人体运动分析综述①

魏 玮, 王丹丹, 刘 命, 刘 静

(河北工业大学 计算机科学与软件学院, 天津 300401)

摘 要: 随着计算机视觉的快速发展, 对运动人体的分析已经成为计算机视觉领域一个很值得关注的课题. 这里对运动人体的分析主要指的全局的运动分析, 主要分析过程包括提取关节点, 再对提取的关节点跟踪, 最后计算特征点的三维坐标并且显示. 本文从这三方面当前的研究现状分别作了介绍.

关键词: 关节点; 跟踪; 模型; 标定

Survey of Human Motion Analysis

WEI Wei, WANG Dan-Dan, LIU Ming, LIU Jing

(School of Computer Science and Software, Hebei University of Technology, Tianjin 300401, China)

Abstract: With the rapid development of computer vision, analysis of the human movement has been widely noted in computer vision areas. This research on the human movement is about the whole human motion, including extracting joint points first, then trcking the feature points and calculating the 3D coordinates of the points and display. The paper introduces from current situation of the three aspects.

Key words: articulation points; track; model; calibration

随着计算机视觉领域中基于视频的人体运动的分析和研究的不断进展, 计算机图形和视觉研究人员对基于视频人体运动捕捉表示出极大的兴趣和希望. 研究人员从丰富的或者是随处可得的普通视频源中获取人体运动数据, 进而生成逼真的人体动画. 这里的视频可以是单摄像机拍摄的, 也可以是多摄像机拍摄的. 从普通图像资料中提取出人体运动立体特征可以分为手动方法和自动方法. 人体模型也有好多种, 不过在本文中的人体模型选择了人体关节骨架模型. 在进行运动人体的立体特征获取的时候一般分为如下几步, 首先是进行二维特征点也就是本文中的关节点的提取, 二维特征点的提取目前有两种情况. 一种是在图像上首帧手动标定出关节点位置; 一种是拍摄时在人体关节点处贴有特殊颜色的标记或者是发光物质[1]. 其次是关节点的运动跟踪, 通常采用的跟踪方法是稀疏光流 L_K 算法、粒子滤波[2]及 Condensation 算法. 再次是进行立体特征的获取, 目前根据对视频拍摄的相机的数目的不同, 有基于单视的和基于多视的, 这样单视的和多视的在立体特征获取上面方法有所不同, 并且从是否需要相机标定也有所不同, 有在相机标定的前提下获取的方法, 也有无需相机标定而获取运动人体立体特征数据的. 本文就在下面的内容中对这些不同的方法进行阐述.

1 特征点提取和跟踪

1.1 特征点提取

人体结构是由各个关节点连接组成的非刚体结构, 这使得人体的运动高度复杂, 不过各个关节点之间的连接又是刚体的. 我们可以认为对运动人体的分析实际上是对关节点运动的分析. 目前人体关节点的提取方法很多:

有标记方法: 运用比较广泛的是光学运动捕获系统, 该类系统是将具有反光标记的物体贴在运动人体关节点上, 之后跟踪反光的标记点的二维坐标. 这种

① 收稿时间:2012-06-27;收到修改稿时间:2012-08-29

方法是最原始的方法.

手工标注方法: 目前比较流行的提取运动人体关节点的方法是对于视频的首帧或者前两帧图像手动标注关节点位置, 后续再对上述标注的关节点跟踪. 由于该方法只需要对首帧图像标注, 实用性强. 但是该方法还是需要人工干预.

基于模型的方法: 针对人体的运动是非刚体运动, 陈坚[3]等人提出了运用人体关节模型的方法, 各个关节点之间构成一段链杆, 人体的 16 个关节点的二维坐标可以构成人体的状态向量, 建立人体的运动模型, 之后利用陈坚等人[3]的方法对运动模型学习. 这种方法的优点是运动人体无需标记, 没有人工干预. 缺点是对于运动模型的学习使得计算复杂.

