# 丝路文化虚拟体验中的多视角立体重建技术研究

李兆歆[1)] 蒋浩[1)] 刘衍青[2)] 王兆其[1)]

[1)](中国科学院计算技术研究所，北京, 100190)

[2)](宁夏师范学院，固原, 宁夏 756000)

**摘 要** 丝路文化是联系一带一路战略的重要纽带，其传承意义重大，但是由于历史地理原因，丝路文化中代表性的历史遗产分散或损坏，难以有效地呈现，因此，本文面向丝路文化的虚拟展示与数字化，提出并实现了基于虚拟现实技术的丝路文化传承平台，通过历史遗迹复原以及基于图像的三维重建，还原了丝路文化中重要节点宁夏固原有关的历史遗迹、文物和事件。特别地，本文提出一种面向高清图像的多视角立体三维重建算法，包括采用 normal-aware PatchMatch stereo 复原高质量的法线图，反映文物表面精细结构，以及提出一种基于 GPU 的增量式的深度融合方法，以较小的显存处理大规模的数据。在公共数据集和本文收集的室内外文物数据上的实验表明，本文提出的三维重建方法可恢复物体表面的精细结构，同时还对大规模数据具有良好的可扩展性。重建的模型可导入到虚拟互动系统中，对丝路文化的传播起到了积极的作用。

**关键词** 多视角立体；深度融合；丝绸之路；文化遗产；虚拟现实系统

**中图法分类号** TP393

# Research on Multi-view Stereo 3D Reconstruction in Virtual Reality System of Silk Road Cultural Inheritance

Li Zhaoxin [1)] Jiang Hao [1)] Liu Yanqing [2)] Wang Zhaoqi [1)]

[1)]( Institute of Computing Technology, Chinese Academy of Science, Beijing 100190)

[2)]( Ningxia Normal University, Guyuan，Ningxia 756000)

**Abstract** Silk Road culture is an important link in the Belt and Road strategy. Its heritage is of great significance. However, due to historical and geographical reasons, the representative historical heritage in the Silk Road culture is scattered or damaged, and it is difficult to present the historical heritage effectively. Therefore, in this work we propose and implement a virtual reality platform for the Silk Road Cultural Heritage. Through historical restoration and image-based 3D reconstruction, we effectively restored the historical sites, cultural relics and events of Guyuan of Ningxia Province in China, one of the important nodes in Silk Road Culture. For outdoor historical sites, we use a DJI Mavic Pro to capture 4K video clips of the giant Buddha of Xumi Mountain in a sunny day. For indoor cultural relics, we use a turntable with digital single lens reflex (DSLR) camera and multiple light sources to capture high-resolution images in 180 degrees. Based on these image data, we propose a simple and efficient multi-view stereo 3D reconstruction method for high-resolution images, which consists of a normal-aware PatchMatch stereo for the high-quality normal recovery to represent the detailed surface of the cultural relics, and a GPU-friendly incremental depth map fusion method which can fuse a large amount of depth maps by leveraging a small size of GPU memory. The high-resolution input images are essential for representing the geometric details in historical sites and cultural relics. However, the

本课题得到国家自然科学基金(No. 61702482和61532002)，以及北京市自然基金（No. L172049）的资助. 李兆歆，博士，助理研究员，计算机学会(CCF)会员（70190M），主要研究领域为三维重建和三维计算机视觉. E-mail: cszli@hotmail.com. 蒋浩（**通信作者**），博士，副研究员，计算机学会(CCF)高级会员（18553S），主要研究领域为群体仿真，E-mail: jianghao@ict.ac.cn. 刘衍青，博士，教授，主要研究领域为固原历史文化. E-mail: 936207408@qq.com. 王兆其，博士，研究员，主要研究领域为虚拟现实. E-mail: zqwang@ict.ac.cn.

state-of-the-art depth map fusion method needs to import all depth maps and normal maps into the GPU memory, and then globally fuse the depth points into the 3D point clouds for each reference image. Nevertheless, the space complexity almost linearly increases when the amount of data and image resolution increase. For instance, doubling image size will result in a fourfold increase in GPU memory. Due to limitation of GPU memory, this kind of global fusion strategy cannot address high-resolution input image data. The proposed incremental depth map fusion method in this paper mainly consists of three steps: a）we first set a reference view and a counter map for cross-view consistency check; b) then, we import α neighboring images of the reference view into GPU memory each time, and perform the cross-view consistency check for the depth points in reference view. And then, depth points are accumulated, and counter map is also updated. We then release the memory of these α images and import another α images into GPU and repeat the above operations. c) When all neighboring views are processed, we can fuse the depth points whose values in counter map are larger than a threshold. The quantitative and qualitative experiment results on the public multi-view stereo benchmark as well as our captured datasets clearly highlight that the proposed method can recover the detailed surfaces while keeping a good scalability for the large-scale image data. The reconstructed high-quality 3D models of historical sites and cultural relics by our method can effectively support immersive virtual reality applications, playing a positive role in the dissemination of Silk Road culture.

**Key words** multi-view stereo; depth map fusion; silk road; cultural heritage; VR system

## 1 引言

丝绸之路在东西方文化交往中起到了十分重要的作用，丝路上留下的文化遗产是丝路历史的实物见证，具有极高的历史文化价值。宁夏固原作为古丝绸之路的重要节点，留下了丰富的丝路文化遗产。对这些历史文化遗产的保护和利用，对服务当地社会文化的发展具有重要意义，一方面可以方便地方历史文化的研究，另外一方面也可以作为文化传承的纽带帮助学生、公众了解当地的历史文化，把更多的人培养成为丝路历史文化的传播者和宣传者。

目前，针对历史文化遗产的传承与弘扬的传统方式主要有两类：一种主要采用文字和多媒体为主的传承与传播方式，这类方法比较灵活，却难以展现完整动态的历史风貌，缺乏用户交互，在自主体验方面有较大的局限性；另一种是利用文化古迹、文物等有形实物的方式，虽然这种方式可以方便人们直观地接触和体会历史文化风貌，但易受自然和人为因素的影响而遭受损坏，而且对丝路文化体验来说，古丝绸之路西北走廊多处于偏远山区，文物遗址分散于丝路沿线，地处偏远、交通不便，在便利性和全面性、以及传播发展和保存保护方面难以平衡。

针对传统方法的不足，基于虚拟现实技术的历史文化虚拟展示与数字化受到广泛关注，在国内外被应用于不同种类的历史文化传承与传播，取得了巨大的社会经济效益。陈颖借助虚拟现实技术，在“天坛神乐署中和韶乐”应用实践中建立演练场所的虚拟现实系统，提升了非物质文化遗产的动态展示效果[1]。针对大型的文化遗址，研究人员开展了相应的数字化技术研究，潘云鹤等人提出了一整套数字化壁画保护修复的技术，用于智能化壁画临摹辅助和石窟壁画文物保护修复辅助[2]，刘箴利用虚拟现实技术开发河姆渡遗址博物馆资源，制作了河姆渡遗址博物馆的三维场景，并在微机上采用 VRML 语言实现了一个三维漫游系统[3]。王圣华等人针对中国传统皮影的文物性保护与场景再现进行了研究，建立了皮影可视化数据模型和渲染方法，实现了逼真的中国皮影戏可视化表达[4]。刘世光等人基于流体力学和向量图表达技术，提出一种石纹纸染艺术图案生成方法，用于这种古老文化艺术形式的保护、记录与传承，可以绘制得到具有花型或特殊特征等具有艺术效果的石纹纸染图案[5]。为了弘扬中国剪纸艺术，张显全等人利用纹样组合，利用计算机辅助生成剪纸形象，可得到具有民族风格的剪纸图案[6]，涂传朋等人则从剪纸作品里面的流水动画效果着手，构建流水波纹模型以及它们的动态控制方法，使得手工剪纸中的流水通过用户少量交互就可以生成动画，用该方法生成的剪纸风格流水动画在视觉效果上自然、流畅，对推广剪纸艺术很有价值[7]。在历史文物的虚拟展示方面，刘晓等人对表面腐蚀或表面残缺不全的文

