

Partial similarity of objects and text sequences

Alexander M. Bronstein*, Michael M. Bronstein*, Alfred M. Bruckstein* and Ron Kimmel*

*Department of Computer Science

Technion – Israel Institute of Technology, Haifa 32000, Israel

Emails: {bron, mbron, freddy, ron }@cs.technion.ac.il

Abstract—Similarity is one of the most important abstract concepts in the human perception of the world. In computer vision, numerous applications deal with comparing objects observed in a scene with some a priori known patterns. Often, it happens that while two objects are not similar, they have large similar parts, that is, they are partially similar. Here, we present a novel approach to quantify this semantic definition of partial similarity using the notion of Pareto optimality. We exemplify our approach on the problems of recognizing non-rigid objects and analyzing text sequences.

I. INTRODUCTION

Similarity is one of the most important abstract concepts in the human perception of the world. We encounter it every day during our interaction with other people whose faces we recognize. Similarity also plays a crucial role in many fields in science. Attempts to understand self-similar or symmetric behavior of Nature led to many fundamental discoveries in physics [1]. In bioinformatics, a fundamental problem is detecting patterns similar to a sequence of nucleotides in given DNA sequences. In computer vision, comparing objects observed in a scene with some *a priori* known patterns is a fundamental and largely open problem.

The definition of similarity is, to a large extent, a semantic question. Judging the similarity of faces, one may say that two human faces are similar if they have a common skin tone, while someone else would require the identity of the geometric structure of facial features like the eyes, the nose, and the mouth.

With slight exaggeration, we can say that all pattern recognition problems boil down to giving quantitative interpretation of similarity (or equivalently, *dissimilarity*) between objects. Since there is no unique definition of similarity, every class of objects and every problem require a specific, problem-dependent similarity criterion. Such criteria have been proposed for images [2], [3], [4], [5], two-dimensional shapes [6], [7], [8], [9], [10], [11], [12], [13], [9], [14], three-dimensional objects [15], [16], [17], [18], [19], text [20], [21], and audio [22], [23]. In the face recognition community, extensive research has been done on similarities insensitive to illumination [24], [25], head pose [26], and facial expressions [27], [28].

In many situation, it happens that, while two objects are not similar, some of their parts are [29], [30], [31], [32]. Such a situation is common, for example, in the face recognition application, where the quality facial images (or surfaces in case of 3D face recognition) can be degraded by acquisition

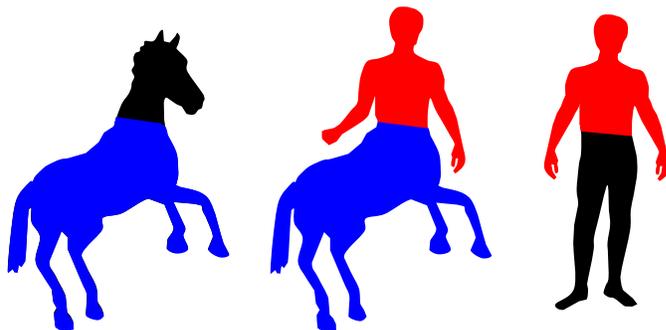


Fig. 1. Is a centaur similar to a horse? A large part of the centaur (blue) is similar to a horse; likely, a large part of the centaur (red) is similar to a human. However, considered as a whole, the centaur is similar neither to a horse, nor to a human.

imperfections, occlusions, and the presence of facial hair [33]. As an illustration that will help to understand the problem of partial similarity, we give an example from the realm of shape comparison. Figure 1 (inspired by Jacobs *et al.* [11]; see also [30], [31]) shows a centaur – a half-horse half-human mythological creature. From the point of view of two-dimensional shape similarity, a centaur is similar neither to a horse nor to a man. However, large part of these objects (the upper part of the human body, marked in red and the bottom part of the horse body, marked in blue) are similar. Semantically, we can say that two object are *partially similar* if they have *large similar parts*. If one is able to detect such parts, the degree of partial similarity can be evaluated [31].

The main purpose of this paper, stated briefly, is to give a quantitative interpretation to what is meant by “similar” and “large”, and derive a consistent relation between these terms, which allows formulate a computationally-tractable problem of finding the largest most similar parts. While presenting a generic framework that may be applied to numerous pattern recognition problems, we try to provide specific examples proving the usefulness of our approach.

The rest of this paper is organized as follows. In Section II, we give formal axiomatic definitions of partiality and similarity and the necessary mathematical background. In Section III, we formulate a multicriterion optimization problem, from which the relation between partiality and similarity is derived. We represent partial similarity as a set-valued distance and study its properties. Sections IV and V are case studies of applications of our partial similarity framework. We study extensively the application of analysis of two- and three-

dimensional non-rigid objects, presenting a practical numerical scheme of its computation. In Section V, we show how the partial similarity approach generalizes classical results in text sequences analysis. All the application-specific mathematical background is defined in the beginning of each section. Section VI concludes the paper.

II. BASIC DEFINITIONS

In order to give a quantitative interpretation to our semantic definition of partial similarity, we first have to define the terms “part”, “large” and “similar”. We start our construction by defining the class \mathcal{C} of objects we wish to compare: these may be, for instance, shapes, pictures, three-dimensional surfaces, or words. An *object* in \mathcal{C} is a set, denoted by X . Given a σ -algebra Σ_X on X (a subset of the powerset 2^X closed under compliment and countable union), we refer to any $X' \in \Sigma_X$ as a *part* of X . The class of all parts of all the objects is denoted as $\mathcal{P}_C = \bigcup_{X \in \mathcal{C}} \Sigma_X$; by properties of σ -algebras, \mathcal{C} itself is also in \mathcal{P}_C .

To judge how similar two objects or their parts are, we define a non-negative function $\epsilon : \mathcal{P}_C \times \mathcal{P}_C \rightarrow \mathbb{R}_+$, obeying the following properties,

- (D1) *Self-similarity*: $\epsilon(X, X) = 0$;
- (D2) *Symmetry*: $\epsilon(X, Y) = \epsilon(Y, X)$;
- (D3) *Triangle inequality*: $\epsilon(X, Y) + \epsilon(Y, Z) \geq \epsilon(X, Z)$;

for all objects X, Y and Z in \mathcal{C} . We call ϵ a *dissimilarity*, since the greater it is, the less similar are the objects. Property (D1) simply states that an object is similar to itself; property (D2) requires similarity to be reflexive; and (D3) expresses the transitivity of similarity: if X is similar to Y , which is in turn similar to Z , then X and Z cannot be dissimilar. Technically speaking, ϵ is a *pseudo-metric* on \mathcal{C} , and a *metric* on the quotient space of \mathcal{C} under the similarity relation. We will encounter some examples of dissimilarities in Sections IV and V.

In order to quantify the *size* of a part of an object X , we define a function $\mu_X : \Sigma_X \rightarrow \mathbb{R}_+$, satisfying the following properties,

- (M1) *Additivity*: $\mu_X(X' \cup X'') = \mu_X(X') + \mu_X(X'')$ for two disjoint parts $X' \cap X'' = \emptyset$.
- (M2) *Monotonicity*: $\mu_X(X'') \leq \mu_X(X')$ for all $X'' \subseteq X' \in \Sigma_X$.

We call μ_X a *measure*; an intuitive example of a measure is the area of a geometric object. Using the measure theory jargon, a property is said to hold *almost everywhere (a.e.)* on X if it holds on entire X , excepting a part with zero measure. A real function $f : X \rightarrow \mathbb{R}$ is called Σ_X -*measurable* if $\{x \in X : f(x) \geq \alpha\} \in \Sigma_X$ for all α .

Let $f : \mathbb{R}^2 \rightarrow \mathbb{R}_+$ be a non-negative bivariate function, satisfying the following conditions,

- (F1) $f(0, 0) = 0$;
- (F2) *Monotonicity*: $f(a', b) \leq f(a, b)$ and $f(a, b') \leq f(a, b)$ for all $a' \leq a, b' \leq b$;
- (F3) *Symmetry*: $f(a, b) = f(b, a)$;

- (F4) *Homogeneity of order ρ* : $f(\alpha a, \alpha b) = |\alpha|^\rho f(a, b)$, where $\rho \geq 0$;

and let X and Y be objects in \mathcal{C} . We define a *partiality*,

$$\lambda(X', Y') = f(\mu_X(X'^c), \mu_Y(Y'^c)), \quad (1)$$

for all $X' \in \Sigma_X$ and $Y' \in \Sigma_Y$, where $X'^c = X \setminus X'$. Partiality quantifies how “small” are the parts X' and Y' ; the larger is the partiality, the smaller are the parts.

