

指导教师： 杨涛 提交时间： 2015.3.28

The task of
Digital Image Processing

数字图像处理

School of Computer Science

No: 1

姓名： 李妙鹏

学号： 2012302430

班号： 10011202

遮挡人群剖析

Golnaz Ghiasi Yi Yang Deva Ramanan Charless C. Fowlkes

Dept. of Computer Science, University of California, Irvine

{gghiasi, yyang8, dramanan, fowlkes}@ics.uci.edu

摘要

由于遮挡模式的多可能性，遮挡问题使得目标识别变得相当困难。本文提出了，采用局部混合模版的非参数遮挡建模方法来产生大量的训练数据。本文提出的模型可以学习不同遮挡模式的外部特征等真实的信息，比如：遮挡部分的轮廓以及相邻遮挡部分的同现数据。底层的交叉结构使得此模型可以捕捉目标相邻部分之间的一致性并且能够很好的预测目标-背景-遮挡物之间的界限。本文在严重遮挡的情况下生成的人体姿态估计模型，实验结果充分地显示了其很好的定位精度。

1. 引言

在如图 1 所示的复杂场景中，遮挡对一个高性能的目标识别系统带来了巨大的挑战。即使在遮挡物的类型已知的情况下（例如，拥挤的街道中的其余人），在一个大型的复杂场景中，被部分遮挡的目标与遮挡物之间的相关布局是不受约束的。已有的许多目标检测和姿态估计研究方法大多在一个假定的遮挡区域而忽略了遮挡物和图像证据。然而，由于没有规律的外部特征或弱分辨力使得这种方法很难进行目标检测。

本文的旨在建立一个可以提取遮挡图像特征的外观模型，而不是简单地忽略它们或者将姿态估计与分类结合。例如，场景、遮挡物的轮廓形状以及自遮挡的初始外形等等都为遮挡问题的研究提供了有力证据。因此，本文利用局部模型，在这些模型中，用大量不同的模版和与之相关的分割来表示局部外形。本系统能够预估场景、目标的



图 1，这幅图像描述了遮挡造成的底层特征描述符的场景。本文旨在对这样的外观特征进行建模，通过用大量的含有遮挡数据的局部混合组成训练模型，进而进行姿态估计和可见性预测。

姿态以及包含人、背景、和遮挡物标记的分类模式。

不幸地，这种高维模型的全连接训练需要大量表示遮挡模式的人工分类的训练数据。鉴于无法得到这些训练数据，本文通过综合生成数据的方法克服了这一困难。本文生成了成千上万个局部被遮挡目标的图像来训练可变的人体模版。每一个生成的样本都带有目标和遮挡物的全部信息。本文采用遮挡物的分割模式作为监督信号将可能的遮挡情况分类并通过枚举法推测遮挡类型。由于本文可产生无限多的训练数据，对少有的遮挡模式的建模仅仅局限于计算问题。

本文所建立的模型是以图结构的分类训练方法来推理和学习的。特别的，我们的模版训练与使用局部混合的局部模型非常相似。然而，捕捉遮挡状态需要量级的有序参数和数量级的训练数据。这使得在学习和推理使得计算复杂度很高，现有的代码【26】需要两周的计算时间。本文做了几点优化使得这种学习方法是可实现的。

文章最后阐述了遮挡模型的作用，使用 H3D【2】中的图像和“*We are Family*”数据集【7】分析遮挡对模型性能的影响。为

了评估定位精度，本文建立遮挡外观的模型，结果显示，此模型比已有的姿态估计技术精度要高。

2.相关工作

本文假定许多遮挡模型的检测的不足在于它们没有对可视化的遮挡外观建模。本文没有把目标的遮挡部分与背景或非目标像素统一建模。这意味着如果某一部分的均值低于阈值，这部分就会被视为遮挡，如果阈值过高，非遮挡部分就会被视为遮挡，如果阈值过高，遮挡目标容易误认为是背景。因此，本文认为只有当有图像证据去支撑时才可以假定为遮挡。

