Journal of Digital Imaging

A New Family of Distance Functions for Perceptual Similarity Retrieval of Medical Images

Joaquim Cezar Felipe,[1] Caetano Traina Jr,[2] and Agma Juci Machado Traina[2]

A long-standing challenge of content-based image retrieval (CBIR) systems is the definition of a suitable distance function to measure the similarity between images in an application context which complies with the human perception of similarity. In this paper, we present a new family of distance functions, called *attribute concurrence influence distances* (AID), which serve to retrieve images by similarity. These distances address an important aspect of the psychophysical notion of similarity in comparisons of images: the effect of concurrent variations in the values of different image attributes. The AID functions allow for comparisons of feature vectors by choosing one of two parameterized expressions: one targeting weak attribute concurrence influence and the other for strong concurrence influence. This paper presents the mathematical definition and implementation of the AID family for a two-dimensional feature space and its extension to any dimension. The composition of the AID family with L_p distance family is considered to propose a procedure to determine the best distance for a specific application. Experimental results involving several sets of medical images demonstrate that, taking as reference the perception of the specialist in the field (radiologist), the AID functions perform better than the general distance functions commonly used in CBIR.

KEY WORDS: Distance function, medical images, content-based image retrieval

INTRODUCTION

Digital images, which are present in the majority of medical systems, serve to support diagnostic activities. However, for the effective and suitable use of images, these systems must include tools for image management, including fast and effective comparison and retrieval. In the image analysis environment, the ability to compare images automatically is important because the number of stored images is usually very large, precluding a radiologist from making individual comparisons of the images in the entire database. Some groups of radiologists prefer to search in reference libraries, which are composed of sets of typical images. However, the possibility of retrieving images containing patient-related information from specific databases enables the analyst to make more in-depth analyses, for instance, to explore and study the patients from a particular geographic region.

The core of content-based image retrieval (CBIR) is the use of intrinsic visual features, which are extracted automatically from the images to describe them while keeping their most relevant characteristics. Such features, usually consisting of numerical values obtained by image-processing algorithms, are used to compare and index images and are usually placed together in a *feature vector*. Each item of a feature vector is also called an *image attribute*. CBIR techniques take advantage of index structures that use the similarity of features to speed up the retrieval, doing this automatically.[1,2] In Figure 1, the user places a query, which is done typically by providing the

[1]*From the Department of Physics and Mathematics, University of São Paulo, Ribeirão Preto, Brazil.*

[2]*From the Department of Computer Science, University of São Paulo, São Carlos, Brazil.*

Correspondence to: Joaquim Cezar Felipe, Department of Physics and Mathematics, University of São Paulo, Ribeirão Preto, Brazil; tel: +55-16-36231558; fax: +55-16-36024887; e-mail: jfelipe@ffclrp.usp.br

Online publication 11 January 2008

doi: 10.1007/s10278-007-9084-x

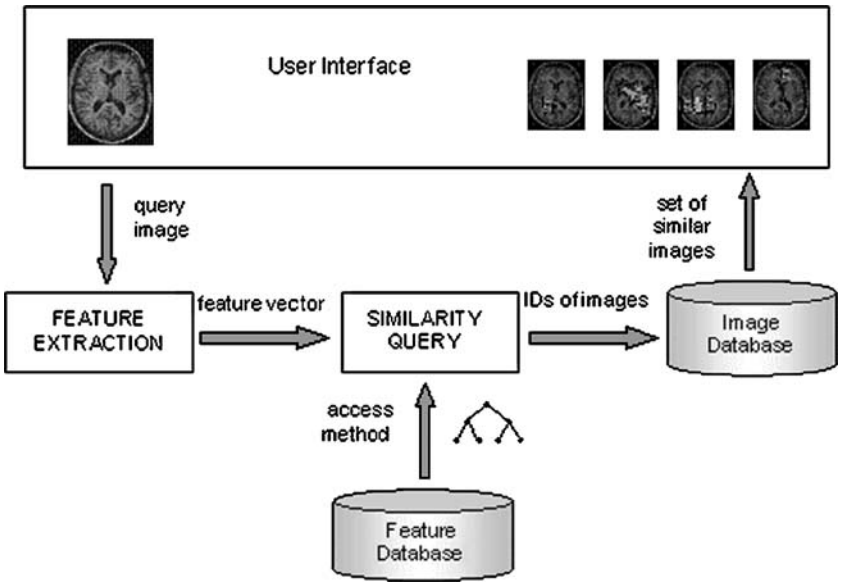

Fig 1. Example of a CBIR environment.

system with an image (query image) from which features will be extracted. In the similarity query process, the features of an image are compared with the features of the images in the database through an access method that uses a distance function to measure the degree of similarity between two images. The identifiers of the selected images are used to retrieve the set of similar images from the image database.

An image can be described according to several *descriptors*. A descriptor is a sensorial stimulus—such as shape, color, texture, and spatial positioning of objects in the image—, which a human takes as reference to judge similarity. Attributes related to specific descriptors can be extracted from stored images to serve as the basis of the similarity calculation.

Several works have been conducted to specify models of image representation. The active appearance model (AAM) is a statistical generative model of a certain visual phenomenon. Considering images, AAM is a model for the shape and gray-level distribution of the image. Matching to an image involves finding model parameters, which minimize the difference between the image and a synthesized model example, projected into the image.[3] The image is decomposed into local regions which are represented, for instance, by a texture descriptor.[4] The match is obtained after a few iterations, but the potentially large number of parameters makes this a difficult problem. It is often used for object recognition (face, for instance) and can be potentially used to generate feature vectors that represent the image in the context of a perceptual study.[5]

The most common types of query based on image similarity are *k-nearest neighbor* and *range queries*. A *k*-nearest neighbor query involves searching for the *k* most similar images to the reference one. A range query consists of searching for all images similar to the reference one up to a given degree (radius or range).

The medical image domain is often considered one of the fields that can benefit the most from the application of CBIR.[6] The use of CBIR techniques applied to medical images eliminates a strong constraint of picture archiving and communication systems (PACS), as searching for stored images with CBIR does not require these images to be linked to a patient or exam identifier, allowing for direct image-to-image comparisons. By enabling the physician to use a system that searches for similar cases based on a reference image, CBIR systems help decision-making and generate opportunities for more accurate diagnoses. On the other hand, concurrent with the evolution of digital imaging, electronic clinical patient records have become commonplace in healthcare centers. The possibility of handling images and related data allows for the development of tools for computer-aided diagnosis (CAD) and increases the quality of the radiologist's tasks.[7–11] In several applications,

combining textual (semantic) and intrinsic (syntactic) information can lead to consistent and valuable results.[12,13] The development of efficient methods to deal with intrinsic features and to perform similarity comparisons is crucial to the achievement of such results.

Several works have focused on adding CBIR functionalities to medical decision systems,[6,14–18] but none of them has tackled the problem of analyzing the concurrent variation of features with a view to keeping the perceptual notion of similarity as the specialists understand it.

Projects and researches involving CBIR methods and their evolution have come up against a long-standing challenge: the so-called *semantic gap*, i.e., the discrepancy between the results of automated methods and the users' expectations. The semantic gap in the field of medical images arises from the difficulty in covering the diversity of circumstances related to image analysis. For instance, a physician analyzing a given image could be involved in several contexts for various diagnostic purposes, which might not be covered suitably by the available methods. To minimize this drawback, an approach that has often been adopted is the delimitation of very specific contexts when defining CBIR methods.

However, although it encompasses specific or wide contexts, the similarity between images can be measured in several ways. The numerical representation and comparison of images by means of their feature vectors and distance functions requires the implementation of extensive sets of image descriptors and distance functions, followed by studies that indicate the best ones for each specific analytical context.

Thus, once the methods for feature extraction are defined, faithfully representing the image descriptors, it is fundamental to define distance functions that are able to reduce the semantic gap by providing the closest approximation between computerized image comparisons and similarity evaluations by the human analyst.

The work reported in this paper consisted of an original study to reduce the semantic gap concerning medical images through the definition of suitable distance functions, taking into account the influence of concurrent variations in the values of the different attributes represented in the feature vectors. The result of this study was the mathematical definition and implementation of a new family of distance functions. Our experiments show that, in some contexts of medical image analysis, the results provided by these functions are closer to the radiologist's findings than those provided by the traditional, most commonly used distance functions. A method was proposed and tested for comparative evaluations to identify the nature of the influence of attribute variations and, hence, the most suitable distance functions for different contexts.

The remainder of this paper is structured as follows. The "Background on Distance Functions" section gives brief background information about distance functions. The "Proposed Method: Attribute Concurrence Influence Distance—Aid Family" section presents the newly proposed distance functions. The "Experiments and Results" section discusses the experiments conducted to corroborate the effectiveness of the proposed functions, and the "Discussion and Conclusions" section concludes the paper.

BACKGROUND ON DISTANCE FUNCTIONS

A computational system that controls the storage and retrieval of images by content must have three main, strongly integrated components, which are:

(a) a set of image feature extractors—which provide the features employed to compare images;
(b) a set of comparative methods making use of distance functions compatible with the features—which compute the degree of similarity between images;
(c) one or more index structures—which are employed to efficiently process the similarity queries between images.

