results

The average value ($\bar{W}_q$) and the variance (σ^2) of the weighted displacement measure for the Euclidean and the FFC distances are reported in Table 5.

TABLE 3
The F Ratio for the Pairwise Comparison of the Similarity Measures

	A	E	TS	FFC 1	FFC 2
FFC2	9.1	7.81	21.68	0.1	0
FFC1	11.83	10.16	25.52	0	
TS	14.3	15.98	0		
E	2.25	0			
A	0				

TABLE 4
The Values for the Pairwise $\hat{\omega}^2$ Calculations

	A	E	TS	FFC 1	FFC 2
FFC2	0.51	0.47	0.69	0.09	0
FFC1	0.49	0.45	0.68	0	
TS	0.54	0.57	0		
E	0.1	0			
A	0				

The quantity $\hat{\omega}^2$ measures the fraction of the variance that is due to actual differences in the experimental conditions, rather than random variations between the subjects. Most of the values are around 0.5 or greater, indicating a strong dependence on the variance on actual differences between the similarity measures.

TABLE 5
Average ($\bar{W}_q$) and Variance (σ^2) of the Weighted Displacement for the Two Measures Considered

	E	FFC
$\bar{W}_q$	1.13	1.058
σ^2	0.58	0.58

E: Euclid. FFC: Fuzzy Feature contrast.

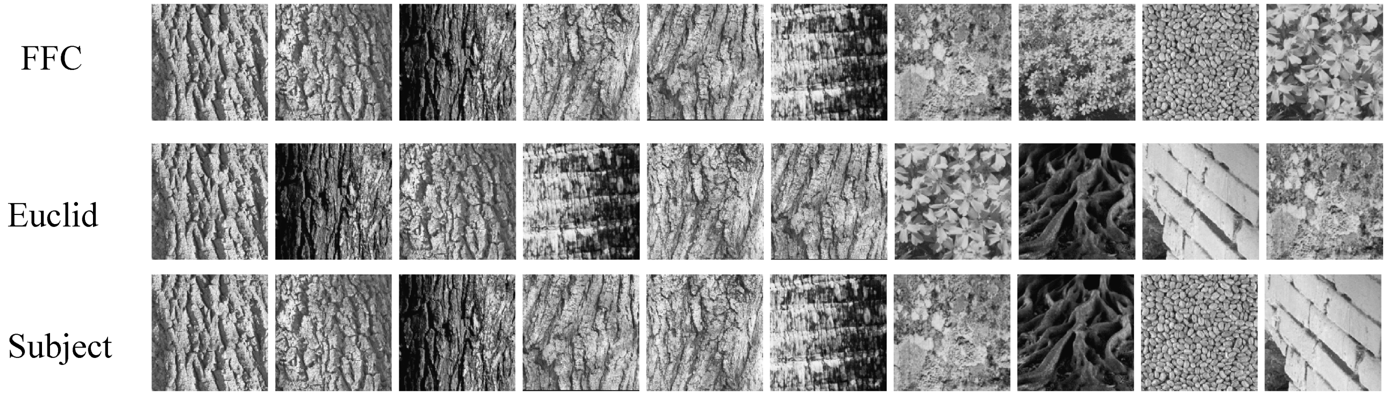

Fig. 7. Similarity results for one of the textures in the database. Orderings obtained by the FFC distance, the Euclidean distance, and a human subject.

This experiment gave us a value $F = 4.86$, which implies that the difference is statistically significant with $\alpha = 0.05$ and $\hat{\omega}^2 = 0.39$, which, conventionally, means that the effect on the distance measure is "large" (a significant portion of the variance is due to the distance measure and not to subject variation).

5 Conclusions

In visual information systems, it is important to define exactly the operation of *similarity assessment*. While matching is defined essentially on logic grounds, the definition of similarity assessment must have a strong psychological component. Whenever a person interrogates a data repository asking for something *close*, or *related*, or *similar* to a certain description or to a sample, there is always the understatement that the similarity at stake is *perceptual similarity*. If our systems have to respond in an "intuitive" and "intelligent" manner, they must use a similarity model resembling the humans'.

One problem with the psychological view is that often we don't have mathematical or computational models that can be applied to artificial domains. In this paper, we have explored the psychological theories that are closer in spirit to the needs of computer scientists.

