

# 浙江大学

## 三维动画课程读书报告



题目 深度相机之服装建模

作者姓名 孙立钢

作者学号 21851026

指导教师 李启雷

学科专业 软件工程

所在学院 软件学院

提交日期 2018 年 12 月 25 日

# Garment Modeling with a Depth Camera

A Dissertation Submitted to

Zhejiang University

in partial fulfillment of the requirements for

the degree of

Master of Engineering

Major Subject: Software Engineering

Advisor: Qilei Li

By

Sun Ligang

Zhejiang University, P.R. China

2018

## 摘要

以往的服装造型技术主要是设计新颖的服装来装扮虚拟人物。我们研究了真正的服装的模型，并开发了一个系统，是直观的使用，即使是新手用户。我们的系统包括服装组件检测器和从人工标注的服装图像数据库中学习的属性分类器。在建模过程中，我们使用 Kinect 对服装进行扫描，根据原始 RGBD 序列，通过 KinectFusion 构建出一个大致的形状。检测器和分类器将从 RGB 图像中识别服装部件(如领子、袖子、口袋、腰带和纽扣)及其设计属性(如荷叶边领子或翻领、泡泡袖或直袖)。我们的系统还包含一个服装组件的 3D 变形模板数据库。一旦确定了组件和它们的设计，我们就会选择合适的模板，将它们缝在一起，并将它们与 KinectFusion 生成的初始服装网格相匹配。对不同款式服装进行实验，均得到了高质量的结果。

**关键字：**服装建模、深度相机、服装解析、3D 模板、语义建模

## Abstract

Previous garment modeling techniques mainly focus on designing novel garments to dress up virtual characters. We study the modeling of real garments and develop a system that is intuitive to use even for novice users. Our system includes garment component detectors and design attribute classifiers learned from a manually labeled garment image database. In the modeling time, we scan the garment with a Kinect and build a rough shape by KinectFusion from the raw RGBD sequence. The detectors and classifiers will identify garment components (e.g. collar, sleeve, pockets, belt, and buttons) and their design attributes (e.g. falbala collar or lapel collar, bubble sleeve or straight sleeve) from the RGB images. Our system also contains a 3D deformable template database for garment components. Once the components and their designs are determined, we choose appropriate templates, stitch them together, and fit them to the initial garment mesh generated by KinectFusion. Experiments on various different garment styles consistently generate high quality results.

**Keywords:** GarmentModeling,DepthCamera,GarmentParsing, 3D Templates,  
Semantic Mo-deling

# 1 深度体感相机的简介

深度相机，即提供一种能提供三维交互信息的计算机视觉设备。三维交互技术的发展促使了一批新型的低成本的体感深度相机设备。如 2010 年微软发布的基于结构光技术的 Kinect Camera。其中 Kinect 是微软开发的体感游戏 XBOX360 的相机设备，它主要包括三个摄像头，其中 IR 发射/接收摄像头主要是用来获取场景中的深度图像，得到深度信息的来源，初次之外还有获取 RGB 图像的摄像头组成，IR 深度摄像头通过结构光技术，将反射回的红外场景图编码成深度信息，以便获取三维空间的信息。

Kinect 支持的主要功能有：获取深度图像源数据和内部麦克风阵列，主要用于数据流交互；骨骼定位跟踪，能实时捕捉镜头中人体三维骨骼信息，方便人机体感交互应用；音频传输，主要支持音频处理。由于以上多种功能的支持，使 Kinect 深度体感相机成为计算机人机交互领域的一个热点。

## 1 深度相机进行服装建模简介

对很多应用程序来说，现实服装的模型是必要的。大多数现有的服装建模工作，专注于开发三维服装模型来装扮虚拟人物。只有很少一部分的工作研究了现有真实服装的建模问题，这个研究已经有了很重要的应用，例如：虚拟试穿或原型，Berthouzoz 等于 2013 年提出可以自动将现有的缝纫模式转换成 3D 服装模型。这个方法会产生复杂的模型，因为服装的缝纫模式并不总是可用的。Zhou 等于 2013 年同时提出，通过服装的单个图像来为服装完成建模工作。这种方法简单易用，但是只能捕获有限的细节。

北京航空航天大学的陈小武教授提出一个系统，致力于为真实服装实现简单而直观的高质量建模。具体方法只需要一个像微软 Kinect 照相机那样的客户级的深度相机。对于那些没有 3D 建模经验的用户来说，这个设置成本低廉并且容易使用。使用这个系统，即使是一个普通用户，也可以快速产生服装的 3D 模型。如图 1 中的实例所示。