A feature extractor is an algorithm responsible for making a low level analysis of the image content, calculating local properties, and generating a generic signature (or feature vector) of the image. CBIR systems usually have a set of different extractors, each one responsible for generating a distinct signature related to a different original image descriptor.

Index methods such as the metric access methods (MAM) create and organize the image signatures hierarchically, serving to process similarity queries.[19,20] In an image database, a specific index tree is built for each type of feature vector and each comparative method. This mechanism allows

for the effective retrieval of information from image databases, following user-defined parameters.

Comparative methods are procedures that define how signatures are compared to find the distance between images. The mathematical formulation used by each method to compare the features is the *distance function*.[21–23] If the dissimilarity is equal to zero, the data items are the same (i.e., the most similar possible) and, as the distance value grows, so does the dissimilarity between them. Table 1 presents the mathematical expressions of distance functions most commonly used in CBIR.[24–26]

The Minkowski family of distances is based on the L_p norm.[25] According to the value of p, specific functions are obtained, such as:

p=1	city block or Manhattan distance (L_1)
p=2	Euclidean distance (L_2)
p=∞	infinity or Chebychev distance (L_∞).

The Euclidean distance is the most well-known and commonly used distance function. It defines the geometric place of all points equidistant from the point representing the query object, i.e., in two-dimensional space, it is a circumference centered at the query object. The L_1 distance, on the other hand, defines the geometric place of all points that have the same value of the sum of absolute differences of each attribute, whereas L_∞ can be approximated by the maximal difference of attributes. The regions where objects at the same distance from the reference object are placed, considering the most common Minkowski distances, are shown in Figure 2, for the positive two-dimensional axis.

Table 1. Distance Functions

Considering two objects (images) Q and C represented by vectors $q = (q_1, q_2, \ldots, q_n)$ and $c = (c_1, c_2, \ldots, c_n)$	
Minkowski distance (L_p)	$d(Q,C) = \sqrt[p]{\sum_{i=1}^{n} \lvert c_i - q_i \rvert^p}$
City block (L_1)	$d(Q,C) = \sum_{i=1}^{n} \lvert c_i - q_i \rvert$
Euclidian (L_2)	$d(Q,C) = \sqrt[2]{\sum_{i=1}^{n} \lvert c_i - q_i \rvert^2}$
Infinity (L_∞)	$d(Q,C) = \max_{i=1}^{n} \lvert c_i - q_i \rvert^p$
Weighted	$d(Q,C) = \sqrt[p]{\sum_{i=1}^{n} w_i \lvert (c_i - q_i) \rvert^p}$
Mahalanobis distance	$d_{MH}(Q,C) = \sqrt{(c-q)^T V^{-1} (c-q)}$
Jeffrey divergence	$d_J(Q,C) = \sum_{i=1}^{n} c_i \log \frac{c_i}{m_i} + q_i \log \frac{q_i}{m_i}$
χ^2 distance	$d_{\chi^2}(Q,C) = \sum_{i=1}^{n} \frac{(c_i - m_i)^2}{m_i}$ where : $m_i = \frac{c_i + q_i}{2}$

The choice of a distance function depends on several characteristics of the related system, such as: (a) the descriptors to be used, (b) the statistical nature of the context, (c) the data types of each attribute, (d) the preprocessing procedure to be applied to data, and (e) the semantic particularities presented by the environment.

The literature contains several works on the definition and use of new distance functions, most of them focusing on specific applications or contexts. Among these distances are: Mahalanobis distance,[27] Kullback–Leibler and Jeffrey divergences,[22] χ^2 measure,[22,28] cosine distance,[29] quadratic distance,[30] histogram intersection,[29] contrast model,[31,32] perception-based distance,[33] Bhattacharyya distance,[34] and Hellinger distance.[28] Reference reviews and evaluates the most commonly used distance functions.[35] Descriptions are given of comparative methods developed specifically to compute the retrieval of similar medical images.

PROPOSED METHOD: ATTRIBUTE CONCURRENCE INFLUENCE DISTANCE—AID FAMILY

The choice of an appropriate distance function is crucial for effective similarity-based comparisons

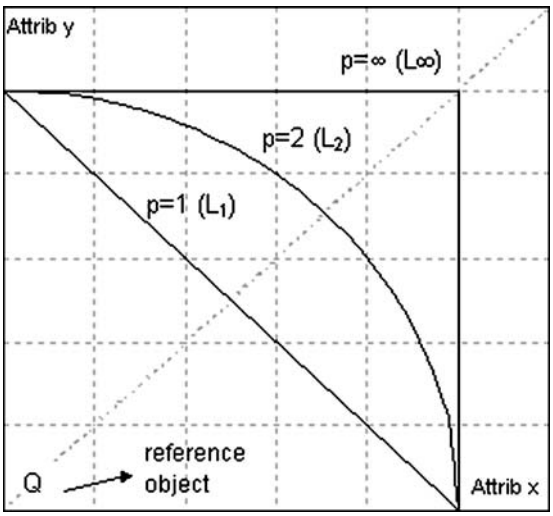

Fig 2. Geometric equidistant places defined by different values of *p*, considering a two-dimensional space (only the first quadrant is shown).

between images. Even considering that the attributes (features) are independent of each other, an important question that should be considered in the process of image comparison is: How do these attributes affect the human perception of similarity when their values vary concurrently? In other words, if images Q and C_1 present a major difference in only one attribute and images Q and C_2 present minor differences but in more than one attribute, which image, C_1 or C_2, will be considered more similar to image Q? In short, how do simultaneous variations in attributes affect the human judgment of similarity?

To identify the best way to combine the attributes that represent the image and thus achieve the most perceptual results, this work investigates distance functions that consider the effects of the concurrent variations of these attributes. Our goal is to answer this question for specific contexts and purposes through the definition of a family of distance functions—attribute concurrence influence distances (AID)—which allow the user to establish parameters and adjust the influence of attribute concurrence, leading to an approximation of human perception and thereby reducing the semantic gap.

Concepts of Attribute Concurrence Influence

For the sake of clarity, we begin our discussion with two-dimensional feature vectors (attributes x and y). Later, we will extend the feature space to any dimension.

Definition 1: Considering two images that are being compared, attribute concurrence (AC) is the proportion of the variation of the attribute values obtained from the respective feature vectors.

$$AC(Q,C) = \frac{\min\left(|q_x - c_x|, |q_y - c_y|\right)}{\max\left(|q_x - c_x|, |q_y - c_y|\right)} \qquad (1)$$

where Q and C are the images being compared with feature vectors $q=(q_x, q_y)$ and $c=(c_x, c_y)$.

This concept is illustrated in Figure 3, over a unitary spatial grid showing attributes x and y, and considering four images, Q, C_1, C_2, and C_3. Images Q and C_1 present the same values of differences of attributes x and y (AC=1=maximum), whereas images Q and C_2 present a difference of attribute x three times that of attribute y (AC=0.33), and images Q and C_3 present differences only for attribute x (AC=0=minimum).

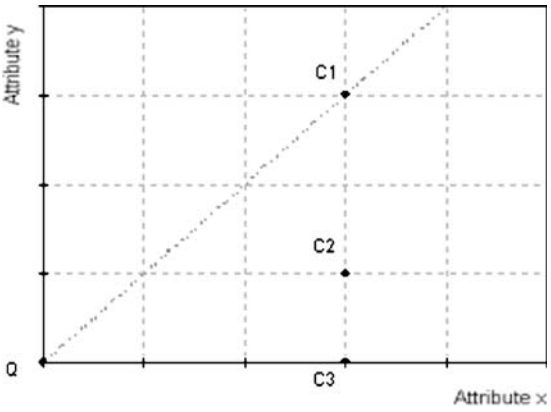

Fig 3. Degrees of AC.

How does a person perceive variations in similarity when two images present high (near 1) or low (near 0) values of AC? We will see that it depends on the features and on the context.

Definition 2: Attribute concurrence influence (ACI) is the effect of AC on the human perception of similarity. In some contexts, high values of AC lead to a perceptual effect of high dissimilarity, in which case we say that the ACI is *strong*. In contexts where high values of AC lead to a perceptual effect of low dissimilarity, we say that the ACI is *weak*.

Let us see an example of these effects by comparing images of uniform colors. Consider the values of red, green, and blue (RGB) components of the color as the attributes that will represent the images having 8 bits of quantization (256 levels). Image Q is composed of RGB values (130, 30, 30), whereas image C_1 is (130, 110, 110) and image C_2 is (130, 130, 30). In the G–B space, the pair (Q, C_1) presents a high value of AC (=1, considering attributes G and B), whereas the pair (Q, C_2) presents no AC (=0, difference only in attribute G). With some tests concerning the human analysis of boards consisting of different uniform colors, we could conclude that this kind of

color comparison occurs in a context of weak ACI, once the person considers that the (Q, C_2) image pair presents greater dissimilarity than the (Q, C_1) pair. In other words, variations in only one attribute make the images more distant than variations in two attributes, so we can say that the concurrence influence is weak. In this context of comparison, a distance function that considers the weak ACI will produce better results than a distance function that considers the strong ACI or one that does not consider the ACI.

