

Understanding ISP Pipeline - Tone Mapping

刘斯宁
Camera技术专家

已关注

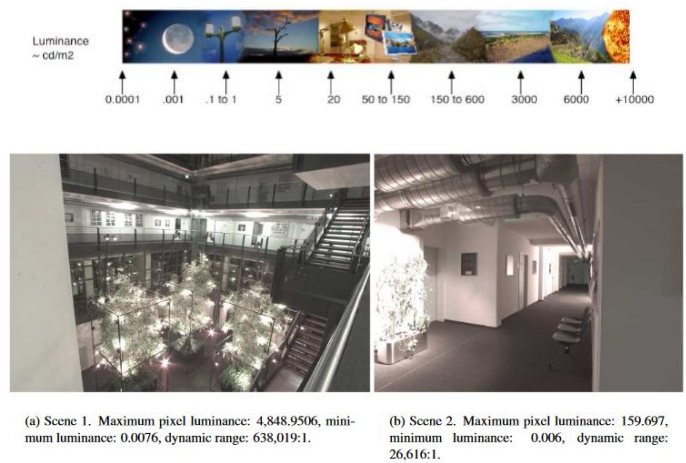
136 人赞同了该文章

色调映射

影像系统的一个核心功能是图像显示，好的显示效果能够真实地再现原始场景，给人的知觉效果就像直接观察原始场景一样。色调映射是影像再现系统的一个重要组成部分，它将原始场景的光照映射成显示设备的发光强度。

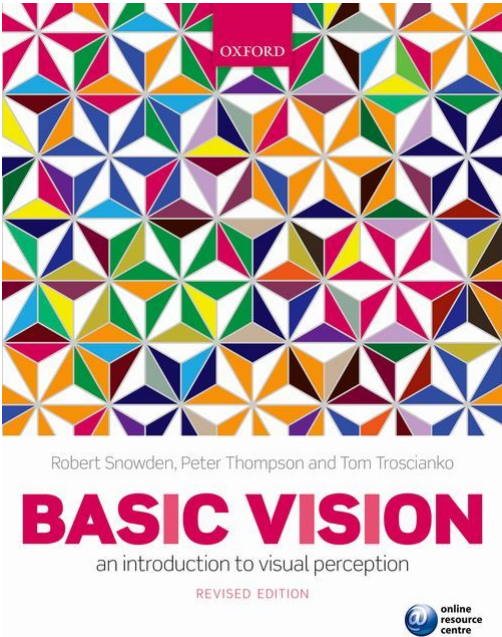
一个好的影像再现系统需要考虑到人类视觉系统（Human Vision System, HVS）是如何处理光线的。场景辐射的光线被人类视网膜上的杆细胞和锥细胞捕捉产生电信号，并传递到视觉神经通道（visual pathway）中做进一步处理。这些信号会被若干层神经网络进行非线性处理，形成的图像被称为**知觉**（percept），此时的知觉图像已经与原始的物理辐射图像相去甚远。

自然场景中可以存在非常广泛的光照条件，从昏暗的夜间场景到明媚的室外场景。下图列举了一些典型场景的光照等级。

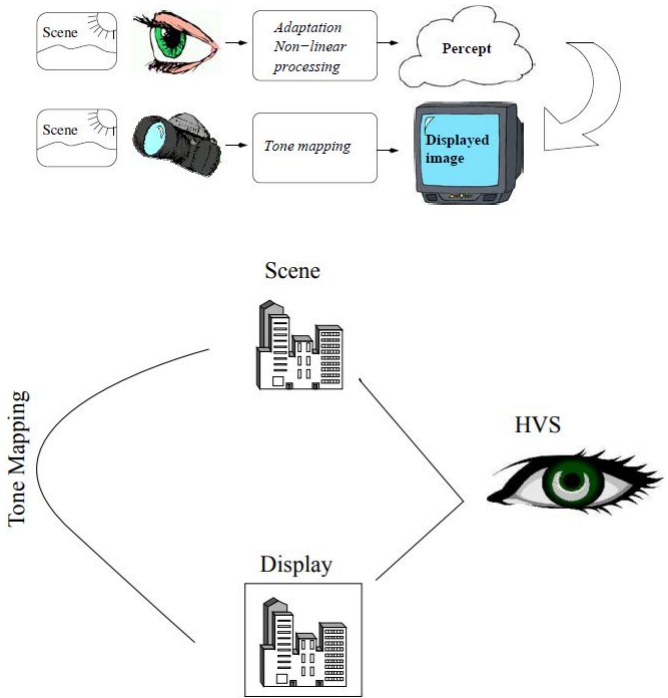


人类为了能够在上述各种场景下都能够正常观察物体并分辨足够的细节，人的视觉系统进化出了多种机制以适应不同的光照条件。首先是瞳孔，它可以改变直径从而控制进入眼睛的光线多少，这是对全局光照强度的一种适应性机制。其次是视锥细胞可以根据视场（Field of View, FOV）上的平均光亮度（即视网膜照度）调节敏感度。最后，当目光注视某一小区域时，视觉系统会自动调节局部对比度。因此，尽管人类生活环境中的光照度变化可以达到8个数量级，而人眼感光细胞的实际动态范围只有4个数量级就能够适应各种环境场景。

关于人类视觉的工作原理可参考Snowden的宝典。



视频捕捉设备生成的图像一般是正比于场景的光辐射强度的，即响应符合线性关系。尤其是当前主流的CMOS Image Sensor（CIS），线性度已经好到可以对入射光子的数量进行准确的计数。CIS的线性特点使人们可以很方便地对摄像机捕捉的原始图像进行处理，从而模拟人类视觉的行为，使设备输出的图像在人眼看来更加自然。这个过程叫做色调映射，如下图所示。



如果不做色调映射处理，摄像机直接输出最原始的线性图像，在人眼看来图像比正常的要暗，如下图所示。



(a) 摄像机捕捉的原始线性图像 (b) 人眼知觉的正常效果

这显然是因为人的视觉系统对视野中的暗区信号做了提升以形成更强烈的知觉，这个特性可以帮助人类识别隐藏在黑暗中的危险，是人类对生存环境的一种适应 (adaptation) 。

场景类型

如果一个场景的动态范围小于影像系统的动态范围，则该场景称为LDR场景，此时需要对图像进行动态范围拉伸，以充分利用设备的动态范围。

如果一个场景的动态范围与影像系统的动态范围大致匹配，则该场景称为SDR场景。

如果一个场景的动态范围大于影像系统的动态范围，则该场景称为HDR场景。由于影响系统的动态范围不足，在图像捕捉过程中很有可能会发生钳位 (clipping)，如果是亮区发生饱和称为高光钳位 (highlight clipping)，如果是暗区发生钳位则称为阴影钳位 (shadow clipping)，在同一场景中两者可能会同时存在。



有人研究了一些典型自然场景的对比度，得出的结论是平均对比度为1:160。室外场景的对比度通常较大，很容易达到甚至超过1:1000，而有雾场景的对比度会比较低。人眼感光细胞的自然动态范围大致是1:10000。

Reinhard, E. and Devlin, K. (2005). Dynamic range reduction inspired by photoreceptor physiology. *IEEE Transactions on Visualization and Computer Graphics*, 11(1):13–24.

