指导教师：杨涛　　提交时间：2015.3.29

The task of Digital Image Processing

数字图像处理

School of Computer Science

No: 01

姓名：张晓桐

学号：2012302622

班号：10011204

大规模动态三维重建的 MAP 可见性估计*

Hanbyul Joo　　Hyun Soo Park　　Yaser Sheikh

卡内基梅隆大学

{hanbyulj,hyunsoop,yaser}@cs.cmu.edu

摘要

三维动作重建中很多传统的挑战，例如匹配广基线和解决遮挡问题。随着新观点的提出，这些难题已经不再重要。但是，要想运用这些观点，又出现了一个新挑战，即准确预测在每一个时刻哪个相机观测哪个点。我们提出通过目标点随时间变化可见性的最大后验概率（MAP）估计来从大量相机中重建一个事件的三维动作。我们的算法以相机位置和图像序列为输入，以相机的时间变化集为输出。在这个变化集中，目标面片和重建轨迹是可见的。我们通过合并包括光照一致性，动作一致性和几何一致性的各种线索对可见性预测建模，即 MAP 预测，结合事先奖励临近相机一致可见性。通过找到相机中生产图的最小割集，可以获得可见性的最优估计。可以证明，我们的估计可见性的方法有很高的精度，同时在很多地点提高了追踪能力，产生了更长的轨迹。比起那些忽略可见性或仅使用光照一致性的方法，我们的方法有更高精度。

1.简介

在世界各地，成千上万的图像成为了最重要的地标。这些图像的实用性推动了大规模三维重建算法的发展。这些算法充分利用了视图的数目来产生密集且精确的三维点云[16, 13, 8]。逐渐地，在重大的体育比赛、音乐会和政治集会中，成百的相机也按比例地捕捉到了这些里程碑事件。但是，类似的可以完全利用一个事件中大量照片来产生长的密集精确三维轨迹的大规模重建算法还没出现。

这些基于视频的三维动作重建是很有难度的，因为自然动作由于遮挡，会导致一个更好测量损失的产生，同时还会导致图像中的人造品（例如，动作模糊和结构变形）。利用大量相机可以解决这些问题，因为它更可能（1）在临近相机中缩短平均基线，（2）减少遮挡的出现，（3）多元视角提供鲁棒的噪声检测。但是，（就重建的轨迹线长度密度和定位精度来说，）先前的方法不能完全利用多样的视角来提高三维追踪的效果。其主要的原因是由于在推导动态三维点随时间变化可见度时产生了错误。如果一个算法没有意识到某个点在交叉图像上是可见的，那么这个算法就无法利用交叉视角，从而使低可见度推导严重地影响追踪效果。进一步说，一个相机中的可见点可以使重建产生偏差，这个错误的结论通常会导致典型的“跳跃”人造品，其中一个点被假设成了另一个地方的点。

在这篇论文中，我们证明了精确推断点的可见性可以是重建算法充分利用大量场景来生成高精度三维长轨迹线。特别地，我们的核心算法贡献有：（1）使用动作一致性作为运动点可见性的线索；（2）使用那个视角规则性作为先验，并用视角接近度来衡量；（3）通过合并光照和几何一致性线索，使用最大后验概率（MAP）估计进行可见度估计。通过场景中的 480 台包含重要遮挡，大移位，改变场景拓扑的相机，我们展示了重建三维动作的经验主义效果。

*http://www.cs.cmu.edu/~hanbyulj/14/visibility.html

2.相关工作

动态三维重建方法可以广义地分为使用轮廓重建（例如，[5, 6, 18, 24, 3]）和使用一致性重建（例如，[22, 4, 17]）两种方法。基于轮廓的方法是典型的实用可视化外形生成高密度重建，但是需要后续处理来估计三维轨迹线[5, 18]。在连续帧之间使用表面匹配算法提供密集通信[20, 17, 21]。在这些方法中，每一帧中的网格模型是使用从图形到轮廓的方法独立地产生的。并且关键网格顶点间的稀疏匹配是通过使用各种线索来展现的，例如形状和外观特征。然后基于稀疏匹配的密集匹配可以使用基于测量距离的归一化代价函数来实现。动作预估计的精度高度取决于初始表面和结构，由顶点分别率界定。基于轮廓的方法还需要静态相机来预测准确的轮廓。

与基于轮廓的重建方法相比，基于一致性的方法可以生成更稀疏的重建，但是不需要静态相机，也不直接生成三维轨迹线。在所有基于一致性的方法中，最相关的方法或许是由 Vedula 等人[22]提出的流动场景重建法。假设可见性是通过先验的物体形状重建给定的，通过多个校准的相机摆成三角状来独立地估计二维视觉流动，从而生成三维流。同时，为了覆盖形状（深度）和动作，有很多子序列算法也提出了[1, 23, 11]。这些方法的基本假设是亮度是连续的（或者光度连续），这可以确定视角中的一致性。空间规律性是用来确定最优性并降噪的。虽然这些方法是针对三维点的，但其他方法更具三维代表性，例如动态表面形状[4,7]和网格[9]。基于网格的方法具有鲁棒的结果，生成的轨迹也持续地更加长。但是需要付出的代价是假定一个好的拓扑结果，即网格，并且需要使用规律性。

