

# 计算机视觉—目标跟踪

申抒含

中国科学院自动化研究所  
模式识别国家重点实验室

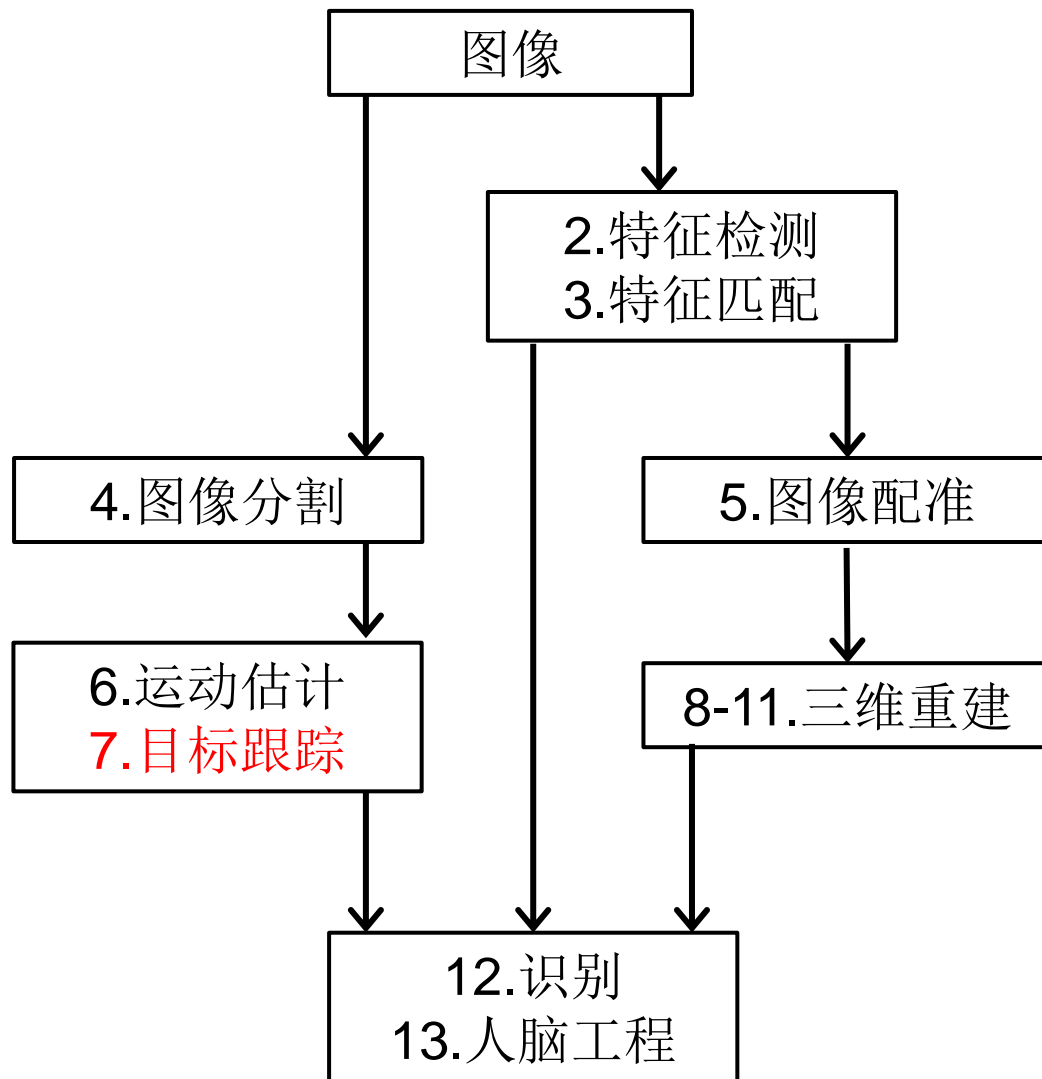


**Robot Vision Group**

National Laboratory of Pattern Recognition

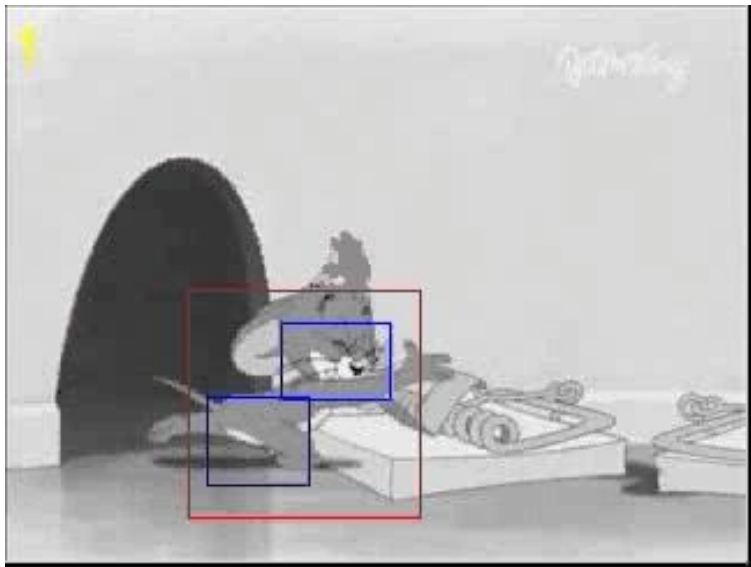
Institute of Automation, Chinese Academy of Sciences

# 计算机视觉课程结构图



# 什么是目标跟踪

目标跟踪：确定图像序列中感兴趣物体的位置和运动



目标跟踪的难点：光照、视角、形状、遮挡、摄像机运动、形变、运动模糊等。

# 目标跟踪的应用

目标跟踪有的广泛的应用领域：

- 运动分析；
- 运动医学；
- 动画特效；
- 智能监控；
- 机器人；
- 人机交互；
- 游戏；
- 行为理解；
- 



# 目标跟踪的分类



单目标跟踪



多目标跟踪

# 目标跟踪的分类



单摄像机跟踪



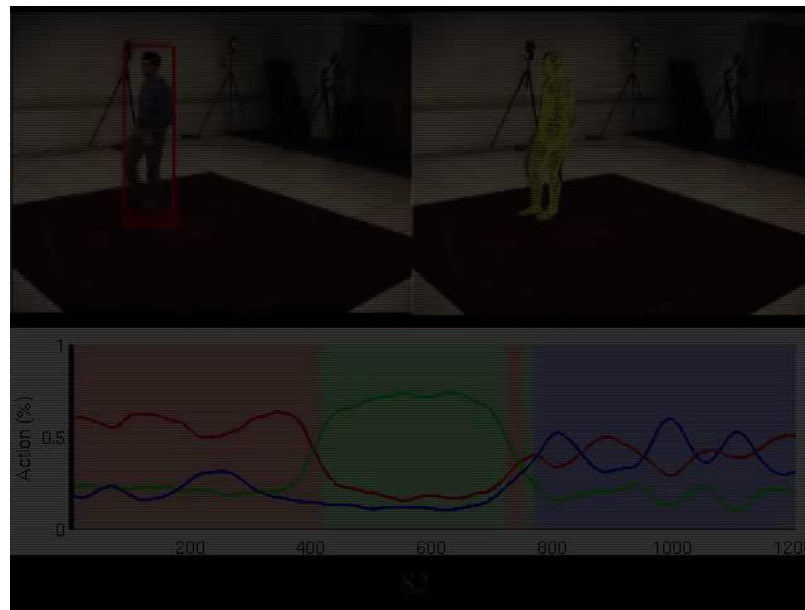
多摄像机跟踪



# 目标跟踪的分类

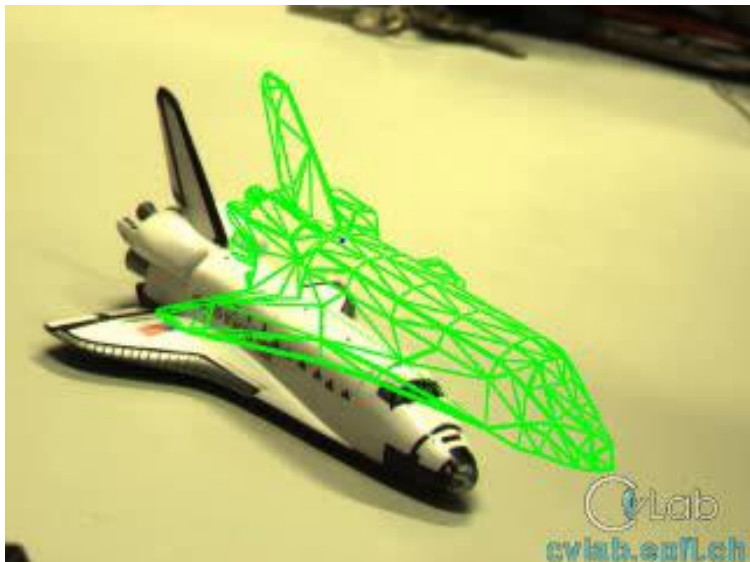


2D跟踪

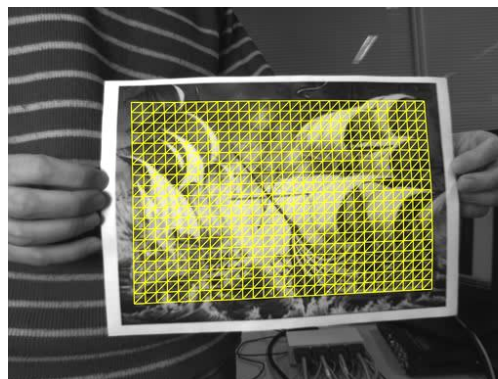


3D跟踪

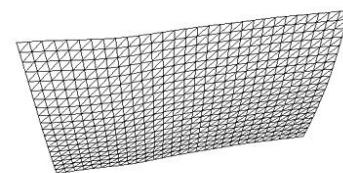
# 目标跟踪的分类



刚体跟踪

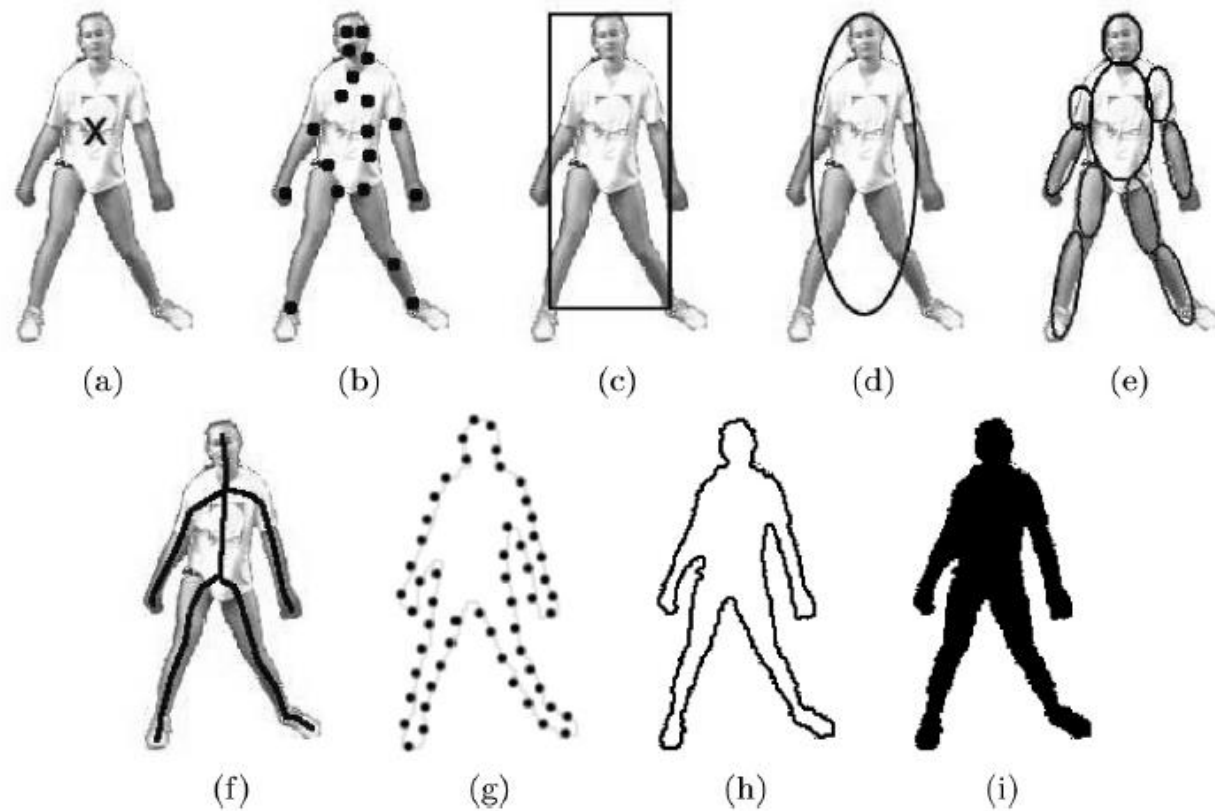


非刚体跟踪





# 运动目标的表达方法



**Fig. 1.** Object representations. (a) Centroid, (b) multiple points, (c) rectangular patch, (d) elliptical patch, (e) part-based multiple patches, (f) object skeleton, (g) complete object contour, (h) control points on object contour, (i) object silhouette.

