基于不完整语义信息的三维人体重建

摘 要: 通过人体尺寸参数进行三维人体重建是在线购物，VR技术，游戏建模等领域的基础研究问题。现有的方法创建一个三维模型通常需要复杂的测量或者高精度的扫描仪。这些方法前者需要大量的精力投入而后者一般人又很难接触到高端的扫描仪。本文提出了一种基于不完整语义信息的三维重建方法。用户只需要提供有限个尺寸信息，避免了复杂的尺寸测量，就可以完成人体重建。我们的算法大概可以分为3步，首先我们将数据集的模型尺寸看做高维空间中的一个点。所有的这些点可以拟合出一个流形。输入有限个尺寸就可以预测出剩余尺寸。其次通过机器学习的方法求出身材参数，通过身材参数重建模型，再对获得的近似解作为优化初值。最后为了能够生成更加精确的人体模型，应该满足用户测量到的尺寸与生成的人体模型上求出的尺寸是没有误差的。为了达到这一目的以上一步结果为优化问题初值，以尺寸为约束进行优化求解。通过大量实验表现，我们的方法能得到生成速度更快，误差更小的模型。

关键词: 人体重建;语义重建;三维重建

**Three-dimensional human reconstruction based on incomplete semantic information**

**Abstract**: 3D human reconstruction through human size parameters is a basic research problem in online shopping, VR technology, game modeling and other fields. Existing approaches for creating a three-dimensional model usually require complex measurements or high-precision scanners. These methods involve heavy expense cost and the latter have little access to high-end scanners. this paper proposes a 3d reconstruction method based on incomplete semantic information. User only need to provide limited anthropometric information, avoid complex anthropometric measurement, can complete human body reconstruction. Our algorithm can be roughly divided into three steps. first we consider the model size of the dataset as a point in a high-dimensional space. All these points can fit a manifold. Input a finite size to predict the remaining size. Secondly, the figure parameters are obtained by machine learning method, and the model is reconstructed by figure parameters, and then the approximate solution obtained is taken as the initial value of optimization. Finally, in order to be able to generate more accurate mannequins, it should be satisfied that the user measured the size and generated mannequins on the size is no error. In order to achieve this goal, the above step result is the initial value of the optimization problem, and the size constraint is used to optimize the solution. Through a lot of experiments, our method can get the model with faster generation speed and less error.

**Key words**: human reconstruction;semantic reconstruction; 3D reconstruction

1. 引言

*【一、研究的理由目的背景】*无论是虚拟试衣[1]、医学研究[2]、游戏创作还是电影制作都有虚拟人物模型的身影。例如，调查表明[1]，网络购物平台加入了虚拟试衣功能后，降低了40%的退货率。并且大大减少了用户等待周期，提升了用户的购物体验。***[1. 同类研究中，现在国内外研究观点和现状]***研究员们提出了很多方法来估计人体身材，大致可为非参数化方法及参数化方法。参数化人体模型是参数化三维人体重建方法的基础，使用一组低维向量（即参数）来表达人体。非参数化方法相对参数化方法而言，其采用原始的顶点或体素来表达三维人体。非参数化三维人体重建通常不使用较强的身材先验，以扫描仪、深度相机或普通相机多角度拍摄下的信息作为输入来重建人体表面的几何位置。因为扫描仪设备昂贵，数据采集投入巨大，近些年越来越多的研究者都采用参数化化方法。Allen B等人[3]使用主成分分析法（Principal Component Analysis）来提取人体身材空间，学习了人体身高、体重与身材参数之间的线性关系，根据用户的身高、体重信息估计身材参数，进而重建三维人体。Anguelov等人[4]结合了以上的两种技术，提出了SCAPE（Shape Completion and Animation of People）模型。它是基于边变形的人体模型中最为经典的一个模型，该模型首次同时描述了姿势以及身材变形的三维人体模型。该模型取得了巨大的成功，被广泛应用到工业，医学，游戏设计等领域。并且后来许多参数化建模的方法都是在SCAPE模型基础上进行改进的。

