BodyPrint: Pose Invariant 3D Shape Matching of Human Bodies

Jiangping Wang[1]*, Kai Ma[2], Vivek Kumar Singh[2], Thomas Huang[1], Terrence Chen[2]
[1]Beckman Institute, University of Illinois at Urbana-Champaign, USA
[2]Medical Imaging Technologies, Siemens Healthcare, Princeton, NJ, USA
[1]{jwang63,huang}@ifp.uiuc.edu, [2]{kai.ma,vivek-singh,terrence.chen}@siemens.com

Abstract

3D human body shape matching has substantial potential in many real world applications, especially with recent advances in 3D range sensing technology. We address this problem by proposing a novel holistic human body shape descriptor called BodyPrint. To compute the bodyprint for a given body scan, we fit a deformable human body mesh, and project the mesh parameters to a low-dimensional subspace which improves discriminability across different persons. Experiments are carried out on three real-world human body datasets to demonstrate that BodyPrint is robust to pose variation as well as missing information and sensor noise. It improves the matching accuracy significantly compared to conventional 3D shape matching techniques using local features. To facilitate practical applications where the shape database may grow over time, we also extend our learning framework to handle online updates.

1. Introduction

Non-rigid 3D shape matching is a fundamental problem that has been widely studied in computer vision and graphics. Given a database of 3D shapes (often represented as meshes), shape matching algorithms return shapes that are similar to the query object. This research area has become particularly active over the last decade due to the availability of low cost 3D acquisition device, such as Microsoft Kinect [2]. The pose-invariant matching of human body shapes is of particular interest, as it finds applications in people re-identification in surveillance [6, 23], biometric authentication and information retrieval in medical imaging [32, 33]. However, human body shape matching is quite challenging as it requires invariance to human body articulations while simultaneously capturing subtle variations of the body shape across different individuals. Furthermore, to enable the aforementioned applications, the matching should be performed rapidly over a large database. Although there have been recent advances[25], this problem is still widely unresolved to the best of our knowledge.

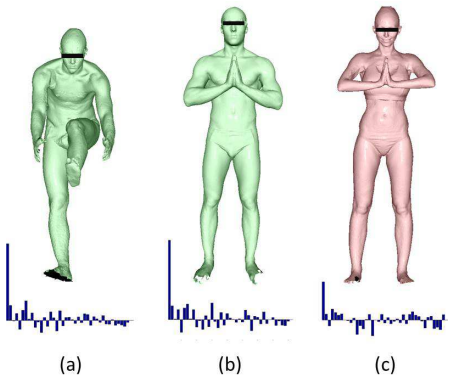

Figure 1. Human body meshes and corresponding *BodyPrint* descriptors (shown as bar chart). (a) and (b) are meshes from the same person in a different pose. (b) and (c) are meshes from different persons in a similar pose. Notice that the bodyprint for (a) is more similar to (b) in comparison to (c).

In this paper we introduce a novel framework for time efficient matching of 3D human body shapes which is invariant to body pose articulations. We also perform extensive experiments to demonstrate that our approach achieves state-of-the-art performance over different datasets. In general, the proposed framework uses a parameterized deformable mesh (PDM) template that can be deformed to fit any human body scans. Based on the mesh parameters (vertices and edges), we derive a compact body shape descriptor that captures the holistic shape information of the human body; we refer to this as *BodyPrint*. The derivation involves a projection of the mesh parameters to a low-dimensional manifold such that the distance between meshes from different persons/categories is maximized, while minimizing the distance between meshes from same person/category. This projection matrix is obtained using a novel ranking based distance metric learning algorithm. Figure 1 shows an example of body scans and their corresponding *BodyPrint* signatures.

*This work was carried out during the author's internship at Medical Imaging Technologies, Siemens Healthcare.

The proposed *BodyPrint* descriptor embodies several key ideas which provide significant benefits over other existing body shape matching approaches. Initially, PDM (which is built on SCAPE [3]) models the pose and body shape parameters independently. Since the bodyprint is derived only from the shape parameters, it is robust to pose perturbations. Following this, *BodyPrint* represents the holistic shape information based on PDM fitting which is robust to missing mesh data as well as local noise, as opposed to local descriptor methods [7, 10, 18, 24, 37]. This is practically relevant since obtaining a complete 360 degree view body scan of individuals is difficult both in surveillance, as well as medical domain (for patients with severe injuries). Last but not least, the projection matrix is obtained using ranking based metric learning which uses similarity constraints over triplets and hence is better suited for matching. We also extend the framework to online learning since in many practical applications, the size of the database may increase from time to time and re-training on the entire database would be time consuming.

2. Related Work

The area of matching 3D shapes has been extensively researched both in computer graphics as well as the vision community. In the case of shape-DNA [26], the intrinsic geometry property of 3D object is captured by using the Laplace-Beltrami spectrum. The compact representation of a sequence of eigenvalues is proven to be an isometry invariant that allows robustness to intra-class variation. Recent research [34] builds on the shape-DNA work and reports impressive performance for brain re-identification over several patient scans. Another similar work [20] aggregates Scale-Invariant Heat Kernel Signature (SI-HKS) [11] as the local spectral feature in a bag-of-feature paradigm. Together with supervised dictionary learning and sparse coding, their method achieves state-of-the-art performance on the latest shape retrieval challenge (SHREC'14 [25]). We adopt a similar strategy that uses a compact yet powerful descriptor to represent the holistic shape information, however, our approach differs in the way the shape descriptor is computed.

