

指导教师： 杨涛

提交时间： 2016.3.13

CVPR2015 Paper Translation

No: 01

姓名： 罗俊民

学号： 2013302502

班号： 10011302



奇怪姿势关节角度限制下的三维人体姿势重建

Ijaz Akhter and Michael J. Black

Max Planck Institute for Intelligent Systems, Tübingen, Germany

{ijaz.akhter, black}@tuebingen.mpg.de

摘要

人们在图像和录像中通过二维的关节位置来分析重建人们的三维姿势。解决事实问题本质上是病态的姿势判断，很多方法对人类前面的姿势来判断。不幸的是，这些方法承认无效是因为他们没有模拟关节限制对姿势的影响。在这里，我们做了 2 个关键的贡献，首先，我们收集了一个运动捕捉数据集，探索了广泛的人类构成。从这一点，我们学习了构成我们先前的共同限制的一个构成依赖模型。这两个数据集和可用于研究目的。其次，我们定义一个新的身体姿态，多级通用参数化，方法估计的三维姿态从二维关节位置使用过完备字典的姿势。我们的方法表现出良好的泛化，同时避免不可能的构成。我们定量比较我们的方法与最近的工作和显示 2D 先进成果的三维姿态估计使用 CMU 动作捕捉数据集。我们也显示出优异的结果使用手动注释对利兹运动姿态数据的真实图像和自动检测。

1. 简介

在三维人体姿态下的精确建模是计算机视觉中的许多问题的根本。以往的先验不适合人类构成的多样性或不足的限制，避免无效的姿势。我们提出了一个物理的动机之前，只允许有效的姿势的人体测量和制约是无效的。

一个可以使用关节角度限制，以评估是否有 2 个连接的骨头是否有效。然而，它是建立在生物力学中，有一定的对骨之间的关节角度限制的依赖关系[12, 17]。例如，一个人的手臂可以弯曲多少取决于它是在前面，还是后面，后面。医学教科书仅提供关节角度限制在几个位置[26, 2]和完整的配置的姿态相关的关节角度限制为完整的身体是未知的。



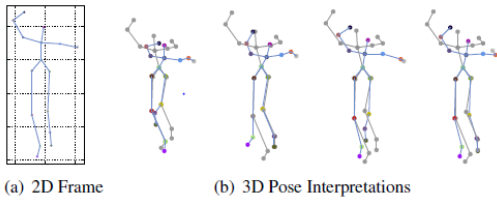
Figure 1. Joint-limit dataset. We captured a new dataset for learning pose-dependent joint angle limits. This includes an extensive variety of stretching poses. A few sample images are shown here.

我们发现，现有的动作捕捉数据（如 CMU 数据集）是不足以了解真正的关节角度的限制，在某些限制造成依赖。因此，我们拍摄的一个新的人类运动，包括一个广泛的各种各样的姿势伸展训练运动员和体操运动员进行数据（见图 1）。我们学习造成相关的关节角度限制，从这个数据，并提出了一种新的基于这些限制。

建议的先验可用于估计三维人体姿态的问题是模棱两可的。我们的姿态参数化是特别简单的，一般的运动骨架的三维姿态是由两个端点定义坐标在直角各骨。约束的三维姿态保持有效的优化过程中，只需要我们的惩罚项的目标函数。我们也证明了我们之前的可以结合稀疏表示的姿态，选择从一个超完备字典，定义人体姿势一般没有准确的参数化。

我们使用我们事先估计三维人体姿势从二维关节位置。图 2 演示了这个问题的主要困难。在图 2(1)中给出了一个单一的视图，三维构成是暧昧[27]，并存在几个可能的三维构成如图 2(b)，所有导致相同的二维观测。因此，没有通用的先验信息静态身体姿势是足够的，以保证一个单一的正确三维姿态。在这里，我们寻求最有可能的，有效的，人类的姿势。

我们在图 2 展示了一个关键的三维知识估计步骤。a 中只给出二维的关节位置，这



里有几个三维姿势通过同一幅二维图像重建出来的。他们中的一些是在 (b) 的颜色点，而灰色点代表地面的真相。在这里，我们展示的姿势从一个不同的三维视图，这样的差异是明确的，但所有这些构成项目的完全相同的二维观测。

二维位置是相机参数估计。由于人体姿势的多样性，不正确的相机参数可以导致一个不正确的姿态估计。为了解决这个问题，我们提出了一个分组的身体部分，称为“扩展躯干，”组成的躯干，头部和腿。开发的事实，即扩展的躯干的姿势的变化是不少于全身，我们估计其三维姿态和相应的相机参数更容易。估计的相机参数，然后用于全身姿态估计。建议的多步解决方案，大大提高了以前的方法。我们评估的三维姿态估计从二维的一个范围广泛的姿势和相机视图使用债务工具中央结算系统的运动捕捉数据集 1 活动。这些都是比较复杂的和不同的数据所使用的数据，比以前的方法，我们表明，以前的方法有麻烦在这种情况下。我们也报告手动注释和利兹运动姿态数据的基础上自动检测 [16] 的部分结果。用于评价和所有软件的数据是可用的其他研究人员与我们的研究结果相比，[1]。

2. 相关工作

文献对模拟人类构成先验和从点、图像、视频、深度数据的三维姿态估计，等广泛。大多数以前的方法对人体姿态建模假设固定关节角度限制 [7, 24, 28]。贺达等人。肘关节角度限制关节角度限制的模型依赖关系 14。他们的模型不能用于我们的 2D 到 3D 的估计问题，因为它需要周围的骨轴而旋转被称。豪伯格等人。【13】建议在关节空间的骨头的端点的分布方面，这种先验模型。我们要进一步确定我们的完全三维骨位置模型。

