人体姿态检测综述

许楚萍

2017 年 5 月 29 日

1 人体模型

人体模型的构建和描述是人体姿态检测流程中一个重要的模块。当前主流的人体模型主要有运动模型（Kinematic Model）、平面模型（Planar Model）和体积模型（Volumetric Model）[61]。

1.1 运动模型

运动模型利用关节点的位置和身体部位的方向来定义人体姿态。运动模型又可以分为预定义模型（predefined model）和图形结构模型（graph structure）。其中，PSM（Pictorial Structures Model）[62,63]是一种常用的图形结构模型，也是当前人体姿态检测领域里相当重要的里程碑。Pictorial Structures[64]早由 Fischler 和 Elschlager 于 1973 年提出，这种方法将人体姿态检测问题转化为能量方程的最小化问题。后来 Felzenszwalb [63]人在此模型的基础上定义了人体姿态。Felzenszwalb 的方法虽然只能应用于前景背景分离的灰度图像上，但仍为后续研究成果的改进和实现提供了诸多帮助。

树结构模型（tree-structured model）是 PSM 的一个特例，无论在 2D 还是 3D 上，都适用于人体姿态检测[65,66,67,68,69,70,71,72,73]。然而运动树结构模型在表示身体部位之间的依赖关系时有一定的局限性。此外，当遮挡存在的情况下，身体部位完全无法检测[74]。针对运动树结构模型的局限性，后续许多研究提出在树结构模型上的改进方法。Wang 等人[75]通过添加不同形状的树结构模型来解决缺少模型描述符的问题。Sapp 等人通过[76]添加模型的状态来增强模型的表达能力。Wang 等人[75]和 Komodakis 等人[77]使用多个树模型而不是单个树模型进行人类姿态估计。

1.2 平面模型

平面模型不仅能够捕获身体部位之间的连接关系，还能够学习人体部位的形状和外观。ASM（Active Shape Models）就是一种平面模型，能够用于表示整个人体结构，并用主成分分析（PCA）从平均形状中获得轮廓变形的统计[78,79]。纸板模型（Cardboard Model）是另一种平面模型，它由前景物体颜色信息和人体部位矩形构成。纸板模型通常包含一个躯干和八个半截肢体，前景信息由颜色直方图表示，每个身体部位由该部位的平均 RGB 来表示。例如 Hao 等人[80]的方法就是采用纸板模型来实现人体姿态检测。

1.3 体积模型

体积模型能够真实地表示 3D 体形和姿态。几何形状和网格是体积模型中常用的两种方式。当使用几何形状作为模型组件时，人体部分用圆柱体、圆锥体和其他近似形状来表示肢

体，例如 Hedvig 等人[81]的方法就是用圆柱体的复合体建模人体姿态。 这种模型能够准确地获得人体四肢的真实形状，并且容许人体结构或服饰的变化，因而比图形基于结构的方法更具有准确性。当使用网格来表示人体姿态时，网格通常被分割成几个身体部位。SCAPE（Shape Completion and Animation of People）是一种常用的 3D 网格模型[82,83]。

1.4 姿势先验

人体姿势受到许多因素的约束，包括运动学、关节操作限制以及特定活动下的运动模式等[84]。通过从数据中学习姿势约束并创建姿势先验，可以为人体姿态检测提供有效的信息。一些研究将这些学习到的知识作为模型的先验知识并用于约束人体关节点[85]。

2 分类算法

通常情况下，人体姿态的检测的分类算法可划分为两类：产生式分类算法和判别式分类算法。产生式分类算法把人体姿态检测模拟为几何映射问题，判别式分类算法则把人体姿态检测视为特定图像处理问题。另一种划分方法根据起始点为高级抽象语义或低级像素，把人体姿态的检测方法划分为自顶向下或自底向上。

2.1 产生式分类和判别式分类

2.1.1 产生式分类算法

产生式分类算法的基本思想是首先建立类别样本的概率密度模型，再利用模型进行推理预测，这种方法要求具有大量的已知样本。在人体姿态检测问题中，产生式分类算法[86]从人体姿势初始化模型开始，将姿势映射到图像平面，再用图像上的信息验证模型的准确性。

2.1.2 判别式分类算法

判别式模型与生成式模型的工作原理恰好相反，它的的基本思想是通过大量的样本学习获得模型分布。在人体姿态检测问题中，判别式分类算法从图像信息开始，利用训练数据为图像信息与人体姿态之间的关系建模。这种方法相比于产生式分类算法相对简单、计算效率高，一旦建模完成，对输入图像的姿态检测效率远远高于产生式分类算法。大部分的人体姿态检测方法采用的是判别式的分类算法。判别式分类算法又可以根据其训练集的来源分为基于学习的分类算法和基于特定样本的分类算法。

基于学习的分类算法不对训练样本进行特定筛选和分类，现已存在大量这一类方法相关的文献。SVM（Support Vector Machines）是当前被最广泛使用的分类器。Remi 等人[87]用 SVM 或相关性向量机（RVM）学习获得的专用检测器替代简单的部位检测器。Ryuzo 等人[88]为了区分姿态簇，使用具有姿势相关特征选择的内核支持向量机，通过最小化半径/边界约束来估计 RBF 内核的尺度参数，从而实现了内核支持向量机的特征选择。Huazhong 等人[89]使用的判别式词袋方法实现单目图像的 3D 人体姿态检测。近年来，随着深度学习的快