在之前的工作中，只考虑了一小部分典型的相机。在流动场景法中，通常会使用立体相机。其他的一些方法通常最多用 10 到 20 个相机（Vedula 等人使用了 17 个，Furukawa 和 Ponce 使用了 22 个，Huguet 和 Devernay 使用了 8 个，Carceroni 和 Kutalakos 使用了 7 个）。在这个规模上，由于动作模糊产生的信息丢失，结构遮挡变形和自遮挡是很严重的，因此，重建中重要的时空规律性是很有必要的。在大部分算法中，精准的相机可见性信息是没有考虑的，因为从少量异常相机中产生的噪声是可以忽略的。相机可见性可以是假设由给定的三维重建算法提供，也可以是通过一份鲁棒的估计量来解决。基于面片的方法使用了光度一致性，通过对比视角中的结构来确定可见性。但是，这些方法都需要三维面片的结构，这个结构高度取决于获得的面片形状精确度。

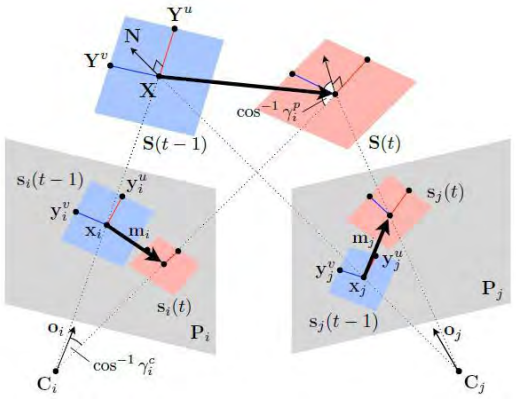


图1. *t*-1到*t*内一个面片的运动通过多相机重建

3.符号

我们的算法需要以 N 个校准和同步的相机在 F 帧内拍出的图片序列作为输入，以 P 个移动点的三维轨迹和他们实时方向作为输出，并且联接了每一个相机帧的可见性。

如图 1 所示，我们跟踪了一个以目标三维点 $\mathbf{X}\in\mathbb{R}^3$,为中心的平行四边形面片，它被定义为两个额外的点 $\mathbf{Y}^u$ 和 $\mathbf{Y}^u\in\mathbb{R}^3$。根据确定的面片上网格位置的数量，用联结了归一化强度值的单位向量定义联结面片的结构信息 $\mathbf{Q}\in\mathbb{R}^m$，其中 m 是网格上的位置数量[1]。面片 $\mathrm{S}(t)$ 用集合 $\{\mathbf{X}(t),\mathbf{Y}^u(t),\mathbf{Y}^v(t),\mathbf{Q}(t)\}$ 表示，它与相机可见性集合 $\mathbf{V}(t)=\{\mathrm{v}_1(t),\cdots,\mathrm{v}_N(t)\}$ 有关。其中，$\mathrm{v}_i(t)$ 是代表第 i 个相机可见性的二元值。三维点投影到第 i 个相机上，组成了一个 3*4 的投影矩阵 $\mathbf{P}_i$。这个投影矩阵以相机

[1]向量$\mathbf{Q}$的结构进行如下归一化：

$$\mathbf{Q}=\frac{1}{\sqrt{\sum_{j=1}^{m}(Q_j-\bar{Q})^2}}\begin{bmatrix}Q_1-\bar{Q}\\ \vdots\\ Q_m-\bar{Q}\end{bmatrix},\qquad(1)$$

其中 $\bar{Q}=\sum_{j=1}^{m}Q_j/m$，$Q_j$ 是结构的第j个强度值。

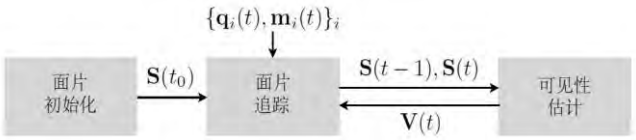


图2. 面片追踪和可见性估计概述

中心向量 $\mathbf{C}_i \in \mathbb{R}^3$ 和 3*3 的旋转矩阵 $\mathbf{R}_i \in SO($ 3为参数。方向向量 o_i 的方向于相机的z轴，也就是 $\mathbf{R}_i^{\mathrm{T}}$ 的第三列。

三维面片投影到相机平面组成了投影面片 $\mathrm{s}_i(t)=\{\mathrm{x}_i(t),\mathrm{y}_i^u(t),\mathrm{y}_i^v(t),\mathrm{q}_i(t)\}$，其中 $\mathrm{x}_i(t)$，$\mathrm{y}_i^u(t)$ 和 $\mathrm{y}_i^v(t)$ 是投影点，即 $\hat{\mathrm{x}}_i(t)\cong\mathbf{P}_i\hat{\mathbf{X}}(t)$，$\hat{\mathrm{y}}_i^u(t)\cong\mathbf{P}_i\hat{\mathbf{Y}}^u(t)$ 和 $\hat{\mathrm{y}}_i^v(t)\cong\mathbf{P}_i\hat{\mathbf{Y}}^v(t)$ 其中 $\hat{\cdot}$ 是每个向量的齐次坐标表示。$\mathrm{q}_i\in\mathbb{R}^m$ 是投影面片的结构信息，由相互联系的所有来自第i个相机的强度所定义，与$\mathbf{S}$的投影网格位置相对应，由公式（1）归一化。在不计算光照变化的理想情况下，如果三维面片$\mathbf{S}$从第i个相机是可见的，那么 $\mathbf{Q}=\mathrm{q}_i$。我们用 m_i 表示第i个相机中在 $\mathrm{x}_i(t-1)$ 时的二维视觉流，如图1所示。

第i个相机和面片之间的关系可以由共同可见性集合 $\boldsymbol{\Gamma}_i=\{\gamma_i^c,\gamma_i^p\}$ 定义，其中

$$\gamma_i^c=\frac{(\mathbf{X}-\mathbf{C}_i)^{\mathrm{T}}\mathrm{o}_i}{\|\mathbf{X}-\mathbf{C}_i\|}\text{ 以及 }\gamma_i^p=\frac{(\mathbf{C}_i-\mathbf{X})^{\mathrm{T}}\mathbf{N}}{\|\mathbf{C}_i-\mathbf{X}\|}$$

