

# Discriminative Parameter Estimation for Random Walks Segmentation: Technical Report

Pierre-Yves Baudin<sup>1-6</sup>, Danny Goodman<sup>1-3</sup>, Puneet Kumar<sup>1-3</sup>,  
Noura Azzabou<sup>4-6</sup>, Pierre G. Carlier<sup>4-6</sup>, Nikos Paragios<sup>1-3</sup>, and  
M. Pawan Kumar<sup>1-3</sup>

<sup>1</sup> Center for Visual Computing, École Centrale Paris, FR

<sup>2</sup> Université Paris-Est, LIGM (UMR CNRS), École des Ponts ParisTech, FR

<sup>3</sup> Équipe Galen, INRIA Saclay, FR

<sup>4</sup> Institute of Myology, Paris, FR

<sup>5</sup> CEA, I<sup>2</sup> BM, MIRCen, IdM NMR Laboratory, Paris, FR

<sup>6</sup> UPMC University Paris 06, Paris, FR

## 1 Literature Survey on the State of the Art Algorithms in Muscle Segmentation

In the following, we present the principal methods addressing the segmentation of skeletal muscles.

*Muscle Segmentation using Simplex meshes with medial representations* A skeletal muscle segmentation method was presented in [8] based on simplex meshes [6]. Considering a 3D surface model, simplex meshes are discrete meshes where each vertex has exactly 3 neighbors. Having a constant connectivity allows to simply parametrize the location of one vertex with respect to its neighbors, and thus parametrize deformation of the shape – translation, rotation, scaling – in a local manner. Indeed, the location of a pixel, denoted as  $\mathbf{x}$  can be expressed as a linear combination of the locations of the three neighbors plus a local elevation term parallel to the local normal:  $\mathbf{x} = \varepsilon_1 \mathbf{x}_1 + \varepsilon_2 \mathbf{x}_2 + (1 - \varepsilon_1 - \varepsilon_2) \mathbf{x}_3 + h \mathbf{n}$ . As a result, many local measurements – including curvature and cell surface – can be computed efficiently and global energy terms enforcing local constraints come up naturally.

Here, the authors impose local smoothing via curvature averaging, which does not tend to reduce the surface like 1-order operators typically do. Prior knowledge is imposed by constraining the local scale changes on the elevation parameter with respect to a reference shape. Denoting the surface of the triangle formed by the three neighbors of a pixel as  $S$ , given the reference shape parameters  $(\tilde{\varepsilon}_1, \tilde{\varepsilon}_2, \tilde{h}, \tilde{S})$ , the new location of the considered pixel is expressed as:

$$\mathbf{x} = \tilde{\varepsilon}_1 \mathbf{x}_1 + \tilde{\varepsilon}_2 \mathbf{x}_2 + (1 - \tilde{\varepsilon}_1 - \tilde{\varepsilon}_2) \mathbf{x}_3 + \tilde{h} \left( S / \tilde{S} \right)^{1/\beta} \mathbf{n}, \quad (1)$$

where  $\beta \in [2, +\infty[$  is a parameter which sets the amount of allowed local deformation: with  $\beta = 2$  this definition is similitude invariant; with  $\beta = +\infty$  this

definition is invariant through rigid transformations only. The model is attached to the target image through either gradient norm maximization in the direction of the gradient at the location of the vertices, or maximization of similarities between the reference and the target images at the vertices location.

A medial representation – similar to the M-reps [12] – is combined with the simplex parametrization to exploit the specific tubular shapes of the skeletal muscles. Medial vertices are added to the model, constrained to remain on the medial axis of the tubular objects. This is achieved by connecting the new vertices to the surface vertices through spring-like forces. This constrains the global structure to resemble its initial reference shape, thus acting as a global shape prior. This medial axis representation also allows efficient collision handling. The model is fit to the image through an iterative process of successive local evolutions. Such model appear to always yield a valid solution, sometimes at the price of an excessive regularization or lack of adaptability to the specifics of the target image.

