



北京航空航天大学
BEIHANG UNIVERSITY

北京航空航天大学计算机学院2015级本科生毕业设计开题答辩

三维场景人体与对象交互的 接触点标注工具

学 生：李奕君

学 号：15061199

指导教师：周 彬

虚拟现实技术与系统国家重点实验室

State Key Laboratory of Virtual Reality Technology and Systems

汇报提纲

一、论文选题背景和意义

二、研究现状及问题分析

三、工作目标与工作内容

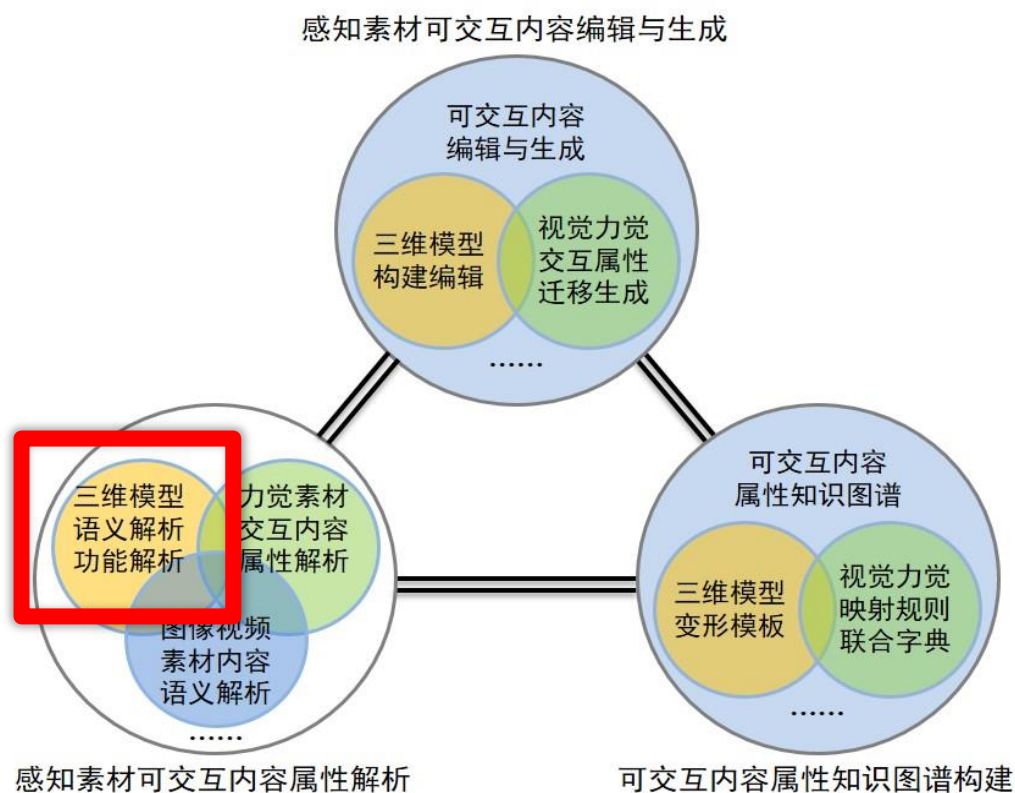
四、技术路线及工作难点

五、工作计划和参考文献

一、论文选题背景和意义 (1/2)

● 论文选题背景

— 国家自然科学基金重点项目：“感知素材可交互内容编辑与生成理论及方法”



一、论文选题背景和意义 (2/2)