***[2. 同行研究者正在做什么，研究的热点和焦点在哪里]***笔者提出的基于语义的方法，也是参数化方法的一种。近年来，许多研究者对以尺寸信息作为输入参数进行了大量的研究。Allen等人[3]学习了身高、体重与身材参数之间的线性关系，根据用户身高体重信息估计身材参数，进而重建三维人体。Zhou等人[5]将身高、体重、腰围及腿长信息线性映射至SCAPE模型身材参数。Wuhrer等人[6]采用34个身材尺寸作为输入，并将尺寸映射为PCA身材空间基的系数来重建三维人体。*【二、理论依据、实验基础和研究方法】****[3. 这些研究目前存在什么问题，或哪些方面需要改善，或哪些方面尚待研究][4. 作者准备解决其中哪方面的哪些问题]***不同的应用领域，对重建后三维人体模型的纹理信息要求不同。譬如在真实的试衣过程中，大部分用户并不了解自己全部的身材尺寸，又譬如在创建虚拟角色时，设计人员对人体形态的精确度要求并不是很高，只需要满足创建之后的角色模型拥有相对真实的外形并且能够很好的融入周围环境中即可。而方法[7] [6] 以全尺寸为输入，对用户输入做了限制，影响用户体验。为此，近来也有研究人员对输入限制做了改善，如，Zhang等人[8]提出了一种基于相关性的方法根据任意个数的身材尺寸预测剩余尺寸的方法。利用皮尔逊相关系数来描述两个变量值之间的线性相关大小。Zeng等人[9]使用基于链式方程的多元估算方法，估算剩余尺寸。Streuber等人[10]将高矮、胖瘦、体型、四肢长短等描述性信息映射至SMPL模型[11]的身材参数。目前的大部分通过学习的方法，数据集都是采集自成人，因此在生成青少年或者婴幼儿的时候会有较大的误差。***[5. 作者的研究思路和准备采用的方法（扣题）]***我们提出一种基于“数据拟合”的思想，来预测尺寸信息。我们提出了一个基于人体部分的参数化语义身材模型。给定任意个数的已知身材尺寸，我们估计所有身材尺寸，并根据每个尺寸学习对应的身材参数，进而重建三维人体。

*【三、预期的结果及其地位、作用和意义】****[6.预期研究能得到什么样的结果][7.这些研究结果在现实中能解决什么问题，其实用性如何]*** 我们的方法主要贡献有：

1. 通过用户提供有限的体型参数作为输入，这确保了在实际场景中更好的用户体验；
2. 能够根据用户提供的精确参数，进行精确的人体重建；
3. 对身体部位能够精准地进行语义控制；
4. 在广泛的验证中，我们的方法是符合用户感知的；
5. 相关工作

