

# 自然图像 matting 的闭式解-TPAMI

Dezeming Family

2023 年 2 月 10 日

DezemingFamily 系列文章和电子书**全部都有免费公开的电子版**，可以很方便地进行修改和重新发布。如果您获得了 DezemingFamily 的系列电子书，可以从我们的网站 [<https://dezeming.top/>] 找到最新的版本。对文章的内容建议和出现的错误也欢迎在网站留言。

## 目录

一 介绍	1
二 公式导出	1
三 约束和求解	1
四 $\epsilon$ 参数	2
五 光谱分析	2
六 恢复前景和背景	3
参考文献	3

# 一 介绍

在《自然图像 matting 的闭式解-CVPR》中我们描述了发布在了 CVPR2006 上的本文 [1]，后来，在 2008 年又发布到了 TPAMI 期刊。本文的公式序号是根据 TPAMI 版本来描述的，但只是对《自然图像 matting 的闭式解-CVPR》进行一些补充。TPAMI 中，用户输入的涂鸦可以不止使用黑白涂鸦，还可以在边界处绘制其他颜色的涂鸦来表示此处是边界区域。

# 二 公式导出

我们定义公式 (10) 所述的损失函数：

$$J(\alpha, a, b) = \sum_{j \in I} \left( \sum_{i \in w_j} \left( \alpha_i - \sum_c a_j^c I_i^c - b_j \right)^2 + \epsilon \sum_c a_j^c \right). \quad (10) \quad \sum_{k|(i,j) \in w_k} \left( \delta_{ij} - \frac{1}{|w_k|} (1 + (I_i - \mu_k)(\Sigma_k + \frac{\epsilon}{|w_k|} I_3)^{-1} (I_j - \mu_k)) \right), \quad (12)$$

$$J(\alpha) = \alpha^T L \alpha. \quad (11)$$

除了数学证明外，我们的损失函数背后的直觉是，matte 可以局部表示为图像颜色通道的线性组合，如下图所示的三个代表性示例所示：

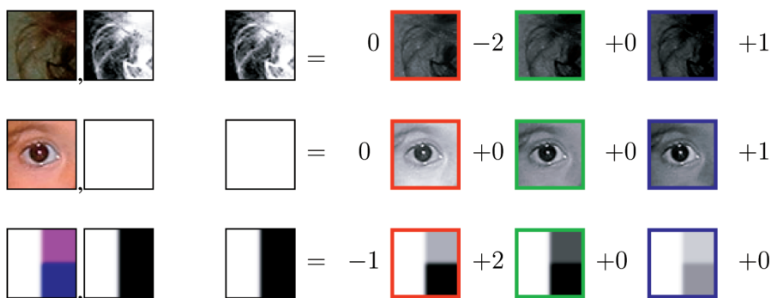


Image and matte pair      Expressing the matte as a linear combination of color channels

第一个例子是一个前景和背景颜色非常一致的窗口。在这种情况下，alpha matte 与图像具有很强的归一化相关性 (normalized correlation)，可以通过将其中一个颜色通道乘以比例因子并添加一个常数来生成。

在第二个示例中，alpha 在整个窗口上是恒定的。无论此窗口中图像纹理的复杂性如何，我们都可以通过将图像通道乘以零并添加一个常数来获得常数 alpha。这个微不足道的例子很重要，因为它展示了 4D 线性模型的一些威力。由于典型的 matte 在大多数图像窗口上是恒定的（零或一），因此无论精确的颜色分布以及颜色线模型是否成立，这些窗口中的遮片都可以以平常的方式表示为图像的线性函数。

最后，我们给出了一个具有非均匀 alpha 的窗口，其中，背景包含一个边缘。由于不同颜色通道中的边缘对比度不同，通过适当缩放颜色通道，我们的模型能够实际消除背景边缘。

该损失函数 (12) 可以化简消去  $a^c$  和  $b$ ，得到公式 (11)。其中  $L$  是 matting Laplacian matrix，是一个  $N \times N$  的矩阵 ( $N$  不是窗口长宽，而是整幅图像的像素数)，每个元素的表示见公式 (12)。可见  $L$  每一行相加都是 0，因此其零空间包含常向量（这里的常向量，即向量里每个元素值都相同。由于每行相加都是 0，所以每行中，每个元素都乘以相同的数再相加，值也是 0）。

既然零空间包含常数向量，也就是说，假设这个常数向量是  $\mathbf{v}$ ，那么  $L\mathbf{v} = \mathbf{0}$ ，则  $\mathbf{v}^T L\mathbf{v} = \mathbf{0}$ ，优化到了最低结果。

