

# Modelling Convex Shape Priors and Matching Based on the Gromov-Wasserstein Distance

Bernhard Schmitzer · Christoph Schnörr

Published online: 14 August 2012  
© Springer Science+Business Media, LLC 2012

**Abstract** We present a novel convex shape prior functional with potential for application in variational image segmentation. Starting point is the Gromov-Wasserstein Distance which is successfully applied in shape recognition and classification tasks but involves solving a non-convex optimization problem and which is non-convex as a function of the involved shape representations. In two steps we derive a convex approximation which takes the form of a modified transport problem and inherits the ability to incorporate vast classes of geometric invariances beyond rigid isometries. We propose ways to counterbalance the loss of descriptiveness induced by the required approximations and to process additional (non-geometric) feature information. We demonstrate combination with a linear appearance term and show that the resulting functional can be minimized by standard linear programming methods and yields a bijective registration between a given template shape and the segmented foreground image region. Key aspects of the approach are illustrated by numerical experiments.

**Keywords** Shape prior · Wasserstein distance · Convex relaxation · Image segmentation

## 1 Introduction

### 1.1 Overview, Motivation

Convex variational approaches have been applied successfully in image processing and computer vision to obtain

nearly global optimizers for models that are originally combinatorial and hard to solve exactly [7, 16, 21]. The energy functionals are usually composed of a data term and a regularizer. In image segmentation the data term can be used to process various kinds of local features whereas the regularizer aims to keep the resulting segmentation contours smooth.

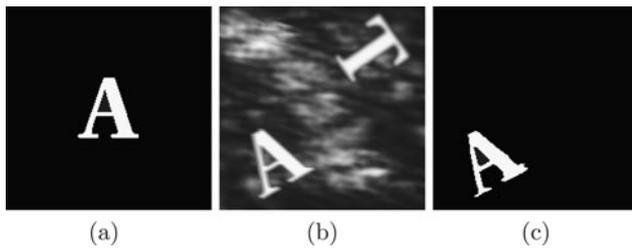
To this date the global shape of the contour has not been incorporated into this framework in a satisfying fashion. Common proposals to formulate shape-prior functionals describe shapes by parametrizing their contour or as the level set of a function, suitable for applying machine learning methods to obtain a notion of the set of allowed contours [8, 9, 27]. However there are several severe drawbacks:

- (i) The contour representation is not computationally compatible with the usual representation of segmentation regions by their indicator functions. The map between the two representations is mathematically complex.
- (ii) Except for the simple case of Gaussian statistics more sophisticated penalty functions employing kernel methods tend to yield highly non-convex functionals.
- (iii) Making the prior functional invariant under reparametrization of the contour or Euclidean isometries is a tedious task. While this alleviates the problem of correspondence between contour points, the mechanisms underlying shape matching and the integration of such a prior with variational segmentation are involved computationally and from the viewpoint of optimization.

For the tasks of shape recognition, classification and for finding meaningful correspondences between two shapes powerful approaches based on the Gromov-Hausdorff Distance and related shape similarity measures have been applied, being able to handle vast classes of transformations

---

B. Schmitzer (✉) · C. Schnörr  
Heidelberg University, Heidelberg, Germany  
e-mail: [Bernhard.Schmitzer@iwr.uni-heidelberg.de](mailto:Bernhard.Schmitzer@iwr.uni-heidelberg.de)



**Fig. 1** Shape prior based convex variational segmentation: (a) Template shape for representing prior knowledge. (b) Input image, *gray* values are interpreted as foreground affinity coefficients (*white* indicating foreground). (c) Globally optimal segmentation: the correct shape is located. Clutter and objects with wrong shape are neglected

of the data (see Sect. 2.2). But the power is paid for by computational complexity. Thus, usually only subsets of the shapes, obtained e.g. by farthest point sampling, can be compared, which is not enough if the segmentation process requires resolution on the pixel level. Also these frameworks consider the shapes to be static in the sense that they are only concerned in computing the distance between fixed shapes and do not address the question of how one shape should be altered to obtain a better matching.

## 1.2 Contribution, Organization

In this paper we propose a novel approach to modelling shape priors for guiding image segmentation. We will rely on metric measure spaces (mm-spaces) [13, 20] for shape representation. Due to its computational complexity the Gromov-Wasserstein Distance cannot be used for the construction of shape prior functionals directly. Its evaluation involves solving a non-convex quadratic assignment problem and it is non-convex as a function of the compared shapes. To overcome these problems we present two suitable approximation steps, arriving at a modified Wasserstein Distance with a particular cost function and relaxed marginal constraints.

The representation of shapes by mm-spaces is compatible to the representation of segmentation regions by relaxed indicator functions which simplifies the application in variational segmentation approaches. In this paper we demonstrate combination with a linear appearance term common to many models. We show that in this case shape optimization and distance computation for the optimal shape can be performed simultaneously by solving a single partial linear assignment problem.

The key aspects of the proposed approach are

- (i) a sound mathematical basis for both shape and matching,
- (ii) its convexity which yields globally optimal results independent of initialization,
- (iii) the generation of a full correspondence between two given shapes,

- (iv) the implementation of large classes of geometric invariances by choosing a suitable metric,
- (v) representation compatibility which simplifies combination with other terms (e.g. appearance model) and application in image segmentation tasks (for a related framework see [25]),
- (vi) the applicability to a wide range of data within the very same framework based on metric measure spaces.

Figure 1 illustrates an example of how the proposed shape prior can be applied to image segmentation in a convex variational framework.

The scope of this paper is to present the mathematical framework and to illustrate the properties above by a range of numerical experiments. The integration of the approach into a full variational segmentation approach is beyond the scope and subject of our future work.

The rest of the paper is organized as follows: Sect. 2 will review related literature, Sect. 3 will introduce the necessary mathematical background for our own developments, described in Sect. 4. Numerical illustrations of the proposed prior will be discussed in Sect. 5. The paper concludes in Sect. 6.

## 2 Related Literature

### 2.1 Wasserstein Distance in Image Processing

The notion of optimal transport dates back to the 18<sup>th</sup> century when Monge considered the question of how to move construction materials in the most efficient way. Kantorovich independently rediscovered the problem in the 20<sup>th</sup> century in a convex formulation apt for the language of linear programming. Optimal transport since then has been studied by numerous researchers and applied to a wide range of problems in various fields (see for example [29] for a modern comprehensive monograph and a brief historical outline). Thus, a variety of labels and names is associated with it: Optimal Transport, Mass Transport, “Earth Mover’s Distance”, Monge, Kantorovich, Rubinstein, Wasserstein, to name a few. For the sake of readability we will in the following use the term *Wasserstein Distance* without intending to take credit away from everyone else.

Also in mathematical image processing the Wasserstein Distance has become a powerful tool: In [14] the  $L^2$  Wasserstein Distance between two grayscale images is computed. The application in mind is to obtain registrations between different images of a non-rigid object in different states of deformation, a task that for example comes up in analyzing medical imaging data of moving organs. The questions of geometric invariance and combination with appearance models are not addressed in this paper: a meaningful registration will only be computed if the images are aligned

properly and gray values are directly converted into mass densities without handling potential detection errors.

The authors of [4] use the Wasserstein Distance as a measure of data-fidelity in an image regularization task. Given noisy input data a target measure is sought after that minimizes the weighted sum of a regularization function and the  $L^2$  Wasserstein Distance between input and target. Naturally the target measure is not known beforehand. This is a common feature of our work presented here although we swap the roles, as we are not concerned with density estimation: The Wasserstein Distance will play the part of the regularizer while another function will determine the fidelity between in- and output.

The idea of [22] is to compute the Wasserstein Distance between point clouds of descriptor vectors taken from two different objects to perform shape classification. For this purpose working with subsamples (in [22] only taken from the contours) of two shapes is sufficient. This will not be possible when a full correspondence is wanted. The generation of the descriptor vectors only uses shape-internal geodesic distances and thus is invariant under the corresponding class of geometric transformations. Hence, the method is able to recognize the same shape in different poses which leads up to the next point to be discussed.

## 2.2 Isometry Invariant Shape Classification

An important challenge in shape related tasks is the incorporation of various geometric invariances into the process. In this context it has proven to be a powerful approach to represent shapes by autonomous metric spaces, detaching them from their embedding spaces and thus providing large classes of invariances, depending on the choice of the metric. Equipping an object with the Euclidean metric of the embedding space renders internal distances invariant under translations, rotations and reflections, the Euclidean isometries. Articulated, non-rigid objects can appear in various poses with very different Euclidean distances in the embedding space. The geodesic metric based on pathlengths on or within the shapes is (in addition to Euclidean isometries) approximately invariant under so-called bendings, deformations that for example correspond to the movement of a joint.

For classification purposes one then has to develop a notion of distance between two such metric spaces. One could compute some kind of shape signature based only on the intrinsic metric information and compare these signatures for classification purposes. See [22, 23] for signatures of different complexity, ranging from a sequence of numbers to point clouds in low dimensional spaces. In [11] the shapes are first detached from their Euclidean embedding space and equipped with their geodesic metrics. Then they are re-embedded into another low dimensional Euclidean space trying to replace the geodesic by the Euclidean metric with

as little distortion as possible (see Multi-Dimensional Scaling). This strips off the vast class of non-Euclidean geodesic isometries from the representations. The remaining isometries can separately be dealt with by the final comparison process.

Alternatively, as proposed in [20] one can use the Gromov-Hausdorff distance (introduced in [13]) to measure directly the similarity of the whole metric structures corresponding to two shapes, yielding equality if and only if the two derived metric spaces are isometric.

Computing the Gromov-Hausdorff distance involves solving a combinatorial problem. This has led to the development of the closely related Gromov-Wasserstein Distance which can be computed by solving a quadratic optimization problem with linear constraints. The objective function however is still non-convex, rendering solving practically impossible for large problem dimensions [18]. The fully discretized case under the above-mentioned constraint of measures of equal weights is known as the Quadratic Assignment Problem in the combinatorial literature [5].

The developers of these methods are naturally aware of such obstacles and thus strive to supply bounds to the exact distances that are quick to compute and thus can be used for preliminary decisions [18]. We will pursue a similar direction in this paper.

It should be mentioned that the choice of the underlying metric is a field of research in itself: The geodesic metric appears to be the obvious choice to make the metric structure independent of the pose. However it exhibits strong sensibility to noise, especially to what one calls *topological noise*, the accidental connection between initially unconnected parts (for example two legs touching at the lower ends). The diffusion distance of two points assigns a weighted average of all available paths between these and thus provides more robustness towards such perturbations [3, 19].

## 2.3 Extracting Features from Metric Structure

Several ideas have been proposed as to how non-local properties can be extracted from the metric structure of two-dimensional silhouettes. A well known example is the distance transform which assigns each point within the silhouette its minimal distance to the boundary. In [12] the Poisson equation is solved within the boundaries. The solution indicates the average time that particles starting at a given point require to reach the boundary by diffusive motion. These two methods therefore stand in a similar relation than geodesic metric and diffusion metric. From these descriptors one can then infer information about the global structure of the objects like part decomposition and orientation of limbs.

Due to approximations our approach necessarily loses some of the descriptiveness of the full Gromov-Wasserstein

Distance. Hence, we will discuss the potential of including additional features from such descriptors to improve the quality of our prior functional while at the same time keeping it computationally feasible.

### 2.4 Heuristic Bipartite Matching

Though different in their technical origin there are other attempts to tackle matching and registration that in their practical implementation somewhat resemble our proposal and thus should be mentioned here. The shape context, introduced in [2], is a local descriptor attached to a point that consist of histograms capturing the relative distribution of the rest of the shape. A heuristic registration between two shapes is then computed based on a similarity measure of the shape contexts. For the concept of comparing local descriptors based on histograms over feature distributions see also for example [1]. Such a step to compute pairwise matching costs between points of two shapes will naturally arise in our approach in the course of making the model feasible by approximations. Unlike in prior work, however, our costs are rigorously derived from the Gromov-Wasserstein Distance leading to a general framework with favorable properties, as discussed in Sect. 1.2. The authors of [17] set out to match two graphs equipped with Euclidean metrics based on their metric and topological structure. The graphs are extracted from image data. To improve robustness local features from the images such as shape context can be included into the matching. The resulting optimization problem is combinatorial in nature but still tractable due to small tree-widths of the underlying graphical models which is owed to the restriction to Euclidean metrics.