三维人体姿态估计从二维点观察到的静态相机有大量的论文。所有这些方法都必

须通过使用附加信息来解决固有的模糊性。方法各不相同。

李和陈 [18] 恢复通过修剪一个二进制的解释树表示所有可能的身体配置。泰勒 [29] 解决了深度模糊，使用人工干预。巴尔和 kakadiaris [4] 利用关节角度的限制约束来解决这个歧义。拉梅瓦朗 Chellappa [21] 使用 3D 模型的不变量来恢复关节角度配置。benabdelkader 和 Yacoob [5] 利用肢体长度估计他们比统计范围。关等。[11] 使用人体测量数据库和一个人的已知性别和身高来预测骨长度。bourdev 和 马利克 [6] 估计姿势要点其次是手动调节。江 [15] 使用泰勒的方法，并提出假设 exemplarbased 修剪方法。室利罗摩克里希纳等人。[23] 提出了一个完整的字典的行动估计三维姿态。这些方法不施加关节角度的限制，可以潜在地估计一个无效的三维姿态。单眼姿态估计中的一些模糊性得到一个序列（但并不总是）来解决。魏和柴 [31] 和 [30] 和 valmadre Lucey 估计三维姿态从多个图像并利用关节角度的限制。要应用关节角度限制，首先必须有一个运动树结构中的坐标轴被明确定义。鉴于每骨仅 2 点，这本身是一个严重的不适应问题，需要先验知识。valmadre 和 依据需要手动解决解决这个问题。

我们的身体表示简化了这一问题，因为它不代表不可观测的轮换的四肢。我们相信我们的第一项工作是提出一个在直角坐标系中的运动骨骼关节角度的限制，其中只有两点是已知的。在计算机图形学中，也存在着人为的二维注释的人为构成动画的方法。grochow 等人。[10] 提出了一个比例的高斯潜变量模型作为一个三维姿态之前。他们的方法的空间复杂度是一个二次的训练数据的大小。卫和柴 [32] 和林等。[19] 需要额外的约束，如已知的关节或接地平面之间的距离，以解决位姿估计中的模糊性。Yoo 等人。[33] 和彩。[8] 提出了一个三维姿态估计的草图界面。他们的方法只适用于训练数据中的姿势。歧视性的方法也存在于文献中，不需要二维点的对应和直接估计人类构成从二维图像测量 [3, 20, 22, 25, 34]。歧视性的方法一般限制从训练数据中学习

的观点。虽然我们的数据集可用于判别方法的培训，它可能会需要培训的每一个新的应用。与之相反，我们以前可以很容易地纳入到生成的姿态估计和跟踪方法。

3. 优先判断姿势

我们观察到，现有的动作捕捉数据并不是设计来探讨造成依赖关节角度的限制。因此，我们捕获了一组新的人体运动进行灵活的人如体操和武术。我们的捕获协议的目的是引起广泛的成对一种运动树中的连通性（图 4）的配置。我们 captured 两类运动。在运动捕捉范围，参与者被要求保持上臂固定，充分伸缩和扩展他们的下臂，然后将它们向内向外。这一动作是重复的一个数字的水平 and 垂直的姿势的上臂。腿采用同样的方法。他们也被要求做一些伸展运动（图 1）。从这个数据，我们估计一个 17 点的运动骨骼和关节角度的限制。我们所代表的人体姿态作为一个串联的 P 的三维坐标点 $x = XT1 \dots XT P T \in R3P \times 1$ 。（一）作为一个运算符，其返回相对于其父在运动骨骼中的相对坐标系。我们扩展的向量和矩阵的点。我们的目标是找到一个函数的有效性

$$\text{isvalid}(x): R3 \times N \rightarrow \{0, 1\}^n,$$

其中 n 是指骨头的数量和价值的 1，如果相应的骨是在一个有效的姿势，否则返回 0。给定一个运动的骨骼，我们首先找到一个本地坐标系，为每个骨，我们讨论下。

3.1. 全局到局部坐标转换

为了估计关节角度的限制，我们需要首先找到所有的关节的局部坐标系。我们可以发现在 3D 方面的两个不平行的向量 u 和 v 的三个坐标轴可以使用克施密特对 U, V 发现一个坐标轴，你告诉我们提出了一个转换从 X 到局部坐标 \tilde{X} 算法 1。为上臂、大腿和头，U 和 V 是指借助躯干的“骨头”（脊椎，左/右髋，左/右肩）（系 3-8）。其他骨，B 的坐标系统的选择是任意的，是一个任意的向量，定义的帮助下，与母骨，PA (B)，B (第 10-11 行)。汝是估计旋转的父骨。改变值

输入向量，一个，可以产生不同的坐标系，并通过保持其价值固定，我们确保本地坐标系的一致性。最后的局部坐标轴使

用的是克施密特发现（12 号线）和局部坐标 \tilde{B} 计算（13 号线）。

Algorithm 1 Global to Local Coordinate Conversion

```

1: Input  $\delta X$  and a constant arbitrary 3D vector  $a$ .
2: for  $b \in \delta X$ 
3:   if (  $b$  is an upper-arm or head)
4:      $u = \text{Left-shldr} - \text{Right-shldr}$ ;
5:      $v = \text{back-bone}$ ;
6:   else if (  $b$  is an upper-leg)
7:      $u = \text{Left-hip} - \text{Right-hip}$ ;
8:      $v = \text{back-bone}$ ;
9:   else
10:     $u = \text{pa}(b)$ ;
11:     $v = R_u a \times u$ ;
12:    $R_b = \text{GramSchmidt}(u, v, u \times v)$ ;
13:    $\tilde{b} = R_b^T b$ ;
14: Return  $\tilde{X} = \{\tilde{b}\}$ ;
```