得益于高精度扫描仪的发展，让实验人员能够更容易获得精度更高的人体模型数据集。数据集的出现，例如CAESAR数据集[12]，MPI数据集[13]，MPII数据集[14]等，激发了大量关于学习三维人体统计模型的研究工作。研究员们提出了很多方法来估计人体身材，大致可为非参数化方法及参数化方法。参数化人体模型是参数化三维人体重建方法的基础，使用一组低维向量（即参数）来表达人体。非参数化方法的命名相对参数化方法而言，其采用原始的顶点或体素来表达三维人体。非参数化三维人体重建通常不使用较强的身材先验，以扫描仪、深度相机或普通相机多角度拍摄下的信息作为输入来重建人体表面的几何位置。因为扫描仪设备昂贵，数据采集投入巨大，近些年越来越多的研究者都采用参数化化方法。比如，Allen B等人[3]使用主成分分析法（Principal Component Analysis）来提前人体身材空间，学习了人体身高、体重与身材参数之间的线性关系，根据用户的身高、体重信息估计身材参数，进而重建三维人体。Anguelov等人[4]结合了以上的两种技术，提出了SCAPE（Shape Completion and Animation of People）模型。它是基于边变形的人体模型中最为经典的一个模型，该模型首次同时描述了姿势以及身材变形的三维人体模型。该模型取得了巨大的成功，被广泛应用到工业，医学，游戏设计等领域。参数化的方法根据输入的不同还可分为基于尺寸的方法[5,6,8]，基于图片的方法[5,15]，基于描述性词语[11,12]等方法。基于人体尺寸的方法大多都将所有的人体尺寸作为输入，Zhou等人[5]将身高、体重、腰围及腿长信息线性映射至SCAPE模型身材参数。采用半自动的方法初始化人体姿势及身材，然后利用隐式马尔科夫模型（HMM）匹配输入图像人体轮廓点与三维人体模型的投影轮廓点，通过最小化对应点之间的距离来优化求解人体参数。Wuhrer等人[6]采用34个身材尺寸作为输入，并将尺寸映射为PCA身材空间基的系数来重建三维人体，随后他们以学习到的重建人体为初始，分两步来完成重建。Yang等人[7]将23个尺寸线性映射为相对于模板人体网格的三角形变形。对于普通用户来说这无疑增加了使用使用难度。因此有研究者开发了能够通过局部尺寸预测完整尺寸的方法。Zhang等人[8]提出了一种基于相关性的方法根据任意个数的身材尺寸预测剩余尺寸的方法。Zeng等人[9]使用基于链式方程的多元估算方法，估算剩余尺寸。通过一些实验对比，我们的方法优于已有的方法。

1. 人体重建

本文算法可通过用户输入的不完整尺寸信息预测完整的尺寸信息，在通过完整的尺寸信息恢复与该尺寸对应的人体模型。即解决两个映射问题。(1) 不完整尺寸信息与完整尺寸信息的映射 (2) 尺寸信息与人体模型的映射。对于问题(1)可以通过拟合的方式解决。 对于问题(2)，在计算机中，存储的三维模型之间并没有形态上的差别，要建立从人体测量尺寸到人体三维模型之间的映射，需要提取出三维模型的特征，这样的形态特征可以使得用户识别出这是人体模型。三维模型形态分析一般有两种，一种是基于顶点位移的方法如【】【】【】，一种是基于三角网格变形的方法如【】【】。我们选择三角网格变形作为分析的特征，建立尺寸信息与网格变形的映射 。采用Sumner等人[17]提出的变形迁移算法，建立网格变形与人体模型之间的映射。有了人体特征作为中间量，就成功建立了尺寸信息到人体模型之间的映射。

* 1. 算法概括
  2. 测量与预测尺寸信息
     1. 构建数据集

在预测缺失尺寸信息时，我们的方法需要大量的人体尺寸数据作为“拟合”基础，在合成人体时，预测初值需要训练人体模型，获取人体参数。所以，这就需要大量的数据作为研究基础。目前比较常见的数据集有SCAPE数据集，MPI数据集，MPII数据集和CAESAR数据集。但是这些数据集都存在不同方面的缺陷。SCAPE数据集分辨率较高，并且经过对齐处理，但是只包括同一个人70个姿势和同一个姿势45种不同身材，因此不够完整。MPI数据集包含114个模型，有一个模型具有35种姿势，其余模型有10种随机姿势。MPII数据集有4301个人体。但是分辨率较低，并且数据量也不够大，因此也在实际应用组存在局限性。CEASAR数据集包含2400个北美人体模型，2000个欧洲人体模型，男女皆有，姿势包括坐姿和站姿。虽然该数据集分辨率很高，但是因为其是商业数据集，价格不菲，并且数据并为经过对齐处理。

