



# Modeling Shapes with Higher-Order Graphs: Methodology and Applications

Chaohui Wang, Yun Zeng, Dimitris Samaras, Nikos Paragios

## ► To cite this version:

Chaohui Wang, Yun Zeng, Dimitris Samaras, Nikos Paragios. Modeling Shapes with Higher-Order Graphs: Methodology and Applications. Shape Perception in Human and Computer Vision: An Interdisciplinary Perspective, Sven J. Dickinson and Zygmunt Pizlo, pp.459-471, 2013. <hal-00858417>

**HAL Id: hal-00858417**

**<https://hal.archives-ouvertes.fr/hal-00858417>**

Submitted on 5 Sep 2013

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Modeling Shapes with Higher-Order Graphs: Methodology and Applications

Chaohui Wang, Yun Zeng, Dimitris Samaras and Nikos Paragios

**Abstract** Extrinsic factors such as object pose and camera parameters are a main source of shape variability and pose an obstacle to efficiently solving shape matching and inference. Most existing methods address the influence of extrinsic factors by decomposing the transformation of the source shape (model) into two parts: one corresponding to the extrinsic factors and the other accounting for intra-class variability and noise, which are solved in a successive or alternating manner. In this chapter, we consider a methodology to circumvent the influence of extrinsic factors by exploiting shape properties that are invariant to them. Based on higher-order graph-based models, we implement such a methodology to address various important vision problems, such as non-rigid 3D surface matching and knowledge-based 3D segmentation, in a one-shot optimization scheme. Experimental results demonstrate the superior performance and potential of this type of approach.

## 1 Introduction

Shape matching and inference aims at determining the correspondence between a source shape instance (or shape model) and a target shape instance (or the observed data where the target shape is embedded). It is a fundamental problem in computer vision, computer graphics, medical image analysis and has been widely investigated in numerous important applications such as 3D surface matching and reconstruction [5, 32, 12, 30, 7, 21], statistical shape modeling and knowledge-based segmentation [16, 15, 22, 34], feature correspondence and image registration [28, 38, 1, 20], shape

---

C. Wang, Vision Lab, UCLA, U.S.A.; Center for Visual Computing, Ecole Centrale Paris; and Equipe GALEN, INRIA Saclay - Île-de-France, France. e-mail: [ch.wang@cs.ucla.edu](mailto:ch.wang@cs.ucla.edu)  
Y. Zeng and D. Samaras, Dept. of Computer Science, Stony Brook University, U.S.A. e-mail: [{yzeng,samaras}@cs.stonybrook.edu](mailto:{yzeng,samaras}@cs.stonybrook.edu)  
N. Paragios, Center for Visual Computing, Ecole Centrale Paris; and Equipe GALEN, INRIA Saclay - Île-de-France, France. e-mail: [nikos.paragios@ecp.fr](mailto:nikos.paragios@ecp.fr)

similarity and object recognition [2, 3, 29]. Let  $S \subset \mathbf{R}^3$  denote a shape<sup>1</sup>. The general idea for solving this problem is usually based on an optimization problem as follows:

$$\tau^{\text{opt}} = \arg \min_{\tau \in \mathcal{T}} \{E_{S_1, S_2}(\tau) = \rho(\tau(S_1), S_2) + \chi(\tau)\} \quad (1)$$

where  $\rho(\tau(S_1), S_2)$  denotes a measure on the geometric and/or photometric difference (often referred to as *data likelihood*) between the transformed source shape (model)  $\tau(S_1)$  and the target shape  $S_2$ .  $\chi(\tau)$  denotes a prior or regularization on the transformation  $\tau$ , and  $\mathcal{T}$  is the feasible solution set (*e.g.*, diffeomorphisms<sup>2</sup>).

One main difficulty in solving shape matching and inference lies in the fact that the shape usually lies in a high-dimensional parameter space and exhibits large and complex deformation/variance in the space where its observed data lies. This poses a challenge to the design of an efficient algorithm for the search of the optimal transformation between two shapes or the optimal shape model from the observed data. Another main difficulty originates from the facts that the problem is inherently ill-posed and that the input data are often noisy and can be partially occluded. That is why prior knowledge on the deformation/variance of the shape is often introduced to address the ill-posedness of the problem and to make the algorithm more robust to noise. However, this raises another challenge in the choice of the representation of prior knowledge, which should be effective in the aspect of modeling and efficient in the aspect of learning and inference.

### Main Obstacle - Extrinsic Factors

A ubiquitous phenomenon in vision perception is that a single object can exhibit infinite geometric variation in the observed data following the change of extrinsic factors such as sensor parameters and global object pose<sup>3</sup>. In the case of 3D data where the observation also lies in a 3D Euclidean space, different sensor parameters and/or global object poses usually lead to observations that differ by a similarity transformation (translation/rotation/scaling). In a broad sense, *extrinsic factors* refer to all that would cause a shape to have different extrinsic manifestations which are nevertheless intrinsically equivalent<sup>4</sup>. An extrinsic factor is often associated with a certain transformation group  $G$  (*e.g.*, the Euclidean group, the similarity group and the isometry group) and globally affect the configuration of a shape. Accordingly, for a shape instance, the set of all intrinsically equivalent shapes is the orbit of that instance under the corresponding transformation group  $G$ .