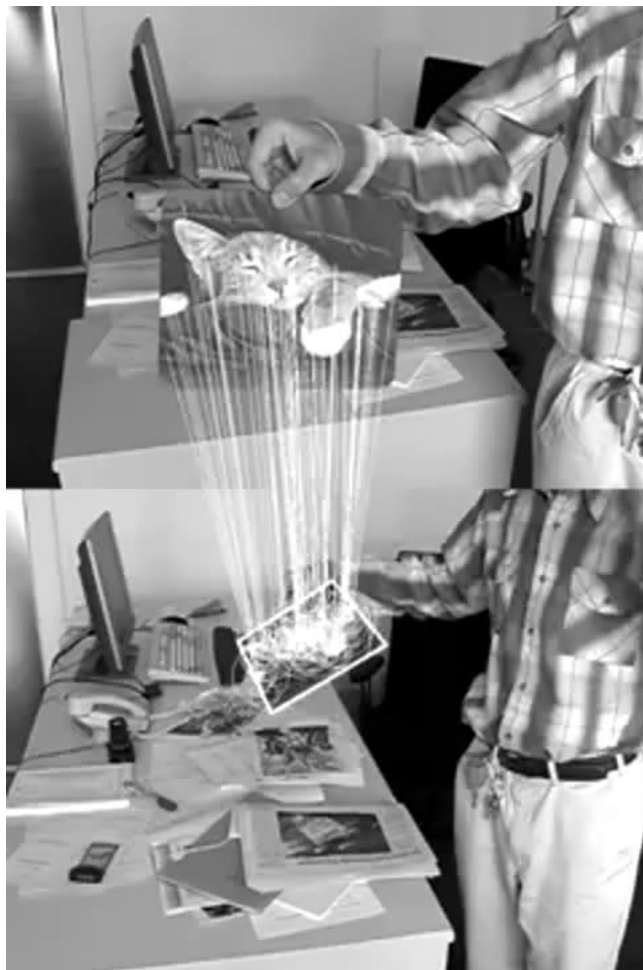
# 运动目标的表达方法

基于区域的跟踪：



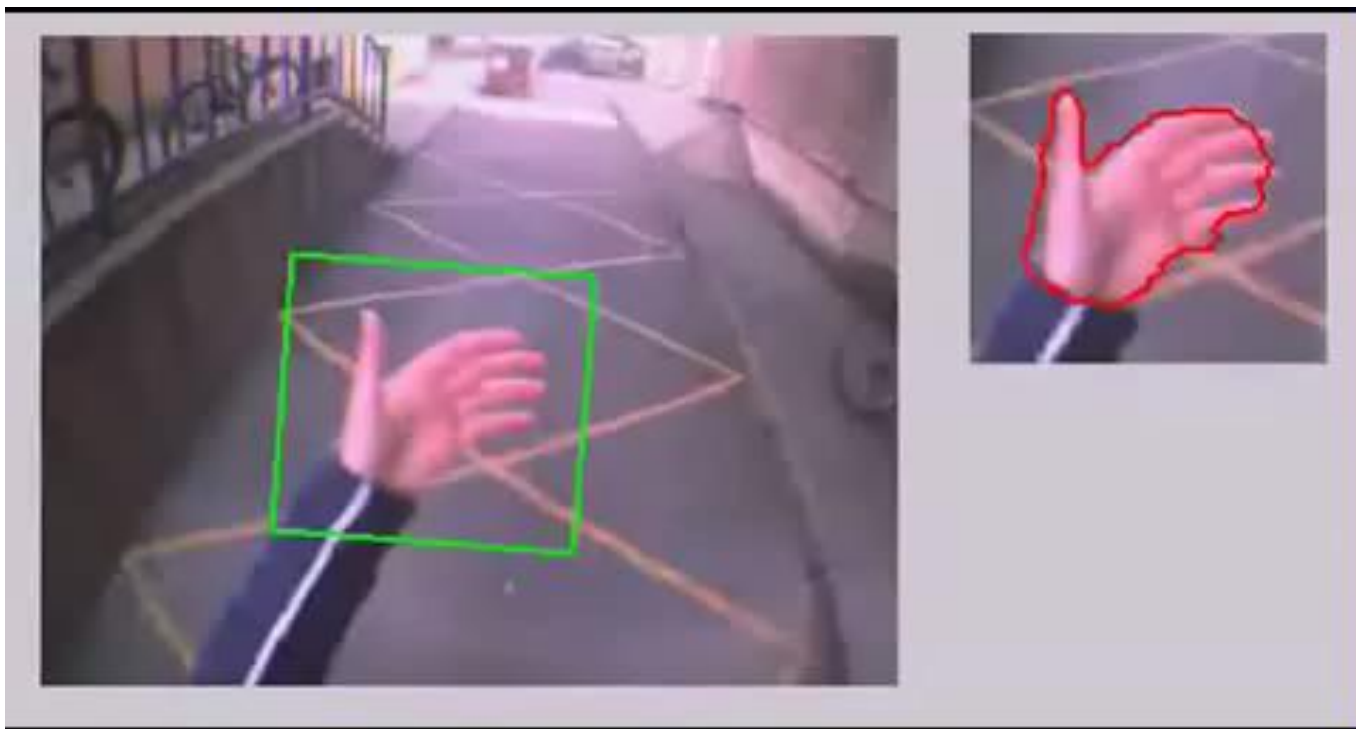
# 运动目标的表达方法

基于特征的跟踪：



# 运动目标的表达方法

基于轮廓的跟踪：



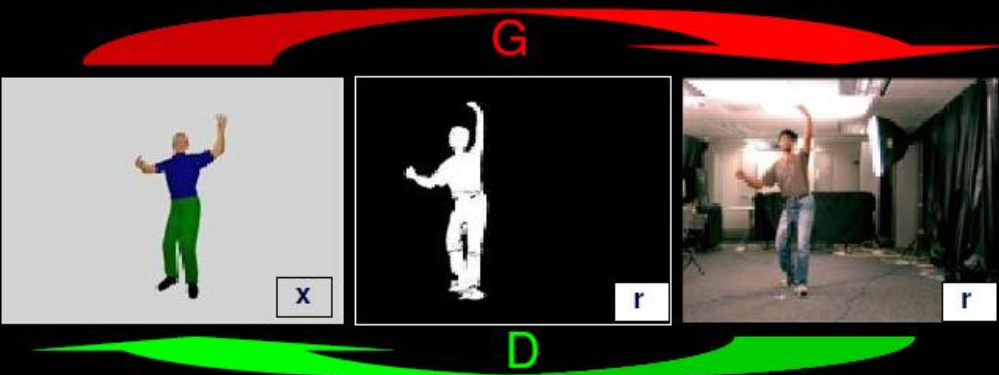
# 运动目标的表达方法

基于模型的跟踪：



# 目标跟踪的基本方法

## Generative vs. Discriminative Modelling



$x$  is the model state  
 $r$  are image observations

**Goal:**  $p_{\theta}(x|r)$

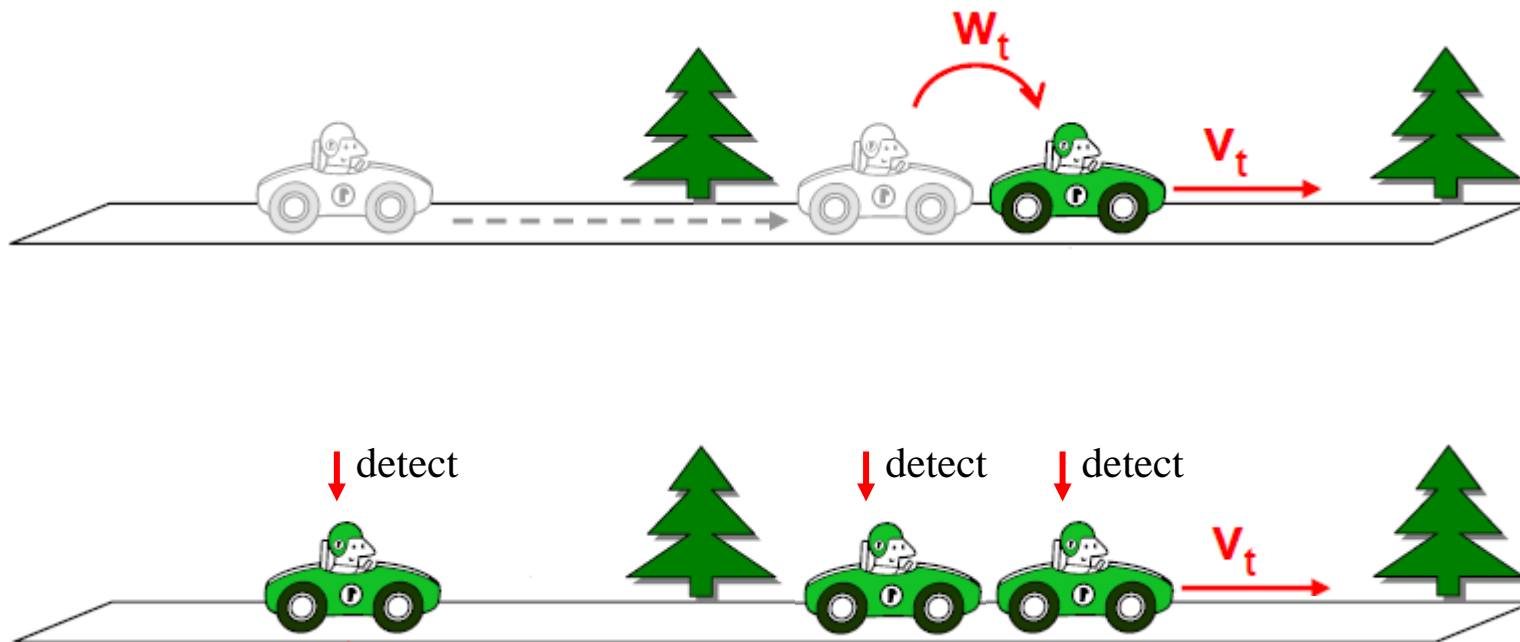
$\theta$  are parameters to learn  
given training set of  $(r, x)$  pairs

$$p_{\theta}(x|r) \propto p_{\theta}(r|x) \cdot p(x)$$

- 判别式方法（自底向上、数据驱动）：不依赖先验
- VS**
- 生成式方法（自顶向下、模型驱动）：依赖于先验



# 目标跟踪的基本方法



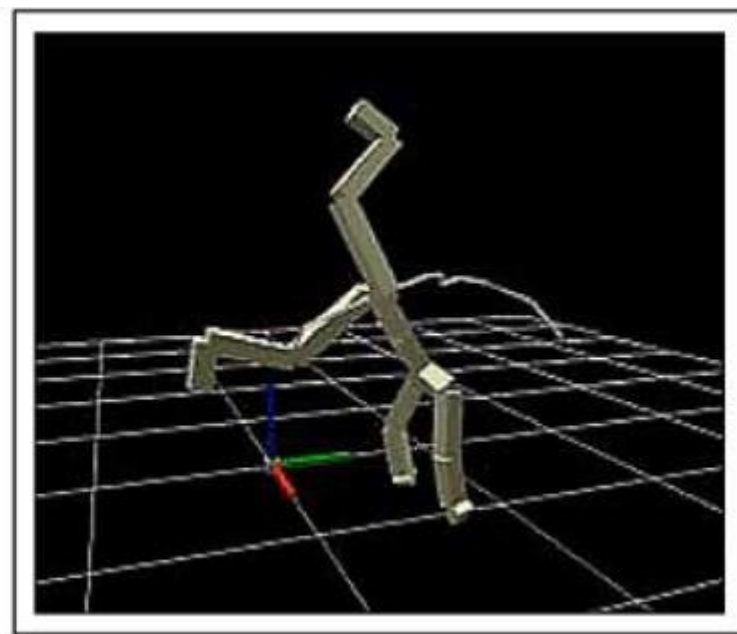
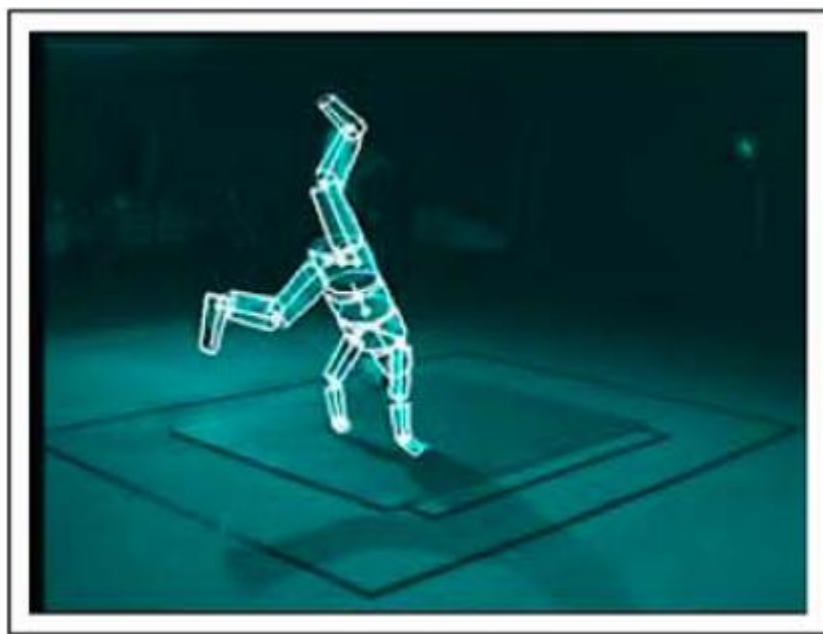
- 递归式跟踪 (Recursive tracking)

VS

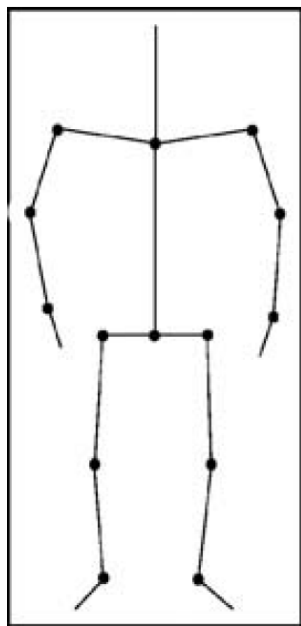
- 单帧检测跟踪 (Tracking by detection)

## 目标跟踪实例——基于退火粒子滤波的人体跟踪

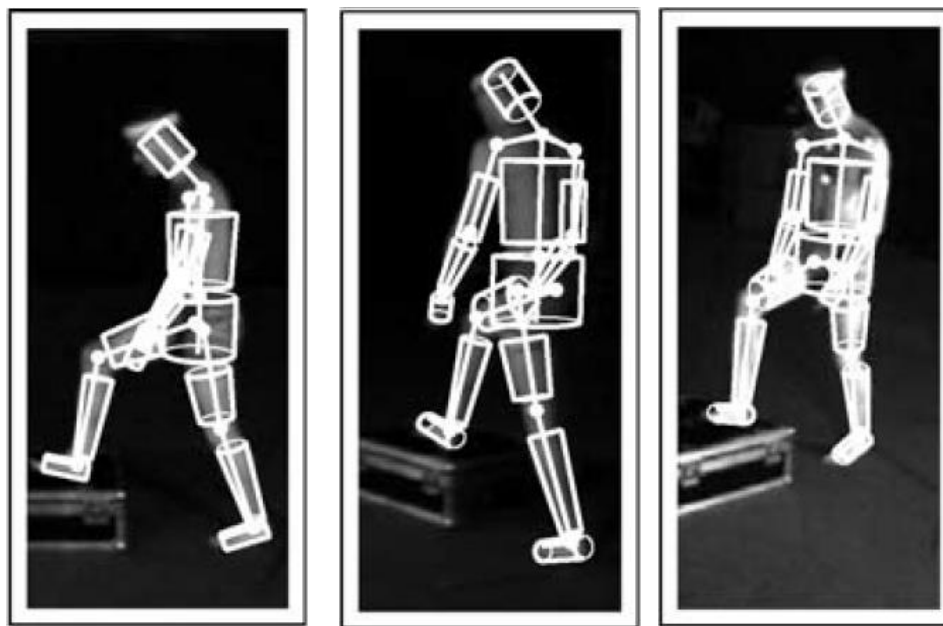
Jonathan Deutscher and Ian D. Reid, **Articulated Body Motion Capture by Stochastic Search**, International Journal of Computer Vision, 61(2): 185–205, 2005.



