# 基于多个关键点对应性的手部动态重建

范 清[1]， 沈旭昆[1,2]

(1. 北京航空航天大学虚拟现实技术与系统国家重点实验室，北京 100191；
2. 北京航空航天大学新媒体艺术与设计学院，北京 100191)

**摘 要**：提出了一种新的基于单视图深度序列的手部运动跟踪和表面重建方法。在假设任意一对关键点的对应性在所有手部姿态上均一致基础上，使用一个平滑的手部网格模板来提供形状和拓扑先验，引入多个能量函数构造输入扫描与模板之间三维关键点到关键点的对应性，并将其整合到一个可用的非刚性配准管线中，以实现精确的表面拟合。通过最小化手部模板和输入深度图像序列之间的误差来捕获非刚性的手部运动。采用迭代求解的方法，通过显式的关键点到关键点之间的对应性，逐步细化手部关节区域的变形，从而达到快速收敛和合理变形的目的。在含有噪声的手部深度图像序列上的大量实验表明，该方法能够重建精确的手部运动，并且对较大的变形和遮挡具有鲁棒性。

**关 键 词**：运动跟踪；表面重建；关键点；非刚性变形；深度序列
**中图分类号**：TP 391 **DOI**：10.11996/JG.j.2095-302X.2020050709
**文献标识码**：A 文 章 编 号：2095-302X(2020)05-0709-07

# Hand dynamic 3D reconstruction using multiple keypoint-to-keypoint correspondences

FAN Qing[1], SHEN Xu-kun[1,2]

(1. State Key Laboratory of Virtual Reality Technology and Systems, Beihang University, Beijing 100191, China;
2. School of New Media Art and Design, Beihang University, Beijing 100191, China)

**Abstract:** We propose a new approach to robustly reconstruct non-rigid hand geometries and motions from single-view depth sequences. On the basis of the assumption that the correspondence between any pair of key points is consistent across all hand poses, a smooth hand mesh template is adopted to provide the shape and topology prior, and several energy functions are introduced to formulate these 3D keypoint-to-keypoint correspondences, which are then incorporated into the available non-rigid registration pipeline to achieve an accurate surface fit. The non-rigid hand motion is captured by minimizing the distortion error between a hand template and the partial input. An iterative solver is employed to gradually refine deformations around joint regions with explicit correspondences, leading to fast convergence and plausible deformations. Extensive experiments on hand depth sequences with noise demonstrate that our approach is capable of reconstructing accurate hand motions, and is robust for large deformations and occlusions.

**收稿日期**：2020-05-09；**定稿日期**：2020-05-22
**Received**：9 May，2020；**Finalized**：22 May，2020
**第一作者**：范 清(1986–)，男，河南信阳人，博士研究生。主要研究方向为计算机视觉、计算机图形学和虚拟现实。E-mail：fanqing@buaa.edu.cn
**First author**：FAN Qing (1986–), male, PhD candidate. His main research interests cover computer vision, computer graphics and virtual reality. E-mail：fanqing@buaa.edu.cn
**通信作者**：沈旭昆(1965–)，男，云南昆明人，教授，博士。主要研究方向为计算机图形学和虚拟现实等。E-mail：xkshen@buaa.edu.cn
**Corresponding author**：SHEN Xu-kun (1965–), male, professor, Ph.D. His main research interests cover computer graphics and virtual reality, Etc. E-mail：xkshen@buaa.edu.cn

**Keywords:** motion tracking; surface reconstruction; key-point; non-rigid deformation; depth sequences

人体的动态重建一直是计算机视觉和计算机图形学领域的研究热点。最近的工作在人体重建方面取得了很好的效果[1]。手作为人体最具特色的器官之一，由于其频繁的自遮挡、丰富的姿态和细节、复杂的形状变化，运动人手的动态重建始终具有挑战性，众多研究人员持续开展研究。

目前已经提出很多基于非刚性变形技术的动态重建方法[2-3]，这些方法通常利用一个模板模型来简化对应性查找和为单视图表面重建提供先验知识，然而，其 3D 对应性建立方式存在低效，阻碍用户运动等问题，其需要在用户身上粘贴光学标记，或依赖用户手工干预，或需要额外强加非平滑的运动约束[4]。鉴于此，本文采用实时的手跟踪系统来推断出准确的 3D 手部关节点位置，进而为手部表面动态重建提供鲁棒的标记信息。

