

适用于基于图像的纹理映射的基于块的最优化技术 (Patch-Based Optimization for Image-Based Texture Mapping)

SIGGRAPH 2017

SAIBI, NIMA KHADEMI KALANTARI, RAVI RAMAMOORTHY (加州大学圣地亚哥分校)

----Translated by 罗辉翔 of FDU for CG PJ

项目地址: <http://cseweb.ucsd.edu/~viscomp/projects/SIG17TextureMapping/>

摘要

在为真实世界中物体的几何模型提供纹理映射方面，基于图像的纹理映射是一种常用的手段。尽管利用经过标定的相机及精确的几何结构可以很容易地算出高质量的纹理映射，这种质量会随着不确定性的提高而显著地降低。在本论文中，我们通过提出一种别致的，全局的，基于块的最优化系统来合成配准的图像。特别地，我们使用了基于块的合成，通过从原图中抽取信息，来重建一个经配准后光照一致的图像集。该最优化系统不仅简便灵活，且相比其他技术（比如局部弯曲）而言，在矫正大范围图像失配上更为合适。该最优化问题包括两大步骤，一是搜索块与投票环节，二是重建。实验结果表明：对于用消费级深度相机（如 Intel RealSense）扫描的物体来说，我们的方法比之现有方法能够提供更高质量的纹理映射。此外，我们证实了该系统可用于纹理编辑的相关工作，比如补洞、重组、多视角物体隐藏。



(左图表示较为精确的几何结构，右图是不太精确的几何结构)

CCS 分类：计算方法论->计算摄影学

附加关键词句：基于图像的纹理映射，基于块的合成

正文

一、引入

对真实世界景物建模是一个重要的视觉任务，在视频游戏、VR、动画设计等方面有着广泛的应用。基于几何结构的重建是大量相关研究的主题，许多相应的算法被研发出来。随着消费级深度相机的普及，普通消费者也能用诸如 KinectFusion 的技术来生成物体的几何模型。

尽管如此，重现真实世界物体的全貌仍需要重建高质量的纹理映射。基于图像的纹理映射，是在一个采集自不同视角的图像集上，构建与视角无关的纹理映射的常用的方式。随之而来的挑战是，几何结构和相机摆放姿态的计算，由于常常受到噪声的影响而不准确；此外，消费级深度相机提供的 RGB 图像，常常会有无法被相机成像模型解释的失真。因此，朴素的投影、融合输入图像就会有模糊的、带重影的人工痕迹，如图 2 所示。

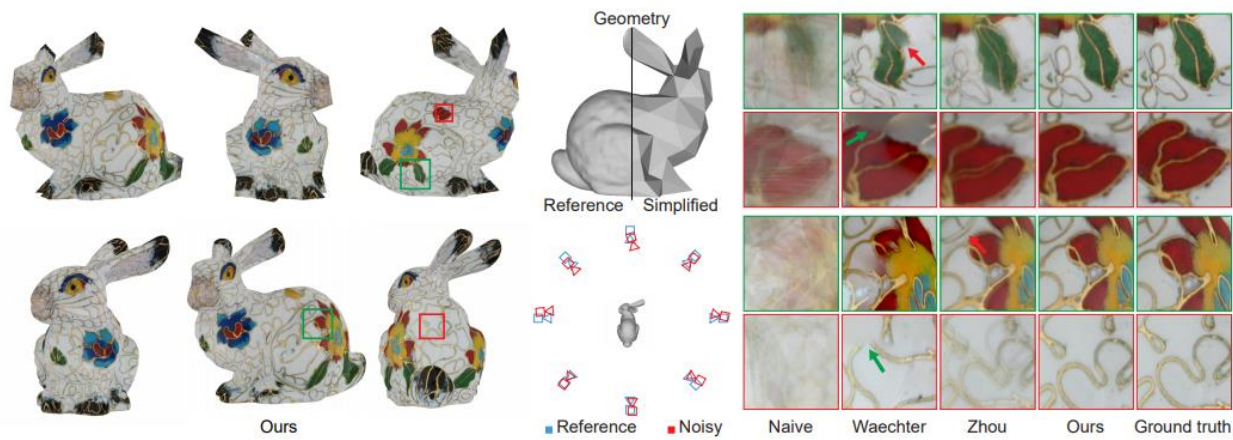


图 2

我们发现，通过对输入的每张图像生成一张配准的图像可以克服绝大多数的不精确问题。我们的方法建立在 Zhou 和 Koltun 的工作之上，他们的工作使用了局部扭曲技术来校正错误配准问题。虽然他们的方法能处理细小的不精确问题，但是对于很不精确的情况很难提供高质量的结果；并且由于局部扭曲技术在校正失配时的局限性能，他们的方法会遗失部分几何特征（见图 1, 2 和 4）。



图 1

受近来基于块的方法在图像和视频编辑任务领域获得成功的影响，我们提出了一种新颖的，全局的，基于块的最优化系统来合成配准好的图像。我们的能量函数（这里的能量函数应该属于“基于能量最优化的图像匹配”这一方法中的技术，参见参考博客①）结合了我们配准后图像的两个主要的需求：（1）包含原始输入图像的绝大部分信息；（2）保留投影矩阵的光学一致性。

为进一步优化能量函数，我们同时将输入的原图像和配准后的图像的局部相似性最大化，并保证了所有配准后的图像与纹理映射的一致性。

我们的系统以基于块的方式直接从原图像中抽取信息，从而可以灵活地处理大范围的不精确问题。此外，我们的方法通过合成缺失的内容来处理缺失几何特征的情形（见图 4），而现有的基于扭曲图像的或是基于图片分割的技术则不能处理。最后，相比于 Zhou 和 Koltun 的方法，我们实现了图像领域的一种优化，使得我们系统性能不受几何复杂性所限。总之，我们做出了以下的贡献：

- （1）对于视点无关的、基于图像的纹理映射，引入了首个基于块的优化系统（见模块 3.1）。该方法通过合成经过配准的图像来校正图像失配问题，而这些经过校正的图像之后能用于生成一个视点独立的纹理映射。
- （2）我们提出了一个简单的两部曲操作来高效地求解我们的能量方程（见模块 3.2）。
- （3）我们证实了本方法比现有的技术效果更好（见模块 5）；此外我们也展现了本方法在其他领域的应用，这些应用也获得了靠当前技术无法实现的效果。

二、相关工作

从真实世界物体的图像集合中重现它的外貌这个主题被广泛地研究。基于图像的着色技术通过产生一个依赖于视角的纹理映射来重现物体的外貌。然而，这种手段只能应对输入图像在相同光照条件下的情况。因此，它们不能被用于某些场景，这些场景使用的是在不同光照环境下被扫描过的物体的图片。此外，由于这些手段并不产生一个全局一致的纹理映射，他们显然无法用于游戏、AR、动画等领域。

视角独立的纹理映射方法，以我们的为例，从不同视角抓取到的图像集中生成一个全局一致的纹理映射，这些图像随后可以被用来在各种光照情况下上色（注意：该情况下最终的纹理仍具有原始的光照条件。然而，该问题可以通过对原图做本征图像分解（Intrinsic Image Decomposition，参见博客②）并使用反照率来生成纹理映射）。这类方法的主要挑战是，在抓取过程中如何解决不精确问题。关于图像的几何配准，有如下几种方法：半自动化的[Franken et al. 2005; Ofek et al. 1997; Pighin et al. 1998]，或者通过优化颜色一致性而自动化的[Bernardini et al. 2001; Pulli and Shapiro 2000]，或者通过配准图像和几何特征来自动化的[Lensch et al. 2001; Stamos

and Allen 2002], 或者通过最大化投影图像间的互信息来自动化[Corsini et al. 2013, 2009]。虽然这些方法在解决由相机标定不精确引起的问题时很有效, 却不能解决几何结构不精确, 或是 RGB 图像中光学畸变引起的问题, 而这些问题在消费级深度相机中是很常见的问题。

