PiGraphs：Learning Interaction Snapshots from Observations

2019/3/19

相关工作

- Work in computer vision and robotics has jointly modeled common human activities and interactions with objects observed in RGB-D data [Koppula et al. 2013; Wei et al. 2013a; Wei et al. 2013b].
- Encode object-object features as presented by ICON[Hu et al.2015]
- Activity model based on agent annotations containing position,orientation and action information[Fisher et al.2015]
- Prior works always treat the environment as input and the poses as output.

目标

- Automatically generate 3D depictions of interactions by modeling how people interact with objects.

- PiGraphs:A human-centric graph-based representation that encodes objects and body parts as nodes,and interactions between nodes as edges.

APPROACH

- A data-driven approach can be tailored to specific domains and is straightforward to scale.
- Perform skeletal tracking of people as they interact with scanned environments,using a stationary RGB-D sensor.
  - iGraphs nodes: human body joints or segments of geometry within the scene.
  - iGraphs edges: specific observed contact or gaze events
- PiGraphs: aggregate iGraphs with the same action annotations to generate a PiGraph. （a set of iGraphs）

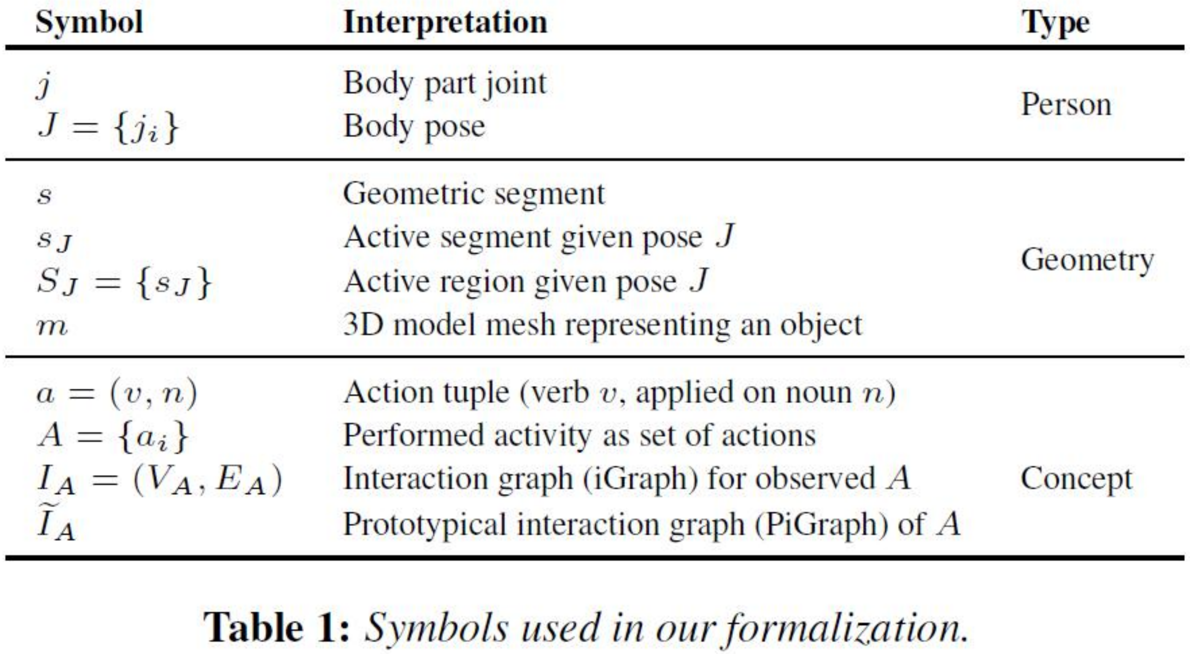
任务定义

- Interaction snapshot generation
  - Generate a interaction snapshot from terse(简洁的) specification.
- Input & output
  - Input:
    - 3D models with categorical label ci for each model mi
    - Interaction A (verb-noun pairs)
  - Output:
    - A snapshot *IS=(J,M) (*posed figure J ,a set of models M*)*
    - *M={(mi,Tj)}* consist of a model mj and associated transform matrix Tj

具体实现

- 获取数据
  - 获取环境的三维重建数据
  - 使用RGB-D相机观察人在环境中的交互
  - 追踪交互过程序列中的人类骨骼，得到骨骼数据，将3D 关节位置映射到重建的场景坐标。
  - 标注动作和物体种类（segment level）
- 数据集：
  - 30个场景，63个观测
    - 视频主体为4男一女
    - 298个动作，43个动名词组合（sit-chair），19个物体种类(couch,bed,keyboard,monitor.etc)

