

# 2021.10.31周报

纪新龙

- 2021.10.31周报
  - 本周计划任务
  - 本周任务完成情况
  - 下周计划任务
  - 具体完成情况
    - 基于L0稀疏优化的本征图像分解
      - 论文的亮点
      - 论文的目标优化函数
      - 论文的模型优化求解
    - 基于多尺度度量和稀疏性的本征图像分解
      - 多尺度度量的思想
      - 多尺度度量的构建过程
      - 目标函数的求解

## 本周计划任务

阅读三篇论文：

4\_丁守鸿\_基于结构分析的可视媒体智能处理技术研究

5\_Ding2017\_Intrinsic\_image\_estimation\_using\_near-L0\_sparse\_optimization

6\_Ding2017\_Intrinsic\_Image\_Decomposition\_Using\_Multi-Scale\_Measurements

## 本周任务完成情况

基本完成。

## 下周计划任务

复现“基于L0稀疏优化的本征图像分解”方法。

# 具体完成情况

论文4\_丁守鸿\_基于结构分析的可视媒体智能处理技术研究是丁守鸿博士的博士毕业论文，里边主要详述了他在读博期间做的研究：图像压缩、本征分解、人脸配准。其中本征分解部分，他提出了两种本征分解算法，第一种方法**基于L0稀疏优化的本征图像分解**恰好是论文

5\_Ding2017\_Intrinsic\_image\_estimation\_using\_near-L0\_sparse\_optimization中的方法，特点是手动构建了一个近似的L0约束函数来约束反射率层，是反射率层更接近理想的**分段常数**先验状态。

而第二个方法**基于多尺度度量和稀疏性的本征图像分解**也恰好是

6\_Ding2017\_Intrinsic\_Image\_Decomposition\_Using\_Multi-Scale\_Measurements的内容，是在前面第一个算法的基础上又引入了**Multi-scale measurements (多尺度度量)**，简称**MMC**。这个MMC稍微复杂一点。实际上是一个逐渐将像素按照反射率的相似度进行聚类、再聚类的过程。每一次聚类都是在根据当前**代表像素**（代表像素是上一次聚类后幸存的像素，他们代表着在之前多次聚类中被去除掉的特征上有冗余的像素）之间的反射率相似度排除掉一半的冗余像素，留取一半作为新的代表像素，然后根据数量减半之前所有像素之间的**反射率相似度、位置特征、亮度特征、色度特征**重新计算出新的位置特征、亮度特征、色度特征值，接着再糅合新的特征和旧的反射率相似度来更新反射率相似度。每次聚类像素数量减半，如果把每次聚类后剩下的代表像素看成一个层，那么层的 **scale(规模)** 随着自底向上多次聚类不断缩小，低层的代表像素的特征信息在反射率相似度的加权影响下不断被抽取为高层代表像素的特征和反射率相似关系。整个过程像个金字塔，这也是用词**多尺度度量**的由来，多尺度指的就是金字塔的一层层、每次聚类的结果。

所以三篇论文的内容最终都总结到了**基于多尺度度量和稀疏性的本征图像分解**这个方法中了。下面我按照他逐渐丰富算法的顺序，先整理“基于L0稀疏优化的本征图像分解”算法，再整理“基于多尺度度量和稀疏性的本征图像分解”。

## 基于L0稀疏优化的本征图像分解

### 论文的亮点

所谓的亮点就是该论文和其他文章最不一样的地方，可能是别人没有想到的某些方法，从而产生的这篇论文的写作。本文最大的亮点是其试图构造了一个L0约束项来更理想地实现反射率层的分段常数（也称局部稀疏性先验）先验。L0项确实是最理想的用来实现稀疏性的约束项了，用它来构建针对反射率梯度的惩罚项，能够使得反射率的非零梯度的尽量得少，大梯度尽量多，克制那些不大不小的梯度。