得益于这些已有的数据集，在生成大量数据集方面，研究者们提出了各种不同的方法。Yang等人[7]使用了一种基于迭代最近点(ICP)的非刚性对齐方法[3]，迭代地将SCAPE的一个模型变形CAESAR数据集中的模型的体型，从而建立一个高分辨率的数据集。为了获取更多的模型数据，Dibra等人[18]合并了Pishchulin等人[14]和Yang等人[7]的数据集，两个数据集虽然都是基于CAESAR数据集生成但是分辨率却不同。Dibra等人[18]也是通过选取一个模板用迭代最近点(ICP) 的非刚性对齐方法，为所有的模型配准。Dibra等人[19]在后续的研究中需要更多的做支持。他们通过改变身材参数的方式生成了100000个模型。文献[4]中，用来描述身材差异的非刚性变形。



其中是通过主成分分析提取的特征向量，为平均身材变形。假设主成分个数取10，数据集个数也为10。为了合成新的网格，可以从PCA空间生成的10维多变量正态分布中采样。对于一个随机样本 ，它服从高斯分布 ，其中为平均变形向量，为协方差矩阵。 通过对随机取样，就可以生成与真实人体相近的模型。为了避免不真实的人体模型，将每一维标准差进行一定限制：。Jiang等人合并了多个已有的数据集，包括SCAPE，SPRING，DYNA[20]，FAUST，MANO等等。但是模型总数也只有11000余个。

* + 1. 尺寸测量

人体测量的尺寸信息能够表征一个人的体型。因此如果要学习测量尺寸与人三维模型之间的关系，我们就需要在计算机中测量的方法要和现实中测量的方法保持一致。实际测量过程中，一般会遇到三种测量方式，在测量身高时需要欧式距离，测量肩宽时用的是测地距离而测量胸围时用圆周距离。圆周可以通过将模型与平面相交来计算，找到包含指定顶点的交点的多边形链，并测量该链的凸包的周长。

我们将数据集的每个模型的测量尺寸看做高维空间中的一个点。所有的点可以拟合出一个流形。基于这样的一个先验常识，我们恢复的尺寸信息要和真实的人体尺寸接近，所以在当前约束下求一个与模板模型尺寸信息最接近的一个点。

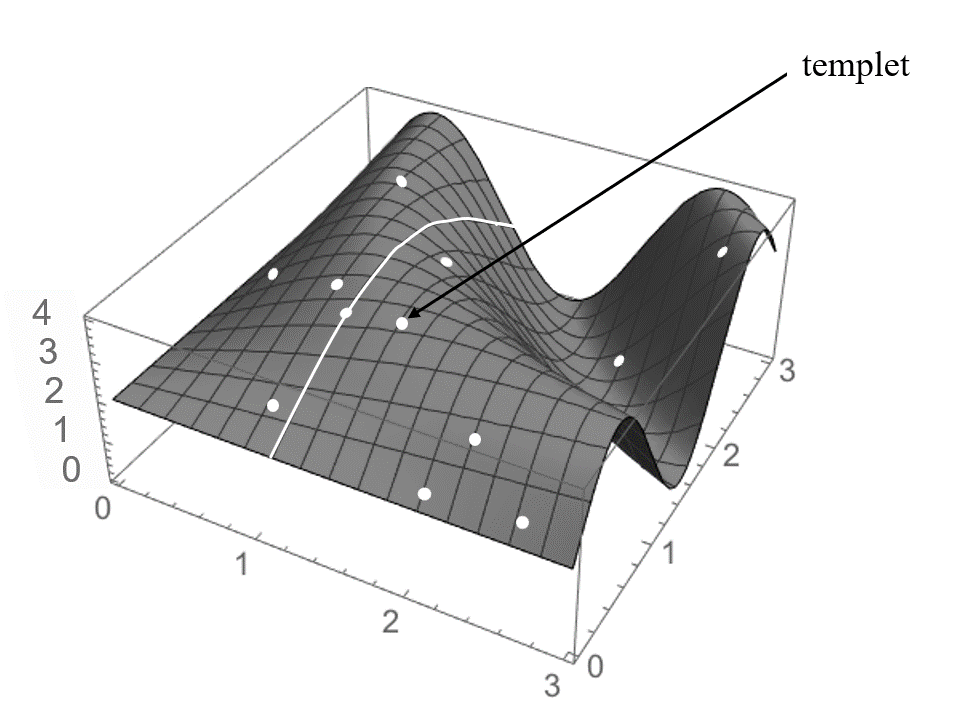


Fig. Schematic diagram of manifold fitting

图 1流形拟合示意图

* 1. 语义学习重建

本算法采用了Sumner等人[17]提出的变形迁移的思想。变形迁移算法的目标是将源网格Source相对于S-Reference的一系列变形迁移到目标网格T-Referecnce上，使得目标网格产生相似的变形，如图 2。应用到本文即找到待生成模型相对于模板模型的变形，即可通过迁移算法生成模型。