3.2. 学习关节角度限制

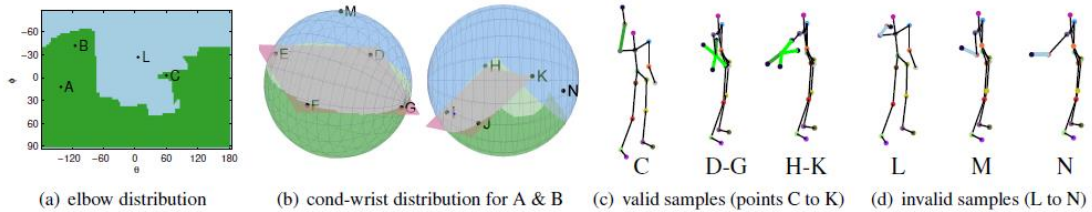
我们将上手臂，上腿和头部的局部坐标转换成球面坐标。使用我们的数据集，我们定义了一个二进制占用矩阵，这些骨骼在离散的方位角和极角，分别 θ 和 ϕ 。骨被认为是在一个有效的位置，如果其方位角和径向角度给出一个值 1，在相应的占用矩阵（图 3 (1)）。所有其他骨骼的有效性都取决于其父母与一个给定的 θ 和 ϕ 的位置。在这种条件下，骨骼只能位于半球甚至较小的部分。利用我们提出了两种类型的约束检查 B。首先我们找到一个半空间的有效性， $BTn + d < 0$ ，由一个分离平面的法向量 N，距离原点 D。二我们项目中的数据集的所有 B 的实例的平面和找到一个边界框包围这些预测。骨被认为是有效的，如果它位于半空间，其投影在边界框（图 3）（图 2）。分离平面是由下面的优化估计，

$$\min_{n,d} d^2 \quad \text{subject to} \quad A^T n < -d1, \quad (1)$$

其中一个是在数据集的所有 B 柱副连接实例。图 3 显示了我们所学到的关节角度范围的可视化。结果表明，腕关节角度的限制是不同的肘关节的不同位置。

3.3. 增强三维姿态稀疏表示

为了表示三维姿态，稀疏表示，提出了在 [23]，使用线性组合的基础构成图 3。姿态相关关节角度极限。（一）在方位角和极角的右肘的占用矩阵：绿色/天蓝色区域代表在我们捕获数据中所观察到的有效/无效的姿势。（乙）在一个和乙的肘部位置，手



腕只能躺在绿色区域的领域。这些有效的手腕位置项目到一个盒子上的平面分离的有效和无效的姿势。图表明，有效的手腕的姿势取决于肘的位置。(丙)及(或)说明有效(绿色)和无效(天空蓝)肘和腕的位置，以供相应选择点(一)和(甲)和(乙)。

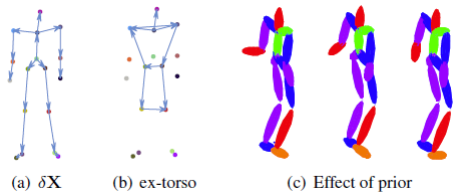


图 4。表示与模糊。(一)运营商通过考虑作为原产地的父计算相对坐标。(乙)扩展躯干的贝叶斯网络利用了躯干和左、右膝关节的相对刚性的位置。(丙)过完备字典表示允许无效的构成。左至右：我)一个三维姿态，在正确的下臂违反关节角度限制。(二)超过完整的字典表示这个无效的三维姿态与少量的基础构成(20 相比，与全维的 51)。三)应用我们的关节角度限制，使无效的姿势有效。

$$\hat{X} = \mu + \sum_{i=1}^K \omega_i B_i = \mu + B^* \omega, \quad (2)$$

$$\{B_i\}_{i \in \mathcal{I}_{B^*}} \in B^* \subset \mathcal{B},$$

B_1, B_2, \dots, B_K , 加上平均的姿态 μ , 哪里! 是构成系数的向量, ! 我, 矩阵 B^* 是一列连接的基础构成双选择 \mathcal{I}_{B^*} 列指数从过完备字典 B 是通过连接许多行动和每个基础的基础计算机学会使用主成分分析 (PCA) 的一个动作类. \hat{x} 表示近似的三维姿态的基础上排列的姿势和相关

要估计的姿态, X , 由摄像机旋转 R , $X \approx (I_{P \times P} \otimes R) \hat{X}$. 这种稀疏表示提供了更好的推广比主成分分析[23]. 我们观察到, 尽管良好的泛化, 稀疏表示也允许无效的姿

势。它是很容易留在空间的基础载体, 但移动以外的空间的有效构成。图 4 (三) 显示, 少量的基础构成可以重建一个无效的三维姿态, 而我们的关节角度限制在防止无效配置。

$$\min_{\omega} \|X - (I \otimes R)(B^* \omega + \mu)\|_2^2 + C_p, \quad (3)$$

我们估计这个构成, 通过求解以下优化问题在 $k.k_2$ 表示 L_2 范数和 $CP=0$, 如果所有的骨头 \hat{X} 是有效的根据功能 $IsValid(\cdot)$ 和 INF 否则。定义该方法是等价的使用非线性不等式约束的有效功能 (\cdot) 。

