

山东大学

毕业论文(设计)

论文(设计)题目: 基于颜色恒常性的图像合成

姓 名 唐超
学 号 201600181073
学 院 计算机科学与技术学院
专 业 人工智能
年 级 2016 级
指导教师 钟凡

2020 年 5 月 24 日

目 录

摘 要.....	I
ABSTRACT.....	II
第 1 章 绪 论.....	1
1.1 选题背景和研究意义.....	1
1.2 文献综述.....	1
1.3 论文主要研究内容.....	2
1.4 论文组织结构.....	3
第 2 章 成像理论.....	4
2.1 颜色感知.....	4
2.1.1 可见光.....	4
2.1.2 物体.....	5
2.1.3 观测者.....	5
2.2 颜色描述.....	6
2.2.1 RGB 颜色空间.....	6
2.2.2 HSL 颜色空间.....	7
2.3 成像模型.....	8
2.3.1 朗伯反射模型.....	8
2.3.2 双色反射模型.....	9
第 3 章 光照一致的图像合成.....	10
3.1 颜色恒常性.....	10
3.2 基于假设统计的颜色恒常性算法.....	11
3.2.1 White-Patch 算法.....	11
3.2.2 Grey-World 算法.....	11
3.2.3 Grey-Edge 算法.....	12
3.3 基于灰色像素的颜色恒常性光照估计.....	12
3.3.1 灰色像素.....	13
3.3.2 寻找灰色像素的算法框架.....	13

3.3.3 灰色指数 (Grey Index)	14
3.3.4 多光源的情况.....	15
3.4 基于光照估计的图像合成	15
第 4 章 实验与分析.....	19
4.1 设计实验	19
4.1.1 数据集.....	19
4.1.2 评价指标.....	19
4.1.3 实验流程.....	20
4.2 结果分析	21
第 5 章 结束语.....	25
5.1 本文总结	25
5.2 工作展望	25
致 谢.....	26
参考文献.....	26
附录 1 代码及相关附件	29
附录 2 文献英文原文	34
附录 3 文献中文译文	38

基于颜色恒常性的图像合成

摘 要

图像合成的一般思路是从一张前景源图像中提取出目标前景,然后将前景根据背景图像做出一定修正,最后合成到背景图像中去,修正内容可能包括大小、颜色、亮度、纹理等,本论文主要研究在提取出目标前景后,如何修正前景使其与背景的光照匹配,增强合成图像的真实感。

颜色恒常性(Color Constancy)是人类颜色感知系统的一个特征,它确保了人在不同光照条件下感知到的物体颜色保持相对恒定,如一个绿色的苹果在中午(白光)和傍晚(红光)看起来都是绿色的,这有助于我们识别物体。颜色恒常性算法试图赋予机器同样的能力,通过估计图像拍摄场景的光照,进而恢复图像中物体的本征颜色。

因此,颜色恒常性对光照不一致的图像合成问题具有一定的启发意义,本论文基于颜色恒常性理论提出了一个有效匹配前景与背景光照的图像合成算法:基于双色反射模型寻找图像中的灰色像素,进而通过灰色像素估计前景和背景图像的光照,最后将前景图像映射到背景光照下。

论文最后在数据集上对提出的图像合成算法做了测试,与 cut-and-paste 相比,大部分图像的合成效果都得到了显著提升,根据合成效果较差的图片分析了算法的局限,并提出了未来的改进方向,即计算图像多光源下的光照热图和单独计算光照亮度。

关键词: 图像合成; 颜色恒常性; 灰色像素; 双色反射模型; 光照一致

ABSTRACT

The general idea of image composition is to extract the target foreground from a foreground source image, then modify the foreground according to the background image, and finally synthesize it into the background image. The correction content may include size, color temperature, brightness, style, etc. This paper mainly studies how to modify the foreground to match the illumination of the background after extracting the target foreground, and enhance the realism of the composite image.

Color constancy is a characteristic of human color perception system, which ensures that the color of objects perceived by people under different lighting conditions remains relatively constant. For example, a green apple looks green at noon (white light) and in the evening (red light). It helps us identify objects. The color constancy algorithm attempts to restore the intrinsic color of the object by estimating the its illumination.

Therefore, the estimation of image illumination in color constancy is of some enlightening significance to the image composition problem of inconsistent illumination. Based on the theory of color constancy, this paper proposes an image composition algorithm that effectively matches the foreground and background illumination: I search the grey pixels in the image based on the Dichromatic Reflection Model, and then use them to estimate the illumination of the foreground and background image. Finally, the foreground image is mapped to the background illumination according to the Von Kries diagonal theory.

The proposed image composition algorithm was tested on the dataset. Compared with cut-and-paste, the effect of most images has been significantly improved. Based on the pictures with poor synthesis effects, the limitations and future improvement directions are analyzed: to calculate the heat map of the image under multiple light sources or to calculate the lightness of the light separately.

Keywords: Image Composition; Color Constancy; Grey Pixels; Dichromatic Reflection Model; Illumination Constancy

第1章 绪 论

1.1 选题背景和研究意义

图像合成的目标是将目标前景合成到背景图像中后,使得合成图像与实际拍摄的照片具有同样的视觉感受。图像合成在拍照修图、媒体制作以及增强现实等领域有着重要的应用。

图像合成涉及目标前景许多特征的调整,如大小、形状、位置、纹理、光照等,其中合成图像的光照是特别重要的一个因素,因为人对这一因素十分敏感。然而生活中遇到的多数目标前景图像与背景源图像所在场景的环境光照并不一致,若直接将目标前景复制到背景源图像当中会造成合成图像呈现出明显的合成痕迹。为解决这一问题,传统方法往往将背景图像的颜色信息转移到前景上,但它合并了光照与物体表面反射率的影响,例如,将一个人作为前景合并到树林这一背景中去,背景中的树是绿的是因为它们本身是绿的,而不是由于光照的影响,因此将人的皮肤修改为绿色显然是不合适的。所以将图像所在场景的光照从图像中物体的颜色中提取出来是十分必要的,在此基础上去除目标前景在源图像下光照,模拟在新背景下的光照效果,从而增强合成图像的真实感。

Photoshop 等图像编辑软件做图像合成的使用门槛一般较高,需要一定的图像处理和软件使用知识,这也使得本论文的研究具有一定的现实意义。

1.2 文献综述

在图像合成中,最简单的方法是 cut-and-paste,即仅调整目标前景图像大小、角度等,而不对前景的光照、颜色、纹理等特征做任何处理,再复制到背景源图像中的合适位置,这样得到的合成图像往往并不真实,为了使合成图像效果更加真实自然,人们发展出了一系列不同的图像合成方法。

传统的图像合成方法主要包括基于 α 通道和基于梯度域的图像合成方法。前者是指直接将目标前景基于 α 分量线性混合到背景图像上,这种方法简单直接,且能解决合成图像前景与背景边界光滑过渡的问题,在前景源与背景源接近时会取得不错的合成效果。后者在图像像素的梯度域上进行操作,可以实现合成图像

上前景与背景亮度、对比度等的融合^[1]，泊松克隆是这一类方法的典型例子^[2]，它首先计算源图像的梯度场，接着计算合成图像的梯度场并进一步计算其散度，最后根据散度和合成图像的边界条件分别求解各通道的泊松方程，即可得到合成图像中目标前景的像素值。

还有一类重要的图像合成算法是基于像素统计信息的，Photoshop 中常用的 color transfer 算法，主要思想就是调整目标前景图像使其像素直方图的均值和方差与背景图像的一致，达到图像合成的目的，但这种方法在很多情况下很难保证调整之后的前景看起来真实并且与背景和谐。Su Xue 等人在这思路上进行改进，从真实图片数据集上寻找使前景和背景匹配的关键指标，包括亮度、色温、饱和度等，再通过人工评价的方式对寻找这些指标的变化对图片真实性的影响，发现不同图片的像素直方图中匹配的区域（ $\{H, M, L\}$ ）不一样，作者基于此设计了一个自动合成图像算法，训练了一个随机森林分类器预测各指标最应匹配的区域，并调整该区域使该指标与背景的匹配^[3]。

还有一些工作尝试用深度学习的方法做图像合成，如 Yi-Hsuan Tsai 等人采用基于 CNN 的 encoder-decoder 并结合图像的语义信息对前景进行调整^[4]，基于生成对抗网络（GAN）的图像合成在近年也取得了非常不错的合成效果，通过生成器生成合成图像并用鉴别器区分真实图像与合成图像，使得生成器的结果越来越接近真实图像，Wenyan Cong 等人基于 GAN 结合注意力机制、域验证鉴别器学习图像合成^[5]，Mathieu Garon 等人也采用深度学习在室内图片光照估计上取得了非常好的效果^[6]，可以用于光照一致的图像合成。

1.3 论文主要研究内容

本论文从颜色恒常性的角度来研究图像合成，通过颜色恒常性算法估计图像拍摄场景的光照，进而匹配目标前景和背景图像的光照，达到合成图像的目的。

由于算法本身的特性，本论文的研究对象主要是前景与背景光照不一致的图像合成问题，不包括一些复杂的图像合成问题，如前景区域极小或极偏，前景背光或一些语义相关的情况，也不涉及前景分割。

通过颜色恒常性算法估计光照是本论文的重点，为了深入理解颜色恒常性这一课题，首先研究了成像理论，接着简述了颜色恒常性理论及相关经典算法，和

本文用以作图像合成的基于灰色像素的颜色恒常性算法,然后提出了基于光照一致的图像合成模型,最后在数据集上验证了这一模型并对实验结果做了细致的分析。

1.4 论文组织结构

本论文内容组织如下:

第 1 章绪论。

第 2 章介绍了与成像相关的基础理论。首先介绍了感知颜色的三大要素,接着阐述了与研究题目相关的几种颜色空间和成像模型。

第 3 章提出了基于颜色恒常性的图像合成完整算法,是本论文的核心。本章从颜色恒常性的基本内容开始,进而介绍了基于假设统计的经典颜色恒常性算法和基于灰色像素的颜色恒常性算法,在此基础上提出了光照一致的图像合成算法,并对多光源的情况做了扩展。