Proposed Distances

To deal with weak and strong concurrence influence, we propose two families of distance functions: weak attribute concurrence influence distances (WAID) and strong attribute concurrence influence distances (SAID). Still considering two attributes, both families, in this proposal, are represented by polynomials of degree 2 that define the geometric place of the objects at the same distance from a reference one. In Figure 4, all the points on the WAID curve are at a distance d_w from Q, whereas all the points on the SAID curve are at a distance d_s from Q. Taking point C, if we decide to work with WAID, its distance from Q (d_w) will be smaller than the distance where we work with SAID (d_s).

The use of WAID considers that objects with high values of AC are closer, whereas the use of SAID considers these objects as more distant. Thus, WAID is designed to compare objects in contexts where ACI is weak, and SAID is designed to compare objects in contexts where ACI is strong. Again analyzing Figure 4, points Q and C present a high value of AC. In a context where ACI is weak, the distance between Q and C should be small (d_w of WAID); and in a context where ACI is weak, the distance should be high (d_s of SAID).

In this section, the curves are described only in a qualitative way. The formal mathematical definition is build up in the "General Expression of SAID and WAID" section.

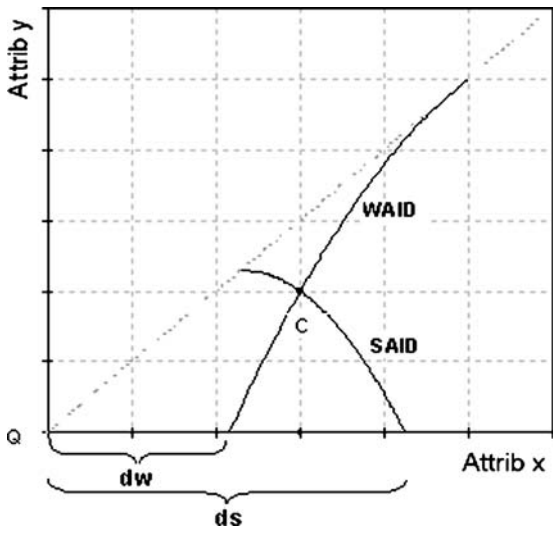

Fig 4. WAID and SAID curves. Only the first octant ($|q_x - c_x| > |q_y - c_y|$) is represented. In the other octants, the identity lines mirror the curves.

Degree of Attribute Concurrence Influence

To quantify and control the effect of ACI, we define a parameter n that represents the *degree of concurrence influence* and determines the elongation of the curves. This leads to a family of curves for WAID and another family for SAID. Figure 5 gives examples of curves for SAID and WAID families.

As a rule of thumb, the stronger the ACI the higher the value of n of the most suitable curve for SAID. In contrast, the weaker the ACI the higher the value of n of the most suitable curve for WAID.

General Expression of SAID and WAID

To simplify the notation, in this section, we call $|q_x - c_x|$ x and $|q_y - c_y|$ y; in other words, x and y are now attribute differences. Considering the first octant, x is always greater than y.

Considering the definitions proposed in the "Proposed Distances" section, the AID families can be described by mathematical functions of degree 2. The equations related to these functions depends on the values of n. Thus, each curve shown in Figure 5 will have a specific equation.

SAID Family

To determine the quadratic expression of the SAID family, considering $y=f(x)$ a polynomial, we use the following constraints (shown in Fig. 6):

- $x=d_s \rightarrow y=f(d_s)=0$ (the point where the curve intercepts the x-axis is the value of the distance)

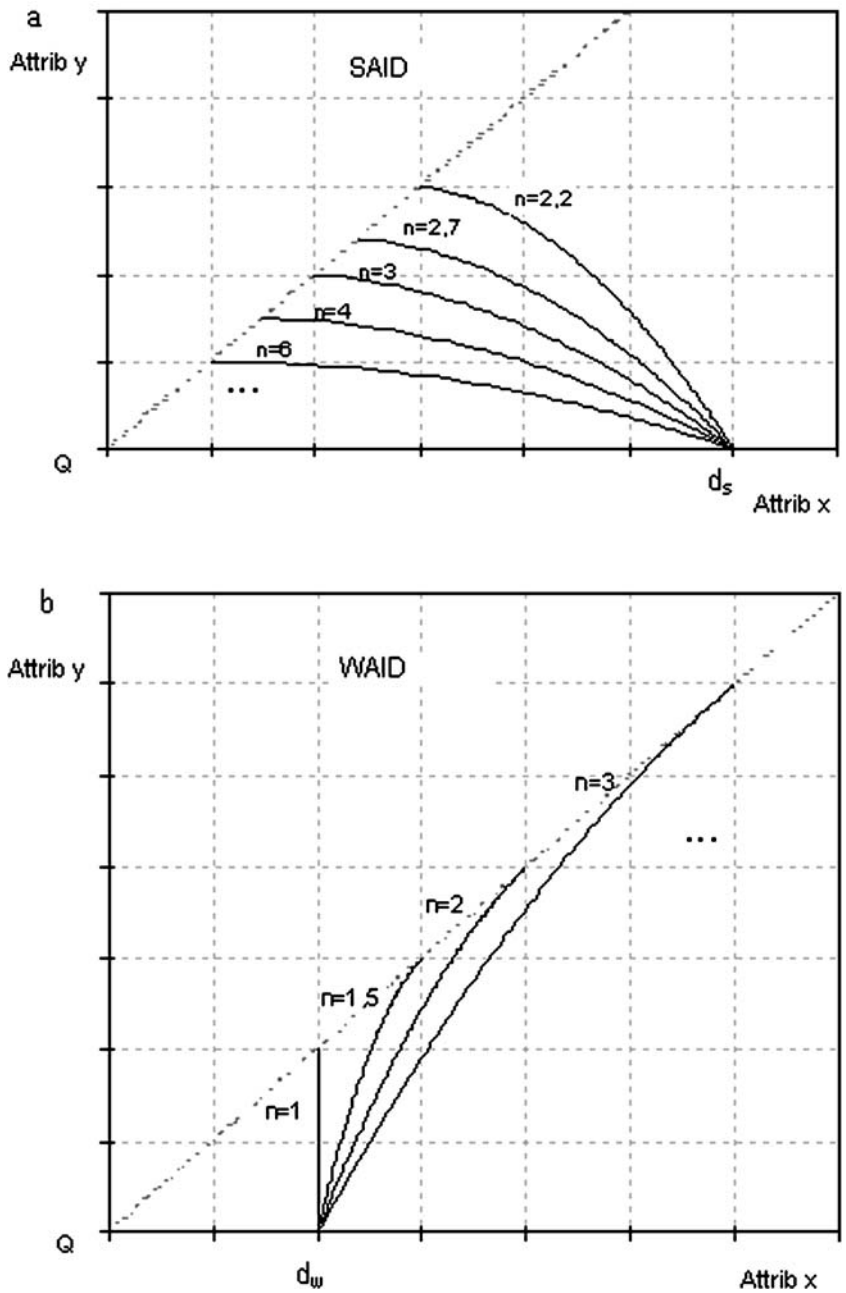

Fig 5. Degree of ACI—geometric place of points located at distance *d* for different values of *n*: a SAID family, b WAID family.

- $x=d_s/n \rightarrow y=f(d_s/n)=d_s/n$ (the elongation of the curve can be defined as the proportion between the point where the curve intercepts the x-axis and the point where the curve intercepts the identity line. Thus, if the elongation is n, then the x coordinate of the point where the curve intercepts the identity line is d_s/n)
- $f'(d_s/n)=0$ (f is maximum at d_s/n)

where d_s is the value of the SAID distance defined for all the points on the curve (Fig. 4) and n is the

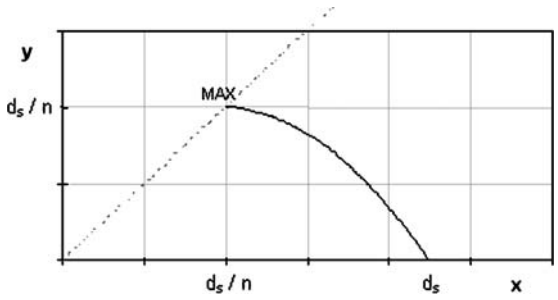

Fig 6. Conditions to determine the general expression of SAID family.

degree of concurrence influence (elongation of the curve) (Fig. 5).

With these constraints, we can determine the value of d_s as follows:

$$d_s = \frac{y(n-1)^2 - 2x + (n-1)\sqrt{4x^2 - 4xy + y^2(n-1)^2}}{2(n-2)} \quad \text{for n} > 2 \tag{2}$$

WAID Family

To determine the quadratic expression of the WAID family, considering $y=f(x)$ a polynomial of degree 2, we use the following constraints (shown in Fig. 7):

- $x=d_w \rightarrow y=f(d_w)=0$ (the point where the curve intercepts the x-axis is the value of the distance)
- $x=n\ d_w \rightarrow y=f(n\ d_w)=n\ d_w$ (n is the elongation of the curve)
- $f'(n\ d_w)=1$ (f tangent to identity at $n\ d_w$)

where d_w is the value of the WAID distance defined for all the points on the curve (Fig. 4) and n is the degree of concurrence influence (Fig. 5).