图 1

这个方案关键的观察点是：大部分的服装都是由标准的组件组成的，例如：衣袖、衣领、纽扣等，而且大部分的组件都是有设计标准的。因此，陈教授他们采用“基于组装建模”的原则来实现快速直观的服装建模。之前，曾经有人使用相似的原则来实现室内场景的建模工作。但是服装具有自己的特点，不能直接使用室内场景建模的方式来实现服装的建模。例如：服装上面的纽扣的口袋通常都很小，很难从通过 3D 扫描区分出来，襟领和定制领的设计属性同样也很难通过 3D 扫描区分出来。所以，这个方案包括一个复杂的图像分析组件，通过这个组件，可以探测出服装的组件和设计属性，这点在早期的作品中是没有使用的。

陈教授和他的团队在 Ingham 和 Covey 首先调查了服装设计规则，这让他们可以通过解析树来表示服装的语义结构，如图 2(c)所示。它通过一组预定的组件来表示服装，如衣领、衣袖、口袋等。每一个组件还可以通过一些设计属性，如尺寸、形状、类型等来标识。每个属性都有一些预定值，例如，对于衣领来说，有襟领、立领、裁缝领等。解析树提供了强有力的结构，为服装建模做好前期工作。

在建模的过程中，系统可以捕捉穿在基于 Kinect 的站立姿势模型上的目标服装。系统程序会基于手动标记的服装图像数据库，通过服装组件探测器和设计属性分类器来自动解析 RGB 图像的前视图。这些单独分析出来的服装组件，再通过贝叶斯网络利用组件语义的兼容性，进一步的联合完善。例如：它会知道泡泡袖经常和荷叶边衣领搭配。如果需要的话（这是罕见的情况），用户可以与系统交互改变检测和分类的结果。

在另一方面，系统有一个专家为服装组件设计的 3D 变形模板数据库。一旦服装的组件和设计属性被确定下来了，我们可以选择具有相同设计的最好模板，

然后将这些模板拼接起来，形成最终的服装模型。系统通过 KinectFusion 构建一个初始的 3D 服装网格，然后将选择好的服装组件模板定位到网格上，以此来指导服装组件模板的选择。由于受模板数据库大小的限制，得到的结果模型并不能完全匹配原来服装的三维扫描。相反，系统目标是创建一个具有相同设计的三维服装模型。

陈教授认为，在这项工作中他们做出了如下贡献。1) 他们提出了一种总结服装建模规则的服装解析树。连同他们服装组件的 3D 模型，形成了一个紧凑的基于规则的服装建模系统。用这个系统，用户通过选择服装组件和它们的设计属性，可以快速构造出一个高质量的服装模型。2) 他们建立了一个具有 6000 张服装图片的标记图片数据库，每一张图片都通过骨架构成、服装组件和设计属性进行了标记。这个数据库和标签工具应该促进未来的研究沿着服装建模、解析或者服装图片搜索的方向来进行。3) 他们设计了服装组件探测器和组件的设计属性分类器，然后在贝叶斯网格中分析整体的结果。解析结果进一步与 3D 模板结合，构建出一个端对端的服装建模系统。

## 2 相关工作

许多服装建模的作品关注的应用是在电影和游戏中装扮虚拟的人物。还有服装目标重定位技术，可以实现服装设计从一个人物形象到另一个体型和姿势具有很大改变的人物形象的转变。基于模式的方法，模仿现实生活中设计和裁剪衣服的过程来创建服装模型，包括 2D 模式创建，基于实物模拟和涉及参数的迭代优化，可以实现高质量的效果。Umetani 等于 2011 年提出了一个交互式的工具：敏感的服装，这个工具提供了服装建模以及模式设计和 3D 褶皱功能。此外，还有一些其他著名的服装建模工具，例如：Marvelous Designer 和 vStitcher and Optitex PDS。这些系统大部分都是劳动密集型的，并且需要很强的专业知识才能使用。

相比较而言，基于草图的放法更加直观，这个方法允许用户直接在人形模具进行绘制，来创建服装模型。Robson 等人进一步整合以前的理论和知识，以提升对宽松衣服的建模能力。虽然效果不错，但是基于草图的方法仍然要求很强的服装设计专业知识，对于非专业人士来说是很艰难的。

只有少部分作品研究真实服装的数字化问题。因为真实的服装通常是以缝纫模式设计的，Berthouzoz 等于 2013 年提出了将现有的缝纫模式转换成 3D 服装

模型。该方法会产生的复杂的腹胀模型，并且服装的缝纫模式并不总是很简单就可以进行分析的。2013 年有人提出了从一张单一的图像创建服装的 3D 模型，这个方法很简单且易于实现，但是只能捕获有限的细节。陈教授他们提出，用深度相机进行 3D 建模，他们的方法非常直观，为新手用户创建高质量的服装提供了可能性。此外，许多方法提出了从视频中捕捉服装动画。陈教授他们的方法专注于服装建模，将捕捉动画的问题留给了未来的研究。

