Pacific Graphics 2017
J. Barbic, W.-C. Lin, and O. Sorkine-Hornung
(Guest Editors)

Volume 36 (2017), Number 7

Regression-Based Landmark Detection on Dynamic Human Models

Deok-Kyeong Jang and Sung-Hee Lee

Korea Advanced Institute of Science and Technology (KAIST)

Abstract

Detecting anatomical landmarks on various human models with dynamic poses remains an important and challenging problem in computer graphics research. We present a novel framework that consists of two-level regressors for finding correlations between human shapes and landmark positions in both body part and holistic scales. To this end, we first develop pose invariant coordinates of landmarks that represent both local and global shape features by using the pose invariant local shape descriptors and their spatial relationships. Our body part-level regression deals with the shape features from only those body parts that correspond to a certain landmark. In order to do this, we develop a method that identifies such body parts per landmark, by using geometric shape dictionary obtained through the bag of features method. Our method is nearly automatic, as it requires human assistance only once to differentiate the left and right sides. The method also shows the prediction accuracy comparable to or better than those of existing methods, with a test data set containing a large variation of human shapes and poses.

Categories and Subject Descriptors (according to ACM CCS): I.3.5 [Computer Graphics]: Computational Geometry and Object Modeling—Geometric algorithms, languages, and systems

1. Introduction

Analyzing and understanding human body shapes are important problems in computer graphics research, with a number of applications such as registration, retargeting, and shape retrieval. Anatomical landmarks on human bodies are essential features for obtaining anthropometric information, but detecting the anatomical landmarks remains a challenging problem due to the high variedness of human shapes and poses.

A conventional approach for landmark detection involves finding correspondences between a template body model with annotated landmarks and a particular body shape, typically through mesh registration methods [CR03]. This registration-based approach is effective for the body shapes similar to the template, but has a limited capability in generalizing to the whole range of human shapes with different poses. A more principled and potentially powerful approach would be to learn and predict the relationship between various body shapes and their landmark locations. In addition, a good method should not require complex preprocessing such as alignments on the input data.

In this paper, we solve the anatomical landmark detection problem on human models with dynamic poses by training a statistical regression model that learn connections between human body shapes and landmarks. When trained with a data set of a wide range of human shapes and poses, the regression-based method achieves a higher accuracy than registration-based approaches. Specifically, we show that the kernel canonical correlation analysis (KCCA) method successfully models the correlation between human body shapes and landmarks.

Under this regression framework, we develop several key ideas that enable robust landmark detection against severe variations in shape and pose. First, we develop a novel method that serves as pose invariant coordinates of landmarks. This is achieved by representing the position of a landmark with the feature vector in the pose invariant, local descriptor space. In addition, since the mapping from vertices to local descriptor space is non-injective (e.g., left and right Stylions have the same feature vector), we augment this representation with the spatial relationship information between landmarks, which is also described in a pose invariant manner.

Second, in order to increase the regression accuracy, we develop a two-level regression method that separately regresses body part-level features and holistic features. The part-level regressor models the connection between the local feature of a landmark and the shape of the body parts that are related with the landmark. To identify such body parts per landmark, we develop a method that uses geometric shape dictionary obtained through the supervised bag of features (S-BoF) method. The segmentation of body parts allows for training the lower-level regressor with respect to only body parts that are related with the landmark, and thus blocks the effect from other unrelated body parts. We will show later that the body part segmentation brings an additional benefit that reduces the search space of landmarks. The global level regressor relates the holistic shape characteristics with the spatial relationship between landmarks, and this overcomes the non-injectivity of local features and allows for the nearly automatic finding of landmarks.

Contributions. The main contribution of our approach is in the novel framework of regression of locating landmarks on challeng-

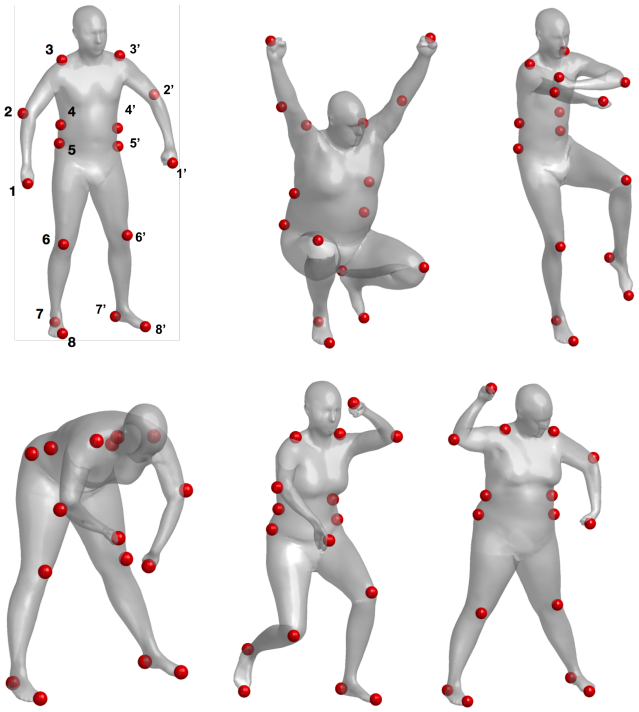

Figure 1: *Human models in dynamic poses in the dataset with annotated landmarks: Proximal phalanges (1), Radiale (2), Acromiale (3), Iliocristale (4), Trochanterion (5), Patella (6), Calcaneus (7) and Hallux (8). See top left figure for the landmark indices.*

ing dynamic human models. The framework provides the following advantages. First, through regression with KCCA, we can not only search for landmarks for a body shape that can be registered by a template model, but also find the landmarks for various body shapes. Second, our shape descriptors that combine the global and body part-level shape features are invariant to isometric deformation, and thus allow for various pose changes. The pose invariance also makes any alignment such as registration, scale normalization, and the origin adjustment unnecessary. Lastly, despite the short online computation time, the accuracy of landmark prediction is reasonably good and comparable to or better than other existing methods.

Data Set. We used a data set from SHREC'14 Track [PSR*14]. The data set consists of 400 human models from 40 human subjects (20 male and 20 female) from slim to obese shapes, each in 10 different dynamic poses. For landmark recognition, we annotated seven landmarks (Proximal phalanges, Radiale, Acromiale, Iliocristale, Trochanterion, Patella, Calcaneus and Hallux) manually on both sides of the body (Fig. 1). Data with this broad spectrum of body shapes and poses can be of great help in training a landmark detector that generalizes to a wide range of human subjects. All the human models and the annotated landmark data can be downloaded from http://www.cs.cf.ac.uk/shaperetrieval/shrec14/ and http://motionlab.kaist.ac.kr/, respectively.

2. Related Work

Due to its importance in shape analysis, researchers have developed various methods to describe local and global features of object shapes. Spin image [JH99], heat kernel signature [SOG09, BK10], and wave kernel signature [ASC11] are some examples of local descriptors that represent the geometric information of a small neighbor or a point. Global descriptors represent the overall features of a shape. Among them, those methods created by combining local descriptors are usually used in shape retrieval problems. Representative global descriptors include ShapeDNA [RWP06, RCB*16], histograms of area projection transforms [GL12], and deep aggregation of localized statistical features [FO16]. The most common way to create a global descriptor is to use the bag of feature (BoF) method [LBBC14, BBGO11, NNT*15, TCF10, LGSX13], which represents the global characteristics of a shape in terms of the occurrence frequencies of local shape features, analogous to the bag-of-words methods in text retrieval. Among them, we use the method of [LBBC14] in our preprocess stage to define a shape descriptor through supervised learning.

The problem of predicting anthropometric landmarks has been receiving growing attention. Giachetti et al. reported state-of-art technologies for landmark prediction [GMP*14]. Among them, the surface-to-surface registration method [CR03] and graphical model method [ASM06] showed better performances than other compared techniques. [CR03] finds the landmark points through a nonrigid spatial mapping from a template mesh to an input mesh. The nonrigid ICP method is used for registration, but in general, such purely geometric mesh registration methods do not guarantee the human-likeness of the registered meshes. The method also requires repetitive alignment steps. [ASM06] uses the Markov network to represent the structure of landmarks. The method shows a high accuracy when trained with a database of uniform poses (A-poses) with a relatively high computation time. Both methods find correspondences between the average data (e.g., template mesh) and the input data. Therefore, the accuracy degrades as the input data gets dissimilar to the average data. Moreover, these methods are not robust to varying poses because of the limitations of the nonrigid ICP method [CR03] or the usage of pose variant local features [ASM06]. In contrast, we develop a regression-based method that models the correlation between the body shapes and the landmark positions by using pose invariant local features. Thus, our method is more robust to shape and pose variations.