4. 三维姿势预估

4.1. 预先说明的知识

$$x = s(I_{P \times P} \otimes R_{1:2})X + t \otimes \mathbf{1}_{P \times 1}, \quad (4)$$

$$C_r(\omega, \mathcal{I}_{B^*}, s, R) = \|x - s(I \otimes R_{1:2})(B^* \omega + \mu)\|_2^2. \quad (5)$$

记得人体姿态被表示为一个串联的三维坐标点 $P x = X T_1 \dots X T_P T \in \mathbb{R}^{3P \times 1}$. 一个比例正交相机模式下, 二维坐标的图像中的点的 $x \in \mathbb{R}^{2P \times 1}$ 和 S, R, T 表示相机的尺度、旋转和平移参数, 这表示 Kronecker 积和下标 1: 2 给出了矩阵的前两行。我们可以做=0, 根据假设的三维质心得到映射到二维质心和这些都是起源的世界和相机坐标系统。一旦三维姿态是已知的, 可以用方程 (4) 估计的实际值。

室利罗摩克里希纳等人。[23]利用方程的稀疏表示 (2) 寻找未知的三维姿态 X . 他们最小化重投影误差找到以下! *, \mathcal{I}_{B^*}, S, R 使用贪婪的正交匹配追踪 (OMP) 算法进行人体测量的正则化, 一旦! *, \mathcal{I}_{B^*}, R 是已知的, 姿态估计:

$$X = (I \otimes R)(B^* \omega + \mu). \quad (6)$$

4.2 目标函数

我们的方法的三维姿态估计给定的二维关节位置，利用提出的姿势之前和骨长度遵循已知比例的事实。学习过完备字典我们同样选择 CMU 动作捕捉序列被室利罗摩克里希纳等人入选。[23]和加两进一步动作类“踢”和“童话剧。”为重点的姿态而不是身体比例将风扇等人的方法。[9]和正常化所有训练机构有相同的平均骨长度和所有机构有相同的比例，给予每一个训练主题相同的骨长度。我们将提出利用 Procrustes 对齐扩展躯干，定义如下。我们了解每个动作类的 PCA 基础和连接基地获得过完备字典。我们也学习主成分分析和协方差矩阵的扩展躯干，我们使用它的姿态估计在下一节。

$$\min_{\omega, s, R} C_r + C_p + \beta C_l, \quad (7)$$

我们估计的三维构成的最小化，这里是一个归一化常数和成本估计 CL 惩罚与骨长度为 k 的平方差 (\hat{x}_i) K2 和归一化平均骨长度

$$C_l = \sum_{i=1}^N \left| \|\delta(\hat{X}_i)\|_2^2 - l_i^2 \right|$$

里， $|\cdot|$ 表示绝对值和 X 估计方程 (2)。我们使用一个轴角参数 R。我们不是为基向量优化但他们分别估计了 4.4 节。一个重要的考虑，在最小化的成本，在方程 (7)，以及在以前的方法中的目标函数[22, 9]，是对初始化的敏感性。特别是一个很好的猜测的相机旋转矩阵的要求来估计正确的三维姿态。为了解决这个问题，我们注意到一个扩展的躯干，躯干，头部和腿的组成部分表现出较少的多样性比全身和它的姿态估计可以给出一个更准确的估计相机矩阵。

4.3. 伸展躯干的姿势估计

为延长躯干的三维姿态估计，我们尽量减少成本类似于方程 (7)，但不是全身曲线，X，我们只考虑点的扩展 X' 躯干。我们学习了一个主成分分析的基础上，为延长躯干的平均。因此，依据排列构成的， $\hat{X}' = B' \omega + s$ 。即使是基于 PCA 扩展躯干建模不足以限制其三维姿态估计从二维模型。我们之前 \hat{X}' 在贝叶斯网络的形式点之间的相互利用 (图 4 (b))。该网络利用的事实，即人体躯干几乎是刚性的，往往是左，右膝盖移动

的相关性。

$$p(\delta \hat{X}') = \prod_i p(\delta \hat{X}'_i | \delta \hat{X}'_T), \quad (8)$$

因此，一个姿势的概率在 \hat{X}' 我表示一个向量通过连接各点的三维坐标，在空调机组的贝叶斯网络的定义。假设对 \hat{X}' 我下， \hat{X}' 我是高斯分布的，我们在附录中显示，造成现有可写为一个线性约束， $AP' = 0$ ，在 AP 的基础上用 B' 和 X' 协方差矩阵的计算。因此，对于扩展躯干之前长期成为 C' $P = KAP' = K^2 2$ 。我们估计扩展躯干姿势通过最小化下面的目的类似于方程 (7)，

$$\min_{\omega', s, R} C_r' + \alpha C_p' + \beta C_l'. \quad (9)$$

我们初始化优化 findingrand 使用关节位置的 2D X' 和' 之间 μ Procrustes 对齐。我们找到解决方案，使用准牛顿优化。估计! 在下一阶段的基础上，对整个身体进行了估计。

