第5章　本征图像分解与图像重光照

重光照的目的在于根据预先采集的图像生成同一场景在新光照条件下的图像，利用有限的图像生成任意光照条件下的图像不仅能够大大降低数据采集所消耗的时间和精力，还可以生成多样的光照效果，在图像视频处理、深度学习训练数据生成、AR 等领域具有广泛的应用前景。物体的外观是场景几何、场景表面材质和场景光照共同作用的效果，因此从图像中恢复上述场景本征信息，再通过修改分解出的光照分量将是实现场景重光照的一种有效手段。本征图像分解技术旨在将一幅自然图像分解成材质图像和光照图像的乘积。借助深度传感器捕获到的深度信息，本章将首先在 5.1 节和 5.2 节介绍两种基于单幅 RGB-D 图像的本征图像分解方法，其中第一个方法利用与用户交互指定的材质类似但光照不同的像素，在考虑光照空间变化性的条件下实现了对室内场景材质图像和光照图像的分离；第二个方法则在上一方法的基础上进一步解决了场景里存在多个不同颜色光源时本征属性的自动估计问题。然后，针对户外场景，5.3 节介绍了一种在场景几何未知的情况下，利用同一场景在不同光照条件下采集到的多幅图像实现场景重光照的方法。最后，5.4 节针对人脸图像介绍了一种基于深度学习的人脸图像光照归一化算法，以降低光照变化对人脸识别等应用的影响。

5.1　基于用户交互的单幅室内 RGB-D 图像本征图像分解

本节将介绍一种基于用户交互的本征图像分解方法，以实现对于单幅室内 RGB-D 图像的本征图像分解。与已有方法相比，该方法考虑了光照的空间变化性，采用了基于物理的光照先验，并且有效地引入了用户交互，最终获得了更为准确的分解结果。该方法将光照图像进一步分解为远距离光照分量和近距离光照分量，以模拟光照的空间变化性，并基于此设计了一个本征图像模型。然后，基于球面调和函数，该方法提出了一种远距离光源分布计算算法，利用计算的远距离光源和深度图像可合成初步的远距离光照图，在计算最终的远距离光照图时可将该光照图像作为约束。考虑到准确恢复场景近距离光照分量的困难性，该方法让用户交互地指出部分材质相同但光照条件不同的像素，并设计算法，将用户的交互信息传递至没有交互的像素，从而实现对场景本征属性的分解。

5.1.1　本征图像模型

将输入的 RGB 图像记为 I，本征图像分解主要解决如何将图像 I 分解为材质图像 A 和光照图像 S 的乘积，即 $I_p = A_p \cdot S_p$。其中，p 表示图像中的像素，本节算法将不同颜色的通道分开处理。为了解决光照的空间变化性，本节将光照图像 S 分解为两部分：远距离光照图像 D 和近距离光照图像 K。远距离光照图像 D 对应于无穷远处光源所生成的光照图

像(忽略投射阴影),用于描述场景中各面的平均亮度;近距离光照图像 K 则对应于场景内的阴影及由场景内光源所引起的光照空间变化。基于上述描述,像素 p 最终可表示为:$I_p = A_p \cdot K_p \cdot D_p$。不同于传统本征图像分解方法,本节方法分为以下 2 步。

(1) 记 A 与 K 的乘积为 V,通过解能量最小化问题求解 D 和 V。

(2) 基于用户交互,通过求解另一个能量最小化问题获得 A 和 K。

5.1.2　远距离光源分布计算

物体的明暗是物体几何和所受光照共同作用的结果,如果能够获得场景光照信息,本征图像分解就会有更为可靠的约束。本节将会讨论远距离光源分布计算算法。

本节假设场景中所有光源具有相似的色度,对于具有彩色光源的场景,可使用白平衡算法消除光源颜色。本节方法仅讨论光源为白色的情况,而计算光源强度便为本节的重点。

设无穷远光源分布为 L,对于场景中一点 ,其表面亮度计算公式为:

$$I_p = R_p \int_{\Omega(\boldsymbol{n}_p)} L(\boldsymbol{\omega})(\boldsymbol{n}_p \cdot \boldsymbol{\omega})\mathrm{d}\boldsymbol{\omega} \tag{5.1}$$

其中,R_p 为 p 处反射系数,$\Omega(\boldsymbol{n}_p)$ 为上半球面,$\boldsymbol{n}_p$ 为 p 点处法向量,$\boldsymbol{\omega}$ 为单位向量。Ramamoorthi 等人提出式(5.1)可表示为球面调和函数 Y_{lm}($l \geqslant 0$,$-l \leqslant m \leqslant l$)的线性组合,即:

$$I_p = R_p \sum_{l,m} \hat{A}_l L_{lm,p} Y_{l,m}(\boldsymbol{n}_p) \tag{5.2}$$

其中,$\hat{A}_l$ 同($\boldsymbol{n}_p \cdot \boldsymbol{\omega}$)相关,$L_{lm,p}$ 为常数,可通过下面的积分求解:

$$L_{lm,p} = \int_0^{\pi}\int_0^{2\pi} L(\theta,\phi) Y_{lm}(\theta,\phi)\sin\theta \mathrm{d}\theta\mathrm{d}\phi \tag{5.3}$$

式(5.3)离散化后,可使用求和的方式进行逼近:

$$L_{lm} = \sum_{(\theta,\phi)\in\Omega} L(\theta,\phi) Y_{lm}(\theta,\phi)\sin\theta\Delta\theta\Delta\phi \tag{5.4}$$

其中,Ω 为单位球面。将式(5.4)代入式(5.2)中可得:

$$I_p = R_p \sum_{l,m} \hat{A}_l Y_{lm}(\boldsymbol{n}_p) \sum_{(\theta,\phi)\in\Omega} L(\theta,\phi) Y_{lm}(\theta,\phi)\sin\theta\Delta\theta\Delta\phi \tag{5.5}$$

式(5.5)经过简单的整理可写为:

$$I_p = R_p \sum_{(\theta,\phi)\in\Omega} L(\theta,\phi)\sin\theta\Delta\theta\Delta\phi \sum_{l,m} \hat{A}_l Y_{lm}(\boldsymbol{n}_p) Y_{lm}(\theta,\phi) \tag{5.6}$$

记 $P_{\theta\phi} = \sin\theta\Delta\theta \sum_{l,m} \hat{A}_l Y_{lm}(\boldsymbol{n}_p) Y_{lm}(\theta,\phi)$,则有:

$$I_p = R_p \sum_{(\theta,\phi)\in\Omega} P_{\theta\phi} L(\theta,\phi) \tag{5.7}$$

式(5.7)等号两边同除以 R_p 得:

$$\sum_{(\theta,\phi)\in\Omega} P_{\theta\phi} L(\theta,\phi) - \frac{I_p}{R_p} = 0 \tag{5.8}$$

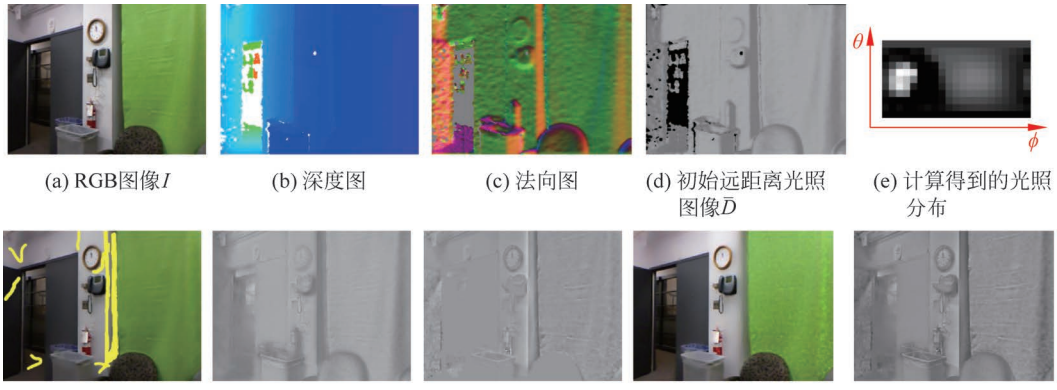
注意到,式(5.8)是一个以 $L(\theta,\phi)$ 和 $1/R_p$ 为未知数的线性方程。如果在场景中选 n 个采样点,便可建立一个线性方程组。然而,求解该线性方程组将会得到值为 0 的解。为了避免

这种情况，令 $1/R_p=1+\rho_p$，其中 ρ_p 是一个非负值(由于 R_p 为小于 1 的材质系数)，可得到如下方程：

$$\sum_{(\theta,\phi)\in\Omega} P_{\theta\phi}L(\theta,\phi)-\rho_p I_p=I_p \tag{5.9}$$

该方程组有 n 个方程，但却有 $n+|\aleph_d|$ 个未知数，其中 $\aleph_d$ 是采样方向的集合，因此该方程组是欠约束的。但是，场景中不同点可能具有相同的材质，这便可极大地减少未知数的个数。为了找到具有相似材质的点，可在图像色度空间进行均值漂移(mean shift)聚类，属于同一类的像素点将具有相同的材质。丢弃像素个数小于 500 的像素集，在剩余的像素中进行均匀采样。

为了进一步提高估计准确度，本节又引入了额外约束。由于光源强度和 ρ_p 都是正值，因此可引入非负约束，即 $L(\theta,\phi)\geqslant 0$ 及 $\rho_p\geqslant 0$。图 5.1(e)展示了计算得到的光照分布。



(a) RGB图像I　(b) 深度图　(c) 法向图　(d) 初始远距离光照图像$\bar{D}$　(e) 计算得到的光照分布

(f) 用户交互选择　(g) 远距离光照图像D　(h) 近距离光照图像K　(i) 材质图A　(j) 光照图像S

图 5.1　分解结果展示

5.1.3　本征图像分解

本节的本征图像分解方法包含 2 步。第一步将输入图像 I 分解为 D 和 V，其中 V 为 A 和 K 之积；第二步则通过用户交互，从 V 中提取 A 和 K。

1. 远距离图像计算

本节将讨论如何计算远距离光照图像 D。通过 5.1.2 节估计得到的远距离光源参数生成初始远距离光照图像 $\bar{D}$。事实上，如果能够使用准确的深度图，$\bar{D}$ 就是要求解的远距离光照图像 D。但是，本节所使用的深度图均具有噪声，这也使得 $\bar{D}$ 存在一定的误差，图 5.1(d)展示了根据 Kinect 捕获的深度图求得的初始远距离光照图像。尽管不能直接将其作为远距离光照图像 D，但 $\bar{D}$ 非常接近 D，所以本节试图结合传统 Retinex 模型和初始远距离光照图像 $\bar{D}$，以得到更为准确的分解结果。

先分析近距离-材质项。用 V 表示 A 与 K 之积，方便起见，称 V 为近距离-材质图像，则得到方程 $I_p=V_pD_p$。按照大多数本征图像分解算法所使用的策略，应将图像转化到对数域上进行分解。对方程两边取对数得到 $i_p=v_p+d_p$。对于 v 和 d 的求解则可通过最小化能量方程 $E(v,d)=E_V(v,d)+E_D(d)$ 实现。其中，E_V 为近距离-材质项，E_D 为远距离光

照项，这两项将在下面详细描述。

根据近距离-材质图像 V 的定义，其包含了场景中光源、阴影及物体材质等因素，而当场景中像素值发生较大变化时，一般均是由阴影、场景中光源产生的高光区域及物体材质本身变化而引起的。因此，当图像中两点的像素值不同时，其在 V 中的值也很有可能不同，特别当这两点的法向相同时，其在 V 中的值一定会不同。据此，将近距离-材质项定义为：

$$E_V(v,d)=\lambda_V\sum_{c\in\{\mathrm{R,G,B}\}}\sum_{p\in\aleph_I}\sum_{q\in\boldsymbol{N}_p}\alpha_{p,q}(v_p^c-v_q^c)^2 \tag{5.10}$$

其中，λ_V 是该项的权重系数，$\aleph_I$ 是图像 I 中所有像素点的集合，$\boldsymbol{N}_p$ 则为以像素 p 为中心的 3×3 邻域。令 $v_p^c=(i_p^c-d_p^c)$，有

$$E_V(v,d)=\lambda_V\sum_{c\in\{\mathrm{R,G,B}\}}\sum_{p\in\aleph_I}\sum_{q\in\boldsymbol{N}_p}\alpha_{p,q}((i_p^c-d_p)-(i_q^c-d_q))^2 \tag{5.11}$$

$\alpha_{p,q}$ 计算公式为：

$$\alpha_{p,q}=\begin{cases}0 & w_{p,q}<0.3\ \text{且}\ \|\boldsymbol{n}_p-\boldsymbol{n}_q\|<0.1\\ w_{p,q} & \text{其他}\end{cases} \tag{5.12}$$

其中，$\boldsymbol{n}_p$ 是 p 处的法向量，$w_{p,q}$ 的定义见式(5.13)：

$$w_{p,q}=\mathrm{e}^{(-100\|\boldsymbol{I}_p-\boldsymbol{I}_q\|)} \tag{5.13}$$

通过权重 $\alpha_{p,q}$ 的定义可知，近距离-材质项可使具有相同 R、G、B 值的像素在 V 中具有相同的值，同时惩罚具有相同法向但 R、G、B 值却有较大差异的像素。

接下来分析远距离光照项。此项包含两个部分：其中一项用以保证远距离光照图像的光滑性，即法向相同的点在远距离光照图像中具有相同的值；另一项则用于保证恢复的远距离光照图 D 接近初始远距离光照图。将这 2 项分别定义为光滑项 E_D^s 与初始约束项 E_D^i：

$$E_D(d)=\lambda_D^sE_D^s(d)+\lambda_D^iE_D^i(d) \tag{5.14}$$

其中，λ_D^s 和 λ_D^i 为权重系数。下面将对这两项进行详细的讨论。

1）光滑项

本节中远距离光照图像所对应的光源均位于无穷远处，并且忽略了场景中的阴影，因此如果两个点具有相同法向，则它们在 D 中的值也相同。光滑项就是为了保证分解得到的远距离光照图像具有这一特性，其定义为：

$$E_D^s(d)=\sum_{p\in\aleph_I}\sum_{q\in\boldsymbol{N}_p}\beta_{p,q}(d_p-d_q)^2 \tag{5.15}$$

式(5.15)中 $\aleph_I$ 和 $\boldsymbol{N}_p$ 的定义与式(5.11)中相同，权重参数 $\beta_{p,q}$ 的定义为 $\beta_{p,q}=\mathrm{e}^{(-100\cdot\|\boldsymbol{n}_p-\boldsymbol{n}_q\|)}$。该权重参数主要用于惩罚具有不同法向的像素对。

2）初始约束项

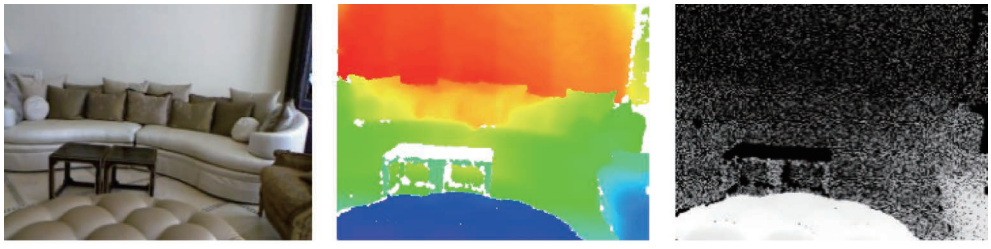
之前已经获得了一幅初始远距离光照图像 $\bar{D}$，尽管该图像受到了噪声干扰，但其与真实的远距离光照图还是非常接近的，因此，该项的主要作用就是保证分解结果同 $\bar{D}$ 相似。另外，为了减小 $\bar{D}$ 中噪声对分解结果的影响，本节仅对部分采样像素进行初始化约束，其影响将通过光滑性传递至其他像素，因此初始约束项的定义为：

$$E_D^i(d)=\sum_{p\in\aleph_{\mathrm{init}}}\gamma_p(d_p-\bar{d}_p)^2 \tag{5.16}$$

其中，$\bar{d}_p$ 为 $\bar{D}_p$ 的对数，权重 γ_p 计算公式为：

$$\gamma_p=\begin{cases}\dfrac{z_p}{t_{\min}}, & z_p<t_{\min}\\ 0.8+0.2\dfrac{t_{\min}-z_p}{t_{\max}-z_{\min}}, & t_{\min}\leqslant z_p<t_{\max}\\ 0.8-0.8\dfrac{2t_{\min}-z_p}{t_{\max}}, & t_{\max}\leqslant z_p<2t_{\max}\\ 0, & z_p\geqslant 2t_{\max}\end{cases}\tag{5.17}$$

其中，$t_{\min}$ 和 $t_{\max}$ 记录了深度传感器的有效工作范围，z_p 为像素 p 处深度。本节采用 Kinect 的有效工作距离 $t_{\min}=1.2\text{m}$、$t_{\max}=4.0\text{m}$。$\aleph_{\text{init}}$ 是采样像素集合，本节根据像素所采集的深度值精确度构建集合 $\aleph_{\text{init}}$。对于深度已知的像素 p，记 $\eta_p=\text{e}^{5(\gamma_p-1)}$，同时按照均匀分布原则生成一个位于区间[0,0.7]的随机变量 σ_p，若 $\eta_p>\sigma_p$ 则将 p 加入 $\aleph_{\text{init}}$，否则丢弃 p，采样结果如图 5.2(c)所示。



(a) 输入图像　(b) 输入深度图　(c) 采样结果

图 5.2　采样结果展示

用于求解远距离光照图像 D 的能量方程的定义为：

$$E(v,d)=\lambda_V\sum_{c\in\{\text{R,G,B}\}}\sum_{p\in\aleph_I}\sum_{q\in\boldsymbol{N}_p}[\alpha_{p,q}((i_p^c-d_p)-(i_q^c-d_q))^2+\lambda_D^s\sum_{p\in\aleph_I}\sum_{q\in\boldsymbol{N}_p}\beta_{p,q}(d_p-d_q)^2+\lambda_D^i\sum_{p\in\aleph_{\text{init}}}\gamma_p(d_p-\bar{d}_p)^2]\tag{5.18}$$

本节中大部分实验将权重参数设置为：$\lambda_V=20$、$\lambda_D^s=30$、$\lambda_D^i=10$。可以看到，若将远距离光照图 D 中所有像素逐行写入一列向量 $\boldsymbol{x}_d$，则式(5.18)可写为以 $\boldsymbol{x}_d$ 为未知数的二次型：

$$E(v,d)=\frac{1}{2}\boldsymbol{x}_d^{\text{T}}\boldsymbol{A}\boldsymbol{x}_d-\boldsymbol{b}^{\text{T}}\boldsymbol{x}_d\tag{5.19}$$