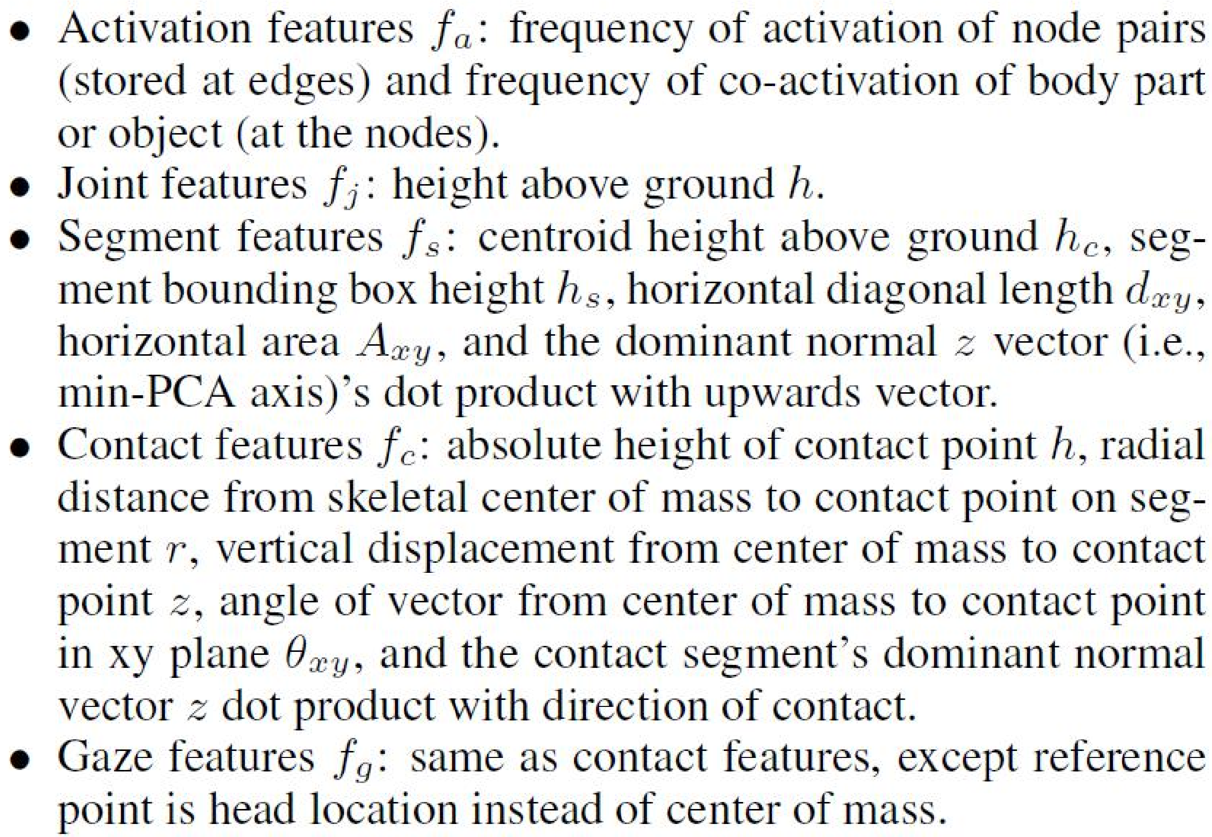
具体实现



Symbol	Interpretation	Type
j	Body part joint	Person
$J = \{j_i\}$	Body pose	
s	Geometric segment	Geometry
s_J	Active segment given pose J	
$S_J = \{s_J\}$	Active region given pose J	
m	3D model mesh representing an object	
$a = (v, n)$	Action tuple (verb v, applied on noun n)	Concept
$A = \{a_i\}$	Performed activity as set of actions	
$I_A = (V_A, E_A)$	Interaction graph (iGraph) for observed A	
$\widetilde{I}_A$	Prototypical interaction graph (PiGraph) of A	

Table 1: *Symbols used in our formalization.*

具体实现



- Activation features f_a: frequency of activation of node pairs (stored at edges) and frequency of co-activation of body part or object (at the nodes).
- Joint features f_j: height above ground h.
- Segment features f_s: centroid height above ground h_c, segment bounding box height h_s, horizontal diagonal length d_{xy}, horizontal area A_{xy}, and the dominant normal z vector (i.e., min-PCA axis)'s dot product with upwards vector.
- Contact features f_c: absolute height of contact point h, radial distance from skeletal center of mass to contact point on segment r, vertical displacement from center of mass to contact point z, angle of vector from center of mass to contact point in xy plane θ_{xy}, and the contact segment's dominant normal vector z dot product with direction of contact.
- Gaze features f_g: same as contact features, except reference point is head location instead of center of mass.

- Attach real-valued features on the nodes and edges of iGraphs depending on the type of node or edge.
- The nodes contain distributions over the segment features and nouns.
- The edges contain a connection probability (i.e., activation probability) and PDFs over the attributes of the connection.