γ_i^c 编码了从相机方向向量 o_i 与面片位置之间的余弦角度，γ_i^p 表示相机位置和三维面片法向$\mathbf{N}$之间的余弦角度。

4.概述

在初始时间 t_0 重建目标三维面片，随着时间的推移，算法交替地估计面片的形状和常态，以及它对所有相机的可见性。需要注意的是，t_0 可以是任意帧，同时通过 t_0 的前向和后向来计算轨迹和可见性。我们只考虑前向追踪，简单来说，就是从 $t-1$ 到t。我们的算法流图如图2所示。

面片初始化。考虑在同一时刻不同相机拍摄的图片，算法通过在RANSAC框架中匹配特征并且使它们呈三角状来重建三维点。三维面片中心X通过最大化可见面片的相机中光度一致性来重建[2]。初始化为 $\mathbf{S}(t_0)$ 和 $\mathbf{V}(t_0)$。

面片追踪。考虑之前获得的三维面片 $\mathbf{S}(t-1)$ 和可见性 $\mathbf{V}(t-1)$，算法根据由为 $\mathbf{V}(t-1)$ 定义的相机的二维视觉流估计下一个三维面片 $\mathbf{S}(t)$。对于在 $\mathbf{V}(t-1)$ 的第i个相机，视觉流由在点 $\mathrm{x}_i(t-1)$, $\mathrm{y}_i^u(t-1)$ 和 $\mathrm{y}_i^v(t-1)$ 的多规模估计得出。为了估计不可靠流，在每个尺度的流由一个后向-前向一致性检验来生成，只保留最可信的流。下一个三维位置，$\mathrm{X}(t)$, $\mathrm{Y}^u(t)$ 和 $\mathrm{Y}^v(t)$ 是由在RANSAC框架中三角化视觉流输出来估计的。RANSAC过程至关重要，因为由于运动，$\mathbf{V}(t-1)$ 在时间t可能不再有效。RANSAC之后，常态通过最大化图像中的光度一致性得到了改善。这些图像属于RANSAC的在面片初始化过程中的内联。

可见性估计。基于重建了的 $\mathbf{S}(t)$ 和由 $\mathbf{S}(t-1)$ 得到的动作，我们的方法通过融合光度一致性，动作一致性和几何一致性，结合马尔可夫随机领域(MRF)先验，找到当前可见性集 $\mathbf{V}(t)$ 的MAP估计。典型的是，追踪过程严重受看不见目标的错误正面相机所影响。RANSAC阶段产生的低可见性会给不同的场景点造成典型的“跳跃”错误，同时也会降低常态优化的效果，在最优化过程中造成频繁的局部最小值。我们的精确可见性估计结果可以使在长追踪轨迹中有高准确率。

面片追踪和可见性估计是相互依赖的过程。在每一个时刻，我们可以将这两个步骤迭代直至收敛。实际中，一个迭代就足够了。

5.可见性估计

在这部分，我们提出一个用光度一致性，动作一致性和几何一致性，根据邻近先验来计算可见性V的最大后验概率（MAP）估计方法。这些线索用二维结构，二维视觉流和共同可见性集来表示。根据这些线索，运用贝

[2]参与RANSAC的相机被用作初始可见集，参考相机 $\mathbf{P}_{\mathrm{ref}}$ 被选为在内联集合离初始三维点最近的相机。以X为中心的三维面片初始化为一个较好规模的正方形面片（40mm*40mm），$\mathbf{N}$平行于 $\mathbf{o}_{\mathrm{ref}}$。我们基于Furukawa和Ponce [10]提出的方法来修改面片，选择了一个离现标准面片最近的相机作为新的参考相机。相应的可见性集合由选中的相机更新，这些相机的归一化交叉相关（NCC）得分比开始的 $\mathbf{P}_{\mathrm{ref}}$ 更高。在面片初始化过程中，常规改善和可见性更新是迭代的。

叶斯理论，可见性的概率是：

$$P(\mathrm{V}|\mathrm{q}_1,\mathrm{m}_1,\Gamma_1,\cdots,\mathrm{q}_N,\mathrm{m}_N,\Gamma_N) \propto P(\mathrm{q}_1,\mathrm{m}_1,\Gamma_1,\cdots,\mathrm{q}_N,\mathrm{m}_N,\Gamma_N|\mathrm{V})P(\mathrm{V}).$$

考虑到每个相机的可见性，我们假设（1）这个相机的线索是条件独立于别的相机中的线索和可见性，（2）同一个相机中的每个线索是互相条件独立的。其可能性可以写成

$$\left(\prod_{i=1}^{N} P(\mathrm{q}_i|\mathrm{v}_i)P(\mathrm{m}_i|\mathrm{v}_i)P(\Gamma_i|\mathrm{v}_i)\right)P(\mathrm{V}). \quad (2)$$

可见性V*的MAP估计可以通过求公式（2）的最大值，即，

$$\mathrm{V}^* = \arg\max_{\mathrm{V}}\left(\prod_{i=1}^{N} P(\mathrm{q}_i|\mathrm{v}_i)P(\mathrm{m}_i|\mathrm{v}_i)P(\Gamma_i|\mathrm{v}_i)\right)P(\mathrm{V}),$$