# 目标跟踪实例——基于退火粒子滤波的人体跟踪



29自由度3D人体模型

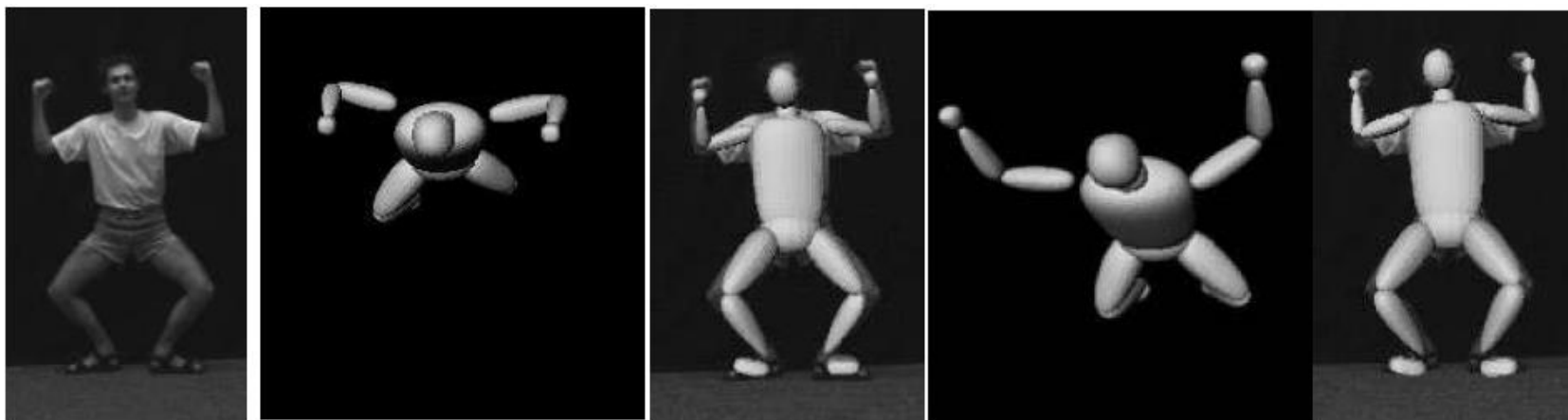


三台摄像机同步拍摄

**算法特征：**单目标、多摄像机、3D、连接刚体、生成式算法、递归式跟踪

# 目标跟踪实例——基于退火粒子滤波的人体跟踪

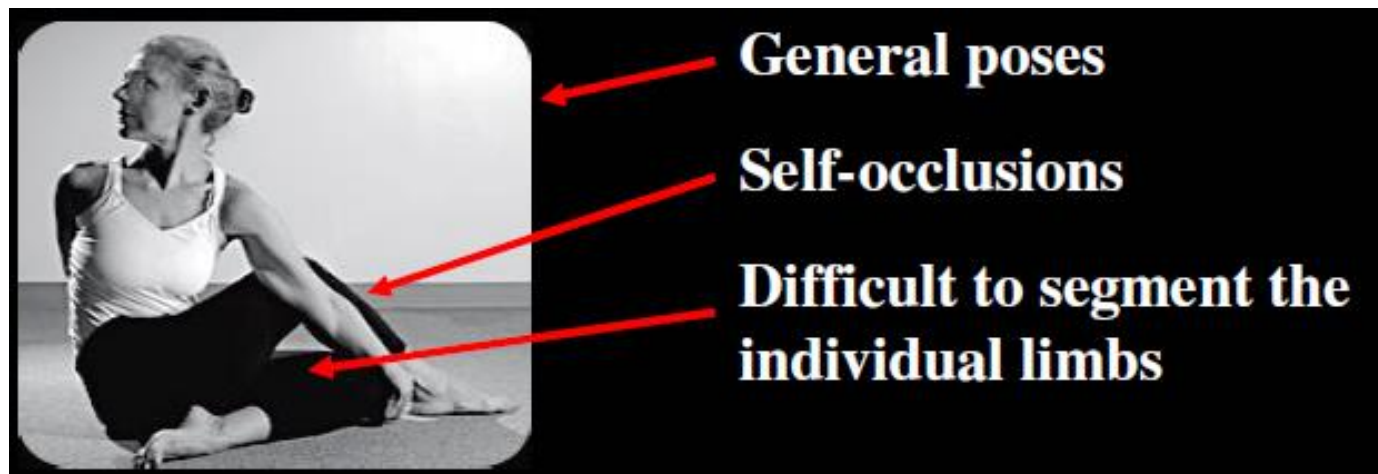
人体三维跟踪的主要难点：



3D到2D投影歧义

# 目标跟踪实例——基于退火粒子滤波的人体跟踪

人体三维跟踪的主要难点：



姿态歧义

# 目标跟踪实例——基于退火粒子滤波的人体跟踪

人体三维跟踪的主要难点：



尺度歧义



# 目标跟踪实例——基于退火粒子滤波的人体跟踪

人体三维跟踪的主要难点：



运动模糊

# 目标跟踪实例——基于退火粒子滤波的人体跟踪

背景知识： 贝叶斯滤波（Bayesian Filtering）

状态：  $\mathbf{x}_t$

观测：  $\mathbf{z}_t$

过程方程：  $\mathbf{x}_t = \mathbf{f}_t(\mathbf{x}_{t-1}, \mathbf{v}_{t-1})$

测量方程：  $\mathbf{z}_t = \mathbf{h}_t(\mathbf{x}_t, \mathbf{n}_t)$

过程噪声：  $\mathbf{v}_{t-1}$

测量噪声：  $\mathbf{n}_t$

贝叶斯滤波： 已知0到 $t$ 时刻的观测 $\mathbf{z}_{1:t} = \{\mathbf{z}_1, \dots, \mathbf{z}_t\}$ 以及0时刻的状态先验 $p(\mathbf{x}_0)$ ，求 $t$ 时刻状态的概率分布 $p(\mathbf{x}_t | \mathbf{z}_{1:t})$

# 目标跟踪实例——基于退火粒子滤波的人体跟踪

背景知识： 贝叶斯滤波（Bayesian Filtering）

根据贝叶斯准则（Bayes Rule）：

$$p(\mathbf{x}_t | \mathbf{z}_{1:t}) = \frac{p(\mathbf{z}_t | \mathbf{x}_t) p(\mathbf{x}_t | \mathbf{z}_{1:t-1})}{p(\mathbf{z}_t | \mathbf{z}_{1:t-1})}$$

- 1) 预测： 计算先验  $p(\mathbf{x}_t | \mathbf{z}_{1:t-1}) = \int p(\mathbf{x}_t | \mathbf{x}_{t-1}) p(\mathbf{x}_{t-1} | \mathbf{z}_{1:t-1}) d\mathbf{x}_{t-1}$   
其中  $p(\mathbf{x}_t | \mathbf{z}_{1:t-1})$  根据过程方程  $\mathbf{x}_t = \mathbf{f}_t(\mathbf{x}_{t-1}, \mathbf{v}_{t-1})$
- 2) 测量： 计算似然  $p(\mathbf{z}_t | \mathbf{x}_t)$ ， 根据测量方程  $\mathbf{z}_t = \mathbf{h}_t(\mathbf{x}_t, \mathbf{n}_t)$
- 3) 更新： 计算后验  $p(\mathbf{x}_t | \mathbf{z}_{1:t}) \propto p(\mathbf{z}_t | \mathbf{x}_t) p(\mathbf{x}_t | \mathbf{z}_{1:t-1})$

# 目标跟踪实例——基于退火粒子滤波的人体跟踪

背景知识：贝叶斯滤波（Bayesian Filtering）

上述贝叶斯滤波给出了状态跟踪问题的最优贝叶斯解，这一过程只是理论上解的形式，实际中往往无法得到解析表达。

在给定一些假设（高斯、线性）时，贝叶斯解具有解析解的形式。

# 目标跟踪实例——基于退火粒子滤波的人体跟踪

背景知识：卡尔曼滤波（Kalman Filtering）

卡尔曼滤波的基本假设：

- 1) 后验概率 $p(\mathbf{x}_{t-1}|\mathbf{z}_{1:t-1})$ 为高斯分布
- 2) 过程方程是线性函数， $\mathbf{x}_t = \mathbf{f}_t(\mathbf{x}_{t-1}, \mathbf{v}_{t-1}) = F_t \mathbf{x}_{t-1} + \mathbf{v}_{t-1}$
- 3) 测量方程是线性函数， $\mathbf{z}_t = \mathbf{h}_t(\mathbf{x}_t, \mathbf{n}_t) = H_t \mathbf{x}_t + \mathbf{n}_t$
- 4) 过程噪声 $\mathbf{v}_{t-1}$ 是高斯白噪声，方差为 $Q_{t-1}$
- 5) 观测噪声 $\mathbf{n}_t$ 是高斯白噪声，方差为 $R_t$

# 目标跟踪实例——基于退火粒子滤波的人体跟踪

背景知识：卡尔曼滤波（Kalman Filtering）

过程方程：  $\mathbf{x}_t = F_t \mathbf{x}_{t-1} + \mathbf{v}_{t-1}$       过程噪声  $\mathbf{v}_{t-1} \sim N(0, Q_{t-1})$

测量方程：  $\mathbf{z}_t = H_t \mathbf{x}_t + \mathbf{n}_t$       测量噪声  $\mathbf{n}_t \sim N(0, R_t)$

$p(\mathbf{x}_{t-1} | \mathbf{z}_{1:t-1}) \sim N(m_{t-1|t-1}, P_{t-1|t-1})$ ,  $t-1$ 时刻状态，已知

$p(\mathbf{x}_t | \mathbf{z}_{1:t-1}) \sim N(m_{t|t-1}, P_{t|t-1})$ , 对 $t$ 时刻状态的进行预测，先验

$p(\mathbf{x}_t | \mathbf{z}_{1:t}) \sim N(m_{t|t}, P_{t|t})$ , 根据 $t$ 时刻的测量进行更新，后验

预测：  $m_{t|t-1} = F_t m_{t-1|t-1}$        $P_{t|t-1} = Q_{t-1} + F_t P_{t-1|t-1} F_t^T$

更新：  $m_{t|t} = m_{t|t-1} + t_t (\mathbf{z}_t - H_t m_{t|t-1})$        $P_{t|t} = P_{t|t-1} - t_t H_t P_{t|t-1}$

$t_t = P_{t|t-1} H_t^T (H_t P_{t|t-1} H_t^T + R_t)^{-1}$  称为卡尔曼增益



# 目标跟踪实例——基于退火粒子滤波的人体跟踪

背景知识：粒子滤波（Particle Filtering）

许多跟踪问题是非线性非高斯的，因此无法使用卡尔曼滤波有效求解。针对非线性非高斯状态跟踪的一种有效方法是粒子滤波。

粒子滤波是对贝叶斯最优解的近似，其基本思想是基于蒙特卡洛法（Monte Carlo Method）。

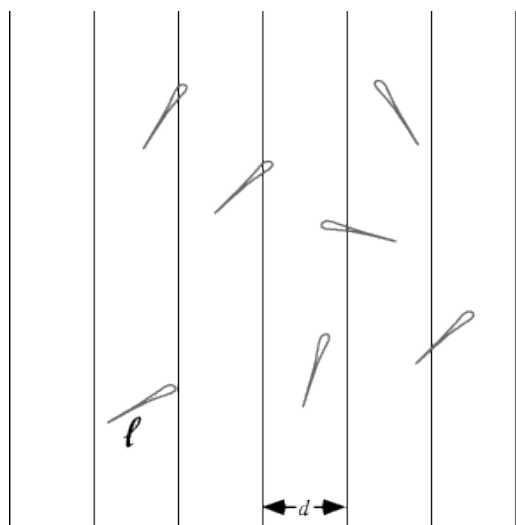
蒙特卡罗方法于20世纪40年代美国在第二次世界大战中研制原子弹的“曼哈顿计划”计划的成员乌拉姆和冯·诺伊曼首先提出，并用驰名世界的赌城——摩纳哥的Monte Carlo——来命名了这种方法。在这之前，蒙特卡罗方法就已经存在。1777年，法国数学家布丰提出用投针实验的方法求圆周率 $\pi$ ，这被认为是蒙特卡罗方法的起源。

# 目标跟踪实例——基于退火粒子滤波的人体跟踪

背景知识：粒子滤波（Particle Filtering）

布丰投针实验（计算 $\pi$ 的最稀奇方法之一）：

1. 取一张白纸，在上面画上许多条间距为 $d$ 的平行线；
2. 取一根长度为 $l$  ( $l < d$ ) 的针，随机地向画有平行直线的纸上掷 $n$ 次，观察针与直线相交的次数 $m$ ；
3. 则针与直线相交的概率与圆周率成正比： $m/n = (\pi - 1)/4$



| 试验者       | 时间(年) | 针长   | 投针次数 | 相交次数 | $\pi$ 的估计值 |
|-----------|-------|------|------|------|------------|
| Wolf      | 1850  | 0.80 | 5000 | 2532 | 3.15956    |
| Smith     | 1855  | 0.60 | 3204 | 1218 | 3.15665    |
| Fox       | 1884  | 0.75 | 1030 | 489  | 3.15951    |
| Lazzarini | 1925  | 0.83 | 3408 | 1808 | 3.14159292 |

# 目标跟踪实例——基于退火粒子滤波的人体跟踪

背景知识：粒子滤波（Particle Filtering）

非高斯非线性的贝叶斯滤波没有解析解，因此粒子滤波使用一组随机样本对贝叶斯滤波器进行蒙特卡洛模拟。

粒子滤波的核心思想是使用一组带有权重的随机样本来近似状态的后验概率。

令  $\{\mathbf{x}_t^i, w_t^i\}_{i=1, \dots, N_s}$  为  $t$  时刻的一组随机样本， $\mathbf{x}_t^i$  为样本采样， $w_t^i$  为权重，则  $t$  时刻状态  $\mathbf{x}_t$  的后验概率可以近似为：

$$p(\mathbf{x}_t | \mathbf{z}_{1:t}) \approx \sum_{i=1}^{N_s} w_t^i \delta(\mathbf{x}_t - \mathbf{x}_t^i)$$

# 目标跟踪实例——基于退火粒子滤波的人体跟踪

背景知识：粒子滤波（Particle Filtering）

标准粒子滤波算法流程：

输入：  $t-1$ 时刻样本及权值  $\{\mathbf{x}_{t-1}^i, w_{t-1}^i\}_{i=1, \dots, N_s}$ ,  $w_{t-1}^i = 1/N_s$

预测：根据过程方程  $\mathbf{x}_t = \mathbf{f}_t(\mathbf{x}_{t-1}, \mathbf{v}_{t-1})$  生成预测样本  $\{\mathbf{x}_t^{-i}\}$