4.4 基础估计

Algorithm 2 Orthogonal Matching Pursuit (OMP)

```

1:  $r_{p0} = x - s(\mathbf{I} \otimes \mathbf{R}_{1:2})\mu$ ;
2:  $r_{d0} = \delta \mathbf{Z}(\mathcal{I}_d) - s(\mathbf{I} \otimes \mathbf{R}_3)\delta\mu(\mathcal{I}_d)$ ;
   while  $t < K$  do
3:    $i_{max} = \arg \max_i (\langle r_{pt}, s(\mathbf{I} \otimes \mathbf{R}_{1:2})\mathbf{B}_i \rangle + \langle r_{dt}, s(\mathbf{I} \otimes \mathbf{R}_3)\delta\mathbf{B}_i(\mathcal{I}_d) \rangle)$ ;
    $\mathbf{B}^* = [\mathbf{B}^* \mathbf{B}_{i_{max}}]$ ;
4:    $\omega^* = \arg \min_{\omega} (\|x - s(\mathbf{I} \otimes \mathbf{R}_{1:2})(\mathbf{B}^*\omega + \mu)\|_2^2 + \|\delta \mathbf{Z}(\mathcal{I}_d) - s(\mathbf{I} \otimes \mathbf{R}_3)(\delta \mathbf{B}^*(\mathcal{I}_d)\omega + \delta\mu(\mathcal{I}_d))\|_2^2)$ ;
    $\mathbf{R} = \arg \min_{\mathbf{R}} \|x - s(\mathbf{I} \otimes \mathbf{R}_{1:2})(\mathbf{B}^*\omega^* + \mu)\|_2^2$ ;
5:   if !isvalid( $\delta(\mathbf{B}^*\omega + \mu)$ )
     remove  $\mathbf{B}_{i_{max}}$  and go to step 4
6:    $r_{pt} = x - s(\mathbf{I} \otimes \mathbf{R}_{1:2})(\mathbf{B}^*\omega^* + \mu)$ ;
    $r_{dt} = \delta \mathbf{Z}(\mathcal{I}_d) - s(\mathbf{I} \otimes \mathbf{R}_3)(\delta \mathbf{B}^*(\mathcal{I}_d)\omega^* + \delta\mu(\mathcal{I}_d))$ ;
7:   Return  $\{\mathbf{R}, \mathbf{B}^*\}$ ;

```

在这一步中，我们估计的基础上 B* 使用 OMP 算法类似于罗摩克里希纳等人。[23]。不同的是，在这里我们已经知道了一些骨骼的深度，通过利用关节角度限制的限制。此外，我们不强加一个硬约束，骨长度必须总和为一个预定义的数字。

让我们 Z 来表示三维构成的所有点的未知深度的向量。给定的平均骨长度李和估计正交尺度，我们估计的相对深度 |Z| 使用泰勒的方法[29]。由于自然人的构成是不是完全任意的，未知的迹象的相对深度可以估计为一些骨骼，通过利用关节角度的限制。

我们产生的所有迹象的骨头在手臂或腿，测试他们是否符合使用功能的有效性有效的姿势 (X)。骨的标志是采取积极的，如果根据我们之前，一个负的迹象是不可能在任何相应的手臂或腿的组合。如果不积极，我们做同样的测试，在另一个方向看，如果标志可以是负面的。如果不可能，我们必须依靠完备的基础。深度的指标估计这种方式表示为 ID。由于关节位置的 2D, X, 相对深度估计, Z (ID), S 和 R, OMP 目前估计, 算法 2, 在贪婪的方式进行。该算法从为 μ 姿态三维电流估计和计算初始残余的二维投影和已知的相对深度 (1、2 号线)。在每次迭代中从 B 的基础矢量的选择和添加到 B^* 是最一致的残余旋转电流估计 (线下 5)。然后给出了 B^* , 姿态系数! *和相机旋转 R 重新估计 (线 6,7)。我们将依据向量如果它使构成无效, 考虑下一个最高的点积的基向量 (线 8,9)。剩余使用 B^* 更新, ! *, 和 R 的新的估计 (线 10,11)。该算法终止时, 乙方 *已经达到一个预定义的大小。最后, 估计 B^* , ! *, R 是用来初始化优化方程 (7)

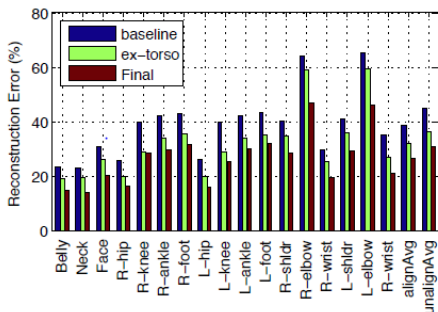


图 5. 扩展躯干的初始化和建议的姿势之前: 重建误差是平均欧氏距离每关节之间的估计和地面真理的三维构成和测量的背骨长度的一小部分。误差随每个模块的增加而减小。

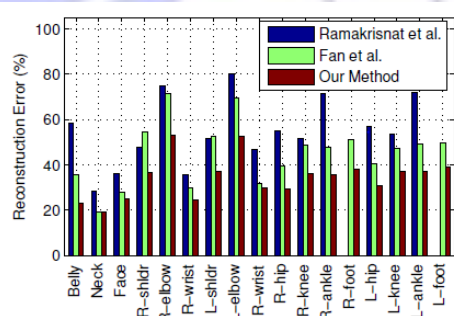


图 6. 所提出的方法给出了一贯较小的重建与其他方法比较的误差。

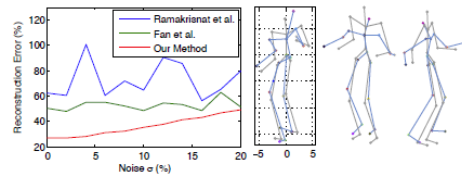


图 7. 所提出的方法是强大的一个相当大的范围与以前的方法比较的噪音。噪音 σ 背骨长度成正比。在 $\sigma = 20\%$ 和我们的估计的三维姿态与误差 = 50% 也显示在意见 (灰色: 地面真理, 有色: 估计)。