速发展及其在数据挖掘和表达能力上的优势，人体姿态检测领域引入了深度学习方法用于学习图像中的关键信息，其中 CNN（Convolutional Neural Networks）被广泛应用于人体姿态检测中。Tompson 等人[90]提出了一种涉及深卷积网络和马尔可夫随机场（MRF）模型的组合架构。Gkioxari 等人[91]用由 CNN 特征表示的区域来训练出具有损失函数的 R-CNN 检测器。Joao 等人[92]采用迭代误差反馈，通过反馈误差预测来改变初始解。

基于特定样本的分类算法中，对于输入的图像，我们用特定姿态的离散图像集合及其对应的特征表示来训练模型并实现人体姿态检测[93]。随机树[94]和随机森林[95,96]、霍夫森林[97]都是可以用于快速处理这种类型问题强大分类技术[98]。此外，稀疏表示（SR）也常用于提取最重要的训练样本，后来所有的估计也都是基于这些样本进行的[99,100,101]。

2.1.3 融合产生式与判别式分类算法

产生式分类将人体模型投射到 2D 图像空间中，并测量它们之间的距离[83]，而判别式分类检测人体的各部位来重建人体姿势。由于产生式分类算法效率低下，而判别式算法对于不存在于训练集中的姿势识别率低[79]，许多研究结合了这两种方法的优势，提出产生式分类与判别式分类相结合的分类算法。这些方法通常用判别式方法[102]的检测结果来初始化姿势，并用产生式方法来优化局部区域内的人体姿态[103,104,105]。

2.2 自顶向下和自底向上

根据人体姿态检测算法的工作方式是开始于高级抽象语义或低级像素，我们又可以把人体姿态检测算法分为自顶向下和自底向上的方法。在计算机视觉的语义层次结构中，图像是最低级别，而人体姿态的配置以及人体姿态所属的行为类型是较高级别。

2.2.1 自底向上

自底向上的方法从图像中采集并描述出具有代表性的特征，将这些特征直接用语检测人体姿态或者用于定位身体部位再组合成人体姿态。根据单位面积的大小，这一类方法包括了基于像素超像素的方法和基于人体部位的方法。

(1)基于像素或超像素的方法

图像的像素或超像素可以用于分割图像并集成到姿态检测中[106,107]基于像素的方法也可以与其他方法相结合。例如，Hernández Vela 等人[108]利用能量最小化框架，通过图切割优化来扩展每像素的分类方法。此外，还可以利用分割的结果来增强像素级的估计。Marcin 等人[107]提出了一种方法，通过在检测到的区域上初始化的“GrabCut”来进一步缩小搜索空间，直到缩小到身体部位的搜索空间[109,110]。基于部件和基于像素的方法也可以在单个优化框架中组合[111]。超像素也可用于限制人体模型中的关节位置[112]。在基于超像素的方法中，我们还可以优化身体部位匹配和前景估计。例如，利用分支绑定算法[113,114,115]来实现这个过程。

(2)基于部件的方法

基于部件的方法通过学习人体部位外观和位置模型来解决人体姿态检测问题。在基于部分的方法中，首先从图像中检测身体部位候选位置，然后逐个检测身体部位使其符合人体外观的正常表现[116]。Yang 等人[72]扩展了可变形部件模型（DPM）[73,117]，提出了用于铰接式 2D 人体姿态检测的 FMP（Flexible mixture of parts ）模型 。 这种方法后来利用组合和/或图形语法模型被进一步改进[118]。

2.2.2 自顶向下

我们这里的自顶向下指的是从高级抽象语义到低级像素的问题解决过程[119]，其中高级语义用于指导低级语义。根据这个概念，自顶向下的方法更多是结合自底向上的方法来使用，而且更高级的语义通常能在效果上提供很大帮助。

2.2.3 融合自顶向下与自底向上

自顶向下与自底向上的方法结合起来比产生式方法与判别式方法结合起来更加灵活。通过将图形运动学模型与检测方法相结合，可以同时获得检测结果和 3D 姿态[116,120,121]。此外，通过图形切割可以找到全局最小能量[122]。 Bray 等人[120]使用图形切割来优化姿态参数以实体人体的集成分割和 3D 姿态估计。

3 特征

从一幅图像中提取并描述关键点是图像处理领域里关键问题，我们需要在图像中提取各种各样的特征，这些特征会直接影响后面所有工作的效果。在人体姿态检测中，首要工作就是提取特征，因为特征包含了图像大部分的信息，后面的工作都是基于此，图像方面的特征主要分成四种：底层特征，中层特征，高层特征和运动特征，下面我们就分别介绍这四种特征。

3.1 底层特征

底层特征大致可以分为三类：剪影[1,2,3,4]，轮廓[5,6]和边缘[7,8]。剪影主要是提取物体的边界，其对纹理和光照具有很好地不变性[2,9,10,11,12]，在以往的工作中，往往是通过将图片进行背景剪除从而得到物体的剪影。轮廓则主要是用于提取身体部位的边界[6],与剪影不同的是，提取轮廓之前需要先对图像进行预处理，比如图形增强，滤波去噪和灰度化等等，经过处理的图像能够突显我们所需要的信息，从而改善了轮廓的提取效果。边缘则是通过卷积计算来突显图像中变化明显的线，我们也可以理解为图像的梯度信息，它可以用滤波器对图片进行扫描卷积得出，它对图片的旋转，光照等因素具有鲁棒性，除此以外，底层特征还有颜色特征[13]和纹理特征[14]。

