中国科学: 信息科学 2011 年 第 41 卷 第 6 期: 659–674
www.scichina.com info.scichina.com
《中国科学》杂志社
SCIENCE CHINA PRESS
论 文

三维动态脚型测量方法

高飞, 王青, 耿卫东*, 潘云鹤

浙江大学 CAD&CG 国家重点实验室, 杭州 310027
* 通信作者. E-mail: gengwd@zju.edu.cn

收稿日期: 2010–05–12; 接受日期: 2010–08–23
国家科技支撑计划 (批准号: 2006BAF01A44) 和浙江省科技攻关项目 (批准号: 2006BAF01A44) 资助

摘要 基于多视几何原理来恢复和重建随时间变化的三维柔性物体是当前计算机视觉和三维扫描领域的研究热点之一. 本文利用主动式标记点方法, 提出并实现了在三维模型指导下鲁棒地从多视点视频中恢复和重建三维脚型的方法. 它首先基于传统立体视觉方法重建出参考帧中的三维脚模型, 然后在此基础上, 通过构建三维模型顶点运动投影的速度矢量场来建立相邻帧之间的标记点对应关系, 鲁棒地跟踪复杂背景中的三维动态目标以及相邻帧间运动幅度较大的三维动态目标. 而对于在运动过程中由于遮挡等原因所造成的标记点缺失, 则在三维参考模型的指导下, 利用数字几何处理方法可靠地在三维空间中估计出相应的标记点, 从而有效保证了三维脚模型在运动过程中的完整性. 实验结果表明: 该方法可以鲁棒、可靠地获得动态的三维脚型.

关键词 动态形状 脚型建模 特征跟踪

1 引言

随时间变化的三维柔性物体的建模和测量在医学、计算机辅助设计、虚拟现实与数字娱乐等领域都有着重要的应用. 就三维脚型的获取而言, 目前一般把静止状态下的柔性脚型近似为刚体的三维脚型, 然后采用激光、结构光、以及立体视觉等技术来建构相应的三维模型 [1−−3]. 但在行走过程中, 人的足部形状会发生显著的变化, 如果只能静态地获取三维脚型, 将会在很多应用中受到明显的限制, 例如, 在足部运动学以及动力学研究中, 需要了解各个相关时刻的足部形状变化情况; 在鞋类产品的个性化定制服务中, 需要通过人脚在运动过程中的变形来分析其在鞋子中的受力情况以进行舒适度的评估等. 而激光扫描方法由于其自身原理的局限性, 一般不适合于动态物体的扫描. 结构光方法通过高速相机同步拍摄投射在目标形状上的结构光栅, 然后根据被目标形状调制后的光栅扭曲情况来推演每一帧的三维模型 [4,5], 但其对环境光要求较高, 一般不能用于开放性环境中. 而且, 目标在运动过程中所形成的自遮挡会导致三维动态形状获取的不完整. 为了克服这些局限性, 本文尝试在基于立体视觉的静态脚型测量系统 [6] 的基础上, 进一步研究在开放环境下的基于多视几何原理的三维动态脚型获取技术.

目前, 基于多视几何原理来获取动态三维物体的测量方法可以分为 3 类: 基于空间分割的方法、基于模型的方法和基于特征的方法. 基于空间分割的方法的基本原理是通过将目标所占据的空间从

引用格式: 高飞, 王青, 耿卫东, 等. 三维动态脚型测量方法. 中国科学: 信息科学, 2011, 41: 659–674

视域空间中分离来获取目标的形状, 例如, 空间雕刻法 [7] 和体素着色法 [8] 分析每个体素的可见性来获取目标所占据的体素集合来获取目标形状; Matusik 等 [9] 和 Moezzi 等 [10] 通过求解多相机目标投影轮廓与灭点构成的视域锥的公共部分, 即 Visual hull 方法 [11], 分别获取前景目标形状的三角网格形式和体素表达形式. 这类方法虽易实现, 但其结果较为粗糙, 精度的提高依赖于视点的数量. 基于模型的形状获取方法通过定义一个包含纹理信息的标准模型, 然后将该标准模型和视频帧图像拟合优化, 把预定义的代价函数最小化, 从而获得目标的三维模型. 例如, Decarlo 等 [12] 和 Pighin 等 [13] 将一个可变形的三维模型和视频目标拟合, Blanz 等 [14] 和 Vlasic 等 [15] 用一组样本几何模型张成了一个形状空间, 通过优化组合系数在形状空间中搜索与目标物体最近似的形状. 然而, 这类方法获取模型的精度依赖于标准模型以及初始条件的质量、精度难以与基于特征的获取方法媲美. 基于特征的形状获取方法通过跟踪特征点的帧间位移, 并拟合重建出的新特征来获取三维形状. Aguiar 等 [16,17] 基于目标自有特征重建目标的动态形状, 但为了获取细致的模型, 需要人工添加大量特征及匹配信息, 并且其精度受限于目标自有特征的数量, 难以保证对于如人脚、茶壶等表面平坦部分较多的目标的获取精度. Pighin 等 [18] 则通过采用稀疏标记点集对人脸进行标记, 并跟踪标记点运动的方式来恢复人脸表情. Herda 等 [19] 和 Moeslund 等 [20] 将标记点运动数据参数化到骨骼上面, 来跟踪基于骨骼的动作. 尽管这类方法具有较高的精度, 但是标记点采样较稀疏, 无法进行目标形状获取. Park 等 [21] 为了克服这个缺点, 采用了几百个标记点来提取人体皮肤的变形, 虽然结果非常逼真, 但该方法所采用的光学标记点的体积相对于脚型较大, 难以用于脚型形状的精确获取, 并且手工布置与清除标记点的工作繁重. 在鲁棒性方面, 上述这些方法普遍要求干净的背景、环境光照的一致性、帧间目标变化幅度小、特征的空间及纹理差异性大等条件, 然而这些条件限制了其使用范围.

本文提出了一种在开放环境下, 基于主动式标记点、具有高鲁棒性的动态脚型形状获取方法, 其主要技术贡献为: (1) 通过基于三维模型的跟踪方式有效地保持了标记点的数量, 提高了测量结果的精度; (2) 通过基于局部微分几何的方法在三维空间中估计缺失标记点的几何位置, 保证了结果模型的完整性; (3) 跟踪过程中运动矢量场的构造避免了对图像二维纹理的依赖, 一方面大大增强了抗噪声能力, 另一方面放宽了对帧间目标运动幅度的要求, 拓展了同帧率相机所能跟踪的目标的运动速度范围, 另外, 算法鲁棒性的提高使整个跟踪重建过程可以自动完成, 避免了传统基于立体视觉的三维测量方法所涉及的特征添加、匹配关系修正等繁琐的交互工作.

2 问题描述及总体方案

动态脚型获取即获取视频中每一帧所对应的三维脚型. 设采集动态脚型运动视频的相机视点集合为 $\{A_1, A_2, \ldots, A_c\}$ (本文中 c=10), 每个视点所采集的视频共 n 帧, 记 A_j 获取的视频帧序列为 $\{I_j^1, I_j^2, \ldots, I_j^c\}$, 本文的目标为从所有视点的同步视频帧 $\{I_1^i, I_2^i, \ldots, I_c^i\}$中求解每一帧对应的三维脚模型 $M^i(1 \leqslant i \leqslant n)$, 其中 M^i 由三角网格表示, 第 i 帧脚模型记为 $M^i = \{\mathcal{P}^i, \mathcal{T}^i\}$, 其中 $\mathcal{P}^i, \mathcal{T}^i$ 分别表示第 i 帧模型的顶点集和三角形集, 而 $\boldsymbol{P}_j^i$ 为第 i 帧网格模型上第 j 个顶点, 每个顶点对应着一个标记点, 不同图像中对应于同一顶点的标记点被称为同源标记点. 类似地, 文中统一用上标表示帧序号, 下标表示集合中的元素序号.

整个动态脚型测量方法由 5 部分组成: 视频采集、参考模型重建、标记点提取、标记点跟踪、模型重建 (如图 1 所示).