如果  $\epsilon = 0$ ，那么  $L$  的零空间就包含了每个  $I$  的颜色通道，因为每个颜色通道都可以表示为自身的线性函数（它的含义我也不是非常了解，意思应该是， $L\mathbf{v} = \mathbf{0}$ ，我们的目标  $\alpha = \mathbf{v}$  得到的  $\alpha$  可以是原图中的三个通道中的任意一个）。

# 三 约束和求解

求解问题，公式 (13) 相比于之前 CVPR 的版本，把硬约束变为了软约束：

$$\alpha = \operatorname{argmin} \alpha^T L \alpha + \lambda(\alpha^T - b_S^T) D_S (\alpha - b_S), \quad (13) \quad (L + \lambda D_S) \alpha = \lambda b_S. \quad (14)$$

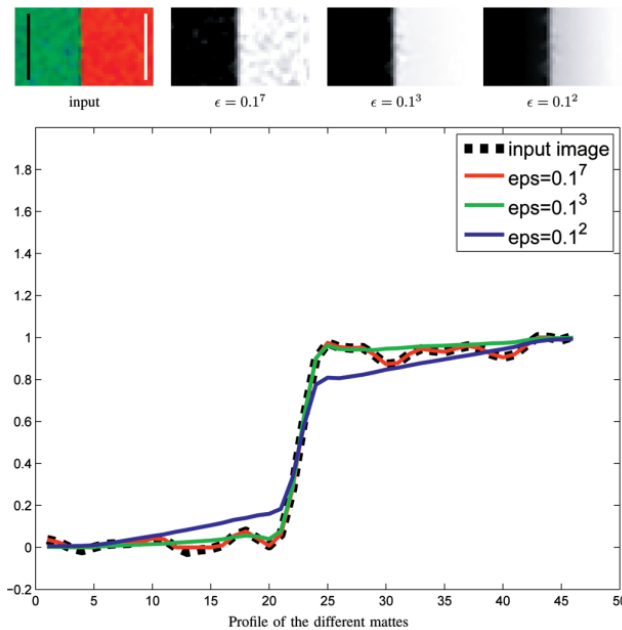
$D_S$  是一个对角矩阵，对于有约束的像素，对角值是 1，其他像素是 0（所以这也是  $N \times N$  的矩阵）。这意味着与 CVPR 原文中的硬约束（保证我们的涂鸦区域设定的值一定满足求解出的 matte 值）不同，这里变成了软约束。 $b_S$  表示一个  $N$  维向量，在指定  $\alpha$  值的区域（涂鸦区）为指定的相应值，在其他区域为 0。有人可能会好奇，如果我们用户指定的某区域  $\alpha$  为 0 该怎么办？其实这就是  $D_S$  的作用了，它把用户指定的区域和没有指定的区域根据其元素值来区分开了。

由于上述损失函数对于  $\alpha$  是二次的，所以可以对公式 (13) 进行微分，取导数为 0，即得到求解公式 (14) 这个稀疏线性系统。

因此，使用软约束使得我们可以很容易地通过多种笔刷进行涂鸦。当然我认为我们还可以进一步改进，比如对设置绝对背景和绝对前景的区域在  $b_S$  矩阵中使用更大权重，这样应该能更好地实现效果。

## 四 $\epsilon$ 参数

$\epsilon$  参数可以控制解的噪声量和平滑度：



输入是两个有一定噪声的区域，并有两个竖线涂鸦，一白一黑。在求解出的图像 mattes 中，我们取图像中间一行，将 mattes 可视化出来。由图看出，不同的  $\epsilon$  值得到不同的 mattes，值越低，绘制出的 mattes 会越接近原图，mattes 会捕捉到图像噪声；值越高，mattes 越平滑，但是可能会过于平滑。

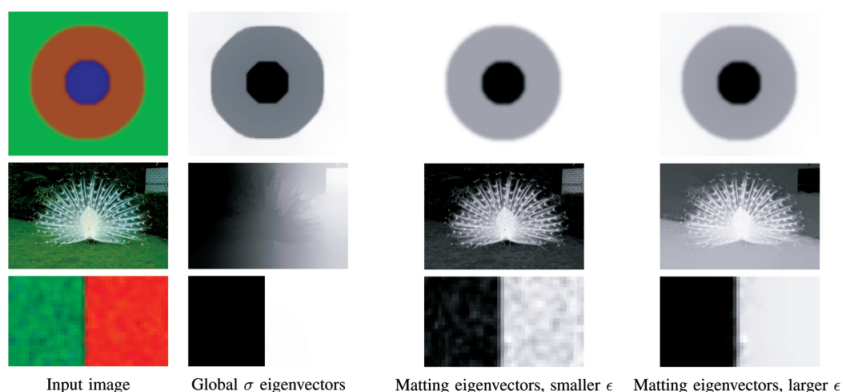
## 五 光谱分析

比较两种分割算法，一种是 normalized cuts（基于 graph Laplacian 的分割算法），另一种是本文的 matting Laplacian。由于我们的 matting Laplacian 也是对称正定的，所以也能写为  $L = D - W$  的形式（有些解释放在了这篇论文 [4] 中）。normalized cuts 和 matting Laplacian 中的  $W$  矩阵分别表示为公式 (15) 和 (16)：

$$W_G(i, j) = e^{-\|I_i - I_j\|^2 / \sigma^2}, \quad (15) \quad W_M(i, j) = \sum_{k|(i,j) \in w_k} \frac{1}{|w_k|} \left( 1 + (I_i - \mu_k) \left( \Sigma_k + \frac{\epsilon}{|w_k|} I_3 \right)^{-1} (I_j - \mu_k) \right). \quad (16)$$