补充：Von Mises distribution(冯·米塞斯分布)

指一种圆上连续概率分布模型，它也被称作循环正态分布

角度X

$$f(x \mid \mu, \kappa) = \frac{e^{\kappa \cos(x-\mu)}}{2\pi I_0(\kappa)}$$

冯·米赛斯分布应用于定向统计中，描绘了来自于相互独立的角度偏差小样本之和的角度分布，样本空间比如有目标感知，或颗粒材料中的颗粒方向。

x = a + bi + cj + dk 高阶复数，描述四维
空间可以执行绕任意过原点的向量的旋转

Pose Representation

- Using a hierarchical joint angle encoding to represent poses.(orientations & positions of 25 joints)
- Encode each joint as quaternions(四元数) and positions relative to parent joints in the kinematic chain.
- Encode the global vertical orientation of the skeleton with respect to the up vector.
- convert the joint orientation quaternions into latitude, longitude, and roll angles for which we fit von Mises distributions
- We also fit a normal distribution to each bone length, normalized by total bone length for stability across individuals
- The von Mises distributions (for orientation angles) and Gaussian distribution (for bone length ratio) at each joint form a total pose distribution under which we evaluate the likelihood of a given pose. We use this distribution to **sample** for **likely poses during snapshot generation**.

Learning from observation

- Give action observation A with pose J, we extract from the scene a set of active geometry segments sJ
    - 1.通过接触（坐着凳子）$r_{act} = 10$ 来找到重建场景mesh的顶点
    - 2.通过视野范围（头部看着电脑屏幕）2m 范围（对于重叠平面有一些问题
- Graph Construction
    - 根据参与的segments {sj}，我们可以得到对于动作A的iGraphs。
        - 在姿态J中的所有关结节点ji都可以表示为一个由关结节点特征表示的点。
        - 人体结构通过用边连接的节点表示
        - Segments 构成的节点由segment feature 构成
        - Contact and gaze edges contain contact and gaze features
- Aggregating iGraphs into PiGraphs
    - Initialization：
        - 新的 $\tilde{I}_A$ 由节点j_i和对应的骨骼边构成。
        - 属于pose的节点和骨骼边只有单个直方图，属于segment的节点和动作边由直方图集合构成，对每一类segment，直方图代表了在segment 特征和接触/视野特征下的条件分布。
    - Aggregation:
        - 动作集合中的每个iGraph，每一个关节点的特征都加入到对应边的直方图中，接触结点、可视结点和对应的边，将观测到的特征归入不同label的segment 直方图中。

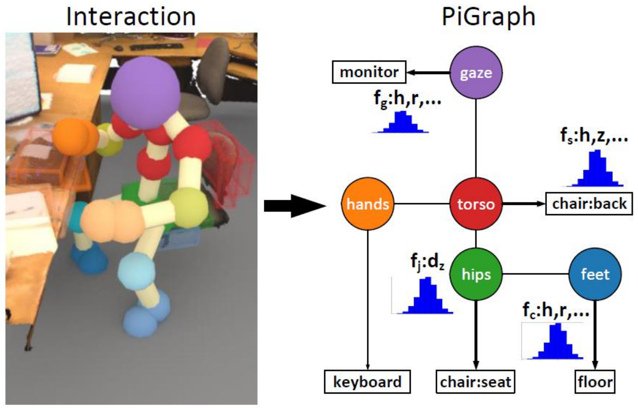


Figure 3: *Aggregation of iGraphs. Left: activated segments highlighted in boxes colored corresponding to body part. Right: features of the segments and their linkage to a body part are computed and aggregated into the nodes and edges of the PiGraph.*

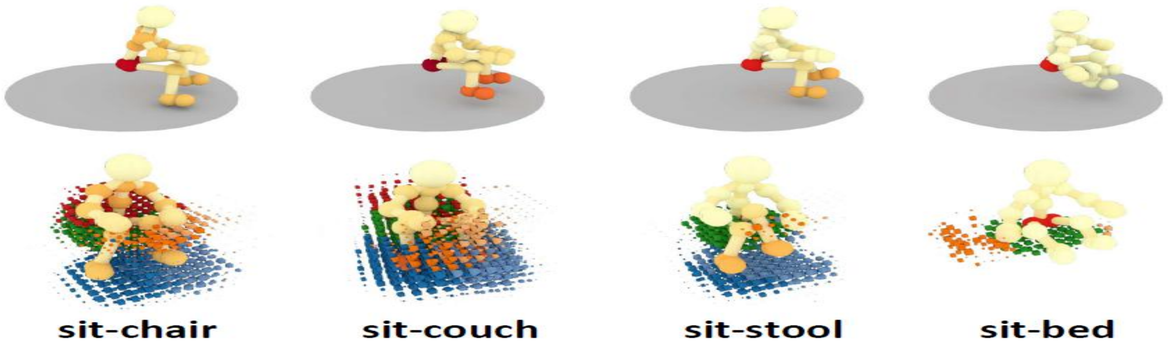


Figure 5: *Comparison of body part interaction weights and interaction volume priors for sitting on different types of objects. Top: torso interaction weight decreases from left to right while hip weight is high overall. Bottom: height and density of the hand interaction (orange voxels) shifts due to the different object categories.*