相当于

$$\mathrm{V}^* = \arg\max_{\mathrm{V}} \sum_{i=1}^{N}\log P(\mathrm{q}_i|\mathrm{v}_i) + \sum_{i=1}^{N}\log P(\mathrm{m}_i|\mathrm{v}_i) + \sum_{i=1}^{N}\log P(\Gamma_i|\mathrm{v}_i) + \log P(\mathrm{V}). \quad (3)$$

我们描述了每一个线索的概率以及在后面阶段中的先验。通过寻找可能相机中的最小割集来计算MAP估计[2]。

5.1.光度一致性

光度一致性在推导可见性中得到了广泛使用[16, 13, 8, 9, 7]。它衡量了三维面片的结构$\mathbf{Q}$和第i个相机中相应面片的结构q_i之间的关系。归一化的交叉相关（NCC）是光度一致性的衡量方法之一，对光照变化具有鲁棒性。由于$\mathbf{Q}$和q_i被公式（1）定义为了归一化的单位向量，$\mathbf{Q}^{\mathrm{T}}\mathrm{q}_i$衡量NCC。我们用围绕$\mathbf{Q}$，即由$\mathbf{Q}^{\mathrm{T}}\mathrm{q}_i$定义的$\mathrm{q}_i \sim \mathcal{V}(\mathbf{Q},\kappa)$的冯Mises-Fisher分布给$\mathrm{q}_i$的可能性分布建模。$\kappa$是控制结构变化度的聚焦参数。降低$\kappa$的值会使$\mathbf{Q}$和$\mathrm{q}_i$之间的变化加大。从分布中，我们可以看出，对给定$\mathrm{v}_i$，$\mathrm{q}_i$可能性的对数是

$$\log P(\mathrm{q}_i|\mathrm{v}_i) \propto \kappa\mathbf{Q}^{\mathrm{T}}\mathrm{q}_i \quad (4)$$

5.2.动作一致性

在动态场景中，动作是用来确定可见性的有效线索。给定面片的三维动作，如果第i个相机可以见到面片，那么它观测到的视觉流与目标面片的三维动作投影肯定是一致的。换句话说，动作一致性需要二维视觉流m_i必须与三维动作$\mathrm{x}_i(t)-\mathrm{x}_i(t-1)$的移位投影是一致的。

我们用三维移位投影正态分布来对可能性分布m_i建模，即$\mathrm{m}_i \sim \mathcal{N}(\mathrm{x}_i(t)-\mathrm{x}_i(t-1),\sigma)$，其中$\sigma$是从像素集合中三维运动估计的确定性得到的标准差。因此，对数可能性可以写成

$$\log P(\mathrm{m}_i|\mathrm{v}_i) \propto -\frac{\|\mathrm{m}_i-(\mathrm{x}_i(t)-\mathrm{x}_i(t-1))\|^2}{2\sigma^2}. \quad (5)$$

动作一致性是一个必要的条件。我们总结了当动作一致性线索是模糊的情况。设$\mathrm{X}(t)$和$\mathrm{X}'(t)$是三维空间中得两个不同点。当且仅当下面两个条件成立时，动作一致性线索是模糊的：

$$\begin{aligned} \mathrm{P}_i\hat{\mathrm{X}}(t) &\cong \mathrm{P}_i\hat{\mathrm{X}}'(t) \\ \mathrm{P}_i\hat{\mathrm{X}}(t+1) &\cong \mathrm{P}_i\hat{\mathrm{X}}'(t+1) \end{aligned} \quad (6)$$

其中$\|\mathrm{X}-\mathrm{C}_i\| > \|\mathrm{X}'-\mathrm{C}_i\|$，即$\mathrm{X}'(t)$遮挡了第$i$个相机的$\mathrm{X}(t)$。在静态的场景中，动作是不存在的。因此，由于$\mathrm{X}(t)=\mathrm{X}(t+1)$和$\mathrm{X}'(t)=\mathrm{X}'(t+1)$，动作一致性线索总是模糊的。另一种实践中出现的情况是当遮挡或被遮挡的面片在一个物体上进行了平移，相机之下的部分接近正投影。

我们总结了当公式（6）成立时的模糊动作集，假设$\mathrm{X}(t)$和$\mathrm{X}'(t)$在帧之间进行了同样的仿射变换：

$$\begin{aligned} \mathrm{X}(t+1) &= \mathrm{A}\mathrm{X}(t)+\mathrm{a} \\ \mathrm{X}'(t+1) &= \mathrm{A}\mathrm{X}'(t)+\mathrm{a}, \end{aligned} \quad (7)$$

其中，$\mathrm{A}\in\mathbb{R}^{3\times3}$和$\mathrm{a}\in\mathbb{R}^3$代表一个三维仿射变换。当且仅当下述情况成立时，动作一致性线索是模糊的：

$$\mathrm{X} \in \mathrm{null}([\mathrm{a}]\times\mathrm{A}), \quad (8)$$

其中，null(·)是·的空集。参见附录的证明。在精确无噪的理想情况下，这个情况很少发生（如果有动作）。

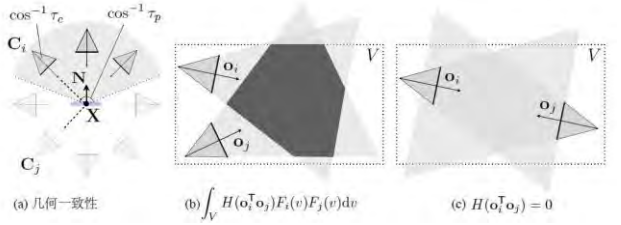


图3. (a) γ_p 和 τ_p 筛选的有效区域如阴影区所示。计算出的关于N的角度限制是 $\cos^{-1}\tau_p$。(b) 图中所示的阴影多边形是由两个相机所计算出的 g_s，其中，$\int_V \mathcal{H}(\mathbf{o}_i^{\mathrm{T}}\mathbf{o}_j)F_i(v)F_j(v)\mathrm{d}v>0$。(c) $\mathcal{H}(\mathbf{o}_i^{\mathrm{T}}\mathbf{o}_j)=0$ 的例子如图所示。一个定向面片不可能被两个面向对方的相机同时看到。