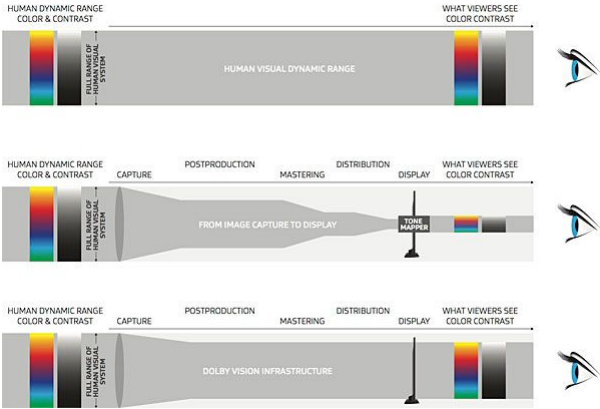
显示类型



当前主流的显示设备有CRT和LCD两种，我们将其归为标准显示设备，将其动态范围定义为SDR，用于区别HDR设备。顾名思义，HDR显示设备的动态范围远高于SDR设备。一般来说，标准显示设备的动态范围为1:100，一些新的LCD设备可以达到1:400。印刷品的动态范围通常为1:50~1:500，取决于采用的技术。较新的HDR显示设备动态范围可以高达1:25000。

当真实场景的动态范围大于显示器的动态范围时，如果不做特殊的处理，超出动态范围的信号就会发生钳位。实际上，由于人类视觉对光照的绝对强度并不敏感，真正起作用的是可视区域内相邻局部之间的亮度对比。根据这个原理，人们可以在图像中过滤掉大部分由光源引起的背景亮度成分，只保留由材质引起的相对亮度成分，这样可以极大地压缩场景的动态范围，使图像在标准显示设备上也能取得与真实场景几乎相同的知觉。

下图对人眼、SDR显示器和HDR显示器的动态范围做了横向对比，并显示了影像处理系统中包括视频捕获、处理、编辑、分发、显示等各个环节上图像动态范围的压缩情况。



Mastering 在音频领域叫母带制作，就是在准备出版CD之前将合成好的多首音乐统一做音频指标修正，以避免在同一张CD内某一首乐曲听起来特别响、而另一首乐曲听起来特别轻，或者一首听起来特别闷，而另一首却特别清脆。推广到视频制作领域，则包括了对图像风格的调整，使视频整体色调、质量统一，在目标媒介上能够取得最佳的效果。

Mastering is the final step of video/audio post-production. The purpose of mastering is to balance video/sonic elements of a stereo mix and optimize playback across all systems and media formats.

全局和局部色调映射

色调映射可以是全局的（Global Tone Mapping, GTM），映射算子具有空间不变性（spatial invariant），也可以是全局结合局部的（Global and Local Tone Mapping, GLTM），映射算子随空间位置而变化（spatial variant）。这两种算子分别对应人类视觉系统的全局适应（global adaptation）和局部适应（local adaptation）特性。

全局色调映射对图像中的所有像素使用同样的映射函数，因此相同的输入像素值会被确定地映射到一个相同的输出像素值。映射函数可以是一个指数函数，对数函数，sigmoid 函数，也可以是由一个由图像本身决定的函数。

局部色调映射对空间位置不同的像素采用不同的映射函数。它的主要特点是，对于同一个输入像素值，由于其空间位置不同或者其周围像素值不同，映射的输出像素值也不同。

在此基础上还可以发展出第三类，即时变色调映射，其映射规则会随时间推移而变化，通常用于处理视频序列。

一般来说，全局色调映射比较简单，计算较快，而局部色调映射算法复杂，计算较慢。在实践中经常会使用查表法加速全局色调映射算法，使其适合摄像机类的实时视频应用。当场景动态范围远大于影响系统的动态范围时，采用全局色调映射会损失较多的对比度和细节，此时需要引入一些局部色调映射的处理，用于改善对比度，保持更多的细节。下图说明了局部色调映射改善对比度和图像细节的实际效果。



(a) 全局色调映射 (b) 全局加局部色调映射

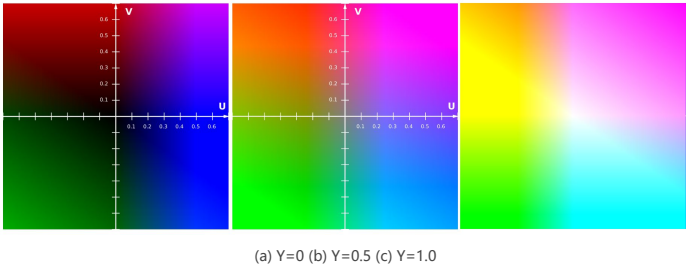


彩色图像的色调映射

一种常用的彩色图像表示方式是RGB格式，其中R，G，B分别代表红、绿、蓝三中颜色通道。对色彩图像进行色调映射，可以理解为对图像的三个通道分别进行三次相同的色调映射运算，对全局色调映射来说，往往能够取得不错的效果。但是局部色调映射则不能这样简单地处理，否则很容易破坏图像的白平衡，在输出图像中引入伪彩。

另一种彩色图像表示方式是将图像从RGB空间变换到YUV 空间，其中Y表示亮度（luminance, luma），UV表示色度（chrominance, chroma）。YUV 空间的主要优势在于实现了亮度和颜色的分离，Y 通道只携带亮度信息，UV通道只携带色度信息。

对于YUV空间的图像，色调映射可以只在Y 通道上处理，UV 通道值保持不变。这种做法的一个缺陷是会影响颜色的饱和度。下图是YUV空间的三个剖面，分别对应Y=0,0.5,1.0，可以看到，如果UV保持不变，则Y值越大颜色的饱和度越低。