第 4 章设计实验验证了上一章提出的图像合成算法,详细介绍了实验数据集、评价指标和实验流程,并对实验结果做了探讨,分析了算法的优点和局限。

第 5 章是研究总结与展望。

第2章 成像理论

这一部分是颜色恒常性理论和光照一致的图像合成的理论基础，分颜色感知、颜色描述和成像模型三个方面进行介绍。

2.1 颜色感知

人对颜色的感知可以概括为以下过程：可见光从光源发出并照射到物体表面上，部分能量被物体吸收，另一部分经物体反射进入人眼，最后被大脑组织接受进而产生视觉效应^[7]。在这一过程中，可见光、物体本身、人的颜色感知系统是必不可少的三大要素。

2.1.1 可见光

作为颜色的本质，光是人感知颜色的基础。一般来说，人眼所能响应的光被称为可见光，其波长范围在 $380nm - 780nm$ 之间。人眼所感知的颜色从紫色到红色，光的波长由短变长。

以波长 λ 为中心的微小波长范围内的辐射量与该波长的宽度之比称为光谱密度，是波长的函数，又称为光谱分布函数，简称光谱。人的视觉可以感受到的光谱称为可见光谱，是整个电磁波谱中的一小部分，如图 2-1 所示。

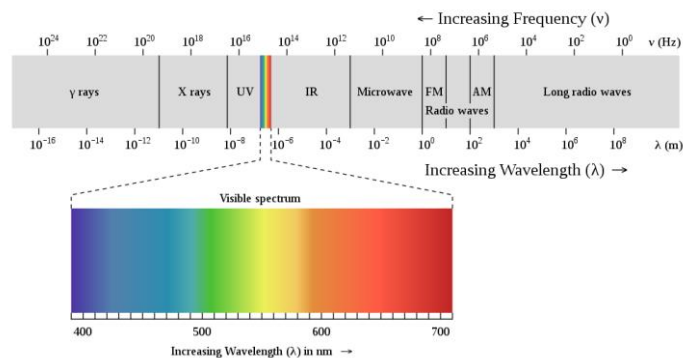


图 2-1 可见光谱

需要注意的是，光源的颜色特性由发出光线中不同波长光的相对能量比例决定，光源的强弱由光谱密度大小决定^[8]。

2.1.2 物体

物体在这一研究课题下可分为荧光物体（自身发光）和非荧光物体（自身不发光）两类，本论文的研究对象是非荧光物体。对于非荧光物体，无外界光照条件下，观测者是无法感知到物体颜色的。观测者之所以能够感知物体颜色，是因为环境光与物体之间发生了反射作用，根据物体表面物理性质的不同，反射可分为三种类型^[9]：

- 1) 镜面反射，遵守反射定律；
- 2) 完全漫反射，这样的理想反射面也称为朗伯表面；
- 3) 同时存在镜面反射和漫反射。

如图 2-2，左图为漫反射，右图为镜面反射。



图 2-2 光在物体表面的反射

2.1.3 观测者

物体颜色不仅受光源和物体的反射（光谱特性）影响，还取决于人眼的生理特性。可见光经过物体表面反射后进入观察者的眼睛，经过人眼视网膜的四种响应细胞，分别为响应红（长波）、绿（中波）、蓝（短波）信号的三种视锥细胞，响应反射光亮度信号的视杆细胞。当这四种细胞接收到反射光并产生响应后，各自会形成物理刺激，这些物理刺激接着被传入大脑中枢，最终使人对物体颜色产生感知。

电子成像设备的成像过程是对人眼成像机制的模拟，数码相机将镜头收集到的光信号转换为电信号，如图 2-3 所示。

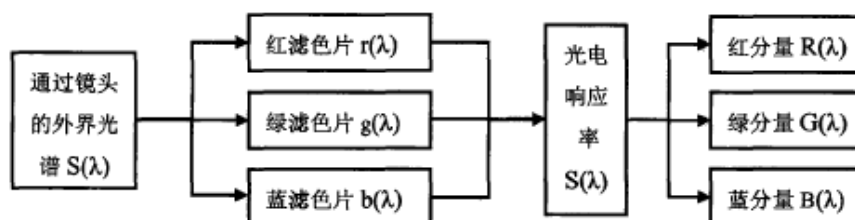


图 2-3 数码相机成像原理

2.2 颜色描述

日常生活中，人们常用天蓝、嫩绿、金黄等名称来表示颜色，但这种方法只是一种定性的描述，在科研中我们需要一个不受人主观影响的对颜色的定量描述。对颜色进行数学定量描述的方法称为颜色空间，关于颜色的研究发展至今，已经发展出了多种多样的颜色空间，包括 RGB 颜色空间、CIE XYZ 颜色空间、CIE Lab 颜色空间，HSL 颜色空间等，这里主要阐述对机器友好的 RGB 颜色空间和更加人性化的 HSL 颜色空间，它们都与本课题的研究相关。

2.2.1 RGB 颜色空间

三色学假说指出，红、绿、蓝三种原色的混合可以表示所有颜色，人眼通过三种视锥细胞对不同颜色做出响应的相对比例，实现对不同颜色的感知。将三种原色混合成其他不同颜色，对混合后的颜色响应值量化，用公式表示如下：

$$C_{\lambda}(C) = r(R) + g(G) + b(B) \quad (2-1)$$

其中， $C_{\lambda}(C)$ 表示波长为 λ 的光的颜色 C ， R 表示波长为 700nm 的光的颜色（红色）， G 表示波长为 540nm 的光的颜色（绿色）， B 表示波长为 430nm 的光的颜色（蓝色）， r, g, b 分别表示它们的比例，即三刺激值。RGB 颜色空间的示意图如图 2-4 所示。

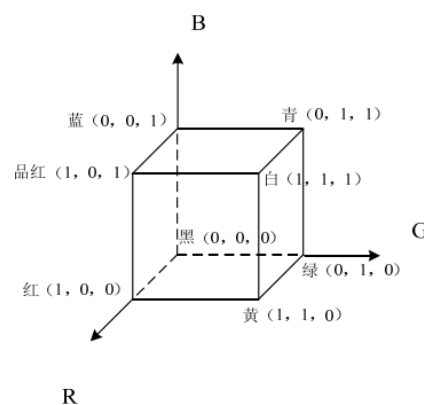


图 2-4 RGB 颜色空间

在计算机中定义颜色或在显示器上显示颜色时，一般采用 RGB 模式，通过 R、G、B 分别表示某一颜色的三刺激值，即通过 RGB 三个通道各灰度值的不同大小描述任意颜色。

RGB 颜色空间直观简单且容易理解，但缺点也是明显的，即 RGB 三个通道相关度较高，单个通道的灰度值对整个颜色影响较大。

2.2.2 HSL 颜色空间

从人的视觉和心理学角度出发，彩色具有三个属性：色相、饱和度、亮度，其中色相和饱和度又统称为色度，我们感受到的任一彩色光都来自这三个属性的综合效果。

色相 (Hue)：也叫色调，是一种颜色区别于其他颜色的特征，由光的频率高低差别所决定。色相只是纯粹地表示色彩相貌的差异，如红、绿、紫、黄等是不同的基本色相。

饱和度 (Saturation)：也叫纯度，指色彩的鲜艳程度。原色纯度最高，颜色混合地越多则纯度越低。饱和度与光波的振幅有关。

亮度 (Lightness)：也叫明度，指颜色的深浅、明暗程度，当反射的光照强弱不一时，就会产生不同程度的明暗。亮度也与光波的振幅有关。

HSL (Hue, Saturation, Lightness) 颜色空间就是根据以上彩色的三大属性建立的一种颜色空间，可以视作 RGB 颜色空间中的点在色相-饱和度-亮度圆柱坐标系中的表示法，如图 2-5 所示。

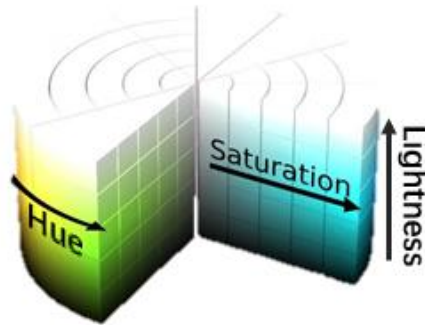


图 2-5 HSL 颜色空间

RGB 颜色空间与 HSL 颜色空间可以互相转换，设 (r, g, b) 分别是一个颜色的 RGB 三刺激值，设 \max 表示 r, g, b 的最大者， \min 表示这些值中的最小者。设 $h \in [0, 360^\circ)$ 表示色相角， $s, l \in [0, 1]$ 分别表示饱和度和亮度，则该像素在 HSL 空间中对应的坐标 (h, s, l) 可通过下列公式计算：

$$h = \begin{cases} 0^\circ & \text{max=min} \\ 60^\circ \times \frac{g-b}{\text{max}-\text{min}} + 0^\circ, & \text{max} = r \text{ and } g \geq b \\ 60^\circ \times \frac{g-b}{\text{max}-\text{min}} + 360^\circ, & \text{max} = r \text{ and } g < b \\ 60^\circ \times \frac{b-r}{\text{max}-\text{min}} + 120^\circ, & \text{max} = g \\ 60^\circ \times \frac{r-g}{\text{max}-\text{min}} + 240^\circ, & \text{max} = b \end{cases} \quad (2-2)$$

$$l = \frac{1}{2}(\text{max} + \text{min}) \quad (2-3)$$

$$s = \begin{cases} 0 & l = 0 \text{ or } \text{max} = \text{min} \\ \frac{\text{max}-\text{min}}{\text{max}+\text{min}}, & 0 < l \leq \frac{1}{2} \\ \frac{\text{max}-\text{min}}{2-(\text{max}+\text{min})}, & l > \frac{1}{2} \end{cases} \quad (2-4)$$

HSL 颜色空间相比 RGB 颜色空间更接近人们对色彩的感知经验，即更加面向用户。

2.3 成像模型

2.3.1 朗伯反射模型

朗伯假设认为生活中所有的表面都是理想的漫反射表面，即不管照明分布如何，该表面在所有方向上接受并发散所有的入射光，结果是从所有视角方向上观测该表面都具有相同的亮度。在朗伯假设下，物体表面的镜面反射被忽略。

