

指导教师：杨涛

提交时间：2015/03/21

# CVPR2015 Paper Translation

No: 01

姓名：曹悦

学号：2013302565

班号：10011305

# 结合运动和骨架信息的复杂运动结构的无监督学习

## 摘要:

在这片论文中，对于单视图图像序列中的复杂相关的物体，我们提出一种新颖的无监督的运动结构学习框架。和先前对相对简单的关节估计得到运动信息的方法不同，我们通过一个逐次迭代的合并过程，可以任意产生复杂的运动骨骼拓扑结构的。反复的合并过程是由骨骼距离函数控制的，这个骨骼距离函数产生于稀疏点中的一种新颖的物体边界生成的方法。我们的主演贡献可以归纳如下：

- (i) 结合运动和骨架信息的复杂运动结构的无监督学习。
- (ii) 反复的由精细到粗糙的合并策略，为可适应的运动分割和结构平滑。
- (iii) 从稀疏的特征点中进行骨骼估计。
- (iv) 一个新的精细的物体数据集，这个数据集包含多极复杂的标注数据。

我们的实验显示，提议的方法在质量和数量上都优于当前最高水平的方法。

## 1. 介绍

此外，大多数已有的运动结构的

学习潜在的组织缜密的物体的运动结构是计算机视觉和机器人学的一项活跃的研究。准确有效的运动结构估测对于很多高层次的工作，例如物体运动识别，人体动作识别，机器人身体计划学习的操纵，组织缜密的物体的操纵，等方面都很有益处。在这片论文中，我们只着眼于利用兴趣点追踪 2D 位置的复杂铰接的运动结构学习特例。

许多从 2D 追踪数据中恢复清晰结构的算法向我们展示了清晰的运动类型（折叠，旋转和平移）的自动检测和运动链的构建。然而他们仅仅应用于相对简单的关节。我们的目标是找到具有精细运动能力的任意对象的运动结构，这种运动能力可以在简单的结构和复杂的结构之间变化，比如人的手。

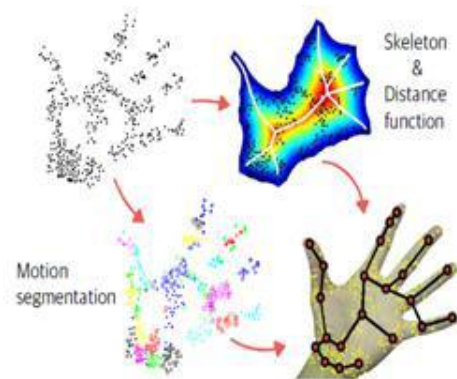


图 1. 提议的框架能从运动和骨架 结合的信息中可靠地学习到复杂物体的潜在运动结构。

生成方法仅仅使用运动信息。这些技

术忽略了实施拓扑或运动学约束的全局优化步骤，这样可以生产难以置信的结构作为输出。另一方面，源于形状信息的清晰结构评估方法已经提出。通常，被估计结构是一种的骨骼结构，这种骨骼结构表现了主体的中轴并且暗示了其拓扑性质。但是这种估计方法不能代表运动结构。

在这篇论文中，我们为源于 2D 特征点轨迹的复杂精细的运动结构估计提出一种新颖的框架。我们结合运动和骨骼信息来生成复杂合理的运动结构（见图 1）。我们假设一个清晰的物体是由一组精确的分割组成的，结构代表了分割之间的联系。当关节很复杂并且输入数据嘈杂时，提前估计分割的段数是一件很难的事情。所以我们介绍一种精细到粗略的策略，这种策略由骨骼拓扑结构和运动相似性作为引导，执行反复的合并和被分割部分之间的平滑化。为了从稀疏的点中得到骨骼距离函数，我们提出一个新颖的物体边界生成方法。最终，我们的方法不需要任何物体的运动分割段数和物体类型这类的先验知识，并且学习到的结构能够充分代表了利用物体的骨骼拓扑结构的复杂关节。我们的实验显示，提议的方法在质量和数量上都优于当前最

高水平的方法。

## 2. 相关工作

一些关于运动物体清晰结构的产生方法已经被提出。RGBD 基于传感器的人类/手骨骼估计的方法也已经成功提出。然而，这些方法仅仅是针对需要计算化先前训练步骤的特定目标骨架。结果是典型的骨骼而不是运动结构。RGB 相机也仍更广泛的用于不同的应用，所以很有必要开发一个从单视图中得到 2D 序列的好的算法。因为我们的方法依赖于这种类型的输入数据，我们将会主要讨论只用二维特征跟踪的相关工作（不采用深度信息）。三大主要的类型可以在文学上加以区别，运动分割和基于因式分解的方法，概率图形模型的方法和基于代价函数的优化方法。

