

A Brief Report on Gradient-Domain Rendering

梁晨 3180102160 计算机学院 本科生院

● Introduction

我从[A Survey on Gradient-Domain Rendering](#)开始阅读，在了解了Gradient-Domain的基本想法和三个阶段（Sampling、Shift-Mapping、Reconstruction）的基本框架后，又选择了一些相关的文章做了延伸阅读，并最终总结、梳理成这份报告。

这份报告中，我将着重介绍Gradient-Domain Rendering的三个阶段，包括现行的算法、算法的局限性和未来的展望。在整篇报告中，我致力于能让读者直观地理解Gradient-Domain Rendering这项技术的相关进展，于是会跳过一些理论性的推导与证明，但我也会尽量做到叙述上的严谨。如果对详细的理论推导感兴趣，建议读者移步原论文的相关章节。

● Basic Idea

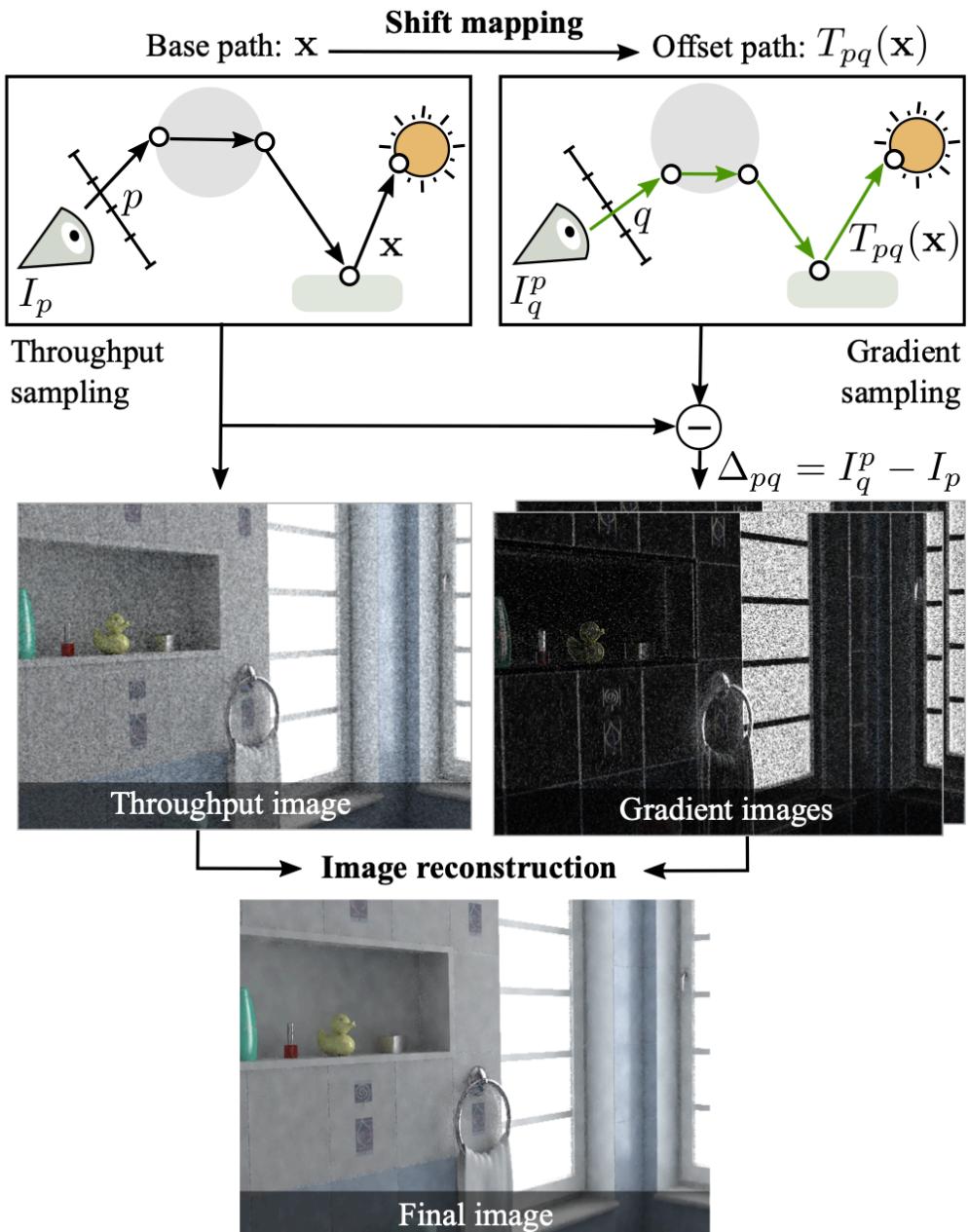
自1986年Kajiya等人在论文[The Rendering Equation](#)提出渲染方程以及用Monte Carlo积分对其进行估计的方法后，学术界、工业界便普遍采用Monte Carlo方法对光线传播问题进行求解，致力于生成物理上正确的、具有全局光照效果的高真实感图片。Monte Carlo积分是一种无偏估计，即无论采样量如何，样本的数学期望与真实结果是相等的。虽然如此，但Monte Carlo估计子的方差却大大依赖于采样量，且随着样本量增大，误差仅按照 $O(1/\sqrt{N})$ 的速率收敛到0，效率不高。低样本量往往导致生成的图片噪点高、质量低，而高样本量的估计子的计算又会带来巨大的时间开销，从几分钟到几小时不等。所以，目前学术界对于真实感渲染的研究，主要目的就是降低Monte Carlo估计子的方差，或者通俗来讲就是降噪。科学家们试图研究出新的方法，在Monte Carlo积分的大框架下，以更低的采样量、更少的时间为代价，换取方差更低、噪点更少、质量更高的图片。Gradient-Domain Rendering正是这些方法中的一类。