With these constraints, we can determine the value of d_w as follows:

$$d_w = \frac{(n-1)^2(x-y) + 2nx + \sqrt{\left((n-1)^2(x-y) + 2nx\right)^2 - 4n^2x^2}}{2n^2} \tag{3}$$

Extending AID to any Dimension of the Feature Space

The expressions of the distances d_s and d_w were determined in the "General Expression of SAID and WAID" section for the families SAID and WAID, considering feature vectors with only two attributes. Now, we extend those expressions to any dimension, simplifying them by using L_1 and L_∞ distances.

In expressions 2 and 3, to compute the distance between two images, the variable x represents the difference $|q_x - c_x|$ in the values of one attribute and y represents the difference $|q_y - c_y|$ in the values of the other attribute. In addition, x represents the maximum difference. If there are more than two attributes, we should analyze the interaction between x (the maximum attribute difference) and each of the other attributes in a separate plane, and then join them in a final expression. To do this, L_1 and L_∞ can be very helpful:

$$L_\infty = x = \max \text{attrib} \tag{4}$$

$$L_1 = \sum \text{all\_attribs} \tag{5}$$

where attrib represents the difference of values of a given attribute for two images that are being compared.

Extending SAID family

SAID presents a behavior that is similar to L_1, i.e., the attributes are summed to compose the final distance. Thus, to determine the SAID distance, y will be replaced by the sum of the other attributes (all the attributes, excluding x) values:

$$y = \text{sum\_other\_attribs} = \sum \text{all\_attribs} - \max \text{attrib} \tag{6}$$

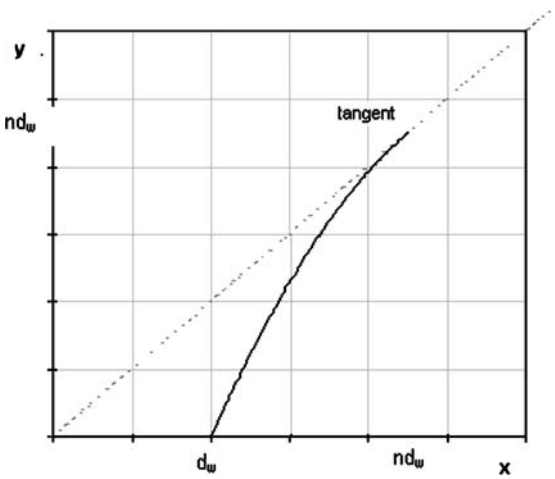

Fig 7. Conditions to determine the general expression of WAID family.

Thus,

$$y = L_1 - L_\infty \tag{7}$$

Now, replacing the values of x (Eq. 4) and y (Eq. 7) in expression 2, we determine the SAID distance between two objects in a multidimensional space, as a function of L_1, L_∞ and n:

$$d_s = \frac{(L_1 - L_\infty)(n-1)^2 - 2L_\infty + (n-1)\sqrt{4L_\infty^2 - 4L_\infty(L_1 - L_\infty) + (L_1 - L_\infty)^2(n-1)^2}}{2(n-2)} \tag{8}$$

Extending WAID family

Analyzing the relationship between x and each of the other attributes, we will take the value of the average to determine the final distance. To do this, y will be replaced by the average of the values of all the attributes, excluding x:

$$y = \text{ave\_other\_attribs} = \left(\sum \text{all\_attribs} - \max \text{attrib}\right) \Big/ (\dim - 1) \tag{9}$$

where dim is the total number of attributes

Thus,

$$y = (L_1 - L_\infty)/(\dim - 1) \tag{10}$$

Now, replacing the values of x (Eq. 4) and y (Eq. 10) in expression 3, we determine the WAID distance between two objects in a multidimensional space, as a function of L_1, L_∞, n and dim:

$$d_w = \frac{(n-1)^2(L_\infty - (L_1 - L_\infty)/(\dim - 1)) + 2nL_\infty + \sqrt{\left((n-1)^2(L_\infty - (L_1 - L_\infty)/(\dim - 1)) + 2nL_\infty\right)^2 - 4n^2L_\infty^2}}{2n^2} \tag{11}$$

Composing AID with L_p

Analyzing the behavior of L_1, L_2, and L_∞ distances, we can see that they present differences in AC. In Figure 8, both points C_1 and C_3 present AC=1, whereas point C_2 presents AC=0. Using L_2, the distances from C_1 and C_3 to Q are r_1 and r_3. As for point C_1 (AC=1), if we choose L_1, the distance from C_1 to Q is r, which is larger than r_1. Thus, L_1 considers point C_1 farther from Q than L_2 does. Concerning point C_3 (AC=1), if we choose L_∞, the distance from C_3 to Q is r, which is smaller than r_3. Thus, L_∞ considers point C_3 closer to Q than L_2 does. As for point C_2 (AC=0), regardless of the distance we choose (L_1, L_2, or L_∞), the distance from C_2 to Q is the same (r). The largest difference in L_1, L_2, and L_∞ occurs in the identity line (AC=1), which decreases in the direction of the axes until it becomes zero over the axes (AC=0). Taking L_2 as reference and considering that L_1

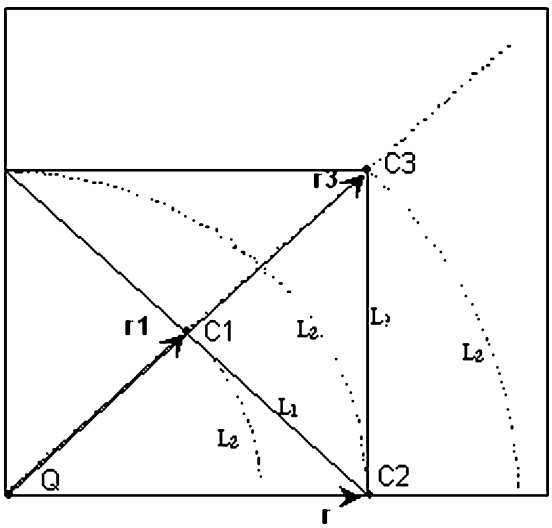

Fig 8. Comparison of L_1, L_2, and L_∞.

places the points with high AC at a greater distance, we can say that L_1 is appropriate for contexts where ACI is strong. On the other hand, considering that L_∞ places the points with high AC at a lesser distance, we can say that L_∞ is appropriate for contexts where ACI is weak.

However, the use of the L_p norm has the following drawbacks: (1) It is restricted to a specific space region; (2) L_1 and L_∞ are linear in terms of the proportion of attribute variation; (3) L_∞ is not suitable for comparing images, as it considers as equal objects those that present high values of dissimilarity in different attributes.

The SAID and WAID families can compose with the L_p family to cover the whole feature space, as shown in Figure 9. The upper bound of SAID is L_1 (SAID with $n=2$) and the lower bound of WAID is L_∞ (WAID with $n=1$). In fact, the L_p family is not limited by L_1, for it can continue in the lower region with values of p smaller than 1, but these functions do not reflect attribute interaction, as the variations that they present along their curves are not consistent with this approach.

The composition of L_p and AID families engenders a scope of distances (which can be extended by varying the parameter n) that allow the AC to be evaluated and quantified precisely for any context. Based on this reasoning, we can determine the distance function (Fig. 10) that best provides the value of perceptual dissimilarity between two objects in that context.

Determining the Best Perceptually Fitted Distance Function

Given the AID and L_p families of distance functions, we define a two-step procedure to determine the best distance for a specific application, as follows:

Step 1: Determine the ACI. With a set of training images, similarity tests with the distances SAID ($n=3$), L_1, L_2, L_∞, and WAID ($n=3$) are run. Analyze the results of each distance and compare them to the results of evaluations by human specialists. With the best distance function obtained for this context, determine the ACI. If SAID ($n=3$) is the most suitable, then we have a strong AC context. If WAID ($n=3$) is the best

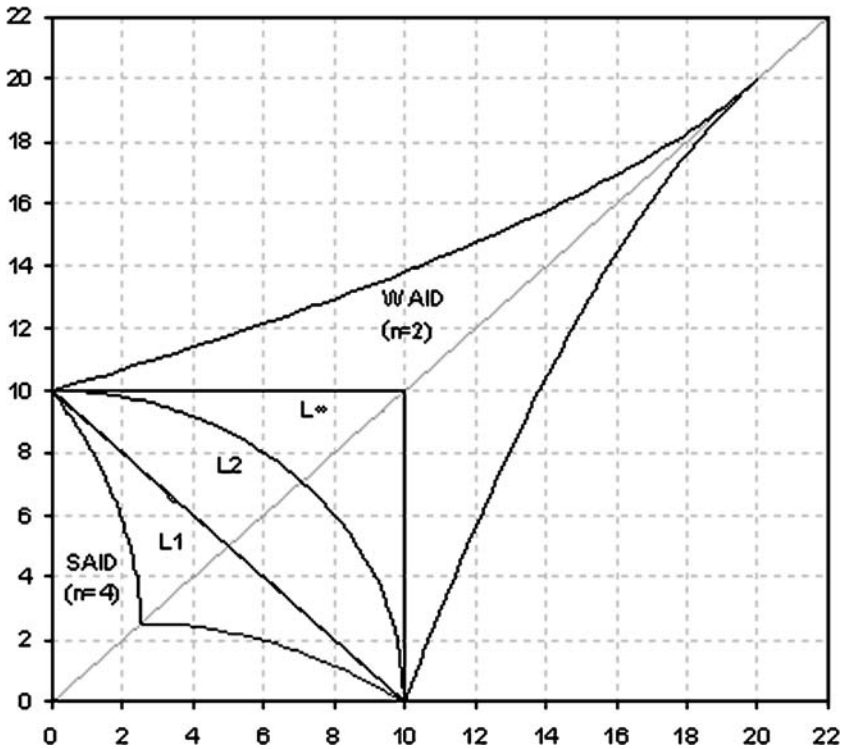

Fig 9. Composition of L_p and AID.

