2021.10.24周报

纪新龙

本周计划任务

研读张磊老师给的三篇本征分解相关论文并跑通论文中的代码：

- 1_基于多曝光图像的本征分解-3120181006-刘聪
- 2_sfsnet learning shape reflectance and illuminance
- 3_Real-Time Global Illumination Decomposition of Videos

本周任务完成情况

基本完成。

下周计划任务

看黄华老师给的三篇论文

具体完成情况

论文1

通篇阅读了论文1，在相关研究部分对现有的本征分解办法有了初步了解：

- 基于单幅图像的本征分解方法
  - 基于局部梯度先验的本征分解方法
  - 基于全局先验的本征分解方法
  - 基于用户交互的本征分解方法
- 基于多幅图像的本征分解方法
  - 基于同一视角图像的本征分解方法
  - 基于多视角图像的本征分解方法
  - 基于RGB-D 图像的本征分解方法
  - 基于深度学习的本征分解方法

在这种分类框架下，我了解了各种方法的起因和优劣。

论文1的算法

此外，对于文章本身，我也了解到本文提出的方法是在最经典的**Retinex变分方法**上增加了*多曝光的反射率一致性时间约束*、*不受光照干扰的反射率空间约束*。沿用了*光照平滑的先验*。其实本文引入的上述两个约束是显而易见的。在经典Retinex变分方法上将一张图像输入变成三张多曝光图像输入，必定要改变保真项为三张图保真项之和，又由于沿用了光照平滑性约束，所以光照层分别平滑，第二项与经典Retinex变分方法一致。第三项才是与经典方法相比有巨大变化的地方，基于不受光照干扰的反射率空间约束，本文提出的M矩阵强调了反射率的局部稀疏性。但是M矩阵的提出其实直接受启发于引文[37]*Yue H, Yang J, Sun X, et al. Contrast enhancement based on intrinsic image decomposition[J]. IEEE Transactions on Image Processing, 2017, 26(8): 3981–3994.*

综上，本文主要的创新点在于多曝光这一方法的提出，同时融合引文[37]中稀疏权值矩阵M的作用。对于多曝光，其实我觉得多曝光方法在日常生活中实用起来还是比较麻烦的，一般用户不可能安装相机三脚架、调节曝光时间来获取此算法需要的多曝光图像，所以应用于p图软件也很难。但是鉴于本算法效果不错，在特定的需要高效本征分解的场景是可以应用的。

论文1的优化求解方法

最简单的**Retinex方法**最先是基于光照平滑先验通过低通高通滤波器来实现的分离的。后来延伸出经典的**Retinex变分方法**，将所有约束都变成一个泛函E中的各个项，变成数值分析中的最优化问题，求解最小的E。又由于能够保存图像边界的稀疏性假设的介入，泛函E中开始出现L1项，导致其不可微分，无法直接求导等于0取极值，因而有人开始使用**Bregman迭代**求解E。
本文提出的E不仅有L1项，而且L1项还和L2过度耦合，所以使用的是**split-Bregman迭代**方法，目的是解耦合，然后分四步迭代求解，各个击破。split-Bregman迭代是在Bregman迭代之后进一步提出的方法，其根基都是Bregman于1966年论文《THE RELAXATION METHOD OF FINDING THE COMMON》中提出的**Bregman距离**及其特性。

- *Bregman距离*:
  函数$J(u)$在u和v之间的*Bregman距离*定义为

  $$D^p(u,v) = J(u) - J(v) - < p, u - v > \geq 0$$

  符合上述条件的p，称为**次梯度**。Bregman距离是一种广义而抽象的距离定义，可以理解为J(u)于其在v点出的一阶泰勒展开的差距。
- *Bregman迭代*
  - 迭代算法

$$
\begin{aligned}
&Initialize: k=0, u^0=0, p^0=0\\
&While\ u^k\ not\ converge:\\
&u^{k+1}=argmin_u D_J^p(u,u^k)+H(u)\\
&p^{k+1}=p^k-\nabla H(u^{k+1})\in\partial J(u^{k+1})\\
&k=k+1\\
&end\ while
\end{aligned}
$$