运动分割和基于因式分解的方法可能是清晰重建的最受欢迎的方法。不同的运动分割的方法已经被提出，比如子空间拟合（GPCA）、子空间聚类和基于多视图的方法。GPCA 被广泛的用于运动分割有关的论文，但是这种方法需要提前知道运动分割的段数。同时，这种方法也不能用于超过几个的子空间，因为所需样本的数量与子空间的数量呈指数级增长。最近，Jung 等人提出了一种新颖

的基于随机投票的刚性运动分割的算法。他们展示了这种算法能够实现合理时间内嘈杂环境下的运动分割实施的当前最高水平。然而，这种算法也需要一个确切数量的运动集群来作为之前的良好表现，这就使得这种算法运用在复杂视频上成为一件很难的事情。

Tresadern, Reid, Yan 和 Pollefeys 为关节式的物体改进了因子分解方法，说明了特性跟踪的矩阵的秩能表明关节的类型。这种算法对于分割相互依赖的运动非常有效，但却不能解决物体间高度密切连接的情况。此外，Yan 和 Pollefeys 通过将关节运动模拟成交叉的运动子空间的方法来预估运动链。联合体的位置可以从连接部分的运动子空间的交叉点获得。这个算法高度依赖于正确检测的轨迹排名，因此对噪声干扰敏感。每一步还有很多调节参数。总体来说，这种算法很难运用到复杂关节上。Jacquet 等人提出了基于线性子空间的相对变换分换法，但是它着重于检测两个受限运动部分之间关节运动的类型。Ross 等人提出了概率图形化模型的方法来从 2D 特征追踪中学习铰接的结构。他们能够自适应地找出关节的数目和它们之间的

连接，但他们的方法对先验知识是敏感的，而且很难从一个比较差的初始分割中恢复出来。同时这种方法难以避免局部极小问题。Strum 等人同样使用了概率的方法来学习运动关节，特别是针对机器人视觉应用程序，身体计划学习和物体操纵。为了得到无噪声的输入数据他们需要基准标记每个对象部分，并且运动分割的段数作为先验知识被给出。Pillai 等人提出了一个无标记的稀疏特性，基于追踪的关节学习，这种学习不需要先验的物体模型。然而他们需要 RGB-D 特征数据，并且不能处理并发关节运动。

Fayad 等人提出了一个用于同时分割和三维重建，基于代价函数的优化方法。我们无需提前知道骨架结构和运动分割段的数量。他们把一组点跟踪分解成重叠的刚体，结构关节来自重叠的区域。然而，将不同段之间的部分重叠，由此产生的段数平滑，像这样复杂的结构是很难正确地估计的。

### 3. 方法论

我们的目标是通过运动和骨架信息来生成复杂的运动结构，同时在复杂并发的动作下能做到准确和合理。最后，我们只使用 2D 轨迹来学

习(假设一个目标对象存在于场景中)。为了提取每个刚性运动部分,我们采用结果最好的运动分割方法:随机投票。为了从二维稀疏的特征点中估计骨架信息,用一个类数据描述方法(支持向量数据描述)来生成对象轮廓。为了骨架提取,我们利用基于距离函数的轮廓修剪的骨架。

所设想的框架的整体概念如图2所示。在3.1节中,我们定义了符号。3.2节中我们讨论自适应的运动分割如何生成。接下来,在3.3节中,我们讨论如何生成物体的边界以及如何从稀疏的特征点中生成骨架距离函数。最后,在3.4节中,我们讨论运动结构的生成,以及利用生成的骨架和运动信息的平滑算法。

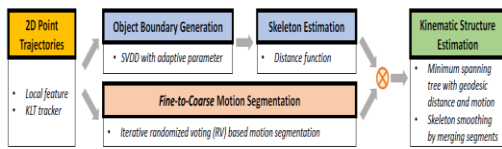


图2.所设想方法的总体流程

### 3.1 符号

二维特征点用  $x_i$  来表示,  $i=1;\dots;N$ 。  $N$  全部的点数。点集  $X$  被定义为  $X = \{x_1, x_2, \dots, x_N\}$ ,  $x_i$  用齐次坐标来表示。轨迹用  $x_i^f$  来表示,  $f=1, \dots, F$  作为顺序索引,  $F$  作为输入视频的帧数。我们在处理轨迹数