<sup>1</sup> The shape can also be associated with a texture model if photometric information is available.

<sup>2</sup> When a bijective mapping between  $S_1 \subset \mathbf{R}^3$  and  $S_2 \subset \mathbf{R}^3$  is required, the feasible solution can be defined as all diffeomorphisms that map  $S_1$  to  $S_2$ .

<sup>3</sup> Photometric variation can be caused by the change of illumination. We mostly focus on the geometric aspect here but the extension to the photometric aspect can be done analogously.

<sup>4</sup> The definition of the intrinsically equivalence depends on the problem to be addressed. For instance, when dealing with non-rigid 3D surface matching, we often assume that two surfaces differing by an isometric transformation (with geodesic metrics) are intrinsically equivalent.

Actually, *extrinsic factors* pose a main obstacle to addressing the aforementioned challenges efficiently, in particular in the following two major aspects.

**Regarding the problem complexity and the algorithm design,** we can see from the above discussion that such extrinsic factors are a main source of shape variability [36], the removal of which will largely reduce the complexity of shape matching and inference. The problem can become much easier if we only need to deal with the *intrinsic shape variability*, which refers to the residual (*e.g.*, intra-class variability, noise) after ruling out the effect of extrinsic factors.

The main issue in the design of the algorithm is how to define and minimize the cost function in Eq. 1 efficiently. To account for the effect of extrinsic factors, the most commonly used scheme in the literature is: decompose the transformation  $\tau$  in Eq. 1 into an transformation  $g \in G$  that corresponds to the extrinsic factors and a residual transformation  $r$  that accounts for the intrinsic shape variability, *i.e.*,  $\tau = g \circ r$ , then optimize  $g$  and  $r$  in a successive or alternating manner (*e.g.*, EM-style approaches). A typical example is the *iterative closest points* (ICP) algorithms [5, 32] for rigid shape matching, which alternates between establishing correspondences given the Euclidean transformation and estimating the Euclidean transformation given the correspondences. Another important example is related to the incorporation of shape priors and will be discussed a bit later.

Such a scheme requires initializing  $g$  and is prone to be trapped at local minima during the alternating search. Therefore, it usually works well only when the two shapes are close enough under the given initialization of  $g$ . Another important limitation is that it cannot directly deal with the case where  $g$  is difficult to be explicitly represented (*e.g.*, the isometric transformation that is often considered in non-rigid 3D surface matching). Last, the search for optimal  $r$  (*i.e.*, the global minimum with respect to  $r$ ) for a fixed  $g$  is actually difficult in general and its complexity increases sharply as  $g$  deviates from the true transformation.

**Regarding the incorporation of the shape prior,** extrinsic factors pose an obstacle for connecting the shape instance and the prior model in the matching and inference process. In fact, the prior information on a shape class lies in the residual transformation  $r$  after factoring out  $g$  corresponding to extrinsic factors from the transformation  $\tau$ . Based on this, most existing shape prior models [22], *e.g.*, the well-known *active shape/appearance models* (ASMs/AAMs) [16, 15], are built by first aligning all the training samples into a reference space (to factor out the similarity group) and then learning the shape distribution on these registered samples.

However, such prior models often exhibit two main limitations. On the one hand, the estimation of the similarity transform  $g$  is required both in the training and the inference stages, since the learned model and an observed shape instance are in different coordinate frames in general. Besides the computational complexity, such an estimation also introduces certain bias on the learned prior model, since the optimal decomposition of  $\tau$  into  $g$  and  $r$  actually is an ill-posed problem. On the other hand, the optimal search in the inference stage with such prior models requires initializing  $g$  and is prone to be trapped at local minima.

### Key Strategy - Encoding Shape Invariance in Higher-order Graphs

In fact, due to the intrinsic equivalence of the shape, the distance function in Eq. 1 should be invariant to extrinsic factors, *i.e.*,  $\rho(\tau(S_1), S_2)$  and  $\chi(\tau)$  should be  $g$ -invariant. Hence, if we can explore shape invariance with respect to extrinsic factors by choosing  $g$ -invariant data term  $\rho(\tau(S_1), S_2)$  and prior model  $\chi(\tau)$ , then we will be able to efficiently search for the optimal transformation  $\tau^{\text{opt}}$  without searching for  $g$ . In particular, when extrinsic factors correspond to a transformation group, such a scheme can be interpreted as representing a shape in an intrinsic shape space that is  $g$ -invariant and the correspondence is then determined in such an intrinsic shape space, where the shape variability is largely reduced.

To this end, we are particularly interested in discrete representations of shapes, which have been widely employed in the literature, where the transformation  $\tau$  in Eq. 1 is represented by the correspondences between the points of two shapes. Then the shape matching and inference problem boils down to determining the correspondence from the target shape (or the observed image data) for each point on the source shape (model). Recent significant development in graph-based methods and inference techniques (*e.g.*, Markov Random Field (MRF) inference algorithms [10, 25, 27] and graph matching [37, 28, 38]) have demonstrated their potential in solving such a correspondence problem. In particular, the newly developed techniques for higher-order models [24, 27, 23, 17] enhance significantly the applicable extent and the performance of graph-based methods. In such a context, we employ higher-order potentials to characterize measures/statistics that are  $g$ -invariant (*e.g.*, similarity-invariant and isometry-invariant) and optimize the energy function using discrete optimization methods to address 3D shape matching and inference (*e.g.*, [42, 39, 40, 41]). One important advantage of such a scheme is that the problem can be solved in a one-shot optimization algorithm with optimality guarantee.