5.3.几何一致性

定向面片只能被方向向量 $\mathbf{o}_i$ 是与面片法向量**N**相反方向，且在它之前的相机看到。我们基于共同可见性集合 Γ_i，包含了这个几何线索。同时，考虑相机位置和面片常规方向之间的关系，以及面片位置和相机方向向量之间的关系。对于给定可见性 $\mathbf{v}_i$，Γ_i 的可能性可以写成：

$$P(\Gamma_i|\mathbf{v}_i)=\begin{cases}\dfrac{1}{(1-\tau_c)(1-\tau_p)} & \text{如果}\gamma_i^c\geq\tau_c\text{且}\gamma_i^p\geq\tau_p\\ 0 & \text{其它}\end{cases} \quad (9)$$

其中，$\tau_c<1$代表相机观测地的余弦角，$\tau_p<1$是确定关于面片常态的带角可见性的极限（余弦角）。图3（a）是线索的一个例子，其中阴影部分代表根据 τ_p 得到的有效区域。

5.4.可见性规律先验

在相机可见性的马尔可夫随机领域先验下，我们把可见性 $P(\mathbf{V})$ 的联合概率分解成成对的可能性，即，

$$P(\mathbf{v}_1,\cdots,\mathbf{v}_N)=\prod_{i,j\in\mathcal{G}(i)}P(\mathbf{v}_i,\mathbf{v}_j) \quad (10)$$

其中，$\mathcal{G}(i)$ 是第i个相机的邻接相机指数集。这个分解获取了可见性的先验分布，表示两个有相似视角的相机可能有一致的可见性的先验。这个临近约束组成了先验知识，当光度一致性和动作一致性线索不充足时（例如，在单个相机中的动作模糊），可以调整有噪声的可见性。我们建立了联合概率对数可能性：

$$\log P(\mathbf{v}_1,\cdots,\mathbf{v}_N)\propto\sum_{i,j\in\mathcal{G}(i)}g_s(\mathbf{v}_i,\mathbf{v}_j), \quad (11)$$

其中 g_s 由两个相机之间的损失定义，使用两个相机平截头体的重叠体积。由下式估计得出：

$$g_s(\mathrm{P}_i,\mathrm{P}_j)=\frac{\int_V\mathcal{H}(\mathbf{o}_i^{\mathrm{T}}\mathbf{o}_j)F_i(v)F_j(v)\mathrm{d}v}{\int_V F_i(v)+F_j(v)-F_i(v)F_j(v)\mathrm{d}v}, \quad (12)$$

其中 v 是工作区 V 内的无穷小体积（见图3（c））。$F_i(v)$ 是一个二值函数，定义如下：

$$F_i(v)=\begin{cases}1 & \text{如果}v\text{在第}i\text{个相机中可见}\\ 0 & \text{其它}\end{cases} \quad (13)$$

$\mathcal{H}$ 是单位阶跃函数，考虑了一对相机朝向一致的情况。公式（11）得到了相机平截头体重叠部分的体积与相机平截头体集之间的比率。图3（b）阐释了 g_s 的含义，其中阴影多边形表示$\int_V\mathcal{H}(\mathbf{o}_i^{\mathrm{T}}\mathbf{o}_j)F_i(v)F_j(v)\mathrm{d}v$。图3（c）举了一个$\mathcal{H}(\mathbf{o}_i^{\mathrm{T}}\mathbf{o}_j)=0$的例子。

实际中，我们用三维像素离散化了有效容积，同时计算了能投影到两个相机中的普通三维像素数量。这使我们可以在临近相机中奖励一致可见性。

5.5.通过图像分割估计MAP可见性

在公式（3）中代入公式（4）（5）（9）和（11），可以计算出可见性V的MAP估计，因此，公式（3）可以重写如下：

$$\mathbf{V}^*=\arg\min_{\mathbf{V}}\sum_{i=1}^{N}E_d(\mathbf{v}_i)+\sum_{i,j\in\mathcal{G}(i)}E_s(\mathbf{v}_i,\mathbf{v}_j), \quad (14)$$

其中，E_d 表示光度一致性，动作一致性和几何一致性。E_s 表示相机之间的先验。

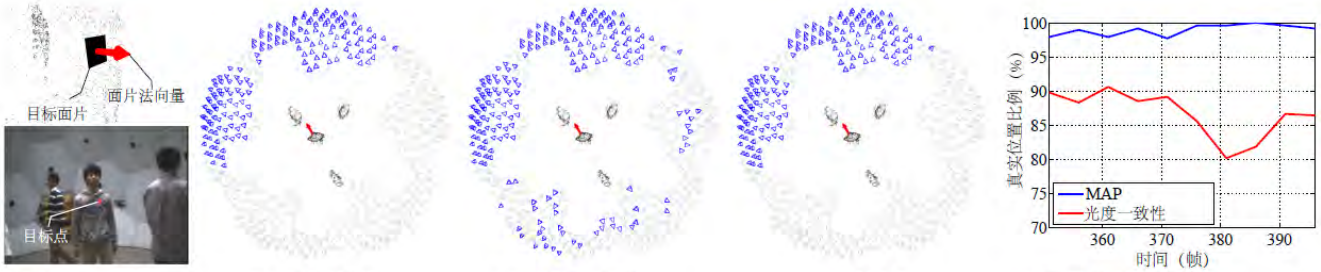
$$E_d(\mathbf{v}_i)=\frac{\|\mathbf{m}_i-(\mathbf{x}_i(t)-\mathbf{x}_i(t-1))\|^2}{2\sigma^2}-\kappa\mathbf{Q}^{\mathrm{T}}\mathbf{q}_i+\delta(\Gamma_i)$$

$$E_s(\mathbf{v}_i,\mathbf{v}_j)=\begin{cases}0 & \text{如果}\mathbf{v}_i=\mathbf{v}_j\\ g_s(\mathrm{P}_i,\mathrm{P}_j) & \text{其它，}\end{cases}$$