人之所以能看到物体，是因为光线经过物体的反射进入了我们的视觉系统。对于摄像头而言，图像传感器将光信号转化为模拟电信号，再通过进一步的后续处理得到我们看到的图片。图像传感器在这一过程中起到的作用是对人类视网膜的简化。基于基本的成像原理，可以将相机图像传感器的成像过程描述为：

$$f_c(x) = \int_{\omega} e(\lambda) s(x, \lambda) c(\lambda) d\lambda \quad (2-5)$$

其中 λ 为可见光波长， x 为像素位置， $e(\lambda)$ 为场景中的光照， $s(x, \lambda)$ 为物体表面光谱反射率，是物体的固有属性。 c 表示图像的 RGB 通道， $c(\lambda) = (R(\lambda), G(\lambda), B(\lambda))$

是图像传感器在各通道上的光响应（镜头敏感度）^[10]。在单一光照下，观测到的光照颜色取决于图像传感器的敏感度 $c(\lambda)$ 及光源的颜色 $e(\lambda)$ ，可用公式（2-6）表示：

$$e = \begin{pmatrix} R_e \\ G_e \\ B_e \end{pmatrix} = \int_{\omega} e(\lambda) c(\lambda) d\lambda \quad (2-6)$$

2.3.2 双色反射模型

日常生活中的物体大多是不透明非均质材料，在环境光的照射下一般同时存在镜面反射与漫反射两种反射过程。Shafer 等人提出了双色反射模型 (Dichromatic Reflection Model) 来描述不透明非均质材料物体的反射规律，该模型同时考虑了镜面反射和漫反射两个独立的物理反射过程，认为物体表面反射光的强度是两者的线性组合^[11]，可用下式表示：

$$I_i^{(x,y)} = \gamma_b^{(x,y)} F_i L_i R_{b,i}^{(x,y)} + \gamma_s^{(x,y)} F_i L_i R_{s,i}^{(x,y)}, i \in \{R, G, B\} \quad (2-7)$$

其中 F 为镜头敏感度， L 为全局光照， R_b, R_s 表示物体的光谱反射率， γ_b, γ_s 是与位置有关的两种反射的几何参数，其值依赖于该处表面的物理性质。

双色反射模型说明了一幅图像中的所有像素值由漫反射和镜面反射两部分组成，每部分又能分解为一个光照分量，即依赖于反射光波长而与物体表面几何结构无关的光谱功率分布：

$$e = [F_R L_R, F_G L_G, F_B L_B]^T \quad (2-8)$$

和一个反射分量，即依赖于物体表面几何结构而与波长无关的物体固有属性：

$$\begin{aligned} W^{(x,y)} &= [\gamma_b^{(x,y)} R_{b,R}^{(x,y)}, \gamma_b^{(x,y)} R_{b,G}^{(x,y)}, \gamma_b^{(x,y)} R_{b,B}^{(x,y)}]^T, \\ V^{(x,y)} &= [\gamma_s^{(x,y)} R_{s,R}^{(x,y)}, \gamma_s^{(x,y)} R_{s,G}^{(x,y)}, \gamma_s^{(x,y)} R_{s,B}^{(x,y)}]^T \end{aligned} \quad (2-9)$$

因此，双色反射模型又可以简化为

$$I^{(x,y)} = W^{(x,y)} \circ e + V^{(x,y)} \circ e \quad (2-10)$$

第3章 光照一致的图像合成

光照一致的图像合成需要估计出前景和背景图像的光照，而前景提取部分不是本论文讨论的重点，因此本章在前景提取完成的假设下，基于颜色恒常性相关算法讨论光照一致的图像合成算法。

3.1 颜色恒常性

颜色恒常性(Color Constancy)是人类主观恒常性(Subjective Constancy)的一个例子(主观恒常性还包括大小恒常性、形状恒常性等)，是人类颜色感知系统的一个特征，它确保了在不同光照条件下感知到的物体颜色保持相对恒定。例如，一个绿色的苹果在主要照明是白光的中午和主要照明是红光的傍晚看起来都是绿色的，这有助于我们识别物体。

在拍摄物体时，由于物体表面反射率、成像传感器的敏感度、光照等的不同，相同场景下拍摄的照片也会呈现出不同的颜色，因此颜色这一图像特征并不具有足够的鲁棒性。假如我们能赋予计算机以人类相同的颜色恒常性能力，即能够将不同场景下的相同物体识别为相同的颜色，许多计算机视觉问题都可以通过颜色恒常性算法得到改善。手机摄像的白平衡功能正是常生活中颜色恒常性的一个重要应用。另外，颜色恒常性还应用于人脸检测、物体识别、视频监控等领域中，本文将利用颜色恒常性的相关算法对待合成图像的光照进行估计，并进一步根据 Von Kries 对角理论对图像作不同光照下的映射。

颜色恒常性算法一般包括两个步骤：(1) 估计图像所在场景的光照；(2) 计算光照不变描述子(illuminant invariant descriptors)^[12]，这通常通过将图像校正到标准白光下完成，如图 3-1 所示。其中第一步估计光照是整个算法的核心，在得到正确的光照估计后，第二步仅需做一个简单变换即可。



图 3-1 颜色恒常性算法步骤

结合公式 (2-10) 可以看出, 颜色恒常性算法就是在给定图像 $I^{(x,y)}$ 下估计场景光照 e , 并进一步恢复物体在标准白光下颜色 w 的过程, 这显然是一个病态问题, 在没有进一步的条件下, 很难估计出图像的准确光照。因此, 经典的图像恒常性算法都提出了一定的假设作为限制条件。下面简述几种经典的基于假设统计的图像恒常性算法。

3.2 基于假设统计的颜色恒常性算法

基于假设统计的颜色恒常性算法直接利用图像本身颜色信息来直接估计图像成像时的光照, 是一种无监督的方法, 由于这类算法不依赖样本的先验知识, 相比基于学习的有监督方法更为简单常用。以下公式的推导基于朗伯反射模型中公式 (2-5) (2-6)。

3.2.1 White-Patch 算法

该算法假设图像中各颜色通道的最大值是由场景中的白色表面导致的, 该表面能够反映场景中的光照颜色, 也被称为 max-rgb 算法。用数学公式表示如下:

$$\max_{x,y} f(x,y) = ke \quad (3-1)$$

其中 k 为常数, e 表示光照, (x,y) 表示像素位置, $f(x,y)$ 表示整幅图像,

$$\max_{x,y} f(x,y) = (\max_{x,y} R(x,y), \max_{x,y} G(x,y), \max_{x,y} B(x,y)) \quad (3-2)$$

White-Patch 算法简单, 计算复杂度低, 但当图像中存在强光或一定的噪声干扰时, 每个通道的最大值很可能不是白色表面, 导致算法的实际应用受到限制。

3.2.2 Grey-World 算法

Grey-World 算法假设场景中所有物体表面的平均反射是无色差的, 即图像中 RGB 三个通道对场景光照的平均反射率是相等的:

$$\frac{\int s(\lambda, x) dx}{\int dx} = k \quad (3-3)$$

其中 $s(\lambda, x)$ 表示波长为 λ 的光在 x 处的反射率, k 为常数, 表示无色差这一概念, 取值范围为 $[0,1]$, 取 0 表示场景中所有物体都没有发生任何反射, 取 1 表

明场景中所有表面发生完全反射。在这一假设下有：

$$\begin{aligned}
 \frac{\int f(x)dx}{\int dx} &= \frac{\int \int_{\omega} e(\lambda)s(\lambda, x)c(\lambda)d\lambda dx}{\int dx} \\
 &= \int_{\omega} e(\lambda)c(\lambda) \frac{\int s(\lambda, x)dx}{\int dx} d\lambda \\
 &= k \int_{\omega} e(\lambda)c(\lambda)d\lambda = ke
 \end{aligned} \tag{3-4}$$

Grey-World 算法将图像的“平均颜色”作为场景光照的颜色，其假设相比 White-Patch 不那么严格，因此适应性更好。

3.2.3 Grey-Edge 算法

与 Grey-World 算法和 White-Patch 算法都是在原始图像特征基础上提出的不同，Grey-Edge 假设场景中所有物体表面平均反射的差分都是无色差的[11]，可通过下式表示：

$$\frac{\int |s_x^{\delta}(\lambda, x)|dx}{\int dx} = k \tag{3-5}$$

式中的下标 x 表示在尺度 δ 上的导数。在这一假设下，光源的颜色就是图像中的平均颜色导数：

$$\begin{aligned}
 \frac{\int |f_x(x)|}{\int dx} &= \frac{1}{\int dx} \int \int_{\omega} e(\lambda) |s_x(\lambda, x)| c(\lambda) d\lambda dx \\
 &= k \int_{\omega} e(\lambda)c(\lambda)d\lambda = ke
 \end{aligned} \tag{3-6}$$

其中 $|f_x(x)| = (|R_x(x)|, |G_x(x)|, |B_x(x)|)^T$ 。

3.3 基于灰色像素的颜色恒常性光照估计

基于灰色像素的颜色恒常性算法也是基于假设统计学的，该算法认为在自然场景图像中广泛存在着灰色或者近似灰色的像素，这些灰色像素可以用来对场景光照作可靠的估计，这一假设在相关数据集上得到了验证^[13]。从而将颜色恒常性问题转换为了寻找图像中灰色像素的问题。

本论文需要对前景图像和背景图像进行光照估计，这一过程主要借鉴了基于灰色像素的颜色恒常性算法，下面对这一算法进行详细阐述。

3.3.1 灰色像素

这一算法的基本假设是，在白色光照下，绝大部分自然图像中包含着一些灰色或近似灰色的像素，需要特别指出的是这里的灰色不是指某一具体的 RGB 值，而是 $R=G=B$ 的一系列像素，一旦找到图像中的这些灰色像素，就可以高效地估计光照，为了说明这一过程，将朗伯反射模型中公式（2-5）进一步简化可以得到

$$I^i(x, y) = C^i(x, y) \cdot R^i(x, y), i \in \{r, g, b\} \quad (3-7)$$

其中 $I^i(x, y)$ 表示图像 I 在 (x, y) 处 i 通道的值， $C(x, y)$ 和 $R(x, y)$ 分别表示在 (x, y) 处的光照和物体表面光谱反射率。