其中，$\boldsymbol{A}$ 为一个 $M\times M$ 的正定稀疏矩阵，$\boldsymbol{b}$ 为一个 $M\times 1$ 列向量，M 表示图像中像素的个数。式(5.19)可通过共轭梯度法最小化，估计得到的远距离光照图如图 5.1(g)所示。

2. 材质图像估计

本节将讨论如何将近距离-材质图像 $\boldsymbol{V}$ 分解为近距离光照图像 K 与材质图像 A 的乘积。由于自动辨别图像像素值变化是由光照条件改变还是材质属性改变引起的具有较大难度，因此，为了保证算法的稳定性，本节选择通过用户交互的方式实现这一目的。让用户在图像 V 上交互地选择材质类似但光照条件不同的像素(如图 5.1(f)所示)，更多具有相似

材质的像素将通过“漫水填充”算法得到，将通过第 i 笔交互得到的具有相似材质的像素记为 $\aleph_S^i$（包括用户交互及自动检测得到的像素）。最终本征属性分解可通过最小化下述能量方程实现：

$$E(a,k)=\lambda_A E_A(a,k)+\lambda_K^s E_K^s(k)+\lambda_K^i E_K^i(k) \tag{5.20}$$

式(5.20)中所有项均是在对数域中进行运算。

材质项 $E_A(a,k)$ 的定义非常类似于式(5.11)，主要区别在于用近距离-材质图像 V 和近距离光照图像 K 分别代替 RGB 图像 I 和远距离光照图 D，权重 $\alpha_{p,q}$ 的定义为：

$$\alpha_{p,q}=\begin{cases}1 & w_{p,q}>0.5 \quad 且 \quad p,q\in\aleph_S^i\\ w_{p,q} & w_{p,q}>0.5 \quad 且 \quad p,q\notin\aleph_S^i\\ 0 & 其他\end{cases} \tag{5.21}$$

其中，$w_{p,q}$ 可通过式(5.13)将 I 变为 V 后进行计算。可以看到，集合 $\aleph_S^i$ 中的像素具有较高的权重，这保证了这些像素可具有相同的材质。

对于近距离光照光滑项 $E_K^s(k)$，除了阴影和高光边缘区域，其余部分也满足光滑性，故该项只需将式(5.15)中的 d 换为 k 即可。而近距离光源在非阴影及高光区域对场景亮度的影响有限，故初始约束项 $E_K^i(k)$ 则需让近距离光照图像 K 接近 1。同时，为了使用户交互对分解结果产生较大影响，设置权重系数 $\lambda_A=70$、$\lambda_K^s=10$、$\lambda_K^i=1$。图 5.1(h)和图 5.1(i)展示了恢复得到的近距离光照图 K 和材质图 A，最终获得的光照图像 S 如图 5.1(j)所示。可以看到，近距离光照图 K 体现了光照的空间变化性。

5.1.4　实验结果

为了验证本节方法的准确性，本节从 MPI-Sintel 数据库中选取了 8 个虚拟场景，然后将本节方法与 Chen 等人提出的方法和 Barron 等人提出的方法均应用于这些场景，接着分别计算 3 种方法分解结果的局部均方误差(LMSE)，其结果如表 5.1 所示。可以看到，即使在不考虑光源颜色的情况下，本节方法仍能比其他两种方法获得更小的均方误差，这证明了本节方法的准确性。图 5.3 展示了在 sleeping 场景下 3 种方法的比较结果。可以看到，本节方法可以得到基本准确的结果，光照图像中包含了场景中的阴影和高光，而材质图像也基本消除了由于光照变化而引起的像素值改变。本节方法假设场景中的光源为白色光源，因此分解得到的本征图像在色调上同真实值有一些偏差，这点可通过引入白平衡算法得以改善。

表 5.1　3 种方法分解结果 LMSE 比较

场　景	Chen 等人提出的方法	Barron 等人提出的方法	本 节 方 法
ally1	0.0615	0.0550	0.0397
ally2	0.0435	0.0506	0.0242
ambush	0.1439	0.1147	0.1023
bamboo	0.0835	0.0869	0.0692
bandage	0.0783	0.0804	0.0727
market	0.0795	0.1254	0.0381
shaman	0.1420	0.1554	0.1019
sleeping	0.0323	0.0438	0.0248

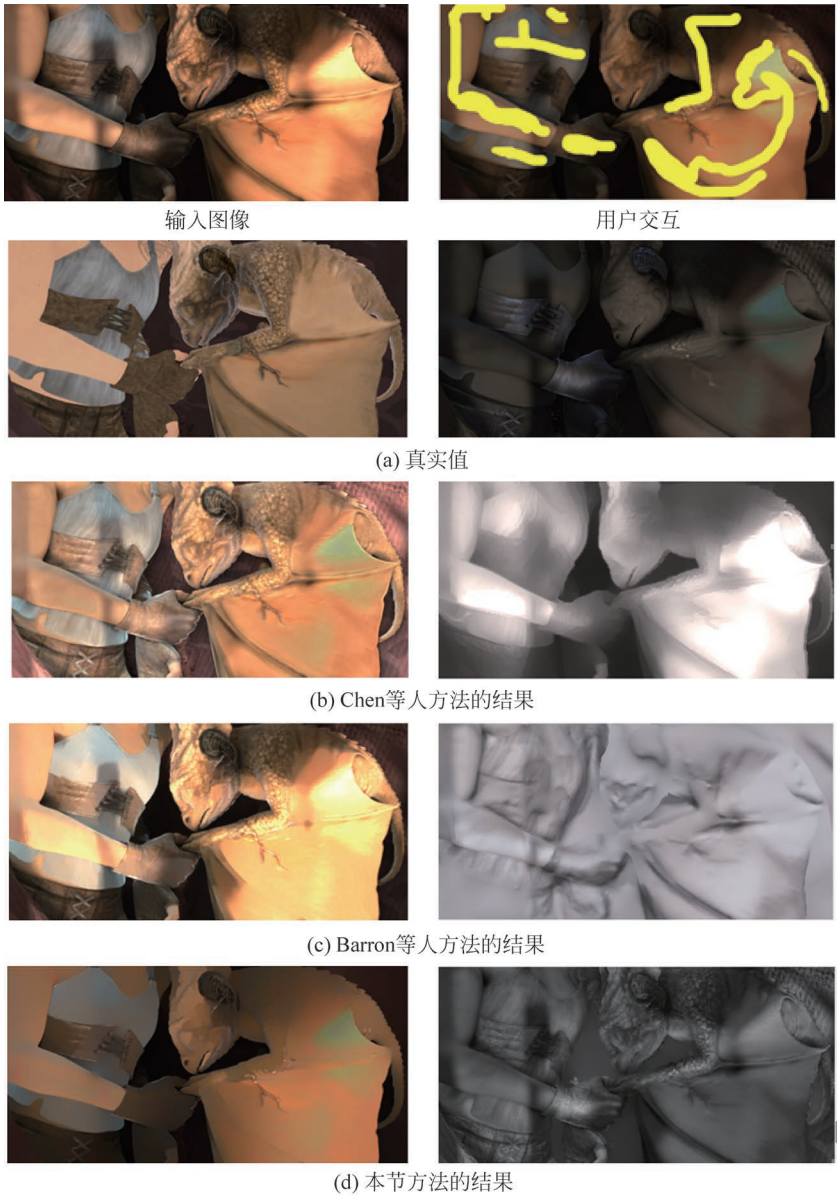


图 5.3　在 sleeping 场景下 3 种方法的比较结果

对于真实场景，本节方法也同 Chen 等人和 Barron 等人的方法进行了比较，比较结果如图 5.4、图 5.5 和图 5.6 所示。可以看到用 Chen 等人的方法得到的光照图像并不是非常准确，图 5.4(b)中的马桶、图 5.5(b)中床上的白色纱巾同周围环境相比均显得过亮，而图 5.6(b)中蓝色的椅子则又过暗。用 Barron 方法得到的结果过多地受到了深度图噪声的影响，因此使得光照图中的一些物体难以识别，如图 5.5(c)中墙上的挂灯有些难以辨识。

本节方法对于材质相似的点可得到非常接近的材质值，光照图像也能正确捕获场景的光照。通过本节方法获得的另外 3 个场景的本征图像分解结果如图 5.7 所示。

本节的本征图像分解方法可应用于图像颜色编辑。改变分解得到的材质图像的颜色，最终颜色编辑的结果为编辑后的材质图像同光照图像的乘积。图 5.8 展示了两个图像颜色

(a) 输入图像　(b) Chen等人方法的结果　(c) Barron等人方法的结果　(d) 本节方法结果

图 5.4　在 restroom(洗手间)场景中本节方法分解结果同其他方法分解结果的比较

(a) 输入图像　(b) Chen等人方法的结果　(c) Barron等人方法的结果　(d) 本节方法结果

图 5.5　在 bed(床)场景中本节方法分解结果同其他方法分解结果的比较

(a) 输入图像　(b) Chen等人方法的结果　(c) Barron等人方法的结果　(d) 本节方法结果

图 5.6　在 kitchen(厨房)场景中本节方法分解结果同其他方法分解结果的比较

编辑的实例。

本节算法测试所用计算机的配置为：Core i5-4570 3.2GHz CPU、16GB 内存。如果忽略用户交互所需时间，本节方法处理完一幅分辨率为 561×427 的图像需要大约 6min 的时间；Chen 等人的方法需要 10 多分钟，而 Barron 等人的方法则需要 1 个多小时。本节方法

(a) 输入图像　(b) 用户交互　(c) 远距离光照图　(d) 近距离光照图　(e) 材质图　(f) 光照图

图 5.7　本节方法获得的本征图像分解结果

(a) 输入图像　(b) 颜色编辑结果

图 5.8　图像颜色编辑结果实例

所需求解的未知数个数大大少于 Chen 等人和 Barron 等人的方法并且仅需求解若干优化问题，因此拥有较高的效率。实际上，用户用于交互指出材质类似但光照不同的像素的时间也仅为 1～2min。

5.2 多光源混合环境下的单幅室内 RGB-D 图像本征图像分解

本节将介绍一种从单幅室内 RGB-D 图像中自动分解场景本征属性的方法，主要解决场景中存在多个不同颜色光源时本征属性的自动估计问题。考虑到光源混合和室内场景光照分布的空间变化性问题，本书提出了一种新的混合光源本征图像模型，将光照图分解为 3 个组成部分，即照明颜色分量、远距离光照分量和局部光照分量。为了实现 3 个光照分

量的求解，本节首先介绍了一种迭代策略，以实现对于场景光源颜色的自动估计；然后利用光照分布的稀疏性和非负性约束进行远距离光照分布的求解，借助恢复的光照分布和深度图共同合成粗糙的远距离光照图像，并将其作为远距离光照分量估计时的附加约束；最后，利用场景反射材质的局部和全局相似性，为材质图像提供可靠的约束，从而实现对各本征属性对应分量的求解。

5.2.1　混合光源本征图像模型

本征图像分解尝试将图像分解成材质图 A 和光照图 S 的乘积。对于 RGB 图像 I：

$$I_p(c)=S_p(c)A_p(c),\quad c\in\{\mathrm{R},\mathrm{G},\mathrm{B}\} \tag{5.22}$$

其中，I_p 表示观察到的像素的 RGB 颜色，S_p 和 A_p 表示每个像素的光照和材质属性。

本节的混合光源模型假设场景中有两种不同的光源，这与大部分室内场景包含的光源种类相符，例如，透过窗户进入房间的室外光照和室内安装的人工光源，或者在使用闪光灯时物体接收到的闪光灯和环境光等。因此有：

$$S_p(c)=S_{\langle 1,p\rangle}L_1(c)+S_{\langle 2,p\rangle}L_2(c) \tag{5.23}$$

其中，L_1 和 L_2 表示光源颜色，$S_{\langle 1,p\rangle}$ 和 $S_{\langle 2,p\rangle}$ 是 p 点在两个灰度光照图像上的值，对应于不同的光源，有：

$$I_p(c)=(S_{\langle 1,p\rangle}L_1(c)+S_{\langle 2,p\rangle}L_2(c))A_p(c) \tag{5.24}$$

大多数本征图像分解方法假设光源远离观察场景，因此，在没有遮挡的情况下，法线方向相同的点将具有相同的像素值。然而，在现实世界中往往存在一些距离场景较近的光源，物体之间也会存在相互的遮挡，实际情况下室内场景不同位置的光照分布可能存在较大的差异，即光照分布具有空间变化性。为了能够处理光照的空间变化性，本节使用远距离光照图 D 来编码仅由远距离光源照明且忽略光源遮挡的场景，因此式(5.25)可以写为：

$$I_p(c)=\left(\frac{S_{\langle 1,p\rangle}}{D_p}L_1(c)+\frac{S_{\langle 2,p\rangle}}{D_p}L_2(c)\right)D_pA_p(c) \tag{5.25}$$

将 $\dfrac{S_{\langle 1,p\rangle}}{D_p}$ 和 $\dfrac{S_{\langle 2,p\rangle}}{D_p}$ 用 $k_{\langle 1,p\rangle}$ 和 $k_{\langle 2,p\rangle}$ 表示，有

$$I_p(c)=(k_{\langle 1,p\rangle}L_1(c)+k_{\langle 2,p\rangle}L_2(c))D_pA_p(c) \tag{5.26}$$

令 $W_p(c)=\dfrac{k_{\langle 1,p\rangle}+k_{\langle 2,p\rangle}}{k_{\langle 1,p\rangle}L_1(c)+k_{\langle 2,p\rangle}L_2(c)}$，在式(5.26)的两边同时乘以 $W_p(c)$，即 $W_p(c)I_p(c)=(k_{\langle 1,p\rangle}+k_{\langle 2,p\rangle})D_pA_p(c)$，所以：

$$I_p(c)=(k_{\langle 1,p\rangle}+k_{\langle 2,p\rangle})D_p\frac{1}{W_p(c)}A_p(c) \tag{5.27}$$

将 $(k_{\langle 1,p\rangle}+k_{\langle 2,p\rangle})$ 和 $1/W_p(c)$ 用 K_p 和 $C_p(c)$ 表示，最终模型为：

$$I_p(c)=D_pK_pC_p(c)A_p(c),\quad c\in\{\mathrm{R},\mathrm{G},\mathrm{B}\} \tag{5.28}$$

根据模型中 4 个项目的定义，C 编码了场景的光照颜色，称为光照颜色图；K 编码了由接近被检查场景的光源和遮挡引起的光照的空间变化，称为局部光照图；D 和 A 分别是远距离光照图和材质图。本节的其余部分将讨论如何计算它们。

5.2.2　远距离光照恢复

对于本征图像分解，如果已知场景的光照条件，那么可以建立一个更可靠的光照约束。

5.1.2 节已经给出了一种远距离光照的计算方法，本节对光照模型进行了一定的简化，提出了一种更为快捷的远距离光照恢复的方法。

在混合光源本征图像模型中，光照的颜色被描述为 3 个相互独立的通道，本节只考虑单一通道，此时计算各通道光源的强度是光照计算算法的主要目标。将远距离光源 L^d 的强度和方向向量分别表示为 $\boldsymbol{L}$ 和 $\boldsymbol{d}$。然后，可以生成与 L^d 相对应的去除阴影的场景图像 I^d，如下所示：

$$I_p^d(c)=\max(0,LA_p(c)\langle \boldsymbol{d},\boldsymbol{n}_p\rangle),\quad c\in\{\mathrm{R},\mathrm{G},\mathrm{B}\} \tag{5.29}$$

其中，$\boldsymbol{n}_p$ 是点 p 的反射率和法向量，$\langle,\rangle$ 表示两个向量的点积。为了方便，称 I^d/L 为基图像，它与光源 L^d 相关。显然，根据光源远离场景的假设，可以将图像表示为与不同远距离光源对应的基图像的线性组合。作为一种近似，可采样入射方向不同的若干光源，因此：

$$I_p(c)=\sum_{q\in\aleph_d}L_qA_p(c)\cdot\max(0,\langle \boldsymbol{d}_q,\boldsymbol{n}_p\rangle) \tag{5.30}$$

其中，$\aleph_d$ 是采样光源的集合。在采样过程中，光源入射方向极角 θ 的采样范围为 $\left[-\frac{\pi}{2},\frac{\pi}{2}\right]$，方位角 ϕ 的采样范围为 $[0,2\pi]$，采样间隔为 $\frac{\pi}{10}$。

在式(5.30)两侧都除以 $A_p(c)$，可得：

$$\sum_{q\in\aleph_d}L_q\cdot\max(0,\langle \boldsymbol{d}_q,\boldsymbol{n}_p\rangle)-\frac{I_p(c)}{A_p(c)}=0 \tag{5.31}$$

注意到，这是一个带有未知数 L_q 和 $1/A_p(c)$ 的线性方程。如果在场景中选择 n 个采样点，就可以构建一个线性方程组。然而，解线性方程组可能得到值为 0 的解。为了避免这种情况，令 $\frac{1}{A_p(c)}=1+\rho_p$，其中 ρ_p 是非负值，因为 A_p 是小于 1 的反射参数。然后有：

$$\sum_{q\in\aleph_d}L_q\cdot\max(0,\langle \boldsymbol{d}_p,\boldsymbol{n}_p\rangle)-I_p(c)\rho_p=I_p(c) \tag{5.32}$$

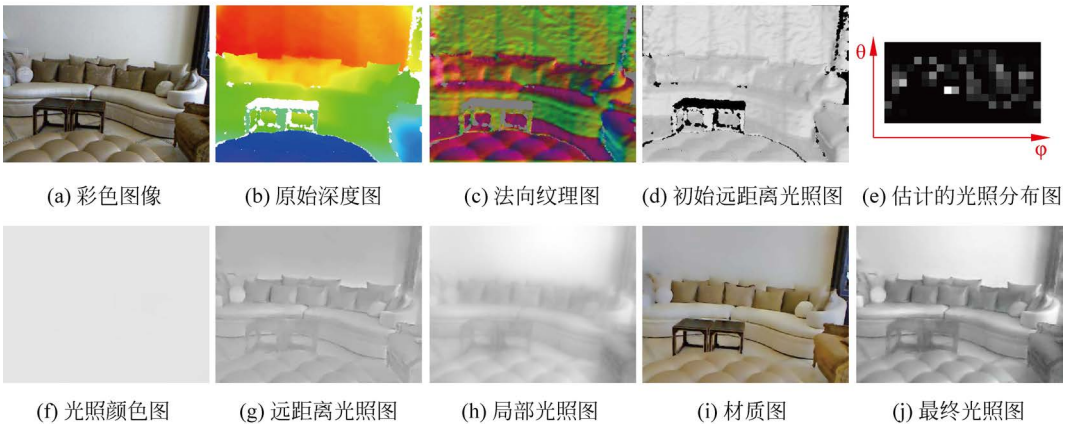
然而，这个线性方程组仍然欠约束，因为有 n 个方程和 $n+|\aleph_d|$ 个未知数。幸运的是，场景中不同的点可能具有相似的材质，这将大幅减少未知数的数量。为了找到具有相似材质的点，可在色度领域中通过 mean-shift 算法对输入图像进行聚类，并假设属于同一类的点具有相似的材质。除此之外，还可采用其他约束来提高近似的准确性，如非负约束，即 $L_p\geqslant 0$(对于 $p\in\aleph_d$)和 $\rho_p>0$；另一个约束则是稀疏约束，即由 Lambertian 场景生成的图像可以通过稀疏光源对应基图像的线性组合表示。