Methods based on spectral analysis are independent of pose changes. However, the accuracies reported in [LH13a, LH13b] are not competitive yet. Wuhrer et al. [WAS10, WSX11] improved the method of [ASM06] to account for varying poses. Compared to these methods, our method shows a higher accuracy, even with more challenging landmarks (e.g., Iliocristale and Trochanterion) and poses (e.g., squatting).

Kernel canonical correlation analysis (KCCA) is a method that finds the relationship between the two multidimensional variables [HSST04]. In computer graphics, KCCA has been used for facial expression recognition [ZZZZ06, Hor07] and facial retargeting [SCSN11]. For example, Fen et al. used KCCA to simultaneously control the deformation of a large number of local regions with a small number of control points [FKY08]. In this paper, we show that KCCA successfully finds connections between the 3D body shapes and landmarks.

3. Overview

Our goal is to detect landmarks on human models with various shapes and poses, and we address this problem by developing a pose invariant regression method of landmarks that depend on body shapes. Our strategy is to perform KCCA for two types of data pairs. One is between SBD (Segmented Body Part Descriptor) and landmark positions in the local descriptor space, and the other is between the global shape descriptor and spatial relationships between landmarks. In order to do this, the connection between the human body shapes and landmark locations is learned from a training data set. The overall process of our method is divided into three stages: preprocess, learning and regression for the locating of landmarks, and landmark detection. Figure 2 illustrates the overview of our method.

In the preprocess stage, we define a global shape descriptor for a human shape and a segmented body part descriptor (SBD) that represents parts of the human body shape related to a particular landmark's (Sec.4.1) and create pose invariant coordinates of landmarks that can define the landmark locations regardless of the dynamic poses (Sec.4.2). In the learning and regression stage (Sec.5), we find the connection between the body descriptors and coordinates of landmarks, which were generated from the preprocess stage for the training of human data set. Our learning strategy is to perform KCCA for two types of data pairs. One is between SBD and landmark position in local descriptor space, and the other is between global shape descriptor and spatial relationships between landmarks. After this stage, we can use a trained predictor to regress the coordinates of landmark corresponding to the input human body shape. Finally, in the landmark detection stage (Sec.6), we use the regressed coordinates of landmarks to predict where the landmarks are actually located in the given human body shape. Each stage is detailed in the following sections.

4. Preprocess

All landmarks in the training human models are labeled $\{1,\dots,8\}$ and $\{1',\dots,8'\}$ in close order from the wrist (Proximal) to the ankle (Hallux) on both sides. Note that training models do not need to be aligned, which is a strong advantage of our method. We assume that various pose models of the same subject are created by isometric deformation. This assumption is needed in calculating the geometric distance when defining the relationship between landmarks. The location of the landmark is only dependent on the human's body shape. Other factors such as pose should not affect landmark detection. To this end, we define pose invariant descriptors for human body shapes and also pose invariant coordinate system for the landmark.

4.1. Pose Invariant Body Shape Descriptors

Since the geometric properties of a landmark is more strongly correlated with its local region than other distant regions (e.g., a landmark on the wrist depends more on the shape of the wrist than on that of the elbow), defining an appropriate descriptor for a local region that includes the landmark is important for the landmark detection. Thus, we define the segmented body part descriptor that characterizes the geometric feature of the body part. As a key idea for this, we use the characteristics of a global shape descriptor that uses the bag of features: Each atom of the geometric dictionary obtained from the bag of features method represents some body parts. Therefore, by collecting only the atoms associated with the body parts that a landmark belongs to, we can define the local body part descriptors.

The first step of this is to define a BoF-based global shape descriptor, which is discussed next, followed by explaining the procedure for defining segmented body part descriptors.

4.1.1. Local Descriptor

A local descriptor represents the geometric structure of a point in the small neighborhood. Among many local descriptors, we use the scale invariant heat kernel signature (SI-HKS), which is widely used to define the intrinsic geometric features of object shapes [BK10]. Its invariance to isometric deformation makes the descriptor independent of human pose changes. The SI-HKS is an improved version of the heat kernel signature (HKS) [SOG09], which solves the heat equation on manifold M with m vertices $(x_1,\dots,x_m)$ that use the cotangent weight.

$$(\Delta_M + \tfrac{\partial}{\partial t})\,u(x,t) = 0$$

where Δ_M denotes the Laplace-Beltrami operator and $u(x,t)$ is the heat distribution on vertex x of manifold M at time t. Given some heat distribution $u(x,0) = \delta$ (dirac-delta) at time $t = 0$ as an initial condition, one solution to the above equation is the *heat kernel*:

$$h_t(x,x) = \textstyle\sum_{k\geq 1} e^{-v_k t}\phi_k^2$$

Here, ϕ_k and v_k are the eigenfunctions and eigenvalues obtained by the eigen-decompostion of Laplacian. The heat kernel can serve as a local feature of the manifold's geometry. For example, a n-dimensional local descriptor for each vertex x can be defined as $(h_{t_1}(x,x), h_{t_2}(x,x),\dots,h_{t_n}(x,x))$. By replacing this local descriptor with a frequency $\omega_1,\dots,\omega_n$ through the Fourier transform, a heat kernel signature becomes invariant to scale (See [BK10] for details):

$$\mathbf{p}(x) = (|H(\omega_1)|,\dots,|H(\omega_n)|)^\top, \tag{1}$$

which is used as our local descriptor for a vertex x. The SI-HKS for all the m vertices of the human body shape M is then defined as a $n \times m$ matrix

$$\mathbf{P}(M) = (\mathbf{p}(x_1),\dots,\mathbf{p}(x_m)). \tag{2}$$

4.1.2. Global Body Shape Descriptor

We use the bag of features as the global shape descriptor. In general, the bag of feature method creates a geometric dictionary matrix $\mathbf{D} = (\mathbf{d}_1 \dots \mathbf{d}_v) \in \mathbb{R}^{n\times v}$ first, and then expresses the local descriptors $\mathbf{P} \in \mathbb{R}^{n\times m}$ (Eq. (2)) as codes $\mathbf{Z}(\mathbf{P},\mathbf{D}) \in \mathbb{R}^{v\times m}$ with respect to $\mathbf{D}$. Finally, a v-dimensional global shape descriptor $\mathbf{f}(\mathbf{P},\mathbf{D})$ is obtained by pooling the codes $\mathbf{f}(\mathbf{P},\mathbf{D}) = \mathbf{Z}(\mathbf{P},\mathbf{D})\mathbf{h}$ where $\mathbf{h}$ is a pooling term, weighting each point proportionally to its area a_i, $\mathbf{h} = (a_1,\dots,a_m)^\top / \sum_{i=1}^m a_i$.