● 论文选题意义

- 交互性 (affordance)：揭示了对象所具有的功能和人可能对其执行的操作。



作用：

- 帮助计算机识别和理解图像和视频中的人类活动，理解人类的世界
- 涉及知识：

- 语义关系、接触点、物体姿态和人类姿态等知识。

体现形式：

- 人与物体接触点的heat map
- 机器人从人与物体交互视频中学习object affordance和social affordance

现状：2D的affordance研究多，3D的affordance研究少

汇报提纲

一、论文选题背景和意义

二、研究现状及问题分析

三、工作目标与工作内容

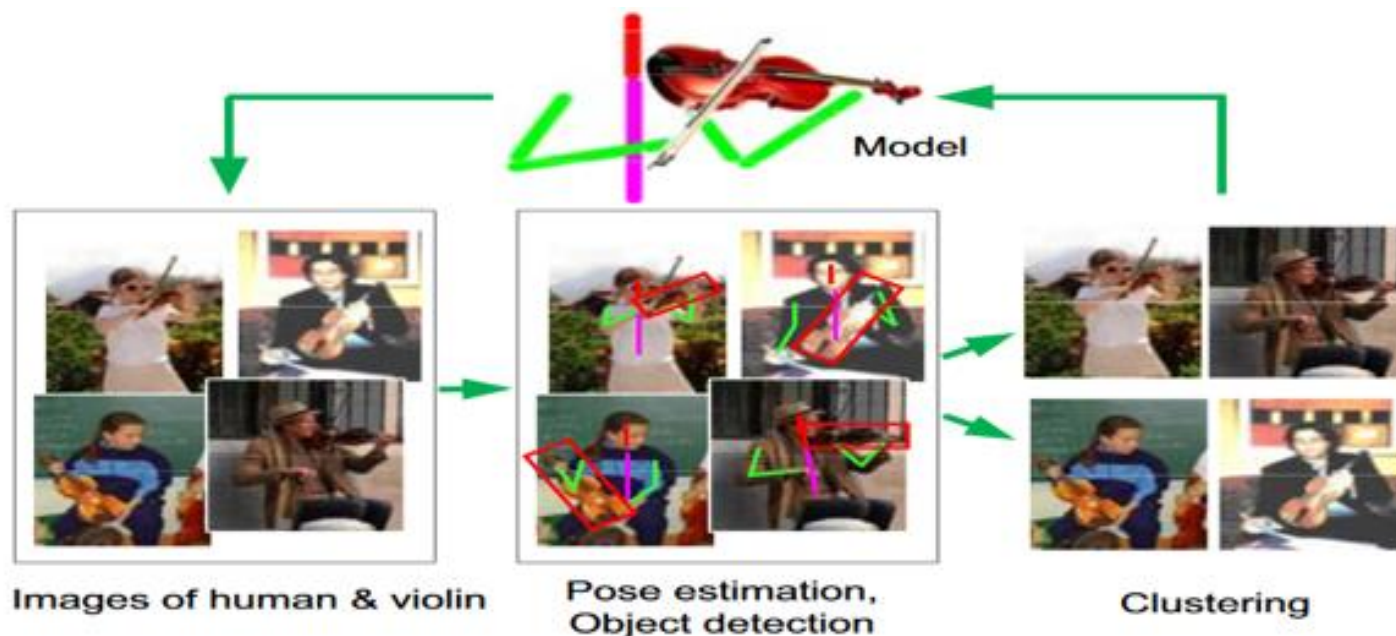
四、技术路线及工作难点

五、工作计划和参考文献

二、研究现状及问题分析 (1/7)

● 2D affordance

本文提出了一种弱监督的方法来发现对象可能的功能，每个功能都是由特定的人-对象交互来表示的。可以根据人体姿势和接触点 预测乐器种类。

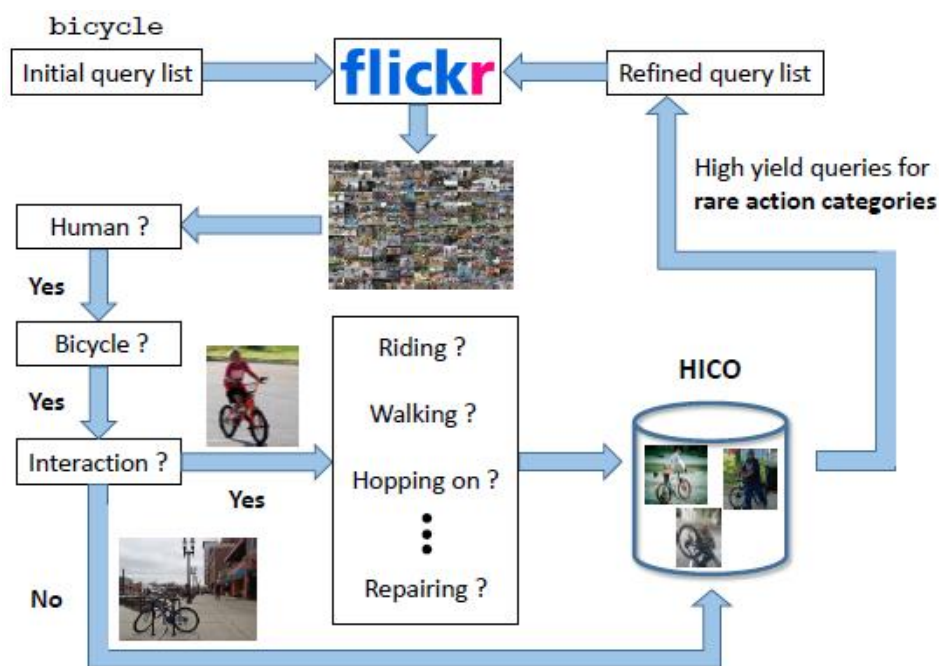


Discovering object functionality . ICCV 2013

二、研究现状及问题分析 (2/7)