理论上讲，可以选择图像的任何通道来创建方程。在实验中发现使用由 Retinex 模型恢复的光照图像作为输入图像能够获得比其他模型更好的性能。图 5.9(e)展示了估计的光照分布图，可以看出估计的光照参数非常稀疏。

根据估计的光照分布就可生成一个初始的远距离光照图像 $\overline{D}$。对于像素 p 有：

$$\overline{D}_p=\sum_{q\in\aleph_d}L_q\cdot\max(0,\langle \boldsymbol{d}_q,\boldsymbol{n}_p\rangle) \tag{5.33}$$

事实上，如果捕获到的场景深度值完全准确，$\overline{D}$ 即是远距离光照图。遗憾的是，目前商用传感器捕捉到的深度图存在噪声且不完整，这使得 $\overline{D}$ 存在较大误差。图 5.9(d)展示了由原始深度图生成的初始远距离光照图，仍需结合 Retinex 模型对其进行进一步优化。



(a) 彩色图像　(b) 原始深度图　(c) 法向纹理图　(d) 初始远距离光照图　(e) 估计的光照分布图

(f) 光照颜色图　(g) 远距离光照图　(h) 局部光照图　(i) 材质图　(j) 最终光照图

图 5.9　多种图像处理结果

5.2.3　本征成分恢复

本节介绍的本征图像分解方法包含两个步骤：首先估计光照颜色图 C；然后求解远距离光照图 K、局部光照图 D 和材质图 A。

1. 光照颜色图计算

首先讨论如何计算光照颜色图 C。采用前面定义的所有符号，对于像素 p，$C_p(c)=\dfrac{k_{\langle 1,p\rangle}L_1(c)+k_{\langle 2,p\rangle}L_2(c)}{k_{\langle 1,p\rangle}+k_{\langle 2,p\rangle}}$，将 $\dfrac{k_{\langle 1,p\rangle}}{k_{\langle 1,p\rangle}+k_{\langle 2,p\rangle}}$ 称为 α_p，它称为光混合参数，则有：

$$C_p=\alpha_p L_1+(1-\alpha_p)L_2 \tag{5.34}$$

式(5.34)说明，光照颜色图的估计被转化为求解 α_p 和光源颜色 L_1、L_2 的问题。类似的问题在 Hsu 等人的论文中已经讨论过，然而在该论文中，光源颜色由用户指定。本节将寻求一种自动估计的方法。

1) 光源颜色计算

对于光源颜色，先使用一个简单的方法来粗略计算 L_1 和 L_2，再对其进行迭代优化。由于白色光源在实际生活中非常常见，故可将 L_1 的初始值设为 $(1,1,1)^{\mathrm{T}}$。还注意到，镜面反射光相较于漫反射光更容易保持入射光的颜色，因此，本节采用高光区域的像素来计算 L_2，先通过主成分分析(PCA)获取所有高光像素的 R、G、B 值矢量的主成分，再将 L_2 设置为第一个主成分。在上述过程中，高光像素是通过 Zhai 等人的方法检测到的。

在完成初始化后，即可开始对光源颜色进行迭代优化，迭代过程分为两步，两个步骤交替进行，直至收敛。迭代过程中的两步具体如下。

(1) 采用 Hsu 等人提出的方法来估算场景的材质颜色，通过一种投票方案来决定图像中像素的最终材质。然而，该方法计算的材质参数与真实材质值相差一个尺度因子，因此，式(5.24)可以写成：

$$I_p(c)=\left(\frac{S_{\langle 1,p\rangle}}{\tilde{k}_p}L_1(c)+\frac{S_{\langle 2,p\rangle}}{\tilde{k}_p}L_2(c)\right)(\tilde{k}_p A_p(c)) \tag{5.35}$$

其中，$c\in\{\mathrm{R},\mathrm{G},\mathrm{B}\}$，$\tilde{k}_p$ 为计算的材质参数与真实材质值间相差的尺度因子。令 $\tilde{A}_p=\tilde{k}_pA_p$、$k'_{\langle 1,p\rangle}=\dfrac{S_{\langle 1,p\rangle}}{\tilde{k}_p}$ 和 $k'_{\langle 2,p\rangle}=\dfrac{S_{\langle 2,p\rangle}}{\tilde{k}_p}$ 有：

$$I_p(c)=(k'_{\langle 1,p\rangle}L_1(c)+k'_{\langle 2,p\rangle}L_2(c))\tilde{A}_p(c) \tag{5.36}$$

显然，$k'_{\langle 1,p\rangle}$ 和 $k'_{\langle 2,p\rangle}$ 也可以计算出来。此外，这种方法无法获得图像中所有像素的材质颜色、但这不会影响光源颜色的计算。

(2) 从图像中随机采样 n 个像素，对于第 i 个像素，将 $\left(\dfrac{I_{p_i}(\mathrm{R})}{\tilde{A}_{p_i}(\mathrm{R})},\dfrac{I_{p_i}(\mathrm{G})}{\tilde{A}_{p_i}(\mathrm{G})},\dfrac{I_{p_i}(\mathrm{B})}{\tilde{A}_{p_i}(\mathrm{B})}\right)$ 设置为 $n\times 3$ 矩阵 $\tilde{\boldsymbol{I}}$ 的第 i 行。根据式(5.36)，$\tilde{\boldsymbol{I}}=\tilde{\boldsymbol{K}}\cdot\tilde{\boldsymbol{L}}$，其中 $\tilde{\boldsymbol{K}}$ 是一个 $n\times 2$ 矩阵，其第 i 行为 $(k'_{\langle 1,p_i\rangle},k'_{\langle 2,p_i\rangle})$，$\tilde{\boldsymbol{L}}$ 是一个 2×3 矩阵，记录了光源颜色。矩阵 $\tilde{\boldsymbol{K}}$ 和 $\tilde{\boldsymbol{L}}$ 可以通过非负矩阵分解(Non-negative Matrix Factorization，NMF)来求解，步骤(1)的结果将用于 NMF 过程的初始化。

2) 光混合参数的计算

根据 Hsu 的推导，可得到表达式：

$$I'_p=\alpha_p\boldsymbol{A}'_p*\boldsymbol{L}'_1+(1-\alpha_p)\boldsymbol{A}'_p*\boldsymbol{L}'_2 \tag{5.37}$$

其中，α_p 是光混合参数，$I'_p=\left[\dfrac{I_p(\mathrm{R})}{I_p(\mathrm{B})},\dfrac{I_p(\mathrm{G})}{I_p(\mathrm{B})}\right]'$，$A'_p=\left[\dfrac{A_p(\mathrm{R})}{A_p(\mathrm{B})},\dfrac{A_p(\mathrm{G})}{A_p(\mathrm{B})}\right]'$，$L'_i=\left[\dfrac{L_i(\mathrm{R})}{L_i(\mathrm{B})},\dfrac{L_i(\mathrm{G})}{L_i(\mathrm{B})}\right]'$ $(i\in\{1,2\})$，符号 $*$ 表示 Hadamard 乘积。注意到，解 α_p 类似于 matting 方法中的经典前景/背景混合问题，因此，可采用 Levin 等人提出的 matting 方法来解决这个问题。

2. 光照和材质分量估计

在式(5.28)的两侧消除估计的颜色图像。方便起见，仍将 $I_p(c)/C_p(c)$ 表示为 $I_p(c)$，因此：

$$I_p(c)=K_pD_pA_p(c) \tag{5.38}$$

与大多本征图像分解一样，分解过程将在对数域中进行。对公式两侧取对数得到：

$$i_p(c)=k_p+d_p+a_p(c) \tag{5.39}$$

然后，分解问题即可表述为下列能量最小化问题：

$$E(k,d,a)=E_{\mathrm{data}}(k,d,a)+E_A(a)+E_D(d)+E_K(k) \tag{5.40}$$

式(5.40)中，将 E_{data} 称为数据项，E_A 称为材质项，E_D 称为远距离光照项，E_K 称为局部光照项，下面将具体介绍这几项。

1) 数据项

此项的目的是确保可以通过恢复的各本征分量重建原始图像 I，其定义为：

$$E_{\mathrm{data}}(k,d,a)=\sum_{c\in\{\mathrm{R},\mathrm{G},\mathrm{B}\}}\sum_{p\in\aleph_I}(i_p-k_p-d_p-a_p(c))^2 \tag{5.41}$$

其中，$\aleph_I$ 是图像 I 所有像素的集合。

2）材质项

本方法考虑了材质的局部相似性和全局相似性，材质项定义为：

$$E_A(a)=\lambda_A^l E_A^l(a)+\lambda_A^g E_A^g(a) \tag{5.42}$$

其中，λ_A^l 和 λ_A^g 是权重参数，$E_A^l(a)$ 和 $E_A^g(a)$ 分别是局部约束项和全局约束项。

局部约束项包括惩罚 I 中相邻像素之间的差异的成对项：

$$E_A^l(a)=\sum_{c\in\{\mathrm{R,G,B}\}}\sum_{p\in\aleph_I}\sum_{q\in\boldsymbol{N}_p}\alpha_{p,q}^l(a_p(c)-a_q(c))^2 \tag{5.43}$$

其中，$\aleph_I$ 是图像 I 中所有像素的集合，$\boldsymbol{N}_p$ 是像素 p 的 5×5 邻域，$\alpha_{p,q}^l$ 由下式计算：

$$\mathrm{e}^{-1\cdot\kappa_1\|\mathrm{ch}(I_p)-\mathrm{ch}(I_q)\|}\min\left(1,\sqrt{\kappa_2\cdot\mathrm{lum}(I_p)\mathrm{lum}(I_q)}\right) \tag{5.44}$$

其中，$\mathrm{ch}(I_p)$ 和 $\mathrm{lum}(I_p)$ 表示 p 的色度和亮度（在 HSV 色彩空间中计算）。式(5.44)中左侧的指数项表达了具有相似色度值的像素可能具有相似材质的事实，κ_1 可以调整本节方法对像素色度变化的敏感性，大多数情况下将其设置为 200，对于具有复杂纹理的场景，应采用较大的值。右侧项用于惩罚具有低亮度值的像素，因为暗区存在大量噪声，因此 κ_2 可以控制惩罚的强度，在本节所有真实场景的示例中可将其设置为 1，而对于虚拟场景，由于渲染图像中的噪声较小，可将其设置为 100。

全局约束项试图保持不相邻但具有相似材质的像素在估计的材质图像中具有相同的值。为了减少求解的维度，仅为图像中的每个像素选择一个匹配的像素。通过将权重参数 λ_A^g 设置为非常大的值，可以通过局部约束来保持两个非相邻区域的材质的相似性。匹配像素选择的策略为：在颜色图像估计时图像中有部分像素的材质颜色已经恢复，对于这些像素，可选择一个具有相似材质颜色和色度值但在图像平面上距离最远的像素作为匹配像素 p；对于其他像素，仅使用色度值的最小差异和最远距离作为判断。全局约束项 $E_A^g(a)$ 的定义如下：

$$E_A^g(a)=\sum_{c\in\{\mathrm{R,G,B}\}}\sum_{p\in\aleph_I}\alpha_{p,p'}^g(a_p(c)-a_{p'}(c))^2 \tag{5.45}$$

其中，$\alpha_{p,p'}^g=\mathrm{e}^{-20\|I_p-I_{p'}\|}$。这个权重参数用于确保具有相似的 RGB、色度和材质颜色的像素对全局材质相似性的贡献更大。

3）远距离光照项

该项包括两部分，一个是用于平滑生成的远距离光照图，一个是用于约束计算得到的远距离光照图 D 与 $\overline{D}$ 具有相近的值，它们称为平滑项 E_D^s 和初始约束项 E_D^i：

$$E_D(d)=\lambda_D^s E_D^s(d)+\lambda_D^i E_D^i(d) \tag{5.46}$$

其中，λ_D^s 和 λ_D^i 是权重参数。

注意到远距离光照图像 D 忽略了场景中的遮挡和高光，因此，如果两个相邻点具有相似的法线，那么它们将在光照图中具有相似的值，平滑项就是用来保证生成的远距离光照图具有上述特性，其形式如下：

$$E_D^s(d)=\sum_{p\in\aleph_I}\sum_{q\in\boldsymbol{N}_p}\beta_{p,q}^s(d_p-d_q)^2 \tag{5.47}$$

其中，$\aleph_I$ 和 $\boldsymbol{N}_p$ 的定义与式(5.43)中的相似。权重 $\beta_{p,q}^s$ 的定义如下：

$$\beta_{p,q}^s=\mathrm{e}^{(-100\|\boldsymbol{n}_p-\boldsymbol{n}_q\|)} \tag{5.48}$$

上述权重函数用于惩罚具有不同法线的像素。

5.2.2 节介绍了一个生成初始远距离光照图像 $\bar{D}$ 的方法。为了降低 $\bar{D}$ 中噪声的干扰，只需让待求解的光照图像的若干个采样像素接近初始图像 $\bar{D}$，其余像素的值可以根据光照图和材质图的平滑性进行估计。为此，使用初始约束项来添加上述约束，其定义为：

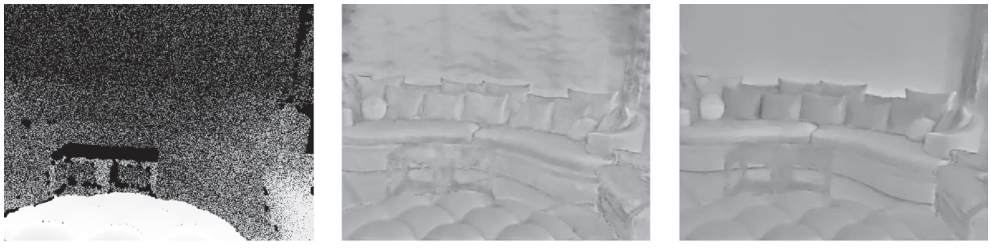
$$E_D^i(d)=\sum_{p\in\aleph_{\text{init}}^d}\beta_p^i(d_p-\bar{d}_p)^2 \tag{5.49}$$

其中，$\bar{d}_p$ 为 $\bar{D}_p$ 的对数，$\aleph_{\text{init}}^d$ 是采样像素。权重参数 β_p^i 的计算公式为：

$$\beta_p^i=\begin{cases}\dfrac{\text{dep}_p}{\text{dis}_{\min}} & \text{dep}_p<\text{dis}_{\min}\\ 0.8+0.2\dfrac{\text{dis}_{\min}-\text{dep}_p}{\text{dis}_{\max}-\text{dep}_{\min}} & \text{dis}_{\min}\leqslant\text{dep}_p<\text{dis}_{\max}\\ 0.8-0.8\dfrac{2\text{dis}_{\min}-\text{dep}_p}{\text{dis}_{\max}} & \text{dis}_{\max}\leqslant\text{dep}_p<2\text{dis}_{\max}\\ 0 & \text{dep}_p\geqslant 2\text{dis}_{\max}\end{cases} \tag{5.50}$$

其中，dep_p 是 p 的深度值，$\text{dis}_{\min}$ 和 $\text{dis}_{\max}$ 共同决定了深度传感器的有效范围。式(5.50)惩罚了超出深度传感器有效工作范围的点，因为它们的深度值非常不准确。本节方法将 $\text{dis}_{\min}$ 和 $\text{dis}_{\max}$ 设置为 1.2m 和 4.0m，对应于 Kinect 传感器的最佳工作范围。

为了获得像素集 $\aleph_d$，本节提出了一种采样策略，首先去除属于深度图像不完整区域的像素；然后使用由式(5.50)计算得到的权重参数 β_p^i 来决定是否应该对像素进行采样。对于像素 p，设 $\eta_p=\text{e}^{-5(1-\beta_p^i)}$，并在[0,0.7]范围内生成一个满足均匀分布的随机变量 σ_p，如果 $\eta_p>\sigma_p$，则将 p 添加到 $\aleph_{\text{init}}^d$ 中。采样结果如图 5.10(a)所示，可以看到深度值较为准确的点具有更高的采样率，而远离摄像机的点具有较低的采样率。图 5.10 展示了使用(图 5.10(b))和不使用(图 5.10(c))上述采样策略计算得到的远距离光照图像，证明了本节提出的采样策略可以有效减少捕获的深度图的噪声。



(a) 采样结果　(b) 不使用采样策略的结果　(c) 使用采样策略的结果

图 5.10　使用和不使用采样策略计算得到的远距离光照图像

4) 局部光照项

与远距离光照项类似，局部光照项也包括两个组成部分：平滑项和初始约束项，因此：

$$E_K(k)=\lambda_K^sE_K^s(k)+\lambda_K^iE_K^i(k) \tag{5.51}$$

其中，平滑项用于保持局部照明的一致性，延用前面符号，$E_K^s(k)$ 的定义为：

$$E_K^s(k)=\sum_{p\in\aleph_I}\sum_{q\in N_p}(k_p-k_q)^2 \tag{5.52}$$

初始约束项则有助于捕捉场景中的遮挡和高光。为此，需要获得一个初始局部光照图像 $\overline{K}$。根据式(5.27)，$k'_{\langle 1,p\rangle}+k'_{\langle 2,p\rangle}=(S_{\langle 1,p\rangle}+S_{\langle 2,p\rangle})/\bar{k}_p$，注意到 $S_{\langle 1,p\rangle}+S_{\langle 2,p\rangle}$ 是 p 在最终光照图中的值，可将初始局部光照图 $\overline{K}$ 近似表示为：

$$K_p=\begin{cases}\dfrac{k'_{\langle 1,p\rangle}+k'_{\langle 2,p\rangle}}{\overline{D}_p} & p\in\aleph^l_{\text{init}}\\ 0 & \text{其他}\end{cases} \tag{5.53}$$

其中，$\overline{D}_p$ 是在 5.2.2 节中计算得到的初始远距离光照图 $\overline{D}$ 在 p 处的值，$\aleph^l_{\text{init}}$ 是可以恢复材质颜色并记录深度值的像素集。初始约束项的定义如下：

$$E^i_K(k)=\sum_{p\in\aleph^i_{\text{init}}}(k_p-\bar{k}_p)^2 \tag{5.54}$$

其中，$\bar{k}_p=\ln(\overline{K}_p)$。由于 $\overline{K}$ 只是真实局部遮挡图像的近似，因此与其他权重参数相比，可将 λ^i_K 设置为一个较小的值。