本工作旨在从消费者级别深度传感器捕获的单目深度序列中动态重建出完整的手部表面。为此，借助一个平滑的手部模板来重构完整的手部运动，并提供形状和拓扑先验(下文将这个模板网格定义为源，而输入深度图像序列定义为目标)。我们发现从鲁棒的手跟踪系统中推断出的 3D 手部关节位置可以作为非刚性曲面拟合的重要线索，为了利用这个线索，本文假设任意一对关键点之间的对应性在不同的手部姿态下保持不变，通过引入多个关键点对到非刚性曲面拟合管线中以帮助获得更合理的手部形变。此外，本文还引入多个能量函数来度量关键点区域的对齐和非刚性曲面拟合的程度，通过最小化源模型和目标模型之间的失真误差来恢复准确的非刚性手部运动。

本文提出了一种鲁棒的基于单目深度序列的非刚性手部动态重建方法，其利用非线性可变形模板来捕捉手部的大尺度变形，并结合关键点对应性约束自动将模板与每个输入深度帧对齐，整个管线不需要用户任何手工干预。此外，还提出了一种多帧稳定策略，每隔一定时间将当前帧的跟踪结果与锚点帧的跟踪结果相融合，以减少跟踪过程中的误差积累。在含有噪声的输入深度序列上进行的大量实验表明，本文方法能够生成准确的动态重建结果，且对于较大的变形和遮挡具有鲁棒性。本文所用模板及部分重建结果如图 1 所示。


**图 1** 手部模板及部分重建结果((a)手部模板及采样节点；(b)和(c)左列为输入，右列为重建网格在不同视点下的渲染结果)

**Fig. 1** Template and partial reconstruction results of our method ((a) Initial template with nodes; (b) ~ (c) The input depth (left) and the reconstructed geometry model (right))

## 1 相关工作

复杂可变形物体的非刚性几何和运动重建是计算机图形学一个长期研究的问题，已有大量文献发表[5]。

(1) 多视图与单视图。目前已经提出了许多多视图非刚性几何和运动重建方法，如由轮廓恢复形状[6]和被动立体匹配方法[7]。文献[8]结合表面模型和体模型对穿着普通衣服的人群进行高质量的重建。研究工作[9]提出将多个非连续的图像序列与复杂的人体和服装运动对齐。文献[10]使用 3 个手持式 Kinect 传感器同时重建人体骨骼姿态、人体表面几何和相机姿态。DOU 等[11]使用围绕对象一圈放置的 8 个 RGBD 相机来重建一个完整的目标模板，用来匹配输入序列以捕捉精确的表面形变。其后续工作[12]通过自动回路检测将对齐误差分布在回路上，以减少跟踪过程中的误差累积和应对非刚性对齐输入序列时产生的漂移问题。紧接着又提出一个多阶段融合方法[13]从多个 RGBD 摄像机实时记录的噪声输入中快速重建具有挑战性的场景。文献[14]引入一个可变形形状模板作为扫描对象的几何和拓扑先验来更好地匹配多视图输入序列。

与多视图方法相比，基于单视图的方法使用更轻便的设备来捕获移动对象。对于关节运动，文献[15]通过在最大后验框架中迭代注册一个已绑定骨骼的人体模型到单视图的深度序列来重建三维人体姿态。文献[16]结合深度、轮廓和时间线索，通过不断拟合三维关节手模型到输入深度图像来重建关节手运动。文献[17]使用骨架对齐来替代模型拟合从而更有效地从单目深度序列中回归关节手运动。然而，这些方法通常只关注人体或手部的

骨架运动的重建，而忽略了表面的非刚性形变。对于手部的非刚性运动和几何重建，非刚性配准技术[18]提供了一个潜在的解决方案。

(2) 标记点与无标记。基于标记点的方法由于具有较高的准确性在早期运动捕获系统中得到广泛应用，近年来，无标记运动跟踪方法由于无需任何其他辅助设备，而更便捷，也因此获得了更多的关注。文献[19]提出了一种从深度扫描序列中重构半刚性对象的方法，但不能处理动态物体。3D 自画像[20]允许普通用户捕获自己的三维模型，但要求用户在扫描过程中保持静止。文献[21]和文献[22]提出在不使用模板的情况下动态重建非刚性的场景，然而，均不能处理快速运动导致的漂移问题。