For a global shape descriptor, we use the method of [LBBC14], which improves the standard BoF that constructs a dictionary in an unsupervised manner by using clustering for all the training human body shapes $\mathbf{M}$, and by learning both the dictionary and the

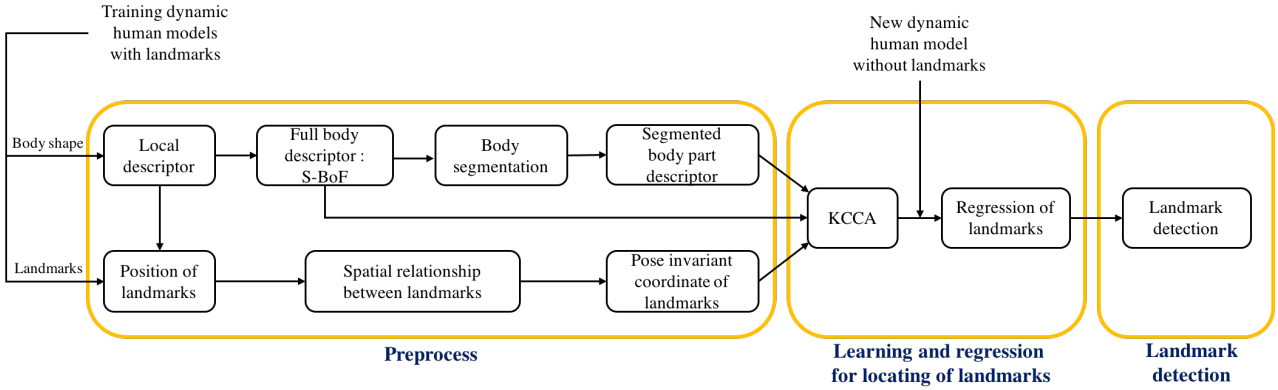

Figure 2: *The pipeline of our method. It consists of three major stages: preprocess, learning and regression, and detection of landmarks.*

sparse codes in a supervised manner through the bi-level optimization with the objective function below:

$$\begin{aligned}
\mathbf{Z}^*(\mathbf{P},\mathbf{D}) &= \underset{\mathbf{D}}{\operatorname{argmin}} \sum_{M\in\mathbf{M}} \ell((\underset{\mathbf{Z}}{\operatorname{argmin}} ||\mathbf{P}-\mathbf{DZ}||_F^2 + \lambda\,||\mathbf{Z}||_1)\cdot\mathbf{1}),\\
\ell &= \alpha\ell_+ + (1-\alpha)\ell_-\\
\ell_+(\mathbf{Z},\mathbf{Z}_+) &= ||\,\mathbf{f}(\mathbf{Z})-\mathbf{f}(\mathbf{Z}_+)||_1\\
\ell_-(\mathbf{Z},\mathbf{Z}_+,\mathbf{Z}_-) &= \max\{0,\mu+||\,\mathbf{f}(\mathbf{Z})-\mathbf{f}(\mathbf{Z}_+)||_1\\
&\quad -||\,\mathbf{f}(\mathbf{Z})-\mathbf{f}(\mathbf{Z}_-)||_1\}
\end{aligned} \tag{3}$$

where $\ell(\cdot)$ is the hinge loss that tries to separate by at least μ between the dissimilarity of the positive and the negative pairs. The codes $\mathbf{Z}$ and $\mathbf{Z}_+$ are from the same subject, affected by some transformation, and $\mathbf{Z}_-$ is from different subject. The weight factor λ is for regularization and α is controlled for trade-off between the false positive and false negative rates. The solution to Eq. (3) produces a dictionary $\mathbf{D}$ that optimally separates between the BoFs of positive and negative pairs, and the optimal code

$$\mathbf{Z}^*(\mathbf{P}) = (\mathbf{z}^*(\mathbf{p}_1),\ldots,\mathbf{z}^*(\mathbf{p}_m)), \tag{4}$$

for each $\mathbf{P}(M)$.

4.1.3. Segmented Body Part Descriptor Corresponding to Landmark

We have defined the global shape descriptor $\mathbf{f}$ that features the whole body shape, independent of its pose. Here, we will define another shape descriptor that expresses the shape of the body parts related with a particular landmark.

The supervised bag of feature method provides the geometric dictionary $\mathbf{D} = (\mathbf{d}_1 \ldots \mathbf{d}_v)$, in which the column vectors $\mathbf{d}_i$ are geometric atoms that represent v number of body parts. In general, the body parts are not mutually exclusive, and thus several atoms are associated with a particular landmark. Our strategy is to collect only those atoms that are related with a landmark and define the landmark's shape descriptor with these atoms.

Suppose that x_i is the vertex of the landmark i, and its code is

$$\mathbf{z}^*(\mathbf{p}_i) = (z_1^*,\ldots,z_v^*)^\top. \tag{5}$$

Note that z_i^* indicates the weight of $\mathbf{p}_i$ corresponding to $\mathbf{d}_i$. Therefore, the atoms that have non-zero values can be considered to be related with $\mathbf{p}_i$. Thus, our method to identify a subset of atoms that are associated with a landmark i is as follows: For every mesh in the training set, we compute $\mathbf{p}_i$ for x_i followed by obtaining $\mathbf{z}^*(\mathbf{p}_i)$, from which we find the atoms with non-zero weights. Among all the meshes in the training set, if an atom has non-zero weights for more than a certain ratio (62.5% in our experiment), the atom is determined to be related to the landmark. Among the 64 atoms in total, the number of atoms related with each landmark are 5 (Proximal phalanges), 4 (Radiale), 6 (Acromiale), 9 (Iliocristale), 8 (Trochanterion), 3 (Patella), 6 (Calcaneus), and 4 (Hallux).

From this analysis, we can define the *segmented body part descriptor*, dubbed SBD_i for a landmark i,

$$\mathrm{SBD}_i(M) = (\mathbf{z}_i^*(\mathbf{p}_1),\ldots,\mathbf{z}_i^*(\mathbf{p}_m))\cdot\mathbf{h} \tag{6}$$

where $\mathbf{z}_i^*(\mathbf{p})$, which is a subvector of $\mathbf{z}^*(\mathbf{p})$, includes only the elements corresponding to the atoms related with a landmark i. Figure 3 shows the vertices that correspond to the atoms related with each landmark by visualizing $||\mathbf{z}_i^*(\mathbf{p}_j)||$ for each vertex x_j with color for a landmark i. The SBD has strong advantages in learning and landmark detection. Since the SBD only describes features of the only body parts that are related with a landmark, predicting the landmark based on SBD is not affected by other unrelated parts, which may be the case with the prediction based on the global shape descriptors. Regarding the landmark detection, SBD significantly reduces the candidate vertices, making detection performed quickly.

4.2. Pose Invariant Coordinates of Landmark

We need appropriate coordinates that are invariant to isometric deformation, pose, as well as the number and order of vertices, in order to detect landmarks, regardless of such changes. The world coordinates are not invariant to isometric deformation and pose changes, and the vertex index is not invariant to the vertex size and order changes. Our key idea to define the landmark coordinates is to use both the landmark's position and the spatial relation between the landmarks in the space of the local shape descriptor. A limitation of this approach, however, is that the coordinates are symmetrical in the sagittal plane, e.g., the left and right ankles have

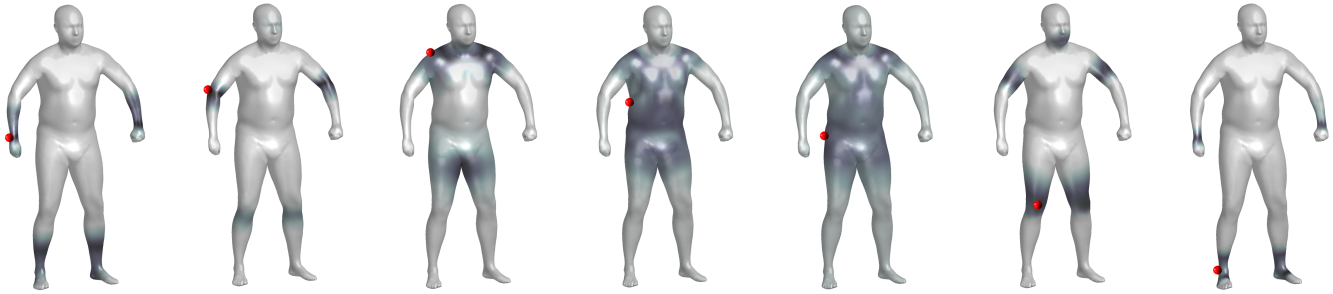

Figure 3: *The segmented body parts corresponding to each landmark. The color of a vertex x_j for a landmark i is set darker proportional to $||\mathbf{z}_i^*(\mathbf{p}_j)||$.*

the same coordinates. To overcome this problem, we additionally use the geodesic distance between the landmarks.

