# Semi-Automatic Prediction of Landmarks on Human Models in Varying Poses

Stefanie WuhrerZouhour Ben AzouzChang ShuNational Research Council of Canada, Ottawa, Ontario, Canada{stefanie.wuhrer, zouhour.benazouz, chang.shu}@nrc-cnrc.gc.ca

### Abstract

We present an algorithm to predict landmarks on 3D human scans in varying poses. Our method is based on learning bending-invariant landmark properties. We also learn the spatial relationships between pairs of landmarks using canonical forms. The information is modeled by a Markov network, where each node of the network corresponds to a landmark position and where each edge of the network represents the spatial relationship between a pair of landmarks. We perform probabilistic inference over the Markov network to predict the landmark locations on human body scans in varying poses. We evaluated the algorithm on 200 models with different shapes and poses. The results show that most landmarks are predicted well.

# 1 Introduction

Currently, full human bodies can be digitized efficiently using off the shelf body scanners. This results in tremendous amounts of information about the shape of the human body. This data can potentially be used to aid in ergonomic design and statistical analysis.

However, processing the surface data collected by body scanners continues to be challenging. To process the data efficiently, intrinsic correspondence information between the body shapes is required. That is, for each point on one surface, one needs to know the corresponding points on the other surfaces. Computing the correspondence information is hard due to incomplete and noisy data and due to pose variations.

Many methods have been proposed to compute intrinsic correspondence information automatically. Methods that do not assume prior information about the shape or pose of the objects [7, 15, 22, 3] are not accurate enough to perform statistical analysis. If a template shape and marker positions are available, reliable correspondences can be obtained. Allen et al. [1] used a template model and marker positions to compute reliable correspondences between human models in similar poses. Xi et al. [21] used a variation

of this approach for statistical analysis. Allen et al. [2] extended their previous approach to allow for changes in pose. Hasler et al. [13] extended Allen et al.'s approach [1] to allow for pose-invariant registration.

Unfortunately, marker positions are not readily available because the marking process significantly increases the time of measurements; from under one minute to over 30 minutes. Hence, this step raises the cost of data collection and is avoided in most projects.

Ben Azouz et al. [5] proposed a method to automatically predict the locations of anthropometric landmarks on human body scans in similar poses. This approach relies heavily on the similarity of poses and fails if the poses vary. This paper introduces a novel method to compute the locations of a small set of landmarks on human body scans in different poses. These landmarks can be used in the landmark-based algorithms, such as Hasler et al. [13] or Allen et al. [2], to compute intrinsic correspondences.

Our method is based on the assumption that different poses of a human model are approximately isometric. This property can be exploited to map an arbitrary pose to a bending-invariant shape, called *canonical form* [11]. Consequently, the problem can be reduced to a single-pose setting. In our method, the single-pose landmark prediction is solved using statistical learning. We model the positions of the landmarks as random variables of a Markov network. The statistical characteristics of the landmarks can be learned through training data. The locations of the landmarks are found by maximizing a joint probability over all possible configurations. This problem can be solved efficiently by using belief propagation [18].

The main contributions of this work are as follows.

- We use bending-invariant canonical forms [11] to learn and predict the positions of landmarks in a postureinvariant way. This results in a semi-automatic way to predict the positions of landmarks on human scans in different poses. The only user-intervention that is required by the algorithm is the choice of the best of eight candidates for the result.
- We propose and use a novel local shape descriptor that

is invariant with respect to isometric deformations of a surface. The descriptor at a vertex v of a triangular mesh S is related to the area of a geodesic circle on S of fixed radius centered at v. The descriptor is related to the descriptor used by Tierny et al. [19] to represent a Reeb chart. However, unlike Tierny et al., we propose to use the geodesic circle for local shape description.

# 2 Related Work

Many methods have been proposed to automatically compute point-to-point correspondences between triangular surfaces. Recently, methods were proposed that do not assume prior marker positions [7, 15, 22, 3]. Some of these methods align bending-invariant canonical forms directly to obtain dense point-to-point correspondences [16, 7]. These methods are currently not accurate enough to perform statistical analysis. When canonical forms are aligned directly, there is no guarantee that close-by points in one shape match close-by points in the other shape.

