On the Role of Context-Awareness in Binary Image Comparison

Sara Iglesias-Rey[1,2], Aitor Castillo-López[1], Carlos Lopez-Molina[1,2,3], Bernard De Baets[3]

[1]*Dept. Estadística, Informática y Matemáticas, Universidad Pública de Navarra, 31006 Pamplona, Spain*
[2]*NavarraBiomed, Complejo Hospitalario de Navarra, Universidad Pública de Navarra, 31008 Pamplona, Spain*
[3]*KERMIT, Dept. Data Analysis and Mathematical Modelling, Ghent University, 9000 Ghent, Belgium*
sara.iglesias@unavarra.es

Abstract

The quantification of image similarity has been a key topic in the computer vision literature for the past few years. Different mathematical theories have been used in the development of these measures, which we will refer to as comparison measures. An interesting aspect in the study of comparison measures is the natural requirement to replicate human behavior. In almost all cases, it is appropriate for a comparison measure to produce results that are consistent with how humans would perform that assessment. However, despite accepting this premise, most of the proposals in the literature ignore a fundamental characteristic of the way in which humans carry out this evaluation: the context of comparison. In this work we present a comparison measure for binary images that incorporates the context of comparison; more precisely, we introduce an approach for the generation of ultrametrics for the context-aware comparison of binary images.

1. Introduction

In the image processing literature there exist a significant number of measures for binary image comparison. This variety is mainly due to the numerous applications of comparison measures for binary images. For example, in many image processing tasks, intermediate or final results are represented using binary images, giving rise to the necessity to compare automatically generated solutions with the ground truth. Comparison measures between binary images would be, in this case, directly used for the evaluation of the results.

The mathematical inspirations to create comparison measures for binary images are diverse, and often driven by the type of elements in the images. The content in binary images may vary greatly, and normally depends on the type of imagery and task. For example, binary images representing meteorological features [1], fungal structures [2] or object silhouettes [3] take very different shapes; hence, comparison operators are often specific for *types* of images or features. In a global perspective, common trends can be recognized in the general foundations of such operators. A relevant portion of comparison operators in the literature are based on metrics, many of them extending the Hausdorff metric [4, 5]. Information theory has also been used to develop binary consensus operators, either based on pixel matching/counting [6] or area coincidence [7]. A perspective that has been rarely studied in the literature is network- or tree-based comparison [8], despite the interest it has sparked in mathematical psychology [9]. Some ideas can be found in the literature, e.g. introducing the concepts of confusion and consensus areas in the field of image segmentation [10], but it has been significantly less explored than metric-based or information-based alternatives. Beyond the wide variety of comparison measures, it is worth mentioning that most of these measures are presented in a parametric form, giving rise to an even wider range of instantiations, exhibiting significantly different behavior.

One of the most unexplored aspects in the development of comparison measures is the modeling of the context of comparison. Most of these measures take into account only the images to be compared, so as to obtain a quantitative evaluation of the (dis)similarity between them. Generally, this is the case of metrics, which define the distance between two elements in a generic universe, satisfying the corresponding metric axioms (*identity of the indiscernibles*, *symmetry* and *triangle inequality*).

Definition 1.1 *U being a generic universe, a function* $d : U \times U \to [0, \infty[$ *is called a metric in U if it satisfies the following properties:*

1. *Identity of the indiscernibles:* $d(x, y) = 0$ *if and only if* $x = y$.

2. *Symmetry:* $d(x, y) = d(y, x)$ *for any* $x, y \in U$.

3. *Triangle inequality:* $d(x, z) \leq d(x, y) + d(y, z)$ *for any* $x, y, z \in U$.

HℏCSS

Metrics are the most used functions for comparison, although there are many others. For example, if instead of the triangle inequality, the function d satisfies the *ultrametric inequality* or *strong triangle inequality*, $d(x, y) \leq \max(d(x, z), d(y, z))$ for any $x, y, z \in U$, the resulting function would be an ultrametric. Obviously, any ultrametric is a metric.

Note that a metric is an ultrametric [11] if and only if any three points in U can be relabeled as x, y, z such that

$$d(x, y) \leq d(x, z) = d(y, z).$$

The use of metrics for image comparison may seem like a natural and mathematically appropriate approach, but in fact it is different from how humans proceed. Humans implicitly incorporate the context in their comparisons, performing a multidimensional analysis of the characteristics of each object. This incorporation of the context may even lead to non-compliance with the axioms when comparing. The context of comparison can be explicit, but if no information is provided, humans will incorporate it implicitly.

The non-necessity of imposing the metric axioms in human similarity judgment, along with the implicit contextualization made by humans, have long been studied. In [12], Yearsley *et al.* studied the implications of the famous example presented by Tversky [13], the *Jamaica-Cuba-Russia* comparison. The example is based on the idea that, d being a measure of dissimilarity between countries, humans normally judge that $d(Jamaica, Russia) > d(Jamaica, Cuba) + d(Russia, Cuba)$. In other words, that the perceived dissimilarity between *Jamaica* and *Russia*, is greater than the sum of the pairwise dissimilarity between *Jamaica* and *Cuba*, and between *Russia* and *Cuba*, hence not fulfilling the triangle inequality. In this example, each country is considered a multidimensional object and, in the comparison of each pair of countries, humans are able to determine the dimension of comparison implicitly, that is, the context of comparison. Even when the context is not explicit, the human evaluator is able to implicitly determine it based on previous experiences. For example, the *Jamaica-Cuba* comparison is normally established in terms of geography, while *Russia-Cuba* is based on its politics; humans selectively alter the role and the setting of the context in the comparisons. This example is used to discredit the need to impose the triangle inequality in human reasoning, as well as to evince the multidimensional nature of the human interpretation. It also serves as an illustration of the interesting role of the notion of context in human comparisons [14].