从RGB空间变换到YUV空间的方法可以参考本专栏的另一主题文章。

GTM的简单形式

- Log 律

人们常用log曲线模拟人类视觉系统HVS的非线性响应。当图像用 $Y' = \log(Y)$ 记录像素的亮度时， Y' 的尺度能够与人的知觉呈线性关系，换句话说，不管 Y' 当前绝对值是多少，只要 Y' 变化1，则人的知觉也认为亮度变化为1。如果以 Y 为尺度则不能做到这一点。



(a) HDR scene in linear scale (b) HDR scene in logarithm scale

人类视觉系统的log响应特性与sensor采用log特性的出发点是一样的，背后都存在同样的逻辑，

1. 光照的动态范围太大，人眼材料和sensor材料都存在动态范围不足的问题，
2. 人眼进化出强光下降低敏感度的应对措施，事实证明没有影响人类生存，具有现实可行性
3. 既然人眼在强光下会降低敏感度，那么sensor在同样场景下同比例降低敏感度不会影响成像质量。

采用log编码方式的视频在普通显示器上看效果是灰灰的颜色，因此也常被称为log灰视频，在后期制作过程中需要进行调色（color grading），从而恢复出光鲜亮丽的正常效果。





- 1/3 指数律

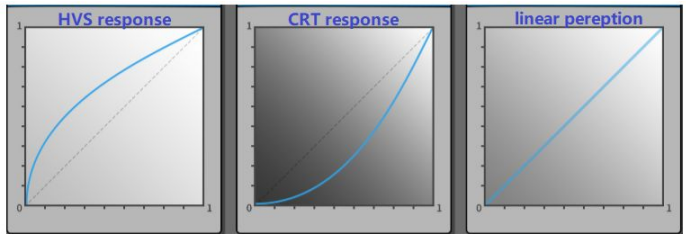
在1~100 范围内，log曲线可以用1/3指数曲线较好地模拟，如下图所示。

CIE Luv 和 CIE Lab (CIE 15:2004 2004) 使用1/3指数曲线模拟HVS系统的非线性特性。

- Gamma

显示设备的输入-输出规律可以用一个指数曲线来描述，该指数一般称为gamma。

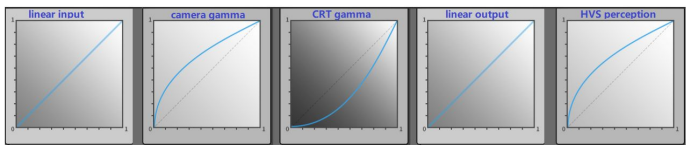
显示设备的gamma 通常设置为2.2，这会使输入的图像变暗。而前文中已经提到，人眼会自动提升暗区的知觉，这相当于对输入信号施加了0.454的gamma，两者叠加的结果，刚好是gamma=1，即**输入图像的亮度会正比于人类的知觉**，如下图所示。



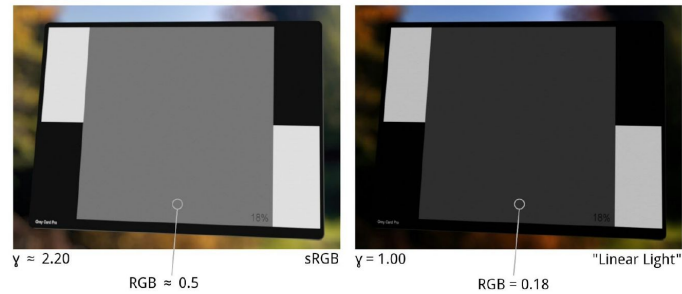
在讨论摄像设备时，摄像机的输出必须与显示设备作为一个系统进行整体设计。在系统内部，显示器具有gamma=2.2，所以摄像机必须设计成gamma=0.454 才能使系统的输出保持线性关系。因此，对于传感器所捕捉到的原始图像I，摄像机需要将其变换为I'，即

其中c表示RGB颜色通道， $\gamma=2.2$ 。

从成像到人类知觉的整个链条中，gamma响应的特性如下图所示。



$\gamma=2.2$ 时，反射系数在18%附近的物体表面会被映射为50%的输出动态范围，这符合人眼的生理知觉特性。



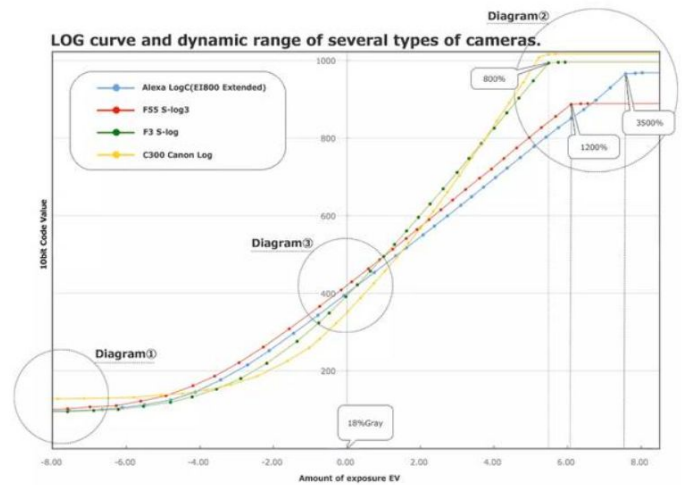
关于本话题的更多内容请参考本专栏的收藏文章

Log律的应用场景

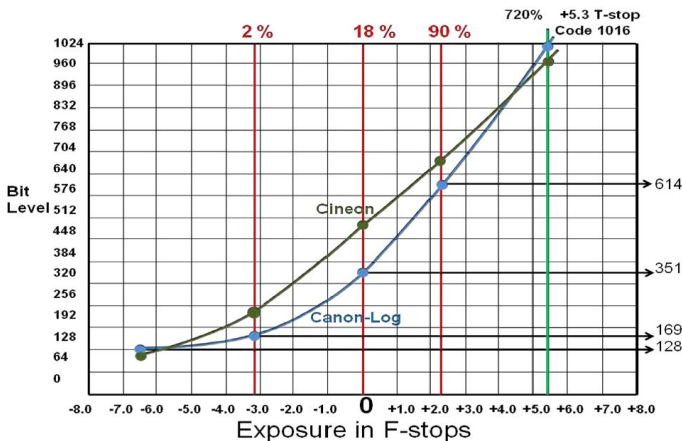
电影机 (Cinema camera) 上使用的CMOS sensor一般是14bit位宽的，如果以RAW格式存储会消耗很大的存储空间。很多电影机产品支持把RAW格式变到RGB空间后压缩成10bit位宽的log格式，进而变换到10bitYUV空间进行编码，可以节省存储和带宽。目前主流电影机品牌都推出了自己的log标准。

