

《高动态视频技术》

目录

《高动态视频技术》	1
第一部分-HDR 技术基础	4
第一章-光学与人类视觉感知特性基础.....	4
1. 1 可见光	4
1. 2 动态范围	5
1. 3 人类视觉感知系统	6
第二章-光电转换与 HDR 图像显示	8
2. 1 Gamma 校正	8
2. 2 HDR 技术中的光电转换	12
第三章-色度学与颜色空间	20
3. 1 色度学	20
3. 2 色域	25
第四章-HDR 主要标准	31
4. 1 HDR 发展现状	31
4. 2 HDR 相关技术组织	32
4. 3 HDR 中的转换曲线	33
4. 4 HDR 元数据规范	35
4. 5 HDR 内容的存储及分发标准	36
第二部分 HDR 处理技术	43
第五章-HDR 生产流程	43
5. 1 实现 HDR 的条件	45
5. 2 HDR 内容制作	46
第六章-色调映射	54
6. 1 色调映射算法综述	54
6. 2 图像色调映射算法	60
6. 3 视频色调映射算法	71
第七章-逆色调映射	80
7. 1 逆色调映射概述	80
7. 2 逆色调映射算法	82
第八章-色域映射	100
8. 1 色域映射概述	100
8. 2 HDR 中的色域映射标准与算法	101
8. 3 色域映射算法	108
第九章-HDR 质量评价技术	114
9. 1 HDR 主观评价方法	114
9. 2 HDR 客观评价方法	122
第三部分 HDR 编码方案	131
第十章-MPEG 及 VCEG 的 HDR 编码优化	131
10. 1 背景	131
10. 2 HDR 发展历程	132
10. 3 HDR 视频编码端的关键技术	135
第十一章-HEVC/H. 265 的 HDR 编码方案	141

11.1 背景	141
11.2 HEVC 编码对 HDR/WCG 的支持.....	141
11.3 其他相关编码方案	150

第一部分-HDR 技术基础

视频与图像，是人类认识自然，感受自然，记录自然的两种最为直观的方式，因此，自从 1839 年第一台照相机被发明出来之后，人类一直致力于更好地，更完善地，更精确地记录自然界的光影。HDR 技术的发明与普及，使得人类前所未有地接近通过人造的显示设备，展示出如同自然界一样的丰富的色彩与丰富的亮度。

传统图像存在的问题是：与现实中的色彩相比，传统图像中的颜色数量太少，这就会导致很多自然界中的图像并不能完美地显示在图像上，从而导致色差的出现；其次，传统图像的动态范围过低，动态范围是指图像中亮度的最大值与最小值之比。图像的动态范围过低会导致两个弊端，第一个是图像中的亮度比现实中的亮度低太多，造成场景还原的失真，另一个就是现实中富有层次感的画面或者场景，记录在图像中后，就失去了层次感，即图像的对比度下降。这种现象产生的原因是现实中的动态范围很大，所以场景中不同的区域会有不同的亮度，不同区域之间亮度的改变就会带给人丰富的层次感；而图像中的动态范围小，就会导致原先现实中不同亮度的区域在图像中的亮度趋于一致，由此产生了层次感的下降。

相比传统的图像（即日常生活中见到的大多数图像），HDR 图像给予人们的直观感受是图像更接近于自然了，图像亮度增强了，图像的色彩与细节更丰富了，图像也更具有层次感。这些现象产生的原因在于：HDR 图像技术，相比于传统图像技术，在两方面进行了提升与拓展：第一是亮度，HDR 技术提供并支持了更加广泛的亮度范围；第二个是色彩，HDR 技术支持更丰富的色彩范围。由此，图像的细节表现与细微差异能够清晰地展示给观看者。

本书的第一部分将从亮度和色彩两个方面，介绍 HDR 技术的理论基础，并在第三章中介绍主流的 HDR 技术标准。

第一章-光学与人类视觉感知特性基础

HDR，即高动态范围是一个光学以及视觉领域的概念，如果想要清楚地理解什么是 HDR，HDR 与 LDR 的差别在哪里，以及 HDR 技术的理论基础，首先需要对光学以及人类视觉感知系统（Human Visual System，HVS）有基本的了解。

1.1 可见光

图像，本质上是对可见光的记录与还原。从电磁波的角度来看，可见光可以看作电磁波谱中人眼可以感知的部分。自然界中，人类熟知的电磁波的波长可以从 $10^{-10}m$ (γ 射线) 一直到 10^3m (无线电)，一般人的眼睛可以感知的电磁波的波长在 $400\sim760nm$ 之间，但还有一些人能够感知到波长大约在 $380\sim780nm$ 之间的电磁波。由此可见，可见光的波长范围相比电磁波的波长，是非常小的范围。。

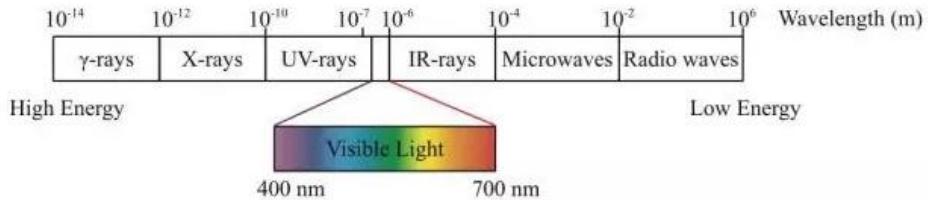


图 1-1. 光谱与可见光范围

从另一个角度来看，光可以被认为是由光子组成的，光子可以被发射，被反射，被吸收以及可以在空间中传输。光子在空间中延直线传播。当遇到物体时，光子可能产生了两种结果：第一种是被物体表面吸收，并产生热能；第二种是被物体表面反射。反射方向依据物体表面情况又可以分为两种：当物体表面粗糙时，光会漫反射，即出射角度可能是任何方向；当物体表面光滑时，光会发生镜面反射，即出射角度基本只有一个方向。光的这些性质，是我们记录图像，显示图像的基础。

从现实中的经验我们可以了解到，镜面反射的光线会对图像的质量产生不良的影响，因为镜面反射的光线强度过大，因此在图像中会显示为白茫茫的一片，从而失去了这一部分图像的精确细节。如何获取，记录这种过曝光区域的图像细节，也是 HDR 技术领域的一个重点问题。

由于人眼的特性，不同波长的可见光，对人眼的刺激是不同的。更进一步，对于同一个波长的可见光，不同的人的敏感程度也是不同的。但是，因为可见光的波长范围较小，所以即使不同的人的敏感程度不同，也可以用一个统一的曲线来大致描述。这种描述曲线由 Commission Internationale de l'Eclairage (CIE) 标准化，并命名为 $V(\lambda)$ ，CIE photopic luminous efficiency curve。曲线图像如图所示：

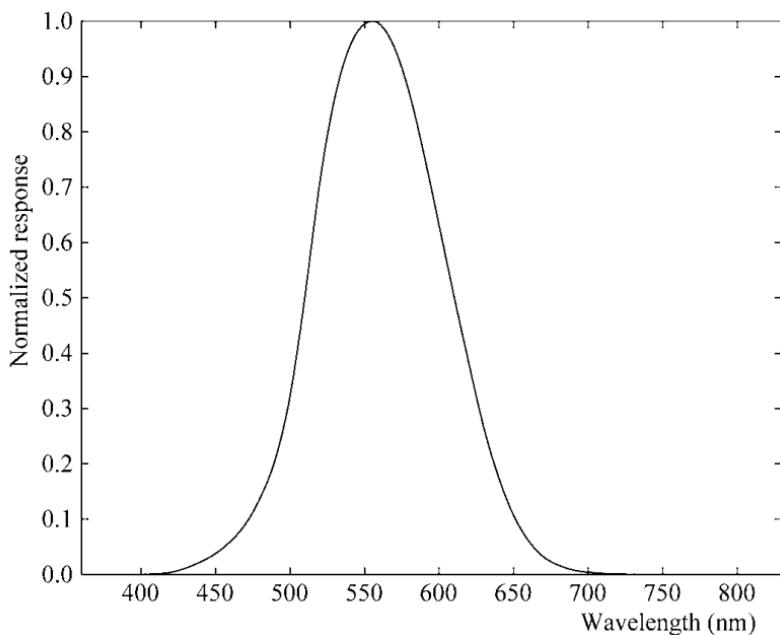


图 1-2：人眼可见光敏感度归一化曲线。

1.2 动态范围

在亮度方面，还有一个重要的特征是相对亮度。因为相对亮度与 HDR 技术中的重要概念：动态范围息息相关。因此，这里我们将详细介绍动态范围，以及动态范围与相对

亮度的关系。

动态范围是指某度量的最大和最小值之间的比值。在 HDR 技术中，动态范围就是指图像的最大亮度与最小亮度的比值。描述相对亮度的物理量我们称之为对比度。对比度描述了场景或显示设备的最大亮度与最小亮度之间的关系。因此我们可以认为，在显示设备上，HDR 图像技术需要实现的高动态范围，实际上就是实现高对比度。

对比度也有几种不同的计算方式，其中最重要的三种对比度分别为：韦伯对比度 (C_W)，麦克森对比度 (C_M) 以及比率对比度 (C_R)，定义为：

$$C_W = \frac{L_{max} - L_{min}}{L_{min}}, C_M = \frac{L_{max} - L_{min}}{L_{max} + L_{min}}, C_R = \frac{L_{max}}{L_{min}}$$

其中 L_{max} 与 L_{min} 分别为场景中的最大与最小亮度值。

1.3 人类视觉感知系统

人类视觉感知系统在图像与视频领域是一个非常重要的环节，因为很多后续的研究与定理都是建立在人类视觉感知系统的基础之上的。

人类视觉感知系统是一个十分抽象的概念，而且其中包含的内容现在仍在改变与拓展，但是，这并不妨碍我们对人类视觉感知系统做出相应的介绍。简而言之，人类视觉感知系统就是人为地对人眼的生理视觉特性进行模拟，从而实现对图像与视频处理方式的优化与流程的简化。但是，人类的生理视觉系统十分复杂，并且到现在也没有完全研究清楚，因此在图像领域我们无法完全模拟人眼的特性。

由于生理视觉的复杂性，现在的人类视觉感知系统，实际上只是对人眼的部分特性的简单归纳与运用。随着人类对视觉的研究不断加深，人类视觉感知系统也将随之不断发展。一个典型的人类视觉感知系统特性的应用实例就是图像压缩。研究发现，人类的视觉对于图像的高频部分不敏感，因此，在图像压缩的过程中，弱化或忽略高频部分可以使图像压缩的性能大幅提高，并且基本不会影响人类的视觉感知。

由于人类视觉系统的极大复杂度，许多视觉物理学现仍处于假设研究阶段。一些研究已经发现了低层的视觉物理学特性，包括：亮度非线性、对比度敏感、掩盖效应、多通道并行及视觉注意力等等。

亮度非线性指相对于观测物体的绝对亮度，人眼对亮度的变化的感知更为敏感。特别地，在一定范围内，人眼对亮度的感觉与亮度 L 的对数线性相关。

对比度敏感表明了人眼视觉系统的频率响应特性，代表了人眼对于强度值差别的区分能力。研究中通过调整正弦光栅的幅度和频率进行了一系列实验，给出了著名的 Campbell-Robson 对比度敏感函数 (Contrast Sensitive Function, CSF)，并提出 CSF 可以视为一个带通滤波过程。

掩盖效应主要指由于某个掩盖因子的出现，图像要素可见性的损失情况。可以形容为当存在激励 A 的情况下，激励 B 的感知将会得到加强或减弱。一些研究指出视觉处理是多通道并行的。即不同的视觉信息通过不同的神经通道预处理后作为视觉中枢的输入。之后被不同类型的皮层细胞处理。例如在初级视觉中枢中，大部分神经元对于具有特定频率的激励表现敏感。

视觉注意力是指我们只对周围环境中特定的场景或行为注意，使得场景中的某些特定点或区域被选为场景特征。注意力分配的过程会加强场景特征的影响并影响观测者的主观感受。

在这些特性之中，我们需要着重介绍的是对比度敏感性，以及与之相关的对比度敏感函数，因为对比度敏感函数在后续的 HDR 技术中，起到了重要的作用。

在日常生活中，人眼需要分辨边界清晰的物体，也需要分辨边界模糊的物体。后一种分辨能力则称为对比敏感度。通过研究，人们发现对比敏感度是与一些参数相关的，通过对这些参数进行建模，我们就可以得到对比度敏感函数（CSF）。

对比敏感度(CS)定义为视觉系统能觉察的对比度阈值的倒数。对比敏感度=1/对比度阈值。对比度阈值低，则对比敏感度高，则视觉功能好。在某一空间频率，视觉系统有一定的对比敏感度；反之，在同一对比度时，视觉系统有一定的空间频率分辨力(形觉)。

第二章-光电转换与 HDR 图像显示

将自然界中的真实场景转换为屏幕上显示出来的图像，往往需要经过两个主要的步骤：第一个是通过摄影设备，将外界的光信息转换为图像信息存储起来，本质上是存储为数字信号；第二个是通过显示设备，将图像信息转换为屏幕输出的光信息。下图展示了从拍摄到现实的电视信号链。在整个过程中，信息流要经过两个重要的非线性映射，才能形成我们在显示设备上看到的图像。其中，相机的非线性映射通常称为光电转换函数（OETF），而显示端的显示器的非线性映射通常称为电光转换函数（EOTF）。通常，OETF 和 EOTF 并不是互逆关系。显示端的 EOTF 往往还包括代表创作者渲染意图的光光转换函数（OOTF），其对观看者在较暗环境下在显示器上观看带来的心理视觉影响进行了修正。

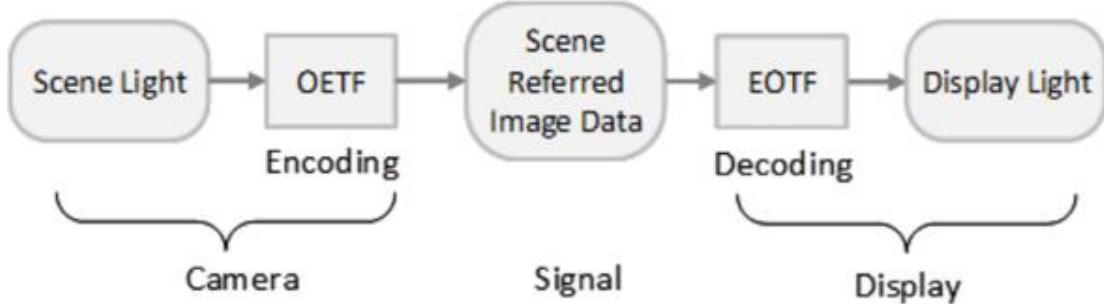


图 2-1：视频制作与显示流程

在图像获取，存储与显示操作的发展过程中，前面提到的光电转换函数与电光转换函数也经历了多次的发展。

在整个发展过程中，最具有指导意义的是 Gamma 校正，这是一种在传统的 LDR 显示设备上被广泛使用的转换函数。这一章，我们将从 Gamma 校正开始，通过对 Gamma 校正的介绍，解释清楚光电转换与电光转换函数的理论依据与在实际场景中是如何应用的。最后，我们再把目光转移到 HDR 领域上，来介绍在 HDR 领域中，如何设计新的光电转换函数与电光转换函数，来实现 HDR 图像的显示。

2.1 Gamma 校正

在最前面必须要强调的是，视觉是人类的主观感受，因此，人眼看到的，感受到的，和自然界中的实际情况与场景可能存在误差。一个最简单的例子就是人眼感受到的，均匀变化的亮度，在自然界中的实际亮度却并不是均匀变化的。这种视觉特性，就是 Gamma 校正产生的原因。人们需要对自然界的亮度进行非线性的映射，才能使人类感受到自然界中的实际景象。

2.1.1 韦伯定律

在详细介绍 Gamma 校正之前，需要先简单了解一个概念：韦伯定律。韦伯定律实际上是一个心理学上的定律。

韦伯定律是用来描述外界刺激的改变与人类感知的改变之间关系的定律，它的公式如下：

$$\Delta\Phi/\Phi = K$$

其中 $\Delta\Phi$ 表示此时的差别阈值，是指人类能够感觉到外界刺激出现改变所需的最小改变量。 Φ 表示原刺激量，K为常数。

简而言之，韦伯定律告诉我们，差别阈值与原刺激量成正比关系。也就是说，外界刺激本身越大，差别阈值就越大。

仔细分析一下韦伯定律能够从中发现，人类对于外界刺激的感知并不是线性的。并非当外界刺激强度为1时，人类感觉到的强度是1；外界刺激强度为2时，人类感觉到的强度为2；外界刺激强度为3时，人类感觉到的强度就为3。实际上，人类的感知与外界的实际刺激之间是一种非线性的关系。从韦伯定律的式子中我们可以发现，当原刺激越大时，由于比例关系，差别阈值也就越大。也就是说，当外界的刺激强度越大时，想让人类能够感知到外界的刺激发生了变化的所需变化量，就越大。所以人类的感知从1到2再到3时，所需的变化量是倍数关系的增加，可能外界刺激强度为1时，人类感知为1，刺激强度为3时，人类感知为2，刺激强度为7时，人类感知到的强度才为3。

举一个现实中的例子就能够更好的理解。在一个黑暗的房间中，点燃一根蜡烛，将让人感知到巨大的亮度改变。而如果房间中有99根蜡烛，再点燃一根蜡烛人类可能根本感觉不到有所差别。所以，如果想要让人类感知到同样程度的亮度改变，所需的蜡烛数量可能是成百上千根才行。

2.1.2 Gamma 校正

根据韦伯定律，我们能够知道人类对于外界刺激的感知是非线性的，人眼对于亮度的感知由此可知，也是非线性的。下面，我们首先介绍人眼对于亮度的实际感受情况，然后以此为基础，介绍Gamma校正的相关内容。

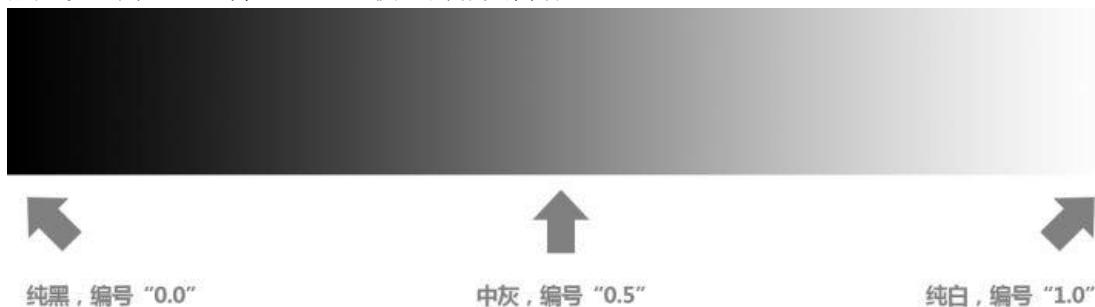
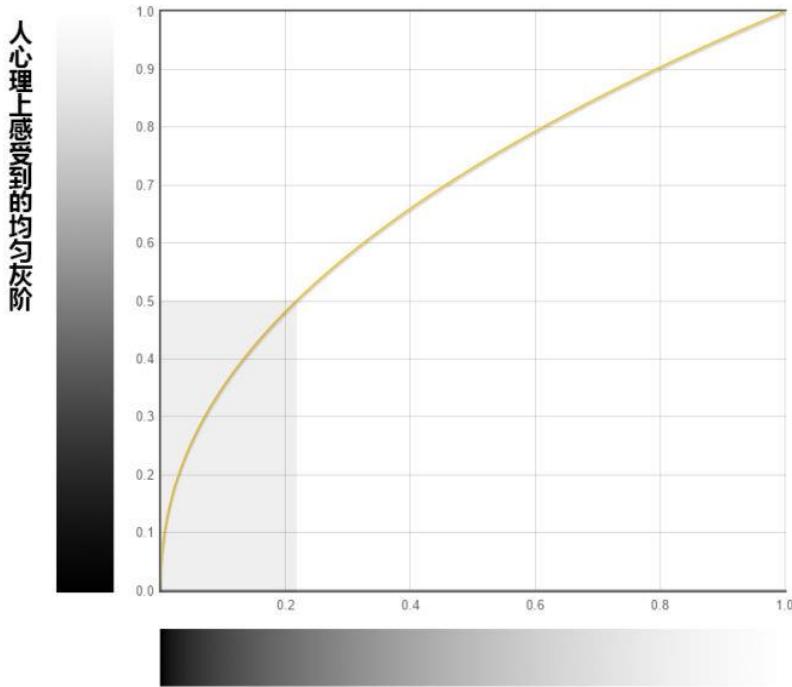


图 2-2：人眼感受到的亮度

经过一系列实验的测定，人们发现人类能够感知到的连续亮度谱如上图所示。也就是说，如果比上图更加精细，则没有必要，因为人眼无法识别；而如果比上图粗糙，则人眼能够感觉到间断。接着，将最左边纯黑编号为“0.0”，最右端纯白编号为“1.0”，中间定义为中灰“0.5”。这就形成了人眼的亮度谱。

但是，上图中的亮度在自然界中并不是同样成比例的，例如图中的中灰，对应自然界中的亮度是纯白的20%。

将人眼的亮度谱与自然界中的实际亮度进行建模，就可以得到下面的图像：



自然界线性增长的亮度

图 2-3：自然界与人眼感知的亮度对应

从上图可以发现，纵轴上是人眼感受的均匀灰阶，而横轴是自然界中线性增长的亮度，它们之间是一种非线性映射的关系。在低亮度区域，人眼的敏感性较高，因此自然界中很小的亮度范围，都能够被人眼察觉；在高亮度范围，人眼的敏感度较低，因此自然界中的大范围的亮度变化，在人眼感知上却并不大。通过经验以及实验，人们发现可以使用幂函数来拟合这种非线性的映射，用公式表示如下：

$$Y = X^\gamma$$

其中 X 是使用 0-1 之间的实数表示的输入亮度， Y 是使用 0-1 之间实数表示的输出亮度， γ 为认为设定的参数，不同的 γ 选择会产生不同的效果。这种通过调整 γ 的取值，来调整输入输出亮度之间映射关系的方式，我们称之为 gamma 校正。

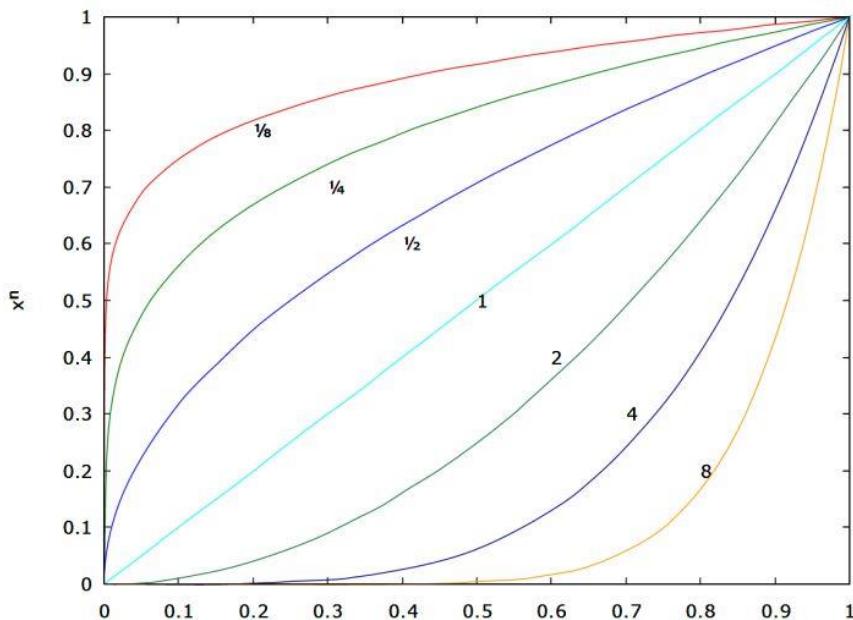


图 2-4: 不同 γ 取值下, 幂函数的图像

根据 γ 取值的不同, 通过上图, 我们能够发现 gamma 校正的不同效果:

1. 当 $\gamma < 1$ 时, 曲线上凸, 输出值大于输入值。
2. 当 $\gamma = 1$ 时, 曲线为一条直线, 输入与输出相等, 不进行校正
3. 当 $\gamma > 1$ 时, 曲线下凸, 输出值小于输入值

2.1.3 Gamma 校正与图像显示

在传统的图像获取与显示过程中, 将自然界中的真实场景转换为屏幕上显示出来的图像, 往往需要经过两个主要的步骤: 第一个是从摄影设备, 将外界的光信息转换为三维数组存储起来, 也就是存储为图像信息, 这一过程如前所述, 需要通过光电转换函数进行非线性映射; 第二个是通过显示设备, 将图像信息转换为屏幕输出的光信息, 同样, 这一过程需要电光转换函数的参与。在这两个阶段中, 光电转换函数与电光转换函数实际上都是幂函数, 也就是都进行 Gamma 校正。

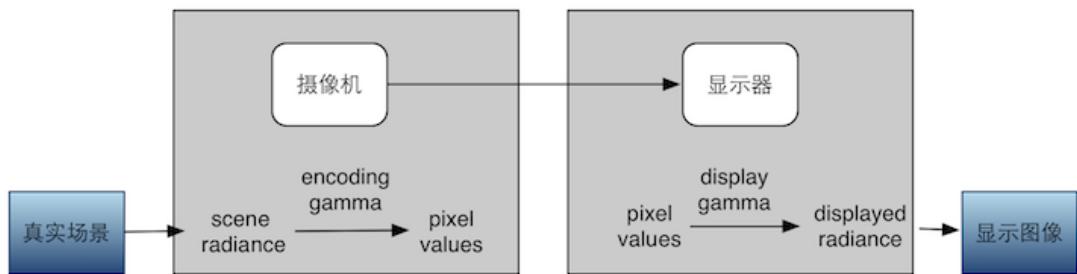


图 2-5: 真实场景到显示图像的两个转换阶段

首先, 因为图像中的每个像素的取值在 0-255 之间, 因此先进行归一化, 将 0-255 这 256 个数线性映射到 0-1 之间。

接着, 在图像获取阶段, 摄像机中会使用一种 encoding gamma 函数进行 gamma 校正, 这里的 gamma 值小于 1。自然界中的亮度, 首先在摄像机中进行归一化, 映射到 0-1 之间, 接着, 将这个经过了归一化的值输入 encoding gamma 函数中。最后将 encoding gamma 函数的输出值与 0-255 的像素值归一化结果进行比较, 就能得到这个像素点的像素值。

显示阶段的 gamma 校正通常使用一种 display gamma 函数来完成, 这里的 gamma 值大于 1。显示设备读取像素值后, 进行归一化, 然后将归一化结果输入 display gamma 函数中。将输出的值对应的自然光线进行输出, 就是显示阶段做的工作。

理论上来说, 上述两个阶段的 gamma 值乘积应该为 1。因为两个 gamma 校正实际上做的是恰好相反的操作, 一个是将自然光线存储为数字信息, 另一个是将数字信息转换为光线。而为了使自然中的光线与显示设备产生的光线一致, 显然这里的 gamma 值的乘积需要为 1。但是, 在实际操作中, 这两个 gamma 值的乘积并不为 1, 原因有以下两点: 1) 对于传统显示设备来说, 屏幕能够显示的亮度范围与颜色范围与实际情况相差好几个数量级; 2) 在真实的场景中, 原始的场景填充了观察者的所有视野, 而显示设备的亮度往往只局限在一个被周围环境包围的屏幕上。这两个差别使得感知对比度相较于原始场景明显下降了。

为了中和这种现象, 所以我们需要的保证显示的亮度结果在感知上和原始场景是一致的两个 gamma 的乘积并不是 1, 而是经过试验得到的其他值。在电影院这种漆黑的环

境中，这个乘积为 1.5，在明亮的室内这个值为 1.125。个人电脑使用的一个标准叫 sRGB，它使用的 encoding gamma 大约是 0.45（也就是 $1/2.2$ ）。这个值就是为了配合 display gamma 为 2.5 的设备工作的。这样，end-to-end gamma 就是 $0.45 * 2.5 = 1.125$ 了。

2.1.4 Gamma 校正的意义

按照上一节所说，两段 gamma 校正实际上是一对逆过程，一个将光信息转化为数字信息，一个将数字信息转化为光信息。因此，按照常理来说，并不需要进行这样复杂的非线性映射，只需要将归一化之后的光信息记录下来，然后再让显示设备进行输出即可。从这种角度来说，gamma 校正是一个不必要的操作。

这是一个合理的质疑。问题的根源是在存储的位宽上，如果存储位宽足够大，显然可以使用上述归一化记录方式进行处理，但是，实际生产与应用中的位宽是有限的。在传统的图像存储与显示标准中，一个像素点有三个通道，而每个通道只有 8 位。这也就意味着，图像的灰阶极其有限，与真实世界相比，差了很多个数量级。因此，想要在有限的灰阶上表现出近似于自然界的亮度，就需要对灰阶的分配进行设计，这也是 Gamma 校正的意义。

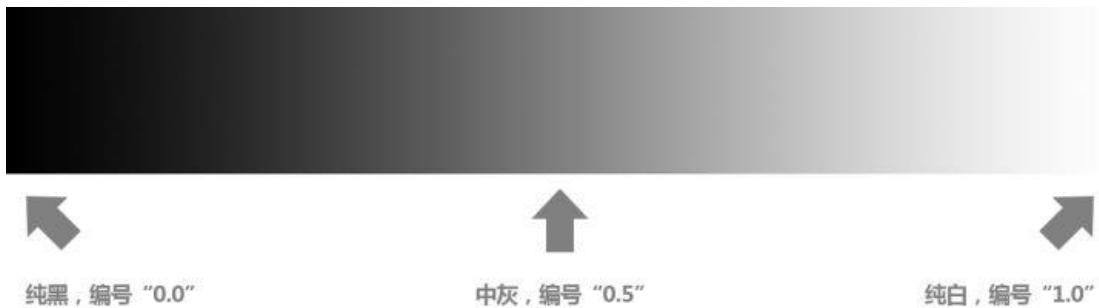


图 2-6：人眼感受到的亮度

由于人类感知的“中灰”，在实际的自然光线中对应于白光的 20% 的亮度。因此，如果完全按照自然光线的数值进行灰度处理，纯黑到中灰这一段在人类感知中占到一半的灰度，将只能由 20% 的灰阶进行表示，而剩下的一半会有 80% 的灰阶进行表示，这显然是不合理的。Gamma 校正就可以解决这个问题。

在实际操作中，中灰对应的自然光线，显然会被映射为 0.5 的取值，也就对应 128 这一灰阶。这样的结果就是高亮度与低亮度部分分别占有一半的灰阶，这就会使得显示的图像的直观感受更接近于现实。

因此，gamma 校正实际上是用来解决显示设备亮度灰阶不足的问题的。如果将图像存储的编码位宽扩大到 32 位，也就是能完全包含自然界亮度范围的时候，自然就不再需要进行 gamma 校正了，这种情况下，只需要将自然光线的亮度值记录下来，然后对应输出即可。但是，使用 32 位位宽来存储图像过于“奢侈”，因此 gamma 校正在实际生活中起到了广泛而巨大的作用。

2.2 HDR 技术中的光电转换

前面提到的传统的 Gamma 校正方式，是针对传统 CRT 显示设备与 LDR 图像设计的，对应的标准是 BT. 1886 标准。而 BT. 1886 标准规定的峰值亮度仅为 100nits，它与传

统 CRT 设备的显示能力相一致。目前的大多数显示设备的显示能力早已突破 100nits 的限制，普通的消费级显示器的亮度一般在 200nits 至 500nits，商业级的监视器亮度可以达到 1000nits 至 2000nits，实验室级别的监视器甚至可以高达 4000nits 至 20000nits。

同时，现代数字运动图像传感器能够捕获动态范围到 80dB 或者比特深度达到 14bit 的线性视频信号。由于实践和历史原因，大多数专业视频制作系统都采用 10 比特，而消费者级别的视频和计算机图形学通常采用 8 比特。在 HDR 技术中，通常采用 10 比特或 12 比特。

由于显示设备的亮度范围得到了提升，同时图像的编码深度也得到了提升，因此传统的 Gamma 校正显然已经不适用于 HDR 技术的光电转换过程。如果强行使用 Gamma 校正进行转换，会在图像与视频中出现明显的分块与分环的情况。

针对上述问题，为了提高显示效果，不同的组织采用了不同的方式进行处理，其中最重要的两种光电转换曲线是：PQ 曲线以及 HLG 曲线。

2.2.1 PQ 曲线

为了研究人类视觉系统及喜好，杜比进行了一系列的用户实验，以确定新的 HDR 技术标准以及新的光电转换曲线。

第一个实验室确定新的显示亮度范围。该实验使用一个 2K 的数字电影放映机投影到一个 21 英寸的单色 LCD 面板上，形成一个高对比度的显示器，峰值亮度可以达到 20000 尼特，最小黑电平可以达到 0.004 尼特，色域为 DCI-P3。测试序列共有三个序列集：一个用于检测黑色电平，一个用于检测白色电平，还有一个用于检测高亮电平。实验中观察者处在光线较暗的影院环境下（暗室），位于距离显示设备屏幕高度三倍远的地方。实验结果如下所示，根据实验结果，在亮度范围为 0–10000nits 时，各种信号水平的用户喜好度均超过了 84%，因此 Dolby 提出的 PQ EOTF 曲线决定使用 0–10000nits 作为支持的显示亮度范围。

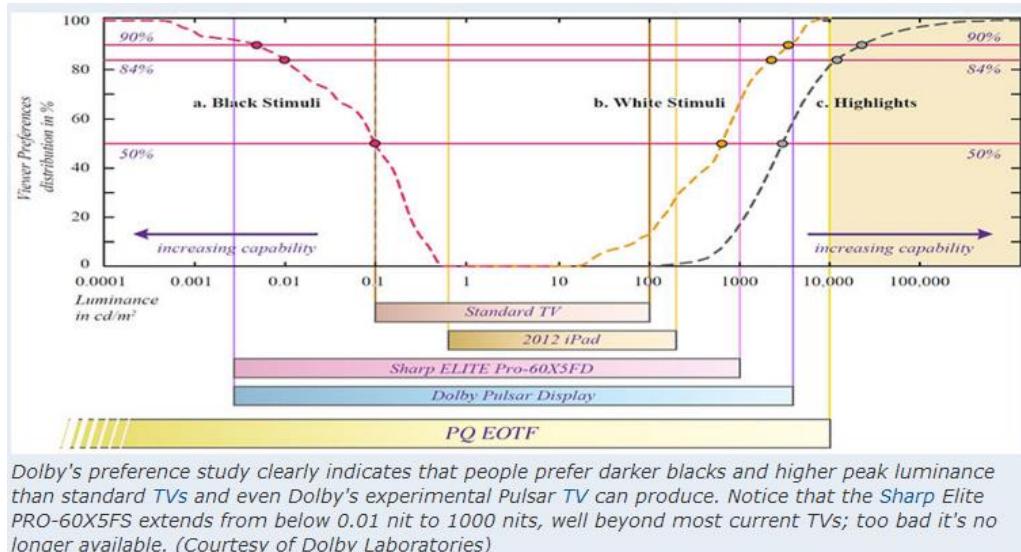


图 2-7：杜比显示亮度实验结果

由于 8 比特的位深限制了使用更广泛的动态范围，因此需要定义新的 OETF 曲线为视频制作和传输提供更高的动态范围。Dolby 实验室提出了新的 OETF 曲线，称为 PQ 曲

线，目前已经被标准化为 SMPTE ST 2084。

Dolby 提出的 PQ EOTF 的表达式为，它基于 Barten 的 CSF 模型：

$$Y = L \left(\frac{V^{1/m} - c_1}{c_2 - c_3 V^{1/m}} \right)^{1/n}$$

其中： $0 \leq V \leq 1$, $L = 1000$, $m = \frac{2523}{4096} \times 128 = 78.8438$, $n = \frac{2610}{4096} \times \frac{1}{4} = 0.1593$,

$$c_1 = \frac{3424}{4096} = 0.8359, c_2 = \frac{2413}{4096} \times 32 = 18.8516, c_3 = \frac{2392}{4096} \times 32 = 18.6875.$$

10 bit PQ and Rec1886 Curves

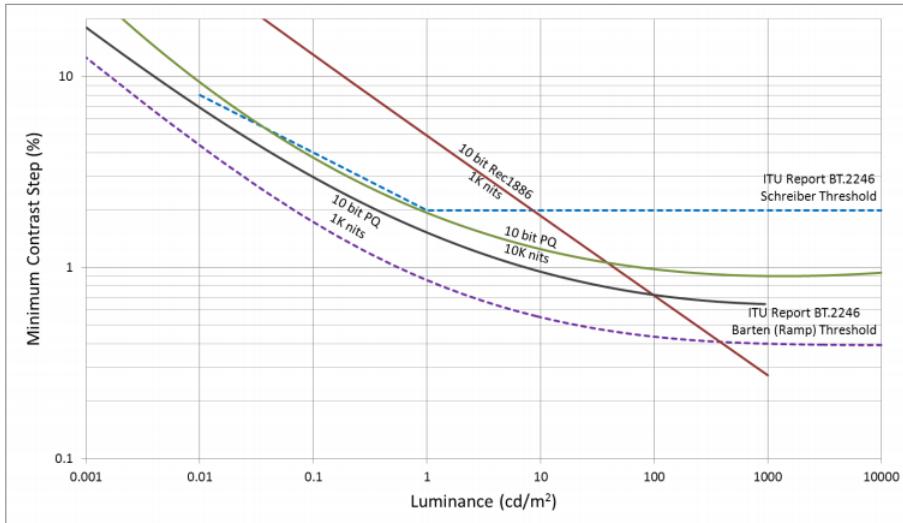


图 2-8：10 bit PQ 曲线与 Gamma 曲线对比图

PQ 曲线的优势

Barten Ramp 作为一般图像的共同阈值，大量实验结果显示，Barten 斜坡曲线规定了不同亮度水平下，不出现可见量化误差的最小对比度步长的上界，也就是允许的最大量化步长。当在这条曲线之下时，编码图像可以实现平滑渐变；而当超过这个临界值，也就是在 Barten 斜坡曲线之上时，会出现人眼可察觉的条带效应等量化失真（banding/contouring）。Barten Ramp 曲线表明，在低亮度区域，人眼的视觉灵敏度较低，因此可以分配较少的比特数进行编码；而人眼的视觉灵敏度在高亮度区域非常灵敏，因此编码时需要分配更多的比特数。新的 EOTF 的选择不仅要满足上述特征，还要尽可能的拟合或者贴近 Barten 曲线。这样不仅能够满足人眼视觉特性需求，不会造成量化失真，同时也能够有效利用有限的编码比特数。

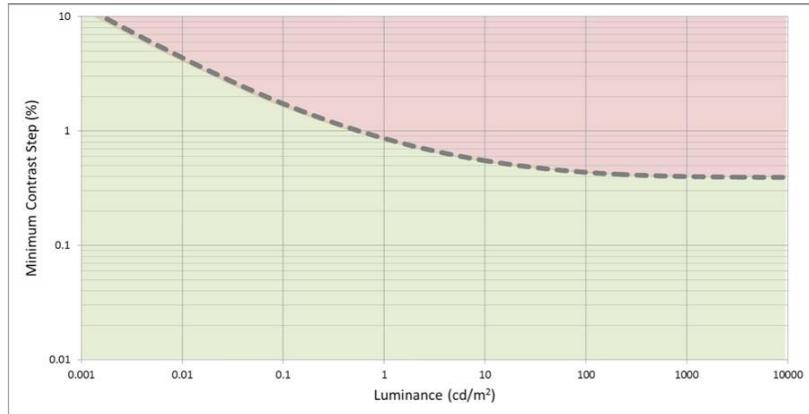


图 2-9: Barten Ramp 量化误差 当曲线位于虚线上方时, 会产生量化误差, 处在下方时, 不会产生误差。

为使用尽可能少的比特数表示尽可能多的信息, 必须要确定不造成量化误差的最小比特数。如下图所示, 当使用量化精度为 10 比特, 值为 2.4 的 gamma 曲线时, 其低于 10nits 的亮度区域均处于 Barten Ramp 曲线之上, 会造成轻微的量化误差, 并且其峰值亮度最大只能为 100nits。因此, 新的 EOTF 曲线不能使用 10 比特的 gamma 曲线。当使用量化精度为 15nits 的 gamma 曲线时, 整条曲线都在 Barten Ramp 之下, 不会造成量化误差, 峰值亮度可以达到 10000nits。但需要面临的另一个问题是, 15 比特的 gamma 曲线会浪费太多的比特数在高亮度区域。除此之外, 目前的视频传输和处理系统绝大多数都是基于 10 比特进行操作的, 采用 15bitsde gamma 编码不仅带来传输压力, 还需要对市场上大部分的视频传输和制作系统进行更新换代。

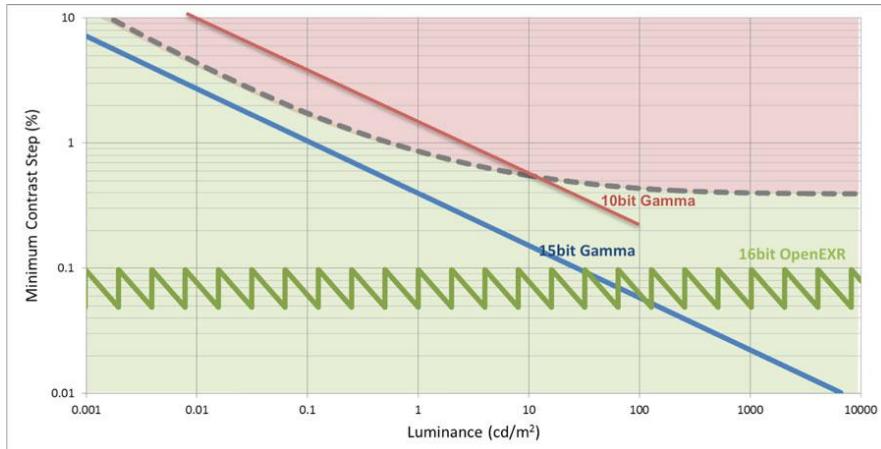


图 2-10: 曲线对比图 (1)

如下图所示, 如果使用 13 比特的对数曲线时, 与 gamma 曲线相反, 又会在低亮度区域分配更多的比特数。同时可以看到, 用于 HDR 图片存储的 16 比特的 OpenEXR 对应曲线在 Barten 阈值之下, 不会造成可见量化误差。当采用 12 比特的 PQ 曲线作为新的 EOTF 时, 由下图可以看出, 其很好地拟合了 Barten Ramp, 因此不仅符合了人眼视觉系统地特性, 而且充分的利用有限的比特数进行高效合理的编码。

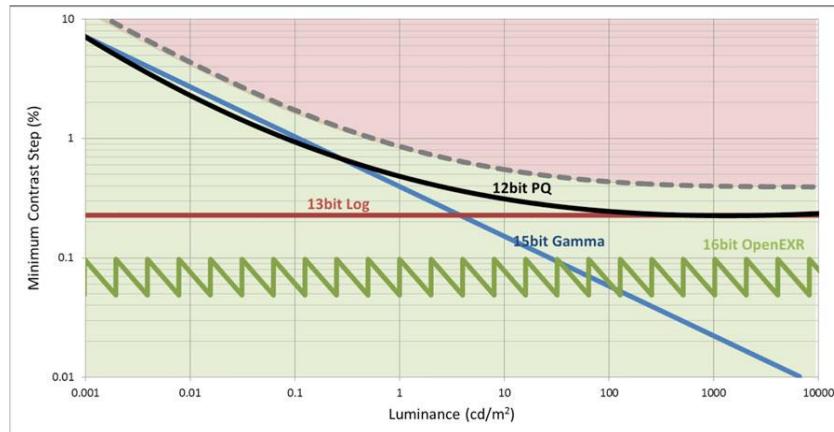
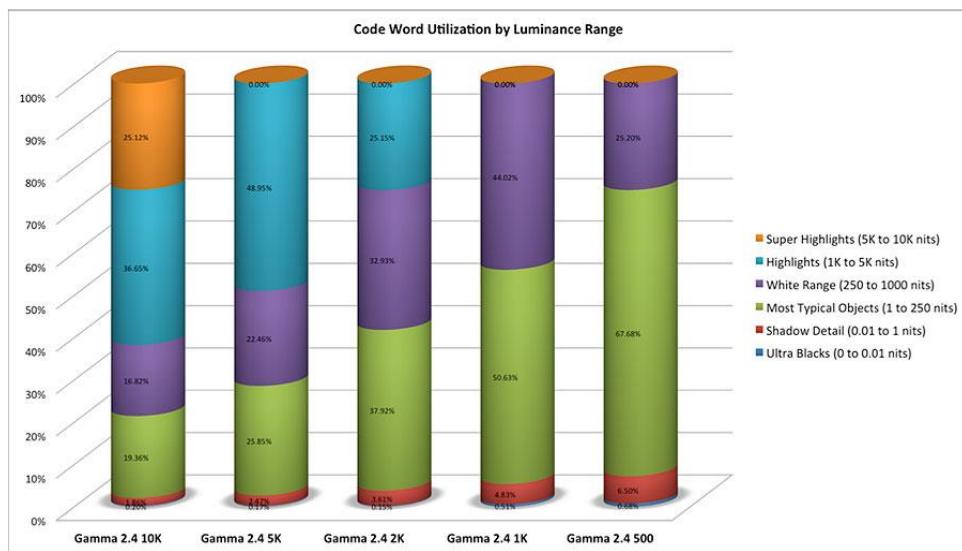