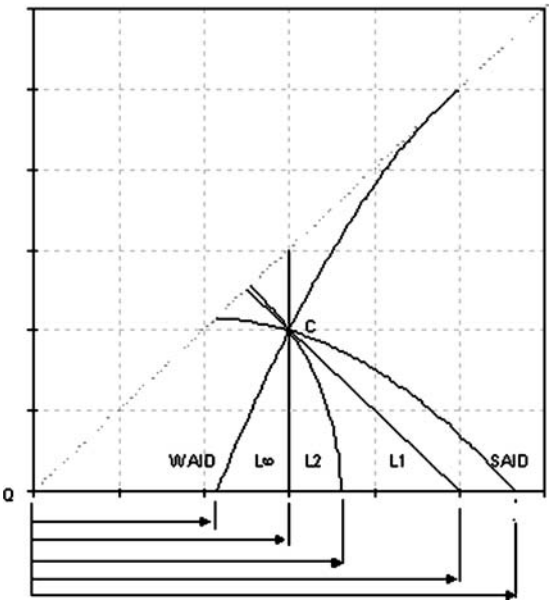

Fig 10. Variations of distance between *C* and *Q* using the distance functions of L_p and AID families represented in the first octant of 2-D space.

choice, considering the human specialists, then we have a weak AC context. If one of the L_p distances is the best, we have an intermediate concurrence influence context.

Step 2: Determine the degree of ACI and the best distance function. If the concurrence influence is intermediate, then the best distance function is the one that achieved the best results in step 1. If the concurrence influence is strong, we conduct new tests with SAID considering different values of n to determine the best one. We carry out the same procedure with WAID if the concurrence influence is weak.

This procedure identifies the best perceptual distance function for a specific context, as we will show in the experiments below.

EXPERIMENTS AND RESULTS

In this section, we describe three experiments dealing with different medical images represented by different descriptors. For each context, we took similarity results obtained with the distance functions of families L_p and AID and compared them with results obtained from humans (radiologists) to determine the ACI and then the distance function that best fits the human judgment.

The process of feature extraction was carried out using a CBIR tool, developed by us, which can extract several features related to texture (homogeneity, variance, energy, entropy, Wavelets, and histogram gradient), color (conventional and metric histograms), and shape (Zernike moments, Fourier transform, and geometric features). In addition, this tool can apply similarity retrieval in image databases. The user can choose a distance function, a set of descriptors, an image database, and a reference image, and see results of k-nearest neighbors.

The images used in the experiments were produced in the Radiology Department of the School Hospital (HC-FMRP) of University of São Paulo at Ribeirão Preto, Brazil. They are stored in the Center of Image Sciences and Medical Physics of the Medicine School of University of São Paulo.

Perceptual Similarity of Texture for Medical Images

In this experiment, we used a set of 30 images that consist of regions of interest (ROIs) extracted from medical images of magnetic resonance, computerized tomography, and mammography (Fig. 11). The tomography images were obtained from a Siemens equipment (Magneton Vision, 1.5 Tesla).

Each ROI in this set of images is a uniform sample of some body tissue: brain, breast, lung, bone, liver. The idea behind this experiment is to evaluate the perception of similarity of the human specialist about texture, disregarding a specific task or a specific class of images, considering only his/her notion of spatial distribution of gray points in each image. Texture is an abstract concept related to the effect of that spatial distribution on the human perception. This effect produces the notion of dissimilarity between images, regarding texture.

Five radiologists (R1, R2, R3, R4, and R5) from HC-FMRP were asked to rank the images in order of similarity with a given reference image, comparing the texture in the images. They were asked to pay attention just to the gray distribution on the images without considering the images' modality or the body region they were acquired from.

We then used the CBIR tool to obtain the texture features employed to ask for the k-nearest neighbors of the same reference image. The texture descriptors used were *uniformity* and *homogeneity*,[36] extracted from the corresponding cooccurrence

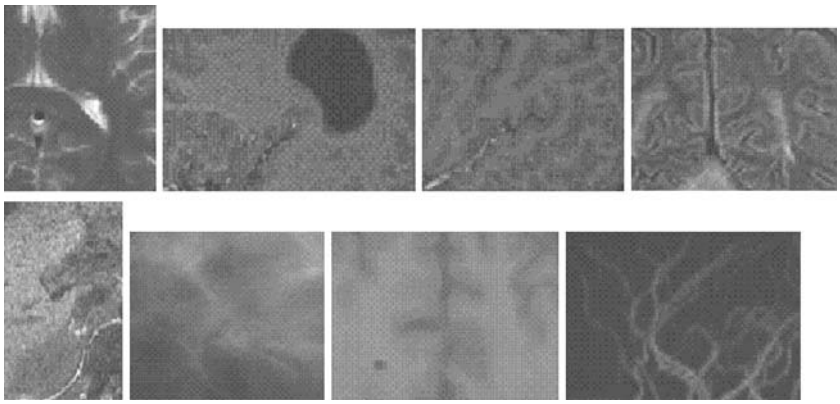

Fig 11. Samples of images from the set of tests with texture attributes.

matrix. The techniques that were used to preprocess the images is described by Felipe.[35] The k-nearest neighbor queries were executed for each distance function of the families AID and L_p as described in step 1 of the method presented in the "Determining the Best Perceptually Fitted Distance Function" section. The algorithm also employed k=30 to identify the rank of the images considering each image of the database. The choice of a small image set was made to enable the physicians to manually rank the images by similarity.

Each sequence ranked by the radiologists was then compared with the results of each distance function. To analyze the results and evaluate the conformity of the automatic sequences to each radiologist's sequence, we calculated the sequence conformity degree (CD) as follows:

- divide the sequences under comparison into subsequences of k images;
- count the number "num" of images present simultaneously in both subsequences (provided by the radiologist and by the distance function);
- vary k from 2 to 29 and calculate CD(k)= num/k;
- calculate the final sequence CD as the average of CD(k).

Under a probabilistic analysis, we can consider that a CD value of 0.50 is the result of a random process. Thus, a distance function must produce a sequence CD value higher than 0.50 to be considered effective. This kind of consideration is valid and often used in the analysis of receiver operating characteristic (ROC) curves.

To calculate the accuracy of each distance function as parameter of conformity with a specific radiologist, we define *precision* as follows:

$$\text{precision} = \frac{\text{CD} - 0.50}{0.50} \tag{12}$$

Figure 12 presents the plots of the precision achieved with each distance function, considering each radiologist. Table 2 presents these values of precision and the gain of average precision for each distance function, taking as reference the L_2 function, which is the most often used one.

An analysis of the curves reveals that despite the large variation in the results of the human specialists, the SAID function consistently reached the best precision for all of them. It indicates that this kind of context is characterized by strong ACI. From Table 2, we can see that SAID (n=3) reached an average precision of 0.49, against WAID (n=3) with an average precision of 0.37. Taking L_2 as reference, we can see that WAID (n=3) presents a loss of 18.6% in precision, whereas SAID (n=3) presents a gain of 8.7%.

In step 2 of the approach, we applied SAID with different values of n to determine its best value. Figure 13 presents the resulting curves for some values of n varying from 3 to 15.

Analyzing the curves, we can see that n=3 and n=4 present similar performances, whereas the performance is reduced at values higher than 4. Thus, our experiments lead us to conclude that, when these descriptors are used to represent the image texture, the best distance function is SAID with n=3.