测量：根据观测似然更新权重  $w_t^{-i} = w_{t-1}^i p(\mathbf{z}_t | \mathbf{x}_t^{-i})$

重采样：对  $N_s$  个样本  $\{\mathbf{x}_t^{-i}\}$  以其权值  $\{w_t^{-i}\}$  为概率进行采样，从而抑制权值较小的样本，增强权值较大的样本，生成新样本  $\{\mathbf{x}_t^i, w_t^i\}_{i=1, \dots, N_s}$ ，其中新权值  $w_t^i$  均设为  $1/N_s$

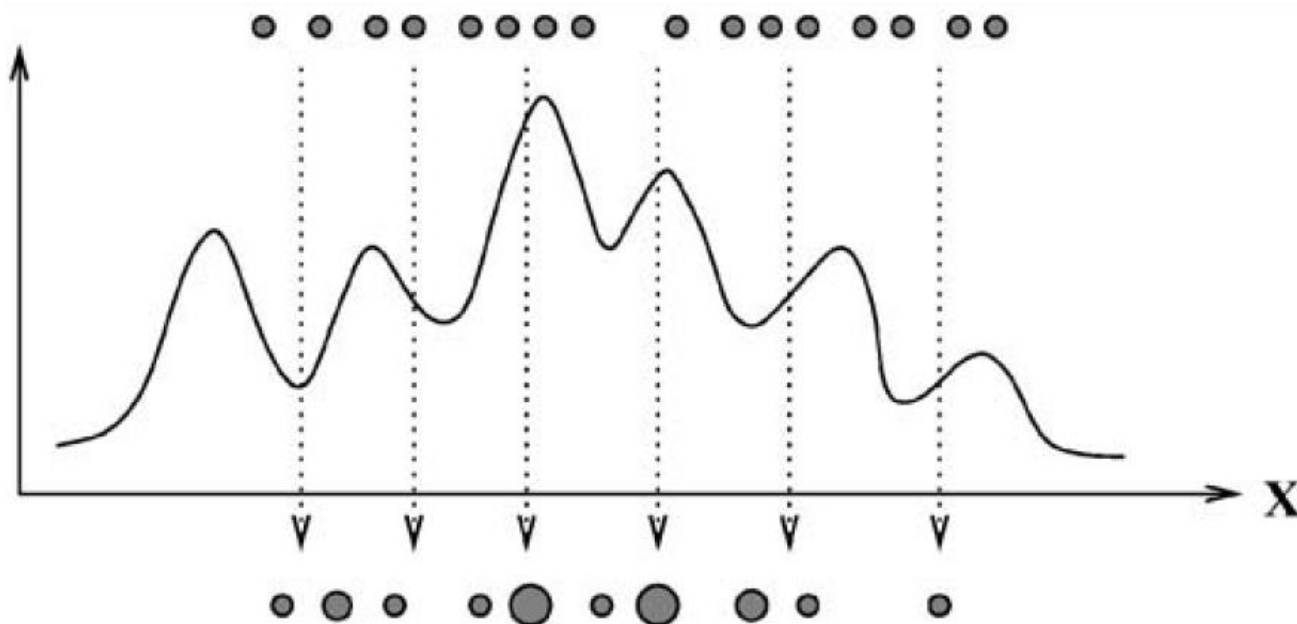
输出：  $t$ 时刻样本及权值  $\{\mathbf{x}_t^i, w_t^i\}_{i=1, \dots, N_s}$ ,  $w_t^i = 1/N_s$

$$t\text{时刻状态估计: } \mathbf{x}_t = \frac{1}{N_s} \sum_{i=1}^{N_s} \mathbf{x}_t^i$$

# 目标跟踪实例——基于退火粒子滤波的人体跟踪

背景知识：粒子滤波（Particle Filtering）

当观测似然函数 $p(\mathbf{z}_t | \mathbf{x}_t^{-i})$ 多峰时，标准粒子滤波算法可能无法准确找到最大后验概率值。



# 目标跟踪实例——基于退火粒子滤波的人体跟踪

背景知识：退火粒子滤波（Annealed Particle Filtering）

使用一组似然函数： $p_0, \dots, p_M$ ，相邻两个函数 $p_m$ 和 $p_{m-1}$ 只有少许不同，定义为：

$$p_m(\mathbf{z} | \mathbf{x}) = \left( p(\mathbf{z} | \mathbf{x}) \right)^{\beta_m}$$

其中 $\beta_0 > \beta_1 > \dots > \beta_M$ ， $p_M$ 的形状比较平坦，能反映出整个状态空间的趋势， $p_0$ 对应真实似然函数，具有较高的峰值（ $\beta_0=1$ ），能够反映局部特征。

退火粒子滤波从比较平坦的 $p_M$ 开始，逐渐收敛到真实观测似然函数 $p_0$ 。



# 目标跟踪实例——基于退火粒子滤波的人体跟踪

背景知识：退火粒子滤波（Annealed Particle Filtering）

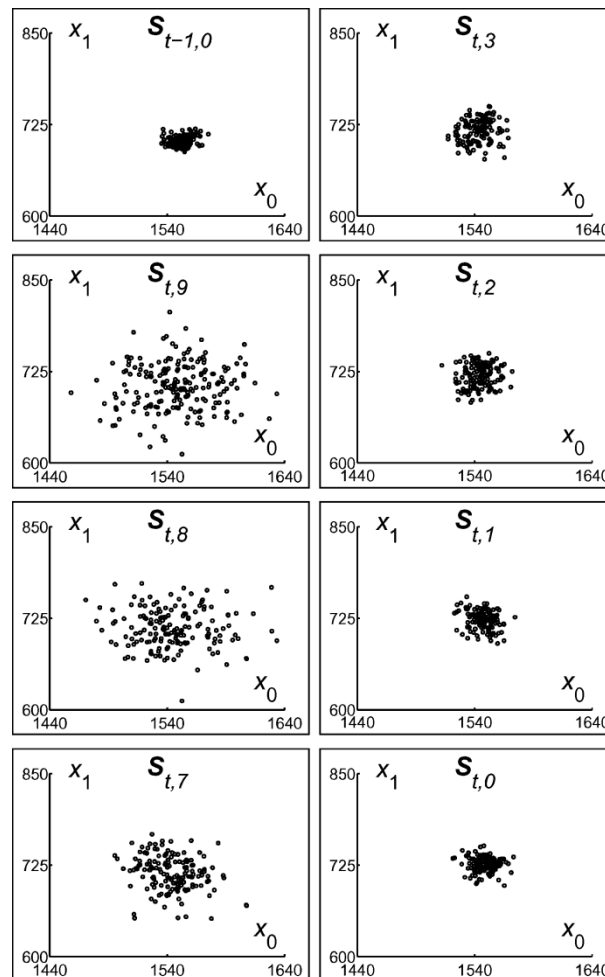
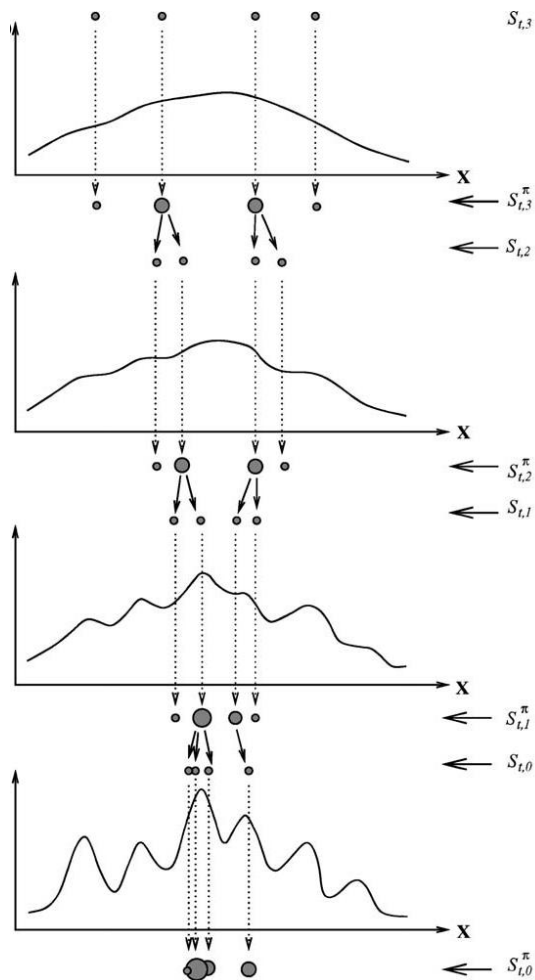
退火粒子滤波流程：

1. 预测：根据过程方程  $\mathbf{x}_t = \mathbf{f}_t(\mathbf{x}_{t-1}, \mathbf{v}_{t-1})$  生成预测样本  $\{\mathbf{x}_t^{-i}\}$
2.  $m=M$ ,  $\{\mathbf{x}_{t,m}^{-i}\} \leftarrow \{\mathbf{x}_t^{-i}\}$
3. 测量：根据观测似然更新权重  $w_{t,m}^{-i} = w_{t-1}^{-i} p_m(\mathbf{z}_t | \mathbf{x}_{t,m}^{-i})$
4. 重采样：对  $N_s$  个样本  $\{\mathbf{x}_{t,m}^{-i}\}$  以其权值  $\{w_{t,m}^{-i}\}$  为概率进行重采样，生成新样本  $\{\mathbf{x}_{t,m-1}^{-i}, w_{t,m-1}^{-i}\}$ ，其中  $\mathbf{x}_{t,m-1}^{-i} = \mathbf{x}_{t,m}^{-i} + \mathbf{B}_m$ 、 $w_{t,m-1}^{-i} = 1/N_s$ 、 $\mathbf{B}_m$  为高斯随机变量
5.  $m \leftarrow m-1$ ，如果  $m < 0$  则迭代结束，否则转步骤3

输出： $t$ 时刻样本及权值  $\{\mathbf{x}_t^i, w_t^i\}_{i=1, \dots, N_s}$ ,  $w_t^i = 1/N_s$

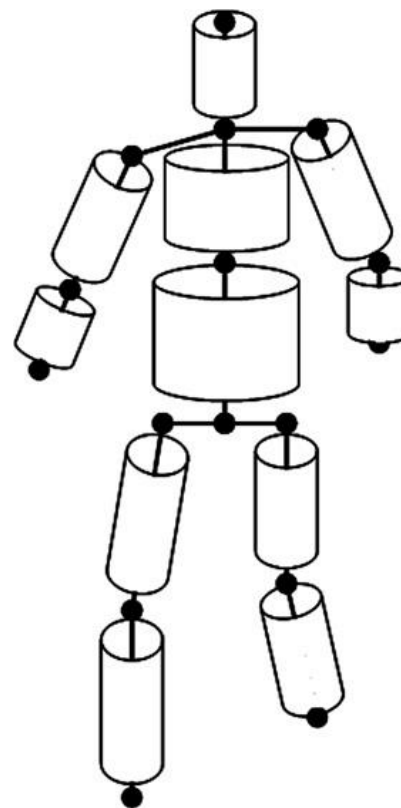
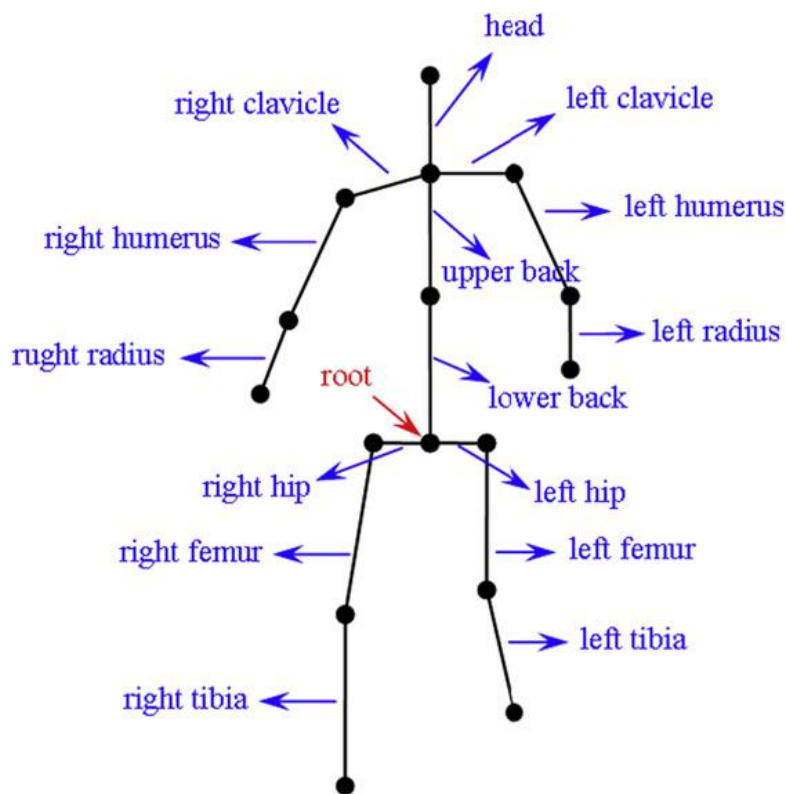
# 目标跟踪实例——基于退火粒子滤波的人体跟踪