基于运动目标分割方法: 对运动图像进行运动目标提取, 提取出人体侧影, 再运用文献[4]中的方法标注头, 躯干和四肢. 这个方法可以用于视频中存在一个或者多个运动目标, 不过保证每个侧影形成单独连通分量. 还可以对提取出来的运动目标细化, 细化后再提取出人体的关节点坐标. 这种方法的缺点是会受到运动目标检测方法的影响, 依赖性强.

1.2 运动跟踪

对于前面利用不同方法提取出来的关节点分别可以采用 L_K 光流法跟踪, 粒子滤波跟踪, Kalman 滤波预测关节点[5], Mean Shift[6]算法来进行跟踪, 还有一些是利用粒子滤波算法结合提取的人体骨骼化后的底层信息, 对人体各关节点进行跟踪和定位, 还有一种常用的方法是利用 L_K 光流法进行跟踪, 再利用 kalman 滤波线性跟踪预测跟踪错误的点.

2 三维点坐标计算

2.1 多目视觉中的立体特征获取

多目视觉指的是拍摄研究所用的视频或者图像的摄像机的数目至少是两个, 双目是多目的一个特例, 两个摄像机下的成像模型如图 1 所示.

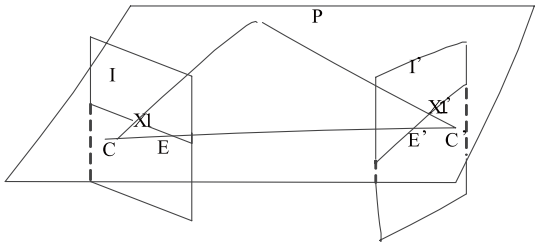

图 1　相机成像模型

其中 C 和 C'是两个摄像机的位置, I 和 I'是两个成像平面, P 为物体所在位置.

2.1.1 最小二乘法

双目视觉的三维重建主要方法是最小二乘法[7,8], 重建之前, 首先知道匹配点对和相机标定出来的内外参数, 计算出两个投影矩阵 P1 和 P2, 对于投影矩阵 P1, 令 P11, P12, P13 对应于 P1 的行向量, $(u_t, v_t, 1)^T$ 对应于 P1 的图像上的第 i 个匹配点图像像素齐次坐标, $\overline{X_i}$ 对应匹配点的空间齐次坐标, S 为常量因子, 则有

$$s\begin{bmatrix} u_i \\ v_i \\ 1 \end{bmatrix} = \begin{bmatrix} P_{11} \\ P_{12} \\ P_{13} \end{bmatrix}\overline{X_i} \Rightarrow \begin{cases} su_i = P_{11}\overline{X_i} \\ sv_i = P_{12}\overline{X_i} \\ s = P_{13}\overline{X_i} \end{cases} \Rightarrow \begin{cases} P_{13}\overline{X_i}u_i = P_{11}\overline{X_i} \\ P_{13}\overline{X_i}v_i = P_{12}\overline{X_i} \end{cases} \Rightarrow$$

$$\begin{cases} P_{13}\overline{X_i}u_i - P_{11}\overline{X_i} = 0 \\ P_{13}\overline{X_i}v_i - P_{12}\overline{X_i} = 0 \end{cases} \Rightarrow \begin{bmatrix} P_{13}u_i - P_{11} \\ P_{13}v_i - P_{12} \end{bmatrix}\overline{X_i} = 0$$

利用 P2 可以得到类似的等式

$$\begin{bmatrix} P_{23}u_i' - P_{21} \\ P_{23}v_i' - P_{22} \end{bmatrix}\overline{X_i} = 0 \tag{1}$$

联立以上两式可以得到

$$\begin{bmatrix} P_{13}u_i - P_{11} \\ P_{13}v_i - P_{12} \\ P_{23}u_i' - P_{21} \\ P_{23}v_i' - P_{21} \end{bmatrix}\overline{X_i} = 0 \tag{2}$$

从上式看出, 通过四个方程式求解三个未知数, 这种情况下, 通过最小二乘法原理求解出 $\overline{X_i}$ 的值, 再用 opengl 显示这些立体特征, 能够反映出人体的运动.