假设场景的光照是统一的，也就是说光照与图像中的位置坐标无关，可以得到

$$I^i(x, y) = C^i \cdot R^i(x, y), i \in \{r, g, b\} \quad (3-8)$$

对于灰色像素，有

$$R^r(x, y) = R^g(x, y) = R^b(x, y) \quad (3-9)$$

因此，为了得到光照 $e = (e_r, e_g, e_b)$ ，只需对灰色像素分别求各通道的均值即可[12]：

$$e_i = \frac{1}{N} \sum_{(x, y) \in GP_n} I^i(x, y), i \in \{r, g, b\} \quad (3-10)$$

其中 GP_n 表示所取灰色像素的集合， N 为灰色像素的个数（当 $n=100\%$ 时，即为 Grey-World 算法）。

3.3.2 寻找灰色像素的算法框架

通过灰色像素可以将颜色恒常性的光照估计问题转化为寻找灰色像素问题，但在一张颜色偏差图像（color-biased image）中准确地定位灰色像素却又是一个挑战。为此，我们需要定义光照不变测度（illuminant-invariant measure, IIM）来鉴别灰色像素，如图像对数空间中局部区域（local patch）的标准差在场景中的该位置不随光照的变化而变化，可以作为 IIM。IIM 在 color-biased image 与 ground-truth image 中保持一致，对于灰色像素，不同颜色通道 IIM 也应保持一致，因此只需定义并计算不同通道 IIM 的“差别”，“差别”越小，该像素是灰色像素的可能性越大，如图 3-2 所示。

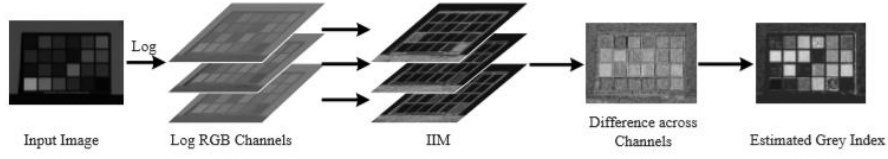


图 3-2 寻找灰色像素的算法流程

3.3.3 灰色指数 (Grey Index) ^[14]

以下算法基于 2.3.1 中的双色反射模型, 对公式 (2-7) 两边同时取对数, 可以得到

$$\log(I_i^{(x,y)}) = \log(F_i L_i R_{b,i}^{(x,y)}) + \log\left(\gamma_b^{(x,y)} + \gamma_s^{(x,y)} \frac{R_{s,i}^{(x,y)}}{R_{b,i}^{(x,y)}}\right), \quad i \in \{r, g, b\}. \quad (3-11)$$

去掉上标 (x, y) , 计算 R 通道与“亮度”(luminance) 在图像像素的对数空间中的差值, 并计算局部对比度 (拉斯高普高斯算子) $C\{\cdot\}$, 可以得到

$$\begin{aligned} C\{\log(I_R) - \log(|I|)\} &= C\{\log(F_R L_R) + \log(\gamma_b R_{b,R} + \gamma_s R_{s,R})\} \\ &\quad - C\{\log(F_R L_R (\gamma_b R_{b,R} + \gamma_s R_{s,R})) \\ &\quad + F_G L_G (\gamma_b R_{b,G} + \gamma_s R_{s,G}) + F_B L_B (\gamma_b R_{b,B} + \gamma_s R_{s,B})\} \end{aligned} \quad (3-12)$$

其中 $|I| = I_R + I_G + I_B$. 对于灰色像素, 有 $R_{j,R} = R_{j,G} = R_{j,B} = \bar{R}_j, j \in \{s, b\}$. 因此上式可以简化为

$$\begin{aligned} C\{\log(I_R) - \log(|I|)\} &= C\{\log(F_R L_R) + \log(\gamma_b \bar{R}_b + \gamma_s \bar{R}_s)\} \\ &\quad - C\{\log((F_R L_R + F_G L_G + F_B L_B)(\gamma_b \bar{R}_b + \gamma_s \bar{R}_s))\}, \end{aligned} \quad (3-13)$$

在小的邻域内, 镜头敏感度、接收到的光照被认为是一致的^[13], 于是

$C\{\log(F_R L_R)\} = 0, C\{\log((F_R L_R + F_G L_G + F_B L_B))\} = 0$, 因此

$$C\{\log(I_R) - \log(|I|)\} = C\left\{\log\left(\frac{\gamma_b \bar{R}_b + \gamma_s \bar{R}_s}{\gamma_b \bar{R}_b + \gamma_s \bar{R}_s}\right)\right\}^{gray} = 0. \quad (3-14)$$

公式 (3-14) 只是作为灰色像素的必要不充分条件, 为了更准确地识别灰色像素, 需要利用另一颜色通道的信息, 即

$$C\{\log(I_R) - \log(|I|)\} = C\{\log(I_B) - \log(|I|)\} = 0 \quad (3-15)$$

由此, 可以定义 $I(x, y)$ 处的灰色指数 (gray index) 如下

$$GI(x, y) = \| [C\{\log(I_R) - \log(I)\}, C\{\log(I_B) - \log(I)\}] \| \quad (3-16)$$

其中 $\|\cdot\|$ 表示 L2 norm。灰色指数越小，该像素是灰色像素的可能性越大。但从公式 (3-16) 可以看出，在一个小的 patch 内，如果所有像素颜色完全统一，即使不是灰色像素，灰色指数也会趋于 0。因此，对这一情况加上限制条件

$$C\{I_i\} > \varepsilon, \forall i \in \{R, G, B\} \quad (3-17)$$

其中 ε 是一个小的对比度阈值，一般取 0.0001。

3.3.4 多光源的情况

在前景或背景图像明显来自多于一个光源的情况下，仍然强制将其视作单光源，求取全局一致的光照是不合适的，因此我们试图得到整个场景光照分布的热图，即每一个像素的光照。

在基于灰色像素的颜色恒常性算法中，IIM 是定义在 local patch 像素的空间位置关系上的，在 local patch 上，光照被认为是统一的，因此这一算法同样可以应用在多光源图像光照估计上 [13]。

首先，利用公式 (3-16) 检测图像中的灰色像素，接着基于像素的空间位置用 K-means 聚类算法将灰色像素分为 M 组，对每组灰色像素利用公式 (3-10) 求取该组所在区域的光照 $(\xi_k^i, i \in \{R, G, B\}, k = 1, 2, \dots, M)$ ，同时可以得到该组灰色像素的中心点 (x_k, y_k) ，那么对图像中任意像素 (x, y) ，光照的计算公式如下：

$$e_i(x, y) = \sum_{k=1}^M \omega_k \xi_k^i, i \in \{R, G, B\} \quad (3-18)$$

$$\omega_k = e^{-\frac{D_k}{2\sigma^2}} / \sum_{m=1}^M e^{-\frac{D_m}{2\sigma^2}}, k = 1, 2, \dots, M \quad (3-19)$$

其中 $D_k = \|(x, y) - (x_k, y_k)\| / \sqrt{W^2 + H^2}$ 表示像素 (x, y) 与第 k 组灰色像素中心 (x_k, y_k) 的归一化空间距离， σ 控制权重的敏感度，一般取 $\sigma=0.2$ 。

3.4 基于光照估计的图像合成

上一节主要说明了如何应用颜色恒常性算法计算图片的光照，本节将会说明

如何应用图片的光照作不同光源图像的合成。在作光照估计时，主要做了以下假设：

假设一 场景中的光照是统一的。

假设二 自然场景图像中存在大量的灰色像素。

假设三 所有物体表面都是理想的漫反射表面。

需要注意的是，颜色恒常性算法估计出的只是光照的色度，即只关注光照 RGB 分量间的比例，而不关注光照的亮度因素，但在作图像合成时，亮度会极大地影响合成效果，另外，由于人对图像的亮度非常敏感，如果只改变目标前景的色度，而不匹配前背景的亮度，会严重影响合成图像的真实感，为了解决这一问题，引入假设四：

假设四 灰色像素的平均亮度与图像场景的光照亮度趋于一致。

这一假设在直观上是合理的，在实验中证实对大部分图像也是成立的。基于这一假设，公式（3-10）计算得到的灰色像素的平均绝对大小就反映了光照的亮度，各通道相对比例反映了光照的色度。虽然从结果上来说，仍旧使用公式（3-10）计算图像的光照，但通过这一假设弥补了理论上的缺陷。

Von Kries 对角理论指出，两种光照条件下的同一物体表面颜色之间的关系可用一个对角矩阵变换来描述^[15]，在不同光照 e_1, e_2 下物体 RGB 值可以作如下转换：

$$\begin{bmatrix} R_2 \\ G_2 \\ B_2 \end{bmatrix} = \begin{bmatrix} k_r & 0 & 0 \\ 0 & k_g & 0 \\ 0 & 0 & k_b \end{bmatrix} \cdot \begin{bmatrix} R_1 \\ G_1 \\ B_1 \end{bmatrix} \quad (3-20)$$

其中 k_r, k_g, k_b 分别 RGB 三个通道的校正系数。

此模型是光照变换的一个近似，也是本文研究在不同光照下图像合成合理性的理论支撑，通过估计出前景及背景光照，再利用 Von Kries 对角理论将前景图像映射到背景光照下。结合朗伯反射定律，目标前景图像关于光照的转换公式为：

$$I_2^i(x, y) = I_1^i(x, y) \cdot (e_b^i / e_f^i), i \in \{R, G, B\} \quad (3-21)$$

其中 e_b^i, e_f^i 分别为前景和背景光照， I_1^i, I_2^i 分别为改变光照前后的图像。

可以预见到，假设四在部分亮度分布差异较大的图片上可能会失效，作为这

一可能情况的补充, 本论文提出了基于 HSL 颜色空间的前景亮度修正算法, 弥补因为光照颜色亮度估计错误而导致合成图像效果较差的问题, 首先需要将图像从 RGB 颜色空间转换为 HSL 颜色空间, 根据公式(2-3)得到每个像素点的亮度 $L(x, y)$ 。接着根据下式计算图像的“平均亮度”^[16]:

$$L_{avg} = \exp \left(\frac{1}{N} \sum_{x,y} \ln(\delta + L(x, y)) \right) \quad (3-22)$$