## 背景知识：退火粒子滤波（Annealed Particle Filtering）



# 目标跟踪实例——基于退火粒子滤波的人体跟踪

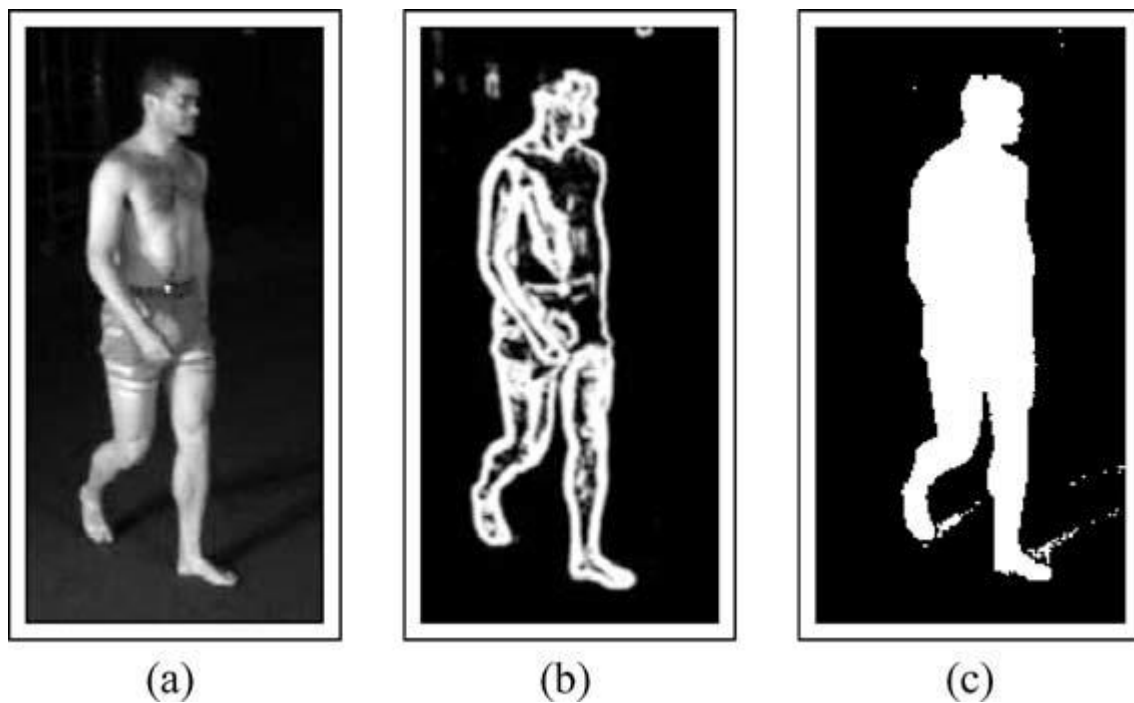
人体结构模型：



29自由度3D人体模型

# 目标跟踪实例——基于退火粒子滤波的人体跟踪

观测似然函数：



图像特征：边缘+剪影

# 目标跟踪实例——基于退火粒子滤波的人体跟踪

观测似然函数：

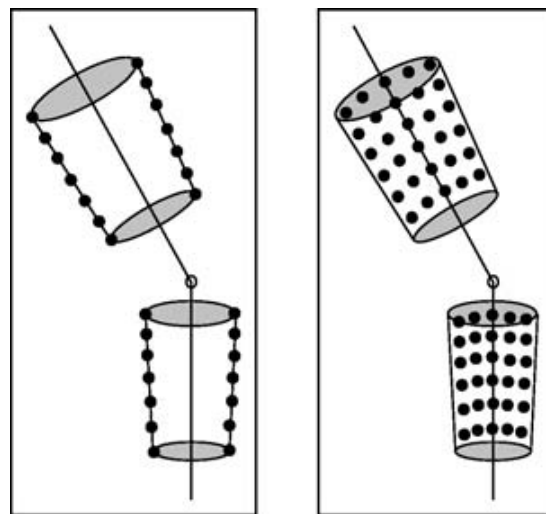
$$p(\mathbf{z} | \mathbf{x}) = \exp - \left( \frac{1}{C} \sum_{i=1}^C (1 - r_i(\mathbf{x}, I_i)) + \frac{1}{C} \sum_{i=1}^C (1 - e_i(\mathbf{x}, I_i)) \right)$$

$C$ : 摄像机数量

$I_i$ : 第 $i$ 个摄像机对应的图像

$r_i(\mathbf{x}, I_i)$ : 模型投影边界采样点处的图像边缘点比例

$e_i(\mathbf{x}, I_i)$ : 模型投影区域采样点处的图像剪影点比例



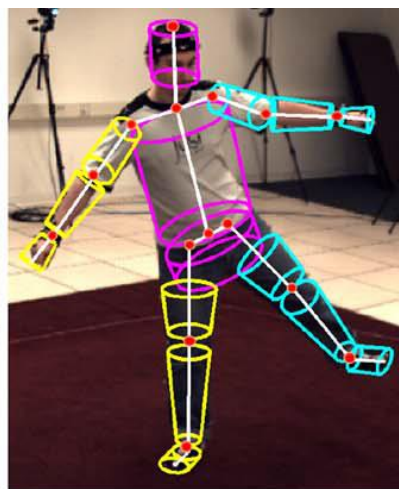
# 目标跟踪实例——基于退火粒子滤波的人体跟踪

观测似然函数：

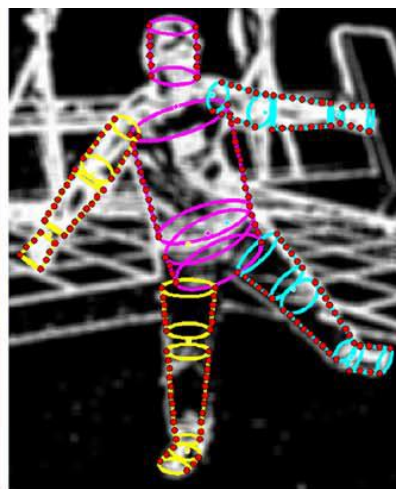
$$p(\mathbf{z} | \mathbf{x}) = \exp - \left( \frac{1}{C} \sum_{i=1}^C (1 - r_i(\mathbf{x}, I_i)) + \frac{1}{C} \sum_{i=1}^C (1 - e_i(\mathbf{x}, I_i)) \right)$$



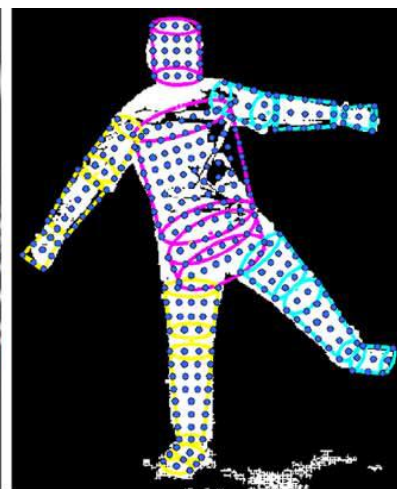
$I_i$



$I_i$  and  $\mathbf{x}$



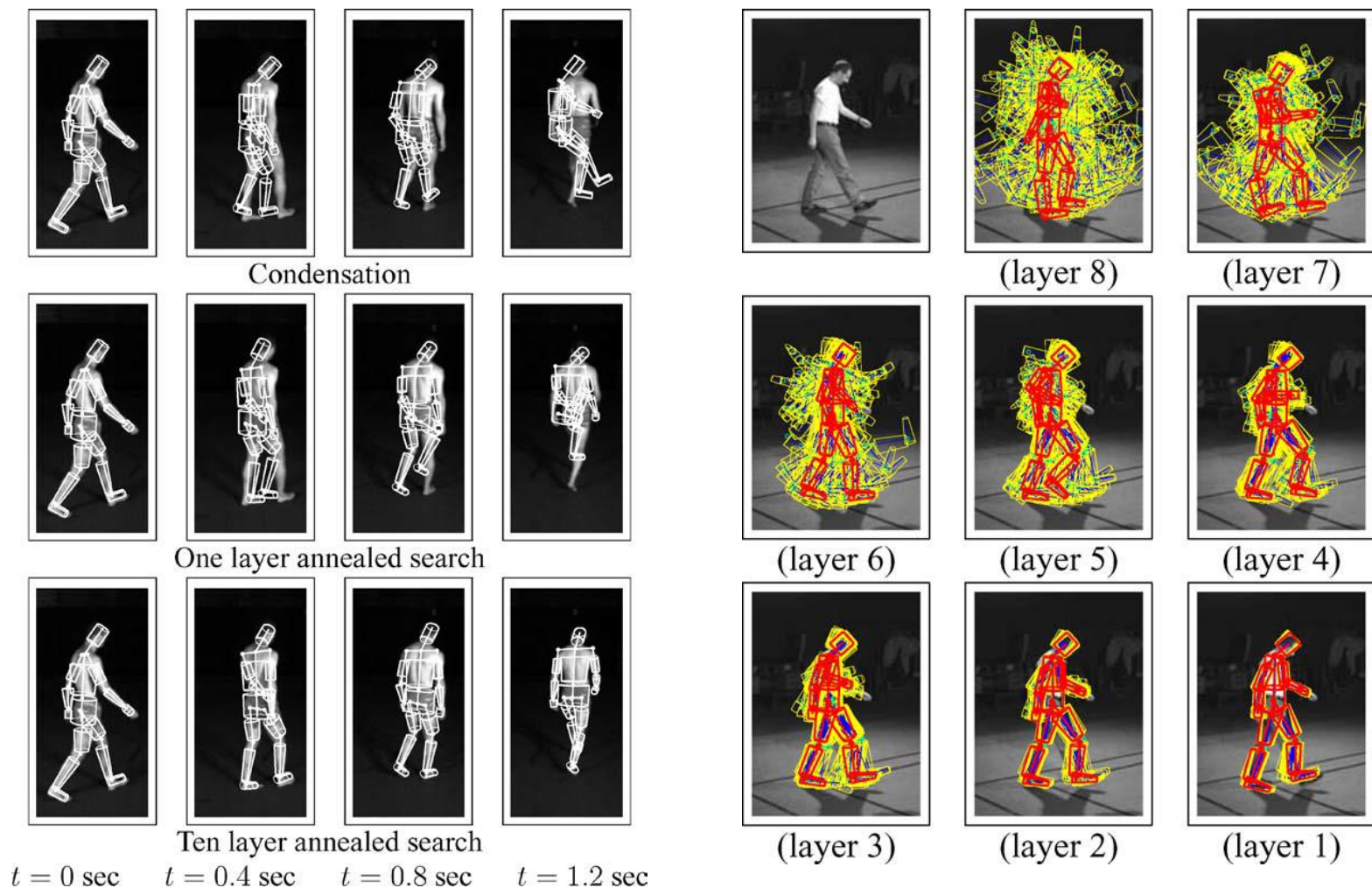
$r_i(\mathbf{x}, I_i)$



$e_i(\mathbf{x}, I_i)$

# 目标跟踪实例——基于退火粒子滤波的人体跟踪

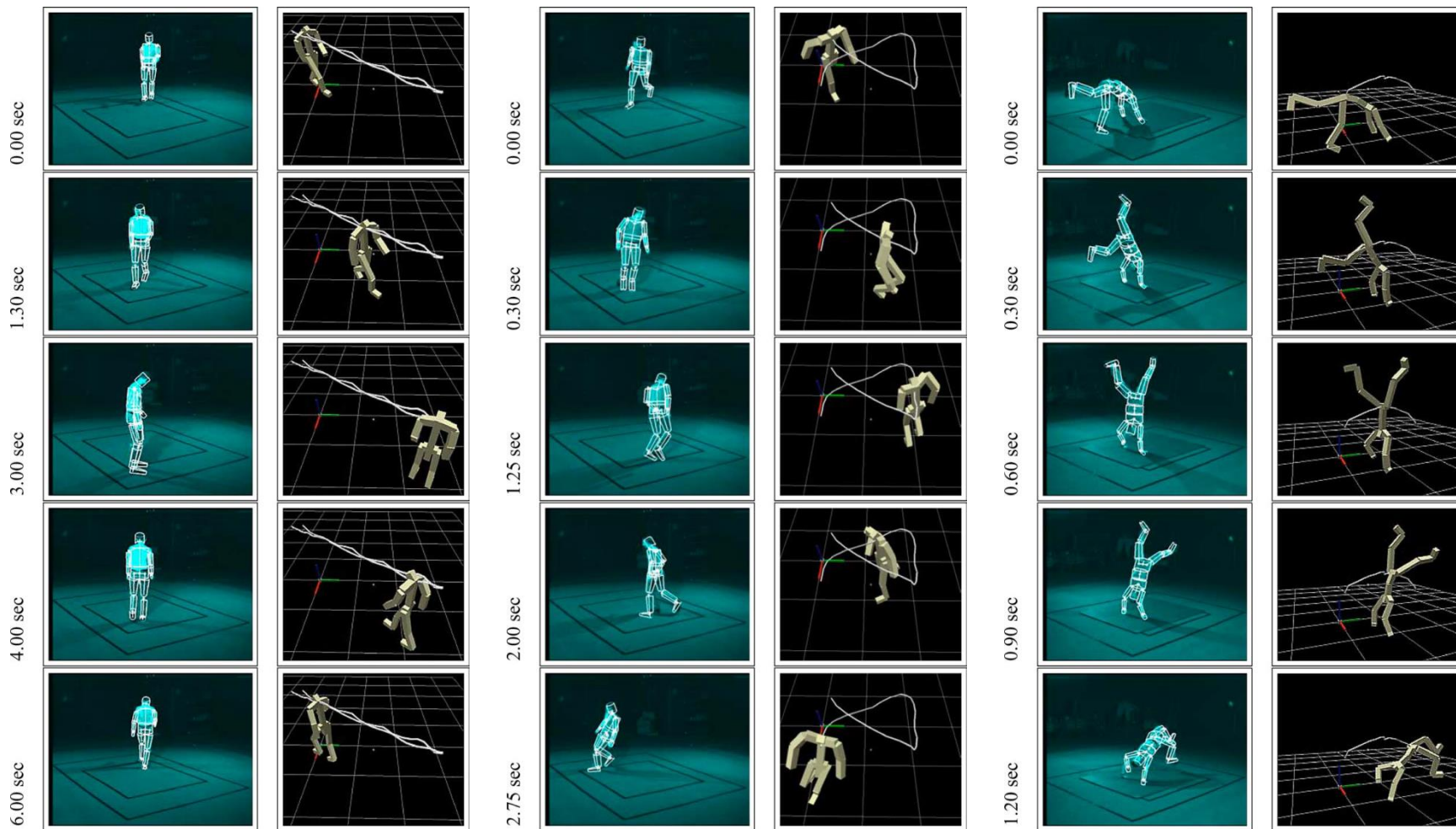
跟踪结果:





# 目标跟踪实例——基于退火粒子滤波的人体跟踪

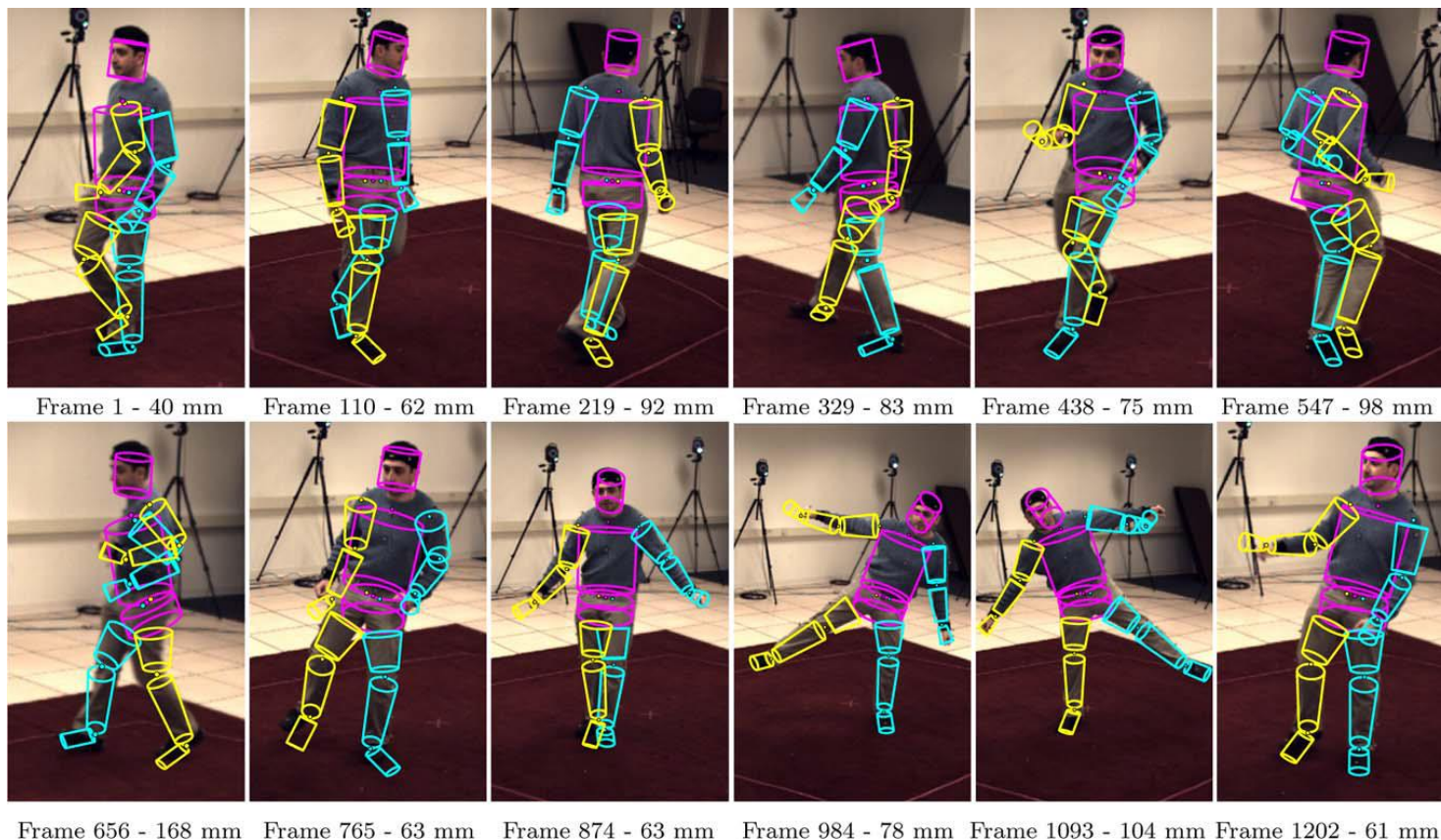
跟踪结果：





# 目标跟踪实例——基于退火粒子滤波的人体跟踪

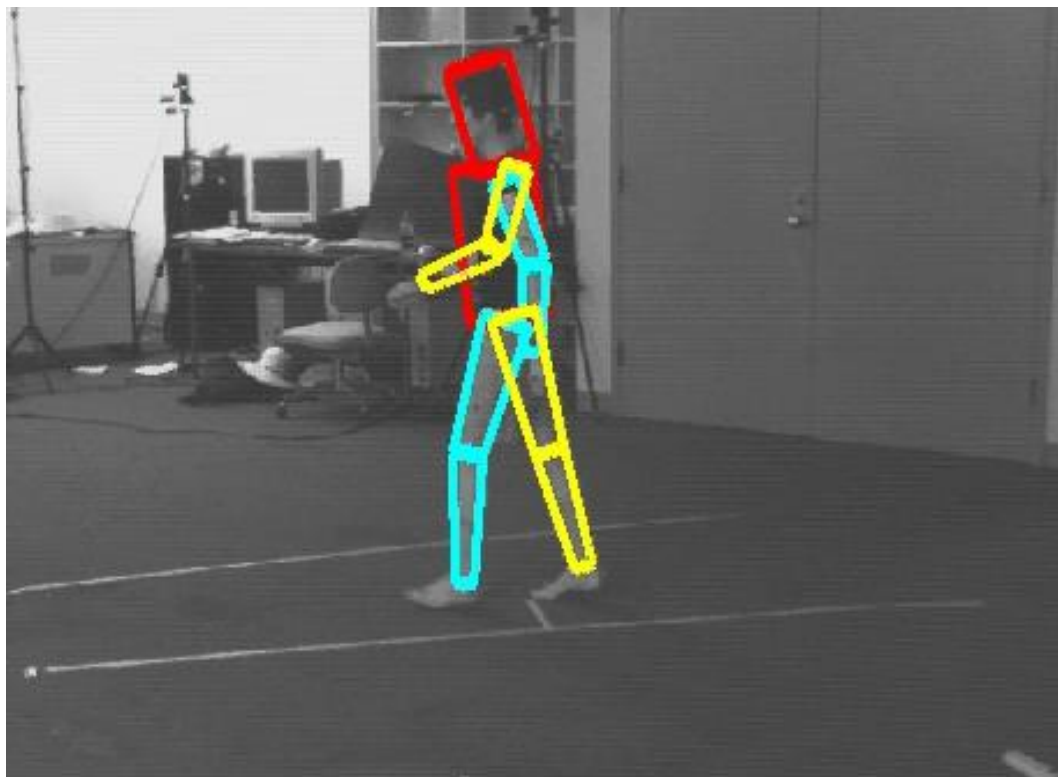
跟踪结果:



Results on HumanEva Dataset

# 目标跟踪实例——基于退火粒子滤波的人体跟踪

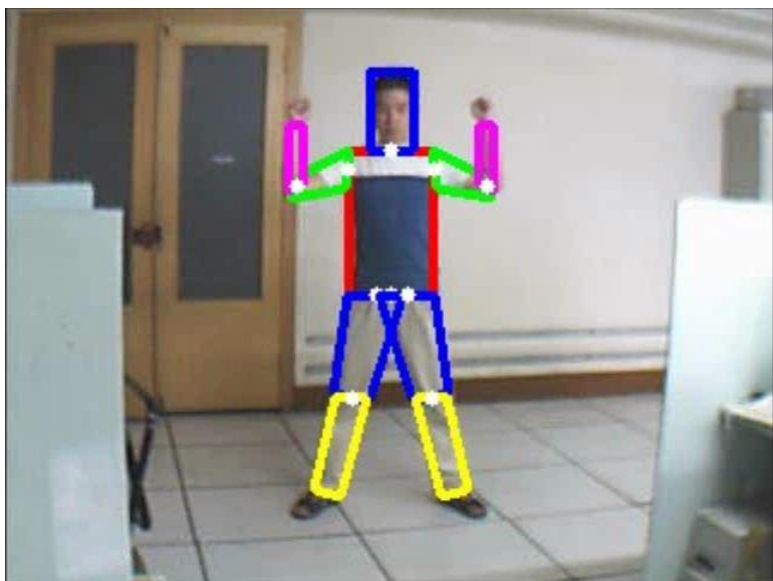
跟踪结果：



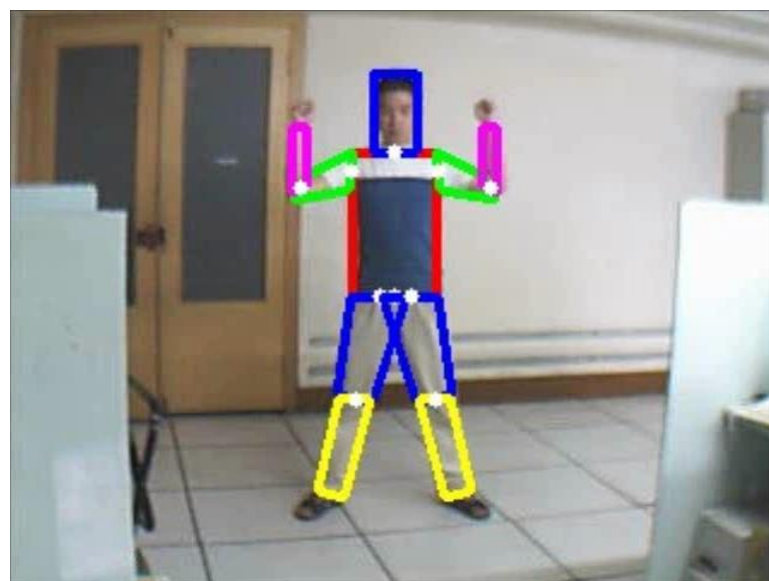
Results on HumanEva Dataset

## 目标跟踪实例——基于退火粒子滤波的人体跟踪

APF是一种递归式跟踪（Recursive tracking）方法，需要前一帧结果作为当前帧的初值，因此在长序列跟踪中有可能会发生跟踪漂移，这是任何递归式跟踪算法都难以避免的问题。



APF跟踪发生漂移



另一种递归式跟踪

# 目标跟踪实例——Kinect人体跟踪

Jamie Shotton, et al., **Real-Time Human Pose Recognition in Parts from Single Depth Images**, CVPR, 2011. (Best Paper Award)



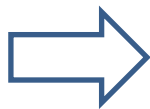
Microsoft Research Cambridge & Xbox Incubation

# 目标跟踪实例——Kinect人体跟踪

通过单幅深度图像计算人体关节点3D空间位置



Kinect



深度图像



关节点3D位置

**算法特征：**多目标、单（深度）摄像机、3D、连接刚体、判别式算法、单帧检测跟踪

## 目标跟踪实例——Kinect人体跟踪

Kinect人体跟踪算法是迄今为止最鲁棒的室内3D人体跟踪算法，其主要特性为：

- 使用单幅深度图像（深度对于视觉感知的重要性）
- 单帧检测跟踪（鲁棒性）
- 在Xbox 360的GPU上运算速度200帧/秒（高效性）
- 可用各种身材各种姿态（通用性）
- 上市60天售出800万台，成为有史以来销售速度最快的电子设备（真实可用）



# 目标跟踪实例——Kinect人体跟踪

Kinect基本原理：



Kinect上红外光源投射出激光散斑（laser speckle），这些散斑具有高度的随机性，而且会随着距离的不同变换图案。不同距离的散斑图案是预先标定好的：每隔一段距离，取一个参考平面，把参考平面上的散斑图案记录下来。需要进行测量的时候，拍摄一副待测场景的散斑图像，将这幅图像和事先保存下来的参考散斑图像依次做互相关运算获得场景深度。

# 目标跟踪实例——Kinect人体跟踪

算法的基本假设：

- 室内环境；
- 人与Kinect距离在1-5米；
- 人体动作包括走、跳、踢、跑等常用游戏动作。





# 目标跟踪实例——Kinect人体跟踪

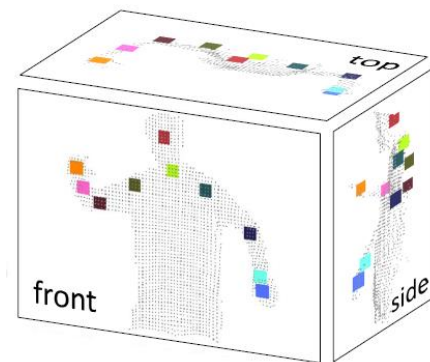
算法的基本流程：



输入深度图像



推断身体各部分



关节点3D位置计算

# 目标跟踪实例——Kinect人体跟踪

算法的基本流程：



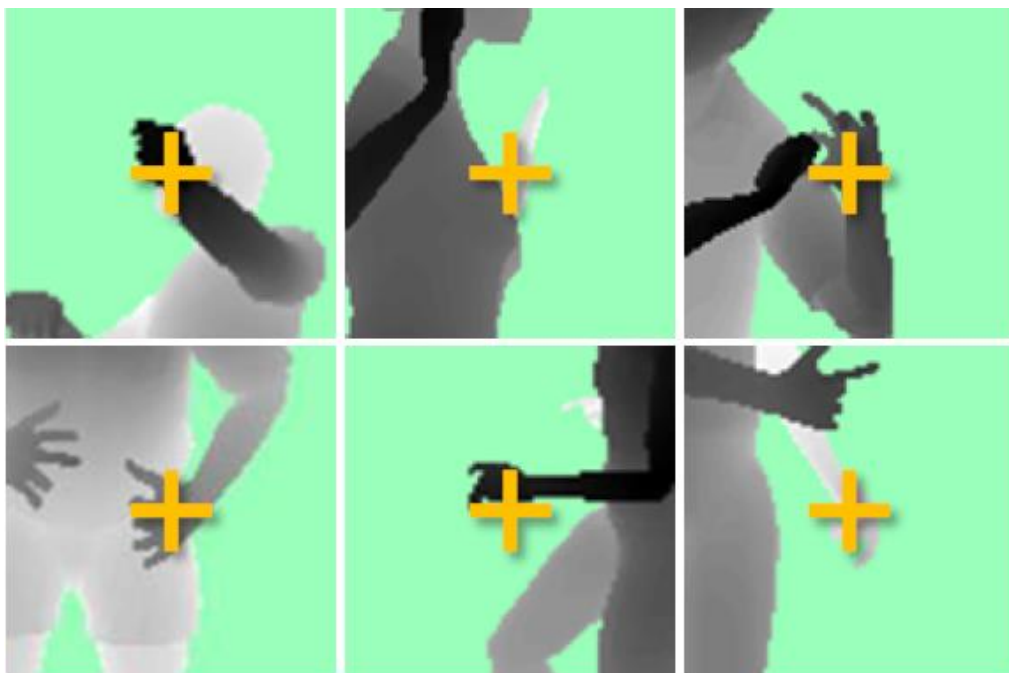
输入深度图像

推断身体各部分

训练一个分类器，输入为深度图中每个像素点（特征），输出为这一像素点所对应的身体部分标签。

## 目标跟踪实例——Kinect人体跟踪

身体上同一部分在不同姿态的深度图中有很大的差别，因此在分类器的训练过程中需要大量带标签的样本。



同一个人的左手在不同姿态时的深度图

# 目标跟踪实例——Kinect人体跟踪

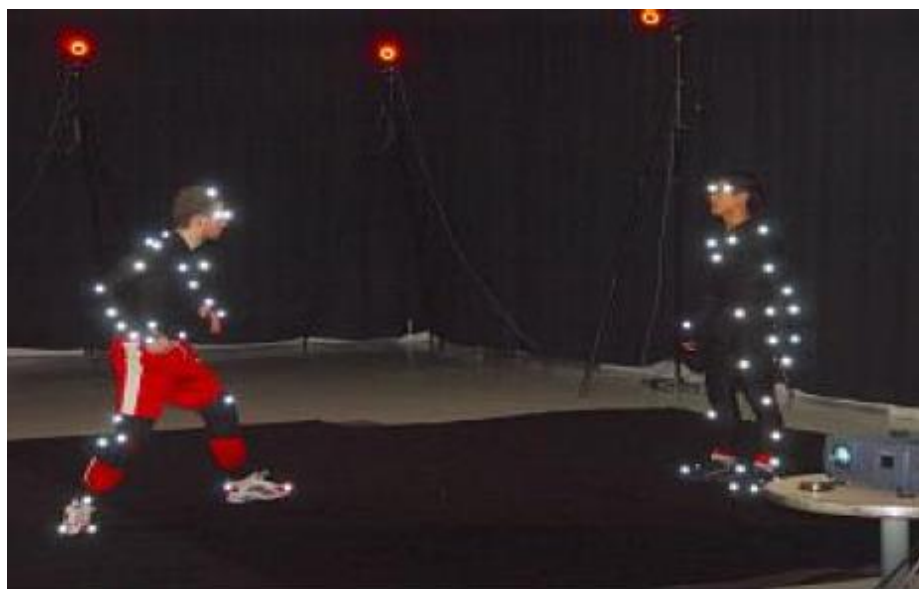
如何产生训练样本？



样本：深度图、标签图（31个身体标签）

## 目标跟踪实例——Kinect人体跟踪

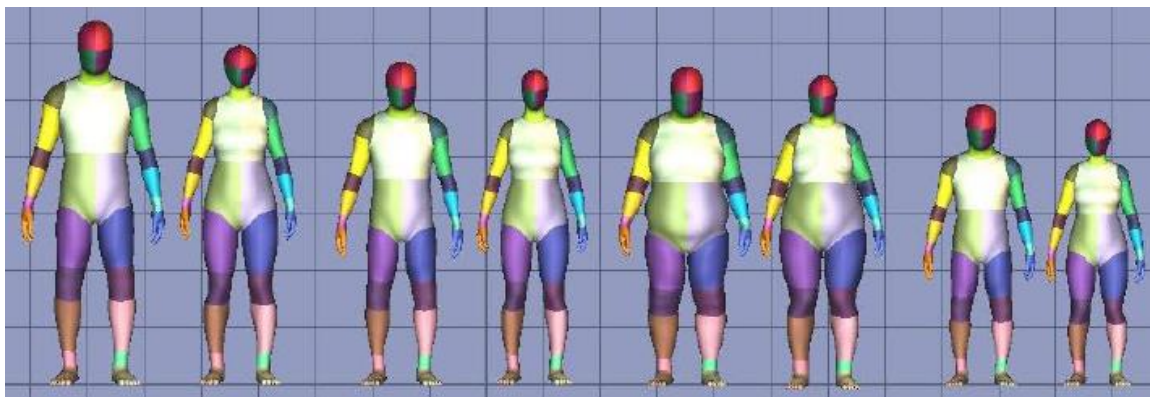
如何产生训练样本？使用运动捕捉（Motion Capture）装置获取人体运动的关节点3D位置。