2.1.2 存在运动不确定性的三维重建

最小二乘法可能会出现计算所得三维点的坐标再投影到图像平面时有很大误差, 罗忠祥[9]等人提出了增加三维点投影到图像平面的点与实际跟踪到的图像点距离最小的约束条件使得三维重建结果更加准确.

2.2 单目视觉中的立体特征获取

单目视觉是拍摄视频或者图像所用的相机数目只有一个, 从而获取立体特征. 一般来说单相机下, 三维信息的恢复往往呈现病态, 因此如何在透视投影的约束下恢复高质量的运动人体立体信息是目前计算机视觉领域的一个具有挑战性的课题. 这样三维重建的方法又分为了基于标定的方法和非相机标定的方法.

2.2.1 基于相机标定的方法

实际应用中选定人体模型上胸部、腹部、左肩、

右肩、左臀和右臀作为算法输入，因为这些点之间的距离在人体运动过程中变化不大. 每个点对(X_{di},Y_{di})来自跟踪结果，而(x_{wi},y_{wi},z_{wi})的取值只要是各个点之间的距离比例符合给定人体模型可以随意给定. 要求是相机在第一帧中对相机标定，约束在拍摄初始采取双手下垂的自然直立姿态.

2.2.2 扩展模型起点选择和立体特征提取

假设刚体 P 上有三个任意点，且第 n 帧相机坐标 $P_i'(X_i',Y_i',Z_i')(i=1,2,3)$是已知的，计算 n+1 帧的相机坐标 $P_i(X_i,Y_i,Z_i)$ $(i=1,2,3)$，P_i 在投影平面中的点坐标 $p(u_i,v_i)$也是已知的，则由公式(1)可知 P_i 可以表示为仅含未知数 Z_i 的形式:

$$P_i=\left(\frac{u_i\bullet Z_i}{f},\frac{v_i\bullet Z_i}{f},Z_i\right) \tag{3}$$

同时由刚体不变性的属性可以知道，任意两点间的距离不便可以得到三个二次方程:

$$\left\|P_i-P_j\right\|=\left\|P_i'-P_j'\right\| \quad (i\neq j\in(1,2,3)) \tag{4}$$

结合式子(3)和(4)，有三个未知数 Z_1,Z_2,Z_3，对应三个非线性方程，通过梯度法求出一组近似解. 由于一些误差的可能，结果有很大不确定性，所以有人提出定义一个搜索空间[1]，用一组基于运动的连续性和刚体不变形假设最优准则求未知数 Z，算法见文献[1].

由于胸部、左肩、右肩组成三角形具有几何学中的刚体不形变的性质，并且不易于遮挡，跟踪效果好，故而将该三角形作为扩展模型起点. 利用上面描述的方法得到每帧模型扩展的起点，下面要做的工作将是从已知的扩展起点出发，扩展出人体模型上剩余的关节点的立体特征，从公式(3)可以知道，当二维图像坐标已知时，在求解对应点的深度信息的情况下，解的空间是一条直线，即连接光心和成像点得到的直线. 直线上所有点在投影平面上都成像于同一个点. 在这条直线上找到三维特征点，利用生理解剖学中的人体骨骼长度作为先验知识. 据上所说，已知了三个点的坐标值 P_0,P_1,P_2，设所求点 P_3，从已知点 P_0,P_1,P_2 中找到与所求点 P_3 相邻的点，比如说是 P_0，在直线上找到与已知点距离等于相应骨骼长度的点，但是从成像几何图 2 看出，会有两个满足条件的点，这就需要一些先验知识来消除这种二义性. 然后再利用文献[1]中的二义性消除方法消除点的二义性，依次进行下去，直到所有的关节点的立体特征都被提取出来.