## 3 Mathematical Background

### 3.1 Notation and Setup

In the following  $(X, d_X)$  and  $(Y, d_Y)$  are discrete, finite metric spaces with the trivial topologies (as induced by the metrics) in which all sets are open.  $\mathcal{P}(\cdot)$  will refer to the set of Borel-measures on a given space.

A triple  $(A, d_A, \mu_A)$  where  $(A, d_A)$  is a metric space and  $\mu_A \in \mathcal{P}(A)$  will be called metric measure space [13]. We will rely on this concept to describe shapes:  $A$  can be thought of as the embedding space with internal structure given by  $d_A$  and  $\mu_A$  could for example be interpreted as measure of certainty as to what regions of  $A$  belong to the shape.

For a measurable map  $f : A \rightarrow B$  and a measure  $\mu \in \mathcal{P}(A)$ ,  $f_{\#}\mu \in \mathcal{P}(B)$  will denote the *pushforward measure* of  $\mu$  via  $f$ , defined by  $(f_{\#}\mu)(\sigma_B) = \mu(f^{-1}(\sigma_B))$  for all measurable  $\sigma_B \subseteq B$ .

For two non-negative measures  $\mu_A \in \mathcal{P}(A), \mu_B \in \mathcal{P}(B)$  with equal mass,  $\mu_A(A) = \mu_B(B)$ , the set of *coupling measures* will be defined as

$$\begin{aligned} \mathcal{M}(\mu_A, \mu_B) = & \left\{ \mu \in \mathcal{P}(A \times B) : \mu(\sigma) \geq 0 \wedge \right. \\ & \mu(\sigma_A \times B) = \mu_A(\sigma_A) \wedge \\ & \mu(A \times \sigma_B) = \mu_B(\sigma_B) \\ & \forall \text{ measurable subsets } \sigma \subseteq A \times B, \\ & \left. \sigma_A \subseteq A, \sigma_B \subseteq B \right\}. \end{aligned} \tag{3.1}$$

$\mathcal{M}(\mu_A, \mu_B)$  is always non-empty as it contains at least the (normalized) product measure of  $\mu_A$  and  $\mu_B$ .

An element in  $\mu \in \mathcal{M}(\mu_A, \mu_B)$  can be understood as multivalued function between the supports of  $\mu_A$  and  $\mu_B$ , where an element  $b \in B$  is assigned to those  $a \in A$  where  $\mu(a, b) > 0$  and the value being a weight of the “strength” of the assignment. A special role play assignments that are deterministic, i.e.  $\forall a \in A$  with  $\mu_A(a) > 0$  there is precisely one element  $b \in B$  such that  $\mu(a, b) > 0$  and vice versa. The existence of such assignments will also be discussed in this paper.

We will not only be concerned with measuring deviations between shapes (for which one must optimize over some set of coupling measures) but also with shape optimization. For a non-negative measure  $\mu_A \in \mathcal{P}(A)$  the set of target measures in a measurable space  $B$  will be defined as

$$\begin{aligned} \mathcal{T}_B(\mu_A) = & \left\{ \mu_B \in \mathcal{P}(B) : 0 \leq \mu_B(\sigma_B) \leq |\sigma_B| \wedge \mu_B(B) \right. \\ & \left. = \mu_A(A) \forall \text{ measurable subsets } \sigma_B \subseteq B \right\}, \end{aligned} \tag{3.2}$$

where the constraint  $0 \leq \mu_B(\sigma_B) \leq |\sigma_B|$  ensures that  $\mu_B$  will always correspond to a relaxed indicator function. When performing shape optimization w.r.t.  $\mu_B$  for fixed  $\mu_A$  this is the feasible set for which coupling measures  $\mathcal{M}(\mu_A, \mu_B)$  exist.

### 3.2 (Gromov-)Wasserstein Distance

Due to space limitations we cannot give a full revision of the mathematical background and thus need to confine ourselves to stating the most central definitions. For a comprehensive discussion of the Wasserstein Distances see for example [29].

**Definition 3.1** (Wasserstein Distance) For a given cost function  $c : X \times Y \rightarrow \mathbb{R}$  and two non-negative measures  $\mu_X \in \mathcal{P}(X), \mu_Y \in \mathcal{P}(Y)$  with equal mass  $\mu_X(X) = \mu_Y(Y)$  define the Wasserstein Distance as follows:

$$D_W(c, \mu_X, \mu_Y) = \inf_{\mu \in \mathcal{M}(\mu_X, \mu_Y)} J_W(c, \mu) \tag{3.3a}$$

with

$$J_W(c, \mu) = \sum_{x,y} \mu(x, y) c(x, y). \tag{3.3b}$$

It is a common setup to choose  $X = Y$  and investigate  $D_W(d^p, \mu_1, \mu_2)^{1/p}$ ,  $1 \leq p < \infty$  where  $d$  is a metric on  $X$ .

*Remark 3.1* If  $c(x, y) = \tilde{c}(x, y) + \Delta c$  where  $\Delta c$  is a constant with respect to  $x$  and  $y$  then  $D_W(c, \mu_X, \mu_Y) = D_W(\tilde{c}, \mu_X, \mu_Y) + \Delta c \cdot \mu_X(X)$ .

For an introduction to the Gromov-Hausdorff Distance and a stepwise motivation and development of the Gromov-Wasserstein Distance see for example [18]. For our purposes the following definition is sufficient:

**Definition 3.2** (Gromov-Wasserstein Distance) For two non-negative measures  $\mu_X \in \mathcal{P}(X)$ ,  $\mu_Y \in \mathcal{P}(Y)$  with equal mass  $\mu_X(X) = \mu_Y(Y)$  define the Gromov-Wasserstein Distance as follows:

$$D_{GW}(\mu_X, \mu_Y)^p = \inf_{\mu \in \mathcal{M}(\mu_X, \mu_Y)} J_{GW}(\mu) \tag{3.4a}$$

with

$$J_{GW}(\mu) = \sum_{x,x',y,y'} \mu(x, y) \mu(x', y') \times |d_X(x, x') - d_Y(y, y')|^p \tag{3.4b}$$

with  $1 \leq p < \infty$ .

We will denote the explicit dependence of  $p$  only in the rare cases where values for different  $p$  are compared.  $1 \leq p < \infty$  is always assumed from now on.

There is a fundamental difference between the Wasserstein Distance and the Gromov-Wasserstein Distance: The former assigns costs to each transport assignment from  $X$  to  $Y$  independently given by  $c(x, y)$ , the latter to *pairs of transport assignments* depending on the function  $\Gamma(x, x', y, y') = |d_X(x, x') - d_Y(y, y')|^p$ , thus making it computationally much more complex but at the same time more suitable for the implementation of invariances: Only the relative position of assignments will matter.

### 3.3 Weighted Bipartite Matching

We would like to interpret the optimal coupling measures that arise from computing the similarity function of two mm-spaces as assignment between the two shapes. Naturally a deterministic assignment is easiest to interpret. Thus we will now provide some mathematical background that will later on allow us to rewrite our optimization problems in a suitable way to prove that deterministic solutions exist.

For two finite sets  $X, Y$  let  $G(V; E) = G(X, Y; E)$  the bipartite graph with disjoint vertex sets  $X, Y$  and edges connecting each vertex  $x \in X$  with all vertices  $y \in Y$ . We assume  $|Y| \geq |X|$  (concerning the plausibility of this assumption see also Remark 4.1). The set of neighbors  $N(X')$  for any subset  $X' \subseteq X$  then satisfies  $|N(X')| \geq |X'|$ , implying existence of a matching covering all points of  $X$  (Theorem of Hall [15, Theorem 10.3]).

Given some weights  $w \in \mathbb{R}_+^{|E|}$  assigned to the edges  $E$ , the *maximum matching problem* asks for a subset  $E' \subset E$  of *non-incident* edges, represented by an edge indicator vector  $z \in \{0, 1\}^{|E|}$ , that maximizes the corresponding weight  $\sum_{e \in E'} w_e$ , i.e. it solves

$$\max_{z \in \mathbb{R}_+^{|E|}} \langle w, z \rangle, \quad \sum_{e \in E(v)} z_e \leq 1 \quad \forall v \in V.$$

In terms of the incidence matrix  $A \in \{0, 1\}^{|V| \times |E|}$  of the graph  $G$ , the problem reads

$$\max_z \langle w, z \rangle \quad \text{s.t.} \quad z \geq 0, \quad Az \leq \mathbb{1}_{|V|}. \tag{3.5}$$

According to the theorem of Hoffman and Kruskal [15, Theorem 5.19], there will be an integral solution  $z$ ,  $z_e \in \{0, 1\}$ ,  $\forall e \in E$ , because the incidence matrix  $A$  of a bipartite graph is totally unimodular [15, Theorem 5.24] and the vector on the r.h.s. is integral.

For further reference, we detail the structure of  $A$ . We order the vertex set  $V = X \cup Y$  by adjoining the linearly ordered set  $Y$  to the linearly ordered set  $X$ , and then edges  $e = xy \in E$ ,  $x \in X$ ,  $y \in Y$ , by running through the set  $X$  for each  $y \in Y$ . Then  $A$  reads

$$A = \begin{pmatrix} \mathbb{1}_{|Y|}^\top \otimes I_{|X|} \\ I_{|Y|} \otimes \mathbb{1}_{|X|}^\top \end{pmatrix} \in \{0, 1\}^{(|X|+|Y|) \times |X||Y|}, \tag{3.6}$$

where  $\otimes$  denotes the Kronecker product [28].

### 3.4 The Quadratic Assignment Problem

Let

$$\text{Perm}(n) = \{ \mu \in \{0, 1\}^{n \times n} : \mu \mathbb{1}_n = \mathbb{1}_n, \mu^\top \mathbb{1}_n = \mathbb{1}_n \}$$

be the set of  $n \times n$  dimensional permutation matrices.

Consider now the Gromov-Wasserstein distance as introduced in Definition 3.2,

$$J_{GW}(\mu) = \sum_{x,x',y,y'} \Gamma(x, y, x', y') \mu(x, y) \mu(x', y'),$$

$$\text{with } \Gamma(x, y, x', y') = |d_X(x, x') - d_Y(y, y')|^p.$$

for the following specific discrete case: Let  $|X| = |Y|$ ,  $\mu_X, \mu_Y$  be the counting measures on  $X, Y$  and fix  $p = 2$ . Then, by confining the feasible set to the permutation matrices  $\text{Perm}(|X|)$  this becomes an instance of a problem class, known as the *Quadratic Assignment Problem (QAP)* [5].

It corresponds to the problem of matching weighted graphs  $X, Y$  having the same number of nodes. Let  $A, B$  denote the weighted adjacency matrices of  $X$  and  $Y$ , respectively, with entries  $A_{x,x'} = d_X(x, x')$ ,  $B_{y,y'} = d_Y(y, y')$  then we find

$$\begin{aligned} & \inf_{\mu \in \text{Perm}(|X|)} \sum_{x,x',y,y'} \Gamma(x, y, x', y') \mu(x, y) \mu(x', y') \\ &= \inf_{\phi} \sum_{x,x'} \Gamma(x, \phi(x), x', \phi(x')) \end{aligned}$$

(where we can represent the set of permutation matrices by the set of bijective assignments  $\phi : X \leftrightarrow Y$ )

$$\begin{aligned} &= \inf_{\phi} \sum_{x,x'} (d_X^2(x, x') + d_Y^2(\phi(x), \phi(x')) \\ &\quad - 2d_X(x, x')d_Y(\phi(x), \phi(x'))) \\ &= \inf_{\mu \in \text{Perm}(|X|)} (c - 2\langle A, \mu B \mu^\top \rangle), \end{aligned}$$

where the constant  $c$  collects the first two terms that do not depend on the variation of  $\phi$  and  $\mu$ , respectively. The resulting objective, for general  $A, B$  defines a QAP problem.