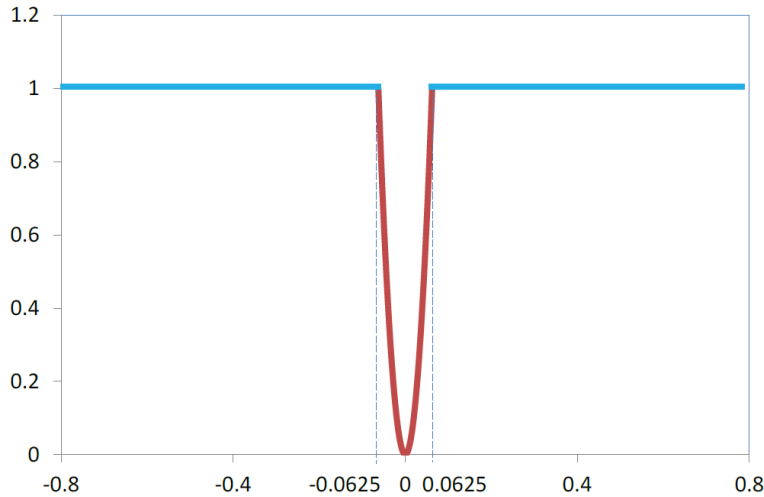
#### 反射率的梯度分类及其成因

**大梯度**：一般是反射率的边界，比如图像上两种截然不同的颜色之间的界限。

**零梯度**：表示反射率没有变化，一段像素的反射率都是某个常数，“属于一个色块”。

**微小的梯度**：既没有大成边界，也没有小到0，在“分段常数”先验下，这些梯度不该存在。在实际图像中，这些梯度更可能是光照、阴影的平滑变化造成的。

能够严格实现分段常数先验的L0项很少被使用，而经常被其他手段近似替代，原因是L0不好构造也不好求解。我看的其他论文中，经常用L1项来实现分段常数先验，L1是对L0效果的近似，但没有L0好。下图就是作者构建的L0约束函数，横坐标是像素反射率的偏导。



下图是L0稀疏约束与不用L0稀疏约束的结果对比。(b)为输入图片，(a)为不用L0稀疏约束分离出的反射率层，(c)为用L0稀疏约束分离出的反射率层，直接可以看出(a)中老虎的左耳朵和后背上光照残留，而(c)分离得更干净。取(a)和(c)图像上同一行，汇出其梯度幅值曲线(d)和(e)，也可以看到(d)中被绿色虚线框框出来地方有微小梯度残留，而(e)没有。可见，L0稀疏约束在分解中作用很大。!

(/images/论文4\_L0项的效果.png?0.2546492072835582) ##### 论文的目标优化函数

本文假设**物体表面平滑并且为朗伯反射体、场景为单一白色光源**，并依据四个约束（**光照平滑约束、Retinex约束、全局反射率稀疏性先验和L0反射率稀疏性先验**），在分析过程中利用概率上的似然估计引入出这几个约束项，最后求解时仍然是**后验概率极大化可以通过对其取对数并取负号转换能量最小化问题，即最小化**

$$E(S, R) = -\log p(S, R|I)$$

，所以最后仍然是要最小化目标泛函，与之前接触过的无异。记 $E(r, u) = E(S, R)$ 为优化的目标函数，则

$$E(r, u) \propto \lambda_1 E_s(r) + \lambda_2 E_{ret}(r) + \lambda_3 E_{L_0}(r) + \lambda_4 E_g(r, u)$$

上面公式的意思是 $E(r, u)$ 正比于**光照能量项、彩色Retinex项、反射率稀疏项和全局反射率稀疏能量项之和**。

其中**光照能量项** $E_s = \sum_i \sum_{j \in N(i)} (\|I_i\|/r_i - \|I_j\|/r_j)^2$ 就是要求光照层梯度积分尽量小， $\|I_i\|/r_i$ 指的是*i*点的像素值与反射率之商，因为有单色光照假设，所以光照可以这么表示。 $N(i)$ 指的是*i*点临域。

彩色Retinex项会人为设定一个色度阈值 $I^{br}$ 和亮度阈值 $I^{chr}$ ，若相邻像素的色度或亮度之差 $I_{ij}^{br}$ 、 $I_{ij}^{chr}$ 任意一个超过阈值则认定他们处在边界两端，此时不增加 $E_{ret}(r)$ ，否则会增加此惩罚项。具体如下：

$$E_{ret}(r) = \sum_i \sum_{j \in N(i)} (\log(r_i) - \log(r_j) - w_{ij})^2$$

其中 $w_{ij} = \begin{cases} \log(\|I_i\|) - \log(\|I_j\|), & \text{if } \|I_{ij}^{br}\| > T^{br} \text{ or } \|I_{ij}^{chr}\| > T^{chr} \\ 0, & \text{otherwise} \end{cases}$

且 $\|I_{ij}^{br}\| = I_i^{br} - I_j^{br}$ ,  $\|I_{ij}^{chr}\| = I_i^{chr} - I_j^{chr}$

