

A 2D Human Body Model Dressed in Eigen Clothing

Peng Guan* Oren Freifeld† Michael J. Black*

*Department of Computer Science †Division of Applied Mathematics
Brown University, Providence, RI 02912, USA
{pguan,black}@cs.brown.edu, freifeld@dam.brown.edu

Abstract. Detection, tracking, segmentation and pose estimation of people in monocular images are widely studied. Two-dimensional models of the human body are extensively used, however, they are typically fairly crude, representing the body either as a rough outline or in terms of articulated geometric primitives. We describe a new 2D model of the human body contour that combines an underlying naked body with a low-dimensional clothing model. The naked body is represented as a Contour Person that can take on a wide variety of poses and body shapes. Clothing is represented as a deformation from the underlying body contour. This deformation is learned from training examples using principal component analysis to produce *eigen clothing*. We find that the statistics of clothing deformations are skewed and we model the *a priori* probability of these deformations using a Beta distribution. The resulting generative model captures realistic human forms in monocular images and is used to infer 2D body shape and pose under clothing. We also use the coefficients of the eigen clothing to recognize different categories of clothing on dressed people. The method is evaluated quantitatively on synthetic and real images and achieves better accuracy than previous methods for estimating body shape under clothing.

1 Introduction

Two-dimensional models of the human body are widely used in computer vision tasks such as pose estimation, segmentation, pedestrian detection and tracking. Such 2D models offer representational and computational simplicity and are often preferred over 3D models for applications involving monocular images and video. These models typically represent the shape of the human body coarsely, for example as a collection of articulated rectangular patches [1–4]. None of these methods explicitly models how clothing influences human shape. Here we propose a new fully generative 2D model that decomposes human body shape into two components: 1) the shape of the naked body and 2) the shape of clothing relative to the underlying body. The naked body shape is represented by a 2D articulated Contour Person (CP) [5] model that is learned from examples. The CP model realistically represents the human form but does not model clothing. Given training examples of people in clothing with known 2D body shape, we

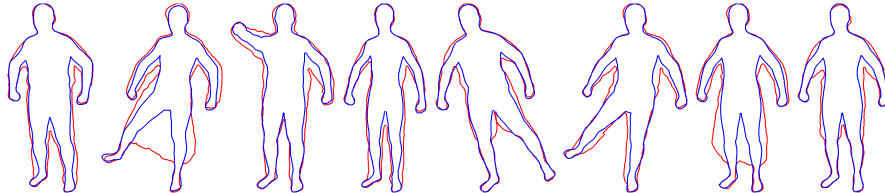


Fig. 1. Samples from the Dressed Contour Person model. Different body shapes and poses (blue) are dressed in different types of eigen clothing (red).

compute how clothing deviates from the naked body to learn a low-dimensional model of this deformation. We call the resulting generative model the *Dressed Contour Person* (DCP) and samples from this model are shown in Fig. 1.

The DCP model can be used just like previous models for person detection, tracking, etc. yet it has several benefits. The key idea is to separate the modeling of the underlying body from its clothed appearance. By explicitly modeling clothing we infer the most likely naked body shape from images of clothed people. We also solve for the pose of the underlying body, which is useful for applications in human motion understanding. The learned model accurately captures the contours of clothed people making it more appropriate for tracking and segmentation. Finally, the model supports new applications such as the recognition of different types of clothing from images of dressed people.

There are several novel properties of the DCP model. First we define *eigen clothing* to model deformation from an underlying 2D body contour. Given training samples of clothed body contours, where the naked shape of the person is known, we align the naked contour with the clothing contour to compute the deformation. The eigen-clothing model is learned using principal component analysis (PCA) applied to these deformations. A given CP model is then “clothed” by defining a set of linear coefficients that produce a deformation from the naked contour. This is illustrated in Fig. 1.

There is one problem, however, with this approach. As others have noted, clothing generally makes the body larger [6, 7]. A standard eigen-model of clothing could generate “negative clothing” by varying the linear coefficients outside the range of the training samples. While non-negative matrix factorization could be used to learn the clothing model, we show that a simple prior on the eigen coefficients addresses the issue. In particular, we show that the eigen coefficients describing clothing deformations are not Gaussian and we model them using Beta distributions that capture their asymmetric nature.

We also demonstrate the estimation of a person’s 2D body shape under clothing from a single image. Previous work on estimating body shape under clothing has either used multiple images [6] or laser range scan data [7]. These previous approaches also did not actually model clothing but rather tried to ignore it. Both of the above methods try to fit a naked body that lies inside the measurements (images or range scans) while strongly penalizing shapes that are “larger” than the observations. We show that there is a real advantage to a principled

statistical model of clothing. Specifically we show accuracy in estimating naked body shape that exceeds that of Bălan and Black [6], while only using one uncalibrated image as opposed to four calibrated views.

