Modeling 3D Environments through

Hidden Human Context

要解决的问题：1.物体检测，2.物体放置

解决办法：Infinite Latent Condition Random Field(ILCRF)

主要思想:将human pose考虑在内

一般的CRF，把每一个物体作为一个node，物体之间的关系作为边

ILCRF；物体作为可见node，人作为潜在的node（latent variable），object affordance作为边，使用potential function表示，连接human和物体的label，由于每个场景中人的个数是不固定的，使用Dirichlet Process来获得，

然后用maximum a posteriori(MAP)来学习object affordances，Gibbs Sampling来学习样本label

Scene arrange,训练阶段，多学了object-object结构

Multi-Target Tracking by Discrete-Continuous Energy Minimization

问题：Muiti-target主要分为两个子问题：1.识别视频中的target（data association），2.得到target的运动轨迹

解决方法：提出一种方法能将data association 和trajectory estimation统一用一个模型来表示，提出了“minimization of a consistent discrete-continuous energy.”

使用连续的曲线来表示轨迹，使用离散的多labelCRF模型来表示data association

Social Grouping for Multi-Target Tracking

and Head Pose Estimation in Video

问题：multi-camera监控摄像头中的Muiti-target tracking 和头部姿势估计

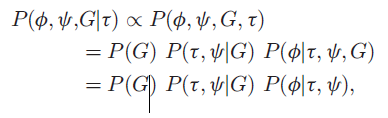
解决办法：social grouping 人们会成群结队的走，而且轨迹速度目的地相同，更有可能互相看或看同一个地方

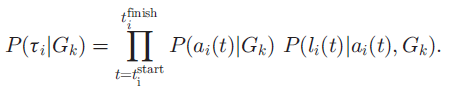
需要解决两个MAP问题：

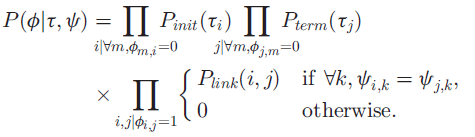


第一个MAP问题：

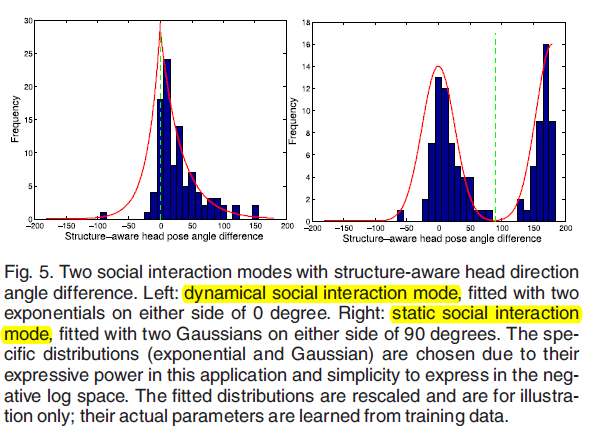
分块来看



表示的是Social Grouping算法得到的相似度，其中分别表示group k出现在camera ai(t)的可能性，后面表示t时间时ai中的group会出现在位置li(t)的概率,使用Gaussian表示，中心是uk,a(t)



算法命名为SGB(Social Grouping Behavior)



Analysing Domain Shift Factors between Videos and Images for Object Detection

目标：比较视频和图片在物体识别中效果的不同

评价标准：得到的结果和gt相比>50%算正确，一类的performance是有平均Precision（AP）来定的，整体performance由中值AP（mAP）来定的

detector：DPM和RCNN

步骤：

1. 提出了个公式来量化factor的影响
2. 找到方法使两个trainingset在公式上更加相似
3. 对比detector在处理前后的表现效果

能够量化由于在不同领域training而产生的performance gap，用以上方法对每一个factor进行处理

Spatial Location Accuracy

TrainingSet分为四种，对视频使用自动识别的算法PRE，FVS，以及手动人工标注的视频和图片

Appearance Diversity

手动将视频set中相同物体，不同viewpoint和遮挡的图片分为一组，appearance diversity用unique samples的数量来表示，修改方法，删掉重复的samples并且让两个domain大小相同（YTO因为重复会删掉很多，VOC因为要和YTO数量保持一致而删掉很多）

影响：VOCtest时，使用VOCtrain在减少了sample之后，mAP有了明显的减少，因为sample减少了一半左右，使用YTO则没有明显变化，因为只有重复的被删除了。YTOtest结果相似。说明相近的samples并没有带来过多的信息，这说明仅需要对视频每一个shot选择一部分很少的镜头就可以了。 同时，在testVOC差距缩小,testYTO差距增大，可以理解，因为对VOC来说效果变差了，而对YTO没有影响。

Image Quality:使用的数据集市Unique Samples

Measurement：计算一个bounding-box里HOG cells的梯度大小的总和，并用size来泛化.平均gradient energy:VOC 4.4 YTO 3.2，原因 视频压缩，运动模糊，低色彩对比度

Equalization：将VOC进行模糊，使得二者的模糊程度相同（Gaussian blur and motion blur）,motion blur的参数（sigema和K）针对每一类都不同，使得gradient energy两类相同

Impact：对两个dataset进行模糊，结果对trainVOC模糊的影响更大，trainYTO对结果影响较小，因为本来YTO已经有了motionblur，再模糊不会丢失更多的信息

两种blur都能减小test VOC 的gap。对testYTO来说motionBlur会增大gape

Aspect Distribution: 在两个领域中一样类型的比较少

Measurement： dKL

Equalization：建立子集，需要找到最大的自己使得满足不等式，使用greedy forward selection algorithm，同时不断向里加欧氏距离最小的一对sample