The QAP problem belongs to the most difficult combinatorial problems. An established benchmark library along with ground truth (global optima) exists [6] for problems whose size is considered as large if the number of nodes exceeds say  $|X| = 50$ . Furthermore, a hierarchy of relaxation bounds has been established ranging from simple spectral approaches to advanced and computationally expensive semidefinite relaxations—cf. [24] and references therein.

### 4 Approximate Gromov-Wasserstein Distance

Having briefly surveyed the necessary mathematical background we will now develop our own contributions.

If computational complexity was not an issue, a potential energy functional for variational image segmentation could be

$$E(\mu_Y) = E_0(\mu_Y) + D_{\text{GW}}(\mu_X, \mu_Y) \tag{4.1}$$

with two mm-spaces  $(X, d_X, \mu_X)$  and  $(Y, d_Y, \mu_Y)$ .  $(X, d_X, \mu_X)$  will play the role of prior knowledge by representing a prototype of the shape that we are after. It will therefore be referred to as *template*.  $(Y, d_Y, \mu_Y)$  will describe the image and a segmentation proposal therein. The

function  $E_0$  will contain other typical components of a segmentation functional (e.g. an appearance model). The density function of  $\mu_Y$  can be interpreted as relaxed indicator function thus making the shape representation by mm-spaces compatible to the region representation by indicator functions. So all terms in the functional will be functions of  $\mu_Y$  and not require any representation conversion.

*Remark 4.1* In the course of this paper we will assume  $|Y| \geq |X|$ . It can be seen here that this is virtually no restriction for the application of shape segmentation:  $(Y, d_Y)$  represents the whole image of which the object we are looking for, its shape described by  $(X, d_X, \mu_X)$  and its location given by  $\mu_Y$ , only takes up a fraction (if the object was larger than the image, we could not make out its shape anyway).

Unfortunately computational complexity is in fact a crucial issue. Computing the value of  $D_{\text{GW}}$  requires solving a non-convex optimization problem and it is non-convex as a function of  $\mu_Y$ , thus rendering the functional (4.1) unfeasible. In the following we will propose two suitable approximations to overcome these obstacles while keeping the favorable properties of the Gromov-Wasserstein Distance like geometric invariance. We show that for a linear appearance term  $E_0(\mu_Y)$  the resulting functional is a modification of the Wasserstein functional with relaxed  $Y$ -marginal constraints. We prove existence of optimal binary  $\mu_Y$  with deterministic optimal coupling measures that provide a bijection between the template  $\mu_X$  and the optimal segmentation region indicated by  $\mu_Y$ .

It should be noted at this point that the shape prior proposed here does not yet contain a model for non-isometric shape variations. We feel however that this is something that can be built on top of the current approach when the more fundamental problems of convexity and isometry invariance have been overcome.

Major properties of our approach, including geometric invariance, absence of the initialization problem, full correspondence, significant noise resistance compared to appearance model alone, will be illustrated in Sect. 5.

#### 4.1 Linear Approximation

Computing the Gromov-Wasserstein Distance implies solving a non-convex quadratic problem which is unfeasible in high-dimensional spaces. Here we discuss a way of obtaining an approximate solution by linearizing the functional properly.

**Definition 4.1** (Linear Approximation of Gromov-Wasserstein Distance)

$$D_1(\mu_X, \mu_Y) = \inf_{\mu \in \mathcal{M}(\mu_X, \mu_Y)} J_1(\mu_X, \mu_Y, \mu) \tag{4.2a}$$

with

$$J_1(\mu_X, \mu_Y, \mu) = \sum_{x,y} \mu(x, y) c_1(\mu_X, \mu_Y; x, y) \tag{4.2b}$$

where

$$c_1(\mu_X, \mu_Y; x, y) = \inf_{\mu' \in \mathcal{M}(\mu_X, \mu_Y)} \sum_{x',y'} \mu'(x', y') |d_X(x, x') - d_Y(y, y')|^p \tag{4.2c}$$

$$= D_W(|d_X(x, \cdot) - d_Y(y, \cdot)|^p, \mu_X, \mu_Y). \tag{4.2d}$$

Note that  $D_1(\mu_X, \mu_Y) = D_W(c_1(\mu_X, \mu_Y; \cdot, \cdot), \mu_X, \mu_Y)$ . It can easily be seen that this provides a lower bound to the exact Gromov-Wasserstein Distance.

**Proposition 4.1**

$$D_1(\mu_X, \mu_Y) \leq D_{GW}(\mu_X, \mu_Y)^p.$$

*Proof* As

$$c_1(\mu_X, \mu_Y; x, y) \leq \sum_{x',y'} \mu(x', y') |d_X(x, x') - d_Y(y, y')|^p$$

for  $\mu \in \mathcal{M}(\mu_X, \mu_Y)$  one finds  $J_1(\mu_X, \mu_Y, \mu) \leq J_{GW}(\mu)$  for  $\mu \in \mathcal{M}(\mu_X, \mu_Y)$  and therefore also  $D_1(\mu_X, \mu_Y) \leq D_{GW}(\mu_X, \mu_Y)$ .  $\square$

The relaxation  $D_1$  can be interpreted as follows: For every potential assignment  $x \leftrightarrow y$  one evaluates how well the rest of the shapes can be matched with respect to the fixed assignment  $x \leftrightarrow y$ . For all pairs in  $X \times Y$  these mismatch-scores are then used as a cost function for a linear Wasserstein functional.

We will give some analytical results that allow for efficient numerical implementation of such matching problems in Sect. 4.4.

4.2 Flexible  $Y$ -Marginals

For fixed  $\mu_Y$  one can easily compute the value of  $D_1(\mu_X, \mu_Y)$ . However it is non-convex as a function of  $\mu_Y$  and thus cannot yet be used in an approach like (4.1). The non-convexity arises from the dependency of  $c_1$  on  $\mu_Y$ . We will now propose a way to estimate a static cost function with the aid of an appearance model. Let  $\Delta : \mathcal{P}(Y) \rightarrow \mathbb{R}$  be a convex function that for a given  $\mu_Y$  rates its plausibility as a segmentation based on local features (for a review on potential local features see for example [10]). Then consider the following definition:

**Definition 4.2**

$$D_2^{\lambda \cdot \Delta}(\mu_X, \mu_Y) = \inf_{\mu \in \mathcal{M}(\mu_X, \mu_Y)} J_2^{\lambda \cdot \Delta}(\mu_X, \mu_Y, \mu) \tag{4.3a}$$

with

$$J_2^{\lambda \cdot \Delta}(\mu_X, \mu_Y, \mu) = \sum_{x,y} \mu(x, y) c_2(\mu_X; \lambda \cdot \Delta; x, y) \tag{4.3b}$$

where

$$c_2(\mu_X; \lambda \cdot \Delta; x, y) = \inf_{\mu_Y \in \mathcal{T}_Y(\mu_X)} \left( \inf_{\mu' \in \mathcal{M}(\mu_X, \mu_Y)} \sum_{x',y'} \mu'(x', y') \cdot |d_X(x, x') - d_Y(y, y')|^p + \lambda \cdot \Delta(\mu_Y) \right) \tag{4.3c}$$

$$= \inf_{\mu_Y \in \mathcal{T}_Y(\mu_X)} \left( D_W(|d_X(x, \cdot) - d_Y(y, \cdot)|^p, \mu_X, \mu_Y) + \lambda \cdot \Delta(\mu_Y) \right). \tag{4.3d}$$

Compared to  $D_1$  the relaxation of  $D_2$  goes one step further: For a hypothetical assignment  $x \leftrightarrow y$  the best potential assignment of the rest of the shapes is sought-after. But now  $\mu_Y$  is no longer fixed but is optimized over while taking the appearance model into account. For  $D_1$  non-convexity is removed by replacing multiple occurrences of  $\mu$  by different variables, in  $D_2$  this step is extended to  $\mu_Y$ . The flaw of these relaxations is that after optimizing the different variables need no longer be consistent. On the other hand, this achieves convexity w.r.t.  $\mu_Y$  which is vital for application in variational frameworks. Also,  $D_2$  gives a lower bound for the Gromov-Wasserstein distance on the basis of Proposition 4.1 and the following Proposition:

**Proposition 4.2**

$$D_2^{\lambda \cdot \Delta}(\mu_X, \mu_Y) \leq D_1(\mu_X, \mu_Y) + \lambda \cdot \mu_X(X) \cdot \Delta(\mu_Y)$$

*Proof*

$$c_2(\mu_X; \lambda \cdot \Delta; x, y) = \inf_{\mu'_Y \in \mathcal{T}_Y(\mu_X)} \left( D_W(|d_X(x, \cdot) - d_Y(y, \cdot)|^p, \mu_X, \mu'_Y) + \lambda \cdot \Delta(\mu'_Y) \right) \tag{4.4a}$$

$$\leq D_W(|d_X(x, \cdot) - d_Y(y, \cdot)|^p, \mu_X, \mu_Y) + \lambda \cdot \Delta(\mu_Y) \tag{4.4b}$$

$$\forall \mu_Y \in \mathcal{T}_Y(\mu_X) = c_1(\mu_X, \mu_Y; x, y) + \lambda \cdot \Delta(\mu_Y) \tag{4.4c}$$

and thus by virtue of Remark 3.1

$$D_2^{\lambda \cdot \Delta}(\mu_X, \mu_Y) = D_W(c_2(\mu_X; \lambda \cdot \Delta; \cdot, \cdot), \mu_X, \mu_Y) \tag{4.4d}$$

$$\leq D_W(c_1(\mu_X, \mu_Y + \lambda \cdot \Delta(\mu_Y); \cdot, \cdot), \mu_X, \mu_Y) \tag{4.4e}$$

$$= D_W(c_1(\mu_X, \mu_Y; \cdot, \cdot), \mu_X, \mu_Y) + \lambda \cdot \mu_X(X) \cdot \Delta(\mu_Y) \tag{4.4f}$$

$$= D_1(\mu_X, \mu_Y) + \lambda \cdot \mu_X(X) \cdot \Delta(\mu_Y). \tag{4.4g}$$

□

Summarizing all presented approximation steps we conclude:

$$D_2^{\lambda_2 \cdot \Delta}(\mu_X, \mu_Y) + \lambda_1 \cdot \Delta(\mu_Y) \tag{4.5a}$$

$$\leq D_1(\mu_X, \mu_Y) + (\lambda_1 + \mu_X(X) \cdot \lambda_2) \cdot \Delta(\mu_Y) \tag{4.5b}$$

$$\leq D_{GW}(\mu_X, \mu_Y) + (\lambda_1 + \mu_X(X) \cdot \lambda_2) \cdot \Delta(\mu_Y) \tag{4.5c}$$

The approximations (4.5c) → (4.5b) → (4.5a) are necessary because one cannot even compute the value of (4.5c), and (4.5b) is still non-convex in  $\mu_Y$ .

Note that this sequence of bounds holds for any convex appearance model  $\Delta$ .

### 4.3 Including Appearance and Unique Shape Matching

In this section we combine the presented shape prior functional with a linear appearance model analogous to (4.1) to illustrate the potential application in variational image segmentation. For a linear appearance model we prove two favorable properties: Existence of binary optimal  $\mu_Y$  and for them existence of optimal deterministic couplings  $\mu \in \mathcal{M}(\mu_X, \mu_Y)$ . This allows the support of the optimal  $\mu_Y$  to be interpreted as segmented foreground region and the coupling as a bijection between the template and the foreground.

First, with the aid of the mathematical background given in Sect. 3.3, we extend the convex relaxation approach for the linear assignment problem to partial assignments between sets with unequal cardinality and then rephrase the joint optimization of  $D_W(c, \mu_X, \mu_Y)$  and  $\Delta(\mu_Y)$  with respect to  $\mu_Y$  for a given cost function  $c$  to match the form of the problem.