● 2D affordance

文章引入“人与普通对象交互 (HICO)”这一新的标准，用来识别人与对象的交互，每个图像中的物体与人的交互标签尽可能的多样化，并展示了DNN在当前识别中的显著优势。



HICO: A Benchmark for Recognizing Human-Object Interactions in Images, ICCV2015

二、研究现状及问题分析 (3/7)

● 2D affordance

引入挖掘语义可供性知识的新问题：给出一个动作和一个物体，确定该动作是否可以应用于物体。

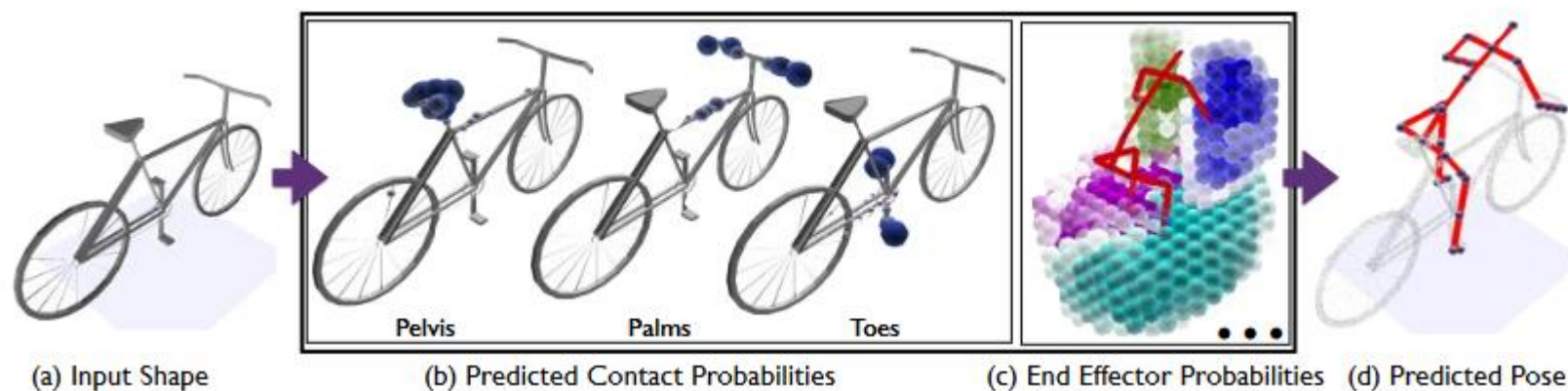
	hurt	accelerate	interview	race	hunt	feed	operate	manufacture	drive	ride	board	...
bicycle												
bird												
boat												
bottle												
car												
cat												
cow												
dining table												
horse												
person												
train												
tv												

Mining Semantic Affordances of Visual Object Categories, cvpr2015

二、研究现状及问题分析 (4/7)