- 文献[1]中详细证明了该迭代优化模型有下界，且在迭代中$D_J^p(u,u^k)$和$H(u)$都单调非递减。

- *split-Bregman迭代*[2]
  如前文所述，*split-Bregman迭代*的精髓在于将L1和L2项分离。例如原优化问题的表达式为

$$
arg\min_u |\phi u|+||Ku-f||^2
$$

  我们可以令$d=\phi u$，原式转化为

$$
arg\min_{u,d} |d|_1+H(u)+\frac{\lambda}{2}||d-\phi(u)||_2^2
$$

  此时L1和L2已经分离，单独求d可以用软阈值方法，单独求u则直接是可微分优化问题。再“加回噪声b”，可以得到如下3步迭代算法。

$$
\begin{aligned}
&Initialize: k=0, u^0=0, b^0=0, d^0=0\\
&While\ u^k\ not\ converge:\\
&Step1: u^{k+1}=arg\min_u H(u)+\frac{\lambda}{2}||d-\phi(u)-b^k||_2^2\\
&Step2: d^{k+1}=arg\min_d |d|_1+\frac{\lambda}{2}||d-\phi(u)-b^k||_2^2\\
&Step3: b^{k+1}=b^k+\phi(u^{k+1})-d^{k+1}\\
&end\ while
\end{aligned}
$$

  本文有两个待优化量，分别是反射率和光照，所以用了四次迭代。

论文1中的其他知识点

比如朗伯体、光照平滑性、反射率稀疏性、反射率的分段连续等本征分解常用先验，我也在阅读论文时了解了。CIElab、稀疏矩阵相乘算法、范数规则化、软阈值等知识也做了了解。为了测试论文中提到的代码，我搭建好了相关C++、opencv、matlab的环境。因为opencv C++版本的编译耽误了一些时间，但最后总算搞好了。

论文2

论文2的思想

论文2与论文1的不同点：

- 专注人脸的本征分解
- 基于神经网络方法
- 除了反射率、光照层，还考虑了人脸的normal层。normal英文直接翻译为法线，根据论文的语境，normal层代表人脸的shape信息，总之和人脸的形状、朝向有关，位于最底层。