其中，如果满足 $\gamma_i^c>\tau_c$ 和 $\gamma_i^p>\tau_p$，那么可得 $\delta=\log(1-\tau_c)(1-\tau_p)$，其他情况 $\delta=\infty$。最小化的问题可以通过图像分割计算最优解[2]。

6.结果

我们在很多高难度场景中验证了我们的算法，包括存在显著遮挡的情况（循环运动和降落的盒子），大平移量的情况（五彩纸屑和流体运动），以及结构改变（下落盒子和排球）。可见性估计使我们可以更好地利用大量相机来生成精确且长的轨迹。表1总结了估计



(a) 选中面片　(b) 真实情况　(c) 光度一致性　(d) MAP　(e) 可见性准确度对比

图4.红色箭头表示选中面片的法向量。金字塔接头表示相机的姿态，其中蓝色相机属于可见集（为了更加形象化，把相机变弯了）。(a)选中面片是用三维视角和二维图像显示的。(b)我们手动生成了真实可见性。(c)通过基线估计的可见性。(d)通过我们方法估计的可见性。(e)我们对比了两种方法可见性估计的准确性。

中用到的数据集，在项目主页上可以找到。序列是在卡内基梅隆全景实验室上获取的，包括480个相机在25Hz下捕获的640*480的视频。相机经过了外在和内在校准，通过一个外在时钟同步。

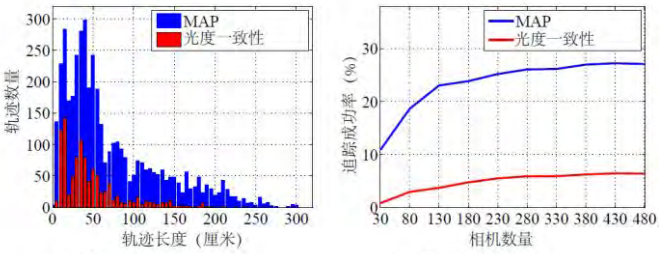
表1. 数据集总结

序列	帧	时长	点数	平均追踪长度
循环运动	250	10.0 sec	10433	404.9 cm
排球	210	8.4 sec	8422	326.4 cm
挥动球棒	200	8.0 sec	3849	224.1 cm
下落盒子	160	6.4 sec	17934	164.7 cm
五彩纸屑	200	8.0 sec	10345	103.0 cm
流体运动	200	8.0 sec	3153	123.1 cm

6.1.定量分析

可见性估计的准确度。我们在某一时刻重建的循环运动序列中选取了任意一个面片，并且在每一个目标面片可见的采样时刻手动地生成了真实的可见性数据。我们比较了可见性估计方法（MAP）和一个先前方法常用的仅基于光度一致性为线索的基线方法[4, 7, 9]。在一个时刻每一个方法产生的可见性估计结果如图4所示。我们以计算出的真实数据和通过两种方法估计出的V(*t*)之间的实际正观测比率作为标准。每一个方法得出的实际正比率如图4（e）所示，表明我们的方法大幅度胜过基线方法。

追踪准确度和长度。我们从追踪准确度和轨迹线长度两方面分析了我们的算法。从Furukawa 和Ponce [9]提出的评价标准中得到启发，测试序列是由把它反向附加到自己结尾而产生的。追踪算法是作用于生成序列



(a) 追踪效果　(b) 相机数量

图5. (a)我们的MAP估计在轨迹的数量和长度上都胜过基线方法。(b)我们的方法利用了大量的视角，并且比基线方法的曲线增长得更快。

上的。如果追踪是准确的，被追踪的面片必须回到初始位置。在这个实验中，三维错误是由目标点初始和结束位置之间的三维距离定义的。我们通过改变确定的初始帧持续时间（10到50帧），使用循环移动序列生成了五组测试序列。为了评估，我们计算了少于2cm漂移的成功重建轨迹数。图5（a）展示了使用480个相机得到轨迹数的直方图。我们的MAP估计方法无论实在轨迹的数量上还是长度上都比基于光度一致性的方法要优越。我们也用不同相机数做了实验来检测其对追踪成功率的影响，这些相机是通过均匀取样的。图5（b）展示了我们的方法是如何利用大量相机的。需要注意的是，成功追踪轨迹数比基于光度一致性的方法增长要快。

6.2定性分析

可见性边界。我们使用球棒挥动序列，通过展示三维可见性集中的相机以及目标面片在图像中的投影，定性证明了MAP可见性估计的效果。如图6所示。由图可见一个干净的可见性边界，展示了被棒球棒遮挡住的视角。

三维轨迹重建。我们对选定的时刻生成了一

图6.我们用球棒挥动序列定性证明了MAP可见性估计的效果。红色箭头是选定面片在三维视角（左）中的垂直方向，每张图片（右）的红色多边形是面片的投影。有蓝色边框的图片是属于可见集中的视角。当球棒遮挡住面片以及它的影响可以看成是可见集中的相机（左）的阴影。