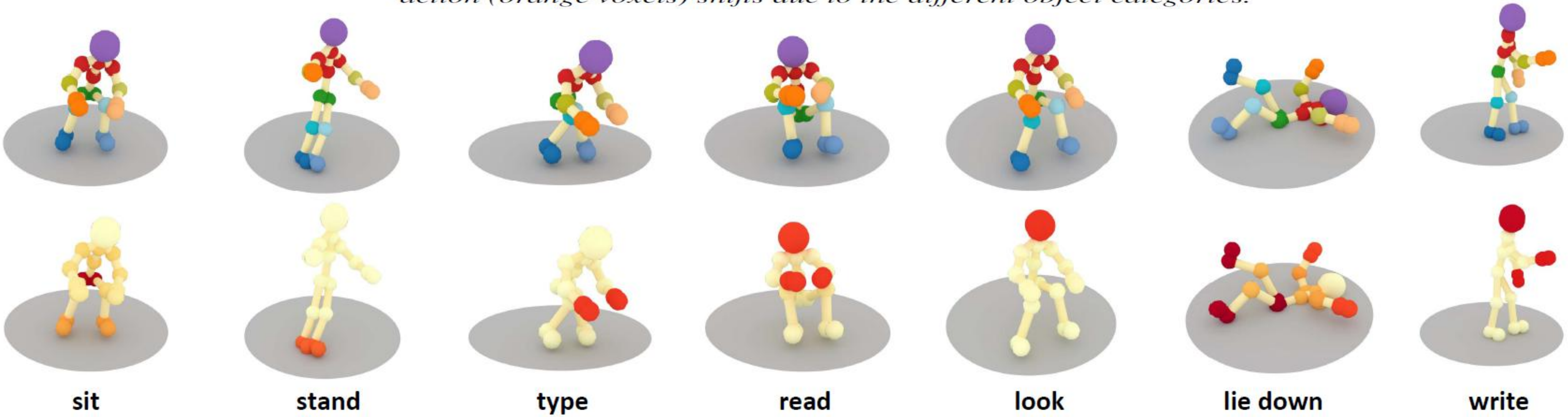


Figure 4: *Top: maximum likelihood poses for aggregated skeleton distributions of some action verbs. Bottom: conditional probabilities of body part interaction with objects during each action (indicated as red saturation). Parts critical to each action have high interaction probability. The somewhat atypical "write" pose is due to all our observations of writing being "writing on whiteboard" interactions.*

Learning from observation

- Aggregating iGraphs into PiGraphs
    - Joint Weight
        - 通过条件概率定义（sit 与 hip gaze与look）
    - Joint Weight 对于不常见的动作例如抓和调换等，不能被学习到（样本量太少）
- Encode Human Pose Distribution
    - 在每个关节点表示为循环正态分布（冯·米塞斯分布）和高斯分布。每个关节点的朝向根据其运动学上的父节点确定。
    - 假设关节点相互独立，所以可以分别学习每个关节点的分布。