无标记方法通常通过各种方法来推断标记点，如姿态估计方法，并隐式地使用预测到的标记点作为伪真值来增强鲁棒性。随着姿态估计技术研究的不断深入，为无标记的运动捕获系统提供可信的标记点成为可能。在之前工作的基础上[17]，本文提出应用手部姿态估计技术来推断出三维手部关节位置作为标记点，并利用每个骨骼的局部坐标将关节投影到每个手部输入扫描的表面，从而在手部形状模板和输入扫描之间建立多个三维关键点到关键点的对应。

## 2　跟踪管线

为了鲁棒的从消费者级别深度传感器捕获的单目深度序列中动态重建出完整的手部表面，本文提出了一个新的跟踪重建管线，以深度序列和一个手部形状模板作为输入，经过处理流程，对每个输入深度图像输出一个重构的完整手部网格。管线如图 2 所示，主要由 4 个组件构成：关键点到关键点对应性、表面形变模型、非刚性曲面拟合和多帧稳定性。

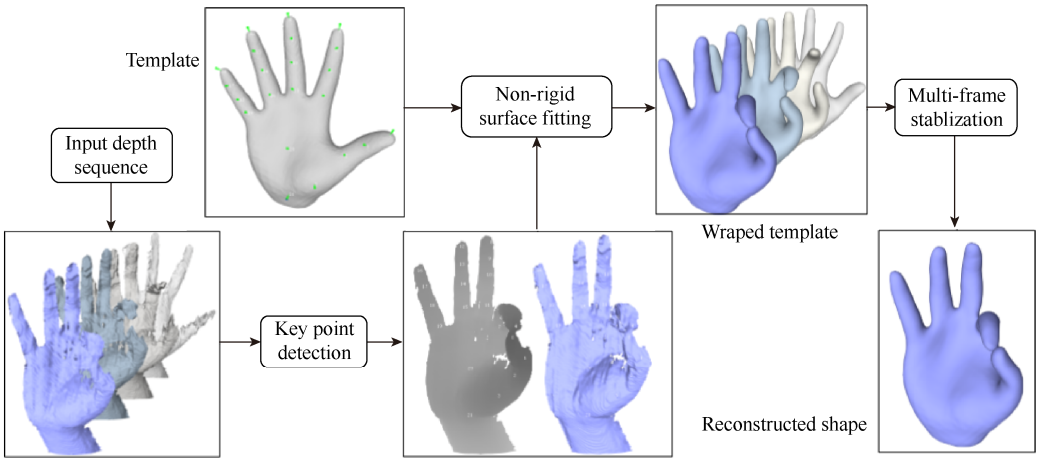


**图 2**　系统流程图(通过多个关键点到关键点的对应注册一个平滑的模板网格到每个输入帧。基于一个可用的手跟踪系统来预测关键点位置。通过多帧稳定性组件减少误差积累以获得鲁棒的重构结果)

**Fig. 2**　Pipeline overview (A smooth template mesh is registered to each of the input depths with multiple keypoint-to-keypoint correspondences. Key point detection is based on an available hand pose estimation baseline. Robust reconstruction is obtained through a multi-frame stabilization component to reduce error accumulation)

### 2.1　关键点到关键点对应性

在手部的非刚性运动过程中，相比于其他区域，手部关节区域的变形更加强烈。关键点到关键点对应性是建立输入深度序列和手部形状模板关节点之间的对应关系，以约束手部关节区域的局部变形，从而提升非刚性运动跟踪的准确性和鲁棒性。关键点到关键点对应性建立包含几个步骤：首先，应用一个可用的手部姿态估计系统来为深度图像序列 $S=\{S^0, S^1, \cdots, S^{n-1}\}$中的每一帧推断出 $J$ 个三维的手部关节位置，并粗略地将初始模板 $T_{\text{init}}$ 与序列首个深度帧 $S^0$ 对齐。然后，沿着每根骨骼的局部坐标的 $Z$ 轴将 3D 手部关节位置投影到输入深度扫描的表面上，并将与交点最接近的顶点设为关键点，如图 3 所示。在自遮挡情况下，将关键点的三维位置设为估计到的 3D 位置沿其当前骨骼局部坐标的 $Z$ 轴减去一个经过数据分析和测试得出的经验值 0.5 cm。手部形状模板上对应的 $J$ 个关键点位置是根据语义对应性手工选取的。最后，这 $J$ 个关键点对作为显式约束被整合到非刚性曲面拟合管线中以强制模板变形到更加准确的位置。本文实验中都设定为 $J$=22。