● 3D affordance

引入了一种新的形状分析模型，通过运动学分析预先计算身体与物体接触点的概率分布，预测人体的姿态参数和接触点。

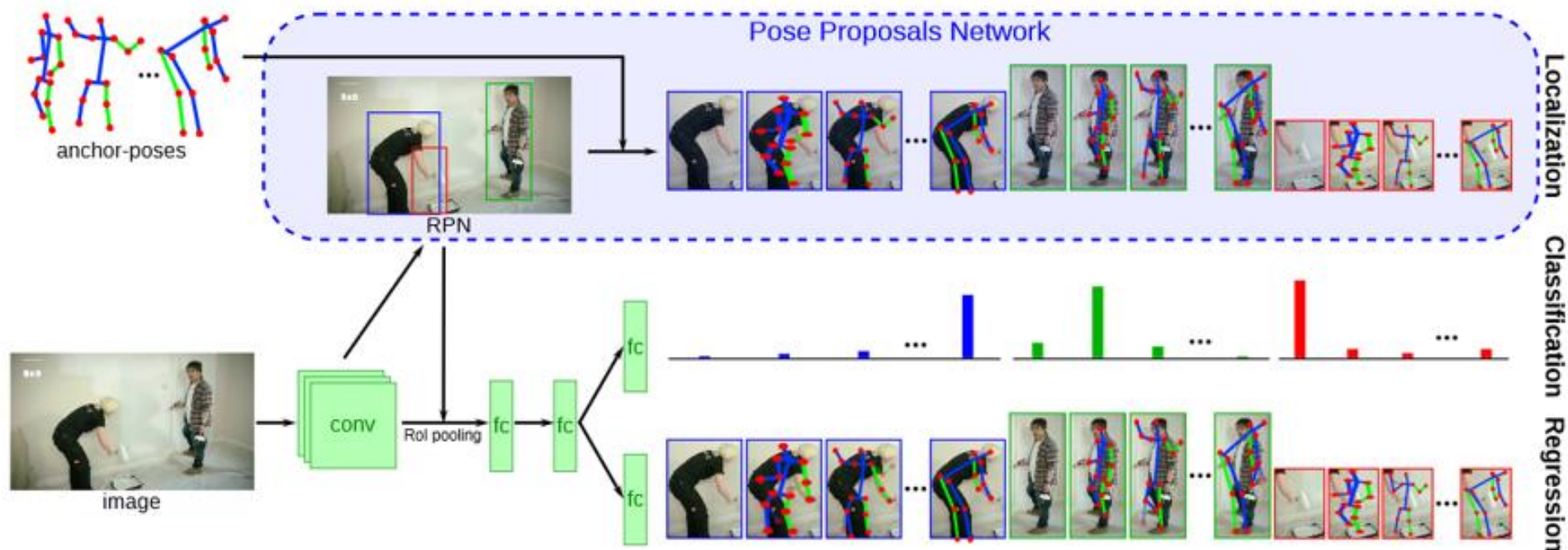


Shape2Pose: Human-Centric Shape Analysis.,siggraph 2012

二、研究现状及问题分析 (5/7)

● 对图像中的人和物体进行建模

2017年, **Gregory**等人提出了LCR-NET架构(由姿势proposal生成器, 分类器及回归器构成), 可以在一张图片中端到端的进行3D的人体姿势估计。