The parametrized deformable mesh used in our work is built on the SCAPE [3] model, which has been widely applied to accurately estimate human body shape and pose under different scenarios [4, 31, 32, 33, 36]. Given a 3D scan of a person, the aforementioned approaches pursue a aligned body mesh that fits closely to the ground truth. Different from those methods, our main focus in this paper is to solve the shape matching problem. To the best of our knowledge, ours is the first method that efficiently proceeds human body shape matching based on the 3D reconstruction of human body. Tsoli *et al*. [33] use the PCA coefficients with linear regressions to predict anthropometric measurements with sufficiently high fidelity, which only further supports our proposal to extend the usage of the PCA coefficients to the human body shape matching.

To enrich the shape representation, we employ metric learning to learn a projection of the mesh parameters to a low-dimensional but more discriminative manifold. In 3D human pose estimation [16], metric learning for achieving robust yet discriminative 3D descriptors has been proven quite effective. Many popular metric learning algorithms adopt pairwise similarity constraints, such as Relevant Component Analysis (RCA) [5], Information Theoretic Metric Learning (ITML) [14], Logistic Discriminant Metric Learning (LDML) [15] and Pairwise Constrained Component Analysis (PCCA) [22]. For supervised metric learning for matching shapes, incorporating pairwise constraints implicitly require a threshold (to determine whether a pair is a match or not); this may introduce ambiguity since different parts of training dataset may be annotated by different individuals. On the contrary, the triplet constraints model the relative information about which pair is closer and hence is better suited for learning the true shape manifold. This is especially helpful when handling the available 3D human dataset which usually does not include many subjects with exactly the same body shape. Online Algorithm for Scalable Image Similarity (OASIS) [12] and Probabilistic Relative Distance Comparison (PRDC) [38] learn similarity using constraints defined over triplets.

3. Parametrized Deformable Mesh for Body Shape Estimation

Given a 3D scan of an object, modern shape matching algorithms find a unique shape characterization of the object, which is usually referred to as a shape descriptor. These descriptors often serve as the key for matching. The ideal shape descriptor should be compact (for fast search) and exhibit invariance to all other deformations beyond shape. For the human body shape, it is particularly important to deal with the variations due to pose changes.

We employ a Parametrized Deformable Mesh (PDM) to model the human body. Our model, inspired by the SCAPE model[3], decouples the human pose and shape perturbations and models them separately. Therefore, the shape model factorizes the deformations caused by changes in intrinsic shape (height, size, belly thickness, etc.) from deformations caused by changes in pose (rigid transformations of body parts). We model the shape deformations using PCA over a large dataset with shape variations, using a commercial software called Poser^{TM} [1]. Besides being able to efficiently generate a large training dataset with accurate point wise correspondences, one additional benefit of using Poser is the reduction in training complexity since Poser allows shape perturbations without changing the pose. The complete training dataset consists of $1,000$ poses and 600 shapes. Figure 2 shows some sample data.

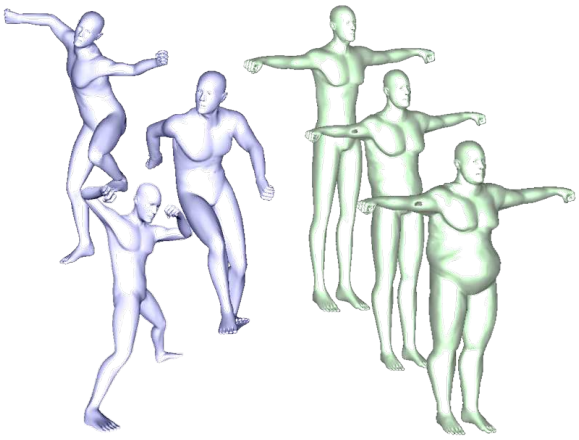

Figure 2. Examples of the synthetic training data of PDM generated by Poser. The pose training data is displayed on the left in blue, and the shape training data is on the right in green.

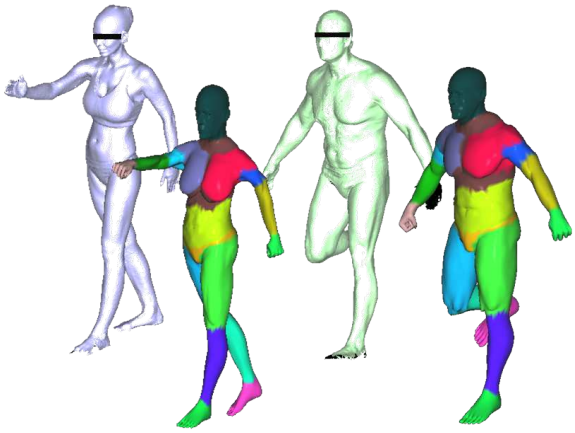

Figure 3. PDM fitting results. The top row shows the input scans from FAUST dataset [9] and the bottom row shows the fitted PDM template mesh. Different body parts of the template mesh are rendered with different colors.

Given the shape training dataset, The shape affine matrix S^i for each tri-angular mesh data i can be obtained by solving the quadratic function:

$$\underset{S^i}{\operatorname{argmin}} \sum_k \sum_{j=2,3} ||S^i_k \hat{v}_{k,j} - v^i_{k,j}||^2 + w \sum_{k_1,k_2 adj} ||S^i_{k_1} - S^i_{k_2}||^2 \tag{1}$$

where k represents the triangle index and $v_{k,j}$ is the j^{th} edges in k^{th} triangle. The affine matrices can be further decomposed into a linear combination of Eigen-vectors U and mean-vector μ by using PCA:

$$S^i = F_{U,\mu}(\beta^i) = U\beta^i + \mu \tag{2}$$

By changing the values of PCA coefficient vector β, we can recover any body shape in the learned manifold. Although there is no explicit interpretation of each dimension to a semantic definition, the first few dimensions of β correspond to the global shape perturbations in the shape training set (gender, height, body size and etc.). The following dimensions of β capture more and more subtle perturbations. The fitting accuracy of PDM depends on the number of PCA coefficients that are used to model the shape parameters. While more dimensions can model body deformations in greater detail (which may be useful for shape matching), it also increases time complexity and the possibility to fit small, noisy perturbations in the data. Hence, the choice of the number of PCA coefficients is important. For our experiments, we use 60 coefficients that retain 98% of the energy; this helps suppress noise without losing most of shape deformation information. (A more detailed experiment with regard to the PCA coefficients is shown in the supplement material.)