随着公众 3D 模型集合的成长以及客户级深度相机的流行，人们提出了很多数据驱动的建模工具。这些方法通过结合现存的模型来数字化现实世界的场景和对象。Fisher 等于 2012 年利用从 3D 场景数据库产生的概率模型来合成新的场景。Chaudhuri and Koltun 于 2010 年构建了一个贝叶斯网格，根据中间形状提出模型组件，以此来辅助交互建模和形状合成。顺着这条主线，Shao 等于 2012 年提出了使用客户级的 RGBD 相机进行室内场景语义建模的交互式方法。从数据库中为每一个对象获取一个最好的 3D 模板，来重构最后的场景。Shen 等于 2012 创建一个结构化的网格模型，这个模型的创建源于具有网格仓库的 Kinect 相机的单一扫描。Chen 等于 2014 年提出了一种室内场景的自动语义建模方法，这里的室内场景来自于使用上下文关系从 3D 室内场景数据库获取的低质量的数据。然而，这些方法中的大部分都是为室内场景或人造物体设计的，例如椅子或自行车。这些方法并不能用于陈教授的服装建模问题。例如，Shao 等于 2012 年和 Chen 等于 2014 年提出的技术，很难用来分割我们的个人服装组件。

最近，在计算机视觉领域，有人提出了服装图像解析方法，它们可以从一个或多个图像中识别简单的服装设计和人物造型，即使视角发生改变，造型多样化，仍然可以做到。直接使用此方法不能解决陈教授他们的问题，因为这个方法会导致噪音效果。陈教授他们专注于运用已知的造型和服装片段来进行服装解析，这样会产生更加精确的探测和分类结果，可以简化建模。

### 3 概述

陈教授他们致力于数字化真正的服装，来产生高质量的服装建模，给新手用于一个直观的体会。具体来说，他们专注于创建一个具有相同设计属性和尺寸的合理的 3D 服装模型。据观察，大部分衣服都是由标准的组件组成的，而且大部分组件都具有标准的设计，陈教授他们调查了服装设计规则，然后拟定了一个紧