本文采用了和我们的静态脚型系统 [6] 中相同的图像采集系统配置, 通过 10 台分辨率为 800*600,

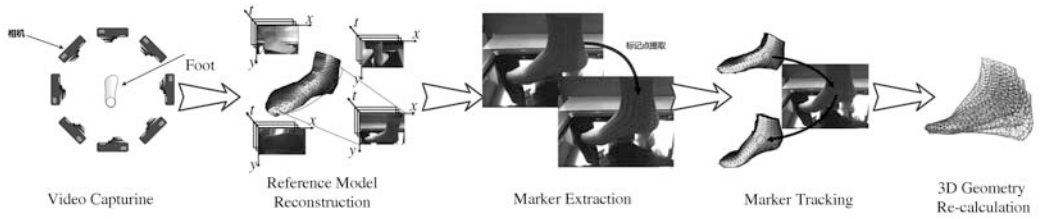

图 1 总体的算法框架

Figure 1 The framework of our method

帧率为 15 f/s 的同步摄像机以图像共享的方式分成 6 对双目相机, 分别对脚型的 6 个区域进行视频采集, 采用 Zhang[22] 提出的平面标定法进行相机标定, 与静态脚型采集系统 [6] 相同. 获得脚型的运动视频后, 首先根据视频的第一帧图像, 通过文献 [6] 所描述的方法重建三维模型 $M^1 = \{\mathcal{P}^1, \mathcal{T}^1\}$. 同时, 用于恢复同一标记点的一对双目相机被定义为该标记点的双目跟踪相机. 然后将 M^1 作为初始参考模型, 对后续帧所对应的脚模型进行恢复. 该动态脚型获取的基本算法步骤如下 (其中, $1 \leqslant i \leqslant n-1$):

步骤 1 基于第 i 帧模型 M^i 跟踪第 $i+1$ 帧的标记点, 跟踪到的标记点集合记为 $\mathcal{F}^{i+1}$(见第 3 节);

步骤 2 根据 $\mathcal{F}^{i+1}$ 重建第 $i+1$ 帧可见标记点对应的模型顶点集合 $\mathcal{Q}^{i+1}$;

步骤 3 计算第 $i+1$ 帧中缺失标记点对应的模型顶点的三维位置, 所构成集合记为 $\mathcal{P}^{i+1}_{\text{invisible}} = \mathcal{P}^{i+1} - \mathcal{Q}^{i+1}$(详见第 4 节);

步骤 4 将 M^i 的拓扑关系移植到 M^{i+1} 中, 即 $\mathcal{T}^{i+1} = \mathcal{T}^i$, 得到 M^{i+1}.

实现上述步骤的主要技术关键包括两个方面:

(1) 各向同性稠密标记点的鲁棒跟踪. 获取结果的精度取决于采样的密度, 当采样标记点变密后, 对于各向同性的标记点跟踪较为困难. 人脚属于柔性物体, 其在行走过程中的形变属于非刚体变化, 这导致每个标记点的运动方式不同, 采用类似 Park 等 [21] 的简单分块求解刚体变形参数的方法来跟踪标记点会使获得的形状失真. Bastiano 等 [23] 和 Wagner 等 [24] 采用基于特征纹理的 SIFT 匹配方法 [25] 建立动态视频帧间特征的对应关系, 但这种方法牺牲了视频帧间的连续性信息, 在特征纹理相似的情况下, 错误率较高. Tsutsui 等 [26] 和 Brox 等 [27] 等大量方法采用光流法 [28] 通过场景图像颜色变化信息的连续性来进行特征跟踪, 然而光流法仅在帧间场景变化极小的条件下效果较好. Bergen 等 [29] 提出层次化光流场来改善这个问题, 但是, 复杂背景中的目标运动交替地遮挡部分背景, 破坏了图像整体颜色变化的连续性, 大大降低了光流法跟踪的正确性. 此外, 图像噪声使特征的自动跟踪问题变得更加复杂.

(2) 缺失标记点的可靠处理. 由遮挡、噪声等因素造成的缺失标记点的位置估计是保持模型完整性的重要问题. 标记点在某些时刻的缺失破坏了时间维的连续性, 这样就需要通过标记点缺失前和重新出现后的轨迹来估计其位置. 在标记点较稠密且邻域属性相近的情况下, 建立正确的轨迹对应关系较为困难. 同时, 标记点在缺失的时间段中, 其运动信息的缺乏使其轨迹拟合的误差控制以及标记点运动的速度估计变得更加困难.

针对上述难点, 本文提出了一个相对完整的解决方案. 在跟踪标记点方面, 同样利用时间维连续性, 与光流法不同的是本文基于前一帧三维模型点运动的连续性而非图像灰度变化的连续性跟踪标记点, 由于不依赖图像灰度属性, 既避免了对帧间目标运动幅度的严苛要求, 又解决了特征间距离相近、纹理差异性不高的问题. 通过引入空间维的连续性对标记点集合进行整体跟踪, 在保证了模型精度的

同时, 提高了跟踪结果的鲁棒性和跟踪过程的效率; 在处理缺失标记点方面, 利用运动目标局部变化的连续性, 通过三维模型的微分信息来估计缺失点的几何位置, 从而避免了在二维空间中对其进行跟踪和估计的不确定性, 保持了模型的完整性. 将估计出的缺失标记点投影到图像空间用于跟踪下一帧的标记点, 保持了缺失标记点运动轨迹的连续性和完整性, 使算法可以鲁棒地跟踪那些" 时隐时现" 的标记点, 有效改善了跟踪质量.

3 三维模型指导下的标记点跟踪

当目标运动超过相邻标记点的间距 (在目前的系统中约为 10 mm 左右) 时, 标记点间的邻域几何和纹理的相似性会造成运动分析的多义性, 因此, 本文算法所能允许的前后帧之间的最大运动幅度为 10 mm 左右. 为了叙述简洁, 本节以基于第 i 帧模型 M^i 跟踪某个相机视频第 $i+1$ 帧中的可见标记点为例对可见标记点的跟踪方法进行说明, 对于其他相机视频采用相同的跟踪方法. 具体算法如下:

步骤 1　提取第 $i+1$ 帧图像标记点集合, 记为 $\mathcal{R}^{i+1}$;

步骤 2　将 $\mathcal{P}^i$ 投影到第 $i+1$ 帧图像中, 在 $\mathcal{R}^{i+1}$ 中搜索与投影点相对应的标记点, 并将 $\mathcal{R}^{i+1}$ 中找到对应关系的元素构成的集合记为 $\mathcal{F}^{i+1}$, 其中的元素记为 f_j^{i+1};

步骤 3　剔除 $\mathcal{F}^{i+1}$ 中的噪声点得到最终的标记点集合.

上述算法中通过模型投影点来跟踪下一帧标记点的优点有两个方面: (1) 对于一对双目相机, 与同一模型顶点相对应的两个 $\mathcal{F}^{i+1}$ 中的元素就自然构成了一对同源标记点, 从而重建出第 $i+1$ 帧模型的可见顶点集合 $\mathcal{Q}^{i+1}$, 避免了单独建立双目同源关系的处理过程; (2) 由于跟踪每一帧图像标记点是通过与前一帧模型顶点的匹配来实现的, 所以如果前一帧图像中对应于某模型顶点的标记点发生了缺失, 它在前一帧图像中的位置则自然地由前一帧模型投影点的位置代替, 从而保证了当其重现时可以及时继续对其跟踪.

传统的标记点提取如 Harris 算法 [30] 和 SUSAN 算法 [31] 等一般基于像素灰度的微分属性进行特征识别, 但本文所采用的标记点由均匀缝制在白色袜子上的直径 1 mm 左右的黑色线扣构成, 其在图像中的投影为一小块区域, 通过基于特征微分属性的提取方法所获得的标记点位置通常位于投影区域的边缘. 为了提高标记点提取的精确性和鲁棒性, 本文将标记点提取转化为图像分割问题, 采用 ELT 算法 [32] 将标记点区域从其背景中分离出来, 经尺度约束后, 将标记点区域中心作为标记点提取位置.