Finally we introduce a new problem of clothing category recognition. We show that the eigen coefficients of clothing deformations are distinctive and can be used to recognize different categories of clothing such as long pants, skirts, short pants, sleeveless tops, etc. Clothing category recognition could be useful for person identification, image search and various retail clothing applications.

In summary, the key contributions of this paper include: 1) the first model of 2D eigen clothing; 2) a full generative 2D model of dressed body shape that is based on an underlying naked model with clothing deformation; 3) the inference of 2D body shape under clothing that uses an explicit model of clothing; 4) shape under clothing in a single image; 5) avoiding “negative clothing” by modeling the skewed statistics of the eigen-clothing coefficients; 6) the first shape-based recognition of clothing categories on dressed humans.

2 Related Work

Very little work in computer vision has focused on modeling humans in clothing. What work there is focuses on modeling 3D human shape *under clothing* without actually *modeling* the clothing itself. Bălan and Black [6] present a system based on the 3D SCAPE [8] body model that uses multiple camera views to infer the body shape. They make the assumption that the estimated body shape belongs to a parametric family of 3D shapes that are learned from training bodies. They fit the body to image silhouettes and penalize estimated body shapes that extend beyond the silhouette more heavily than those that are fully inside. This models the assumption that body shape should lie inside the visual hull defined by the clothed body. In essence their method attempts to be robust to clothing by ignoring it. More recently, Hasler *et al.* [7] take a similar approach to fitting a 3D body to laser range scans of dressed humans. Rosenhahn *et al.* [9] model clothing explicitly on a 3D mesh but do so for the purpose of tracking, not body shape estimation. Our approach differs from the above by focusing on 2D models and explicitly modeling clothing deformations on the body using eigen clothing.

The vast majority of work on modeling clothing has focused on the recovery of 3D mesh models of the clothing itself (e.g. [10]). We know of no work on modeling eigen clothing or 2D clothing deformation models. Of course other types of 2D deformable shape models (e.g. active shape models [15]) have been widely used in vision applications and a review is beyond our scope.

Almost all work on 2D person detection and pose estimation implicitly assumes the people are clothed (though [11] is a notable counterexample). Despite this, few authors have looked at using clothing in the process [12] or at actually using a model of the clothing. Recent work by Bourdev and Malik [13] learns body part detectors that include upper and lower clothing regions. They do not model the clothing shape or body shape underneath and do not actually recognize different types of clothing. One recent paper does try to recognize clothing

types for Internet retail applications [14]. The garments, however, are assumed to lie in a plane and are hence not actually on a human body.

3 The Contour Person Model

We start with a Contour Person (CP) model [5], which is a low-dimensional, realistic, parameterized generative model of 2D human shape and pose. The CP model is learned from examples created by 2D projections of multiple shapes and poses generated from a 3D body model such as SCAPE [8]. The CP model is based on a template, T , corresponding to a reference contour that can be deformed into a new pose and shape. This deformation is parameterized and factors the changes of a person’s 2D shape due to pose, body shape, and the parameters of the viewing camera. This factorization allows different causes of the shape change to be modelled separately. Let $B_T(\Theta) = (x_1, y_1, x_2, y_2, \dots, x_N, y_N)^T$ denote the parametric form of the CP, where N is the number of contour points and Θ is a vector of parameters that controls the deformation with respect to T . The CP model represents a wide range of 2D body shapes and poses, but only does so for naked bodies. Examples of such body contours, $B_T(\Theta)$, are shown in blue in Fig. 1. See Freifeld *et al.* [5] for mathematical details.

The CP model may contain internal or occluded portions of the body contour. However, here our clothing training data consists only of silhouettes with no internal structure. Consequently, we restrict the poses we consider and define $B_T(\Theta)$ to be a CP model corresponding to a bounding body contour without holes. In future work, we will generalize the DCP model to take advantage of the ability of the CP to accommodate self occlusions.

4 Clothing Model

We directly model the deformation from a naked body contour to a clothed body by virtually “dressing” the naked contour with clothing. We start with a training set (described below) of clothing outlines and corresponding naked body outlines underneath. The CP model is first fit to the naked body outline to obtain a CP representation. For each point on the CP, we compute the corresponding point on the clothing outline (described below) and learn a point displacement model using PCA [15]. We further learn a prior over the PCA coefficients using a Beta distribution to prevent infeasible displacements (i.e. “negative clothing”).