图 2-11：曲线对比图（2）

下边两张图展示了不同亮度水平下分别采用 Gamma 曲线和 PQ EOTF 进行编码的编码字的利用率。纵轴使用百分比作为度量，100% 对应的峰值亮度为 10000nits。Gamma 2.4 编码在图像亮度范围为 0–0.01nits 的部分几乎不分配编码字，图像阴影部分分配的编码字也非常少。并且，随着峰值亮度的增加，其在大多数物体所在的亮度范围内以及白光范围内分配的编码字逐渐减少；而使用 PQ 编码，不同峰值亮度水平下很暗的部分（0–0.01nits）以及阴影部分的细节处都分配了一定的编码字。但是在峰值亮度变化时，亮度范围 1–1000nits 的范围分配的编码字差值很小。



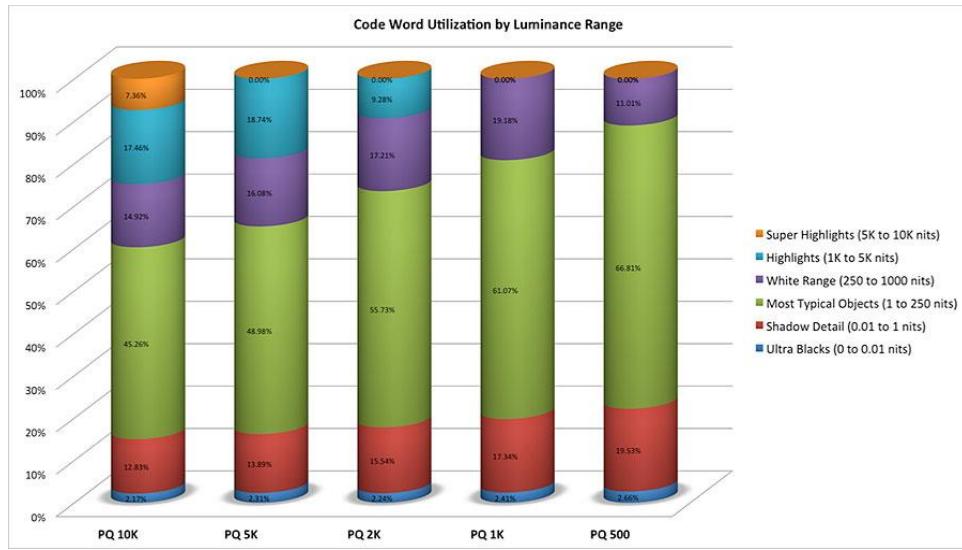


图 2-12: PQ 曲线与 Gamma 曲线的码字利用率对比

2.2.2 HLG 曲线

另外一种与 PQ 相对应的适用于 HDR 的光电转换函数是 HLG (Hybrid-Log Gamma)。与 PQ 的不同在于，PQ 是一种绝对的映射，亮度值与编码是绝对的一一映射，而 HLG 是一种相对的映射。

为了增加高光部分的对比度，通常会对 Rec. 709 OETF 曲线加入一个拐点，如下图所示。当拐点位置在 85% 的参考白电平，最大曝光可以改善为原来的 5 倍。这种方法虽然改变了最大曝光，但是由于拐点位置并没有标准化过，而且对相机调校的要求很高，所以这种方法实现起来还是有一定困难。另外，Rec. 709 标准中的 gamma 曲线主要是使噪声在不同的亮度水平下表现一致。

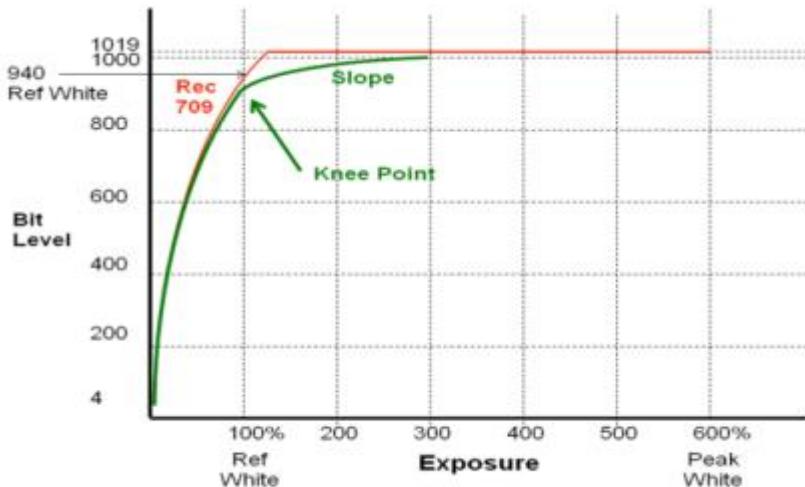


Figure 2: Example of Knee Transfer Characteristic

图 2-13: 拐点图

根据韦伯定律，图像中的高亮区域的感知量化的阈值基本是恒定的一个常数。因此，在图像的高亮区域使用对数形式的 OETF 能够利用给定的比特数获得最大的动态范围。

而在图像的暗部，如德维斯-罗斯定律所定义的，当亮度逐渐变小时，可见量化误差会越来越明显。因此，低亮度区域部分的 OETF 可以考虑使用 gamma 曲线，因为 gamma 曲线非常符合德维斯-罗斯定律的约束。综上所述，当动态范围增加时，一个理想的 OETF 应当由两部分组成：低光部分使用传统的 gamma 曲线；而高亮区域使用对数形式的 OETF。这样就形成了新的可用于更高动态范围的 HLG OETF 曲线

$$\text{Weber Fraction} = \frac{1}{N \cdot L} \frac{dL}{dV} = \frac{\gamma}{N} \frac{1}{L^{1/\gamma}}$$

其中 L 为显示亮度， V 为信号电平值， N 为亮化水平（对于 8 比特的视频 $N=220$ ；对于 10 比特的视频 $N=876$ ）

HLG (Hybrid-Log Gamma)，是由 BBC 和 NHK 提出的一种新的标准，规定了输入端的光电转换特性 (OETF)——Hybrid-Log Gamma 曲线，目前已被日本电波产业协会 (ARIB) 采纳为 ARIB STD-B67 标准。HLG 曲线分为两段，在线性光亮度较小时，其与 SDR 的 gamma 曲线重合，当亮度较高时，其曲线呈现对数特性。在输出端，EOTF 定义为加入了系统 gamma 和亮度调整的 OETF 的逆。支持 10 比特量化和最大亮度 4000nit。

HLG 曲线的数学表达式如下：

$$E' = \begin{cases} r\sqrt{E} & 0 \leq E \leq 1 \\ a\ln(E - b) + c & 1 < E \end{cases}$$

其中， E' 是电平信号， E 表征光照强度。变量 r 为参考电平，一般取 0.5。 a 、 b 、 c 均为常量，其值分别为 0.17883277、0.28466892、55991073。

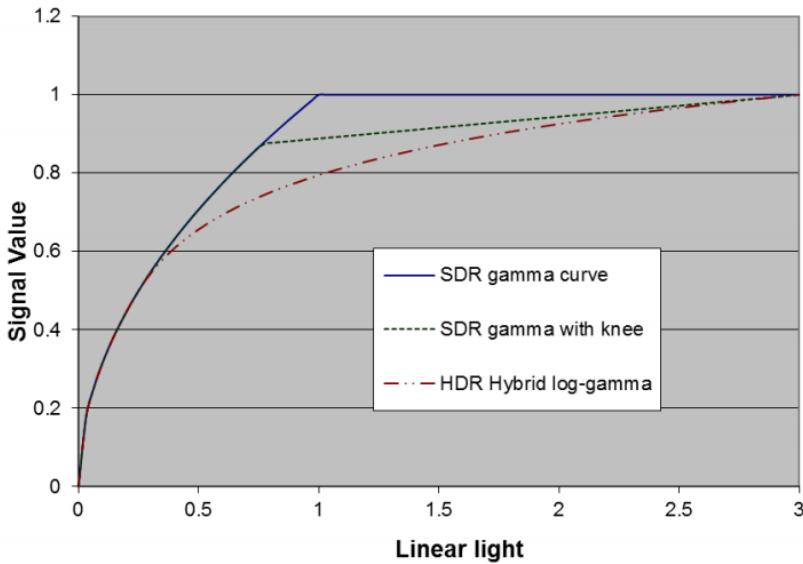


图 2-14：HLG 曲线与传统 SDR-Gamma 曲线对比 (BBC)

在拍摄端使用 HLG OETF，那么在显示端需要定义对应的 EOTF 才能更好的再现场景。为了很好的适配 HLG OETF，相应的 EOTF 应当满足以下条件：(1) 保留创作者的渲染意图；(2) 能够适配从黑电平到峰值白电平的整个动态范围；(3) 减小量化误差。因此，相应的 EOTF 定义为：

$$Y_d = \alpha Y_s^\gamma + \beta$$

其中 Y_d 为显示面板的显示亮度； Y_s 表示场景的相对亮度； γ 表示系统 gamma 值，

取值大小与观看环境相关；而参数 α 和 β 分别控制对比度和亮度，它们共同控制着显

示的峰值亮度和黑电平水平。

$$E = \begin{cases} (E'/r)^2, & 0 \leq E' \leq r \\ \exp((E' - c)/a) + b, & r < E' \end{cases}$$

根据 HLG OETF 的数学模型可以看出，HLG 的低亮度部分与传统的 gamma 曲线完全吻合，因此可以不经过任何处理直接显示在现有的 SDR 显示器上，其 10 比特的信号传输制作过程与现有的 SDR 系统完美兼容。另外，由于不需要元数据，因此信号是独立于显示器的。

第三章-色度学与颜色空间

HDR 技术的第二个理论基础是色度学。从前面的内容中可以了解到，光学以及人类视觉感知模型为人类提供了解释与分析人类感知亮度的理论基础，但是 HDR 技术不仅仅关注于提升图像与视频的亮度范围，同时也关注于提供更加丰富的色彩。因此，本章将首先介绍人眼与色度学相关的生理特征以及人类对颜色的识别方式，然后介绍颜色空间的概念，最后再回到 HDR，介绍与 HDR 相关的颜色标准。

3.1 色度学

由于视觉是人类的主观感受，因此无论是亮度还是颜色的研究，最根本的研基础都在于人眼的生理特征。人眼的最内层是视网膜，它布满了眼球后部的内壁。当眼球聚焦时，外部的光线在视网膜上聚焦。视网膜表面分布许多视觉细胞，人眼靠这些细胞感知外部光线，然后传递给大脑，形成视觉感知。这些视觉细胞分为两大类：视觉锥细胞以及视觉杆细胞。其中，视觉锥细胞对颜色非常敏感，而视觉杆细胞没有颜色感觉，因此人眼靠视觉锥细胞来产生颜色感知。视觉锥细胞又可以三类，它们分别对红色光，绿色光，蓝色光敏感。但是需要额外注意的是，这里提到的对某种色光敏感，不代表这种视觉锥细胞只能感知到该中色光。也就是说，在可见光范围内的任意波长的色光，都能刺激三种视觉锥细胞，而某些特殊波长的光线能够对视觉锥细胞产生最大的刺激。

色度学研究的内容就在于如何在显示设备上产生与自然界类似的刺激，使人眼能够感受到相同的颜色。显而易见的是，最简单的方法就是对自然界的每一种颜色生产一种对应的发光原件，可以发出对应颜色的光线。但是，人眼可以识别上万中不同的颜色，而在每一个像素点设置上万个不同的发光元件基本是不可能的，而且生产上万种不同的发光元件的代价也过于高昂，因此，我们需要其他简单的，可行的方法来实现对自然界中颜色的模拟。

可以想到的最直接的，也是人们一直以来使用的解决方式就是：利用人眼中只有三种视觉锥细胞的特性，只使用红色（r），绿色（g）和蓝色（b）三种光线，来模拟所有其他的颜色。因此，这三种颜色也被称为光的三原色。

这里需要思考一个问题：以黄色为例。在自然界中，显然将红色光与绿色光叠加，得到的新的光线的频率并不是黄色光的频率，但是，红色光与绿色光的叠加，给人眼的感受就是黄色。这种现象的原因是什么？

人眼的三种锥细胞，不同波长的光以及这些光的亮度会对三种锥细胞产生不同的刺激。同样以黄色光为例子，当人眼看见黄色光时，红视锥细胞与绿视锥细胞会产生某种程度的刺激，蓝视锥细胞则基本不产生刺激。而大脑接受到这种刺激后，产生了黄色的视觉。而通过改变红色光与绿色光的亮度，也可以使红视锥细胞与绿视锥细胞产生相同程度的刺激，而大脑接收到这样的刺激之后，同样会认为这是黄色，而非绿色与红色的混合色。

也就是说，人眼对颜色的感知，实际上是通过色光及其亮度，对三种视觉锥细胞产生刺激，然后神经系统将不同的刺激，映射为人感知中的不同的颜色。因此，模拟自然界中的颜色，实际上就是模拟它们对三种视觉锥细胞的刺激。

为了实现使用三种原色光线生成其他颜色光线的目的，CIE（国际照明委员会）在 1931 年组织了一项实验，通过一些观测者的观察，来确定生成不同的颜色，所需的三原色光线的强度。他们将三种波长的光：700nm（红色光），546.1nm（绿色光），435.8nm

(蓝色光)作为基色。在实验过程中,首先将三种基色的光投影到白色屏幕的一侧,接着,再讲第四束光-待测光投影到白色屏幕的另一侧。调节三种基色光的强度直到观测者感受到两侧的颜色一致时,就能够确定如何使用三种基色光得到各种颜色的光。

实验的最终结果是得到了三个颜色匹配函数 $\bar{r}(\lambda)$, $\bar{g}(\lambda)$, $\bar{b}(\lambda)$ 。这三个函数分别表示产生波长为 λ 的光,所需的红,绿,蓝三原色光的强度。

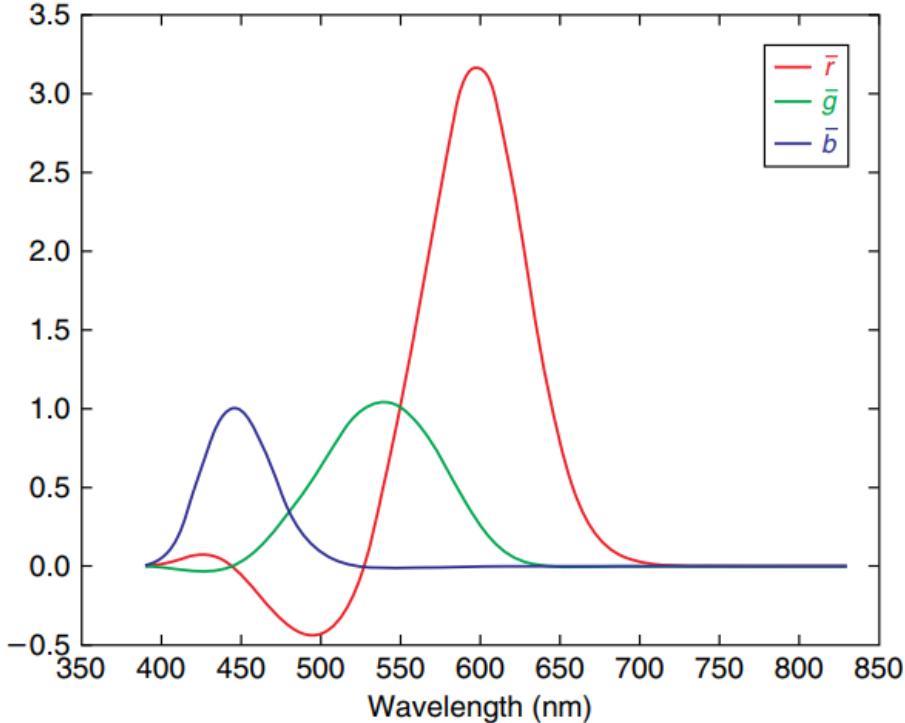


图 3-1 三种颜色匹配函数曲线

实验结果的图像如上图所示。对上图中的单位做如下规定:当三原色光配比出等能白光时,三原色光各自的量称为一单位的红光[R],一单位的绿光[G]与一单位的蓝光[B]。此时三个原色光的通光量之比为 1:4.5907:0.0601。此时,任意波长的光可以通过三原色光的线性组合得到:

$$C(\lambda) = \bar{r}(\lambda)[R] + \bar{g}(\lambda)[G] + \bar{b}(\lambda)[B] \quad (1)$$

这就是 1931 CIE-RGB 计色系统。对于任意单色光,理论上可以全部使用 RGB 三色光表示出来。但是由于 435.8 nm–546.1 nm 这一段的红色光通量为负值,因此在实际生产中无法实现,所以实际上并非所有单色光(只有一个频率,不会发生色散)都可以由 RGB 三原色表示出来。这里出现负值的原因是,待配色光为单色光,其饱和度很高,而 RGB 三色混合后,饱和度会降低。因此,想要配出一些单色光,需要将 RGB 三色光中的某一个与待配色光混合,才能实现。此时就会出现负值。

而对于非单色光(能产生色散的色光),任意光谱分布 $\Phi_e(\lambda)$ 的 RGB 值都可以由积分得出。

$$\begin{aligned} R &= \int_{380}^{780} \Phi_e(\lambda) \bar{r}(\lambda) d\lambda \\ G &= \int_{380}^{780} \Phi_e(\lambda) \bar{g}(\lambda) d\lambda \\ B &= \int_{380}^{780} \Phi_e(\lambda) \bar{b}(\lambda) d\lambda \end{aligned} \quad (2)$$

通过这种方法就可以将单色光与非单色光用 RGB 三原色表示出来。但是,此时出现

的另一个问题是，人类对于颜色的感知包括色度与亮度两部分。例如，绿色会因为亮度的不同，带给人亮绿色，绿色，暗绿色等不同的感受，但是从色度角度来说，它们都是绿色，差别仅仅体现在亮度上。因此，想要表示准确的色度信息，就需要消除亮度带来的影响。通过对颜色匹配函数归一化，可以帮助消除亮度带来的影响。

归一化的结果如下：

$$\begin{aligned} r(\lambda) &= \frac{\bar{r}(\lambda)}{\bar{r}(\lambda) + \bar{g}(\lambda) + \bar{b}(\lambda)} \\ g(\lambda) &= \frac{\bar{g}(\lambda)}{\bar{r}(\lambda) + \bar{g}(\lambda) + \bar{b}(\lambda)} \\ b(\lambda) &= \frac{\bar{b}(\lambda)}{\bar{r}(\lambda) + \bar{g}(\lambda) + \bar{b}(\lambda)} \end{aligned} \quad (3)$$

由上面的三个式子我们可以发现， $b(\lambda) = 1 - r(\lambda) - g(\lambda)$ ，因此，使用 $r(\lambda)$ 与 $g(\lambda)$ 就可以表示整个色域。

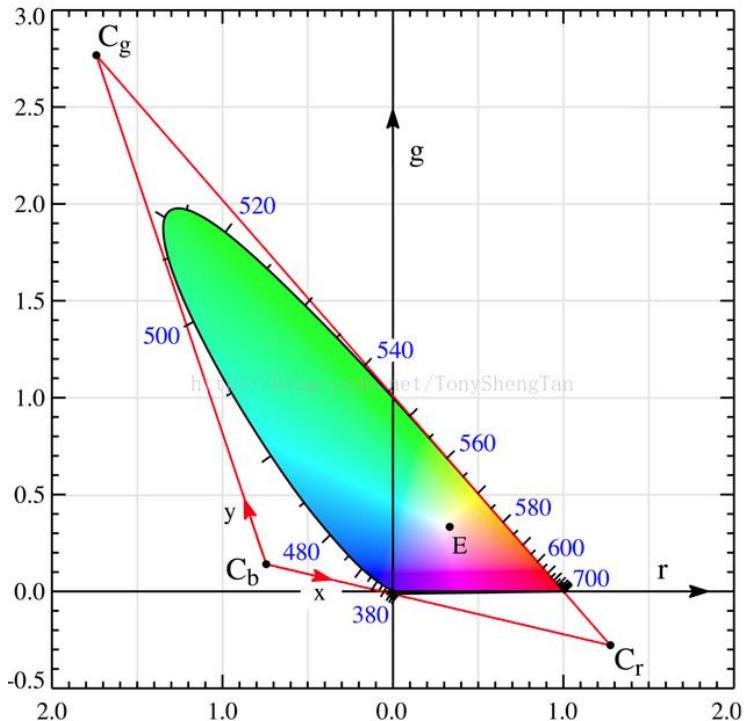


图 3-2：使用 r, g 表示的色域

理论上来说，通过上述的函数，我们已经可以通过三原色光表示所有颜色了，但是从图中我们可以发现，在系数的取值中存在负值。例如上图中 r 轴在负数部分也有取值。

这会在实际的工业生产中带来不便，因为负值的光强是无法制造出来的。因此，CIE 对此进行了修改，提出了 1931 CIE-XYZ 计色系统。目的是使得系数不再出现负值。

为了实现系数不再出现负值的目的，CIE 制定了一系列规则，通过坐标变换的方式，将 RGB 表示法转变成了 XYZ 表示法。简单来说，可以将 XYZ 表示法看做 RGB 表示法的一种非线性变换，这种非线性变换的结果是系数不再出现负值。此时的 XYZ 与 RGB 这种自然界中存在的红光，绿光，蓝光不同，它的人为定义出的具有特殊匹配函数的三种“基色”，它的匹配函数 $\bar{x}(\lambda)$, $\bar{y}(\lambda)$, $\bar{z}(\lambda)$ 如下图所示：

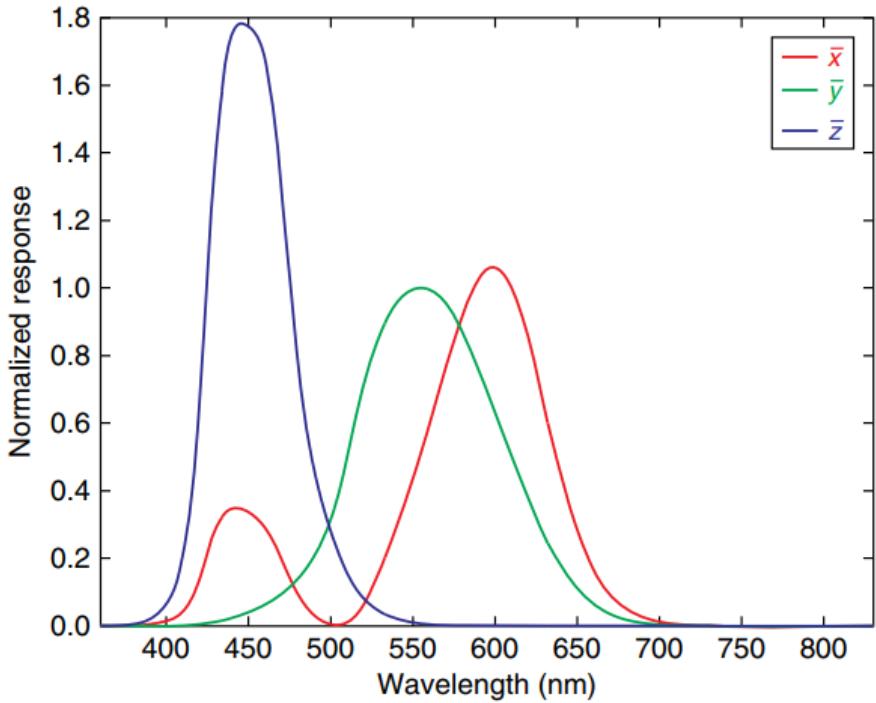


图 3-3. CIE-XYZ 颜色匹配函数

这里我们不需要尝试将 XYZ 与自然界中的某种波长的光线对应，这是没有必要的，我们只需要知道它们是具有如上图所示的匹配函数，并且是由 RGB 通过非线性变换得到的新的“基色”即可。

与 RGB 色度表示法类似，将 $\bar{x}(\lambda)$, $\bar{y}(\lambda)$, $\bar{z}(\lambda)$ 进行归一化：

$$\begin{aligned} x(\lambda) &= \frac{\bar{x}(\lambda)}{\bar{x}(\lambda) + \bar{y}(\lambda) + \bar{z}(\lambda)} \\ y(\lambda) &= \frac{\bar{y}(\lambda)}{\bar{x}(\lambda) + \bar{y}(\lambda) + \bar{z}(\lambda)} \\ z(\lambda) &= \frac{\bar{z}(\lambda)}{\bar{x}(\lambda) + \bar{y}(\lambda) + \bar{z}(\lambda)} \end{aligned} \quad (4)$$

归一化后， $x(\lambda)$, $y(\lambda)$, $z(\lambda)$ 与 $r(\lambda)$, $g(\lambda)$, $b(\lambda)$ 的关系为：

$$\begin{aligned} x(\lambda) &= \frac{0.49000r(\lambda) + 0.31000g(\lambda) + 0.20000b(\lambda)}{0.66697r(\lambda) + 1.13240g(\lambda) + 1.20063b(\lambda)} \\ y(\lambda) &= \frac{0.17697r(\lambda) + 0.81240g(\lambda) + 0.01063b(\lambda)}{0.66697r(\lambda) + 1.13240g(\lambda) + 1.20063b(\lambda)} \\ z(\lambda) &= \frac{0.00000r(\lambda) + 0.01000g(\lambda) + 0.99000b(\lambda)}{0.66697r(\lambda) + 1.13240g(\lambda) + 1.20063b(\lambda)} \end{aligned} \quad (5)$$

我们也可以发现 $z(\lambda) = 1 - x(\lambda) - y(\lambda)$ 。因此，XYZ 表示法也可以只是用 X 与 Y 表示整个色域，XYZ 表示法下的色域图如下图所示：

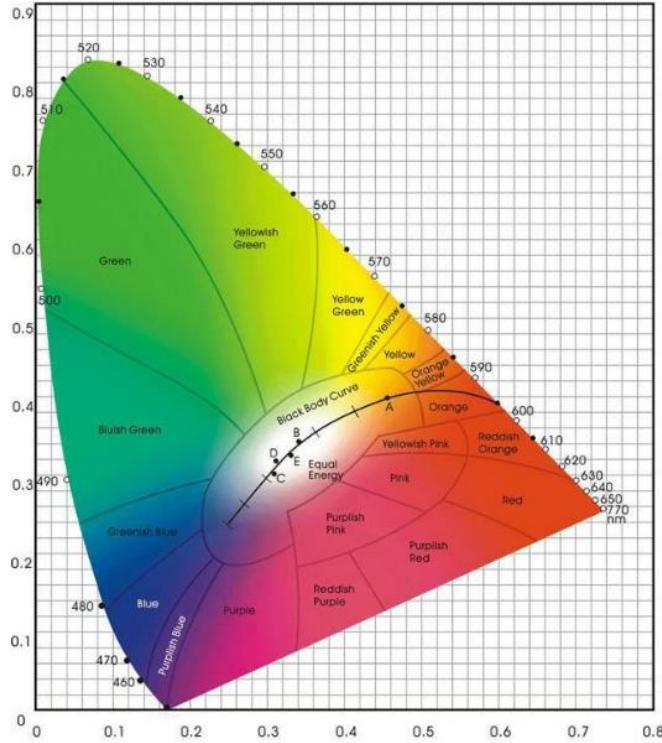


图 3-4: CIE-XYZ 1931

XYZ 表示法与 RGB 表示法存在一定的联系并且有一些重要的性质:

- XYZ 色度图中的原点, 对应图 2, RGB 表示法中的 C_b 点。同时, 图 2 中的直线 C_bC_g 对应于 XYZ 表示法中的 y 轴, 且经过变换后, 在图 4 中的对应长度为 1, 同理, C_bC_r 对应于 x 轴, 且对应的长度也为 1。
- 图 2 中 C_rC_g 线段上的点, 在新的 XY 坐标系中, 仍然在这条线段上。
- XYZ 表示法中不会出现负值。
- 原坐标系中的等能白光坐标 $(\frac{1}{3}, \frac{1}{3})$ 在新的坐标系中也为这个坐标。
- 所有的单色光都出现在图 4 的边缘处。
- 在 XYZ 色度图中选择任意两点, 位于两点连线之间的任何颜色都可以由这两种颜色表示出来。这也意味着 XYZ 色度图必然是一个凸形。将这个性质进行推广, 任意三个光源可以混合产生的颜色, 必然在这三个光源对应点形成的三角形内部。多个光源亦然。
- 最后, 很重要的一点是从图 4 我们可以发现, 任何三个给定光源, 都不能覆盖整个色域。几何上说, 色域中不存在能包含整个色域的三角形, 即人类视觉的色域不是三角形的, 使用三色表示法无法表示整个人类色域, 而只能表示局部。

XYZ 表示法是用来表示色度的, 也就是说, 使用 XYZ 表示法可以表示人眼所能看到的全部颜色, 但是还不能表示亮度, 因为 XYZ 表示法是归一化以后的结果。因此, 通过定义 $\bar{y}(\lambda) = V(\lambda)$, 就可以解决这个问题, 其中 $V(\lambda)$ 是可见光的光谱光效率函数, 即人眼对等能量的各种色光的敏感程度函数。由此, 就可以通过 $\bar{y}(\lambda)$ 来表示出 $\bar{x}(\lambda), \bar{z}(\lambda)$ 。

$$\begin{aligned}\bar{x}(\lambda) &= \frac{x(\lambda)}{y(\lambda)} V(\lambda) \\ \bar{y}(\lambda) &= V(\lambda) \\ \bar{z}(\lambda) &= \frac{z(\lambda)}{y(\lambda)} V(\lambda)\end{aligned}\tag{6}$$

这三个刺激值就可以使人眼既感受到颜色，又感受到亮度。它们也被称为“CIE 1931 标准观察观察者标准”。

而对于任意一个知道光谱分布的光 $\Phi_e(\lambda)$ ，如果想要得到它在“CIE 1931 标准观察观察者标准”下的表示方法，可以先求出对应的 XYZ 值：

$$\begin{aligned} X &= \int_{380}^{780} \Phi_e(\lambda) \bar{x}(\lambda) d\lambda \\ Y &= \int_{380}^{780} \Phi_e(\lambda) \bar{y}(\lambda) d\lambda = \int_{380}^{780} \Phi_e(\lambda) V(\lambda) d\lambda \\ Z &= \int_{380}^{780} \Phi_e(\lambda) \bar{z}(\lambda) d\lambda \end{aligned} \quad (7)$$

这里的 Y 值就是颜色的亮度。然后进行归一化：

$$\begin{aligned} x &= \frac{X}{X + Y + Z} \\ y &= \frac{Y}{X + Y + Z} \\ z &= \frac{Z}{X + Y + Z} \end{aligned} \quad (8)$$

这里的小写 x, y 就是颜色在色度图中对应的色坐标。有了色坐标 x, y 以及亮度 Y，一个光的颜色就可以被确定下来了。因此，“CIE 1931 标准观察观察者标准”也被称为 CIE-xyY 表示法。

3.2 色域

上一节介绍了颜色空间，并且给出了颜色空间的常用表示方法。如图 4 中所示的舌型图，表示了人眼所能感知到的所有颜色的范围。我们需要注意的是，这些颜色是指自然界中人类所能感知到的所有颜色。在图像与视频处理领域，显示设备所能显示的颜色，远远没有自然界中的丰富多彩，产生这种情形的原因主要是生产力与技术的限制。上一节我们提到，显示设备通过在三种基色混合，来表示其他的颜色，并且我们提到“几何上说，色域中不存在能包含整个色域的三角形，即人类视觉的色域不是三角形的，使用三色表示法无法表示整个人类色域，而只能表示局部。”因此，显示设备的制作原理导致我们无法使用显示设备来显示出自然界中存在的所有颜色，而只能选择其中的一部分颜色进行显示，如何选择则取决于颜色标准。

图像领域中通常使用的色域标准来自 ITU-R，使用最广泛的是 BT. 709 [1] 和 BT. 2020 [2] 标准。BT. 709 是 HDTV 的标准，也是目前普通显示设备使用的色域标准。BT. 2020 是 UHDTV 的颜色标准，也是 HDR 技术所支持的色域标准。

3.2.1 ITU-R BT. 709 标准

ITU-R BT. 709 也称为 Rec. 709，该标准的第一版于 1990 年发布。BT. 709 定义了 $R'G'B'$ 和 $Y'C_bC_r$ 的两种编码模式，每种模式的各通道都有 8 或 10 比特两种类型，在 8 位编码中， $R'G'B'$ 通道和 Y' 通道的标称范围是 16~235， C_b 和 C_r 通道的标称范围是 16~240，中性值为 128。因此，在 $R'G'B'$ 中，黑色的值是 (16, 16, 16)，白色的值是 (235, 235, 235)；在 $Y'C_bC_r$ 中黑色的值是 (16, 128, 128)，白色的值是 (235, 128, 128)。标

准 RGB 空间的 0 和 255 仍保留作为定时参考，可能不包含颜色数据。

BT. 709 标准同时支持模拟信号与数字信号，因此在 BT. 709 标准内部，存在两个重要的转换： $R'G'B'$ 编码到 $Y'C_bC_r$ 编码的转换，以及模拟信号到数字信号的转换。

$R'G'B'$ 编码下的模拟信号到 $Y'C_bC_r$ 编码下的模拟信号之间的转换为：

$$\begin{aligned} E_Y &= 0.2126E_R + 0.7152E_G + 0.0722E_B \\ E_{CB} &= \frac{-0.2126E_R - 0.7152E_G + 0.9278E_B}{1.8556} \\ E_{CR} &= \frac{0.7874E_R - 0.7152E_G - 0.0722E_B}{1.5748} \end{aligned} \quad (9)$$

其中 E_Y, E_{CB}, E_{CR} 分别表示亮度信号和两种色差信号， E_R, E_G, E_B 分别表示 R', G', B' 三个通道的模拟信号值。

两种编码方式的数字图像的量化方法是：

$$\begin{aligned} D_{R,G,B} &= INT[(219E_{R,G,B} + 16) \cdot 2^{n-8}] \\ D_Y &= INT[(219E_Y + 16) \cdot 2^{n-8}] \\ D_{CB} &= INT[(224E_{CB} + 128) \cdot 2^{n-8}] \\ D_{CR} &= INT[(224E_{CR} + 128) \cdot 2^{n-8}] \end{aligned} \quad (10)$$

其中 n 表示量化信号的比特长度，INT 作为将小数化整的四舍五入功能。因此我们可以得出 $R'G'B'$ 编码下的数字信号到 $Y'C_bC_r$ 编码下的数字信号之间的转换方法：

$$\begin{aligned} D_Y &= INT[0.2126D_R + 0.7152D_G + 0.0722D_B] \\ D_{CB} &= INT\left[\left(-\frac{0.2126}{1.8556}D_R - \frac{0.7152}{1.8556}D_G + \frac{0.9278}{1.8556}D_B\right) \cdot \frac{224}{219} + 2^{n-1}\right] \\ D_{CR} &= INT\left[\left(\frac{0.7874}{1.5748}D_R - \frac{0.7152}{1.5748}D_G - \frac{0.0722}{1.5748}D_B\right) \cdot \frac{224}{219} + 2^{n-1}\right] \end{aligned} \quad (11)$$

3. 2. 2 ITU-R BT. 2020 标准

ITU-R BT. 2020 定义了 UHDTV 的各个方面，比如显示分辨率、帧速率、色度子采样、比特深度和色彩空间，它是由国际电信联盟在 2012 年发布的。相比于 BT. 709，BT. 2020 提供了更加广泛的颜色范围。BT. 2020 也是 HDR 技术的颜色空间。

BT. 2020 标准的重要性是毋庸置疑的，正如 BT. 709 对于高清视频传输与高清显示设备制造方面起到的引导性作用，BT. 2020 标准同样也深刻地影响着消费领域超高清显示设备的设计与制造，例如目前绝大部分的 4K 平板电视的物理分辨率都是采用 BT. 2020 标准的 3840×2160 ，但 BT. 2020 标准绝不仅仅在分辨率方面有所提升，在色彩和刷新频率和信号格式与分析等方面也进行了相关的规定。

相对于 BT. 709 标准，BT. 2020 标准大幅度提升了视频信号的性能规范，色彩深度方面提高至 10bit 或 12bit，很大程度上增强了整个影像的色彩层次，色域的广度也大大增强，两种标准的色域对比如图 1.13 所示；显示分辨率方面，BT. 2020 标准定义了 4K 和 8K 两种分辨率，这些分辨率的宽高比为 16: 9，支持的刷新率包括 120p、60p、

59.94p、50p、30p、29.97p、25p、24p、23.976p。不得不提的是，在BT.2020标准中取消了隔行扫描，所有超高清标准下的影像都是基于逐行扫描的，无疑是历史性的突破，进一步提升了超高清影像的细腻度与流畅感。

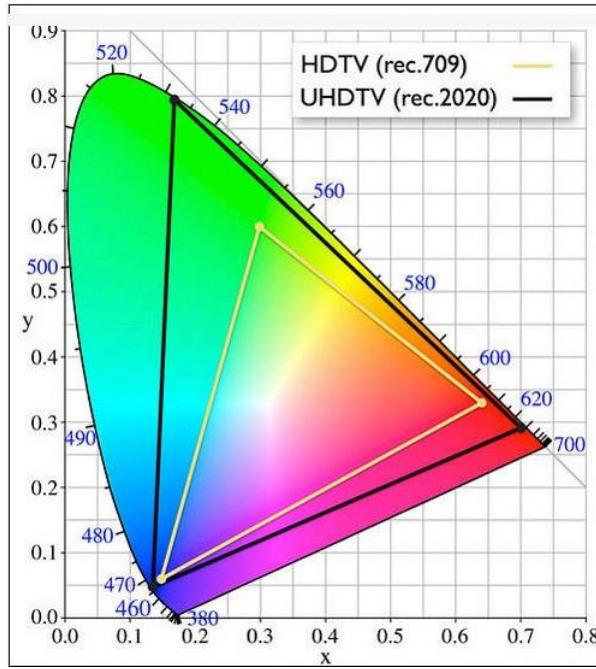


图 3-5：BT.709 与 BT.2020 对比

与 BT.709 一致，BT.2020 也支持了 $R'G'B'$ 和 $Y'C_bC_r$ 两种编码模式，以及支持模拟与数字两种信号。同样的，BT.2020 标准内部也存在两个重要的转换：两种编码模式之间的转换以及模拟信号到数字信号的转换。

$R'G'B'$ 编码下的模拟信号到 $Y'C_bC_r$ 编码下的模拟信号之间的转换：

$$\begin{aligned} E_Y &= 0.2627E_R + 0.6780E_G + 0.0593E_B \\ E_{CB} &= \frac{-0.2627E_R - 0.6780E_G + 0.9407E_B}{1.8814} \\ E_{CR} &= \frac{0.7373E_R - 0.6780E_G - 0.0593E_B}{1.4746} \end{aligned} \quad (12)$$

$R'G'B'$ 编码下的数字信号到 $Y'C_bC_r$ 编码下的数字信号之间的转换：

$$\begin{aligned} D_Y &= INT[0.2627D_R + 0.6780D_G + 0.0593D_B] \\ D_{CB} &= INT\left[\left(-\frac{0.2627}{1.8814}D_R - \frac{0.6780}{1.8814}D_G + \frac{0.9407}{1.8814}D_B\right) \cdot \frac{224}{219} + 2^{n-1}\right] \\ D_{CR} &= INT\left[\left(\frac{0.7373}{1.4746}D_R - \frac{0.6780}{1.4746}D_G - \frac{0.0593}{1.4746}D_B\right) \cdot \frac{224}{219} + 2^{n-1}\right] \end{aligned} \quad (13)$$

除了颜色空间和色域的转换，在处理 HDR 源视频或通过图片合成 HDR 视频时，通常还要经过采样方式和光电转换函数等步骤的转化，常见的 HDR 视频符合 SMPTE ST 2084 规定的转换方程的要求，按照不同要求经过 PQ 曲线或 HLG 的处理。具体可以参考本系列前面的内容。

3.2.3 ITU-R BT. 709 到 ITU-R BT. 2020 的颜色转换

UHDTV (Ultra-high Definition Television, 超高清电视) 是未来电视与显示领域的发展趋势, 现在, 越来越多的电视和节目制作人正开始制作 UHDTV 节目。由于 UHDTV 还在刚刚起步的阶段, 因此适配于 UHDTV 的节目内容数量仍然不足, 因此, 将原有的大量的 HDTV 的节目, 通过一定的转换, 转变为 UHDTV 节目, 就是一个解决 UHDTV 节目数量不足的可行的方案。由于 UHDTV 支持的色域要宽于 HDTV 的色域, 在转换的过程中, 不可避免的需要对节目从 ITU-R BT. 709 标准向 ITU-R BT. 2020 标准进行颜色转换。

BT. 709 到 BT. 2020 的转换有一个重要的限制, 即 BT. 709 标准下的颜色, 在 BT. 2020 标准下要是同样的颜色。换句话说, 如果这个颜色是两种色域中都存在的, 那么这个颜色在转换过程中应当保持不变; 而如果这个颜色是 BT. 2020 中独有的, 这就需要进行颜色转换。BT. 709 到 BT. 2020 的转换可以大致分为三个主要步骤: 第一个步骤是逆量化, 得到 BT. 709 标准下的模拟信号值 E_R, E_G, E_B 。第二个是色域转换, 将 BT. 709 下的模拟信号值, 转换为 BT. 2020 标准下的模拟信号值。最后一个步骤是量化, 得到 BT. 2020 标准下的数字信号。选择在模拟信号域进行色域转换而不是直接在数字信号域进行色域转换的原因是为了保留更多的信息与细节。由于量化过程会存在信息的丢弃, 并且 BT. 709 的色域比 BT. 2020 小, 如果直接用数字信号进行转换, 无疑会使得 BT. 2020 色域下的图像效果变差。因此, 通过逆量化过程, 恢复图像的相关细节, 然后进行模拟信号的色域转换, 就可以解决这个问题了。

对于原视频/节目的不同形式, 转换方式也略有不同。根据信号源的亮度是否恒定, 转换方式可以分为以下两种:

1. 非恒定亮度信号的转换

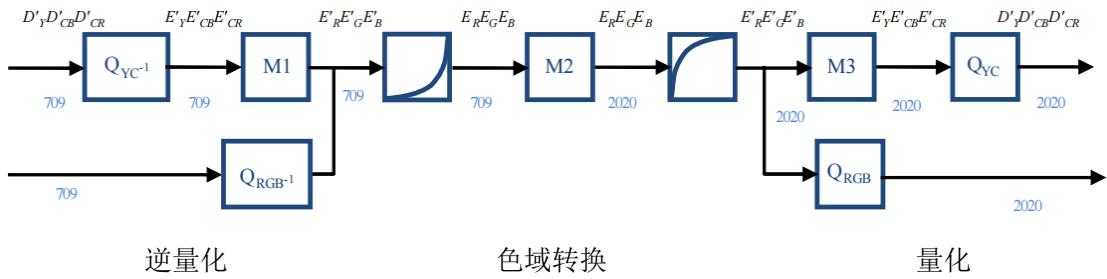


图 3-6. BT. 709 到 BT. 2020 在非恒定亮度信号格式下的转换过程

上图是非恒定亮度信号的转换流程图。由于 BT. 709 与 BT. 2020 均支持两种编码方式: $R'G'B'$ 和 $Y'C_bC_r$, 因此逆量化过程的输入与量化过程的输出均有两条路径。根据需求的不同, 可以选择不同的输入输出模式。

每一个区块的具体实现过程以及相关函数定义, 可以参考 BT. 2087 [3]。

2. 恒定亮度信号转换

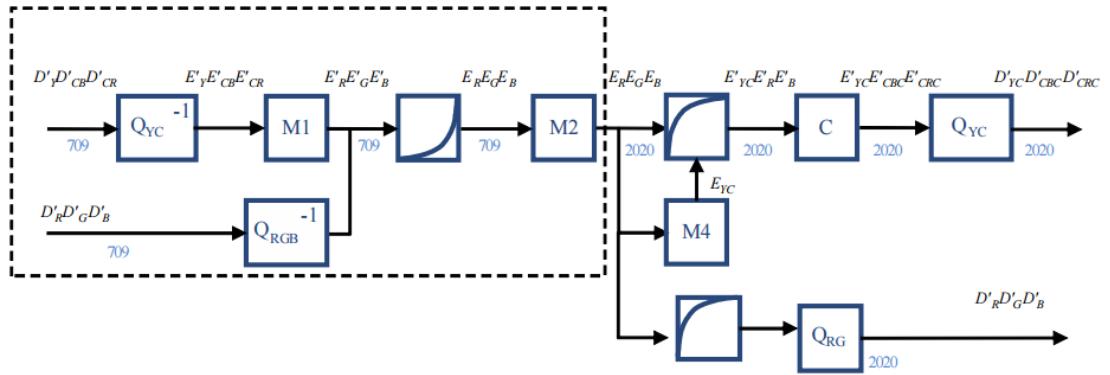


图 3-7. BT. 709 到 BT. 2020 在恒定亮度信号格式下的转换过程

上图是恒定亮度信号的转换流程图。由于 BT. 709 与 BT. 2020 均支持两种编码方式： $R'G'B'$ 和 $Y'C_bC_r$ ，因此逆量化过程的输入与量化过程的输出均有两条路径。根据需求的不同，可以选择不同的输入输出模式。