由于背景以及光照的复杂性, 一般无法得到干净的待匹配标记点集合, $\mathcal{R}^{i+1}$ 中伪标记点的存在以及部分标记点的缺失使对第 $i+1$ 帧图像标记点的跟踪问题变为在 $\mathcal{R}^{i+1}$ 中寻找一个最大子集, 使其与第 i 帧模型投影点集合的一个子集构成一一对应关系. 这种情况下, 标记点间纹理差异性不高及局部几何关系的高度近似性会使单独为每个标记点寻找对应点的方法正确率低, 而且抗噪声能力差. 为了提高算法的稳定性和匹配的正确率, 本文将运动中的脚型变形过程定义为一个连续的运动矢量场, 利用标记点间的几何结构来对标记点进行整体跟踪. 矢量场的连续性使算法的抗干扰能力大大增强 (在标记点匹配错误的情况下运动矢量场会呈现出局部极不光滑的特性), 同时避免了穷举搜索带来的时间复杂度. 运动矢量场定义如下:

$$\boldsymbol{V}(x,y,t)=(u,v)^{\mathrm{T}}, \tag{1}$$

其中, $\boldsymbol{V}(x,y,t)$ 定义了图像点 (x,y) 在 t 时刻的速度, $(\frac{\partial \boldsymbol{V}(x,y,t)}{\partial x}, \frac{\partial \boldsymbol{V}(x,y,t)}{\partial y})$ 定义了速度在二维空间中的变化率. 公式 (1) 定义了二维投影平面上的运动矢量场, 它是三维空间矢量场进行比例缩放后的线性投影结果, 其依然保持了三维空间矢量场的连续性. 在离散的情况下, 袜子上的标记点可以被看作是

对矢量场在空间维的采样, 视频中的每一帧可以被看作是对矢量场时间维的采样. 标记点的跟踪可以通过分别插值矢量场的空间维和时间维实现.

本文基于局部优化来迭代重建第 i 帧的运动矢量场并搜索第 $i+1$ 帧的对应标记点. 步骤如下:

步骤 1 建立标记点间的邻域关系. 在第 i 帧模型中, 将属于同一跟踪相机的每个顶点的邻域顶点定义为其一环邻域内属于该相机的顶点中距离其最近的 4 个顶点 (当一环邻域少于 4 个顶点时, 则全部定义为邻域顶点), 并将该邻域关系定义到它们在该跟踪相机中所对应的投影点上;

步骤 2 根据矢量场的时间维连续性, 在第 i 帧中选择部分投影点, 建立与第 $i+1$ 帧原始标记点集合 $\mathcal{R}^{i+1}$ 元素间的匹配关系;

步骤 3 从第 i 帧所有未处理的投影点中选取拥有最多的已处理邻居的投影点作为本次迭代的待处理点. 如图 2 所示, 设所选定的待处理点为 $\boldsymbol{P}_m^i$ 的投影点 $\boldsymbol{p}_m^i$, $\boldsymbol{p}_m^i$ 的邻域集合记为 $\mathcal{N}(m)$, 当前已经处理完毕的投影点集合记为 $\mathcal{W}(i,m)$, 图中位于 $\mathcal{W}(i,m)$ 区域中带箭头线段标示的点表示当前已处理的投影点, 箭头线段表示已处理点的速度矢量;

步骤 4 根据已处理邻居的速度来估计所选点 $\boldsymbol{p}_m^i$ 的初始速度 $\boldsymbol{v}_m^i$:

$$\boldsymbol{v}_m^i = \frac{\sum_{n\in\mathcal{N}(m)\bigcap\mathcal{W}(i,m)}\frac{1}{L^i(m,n)}\boldsymbol{v}_n^i}{\sum_{n\in\mathcal{N}(m)\bigcap\mathcal{W}(i,m)}\frac{1}{L^i(m,n)}}, \tag{2}$$

其中, $L^i(m,n)$ 表示编号为 n 的投影点 $\boldsymbol{p}_n^i$ 在此时到 $\boldsymbol{p}_m^i$ 的距离;

步骤 5 在 $\mathcal{R}^{i+1}$ 中搜索距离 $\boldsymbol{p}_m^i+\boldsymbol{v}_m^i$ 最近的标记点, 当所找到的标记点到 $\boldsymbol{p}_m^i$ 的距离小于标记点平均间距的 1.5 倍时, 将其作为 $\boldsymbol{p}_m^i$ 对应的下一帧标记点 $\boldsymbol{f}_n^{i+1}$, 并将其加入最终标记点集合 $\mathcal{F}^{i+1}$ 中, 否则将 $\boldsymbol{p}_m^i$ 作为下一帧时刻位置不可跟踪的点, 其对应的 $\boldsymbol{P}_m^{i+1}$ 采用第 4 节的方法进行估计;

步骤 6 当 $\boldsymbol{p}_m^i$ 为可跟踪点时, 修正 $\boldsymbol{v}_m^i$,

$$\boldsymbol{v}_m^i = \boldsymbol{f}_n^{i+1} - \boldsymbol{p}_m^i, \tag{3}$$

否则暂置 $\boldsymbol{v}_m^i=0$, 在 M^{i+1} 重建完毕后, 令 $\boldsymbol{v}_m^i=\boldsymbol{p}_m^{i+1}-\boldsymbol{p}_m^i$;

步骤 7 检查第 i 帧中是否还有未处理过的投影点, 如果有则返回步骤 3 开始执行, 否则结束迭代过程, 投影点集合元素在第 $i+1$ 帧时刻的位置搜索完毕.

在上述迭代过程的步骤 2 中, 第 i 帧与第 $i+1$ 帧的初始匹配关系可以通过选取种子标记点, 在种子标记点位置的时间维插值运动矢量场并搜索种子标记点在下一帧时刻的对应点来实现. 为了包含更多的目标运动信息, 本文从距离第 i 帧投影标记点重心最近的、在时间维运动连续的 s 个点中选择一个投影点作为种子标记点 (本文中 s 取 5), 被选的 s 个标记点记为 $\boldsymbol{p}_{j_k}^i(1\leqslant k\leqslant s)$. 如图 3 所示, 对每一个 $\boldsymbol{p}_{j_k}^i$ 沿着 $\boldsymbol{v}_{j_k}^{i-1}$ 的方向的小幅张角范围内 (图中棕色区域范围), 在 $\mathcal{R}^{i+1}$ 的元素中 (图中灰色亮点) 搜索距离其最近的点作为 $\boldsymbol{f}_{j_{k'}}^{i+1}$ 并修正 $\boldsymbol{v}_{j_k}^i=\boldsymbol{f}_{j_{k'}}^{i+1}-\boldsymbol{p}_{j_k}^i$. 然后, 对这些待选种子标记点的速度进行聚类, 剔除不满足速度的空间连续性的异常点. 在剩下的待选种子标记点中选取速度模的中值点作为最终的种子标记点. 当 $i=1$ 时, $\boldsymbol{v}_{j_k}^0$ 初始化为指向脚型运动初始方向的单位向量.

对于那些不可见顶点的投影跟踪, 会干扰运动矢量场的建立, 所以在进行模型投影时, 我们将利用 $\mathcal{P}^i$ 中待投影顶点的法向与相机视线方向的夹角来对每个待投影点的可见性进行检查, 当顶点运动到跟踪相机对应的物体背面时, 暂不对其进行投影跟踪直到其再次可见为止.

$\mathcal{R}^{i+1}$ 中的噪声点主要来自两个部分, 其中一部分是未建立对应关系的标记点, 它们由于不满足运动的时间维连续性, 被自然地排除在 $\mathcal{F}^{i+1}$ 外; 另一部分是由于实际的光照条件和袜子褶皱造成的