5.2.4　实验结果

本节使用 MPI-Sintel 数据集对所提方法进行评估，权重参数的设置为：$\lambda^l_A=15$，$\lambda^g_A=200$，$\lambda^s_D=1$，$\lambda^i_D=1$，$\lambda^s_K=0.5$，$\lambda^i_K=0.01$。实验共选择了 8 个不同的场景，使用本节方法获得了材质和光照图像，并与其他方法进行了比较，其他方法包括采用本节方法但不考虑光照颜色的情况(记为 NC)、Chen 等人的方法、Barron 等人的方法和 Bell 等人的方法。为了进行比较，本节对不同方法获得的结果进行了定量评估，这些方法计算得到的最小均方误差(Least Mean Squared Error，LMSE)如表 5.2 所示。从中可以看出，本节方法表现得比其他方法更好，考虑光照颜色可使算法更准确。需要注意的是，bamboo 场景在不考虑光照颜色时会有更小的 LMSE，这主要是由于场景中的竹子使恢复的光照颜色略带绿色，进而增大了误差。

表 5.2　不同方法计算得到的最小均方差

场　景	Bell 等人的方法	Barron 等人的方法	Chen 等人的方法	NC 方法	本 节 方 法
ally1	0.0597	0.0550	0.0615	0.0399	0.0397
ally2	0.0459	0.0506	0.0435	0.0317	0.0242
ambush	0.1470	0.1147	0.1439	0.1664	0.1423
bamboo	0.0480	0.0869	0.0835	0.0669	0.0692
bandage	0.0762	0.0804	0.0783	0.0744	0.0727
market	0.0572	0.1254	0.0795	0.0403	0.0381
shaman	0.1122	0.1554	0.1420	0.1058	0.1019
sleeping	0.0300	0.0438	0.0323	0.0348	0.0248

利用绘制出的虚拟场景，对按照本节方法估计的本征图像与本征图像真值进行对比，对比结果如图 5.11 所示(左边是 market 场景，右边是 bamboo 场景)。可以看到，本节方法产生的本征分量与实际情况非常接近，尽管 bamboo 场景的光照图颜色与真实情况存在差异，但是绿色竹子之间存在的相互反射使得本节方法获得的结果仍然是合理的。

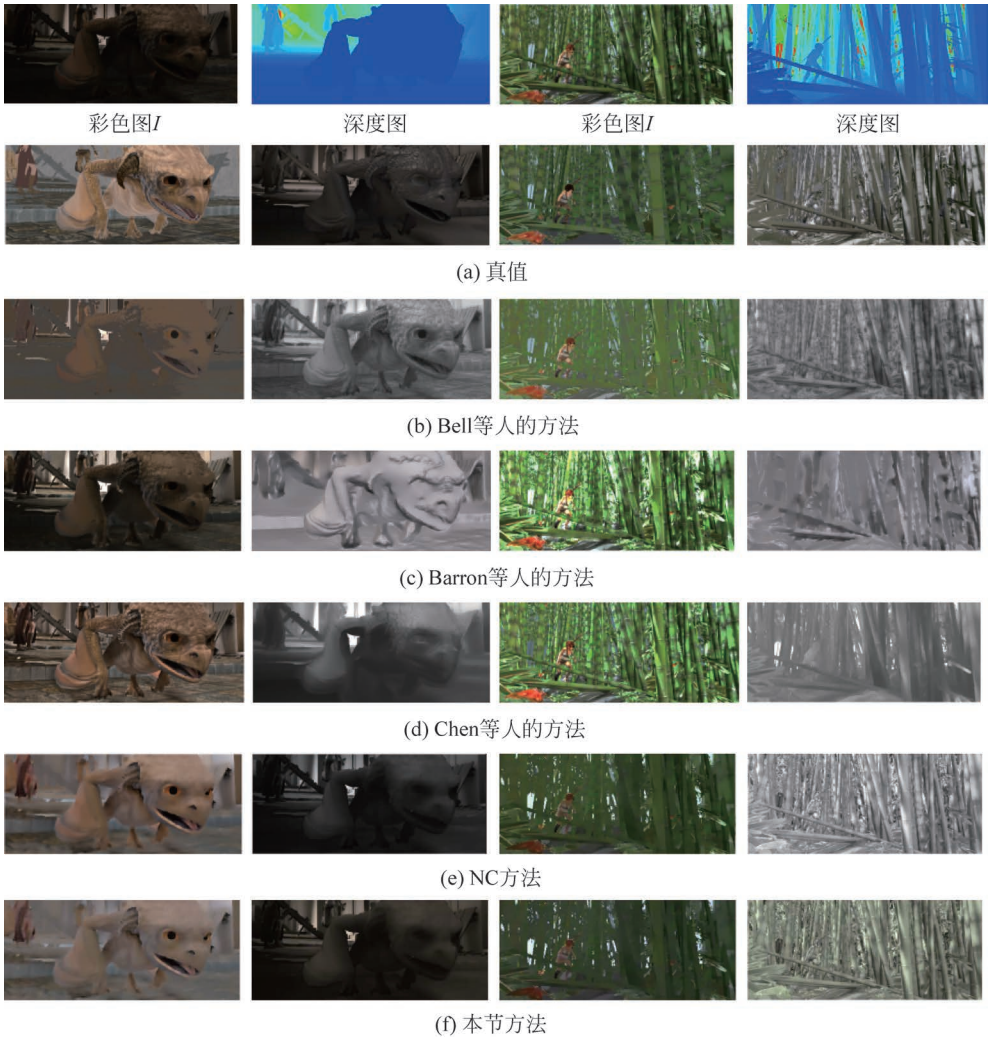
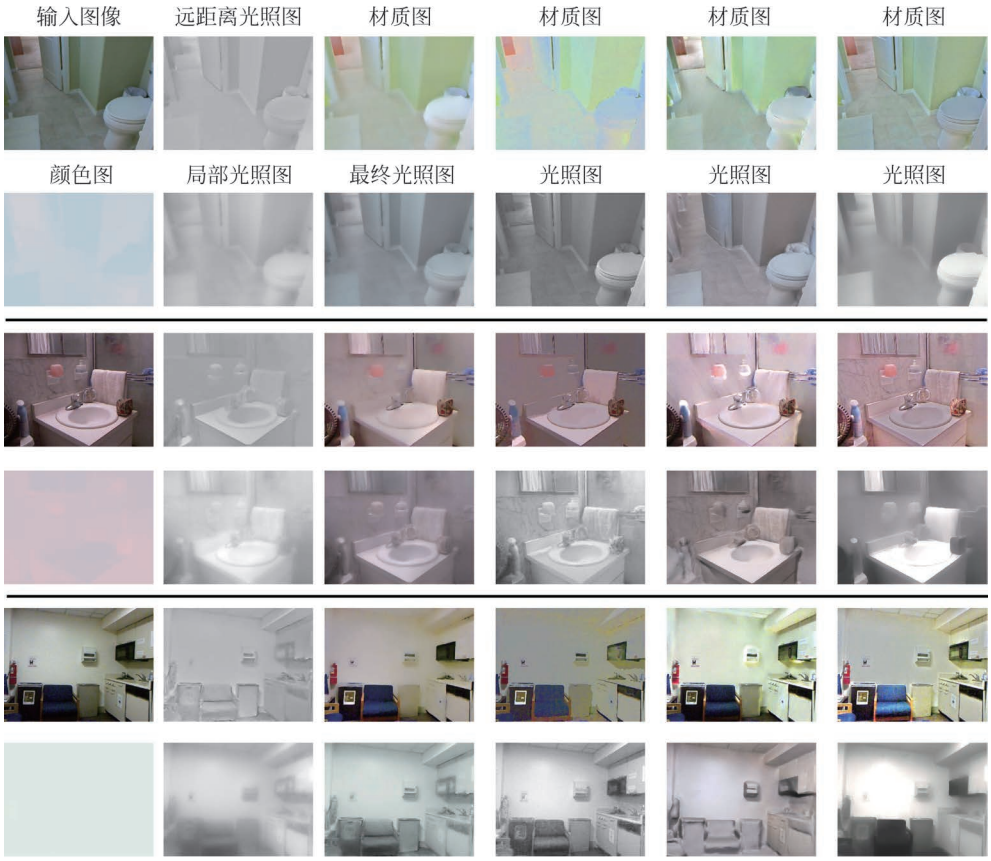


图 5.11　不同方法下的本征图像对比

本节方法还在由 Kinect 捕获的 RGB-D 图像构成的 NYU 数据集上进行了测试。图 5.12 展示了本节方法在 NYU 数据集上与 Bell 等人、Chen 等人和 Barron 等人的方法在获得本征图像分解结果方面的比较。图 5.13 展示了本节方法的本征图像分解结果，以及同 Bell 等人、Chen 等人和 Barron 等人的方法的比较。从图 5.12 和图 5.13 中可以看出，Bell 等人的方法恢复的材质图像太平滑，使光照图像具有较大误差；Chen 等人的方法使具有小 R、G、B 值的物体在恢复的光照图像中太暗，而具有大 R、G、B 值的物体太亮；Barron 等人的方法在分解图像中产生了不准确的区域，例如，图 5.12 中墙上的饰品很难根据光照图像进行识别。本节方法可以为具有相似反照率的点产生几乎均匀的反射率，并且光照图像仍然符合常识。光照颜色图像可以捕捉场景中的多光源混合现象，以图 5.13(a)中第一个场景为例，洗手间外部区域的光照颜色是白色，而洗手间略带蓝色，这与恢复的颜色图像一致。图 5.14 展示了另外 3 个场景的本征图像分解结果，可以看到，图 5.14(b)中的颜色图像可以捕捉室内光源和窗户光的颜色。

(a) 输入图像　(b) Bell　(c) Barron　(d) Chen　(e) 本节方法

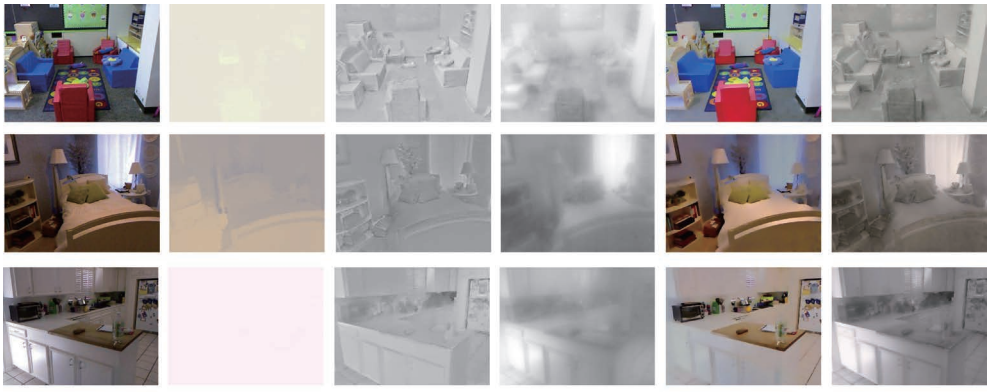
图 5.12　在 NYU 数据集中本节方法与 Bell 等人、Chen 等人和 Barron 等人的方法所获得的本征图像分解结果的比较



(a) 本节方法　(b) Bell　(c) Barron　(d) Chen

图 5.13　在 NYU 测试集中本节方法与 Bell 等人、Barron 等人和 Chen 等人的方法所获得的本征图像分解结果的比较

本节方法在一台配备为 Core i7-4790 4.0GHz CPU 和 16GB RAM 的计算机上实现。本节方法和 Bell 等人的方法的平均耗时接近 4min(对于分辨率为 561×427 的图像)，而 Chen 等人的方法需要 10min，Barron 等人的方法需要 40min。本节方法只需要解决几个优化问题，而且未知参数的数量远远少于 Chen 等人和 Barron 等人的方法，这使得算法更高效。



(a) 输入图　(b) 颜色图　(c) 远距离光照图　(d) 局部光照图　(e) 材质图　(f) 最终光照图

图 5.14　另外 3 个场景的本征图像分解结果

5.3 基于稀疏图像的室外场景重光照

本节介绍了一种基于稀疏图像的室外场景重光照方法，可通过视点固定相机采集到的 4 幅室外场景图像生成该场景在一天内任意时刻任意光照条件下的图像。图 5.15 为重光照方法的流程图，该方法基于基图像理论，根据采样图像计算出相应的太阳光基图像和天空光基图像，并采用迭代优化策略来获得较为准确的结点基图像，任意新的太阳位置及光照条件下的场景图像可利用结点基图像的线性组合来计算。图 5.15 的绿色部分为需要用户交互的步骤。与已有的重光照算法相比，本节方法的主要贡献如下：

(1) 该方法无须在预设光照条件下获取采样图像，也无须重建场景的 3D 模型，因此更加实用。

(2) 该方法能够从 4 幅采样图像中估计出对应于新的太阳位置的基图像而无须其他条件；

(3) 该方法无须对场景进行渲染，非常高效。

在本节下面的内容中，首先将介绍室外场景光照模型，阐述重光照算法的基本思想；然后详细讨论如何从 4 幅满足指定条件的采样图像中计算太阳光基图像和天空光基图像；最后给出基于基图像合成同一场景在新光照条件下的重光照图像的方法。

5.3.1 室外场景光照模型

本节采用与 4.3 节相同的室外场景光照模型，该模型将室外场景的光照环境近似认为是太阳光和天空光共同作用的结果。太阳光为平行光，天空光为呈半球面均匀分布的面光源。3D 空间中的一点 p 的光亮度可以表示为：

$$I(p,\lambda,t)=L_{\text{sun}}(\lambda,t)C_{\text{sun}}(p,\lambda,t)+L_{\text{sky}}(\lambda,t)C_{\text{sky}}(p,\lambda,t) \tag{5.55}$$

其中，C_{sun} 和 C_{sky} 的定义为：

$$\begin{aligned} C_{\text{sun}}(p,\lambda,t)&=\rho(p,\theta_{\text{in}},\theta_{\text{out}},\lambda)S_{\text{sun}}(p,t)\cos\theta(p,t)\omega_{\text{sun}} \\ C_{\text{sky}}(p,\lambda,t)&=\int_{\Omega_{\text{sky}}}\rho(p,\theta_{\text{in}},\theta_{\text{out}},\lambda)S_{\text{sky}}(p,\theta_{\text{in}})\cos\theta(p,\theta_{\text{in}})\mathrm{d}\omega \end{aligned} \tag{5.56}$$

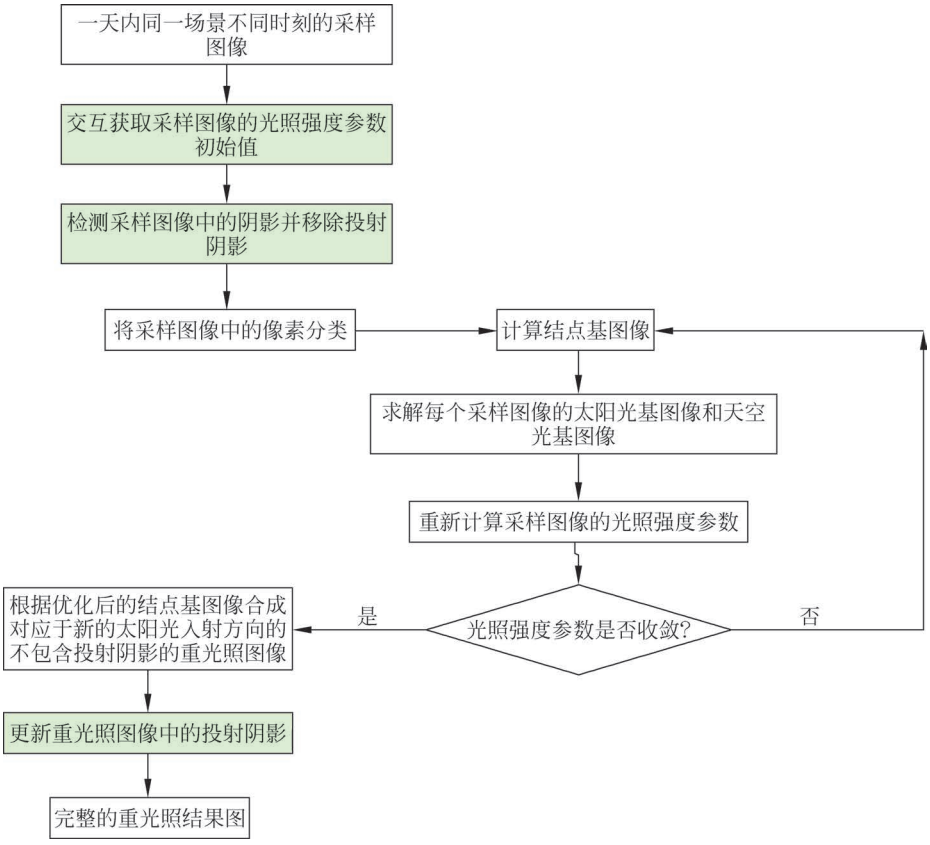


图 5.15　重光照方法的流程图

式(5.55)中的 λ 为入射光波长，ρ 代表点 p 的 BRDF，ω_{sun} 为太阳立体角，并且 $\omega_{\text{sun}}=2\pi(1-\cos(d/2))$，$d=0.53$；$S_{\text{sun}}(p,t)$ 和 $S_{\text{sky}}(p,\theta_{\text{in}})$ 为 p 点处太阳光和天空光的遮挡因子，取值为 0 或 1；L_{sun} 和 L_{sky} 分别为太阳光和天空光的入射强度；Ω_{sky} 是天空光立体角。可以看到，C_{sun} 和 C_{sky} 集合了场景的材质、几何、遮挡信息，这两项分别被称为太阳光基图像和天空光基图像。由于天空区域并不符合室外场景光照模型，因此本节算法忽略天空区域。

5.3.2　室外场景重光照方法

1. 重光照算法基本思路

本算法假设天空光是一个均匀分布的半球面光源，根据式(5.55)可知，同一天中天空光基图像保持不变。对于漫反射表面来说，式(5.55)中的 C_{sun} 可以改写为：

$$C_{\text{sun}}(p,\lambda,t)=k_d(p,\lambda)S_{\text{sun}}(p,t)\cos\theta(p,t)\omega_{\text{sun}} \tag{5.57}$$