- Sony ~ S-Log2 和 S-Log3
- Canon ~ C-log
- Panasonic ~ V-Log , V-Log L
- ARRI ~ Log C
- DJI ~ D-Log 。

下图显示的是一些常用的log曲线，横坐标是以stop为单位的EV值，纵坐标是10bit 编码的log值。各种log方案中，18%中性灰的曝光值都在400左右。



从上图可以看到，采用log方案可以显著地压缩输入信号动态范围，输入、输出信号的动态范围比值可以达到800%~3500%。



C-Log 与 Cineon 比较



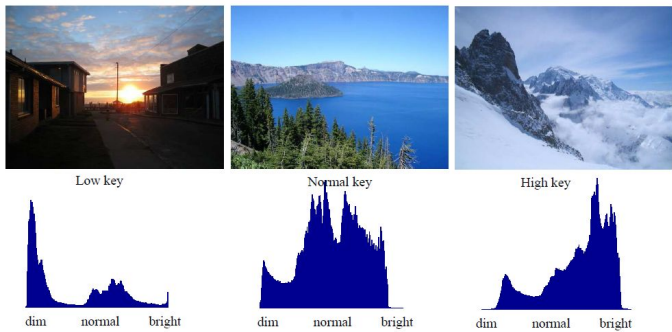
Log C “Log C” 里面的字母 “C” 代表的是 “Cineon” 。“Cineon” 是上世纪九十年代柯达（Kodak）公司开发的胶片数码扫描、处理和记录系统，同时也是一种文件格式的名称（包含了扫描负片的密度数据）。密度是胶片感光特性的对数测量标准，密度与胶片曝光度（以对数单位测量）的关系被称为胶片的特性曲线。每一种胶片都有它自己的特性曲线，但整体看来曲线是一致的。对于ALEXA和 D-21而言，ARRI推出的log编码方式与扫描负片的密度数据相似。因此，它被称为 “Log C”。Log C为后期调色提供了极大的灵活性，它在未经任何处理时是消色的，画面非常平淡，所以需要使用查色表（LUT）进行处理，使其在人眼中呈现出自然的状态。

Log C拍摄当扫描的胶片素材进入调色阶段，调色师会使用一个预览查色表（LUT）对文件进行色彩渲染，使监视器上显示的图像与以胶片方式显示的图像一致。来自ALEXA 摄影机的Log C素材也可以像胶片扫描的画面一样，通过预览查色表（LUT）进行颜色修正。在制作以及其他诸多应用中，查色表（LUT）需要预先准备好，并应用于以下一些情况：拍摄现场来自摄影机的实时信号以及用来回放审看的录制素材，通常需要利用查色表（LUT）来转换，并在监视器上显示正常的图像。实际上，此时已经不是在模仿印片的查色表，而是将LUT作为转换画面的一种方式，好像摄影机正在拍摄 “所见即所得” 的高清视频（Rec 709）效果。

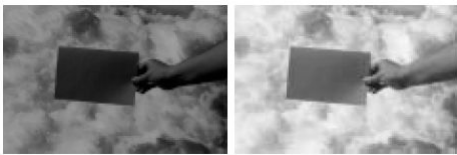
关于gamma和log有一个B站视频讲得挺不错，值得推荐。

图像的主色调

对于任一幅图像，可以按照其主色调（多数像素所在的亮度区间）将其归纳为三类色调中的一类，即低调、中调、高调。三种色调的图像在直方图上存在明显的特征。



在对图像进行色调映射时，映射算法需要考虑图像的色调特点，将更多的动态范围分配主色调，这样给才能取得最佳的映射效果。下图是一个高调图像的例子，画面中的18度灰卡应该映射为灰度值128，背景中的白雪应该映射为灰度值200以上。如果把它错误地映射为一个中调图像，就会变成(a)的效果。



(a) normal-key map (b) high-key map

GTM的复杂形式

很多色调映射方法基于分区的思想实现，大致思路是把输入图像（设备）的动态范围划分为若干个区间，输出图像（设备）的动态范围也划分为若干个区间，然后寻找一种方法将输入区间映射到输出区间。



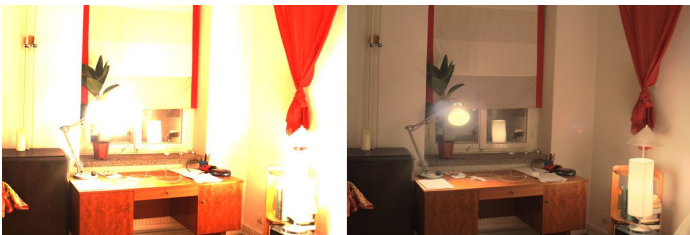
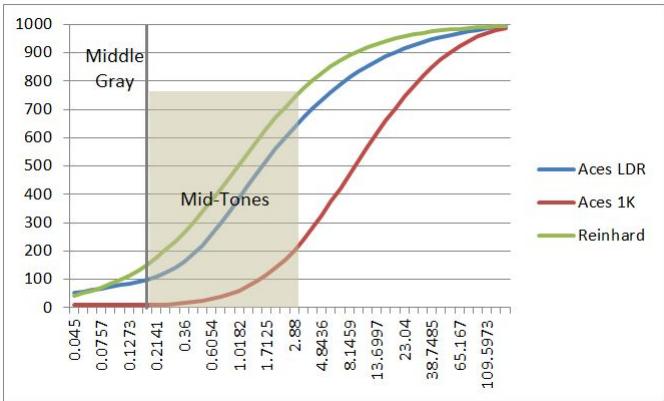
Holm 定义了一条S形的sigmoid映射曲线，而曲线的参数取决于场景各分区的特性（最大/平均/最小亮度），以及输出介质的动态范围。这种方法对SDR图像效果较好，但是对HDR图像有时会多度压缩局部细节。

Holm, J. (1996), Photographic tone and colour reproduction goals, *in* Proc. CIE Expert Symposium'96 on Colour Standards for Image Technology, Vienna, Austria, pp. 51–56.

