Ranking Invariance Based on Similarity Measures in Document Retrieval

Jean-Francois Omhover, Maria Rifqi, and Marcin Detyniecki

LIP6 – Pole IA, 8, rue du Capitaine Scott, 75015 Paris, France

Abstract. To automatically retrieve documents or images from a database, retrieval systems use similarity measures to compare a request based on features extracted from the documents. As a result, documents are ordered in a list by decreasing correspondance to the request. Several comparison measures are used in the field and it is difficult to choose one or another. In this paper, we show that they can be grouped into classes of equivalent behavior. Then, in a query by example process, the choice of these measure can be reduced to the choice of a family of them.

Keywords: Fuzzy Similarity Measures, Image Retrieval, Aggregation, Segmentation.

1 Introduction

Decriptions, queries and similarity measures are the three basic components of a document retrieval system. In the case of an image retrieval system, an image is indexed through a visual description (color, shape, texture,...) by means of a vector or a set of features. The query part of a retrieval system consists, for the user, in choosing an image or a part of image example. Then the image retrieval system evaluates the similarity between this request and each image of the database (or a part of it). This is done by computing a similarity measure between pairs of descriptions. The comparison of two image descriptions is therefore a fundamental operation for such systems. Obviously then, the choice of a particular similarity measure is a crucial point. In response to its request, the user gets a list of images. This list is ordered by the decreasing degree of similarity to the request.

When a user obtains this list, he tends to ignore the similarity degrees of the images and focuses on the images themselves. What is important to him is the order of this list of documents [10, 8], because he will evaluate the relevance of each resulting images in their order of arrival. The similarity degree is not examined by the user. Sometimes, it is not even displayed. This similarity degree is also discarded in the measures evaluating the efficiency of information retrieval systems. In particular, the recall and precision measures are based on the number of relevant documents in the first results of the system. Hence, they depend on the order in which these relevant documents appear, and not on the relevance values computed by the system for these documents.

M. Detyniecki et al. (Eds.): AMR 2005, LNCS 3877, pp. 55–64, 2006.

As a matter of fact, the value of the similarity itself is unimportant for both the user and the system. The system only uses the value to order the results. The user barely notices this value because his attention is focused on the content of the result images. Based on this fact, choosing between one measure or another to compare visual descriptions is of little interest if two measures lead to the same ordered list.

Very different kind of similarity measures are used in the field. In our image retrieval system [6], many similarity measures can be used to support various queries. In particular, we use the fuzzy similarity measures to compute the similarity between gradual visual descriptors as those used classically in the field. This family of measures, first introduced in a psychological context [14], were adapted to gradual sets [2]. Fuzzy similarity measures can cover various intuitive user needs. They can also be adapted to various descriptors. As an example of the measures falling under this formalism, the classical histogram intersection [12] has widely spread in the image retrieval community.

In this paper we study the set of similarity measures in the perspective of ordering documents relatively to a request. We show that, if the order of the results of a system is indeed the only information to be considered for its use and evaluation, then similarity measures can be grouped in classes of equivalent behavior. For any given request, measures belonging to a same group do provide the exact same result order. We also show that the value of one measure can be predicted from the value of one of its equivalent measures so that we can predict the outcome of a procedure based on similarity values rather than results order. We also study the consequences of this result in the context of image retrieval.

We first introduce the similarity measures in a CBIR perspective (see section 2). We then build a formal theory about order invariance for fuzzy similarity measures (see section 3). The three definitions introduced in this section let us draw equivalence classes that group the similarity measures leading to the same orders. Then, as an application and an illustration, we focus on a specific set of similarity measures, that are Tversky's ratio model measures (see section 4). In this set of measures, we entirely describe the families of equivalent measures. In the last section (section 5) we discuss the consequences of considering equivalent measures on the issue of document retrieval by similarity.

2 Similarity Measures for Image Retrieval Systems

Any image retrieval system bases is action on the computation of similarity measures. Commonly, histogram intersection measures [12], or distance models are widely used in this field [3, 5, 11]. Generalising and covering different similarity measures (as histogram intersection), we use fuzzy similitude measures as a tool to compare image indexes and to evaluate visual similarities.

We first introduce the definition of fuzzy similitude measures, then we apply some of these measures to the comparison of global histograms. Finally, we observe an invariance in the ranking provided by different fuzzy similarity measures. It is not the purpose of this paper to focus on the image representation we use to compare the images.