3.2 中层特征

中层特征可分为局部特征，全局特征，多级分层编码和自提取特征四种，下面分别介绍

这几种特征。

3.2.1 局部特征

上面的底层特征中，被提取的剪影被编码成傅里叶描述符[15]，形状上下文[16]，几何特征[17]和泊松特征[18]等等。其中最常用的的形状上下文描述符，它得到了当前的点跟其他点的相对分布的描述，并利用对数极坐标来计算直方图，通常会把空间分成多个相同的角度和方向，此举也是为了方便计算直方图。每一个点经过计算都会成为直方图的其中一部分，最后将所有部分累积起来得到的就是整副图像的直方图。形状上下文将分布的点转化成了多维的描述符，这对于图像分割错误具有鲁棒性。

基于边缘或者梯度的底层特征则被编码成了梯度直方图（HOG）[19,20,21]，HOG 通常是表示身体的不同部位，还有相对边缘分布[22],尺度不变特征变换（SIFT）[23,24]和类 SIFT 特征[25,26]，由于可以适应不同大小的图片，在卷积神经网络出现之前，是最受欢迎的一种特征，除此以外还有边界特征[27]和形变特征[28]，边界特征和形变特征是梯度与边缘的组合。

3.2.2 全局特征

除了上述的局部特征以外，还存在全局特征，这些特征就是以整张图为单位来提取信息，比如物体前景图和密集特征网络，密集网格特征有 HOG 特征[19]的网络和 SIFT 特征的网络[23,29]，实验证明，特征网络往往能取得比原特征更好的效果。

3.2.3 多级分层编码

多级分层编码，包括多级模型和 HMAX[30]，还有超特征[31]和特征金字塔[32]，词汇树以及多层空间块[33]都在保持空间几何不变方面具有鲁棒性，其他特征如本地路径[34]，预测流水线[35]以及极端人体曲线[36]等也是常用的特征。

3.2.4 自提取特征

自提取特征的一个典型是卷积神经网络（CNN）目前在计算机视觉，机器学习和人工智能领域非常流行，CNN 是神经网络(NN)的延伸，其基本结构包括两层，其一为特征提取层，每个神经元的输入与前一层的局部接受域相连，并提取该局部的特征。一旦该局部特征被提取后，它与其它特征间的位置关系也随之确定下来；其二是特征映射层，网络的每个计算层由多个特征映射组成，每个特征映射是一个平面，平面上所有神经元的权值相等。特征映射结构采用影响函数核小的 sigmoid 函数作为卷积网络的激活函数，使得特征映射具有位移不变性。此外，由于一个映射面上的神经元共享权值，因而减少了网络自由参数的个数。卷积神经网络中的每一个卷积层都紧跟着一个用来求局部平均与二次提取的计算层，这种特有的两次特征提取结构减小了特征分辨率。

CNN 主要用来识别位移、缩放及其他形式扭曲不变性的二维图形。由于 CNN 的特征检

测层通过训练数据进行学习，所以在使用 CNN 时，避免了显示的特征抽取，而隐式地从训练数据中进行学习；再者由于同一特征映射面上的神经元权值相同，所以网络可以并行学习，这也是卷积网络相对于神经元彼此相连网络的一大优势。卷积神经网络以其局部权值共享的特殊结构在语音识别和图像处理方面有着独特的优越性，其布局更接近于实际的生物神经网络，权值共享降低了网络的复杂性，特别是多维输入向量的图像可以直接输入网络这一特点避免了特征提取和分类过程中数据重建的复杂度，目前许多工作都是使用卷积神经网络[37,38,39]。

3.3 高层特征

几个描述符具有高级特征，例如身体部位块，几何描述符或上下文特征。身体部位块可以描述身体部位的任何一个方向和位置，与身体部位相比，身体部位块是一个更容易理解的描述符，它被限制在肢体与肢体之间，关节与关节之间或者是关节附近。

几何描述符则是描述了单个部位之间的语义关系[40,41,42]，通常是将两种特征组合在一起，比如组合身体部位的方向和位置。高级特征编码组合单位之间的语义共生。与预期模式中的空间或时间编码的中级特征相比，高级别将开发训练数据的相关性，并使数据自身发挥作用。

一些上下文特征有哈尔特点 LBP 纹理特征[43]，这是基于树结构所提出的一种新的特征，HOG 和颜色融合特征[44]，随机森林深度特征[45]和二进制相位滤波器特征[46]，这个特征可以用在深度图像。除此以外还有 3D 云特征和[47]和刚性身体部件特征[48]。

3.4 运动特征

如上所述，单眼图像的估计姿势可以用作智能监控系统中姿态跟踪的初始化。其时间和空间一致性非常有用，例如，它可以在一个帧中发现错误的姿态并矫正。

密集光流[49]，鲁棒光流[50]，边缘能量和运动边界及其组合[51]等运动特征通过时间对应增强了估计性能。光流[52]是由观察者和场景之间的相对运动引起的物体，表面和边缘运动的图案。光流中的梯度与运动有关，可用于跟踪姿势[53,54]。特征还表达了局部运动相似性，例如运动框架[55,56]和基于图像差异的运动和外观斑块[57]。由于单一功能对背景变化不敏感，从而导致模糊。因此可以组合特征来提高姿态估计的性能[58,59]。此外，通过组合具有不同特征的多个图像提示，如边缘提示，脊线提示和运动提示，可以更精确地估计单目图像中的人体姿态[60]。

参考文献

[1]Inferring Body Pose without Tracking Body Parts. Available online: http://ieeexplore.ieee.org/document/854946/ (accessed on 10 November 2016)

[2]Agarwal, A.; Triggs, B. Recovering 3D Human Pose from Monocular Images. *IEEE Trans. Pattern Anal. Mach. Intell.* **2006**,*28*, 44–58.