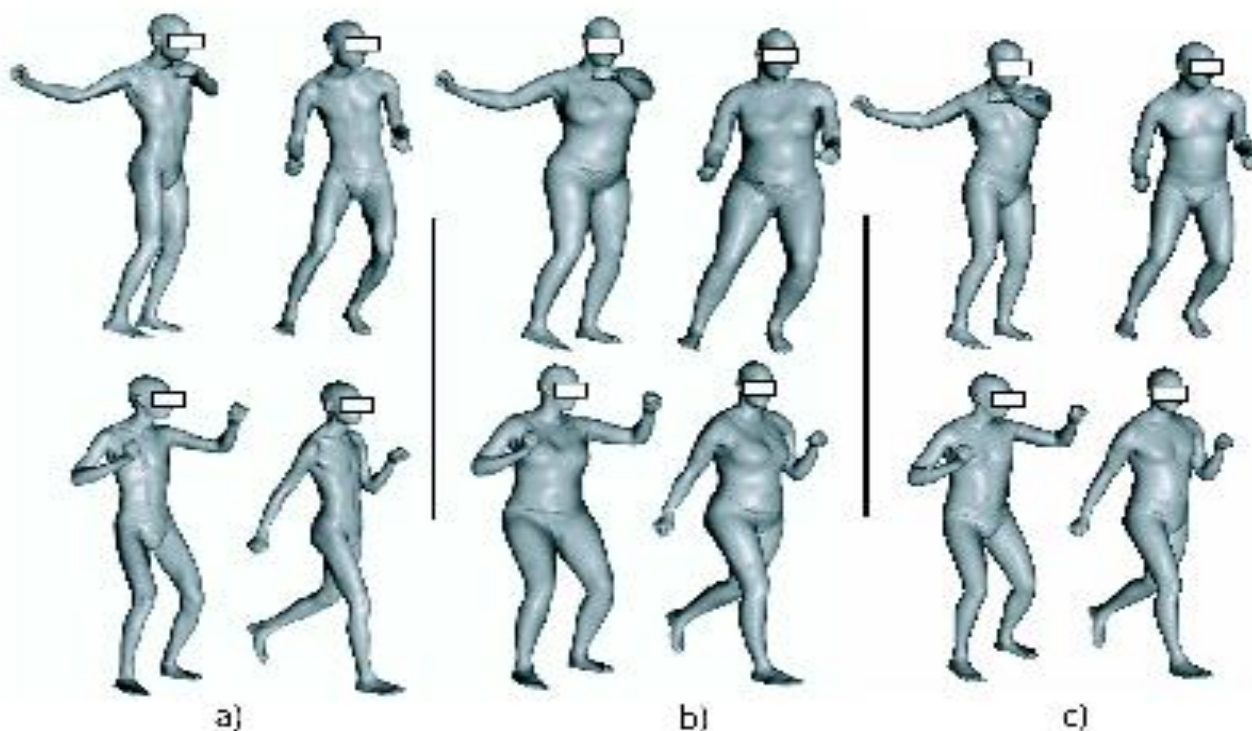


LCR-Net: Localization-Classification-Regression for Human Pose, cvpr2017

二、研究现状及问题分析 (6/7)

● 对图像中的人和物体进行建模

介绍了SCAPE方法,一种数据驱动的方法,用于构建跨越主体形状和姿势变化的人体形状模型。

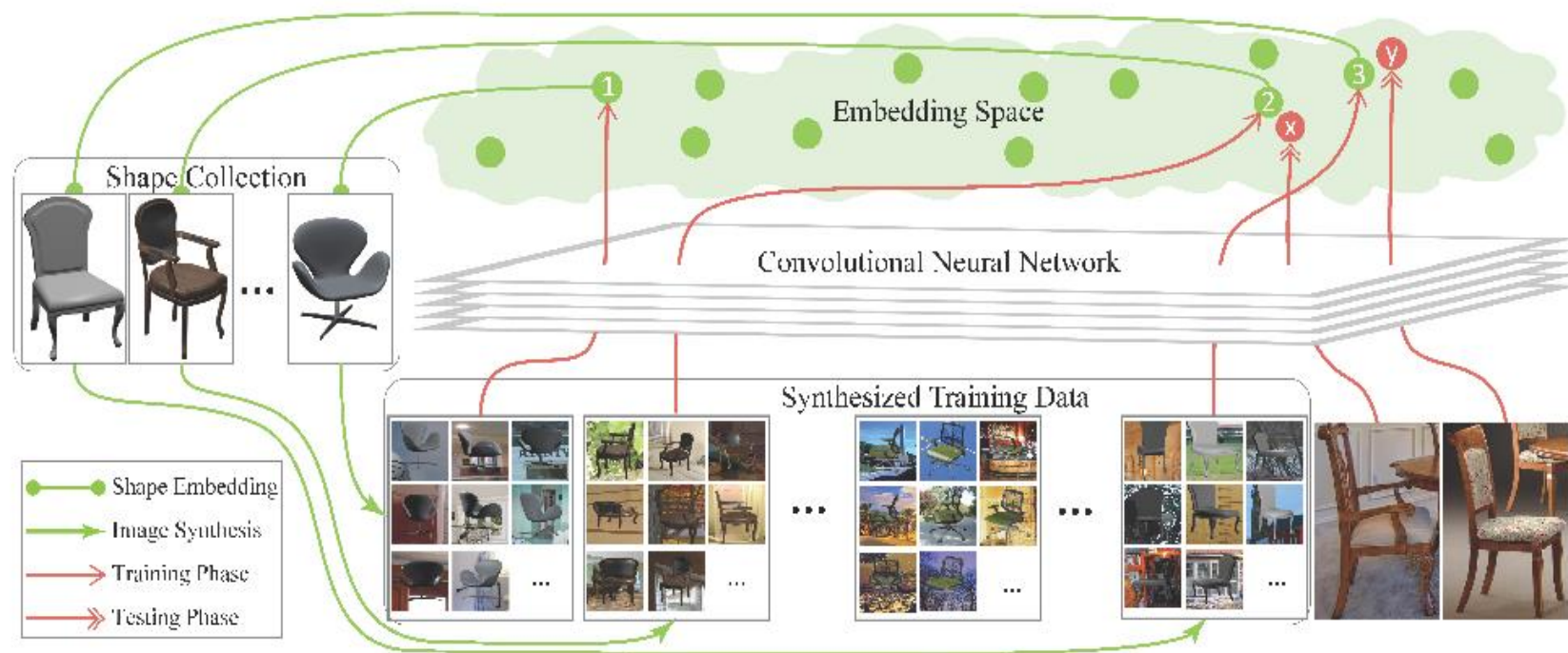


SCAPE: Shape Completion and Animation of People, SIGGRAPH, 2005

二、研究现状及问题分析 (7/7)

