

Path length regularization

在单个 $w \in W$ 时，生成元映射 $g(w): W \rightarrow Y$ 的局部度量标度性质由 Jacobian 矩阵 $J_w = \partial g(w) / \partial w$ 捕获

$$\mathbb{E}_{w, y \sim \mathcal{N}(0, \mathbf{I})} \left(\left\| \mathbf{J}_w^T \mathbf{y} \right\|_2 - a \right)^2,$$

y 是随机图像，其像素（数值）强度服从正态分布； $w \sim f(z)$ (w 是 z 的函数)， z 是正态分布的。

我们希望鼓励 W 中的一个固定大小的步骤导致图像中的非零、固定大小的变化。我们可以通过进入图像空间中的随机方向并观察相应的 w 个梯度来经验地衡量偏离这个理想的程度。这些梯度不管是 w 还是图像空间方向，都应该具有接近相等的长度，表明从潜在空间到图像空间的映射是有条件的

由于 $y = g(w)$ ，则 $\Delta y = \partial g(w) / \partial w \cdot \Delta w$ ，即： $\Delta y = J_w \cdot \Delta w$ 同时，由于 J_w 是正交矩阵， $J_w^T \cdot J_w = E$ (E 为单位矩阵)，因此： $J_w^T \cdot \Delta y = J_w^T \cdot J_w \cdot \Delta w = \Delta w$ 。当 y 是随机图像，其像素（数值）强度服从正态分布时， w 也表现为服从正态分布。也就是说， w 向量的微小扰动，表现为图像 y 的微小变化，二者的变化模式趋同。

若基于训练数据得出的 J_w 在任何 w 向量处都正交时，则公式（左）所表达的先验（期望）是最小的，这意味着此时 w 向量的微小扰动，所带来的图像 y 的变化是最小的，即：生成器的映射平滑度为最优。

为了避免雅克比矩阵的显式计算，我们使用恒等式：

$$\mathbf{J}_w^T \mathbf{y} = \nabla_w (g(w) \cdot \mathbf{y})$$

其中，哈密顿算子使用标准的反向传播算法即可高效计算。在优化过程中，常数 a 被动态地设置为长度的长期指数移动平均值，使得优化过程可以仅凭自己就能找到合适的全局尺度。也就是说，公式（左）中的常数 a 引用的是当前优化周期得到的欧式距离长度 $\|J_w^T \cdot y\|$ 的平均值，减去 a 之后，则公式（左）所表达的先验（期望）则接近于一个均值为 0 的正态分布，该正态分布越是呈现出一种窄而高的尖峰时，则生成器映射的期望方差越小，生成的图像质量越稳定。

Inception Score

Inception Score使用图片类别**分类器**来评估生成图片的质量。其中使用的图片类别分类器为Inception Net-V3。这也是Inception Score名称的由来。

清晰度，IS 对于生成的图片 x 输入到Inception Net-V3中产生一个1000维的向量 y 。其中每一维代表数据某类的概率。对于清晰的图片来说， y 的某一维应该接近1，其余维接近0。即对于类别 y 来说， $p(y|x)$ 的熵很小（概率比较确定）。

多样性：对于所有的生成图片，应该均匀分布在所有的类别中。比如共生成10000张图片，对于1000个类别，每一类应该生成10张图片。即 $p(y) = \sum p(y|x(i))$ 的熵很大，总体分布接近均匀分布。

Inception Score的公式

IS是对生成图片清晰度和多样性的衡量，IS值越大越好

$$IS(G) = \exp(\mathbb{E}_{x \sim p_g} D_{KL}(p(y|x) || p(y)))$$

- $\mathbb{E}_{x \sim p_g}$ 遍历所有的生成样本，求平均值
- D_{KL} : KL散度， $D_{KL}(P || Q)$ 用于衡量分布 P 和 Q 之间的近似程度
- $p(y|x)$: 对于图片 x ，属于所有类别的概率分布。对于给定图片 x ，表示为一个1000维数向量。
- $p(y)$: 边缘概率，具体实现为对于所有的验证图片 x ，计算得到 $p(y|x)$ ，再求所有向量平均值。

Inception Score的问题

(1) 数据集问题

不能使用在一个数据集上训练分类模型，评估在另一个数据集上训练的生成模型

(2) Inception Score敏感性问题

使用pytorch、tensorflow、keras等不同框架下的Inception Net权值，在同样的分类精度下，计算同一个数据集的IS。IS的差别很大，仅仅由于使用的框架不同，IS分值可以相差11.5%