There is an important remark regarding these results: a huge variability between the results for

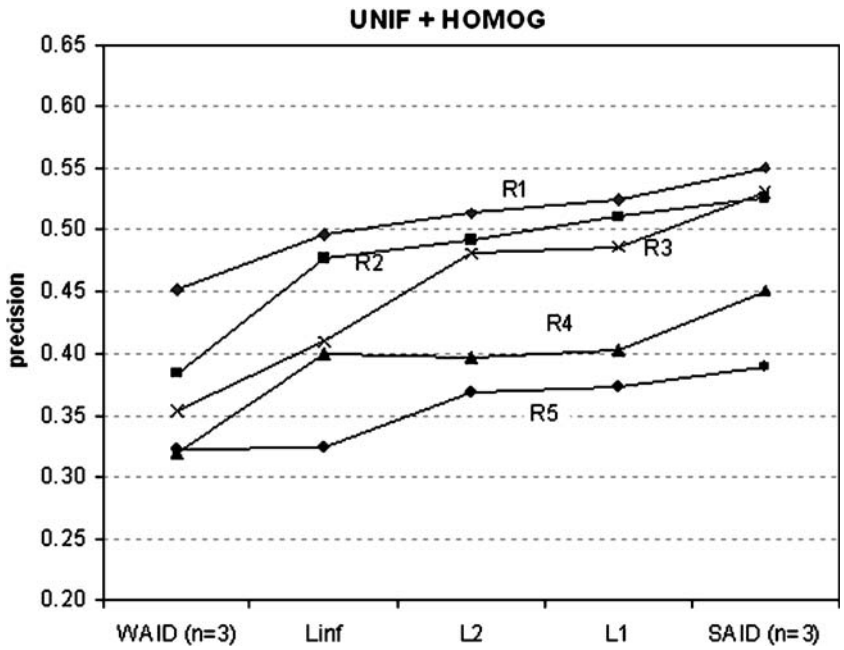

Fig 12. Curves of precision in evaluating the accuracy of distances. Each curve corresponds to the results of one radiologist (*R1* to *R5*).

each radiologist. These differences occur because of the way the analysis was conducted where the radiologist compared the images only attempting to their intrinsic texture effect without following any analysis pattern. In spite of this variability, it is important to observe that the performances of all functions are the same for all radiologists. In the "Similarity of Medical Images Based on Wavelets and Histogram Gradient Descriptors" section, we describe an experiment based on the radiologist diagnosis.

After this, to work with four features, we added two new texture attributes to the experiment: *variance* and *entropy*, which are also obtained from the cooccurrence image matrix. The results reached in step 1 are shown in Figure 14.

The values of average precision for this experiment were: WAID=0.37, L_∞=0.45, L_2=0.45, L_1=0.47, and SAID=0.47. Analyzing these values, we see that in this context and with these attributes, the differences in precision of all functions are similar and not expressive. The best precision is reached by L_1 and SAID with n=3 functions, which leads to the same average performance. To study the performance of other values of n, we executed step 2 varying n from 3 to 15, and found that for n values higher than 3, SAID does not generate better results.

Similarity of Medical Images Based on Region Segmentation

This experiment was aimed at evaluating the ability of the distance functions to discriminate images from different regions of the human body. We used a medical image database with 704 magnetic resonance images, separated into 8 classes: axial head, coronal head, sagittal head, axial pelvis, axial abdomen, coronal abdomen, angiogram, and sagittal spine. These classes were predefined by the exam categories.

To analyze the results and to evaluate the efficacy of the distance functions, the well-known concepts of *precision* and *recall* were applied.[37] A

Table 2. Values of Precision in Evaluating the Accuracy of Distances

Radiologist	WAID (n = 3)	L_∞	L_2	L_1	SAID (n = 3)
R1	0.45	0.50	0.51	0.52	0.55
R2	0.38	0.48	0.49	0.51	0.53
R3	0.35	0.41	0.48	0.49	0.53
R4	0.32	0.40	0.40	0.40	0.45
R5	0.32	0.32	0.37	0.37	0.39
Average	0.37	0.42	0.45	0.46	0.49
Gain % (ref = L_2)	−18.6	−6.4	0.0	2.0	8.7

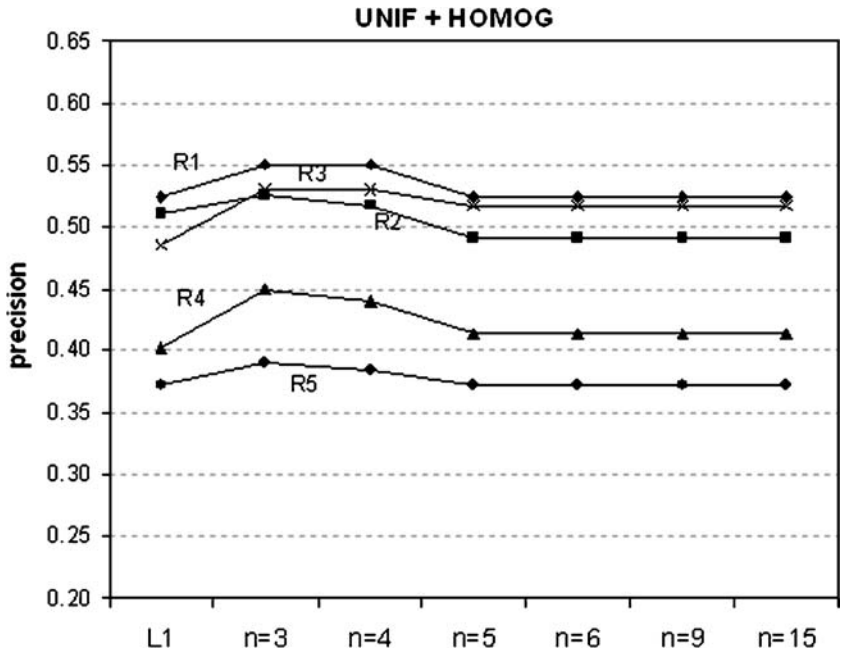

Fig 13. Precision in determining the best function regarding parameter *n* (the five radiologists were considered).

rule of thumb in analyzing such curves is that the closer to the top the better the method. The curves were plotted using the average of the precision and recall values in 20 queries.

Each image was segmented into five regions employing a stochastic approach based on EM/MPM.[38] This number of regions is determined by the calibration of the algorithm based on the search for distinct and significant regions. Six features were extracted from each region: the fractal dimension, the x and y coordinates of the center of mass, the mass, the average gray level, and the

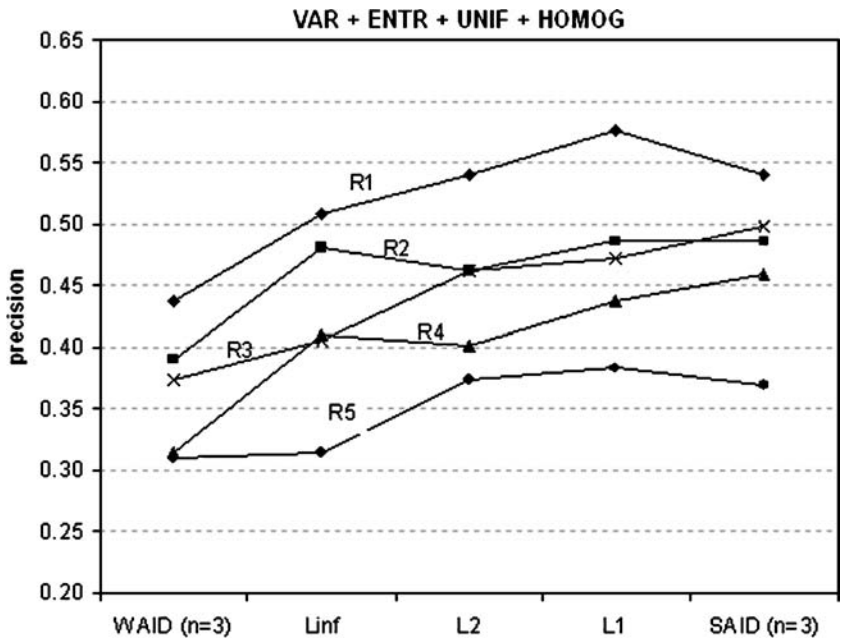

Fig 14. Precision of four texture attributes.

linear coefficient of the fitting line used to obtain the fractal dimension. Therefore, for each image, a feature vector consisting of 30 attributes was defined.

k-nearest neighbor queries were executed taking sampled images randomly from all classes as reference and using the distance functions of families AID and L_p. Individual precision and recall values were calculated and the averages of all classes were used to generate precision vs recall curves for each distance function. Figure 15 shows these curves.

In Figure 15 we can see that the best graph was provided by SAID. For a value of recall of 0.50, the values of precision of WAID (n=3), L_∞, L_2, L_1, SAID (n=3), and SAID (n=6) are 0.41, 0.50, 0.73, 0.86, 0.93, and 0.90, respectively. Taking L_2 as reference, we see that SAID (n=3) presents a gain in accuracy of 27.4% for a recall of 0.50.

Once the best distance function is SAID, we conclude that this context is also characterized by strong ACI. We also see that SAID with n=3 is better than SAID with n=6 (we have processed the data with other values of n, but the best results were obtained with n=3).