For inference (i.e. to deform the template mesh to an input 3D scan), we develop upon the iterative optimization technique presented in [3]. Such techniques have already demonstrated to fit 3D data quite well [4, 36]. For faster inference, the optimization is done in a coarse-to-fine manner. First the pose and shape of the template model is initialized by several pre-determined landmarks. A robust non-rigid ICP algorithm [28] is then applied to identify a set of point-to-point correspondences between the input and the mesh template. Given the correspondences, the template mesh is properly deformed to minimize the ℓ_2 norm. The deformed template mesh is then used to determine new correspondences and this process of registration and optimization is repeated until convergence (i.e. average distance is below a certain score or maximum number of iterations are reached). Figure 3 shows some PDM fitting results of different person.

4. Learning BodyPrint Descriptor for Fast Matching

Given a human body scan, the PDM module decouples the pose and shape and projects the holistic body shape information onto a low dimension PCA subspace, yielding a set of coefficients β in Eq. (2). These coefficients can themselves be used as a shape descriptor for matching body shape with Euclidean metric, but its discriminability may not be sufficient to enable practical applications. To this end, we project shape coefficients to another manifold such that the distance between meshes from different persons is maximized, while the distance between meshes from same person is minimized. *BodyPrint* is the descriptor obtained from projecting the PCA coefficient vector to a more discriminative manifold. We compute this projection matrix using a novel ranking based metric learning framework.

4.1. Ranking Based Metric Learning

For a pair of descriptors (x_j, x_k), the Mahalanobis distance is measured by

$$D_M^2(x_j, x_k) = (x_j - x_k)^T M (x_j - x_k) \tag{3}$$

where M is a symmetric positive semi-definite (PSD) matrix. Since the matrix M can be decomposed as $M = L^T L$, the Mahalanobis distance can be interpreted as the Euclidean distance in a linearly transformed feature space, i.e., $D_M^2(x_j, x_k) = \|L(x_j - x_k)\|_2^2$.

In order to learn a metric[1] for 3D shape matching, we consider using triplet constraints for supervision. Triplet constraint carries pairwise similarities between three data items of a set. In a triplet (x_i, x_i^+, x_i^-) annotation, x_i is considered more similar to x_i^+ than x_i^-. One reason to use triplets is that in most cases we only have very few (if not one) 3D shape data for the same subject and thus lack pairwise constraints; the other reason is triplet constraints contain more similarity side information than pairwise constraints on the same dataset. In the experiment section, we show how we derive triplet constraints for CAESAR data [27] using anthropomorphic measurements.

Given the triplet constraints, we propose a batch version and an online version of metric learning. Our formulation is based on the maximum margin criterion; that is, the distance between more similar pairs (x_i, x_i^+) is less than that between (x_i, x_i^-) by a large margin. This idea is similar to Large Margin Nearest Neighbor (LMNN) [35] and the approach in [19]; however, in our problem, we do not require class labels of the data and aim to learn a metric purely based on triplet constraints, while LMNN optimizes kNN for classification. As opposed to [19] that simplifies the problem by assuming distance metric to be a diagonal matrix, we directly solve for a PSD matrix. As we will show later, our algorithms are effective yet efficient, and have theoretical justification.

In an ideal setting, there might exist a matrix M, such that for any triplet (x_i, x_i^+, x_i^-),

$$D_M^2(x_i, x_i^-) > D_M^2(x_i, x_i^+) + 1 \tag{4}$$

Similar to Support Vector Machine (SVM), we consider soft margin for inseparable case which amounts to minimizing hinge loss

$$\ell_i(M) := [1 + D_M^2(x_i, x_i^+) - D_M^2(x_i, x_i^-)]_+ \tag{5}$$

where $[z]_+ = max(0, z)$. On the other hand, we employ low-dimensional PDM shape descriptors after PCA, so the learned metric should not be distorted from the identity matrix too much. Taking into account the above constraints, our batch algorithm is given by

$$\min_{M \succeq 0} \frac{\lambda}{2}\|M - I\|_F^2 + \frac{1}{N}\sum_{\mathcal{S}} \ell_i(M) =: F(M) \tag{6}$$

where $\mathcal{S} = \{(x_i, x_i^+, x_i^-)\}$ is the set of triplets, $|\mathcal{S}| = N$.

Generic solver for semi-definite programming employs an interior point and does not scale well with a large number of constraints, as is the case in Eq. (6). We develop an efficient stochastic subgradient descent algorithm to solve the optimization, as shown in Algorithm 1, where η_t is the learning rate and output M is the minimizer of $F(M)$.[2]

Algorithm 1 Mini-batch stochastic subgradient descent algorithm for solving optimization in Eq. (6)

Input: $\mathcal{S}$, λ and T.
Initialization: $M_1 = I$
for $t = 1, 2, \ldots, T$ **do**
  Randomly choose $\mathcal{S}_t \subseteq \mathcal{S}, |\mathcal{S}_t| = K$
  Set $\mathcal{S}_t^+ = \{(x_i, x_i^+, x_i^-) \in \mathcal{S}_t : 1 + D_{M_t}^2(x_i, x_i^+) - D_{M_t}^2(x_i, x_i^-) > 0\}$
  $\bigtriangledown_t = \lambda(M_t - I) + \frac{1}{K}\sum_{\mathcal{S}_t^+}(C_t(x_i, x_i^+) - C_t(x_i, x_i^-))$, where $C_t(x_i, x_i^+) := (x_i - x_i^+)(x_i - x_i^+)^T$
  $M_{t+1} = M_t - \eta_t \bigtriangledown_t$
  Decompose $M_{t+1} = U\Lambda U^T$
  Project M_{t+1} onto PSD cone, $M_{t+1} \leftarrow U\Lambda^+ U^T$, where $\Lambda^+ = max(0, \Lambda)$.
end for
Output: $\text{argmin}_{M \in \{M_1, M_2, \ldots, M_{T+1}\}} F(M)$