这种方法需要相机标定来计算出相机的内外参数，但是和上面的方法相比，是单目视觉，比上面的方法相对简单. 不过提取出来的立体特征数据没有前面提到的方法精确.

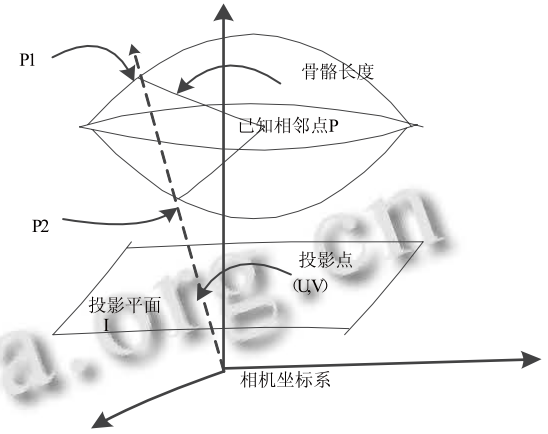

图 2 二义性的消除

2.2.3 非相机标定的方法

当目标和摄像机的距离足够大，目标反射出光线几乎是平行光. 当焦距 $f\to\infty$，使得 $f/z_c\to 1$. 根据(2)式得到$u=x_c$，$v=y_c$. 当目标位置接近于摄像机光轴且目标表面各点与摄像机之间的平均距离 $\bar{z}$ 远大于目标自身维度时也就满足了正交投影模型的条件，即 $\bar{z}>>dz$，根据(2)式可以得到目标任意关节点的空间三维坐标(x,y,z)与其投影点在图像坐标系下的二维坐标(u,v)间的关系为: $u=fx/z\approx fx/\bar{z}$，$v=fy/z\approx fy/\bar{z}$. 令 $s=f/\bar{z}$ 表示比例因子，可以表示为[10]:

$$\begin{bmatrix}u\\v\end{bmatrix}=s\begin{bmatrix}100\\010\end{bmatrix}\begin{bmatrix}x\\y\\z\end{bmatrix} \tag{5}$$

设有一肢体段 mn，其比例正交投影示意图如图 3.

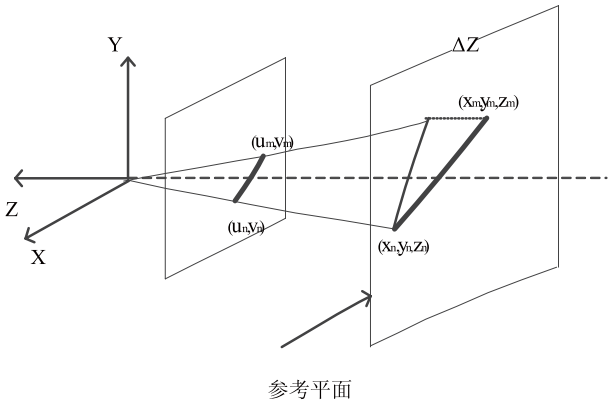

图 3 正交比例因子条件下一段肢体投影

经过运动人体跟踪，得到了 14 个关节点的二维坐标值 $\overline{X_k}=(u_k,v_k)$，$k=1,\ldots 14$，以其作为输入，根据比例正交投影模型求取各个关节点对应的相对三维坐标值 $\overline{X_k}=(x_k,y_k,z_k)$，根据上面的式子推导可以得到

$$\begin{cases} x_k = u_k / s \\ y_k = v_k / s \end{cases} \tag{6}$$

其中 u_k 和 v_k 已知. 只需 s 便得 x_k 和 y_k，深度信息 z_k 需要肢体长度信息. l_{mn} 表示关节点 mn 的肢体段长度[11]，由空间点距离公式得:

$$l_{mn}^2=(x_m-x_n)^2+(y_m-y_n)^2+(z_m-z_n)^2 \tag{7}$$

由(3)(4)两式得到关节点 m 和 n 的相对深度为:

$$\Delta z=|z_m-z_n|=\sqrt{l_{mn}^2-\left((u_m-u_n)^2+(v_m-v_n)^2\right)/s^2} \tag{8}$$