其中 δ 是一个较小的常数, 避免计算黑色像素时出现错误, δ 一般可取 0.0001。

分别计算前景源图像的平均亮度 L_f , 背景源图像的平均亮度 L_b , 根据两者的比值对图像亮度作线性调整, 同时将上述由颜色恒常性算法估计到的光照作归一化, 仅用以调整前景的色度, 因此待合成图像前景的转换公式如下

$$I_2^i(x, y) = I_1^i(x, y) \cdot (L_b^i / L_f^i) \cdot (e_{normalized,b}^i / e_{normalized,f}^i), i \in \{R, G, B\} \quad (3-23)$$

其中 $e_{normalized}^i$ 表示归一化的光照颜色分量, 计算公式如下

$$e_{normalized}^i = \frac{e^i}{e^R + e^G + e^B} \quad (3-24)$$

另外一种思路式根据 RGB 颜色空间向 HSL 颜色空间转换的公式求取到前景源和背景源图像中每个像素的亮度, 再采用泊松克隆的方式确定前景在新的背景中具体各像素的亮度。

本算法合成图像的一个例子如图 3-3 所示:

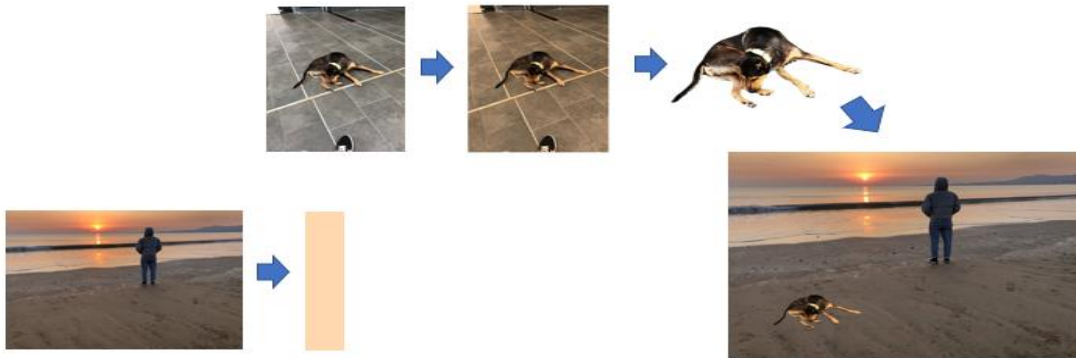


图 3-3 光照一致图像合成实例

基于光照一致的图像合成的整体流程如图 3-4 所示:

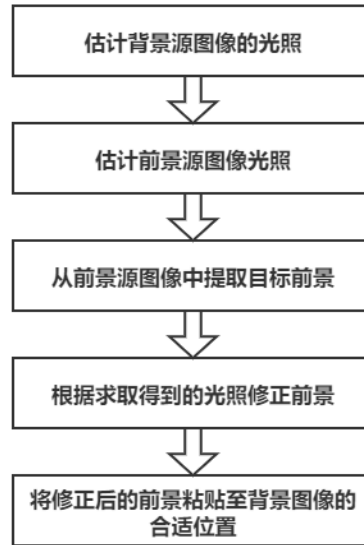


图 3-4 光照一致的图像合成算法流程

第4章 实验与分析

4.1 设计实验

4.1.1 数据集

图像和谐化数据集一般将实际拍摄的图像作为 ground-truth, 即合成的目标, 将其前景分割出来并作一定调整, 作为待处理的图像, 但这样的数据集前景图像和背景图像都不完整, 缺失了用于估计图像光照的部分信息, 并不适用于本论文的基于颜色恒常性的图像合成算法。

本实验所采用的数据集共 73 组图片, 包括 5 个部分: 背景源图像 bg, 前景源图像 fg, 直接粘贴目标前景至背景源的结果图像 com (一般称之为 cut-and-paste), 目标前景在 com 上的掩模 mask, 标准合成图像 gt (实验中假设 gt 是最优的合成结果), 如图 4-1 所示。

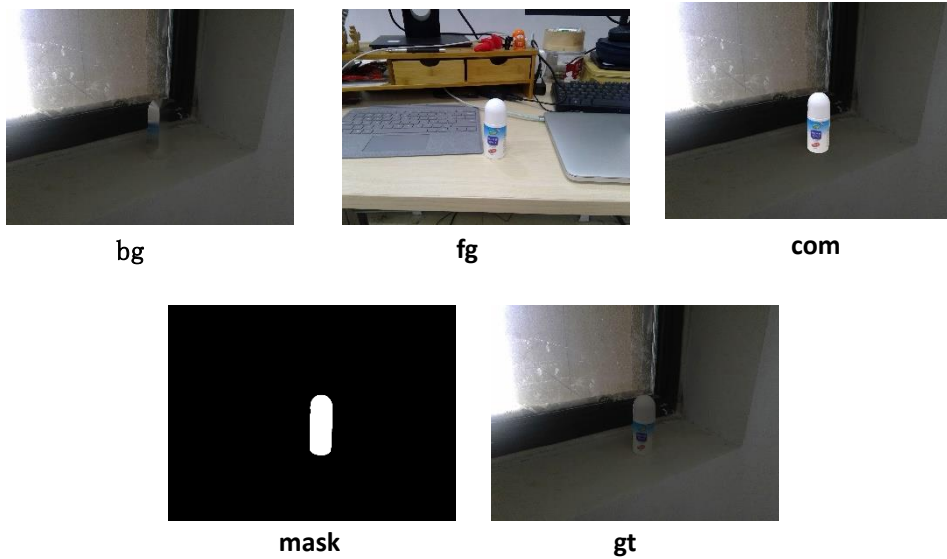


图 4-1 测试数据集

4.1.2 评价指标

为了定量地评价图像的合成效果, 需要定义合成图像与标准图像间的误差, 这里采用平均绝对误差 MAE:

$$MAE = \frac{1}{N} \sum_{i \in \{R, G, B\}} \sum_{x, y} |I_1^i(x, y) - I_2^i(x, y)| \quad (4-1)$$

其中 I_1, I_2 是待比较的两幅图像, N 为图像的像素数。

图像合成的结果很难达到 $MAE = 0$, 为了更直观地评价图像修正的程度, 同时计算 com 与 gt 的平均绝对误差 MAE_{com} , 修正图像(corrected image)与 gt 的平均绝对误差 MAE_{cor} , 将前者相对于后者的改动程度作为评价图像合成效果的指标 R:

$$R = \frac{MAE_{com} - MAE_{cor}}{MAE_{com}} \times 100\% \quad (4-2)$$

R 的取值范围为 $(-\infty, 1]$, 当 $0 < R \leq 1$ 时, 表明算法改进了图像的合成效果, R 越接近 1 表示修正图像越接近 gt 图像, 合成效果越好; 当 $R = 0$ 时, 表明修正图像相比 com 没有改进; 当 $R < 0$ 时, 说明表明修正图像相比 com 合成效果更差。

4.1.3 实验流程

实验执行流程如图 4-2 所示, 具体如下:

- 1) 读入前景源图像 fg, 背景源图像 bg, 待合成图像 com, 掩模 mask, 标准合成图像 gt, 并转换为范围 $[0, 1]$ 的 double 类型;
- 2) 根据公式 (3-16) (3-17) 计算图像中每个像素的 Grey Index, 取 GI 最小的 n% 的像素作为灰色像素, 接着由公式 (3-10) 分别计算 fg 的光照 e_f , bg 的光照 e_b ;
- 3) 根据上一步得到的 e_f, e_b 由公式 (3-21) 修正 com 的 mask 区域 (即前景), 得到修正图像 (cor);
- 4) 计算 cor 与 gt 的平均绝对误差 MAE_1 ;
- 5) 计算 com 与 gt 的平均绝对误差 MAE_0 ;
- 6) 根据 MAE_0 和 MAE_1 计算图像合成效果评价指标 R。

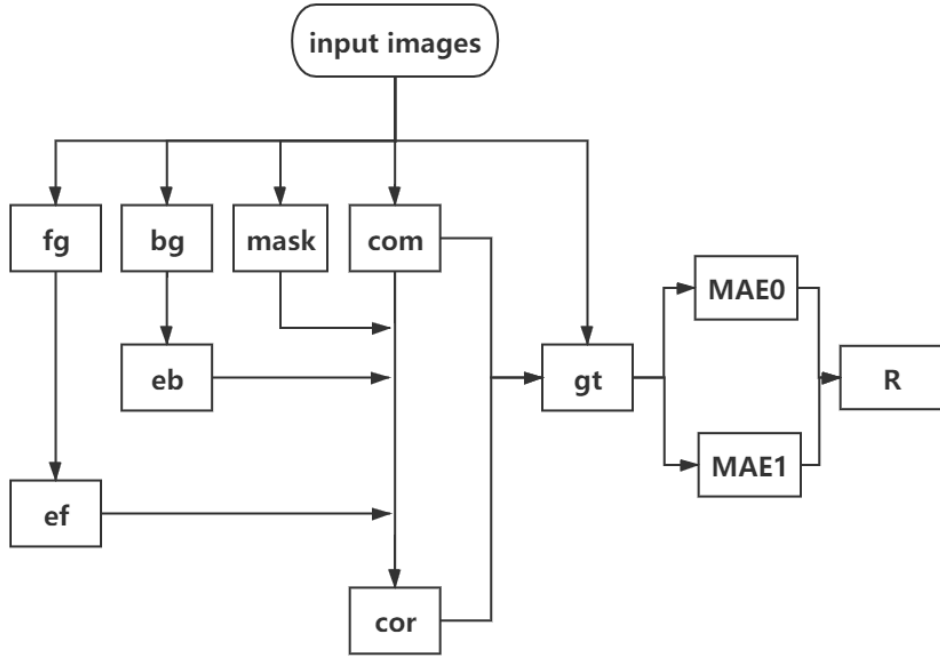


图 4-2 实验执行流程

4.2 结果分析

取参数 $n=10, \varepsilon=0.0001$ ，数据集 73 组图片中 58 组的合成结果与 cut-and-paste 对比都取得了更好的合成效果，平均绝对误差数据如表 1 所示，

表 1 平均绝对误差比较

MAE ($\times 10^{-3}$)	Mean	Median	Mean of Best 25%	Mean of Worst 25%
Cut-and-paste	25.8	18.8	6.2	53.1
Ours	13.8	9.3	2.2	31.9

表 1 表达了本论文图像合成算法与 cut-and-paste 合成图像的平均绝对误差分布差异，本论文算法整体上使得合成图像的 MAE 值大幅降低，说明了本文算法是有一定的改进效果的。