物表面图案进行建模[8]，用计算机修复虚拟文物，这样可以重现古文物的原来面貌，这对文物资源的立体展示以及文物资源考古研究等将起到重要的推动作用，但这种方法构建的文物三维模型只能重现原始文物，无法改变文物器型展示效果，而胡晏秋等人面向青铜器建模与绘制，通过区域几何特征对三维网格模型进行分割和参数化处理，实现青铜器纹的凹凸绘制以及表面锈蚀效果绘制，在三维展示、艺术设计等领域有较好的应用前景[9]。此外，增强现实技术也被广泛应用于文化遗产的数字化保护与传承[10]，研究人员针对不同类型文化遗产的特点，开展了有针对性的应用研究（例如，数字圆明园增强现实系统[11]）。

文物的数字化需要借助三维重建技术。相比基于激光扫描的方式[12]，基于多视角立体（Multi-view stereo, MVS）的方式具有低成本，和对室内外场景的广泛适应性。特别地，由于捕获高清图像数据越来越便利，多视角重建的应用越来越广泛。随着多个评估数据库的提出 [13-15], 多视角重建方法的精度和完整度不断提升。当前主流方法主要是基于深度图的 GPU 友好的 MVS 方法 [16-19], 这类方法首先重建每个视角的深度图，然后将不同视角的深度图导入到 GPU，通过跨视角一致性检验和一致性深度的平均，进行深度图融合，形成统一的三维点云 [17]。然而，由于噪声和视差范围的影响，重建的深度图往往不能很好地恢复物体表面的精细结构。通过增加正则化项[20, 21] 虽然可以减弱噪声带来的影响，但会增加计算成本，并破坏物体表面的细节。相比低分辨率图像，高分辨图像数据蕴含了物体表面可辨识的细节信息，可增加视角间的匹配精度，但是由于重建的高分辨率图像的深度图包含大量的数据，融合深度图将需要较大的显存资源，限制了方法在高清图像数据集的应用。

本文主要面向丝路重镇宁夏固原的代表性历史遗迹和文物等丝路文化遗产（如图 1），基于虚拟现实技术，综合建模技术和交互技术，通过高清图像数据采集与多视角三维重建，将丝绸之路中具代表性的文物古迹进行数字建模与虚拟化，提供一个高真实感、沉浸感和多种交互方式的虚拟现实互动环境，通过互动体验的方式，在宁夏、西部乃至世界范围内让更多的人能够更直观、全面的了解固原丝路文化。

本文的主要贡献包括：

（1）提出一种针对古迹和文物高清图像数据集的简单高效的三维重建算法，包括一种具有较低显存占用率的增量式深度融合算法，一方面显著提升了古迹和文物三维几何细节复原精度，另外一方面对图像分辨率和图像数据规模具有较好的可扩展性。

（2）通过多视角三维重建技术将丝路文化历史遗产进行数字化与虚拟化，提供一个高真实感的虚拟现实互动体验环境，提出的整个系统框架，特别是三维重建方法，可潜在应用于其它面向文化保护和传播的虚拟现实系统中。

（a） （b）

图 1. 固原境内的丝路历史遗迹和文物：（a）须弥山大佛造像；（b）固原市博物馆馆藏文物。

## 2 相关工作

本文面向室外大型古迹和室内馆藏文物提出一种有效的多视角立体重建算法。多视角立体重建旨在从场景的一组图像集中复原场景三维结构，是图像生成的逆过程。基于三维表面的表达形式，多视角立体重建方法可分为基于体素 [20, 21]、三角面片 [22, 23] 和深度图的方法 [16-19]，近年来基于深度学习的端到端重建方法也逐渐获得广泛关注 [24, 25]。关于多视角方法的详细综述可参见[26]， 本节主要讨论和本文工作密切相关的方法，即面向大尺度数据的基于深度图的多视角立体重建方法。

基于深度图的重建方法尝试估计输入图像的深度信息，取决于深度点的稠密度，可分为面向稠密深度图的重建方法和面向半稠密深度图的重建方法。半稠密深度图的重建方法主要通过提取图像中的显著特征点进行多视图间的匹配，也称为基于特征点的方法，典型工作如 PMVS [27]。这类方法对强纹理表面非常有效，然而容易在弱纹理区域处产生不完整的重建结果。稠密深度图的重建方法尝试估计每一参考图像的稠密深度信息，其主要步骤包括了深度图估计和深度图融合两个子问题。由于这类方法可以重建相对完整的场景表面并可借助 GPU 加速提升重建效率，是当前多视角重建的主流

方法[16-19]。Galliani 等[17] 首次将 Black-Red 网格传播方法应用于 PatchMatch Stereo, 方法通过 GPU 加速交替传播黑格和红格像素区域的深度和法线，显著提升了重建每一视角深度图的速度。Schönberger 等在[17]的基础上将遮挡关系作为约束项进行联合优化[16]，获得了更高质量的重建结果。[28] 提出一种非对称的网格传播方法，同时采用由粗到细（coarse-to-fine）的策略提升方法在弱纹理区域的表现。 基于深度学习的端到端的深度估计方法尝试使用神经网络刻画从图像到深度图的过程[24, 25], 对弱纹理较为鲁棒。[24] 提出一种面向稠密图像采样的小场景深度图预测网络，其网络需要固定输入的视点数量，在处理更大场景时无法有效的处理图像间可见性信息。[25] 提出一种面向大场景非结构化数据的网络架构，但是由于网络结构需要较大的显存，只能应用于低分辨率的深度图恢复，且无法有效重建物体表面的精细结构，限制了方法在高清图像数据和大尺度场景下的应用。而在深度图融合问题方面，为了高效的融合大尺度数据，基于 GPU 的逐像素滤波和融合 [17] 已经在各类主流方法上获得广泛应用 [19, 24, 28]。然而此类方法需要将目标场景的深度图、法线和原始图像数据全部导入显存，随着图像规模和分辨率的增大，其对显存的需求量迅速增加，限制了方法在大场景和高清图像数据集上的应用。

针对古迹和文物往往包含精细三维结构这一特点，本文提出一种针对高清图像数据集的简单高效的三维重建算法，一方面显著提升了古迹和文物三维细节复原精度，另外一方面对图像分辨率和数据规模具有较好的可扩展性。

## 3 总体设计

虚拟现实技术综合运用图形学、计算机视觉、多媒体等技术，将现实场景在计算机上真实呈现，日益成为一个研究热点。虚拟现实数字化互动体验系统将传统的丝路文化遗产展示从二维转向三维，能更逼真、更准确地呈现丝路文化遗产资源，使用户可以利用虚拟现实设备实现丝路文化遗产数字资源的 3D 浏览,有如身临其境的感觉。突破了时空的限制，让人们从能“走进”这一场景，真正做到方寸之间就能直接通过计算机全方位和清晰地体验丝路文化遗产数字资源。

丝路文化虚拟现实互动体验系统由 2 个底层基础模块构成，分别为数字资源采集，多视角立体重建与模型植入模块。 在此基础上建立数字资源库，协助虚拟现实互动体验平台的建立，实现人机交互以及场景的真实感展示。提出的系统平台示意图如图 2 所示。

系统使用的硬件资源主要分为数据采集、三维重建和虚拟展示三部分。为了在虚拟现实体验系统中能同步浏览丝路文化代表性历史遗产，需要采集历史遗迹或文物的三维数字资源，并根据采集对象的特点选用不同的采集方式。在本文中，主要考虑较低部署成本的、以图像采集为主的方案，包括无人机和自动控制的环绕采集系统。无人机比较灵活，适合采集较大范围的图像数据，比如人们难以获得完整数据的山体、大型像等。环绕采集系统可以配置好之后由软件自动控制，适合近距离采集物品清晰的图像序列，主要针对文物等体积较小的需要精细展示的代表性物品。此外，采集并进行标定后的数据，通过本文提出的三维重建流程生成三维