In the next two sections, we will show via our recent works [42, 39] how this methodology can be implemented for two typical problems: non-rigid 3D surface matching and knowledge-based 3D segmentation, and demonstrate the superior performance of our approaches. Finally we will conclude the chapter with a discussion of future directions in section 4.

## 2 Non-rigid 3D Surface Matching

We present our approach [42] to robustly establish correspondences between two surfaces via a higher-order graph-based formulation, where the similarity between local structures and the distortion of global structures are isometry-invariant and incorporated together via singleton terms and third-order interactions, respectively.

Let us denote by  $\mathcal{P}_1$  and  $\mathcal{P}_2$  the two point sets from surfaces  $S_1$  and  $S_2$ , respectively. Our goal is to find the correspondence from  $\mathcal{P}_2$  for each point of  $\mathcal{P}_1$ , if it exists. This can be formulated as selecting a subset (referred to as *matching*)  $\mathcal{M}$  from the set of all possible correspondences  $\mathcal{A} \triangleq \mathcal{P}_1 \times \mathcal{P}_2$  that leads to the least

dissimilarity while respecting matching constraints (*e.g.*, one-to-one mapping). For each correspondence  $a = (i, j) \in \mathcal{A}$ , we assign a Boolean variable  $x_a$  to indicate if  $a$  is included in the matching  $\mathcal{M}$  ( $x_a = 1$ ) or not ( $x_a = 0$ ). By doing so, the matching  $\mathcal{M}$  can be represented by a tuple of Boolean variables  $\mathbf{x} = (x_a)_{a \in \mathcal{A}}$ . The feasible solution space  $\mathcal{X}$  of  $\mathbf{x}$  depends on the matching constraints. Here, we impose the constraint that each point in  $\mathcal{P}_1$  is mapped to at most one point in  $\mathcal{P}_2$  and *vice versa*, leading to the following feasible solution space  $\mathcal{X}$ :

$$\mathcal{X} = \{\mathbf{x} \in \{0, 1\}^{|\mathcal{A}|} \mid \sum_{i \in \mathcal{P}_1} x_{i,j} \leq 1, \sum_{j \in \mathcal{P}_2} x_{i,j} \leq 1, \forall i \in \mathcal{P}_1 \text{ and } \forall j \in \mathcal{P}_2\} \quad (2)$$

The dissimilarity induced by a matching between two surfaces can be defined based on the distortion encoded within various numbers of correspondences. We then formulate the surface matching problem as finding the optimal matching that minimizes the dissimilarity function as follows:

$$\mathbf{x}^{\text{opt}} = \arg \min_{\mathbf{x} \in \mathcal{X}} \{E(\mathbf{x}) = \sum_{a \in \mathcal{A}} \theta_a x_a + \sum_{(a,b) \in \mathcal{A}^2} \theta_{ab} x_a x_b + \sum_{(a,b,c) \in \mathcal{A}^3} \theta_{abc} x_a x_b x_c\} \quad (3)$$

In the following, we discuss the definitions of the potential functions in Eq. 3, which capture the information of both local structures and global deformation.

**The singleton potential** encodes geometric and/or photometric compatibility between the local structures of each correspondence. For simplicity, we use the Gaussian curvature  $\text{curv}(i)$  at point  $i$  as geometric descriptor, which is invariant to isometric transformation [14], as well as the texture value  $\text{tex}(i)$  at point  $i$  as photometric descriptor if texture information is available. Then, the singleton potential  $\theta_a$  for a correspondence  $a = (i, j)$  is defined as follows:

$$\theta_a = (\text{curv}(i) - \text{curv}(j))^2 + \lambda_0 (\text{tex}(i) - \text{tex}(j))^2 \quad (4)$$

where  $\lambda_0$  is a positive weight that balances the contribution between curvature and texture information. Similarly, other features can also be considered within such potentials, such as multiscale heat kernel signatures [35] and eigenfunctions of the Laplace-Beltrami operator [33].

**The higher-order potential** encodes the intrinsic deformation priors of global structures which are invariant to isometric transformation. Theories in Riemann surface [19] reveal that when two surfaces are isometrically deformed from one to the other, the correspondences (mapping) between them can be sufficiently characterized by a *Möbius transformation*, which has only six degrees of freedom and can be uniquely determined by a triplet of point-wise correspondences. Hence, we can measure the deviation from isometry for the mapping (implied by the Möbius transformation) between two surfaces determined by a triplet of point-wise correspondences, which serves as an intrinsic deformation prior term that can be encoded in a third-order potential.