采集了约50万帧人体运动数据（跳、走、跑、划等），从中选出约10万帧作为训练样本（任意两帧关节点间距离均大于5cm）。

# 目标跟踪实例——Kinect人体跟踪

将采集的10万帧人体运动数据用于虚拟人体运动（Autodesk MotionBuilder）



不同高矮胖瘦



添加衣物发型

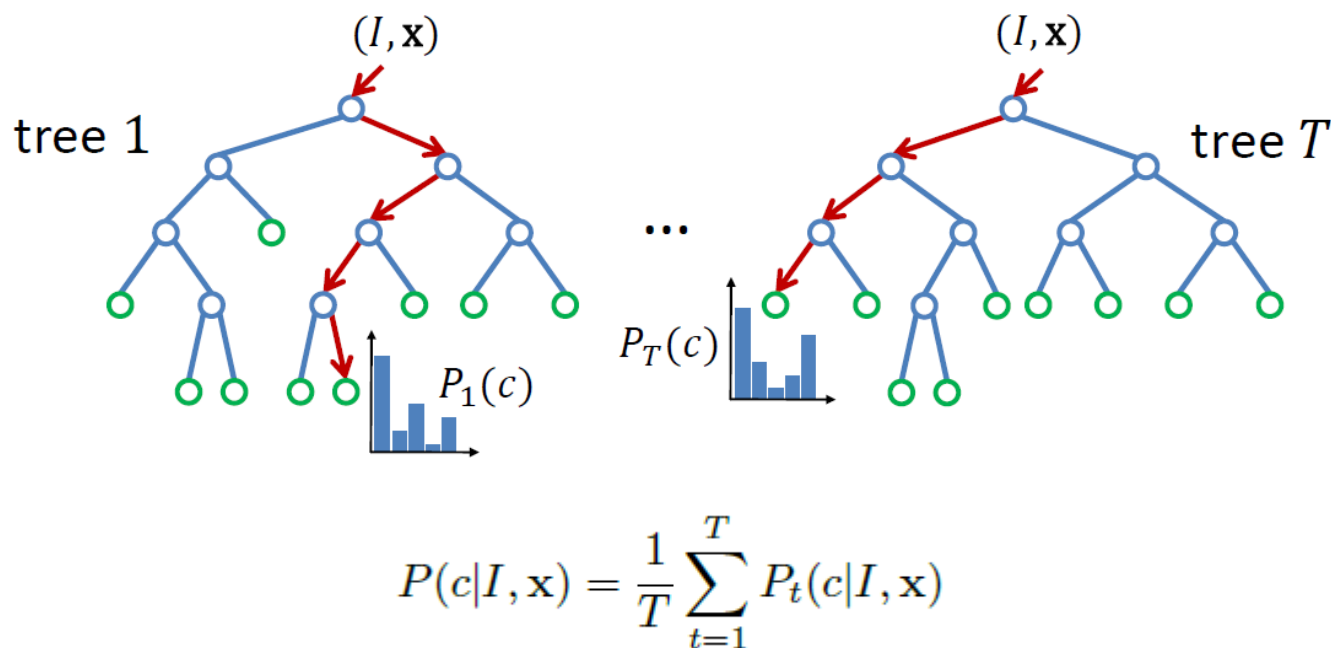
# 目标跟踪实例——Kinect人体跟踪

将虚拟运动人体投影到虚拟深度相机，获得训练样本。



# 目标跟踪实例——Kinect人体跟踪

使用随机决策森林作为分类器：



- Y. Amit and D. Geman. Shape quantization and recognition with randomized trees. **Neural Computation**, 9(7):1545–1588, 1997
- V. Lepetit and P. Fua. Keypoint Recognition Using Randomized Trees, **IEEE T-PAMI**, 28(9):1465-1479, 2006.

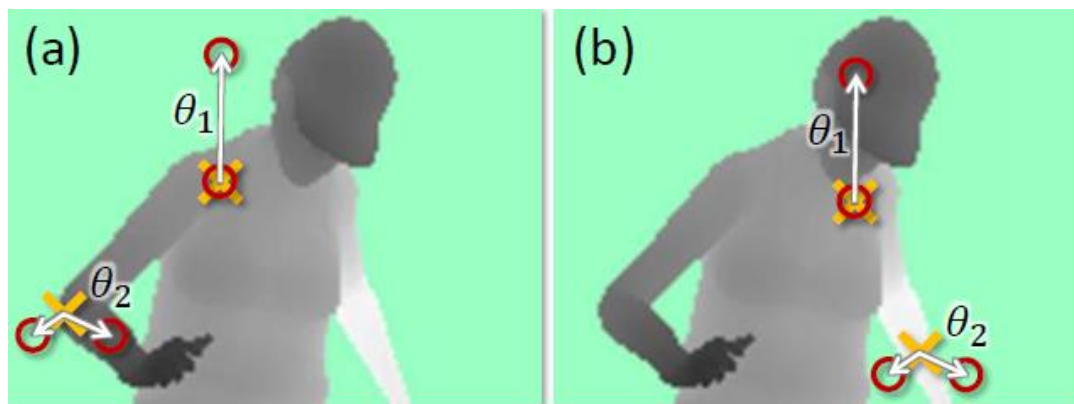


# 目标跟踪实例——Kinect人体跟踪

随机决策森林节点使用的特征：

$$f_{\theta}(I, \mathbf{x}) = d_I \left( \mathbf{x} + \frac{\mathbf{u}}{d_I(\mathbf{x})} \right) - d_I \left( \mathbf{x} + \frac{\mathbf{v}}{d_I(\mathbf{x})} \right)$$

$I$ 为深度图像， $\mathbf{x}$ 为一个像素点， $\theta=(\mathbf{u}, \mathbf{v})$ 为两个偏移向量



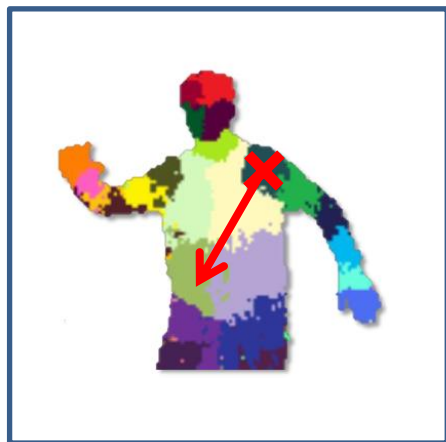
深度图中不同位置像素点使用同样的偏移向量所产生的 $f_{\theta}(I, \mathbf{x})$ 不同

# 目标跟踪实例——Kinect人体跟踪

随机决策森林节点使用的特征：

$$f_{\theta}(I, \mathbf{x}) = d_I \left( \mathbf{x} + \frac{\mathbf{u}}{d_I(\mathbf{x})} \right) - d_I \left( \mathbf{x} + \frac{\mathbf{v}}{d_I(\mathbf{x})} \right)$$

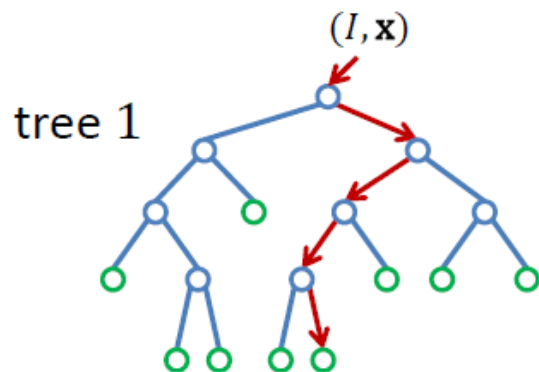
$I$ 为深度图像， $\mathbf{x}$ 为一个像素点， $\theta=(\mathbf{u}, \mathbf{v})$ 为两个偏移向量



$f_{\theta}(I, \mathbf{x})$ 中偏移量除以深度值的目地：使特征对深度具有不变性

## 目标跟踪实例——Kinect人体跟踪

## 随机决策森林训练方法:



$$f_{\theta}(I, \mathbf{x}) = d_I \left( \mathbf{x} + \frac{\mathbf{u}}{d_I(\mathbf{x})} \right) - d_I \left( \mathbf{x} + \frac{\mathbf{v}}{d_I(\mathbf{x})} \right)$$

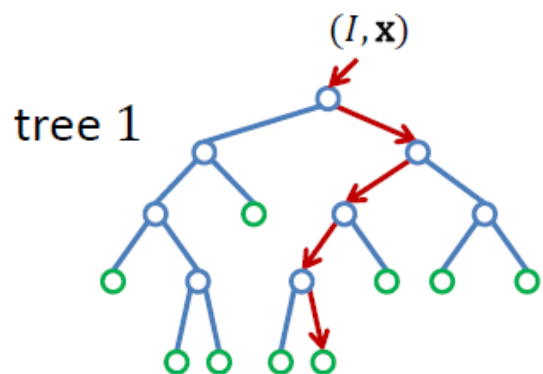
|                                       |                   |
|---------------------------------------|-------------------|
| If $f_{\theta}(I, \mathbf{x}) < \tau$ | go to left child  |
| otherwise                             | go to right child |

**随机选择法:** 在每个节点完全随机选择参数 $\theta$ 和 $\tau$

- V. Lepetit and P. Fua. Keypoint Recognition Using Randomized Trees, **IEEE T-PAMI**, 28(9):1465-1479, 2006.

# 目标跟踪实例——Kinect人体跟踪

随机决策森林训练方法：



$$f_{\theta}(I, \mathbf{x}) = d_I \left( \mathbf{x} + \frac{\mathbf{u}}{d_I(\mathbf{x})} \right) - d_I \left( \mathbf{x} + \frac{\mathbf{v}}{d_I(\mathbf{x})} \right)$$

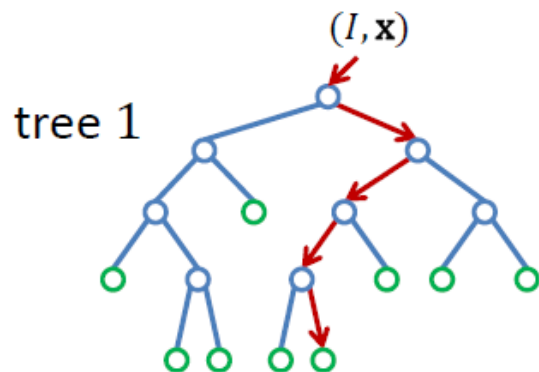
If  $f_{\theta}(I, \mathbf{x}) < \tau$       go to left child  
otherwise                      go to right child

**贪心选择法：** 在每个节点选择能最大化区分样本的参数 $\theta$ 和 $\tau$

- Y. Amit and D. Geman. Shape quantization and recognition with randomized trees. **Neural Computation**, 9(7):1545–1588, 1997.

# 目标跟踪实例——Kinect人体跟踪

## 随机决策森林训练方法：



$$f_{\theta}(I, \mathbf{x}) = d_I \left( \mathbf{x} + \frac{\mathbf{u}}{d_I(\mathbf{x})} \right) - d_I \left( \mathbf{x} + \frac{\mathbf{v}}{d_I(\mathbf{x})} \right)$$

1. Randomly propose a set of splitting candidates  $\phi = (\theta, \tau)$  (feature parameters  $\theta$  and thresholds  $\tau$ ).
2. Partition the set of examples  $Q = \{(I, \mathbf{x})\}$  into left and right subsets by each  $\phi$ :

$$Q_l(\phi) = \{ (I, \mathbf{x}) \mid f_{\theta}(I, \mathbf{x}) < \tau \} \quad (3)$$

$$Q_r(\phi) = Q \setminus Q_l(\phi) \quad (4)$$

3. Compute the  $\phi$  giving the largest gain in information:

$$\phi^* = \underset{\phi}{\operatorname{argmax}} G(\phi) \quad (5)$$

$$G(\phi) = H(Q) - \sum_{s \in \{l, r\}} \frac{|Q_s(\phi)|}{|Q|} H(Q_s(\phi)) \quad (6)$$

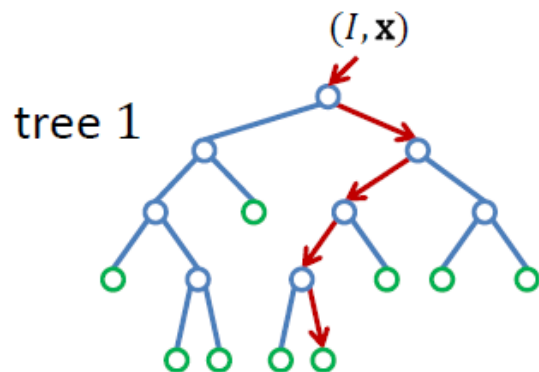
where Shannon entropy  $H(Q)$  is computed on the normalized histogram of body part labels  $l_I(\mathbf{x})$  for all  $(I, \mathbf{x}) \in Q$ .