If a template shape and marker positions are known a priori, accurate correspondences can be computed [1, 2]. The correspondences computed using marker based approaches have the property that close-by points in one shape match close-by points in the other shape, which is desirable. Allen et al. [1] deform a known template mesh of a generic human body to fit a range scan of a human body. The deformation is guided by a small set of known marker positions on the object and it is ensured that the deformation is smooth in the neighborhood of each vertex. Allen et al. [2] extend this approach to work for varying poses by using a skeleton model.

Unfortunately, marker positions are often not available. Ben Azouz et al. [5] propose to find reliable correspondences by automatically predicting marker positions and by using these marker positions to find correspondences. Their method is based on statistical learning. It works for models in similar poses, but fails if the pose variation is large.

This paper introduces a novel method to semiautomatically compute the locations of a small set of landmarks on human body scans in different poses. These semiautomatically placed landmarks can be combined with the algorithm by Hasler et al. [13] or Allen et al. [2] to compute intrinsic correspondences that have the property that closeby points in one shape match close-by points in the other shape.

Our approach makes use of *canonical forms* [11]. Elad and Kimmel define the canonical form X of a surface S as the mapping of S to  $\mathbb{R}^3$ , such that the Euclidean distances between the mapped vertices approximate the geodesic distances between the original vertices well. The canonical form is computed via multi-dimensional scaling with the geodesic distances between vertices on the triangular manifold as dissimilarities. Hence, the canonical form of a nonrigid body is pose-invariant. We use fast marching [17] to compute geodesic distances on S. We then use least-squares multi-dimensional scaling [6, p.146-155] to compute the canonical form in  $\mathbb{R}^3$ . For increased space efficiency, we compute the canonical form using a coarse-to-fine strategy as outlined by Wuhrer et al. [20]. The approach by Wuhrer et al. consists of two steps. First, n' vertices of S are used to compute a canonical form at low resolution. Second, the remaining vertices of S are added to the canonical form one by one by minimizing a least-squares energy function.

We use the database of humans in varying poses by Hasler et al. [13] to learn and predict the landmarks. The database contains 550 laser scans of 114 subjects in up to 35 poses. Hasler et al. computed the correspondences between the subjects of the database using a variation of the approach by Allen et al. [1] by manually placing landmarks on the scans. We use the registered database to manually select eight landmarks. We use the selected landmarks for training and as ground truth.

Our approach is conceptually similar to the approach by Fergus et al. [12] to locate features in two-dimensional images for object recognition in that a Bayesian approach is used to model the properties of features and the spatial relationship between features. While we only model pairwise spatial relationships between features, it is possible to model spatial relationships between features using the star model by Crandall et al. [9].

### **3** Overview

We model the problem of predicting the locations of landmarks as an inference problem on a Markov network. We first learn the properties of a small set of manually placed landmarks. With this knowledge, we can then predict the locations of the landmarks on a surface S using statistical inference.

To learn the location of the landmarks on a surface S in a pose-invariant way, we use the canonical form of S. To learn the properties of the landmarks on a surface S in a pose-invariant way, we represent each vertex of S by a descriptor that depends on the intrinsic geometry of S. The descriptor at a vertex v of S uses a measure related to the area of the geodesic circle of radius r centered at v.

# 4 Markov Network

We model the problem of predicting the locations of a set of k landmarks on the surface S of a human body as a probabilistic inference problem on a pairwise Markov network [10]. A Markov network is an undirected graph

G(V, E) that models the joint probability distribution of the set of random variables. In our case, the random variables are the locations of the k landmarks  $L = l_0, \ldots, l_{k-1}$ . Each node v in V represents one of the random variables. Each edge e in E represents a dependency between the random variables represented by the endpoints of e. We associate a potential  $\phi_i(l_i)$  corresponding to the likelihood that landmark  $l_i$  is located at a given vertex of S with each node  $v_i$  in V. Furthermore, we associate a potential  $\psi_{i,j}(l_i, l_j)$  corresponding to the joint likelihood that landmark  $l_i$  is located at another given vertex of S with each edge e at another given vertex of S with each edge  $e = (l_i, l_j)$  in E. The joint probability of the network is

$$p(L) = \frac{1}{Z} \prod_{i} \phi_i(l_i) \prod_{i,j} \psi_{i,j}(l_i, l_j), \qquad (1)$$

where Z is a normalizing factor.