Proposition 1: λ satisfies the following properties,

- (P1) *Partiality of the whole*: $\lambda(X, Y) = 0$.
- (P2) *Symmetry*: $\lambda(X', Y') = \lambda(Y', X')$.
- (P3) *Partial order*: $\lambda(X'', Y'') \geq \lambda(X', Y')$ for every $X'' \subseteq X' \in \Sigma_X$ and $Y'' \subseteq Y' \in \Sigma_Y$.

The proof is straightforward by using properties (M1)–(M2) and (F1)–(F4). Hereinafter, we will encounter the following examples of partialities:

$$\begin{aligned} \lambda_{\text{SUM}}(X', Y') &= \mu_X(X'^c) + \mu_Y(Y'^c); \\ \lambda_{\text{MAX}}(X', Y') &= \max\{\mu_X(X'^c), \mu_Y(Y'^c)\}; \\ \lambda_{\text{NMAX}}(X', Y') &= \max\left\{\frac{\mu_X(X'^c)}{\mu_X(X)}, \frac{\mu_Y(Y'^c)}{\mu_Y(Y)}\right\}. \end{aligned}$$

All the above partialities are homogeneous of order $\rho = 1$.

A. Fuzzy formulation

The above definitions can be extended using the fuzzy set theory [34]. This formulation is useful in numerical computations, as will be shown in Section IV-E. As a notation convention, we will use tilde to denote all the fuzzy quantities.

We define a *fuzzy part* of X as a collection of pairs of the form $\{(x, m_X(x)) : x \in X\}$, where $m_X : X \rightarrow [0, 1]$ is referred to as a *membership function* and measures the degree of inclusion of a point into the subset. A fuzzy part is completely described by its membership function m_X ; hereinafter, we use m_X referring to fuzzy parts. The complement of the fuzzy part is defined as $m_X^c = 1 - m_X$. A subset $X' \subseteq X$ in the traditional set theoretic sense (called *crisp* in fuzzy set theory) can be described by a membership function $\mathbb{I}_{X'}(x)$, equal to one if $x \in X'$ and zero otherwise. More generally, a fuzzy part m_X can be converted into a crisp one by *thresholding*, $\tau_\delta \circ m_X$, where

$$\tau_\delta(x) = \begin{cases} 1 & x \geq \delta \\ 0 & \text{otherwise.} \end{cases} \quad (2)$$

and $0 \leq \delta \leq 1$ is some constant. The corresponding crisp set is denoted by $T_\delta m_X = \{x \in X : \tau_\delta \circ m_X = 1\}$.

Given a Σ_X , we define M_X as the set of all the fuzzy parts whose membership functions are Σ_X -measurable. It follows straightforwardly that $T_\delta m_X$ is in Σ_X for all $m_X \in M_X$ and $0 \leq \delta \leq 1$. As previously, we define $\tilde{\mathcal{P}}_C = \bigcup_{X \in \mathcal{C}} M_X$. Since crisp parts are a particular case of fuzzy parts, $\mathcal{P}_C \subseteq \tilde{\mathcal{P}}_C$.

We define a *fuzzy dissimilarity* as a function $\tilde{\epsilon} : \tilde{\mathcal{P}}_C \times \tilde{\mathcal{P}}_C \rightarrow \mathbb{R}_+$ satisfying properties (D1)–(D3) with crisp sets replaces by fuzzy ones. We require $\tilde{\epsilon}$ to coincide with ϵ on $\mathcal{P}_C \times \mathcal{P}_C$, or in other words, $\epsilon(X', Y') = \tilde{\epsilon}(\mathbb{I}_{X'}, \mathbb{I}_{Y'})$. The *fuzzy measure* is

defined as

$$\tilde{\mu}_X(m_X) = \int_X m_X(x) d\mu_X, \quad (3)$$

for all $m_X \in M_X$, where μ_X is a (crisp) measure on X . Given a function f obeying properties (F1)–(F4) and homogeneous of order ρ , we define the *fuzzy partiality* as

$$\tilde{\lambda}(m_X, m_Y) = f(\tilde{\mu}_X(m_X^c), \tilde{\mu}_Y(m_Y^c)), \quad (4)$$

similarly to definition (1). The following relation between the fuzzy and the crisp partialities holds,

$$\text{Proposition 2: (i) } \lambda(X', Y') = \tilde{\lambda}(\mathbb{I}_{X'}, \mathbb{I}_{Y'}); \quad \text{(ii) } \lambda(T_\delta m_X, T_\delta m_Y) \leq \left(\frac{1}{1-\delta}\right)^\rho \tilde{\lambda}(m_X, m_Y), \text{ for all } 0 < \delta < 1.$$

III. PARETIAN FORMULATION OF PARTIAL SIMILARITY

Using the definitions of Section II, we can now give a quantitative expression to our semantic definition of partial similarity: X and Y are *partially similar* if they have parts X' and Y' with small partiality $\lambda(X', Y')$ (“large”) and small dissimilarity $\epsilon(X', Y')$ (“similar”).

We therefore formulate the computation of partial similarity as a *multicriterion optimization* problem: minimization of the *vector objective function* $\Phi(X', Y') = (\epsilon(X', Y'), \lambda(X', Y'))$ with respect to the pair (X', Y') over the *feasible set* $\Omega = \Sigma_X \times \Sigma_Y$. The values of the criteria $\epsilon(X', Y')$ and $\lambda(X', Y')$ for every $(X', Y') \in \Omega$ can be associated with a point with the coordinates $\Phi(X', Y')$ in the *criteria space* \mathbb{R}^2 . The set of possible criteria values correspond to the region $\Lambda = \Phi(\Omega)$ in \mathbb{R}^2 , referred to as the *attainable set* (see Figure 2a). The point $(0, 0)$ is usually not achievable, unless X and Y are completely similar. For this reason, it is called the *utopia point*.

Since the two criteria are competing, no solution simultaneously optimal for both (i.e., the utopia point) can be found.¹ Thus, the notion of optimality used in traditional scalar optimization must be replaced by a new one, adapted to the multicriterion problem. Since there does not exist a *total order* relation in \mathbb{R}^2 , we generally cannot say which solution is better, for example: is the point $(0.5, 1)$ better than $(1, 0.5)$? Yet, we can introduce *partial order* by coordinate-wise comparison: $\Phi(X', Y')$ is better than $\Phi(X'', Y'')$ if both $\lambda(X', Y') \leq \lambda(X'', Y'')$ and $\epsilon(X', Y') \leq \epsilon(X'', Y'')$, e.g., the point $(0.5, 0.5)$ is better than $(1, 1)$. By writing $\Phi(X', Y') \leq \Phi(X'', Y'')$, this partial order relation is implied hereinafter.

A solution (X^*, Y^*) is called a *Pareto optimum* [35], [36], [37] of the multicriterion optimization problem, if at least one of the following holds,

$$\begin{aligned} \epsilon(X^*, Y^*) &\leq \epsilon(X', Y'); \quad \text{or,} \\ \lambda(X^*, Y^*) &\leq \lambda(X', Y'), \end{aligned} \quad (5)$$

for all $(X', Y') \in \Omega$. An intuitive explanation of Pareto optimality is that no criterion can be improved without compromising the other. The set of all the Pareto optima, referred

¹In information theory, such multicriterion optimization problems are widely known. For example, in statistical estimation, the bias and the variance of an estimator are two competing criteria. In lossy signal compression, distortion and bitrate are competing.