(3) Inception Score低的图片不一定差

如果给出一张真实的图片，但并不属于Inception Net-V3的1000分类中的任何一类。分类器无法判别，那么Inception Score分数不高，但图像是真实的。

(4) Inception Score的多样性检验有局限性

Inception Score检测生成图片是否多样，是根据生成的类别进行检验判断。如果模型输出图片，类别是平均分配的。但每一类中，图片都一样，也就是mode collapse。这种情况Inception Score是无法检测的

(5) Inception Score不能反映过拟合

如果神经网络只是单纯的拷贝训练集的图片，那么Inception Score肯定是很高的，但这样的生成模型是没有意义的。

Fréchet Inception Distance (FID)

基本思想：直接考虑生成数据和真实数据在feature层次的距离，不再额外的借助分类器。因此来衡量生成图片和真实图片的距离。

FID使用Inception Net-V3全连接前的2048维向量作为图片的feature。

FID是反应生成图片和真实图片的距离，数据越小越好。

$$FID = \|\mu_r - \mu_g\|^2 + \text{Tr}(\Sigma_r + \Sigma_g - 2(\Sigma_r \Sigma_g)^{1/2})$$

- μ_r ：真实图片的特征均值
- μ_g ：生成图片的特征均值
- Σ_r ：真实图片的协方差矩阵
- Σ_g ：生成图片的协方差矩阵
- Tr ：迹（一个 $n \times n$ 的对角矩阵 \mathbf{A} 的主对角线（从左上方至右下方的对角线）上各个元素的总和被称为矩阵 \mathbf{A} 的迹）

Fréchet Inception Distance优点

- 生成模型的训练集可以和Inception Net-V3不同
- 刷分不会导致生成图片质量变差

Fréchet Inception Distance的缺点

- FID是衡量多元正态分布之间的距离，但提取的图片特征不一定是符合多元正态分布的
- 无法解决过拟合问题，如果生成模型只能生成和训练集一模一样的数据无法检测

Stylegan的缺陷与改进方向

1.主要改进均集中在生成器方面，没有对判别器进行改进，作者认为对判别器的改进可能会进一步提升等变性。

2.对一些特别的数据集训练起来非常困难，比如黑白卡通图片，只有黑色或白色的像素，还有一些图片质量比较低的数据集。

3.计算资源消耗巨大，stylegan3消耗了92个GPU年（V100）和225Mwh的电力，如何节约训练成本也是一个问题。

Survey on 3D face reconstruction

Why we need 3D face reconstruction?

由于人脸的3D性质，一幅2D图像不足以准确捕捉其几何形状，因为它坍塌了一个维度。此外，3D成像提供了对姿态和光照不变的面部几何结构的表示，这是2D成像的两个主要缺点。

How to do 3D face reconstruction ?

3D - from - 2D重建方法成功的关键是加入先验知识来解决解中的模糊问题。在过去的十年中，我们可以区分出三种加入这种先验信息的策略，即**统计模型拟合 (statistical model fitting)**、**光度立体视觉 (photometric stereo)** 和**深度学习(deep learning)**。在第一种方法中，先验知识被编码在三维人脸模型中，该模型由一组三维人脸扫描数据构建，并与输入图像拟合。第二种方法是将3D模板人脸或3D人脸模型与光度立体方法相结合来估计人脸表面法线。该策略下的方法一般使用多幅图像的信息，进一步约束了问题。在第三种方法中，2D - 3D映射是通过深度神经网络实现的，在给定合适的训练数据的情况下，可以学习到关联人脸几何和外观所需的先验。

统计模型拟合 (statistical model fitting)

1. Construction of a 3D facial model

1.1 3D Morphable Models(3DMM)

前提所有三维人脸是已经进行稠密对齐(3D face registration)的，即所有的三维人脸都能用相同的点云数或面片数来表示，且相同序号的点代表相同的语义。在稠密对齐的情况下，文章提出每个带纹理的三维人脸都能用如下公式表示：

$$S_{model} = \bar{S} + \sum_{i=1}^{m-1} \alpha_i S_i, \quad T_{model} = \bar{T} + \sum_{i=1}^{m-1} \beta_i T_i,$$