[3]Gavrila, D.M. A Bayesian, Exemplar-Based Approach to Hierarchical Shape Matching. *IEEE Trans. Pattern Anal. Mach. Intell.* **2007**, *29*, 1408–1421.

[4]Elgammal, A.; Lee, C.S. Inferring 3D Body Pose from Silhouettes Using Activity Manifold Learning. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Washington, DC, USA, 27 June–2 July 2004; pp. 681–688

[5]Viola, P.; Jones, M.J.; Snow, D. Detecting Pedestrians Using Patterns of Motion and Appearance. *Int. J. Comput. Vis.* **2005**,*63*, 153–161.

[6]Sapp, B.; Toshev, A.; Taskar, B. Cascaded Models for Articulated Pose Estimation. In Proceedings of the European Conference on Computer Vision, Crete, Greece, 5–11 September 2010; pp. 406–420.

[7]Dimitrijevic, M.; Lepetit, V.; Fua, P. Human Body Pose Detection Using Bayesian Spatio-Temporal Templates. *Comput. Vis. Image Underst.* **2006**, *104*, 127–139.

[8]Weinrich, C.; Volkhardt, M.; Gross, H.M. Appearance-based 3D Upper-Body Pose Estimation and Person Re-Identification on Mobile Robots. In Proceedings of the IEEE International Conference on Systems, Man, and Cybernetics, Manchester, UK, 13–16 October 2013; pp. 4384–4390.

[9]Sminchisescu, C.; Kanaujia, A.; Li, Z.; Metaxas, D. Discriminative Density Propagation for 3D Human Motion Estimation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, San Diego, CA, USA, 20–26 June 2005; pp. 390–397.

[10]Grauman, K.; Shakhnarovich, G.; Darrell, T. Inferring 3D Structure with a Statistical Image-Based Shape Model. In Proceedings of the IEEE International Conference on Computer Vision, Madison, WI, USA, 16–22 June 2003; pp. 641–647.

[11]Kehl, R.; Bray, M.; Van Gool, L. Full Body Tracking from Multiple Views Using Stochastic Sampling. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, San Diego, CA, USA, 20–26 June 2005; pp. 129–136.

[12]Sminchisescu, C.; Telea, A. Human Pose Estimation from Silhouettes-A Consistent Approach Using Distance Level Sets. In Proceedings of the International Conference in Central Europe

on Computer Graphics, Visualization and Computer Vision, Bory, Czech Republic, 4–8 February 2002.

[13]Wren, C.R.; Azarbayejani, A.; Darrell, T.; Pentland, A.P. Pfinder: Real-Time Tracking of the Human Body. *IEEE Trans. Pattern Anal. Mach. Intell.* **1997**, *19*, 780–785.

[14]Felzenszwalb, P.F.; Girshick, R.B.; McAllester, D.; Ramanan, D. Object Detection with Discriminatively Trained Part-Based Models. *IEEE Trans. Pattern Anal. Mach. Intell.* **2010**, *32*, 1627–1645.

[15]Zahn, C.T.; Roskies, R.Z. Fourier Descriptors for Plane Closed Curves. *IEEE Trans. Comput.* **1972**, *100*, 269–281.

[16]Mori, G.; Belongie, S.; Malik, J. Efficient Shape Matching Using Shape Contexts. *IEEE Trans. Pattern Anal. Mach. Intell.***2005**, *27*, 1832–1837.

[17]Arkin, E.M.; Chew, L.P.; Huttenlocher, D.P.; Kedem, K.; Mitchell, J.S. An Efficiently Computable Metric for Comparing Polygonal Shapes. *IEEE Trans. Pattern Anal. Mach. Intell.* **1991**, *13*, 209–216.

[18]Gorelick, L.; Galun, M.; Sharon, E.; Basri, R.; Brandt, A. Shape Representation and Classification Using the Poisson Equation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2006**, *28*, 1991–2005.

[19]Dalal, N.; Triggs, B. Histograms of Oriented Gradients for Human Detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, San Diego, CA, USA, 20–26 June 2005; pp. 886–893.

[20]Zhu, Q.; Yeh, M.C.; Cheng, K.T.; Avidan, S. Fast Human Detection Using a Cascade of Histograms of Oriented Gradients. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, New York, NY, USA, 17–22 June 2006; pp. 1491–1498.

[21]Shakhnarovich, G.; Viola, P.; Darrell, T. Fast Pose Estimation with Parameter-Sensitive Hashing. In Proceedings of the IEEE International Conference on Computer Vision, Madison, WI, USA, 16–22 June 2003; pp. 750–757.

[22]Nayak, S.; Sarkar, S.; Loeding, B. Distribution-Based Dimensionality Reduction Applied to Articulated Motion Recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* **2009**, *31*, 795–810.

[23]Lowe, D.G. Object Recognition from Local Scale-Invariant Features. In Proceedings of the IEEE International Conference on Computer Vision, Corfu, Greece, 20–25 September 1999; pp. 1150–1157.