除此之外，我们也需要了解平均每张图片合成效果相比 cut-and-paste 增强了多少，根据公式 (4-2) 进一步计算每幅合成图像的 R 值，如表 2 所示。

表 2 合成图像的 R 值分布

R	Mean	Median	Mean of Best 25%	Mean of Worst 25%
Ours	34.32%	44.92%	85.53%	-29.64%

表 2 更关注单张图像的合成效果，可以看出表中 4 个统计值差异很大，特别是合成结果最好与最差的 25% 的图片（19 张）的 R 值，Mean of Worst 25% 的 R 值是负的，即合成结果相比 cut-and-paste 更差。这是由于本论文图像合成算法是基于场景光照估计的，受不同图像拍摄场景影响较大，在同一场景（前景源、背景源）下，即使前景图像不同、位置不同，也会有相近的 R 值。

通过对实验结果加以分析，可以发现合成效果较好的图像往往前景光照与背景光照差别较大（部分实验结果如图 4-3 所示，从左至右分别为 com, mask, 背景光照，前景光照，合成结果，gt）。然而数据集中多数图片都是处于不同明暗的自然白光下，前景源与背景源光照差别并不明显，因此算法对于部分图像并没有完全发挥作用。

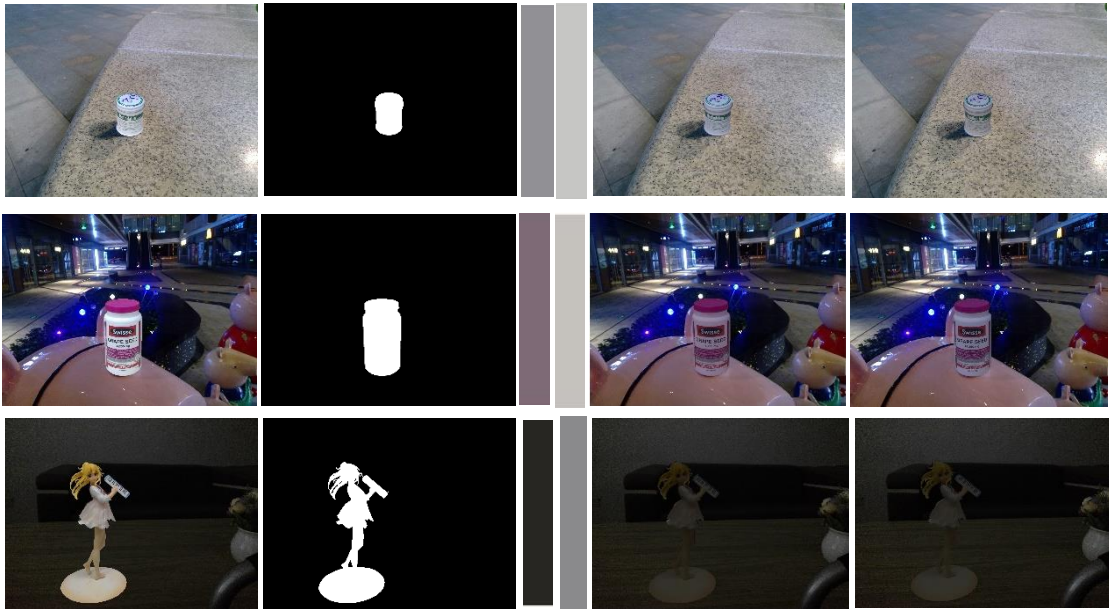


图 4-3 实验结果：前景与背景光照差别较明显

分析部分图像合成效果较差 ($R < 0$) 的原因主要包括以下两个方面：

1. MAE_0 较小

本论文图像合成算法的核心是图像拍摄场景的光照估计，再根据估计得到的前背景光照修正目标前景图像，整个过程是基于 3.4 中四个假设的，因此算法本身并不足够精确，当 cut-and-paste 的结果与标准合成图像 gt 的差别较小时，可能使得合成效果反而变得更差。另一方面，由于 MAE_0 较小， MAE_1 相对于 MAE_0 的一点改变容易被放大。如图 4-4， $MAE_0 = 0.0022$, $MAE_1 = 0.0031$, $R = -43.83\%$ 。

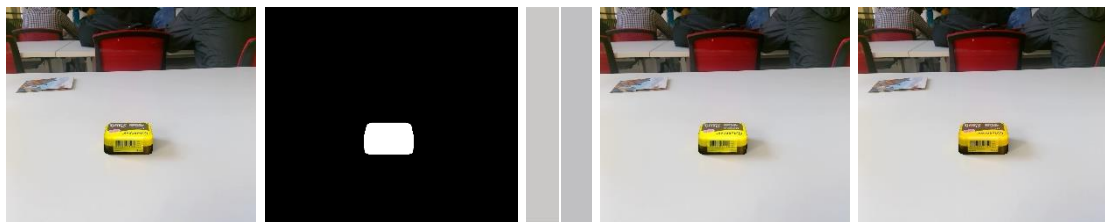


图 4-4 实验结果： MAE_0 较小

$MAE_0 - R$ 散点图如图 4-5 所示，合成效果较差 ($R < 0$) 的点绝大部分集中在左下角 $MAE_0 < 0.02$ 的区域，也说明了这一问题。

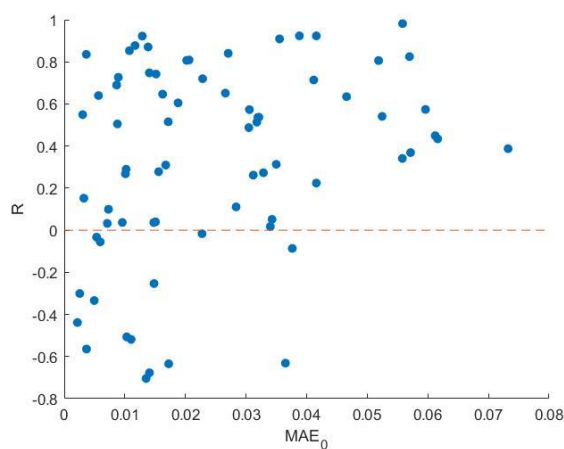


图 4-5 $MAE_0 - R$ 散点图

2. 光照估计错误

在一些图片中，灰色像素占比较低，再经过公式 (3-17) 等限制条件排除后，剩下的灰色像素更少，此时强制取固定比例像素会造成取到的灰色像素很可能并不是灰色的，虽然利用前背景的源图像估计光照一定程度上改善了这一问题，但某些情况下，这仍会导致光照色度估计错误，进而导致图像合成效果较差，如图 4-6 所示。

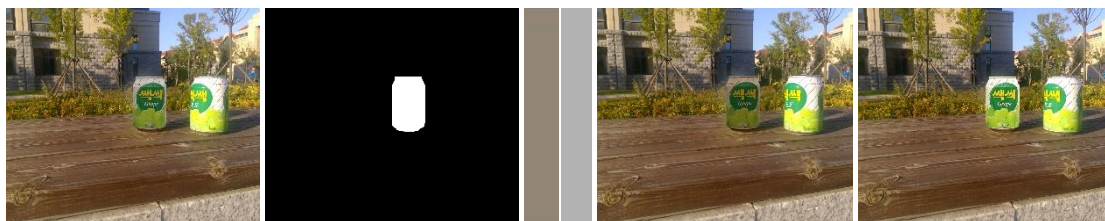


图 4-6 实验结果：光照色度估计错误

除此之外，假设四在某些图像场景下会失效，造成光照亮度错误，进而导致

图像合成效果差，如图 4-7 所示。光照的亮度估计出错，这与参数（主要是取灰色像素比例 n ）有关，对于不同场景图像，这一参数的最优值不同，即合成效果关于这一参数较为敏感。

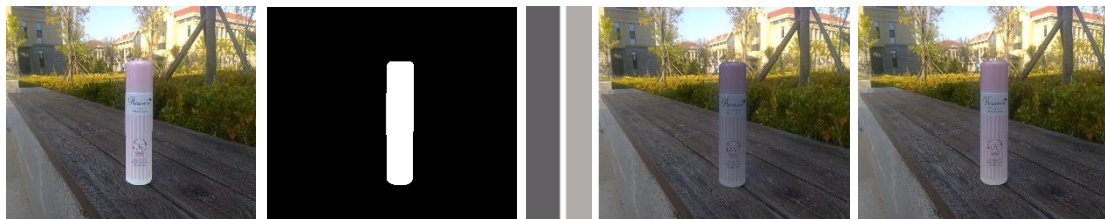


图 4-7 实验结果：光照亮度估计错误

第5章 结束语

5.1 本文总结

本论文基于颜色恒常性提出了解决图像合成时前景与背景光源不一致问题的算法,通过相关理论说明了这一方法的合理性,通过实验验证了方法的可行性,在数据集上定量测试了这一方法的图像合成效果,最后分析了这一方法的合成结果和局限性。新的图像合成算法使得调整后的目标前景的光照更接近背景图像,缓解了直接拷贝目标前景到背景源图像中的不真实感。

5.2 工作展望

4.2 中提到了图像合成效果较差的两个原因,这两方面都源于算法本身的局限性,其一是对于精度要求稍高的图像合成,可以考虑计算图像多光源下的光照分布热图,用以更准确地改变前景光照,3.3.4 就是在这一方向上的尝试。其二是光照估计错误,可以考虑通过单独计算光照的亮度,即公式(3-22)(3-23)(3-24)来改善这一问题。这一图像合成算法还有很多局限性,需要进一步深入研究。