其中S和T表示人脸平均形状和平均纹理部分，每张人脸的判别特性体现在加号右边一组基的线性组合， S_i 和 T_i 分别表示按照特征值降序排列的协方差阵的特征向量。这也是3DMM的关键。在文章中，作者采集了200个三维人脸的头部，用这组数据作为PCA（主成分分析）的基础，来获得能代表人脸形状和纹理信息的主成分即 S_i 和 T_i 。

求解过程，实际就是对参数 α_i 和 β_i 的求解。在文中作者告诉我们这两组系数的分布满足，

$$p(\alpha) \sim \exp\left[-\frac{1}{2} \sum_{i=1}^{m-1} (\alpha_i / \sigma_i)^2\right] \quad \text{其中} \sigma_i \text{是形状协方差阵的特征值。满足多元正态分布。} \beta \text{同理。}$$

3DMM的缺陷

虽然现有的大多数3D统计人脸模型都是基于上述过程实现的，但是这种方法有两个局限性已经被一些研究者注意到。首先，PCA估计的基向量对输入数据进行全局建模，因此无法捕捉诸如皱纹等细微信息，因此通过拟合3DMM重建面部细节成为一项艰巨的任务。一些工作强调了局部建模面部形状的重要性，并提出了不同的方法。

Neumann等和Ferrari等提出通过施加稀疏分量来分解训练形状的矩阵。布伦顿等人对每个训练形状应用小波变换，获得表面的多尺度分解，并在估计的小波系数上计算局部多线性模型。Jin等使用非负矩阵分解(NMF)，因为它将形状分解为局部特征。最后，Lüthi等人利用高斯过程对形状变化进行建模，提供了一种将局部模型添加到全局模型中的方式，从而结合了多个尺度上的信息。

3DMMs的第二个缺点，即面部形状变化不是完全线性的，因此不能用线性模型精确建模。一些解决方法是学习面部形状的潜在空间，Ranjan等和Bouritsas等建模了单个潜在空间中的所有形状变化，而Jiang等估计了两个分离的潜在空间，一个对应于与身份相关的变形，另一个对应于与表情相关的变形。Ranjan等和Jiang等使用谱卷积算子，Bouritsas等提出了一种使用各向异性滤波器的螺旋卷积，允许顶点的邻域和局部滤波器的参数之间进行一对一的映射。

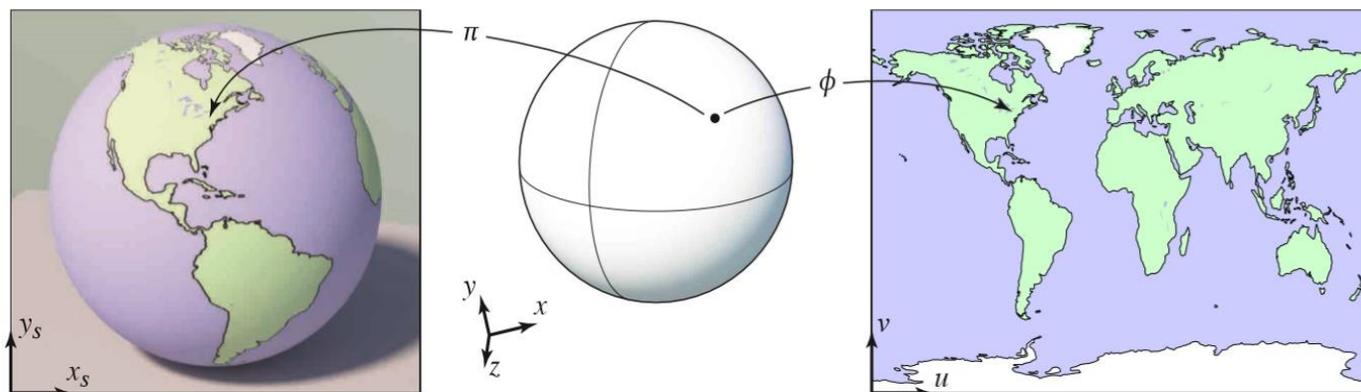
3DMMs

3DMM: Blanz和Vetter 构建了一个具有200个激光扫描年轻成年人头部的可变形模型(男100例,女100例)。他们将训练集的3D人脸使用基于2D中3D人脸平坦化到UV空间的光流算法进行点对点对应。每个3D人脸通过双射映射与一个2D圆柱参数化相关联。