The DCP model can be thought of as having two “layers” that decouple the modeling of body pose and shape from the modeling of clothing. The first layer generates a naked body deformation from the template contour and the second layer models clothing deformation from this deformed naked contour. The first layer is the CP model, which is compositional in nature and based on deformations of line segments (see [5]). The second layer, described here, is simpler and is based directly on displacements of contour points. The layered representation is desirable because it allows constraints to be imposed independently on the body and the clothing.

4.1 Data sets

Our method requires training contours of people in clothing for which we know the true underlying naked body shape. We describe two such training sets below. **Synthetic data set.** Synthetic data provides ground truth body shapes that enable accurate quantitative evaluation. We use 3D body meshes generated from the CAESAR database (SAE International) of laser range scans and dress these bodies in simulated clothing (Fig. 2). We used 60 male and 100 female bodies spanning a variety of heights and weights and use commercial software (OptiTex International, Israel) to generate realistic virtual clothing. The clothing simulation produces a secondary 3D mesh that lies on top of the underlying body mesh by construction. Given a particular camera view, we project the body mesh into the image to extract the body outline and do the same for the combined body and clothing meshes. This provides a pair of training outlines.

For the synthetic dataset we restrict the clothing to a single type (Army Physical Training Uniforms) but in different sizes, as appropriate for the body model. While narrow, this dataset provides nearly perfect training data and ground truth for evaluation.

Real data set. To model real people in real clothing we use the dataset described by Bălan and Black in [6] (Fig. 2) which contains images of 6 subjects (3 males, 3 females) captured by 4 cameras in two conditions: 1) the “naked condition” in which the subjects wear tight fitting clothing; 2) the “clothed condition” in which they wear different types of “street” clothing. The dataset contains four synchronously captured images of each subject, in each condition, in a fixed set of 11 postures. For each posture the subjects are dressed in 6-10 different sets of clothing (trials). Overall there are 47 trials with a total of 235 unique combinations of people, clothing and poses.

For each image of a dressed person, we use standard background subtraction [6] to estimate the clothed body silhouette and extract the outline. To obtain the underlying naked body contours, we fit a 3D parametric body model using the 4 camera views in the naked condition [6]. We take this estimated 3D body shape to be the true body shape. We then hold this body shape fixed while estimating the 3D pose of the body in every clothing trial using the method of [6] which is robust to clothing and uses 4 camera views.

The process produces a 3D body of the “true” shape, in the correct pose, for every trial. We project the outline of this 3D body into a selected camera view to produce a training 2D body contour. We then pair this with the segmented clothed body in that view. Note that the fitting of the 3D body to the image data is not perfect and, in some cases, the body contour actually lies outside the clothing contour. This does not cause significant problems and this dataset provides a level of realism and variability not found in the synthetic dataset.

4.2 Correspondence

Given the naked and clothed outlines defined above, we need to know the correspondence between them. Defining the correspondence between the naked outline

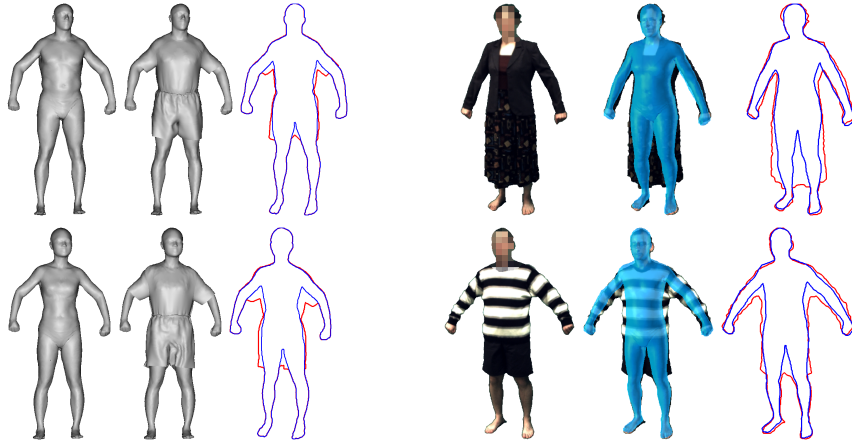


Fig. 2. Example training data. Left: Pairs of synthetic 3D bodies, unclad and clothed. Projecting the silhouette contours of these pairs produces training contours. Right: Training contours derived from multi-camera data (see text); the estimated ground truth 3D body is shown as a translucent overlay.

and the clothing outline is nontrivial and how it is done is important. Baumberg and Hogg, for example, model the outline of pedestrians (in clothing) using PCA [17]. In their work, correspondence is simply computed by parameterizing all training contours with a fixed number of evenly sampled points. Incorrect correspondence (i.e. sliding of points along the contour) results in eigen shapes that are not representative of the true deformations of the contours.