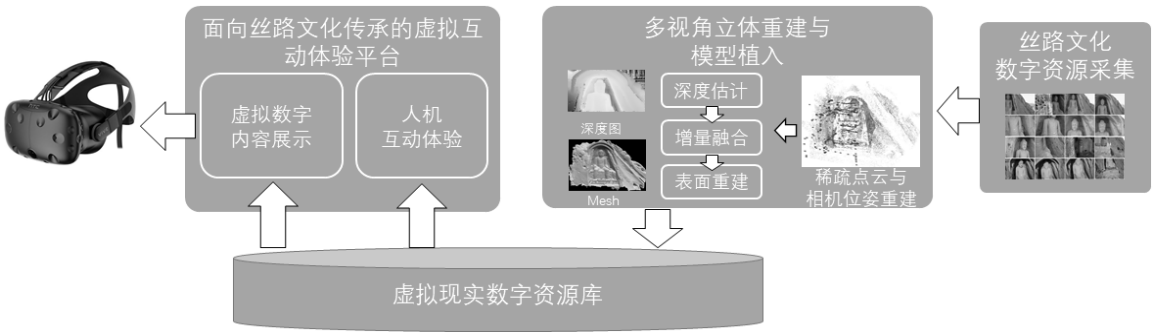



图 2 系统框架图

数字资源，再通过 VR 设备进行展示与体验，为获得更好的用户体验，本文主要使用 VR 头盔与 VR 自由行动平台相结合的方式，用户在使用头盔进行观察的同时，还可以在三维场景内自由行走，并通过手柄与系统进行互动，并辅助与背景资料、语音解说等相关说明材料进行关联。本文中的虚拟现实互动体验系统选择 Unity3D 作为软件开发平台。通过融合三维重建的虚拟场景，在系统中实时渲染沉浸画面和实时反馈用户交互信息，可实现高质量的虚拟体验。用户通过佩戴 VR 头盔，处在一个完全被包裹的环境中，大大提高用户的体验感和沉浸感。

作为 VR 系统的核心模块，本文系统中的三维重建模块可以处理高清图像的输入，并生成高精度的三维数字模型。接下来将对本文采用的三维重建方法的具体流程进行详细描述。

时间飞行法和结构光法相比，多视角立体方法具有更好的可扩展性，能够应用于大尺度场景的重建任务，并且只需使用相机对场景进行拍照，具有较低的应用成本。本文针对固原历史文化遗产中的两类代表性文物进行三维数字化：须弥山景区石窟遗址和固原馆藏文物（包括固原市博物馆、彭阳博物馆和农耕博物馆）。

为了真实地数字化文物数据，本文捕获高分辨率的图像数据以真实反映物体表面的精细结构。本文分别设计了基于 normal-aware PatchMatch stereo 和增量点云融合的方法来处理高分辨率图像数据。方法示意图如图 3 所示。下面将按照数据获取、相机标定和多视角立体重建的步骤，对本文方法进行详细介绍。

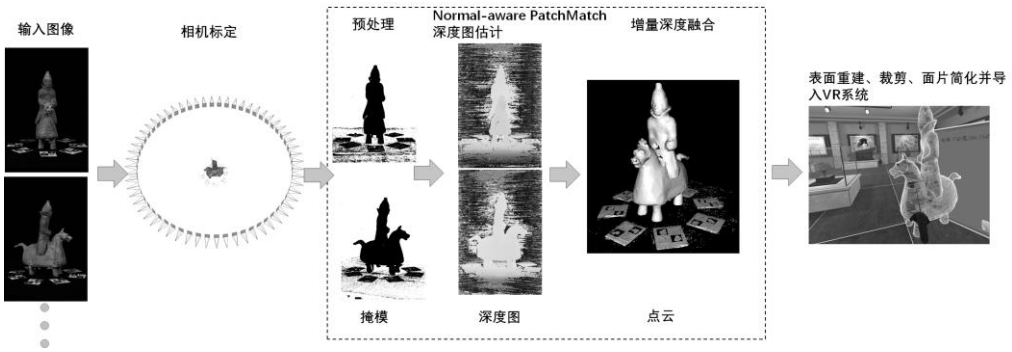
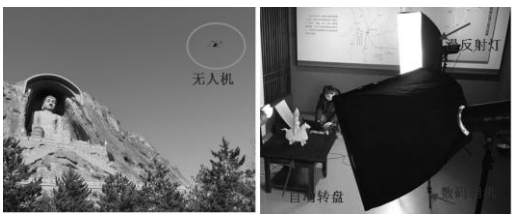



图 3　三维重建方案示意图

## 4 文化遗产多视角立体三维建模

文化遗产的三维数字化可用于文化的宣传与传承，三维展示使得大众通过交互应用或网络就能获得对文化瑰宝的身临其境的感受。文化遗产不仅包含了小尺度的文物藏品还包含室外大型建筑古迹。虽然基于结构光扫描的方式可以实现小尺度文物的精细重建，但此类方法成本较高，且易受光线干扰，不适用于大尺度的古迹重建。本文提出一种基于高清图像为输入的多视角立体（Multi-view stereo, MVS）重建方法，实现文物古迹的低成本高质量重建。数字化的三维模型可导入虚拟现实场景，使用户获得身临其境的体验效果。给定目标场景下的多视角图像数据，多视角立体方法尝试寻找图像间的对应点并推断场景的三维形状[13-15]。与




(a)　(b)

图 4　数据获取：(a) 室外无人机航拍；(b)室内自动拍摄

### 4.1 数据获取

多视角立体重建方法的输入是被测物的多角度 RGB 图像数据。针对须弥山石窟和固原馆藏文物的不同特点，本文采用不同的数据采集方案。

须弥山石窟是固原的代表性文化遗址，由于此遗址为大尺度场景，本文使用无人机航拍进行数据的采集，如图 4（a）。针对日照强度和方向，以及

风力对拍摄的影响，航拍时间选择晴朗无风的上午。本文使用 Mavic Pro 无人机，并在无人机移动过程中拍摄 4K 分辨率视频（分辨率 3840×2160），无人机每次飞行的有效航拍时间约为 20 分钟，为了尽可能让拍摄的图片覆盖大佛的不同区域，本文进行多次航拍。对拍摄的视频数据进行关键帧抽取，提取一组相互交叠且不同角度的视频帧，并手动剔除那些模糊的视频帧。

固原各博物馆收藏了大量与丝绸之路文化密切相关的珍贵文物。本文面向文物的拍摄系统包括可自动控制转动步长的电动转台和高清相机，以及若干漫反射光源。如图 4（b）所示，被拍摄的物体放置于转盘中心，当转盘匀速缓慢转动时，自动触发相机进行等时间间隔的拍摄，共拍摄以物体为中心的不同角度的 60 张高清图像（平均每 6 度拍摄一张图像，图像分辨率：5184×3456），为了清晰捕获文物表面的精细纹理避免散焦导致的模糊，相机在拍摄过程中采用自动变焦模式。值得指出的是，在上述两类不同特点场景的数据获取过程中，本文均是捕获高清图像。通过捕获高清图像可以让物体表面的纹理特征更具辨识度，一方面减少由于弱纹理和重复纹理导致的匹配二义性，增加了后续相机标定和三维重建的精度，另一方面也可以提升物体表面的小尺度几何结构的复原效果，而这些小尺度的精细几何结构往往是文物高超艺术性的体现。

### 4.2 相机标定

给定一组输入 RGB 图像$\{I_1 \ldots I_i \ldots I_n\}$，下一步将是对每一图像所对应的相机参数进行求解。考虑到对场景的适应性，本文采用 structure from motion (SFM)方法求解相机的内外参数。 对于一幅观察图像$I_i$，相机$i$的参数包括了相机的内部参数和相机的位姿。当假设场景符合针孔模型(pinhole model)时，其内参数矩阵可表达为 3×3 矩阵（公式（1））：

$$K=\begin{bmatrix} f_x & 0 & c_x \\ & f_y & c_y \\ & & 1 \end{bmatrix} \tag{1}$$