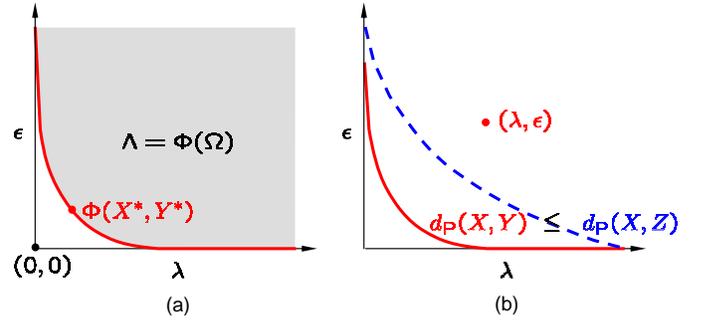


Fig. 2. (a) Illustration of the notion of Pareto optimality. (b) Illustration of order relations between partial dissimilarities. We can say that $d_P(X, Y) < d_P(X, Z)$ because the entire $d_P(X, Y)$ is below $d_P(X, Z)$. Since the point (λ, ϵ) is below $d_P(X, Y)$, the point-wise order relation $(\lambda, \epsilon) < d_P(X, Y)$ holds.

to as the *Pareto frontier* in economics, is our criterion of partial similarity. We denote the partial dissimilarity by $d_P(X, Y)$, thinking of it as a generalized, *set-valued distance*. By writing $(\lambda, \epsilon) \in d_P(X, Y)$, we mean that there exist two parts X' and Y' with partiality $\lambda(X', Y')$ and dissimilarity $\epsilon(X', Y')$, such that any two other parts with smaller partiality have larger dissimilarity, or two other parts with smaller dissimilarity have larger partiality. For brevity, we express the latter fact by saying that X and Y are (λ, ϵ) -dissimilar.

Since there exists only a partial order relation between our criteria, not all partial dissimilarities are mutually comparable. In this sense, the notion of partial similarity is notably different from the standard “full” similarity. We can say that X is *more similar to Y than to Z* (expressed as $d_P(X, Y) < d_P(X, Z)$) only if $d_P(X, Y)$ is entirely below $d_P(X, Z)$. Otherwise, only point-wise comparison is possible: we write $(\lambda_0, \epsilon_0) < d_P(X, Y)$, implying that the point (λ_0, ϵ_0) is below $d_P(X, Y)$ (see Figure 2b).

A. Scalar-valued partial dissimilarities

The set-valued distance d_P conceals much information about the partial similarity of objects, however, it is sometimes inconvenient to work with due to the lack of total order relation, and as the result, the inability to compare two such distances. In order to define a *total order* between partial dissimilarities, we have to convert the set-valued distance into a scalar-valued one. A naïve approach is selecting a point on the Pareto frontier with fixed value of partiality or dissimilarity. This allows us define the following scalar partial dissimilarities,

$$\begin{aligned} d_{\lambda_0\text{-PART}}(X, Y) &= \inf_{(\lambda_0, \epsilon) \in \Lambda} \epsilon \\ &= \inf_{(X', Y') \in \Omega} \epsilon(X', Y') \quad \text{s.t. } \lambda(X', Y') \leq \lambda_0, \end{aligned}$$

and

$$\begin{aligned} d_{\epsilon_0\text{-DIS}}(X, Y) &= \inf_{(\lambda, \epsilon_0) \in \Lambda} \lambda \\ &= \inf_{(X', Y') \in \Omega} \lambda(X', Y') \quad \text{s.t. } \epsilon(X', Y') \leq \epsilon_0. \end{aligned}$$

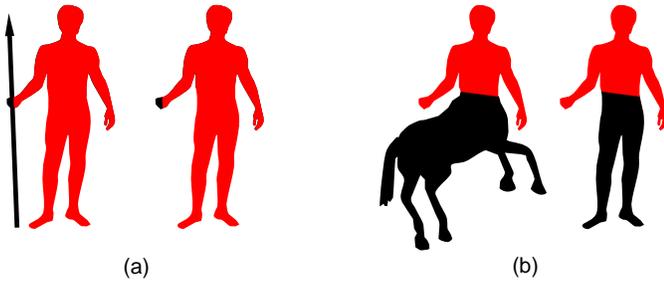


Fig. 3. Converting set-valued partial dissimilarity into a scalar-valued one: in order to make a spear-bearer similar to a man (a), we have to remove the spear (i.e., $d_{\epsilon_0\text{-DIS}}$ is small). In order to make a centaur similar to a man (b) we have to remove the horse body from the centaur and the legs from the man ($d_{\epsilon_0\text{-DIS}}$ is large).

$d_{\lambda_0\text{-PART}}(X, Y)$, to which we refer as the λ_0 -*partiality distance*, equals the minimum possible dissimilarity between parts which have partiality λ_0 . $d_{\epsilon_0\text{-DIS}}(X, Y)$, referred to as the ϵ_0 -*dissimilarity distance*, quantifies how large are the parts we have to remove from the objects X and Y in order to make the remaining parts ϵ_0 -dissimilar.

The disadvantage of $d_{\lambda_0\text{-PART}}$ and $d_{\epsilon_0\text{-DIS}}$ is that it is usually impossible to fix a threshold suitable for all the objects. Such a situation is illustrated in Figure 3: in order to make a spear-bearer similar to a man we have to remove only a small part (the spear), whereas in order to make a centaur similar to a man we have to remove large parts (the horse body from the centaur and the legs from the man).

To overcome this problem, we can select a point on $d_P(X, Y)$ which is the closest, in the sense of some distance, to the utopia point $(0, 0)$. A Pareto optimum corresponding to such a point is called *Salukwadze optimal* [38]. We define the *Salukwadze distance* between X and Y as

$$\begin{aligned} d_S(X, Y) &= \inf_{(\lambda, \epsilon) \in \Lambda} \|(\lambda, \epsilon)\| \\ &= \inf_{(X', Y') \in \Omega} \|\Phi(X', Y')\|. \end{aligned} \quad (6)$$

Here, $\|\cdot\|$ denotes some norm on \mathbb{R}_+^2 . One example is the family of weighted norms $\|\Phi\|_{\mathbf{w}} = \Phi^T \mathbf{w}$ ($\mathbf{w} \in \mathbb{R}_+^2$). The particular case $\|\cdot\|_{(1,1)}$ coincides with the L_1 -norm.

IV. CASE STUDY I – NON-RIGID OBJECTS

Our first case study deals with non-rigid two- and three-dimensional objects. Analysis of such objects is an important field emerging in the last decade in the pattern recognition community and arising in applications of face recognition [27], [28], matching of articulated objects [7], [11], [17], [39], [40], [18], [14], shape watermarking [41], texture mapping and morphing [42], [43], to mention a few. Often, natural deformations of non-rigid objects can be approximated as isometries, hence, recognition of such objects requires an isometry-invariant criterion of similarity. Moreover, in many situations (e.g. in face recognition [33]), due to acquisition imperfections, the objects are given only partially, i.e., have similar overlapping parts. This makes our partial similarity framework especially useful for such problems.

A. Background

A non-rigid object is modeled as a pair (X, d_X) , where X is a two-dimensional smooth compact connected and complete Riemannian manifold with boundary, and $d_X : X \times X \rightarrow \mathbb{R}$ is the *geodesic metric* (measuring the lengths of the shortest paths on the manifold), induced by the Riemannian structure. For the brevity of notation, we will write simply X , implying (X, d_X) . Two-dimensional objects (silhouettes) are obtained as a particular case of flat manifolds.

We broadly refer to d_X as to the *intrinsic geometry* of X . From the intrinsic point of view, two objects X and Y are similar (*isometric*) if there exists a bijective distance-preserving map (*isometry*) between X and Y . More generally, X and Y are said to be ϵ -*isometric* if there exists an ϵ -*surjective* map $\varphi : X \rightarrow Y$ (i.e., $d_Y(y, \varphi(x)) \leq \epsilon$ for all $y \in Y$), which has *distortion*

$$\text{dis } \varphi = \sup_{x, x' \in X} |d_X(x, x') - d_Y(\varphi(x), \varphi(x'))| = \epsilon.$$

Such φ is called an ϵ -*isometry*.

In [44], Mikhail Gromov introduced a criterion of intrinsic geometric similarity between metric spaces, commonly known today as the *Gromov-Hausdorff distance*, defined in our setting as

$$d_{\text{GH}}(X, Y) = \frac{1}{2} \inf_{\varphi: X \rightarrow Y} \max_{\psi: Y \rightarrow X} \{\text{dis } \varphi, \text{dis } \psi, \text{dis } (\varphi, \psi)\}.$$

Here, $\text{dis}(\varphi, \psi) = \sup_{x \in X, y \in Y} |d_X(x, \psi(y)) - d_Y(y, \varphi(x))|$. The Gromov-Hausdorff distance is a metric on the quotient space of non-rigid objects under the isometry relation. Particularly, this implies that $d_{\text{GH}}(X, Y) = 0$ if and only if X and Y are isometric. More generally, if $d_{\text{GH}}(X, Y) \leq \epsilon$, then X and Y are 2ϵ -isometric and conversely, if X and Y are ϵ -isometric, then $d_{\text{GH}}(X, Y) \leq 2\epsilon$ [45]. Mémoli and Sapiro [18] introduced this criterion to the realm of three-dimensional shape matching and proposed its probabilistic approximation.