In general, the comparison measures in the literature do not make use of the notion of context. Normally,

these measures only consider the two images to be compared, assuming a global context and thus providing an absolute quantification of (dis)similarity. While this may be advantageous in some tasks, the process moves away from the way in which humans compare. Nowadays there is a general trend to replicate how humans naturally perform image processing tasks, since the intermediate or final results are compared to hand-labelled ground truth images; this is the concept behind supervised learning [15], where the training sets are human-labelled data, with the workload and waste of time that entails. A perfect simulation of human behaviour would then be considered a perfect performance, avoiding the expenses of human labelling. The aim of this paper is to create a context-aware comparison measure for binary images. More precisely, we intend to design an ultrametric that quantifies the dissimilarity between any two images within the context of comparison. The quantified distance between two images not only depends on their coincidences and divergences, but also on the characteristics of the remaining images within the context. By including the context of comparison, our measure brings the process closer to how humans perform comparisons.

The rest of this article is organized as follows. Section 2 presents the ultrametrics in context-aware comparison measures. The methodology for the construction of the ultrametric for the comparison of binary images is presented in Section 3, and is evaluated in Section 4. Finally, Section 5 presents the conclusions of the article.

2. Ultrametrics for object comparison

This section presents the concept of ultrametrics. First, a taxonomy of comparison measures and their main characteristics is presented (Section 2.1). Next, we focus on a network-based approach for object comparison (Section 2.2) and finally we present a methodology for the ultrametric construction through tree-based comparison measures (Section 2.3).

2.1. Taxonomy of comparison measures

Comparison measures, whether metrics, ultrametrics or any other type of function, are crucial in most of the scientific areas. These measures are decisive in tasks such as quality assessment, classification or optimization. In Tversky's terms [16], comparison measures can be divided into two main classes: (*i*) the ones based on geometrical interpretations, called *spatial models*, and (*ii*) those based on graph theory, called *network models*. Spatial models represent each object as a point in a coordinate space, so the distance between

points represents the closeness between objects. Most of the metrics and measures of dissimilarity adhere to this strategy, like the different variants of the Hausdorff metric [17, 18]. As opposed to spatial models, network models generate a graph-like representation of the relations between the objects to be compared. Each object is defined as a node in a connected acyclic graph, generally a rooted and labeled binary tree, while edges or *branches* (and their weights) are used to represent dissimilarity (or distance) between each pair of objects. When using trees to generate hierarchical structures, the nodes of degree 1 are called *leaves* and the node of degree 2 *root*.

In Tversky's double taxonomy, network models are more suitable for modeling the notion of context than spatial models. Spatial models may potentially establish a coordinate space depending on some definition of the context (however this is almost absent in the image processing literature, except in specific image retrieval applications through distribution comparison [19]). Generally it is assumed that the context is sufficiently large and nonspecific, and containing all possible elements. Network models are easily adapted to the context when establishing the topology of the graph, since they consider all possible interrelationships between elements. Therefore, the set of objects to be compared affects the network topology and, thus, the quantification of the dissimilarity between any two nodes.

In this work we introduce a network-based ultrametric that allows for a simple but meaningful modeling of the comparison context. The generation of a graph allows to model the relationships between images of a set, building an ultrametric in the process. This alternative enables better modeling of the human behavior when comparing and evaluating dissimilarity, as it implicitly includes the context. In addition, by satisfying the properties of the ultrametric, we guarantee that our comparison measure is a metric and, therefore, it can be applied in different applications that require metrics. The main disadvantage of this alternative is that the network topology cannot establish distances from or towards elements that were not present when the structure was settled, so all the elements to be compared must be known in advance.

2.2. Network-based comparison measures

Network-based comparison measures are obtained by establishing a network topology on a set of elements. Although this network topology can be any type of graph (distances on cyclic road maps are an example of this), generally rooted trees are the preferred
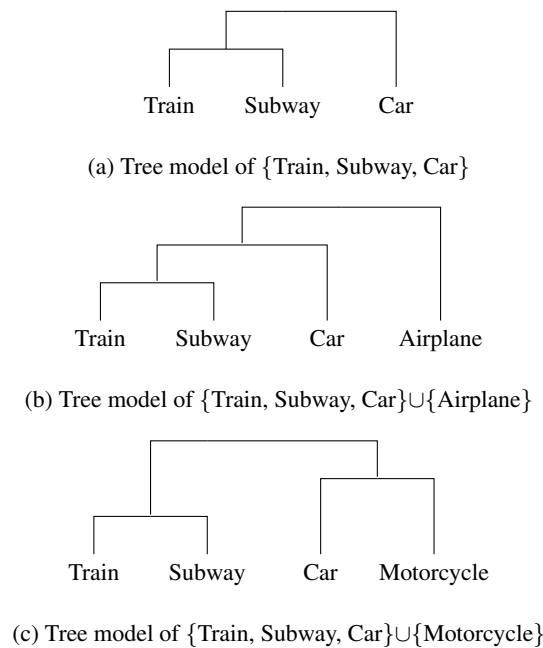
option, since they impose an appropriate hierarchical structure on the data. This choice is fundamental to many algorithms, especially in hierarchical clustering algorithms [20]. This work focuses on the use of trees as network topologies.