4.2. Extension to Online Learning

We propose an online algorithm based on the Passive-Aggressive (PA) family of learning algorithms introduced by Crammer *et al.* [13]. In the online setting, we assume a triplet (x_i, x_i^+, x_i^-) is observed at each time step i, which suffers a loss defined in Eq. (5). If $\ell_i(M) = 0$, we suffer no loss; otherwise the metric should be updated. Denote by M_i the matrix used for prediction at time step i.

4.2.1 Separable Case

We first consider the separable case, which assumes that there exists a matrix M^* such that $\ell_i(M^*) = 0$ for all i. Following the method in Pseudo-metric Online Learning Algorithm (POLA) [29], we derive our algorithm based on the orthogonal projection operation. Given $\forall W \in \mathbb{R}^{d \times d}$ and a closed convex set $\mathcal{C} \subset \mathbb{R}^{d \times d}$, the orthogonal projection of W onto $\mathcal{C}$ is defined by

$$\mathcal{P}_{\mathcal{C}}(W) = \underset{W' \in \mathcal{C}}{\text{argmin}} \|W - W'\|_F^2 \tag{7}$$

[1] Strictly speaking, we learn pseudo-metric. The term "metric" is used in the paper for simplicity.

[2] In practice, $F(M)$ is computed every j epochs in the late stage of learning, where j ranges from dozens to hundreds, depending on the total size of mini-batches.

For each time step i, the set $\mathcal{C}_i \subset \mathbb{R}^{d\times d}$ is defined as

$$\mathcal{C}_i = \{M \in \mathbb{R}^{d\times d} : \ell_i(M) = 0\} \tag{8}$$

where $\ell_i(M)$ is defined in Eq. (5). Another constraints on M is that $M \succeq 0$. Denote by $\mathcal{C}_a$ the set of PSD matrices,

$$\mathcal{C}_a = \{M \in \mathbb{R}^{d\times d} : M \succeq 0\} \tag{9}$$

With the above definitions, our online algorithm is comprised of two successive projections as below

$$M_{\tilde{i}} = \mathcal{P}_{\mathcal{C}_i}(M_i), \tag{10}$$
$$M_{i+1} = \mathcal{P}_{\mathcal{C}_a}(M_{\tilde{i}}). \tag{11}$$

First, we project the current matrix M_i onto $\mathcal{C}_i$ so the resulting $M_{\tilde{i}}$ will be the closest one to M_i while achieving a zero loss on the received triplet at time step i. Second, we project $\mathcal{C}_i$ onto $\mathcal{C}_a$ to ensure it is a metric. Now we show how the projections can be performed analytically. The first projection is equivalent to solving the following constrained optimization problem,

$$M_{\tilde{i}} = \underset{M}{\operatorname{argmin}} \frac{1}{2}\|M - M_i\|_F^2, \text{ s.t. } \ell_i(M) = 0 \tag{12}$$

which has a simple closed-form solution by using KKT condition

$$M_{\tilde{i}} = M_i + \alpha_i V_i \tag{13}$$

where

$$V_i = (x - x^-)(x - x^-)^T - (x - x^+)(x - x^+)^T \tag{14}$$

$$\alpha_i = \frac{\ell_i(M_i)}{\|V_i\|_F^2} \tag{15}$$

Since we initialize M_1 to be identity matrix I, $M_{\tilde{i}}$ is always symmetric and thus can be decomposed as $M_{\tilde{i}} = U_i\Lambda_i U_i^T$ By projecting $M_{\tilde{i}}$ onto PSD cone, M_{t+1} can be derived as $M_{t+1} = U_i\Lambda_i^+ U_i^T$, where $\Lambda_i^+ = max(0, \Lambda_i)$.

Theorem 1. *Let $(x_i, x_i^+, x_i^-)_{i=1}^T$ be a sequence of triplet instances. Assume that there exists $M^* \succeq 0$ such that $\forall i \geq 1, \ell_i(M^*) = 0$. Let R be an upper bound that satisfies $\forall i : R \geq \|V_i\|_F^2$. Then the following bound holds for any $T \geq 1$*

$$\sum_{i=1}^{T} \ell_i^2(M_i) \leq R\|M^* - I\|_F^2 \tag{16}$$

See proof in the supplemental material.

4.2.2 Inseparable Case

For the inseparable case, there is no metric that separates the triplet instances by a large margin perfectly. We relax the assumption and modify Eq. (10) by solving the following optimization problem

$$M_{\tilde{i}} = \underset{M}{\operatorname{argmin}} \frac{1}{2}\|M - M_i\|_F^2 + C\ell_i^2(M) \tag{17}$$

where C is aggressiveness parameter that controls the trade-off between the loss on the triplet and the regularization. The above learning problem can be regarded as a matrix version of PA-II algorithm in [13], and has closed-form solution as

$$M_{\tilde{i}} = M_i + \alpha_i V_i \tag{18}$$

where

$$\alpha_i = \frac{\ell_i(M_i)}{\|V_i\|_F^2 + \frac{1}{2C}} \tag{19}$$

and V_i is defined in Eq. (14).

Essentially the solution has the same form as in Eq. (13) for separable case, with some modification in α_i. M_{t+1} is obtained by projecting $M_{\tilde{i}}$ onto PSD cone, following the same procedure in Eq. (11).