据,所以我们用上角标来表示顺序索引。为了表达运动段数,我们用  $S_k$  表示属于  $k^{th}$  段的不相交的点集,  $k=1, \dots, c$ ,  $c$  表示分割的总段数,  $X = S_1 \cup S_2 \cup \dots \cup S_c$ 。  $y_k$  代表  $S_k$  分割段的中心位置,是通过计算点的平均值得到的。我们用  $\Omega$  表示物体的区域,用  $\partial\Omega$  表示其边界。术语对象边界和轮廓是可交替使用的。

### 3.2 运动分割

精准的估计动作分割的段数是很难的,尤其是当动作高度复杂并且输入数据含有噪声时。为了解决这些复杂的情况,我们提出一个迭代的 fine-to-coarse 推理策略自适应地估计初始化运动分割的数量上限。我们使用随机投票(RV)的方法,这种方法是目前为止执行效果最好的,对于运动分割的噪声干扰很健壮,但是需要提前知道分割段数。

因为 RV 利用迭代的基本矩阵估计,每段内要有八个点才能开始这个算法。因此,我们将初始段数估计为  $\hat{c}^{init} = \lceil N/8 \rceil$ 。每段都初始化地分配了多余八个点,有些段还是会因为随机投票过程得到少于8个点。在合成段,少于8个点( $c < 8$ )的段数被统计出来,

段数重置为  $\hat{c}^{t+1} = \hat{c}^t - c_{<8}$ 。然后我们用减少后的段数  $\hat{c}^{t+1}$  反复执行 RV 算法，直至所有段都含有多余 8 个点 ( $c_{<8} = 0$ )。反复迭代的 fine-to-coarse 分割过程如算法 1 描述。

---

**Algorithm 1** Fine-to-coarse Motion Segmentation

---

**Input:**  $x_i, i = 1, \dots, N$  ▷ Point trajectories  
**Output:**  $S_k, k = 1, \dots, \hat{c}$

- 1:  $t \leftarrow 1$
- 2:  $\hat{c}^t \leftarrow \lfloor N/8 \rfloor$  ▷ Initialise the number of segments
- 3: **repeat**
- 4:      $S_k^t \leftarrow \text{RV motion segmentation}(\{x_i\}_{i=1}^N, \hat{c}^t)$
- 5:      $c_{<8} \leftarrow 0$
- 6:     **for**  $k = 1, \dots, \hat{c}^t$  **do**
- 7:         **if**  $|S_k^t| < 8$  **then**
- 8:              $c_{<8} \leftarrow c_{<8} + 1$
- 9:      $\hat{c}^{t+1} \leftarrow \hat{c}^t - c_{<8}$
- 10:     $t \leftarrow t + 1$
- 11: **until**  $c_{<8} = 0$

---

### 3.3 从稀疏的特征点中提取框架

使用骨架作为一个物体抽象的有很大的好处。它可以以低维形式包含必要的形状特征，也可以包含一个物体的拓扑结构。已经有无数从目标物体的二进制轮廓图像来估计骨架的算法。依靠背景减法，微分的整体特性或人体检测技术，轮廓可以从 RGB 图像中提取出来。不幸的是，这些方法对于从稀疏的二维特征点中产生轮廓和骨架并不是很合适。

#### 3.3.1 物体边界生成

我们现在提出一种基于支持向量数据描述 (SVDD)，从稀疏特征点

$X^f$  自适应生成物体边界 ( $\partial\Omega$ ) 的方法。SVDD 试图找到一个覆盖所有目标数据，将多余空间最小化的严格的描述。我们把描述边界当做对象边界 ( $\partial\Omega$ ) 点。

为了制定多余空间最小化的覆盖描述，描述的形状是一个最小体积的球体。结果，它获得一个球体状的，包含所有目标数据的封闭边界 (一个超球面)。类推到支持向量机，通过运用核函数，边界可以很灵活。球的特点是有一个中心  $a$  和半径  $R > 0$ 。球体的体积通过将  $R^2$  最小化从而达到最小化。将  $R^2$  最小化的目标函数有松弛变量  $\xi_i \geq 0$ ，惩罚参数  $C$  被定义为：

$$F(R, a) = R^2 + C \sum_i \xi_i \quad (1)$$

受到以下限制：

$$\|x_i - a\|^2 \leq R^2 + \xi_i, \quad \xi_i \geq 0 \quad \forall i \quad (2)$$

类推到支持向量机，方程式 (1) 和方程式 (2) 可以通过引入拉格朗日乘数  $\alpha_i \geq 0$  得到合并；

$$L = \sum_i \alpha_i (x_i \cdot x_i) - \sum_{i,j} \alpha_i \alpha_j (x_i \cdot x_j) \quad (3)$$

$$s.t. 0 \leq \alpha_i \leq C$$

此外，非线性边界可以通过使用核函数  $K(x_i, x_j) = \Phi(x_i) \cdot \Phi(x_j)$  更换

内积  $(x_i, x_j)$  来找到  $(\Phi$  是数据向高维特征空间的隐式映射, 我们指示一个内核参数  $\delta$ )。我们使用一个指数径向的基函数内核  $K(x_i, x_j) = \exp(-\|x_i - x_j\|/2\delta^2)$ , 它能产生一个严格的分段线性的解决方案。