将源网格相对于的变形看做是每个三角面片仿射变换的集合。然后三角形只有三个顶点，无法表示垂直于三角面片的变形。于是在垂直于三角面片的方向上加入第四个顶点。记分别为模板模型的顶点和变形后的顶点。由计算可以得出第四个顶点的位置：



定义仿射变换是变形矩阵和平移向量的线性组合，那么变形后的顶点位置可由下式得出：



对的方程都分别减去的方程，可以消去平移向量。因此可以将方程转化为：



将其向量化可得：





在算出数据集中所有物体的变形后，我们将每个人体模型网格中的每个面的变形表示为3×3的变形矩阵：，记所有模型变形的集合为，其中为模型个数，为模型三角面片数。

在计算出所有模型的变形后，就可以建立尺寸信息和变形之间的线性回归，当输入新的尺寸时，就可以预测出对应模型的变形矩阵。但是实际生活中，腰围可能受到胸围的影响，但是小腿围对其影响就会很小。所以每个三角面片的变形并不是受所有尺寸变换的影响，为了解决这个问题，需要为每个三角面片绑定与之相关的个尺寸信息。我们通过递归消除特征法，建立尺寸信息与面片变形之间的线性回归。递归消除特征法使用一个机器学习模型来进行多轮训练，每轮训练结束后，消除若干权值系数对应的特征，再基于新的特征集进行下一轮训练。直到剩下的特征数达到要求。

记个模型的19个尺寸构成的矩阵为:



由于是线性变换的矩阵，所有其行列式值可以反映面片的面积变化情况，某一面片在所有模型中的变形的行列式可以如下表示：



由此建立和的线性回归，每一轮训练中，删去系数最小的对应的特征。然后将剩余的特征重新进行下一轮训练。直到剩余特征为个。当递归消除特征算法完成后，会得到一个：



表示第个面片的第个尺寸在递归消除时被选择，也就是该面片受第个尺寸影响。设为选择后的尺寸参数，可以建立到的线性回归，为方便表示，以一个三角形面片为例，线性回归模型为：