In this paper, we consider the landmarks shown as red spheres in Figure 1. That is, we consider as landmarks the hands, elbows, feet, head, and crotch of the models. These landmarks are usually sufficient to allow a satisfactory registration [14]. We model the correlations between the positions of the landmarks using the graph structure G(V, E) shown as black lines in Figure 1. This graph is obtained manually and is not necessarily optimal with respect to the correlations between the landmarks. Note that the proposed method does not consider the graph on the original mesh, but on the bending-invariant canonical form of the original mesh.



Figure 1. Landmarks and structure of landmark graph.

It remains to define the potentials  $\phi_i(l_i)$  and  $\psi_{i,j}(l_i, l_j)$ .

#### 4.1 Node Potential

Our goal is to construct a node potential that characterizes the surface properties of the location of a landmark in a pose-invariant way. We achieve this goal by using a novel local surface descriptor that is based on the intrinsic geometry of S. That is, we choose as node potential

$$\phi_i(l_i = v_j) = \mathcal{N}(D(v_j), \mu_i^{node}, \Sigma_i^{node}), \qquad (2)$$

where  $\mathcal{N}$  is a multivariate Gaussian distribution with mean  $\mu_i^{node}$  and covariance  $\Sigma_i^{node}$ , and  $D(v_j)$  is a pose-invariant surface descriptor. The mean  $\mu_i^{node}$  and the covariance  $\Sigma_i^{node}$  are learned based on a training set as outlined in Section 5.

We use as surface descriptor  $D(v_k)$  a measure related to the area A(c) of the geodesic circle c of radius r centered at  $v_k$ . The geodesic circle c is topologically either equivalent to a disk or to a torus. Following Tierny et al. [19], we compute the distortion

$$d(r) = \begin{cases} \frac{A(c)}{\pi r^2} \text{ if } c \text{ is topologically equivalent to a disk,} \\ \frac{A(c)}{\pi ((1+r)^2 - 1)} \text{ otherwise.} \end{cases}$$

The surface descriptor  $D(v_k)$  is a vector of distortions  $d(r_i)$  obtained by varying the radius of the geodesic circle. In our experiments,  $D(v_k)$  has dimension twenty. Figure 2 visualizes the node descriptor at the elbow of one of the models. The node descriptor is obtained by computing the distortion of the geodesic disks shown in red with respect to the corresponding Euclidean disks shown in blue. Note that we use d(r) to locally describe landmark locations while Tierny et al. use a set of distortions d(r) to globally characterize a shape.

The left of Figure 3 shows the geodesic isolines at the landmark positions on one of the human models. The right of Figure 3 shows projections of the descriptor values  $D(v_k)$  on 50 models of the database in different postures. We project along a fixed random direction to obtain this figure. Descriptors of corresponding landmarks are shown in the same color. We can see that points of the same color form clusters. This shows that the descriptor describes the landmark positions well.

#### 4.2 Edge Potential

We choose an edge potential based on spatial relationships between vertices on the canonical form X. That is,

$$\psi_{i,j}(l_i = v_l, l_j = v_m) = \mathcal{N}(D(v_l, v_m), \mu_{i,j}^{edge}, \Sigma_{i,j}^{edge}),$$
(3)

where  $\mathcal{N}$  is a multivariate Gaussian distribution with mean  $\mu_{i,j}^{edge}$  and covariance  $\Sigma_{i,j}^{edge}$ , and  $D(v_l, v_m)$  is a descriptor



**Figure 3.** Properties of node potential. Left: geodesic isolines at the landmark positions. Middle: detail view of geodesic isolines at right elbow. Right: projections of descriptor values on 50 human models. Descriptors of corresponding landmarks are shown in the same color.



**Figure 2.** Visualization of the node descriptor. Geodesic circles are shown in red and Euclidean circles are shown in blue.

on X. The mean  $\mu_{ij}^{edge}$  and the covariance  $\Sigma_{i,j}^{edge}$  are learned based on a training set as outlined in Section 5.