5. 实验

我们比较了姿势之前, 从我们的数据同之前从 CMU 集中学习。我们把所有的姿势在我们的数据集为有效或无效使用从 CMU 学习之前。我们发现了一个 110 分钟的数据约 12% 不是由 CMU 之前解释的基础。这表明, 债务工具中央结算系统的数据集不包含人类运动的所有范围。一个类似的实验表明, 在 9.5 小时内约 8% 的债务工具中央结算系统的数据不是由我们之前的解释。仔细研究发现, CMU 包含许多贴错标签标记。这将有效的空间构成包括 CMU 无效的。去除无效的姿势可能会增加我们的姿态, 不解释, 会减少我们的债务工具中央结算系统的数据之前无法解释量百分比。

我们定量评估的方法, 采用四位演员 CMU 动作捕捉序列 (103, 111, 124, 和 125) 共 69 个序列。我们创造了两套合成图像, 称为 testset1 和 testset2, 随机选择 3000 和 10000 从这些序列帧和投影使用随机相机的观点。我们报告的平均欧氏距离之间的估计和地面真理构成的平均欧氏重建误差。像以前的方法 [9, 23] 我们的 Procrustes 对齐三维姿态估计与地面真实计算误差。为了解决任意规模, 将地面的背骨长度和 Procrustes 对齐与估计的姿态。

我们还评估了相机矩阵估计。我们首先评估的影响, 延长躯干的初始化和关节角在所提出的方法对 testset1 之前。我们开始与一个基线组成的只是投影匹配追踪算法和测试其准确性。我们将通过寻找 R 和 S 的 Procrustes 对齐 X 和 μ 之间。然后, 我们包

括初始化使用姿态估计的扩展躯干和最终的联合优化，以执行长度的限制。最后，我们包括深度估计使用关节角度的限制和建议提出的联合优化。在图 5 中，我们报告了这个实验的平均重建误差。结果表明，单调减少错这些模块中的每一个。我们还报告总体平均重建误差和 Procrustes 对齐。对于后来的情况下，我们乘摄像机旋转的三维姿态，并采用一个典型的相机公约。观察两个错误是大致相等的，我们得出这样的结论估计相机矩阵是正确的。

接下来我们比较我们的方法的准确性对以前的方法对 testset2。作者提供的源代码，以前的方法是友好的。请注意，该方法由风扇等。为数类动作进行定制，包括行走、跑步、跳跃、拳击和攀爬，其精度预计会降低其他类型的动作。图 6 显示，该方法优于其他 2 种方法。在图 7 中我们检验我们对高斯噪声算法的灵敏度和比较反对的罗摩克里希纳等人的方法。[23]和风扇等。[9]。我们添加噪声的骨干长度在三维，项目的噪声点，使用随机相机矩阵和报告我们的姿态估计精度。结果表明，所提出的方法是显著更强大的比以前的方法。我们的实验表明，该方法提供了一个小的投影误差和人体测量有效的三维姿态的诠释，而以前的方法往往无效的三维姿态估计。进一步的调查结果显示，以往的方法在典型相机公约的重建误差比有显著差 Procrustes 对准（55%和 44%和分别为 143%和 145%），而我们的方法的误差没有明显差异（34% vs 45%）。

这意味着以前的方法的失败的一个重要原因是不正确的估计的相机矩阵。这突出贡献的扩展躯干初始化。重要的是要提到的三维姿态估计（见图 2），这意味着给定的二维点位置，一个正确的姿态估计不能被保险，只有一个可能的三维姿态估计固有的模糊性。结果表明，所提出的方法满足这个标准。

图 8 显示的是真实的图像与关节的关节的真实图像的结果，并与以前的方法相比，他们与以前的方法，显示出三维姿态在任意视图。再次的结果表明，我们的方法给出了一个有效的三维姿态，而以前的方法往往不。图 9 显示了在利兹运动检测[16]自动部

分结果构成的数据集上几帧。结果表明，尽管显著的噪声检测，所提出的方法是能够恢复一个有效的三维姿态。为了更好的结果，请参阅补充材料[1]。

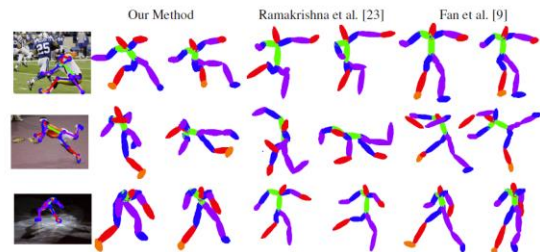


图 8。手工标注的实际结果。我们表现出大幅改善，比以前的方法。所提出的方法给出了二维关节位置的人体测量有效解释以往的方法常常会无效的姿态。

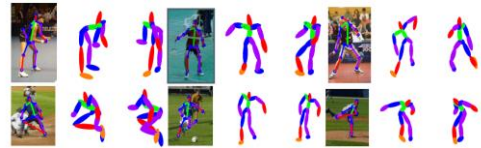


图 9。在利兹运动检测[16]自动部分真实的结果构成的数据集上几帧。尽管在检测中的异常值，我们的方法给出了有效的三维姿态的解释。请注意，脚没有检测到的图像，但我们的姿势的帮助下，他们的三维位置估计。

6. 结论

我们提出了构成条件关节角度限制，并制定一个事先的人为构成。我们认为，这是第一个一般优先考虑的联合限制造成的依赖。我们证明这之前限制无效是 2D 到 3D 人体姿态重构。此外，我们提供了一个新的算法，估计三维姿态，利用我们的先验。我们的方法显著优于目前的状态的艺术方法，定量和定性。我们以前的优化框架，可以应用到许多问题，在人类姿态估计超出了应用程序所描述的。在未来，我们会考虑在关节角度限制的运动树的兄弟姐妹的依赖关系。我们也对二维到三维姿态估计，进一步减少歧义的时空模型。未来的工作还应该考虑的时间依赖性关节限制，因为在运动过程中，身体可以达到的状态，可能不可能是静态的。