输入所有尺寸信息，以一个三角形面片为例，通过可以选择出该面片最相关的个尺寸信息的矩阵。由回归模型即可求得对应的三角面片变形矩阵：



 当获得所有面片的变形后，通过Sumner[17]提出的方法，可以从一系列变形还原出顶点位置。

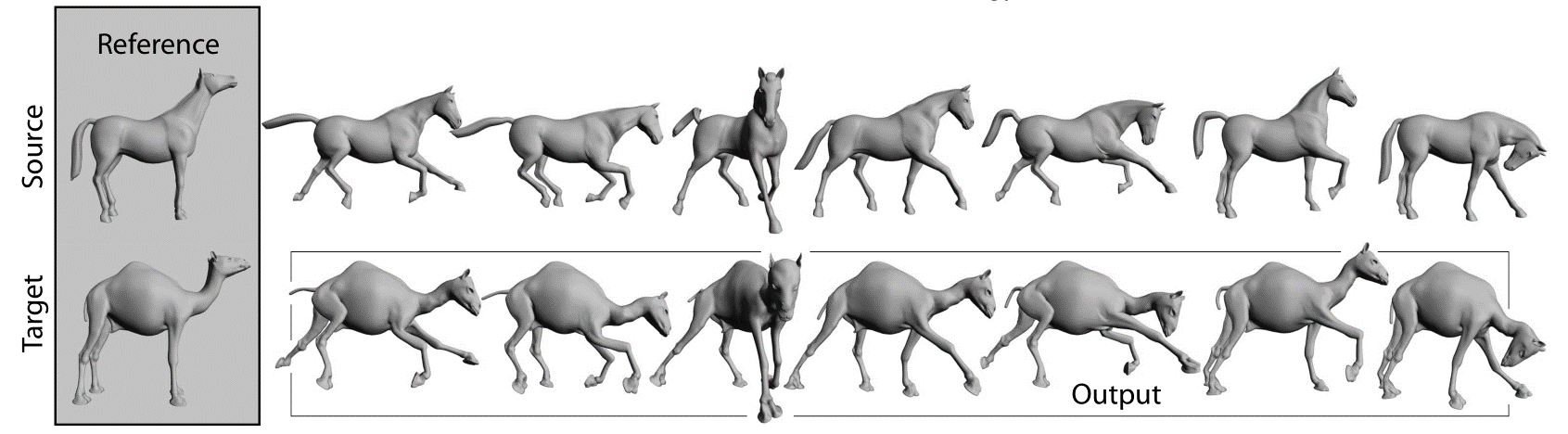


Fig. Deformation transfer copies the deformations exhibited by a source mesh onto a different target mesh.

图 2 变形迁移算法将源网格的变形转移到目标网格上，使得目标网格产生相似的变形。

* 1. 精确重建

以上方法因为采取机器学习的方法，所生成的模型仍然与真实测量结果有较大的误差。本节针对测量尺寸进行精确重建。

首先对于基于欧式距离测量的尺寸，目标长度表示线段两个端点的欧式距离。那么目标网格应该满足以下函数的约束：



其中表示所有基于欧式距离测量的尺寸集合。

对于测地距离，目标长度表示测地路径两个端点的测地距离。我们假定在变形前后，路径上的每个边长相对长度保持不变，有了这个假设，我们就可以得到每个边的目标边长，就可以将问题化成一个优化每条边长的问题。能量函数可以表示为：



其中表示所有基于测地距离测量的尺寸集合，，即当前尺寸的实际测量长度。

最后是基于围长测量的尺寸，围长即过空间一点且法向垂直于地面的平面与网格模型相切所得切面的凸包。和测地距离一样，我们假定在变形前后，围长上每条边相对长度仍然保持不变，即，即当前围长尺寸实际测量长度。能量函数可以表示为：



记所有约束能量为，为了保持原有拓扑和连接，采用拉普拉斯能量项保形：



以欧式距离为例，将能量方程降为2次，设，重写能量函数:



由反三角不等式可得：



将替换不等式右边的:



因此当时，方程左右都有相同的极小值。因此重写能量函数为：



1. 实验结果与分析
   1. 尺寸预测误差对比

桌子上有黑色的咖啡杯

描述已自动生成

Fig. Error comparison of this method with MICE and correlation-based methods

图 3 本文方法与MICE和基于相关性方法的误差比较

* 1. 模型生成误差对比

桌子上有黑色的咖啡杯

描述已自动生成

Fig. Our method compared with models generated by different methods

图 4 本文方法与不同方法生成的模型对比

Table Comparison of reconstruction error between different methods

表 1 不同方法的重建误差比较

|  |  |  |
| --- | --- | --- |
|  |  |  |
|  |  |  |
|  |  |  |

1. 总结

本文处理的模型都是人体这一类物体，未来可以考虑将本文的一些方法推广到通用的动物，从而提高本文方法的适用性。 本文的方法都是在个人计算机上实现，随着高性能计算技术的发展，探索采用 GPU 或者集群系统来提高算法的性能可能也是未来有益的尝试。

**Reference**

[1] Kristensen K, Borum N, Christensen L G, 等. Towards a Next Generation Universally Accessible ‘Online Shopping-for-Apparel’ System[A]. 见: M. Kurosu. Human-Computer Interaction. Users and Contexts of Use[M]. Berlin, Heidelberg: Springer Berlin Heidelberg, 2013, 8006: 418–427.