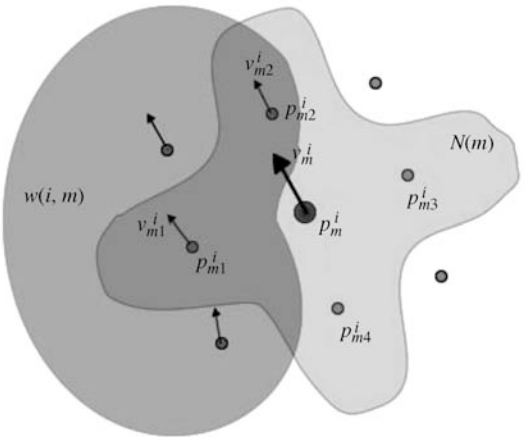

图 2　同一帧投影标记点的初始速度估计

Figure 2　Initial velocity estimation at $\boldsymbol{p}_m^i$

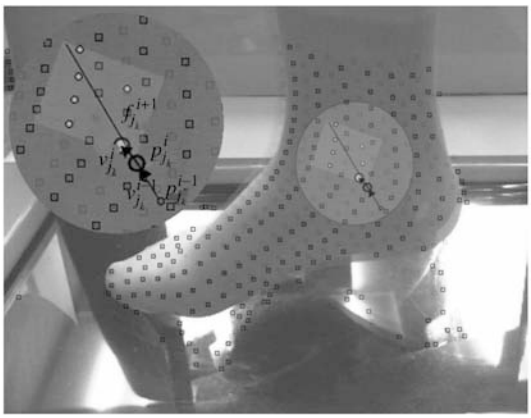

图 3　待选种子标记点集合元素的初始匹配关系建立

Figure 3　Searching for the matching markers for candidate seeds

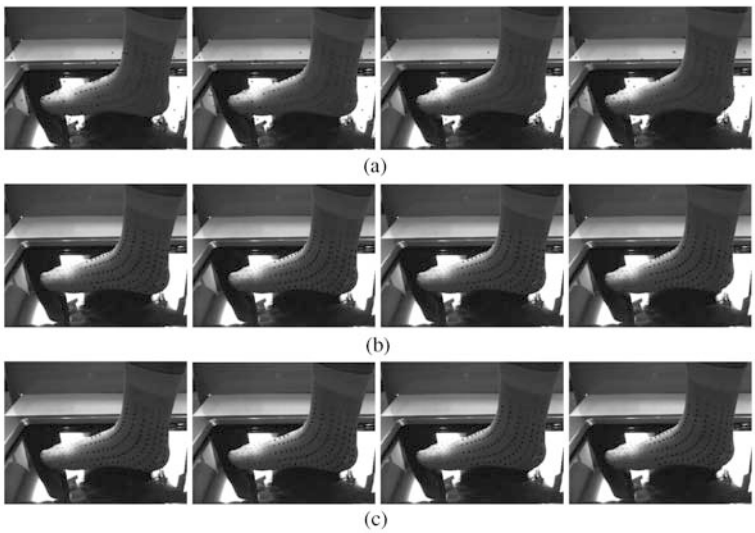

图 4　连续帧标记点跟踪结果

Figure 4　Marker tracking results for successive frames

(a) The markers tracked by SIFT algorithm; (b) the markers tracked by hierarchy optical flow algorithm; (c) the markers tracked by our method

$\mathcal{R}^{i+1}$ 中元素位置距离真实标记点偏差较大的点, 它们易被作为 $\mathcal{F}^{i+1}$ 的元素. 当第二部分点的误差严重时, 会给重建的三维模型带来极大的精度损失. 因此, 在视觉重建前需要先检验标记点位置的正确性. 根据极线约束, 一对同源标记点会分别落在彼此对应的极线上. 可通过每对同源标记点到各自对应极线的距离和来评价标记点位置的偏差程度, 当距离和超过预先指定的阈值时, 该对同源标记点将被剔除. 相关阈值的大小由相机标定参数的精度决定.

图 4 分别显示了采用本文方法以及 SIFT 算法和层次化光流法对于同一组数据 (脚后跟抬起过程

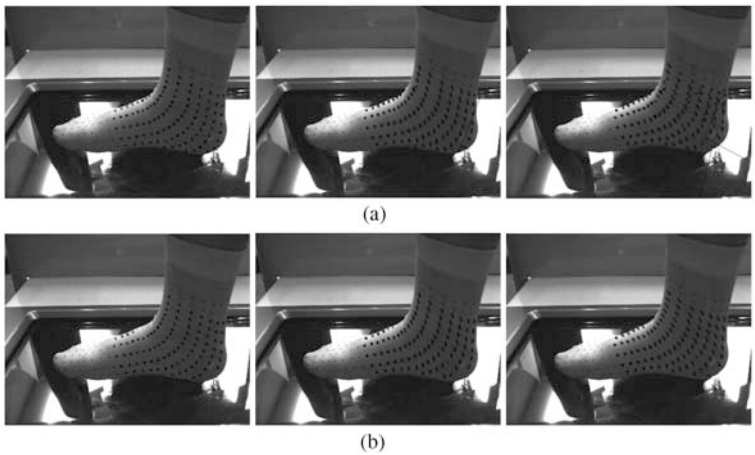

图 5 连续帧标记点对应关系

Figure 5 Matching the markers from consecutive frames

(a) The corresponding relations by hierarchy optical flow algorithm; (b) the corresponding relations obtained by our method

的帧序列) 的跟踪结果. 其中, 图中的方框中心标示了所跟踪到的标记点位置. SIFT 算法主要依赖特征的纹理信息在运动过程中的不变性对特征进行匹配, 在本文的问题中, 由于脚型的特征相对于整个场景的纹理并不明显, 且属于各向同性纹理, 所以图 4(a) 中 SIFT 算法所跟踪到的标记点较少. 图 4(b) 是基于层次化光流 (5 层光流场) 的跟踪结果, 前两帧由于目标的帧间变化幅度较小, 该方法的跟踪结果较为理想, 但在后面两帧, 帧间目标运动的距离变大, 导致错误点增多了. 而本文的算法在跟踪过程中较少依赖纹理信息, 从而稳定性较好. 图 5 通过标记点的帧间对应关系来检验光流法和本文方法跟踪结果的正确性, 其中, 直线相连的两个方框点表示跟踪结果中前后帧对应的两个标记点. 光流法对于标记点的跟踪过程部分地依赖于纹理信息, 这样在各向同性的特征匹配问题中, 当目标运动幅度慢慢增大时渐渐出现了匹配错误的情况 (如图 5(a)), 而本文算法在这种情况下保持了正确的跟踪结果 (图 5(b)).

4 基于微分几何的缺失标记点估计及三维模型重建

在柔性物体运动过程中, 当采样帧率相对于目标形变速度较高时, 可以假设模型的局部几何属性在相邻帧间变化相对很从而可以利用 $\mathcal{P}^i$ 与 $\mathcal{Q}^{i+1}$ 的对应关系将 M^i 的微分属性移植到 $\mathcal{Q}^{i+1}$ 所在的模型中对 $\mathcal{P}^{i+1}_{\text{invisible}}$ 进行估计. 同时, 将 M^i 的拓扑关系移植到 $\mathcal{P}^{i+1}$ 上就得到了 M^{i+1}, 如图 6 所示. 通过 $\mathcal{P}^i$ 元素的投影与 $\mathcal{F}^{i+1}$ 元素间的对应关系容易得到 $\mathcal{P}^i$ 与 $\mathcal{Q}^{i+1}$ 元素间的对应关系, 利用该对应关系, 本节将通过 M^i 和 $\mathcal{Q}^{i+1}$ 估计 $\mathcal{P}^{i+1}_{\text{invisible}}$ 和 M^{i+1}.

Laplacian 坐标是记录顶点局部微分属性的一种较好的形式. M^{i+1} 的微分表达形式可以由 M^i 中每个顶点所对应的 Laplacian 坐标以及 $\mathcal{Q}^{i+1}$ 近似表示, 即 M^{i+1} 的微分方程形式. 通过求解该微分方程就可以求解 M^{i+1} 中那些未知的顶点. 本文采用对偶 Laplacian 坐标 [33] 在原始模型的对偶空间中编码模型的局部微分信息以防止局部几何在从第 i 帧迁移到第 $i+1$ 帧的过程中发生法向漂移, 从而减小模型的扭曲.