遮挡模型：一种成熟的方法是在测试时将可见度视为一个二元变量。建立局部遮挡模型适用于有遮挡部位详细信息的前提下。例如，Weber et. al.的模型和 Fergus et. al.枚举所有可能遮挡情况。【1】中的模型监督部分包含了遮挡类型的模版但是没有不同部分可见度的成对约束。【12】中描述的基于规则的模型也包含了遮挡类型的模版的但是更多地强调遮挡模式的结构，将人体检测具体化为跟随遮挡者动态可变的纵向链。语法的实现与本文中局部混合模型形式相似。规则使得模型之间的参数模式可共享。与本文模型最大的不同在于【12】要求遮挡模式必须可以被设计为规则。本文所描述的方法避免了这一结构学习问题，而是以非参数的形式从数据中自动学习遮挡模式。

部分程度遮挡的一个缺点在于不能够在部分图像中捕捉微型遮挡模式。一个交叉技术将遮挡原因附加到构成模版的图像特征图或者单个像素【21, 11, 22】。这些采用带有像素灰度的马尔可夫随机领域理论的模型强化了空间一致性。然而，在自然场景中，遮挡模式的空间数据是变量并且与环境 and 图像的集合特征有关（参考【13】中的结果）。本文的建模方法在不同的局部模版上描述遮挡但是在部分而非像素或者 HOG 元的层次上强调一致性，这使得遮挡之间的结构独立性可以适应连续的形状。因此，本



图 2，本文在一个原始图像上选取随机的位置放置目标来综合生成大量的训练数据。遮挡物的位置的尺寸与被遮挡目标之间通过一个弱初步假设相关联。如果与原始图像的的关键点注释以及分割相结合，就能生成无限多的包含关键点可见度和分割掩膜的强监督训练数据。

文模型利用树结构分布的有点学习遮挡的空间数据，提高了枚举的计算高效性。

图像解析：典型的转变是从但目标检测到全图像的解析。场景中的遮挡现象可以用遮挡的目标开解析。例如，【25】采用层次分割模型在像素级上解析探测目标之间的遮挡。【11】强调 HOG 元任务之间的互排斥。【3】中采用与检测器相联系的形状模版之间的平滑方法。本文的工作受多目标交互对复杂外观建模的方法启发。包括捕捉目标间遮挡交互的多人【24,7,20,16】或其他的多目标【18,5】模型。

图像解析方法固有的困难在于需要检测遮挡的目标。现实世界中，遮挡者可以使任意的，使得目标检测器必须在各个图像上进行训练和评估。多目标的视觉检测器的学习对小范围的人与目标交互也许是可行的，但是对于交叉弱约束的遮挡而言似乎是不可能的。只有有限的方法可以解决这一问题而更多的方法却回避了这个问题。本文的建模方法的复杂性在于单目标和多目标模型。本文为单目标分类训练检测器独立于场景中的其他检测运行。然而，采用不同方法给遮挡物建模是，依赖于大量的训练数据库来获取遮挡外观的数据。

综合训练数据：一些文章已经提出利用训练系统中的综合数据来识别和姿态估计。【19】中，为了实现最近姿态回归，作者采

用大量的渲染姿态生成连接空间。【15】采用绿色背景来增强杂乱背景下中真实目标的训练数据并生成了三百万真实背景中综合图像的数据集。我们的工作在于基于图形的渲染是不同于剪切和粘贴已有的图像来生成一个新图像。与【17】非常相似，将三维模型应用到真实图像，并通过稍微打乱拼接角度生成了综合的渲染图像。

3. 局部遮挡模式建模

类似于【26】中的可变局部模型，本文利用局部混合的图结构对遮挡人群的外观建模。本节描述了如何为每一部分生成一个局部混合标签。下一节描述外观模版如何通过学习得到以及拼接成整个模型。