S形曲线相对简单，它的一般特点是中间段比较陡（斜率大，上升快），两边（暗区、亮区）比较缓。对于很多图像来说，中间色调（反射率18%附近）往往处于S型曲线比较平缓的区域，因而受到较大的压缩，信息损失较大。

Reinhard 等人提出一种方法，首先将输入图像的log均值映射到输出空间的某个值。如果输入图像的主色调（key）是中调的，则映射为输出空间的中性灰，如果输入主色调是低调或者高调的，则映射到输出空间的值在中性灰的基础上适当地减小或者增加。Reinhard 方法没有使用S形曲线，而是使用了另一种方法只压缩亮区，并且提出了高光钳位的方法，因为在很多时候其实并不需要关注图像亮部的细节。

Reinhard, E., Stark, M., Shirley, P. and Ferwerda, J. (2002), Photographic tone reproduction for digital images, *in* Proc. ACM SIGGRAPH 2002, Annual Conference on Computer Graphics, San Antonio, CA, pp. 267–276.



(a) HDR scene in linear scale (b) HDR scene with Reinhard GTM

Ward 等人提出了一种基于直方图均衡化的映射方法，它按照一定的方法将像素值重新进行分布，使输出图像的直方图接近均匀分布。直方图均衡算法一方面需要保持图像的对比度，另一方面还要注意，如果某些细节在输入图像中对人眼不可见，那么在输出图像中也需要保持不可见。简而言之，就是把真实世界的最小可辨差（Just Noticeable Difference, JND）映射为输出设备的最小可辨差。

Ward, G., Rushmeier, H. and Piatko, C. (1997), A visibility matching tone reproduction operator for high dynamic range scenes, *IEEE Transactions on Visualization and Computer Graphics* 3(4), 291–306.

黑白校正



任何色调映射算法都需要包含一个过程，就是把输入图像的黑点映射为显示设备的最暗值，把输入图像的白点映射为显示设备的最亮值，这个过程就叫做黑白校正，它对图像还原是不可或缺的过程。

对于有些色调映射算法，如基于直方图均衡化的算法，黑白校正已经自动包含在映射算法当中。而对于另外一些算法，黑白校正需要作为映射算法的预处理或后处理环节单独完成。当黑白校正作为独立步骤时，可以参考以下公式对像素进行处理。

其中，

b 为黑点阈值，不大于此值的像素都被钳位至0；

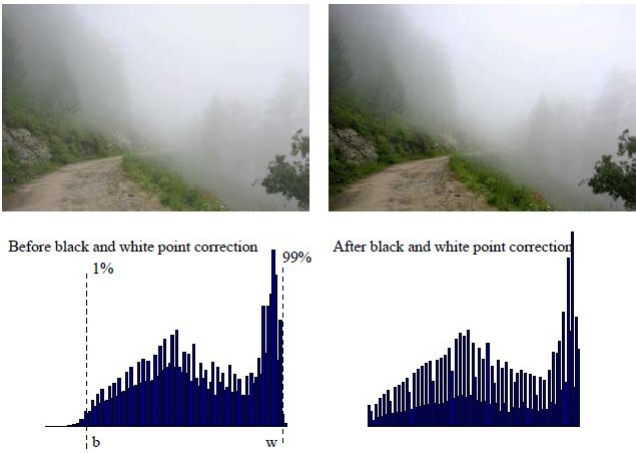
w 为白点阈值，大于 w 值的像素都会被钳位至 w ；

c 代表图像的RGB通道；

$I(p)$ 为输入、输出图像的像素值；

b, w, I 等参数的最大值均归一化至1。

下图是黑白校正的一个例子，取 $b=0.01$ ， $w=0.99$ 。



黑白校正有利于改善低动态（LDR）图像的对比度，用一个比较常用的主观词语描述就是改善了画面的“通透性”（clarity）。通常人们感觉画面不通透的主要原因就是暗的像素不够黑，感觉画面像是蒙了一层雾（有时真的是雾），而黑白校正能够在一定程度上起到去雾的作用。目前有些ISP 已支持此功能，并允许通过软件配置 w, b 参数的值。

Retinex 理论

Edwin Land 等人于1964年提出了Retinex 理论，试图解释人类视觉系统（HVS）如何在光照变化的条件下稳定可靠地提取场景中的信息。基于这个理论，人们发展了一些行之有效的色调映射算法，这也是本文的重点。

黑白蒙德里安实验

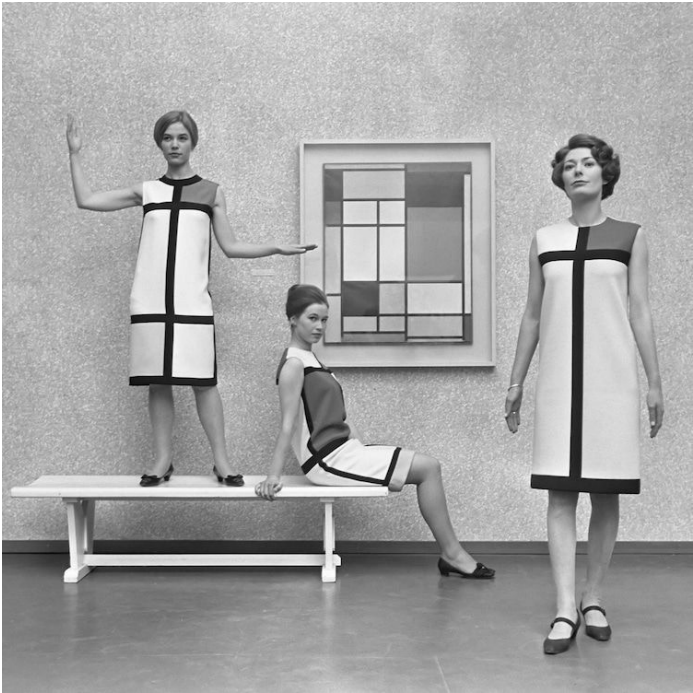
Edwin Land 等人的实验主要采用了下图所示的测试图案，这个图案被命名为蒙德里安图案，其灵感来自于荷兰抽象派画家Pieter Cornelis Mondrian（1872-1944）。

Land 设计的第一个实验使用了黑白蒙德里安图案，通过控制光照使图案从上到下亮度逐渐且均匀地增强。尽管图案下方的实际亮度显著高于图案上方，但是在受试者看来，图案下方圆环的亮度知觉与上方圆环的亮度知觉并无不同。这个实验说明人类视觉能够自动过滤由光源引起的全局亮度（global illumination），从而使注意力集中在由材料表面反射引起的相对亮度（reflectance illumination）上。



P.S.