[24]Lowe, D.G. Distinctive Image Features from Scale-Invariant Keypoints. *Int. J. Comput. Vis.* **2004**, *60*, 91–110.

[25]Agarwal, A.; Triggs, B. A Local Basis Representation for Estimating Human Pose from Cluttered Images. In Proceedings of the Asian Conference on Computer Vision, Hyderabad, India, 13–16 January 2006; pp. 50–59.

[26]Scovanner, P.; Ali, S.; Shah, M. A 3-Dimensional Sift Descriptor and Its Application to Action Recognition. In Proceedings of the International Conference on Multimedia, Augsburg, Bavaria, Germany, 23–28 September 2007; pp. 357–360.

[27]Wu, Y.; Lin, J.; Huang, T.S. Analyzing and Capturing Articulated Hand Motion in Image Sequences. *IEEE Trans. Pattern Anal. Mach. Intell.* **2005**, *27*, 1910–1922.

[28]Sabzmeydani, P.; Mori, G. Detecting Pedestrians by Learning Shapelet Features. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Minneapolis, MN, USA, 18–23 June 2007; pp. 1–8.

[29]Ionescu, C.; Li, F.; Sminchisescu, C. Latent Structured Models for Human Pose Estimation. In Proceedings of the IEEE International Conference on Computer Vision, Barcelona, Spain, 6–13 November 2011; pp. 2220–2227.

[30]Serre, T.; Wolf, L.; Poggio, T. Object Recognition with Features Inspired by Visual Cortex. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, San Diego, CA, USA, 20–26 June 2005; pp. 994–1000.

[31]Agarwal, A.; Triggs, B. Hyperfeatures–Multilevel Local Coding for Visual Recognition. In Proceedings of the European Conference on Computer Vision, Graz, Austria, 7–13 May 2006; pp. 30–43.

[32]Lazebnik, S.; Schmid, C.; Ponce, J. Beyond Bags of Features: Spatial Pyramid Matching for Recognizing Natural Scene Categories. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, New York, NY, USA, 17–22 June 2006; pp. 2169–2178.

[33]Kanaujia, A.; Sminchisescu, C.; Metaxas, D. Semi-Supervised Hierarchical Models for 3D Human Pose Reconstruction. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Minneapolis, MN, USA, 18–23 June 2007; pp. 1–8.

[34]Hara, K.; Kurokawa, T. Human Pose Estimation Using Patch-Based Candidate Generation and Model-Based Verification. In Proceedings of the IEEE International Conference on Automatic Face and Gesture Recognition, Santa Barbara, CA, USA, 21–25 March 2011; pp. 687–693.

[35]Lallemand, J.; Szczot, M.; Ilic, S. Human Pose Estimation in Stereo Images. In Proceedings of the International Conference on Articulated Motion and Deformable Objects, Palma de Mallorca, Spain, 16–18 July 2014; pp. 10–19.

[36]Slama, R.; Wannous, H.; Daoudi, M. Extremal Human Curves: A New Human Body Shape and Pose Descriptor. In Proceedings of the IEEE International Conference and Workshops on Automatic Face and Gesture Recognition, Shanghai, China, 22–26 April, 2013; pp. 1–6.

[37]Li, S.; Chan, A.B. 3D Human Pose Estimation from Monocular Images with Deep Cnvolutional Neural Network. In Proceedings of the Asian Conference on Computer Vision, Singapore, 1–5 November 2014; pp. 332–347.

[38]Li, S.; Liu, Z.Q.; Chan, A. Heterogeneous Multi-Task Learning for Human Pose Estimation with Deep Convolutional Neural Network. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, Columbus, OH, USA, 23–28 June 2014; pp. 482–489.

[39]Pfister, T.; Charles, J.; Zisserman, A. Flowing Convnets for Human Pose Estimation in Videos. In Proceedings of the IEEE International Conference on Computer Vision, Santiago, Chile, 13–16 December 2015; pp. 1913–1921.

[40]Roberts, T.J.; McKenna, S.J.; Ricketts, I.W. Human Pose Estimation Using Learnt Probabilistic Region Similarities and Partial Configurations. In Proceedings of the European Conference on Computer Vision, Prague, Czech Republic, 11–14 May 2004; pp. 291–303.

[41]Bourdev, L.; Maji, S.; Brox, T.; Malik, J. Detecting People Using Mutually Consistent Poselet Activations. In Proceedings of the European Conference on Computer Vision, Crete, Greece, 5–11 September 2010; pp. 168–181.

[42]Gkioxari, G.; Hariharan, B.; Girshick, R.; Malik, J. Using k-Poselets for Detecting People and Localizing Their Keypoints. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 23–28 June 2014; pp. 3582–3589.

[43]Yuan Zihua; Li Feng; Zhou Shuren.human pose estimation Based on Haar Characteristics LBP Texture Feature.Computer Engineering.15 April 2015;pp.199-204.

[44]Han Guijin; Zhu Hong;Human Pose Estimation Based on Fusion of HOG and Color Feature.Pattern Recognition and Artificial Intelligence.2014.pp:769-777.

[45]Zhu Jueyu; Cao Yawei; Zhou Shuren; Li Feng.Human pose estimation based on random forest depth feature selection.Computer Engineering and Application.2017,pp:172-176.

[46]Cao Yawei; Zhou Shuren.Research of human pose estimation based on BPOF feature and depth image;Computer Engineering and Applications,2016,pp:168-172.