*Muscle Segmentation using Deformable Models and Shape Matching* A shape prior for muscle segmentation in 3D images was presented in [9], deriving from a computer animation technique, called *shape matching*, used to efficiently approximate large soft-tissue elastic deformations. This method was applied to muscle segmentation with some success. In this approach, discrete meshes are used to parametrize the moving surface. Let  $\mathbf{x}^0$  be the vector containing the initial position of the control points of the parametric surface. Clustering is performed on  $\mathbf{x}^0$  such that each cluster  $\zeta_i$  contains at least a certain number of vertices (set by the user). During segmentation, the evolution of the active surface is performed according to the following iterative procedure:

1. Shift vertices according to the external force:  $\tilde{\mathbf{x}}^t = \mathbf{f}_{\text{ext}} + \mathbf{x}^t$ . The external “force”  $\mathbf{f}_{\text{ext}}$  is computed as the maximal gradient search in the gradient direction.
2. Regularize vertex positions:
  - (a) Compute rigid registration for each clusters:

$$\mathbf{T}_i = \arg \min \sum_{j \in \zeta_i} \|\mathbf{T}_i \mathbf{x}_j^0 - \tilde{\mathbf{x}}_j^t\|^2, \quad (2)$$

- (b) Average target position for each vertex:

$$\mathbf{x}_i^{t+1} = \frac{1}{|\zeta_i|} \sum_{j \in \zeta_i} \mathbf{T}_j \mathbf{x}_j^0. \quad (3)$$

Single reference prior models are convenient in that they require only one annotated example of the objects of interest. However, when segmenting a class of objects whose shape varies a lot, such approach becomes too constraining and does not allow the model to adopt valid shapes which are too different from the single reference.

*Muscle segmentation using a hierarchical statistical shape model* A hierarchical prior model using Diffusion Wavelets was proposed to segment organs in [7] – including one calf muscle – in MRI. This model builds on the formulation of the ASM [4], using a different basis for the subspace of valid solutions. One of the main drawbacks of ASMs, is that they often require a large number of training data in order to obtain relevant decomposition modes. Indeed, some non-correlated shape features – such as global and local shape configurations – are often modeled by the same deformation modes. Thus, desired shape behaviors are often mixed with unwanted shape behaviors when optimizing the shape parameters for segmenting a new image. The hierarchical approach allows to uncorrelate small and large scale shape behaviors. Moreover, the presented method also uncorrelates long-range shape behaviors, thus ensuring that deformation mode are spatially localized.

We give a brief summary of this method. First, a graph  $\mathcal{G}(\mathcal{V}, \mathcal{E})$  is built on the set of landmarks:  $\mathcal{V}$  is the set of nodes and each landmark corresponds to a node in  $\mathcal{V}$ ;  $\mathcal{E}$  is the set of edges, whose weights are determined through a statistical analysis of the mutual relations between the landmarks in the training set  $\{\mathbf{x}_k\}_{k=1\dots K}$  (cf. Shape Maps [11]). As a result, landmarks with independent behaviors will be connected by edges with a small weight, whereas nodes with strongly related behaviors – such as neighboring points – will be connected by large weight edges.

Second, a Diffusion Wavelet decomposition of  $\mathcal{G}$  is performed. This process involves computing the diffusion operator  $\mathbf{T}$  of graph  $\mathcal{G}$ , which is the symmetric normalized Laplace-Beltrami operator, and computing and compressing the dyadic powers of  $\mathbf{T}$ . The output of this decomposition is a hierarchical orthogonal basis  $\{\Gamma_i\}$  for the graph, whose vectors correspond to different graph scales; considering the vector of landmark positions when decomposed on the new basis:

$$\mathbf{x} = \bar{\mathbf{x}} + \Gamma \mathbf{p}, \quad (4)$$

global deformations – i.e. global relations between all the nodes – are controlled by some of the coefficients in  $\mathbf{p}$ , while local interactions – i.e. local interactions between close-by nodes – are controlled by some other coefficients in  $\mathbf{p}$ . Projecting all the training examples onto this new basis, a PCA is performed *at each scale* of the decomposition. Finally, during the segmentation process, the landmarks are positioned on the target image in an iterative manner: 1) the position of the landmarks is updated according to a local appearance model; 2) they are projected into the hierarchical subspace defined previously.