4.2.1. Local Descriptor as Landmark Position

The local descriptor $\mathbf{p}(x_i)$ is a point in $\mathbb{R}^n$ and it satisfies some important properties as coordinates. Figure 4 shows the first three coordinates of $\mathbf{p}$ from the landmarks of 160 human subjects. One can see that the points form clusters per landmark, which indicates that the local descriptor, if used as the landmark coordinate, can discriminate between the landmarks. In addition, since $\mathbf{p}$ is invariant to isometric deformation, it is independent of pose changes. However, the map $x_i \rightarrow \mathbf{p}(x_i)$ from a vertex x_i to $\mathbf{p}(x_i)$ is not injective, i.e., many points on a body can be mapped to the same position in the local descriptor space. To cope with this, we introduce additional elements to the coordinates, namely the spatial relationship between landmarks, as discussed next.

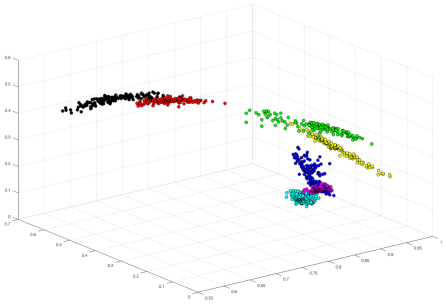

Figure 4: *The first three dimensions of the local descriptors of landmarks. Same colors indicate the same landmarks.*

4.2.2. Spatial Relationship between Landmarks

Due to the similarity in shapes throughout humans, the spatial relationships between landmarks show consistent patterns. We use these characteristics to detect the landmark positions. For this, we consider three types of spatial relationships.

First, we use the relative positions between landmarks in the local descriptor space. Specifically, we use

$$\mathbf{V}_i = ((\mathbf{p}_i - \mathbf{p}_1)^\top,\ (\mathbf{p}_i - \mathbf{p}_{i-2})^\top,\ (\mathbf{p}_i - \mathbf{p}_{i-1})^\top)^\top, \tag{7}$$

where $\mathbf{p}_i$ is the local descriptor corresponding to landmark i. The positions of $\mathbf{p}_i$ relative to $\mathbf{p}_{i-2}$ and $\mathbf{p}_{i-1}$ are selected by the assumption that neighboring landmarks show a higher correlation than distant landmarks. The relative position between $\mathbf{p}_i$ and $\mathbf{p}_1$ is considered to be consistent with the geodesic distance, which will be discussed next. Note that we compare $\mathbf{p}_i$ with its predecessors $\mathbf{p}_{i-1}$, $\mathbf{p}_{i-2}$, and $\mathbf{p}_1$ because we detect landmarks in increasing order from the landmark 1. Thus, for example $\mathbf{p}_i - \mathbf{p}_{i-2}$ in Eq. 7 is ignored for landmark 2.

Since $\mathbf{p}$ is the same for the landmarks on the left and right sides, $\mathbf{V}_i$ is also symmetrical in a mid-sagittal plane. To distinguish the symmetrical landmarks, we consider an additional distance measure, which is the geodesic distance between the landmarks. The geodesic distance is invariant to isometric deformation and thus this feature is sensible to our purpose as changes in human poses are considered to induce near isometric skin deformation. Similarly to the relative distances in the local descriptor space, we consider geodesic distances from a landmark i to $i-1$, $i-2$, and 1.

$$\mathbf{g}_i = (d(x_i,x_1),\ d(x_i,x_{i-2}),\ d(x_i,x_{i-1}))^\top, \tag{8}$$

where x_i is a vertex in shape M that corresponds to a landmark i. The geodesic distance between $\mathbf{a}$ and $\mathbf{b}$ is denoted by $d(\mathbf{a},\mathbf{b})$.

Lastly, in order to differentiate the symmetrical landmarks in the detection stage, we consider the geodesic distances between them.

$$\mathbf{u} = (d(x_1,x_{1'}),\ \ldots,\ d(x_s,x_{s'}))^\top, \tag{9}$$

where x_i and $x_{i'}$ are locations of the symmetrical landmarks. Figure 5 shows the geodesic distances between landmarks, and the values of the first three dimensions of $\mathbf{p}(x_i)$ for each vertex x_i.

Our method to define the spatial relationships between landmarks is similar to the methods in [ASM06, WAS10, WSX11]. [ASM06] defines the spatial relationship in terms of Euclidean distance in $\mathbb{R}^3$ space, which is not invariant to rigid transformations or pose changes. [WAS10, WSX11] use both the Euclidean distance and the difference vector between local features in the descriptor space. This choice is somewhat redundant because the difference between two vectors contains the notion of distance already. Compared with these approaches, our method uses two largely independent components, the difference vectors in a local descriptor space and geodesic distance, which makes more sense.

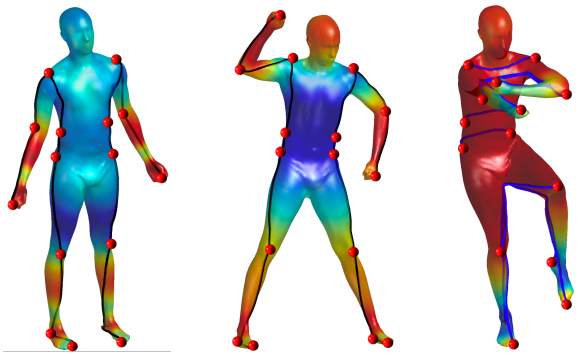

Figure 5: *The colors plotted on the body shape are the values of vertices in the first three dimensions of the local descriptor space, from the left to right. The values show a left-right symmetry even for asymmetric poses. The black line represents the geodesic distances between the closest landmarks i and $i-1$ on the same side. The blue line represents the geodesic distance between the symmetric landmarks i and i'.*

5. Regression for Locating Landmarks

The global and segmented body part shape descriptors, as well as the pose invariant coordinates defined in the previous section are used to locate landmarks. Our approach to this problem is to learn a regression model between the two. After learning, the global and body part shape descriptors for an input body shape allow us to find suitable landmark coordinates, which are then used to locate the landmark.

Since the shape descriptors and landmark coordinates are both high dimensional and nonlinear, we use the kernel canonical correlation analysis (KCCA) technique as a regression model. In KCCA-based regression, the estimation of landmark coordinates is performed using the connection parameter found in KCCA. That is, when a new human shape is given, we perform the regression based on learned connection parameter. Then, we can get the landmark coordinates that consist of the position factor $\mathbf{p}_i$ and the spatial relationship factors $\mathbf{V}_i$, $\mathbf{g}_i$ $(i = 1,\ldots,s)$, and $\mathbf{u}$. Regression process is performed respectively on the left and right sides.

5.1. Two-Level Nonlinear Correlation Analysis

Kernel CCA maps nonlinear variables to a feature space of a higher dimension and examines the correlation between two variables in that space. Our strategy is to perform KCCA for two types of data pairs. One is between SBD and landmark positions in the local descriptor space, and the other is between the global shape descriptor and spatial relationships between landmarks. The former investigates the relation between local features while the latter focuses on the relation between the global features. The detailed procedure for our KCCA is provided next.

5.1.1. Analysis on Position

Before performing the KCCA, we define a kernel function, $k(\mathbf{x},\mathbf{y}) = \phi(\mathbf{x})^\top \phi(\mathbf{y})$, where ϕ is a projection mapping that transforms $\mathbf{x}$ into a higher dimensional space. Our goal for each landmark i, to find the connection between the SBD_i that represents the body part shape and the landmark position in local shape descriptor space. Given $|M_{\text{train}}|$ pairs of training data $\{\mathrm{SBD}_i,\ \mathbf{p}_i\}_{s=1}^{|M_{\text{train}}|}$ for each landmark i, we construct matrices $\mathbf{Q}_i = (\mathrm{SBD}_i^1,\ldots,\mathrm{SBD}_i^{|M_{\text{train}}|})$ and $\mathbf{R}_i = (\mathbf{p}_i^1,\ldots,\mathbf{p}_i^{|M_{\text{train}}|})$ as data with each column that correspond to each training data. Then, the kernel is computed over $\mathbf{Q}_i$ and $\mathbf{R}_i$.