Instead, we start with the trained parametric CP representation $B_T(\theta)$ and optimize it to fit the 2D naked body that minimizes the difference between the CP silhouette and the naked body silhouette. This gives a CP representation of the naked body that consists of $N = 1120$ points. We then densely sample M points on clothing outline, where $M \gg N$ and select the N clothing contour points that best correspond to the CP points. During matching, the relative order of the points is maintained to guarantee the coherence of the match. Let the CP contour be represented by a list of points $P = \{p_1, p_2, \dots, p_N\}$ and let the sampled clothing outline be represented by $Q = \{q_1, q_2, \dots, q_M\}$. We pick a subset of N points $G = \{q_{k_1}, q_{k_2}, \dots, q_{k_N}\}$ from Q that minimizes $\sum_{i=1}^N \|p_i - q_{k_i}\|^2$ over the indices k_i such that the ordering, $k_r < k_s$, is preserved for $1 \leq r < s \leq N$. We use the dynamic programming method proposed in [18]. Example alignments are shown in Fig. 3.

4.3 Point displacement model

We convert the point list G to a vector $\hat{G} = (x_1, y_1, x_2, y_2, \dots, x_N, y_N)^T$ and now we have $B_T(\theta)$ for the naked body contour and \hat{G} for clothing contour, both of which have N corresponding points. The clothing displacement for a

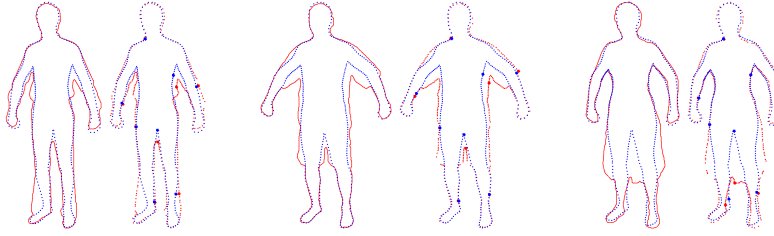


Fig. 3. Correspondence between body and clothing contours. In each pair: the left image shows the sample points of the body contour in blue and the densely sampled clothing contour in red. The right image shows the final sub-sampled clothing contour with a few matching points highlighted as larger dots. Nearby dots illustrate corresponding points (in some cases they are on top of each other).

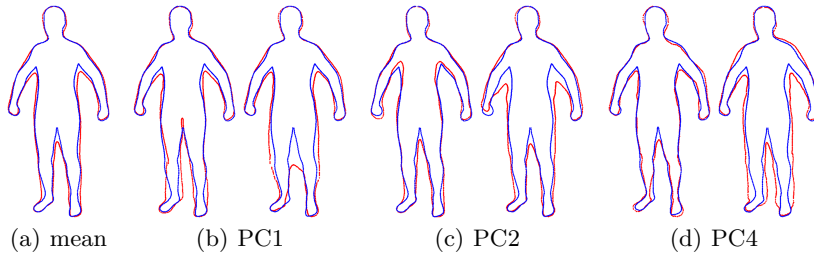


Fig. 4. Eigen clothing. The blue contour is always the same naked shape. The red contour shows the mean clothing contour (a) and ± 3 std from the mean for several principal components (b)-(d).

particular training example, i is then defined by $\delta_i = \hat{G}_i - B_T(\Theta_i)$. We collect the training displacements into a matrix and perform PCA. We take the first 8 principal components accounting for around 90% of the variance to define the eigen-clothing model. Figure 4 shows the mean and first few clothing eigenvectors for the real data set. This illustrates how the principal components can account for various garments such as long pants, skirts, baggy shirts, etc. Note that simply varying the principal components can produce “negative clothing” that extends inside the blue body contour. We address this in the following section.

Using this model we generate new body shapes in new types of clothing by first sampling CP parameters Θ to create a naked body contour $B_T(\Theta)$ and then using the following equation to generate a clothed body

$$C(\Theta, \eta) = B_T(\Theta) + \Delta_{mean} + \sum_{i=1}^{N_\eta} \eta_i \cdot \Delta_i \quad (1)$$

where N_η is the number of eigenvectors used, the η_i ’s are coefficients, Δ_{mean} is the mean clothing displacement, and Δ_i is the i^{th} eigen-clothing vector.