7.附录

我们模型的姿势之前的扩展躯干作为以下的贝叶斯网络，

$$p(\hat{\mathbf{X}}) = \prod_i p(\delta\hat{\mathbf{X}}'_i | \delta\hat{\mathbf{X}}'_I), \quad (10)$$

我是指关节指数在空调机组由图 4 所示的贝叶斯网络的定义 (B) 和 X' 我是通过连接他们的 3D 获得矢量坐标。我们认为联合高斯分布的一个共同的我和它的空调设置为，

$$\begin{pmatrix} \delta\hat{\mathbf{X}}'_i \\ \delta\hat{\mathbf{X}}'_I \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \delta\mu'_i \\ \delta\mu'_I \end{pmatrix}, \begin{pmatrix} \Sigma'_{ii} & \Sigma'_{iI} \\ \Sigma'_{Ii} & \Sigma'_{II} \end{pmatrix}\right), \quad (11)$$

那里的相对位姿 X' 满足，

$$\delta\hat{\mathbf{X}}' = \delta\mathbf{B}'\omega' + \delta\mu'. \quad (12)$$

给出了方程 (11) 的条件分布可以写成，

$$\left(\delta\hat{\mathbf{X}}'_i | \delta\hat{\mathbf{X}}'_I = \mathbf{a}\right) \sim \mathcal{N}\left(\delta\bar{\mu}'_i, \bar{\Sigma}'_{ii}\right),$$

，其中

$$\begin{aligned} \delta\bar{\mu}'_i &= \delta\mu'_i + \Sigma'_{iI}\Sigma'^{-1}_{II}(\mathbf{a} - \delta\mu'_I), \\ \bar{\Sigma}'_{ii} &= \Sigma'_{ii} - \Sigma'_{iI}\Sigma'^{-1}_{II}\Sigma'_{Ii}. \end{aligned} \quad (13)$$

上述姿势之前，可以结合方程 (12) 注意

$$\left(\mathbf{a} - \delta\mu'_I\right) = \delta\mathbf{B}'_I\omega' \quad \text{在这里，}$$

我由一行对应于点的行组成，利用这个关系，可以将完整的向量

$$\delta\bar{\mu}' - \delta\mu' = \mathbf{G}\omega', \quad (14)$$

$$\bar{\mu}_{\omega'} = \delta\mathbf{B}'^T\mathbf{G}\omega', \quad \text{and} \quad \bar{\Sigma}_{\omega'} = \delta\mathbf{B}'^T\bar{\Sigma}'\delta\mathbf{B}'^T, \quad (15)$$

$$\mathbf{x}' = s(\mathbf{I} \otimes \mathbf{R}_{1:2})(\mathbf{B}'\omega' + \mu'). \quad (16)$$

$$c(\omega') = \frac{1}{\alpha}\|\mathbf{A}\omega' - \mathbf{b}\|^2 + (\omega' - \bar{\mu}_{\omega'})^T \bar{\Sigma}_{\omega'}^{-1} (\omega' - \bar{\mu}_{\omega'}).$$

$$\left(\mathbf{A}^T\mathbf{A} + \alpha\mathbf{D}^T\bar{\Sigma}_{\omega'}^{-1}\mathbf{D}\right)\omega' = \mathbf{A}^T\mathbf{b}, \quad (17)$$

$$\sqrt{\alpha}\mathbf{A}_p\omega' = \mathbf{0}, \quad (18)$$

8.致谢

我们感谢安德列凯勒，索菲型，斯蒂芬 streuber，和 naureen 马哈茂德关节极限数据准备。我们与泰勒乔纳森的讨论中获益，彼得 gehler 脑桥，热拉尔摩尔，Varun jampani，

和该 murtza。我们也感谢 Varun Ramakrishna 和周小川为我们提供他们的源代码。

参考文献

- [1] <http://poseprior.is.tue.mpg.de/>.
- [2] U. S. N. Aeronautics and S. Administration. *NASA-STD-3000: Man-systems integration standards*. Number v. 3 in NASA-STD. National Aeronautics and Space Administration, 1995.
- [3] M. Andriluka, S. Roth, and B. Schiele. Monocular 3D pose estimation and tracking by detection. In *Computer Vision and Pattern Recognition*, pages 623–630, 2010.
- [4] C. Barron and I. Kakadiaris. Estimating anthropometry and pose from a single uncalibrated image. *Computer Vision and Image Understanding*, 81(3):269–284, March 2001.
- [5] C. BenAbdelkader and Y. Yacoob. Statistical estimation of human anthropometry from a single uncalibrated image. In *Methods, Applications, and Challenges in Computerassisted Criminal Investigations*, Studies in Computational Intelligence. Springer-Verlag, 2008.
- [6] L. Bourdev and J. Malik. Poselets: Body part detectors trained using 3D human pose annotations. In *International Conference on Computer Vision*, pages 1365–1372, Sept. 2009.
- [7] J. Chen, S. Nie, and Q. Ji. Data-free prior model for upper body pose estimation and tracking. *IEEE Trans. Image Proc.*, 22(12):4627–4639, Dec. 2013.
- [8] M. G. Choi, K. Yang, T. Igarashi, J. Mitani, and J. Lee. Retrieval and visualization of human motion data via stick