We define a part of (X, d_X) as a pair $(X', d_X|_{X'})$, where $X' \subseteq X$ and $d_X|_{X'}(x, x') = d_X(x, x')$ for all $x, x' \in X'$ is the *restricted metric*. As the measure μ_X , we use the area derived from the Riemannian structure of the manifold; technically, the corresponding σ -algebra is assumed to be a Borel algebra.

B. Partial similarity of non-rigid objects

As the dissimilarity in our framework, we use $\epsilon = d_{\text{GH}}$. Roughly speaking, it measures how non-isometric two objects or their parts are. On the other hand, there are many meaningful ways to define the partiality. For example, λ_{SUM} measures the area of the regions cropped from the objects; λ_{MAX} measures the maximum area of the cropped regions and λ_{NMAX} measures the maximum relative area of the cropped regions.

The partial dissimilarity $d_P(X, Y)$ measures the tradeoff between the intrinsic dissimilarity (Gromov-Hausdorff distance) and the area cropped from the objects (one of the partialities defined above). By properties of the Gromov-Hausdorff

distance, $(\lambda, \epsilon) \in d_P(X, Y)$ implies that there exist $X' \in \Sigma_X$ and $Y' \in \Sigma_Y$ with partiality $\lambda(X', Y')$, such that $(X', d_X|_{X'})$ and $(Y', d_Y|_{Y'})$ are 2ϵ -isometric; and if $(X', d_X|_{X'})$ and $(Y', d_Y|_{Y'})$ are ϵ -isometric, then $(\lambda(X', Y'), 2\epsilon) \in d_P(X, Y)$. If $(0, 0) \in d_P(X, Y)$, then X and Y are a.e. isometric [46].

C. Fuzzy approximation of d_P

The partial similarity computation problem requires optimization over subsets of X and Y , which, in the discrete setting, gives rise to an NP-hard combinatorial optimization problem. Using fuzzy formulation, we can pose this computation as continuous optimization.

Given two fuzzy parts $m_X \in M_X$ and $m_Y \in M_Y$, we define the fuzzy partiality according to (4). The fuzzy dissimilarity is constructed as a fuzzy version of the Gromov-Hausdorff distance,

$$\tilde{\epsilon}_\delta(m_X, m_Y) = \frac{1}{2} \inf_{\substack{\varphi: X \rightarrow Y \\ \psi: Y \rightarrow X}} \left\{ \begin{array}{l} \sup_{x, x' \in X} m_X(x)m_X(x') |d_X(x, x') - d_Y(\varphi(x), \varphi(x'))| \\ \sup_{y, y' \in Y} m_Y(y)m_Y(y') |d_Y(y, y') - d_X(\psi(y), \psi(y'))| \\ \sup_{x \in X, y \in Y} m_X(x)m_Y(y) |d_X(x, \psi(y)) - d_Y(\varphi(x), y)| \\ \sup_{x \in X} D(1 - m_Y(\varphi(x)))m_X(x) \\ \sup_{y \in Y} D(1 - m_X(\psi(y)))m_Y(y) \end{array} \right\},$$

where $D = \max\{\text{diam } X, \text{diam } Y\}/\delta(1 - \delta)$ and $0 < \delta < 1$ is a parameter.

Proposition 3: (i) $\tilde{\epsilon}_\delta(\mathbb{I}_X^*, \mathbb{I}_Y^*) = \epsilon(X', Y')$; (ii) $\tilde{\epsilon}(T_\delta m_X, T_\delta m_Y) \leq \frac{1}{\delta^2} \tilde{\epsilon}_\delta(m_X, m_Y)$, for all $0 < \delta < 1$.

The multicriterion optimization problem is defined as the minimization of the vector objective $\tilde{\Phi} = (\tilde{\lambda}, \tilde{\epsilon}_\delta)$ over the set $\tilde{\Omega} = M_X \times M_Y$. A Pareto optimum is a point (m_X^*, m_Y^*) , for which at least one of the following holds,

$$\begin{aligned} \tilde{\epsilon}_\delta(m_X^*, m_Y^*) &\leq \tilde{\epsilon}_\delta(m_X, m_Y); \text{ or,} \\ \tilde{\lambda}(m_X^*, m_Y^*) &\leq \tilde{\lambda}(m_X, m_Y), \end{aligned} \quad (7)$$

for all $(m_X, m_Y) \in \tilde{\Omega}$. The *fuzzy partial dissimilarity* $\tilde{d}_P(X, Y)$ is defined as the Pareto frontier, similarly to our previous crisp definition. Combining the results of Propositions 2 and 3, we can connect the crisp and fuzzy partial dissimilarities,

$$\tilde{d}_P(X, Y) \leq ((1 - \delta)^{-\rho}, \delta^{-2}) \cdot d_P(X, Y). \quad (8)$$

This result allows us use $\tilde{d}_P(X, Y)$ as an approximation of $d_P(X, Y)$.

D. Discretization

In practice, the computation of the partial dissimilarity is performed on discretized manifolds. The manifold X is sampled at N points $X_r = \{x_1, \dots, x_N\} \subseteq X$, constituting an r -covering, i.e., $X = \bigcup_{n=1}^N B_X(x_n, r)$ (here, $B_X(x, r)$ is a metric ball of radius r around x). The discrete manifold

is represented as a triangular mesh with N vertices x_1, \dots, x_N and T_X triangles; each triangle is a triplet of indices of vertices belonging to it. A point on the mesh is represented as a vector $\mathbf{x} = (t, \mathbf{u})$, where $t \in \{1, \dots, T_X\}$ is the index of the triangular face enclosing it, and $\mathbf{u} \in [0, 1]^2$ is the vector of *barycentric coordinates* with respect to the vertices of that triangle. We denote $U_X = \{1, \dots, T_X\} \times [0, 1]^2$.

The geodesic metric d_X is discretized by numerically approximating the geodesic distances $d_X(x_i, x_j)$ between the manifold samples, e.g., using the *fast marching method* (FMM) [47], [48]. Geodesic distances $d_X(\mathbf{x}, \mathbf{x}')$ between two arbitrary points \mathbf{x}, \mathbf{x}' on the mesh are interpolated from the values of $d_X(x_i, x_j)$ using the *three-point interpolation* presented in [49]. The error of the distance discretization is $\mathcal{O}(r^2)$.

The measure μ_X is discretized by assigning to $\mu_X(x_i)$ the area of the *Voronoi cell* around x_i and represented as a vector $\boldsymbol{\mu}_X = (\mu_X(x_1), \dots, \mu_X(x_N))^T$.

E. Iterative computation of partial similarity

Given two meshes X_r and Y_r sampled at N and M points, respectively, the set-valued partial dissimilarity $d_P(X_r, Y_r)$ is found by computing $d_{\lambda\text{-PART}}(X_r, Y_r)$ for a range of values of λ , each giving a point on the Pareto frontier. The numerical solution is similar to the generalized multidimensional scaling (GMDS) [19], [49] and, in general, to the spirit of multidimensional scaling (MDS) problems [50], [17]. Instead of optimizing over the embeddings $\varphi: X_r \rightarrow Y_r$ and $\psi: Y_r \rightarrow X_r$, optimization is performed over the images $y'_i = \varphi(x_i)$ and $x'_j = \psi(y_j)$, with $i = 1, \dots, N$ and $j = 1, \dots, M$ (note that in our notation, x'_i is a point on Y and y'_j is a point on X). The points y'_i and x'_j are represented in baricentric coordinates, as matrices $\mathbf{Y}' \in U_X^N$ and $\mathbf{X}' \in U_Y^M$.