个初始面片，并且进行了高达150帧的前向和后向面片追踪，所有的序列见表1。图7展示了重建的轨迹。重建的时刻进行了颜色编码。注意，我们的方法可以多次应用在不同的时刻来提高轨迹的密度。

循环移动。三个人绕中间的人旋转（图7（a））。这个实验是用于在可见性推理方面分析我们的方法。

排球。两个人在打排球（图7（b））。我们展示了动作很快以及遮挡严重的情况。在这种情况下，我们仍能够重建求和运动员的轨迹。

挥动球棒。一个人在挥动一个棒球棒。重建的长轨迹可以给运动分析，捕获微小运动提供计算基础。（图7（c））

下落的盒子。一个人使堆叠的盒子相撞，盒子倒塌了。这个场景包含了严重的遮挡和结构的拓扑变化。（图7（d））

五彩纸屑。一个人往空中撒五彩纸屑。三维重建的每一个序列都是很有挑战性的，因为遮挡和出现的变化。可见性估计也是很难得，由于纸屑很小并且他们的外观会发生突然变化。（图7（e））

流体运动。我们使用扇子和小纸屑在一个屋子里制造了混乱流（图7（e））[3]。

7.讨论

我们利用大量视角，展示了三维轨迹重建在时间变化可见性下的评价方法。我们为可见性估计提出了新线索（动作一致性，集合一致性和可见性规律先验），并将它们在MAP估计框架下与常见的广度一致性线索进行融合。我们证明了这个算法有更高的可见性准确率，并最终能比仅使用光度一致性的基线方法生成更长更密的轨迹线。与光度一致性线索不同，动作一致性线索是对光度线索的补充。因为它不需要目标三维面片的结构和清晰的形状。尽管动作一致性线索可能有些模糊，这在实际中通常发生在目标面片之后的相机上，此时整个目标体（包括面片）经历了一个纯转变。这种情况可以通过面片和相机的由几何一致性进行处理。

我们方法一个关键的好处是它在点的位置上不需要使用任何空间或时间正则化。我们方法使用的标度是可见性。这个结果真实地重建了三维点运动，不会因为变形的先验模型而产生偏差或不平缓。最常见的失败原因是出现后图像伪影，例如动作不清楚或饱和度不好。这失真是不可避免的，尤其在室外环境中。未来的一个重要发展方向是研究相应的技术来重关联点。

附录

公式（8）的证明：不失一般性，我们可以定义投影矩阵为$\mathrm{P}=[\mathrm{I}\ \ 0]$，然后，公式（6）可以重写如下，

$$[\mathrm{X}'(t)]_{\times}\,\mathrm{X}(t)=0 \qquad (15)$$

$$[\mathrm{X}'(t+1)]_{\times}\,\mathrm{X}(t+1)=0, \qquad (16)$$

给定$\mathrm{P}=[\mathrm{I}\ \ 0]$其中$[\cdot]_{\times}$是交叉乘积的倾斜对称表示。因为公式（15），$\mathrm{X}'(t)$与$\mathrm{X}(t)$呈线性比例。因此，$\mathrm{X}'(t)=\alpha\,\mathrm{X}(t)$。其中$\alpha$是标量。因为$\mathrm{X}'(t)\neq\mathrm{X}(t)$，$\alpha\neq 1$。

根据公式（7），公式（16）可以重写如下，

$$\begin{aligned} 0 &= [\mathrm{AX}'(t)+\mathrm{a}]_{\times}(\mathrm{AX}(t)+\mathrm{a}) \\ &= [\mathrm{AX}'(t)]_{\times}\,\mathrm{AX}(t)+[\mathrm{a}]_{\times}\,\mathrm{AX}(t) \\ &+ [\mathrm{AX}'(t)]_{\times}\mathrm{a}+[\mathrm{a}]_{\times}\mathrm{a} \\ &= (1-\alpha)[\mathrm{a}]_{\times}\,\mathrm{AX}(t), \qquad (17) \end{aligned}$$