**Partial Weighted Bipartite Matching** It is well known that generic solutions to the linear assignment problem

$$\min_{\mu \in \mathbb{R}^{n \times n}} \langle c, \mu \rangle \quad \text{s.t.} \quad \mu \geq 0, \quad \mu \mathbb{1}_n = \mathbb{1}_n, \quad \mu^\top \mathbb{1}_n = \mathbb{1}_n, \quad \mu^\top \mathbb{1}_n = 1 \tag{4.6}$$

for some  $c \in \mathbb{R}_+^{n \times n}$  correspond to permutation matrices as extrem points of the feasible set of doubly stochastic matrices (Birkhoff-von-Neumann Theorem, [15, Corollary 11.3]), hence constitute a one-to-one mapping between  $\{1, \dots, n\}$  and itself.

We next consider the partial assignment problem with upper bound constraint

$$\min_{\mu \in \mathbb{R}^{m \times n}} \langle c, \mu \rangle \quad \text{s.t.} \quad \mu \geq 0, \quad \mu \mathbb{1}_n = \mathbb{1}_m, \quad \mu^\top \mathbb{1}_m \leq \mathbb{1}_n, \tag{4.7}$$

with  $m \leq n$ .

**Proposition 4.3** *There are integral solutions  $\bar{\mu}$  to the linear program (4.7):  $\bar{\mu} \in \{0, 1\}^{m \times n}$ .*

*Proof* We show that the constraints (4.7) forms a linear system whose matrix has the structure (3.6). Using the vec operator [28] that stacks column vectors when applied to a matrix, and the corresponding relationship for matrices  $B, C, X$ ,

$$\text{vec}(CXB^\top) = (B \otimes C) \text{vec}(X),$$

we have

$$\text{vec}(\mu \mathbb{1}_n) = \text{vec}(I_m \mu \mathbb{1}_n) = (\mathbb{1}_n^\top \otimes I_m) \text{vec}(\mu) = \mathbb{1}_m, \tag{4.8a}$$

$$\text{vec}(\mathbb{1}_m^\top \mu) = \text{vec}(\mathbb{1}_m^\top \mu I_n) = (I_n \otimes \mathbb{1}_m^\top) \text{vec}(\mu) \leq \mathbb{1}_n. \tag{4.8b}$$

$\text{vec}(\mu)$  can be identified with the edge-indicator vector  $z$  in (3.5). The left-hand side corresponds to (3.6), and the equality sign in (4.8a) restricts the feasible set to a face of the integral polyhedron  $Ax \leq \mathbb{1}$ , that is also integral. □

**Combining Shape Prior and Linear Appearance Term** Recall that we confined ourselves to  $X$  and  $Y$  being discrete finite metric spaces. The sets of measures thereon thus correspond to the vector spaces  $\mathbb{R}^{|X|}$  and  $\mathbb{R}^{|Y|}$ , dimensions indexed by elements  $x \in X$  and  $y \in Y$ . Let the *template-space*  $X$  consist only of points belonging to our sample shape with  $\mu_X$  being the counting measure on  $X$ ,  $\mu_X(x) = 1$  for all  $x \in X$ . Assume also  $|Y| \geq |X|$ , this means the image to be segmented must at least have  $|X|$  pixels, i.e. “that there is enough space for  $X$  in  $Y$ ”.

Consider now, analogous to (4.1), the optimization problem

$$\inf_{\mu_Y \in \mathcal{T}_Y(\mu_X)} E(\mu_Y) \quad \text{with} \tag{4.9a}$$

$$E(\mu_Y) = D_W(c, \mu_X, \mu_Y) + \lambda \cdot \Delta(\mu_Y)$$

where  $D_W$  will become the shape prior functional for proper choice of  $c$ . For the appearance model  $\Delta$  we choose

$$\Delta(\mu_Y) = \sum_y f(y) \mu_Y(y) = \langle f, \mu_Y \rangle \tag{4.9b}$$

where  $f(y) > -\infty$  gives the affinity of the pixel  $y$  to be part of the foreground, based on local features ( $f(y) < 0 \Rightarrow y$  tends to be part of the foreground).

**Proposition 4.4** *Given  $\mu_X(x) = 1$  for all  $x \in X$ ,  $|Y| \geq |X|$ , a cost function  $c : X \times Y \rightarrow \mathbb{R}$  bounded from below and a function  $\Delta : \mathcal{P}(Y) \rightarrow \mathbb{R}$  bounded from below, as defined by (4.9b), the problem defined in (4.9a) is equivalent to solving a partial assignment problem of the form (4.7).*

*Proof* Note first that

$$\inf_{\mu'_Y \in \mathcal{T}_Y(\mu_X)} g(\mu'_Y) = \inf_{\mu'_Y \in \mathcal{T}_Y(\mu_X)} \inf_{\mu \in \mathcal{M}(\mu_X, \mu'_Y)} g(\mu(X \times \cdot))$$

for any function  $g$  since the inner optimization over  $\mathcal{M}(\mu_X, \mu'_Y)$  is trivial as all  $\mu$  therein have the same  $Y$ -marginal on which only the function  $g$  depends here. This double optimization can then again be rewritten as

$$= \inf_{\mu \in \mathcal{T}\mathcal{M}_Y(\mu_X)} g(\mu(X \times \cdot))$$

where we define the feasible set

$$\begin{aligned} \mathcal{T}\mathcal{M}_Y(\mu_X) &= \bigcup_{\mu'_Y \in \mathcal{T}_Y(\mu_X)} \mathcal{M}(\mu_X, \mu'_Y) \\ &= \{ \mu \in \mathcal{P}(X \times Y) : \mu(v) \geq 0 \wedge \mu(v_X \times Y) = \mu_X(v_X) \\ &\quad = |v_X| \wedge \mu(X \times v_Y) \leq |v_Y| \\ &\quad \forall \text{ measurable subsets } v \subseteq X \times Y, v_X \subseteq X, v_Y \subseteq Y \} \end{aligned} \tag{4.10a}$$

as the set of all possible couplings to  $\mu_X$ .

This allows us to rewrite (4.9a) as

$$\begin{aligned} &\inf_{\mu_Y \in \mathcal{T}_Y(\mu_X)} D_W(c, \mu_X, \mu_Y) + \lambda \cdot \Delta(\mu_Y) \\ &= \inf_{\mu \in \mathcal{T}\mathcal{M}_Y(\mu_X)} \sum_{x,y} c(x, y)\mu(x, y) + \lambda \cdot \Delta(\mu(X \times \cdot)). \end{aligned} \tag{4.10b}$$

Plugging in the choice of  $\Delta$  yields

$$= \inf_{\mu \in \mathcal{T}\mathcal{M}_Y(\mu_X)} \sum_{x,y} c(x, y)\mu(x, y) + \lambda \sum_{x,y} \mu(x, y) \cdot f(y) \tag{4.10c}$$

$$= \inf_{\mu \in \mathcal{T}\mathcal{M}_Y(\mu_X)} \sum_{x,y} (c(x, y) + \lambda \cdot f(y))\mu(x, y). \tag{4.10d}$$

We can rewrite the feasible set as

$$\begin{aligned} \mathcal{T}\mathcal{M}_Y(\mu_X) &= \{ \mu \in \mathbb{R}^{|X| \times |Y|} : \mu \geq 0 \wedge \mu \mathbb{1}_{|Y|} = \mathbb{1}_{|X|} \\ &\quad \wedge \mu^\top \mathbb{1}_{|X|} \leq \mathbb{1}_{|Y|} \} \end{aligned} \tag{4.10e}$$

where we have now used the equivalence of  $\mathcal{P}(\cdot)$  and  $\mathbb{R}^{|\cdot|}$  for  $X, Y$  and the product space and rephrased the conditions in vector notation. Now the equivalence of (4.10d), (4.10e)

and (4.7) is manifest with an appropriate edge weight vector  $c(x, y) + \lambda \cdot f(y)$ .

The edge weights implied by (4.10d) may not be non-negative but by virtue of the assumptions they are bounded from below. Hence they can be made non-negative by a constant shift which does not change the minimizing set.  $\square$

This proposition holds for any bounded cost function  $c$  and thus applies to  $c_2$  as defined in (4.3c). This implies that for a linear  $\Delta$  the functional (4.5a) has a binary optimizer and a respective deterministic coupling.

#### 4.4 Radial distribution comparison

For large metric spaces the computation of  $c_1(\mu_X, \mu_Y; x, y)$  or  $c_2(\mu_X, \lambda \cdot \Delta; x, y)$  for all  $x, y$  can be quite costly. A reformulation of  $c_1$  will now be derived that only depends on the “radial” mass distributions of  $\mu_X$  and  $\mu_Y$  relative to  $x$  and  $y$  respectively. Whatever the internal structure of  $X$  and  $Y$  might be,  $c_1$  can be computed by solving a mass transport problem between two subsets of the real line. A similar simplification exists for  $c_2$ . This will give a clear insight about the nature of the first relaxation step and be particularly important for efficient numerical implementation.

The results of this section are based on the following Proposition:

**Proposition 4.5** *For two discrete sets  $S_X$  and  $S_Y$  and two measurable maps  $\phi_X : X \rightarrow S_X, \phi_Y : Y \rightarrow S_Y$  denote by  $\phi$  the product map  $\phi(x, y) = (\phi_X(x), \phi_Y(y))$ . Then one finds*

$$\phi_{\#}\mathcal{M}(\mu_X, \mu_Y) = \mathcal{M}(\phi_{X\#}\mu_X, \phi_{Y\#}\mu_Y).$$

*Proof* For any  $\mu \in \mathcal{M}(\mu_X, \mu_Y)$  get

$$\begin{aligned} (\phi_{\#}\mu)(\sigma) &= \mu(\phi^{-1}(\sigma)) \geq 0 \\ (\phi_{\#}\mu)(\sigma_{S_X} \times S_Y) &= \mu(\phi_X^{-1}(\sigma_{S_X}) \times Y) \\ &= \mu_X(\phi_X^{-1}(\sigma_{S_X})) = (\phi_{X\#}\mu_X)(\sigma_{S_X}) \end{aligned}$$

and analogous

$$(\phi_{\#}\mu)(X \times \sigma_{S_Y}) = (\phi_{Y\#}\mu_Y)(\sigma_{S_Y})$$

for all measurable  $\sigma \subseteq S_X \times S_Y, \sigma_{S_X} \subseteq S_X, \sigma_{S_Y} \subseteq S_Y$ . Thus  $\phi_{\#}\mathcal{M}(\mu_X, \mu_Y) \subseteq \mathcal{M}(\phi_{X\#}\mu_X, \phi_{Y\#}\mu_Y)$ . For reasons of readability we have moved the step to show that  $\mathcal{M}(\phi_{X\#}\mu_X, \phi_{Y\#}\mu_Y) \subseteq \phi_{\#}\mathcal{M}(\mu_X, \mu_Y)$  and thus the two sets are in fact equal to the Appendix.  $\square$

This can be applied to simplify the computation of

$$c_1(\mu_X, \mu_Y; x, y) \quad \text{and} \quad c_2(\mu_X, \lambda \cdot \Delta; x, y)$$

as defined in (4.2c), (4.3c). Several new symbols will be introduced which depend on two non-negative measures  $\mu_X \in \mathcal{P}(X)$ ,  $\mu_Y \in \mathcal{P}(Y)$  and two elements  $x \in X$ ,  $y \in Y$ . We will consider everything that follows for a fixed choice of  $\mu_X, \mu_Y, x, y$  and for the sake of legibility will not always denote the dependence on this choice.

Let

$$\begin{aligned} L_X &= \{d_X(x, x') : x' \in X\} \quad \text{and} \\ L_Y &= \{d_Y(y, y') : y' \in Y\} \end{aligned} \tag{4.11a}$$

be the (discrete) sets of appearing distances in  $X$  and  $Y$  relative to the elements  $x$  and  $y$ . Let  $\Pi_X : X \rightarrow L_X$ ,  $\Pi_X(x') = d_X(x, x')$  and  $\Pi_Y : Y \rightarrow L_Y$ ,  $\Pi_Y(y') = d_Y(y, y')$  be the corresponding maps onto these sets and denote by

$$\Pi = \Pi_X \times \Pi_Y : X \times Y \rightarrow L_X \times L_Y \tag{4.11b}$$

their product.