As an illustration, Figure 1 shows a tree organization of linguistic terms based on their dissimilarity. This graph contains, as leaves, the initial terms Train, Subway and Car, and the resulting groupings of the elements. In Fig. 1(a), {Train, Subway} are the first grouped, since they are the closest terms in the set. Then, {Train, Subway} are grouped with {Car}, forming the root of the tree. Numerical details are omitted, so we assume that the dissimilarity between two elements is proportional to the height of the node where they are grouped for the first time.

Figures 1(b) and 1(c) show the resulting trees when adding two different terms to the candidate set of Fig. 1(a): Airplane and Motorcycle. In the first case, by including an element perceptually far from all the elements of the original set, Airplane, the tree topology does not change significantly; the new element, Airplane, is grouped at the highest level of the tree with the other elements of the set. On the contrary, in Fig. 1(c), by including the element Motorcycle, which is perceived as close to one of the elements of the original set, Car, the tree changes significantly; the new element Motorcycle is grouped with Car at the second level, altering completely the structure of the tree. This represents the fact that humans understand that terms are similar in a variable way, depending on the comparison context.

2.3. Tree-based comparison measures

The three examples in Fig. 1 provide an idea of how to establish a hierarchy of the relations between terms according to their perceived dissimilarity, but they do not explain how to generate an ultrametric from them, or any other comparison measure. The construction process of the tree begins by considering all the objects to be compared and setting the first combined node as the *closest pair* of terms among all candidate pairs. This choice is not trivial; in computational terms, multiple different paradigms may be involved: minimum entropy of the resultant set, conditional probability, lower dissimilarity, etc. In the tree construction, it is necessary to iteratively decide which is the next node to be created, until the root is reached. Note that the strategy used to select the next combined node should not only consider the grouping of leaves, but it must also be prepared to consider the grouping of intermediate nodes that are generated in the

(a) Tree model of {Train, Subway, Car}

(b) Tree model of {Train, Subway, Car}∪{Airplane}

(c) Tree model of {Train, Subway, Car}∪{Motorcycle}

Figure 1. Variants of a tree model of the similarity between linguistic terms for three different sets of terms.

tree construction, *i.e.*, sets of elements.

Once the topology of the tree is established, it is important to set weights for the edges or branches. Depending on the nature of the strategy used to merge nodes, there are two different situations. If the strategy is qualitative, *i.e.*, it determines which is the next node that will be generated but does not return a numerical evaluation of its cost (dissimilarity), each new node is located at an increasing height of the tree. The cost of the first node created by combining the leaves will be 1, the second will have a cost of 2 and so on. If the strategy used for the node grouping is quantitative, each new node has a height equal to the cost of its generation. It is important to remark that this height is not an intrinsic property of the nodes, but it is represented by the length of the edges that connect each node to the leaves.

Quantitative strategies must ensure certain properties, for example, that nodes are created with increasing height, so that each new node cannot have a lower cost than older nodes. These strategies must also consider the situation in which different candidate nodes present the same cost during the construction process of the tree; there are multiple strategies to overcome this situation, like creating several nodes with the same cost at the same time or randomly selecting.

Whichever the strategy chosen to design the tree (qualitative or quantitative), the graph can be used to create an ultrametric. The distance between any two elements in the set of original elements,

in terms of the ultrametric, is the height of the lowest node that includes both of them. This agrees with the *strong triangle inequality* property in Def. 1.1 since the distance between two elements fulfills $d(x, y) \leq \max(d(x, z), d(y, z))$ for any $x, y, z \in U$. Also, this measure depends on the context of comparison, since different groups of elements will give rise to different tree topologies and, therefore, the ultrametric will yield different values.

3. Ultrametrics for binary image comparison

In this section we present the construction of an ultrametric for the comparison of binary images. First, we review the existing literature (Section 3.1) and second we introduce ultrametric trees for binary image comparison (Section 3.2).

3.1. Binary image comparison

The widespread use of binary images as intermediate (or final) results in image processing tasks led to the proliferation of a wide variety of measures for binary image comparison. Much of the literature is devoted to measures based on the confusion matrix [6, 21, 22], but there is also a significant number of measures that adopt a geometrical interpretation of the space of binary images [23]. The most popular options in this regard are the Hausdorff metric [4] and different generalizations

oriented to specific tasks [17, 18], or the Symmetric Difference [24, 25]. Other measures like Pratt's FoM (PFoM) [26] or Haralick's Measure [27], even if not satisfying the metric axioms, are partially based on metrics.

Despite satisfying certain properties, metric axioms or not, these measures do not provide an easy and intuitive way to incorporate the context of comparison. As far as we know, there are no references in the image processing literature that present context-aware comparison measures, where each comparison is made taking into account, not only the given pair of images to be compared, but all binary images within a set. Generally, when used, the term context refers to the idea that different regions of the image present local environments and discriminatory global standards, providing an adaptive comparison measure for each image; therefore, the concept of context refers to the relationships of a pixel with its neighborhood. Examples can be found in a wide variety of applications such as shape matching [28], edge detection [29], or tracking [30].