UVmap基本概念

"UV"这里是指u,v纹理贴图坐标的简称(它和空间模型的X, Y, Z轴是类似的)。它定义了图片上每个点的位置的信息。这些点与3D模型是相互联系的,以决定表面纹理贴图的位置。UV就是将图像上每一个点精确对应到模型物体的表面。在点与点之间的间隙位置由软件进行图像光滑插值处理。这就是所谓的UV贴图。

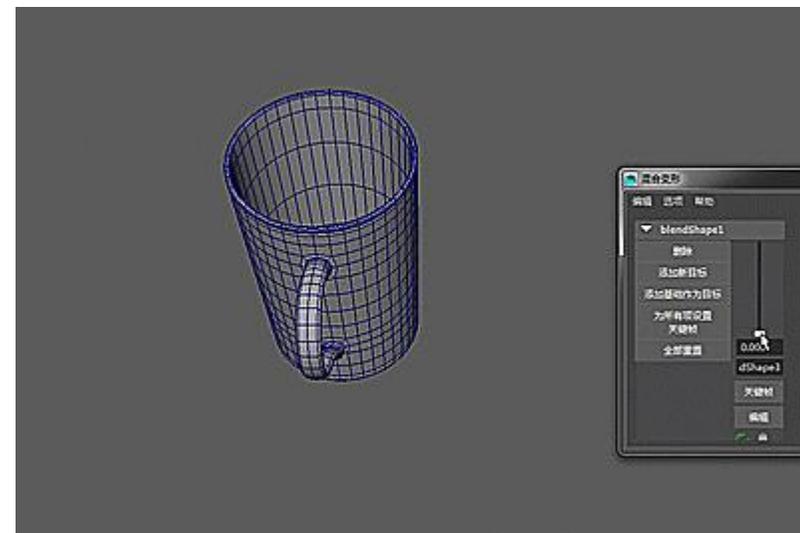
简单的来说,就是将三维物体上的点映射到2维空间上。



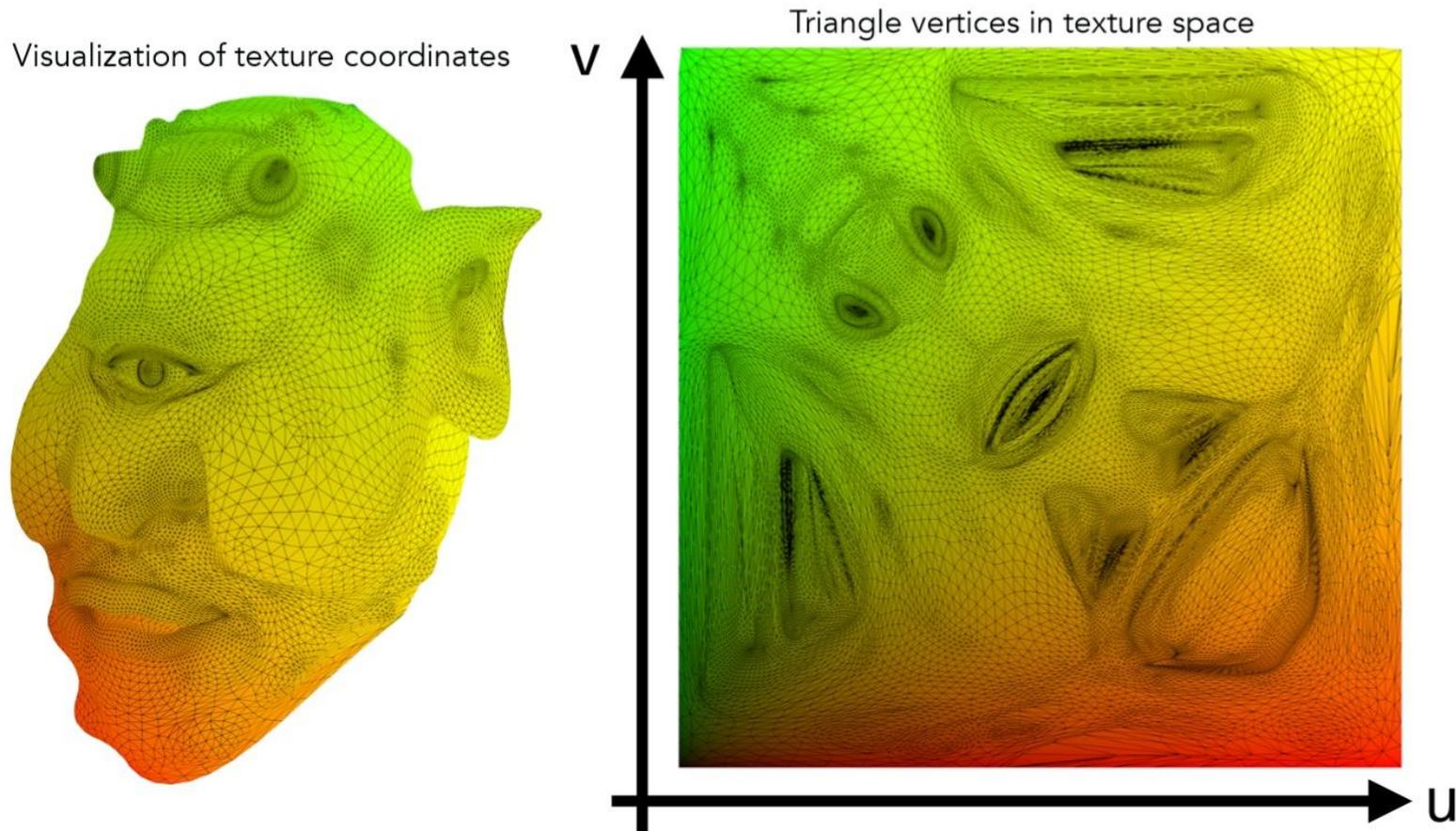
Screen space

World space

Texture space



UV坐标为纹理坐标，在纹理空间中，任意一个二维坐标都在【0, 1】内。

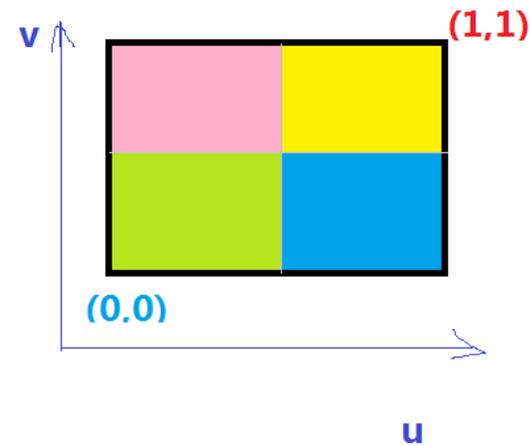


Mesh用于描述物体的形状。使用三角面来存储。

储存的信息：

- 1.顶点。每个三角形的三个顶点。
- 2.边。每个三角形的边。
- 3.面。每个三角形为一个面。

映射就是通过 (u, v) 坐标去查找对应的颜色值，例如下图：



纹理坐标的U和V的最大值均为1，每张纹理图Texture map的 (u, v) 坐标均可存储在三维Mesh的各个顶点信息中，这个存储操作即为两者之间的映射关系。在UV坐标系（或者UVW坐标系）中，显示器水平方向是U（大小0-1），显示器垂直方向是V（大小0-1），（W的方向垂直于显示器表面）纹理的映射是将纹理图贴到多边形表面的过程。