2.1 单视角选择

这类方法不去融合投影后的输入图像, 由于失配的缘故这种操作可能产生模糊的效果; 而是对每个面只选择一个视角。为避免在面与面之间产生可见的缝隙, 通常要解决离散标记的问题 ([Lempitsky and Ivanov 2007; Sinha et al. 2008; Velho and Sossai Jr. 2007; Waechter et al. 2014])。

例如, 当下效果最好的方法 (Waechter et al. [2014]) 解决了条件随机场能量方程的问题, 该方程包括两项: 一个数据项“偏好”与当前的面接近的、不模糊的面, 一个平滑项“惩罚”相邻面不连续的情况 (是不是把参考博客①中的像素改成了面呢?)。可是正如图 2 所示, 即使是这种方法也无法解决在不精确程度很大时许多挑战性的情况, 从而在最终的纹理映射中产生可见的缝隙。

2.2 图像配准

本类别中的方法直接通过配准输入图像来解决不精确问题。Tzur and Tal [2009]提出估计几何结构的每个顶点的局部的相机投影来解决相机标定、几何结构等的精确问题。然而, 该方法需要用户的交互, 来产生看上去合理的结果。Aganj et al. [2010]通过寻找不同视角中的 SIFT 特征、扭曲输入图像来解决失配问题, 而[Dellepiane et al. 2012; Eisemann et al. 2008]则用光流处理处理扭曲。这些方法并未要求失真全局最小化, 且在一对图片上进行处理, 因而是次优解。Gal et al. [2010]给一张输入图像指定一个三角形, 找到每个三角形的最优转换 (方式/矩阵) 来消除缝隙, 然而这种优化方式需要大量的运算。

我们的方法是基于 Zhou and Koltun [2014]近期的工作的, 该工作同时解决了寻找最优的相机姿态和非刚性的对输入图像的矫正。它们使用了局部扭曲来实现非刚性的配准, 并提出一个交互的优化来最小化它们的目标函数。然而, 局部扭曲并不能矫正大规模的失配, 而且会产生重影、模糊的人工痕迹, 如图 2 所示。为避免该问题, 我们提出了一种相比于局部扭曲更为灵活的机制来处理非刚性的配准, 将此机制用于我们的优化系统。

2.3 基于块的合成

我们的方法受近来基于块的合成方法在大量应用, 例如填洞 ([Wexler et al. 2007])、图像重定向 (image retargeting, 参见博客③) 与图像编辑 ([Barnes et al. 2009; Simakov et al. 2008]), 影像变形 (参见资料⑨), HDR 重建 ([Kalantari et al. 2013; Sen et al. 2012]) 和风格迁移 ([Bénard et al. 2013; Jamriška et al. 2015])。基于块的合成, 在找出几张图像的一致性 (比如影像变形和 HDR 重建) 很困难的应用场景中已经特别成功了。在我们的应用中, 合成的配准图像需要在物体的几

何结构上具有一致性，因此基于块的合成在我们的问题上无法直接应用。我们通过提出一个新颖的，基于块的能量方程来应对这个挑战，该方程将几何结构融合到公式之中。

三、算法

绝大多数纹理映射的生成方式是使用一个包含 N 张源图片(记作 S_1, \dots, S_N)的集合来生成一个高质量的视角独立的纹理映射，这些源图片从不同的视角中拍摄得到。这些方法常常假设源图片对应的物体的近似的几何结构和粗略的相机姿态（即相机的内参、外参矩阵）用现有的方法可以估计得到（[Newcombe et al. 2011--获取相机参数; Seitz et al. 2006--三维重建物体]；一个很有意思的论文是只需要给两张相关照片就能进行三维重建：Novel View Synthesis in Tensor Space），一旦纹理映射产生，有着视角独立的纹理的物体就可以在任何新视角中着色了。

一种生成纹理映射的简单方式是将原图像投影到几何结构上，并将所有的投影后的图像结合在一起。理想情况下，这些投影后的图像是光照一致的，因此融合这些照片能产生高质量的纹理图片。然而实际上由于不精确的问题，投影图像常常失配。因此这种简单的做法会产生重影的效果。

在图 3 中的顶行我们演示了该做法，可以看到给定源图片 S_1, S_2 ，展现出的合成后的效果。为了观察失配问题，我们将原图像投影到一个新的视角 i 。注意从原图像 S_j 投影到新视角 i 可通过重映射原图像的像素颜色 $S_j(y)$ 来实现。 y 是像素从图像 i 投影到 j 的位置，记作： $y = P_j(G_i(x))$ 。 x 是图像 i 上像素的位置， G_i 将 i 图像上的一个像素投影到全局三维空间， P_j 将三维空间中的点投影到 j 图像。在本论文中，为了符号的清晰简洁考虑，我们使用 x_i 表示图像 i 上的像素，用 $x_{i \rightarrow j}$ 表示这个像素投影到图像 j 后的新像素。在该定义下， $y = x_{i \rightarrow j}$ ， $S_j(x_{i \rightarrow j})$ 是将图像 S_j 投影到视角 i 的结果。查表 1 可得本论文中用到的完整的符号表。

S_1, \dots, S_N	source images (input)
T_1, \dots, T_N	target (aligned) images (output)
M_1, \dots, M_N	texture at different views (output)
x_i	pixel position on image i
$x_{i \rightarrow j}$	pixel position projected from image i to j
$T_j(x_{i \rightarrow j})$	RGB color of the j^{th} target image at pixel $x_{i \rightarrow j}$, i.e., the result of projecting target j to camera i

Table 1. Notation used in the paper.

正如图 3（顶行）所示，由于估计的几何结构与相机姿态（即相机内外参）的不精确，投影后的源图片 $S_j(x_{i \rightarrow 1})$ 和 $S_j(x_{i \rightarrow 2})$ 都失配了。因此，由简单的投影和混合方法产生的纹理映射有着重影的效果（如最右边那一列所示）。这里（图中）的 M_i 指的是在 i 相机视角下，最终的全局一致的纹理映射。注意 M_j 是从所有的源图像中重建得到的映射，因此与投影后的源图像是不同的。

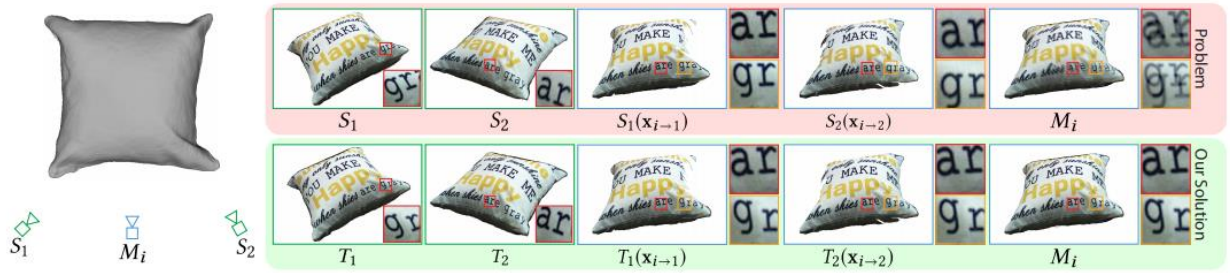


图 3

为了克服这种失配问题，我们为每张源图像 S_i 合成匹配后的（目标）图像 T_i 。正如图 3 所示，目标图像是通过移动源图像的内容来矫正失配问题。作为结果，所有的目标图像都是光照一致的，因此将它们投影到几何结构上并结合在一起产生了高质量的结果。下个章节将解释我们的基于块的优化系统如何合成这些目标图像。