● 对图像中的人和物体进行建模

文章提出了一个关于3D形状的联合嵌入空间的，其中嵌入实体之间的距离反映了基础对象之间的相似性。



Joint Embeddings of Shapes and Images via CNN Image Purification, Acm Transactions on Graphics, 2015

二、研究现状及问题分析

● 问题分析

- 如何在三维空间体现affordance?
 - 如何研究人和物体的三维接触点?
 - 如何为研究affordance提供数据?
-
- 可以利用大量图像资源研究affordance

汇报提纲

- 一、论文选题背景和意义
- 二、研究现状及问题分析
- 三、工作目标与工作内容
- 四、技术路线及工作难点
- 五、工作计划和参考文献

工作目标与工作内容（1/4）

● 工作目标

- 设计并实现一个标注工具，输入一个人体与对象交互的三维场景（或图像场景），能够计算和标注出人体与对象的接触点（二维接触点、三维接触点）。

● 工作内容

- 工作内容一：人体与对象三维交互场景接触点自动粗略解析；
- 工作内容二：人体与对象三维交互场景接触点手动标注编辑；
- 工作内容三：基于单幅图像的人体与对象交互接触点标注。

三、研究目标与研究内容（2/4）

- 研究内容一：人体与对象三维交互场景接触点自动粗略解析
 - 输入一个人和物体交互的三维场景文件
 - 通过特定算法计算人和物体接触点
 - 输出人和物体的接触点信息（用球、长方体等形状表示）