其中$f_x$与$f_y$为相机焦距在像平面 X 方向与 Y 方向的度量，$c_x$与$c_y$为相机主点位置，即相机光心在像平面上的投影位置，单位均为像素。相机的外参数包含了相机相对于参考世界坐标系的旋转和平移，分别用 3×3 旋转矩阵$R$和 3×1 平移向量$\boldsymbol{t}$表示。

使用 SFM 实现相机标定的步骤主要包含 3 步：特征点提取，多视角模型估计和捆绑调整（bundle adjustment）。SFM 效果很大程度依赖特征点的提取和匹配精度，通过尺度不变特征点检测与描述算法 SIFT，可以在每一高清图像上提取一些辨识度较高的特征点，并建立视角间的特征点对应。由于本文采用了高分辨率的图像作为输入，保证了特征点提取的精度，减少了噪声的影响。多视角模型基于特征点对应估计局部相机组的内外参数，同时剔除特征点对应中的误匹配（outliers）。经由上述两步，本文获得特征点对应、初始的相机内外参数和一组稀疏的三维点集$\Lambda$。为了进一步最小化计算误差，采用捆绑调整对相机参数进行进一步优化。

假设在世界坐标系三维点$\boldsymbol{X}_k \in \Lambda$在其可见相机集合$i \in N(\boldsymbol{X}_k)$的二维重投影$\boldsymbol{p}_i'$为$\phi_z\left(K_i\left(R_i\boldsymbol{X}_k+\boldsymbol{t}_i\right)\right)$，其中$\phi_z(\mathbf{a})$为深度归一化操作，即$\mathbf{a}/z$。图像$I_i$中与$\boldsymbol{X}$对应的特征点位置为$\boldsymbol{p}_i$，捆绑调整可表述为（公式（2））：

$$\varepsilon(\boldsymbol{X},K,R,\boldsymbol{t})=\sum_{\boldsymbol{X}_k\in\Lambda}\sum_{i\in N(\boldsymbol{X}_k)}\rho\left(\boldsymbol{p}_i-\phi_z\left(K_i\left(R_i\boldsymbol{X}_k+\boldsymbol{t}_i\right)\right)\right) \tag{2}$$

其中$\rho$为鲁棒惩罚函数用于度量重投影误差。公式（2）可通过迭代优化方法 Levenberg-Marquardt 方法求解。

SFM 的重建具有尺度二义性，即无法确定场景的真实物理尺寸。对于基于转盘的拍摄系统，本文在拍摄时在文物周围放置若干标定板（图 5 (a)和图 5 (b)分别为两个被测物体和放置的标定板），从而增加了 SFM 的精度和成功率，同时由于可以测量标志物的真实物理尺寸（毫米），从而可以获得 SFM 场景与真实物理场景之间的尺度比例因子$s$，帮助实现真实物理尺寸的三维数字化。

在研究当中，本文综合比较了几种主流 SFM 方

法[29-31], 发现[31]在多数情况下能够获得最好的效果。[31]采用一种局部增量捆绑调整策略，逐步减少匹配的误差，增强了算法的鲁棒性，因此在实验中本文基于[31]完成相机内外参数的标定。

(a) (b)

图 5. 围绕物体放置的标定物

### 4.3 深度估计

给定一组多角度的观察图像 $\{I_1 \ldots I_i \ldots I_n\}$ 和对应的相机投影矩阵 $\{P_1 \ldots P_i \ldots P_n\}$，多视角立体重建（Multi-view stereo）的目标是复原被测物体的三维表面 $S \subset \mathrm{R}^3$，可以看做是图像生成的逆问题。与体素方法[20, 21]相比，点云方法数据冗余较少，对内存的需求低，而与基于三角网格的方法[22, 23]相比，点云方法能够灵活的处理拓扑结构变化。基于点云的多视角重建方法可分为基于深度图的方法(depth map-based MVS) [16-19]和基于特征点的方法(feature-based MVS) [27]。 基于特征点的方法依赖场景包含丰富的纹理，在缺乏纹理的区域容易产生重建缺失，而基于深度图的方法在弱纹理表面有更好的效果，能够重建更完整稠密的表面，并且可以恢复物体表面的高频细节，增加物体表面的几何细节辨识度。文化遗址和文物的重建需要能够清晰复原几何细节，因为这些细节是这些杰出艺术品的精湛艺术性的体现，因此本文采用基于深度图的方法，以便重建模型完整且包含精细几何结构。 为了有效处理大尺度图像数据集，本文基于改进的 Gipuma 算法[17], 进行三维点云的重建。算法[17]包含两个基本步骤：基于 GPU 加速的 PatchMatch stereo 深度图生成，以及基于 GPU 的深度图的融合。

#### 4.3.1 基于 Normal-aware PatchMatch 的深度图估计

在多视角深度图估计（Multi-view depth estimation）阶段，目标是估计各个输入图像的对应深度图。在[17]中，除了可以估计深度图还可以估计法线图，实际是为每一个像素估计一个 3D 平面。具体地，以任意图像 $I_i$ 为参考图像， $I_j \in V(i)$ 为参考图像 $I_i$ 的邻域图像集中的任意图像。对于像素 $\boldsymbol{p} \in I_i$， 设它当前的深度假设值为 $d_p$ 与法线假设值为 $\boldsymbol{n}_p$， 经由深度 $d_p$ 和法线 $\boldsymbol{n}_p$ 所构成的平面 $\Pi$ 和相机参数 $P_i$, $P_j$ ,可计算 $\boldsymbol{p}$ 在邻域图像 $I_j$ 的对应点 $\boldsymbol{q}_j$。通过参考图像中以 $\boldsymbol{p}$ 为中心的 $r \times r$ 的图像块 $R(\boldsymbol{p})$ 与邻域图像 $I_j \in V(i)$ 的对应匹配图像块 $R(\boldsymbol{q}_j)$ 计算匹配代价选择最优的深度和法线假设。匹配相似性函数定义为： $\rho_j = \rho\left(R(\boldsymbol{p}), R(\boldsymbol{q}_j)\right)$。方法[17]采用了 intensity+gradient 的成本函数（公式（3）和公式（4））：

$$m_j(\boldsymbol{s},\boldsymbol{t}) = (1\text{-}\alpha)\cdot \min\left(\left\|I_i(\boldsymbol{s}) - I_j(\boldsymbol{t})\right\|, \tau_c\right) + \alpha \cdot \min\left(\left\|\nabla I_i(\boldsymbol{s}) - \nabla I_j(\boldsymbol{t})\right\|, \tau_g\right) \quad (3)$$

$$\rho_j = \sum_{s\in R_p, t\in R_{qj}} w_p^s m_j(\boldsymbol{s},\boldsymbol{t}) \quad (4)$$

其中 $\boldsymbol{s} \in R_p$ 和 $\boldsymbol{t} \in R_{q_j}$ 是参考图像和邻域图像的对应点, 参数 $\alpha$ 调节图像块的灰度值差异和梯度值差异的权重， $\tau_c$ 和 $\tau_g$ 为两个控制最大差异的常数。仿射权重函数定义为 $w_p^s = e^{-\frac{\|I(\boldsymbol{p})-I(\boldsymbol{s})\|_1}{\gamma}}$,其中 $\gamma$ 是参数，而 $\|I(\boldsymbol{p})-I(\boldsymbol{s})\|_1$ 计算 $I(\boldsymbol{p})$ 和 $I(\boldsymbol{s})$ 之间的 $L_1$-距离，其中 $\boldsymbol{s}$ 为像素位置 $\boldsymbol{p}$ 的邻域。 仿射权重 $w$ 减少远离中心像素的像素影响。