神经网络方法的好处：

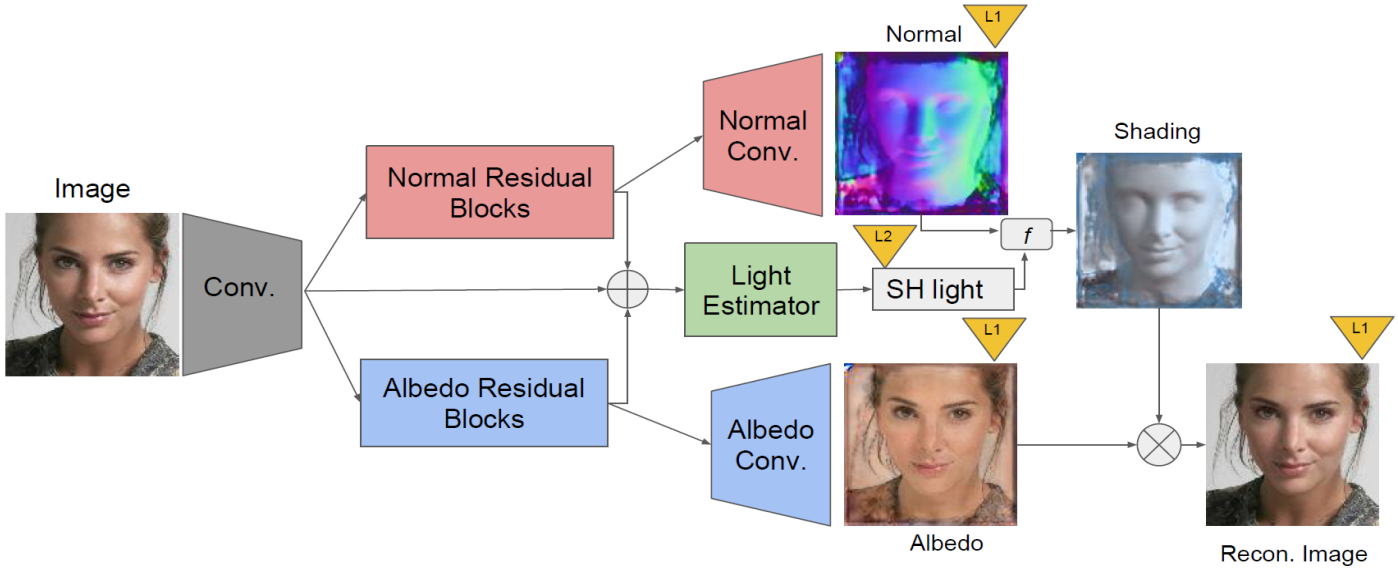
- 与其他方法相比，并不强调太多先验（例如单色光照这样的先验）

神经网络方法的坏处：

- 监督学习，需要大规模的标记好的数据集，这些数据集需要有原图像以及图像本征分解结果的真实值。而目前的数据集分为要么规模不够大，要么是真实原图像但是缺少结果的真实值，要么是有分解结果真实值的人造图像数据集。人造图像跟真实世界收集到的图像相比少那么一点逼真感，缺少一些细节。

本文提出了一种解决此数据集困境的方法，提出用真实数据和人造数据联合训练sfsNet，得到了较好的效果。

论文2的算法



上图就是sfsNet网络，简单的说就是先用Normal Residual Blocks和Albedo Residual Blocks学习出输入图像(image)中的法向层（Normal）和反射率层（Albedo），然后联合输入图像求解出光照层（light）。用光照层和法向层合成阴影层（shading）。最后加上反射率的影响，重建出人脸图像(recon. image)。

先来了解一些知识：

- 人造数据比真实数据缺少一些细节，比如胡子、皱纹。这些细节通常是高频信息。
- 现有两种卷积神经网络：
  - *skip-connection based convolutional encoder-decoder*： 跳跃连接的卷积编码解码结构，这种网络能直接将高频信息从输入传到输出。但是也因此有可能让大部分高频信息直接跳过所有中间层，使得中间的一些层对高频信息的学习不够好。
  - *residual block based network*基于残差块的网络，该网络严格地将高频信息一层一层传下去，保证每一层都充分学习高频信息。

文章提出的训练模式

1. 先用人造数据训练出一个网络。在有标签的合成数据上训练一个简单的基于跳跃连接的编码—解码网络。
2. 将此网络应用于真实数据，以获得法向，反照率和照明估计。 这些元素将在下一阶段用作“伪监督”。
3. 然后将人造数据和真实数据（含前一步得到的伪监督标签）同时喂给上图中的sfsNet网络，通过三层的误差最小进行训练，同时也要求重建图像与真实图像之间的误差最小。**注意，这里的sfsNet使用的是残差快网络，而不是基于跳跃连接的网络。**

简单来说，就是前两步是在为第三步准备带有“伪监督”标签的数据。利用了*跳跃连接*网络不严格学习高频信息、人工合成数据不含有高频信息的特性，在人工数据上训练了*跳跃连接*网络分辨低频信息的能力。然后将此跳跃连接网络来正确分离出真实数据的三层低频信息，并直接保留高频信息到法向层和反射率层中。这样得出的“伪监督”是近似逼真的。
而在真正的sfsNet中又恰恰不使用*跳跃连接*网络，而使用的是残差块。这里看重的是残差块学习高频信息的“认真”。实际上如果在sfsNet中继续使用跳跃连接网络的话（即文末对比试验中的SkipNet+），分离效果跟使用残差块的sfsNet是差不多的。但是SkipNet+得出的光照层没有sfsNet得出的光照层准确。原因可能是残差块本身对待高频信息是*更为严谨的*，尽管使用的伪监督标签不是真正的真实标签，它还是学到了法向层和反射率层之间的*某种子空间*，因此sfsNet根据输入图、法向层和反射率层得到的光照层也更为准确了。