Similarity of Medical Images Based on Wavelets and Histogram Gradient Descriptors

This experiment served to evaluate the ability of the distance functions to discriminate pathological pulmonary images from normal ones. We used a medical image database consisting of 150 ROIs extracted from lung radiographs, 50 of which were analyzed as normal regions and 100 presented one of the following pathologies: fibrosis, sclerosis, lymphangitis, mycosis, sarcoidosis, and silicosis (Fig. 16). The analysis and diagnosis of these images were carried out by a senior radiologist from HC-FMRP.

Each image was subjected to the following procedure:

- The Haar wavelets with two levels were extracted, generating seven subbands;
- for each subband, the gradient histogram equalized to 16 ranges of gray level was calculated,[39] i.e., 16 attributes;
- the feature vector concatenating the 7 gradient histograms was composed, resulting in 112 attributes;

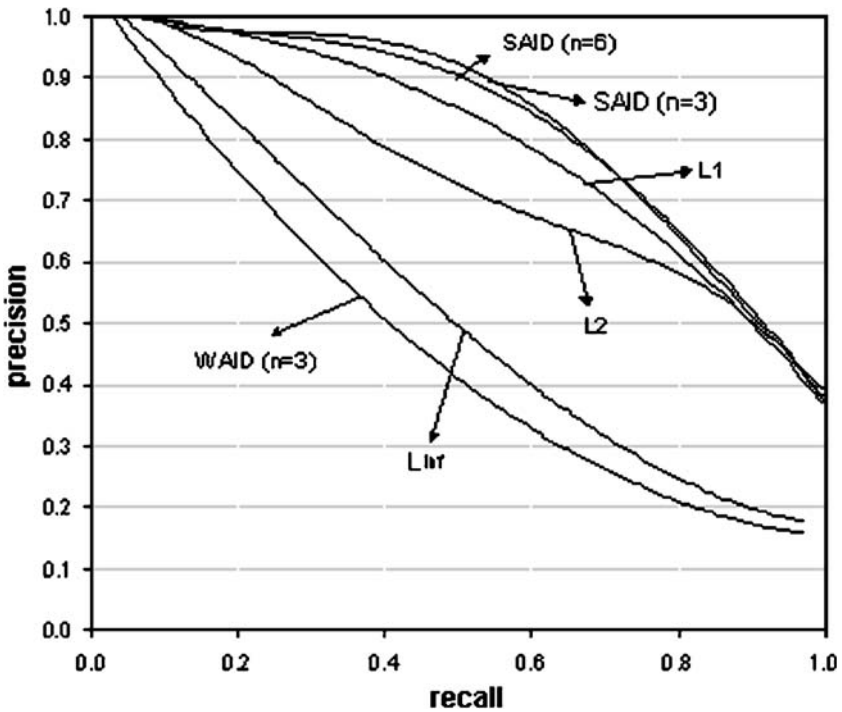

Fig 15. Precision vs recall for image segmentation.

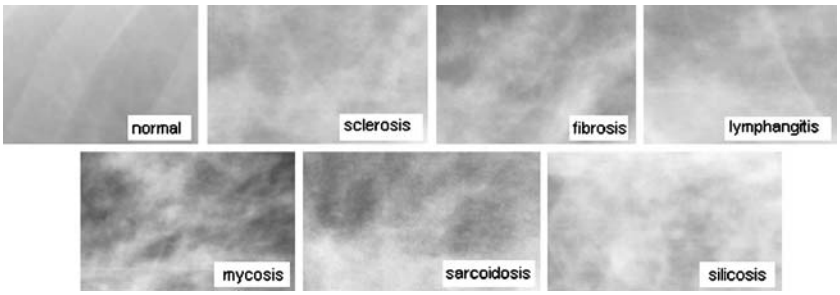

Fig 16. Samples of lung ROIs presenting specific pathologies.

- the *relief attribute evaluation* algorithm was applied to determine the most relevant attributes for discriminating the images as either normal or pathological. The free software Weka (University of Waikato, New Zealand) was used, and 16 attributes were selected;
- these 16 attributes were used to represent the image.

k-nearest neighbor queries were then executed taking all images from both classes as reference and using the distance functions of families AID and L_p. Individual precision and recall values were calculated and the averages for both classes were used to generate precision vs recall curves for each distance function. Figure 17 shows these curves.

Figure 17 indicates that the best graph was again provided by SAID. For a value of recall of 0.50, the values of precision of WAID (n=3), L_∞, L_2, L_1, SAID (n=3), and SAID (n=6) are 0.53, 0.54, 0.57, 0.60, 0.65, and 0.64, respectively. Taking L_2 as reference, we see that SAID (n=3) presents a gain in accuracy of 14.0% for a recall of 0.50.

So this context is also characterized by strong ACI. It also shows that SAID with n=3 is slightly

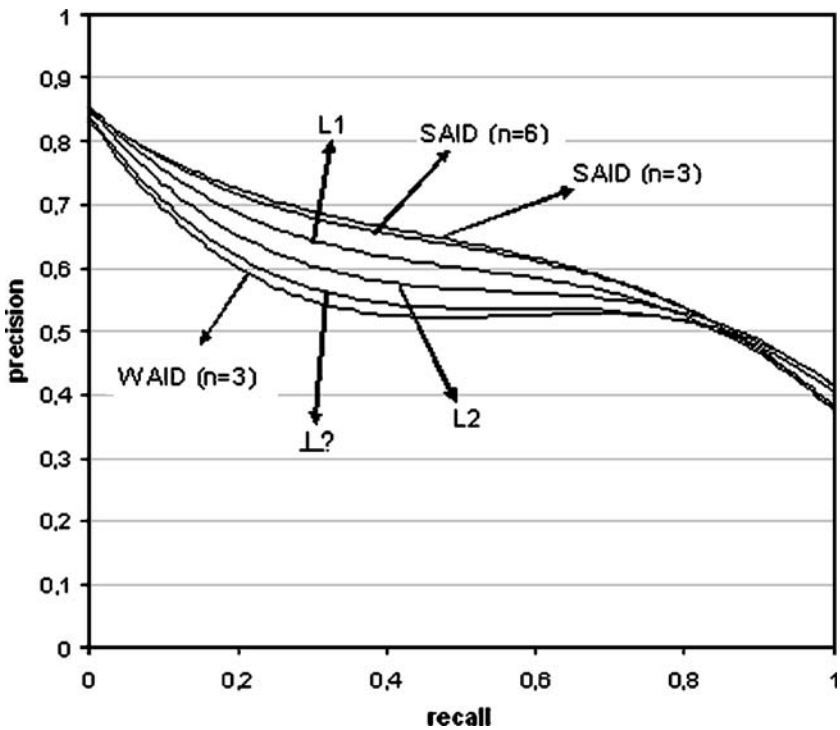

Fig 17. Precision vs recall for wavelets and histogram gradient descriptors.

better than SAID with $n=6$ (we processed the data with other values of n, but the best results were attained with $n=3$).

DISCUSSION AND CONCLUSIONS

Two new families of distance functions for image comparison were presented in this paper. These families—SAID and WAID—take into account the effects of concurrence among attribute differences when two images are compared, reflecting the human perception of similarity. Thus, SAID and WAID represent a new way to summarize one aspect of human perception: the influence of AC in similarity judgments.

The AC and ACI concepts were formally defined. The proposed families of functions were developed mathematically to define the formal expressions needed to calculate the distance between objects.

The composition of AID and L_p families was evaluated, and an experimental approach was described to determine the most suitable function to perform similarity retrieval in a specific context.

Results of experiments of the proposed distance functions were presented. The SAID and WAID families were compared to the well-known distances L_1, L_2, and L_∞ because these distances are most often used in general contexts of feature vector comparison for similarity evaluation in CBIR environments. Other distance functions—such as the Mahalanobis distance and the Jeffrey divergence—have presented good results in specific contexts, as shown by other researches, but the majority of them require special information—such as a matrix of covariance of attributes for the Mahalanobis distance—in addition to the feature vectors, or need special conditions of the feature vectors content to be applied. Even so, the comparison between AID family and other distances than the L_p family is subject for a future work.

Using the results of tests involving five radiologists, we determined that the best function for representing a general concept of texture of medical images is SAID with an interaction coefficient of 3. The results show that with the use of texture descriptors uniformity and homogeneity, the SAID distance outperforms the L_2 distance with a gain of 8.7% in precision. The purpose of comparing images from different modalities using a general perception of texture conducted to a variability among the results of radiologists, but it also conducted to an interesting result, where we can see that the comparative degree of accuracy presented by the distances was the same for all radiologists.

Two other experiments were conducted, one employing six features from a database of segmented images from different human body parts and the other using gradient histograms over Haar wavelets from a database of normal and pathological lung images. Precision vs recall curves were plotted to evaluate the distance functions performance in each context. In the first experiment, taking L_2 as reference, SAID ($n=3$) presented a gain in accuracy of 27.4% for a recall of 0.50, and we can conclude that the evaluation of L_p and AID distances again showed that the best function in terms of discriminative power is SAID with a coefficient of 3. In the experiment with lung images, taking L_2 as reference, SAID ($n=3$) presented a gain in accuracy of 14.0% for a recall of 0.50, and we can again conclude that the best function for this context is SAID with a coefficient of 3.