The notion of context-awareness as presented in this work is only present in certain specific tasks like classification [31], concept comparison [32] or recommender systems [33]. Most of these measures emerge as a fundamental notion of cognitive theories in psychology [34]. Despite considering the notion of context as presented in this work, these measures are not suitable for image comparison. There are no examples in the image processing literature that attempt to include this notion of context when comparing binary images.

Our proposal is a context-aware comparison measure for binary images, applying trees to obtain a quantification of dissimilarity by means of an ultrametric.

3.2. Tree-based comparison for binary images

The construction of the tree-based ultrametric presented in Section 2 can be extended to the case of binary images. In order to achieve it, a mathematical tool that allows for node grouping between objects is required, that is, a quantitative method to carry out comparisons between images. Due to the hierarchical nature of the trees, this tool will yield the order in which different images will be grouped in nodes. In our proposal, the comparisons between images are performed using Baddeley's Delta Metric (BDM) [35, 36], a very popular comparison measure derived from the Hausdorff metric [4], and the Symmetric Difference (SD_k) [24], an error measure for binary images based on distances.

In this work we consider images to have fixed dimensions $\mathcal{M} \times \mathcal{N}$, such that $\Omega = \{1, \dots, \mathcal{M}\} \times \{1, \dots, \mathcal{N}\}$ represents the set of positions in an image. The set of all possible binary images is denoted as $\mathbb{B}$, *i.e.*, the set of mappings $\Omega \mapsto \{0, 1\}$.

$A, B \in \mathbb{B}$ being two binary images in Ω and m a metric in Ω. The distance between them in terms of BDM is given by

$$\Delta^k(A, B) = \left[\frac{1}{|\Omega|} \sum_{p \in \Omega} |w(\mathcal{T}_m[A](p)) - w(\mathcal{T}_m[B](p))|^k \right]^{\frac{1}{k}},$$

(1)

where $w : \mathbb{R}^+ \mapsto \mathbb{R}^+$ is a concave function with $w(x) = 0$ if and only if $x = 0$, $k \in \mathbb{R}^+$ and $\mathcal{T}_m$ is an image distance transformation defined by

$$\mathcal{T}_m[I](p) = \min_{p' \in I} m(p, p'),$$

(2)

for all $p \in \Omega$.

For the same two binary images $A, B \in \mathbb{B}$ in Ω and m a metric in Ω, the distance between them, in terms of SD_k is given by

$$SD_k(A, B) = \frac{(\sum_{p \in B} \mathcal{T}_m^k[A](p) + \sum_{p \in A} \mathcal{T}_m^k[B](p))^{1/k}}{(|A \cup B|)^{1/k}},$$

(3)

where $k \in \mathbb{R}^+$ and $\mathcal{T}_m$ is an image distance transformation defined by Eq. 3.2.

These measures provide a quantification of the distance between images in the set and also a valid strategy for designing hierarchical tree-based dissimilarity measures, obtaining a proportionality relation between the distances between elements and the topology of the tree. This strategy also allows for the creation of an ultrametric, since the distance between two images in the set is the height of the lowest node that includes both of them. As a result, a graphical and intuitive representation of the proximity of the elements to be compared from the original set is obtained. The addition of new images to the original set will give rise to different quantifications of dissimilarity and, consequently, to different configurations of the ultrametric tree, due to its context-awareness.

The construction of an ultrametric tree requires the comparison not only of images, but also of sets of images in the top nodes. However, very few measures in the literature allow to perform N-to-M comparisons of binary images; 1-to-1 comparison measures can be used for the generation of N-to-M comparison measures. From BDM and SD_k we can obtain comparison measures between sets of binary images

A,B such that

$$C(\mathbf{A}, \mathbf{B}) = \bigwedge_{i,j} \Delta^k(A_i, B_j) , \qquad (4)$$

and

$$D(\mathbf{A}, \mathbf{B}) = \bigwedge_{i,j} \mathrm{SD}_k(A_i, B_j) , \qquad (5)$$

respectively.

4. Experiments

In this section we conduct the experiments aimed at testing the performance of our proposed comparison measure. In this regard, we first present the binary image dataset employed for the experiments (Section 4.1) and second, we present a separability analysis based on quantitative data (Section 4.2).

4.1. BSDS500 dataset

The Berkeley Segmentation Data Set and Benchmark 500 (BSDS500) [37] is a very popular dataset for edge detection and image segmentation applications. The dataset contains a large set of images, each one associated with a collection of hand-labelled ground truth images of edges. Figure 2 shows ground truth images associated with two different images of the BSDS500 dataset, *i.e.*, belonging to different classes.

Despite the fact that humans incur into great variability when labelling images, generally any human can group labelled images that come from the same original image (intra-class), and differentiate them from those that do not (inter-class). Any comparison measure should be able to replicate this human behavior, producing lower values for the comparison of intra-class images than for inter-class images. In Section 4.2 we will evaluate the ability to discriminate between pairs of intra-class and inter-class images of four different comparisons measures based on BDM and SD_k.