由(5)式来看，有了肢体长度，两关节点的相对深度是关于比例因子 s 的函数. 令图像中两关节点之间的以像素为单位的初始化肢体长度为 d_{mn}，有比例正交投影关系得到

$$l_{mn}=d_{mn}/s \tag{9}$$

根据等式(5)和(6)得到

$$\Delta z=|z_m-z_n|=\sqrt{d_{mn}^2-\left((u_m-u_n)^2+(v_m-v_n)^2\right)}/s \tag{10}$$

由(7)式可以看出只要确定了比例因子 s，以关节点 m 或者 n 中任意一点作为参考点，便可以求出另一点的相对三维坐标值，并且在三维人体运动行为分析时，也是只要知道关节点之间的相对坐标值就可以进行分析. 这里把人体的模型表示成了一个树模型，选择了 $\overline{X_{root}}=(x_{root},y_{root},z_{root})$，并将其放置于参考平面上 $(z_{root}=0)$. 如此运动人体立体特征可以提取出来，这样利用 opengl 将三维人体显示出来.

这种方法计算相机的内外参数，方法相对比较简单，结果也可以很直观的体现人体的运动，不过这种方法求出来的是人体的相对立体数据，不是精确的立体特征数据.

实验结果如下，其中图 4 是双目视觉下手动获取关节点的坐标位置，图 5 是单目视觉下的效果图.

3 结语

对运动人体分析的实用性已经随着电影动画等的发展逐渐增强. 通过分析，我们发现这项研究已经从之前的依靠硬件发展到现在研究趋于自动化. 从实验结果我们不难发现，在人体关节点处贴上发光或者反光的标记很容易找到人体关节点的位置，但是硬件设施比较贵并且不方便；在运动人体的首帧标注关节点的位置的方法准确并且不需要硬件设施的帮助；双目标定下的运动人体分析得出的分析结果比较准确，过程比较复杂，并且普通的图像帧都无法相机标定；单目非标定的方法简单，不过精确度不高. 目前比较适用的是单目非标定的方法，今后对其进一步研究并且提高精确度.

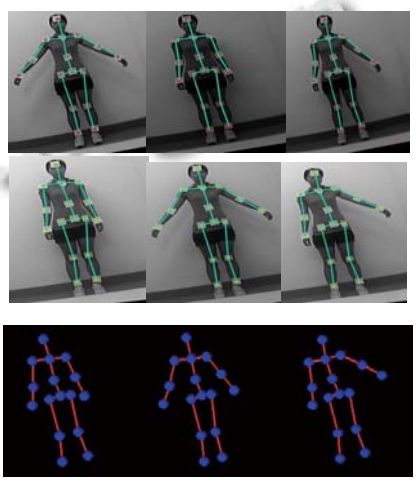

图 4　双目标定下的三维显示

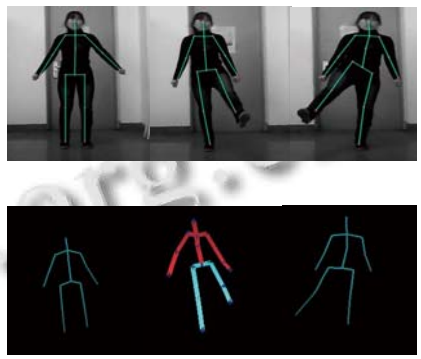

图 5　单目非标定下的三维显示

参考文献

1 陈家实,庄越挺,朱强等.透视投影下三维运动重建.计算机辅助设计与图形学学报,2002,11(14):1041–1046.

2 邓宇,李振波,李华.基于视频的三维人体运动跟踪系统的设计与实现.计算机辅助设计与图形学学报,2007,19(6):769–780.