$L_0$ 稀疏约束项目的在于严格约束反射率使其满足分段常数先验：

$$E_{L_0}(r) = \sum_i (\phi(\partial_x r_i) + \phi(\partial_y r_i))$$

其中的 $\phi$ 函数是作者自己构建的，[图](#)在前面。

全局反射率稀疏能量项 $E_g(r, u) = \sum_i \|r_i \vec{R}_i - K(u_i)\|^2$ 也和其他大多数提到反射率全局稀疏的论文一样，场景主要由少量的材质颜色控制的。对全局反射率进行聚类，找出C个色块，这些色块记为 $c = 1, 2, \dots, C$ 。 $u_i$ 表示像素i对应的色块编号， $K(u_i)$ 表示像素i对应的色块编号的反射率（三通道，是个三维向量）。因此，上面的约束项 $E_g(r, u)$ 目的就是让每个像素的反射率估计值和它所属的颜色聚类对应的代表颜色相互接近。

## 论文模型优化求解

### • r的初始化

总的能量方程关于r是非凸的，因此用共轭梯度下降法求能量的最小化可能会陷入局部极小值。作者通过给r提供一个合适的初始值来有效避免陷入局部极小。



面对非凸优化问题，合理选择一个好的初始值也是一种办法。

对输入图像进行亮度归一化，并用均值漂移算法（Mean Shift）对图像进行分割，将具有相似性质的图像聚集到同一个区域。共有g个区域。假如两个相邻的像素分属于两个区域p和q，共用一条边。由于他们距离很近，可以认为他们受到的光照相同。设他们的发射率分别为 $m_p$ 、 $m_q$ ，已知像素的范数分别为 $X_{pq}$ 、 $X_{qp}$ 那么他们的光照相等可表示为：