The descriptor  $D(v_l, v_m)$  consists of the distance and the difference vector between the vertices corresponding to  $v_l$  and  $v_m$  on X. Let  $X(v_i)$  be the vertex of X corresponding to  $v_i$  and let  $X(v_i)$  be the position vector of that vertex. The descriptor  $D(v_l, v_m) =$  $\left[ \|X(v_l) - X(v_m)\|, (X(v_l) - X(v_m))^T \right]^T$  is of dimension four, where  $\|X(v_l) - X(v_m)\|$  denotes the Euclidean distance between  $X(v_l)$  and  $X(v_m)$ .

Note that the edge potential is purely based on the

canonical form X and that the node potential is isometryinvariant. This has the effect that the potentials are invariant with respect to deformations that preserve geodesic distances.

### 5 Learning

We use 200 models  $S_0, \ldots, S_{199}$  of the database by Hasler et al. [13] to learn the variables of the node and edge potentials given in Equations 2 and 3. The training set contains models in all 35 poses. We learn all of the means  $\mu_i^{node}$  and  $\mu_{i,j}^{edge}$  and covariances  $\Sigma_{i,ode}^{node}$  and  $\Sigma_{i,j}^{edge}$  using maximum likelihood estimation [10].

Furthermore, we learn the average positions  $\overline{p}(l_i)$  of the landmarks on the canonical forms  $X_0, \ldots, X_{199}$ . These positions are helpful to restrict the search space. To learn the average positions, we need to align the canonical forms. A canonical form is invariant with respect to translation, rotation, and reflection [8]. Hence, we need to consider multiple alignments. We do this by aligning all of the canonical forms  $X_0, \ldots, X_{199}$  by their respective eigenvectors and by normalizing the height of the canonical forms. It remains to consider eight alignments because each canonical form  $X_i$  in  $\mathbb{R}^3$  is invariant with respect to reflection and because there are  $2^3 = 8$  possible reflections of X. We align  $X_1, \ldots, X_{199}$  to  $X_0$  in turn as follows. For  $X_i$ , we consider all eight sign assignments of the eigenvectors. For each sign assignment, we compute the sum of squared differences between the corresponding landmark positions on  $X_i$  and  $X_0$ . We find the best alignment as the one that minimizes the sum of squared differences. The learned positions  $\overline{p}(l_i)$  are shown in Figure 4. Note that the graph is nearly

symmetric in three dimensions. The legs in the graph are almost contained in the x, y-plane and the arms in the graph are almost contained in the x, z-plane.



**Figure 4.** *Left: a model in the database. Middle: canonical form of the model. Right: Projected land-mark graph obtained using 200 subjects.* 

### 6 Predicting Locations of Landmarks

When predicting landmarks on the surface S of a human body, our goal is to compute the landmark locations that maximize Equation 1. We use Bayesian belief propagation to find a solution. The Bayesian belief propagation algorithm was introduced by Pearl [18]. The algorithm can be applied to a Markov network G(V, E) if each node  $v_i$ in V is associated with a discrete number of possible labels  $F_i$ . The labels of node  $v_i$  represent candidate locations on S for landmark  $l_i$ . When running Belief propagation on a Markov network, beliefs, or probabilities of certain hypotheses given some evidence, are propagated through the network by local update rules. Note that if the Markov network has a tree structure as in our application, the Bayesian belief propagation algorithm finds the exact solution to the problem of finding landmark locations that maximize Equation 1. The size of the label sets  $F_i$  has a significant influence on the efficiency of the belief propagation algorithm. It is therefore important to choose appropriate label sets  $F_i$ .

We use the learned average positions  $\overline{p}(l_i)$  to find label sets  $F_i$  as follows. We first align the canonical form X along its eigenvectors. As X is invariant with respect to rotation, translation, and reflection, we need to consider eight possible alignments. We predict the locations of the landmarks using all eight possibilities and let the user choose the correct result. Note that this is the only user-intervention required by the algorithm.