- figures.
Computer Graphics Forum, 31(7):2057–2065, 2012.
- [9] X. Fan, K. Zheng, Y. Zhou, and S. Wang. Pose locality constrained representation for 3d human pose reconstruction. In *Computer Vision–ECCV 2014*, pages 174–188. Springer, 2014.
- [10] K. Grochow, S. L. Martin, A. Hertzmann, and Z. Popović. Style-based inverse kinematics. *ACM Transactions on Graphics (TOG)*, 23(3):522–531, 2004.
- [11] P. Guan, A. Weiss, A. Balan, and M. J. Black. Estimating human shape and pose from a single image. In *Int. Conf. on Computer Vision, ICCV*, pages 1381–1388, Sept. 2009.
- [12] H. Hatze. A three-dimensional multivariate model of passive human joint torques and articular boundaries. *Clinical Biomechanics*, 12(2):128–135, 1997.
- [13] S. Hauberg, S. Sommer, and K. Pedersen. Gaussian-like spatial priors for articulated tracking. In K. Daniilidis, P. Maragos, and N. Paragios, editors, *Computer Vision ECCV 2010*, volume 6311 of *Lecture Notes in Computer Science*, pages 425–437. Springer Berlin Heidelberg, 2010.
- [14] L. Herda, R. Urtasun, and P. Fua. Hierarchical implicit surface joint limits for human body tracking. *Computer Vision and Image Understanding*, 99(2):189–209, 2005.
- [15] H. Jiang. 3D human pose reconstruction using millions of exemplars. In *Pattern Recognition (ICPR), 2010 20th International Conference on*, pages 1674–1677. IEEE, 2010.
- [16] M. Kiefel and P. Gehler. Human pose estimation with fields of parts. In D. Fleet, T. Pajdla, B. Schiele, and T. Tuytelaars, editors, *Computer Vision – ECCV 2014*, volume 8693 of *Lecture Notes in Computer Science*, pages 331–346. Springer International Publishing, Sept. 2014.
- [17] T. Kodek and M. Munich. Identifying shoulder and elbow passive moments and muscle contributions. In *IEEE Int. Conf. on Intelligent Robots and Systems*, volume 2, pages 1391–1396, 2002.
- [18] H. J. Lee and Z. Chen. Determination of 3D human body postures from a single view. *Computer Vision Graphics and Image Processing*, 30(2):148–168, 1985.
- [19] J. Lin, T. Igarashi, J. Mitani, M. Liao, and Y. He. A sketching interface for sitting pose design in the virtual environment. *Visualization and Computer Graphics, IEEE Transactions on*, 18(11):1979–1991, 2012.
- [20] G. Mori and J. Malik. Recovering 3D human body configurations using shape contexts. *Pattern Analysis and Machine Intelligence*, 28(7):1052–1062, 2006.
- [21] V. Parameswaran and R. Chellappa. View independent human body pose estimation from a single perspective image. In *Computer Vision and Pattern Recognition*, pages 16–22, 2004.
- [22] I. Radwan, A. Dhall, and R. Goecke. Monocular image 3D human pose estimation under self-occlusion. In *International Conference on Computer Vision*, pages 1888–1895, 2013.
- [23] V. Ramakrishna, T. Kanade, and Y. Sheikh. Reconstructing 3D human pose from 2D image landmarks. *European Conference on Computer Vision*, pages 573–586, 2012.

- [24] J. M. Rehg, D. D. Morris, and T. Kanade. Ambiguities in visual tracking of articulated objects using two-and three-dimensional models. *The International Journal of Robotics Research*, 22(6):393–418, 2003.
- [25] G. Rogez, J. Rihan, C. Orrite-Uruñuela, and P. H. Torr. Fast human pose detection using randomized hierarchical cascades of rejectors. *International Journal of Computer Vision*, 99(1):25–52, 2012.
- [26] M. Schünke, E. Schulte, and U. Schumacher. *Prometheus: Allgemeine Anatomie und Bewegungssystem : LernAtlas der Anatomie*. Prometheus LernAtlas der Anatomie. Thieme, 2005.
- [27] C. Sminchisescu and B. Triggs. Building roadmaps of local minima of visual models. In *European Conference on Computer Vision*, volume 1, pages 566–582, Copenhagen, 2002.
- [28] C. Sminchisescu and B. Triggs. Estimating articulated human motion with covariance scaled sampling. *The International Journal of Robotics Research*, 22(6):371–391, 2003.
- [29] C. J. Taylor. Reconstruction of articulated objects from point correspondences in a single uncalibrated image. *Computer Vision and Image Understanding*, 80(10):349–363, October 2000.
- [30] J. Valmadre and S. Lucey. Deterministic 3D human pose estimation using rigid structure. In *European Conference on Computer Vision*, pages 467–480. Springer, 2010.
- [31] X. K. Wei and J. Chai. Modeling 3D human poses from uncalibrated monocular images. In *International Conference on Computer Vision*, pages 1873–1880, 2009.
- [32] X. K. Wei and J. Chai. Intuitive interactive human-character posing with millions of example poses. *Computer Graphics and Applications, IEEE*, 31(4):78–88, 2011.
- [33] I. Yoo, J. Vanek, M. Nizovtseva, N. Adamo-Villani, and B. Benes. Sketching human character animations by composing sequences from large motion database. *The Visual Computer*, 30(2):213–227, 2014.
- [34] T.-H. Yu, T.-K. Kim, and R. Cipolla. Unconstrained monocular 3D human pose estimation by action detection and crossmodality regression forest. In *Computer Vision and Pattern Recognition*, pages 3642–3649, 2013.