Gradient-Domain Rendering的基本思想可以理解为把图像空间的梯度图作为一个Feature，去和低spp的原图做重建，从而得到方差更小的结果图。这样做之所以能奏效是因为图像空间存在着某种连续性，我们认为相邻（可以是图像空间上的相邻，也可以是其他意义上的“相邻”，在下面Sampling的部分我会展开说明）的像素间，它们Radiance，或者说RGB值，是非常接近的。而低spp的原图，方差之所以很大，则是因为对这样的连续性，或说是一致性造成了某种破坏。于是我们将低ssp的原图与梯度图做重建，希望得到的结果图，可以同时满足两者的条件，则可以修复这样的破坏，从而得到方差更小的图片。

于是自然地，Gradient-Domain Rendering这项技术的研究就可以拆解为以下三个阶段：

- 对于一个像素，选择哪些其他像素去和它计算梯度？(Sampling)
- 梯度要如何计算？(Shift-Mapping)
- 得到梯度图后，它和低ssp的原图要用哪种方法做重建？(Reconstruction)

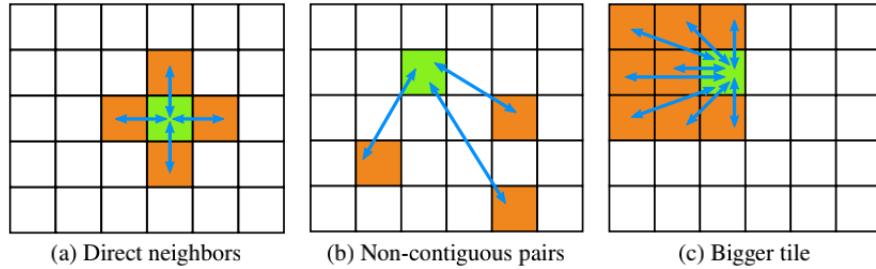
具体流程图表达如下，我的报告也将按照这个流程顺此展开。



- Sampling

对于一个像素，我们如何要采其他哪些像素来计算它的梯度呢？最自然的想法肯定是在图像空间上的相邻像素，在最初的论文[Grandient-Domain Metropolis Light Transport](#)中，作者也的确是这么做的。但这样做有一个问题，就是当图像空间不够连续，或者说不够光滑时，采出来的梯度图就不再是修复低spp原图的一个好的feature。举例来说，在上面流程图中，我们看到梯度图上有Intensity非常大的区域，主要集中在窗户（光源）与其他物体的过度处。这样的高梯度feature在重建过程中，会让重建算法分不清低spp的原图中相对区域的高梯度是符合实际的，还是方差带来的。