生成综合图形：原始的训练数据集已经注释有局部信息和真实的模版，本文通过合成这些数据集上的分割目标产生大量的遮挡数据。这一过程自动产生了带有监督信息的遮挡外观范例，其中，监督信息包括遮挡目标在像素级上的支撑。本文在测试时采用H3D数据集，它提供了分割模版以及大约1500人的连接方式。基于原始图像中的目标注释，本文按照比例生成真实的空间分布（例如，人的头不可能被别人的脚遮挡）。使遮挡物的底部低于原始图像并且按照纵轴的线性函数进行缩放变换。图2显示了这种训练图像的样本。

局部外观学习：本文将这种含有大量注释的综合训练数据转化为寻找可以捕捉不同姿态和局部遮挡情况外观特征的训练样本聚类。生成大量的聚类是一大难点；典型的方法有基于图像修补【6】，关键点注释【2】，甚至手工分组【27】等，这些方法显示，只有增加聚类数量才可以提升系统性能。这些困难是由于数据的不充足（有限的数据集，随着聚类的增加每个聚类的训练数据减少）以及不合适聚类指标。我们发现，由于遮挡物的任意放置以及外观造使得遮挡模式的空间特征是高维的，简单的采用基于外观的聚类方法的性能不高。

综合训练数据采用两种方法解决了这些困难。第一点，生成综合训练数据可以增

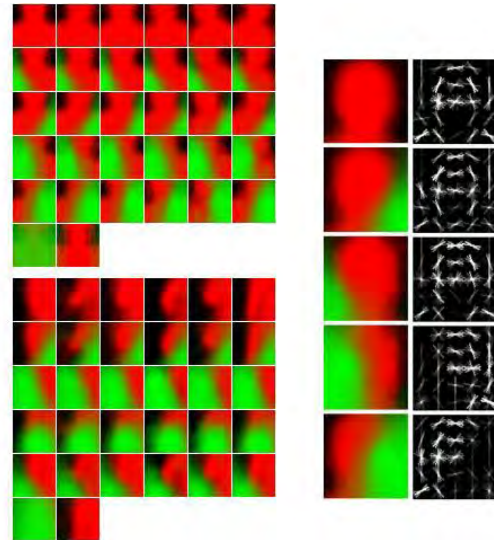


图3，本文利用一个可获取姿态变量（关键点位置聚类）和遮挡模式（分割聚类）的因子模型对局部外观进行分类。本模型也包括全遮挡和自遮挡类型。左边显示的是脖子和肘的聚类样本。每行对应一种遮挡模式，每列对应一种姿态。颜色表示与其他聚类的平均分割掩膜关系。右图显示了模版在不同的头部遮挡情况下的可见性。与没有遮挡的头相比，含有局部遮挡的模版在遮挡轮廓周围含有更多的边缘，这些信息有助于检测。

加每帧的训练数据。第二，采用带有形如遮挡物-目标-背景这种监督信息的综合数据为聚类提供了更强的依据。

含因子的遮挡姿态分类：本文利用每一个关键点对训练片段单独分类。采用集合姿态特征 g_i 以及图像-背景-遮挡物分割 o_i 来标记第 i 个训练片段。姿态特征向量在图结构中描述了某一部分与相邻部分的空间特征。

分割 o_i 是三个表明目标分割关键点的二元掩膜的集合（如图2所示）。一种简单的聚类算法是将多行索引描述符标准聚类算法（e.g., K-means）和两种特征的相关测量结合起来。然而，由于所有的训练数据都有严重遮挡，造成了训练集偏差。我们不希望这种偏差严重影响分类。进一步，从生成的角度来看，我们希望遮挡模式和目标姿态是相对独立的（一种非常重要的例外）。基于此，本文采用一种因子聚类算法。利用 K-means 算法以及几何姿态特征生成 K_g 个聚类。其