Mondrian 的绘画作品主要是极简主义风格，只使用最基本的线条、方块和简单颜色组合成抽象的几何图案。下图摄于海牙Gemeente博物馆，穿着Mondrian 风格服装的模特在展示Mondrian 的代表作。



彩色蒙德里安实验

第二个实验采用了两个完全一样的由彩纸制成的蒙德里安图案（A和B），并使用短、中、长三种波长的光源照明，每种光源的功率可独立调整。实验过程中，A图案光源可变，B图案光源不变作为参考。Land 调整了A图案的照明，使A图案中某颜色块的RGB光辐射功率与B图案中白色块的RGB光辐射功率完全相同。尽管事实如此，但是当受试者同时看到A、B两个图案时，B图案中的白色块仍然知觉为白色，而A图案中的颜色块仍然知觉为其本来的颜色。

以上实验证明，人类关于亮度的知觉与入射到视网膜上的辐射功率无直接关系。并且人类关于某物体颜色的知觉不但与该物体有关，还与其周围其它物体有关，但是与光源的频段和强度无关。换句话说，真正决定物体颜色的是不同物体表面反射率之间的比值，而不是单个物体反射率的绝对值。后人进一步发展完善了这个理论，最终形成了关于颜色的Retinex理论。

Retinex 计算模型

Retinex 理论是HVS 模型的一个近似，以取得计算上的可行性。目前基于Retinex 理论已经发展出一系列算法，它们共同的思想是将输入图像的每个像素用一个新的值来替代，而替代值根据输入像素的空间位置以及周围一个邻域内的像素计算得到，而周围像素的权重与距离中心像素的远近有关，也与权重函数的选择有关。

Path Retinex

Retinex 算法的核心问题是将输入图像分解为反射图像（reflectance image）和光源图像（illuminant image）两个分量，从输入图像中去除光源分量后就得到了物体本身的相对亮度。



Land 提出了一种基于路径的Retinex 算法称为path-based Retinex。

Land, E. (1964), The Retinex, *American Scientist* 52(2), 247–264.

这个算法大致分为两个步骤，对于一个像素 p ，算法在以 p 为中心的一个邻域 σ 内选择一条随机的路径 Γ ，并计算该路径的Retinex

其中

I 为输入图像，

下标 c 代表图像的RGB通道，

下标 k 代表路径 Γ 上的一个输入像素，

R 代表路径 Γ 所对应的Retinex 值，

而函数

重复以上步骤 K 次后可以得到 K 个不同路径的Retinex 值，将他们取算术平均就得到像素 p 的Retinex 值，

路径算法的一个主要挑战在于随机路径的选择，采用不同的路径将得出不同的结果，这使算法的效果变得难以预测。一个不能保证收敛的算法在工程上的价值是非常有限的，因此很多人开始探索将路径算法工程化、收敛化的方法。

Horn 提出，Land 路径算法中可以使用2D拉普拉斯算子计算光源分量。

Horn, B. K. P. (1974), Determining lightness from an image, *Computer Graphics and Image Processing* 3, 277–299.

Brainard 等人重点研究了路径生成算法，他们指出，随机路径的选择符合用马尔科夫随机过程模型，当路径的数量足够多、路径足够长的时候，结果等效于使用最大像素值做归一化。

Brainard, D. and Wandell, B. (1986), Analysis of the Retinex theory of color vision, *Journal of the Optical Society of America A* 3(10), 1651–1661.

Hurlbert 指出，Retinex 模型在数学上其实等效于求解泊松方程。她将若干个Retinex 算法整理抽象成一个统一的数学形式，并推定完整的解可以使用格林公式（Green's Theorem）求得，而之前的几个Retinex 算法可以理解格林公式的近似解。

Hurlbert, A. (1986), Formal connections between lightness algorithms, *Journal of the Optical Society of America A* 3, 1684–1692.

除了收敛性问题，Retinex 算法的另一个挑战是计算量大，不易工程化。Elad 等人总结了一些方法，包括使用查找表，线性/非线性滤波器，裁剪基底数量等方法，可以在一定程度上减少计算量。

Elad, M., Kimmel, R., Shaked, D. and Keshet, R. (2003), Reduced complexity Retinex algorithm via the variational approach, *Journal of visual communication and image representation* 14(1), 369–388.

迭代Retinex

迭代Retinex 是路径Retinex 算法的二维扩展。与路径Retinex 类似，它也采用四种基本运算，即 ratio, product, reset, average，基本方程如下。

$$\mathcal{R}_c^j(p) = \frac{\Omega[\log(I_c(p)) - \log(I_c(p + \delta))] + \mathcal{R}_c^{j-1}(p + \delta)] + \mathcal{R}_c^{j-1}(p)}{2},$$

其中

δ 是邻域像素与中心像素 p 的距离，从 $w/2$ 和 $h/2$ 开始，每轮迭代后减半直到 $\delta=1$

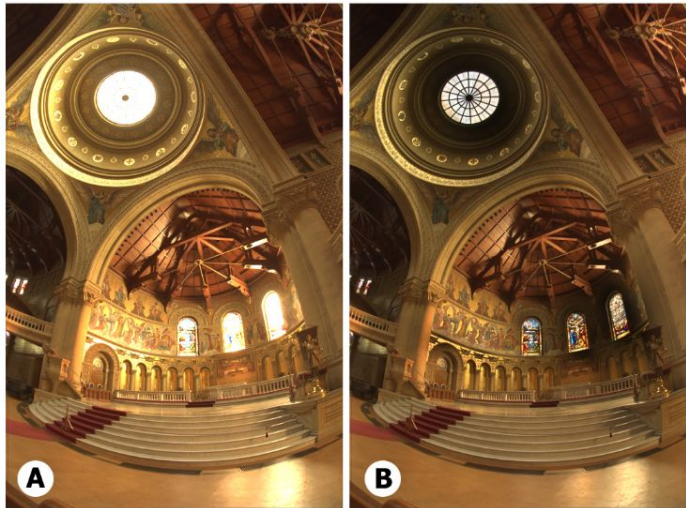


j 是迭代次数，对于每个 δ 分辨沿四个方向迭代四次，

Ω 是reset 函数，

这个算法的主要问题在于迭代的具体方式并不十分明确，不同的迭代次数（方法）可以导致非常不同的结果，所以算法本身仍然存在较大的不确定性和收敛问题。

下图是迭代算法的一个例子，图片摄于The Stanford Memorial Church，场景动态范围约为350,000 : 1



(a) linear mapping (b) Retinex mapping, 4 iterations

环绕Retinex

Land 又提出了一种新的Retinex 算法，称为center/surround-based Retinex，这是一个不需要迭代的算法，每个像素只处理一次，输出像素值由输入像素和周围像素的加权平均值共同决定。

Land, E. (1986), An alternative technique for the computation of the designator in the Retinex theory of color vision, *in* Proc. National Academy of Sciences of the United States of America, Vol. 83, pp. 3078–3080.