然而, 不同于两类支持向量机, 选择一个合适的控制边界密合的内

核参数  $\delta$  是很难的, 因为数据中没有野值。图三展示了不同内核参数生成的边界。有一些解决内核参数选择问题的方法。然而, 理论分析方法给的选择范围过于宽松, 遗传算法的启发式求解太耗时。在人工离群值生成方法中, 生成良好的离群值是一个问题。

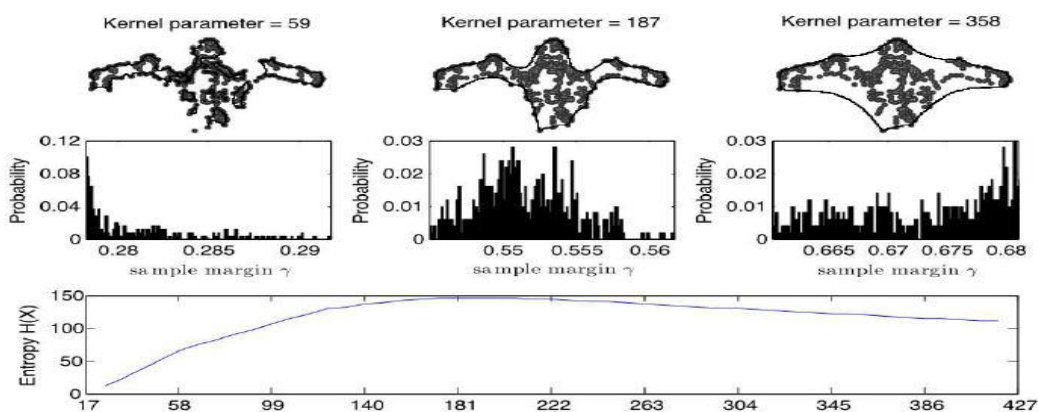


图 3. 不同内核参数的物体边界生成结果。一个小的参数值产生了超过估计, 边界区域分离的结果。一个大的参数值产生了低于估计值的边界结果。样本边缘的分布显示在每个内核值的中间。正如我们可以看到, 最合适的边界由给出最大熵的内核值产生。

在这项工作方面, 我们介绍一种新颖的, 利用样本边缘进行最佳内核参数选择的方法。样品边缘是一个从基准到超平面的距离, 它穿越了内核空间中的超球面的中心。样本边缘反映内核空间中图像数据的分布, 并且对于每个数据点  $x_i$  能通过公式

$$\gamma(x_i) = \frac{a \cdot \Phi(x_i)}{\|a\|} \quad (0 \leq \gamma(x_i) \leq 1 \quad \forall i)$$

计算出来,  $a = \sum_i \alpha_i \Phi(x_i)$ 。每个样本边缘反映了规范化的中心和超球面边界之间的相对位置, 所以不同的内核参数给出不同的样品边缘分布以及不同描述边界 (如图 3 所示)。

在这片论文中, 通过计算样本样

本边缘分布的熵值，我们为内核参数选择提出一个标准。如果描述过度适合，样本边缘将会向超球面

边界方向分布。如果边界不够充分拟合，则分布偏向中心。通过寻找一个最大熵的内核参数，我们避免过度拟合与不够充分拟合的情况。而且，根据最大熵原则，如果没有关于分布的可用先验知识，最大熵的概率分布最能代表当前知识的状态。所以当前帧特征点  $X^f$  的最优内核参数  $\delta^f$  可以通过如下的式子估计：

$$\mathcal{X} = \left\{ q \mid \forall q \in I, \|q - a\|^2 = 1 - 2 \sum_i \alpha_i K(q, x_i) + \sum_{i,j} \alpha_i \alpha_j K(x_i, x_j) = R^2 \right\} \quad (5)$$