$$X_{pq}/m_p \approx X_{qp}/m_q$$

全图一共g个区域，只有g种反射率需要估计，而这种共用一条边界的点对是很多很多的。用最小二乘的

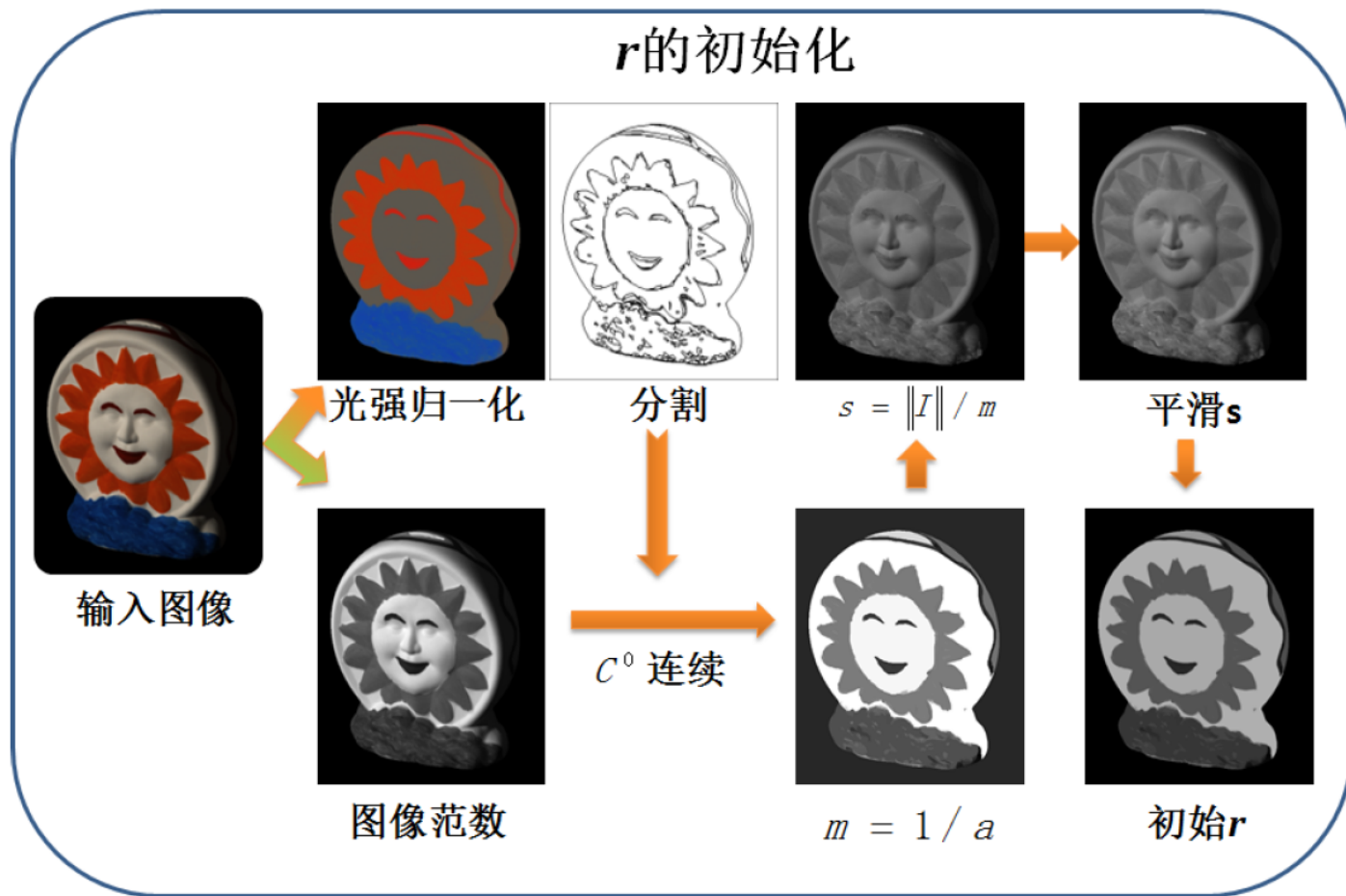
思维将这个方程组改写成对

$$J(\mathbf{a}) = \mathbf{a}^T H \mathbf{a}$$

其中  $\mathbf{a} = (a_1, a_2, \dots, a_g)^T, a_i = 1/m_i$

的最小化，却发现 $H$ 为对称半正定矩阵。原因是，表面上看等式数量远大于未知数的数量，但是很多等式提供的数据是冗余甚至有微小冲突的。因此包括0向量在内的很多组值都是该方程的解。为了求出一个确定的有意义的解，作者令边界最长的那个区域的反射率为1，该设定会在上述 $J(\mathbf{a})$ 的限制下扩散到所有区域，使所有区域都有一个初始的反射率估计值。

这种对 $r$ 初始估计的方法的依据很显然是反射率的全局稀疏性。所以得到的 $r$ 的初始值就很容易过度平滑，边界处也很陡峭。而现实情况是很多图片的真实反射率边界是平滑过渡的。这就导致本应该属于反射率层边界处的一些梯度被挤压到了光照层中，形成了光照层中的噪声。为了避免这些问题的出现，在直接求解得到光照分量后，可对靠近分割边界的光照分量进行平滑操作，得到一个在区域边界处平滑的光照图，然后再利用此光照图产生初始的 $r$ 。下图是上述描述的形象化示例。



### • 对不可微分L0项的处理

总的优化目标中含有两个未知数：

$$E(r, u) \propto \lambda_1 E_s(r) + \lambda_2 E_{ret}(r) + \lambda_3 E_{L_0}(r) + \lambda_4 E_g(r, u)$$

类似所有使用了 $L_0$ 或者 $L_1$ 会遇到的函数不可微分的困境，本文也使用了变量代换的技巧，引入一个误差变量 $\mu$ 将原本 $L_0$ 项中的函数调换到一个 $L_2$ 惩罚项中。

$$E_{L_0}(r, \mu) = \sum_i (\|\mu_x\| + \|\mu_y\| + \frac{1}{\epsilon^2} (\|\partial_x r_i - \mu_x\|_2^2 + \|\partial_y r_i - \mu_y\|_2^2))$$

$r$ 被放到 $L_2$ 项中之后，该项对 $r$ 可微，当其他两个变量 $\mu$ 、 $u$ 固定后，可以用梯度下降法来求解 $r$ 了。这种做法十分像split-bregman方法。

### • 总的迭代步骤：

- 固定 $u$ 交替更新 $r$ 和 $\mu$ 
  - 非凸但是可微:结合 $r$ 的初始化，用梯度下降法更新 $r$
  - $\mu$ 直接用阈值法求解
- 固定 $r$ 求解 $u$ ：  
更新 $u$ 是一个简单的分配问题，即已知色块聚类的结果和所有像素的反射率，把每个像素分配到一个最相近的色块。

## 基于多尺度度量和稀疏性的本征图像分解

该算法保留了上一个算法的绝大部分先验、假设、目标优化项，去掉了上一个算法中的反射率全局先验，并引入了一个自底向上构建的金字塔结构，用于不断抽取下层聚合本身的位置、亮度、色度特性和聚合间反射率相似度构成上层新的位置、亮度、色度特性和聚合间反射率相似度