2.1 Tversky's Ratio Model

Measures of similarity (or dissimilarity) are distinguished into two classes: the geometric one and the set-theoretic one.

Geometric distance models are the most commonly used approach. Objects to be compared are considered as points in a metric space. These models are constrained by 4 properties that a distance has to satisfy: positivity, symmetry, minimality and triangular inequality.

These axioms were in particular studied by [14], who proposed an approach based on more psychological considerations. His study concludes on the very questionable character of the distance axioms. Other studies pointed out the problematic behavior of distances in high dimensional feature spaces [4, 1]. Tversky proposed a set-theoretic definition of similarity. In his scheme, objects to be compared are described by means of sets of binary features. Let a and b be two objects described respectively by the sets of features A and B, and s a measure of similarity, then $s(a, b) = F(A \cap B, A - B, B - A)$ with F a real function of three arguments: the common features $(f(A \cap B))$ and the distinctive features $(f(B - A), f(A - B))$.

His mathematical formulation (called the Ratio Model) introduces a ratio between common features and distinctive features. The two parameters α and β are real positive weights that balance the influence of each part of the distinctive features (those shared by A and not by B and those shared by B and not by A):

$$S(a, b) = \frac{f(A \cap B)}{f(A \cap B) + \alpha f(B - A) + \beta f(A - B)}$$

with α, $\beta \geq 0$, and f an additive interval scale.

2.2 Similitude Measures for Fuzzy Sets

Because of the restriction of the contrast model to binary features, we proposed in [2, 9], a generalisation of Tversky's model to fuzzy features [1]. Furthermore, our framework enables to study particular families of similarity measures according to additional properties corresponding to particular needs and behaviours.

In this framework, for any set Ω of elements, $P_f(\Omega)$ denotes the set of fuzzy subsets of Ω and a fuzzy set measure M is supposed to be given such that $M : P_f(\Omega) \to I\!R^+$ and $M(\emptyset) = 0$ and M is monotonous with respect to $\subseteq$ (for instance $M(A) = \sum_{count} f_A(x)$).

Definition 1. *An M-*measure of comparison *S on Ω is a mapping $S : P_f(\Omega) \times P_f(\Omega) \to [0, 1]$ such that*

$$S(A, B) = F_S(M(A \cap B), M(B - A), M(A - B))$$

where F_S is a mapping $F_S : I\!R^+ \times I\!R^+ \times I\!R^+ \to [0, 1]$ and M a fuzzy set measure on $P_f(\Omega)$.

We denote $X = M(A \cap B)$, $Y = M(B - A)$, $Z = M(A - B)$.

[1] Latter attempts of generalisation to fuzzy features can be found after the publication of our work: see [10], [13].

A measure of comparison captures various families of measures. We are interested in those which evaluate the likeliness of two descriptions. We have called them *measures of similitude.*

Definition 2. *An M*-measure of similitude *S on Ω is an M-measure of comparison S such that $F_S(X,Y,Z)$ is :*

- *monotonous non decreasing with respect to X*
- *monotonous non increasing with respect to Y and Z.*

2.3 Image Retrieval by Fuzzy Ressemblance

Our image retrieval system [7, 6] is based on a regional representation of the images. Regions are extracted automatically by a segmentation algorithm. For each image of an image database, a description of each of its region is computed and stored. The image retrieval can then be driven on the basis of regional queries. The comparison between the descriptors of these regions is computed using an M-measures of similitude. These measures let us build intuitively modifiable queries, and can be aggregated homogeneously to form composite requests (requests for images containing several regions of interest) [6].

For the sake of simplicity, we focus here on the classical global representation (i.e. descriptions of a non segmented image). We represent each image by its global histogram [12], which gives the distribution of pixels' colors for a given color palette $C_1, ..., C_n$. It can be considered as a fuzzy set membership function H_I on the universe $C_1, ..., C_n$. To compute similarities between images, we used fuzzy M-measures of ressemblance to compare their histograms. For two images I_1,I_2, the comparison is done by computing one of the four following measures based on their histograms H_{I_1},H_{I_2}, we denote $X = M(H_{I_1} \cap H_{I_2})$, $Y = M(H_{I_2} - H_{I_1})$, $Z = M(H_{I_1} - H_{I_2})$, M being the area of the given set:

$$S_{jaccard}(X,Y,Z) = \frac{X}{X+Y+Z}$$
$$S_{dice}(X,Y,Z) = \frac{2X}{2X+Y+Z}$$
$$S_{ochiai}(X,Y,Z) = \frac{X}{\sqrt{X+Y}\sqrt{X+Z}}$$
$$S_{Fermi-Dirac}(X,Y,Z) = \frac{F_{FD}(\phi)-F_{FD}(\frac{\pi}{2})}{F_{FD}(0)-F_{FD}(\frac{\pi}{2})}$$

with $F_{FD}(\phi) = \frac{1}{1+\exp^{\frac{(\phi-\phi_0)}{\Gamma}}}$ and

$\phi = \arctan(\frac{Y+Z}{X})$, Γ is a positive real and $\phi_0 \in [0, \frac{\pi}{2}]$. The parameter ϕ_0 controls the point where the decrease will occur whereas Γ controls the decrease speed of the measure. These measures generalize the classical similarity measures to fuzzy sets, well-known in information retrieval particularly.

2.4 Order Induced by a Similarity Measure

In the following sections, we will study the invariance of the order observed within images ranked by their similarity to a query. Such an invariance can be observed on the two simple requests shown on the figure 1. Images are described by a simple global histogram, then two different image requests are given. To answer these two different requests we evaluate In the following sections, we will study the invariance of the order observed within images ranked the similarity of each entry of our image database (Washington Groundtruth) by means of two different similarity measures, that are Jaccard and Dice measures presented in section 2.3. For each query on the figure 1, we clearly see that the results obtained by the two different similarity measures are ordered the same way. Practically, for one user these two results are the same.

Formally, for a database of images $I_1, ..., I_n$, the value returned by a measure S for the comparison of each entry to a query R is used to order $I_1, ..., I_n$ by decreasing ressemblance to R.

A similarity measure as defined in 2.2 is computed based on three real values X,Y and Z (as denoted in definition 1). Then the problem of ordering pairs of images by their similarity is extended to the problem of ordering real triplets of values (X, Y, Z) by a measure of similarity S.

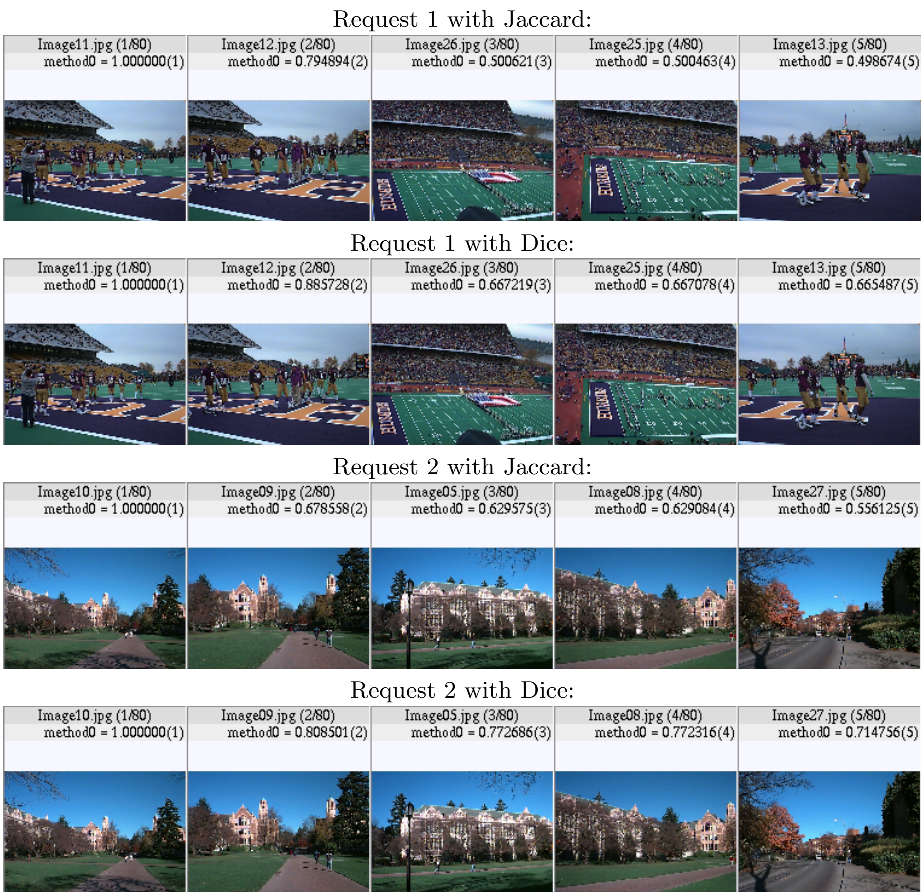

Fig. 1. Two 5-best-results lists using Jaccard and Dice measures applied to image histograms

This problem is not only relevant for the simple purpose of global histogram comparison, but also for any fuzzy set representation, or multiple sets of features. For example, in our CBIR system, similarity measures are computed to match regions in each image I_i with regions in a given query R. As we use a best-matching mecanism, the order induced by the similarity measure between pairs of regions in I_i and R significantly influences the result of the matching.