3.1 基于块的能量函数

我们观察到，为了产生高质量的纹理映射，目标图像应当有两个主要特性：

- 1) 每张特征图像应该与它的相应的源图像相似；
- 2) 投影后的目标图像应当光照一致。

我们的目标是提出一个全局能量函数，能兼顾这两个主要特性。为满足第一个特性，我们确保了在视觉一致性的角度上，每张特征图像都包含了相应的源图像的绝大多数的信息。为了实现它，我们使用双向相似性(BDS)(见 Simakov et al. [2008])。基于块的能量函数定义为：

$$E_{\text{BDS}}(S, T) = \underbrace{\frac{1}{L} \left(\sum_{s \in S} \min_{t \in T} D(s, t) \right)}_{\text{completeness}} + \alpha \underbrace{\sum_{t \in T} \min_{s \in S} D(s, t)}_{\text{coherence}},$$

α 是前后两项的影响因素占比的调整参数， s 和 t 则分别是源图像和目标图像的块， D 是在 RGB 彩色空间上块 s 和 t 的所有像素值平方差之和。此外， L 是每个块的像素数目，例如对于 7×7 的块来说 $L = 49$ 。

公式中第一项（完整性）确保了每个源块在目标图像中都有与之相似的块，第二项（一致性）的功能反过来与之类似。完整性这一项度量了目标图像中保留了多少的源图像信息，而一致性这一项则度量了目标图像中是否存在着一些新的视觉结构（人工痕迹）。最小化能量函数保证了绝大多数的源图像中的信息，能在视觉上一致的情况下被保留在目标图像之中。

注意到上面的第一个等式是为一对源图像和目标图像设计的。为了将这种相似特性应用于所有的图像，我们将其扩展为：

as:

$$E_1 = \sum_{i=1}^N E_{\text{BDS}}(S_i, T_i).$$

基于块的合成技术，相比局部扭曲技术 ([Zhou and Koltun 2014]) 更为灵活，因而当几何结构和相机姿态出现大规模不精确的时候更适合于使用该技术。此外，虽然局部扭曲技术内在地保留了视觉一致性，但它同时也在目标图像中保留了我们不需要的、来自源图像的其他信息。如果几何模型不包含特定的特征，源图像中对应这些特征的区域将不应该被包含，从而也不在纹理映射中被包含。因此，该方法在这些区域会产生模糊/重影的人为合成痕迹，如图 4 所示。Waechter et al.[2014]的方法在每个面中选取一个视角，从而避免了这种情况下的糟糕效果。然而他们的方法不能将缺失特征部分对应的纹理给删去，因为这些特征在所有的源图像中都是存在的。注意到缺失的几何特征在绝大多数情况下都会带来不精确的几何结构（如图 9 所示），这正是现有的技术很难处理大规模不精确情况的原因。



图 4

尽管目标图像与源图像的相似性是生成高质量纹理映射的必要条件，却不是充分条件，如图 5 所示。

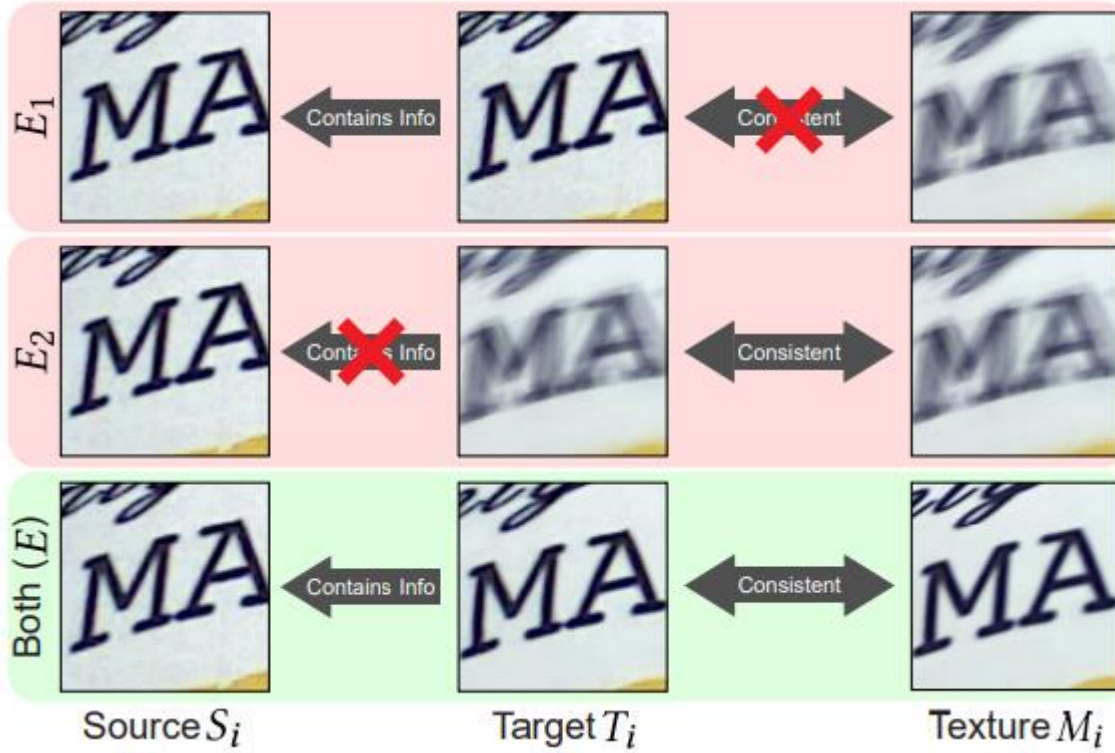


图 5

因此，我们需要将通过确保目标图像的一致性来实现第二项的效果。这个约束可以由以下几种方式实现。例如，我们可以通过保证投影后的目标图像与当前的目标图像很接近来实现一致性，例如使得 $T_j(x_{i \rightarrow j}) = T_i(x_i)$ 。这一约束可用 $T_j(x_{i \rightarrow j})$ 与 $T_i(x_i)$ 之间的二范数距离最小化的最小二乘思想来表述。

另一种选择是，该约束可通过保证当前目标与所有投影目标均值的一致性来实现，例如保证 $\frac{1}{N} \sum_{j=1}^N T_j(x_{i \rightarrow j}) = T_i(x_i)$ 成立。类似地，要满足该约束，可以令视角 i 处的纹理与投影后的目标图像的纹理保持一致，即满足等式 $T_j(x_{i \rightarrow j}) = M_i(x_i)$ 。由于所有的目标图像都将保持一致，且与优化后的纹理映射保持一致，应用上述两种不同的方法中任意一种会产生类似的最佳目标图像。然而，为利用交互优化（详见 3.2 模块），我们使用了后一种策略（ $T_j(x_{i \rightarrow j}) = M_i(x_i)$ ），一致性能量等式写作：

$$E_C(\{T_j\}_{j=1}^N, M_i) = \frac{1}{N} \sum_{\mathbf{x}_i} \sum_{j=1}^N w_j(\mathbf{x}_{i \rightarrow j}) (T_j(\mathbf{x}_{i \rightarrow j}) - M_i(\mathbf{x}_i))^2, \quad (3)$$