Other 3DMMs

BFM:BFM同样由100名8 - 62岁的女性和100名男性被试组成，平均年龄24.97岁。与Blanz和Vetter的3DMM相比，BFM的一个技术改进是扫描仪，它能够在更短的时间内以更高的分辨率和精度捕获面部几何结构。Gerig等人对BFM进行了扩展，在Opfer中引入了以平均变形为零函数、多尺度B样条核为变形基的高斯过程变形模型，代替NICP建立了稠密对应关系。

LSFM:由9,663个个体组成，涵盖了广泛的年龄(图2)、性别(48%男性)和种族(白人占82%，亚裔占9%，混血占5%，黑人占3%，其他民族占1%)，从所有拟合的初始PCA模型中，错误的对应关系可以通过表现为异常值的相应形状向量来识别。剔除异常值后对训练集进行主成分分析，得到最终模型。

FaceWarehouse: 与之前的所有模型只捕获由于形状引起的几何变化不同，其他工作将与表情变化相关的面部形变从身份中单独建模。Cao等从150个7 ~ 80岁个体的20个表情的深度图中构建了FaceWarehouse模型。为了得到稠密对应的网格，将Blanz和Vetter面模型拟合到深度图中。然后，通过对矢量化网格构造的3阶数据张量(顶点 \times 身份 \times 表达式)应用高阶奇异值分解(HOSVD)构建双线性人脸模型。