为了定量地度量用新标准生成边界的优点，我们遵循离群值均匀生成分布在目标数据周围。使用离群值，损失函数平衡分类和扰动抑制性能，被用作衡量是否合适的标准。最终，我们发现类似但 3.2 ~ 7.8 倍快的描述。

### 3.3.2 骨架距离函数生成

物体的骨架  $\Upsilon(\Omega)$ ，被定义为包含在一个物体  $\Omega$  中，一组最大圆的全部中心点，是一个物体的中轴。它可以用公式表示为，源于至少两个边界点，相同距离的点的轨迹：

$$\delta^f = \arg \max_{\delta} H(\gamma(X^f)) = \arg \max_{\delta} \sum_i -p_i \log(p_i) \quad (4)$$

$H$  代表熵值， $p_i$  表示概率分布，

$$p_i = P_r(\gamma(x_i^f))。$$

物体边界  $\mathcal{X}\Omega$  可以被认为是一组图像空间  $I$  中，位于距离超球面中心  $a$  同样的距离（半径  $R$ ）的全部点。

它能由选择好的内核参数  $\delta$  生成：

$$\Upsilon(\Omega) = \left\{ p \in \Omega \mid \exists q, r \in \mathcal{X}\Omega, q \neq r : \text{dist}(p, q) = \text{dist}(p, r) \right\} \quad (6)$$

骨架既包含形状特征，又包含原始对象的拓扑结构。作为优秀的骨架代表，距离转换被定义为返回每个点  $p$  离边界最近的距离的函数。利用得到的物体边界， $\Omega$  的距离函数 ( $\Psi(p)$ ) 定义如下： $\Psi(p) = \min_{q \in \mathcal{X}\Omega} (\text{dist}(p, q))$

$$(7)$$

对于的每个点。距离度量通常使用欧几里得距离  $\text{dist}(p, q) = \|p - q\|^2$ 。使用的距离函数是由吸引力的，因为其计算相对简单，并且骨架可以生成脊距离函数。



### 3.4 运动结构估计和结构平滑

在这个章节中，我们提出如何用运动节段( $S$ )和骨架距离函数( $\Psi$ )的结果，生成关节式物体的运动结构。我们假设运动结构并不像大多数复杂对象那样循环。我们利用一个图解模型  $G=(V,E)$  来决定运动段之间的拓扑连接。在一幅完整地图中，所有的运动段的中心  $y_1, \dots, y_c$  被看做节点  $V$ 。 $y_k$  和  $y_l$  段之间的接近程度定义如下：

$$E(y_k, y_l) = \underset{f \in F}{\text{median}} \left\{ \left( \zeta(y_k^f - y_l^f) \times \|y_k - y_l\| \right) \right\} \quad (8)$$

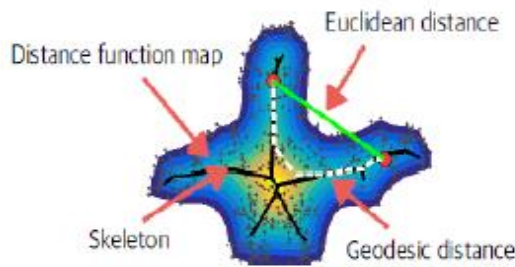
它结合骨架距离转换中的测地距离和移动速度的差异。为了对所有帧中两段之间的临近性程度进行最终估

计，我们采用中位数来增强对离群值的鲁棒性。

给出了距离函数  $\Psi$ ，两点  $p$  和  $q$  之间的测地距离定义如下：

$$\zeta(p, q; \Psi^f) = \min_{\Gamma \in P_{p,q}} \sum_{n=1}^{k(\Gamma)} \frac{1}{\Psi^f(P_n)} \quad (9)$$

$\Gamma$  是一条连通两点的路径， $P_{p,q}$  是一组所有可能的路径。所以公式 (9) 通过图 4 的骨骼拓扑路径，定义了目标区域中两个点的最小距离。



图四. 两点之间的测地距离和欧几里得距离。白色的虚线是测地距离，绿色的实线是几里得距离。黑色的实线是物体的骨架。测地距离代表延骨架的最小距离。

提出的临近性度量将拓扑结构上独立的和以不同速度移动的段分离开来。两条边权重较小的段有很大

的可能性连接起来。我们以物体的运动结构来生成图的最小生成树。

然而，最初生成的结构高度弯

曲，因为许多小的运动段偏离了中轴。所以我们进一步通过反复合并的过程进行结构平滑化，这个过程是由骨架距离函数指导的。如果  $S_k$  段很大程度上偏离中轴，那么就很小(例如  $\Psi(y_k) < \tau$ )。我们设定了阈值  $\tau$  作为骨架  $\Upsilon$  最小的距离函数值； $\tau = \min \Psi(\Upsilon)$ 。偏离的  $S_k$  合并到周围  $\Psi$  值更大的段 ( $\hat{c} = \hat{c} - 1$ )，然后我们重建结构直到所有段中心靠近骨架(见图 5)。