$$\begin{aligned}\mathbf{K}_{Q,i} &= \phi(\mathbf{Q}_i)^\top \phi(\mathbf{Q}_i),\\ \mathbf{K}_{R,i} &= \phi(\mathbf{R}_i)^\top \phi(\mathbf{R}_i).\end{aligned} \tag{10}$$

According to [HSST04], the projection directions $\mathbf{w}_{Q,i}$ and $\mathbf{w}_{R,i}$ for $\mathbf{Q}_i$ and $\mathbf{R}_i$ satisfy $\mathbf{w}_{Q,i} \in span(\phi(\mathbf{Q}_i))$ and $\mathbf{w}_{R,i} \in span(\phi(\mathbf{R}_i))$. So we can represent a pair of directions as $\mathbf{w}_{Q,i} = \phi(\mathbf{Q}_i)\alpha_i$ and $\mathbf{w}_{R,i} = \phi(\mathbf{R}_i)\beta_i$, where α_i and β_i are weights. Therefore, we can write the nonlinear version of CCA using the kernel functions as follows:

$$\max_{\alpha_i,\beta_i} \rho = \max_{\alpha_i,\beta_i} \frac{\alpha_i^\top \mathbf{K}_{Q,i}\mathbf{K}_{R,i}\beta_i}{\sqrt{\alpha_i^\top (\mathbf{K}_{Q,i})^2\alpha_i \beta_i^\top (\mathbf{K}_{R,i})^2\beta_i}} \tag{11}$$

The solution to this correlation problem yields:

$$\mathbf{A}_i = [\alpha_i^1,\ldots,\alpha_i^t],\ \mathbf{B}_i = [\beta_i^1,\ldots,\beta_i^t]$$

where $\{\alpha^j,\beta^j\}_{j=1}^t$ are t pair directions corresponding to $\rho_i^1,\ldots,\rho_i^t$ correlation values (see [HSST04] for the detailed solution). Let $\mathrm{proj}(\mathbf{K}_{Q,i})$ and $\mathrm{proj}(\mathbf{K}_{R,i})$ be the projections of kernel $\mathbf{K}_{Q,i}$ and $\mathbf{K}_{R,i}$ onto the projection matrices $\mathbf{A}_i$ and $\mathbf{B}_i$:

$$\mathrm{proj}(\mathbf{K}_{Q,i}) = \mathbf{K}_{Q,i}\mathbf{A}_i, \quad \mathrm{proj}(\mathbf{K}_{R,i}) = \mathbf{K}_{R,i}\mathbf{B}_i$$

Assume $\rho_i^1,\ldots,\rho_i^t$ is similar to 1. Then, we suppose that $\mathrm{proj}(\mathbf{K}_{Q,i})$ and $\mathrm{proj}(\mathbf{K}_{R,i})$ are linearly correlated. Finally, we can compute linear coefficient matrix $\mathbf{C}_{\text{PO}}^i$ such that

$$\mathbf{C}_{\text{PO}}^i = \mathrm{proj}(\mathbf{K}_Q)^\top\ \mathrm{proj}(\mathbf{K}_R). \tag{12}$$

The above $\mathbf{C}_{\text{PO}}^i$ is for landmark i, so we perform KCCA for each landmark and obtain $\mathbf{C}_{\text{PO}}^1$, …, $\mathbf{C}_{\text{PO}}^s$.

5.1.2. Analysis on Spatial Relationships

Now we analyze the correlation between the full body shape descriptor $\mathbf{f}(\mathbf{P})$ (Sec.4.1.2) and the spatial relationships $\mathbf{V}$, $\mathbf{g}$, $\mathbf{u}$. The analysis takes the same process as the one for landmark coordinates. However, the only difference is that KCCA is only performed twice here whereas the previous analysis requires multiple KCCAs per landmark. Specifically, given $|M_{\text{train}}|$ pairs of training data $\{\mathbf{f}(\mathbf{P})^s,\mathrm{SP}^s\}_{s=1}^{|M_{\text{train}}|}$ and $\{\mathbf{f}(\mathbf{P})^s,\mathbf{u}^s\}_{s=1}^{|M_{\text{train}}|}$, where $\mathrm{SP} = (\mathbf{V}_1^\top,\mathbf{g}_1^\top,\ldots,\mathbf{V}_s^\top,\mathbf{g}_s^\top)^\top$, we construct matrices $\mathbf{Q} = (\mathbf{f}(\mathbf{P})^1,\ldots,\mathbf{f}(\mathbf{P})^{|M_{\text{train}}|})$, $\mathbf{R} = (\mathrm{SP}^1,\ldots,\mathrm{SP}^{|M_{\text{train}}|})$ and $\mathbf{U} = (\mathbf{u}^1,\ldots,\mathbf{u}^{|M_{\text{train}}|})$ as data. After the analysis we obtain the linear coefficient matrix $\mathbf{C}_{\text{SP}}$ and $\mathbf{C}_{\text{SM}}$ between the full body shape descriptor $\mathbf{f}(\mathbf{P})$ and the spatial relationships.

5.2. Regression of Landmark Features

As the output of correlation analysis on the training data, we have obtained the linear coefficients $\mathbf{C}_{\text{PO}}^1$, …, $\mathbf{C}_{\text{PO}}^s$ between the segmented body part and landmark positions, and $\mathbf{C}_{\text{SP}}$, $\mathbf{C}_{\text{SM}}$ between

the full body descriptor and the spatial relationship between the landmarks. For $\mathbf{C}_{\mathrm{PO}}$ and $\mathbf{C}_{\mathrm{SP}}$, KCCA is performed on each of the left and right sides, and thus $2(s+1)$ coefficients are obtained. Together with a coefficient for $\mathbf{C}_{\mathrm{SM}}$, the total number of coefficients is $2(s+1)+1$. Now, given a new human model M_{new}, we will predict the landmark position (in a local shape descriptor space) and their spatial relationships by regression using these coefficients. Similarly to the analysis process, the regression is performed in two stages.

Algorithm 1 shows the procedure for predicting landmark positions $\mathbf{p}_i^{\mathrm{new}}$ for a new input shape M_{new}. For each landmark, it first computes $\mathrm{SBD}^{\mathrm{new}}$, and then kernel $\mathbf{K}_Q^{\mathrm{new}}$, which is projected to $\mathbf{A}$ and multiplied with $\mathbf{C}_{\mathrm{PO}}$ to predict the projection of $\mathbf{K}_R^{\mathrm{new}}$ (line 4). This regression is performed in the a high dimensional feature space and not for the local shape descriptor space. The inverse mapping from $\mathbf{K}_R$ to $\mathbf{p}$ is achieved by training a radial basis function network (RBFN), which is a widely used model for function approximation.

Algorithm 1 Regression for Landmark Coordinates

Require: $\mathbf{C}_{\mathrm{PO}}^i$ and M_{new}
1: **for all** landmarks $i = 1,\ldots,s$ **do**
2: compute $\mathrm{SBD}_i^{\mathrm{new}}$ (Sec.4.1.3)
3: compute $\mathbf{K}_{Q,i}^{\mathrm{new}}$
4: $\mathrm{proj}(\mathbf{K}_{R,i}^{\mathrm{new}}) \leftarrow \mathbf{C}_{\mathrm{PO}}^i\ \mathrm{proj}(\mathbf{K}_{Q,i}^{\mathrm{new}})$ (Sec. 5.1.1)
5: $\mathbf{p}_i^{\mathrm{new}} \leftarrow f_{\mathrm{RBFN}}(\mathrm{proj}(\mathbf{K}_{R,i}^{\mathrm{new}}))$
6: **return** $\mathbf{p}_i^{\mathrm{new}}$

Regression for spatial relationships is performed in the same manner, with $\mathrm{SBD}^{\mathrm{new}}$ replaced with $\mathbf{f}(\mathbf{P}^{\mathrm{new}})$, and $\mathbf{C}_{\mathrm{PO}}^i$ with $\mathbf{C}_{\mathrm{SP}}$ and $\mathbf{C}_{\mathrm{SM}}$.