3 Classes of Equivalent Similarity Measures

Three formal definitions can be proposed for the equivalence relations between similarity measures based on order conservation. These are the two first :

Definition 3. *For any similarity measures S_a and S_b, S_a is "equivalent in order" to S_b if and only if*

$$\begin{array}{l} \forall (X,Y,Z) \in I\!R^{+3}, \forall (X',Y',Z') \in I\!R^{+3} \\ S_a(X,Y,Z) \leq S_a(X',Y',Z') \\ \iff S_b(X,Y,Z) \leq S_b(X',Y',Z') \end{array}$$

Definition 4. *For any similarity measures S_a and S_b, S_a is "equivalent by a function" to S_b if and only if there exists a strictly increasing function:*

$$f : \begin{cases} Im(S_a) \to Im(S_b) \\ x \mapsto f(x) \end{cases}$$

such as $S_b = f \circ S_a$
and $Im(S_t) = \{\alpha / \exists (X,Y,Z) \in I\!R^{+3}, \alpha = S_t(X,Y,Z)\}$.

These two relations are obviously reflexive, symmetrical and transitive. They define equivalence classes within similarity measures. It is also a well known fact that these two definitions are equivalent, meaning that two measures are equivalent in order if and only if they are equivalent by a function.

In a CBIR perspective, what is interesting here is that two measures S_a and S_b that obtain the same results order are linked one to the other by a bijective

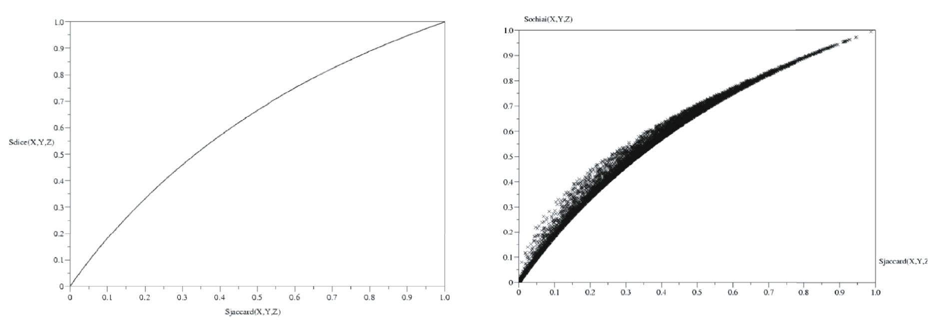

Fig. 2. (a) $S_{jaccard}(X,Y,Z)$ vs $S_{dice}(X,Y,Z)$ (b) $S_{jaccard}(X,Y,Z)$ vs $S_{ochiai}(X,Y,Z)$

function. This means that the value computed by S_b for the evaluation of the similarity of two documents can be predicted from the value computed by S_a for the same purpose. This observation puts this reflexion beyond our preliminary hypothesis : if, ever, the order is *not* the only important information in a result list (as if the results are thresholded by their similarity value, for example), we can still predict the values taken by a measure from any of its equivalent measures. We can then predict the outcome of a procedure relying on the value of the similarity measure (see section 5).

Thanks to these first two definitions of equivalence, we can show that Jaccard, Dice and Fermi-Dirac measures belong to the same equivalence class and that Ochiai's measure belongs to a different class. The figure 2 illustrates the fact that $S_{jaccard}$ can be written as a function of S_{dice} and can not be written as a function of S_{ochiai} : a single value of $S_{jaccard}$ corresponds to many values of S_{ochiai} (and vice versa).

A "value versus value" plot such as figure 2 offers a convenient representation of two measures equivalence. By this kind of graphic, we can simply detect the equivalence of two similarity measures.