4.2. Separability analysis

The evaluation of the performance of comparison measures is not an easy task. In this work we analyze whether different comparison measures for binary images are able to replicate the human ability to discriminate when two boundary images in the BSDS500 are generated from the same original image or not; quantitatively, whether the comparison measure yields greater dissimilarity values for inter-class comparisons than for intra-class comparisons. A pair of images is considered intra-class if they are generated from the same original image (by different humans),
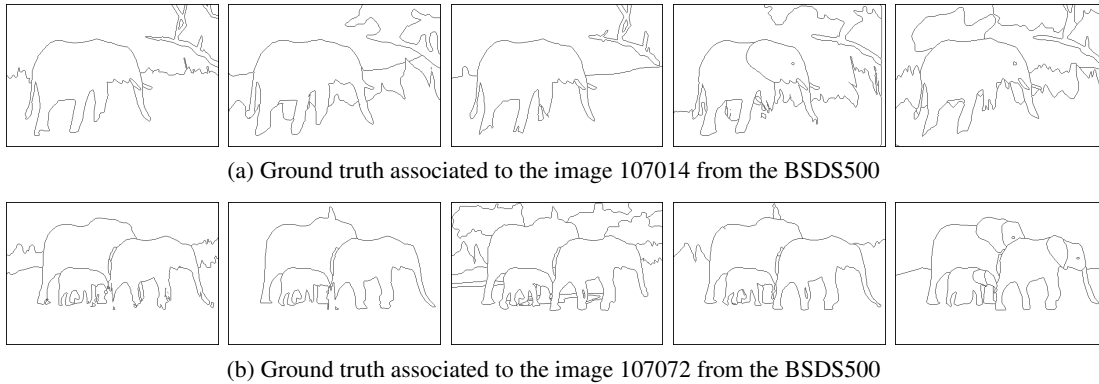
and inter-class if they are not. Since this is a task that humans can perform with little effort, it can be used as a simple measure of human behavior replication.

In this experiment, we compared all pairs of images using different configurations of BDM and SD_k, and ultrametric trees based on them. For these experiments for BDM, we set $w(x) = x$, $k = 2$, and m the bounded Euclidean distance $m_t(p, p') = \min(t, m(p, p'))$, $p, p' \in \Omega$, with $t = \{2.5, 5, 10\}$; we refer to the direct measures and those generated with ultrametric trees as Δ_t^2 and UMT-Δ_t^2, respectively. As for SD_k, we set $k = 2$, and m the bounded Euclidean metric $m_t(p, p') = \min(t, m(p, p'))$ for any $p, p' \in \Omega$, with $t = \{2.5, 5, 10\}$; we refer to the direct measures and those generated with ultrametric trees as $\mathrm{SD}_{2,t}$ and UMT-$\mathrm{SD}_{2,t}$, respectively. The BSDS500 Test dataset contains 200 different classes, giving rise to 1063 images in total. This implies that the number of inter-class comparisons is several orders of magnitude greater than the number of intra-class comparisons. We also calculated the accuracy (Acc) of discrimination of the distributions for each possible threshold. Ideally, Acc = 1 for at least one threshold, if the two distributions do not overlap, resulting in total separability.

In Figures 3 and 4 we present the distributions of the intra-class and inter-class comparisons for the direct comparison measures Δ_t^2 and $\mathrm{SD}_{2,t}$ (upper rows) and the ultrametric-tree-based comparison measures UMT-Δ_t^2 and UMT-$\mathrm{SD}_{2,t}$ (bottom rows). In addition, the accuracy is presented for each given threshold in both figures. Note that if the distributions were non-overlapping, the accuracy would reach the value 1. The results are replicated with $t \in \{2.5, 5, 10\}$. It is relevant to point out that, due to the great difference between the number of intra-class and inter-class comparisons, the distributions are presented in percentage terms. Note also that the distributions are shown on the left axis, while Acc is measured on the right axis, taking values between 0.5 and 1.

In general, the distributions of the intra-class and inter-class comparisons in Figs. 3 and 4 are quite separable for the four comparison measures, both the direct and the ultrametric tree-based measures. In general, the inter-class comparisons generate higher values than intra-class comparisons for all configurations, correctly simulating the human behavior. However, it can be seen at first glance that depending of the comparison measure, the intra-class and inter-class distributions are visually less overlapping, *i.e.*, more distant than others.

When comparing the standard (direct) and tree-based comparison measures, very interesting facts can be

(a) Ground truth associated to the image 107014 from the BSDS500

(b) Ground truth associated to the image 107072 from the BSDS500

Figure 2. Binary ground truth images from the BSDS500 dataset. Images that belong to the same class present a high variability due to different humans labelling.

pointed out. We can observe that both UMT-Δ_t^2 and UMT-SD$_{2,t}$ present a notable improvement compared to Δ_t^2 and SD$_{2,t}$ for all values of t. Direct comparison measures show a lower separability between the intra-class and inter-class distributions; on the contrary, the tree-based configurations not only show higher peaks of Acc, but also a larger area under the curve. Therefore, we can affirm that the use of trees not only leads to a qualitative improvement through the inclusion of the context of comparison, but it also leads to a quantitative improvement of the separability.