其中，$k_d(p,\lambda)$ 为场景表面漫反射系数。为表达方便，暂时不考虑太阳光遮挡因子并记 $\overline{C}_{\text{sun}}(p,\lambda,t)=k_d(p,\lambda)\cos\theta(p,t)\omega_{\text{sun}}$，称 $\overline{C}_{\text{sun}}(p,\lambda,t)$ 为简化的太阳光基图像，即 s-基图像。

记 3D 空间中点 p 的单位法向量 $\boldsymbol{n}$ 为 (x_p,y_p,z_p)，t 时刻太阳光入射方向的单位向量 $\boldsymbol{e}$

为$(x_{\text{sun}}(t),y_{\text{sun}}(t),z_{\text{sun}}(t))$。式(5.57)中的余弦项可以表示：

$$\cos\theta(p,t)=\boldsymbol{n}(p)\cdot\boldsymbol{e}(t) \tag{5.58}$$

3个线性无关的向量都能构成3D空间的一组基。将相机位置作为3D世界的坐标原点，太阳光入射向量$\boldsymbol{e}(t)$便能够由3个线性无关的基向量$\boldsymbol{e}(t_i)$的线性组合来表示，即$\boldsymbol{e}(t)=\sum_i k_i\boldsymbol{e}(t_i)(i=1,2,3)$，其中$k_i$为组合系数，$\boldsymbol{e}(t)$为时刻$t$的太阳光单位入射向量。因此，式(5.58)可改写为：

$$\cos\theta(p,t)=\boldsymbol{n}(p)\cdot\sum_{i=1}^{3}k_i\boldsymbol{e}(t_i) \tag{5.59}$$

由此，可得：

$$\begin{aligned}\overline{C}_{\text{sun}}(p,\lambda,t)&=k_d(p,\lambda)\cos\theta(p,t)\omega_{\text{sun}}\\&=k_d(p,\lambda)\omega_{\text{sun}}\boldsymbol{n}(p)\cdot\sum_{i=1}^{3}k_i\boldsymbol{e}(t_i)\\&=\sum_{i=1}^{3}k_ik_d(p,\lambda)\omega_{\text{sun}}\cos\theta(p,t_i)\\&=\sum_{i=1}^{3}k_i\overline{C}_{\text{sun}}(p,\lambda,t_i)\end{aligned} \tag{5.60}$$

式(5.60)表明了时刻t的s-基图像能够由预先计算的3个线性无关的太阳光入射向量对应的s-基图像的线性组合来表示，本节将这3个预计算的s-基图像记为结点基图像，接下来的工作就是如何从4幅采样图像中获取结点基图像。

2. 结点基图像计算

由前述内容可知，天空光基图像在一天内保持不变。因此，对于一天内不同时刻拍摄的两幅采样图像，由式(5.55)和式(5.57)可得：

$$\begin{aligned}&\frac{L_{\text{sky}}(\lambda,t_2)}{L_{\text{sky}}(\lambda,t_1)}I(p,\lambda,t_1)-I(p,\lambda,t_2)\\=&\frac{L_{\text{sky}}(\lambda,t_2)}{L_{\text{sky}}(\lambda,t_1)}L_{\text{sun}}(\lambda,t_1)S_{\text{sun}}(p,t_1)k_d(p,\lambda)\cos\theta(p,t_1)\omega_{\text{sun}}-\\&L_{\text{sun}}(\lambda,t_2)S_{\text{sun}}(p,t_2)k_d(p,\lambda)\cos\theta(p,t_2)\omega_{\text{sun}}\end{aligned} \tag{5.61}$$

对于场景中某点p，假设它在两幅采样图像中均不处于阴影区域，那么由式(5.61)可得：

$$\begin{aligned}&\frac{aI(p,\lambda,t_1)-I(p,\lambda,t_2)}{aL_{\text{sun}}(\lambda,t_1)}\\=&k_d(p,\lambda)\omega_{\text{sun}}\left[\cos\theta(p,t_1)-\frac{b}{a}\cos\theta(p,t_2)\right]\end{aligned} \tag{5.62}$$

其中，$a=L_{\text{sky}}(\lambda,t_2)/L_{\text{sky}}(\lambda,t_1)$，$b=L_{\text{sun}}(\lambda,t_2)/L_{\text{sun}}(\lambda,t_1)$。结合式(5.58)，可得到如下等式：

$$\frac{aI(p,\lambda,t_1)-I(p,\lambda,t_2)}{aL_{\text{sun}}(\lambda,t_1)}$$

$$=k_d(p,\lambda)\omega_{\text{sun}}\boldsymbol{n}(p)\cdot\left(\boldsymbol{e}(t_1)-\frac{b}{a}\boldsymbol{e}(t_2)\right) \tag{5.63}$$

记：

$$\boldsymbol{e}(t_{\text{inter}})=\boldsymbol{e}(t_1)-\frac{b}{a}\boldsymbol{e}(t_2) \tag{5.64}$$

那么式(5.63)可改写为：

$$\begin{aligned}k_d(p,\lambda)\cos\theta(p,t)\omega_{\text{sun}}&=k_d(p,\lambda)\omega_{\text{sun}}\boldsymbol{n}(p)\cdot\frac{\boldsymbol{e}(t_{\text{inter}})}{l}\\&=\frac{aI(p,\lambda,t_1)-I(p,\lambda,t_2)}{aL_{\text{sun}}(\lambda,t_1)l}\end{aligned} \tag{5.65}$$

其中，l 为 $\boldsymbol{e}(t_{\text{inter}})$的模长。从式(5.65)可以看出，在采样图像的太阳光入射方向和光照强度参数已知的情况下，便能够由两幅采样图像计算出一个新的太阳光 s-基图像，这个图像对应的太阳光入射方向为 $\boldsymbol{e}(t_{\text{inter}})$。

然而，当 p 点在采样图像中处于阴影区域时，上述讨论并不成立。对于 p 只在一幅采样图像中有阴影的情况，即 $S_{\text{sun}}(p,t_1)=0$，$S_{\text{sun}}(p,t_2)=1$，式(5.61)可以表示为：

$$\begin{aligned}&k_d(p,\lambda)\cos\theta(p,t_2)\omega_{\text{sun}}\\&=\frac{I(p,\lambda,t_2)}{L_{\text{sun}}(\lambda,t_2)}-\frac{L_{\text{sky}}(\lambda,t_2)}{L_{\text{sun}}(\lambda,t_2)L_{\text{sky}}(\lambda,t_1)}I(p,\lambda,t_1)\end{aligned} \tag{5.66}$$

式(5.66)即为所需的 s-基图像。

3. 合成目标太阳光 s-基图像

由地理知识可知，在本节的相机坐标系统中，一年中除了某些特殊时刻外，通常情况下一天中不同时刻太阳的坐标向量并不处于同一平面，太阳运动轨迹示意如图 5.16 所示。

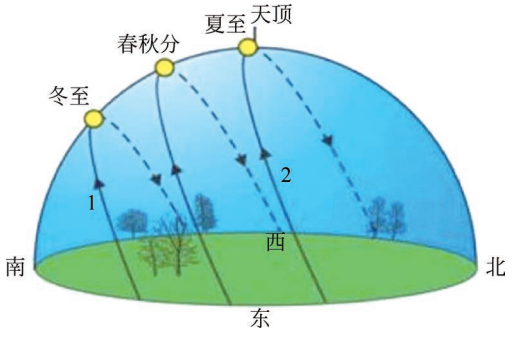


图 5.16　太阳运动轨迹示意

由对应于不同太阳位置的 4 幅采样图像，可以建立一棵最小生成树(Minimum Spanning Tree，MST)，如图 5.17 所示。其中图 5.17(b)和图 5.17(c)为图 5.17(a)所示的无向图的 MST。其中，每个结点 $Nd_i(i=1,2,3,4)$代表一个采样图像，每条边$\boldsymbol{v}_{ij}=\boldsymbol{e}(t_i)-(L_{\text{sun}}(t_j)L_{\text{sky}}(t_i)/(L_{\text{sun}}(t_i)L_{\text{sky}}(t_j)))\boldsymbol{e}(t_j)$代表由式(5.64)计算所得的一个中间时刻的太阳光入射方向向量。

在图 5.17(a)中，$\boldsymbol{v}_{12}$、$\boldsymbol{v}_{13}$、$\boldsymbol{v}_{23}$ 构成了一个回路，通过计算可知$(\boldsymbol{v}_{12}\times\boldsymbol{v}_{13})\cdot\boldsymbol{v}_{23}=0$，因

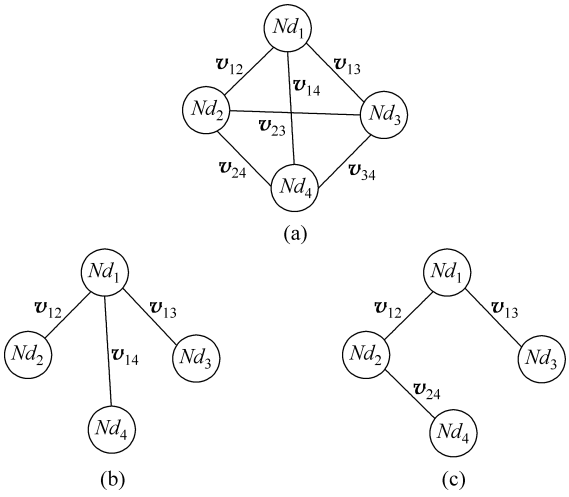


图 5.17　最小生成树

此可证明构成回路的 3 条边是共面的。同样，以图 5.17(b)中的 MST 为例，可计算得$(\boldsymbol{v}_{12}\times\boldsymbol{v}_{13})\cdot\boldsymbol{v}_{23}\neq 0$，即 MST 中的 3 条边所代表的向量是线性无关的，因此根据式(5.65)便可以计算出 3 张线性无关的结点基图像，这样做的前提是 4 幅图像的采样间隔尽可能大，并且采样图像中无阴影区域。

基于同样 4 幅采样图像的不同形式的 MST 对于生成结点基图像来说是等价的。以图 5.17 中的两个 MST 为例，二者的差别仅在于$\boldsymbol{v}_{14}$ 和$\boldsymbol{v}_{12}$。然而，根据图 5.17(a)，$\boldsymbol{v}_{12}$、$\boldsymbol{v}_{14}$、$\boldsymbol{v}_{24}$ 构成了一个回路，可证明这 3 个向量共面。因此，$\boldsymbol{v}_{24}$ 可以由$\boldsymbol{v}_{12}$ 和$\boldsymbol{v}_{14}$ 的线性组合表示，这样就表明了向量组$(\boldsymbol{v}_{12},\boldsymbol{v}_{13},\boldsymbol{v}_{14})$与$(\boldsymbol{v}_{12},\boldsymbol{v}_{13},\boldsymbol{v}_{24})$属于同一个空间，并且各自的结点基图像能够相互转化，故能够选取任意 MST 对应的边向量来计算结点基图像。

前述讨论中均假设遮挡因子非 0 即 1，但是，在实际室外场景中通常有软影的存在。处于软影区域中的像素点的遮挡因子处于 0～1，与本节假设有冲突。为了避免软影区域对基图像估计结果的影响，本节在预处理阶段检测采样图像中的阴影区域并移除其中的软阴影。本节采用 Helor 等人提出的方法来实现阴影检测及移除，该方法要求用户交互地指定一些阴影区域和非阴影区域的像素点，然后基于支持向量机(Support Vector Machine，SVM)和区域增长的方法检测出场景中的阴影区域。在阴影移除阶段，用户交互地画出包围所需移除的阴影区域的多边形，通过求解能量最小化方程来获取阴影移除的结果。

尽管一天中太阳入射方向向量不共面，但是偏移量并不大。因此，本节根据 4 幅采样图像的太阳入射方向向量拟合出一个平面，并将一天中太阳的运动轨迹固定在该平面上。

令变量χ_i 来表示点 p 在第 i 幅采样图像中是否处于自遮挡阴影区域：

$$\chi_i=\begin{cases}0, & p\text{ 处于自遮挡区域}\\ 1, & p\text{ 不处于自遮挡区域}\end{cases}\tag{5.67}$$

将采样图像中的投射阴影移除后剩余的阴影区域即为自遮挡区域。根据χ_i 的值将场景中的像素点分为 5 个集合$\psi_i(i=0,1,2,3,4)$，分类规则如表 5.3 所示。

表 5.3　场景中像素点的分类规则

所属集合	$\sum_{i=0}^{4}\chi_i$	所属集合	$\sum_{i=0}^{4}\chi_i$
ψ_0	4	ψ_3	1
ψ_1	3	ψ_4	0
ψ_2	2		

对于 ψ_0 中的像素点，可以计算出 3 个结点基图像，从而合成对应于新的太阳光入射方向的目标太阳光 s-基图像。

集合 ψ_1 中的像素点 p 仅在一幅采样图像中处于阴影区域，假设该图像为 $I(t_1)$，那么 $<I(t_1),I(t_2)>$、$<I(t_1),I(t_3)>$、$<I(t_1),I(t_4)>$构成 3 个有效图像对，运用式(5.66)计算出 3 个结点基图像。

对于集合 ψ_2 中的像素点 p，通过计算只能得出两个有效的结点基图像，因此，本节将新的太阳光入射向量固定于这两个有效结点基图像对应的中间入射向量所构成的平面上，这样，目标太阳光 s-基图像便表达为两个有效结点基图像的线性组合。

对于集合 ψ_3 中的像素点 p，通过计算只能得出一个有效的结点基图像 $\bar{C}_{\text{sun}}(p,\lambda,t)$，无法通过合成实现重光照。然而，在一天中一定能够找到 p 点的表面法向垂直于太阳光入射方向的时刻，根据采样图像的太阳位置便能够近似估计 p 点的表面法向 $\boldsymbol{n}_p$。根据式(5.58)，目标太阳光 s-基图像 $\bar{C}_{\text{sun}}$ 可表示为 $\dfrac{\boldsymbol{n}_p \cdot \boldsymbol{e}_{\text{new}}}{\boldsymbol{n}_p \cdot \boldsymbol{e}_t}\bar{C}_{\text{sun}}(p,\lambda,t)$，其中 $\boldsymbol{e}_{\text{new}}$ 为目标太阳光入射方向。

最后，由于采样图像中的投射阴影在预处理阶段已被移除，集合 ψ_4 中的像素点在整个采样间隔中一定均位于背阴处，因此这些点的像素值仅由天空光决定，无须考虑太阳光基图像。图 5.18 以不同的颜色展示了场景中不同集合的点的分布，不考虑天空区域。由式(5.65)和式(5.66)可知，结点基图像的精确度依赖于 $L_{\text{sun}}(\lambda,t_i)$ 和 $L_{\text{sky}}(\lambda,t_i)$ 值的准确度，本节采用迭代优化的策略得到较为精确的结点基图像，该过程具体如下。

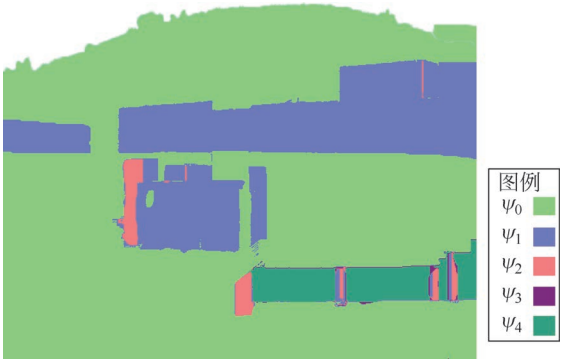


图 5.18　分类结果

(1) 运用 Nakamae 等人的方法获取 $L_{\text{sun}}(\lambda,t_i)$ 和 $L_{\text{sky}}(\lambda,t_i)$ 的初始值。

(2) 已知 $L_{\text{sun}}(\lambda,t_i)$ 和 $L_{\text{sky}}(\lambda,t_i)$，计算结点基图像，然后基于结点基图像合成每个采样图像的太阳光基图像 $C_{\text{sun}}(p,\lambda,t_i)$。

(3) 根据式(5.55)计算采样图像的天空光基图像 $C_{\text{sky}}(p,\lambda)$。

(4) 选取 n 个像素点,求解以下线性方程组:

$$\begin{cases} I(p_1,\lambda,t_i)=L_{\text{sun}}(\lambda,t_i)C_{\text{sun}}(p_1,\lambda,t_i)+L_{\text{sky}}(\lambda,t_i)C_{\text{sky}}(p_1,\lambda) \\ I(p_2,\lambda,t_i)=L_{\text{sun}}(\lambda,t_i)C_{\text{sun}}(p_2,\lambda,t_i)+L_{\text{sky}}(\lambda,t_i)C_{\text{sky}}(p_2,\lambda) \\ \vdots \\ I(p_n,\lambda,t_i)=L_{\text{sun}}(\lambda,t_i)C_{\text{sun}}(p_n,\lambda,t_i)+L_{\text{sky}}(\lambda,t_i)C_{\text{sky}}(p_n,\lambda) \end{cases} \tag{5.68}$$

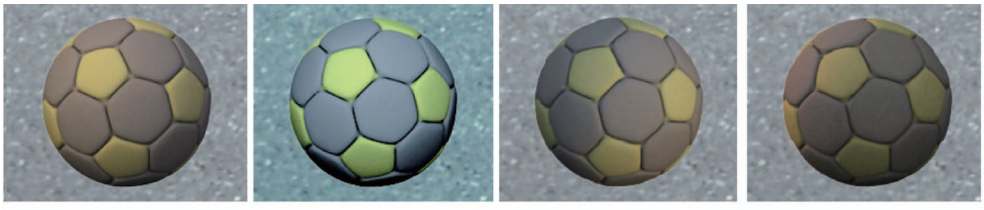
(5) 将 $L_{\text{sun}}(\lambda,t_i)$、$L_{\text{sky}}(\lambda,t_i)$与初值相比,若收敛,则执行步骤(6),否则,返回步骤(2)。

(6) 用 $L_{\text{sun}}(\lambda,t_i)$和 $L_{\text{sky}}(\lambda,t_i)$计算结点基图像,并基于结点基图像合成目标太阳光基图像,结束。