Theorem 2. *Let $(x_i, x_i^+, x_i^-)_{i=1}^T$ be a sequence of triplets, and let R be an upper bound such that $\forall i : R \geq \|V_i\|_F^2$. Then, for any matrix $Q \succeq 0$, the following bound holds for any $T \geq 1$*

$$\sum_{i=1}^{T} \ell_i^2(M_i) \leq (R + \frac{1}{2C})(\|Q - I\|_F^2 + 2C\sum_{i=1}^{T} \ell_i^2(Q)) \tag{20}$$

See proof in the supplemental material.

4.3. Remark on Metric Learning

Our batch and online algorithms can be applied either independently or in conjunction with each other. In a practical setting where the human subjects in the dataset may change or grow over time, an initial metric can be obtained using the batch algorithm and then subsequently adapted using the online algorithm as new data becomes available.

Our online algorithm advances POLA [29] in several aspects. To start with, our algorithm handles triplet annotations and learns the metric based on relative similarity constraints, as opposed to POLA that works with pairwise constraints. Furthermore, we derive algorithms and loss bounds for both separable and inseparable cases, while POLA mainly focus on the analysis of separable case. On the other hand, same as OASIS [12], our online algorithm is also based on PA family of algorithms [13]; however, we aim to obtain a metric in the form of a symmetric and PSD matrix, while OASIS learns a bilinear similarity matrix.

5. Experiments

To validate the proposed *BodyPrint* descriptor, we conduct extensive experiments on three real-world datasets including two public 3D human body scan datasets, CAESAR [27] and MPI-FAUST [9], as well as a new Kinect

body scan dataset. Our baseline is a local feature based 3D shape matching workflow that includes local descriptor extraction (existing 3D Descriptors in PCL [28] such as SHOT, 3DSC and PFH), 3D shape representation and metric learning using pairwise constraints (e.g., KISSME [17]). We also compare the performance of our metric learning algorithm with state-of-the-art methods - ITML [14], LMNN [35], KISSME [17] and LDMLT [21] which are generic as opposed to algorithms that are designed to target a specific use case such as face verification. To further justify the performance of the proposed method, we also run evaluations with the current state-of-the-art spectral method [20] and a statistically adapted method [8] on FAUST and Kinect dataset. For all experiments, we set $\lambda = 0.01$, $\eta_t = 10^{-3}/\sqrt{t}$ in Algorithm 1 and $C = 0.01$ in Eq. (17). For other metric learning algorithms as well as the spectral method, we use default parameters provided by the authors.

5.1. Dataset

The CAESAR dataset includes scans of $2,400$ human subjects in 3 different poses. It also comes with 40 precise anthropometric attributes that were directly measured on human subjects, which are used here to rank the similarities among the scan data. Among all the 40 attributes, we carefully select 22 of them that represent uncorrelated measurements (8 of these attributes are shown in Table 1).

The FAUST dataset, although designed as a benchmark for human body model fitting, also serves as a good benchmark for pose-invariant human re-identification. The full dataset includes 300 scans acquired from 10 subjects in 30 different poses per subject. Although the FAUST dataset has significant pose variations, the data only has label (person identity) information, unlike CAESAR which also includes additional geometry based attributes. Hence, we design the experiments for the use case of person re-identification and study the robustness of various approaches to body pose perturbations. In our experiments, we train on 10 poses per subject and test on the rest to evaluate pose invariance.

We also acquired a new Kinect dataset that includes $1,200$ depth images collected from Microsoft Kinect 1.0 to evaluate the robustness of our approach regarding to the effects of sensor noise, clothing and partial occlusion. Unlike some of the existing Kinect dataset that are well suited for surveillance applications[6, 23], we geared towards biometric authentication of a cooperative user in an office-like environment and designed our experiment accordingly. The dataset contains snapshots (single depth image) of 20 human subjects in 30 different poses (frontal, upright pose with casual limb positions) acquired at the distance of 1.5 to 3 meters from the sensor. Each snapshot covers head-to-toe information of the subjects and segments the subjects from background. Since we are interested in shape based matching, color information is not used in this experiment.

Acromial Height	Chest Circumference
Buttock-Knee Length	Crotch Height
Hip Circumference	Shoulder Breadth
Sitting Height	Waist Circumference

Table 1. Examples of biometric attributes for CAESAR dataset.

5.2. Evaluation Metric

We follow the same evaluation protocol in [25]. During the query step, each individual input is queried within the full dataset to get a list of all other shapes ranked in descending order according to the shape similarities. We evaluate the results using various statistical measurements: nearest neighbor (rank-1), e-measure (E-M), discounted cumulative gain (DCG), and precision/recall curves. Definitions of these evaluation metrics are listed in [30].

5.3. Results on CAESAR Dataset

Triplet Annotation. As mentioned in 4.1, our metric learning benefits from the triplet annotation. We build triplets based on the similarity ranking in the biometric attribute space. For each scan in the training set, we organize the rest of the data in descending order based on the accumulated errors of all 22 given attributes. Then for a given triplet, the positive x_i^+ and negative x_i^- labels can be efficiently determined by the data indices in the queue of x_i. This "soft" annotation works well since the 22 attributes were precisely measured by the data provider and were sufficient to reflect the actual body shape. Note that for the CAESAR data, the pairwise similarity constraints based metric learning algorithms such as KISSME [17] cannot directly use similarity ranks for training, and we obtain the pairwise labels by thresholding the distance among the data in biometric attribute space. It is hard and ambiguous to select optimal fixed thresholds to generate pairwise similarity constraints for different data or setups, which is another reason we utilize ranking based metric learning for our *BodyPrint*.