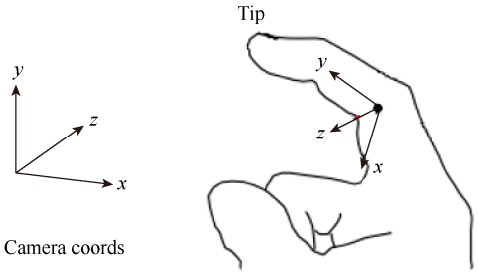


**图 3** 关键点定义示意图

**Fig. 3** Illustration of key-points definition

## 2.2 表面形变模型

本文表面形变模型基于一个扩展的嵌入式变形框架，并通过使用一个变形图来计算变形场以离散化底层空间。该模型自动适应捕获对象的运动，其超越了基于骨架的线性混合蒙皮的表达能力，可以进一步重建复杂的表面形变。依据当前的研究[4]，网格 $M$ 的变形由一系列离散的图节点的仿射变换来表示。每个节点 $x_i$ 诱导一个半径为 $r_i$ 的局部影响区域的变形，其由一个 3×3 的矩阵 $\boldsymbol{A}_i$ 和一个 3×1 的平移向量 $\boldsymbol{t}_i$ 表示。当 2 个图节点影响网格的同一个顶点时，用 1 条边连接这 2 个节点。变形后，网格的第 $j_{th}$ 个顶点 $v_j$ 将被映射到新位置 $v'_j$。

$$\begin{cases} v'_j = \sum_{x_i} \omega(v_j, x_i, r_i)[\boldsymbol{A}_i(v_j - x_i) + x_i + \boldsymbol{t}_i] \\ \omega(v_j, x_i, r_i) = \max(0, (1 - d^2(v_j, x_i)/r_i^2)^3) \end{cases} \tag{1}$$

其中，$d(v_j,x_i)$为 $v_j$ 和 $x_i$ 之间的测地线距离。更详细的关于如何在网格上提取图节点 $x_i$ 以及如何为所有顶点计算 $\omega$，请参阅文献[2]。本文使用的手部模板和抽取的图节点如图 1(a)所示。

## 2.3 非刚性曲面拟合

非刚性曲面拟合通过连续匹配一个手部形状模板到输入序列的每一帧来捕获大尺度的手部运动和非刚性形变。在时刻 $t$，给定深度帧 $S^t$ 和一个已与上一帧对齐的已变形模板网格 $W^{t-1}$，非刚性曲面拟合组件在模板与跟踪对象之间关键点到关键点对应性的指导下变形 $W^{t-1}$ 以拟合 $S^t$。后续为了简洁起见，时间戳 $t$ 将被省略。本组件通过最小化一个总能量 $E_{tol}$ 来求解精确的曲面拟合结果，即

$$E_{tol} = E_{\text{point}} + E_{\text{plane}} + E_{\text{keypoint}} + E_{\text{weight}} + E_{smo} + E_{\text{rigid}} \tag{2}$$

### 2.3.1 拟合能量

拟合能量惩罚手部模板与观测数据之间的偏差，其中包括点到点、点到面和关键点到关键点的差异，以避免错误的对应，特别是对于较大的特征不明显或遮挡的手部区域。$E_{\text{point}}$ 为点到点的对应，其定义为

$$E_{\text{point}} = \lambda_{\text{point}} \sum_{v_j \in C} \eta_i w_j^2 \left\| v'_j - c_j \right\|_2^2 \Big/ \sigma_d \tag{3}$$

其中，$\lambda_{\text{point}}$ 为权重因子；$\sigma_d$ 为传感器噪声；$C$ 为包含在输入扫描中有对应最近点的所有顶点的集合；$w$ 为每个最近点对的信心；$v_j$ 为可见性用一个二进制变量 $\eta_i$ 来表示，其可以通过在当前相机视角和参数下渲染形状模板获取。

为了使最近点对计算的更加准确，本文还在能量函数中采用了一个点到面的拟合项，其定义为

$$E_{\text{plane}} = \lambda_{\text{plane}} \sum_{v_j \in C} \eta_i \, w_j^2 \left| n_{c_j}^{\mathrm{T}} \left( v'_j - c_j \right) \right|^2 \Big/ \sigma_n \tag{4}$$

其中，$\lambda_{\text{plane}}$ 为权重因子；$\sigma_n$ 为噪声估计。上述 2 个数据拟合项结合在一起，迫使形状模板中的每个顶点 $v_j$ 沿其法向 $n_{c_j}$ 拟合到对应深度扫描上的最近点 $c_j$。在本文实验中，$\sigma_d$ 和 $\sigma_n$ 被固定为测量值 1，并且所有的信心权重均被经验性地初始化为 1。