According to the uniformization theorem [19], any 3D surface can be flattened conformally to a canonical 2D domain. Then for any triplet of correspondences,  $(p_i^1, p_j^1, p_k^1) \in \mathcal{P}_1$  and  $(p_i^2, p_j^2, p_k^2) \in \mathcal{P}_2$ , we first recover the associated Möbius transformation  $m^1(z)$  and  $m^2(z)$  that maps each triplet to a constant configuration  $(e^{i\frac{2\pi}{3}}, e^{i\frac{4\pi}{3}}, e^{i2\pi})$ . Under this transformation, each point  $p$  in the sets  $\mathcal{P}_1$  and  $\mathcal{P}_2$  is equipped with coordinates in  $\hat{\mathbb{C}}$  (*i.e.*, the complex plane  $\mathbb{C} \cup \{\infty\}$ ) denoted by  $z(p) \in \hat{\mathbb{C}}$ . Similar to [30], we establish correspondences between  $\mathcal{P}_1$  and  $\mathcal{P}_2$  by searching the mutually closest point correspondences set  $\mathcal{M}_{ijk}$  under the new coordinates, and define the deformation deviation from isometry as:

$$E_{ijk} = \sum_{(p_1, p_2) \in \mathcal{M}_{ijk}} |z(p_1) - z(p_2)|^2 \quad (5)$$

Then we define the intrinsic deformation prior term as follows:

$$\theta_{ijk}^{\text{Möbius}} = \begin{cases} E_{ijk}/|\mathcal{M}_{ijk}|^2 - 1 & \text{if } E_{ijk}/|\mathcal{M}_{ijk}|^2 < \delta \\ 1/|\mathcal{M}_{ijk}| & \text{otherwise} \end{cases} \quad (6)$$

Here  $\delta$  is a lower bound value to rule out unlikely correspondences (in our experiment  $\delta = 0.1$ ). Intuitively, if there are more matching pairs and the distances between those matching pairs are smaller, the potential will be lower. Such a prior term is invariant with respect to isometric transformation, due to the fact that  $E_{ijk}$  is computed in the canonical 2D domain and an isometric transformation applied to a surface will not change the representation of the surface in the canonical domain.

Since the mirror symmetry group is a subset of the isometry group, the intrinsic deformation prior term in Eq. 6 cannot resolve symmetry ambiguity. In practice we often want to eliminate such an ambiguity, for which we can define another type of third-order terms based on the Gaussian map of the surface. The Gaussian map is defined as the mapping of the normal at each point on the surface to the unit sphere [14]. Due to the fact that two triplets have the same orientation of the Gaussian maps if and only if the determinant of their normals have the same sign, we can define the below higher-order term as a penalty for extrinsic orientation inconsistency:

$$\theta_{ijk}^{\text{Gaussian}} = \begin{cases} 0 & \text{if } \det(\mathbf{n}_i^1, \mathbf{n}_j^1, \mathbf{n}_k^1) \cdot \det(\mathbf{n}_i^2, \mathbf{n}_j^2, \mathbf{n}_k^2) \geq 0 \\ 1/|\mathcal{M}_{ijk}| & \text{otherwise} \end{cases} \quad (7)$$

where  $\mathbf{n}_i \in \mathbb{R}^3$  denotes the normal at point  $i$ , and  $\det(\mathbf{n}_i, \mathbf{n}_j, \mathbf{n}_k)$  denotes the determinant of the  $3 \times 3$  matrix  $[\mathbf{n}_i, \mathbf{n}_j, \mathbf{n}_k]$ . With such an additional term, the third-order potential for each triplet of correspondences  $(p_i^1, p_j^1, p_k^1) \rightarrow (p_i^2, p_j^2, p_k^2)$  is defined as a weighted sum of the two types of potentials, *i.e.*,

$$\theta_{ijk} = \lambda_1 \theta_{ijk}^{\text{Möbius}} + \lambda_2 \theta_{ijk}^{\text{Gaussian}} \quad (8)$$

Here, only singleton and third-order terms are considered for simplification. Pair-wise potentials defined based on different metrics (*e.g.*, geodesic [12], diffusion met-

rics [13] and commute time [31]) can also be considered in this general formulation to integrate more geometric information towards improving the performance.

**Dual-decomposition-based optimization** An advantage of the pseudo-boolean formulation is that higher-order terms can be reduced into a quadratic term and then be solved by existing efficient optimization algorithms such as QPBO techniques [9, 26]. Inspired by [38], the dual-decomposition optimization framework [4, 27] and the order-reduction technique proposed in [23] are adopted to deal with the problem in Eq. 3. More specifically, the original problem is decomposed into a *linear subproblem*, a *higher-order pseudo-boolean subproblem* and a set of *local subproblems*. The linear subproblem and the local subproblems used in the experiments are similar to those of [38]. Then, a higher-order pseudo-boolean subproblem is introduced to deal with the higher-order terms in Eq. 3. After solving the subproblems, the dual variables are updated using a projected subgradient method [27, 38] to maximize the lower bound.