*Muscle segmentation using a continuous region model* A region-based segmentation method, proposed in [5], was extended to multi-label segmentation in [2], and applied to skeletal muscle segmentation [1]. Before performing the PCA on the training samples, an Isometric Log-Ratio (ILR) transform is applied to the assignment vectors. The reason for using this transform is that multi-label segmentation requires to have probabilities at all times, which the previous method does not achieve. Here, the PCA is performed in the ILR space and its output

is projected back into the initial probability space. Denoting  $\eta_\gamma = \mu + F\gamma$  a segmentation in the subspace of valid solution spanned by the PCA in the ILR space, the following functional is proposed:

$$E(\eta_\gamma) = d(\eta_{BG}, \eta_\gamma)^2 + \int (1 - h(x)) |\nabla \eta_\gamma|^2 + \gamma^T \Sigma^{-1} \gamma, \quad (5)$$

where  $d(\eta_{BG}, \cdot)$  is an intensity prior functional for separating muscle voxels from background voxels, and  $h(x)$  is an edge-map of the target image such that the energy is minimal when the boundaries of the model match the edges in the image.

## 2 The Random Walks Algorithm

The Random Walks algorithm is graph-based: consider a graph  $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ , where  $\mathcal{V}$  is a set of nodes – corresponding to each voxel in the 3d image – and  $\mathcal{E}$  is a set of edges – one per pair of adjacent voxels. Let us also denote a set of labels – one per object to segment – as  $\mathcal{S}$ .

The aim of the Random Walks Algorithm [?] is to compute an assignment probability of all voxels to all labels. These probabilities depend on: i) contrast between adjacent voxels, ii) manual – thus deterministic – assignments of some voxels, called *seeds*, and iii) prior assignment probabilities.

The probabilities, contained in vector  $\mathbf{y}$ , can be obtained by minimizing the following functional:

$$\begin{aligned} E_{\text{RWprior}}(\mathbf{x}, \mathbf{y}) &= \mathbf{y}^\top L \mathbf{y} + w \|\mathbf{y} - \mathbf{y}_0\|_{\Omega(\mathbf{x})}^2 \\ &= \mathbf{y}^\top \left[ \sum_{\alpha} w_{\alpha} L_{\alpha} \right] \mathbf{y} + \sum_{\beta} w_{\beta} \|\mathbf{y} - \mathbf{y}_{\beta}\|_{\Omega_{\beta}}^2, \end{aligned}$$

which is a linear combination of Laplacians and prior terms.

The Laplacian matrices  $L_{\alpha}$  contain the contrast terms. Its entries are of the form:

$$L_{i,j} = \begin{cases} \sum_k \omega_{kj} & \text{if } i = j, \\ -\omega_{ij} & \text{if } (i, j) \in \mathcal{E}, \\ 0 & \text{otherwise.} \end{cases} \quad (6)$$

Here,  $\omega_{ij}$  designates the weight of edge  $(i, j)$ . It is usually computed as follows:

$$\omega_{ij} = \exp \left( -\beta (I_i - I_j)^2 \right), \quad (7)$$

where  $I_i$  is the intensity of voxel  $i$ .

In our experiments, we used three different Laplacians using this formulation, with three values of  $\beta$ : 50, 100 and 150 (with the image voxel values normalized with their empirical standard deviation).

We also implemented the lesser used alternate formulation:

$$w_{ij} = \frac{1}{\beta |I_i - I_j| + \varepsilon}, \quad (8)$$

which we employed in one additional Laplacian term with  $\beta = 100$  and  $\varepsilon = 1$  for comparison purposes, since the selected values do give good results on their own.