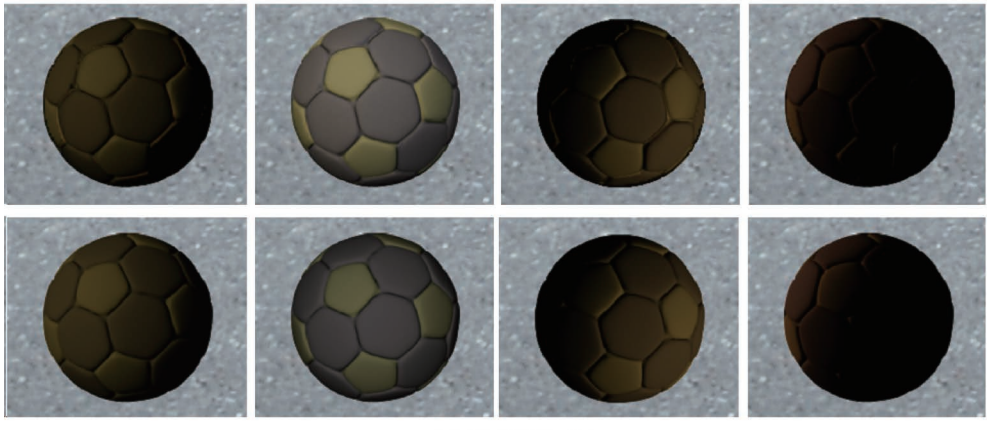
5.3.3 实验结果

本节用一个虚拟场景和两个真实场景来验证算法的有效性。对于每个真实场景,均有固定视点在一天中不同时刻拍摄的 4 幅采样图像作为算法的输入。本节所有方法均在一台配置为 Core2 Duo E7400 2.80GHz GPU 和 3GB RAM 的计算机上进行测试。

方法在包含一个足球的虚拟场景中进行测试,该场景的光源配置与室外场景相似,用平行光代替太阳光,用半球面光源代表天空光。选取用 4 幅不同光源位置和光照强度渲染出来的场景图像作为输入,将计算出的采样图像的太阳光基图像和天空光基图像与基准值进行对比,结果如图 5.19 所示。其中,图 5.19(a) 为采样图像。图 5.19(b)和图 5.19(c)分别为太阳光基图像和天空光基图像的对比图,第一行为计算所得结果,第二行为基准值。太阳光基图像和天空光基图像通过绘制得出。对比图证明本节方法的结果非常接近于真实绘制结果。

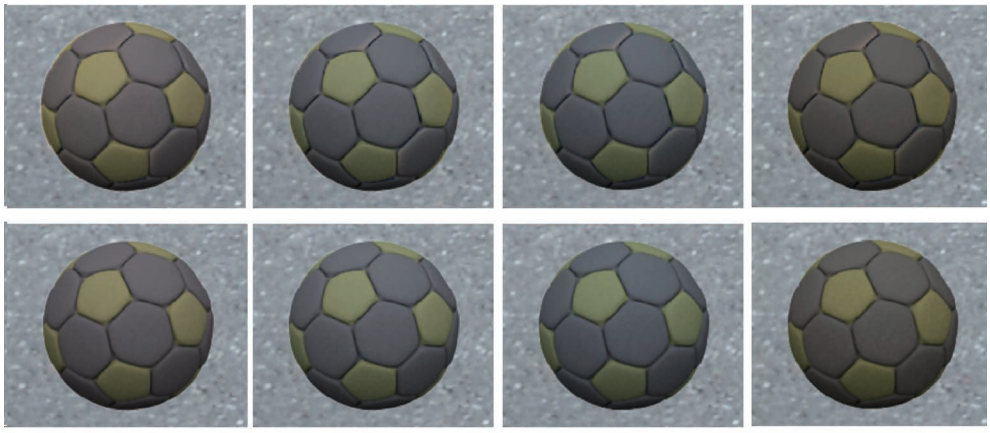


(a) 采样图像



(b) 太阳光基图像对比

图 5.19 对比结果

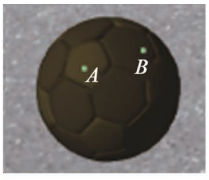


(c) 天空光基图像对比

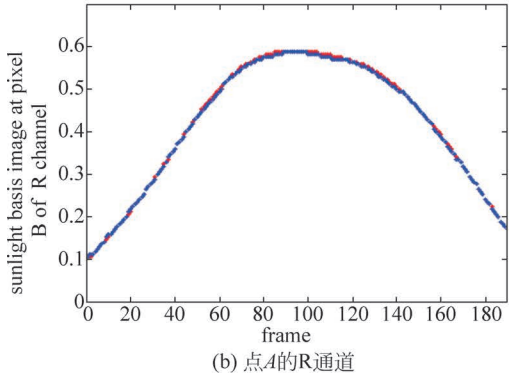
图 5.19　（续）

另外，本节还生成了一段太阳光基图像随光源位置及强度变化而变化的视频，图 5.20 展示了合成的太阳光基图像中两个采样点的像素值随时间变化的曲线与基准曲线的对比。其中，红色曲线为估算值，蓝色曲线是基准值，图 5.20(b)～图 5.20(d)分别代表像素点 A 的 R、G、B 通道，图 5.20(e)～图 5.20(g)分别代表像素点 B 的 R、G、B 通道。

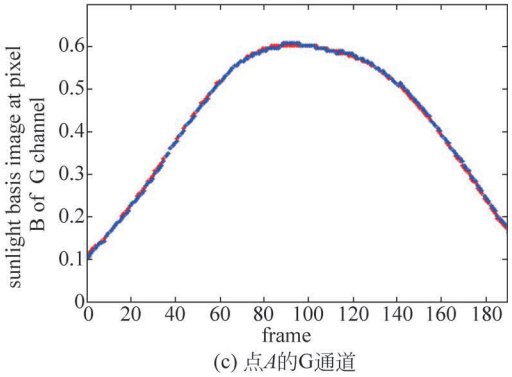
图 5.21 展示了对真实场景进行重光照的结果，图 5.21(a)为真实采样图像，图 5.21(b)为在预处理阶段移除采样图像中投射阴影后的结果。图 5.21(c)为计算得出的每个采样图像所对应的太阳光基图像，图 5.21(d)最左边的图为天空光基图像，其他为新的太阳位置和光照强度下的重光照结果，其中太阳光强度 $L_{\text{sun}}(t_i)$ 的 R、G、B 的比例分别为(1.0，0.788，0.5379)、(1.0，0.9256，0.759)、(1.0，0.6716，0.3358)。另外，用相同光照强度参数 $L_{\text{sun}}(\lambda, t_i)$ 和 $L_{\text{sky}}(\lambda, t_i)$ 合成的重光照结果图可以用来验证颜色恒常性，如图 5.21(e)所示。



(a) 合成图像

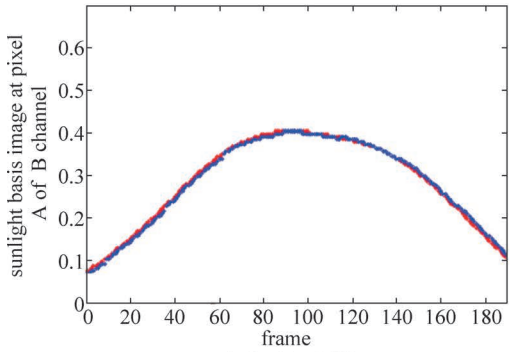


(b) 点A的R通道

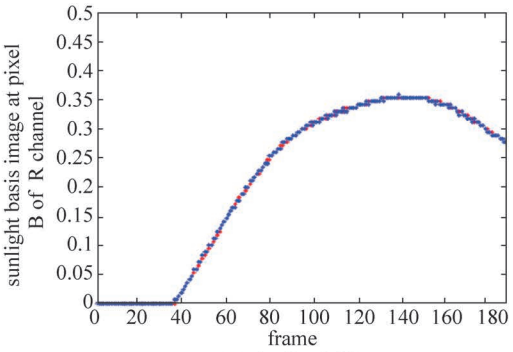


(c) 点A的G通道

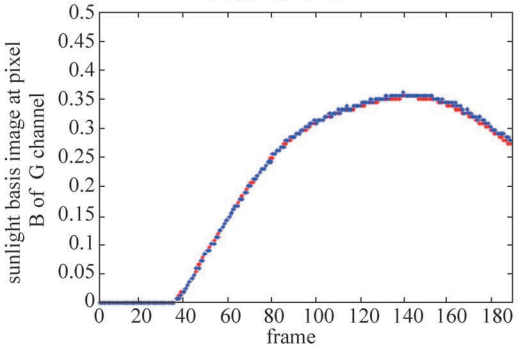
图 5.20　合成的太阳光基图像中两采样点的像素值随时间变化的曲线与基准曲线的对比



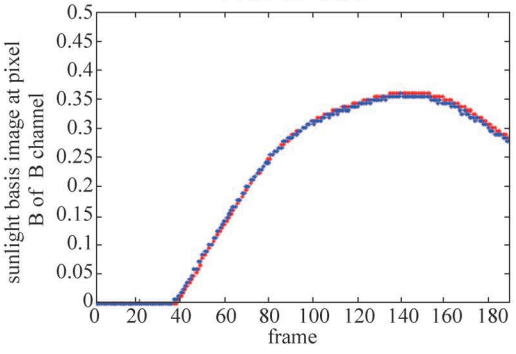
(d) 点A的B通道



(e) 点B的R通道



(f) 点B的G通道



(g) 点B的B通道

图 5.20 （续）

(a) 真实采样图像

(b) 移除投射阴影结果

(c) 太阳光基图像

图 5.21 对真实场景重光照的结果

(d) 天空光基图像(左一)及重光照结果

(e) 用重光照结果图验证颜色恒常性

图 5.21　（续）

图 5.22 展示了真实场景的重光照结果与基准图像的对比及误差。图 5.22(a)为基准图像，即去除投射阴影后的采样图像；图 5.22(b)为重光照结果，其光照参数与基准图像保持一致，光照强度参数 L_{sun} 和 L_{sky} 中 R、G、B 分量的比值分别为(1.0,0.7563,0.6181)和(1.0,1.059,1.069)，图 5.22(c)～图 5.22(e)分别展示了重光照结果中 R、G、B 分量的绝对误差。在本例中，基准图像由相机拍摄而来，相应的太阳位置可以由拍摄时间、地点得出，光照强度参数 $L_{sun}(\lambda)$ 和 $L_{sky}(\lambda)$ 可由第 4 章中的方法估计得出。这样，根据本节提出的算法便可合成具有相同太阳位置和光照强度的重光照结果。该误差对比图同样证明了本节算法的有效性及准确性。

(a) 基准图像　　(b) 重光照结果

(c) R通道误差　　(d) G通道误差　　(e) B通道误差

图 5.22　真实场景的重光照结果与基准值的对比及误差

5.4 基于非对称生成对抗网络的人脸图像光照归一化

人脸图像光照归一化试图将非均匀光照条件下的人脸图像转换为均匀光照条件下的人脸图像，是提高非均匀光照环境下人脸识别正确率的有效手段。然而，人脸结构特征对图像的扭曲极为敏感，使用传统的图像转换或者翻译方法不可避免地会造成人脸面部结构的扭曲，在实际应用中即使轻微的扭曲都会对人脸识别算法产生干扰。为了解决这一问题，本节介绍了一种新的卷积神经网络——非对称联合生成对抗网络（Asymmetric Joint Generative Adversarial Network，AJGAN）。该网络可以在不需要人脸先验知识的情况下，对任意光照环境下采集的人脸图像进行归一化。

AJGAN 具有两组生成对抗网络结构——生成器 G_1 和生成器 G_2。其中，G_1 将不同光照类型的人脸图像进行光照归一化，G_2 则将归一化的人脸图像映射回不同的光照条件。G_1 和 G_2 是一组非对称性的网络结构，G_2 的引入保证了人脸身份属性在光照归一化的过程中恒定。为了解决 G_2 在重光照过程中无法控制光照条件，可在 AJGAN 中引入一个光源标签，以应对不同的光照条件。在标签的引导下，G_2 能够在众多的光照条件下，找到映射至特定光照下人脸图像的准确路径。此外，不同的光照标签组合，能够动态地扩展原始训练库中不具备的光照条件，在一定程度上减小了深度学习对数据的依赖。AJGAN 在 3 个公开数据库中进行定性和定量的测试。实验表明，该方法明显优于其他光照归一化方法。

5.4.1 算法思路

给定一个光照不均匀的人脸图像 I 和对应的均匀光照图像 J 作为光照归一化的图像对（I，J）。根据本征图像理论，图像是由材质图像 R 和光照图像 L 的像素乘积得到的。因此，可以将 I 和 J 表示为：

$$\begin{aligned} I(x,y) &= R_I(x,y)L_I(x,y) \\ J(x,y) &= R_J(x,y)L_J(x,y) \end{aligned} \tag{5.69}$$

其中，$I(x,y)$、$J(x,y)$分别是 I、J 在(x,y)位置处像素的 R、G、B 值，$R_I(x,y)$和 $R_J(x,y)$是 I、J 的反射率，$L_I(x,y)$和 $L_J(x,y)$是 I 和 J 的光照分量。对于固定的人脸姿势，反射率 R 是固定的。换句话说，同一个人的反射率在不同光照条件下是相似的，即 $R_I \approx R_J$。因此，式(5.69)可以简单地写成：

$$\| \log I(x,y) - \log J(x,y) \|_1 = \| \log L_I(x,y) - \log L_J(x,y) \|_1 + N(0,\sigma^2) \tag{5.70}$$

其中，高斯噪声解释了在姿态对齐过程中产生的 R_I 和 R_J 之间的微小差异。因此，从非均匀光照图像 I 到对应的均匀光照图像 J 的映射就是光照类型的相互变换。

此外，从映射的角度来看，在固定视点捕获的静态场景中，不同光照条件的图像与高维非线性函数有关。在基于图像的重光照中将光照的传输建模为像素坐标和光源位置的函数，并用已知的不同光照条件下拍摄的图像来训练神经网络，从而逼近该非线性函数。受此启发，不同光照条件下的同一个人之间的图像转换映射为非线性映射，这可以由生成器

G 来近似表达：

$$I \xrightarrow{G} J \approx L_I(x) \xrightarrow{G} L_J(x) \tag{5.71}$$

5.4.2　网络结构

AJGAN 的主要目标是将任意光照条件下的人脸图像转换为均匀光照下的图像。同时，在转换过程中，必须保证人脸的身份不变，因为这对于人脸识别是非常重要的。为了实现这一目标，AJGAN 具有两个闭环结构：光照归一化 G_1 和重光照 G_2。图 5.23 详细显示了 AJGAN 结构。橙色流程图显示了光照归一化 GANG_1，蓝色流程图显示重光照 GAN G_2，光照标签作为输入。在每一个网络流中，输入图像首先根据相应的真实图像和其光照标签进行变换，然后在重建损失的引导下重建图像。D_1 的功能是判断图像是真是假，D_2 不仅判断图像是真是假，还对光照标签进行分类。

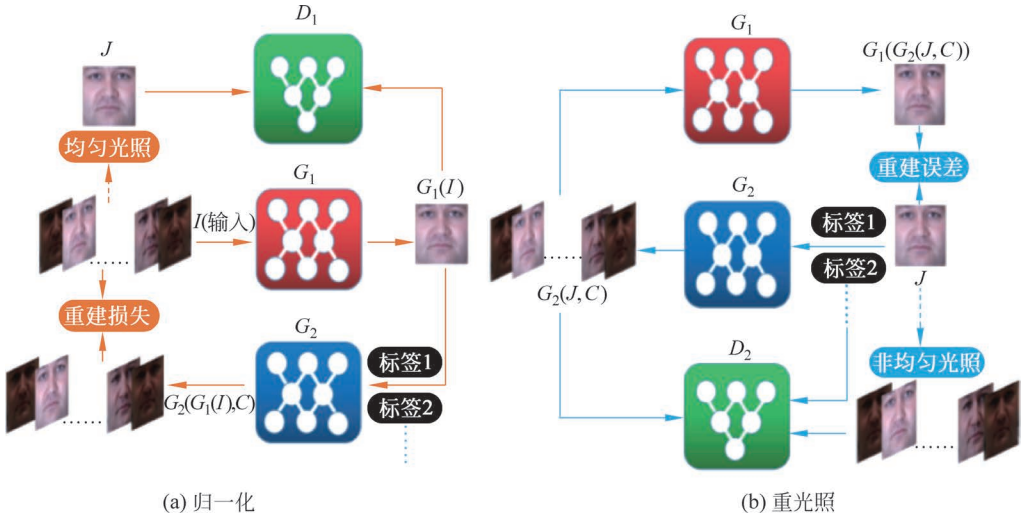


图 5.23　AJGAN 的光照归一化和重光照的数据流

在第一个橙色环路 L_1 中，光照归一化的输入是具有不均匀光照的人脸图像 I，而 J 是 I 所对应的均匀光照的真实人脸图像。循环过程如下：G_1 接收输入 I 并生成 $G_1(I)$，判别器 D_1 判别 $G_1(I)$和真实图像 J 孰真孰假；然后，$G_1(I)$通过 G_2 进行重光照，生成的 $G_2(G_1(I),C)$被约束为与输入 I 保持一致。其中，C 是 I 对应的光照标签。即 $G_1(I)$在 C 的指引下通过 G_2 重建 I。

在第二个蓝色闭环 L_2 中，J 是 G_2 的输入，而 I 是对应的不均匀光照图像。蓝色箭头显示了环路 L_2 的过程：G_2 的生成图像 $G_2(J,C)$在判别器 D_2 的引导下重建 I；然后，$G_2(J,C)$通过 G_1 进行光照归一化；最后，归一化图像 $G_1(G_2(J,C))$被约束为与原始的输入图像 J 保持一致。作为另一个重要的判别器 D_2，其结构不同于 D_1，因为 D_2 不仅要判定图像的真实性，还要对光照标签 C 进行分类。另外，在 G_1 中，I(具有任意的光照类型)和 J(均匀光照)之间的映射是多对一的映射，而将 J 映射回 I 的重光照过程是一对多的映射。重光照的过程依赖于光照标签 C，G_2 重光照示意图如图 5.24 所示。

在图像重建方面，环路 L_1 突出光照归一化的过程 G_1，重光照 G_2 辅助归一化网络 G_1

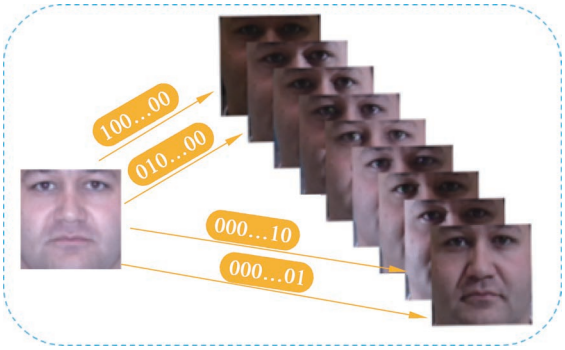


图 5.24　G_2 重光照示意图

重建输入图像，用以确保人脸身份特征不变。而 L_2 突出重光照过程 G_2，归一化网络 G_1 辅助 G_2 重建 G_2 的输入图像。在环路 L_1 和 L_2 中，重光照 G_2 的作用体现在两个方面，具体如下。

（1）通过 G_1 和 G_2 构造的两个闭环结构，使得 AJGAN 能够将不同光照的人脸图像和光照均匀的人脸图像之间的差异的消除强制为光照的转换。利用 G_1 和 G_2 的合作，AJGAN 可以更好地保护人脸结构。