第一个求和符是对图像 i 上所有位于 \mathbf{x}_i 处的像素求和。权重 w_j 使得约束条件和第 j 个投影后的目标图像的贡献成比例。在具体实现时， $w_j = \cos(\theta)^2 / d^2$ ， θ 是表面法线和图像 j 的视角的夹角， d 表示相机与表面的距离。该权重在相机离物体较远，或者当相机视角与表面擦边时值较

小。最小化能量函数确保了所有的目标图像与从 i 相机视角看到的最终纹理映射保持一致。我们将该等式扩展成如下形式来确保这种对于所有图像的一致性约束：

$$E_2 = \sum_{i=1}^N E_C(\{T_j\}_{j=1}^N, M_i)$$

为满足这两个特性（等式的两项分别体现两个特性），我们将完整的目标函数记做带权的 E_1 与 E_2 之和：

$$E = E_1 + \lambda E_2, \quad (5)$$

λ 表示一致性那一项的权重，在我们的试验中设定为 0.1。优化我们提出的基于块的能量函数将产生包含源图像大多数信息，而且在视觉上具有一致性，而且保留了投影一致性的目标图像。一旦获得了优化的目标图像 T_i ，就能用不同的方式利用它们产生单一的一致性纹理。例如，通过先把所有的目标图像投影到几何结构上可以实现该目标。在该过程之后，每个顶点从不同的目标图像获取一个颜色样例的集合。每个顶点的最终颜色可通过计算这些样例的加权均值来得到。

图 5 是我们对优化系统中各项的效果评估。仅仅优化第一项可以产生于源图像有着一致视觉外表的配准后的图像，但是却不具有一致性。仅优化第二项能产生具有一致性的目标图像，可这些图像包含了在源图像中不存在的信息。优化我们提出的完整的能量函数可通过满足上述两个特性来产生高质量的纹理映射。

3.2 优化

为了高效地优化等式 5 中的能量函数，我们提出了替代性的优化方式，它在不同视角 M_1, \dots, M_N 下能同时优化目标图像 T_1, \dots, T_N ，以及纹理。特别地，我们通过轮流使用两个变量集来最小化能量函数。我们将目标图像和纹理分别初始化成他们相应的源图像，例如 $T_i = S_i, M_i = S_i$ 。接着我们交替地执行配准和重建两步，直到收敛。我们的算法的架构参见图 6。

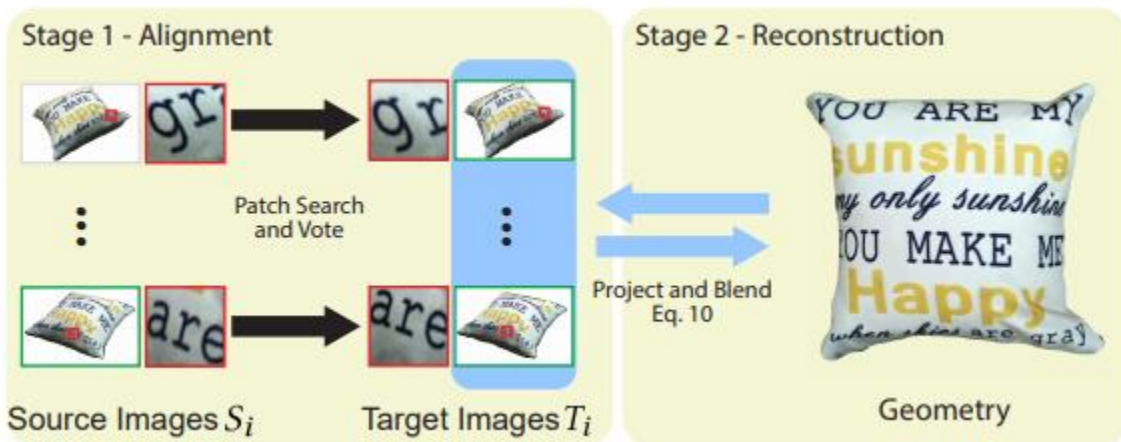


图 6

接下来我们解释以下这两个步骤：

1) 配准。

在这个环节，我们固定 M_1, \dots, M_N ，通过寻找最优的 T_1, \dots, T_N 来最小化等式 5。通过迭代的搜索和投票过程（与 Simakov et al. [2008] 类似）即可实现。在第一步中，我们进行块搜索的步骤，来寻找具有最小的 $D(s, t)$ 的块（参见等式 1）， D 表示平方差之和。在下一步中，我们实现了投票步骤，在上一步给定的算得的块的基础上，获取最小化等式 5 的 T_1, \dots, T_N 。注意，正如我们接下来要提到的，我们的投票机制与 Simakov et al. [2008] 是不同的，这是由于我们添加了额外的一致性约束 E_C 。

方便起见，我们通过先单独讨论等式 5 中的每个项来解释我们的投票机制。

第一项（相似性）：与 Simakov et al. [2008] 类似，我们通过使用在搜索过程中获得的块来重新定义 BDS 能量函数 (E_1)：

$$E_1(i, \mathbf{x}_i) = \frac{1}{L} \left[\sum_{u=1}^U (s_u(\mathbf{y}_u) - T_i(\mathbf{x}_i))^2 + \alpha \sum_{v=1}^V (s_v(\mathbf{y}_v) - T_i(\mathbf{x}_i))^2 \right]. \quad (6)$$

$E_1(i, x_i)$ 是指指定相机 i （这个 i 是）和像素 x_i 对应误差的 E_1 。 s_u 和 s_v 分别是重叠部分包含了像素 x_i 的源块，用以确保目标图像的完整性和一致性。此外， y_u 与 y_v 分别指 s_u 和 s_v 中的，对应于目标图像的第 x_i 个像素位置的单个像素。最后， U 和 V 分别指代为保证完整性和一致性所需的块的数目。注意到绝大多数的这些变量是当前像素 x_i 的函数，但是我们为标记符号的简单原则而省略了这部分。等式的推导参见 Simakov et al. [2008] 的论文。为获取能最小化上述等式的 T_i ，我们需要对未知的颜色 $T_i(x_i)$ 求导，令其值为 0，从而得到：

$$T_i(\mathbf{x}_i) = \frac{\frac{1}{L} \sum_{u=1}^U s_u(\mathbf{y}_u) + \frac{\alpha}{L} \sum_{v=1}^V s_v(\mathbf{y}_v)}{\frac{U}{L} + \frac{\alpha V}{L}}. \quad (7)$$

目标图像通过计算源块的集合中，包含目标图像的第 x_i 个像素的所有块的像素颜色加权均值。注意到尽管标准化系数 $\frac{1}{L}$ 可被约去，我们将它保留在等式中，以便使其可以在等式 9 中与等式 8 很容易结合。

第二项（一致性）：第一项是标准的投票过程，如 Simakov 等人所述，主要用于从源图像中抽取信息来重建目标。我们的主要区别在于第二项，它通过保证目标图像与纹理一致确保了一致性约束。正如附录中所示，最小化等式 5 中的第二项的目标值的计算方式为：

$$T_i(\mathbf{x}_i) = \frac{\frac{1}{N} w_i(\mathbf{x}_i) \sum_{k=1}^N M_k(\mathbf{x}_{i \rightarrow k})}{w_i(\mathbf{x}_i)}. \quad (8)$$