The discrete optimization problem is formulated as follows,

$$\begin{aligned} d_{\lambda\text{-PART}}(X_r, Y_r) = \min_{\substack{\mathbf{Y}' \in U_X^N \\ \mathbf{X}' \in U_Y^M \\ \mathbf{m}_X \in [0, 1]^N \\ \mathbf{m}_Y \in [0, 1]^M}} s(\mathbf{Y}', \mathbf{X}', \mathbf{m}_X, \mathbf{m}_Y) \\ \text{s.t. } \mathbf{m}_X^T \boldsymbol{\mu}_X \geq 1 - \lambda, \\ \mathbf{m}_Y^T \boldsymbol{\mu}_Y \geq 1 - \lambda, \end{aligned} \quad (9)$$

where

$$s(\mathbf{Y}', \mathbf{X}', \mathbf{m}_X, \mathbf{m}_Y) = \max \left\{ \begin{array}{l} \max_{i,j} m_X(x_i)m_X(x_j) |d_X(x_i, x_j) - d_Y(\mathbf{y}'_i, \mathbf{y}'_j)| \\ \max_{k,l} m_Y(y_k)m_Y(y_l) |d_Y(y_k, y_l) - d_X(\mathbf{x}'_k, \mathbf{x}'_l)| \\ \max_{i,k} m_X(x_i)m_Y(y_k) |d_X(x_i, \mathbf{x}'_k) - d_Y(y_k, \mathbf{y}'_i)| \\ \max_i D(1 - m_X(\mathbf{x}'_i))m_X(x_i) \\ \max_k D(1 - m_X(\mathbf{y}'_k))m_Y(y_k) \end{array} \right\},$$

and $\mathbf{m}_X = (m_X(x_1), \dots, m_X(x_N))$ and $\mathbf{m}_Y = (m_Y(y_1), \dots, m_Y(y_M))$ represent the discretized membership functions. The values $m_X(\mathbf{x}'_i)$ and $m_Y(\mathbf{y}'_i)$ at arbitrary points of the triangular mesh are computed by interpolation.

The geodesic distances $d_X(x_i, x_j)$ and $d_Y(y_k, y_l)$ are pre-computed by FMM; on the other hand, the distances $d_Y(\mathbf{y}'_i, \mathbf{y}'_k)$, $d_X(\mathbf{x}'_k, \mathbf{x}'_l)$, $d_X(x_i, \mathbf{x}'_k)$ and $d_Y(y_k, \mathbf{y}'_i)$ are interpolated.

Problem (9) is solved using alternating minimization, consisting of two stages, repeated until convergence: First, the values of $\mathbf{m}_X, \mathbf{m}_Y$ are fixed and minimization is performed over \mathbf{X}' and \mathbf{Y}' , cast as the constrained minimization problem,

$$\begin{aligned} & \min_{\substack{\epsilon \geq 0 \\ \mathbf{Y}' \in U_X^N \\ \mathbf{X}' \in U_Y^M}} \epsilon & (10) \\ \text{s.t.} & \begin{cases} m_X(x_i)m_X(x_j)|d_X(x_i, x_j) - d_Y(\mathbf{y}'_i, \mathbf{y}'_k)| \leq \epsilon \\ m_Y(y_k)m_Y(y_l)|d_Y(y_k, y_l) - d_X(\mathbf{x}'_k, \mathbf{x}'_l)| \leq \epsilon \\ m_X(x_i)m_Y(y_k)|d_X(x_i, \mathbf{x}'_k) - d_Y(y_k, \mathbf{y}'_i)| \leq \epsilon \\ D(1 - m_X(\mathbf{x}'_i))m_X(x_i) \leq \epsilon \\ D(1 - m_Y(\mathbf{y}'_k))m_Y(y_k) \leq \epsilon \end{cases} \end{aligned}$$

with $i, j = 1, \dots, N$ and $k, l = 1, \dots, M$. Numerical solution of problem (10) requires the ability to perform a step in a given direction on a triangulated mesh (such a path is polylinear if it traverses more than one triangle), computed using an *unfolding* procedure described in [49]. The second stage is performed by fixing \mathbf{X}' and \mathbf{Y}' and minimizing with respect to $\mathbf{m}_X, \mathbf{m}_Y$,

$$\begin{aligned} & \min_{\substack{\epsilon \geq 0 \\ \mathbf{m}_X \in [0,1]^N \\ \mathbf{m}_Y \in [0,1]^M}} \epsilon & (11) \\ \text{s.t.} & \begin{cases} m_X(x_i)m_X(x_j)|d_X(x_i, x_j) - d_Y(\mathbf{y}'_i, \mathbf{y}'_k)| \leq \epsilon \\ m_Y(y_k)m_Y(y_l)|d_Y(y_k, y_l) - d_X(\mathbf{x}'_k, \mathbf{x}'_l)| \leq \epsilon \\ m_X(x_i)m_Y(y_k)|d_X(x_i, \mathbf{x}'_k) - d_Y(y_k, \mathbf{y}'_i)| \leq \epsilon \\ D(1 - m_X(\mathbf{x}'_i))m_X(x_i) \leq \epsilon \\ D(1 - m_Y(\mathbf{y}'_k))m_Y(y_k) \leq \epsilon \\ \mathbf{m}_X^\top \boldsymbol{\mu}_X \geq 1 - \lambda \\ \mathbf{m}_Y^\top \boldsymbol{\mu}_Y \geq 1 - \lambda \end{cases} \end{aligned}$$

The entire iterative optimization algorithm can be summarized as follows:

```

1 for  $\lambda = 0, \Delta\lambda, \dots, \lambda_{\max}$  do
2   Initialization: set  $k \leftarrow 0$ ;  $\mathbf{m}_X^{(0)} = 1, \mathbf{m}_Y^{(0)} = 1$ ;
    $\mathbf{X}'^{(0)}, \mathbf{Y}'^{(0)}$ .
3   repeat
4     Compute the  $(k + 1)$ st iteration solution
      $\mathbf{X}'^{(k+1)}, \mathbf{Y}'^{(k+1)}$  by solving problem (10) with
     fixed  $\mathbf{m}_X^{(k)}, \mathbf{m}_Y^{(k)}$ .
5     Compute the  $(k + 1)$ st iteration solution
      $\mathbf{m}_X^{(k+1)}, \mathbf{m}_Y^{(k+1)}$  by solving problem (11) with
     fixed  $\mathbf{X}'^{(k)}, \mathbf{Y}'^{(k)}$ .
6     Set  $k \leftarrow k + 1$ .
7   until convergence
8    $d_{\lambda\text{-PART}}(X_r, Y_r) = s(\mathbf{Y}'^{(k)}, \mathbf{X}'^{(k)}, \mathbf{m}_X^{(k)}, \mathbf{m}_Y^{(k)})$ .
9 end

```

Convergence at Step 7 is determined when the optimization variables on subsequent iterations do not change significantly, i.e., the values of $\|\mathbf{m}_X^{(k)} - \mathbf{m}_X^{(k-1)}\|$, $\|\mathbf{X}'^{(k)} - \mathbf{X}'^{(k-1)}\|$ and $\|\mathbf{Y}'^{(k)} - \mathbf{Y}'^{(k-1)}\|$ are below some threshold. At Steps 4 and 5, convex optimization algorithms are used. Since the problems are non-convex, such algorithms are liable to converge to a local minimum, a caveat widely known in MDS problems [50]. Local convergence can be avoided in practice by using a multiresolution scheme [51], [52], in which a hierarchy of problems is constructed, starting from a coarse version of the problem containing a small number of points. The coarse level solution is interpolated to the next resolution level, and is used as an initialization for the optimization at that level. The process is repeated until the finest level solution is obtained. We use this multiresolution scheme at Stage 5 of the algorithm.

F. Results

In order to exemplify the presented method, two experiments were performed. Our first experiment is a numerical demonstration of the mythological creatures example we used as a motivation. The two-dimensional objects (human, horse and centaur) were represented as binary images and subsampled using the *farthest point strategy* [53], [54], [17] at approximately 3000 points. The geodesic distances between the samples were approximated numerically using FMM. Thirteen values of λ were used to compute the Pareto frontier. Figure 4 shows the partial dissimilarities between the objects. We can say that a human is more similar to a centaur than to a horse, because the Pareto frontier corresponding to the human-centaur comparison (green) is below that corresponding to the human-horse comparison (red). Additional examples with two-dimensional objects are shown in [55].

In the second experiment, we used five three-dimensional non-rigid objects from [17]. Each object appeared in five different deformations with different parts missing (see Figure 6). The objects were represented as triangular meshes, sampled between 1500 to 3000 points, with geodesic distances computed using FMM. Partial dissimilarities were computed between all the objects using 13 values of λ . We used the scheme described in Section IV-E with six-level multiresolution optimization. The finest resolution level contained 50 points. Figure 5 shows the Pareto frontiers arising from partial comparison of the dog object to different objects. One can observe that a dog and dog-man and the dog-giraffe comparisons (red) result in curves above those obtained for the comparison of different instances of the dog (black). Figure 6 depicts the L_1 -Salukwadze distance between the objects, represented as Euclidean similarities using MDS [50]. Clusters corresponding to different objects are clearly distinguishable. For additional results and examples, the reader is referred to [46].