Generate Interaction snapshot

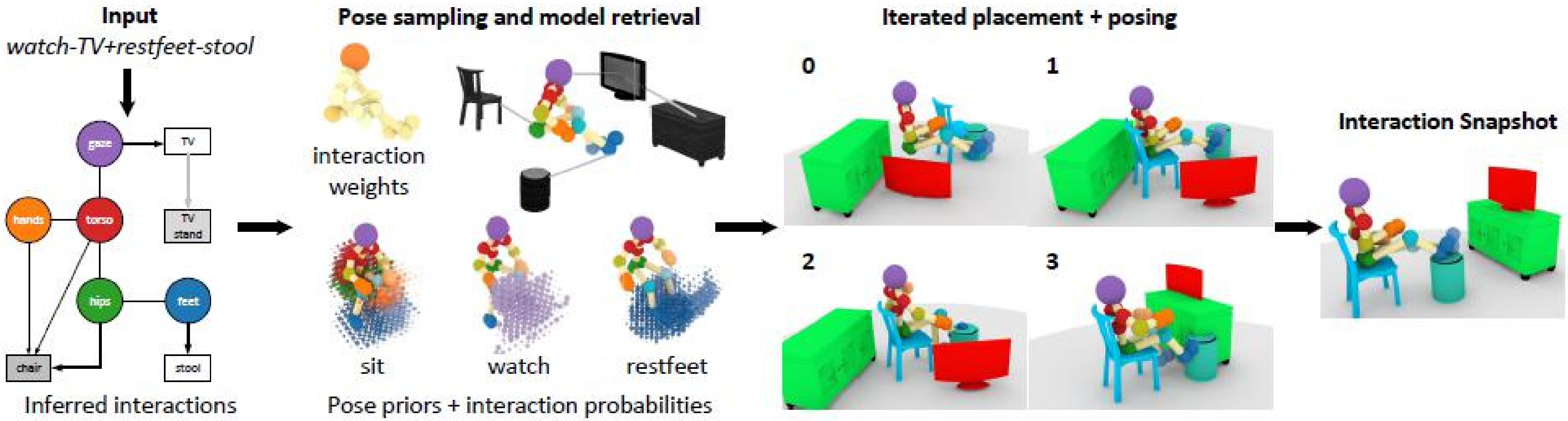


Figure 6: *Overview of pipeline for generating interaction snapshots. Left: inference on the input specifications expands the set of objects and relations. Middle: pose, 3D models and interaction priors are retrieved from a matching PiGraph. Right: iterative re-posing and object arrangement to optimize likelihood under PiGraph and generate final interaction snapshot.*

1. THE INPUT SPECIFICATIONS ARE USED TO RETRIEVE THE CORRESPONDING PIGRAPH.
2. INFER THEIR SUPPORT RELATIONS AND CONTACT&GAZE LINKAGE.
3. RETRIEVE THE MODEL MATCHING THE OBJECT.
4. SAMPLING FROM DISTRIBUTION OF THE PIGRAPH TO GET A POSE.
5. USE PRIORS PIGRAPH ITERATELY TO SCORE THE POSE AND THE OBJECT ARRANGEMENT.

Sampling approach

- 在每个步骤中，相应地使用PiGraph的pose分布和对应的物体先验对pose和物体摆放可能性进行评分。 检索模型并按支持层次结构的顺序从支撑对象到被支撑的对象，从最大到最小。
- 然后对物体支撑平面、位置、朝向和Pose 关节角度进行采样。每次采样根据PiGraph中的边和节点的分布进行评分
- 先对pose采样，此时固定物体摆放。再对物体摆放采样，固定pose。

Interaction Graph and Support Prediction

- To infer an iGraph, we take the set of verb-noun tuples and determine the set of objects and their interacting joints
- 确定物体之间的支撑关系和支撑人的物体
- 推断交互连接
  - 假定一个关节最多和一个物体交互
  - 假定在交互过程中同一种类的物体只有一个（避免歧义 坐着椅子，靠着椅子（意味着只有一把椅子））
  - 通过关节和对应物体出现的条件概率确定物体种类。
  - 通过设定阈值为0.05 筛选掉不可能出现的组合
- 推断支撑关系
  - 从数据中获取曾经出现的支撑关系，其中具有最大的性的即设定为物体之间的支撑关系
  - 膝盖和臀部之间的垂直距离大于小腿的平均长度即定义为站立，否则将与臀部交互的物体作为pose的支撑。
  - 出现物体没有被识别为支撑或者被支撑，则假定其被交互关节所支撑（拿着书看）

Interaction Graph and Support Prediction

- **Model Retrieval and Object Placement**
  - 从类别c中随机挑选一个模型
  - 支撑人的物体先放，再根据支撑关系，自底向上放置
  - 在父支撑面朝上的水平支撑面上采样，maximize 给定条件下点到交互关节距离的概率。确定后这些物体正向朝着人。
- **Generating Poses**
  - 根据给定的PiGraph从pose 分布中采样得到可能的poses，对骨长度和骨关节角度取均值
  - 剩余能转动的关节可以采样作为关节点的朝向

Scoring

$$L_A(J, M) = w_p L_{p_A}(J) + w_o L_o(M) + w_i L_{iA}(J, M)$$

$L_A(J,M)$: 总得分

$L_{p_A}(J)$：pose 得分 $L_{p_A}(J) = \sum_i V_i(j_i) - C(J)$

$L_o(M)$：物体摆放得分 $L_o(M) = \sum_{m_i, m_j, j \neq i} (1 - C(m_i, m_j))$

$L_{iA}(J,M)$：交互得分 $L_{iA}(J, M) = sim(I, \tilde{I}_A)$ $sim(I, I') = w_{gaze} sim_{gaze} + w_{acon} sim_{acon}$

Pose 得分

对于每一个关节点的朝向

$$L_{p_A}(J) = \sum_i V_i(j_i) - C(J)$$

V_i: log likelihood function for von Mises distribution at joint j_i orientation
$C(J)$: a [0,1] normalized measure of self collision by by point sampling the oriented bounding box of each bone and checking how many points are contained by other bones (a cross section radius of 10 cm is used for all bones except the torso which has a radius of 15 cm).