$E_{\text{keypoint}}$ 表示关键点到关键点的对应，其定义为

$$E_{\text{keypoint}} = \lambda_{\text{keypoint}} \sum_{k=1}^{K} w_k^2 \left\| v'_k - c_k \right\|_2^2 \tag{5}$$

其中，$\lambda_{\text{keypoint}}$ 和 $K$ 分别为权重因子和关键点对的数量；$E_{\text{keypoint}}$ 项在本节的表面匹配组件中施加了一个硬约束，以提高频繁被遮挡关节区域局部变形的准确性。值得注意的是此项中没有考虑关键点的可见性，因为本文使用的关键点检测系统即使在遮挡情况下也可以产生准确的关键点位置。

### 2.3.2 正则化能量

本文引入一个信心权重正则化项 $E_{\text{weight}}$，其定义为

$$E_{\text{weight}} = \lambda_{\text{weight}} \sum_{v_j \in C} \eta_i \left( 1 - w_j^2 \right) \tag{6}$$

最小化该项相当于最大限度地增加可靠对应的数量，并抑制不可靠的对应。$w$ 的值接近于 1 表示可靠的对应关系，而接近于 0 表示未找到合适的对应关系。这一项与 LI 等[2]的类似，但该方法未考虑对应点的可见性。

$E_{smo}$ 描述了一个平滑变形约束，其强制一个节点的仿射变换与其邻域尽可能相似，定义为

$$E_{smo} = \lambda_{smo} \sum_{x_i} \sum_{x_j \in N(x_i)} \omega(x_i, x_j) \Big\| A_i(x_j - x_i) + x_i + t_i - (x_j + t_j) \Big\|_2^2 \tag{7}$$

一个特征保持变形场 $E_{\text{rigid}}$ 限制每个节点的仿射变换尽可能的刚性，其定义为

$$E_{\text{rigid}}=\lambda_{\text{rigid}}\sum_{x_i}\left(\left(a_1^{\mathrm{T}}a_2\right)^2+\left(a_1^{\mathrm{T}}a_3\right)^2+\left(a_2^{\mathrm{T}}a_3\right)^2+\left(1-a_1^{\mathrm{T}}a_1\right)^2+\left(1-a_2^{\mathrm{T}}a_2\right)^2+\left(1-a_3^{\mathrm{T}}a_3\right)^2\right) \quad (8)$$

其中，$a_1$，$a_3$，$a_3$ 分别为矩阵 $\boldsymbol{A}_i$ 的第 1~3 列。该项强制矩阵 $\boldsymbol{A}_i$ 尽可能是一个正交阵。

总能量 $E_{tol}$ 的最优化通过使用一个高斯-牛顿法迭代求解一个非线性最小二乘问题实现。实验参数设置见表 1。

**表 1　权重参数设置**

**Table 1　Initial values of weight parameters**

| 参数 | $\lambda_{\text{point}}$ | $\lambda_{\text{plane}}$ | $\lambda_{\text{keypoint}}$ | $\lambda_{\text{weight}}$ | $\lambda_{smo}$ | $\lambda_{\text{rigid}}$ |
|---|---|---|---|---|---|---|
| 数值 | 0.1 | 1.0 | 1 000.0 | 1.0 | 1 000.0 | 1 000.0 |

### 2.4　多帧稳定策略

多帧稳定性保证可靠跟踪对象被扫描到的表面区域，并通过 2 个已变形模板形状的线性混合来减少跟踪对象被遮挡部分的误差累积。在时刻 $t$–1，已与上一帧对齐的已变形的模板网格 $W^{t-1}$ 处于零能量状态，在时刻 $t$，跟踪管线强制将 $W^{t-1}$ 对齐到当前输入 $S^t$。在运动跟踪过程中，可见手部区域的运动由于包含丰富的观测信息可以被精确地捕获。然而，当连续地将已变形模板 $W$ 与来自单个深度传感器的包含噪声与遮挡数据进行对齐时，对齐误差累积的很快，最终会产生严重的漂移问题。为此，本文引入了一个简单但有效的多帧稳定策略来精确跟踪可见的表面区域运动，并减少跟踪过程中被遮挡区域的误差累积。