4.4 Prior on point displacement

Although the PCA model captures clothing deformation, it allows point displacements in both inward and outward directions, which violates our assumption that clothing only makes the body appear bigger. This assumption is confirmed by examining the statistics of the linear eigen coefficients in the training data. Figure 5 shows several such distributions, which may be skewed or symmetric. In particular we find that coefficients for the principal components that capture the most variance are typically positively or negatively skewed while coefficients for the lower-variance components tend to be more normally distributed. The first few eigenvectors capture the gross clothing displacements, which are always away from the body. Of course clothing also exhibits many fine details and folds and these are captured by the lower variance eigenvectors. These “detail” eigenvectors modify the main clothing contour both positively and negatively (e.g. out and in) and hence tend to have more symmetric statistics.

Based on the observation of natural clothing statistics, we learn a prior on the PCA coefficients to penalize infeasible clothing displacements. We make the assumption that the eigenvectors are independent (not necessarily true since the data is not Gaussian) and independently model a prior on each coefficient using a Beta distribution. The Beta distribution is defined on $[0, 1]$ and is characterized by two parameters α and β that can be varied to capture a range of distributions including positively skewed, negatively skewed and symmetric shapes:

$$\mathbf{Beta}(x; \alpha, \beta) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1} (1-x)^{\beta-1}. \quad (2)$$

Given L training body/clothing pairs, and the associated clothing displacements, we project each displacement onto the PCA space to obtain coefficient η_m^l for instance l , ($l \in [1, L]$), on eigenvector m . We normalize $\eta_m^1, \eta_m^2, \dots, \eta_m^L$ to $[0, 1]$ to obtain $\tilde{\eta}_m^1, \tilde{\eta}_m^2, \dots, \tilde{\eta}_m^L$ and fit these with the Beta distribution. The probability of observing a normalized coefficient \tilde{x}_m for the m^{th} eigenvector is given by $\mathbf{Beta}(\tilde{x}_m, \alpha_m, \beta_m)$, where α_m and β_m are the estimated parameters of the Beta distribution. If we observe a coefficient during testing that is out of the scope of the training coefficients, we threshold it to be between the minimal and maximal value in the training set and normalize it to compute its prior probability. If thresholded, however, we still use the original value to reconstruct the shape. Figure 5 shows how the Beta function can represent a variety of differently shaped distributions of clothing displacement coefficients.

4.5 Inference

The inference problem is to estimate the latent variables Θ and η by only observing a single image of a person in clothing. We define a likelihood function in terms of silhouette overlap. We adopt a generative approach in which $C(\Theta, \eta)$, the clothed body (Eq. 1), defines an estimated silhouette, $S^e(C(\Theta, \eta))$, and compare it with the observed image silhouette S^o . We follow [6] and define the asymmetric distance between silhouettes S^r and S^t as $d(S^r, S^t) = \frac{\sum_{i,j} S_{i,j}^r H_{i,j}(S^t)}{\sum S_{i,j}^r}$, where

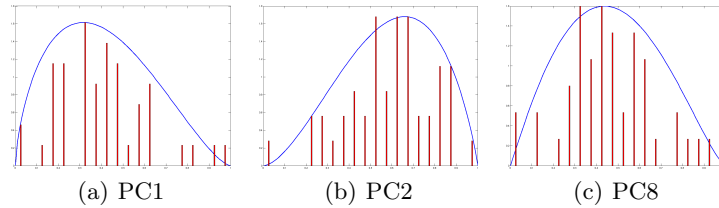


Fig. 5. The statistics of clothing displacements. Example histograms and Beta distribution fits to linear eigen-clothing coefficients. Note the skew that results from the fact that clothing generally makes the body appear larger.

$S_{i,j}^r$ is a pixel inside silhouette S^r and $H_{i,j}(S^t)$ is a distance function which is zero if pixel (i,j) is inside S^t and is the distance to the closest point on the boundary of S^t if it is outside.

We then define the data term as the following symmetric data error function

$$E_{data}(\Theta, \eta) = d(S^e(C(\Theta, \eta)), S^o) + d(S^o, S^e(C(\Theta, \eta))). \quad (3)$$

The first part of Eq. 3 penalizes the regions of the synthesized clothing instance $S^e(C(\Theta, \eta))$ that fall outside the observed clothing silhouette S^o , and the second part makes $S^e(C(\Theta, \eta))$ explain S^o as much as possible.

E_{data} alone is not sufficient to estimate Θ and η correctly, because there are ambiguities in estimating smaller bodies with larger clothing and larger bodies with smaller clothing. As was mentioned in Sec. 4.4, we use the Beta prior to penalize unlikely displacements. Recall that $\tilde{\eta}_m$ represents the normalized coefficient for the m^{th} basis. The prior term is defined as

$$E_{prior}(\eta) = - \sum_m \log(\mathbf{Beta}(\tilde{\eta}_m, \alpha_m, \beta_m)). \quad (4)$$

The final energy function we minimize is

$$E(\Theta, \eta) = E_{data}(\Theta, \eta) + \lambda E_{prior}(\eta) \quad (5)$$

where λ indicates the importance of the prior. Problems with “negative clothing” and clothing that is unusually large are avoided due to the prior. Optimization is performed using MATLAB’s `fminsearch` function.