另外一种常用的匹配成本为自适应的零均值交叉相关 ANCC [32]，其定义如公式（5）：

$$\rho_j = 1\text{-}\frac{\sum_{s\in R_p, t\in R_{qj}} w_p^s w_{q_j}^t \left(I_i(\boldsymbol{s}) - \bar{A}_p\right)\left(I_j(\boldsymbol{t}) - \bar{A}_{q_j}\right)}{\sqrt{\sum_{s\in R_p}\left|w_p^s\left(I_i(\boldsymbol{s}) - \bar{A}_p\right)\right|^2}\sqrt{\sum_{t\in R_{q_j}}\left|w_{q_j}^t\left(I_j(\boldsymbol{t}) - \bar{A}_{q_j}\right)\right|^2}} \quad (5)$$

其中， $\bar{A}_p = \sum_{s\in R_p} w_p^s I(\boldsymbol{s})$ 和 $\bar{A}_{q_j} = \sum_{t\in R_{q_j}} w_{q_j}^t J(\boldsymbol{t})$ 是窗口 $R_p$ and $R_{q_j}$ 内图像灰度的加权均值。

对于采用转台和固定光源拍摄的室内文物场

景，本文观察到采用 intensity+gradient 成本可产生更平滑精细的重建效果，而对于室外场景，为了增加对于场景光照变化的鲁棒性，本文采用 ANCC 作为匹配成本。

为了避免遮挡的影响，选择匹配成本最小的 K 个邻域图像子集$V^*(i)$，则关于像素$\boldsymbol{p}$的深度$d_p$和法线$\boldsymbol{n}_p$的多视角累积成本如公式（6）：

$$g^*\left(d_p,\boldsymbol{n}_p\right)=\frac{1}{\left|V^*(i)\right|}\sum_{I_j\in V^*(i)}\rho_j \qquad (6)$$

对于参考图像$I$，邻域图像集合$V$的选择需要保证场景内容尽可能交叠的同时，使得图像之间的基线尽可能的大。对于采用转台拍摄的小体积文物场景，由于图像围绕物体成圆形分布，本文采用参考图像的视点方向$\boldsymbol{v}_i$与邻域图像的视点方向$\boldsymbol{v}_j$的角度差异来选择邻域图像：$V(i)\leftarrow\left\{v_j\,|\,\boldsymbol{v}_i\cdot\boldsymbol{v}_j>\cos(\tau)\right\}$，当角度差异的阈值$\tau$设置为 45 度时，选择的邻域图像的数量设置为 7。而对于无人机航拍的大佛造像场景，由于包含了不同距离下多次拍摄的数据，本文采用对尺度鲁棒的视点选择方法[18]来进行邻域视点的选择，邻域图像的最大数量设置为 9。

PatchMatch stereo 首先随机初始化深度和法线值，方法通过交替执行邻域传播(propagation)和精化（refinement）步骤来不断的优化深度和法线假设。在文献[17] 所提出的方法中， 采用一种基于 GPU 的红黑棋盘格传播算法，可以在整幅图像上并行的传播邻域的深度与法线假设。其中精化操作采用二分法，不断的在更小的区间内随机寻找更优的深度和法线。 然而相比深度，由于法线包含 3 个自由度，通过在三个分量上同时搜索不容易找到最优值。同时低精度的法线将导致重建的结果无法恢复文物表面的精细结构。为了提升法线的精度，本文根据前期的工作，采用一种 normal-aware 的 PatchMatch stereo 方法[19]，基于当前估计的深度图计算出新的法线假设，提升法线估计的准确度。

同时对于室内拍摄场景，在所有像素上进行深度精化是不必要的，因为部分场景实际对应的是背景区域，无需对其进行精细的估计。 为此，本文首先检测背景区域，并在此区域上取消精化操作。检测背景采用三组阈值实现，如公式（7）：

$$M(\boldsymbol{p})=\Lambda\left(I^r(\boldsymbol{p}),\ \text{蝌}\right)\cdot\Lambda\left(I^g(\boldsymbol{p}),\ {}_g\right)\cdot\Lambda\left(I^b(\boldsymbol{p}),\ ?_b\right) \qquad (7)$$

其中$I^r(\boldsymbol{p})$, $I^g(\boldsymbol{p})$和$I^b(\boldsymbol{p})$分别表示像素$I(\boldsymbol{p})$的 R, G, B 分量。$\Lambda(x,\ t)$是一个二值函数，当$x>=t$时值为 1，否则值为 0。

### 4.4 面向大尺度图像数据的增量深度图融合

当获得了一组输入图像的深度图$\{D_1\ldots D_i\ldots D_n\}$后，下一步的工作将是对深度图进行噪声的剔除和融合，生成三维点云。[17] 提出一种全局融合方法， 对于任意的深度值$d_p$和法线$\boldsymbol{n}_p$，检测其与邻域深度图的一致性，如公式（8）：

$$f_j(d_p,\boldsymbol{n}_p)=\Lambda\left(\left\|\boldsymbol{p}-\boldsymbol{p}'_j\right\|<\text{蝌}\cap\left|\boldsymbol{n}_{q_j}\cdot\boldsymbol{n}_p\right|<{}_n\right) \qquad (8)$$

其中$\boldsymbol{p}'_j$是$\boldsymbol{p}$经由邻域深度图$D_j$的重投影位置，其可通过将$\boldsymbol{p}$在邻域图像$I_j$的对应点$\boldsymbol{q}$经由邻域深度投影回$D_i$而得到。$\boldsymbol{n}_{qj}$是在像素位置$\boldsymbol{q}_j$处的法线值。参考图像上的深度和法线估计值将被认为是正确，仅当满足公式（9）：

$$\sum_{j\in\Omega(i)}f_j(d_p,\boldsymbol{n}_p)\geq\beta \qquad (9)$$

其中$\Omega(i)$为除了参考视角 $i$ 之外的所有邻域视角。$\beta$为预定义的阈值，反映视角一致性。假如深度值$d_p$和法线$\boldsymbol{n}_p$通过上述一致性检查过程，并且一致性邻域图像集为$\Omega'(i)\subset\Omega(i)$，则其对应的三维点$\boldsymbol{X}_p$与所有与该三维点一致的视点$j\in\Omega'(i)$中的三维点$\boldsymbol{X}_{qj}$平均，如公式（10）所示：

$$\bar{\boldsymbol{X}}_p = \frac{\left(\boldsymbol{X}_p + \sum_{j\in\Omega'(i)} \boldsymbol{X}_{qj}\right)}{1+\left|\Omega'(i)\right|} \quad (10)$$

由于上述融合方法的精确性和高效性，已经被广泛应用到最近的三维重建主流工作中。然而在上述实现过程中，方法[17] 需要将所有的深度图和法线图全部调入到显存中，然后逐个参考图像执行上述融合操作。其空间复杂度随着图像分辨率和图像集规模近似线性增长。高清图像的深度图和法线图的处理需要占用大量的显存，随着图像尺寸增大一倍，所需要显存将增大 4 倍，随着输入图像数量的增加，显存占用率也将迅速增加。然而，硬件资源，特别是显存资源是非常有限的，这限制了上述方法的应用范围。 为此本文提出一种增量式融合策略以减少对于显存的占用：

（1）设视角 $i$ 为参考视角，令$C$ 为与深度图同大小的一致性计数图像，初始值设置为 0.