每一个区块的具体实现过程以及相关函数定义，可以参考 BT. 2087。

3.2.4 ITU-R BT. 2020 到 ITU-R BT. 709 的颜色转换

在现在的实际 UHDTV 节目的广播过程中，特别是在同时进行 UHDTV 和 HDTV 广播的情况下，从 UHDTV 到 HDTV 的实时高质量色域转换就至关重要。也就是说，需要一些方法，来使得 UHDTV 也能够在 HDTV 设备上成功播放，并且不影响观看。这就需要从 ITU-R BT. 2020 到 ITU-R BT. 709 的颜色转换，也就是 UHDTV 到 HDTV 的转换。理想情况下转换应该满足以下要求：BT. 709 色域内的颜色应该保持不变；转换方法适用于 BT. 2020 和 BT. 709 之间的多次转换；感知的色调变化必须尽可能小；空间细节没有重大损失；不会在颜色上引入可见的不连续性；映射方法在数学上是可定义的。然后没有可以同时满足所有这些要求的通用色域映射方法。在从更广泛的色域转换到更小的色域时，BT. 709 色域外的颜色修改是不可避免的。此转换必然是不同要求之间的妥协，可能因应用程序而异。色域映射算法通常受艺术创作，人类视觉，技术约束和经验等方面的影响。下面我们介绍一种从 BT. 2020 到 BT. 709 的基于线性矩阵的简单变换。

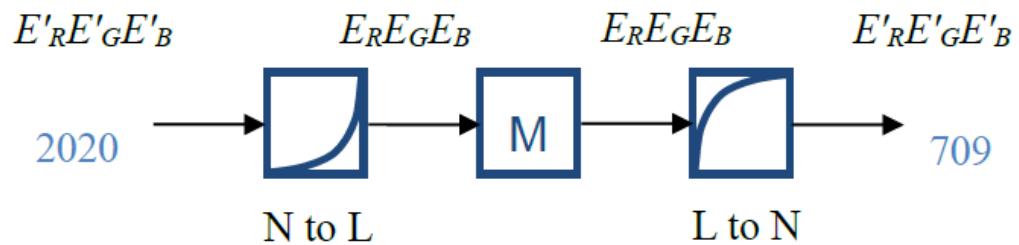


图 3-8 基于线性矩阵变换从 BT. 2020 到 BT. 709 的颜色转换

图 8 是从 BT. 2020 到 BT. 709 的基于线性矩阵变换的颜色转换的框图。除了输出信号被严格限制之外，这是 ITU-R BT. 2087 标准中对从 BT. 709 转换到 BT. 2020 所规定的操作的逆操作。这种方法在条件限制的情况下是最直接简单的。

这种方法首先使用 ITU-R BT. 2087 中规定的两个非线性到线性的传递函数（一种是用于使用电光转换功能的显示参考方法（EOTF），一种是用于使用反光电转移函数

(OETF)) 之一将归一化的非线性 RGB 信号转换为归一化的线性 RGB 信号。接着通过矩阵运算，将 BT. 2020 颜色空间的 RGB 信号转换到 BT. 709 颜色空间的 RGB 信号。最后再通过使用 ITU-R BT. 2087 建议书规定的两种线非线性传递函数（一种是使用反向 EOTF 的显示参考方法，另一种是使用 OETF）之一将线性 RGB 信号转换为非线性 RGB 信号。从归一化线性 RGB 信号 (EREGB) 到归一化非线性 RGB 信号 (E' RE' GE' B) 的转换。

该方法具有非常理想的特性，即使在 BT. 2020 和 BT. 709 之间进行多次转换之后，它也不会改变 BT. 709 色域内的颜色。然而，BT. 709 色域之外的色彩是硬限幅的，即小于零或大于 1 的 RGB 信号 (EREGB) 会被限幅为零或 1，这会导致明显的色调偏移和空间细节的损失。虽然大多数内容看起来都很好，但色调和空间细节方面还是会有所损失。

色域转换的更多内容，将在本书第二部分的《色域映射》章节中进行介绍。

第四章-HDR 主要标准

HDR 是 UHD 技术中最重要维度之一，带来新的视觉呈现体验。HDR 技术涉及到采集、加工、传输、呈现等视频流程上的多个环节，需要定义出互联互通的产业标准，以支持规模化应用和部署。本文整理当前 HDR 应用中的一些代表性的国际标准。

4. 1 HDR 发展现状

HDR 技术的发展，基于 UHDTV 技术的革新。相较于 HDTV，UHDTV 在三个方面进行了巨大的提升：

1. 视频分辨率。UHDTV 视频的分辨率达到了 $3840*2160$ ，也就是“4K”分辨率。相较于 HDTV 视频 $1920*1080$ 的分辨率，UHDTV 提供了 4 倍的像素点。
2. 显示设备的动态范围。显示设备的动态范围是指其最大亮度与最小亮度的比值。在新的 UHDTV 显示设备上，可显示亮度的范围为： $0.00005\text{ cd}/\text{m}^2 \sim 1000\text{ cd}/\text{m}^2$ ，相比之下，HDTV 可显示亮度范围仅为： $0.0002\text{ cd}/\text{m}^2 \sim 100\text{ cd}/\text{m}^2$ 。显示设备亮度范围的提升，为 HDR 技术的应用提供了基础。
3. 显示设备的色域范围。UHDTV 中应用的是广色域技术(WCG, Wide Color Gamut)。传统的 HDTV 显示设备支持的是 BT. 709 或 BT. 601 标准下的色域，这两种标准仅仅包含了 33.5% 的可见颜色。UHDTV 显示设备支持的是 DCI-P3 与 BT. 2020 标准下的色域，它们分别包含了 41.8% 与 57.3% 的可见颜色。可见，色域范围的提升是非常明显的。

显示设备性能的提升为 HDR 技术的发展提供了基础。HDR 技术的目的，是在充分利用 UHDTV 亮度范围的基础上，提供更加接近现实场景的视频。在常规亮度区域，HDR 视频与 SDR 视频的亮度基本一致；在低亮度区域，由于 HDR 技术提供了更多的亮度等级，因此 HDR 视频在低亮度区域的细节会得到增强；在高亮度区域，由于 HDR 视频的亮度范围远远高于 SDR 视频，因此大部分 SDR 视频中的峰值亮度在 HDR 视频中能够得到正常显示，只有类似于镜面反射这种亮度的像素点，才会被设计为最高亮度。

HDR 技术包含两个重要方面，第一个是内容制作，第二个是分发与显示。这两方面 HDR 技术与传统电视技术都存在差异。

内容制作方面，HDR 电视的视频源与传统视频源的摄影设备存在区别。传统的摄影设备的曝光度有六阶，而现在的专业 HDR 摄影设备已经具有 14 阶的曝光度（例如电影摄影设备），普通的民用 HDR 摄影设备也普遍具有 8-10 阶的曝光度。这就使得 HDR 视频源具有更好的细节表现能力。

内容分发方面，由于 HDR 视频的 bit 深度要大于传统视频，因此 HDR 技术通常需要更宽的传输带宽或更优质的传输通道。但是，为了实现向下的兼容，部分 HDR 标准中使用了一些技术（例如动态元数据）实现了带宽的压缩，使 HDR 信号在传统的传输标准仍然可以传输。

内容显示方面，HDR 电视能够提供更加广的显示亮度，相较于传统显示设备 0-100nits 的显示亮度，HDR 显示设备通常可以达到 0-10000nits 的显示亮度。由此可以发现，HDR 电视显示的视频是真正的 HDR 视频，不仅仅能够显示更多的图像细节，同时还能显示更广的亮度；而 HDR 摄影技术拍摄出的 HDR 图像或视频仅仅只能显示更多图像细节，而无法对显示亮度范围做出提升。

4. 2 HDR 相关技术组织

本节对主要 HDR 技术规范进行总结。由于 HDR 技术本身基于 UHDTV，因此在 HDR 技术中，除了对显示设备的亮度范围进行了提升之外，对显示设备的颜色范围也作了提升。在 HDR 技术规范的建立过程中，以下组织对 HDR 的各方面技术与参数标准进行了规定：The International Telecommunications Union (ITU)，Society of Motion Picture and Television Engineers (SMPTE)，Consumer Electronics Association (CEA)，Ultra HD Alliance，and the Blu-ray Disc Association (BDA)。

SMPTE

SMPTE 在 HDR 方面的标准主要集中在两方面：EOTF（电光转换函数）曲线的标准以及元数据的标准。由于 EOTF 曲线以及元数据是 HDR 技术中非常重要的两部分，因此在后面一节中会对这两部分做出进一步的介绍。

- SMPTE Standard ST 2084:2014 [1]：该标准中对 HDR 技术中使用的 EOTF 曲线进行了定义，其中规定的 EOTF 曲线是 Dolby Vision 测定的 PQ (Perceptual Quantization) 曲线。
- SMPTE Standard ST2086:2014 [2]：该标准中对 HDR 技术中使用的静态元数据进行了定义。
- SMPTE Standard ST 2094:2016 [3]：该标准中对 HDR 技术中使用的动态元数据进行了定义。

ITU

ITU 主要对 HDR 技术中的颜色范围进行了规定。

- ITU-R Recommendation BT.2020 (International Telecommunication Union, 2015)：该标准对 UHDTV 系统以及 HDR 技术中使用的色域进行了定义。
- ITU-R Report BT.2381-0 [5]：该标准对 HDR 电视分发系统的性能做出了规定。同时，该标准中也包括了对了 HDR 内容向后兼容 SDR 显示设备的相关标准做出了规定。

CTA

The Consumer Technology Association (CTA)，前身为 CEA，对 HDR 显示设备应当达到的最低标准进行了限制：

- 包含至少一个 HDR 信号接口。
- 能够接受并处理 HDR 静态元数据 (HDMI 接口需要达到 HDMI2.0a 以传递 HDR 元数据)。
- 能够接受并处理由 IP, HDMI 或其他接口得到的 HDR 10 媒体文件。其他 HDR 媒体文件也可进一步得到支持。
- 在显示图像之前能够使用合适的 EOTF 曲线对媒体文件进行处理。

UHD Alliance

UHD 联盟中包括了所有的主要 UHDTV 的内容制造商，内容分发机构以及显示设备制造商。他们对 UHDTV 的相关最低标准进行了限制：

- 分辨率：内容，分发与显示的分辨率都应达到 3840*2160。
- 颜色的 bit 深度：内容，分发与显示过程中使用的 bit 深度最小为 10bits。
- 色域：支持 BT. 2020 标准中的色域。
- 母带制作标准：转换函数为 SMPTE ST2084 标准中的 EOTF 的反函数。颜色空间最小为 100% 的 P3 颜色空间。峰值亮度超过 1000nits；最暗亮度低于 0.03nits。
- 内容转换函数：遵循 SMPTE ST 2084。
- 显示内容标准：转换函数为 SMPTE ST 2084 标准中的 EOTF 函数。颜色空间最小为 90% 的 P3 颜色空间。峰值亮度超过 1000nits 并且最暗亮度低于 0.05nits 或者峰值亮度超过 540nits 并且最暗亮度低于 0.0005nits。

UHD 联盟使 UHDTV 内容能够向下兼容 SDR 显示设备。

4. 3 HDR 中的转换曲线

由于 EOTF 函数以及元数据是 HDR 标准中十分重要的组成部分，因此在本节以及下节中，分别对这两部分的相关标准进行介绍。

图像或视频的内容由场景中获取，再到显示设备上显示出来，要经历两个过程：一个是由场景中的自然光线转换到图像编码的光电转换过程；另一个是从图像编码转换到显示设备显示亮度的电光转换过程。由于 HDR 文件的编码长度较长，因此在光电转换过程中通常可以将场景中的亮度高精度地储存在图像编码当中，所以 HDR 转换函数的重点，在于图像编码到显示设备的电光转换过程，这一过程，由光电转换函数（EOTF）来执行。

不同显示设备的峰值亮度和最暗亮度不同，并且亮度分级也不同，因此，不同显示设备有不同的 EOTF。传统的 BT. 709 与 BT. 1886 标准中规定的指数 Gamma 曲线就是 SDR 显示设备的 EOTF。Gamma 曲线是对相对亮度的量化，这些曲线规定了相对亮度，它们将图像编码最大值与显示设备最大亮度进行映射，最小值与最低亮度进行映射，中间也是类似的量化机制。由于传统 Gamma 曲线是根据老式 CRT 显示设备测算的，因此 Gamma 曲线只是对人眼在低亮度区域特性的总结，也就是在 0–100nits 范围内，Gamma 曲线能够准确转换编码与亮度。

在 HDR 技术中，显示设备亮度能达 1000nits，如果仅将 Gamma 曲线拉伸 10 倍，那么在图像中就会出现明显的块状、环状瑕疵现象。因为在高亮度区域（大于 100nits 的区域）人眼有着不同的视觉特性。解决此问题有两种不同设计思路，进而形成两类 HDR EOTF 曲线。

4. 3. 1 PQ 曲线 (SMPTE ST 2084)

在 SMPTE ST 2084 中规定的 EOTF 曲线没有采用之前的相对亮度量化的方式，而是采用了绝对亮度量化方式。规定的 EOTF 曲线将 10–12bits 的编码转换到固定的亮度上，这就要求显示设备的 EOTF 曲线要与母片制作过程中使用的 EOTF 曲线完全一致。

ST 2084 标准中的 EOTF 曲线为了能够满足以后技术的发展，定义了 0.00005nits 到 10000nits 范围内的映射曲线。ST 2084 将每一个视频编码映射到一个在每一个显示设备上都相同的绝对亮度值（例如 10bits 下的 691 码字永远对应于 500nits 的亮度）。

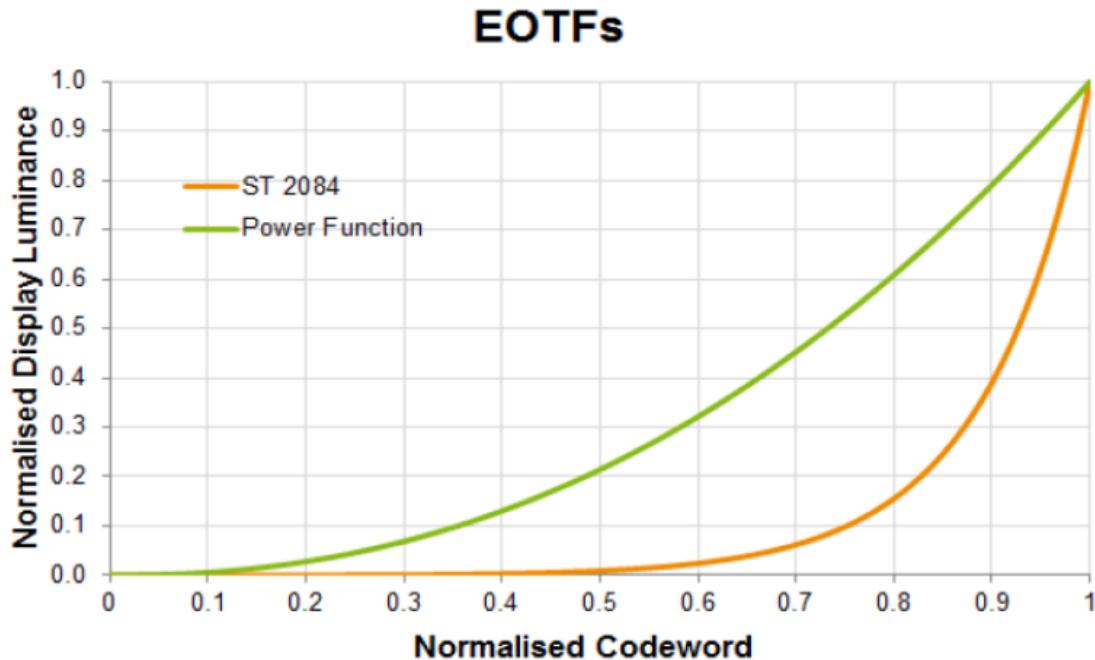


图 4-1 BT. 709 与 SMPTE ST 2084 标准下的 EOTF 曲线

SMPTE ST 2084 标准中的 EOTF 曲线基于人眼的对比度敏感函数，它由 Barten 测定，并记录于 ITU-R BT.2246-6 [6]。这种 EOTF 曲线也称为 PQ 曲线。由于 PQ 曲线充分利用了人眼视觉特性，因此它对 bit 深度做了充分利用。PQ 曲线仅仅使用很少码字就可以表示高达 10000nits 的亮度（如图 1 所示，高亮度范围仅仅占用了很小的 bits 范围）。因为人眼在低亮度区域有更高的敏感性，因此在低亮度区域中使用更多码字，可以显著提升人眼的视觉体验，而高亮度区域人眼敏感度较低，因此不需要使用过多码字。

如果仅仅使用上述 PQ 曲线进行转换，会发现实际显示图像的对比度远弱于原始场景。这是因为原始场景的亮度要高于实际设备的亮度，因此经过 PQ 曲线映射后，显示对比度要弱于原始场景对比度。为了解决此问题，需要使用 OOTF（Optical to Optical Transform Function，光光转换函数）将极亮区域与极暗区域的亮度进行压缩，并增强中间区域的对比度。

在 SMPTE ST 2084 规定的 PQ 系统中，OOTF 过程在拍摄或制作过程中进行。

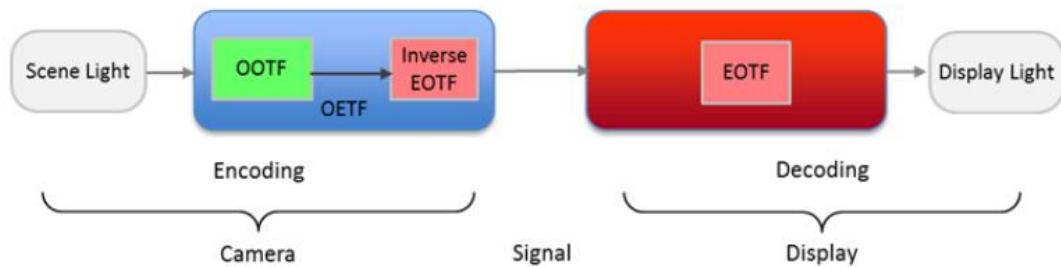


图 4-2 PQ 系统中，OOTF 在拍摄或制作过程中进行

4.3.2 HLG 曲线 (Hybrid Log Gamma)

另一种 EOTF 曲线的设计思路与 PQ 曲线不同。由于 PQ 曲线的设计思路与传统 Gamma 曲线完全不同，因此，PQ 曲线不能向下兼容传统的 SDR 视频的编码。但是，对于电视运营商而言，如果使用 PQ 曲线进行转换，就需要为 HDR 和 SDR 内容分别设计数据流并分别传输数据，这意味着成本的上升。因此，BBC 与 NHK 合作推出了一种新的 EOTF 曲线—HLG 曲线。HLG 曲线的优势在于它能够兼容 SDR 图像的编码。当一个传统显示设备接收到视频信号时，它可以使用传统的 Gamma 曲线将视频显示为 SDR 视频。而一个 HDR 显示设备能够将视频信号通过 HLG 曲线转换为 HDR 视频。这意味着电视运营商能够只传输一种数据流，使得 SDR 用户与 HDR 用户都能够获得电视信号。

HLG 曲线是运营商为了节约传输成本而做出的妥协，它能够很好地兼容 SDR 与 HDR 的数据，但会使得两种视频的观看体验都无法达到最佳。为了实现对 SDR 的兼容，HLG 曲线在低亮度区域的映射曲线就是传统的 Gamma 曲线，而在高亮度区域使用对数曲线进行映射。这也是 Hybrid Log Gamma 名称的由来。

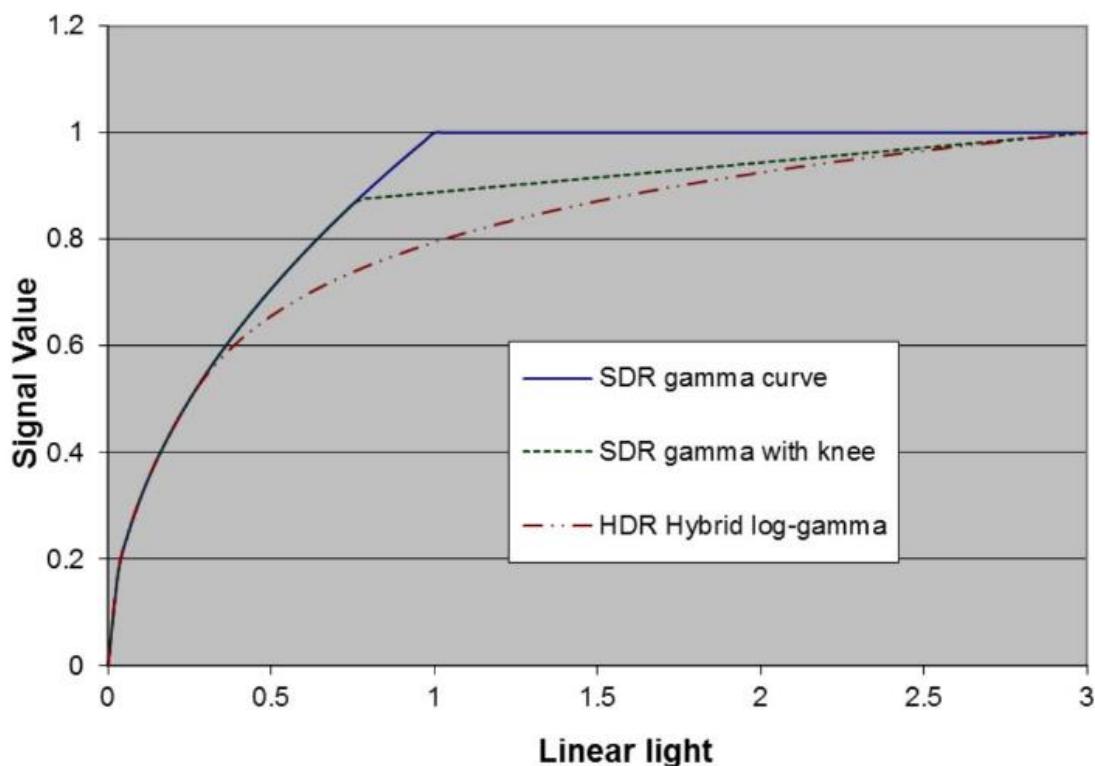


图 4-3 HLG 曲线图像

BBC 公司利用 HLG 曲线进行了小规模的 HDR 电视节目制作与传输实验。BBC 制作了自然纪录片“Planet Earth II”的 HDR 版本，并在松下的 HDR 电视上进行了传输与显示实验。同时，由于 UHD 蓝光标准中不支持 HLG 曲线，因此在其节目蓝光版本中，使用 HDR10 标准来替代。

4.4 HDR 元数据规范

HDR 的元数据分为静态元数据与动态元数据两类。

静态元数据

SMPTE ST 2086 标准中，对 HDR 技术中的静态元数据所应遵循的标准进行了定义。在静态元数据中，包含了 HDR 内容的颜色信息以及亮度信息。HDMI 2.0a 支持这种静态元数据的传输。

静态元数据信息中包含了以下内容：以红，绿，蓝三色为基色的颜色信息；HDR 母带的白色点信息；最暗亮度以及峰值亮度。ST 2086 同时也会传输 HDR 内容的下列几种亮度信息：

- 最大内容亮度等级 (MaxCLL, Maximum Content Light Level)：MaxCLL 的单位是 cd/m^2 。表示 HDR 内容中最亮的像素点的亮度值。
- 最大平均帧亮度等级 (MaxFALL, Maximum Frame-Average Light Level)：单位为 cd/m^2 。平均帧亮度表示一帧内所有像素点的平均亮度。MaxFALL 表示 HDR 内容中所有帧的最大的平均帧亮度。

动态元数据

SMPTE ST 2094 标准中对 HDR 技术的动态元数据进行了标准定义。动态元数据是指，在传输过程中，传递的是帧与帧之间，或者场景与场景之间的颜色重映射信息 (CRI, Color Remapping Information)。也就是说，动态元数据传递的并非每一帧或每一个场景的绝对亮度与颜色，而是随着时间线传输颜色的变化与亮度的变化。这种做法的优势在于能够有效节约带宽。

动态元数据包含的种类较多，具体可参见 [3]。

4.5 HDR 内容的存储及分发标准

上述各种 HDR 技术规格是 SMPTE, ITU-R 等组织对 HDR 内容应符合的技术要求和限制。工业界与产业界根据这些内容标准，制定出了一些成型的 HDR 格式标准。这些格式标准，包含了 HDR 内容从制作，到传输，再到显示的整个流程中所应遵循的标准。HDR 的整个流程，通常分为两大部分，第一部分是母带制作，第二部分是分发与重放。下面介绍 HDR10, Dolby Vision, BBC/NHK HLG 三个主要 HDR 存储及分发标准。

4.5.1 HDR10

HDR10 标准是由 MPEG 最先发起的。HDR10 格式的设计目的，是为了建立一个“最低限度”的 HDR 标准。MPEG 的研究认为 HEVC main 10 (H.265) 已经足以保证 HDR 内容传输过程中的内容质量以及传输效率。同时 10bits 下的 4:2:0 的编码格式也足以满足 HDR 内容的传输。

基于 MPEG 的结果，Blu-ray Disc Association (BDA), HDMI Forum, UHD Alliance 等组织合力推出了利用 HEVC main 10 压缩与传输 HDR 内容的 HDR 标准。这种最基本的 HDR 标准被称为“HDR10”。HDR10 目前为止并非一个完全的标准，它目前仍然是各种标准与技术的集合体，并且仍然在随着 HDR 技术的发展而拓展。

HDR10 的母版制作过程如图 4 所示。首先，来自相机的线性光图像数据要进行某种形式的实时或离线分级，如通过 HDR 兼容的母片显示器上的视觉结果判断，内容的色彩体积在 BT.2020 规定的颜色空间容器内表示；随后，使用 SMPTE ST2084 中规定的 PQ 曲线[2]将 RGB 4:4:4 格式的图像数据编码为最能体现人类视觉特征的码字（图 5）；最后，将编码的数据分通道量化为 10 比特，执行颜色转换和色度子采样，将信号转化为 Y'CbCr 4:2:0 格式，并输送到 HEVC 编码器。

传统的 SDR 电视节目使用的 Gamma 曲线是基于电视显像管的特性定义的，无法承载高动态范围的信号，而 PQ 曲线表示的是光的绝对值，在从纯黑到 10000 尼特的范围内明确定义了光的值，因为 EOTF 的作用就是把数字视频信号的编码值换算成光强，所以 PQ 曲线具有更好的可扩展性，另外 PQ 曲线是基于人眼特性的，在人眼敏感的亮度范围内分配较多的码率，而在人眼不太敏感的高光部分则分配较少的码率，既符合人眼的视觉特征，又可以确保数据被充分利用。

ST 2086 标准的静态元数据将最暗亮度与峰值亮度以及其余亮度参数传递给显示设备。如果显示设备的色域小于母带的色域或者显示设备的亮度范围要小于母带的亮度范围，那么，设备就会根据这些静态元数据将色域和亮度范围相对地缩小到显示设备的范围内。

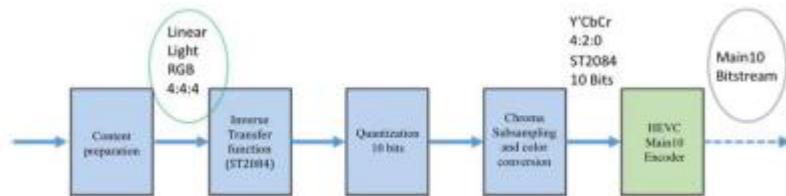


图 4-4: HDR10 内容的母版制作过程[1]

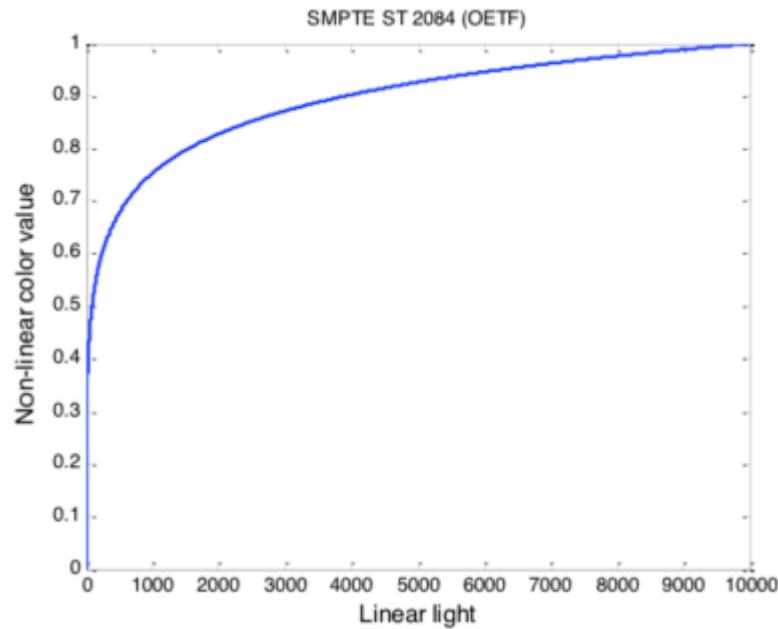


图 4-5: PQ 曲线[1]

HDR10 的传输与播放标准：

- 最小信号通道：HDMI 2.0a
- 重放色域：BT. 2020
- EOTF 曲线：SMPTE ST2084 (PQ) 曲线

- 静态元数据: SMPTE ST2086
- Bit 深度: 10bits

HDR10 显示器的播放过程几乎就是信号母片制作的逆过程, 对信号进行解码, 然后进行反颜色转换和子采样以恢复出 RGB 4:4:4 信号(图 6), 最后将该信号应用于 SMPTE ST2084 的播放转换函数(EOTF), 恢复出线性光值。

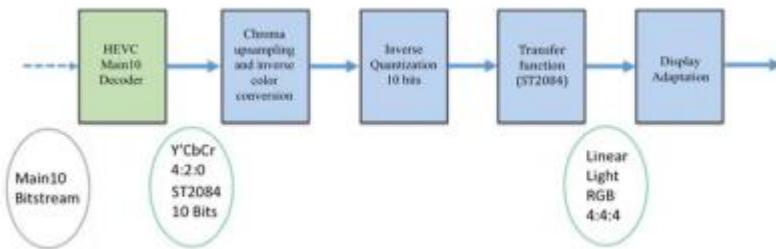


图 4-6: HDR10 内容解码过程[1]

ST2086 元数据中包含了母片显示器的黑电平和峰值亮度以及母片内容的亮度, 如果播放器的色域比视频内容窄或者播放器的峰值亮度低于视频内容, 播放器就可以根据元数据把视频内容映射到更窄的色域或更低的亮度, 同时播放器只需要重映射到视频内容的峰值亮度即可, 而不必达到 ST2084 中的最大值 10000 nit。

4.5.2 Dolby Vision

Dolby Vision 是一种高质量的 HDR 标准。与 HDR10 相比, Dolby Vision 对获得了许可的付费用户提供支持, 并且 Dolby Vision 的 HDR 标准更加严格。Dolby Vision 符合 4K UHD Blu-ray 规格, 是一个端到端、可扩展的系统, 使用 PQ 曲线映射线性光图像数据, 可以按需向下兼容 HDR10 和 SDR 显示器。

Dolby Vision 是杜比 Dolby Atmos 的补充。电影、电视和游戏创作者借助它可以实现自己的创意, 并让消费者体验到沉浸式的内容。DV 关键的一步是 HDR 调色, 包括高动态范围、宽色域以及可能的高分辨率、高帧率。相比于 SDR 动态范围更高、色域更宽的图像, 杜比称为 EDR (Extended Dynamic Range)。

Dolby 的 EDR 信号可通过 HDMI 进行呈现, 并且 Dolby 通过将元数据嵌入内容信号本身, 从而实现了将 12bits 内容深度的文件与元数据同时通过 HDMI 2.0 进行传输, 而不需要像 HDR10 一样, 只可以通过 HDMI 2.0a 进行传输。Dolby 利用专有的动态元数据, 实现亮度与颜色的端到端传输。动态元数据能够传输特定场景的亮度、颜色信息。

Dolby Vision 的母带制作过程有两种不同的设计方式: 单层文件方式与双层文件方式。其中双层文件方式通过将内容分为 base-layer 以及 enhancement-layer 两层, 使得该制作方式制成的母带能够向后兼容 SDR HDTV 以及 HDR10 的显示设备。另一种单层文件方式不提供向后的兼容, 但却能够节约大量的带宽。

- 双层配置包括基本层和增强层, 基本层采用了 10-bit 的 SDR HDTV (非 UHD Blu-ray) 或 HDR10 (可选择 UHD Blu-ray) 规格, 是为了与 SDR 设备兼容, 2-bit 的增强层包含了高动态范围和宽色域的部分 (EDR 增强层), 用于重建 EDR 内容, 当然增强层的传输会消耗一定的带宽。离线内容制作中经常使用双层配置。
- 单层配置常用于实时 EDR 内容制作中, 如在直播和 OTT 中的应用, 这种向后不兼容方案可以节省更多的码率。

Dolby Vision 母带中同时使用了静态元数据与动态元数据。其中使用的静态元数据包括：

- 最暗与峰值亮度
- 颜色表示以及
- MaxCLL
- MaxFALL

Dolby Vision 母带中使用的动态元数据用于描述帧与帧之间的亮度，并将内容与特定显示设备亮度对应。在双层形式的母带制作过程中，动态元数据反应了视频每一帧当中 HDR 内容与 SDR 内容的差异。动态元数据储存在 enhancement-layer 中。单层形式的母带中不包含动态元数据。

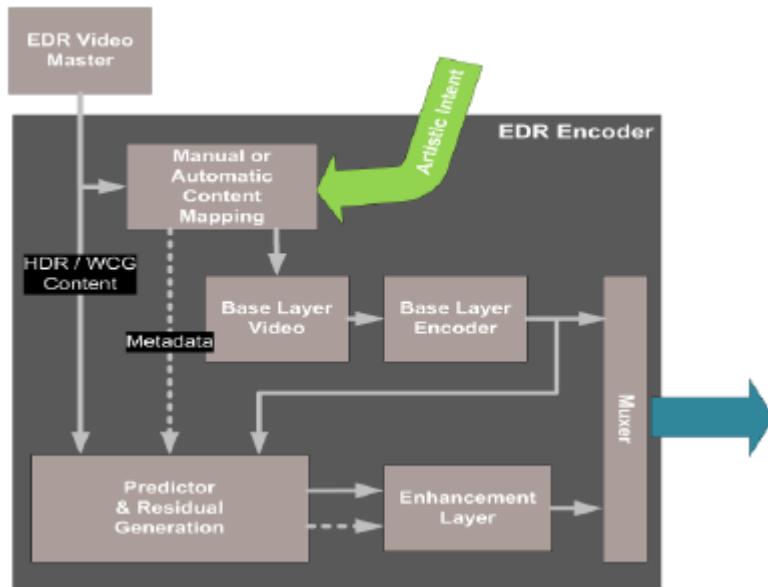


图 4-7: Dolby Vision 母片处理流程[1]

Dolby Vision 的传输与重放标准：

- 最小信号通道：HDMI 1.4
- 重放色域：BT. 2020
- EOTF 曲线：SMPTE ST2084 (PQ) 曲线
- 最大显示亮度：10000nits

所有的 Dolby 显示设备，都应能兼容双层与单层的母带。在解码过程中，对于双层文件，Dolby Vision 显示设备将 base-layer 与 enhancement-layer 分别解码，将 12bit 的文件解码为 10bits 的 base-layer 与 2bits 的 enhancement-layer，然后将两层与动态元数据并进行组合，形成 HDR 视频，然后进行播放。而如果普通显示设备接收到双层文件，则将 base-layer 部分直接进行解码放映即可。解码流程如图 8 所示。

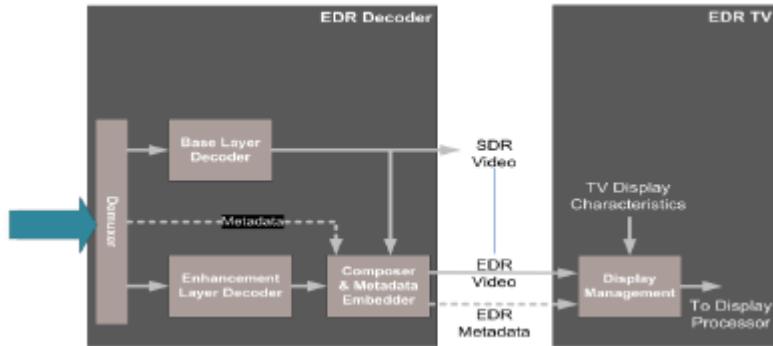


图 4-8: Dolby Vision 内容的解码流程[1]

4. 5. 3 BBC/NHK HLG

前述的 HDR10 与 Dolby Vision 都是基于 PQ 曲线的 HDR 标准。BBC/NHK HLG 则是基于 HLG 曲线的 HDR 标准。由于 HLG 曲线提供了对 SDR 内容的编码兼容，因此，对于广播公司来说，利用 HLG 曲线可以将 HDR 内容与 SDR 内容流使用同一个数据流进行传输，进而显著降低成本。HLG 系统已经被日本组织 ARIB 纳入 STD-B67 标准中。与 HDR10 一样，HLG 也是一种免费的 HDR 标准。

HLG 系统中使用了混合的 OETF 曲线，在传统 SDR 的较低亮度区域使用标准的 Gamma 曲线，在较高亮度区域使用对数曲线，HLG 的 OETF 曲线和传统 SDR 内容使用的 OETF 曲线如图 6 所示，有了这种混合的 OETF 曲线，就可以播放兼容 SDR 和 HDR 电视的单一的信号/视频流，并且 HLG 系统不依赖于元数据，信号与显示器是相互独立的。使用 HLG 技术不需要付版权费，也不需要经过广播公司和显示设备生产厂家的许可。

HLG 的母带制作过程非常简便，将原始的拍摄的文件利用 HLG 曲线作为 OETF 曲线，并且将色域映射到 BT. 2020 空间内，就可以制成相应的 HLG 文件。由于 HLG 文件同样严格遵循现有的 10bits 的数据流，因此 HLG 可以无缝的通过现有的广播传输方式进行传播，例如 IP、卫星等。HLG 信号被接收后，取决于显示设备的类别，HDR 设备能够将 HLG 信号转换为 HDR 视频播出，SDR 显示设备能够将 HLG 信号转换为 SDR 视频播出。

HLG 的传输与重放标准：

- 最小信号通道：HDMI 1.4
- 重放色域：BT. 2020
- EOTF 曲线：逆 HLG 曲线与 OOTF
- 元数据：无
- Bit 深度：10bits

HLG 格式的亮度也是一种相对的亮度形式，它将最大码字对应于显示设备的最大显示亮度。传统显示设备的 EOTF 并非 OETF 的逆曲线，为了补偿场景亮度，通常还需要使用 OOTF（通常为 1.1 或 1.2 的指数函数）。在 HLG 的重放过程当中，也需要相应的 OOTF 函数，才能够使得画面的对比度不会过低。HLG 中的 OOTF 是一个自动校准的指数函数。根据显示设备动态范围的不同以及场景亮度的不同，指数函数的值会发生改变。对于高动态范围的显示设备以及高场景亮度，通常会使用大的指数值，例如 1.6。而对于低动态范围的显示设备与低场景亮度，通常会选择 1.3 左右的指数值。在实际操作过程中，显示设备将收到的信号经过逆 HLG 曲线，然后计算每个像素点的相对值 Y，以表示像素点的亮度。在后续的 OOTF 校准过程中，利用相对值 Y、最高及最低亮度、场景亮度等

信息进行参数确定，最后再进行亮度映射。

利用相对值 Y 进行亮度映射的优势在于避免了颜色的改变。因为 HLG 标准中不使用元数据，因此直接进行 OOTF 处理可能会造成颜色改变。

整个 HLG 的制作与传输重放流程如下图：

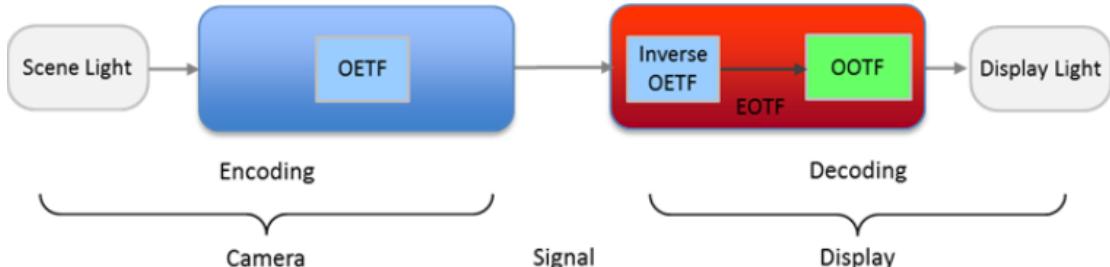


图 4-9 HLG 的制作与重放流程

4.5.4 飞利浦/特艺 SL-HDR1

特艺 (Technicolor) 开发了两种 HDR 技术：第一种是从摄像机获取 10-bit HDR 视频信号，然后生成可以兼容 SDR 和 HDR 显示器的视频信号；另一种称为智能音调管理，提供了一种将 SDR 素材“升级”为 HDR 内容的方法。在 CES 2016 上，特艺和飞利浦 (Philips) 合并了它们的 HDR 技术，即 SL-HDR1 标准。

SL-HDR1 分发系统的主要特征：

- 包含元数据的单层配置：该 HDR 系统是带有附加元数据的单层编码过程，这些附加元数据（相当于每一帧或者每个场景包含几个字节）可以用于后处理阶段，用来重建 HDR 信号。
- 编解码器无关：该 HDR 系统独立于编解码器（推荐使用 10 位编解码器）。
- 兼容 SDR：解码的视频流可以被直接显示在 SDR 显示器上。由于包含元数据，可以通过后处理将解码的 SDR 视频转化为 HDR 视频。
- 保证 HDR 内容的质量：与 HLG 系统相比，编码性能有所提升。
- 根据显示器的性能自适应调整 HDR 内容：当 HDR 内容的峰值亮度高于显示器的峰值亮度时，后处理过程可以在保留所有细节和艺术意图的前提下，将 HDR 内容的峰值亮度调整为显示器的峰值亮度。
- 附加的复杂度有限：预处理和后处理步骤只带来了有限的复杂度。
- 独立于 OETF 函数：预处理和后处理操作是在线性光域中进行的，因此与 OETF 函数是独立的。

SL-HDR1 分发系统如图 8 所示，原始内容首先要经过预处理、后处理和母片处理操作生成 HDR 母片信号和内容元数据，在编码阶段，HDR 母片信号分解为 SDR 信号、内容元数据和动态元数据，SDR 信号通过任意一种编码器（比如 HEVC）进行编码，然后通过现有的 SDR 分发网络进行传输，同时传送的还有动态元数据，在解码端可以同时获得 SDR 视频和 HDR 视频，经过一定的后处理后即可进行显示。

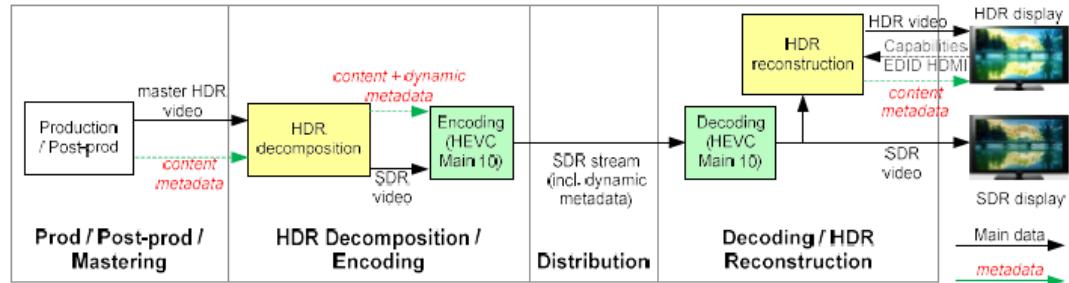


图 4-10: 单层 SL-HDR1 分发系统

图 9 中的框图更加详细地描述了 HDR 分解和重建的过程，HDR 分解的核心部分是 HDR 到 SDR 的转换，HDR 重建则是将 SDR 视频转换为 HDR 视频，视频的峰值亮度需要适应显示器的性能。

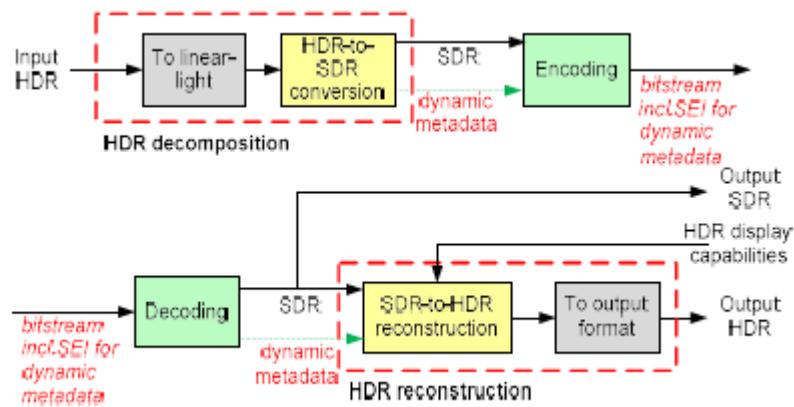


图 4-11: HDR 内容分解和重建过程

第二部分 HDR 处理技术

本书的第二部分主要关注 HDR 处理技术。HDR 图像与视频的处理通常分为以下几个方面：（1）制作，分发与显示 HDR 内容。（2）HDR 内容与对应的 LDR 内容之间的相互转化。（3）HDR 内容的质量评价标准与方法。针对以上问题，本书第二部分将分为 5 个章节。第五章将介绍 HDR 内容的生产流程。第六章将介绍 HDR 内容到 LDR 内容的转换技术-色调映射。第七章将介绍 LDR 内容到 HDR 内容的转换技术-逆色调映射。第八章将介绍色域映射，即 LDR 内容与 HDR 内容转换过程中，颜色应当如何改变。第九章将介绍 HDR 内容的质量评价标准。