3 陈坚,王文成,吴恩华.单目视频中无标记的人体运动跟踪.计算机辅助设计与图形学学报,2005,17(9):2033–2039.

(下转第 52 页)

越的时频聚集性. 同时, 由于 LPFT 的线性特性, 它在处理多分量语音信号时不会受到交叉项的干扰. LPFT 的优越性在语音的过渡带 0.2-0.3 范围内更为明显, 因为这个部分语音的频率变化较为迅速.

3 结语

本文利用局部多项式傅里叶变换 LPFT 处理语音信号, 并建立了基于LPFT的语音处理GUI系统. 直观地将 LPFT 分析与时域、频域、STFT 和 WVD 分析进行对比, 从而验证了 LPFT 是一种更为优化的处理语音信号的工具. 该系统界面友好, 易于操作, 是有效处理语音信号的可视化平台.

致谢

该论文还受到以下项目资助: 杭州师范大学科研启动基金(2011QDL021), 杭州师范大学本科生创新能力提升工程项目, 杭州师范大学实验室开放项目.

参考文献

1 胡航.语音信号处理.第 2 版.哈尔滨:哈尔滨工业大学出版社,2005.

2 Cohen L. Time-Frequency Analysis. New Jersey: Prentice-Hall, 1995.

3 张贤达.现代信号处理.第 2 版.北京:清华大学出版社,2002.

4 Katkovnik V. A new form of Fourier transform for time-varying frequency estimation. Signal Processing, 1995,47(2): 187–200.

5 Djurovic I, Thayaparan T, Stankovic L. SAR imaging of moving targets using polynomial Fourier transform. IET Signal Processing, 2008,2(3):237–246.

6 Li XM, Bi GA, Ju YT. Quantitative SNR analysis for ISAR imaging using LPFT. IEEE Trans. on Aerospace and Electronic Systems, 2009,45(3):1241–1248.

7 Djukanovic S, Dakovic M, Stankovic L. Local polynomial Fourier transform receiver for nonstationary interference excision in DSSS communications. IEEE Trans. on Signal Processing, 2008,56(4):1627–1636.

8 Katkovnic V, Gershman A. A local polynomial approximation based beamforming for source localization and tracking in nonstationary environments. IEEE Signal Processing Letters, 2000,7(1):3–5.

9 王光艳,赵晓群,王霞.语音信号时频特征显示系统的设计和仿真.计算机工程与应用,2010,46(29):73–75.

10 题原,张劲松.基于 MATLAB 的语音信号采集和分析系统的可视化设计.齐齐哈尔大学学报,2006,22(6).

(上接第 4 页)

4 陈成,肖俊,庄越挺.单目视频人体三维运动高效恢复.计算机辅助设计与图形学学报,2009,21(8):1118–1126.

5 雷涛,罗薇薇,樊养余,等.复杂环境下的运动人体骨架提取算法.计算机应用研究,2010,27(8):3194–3197.

6 于若飞.基于 MeanShift 的运动人体骨架重构方法.科学技术与工程,2011,11(21):5220–5222.

7 马颂德.计算机视觉——计算理论与算法基础.北京:科学出版社,1997.

8 Furukawa Y, Ponce J. Accuate, dense and robust multi-view stereopsis. IEEE Pattem Analysis and Machine Intelligence, Aug, 2008,11(1):1–14.

9 罗忠祥,庄越挺,潘云鹤,等.基于视频的运动捕获.中国图像图形学报,2002,7(8):752–758.

10 Taylor CJ. Reconstruction of Articulated Objects from point Correspondences in a Single Uncalibrated Image. Computer Vision and Image Understanding, 2000,349–363.

11 Zou BJ, Chen S, Shi C, et al. Automatic reconstruction of 3D human motion pose from uncalibrated monocular video sequences based on markerless human motion tracking. Pattern Recognition, 2009: 1559–1571.