V. CASE STUDY II – CHARACTER SEQUENCES

Another application of our partial similarity is the analysis of character sequences. Problems requiring the comparison of such sequences arise in linguistics [56], web search [57], [58], spell checking [59], plagiarism detection [60], speech

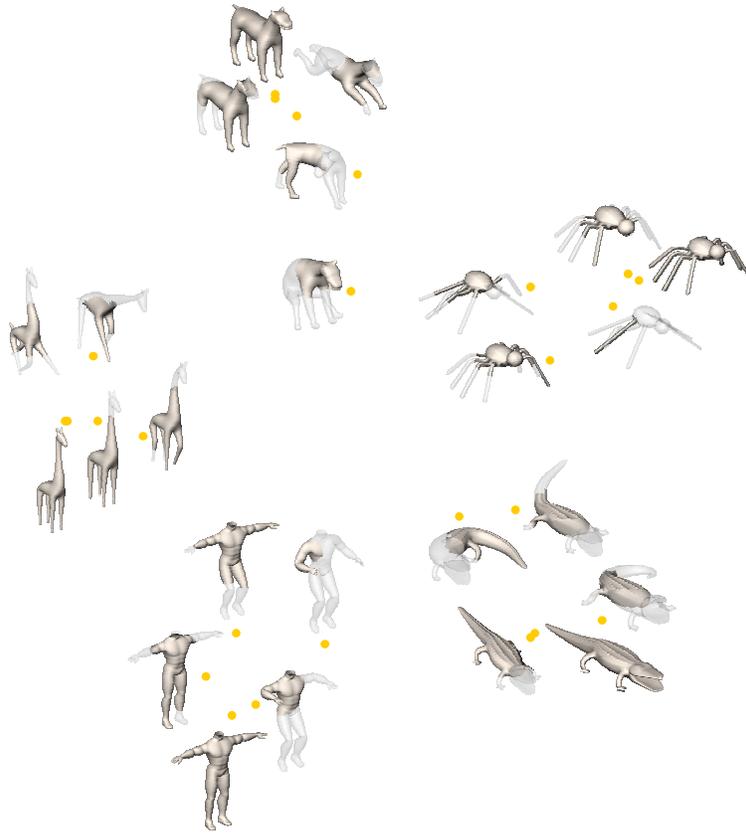


Fig. 6. Comparison of partially-overlapping deformable objects. In this example, we show a Euclidean representation of the Salukwadze distance, which captures the partial similarity between such objects (light shading represent the missing parts). Each point in the plane represents an object, and Euclidean distances between pairs of points approximate the corresponding Salukwadze distances.

recognition, and bioinformatics [61], [62]. The basic problem in this field is finding subsets of sequences that are similar to some give pattern – again, a problem fitting nicely into the concept of partial similarity.

A. Background

The object used in text analysis is a *sequence* $X = (x_n)_{n=1}^N$. Each x_k (called *character*) is an element in some set A , referred to as the *alphabet*. For example, in text analysis A can be the Latin alphabet, and in bioinformatics, A is the set of four nucleotides. A part is a sequence of the form $X' = (x_{n_k})$ (called a *subsequence* of X), where n_k is a strictly increasing sequence in the index set $\{1, \dots, N\}$. The σ -algebra Σ_X in this problem is defined as the collection of all the subsequences of X . The measure is the subsequence length, $\mu_X(X') = |X'|$.

Given two sequences X and Y , a *longest common subsequence* (LCS) of X and Y is defined as

$$lcs(X, Y) = \operatorname{argmax}_{Z \in \Sigma_X \cap \Sigma_Y} |Z|. \quad (12)$$

Note that the LCS may not be unique; for example, the longest common subsequences of AATCC and ACACG are the sequences ACC and AAC.

If X and Y are of equal length, we can define the *Hamming distance* between X and Y as the number of character

substitution edits required to transform one string to another,

$$d_{\text{HAM}}(X, Y) = \sum_{n=1}^{|X|} \mathbb{I}_{x_n \neq y_n}. \quad (13)$$

For sequences of non-equal length, the Hamming distance can be extended by considering not only the substitution edits, but also character insertions and deletions. A classical tool in text analysis, known as the *edit-* or *Levenshtein distance* and denoted here by $d_E(X, Y)$, is defined as the minimum number of edits required to transform one string to another, where an edit is either a single character deletion or insertion (which adds 1 to the distance), or character substitution (which adds 2)² [20], [63]. The edit distance is related to the longest common subsequence,

$$d_E(X, Y) = |X| + |Y| - 2|lcs(X, Y)|. \quad (14)$$

B. Partial similarity of text sequences

To define the partial similarity between character sequences, we use $\epsilon = d_{\text{HAM}}$ as the dissimilarity. If the subsequences are not of equal length, ϵ is undefined. The partiality is defined as the total number of characters dropped from the sequences X and Y to obtain the two sub-sequences X' and

²In some definitions, character substitution adds up 1 to d_E .

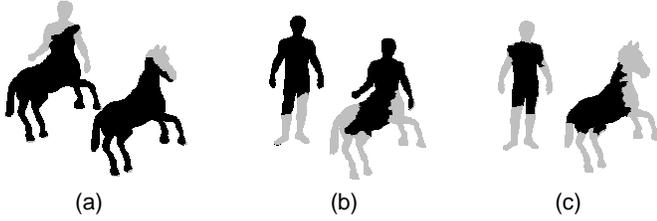
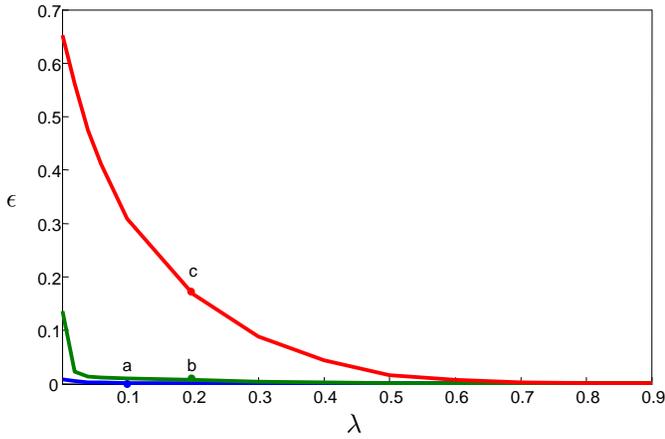


Fig. 4. Partial similarity between mythological creatures. Shown are the set-valued distances between human and horse (red), human and centaur (green) and centaur and horse (blue).

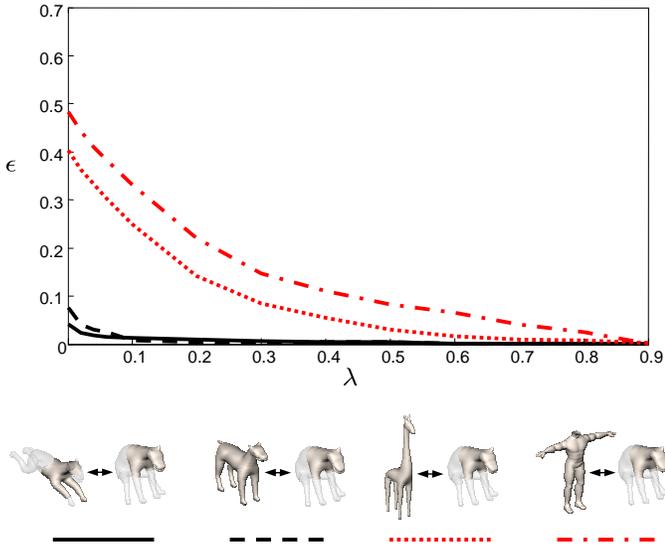


Fig. 5. Comparison of the partially-missing dog object to other objects. The set-valued partial dissimilarity clearly indicates that dog is more similar to other instances of dogs than to other objects.