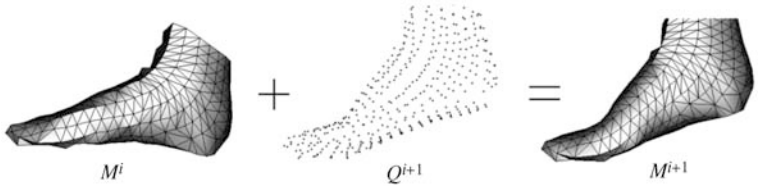

图 6　基于三维微分信息估计缺失顶点的三维几何位置

Figure 6　Estimating the invisible markers' 3D positions by 3D differential information

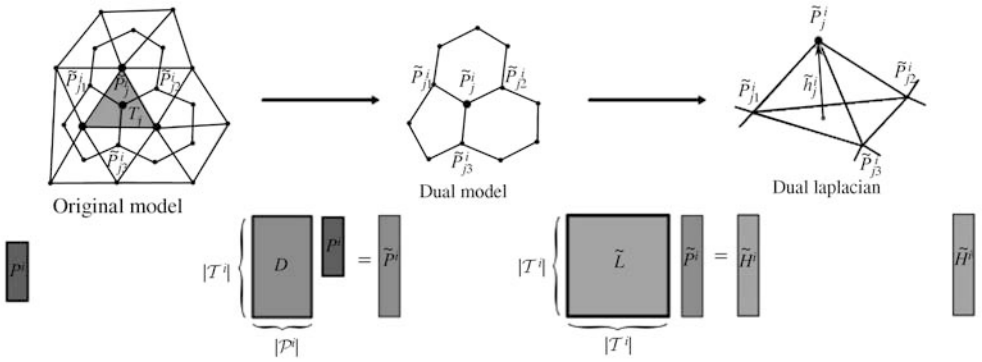

图 7　一环邻域上对偶 Laplacian 向量的构建过程

Figure 7　Illustration of the construction of dual Laplacian coordinates on the one-ring structure

首先将 M^i 转化为对偶模型. 取 M^i 中每一个面的重心作为对偶模型的顶点, 并根据原始模型面片间的相邻关系连接对偶顶点得到对偶模型. 该过程等价于一个线性变换, 即将对偶变换矩阵 $\boldsymbol{D}$ 左乘 M^i 的顶点坐标向量 $\boldsymbol{P}^i = \left[\boldsymbol{P}_1^i \cdots \boldsymbol{P}_j^i \cdots \boldsymbol{P}_{|\mathcal{P}^i|}^i\right]^{\mathrm{T}}$, 其中 $\boldsymbol{D}$ 是一个规模为 $|\mathcal{T}^i| * |\mathcal{P}^i|$ 的矩阵, 它的每一行对应着原始模型中的一个三角形, 每行中仅有 3 个非零元素, 分别对应着相应三角形的 3 个顶点. $\boldsymbol{D}$ 中的所有非零元素均为 1/3.

将 M^i 转换为对偶模型后, 计算对偶模型中每个顶点的 Laplacian 坐标. 每个对偶顶点的一环邻域为 3 的属性使其所在处的一环法向和切空间是唯一的, 这方便了在对偶模型中计算每个顶点的 Laplacian 坐标. 图 7 为将原始模型转换为对偶模型并构造一个对偶顶点的 Laplacian 坐标的过程示意图, 其中 $\widetilde{\boldsymbol{L}}$ 为对偶模型中的 Laplacian 算子矩阵, 它在整个形状跟踪的过程中是不变的, 本文采用了 Au 等 [33] 的方法来构造 $\widetilde{\boldsymbol{L}}$, $\widetilde{\boldsymbol{H}}^i = [\widetilde{\boldsymbol{h}}_1^i \cdots \widetilde{\boldsymbol{h}}_j^i \cdots \widetilde{\boldsymbol{h}}_{|\mathcal{T}^i|}^i]^{\mathrm{T}}$ 为对偶模型中各顶点对应的 Laplacian 坐标构成的列向量, $\widetilde{\boldsymbol{h}}_j^i$ 为 M^i 的对偶模型中第 j 个顶点的 Laplacian 坐标.

第 i 帧的对偶模型和 Laplacian 算子矩阵构造完毕后, 优化 (4) 式求解未跟踪到的模型顶点集 $\mathcal{P}_{\text{invisible}}^{i+1}$

$$\arg\min_{\boldsymbol{P}^{i+1}} \left|\left|\widetilde{\boldsymbol{L}}\boldsymbol{D}\boldsymbol{P}^{i+1} - \widetilde{\boldsymbol{L}}\boldsymbol{D}\boldsymbol{P}^{i}\right|\right|^2. \tag{4}$$

最后, 综合所求出的 $\mathcal{P}^{i+1} = \{\mathcal{Q}^{i+1}, \mathcal{P}_{\text{invisible}}^{i+1}\}$ 和 $\mathcal{T}^{i+1} = \mathcal{T}^i$ 即可得到第 $i+1$ 帧模型 M^{i+1}. 图 8 为本文算法所获取的步态周期中不同时刻的脚型形状.

5　实验与讨论

获取结果的精确性、完整性和获取过程的鲁棒性是评价动态形状获取技术的 3 个重要方面, 针对

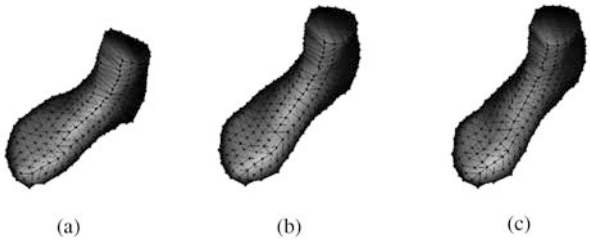

图 8 图中依次为第 1, 40, 80 帧的三维稀疏网格脚模型

Figure 8 The 3D sparse grid foot models at the 1th frame (a), at the 40th frame (b) and at the 80th frame (c)

这 3 个方面, 分别设计了两组实验对本文算法的结果模型的精度、完整性以及获取过程的鲁棒性进行了测试. 由于实际的脚型形状和所获取的模型无法直接进行对比, 第 1 组实验首先基于在标记点跟踪完备的情况下所获取的模型精度和参考模型精度相同的事实, 通过检验标记点跟踪的完备性来验证本文结果的精确性. 然后, 通过将由残缺图像获取的脚型重投影, 并与原始图像标记点对比来验证缺失部分重建结果的正确性, 最后通过与由原始图像获得的三维脚型进行对比来验证缺失点估计的精确性. 第 2 组实验通过分别检验算法在低帧率视频、复杂背景和目标被噪声干扰等条件下获取结果的正确性来测试形状跟踪算法的鲁棒性.

5.1 模型精确性与完整性

可跟踪到的标记点数量相当于各时刻对目标脚型形状的采样密度, 其对最终模型的精度有非常重要的影响. 当一个局部范围内, 缺失的模型顶点过多时, 缺失标记点的跟踪将由一个插值问题变为一个外推问题, 该区域的模型精度将难以保证. 因此标记点的数量是获得一个精确性较高模型的重要条件. 当对视频标记点的跟踪结果可以保持参考帧中的标记点数量时, 其重建精度和参考模型的重建精度是相同的, 从静态脚型获取系统 [6] 的实验结果可以得到本文参考模型的精度误差为 1 mm. 图 9 为 80 帧视频的参考帧以及视频中后段中的 2 帧图像及在这 3 帧中本文标记点跟踪的结果. 作为对比, 图 9(a),(b),(c) 显示了将本文第 3 节的基于矢量场的点集匹配方法用于迭代匹配 $\mathcal{F}^i$ 和 $\mathcal{R}^{i+1}$ 的跟踪结果. 在这类直接建立帧间对应关系的跟踪方法中, 一旦有部分点丢失了, 其对应轨迹中的后续点的前驱就不存在了, 从而这些标记点也不会被顺序跟踪到, 导致标记点的数量越来越少. 相反, 本文方法用前一帧的模型顶点和后一帧所提取出的初始标记点做对应, 由于模型顶点不会缺失, 故可以很好地保持标记点的数量. 对比图 9(d),(e),(f) 中本文方法的迭代匹配结果可以看出, 基于三维模型的匹配过程较好地保持了参考帧的标记点. 图 10 通过将所重建出的模型顶点重投影到图像中, 将投影点和原始图像标记点的位置对比来验证模型的精度. 图中的网格点为重建出的模型点的投影位置. 图 10(a),(b),(c) 为采用图 9 (a),(b),(c) 中的标记点重建的模型, 而图 10(d),(e),(f) 的结果为采用图 9(d),(e),(f) 中的标记点重建出的三维模型. 比较投影点和图像中真实标记点的位置, 随着帧数的增加, 采用结合三维模型的跟踪方法提供的脚型点的投影位置和真实标记点的位置基本一致, 其从另一角度表明在标记点不缺失的情况下可以达到和参考模型同样的精度水平, 而仅在二维图像空间进行跟踪的方法在视频的后面部分由于所提供的标记点较少, 导致其重建出的模型的投影明显偏离了真实标记点的位置.