For each of the eight alignments, we project the canonical form X of S into the plane  $\pi_{x,y}$  defined by the first two principal components of X and into the plane  $\pi_{x,z}$  defined by the first and the third principal components of X. Furthermore, we project  $\overline{p}(l_i)$  into  $\pi_{x,y}$  and  $\pi_{x,z}$ . Denote the projected canonical forms by  $X^{\pi_{x,y}}$  and  $X^{\pi_{x,z}}$  and the projected average positions by  $\overline{p}(l_i)^{\pi_{x,y}}$  and  $\overline{p}(l_i)^{\pi_{x,z}}$ . For the crotch landmark, we compute the label set  $F_i$  as the nearest neighbors of  $\overline{p}(l_i)$  in X. For the landmarks at the heels, we compute a label set  $F_i$  as the nearest neighbors of  $\overline{p}(l_i)^{\pi_{x,y}}$ in  $X^{\pi_{x,y}}$ . For all the other landmarks, we compute a label set  $F_i$  as the nearest neighbors of  $\overline{p}(l_i)^{\pi_{x,z}}$ . We use three kd-trees to speed up the nearest neighbor search [4].

We compute  $F_i$  in  $\pi_{x,y}$  and  $\pi_{x,z}$  for most of the landmarks to avoid problems due to canonical forms that are not rigid to each other. For some models, the left leg is in the front in the canonical form and for other models, the left leg is in the back. When computing  $F_i$  using nearest neighbors in three dimensions, the label set for a landmark on the left leg may therefore only include vertices on the right leg. A similar problem may happen with the arms of the model. These problems are avoided when computing  $F_i$  in  $\pi_{x,y}$  and  $\pi_{x,z}$ .

Figure 5 illustrates the label sets  $F_i$ . Figure 5(a) shows the landmarks on a model of the database. Each landmark  $l_i$  is assigned a unique color. Figure 5(b) shows the label sets  $F_i$  on a model in a different pose. Each vertex of the mesh belonging to  $F_i$  is shown in the same color as  $l_i$  in Figure 5(a). If a vertex belongs to more than one label set, the corresponding colors are interpolated.

## 7 Experiments

We predict the landmarks on 200 models of the database in different poses that were not used for learning. Figure 6 shows the predicted landmarks on five different models. Red spheres correspond to the manually marked true landmarks and green spheres correspond to the predicted landmarks. For some landmarks, only the green spheres are visible. In these cases, the learned and predicted landmarks are identical.

We use node descriptors  $D(v_k)$  of dimension twenty. Experimentally, we found that the higher the dimension of the node descriptor, the higher the accuracy of the result and the higher the running time of the algorithm. This gives a way to trade off accuracy and running time.

Table 1 shows the average, the standard deviation, and the maximum of the location errors for each landmark. We notice that for all of the landmarks, the average error is less than 10cm.

### 8 Conclusions

This paper presented a new approach to predict the locations of landmarks in a pose-invariant way. The approach combines statistical learning, bending-invariant canonical forms, and a novel local surface descriptor.

We leave the following ideas for future work:



**Figure 5.** Illustration of label sets  $F_i$ .

Landmark	Average	Standard	Maximum
	(mm)	Deviation	(mm)
		(mm)	
1 Head	58.28	3.128	147.3
2 Crotch	77.34	0.315	205.6
3 Right Hand	14.92	1.277	66.7
4 Right Elbow	85.65	1.136	200.3
5 Left Hand	10.12	2.091	57.06
6 Left Elbow	93.83	0.4291	197.7
7 Right Heel	42.59	0.7731	165.2
8 Left Heel	37.88	0.5053	148.3

**Table 1.** Error of prediction computed over 200 test human scans.

- The canonical forms of two human bodies with different topologies, large holes, or additional parts are not similar. Hence, manual preprocessing may be required in these cases. It is possible to automatically compute a label set in these cases.
- We designed a graph that connects the landmarks manually. This graph may be computed automatically.
- We model the node and edge potentials as Gaussian distributions. Other distributions may be considered.

## Acknowledgments

We thank Nils Hasler for providing us with the MPI database.