补充：Von Mises distribution(冯·米塞斯分布)
指一种圆上连续概率分布模型，它也被称作循环正态分布

角度x

$$f(x \mid \mu, \kappa) = \frac{e^{\kappa \cos(x-\mu)}}{2\pi I_0(\kappa)}$$

冯·米赛斯分布应用于定向统计中，描绘了来自于相互独立的角度偏差小样本之和的角度分布，样本空间比如有目标感知，或颗粒材料中的颗粒方向。

Object Placement Score

$$L_o(M) = \sum_{m_i, m_j, j \neq i} (1 - C(m_i, m_j))$$

添加惩罚项

因为这里的组合情况很少，所以只使用了简单的惩罚项来获取支撑回报（model i 和model j 的bounding box 是否有重合已经重合程度）

Interaction Score

$L_{iA}(J,M) = sim(I,\tilde{I}_A)$ (1) $sim(I,I') = w_{gaze} sim_{gaze} + w_{acon} sim_{acon}$ (2)

两个iGraphs 之间的相似度为（2）

$$\mathrm{sim}_{gaze} = max_{(s,s')}(sim(f_s, f'_s) \times sim(f_e, f'_e))$$

s:gazed segment in I
e:gaze edge for s
使用角相似性衡量

$$\mathrm{sim}_{con_j} = max_{(s,s')}(sim(f_s, f'_s) \times sim(f_e, f'_e))$$

s:contacted segment
e:contact edge

$$sim_{acon} = \sum_{j \in J} sim_{con_j}$$

设置$w_{gaze} = w_{acon}$
Use the probability that a feature vector f from I_{J,S_J} is drawn from an aggregated histogram hist of $\tilde{I}_A$ $sim_{hist}(f) = P_{hist}(f)$