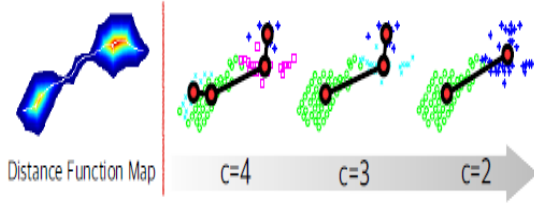


图 5. 通过反复的段合并实现结构平滑

#### 4. 实验

**数据集** 该提议的方法已通过诸如‘手臂’、‘玩具’、‘舞蹈’这样著名的序列进行了评估。但常见的数据序列相对简单。我们引进了具有挑战性的，由复杂动作组成的序列。我们已经试图避免严重阻碍的运动，因为二维特征点跟踪器被遮挡时不能保持跟踪。然而，新的序列仍然包含多样化复杂的动作，比如连接、并发、旋转、仿射和缩放。我们总结数据集的

属性如表一。因为在一帧内有很多运动，我们要密集地提取特征点，不能遗漏微小的动作。我们为每一个动作段手动地标上了标注。

Dataset	# of seg.	# of points	# of frames	motion concur.
arm	2	77	30	no
toy	3	93	69	no
dance	6	236	60	yes
<b>robot arm</b>	8	144	737	yes
<b>iCub body</b>	7	573	250	yes
<b>iCub hand</b>	8	154	280	yes
<b>Baxter</b>	11	484	454	yes
<b>human hand</b>	20	450	634	yes

表一. 数据集的属性。新介绍的数据集更具有挑战性，因为他们由并发的和复杂的运动组成。

#### 4.1 自身对照

我们已经进行了不同的实验来证明所提的框架。我们的方法是基于随机投票的，所以不同的尝试结果并不完全一样。所有的实验统计是从一百次尝试中得到的。

为了定量地评估效果，我们如下定义了错误度量：

$$error = \frac{1}{c \cdot F} \sum_{k=1}^{\hat{c}, F} \left( \min_{g=1 \dots c_{GT}} \|y_k^f - y_g^f\| \right) \times \left( 1 + \frac{|\hat{c} - c_{GT}|}{c_{GT}} \right) \quad (10)$$

$c_{GT}$  和  $y_k$  分别代表真实状态数据的总数和他们各自的中心。有了这一度量，我们可以考虑结构复杂性的差异每一部分的空间偏差。

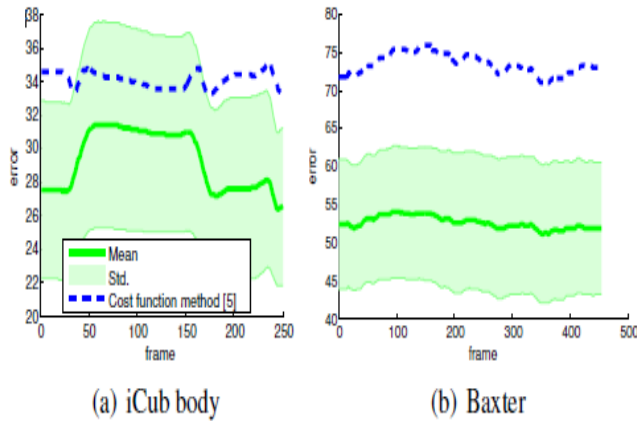


图 7. 帧的错误水平比较

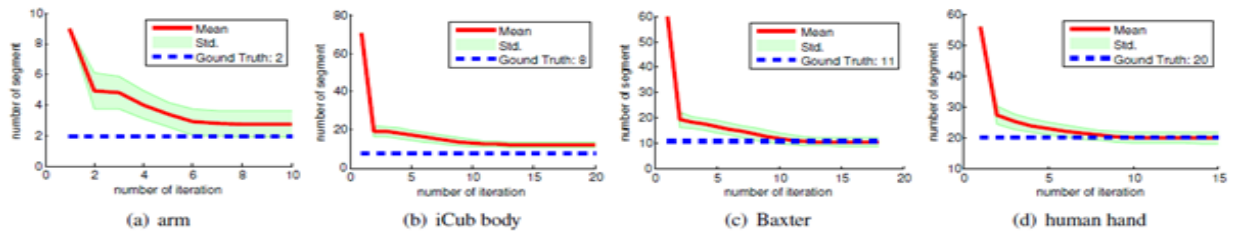


图 6. 运动段数紧密收敛于通过反复合并过程所得的真实状态数据

首先，我们验证所提的 fine-to-coarse 反复合并过程是否能找出正确的分割段数。正如图 6 所示，迭代所得结果段数紧密收敛于真实状态数据。此外，我们还测量了帧的误差变化。通过这些实验，我们可以测试评估段的运动准确性。正如我们图 7 所看到的，我们的方法比其他方法发现更多的准确运动点。