Since the objective function is quadratic in  $\mathbf{y}$ , its minimum can be computed by minimizing a linear system. The quadratic term, composed of a sum of Laplacians and diagonal matrices due to the prior term, is very sparse and has a specific structure due to the fact that only adjacent voxels are connected with an edge.

Given the size of the problem (several millions of variables), this system has to be solved with iterative methods, such as Conjugate Gradient. The specific structure of the problem and the existence of parallelized algorithms (such as multigrid Conjugate Gradient) allow for an efficient optimization. For instance, our own implementation takes less than 20s for volumes of size  $200 \times 200 \times 100$  on a regular desktop machine.

### 3 Derivation of the Latent SVM Upper Bound

Given a dataset  $\mathcal{D} = \{(\mathbf{x}_k, \mathbf{z}_k), k = 1, \dots, N\}$ , which consists of inputs  $\mathbf{x}_k$  and their hard segmentation  $\mathbf{z}_k$ , we would like to estimate parameters  $\mathbf{w}$  such that the resulting inferred segmentations are accurate. Here, the accuracy is measured using the loss function  $\Delta(\cdot, \cdot)$ . Formally, let  $\tilde{\mathbf{y}}_k(\mathbf{w})$  denote the soft segmentation obtained by minimizing the energy functional  $E(\cdot, \mathbf{x}_k; \mathbf{w})$  for the  $k$ -th training sample, that is,

$$\tilde{\mathbf{y}}_k(\mathbf{w}) = \arg \min_{\mathbf{y}} \mathbf{w}^\top \psi(\mathbf{x}_k, \mathbf{y}). \quad (9)$$

We would like to learn the parameters  $\mathbf{w}$  such that the empirical risk is minimized over all samples in the dataset. In other words, we would like to estimate the parameters  $\mathbf{w}^*$  such that

$$\mathbf{w}^* = \arg \min_{\mathbf{w}} \frac{1}{N} \sum_k \Delta(\mathbf{z}_k, \tilde{\mathbf{y}}_k(\mathbf{w})). \quad (10)$$

The above objective function is highly non-convex in  $\mathbf{w}$ , which makes it prone to bad local minimum solutions. To alleviate this deficiency, the latent SVM

formulation upper bounds the risk for a sample  $(\mathbf{x}, \mathbf{z})$  as follows:

$$\Delta(\mathbf{z}_k, \tilde{\mathbf{y}}_k(\mathbf{w})) = \Delta(\mathbf{z}_k, \tilde{\mathbf{y}}_k(\mathbf{w})) + \mathbf{w}^\top [\psi(\mathbf{x}_k, \tilde{\mathbf{y}}_k(\mathbf{w})) - \psi(\mathbf{x}_k, \tilde{\mathbf{y}}_k(\mathbf{w}))], \quad (11)$$

$$\leq \min_{\Delta(\mathbf{z}_k, \hat{\mathbf{y}})=0} \mathbf{w}^\top \psi(\mathbf{x}_k, \hat{\mathbf{y}}) \quad (12)$$

$$\begin{aligned} & - [\mathbf{w}^\top \psi(\mathbf{x}_k, \tilde{\mathbf{y}}_k(\mathbf{w})) - \Delta(\mathbf{z}_k, \tilde{\mathbf{y}}_k(\mathbf{w}))], \\ & \leq \min_{\Delta(\mathbf{z}_k, \hat{\mathbf{y}})=0} \mathbf{w}^\top \psi(\mathbf{x}_k, \hat{\mathbf{y}}) \quad (13) \\ & - \min_{\bar{\mathbf{y}}} [\mathbf{w}^\top \psi(\mathbf{x}_k, \bar{\mathbf{y}}) - \Delta(\mathbf{z}_k, \bar{\mathbf{y}})]. \end{aligned}$$

The first inequality follows from the fact that the prediction  $\tilde{\mathbf{y}}_k(\mathbf{w})$  has the *minimum possible energy* (see equation 9). Thus, its energy has to be less than or equal to the energy of any compatible segmentation  $\hat{\mathbf{y}}$  with  $\Delta(\mathbf{z}_k, \hat{\mathbf{y}}) = 0$ . The second inequality is true since it replaces the *loss augmented* energy of the prediction  $\tilde{\mathbf{y}}_k(\mathbf{w})$  with the minimum loss augmented energy.