在此基础上，Jobson 和Rahman 等人发展出了多尺度Retinex算法，即Multi-Scale Retinex with Color Restoration，简称MSRCR。

Jobson, D., Rahman, Z.-U. and Woodell, G. (1997), A multiscale Retinex for bridging the gap between color images and the human observation of scenes, *IEEE Transactions on Image Processing* 6(7), 965–976.

Rahman, Z.-U., Jobson, D. D. and Woodell, G. A. (2004), Retinex processing for automatic image enhancement, *Journal of Electronic Imaging* 13(1), 100–110.

MSRCR 算法首先按照以下公式算出一组（三个）单尺度Retinex 图像，每个尺度对应不同的邻域大小。

其中

p 代表像素，

R 代表尺度 σ 对应的Retinex 图像，

下标 c 代表图像的RGB 通道，

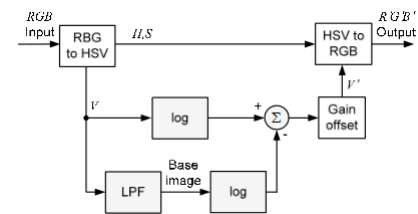
下标 σ 代表邻域的尺度，通常由图像尺寸（像素数）定义，

G 代表2D高斯滤波器，

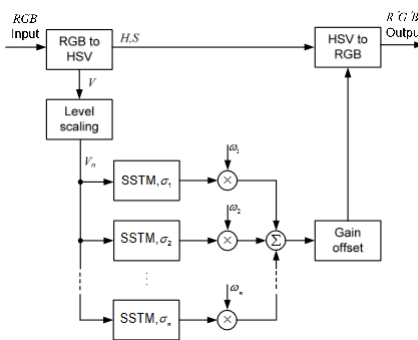
三个单尺度对应的高斯滤波器如下图所示，



单尺度Retinex 常简写为SSTM (Single Scale ToneMapping), 功能框图如下图所示。



将三个SSTM 取算术平均即得到多尺度Retinex,

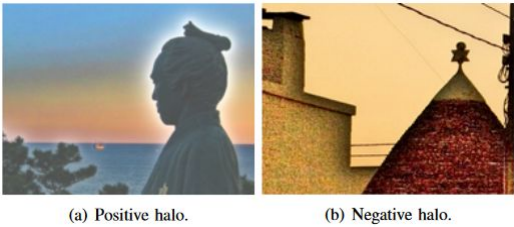


Meylan 等人对MSRCR 进行了一定的优化，将三个尺度所需的所有参数打包合并到一个滤波函数中，并且输入图像为YUV，算法只处理Y 通道，不处理UV 通道，因此计算效率可以大幅提高。

Meylan, L. and Süsstrunk, S. (2004a), Bio-inspired image enhancement for natural color images, *in Proc. IS&T/SPIE Electronic Imaging 2004. The Human Vision and Electronic Imaging IX*, Vol. 5292, San Jose, CA, pp. 46–56.

Meylan 算法使用的滤波器如下图所示。

此类Retinex 算法一个共同的问题是在图像明暗交界的地方容易发生光晕效应（halo artifacts），如下图所示。



另外一个常见问题是图像中的低对比度区域容易变灰，如下图所示。





变分模型 Variation Model

在数学领域，一个问题是不适定问题(well-posed problem)需要同时满足以下三个条件

- 1. 解是存在的 (a solution exists)
- 2. 解是唯一的 (the solution is unique)
- 3. 解是稳定的 (the solution's behavior changes continuously with the initial conditions)

只要任何一个条件不满足，这个问题就是一个不适定问题(ill-posed problem)。

在计算机视觉中，有很多不适定问题通常是不满足第二条和第三条。比如做图像超分辨率，或者对图像去雨去雾去模糊等任务，这些都没有一个标准答案，解有无数种。更重要的是，这些解都是不稳定的，选择不同的初始条件可以得到差异很大的结果。由于Retinex 模型并未规定一个求解光源图像 (Illumination Image) 的标准方法，所以Retinex 本质上也是不适定问题。目前已经有多种方法可以求解光源图像，其中比较有代表性的是变分模型，它在算法复杂度和图像效果之间取得了比较好的平衡。

变分模型的核心是一个如下所示的代价函数 (cost function) ，

$$F[l] = \int_{\Omega} (|\nabla l|^2 + \alpha(s - l)^2 + \beta|\nabla(s - l)|^2)dx dy,$$

其中，

α, β 为权重系数，

$s(x,y), L(x,y)$ 分别代表输入图像和光源图像，以像素值的对数形式表示，

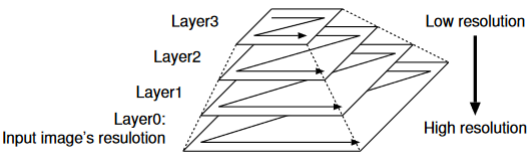
∇ 为梯度算子，图像越平滑则梯度越小，

令 $r(x,y)=s(x,y) - L(x,y)$ 为反射图像，则

- 代价函数第一、三项的意义是，光源图像 $L(x,y)$ 和反射图像 $r(x,y)$ 越光滑，整体代价就越小。
- 代价函数第二项的意义是，反射图像 $r(x,y)$ 值越小，整体代价就越小。

采用最速下降算法 (Projected Normalized Steepest Descent, PNSD) 可以求出使 $F(L)$ 最小的光源图像 $L(x,y)$ 。

算法需要构造一个如下图所示高斯金字塔，



金字塔使用使用以下滤波核

令 k 为金字塔的层级 (0层为无压缩的原始图像) , j 为迭代次数，梯度函数定义为

其中 v 为拉普拉斯梯度算子

进一步定义 μ 因子，



于是迭代公式为

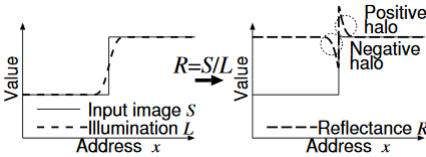
$$l_j^{(k)}(x,y) = \max \left(l_{j-1}^{(k)}(x,y) - \mu_{NSD} G(x,y), s^{(k)}(x,y) \right)$$