致 谢

在完成本论文的过程中，得到了钟凡老师和宋双兵学姐的许多指导，实验所使用的数据集也是由宋双兵学姐提供的，在此向他们致谢。

除此之外，在即将毕业之际，我要感谢我的母校，在山大的四年时光里，我成长了很多，留下了许多珍贵的回忆。希望山大越来越好，成为真正的世一大。

我要感谢给我上过课的每一位老师，他们中的很多老师不仅教会了我专业知识，还教给了我许多比知识本身更重要的东西。

我要感谢在山大遇到的每一位同学，是优秀的他们让我在懒惰时没有一蹶不振，催促我努力奋进，是他们让我明白每个人都可以走出自己的可能性。

最后，我要感谢我的父母，他们从不曾对我有任何约束，却永远默默地支持我。

参考文献

- [1] Perez P, Gangnet M, Blake A. Poisson image editing[C]//Proceedings of ACM SIGGRAPH 2003, New York, USA: ACM, 2003: 313-318.
- [2] Perez P , Gangnet M , Blake A . Poisson Image Editing[J]. ACM Transactions on Graphics, 2003, 22(3):p.313-318.
- [3] Xue, Su, et al. "Understanding and improving the realism of image composites." ACM Transactions on graphics (TOG) 31.4 (2012): 1-10.
- [4] Tsai, Yi-Hsuan, et al. "Deep image harmonization." *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2017.
- [5] Cong, Wenyan, et al. "Deep Image Harmonization via Domain Verification." *arXiv preprint arXiv:1911.13239* (2019).
- [6] Garon, Mathieu, et al. "Fast Spatially-Varying Indoor Lighting Estimation." *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2019.
- [7] 李亨. 颜色技术原理及其应用[M]. 北京: 科学出版社, 1994.
- [8] 张雅媛. 基于光照一致的图像合成方法研究[D]. 2017.
- [9] J. H. Lambert, *Pilottometria sive de mensura de gratibus luminis, colorum et umbrae*(Eberhard Klett, Augsburg, Germany, 1760).
- [10] Gijsenij A, Gevers T, Van De Weijer J. Computational color constancy: Survey and experiments[J]. IEEE Transactions on Image Processing, 2011, 20(9): 2475-2489.
- [11] Shafer, Steven A. "Using color to separate reflection components." *Color Research & Application* 10.4 (1985): 210-218.
- [12] Weijer J V D , Gevers T . Color Constancy based on the Grey-Edge Hypothesis[C]// IEEE International Conference on Image

Processing. IEEE, 2005.

- [13] Yang K F, Gao S B, Li Y J. Efficient illuminant estimation for color constancy using grey pixels[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2015: 2254-2263.
- [14] Qian Y, Kamarainen J K, Nikkanen J, et al. On finding gray pixels[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2019: 8062-8070.
- [15] 朱贵冬, 沈理, 王今觉. 基于 von-Kries 色适应的分区颜色校正方法[J]. 计算机工程与科学, 2007, 29(2):50-52.
- [16] Reinhard E, Stark M, Shirley P, et al. Photographic tone reproduction for digital images[C]//Proceedings of the 29th annual conference on Computer graphics and interactive techniques. 2002: 267-276.

附录 1 代码及相关附件

```

clc, clear, close all

% 待处理图片的路径、名称及图片数量
path = "C:\Users\19843\Desktop\毕设\dataset\fromSong";
File = dir(fullfile(path, '\gt\','*.png'));
imgNames = strrep(string({File.name}'),'.png','');
numImg = size(imgNames,1);

% 测试结果存储
outFilePath = "C:\Users\19843\Desktop\毕设\test results on
dataset\5.14_Yanlin Qian_addfg&bg_Song_1_4";
R = zeros(numImg,3);
IlluminantB = zeros(numImg,3);
IlluminantF = zeros(numImg,3);

% 设置参数
param.Npre = 10.^ (1);% 光照对这一参数极为敏感
param.delta_threshold=10.^ (-4);%[-5,-4,-3];

for k = 1:numImg
    name = imgNames(k);
    close;

    % 读取图片
    com = im2double(imread(path+'\com\' +name+'.png')); % 此时读取进去
    的像素值范围变为 double 类型 0~1, 有更多计算空间
    gt = im2double(imread(path+'\gt\' +name+'.png'));
    fg = im2double(imread(path+'\fg\' +name+'.png'));
    bg = im2double(imread(path+'\bg\' +name+'.png'));
    mask = imread(path+'\mask\' +name+'.png'); % 0 or 255
    mask = im2bw(mask);% 将 mask 转换为二值图

    % 前背景光照, 由于图像中的像素值为 0~1, 对检测到的灰色像素取均值
    后, 光照 B, F 各分量的范围也为 0~1
    B = illuminant_estimation(bg, param);
    F = illuminant_estimation(fg, param);

```

```

% 改变前景的光照
outimg = correctForeground(com, mask, B, F);

% 展示合成效果
figure('name', "composite"+name);
subplot(2,3,1); imshow(com./max(com(:)), []); title('toCom')
subplot(2,3,2); imshow(outimg); title('corrected image')
subplot(2,3,3); imshow(gt./max(gt(:)), []); title('gt')
subplot(2,3,4); imshow(mask); title('mask')

showB = zeros(size(com));
showB(:, :, 1) = B(1);
showB(:, :, 2) = B(2);
showB(:, :, 3) = B(3);
subplot(2,3,5); imshow(showB); title('Illuminant B')

showF = zeros(size(com));
showF(:, :, 1) = F(1);
showF(:, :, 2) = F(2);
showF(:, :, 3) = F(3);
subplot(2,3,6); imshow(showF); title('Illuminant F')

% 计算误差
Err0 = getError(com, gt, 'MAE');
Err1 = getError(outimg, gt, 'MAE');
reducedBy = (Err0-Err1)/Err0;
fprintf(name + ": MAE0 = %f\t\t MAE1 = %f\t\t reduced
by %f\n", Err0, Err1, reducedBy);

% 存储实验结果
R(k, 1) = Err0;
R(k, 2) = Err1;
R(k, 3) = reducedBy;

IlluminantB(k, :) = B;
IlluminantF(k, :) = F;

set(gcf, 'visible', 'off');
print(gcf, '-djpeg', outFilePath+'\' +imgNames(k))

break

```

end

```
fprintf('average_MAE0 = %f\t average_MAE1 = %f\t average reduced
by %f', sum(R(:,1))/numImg, sum(R(:,2))/numImg, sum(R(:,3))/numImg);
fprintf('\tpositive effect: %d ', sum(R(:,3)>0));
save(outFilePath+'R.mat', 'R');
save(outFilePath+'Ib.mat', 'IlluminantB');
save(outFilePath+'If.mat', 'IlluminantF');
```

```
function Evalum = illuminant_estimation(input_im, param)

%在参数 param 条件下, 输出 input_im-mask 的光照
Npre=param.Npre;
delta_threshold=param.delta_threshold;

%choose Npixels.
Npixels = size(input_im,1)*size(input_im,2);
numGPs = floor(Npre*Npixels/100);

% mask saturated pixels and mask very dark pixels
mask=(max(input_im,[],3)>=0.95) | (sum(input_im,3)<=0.0315);

img_column=reshape(input_im,[],3);
r=input_im(:, :, 1); g=input_im(:, :, 2); b=input_im(:, :, 3);

%denoise
hh = fspecial('average', [7 7]);
r = imfilter(r, hh, 'circular'); g = imfilter(g, hh, 'circular'); b =
imfilter(b, hh, 'circular');

%mask 0 elements
mask=mask | (r==0) | (g==0) | (b==0);
r(r==0)=eps; g(g==0)=eps; b(b==0)=eps; norml=r+g+b;

%mask low contrast pixels
delta_r=DerivGauss(r,.5); delta_g=DerivGauss(g,.5);
delta_b=DerivGauss(b,.5);
mask=mask | delta_r<=delta_threshold & delta_g<=delta_threshold &
delta_b<=delta_threshold;

log_r=log(r)-log(norml); log_b=log(b)-log(norml);
```

```

delta_log_r=DerivGauss(log_r,.5);
delta_log_b=DerivGauss(log_b,.5);
mask=mask | (delta_log_r==Inf) | (delta_log_b==Inf);

% valuedNumMask = size(find(mask==0),1)

data=[(delta_log_r(:)), (delta_log_b(:))];%, Mr(:)-
norml_M(:), Mg(:)-norml_M(:), Mb(:)-norml_M(:)];
mink_norm=2;
norm2_data=power(sum(power(data,mink_norm),2),1/mink_norm);
map_uniquelight=reshape(norm2_data,size(delta_log_r));

%(1)mask out some pixels
map_uniquelight(mask==1)=max(map_uniquelight(:));

%(2)denoise
hh = fspecial('average',[7 7]);
map_uniquelight = imfilter(map_uniquelight,hh,'circular');

%filter by map_uniquelight
Greyidx_unique=map_uniquelight;
% figure(2);imshow(Greyidx_unique)
sort_unique=sort(Greyidx_unique(:));
Gidx_unique = zeros(size(Greyidx_unique));

Gidx_unique(Greyidx_unique<=sort_unique(floor(numGPs))) = 1;
choosen_pixels=img_column(Gidx_unique==1,:);

Evalum=mean((choosen_pixels),1);

end

```

```

function outimg = correctForeground(com, mask, B, F)
% 对 com 的 mask 区域作修正

c = B./F;
outimg = zeros(size(com));
for i=1:3
    m = ones(size(mask));
    m(mask==1) = c(i);
    outimg(:, :, i) = com(:, :, i).* m;
end

```

```
end  
end
```

```
function error = getError(img1, img2, errortype)  
% img1, img2 --- 待比较图片  
% errortype --- 误差类型，包括均方误差（MSE）、平均绝对误差（MAE）  
  
    d = img1 - img2;  
  
    if errortype == 'MSE'  
        s = d.*d;  
    end  
  
    if errortype == 'MAE'  
        s = abs(d);  
    end  
  
    error = sum(s(:))/(size(img1,1)*size(img1,2));  
end
```

附录 2 文献英文原文

Efficient Illuminant Estimation for Color

Constancy Using Grey Pixels

Kai-Fu Yang Shao-Bing Gao Yong-Jie Li

University of Electronic Science and Technology of China, Chengdu,
China

Illuminant estimation is a key step for computational color constancy. Instead of using the grey world or grey edge assumptions, we propose in this paper a novel method for illuminant estimation by using the information of grey pixels detected in a given color-biased image. The underlying hypothesis is that most of the natural images include some detectable pixels that are at least approximately grey, which can be reliably utilized for illuminant estimation. We first validate our assumption through comprehensive statistical evaluation on diverse collection of datasets and then put forward a novel grey pixel detection method based on the illuminant-invariant measure (IIM) in three logarithmic color channels. Then the light source color of a scene can be easily estimated from the detected grey pixels. Experimental results on four benchmark datasets (three recorded under single illuminant and one under multiple illuminants) show that the proposed method outperforms most of the state-of-the-art color constancy approaches with the inherent merit of low computational cost.