This inequality leads to the following minimization problem:

$$\begin{aligned} & \min_{\mathbf{w}, \xi_k \geq 0} \lambda \|\mathbf{w}\|^2 + \frac{1}{N} \sum_k \xi_k, \quad (14) \\ & \text{s.t.} \quad \min_{\Delta(\mathbf{x}_k, \hat{\mathbf{y}})=0} \mathbf{w}^\top \psi(\mathbf{x}_k, \hat{\mathbf{y}}) \leq \mathbf{w}^\top \psi(\mathbf{x}_k, \bar{\mathbf{y}}) - \Delta(\mathbf{z}_k, \bar{\mathbf{y}}) + \xi_k, \forall \bar{\mathbf{y}}, \forall k, \end{aligned}$$

where  $\lambda \|\mathbf{w}\|^2$  is a regularization term, preventing overfitting the parameters to the training data.

## 4 Dual-Decomposition Algorithm for the ACI

Briefly, dual decomposition allows us to iteratively solve a convex optimization problem of the form

$$\mathbf{y}^* = \operatorname{argmin}_{\mathbf{y} \in \mathcal{F}} \sum_{m=1}^M g_m(\mathbf{y}). \quad (15)$$

At each iteration  $t$  it solves a set of slaves problems

$$\mathbf{y}_m^* = \operatorname{argmin}_{\mathbf{y}_m \in \mathcal{F}} (g_m(\mathbf{y}_m) + \rho_m^t \mathbf{y}_m), \quad (16)$$

where  $\rho_m^t$  are the dual variables satisfying  $\sum_m \rho_m^t = 0$ . The dual variables are initialized as  $\rho_m^0 = 0, \forall m$ , and updated at iteration  $t$  as follows:

$$\rho_m^{t+1} \leftarrow \rho_m^t + \eta^t (\mathbf{y}_m^* - \sum_n \mathbf{y}_n^* / M), \quad (17)$$

where  $\eta^t$  is the learning rate at iteration  $t$ . Under fairly general conditions, this iterative strategy converges to the globally optimal solution of the original problem, that is,  $\mathbf{y}^* = \mathbf{y}_m^*, \forall m$ . We refer the reader to [3,10] for details.

In order to specify our slave problems, we divide the set of voxels  $\mathcal{V}$  into subsets  $\mathcal{V}_m, m = 1, \dots, M$ , such that each pair of neighboring voxels  $(i, j) \in \mathcal{N}$

appear together in exactly one subset  $\mathcal{V}_m$ . Given such a division of voxels, our slave problems correspond to the following:

$$\min_{\mathbf{y}_m \in \mathcal{C}(\mathcal{V}_m)} \mathbf{y}_m^\top L_m(\mathbf{x}; \mathbf{w}) \mathbf{y}_m + E_m^{\text{prior}}(\mathbf{y}_m, \mathbf{x}; \mathbf{w}) + \rho_m^t \mathbf{y}_m, \quad (18)$$

where  $L_m(\mathbf{x}; \mathbf{w})$  is the Laplacian corresponding to the voxels  $\mathcal{V}_m$ . The prior energy functions  $E_m^{\text{prior}}$  modify the original prior  $E^{\text{prior}}$  by weighing each voxel  $i \in \mathcal{V}_m$  by the reciprocal of the number of subsets  $\mathcal{V}_n$  that contain  $i$ . In other words, the prior term for each voxel  $i \in \mathcal{V}_m$  is multiplied by  $1/|\{\mathcal{V}_n, i \in \mathcal{V}_n\}|$ .