这里尽管权重 $w_i(x_i)$ 会被约去，我们在等式中仍然将其保留，以便在等式 9 中融合两项。在此通过计算当前纹理映射在不同视角的均值来计算目标。这是凭直觉得出的等式，因为约束主要是要求配准后的图像与纹理尽可能一致。

结合项：直觉上，求解结合项的目标值，应当通过在保证与纹理的相似性的同时，抽取源图像中信息，然后重建得到。由于这两项通过 λ 参数结合在一起，结合的结果应当能通过分别在分子、分母上添加等式 7、8 中的项来得到，记做：

$$T_i(\mathbf{x}_i) = \frac{\frac{1}{L} \sum_{u=1}^U s_u(y_u) + \frac{\alpha}{L} \sum_{v=1}^V s_v(y_v) + \frac{\lambda}{N} w_i(\mathbf{x}_i) \sum_{k=1}^N M_k(\mathbf{x}_{i \rightarrow k})}{\frac{U}{L} + \frac{\alpha V}{L} + \lambda w_i(\mathbf{x}_i)}. \quad (9)$$

最终版的目标值是加权的，包含常规的投票过程（等式 7）和所有当前纹理映射均值（等式 8）的结果。这意味着一致性那一项主要保证了我们的目标值与当前的纹理映射具有一致性。该能量函数的最小化步骤就是反复地做搜索、投票过程，直到收敛。该迭代过程通过将投票后的，经过更新的目标值作为新一轮迭代过程的输入。（这篇 paper 为何有这么多冗余的文字?!）我们实验表明只需要一轮迭代就足以获取高质量的结果，如图 7 所示。



图 7

2) 重建。

在该步骤中，我们固定 T_1, \dots, T_N ，计算出不同视角下最优的纹理 M_1, \dots, M_N 来最小化等式 5。由于纹理只在第二项（ E_C ）这个二次式中出现，最优纹理可通过以下方式求得：

$$M_i(\mathbf{x}_i) = \frac{\sum_{j=1}^N w_j(\mathbf{x}_{i \rightarrow j}) T_j(\mathbf{x}_{i \rightarrow j})}{\sum_{j=1}^N w_j(\mathbf{x}_{i \rightarrow j})}. \quad (10)$$

这是我们的纹理生成等式，它主要说明了最优纹理是由所有投影后的目标的加权均值算得的。在目标失配这种优化初期很常见的情况之下，该过程会产生带重影和模糊的纹理。下一个配准过程的迭代就会试着缓和目标间的失配程度，最终使得在重建之后纹理映射中出现的人工痕迹更少。

我们迭代地执行配准和重建两步，直至收敛。与[Barnes et al. 2009; Wexler et al. 2007]基于块的配准方式相同，我们在多尺度上执行这两步来避免局部最小值，同时加速收敛（参见模块4）。注意到这里的迭代是在我们的配准、重建这两大步骤之间进行的。我们还有一个内层迭代，它位于配准步骤中，反复进行的是搜索和投票步骤（不过这个迭代只需进行1次）。

一经收敛，我们的算法就生成配准的图像 T_1, \dots, T_N 以及在不同视角下的最优纹理 M_1, \dots, M_N ，这两者非常相似。由于目标图像具有一致性，通过将所有的目标图像投影到集合结构上，并求出他们的颜色样本均值，以获得每个顶点上的最终颜色，完成这些步骤后最终仅生成一个全局纹理。

四、实现细节

获取输入数据。我们使用 Intel RealSense R200 相机来抓取我们的 RGB-D 序列。该相机以 628X468，或 1920X1080 的分辨率记录深度及色彩序列，帧率为 30。为了最小化颜色差异，我们固定了曝光及白平衡（参见博客⑬）。我们使用 KinectFusion 的算法 ([Izadi et al. 2011]) 来估计每一帧的集合结构和相机姿态。注意到该方法估计了相机姿态的深度结构，我们也将此估计赋给相应的彩图结构（有种获取彩图的相机结构的方法，需要将严格的转换用到深度相机的姿态之上，但是该策略不能显著地起帮助作用，主要是因为以下两个原因：一是深度相机和彩度相机的快门不是同步的，二是我们的深度和彩度相机非常靠近，所以有相似的姿态）。

关键帧选择。为了减少我们输入的图像的数目，我们用与 Koltun 的方法 [2014] 类似的贪心的方式来选取一个图像的子集。特别地，给定一个已经选择好关键帧的集合，我们使用 Crete et al. [2007] 的方法来寻找在已选择的最后一个关键帧之后的，有着最低模糊度的，介于区间 $(t, 2t)$ 的一帧。在实现时， t 根据场景可在 30 到 60 帧之间做选择。

配准。为加速搜索过程，我们使用块匹配的算法 (Barnes et al. [2009])，块尺寸参数设置为默认值 7。此外，为了避免目标图像与源图像的偏差过大，我们将搜索框的尺寸限制在 $0.1\sqrt{w \cdot h}$ ， w 和 h 分别表示源图像的宽度与高度。

多尺度优化。我们通过将优化过程在多尺度上应用来求解等式 5 中的能量函数。特别地，我们首先对所有的源图像下采样至最粗糙的尺度（上下采样的定义参见参考博客④）。我们先用低分辨率的源图像初始化 T_1, \dots, T_N 和纹理 M_1, \dots, M_N ，并且迭代地进行配准和重建步骤直至收敛。我们

接着对所有的目标和纹理上采样至下个尺度的清晰度，并且在新的尺度上迭代地做这两个步骤。注意到我们并不是在更粗糙的尺度上对源图像上采样，而是直接对原先的高分辨率源图像做下采样至当前的尺度。这就使得系统可以将高频率出现的细节插入到目标图像和纹理中。我们在所有的更精细的尺度上继续做该步骤，从而在最精细的尺度上获取最终结果。在最粗糙的尺度上，输入的图像在更小的维度上有 64 个像素我们总共有 10 个尺度，尺度参数为 $\sqrt[3]{x/64}$ ， x 是原先的源图像中较小的维度。我们在最粗糙的尺度上做 50 次配准、重建的迭代，在更精细的尺度上则减少至 5 次。

正如图 8 所示，这种多尺度的方式对于避免局部极小值来说是很有必要的，因而会生成高质量的结果。直觉上说，我们的优化系统在更粗糙的尺度上配准全局结构，在更精细的尺度上复原细节。我们在附件视频中演示了我们算法在不同尺度上的收敛算法。



图 8

五、结果（配图灰色部分就是物体的（较为精确的）三维几何结构）

我们用 MATLAB/C++ 实现了框架算法，并与当下最好方法（Eisemann et al. [2008]; Waechter et al. [2014]; Zhou and Koltun [2014]）做对比。我们直接使用了前两者的源码，而第三个方法由于没有开源代码，因此我们自己编程实现了。注意到对于 Eisemann 等人的算法，我们使用了静态的场景，并生成了视角独立的纹理，来做公平的比较。我们通过展示每个物体的一至两个视角来演示结果，在附录的视频中可以找到来自不同视角的，体现了纹理映射效果的物体的段落。注意到我们的应用场景通常比 Zhou and Koltun 的更具挑战性。这（更具挑战性的原因）主要是由于我们随机地在典型的光照条件下选择场景，因此我们的几何结构精确度较低。我们在 Zhou and Koltun 论文中提到的 FOUNTAIN（喷泉）场景中测试了我们的方法，得到了有可比性的效果，参见图 14（配准目标）。