在跟踪过程中，每 30 帧选择出一帧作为锚点帧 $S^{an}(an=30t,\ t=1,2,\cdots,n)$。基于时间相关性，通过变形模板 $W^{an-1}$ 来匹配 $S^{an}$ 得到一个对齐后的模板 $W^{an}$，其顶点表示为 $v^{an}$。当不采用时间相关性，通过变形 $T_{\text{init}}$ 来匹配当前帧，也可以得一个对齐后的模板 $\bar{W}^{an}$，其顶点表示为 $\bar{v}^{an}$。然后基于顶点可见性，为每个顶点定义一个硬的位置约束，通过线性混合 $W^{an}$ 和 $\bar{W}^{an}$ 生成一个最终修正后的网格 $R^{an}$，其顶点定义为

$$v^r=\eta^{an}v^{an}+(1-\eta^{an})\bar{v}^{an} \quad (9)$$

## 3　实验结果

本文提出方法在 2 个挑战的手部深度序列上做出评价。第一个是文献[4]中所提供的一个手部深度序列(序列 1)，其包含使用 Intel IVCam 摄像机记录的 400 帧深度图像和一个具有约 9 000 个顶点的可变形手部形状模板。另一个序列(序列 2)包含从 NYU 手部姿态数据集[23]抽取的 500 帧深度图像序列。这 2 个序列均包含快速的手部运动与严重遮挡案例，更多信息见表 2。在粗略对齐手部形状模板与第一个深度帧后，在这些序列上，本文的表面运动跟踪管线以每分钟 20 帧的速度运行，采用 C++ 实现，运行环境为具有 3.20 GHz，4 核 CPU 和 8 G 内存的个人计算机。

**表 2　实验所用测试序列**

**Table 2　Statistics of two test sequences in the experiments**

| 序列 | 帧数 | 锚点数 | 顶点数 | 存在遮挡 | 捕获设备 |
|---|---|---|---|---|---|
| 1 | 400 | 15 | 44~72 k | 是 | IVCam |
| 2 | 500 | 18 | 9~15 k | 是 | Kinect |

为了定量分析实验结果，将这些序列输入到一个现有的 Vicon 光学运动捕获系统，并手动配准系统标记点与手部模板网格上的对应顶点，以此生成准确的标注数据。图 4 给出了本文方法、文献[12]和文献[4]在一个非遮挡输入案例上的误差分布对比结果。其中使用每个顶点与其真实标注值之间的欧式距离来编码单个顶点的颜色。从图 4 可看出，在非遮挡情况下本文方法重建误差分布稳定且无异常值。

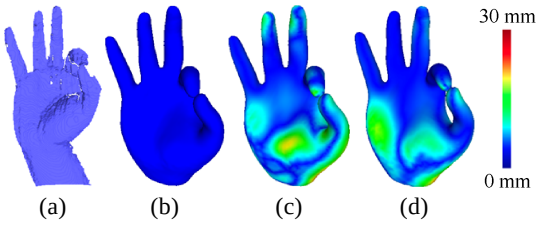


**图 4**　在序列 1 上的误差分布对比((a)为输入深度帧；(b)~(d)分别为本文、文献[12]和文献[4]的误差分布结果)

**Fig. 4**　Comparison of error distribution on sequence 1st ((a) Presents the input depth; (b)~(d) Present maps of error distributions of our method, [12] and [4], respectively the color-coded per-vertex error is the Euclidean distance between a vertex and its ground truth position)

图 5 给出了本文方法、文献[12]和文献[4]在另一个存在遮挡的输入案列上的误差分布对比结果。从图 5 可看出，在被扫描对象部分被遮挡情况下，文献[12]和文献[4]方法在虎口处均出现异常值，而本文方法依然能得到准确的重建结果。

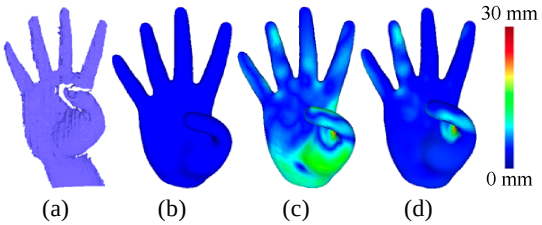


**图 5** 在一个遮挡案列上的误差分布对比((a)为输入深度帧；(b)~(d)分别为本文、文献[12]和文献[4]的误差分布结果

**Fig. 5** Comparison of error distribution on a case with occlusion ((a) Input depth; (b)~(d) Maps of error distributions of our method, literature [12] and [4], respectively. The color-coded per-vertex error is the Euclidean distance between a vertex and its ground truth position)