Statistical model fitting methods

利用统计模型对三维人脸结构的先验知识进行编码，通过寻找最适合给定二维图像的模式库的线性组合，可以从一张或多张照片中重建出新的三维人脸。本质上，将三维人脸模型拟合为二维图像意味着优化一个非线性的代价函数。

从一张新的图像中重建一个人的三维面部几何和外观比使用多张图像更具挑战性。这就是为什么大多数工作主要集中在从单幅图像重建三维人脸。即便如此，一些研究者通过观测不同姿态和光照条件下的被试人脸，考虑了多张图像来提高重建精度。其他人也已经提出使用2D视频序列作为从同一个人获得一组图片的简单方式。然而，在使用视频时，必须考虑帧与帧之间的时间关系。

3DMM拟合工作中确定的三种不同方法：代价函数的非线性优化（non-linear optimisation of a cost function）、线性方法（linear approaches）和概率方法（probabilistic approaches）。

代价函数的非线性优化 (non-linear optimisation of a cost function)

Blanz和Vetter (3DMM的提出者) 不仅将3DMMs引入, 而且还提出了一种将3DMMs拟合到单张人脸图像的方法, 表明了一种将其扩展到多张图像的方法, 这启发了很多人。他们的拟合过程基于一个简单的想法: 如果输入图像和从拟合的三维人脸渲染的图像相似, 则三维重建可以认为是输入图像好的拟合。具体来说, 他们优化了形状 (shape) α 和反射率 (albedo) β 参数以及一组渲染参数 ρ (包括摄像机内参数、投影参数以及光照参数), 使得他们生成的图像 I_{mod} 尽可能接近输入 I_{in} ; 即最小化输入图像和重建图像的每个像素 (x, y) 之间的欧氏距离之和:

$$\mathbb{E}_{\mathbf{I}}(\alpha, \beta, \rho) = \sum_{x,y} \|I_{\text{in}}(x, y) - I_{\text{mod}}(x, y)\|_2^2. \quad I_{\text{mod}} \text{采用Phong反射模型渲染}$$

仅通过最小化输入图像和渲染图像之间的距离, 由于最小化是 (ill-posed) 的, 因此无法保证面形表面。为了解决这个问题, Blanz和Vetter利用模型系数的先验概率在成本函数中加入了参数的正则化项

$$\mathbb{E}_{\text{reg}}(\alpha, \beta, \rho) = \sum_{i=1}^{\bar{M}_x} \frac{\alpha_i^2}{\sigma_{\alpha_i}^2} + \sum_{i=1}^{\bar{M}_c} \frac{\beta_i^2}{\sigma_{\beta_i}^2} + \sum_j \frac{(\rho_j - \bar{\rho}_j)^2}{\sigma_{\rho_j}^2}.$$

后来, , 他们对他们的工作进行了扩展, 在重建的面部形状中增加了一个固定的面部特征点(地标)集合 $\{\ell_i^{3D}\}_{i=1}^L$, 使得它们的投影 $\hat{\ell}_i^{2D}$ 位于输入图像中相应的人工标注的地标集合 ℓ_i^{2D} .

$$\mathbb{E}_{\text{lmk2D}}(\xi) = \sum_{i=1}^L \|\ell_i^{2D} - \hat{\ell}_i^{2D}\|_2^2.$$

Redefining the image error

与上述工作不同的是，上述工作是测量了输入图像与渲染图像之间的欧氏距离。一些工作完全重新定义了图像误差：有人利用图像的梯度计算图像误差，有人则在特征空间中度量图像之间的差异。特别地，Sariyanidi等计算了输入图像和渲染图像之间的梯度相关性， $E_I = \nabla I_{in}^T \nabla I_{mod}$ ，他们认为该方法对光照变化和遮挡具有更强的鲁棒性。