我们把  $W$  叫做亲和性 (affinity) 函数，尤其是我们的亲和性函数，叫做 matting affinity，虽然亲和性函数一般是非负的，但是我们的 matting affinity 可能是负数。因为  $W_G$  和  $W_M$  的特征向量被光谱分割算法用于分割图像，我们测试一下对应的特征值。注意  $W_G$  使用了全局的统一的  $\sigma$  值。

由于最小的特征值对应的特征向量是常数向量，我们显示第二小的特征值：



由此看出，当  $\epsilon = 0.1^5$  时，分割效果很好。当  $\epsilon = 0.1^7$  时，特征图非常接近于原图，并继承了原图的一些噪声。硬分割中，使用精心选择的  $\sigma$  值可以得到很好的效果，但不适用于 matting。

注意在特征图中，如果区域颜色相近，那么 matte 值很容易在这个区域传播；如果区域颜色不相近，那么 matte 值就不容易在这个区域传播。比如同样是前景，但是包含了人的头发和皮肤，那么最好要把头发和皮肤都进行一些涂鸦，这跟这样一个事实有关：如果  $\epsilon = 0$ ，则  $L$  的零空间包含了原图所有三个颜色通道。零空间又可以理解为即特征值为 0 的特征向量构成所在的空间，对于  $L$  这个半正定矩阵而言，其特征值都是非负的，因此为最小的特征值对应的特征向量，当然我们前面也知道了，当  $\epsilon$  足够小的时候，计算得到的 matte 值（再提一句，这是  $N$  维的）就会越接近于原图值，所以当设置为 0 时，自然零空间包含 RGB 三个通道也不足为怪。对不同颜色的区域都进行标注，也是有助于这个系统 matting 效果更好。

更准确地说，alpha matte 可以通过检查 matting Laplacian 的一些较小特征向量来预测，因为 (13) 的最优解将在很大程度上由较小特征向量覆盖。下图可视化了最小的两个特征值对应的特征向量（注意每张图都是一个特征向量，把值归一化为  $(0 - 1)$  之间），可以看到，对于背景区  $\alpha \approx 0$ ，符合最小特征向量图中值非常低的位置；对于前景区  $\alpha \approx 1$ ，符合最小特征向量图中值非常高的位置；对于过渡区，其值介于前景  $\alpha$  值和背景  $\alpha$  值之间：



事实上，有可能将最优解中较大特征向量的权重限制为对应特征值比率的函数，文中描述的“较大的特征向量”和“较小的特征向量”，这里应该是表示“较大的特征值对应的特征向量”，这里我不太明白作者的意思，究竟是要限制什么的权重，不清楚是不是公式 (13) 里的  $D_S$  中的对应像素的权重。更详细的解释在论文的 Theorem 4 和 Corollary 1 中。

## 六 恢复前景和背景

求出 mattes 以后，就可以通过最小化损失函数 (19) 来求解得到  $\mathbf{F}$  和  $\mathbf{B}$ ：

$$\alpha_i \approx \sum_c a^c I_i^c + b, \quad \forall i \in w, \quad (9) \quad \min \sum_{i \in I} \sum_c (\alpha_i F_i^c + (1 - \alpha_i) B_i^c - I_i^c)^2 \quad (19)$$

$$+ |\alpha_{i_x}| \left( (F_{i_x}^c)^2 + (B_{i_x}^c)^2 \right) + |\alpha_{i_y}| \left( (F_{i_y}^c)^2 + (B_{i_y}^c)^2 \right),$$

仍然是二次损失函数，用求解线性系统的方法就能够得到最优解。

## 参考文献

- [1] Levin, A. . "A closed form solution to natural image matting." IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), 2006 IEEE Computer Society, 2006.
- [2] Radke R J . Computer vision for visual effects. Cambridge University Press, 2010.
- [3] Szeliski R . Computer Vision: Algorithms and Applications[M]. Springer-Verlag New York, Inc. 2011.
- [4] Shi, Jianbo, and Jitendra Malik. "Normalized cuts and image segmentation." IEEE Transactions on pattern analysis and machine intelligence 22.8 (2000): 888-905.