（2）当面对不充分的训练数据时，通过给出新的光照标签 C，可以指导均匀光照的人脸图像 J 转换为新的光照类型的人脸图像。

该方法动态地扩展了训练数据库，减少了模型对数据集大小的依赖。图 5.25 显示了 G_2 的两个角色。同时，该不对称的闭环结构为其他深度学习方法在遇到数据量不足时提供了一个很好的策略。

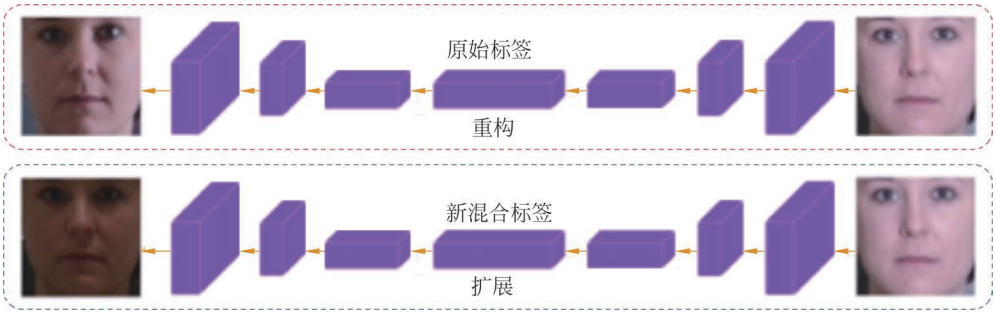


图 5.25　G_2 的两个角色

5.4.3　损失函数

AJGAN 的损失函数主要有 4 项：用于保证转换后的人脸图像的身份特征不变的循环损失、用于保证生成图像更接近真实图像的生成损失、用于保证生成图像的分布与真实图像分布相匹配的对抗损失、用于保证不同标签具有不同光照类型的标签分类损失。

1. 循环损失

将输入图像 I 转换为均匀光照图像 J，要保证生成的图像只是光照的改变，而输入图像 I 的身份特征不改变。基于重建的循环损失可以有效地保护人脸特征。在光照归一化中，

循环损失将光照归一化后的图像再次转换为原始的输入图像，该重建过程形成一个闭环结构：

$$I \to G_1(I) \to G_2(G_1(I),C) \approx I \tag{5.72}$$

循环损失也存在于重光照的过程中，该损失可保证重光照后的图像能够重建 G_2 的输入图像：

$$J \to G_2(J,C) \to G_1(G_2(J,C)) \approx J \tag{5.73}$$

这对于人脸图像的转换有着重要的意义，因为循环损失能够在光照改变的同时保证人脸的身份属性，其损失项表达如下：

$$\begin{aligned} L_{\mathrm{LOOP}}(G_1,G_2) = {} & E_{I\sim p_{\mathrm{data}}(I)}[\,\|G_2(G_1(I),C)-I\|_1\,]+ \\ & E_{J\sim p_{\mathrm{data}(J)}}[\,\|G_1(G_2(J,C))-J\|_1\,] \end{aligned} \tag{5.74}$$

2. 生成损失

AJGAN 期望光照归一化后的生成图像 $G_1(I)$ 与 I 所对应的真实均匀光照图像 J 相似，希望它们具有相同的均匀光照特性。同样，在光照标签 C 的控制下，重光照过程中的生成图像 $G_2(J,C)$ 应与原始图像 I 的光照类型保持一致。同时，为了生成更清晰的人脸图像，使用 L_1 距离来度量生成损失：

$$L_{\mathrm{GEN}}(G_1,G_2) = E_{I,J\sim p_{\mathrm{data}}}[\,\|J-G_1(I)\|_1 + \|I-G_2(J,C)\|_1\,] \tag{5.75}$$

3. 对抗损失

AJGAN 有两组生成-判别结构：生成器 G_1 和相应的判别器 D_1，生成器 G_2 和相应的判别器 D_2。D_1 尝试区分 G_1 的生成图像 $G_1(I)$和对应的真实图像 J。同样地，D_2 尝试区分 G_2 的生成图像 $G_2(J,C)$和真实图像 I。对于每组生成-判别结构，训练过程可以看作一个博弈的过程，其目标函数为：

$$\begin{aligned} L_{\mathrm{ADV}}(D,G) = {} & E_{J\sim p_{\mathrm{data}}(J)}[\log D_1(J)] + E_{I\sim p_{\mathrm{data}}(I)}[\log(1-D_1(G_1(I)))] + \\ & E_{I\sim p_{\mathrm{data}}(I)}[\log D_2(I)] + E_{J\sim p_{\mathrm{data}}(J)}[\log(1-D_2(G_2(J,C)))] \end{aligned} \tag{5.76}$$

4. 标签分类损失

G_2 通过均匀光照图像 J 和光照类型标签 C 生成不均匀的光照图像 $G_2(J,C)$。在训练阶段，不同光照类别的 I 对应不同的光照标签 C，可以使用一个 n 维的 one-hot 标签表示 C，n 表示光照类型数目。若在 D_2 的结构上加上辅助分类的功能，那么可以在优化 D_2 和 G_2 的同时强制对标签进行分类。判别器 D_2 不仅可以检测“伪造图像”，而且可以使得光照标签 C、生成的图像 $G_2(J,C)$及原始图像 I 更加紧密地联系起来。AJGAN 通过最大化光照标签 C 和不同光照类型的图像(包括生成图像 $G_2(J,C)$和输入图像 I)之间的互信息来实现标签与光照类型的联系。通过变分引理，光照标签 C 和光照类型图像之间的互信息推导如下：

$$M(X,Y) = \int_{x\in X}\int_{y\in Y} P(x,y)\log\frac{P(x,y)}{P(x)P(y)}\mathrm{d}x\mathrm{d}y$$

$$= \int_{x\in X}\int_{y\in Y} P(x,y)\log\frac{P(x,y)}{P(x)}\mathrm{d}x\mathrm{d}y - \int_{x\in X}\int_{y\in Y} P(x,y)\log P(x,y)\log P(y)\mathrm{d}x\mathrm{d}y$$
$$= \int_{x\in X}\int_{y\in Y} P(x)P(y\mid x)\log P(y\mid x)\mathrm{d}x\mathrm{d}y - \int_{y\in Y}\log P(y)\int_{x\in X} P(x,y)\mathrm{d}x\mathrm{d}y$$
$$= \int_{x\in X} P(x)\int_{y\in Y} P(y\mid x)\log P(y\mid x)\mathrm{d}x\mathrm{d}y - \int_{y\in Y}\log P(y)P(y)\mathrm{d}x\mathrm{d}y - H(X\mid Y)+H(Y) \tag{5.77}$$

$$\begin{aligned} M(c,G(z,c)) &= -H(c\mid G(z,c))+H(c)\\ &= E_{x\sim G(z,c)}[E_{c'\sim P(c\mid x)}[\log P(c'\mid x)]]+H(c)\\ &= E_{x\sim G(z,c)}[D_{KL}(P(\cdot\mid x)Q(\cdot\mid x))+E_{c'\sim P(c\mid x)}[\log Q(c'\mid x)]]+H(c)\\ &\geqslant E_{x\sim G(z,c)}[E_{c'\sim P(c\mid x)}[\log Q(c'\mid x)]]+H(c)\\ &= E_{c\sim P(c),x\sim G(z,c)}[\log Q(c\mid x)]+H(c) \end{aligned} \tag{5.78}$$

$$\begin{aligned} \min M(C;I,J) = {} & E_{I\sim P_{\text{data}}(I)}[-\log D_2(C\mid I)]+\\ & E_{J\sim P_{\text{data}}(J)}[-\log D_2(C\mid G_2(J,C))] \end{aligned} \tag{5.79}$$

其中，$M(X,Y)$表示 X 和 Y 之间的互信息。

通过对这些公式的分析，可知需要将 C 和$G_2(J,C)$、C 和 I 之间的互信息最大化，即标签的分类损失的定义为：

$$\begin{aligned} L_{\text{CLS}}(G_2,D_2) = {} & E_{I\sim p_{\text{data}}(I)}[-\log D_2(C\mid I)]+\\ & E_{J\sim p_{\text{data}}(J)}[-\log D_2(C\mid G_2(J,C))] \end{aligned} \tag{5.80}$$

其中，$\log D_2(C|I)$和 $\log D_2(C|G_2(J,C))$表示由 D_2 计算的光照标签的概率分布。换句话说，D_2 将具有不同光照类型的图像分类到其对应的光照标签 C。

最终目标函数见式(5.81)：

$$\begin{aligned} L^{*} = {} & \arg\min_{G_1,G_2}\max_{D_1,D_2} \alpha L_{\text{LOOP}}(G_1,G_2)+\beta L_{\text{GEN}}(G_1,G_2)+\\ & \gamma L_{\text{ADV}}(D,G)+L_{\text{CLS}}(G_2,D_2) \end{aligned} \tag{5.81}$$

其中，α、β 和 γ 是超参数，可根据实际情况进行调整。一般情况下，其范围为[1～100]，推荐的参数值分别为 10、20 和 1。

5.4.4 网络的实现和优化

1. 实现细节

AJGAN 中 G_1 和 D_1 采用了 Radford 等人提出的生成器和判别器结构，使用了 convolution-BatchNorm-ReLU 模块，该结构在图像的翻译的任务中有着很好的转换结果。对于生成器 G_1 和 G_2，AJGAN 没有使用传统的编码器-解码器网络，而是使用 U-Net 结构，该结构在每个 i 层和 $n-i$ 层之间添加连接，其中 n 是卷积层的总数。这样的连接层为网络结构提供了可以跳过编码-解码的选项。作为非对称的另一组网络结构 G_2 和 D_2，G_2 的输入不仅有光照均匀的人脸图像，还有引导该图像进行转换的光照标签。由于判别器 D_2 不仅可以检测“伪造图像”，还可以对光照类型进行分类，因此 D_2 的结构不同于 D_1，D_2 采用 StarGAN 的结构，G_2 和 D_2 的网络结构如表 5.4 和表 5.5 所示。

表 5.4　G_2 的网络结构

网络层级	激活大小
输入	31×28×128
4×4×64 卷积层，步幅 2，填充 1	64×64×64
4×4×128 卷积层，步幅 2，填充 1	128×32×32
4×4×256 卷积层，步幅 2，填充 1	256×16×16
4×4×512 卷积层，步幅 2，填充 1	512×8×8
4×4×1024 卷积层，步幅 2，填充 1	1024×4×4
4×4×2048 卷积层，步幅 2，填充 1	2048×2×2
3×3×1 卷积层，步幅 1，填充 1	1×2×2
2×2×nd 卷积层，步幅 1，填充 1	nd×1×1

表 5.5　D_2 的网络结构

网络层级	激活大小
输入图片＋标签	3＋nd×128×128
7×7×64 卷积层，步幅 1，填充 3	64×128×128
4×4×128 卷积层，步幅 2，填充 1	128×64×64
4×4×256 卷积层，步幅 2，填充 1	256×32×32
残差模块：3×3×256 卷积层，步幅 1，填充 1	256×32×32
残差模块：3×3×256 卷积层，步幅 1，填充 1	256×32×32
残差模块：3×3×256 卷积层，步幅 1，填充 1	256×32×32
残差模块：3×3×256 卷积层，步幅 1，填充 1	256×32×32
残差模块：3×3×256 卷积层，步幅 1，填充 1	256×32×32
残差模块：3×3×256 卷积层，步幅 1，填充 1	256×32×32
4×4×128 卷积层，步幅 2，填充 1	128×64×64
4×4×64 解卷积层，步幅 2，填充 1	64×128×128
7×7×3 卷积层，步幅 1，填充 3	3×128×128

2. 优化方法

AJGAN 并没有遵循原始 GAN 中提出的训练方式，因为这可能导致训练过程中梯度消失。Mao 等人使用最小二乘损失替代了 Sigmoid 交叉熵损失，其实验结果表明，与原始的 GAN 相比，最小二乘损失减小了训练的不稳定性问题，提高了生成图像的质量。因此，将式(5.76)中的负对数似然函数替换为最小二乘法损失，如式(5.82)所示。在实验中，分别将最小二乘法的参数 a、b、c 设置为 0、1、1。

$$\min_{D_1} L_{\mathrm{ADV}}(D_1)=\frac{1}{2}E_{J\sim p_{\mathrm{data}}(J)}[(D_1(J)-b)^2]+\frac{1}{2}E_{I\sim p_{\mathrm{data}}(I)}[(D_1(G_1(I))-a)^2]$$

$$\min_{G_1} L_{\mathrm{ADV}}(G_1)=\frac{1}{2}E_{I\sim p_{\mathrm{data}}(I)}[(D_1(G_1(I))-c)^2]$$

$$\min_{D_2} L_{\mathrm{ADV}}(D_2)=\frac{1}{2}E_{I\sim p_{\mathrm{data}}(I)}[(D_2(I)-b)^2]+\frac{1}{2}E_{J\sim p_{\mathrm{data}}(J)}[(D_2(G_2(J))-a)^2]$$

$$\min_{D_1} L_{\mathrm{ADV}}(D_1)=\frac{1}{2}E_{J\sim p_{\mathrm{data}}(J)}[(D_2(G_2(J))-c)^2] \tag{5.82}$$

因为 AJGAN 的主要目的是光照归一化，而不是重光照的过程，因此，先训练光照归一化的过程 G_1 和 D_1。具体而言，梯度下降法在 D_1 和 G_1 上交替进行。为了降低 D_1 相对于 G_1 的学习速率，在优化 D_1 时，将损失函数除以 2。AJGAN 的优化采用小批量随机梯度下降法和 Adam，学习率为 0.0002，动量参数 $\beta_1=0.5$、$\beta_2=0.999$。实验表明，先训练 G_1 和 D_1，再联合训练 G_2 和 D_2 可以得到比同时联合训练更好的效果。AJGAN 在单一的 GeForce 1080 GPU 上训练大约 500 轮次。G 和 D 的学习率为 2×10^{-4}。300 轮次后，每个轮次的学习率衰减系数为 2×10^{-6}。

5.4.5 实验结果

本小节详细讨论了实验的相关设置及定性和定量结果。为了使 AJGAN 更具说服力，将测试数据集分为灰度图像集和彩色图像集。在彩色图像集中，使用目前非常优秀的无监督和有监督的深度学习方法进行对比。在灰度图像集中，添加了两个非常优秀的光照归一化算法。

1. 实验对比方法

目前缺少基于 GAN 的人脸光照归一化工作，在实验中，本节利用 CycleGAN 的闭环结构和 Pix2Pix 的监督学习形成有监督的 SupervisedCycleGAN，并将其作为对比方法。为了公平比较，使用相同的数据集训练 Pix2Pix、CycleGAN、SupervisedCycleGAN 和 AJGAN。Pix2Pix、CycleGAN 的网络参数是其文献提供的默认参数，而 SupervisedCycleGAN 与 CycleGAN 相比，只是训练时使用真实的图像，因此其参数保持和 CycleGAN 的一致。为了减少随机性，每一个网络都经过 5 次训练，并选择其最佳结果进行比较。

除了基于深度学习的方法外，实验还比较了 Xie 等人和 Matsukawa 等人提出的两种先进的手动提取人脸特征的光照归一化方法。不同于直接对原始人脸图像进行归一化，Xie 等人的方法提出将输入图像分解为大小尺度特征的分量，然后分别对大小尺度特征进行归一化。Matsukawa 通过考虑投射阴影对该方法进行了扩展，也取得了很好的归一化效果，即使在极端光照条件下也能有效提取人脸特征。

2. 测试数据库

目前，可用的人脸光照公共数据集有限。本节使用 Multi-PIE、CAS-PEAL 和 Extended-YaleB 数据库对 AJGAN 进行了验证，并与现有方法进行了比较。

Multi-PIE 数据库在人脸识别中得到了广泛的应用。在实验中，使用相机正面拍摄的图像来训练和测试人脸的光照归一化方法。在训练阶段，随机抽取 230 人作为训练数据。对于每个人，将对应的真实均匀光照图像作为真实图像 J，剩下的严格对齐的 19 幅图像作为任意光照下的图像。因此，每个人有 19 个光照图像对，共使用 $19\times230=4370$ 幅图像进行训练。同时，使用 20 个不同人脸，每人 19 种光照条件，分别测试 AJGAN 和对比方法。为了平衡视觉效果和计算成本，利用人脸对齐工具将人脸图像裁剪为 128×128 的大小。

CAS-PEAL 数据库包含多种类型的人脸图像，这些人脸具有不同的朝向、表情、佩饰、光照、背景、距离和年龄。对于光照变化，该数据库记录了 200 多个不同人脸的 10 多种不同光照条件。光源的位置变化引起了光照的变化。在极端光照条件下，许多 CAS-PEAL 图像

的人脸特征丢失。此外，由于该数据库中的图像是在 2min 内拍摄得到的，因此在图像中可以明显地观察到包括朝向、表情在内的变化。这些因素给光照归一化带来了相当大的挑战。该数据库共使用 1468 幅人脸图像进行训练，198 幅图像进行测试。

Extended-YaleB 数据库包含 38 个不同人脸在 64 种光照条件下的正面姿势图像。该数据库中的人脸图像的光照极端的样例较多，很多人脸特征几乎无法用肉眼观察。实验采用 64×32＝2048 幅图像进行训练，384 幅图像进行测试。

3. 彩色图像集的视觉效果

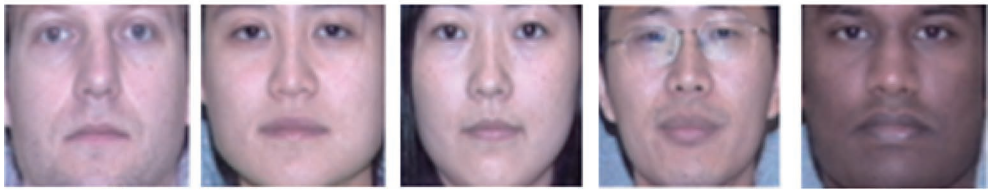
图 5.26 显示了白色、黄色和黑色肤色的人脸图像在黑暗光照条件下的光照归一化结果。尽管使用了真实的成对图像作为训练集，但 Pix2Pix 除了丢失许多精细的面部纹理外，