作者强调，在成本函数中加入正则化项。可能导致不满意的结果，要么是因为过度平滑的3D人脸，要么相反，因为正则化在最小化问题上的权重由一个被选择的参数控制。

Booth等人使用基于稠密特征的表示构建纹理模型，而不是使用三维人脸的每个顶点的RGB值。用于构建纹理模型的训练样本为图像特征向量 f_i ，通过对训练集中的每幅图像应用稠密特征提取函数得到。然后，假设每张训练图像的形状参数 α （因此，已知三维人脸形状， $x(\alpha)$ ）和投影参数 ζ 已知，根据 ζ 计算三维人脸 $x(\alpha)$ 在图像平面 $P(x(\alpha), \zeta)$ 上的投影，并在三维人脸投影的位置上采样特征向量。因此，对于每一张训练图像，他们得到一个由对应投影三维人脸位置上的特征向量组成的向量 f_i ，图像误差 E_I 被重新定义为

$$E_I(\alpha, \beta, \zeta) = \|F(P(x(\alpha), \zeta)) - f(\beta)\|_2^2,$$

其中 F 是从输入图像中对应于投影重建三维人脸 $P(x(\alpha), \zeta)$ 位置的像素中提取特征的函数， $f(\beta)$ 是对应于参数 β 的基于特征的纹理模型的实例。

Multi-stage optimisation

在不同阶段通过最小化代价函数来重建三维人脸形状。有研究者提出分阶段重建3D人脸，每个阶段逐步添加细节以细化前一个阶段，另有研究者训练了一个级联回归器，将3D人脸重建过程拆分为对受试者面部几何形状的估计和对包括光照在内的外观的估计。

$$\mathbb{E}_{\text{Imk2D}}(\zeta) = \sum_{i=1}^L \left\| \ell_i^{2D} - \hat{\ell}_i^{2D} \right\|_2^2.$$

在由粗到精的类别中，Qu等使用多幅图像细化了通过最小化路标误差得到的单一意象重建。相比之下，将拟合过程分为全局和精细重建两个阶段。Jin等和Shi等通过最小化特征点误差，然后，他们从3DMM出发，通过使用图像误差 $\mathbb{E}_I(\alpha, \beta, \rho) = \sum_{x,y} \|\mathbf{I}_{\text{in}}(x, y) - \mathbf{I}_{\text{mod}}(x, y)\|_2^2$ ，并增加一个强制精细化形状与粗化形状相似的项。表面法线的细化允许他们以显著的精度获得详细的重建。尽管两个工作遵循非常相似的方法，但主要区别在于使用的模型类型：shi等使用了基于PCA的模型，Jin等人使用了基于非负矩阵分解的人脸模型。

Bas等人遵循由粗到精的策略，首先以线性方式估计形状和投影参数，然后用非线性代价函数对其进行精化。在线性阶段，通过求解线性方程组来估计形状 α 和投影 $\zeta = \{R, \tau, s\}$ 参数 $\underline{p}_i = \mathcal{P}(\mathbf{v}_i, \zeta)$

对于三维人脸中的一组对应顶点 $\mathbf{v}_i \in \mathbb{R}^3$ 和图像中的像素点 $\mathbf{p}_i \in \mathbb{R}^2$ ，其中 $\mathcal{P}(\mathbf{v}, \zeta)$ 为 $\mathbf{v} \in \mathbb{R}^3$ 根据 ζ 向图像平面的弱透视投影。为了进一步约束重建，他们计算了纹理边缘与三维人脸遮挡边界之间的对应关系，定义为位于网格边缘的顶点集合，其相邻的网格面具有可见性变化。这样，由于在人脸边界处得到了密集的对对应关系，从而更准确地恢复了人脸的全局结构。然而，由于内层人脸中使用的特征点构成了一个稀疏集，不足以捕捉细微的细节，因此精细的细节很难被捕获。

Linear approaches

前面提到的所有工作都利用了非线性误差函数的优化，类似于最初由Blanz和Vetter提出的方法。然而，Aldrian和Smith观察到Blanz和Vetter提出的优化问题是高度复杂的，计算代价高昂，并且由于 (ill-posed) 容易陷入局部极小。为了避免这种情况，他们提出了一种替代方案：以线性方式拟合3DMM，分别用于投影参数、面部几何和纹理。与以非线性的方式改进线性阶段的重建结果不同，Aldrian和Smith提出了一种完全线性的方法。