如何解决这个问题呢？直觉上，我们可以采样更大的区域，来平均梯度图上过高的Intensity；另一种思路是干脆放弃在图像空间上的相邻区域做采样，而是追求“另一种空间上的相邻”。在图像空间上体现出来，则是在做不连续的采样来计算梯度。

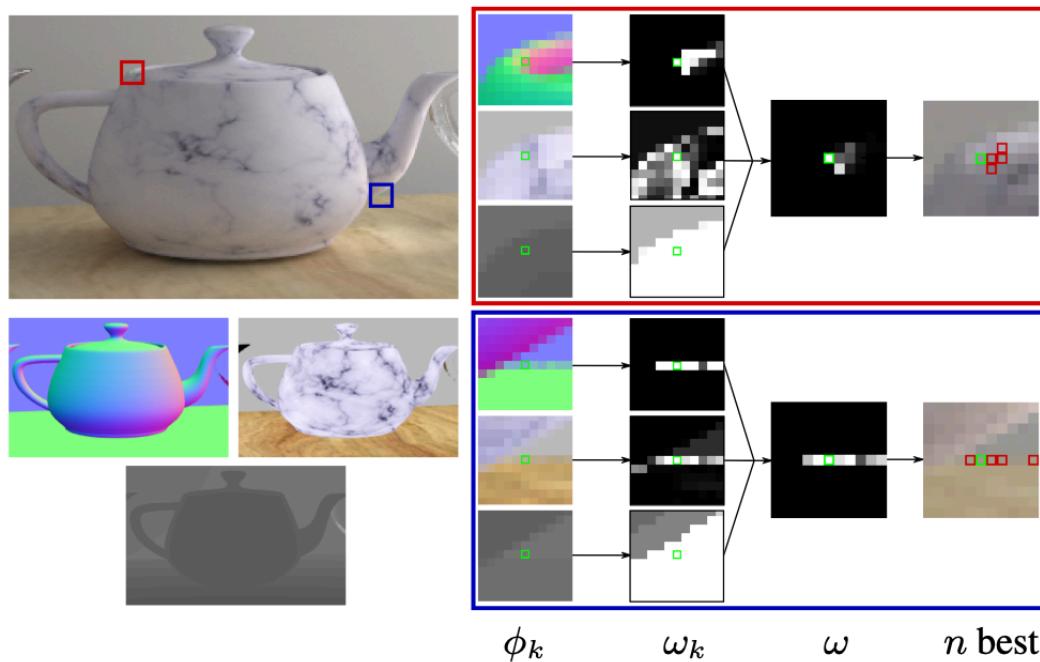


在论文[Improved Sampling for Gradient-Domain Metropolis Light Transport](#)中，作者提出了一种人为定义的距离度量函数，来决定哪个像素和现在要采梯度的像素是“相邻”的。