5 Results

We consider two novel applications of the proposed method. The first is the estimation of 2D body shape under clothing given a single image of a clothed person. The second is the recognition of different clothing categories by classifying the estimated clothing deformation parameters. We evaluate our model on three tasks: body shape estimation from synthetic data, body shape estimation from real data, and clothing type classification from real data. We compare

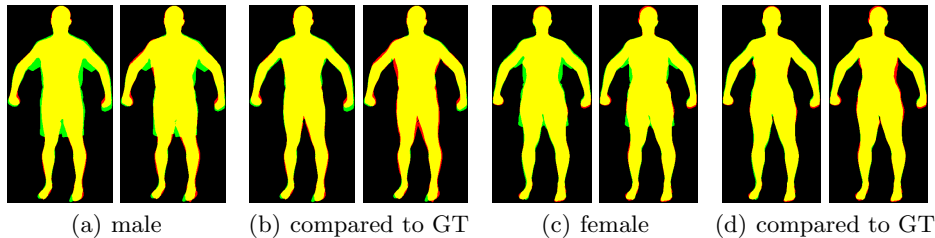


Fig. 6. Synthetic data results. For each pair of images, the DCP result is on the left and NM result is on the right. The first pair shows an estimated body silhouette (red) overlaid on the clothing silhouette (green); overlapped regions are yellow. The second pair shows the estimated body (red) overlaid on the ground truth (GT) body (green). The third and fourth pairs show the same but for a female. NM typically overestimates the size of the body.

the results of the first two tasks with approaches that do not explicitly model clothing deformation.

Body estimation under clothing from synthetic data. We use the synthetic dataset of 60 males and 100 females, in and out of synthetic clothing, as described above. We randomly select 30 males and 50 females as the training set and the remaining 80 bodies as the test set. A gender-specific CP model is learned for males and females separately while a gender-neutral eigen model is learned for clothing deformations. We estimate the underlying bodies for the test samples using the *Dressed Contour Person* (DCP) and measure the estimation error as

$$err(S^{EST}, S^{GT}) = \frac{\sum_{i,j} |S_{i,j}^{EST} - S_{i,j}^{GT}|}{2 \sum_{i,j} S_{i,j}^{GT}} \quad (6)$$

where S^{EST} is a silhouette corresponding to the estimated naked body contour and S^{GT} is the ground truth underlying naked body silhouette. The results of DCP are also compared with a naive method (NM) in which we simply fit the CP model to the image observations of clothed people. As in [6], the NM attempts account for clothing by penalizing contours more if the estimated body silhouette falls outside of the clothing observation than if it does not fully explain the clothing observation. The average estimation errors obtained with NM for males and females are 0.0456 and 0.0472 respectively while DCP achieves 0.0316 and 0.0308. Our DCP model improves accuracies over NM by 30% (male) and 35% (female) relatively. While the synthetic dataset has only one clothing type, the bodies span a wide range of shapes. The results show a principled advantage to modeling clothing deformation compared with ignoring clothing. Figure 6 shows some representative results from the test set.

Body estimation under clothing from real data. Figure 7 shows 8 different poses from the real dataset (Sec. 4.1). For each pose there are 47 examples having unique combinations of subjects and clothing types. Since the number of body/clothing pairs is limited in each pose, we use a leave-one-out strategy where

Table 1. Comparison on real data: DCP, NM, and NP3D methods (see text)

Method, AEE	Pose1	Pose2	Pose3	Pose4	Pose5	Pose6	Pose7	Pose8	Average
DCP	0.0372	0.0525	0.0508	0.0437	0.0433	0.0451	0.0503	0.0668	0.0487
NP3D	0.0411	0.0628	0.0562	0.0484	0.0494	0.046	0.0472	0.0723	0.0529
NM	0.0865	0.0912	0.0846	0.0835	0.0877	0.0921	0.0902	0.1184	0.0918
Significance (<i>p</i> -value)									
DCP vs NP3D	0.38	0.13	0.34	0.46	0.36	0.89	0.66	0.54	0.07
DCP vs NM	6.4e-7	4.9e-4	2.1e-4	2.1e-4	6.7e-8	1.0e-5	1.0e-6	2.3e-4	9.9e-17

we estimate the body of instance i using the eigen-clothing model learned from all remaining 46 instances excluding i . We use DCP to estimate the underlying body shape for a total of $47 * 8 = 376$ instances (Fig. 7) and compare the results with two other methods: 1) NM described in the previous experiment; and 2) “Naked People estimation in 3D”(NP3D) proposed in [6]. Since DCP and NM are 2D methods using a 2D CP model, they only use one camera view. NP3D, however, estimates a 3D body model from four camera views [6]. To compare with NP3D we project the estimated body from NP3D into the image corresponding to the camera view used by our method.