第五章-HDR 生产流程

在介绍 HDR 的生产流程之前，我们先介绍下视频制作与传输的一些基本知识。

- 内容类型：
 - 直播内容（**live content**）——所谓的直播内容即没有后处理过程以及创作者意图。分发给用户的信息是实时产生并且实时制作并派发的。常见的应用节目类型包括体育类比赛，新闻等。
 - 预先录制内容（**pre-recorded content**）——该类内容并不是采集后实时处理并派发的。它们需要进行后期处理，并可能被加入创作者意图，录制的视频全部制作好之后再将其分发给用户。
- 视频服务类型：
 - 实时节目服务——实时节目服务内容包括直播内容和与线路之内容，分发时会将这类内容实时汇集并叠加 logo 等文字图形类信息，然后实时将内容派发到用户。
 - 点播服务（**VOD**）——点播服务顾名思义是根据用户的需求分发相应的内容，而不是让用户被动式的接受根据节目表分发的内容。

下图为一个传统的 SDR 视频内容制作的基本流程图。图中包括了制作，后期处理、内容分发及多类型发行等过程，同时包含了实时制作和离线制作，包含了卫星、地面广播、广域网、光盘等常见的内容分发途径，以及 IPTV、电脑、移动设备等终端显示设备。

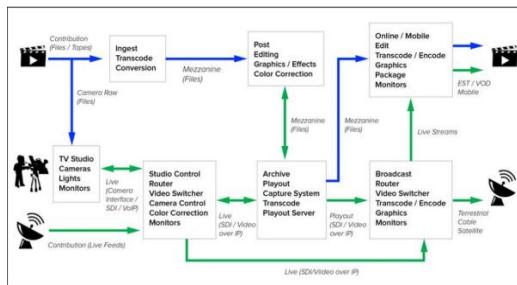
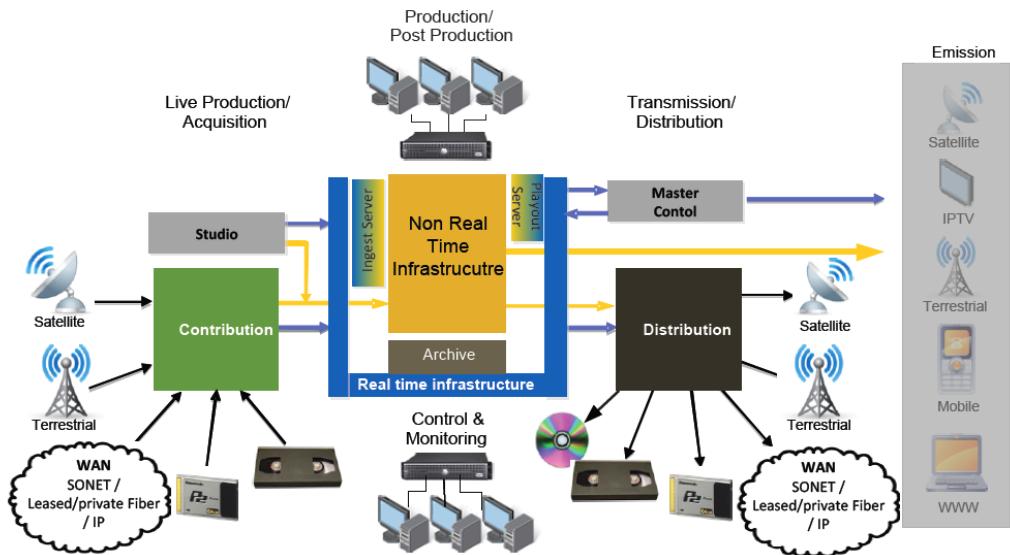
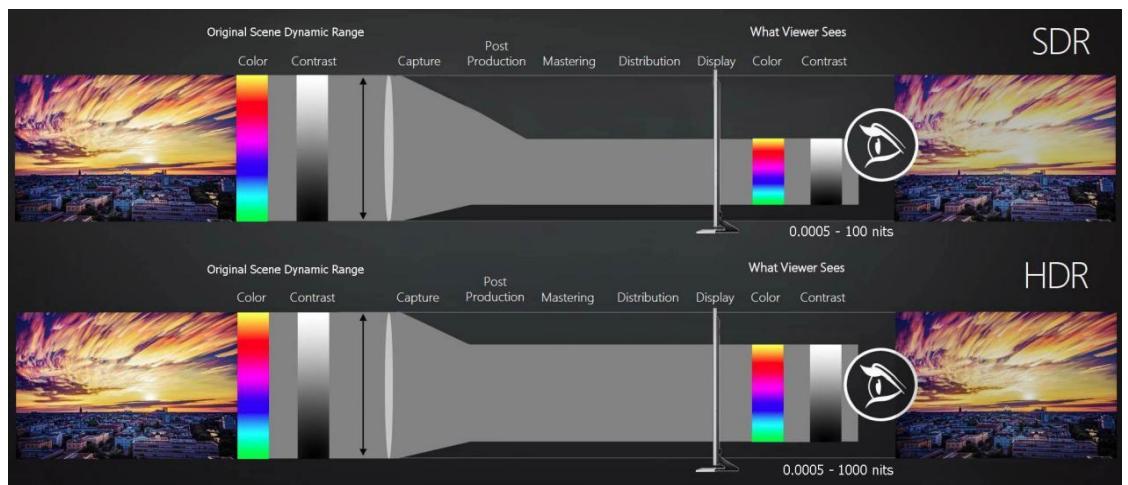


Figure 1 Content Production and Distribution Workflow

图 5-1 Ultra-HD-Forum Deployment (备用)

与 SDR 类似，从 HDR 的拍摄到最终的显示，也包括了拍摄、制作、后期处理、传输、显示等多个环节，是一个端到端的完整系统。



下图所示为兼容 SDR 的 HDR 端到端系统示意图，其描绘了 HDR 制作从输入端到头端显示的整个流程。该制作流程的输入端有三个来源，包括 HDR 摄像机、SDR 摄像机以及传统的 SDR 内容。HDR 摄像机采集到的视频为原生 HDR，通常经过简单处理和编码后就可以输送到电视机或者 STB 等进行内容呈现。SDR 摄像机拍摄的内容直接将其上变

换为 HDR 后再进行类似的处理。而对于目前占比非常大的传统的 SDR 视频源，由于其并不像 SDR 摄像机拍摄的原素材那样保留很多较多的信息，因此其需要更加复杂和精细的处理以获得质量较好的 HDR 片源。该类输入需要使用播出服务器将其进行映射或者上变换，以获得 HDR 格式的内容。其中，SDR 到 HDR 的上变换过程称为 *inverse tone mapping (ITM)*，会在本书的第七章中详细介绍。HDR 使用的是 HEVC Main10 编码器，在 SDR 编码的基础上对一些参数进行了调整，并对 HDR 进行了专门的优化处理，该部分将会在本书的第三部分分为几章进行详细描述。

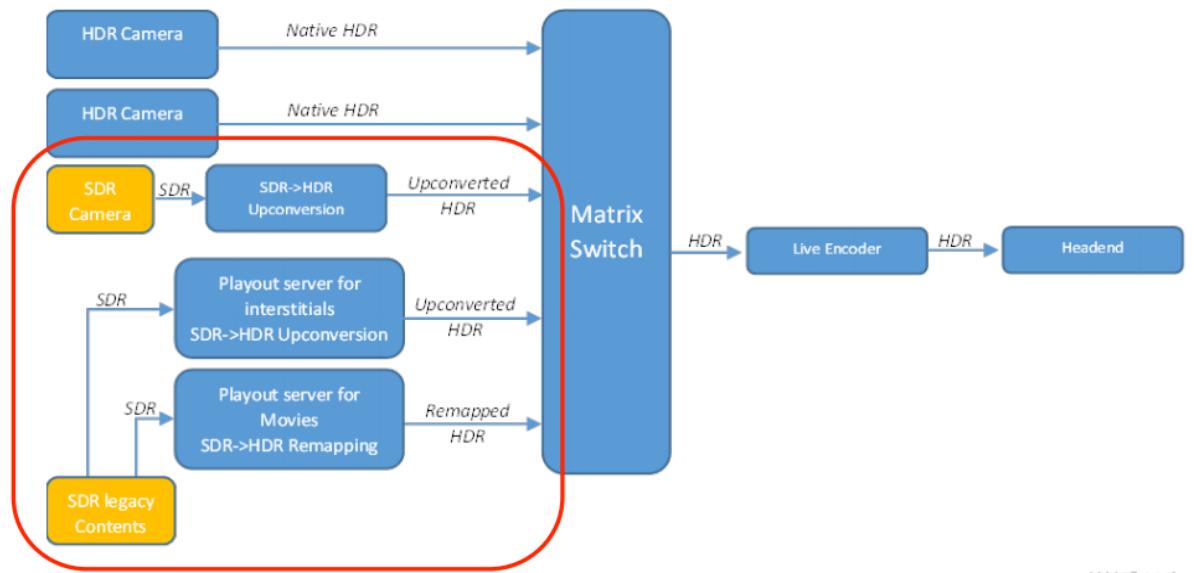


图 5-2 《UHD Forum Phase A》

5.1 实现 HDR 的条件

在介绍 HDR 的生产流程前，需要介绍下 HDR 制作的一些前置条件。首先，HDR 使用的拍摄设备与传统的 SDR 设备相比有更高的要求，需要能够达到至少 16.7 档光圈的动态范围，并且支持的色域范围应该远大于 BT.709。其次，HDR 的制作都是基于 10 比特或者 12 比特的，因此 HDR 显示面板至少要支持 10 比特或者 12 比特。

其次，为了能完美再现真实场景，显示面板的动态范围也要至少在 16.7 档光圈以上，并且绝对亮度应大于 1000 尼特，而事实上市面上普通的 LCD 显示面板的绝对亮度一般在 400 尼特左右。超高清联盟提出的“Ultra HD Premium”标准是对 HDR 显示面板的一个认证标准，可以保障消费者更够购买到真正达标的 HDR 设备。在该标准中，对 HDR 显示面板做出如下规定：图像分辨率 3840x2160，色深 10 比特，支持 BT.2020 格式的信号输入并且可以再现 DCI-P3 色域的 90%，支持 SMPTE ST2084 电光转换曲线，以及相应的峰值亮度和背光水平（规定共有两种：一种是峰值亮度高于 1000 尼特，背光水平低于 0.05 尼特；另一种是峰值亮度高于 540 尼特，背光水平低于 0.0005 尼特）。

关于内容制作，超高清联盟要求图像分辨率达到 3840x2160，色域深度不小于 10 比特，颜色空间 BT.2020，支持 SMPTE ST2084 EOTF。如果要制作用于制作数字电影发行版的电影的数字电影母版，则需要相应的提高要求。

考虑到目前 HDR 的市场占比仍然非常小，因此在 HDR 的生产流程中兼容到现有的 SDR 设备和内容是非常有必要的。这就要求需要从 HDR 的内容制作到内容分发，再到显示端都要考虑到 SDR 的兼容性。

5.2 HDR 内容制作

该部分将详细介绍 HDR 生产流程的每一部分。HDR 的拍摄一节会首先介绍一些 HDR 相关的基础知识，以便于对后续内容的理解。其次，该节中会介绍一些 HDR 拍摄常用的设备以及一些 HDR 拍摄的技巧或注意事项。在设备及制作一节中，我们会先对市面上常用的监视器，接着介绍完整的 HDR 内容制作的设备及流程，然后会详细介绍一些在缺少一些高端设备时低成本制作 HDR 视频的实践方案。内容分发会介绍不同的 HDR 内容分发方式以及 HDR 传输相关的一些方案。后期制作主要介绍一些调色相关的知识。解码和渲染则主要介绍部分 HDR 显示器以及 HDR 的适配方案等。最后，在兼容性一节中，我们会介绍 HDR 与传统 SDR 的兼容性解决方案、HDR10 与 HLG10 兼容性解决方案等。

该部分以理论知识讲解为基础，并介绍一些可行的实践方案，旨在提供一些 HDR 内容制作的实践指导。

5.2.1 拍摄

所谓的 HDR 摄影学主要是通过提升图像的局部对比度，同时将同场景下同时刻不同曝光水平的图像进行叠加以达到同时保持图像亮部和暗部细节的目的。这种方法是最早的 inverse Tone Mapping 方法，主要用于图像，一般手机的照相机中都有 HDR 的功能，其基本原理就是通过多曝光图像的叠加以获得具有 HDR 特性的图像。HDR 摄影学记录的是 SDR 图像，与 SDR 视频拍摄一样，都将真实场景中的自然光进行了动态范围压缩，该过程中丢失了部分信息，因此在动态范围、色域等方面无法与真实的 HDR 相比。虽然这种方法可以一定程度上提升图像的视觉质量，但有时得到的图像往往不自然，影响视觉体验。

HDR 视频的拍摄则与上述的 HDR 摄影学完全不同。HDR 视频拍摄的初衷是尽可能真实地记录自然场景，而不对动态范围进行压缩，以便后期制作时有较大的调整余量，也能更加真实地呈现在显示设备上。同时，HDR 视频位深一般为 10 比特或者 12 比特，位深的对图像动态范围和颜色数的提升都有很大帮助。10 比特可以记录的颜色数可以达到 $1024*1024*1024$ ，而 8 比特的 SDR 可以记录的颜色数仅为 $256*256*256$ ；而位深为 10 比特的视频的灰阶过渡效果显然要比 8 比特的 SDR 平滑。人眼可察觉的动态范围一般在 0.0001~10000 之间，当使用 HDR 摄像机记录场景，同时显示设备支持 HDR 时，那么就可以在终端真实再现自然场景。

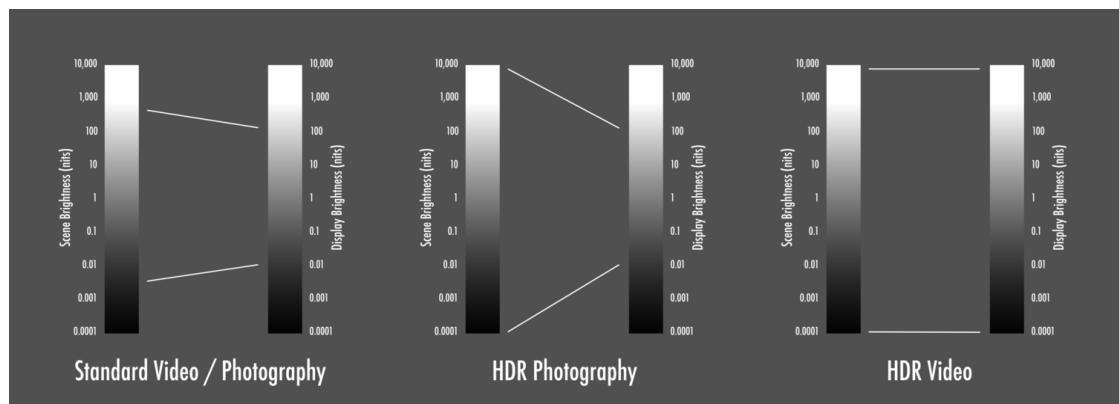


图 5-3 拍摄于显示过程中的动态范围

HDR 摄像机比较多，常用的有 Sony F65、Sony PMW-F55、Arri 的 ALEXA 65 等。HDR 摄像机首先必须支持比 BT.709 更广的色域，其次，拍摄的 HDR 素材可以采用 Log 或者 RAW 格式记录，能比线性格式保持更多的暗区细节，使得暗区更加深邃，同时保持高光区域的亮度，从而使整体画面更加真实动人。在条件允许的情况下，使用极可能多的比特数记录更多的信息，这样可以为后期处理和调色留下更大的余量。（Polyfill：如果你只有 8 比特的相机，并且想要体验 HDR 的效果，那么你可以外接 Avid DNxHR 或者 Apple ProRes 进行处理，以减小多次生成造成质量退化。因为 DNxHR 和 ProRes 会将这些 RAW 格式的序列存储为 Log 格式，同时使用 10 比特的比特深度）。需要注意的是，不同摄像机的动态范围有所不同，有些会在高亮区域分配更多的比特数，但是有些则会在暗区分配更多的比特数，因此可以根据实际的拍摄场景决定相机的使用，或者可以在拍摄时进行光补偿。

HDR 的特点，或者说是优势就是在高光区域和暗部细节的表现力。由于传感器物理特性的限制，所有相机记录曝光度的能力都是有限的，超过该限制则会被截断，不会被记录。因此，HDR 拍摄的关键就是如何根据摄像机的特性，更好的获取并且保持这些区域的信息。如果可能，现场使用 HDR 监视器能够及时的查看真实曝光情况并进行相应的调整。另外，需要注意的是高的对比度带来的噪声放大等问题也要有相应的处理。

5.2.2 设备及制作

在 HDR 专业制作流程中一般需要使用到专业级的参考显示器（监视器）。HDR 监视器的价格普遍比较昂贵，在 HDR 相关的测试中使用较多的是 DR-37P，但是市面上比较容易购买到的是 Sony 的 BVM-X300 和 Dolby PRM-4220，两者皆覆盖了 100% 的 DCI-P3 颜色空间。其中 BVM-X300 的第二代已经发售，市价在 30 万元人民币左右；Dolby PRM-4220 的分辨率为 1920*1080，屏幕大小为 42 英寸，最大亮度仅为 600nits，市场价为 \$40000。

如要将拍摄的图像渲染到 HDR 专业监视器上，需要借助一些相应的软件。此处先讲述使用相关软件从拍摄端到最终的渲染到 HDR 专业级监视器上的具体流程。

目前使用较多的是 DaVinci Resolve。DaVinci Resolve Studio 可以将 HDR 和 BT.2020 的标志注入 DeckLink 或 UltraStudio 硬件的 HDMI 输出。使用具有 HDMI 输入功能的 HDR 电视机对 HDR 进行分级时，只需将 HDMI 显示器连接起来，切换 DaVinci 设置中的选项，显示屏将自动切换到 HDR 模式：

HDR 专业监视器价格昂贵，加上相应的调色台和配套的商业软件，并不是任何团队和个人能负担得起的。那么，如何在没有 HDR 监视器的情况下仿真出 HDR 视频的效果呢？

如果只有 SDR 显示器，并且不支持 BT.2020 色域，并且想要获得 HDR 的效果，可以考虑在输出上使用 LUT 进行映射。DaVinci Resolve Studio 软件中可以使用多个 LUT，其中一个作用于自身输出，另一个作用于 HDR 监视器。作为最后的手段，您甚至可以为不同的颜色分级随意附加任意数量的 LUTs。当然，使用 Scratch 或者 Nuke 等相应的颜色分级软件也可以达到你的目的。

一个基本的离线 HDR 生产流程应当如下图所示，包含以下设备：HDR 相机、HDR 监视器、调色台及 HDR 显示器。调色台功能强大，操作方便灵活，同时赋予了调色师

最大限度的创作和调整空间。借助调色台和 HDR 监视器，你可以在较短时间内很轻松高效地制作出一个在 HDR 消费级显示器上渲染效果很好的 HDR 视频，有效地避免了不必要的中间产物和繁琐的离线反复调整查看过程。



HDR & SDR Simultaneous Production (Parallel System)

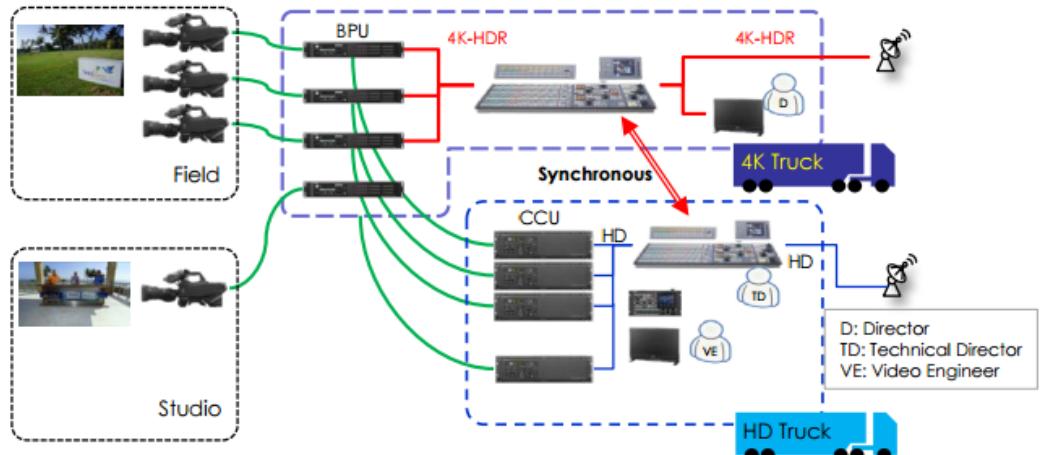


Figure 2

图 5-4 HDR 与 LDR 平行制作系统

目前市场上很多的 LCD、OLED 等显示器的最大显示亮度早已突破 100 尼特，有的可以到达 300-400 尼特，部分也已经支持 BT.2020 色域。因此我们可以使用这类高亮度的消费级的 HDR 显示器仿真 HDR 视频的效果。

既然没有调色台，也没有专业级的 HDR 监视器，仅有一些高端或者消费级的 HDR 电视机时，可以使用 ATEME TITAN KFE Job Composer 等一类商业级的编辑软件进行 HDR 视频的制作。目前该类软件已经能够较好的支持 HDR 视频制作的工作，不仅提供 HDR 相关的多种详细的编码参数设置选项，而且也可以选择和调整静态元数据，同时其可以支持 PQ 和 HLG 曲线，并且处理速度也非常可观，

一些开源的工具和软件也可以用于制作 HDR 视频。HDRTools 是由国际电信联盟电信标准化部门（ITU-T）和 ISO/IEC 提供的开源软件，将 HDR 相关的格式转换、编辑、制作和评价等相关功能集成在一起，可以满足大多数的 HDR 编辑需求。其可以接受的输

入序列的格式包括 EXR、DPX、TIFF 等，可以进行 YUV 的降采样、位深转换、缩放等，可以进行不同颜色空间的转换及不同并且提供了 mPSNR、RPSNR、MS-SSIM、VIF 及 HDR-VQM 等质量评价度量标准。

5.2.3 后期制作

后期制作一般包括调色、特效制作、声音制作、字幕制作等步骤。HDR 后期处理的工作流程其实与 SDR 的后期处理差异并不大。

后期调色的第一步是将拿到的素材通过 Gamma 校正进行线性化，然后进行颜色分级处理。颜色分级包括了颜色校正和添加创作者意图，用于改变和增强视频序列的颜色表现。颜色分级通常是调色师根据不同的观看环境进行个人意图的颜色调整，调整的过程是调色师根据自己的经验进行的。当然，ACES(the Academy Color Encoding System)也提出了一个更加统一的工作流程用于指导调色过程，同样可以适用于 HDR 的调色过程。电影制作中常用的调色系统有达芬奇调色系统、英国 Film Light 公司出品的 Baselight Eight 调色系统等。

特效制作是指通过数字影像合成技术制作出现实中不存在的事物或者实现难度很大的场景。特效制作的文件导出格式应当与调完色的拍摄素材使用相同格式，然后与后者进行叠加。

按照其存在方式，字幕一般可以分为三类：硬字幕、外挂字幕和软字幕。硬字幕是将字幕直接烧在视频中，制作完成后不可修改调整。外挂字幕是指将字幕制作成一个单独的文件，后期进行载入并可进行编辑和调整。软字幕则是将外挂字幕和视频一块打包，使用时直接选择即可。

在电影制作中，通常需要发行不同格式和版本的发行片，因此通常会引入一些方便制作的中间格式。IMF (Interoperable Master Format)，可交互模板格式，是一项 SMPTE 的媒体文件标准，用于后期发行时输出不同文件格式、分辨率或者音频和字幕等。

DI (digital intermediate)：数字中间片指的是将拍摄的电影素材进行素材编辑、色彩调整、后期合成、特效处理、字幕混合等一系列处理时采用数字化处理，其通常在拍摄完片和放映出片之间，因此被称为数字中间片。

目前常用的 HDR 母版标准主要有 HDR10、Dolby Vision 和 HLG10。同时，为了考虑兼容性问题，一般需要制作相应的 SDR 版本。这些不同格式的、考虑兼容性的相应版本都需要在后期制作中进行相应的处理。关于兼容性的问题，将会在本章第 6 小节中详细介绍。

5.2.4 内容分发

后期调色完成后，我们相当于得到了可以用于制作不同应用场景的不同格式的 IMF 文件，然后可以进一步制作成适用于影院、广播、电视等不同应用场景的视频文件。然后，将这些文件通过相应的途径分发给相应的用户群体。

一般调色后输出的成品主要包括三中类型：胶片输出、数字电影包 (DCP, Digital Cinema Package)。胶片输出主要是用于使用胶片的老电影，但在现在的数字电影时代已经近乎淘汰；DCP 输出是提供给数字影院进行放映的一种数字文件集；根据不同的应用场景，后期的视频输出通常有许多版本，而处理不同版本之间的交互是一个高成本的重复性工作。基于此，SMPTE 提出了 IMF 格式，这是一个通用的标准化的文件格式，相当

于所有交互文件的母版。IMF 的每个包就是以恶文件夹，里边存储量数据流、用于创建不同版本的合成播放列表（CPL，Composition Play Lists）、资源映射（Asset Map）、打包列表以及包含转码参数的 XML 格式的输出配置文件列表（OPL，Output Profiles Lists）。因此，后期可以通过 IMF 转码得到适应不同应用场景的文件格式（不同分辨率、帧率、码率、字幕等）。

下图描述了 IMF 在 HDR 制作流程中的作用。可以使用 IMF 导出 ProRes、MPEG2、H264、H265 等多种格式。

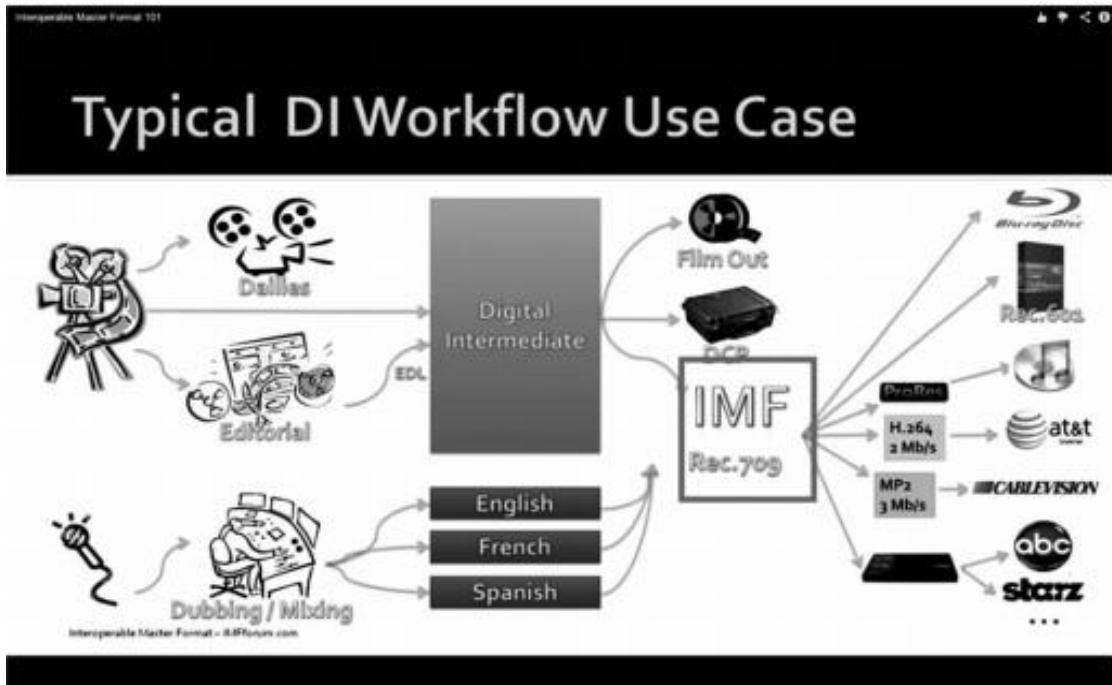
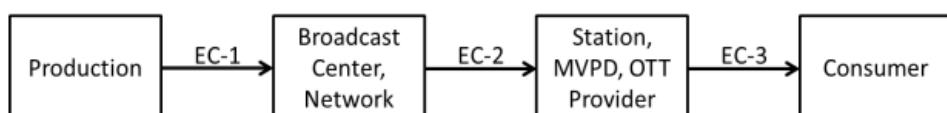


图 5-5 IMF 在 DI 流程中的作用_支持 4K 电影制作的新技术

下图所示为可以适用于节目服务和点播服务的内容分发流程图，成片制作完成后交付给广播中心或者网络媒体，然后进一步交付给多频道视频内容传输商（MVPD）、OTT 等内容服务运营商，最后通过不同的途径到达用户的终端设备（电视机、机顶盒、电脑、移动设备等）。不同的分发方式对视频的格式以及编码等要求也不尽相同。

Distribution Chain



- Guidelines covers the following at each point in the chain:
 - compression technologies
 - metadata carriage options
 - audio, captions and subtitles
 - content manipulation: ad insertion, graphic overlays, etc.
 - sample bitrate ranges

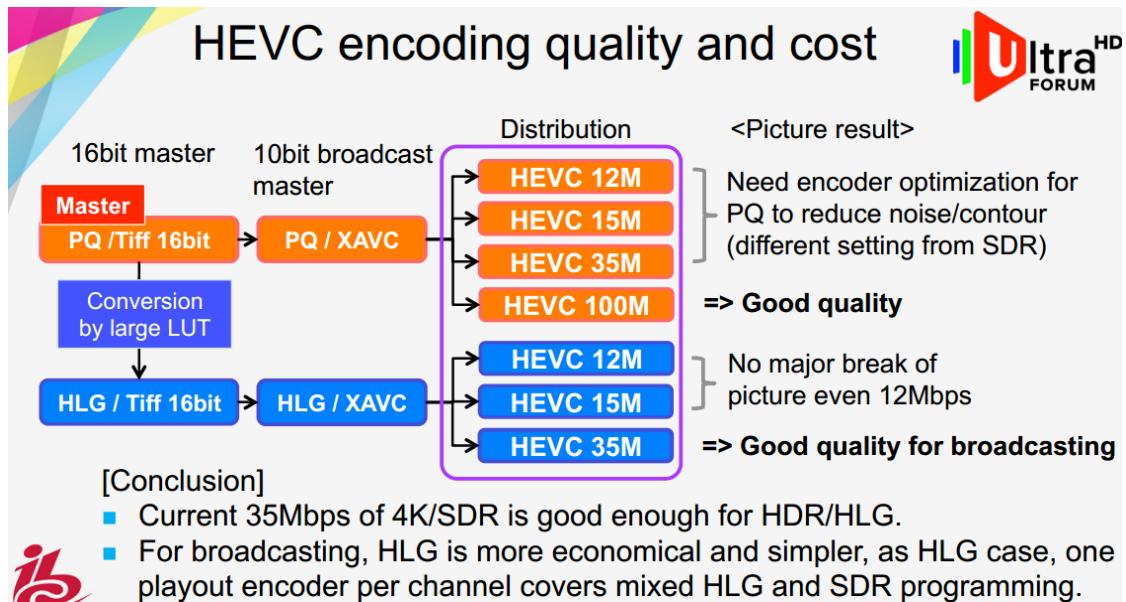


图 5-6 HEVC 编码质量与花费

5.2.5 解码&渲染

HDR 的解码和渲染主要有两种方式。一种是使用支持 HDR 解码功能的 STB，同时将其连接到电视机上；另外一种是直接将支持 HDR 解码的解码器集成到电视机中，即 HDR 电视机。

第一种方法是目前从 SDR 主导的视频生产流程平滑过渡到 HDR 的一种有效解决方案。根据统计显示，2016 年中国 IPTV 用户总数已经超过 9000 万用户，而 IPTV 服务提供的内容主要是传统的 SDR 内容。如果运营商想要提高用户体验，将部分内容升级到 HDR，那么必须要考虑到这些用户的利益。让所有 IPTV 用户直接更换 HDR 电视机，显然是成本巨大并且不宜实行的。因此，在终端使用 STB 实现 SDR 与 HDR 的兼容性可以说是一种成本较低并切实可行的方案。而这里提到的 STB 根据功能又可以分为两类，一类是支持 SDR 转 HDR 的 STB，可以将现有的 SDR 内容通过一定的处理转换为带有 HDR 效果的视频，支持 SDR 视频，可以连接 SDR 或者 HDR 电视机。另外一种则是支持 HDR 转 SDR 的 STB，支持 HDR 视频，并可以根据显示器类型自动判断是否转为 SDR。2017 年 5 月，由电信和联通及国内业界领先的多家机顶盒厂商组成的“全 4K 联盟”在上海交通大学电子系图像所进行了业内的第一次 HDR 转 SDR 芯片标准化测试，推动着 IPTV 向 HDR 方向的过渡。

第二种方案比较直接，当然代价比较高，个人用户可以按需选择。

5.2.6 兼容性

这里提到的 HDR 的兼容性涉及到两方面，一方面是由于 HDR 多标准引起的 HDR 格式兼容性问题，主要有 HDR10、HLG10 以及 Dolby Vision；另一方面是 SDR 与 HDR 的兼容性问题。

目前，国内应用较多的 HDR 格式为 HDR10 和 Dolby Vision，Dolby Vision 的双层编码方案由于涉及到商业版权及动态元数据等，目前普及程度并不高。HDR10 与 HLG10 的

转换的实质是 PQ 曲线与 HLG 曲线的相互转换，MPEG 已经有相应的解决方案[JCTVC-X0049: Conversion between PQ and Hybrid Log-Gamma (HLG)]，或者可以参考 ITU 的 BT.2390 标准。下图所示为 HLG 与 HDR10 之间的相互转换，有图可以看到，HDR10 与 HLG 的转换可以在内容分发前进行处理，当然也可以直接把这个任务交给 STB 完成。

由于兼容性问题还涉及到 HDR 的编码方案，包括传输时使用 SDR 还是 HDR，是否使用元数据以及使用何种元数据等，都与兼容性相关。因此更多关于兼容性的内容可以参见本书第三部分。

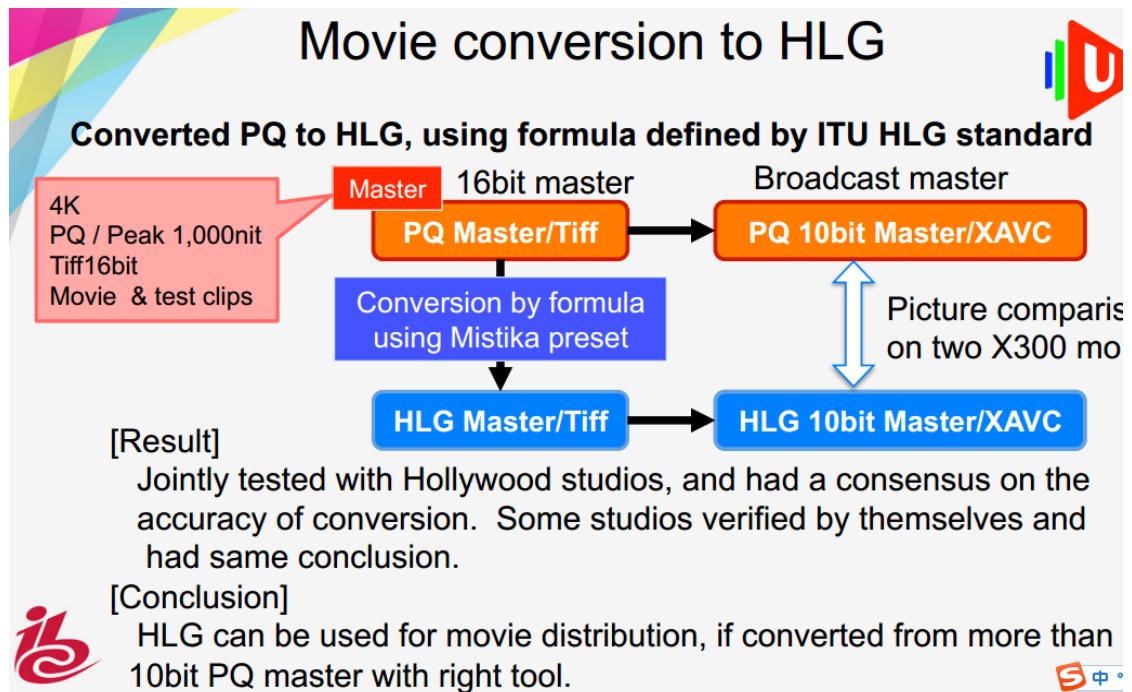


图 5-7 UHD_Forum Master

5.2.7 显示适配

由于后期处理和调色时，生成的 HDR 内容都是根据参考监视器和相应的参考观看环境制作的。HDR 监视器可以支持的动态范围更大，可以达到 0.005-4000nits 或者更大，同时支持更宽的色域。显然，普通的 HDR 显示器不管是背光水平还是峰值亮度均不能与 HDR 监视器匹敌，并且用户的观看环境迥异。如果将参考 HDR 监视器制作的内容直接让用户的 HDR 显示器显示，超过动态范围的信息都会发生截断，显然达不到理想的效果。因此，HDR 的显示特征需要相应的调整。

对于 HDR10 的内容，显示适配是通过电-电转换函数（EETF, electrical-electrical transfer function）实现的。如下图所示，HDR10 格式的视频通常按照 0 - 10000cd/m² 的动态范围制作的，通过将 EETF 曲线的暗区和高亮区域进行非线性压缩，使 EETF 由线性转为 S 型，能够使内容的动态范围映射到 0.01 - 1000cd/m²。这样不仅能够很好的保留创作者的创作意图，同时也能保持图像细节，是一个折衷的解决方案。

FIGURE 18
Example EETF From 0 - 10 000 cd/m² to 0.01 - 1 000 cd/m²

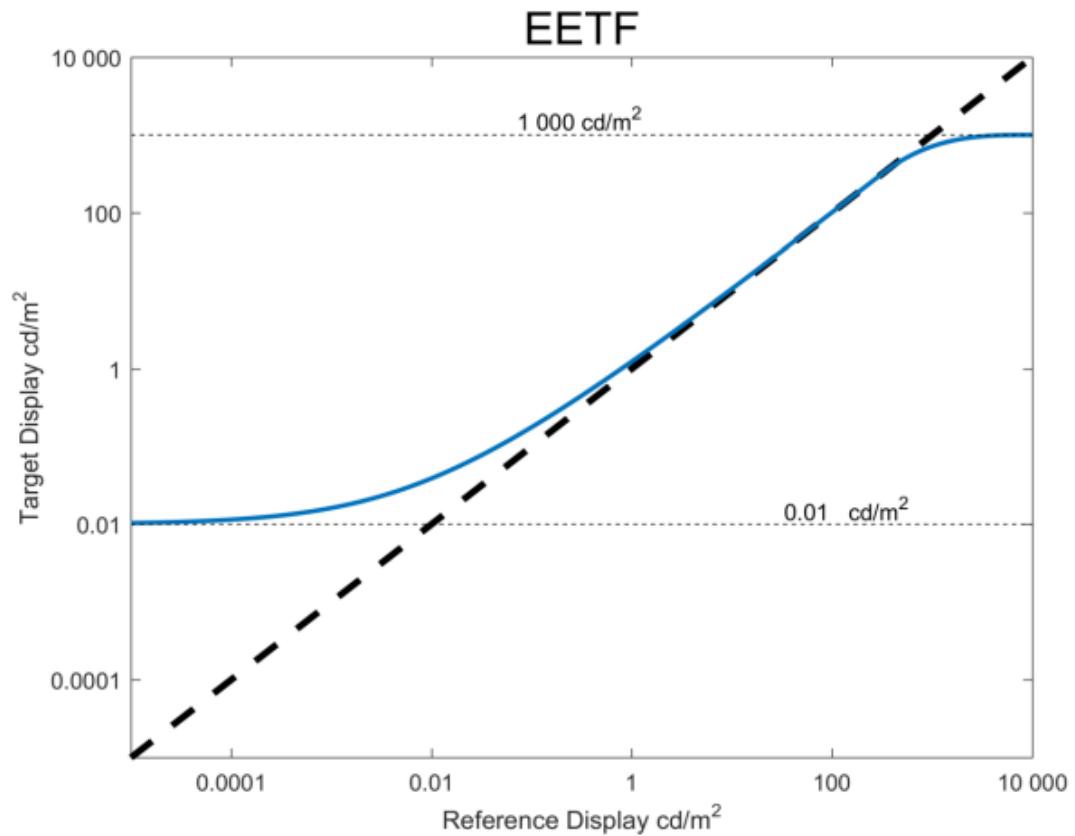


图 5-8 显示适配——BT.2390

EETF 与 EOTF 及 OOTF 的关系如下图所示。EETF 处于 OOTF 和 EOTF 的中间阶段，而 PQ 又是 display-referred，因此可以将 EETF 的映射合并到 PQ 中一块处理。关于 EETF 的计算及相关知识，可以查看 BT.2390 标准。

FIGURE 19
Block diagram of signal chain showing location of EETF application



图 5-9 EETF——BT.2390

第六章-色调映射

图像显示技术的最终目的就是使得显示的图像效果尽量接近人们在自然界中观察到的对应的场景。HDR 图像与视频有着更高的亮度、更深的位深、更广的色域，因此它无法在常见的普通显示器上显示。入门级的显示器与播放设备（例如普通人家使用的电视，常见的电脑、智能手机屏幕等）的对比度很低，只有大约 200: 1。相对性能更好的 LCD 显示器能达到更高的对比度，大约 10000: 1。但是，这些设备通常都会将一个色彩通道离散化到 8-bit，少数 10-bit 的色度区间内。这意味着色度区间只有 255 个层级，这样的显示设备，我们称之为 LDR 显示设备。很显然，HDR 图像是无法在这些显示设备上直接显示的，为了使得 HDR 图像与视频也能兼容地放映在 LDR 设备上，就必须使用色调映射（Tone Mapping）技术将 HDR 图像与视频映射为 LDR 图像与视频，进而可以正常显示在 LDR 显示设备上。

尽管近年来 HDR 显示设备的研究与生产已经逐步进入正轨，并且部分厂商生产的 HDR 电视或者屏幕已经投入市场，但是不可忽视的是，如今大多数显示设备仍然属于 LDR 显示设备，HDR 显示设备想要实现对 LDR 显示设备的全面取代，仍然需要经历一个漫长的过程。因此，色调映射技术在这样的背景下，就显得尤为重要。

色调映射的目的是使高动态范围 HDR 图像能够适应低动态范围 LDR 显示器。想要实现这一目的，就需要将 HDR 图像的色度，亮度，动态范围等，全部映射到 LDR 图像的标准范围内。在这个过程中，不能简单使用线性映射，因为这样会丢失原始图像的一些重要信息，例如全局与局部的对比度和重要的细节等，最重要的一点是，简单线性映射产生的图像相比于自然界中的场景，会出现严重的失真情况。因此，色调映射技术需要经过深入的研究与精密的设计，才能使得产生的图像在视觉效果上与实际场景中的效果一致。本章的前半部分，我们将首先对色调映射算法的整体思想与相关内容作出综述，接着通过图像色调映射算法，介绍清楚色调映射算法的基本思想与设计方法，以及常见或常用的图像色调映射算法。

随着图像色调映射算法的发展日益成熟，另一个更加复杂但是却同样具有实际意义的问题也随之出现，那就是视频色调映射算法。视频色调映射算法本质上可以看作对视频中的每一帧应用图像色调映射算法，但是，由于视频独有的时域特性以及时域联系，简单地将图像色调映射算法应用于视频当中是不合理的，并且也会使得视频质量出现很大的问题，例如帧间闪烁现象。因此，视频色调映射算法与图像色调映射算法虽然存在联系，但是也有所区别。在本章的后半部分，我们会介绍色调映射算法在视频方面的独特运用。

6. 1 色调映射算法综述

色调映射算法的目的在于将 HDR 图像的亮度进行压缩，进而映射到 LDR 显示设备的显示范围之内，同时，在映射的过程中要尽量保持原 HDR 图像的细节与颜色等重要信息。所以色调映射算法需要具有两方面的性质：1. 能够将图像亮度进行压缩。2. 能够保持图像细节与颜色。

对于 HDR 图像与 LDR 图像来说，两者的亮度是存在区别的，这种区别不仅仅在于范围上，也在于亮度的意义上。对于 HDR 图像来说，亮度值就代表绝对的亮度值。由于 HDR 显示设备的亮度范围足够大，因此，HDR 图像每一个像素点记录的亮度值，就是显示在 HDR 显示设备上的亮度值。以 OpenEXR 格式的 HDR 图像为例：OpenEXR 格式

的图像的亮度范围在 $6.14 \times 10^{-5} \sim 6.41 \times 10^4$ 。这些亮度信息记录在一个单独的亮度信道中，显示在 HDR 显示器上的亮度，就是亮度信道中记录的亮度。而 LDR 图像则有所不同，因为 LDR 显示设备的亮度范围很小，并且编码格式也是 0-255 的整数，因此能够显示的颜色与亮度的范围都很有限。为了让有限的亮度空间都能被合理利用，LDR 图像的亮度实际上并不对应于显示场景中的亮度，而是一种“相对亮度”。像素值通过 Gamma 映射，映射到显示器支持的亮度范围当中后，再显示出来。因此，LDR 图像的亮度是由显示设备决定的，而 HDR 图像的亮度记录在图像当中，由图像决定的。