本文方法的另一个优点是较好地解决了缺失标记点的估计问题, 保证了结果模型的完整性. 图 11 比较了本文的结果与结构光法的跟踪结果. 图 11(a) 为由 Bruce 等 [4] 提出的基于结构光的人体测量方法中动态脚型测量部分的结果. 其中某些区域由于在某些时刻是不可见的, 所以在最终的结果中, 这

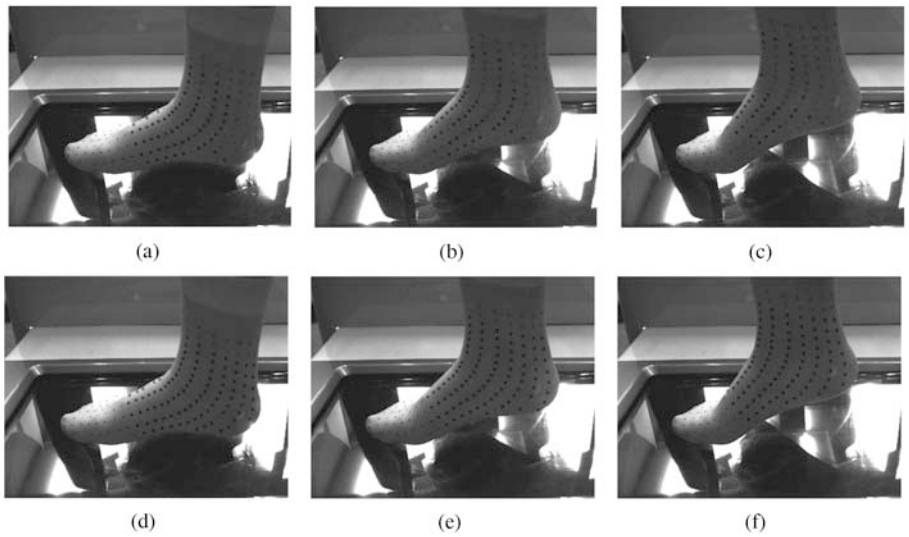

图 9　标记点跟踪结果

Figure 9　Marker point tracking results

(a),(b),(c) Results by only-2D-spatial-space-based method; (d),(e),(f) results by our method

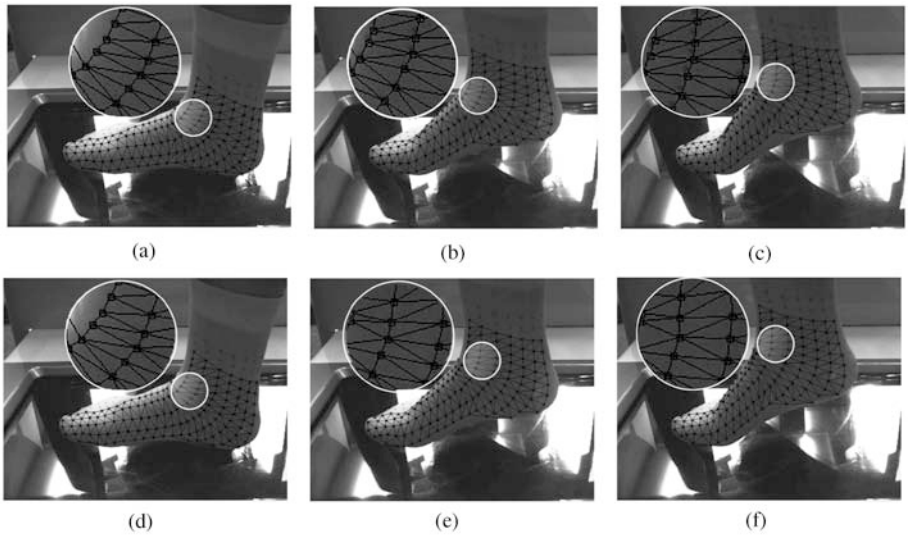

图 10　模型重投影结果

Figure 10　Projections of models

(a),(b),(c) The projected models reconstructed from the markers tracked by only-2D-spatial-space-based method; (d),(e),(f) the projected models reconstructed by our method

部分区域不能被完整地重建出来. 而我们获取的模型则较为完整. 图 11(b) 显示了由本文方法得到的经过细分处理后的脚型形状.

图 12 验证了缺失标记点重建的正确性. 图 12(a), (b), (c) 是将部分原始图像擦去后的残缺图像, 图 12(d), (e), (f) 为利用图 12(a), (b), (c) 获取的脚模型在相应图像中的投影, 从图中可以看出, 缺失

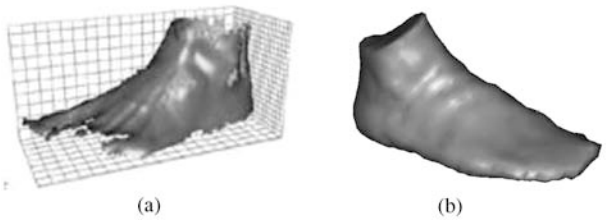

图 11 (a) Bruce 等 [4] 基于结构光的动态重建方法; (b) 本文方法所得到的结果

Figure 11 Results obtained by structure light based method of Bruce, et al. [4] (a) and by ours (b)

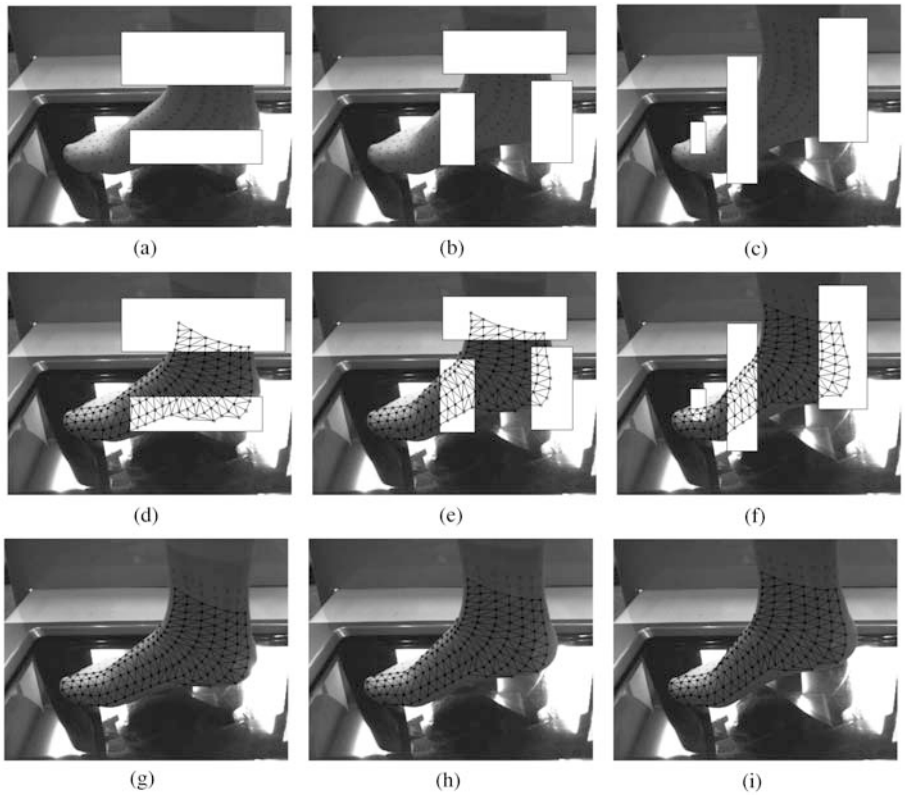

图 12 缺失标记点估计测试实验

Figure 12 Accuracy testing for the estimation quality of missing vertices

(a),(b),(c) The input images with some parts erased; (d),(e),(f) projected result model reconstructed from the corresponding erased images; (g),(h),(i) the projections of the result models to the original images to compare the projected missing vertices with the original markers

部分的脚型依然被正确地重建出来. 图 12(g), (h), (i) 为用残缺图像重建出的模型在原始图像中的投影, 对比真实标记点位置, 重建所获得的模型投影位置的偏差相对较小. 为了验证缺失标记点所对应的模型顶点的精确性, 本文对 80 帧图像的每一帧进行了类似图 12 式的随机处理, 通过将重建出的各帧脚模型与原始图像重建出的模型进行对比来验证缺失点的精度. 图 13 以统计图的形式显示了我们的实验结果. 从图中可以看出, 当缺失标记点增多时, 平均估计误差会增大, 但 80 帧中的缺失点的平均估计偏差普遍小于 1 mm.