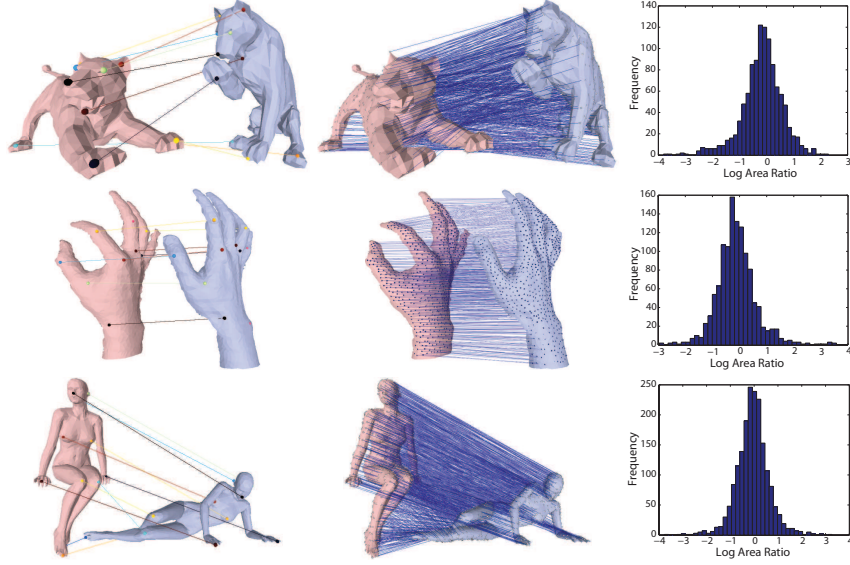
**Towards efficient dense surface matching**, we propose a two-stage optimization pipeline which consists of *sparse feature matching* and *dense point matching*. In the sparse matching stage, we establish the correspondences between two small sets of sparse features using the high-order graph matching algorithm presented above. Since any three correspondences determine a mapping between the two surfaces and provide a correspondence candidate on  $S_2$  for each point on  $S_1$ , a large number of correspondence candidates can be obtained for each point by considering all distinct triplets of correspondences in the sparse feature matching result. This can be followed by a clustering process to find the modes of the candidates so as to significantly reduce the number of candidates. Finally, a similar high-order graph matching scheme is employed to determine the optimal dense surface matching.

**Experimental results** The evaluation of our framework is done based on a number of challenging examples, which demonstrates its accuracy and efficiency, notably in challenging cases of very large deformations, or meshes that are partially occluded (see sample results in Fig. 1). Due to the lack of a ground truth regarding the dense correspondence, we quantitatively measure the quality of dense registration as follows: after performing the Delaunay triangulation of the points on the source surface, we compute the ratio of the area of each facet to the area of its matched facet (see Fig. 1). For natural deformations (*e.g.*, expression change, stretched arms or bending figures) such as those in our experiments, the local area is not expected to undergo abrupt change. Therefore the log area ratio is expected to be close to 0.

### 3 Pose-Invariant Prior and Knowledge-based Segmentation

In this section, higher-order interactions are considered to build pose-invariant shape priors and are exploited for the development of a novel one-shot optimization approach for knowledge-based 3D segmentation in medical imaging [39].





**Fig. 1** Matching results (from left to right: sparse matching, dense matching and log area ratio).

**Pose-invariant shape modeling** The shape model consists of a set  $\mathcal{V}$  of landmarks on the boundary of the object of interest. In the 3D case, we use  $x_i$  ( $i \in \mathcal{V}$ ) to denote the 3D position of landmark  $i$  and  $\mathbf{x} = (x_i)_{i \in \mathcal{V}}$  denote the positions of all the landmarks. Our goal is to learn priors on  $\mathbf{x}$  from the training data that consists of a set of  $M$  shapes. Instead of registering all the surfaces into a reference space, we only assume that point-wise correspondences have been determined for the landmarks in the training set. We propose to learn statistics on similarity invariants, such as the relative distances between pairs of landmarks in a clique. Let  $\mathcal{P}_c = \{(i, j) | i, j \in c \text{ and } i < j\}$  denote all the pairs for a clique  $c$  ( $c \subseteq \mathcal{V}$  and  $|c| \geq 3$ ) of landmarks, and  $d_{ij} = \|x_i - x_j\|$  denote the Euclidean distance between points  $i$  and  $j$  ( $(i, j) \in \mathcal{P}_c$ ). We compute the *relative distance*  $\hat{d}_{ij}$  by normalizing  $d_{ij}$  over the sum of the distances between all the pairs of points involved in the clique  $c$ , *i.e.*,

$$\hat{d}_{ij} = \frac{d_{ij}}{\sum_{(i,j) \in \mathcal{P}_c} d_{ij}} \quad (9)$$

Since the distance  $\hat{d}_{ij}$  is normalized (*i.e.*,  $\sum_{\{i,j\} \in \mathcal{P}_c} \hat{d}_{ij} = 1$ ), it is sufficient to consider a vector  $\hat{\mathbf{d}}_c$  of relative distances corresponding to  $|\mathcal{P}_c| - 1$  pairs of points. For instance, let us consider a third-order clique  $c = \{i, j, k\}$  ( $i, j, k \in \mathcal{V}$  and  $i < j < k$ ), the corresponding three points compose a triangle  $\Delta_{ijk}$  and  $\hat{\mathbf{d}}_c$  denotes the relative lengths ( $\hat{d}_{ij}, \hat{d}_{jk}$ ) of the sides  $(i, j)$  and  $(j, k)$ , *i.e.*,

$$\hat{\mathbf{d}}_c = \left( \frac{d_{ij}}{d_{ij} + d_{jk} + d_{ki}}, \frac{d_{jk}}{d_{ij} + d_{jk} + d_{ki}} \right) \quad (10)$$