Result

- 交互可视化

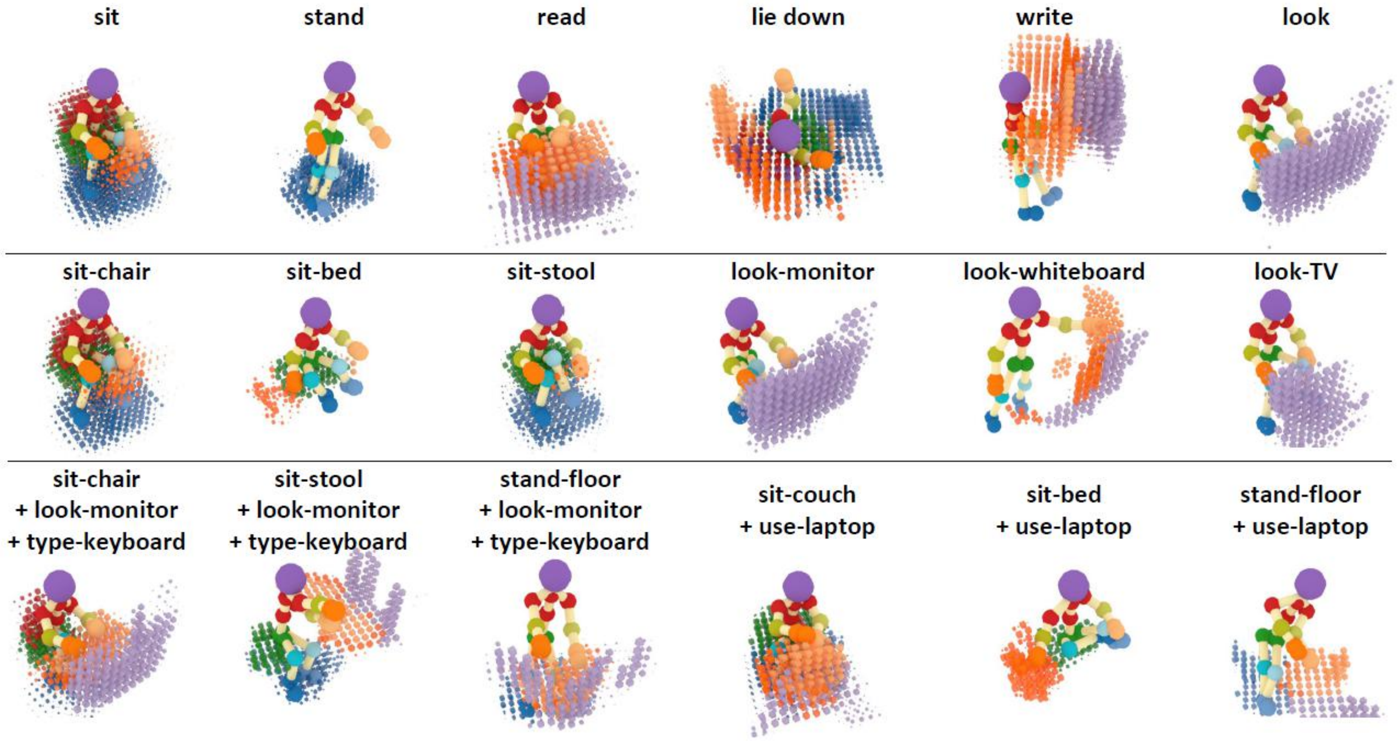
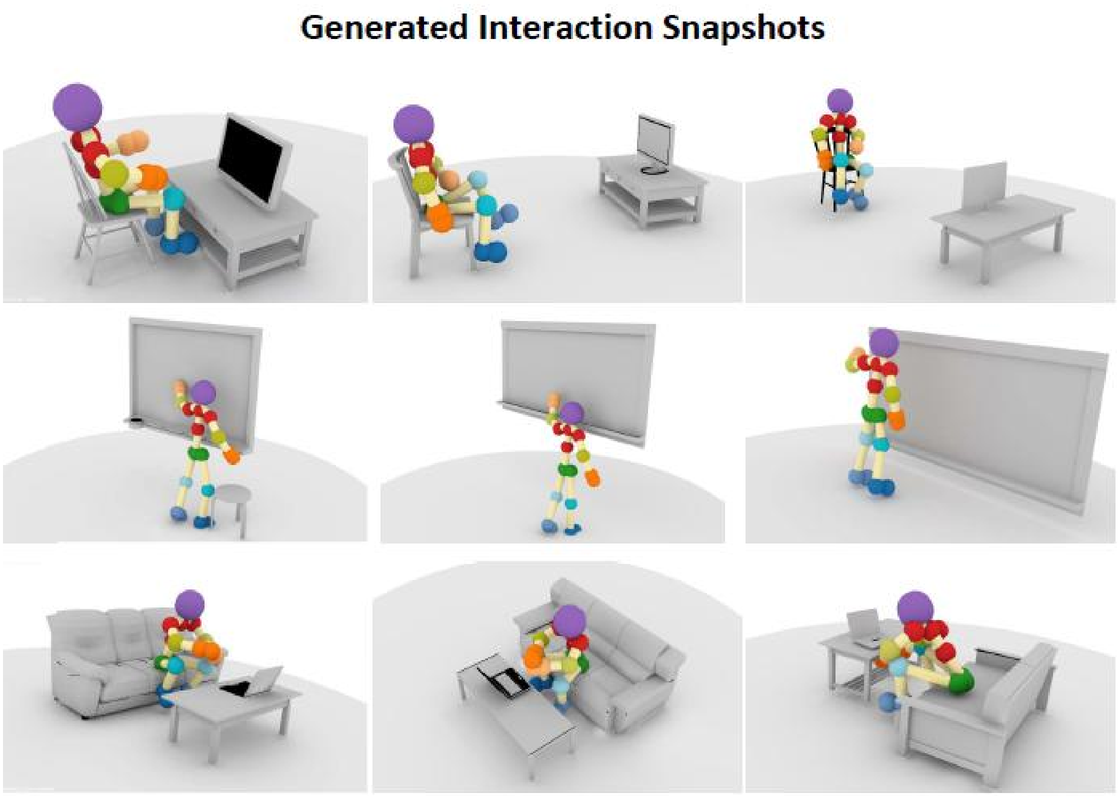


Figure 9: *PiGraph interaction volumes for some actions. The voxels around the pose are colored corresponding to body part, and their size is determined by the probability of a contact or gaze linkage occurring with the given body part. These priors summarize intuitive but rarely stated and represented facts about human pose–geometry coupling. For example: looking means gazing at geometry in front of one's head; when sitting in chairs there is usually a backrest part behind the torso, in contrast to sitting in bed and sitting on a stool.*