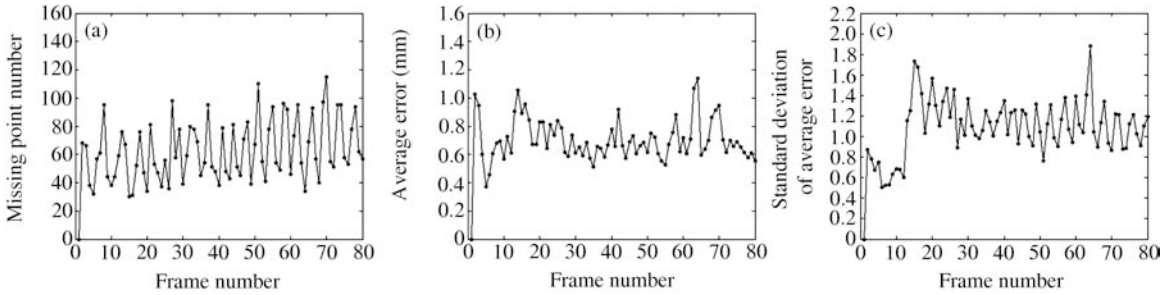

图 13　缺失点重建精度验证结果

Figure 13　The estimation error of missing vertices

(a) The quantity of missing vertices at every frame of a 80 frames video; (b) average deviation between the estimated locations and the real locations of the missing vertices; (c) the standard deviation of these errors

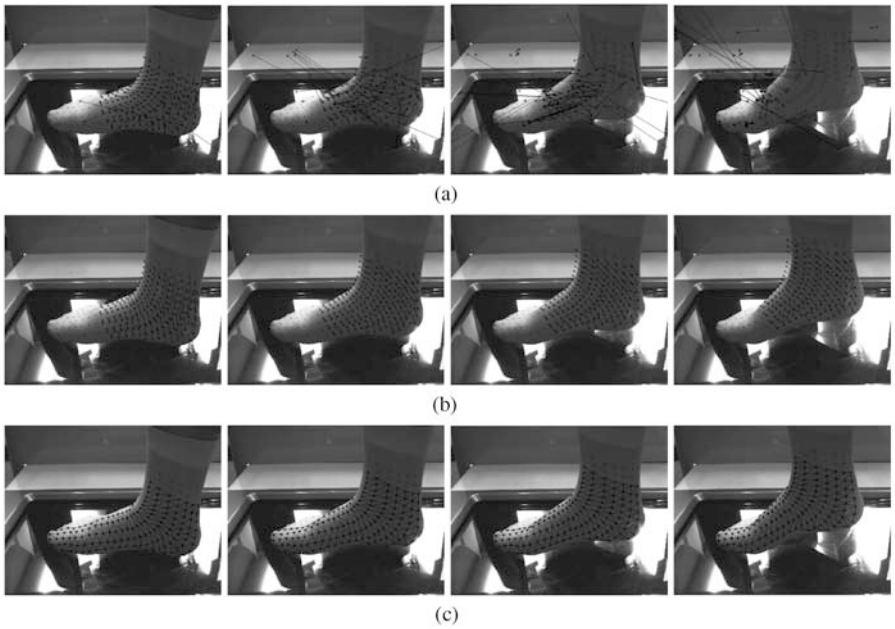

图 14　帧间目标变化幅度较大的连续帧跟踪对比结果

Figure 14　Comparative results of tracking the object with a larger range of variations between consecutive frames by optical flow algorithm (a) and ours (b); (c) projection results of 3D model obtained by our method

5.2　形状跟踪的鲁棒性

三维动态目标形状跟踪方法的鲁棒性主要由两个方面构成, 一方面是允许测量目标在帧间的运动幅度大小, 另一方面是对图像噪声的抗干扰能力. 我们首先通过加大目标在帧间的运动幅度来测试本文算法和层次化光流法的鲁棒性; 然后在抗噪声能力的测试实验中, 采用添加了人工非均匀噪声的 3 帧图像来重建相应帧模型, 通过验证测量结果的正确性来测试本文算法的抗干扰性能.

图 14 通过对本文所获取的视频帧序列每隔 15 帧进行重采样来模拟低帧率视频的情况, 其中显示了目标在相邻帧间变化幅度较大的情况下光流法和本文方法的跟踪结果, 从实验结果中可见, 对于同

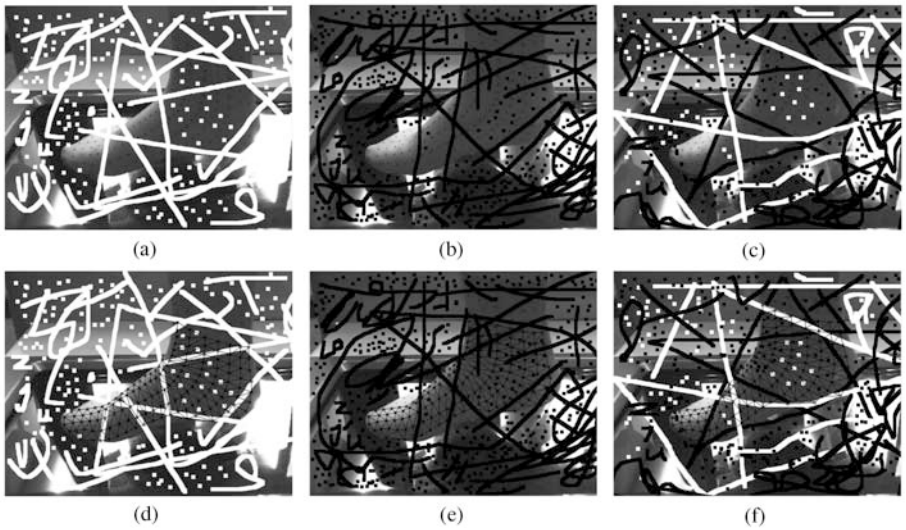

图 15 在极端噪声干扰下本文算法的脚型获取结果

Figure 15 Foot acquisition result under extreme severe conditions

(a),(b),(c) The original polluted images; (d),(e),(f) the projective result of restructured foot from the polluted images

图 16 楦脚匹配仿真穿鞋状态

Figure 16 Simulation of wear test by superpositioning the 3D foot with the candidate shoe-last

样的数据, 图 14(a) 中光流法的跟踪结果正确率很低, 而本文的方法则保持了较高的正确率 (图 14(b)). 图 14(c) 通过将本文算法重建出的三维模型投影到图像中, 验证了在目标运动幅度如此大的情况下其依然能够获得正确的重建结果.

图 15 是一组测试算法抗噪声能力的实验结果, 其中图 15(a),(b),(c) 为 3 幅添加了人工噪声的原始图像, 图 15(d),(e),(f) 显示了利用图 15(a),(b),(c) 重建出的结果在图像中的投影. 从图中可以看出, 在这样极其混乱的图像条件下, 本文的算法依然能够重建出正确的脚型形状.

6 结论

本文提出了一种从视频中获取动态脚型形状的新方法, 它可以在开放环境中获取运动着的脚型在步态周期中离开地面前任意时刻的形状, 并且已经被用于鞋类产品个性化定制的舒适度评测中 (如图 16). 精度验证实验的结果说明模型形状的精度在其他条件相同的情况下取决于标记点的数量, 本文通过迭代匹配指导模型顶点与所提取到的标记点的跟踪方式, 使标记点的数量得到了较好的保持, 提高了测量结果的精度. 将缺失标记点的估计问题在三维空间中求解, 避免了标记点缺失前和重新出现后的轨迹对应问题以及缺失时段标记点运动速度的估计问题, 改善了缺失标记点估计的质量, 既增强了

后续跟踪过程的鲁棒性, 也保证了形状的完整性. 通过构建三维模型顶点运动投影的速度矢量场来指导建立帧间标记点的对应关系, 并将标记点集合作为整体进行跟踪的方法, 有效地避免了对特征纹理的过多依赖, 从而在复杂背景下, 可以较好地跟踪邻域纹理和几何相近的特征. 即使在相邻帧间目标运动幅度较大以及噪声遮挡等干扰严重的条件下, 也能鲁棒地恢复目标的三维形状. 整个方法除了在相当极端的情况下需要少量交互外, 可以自动提取、跟踪标记点并完成目标建模, 避免了传统方法中繁琐的交互修正工作.