Most of the similarity theories proposed in literature reject some or all the geometric distance axioms. The more troublesome axiom is the triangle inequality, but other properties, like symmetry and the constancy of self-similarity have been challenged. Also, nonlinearities enter in the similarity judgment both at the feature level (Fechner's law [12]) and during similarity measurement.

One of the most successful models of similarity is Tversky's *feature contrast* which, incidentally, is also the most radical in the refusal of the distance axioms. In this paper, we have used fuzzy logic to extend the field of applicability of the model. Also, the use of fuzzy logic allows us to model the *interference* between the features upon which the similarity is based. By interference, we mean that the judged truth of a property, like the fact that a line is long, does not depend only on the measured length of the line, but also on the relationships between the line and the other elements in the image. We have shown that it is possible to model this interference using a suitable *fuzzy measure*.

An important problem that we could not address in this paper is the determination of the parameters of the similarity measure. The parameters α and β in (23), the constants γ in (47), and the parameters of the membership function influence the similarity measure. This topic is considered in [18].

Acknowledgments

The authors gratefully acknowledge the anonymous reviewers for the many helpful comments and criticism on earlier drafts of the paper. This work was supported in part by the U.S. National Science Foundation under grant NSF IRI-9610518.

References

[1] P. Aigrain, H.-J. Zhang, and D. Petkovic, "Content-Based Representation and Retrieval of Visual Media: A State-of-the-Art Review," *Multimedia Tools and Appliactions*, vol. 3, pp. 179-202, 1996.
[2] F.G. Ashby and N.A. Perrin, "Toward a Unified Theory of Similarity and Recognition," *Psychological Review*, vol. 95, no. 1, pp. 124-150, 1988.
[3] F. Attneave, "Dimensions of Similarity," *Am. J. Psychology*, vol. 63, pp. 516-556, 1950.
[4] E. Brunswik, *Perception and the Repreentative Design of Psychological Experiments*. Univ. of California Press, 1956.
[5] G. Debreu, "Topological Methods in Cardinal Utility Theory," *Mathematical Models in the Social Sciences*, K. Arrow, S. Karlin, and P. Suppes, eds. Stanford Univ. Press, 1960.
[6] A.D. Narasimhalu, M.S. Kankanhalli, and J. Wu, "Benchmarking Multimedia Databases," *Multimedia Tools and Applications*, vol. 4, no. 3, pp. 333-355, May 1997.
[7] D.M. Ennis, J.J. Palen, and K. Mullen, "A Multidimensional Stochastic Theory of Similarity," *J. Math. Psychology*, vol. 32, pp. 449-465, 1988.
[8] A.K. Jain, *Fundamental of Digital Image Processing*. Englewood Cliffs, N.J.: Prentice Hall, 1989.
[9] R. Jain, R. Kasturi, and B. Shunk, *Machine Vision*. McGraw-Hill, 1995.
[10] G. Keppel, *Design and Analysis. A Researcher's Handbook*. Upper Saddle River, N.J.: Prentice Hall, 1991.
[11] C.L. Krumhansl, "Concerning the Applicability of Geometric Models to Similarity Data: The Interrelationship between Similarity and Spatial Density," *Psychological Review*, vol. 85, pp. 445-463, 1978.
[12] S.W. Link, *The Wave Theory of Difference and Similarity*. Lawrence Erlbaum Assoc., 1992.