论文2中的实验

- sfsNet的重大优点在于真实数据的伪标签中含有高频信息，因而sfsNet分离出的反射率层也含有高频信息。与‘MoFA’的对比实验中，‘MoFA’分离出的男人面孔的反射率层没有胡子。
- 现有的Pix2Vertex网络用于分解高分辨率图像，能够学习到高频信息。与Pix2Vertex的对比试验中，作者指责Pix2Vertex网络的结果太过于细节以至于“不真实”了，同时指出sfsNet比Pix2Vertex网络快2000倍。
- 在重光照、光照迁移的实验中，sfsNet优于其他网络，效果良好。
- 消融实验
  通过改变sfsNet本身的结构和训练方式，
  - 比如将sfsNet中的残差模块换成跳跃网络（SkipNet+），
  - 比如只用人工数据来训练sfsNet（SfSNet-syn），

- 比如将sfsNet中的残差模块换成跳跃网络且只在人工数据上训练（SkipNet-syn）

将效果与本文提出的训练模式、sfsNet对比，得出本文的模式和网络结构是最优的这一结论。

论文3

论文3的概况

论文3做的工作是从视频流中实时光照分解。论文3的方法也是在很多先验下构造一个泛函，求解这个变分目标函数的最小值。同时，为了达到实时处理视频的速度，根据目标函数的特征精心设计求解是在GPU上求解的。这两点和论文1很像。和前两篇一样，论文3也设计了消融实验证明目标函数中各个项的作用。

论文3的亮点

1. 对间接光照的考虑

分离的光照层不仅有直接光照，还有**间接光照**。间接光照就是图像中的物体之间相互反射的光，一般都带有产生间接光照的物体表面的色彩。所以如果能分离出间接光照，将会使物体重新着色更加逼真，或根据需要消除间接光照的影响。

2. 更全面的图层分解目标函数

为了细致分解出图像的反射率、直接光照、间接光照层，论文3引入了大量的能够解释得通得规律、先验来构建图层分解目标函数：

总约束

$$E_{decomp}(\chi) = E_{data}\ (\chi)\ + E_{reflectance}\ (\chi)\ + E_{illumination}\ (\chi)$$

数据保真项

$$E_{data} = \lambda_{data} \cdot \sum_{x} ||I(x) - R(x) \circ \sum_{k=0}^{K} b_k T_k(x)||_2^2$$

反射层约束

$$E_{reflectance}(\chi) = E_{clustering}\ (\chi)\ + E_{r-sparsity}\ (\chi)\ + E_{r-consistency}\ (\chi)$$

$$E_{clustering} = \lambda_{clustering} \cdot \sum_{x} ||r(x) - r_{cluster}(x)||_2^2$$

$$E_{r-sparsity} = \lambda_{r-sparsity} \cdot \sum_{x} ||\nabla r(x)||_2^p$$

光照层约束

$$E_{illumination}(\chi) = E_{monochrome}\ (\chi)\ + E_{i-sparsity}\ (\chi)\ + E_{smoothness}\ (\chi)\ + E_{non-neg}\ (\chi)$$

$$E_{monochrome} = \lambda_{monochrome} \cdot w_{SR} \sum_{x} \sum_{c} (S_c(x) - |S(x)|)^2$$

$$w_{SR} = 1 - exp(50 \cdot \delta C)$$

$$E_{i-sparsity} = \lambda_{i-sparsity} \cdot \sum_{x} ||T_k(x)_{k=1}^{K}||_1$$

$$E_{smoothness} = \lambda_{smoothness} \cdot \sum_{x} \sum_{k=0}^{K} ||\nabla T_k(x)^2||_1$$

$$E_{non-neg} = \lambda_{non-neg} \cdot \sum_{x} \sum_{k=0}^{K} \max(-T_k(x), 0)$$