Through the regression procedure, we have obtained the landmark positions in the local shape descriptor space $\mathbf{p}_1^{\mathrm{new}},\ldots,\mathbf{p}_s^{\mathrm{new}}$ and the spatial relationships between the landmarks $\{\mathbf{V}_i^{\mathrm{new}},\mathbf{g}_i^{\mathrm{new}}\}_{i=1}^s$, $\mathbf{u}^{\mathrm{new}}$. We additionally have obtained $\mathbf{p}_{1'}^{\mathrm{new}},\ldots,\mathbf{p}_{s'}^{\mathrm{new}}$ and $\{\mathbf{V}_{i'}^{\mathrm{new}},\mathbf{g}_{i'}^{\mathrm{new}}\}_{i'=1}^{s'}$ for other side of the body. All of them will be used to detect landmark vertices as explained next.

6. Landmark Detection

We detect vertices x_i and $x_{i'}$ simultaneously for the landmarks i and i' by solving the optimization problem:

$$\begin{aligned} &\underset{x_i\in\mathbf{X}_i}{\operatorname{argmin}} ||\mathbf{p}_i^{new}-\mathbf{p}(x_i)||^2 \quad \text{subject to} \\ &||\mathbf{g}_i^{new}-\mathbf{g}(x_i)||_w^2 \le \varepsilon_{\mathrm{geo}},\ ||\mathbf{V}_i^{new}-\mathbf{V}(x_i)||_w^2 \le \varepsilon_{\mathrm{vec}}, \\ &||\mathbf{u}_i^{new}-\mathbf{u}_i(x_i,x_{i'})||_w^2 \le \varepsilon_{\mathrm{sym}}, \end{aligned} \tag{13}$$

where $\mathbf{X}_i$ is a set of vertices in the segmented body part corresponding to the landmark i, and $\varepsilon_{\mathrm{geo}}$, $\varepsilon_{\mathrm{vec}}$ and $\varepsilon_{\mathrm{sym}}$ are thresholds. $\mathbf{u}_i$ denotes a i's element of $\mathbf{u}$ (Eq. 9). The weighted Euclidean distance $||\cdot,\cdot||_w^2$ is computed as

$$||\mathbf{V}^{new}-\mathbf{V}(x)||_w^2 = \sum_i w_i ||\mathbf{v}_i^{new}-\mathbf{v}_i(x)||^2, \tag{14}$$

where $\sum_i w_i = 1$ and $\mathbf{v}_i(x)$ denote a column vector of $\mathbf{V}$ (Eq. 7). The weights w_i are set higher for a closer landmark as it has higher correlation than more distant landmarks.

The detection procedure consists of two steps. We first collect the segmented body part vertex set $\mathbf{X}_i$ that corresponds to the landmark i, and then find the landmark vertex from $\mathbf{X}_i$. The pseudocode for constructing $\mathbf{X}_i$ is provided in Algorithm 2. Note that the segmented body part function $\mathbf{z}_i^*$ is used to check whether a vertex x is to be included in $\mathbf{X}_i$ or not (line 4).

Algorithm 2 Collecting vertices contained in segmented body part

Require: M_{new} and $\mathbf{z}_i^*$(Sec.4.1.3)
1: Compute $\mathbf{p}(M_{\mathrm{new}})$
2: $\mathbf{X}_i \leftarrow \emptyset$
3: **for all** vertices x included in M_{new} **do**
4: **if** $\left|\left|\mathbf{z}_i^*(\mathbf{p}(x))\right|\right| > 0$ **then**
5: $\mathbf{X}_i \leftarrow \{x\} \cup \mathbf{X}_i$
6: **return** $\mathbf{X}_i$

The segmented body part $\mathbf{X}_i$ obtained in Algorithm 2 significantly reduces the number of candidate vertices for finding a landmark. In addition, it increases the detection accuracy by precluding vertices in unrelated body parts. After having found $\mathbf{X}_i$, Eq. 13 is solved to find the position of a landmark. Specifically, we compute $\mathbf{p}(x_i)$ for all $x_i \in \mathbf{X}_i$ and find a vertex that satisfies the constraints in the increasing order of $||\mathbf{p}_i^{new}-\mathbf{p}(x_i)||$.

Detection for other side of the landmark is performed in the same manner, with the index i being replaced with i'. Once this detection procedure is completed, landmarks i and i' are identified.

Detecting the First Landmark Landmarks are found in the increasing order from $(1,1')$ to (s,s') by solving Eq. 13. When finding the first landmark (proximal phalanges), we solve Eq. 13 using only symmetry constraint term since we cannot use the spatial relations. However, finding a landmark only with $\mathbf{p}$ does not differentiate between the left and right sides. Thus, for the first landmark only, our system visualizes a total of 30 candidate landmarks, from which a user selects one from the right side. The locations of candidate landmarks are typically very concentrated. Thus, a user is practically required to select either a left or right side. All the subsequent landmarks are then found in a fully automatic manner.

7. Experiment

Our data set includes 400 dynamic human models from 20 male and 20 female subjects, each making 10 poses. Each model consists of nearly 15K vertices. We constructed a training data set of 160 models by randomly selecting four poses from each subject. After that, we manually marked seven landmarks (i.e., proximal phalanges, radiale, acromiale, iliocristale, trochanterion, patella, calcaneus and hallux) on both sides of the human body. We created a test data set with 210 poses by selecting a pose randomly from each subject from the remaining 240 models.

Finding landmarks from the data set is very challenging since human models have a variety of shapes and some of the landmarks are difficult to find, even with the human eye. In addition, since the data sets include dynamic poses, the positions of the landmarks vary significantly.

All parameters of our approach were obtained through a standard 10-fold cross validation procedure. When calculating the S-BoF in the preprocess stage, all the parameter values are the same

as in [LBBC14]. The difference is that we used a higher resolution data, so the dictionary size v is 64 and λ is 0.25. In the learning and regression stages, the kernel function used in the KCCA was a *Gaussian kernel*, whose width was set to the maximum of the Euclidean distances of input variables among all the training data set. Lastly, the thresholds ε_{geo}, ε_{sym} and ε_{vec} were set to 7, 7 and 0.09, respectively after some experiments. The weights w_1, w_2, and w_3 representing the degree of influence depending on the spatial relationship between landmarks were set to 0.25, 0.25, and 0.5 respectively.

We applied our method to 210 test human models. Figure 6 shows the results of the located landmarks for dynamic poses of a variety of human models. One can see that our method estimates landmark positions with reasonably good accuracy. For each landmark, the difference between the ground truth and the predicted landmarks was calculated using both geodesic and Euclidean distances.

Table 1: *Average, standard deviation, and median of geodesic (left) and Euclidean (right) distances between the estimated landmarks and the ground truth data.*

Landmark	mean(mm)		st.dev.(mm)		median(mm)	
R. proximal phalanges	51.66	39.66	26.36	17.44	51.58	41.05
L. proximal phalanges	61.51	50.22	23.54	15.06	58.85	51.65
R. radiale	71.15	58.75	39.67	24.58	68.31	61.85
L. radiale	64.75	56.17	37.49	26.47	62.29	58.19
R. acromiale	66.76	62.43	37.79	33.66	62.86	59.99
L. acromiale	71.21	65.57	48.99	42.09	56.76	53.96
R. iliocristale	68.61	57.51	60.35	28.42	54.74	52.74
L. iliocristale	51.74	49.70	31.79	30.38	45.82	44.43
R. trochanterion	58.21	55.75	36.44	34.86	50.06	48.27
L. trochanterion	53.18	50.86	30.80	29.26	46.20	44.30
R. patella	86.48	77.29	51.95	42.89	85.00	78.53
L. patella	90.06	81.57	43.58	36.82	87.41	82.83
R. calcaneus	32.85	30.09	14.20	12.53	30.61	29.01
L. calcaneus	24.78	23.05	12.35	11.03	23.18	21.74
R. hallux	16.06	14.15	10.06	8.58	13.36	11.85
L. hallux	16.82	15.26	14.19	12.80	12.45	11.09

Table 1 shows the average, standard deviation, and median of distance errors for each landmark. The average error of all landmarks is less than 9*cm* in geodesic distance and 8.2*cm* in Euclidean distance. Overall, our method is mostly superior with respect to performance and computation time to [WSX11] that share a similar purpose with our paper although [WSX11] tested against less challenging landmarks for less varying poses (See Fig. 8)[†]. Figure 7 compares the accuracy of [WSX11] and ours in terms of the errors in the Euclidean distance. Our method marks lower error for radiale, acromiale, and patella while [WSX11] is better for proximal phalanges and hallux. Compared with other landmarks, in our database, the local areas near proximal phalanges and hallux are rugged and show high variedness caused by the mesh reconstruction of thin parts (See Fig. 9). This is in contrast with the data of [WSX11] in which the reconstructed meshes are smoother in these areas. We conjecture that the different complexities of the data might have caused our accuracy for these landmarks to be lower than that of [WSX11].