For two non-negative measures  $\mu_X \in \mathcal{P}(X)$ ,  $\mu_Y \in \mathcal{P}(Y)$  and two elements  $x \in X$ ,  $y \in Y$  define the radial mass distributions

$$\rho_{L_X} = \Pi_{X\#}\mu_X \quad \text{and} \quad \rho_{L_Y} = \Pi_{Y\#}\mu_Y. \tag{4.11c}$$

Now we express  $c_1(\mu_X, \mu_Y; x, y)$  in terms of the radial distributions  $\rho_{L_X}$  and  $\rho_{L_Y}$ :

**Corollary 4.6**  $c_1(\mu_X, \mu_Y; x, y)$  can be expressed by a comparison of the radial distributions  $\rho_{L_X}$  and  $\rho_{L_Y}$ :

$$c_1(\mu_X, \mu_Y; x, y) = D_W(|\cdot - \cdot|^p, \rho_{L_X}, \rho_{L_Y}) \tag{4.12}$$

*Proof* By virtue of Proposition 4.5, where we choose  $S_{X/Y} = L_{X/Y}$ ,  $\phi_{X/Y} = \Pi_{X/Y}$  and thus  $\Pi = \phi$ , we find

$$\Pi_{\#}\mathcal{M}(\mu_X, \mu_Y) = \mathcal{M}(\rho_X, \rho_Y).$$

So one obtains

$$\begin{aligned} c_1(\mu_X, \mu_Y; x, y) &= \inf_{\mu \in \mathcal{M}(\mu_X, \mu_Y)} \sum_{x', y'} |d_X(x, x') - d_Y(y, y')|^p \mu(x', y') \\ &= \inf_{\mu \in \mathcal{M}(\mu_X, \mu_Y)} \sum_{l_X, l_Y} |l_X - l_Y|^p (\Pi_{\#}\mu)(l_X, l_Y) \end{aligned}$$

(where the sums  $l_X, l_Y$  range over the sets of distances  $L_X, L_Y$ )

$$\begin{aligned} &= \inf_{\rho \in \Pi_{\#}\mathcal{M}(\mu_X, \mu_Y)} \sum_{l_X, l_Y} |l_X - l_Y|^p \rho(l_X, l_Y) \\ &= \inf_{\rho \in \mathcal{M}(\rho_{L_X}, \rho_{L_Y})} \sum_{l_X, l_Y} |l_X - l_Y|^p \rho(l_X, l_Y) \\ &= D_W(|\cdot - \cdot|^p, \rho_{L_X}, \rho_{L_Y}). \quad \square \end{aligned}$$

Corollary 4.6 allows for a transparent interpretation of the relaxed functional  $D_1$ : We find

$$\begin{aligned} D_1(\mu_X, \mu_Y) &= D_W(c_1(\mu_X, \mu_Y; \cdot, \cdot), \mu_X, \mu_Y) \tag{4.13a} \\ &= D_W(D_W(|\cdot - \cdot|^p, \rho_{L_X}(\mu_X, \cdot), \rho_{L_Y}(\mu_Y, \cdot)), \mu_X, \mu_Y) \tag{4.13b} \\ &= D_W(D_{W,p}, \rho_{L_X}(\mu_X, \cdot)_{\#}\mu_X, \rho_{L_Y}(\mu_Y, \cdot)_{\#}\mu_Y) \tag{4.13c} \end{aligned}$$

where  $D_{W,p}$  is the Wasserstein Distance on  $\mathcal{P}(\mathbb{R}_+)$  with respect to the  $p^{\text{th}}$  power of the Euclidean metric on  $\mathbb{R}_+$  as cost function. This means the measures  $\mu_X$  and  $\mu_Y$  are transformed into two measures of radial mass distributions in  $\mathcal{P}(\mathcal{P}(\mathbb{R}_+))$ . We equip this space with the cost function that is given by the standard  $p$ -Wasserstein distance on  $\mathcal{P}(\mathbb{R}_+)$ . That is,  $D_1$  measures the deviation in radial distributions between  $\mu_X$  and  $\mu_Y$ .

Now we discuss the reformulation of  $c_2$ :

*Remark 4.2* Consider  $\Delta$  as defined in (4.9b). Let  $(\Pi, f)_Y(y) = (\Pi_Y(y), f(y))$  and let  $F_Y = \{(\Pi, f)_Y(y') : y' \in Y\}$  be the set of pairs of distances and affinity coefficients.

Applying Proposition 4.5 with  $S_Y = F_Y$ ,  $\phi_Y = (\Pi, f)_Y$  and  $S_X = L_X$ ,  $\phi_X = \Pi_X$  as before, we find

$$(\Pi_X \times (\Pi, f)_Y)_{\#}\mathcal{M}(\mu_X, \mu_Y) = \mathcal{M}(\rho_X, (\Pi, f)_{Y\#}\mu_Y). \tag{4.14a}$$

So we obtain

$$\begin{aligned} c_2(\mu_X, \lambda \cdot \Delta; x, y) &= \inf_{\mu_Y \in \mathcal{T}_Y(\mu_X)} (D_W(|d_X(x, \cdot) - d_Y(y, \cdot)|^p, \mu_X, \mu_Y) \\ &\quad + \lambda \cdot \Delta(\mu_Y)) \\ &= \inf_{\mu_Y \in \mathcal{T}_Y(\mu_X)} \left( \inf_{\mu \in \mathcal{M}(\mu_X, \mu_Y)} \sum_{x', y'} (|d_X(x, x') - d_Y(y, y')|^p \right. \\ &\quad \left. + \lambda \cdot f(y')) \mu(x', y') \right) \\ &= \inf_{\mu_Y \in \mathcal{T}_Y(\mu_X)} \left( \inf_{\mu \in \mathcal{M}(\mu_X, \mu_Y)} \sum_{l_X, (l_Y, f_Y) \in F_Y} (|l_X - l_Y|^p \right. \\ &\quad \left. + \lambda \cdot f_Y)((\Pi_X \times (\Pi, f)_Y)_{\#}\mu)(l_X, (l_Y, f_Y)) \right) \end{aligned}$$

(where the sum  $(l_Y, f_Y)$  runs over the set of pairs  $F_Y$ )

$$\begin{aligned} &= \inf_{\mu_Y \in \mathcal{T}_Y(\mu_X)} \left( \inf_{\rho \in \mathcal{M}(\rho_X, (\Pi, f)_{Y\#}\mu_Y)} \sum_{l_X, (l_Y, f_Y) \in F_Y} (|l_X - l_Y|^p \right. \\ &\quad \left. + \lambda \cdot f_Y)\rho(l_X, (l_Y, f_Y)) \right) \\ &= \inf_{\rho \in \mathcal{S}} \sum_{l_X, (l_Y, f_Y) \in F_Y} (|l_X - l_Y|^p + \lambda \cdot f_Y)\rho(l_X, (l_Y, f_Y)) \end{aligned} \tag{4.14b}$$

with  $\mathcal{S} = \bigcup_{\mu_Y \in \mathcal{T}_Y(\mu_X)} \mathcal{M}(\rho_X, (\Pi, f)_{Y\sharp} \mu_Y)$ . This is a partial linear assignment problem on  $L_X \times F_Y$  which is a subset of  $\mathbb{R} \times \mathbb{R}^2$ .

One can show in a fashion similar to Proposition 4.5 that

$$\begin{aligned} \mathcal{S} &= \{ \rho \in \mathcal{P}(L_X \times F_Y) : \rho(\sigma) \geq 0, \rho(\sigma_{L_X \times F_Y}) \\ &= \rho_{L_X}(\sigma_{L_X}), \rho(L_X \times \sigma_{F_Y}) \leq |(\Pi, f)_Y^{-1}(\sigma_{F_Y})| \\ &\text{for all measurable } \sigma \subseteq L_X \times F_Y, \sigma_{L_X} \subseteq L_X, \\ &\sigma_{F_Y} \subseteq F_Y \}. \end{aligned} \tag{4.14c}$$

Here the parameter  $\lambda$  has a very intuitive interpretation: When comparing radial distributions  $(\lambda \cdot (f_1 - f_2))^{1/p}$  is the maximal distance  $|l_X - l_Y|$  that mass is transported to obtain a better feature match  $f_2 < f_1$ .

### 4.5 The Choice of Metric

The choice which metric to impose on the mm-spaces is crucial as to which geometric invariances should be implemented into the approach. For rigid objects the Euclidean metric is the obvious option enabling recognition of an object in any translated or rotated state. Similarly for articulated objects that may appear in different poses the geodesic metric can be applied. The shortest paths can quickly be computed by fast marching algorithms [26]. To increase robustness to topological noise the diffusion metric might be considered.

Having said this it must be pointed out that there is a fundamental issue about the geodesic (and diffusion) metric in the context of our applications: Reasonably defined the geodesic metric on  $Y$  depends on  $\mu_Y$  as the measure indicates the actual location of the shape and only within this the shortest paths are to be routed. This raises the question how the metric should deal with the facts that  $\mu_Y$  can be non-binary and changes during optimization: Will there be a threshold for  $\mu_Y(y)$  above which a point  $y$  will be considered as “path permeable”? Will points with low  $\mu_Y(y)$  contribute longer piecewise path lengths? Moreover updating  $d_Y$  with  $\mu_Y$  will certainly render the optimization problem unfeasible again. Ad hoc proposals are conceivable to tackle this. Yet we consider this problem too delicate for attempting to solve it in passing-by and leave a corresponding more thorough study for future work. In our numerical experiments we present application of the geodesic metric in a confined setup that allows to circumvent these issues and to demonstrate its potential up to the mentioned caveat.

#### 4.5.1 Metric Enhancement

In this paper we approximate the quadratic Gromov-Wasserstein Distance by a linear problem. When computing the assignment the only remaining interaction between different

pixels is via the constraints. To somewhat make up for this loss of non-locality we now present a heuristic way to incorporate additional geometric information into the matching process. In this way one can exploit a large class of additional features.

Let  $\mathcal{F}$  be a feature space (for example a set of labels) with a mismatch cost function  $c_{\mathcal{F}}$  and let  $\varphi_{X,Y} : X, Y \rightarrow \mathcal{F}$  be two functions that assign these features to the elements in  $X$  and  $Y$ . This induces an assignment cost function  $c_{\mathcal{F}}(\varphi_X(\cdot), \varphi_Y(\cdot))$  on  $X \times Y$ .

One can interpret the linear appearance model  $\Delta$  (4.9b) as a degenerate example of such a function where  $\mathcal{F} = \mathbb{R}$ ,  $\varphi_Y(y) = f(y)$  and  $c_{\mathbb{R}}(\varphi_X, \varphi_Y) = \varphi_Y$  solely depends on  $\varphi_Y$ . But more sophisticated choices are at hand: consider an object that has different characteristic appearances in different regions. One might then consider an appearance term that depends not only on  $\mu_Y$  but on the coupling  $\mu$  directly and thus can incorporate such additional information:

$$\begin{aligned} \Delta(\mu) &= \sum_{x,y} f(x, y) \mu(x, y) \\ &= \sum_{x,y} c_{\mathcal{F}}(\varphi_X(x), \varphi_Y(y)) \mu(x, y) \end{aligned} \tag{4.15}$$

where  $c_{\mathcal{F}}$  determines how well the appearance feature  $\varphi_Y(y)$  found at  $y$  matches the expected appearance  $\varphi_X(x)$  at  $x$ .

The definition of  $c_2(\mu_X, \lambda \cdot \Delta; x, y)$  can be generalized to such a  $\Delta$ . The simplifications discussed in Sect. 4.4 can be applied analogously where, by means of Proposition 4.5, for the computation of the modified  $c_2$  one arrives at a relaxed optimal transport problem between subsets of  $L_X \times \mathcal{F}$  and  $L_Y \times \mathcal{F}$  with a cost function  $c((l_X, \varphi_X), (l_Y, \varphi_Y)) = |l_X - l_Y|^p + c_{\mathcal{F}}(\varphi_X, \varphi_Y)$ . Both the modified cost function  $c_2$  and  $\Delta$  can then be plugged into an approach as (4.9a) and one will find that Proposition 4.4 still holds.