We argue that most of the natural images contains some intrinsic grey pixels (or at least very approximate to grey) when under a white light source. Once these grey pixels are efficiently retrieved, they can be used for efficient light source color estimation. However, the challenging problem is how to find the grey pixels efficiently from RGB values of original color-biased images since the changing external illuminant usually alters their color appearance.

Studies on color-based recognition seek for color descriptors independent of the color, intensity, and geometry of scene light source. Note that these descriptors may be roughly attributed as one of a certain category of color constancy algorithms, but they

are not aimed at recovering the real color appearance of scenes. Motivated by this, we consider to define a certain illuminant-invariant measure (IIM) to identify the grey pixels in images.

If we define the IIM based on the spatial relationship of pixels within local image patches, they can be used in the both situations of uniform and non-uniform illuminant. For example, the standard deviation of local patches calculated in logarithmic space can be used as an IIM (see more details later). However, the true grey pixels being isolated or within a uniform region will have little chance to be detected, since no reliable local difference can be used for IIM computation. Therefore, in order to accurately find the grey pixels, we need to further define the detectable grey pixels by adding following constraints on IIM.

(a) The grey pixels should be located in small grey patches with the minimum size of 3×3 pixels.

(b) These small grey patches are not uniform, i.e., the contrasts of grey patches are non-zero.

Based on the constraint (a), we can compute IIM in each color channel of the color-biased images. Because IIM is required to be independent of illuminant, the grey pixels can be detected by searching for the points with equal IIM across three color channels. Meanwhile, constraint (b) excludes the points with zero IIM.

Once we find these grey pixels, we can easily extract the illuminant of a color-biased image from them. However, it is worthy to note that Equation 4 is just a necessary condition for pixel (x, y) being grey. This means that some pixels meeting Equation 4 may be colored. In this work, in order to obtain more robustly the grey pixels, we further exclude the pixels with quite low luminance in original space (RGB space) and relatively isolated in spatial location.

Hypothesis Validation : In order to verify whether most of the natural images contain more or less grey pixels, we evaluate the possibility of detectable grey pixels in ground-truth images, which are produced by transforming the raw color-biased images into the canonical images under white light source based on the ground-truth illuminant provided in the benchmark datasets. In addition, we also evaluate our hypothesis on

two irrelevant natural datasets collected for other applications, e.g., saliency region detection. These datasets include thousands of images captured by different cameras in various natural scenes, presuming that these images have been well white-balanced by cameras themselves.

To measure how close a pixel approximates to grey, we define a Grey Index (GI) for each pixel, which is computed as follows. For each pixel (x, y) , we first compute the local mean and local contrast (defined by standard deviation in this paper) within a local window with the size of $\eta \times \eta$

pixels in each channel. Then, we denote the locally averaged maps and local contrast maps. Then GI is computed as

$$GI(x, y) = \sqrt{\frac{1}{3} \sum_{i \in \{r, g, b\}} \frac{(GI_{\eta}^i(x, y) - GT_{\eta}(x, y))^2}{GT_{\eta}(x, y)}}$$

It is obvious that a pixel with lower GI has higher probability of being grey.

The detectable grey points should have very low GI, while CT_{η} , CTg_{η} , and CTb_{η} are non-zero. Figure 1 shows the percentage of images containing detectable grey pixels on three benchmark datasets and two irrelevant natural datasets. Note that the gray ball in SFU Gray Ball dataset and the color board in Color-Checker dataset were masked out in this experiments. This figure shows that more than 95% of images in each dataset contain detectable grey pixels with GI lower than 0.02 and almost all images contain detectable grey pixels with GI lower than 0.1. This strongly supports the proposed hypothesis that most natural scenes under white light source indeed contain detectable pixels that are approximately grey.

The next question is whether these grey pixels with low GIs could be utilized as the accurate indicators for illuminant estimation. To answer it, we first estimated the illuminant by summing the detected pixels in a color-biased image, which have low GI values computed from its corresponding ground-truth image. Then, the angular error is computed to measure the error between the estimated and the ground-truth illuminants. Figure 2 shows the angular errors evaluated on three benchmark datasets when using the detected grey pixels for illuminant estimation. It is obvious that the median and mean angular errors on the three datasets are smaller than or around 1.0 at various GI

levels, which are significantly better than all existing color constancy models as far as our knowledge. This indicates that perfect illuminant estimation can be obtained based on the grey pixels of a color-biased image if these grey pixels could be correctly detected out.

So far, the above analysis has grounded our hypothesis that most natural images always include some detectable grey (or approximately grey) pixels, which can be reliably utilized for accurate illuminant estimation.

附录 3 文献中文译文

使用灰色像素的颜色恒常性高效光源估计

光源估计是计算颜色恒定性的关键步骤。代替使用灰色世界或灰色边缘假设，我们在本文中提出了一种通过使用在给定的彩色图像中检测到的灰色像素信息来进行光源估计的新方法。基本假设是，大多数自然图像都包含一些至少近似为灰色的可检测像素，这些像素可以可靠地用于光源估计。我们首先通过对各种数据集进行综合统计评估来验证我们的假设，然后提出了一种基于光源不变量度（IIM）在三个对数颜色通道中的新型灰度像素检测方法。然后，可以根据检测到的灰色像素轻松估算场景的光源颜色。对四个基准数据集（三个记录在单个光源下，一个记录在多个光源下）的实验结果表明，所提出的方法优于大多数最新的色彩恒定性方法，其固有优点是计算成本低。

我们认为，在白光源下时，大多数自然图像都包含一些固有的灰色像素（或至少非常接近灰色）。一旦有效检索了这些灰色像素，即可将其用于有效的光源颜色估计。然而，具有挑战性的问题是如何从原始的色彩偏移图像的 RGB 值中有效地找到灰色像素，因为变化的外部光源通常会改变其颜色外观。

基于颜色的识别研究寻求独立于场景光源的颜色，强度和几何形状的颜色描述符。请注意，这些描述符可以粗略地归为某种颜色恒定性算法中的一种，但它们的目的是恢复场景的真实颜色外观。因此，我们考虑定义某种光源不变量度（IIM）以识别图像中的灰色像素。

如果我们根据局部图像块内像素的空间关系来定义 IIM，则它们可用于均匀和非均匀光源的两种情况。例如，在对数空间中计算的局部面片的标准偏差可以用作 IIM（请参阅后面的详细信息）。但是，由于没有可靠的局部差异可用于 IIM 计算，因此孤立的真实灰色像素或在均匀区域内的真实灰色像素将很少被检测到。因此，为了准确地找到灰色像素，我们需要通过在 IIM 上添加以下约束来进一步定义可检测的灰色像素。

(a) 灰色像素应位于最小 3×3 像素的灰色小块中。

(b) 这些小的灰色斑块不均匀，即，灰色斑块的对比度不为零。

基于约束 (a)，我们可以计算出色彩偏置图像的每个色彩通道中的 IIM。由于要

求 IIM 独立于光源，因此可以通过在三个颜色通道中搜索具有相等 IIM 的点来检测灰色像素。同时，约束 (b) 排除了 IIM 为零的点。

一旦找到这些灰色像素，就可以轻松地从中提取偏色图像的光源。但是，值得注意的是，公式 4 只是像素 (x, y) 为灰色的必要条件。这意味着某些满足公式 4 的像素可能会被着色。在这项工作中，为了更稳健地获得灰色像素，我们进一步排除了原始空间 (RGB 空间) 中亮度非常低且空间位置相对孤立的像素。

假设验证：为了验证大多数自然图像是否包含或多或少的灰色像素，我们评估了地面真实图像中可检测到的灰色像素的可能性，这些事实是通过将原始的偏色图像转换为标准图像而产生的。基于基准数据集中提供的地面光源的白光源。此外，我们还针对为其他应用（例如，显著区域检测）收集的两个不相关的自然数据集评估了我们的假设。这些数据集包括在不同自然场景下由不同相机捕获的数千张图像，假定这些图像已被相机本身很好地进行了白平衡。

为了测量像素接近灰色的程度，我们为每个像素定义了灰度指数 (GI)，其计算方法如下。对于每个像素 (x, y) ，我们首先在大小为 $\text{and} \times \eta$ 的局部窗口内计算局部均值和局部对比度（由本文中的标准差定义）、每个通道中的像素。然后，我们表示局部平均图和局部对比图。然后 GI 计算为

$$GI(x, y) = \sqrt{\frac{1}{3} \sum_{i \in \{r, g, b\}} \frac{(GI_{\eta}^i(x, y) - GT_{\eta}(x, y))^2}{GT_{\eta}(x, y)}}$$

显然，GI 较低的像素变灰的可能性更高。

可检测的灰点应具有非常低的 GI，而 CT_{η} 、 $CT_{g\eta}$ 和 CT_b 则不为零。图 1 显示了三个基准数据集和两个不相关的自然数据集上包含可检测的灰色像素的图像的百分比。请注意，此实验中掩盖了 SFU grey ball 数据集中的灰球和 Color-Checker 数据集中的色板。该图表明，每个数据集中超过 95% 的图像包含 GI 低于 0.02 的可检测灰色像素，几乎所有图像包含 GI 低于 0.1 的可检测灰色像素。这有力地支持了以下假设：白色光源下的大多数自然场景确实包含大约为灰色的可检测像素。

下一个问题是这些低 GI 的灰色像素是否可以用作照明估计的准确指标。为了回答这个问题，我们首先通过对色彩偏移图像中检测到的像素求和来估计光源，这些像素具有从其对应的地面真实图像计算出的低 GI 值。然后，计算角度误差

以测量估计的和真实的光源之间的误差。图 2 显示了使用检测到的灰色像素进行光源估计时在三个基准数据集上评估的角度误差。显然，在不同的 GI 水平下，三个数据集的中值和平均角度误差均小于或约为 1.0，就我们所知，这明显优于所有现有的颜色恒定性模型。这表明如果可以正确地检测出这些灰度像素，则可以基于偏色图像的灰度像素获得理想的光源估计。

到目前为止，以上分析已经建立了我们的假设，即大多数自然图像始终都包含一些可检测到的灰色（或近似灰色）像素，这些像素可以可靠地用于准确的光源估计。