中，这些聚类不受综合训练数据数量的影响。本文也生成了遮挡物独立的 K_o 个聚类。

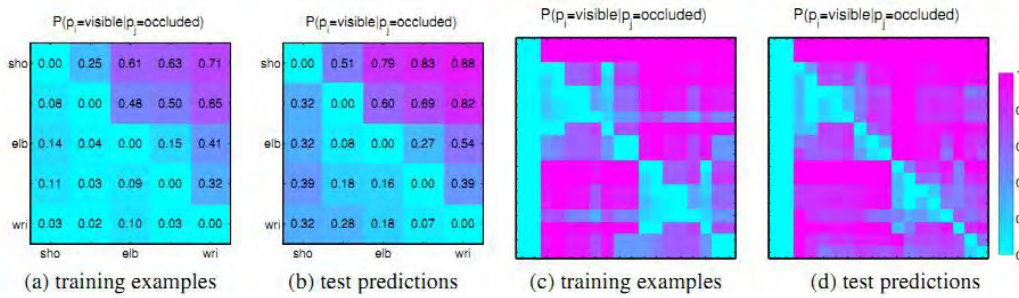


图 4，此图反映了各部分遮挡状态（可见度）的同现性，第 j 列反映了在 j 部分是被遮挡的情况下第 i 部分是可见的可能性。图(a)和图(b)反映了 5 部分 32 种混合模式对应的同现性。图(a)给出了测试数据的真实可见度值，图(b)给出了根据模型在测试数据上生成的标记数据。随着可见度增加，通过 SVM 训练产生的数据与偏差之间有很好的匹配关系。图(c)(d)反映了上半身模型的相似数据。其中，显著地块结构对应着头部，以及身体的左右半部分。在一个肢体内，上臂（例如肩）遮挡使得下臂也不可可见，然而，手腕的遮挡对上臂的可见性却无强约束力。

最终将每一个训练样本与 $K_g \times K_o$ 个交叉聚类空间中的一个元素或全遮挡或自遮挡相匹配。

图三显示了几种不同目标部分的分类示意图。每一行对应不同的遮挡聚类，每一类对应不同的姿态聚类。本文也包含了两种特殊的聚类，一个全遮挡和一个自遮挡聚类。由于全遮挡仅仅包含来自遮挡物的图像特征，因此假定全遮挡的外观与姿态相对独立。一个样本属于自遮挡聚类是指即使图像中没有遮挡物，图像中的一部分也是不可见的（例如，手放在头上）。多姿态聚类有利于自遮挡但是需要大量的训练数据而我们的合成方法不能自动的生成这样的自遮挡。

聚类数据：部分遮挡是不独立的。如果一个人的肘被另一个人的肘挡住了，那么他的肩膀可也被挡住了（被那个遮挡者的肩膀）。这表明相邻部分之间的聚类标签有一定的同现数据。图 4 和图 5 将样本中的这些数据可视化。

4.遮挡局部模型

接下来阐述一种用成千上万的人体姿态图像（有严重遮挡）训练可变图结构的方法。

可变模版：本文的模版有片段集合 V 以

及带有局部位置和外观约束的双元关系 E 组成。设 I 是一个图像 $p_i = (x, y)$ 是 $i \in V$ 部

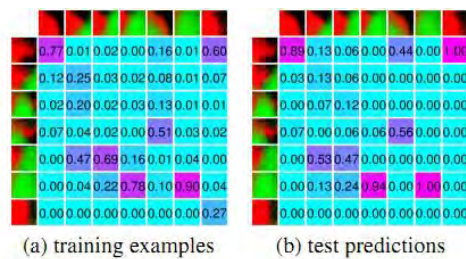


图 5，图(a),(b)分别给出肩膀在真实的训练数据和模型利用测试数据选择的混合模式下，肘状态的条件概率。模型可以学习分割模式之间的一致性。例如，如果左肩膀被遮挡（第 5 列），那么左肘要么课件（第 1 行）要么被遮挡（第 4 行），第 7 行对应于自遮挡。