The slave problems defined above can be shown to provide a valid reparameterization of the original problem:

$$\min_{\mathbf{y} \in \mathcal{C}(\mathcal{V})} \mathbf{y}^\top L(\mathbf{x}; \mathbf{w}) \mathbf{y} + E^{\text{prior}}(\mathbf{y}, \mathbf{x}; \mathbf{w}). \quad (19)$$

By using small subsets  $\mathcal{V}_m$  we can optimize each slave problem in every iteration using a standard quadratic programming solver. In our experiments, we used the Mosek solver. To the best of our knowledge, this is the first application of dual decomposition to solve a probabilistic segmentation problem under linear constraints.

## References

1. Andrews, S., Hamarneh, G., Yazdanpanah, A., HajGhanbari, B., Reid, W.D.: Probabilistic multi-shape segmentation of knee extensor and flexor muscles. In: MICCAI, vol. 14, pp. 651–8 (Jan 2011), <http://www.ncbi.nlm.nih.gov/pubmed/22003755>
2. Andrews, S., McIntosh, C., Hamarneh, G.: Convex Multi-Region Probabilistic Segmentation with Shape Prior in the Isometric Log-Ratio Transformation Space. Medical Image Analysis pp. 1–17 (2011), <http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:Convex+Multi-Region+Probabilistic+Segmentation+with+Shape+Prior+in+the+Isometric+Log-Ratio+Transformation+Space#0>
3. Bertsekas, D.: Nonlinear Programming. Athena Scientific (1999)
4. 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), [http://www.isbe.man.ac.uk/~bim/Papers/cootes\\_cviu95.pdf](http://www.isbe.man.ac.uk/~bim/Papers/cootes_cviu95.pdf)
5. Cremers, D., Schmidt, F.R., Barthel, F.: Shape priors in variational image segmentation: Convexity, Lipschitz continuity and globally optimal solutions. 2008 IEEE Conference on Computer Vision and Pattern Recognition pp. 1–6 (Jun 2008), <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=4587446>
6. Delingette, H.: Modélisation, Déformation et Reconnaissance d’objets tridimensionnels a l’aide de maillages simplexes. Thèse de sciences, Ecole Centrale de Paris (July 1994), <http://www.inria.fr/sophia/asclepios/Publications/Herve.Delingette/These-Delingette.pdf>
7. Essafi, S., Langs, G., Paragios, N.: Hierarchical 3D diffusion wavelet shape priors. In: ICCV. pp. 1717–1724. IEEE (Sep 2009), [http://dspace.mit.edu/handle/1721.1/59305http://ieeexplore.ieee.org/xpl/freeabs\\_all.jsp?arnumber=5459385](http://dspace.mit.edu/handle/1721.1/59305http://ieeexplore.ieee.org/xpl/freeabs_all.jsp?arnumber=5459385)

8. Gilles, B., Magnenat-Thalmann, N.: Musculoskeletal MRI segmentation using multi-resolution simplex meshes with medial representations. *Medical image analysis* 14(3), 291–302 (Jun 2010), <http://www.ncbi.nlm.nih.gov/pubmed/20303319>
9. Gilles, B., Pai, D.K.: Fast musculoskeletal registration based on shape matching. In: *MICCAI, Lecture Notes in Computer Science*, vol. 5242, pp. 822–829 (Jan 2008), <http://www.ncbi.nlm.nih.gov/pubmed/18982681>
10. Komodakis, N., Paragios, N., Tziritas, G.: MRF optimization via dual decomposition: Message-passing revisited. In: *ICCV* (2007)
11. Lings, G., Paragios, N., Mas, L., Paris, E.C.D., Galen, E., Saclay, I.: Modeling the structure of multivariate manifolds: Shape Maps (2008)
12. Pizer, S.M., Fletcher, P.T., Joshi, S., Thall, A., Chen, J.Z., Fridman, Y., Fritsch, D.S., Gash, A.G., Glotzer, J.M., Jiroutek, M.R., et al.: Deformable m-reps for 3d medical image segmentation. *International Journal of Computer Vision* 55(2-3), 85–106 (2003)