[47] Chan, Kai-Chi; Koh, Cheng-Kok; Lee, C. S. George.A 3D-Point-Cloud Feature for Human-Pose Estimation.In Proceedings of theIEEE International Conference on Robotics and Automation (ICRA) , Karlsruhe, GERMANY ,06-10 MAY , 2013.pp:1623-1628.

[48]Jalal, A.; Yeonho Kim; Daijin Kim.Ridge body parts features for human pose estimation and recognition from RGB-D video data.2014 5th International Conference on Computing, Communication and Networking Technologies (ICCCNT) ,Hefei, China ,11-13 July 2014;pp:6

[49]Zuffi, S.; Freifeld, O.; Black, M.J. From Pictorial Structures to Deformable Structures. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Providence, RI, USA, 16–21 June 2012; pp. 3546–3553.

[50] Lu, Y.; Jiang, H. Human Movement Summarization and Depiction from Videos. In Proceedings of the IEEE International Conference on Multimedia and Expo, San Jose, CA, USA, 15–19 July 2013; pp. 1–6.

[51] Sminchisescu, C.; Triggs, B. Covariance Scaled Sampling for Monocular 3D Body Tracking. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Kauai, HI, USA, 8–14 December 2001; pp. I-447–I-454.

[52]Dalal, N.; Triggs, B.; Schmid, C. Human Detection Using Oriented Histograms of Flow and Appearance. In Proceedings of the European Conference on Computer Vision, Graz, Austria, 7–13 May 2006; pp. 428–441.

[53]Bregler, C.; Malik, J. Tracking People with Twists and Exponential Maps. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Santa Barbara, CA, USA, 23–25 June 1998; pp. 8–15.

[54]Scene Constraints-aided Tracking of Human Body. Available online:http://ieeexplore.ieee.org/document/ 855813/ (accessed on 23 November 2016).

[55]Wang, J.; Yang, J.; Yu, K.; Lv, F.; Huang, T.; Gong, Y. Locality-Constrained Linear Coding for Image Classification. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, San Francisco, CA, USA, 13–18 June 2010; pp. 3360–3367.

[56] Sun, L.; Song, M.; Tao, D.; Bu, J.; Chen, C. Motionlet LLC Coding for Discriminative Human Pose Estimation. Multimed. Tools Appl. 2014, 73, 327–344.

[57]Yang, M.H.; Bissacco, A. Fast Human Pose Estimation Using Appearance and Motion via Multi-Dimensional Boosting Regression. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, San Francisco, CA, USA, 13–18 June 2010; pp. 1–8.

[58]Sedai, S.; Bennamoun, M.; Huynh, D.Q.; Crawley, P. Localized Fusion of Shape and Appearance Features for 3D Human Pose Estimation. In Proceedings of the British Machine Vision Conference, Aberystwyth, UK, 31 August–3 September 2010; pp. 1–10.

[59] Sedai, S.; Bennamoun, M.; Huynh, D.Q. Discriminative Fusion of Shape and Appearance Features for Human Pose Estimation. Pattern Recognit. 2013, 46, 3223–3237.

[60] Sidenbladh, H.; Black, M.J. Learning Image Statistics for Bayesian Tracking. In Proceedings of the IEEE International Conference on Computer Vision, Vancouver, BC, Canada, 7–14 July 2001; pp. 709–716.

[61]Gong, W. et al. Human pose estimation from monocular images: A comprehensive survey. *Sensors (Switzerland)* **16,** 1–39 (2016).

[62]Felzenszwalb, P. F., Huttenlocher, D. P. & Pedro F. Felzenszwalb. Efficient maching of pictorial structures. *Cvpr* 66–75 (2000).

[63]Felzenszwalb, P. F. & Huttenlocher, D. Pictoral structures for object recognition. *Ijcv* **61,** 55–79 (2005).

[63]Fischler, M. A. & Elschlager, R. A. The Representation and Matching of Pictorial Structures Representation. *IEEE Trans. Comput.* **C-22,** 67–92 (1973).

[65]Andriluka, M., Roth, S. & Schiele, B. Pictorial structures revisited: People detection and articulated pose estimation. *2009 IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. Work. CVPR Work. 2009* 1014–1021 (2009). doi:10.1109/CVPRW.2009.5206754

[66]Felzenszwalb, P., McAllester, D. & Ramanan, D. A Discriminatively Trained, Multiscaled, Deformable Part Model. *Cvpr* 1–8 (2008).doi:10.1109/CVPR.2008.4587597

[67]Andriluka, M., Roth, S. & Schiele, B. Monocular 3D Pose Estimation and Tracking by Detection. *2010 Ieee Conf. Comput. Vis. Pattern Recognit.* 623–630 (2010). doi:10.1109/CVPR.2010.5540156

[68]Yang, Y. & Ramanan, D. Articulated pose estimation with flexible mixtures-of-parts. *Comput. Vis. Pattern ...* (2011).

[69]Sapp, B., Jordan, C. & Taskar, B. Adaptive pose priors for pictorial structures. *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.* 422–429 (2010). doi:10.1109/CVPR.2010.5540182

[70]Chen, X. & Yuille, A. Parsing occluded people by flexible compositions. *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.* **07–12–June,** 3945–3954 (2015).