#### 4 4.2 与先进水平比较

我们将提议的方法和当前最先

进的方法进行了比较。我们实现了基于随机抽样一致性和的因子分解的方法。第三种我们进行比较的方法是基于代价函数的优化方法，这是当前效果最好的方法。所有的方法都像各自的论文中描述的那样使用工具箱得到了实现。

为了得到合理的结果，我们为每个数据序列手动地调整了一些参数，比如动作段的数目和等级检测参数。J. Fayad, C. Russell 和 L. Agapito 通过平均两段之间的交叉点找到结

构节点，并把交叉点连接了起来。然而，没有更多关于终止节点（无重叠）和连接过程的详尽描述。所以我们手动选择端节点，并应用最小生成树的方法进行连接。为了比较，我们在图 9 中也展示了没有人工参与终止关节选择的结构。

要强调的是，我们不会对任意特定序列的参数的粗略初始值作特地调整。所有的比较通过在固定参数下完全自适应算法（无需人工干预）得到的。

在表 2 和图 8 中，我们展示了平均的关节错误。我们的方法对于简单关节，实现了相比降低的平均错误；对于复杂运动序列极大地优于其他当前先进方法。另外，在图 9 中，我们展示了一些定量的成果。

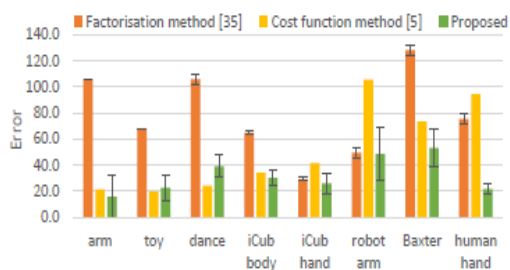


图 8. 定量的错误比较图

Dataset	RANSAC method [34]	Factorisation method [35]	Cost function method [5]	Proposed
arm	561.2 (176.4)	105.8 (0.0)	<b>21.5</b>	<b>15.7</b> (16.2)
toy	2357.0 (0.0)	68.2 (0.0)	<b>20.0</b>	22.6 (9.4)
dance	3041.1 (320.9)	105.8 (3.5)	<b>24.3</b>	39.3 (8.2)
iCub body	6357.4 (1482.6)	65.0 (1.8)	<b>34.1</b>	<b>30.4</b> (5.9)
iCub hand	975.6 (0.0)	29.2 (1.3)	41.8	<b>26.2</b> (7.9)
robot arm	1305.4 (120.1)	49.5 (3.5)	105.3	<b>48.5</b> (19.9)
Baxter	6606.9 (1108.7)	127.9 (4.3)	73.4	<b>53.2</b> (14.4)
human hand	3127.2 (226.2)	75.5 (3.7)	94.1	<b>21.9</b> (3.5)

表 2. 估计的与先进方法的关节精确度比较。所有的值都来自一百次尝试，除了[5]给出了一个最优化方法得到的固定结果。以上数据是一个平均数，括号里的数据代表标准偏差。

基于随机抽样一致性算法的方法，和基于因子分解的方法对噪声和参数设置很敏感，并且噪声影响随着运动复杂性的增大而增大。在[5]中，代价函数权衡了模型整体复杂性和局部运动误差，当结构简单时执行效果很好。然而，当运动复杂性增加时，相对于一个精确详细的结构，它会找到一个适中的结构。此外，代价函数使得相关运动段产生重叠，这样输出的结果可能离我们正常预估的差很远。我们的 fine-to-coarse 过程能找到详细的结构，骨架信息减少噪声影响。所以我们学习到的结构更加详细可靠。而且，我们的方法比平均情况快 1.8 倍、需要注意的是，我们的方法和[5]都是在 matlab 上实现的，单线程，没有任何 CPU/GPU 并程序。