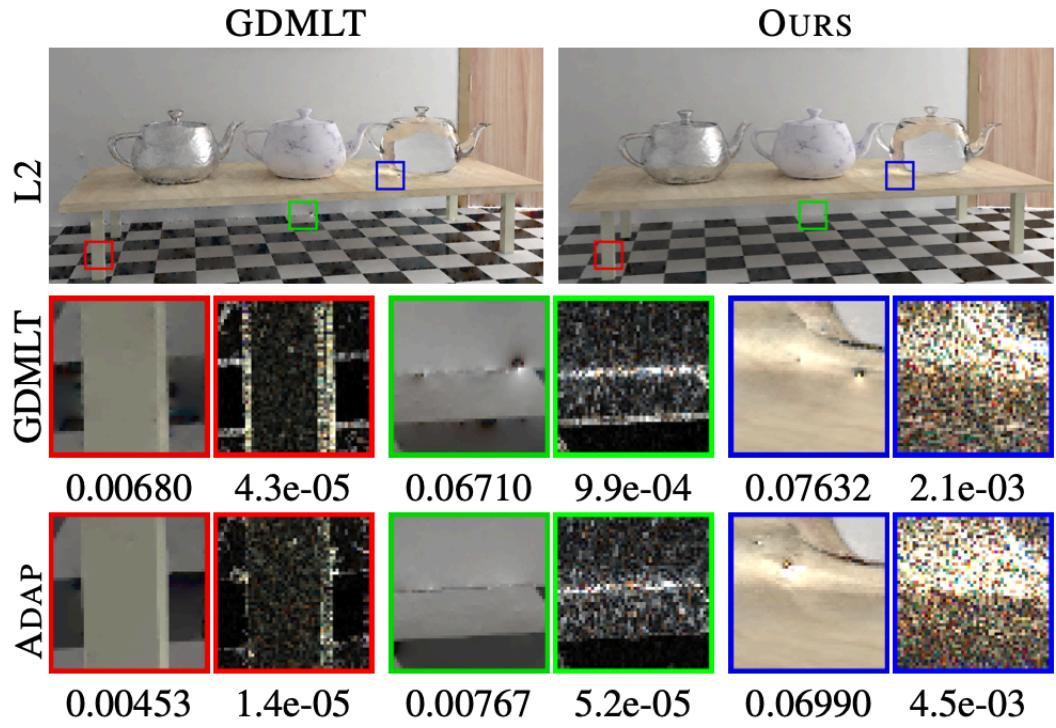
$$\omega_k = \exp \left(-\frac{(\max(0, |\phi_k(i) - \phi_k(j)|^2 - \tau_k)}{\sigma_k^2} \right)$$

首先，我们提前导入 n 个较易获取的feature，比如深度图、albedo图等等。以上公式定义了在第 k 个feature的度量下，第*i*个pixel与第*j*个pixel之间的距离。其中 $\phi_k(i)$ 是第*i*个pixel的第 k 个feature的值， τ_k 和 σ_k 是两个人为定义的参数，对于理解公式意义不大。可以看到，以上公式可以理解成在计算两个pixel第 k 个feature的差的二范数，之后将这个二范数做了某种映射（取指数之类），反应的恰恰是两个pixel之间的在第 k 个feature的度量下的“距离”。

很自然地，我们定义两个pixel之间的距离为 $\omega(i, j) = \min_k[\omega_k]$ 。对于第*i*个pixel，取与它距离最小的4个pixel(*j*)来采梯度。



在上图中，左半边说明了我们根据三个feature，即albedo、texture和depth作为feature，来决定选择哪些“相邻”像素采梯度。右半边反应了我们根据度量函数，决定哪个feature作为参考，在红框、蓝框里我们均选用了texture。作者说明，在rmse的准则下，在其他条件不变的情况下，使用这种Sampling的策略(ADAP)比使用Naive的采图像空间相邻像素的Sampling策略(GDMLT)，有如下的提升：



- Shift-Mapping

Shift-Mapping其实是涉及到梯度要如何计算的问题。最简单的方法，我们容易想到就是两个相邻像素做差。然而这样并不能达到减小方差的效果。因为如果我们取低spp原图的相邻两像素做差，相当于没有加进去新的feature，达到的是两像素差的方差等于两像素方差的差这一trivial的效果。

为了减小方差，得到增添了信息量的梯度图，我们要进一步探索场景空间，即path space。假设在第 p 个像素的Radiance是采样 \mathbf{x}_k 这条path得来的，考虑图像空间的渲染方程，我们有：

$$\langle I_p \rangle = \frac{1}{N} \sum_{k=1}^N \frac{h_p(\mathbf{x}_k) f(\mathbf{x}_k)}{p(\mathbf{x}_k)},$$

其中 f 是贡献度量函数 (measurement contribution function)， h 是重建滤波函数 (reconstruction filter)， p 是采样路径 \mathbf{x}_k 的概率密度函数。我们在计算第 p 个像素与第 q 个像素的radiance的差值时，希望通过一个变换函数 T ，把路径 \mathbf{x}_k 变换到路径 $T(\mathbf{x}_k)$ ，去算出第 q 个像素的radiance，然后再做差，这样就得到了“增添了信息”的梯度图。表达成方程的形式即为：

$$\langle \Delta_{pq} \rangle = \frac{1}{N} \sum_{k=1}^N h_p(\mathbf{x}_k) \left(\frac{f(\mathbf{x}_k)}{p(\mathbf{x}_k)} - \frac{f(T_{pq}(\mathbf{x}_k))}{p(T_{pq}(\mathbf{x}_k))} \right).$$