Generated Interaction Snapshot



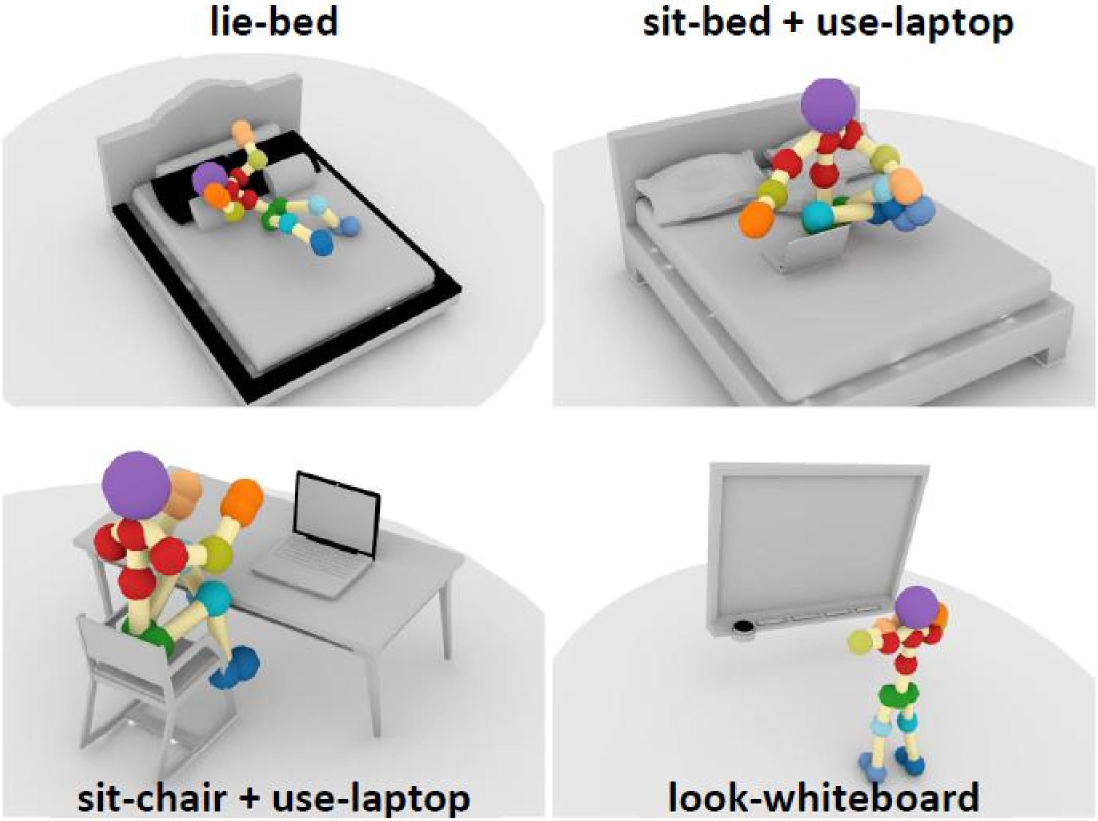


Figure 8: *Interaction snapshots generated by sampling PiGraphs learned from various action observations.*

Evaluation

- Human Judgement Study

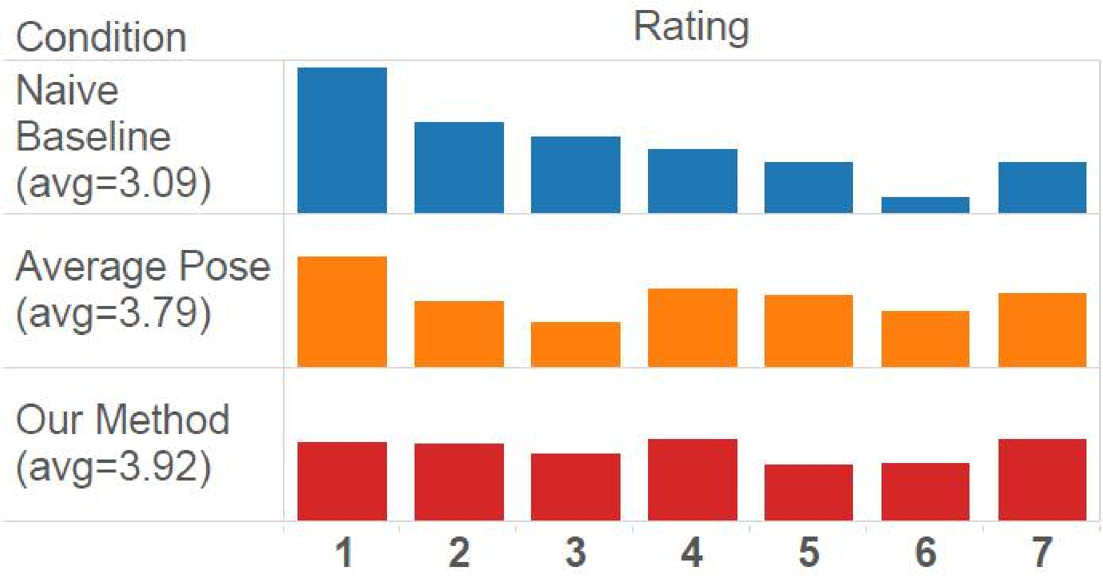


Figure 10: *Rating distributions for the interaction snapshot quality study (higher is better, 2072 judgments across conditions). The naive baseline performs worst as expected. Using the average pose instead of sampling and scoring poses results in a lower perceived quality. This illustrates the benefit of the pose priors.*

Conclusion

- We introduced the PiGraph representation to connect geometry and human poses during static interactions. We showed that PiGraphs can be learned from RGB-D data and used to generate a variety of interaction snapshots
- We presented a new framework for jointly modeling human pose and object arrangements during common interactions. Our method offers a novel view of geometry through the lens of interactions

THANKS