由于 LDR 显示设备的亮度范围远远小于 HDR 图像以及自然界中的亮度范围。因此，在设计色调映射算法时，如果仅仅使用简单的线性映射的方法，将会产生的问题是过于明亮与过于阴暗区域的细节无法保持，这显然是不可接受的。所以，在设计算法时，必须考虑图像中场景的显示一个场景时，主要需要考虑其光照水平和对比度范围，一些需要普遍关注的例子有：在阳光充足的场景下，事物会显得更加多彩和对比强烈；而夜晚时，多彩的事物会显得灰暗的多。因此，对像素的简单缩放和压缩其光照强度水平和对比度范围不足以精确再现场景的视觉效果。Tumblin 和 Rushmeier 正式提出了这个问题，如图 1 所示，并提出了解决这一问题的视觉模型，从此视觉模型就成为了色调映射中的一部分。

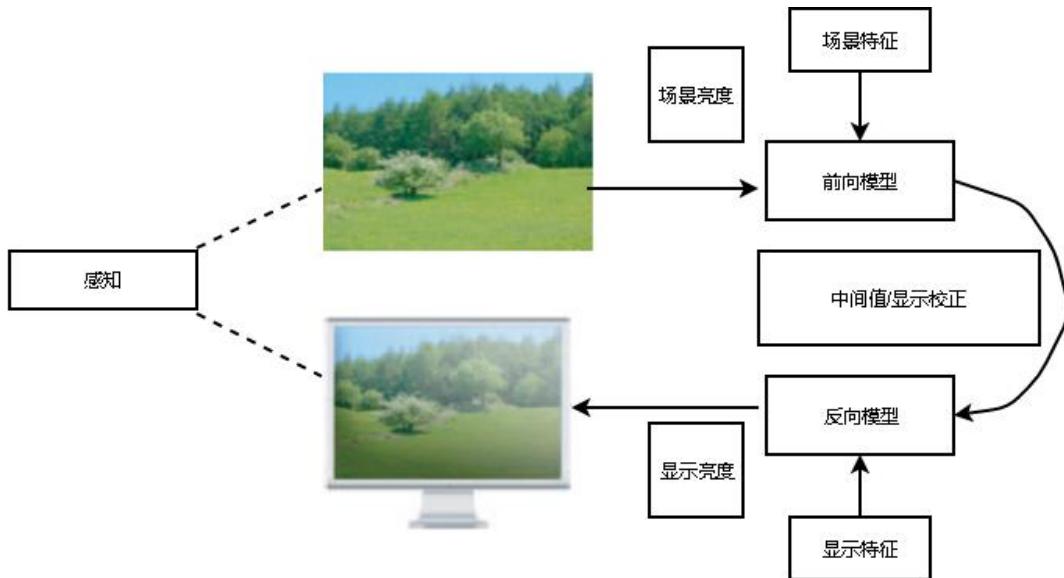


图 6-1 结合视觉模型的色调映射方案框图

根据我们上一部分介绍的人类视觉感知系统的相关内容，人类对于自然场景的感知是由视网膜上的杆状细胞和锥状细胞捕捉，并进一步通过视觉通路产生视觉信息。这些信号是由多层神经元进行非线性处理，进而形成人对于场景的感知。但是，按照我们在 gamma 校正这一节提到的人眼“视觉亮度非线性”特性，人眼对于亮度的感知与亮度的实际物理辐射亮度之间的关系不是线性映射，这就需要我们在色调映射算法设计的过程中，对这种非线性映射做出相应的处理。

自然场景有着亮度范围非常大的光照条件，从夜晚几乎没有光照、到艳阳天亮度很高的场景。如果在显示器上显示时，只是简单的将真实场景的整个亮度域线性压缩到显示器的范围内，由于“视觉亮度非线性”，图像会在明暗两端损失大量细节，人眼无法感知这一部分的细节变化，这不是场景重现希望得到的效果。人眼是如何解决这个问题的呢？当我们从一个很亮的区域，进入一个很暗的场景时，眼睛所看到的场景会出现一段时间的模糊，看不清，接着才会变得清晰。这种现象产生的原因是：实际上，虽然人

眼能够分辨的亮度域很大，但是这个亮度域中的所有亮度同时出现在一个场景中时，人眼也无法感知其中的全部细节。而人眼的处理方式是：既然无法感知整个亮度域的细节，那么就只关注局部的亮度范围。人眼通过瞳孔的调节，找到场景中亮度的主要分布范围，人眼会对于这个亮度范围内的亮度敏感，能够分辨其中的细节，而对于其他亮度，在人眼的感知中就是过于明亮或者过于阴暗，人眼就无法感知其中的细节。色调映射如前所述，也需要解决这个问题。不同的色调映射算法选择的处理方式有所不同。

色调映射技术可以使用以下的方式做一个统一的定义：

$$f(I): I_i^{w*h*c} \rightarrow D_o^{w*h*c} \quad (1)$$

其中 f 代表的是色调映射算子， I 表示待操作的图像， w, h 代表图像的宽和高， c 代表图像的通道数，通常为 RGB 三通道。同时，因为色调映射算子此时只对亮度进行操作，而不改变颜色，所以上式可以简化为：

$$f(I) = \begin{cases} L_d = f_L(L_w): I_i^{w*h} \rightarrow [0, 255] \\ \begin{bmatrix} R_d \\ G_d \\ B_d \end{bmatrix} = L_d \left(\frac{1}{L_w} \begin{bmatrix} R_w \\ G_w \\ B_w \end{bmatrix} \right)^s \end{cases} \quad (2)$$

这里 $s \in (0,1]$ 代表饱和度因子，用来降低色调映射后的饱和度，因其在色调映射过程中饱和度通常会增加。操作完成后，通常还会对每个颜色通道进行 Gamma 校正，使它的色度范围固定在 [0,255] 范围内。

从上面的式子我们可以发现，色调映射算法的关键，就是设计出适合的色调映射因子 (Tone Mapping Operator)。色调映射算法的区别，本质上是设计并使用了不同的色调映射因子。

图 2 展示了一个场景的 HDR 图片，该图片在 SDR 显示器上显示时需要进行一些本地处理。当我们对该 HDR 图片进行全局色调映射算法处理后，我们可以得到左侧图片，我们可以看出得到的图片在暗处细节损失严重，很难看清其局部细节；因此我们对 HDR 的图片同时应用全局色调映射算法和局部色调映射算法相结合的方法来渲染正确的图像，也就是图 6.1 中右侧的图片，我们可以看到不论是明亮处的内容、还是暗部细节都可以很好的呈现出来，而单独使用全局映射算法，阴影部分则无法清晰的显示出来。



图 6-2 HDR 场景示例：左侧使用全局色调映射，右侧使用全局和局部色调映射
 尽管市场上已经开始逐渐出现支持 HDR 技术的显示器和电视等，其显示能力已经越来越接近真实世界中的场景，这使得新出现的显示器能够在少量压缩 HDR 的动态范围的情况下显示 HDR 内容。但是，色调映射在 HDR 技术的普及上仍然非常重要，该技术可以很大程度上增强视频内容的效果。

6.1.1 色调映射中的颜色映射

在色调映射过程中，如我们之前所述，在色调映射的过程中，不仅仅图像亮度的范围减小了，图像的颜色范围也减小了。在色调映射的过程中，通过对三个通道进行统一的操作，进而将图像的亮度进行适当的映射；也可以分别对单个通道进行不同的操作，进而在对亮度进行映射的同时，也对图像的颜色进行映射。

由于分别对图像的三个通道进行操作，会造成颜色的饱和度降低，色调映射过程中，通常的做法是先对图像的亮度进行映射，然后再通过原图像的信息，获取新的颜色信息 (SCHLICK)。也就是说，色调映射过程分为两个步骤，(1) 亮度映射，(2) 颜色映射。

颜色映射的方式较为统一，通常使用如下式子的方式进行映射：

$$c = \left(\frac{C}{L}\right)^s T \quad (3)$$

其中， C 和 c 分别是色调映射前和色调映射后的颜色。 L 是 HDR 图像的亮度， T 是色调映射后对应的亮度值。 s 是用来调节饱和度的参数，当 $s < 1$ 时，颜色的饱和度可以得到调节。

因此，大多数的色调映射算法，都是通过这样的方式进行颜色的映射。色调映射算法的主要内容，也就变为了如何设计更好的亮度映射算法。在本节的后续部分当中，我们提到的色调映射算法，如没有额外叙述，都将表示亮度映射算法。

6.1.2 色调映射算法流水线

不同的色调映射算法有不同的关注点，也有不同的算法设计思路，想要总体地抽象概括全部的色调映射算法的性质与设计思想是十分困难的。但是，随着色调映射算法的研究逐渐成熟与深入，现在的色调映射算法通常都遵循着一种算法设计流水线。现在的大多数算法的不同，也只是体现在流水线中不同步骤的实现方式存在差异，或者额外添加了一些新的限制，但是算法的主要部分，依然遵循着这样一种流水线。在本小节当中，我们就来逐步分析与介绍这种色调映射算法流水线。

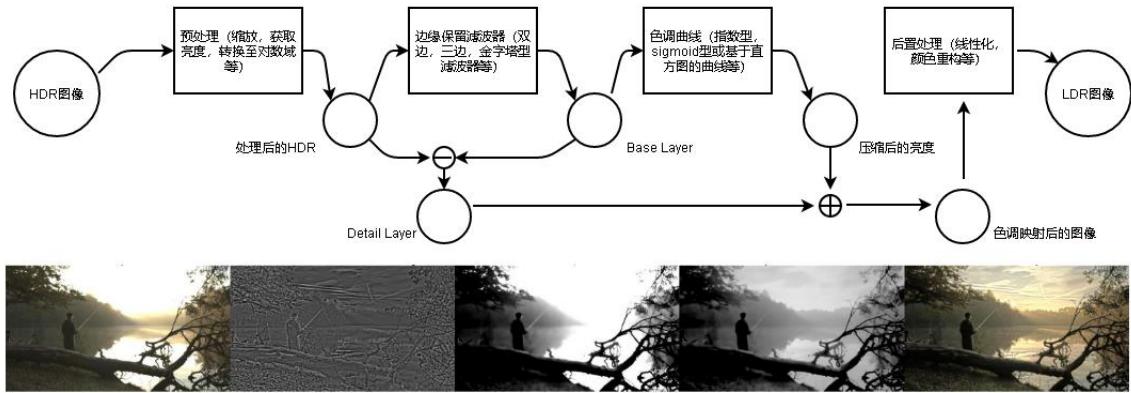


图 6-3 色调映射算法流程图

这个流水线接收 HDR 图像，输出 LDR 图像。首先，先对输入的 HDR 图像进行预处理，主要包括测算图像的亮度信息，并将亮度信息转换到对数域。接着，处理后的图像通过边缘保留滤波器（例如双边，三边，金字塔型滤波器），提取出图像的亮度信息，也就是图中的 **base layer**。然后从处理后的图像中减去 **base layer**，就可以得到 **detail layer**，也就是图像的边缘信息，或者称之为图像细节。也就是说，第二步操作，将图像的亮度信息与图像细节分开，以供后续步骤进行不同的操作。再之后，第三步，提取出的亮度信息通过不同的 **tone curve**（色调曲线）进行压缩，或称之为映射，进而映射到 LDR 图像的亮度域之中。最后，将新的，映射后的亮度与之前提取出的 **detail layer** 叠加，并通过前一节提到的方式，将图像的颜色还原添加，就得到了最终的 LDR 图像。

色调曲线

色调曲线以及局部处理（Local Processing）是色调映射算法设计的核心问题。在上面的色调映射算法流程图中我们提到，提取出的亮度信息，要通过色调曲线，映射到 LDR 图像的亮度域当中。本小节，我们就来看一下什么是色调曲线，以及色调曲线部分有哪些重要的内容。下一小节，则主要介绍局部处理。

在色调映射算法中，最核心的问题就是如何将亮度从 HDR 亮度域映射到 LDR 亮度域。流水线中的其他操作与处理，实际上都是在为这个核心目的服务。在大多数情况下，HDR 到 LDR 的亮度映射，是通过一种连续函数-亮度曲线实现的。

首先，我们公式化地描述色调映射问题。色调映射问题，就是将输入的 HDR 图像 $L \in R^+$ 映射为 LDR 图像的像素值 $I \in N_d = \{0, 1, 2, \dots, 255\}$ 。转换的方式我们记做 $M: L \rightarrow I$ 。 L 亮度的计量单位或者存储在图像中的亮度值的计量单位不会影响算法的设计，只要这些亮度与记录的对应场景的亮度之间是线性相关的即可。但是，不同的色调映射算法的设计存在不同，有些算法直接将 L 的亮度值映射为 0-255 的像素值，而有些则将 L 的

亮度值映射为对应的亮度信息 T , 然后由显示设备将亮度信息 T 映射为 LDR 图像的像素值。针对这种情况, 我们给出色调曲线的定义: 我们将色调曲线记做 $V: L \rightarrow T$, 其中 $T \in R^+$ 。也就是说, 色调曲线是 $R^+ \rightarrow R^+$ 的映射, 并且是一段连续的映射, 它实现了亮度域的压缩转换。而为了保证算法的完备性, 色调映射算法还应当在色调曲线之后, 补充亮度信息 T 到 LDR 图像像素值 I 的映射。

色调曲线函数 V 的设计取决于一些自定义的参数 $\alpha = \{\alpha_1, \dots, \alpha_{K_\alpha}\}$, 例如整体亮度, 对比度, 饱和度; 同时, 也取决于一些图像的数据 $\phi(L) = \{\phi_1(L_1, \dots, L_N), \dots, \phi_{K_\phi}(L_1, \dots, L_N)\}$ 。图像数据可以包括图像均值: $\phi = \mu = \frac{1}{N} \sum_s L_s$, 图像的最大, 最小值以及图像的直方图 $\phi_i = p_L(L_i)$ 。

利用 α 与 ϕ , 我们就可以给出空间不变的像素级的色调曲线的定义: $T_s = V(L_s; \phi(L); \alpha)$ 。其中 s 表示图像中的像素。

例如, 一个简单的曝光补偿可以描述为 $T_s = V(L_s; \phi(L); \alpha) = \alpha L_s$ 。另一个例子是 Sigmoid 曲线, 它可以适应性地调节图像整体亮度, 假设输入图像的均值 $\phi = \mu$, 此时的色调曲线就可以记做: $T_s = V(L_s; \phi(L); \alpha) = \frac{L_s^\alpha}{L_s^\alpha + \phi^\alpha}$ 。

上述的两个例子实际上就是简单的色调映射算法, 它们的效果虽然并不好, 但都实现了亮度域从 HDR 到 LDR 的转换。

局部适应

在上一节中提到, 色调曲线可以压缩图像的动态范围, 或者说亮度范围, 但是, 在压缩的过程中, 会使得图像整体的对比度出现损失。由于全局对比度在亮度被压缩到较小的亮度范围时不可避免的会出现损失, 提升图像的品质就可以从保持图像的局部对比度, 或者说细节入手。

我们同样公式化地描述局部对比度。类似于色调曲线中提到的 $\phi(L)$, 我们使用一系列参数 $\xi_s(L)$, 来描述局部对比度。这些参数 $\xi_s(L)$ 取决于图像 L , 并且它们的取值取决于像素 s 周围的领域 Ω_s : $\xi_s(L) = \xi(L_{\Omega_s}) = \{\xi_1(L_{\Omega_s}), \dots, \xi_{K_\xi}(L_{\Omega_s})\}$ 。通过 $\xi_s(L)$ 来获取像素点 s 附近的局部特征, 色调映射算法可以在空间上自适应的维持或者增强局部对比度。局部的色调曲线现在可以记做如下形式: $T_s = V(L_s; \xi_s(L); \phi(L); \alpha)$

也就是说, 利用 $\xi_s(L)$, 可以有效地获取图像的局部信息。因此, 考虑我们色调映射流水线中的思想, 我们可以利用 $\xi_s(L)$ 获取图像的 base_layer, $B_s = \xi_s(L)$, 以及 detail_layer, $D_s = L_s - B_s$ 。利用色调曲线来处理 base_layer, 然后将 detail_layer 加回图像中, 由此来保持图像的局部对比度:

$$\begin{aligned} T_s &= V(L_s; \xi_s(L); \phi(L); \alpha) \\ &= \tilde{V}(B_s; \phi(L); \alpha) + D_s \\ &= \tilde{V}(\xi_s(L); \phi(L); \alpha) + L_s - \xi_s(L) \end{aligned} \quad (4)$$

这里的 V 表示局部色调曲线, 而 V 可以被分解为两个部分, 一个是用于处理全局图像对比度 B 的全局色调曲线, 第二个是细节 D , 用于保留细节到最后。

6.2 图像色调映射算法

图像色调映射算法可以分为全局（空间不变）算法和局部（空间变化）算法，他们的建模方式的区别在于部分算法只基于图像全局特性进行调整，另外一部分算法结合图像的全局和局部的视觉特性进行不同的调整。全局色调映射算法对整幅图像的所有像素点采用相同的转换函数，也就是说整幅图像中的所有相同的像素点，在转换后的像素值也是相同的，它们可以是幂函数、对数函数、**Sigmoid** 或者是基于图像内容的函数。局部色调映射算法在图像中不同的空间区域采用不同的转换函数，这种情况下，映射前颜色相同的区域，在映射后的颜色可能不同，这与其所在位置和周围的像素点值有关。总的来看，全局色调映射算法相对于局部算法来看，有着更快的速度，因局部算法有着更高的计算复杂度。全局算法中，查找表可以更快的处理图片，因此更适合用于照相机中和视频处理。

全局色调映射方法适用于动态范围与显示设备支持的动态范围相近的场景，或者更低。当场景的动态范围远远超过显示设备的显示能力时，全局色调映射方法过度压缩色调范围，导致对比度和可视细节上的无法避免的损失。因此，除了全局色调映射算法外，我们还需要引入局部处理使其具有更好的视觉效果。局部处理算法允许增加局部对比度，从而提高图像某些部分的细节可见性，而全局处理算法将图像的整体动态范围缩放到显示设备支持的动态范围。

6.2.1 全局色调映射

使用全局色调映射算法，就意味着对整幅图像的所有像素点采用一个相同的映射函数来进行处理。这样做的好处是，对所有的像素点进行相同的操作可以保留整幅图片的全局对比度。映射函数有时可能会对输入的图像先做一次处理来计算得到一些重要的全局信息，然后使用计算得到的全局信息来进行色调映射。色调映射算法中常用的全局信息包括最大亮度、最小亮度、对数平均值和算数平均值等。为了增强算法的鲁棒性并减少结果中的异常值，这些全局信息一般使用百分数计算，尤其是最大亮度和最小亮度信息。在时域上应用全局映射算法也是十分直观的，在多数情况下我们需要暂时过滤计算完成的图像数据，因为这可以避免因为序列中的结构不连续导致的闪烁问题。全局映射算法的主要缺陷在于因为操作过程中使用的是图像的全局信息，因此无法获得图像的全局对比度以及原始 HDR 图像中良好的细节。

全局色调映射方法的设计通常遵循两种不同的思路：第一种是从传统的数字图像处理的方法出发，通过计算得到图像的某些全局信息，来对像素点进行操作；第二种是从人类视觉感知系统的某些特性出发，通过人类视觉感知系统的某些结论，来设计色调映射算法。基于数字图像处理方法是最先应用到色调映射算法当中的，而随着人类对视觉的研究越发深入，基于人类视觉感知系统的色调映射方法才逐渐进入人们的视野。通常情况下，基于人类视觉感知系统的色调映射算法得到的结果要更加优秀。

本节中，我们将首先按照介绍两种基于数字图像处理的算法：简单映射算法以及直方图校准算法。接着，我们将介绍两种主要的基于人类视觉感知系统的算法：视觉适应模型算法以及变视觉适应算法。这些方法各有优劣，并且已经很少在实践中使用了，但是这些方法的设计思路以及应用到的原理，仍然能够带给我们一些启发，因此本节介绍的这些算法中，我们不用过于关注具体实现以及数学推导，我们需要关注的重点是这些算法的设计思路与想法。

1. 简单映射算法

我们介绍的第一种基于数字图像处理的色调映射算法称为简单映射算法。全局色调映射技术是将 HDR 图像的每一个像素点通过同一个映射因子映射到 LDR 图像的范围内。因此，本质上来说，色调映射就是两个图像空间的映射，而最简单也是最直观的映射方式就是使用一些基本初等函数进行映射。基于基本初等函数的色调映射算法就被称为简单映射算法。

简单映射算法中使用的基本函数通常为线性函数、对数函数和指数函数。尽管这些基本函数可以简单而快速的实现整个映射过程，但他们却无法将图像的动态范围准确的压缩。线性曝光是显示 HDR 图像的一种非常直接的方法，初始图像被乘以一个因子 e ，就像数码相机处理曝光度的方法一样：

$$L_d(x) = eL_w(x) \quad (5)$$

使用者可以根据关注的信息来选择 e 的取值，当 $e = \frac{1}{L_{w,max}}$ 时，上式被称为标准化

形式，此时得到的结果是图像会变得很暗。当 e 的取值能使图像中有着最多的曝光良好的像素点时，上式可以被称为自动曝光形式。但是，一个简单的线性变换无法很好的完成动态范围的压缩任务，因为线性变换只能良好的显示其中一小部分的信息。

对数映射则是利用对数函数对 HDR 图像进行映射操作，对数映射将会以 HDR 图像中的最大值为标准，将整幅图像非线性地映射到 $[0,1]$ 区间内，其映射函数可以写为：

$$L_d(x) = \frac{\lg(1+qL_w(x))}{\lg(1+kL_{w,max})} \quad (6)$$

其中 $q, k \in [1, +\infty)$ 是由使用者定义的参数，这两个参数可以决定映射算法的具体表现。

指数映射是利用指数函数对 HDR 图像进行映射操作，指数映射将图像各像素点的值通过由指数函数构成的函数映射到 $[0,1]$ 区间内，具体的映射函数定义如下：

$$L_d(x) = 1 - e^{\left(-\frac{qL_w(x)}{kL_{w,H}}\right)} \quad (7)$$

同样， q, k 也是使用者需要调节的参数。使用上述几种方法得到的效果如图 4 所示。

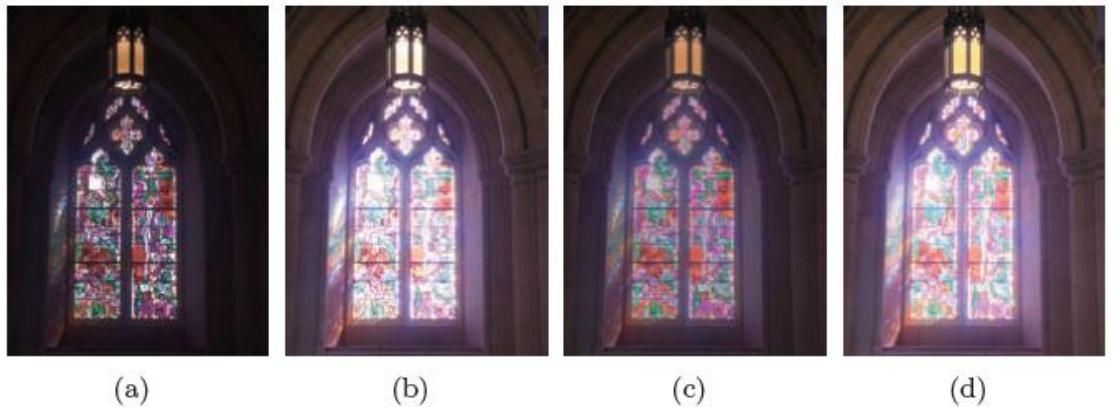


图 6-4 简单映射方法的结果图示例：(a) 标准化形式；(b) 自动曝光形式；(c) 对数映射，其中 $q=0.01, k=1$ ；(d) 指数映射，其中 $q=0.1, k=1$ 。

对数映射和指数映射都可以在处理出于中段动态范围的内容时，获得很好的效果。但是，这两种方法在计算整幅 HDR 图片时却效果欠佳，这将导致图像过于明亮或过于

阴暗，全局对比度也会有失真并且图像细节不够自然。

2. 直方图校准

最后一种基于数字图像处理的算法是直方图校准算法，在这种算法中，也加入了一些人类视觉感知系统的理论应用。**Larson** 将传统的直方图均衡技术进行了修改和调整，使它能够用于色调映射技术上，**Larson** 同时也在其映射算法中模拟了人类视觉系统的一些特性。首先，这种方法要计算输入图像的灰度直方图 I ，并在对数域内使用二进制数 n_{bin} 表示。**Larson** 通过实验证明，最多需要 100 个二进制数就足够准确地表示结果，此时，累计直方图 P 表示如下：

$$P(x) = \sum_{i=1}^x \frac{I(i)}{T}, T = \sum_{i=1}^{n_{bin}} I(i) \quad (8)$$

其中 x 是二进制数，这里需要注意的是累计直方图是一个积分形式，而灰度直方图是它在适宜尺度下的导数：

$$\frac{\partial P(x)}{\partial x} = \frac{I(x)}{T \Delta x}, \Delta x = \frac{[\log(L_{w,max}/L_{w,min})]}{n_{bin}} \quad (9)$$

随后，灰度直方图需要均衡化，传统的均衡化对比度方法如下：

$$\log(L_d(x)) = \log(L_{d,min}) + P(\log L_w(x)) \log(L_{d,max}/L_{d,min}) \quad (10)$$

这种操作因为只利用了很少的点进行区域内的色度域压缩并用最中间的点进行拓展，所以会引起图像中一大片区域的对比度失真，因此他们使用了一种直接的方法，可以用下式表示：

$$\frac{\partial L_d}{\partial L_w} \leq \frac{L_d}{L_w}$$

综合上面的三个式子，我们可以得到：

$$e^{\log(L_d) \frac{f(\log(L_w))}{T \Delta x} \frac{\log(L_{d,max}/L_{d,min})}{L_w}} \leq \frac{L_d}{L_w} \quad (11)$$

上式可以简化为：

$$f(x) \leq c, c = \frac{T \Delta x}{\log(L_{d,max}/L_{d,min})} \quad (12)$$

因此，当上式的条件无法满足时，对比度的失真就产生了，解决方法就是截断 $f(x)$ ，不过这种操作需要迭代的进行，以避免改变 T 和 c 。这种算法引入了一些模拟人类视觉系统的机制，例如对比度、锐度和颜色敏感度等，这些都是受启发于 **Ferwerda** 之前的工作。

总的来说，这种色调映射算法提供了一种改进 HDR 图像的直方图均衡方法，可以提供效果更好的动态范围压缩以及良好的整体对比度。

3. 视觉适应模型

Ferwerda 提出了一种视觉适应模型色调映射算法，这种算法根据人类视觉系统的适应机制，对人类视觉系统的部分方面进行建模，如阈值可见性、颜色感知、视觉灵敏度和时间灵敏度等，通过分段函数来实现对于视杆细胞 (T_p) 和视锥细胞 (T_s) 的建模，模型函数可以定义为：

$$\log_{10} T_p(x) = \begin{cases} -0.72 & \log_{10} x \leq -2.6 \\ \log_{10} x - 1.255 & \log_{10} x \geq 1.9 \\ (0.249 \log_{10} x + 0.65)^2 - 0.72 & \text{otherwise} \end{cases} \quad (13)$$

$$\log_{10} T_s(x) = \begin{cases} -2.86 & \log_{10} x \leq -3.94 \\ \log_{10} x - 0.395 & \log_{10} x \geq -1.44 \\ (0.405 \log_{10} x + 1.6)^{2.18} - 2.86 & \text{otherwise} \end{cases} \quad (14)$$

映射过程是对每个颜色通道使用一个简单的线性方程来重建光照效果，同时加入了一个用于适应暗条件的消色差项，因此方程可以表示为：

$$\begin{bmatrix} R_d(x) \\ G_d(x) \\ B_d(x) \end{bmatrix} = m_c(L_{da}, L_{wa}) \begin{bmatrix} R_w(x) \\ G_w(x) \\ B_w(x) \end{bmatrix} + m_r(L_{da}, L_{wa}) \begin{bmatrix} L_w(x) \\ L_w(x) \\ L_w(x) \end{bmatrix} \quad (15)$$

其中 L_{da} 表示显示屏的亮度适应范围， L_{wa} 表示图像的亮度适应范围， m_r, m_c 是两个取决于分段函数的权重因子，通过以下等式定义：

$$m_r(L_{da}, L_{wa}) = \frac{T_p(L_{da})}{T_p(L_{wa})}, m_c(L_{da}, L_{wa}) = \frac{T_s(L_{da})}{T_s(L_{wa})} \quad (16)$$

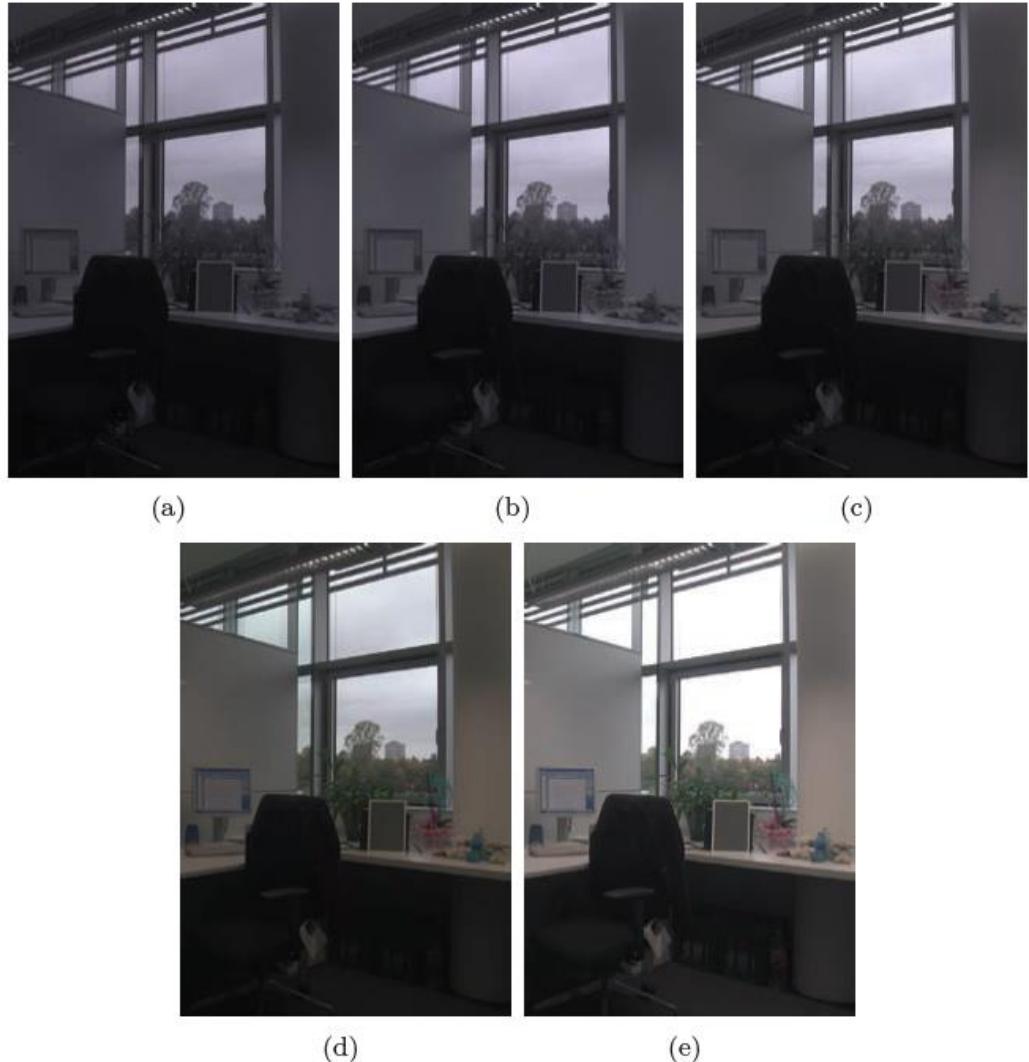


图 6-5 Ferwerda 的方法下不同平均亮度设定时的结果。(a)0.01 cd/m²; (b) 0.1

cd/m²; (c) 1 cd/m²; (d) 10 cd/m²; (e) 100 cd/m²。

当然，本色调映射方法提供了多种引入人类视觉系统仿真的想法，在缩放动态范围的同时，通过对人类视觉系统部分特性的建模，实现了暗环境细节在 LDR 显示器上显示时的适应能力，但是其动态范围的减少是通过一个简单的线性缩放，因此无法达到更好的动态范围压缩效果。

6.2.2 局部色调映射

相比于全局色调映射因子，局部色调映射因子更能提高色调映射图片的质量，因为局部色调映射因子不仅着眼于全局的对比度构建，也着眼于局部的对比度构建。映射因子 f 通过将正在映射的像素点，以及这个像素点周围的像素点的强度值同时纳入计算，进而得到映射后的图像的像素值。尽管局部色调映射方法在理论上产生的图像效果要好于全局色调映射因子，但是，在局部色调映射算法的实际应用中通常会出现一个问题，进而影响图像的质量。这个问题就是光晕现象。

在局部色调映射算法的设计过程中，如果周围的像素点的选择不够好，进行的映射方法的设计不够好，就有可能在图像某些区域的边缘出现明显的光晕。尽管在一些时候使用者希望认为产生光晕，因为这可以提醒人们去注意某一特定区域，但是，由于光晕的产生无法被控制，并且通常情况下会给图像效果带来不良影响，因此，在实际算法设计过程中，通常都需要考虑如何避免光晕的产生。

局部色调映射算法的理论基础较为复杂。大多数局部色调映射算法都是综合了很多人类视觉感知系统的结论或是基于相关研究而得出的。在本节我们将给出四种经典的局部色调映射算法，我们在了解算法的同时，也应当注意在这些算法中是如何消除光晕的。

1. 空间不均匀缩放

Chiu 最先提出了一种保持局部对比度的方法。这种色调映射因子通过一个像素点周围的其他像素点的均值来衡量这个像素点的亮度。定义如下：

$$L_d(x) = L_w(x)s(x) \quad (17)$$

这里的 $s(x)$ 是用来衡量周围像素点局部平均值的测量函数。定义如下：

$$s(x) = (k(L_w \otimes G_\sigma)(x))^{-1} \quad (18)$$

其中 G_σ 是一个高斯滤波器， k 是用来衡量最终结果的常数。这种色调映射算法存在的一个问题是如果 σ 过小，那么产生的图像的对比度就会很低，效果不好；如果 σ 过大，那么产生的图像中会出现光晕。光晕通常会出现在明亮区域与阴暗区域的交界处，这意味着 $s(x) > L_w(x)^{-1}$ 。为了减轻这种情况带来的影响，我们将在 $s(x) > L_w(x)^{-1}$ 时，把 $s(x)$ 的值固定为 $L_w(x)^{-1}$ 。在 $s(x) > L_w(x)^{-1}$ 的点上， s 仍然会有些人为操作的痕迹在里面，主要的表现形式是会有陡坡的出现。一个解决的方法是使用一个 3×3 的高斯滤波器来迭代地平滑 $s(x)$ 。最后，该算法再使用一个低通滤波器来掩盖那些引人注意的人为产生的光晕。

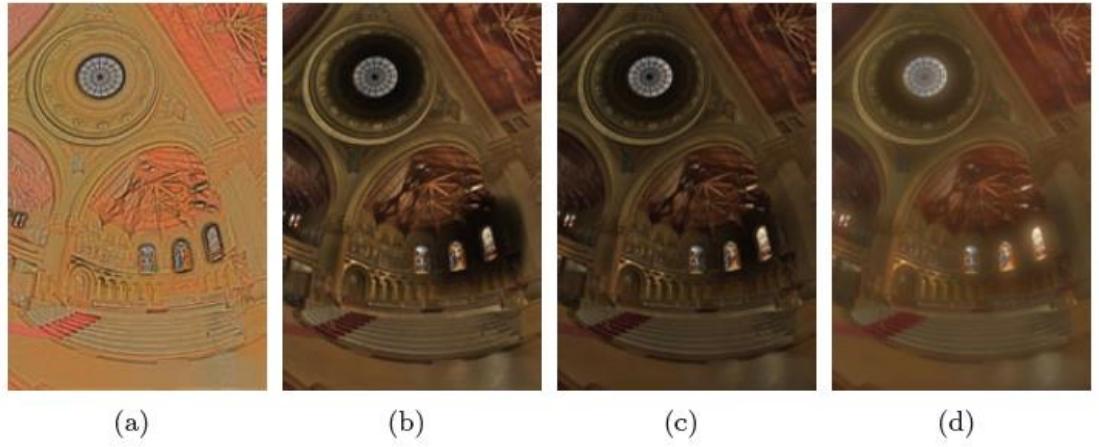


图 6 Chiu 局部色调映射算法的示例。

如图 6-6 所示, 是本方法的结果图示例, 其中(a)是 $\sigma = 3$ 的简单算法, 可以看出局部信息保留着, 但全局信息丢失严重; (b)代表 $\sigma = 27$ 的简单算法, 这种情况下局部信息和全局信息都被保留了, 但图中的光晕非常明显; (c)是 $\sigma = 27$ 并固定 s 值, 光晕有所消减; (d) $\sigma = 27$ 并有眩光模拟, 消除了大量光晕。

这是人们提出的第一个局部色调映射操作因子, 但是在减轻光晕带来的影响时, 这种操作因子产生的计算代价过于高昂(在平滑阶段大约需要进行 1000 次迭代)。有太多的参数需要去被调整, 而且在一系列的操作之后, 光晕仍然只能被减轻, 而不能被完全消除。

2. 摄影学色调重现

Reinhard 提出了一种基于摄影原理的局部色调映射因子, 这种方法模拟了摄影技术中使用了超过一个世纪的 burning&dodge 效应, 这种算法灵感来源于 Adams 提出的 Zonal 系统。这种算法的全局分量主要对高亮度的部分进行压缩:

$$L_d(x) = \frac{L_m(x)}{1+L_m(x)} \quad (19)$$

其中 L_m 是对 $aL_{w,h}^{-1}$ 进行缩放的原始亮度, a 是选定的曝光度, $L_{w,h}$ 是场景键值的对数平均数估计值。键值主观地认定该场景是明亮的、正常的还是暗的, 并且会被用在区域系统中来预测一个场景亮度是如何映射到打印区域中的。注意到在上式中, 高亮度被压缩, 而其他亮度则是被线性缩放的。但是, 上式并不会出现明亮的区域变得更亮, 但在摄影中, 摄影人很可能会为了强调某些场景而加大曝光, 加大对比度。因此, 上式可以被修改成下面这种形式:

$$L_d(x) = \frac{L_m(x)(1+L_{white}^{-2}L_m(x))}{1+L_m(x)} \quad (20)$$

这里的 L_{white} 表示会被映射到白色的最小亮度值, 默认情况下它与 $L_{m,max}$ 相等, L_{white} 作为截断值, 会将超过该值的像素点值截断为 L_{white} 。



图 6-7 Reinhard 摄影学色调重现的示例。(a) $\Phi = 4, \alpha = 0.05, L_{white} = 10^6 cd/m^2$; (b) $\Phi = 4, \alpha = 0.05, L_{white}$ 与窗口亮度接近。

通过之前的叙述可以定义一个局部色调映射算法,具体的操作方式是找到一个最大的没有明显边界的局部区域,这样可以避免产生光晕效应。这样的区域可以通过比较不同尺寸的高斯滤波之后的图片 L_m ,如果其差别很小甚至趋近于零,那就说明没有明显边界,否则有明显边界,判别方程如下:

$$\left| \frac{L_\sigma(x) - L_{\sigma+1}(x)}{2^\Phi a \sigma^{-2} + L_\sigma(x)} \right| \leq \varepsilon \quad (21)$$

当所有像素点的最大 σ 都满足上式,也就是都没有明显边界时,全局操作因子就会被更改为局部操作因子,具体如下:

$$L_d(x) = \frac{L_m(x)}{1 + L_{\sigma_{min}}(x)}$$

$$L_d(x) = \frac{L_m(x)(L_{white}^2 L_m(x))}{1 + L_{\sigma_{min}}(x)} \quad (22)$$

式中的 $L_{\sigma_{min}}(x)$ 是图像像素周围最大区域 σ_{max} 的平均亮度值。摄影学色调重现是一种保留了边界效应,并且避免了光晕效应的局部色调映射算法,除此之外它还有个优势,不需要输入校正后的图片。

6.2.3 频域算法

前面两节介绍的全局与局部色调映射算法可以统一地看做在空间域上,对图像进行色调映射的操作,本节,我们将介绍两种在频域下对图像进行色调映射操作的方法:快速双边滤波算法以及梯度域压缩算法,以供读者参考。

基于频域的算法具有与局部映射算法相同的目标,也就是保持边缘和局部对比度的目标,在频率运算符的情况下,通过在频域下而不是空间域下的计算来实现。这种方法

当且仅当大型特征和细节直接完全分离时，边缘信息和局部对比度信息才能得以保留。因此，频域算法的主要任务，实际上是将图像滤波出高频与低频部分，然后分别进行处理。

1. 快速双边滤波算法

双边滤波（Bilateral Filter）是一种非线性的滤波，可以将图像分离为高频图像，被称为细节层，并将保留边缘的低频图像成为基本层。Durand 和 Dorsey 利用这个特性设计了一个通用且高效的色调映射框架，保留了局部对比度特性。

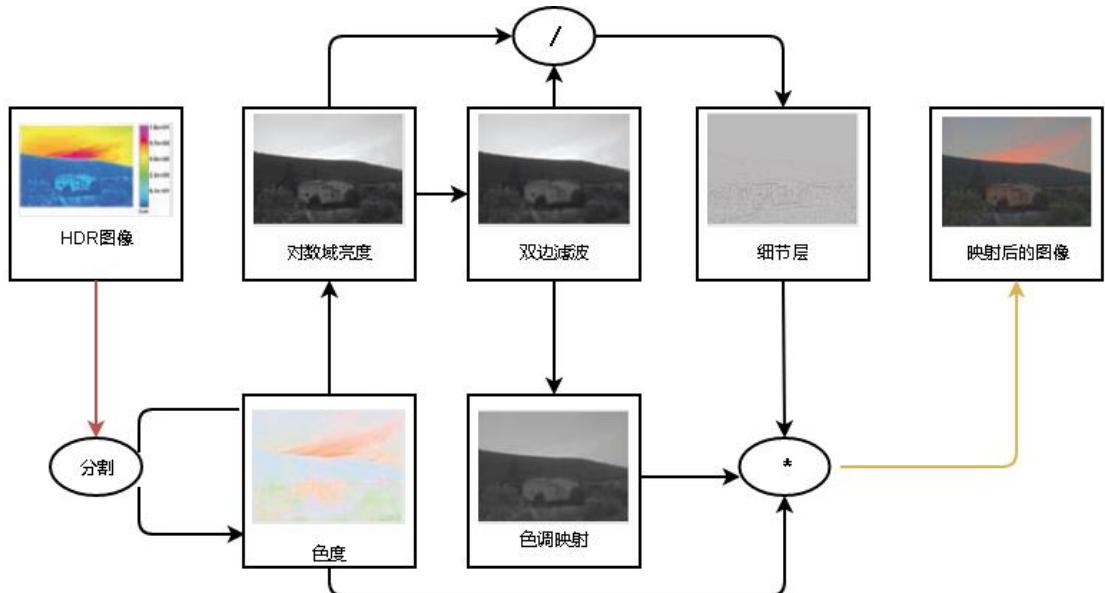


图 6-8 快速双边滤波算法的流程框图

方法的流程框图如图 6-8 所示，框架的第一步是将 HDR 图像分解为亮度分量和色度分量，此时用双边滤波器对对数域内的亮度进行滤波，并通过亮度与滤波后的商来计算细节层，滤波后的亮度随后使用一个全局算法进行色调映射。最后，将色调映射的基本层、细节层和色度分量重组，形成色调映射的最终结果。



图 6-9 使用双边滤波结构与否的效果对比。(a)亮度再现方法的结果；(b)亮度再现方法与双边滤波结构结合方法的效果。

Durand 和 Dorsey 提出了一个使用近似双边滤波和降采样的加速滤波器，然而这种技术已经被后来引入的加速双边滤波器淘汰了，这种加速方法可以应用于任何一种全局色调映射算法。图 9 显示了一种例子，与这种方法相关的一个问题是，光晕并没有被完全消除，因此对该种方法的改进，有使用三边滤波器等方法出现。

6.2.4 近期一些新算法

近年来，随着人们对于图像以及视觉的研究不断深入，色调映射算法的设计也逐渐变得更加复杂，适应性与通用性也逐渐加强。本节，我们将介绍两种近年来表现效果良好的三种色调映射算法：分割色调映射算法，Reinhard 色调映射算法以及自动淡化与加深算法。

1. 分割色调映射算法

最近，一种新的色调映射算法将分割算子的思想纳入其中。强边缘和大多数局部对比如觉位于大均匀区域的边界上。分割算子将图像分割成均匀的段，在每个段上应用一个全局算子，最后合并它们。这种方法的另一个优点是，最小化色域修改，因为在每个情况下，每个片段的线性算子有时都是足够的。

色调映射中的明度感知

Krawezyk 提出了一种基于明度感知锚定理论的色调映射算法，这种理论指出人类视觉系统可以将视野中的最高亮度值也叫锚点，视为白色，这种感知方法受到相对区域的影响。当最高亮度覆盖一小个区域时，它看起来是自发光的。为了将光度理论应用于复杂的图像，Gilchrist 等人提出在可以应用锚定理论的区域中分解图像，被称为框架。

算法的第一步是确定框架，然后在 \log_{10} 的对数域下计算图像直方图，使用 k-means 聚类算法确定直方图中的质心 G_i ，通过基于像素计数的加权平均合并接近质心。为了避免接缝或不连续性，使用软分割生成框架。每个框架定义一个概率函数，确定这个像素是否属于基于质心的框架，如下所示：

$$P_i(x) = \exp\left(-\frac{(C_i - \log_{10}(L_w(x)))^2}{2\sigma^2}\right) \quad (23)$$