其中函数 T 即是Shift-Mapping。我们的目标是通过设置合适的Shift-Mapping，使得以上Monte Carlo估计子的方差值尽可能小。容易推导得到，要达成这一目标，要满足的条件是 $T(\mathbf{x}_k) \approx \mathbf{x}_k$ 。我们将 \mathbf{x}_k 称作base path，将 $T(\mathbf{x}_k)$ 称作offset path。

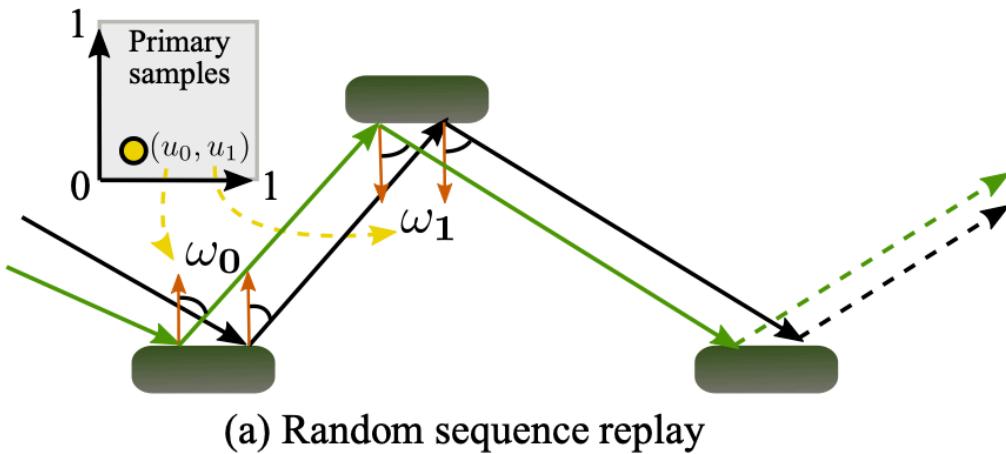
目前做Shift-Mapping的方法主要有以下四种：

- Random Sequence Replay
- Half-Vector Copy

- Path Reconnection
- Manifold Exploration

除最后一种外，另外三种都能顾名思义，如果对Monte Carlo Path Tracing有基本的了解，就能大概从名字猜出其做法，下面我一一介绍。

Random Sequence Replay即是在采样offset path时，使用和base path相同的一组随机数，使得在局部上看，对于path上的每一小段，base path和offset path在空间中局部坐标系下的角度是相同的。但这样只保证了局部的相似性，并不能保证全局的相似性。



上图的例子中描述了一种非常理想的情况，即offset path的每一小段都和base path的对应小段打到了相同物体上，即**局部标架**是相同，再加上采样的时候Random Number是相同的，导致全局上依然保持了offset path和base path的相似性。但多数情况下，因为offset path和base path从空间中的不同位置出发，很有可能第一次打到的的物体就是不同的，**局部标架**也是不同的，使用相同的Random Number，使得offset path和base path并不能具备很好的相似性。且随着每一段的采样，这种offset path和base path间的偏移是会被“逐步放大”的，导致虽然局部上两者相似，但放到全局，两者可能相差甚远。场景中就算全是Diffuse材质的物体都会如此，更莫说材质情况复杂，导致光线情况复杂，有Specular甚至Specular Chain的情况。

Random Sequence Replay的优点是保证了物理上的正确性，即采样出的offset path一定是真实中确实存在的。同时，Random Sequence Replay是一种可逆的策略。