分的位置， m_i 是第 i 部分的混合组成。每个 m_i 与从该部分训练数据学习得到的遮挡-姿态聚类相关联。如第 i 本分与左肘相关联，那么 m_i 就选取肘方位的粗糙集和特定的遮挡模式（包括全遮挡和自遮挡）。 m_i 的每种选择与人体--地--遮挡物掩膜的均值有关，它可以再测试阶段预测关键点的可见性和

分割。

给定一幅图像，可采用下式计算采集的假定位置和混合选择：

$$S(I, p, m) = \sum_{i \in V} [\alpha_i^{m_i} \cdot \phi(I, p_i)] + \sum_{ij \in E} [\beta_{ij}^{m_i, m_j} \cdot \psi(p_i - p_j) + \gamma_{ij}^{m_i, m_j}]$$

第一部分计算使用第 i 部分 $\alpha_i^{m_i}$ 模版的外观证据来调整 p_i 位置的混合组成 m_i 。

$\phi(I, p_i)$ 表示从 I 图像 p_i 位置提取的特征向量。注意，本文为每一个混合组成定义各自的模版，这种模版的遮挡状态可以获取与遮挡有关的视觉特征。

第二部分计算两部分之间的关系约束， $\psi(p_i - p_j) = [dx \ dx^2 \ dy \ dy^2]$ 是第 i 部分和第 j 部分之间的关系向量。参数 $\beta_{ij}^{m_i, m_j}$ 表示相关的其余位置的二次偏差惩罚系数。波动和偏差都依赖于从第 i 部分和第 j 部分选取的混合组成 m_i 和 m_j 。这使得此模型可获取相邻部分可见度之间的独立性（如图 4），相比于单模版，这种模型提供了更充足的对姿态和外观的非参数描述。

学习和推理：给定一个测试图像，本文寻找分值最大的部分分配 p 以及混合组成。在 E 是树结构的情况下，这种方法可以采用动态程序【9,26】高效的计算。

设 (p^n, m^n) 表示第 n 个正样本的真实的部分坐标和混合组成标签。利用在结构化的 SVM 中的一个变量来学习得到模型参数 $w = (\alpha, \beta, \gamma)$ 。

$$\arg \min \quad \frac{1}{2} \|w\|^2 + C \sum_n \xi_n \quad (2)$$

$$s.t. \quad \forall n \in pos \quad w \cdot \Phi(I^n, p^n, m^n) \geq 1 - \xi_n$$

$$\forall n \in neg \quad \forall p, m \quad w \cdot \Phi(I^n, p^n, m^n) \geq -1 + \xi_n$$

上述的二次规划（QP）试图通过学习使正样本得估值大于去 1（对真实的部分坐标和混合标签评价），负样本的估值小于 -1（任意组合的部分坐标和混合模式）。本文使用 ige 标准的截面方法通过利用负样本运行检测系统逐渐增加负约束条件以找到最佳的约束集合。

误判的正样本：为了更好地实现定位和遮挡预测，本文将误判的正样本添加到负样本。这样，在二次规划中增加如下可负约束条件：

$$\forall p, n \in pos, m \neq m' \quad w \cdot \Phi(I^n, p^n, m') \geq 1 - \xi_n \quad (3)$$

如果，两个部分混合 m 和 m' 的对相同部分集的预测是可见的，我们就认为这两部分属于同一类。用记号 ξ 表示混合组成 m 在不同等价类。这一约束使得对于给定的正样本 n ，所有与错误的混合方式相关姿态的遮挡预测值低于 -1，这些约束与传统的结构型预测（正确标签估值高于错误标签）相似，分为正约束和负约束。我们发现，这种分类在没有降低性能的情况下加速了优化。

半隐藏学习：尽管正训练数据给出了部分位置和混合标签。部分定位的重估计是非常有用的。对于一些遮挡区域而言，在遮挡物的位置学习可区别的外部特征（遮挡轮廓）比在部分关键点定位依赖性更强，这种方法就显得特别重要。