图 9 是我们的方法与其他方法在 6 个有挑战性物体上应用效果的比较，这些物体的经估计的几何结构参见图 10。TRUCK（卡车）是有着复杂几何结构的有挑战性的景物，它无法通过消费级深度相机直接获取几何结构。Eisemann et al. [2008]在一对图像上做处理，使用没有优化全局能量函数的光流来校正失配，这种解法是求次优解的。由于他们的扭曲后的图像包含了局部失配的情况，他们的方法会生成模糊的纹理。Waechter et al. [2014]的方法通过求解优化系统从每个表面选择一个视角，来隐藏相邻面的接缝。然而它们的方法无法在该情形下生成令人满意的结果，这是

因为他们在精确度很低的情况下会将不连续的纹理赋给邻接面。参见上一行插图中撕裂的人工痕迹，和下一行插图的失真的熊脸。



此外由于不精确的几何结构（见图 10），Zhou and Koltun 的方法中的局部扭曲无法对该情况中的显著的失配做校正。因此他们的结果饱受重影和模糊之苦。我们的方法将配准后的目标图像进行合成，从而能够产生人工痕迹很少的纹理映射图像。

没有其他方法能够处理枪的场景。特别注意在下一行插图中只有我们的方法能够重建细的黑的结构。

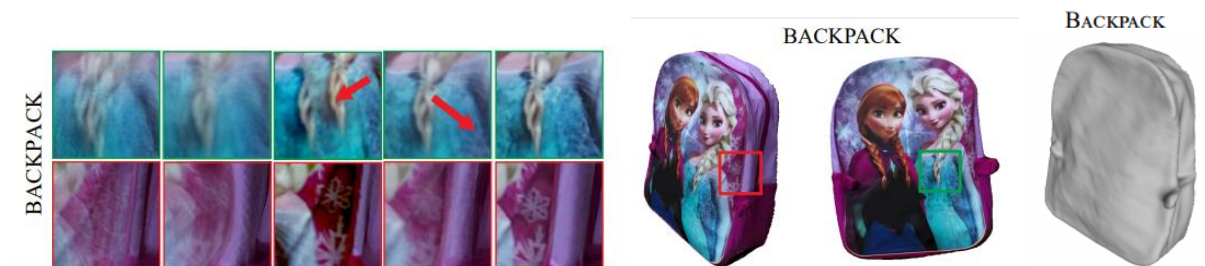


由于光流估计的不精确性，Eisemann 等人的方法产生的结果具有撕裂的人工痕迹。Waechter 等人的方法通过校正颜色的做法来解决邻接面之间颜色差异的问题，这种做法价值并不大。由于在这种情况下图像是显著地失配的，邻接面可能有不连续的纹理。因此在两行插图中可以看到，颜色校正会带来肉眼可见的变色情况。另外，我们尝试了具有复杂几何结构的房屋景物。



Waechter 等人的方法产生了撕裂的人工痕迹，而 Eisemann、Zhou and Koltun 的结果则会有重影。这主要是由于景物的复杂性以及集合结构的不精确造成的。不过我们的方法可以在这些挑战性的场景下产生高质量的结果。

书包景物的上一行插图展示了一个几何结构相对平滑的区域。然而 Eisemann 等人的方法仍然无法正确地配准图像，他们生成的还是模糊的纹理。此外，Waechter 等人的方法因为相机姿态数据不正确的缘故，产生了具有撕裂的人工痕迹的结果。



尽管 Zhou and Koltun 的方法在该样例中矫正了大部分的失配，他们的结果相比我们的还是要模糊一些。下一行插图展示了有着复杂纹理的书包的一个侧面区域。在该区域中，Waechter 等人的方法会显示出褪色的人工痕迹，而 Zhou and Koltun 或是 Eisemann 的方法则会造成重影的结果。这些方法在处理枕头样例的边沿时，由于区域的几何结构很复杂，也无法提供正确重建的纹理。值得一提的是，Waechter 等人的做法在枕头的阴暗面也会产生褪色的人工痕迹（参见附件视频）。最终我们的方法在牛的样例中恰当地重建了眼和爱心图案的部分，以及蓝色和棕色的结构部分。



在人的样例上，将我们的方法与其他方法做比较，如图 11 所示：由于在抓取环节物体一致在动，所以这个景物对所有方法来说都很有挑战性。虽然其他的所有方法产生的结果都具有重影和模糊的人工痕迹，我们的方法恰当地处理了所有的不精确性，产生了高质量的纹理。

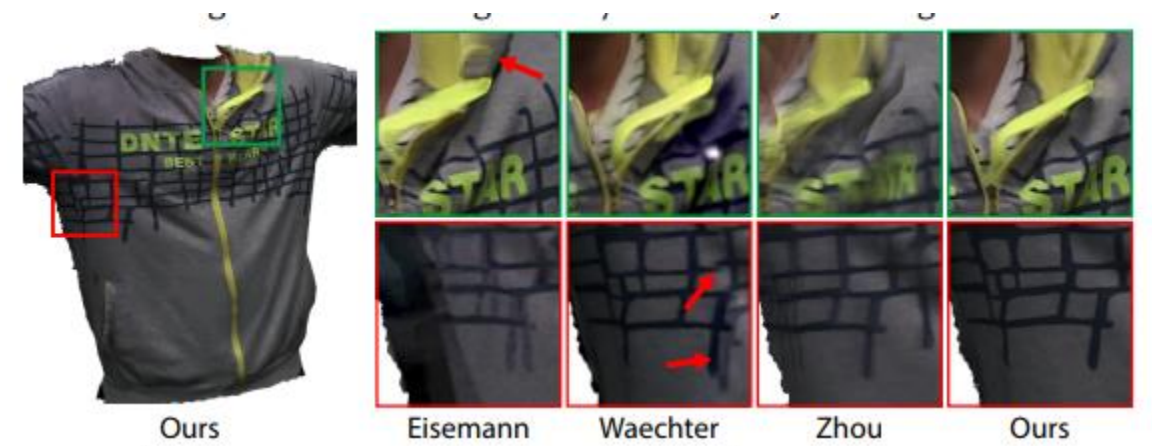


图 11

局限性。我们的方法的主要局限在于，基于块的合成通常产生貌似真实的结果，但是在某些场景下却无法保留语义信息，正如图 12 所示。这里尽管我们的方式矫正了显著的失配，生成了看似合理的结果，却不能保留洞的结构。



图 12

六、其他应用

在本章节，我们讨论本系统的几个应用，包括纹理的修补孔洞，图像重组，多视角图像伪装。注意到尽管基于块的合成在之前已经被用于图像补洞和重组 ([Barnes et al. 2009; Simakov et al. 2008])，这些方法因为缺少一致性所以在我们的应用中不合适。

6.1 纹理补洞

在一些情形下，真实世界中物体的纹理可能包含一些我们不想要的区域，我们需要用填洞的方式将它修饰掉。图 13 就是一个例子：枕头上的标签是我们不想要的，应当在最终的纹理映射中被移除。为了实现该目标，我们首先用我们的系统合成配准后的目标图像。接着在一张配准图像中标记需要填补的区域（用蓝色标记）。该区域可以被简单地投影到其他视角，从而在所有的目标图像中生成这个洞。这些标记过的区域将每张目标图像分为 H_i 和 I_i (洞外区域)。

这里的目标是从每个输入 I_i 中抽取信息来填补 H_i ，同时保证被填充区域的光学一致性。这与等式 5 中能量函数的主要性质很相似，所以我们的系统能用于实现补洞操作。注意到这个问题与多视角补洞这个课题有关，在[Baek et al. 2016; Thonat et al. 2016]中也有不少技术用于解决这些问题，但是我们提出了一种使用纹理映射框架来实现该工作的方法。