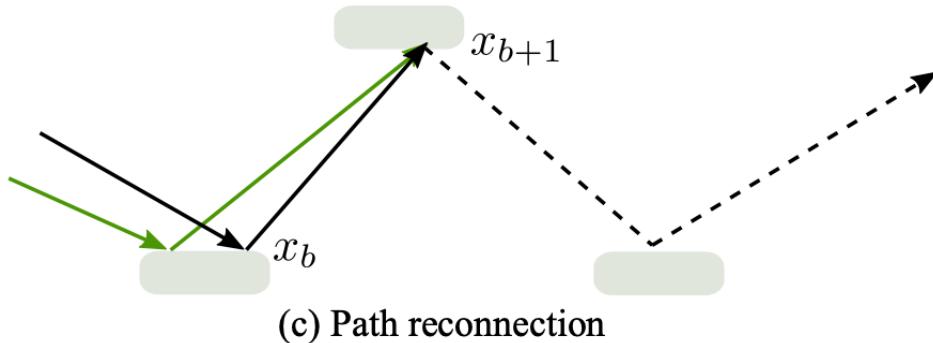
Half Vector Copy顾名思义，即保持base path的半角向量不变，然后根据这个相同的半角向量去采offset的下一个点。

$$h = \frac{\omega_i + \omega_o}{\|\omega_i + \omega_o\|} = \frac{\omega'_i + \omega'_o}{\|\omega'_i + \omega'_o\|} = h'$$

这样做的好处是，因为half vector是一个在**世界坐标系**下不变的量，导致base path和offset path可以在一定程度上保持全局的相似性，即便它们每次打到不同的物体。

但Half Vector Copy的劣势是不一定能采出物理上正确的offset path，且是不可逆的。原因是根据保留下来的半角向量算出的offset path的出射光的值，很有可能和物体本身的材质相矛盾。举一个比较极端的例子来说，如果base path打到的是完全Diffuse的物体，而offset path打到的是完美镜面，那么对于offset path来说，出射光在物理上正确的只有一种位置，就是半角向量就是当前表面法向量的情况，而Copy过来的半角向量如果不是法向量，则导致采样失败，或说是采样不到物理上正确的offset path。

Path Reconnection就是先采用上面两种策略中的任意一种对offset path进行采样，并在适当的时候，将offset path连到base path上，以保证全局的相似性。



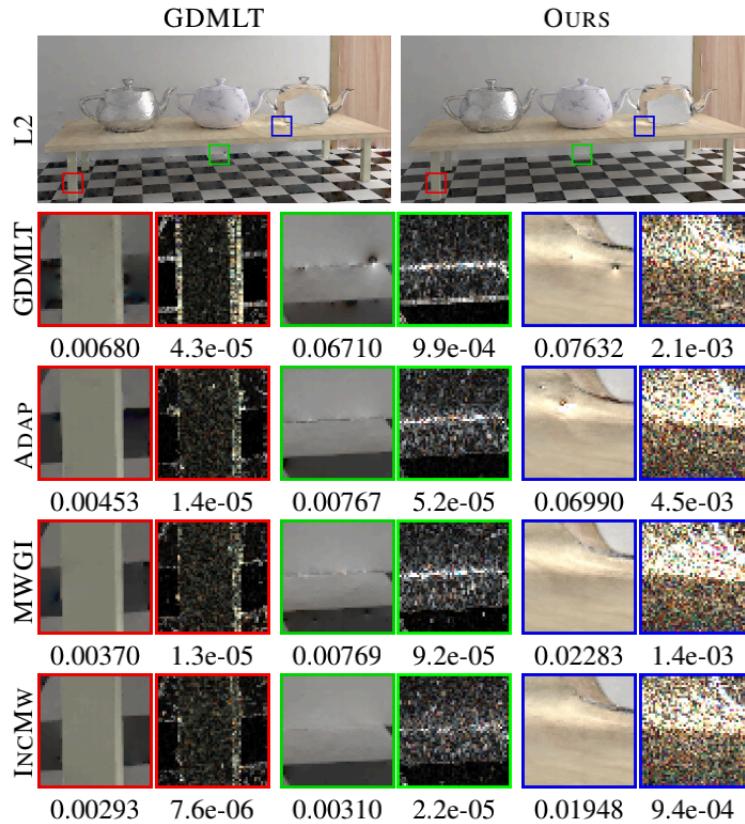
(c) Path reconnection

这样做的好处在于保证了全局的相似性，而局限在于这种方法奏效当且仅当 x_b , x_{b+1} 和 x'_{b+1} 的材质“差不多”。可以想像，如果三点的材质相差甚远，那么这样的连接会造成两种结果：