实验结果表明, 文中的方法在用于跟踪运动过程中的脚型形状时获得了较好的精确性与完整性, 同时方法本身具有良好的鲁棒性和可靠性, 但是其对更加复杂的变形目标的跟踪问题有待进一步研究. 虽然通过袜子对目标进行标记的方法和标记点的数量使本文技术存在一定的局限性, 但其在类似本文的问题中可以较好地满足应用的要求. 基于本文的技术, 潜在的未来工作还包括解决当目标在变形过程中经历面向相机的正反面完全互换的跟踪问题, 以及目标在变形及移动过程中交替跨越多个相机视野范围的多相机协同跟踪问题等.

致谢 感谢王彬等同学在实验数据测试、文章校对方面的辛勤工作.

参考文献

1 Makiko K, Masaaki M. Development of a low cost foot-scanner for a custom shoe making system. In: Proceedings of the 5th Symposium on Footwear Biomechanics. Paris: Footwear Biomechanics Group, 2001. 58–59
2 Luximon A, Goonetilleke R S, Zhang M. 3D foot shape generation from 2D information. Ergonomics, 2005, 48: 625–641
3 Witana C P, Xiong S P, Zhao J H, et al. Foot measurements from three-dimensional scans: a comparison and evaluation of different methods. Int J Industr Ergonom, 2006, 36: 789–807
4 MacWilliams B A, Cowley M, Nicholson D E. Foot kinematics and kinetics during adolescent gait. Gait Post, 2003, 17: 214–224
5 Boehnen C, Flynn P J. Accuracy of 3D scanning technologies in a face scanning context. In: Fifth International Conference on 3D Imaging and Modeling(3DIM2005). Washington: IEEE Computer Society Press, 2005. 310–317
6 Gao F, Chu J, Li M J, et al. 3D Scanning of foot by stereo vision based on explicit markers. J Comput Aid Des Comput Graph, 2009, 21: 1412–1419
7 Kutulakos K N, Seitz S M. A theory of shape by space carving. Int J Comput Vis, 2000, 38: 197–216
8 Culbertson W, Malzbender T, Slabaugh G. Vision Algorithms: Theory and Practice. 1st ed. Berlin: Springer, 1999. 100–115
9 Matusik W, Buehler C, McMillan L. Polyhedral visual hulls for real-time rendering. In: Proceedings of the 12th Eurographics Workshop on Rendering Techniques. London: Springer-Verlag, 2001. 116–126
10 Moezzi S, Tai L C, Gerard P. Virtual view generation for 3D digital video. IEEE MultiMed, 1997, 4: 18–26
11 Laurentini A. The visual hull concept for silhouette-based image understanding. IEEE Trans PAMI, 1994, 16: 150–162
12 Decarlo D, Metaxas D. The integration of optical flow and deformable models with applications to human face shape and motion estimation. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Washington: IEEE Computer Society Press, 1996. 231–238
13 Pighin F H, Szeliski R, Salesin D. Resynthesizing facial animation through 3D model-based tracking. In: Proceedings of International Conference on Computer Vision (ICCV). Washington: IEEE Computer Society Press, 1999. 143–150
14 Blanz V, Basso C, Poggio T, et al. Reanimating faces in images and video. Comput Graph Forum, 2003, 22: 641–650
15 Vlasic D, Brand M, Pfister H, et al. Face transfer with multilinear models. ACM Trans Graph, 2005,24: 426–433
16 De Aguiar E, Theobalt C, Stoll C, et al. Marker-less deformable mesh tracking for human shape and motion capture. In:

Proceedings of IEEE Conference on Computer Vision and Pattern Recognition(CVPR). Washington: IEEE Computer Society Press, 2007. 1–8
17 De Aguiar E, Stoll C, Theobalt C, et al. Performance capture from sparse multi-view video. ACM Trans Graphics (TOG), 2008, 27: 98
18 Pighin F, Hecker J, Lischinski D, et al. Synthesizing realistic facial expressions from photographs. In: Proceedings of ACM SIGGRAPH1998. New York: Association for Computing Machinery SIGGRAPH, 1998. 75–84
19 Herda L, Fua P, Plankers R, et al. Skeleton-based motion capture for robust reconstruction of human motion. In: Proceedings of Computer Animation 2000(CA'00). Washington: IEEE Computer Society Press, 2000. 77–93
20 Moeslund T B, Granum E. A survey of computer vision-based human motion capture. Computer Vision and Image Understanding (CVIU), 2001,81(3):231–268
21 Park S I, Hodgins J K. Capturing and animating skin deformation in human motion. ACM Trans Graphics (TOG), 2006, 25: 881–889
22 Zhang Z. A flexible new technique for camera calibration. IEEE Trans Patt Anal Mach Intell, 2000, 22: 1330–1334
23 Battiato S, Gallo G, Puglisi G, et al. SIFT features tracking for video stabilization. In: Proceedings of 14th International Conference on Image Analysis and Processing. Berlin: Springer, 2007. 825–830
24 Wagner D, Reitmayr G, Mulloni A, et al. Pose tracking from natural features on mobile phones. In: Proceedings of the 7th IEEE/ACM International Symposium on Mixed and Augmented Reality. Washington: IEEE Computer Society Press, 2008. 125–134
25 David G L. Distinctive image features from scale-invariant keypoints. Int J Comput Vis, 2004, 60: 91–110
26 Tsutsui H, Miura J, Shirai Y. Optical flow-based person tracking by multiple cameras. In: Proceedings of International Conference on Multisensor Fusion and Integration for Intelligent Systems. Berlin: Springer, 2001. 91–96
27 Brox T, Rosenhahn B, Cremers D, et al. High accuracy optical flow serves 3-D pose tracking: exploiting contour and flow based constraints. In: Proceedings of European Conference on Computer Vision, ECCV'06(volume 3952 of LNCS). Berlin: Springer, 2006. 98–111
28 Lucas B, Kanade T. An iterative image registration technique with an application to stereo vision. In: Proceedings of the International Joint Conference on Artificial Intelligence. San Francisco: Morgan Kaufmann, 1981. 674–679
29 Bergen J R, Anandan P, Hanna K J, et al. Hierarchical model-based motion estimation. In: Poceedings of the European Conference on Computer Vision(volume LNCS 588). Berlin: Springer, 1992. 237–252
30 Harris C, Stephens M. A combined corner and edge detector. In: Proceedings of the Fourth Alvey Vision Conference. Manchester: Organising Committee AVC 88, 1988. 147–151
31 Smith S M, Brady J M. SUSAN-A new approach to low level image processing. Int J Comput Vis, 1997, 23: 45–78
32 Parker J R, Jennings C, Salkauskas A G. Thresholding using an illumination model. In: Proceedings of the 2nd International Conference on Document Analysis and Recognition. Washington: IEEE Computer Society Press, 1993. 270–273
33 Au O K C, Tai C L, Liu L G, et al. Dual Laplacian editing for meshes. IEEE Trans Visual Comput Graph, 2006, 12: 386–395

Acquisition of time-varying 3D foot shapes from video

Gao Fei, Wang Qing, Geng Weidong* & Pan Yunhe

State Key Lab of CAD&CG, Zhejiang University Hangzhou 310027*; China*
*E-mail: gengwd@zju.edu.cn

Abstract The 3D scanning of time-varying objects based on multi-view geometry has been a hot topic in

computer vision in recent years. This paper presents an active approach to acquiring 3D shapes of moving foot from multi-view video sequences based on the explicit marker points woven in socks. The reference 3D model of foot is firstly recovered from the starting frames in the multi-view video clips, and then the markers in each view are traced as a whole based on a continuous motion vector field built from the reference 3D model. The missing vertices in the candidate 3D model caused by self-occlusion, non-uniform illumination and random noises are reliably estimated and reconstructed in 3D space by minimizing the changes of differential features in 3D geometry of the reference model. The experimental results show that our method can robustly recover the 3D model of foot in complex background even if there is a relatively large movement between the adjacent frames, with an acceptable accuracy of the resulting 3D model.

Keywords Time-varying shapes, 3D Foot modeling, Feature tracking, Multi-view geometry