[13] F. Liu and R.W. Picard, "Periodicity, Directionality, and Randomness: Wold Features for Perceptual Pattern Recognition," *Proc. 12th Int'l Conf. Pattern Recognition,* Oct. 1994.
[14] B.S. Manjunath and W.Y. Ma, "Texture Features for Browsing and Retrieval of Image Data," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 18, no. 8, pp. 837-842, Aug. 1996.
[15] K. Miyajima and A. Ralescu, "Modeling of Natural Objects Including Fuzziness and Application to Image Understanding," *Proc. Second IEEE Int'l Conf. Fuzzy Systems,* pp. 1,049-1,054, 1993.
[16] E. Rosh, "Cognitive Reference Points," *Cognitive Psychology,* vol. 7, pp. 532-547, 1975.
[17] E.Z. Rothkopf, "A Measure of Stimulus Similarity and Errors in Some Paired-Associate Learning Tasks," *J. Experimental Psychology,* vol. 53, pp. 94-101, 1957.
[18] S. Santini and R. Jain, "The Use of Psychological Similarity Measure for Queries in Image Databases," technical report, Visual Computing Laboratory, Univ. of California, San Diego, 1996, available at http://www-cse.ucsd.edu/users/ssantini.
[19] S. Santini and R. Jain, "Similarity Is a Geometer," *Multimedia Tools and Applications,* vol. 5, no. 3, Nov. 1997, available at http://www-cse.ucsd.edu/users/ssantini.
[20] R.N. Shepard, "The Analysis of Proximities: Multidimensional Scaling with Unknown Distance Function, Part I," *Psychometrika,* vol. 27, pp. 125-140, 1962.
[21] R.N. Shepard, "Toward a Universal Law of Generalization for Physical Science," *Science,* vol. 237, pp. 1,317-1,323, 1987.
[22] M. Swain and D. Ballard, "Color Indexing," *Int'l J. Computer Vision,* vol. 7, no. 1, pp. 11-32, Nov. 1991.
[23] L.L. Thurstone, "A Law of Comparative Judgement," *Psychological Review,* vol. 34, pp. 273-286, 1927.
[24] W.S. Torgerson, "Multidimensional Scaling of Similarity," *Psychometrika,* vol. 30, pp. 379-393, 1965.
[25] A. Tversky, "Features of Similarity," *Psychological Review,* vol. 84, no. 4, pp. 327-352, July 1977.
[26] A. Tversky and I. Gati, "Similarity, Separability, and the Triangle Inequality," *Psychological Review,* vol. 89, pp. 123-154, 1982.
[27] A. Tversky and D.H. Krantz, "The Dimensional Representation and the Metric Structure of Similarity Data," *J. Math. Psychology,* vol. 7, pp. 572-597, 1970.
[28] Vision Texture, Web Page, http://www-white.media.mit.edu/vismod/imagery/VisionTexture/vistex.html.

Simone Santini (M'98) received the Laurea degree in electrical engineering from the Università di Firenze, Firenze (Italy) in 1990, the MSc degree in computer science from the University of California, San Diego, in 1996, and the PhD degree from the University of California, San Diego, in 1998. Currently, he is a project scientist at the Visual Computing Laboratory, Department of Electrical and Computer Engineering at the University of California, San Diego. From 1990 until 1994, Dr. Santini was with the Dipartimento di Sistemi e Informatica, Università di Firenze. He was a visiting scientist at the Artificial Intelligence Lab of the University of Michigan in 1990 and at the IBM Almaden Research Center, San Jose, California, in 1993. He is on the editorial board of the journals *Multimedia Tools and Applications* and *Pattern Recognition*. His current research interests are the development of alternative query models for image databases, behavior characterization and recognition from multisensor input, and the development of event definition and query languages for storage and access of video events.

Ramesh Jain (F'92) received the BE degree from Nagpur University in 1969 and the PhD degree from the Indian Institute of Technology, Kharagpur, in 1975. He is currently a professor of electrical and computer engineering and computer science and engineering at the University of California, San Diego (UCSD). Before joining UCSD, he was a professor of electrical engineering and computer science and the founding director of the Artificial Intelligence Laboratory at the University of Michigan, Ann Arbor. He also has been affiliated with Stanford University, IBM Almaden Research Labs, General Motors Research Labs, Wayne State University, the University of Texas at Austin, the University of Hamburg, West Germany, and the Indian Institute of Technology, Kharagpur, India. His current research interests are in multimedia information systems, image databases, machine vision, and intelligent systems.

Dr. Jain is a fellow of the IEEE, the AAAI, and the Society of Photo-Optical Instrumentation Engineers, and a member of the ACM, Pattern Recognition Society, Manufacturing Engineers. He has been involved in the organization of several professional conferences and workshops, and served on the editorial boards of many journals. He was editor-in-chief of *IEEE Multimedia* and is on the editorial boards of *Machine Vision and Applications*, *Pattern Recognition*, and *Image and Vision Computing*.