Table 1 shows the Average Estimation Error (AEE) computed by averaging $err(\cdot, \cdot)$ (Eq. 6) over the 47 instances for each pose (or over all poses in the last column). Figure 8 shows details of the fitting results. We find that DCP has lower error than both NM and NP3D. In the case of NM these differences are statistically significant (paired t-test, $p < 0.05$) for all poses and in the aggregate. While DCP has lower error than NP3D in all but one pose, and lower error overall, the differences are not significant at the $p < 0.05$ level. Recall that NP3D is using significantly more information. These results suggest that using a learned statistical model of clothing is preferable to simply trying to ignore clothing [6].

Clothing category recognition. We now ask whether the clothing deformation coefficients contain enough information about clothing shape to allow the classification of different types of clothing. Note that this task involves recognizing clothing *on the body* as it is worn by real people. We separate upper clothing and lower clothing and define 7 different categories (as color coded in Fig. 9).

We use a simple nearest neighbor (NN) classifier with Euclidean distances computed from the coefficients along the first 8 principal components. Since we have a limited number of clothing instances (47) for each pose, we use a leave-one-out strategy and assume that we know the categories of all the instances except the one we are testing. Each instance is then assigned a category for both upper clothing and lower clothing based on its nearest neighbor. Classification results are shown in Fig. 9 along with chance performance for this task.



Fig. 7. Sample DCP results of estimated body shape overlaid on clothing. The estimated body contour and synthesized clothing contour are depicted by blue and red outlines respectively. Body shape is the transparent region encompassed by the body contour. Results are shown for a variety of poses (left to right: 1-8) and viewing directions.

6 Conclusions

We have presented a new generative model of the 2D human body that combines an underlying Contour Person representation of the naked body and layers on top of this a clothing deformation model. This goes beyond previous work to learn an eigen model of clothing deformation from examples and defines a prior over possible deformations to prevent “negative clothing”. While previous work has examined 3D body models captured with multiple cameras or laser range scanners, we argue that many computer vision applications use 2D body models and that these applications will benefit from a more realistic generative model of clothed body shape. By modeling clothing deformations we estimate 2D body shape more accurately and even out-perform previous multi-camera systems on estimating shape under clothing. Finally we define a new problem of clothing category recognition on the human body and show how the coefficients of the estimated eigen clothing can be used for this purpose. This new dressed person model is low dimensional and expressive, making it applicable to many problems including 2D human pose estimation, tracking, detection and segmentation.

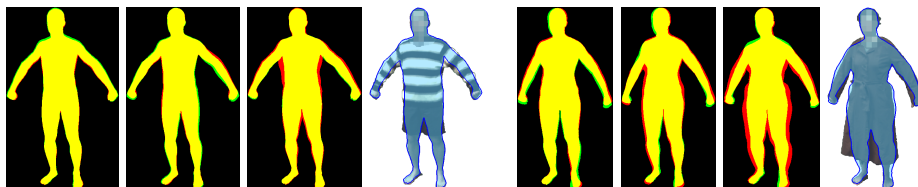


Fig. 8. Comparisons of DCP, NM, and NP3D. For each group of images: the first 3 images (left to right) show overlap of the estimated silhouette (red) and the ground truth silhouette (green) for DCP, NP3D, and NM (yellow is overlap); the 4th image shows the body estimated by NM overlaid on a clothing image. NM overestimates body shape as expected.

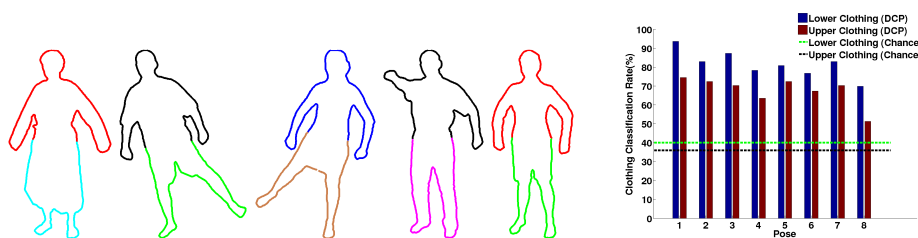


Fig. 9. Color coded clothing type. We consider three types of upper clothing: long sleeves (red), short sleeves (black) and sleeveless tops (blue) and four types of lower clothing: short pants (green), long pants (magenta), short skirts (cyan), and long skirts (cyan). Classification results for the 7 clothing types in all 8 poses are shown in the right figure compared to “Chance”.