本文采用标准的学习方法【8】利用最凸优化和区域集合位置的 p^n 重估计来训练模版。对于每一个正样本 n ，设 $\Omega_n = \{p : |p_i - p_i^n| < r\}$ 表示一系列于真实位置 p^n 相近的可能位置。本文通过下列的坐标下降算法来学习模型 w 并更新部分位置 p^n ：

1.模型更新：利用参考的正样本位置 p^n 通过二次规划学习 w

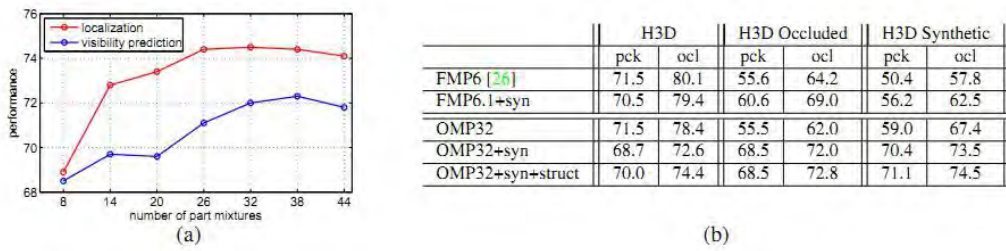


图 6, (a) 图表示利用综合的遮挡数据, 定位和可见性预测精度与混合模型组成树之间的函数关系。不难发现, 系统性能在混合组成数为 32 时达到最佳。(b) 反应了不同方法在 H3D 测试图像上的性能, H3D 数据集包含了严重遮挡和一系列综合遮挡样本。在综合和严重遮挡的数据集中, 遮挡建模的优点更加突出。FMP6.1 是一个典型的单遮挡基准模型, 因此, 可用来生成综合的训练数据。

2. 隐式任务: 计算

$$p^n = \max_{p \in \Omega_n} w \cdot \phi(I^n, p, m^n) \text{ for } n \in pos$$

计算瓶颈: 隐式的 SVM 训练【8】瓶颈在于第一步, 需要传递大量的负样本(一千个)并进行负样本挖掘。在我们的数据集里有成千上万个正样本, 因而本文的瓶颈在于

5. 实验结果:

数据集: 本文使用 H3D【26】作为基本的训练和测试数据来源。由于 H3D 数据集包含了许多不同视角具有挑战性的姿态而我们的基准模型采用非正面姿态, 我们选取了 668 张正面人物图像, 每一个原始训练图像用 100 个不同的遮挡来增强, 生成了 50 万个正样本。本文采用来自 INRIAPerson【4】的负训练样本。另外, 我们采用” We Are Family” 数据集(WAF)和基准【7】测试了我们的模型。为了更好地了解模型性能, 我们采用含有很多遮挡的测试数据集, 在 H3D 的三个不同的变体中进行评估。原始的 190 个测试图像, 其中 60 个含有严重遮挡(许多不可见的真实关键点)以及遮挡类型。针对 WAF, 我们考虑与可见关键点成比例的六个数据子集。在 WAF 数据集中, 六个部分(头, 主躯干, 上臂, 右臂)用一个可见变量做标记。根据整体的可见部分建立六个子集。

评估: 本文使用【26】中的正确定位的关键点的百分比(PCK)来评估 H3D。当预测的关键点与真实关键点的距离在一个

第二步。一个正样本更新的标准方法是在每个正样本上评估模型 w 检测器。一个简单但是关键的观测结果是指混合标签 已知但没有隐式更新的结果, 因此每部分只有一个单滤波器需要评估。修改【26】中的代码加速半隐式更新, 将训练时间从一周减少到一天之内。本文表明, 隐式任务可以被优化。归一尺度范围内的时候认为预测是正确的。为了对遮挡人体部分扩展 PCK 标准。本文要求按照此模型任意一个真实存在的被标记为不可见的关键点必须被正确的预测为遮挡。为了区分定位错误与可见性预测错误, 本文将计算局部可见性预测作为二元分类任务。