### 多尺度度量的思想

构建多尺度金字塔是一个较为复杂的过程，我不知道作者是怎么想到这个方法，但是从论文的引言中或许可以看出一点该方法的启发点：

我们观察到图像的反射率并不是一种单纯基于像素的特性，也不是单纯基于区域的特性。首先，给定单个像素的像素值，人很难确定该像素的颜色（反射率值）。人需要将它与其周围的像素进行对比，考虑整张图像的内容后，方可估计出它的颜色。其次，如果按区域进行反射率推理，势必会造成过渡的平滑，导致反射率估计的错误。因此，将图像的反射率看成一种具有多尺度特性的结构化信息，将是一个更好的选择。

其实就是希望用一种方法将像素的特性和区域的特性紧密融合起来，避免单纯使用像素特性时的信息不足和单纯使用区域特征导致反射率层过于平滑。

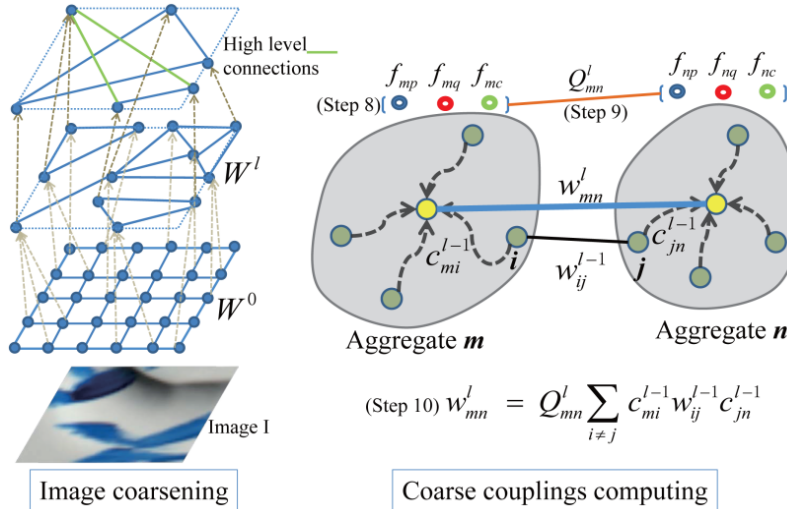
### 多尺度度量的构建过程



在最前面，我用尽量简洁易懂的方式对构建多尺度度量的过程进行了描述。下图是金字塔构建的算法流程图和构建过程更形象的展示图。

**Algorithm 1** Irregular pyramid structure building

- 1: Inputs: image  $I$ , the level number of the pyramid  $T$
- 2:  $l = 0$
- 3: Compute  $W^l$  and  $L^l$  according to Eq.(6)
- 4: Compute pixel feature  $f_i^l, i = 1, 2, \dots, n$
- 5: **while**  $l < T$  **do**
- 6:   Select representative nodes  $B$
- 7:   Compute interpolation matrix  $C^l$
- 8:   Accumulate region properties  $f_s^{l+1}$  using Eq.(9)
- 9:    $Q_{mn}^{l+1} = \exp(-(\alpha_p D_{mn}^p + \alpha_q D_{mn}^q + \alpha_c D_{mn}^c))$
- 10:   Compute the coarse couplings  $W^{l+1}$  using Eq.(8)
- 11:   Compute the Laplacian matrix of  $W^{l+1}$   $L^{l+1}$
- 12:    $l = l + 1$
- 13: **end while**
- 14: //Incorporate high-level information
- 15: For any  $Q_{mn}^T > t_h$ , set  $w_{mn}^T = 1$



- 1~4: 初始化相似度第0层的关联矩阵。  $W^0 = W$ ，其中上标为尺度记号， $W$ 为色度图的关联矩阵，记录了相邻像素的相似度。相似度的定义：

$$w_{ij} = \begin{cases} \exp(-\frac{\|\hat{R}_i - \hat{R}_j\|^2}{0.3}), & \text{if } \|\hat{R}_i - \hat{R}_j\| < t_c (\text{阈值}) \\ 0, & \text{otherwise} \end{cases}$$

其中， $\hat{R}_i$ 和 $\hat{R}_j$ 分别为像素 $i$ 和像素 $j$ 的色度， $t_c$ 为阈值，此处的像素 $i$ 和像素 $j$ 为相邻像素。