(a) 输入图像

(b) Pix2Pix

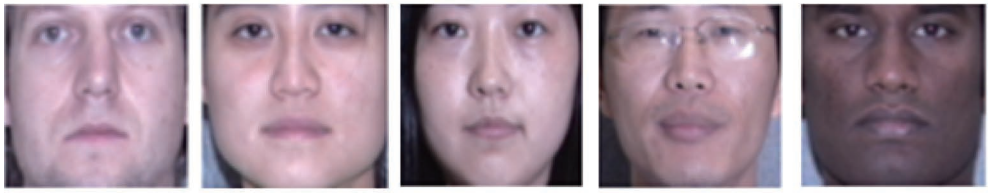
(c) CycleGAN

(d) SupervisedCycleGAN



(e) AJGAN

图 5.26　光照归一化结果



(f) 统一光照图

图 5.26 （续）

甚至明显地改变了人脸的形状，例如，第二列中的图像，人脸的“胖瘦”改变了，这种变化导致难以判断输入图像和输出图像是同一个人，这表明了如果没有闭环结构来加强生成的图像与原始图像之间语义的一致性，就很难保证输入图像的人脸结构特征不变。在闭环结构下，无监督的 CycleGAN 能够生成更细致的纹理，在一定程度上保留了原始图像的人脸特征，这进一步验证了循环重建损失对人脸识别的重要作用。但是，在 CycleGAN 的归一化结果中，人脸的肤色会发生变化，因为如果没有相应的真实图像作为引导，生成的图像在映射回原始图像时，原始图像的光照条件被误认为是肤色。SupervisedCycleGAN 集合了 Pix2Pix 和 CycleGAN 的优点，在某种程度上保护了人脸的身份特征，避免了错误的肤色。然而，在重光照过程中，即将具有均匀光照的人脸图像映射回不同类型的光照图像时，这是一个一对多的过程，本质上是一个欠约束问题，因此，生成的图像是模糊的。相比之下，AJGAN 可在光照标签 C 的引导下找到正确的重建路径，从而生成的图像足够清晰。此外，AJGAN 具有很好的归一化效果，并保留了人脸的身份特征。

图 5.27 所示的光照归一化结果进一步显示了 CycleGAN 可能导致的误映射及 AJGAN 的优越性。CycleGAN 有时会将输入图像映射到不理想的分布。在训练数据库中，某些人脸图像在微笑时露出了牙齿，由于缺乏真实图像的引导，CycleGAN 的归一化结果就会在牙齿附近显示出明显的伪影。虽然归一化的结果表明是同一个人，但嘴巴的形状是扭曲的。SupervisedCycleGAN 完全避免了这一问题，但由于缺乏相应的光照标签，图像比较模糊。相比之下，AJGAN 清楚地展示了光照归一化后的图像。

图 5.28 显示了不同方法对戴眼镜的人脸的归一化结果。Pix2Pix 生成的图像虽然平滑，但眼镜的轮廓却是模糊的。尽管第三列中的 CycleGAN 比 Pix2Pix 拥有更清晰的眼镜轮廓和人脸特征，但是在黑暗的光照条件下，肤色发生了改变。同时，由于缺少真实图像的引导，人脸脸型发生了扭曲。SupervisedCycleGAN 在保护人脸结构特征方面具有优势，而且也没有改变肤色，但生成的图像模糊。AJGAN 取得了非常好的归一化结果，同时保留了大部分的人脸纹理和清晰的眼镜轮廓。

4. 灰度图像集的视觉效果

在这一小节中，使用 Extended-YaleB 和 CAS-PEAL 数据库中的人脸图像来验证灰度集的归一化。此外，实验增加了两种经典的算法 NPL-QI(S&L)和 NPLE-QI(S&L)以进行比较。Extended-YaleB 数据库中的人脸图像归一化结果如图 5.29 所示。

从图 5.29 可以看出，NPL-QI(S&L)和 NPLE-QI(S&L)在保护人脸结构方面具有绝对优势，因为它们直接对源图像进行归一化而不是“制造”均匀光照图像。然而，它们的视觉效果总体上比基于深度学习的方法差。尽管与 NPL-QI(S&L)和 NPLE-QI(S&L)相比，

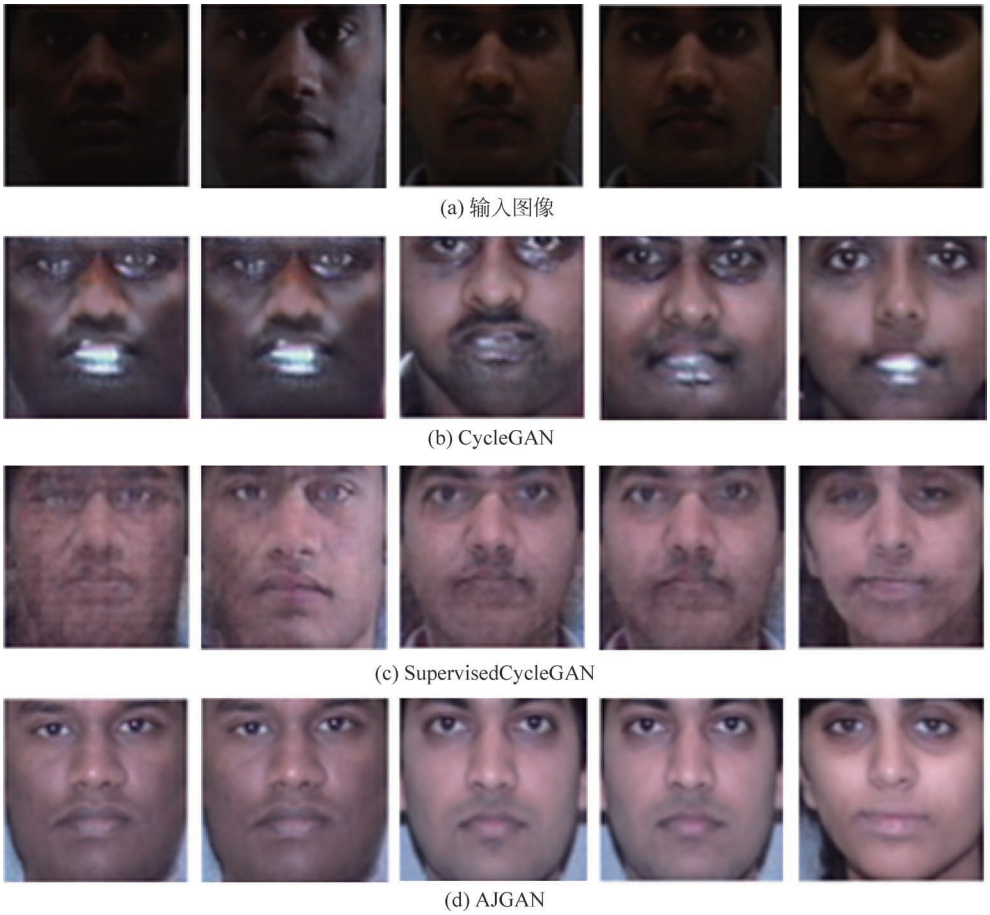
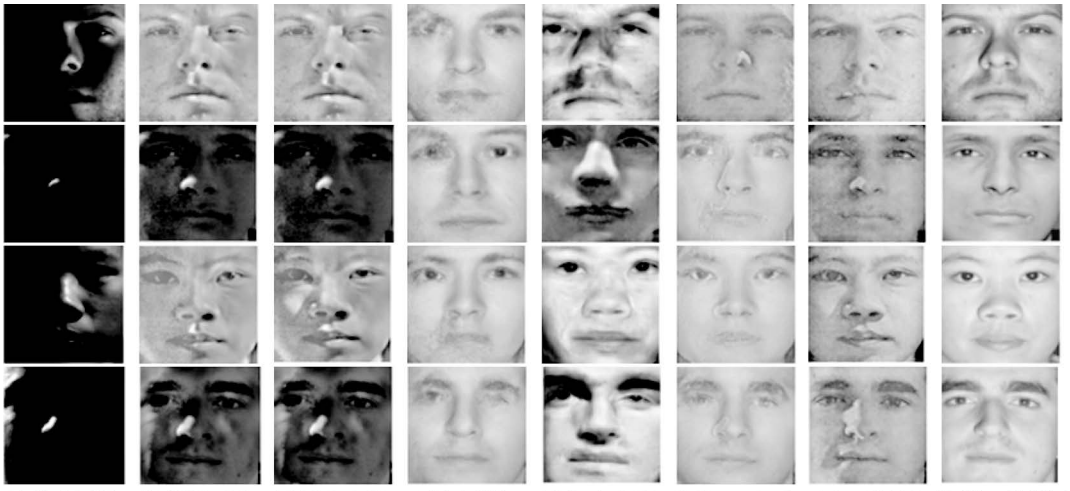


图 5.27　光照归一化结果

图 5.28　不同方法对戴眼镜的人脸的归一化结果

Pix2Pix 能够实现更均匀的光照条件，但它明显改变了输入人脸的形状，肉眼可以观察到人脸形状的明显改变。CycleGAN 在保护人脸结构方面优于 Pix2Pix。但是，CycleGAN 归一化后的人脸形状发生扭曲，并且在眉毛、嘴和鼻子上引入了伪影。SupervisedCycleGAN 很

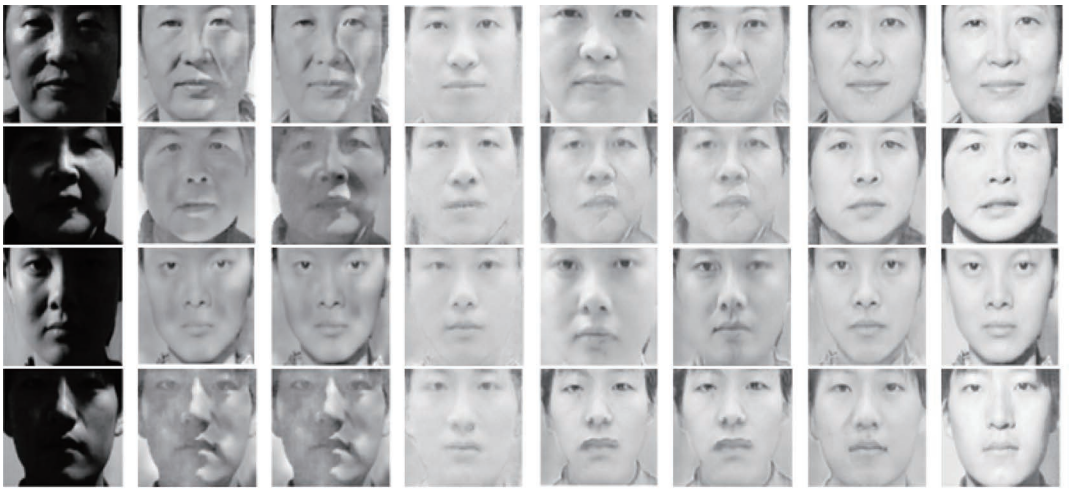


(a) 输入图像 (b) NPL- QI (S&L) (c) NPLE-QI (S&L) (d) Pix2Pix (e)CycleGAN (f) Supervised-CycleGAN (g) AJGAN (h) 统一光照图

图 5.29 Extended-YaleB 数据库中的人脸图像归一化结果

好地保留了人脸特征，但由于在重光照中缺乏指导，因此生成的图像相对模糊。相比之下，尽管在极端的光照条件下，与 Pix2Pix 和 CycleGAN 相比，AJGAN 并没有出现人脸形状的变化。此外，AJGAN 比 NPL-QI(S&L)和 NPLE-QI(S&L)具有更好的光归一化效果。

CAS-PEAL 数据库中的人脸图像光照归一化比较结果如图 5.30 所示。由于滤波操作直接作用在原图的像素值上，因此在极端的光照条件下，NPL-QI(S&L)和 NPLE-QI(S&L)的归一化效果并不是很好。在该数据库中，由于拍摄时间长，因此用于训练的图像对没有完全对齐，这给有监督的学习方法带来了相当大的挑战。例如，在图 5.30(a)列中输入的测试图像和图 5.30(h)呈现的真实均匀光照图像之间，可以观察到面部表情和朝向的差异。在这种情况下，Pix2Pix 完全改变了原始的人脸身份，CycleGAN 和 SupervisedCycleGAN 无法



(a) 输入图像 (b) NPL-QI (S&L) (c) NPLE-QI (S&L) (d) Pix2Pix (e) CycleGAN (f) Supervised-CycleGAN (g) AJGAN (h) 统一光照图

图 5.30 CAS-PEAL 数据库中的人脸图像的归一化结果

生成清晰的图像。但是，AJGAN 明显优于其他方法，光照归一化效果比较理想的同时保证了人脸的身份属性。

5. 定量比较

为了定量比较 AJGAN 和其他方法，先使用 PSNR 来衡量光照归一化后生成图像的质量，再使用 RMSE、SSIM 来测量生成图像和真实图像的相似性。对于每个数据库，随机选取 50 个测试图像进行定量分析，并计算其平均值。由于光照归一化的目的是便于人脸识别，为了定量地评价 AJGAN 的通用性和鲁棒性，可在光照环境变化较大的人脸数据库上进行人脸识别实验。

表 5.6 展示了 AJGAN 和对比方法在 PSNR 上的比较结果。PSNR 用于衡量光照归一化后生成图像的质量。PNSR 越高，图像质量越好。从表 5.6 中可以看出，在不同的数据库中，Multi-PIE 总体上要比其他两个数据库的 PSNR 高，这是因为该数据库的光照条件没有那么极端。然而，Pix2Pix 的测试结果并不能令人满意，这侧面证明了该方法生成的图像丢失了许多细节特征。SupervisedCycleGAN 比 CycleGAN 高，这是因为 CycleGAN 没有使用真实的图像进行训练，一定程度上导致人脸形状变形。此外，SupervisedCycleGAN 和 CycleGAN 整体上优于 NPLE-QI(S&L)。从表 5.6 中可以看出，AJGAN 比其他方法好得多。在光照标签的引导下，AJGAN 可以在重建过程中准确地映射回源图像，从而生成比 SupervisedCycleGAN 更清晰的图像。

表 5.6　AJGAN 和对比方法在 PSNR 上的比较结果

对 比 方 法	PSNR 值		
	MultiPIE	CAS-PEAL	Extended-YaleB
NPL-QI(S&L)	25.07	24.88	24.09
NPLE-QI(S&L)	25.10	24.91	24.18
Pix2Pix	28.16	24.42	23.58
CycleGAN	29.15	24.57	24.11
SupervisedCycleGAN	31.02	24.82	24.15
AJGAN	34.90	25.18	24.23

表 5.7 展示了 AJGAN 和对比方法在 RMSE 上的比较结果。由于 RMSE 用于测量归一化图像与真实图像的相似程度，因此它在一定程度上反映了光照归一化的有效性。很明显，较小的 RMSE 意味着归一化后的图像更接近于真实情况。如表 5.7 所示，NPL-QI(S&L)和 NPLE-QI(S&L)的 RMSE 值大于基于 GAN 的方法，这是因为 NPL-QI(S&L)和 NPLE-QI(S&L)方法的滤波操作极大地改变了输入图像的像素值。基于 GAN 的方法中的对抗性损失使得输出图像的分布与均匀光照图像的分布相匹配。进一步地，基于 GAN 的方法利用真实图像作为生成图像的真值，并根据二者之间的差异构建损失函数迭代优化以生成图像，从而使得生成图像中的光照分布与真实图像的光照分布更为接近。从表 5.7 中可以看到，SupervisedCycleGAN 和 Pix2Pix 优于 CycleGAN。显然，AJGAN 获得了最低的 RMSE，证明了输入图像的光照被很好地归一化。

表 5.7　AJGAN 和对比方法在 RMSE 上的比较结果

对比方法	RMSE 值		
	MultiPIE	CAS-PEAL	Extended-YaleB
NPL-QI(S&L)	46.55	57.93	78.92
NPLE-QI(S&L)	44.98	56.77	78.09
Pix2Pix	28.29	38.79	31.61
CycleGAN	42.08	46.27	54.06
SupervisedCycleGAN	22.54	39.77	23.68
AJGAN	20.12	31.13	22.08

SSIM 是一种用于测试两幅图像之间结构相似性的指标，被认为与人类视觉系统的感知有关。SSIM 值越接近 1，两幅图像的结构越相似。如表 5.8 所示，NPL-QI(S&L)、NPLE-QI(S&L)和 Pix2Pix 优于 CycleGAN，但低于 SupervisedCycleGAN 和 AJGAN。这是因为 NPL-QI(S&L)和 NPLE-QI(S&L)的图像尺度分解操作在一定程度上保留了图像结构。同理，Pix2Pix 是监督学习的，在训练过程中，也保留了结构特征。相反，CycleGAN 的无监督的光照归一化方法对输入图像的结构信息的保存能力较差。然而，监督学习能克服这种不足。AJGAN 取到的最大 SSIM 值证明了 AJGAN 在光照归一化过程中有效地保护了人脸结构。

表 5.8　AJGAN 和对比方法在 SSIM 上的比较结果

对比方法	SSIM 值		
	MultiPIE	CAS-PEAL	Extended-YaleB
NPL-QI(S&L)	62.12	59.11	47.74
NPLE-QI(S&L)	63.37	61.04	45.88
Pix2Pix	64.21	61.70	57.09
CycleGAN	55.07	42.50	39.32
SupervisedCycleGAN	74.59	61.78	64.63
AJGAN	75.9	70.95	70.44

6. 人脸识别测试

在本实验中，不同方法的归一化图像作为识别对象，真实均匀光照的图像作为被对比的图像。从 3 个数据库中随机选择 500 幅图像计算最后的识别率，并利用 VGGFace2 算法提取人脸特征，采用其预训练模型，然后使用归一化相关系数作为相似度度量。表 5.9 显示了 3 个数据库的识别率的比较，从表 5.9 中可以发现，基于 GAN 的方法，包括 CycleGAN 和 Pix2Pix，都弱于直接对原图像操作的 NPL-QI(S&L)和 NPLE-QI(S&L)方法，这是因为 CycleGAN 和 Pix2Pix 都会导致人脸形状的变形并失去个性化的精细特征。此外，由于 CycleGAN 的循环结构有助于保护人脸身份，因此 CycleGAN 的识别率远高于 Pix2Pix。在监督学习和循环结构的帮助下，SupervisedCycleGAN 在 3 个数据库上的识别率均高于 Pix2Pix 和 CycleGAN，但由于缺乏光照条件的指引，因此其识别率要低于 AJGAN。尽管使用非严格对齐的图像对作为训练数据，但在 CAS-PEAL 数据库上，AJGAN 将 NPLE-QI

(S&L)的识别率提高了3%,这证明了该算法的适用性。

表5.9　3个数据库的识别率的比较结果

对比方法	人脸识别率/%		
	MultiPIE	CAS-PEAL	Extended-YaleB
NPL-QI(S&L)	98.1	84.9	86.1
NPLE-QI(S&L)	98.3	84.9	86.2
Pix2Pix	95.6	72.1	51.7
CycleGAN	97.2	82.5	80.7
SupervisedCycleGAN	98.1	84.1	81.8
AJGAN	99.9	87.8	87.2