When comparing the results in Figs. 3 and 4, we can state that both SD$_{2,t}$ and UMT-SD$_{2,t}$ lead to higher separability values than Δ_t^2 and UMT-Δ_t^2, even if the two direct measures present a notable improvement when incorporating the ultrametric trees. Despite the difference between the peaks of the distributions, we can still affirm that including the use of trees provides better separability values independently of the measure, in addition to the incorporation of the context of comparison.

5. Conclusions

In this work we present a context-aware comparison measure based on ultrametric trees for binary images. Specifically we intend to create a comparison measure that takes into account all binary images of a set, replicating how humans naturally perform comparisons. In order to achieve it, we propose a tree-based comparison measure that allows for an easy adaptation to the context through the setting of the topology of the graph; our method also leads to the creation of an ultrametric, since the distance between any two element in the set of original elements is the height of the lowest node that includes both of them, fulfilling the *strong triangle inequality*.

To obtain a quantitative comparison between binary images, we apply BDM [35, 36], a comparison measure derived from the Hausdorff metric [4], and the Symmetric Difference (SD$_k$) [24], a distance-based error measure for binary images. We apply our algorithm to the BSDS500 dataset and we perform a separability analysis, obtaining the accuracy and the corresponding distributions.

As a conclusion, we can affirm that the construction of an ultrametric tree applying simple mathematical notions, allows to model the context when comparing binary images. However, this construction must be supported by a comparison measure that provides a dissimilarity quantification for the selection of the nodes. Using BDM and SD$_k$ provides high separability values, so we can affirm that our algorithm not only provides context-aware comparison measure, but also a better simulation of human behavior.

In future work, we shall explore the performance of ultrametric trees based on comparison measures that allow to perform N-to-M comparisons of binary images naturally, and not through the adaptation of 1-to-1 comparison measures. Due to the almost absence of comparison measures that allow to perform comparisons involving different cardinalities of images, the incorporation of these measures could lead to an interesting new analysis of the separability when including the context of comparison.
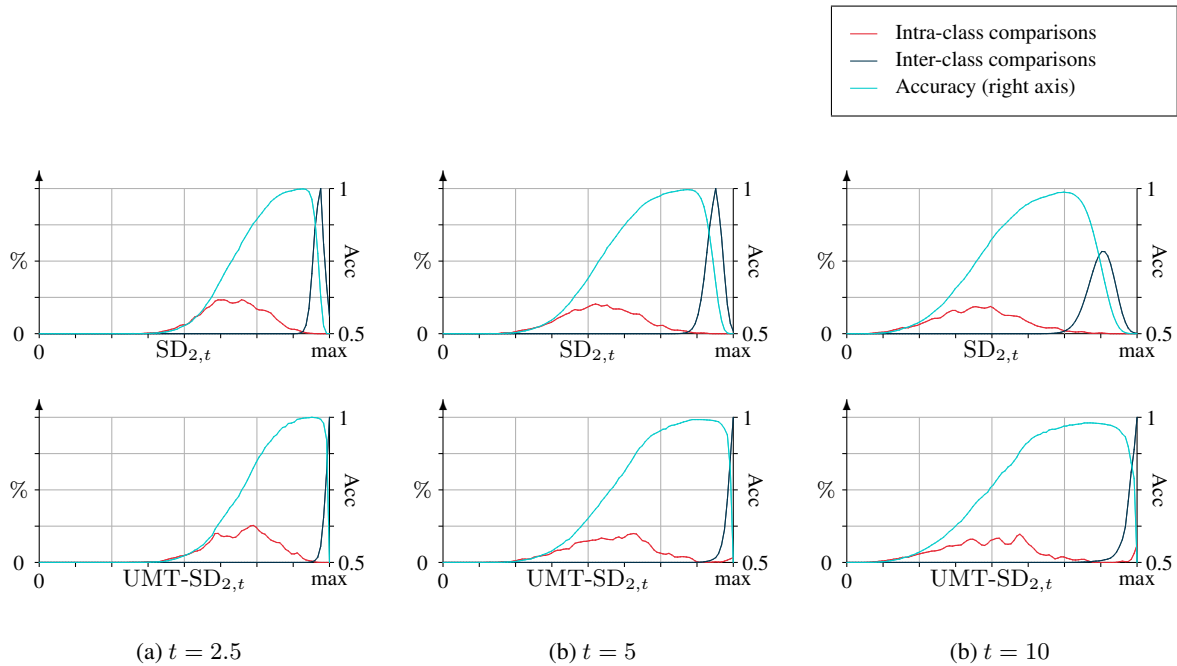
Acknowledgments
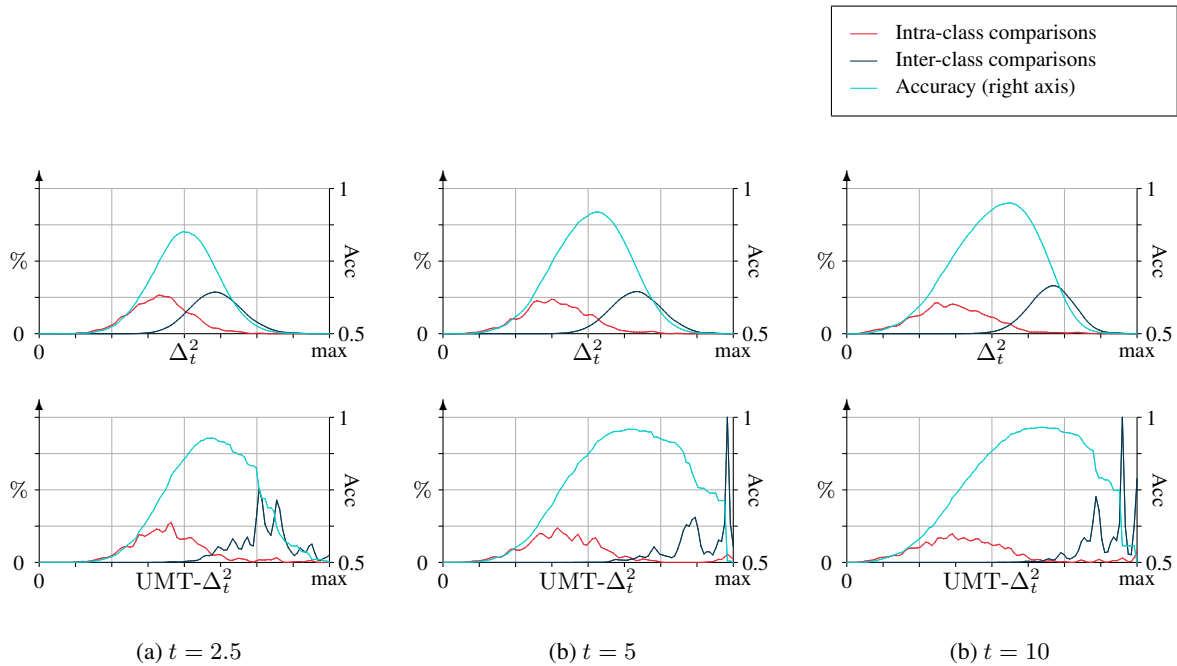
Figure 3. Intra- and inter-class distributions of the values obtained by Δ_t^2 and $\mathrm{UMT}\text{-}\Delta_t^2$, for the images in the BSDS500 Test set [37]. The accuracy of the separability of the distributions is also included for each threshold in the interval. Distributions are configured with 100 bins.