[2] Keizer A, van Elburg A, Helms R, 等. A Virtual Reality Full Body Illusion Improves Body Image Disturbance in Anorexia Nervosa[J]. A. Stengel. PLOS ONE, 2016, 11(10): e0163921.

[3] Allen B, Curless B, Curless B, 等. The Space of Human Body Shapes: Reconstruction and Parameterization from Range Scans[A]. ACM SIGGRAPH 2003 Papers[C]. New York, NY, USA: ACM, 2003: 587–594.

[4] Anguelov D, Srinivasan P, Koller D, 等. SCAPE: Shape Completion and Animation of People[J]. : 9.

[5] Zhou S, Fu H, Liu L, 等. Parametric reshaping of human bodies in images[J]. ACM Transactions on Graphics, 2010, 29(4): 1.

[6] Wuhrer S, Shu C. Estimating 3D Human Shapes from Measurements[J]. arXiv:1109.1175 [cs], 2012.

[7] Yang Y, Yu Y, Zhou Y, 等. Semantic Parametric Reshaping of Human Body Models[A]. 2014 2nd International Conference on 3D Vision[C]. Tokyo: IEEE, 2014: 41–48.

[8] Zhang Y, Zheng J, Magnenat-Thalmann N. Example-guided anthropometric human body modeling[J]. The Visual Computer, 2015, 31(12): 1615–1631.

[9] Zeng Y, Fu J, Chao H. 3D Human Body Reshaping with Anthropometric Modeling[A]. 见: B. Huet, L. Nie, R. Hong. Internet Multimedia Computing and Service[M]. Singapore: Springer Singapore, 2018, 819: 96–107.

[10] Streuber S. Body Talk: Crowdshaping Realistic 3D Avatars with Words[J]. 2016, 35(4): 15.

[11] Loper M, Mahmood N, Romero J, 等. SMPL: a skinned multi-person linear model[J]. ACM Transactions on Graphics, 2015, 34(6): 1–16.

[12] Robinette K M, Blackwell S, Daanen H, 等. Civilian american and european surface anthropometry resource (caesar), final report. volume 1. summary[R]. SYTRONICS INC DAYTON OH, 2002.

[13] Hasler N, Stoll C, Sunkel M, 等. A Statistical Model of Human Pose and Body Shape[J]. Computer Graphics Forum, 2009, 28(2): 337–346.

[14] Pishchulin L, Wuhrer S, Helten T, 等. Building Statistical Shape Spaces for 3D Human Modeling[J]. Pattern Recognition, 2017, 67: 276–286.

[15] Lassner C, Romero J, Kiefel M, 等. Unite the People: Closing the Loop Between 3D and 2D Human Representations[A]. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)[C]. Honolulu, HI: IEEE, 2017: 4704–4713.

[16] Seo H, Cordier F, Magnenat-Thalmann N. Synthesizing Animatable Body Models with Parameter- ized Shape Modifications[J]. : 6.

[17] Sumner R W, Popovic J. Deformation Transfer for Triangle Meshes[J]. : 7.

[18] Dibra E, Jain H, Oztireli C, 等. HS-Nets: Estimating Human Body Shape from Silhouettes with Convolutional Neural Networks[A]. 2016 Fourth International Conference on 3D Vision (3DV)[C]. Stanford, CA, USA: IEEE, 2016: 108–117.

[19] Dibra E, Öztireli C, Ziegler R, 等. Shape from Selfies: Human Body Shape Estimation Using CCA Regression Forests[A]. 见: B. Leibe, J. Matas, N. Sebe, 等. Computer Vision – ECCV 2016[M]. Cham: Springer International Publishing, 2016, 9908: 88–104.

[20] Pons-Moll G, Romero J, Mahmood N, 等. Dyna: a model of dynamic human shape in motion[J]. ACM Transactions on Graphics, 2015, 34(4): 120:1-120:14.