Shape Matching. We randomly select 100 body scans as the training data and another 200 for testing (both are gender balanced, 50% male or female). For test data, the ground truth similarity ranking is also built on biometric attributes, similar to triplet annotation. Each shape matching method will generate the similarity ranking for every scan in the test dataset, as described in the subsection 5.2. In evaluation protocol on CAESAR data for each body scan query, the top 20 scans in its similarity rank are considered correct matches, since matching only the most similar shape is

Method	rank-1	E-M	DCG	τ	ρ
PDM+Euclidean	0.705	0.438	0.742	0.567	0.743
PDM+ℓ_1	0.620	0.412	0.718	0.498	0.672
PDM+Mahalanobis	0.240	0.211	0.535	0.190	0.273
PDM+KISSME	0.765	0.536	0.812	0.701	0.863
PDM+LDMLT	0.730	0.570	0.823	0.732	0.889
BodyPrint	**0.820**	**0.574**	**0.843**	**0.741**	**0.892**

Table 2. Performance comparison of shape matching on the CAESAR dataset, with 100 scans for training and 200 for testing.

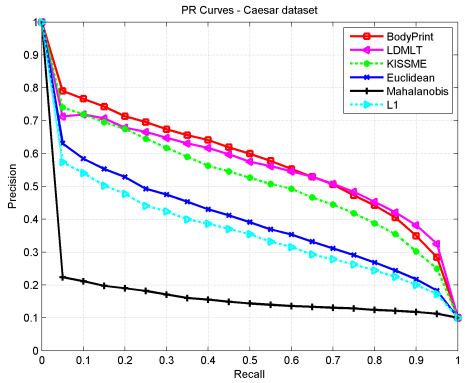

Figure 4. Precision/recall curves of shape matching on CAESAR.

	Training accuracy(%)	Testing accuracy(%)
Gender	100	97.5
Short/Tall	96	94
Light-weight/Heavy	94	92

Table 3. Semantic classification on CAESAR, using obtained *BodyPrint* as a feature.

too restrictive for evaluation purposes.[3] Thus the evaluation metrics introduced in the subsection 5.2 can be computed accordingly. We also calculate the Kendall's coefficient (τ) and Spearman's coefficient (ρ) to measure the rank correlation with the ground truth. Both τ and ρ take values between $[-1, 1]$, where $-1/1$ indicates the ground truth rank and predicted rank are completely reverse/same. Table 2 shows the performance of *BodyPrint* shape matching on CAESAR compared with several baselines with the aforementioned evaluation metrics. Figure 4 displays the precision-recall curves. Overall, our PDM with metric learning framework works very well in the task of shape retrieval with CAESAR data. Using raw PDM coefficients with plain Euclidean distance, about 70% rank-1 accuracy can be achieved. Our (batch) *BodyPrint* algorithm performs the best among all the tested methods.

Semantic Classification. As mentioned in Section 3, there is no explicit interpretation of each dimension in the PDM coefficients to a semantic definition. It is interesting to investigate if general semantic body shape information such as male/female, short/tall can be captured by *BodyPrint*. To this end, we conduct experiments for semantic attribute classification using *BodyPrint* with a linear SVM. We experiment with three semantic attributes: **gender**, **height** and **weight**. Table 3 summarizes the classification results. The high classification accuracy confirms that *BodyPrint* indeed carries shape as well as semantic information.

[3]For top 1 matching accuracy, our method still performs best; it is more meaningful to measure top 20 matching accuracy here because soft annotation is used.

Method	rank-1	E-M	DCG
SHOT[28]+ITML	0.450	0.303	0.647
SHOT[28]+KISSME	0.550	0.318	0.692
3DSC[28]+ITML	0.680	0.329	0.735
3DSC[28]+KISSME	0.645	0.355	0.730
PFH[28]+ITML	0.710	0.372	0.772
PFH[28]+KISSME	0.715	0.394	0.763
Litman et al.[20]	0.875	0.423	0.834
PDM+Euclidean	0.767	0.382	0.739
PDM+Blanz et al.[8]	0.550	0.331	0.650
PDM+KISSME	0.900	**0.488**	0.823
PDM+LMNN	0.900	0.444	0.841
PDM+ITML	0.875	0.452	0.831
BodyPrint(batch)	**0.933**	0.442	**0.881**

Table 4. Performance comparison on FAUST, with 10 persons and 10 poses each for training.

5.4. Results on FAUST Dataset

Experimental Setting. For the FAUST dataset, both triplet and pair constraints can be easily generated from the subject identities. We vary the number of person and poses in the training set to test the robustness and generalization of different approaches. We experiment with both batch and online *BodyPrint* under different setups. For each person, we use 10 poses for training, and compare our PDM framework with the baselines. Within PDM framework, we also compare our (batch) *BodyPrint* with PDM based shape signatures using other metric learning algorithms. To demonstrate the performance of our online learning method, we compare with the LDMLT [21] algorithm. Here, we reduce the number of person in training set from 10 to 6 to evaluate the generalization ability of the algorithms.

The implementations of our baseline methods, the local feature based 3D shape descriptors such as SHOT, 3DSC and Point Feature Histograms (PFH), are provided by the widely diffused PCL [28] library. The default parameters are applied when available. The radius of surface normal calculation and local descriptor are set to $20mm$ and $50mm$ individually for all baseline. Once the descriptors get extracted, we concatenate all the feature vectors and apply PCA to reduce the dimensionality to 60. Two metric learning frameworks, ITML and KISSME, are applied to baseline descriptors to measure the performance.

Analysis. From Table 4, we can see that PDM with plain Euclidean metric obtains quite competitive results, outperforming the baseline methods. This indicates the superior ability of PDM in capturing essential shape information of body scan. All the implemented metric learning algorithms

Method	rank-1	E-M	DCG
PDM+Euclidean	0.763	0.375	0.730
PDM+LDMLT	0.881	0.467	0.818
BodyPrint(online)	**0.919**	**0.472**	**0.826**

Table 5. Performance comparison on FAUST, using 6 persons each with 10 poses for training and total 160 body scans for testing.