（2）对于图像集中的除视角 $i$ 之外所有其它图像集 $\Omega$，每次增量的导入 α 张图像进入 GPU 的显存，执行跨视角一致性检查。若参考视点的深度 $d_p$ 和法线 $\boldsymbol{n}_p$ 在与这 α 张图像比较时，可通过公式（8）与公式（9）所确立的一致性判断，则：

(a) $\boldsymbol{p}$ 位置的一致性计数$C(\boldsymbol{p}) = C(\boldsymbol{p}) + 1$；

(b) 对应的点云 $\boldsymbol{X}(\boldsymbol{p})$ 执行增量累加操作 $\boldsymbol{X}_p = \boldsymbol{X}_p + \boldsymbol{X}_{qj}$。

当执行完 α 幅图像的视角一致性检查后，释放其显存，继续导入后续的 α 幅图像，直至处理完图像集 $\Omega$ 中所有图像。

（3）若已处理完 $\Omega$ 中的所有图像，则对于所有 $C(\boldsymbol{p}) \geq \tau$ 的像素位置，获得其融合三维点坐标 $\boldsymbol{X} = \boldsymbol{X}_p / C(\boldsymbol{p})$。

### 4.5 表面重建

给定一组三维点云，表面重建算法用于从点云重建三角网格表面（mesh）。本文基于 Screened Poisson surface reconstruction (SPSR) [33]进行表面重建，并剪枝掉三角形面积过大的区域。经由 SPSR 获取的模型包含数十万到数百万的三角面片，如此稠密的模型将会限制在虚拟环境下的渲染效率。为此需要在不损失模型质量的情况下尽可能减少面片的数量。本文对 SPSR 输出的三维网格表面， 使用 Qslim 算法 [34] 进行必要的网格精简，从而保证简化的三维表面包含的三角面片数量在十万面以下，同时保留物体表面的几何结构。

## 5 实验结果

本文实验平台为联想 P500 工作站，搭载 CPU E5-1620v3 (3.5Hz), 64GB 内存，Nvidia GTX1080 显卡。 固原须弥山石窟包括大量与丝绸之路文化相关的珍贵遗址，如须弥山大佛（第五窟），石门关和相国寺等，其中须弥山大佛是该景区的代表建筑，始凿于北魏孝文帝太和年间（公元 477 至 499 年），高 20.6 米。针对须弥山大佛采集的部分航拍图像如图 6 所示。经过 SFM 标定相机参数后，重建的稀疏点云和相机位姿如图 7 所示。 经由本文多视角立体重建方法实现的须弥山大佛重建结果如图 8 所示。可以看出本文方法基于高清航拍数据，在保证三维形状完整重建的同时可复原表面精细几何结构，如大佛的脸部和衣服的褶皱等。

固原地区馆藏文物的部分多角度采集图像如图 9 所示。基于这些图像数据，实现的三维重建结果如图 10 所示。可以看出重建模型忠实复原文物的几何形状，模型中几何细节清晰可见。 同时本文验证 normal-aware 的 PatchMatch stereo 方法在法线恢复方面的能力。如图 11 所示，相比基准方法，

图 6. 须弥山大佛部分航拍图像

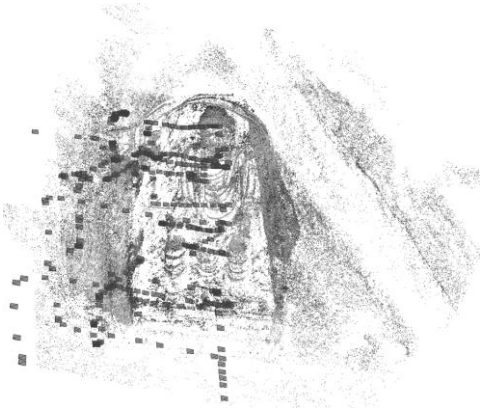


图 7. 经由 SFM 重建的须弥山大佛稀疏点云和航拍相机参数

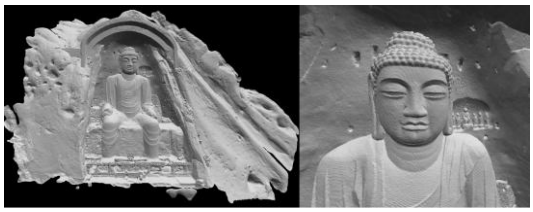


图 8. 须弥山大佛重建效果图。从左至右，依次是全局重建效果展示和局部几何细节展示。

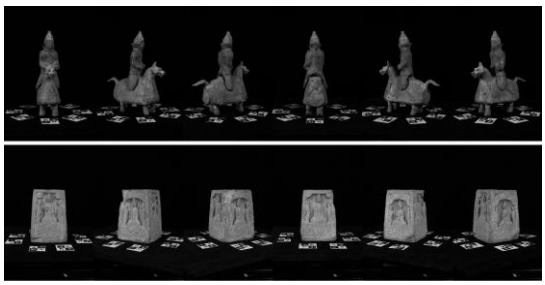


图 9. 固原馆藏文物部分多视角图像

normal-aware PatchMatch stereo 可以更精准的恢复表面的法线信息。

本文进一步在公共数据集 ETH3D 上评测本文提出的重建算法的有效性。ETH3D [15]包含了室内外场景的高清图像数据集，原始图像大小为 6,048×4,032，这些高分辨率图像能够较好地验证本文提出的面向高分辨率图像三维重建算法的有效性。ETH3D 的真值数据使用专业的激光扫描仪获得，经由 MVS 方法重建的点云可与真值点云数据进行量化的比对，评价方法的有效性。本文在其 13 个训练集上进行量化实验。

首先本文验证提出的增量融合的有效性，由于在高分辨率图像的数据集上，全局方法将在一些图片数量较多的数据上导致显存不足的问题，因此本文将图像降采样到 1,600×1,064，在相同的深度图计算方法的基础上，分别采用全局融合和提出的增量融合进行深度图的融合。 ETH3D 采用 $F_1$ 度量重建的综合质量，数值越高，质量越好。在实验中发现，全局方法随着图像集规模的增加，显存占用增加，最小显存占用为 1828.5 MB（14 张图像），最大显存的占用率为 4429.0 MB（76 张图像）。可以预见，如果将输入图像大小增加到 3,200×2,128, 则显存将增加到 16GB 以上，由此限制了方法对于大尺度数据的应用。而本文提出的增量方法可将最大显存占用率控制在 1821MB 以内（α=10），因此可以在有限的显存下处理更多的数据。通过表 1 的量化质量评价可以看到，提出的增量方法的重建质量基本和全局方法一致。以上定量实验证明了增量融合方法的有效性。 虽然通过增加显卡的数量可增强全局融合方法的适应性，但是全局方法的显存瓶颈将很快限制它在更大规模数据的应用。例如，对于大佛数据，由于包含大量的大尺寸深度图，全局方法在两块显卡上也将导致显存不足，无法完成重建，而本文提出的增量融合方法可以在单个 8G 显存的显卡上产生理想的融合结果。

(a) (b)

图 10. (a) 甲骑具装俑和 (b) 石雕佛造像塔点云模型。

为了验证本文提出的重建算法的整体有效性，本文在 ETH3D 数据集上与当前主流方法进行了对比，包括 Gipuma [17], PMVS [27]，COLMAP [16]，CMPMVS[35]和 ACMH [28]。需要指出的是，这些方法和本文一样都是不包含正则化项的方法。一些包含正则化项的方法虽然可提升重建在弱纹理区域的完整度，但是也会导致物体表面细节被抹去，并且显著增加计算时间。文物和古迹数据的纹理比较丰富，且细节是其艺术价值的体现，因此非正则化的方法更能适应这一特定应用。在表 2 中本文给出了量化评价的结果。 ETH3D 高分辨数据集鼓励高清图像进行重建，如 ACMH [28]和 COLMAP[16]

将图像缩放为 3,200×2,128 大小进行重建。对于本文提出的方法，本文分别给出了低分辨率版本（1,600×1,064）和高分辨率版本（3,200×2,128）的重建结果。从表 2 可以看出，本文的高分辨率版本可以产生最好的重建结果，而低分辨率版本也在多数指标上优于主流算法，由此验证了方法的有效性。一般来说，较高的图像分辨率对计算机显卡的性能有更高的要求，如 ACMH 采用了 2 块 GTX Titan X，总共 24GB 显存， 而本文的计算机仅包含一个 GTX 1080 显卡（8GB 显存）。 图 12 给出了本文方法（高分辨率版本）和对比方法在 ETH3D 数据集上的定性(主观)评价结果，可以看出本文方法明显提升了重建的完整度。值得指出的是，表 2 的结果还说明了，通过在高分率图像上执行重建，可以提升重建质量，验证了本文提出的面向高分辨率图像三维重建方法的潜在应用价值。