[71]Andriluka, M., Roth, S. & Schiele, B. Monocular 3D pose estimation and tracking by detection. *2010 IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.* 623–630. doi:10.1109/CVPR.2010.5540156

[72]Yang, Y. & Ramanan, D. Articulated Human Detection with Flexible Mixtures-of-Parts. *IEEE Trans. Pattern Anal. Mach. Intell.* 1–15 (2012). doi:4E81F44D-B885-4A86-B3BB-ACD954C51B4D

[73]Felzenszwalb, P. F., Girshick, R. B., Mcallester, D. & Ramanan, D. Object Detection with Discriminatively Trained Part Based Models. *IEEE Trans. Pattern Anal. Mach. Intell.* **32,** 1–20 (2009).

[74]Kanade, T., Ramakrishna, V., Sheikh, Y., Kanade, T. & Sheikh, Y. Tracking Human Pose by Tracking Symmetric Parts. *Comput. Vis. Pattern Recognit. (CVPR), 2013 IEEE Conf.* 3728–3735 (2013). doi:10.1109/CVPR.2013.478

[75]Wang, Y. & Mori, G. Multiple Tree Models for Occlusion and Spatial Constraints in Human Pose Estimation. *Constraints* 710–724 (2008).

[76]Sapp, B. & Taskar, B. MODEC: Multimodal decomposable models for human pose estimation. *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.* 3674–3681 (2013). doi:10.1109/CVPR.2013.471

[77]Komodakis, N., Paragios, N. & Tziritas, G. MRF energy minimization and beyond via dual decomposition. *IEEE Trans. Pattern Anal. Mach. Intell.* **33,** 531–552 (2011).

[78]Freifeld, O., Weiss, A., Zuffi, S. & Black, M. J. Contour people: A parameterized model of 2D articulated human shape. in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition* 639–646 (2010). doi:10.1109/CVPR.2010.5540154

[79]Cootes, T. F., Taylor, C. J., Cooper, D. H. & Graham, J. Active Shape Models - Their Training and Application. *Comput. Vis. Image Underst.* **61,** 38–59 (1995).

[80]Jiang, H. Finding human poses in videos using concurrent matching and segmentation. in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* **6492 LNCS,** 228–243 (2011).

[81]Sidenbladh, H., De La Torre, F. & Black, M. J. A framework for modeling the appearance of 3D articulated figures. in *Proceedings - 4th IEEE International Conference on Automatic Face and Gesture Recognition, FG 2000* 368–375 (2000). doi:10.1109/AFGR.2000.840661

[82]Anguelov, D. *et al.* SCAPE: Shape Completion and Animation of People. in *SIGGRAPH '05 ACM SIGGRAPH 2005 Papers* 408–416 (2005). doi:10.1007/978-3-642-37484-5_12

[83]Ge, S. & Fan, G. Non-rigid articulated point set registration for human pose estimation. in *Proceedings - 2015 IEEE Winter Conference on Applications of Computer Vision, WACV 2015* 94–101 (2015). doi:10.1109/WACV.2015.20

[84]Rius, I., Gonzàlez, J., Varona, J. & Xavier Roca, F. Action-specific motion prior for efficient Bayesian 3D human body tracking. *Pattern Recognit.* **42,** 2907–2921 (2009).

[85]Urtasun, R., Fleet, D. J., Hertzmann, A. & Fua, P. Priors for people tracking from small training sets. in *Proceedings of the IEEE International Conference on Computer Vision* **I,** 403–410 (2005).

[86]Lepetit, V. & Fua, P. Monocular Model-Based 3D Tracking of Rigid Objects: A Survey. *Found. Trends® Comput. Graph. Vis.* **1,** 1–89 (2005).

[87]Ronfard, R., Schmid, C. & Triggs, B. Learning to parse pictures of people. *Proc. Eur. Conf. Comput. Vis.* **2353,** 700–714 (2002).

[88]Okada, R. & Soatto, S. Relevant feature selection for human pose estimation and localization in cluttered images. in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* **5303 LNCS,** 434–445 (2008).

[89]Ning, H. N. H., Xu, W. X. W., Gong, Y. G. Y. & Huang, T. Discriminative learning of visual words for 3D human pose estimation. *2008 IEEE Conf. Comput. Vis. Pattern Recognit.* 1–8 (2008). doi:10.1109/CVPR.2008.4587534

[90]Tompson, J., Jain, A., LeCun, Y. & Bregler, C. Joint Training of a Convolutional Network and a Graphical Model for Human Pose Estimation. 1–9 (2014).

[91]Gkioxari, G., Hariharan, B., Girshick, R. & Malik, J. R-CNNs for Pose Estimation and Action Detection. 1–8 (2014).

[92]Carreira, J., Agrawal, P., Fragkiadaki, K. & Malik, J. Human pose estimation with iterative error feedback. in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition* **2016–Decem,** 4733–4742 (2016).

[93]Flitti, F., Bennamoun, M., Huynh, D. Q. & Owens, R. A. Probabilistic human pose recovery from 2D images. in *Proceedings - International Conference on Image Processing, ICIP* 1517–1520 (2010). doi:10.1109/ICIP.2010.5652502

[94]Amit, Y. & Geman, D. Shape Quantization and Recognition with Randomized Trees. *Neural Comput.* **9,** 1545–1588 (1997).

[95]Chang, J. Y. & Nam, S. W. Fast random-Forest-Based human pose estimation using a multi-Scale and cascade approach. *ETRI J.* **35,** 949–959 (2013).