It should be noted at this point, that this does not simply imply changing  $\Delta$  in the global matching according to (4.9a). Via the computation of the cost function  $c_2$  these additional features are also considered in the first term in (4.9a).

In Sect. 5 we will present two applications of this extension: the incorporation of an inhomogeneous appearance model and the usage of the distance transform as an additional feature to implicitly take into account non-local geometric information.

## 5 Numerical Examples

In this section we want to demonstrate the potential of the proposed shape prior functional combined with a linear appearance model, as discussed in the previous sections, for

variational image segmentation. We give some numerical examples to illustrate the favorable properties of our approach and also the limitations implied by the involved approximations.

The exact approach 4.1, without approximations, is considerably more general than the QAP problem, in several ways. In particular,  $|Y| \gg |X|$  is an essential relation covering the image segmentation scenario (recall Remark 4.1). As a consequence, performing ground truth experiments from the viewpoint of optimization is elusive. Our numerical experiments are merely supposed to demonstrate the extent to which invariant matching of metric measure spaces can be enforced by our convex relaxation approach to shape prior design. A systematic study of further suboptimality bounds and the application to specific segmentation problems is beyond the scope of the present paper.

Before presenting numerical results, the next section describes technical details of the implementation, in particular how computational effort can be reduced (including using the results presented in Sect. 4.4).

### 5.1 Implementation Details and Computational Complexity

The prior mm-spaces  $(X, d_X, \mu_X)$  were created from binary images, depicting the template shapes. All pixels with value 0 were removed from the space. The remaining pixels were equipped with the Euclidean metric and  $\mu_X$  was set to be the counting measure on these points. For a given experiment  $(Y, d_Y)$  represents the test-image grid with Euclidean metric. The function  $f$  that defines the appearance model  $\Delta$  was constructed from the gray values of the test-image.

To compute the cost function  $c_2$  for some  $(x, y)$  one needs to compute a modified mass transport problem on  $L_X \times F_Y$  (see Remark 4.2). For this  $L_X$  was approximated by a set of equally sized bins on the real line and  $F_Y$  by a set of rectangular bins on  $\mathbb{R}^2$ . While only inflicting a small discretization error this reduced the involved problem sizes by several orders of magnitude. This method becomes particularly efficient when the affinity coefficients  $f$  are binary (e.g.  $\pm 1$ , indicating unweighted preference for yes/no only) and  $F_Y$  can be approximated by two discretized real lines. Also, it is straightforward to parallelize the computation of  $c_2$  for all  $(x, y)$ .

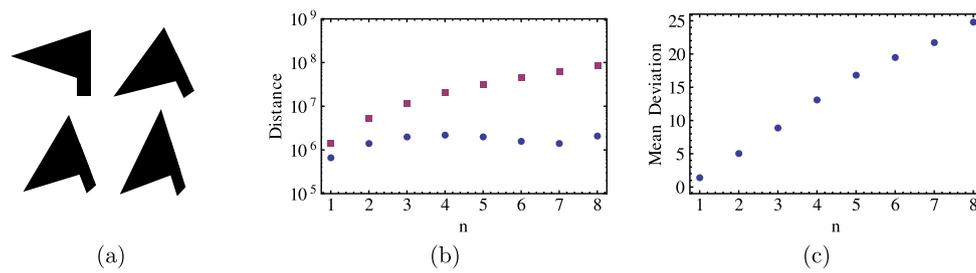
For solving the global matching between  $X$  and  $Y$  we experimented with constraining the full coupling space  $X \times Y$  to smaller subsets to keep the problem size low, while still solving the global problem. Consider the following modification to the partial assignment problem (4.7): For any  $x \in X$  include only a subset  $Y_x \subset Y$  with the lowest assignment costs. Then, for every  $x$ , add an additional overflow element  $y_{\text{of},x}$  and a corresponding variable  $\mu(x, y_{\text{of},x})$  whose assignment cost  $c(x, y_{\text{of},x})$  is chosen such that  $c(x, y_1) \leq$

$c(x, y_{\text{of},x}) \leq c(x, y_2)$  for all  $y_1 \in Y_x, y_2 \in (Y \setminus Y_x)$ . For each  $x$  the coupling value  $\mu(x, y_{\text{of},x})$  will be taken into account when computing the  $X$ -marginal constraint, but there will be no  $Y$ -constraints on any of the  $y_{\text{of},x}$ . Then for any feasible coupling  $\mu$  in the original problem that is non-zero outside of the constrained coupling set, one can create a feasible coupling in the modified problem with non-zero overflow variables, which will yield a lower score. This implies that when solving the restricted partial assignment problem and one gets an optimizer where  $\mu(x, y_{\text{of},x}) = 0$  for all  $x$  then one knows to have found an optimizer for the original problem with the full coupling space  $X \times Y$ . In “easy” problems this enabled us to find global minimizers while considering only  $< 5\%$  of the coupling space, “harder” problems were still generally  $< 25\%$ . In the special case  $p = 1$  it is easy to show that both  $c_{1/2}$  are Lipschitz. Then one can estimate a suitable subset of  $X \times Y$  by subsampling and lower bounds via the Lipschitz condition without scanning all possible pairs.

In the presented experiments  $|X|$  is of the order  $10^3$  and  $|Y|$  up to the order of several  $10^4$ . We have set  $p = 2$  but we did not observe a substantial change of results for other values  $p \geq 1$ .

### 5.2 Experiments and Discussion

*Approximation Quality  $D_{\text{GW}} \rightarrow D_1$*  The purpose of the first experiment is to gain an insight into the quality of the relaxation  $D_{\text{GW}} \rightarrow D_1$ , see Proposition 4.1. We take a simple shape, rotate it, distort it by non-isometric but mass preserving scalings with factors  $q^n, q^{-n}$  along the vertical and horizontal axis and then compute the optimal assignment according to  $D_1$  between the original and the distortion for various  $n > 0$ . As an estimate for ground truth we use the assignment induced by the distortion map  $(r_1, r_2) \rightarrow (q^n \cdot r_1, q^{-n} \cdot r_2)$ . The results are summarized in Fig. 2. For low distortions one can see how  $D_1$  is a good measure for increasing non-isometry, although growing slower than the functional value of the distortion map. For high  $n$  the deviation becomes more significant as  $D_1$  decreases, while the upper bound grows further. Here one can assume that the distortion map is no longer the optimal assignment and thus the estimated “ground truth” is in fact too high. There is an additional subtlety in this experiment:  $D_1$  was computed between two rasterizations of a vector graphic, one as is and one undergoing the distortion transformation. Thus even for  $n = 0$  (applying only a rotation) the two resulting metric spaces would not be isometric due to different rasterization. When estimating the ground truth this rasterization cannot be taken into account, since it is unclear how to match the two rasterized graphics. It has thus been estimated on the vector graphics level. The fact that such problems appear even for such simple shapes is a clear indicator of how hard it is to solve the full quadratic problem.



**Fig. 2** Linearization of  $D_{GW}$ : (a) top-left to bottom right: original shape and three distortions  $q^n$  for  $n = 2, 4, 6$  and  $q = 0.95$  (dimensions:  $\approx 60 \cdot 70 \cdot (\text{length units})^2$ ). (b) Circles:  $D_1$  between original and  $q^n$ -distortion, squares: Gromov-Wasserstein functional evaluated for assignment induced by distortion map. (c) Mean metric deviation between underlying distortion assignment and assignment computed by  $D_1$  (averaged over all assigned pairs). For small  $n \leq 4$ ,  $D_1$  grows with increasing metric distortion, although slower than the estimated “true” Gromov-Wasserstein distance. For  $n > 4$ ,  $D_1$  first starts to decrease a little, before eventually growing again. The assignment com-

puted by  $D_1$  is (up to rasterization errors on the pixel level) identical to the underlying distortion transformation for  $n = 1$ , deviation grows with increasing non-isometry. From  $n \leq 4$  we learn that  $D_1$  is a lower bound that grows with increasing level of non-isometry, which is a favorable property for the functional. For  $n > 4$  presumably the distortion map itself is no longer the best distance-preserving assignment between the two shapes and thus the estimated ground truth value is in fact too high (note how the triangle transforms from being horizontally elongated to vertically elongated). This is an illustration for the difficulty of obtaining ground truth data and the need for relaxations

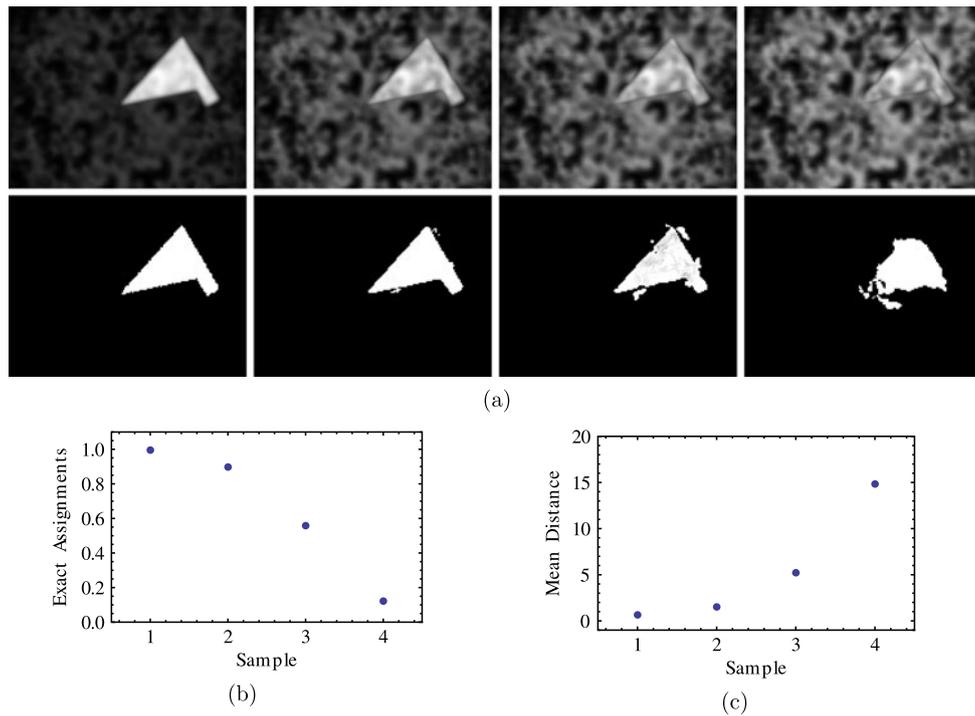
*Linear Appearance Term and Matching* We want to go beyond measuring the distance between two fixed shapes and perform shape optimization according to the two criteria shape and appearance. The affinity coefficients  $f(y)$  for the linear appearance model  $\Delta$  are generated from a grayscale image: First an imprint of the template shape is created somewhere in the image and then noise is added in different levels. For low noise level the shape restoration works perfectly, without any prior knowledge or the need for proper initialization. Increasing noise will cause an increasing number of assignments to become inaccurate, while remaining roughly correct. Eventually only the coarse location, but no longer the contours of the shape can be recovered (see Fig. 3). In Fig. 4 the influence of the global weighting parameter  $\lambda$  (cf. (4.9a)) is demonstrated: For high values the optimal  $\mu_Y$  is determined locally by the appearance coefficients, for low values the shape prior becomes more dominant and leads to a more accurate restoration of the original contours.