其中 σ 等于两个框架之间的最大距离， $P_i(x)$ 使用双边滤波平滑处理，以删除局部的小变化，如图 10 所示。一个框架的局部锚点 ω_i 通过计算框架中亮度第 95% 的点来确定。最终色调映射的图片可以被计算为：

$$L_d(x) = \log_{10}(L_w(x)) - \sum_{i=1}^n \omega_i P_i(x) \quad (24)$$

色调映射的最终图片示例可以从图 10 看到，算法通过将其与摄影色调再现算法和快速双边滤波算法的结合来比较验证 Gelb 效应，这是一种与明度无关的错觉。试验结果表明，基于亮度的算法可以重现这种效果，算法可以快速直观地进行实施，但在将其应用于动态场景时需要特别小心，以避免出现重影伪像等情况。

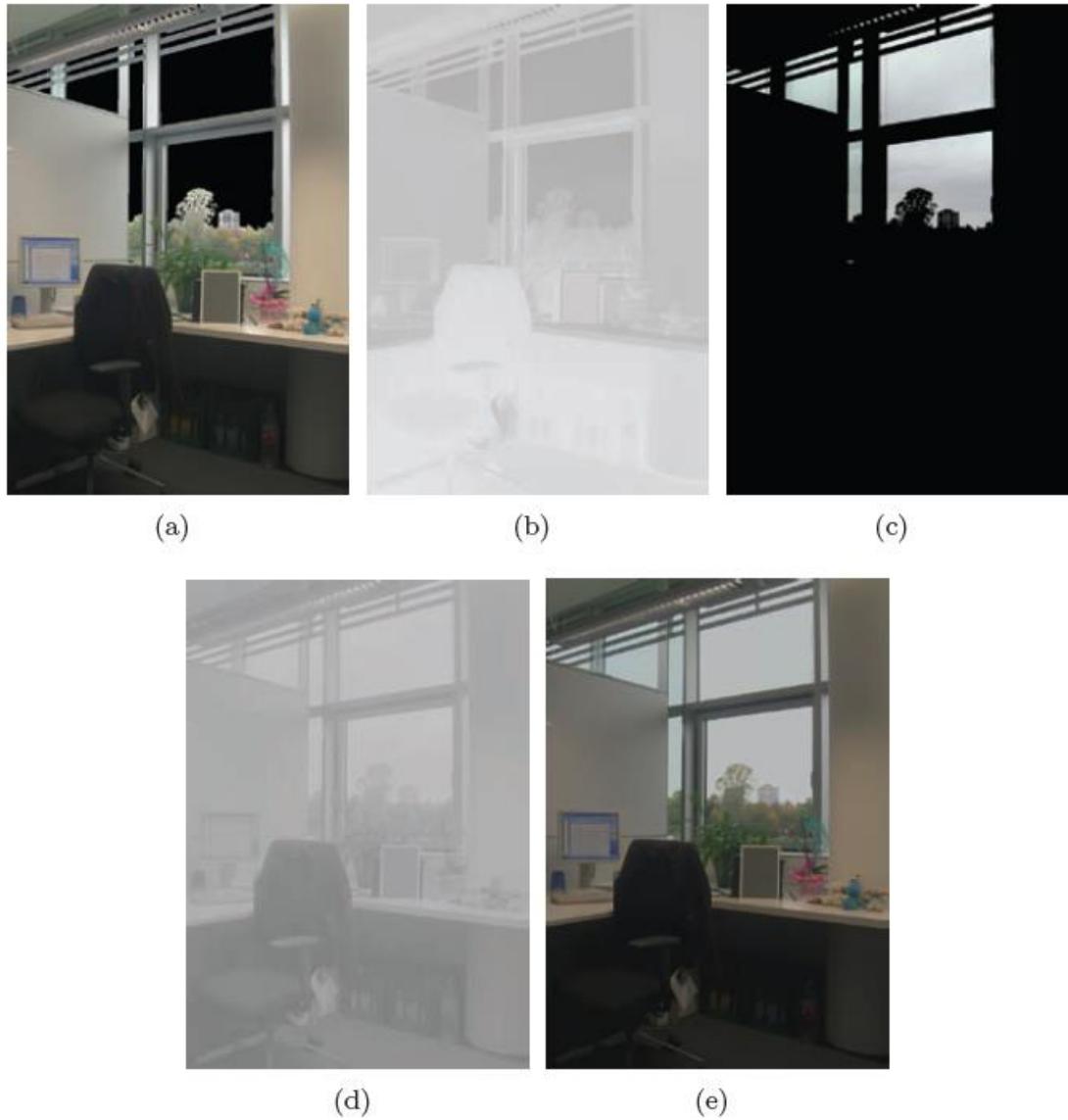


图 6-10 Krawczyk 色调映射算法结果图示例。其中(a)、(c)框架中使用锚点，(b)、(d)将(a)、(c)的概率分布平滑处理，(e)是通过合并框架得到的最终色调映射后的图像。

曝光融合

在色调映射之前，HDR 图像通常需要由一系列 SDR 图像组合而成，Mertens 提出来一种可以避免色调映射步骤的新方法，该算法的核心概念是合并每次曝光中曝光良好的像素点。

算法第一步是分析每个 SDR 图像以确定在合并时需要使用哪些像素，这是通过计算每个像素的三个度量来实现的：对比度 C 是指图像中渐变的绝对值，饱和度 S 定义为红色、绿色、蓝色三通道的标准偏差，亮度 L 的良好曝光度 E 确定像素是否以模糊方式良好曝光。

$$E(L) = \exp(-0.5(L - 0.5)^2\sigma^{-2}) \quad (25)$$

这三种度量结合，得到了一个权重和 $W_i(x)$ ，它决定这个像素点在当前曝光度下的重要性：

$$W_i(x) = C_i(X)^{\omega_c} \times S_i(x)^{\omega_s} \times E_i(x)^{\omega_e} \quad (26)$$

其中 i 指第 i 张图像， $\omega_c, \omega_s, \omega_e$ 是增加度量标准对其他标准影响的指数， N 个权重

图都被归一化处理过，使得它们的总和在每个像素位置处等于 1，以便获得一致的结果。

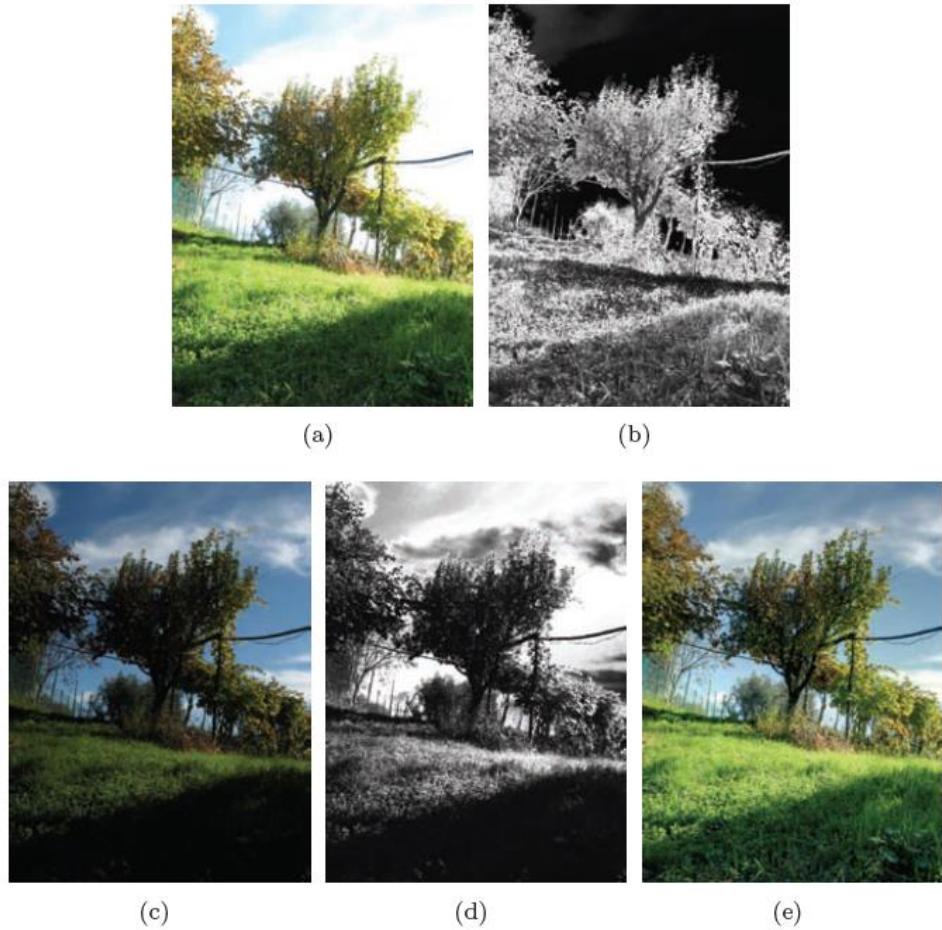


图 6-11 Mertens 的曝光融合算法示例。(a)是 HDR 图像的第一个曝光度；(b)是(a)的权重图；(c)是 HDR 图像的第二个曝光度；(d)是(c)的权重图；(e)是使用拉普拉斯金字塔融合后的色调映射图像。

分析之后，曝光结合在最终图像中，为了避免接缝和不连续性，图像混合使用拉普拉斯金字塔。虽然如用运算符 G 表示的那样在高斯金字塔中分解权重，但是如运算符 L 所示，曝光图像 I_d 被分解为拉普拉斯金字塔，因此混合计算可以定义为：

$$L^l(I_d)(x) = \sum_{i=1}^n L^l(I_i)(x)G^l(W_i)(x) \quad (27)$$

其中 l 表示拉普拉斯或高斯金字塔的第 l 层。最终 $L(I_d)$ 消掉了，得到了色调映射后的图片 I_d ，如图 11 所示。

该算法的主要优点是使用者不需要生成 HDR 图像，此外，它还可以最大限度地减少传统色调映射算法中可能出现先的颜色偏差，这是因为曝光良好的像素是在没有应用实际压缩功能的情况下拍摄的，只是线性关系。

2. Reinhard 色调映射算法

Reinhard 算法是一种基于区域系统的算法。在具体的操作过程中，算法会利用到一些摄影领域的技术，主要的应用就是利用区域系统的基本概念模型来对色调重构进行选择。操作过程中，我们首先会对图像进行缩放处理，就像照相机的曝光那样。然后，如果需要的话，我们会对图像进行自动淡化或加深处理，这样可以实现动态范围的压缩。

首先要做的是如何基于场景的关键值来设定输出图像的色调范围。与很多其他的色调映射方法类似，我们也认为对数平均亮度是场景关键值的有效近似。这个值 \bar{L}_w 可由下式得到：

$$\bar{L}_w = \frac{1}{N} \exp(\sum_{x,y} \log(\delta + L_w(x,y))) \quad (28)$$

$L_w(x,y)$ 是真实世界中 (x, y) 这个像素点的亮度值。 N 是图像中的像素总数。 δ 是用来避免奇异化的一个很小的量。如果场景的关键值是正常的，那么我们会把亮度映射到显示图形的中间灰度去，或者 0.18. 从 0 到 1 进行缩放。如下式： $L(x,y) = \frac{a}{\bar{L}_w} L_w(x,y)$

$L(x,y)$ 是缩放后的亮度， $a=0.18$ 。对于低关键值和高关键值的图像我们允许使用者将对数平均值映射到 a 的不同取值上。一般地， a 向上会取 0.36, 0.72；向下会取 0.09, 0.045。

上面的式子存在的一个问题，虽然大多数场景的主要部分都属于正常的亮度范围内，但是仍然有少数高亮的部分以及天空部分属于高亮度范围。在传统摄影技术中这种问题是通过同时压缩高低亮度部分来完成的。

但是，在现在的图像技术中，我们不再继续使用这种同时压缩高低亮度的“s”型的映射曲线，而是主要对高亮度部分进行压缩。一个符合这种要求的简单操作因子如下：

$$L_d(x,y) = \frac{L(x,y)}{1 + L(x,y)}$$

高亮度部分的缩放比约为 $1/L$ ，而低亮度部分则为 1。这种方法在两种缩放方式之间达成了很好的协调。而且这个式子可以保证将所有的亮度都映射到可以显示的亮度范围内。但是，并不是所有时候都需要这样的效果，上式可以经过适当的修改，然后可以让高亮度部分在可控范围内得到加强：

$$L_d(x,y) = \frac{L(x,y) \left(1 + \frac{L(x,y)}{L^2_{white}} \right)}{1 + L(x,y)} \quad (29)$$

L_{white} 是会被映射为白色的最小的亮度值。这个式子可以看做前面的操作因子与线性映射的混合。

对于大多数 HDR 图像来说，上面的方法提供的动态范围压缩方式已经足够保存低对比度区域的细节了。但是对于那些有着非常高的动态范围的图像来说，这种方法不足以保存它们存在于高动态范围的细节，因此对这些图像来说，淡化与加深算法是需要的。

6.3 视频色调映射算法

视频色调映射算法与图像色调映射算法存在很大的不同。视频色调映射算法实际上是对一段连续的帧进行色调映射，与图像色调映射算法相比，两者之间最大的不同是视频色调映射算法需要考虑到，并且准确地处理相邻帧之间的相关性，即考虑时域特性。如果这种帧间的相关性，或者换言之，这种时域的联系没有被考虑到，或者设计的算法不能准确处理这种时域联系，那么，在经过色调映射之后，视频就会出现一系列的问题，例如帧间闪烁现象。这种情况下，视频的人工痕迹就会非常明显，视频的质量与观看体验就会大幅下降。而这种时域联系并非是视频色调映射算法面临的唯一问题。由于视频处理要处理大量的帧，因此计算的效率也是十分重要的一个方面。另外，由于不同的 HDR 获取设备使用不同的方式获取 HDR 视频，因此这些视频本身也存在差异。所以，想要设计出适应多种情况的高质量的视频色调映射算法，需要综合考虑多方面的情况。

6.3.1 背景

尽管 HDR 视频的获取技术在 2010 年左右才相对成熟，视频色调映射算法早在这之前 15 年，就已经开始出现了。早期的视频色调映射算法都会涉及出固定的时域处理算法或者模型来处理视频的时域特征，但是这是不全面的。视频色调映射算法的最终目的，是建立一套自适应的模型，或者算法，来解决所有视频的时域问题。但是，由于早期缺乏高质量的 HDR 视频源，因此在当时无法精确地判断一个视频色调映射算法的性能。

随着 HDR 视频获取技术的发展以及新的 HDR 视频采集设备的使用，视频色调映射算法的性能可以得以检验。经过相关的调研 (EILERTSEN G.)，我们发现，大多数的视频色调映射算法都存在相同的几类问题：人工痕迹过重，对比度缺失以及细节缺失。因此，在 2010 年之后，视频色调映射算法的设计已经逐渐转变为针对具体存在的问题，做出相应的调整与完善。而视频色调映射算法的设计目标，也被明确为：设计出与视频内容无关的高质量色调映射算法。

时域特性

Boitard 曾经提出过视频色调映射过程中可能产生的时域问题的类型，其中主要包括：帧间闪烁，时域对比度缺失以及物体不连贯 (BOITARD R. C. R., 2014)。这些问题当中，我们主要关注全局与局部的闪烁现象以及鬼影现象，因为这三个问题，是视频色调映射过程中最常出现的时域问题。而其他的问题，例如时域对比度缺失以及物体不连贯，相较之下就容易处理得多，因为人们对于这些现象相对比较习惯。只要物体与对比度的变化是平滑的，那么，这些变化在人们看起来就是自然的，可以接受的，也就是说，这些问题并不会影响到人们的观看体验。而前面提到的三个主要的问题，全局与局部的闪烁现象以及鬼影现象，就会给人们带来很差的观看体验。因此，视频色调映射算法需要注重处理全局与局部的闪烁现象以及鬼影现象。

假设 $V: L \rightarrow T$ 代表一个由 HDR 视频到 LDR 视频的映射，如果这个映射 $V = V(L_s; \phi(V))$ ，其中 L_s 表示映射的输入内容是 HDR 视频， $\phi(V)$ 是通过某些方式得到的图像与视频信息。在这种情况下，这个映射 V 就是一个与图像内容相关的映射。由于 $\phi(V)$ 在部分帧与帧之间会剧烈的改变，这样就会影响到色调映射算法的效果。体现在输出的视频上，就是会出现帧间闪烁的情况。

对于视频时域特性的处理，第一种方式是进行像素级的改变，也就是说对于图像的操作实直接对像素值进行改变。如我们在图像色调映射算法中所说的那样，全局色调映射算法会对所有的像素点使用同样的色调曲线进行映射。因此，想要在全局处理中保持帧间变化的平滑性，只需要将连续的帧的信息通过一个时域低通滤波器，其中使用的核记做 G_σ 。

$$\tilde{\phi}(L_t) = \sum_{r=t-d}^{t+d} G_\sigma(\phi(L_r)) \quad (30)$$

从上面的处理中我们能够发现，这种时域低通滤波，是一种非因果的操作。如果希望保证算法的因果性，我们可以只对时间 t 之前的帧，进行低通滤波。

而对于局部色调映射算法来说，情况就更加复杂。如果局部色调映射算法使用了一些局部的图像信息 $\xi_s(L)$ ，并且没有适当的处理这些信息，那么，就会造成图像在某些区域剧烈地改变，这就会造成局部视频闪烁的发生。而如果向上面的式子所述的那样，直接对每个像素进行低通滤波，去除全局闪烁，生成的视频往往会产生鬼影现象。一种可以选择的改变方式是不直接对图像信息进行滤波，而是对色调映射后的亮度信息进行低

通滤波： $\tilde{T}_s = (G_\sigma * l)_s$ 。这种操作既可以保证时域相关性，又可以尽量避免噪声干扰，但是，这却会产生动作的模糊。

为了使像素级的时域滤波避免动作模糊的影响，可以对滤波器的核用很大的时间梯度进行限制，例如使用边缘截止滤波器。但是，这种操作会造成更大的时间梯度上出现少量闪烁。也就是说，像素级的时域滤波天然存在很多问题与缺陷。往往解决了一个问题之后，都会产生新的问题影响视频的质量。因此，对像素级的时域滤波的一种直接的拓展，就是使用动作补偿时域滤波器，这种滤波器可以有效的过滤出动作在时间线上的轨迹。如果动作的预估是正确的，那么这样的滤波器就能够在时域滤波器的平稳阶段以及边缘都能够准确过滤出动作，并且防止动作模糊现象的产生。但是，在一些情况下，动作预测可能会出现错误，这种情况下，这种滤波方式就会产生失真。例如在产生了镜头突然切换的两帧之间，就会出现动作预测的错误。动作补偿滤波器的例子包括动作向量的块状匹配 (LEE C.)，以及运用光流预测的滤波 (AYDIN T. O., 2014)。

通过使用上述的这些滤波器，理论上来说，现在使用的图像色调映射算法都可以通过增加这些滤波环节，拓展为视频色调映射算法。这也是大多数现在的视频色调映射算法使用的手段。

除了滤波手段之外，第二种利用时间相关性消除视频色调映射过程中出现的失真现象的思路是将时域处理作为一种色调映射的后置操作。也就是说，在这种想法下，使用何种色调映射算法已经无关紧要了，在色调映射之后，利用连续的帧之间的全部与局部信息等得到时域的相关性，然后处理得到的视频，就可以得到消除了各种失真现象的视频。其中一种控制闪烁现象的方式是计算色调映射后每一帧的均值，然后在时域上对每一帧之间的改变做出限制 (GUTHIER B., 2011)，或者将映射后的序列与原 HDR 序列进行比较 (BOITARD R. B. K., 2012)。这些方式通过对不同的帧进行亮度的不同比例的缩放，达到平衡与消除闪烁的目的。但是，这些方法的局限性在于这些亮度的改变只能作用于整幅图像，所以只能调节全局亮度。如果想要调节局部亮度与局部对比度，那么，就需要使用其他的方法进行调整。

大部分调节局部对比度的后置操作方法都是基于动作域的光流预测 (LANG M., 2012), (DONG X., 2015), (BONNEEL N., 2015)。这些后置操作算法的好处在于，我们不用像之前一样去考虑色调映射过程是如何完成的，不用去考虑对于不同的色调曲线，需要对滤波器进行哪些改变或者是否需要增加其他操作。特别是对于一些已经进行了色调映射，但是存在失真现象的视频，由于不知道色调曲线是什么，所以很难去针对性地设计滤波器，但是后置操作对于这种情况就可以很好地适应。而这类方法的缺陷在于，对于局部对比度的呈现以及视频的实际效果来看，要弱于滤波器的方法。

噪声

在前面的内容当中，我们把视频色调映射算法的设计问题基本介绍完了，我们可以通过两大类不同的方法来获取与利用视频的时域特性，然后将这种时域特性与图像色调映射算法结合起来，就可以实现视频的色调映射。但是，前面的算法与介绍都是在理想情况下进行的。在实际的生产与应用当中，我们还会遇到其他很多不同的问题，其中一个非常重要的问题就是如何处理噪声。

在 HDR 视频的获取过程中，HDR 摄影设备会产生很多噪声，特别是在阴暗的环境下，产生的噪声更加多。这是由于为了获得更高的动态范围，摄像的曝光时间需要被严格限制。因此，在阴暗的环境下，想要在有限的光照条件下拍摄与捕捉场景中的阴影与

黑暗的区域，就需要将摄影设备传感器的灵敏度调节得非常高。由于传感器灵敏度很高，这就会产生很强的传感器噪声。这些噪声在 HDR 视频当中基本是不可见的，所以并不会影响 HDR 视频的质量与观看体验。但是，经过了色调映射之后，这些噪声在 LDR 视频中会被大幅放大，从而十分影响视频的质量以及观看体验。噪声会在色调映射过程中被放大的原因在于，色调映射也是一种非线性的映射，它会将明亮区域的像素点的亮度降低，然后提升阴暗区域像素点的亮度。也就是说，在色调映射过程中，暗区的噪声被放大了。

为了缓解这些视频噪声，需要对视频进行相应的降噪处理。上一节提到的滤波器方法，可以在执行过程中增加部分操作与设计，这样就可以在色调映射的滤波过程中降低部分噪声 (BENNETT E. P., 2005), (H., 2006)。但是，这种方法的问题是会额外增加很多的运算，并且并不能够在不使得画面失真的情况下完全去除噪声。并且，如果算法中使用的色调曲线对暗区的增益非常高，那么就会使得没有被去除干净的噪声，以及降噪过程中产生的失真现象被进一步放大。

为了使得噪声不会在色调曲线的映射过程当中显现，一种合理的做法是在了解了噪声特征的情况下，可以对色调曲线的形状进行严格的控制 (EILERTSEN G. M. R., 2015), (LI J., 2016)。这种情况下，噪声虽然没有被消除，但是却隐藏在了暗区当中，不会被人眼发现，也就不会影响到人们的观看体验。这种方法可以与前面提到的降噪方法同时使用，这样就可以进一步的提高图像的质量与观看效果。

6.3.2 算法综合介绍

视频色调映射算法的不同总体上体现在四个方面，包括：色调曲线，处理方式，算法目标以及时域滤波。同时，由于视频色调映射所需的运算量非常庞大，因此，运算速度也是一个重要的评价指标。

1. 色调曲线：这是色调映射算法的核心。大部分色调映射算法，以及后面我们将要介绍的算法，主要使用以下几种函数作为色调曲线：线性函数，指数/对数函数，sigmoid/s-shaped 函数以及基于灰度图的色调曲线。同时，还要少量色调映射算法在梯度域进行映射。
2. 处理方式：处理方式主要分为两种。一种是全局映射，即对所有的像素使用同样的色调曲线进行映射。第二种是局部映射，在不同的空间位置，使用不同的色调曲线进行映射。
3. 算法目标：色调映射的主要目的分为三类：visual system simulators (VSS), best subjective quality (BSQ) operators and scene reproduction (SRP) operators。(EILERTSEN G. W. R., 2013), (DUFaux F., 2016)
4. 时域滤波：时域滤波通常分为两类。第一类是在时域上对色调映射参数进行整体的滤波。第二类是在局部上对每个像素点进行时域滤波。
5. 速度：视频色调映射算法的速度大致分为三个档次：离线，交互式，实时。划分的依据是算法的处理能力以及运算复杂度。实时的算法，可以提供超过 20 帧/秒的运算速度。交互式算法相对弱一些，每秒可以处理几帧。离线法则速度更慢。由于算法的执行速度与图像的分辨率以及计算机的硬件水平也息息相关，因此这种划分方式只是一个大概的划分，仅供参考。

6. 3. 3 代表性算法简介

上一节中, 我们对于视频色调映射算法的总体思路以及主要组成部分做了分别的介绍。本节, 我们就要将注意力转移到具体的视频色调映射算法上, 我们将会介绍一些具体的视频色调映射算法的思路以及实现方法以及主要特点。

1. 视觉适应算法 (Visual adaptation TMO)

(FERWERDA J. A., 1996)这是第一个使用时域适应机制的视频色调映射算法。视觉适应算法利用了一系列视觉领域的心理与生理学研究结论建立了一个视觉相应模型。算法的核心方法是: threshold-versus-intensity 方程, 这个方程之前大量的实验结果与数据, 可以计算出人类视觉范围内的颜色变化, 视觉敏感度(能够发现空间上的细节)变化以及时域适应性(时域上明暗变化)变化的阈值。

视觉适应算法的核心思想是, 通过在背景亮度下预测出人眼能够感知的颜色, 场景清晰度变化阈值, 然后只对人眼能够感知到的变化部分, 进行色调映射, 从而在一定程度上减小了算法的运算量。同时, 算法会对视觉敏感度以及时域适应性变化进行额外的处理。

2. 时域适应算法 (Time adaptation TMO)

(PATTANAIK S. N., 2000)这种算法充分利用了人类视觉模型的时间适应机制。通过在适应模型后增加一个表现模型, 就可以模拟出一幅图像的响应。随后再使用逆表现模型与逆适应模型, 就可以将图像显示在 LDR 显示设备上。

在适应模型中, 图像中视锥细胞与视杆细胞的静态响应通过 sigmoid 函数被分别计算出, 而时域适应性则通过指数平滑滤波器进行建模。并且算法加入了染色与漂白因子, 用来对时域上的特征进行着色与漂白。

3. 时域曝光算法 (Temporal exposure TMO)

(KANG S. B., 2003), 这种算法首先提出了一种新的获取 HDR 视频的方式: 在帧与帧之间, 交替的改变曝光时间, 从而增大视频的亮度范围。之后, 为了创建出 HDR 视频, 每一帧利用光流信息进行内插, 从而为每一帧提供了两种不同的曝光度, 实现了动态范围的提升。并且, 针对这类 HDR 图像, 文章中也提供了一种算法来对其进行色调映射, 从而显示在 LDR 显示设备上。

这种视频色调映射算法基于一种经典的图像色调映射算法: Photographic TMO (REINHARD E., 2002)。通过使用这种算法, 再进行时域滤波, 从而实现视频色调映射。Photographic TMO 利用了 sigmoid 函数并设计了如下形式的色调曲线:

$$T = \frac{L_s}{1+L_s} \left(1 + \frac{L_s}{L_{white}^2} \right) \quad (31)$$

其中 $L_s = a L/k$, k 表示几何平均数。由于帧与帧之间的平均数会剧烈的改变, 从而造成闪烁现象。因此, 为了解决闪烁现象, k 会通过固定数量的帧得到, 从而缓解闪

烁现象，并且可以在时域上实现有效的统一滤波。

$$k = \exp\left(\frac{1}{N} \sum_{s,t} \log(\varepsilon + L)\right) \quad (32)$$

这里的 s 表示每幅图像中的每一个像素点， t 表示一组连续的帧。 ε 是一个很小的常数，用来避免奇异值。同时，由于 **photographic TMO** 的计算成本很大，因此，**Goodnight** 为这种算法设计了相关的 GPU 加速算法 (**GOODNIGHT N.**, 2003)，使得这种色调映射方法的计算时间大大缩短。

4. 局部适应算法 (Local adaptation TMO)

(LEDDA P., 2004) 与时域视觉模型类似，局部适应算法同样利用了人眼的时域特性，并且使用了同样的方法，即利用 **sigmoid** 函数将视觉锥细胞与视觉杆细胞的响应分开计算。但是，两者的区别是，为了获得局部响应，也就是局部细节，本算法中对每一个像素单独进行这种操作，而时域视觉模型针对每一帧整体进行操作。为了对局部适应性的时域特征进行建模，算法在时域上利用指数滤波器进行建模。并且对于视锥细胞与视杆细胞，明亮区域与阴暗区域，使用了不同的滤波器。

5. 感知效应算法 (Perceptual effects TMO)

(KRAWCZYK G., 2005) 这种算法同样是建立在 **photographic TMO** 的基础上。这种算法的主要目的之一，是实现实时的局部色调映射。这种算法利用一些感知效应增强了 **photographic TMO** 的效果。由于 **photographic TMO** 需要通过建立高斯金字塔来获取色调映射的局部适应等级，基于这种框架的一些效应处理，只需要增加很少的计算量，就可以实现算法效果的增强。基于这种框架的优势，视觉敏锐度以及光幕照明等算法都被添加进了框架当中。

时域适应性则是通过指数衰减方程 (**exponential decay function**) 在时域的适应性等级上进行滤波。具体做法类似于 **Interactive walk-through TMO** (DURAND F., 2000)。最后，为了提升算法的计算速率以实现实时映射，最终的算法实现使用了 GPU 加速。

6. 梯度域算法 (Gradient domain TMO)

(WANG H., 2005) 该算法首先也是提供了一种全新的 HDR 视频获取方式。通过一种称之为：分离光圈相机的拍摄设备，这种设备包括 3 个 CCD 传感器相机，可以同时拍摄三种不同亮度范围下同一场景的图像，然后场景的实际亮度可以利用拍摄出的图像通过传统方法重现。而算法的实际目的是通过这种方式获得的 HDR 视频，来研究 HDR 视频在空间域与时域的特性。

对于色调映射算法来说，梯度域算法之前就有所研究。这种算法在不同的位置 s 处，以不同的比例 k ，利用比例因子 φ_s^k 减弱图像的梯度 ∇H_s^k 。

$$\varphi_s^k = \frac{\alpha}{\|\nabla H_s^k\|} \left(\frac{\|\nabla H_s^k\|}{\alpha} \right)^\beta \quad (33)$$

改变后的梯度域 $G_s = \Phi_s \nabla H_s^k$ (Φ 是将所有 k 累加起来的比例因子) 之后被用于利用泊松方程来计算色调映射后的 T 。

$$\nabla^2 V = \operatorname{div} G \quad (34)$$

与原方法相比,这种算法的差别在于,这里的梯度域是三维的,其中两维表示空间,一维表示时间。解决这种三维泊松问题会使得结果具有时域一致性。但是,对于色调映射问题来说,处理过程中只对空间特性进行梯度减弱,因为如果进行了时域的减弱,就会产生运动模糊。

7. 区域匹配算法 (Block match TMO)

(LEE C., Gradient domain tone mapping of high dynamic range videos, 2007)这种算法的思想与梯度域算法类似,即色调映射在梯度域执行,并且梯度是三维的,两维表示空间,一维表示时间,梯度在计算过程中通过压缩得到减弱。与梯度域算法不同的是,梯度域算法只在空间域的两维上进行梯度减弱,区域匹配算法在时间域的维度上也进行了梯度减弱。为了在实现时间域梯度减弱的过程中不产生运动模糊,算法中又利用区域匹配以实现运动补偿,从而消减运动模糊的影响。

记像素点 s 处,从第 $t-1$ 到第 t 帧的动作的估计向量为 $v_{s,t}$, 对应的代价函数为:

$$C = \sum_s (L_{s,t} - L_{s-v_{s,t},t-1})^2 \quad (35)$$

当上一帧的像素点的值与这一像素点在下一帧预计应该在的位置的像素点值存在差异时,代价函数的值就会增大,也就意味着这个运动估计是不可靠的。代价 C 会被用于泊松方程以平滑动作路径上的点的梯度。这样操作之后,就可以增强相邻帧之间的时域相关性,并减弱噪声。也就是说,这种算法不仅不会像之前的梯度域算法一样产生运动模糊,并且还可以减弱闪烁与时域不相关问题。

8. 视网膜模型算法 (Retina model TMO)

(BENOIT A., 2009)这种算法的思想来源与模拟人类视觉系统的处理方式,也就是视网膜的处理方式。在人眼对图像的处理过程中,眼部细胞首先接受光感受器的响应,然后将刺激传递到神经节细胞,最后由视网膜进行空间域与时域的滤波,进而实现视觉的生成。这种算法模拟了视网膜的底层处理,包括局部适应性。

首先,利用 sigmoid 函数可以获得图像的局部响应,通过空间低通滤波器可以获得图像的局部适应等级。局部响应通过时空滤波器,可以模拟出 Outer Plexiform Layer (OPL) (HÉRAULT J., 2007)的处理方式。OPL 滤波器可以被认为是两个低通时空滤波器的差,并且将频谱进行了白化滤波。也就是说,它将高频部分中的低频信号与低频部分中的高频信号进行了过滤。这种滤波器通过过滤高频部分,从而消减了噪声;并且通过过滤时域的局部适应等级,从而提供了更好的时域稳定性。在通过了时空滤波器之后,响应会通过神经节细胞响应模型,这个模型与算法的第一步很类似,都是通过 sigmoid 函数,区别是使用了不同的参数。

整个算法的最后一步是颜色的重建。在算法的最一开始,颜色信息会在初始化阶段,通过多路处理被提取出来。而在最后一步,颜色信息会被还原到图像当中。整个操作的目的是实现颜色的一致性,防止在算法中间的滤波过程中,对颜色信息产生不可逆的修改。这种思想,来源于视网膜的处理方式 (MEYLAN L., 2007)。

9. 闪烁消除模型 (Flicker reduction TMO)

(GUTHIER B., Flicker reduction in tone mapped high dynamic range video, 2011)这种算法实际上并非完整的视频色调映射算法。这种算法的应用对象是已经经过色调映射的视频，模型通过检测与处理，可以将视频中存在的闪烁现象减弱或去除。

在色调映射后的视频序列中，首先检测是否存在闪烁现象。算法的检测方法是通过检测连续帧之间的 k 值之差是否超过了阈值 H ，来进行判断。其中， K 值就是时域曝光模型中的 k ，而阈值则通过 Steven's Power Law (S., 1962) 计算， $H = \Delta L = aL^{1/3}$ ， a 通过实验决定。

$$k = \exp\left(\frac{1}{N} \sum_{s,t} \log(\varepsilon + L)\right) \quad (36)$$

对于大于阈值的帧，亮度迭代地下降，知道帧的 k 值范围落在合理范围之内 $k_t \in \{k_{t-1} - H, k_{t-1} + H\}$ 。对于只存在全局亮度问题，也就是全局闪烁问题的色调映射后的视频序列，这种算法都是有效的。

10. 时域相关模型 (Temporal coherence TMO)

(BOITARD R. B. K., 2012)除了时域闪烁之外，视频色调映射中还存在两个潜在的需要注意的问题，一个是物体不连续问题，用于描述在色调映射物体的亮度变化问题以及由于不同的帧的统计信息不同，而导致的与原视频序列中物体的变化不一致的问题。第二个问题是亮度不连续，用于描述视频整体的亮度在时域上的改变，例如经过了色调映射后，原视频中亮度最高的一帧可能就不再是最亮的一帧。

为了使色调映射算法能够克服这些问题，Boitard 等人提出了时域相关模型。这种模型与前一个提到的闪烁消除模型一样，也不是完整的色调映射算法，而是用于从色调映射后的视频序列中消除时域问题的一类附加修正算法。对于每一帧 t ，首先计算它在色调映射之前的 k 值 k_t^L 以及色调映射之后的 k 值 k_t^T 。计算完成后，利用全局系数 τ 将色调映射后的视频的每一帧的 k 值都固定为原 HDR 视频序列中 k 值的最大值， $\tilde{T} = \tau T$ 。

$$k = \exp\left(\frac{1}{N} \sum_{s,t} \log(\varepsilon + L)\right) \quad (37)$$

$$\tau = \frac{k_t^L k_{max}^T}{k_{max}^L k_t^T} \quad (38)$$

接着再用一个调制系数，使得 τ 能够在很小的范围内浮动

$$\tau = \tau_{min} + (1 - \tau_{min})\tau \quad (39)$$

由于这种算法也是一种附加操作，因此，这种算法也使用与所有只存在全局亮度问题的色调映射算法。这种算法利用了整个视频序列的信息来解决亮度问题，并且将每一帧都与原 HDR 视频进行比较。这样的操作会造成视频的整体亮度出现下降，因为为了保持视频亮度的相对不变，视频中的大多数帧的亮度在操作过程中都会下调。

11. 混合模型 (Hybrid TMO)

(SHAHID H., 2015)混合模型顾名思义，利用了两种不同的色调映射算法对同一帧进

行处理。首先，每一帧会通过 directional filter bank edge-detection algorithm (ANAND S., 2012) 进行处理，将图像分为包含剧烈改变的区域以及不包含剧烈改变的区域。包含剧烈改变的区域随后使用 iCAM06 TMO (KUANG J., 2007) 算法进行色调映射，这种算法通过局部操作保留了图像的细节。图像的剩余区域通过直方图校正算法 (WARD LARSON G., 1997) 进行处理。最后，为了保持视频的时域性质，映射后的帧的亮度会进行比例缩放，进而使得连续的帧之间的强度差值不会超过预先设定的阈值，具体操作类似于闪烁消除模型。

12. 噪声感知模型 (Noise-aware TMO)

(EILERTSEN G. M. R., Real-time noise aware tone mapping, 2015) 这种算法用于处理视频获取过程中出现的噪声。算法的目的是利用一种有效的时域稳定的视频色调映射算法，在加强视频局部对比度的同时，不提升视频噪声，使视频噪声保持不可见。

算法中，利用式子 (4) 实现局部色调映射，并且算法中还使用了一种特别设计的滤波器。这种滤波器基于标准的用于噪声消除的边缘保留滤波器，它可以在一些情况下找出图像中的失真区域。色调映射中的细节抽取则使用了不同的思路。为了能够抽取出图像的细节并且不引起失真，算法中使用了一种简化版的双边滤波以及各向异性扩散滤波。这种方法可以解决与 0 阶各向异性相关的重构问题。

对于色调映射来说，色调曲线用于将原始图像序列的全局对比度尽可能地保存，因此在设计色调曲线的时候，实际上是在解决一个最优化问题，优化的目标是减少由于亮度映射带来的对比度的畸变。对于输入的对比度 G 以及映射后的对比度 \tilde{G} ，畸变的平方的期望 $(G - \tilde{G})^2$ 可以利用对比度相对于对数域的亮度的分布 $p(G|L)$ 进行表示：

$$E[\|G - \tilde{G}\|_2^2] = \int p(L) \int (G - \tilde{G})^2 p(G|L) dG dL \quad (40)$$

上式提供了一个最优化问题，可以通过将色调曲线参数化表示以及使用图像直方图来表示 $P(L)$ 来很快地解决。整体的概念与显示适应模型 (MANTIUK R., 2008) 中的对比度优化过程十分类似，但是本算法中使用的方式减少了很多计算复杂度。为了保证时域的平滑改变，色调曲线的节点会在时域上利用低通 IIR 或者边缘截止滤波器进行滤波。

由于色调曲线的映射会将原视频中不可见的噪声变得可见，为了防止这一现象的发生，算法中使用了噪声感知技术。这一技术的想法是在色调映射的优化模型中，也就是式 (22) 的优化模型中，引入表示图像噪声的项，在最优化的过程中，这一项就会被不断削弱。同时，算法中还将提取出的细节层进行了缩放，以减弱可见噪声。基于这两种方法，算法能够将可见的噪声减弱至人眼无法观察到的程度，进而提升了图像的质量。

第七章-逆色调映射

HDR 技术近年来发展迅猛，在未来将会成为图像与视频领域的主流。当前 HDR 内容非常短缺，限制了 HDR 视听节目的广泛应用。逆色调映射 (Inverse Tone Mapping) 应运而生，它是一种用来将 SDR 源信号转换为 HDR 源信号的技术，可以应用于生产端或终端设备，在一定程度上实现对现有 SDR 节目的 HDR “还原” 及向上兼容。本系列中，我们将会详细分类介绍逆色调映射算法。分为两个部分：（一）逆色调映射概述及一些经典算法，包括全局算法，分类算法以及拓展映射算法；（二）介绍最近的研究趋势，特别是基于机器学习的逆色调映射算法。

7.1 逆色调映射概述

HDR 技术说是继 4K 之后视频领域的又一次重大技术突破，不仅提高图像的动态范围，同时也提供了更宽的色域，并提升了画面在高光和暗部的细节信息，带给观看者更加真实的视觉体验。但作为 HDR 产业链的核心，内容短缺的问题仍然是 HDR 普及和成熟的重要限制因素。众多电视厂商将 HDR 技术作为其产品规划中的重要技术，主要是在硬件方面进行提升。而由于 HDR 内容制作对于设备要求较高，对采集、调色、显示等多方面都需要较高的质量要求，内容制作的复杂度和周期长等因素，导致了 HDR 内容的匮乏。让内容跟上硬件发展的速度，让内容不再限制整个 HDR 产业链的发展依然是当前 HDR 发展完善的重要任务。

如果能充分利用已有的大量 SDR（部分文献中使用 LDR，两者意义一致，本文中均使用 SDR）视频源，将其转换为 HDR 内容，那么不仅可以改善 HDR 片源不足的问题，同时也可以有效利用已有的资源进行内容重置。为此，逆色调映射方法 (inverse Tone Mapping Operator, iTMO 或 reverse Tone Mapping Operator, rTMO) 应运而生。通常，也将该 SDR 转化 HDR 过程的算法称为扩展算子 (Expand Operator)，它可以被定义为一个 SDR 图像 I 的函数 g :

$$g(I) = D_i^{w \times h \times c} \longrightarrow D_o^{w \times h \times c} \quad (1)$$

其中 $D_i \subset D_o$ ， w 、 h 、 c 分别是图像的高度、宽度和通道数， D_i 是 SDR 图像，在 8 比特量化的情况下， $D_i = [0, 255]$ ； D_o 是 HDR 图像，在使用单精度浮点数时， $D_o = [0, 3.4 \times 10^{38}] \subset \mathbb{R}$ 。

由于在图像的过曝光和欠曝光部分存在大量的信息缺失，因此从 SDR 的单曝光图片，通过逆色调映射算法进行 HDR 内容的重构是一种病态问题 (ill-posed)，因为这需要算法预测或估计出过曝光或欠曝光部分缺失的信息是什么。因此，过曝与欠曝区域信息的重构，是逆色调映射技术中，需要着重设计的部分。一般情况下，逆色调映射算法将 SDR 内容变换为 HDR 内容需要经过以下几个步骤：

1. 线性化：图片通常已经经过 gamma 校正，也就是说其电平信号与记录的真实场景中的辐射率 (radiance) 并不是线性关系。为了后续处理方便，会先将图像进行线性化处理。
2. 动态范围扩展：该步骤是逆色调映射算法的核心部分，动态范围扩展方式决定着映射后图像的最终质量。对高光部分进行扩展，对低光部分进行压缩，可以达到目的。

3. 过曝/欠曝区域重建：存在过曝或者欠曝现象的 SDR 图像相应区域会损失很多细节，因此需要通过一定的方式恢复或者重建该区域以获得更多的细节。

4. 去噪：逆色调映射算法处理和存储使用的是 16 比特的浮点数，因此需要通过去噪消除量化或压缩而引入的噪声。

5. 颜色校正：进行逆色调映射处理后图像的饱和度会降低，因此需要进行颜色校正让色彩表现（color appearance）和 SDR 源相一致。

需要说明，第一步线性化图像非常重要，在线性空间中操作图像的主要原因是对亮度和色域拓展有更大程度上的控制和可预测性。在一个未知的空间中，很难预测扩展如何表现，此外为了恢复场景属性，例如平均值、几何平均值、标准偏差等，需要精确估计场景的辐射度量。

当可以得到拍摄设备信息时，线性化可以通过设备的 CRF 来计算，但当图像储存为 RAW 格式时不需要这个操作，因为 RAW 格式本身就是存储为线性格式。然而，一般情况下我们并不能得到拍摄设备的 CRF，并且视频和图像都是通过 8bit 非线性储存，这种情况下我们需要估计 CRF 来实现线性化操作。一般实现这种线性化的方法，最常用的是 Gamma 校正，除此之外还可以使用辐射校准等。

7.1.1 逆色调映射中的颜色转换

SDR 图像与 HDR 图像的区别不仅体现在亮度范围上，也体现在色域差别上。常见的 SDR 图像/视频的颜色空间符合 BT. 709 标准中规定的色域，而 HDR 图像/视频的颜色空间则符合 BT. 2020 标准中规定的色域。这种色域的拓展，使得 HDR 图像/视频在可见光中的颜色覆盖率提高了近 40%。因此，现今的逆色调映射算法不仅仅要对图像的亮度信息与对比度进行增强，还需要进行色域映射。传统的逆色调映射算法很少涉及颜色的处理，忽视这点可能引起增强后的图像/视频出现颜色偏移等现象。

逆色调映射算法中色域转换的常见方法是色彩校正（Color Correction），这种方法可以使映射后得到的 HDR 图像获得视觉上与原生 HDR 接近的效果；另一种方法是使用色彩或图像外观模型进行色域转换，一般这种模型都是基于人类视觉系统的特性来实现的，但是这种方法能否准确地预测复杂场景或任意观看条件下的转换参数是值得商榷的。