Our method does have some limitations. The method assumes there is a correspondence between body contour points and clothing contour points. When there is significant limb self occlusion, the clothing silhouette may not contain features that *correspond to* that limb. Dealing with significant self occlusion is future work. Also, here we assume that the rough viewing direction (frontal or side) and rough pose are known.

There are several directions for future work. First, we plan to model clothing deformation as a function of human movement. This may require a model more like the original CP model in which deformations are defined as scaled rotations of contour line segments [5]. This representation allows the factoring of contour changes into different deformations that can be composed. Second, we will explore what we call “eigen separates”; that is, separate eigen models for tops and bottoms as well as for hair/hats and shoes. Having separate eigen spaces reduces the amount of training data required to capture a wide range of variations. Finally we plan to extend these methods to model 3D clothing deformations from a 3D body model. Again data acquisition for 3D clothed and unclothed training data is very difficult, and we plan to use realistic physics simulation of clothing.

Acknowledgements. This work was supported in part by NIH EUREKA 1R01NS066311-01 and NSF IIS-0812364. We thank L. Reiss for generating the synthetic training data, A. Balan for assistance with the real dataset, and S. Zuffi for helpful comments.

References

1. Mori, G., Ren, X., Efros, A., Malik, J.: Finding and tracking people from the bottom up. *CVPR*. Vol. 2. (2003) 467–474
2. Felzenszwalb, P.F.: Representation and detection of deformable shapes. *PAMI*. **27** (2004) 208–220
3. Hinton, G.E.: Using relaxation to find a puppet. *Proc. of the A.I.S.B. Summer Conference*. (1976) 148–157
4. Ju, S.X., Black, M.J., Yacoob, Y.: Cardboard people: A parameterized model of articulated motion. *Face and Gesture*. (1996) 38–44
5. Freifeld, O., Weiss, A., Zuffi, S., Black, M.J.: Contour people: A parameterized model of 2D articulated human shape *CVPR*. (2010)
6. Balan, A., Black, M.: The naked truth: Estimating body shape under clothing. *ECCV*. (2008) 15–29
7. Hasler, N., Stoll, C., Rosenhahn, B., Thormählen, T., Seidel, H.: Estimating body shape of dressed humans. *Computers & Graphics* **33** (2009) 211–216
8. Anguelov, D., Srinivasan, P., Koller, D., Thrun, S., Rodgers, J., Davis, J.: SCAPE: shape completion and animation of people. *ACM TOG (SIGGRAPH)* **24** (2005) 408–416
9. Rosenhahn, B., Kersting, U., Powell, K., Klette, R., Klette, G., Seidel, H.: A system for articulated tracking incorporating a clothing model. *Mach. Vis. App.* **18** (2007) 25–40
10. White, R., Crane, K., Forsyth, D.: Capturing and animating occluded cloth. *ACM TOG (SIGGRAPH)* **26** (2007) 34
11. Fleck, M., Forsyth, D., Bregler, C.: Finding naked people. *ECCV*. Vol. 2. (1996) 592–602
12. Sprague, N., Luo, J.: Clothed people detection in still images. *ICPR*. Vol. 3. (2002) 585–589
13. Bourdev, L., Malik, J.: Poselets: Body part detectors trained using 3D human pose annotations. *ICCV*. (2009)
14. Tseng, C.H., Hung, S.S., Tsay, J.J., Tsaih, D.: An efficient garment visual search based on shape context. *W. Trans. on Comp. (WSEAS)* **8** (2009) 1195–1204
15. Cootes, T., Taylor, C.: Active shape models—smart snakes. *BMVC* (1992) 266–275
16. Balan, A., Sigal, L., Black, M., Davis, J., Haussecker, H.: Detailed human shape and pose from images. *CVPR*. (2007) 1–8
17. Baumberg, A., Hogg, D.: Learning flexible models from image sequences. *ECCV*. (1994) 299–308
18. Oliveira, F., Tavares, J.: Algorithm of dynamic programming for optimization of the global matching between two contours defined by ordered points. *Computer Modeling in Eng. & Sciences* **31** (2008) 1–11
19. Sigal, L., Balan, A., Black, M.: Combined discriminative and generative articulated pose and non-rigid shape estimation. *Advances in Neural Information Processing Systems 20*, (NIPS). MIT Press. (2008) 1337–1344