(a) $t = 2.5$ (b) $t = 5$ (b) $t = 10$

References

[1] D. Brunet and D. Sills, "A generalized distance transform: Theory and applications to weather analysis and forecasting," *IEEE Trans. on Geoscience and Remote Sensing*, vol. 55, no. 3, pp. 1752–1764, 2016.

[2] G. Vidal-Diez de Ulzurrun, T.-Y. Huang, C.-W. Chang, H.-C. Lin, and Y.-P. Hsueh, "Fungal feature tracker (fft): A tool for quantitatively characterizing the morphology and growth of filamentous fungi," *PLoS Computational Biology*, vol. 15, no. 10, p. e1007428, 2019.

[3] P. Arbelaez, M. Maire, C. Fowlkes, and J. Malik, "Contour detection and hierarchical image segmentation," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 33, pp. 898–916, 2011.

[4] D.-G. Sim, O.-K. Kwon, and R.-H. Park, "Object matching algorithms using robust Hausdorff distance measures," *IEEE Trans. on Image Processing*, vol. 8, no. 3, pp. 425–429, 1999.

[5] J. Paumard, "Robust comparison of binary images," *Pattern Recognition Letters*, vol. 18, no. 10, pp. 1057–1063, 1997.

[6] F. J. Estrada and A. D. Jepson, "Benchmarking image segmentation algorithms," *International Journal of Computer Vision*, vol. 85, no. 2, pp. 167–181, 2009.

[7] C. Lopez-Molina, B. De Baets, H. Bustince, J. Sanz, and E. Barrenechea, "Multiscale edge detection based on Gaussian smoothing and edge tracking," *Knowledge-Based Systems*, vol. 44, pp. 101–111, 2013.

[8] B. Zieliński and M. Iwanowski, "Binary image comparison with use of tree-based approach," in *Image Processing and Communications Challenges 4* (R. S. Choraś, ed.), (Berlin, Heidelberg), pp. 171–177, Springer Berlin Heidelberg, 2013.

[9] A. Tversky and D. H. Krantz, "The dimensional representation and the metric structure of similarity data," *Journal of mathematical psychology*, vol. 7, no. 3, pp. 572–596, 1970.

[10] A. Asman and B. Landman, "Robust statistical label fusion through consensus level, labeler accuracy and truth estimation (COLLATE)," *IEEE Trans. on Medical Imaging*, vol. 30, pp. 1779–94, 2011.

[11] T. Faver, K. Kochalski, M. K. Murugan, H. Verheggen, E. Wesson, and A. Weston, "Roundness properties of ultrametric spaces," *Glasgow Mathematical Journal*, vol. 56, no. 3, pp. 519–535, 2014.

[12] J. M. Yearsley, A. Barque-Duran, E. Scerrati, J. A. Hampton, and E. M. Pothos, "The triangle inequality constraint in similarity judgments," *Progress in Biophysics and Molecular Biology*, vol. 130, pp. 26–32, 2017.

[13] A. Tversky, "Features of similarity," *Psychological Review*, vol. 84, no. 4, pp. 327–352, 1977.