4. If the largest gain  $G(\phi^*)$  is sufficient, and the depth in the tree is below a maximum, then recurse for left and right subsets  $Q_l(\phi^*)$  and  $Q_r(\phi^*)$ .

- Y. Amit and D. Geman. Shape quantization and recognition with randomized trees. **Neural Computation**, 9(7):1545–1588, 1997.

# 目标跟踪实例——Kinect人体跟踪

随机决策森林训练方法：



决策森林包含3棵决策树，树的最大深度为20，每棵树使用30万幅深度图像、每幅上2000个像素点进行训练，每个节点有2000个候选 $\theta$ 值和50个候选 $\tau$ 值。

在一个1000核的集群上训练1天。

$$f_{\theta}(I, \mathbf{x}) = d_I \left( \mathbf{x} + \frac{\mathbf{u}}{d_I(\mathbf{x})} \right) - d_I \left( \mathbf{x} + \frac{\mathbf{v}}{d_I(\mathbf{x})} \right)$$

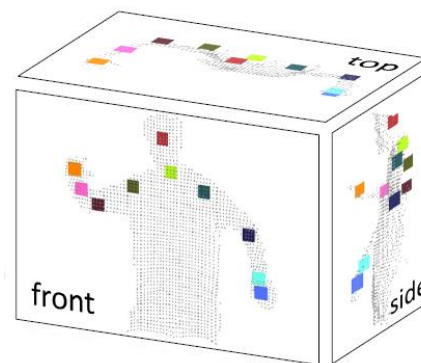
- Y. Amit and D. Geman. Shape quantization and recognition with randomized trees. **Neural Computation**, 9(7):1545–1588, 1997.

# 目标跟踪实例——Kinect人体跟踪

关节点3D位置计算：



推断身体各部分



关节点3D位置计算

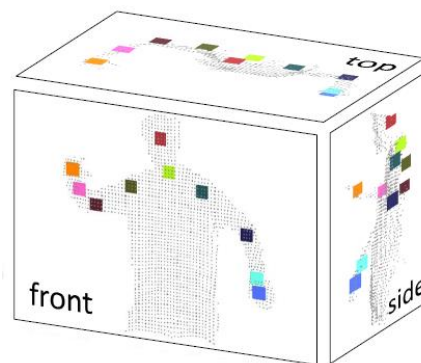
最简单的方法：将深度图中身体每一部分所对应像素点的空间均值作为关节点3D位置，但这一方法易受外点影响。

# 目标跟踪实例——Kinect人体跟踪

关节点3D位置计算：



推断身体各部分



关节点3D位置计算

改进方法：使用加权高斯核函数进行mean shift聚类。



# 目标跟踪实例——Kinect人体跟踪

关节点3D位置计算：

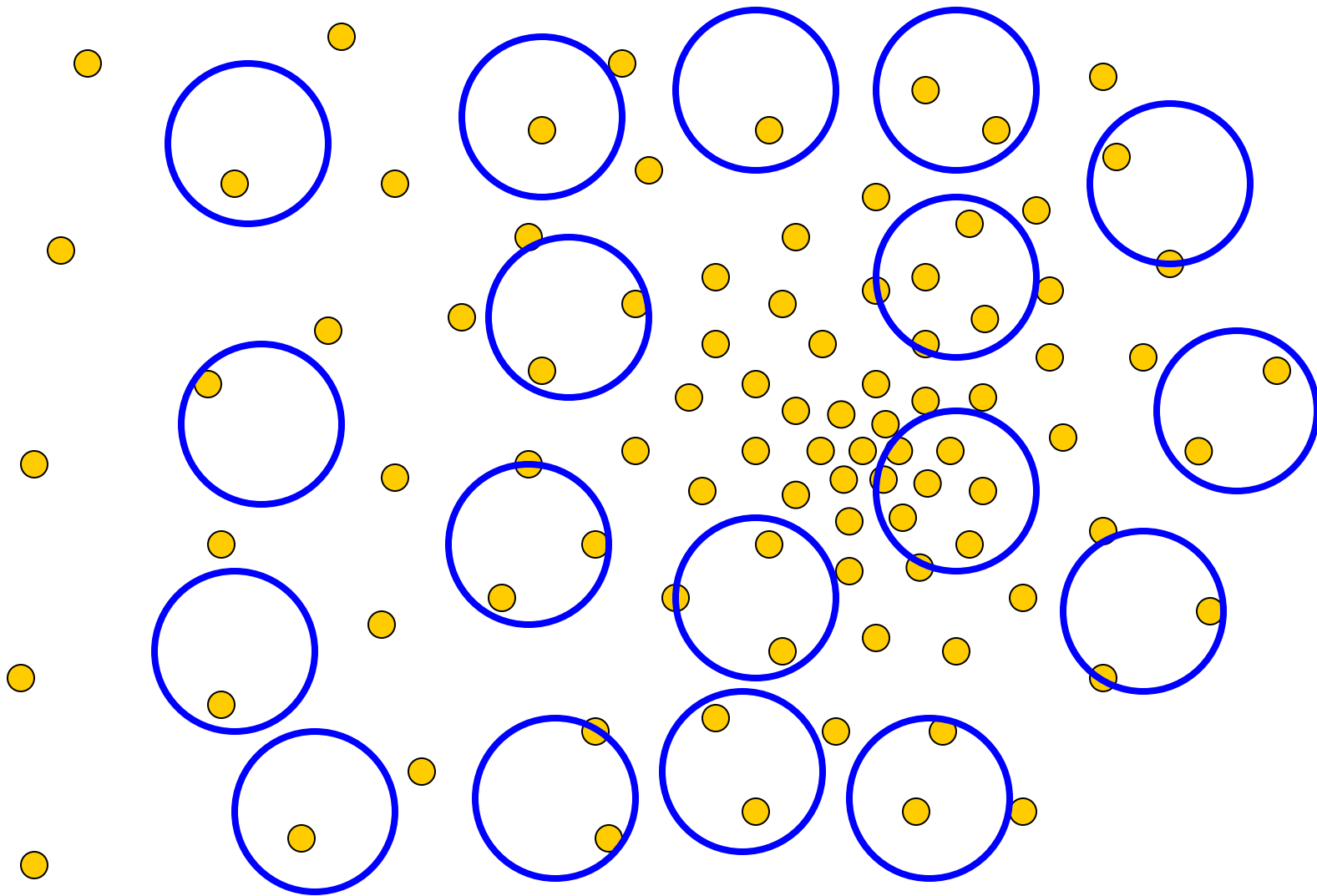
使用加权高斯核函数进行mean shift聚类，即寻找密度分布函数的极值：

$$f_c(\hat{\mathbf{x}}) \propto \sum_{i=1}^N w_{ic} \exp \left( - \left\| \frac{\hat{\mathbf{x}} - \hat{\mathbf{x}}_i}{b_c} \right\|^2 \right) \quad w_{ic} = P(c|I, \mathbf{x}_i) \cdot d_I(\mathbf{x}_i)^2$$

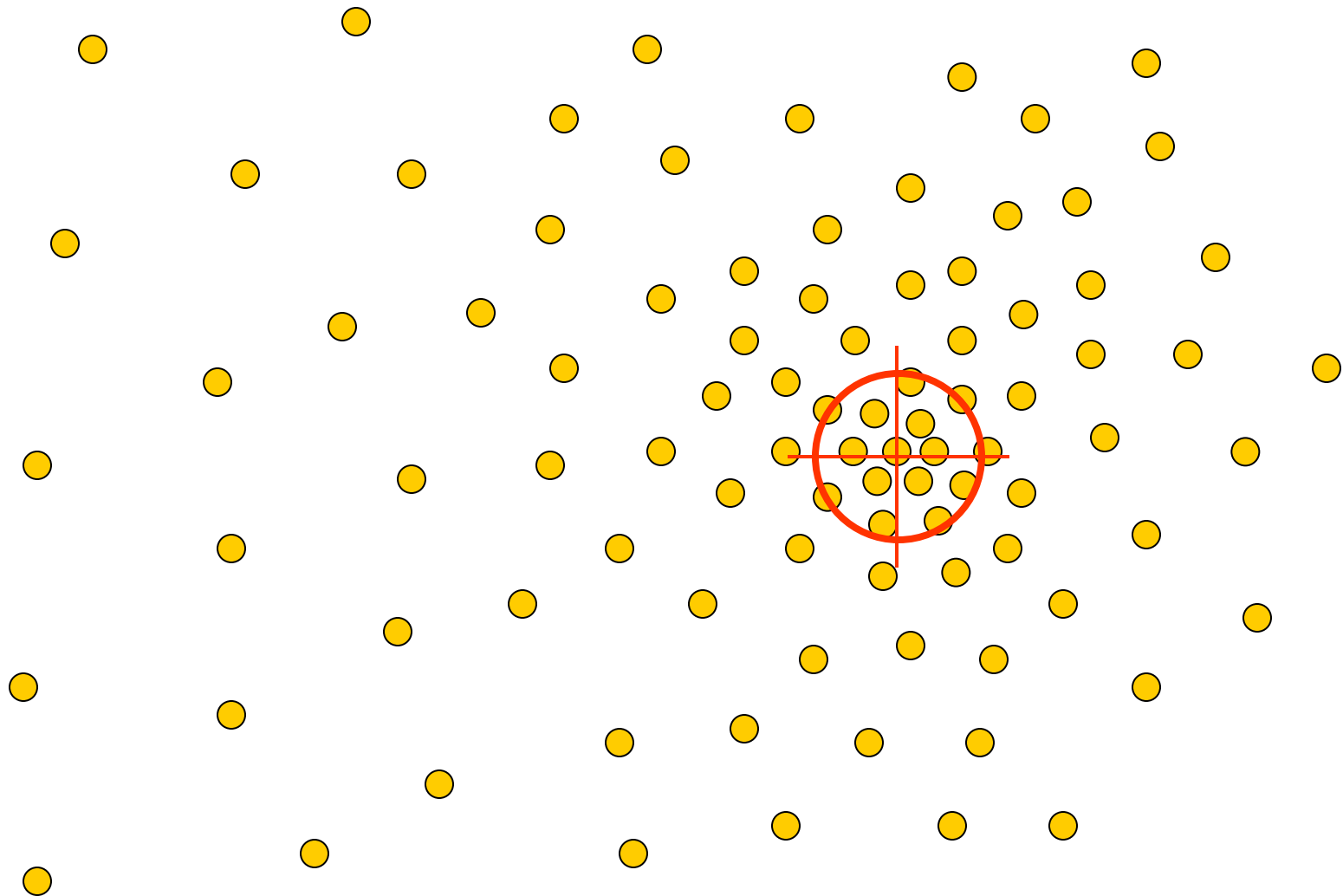
$\hat{\mathbf{x}}$  为待估计的关节点3D位置,  $\hat{\mathbf{x}}_i$  为深度图中像素点  $\mathbf{x}_i$  对应的空间点。

以所有概率估计  $P(c|I, \mathbf{x}_i)$  大于一定阈值的  $\hat{\mathbf{x}}_i$  为初值进行mean shift聚类。

# Mean Shift 聚类



# Mean Shift 聚类



# 目标跟踪实例——Kinect人体跟踪

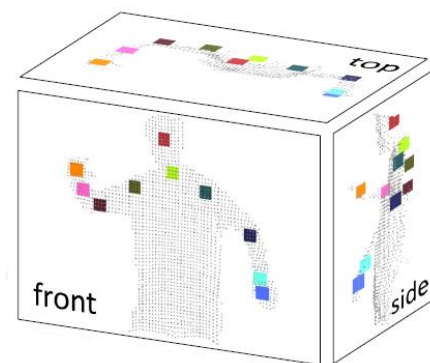
算法的整体计算速度：



输入深度图像



推断身体各部分

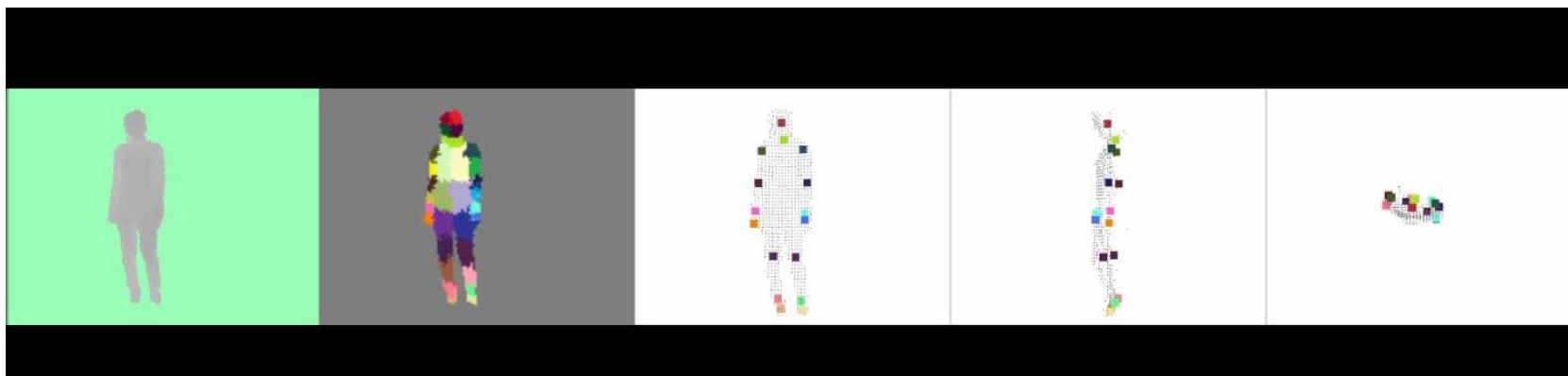


关节点3D位置计算

- 在Xbox 360的GPU上运算速度200帧/秒（使用一般聚类）
- 在8核台式机上运算速度50帧/秒（使用mean shift聚类）

# 目标跟踪实例——Kinect人体跟踪

Kinect人体跟踪结果：



# 总结—本节课内容回顾

目标跟踪的分类：

- 单目标、多目标
- 单摄像机、多摄像机
- 2D、3D
- 刚体、连接刚体、变形体

目标的表达方法：

- 区域、特征、轮廓、模型

跟踪方法的分类：

- 生成式、判别式
- 递归式跟踪、单帧检测跟踪

## 总结—参考文献

- Alper Yilmaz, Omar Javed and Mubarak Shah, **Object Tracking: A Survey**, ACM Computing Surveys, 38(4): 1-45, 2006.
- Vincent Lepetit and Pascal Fua, **Monocular Model-Based 3D Tracking of Rigid Objects: A Survey**, Foundations and Trends in Computer Graphics and Vision, 1(1): 1–88, 2005.
- Jonathan Deutscher and Ian D. Reid, **Articulated Body Motion Capture by Stochastic Search**, International Journal of Computer Vision, 61(2): 185–205, 2005.
- Jamie Shotton, et al., **Real-Time Human Pose Recognition in Parts from Single Depth Images**, CVPR, 2011.