Y' , $\lambda_{\text{SUM}}(X', Y') = |X| + |Y| - (|X'| + |Y'|)$. As the result of the tradeoff between $d_{\text{HAM}}(X', Y')$ and $\lambda_{\text{SUM}}(X', Y')$, a discrete Pareto frontier $d_{\text{P}}(X, Y)$ is obtained. If the value of $|X| + |Y|$ is even, $d_{\text{P}}(X, Y)$ exists only at even λ_{SUM} points³; otherwise, it is defined only at odd values of λ_{SUM} .

³Subsequences X' and Y' must be of equal length in order for d_{HAM} to be defined, such that $|X'| + |Y'|$ is always even. If $|X| + |Y|$ is even, an odd value of $\lambda_{\text{SUM}}(X', Y')$ implies that X' and Y' are of unequal length and consequently, the Pareto frontier is not defined at this point.

We can establish the following relation between the zero-dissimilarity distance and the edit distance:

Theorem 1: (i) $d_{0\text{-DIS}}(X, Y) = d_{\text{E}}(X, Y)$;
(ii) $d_{0\text{-DIS}}(X, Y)$ is realized on subsequences $X' = Y' = \text{lcs}(X, Y)$.

In other words, the edit distance is a particular case of our partial similarity framework, obtained by selecting a specific point on the Pareto frontier, corresponding to the minimum partiality obtained requiring that d_{HAM} is zero. However, we may allow for subsequences which are not similar ($d_{\text{HAM}} > 0$), yet, have smaller partiality. This brings us to the definition of the Salukwadze distance $d_{\text{S}}(X, Y)$, which may better quantify the partial similarity between two sequences.

C. Numerical example

To demonstrate the partial similarity concept in text analysis, we compare two sequences: $X = \text{PARTIAL SIMILARITY}$ and $Y = \text{PARETO OPTIMUM}$. The obtained discrete Pareto frontier is shown in Figure 7. Point marked as (a) on the Pareto frontier in Figure 7 corresponds to the smallest Hamming distance with the smallest possible partiality ($\lambda_{\text{SUM}} = 4$). It is realized on subsequences $X' = \text{PARTIAL SIMILARITY}$ and $Y' = \text{PARETO OPTIMUM}$, the Hamming distance between which is $d_{\text{HAM}}(X', Y') = 9$. Point (b) corresponds to the L_2 -Salukwadze distance ($d_{\text{S}}(X, Y) = \|(6, 7)\|_2 = \sqrt{85}$). It is realized on subsequences $\text{PARTIAL SIMILARITY}$ and PARETO OPTIMUM (highlighted in red in Figure 7b). Point (c) is the smallest value of partiality ($\lambda_{\text{SUM}} = 18$), for which d_{HAM} is zero, i.e., $d_{0\text{-DIS}}(X, Y) = 18$. According to Theorem 1(ii), it is realized on a LCS, which in our example is $\text{PARTIAL SIMILARITY}$ (highlighted in Figure 7c). It is easy to verify that $d_{\text{E}}(X, Y) = 18$ as well, which is an empirical evidence that Theorem 1(i) holds in this case.

VI. CONCLUSIONS

We presented a method for quantifying the partial similarity between objects, based on selecting parts of the objects with the optimal tradeoff between dissimilarity and partiality. We use the formalism of Pareto optimality to provide a definition to such a tradeoff. We demonstrated our approach on two problems, comparison of non-rigid objects and analysis of text sequences. In both cases, our construction has a meaningful interpretation. The set-valued distances resulting from it have appealing theoretical and practical properties. Particularly, in shape matching and text analysis, they can be viewed as a generalization of prior results. The presented framework of partial similarity is generic and can be applied to different pattern recognition problems.

REFERENCES

- [1] B. Greene, *The elegant universe*. Vintage Books New York, 2000.
- [2] M. Stricker and M. Orengo, "Similarity of color images," *Proc. SPIE*, vol. 2420, pp. 381–392, 1995.
- [3] V. N. Gudivada and V. V. Raghavan, "Content based image retrieval systems," *Computer*, vol. 28, no. 9, pp. 18–22, 1995.
- [4] Y. Chen and E. Wong, "Augmented image histogram for image and video similarity search," *Proc. SPIE*, vol. 3656, p. 523, 2003.

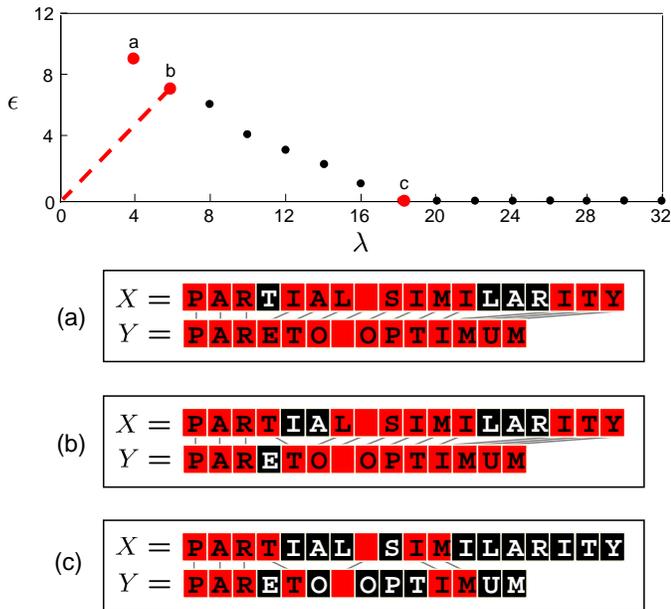


Fig. 7. Comparison of sequences of characters. Red denotes the subsequences. Point (b) corresponds to the L_2 -Salukwadze distance. Point (c) is realized on the LCS. At this point, λ equals the value of d_E and $d_{0-DIS}(X, Y)$.