There is not a specific distance function that is the best one for any context of CBIR involving medical images. We proposed new distance functions and we propose that, for any different context of image modality and feature descriptor, a comparative study has to be carried out to determine the best distance function for that context. Inside a specific context, the scale of the attributes' values remains constant; thus, the study comparing distance functions does not need to consider variations in scale. Although the AID family can be employed as a general approach to compare images by similarity, the focus of this work was the context of medical image analysis, which showed a tendency to present strong ACI where the SAID family of functions achieved the best results when compared to the L_p family.

Otherwise, the study using four texture descriptors (uniformity, homogeneity, variance, and entropy) was not conclusive regarding the best distance function, as the results did not present statistical relevance. The results suggest that the SAID distance is the best, but future experiments need to be done with different amount of attributes to evaluate the effectiveness of the proposed

formulation of the AID family extended to any dimension of feature space.

Another promising study that can be carried out in a future work is the application of the AID family to feature vectors generated under appearance model techniques. This study could lead to an interesting analysis of the convergence of two distinct methodologies, both of them based on the modeling of human perception.

The proposed family of distance functions helps answer the question: "How does the human perception of similarity change when image attributes vary together?" Moreover, our experiments showed that the use of a well-suited distance function involving AC in a specific context allows for a reduction of the semantic gap in automated methods of similarity-based image retrieval.

ACKNOWLEDGMENTS

The authors are grateful to the radiologists from HC-FMRP, Ribeirão Preto, Brazil who participated in the experimental studies and the researchers from the CCIFM-FMRP who provided the images for the experiments. This research has been supported, in part, by the Brazilian National Research Council (CNPq) under grants 52.1685/98-6, 860.068/00-7 and 35.0852/94-4 and by the Sao Paulo State Research Foundation (FAPESP) under grant 04/02215-5.

REFERENCES

1. Lehmann TM, Güld MO, Thies C, Fischer B, Keysers D, Kohnen M, Shubert H, Wein BB: Content-based image retrieval in medical applications for picture archiving and communication systems. Proc SPIE 5033:109–117, 2003

2. Zhang H, Su Z: Relevance feedback in CBIR. In: Zhou X, Pu P Eds. Visual and Multimedia Information Management. 1st edition. New York: Springer, 2002, pp 21–36

3. Mathews I, Baker S: Active appearance models revisited. Int J Comput Vis 60(2):135–164, 2004

4. Bregler C, Malik J: Learning Appearance Based Models: Hierarchical Mixtures of Experts Approach Based on Generalized Second Moments. Technical Report No. UCB/CSD-96-897, EECS Department, University of California, Berkeley, 1996

5. Cootes TF, Taylor CJ: Statistical models of appearance for medical image analysis and computer vision. Proc SPIE Medical Imaging 4322:238–248, 2001

6. Müller H, Michoux N, Bandon D, Geissbuhler A: A review of content-based image retrieval systems in medical applications—clinical benefits and future directions. Med Inform 73:1–23, 2004

7. Doi K: Current status and future potential of computer-aided diagnosis in medical imaging. Br J Radiol 78:3–19, 2005

8. Yin T-K, Chiu N-T: A computer-aided diagnosis for locating abnormalities in bone scintigraphy by a fuzzy system with a three-step minimization approach. IEEE Trans Med Imag 23(5):639–654, 2004

9. Doi K: Diagnostic Imaging over the last 50 years: research and development in medical imaging science and technology. Phys Med Biol 51(13):5–27, 2006

10. Giger ML, Karssemeijer N, Armato SG: Computer-aided diagnosis in medical imaging. IEEE Trans Med Imag 20 (12):1205–1208, 2001

11. Kinoshita SK, Marques PMA, Pereira Jr RR, Rodrigues JAH, Rangayyan RM: Content-based retrieval of mammograms using visual features related to breast density patterns. J Digit Imaging 20(2):172–190, 2007

12. Antani SK, Natarajan M, Long JL, Long LR, Thoma GR: Developing a comprehensive system for content-based image retrieval of image and text from a national survey. SPIE Med Imaging 5748:152–161, 2005

13. Inoue M, Ueda N: Retrieving lightly annotated images using image similarities. Proceedings of the ACM Symposium on Applied Computing—SAC, Santa Fé, USA, 2005, pp 1031–1037

14. Bueno JM, Chino F, Traina AJM, Traina Jr C, Marques PMA: How to add content-based image retrieval capability in a PACS. Proceedings of the IEEE International Conference on Computer Based Medical Systems—CBMS, Maribor, Slovenia, 2002, pp 321–326

15. Antani SK, Long LR, Thoma GR: Content-based image retrieval for large biomedical image archives. Proceedings of the 11th World Congress on Medical Informatics (MEDINFO), San Francisco, USA, 2004, pp 829–833

16. Lee CS, Tschai HJ, Kuo YH, Ko WT, Cheng YH, Wang CC: PACS: construction and application to medical image enhancement. Proceedings of the Third International Conference on Knowledge-Based Intelligent Information Engineering Systems, 1999, pp 246–249

17. Kak A, Pavlopoulou C: Content-based image retrieval from large medical databases. Proceedings of the First International Symposium on 3D Data Processing Visualization and Transmission, Padova, Italy, 2002, pp 138–147

18. Lehmann TM, Wein BB, Greenspan H: Integration of content-based image retrieval to picture archiving and communication systems. Proceedings of the Med Inf Europe (MIE), St Malo, France, CD-ROM, 2003

19. Traina Jr C, Traina AJM, Seeger B, Faloutsos C: Slim-trees: high performance metric trees minimizing overlap between nodes. Proceedings of the International Conference on Extending Database Technology, Konstanz, Germany, 2000, pp 51–65

20. Chávez E, Navarro G, Baeza-Yates R, Marroquin JL: Searching in metric spaces. ACM Comp Surv 33(3):273–321, 2001

21. Santini S, Jain R: Similarity measures. IEEE Trans Pattern Anal Machine Intell 21(9):871–883, 1999

22. Rubner Y, Tomasi C: Perceptual metrics for image database navigation. Boston, USA: Kluwer, 2001

23. Hiransakolwong N, Hua K A, Koompairojn S, Vu K, Lang SD: An adaptive distance computation technique for image retrieval systems. Proceedings of the ACM Symposium on Applied Computing (SAC), Santa Fe, USA, 2005, pp 1195–1199

24. Vasconcelos N: On the efficient evaluation of probabilistic similarity functions for image retrieval. IEEE Trans Inf Theory 50 (7):1482–1496, 2004

25. Akleman E, Chen J: Generalized distance functions. Proceedings of the International Conference on Shape Modeling and Applications, Aizuwakamatsu, Japan, 1999

26. Aggarwal CC, Hinneburg A, Keim DA: On the surprising behavior of distance metrics in high dimensional spaces. Proceedings of the 8th International Conference on Database Theory (ICDT), London, England, 2001, pp 420–434

27. Theodoridis S, Koutroumbas K: Pattern Recognition. New York: Academic, 1999

28. Gibbs AL, Su FE: On choosing and bounding probability metrics. Int Stat Rev 70(3):419–435, 2002

29. Zhang DS, Lu G: Evaluation of similarity measurement for image retrieval. Proceedings of the IEEE International Conference on Neural Networks & Signal Processing, Nanjing, China, 2003, pp 928–931

30. Wilson DR, Martinez TR: Improved heterogeneous distance functions. J Artificial Intelligence Res 6:1–34, 1997

31. Tversky A: Features of similarity. Psychol Rev 84 (4):327–352, 1977

32. Tversky A, Gati I: Similarity, separability and the triangle inequality. Psychol Rev 89(1):123–154, 1982

33. Chang EY, Li B, Li C: Toward preception-based image retrieval. Proceedings of the IEEE Workshop on Content-Based Access of Image and Video Libraries, Hilton Head, USA, 2000, pp 101–105

34. Vasconcelos N, Lippman A: A unifying view of image similarity. Proceedings of the International Conference on Pattern Recognition (ICPR), 2000

35. Felipe JC: Development of methods for extraction, comparison and analysis of intrinsic features of medical images, aiming at content-based perceptual retrieval. Ph.D. Thesis, Department of Computer Sciences—ICMC, University of São Paulo at São Carlos, Brazil, 163 p (in Portuguese), 2005

36. Haralick RM, Shanmugan K, Dinstein I: Textural features for image classification. IEEE Trans Syst Man Cybernet 3(6):610–621, 1973

37. Baeza-Yates R, Ribeiro-Neto B: Modern Information Retrieval. New York: ACM Press, 1999

38. Balan AGR, Traina AJM, Traina Jr C: Fractal analysis of image textures for indexing and retrieval by content. Proceedings of the 18th IEEE Symposium on Computer-Based Medical Systems (CBMS), Dublin, Ireland, 2005, pp 581–586

39. Felipe JC, Traina AJM, Traina Jr C: Retrieval by content of medical images using texture for tissue identification. Proceedings of the 16th IEEE Symposium on Computer-Based Medical Systems (CBMS), New York, 2003, pp 175–180