凑的服装解析层次，如图 2 所示。与解析树相对应，他们完整的系统包括一个后端和一个前端。

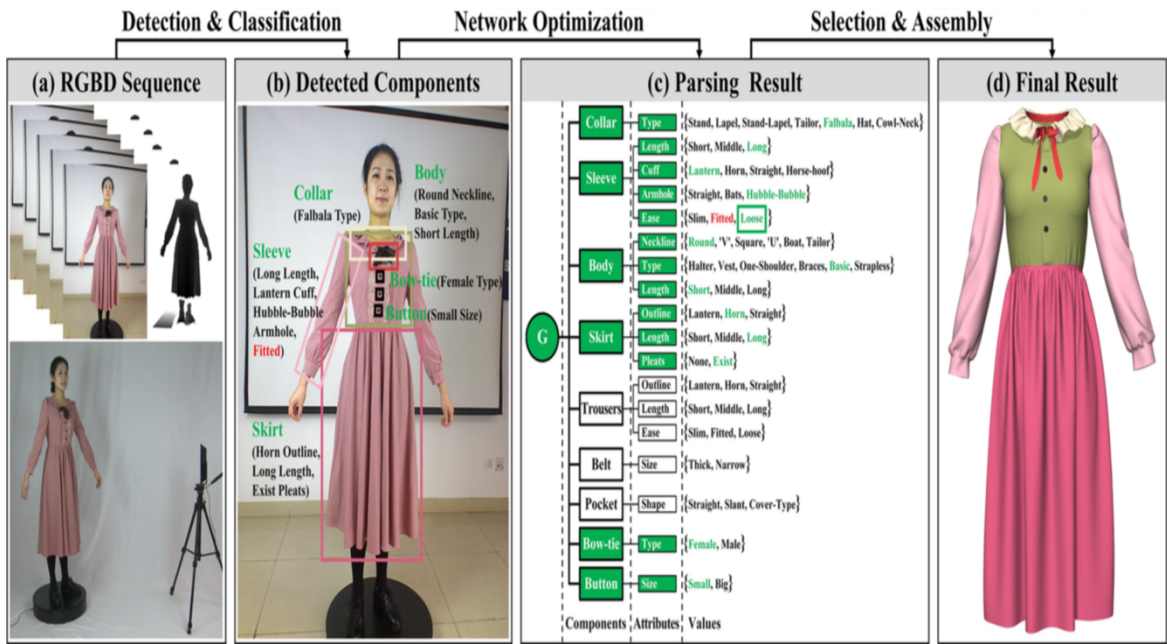


图 2

后端包括两个服装数据库。第一个数据库里面的存储的数据，是由专家设计的 3D 服装组件模板，这个数据库一共包含 1000 多张不同的模板，如图 3 所示。

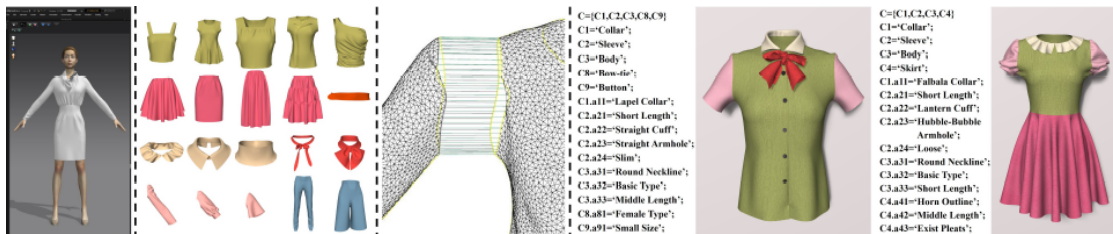


图 3

这些模板是在一个标准的 3D 人体模型上设计的，这个人体模型可以改变其骨架，形成不同的姿势，与扫描得到的模型相匹配。每个模板都使用和组件的名称和各种属性进行标记，并且使用“插槽接缝”来连接相邻的组件。结合服装解析树一起使用，陈教授他们构建了一个有效的基于规则的服装建模系统。用户通过选择服装组件和组件的设计属性，可以快速建立一个高质量的服装模型。第二个数据库是手动标记的图像数据库，这个数据库中包含了 6000 张服装的图片。每一张图片都用骨架造型、服装片段和服装组件进行了标记，其中服装组件包括它们的设计属性和边界。他们分别使用了使用了 DPM 方法和随机森林算法，基于被标记的 2D 图像来训练服装组件探测器和设计属性分类器。陈教授他们又进一步研

究了贝叶斯网络，捕捉服装组件之间的语义兼容性。

前端的总结如图 2(a)所示，陈教授他们使用一个固定的 Kinect 照相机来扫描穿着目标服装的模型。目标站在一个旋转的舞台上，旋转大概一分钟一次。这个设置成本低，新手用户可以直观的使用。作为预处理，陈教授他们首先使用 KinectFusion 系统，从原始的 RGBD 序列中构建了一个初始化的网格。然后他们选择一张参考图片，要求图片中模型的脸是向前的。在这个参考帧中，使用一个交互式工具对服装进行分割，模型的骨架姿势也被交互式的标记了。分段服装被用于减少初始网格交互，获取一个初始化的服装网格。陈教授他们将标记的 2D 骨架反投影到初始化的服装网格上，进一步转化成 3D 模型，包括模型的 3D 骨架。他们就爱那个组件探测器和属性分类器应用到参考图片上，结果如图 2(b)所示。贝叶斯网络被用于优化在一起使用的单个确定的服装组件，如图 2(c)所示。根据服装解析的结果，陈教授他们设计的系统可以选择最精确的组件模板，然后自动将这些组件模板组合起来，构成一个 3D 模型，如图 2(d)所示。

陈教授他们的系统包含几个组件，这些组件中的大部分在 3D 建模作品作品中都是标准和普遍的，例如：Kinect 扫描，融合、分割背景。除此之外，他们的主要包括服装分析器（检测器、分类器和贝叶斯网络）和服装合成器（变形和拼接模板）。虽然单个图像（或视频）更容易采集数据，但很难得到高品质的结果，因为缺少 3D 信息。

## 4 总结

陈教授他们为服装建模提出了一个直观的系统，使用这个系统，只需要一个深度相机。根据目标服装的 RGBD 序列，通过从专家设计的 3D 模板数据库中选择精确的组件模板，再将它们组合起来，系统就可以产生一个高质量的服装模板。陈教授他们通过具有大量手动标记的服装图片的数据库，对系统的组件探测器和属性分类器进行训练，使它们可以更加精确。这些探测器和分类器建立在强大的服装结构知识的基础上，来实现良好的表现。然后在贝叶斯网络中对选择的组件模板结果进行分析，以此来指导模板的选择。使用陈教授他们的系统，一个普通的用户可以快速产生一个详细的 3D 服装模型。当然，他们的系统也是有缺陷的，但这些缺陷也指出了这个系统的发展方向。首先，这个系统不会产生目标服装的精确扫描。相反，它输出一个貌似真实的具有相似语义结构和形状的服装模型。

他们的系统假设知道服装组件的清晰结构，而且他们不能处理多层的衣服，例如T恤在背心外面，因为太多的遮挡，而不好处理。第二，他们当期工作的重点是对衣服的几何形状进行建模，而不是捕捉布的纹理模式。第三，他们的工作不捕捉布的变形和动画参数。

总的来说，虽然陈教授他们的系统仍有不足之处，但他们取得的成就也是了不起的，希望他们可以克服不足之处，让系统更加完善。