We can learn the statistics  $\psi_c(\hat{\mathbf{d}}_c)$  of  $\hat{\mathbf{d}}_c$  from the training data, with standard probabilistic models such as Gaussian Distributions, Gaussian Mixtures and Parzen-Windows. Finally, we build the higher-order shape model  $\mathcal{S} = (\mathcal{V}, \mathcal{C}, \{\psi_c(\cdot)\}_{c \in \mathcal{C}})$ , where  $\mathcal{V}$  and  $\mathcal{C}$  determine the topology of the model and  $\{\psi_c(\cdot)\}_{c \in \mathcal{C}}$  characterizes the statistical geometric priors between the points contained in each clique  $c \in \mathcal{C}$ . In the case where third-order cliques are used,  $\mathcal{C}$  is defined as  $\mathcal{C} = \{\{i, j, k\} | i, j, k \in \mathcal{V} \text{ and } i < j < k\}$ . Such statistical constraints can be easily encoded in a higher-order MRF with a set of cliques that includes  $\mathcal{C}$ , leading to a prior probability on the 3D configuration of the shape model as follows:

$$p(\mathbf{x}) \propto \prod_{c \in \mathcal{C}} \psi_c(\hat{\mathbf{d}}_c(x_c)) \quad (11)$$

where  $\hat{\mathbf{d}}_c(x_c)$  denotes the mapping from the 3D positions  $x_c$  of the three points contained in the clique  $c$  to the relative distance vector  $\hat{\mathbf{d}}_c$ . It is easy to verify that  $\hat{\mathbf{d}}_c$  is similarity-invariant. However, other similarity invariants (such as angles of a triangle) can also be adopted in the above shape prior model.

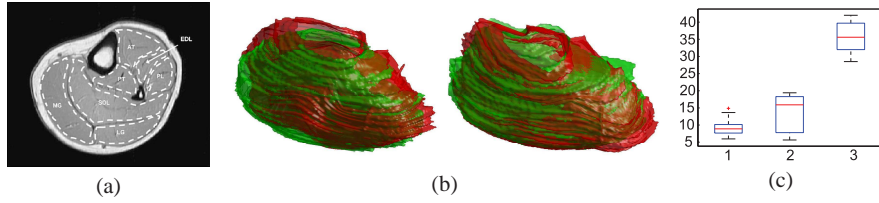
**Landmark candidate detection** In order to explore image support through feature vectors and to avoid a prohibitive computational complexity, we perform landmark detections to find a set of correspondence candidates in the observed image for each landmark  $i$  ( $i \in \mathcal{V}$ ) in the 3D shape model. To this end, we first learn a classifier for each landmark, then compute a score for each possible location, and finally select the  $L$  positions that have the best scores to compose the candidate set for the landmark. We employed *Random Forests* [11] to perform the classification.

**Higher-order MRF segmentation formulation** The shape model, together with the evidence from the image support, is formulated within a higher-order MRF to perform image segmentation. To this end, we associate each node of the MRF with a landmark  $i$  ( $i \in \mathcal{V}$ ), and the latent variable  $X_i$  corresponding to the node  $i$  denotes the 3D position of the associated landmark. The candidate set of each variable  $X_i$  is denoted by  $\mathcal{X}_i$ , which consists of the detected landmark candidates. Thus, the Cartesian product  $\mathcal{X} = \prod_{i \in \mathcal{V}} \mathcal{X}_i$  denotes the candidate set of the configuration  $\mathbf{x} = (x_i)_{i \in \mathcal{V}}$  of the MRF model. In order to introduce the pose-invariant shape prior (of third order) into the MRF formulation, we associate a triplet of landmarks to a third-order clique  $c$  and use the potential function of the clique  $c$  to encode the statistical geometric constraints between the three landmarks. Finally, the segmentation problem is transformed into estimating the optimal configuration  $\mathbf{x}^{\text{opt}}$  of the higher-order MRF, which is formulated as a minimization of the MRF energy  $E(\mathbf{x})$ :

$$\mathbf{x}^{\text{opt}} = \arg \min_{\mathbf{x} \in \mathcal{X}} E(\mathbf{x}) \quad (12)$$

The energy of MRF is defined as a sum of singleton potentials  $U_i(x_i)$  ( $i \in \mathcal{V}$ ) and third-order potentials  $U_c(x_c)$  ( $c \in \mathcal{C}$ ), *i.e.*,

$$E(\mathbf{x}) = \sum_{i \in \mathcal{V}} U_i(x_i) + \sum_{c \in \mathcal{C}} H_c(x_c) \quad (13)$$



**Fig. 2** Muscle segmentation. (a) A slice of a 3D MRI image of calf muscle with expert annotation. (b) MG muscle segmentation results (green: reference; red: result). (c) Boxplots of the average landmark error measure in voxel (1. our method, 2. method in [18], 3. standard ASM method.). On each box, the central mark in red is the median, the edges of the box are the 25th and 75th percentiles, the whiskers extend to the most extreme data points.