[96]Breiman, L. Random forests. *Mach. Learn.* **45,** 5–32 (2001).

[97]Girshick, R., Shotton, J., Kohli, P., Criminisi, A. & Fitzgibbon, A. Efficient regression of general-activity human poses from depth images. in *Proceedings of the IEEE International Conference on Computer Vision* 415–422 (2011). doi:10.1109/ICCV.2011.6126270

[98]Lepetit, V. & Fua, P. Keypoint recognition using randomized trees. *IEEE Trans. Pattern Anal. Mach. Intell.* **28,** 1465–1479 (2006).

[99]Huang, J. Bin & Yang, M. H. Fast sparse representation with prototypes. in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition* 3618–3625 (2010). doi:10.1109/CVPR.2010.5539919

[100]Wright, J., Yang, A. Y., Ganesh, A., Sastry, S. S. & Ma, Y. Robust face recognition via sparse representation. *IEEE Trans. Pattern Anal. Mach. Intell.* **31,** 210–227 (2009).

[101]Chen, C., Yang, Y., Nie, F. & Odobez, J. M. 3D human pose recovery from image by efficient visual feature selection. *Comput. Vis. Image Underst.* **115,** 290–299 (2011).

[102]Kanaujia, A. Coupling Top-down and Bottom-up Methods for 3D Human Pose and Shape Estimation from Monocular Image Sequences. 13 (2014).

[103]Agarwal, a. & Triggs, B. Monocular Human Motion Capture with a Mixture of Regressors. *2005 IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. - Work.* **3,** 72–72 (2005).

[104]Sigal, L., Balan, A. O. & Black, M. J. Combined discriminative and generative articulated pose and non-rigid shape estimation. in *NIPS* **20,** 1–8 (2007).

[105]Orrite-Uru??ueta, C., Del Rinc??n, J. M., Herrero-Jaraba, J. E. & Rogez, G. 2D silhouette and 3D skeletal models for human detection and tracking. in *Proceedings - International Conference on Pattern Recognition* **4,** 244–247 (2004).

[106]Lu, H., Shao, X. & Xiao, Y. Pose estimation with segmentation consistency. *IEEE Trans. Image Process.* **22,** 4040–4048 (2013).

[107]Eichner, M. & Ferrari, V. Better appearance models for pictorial structures. in *Procedings of the British Machine Vision Conference 2009* 3.1-3.11 (2009). doi:10.5244/C.23.3

[108]Hernandez-Vela, A. *et al.* Graph cuts optimization for multi-limb human segmentation in depth maps. in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition* 726–732 (2012). doi:10.1109/CVPR.2012.6247742

[109]Ferrari, V., Mar??n-Jim??nez, M. & Zisserman, A. 2D human pose estimation in TV shows. in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* **5604 LNCS,** 128–147 (2009).

[110]Ferrari, V., Marin-Jimenez, M. & Zisserman, A. Pose search: Retrieving people using their pose. in *2009 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, CVPR Workshops 2009* 1–8 (2009). doi:10.1109/CVPRW.2009.5206495

[111]Ladicky, L., Torr, P. H. S. & Zisserman, A. Human Pose Estimation Using a Joint Pixel-wise and Part-wise Formulation. *Comput. Vis. Pattern Recognit. (CVPR), 2013 IEEE Conf.* 3578–3585

(2013). doi:10.1109/CVPR.2013.459

[112]Mori, G. Guiding model search using segmentation. in *Proceedings of the IEEE International Conference on Computer Vision* **II,** 1417–1423 (2005).

[113]Nakariyakul, S. A comparative study of suboptimal branch and bound algorithms. *Inf. Sci. (Ny).* **278,** 545–554 (2014).

[114]Tian, T. P. & Sclaroff, S. Fast globally optimal 2D human detection with loopy graph models. in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition* 81–88 (2010). doi:10.1109/CVPR.2010.5540227

[115]Jiang, H. Finding Human Poses in Videos. 228–243 (2011).

[116]Rogez, G., Rihan, J., Ramalingam, S., Orrite, C. & Torr, P. H. S. Randomized trees for human pose detection. in *26th IEEE Conference on Computer Vision and Pattern Recognition, CVPR* (2008). doi:10.1109/CVPR.2008.4587617

[117]Forsyth, D. Object detection with discriminatively trained part-based models. *Computer (Long. Beach. Calif).* **47,** 6–7 (2014).

[118]Rothrock, B., Park, S. & Zhu, S. C. Integrating grammar and segmentation for human pose estimation. in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition* 3214–3221 (2013).doi:10.1109/CVPR.2013.413

[119]Torres Garcia, F. & Kropatsch, W. Top-down 3D Tracking and Pose Estimation of a Die Using Check-points. in *Proceedings of the 18th Computer Vision Winter Workshop* 102–109 (2013).

[120]Bray, M., Kohli, P. & Torr, P. PoseCut: simultaneous segmentation and 3D pose estimation of humans using dynamic graph-cuts. *Network* **1,** 642–655 (2006).

[121]Ionescu, C., Li, F. & Sminchisescu, C. Latent structured models for human pose estimation. in *Computer Vision (ICCV), 2011 IEEE International Conference on* 2220–2227. doi:10.1109/ICCV.2011.6126500

[122]Kolmogorov, V. & Zabih, R. What Energy Functions Can Be Minimized via Graph Cuts? IEEE Trans. Pattern Anal. Mach. Intell. 26, 147–159 (2004).