输入三维场景

通过特定算法自动
标注接触点



输出接触点信息

三、研究目标与研究内容（3/4）

- 研究内容二：人体与对象三维交互场景接触点手动标注编辑
 - 使用者能够在粗略求得的接触点上进行手工调整
 - 调整接触点位置、大小、形状、数量等信息



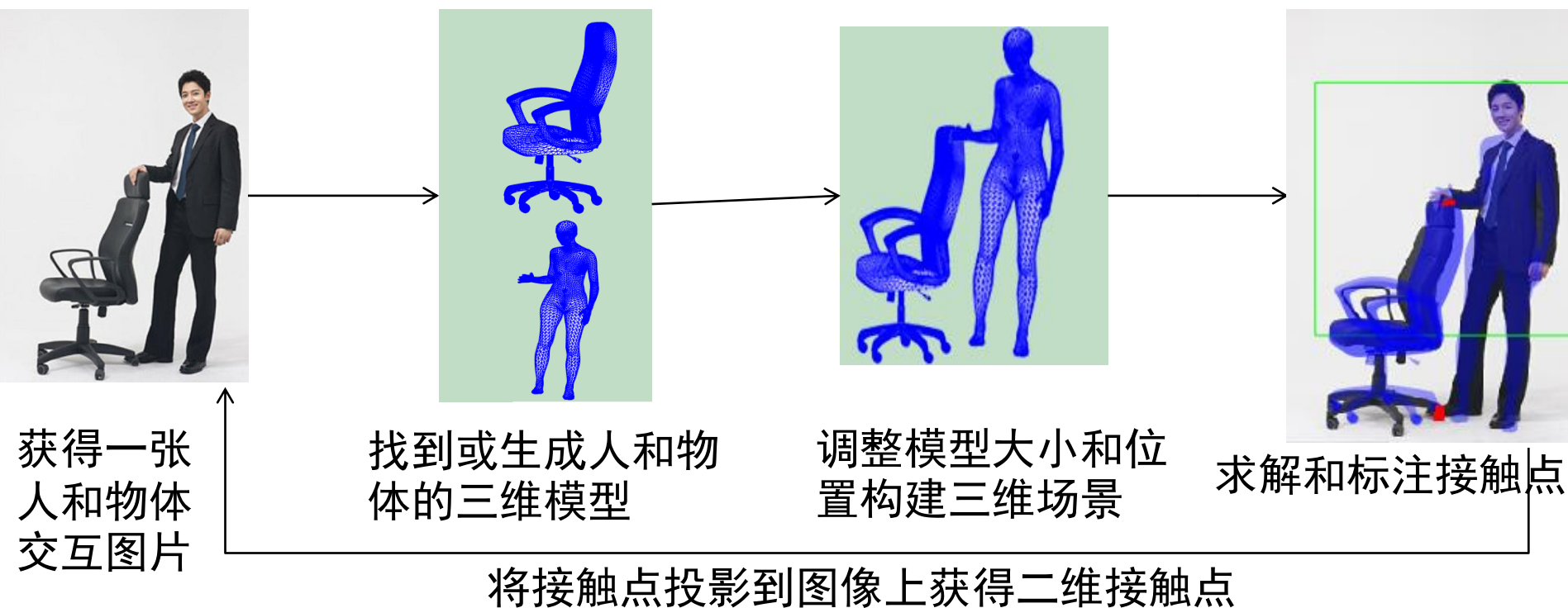
在自动求解接触点的基础上手动调整接触点



三、研究目标与研究内容（4/4）

● 研究内容三：基于单幅图像的人体与对象交互接触点标注

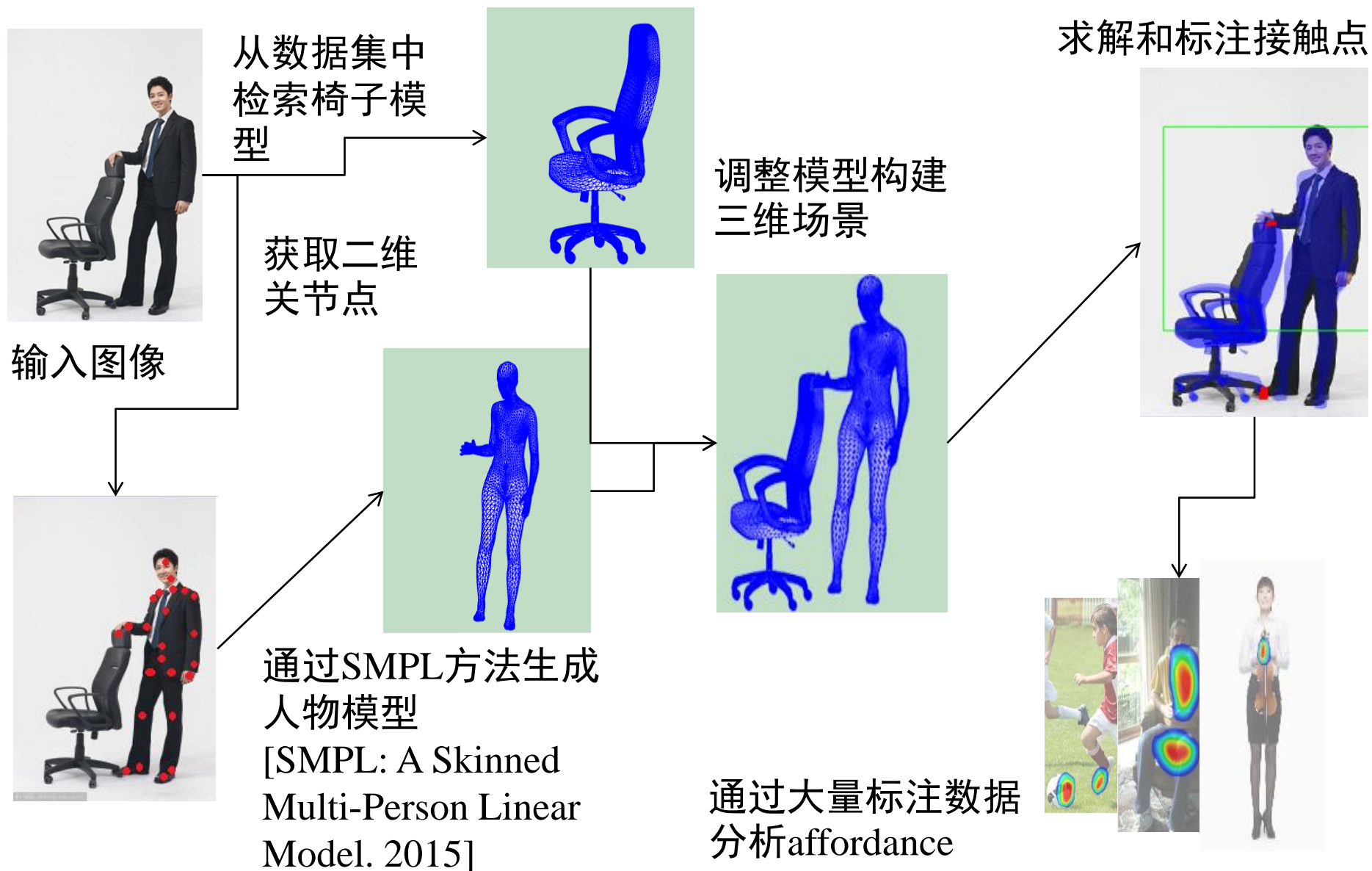
- 根据单幅图像，交互式构建出人体与对象三维交互场景
- 集成接触点自动解析、接触点手动标注编辑等功能
- 实现三维接触点到输入图像的投影，得到二维接触点