我们令等式 5 中的 $S_i = I_i$, $T_i = H_i$ 。在此情形下，我们的优化以连续的方式从源中（洞外区域）抽取信息来填补目标（洞）。这是通过对洞内外区域做块搜索，对搜出的区域做投票，然后只对洞的区域进行重建而成的。初始化的时候，我们并没有使用源图像，而是使用 MATLAB 的 `roifill` 函数（参见博客⑧）对洞的边界像素做平滑的填充。我们也省略了 BDS 能量项中的完整性那一项（见等式 1），它用于在目标图像中保留绝大部分源图像中的信息。注意到虽然对于配准是必要的，对于填洞来说不是必须的，因为我们只需要输入的部分，而不是全部信息来填洞。

在图 13 中包含了我们的方法与[Wexler et al. 2007]基于块的图像补洞技术的比较。虽然独立地补洞在每个视角都能产生看似合理的结果（顶行），将它们结合起来的时候就会产生重影的效果（底部左列），这是因为它们缺乏一致性。Zhou and Koltun [2014]的方法可用于在不同的视角配准填洞后的图像（底部中间）。然而最终的纹理仍然包含了重影的效果，因为通过扭曲无法校正

不一致性。我们的方法能够在不同的视角生成一致的填洞后的结果，进而生成高质量填洞后的纹理。

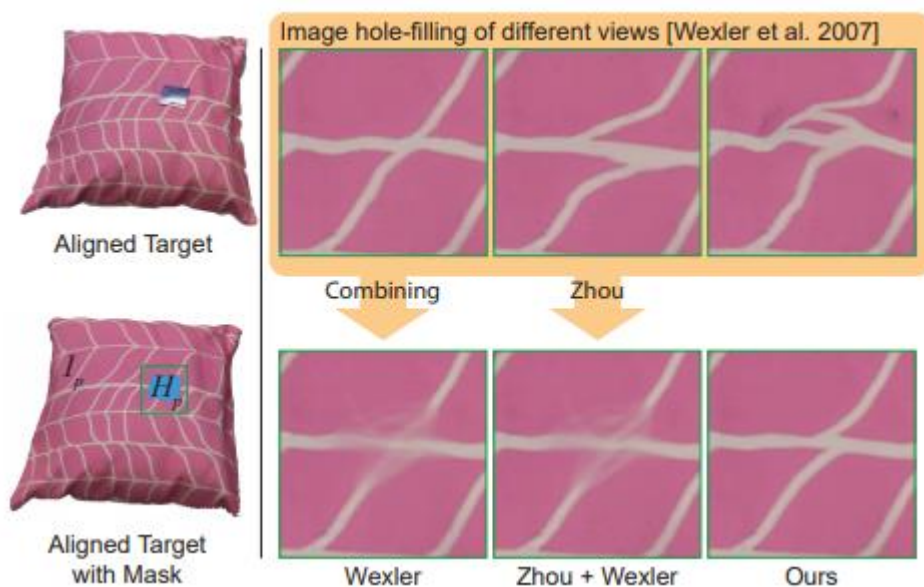


图 13

不过该技术不能用于几何结构填洞。因此我们的方法只能填补哪些底层几何结构不复杂的纹理的洞，就像图 13 中的情形。如何将我们的系统扩展到几何结构的填补是未来有趣的一个研究方向。

6.2 纹理重组

如图 14 所示，我们的方法也能被用来复制部分的纹理（用红色记号标记）并拷贝到其他区域上（用黄色记号标记）。在开始重组之前，我们用系统合成配准后的目标图像。接着我们在一张目标图像上（需要重组的目标图像）标记一些区域，我们的目的是在需要的区域以看似真实的方式合理地替代它们（图 14 中标记成黄色的部分）。此外，在要重组的目标图像的新位置处，合成的内容需要与其他的目标图像一致。

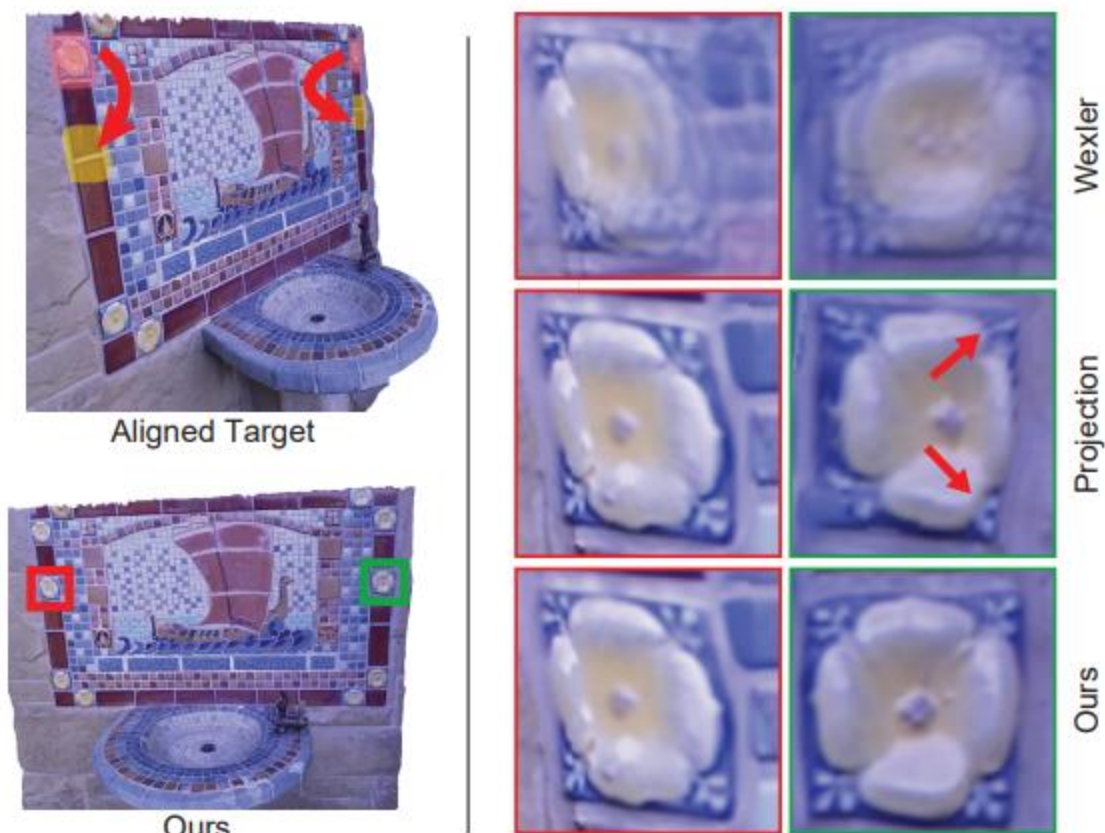


图 14

如何实现呢？我们首先对单张图像做重组 ([Simakov et al. 2008])，在新位置合成 ROI (region of interest 感兴趣区域) 区域的复制品。注意到该过程与 Simakov et al. [2008] 的是类似的，不过只是对于需要重组的目标进行的。此刻其他目标图像与该图像，在执行了重组操作的区域 (黄色部分) 是不一致的。

如何解决该问题？首先将黄色部分投影到其他目标图像上。接着我们用我们描述的填洞系统，在其他目标图像上填充黄色投影部分。注意到在此我们并未修改需要重组的目标，他只能用于强制其他目标图像，在由黄色投影部分定义的区域产生保持一致性的内容。之前讨论过，这就意味着我们移除了等式 5 中重组目标对应的 E_{BDS} 项。