具体来说，Aldrian和Smith通过求解线性方程组来估计投影参数。 $\underline{p}_i = \mathcal{P}(\underline{v}_i, \zeta)$ 为一组对应的3D - 2D地标，并假设一个仿射相机模型，这允许他们重写方程。公式变为 $\underline{p}_i = \mathbf{C}\underline{v}_i$ ，其中 $\mathbf{C} \in \mathbb{R} (3 \times 4)$ 为相机投影矩阵。然后，给定图像 \mathcal{L}^{2D} 中的二维地标集合 $p(\alpha | \mathcal{L}^{2D})$ ，通过最大化形状参数 α 的后验概率来估计形状模型参数， $p(\alpha | \mathcal{L}^{2D})$ 是可微的，从而得到关于 α 的线性方程组。最后，为了估计纹理，他们假设白光照明和朗伯反射模型(漫反射)。这些假设允许他们通过将输入图像中颜色通道对之间的比例与三维人脸对应顶点的纹理比例相等来计算纹理。

Ding等认为，Aldrian和Smith提出的线性估计几何形状的方法存在两个主要缺陷：一是在相机投影矩阵估计中，没有考虑旋转矩阵的正交性；二是以两步的方式分别估计相机和形状参数，并不能保证每一步都减少代价函数。为了解决第一个缺点，Ding等人提出对Aldrian和Smith算法得到的相机投影矩阵进行缩放和正交化，以增强旋转矩阵的正交性。第二个缺点通过联合估计相机投影矩阵和形状参数来解决。

Probabilistic approach

Schönborn等提出了一种非常不同的拟合3DMM的方法，他们将拟合3DMM的过程重新建模为概率推断问题。他们的方法是在给定输入图像 lin ， $p(\theta | lin)$ 的情况下，从包含3DMM参数(形状 α 和纹理 β)、光照和投影参数 ζ 的所有参数 θ 的集合上的后验分布中抽取样本。

他们采用Metropolis - Hastings算法，通过随机接受或拒绝样本来抽取参数 θ 按照后验分布 $p(\theta | lin)$ 分布的样本。这种后验分布是通过贝叶斯定理 $p(\theta | lin) \propto p(lin | \theta) p(\theta)$ 来计算的。

第一项 $p(lin | \theta)$ 是高斯分布，考虑到前景和背景像素的不同分布，每个像素独立处理。在输入图像的所有背景像素上训练背景分布，前景分布由3DMM已知。在第二项 $p(\theta)$ 中，他们综合所有的信息来源来估计参数的分布。基本上，他们观察到，给定一个完全自动化的pipeline，其中需要一个人脸检测器和一个地标检测器；由于强烈的姿态和光照变化，两个探测器都被迫做出可能不可靠的早期决策。

因此，他们将人脸和地标检测器的结果作为检测图，为图像中的每个像素分配在该位置有人脸(或者特定的地标)的可能性。这样，对于每一个检测到的人脸 box_i ， $p(\theta)$ 与框的位置和大小以及框 L_i 的地标检测图 $p(\theta | box_i, L_i)$ 是有偏的。然后，将所有的候选分布(每个检测到的人脸一个)组合起来，通过计算人脸框分布的均值来构造一个全局分布 $p(\theta | allboxes, L)$ ，该分布包含了所有检测的知识。

Local approaches

大多数提出的重建一个人的三维人脸的技术都以全局的方式来处理问题，将整个面部模型拟合为图像。然而，Ding等和Piotraschke和Blanz在考虑面部子区域的情况下，以局部的方式处理拟合过程，尽管方式略有不同。前者从同一输入图像中局部重建每个子区域，确保每个子区域都是最优拟合的，而不考虑其他区域。相比之下，Piotraschke和Blanz从不同的输入图像中重建全局3D人脸，然后根据每个子区域的精度将重建结果进行融合。

具体而言，Ding等得到了K个形状参数向量 α_k ，分别对其中的一个子区域进行优化。然后，将形状参数向量与依赖于顶点的混合权重进行线性组合： $x = \bar{x} + \sum_{k=1}^K w_k (\Phi \alpha_k)$ ，其中 $\sum_{k=1}^K w_{kj} = 1$ ，j为顶点。通过这种方式，他们增加了3DMM的灵活性，并确保每个区域都是最佳拟合的，从而获得更准确的重建结果。

另一方面，如上所述，Piotraschke和Blanz使用了一个人的多个图片，并使用从每个人身上重建了个人的3D形状。对每个子区域进行局部评估，获得每个3D人脸的每个子区域的质量度量。这些措施通过计算最佳子区域的加权线性组合来获得每个子区域的单次重建，最终合并成单个3D人脸。因此，其主要贡献是在不需要真实人脸的情况下，通过合并几个基于质量的次优3D人脸重建来提供更精确的3D人脸。