*Binary Appearance Term* If the appearance coefficients  $f$  are binary, for example  $\pm 1$ , indicating unweighted preference for fore- or background, one can interpret the region  $\{y \in Y: f(y) < 0\}$  as a noisy foreground proposal and extract additional information from this region. An example for this would be the distance transform (for a noise resistant alternative see [12]). In a scenario with strong appearance classifier one might assume that this preliminary foreground region already resembles the true sought-after region. Thus the distance transform might yield similar values in corresponding places of the template and the image and can therefore be used as a matching criteria as discussed in Sect. 4.5.1, with  $\varphi_X, \varphi_Y$  being the distance transformations,  $\mathcal{F}$  the real line and  $c_{\mathcal{F}}$  for example the  $L_1$  norm. The experiment presented in Fig. 5 has been specifically designed

to demonstrate how the local  $Y$ -marginal estimation during the computation of  $c_2$  via the appearance model can fail: For points near the center of the cross of the template the outer regions of the “blob” on the right of the input image appear more suitable than the center of the actually corresponding cross, where one “arm” has been shortened. By including the additional information encoded in the distance transform this mismatches can be fixed.

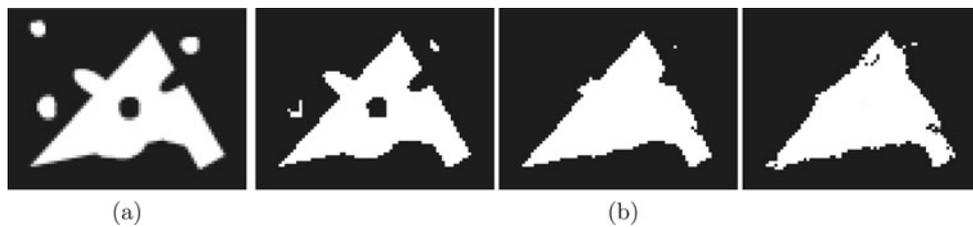
The setup of Fig. 5 is also well suited to discuss the implications of the convexity of the functional. A major advantage is the independence of initialization. An approach based on active contours would, if initialized around the blob, be stuck on the right hand side no matter how bad the matching cost will be. The contour could not leave the blob and move through an area without any mass (and thus without reasonable gradient information). The proposed approach does not suffer from this issue (up to the discussed level of confusion caused by approximations of the GW-functional).

The question then arises how the optimal coupling measure looks like if there are multiple (approximately) equivalent optimal solutions. Up to rasterization artifacts there is no preferred choice how to map the template cross onto the input: Eight orientations (rotations, reflections) are equally possible. Each corresponding to one local extremum for an active contour approach that one would consider as valid solution. For the proposed approach such symmetries cause degeneration of the space of optimal couplings, making a whole facet of the feasible polytop extremal. Interior point methods then usually do not lead to integer solutions. Integer solutions exists and applying a simplex algorithm will produce one. Some may correspond to one of the eight possible assignments, some may be highly discontinuous (meaning that adjacent pixels are assigned to very different target pixels), but from the viewpoint of the functional they are all equivalent and the choice is arbitrary.



**Fig. 3** Appearance model and noise: (a) top row: appearance coefficients  $f$  for the image (bright indicating higher foreground-affinity), noise increasing from left to right. Bottom row: corresponding optimal  $\mu_Y$  according to  $D_2$ , brightness indicating mass density. ( $X, d_X, \mu_X$ ) as in Fig. 2,  $Y$ -image dimensions:  $160 \times 120$ (length units)<sup>2</sup>. (c) Fraction of computed assignments that is closer than 3 pixels to the underlying transformation (c) mean metric deviation between true underlying transformation and assignment computed by  $D_2$  (averaged over all assigned pairs). For low noise levels the appearance model in combination with the metric information can compensate for noisy

appearance data and correctly restore the original shape. Although the location and orientation of the shape within  $Y$  is not known a priori, the isometry invariant approach can extract the correct transformation. With higher noise levels the number of assignments that is led astray increases, starting to erode the shape contours, although the majority remains correct. Finally, for very high noise levels the relaxation breaks down completely and hardly any of the assignments are correct, indicating that the local  $Y$ -marginal estimation during the computation of  $c_2$  is no longer powerful enough. Here a more global approach would be required

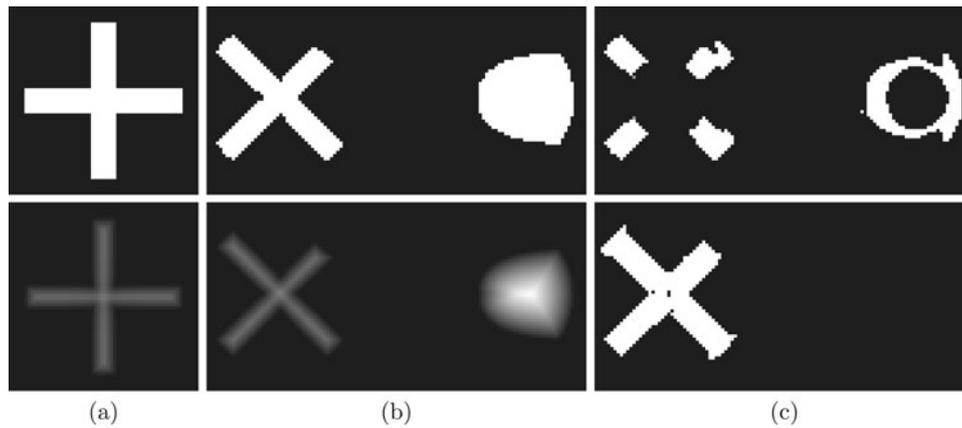


**Fig. 4** Influence of the global regularization parameter  $\lambda$ : (a) appearance coefficients  $f$ , representing a transformed, distorted version of the original shape (see Fig. 2). (b) From left to right: optimal  $\mu_Y$  for  $\lambda = 10^4, 10^3, 10^2$ . With decreasing  $\lambda$  the shape prior becomes more influential and pushes towards restoration of the original shape. It should be noted here, that tiny holes or jagged contours in the optimal segmen-

tation regions are not due to numerical instabilities of the optimization implementation but due to discretization artifacts. Sometimes from, the metric point of view, it is better to drop single pixels when matching two different rasterizations of the same shape. Spatial regularity of  $\mu_Y$  on the rasterization scale is not enforced by the used functional, so this does not increase the functional value

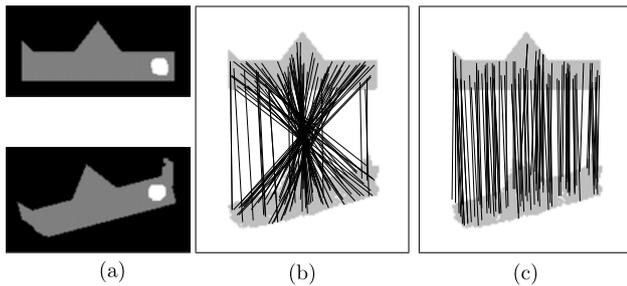
*Inhomogeneous Appearance Term* The extension discussed in Sect. 4.5.1 can also be used to incorporate an inhomogeneous appearance model where different regions of the shape are associated with different characteristic appearances. See for example Fig. 6: the shape itself is almost mirror-symmetric and in fact the noise was chosen such that the matching purely based on background  $\leftrightarrow$

“fish” confuses front and back of the schematic fish. Assume now from the underlying image data there is additional information available, like a dedicated detector for the eye. Then this can be exploited, leading to the desired effect. Also note that in both cases, corresponding to the assumed orientation of the fish, the appropriate shape is restored.



**Fig. 5** Breakdown of second approximation step, enhancement by additional features: (a) template (*top*, *white* indicates mass) and corresponding distance transform. (b) Input: binary appearance coefficients *f* (*top*, *white* indicates foreground) and distance transform. (c) *Top*: optimal  $\mu_Y$  according to  $D_2$ , *bottom*: optimal  $\mu_Y$  according to matching according to Sect. 4.5.1 with distance transform as additional feature.

In the input image, one arm of the cross has been shortened and the “blob” on the right-hand side of the input has been designed to confuse the local  $\mu_Y$  estimation during the  $c_2$  computation, thus causing faulty assignments. This demonstrates the limitations of the second approximation step. By including additional information like the distance transform according to Sect. 4.5.1 this confusion can be resolved



**Fig. 6** Inhomogeneous appearance model: (a) *top*: template with features, black indicating background, *grey*  $\rightarrow$  “body” and *white*  $\rightarrow$  “eye”; *bottom*: input image with detected features. (b) Optimal  $\mu_Y$  (*gray shading bottom*) and assignment (*black lines*, subsampling) of homogeneous appearance model: distinguish only background  $\leftrightarrow$  fish (=body & eye). Due to the approximate mirror symmetry and the noise in the image features, front and back are confused (while still “correctly” reconstructing the edges according to the mixup). (c) Optimal  $\mu_Y$  and assignment with an appearance model that penalizes the matching  $x \leftrightarrow y$  between different feature classes (see Sect. 4.5.1). The confusion between back and front is remedied

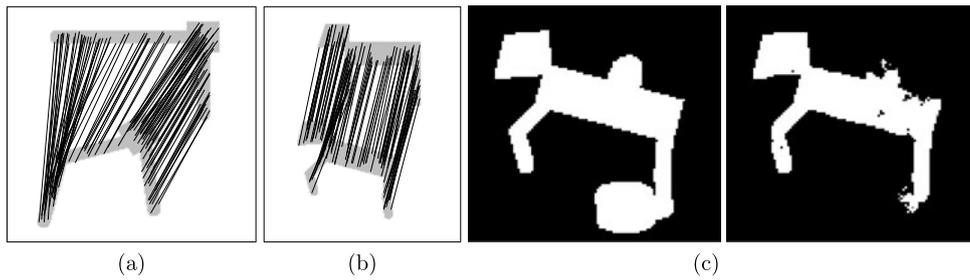
coefficients, as introduced earlier, and assume that all true foreground pixels are in fact also labeled as foreground by the appearance model. In addition some false positive detections are possible, i.e. regions in the test image that are wrongfully indicated to be foreground. The template shape and the apparent foreground region are then equipped with their respective geodesic metrics, efficiently computed by fast marching methods [26]. This requires that the false positive detections are rare enough to keep the geodesic metric of the underlying true foreground approximately unchanged. Figure 7(c) shows an image of binary appearance-coefficients with such superfluous false positive foreground labels and the computed optimal  $\mu_Y$  via  $D_2$ . Although the reconstruction is by no means perfect, the method still tends to neglect the false positive foreground-detections.

### 6 Conclusion, Outlook

In this paper we have proposed and developed a novel convex shape prior functional with potential for application in variational image segmentation and demonstrated its combination with a linear appearance term. The starting point is the Gromov-Wasserstein Distance which has been applied successfully in the field of shape recognition and classification. From this the approach inherits the ability to incorporate vast classes of geometric invariances. Since the Gromov-Wasserstein distance is computationally too complex to be computed, let alone to perform shape optimization, we proposed and discussed two successive approximation steps to overcome these two obstacles.

In combination with a linear appearance model we proved for the resulting matching problem the existence of

**Geodesic Metric and Pose Invariance** The key to recognizing the same object in different poses is to equip shapes with the geodesic metric. In Figs. 7(a) and 7(b) the geodesic metric is used to compute the optimal assignments between two pairs of objects in different poses via  $D_1$ . In Sect. 4.5 the problems were discussed that arise when one wants to port the concepts of mm-spaces to image segmentation and faces the involved shape optimization task. The estimation of a static cost function  $c_2$  can in general not be performed in a straightforward fashion. Here we demonstrate the potential of the geodesic metric for a pose invariant shape prior functional in a restricted setup where the aforementioned difficulties can be avoided. Consider binary appearance coeffi-



**Fig. 7** Geodesic metric and pose invariance: **(a)** assignment between two different poses of an object, computed by  $D_1$ . **(b)** Assignment between two different poses of a schematic “horse” via  $D_1$ . Both assignments correctly associate the corresponding parts of the objects.