图 6 为本文、文献[12]和文献[4]等 3 种方法在序列 1 上单帧平均误差定量对比结果。单帧平均误差定义为每一个重建的网格与其真实标注之间所有顶点对欧氏距离的平均值。从图 6 可明显看出，红色、绿色和蓝色曲线分别代表本文、文献[4]和文献[12]的平均误差结果，相比于其他 2 种方法，本文方法得到了更低的平均误差。

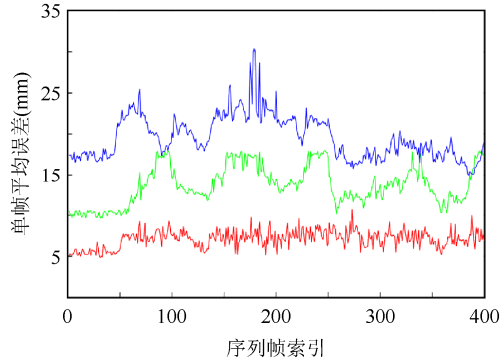


**图 6** 在序列 1 上单帧平均误差对比

**Fig. 6** Comparison of per-frame average numerical errors on sequence 1

图 7 给出了本文方法与文献[4]的定性比较结果。值得注意的是，文献[4]更关注于人体及一些非关节式运动物体的动态重建，而本文方法更局限于人手。此外，本文方法只恢复大尺度的表面运动，而文献[4]更进一步重建表面细节。从图中的最后 3 列可以看出，在严重遮挡发生的情况下，本文方法相比文献[4]产生了更准确的重建结果。

图 8 给出了本文方法在序列 2 上的部分重建结果。从图 8 可以看出，本文方法能够从包含各种复杂手部姿态并存在严重遮挡情况的单目深度序列中重建出准确的手部表面。


**图 7** 在序列 1 上本文与文献[4]方法的定性结果对比

**Fig. 7** Visual comparisons with literature [4] on sequence 1

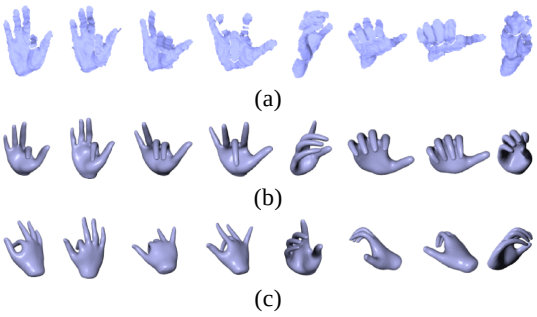


**图 8** 本文方法在序列 2 上的一些重建结果((a)为输入深度图像；(b)和(c)为重建的网格在不同视点下的渲染结果)

**Fig. 8** Some reconstruction results on sequence 2 ((a) The input depth image; (b) ~ (c) Rendering results of reconstructed hand from different viewpoints)

## 4 结 论

本文通过引入多个三维关键点到关键点的对应性，提出了一种新的非刚性手部动态重建方法。可发现这些三维关键点之间的对应关系可以作为手部关节区域局部变形和大尺度运动重构的重要线索。本文使用几个能量项描述这些对应关系，并将其整合到一个非刚性配准管线中，以实现精确的曲面拟合。之后借助于一个简单的多帧稳定策略来减少误差累积，最终获得了高质量的重建结果。在 2 个具有噪声的手部深度序列上进行的实验表明，本文方法不但能够准确的重建出手部运动和表面，而且对于大的变形和遮挡鲁棒。

本文方法虽然能够重建大尺度的表面运动，但仍然受限于恢复手部精细尺度的表面细节，如皱纹和褶皱等。未来工作中可以考虑引入一种基于自适应可见性技术的细节保持方法[24]来尝试解决这一问题。

## 参 考 文 献
## References


[1] ZUFFI S, BLACK M J. The stitched puppet: a graphical model of 3D human shape and pose[C]//2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). New York: IEEE Press, 2015: 3537-3546.

[2] LI H, ADAMS B, GUIBAS L J, et al. Robust single-view geometry and motion reconstruction[J]. ACM Transactions on Graphics, 2009, 28(5):175.1-175.9.

[3] ZOLLHÖFER M, NIESSNER M, IZADI S, et al. Real-time non-rigid reconstruction using an RGB-D camera[J]. ACM Transactions on Graphics, 2014, 33(4): 156.1-156.12.