The singleton potential  $U_i(x_i)$  ( $i \in \mathcal{V}$ ) consists of the negative log-likelihood, imposing penalty for landmark  $i$  to be located at position  $x_i$  in image  $\mathbf{I}$ , *i.e.*,

$$U_i(x_i) = -\log p(\mathbf{I}|x_i) \quad (14)$$

$p(\mathbf{I}|x_i)$  is defined using the classifier’s output probability value for landmark  $i$  to be located at  $x_i$ . The higher-order clique potential  $U_c(x_c)$  ( $c \in \mathcal{C}$ ) encodes the statistical geometrical constraints on the triplet  $c$  of points and is defined as:

$$U_c(x_c) = -\alpha \cdot \log \psi_c(\hat{\mathbf{d}}_c(x_c)) \quad (15)$$

where  $\alpha > 0$  is a positive weight,  $\hat{\mathbf{d}}_c(x_c)$  and  $\psi_c(\cdot)$  have been presented previously. Regional terms can also be factorized and incorporated in such an MRF model [41].

**Higher-order MRF inference** We adopt the dual-decomposition optimization framework [4, 27] to solve the inference (Eq. 12). More specifically, we decompose the original problem into a set of subproblems, each corresponding to a factor-tree [6] and perform the exact inference efficiently in each subproblem in polynomial time using the max-product belief propagation algorithm [6], with complexity  $O(NL^K)$ , where  $N$ ,  $L$  and  $K$  denote the number of nodes, the number of candidates for each node, and the maximum order of the factors, respectively. The solutions of the subproblems are combined using projected subgradient method to solve the Lagrangian dual so as to obtain the solution of the original problem [27].

**Experimental results** The dataset for experimental validation consists of 3D MRI scans of the calf muscles of 25 subjects (Fig. 2(a)). Standard of reference was available, consisting of annotations provided by experts for the Medial Gastrocnemius (MG) muscle. To segment MG muscle from such images is challenging since there is no evident difference of tissue properties between neighbor muscles and boundaries between adjacent muscles are visible very sparsely and heterogeneously. We performed a leave-one-out cross-validation on the whole dataset. For comparison purposes, we considered those methods presented in [18]. Fig. 2(b) shows two examples of the surface reconstruction results obtained using the estimated position

of landmarks and *thin plate spline* (TPS) [8]. Fig. 2(c) presents the average distance between the real and estimated landmark positions using different methods, which confirms the superior performance of the proposed similarity-invariant shape prior and inference using higher-order MRFs.

## 4 Conclusion

We have shown, via two specific applications, the idea of encoding shape invariance in higher-order graphs for shape matching and inference, resulting in a one-shot optimization algorithm without initializing and estimating extrinsic factors. Similar ideas can be applied to address other extrinsic factors. For example, we introduced in [40] a unified paradigm for 3D landmark model inference from monocular 2D images to simultaneously determine both the optimal 3D model and the corresponding 2D projections without explicit estimation of the camera viewpoint. As the next step, it is interesting to study the optimal invariants and to recover the optimal subset of higher-order interactions that can best express the 3D geometric manifold. Besides, faster optimization algorithms of higher-order MRFs could be beneficial both in terms of the considered applications as well as in terms of modularity with respect to other shape matching and inference applications.

## References

1. Ashburner, J.: A Fast Diffeomorphic Image Registration Algorithm. *NeuroImage* **38**(1), 95–113 (2007)
2. Belongie, S., Malik, J., Puzicha, J.: Shape Matching and Object Recognition Using Shape Contexts. *IEEE TPAMI* **24**(4), 509–522 (2002)
3. Berg, A.C., Berg, T.L., Malik, J.: Shape Matching and Object Recognition Using Low Distortion Correspondences. *CVPR* (2005)
4. Bertsekas, D.P.: *Nonlinear Programming* (Second Edition). Athena Scientific (1999)
5. Besl, P.J., McKay, N.D.: A Method for Registration of 3-D Shapes. *IEEE TPAMI* **14**(2), 239–256 (1992)
6. Bishop, C.M.: *Pattern Recognition and Machine Learning* (Information Science and Statistics). Springer (2006)
7. Blanz, V., Vetter, T.: A Morphable Model for the Synthesis of 3D Faces. *SIGGRAPH* (1999)
8. Bookstein, F.L.: Principal Warps: Thin-Plate Splines and the Decomposition of Deformations. *IEEE TPAMI* **11**(6), 567–585 (1989)
9. Boros, E., Hammer, P.L., Sun, X.: Network Flows and Minimization of Quadratic Pseudo-Boolean Functions. Tech. rep., RRR 17-1991, RUTCOR Research Report (1991)
10. Boykov, Y., Veksler, O., Zabih, R.: Fast Approximate Energy Minimization via Graph Cuts. *IEEE TPAMI* **23**(11), 1222–1239 (2001)
11. Breiman, L.: Random Forests. *Machine learning* **54**(2), 5–32 (2001)
12. Bronstein, A.M., Bronstein, M.M., Kimmel, R.: Generalized Multidimensional Scaling: A Framework for Isometry-Invariant Partial Surface Matching. *PNAS* **103**(5), 1168–1172 (2006)
13. Bronstein, A.M., Bronstein, M.M., Kimmel, R., Mahmoudi, M., Sapiro, G.: A Gromov-Hausdorff Framework with Diffusion Geometry for Topologically-Robust Non-rigid Shape. *IJCV* **89**(2–3), 266–286 (2010)