- 6: 当前尺度上所有的点组成的点集叫 $V^l$ ，要从中选出一半最有代表性的点，组成点集 $B$ ，使得 $V^l \setminus B$ 与 $B$ 中的点有很强的连接关系。为了选择有代表性的节点，我们首先对节点基于其值进行降序排列。选择第一个节点作为代表性节点。其次，我们根据这个顺序扫描节点，并检查它们与之前被选中点的最大权重。如果这个最大权值小于它与 $V^l$ 中所有节点的权值之和的一个比例(在我们

的实现中为0.15)，我们将该节点添加到代表节点列表中。在这个过程的最后，几乎 $V^l \setminus B$ 中的每个节点都与代表集合 $B$ 中的节点强耦合，然后将剩下的几个节点添加到集合 $B$ 中。新生成的代表点集 $B$ 与被淘汰的 $V^l \setminus B$ 之间的强耦合关系，正是前者足以代表后者进入更高层的原因。

- 7: 计算插值矩阵 $C^l$ 。宏观来看，插值矩阵反映了反射率 $r$ 在不同层次之间的转换关系；微观来看 $c_{ik}^{l-1}$ 暗示了 $l-1$ 层上的聚合（点） $i$ 属于 $l$ 层上聚合（点） $k$ 的程度。

$$r^{l-1} = C^{l-1} r^l$$

$$c_{ik}^{l-1} = \begin{cases} w_{ik} / \sum_j w_{ij}, & \text{if } i < N \\ 1, & \text{if } k = i \\ 0, & \text{otherwise} \end{cases}$$

- 8: 从上一步中的 $C^l$ 可以算得一个 $P^{l-1}$ 。 $f_s^l = P^{l-1} f_s^{l-1}$ 计算新的 $B$ 中每个点的在 $s$ 特征上的取值。
- 9~11: 有了所有点的新特征值之后，计算这些点之间在 $p$ 、 $q$ 、 $c$ 三种特征上的距离的加权 $Q$ 。然后计算出点对之间新的相似度：

$$w_{mn}^i = Q_{mn}^l \sum_{i \neq j} c_{mi}^{l-1} w_{ij}^{l-1} c_{jn}^{l-1}$$

点 $m$ 、 $n$ 之间新的相似度由与他们有联系的点 $i$ 、 $j$ 之间的相似度以及 $m$ 和 $i$ 、 $n$ 和 $j$ 的从属强度共同决定。计算出 $W^{l+1}$ 之后计算它的拉普拉斯矩阵 $L^{l+1}$

- 15: 在前面的循环中，金字塔从底层到顶层被构建。本来最底层的每一个像素的特征（位置、亮度、色度）以及他们之间的相似度是已知的，向上构建的过程中每一层中这些也都被重新计算一遍。到了顶层之后，顶层的点凝聚了大区域的特征信息。此时如果顶层的某两个点之间的特征值很相似，**那么我们就从顶层上认为这两个点的反射率是一样的，改写其相似度为1**，从而实现了远距离像素点之间、区域与区域之间的像素特征对比、联通。注意，金字塔中的每层的反射率 $r$ 都还是未知的，但是

- 层与层的反射率 $r$ 之间有 $C$ 来约束

$$C = \begin{pmatrix} E & -C^1 & \cdots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & & E & -C^l \end{pmatrix}$$

- 层各个点的反射率 $r$ 受 $W$ 约束

$$\tilde{L} = \begin{pmatrix} L^1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & & L^l \end{pmatrix}$$

顶层的 $W$ 被调整之后，这种调整可以自上而下的传达到每一层。 $\tilde{r} = (r_1, r_2, \dots, r_l)'$ 最后总的约束项是： $F_r^M = \tilde{r}' \tilde{L} \tilde{r} + \alpha_M \|C' \tilde{r}\|^2$



## 目标函数的求解

由于大部分继承自前一个算法（去掉了上一个算法中的反射率全局先验），此算法的未知量还要少一个 $u$ 。且加入的多尺度约束项只和 $r$ 有关。所以此算法的求解只需交替迭代求解 $r$ 、 $\mu$ 。

值得一提的是，可以将多尺度度量直接约束到 $r$ 上，来减少变量维度，提高求解效率。回顾在构建图像金字塔结构的过程中，高层的节点总是来自底层的代表性节点。这意味着，高层的节点可对应于原始输入图片中的像素点。因此可以把多尺度的度量直接约束到反射率图 $r$ 上，较小求解方程组的规模，提高求解的效率。分解耗时25s，只用CPU。