[3]为了得到这个结果，我们置$\tau_c=0$和$\tau_p=0$，忽略了几何一致性，由于物体和平面很接近。

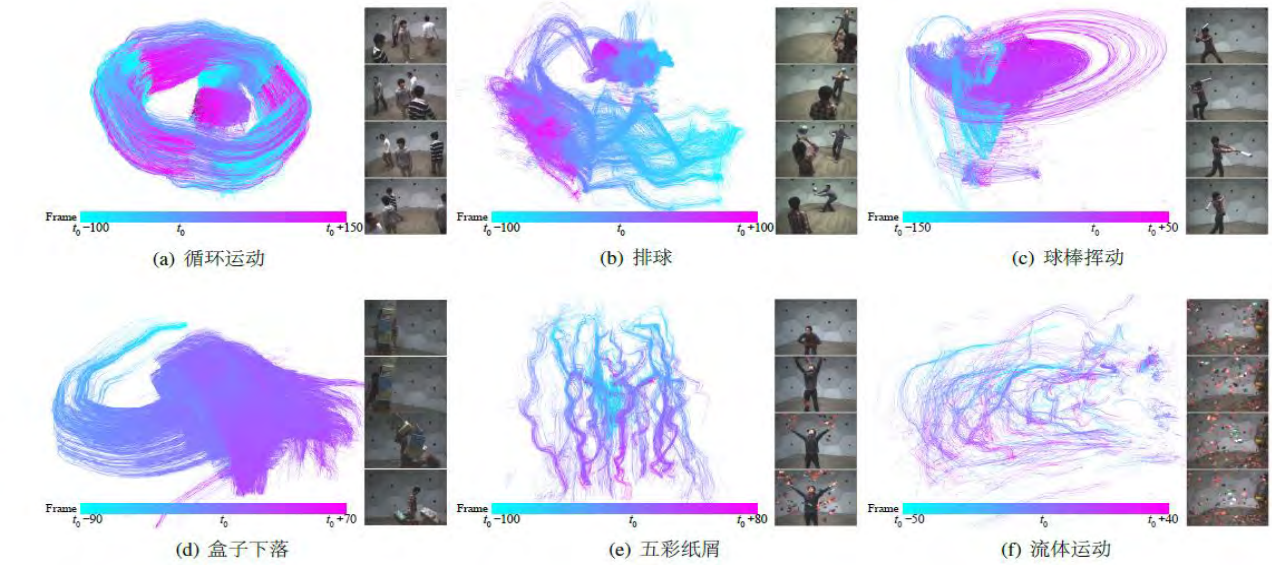


图7. 我们重建了有显著遮挡、大位移量和拓扑改变的现实场景中的三维轨迹。颜色编码了轨迹点重建的时间。需要注意的是每一个轨迹都是独立重建的，没有任何时间或空间的正则化。

其中，$[\mathrm{AX}'(t)]_\times \mathrm{AX}(t) = \alpha [\mathrm{AX}(t)]_\times \mathrm{AX}(t) = 0$。公式（17）蕴含了公式（8）。

致谢

这份材料是基于由国家自然基金支持，授予1353120和1029679号的项目。三星奖金对Hanbyul Joo提供了部分支持。

引用

[1] T. Basha, Y. Moses, and N. Kiryati. Multi-view Scene Flow Estimation:A View Centered Variational Approach. IJCV, 2012. 2

[2] Y. Boykov, O. Veksler, and R. Zabih. Fast approximate energy minimization via graph cuts. TPAMI, 2001. 5, 5.5

[3] C. Budd, P. Huang, M. Klaudiny, and A. Hilton. Topology-adaptive mesh deformation for surface evolution, morphing, and multiview reconstruction. IJCV, 2013. 2

[4] R. Carceroni and K. Kutalakos. Multi-view scene capture by surfel sampling: From video streams to non-rigid 3D motion, shape and reflectance. IJCV, 2002. 2, 6.1

[5] E. de Aguiar, C. Stoll, C. Theobalt, N. Ahmed, H.-P. Seidel, and S. Thrun. Performance capture from sparse multi-view video. TOG, 2008. 2

[6] E. de Aguiar, C. Theobalt, C. Stoll, and H.-P. Seidel. Marker-less deformable mesh tracking for human shape and motion capture. In CVPR, 2007. 2

[7] F. Devernay, D. Mateus, and M. Guilbert. Multi-Camera Scene Flow by Tracking 3-D Points and Surfels. In CVPR, 2006. 2, 5.1, 6.1

[8] Y. Furukawa, B. Curless, S. M. Seitz, and R. Szeliski. Towards internet-scale multi-view stereo. In CVPR, 2010. 1, 5.1

[9] Y. Furukawa and J. Ponce. Dense 3D motion capture from synchronized video streams. In CVPR, 2008. 2, 5.1, 6.1

[10] Y. Furukawa and J. Ponce. Accurate, dense, and robust multiview stereopsis. TPAMI, 2010. 2

[11] F. Huguet and F. Devernay. A variational method for scene flow estimation from stereo sequences. In ICCV, 2007. 2

[12] B. D. Lucas and T. Kanade. An iterative image registration technique with an application to stereo vision. IJCAI, 1981. 4

[13] J. michael Frahm, P. Fite-georgel, D. Gallup, T. Johnson, R. Raguram, C. Wu, Y. hung Jen, E. Dunn, S. Lazebnik, and M. Pollefeys. Building rome on a cloudless day. In ECCV, 2010. 1, 5.1

[14] J. Quiroga, F. Devernay, and J. Crowley. Scene flow by tracking in intensity and depth data. In CVPR Workshop, 2012. 2

[15] Y. Sheikh, S. Nobuhara, H. Joo, H. Liu, L. Tan, L. Gui, M. Vo, B. Nabbe, I. Matthews, and T. Kanade. The panoptic studio. Technical Report, CMU-RI-TR-14-04, 2014. 6

[16] N. Snavely, S. M. Seitz, and R. Szeliski. Photo tourism: Exploring photo collections in 3D. TOG, 2006. 1, 5.1

[17] J. Starck and A. Hilton. Spherical matching for temporal correspondence of non-rigid surfaces. In ICCV, 2005. 2

[18] J. Starck and A. Hilton. Surface capture for performance-based animation. CGA, 2007. 2

[19] N. Sundaram, T. Brox, and K. Keutzer. Dense point trajectories by gpu-accelerated large displacement optical flow. In ECCV, 2010. 4

[20] T. Tung and T. Matsuyama. Dynamic surface matching by geodesic mapping for 3D animation transfer. In CVPR, 2010. 2

[21] K. Varanasi and A. Zaharescu. Temporal surface tracking using mesh evolution. In ECCV, 2008. 2

[22] S. Vedula, S. Baker, P. Rander, R. Collins, and T. Kanade. Threedimensional scene flow. TPAMI, 2005. 2

[23] C. Vogel, K. Schindler, and S. Roth. 3D scene flow estimation with a rigid motion prior. In ICCV, 2011. 2

[24] A. Zaharescu, E. Boyer, and R. Horaud. Topology-adaptive mesh deformation for surface evolution, morphing, and multiview reconstruction. TPAMI, 2011. 2