14. do Carmo, M.P.: *Differential Geometry of Curves and Surfaces*. Prentice Hall (1976)
15. Cootes, T.F., Edwards, G.J., Taylor, C.J.: Active Appearance Models. *IEEE TPAMI* **23**(6), 681–685 (2001)
16. Cootes, T.F., Taylor, C.J., Cooper, D.H., Graham, J.: Active Shape Models-Their Training and Application. *Computer Vision and Image Understanding* **61**(1), 38–59 (1995)
17. Duchenne, O., Bach, F., Kweon, I., Ponce, J.: A Tensor-based Algorithm for High-order Graph Matching. In: *CVPR* (2009)
18. Essafi, S., Langs, G., Deux, J.F., Rahmouni, A., Bassez, G., Paragios, N.: Wavelet-Driven Knowledge-Based MRI Calf Muscle Segmentation. In: *ISBI* (2009)
19. Farkas, H.M., Kra, I.: *Riemann Surfaces*. Springer (2004)
20. Glocker, B., Komodakis, N., Tziritas, G., Navab, N., Paragios, N.: Dense Image Registration through MRFs and Efficient Linear Programming. *Medical Image Analysis* **12**(6), 731–741 (2008)
21. Gu, L., Kanade, T.: 3D Alignment of Face in a Single Image. In: *CVPR* (2006)
22. Heimann, T., Meinzer, H.P.: Statistical Shape Models for 3D Medical Image Segmentation: A Review. *Medical Image Analysis* **13**(4), 543–63 (2009)
23. Ishikawa, H.: Higher-Order Clique Reduction in Binary Graph Cut. In: *CVPR* (2009)
24. Kohli, P., Pawan Kumar, M., Torr, P.H.S.: P3 & Beyond: Move Making Algorithms for Solving Higher Order Functions. *IEEE TPAMI* **31**(9), 1645–1656 (2009)
25. Kolmogorov, V.: Convergent Tree-Reweighted Message Passing for Energy Minimization. *IEEE TPAMI* **28**(10), 1568–1583 (2006)
26. Kolmogorov, V., Rother, C.: Minimizing Nonsubmodular Functions with Graph Cuts-A Review. *IEEE TPAMI* **29**(7), 1274–1279 (2007)
27. Komodakis, N., Paragios, N., Tziritas, G.: MRF Energy Minimization and Beyond via Dual Decomposition. *IEEE TPAMI* **33**(3), 531–552 (2011)
28. Leordeanu, M., Hebert, M.: A Spectral Technique for Correspondence Problems using Pairwise Constraints. In: *ICCV* (2005)
29. Lin, L., Liu, X., Zhu, S.C.: Layered graph matching with composite cluster sampling. *IEEE TPAMI* **32**(8), 1426–1442 (2010)
30. Lipman, Y., Funkhouser, T.: Möbius Voting for Surface Correspondence. *ACM TOG* **28**(3), 72:1–72:12 (2009)
31. Qiu, H., Hancock, E.R.: Clustering and Embedding Using Commute Times. *IEEE TPAMI* **29**(11), 1873–1890 (2007)
32. Rusinkiewicz, S., Levoy, M.: Efficient Variants of the ICP Algorithm. In: *International Conference on 3-D Digital Imaging and Modeling* (2001)
33. Rustamov, R.M.: Laplace-Beltrami Eigenfunctions for Deformation Invariant Shape Representation. In: *SGP*, pp. 225–233 (2007)
34. Seghers, D., Loeckx, D., Maes, F., Vandermeulen, D., Suetens, P.: Minimal Shape and Intensity Cost Path Segmentation. *IEEE TMI* **26**(8), 1115–1129 (2007)
35. Sun, J., Ovsjanikov, M., Guibas, L.J.: A Concise and Provably Informative Multi-Scale Signature. Based on Heat Diffusion. In: *Computer Graphics Forum*, vol. 28, pp. 1383–1392 (2009)
36. Sundaramoorthi, G., Petersen, P., Varadarajan, V.S., Soatto, S.: On the set of images modulo viewpoint and contrast changes. In: *CVPR* (2009)
37. Torr, P.H.S.: Solving Markov Random Fields using Semi Definite Programming. In: *International Workshop on Artificial Intelligence and Statistics* (2003)
38. Torresani, L., Kolmogorov, V., Rother, C.: Feature Correspondence via Graph Matching: Models and Global Optimization. In: *ECCV* (2008)
39. Wang, C., Teboul, O., Michel, F., Essafi, S., Paragios, N.: 3D Knowledge-Based Segmentation Using Pose-Invariant Higher-Order Graphs. In: *MICCAI* (2010)
40. Wang, C., Zeng, Y., Simon, L., Kakadiaris, I., Samaras, D., Paragios, N.: Viewpoint Invariant 3D Landmark Model Inference from Monocular 2D Images Using Higher-Order Priors. In: *ICCV* (2011)
41. Xiang, B., Wang, C., Deux, J.F., Rahmouni, A., Paragios, N.: 3D Cardiac Segmentation with Pose-Invariant Higher-Order MRFs. In: *ISBI* (2012)
42. Zeng, Y., Wang, C., Wang, Y., Gu, X., Samaras, D., Paragios, N.: Dense Non-rigid Surface Registration Using High-Order Graph Matching. In: *CVPR* (2010)