- 物理上不正确
- 虽然物理上正确，但违反了对BSDF的Importance Sampling，从而引入更多方差。

于是在使用这种策略时，我们先要根据BSDF的Roughness值对于base path的顶点进行分类，足够接近Diffuse的点分类为connectable。

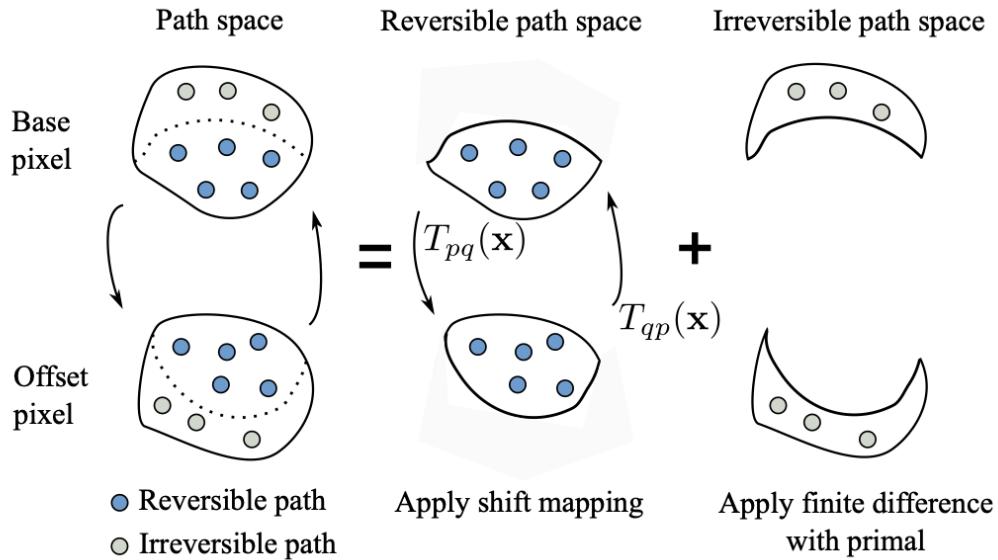
Manifold Exploration在论文[Improved Sampling for Gradient-Domain Metropolis Light Transport](#)中被提出，其基本思想其实也很简朴。它的初衷也是为了在保证全局相似性的同时，做到也保证物理的正确性且可逆。具体方法类似于Half Vector Copy，只不过在考虑offset path的下一个点时，不再仅仅考虑base path的当前点，而是考虑base path之前的所有点，保存它们的Half Vector，并把这些Half Vector作为约束条件，同作为约束条件的当前offset base点的材质一起丢给Newton Solver去求解。这样做则可以满足全局相似、物理正确、可逆三个条件，可以解决非常复杂的光线情况，比如Specular Chain。唯一的缺点就是耗时。下面一张图是论文作者对这种方法做的评测，最后一行是运用了Manifold Exploration，且采取了之前所说的根据Feature决定采样像素的策略，可以看出其比起baseline的算法有非常大的提高，尤其是在处理复杂材质时（参看最后一列）：



对各方法的优劣总结如下表：

	Random Sequence Replay	Half Vector Copy	Path Reconnection	Manifold Exploration
Similarity		T	T	T
Correctness	T			T
Inversibility	T			T
Efficiency	T	T	T	

目前在具体实现时，因为Manifold Exploration太耗时，我们又想保证物理的正确性和可逆性，所以我们通常将路径分为两部分，可逆和不可逆。对于可逆的路径，我们使用Half Vector Copy或者Path Reconnection，而对于不可逆的路径，我们使用直接做差的方法：



- Reconstruction

解决重构的问题比起以上两个部分相对简单，比较套路。总体来说有两类方法，第一类是使用 Poisson Reconstruction：

$$I = \arg \min_I \|D_x(I) - G_x\|_n^n + \|D_y(I) - G_y\|_n^n + \lambda \|I - I_0\|_n^n$$

这个方程比较好理解，D是要求解的结果图的梯度图，G是我们根据一定的采样策略和Shift-Mapping构造出来的梯度图，I是要求解的结果图，I0是低spp的原图。