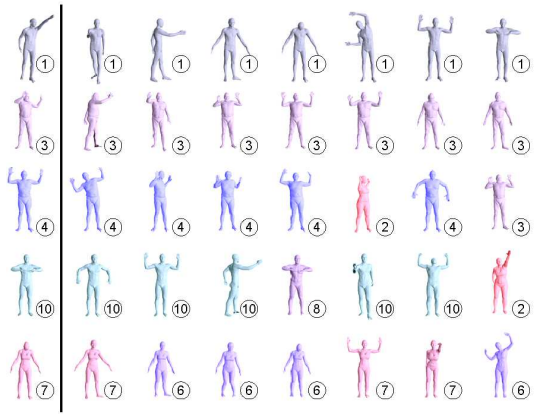

Figure 5. Re-identification result on FAUST dataset. The input meshes are shown on the left and the retrieved nearest neighbors are shown on the right in a descending order of similarity.

can boost the performance of PDM. Overall, *BodyPrint* and PDM+KISSME perform best: *BodyPrint* achieves the highest accuracy on rank-1 metric, and the latter attains best performance on the other two metrics. From Table 5, we can see that our online algorithm outperforms LDMLT on all the evaluation metrics. Figure 5 provides a qualitative impression of the matching performance. Column 1 shows the body scan used to query the database and the rest of the columns show the top 7 results sorted from left to right, based on the matching score. As shown in Figure 5, our *BodyPrint* demonstrates the property of pose invariance.

5.5. Results on Kinect Dataset

In the previous experiments, we have shown that our method achieves state-of-the-art performance on dense, 360-degree scans of human subjects. Kinect data, on the other hand, is noisy and has significantly inferior resolution as well as has only partial body surface information. This makes the dataset particularly challenging for the people re-identification problem.

Experimental Setting. The set up is similar to the FAUST experiment setup. For each of the 20 subjects, we randomly select 20 scans for training and another 40 for testing. As in FAUST, we use the person identity information to generate triplets for metric learning (and pairwise constraints for the baseline methods). To increase efficiency, we only use randomly sampled subsets of the triplets and employ our online *BodyPrint* metric learning algorithm.

Analysis. From Table 6, we can see that the proposed *BodyPrint* method achieves high re-identification rank-1 accuracy (about 90%). This suggests that our approach is able to deal with noisy, partial depth information for person re-identification tasks. We also observe that PDM based shape matching methods (including *BodyPrint*) notably outperform traditional local feature based methods as well as the spectrum method [20]. This may be due to the fact that the detailed surface information may not be easily distinguishable due to noisy depth data at 2-3 meters distance, while the holistic body shape information is likely to be more stable and hence better suited for re-identification. Notice also that among the PDM based methods, *BodyPrint* (Online) improves on the results of the rest.

Method	rank-1	E-M	DCG
PFH[28]+ITML	0.543	0.244	0.645
PFH[28]+KISSME	0.603	0.292	0.687
Litman et al.[20]	0.609	0.137	0.549
PDM+Euclidean	0.741	0.225	0.638
PDM+Blanz et al.[8]	0.743	0.241	0.653
PDM+KISSME	0.858	**0.519**	0.837
PDM+LMNN	0.884	0.502	0.837
PDM+ITML	0.809	0.491	0.819
PDM+LDMLT	0.861	0.414	0.784
BodyPrint(Online)	**0.891**	0.516	**0.843**

Table 6. Performance comparison on Kinect dataset, with 20 persons and a randomly selected 20 scans each for training.

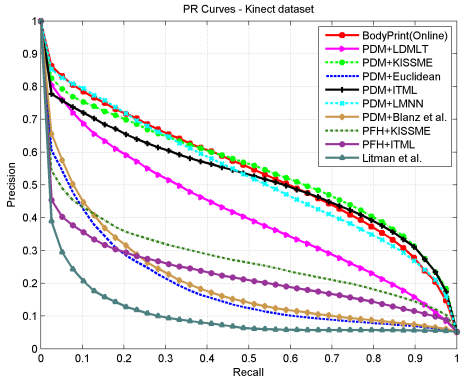

Figure 6. Precision/recall curves of re-identification on Kinect dataset.

6. Conclusion

We presented a novel shape descriptor for 3D human body shape matching. The proposed descriptor is demonstrated to be robust to pose variations, as well as sensor noise and missing data. It improves the matching accuracy significantly over existing 3D shape matching algorithms. The low dimensionality of *BodyPrint* also allows fast search, which is important to practical applications. We also extend our learning framework to allow online updates and hence the *BodyPrint* computation can be adapted to specific scenarios for better performance.

References

[1] http://poser.smithmicro.com/poser10-poserpro2014. 2

[2] https://www.microsoft.com/en-us/kinectforwindows. 1

[3] D. Anguelov, P. Srinivasan, D. Koller, S. Thrun, J. Rodgers, and J. Davis. SCAPE: shape completion and animation of people. *ACM Trans. Graph*, 2005. 2, 3

[4] A. Balan and M. J. Black. The naked truth: Estimating body shape under clothing,. In *ECCV*, 2008. 2, 3

[5] A. Bar-Hillel, T. Hertz, N. Shental, and D. Weinshall. Learning distance functions using equivalence relations. In *ICML*, 2003. 2

[6] B. I. Barbosa, M. Cristani, A. Del Bue, L. Bazzani, and V. Murino. Re-identification with rgbd sensors. In *1st Intl Workshop on Re-identification*, October 2012. 1, 6

[7] S. Bauer, J. Wasza, S. Haase, N. Marosi, and J. Hornegger. Multi-modal surface registration for markerless initial patient setup in radiation therapy using microsoft's kinect sensor. In *ICCV Workshops*, 2011. 2