[4] GUO K W, XU F, WANG Y G, et al. Robust non-rigid motion tracking and surface reconstruction using L0 regularization[C]//2015 IEEE International Conference on Computer Vision (ICCV). New York: IEEE Press, 2015: 3083-3091.

[5] MOESLUND T B, HILTON A, KRUGER V, et al. A survey of advances in vision-based human motion capture and analysis[J]. Computer Vision and Image Understanding, 2006, 104(2): 90-126.

[6] WASCHBÜSCH M, WÜRMLIN S, COTTING D, et al. Scalable 3D video of dynamic scenes[J]. The Visual Computer, 2005, 21(8-10): 629-638.

[7] STARCK J, HILTON A. Surface capture for performance-based animation[J]. IEEE Computer Graphics and Applications, 2007, 27(3): 21-31.

[8] THEOBALT C, AGUIAR E, STOLL C, et al. Performance capture from multi-view video[J]. Image and Geometry Processing for 3-D Cinematography, 2010, 5: 127-149.

[9] BUDD C, HUANG P, KLAUDINY M, et al. Global non-rigid alignment of surface sequences[J]. International Journal of Computer Vision, 2013, 102(1-3): 256-270.

[10] YE G Z, LIU Y B, HASLER N, et al. Performance capture of interacting characters with handheld kinects[C]//European Conference on Computer Vision. Heidelberg: Springer, 2012: 828-841.

[11] DOU M S, FUCHS H, FRAHM J M. Scanning and tracking dynamic objects with commodity depth cameras[C]//2013 IEEE International Symposium on Mixed and Augmented Reality (ISMAR). New York: IEEE Press, 2013: 99-106.

[12] DOU M S, TAYLOR J, FUCHS H, et al. 3D scanning deformable objects with a single RGBD sensor[C]//IEEE Conference on Computer Vision and Pattern Recognition (CVPR). New York: IEEE Press, 2015: 493-501.

[13] DOU M S, KHAMIS S, DEGTYAREV Y, et al. Fusion4D: real-time performance capture of challenging scenes[J]. ACM Transactions on Graphics, 2016, 35(4): 114.1-114.13.

[14] CAGNIART C, BOYER E, ILIC S. Probabilistic deformable surface tracking from multiple videos[C]// European Conference on Computer Vision. Heidelberg: Springer, 2010: 326-339.

[15] WEI X, ZHANG P, CHAI J. Accurate realtime full-body motion capture using a single depth camera[J]. ACM Transactions on Graphics, 2012, 31(6): 188.1-188.12.

[16] TAGLIASACCHI A, SCHRÖDER M, TKACH A, et al. Robust articulated-ICP for real-time hand tracking[J]. Computer Graphics Forum, 2015, 34(5): 101-114.

[17] FAN Q, SHEN X K, HU Y. Detail-preserved real-time hand motion regression from depth[J]. The Visual Computer, 2018, 34(9): 1145-1154.

[18] TAM G K L, CHENG Z Q, LAI Y K, et al. Registration of 3D point clouds and meshes: a survey from rigid to nonrigid[J]. IEEE Transactions on Visualization and Computer Graphics, 2013, 19(7): 1199-1217.

[19] ZENG M, ZHENG J X, CHENG X, et al. Templateless quasi-rigid shape modeling with implicit loop-closure[C]//2013 IEEE Conference on Computer Vision and Pattern Recognition. New York: IEEE Press, 2013: 145-152.

[20] LI H, VOUGA E, GUDYM A, et al. 3D self-portraits[J]. ACM Transactions on Graphics, 2013, 32(6): 1-9.

[21] NEWCOMBE R A, FOX D, SEITZ S M. DynamicFusion: Reconstruction and tracking of non-rigid scenes in real-time[C]//2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). New York: IEEE Press, 2015: 343-352.

[22] INNMANN M, ZOLLHÖFER M, NIEßNER M, et al. VolumeDeform: real-time volumetric non-rigid reconstruction[C]//European Conference on Computer Vision. Heidelberg: Springer, 2016: 362-379.

[23] TOMPSON J, STEIN M, LECUN Y, et al. Real-time continuous pose recovery of human hands using convolutional networks[J]. ACM Transactions on Graphics, 2014, 33(5): 1-10.

[24] ZHOU Y, SHEN S H, HU Z. Detail preserved surface reconstruction from point cloud[J]. Sensors, 2019, 19(6): 1278.