[14] J. Fumanal-Idocin, A. Alonso-Betanzos, O. Cordón, H. Bustince, and M. Minárová, "Community detection and social network analysis based on the Italian Wars of the 15th century," *Future Generation Computer Systems*, vol. 113, pp. 25–40, 2020.

[15] C. M. Bishop, *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Berlin, Heidelberg: Springer-Verlag, 2006.

[16] A. Tversky, *Preference, belief, and similarity : selected writings by Amos Tversky; edited by Eldar Shafir*. Cambridge, Mass: MIT Press, 2004.

(a) $t = 2.5$ (b) $t = 5$ (b) $t = 10$

Figure 4. Intra- and inter-class distributions of the values obtained by $\mathrm{SD}_{2,t}$ and $\mathrm{UMT\text{-}SD}_{2,t}$, $t \in \{2.5, 5, 1\}$, for the images in the BSDS500 Test set [37]. The accuracy of the separability of the distributions is also included for each threshold in the interval. Distributions are configured with 100 bins.

[17] M. Dubuisson and A. K. Jain, "A modified Hausdorff distance for object matching," in *Proc. on Pattern Recognition*, vol. 1, pp. 566–568, 1994.

[18] B. Takács, "Comparing face images using the modified Hausdorff distance," *Pattern Recognition*, vol. 31, no. 12, pp. 1873–1881, 1998.

[19] F. Perronnin, Y. Liu, and J.-M. Renders, "A family of contextual measures of similarity between distributions with application to image retrieval," in *2009 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2358–2365, 2009.

[20] Y. Zhao, G. Karypis, and U. Fayyad, "Hierarchical clustering algorithms for document datasets," *Data Mining and Knowledge Discovery*, vol. 10, pp. 141–168, 2005.

[21] K. Bowyer, C. Kranenburg, and S. Dougherty, "Edge detector evaluation using empirical ROC curves," *Computer Vision and Image Understanding*, vol. 84, no. 1, pp. 77–103, 2001.

[22] B. Magnier and H. Abdulrahman, "A study of measures for contour-based recognition and localization of known objects in digital images," in *Proc. of the IEEE International Conf. on Image Processing Theory, Tools and Applications (IPTA 2018)*, 2018.

[23] J. Gimenez, J. Martinez, and A. G. Flesia, "Unsupervised edge map scoring: A statistical complexity approach," *Computer Vision and Image Understanding*, vol. 122, pp. 131–142, 2014.

[24] C. Lopez-Molina, B. De Baets, and H. Bustince, "Quantitative error measures for edge detection," *Pattern Recognition*, vol. 46, no. 4, pp. 1125–1139, 2013.

[25] C. Lopez-Molina, H. Bustince, and B. De Baets, "Separability criteria for the evaluation of boundary detection benchmarks," *IEEE Trans. on Image Processing*, vol. 25, no. 3, pp. 1047–1055, 2016.

[26] I. Abdou and W. Pratt, "Quantitative design and evaluation of enhancement/thresholding edge detectors," in *Proc. of the IEEE*, vol. 67, pp. 753–763, 1979.

[27] R. M. Haralick, "Digital step edges from zero crossing of second directional derivatives," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 6, no. 1, pp. 58–68, 1984.

[28] S. Belongie, J. Malik, and J. Puzicha, "Shape matching and object recognition using shape contexts," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 24, no. 4, pp. 509–522, 2002.

[29] R. M. Haralick and J. S. J. Lee, "Context dependent edge detection and evaluation," *Pattern Recognition*, vol. 23, no. 1-2, pp. 1–19, 1990.

[30] M. Mueller, N. Smith, and B. Ghanem, "Context-aware correlation filter tracking," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1387–1395, 2017.

[31] H. Cao, D. H. Hu, D. Shen, D. Jiang, J.-T. Sun, E. Chen, and Q. Yang, "Context-aware query classification," in *Proceedings of the 32nd International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '09, (New York, NY, USA), p. 3–10, Association for Computing Machinery, 2009.

[32] S. Bordag, "A comparison of co-occurrence and similarity measures as simulations of context," in *Computational Linguistics and Intelligent Text*

Processing (A. Gelbukh, ed.), (Berlin, Heidelberg), pp. 52–63, Springer Berlin Heidelberg, 2008.

[33] U. Panniello and M. Gorgoglione, "Context-aware recommender systems: A comparison of three approaches," vol. 771, 09 2011.

[34] M. Bazire and P. Brézillon, "Understanding context before using it," in *Modeling and Using Context* (A. Dey, B. Kokinov, D. Leake, and R. Turner, eds.), (Berlin, Heidelberg), pp. 29–40, Springer Berlin Heidelberg, 2005.

[35] A. J. Baddeley, "An error metric for binary images," in *Robust Computer Vision: Quality of Vision Algorithms* (W. Förstner and S. Ruwiedel, eds.), (Karlsruhe), pp. 59–78, Wichmann Verlag, 1992.

[36] A. J. Baddeley, "Errors in binary images and an L^p version of the Hausdorff metric," *Nieuw Archief voor Wiskunde*, vol. 10, pp. 157–183, 1992.

[37] D. Martin, C. Fowlkes, D. Tal, and J. Malik, "A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics," in *Proc. of the IEEE International Conf. on Computer Vision*, vol. 2, pp. 416–423, 2001.