The third definition is based on the level sets of the similarity measures. We denote the level set of a similarity measure S_i at the level λ by S_i^λ:

$$S_i^\lambda = \{(X, Y, Z)/S_i(X, Y, Z) = \lambda\}$$

Definition 5. *For any similarity measures S_a and S_b, S_a is said "equivalent in level-sets" to S_b if:*

$$\forall \beta \in Im(S_b),\ \exists!\ \alpha \in Im(S_a) \text{ such that } S_a^\alpha = S_b^\beta$$

This definition means that measures that are equivalent in level sets have a common structure of level sets : they rely on the same level sets but maybe on different levels. Here again we rely on the fact that the value of the similarity is unimportant. This relation is obviously reflexive. It can easily be proven that it is also symmetrical and transitive. We have shown that in the case of continuous measures (as are most of the similarity measures used in the field), this third definition was equivalent to the two first definitions. This can be shown thanks to the monotonicity property of the similarity measures for each of their variables (X,Y,Z).

As a result, we have three definitions leading to the same notion of equivalence. It means that measures that induce the same order in every result list are related one to the other by a function, and their level sets have identical shapes.

The notion of equivalence between similarity measures leads to the construction of equivalence classes. Each class gathers similarity measures that are equivalent one to the other. As a result, families of similarity measures can be drawn ; these families correspond to the sets of similarity measures that induce the same order within the results of a query-by-example process.

4 Application to the Equivalence of Tversky's Similarity Measures

As an application of our theory, we propose to study the form of equivalence classes within a particular set of measures: the similarity measures of Tversky's ratio model. This model proposed in [14] gathers different behaviors by means of two weights balancing the influence of the two sets of distinctive features in the similarity measure. In this section, we give, for this family of measures, a complete caracterisation of the equivalence classes defined in the previous section.

4.1 Tversky's Ratio Model Measures and Their Behaviours

As introduced in section 2.1, Tversky proposed a general expression for the computation of the similarity, the ratio model:

$$S_{(\alpha,\beta)}(X,Y,Z) = \frac{X}{X+\alpha Y+\beta Z}$$

This formulation gives two free parameters (α, β). The choice of a given couple of parameters (α, β) leads to a particular measure behaviour, for example:

- if $\alpha = \beta$, the measure is symmetrical. Actually, it is a ressemblance measure (see [2]). Jaccard's measure (histogram intersection) and Dice's measure are two examples of Tversky's ratio model measures ($S_{(1,1)} = S_{jaccard}$, $S_{(\frac{1}{2},\frac{1}{2})} = S_{dice}$).
- for any α, if $\beta = 0$, the measure is called an inclusion measure (see [2]), and it evaluates the degree of inclusion of A in B.
- for any β, if $\alpha = 0$, the measure is called a satisfiability measure (see [2]), and it evaluates the inclusion of B in A. This kind of measure is used for instance, in decision analysis to evaluate the satisfiability of the observation of a fact B for the premise A of a rule.

4.2 Balancing Parameter of Tversky's Ratio Measures

As we have shown in section 3, the equivalence of two measures can be determined by a study of their level sets. Let us consider two measures $S_{(\alpha,\beta)}$, $S_{(\alpha',\beta')}$, and two levels h, h'.

To prove the equivalence of $S_{(\alpha,\beta)}$ and $S_{(\alpha',\beta')}$, we have to show that, for any h', we can find an unique h such as $S^h_{\alpha,\beta} = S^{h'}_{\alpha',\beta'}$. We can show that it happens if and only if $\alpha.\beta' - \alpha'.\beta = 0$.

So, two Tversky's measures $S_{(\alpha,\beta)}$ and $S_{(\alpha',\beta')}$ are equivalent of and only if $\alpha.\beta' = \alpha'.\beta$. In other words, two Tversky's Ratio measures are equivalent if their parameters have the same ratio $\frac{\alpha}{\beta}$.

If we are interested only in the order induced by a similarity measure taken within Tversky's ratio model, then the choice of the parameters α and β can be reduced to the choice of a single parameter $k = \frac{\alpha}{\beta}$:

- $k = 0$ for an inclusion measures.
- $k = 1$ for a ressemblance measure.
- $k = +\inf$ for a satisfiability measure.

5 Discussion

This section studies the consequences of using measures taken in a same given equivalence class, in other words, the implications of order invariance in some applications.

5.1 Order Based Procedures Such as Recall and Precision

The first consequence lies in the document retrieval context, where resemblance measures are used to compare features extracted from the documents. As shown on figure 1 for Jaccard and Dice, the user will obtain exactly the same results for any of the measures belonging to a given equivalence class.