随着色域的标准化，颜色空间由 BT. 709 向 BT. 2020 的转换过程也被纳入了 ITU 的标准之中，ITU 发布的色域转换建议书 BT. 2087 [25] 中提出了由 BT. 709 到 BT. 2020 的色域转换流程，BT. 2407[26] 建议书中提出了一种从 BT. 2020 到 BT. 709 色域转换的方法。BT. 2087 中提到的转换方法主要应用于解决 BT. 709/HD 电视内容在 BT. 2020/UHDTV 上播放的兼容性问题，方法是基于光电转换函数和电光转换函数及其反函数，具体转化方法可以参照 ITU 发布的文档。在逆色调映射过程中的颜色转换，可参考 BT. 2087 建议。此外，更先进的色域映射算法考虑如何高效利用图像的空间属性。在空间色域映射中，不再存在从输入色域到输出色域的一对一颜色转换，而是考虑图像的局部背景，使得算法不仅可以准确地再现颜色，而且还能同时再现纹理和局部对比度，其主要方法有两类，一种是将传统映射过程中丢失的信息迭代添加回到图像中；另一类方法，最小化刻画色域映射图像和原始图像之间感知误差的损失函数，一般通过机器学习的方法实现。

逆色调映射算法，可以分为三大类算法：全局算法，分类算法以及拓展映射算法。在本文的后续部分当中，将介绍这三大类算法中的代表性方法，以及最新的基于机器学习的逆色调映射方法。

7.2 逆色调映射算法

7.2.1 全局模型

全局模型顾名思义，就是对 SDR 内容中的所有像素点都使用相同的扩展方程，该扩展方程对全局有着相同的操作。

1. 幂函数模型

这种方法是最早提出的逆色调映射算法之一，最初由 Landis [1] 提出主要从图像重新获取基于幂函数的数字三维模型，主要考虑到幂函数良好的范围拓展能力，用于将小范围的亮度拓展到高动态范围，亮度扩展可以定义为：

$$L_w(x) = \begin{cases} (1-k)L_d(x) + kL_{w,\max}L_d(x) & L_d(x) \geq R \\ L_d(x) & \text{otherwise} \end{cases} \quad (2)$$
$$k = \left(\frac{L_d(x) - R}{1 - R} \right)^\alpha$$

其中 R 表示扩展的阈值，原文章中使用的是 0.5； $L_{w,\max}$ 表示用户需要在输出图片中展示的最大亮度； α 是控制色调曲线伸展的衰减指数。

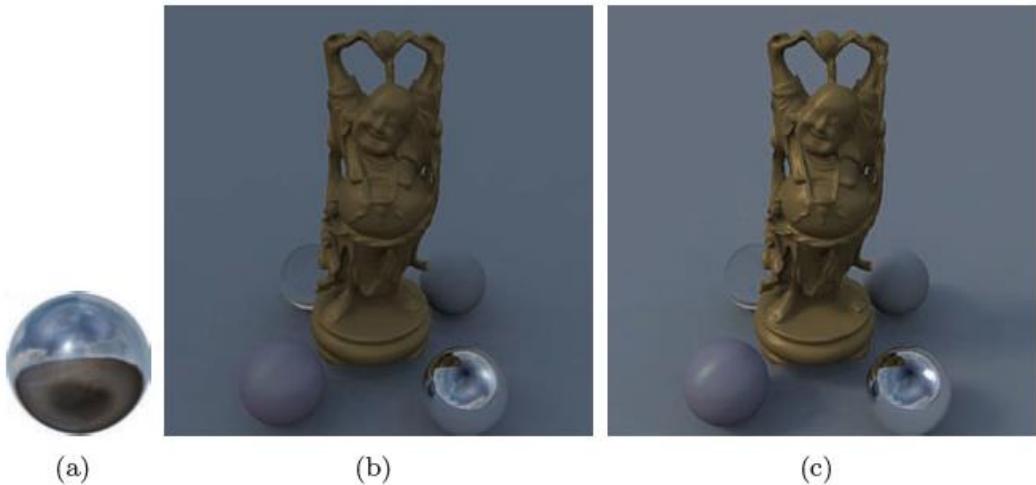


图 7-1 Landis 方法示例：(a) 初始 SDR 环境贴图；(b) 仅使用(a)的重建图像；(c) 使用扩展过的(a)的重建图像。

虽然这种方法为 IBL (image based lighting) 生成了合适的 HDR 环境贴图，如图 1 所示，但它可能无法生成可在 HDR 显示器上可视化的高质量图像和视频。这是因为这种方法不处理伪像，例如对压缩或量化产生伪像不作处理，全局的亮度拓展对于整体亮度效果有着较好的改善，但是无法进行细节增强。

2. 对 HDR 监视器的线性缩放

为了研究 HDR 显示器支持 SDR 内容的效果，Akyuz [2] 等人进行了一系列有关显示效果的主观测试，测试实验通过使用杜比 DR-37P HDR 显示器评估色调映射图像、单幅曝光图像和 HDR 图像。实验设计 20-40 岁之间的 22 名测试者，在所有实验中均使用了从室内到室外、从昏暗到非常明亮的不同光线条件下的 10 张 HDR 图像，HDR 图像亮度约为 5 个数量级，通过线性缩放处理得到不同动态范围的图片，以便直接映射到杜比 DR-37P HDR 显示器上。

第一个实验中，使用各种色调映射算法由 HDR 得到 SDR，再比较两个图像，另外也比较了一种自动曝光和测试者在试点研究中选择的曝光，其中自动曝光是为了最大限度上减少曝光过度和曝光不足的像素数量。图像显示在 DR-37P 显示器上，使用经过校准的 HDR 图像和 SDR 图像匹配 Dell UltraSharp 2007FP 20.1 英寸 LCD 显示器，测试者的任务主要是将图像按照主观效果排序。对于每个原始测试图像，测试者必须观看测试图像两秒钟，这是从不同类型的图像中随机选择的。实验结果表明，测试者更喜欢 HDR 图像，作者没有发现色调映射和单次曝光图像在测试者之间的偏好方面存在很大的差异。

第二个实验中，作者比较了从单次曝光图片扩展得到的 HDR 图像和单次曝光的图像，为了扩大单次曝光图像，扩展方法可以表示为：

$$L_w(x) = k \left(\frac{L_d(x) - L_{d,\min}}{L_{d,\max} - L_{d,\min}} \right)^\gamma \quad (3)$$

其中 k 表示 HDR 显示器的最大亮度强度， γ 表示非线性缩放因子。对于这个实验，分别生成了 γ 为 1、2.2 和 0.45 的图像，此外设置和排名方式与第一个实验相同。结果表明，选择较亮的曝光扩展图像比 HDR 图像更受欢迎，当它们具有相同的平均亮度时反之亦然。Akyuz 等人表示平均亮度比对比度更受欢迎。最后，另一个重要结论是 $\gamma = 1$ 时产生的线性缩放图像最受欢迎，这表明线性缩放足以满足 HDR 图像和视频的体验。

作者仅使用高分辨率 HDR 图像，并且没有产生压缩伪像。虽然这在理想条件下工作良好，但在实际情况下，例如使用压缩的电视节目或 DVD 中，情况可能并非总是如此。在这些情况下，需要进行更准确的扩展以避免压缩伪像和轮廓的放大。

3. 过曝 SDR 图像的 Gamma 扩展

Masia [7] 等人进行了两项心理物理学研究，以分析扩展函数在各种曝光度水平下的表现。作者随后使用这些实验的结果来开发过曝内容的扩展技术。在这些实验中，比较了三种逆色调映射算法，包括 Rempe1[8]、线性缩放[2] 和 Banterle[13] 提出的方法（这些算法后文中将介绍），并且结果显示随着内容曝光度的增加，结果下降的趋势。作者还观察到，逆色调映射函数制造出一些在图像 SDR 格式中可见的伪像，原因在于算法产生不正确的光照强度值，并造成空间内容失真。了解逆色调映射函数如何影响最终输出图像的观看效果可能有助于开发更好的算法。

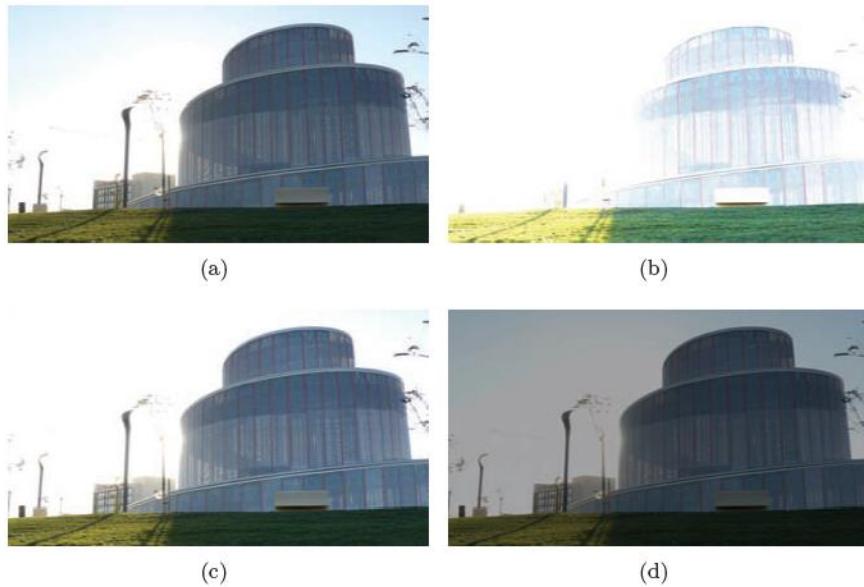


图 7-2 Masia 方法中过曝图片案例：(a) 原始 SDR 图像；(b) (c) (d) 扩展后不同光圈下的图片。

如图 2 所示，作者指出：当图片的曝光度提高时，细节会丢失，这是由于像素点的饱和度增加使得对应的颜色褪色至白色。基于这种发现，他们提出通过将剩余细节突出更加明显来映射 SDR 图像，这不同于常见的逆色调映射算法只是增强饱和的区域。

一种实现这种方法的直观的方式，就是使用 Gamma 扩展，Masia 提出一种自适应的方式，通过输入 SDR 图像的动态内容来确定 γ 的值，与 Akyuz 的方法类似， k 值可以通过以下方式计算：

$$k = \frac{\log L_{d,H} - \log L_{d,min}}{\log L_{d,max} - \log L_{d,min}} \quad (4)$$

其中 $L_{d,H}, L_{d,min}, L_{d,max}$ 分别是对数平均数、最小亮度值和最大亮度值， k 值是一个统计数据，有助于澄清输入图像是主观黑暗还是明亮。为了自动预测 Gamma 值，开展了一项试点研究，要求用户手动调整一组图像中的 γ 值，数据是经验拟合关系的线性回归：

$$\gamma(k) = ak + b \quad (5)$$

由拟合得到， $a = 10.44, b = -6.282$ 。这种逆色调映射算法的主要缺点之一是它可能无法充分利用动态范围，而且 a 和 b 的值只能在该组测试图像上正常工作，在一些不属于原始图像集的图像中， γ 可能为负值。

7.2.2 分类模型

1. HDR 显示器的高光重现

Meylan [9, 10] 提出一种专门用于处理包含镜面高光的图像的逆色调映射算法，是一种

典型的分类模型算法。其主要思想是检测图像中的漫反射和镜面部分，并使用不同的线性方程对其进行扩展。

算法的第一步是计算一个阈值 ω ，用于在亮度通道区分高光部分和漫反射部分，如图 3 所示。首先，图像通过使用一个大小为 $m = \max(\text{width}, \text{height}) / 50$ 的滤波器，来计算 t_1 作为滤波后亮度通道的最大值，重复使用上述操作可以使用一个 $2m+1$ 大小的滤波器来计算 t_2 。其次， t_1 被用于在原始亮度上的阈值，用于生成一个 mask，随后使用 t_2 对 mask 应用腐蚀和膨胀滤波器。这些可以应用于一些迭代以获得稳定的 mask，一般情况下几次迭代就可以达到效果。此时，mask 中像素值为 1 的点被视为镜面像素，而黑色像素被视为漫反射像素。因此， ω 被作为 L_d 中的镜面像素的最小亮度值。

在计算 ω 之后，亮度通道可以通过以下方式扩展：

$$L_w(x) = f(L_d(x)) = \begin{cases} s_1 L_d(x) & L_d(x) \leq \omega \\ s_1 \omega + s_2 (L_d(x) - \omega) & \text{otherwise} \end{cases} \quad (6)$$

$$s_1 = \frac{\rho}{\omega}, s_2 = \frac{1-\rho}{L_{d,\max} - \omega}$$

其中 $L_{d,\max} = 1$ 是由于图像被归一化处理过， ρ 表示 HDR 显示器分配给漫反射部分的百分比，这是由用户定义的。

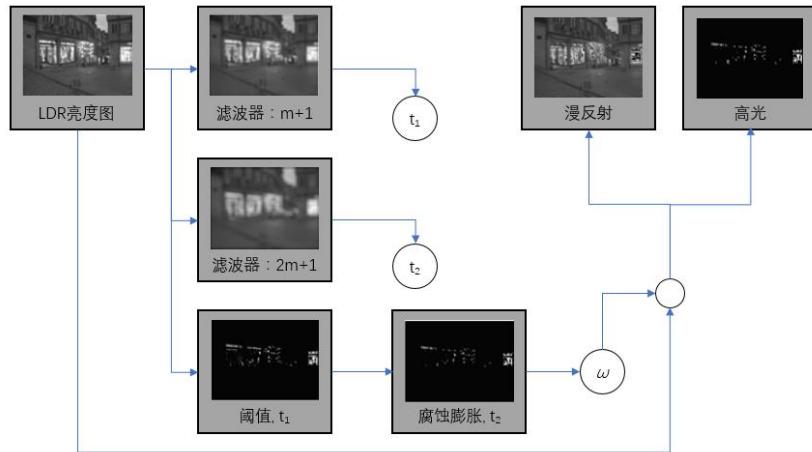


图 7-3 计算最大漫反射亮度值 ω 的计算框图

f 的全局应用可能会导致增强高光部分的量化伪像，通过在镜面反射区域应用选择性率可以减少这些伪像，图 4 显示了算法整体的流程。首先使用 $5*5$ 平均滤波器对扩展后的亮度值 $f(L_d(x))$ 进行滤波，得到 f' 。随后， $f(L_d(x))$ 和 f' 使用线性内插和 mask 进行混合，该 mask 是通过使用 ω 对 SDR 亮度进行阈值化处理并应用扩张和 $5*5$ 平均滤波器来计算。

最后，作者使用杜比 DR-37P HDR 显示器进行了一系列心理物理学实验，以确定 ρ 值。结果显示，对于室外场景，用户偏好较高的 ρ 值，这意味着分配给高光的动态范围的百分比

很小，而室内场景的百分比则相反。对于具有相同漫反射亮度的室内和室外场景，用户选择较低的 ρ 值，因此他们更喜欢将更多范围分配给高光，数据分析表明 $\rho=0.66$ 是一个很好的一般估计值。

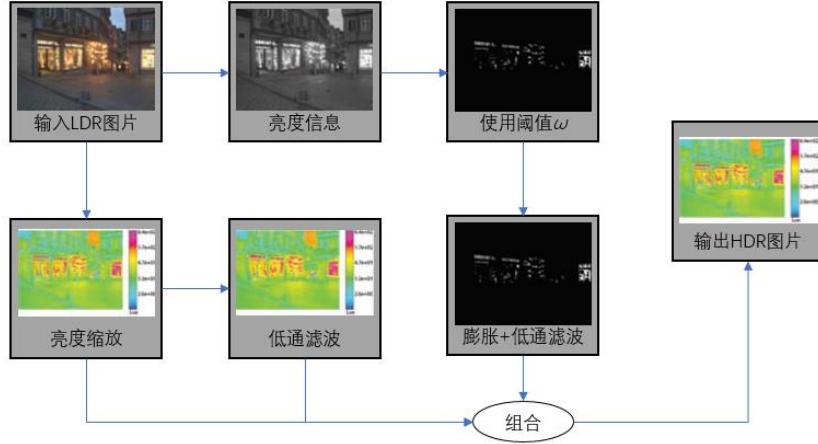


图 7-4 Meylan 方法整体框图

镜面高光相比漫反射高光部分具有更高的亮度，通常在图像采集或者处理阶段往往进行压缩或者截断，损失了一定的图像细节和对比度。Meylan 通过心理视觉实验发现，对于传统的 SDR 内容，如果直接采用线性 iTMO 进行动态范围扩展，必然会引起图像镜面高光部分泛白，图像饱和度不足，图像整体视觉表现不自然。该算法转为特定服务而设计，在 HDR 监视器上重现高光点，其他任务如增强食品的使用需要更多处理和分类操作，作者的评估实验强调了这一点。

2. HDR 显示器的明亮视频特征增强

Didyk [11] 等人提出了一个用于增强 SDR 视频亮度的交互式系统，主要用于 DVD 内容的显示和增强。这个系统的主要思想是将场景分为三个部分：漫反射、反射和光源，然后仅对反射和光源进行增强。作者提到，漫反射部分在不产生视觉伪像的情况下难以增强，并且电影制作者可能会向他们展示饱和度，而不是光源和截断反射。该系统适用于非线性值，因为目标是增强而非物理准确性。

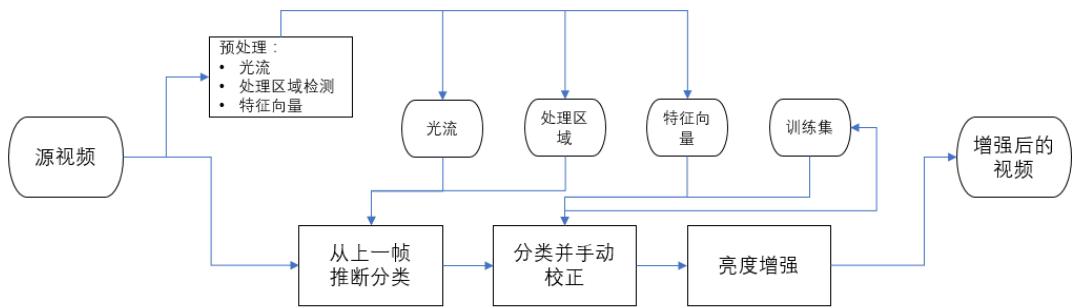


图 7-5 Didyk 方法流程图

该系统由三个主要部分组成：预处理、分类和增强裁剪区域，流程框图如图 5 所示。预

处理步骤会生成分类过程中所需的数据，特别是它使用 flood-fill 算法来确定剪切区域，至少有一个通道必须饱和，而且亮度值须大于 222。此外，在此阶段，还会计算光流以及其他功能，如图像统计、几何特征和邻域特征。

分类决定了框架中的亮区、反射和漫反射区域，并依赖于 2000 个人工分类区域的训练集。首先，具有内核 $k(z, z') = \exp(-\gamma \|z - z'\|^2)$ 的支持向量机执行区域的初始分类；随后，使用基于欧几里得度量的最近邻分类器，运动跟踪改进了初始估计：

$$d^2((z, x, t), (z', x', t')) = 50 \|z - z'\|^2 + \|z - z'\|^2 + 5(t - t')^2 \quad (7)$$

其中 z 表示区域特征， x 表示图像中的坐标， t 表示帧数。这允许在测试中使用的所有区域达到 12.6% 的分类误差，使用运动补偿跟踪修剪区域进一步减少了需要手动修正的对象的百分比，达到 3%。

通过应用非线性自适应色调曲线来增强修剪区域，这是基于存储在直方图 H 中的修剪区域内的偏导数计算的，色调曲线被定义为 H 的倒数的直方图均衡：

$$f(b) = k \sum_{j=2}^b (1 - H[j]) + t_2 \quad (8)$$

这里 t_2 是修剪区域的最低亮度值， k 是限制到最大提升值 m 缩放因子（对于亮光为 150%，对于反射为 125%）：

$$k = \frac{m - t_2}{\sum_{j=1}^N (1 - H[j])} \quad (9)$$

值 N 是 H 中的 bin 的数量，为了避免在增强期间产生轮廓，使用双边滤波对亮度通道进行滤波，将其分离为精细细节和基础层，其在亮度扩展之后合并，该方法是半自动的，因为需要用户的介入。

7.2.3 扩展映射模型

扩展映射模型（Expand Map Models）最早是 2006 年 Banterle 在其一篇文章“inverse tone mapping” [13] 中提出的术语，表示在将 SDR 内容扩展为 HDR 的过程中使用一个引导函数指导该过程的转换。该模型在 Banterle 及 Rempe1 的文章中均有使用。

1. Banterle——使用扩展映射模型进行非线性扩展

Banterle [13] 提出了扩展 SDR 内容到 HDR 监视器的基本流程“inverse tone mapping/a framework for inverse tone mapping”，如下图 6 所示。该方法的核心即如何使用反向的 TMO 算子，并结合一定的平滑处理算法重建出过曝区域中丢失的图像信息。

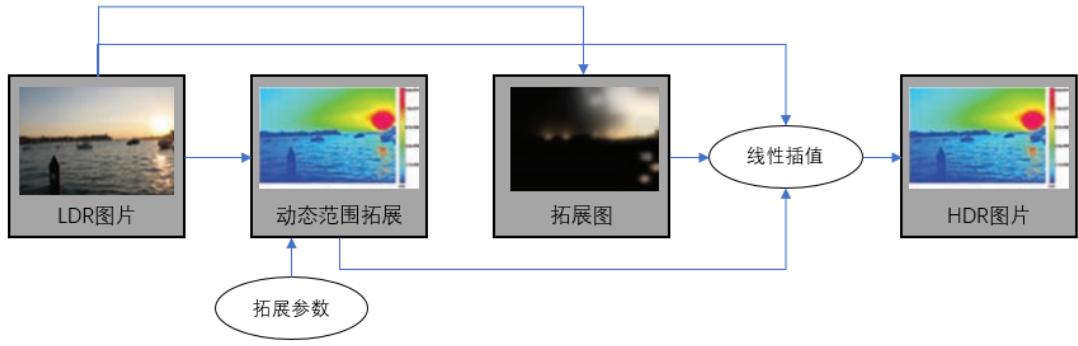


图 7-6 Banterle 方法的流程框图

本方法的逆色调映射方程可以被描述为：

$$L_w(x) = \frac{1}{2} L_{w,\max} L_{white} \left(L_d(x) - 1 + \sqrt{(1 - L_d(x))^2 + \frac{4}{L_{white}^2} L_d(x)} \right) \quad (10)$$

其中 $L_{w,\max}$ 表示输出亮度的最大值，以 cd/m^2 为单位， $L_{white} \in (1, +\infty)$ 是一个决定扩展曲线形状的参数，与对比度成比例。作者建议让这两个值取等，以此来在限制扩展导致的伪像，同时增加对比度。

在范围拓展之后，计算扩展映射是表示高亮度区域中图像的低频版本的平滑区域，它有两个主要目标，首先是重建图像曝光过度区域丢失的亮度分布；第二个目标是减弱扩展期间可以增强的量化或压缩伪像。扩展映射通过对使用重要性采样生成的样本应用密度估计来实现。最后，扩展 SDR 图像和原始 SDR 图像通过线性插值进行组合，其中扩展映射充当权值加权。需要注意的是，低亮度值保持原始值，这样可以避免在 L_{white} 设置为较高值时对低值进行压缩，否则可能会出现轮廓等伪像。

为了方便后期处理，首先需要把图像进行线性化，然后使用反向的 TMO 算子进行范围扩展。得到线性化的输入图像之后，根据输入图像，利用中位数采样强度估计的方法计算扩展映射图。其包含了图像的高亮区域的低频分量，提取高亮区域的低频分量是用于重构高亮区域中丢失的图像信息。另外，在范围扩展的过程中，量化和压缩过程必然会带来失真。使用扩展映射的另一个目的是为了尽量消除该影响。最后，将扩展映射作为插值权重，将原始的 SDR 图像和范围扩展得到的图像进行线性插值，得到最终的扩展图像。

Banterle 将该模型进行了扩展，使其可以用于自动处理图像和视频，扩展参数需要通过强度估计和试探法确定以重构高亮区域丢失的图像信息。该方法的优势在于使用扩展映射很好的保持了原始图像中的边缘信息，从而减小了最终生成的图像的边缘失真。该方法也有一定的局限性：当图像中的过曝区域过大（大于 30% 左右时），没有足够的信息用于图像区域的重建，不能通过强度估计等方法进行图像信息恢复。

2. RempeI——SDR2HDR

由于 Banterle 提出的扩展映射模型在速度方面具有一定的限制，如果不使用高端的硬件支持，很难满足实时性要求。2007 年 RempeI [8] 提出了一种基于扩展映射模型的简单方

法，该方法可以实时地将 SDR 图像进行扩展得到 HDR 图像。该方法的流程图如下图所示：

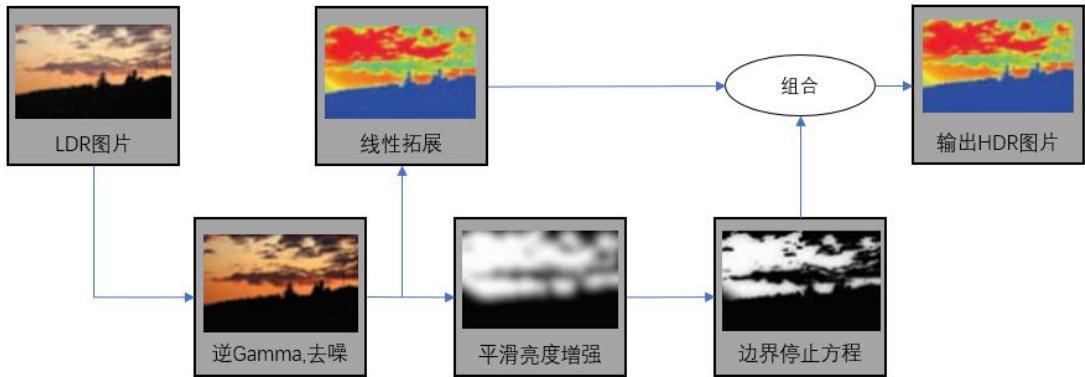


图 7-7 SDR2HDR 方法流程框图

该算法首先对输入 SDR 图像进行去线性和对比缩放的操作，一般我们看到的 SDR 图片都是经过 Gamma 校正后的图像，在逆色调映射中通常需要将图像还原至线性空间。但是经过实验发现，数码相机拍摄的照片和视频中通常还经过了传感器本身的曲线的处理，比常见的 Gamma 校正更为复杂，算法在本步骤的处理就是为了模拟数码相机的反应曲线，使得得到的图片更接近模拟胶片的效果。实现是通过实验验证设定了一个简单的 Gamma 曲线，从而得到了一种视觉上令人满意的效果。

得到线性空间的图像之后，为了适配 HDR 显示设备，将图像的亮度信息缩放到整个 HDR 显示设备支持的亮度范围内。但是这种缩放不能改变过多，否则会造成图像看起来不够自然，最重要的是将高亮部分变得更强，使其产生类似“发光”的效果。大量实验结果得出，算法设定的对比度大概在 5000:1，亮度范围在 $0.3\sim 1200\text{cd}/\text{m}^2$ 之间。随后，为了去除这种缩放所产生的噪声和量化误差，使用一个简单的滤波操作来处理图像。

算法的第二步是对图像的亮区进行增强，首先将亮区分割出来，算法认为视频中像素值高于 230、图像中像素值高于 254 的区域为亮区，分割出来存为 mask，再获得图像的轮廓图用于之后的处理。

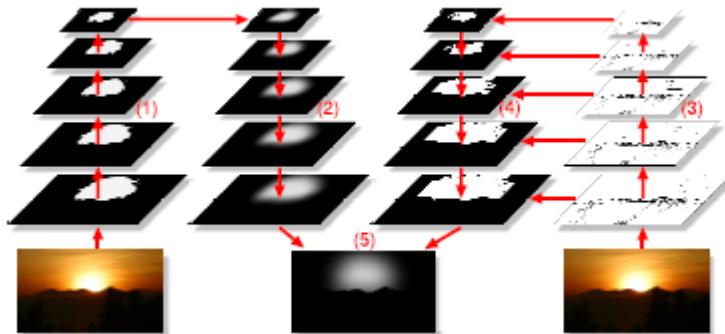


图 7-8 使用图像金字塔进行亮度增强

如图 8 所示，亮度增强是通过图像金字塔进行的。首先是过程(1)对图像的亮区 mask 进行下采样，得到光源的信息，然后如过程(2)对结果使用最近邻插值上采样，得到平滑后的亮区光源信息；同时(3)对轮廓图也同样进行下采样处理，得到图像的强轮廓信息，在过程(4)上采样亮度 mask 信息时，使用同分辨率下的轮廓信息，在上采样的过程中，遇到强轮廓就停止采样，从而使得亮区不会产生过量扩散的错误。最后(5)就是将平滑过的亮度信息与 mask 融合，得到最终 HDR 图像的高亮区域亮度信息。

该算法本质上可以算作是一种基于 Expand Map 的方法，主要对 SDR 图像上无法显示的

高亮区域进行增强，使得图像在 HDR 显示器上可以更为自然的显示出来。由于算法的计算复杂度较小，可以达到实时处理的效果，这相比之前 Banterle 的实时方法对硬件的要求更低。但是本算法只对高亮区域处理，而对正常亮度区域的增强不明显，在正常场景拍摄的图片中增强效果不明显，且算法没有考虑色域增强，只是简单地增加饱和度来实现更鲜艳的颜色。

7.2.4 近年提出的逆色调映射算法

1. 适用于各种曝光的高质量逆色调映射

Kovaleski 和 Oliveira [22] 在 2014 年提出了一种增强算法，这种方法基于交叉双边滤波，并能够很好的处理曝光不足或过量的情况。

双边滤波是一种非线性滤波器，同时对空间域和亮度范围进行滤波。使用一个高斯方程 G 作为下降函数，则双边滤波后的 I^b 可以被定义为：

$$\begin{aligned} I_p^b &= \frac{1}{W_p^b} \sum_{q \in S} G_{\sigma_s}(\|p - q\|) G_{\sigma_r}(|I_p - I_q|) I_q \\ W_p^b &= \sum_{q \in S} G_{\sigma_s}(\|p - q\|) G_{\sigma_r}(|I_p - I_q|) \end{aligned} \quad (11)$$

其中 I_p, I_q 代表在图像在位置 p 和 q 上对应的像素值， σ_s, σ_r 表示空间和距离上的高斯函数标准差， W_p^b 是一个归一化因子。将其中的 $G_{\sigma_r}(|I_p - I_q|)$ 部分替换为另一幅图片，那么上式就可以得到交叉双边滤波：

$$\begin{aligned} I_p^b &= \frac{1}{W_p^b} \sum_{q \in S} G_{\sigma_s}(\|p - q\|) G_{\sigma_r}(|E_p - E_q|) I_q \\ W_p^b &= \sum_{q \in S} G_{\sigma_s}(\|p - q\|) G_{\sigma_r}(|E_p - E_q|) \end{aligned} \quad (12)$$

直观地认为，高斯滤波器的权重基于图像中位置 p 和 q 之间的空间距离和图像中 p 和 q 位置像素值的差。由于该滤波器的非线性特性，在实时应用受到限制。有些环节可以通过使用不同的方法来近似结果，以交互式速率进行计算。Durand 和 Dorsey 通过线性内插几个离散化的强度值来计算滤波器的响应，这些强度值在频域中用高斯内核滤波。Paris 和 Durand 将这项工作解释为一种高维线性卷积，其后是三线性插值和分裂，从而扩展了这项工作。这种想法的推广是使用双边网格，双边网格 Γ 是一个有规则取样的三维阵列，其中前两个维度定义图像的空间域，而第三个维度表示图像的范围。使用双边网格进行滤波操作，分为三个步骤进行：网格创建、处理和切片。为了处理双边网格，可以将三维函数应用其中，我们得到一个新的网格 $\Gamma' = f(\Gamma)$ 。对于双边滤波器， f 是高斯核的卷积，分别使用 σ_s, σ_r 作为域和距离维度上的方差。切片操作用于从双边网格中提取分段平滑的二维图像。

在双边网格中，使用图像 E 来执行交叉双边滤波以确定网格位置，同时存储来自数据图像 I 的像素值。切片操作随后使用边缘图像来恢复结果。Adams 等人通过统一单形有效地实现了 5D 彩色双边滤波，与 Paris 和 Durand 的方法相比，他们使用了更简单的插值操作。最近 Gastal 和 Oliveira 提出了一种计算适应信号的流形，这些流形之间的加权插值可以得到更高精度和更快速率的输出，并使用比之前方法更少的采样点数量。

Kovaleski 和 Oliveira 提出的方法通过使用双边网格生成一个亮度增强函数，网格使

用两个不同的图片生成，其意义上更像是交叉双边滤波。但是切片操作通过使用第三幅图片来实现，其方法可以被写为：

$$bf(I) = s_B(G_{\sigma_s, \sigma_r} \otimes g_A(C)) \quad (13)$$

其中 A, B, C 都是单通道图片， A 中的每个像素都包含对应 RGB 通道的图像中的最大强度值，图像 B 包含图像 I 的亮度信息，图像 C 包含图像 A 中所有像素值超过某阈值 t 的所有像素点。进一步，RGB 通道中的最大强度值（图像 A ）被用作检测饱和像素的默认方式，正如我们将要展示的那样，这个选择会影响该算法处理过度曝光内容的能力。在生成图像 A, B, C 之前，图像使用 Gamma 2.2 曲线进行线性化处理， $g_A(C)$ 是一个双边网格，使用图像 I 的边缘信息进行过滤，同时存储图像 C 的值。因此双边滤波公式可以被写为：

$$\begin{aligned} I_p^b &= \frac{1}{W_p^b} \sum_{q \in S} G_{\sigma_s}(\|p - q\|) G_{\sigma_r}(|B_p - A_q|) C_q \\ W_p^b &= \sum_{q \in S} G_{\sigma_s}(\|p - q\|) G_{\sigma_r}(|B_p - A_q|) \end{aligned} \quad (14)$$

一个关键点在于，上式中使用最大强度值图像 A 倾向于导致 Kovaleski 和 Oliveira 的算法无法增强过度曝光的 SDR 图像的细节。通过使用最大强度信道值，他们的算法尽管在感知上存在差异，但他们对 RGB 三通道的信息平等处理，这种不一致反映在所得到的亮度增强函数中。这个关键点同样可以解决这个问题：用亮度图像 B 代替最大强度通道图像 A ，这种操作不仅解决了曝光过度的图像的问题，而且极大地简化了亮度增强函数生成过程，将以上公式变为交叉双边滤波形式：

$$\begin{aligned} I_p^b &= \frac{1}{W_p^b} \sum_{q \in S} G_{\sigma_s}(\|p - q\|) G_{\sigma_r}(|B_p - B_q|) C_q \\ W_p^b &= \sum_{q \in S} G_{\sigma_s}(\|p - q\|) G_{\sigma_r}(|B_p - B_q|) \end{aligned} \quad (15)$$

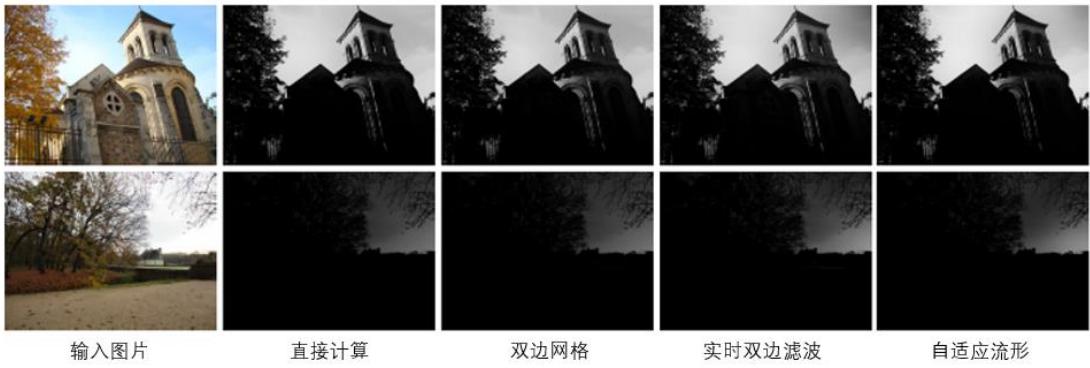


图 7-9 不同双向滤波算法的结果对比

公式(15)相比与之前公式来说，可以使用多种不同的双边滤波加速方法，如图 9 所示是使用不同种加速方法的效果，对比了不使用加速的双边滤波、双边网格、实时双边滤波和自适应流形四种结果。此外，滤波参数中本算法取 $\sigma_s = 150$ ，这时模糊滤波器的滤波结果不容易被人类视觉系统察觉；同时，另一个滤波器参数 $\sigma_r = 25$ 是通过大量实验测量得出，实验表明，这个取值更有利得到亮度增强函数。

根据计算存储的亮度增强函数，作者使用和 Rempe1 算法类似的步骤：输入图像 I 使用 Gamma 2.2 曲线进行线性化，使用上面公式计算得到亮度增强函数，随后将函数缩放至范围 $[1..\alpha]$ 内，在实验中 α 取值为 4。最后，线性化输入图像的值根据目标显示能力和选择的 α

值进行性缩放，最低亮度为 $0.3\text{cd}/\text{m}^2$ ，最高亮度为 $1200\text{cd}/\text{m}^2$ 。然后通过重新调整的亮度增强函数将该缩放的图像逐点相乘来获得最终的 HDR 图像。

生成效果图如图 10 所示，结果图使用 DRIM 评价方法进行展现。



图 7-10 DRIM 对结果图进行评价

2. 基于视网膜反应的逆色调映射

Huo [24] 在 2014 年提出一种基于人类视觉系统视网膜反应的逆色调映射算法，通过生理学上的研究，该方法可以避免出现很多现有方法都会发生的伪像问题。受到人类视觉系统特性的启发，这种逆色调映射方案的计算复杂度低，参数数量较少，并能获得高质量的 HDR 图像输出。

与真实世界场景的动态范围相比，人类视网膜的感光细胞具有较窄的动态范围，这要归功于人类视觉系统的适应机制，它使得眼睛首先适应一些亮度值，然后以相当小的动态感知图像范围围绕该亮度值。因此，人类视觉的基本过程是对整个场景的全局色调映射，这种全局功能可以用视网膜的反应和光强刺激之间的关系来描述：

$$\frac{R}{R_{\max}} = \frac{I^n}{(I^n + \sigma^n)} \quad (16)$$

其中 R 表示视网膜对光强 I 的刺激反应， R_{\max} 是反映的最大值， σ 是全局适应等级，它表示产生的相应强度是 R_{\max} 振幅的一半所需的强度，参数 n 是灵敏度控制指数。

实验证明，场景的绝对亮度信息对于人类视觉系统来说是次要的，通过亮度很定性的机制在视觉处理的非常早期阶段往往会被大量舍弃，而是用局部对比来传达广域场景的大量信息。因此，视网膜反应的更恰当描述应该是局部映射，局部自适应属性可以通过将全局自适应水平 σ 改编为局部像素 p 的自适应水平 σ_p 来实现。设 I_p 表示像素 p 在实际场景中的辐射

强度值，视网膜反应 R_p 被认为是眼睛感知到的 SDR 图像的强度值，可以表示为：

$$R_p = \frac{R_{\max} I_p^n}{(I_p^n + \sigma_p^n)} \quad (17)$$

对于逆色调映射， I_p 可以通过以下方式计算， δ 是为了避免分母为 0：

$$I_p = \left(\frac{R_p \sigma_p^n}{(R_{\max} - R_p + \delta)} \right)^{1/n} \quad (18)$$

之前的研究认为灵敏度参数 n 通常取值在 $0.7 \sim 1.0$ 之间，作者通过大量实验得出，取值为 0.9 时结果更好。 σ_p 是局部适应水平，它描述了一个像素邻域强度信息，通常可以使算术平均值、几何平均值或图像局部区域的高斯模糊来确定该参数。这里，HDR 图像中像素 p 周围的局部强度 I_{ps} 用来作为 σ_p 的取值。在仅有 SDR 图像的情况下，基于 SDR 图像的最大亮度 255 被映射到 HDR 显示器的最大亮度的假设，算法首先计算 SDR 图像的 L_{ps} ，随后将其乘以 HDR 显示器最大亮度与 255 的比值。

很多方法可以计算得到 L_{ps} ，论文使用 Durand 和 Dorsey 提出的双边滤波，逆色调映射公式可以写为：

$$I_p = \left(\frac{R_p (I_{ps}(\sigma_m, \sigma_d))^n}{(R_{\max} - R_p + \delta)} \right)^{1/n} \quad (19)$$

输出的像素 p 的双边滤波 L_{ps} ：

$$L_{ps}(\sigma_m, \sigma_d) = \frac{1}{W_p \sum_{q \in \Omega} f_{\sigma_m}(q-p) g_{\sigma_d}(R_q - R_p) R_q} \quad (20)$$

其中 W_p 是归一化因子：

$$W_p = \sum_{q \in \Omega} f_{\sigma_m}(q-p) g_{\sigma_d}(R_q - R_p) \quad (21)$$

其中 σ_m 表示高斯函数 f 在空间域的标准差：

$$f_{\sigma_m}(p | p = (x, y)) = K_m \exp \left\{ -\frac{(x^2 + y^2)}{\sigma_m^2} \right\} \quad (22)$$

其中 σ_d 表示高斯函数 g 在亮度域下的标准差， K_m 是归一化因子， Ω 表示整幅图像。 σ_m 和 σ_d 分别取值为 16 和 0.3。如图 10 所示为方法计算结果与其他两种方法的对比。

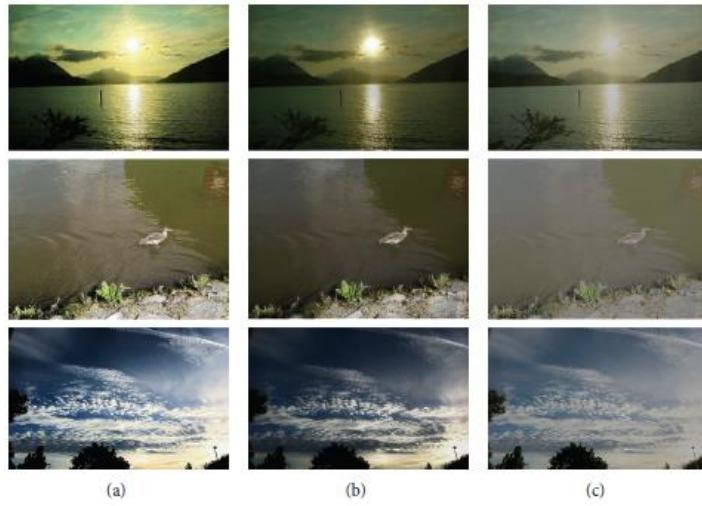


图 7-11 算法效果图和与其他方法的对比: (a) 论文方法, (b) iPG, (c) SDR2HDR

3. 基于深度学习的逆色调映射

Endo [23]针对逆色调映射中过度曝光和曝光过少、细节损失的情形进行优化，提出了一种基于深度学习的方法，其网络基于卷积神经网络，将输入图像转化为多幅不同曝光度的图片，再组合为一张 HDR 图像输出。

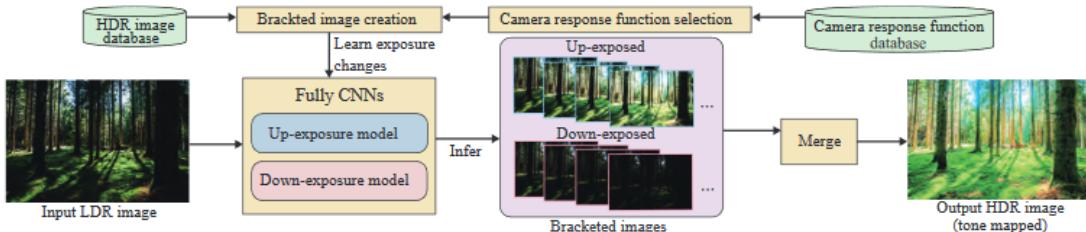


Fig. 2. Overview of the proposed method. The flow is decomposed into the learning and inference phases. In the learning phase, the bracketed LDR images are first created from HDR databases by simulating cameras. Next, we let our fully CNNs learn the changes in the exposures of the bracketed images. In the inference phase, the learned CNNs compute LDR images with different exposures from a single input LDR image. The final HDR image is then generated from these bracketed LDR images.