**(c) Left:** modification of the lower horse from **(b)** by adding additional false positive foreground detections; **right:** optimal marginal  $\mu_Y$  for matching between the modified lower and the original upper horse via  $D_2$ . Excess detections are mostly removed at the correct locations

optimizers that imply a well defined segmentation region and a bijective assignment to the reference shape. In the final problem shape optimization and computation of the approximate shape distance are performed in a single pass. Some analytic results that concern efficient numerical implementation and help to better understand the approximations were given. Ways were proposed to counterbalance the loss due to approximations and even to process additional feature information. Key aspects of the approach were illustrated and discussed based on numerical examples.

Future work will include incorporation of the functional into a full variational segmentation framework. Application to other types of data that can be described by mm-spaces (e.g. weighted point data) and a more detailed study of the potential of the matching enhancement by additional features. Also the delicate question as to how the geodesic framework is best extended to dynamic shapes remains open.

**Acknowledgement** This work was supported by the German National Science Foundation (DFG) grant GRK 1653.

**Appendix**

*Step 2 of the Proof of Proposition 4.5* We show by construction for any  $\rho \in \mathcal{M}(\phi_{X\#}\mu_X, \phi_{Y\#}\mu_Y)$  the existence of some  $\mu \in \mathcal{M}(\mu_X, \mu_Y)$  such that  $\rho = \phi_{\#}\mu$ . For any element  $(s_X, s_Y) \in S_X \times S_Y$  construct the pre-image measure

$$\mu_{(s_X, s_Y)}(x, y) = \begin{cases} 0 & \text{if } \rho(s_X, s_Y) = 0 \vee (s_X, s_Y) \neq \phi(x, y) \\ \frac{\mu_X(x)\mu_Y(y)}{(\phi_{X\#}\mu_X)(s_X)(\phi_{Y\#}\mu_Y)(s_Y)}\rho(s_X, s_Y) & \text{else} \end{cases}$$

where this element wise definition for each  $(x, y)$  is extended to all subsets of  $X \times Y$  by

$$\mu_{(s_X, s_Y)}(\sigma) = \sum_{(x, y) \in \sigma} \mu_{(s_X, s_Y)}(x, y).$$

Now consider  $\mu = \sum_{(s_X, s_Y) \in S_X \times S_Y} \mu_{(s_X, s_Y)}$ : First verify that it is indeed contained in  $\mathcal{M}(\mu_X, \mu_Y)$ :

$$\mu(\sigma) \geq 0$$

since  $\mu(x, y) \geq 0$  for all  $(x, y)$ . Furthermore

$$\begin{aligned} \mu(\sigma_X \times Y) &= \sum_{x \in \sigma_X} \sum_{\substack{s_X, s_Y: \\ y \in Y, \phi(x, y) = (s_X, s_Y), \\ \rho(s_X, s_Y) > 0}} \frac{\mu_X(x)\mu_Y(y)}{(\phi_{X\#}\mu_X)(s_X)(\phi_{Y\#}\mu_Y)(s_Y)} \\ &\quad \times \rho(s_X, s_Y) \\ &= \sum_{x \in \sigma_X} \sum_{\substack{s_Y: \\ \rho(\phi_X(x), s_Y) > 0}} \frac{\mu_X(x)(\sum_{y: \phi_Y(y) = s_Y} \mu_Y(y))}{(\phi_{X\#}\mu_X)(\phi_X(x))(\phi_{Y\#}\mu_Y)(s_Y)} \\ &\quad \times \rho(\phi_X(x), s_Y) \\ &= \sum_{x \in \sigma_X} \sum_{\substack{s_Y: \\ \rho(\phi_X(x), s_Y) > 0}} \frac{\mu_X(x)\mu_Y(\phi_Y^{-1}(s_Y))}{(\phi_{X\#}\mu_X)(\phi_X(x))(\phi_{Y\#}\mu_Y)(s_Y)} \\ &\quad \times \rho(\phi_X(x), s_Y) \\ &= \sum_{x \in \sigma_X} \frac{\mu_X(x)}{(\phi_{X\#}\mu_X)(\phi_X(x))} \sum_{\substack{s_Y: \\ \rho(\phi_X(x), s_Y) > 0}} \rho(\phi_X(x), s_Y) \\ &= \sum_{x \in \sigma_X} \mu_X(x) = \mu_X(\sigma_X) \end{aligned}$$

and likewise

$$\mu(X \times \sigma_Y) = \mu_Y(\sigma_Y)$$

for all measurable subsets  $\sigma \subseteq X \times Y, \sigma_X \subseteq X, \sigma_Y \subseteq Y$ .

Now check whether  $\phi_{\#}\mu = \rho$ :

$$\begin{aligned} (\phi_{\#}\mu)(\sigma) &= \mu(\phi^{-1}(\sigma)) = \sum_{(x, y) \in \phi^{-1}(\sigma)} \mu(x, y) \end{aligned}$$

$$\begin{aligned}
 &= \sum_{\substack{(x,y) \in \phi^{-1}(\sigma): \\ \rho(\phi(x,y)) > 0}} \frac{\mu_X(x)\mu_Y(y)}{(\phi_{X\#}\mu_X)(\phi_X(x))(\phi_{Y\#}\mu_Y)(\phi_Y(y))} \\
 &\quad \times \rho(\phi(x,y)) \\
 &= \sum_{\substack{(s_X,s_Y) \in \sigma \\ \rho((s_X,s_Y)) > 0}} \sum_{(x,y) \in \phi^{-1}((s_X,s_Y))} \frac{\mu_X(x)\mu_Y(y)}{(\phi_{X\#}\mu_X)(s_X)(\phi_{Y\#}\mu_Y)(s_Y)} \\
 &\quad \times \rho(s_X,s_Y) \\
 &= \sum_{\substack{(s_X,s_Y) \in \sigma \\ \rho((s_X,s_Y)) > 0}} \frac{(\sum_{x \in \phi_X^{-1}(s_X)} \mu_X(x)) (\sum_{y \in \phi_Y^{-1}(s_Y)} \mu_Y(y))}{(\phi_{X\#}\mu_X)(s_X)(\phi_{Y\#}\mu_Y)(s_Y)} \\
 &\quad \times \rho(s_X,s_Y) \\
 &= \sum_{\substack{(s_X,s_Y) \in \sigma \\ \rho((s_X,s_Y)) > 0}} \rho(s_X,s_Y) = \rho(\sigma).
 \end{aligned}$$

Consequently any  $\rho \in \mathcal{M}(\phi_{X\#}\mu_X, \phi_{Y\#}\mu_Y)$  is also contained in  $\phi_{\#}\mathcal{M}(\mu_X, \mu_Y)$  and the two sets are equal.

**References**

1. Aherne, F., Thacker, N.A., Rockett, P.: Optimal pairwise geometric histograms. In: Proc. British Machine Vision Conf. (BMVC), pp. 480–490 (1997)
2. Belongie, S., Malik, J., Puzicha, J.: Shape matching and object recognition using shape contexts. *IEEE Trans. Pattern Anal. Mach. Intell.* **24**, 509–522 (2002)
3. Bronstein, A., Bronstein, M., Kimmel, R., Mahmoudi, M., Sapiro, G.: A Gromov-Hausdorff framework with diffusion geometry for topologically-robust non-rigid shape matching. *Int. J. Comput. Vis.* **89**, 266–286 (2010)
4. Burger, M., Franek, M., Schönlieb, C.-B.: Regularised regression and density estimation based on optimal transport. *Appl. Math. Res. eXpress*, March (2012)
5. Burkard, R.E., Çela, P.P., Pitsoulis, L.: The Quadratic Assignment Problem. *Handbook of Combinatorial Optimization*. Kluwer Academic, Dordrecht (1998)
6. Burkard, R., Karisch, S., Rendl, F.: QAPLIB—a quadratic assignment problem library. *J. Glob. Optim.* **10**, 391–403 (1997)
7. Chan, T.F., Esedoglu, S., Nikolova, M.: Algorithms for finding global minimizers of image segmentation and denoising models. *SIAM J. Appl. Math.* **66**(5), 1632–1648 (2006)
8. Charpiat, G., Faugeras, O., Keriven, R.: Shape statistics for image segmentation with prior. In: *Computer Vision and Pattern Recognition (CVPR 2007)*, pp. 1–6 (2007)
9. Cremers, D., Kohlberger, T., Schnörr, C.: Shape statistics in kernel space for variational image segmentation. *Pattern Recognit.* **36**(9), 1929–1943 (2003)

10. Cremers, D., Rousson, M., Deriche, R.: A review of statistical approaches to level set segmentation: integrating color, texture, motion and shape. *Int. J. Comput. Vis.* **72**(2), 195–215 (2007)
11. Elad, A., Kimmel, R.: On bending invariant signatures for surfaces. *IEEE Trans. Pattern Anal. Mach. Intell.* **25**(10), 1285–1295 (2003)
12. Gorelick, L., Galun, M., Sharon, E., Basri, R., Brandt, A.: Shape representation and classification using the Poisson equation. *IEEE Trans. Pattern Anal. Mach. Intell.* **28**(12), 1991–2005 (2006)
13. Gromov, M.: *Metric Structures for Riemannian and Non-Riemannian Spaces*. Birkhäuser, Boston (2007)
14. Haker, S., Zhu, L., Tannenbaum, A., Angenent, S.: Optimal mass transport for registration and warping. *Int. J. Comput. Vis.* **60**, 225–240 (2004)
15. Korte, B., Vygen, J.: *Combinatorial Optimization*, 4th edn. Springer, Berlin (2008)
16. Lellmann, J., Schnörr, C.: Continuous multiclass labeling approaches and algorithms. *SIAM J. Imaging Sci.* **4**(4), 1049–1096 (2011)
17. McAuley, J.J., de Campos, T., Caetano, T.S.: Unified graph matching in Euclidean spaces. In: *Computer Vision and Pattern Recognition (CVPR 2010)*, pp. 1871–1878 (2010)
18. Mémoli, F.: Gromov-Wasserstein distances and the metric approach to object matching. *Found. Comput. Math.* **11**, 417–487 (2011)
19. Mémoli, F.: A spectral notion of the Gromov-Wasserstein distance and related methods. *Appl. Comput. Harmon. Anal.* **30**(3), 363–401 (2011)
20. Mémoli, F., Sapiro, G.: A theoretical and computational framework for isometry invariant recognition of point cloud data. *Found. Comput. Math.* **5**, 313–347 (2005)
21. Pock, T., Chambolle, A., Cremers, D., Bischof, H.: A convex relaxation approach for computing minimal partitions. In: *Computer Vision and Pattern Recognition (CVPR 2009)*, pp. 810–817 (2009)
22. Rabin, J., Peyré, G., Cohen, L.: Geodesic shape retrieval via optimal mass transport. In: *Computer Vision—ECCV 2010*. LNCS, vol. 6315, pp. 771–784. Springer, Berlin (2010)
23. Reuter, M., Wolter, F.E., Peinecke, N.: Laplace-Beltrami spectra as “shape-dna” of surfaces and solids. *Comput. Aided Des.* **38**(4), 342–366 (2006)
24. Schellewald, C., Roth, S., Schnörr, C.: Evaluation of a convex relaxation to a quadratic assignment matching approach for relational object views. *Image Vis. Comput.* **25**, 1301–1314 (2007)
25. Schmitzer, B., Schnörr, C.: Convex coupling continuous cuts and shape priors. In: *Bruckstein, A.M., ter Haar Romeny, B.M., Bronstein, A.M., Bronstein, M.M. (eds.) Scale Space and Variational Methods in Computer Vision (SSVM 2011)*. LNCS, vol. 6667, pp. 423–434. Springer, Berlin (2011)
26. Sethian, J.A.: A fast marching level set method for monotonically advancing fronts. In: *Proc. Nat. Acad. Sci.*, pp. 1591–1595 (1995)
27. Sundaramoorthi, G., Mennucci, A., Soatto, S., Yezzi, A.: A new geometric metric in the space of curves, and applications to tracking deforming objects by prediction and filtering. *SIAM J. Imaging Sci.* **4**(1), 109–145 (2011)
28. Van Loan, C.: The ubiquitous Kronecker product. *J. Comput. Appl. Math.* **123**, 85–100 (2000)
29. Villani, C.: *Optimal Transport: Old and New*. Springer, Berlin (2009)