(a) (b)

图 11. Normal-aware 重建效果展示。(a) 基准方法的法线图；(b) Normal-aware PathMatch 估计的法线图。

本文的重建算法主要包含 2 部分，深度图估计和增量深度融合。以 ETH3D 的 *relief* 数据集为例，其包含 31 张图像，本文采用 1,600×1,064 分辨率为输入，计算每一视角图像的深度图的平均计算时间为 11.6 秒，采用全局深度融合的时间为 58.4 秒，采用本文增量融合（α=15）的时间为 68.2 秒，采用本文增量融合（α=10）的时间为 69.6 秒。以 ETH3D 的 *electro* 为例, 其包含 45 张多视角图像， 计算每一图像深度的平均时间为：11.2 秒，采用全局深度融合的时间为 80.6 秒， 采用本文增量融合（α=15）的时间为 99.6 秒，本文增量融合（α=10）的时间为 113.4 秒。本文增量融合相比全局融合算法时间开销略多，分析原因为数据从 CPU 内存到 GPU 内存的传输带来了额外的时间开销。整体来说，本文提出的重建算法具有较高的计算效率。相比全局融合算法，本文提出的增量融合算法在减少显存的同时，并没有显著增加计算时间，保证了算法的执行效率和实用性。

最后，本文在文物数据“甲骑具装俑”上对比了本文方法和著名的 PMVS 算法[27]的重建效果，如图 13 所示。可以看出，本文方法不仅产生更平滑的表面，同时还有效复原了文物表面的精细结构，验证了方法的优越性。重建的模型经过表面重建，可导入到虚拟互动体验平台，进行内容展示和人机互动体验，如图 14 所示。

**表 1 在 ETH3D 上评测不同融合方法的显卡显存消耗（MB）与重建质量（$F_1$）。评估阈值为 0.02。**

| 指标 | 全局融合 | 增量融合 (α=20) | 增量融合 (α=15) | 增量融合 (α=10) |
|---|---|---|---|---|
| 重建质量 $F_1$ (%) | 72.88 | 72.75 | 72.78 | 72.63 |
| 最大显存 (MB) | 4429 | 2400 | 2018 | 1821 |


(a) (b) (c) (d)

图 12 ETH3D 数据集的定性评估结果。(a)某一视角的输入图像；(b) COLMAP[19]重建结果；(c) ACMH[31]重建结果；(d)本文方法重建结果。相比对比方法，可以看出本文方法明显提升了重建的完整度。

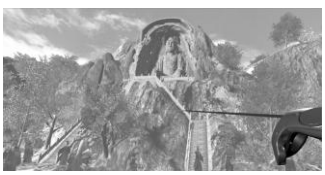
表 2 在 ETH3D 对比本文方法和主流 MVS 重建算法。其中本文方法分别给出了低分辨率和高分辨率的重建结果。其中加粗为最好的结果，下划线为第二好的结果。

| 方法/阈值 | 0.01 | 0.02 | 0.05 | 0.1 | 0.2 |
|---|---|---|---|---|---|
| Gipuma | 25.87 | 36.38 | 49.17 | 58.52 | 68.35 |
| PMVS | 33.81 | 46.06 | 55.81 | 61.27 | 67.27 |
| COLMAP | 51.99 | 67.66 | 80.50 | 87.61 | **93.27** |
| CMPMVS | 49.24 | 62.49 | 73.84 | 79.86 | 84.76 |
| ACMH | 58.24 | 70.71 | 81.86 | 87.31 | 91.51 |
| Ours(low-res) | 56.97 | 72.78 | 84.16 | 89.27 | 92.90 |
| Ours(high-res) | **64.86** | **75.43** | **84.79** | **89.52** | 93.02 |

(a)

(b)

图 13. 在甲骑具装俑文物数据上，对比本文方法(a)与 PMVS [27] (b)。可以看出，本文方法不仅可产生更平滑的表面，同时保留表面的细节，更真实地复原了文物的三维模型。



(a) (b)

图 14 重建模型集成到沉浸式交互虚拟漫游系统。

## 6 结论

丝路文化传承等文化科技领域对虚拟现实互动体验系统的需求正在快速增长，本文基于虚拟现实技术，围绕固原丝路文化遗产，通过高清图像数据采集与面向大尺度图像数据的三维重建，提供了一个高真实感、沉浸感和多种交互方式的虚拟现实互动环境，可以为公众提供全新的交互体验方式，打造动静相宜、内容丰富的互动漫游体验效果，使人们能够身临其境般地感受中华传统文化的博大内涵，提升丝路文化遗产的推广、普及和数字化服务能力。

此外提出的整个系统框架，特别是三维重建方法，还可潜在应用于其它面向文化保护和传播的 VR 漫游系统中，如虚拟博物馆等。

**致 谢** 感谢国家自然科学基金(No. 61702482 和 61532002)，以及北京市自然基金（No. L172049）对本研究的资助。

## 参 考 文 献

[1] Chen Ying. The application of virtual reality technology in the Inheritance and dissemination of non heritage culture. Study on Natural and Cultural Heritage, 2017,(5):146-148 (in Chinese)
(陈颖.虚拟现实技术在非遗文化传承与传播中的应用.遗产与保护研究,2017,(5):146-148).

[2] Pan Yun-He, Lu Dong-Ming. Digital protection and restoration of dunhuang mural. Journal of System Simulation, 2003, 15(3):310-314 (in Chinese)
(潘云鹤, 鲁东明. 古代敦煌壁画的数字化保护与修复. 系统仿真学报, 2003, 15(3):310-314).

[3] Liu Jian. Research on virtual exhibition system for hemudu site museum. Journal of System Simulation, 2017,(5):146-148 (in Chinese)
(刘箴. 河姆渡遗址博物馆虚拟展示系统的研究. 系统仿真学报, 2009, 21(7):1945-1949).

[4] Wang Sheng-Hua, Tan Jian. Chinese shadow play simulation based on volume visualization. Journal of System Simulation, 2015,27(09):2126-2134 (in Chinese)
(王圣华,谭剑.中国皮影体可视化仿真方法研究.系统仿真学报,2015,27(09):2126-2134.)

[5] Liu Shi-Guang, Chen Di. Simulation technique for marbled paper patterns. Journal of Software, 2012, 23(Suppl.(2)):1-7 (in Chinese)
(刘世光,陈迪.一种石纹纸染艺术图案仿真技术.软件学报, 2012, 23(Suppl.(2)):1-7.)

[6] Zhang Xian-Quan, Yu Jin-Hui, Jiang, Ling-Lin, Tao Xiao-Mei. Computer assisted generation of paper cut-out images. Journal of Computer-Aided Design & Computer Graphics, 2005, 17(6):1378-1382 (in Chinese)

(张显全, 于金辉, 蒋凌琳, 陶小梅. 计算机辅助生成剪纸形象. 计算机辅助设计与图形学学报, 2005, 17(6):1378-1382.)

[7] Tu Chuan-Peng, Peng Ren, Chen Hai-Ying. Computer generation of water animation with the style of paper-cuts. Journal of Computer-Aided Design & Computer Graphics, 2009, 21(7):949-953 (in Chinese)

(涂传朋, 彭韧, 陈海英, 于金辉. 计算机生成剪纸风格流水动画. 计算机辅助设计与图形学学报, 2009, 21(7):949-953.)

[8] Liu Xiao, Wu Xun-Wei. Computer-rebuilding damaged cultural relics in three dimensions. Journal of Electronics & Information Technology, 2001,23(7):650-656 (in Chinese)

(刘晓，吴训威．破损古文物的计算机三维重构．电子与信息学报，2001,23(7):650-656.)