针对 WAF, 本文采用正确定位部分的百分比(PCP)来评估【7】, 类似于(PCK), 但是测量的是块而非点的位置。类似于 H3D, 我们也测量了真实标记的基于块可见性的遮挡预测。利用 WAF 提供的一个组合的面部和肢体检测器, 我们的一套上半身检测器的检测率为 86%。

模型复杂度: 为了给每部分选择一定数量的混合组成, 我们在区别 个哦每部分产生 44 中混合组成的前提下评价系统性能。图 6 显示了测试位置和可见性预测精度与混合组成数量之间的函数关系。选取 $K_o = 5$ 时, 对应 32 中混合组成并且性能和运行时间之间得到了很好的平衡。增加训练数据的数量似乎并不改变饱和点。事实上, 在很多情

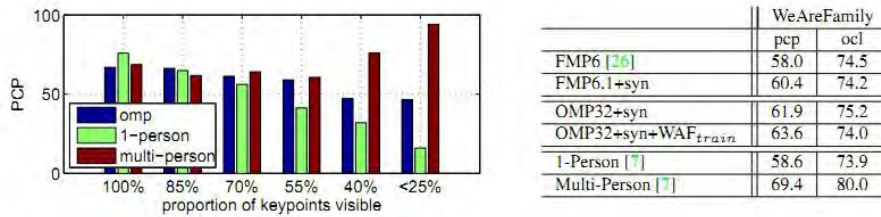


图 7, 在 WeAreFamily 数据子集中测试的系统性能与遮挡程度之间的函数。本文模型 (OMP) 能更好地利用遮挡样例使得性能明显优于【7】中的单人模型。对于严重遮挡 (遮挡 50% 以内的关键点), 【7】中的相对复杂的多人模型性能更好一些, 但是, 与单人模型一样, 定位精度大大下降。表格给出了以上所有模型性能参数和遮挡预测精度。

况下, 使用 10x 个数据和 100x 个数据的性能是相近的。这表明, 我们可以增加遮挡者的形状种类来进一步增加混合组成的数量。

综合训练数据: 图 6 显示了在不同训练和测试条件下 32 中混合模型的性能 (OMP32)。在原始的 H3D 数据中没有综合遮挡数据 (OMP32) 表现出很好的性能而且包含综合遮挡 (OMP32+syn) 的数据显示了本质的提升。在含少量的原始数据集上利用误标的正样本训练模型 (OMP32+syn+struct) 定位精度有了很大的提高。

性能比较: 图 6 比较了本文建立的模型 (OMP32+syn+struct) 与其他几个模型的性能。FMP6 是基于【26】中的代码。在 H3D 数据集上训练 FMP 的模型也可以预测所有可见的部分。在不包含遮挡数据时训练得到的模型性能较差 (PCK=69.6)。为了使 FMP 能够预测可见性, 本文用额外的遮挡混合组成对每一部分进行训练。这个模型在测试阶段可以预测遮挡并且提高了精度。仅仅使用一种遮挡混合组成, 其他的数据使得在遮挡预测时性能进一步提高但是确实与定位精度为代价的。

图 7 在 “We are Family” 数据集上比较了包括 FMP, 单人或多人模型【7】在内的几个模型的性能, 通过用综合数据训练 OMP 模型, 尽管在完全不同的数据集上, 我们的模型性能由于单人模型。包括 WAF 正样本在内的训练数据集使得 OMP 模型的性能进一步提高。实验也测试了系统性能与遮挡程度之间的函数关系。对于所有或者大部分全遮挡, 本文模型与多人模型的 PCP 相似。由于多人检测器才有连续推理而不是