Another consequence for this field is more profund and concerns the evaluation of the retrieval. For two equivalent measures, any recall/precision comparison based on the lists of the entries retrieved will conclude to the exact same accuracy. In particular, if order is invariant between two lists of results, the question of comparing the numbers of pertinent documents in the first 10 results is void.

Furthermore, as we have pointed it out in section 2.3, if a best-matching procedure uses comparisons of pairs of objects (in our case, regions), the result will be depending only on the order of the resemblance values. If that order is conserved from a measure to another, the final matching will be identical.

5.2 Value-Based Procedures

As we have discovered in section 3, two equivalent measures S_a and S_b are linked by composition of a strictly increasing function f. This extends the result of our discussion to value-based procedures such as thresholding, aggregation, etc. For some purpose beyond the simple query-by-example scheme, we may take into account the value computed by S_a in a specific operation (averaging with some other result, thresholding). In this case, we can predict the outcome of this operation using S_b instead of S_a (every other thing being equal) by using the equivalence function f to predict the value of S_b.

As an example, if we filter our query results by some lower threshold (all results must have a similarity value above 0.5), we know that using S_dice (see section 2) rather than $S_jaccard$ (histogram intersection) will lead the system to obtain the same ordered results but to present more results because the value of S_dice depends on the value of $S_jaccard$ and is globally larger.

6 Conclusion

In this paper, we have described both theoretically and empirically existing families of similarity measures that are bound to obtain the same results in a query-by-example perspective. As discussed in the paper, the choice of a similarity measure in this perspective can be reduced to the choice of a family of them. As an example, we have reduced the choice of the parameters of the Tversky's Ratio measures to one unique parameter.

We also discussed the consequences regarding the order of the values issued from a comparison between pairs of objects used in some applications. We have shown that we can practically predict the behavior of a system using one or the other of two equivalent measures.

References

1. C.C. Aggarwal, A. Hinneburg, and D.A. Keim. On the surprising behavior of distanc metrics in high dimensional space. In *ICDT 2001 (LNCS 1973)*, pages 420–434, London, UK, 2001.
2. B. Bouchon-Meunier, M. Rifqi, and S. Bothorel. Towards general measures of comparison of objects. *Fuzzy Sets and Systems*, 84(2):143–153, 1996.
3. M.D. Flickner, H. Sawhney, W. Niblack, J. Ashley, Q. Huang, B. Dom, M. Gorkani, J. Hafner, D. Lee, D. Petkovic, D. Steele, and P. Yanker. Query by image and video content: The qbic system. *Computer*, 28(9):23–32, September 1995.
4. A. Hinneburg, C.C. Aggarwal, and D.A. Keim. What is the nearest neighbor in high dimensional spaces? In *26th International Conference on Very Large Data Bases, VLDB2000*, pages 506–516, Cairo, Egypt, September 10-14, 2000.
5. A.K. Jain and A. Vailaya. Image retrieval using color and shape. *Pattern Recognition*, 8(29):1233–1244, 1996.
6. J.F. Omhover and M. Detyniecki. Strict: an image retrieval platform for queries based on regional content. In *CIVR'2004 (LNCS 3115)*, pages 473–482, Dublin, Ireland, August, 2004.
7. J.F. Omhover, M. Detyniecki, and B. Bouchon-Meunier. A region-similarity-based image retrieval system. In *IPMU'2004*, pages 1461–1468, Perugia, Italy, July, 2004.
8. J. S. Payne, L. Hepplewhite, and T. J. Stonham. Perceptually based metrics for the evaluation of textural image retrieval methods. In *Int. Conf. on Multimedia Computing and Systems*, volume 2, pages 793–797, Italy, 1999.
9. M. Rifqi, V. Berger, and B. Bouchon-Meunier. Discrimination power of measures of comparison. *Fuzzy Sets and Systems*, 110(2):189–196, March 2000.
10. S. Santini and R. Jain. Similarity measures. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 21(9):871–883, September 1999.
11. M. Stricker and M. Orengo. Similarity of color images. In *Proc. of SPIE Storage and Retrieval for Image and Video Databases*, pages 381–392, San Diego, USA, 1995.
12. M.J. Swain and D.H. Ballard. Color indexing. *International Journal of Computer Vision*, 7(1):11–32, 1991.
13. Y. A. Tolias, S. M. Panas, and L. H. Tsoukalas. Generalized fuzzy indices for similarity matching. *Fuzzy Sets and Systems*, 120(2):255–270, 2001.
14. A. Tversky. Features of similarity. *Psychological Review*, 84:327–352, 1977.