汇报提纲

- 一、论文选题背景和意义
- 二、研究现状及问题分析
- 三、工作目标与工作内容
- 四、技术路线及工作难点
- 五、工作计划和参考文献

四、技术路线及研究难点 (1/2)



四、技术路线及研究难点（2/2）

● 研究难点

- 如何比较好地定义接触点（二维、三维）？
- 如何从3D的交互场景，估计3D的人与物体的接触点？
- 如何从单张图片，生成3D的交互场景？

汇报提纲

- 一、论文选题背景和意义
- 二、研究现状及问题分析
- 三、工作目标与工作内容
- 四、技术路线及工作难点
- 五、工作计划和参考文献

五、研究计划和参考文献（1/2）

● 工作计划

时间节点	研究计划
2018.11.01~2018.11.10	制定进度表，规划未来工作。
2018.11.11~2018.11.20	查阅收集相关资料。
2018.11.21~2018.12.21	实现自动求解三维接触点功能。
2018.12.22~2019.01.15	实现手工标注三维接触点功能。
2019.01.16~2019.03.31	实现物体模型检索和人物模型生成功能。
2019.04.01~2019.04.30	对标注工具进行测试和完善。
2019.05.01~2019.06	准备毕设结题论文

五、研究计划和参考文献 (2/2)

● 主要参考文献

1. Yu Xiang, Wonhui Kim, Wei Chen, Jingwei Ji, Christopher Choy, Hao Su, Roozbeh Mottaghi, Leonidas Guibas and Silvio Savarese. ObjectNet3D: A Large Scale Database for 3D Object Recognition. 2016
2. Marszalek M, Schmid C. Semantic Hierarchies for Visual Object Recognition. 2007
3. Fellbaum C, Miller G. WordNet:An Electronic Lexical Database. 1998
4. Chao Y W, Wang Z, He Y, et al. HICO: A Benchmark for Recognizing Human-Object Interactions in Images. 2015
5. Yang Y, Ramanan D. Articulated pose estimation with flexible mixtures-of-parts. 2011
6. Carreira J, Agrawal P, Fragkiadaki K, et al. Human Pose Estimation with Iterative Error Feedback. 2016
7. Camillo Jose Taylor. Reconstruction of Articulated Objects from Point Correspondences in a Single Image. 2000
8. Guan P, Weiss A, Bălan A O, et al. Estimating Human Shape and Pose from a Single Image. 2009
9. Li Y, Su H, Qi C R, et al. Joint embeddings of shapes and images via CNN image purification. 2015
10. Ionescu C, Papava D, Olaru V, et al. Human3.6M[DB/OL]. 2014
11. Girshick R. Fast r-cnn. 2015
12. Girshick R, Donahue J, Darrell T, et al. Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation. 2014
13. Deng J, Dong W, Socher R, et al. ImageNet: A large-scale hierarchical image database. 2009
14. Valmadre J, Lucey S. Deterministic 3D Human Pose Estimation Using Rigid Structure. 2010
15. CMU. Graphics Lab Motion Capture Database[DB/OL]. 2007
16. Chao Y W, Wang Z, He Y, et al. HICO: A Benchmark for Recognizing Human-Object Interactions in Images. 2015



北京航空航天大学
BEIHANG UNIVERSITY

谢谢!

虚拟现实技术与系统国家重点实验室

State Key Laboratory of Virtual Reality Technology and Systems