迭代从金字塔的顶层开始（最小分辨率，k=N），经过T(k)次迭代后L(k)趋于收敛，使用L(k)构造第k-1层图像，进行新一轮迭代，直到得到第0层图像就是所需的光源图像。

由于s(x,y)和 L(x,y) 都是对数形式表示，对s(x,y) - L(x,y) 求指数就是最终的反射图像。

光晕效应

由于L是从s计算得到的，所以L相对s总会存在一定的滞后，在图像中较陡的边缘处，s-L会产生绝对值非0的分量，求指数后就出现了或亮或暗的过渡带，这就是光晕效应（halo artifacts）的来源，如下图所示。



Hiroshi 等人提出了一些可以抑制光晕的方法。

Halo Artifacts Reduction Method for Variational based Realtime Retinex Image Enhancement

Hiroshi Tsutsui*, Satoshi Yoshikawa[†], Hiroyuki Okuhata[‡], and Takao Onoye[‡]
*Dept. of Communications and Computer Engineering,
Graduate School of Informatics, Kyoto University
Yoshida-Honmachi, Sakyo, Kyoto 606-8501, Japan
[†]Dept. of Information Systems Engineering,
Graduate School of Information Science and Technology, Osaka University
1-5 Yamadaoka, Suita, Osaka 565-0871, Japan
[‡]Synthesis Corporation
Awajimachi TC Bldg. 4th Floor, Awajimachi 2-6-9, Chuo-ku, Osaka 541-0047 Japan

降噪Retinex

Retinex 算法主要使用梯度和除法运算，当场景比较暗时，图像信号的幅值常常小于噪声幅值，而梯度和除法运算会放大噪声，恶化信噪比。



(a) original image (b) Retinex output

北航Zhenwei 团体于2013年提出了一种方案，思路是给Retinex 图像分配一个权重W

$$I_i(x,y) = F_i(x,y) \cdot W_i^1(x,y) + S_i(x,y) \cdot (1 - W_i^1(x,y))$$

当输入像素S(x,y)亮度较高时，Retinex 图像F(x,y)的权重W(x,y) 较大，输出值以Retinex 为主；当输入像素暗度较低时，权重W(x,y) 较小，输出值以S(x,y) 为主。

Multi-scale retinex improvement for nighttime image enhancement[☆]

Haoning Lin, Zhenwei Shi*

Image Processing Center, School of Astronautics, Beihang University, Beijing, China

北大Jiaying 团队于2017年提出了另一种思路，他们使用以下公式，将空间滤波器合并到Retinex 代价函数中，取得了较好的效果。



$$\operatorname{argmin}_{\mathbf{R}, \mathbf{L}} \|\mathbf{R} \circ \mathbf{L} - \mathbf{I}\|_F^2 + \alpha \|\nabla \mathbf{R}\|_F^2 + \beta \|\nabla \mathbf{L}\|_1 + \omega \|\nabla \mathbf{R} - \mathbf{G}\|_F^2,$$

其中

\mathbf{I} , \mathbf{R} , \mathbf{L} 分别为输入图像, 反射图像和光源图像,

$\mathbf{R} \circ \mathbf{L}$ 代表逐元素相乘操作,

$\|\cdot\|_F$ 代表Frobenius 范数

$\|\cdot\|_1$ 代表L1 范数

α, β, ω 为代价项的权重参数

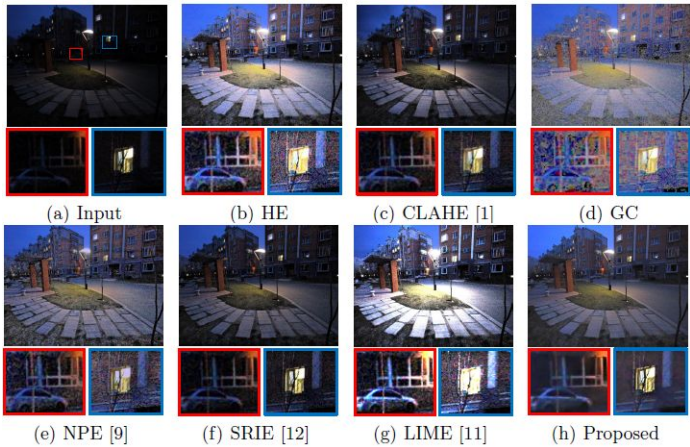
定义关于 \mathbf{I} 的梯度算子

当梯度 $\nabla \mathbf{I}$ 小于某阈值 ϵ 时, 将其视作噪声钳位至0, 以起到空域低通滤波器的作用。

于是可以进一步定义Retinex 的梯度算子 \mathbf{G} ,

其中 \mathbf{K} 的作用调节梯度 $\nabla \mathbf{I}$ 的权重, 当 $\nabla \mathbf{I}$ 较小时, \mathbf{K} 将其放大, 当 $\nabla \mathbf{I}$ 较大时 \mathbf{K} 趋近于单位算子。 \mathbf{K} 的调节作用使 \mathbf{G} 的取值趋同, 即压缩了 \mathbf{G} 的动态范围。因此从趋势上看, 代价项 $\|\nabla \mathbf{R} - \mathbf{G}\|$ 可以理解, 当梯度 $\nabla \mathbf{I}$ 小于某阈值 ϵ 时 $\nabla \mathbf{R}=0$ (降噪作用), 否则 $\nabla \mathbf{R}=\nabla \mathbf{I}$ (保留对比度)。

使用以上算法可以取得较好的降噪效果, 下图是各种Tone Mapping算法的效果比较



其中

(a) 是输入图像, (b) 是直方图均衡化的方法, (d) 是gamma校正方法, 其它对比方法在原文的参考文献中给了出处。

Joint Denoising and Enhancement for Low-Light Images via Retinex Model

Mading Li, Jiaying Liu*, Wenhan Yang, and Zongming Guo

Institute of Computer Science and Technology, Peking University, Beijing,
P. R. China, 100871

综述

上文中介绍了色调映射的基本原理和一些典型算法, 关于此领域还有很多有代表性的算法没有涉及到, 下面的文章对2016以前的相关工作做了总结和对比, 可以提供很多有价值的信息。



Which Tone-Mapping Operator Is the Best? A Comparative Study of Perceptual Quality

XIM CERDÁ-COMPANY, C. ALEJANDRO PÁRRAGA and XAVIER OTAZU
Computer Vision Center,
Computer Science Department,
Universitat Autònoma de Barcelona,
Spain

编辑于 2020-11-18 11:47

[Camera](#) [图像信号处理器ISP \(Image Signal Processor\)](#) [图像处理](#)