[†] Note that we cannot make exact comparison with ours and [WSX11] because the landmarks and the complexities of the human models in the tested databases differ. We made a rough comparison with respect to twelve landmarks (proximal phalanges, radiale, acromiale, patella, calcaneus, and hallux) that overlapped with our landmarks.

Computation Time. The method was implemented in MATLAB. For 40 test human models, it took approximately 6 hours to complete the preprocess stage with the majority of time spent for learning a supervised bag of features, 0.9 seconds for KCCA learning, 0.2 seconds for regression, and 7099 seconds for landmark detection using 2.6 GHZ Intel Core i7 CPU and 16GB 2133MHz memory Mac. The online learning, regression, and detection stages were performed reasonably fast; it took 33.3 seconds per human model for 16 landmarks. [WSX11] takes about three minutes per human model to detect 6 landmarks.

8. Conclusion and Future Work

In this paper, we introduced novel methods for landmark detection using BoF and KCCA. The main contributions of our work are the segmented body part descriptor using geometric dictionary, and the pose invariant landmark coordinates combining the landmark position in the local descriptor space with the spatial relationships between landmarks. Our method allows for efficient detection of landmarks independently of alignment, vertex ordering, and pose changes.

Our method has several limitations in the following aspects. First, it detects landmarks in a sequential manner. This strategy allows faster performance than finding all the landmarks at once, but may have negative effects on the accuracy because errors can accumulate in detecting the latter part of landmarks. Second, our assumption on the isometry of human deformation is not perfectly valid. Human body deformation is largely isometric, but our examination on the real human scan data shows that the degree of non-isometry is not negligible. This non-isometric deformation induced in pose changes degrades the accuracy of predicting geodesic distances between landmarks. Developing appropriate methods that overcome this error would require investigating the features of which behaviors can be accurately predicted from dynamic poses. If we find this feature, our approach will have a much higher accuracy and may be completely automatic. This remains our future research goal.

There are many additional future research directions. While we only deal with the landmarks that are located on the surface, our method can be used for the regression of points inside the body as well. A good application would be to predict the locations of joints from the body shape, e.g., for skinning purposes [BCBiR*15]. In addition, our landmark detection method can be used for various applications, such as global shape alignment that can be performed around the detected landmarks, as well as registration and one-to-one correspondence detection using the matched landmarks. Lastly, it may help subject-specific simulation [KPMP*17] through finding joint positions through regression.

Acknowledgement

This work was supported by the Seoul Olympic Sports Promotion Foundation, Ministry of Culture, Sports and Tourism, Republic of Korea, and by the "Cross-Ministry Giga KOREA Project" funded by MSIT, Republic of Korea (GK17P0200). We thank Roee Litman for sharing his MATLAB codes for S-BoF.

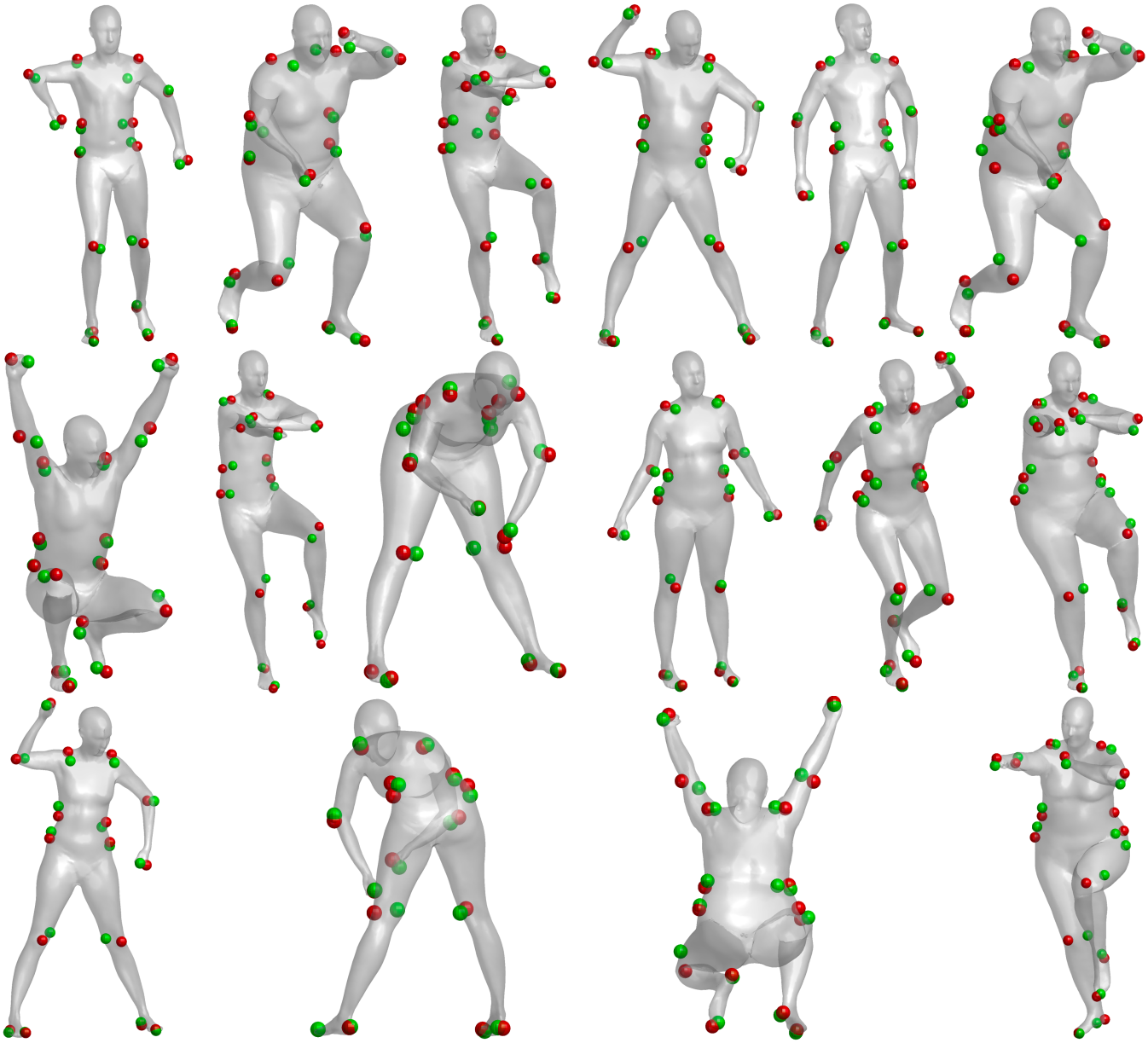

Figure 6: *Results of landmark detection on 16 subject human models (8 female and 8 male) with dynamic poses. Red markers are ground truth landmarks annotated manually and green markers are estimated landmarks using our method.*

References

[ASC11] AUBRY M., SCHLICKEWEI U., CREMERS D.: The wave kernel signature: A quantum mechanical approach to shape analysis. In *Computer Vision Workshops (ICCV Workshops), 2011 IEEE International Conference on* (2011), IEEE, pp. 1626–1633. 2

[ASM06] AZOUZ Z. B., SHU C., MANTEL A.: Automatic locating of anthropometric landmarks on 3d human models. In *3D Data Processing, Visualization, and Transmission, Third International Symposium on* (2006), IEEE, pp. 750–757. 2, 5

[BBGO11] BRONSTEIN A. M., BRONSTEIN M. M., GUIBAS L. J., OVSJANIKOV M.: Shape google: Geometric words and expressions for invariant shape retrieval. *ACM Transactions on Graphics (TOG) 30*, 1 (2011), 1. 2