## 5. 总结和未来的工作

在这篇论文中，我们介绍了一种新颖的运动结构评估框架，它可以合理地表示复杂的结构。我们已经表明，将先进的运动分割应用到骨架信息指导的合并过程中，遇到的挑战可以有效地解决。我们采用 fine-to-coast 凝结的合并计划，即我们开始 over-segmented 的运动分割。已经提出了用稀疏特征点生成物体轮廓，骨架距离函数就是从轮廓中产生的。反过来，结构学习期间，我们利用了运动段和骨架的距离函数，考虑了运动相似和拓扑学，建立一个连接树。我们的方法通过了公共数据集和我们新的具有挑战性的复杂关节运动数据集。先前的工作需要手动的干预（比如，段数和端节点位置），而我们无需调优参数，自适应地就能找到运动部分和骨架。因此除了准确运动关节检测的结果，我们可以获得一个高度可信的典型代表结构来促进进一步的任务，比如对象操作、物体识别、机器人身体方案理解等等。该方法在处理遮挡时有一个限制，因为当发生阻塞时二维特征点跟踪失败。未来我们计划使用 RGB-D 相机用于追踪，物体分割和 3D 轮廓生成所提的轮廓生成方法也可以作为运动

分割的先行知识。

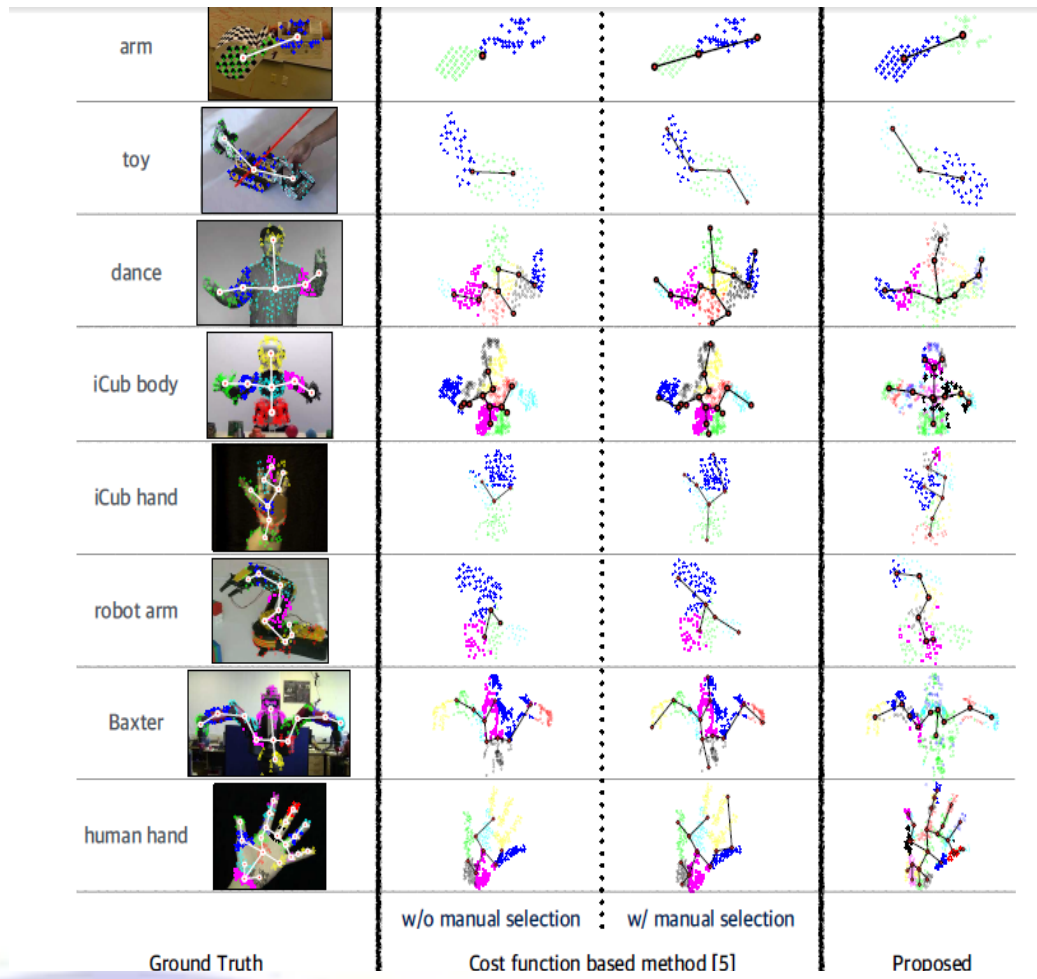


图 9. 基于代价函数的方法需要人工选择端节点。用所提议的方法学习到的结构更加详细、可信。