检测器输出采集并且为了寻找目标深度信息合并了其他图形线索, 因此, 这个结果是非常出乎意料的。

6. 结论

本文阐述了一个遮挡建模的方法, 意在显示的学习遮挡模式的外部特征及数据。相比于已有的方法, 本系统生成了一个在严重遮挡情况下更为鲁棒的模型。除此之外, 本文的模型可以显示表示遮挡区域, 因此不仅能够区域位置而且可以预测局部分割掩膜。相比于多参数的方法, 综合训练数据和含有大量外观混合的动态模型有点野蛮。然而, 它显著地优势在于能够区别学习, 并且, 众所周知, 它不仅可有学习详细的遮挡数据, 也能够人体姿态估计方面有非常好的性能。

7. 致谢

这项工作得到了 NSF IIS-1253538 和 IIS0954083 的支持。

参考文献

- [1] H. Azizpour and I. Laptev. Object detection using strongly supervised deformable part models. In ECCV, pages 836-849. 2012. 2
- [2] L. Bourdev and J. Malik. Poselets: Body part detector trained using 3d human pose annotations. In CVPR, 2009. 2, 3
- [3] T. Brox, L. Bourdev, S. Maji, and J. Malik. Object segmentation by alignment of poselet activations to image contours. In CVPR, 2011. 2



图 8, H3D 数据集中的姿态样本以及遮挡估计图。每张图显示了关键点的位置 (虚线表示遮挡), 变形的 HOG 模版的可视化以及通过组合单个部分混合模式得到的遮挡物—人体—背景分割图。这 18 张图都代表了 32 中混合 (遮挡) 状态产生的 3218 中可能的遮挡模式之一。前两行显示了来自 H3D 数据集的样本, 后两行来自 H3D 经综合分析得到的数据。

- [4] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. In CVPR, pages 886 – 893, 2005. 5, 6
- [5] C. Desai and D. Ramanan. Detecting actions, poses, an objects with relational phraselets. ECCV, 2012. 1, 2
- [6] S. K. Divvala, A. A. Efros, and M. Hebert. How important are deformable parts in the deformable parts model? I ECCV Parts and Attributes Workshop, 2012. 3
- [7] M. Eichner and V. Ferrari. We are family: Joint pose estimation of multiple persons. In ECCV, pages 228 – 242. Springer 2010. 2, 6, 7
- [8] P. Felzenszwalb, R. Girshick, D. McAllester, and D. Ramanan. Object detection with discriminatively trained part based models. IEEE TPAMI, 32(9):1627 – 45, Sept. 2010. 5
- [9] P. Felzenszwalb and D. Huttenlocher. Pictorial structures for object recognition. IJCV, 61(1):55 – 79, 2005. 5
- [10] R. Fergus, P. Perona, and A. Zisserman. Object class recognition by unsupervised scale-invariant learning. In CVPR, 2003. 2
- [11] T. Gao, B. Packer, and D. Koller. A segmentation-aware object detection model with occlusion handling. In CVPR, pages 1361 – 1368, 2011. 2
- [12] R. B. Girshick, P. F. Felzenszwalb, and D. A. McAllester. Object detection with grammar models. In NIPS, pages 442 – 450, 2011. 2
- [13] E. Hsiao and M. Hebert. Occlusion reasoning for object detection under arbitrary viewpoint. In CVPR, pages 3146 – 3153, 2012. 2
- [14] C. Ionescu, D. Papava, V. Olaru, and C. Sminchisescu. Human3.6m: Large scale datasets and predictive methods for 3d human sensing in natural environments. Technical report, Institute of Mathematics of the Romanian Academy and University of Bonn, September 2012. 3
- [15] Y. LeCun, F. J. Huang, and L. Bottou. Learning methods for generic object recognition with invariance to pose and lighting. In CVPR, volume 2, page 97, 2004. 3
- [16] W. Ouyang and X. Wang. Single-pedestrian detection aided by multi-pedestrian detection. In CVPR, 2013. 2
- [17] L. Pishchulin, A. Jain, M. Andriluka, T. Thormahlen, and B. Schiele. Articulated people detection and pose estimation: Reshaping the future. In CVPR, pages 3178 – 3185, 2012. 3