这些全是约束项，根据下标都可以看出是对什么层、什么特性的约束。比如最后一个公式是对光照层非负性的约束，就是说直接光照和各种间接光照对成像的影响应该是正向的，会增加被照射区域的能量。可见，这篇论文一大特点就是对反射层、光照层的特点做了细致的分析，用大量的约束来使得分解更准确。

3. 在图层分解的同时优化对色块的估计

论文中的一个先验就是色彩的稀疏性，即图像中只有几种色彩。在该先验下，设定该图中有K种色彩，初始化的时候通过全图的灰度直方图将图像所有像素点的颜色分为K类，每一类的初始是该类别所有像素点的平均色度。这就有了$b_1, b_2 .... b_k$共k种初始色块。**这种基于灰度直方图的聚类方法是优于同类聚类方法的，比其他聚类分割方法更简单。**

此时需要人类介入，看一眼当前图像的聚类情况，进行一个**Misclustering Correction**。比如物体是绿色的，地面是白色的，那么物体上反射的光落在地面上会把地面冉绿，初始化的时候很容易将那一小块绿色地面划分到物体一起。此时就需要人类看一眼，点击这片被错误分类的区域，接下来会有纠偏算法用区域扩散的方法从人类点击的地方逐渐扩散，直到将整个聚类出错的区域都重新估计。**与同类算法相比，这种**

纠偏方式更智能，需要更少的人类互动（同类算法是需要更多的笔画）
k个色块仍会在后续对$E_{decomp}(\chi)$的迭代优化过程中随同其进行优化。**而同类算法会一致沿用第一次的计算出的色块**

$$E_{refine}(\chi) = \lambda_{data} \sum_{x} ||I(x) - R(x) \circ \sum_{k=0}^{K}(b_k + \Delta b_k)T_k(x)||_2^2)$$
$$+\lambda_{IR} \sum_{k=1}^{K} ||\Delta b_k||_2^2 + \lambda_{CR} \sum_{k=1}^{K} ||(C(b_k) + \Delta b_k)||_2^2$$

这是色块估计的目标优化函数。第一项是数据保真约束，第二项约束要求每一步迭代色块的估计值的强度不能变化太大，第三项要求色块的色度不能变化太快。

论文3的求解策略

前一部分中的更全面的图层分解目标函数和在图层分解的同时优化对色块的估计是交替迭代的。
原优化目标函数：
$\chi^* = arg\min_{\chi} E(\chi)$
转化成非线性最小二乘中的高斯牛顿法来求解：
$\delta_k^* = arg\min_{\delta_k} |F(\chi_{k-1}) + \delta_k J(\chi_{k-1})||_2^2$
其中$E(\chi) = ||F(\chi)||_2^2$,$F(\chi)$是保存残差的向量，$\delta_k$是待估计的$\chi$的更新，$\chi_k = \chi_{k-1} + \delta_k^*$,$J$是Jacobian矩阵。
据观察，可以发现图层分解目标函数中包含了大量的约束项。这会导致其J矩阵是一个非常大的稀疏矩阵。
类似于论文1，论文3也采用GPU来并行求解。
对色块估计的优化不同，J矩阵是一个小而稠密的矩阵。用CPU对其进行奇异值分解，可求解。

论文3的一些问题

对于图像颜色类别较少、大片区域同色的视频，该分解算法效果很好。除此以外，还要求视频中的场景移动较慢，没有新的物体加入视频。最后，还需要用户在视频开始时做一些交互。
本文还要求其他一些经典先验：

- 直接光照是单色光
- 反射是漫反射

这些先验和假设简化了问题，同时也限制了算法的应用场景。

1. *1_3_An iterative regularization method for total variation-based image restorationBregman* ↩
2. *1_1_The_Split_Bregman_Method_for_L1-Regularized_Proble* ↩