该过程产生的目标图像与需要重建的目标图像具有一致性，正如图 14 所示。只是，通过使用 Wexler 方法对每个目标独立地使用填洞得到纹理仍然包含重影的人工痕迹。此外，将黄色部分从重组目标图像投影到其他图像上会产生模糊。我们的方法则能产生高质量的结果。

6.3 多视角图像隐藏

我们的方法也能用于在多视角中隐藏一个 3D 物体。输入一个场景的图像集合，以及一个需要插入到场景中并隐藏起来的，物体的 3D 几何结构。通过生成几何结构的一致性纹理映射，可以使其在不同的视角变得不可见。该问题可被视作一个高度不精确的几何结构的基于图像的纹理映射，这个场景中的几何结构则通过 3D 物体建模。我们将自己的技术的结果与 Owens 等人的方法对隐藏一个盒子这个任务的效果进行比较，参见图 15.注意到他们的方法是专门为这个任务设计的，只能用来隐藏盒子。因此他们的方法在这种情况下能生成高质量的图像。经过比较，我们的框架也能处理这个外延任务，产生合理的结果。此外我们的方法可以处理不限于盒子的各类物体（参见附件视频）。



图 15

七、结论与展望

在本论文中，我们为基于图像的纹理映射提出了一种新颖的全局的基于块的优化系统。我们通过为每张源图像合成配准后的图像，矫正了由几何结构、相机姿态、输入图像光学扭曲等造成的不精确部分。我们使用了新颖的基于块的能量函数，通过从源图像中重建具有光学一致性的配准后的图像，来实现该过程。为了高效地求解我们的能量函数，我们提出了一个包含两个步骤的策略，一个步骤是修改过的块搜索，之后一个步骤是重建。我们展示了该方法在处理很不精确的情形时的有效性，以及在效果方面超过当前最好方法的结果。此外我们也演示了系统在纹理编辑、多视角隐藏物体等其他方面的应用价值。

未来我们将试着扩展该系统的应用面，直接校正几何结构的相机姿态后生成配准好的图像。此外，我们将研究该系统在宽基线视角（参见参考博客⑤）的插值问题，在这种情况下图像集合中的信息需要结合起来用以生成具有一致性的新视角下的图像。

附录

这里我们讨论一下等式 8 中的求导，它计算出了使得等式 5 中第二项的取得最小值的目标变量取值。我们首先将 E_2 重新记做：

$$E_2 = \frac{1}{N} \sum_{k=1}^N \sum_{\mathbf{x}_k} \sum_{j=1}^N w_j(\mathbf{x}_{k \rightarrow j}) \left(T_j(\mathbf{x}_{k \rightarrow j}) - M_k(\mathbf{x}_k) \right)^2. \quad (11)$$

为了计算最优的结果，我们首先需要对每个目标值求偏导，即为：

$$\frac{\partial E_2}{\partial T_i(\mathbf{x}_i)} = \frac{\partial \sum_{k=1}^N \sum_{\mathbf{x}_k} w_i(\mathbf{x}_{k \rightarrow i}) (T_i(\mathbf{x}_{k \rightarrow i}) - M_k(\mathbf{x}_k))^2}{\partial T_i(\mathbf{x}_i)}, \quad (12)$$

我们用到了一个事实： $x_{i \rightarrow k \rightarrow i} = x_i$ 。通过对上式进行求导计算，令其为 0，即可求得等式 8。注意到由于导函数与目标图像的单个像素 x_i 有关，我们在求导之前去除了对所有像素求和的符号。

一些参考资料

- 1.我认为本文的 image alignment 就是 image registration，可解释为图像对齐/图像配准，参见博客⑥
- 2.本文中的 patch 是图像的块，比如一个 3x3 或 5x5 的正方形像素区域，而不是补丁的意思。在图形学中有 patch-match 算法，利用图片中的其他区域来恢复边缘区域的做法。参见博客⑦
- 3.图像重组-reshuffle：将一张图片进行重新组合，比如讲图片右上角的部分贴到左边，并且使得图片看起来很自然。参见博客⑦
- 4.view-independent:视点无关的，是指从任意角度拍摄的图像都可以
- 5.novel view synthesis 也是一个术语：Given a single image depicting an object, novel-view synthesis is the task of generating new images that render the object from a different viewpoint than the one given.参见资料⑩我翻译为：新视角图像生成
- 6.multiview camouflage 应该是指 multiview image camouflage（在图像中隐藏信息的技术），参考论文⑪，⑫等文献。下面给出其中一篇文献对该术语的界定：

Camouflage image is also referred as hidden image. It is an instance of recreational art.（图像伪装属于娱乐艺术中的一种。）Such an image contains one or more hidden objects.（一张图像包含了一个或

多个的被遮挡的物体。) The color plays an important role in vision. (在视觉上颜色占据了重要的地位。) Camouflaging technique is also called cryptic coloration and it is trick that is used to disguise its appearance, usually to blend with its surroundings. (伪装技术也被称作隐蔽上色, 其欺骗性在于它被用于伪装其外表, 混合在它的周围环境之中。) This technique is used to mask location, identity and movement while it is the mixture of resources, design or information which is hidden within an image. (这种技术被用来遮蔽位置、标志和移动, 它是图像内隐藏的资源、设计或信息的混合。) Our proposed solution is identified to apply these color extraction technique which increases high visual quality of image. (我们提出的解决方式可视作应用了这些增加图像高视觉质量的颜色抽取技术。) Our system can be used in many fields such as military security, gaming, automotive manufactures, animation and graphic designing purposes.

7.附件视频在项目地址中可以找到

8.参考资料:

①【深度相机系列三】深度相机原理揭秘--双目立体视觉:

<https://blog.csdn.net/electech6/article/details/78526800>

②本征图像分解方法与应用研究:

<http://www.docin.com/p-1869698117.html?docfrom=rrela>

③ image retargeting (图像缩略图、图像重定向):

<https://blog.csdn.net/u010922186/article/details/41652703>

④图像的上采样 (up-sampling) 和下采样(down-sampling):

<https://blog.csdn.net/ccblogger/article/details/72875497>

⑤宽基线(Wide Baseline)和窄基线(Short Baseline):

https://blog.csdn.net/Ysm_Shui/article/details/50530976

⑥ 机器视觉: 基于特征的图像对齐 (使用 opencv 和 python) :

<https://blog.csdn.net/yuanlulu/article/details/82222119>

⑦ PatchMatch 分析:

<https://blog.csdn.net/z6491679/article/details/50807689>

⑧6. 特定区域处理:

https://blog.csdn.net/matlab_matlab/article/details/54015331

其中提到 roifill func 的作用为: 通过求解边界的拉普拉斯方程, 利用多边形边界点的灰度平滑的插值得到多边形内部的点。通常可以利用对指定区域的填充来“擦”掉图像中的小块区域

⑨影像变形的 wiki 百科:

<https://en.wikipedia.org/wiki/Morphing>

⑩Novel Views of Objects from a Single Image:

<https://arxiv.org/abs/1602.00328>

⑪ 论文“Camouflage images”: <https://dl.acm.org/citation.cfm?id=1778788>

⑫ 论文“Camouflage Image Generation System For Security”,
<https://ieeexplore.ieee.org/document/7824827>

⑬ 摄像头自动曝光相关基础知识:

<https://blog.csdn.net/u011776903/article/details/78783975>