[BCBiR*15] BANG S., CHOI B., BLANCO I RIBERA R., KIM M., LEE S.-H., NOH J.: Interactive rigging with intuitive tools. *Computer Graphics Forum 34*, 7 (2015), 123–132. 8

[BK10] BRONSTEIN M. M., KOKKINOS I.: Scale-invariant heat kernel signatures for non-rigid shape recognition. In *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on* (2010), IEEE, pp. 1704–1711. 2, 3

[CR03] CHUI H., RANGARAJAN A.: A new point matching algorithm for non-rigid registration. *Computer Vision and Image Understanding 89*, 2 (2003), 114–141. 1, 2

[FKY08] FENG W.-W., KIM B.-U., YU Y.: Real-time data driven defor-

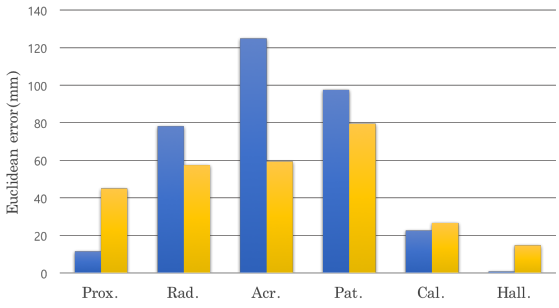

Figure 7: *Comparison of the Euclidean error of landmark detection between [WSX11] (blue) and ours (yellow).*

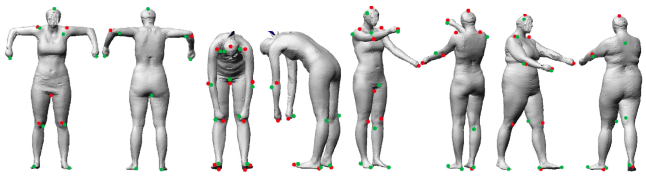

Figure 8: *Result figure from [WSX11]. Landmarks are less challenging and pose variations are not as severe as our data set.*

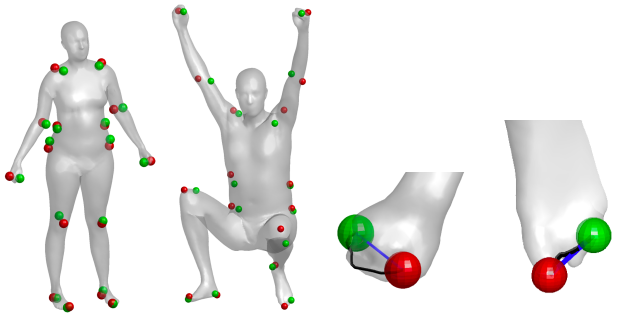

Figure 9: *Cases with large errors in detecting proximal phalanges and hallux. Geodesic (black) and Euclidean (blue) distance errors are shown in the right.*

mation using kernel canonical correlation analysis. In *ACM Transactions on Graphics (TOG)* (2008), vol. 27, ACM, p. 91. 2

[FO16] FURUYA T., OHBUCHI R.: Deep aggregation of local 3d geometric features for 3d model retrieval. 2

[GL12] GIACHETTI A., LOVATO C.: Radial symmetry detection and shape characterization with the multiscale area projection transform. In *Computer Graphics Forum* (2012), vol. 31, Wiley Online Library, pp. 1669–1678. 2

[GMP*14] GIACHETTI A., MAZZI E., PISCITELLI F., AONO M., HAMZA A. B., BONIS T., CLAES P., GODIL A., LI C., OVSJANIKOV M., ET AL.: Shrec'14 track: automatic location of landmarks used in manual anthropometry. In *Eurographics Workshop on 3D Object Retrieval* (2014), pp. 93–100. 2

[Hor07] HORIKAWA Y.: Facial expression recognition using kcca with combining correlation kernels and kansei information. In *Computational Science and its Applications, 2007. ICCSA 2007. International Conference on* (2007), IEEE, pp. 489–498. 2

[HSST04] HARDOON D. R., SZEDMAK S., SHAWE-TAYLOR J.: Canonical correlation analysis: An overview with application to learning methods. *Neural computation 16*, 12 (2004), 2639–2664. 2, 6

[JH99] JOHNSON A. E., HEBERT M.: Using spin images for efficient object recognition in cluttered 3d scenes. *IEEE Transactions on pattern analysis and machine intelligence 21*, 5 (1999), 433–449. 2

[KPMP*17] KIM M., PONS-MOLL G., PUJADES S., BANG S., KIM J., BLACK M. J., LEE S.-H.: Data-driven physics for human soft tissue animation. *ACM Trans. Graph. 36*, 4 (July 2017), 54:1–54:12. 8

[LBBC14] LITMAN R., BRONSTEIN A., BRONSTEIN M., CASTELLANI U.: Supervised learning of bag-of-features shape descriptors using sparse coding. In *Computer Graphics Forum* (2014), vol. 33, Wiley Online Library, pp. 127–136. 2, 3, 8

[LGSX13] LIAN Z., GODIL A., SUN X., XIAO J.: Cm-bof: visual similarity-based 3d shape retrieval using clock matching and bag-of-features. *Machine Vision and Applications 24*, 8 (2013), 1685–1704. 2

[LH13a] LI C., HAMZA A. B.: Intrinsic spatial pyramid matching for deformable 3d shape retrieval. *International Journal of Multimedia Information Retrieval 2*, 4 (2013), 261–271. 2

[LH13b] LI C., HAMZA A. B.: A multiresolution descriptor for deformable 3d shape retrieval. *The Visual Computer 29*, 6-8 (2013), 513–524. 2

[NNT*15] NGUYEN V.-T., NGO T. D., TRAN M.-T., LE D.-D., DUONG D. A.: A combination of spatial pyramid and inverted index for large-scale image retrieval. *International Journal of Multimedia Data Engineering and Management (IJMDEM) 6*, 2 (2015), 37–51. 2

[PSR*14] PICKUP D., SUN X., ROSIN P. L., MARTIN R., CHENG Z., LIAN Z., AONO M., HAMZA A. B., BRONSTEIN A., BRONSTEIN M., ET AL.: Shrec'14 track: Shape retrieval of non-rigid 3d human models. In *Proceedings of the 7th Eurographics workshop on 3D Object Retrieval* (2014), vol. 1, p. 6. 2

[RCB*16] RODOLÀ E., COSMO L., BRONSTEIN M. M., TORSELLO A., CREMERS D.: Partial functional correspondence. In *Computer Graphics Forum* (2016), Wiley Online Library. 2

[RWP06] REUTER M., WOLTER F.-E., PEINECKE N.: Laplace–beltrami spectra as 'shape-dna'of surfaces and solids. *Computer-Aided Design 38*, 4 (2006), 342–366. 2

[SCSN11] SONG J., CHOI B., SEOL Y., NOH J.: Characteristic facial retargeting. *Computer Animation and Virtual Worlds 22*, 2-3 (2011), 187–194. 2

[SOG09] SUN J., OVSJANIKOV M., GUIBAS L.: A concise and provably informative multi-scale signature based on heat diffusion. In *Computer graphics forum* (2009), vol. 28, Wiley Online Library, pp. 1383–1392. 2, 3

[TCF10] TOLDO R., CASTELLANI U., FUSIELLO A.: The bag of words approach for retrieval and categorization of 3d objects. *The Visual Computer 26*, 10 (2010), 1257–1268. 2

[WAS10] WUHRER S., AZOUZ Z. B., SHU C.: Semi-automatic prediction of landmarks on human models in varying poses. In *Computer and Robot Vision (CRV), 2010 Canadian Conference on* (2010), IEEE, pp. 136–142. 2, 5

[WSX11] WUHRER S., SHU C., XI P.: Landmark-free posture invariant human shape correspondence. *The Visual Computer 27*, 9 (2011), 843–852. 2, 5, 8, 10

[ZZZZ06] ZHENG W., ZHOU X., ZOU C., ZHAO L.: Facial expression recognition using kernel canonical correlation analysis (kcca). *IEEE Transactions on Neural Networks 17*, 1 (2006), 233–238. 2