# References

- B. Allen, B. Curless, and Z. Popović. The space of human body shapes: reconstruction and parameterization from range scans. *ACM Transactions on Graphics*, 22(3):587– 594, 2003. Proceedings of SIGGRAPH.
- [2] B. Allen, B. Curless, Z. Popović, and A. Hertzmann. Learning a correlated model of identity and pose-dependent body shape variation for real-time synthesis. In ACM SIG-GRAPH/Eurographics Symposium on Computer Animation, pages 147–156, 2006.
- [3] D. Anguelov, P. Srinivasan, D. Koller, S. Thrun, H.-C. Pang, and J. Davis. The correlated correspondence algorithm for unsupervised registration of nonrigid surfaces. In *Neural Information Processing Systems*, 2004.
- [4] S. Arya and D. M. Mount. Approximate nearest neighbor queries in fixed dimensions. In ACM-SIAM Symposium on Discrete Algorithms, pages 271–280, 1993.
- [5] Z. B. Azouz, C. Shu, and A. Mantel. Automatic locating of anthropometric landmarks on 3d human models. In *3D Data Processing, Visualization and Transmission*, 2006.
- [6] I. Borg and P. Groenen. *Modern Multidimensional Scaling Theory and Applications*. Springer, 1997.
- [7] A. M. Bronstein, M. M. Bronstein, and R. Kimmel. Generalized multidimensional scaling: a framework for isometryinvariant partial surface matching. *Proceedings of the National Academy of Science*, 103(5):1168–1172, 2006.
- [8] T. Cox and M. Cox. Multidimensional Scaling, Second Edition. Chapman & Hall CRC, 2001.
- [9] D. Crandall, P. Felzenszwalb, and D. Huttenlocher. Spatial priors for part-based recognition using statistical models. In *Conference on Computer Vision and Pattern Recognition*, pages 10–17, 2005.
- [10] R. Duda, P. Hart, and D. Stork. Pattern Classification, Second Edition. John Wiley & Sons, Inc., 2001.
- [11] A. Elad and R. Kimmel. On bending invariant signatures for surfaces. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(10):1285–1295, 2003.



**Figure 6.** Results of landmark prediction on five models. Red spheres correspond to the manually marked true landmarks and green spheres correspond to the predicted landmarks. For some landmarks, only the green spheres are visible. In these cases, the learned and predicted landmarks are identical.

- [12] R. Fergus, P. Perona, and A. Zisserman. Object class recognition by unsupervised scale-invariant learning. In *Conference on Computer Vision and Pattern Recognition*, pages 264–271, 2003.
- [13] N. Hasler, C. Stoll, M. Sunkel, B. Rosenhahn, and H.-P. Seidel. A statistical model of human pose and body shape. In P. Dutré and M. Stamminger, editors, *Computer Graphics Forum (Proc. Eurographics 2008)*, volume 2, 2009.
- [14] N. Hasler, C. Stoll, T. Thormählen, B. Rosenhahn, and H.-P. Seidel. Estimating body shape of dressed humans. *Comput*ers & Graphics, 2009. (to appear).
- [15] Q. Huang, B. Adams, M. Wicke, and L. J. Guibas. Nonrigid registration under isometric deformations. *Computer Graphics Forum (Special Issue of Symposium on Geometry Processing 2008)*, 27(5), 2008.
- [16] V. Jain, H. Zhang, and O. van Kaick. Non-rigid spectral correspondence of triangle meshes. *International Journal* on Shape Modeling, 13(1):101–124, 2007.

- [17] R. Kimmel and J. Sethian. Computing geodesic paths on manifolds. *Proceedings of the National Academy of Science*, 95:8431–8435, 1998.
- [18] J. Pearl. Probabilistic Reasoning in Intelligent Systems. Morgan Kaufmann, 1988.
- [19] J. Tierny, J.-P. Vandeborre, and M. Daoudi. Reeb chart unfolding based 3D shape signatures. In *Eurographics*, pages 13–16, 2007.
- [20] S. Wuhrer, C. Shu, Z. B. Azouz, and P. Bose. Posture invariant correspondence of incomplete triangular manifolds. *International Journal on Shape Modeling*, 13(2):139–157, 2007.
- [21] P. Xi, W.-S. Lee, and C. Shu. Analysis of segmented human body scans. In *Graphics Interface*, 2007.
- [22] H. Zhang, A. Sheffer, D. Cohen-Or, Q. Zhou, O. van Kaick, and A. Tagliasacchi. Deformation-driven shape correspondence. *Computer Graphics Forum (Special Issue of Symposium on Geometry Processing 2008)*, 27(5), 2008.