- [5] C. Kim and C. Chung, "XMage: An image retrieval method based on partial similarity," *Information Processing and Management*, vol. 42, pp. 484–502, 2006.
- [6] A. Bruckstein, N. Katzir, M. Lindenbaum, and M. Porat, "Similarity-invariant signatures for partially occluded planar shapes," *International Journal of Computer Vision*, vol. 7, no. 3, pp. 271–285, 1992.
- [7] D. G. R. Basri, L. Costa and D. Jacobs, "Determining the similarity of deformable shapes," *Vision Research*, vol. 38, pp. 2365–2385, 1998.
- [8] L. J. Latecki and R. Lakamper, "Shape similarity measure based on correspondence of visual parts," *IEEE Trans. PAMI*, vol. 22, no. 10, pp. 1185–1190, 2000.
- [9] S. W. Cheng, H. Edelsbrunner, P. Fu, and K. P. Lam, "Design and analysis of planar shape deformation," *Computational Geometry: Theory and Applications*, vol. 19, no. 2-3, pp. 205–218, 2001.
- [10] D. Geiger, T. L. Liu, and R. Kohn, "Representation and Self-Similarity of Shapes," *IEEE Trans. PAMI*, vol. 25, no. 1, pp. 86–99, 2003.
- [11] D. Jacobs, D. Weinshall, and Y. Gdalyahu, "Class representation and image retrieval with non-metric distances," *IEEE Trans. PAMI*, vol. 22, pp. 583–600, 2000.
- [12] D. J. H. Ling, "Deformation invariant image matching," in *Proc. ICCV*, 2005.
- [13] P. F. Felzenszwalb, "Representation and detection of deformable shapes," *IEEE Trans. PAMI*, vol. 27, no. 2, pp. 208–220, 2005.
- [14] A. M. Bronstein, M. M. Bronstein, A. M. Bruckstein, and R. Kimmel, "Matching two-dimensional articulated shapes using generalized multi-dimensional scaling," in *Proc. AMDO*, 2006, pp. 48–57.
- [15] Z. Y. Zhang, "Iterative point matching for registration of free-form curves and surfaces," *IJCV*, vol. 13, pp. 119–152, 1994.
- [16] A. Tal, M. Elad, and S. Ar, "Content based retrieval of VRML objects - an iterative and interactive approach," in *Proc. Eurographics Workshop on Multimedia*, 2001.
- [17] A. Elad and R. Kimmel, "Bending invariant representations for surfaces," in *Proc. CVPR*, 2001, pp. 168–174.
- [18] F. Mémoli and G. Sapiro, "A theoretical and computational framework for isometry invariant recognition of point cloud data," *Foundations of Computational Mathematics*, vol. 5, no. 3, pp. 313–347, 2005.
- [19] A. M. Bronstein, M. M. Bronstein, and R. Kimmel, "Generalized multidimensional scaling: a framework for isometry-invariant partial surface matching," *Proc. National Academy of Sciences*, vol. 103, no. 5, pp. 1168–1172, January 2006.
- [20] V. I. Levenshtein, "Binary codes capable of correcting deletions, insertions, and reversals," *Doklady Akademii Nauk SSSR*, vol. 163, no. 4, pp. 845–848, 1965.
- [21] V. Hatzivassiloglou, J. Klavans, and E. Eskin, "Detecting text similarity over short passages: Exploring linguistic feature combinations via machine learning," *Proceedings of the Joint SIGDAT Conference on Empirical Methods in Natural Language Processing and Very Large Corpora*, 1999.
- [22] J. Foote, "Content-based retrieval of music and audio," *Proc. SPIE*, vol. 3229, p. 138, 1997.
- [23] J. Foote, M. Cooper, and U. Nam, "Audio retrieval by rhythmic similarity," *Proc. International Conf. Music Information Retrieval*, vol. 3, pp. 265–266, 2002.
- [24] P. Hallinan, "A low-dimensional representation of human faces for arbitrary lighting conditions," in *Proc. CVPR*, 1994, pp. 995–999.
- [25] A. M. Bronstein, M. M. Bronstein, E. Gordon, and R. Kimmel, "Fusion of 3D and 2D information in face recognition," in *Proc. ICIP*, 2004, to appear.
- [26] A. S. Gheorghides, P. N. Belhumeur, and D. J. Kriegman, "From few to many: illumination cone models for face recognition under variable lighting and pose," *IEEE Trans. PAMI*, vol. 23, no. 6, pp. 643–660, 2001.
- [27] A. M. Bronstein, M. M. Bronstein, and R. Kimmel, "Expression-invariant 3D face recognition," in *Proc. Audio and Video-based Biometric Person Authentication*, ser. Lecture Notes on Computer Science, no. 2688. Springer, 2003, pp. 62–69.
- [28] —, "Three-dimensional face recognition," *IJCV*, vol. 64, no. 1, pp. 5–30, August 2005.
- [29] S. Berchtold, D. Keim, and H. Kriegel, "Using extended feature objects for partial similarity retrieval," *International Journal on Very Large Data Bases*, vol. 6, no. 4, pp. 333–348, 1997.
- [30] R. C. Veltkamp, "Shape matching: similarity measures and algorithms," in *International Conference on Shape Modeling and Applications*, 2001, pp. 188–197.
- [31] L. Latecki, R. Lakaemper, and D. Wolter, "Optimal Partial Shape Similarity," *Image and Vision Computing*, vol. 23, pp. 227–236, 2005.
- [32] O. Boiman and M. Irani, "Similarity by composition," in *International Conference on Neural Information Processing Systems*, 2006.
- [33] A. M. Bronstein, A. M. Bronstein, and R. Kimmel, "Robust expression-invariant face recognition from partially missing data," in *Proc. ECCV*, 2006, pp. 396–408.
- [34] L. A. Zadeh, "Fuzzy sets," *Information and Control*, vol. 8, pp. 338–353, 1965.
- [35] V. Pareto, *Manuale di Economia Politica*, 1906.
- [36] S. de Rooij and P. Vitanyi, "Approximating rate-distortion graphs of individual data: Experiments in lossy compression and denoising," *IEEE Trans. Information Theory*, 2006, submitted.
- [37] R. M. Everson and J. E. Fieldsend, "Multi-class ROC analysis from a multi-objective optimization perspective," *Pattern Recognition Letters*, vol. 27, no. 8, pp. 918–927, 2006.
- [38] M. E. Salukwadze, *Vector-Valued Optimization Problems in Control Theory*. Academic Press, 1979.
- [39] J. Zhang, R. Collins, and Y. Liu, "Representation and matching of articulated shapes," in *Proc. CVPR*, vol. 2, June 2004, pp. 342–349.
- [40] H. Ling and D. Jacobs, "Using the inner-distance for classification of articulated shapes," in *Proc. CVPR*, 2005.
- [41] M. Reuter, F.-E. Wolter, and N. Peinecke, "Laplace-Beltrami spectra as shape-DNA of surfaces and solids," *Computer-Aided Design*, vol. 38, pp. 342–366, 2006.
- [42] A. M. Bronstein, M. M. Bronstein, and R. Kimmel, "Face2face: an isometric model for facial animation," in *Proc. AMDO*, 2006, pp. 38–47.
- [43] —, "Calculus of non-rigid surfaces for geometry and texture manipulation," *IEEE Trans. Visualization and Computer Graphics*, accepted.
- [44] M. Gromov, *Structures métriques pour les variétés riemanniennes*, ser. Textes Mathématiques, 1981, no. 1.
- [45] D. Burago, Y. Burago, and S. Ivanov, *A course in metric geometry*, ser. Graduate studies in mathematics. American Mathematical Society, 2001, vol. 33.
- [46] M. M. Bronstein, A. M. Bronstein, A. M. Bruckstein, and R. Kimmel, "Paretian similarity of non-rigid objects," in *Scale Space*, submitted.
- [47] J. A. Sethian, "A review of the theory, algorithms, and applications of level set method for propagating surfaces," *Acta numerica*, pp. 309–395, 1996.

- [48] R. Kimmel and J. A. Sethian, "Computing geodesic on manifolds," in *Proc. US National Academy of Science*, vol. 95, 1998, pp. 8431–8435.
- [49] A. M. Bronstein, M. M. Bronstein, and R. Kimmel, "Efficient computation of isometry-invariant distances between surfaces," *SIAM J. Sci. Comp.*, 2006, to appear.
- [50] I. Borg and P. Groenen, *Modern multidimensional scaling - theory and applications*. Springer-Verlag, Berlin Heidelberg New York, 1997.
- [51] M. M. Bronstein, A. M. Bronstein, R. Kimmel, and I. Yavneh, "A multigrid approach for multidimensional scaling," in *Proc. Copper Mountain Conf. Multigrid Methods*, 2005.
- [52] —, "Multigrid multidimensional scaling," *Numerical Linear Algebra with Applications (NLAA)*, vol. 13, pp. 149–171, March-April 2006.
- [53] Y. Eldar, M. Lindenbaum, M. Porat, and Y. Y. Zeevi, "The farthest point strategy for progressive image sampling," *IEEE Trans. Image Processing*, vol. 6, no. 9, pp. 1305–1315, 1997.
- [54] A. M. Bruckstein, R. Holt, and A. N. Netravali, "Holographic representations of images," *IEEE Trans. Image Processing*, vol. 7, no. 11, pp. 1583–1597, 1998.
- [55] A. M. Bronstein, M. M. Bronstein, A. M. Bruckstein, and R. Kimmel, "Analysis of two-dimensional non-rigid shapes," *IJCV*, submitted.
- [56] C. Gooskens and W. Heeringa, "Perceptive evaluation of Levenshtein dialect distance measurements using Norwegian dialect data," *Language Variation and Change*, vol. 16, pp. 189–2007, 2004.
- [57] C. L. Giles, K. D. Bollacker, and S. Lawrence, "CiteSeer: an automatic citation indexing system," *Proc. 3rd ACM conference on Digital libraries*, pp. 89–98, 1998.
- [58] G. A. Di Lucca, M. Di Penta, and A. R. Fasolino, "An approach to identify duplicated web pages," *Proc. COMPSAC*, pp. 481–486, 2002.
- [59] F. J. Damerau, "A Technique for Computer Detection and Correction of Spelling Errors."
- [60] M. J. Wise, "YAP3: improved detection of similarities in computer program and other texts," *Proc. 27th SIGCSE Technical Symposium on Computer science education*, pp. 130–134, 1996.
- [61] J. B. Kruskal, *An overview of sequence comparison*. CSLI Publications, 1999, ch. Time Warps, String Edits, and Macromolecules: The Theory and Practice of Sequence Comparison.
- [62] S. Bonhoeffer, C. Chappey, N. T. Parkin, J. M. Whitcomb, and C. J. Petropoulos, "Evidence for positive epistasis in HIV-1," *Science*, vol. 306, pp. 1547–1550, 2004.
- [63] R. A. Wagner and M. J. Fischer, "The string-to-string correction problem," *JACM*, vol. 21, no. 1, pp. 168–173, 1974.