[8] V. Blanz, S. Romdhani, and T. Vetter. Face identification across different poses and illuminations with a 3d morphable model. In *AFGR*, 2002. 6, 7, 8

[9] F. Bogo, J. Romero, M. Loper, and M. J. Black. FAUST: Dataset and evaluation for 3D mesh registration. In *CVPR*, 2014. 3, 5

[10] A. M. Bronstein, M. M. Bronstein, L. J. Guibas, and M. Ovsjanikov. Shape google: Geometric words and expressions for invariant shape retrieval. *ACM Transactions on Graphics (TOG)*, 30(1):1, 2011. 2

[11] M. Bronstein and I. Kokkinos. Scale-invariant heat kernel signatures for non-rigid shape recognition. In *CVPR*, 2010. 2

[12] G. Chechik, V. Sharma, U. Shalit, and S. Bengio. Large scale online learning of image similarity through ranking. *The Journal of Machine Learning Research*, 2010. 2, 5

[13] K. Crammer, O. Dekel, J. Keshet, S. Shalev-Shwartz, and Y. Singer. Online passive-aggressive algorithms. *The Journal of Machine Learning Research*, 2006. 4, 5

[14] J. V. Davis, B. Kulis, P. Jain, S. Sra, and I. S. Dhillon. Information-theoretic metric learning. In *ICML*, 2007. 2, 6

[15] M. Guillaumin, J. Verbeek, and C. Schmid. Is that you? metric learning approaches for face identification. In *ICCV*, 2009. 2

[16] A. Kanaujia, C. Sminchisescu, and D. Metaxas. Semi-supervised hierarchical models for 3d human pose reconstruction. In *CVPR*, 2007. 2

[17] M. Koestinger, M. Hirzer, P. Wohlhart, P. M. Roth, and H. Bischof. Large scale metric learning from equivalence constraints. In *CVPR*, 2012. 6

[18] I. Kokkinos, M. M. Bronstein, R. Litman, and A. M. Bronstein. Intrinsic shape context descriptors for deformable shapes. In *CVPR*, 2012. 2

[19] J. E. Lee, R. Jin, and A. K. Jain. Rank-based distance metric learning: An application to image retrieval. In *CVPR*, 2008. 4

[20] R. Litman, A. Bronstein, M. Bronstein, and U. Castellani. Supervised learning of bag-of-features shape descriptors using sparse coding. In *Computer Graphics Forum*, volume 33, pages 127–136. Wiley Online Library, 2014. 2, 6, 7, 8

[21] J. Mei, M. Liu, H. Karimi, and H. Gao. Logdet divergence-based metric learning with triplet constraints and its applications. *Image Processing, IEEE Transactions on*, 23(11):4920–4931, Nov 2014. 6, 7

[22] A. Mignon and F. Jurie. Pcca: A new approach for distance learning from sparse pairwise constraints. In *CVPR*, 2012. 2

[23] M. Munaro, A. Basso, A. Fossati, L. Van Gool, and E. Menegatti. 3d reconstruction of freely moving persons for re-identification with a depth sensor. In *ICRA*, 2014. 1, 6

[24] A. Petrelli and L. Di Stefano. On the repeatability of the local reference frame for partial shape matching. In *ICCV*, 2011. 2

[25] D. Pickup, X. Sun, P. L. Rosin, R. R. Martin, Z. Cheng, Z. Lian, M. Aono, A. Ben Hamza, A. Bronstein, M. Bronstein, and 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. 1, 2, 6

[26] M. Reuter, F.-E. Wolter, and N. Peinecke. Laplace-beltrami spectra as 'shape-dna' of surfaces and solids. *Comput. Aided Des.*, 2006. 2

[27] K. Robinette, S. Blackwell, H. Daanen, M. Boehmer, S. Fleming, T. Brill, D. Hoeferlin, and D. Burnsides. Civilian american and european surface anthropometry resource final report. *AFRL-HE-WP-TR*, 2002. 4, 5

[28] R. B. Rusu and S. Cousins. 3d is here: Point cloud library (pcl). In *International Conference on Robotics and Automation*, 2011. 3, 6, 7, 8

[29] S. Shalev-Shwartz, Y. Singer, and A. Y. Ng. Online and batch learning of pseudo-metrics. In *ICML*, 2004. 4, 5

[30] P. Shilane, P. Min, M. Kazhdan, and T. Funkhouser. The princeton shape benchmark. In *Shape modeling applications Proceedings*, 2004. 6

[31] L. Sigal, M. Isard, H. Haussecker, and M. Black. Loose-limbed people: Estimating 3d human pose and motion using non-parametric belief propagation. *IJCV*, 2012. 2

[32] V. Singh, Y.-j. Chang, K. Ma, M. Wels, G. Soza, and T. Chen. Estimating a patient surface model for optimizing the medical scanning workflow. In *MICCAI*, 2014. 1, 2

[33] A. Tsoli, M. Loper, and M. Black. Model-based anthropometry: Predicting measurements from 3d human scans in multiple poses. In *WACV*, 2014. 1, 2

[34] C. Wachinger, P. Golland, and M. Reuter. Brainprint : Identifying subjects by their brain. In *MICCAI*, 2014. 2

[35] K. Weinberger and L. Saul. Distance metric learning for large margin nearest neighbor classification. *The Journal of Machine Learning Research*, 10:207–244, 2009. 4, 6

[36] A. Weiss, D. Hirshberg, and M. Black. Home 3d body scans from noisy image and range data. In *ICCV*, 2011. 2, 3

[37] A. Zaharescu, E. Boyer, K. Varanasi, and R. Horaud. Surface feature detection and description with applications to mesh matching. In *CVPR*, 2009. 2

[38] W.-S. Zheng, S. Gong, and T. Xiang. Reidentification by relative distance comparison. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 35(3):653–668, 2013. 2