[9] Hu Yan-Qiu, Yu Jin-Hui, Jiang Wei, Peng Ren. Modeling and rendering of bronze articles. Journal of Computer-Aided Design & Computer Graphics, 2008, 20(9):1140-1145 (in Chinese)

(胡晏秋, 于金辉, 姜威, 彭韧. 面向青铜器的建模与绘制. 计算机辅助设计与图形学学报, 2008, 20(9):1140-1145.)

[10]Kong Li-Ming, Rong Xiao-Man, The review of application of augmented Reality in culture presentation. China Cultural Heritage, 2017(02):62-69 (in Chinese)

(孔黎明,荣晓曼.增强现实技术在文化遗产展示中应用综评.中国文化遗产,2017(02):62-69.)

[11]Shi-Guo-Wei, Wang Yong-Tian, Liu Yue, Zheng Wei. Digital conservation of cultural heritage using augmented reality. Journal of System Simulation, 2009, 21(7):2090-2093. (in Chinese)

(师国伟, 王涌天, 刘越, 郑伟. 增强现实技术在文化遗产数字化保护中的应用. 系统仿真学报, 2009, 21(7):2090-2093.)

[12]Levoy M., Pulli K., Curless B. et al. The Digital michelangelo project: 3D scanning of large statues//Proceedings of the Siggraph. New Orleans, Louisiana, USA, 2000: 1-14.

[13]Seitz S., Curless B., Diebel J., Scharstein D., Szeliski R. A comparison and evaluation of multi-view stereo reconstruction algorithms//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. New York, USA, 2006: 519-526.

[14]Strecha C., Hansen W. von, Van Gool L., Fua P., Thoennessen U., On benchmarking camera calibration and multiview stereo for high resolution imagery//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Anchorage, USA , 2008: 1-8.

[15]Schöps T., L. Schönberger J., Galliani S., Sattler T., Schindler K., Pollefeys M., Geiger A., A multi-view stereo benchmark with high-resolution images and multi-camera videos// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, USA , 2017:1-8.

[16]Schönberger J. L., Zheng E., Pollefeys M., Frahm J. M., Pixelwise view selection for unstructured multi-view stereo// Proceedings of European Conference on Computer Vision. Amsterdam, The Netherlands, 2016: 1-17.

[17]Galliani S., Lasinger K., Schindler K.. Massively Parallel Multiview Stereopsis by Surface Normal Diffusion//Proceedings of the IEEE International Conference on Computer Vision. Santiago, Chile, 2015:873-881.

[18]Bailer C., Finckh M., Lensch H. P. A., Scale robust multiview stereo// Proceedings of European Conference on Computer Vision. Florence, Italy, 2012:1-14.

[19]Li, Z., Zuo W., Wang Z., Zhang L., Confidence-based large-scale dense multi-view stereo. IEEE Transactions on Image Processing, 2020，29: 7176-7191

[20]Kostrikov I., Horbert E., Leibe B., Probabilistic labeling cost for high-accuracy multi-view reconstruction//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Columbus, USA , 2014：1-8

[21]Hane C., Zach C., Cohen A., Pollefeys M., Dense semantic 3D reconstruction. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017，39(9): 1730-1743

[22]Delaunoy A., Prados E., Gradient flows for optimizing triangular mesh-based surfaces: applications to 3D reconstruction problems dealing with visibility. International Journal of Computer Vision, 2011, 95(2):100-123.

[23]Li Z., Wang K., Zuo W., Meng D., Zhang L., Detail-preserving and content-aware variational multi-view stereo Reconstruction. IEEE Transactions on Image Processing, 2016, 25(2): 864-877

[24]Yao, Y., Luo Z., Li S., Shen T., Fang T., Quan L., Recurrent mvsnet for high-resolution multi-view stereo depth inference //Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Long Beach, USA , 2019:5520-5529.

[25]Im S., Jeon H., Lin S., Kweon I., Dpsnet: End-to-end deep plane sweep stereo//Proceedings of International Conference on Learning Representations. New Orleans, USA, 2019: 1-12.

[26]Furukawa Y., Hernndez C., Multi-view stereo: a tutorial. Foundations and Trends in Computer Graphics and Vision, 2015, 9(6): 1-148

[27]Furukawa Y., Ponce J., Accurate, dense, and robust multiview stereopsis. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2008，32(8): 1362-1376

[28]Xu Q., Tao W., Multi-scale geometric consistency guided multiview stereo//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Long Beach, USA , 2019: 5483–5492.

[29]Snavely N., Seitz S. M., Szeliski R., Photo Tourism: Exploring image collections in 3D. ACM Transactions on Graphics, 2006(1): 835-846.

[30]Wu C., Towards Linear-time incremental structure from motion//Proceedings of the 3D computer vision. Seattle, USA, 2013:127-134.

[31]Johannes Lutz S., Jan-Michael F., Structure-from-motion revisited//Proceedings of the IEEE Conference on Computer Vision and

Pattern Recognition. Las Vegas, USA, 2016:4104-4113.

[32]Heo Y., Lee K., Lee S., Robust stereo matching using adaptive normalized cross-correlation. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2011, 33(4): 807-822

[33]Kazhdan M., Hoppe H.. Screened poisson surface reconstruction. ACM Trans. Graphics, 2013, 32(3):1-13.

[34]Garland M., Heckbert P. S., Surface simplification using quadric error metrics//Proceedings of the ACM SIGGRAPH. New York, USA, 1997: 209-216.

[35]Jancosek M., Pajdla T., Multi-view reconstruction preserving weakly-supported surfaces// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Colorado, USA:Springs 2011: 3121–3128.

**Li Zhaoxin**, Ph.D. ，assistant Professor. His research interests include 3D reconstruction and 3D computer vision

**Jiang Hao**, Ph.D., associate professor.. His research interest is crowd simulation.

**Liu Yanqing**, Ph.D., professor. Her research interest is Guyuan historic culture.

**Wang Zhaoqi**, Ph.D., professor . His research interest is virtual reality.

**Background**

This research focuses on 3D reconstruction and visualization of the historical sites and cultural relics. It is overlapping between cultural heritage protection and computer science. Silk Road culture is an important link in the Belt and Road strategy. Its heritage is of great significance. However, due to historical and geographical reasons, the representative historical heritage in the Silk Road culture is scattered or damaged, and it is difficult to present it effectively. Through historical restoration and image-based 3D reconstruction, we restored the historical sites, cultural relics and events of Guyuan, one of the important nodes in Silk Road Culture. Especially, we first reconstruct the historical sites and cultural relics, and then visualize them in the virtual reality application. The 3D reconstruction of the historical sites and cultural relics is an important research area in computer vision, architecture and archaeology. The traditional 3D reconstruction methods for the historical sites and cultural relics rely on laser scanner, which is bulky and expensive. In this paper, we propose an effective and low-cost 3D reconstruction method based multi-view stereo. The proposed method fully explores multi-view high-resolution images and can generate the high-quality 3D surface with detailed geometric features for both outdoor large-scale historical sites and indoor small-scale cultural relics. We also propose and implement a virtual reality platform for the Silk Road Cultural Heritage.

The proposed MVS 3D reconstruction method for high-resolution images, which consists of a normal-aware PatchMatch for the high-quality normal recovery to represent the detailed surface of the cultural relics, and a GPU-friendly incremental depth fusion method which can fuse a large amount of depth maps by a small size of GPU memory. The experiment results on the public MVS benchmark and our captured datasets highlight that the proposed method can recover the detailed surfaces while keeping a good scalability for the large-scale data. The reconstructed models can effectively support VR application, playing a positive role in the dissemination of Silk Road culture.

This work was supported by National Natural Science Foundation of China under Grants No. 61702482. This NSFC project mainly focus on the complete and detailed 3D reconstruction based multi-view stereo. The research team has worked on 3D reconstruction research for more than 10 years and has already built a complete pipeline for the entire reconstruction procedure, including depth map estimation, volumetric fusion and mesh-based refinement. The related publications include IEEE Trans. on Image processing, Image and Vision Computing, Neurocomputing and The Visual Computer.