当我们取n=2时，即采用L2 Poisson重构，问题等价于解如下的线性方程：

$$(D_x^\top D_x + D_y^\top D_y + \lambda)I = D_x^\top G_x + D_y^\top G_y + M_0,$$

当我们取n=1时，即采用L1 Poisson重构，其计算过程相对复杂，但结果更Robust。结果更Robust的原因是对梯度项采用一范数而不是二范数，权重更低地考虑了梯度项，从而对梯度图G的outliers(过亮的点)不再敏感。这样的方法与上述的考虑多个Feature的Sampling策略相结合，可以有效地消除图片中的Singularities。

第二种方法是Control Variates(CVs)，在论文[Image-space Control Variates for Rendering](#)中被提出。其把梯度图当成Control Variate，用一种迭代的策略求解结果图：

$$I_p^{(n+1)} = I_q^{(n)} + \langle I_p - I_q \rangle,$$

写成带权的形式为：

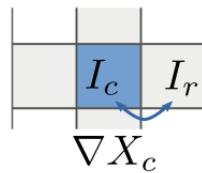
$$\langle I_p \rangle = \sum_{q \in \mathcal{N}_p} w_{p,q} (\langle I_q \rangle + \langle I_p - I_q \rangle)$$

如对几个相邻像素平均赋权，则有：

$$\begin{aligned} \langle I_c \rangle &= \frac{1}{5} (\langle I_c \rangle + (\langle I_r \rangle - \nabla X_c) + (\langle I_l \rangle + \nabla X_l) \\ &\quad + (\langle I_b \rangle - \nabla Y_c) + (\langle I_t \rangle + \nabla Y_t)). \end{aligned}$$

其中：

$$\begin{array}{ll} \nabla X_c = \langle I_r - I_c \rangle & \nabla X_l = \langle I_c - I_l \rangle \\ \nabla Y_c = \langle I_b - I_c \rangle & \nabla Y_t = \langle I_c - I_t \rangle \end{array}$$



在论文[Image-space Control Variates for Rendering](#)中，作者还给出了一种优化权值的方法，但是这种方法中的一部分是基于经验公式，导致效果并不一定最优。另外，作者还证明了Poisson重构和Control Variates是一种更广泛的重构方法的两种表现形式，这在一定程度上解释了为什么Control Variate重构出的图片看起来与用L1 Poisson重构做出的图片十分相似。同时作者还提出，CVs是一种有偏的重构方法，这也是它比起Poisson重构的劣势所在。

	CVs	L1	L2
Efficient	T		T
Robust	T	T	
Unbiased		T	T



在上图中可以看到，用Weighted CVs可以得到最好的重构效果。

- Future Work

- Sampling: 研究更有效的Feature，即改进度量像素间的距离公式，考虑可以用网络解决这个问题。
- Shift-Mapping: 个人认为这一块可改进余地是最大的，但也是最难的，首先两个path其实是两个折线段，要两个折线段“相对靠近”，怎么定义折线段之间的“近”就是一个大问题。目前的四种方法都只考虑了折线段每一段在世界坐标系的角度足够接近，但单纯角度接近是否足够？运用类似退火的算法，在局部上牺牲最优是否能换来全局最优？如何在开销更小的情况下保证物理上的正确性、可逆性和相似性三者同时成立？[A Survey on Gradient-Domain Rendering](#)指出运用Path Guiding解决Shift-Mapping的问题可能是一个好思路。
- Reconstruction: CVs已经能做到比较好的重构效果，目前的问题就是计算最优权值的公式是一个经验公式，能否有更准确，理论上更说得过去的公式？是否可以考虑用网络解决这个问题？

- Conclusions

本报告对Gradient-Domain Rendering这项技术做了相对全面，但深度较浅的介绍，尽量形象地描述了我对这项技术的理解。论文[Improved Sampling for Gradient-Domain Metropolis Light Transport](#)中指出，虽然Gradient Domain Rendering目前的效果还不如直接使用Monte Carlo降噪器，但其具有的理论重要性和未来发展的广阔空间不容忽视。目前我和组里的梁宇智学长也将主要精力放在研究这项技术上面，希望我们能做出一些的贡献来完善这项技术。