图 7-12 Endo 提出的方法流程图

12 所示是方法的流程框图。首先，作者通过使用不同的非线性相机反应曲线来改变图像的曝光度，由此形成了多幅不同曝光度的图片，数据集就是以 SDR 图像为输入，多幅不同曝光度的图片为输出组成。其计算方法如下：

$$Z_{i,j} = f(E_i \Delta t_j) \quad (23)$$

$$\Delta t_j = \frac{1}{\tau^{T/2}}, \dots, \frac{1}{\tau^2}, \frac{1}{\tau}, 1, \tau, \tau^2, \dots, \tau^{T/2}$$

其中 $Z_{i,j}$ 表示 SDR 图像对应点 i 在曝光指数 j 下的像素值， f 表示相机反应函数， E_i 表示 HDR

图像在点 i 处的像素值。 Δt_j 为曝光时间，作者通过指数定义，实验中采用 $T = 8, \tau = \sqrt{2}$ 。随后，在使用得到的多张不同曝光度的照片，通过合并算法进行计算，从而得到一张完整的 HDR 图片。

Endo 使用的是编解码器结构，并在对应层间有跳跃式连接。先将一张 SDR 图片编码至深度网络，这里通过 2D 卷积神经网络实现，随后再将图像的深度语义特征解码为不同曝光度的 SDR 图片，这时使用的是 3D 卷积神经网络。这种网络用于将图像的曝光度提高或降低，从而得到多幅图片用于之后合成 HDR 图片。

在具体网络结构方面，编码器网络有 9 层卷积核大小为 4 的卷积层，并包含批归一化和 Leaky-ReLU。解码器网络使用 3D 去卷积神经网络，其中前三层的卷积核为 $4 \times 4 \times 4$ ，其余层为 $3 \times 3 \times 3$ 。类似地，出最后一层外，其他层都包含批归一化和 Leaky-ReLU，而最后一层的激活函数为 Sigmoid。

在这个编解码网络中，解码器使用完全编码的向量，这意味着整个图像的潜在特征都包含在内，为了将输入图像中的局部和底层信息逐步融入解码器，作者在编解码器的对应层之间添加了跳跃式连接，具体网络结构图如图所示：

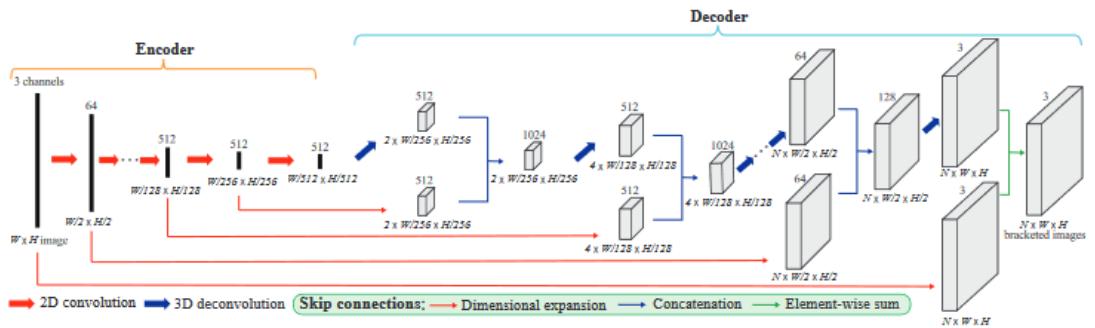


Fig. 5. Our network architecture. The same architecture is used both for up- and down-exposure models.

图 7-13 网络结构图

通过长达一个月的训练，取得了满意效果。其与一些传统算法的比如下图所示。

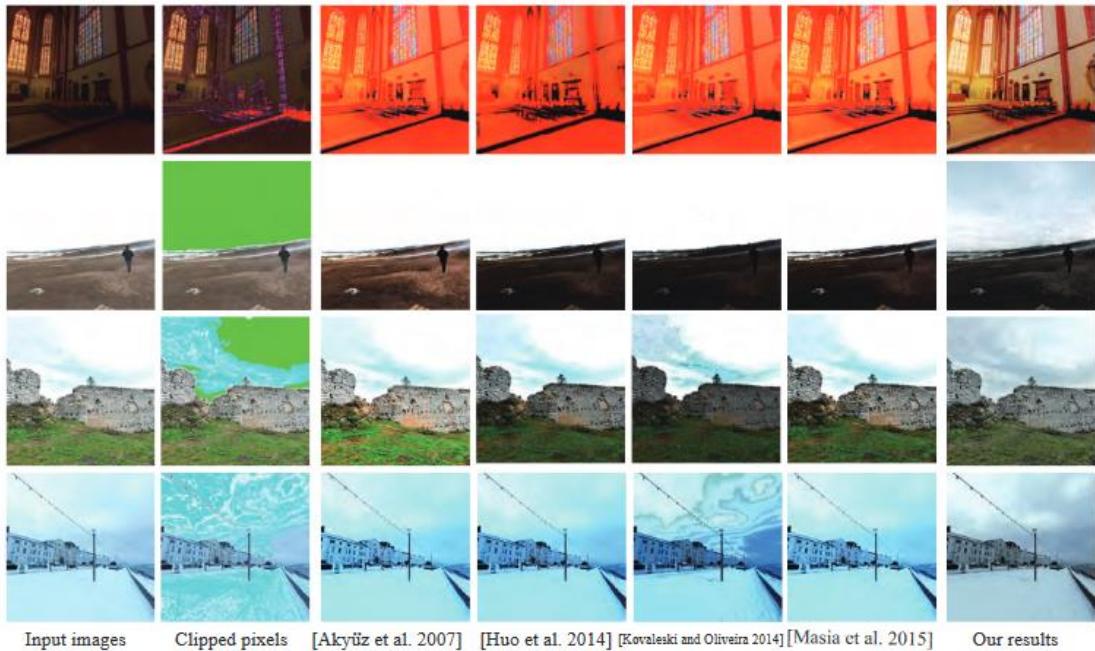


Fig. 6. Tone-mapped results of HDR images generated by each rTMO. To visualize the clipped pixels in the input LDR images, pixels of the images in the second column are painted in red/green if they are completely under/overexposed (i.e., black/white) or in magenta/cyan if one or two channels are under/overexposed, respectively. While there is amplified noise in the dark regions of other results, our results exhibit natural tone reproduction with the least noise. Please enlarge images in the electronic version.

图 7-14 Endo 算法效果图和与其他方法的对比

4. 基于生成对抗网络的逆色调映射模型

S. Ning 等人[3]认为由于逆色调映射过程相当于是向低动态范围的图像中添加暗区和亮区的细节，希望可以达到一个信息增加的过程，因此采用生成对抗网络的对抗机制可以很好的完成生成任务。生成对抗网络是由 I. Goodfellow 提出的基于博弈的模型，其模型本身由两个网络构成，一个是生成网络，在图像任务中用于生成图像，实现图像的转换任务；另一个是判别网络，用于判别输入给判别网络的图像是由生成器生成的还是源图像。生成对抗网络的训练过程就是，通过训练生成网络使得生成的图像越来越接近源图像，从而可以“蒙骗”判别网络进行错误的决策；训练判别网络的过程就是使判别网络可以更准确的分辨出生成的图像和源图像，从而使得这两部分网络的训练过程成为一种博弈式对抗过程，期望的训练结果是能够得到一个稳态。

算法上讲，GAN 的判别网络可以看成是将图像数据映射到判别概率的函数 $D: D(x) \rightarrow (0,1)$ ，对于一个固定的生成器 G ，判别器 D 可能被训练用于粪便图像是来自训练数据（概率接近 1）还是来自生成器（概率接近 0）。这种方法的优点是其分类边界灵活，使用纯概率的方法可以得到挖掘信息更深的模型，这种模型可以清晰的判别出多类别或某一类与其他类之间存在的差异特征。

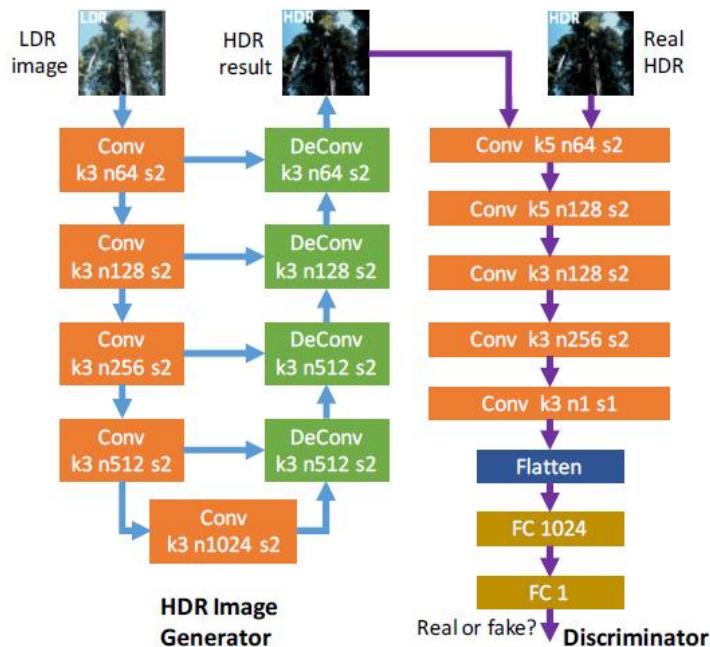


图 7-15 网络结构图

作者在论文提出了一种 iTMN 可以用于将 SDR 转换为 HDR 的生成对抗网络，网络结构如图 15 所示，生成网络采用 U-Net 类似结构，这是因为 U-Net 可以在图像重构的过程中同时使用深层信息和浅层信息一步一步地重建图像，因此 U-Net 会很适合本任务。用于约束网络的损失函数是一种混合损失函数，包括生成对抗网络定义相关的对抗损失函数和生成网络内容相关的内容损失函数：

$$\min_G \max_D \underbrace{\lambda L_G(L, H)}_{\text{content loss}} + \underbrace{R_{G,D}(L, H)}_{\text{adversarial regularizer}} \quad (24)$$

其中内容损失函数包含两个部分，分别是均方差和差分均方差，希望通过均方差从像素值方面上约束网络，通过差分均方差在周围像素点的关系上进行约束：

$$E_{(L,H) \sim P_{data}} \left[\|G(L) - H\|_F^2 + \alpha \left(\|d_x G(L) - d_x H\|_F^2 + \|d_y G(L) - d_y H\|_F^2 \right) \right] \quad (25)$$

除此之外，对抗损失函数即由 GAN 网络的定义得出，可以被归纳为：

$$E_{H \sim P_{data}} [\log(1 - D(H))] + E_{L \sim P_{data}} [\log(D(G(L)))] \quad (26)$$

数据使用 HDR 视频切片单帧图片，经过多个色调映射算法处理后选取效果最好的 SDR 图像与原生 HDR 图像组成数据集，训练使用的训练集包含 2660 对图像，测试集包含 140 对图像，实现采用 TensorFlow 框架。训练后网络在测试数据集上做验证，评测方法采用 HDR-VDP-2、mPSNR 和 SSIM，结果与 Huo [24]、KO [22] 和 DrTMO [23] 做对比，并同时与提出结构的两种变式做对比，其中 NoDMSE 是提出的结构中除去 L2 部分，NoAdvReg 是提出结构除去对抗部分，即使用单独的 U-Net，评价结果如下：

Table 1: Comparison for various methods

Method	iTMN	NoDMSE	NoAdvReg	Huo [2]	KO [3]	DrTM [4]
HDR-VDP-2	73.77	72.47	70.52	60.84	56.45	73.21
mPSNR	21.61	20.62	20.07	20.64	14.25	18.73
SSIM	0.8098	0.7298	0.7242	0.7707	0.6459	0.7968

该方法在三种不同的评价指标效果均有提升，说明该算法在对比度提升、图像内容增强和图像结构信息上都有一定的进步。除此之外，也对网络的参数进行了一些实验，关于是否需要运用 GAN 模型在发色调映射任务上的问题，从之前的数据和图 16 所示的效果图中都显示，使用 GAN 可以得到更好的转换效果。

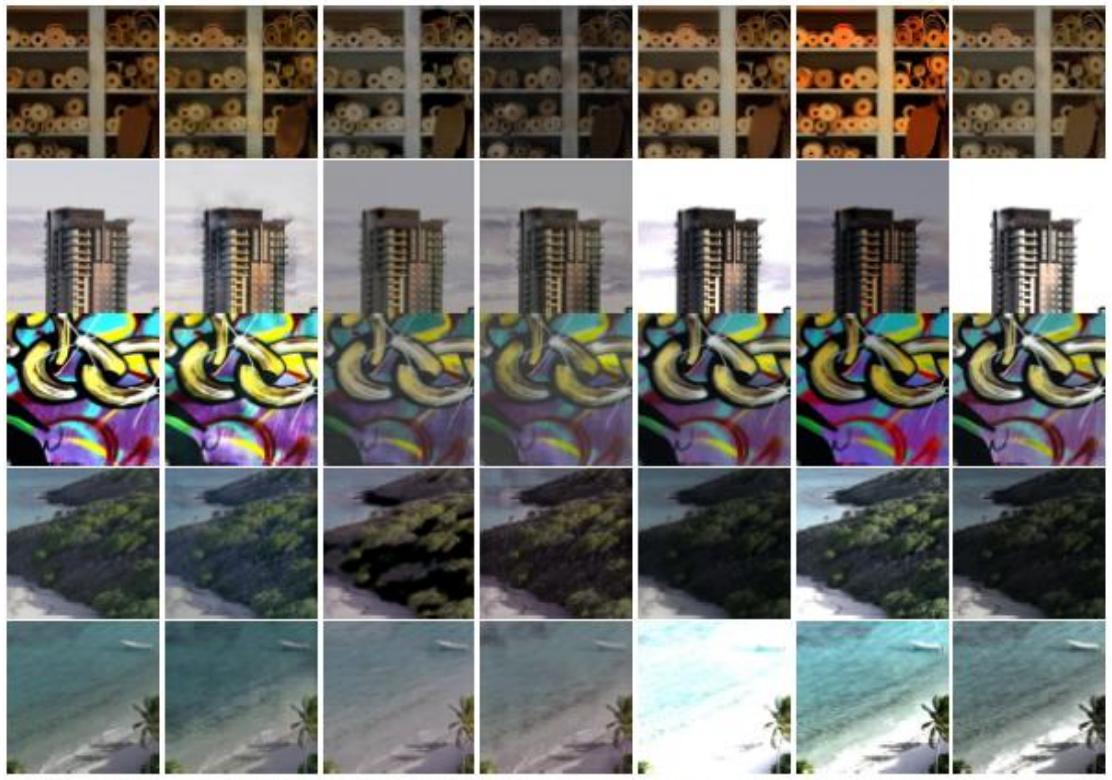


图 7-16 S. Ning 算法效果图和与其他方法的对比

5. 利用深度卷积网络的单次曝光 HDR 重建

相机的感光元件只能一时间捕捉范围有限的亮度，因此传统方法中很多都是使用多个曝光度的照片来合成 HDR 图片。Eilertsen 在文章[27]中提出了一种使用单次曝光的图像来进行 HDR 图像重建，本质是在亮度饱和区域上预测丢失的信息，这种任务很适合用深度学习的算法来解决，因此算法是基于一种深度卷积网络（CNN）来预测丢失的 HDR 信息。

算法包括以下步骤：

首先，算法需要对输入的 SDR 图片进行线性化、颜色校正等预处理。线性化方法中，为了降低复杂度而使用了一种平均的相机响应函数来线性化图像；颜色校正上，也是使用固定的白平衡转换来简化预处理的过程；量化恢复若使用 CNN 来完成则更接近超分辨率等工作，因此本方法中并未进行量化恢复的处理。

随后，算法的模型使用自动编解码器结构实现，对应层也使用了跳跃式连接，但与 U-Net 不同的是，网络的解码器部分是在图像的对数域进行处理，因此跳跃式连接还需要将图像的特征转换到对数域再与解码器的各层输出结合，可以通过公式(27)表示：

$$\bar{h}_i^D = \sigma \left(W \left[\log(f^{-1}(h_i^E)) + \varepsilon \right] + b \right) \quad (27)$$

其中 h_i^E 和 h_i^D 分别表示编码网络和解码网络对应层的输出， f^{-1} 是将图片线性化的处理函数，随后的操作就是将两部分信息组合起来作为混合信息送给下一卷积层处理。网络的结构如图 17 所示：

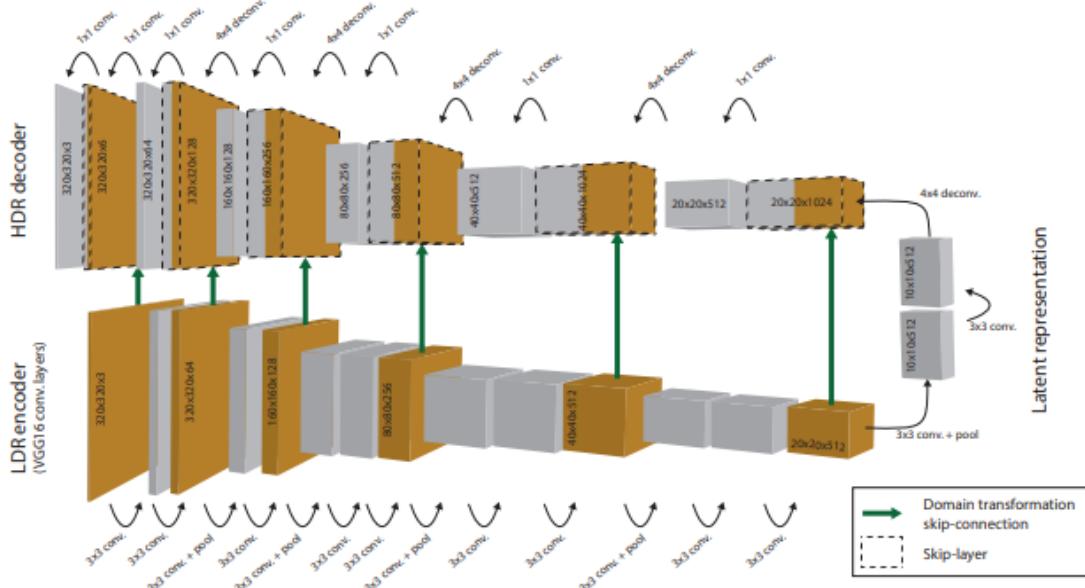


图 7-17 Eilertsen 提出方法的网络结构

最终输出的 HDR 图像也是由网络输出的对数域图片和输入图片线性化后加权得到，可以由公式(28)表示：

$$H_{i,c} = (1 - \alpha_i) f^{-1}(D_{i,c}) + \alpha_i \exp(\hat{y}_{i,c}) \quad (28)$$

其中 $D_{i,c}$ 表示输入的 SDR 图像， $\hat{y}_{i,c}$ 表示网络输出的对数域的 HDR 图像， i, c 分别表示图像

的标号和颜色通道， α_i 是用户自定义取值，用于调整两种图像来源的权重。

最后，网络的损失函数是通过将图像分解为亮度分量（Illumination）和漫反射分量（Reflectance），用于得到图像不同维度的信息，然后对两种分量转换至对数域后，在 L2 范数下进行约束，最终的损失函数由两个分量在对数域上的 L2 范数差的加权和来确定，可以由公式(29)表示：

$$L(\hat{y}, H) = \frac{\lambda}{N} \sum_i \left| \alpha_i (\log(\hat{I}_i) - \log(I_i^y)) \right|^2 + \frac{1-\lambda}{3N} \sum_{i,c} \left| \alpha_i (\log(\hat{R}_{i,c}) - \log(R_{i,c}^y)) \right|^2 \quad (29)$$

其中 I, R 分别表示图像的亮度分量和漫反射分量， \hat{y}, y 分别表示网络的输出图片和真实图片， λ 是用户指定的权值参数。训练使用的是预训练的模型，实验表明经过预训练的模型可以重建出局部亮度更高的 HDR 结果，这与真实 HDR 图像更为接近，随后对网络使用常用参数在与 Endo[23]相同方法下构建的数据集进行训练，训练结果如图 18 所示，可以看出得到的模型能够较好的重建 HDR 图像的高亮部分，这与真实的 HDR 图片很接近。



图 7-18 Eilertsen 使用的深度学习方法的重建结果

第八章-色域映射

在之前的色调映射章节中提到：在色调映射环节，为了便于操作，且不使图像颜色产生巨大失真，色调映射算法通常会仅处理图像亮度信息，将 HDR 图像亮度映射到 SDR 图像亮度域中，通过原 HDR 图像的颜色信息，恢复并重建 SDR 图像的颜色信息。由于前面的主题是色调映射，因此颜色转换相关技术没有深入介绍。但颜色转换或色域映射问题（Color Transfer or Gamut Mapping），也是 HDR 的重要环节。本章将介绍 HDR 中颜色转换(或色域映射)技术。本章分为两个部分，第一部分介绍色域映射的定义以及相关背景知识；第二部分将介绍代表性的色域映射算法，特别对 ITU 中相关标准进行浅析。

8.1 色域映射概述

在日常生活中，人们会在多种场景和设备上观看图像或视频。由于显示设备的多样性，自然会引发一个问题：给定一幅图像，如何能够保证人们在这些不同的显示设备上看到的图像是一致的？更进一步，如何使得不同显示设备上看到的图像，与这幅图像的原场景是一致的？

由于不同显示设备原理、设计、颜色标准不尽相同，想要实现这一目标是非常困难的。如果希望不同的人在不同的显示设备上看到的图像一致，首先需要了解哪些原因造成不一致。原因可以大致分为三类：

1. 不同人视觉特性不同。不同的人，对于进入眼睛的光线的响应是存在轻微的不同的，因此，即使在同一场景下观察同一图像，不同的人对于图像的感知（例如颜色，对比度，饱和度等）都是不同的。
2. 观测场景不同。同样的图像，在不同的场景下观测，同样观测者感受也是不同。
3. 显示方式不同。对于一幅数码相机拍摄的图像，将存储的图像文件使用不同的显示设备显示，人们观察到的图像也是不同的。

色域，来用于描述不同显示设备，或者不同标准所使用或规定的颜色范围。而在两个不同色域之间构建的映射，称为色域映射。由于视觉特性以及场景原因带来的差异难以建模并且处理，因此色域映射问题通常是指显示设备的显示方式不同产生的色域差异。一般来说，可以通过改变像素值将两个色域中均含有的颜色进行映射，以实现颜色的一致。但当两个色域范围有差别，也就是说一个色域中包含了另一个色域没有的颜色时，色域映射就比较复杂，通常的做法是选择最接近于原图像颜色的颜色，进行显示。

在 CIE(2004d) [2] 标准中，给出色域映射及其相关内容的官方定义：

色域映射：一种将待显示媒介的颜色用原始图像或媒介中的颜色表示的方法

图像：一种包含了图形与形象的二维视觉刺激。重新生成的图像可以通过某些特征与原始图像进行比较。

颜色媒介：能够显示或者获取颜色的媒介物。例如 CRT 显示器，数码相机。

色域：在给定的观测条件下，一个给定的颜色媒介能够显示或获取的颜色范围。

由此可以明确，色域映射是将两种不同的色域联系起来的核心方法。色域映射的分类主要包括四大类：独立（两个色域之间没有交集），交叉（两个色域相交，但不包含），缩小（目标色域小于原色域），扩张（目标色域大于原色域），具体见图 1。其中 S 表示原色域， D 表示目标色域。

前面部分介绍了色域映射的基础，后续部分中将把注意力转向 HDR 中是如何进行色域映射的。更多关于传统色域映射的内容与算法，可参阅文献 [1]，该书包含了关

于传统色域映射（即不同显示设备之间）的丰富内容。

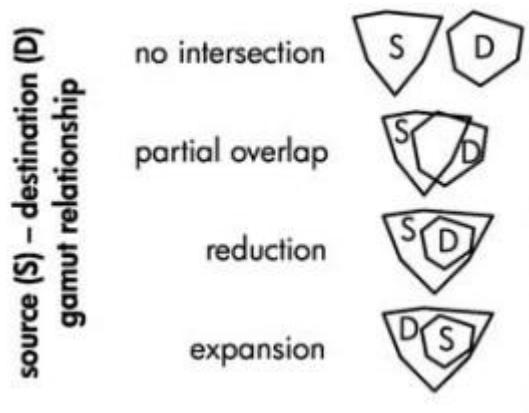


图 8-1 色域映射的分类[1]

8. 2 HDR 中的色域映射标准与算法

在 HDR 技术中，色域映射的目的有所不同。传统的色域映射方法，是为了解决不同显示设备之间色域不同的问题，但是归根结底，传统显示设备的色域都在 BT.709 (ITU, 1990) 标准范围内。而在 HDR 中，由于 HDR 图像的亮度与颜色是与场景亮度与颜色一致的，因此理论上来说，HDR 图像显示在不同的 HDR 显示设备上时，颜色与亮度也不会存在差异，因此 HDR 的色域映射方法并不是用于解决不同显示设备的颜色不一致问题的，而是主要用于色调映射过程中的颜色处理。由于色调映射过程中，HDR 图像映射成了 SDR 图像，因此，图像的亮度与颜色都发生了较大的改变。而对于颜色来说，由于 HDR 图像的色域标准遵循 BT.2020 (International Telecommunication Union, 2015) 标准，而 SDR 图像遵循 BT.709 标准，因此，本质上来说，HDR 中的色域映射问题，实际上就是 BT.2020 色域与 BT.709 色域之间的转换问题。下面，我们将从 ITU 标准入手，介绍 HDR 的色域映射标准与相关算法。

8. 2. 1 BT. 709 与 BT. 2020 颜色空间

在 1931 年，CIE 根据相关实验，测定了人眼能够观察到的颜色范围，由此确定了“CIE 1931 标准观察者标准”，也被称为 CIE-XYZ 表示法。而在图像与视频处理领域，显示设备所能显示的颜色，远少于自然界中的颜色范围，由于技术的限制以及显示设备的制作原理的限制，现在的显示设备无法显示出自然界中存在的所有颜色，而只能选择其中的一部分颜色进行显示。

在 HDR 技术中，通常使用的颜色标准是 ITU-R BT. 2020 标准。BT. 709 是 HDTV 的标准，也是目前运用最为广泛的颜色标准。BT. 2020 是 UHDTV 的颜色标准，也是 HDR 技术所支持的颜色标准。HDR 中的色域映射，就发生在这两个色域之间。

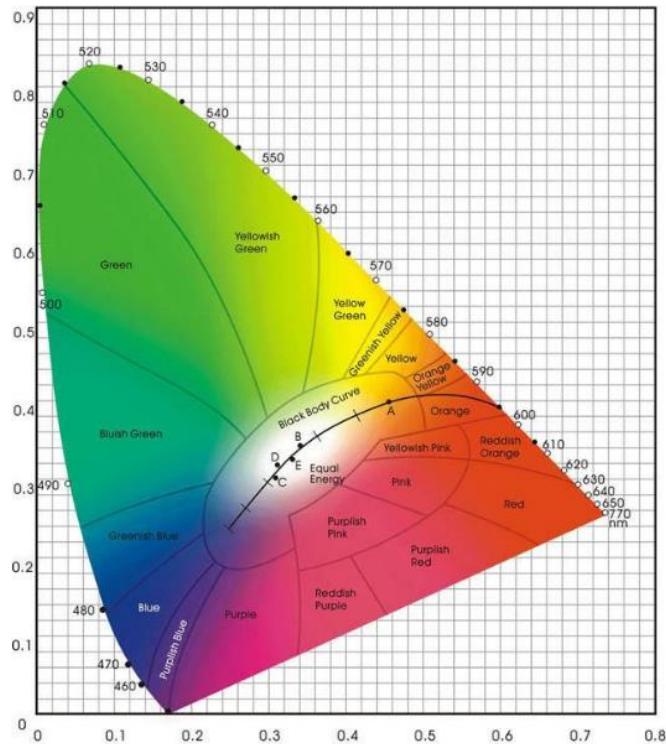


图 8-2 CIE-XYZ 1931

BT. 709 色域

BT. 709 标准中定义了 SDR 图像的编码方式。具体分为 $R'G'B'$ 和 $Y'C_bC_r$ 的两种编码模式，每种模式的各通道都有 8 或 10 比特两种类型，在 8 位编码中， $R'G'B'$ 通道和 Y' 通道的标称范围是 16~235， C_b 和 C_r 通道的标称范围是 16~240，中性值为 128。因此，在 $R'G'B'$ 中，黑色的值是 (16, 16, 16)，白色的值是 (235, 235, 235)；在 $Y'C_bC_r$ 中黑色的值是 (16, 128, 128)，白色的值是 (235, 128, 128)。标准 RGB 空间的 0 和 255 仍保留作为定时参考，可能不包含颜色数据。BT. 709 色域的范围见图 3。

BT. 709 标准同时支持模拟信号与数字信号，因此在 BT. 709 标准内部，存在两个重要的转换： $R'G'B'$ 编码到 $Y'C_bC_r$ 编码的转换，以及模拟信号到数字信号的转换。

$R'G'B'$ 编码下的模拟信号到 $Y'C_bC_r$ 编码下的模拟信号之间的转换为：

$$\begin{aligned} E_Y &= 0.2126E_R + 0.7152E_G + 0.0722E_B \\ E_{CB} &= \frac{-0.2126E_R - 0.7152E_G + 0.9278E_B}{1.8556} \\ E_{CR} &= \frac{0.7874E_R - 0.7152E_G - 0.0722E_B}{1.5748} \end{aligned} \quad (1)$$

其中 E_Y, E_{CB}, E_{CR} 分别表示亮度信号和两种色差信号, E_R, E_G, E_B 分别表示 R', G', B' 三个通道的模拟信号值。

两种编码方式的数字图像的量化方法是：

$$\begin{aligned}
D_{R,G,B} &= INT[(219E_{R,G,B} + 16) \cdot 2^{n-8}] \\
D_Y &= INT[(219E_Y + 16) \cdot 2^{n-8}] \\
D_{CB} &= INT[(224E_{CB} + 128) \cdot 2^{n-8}] \\
D_{CR} &= INT[(224E_{CR} + 128) \cdot 2^{n-8}]
\end{aligned} \tag{2}$$

其中 n 表示量化信号的比特长度，INT 作为将小数化整的四舍五入功能。因此可以得出 $R'G'B'$ 编码下的数字信号到 $Y'C_bC_r$ 编码下的数字信号之间的转换方法：

$$\begin{aligned}
D_Y &= INT[0.2126D_R + 0.7152D_G + 0.0722D_B] \\
D_{CB} &= INT\left[\left(-\frac{0.2126}{1.8556}D_R - \frac{0.7152}{1.8556}D_G + \frac{0.9278}{1.8556}D_B\right) \cdot \frac{224}{219} + 2^{n-1}\right] \\
D_{CR} &= INT\left[\left(\frac{0.7874}{1.5748}D_R - \frac{0.7152}{1.5748}D_G - \frac{0.0722}{1.5748}D_B\right) \cdot \frac{224}{219} + 2^{n-1}\right]
\end{aligned} \tag{3}$$

BT. 2020 色域

ITU-R BT. 2020 定义了 UHDTV 的各个方面，比如显示分辨率、帧速率、色度子采样、比特深度和色彩空间，它是由 ITU 在 2012 年发布的。相比于 BT. 709，BT. 2020 提供了更加广泛的颜色范围，BT. 2020 也是 HDR 技术的颜色空间。

BT. 2020 标准的重要性是毋庸置疑的，正如 BT. 709 对于高清视频传输与高清显示设备制造方面起到的引导性作用，BT. 2020 标准同样也深刻地影响着消费领域超高清显示设备的设计与制造，例如目前绝大部分的 4K 平板电视的物理分辨率都是采用 BT. 2020 标准的 3840×2160 ，但 BT. 2020 标准绝不仅仅在分辨率方面有所提升，在色彩和刷新频率和信号格式与分析等方面也进行了相关的规定。本文仅关注 BT. 2020 标准中关于颜色空间的定义。

相对于 BT. 709 标准，BT. 2020 标准大幅度提升了视频信号的性能规范，色彩深度方面提高至 10bit 或 12bit，很大程度上增强了整个影像的色彩层次，色域的广度也大大增强，两种标准的色域范围对比如图 3 所示。

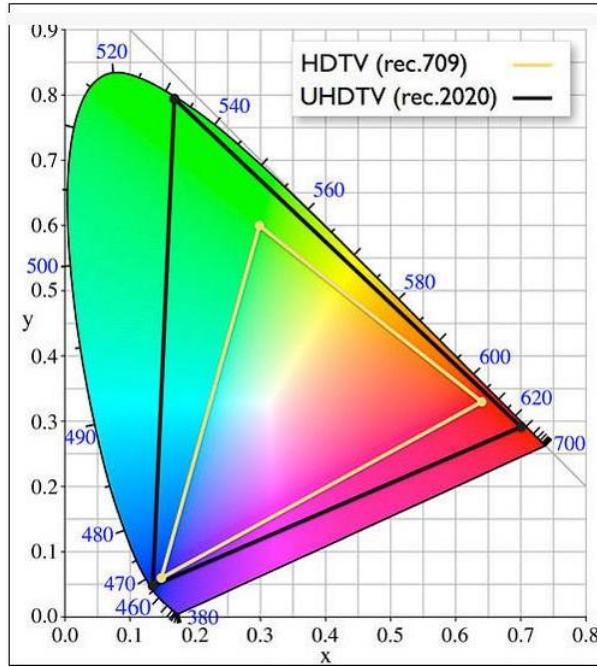


图 8-3: BT. 709 与 BT. 2020 对比

与 BT. 709 一致, BT. 2020 也支持了 $R'G'B'$ 和 $Y'C_bC_r$ 两种编码模式, 以及支持模拟与数字两种信号。同样的, BT. 2020 标准内部也存在两个重要的转换: 两种编码模式之间的转换以及模拟信号到数字信号的转换。

$R'G'B'$ 编码下的模拟信号到 $Y'C_bC_r$ 编码下的模拟信号之间的转换:

$$\begin{aligned} E_Y &= 0.2627E_R + 0.6780E_G + 0.0593E_B \\ E_{CB} &= \frac{-0.2627E_R - 0.6780E_G + 0.9407E_B}{1.8814} \\ E_{CR} &= \frac{0.7373E_R - 0.6780E_G - 0.0593E_B}{1.4746} \end{aligned} \quad (4)$$

$R'G'B'$ 编码下的数字信号到 $Y'C_bC_r$ 编码下的数字信号之间的转换:

$$\begin{aligned} D_Y &= INT[0.2627D_R + 0.6780D_G + 0.0593D_B] \\ D_{CB} &= INT\left[\left(-\frac{0.2627}{1.8814}D_R - \frac{0.6780}{1.8814}D_G + \frac{0.9407}{1.8814}D_B\right) \cdot \frac{224}{219} + 2^{n-1}\right] \\ D_{CR} &= INT\left[\left(\frac{0.7373}{1.4746}D_R - \frac{0.6780}{1.4746}D_G - \frac{0.0593}{1.4746}D_B\right) \cdot \frac{224}{219} + 2^{n-1}\right] \end{aligned} \quad (5)$$

BT. 2020 色域到 BT. 709 色域的转换

HDR 中的色域转换问题, 特别是色调映射之后的色域映射, 主要是 ITU-R BT. 2020 色域到 ITU-R BT. 709 色域的颜色转换。理想情况下转换应该满足以下要求: 1. BT. 709 色域内的颜色应该保持不变; 2. 转换方法适用于 BT. 2020 和 BT. 709 之间的多次转换; 3. 能够感知到的色调变化必须尽可能小; 4. 空间细节没有重大损失; 5. 不会在颜色上引入可见的不连续性; 6. 映射方法在数学上是可定义的。

需要注意的是，没有可以同时满足上述这些要求的通用色域映射方法。在从更广泛的色域转换到更小的色域时，BT. 709 色域外的颜色出现改变是不可避免的。不同的转换方式会在不同要求之间进行妥协，具体的做法可能因应用场景而异。下面介绍一种 ITU-R BT. 2407 [5] 标准中介绍的基于线性矩阵的从 BT. 2020 到 BT. 709 的简单变换。

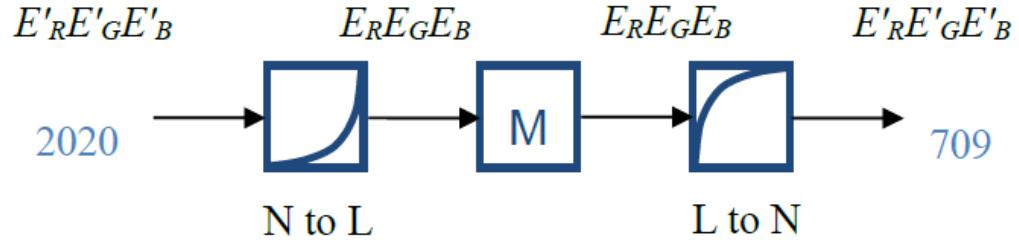


图 8-4 基于线性矩阵变换从 BT. 2020 到 BT. 709 的颜色转换

ITU-R BT. 2407 中的简单色域映射方式

图 4 是从 BT. 2020 到 BT. 709 的基于线性矩阵变换的颜色转换的框图。除了输出信号被严格限制之外，这是 ITU-R BT. 2087 [6] 标准中对从 BT. 709 转换到 BT. 2020 所规定的操作的逆操作。这种方法在条件限制的情况下是最直接简单的。

这种方法首先使用 ITU-R BT. 2087 中规定的两个非线性到线性的传递函数（一种是电光转换功能函数（EOTF），一种是逆光电转移函数（OETF）之一将归一化的非线性 RGB 信号转换为归一化的线性 RGB 信号。接着通过矩阵运算，将 BT. 2020 颜色空间的 RGB 信号转换到 BT. 709 颜色空间的 RGB 信号。最后再通过使用 ITU-R BT. 2087 建议书规定的两种线非线性传递函数（一种是逆 EOTF 方法，另一种是使用 OETF）之一将线性 RGB 信号转换为非线性 RGB 信号。从归一化线性 RGB 信号 (EREGB) 到归一化非线性 RGB 信号 (E' RE' GE' B) 的转换。

该方法具有非常理想的特性，即使在 BT. 2020 和 BT. 709 之间进行多次转换之后，它也不会改变 BT. 709 色域内的颜色。然而，BT. 709 色域之外的色彩是硬限幅的，即小于零或大于 1 的 RGB 信号 (EREGB) 会被限幅为零或 1，这会导致明显的色调偏移和空间细节的损失。虽然大多数内容看起来都很好，但色调和空间细节方面还是会有所损失。

ITU-R BT. 2407 中规定的色域映射算法原则

为了提升简单色域映射算法的结果，BT. 2407 提供了一套色域映射算法的设计流程与设计原则。图 5 显示了从 BT. 2020 色域到 BT. 709 色域的算法抽象框图。首先，选定一个映射颜色空间，然后在算法的第一步，将输入的符合 BT. 2020 标准的 RGB 表示的图像转换为映射颜色空间的形式。接着，使用色域映射算法，将映射颜色空间表示的颜色转换到 BT. 709 色域内。最后，再将颜色空间转换为 RGB 表示。到此，色域映射流程结束。

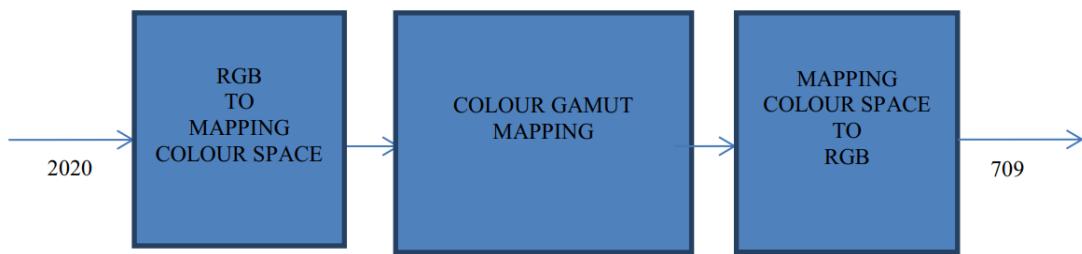


图 8-5 BT. 2020 色域到 BT. 709 色域转换算法流程图

其中，映射颜色空间的选取是算法的一个重要方面，可供选择的映射颜色空间包括：

1. CIE xyY
2. CIE u' v' Y
3. CIE L*a*b*
4. 简化版 Lab
5. 基于 CIECAM02 的 Uniform Color Space

BT. 2020 色域与 BT. 709 色域的 RGB 值(E_R, E_G, E_B)可以通过线性转换与投影，转换到到 CIE xyY 以及 CIE u' v' Y 的颜色空间中。CIE xyY 表示法中的 xy 平面代表传统的色度 (Chromaticity)，CIE u' v' Y 表示法中的 u' v' 平面也表示色度，但是当 Y 一定时，相对于 xy 平面来说，u' v' 表示法下的色度更加的一致 (Uniform)。

当使用 CIE xyY 以及 CIE u' v' Y 作为映射颜色空间时，可以通过线性变换来简化色域映射算法。但是，需要注意的是，简化后的某条映射路径（即从 BT. 2020 色域的某一个颜色映射到 BT. 709 色域的某一个颜色）是直线时，颜色的色度，色相和明度会同时出现改变。

CIE L*a*b* 颜色空间相对于 CIE xyY 以及 CIE u' v' Y 颜色空间说，提供了更加优秀的视觉统一性。而简化版 Lab 是 CIE L*a*b* 的简化版。但是将 BT. 2020 与 BT. 709 的 RGB 值转换到这两种表示法时，相对于 CIE xyY 以及 CIE u' v' Y 表示法来说，需要使用更加复杂与非线性的操作。这两种颜色空间的优势在于由于将明度纳入了表示法的设计当中，因此他们在感知上提供了更好的一致性。当使用这两种表示法时，线性映射路径能够保留色度，色相以及明度中的一个或两个属性。通过这些映射路径，BT. 2020 的颜色就能够映射为 BT. 709 的颜色。这种颜色的修改过程会使用到线性或非线性的方程。

在从 BT. 2020 到 BT. 709 的转换过程中，一些色域映射算法将 BT. 709 色域内的颜色完全保留，不加改变，所做的颜色映射只针对那些在 BT. 709 之外，BT. 2020 之内的颜色。但是另外一下算法对 BT. 709 内部的颜色也作了一些改变，特别是对处于 BT. 709 色域边缘的颜色。而当颜色在 BT. 709 色域与 BT. 2020 色域之间不断循环转换时，这种对 BT. 709 色域内部颜色也做出改变的算法就会产生非常严重的人工痕迹。

从上述内容可以发现，BT. 2407 标准提供了一种简单的色域映射的算法，同时，也给出了一套设计流程。不同的设计者可以通过选择不同的色域表示法，设计不同的色域映射算法，来实现 BT. 2020 色域与 BT. 709 色域之间的转换。

ITU-R BT. 2407 中的其他色域映射方法

BT. 2407 中提供了 6 种较新的，效果更好的色域映射方法，本文中只给出这些色域映射方法的算法流程图或算法思路以供参考。算法的具体实现方法以及具体细节，可以参见 [5]。

算法 1

这种算法在三个步骤中，分别对源 HDR 图像的色相，亮度与色度进行映射。算法框图如下：



图 8-6 算法 1 框图

算法 2

这种算法对图像的色相、亮度与色度同时进行映射。同时，在使用算法时，需要将图像使用 CIECAM02 表示法进行表示。

算法 3

算法 3 可以看做算法 1 的简化版本。它的算法流程图与图 6 中算法 1 的流程图一致，两者的区别是算法 3 在 CIE LAB 颜色空间中进行，同时利用简单的线性化操作来简化计算量，从而实现运算效率的加强。

算法 4

算法 4 提供了一种自适应的色域映射算法，这种算法的目的是尽可能保留原图的整体感知。具体流程图见图 4.

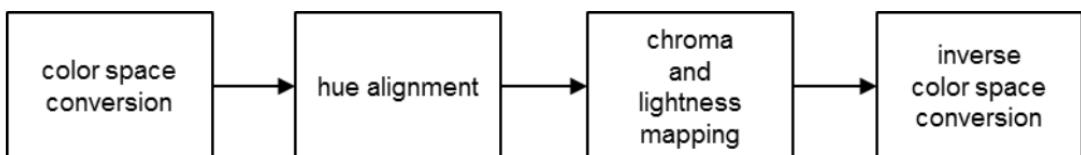


图 8-7 算法 4 流程图

算法 5

算法 5 是一种保留亮度信息的色域转换。通过改变图像的颜色空间，将图像的色度信息与亮度信息分开，从而仅仅处理图像的色度信息

算法 6

算法 6 的设计思路是在考虑显示设备色域的基础上，实现 BT.2020 空间中的颜色到 BT.709 空间中的颜色的比色法转换

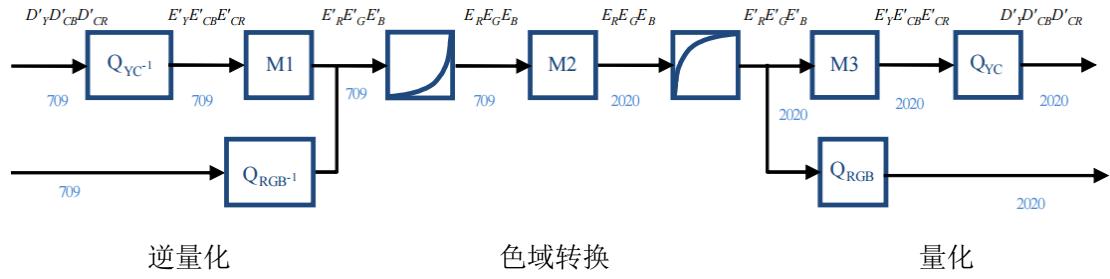
BT. 709 色域到 ITU-R BT. 2020 色域的转换

HDR 色域转换的另外一个角度就是从 BT. 709 色域到 BT. 2020 色域的转换。在将 SDR 内容适配到 HDR 显示设备上播放的过程中，通常需要将 BT. 709 色域中的颜色转换到 BT. 2020 色域中。

BT. 709 到 BT. 2020 的转换有一个重要的限制，即 BT. 709 标准下的颜色，在 BT. 2020 标准下要是同样的颜色。换句话说，如果这个颜色是两种色域中都存在的，那么这个颜色在转换过程中应当保持不变；而如果这个颜色是 BT. 2020 中独有的，这就需要进行颜色转换。BT. 709 到 BT. 2020 的转换可以大致分为三个主要步骤：第一个步骤是逆量化，得到 BT. 709 标准下的模拟信号值 E_R, E_G, E_B 。第二个是色域转换，将 BT. 709 下的模拟信号值，转换为 BT. 2020 标准下的模拟信号值。最后一个步骤是量化，得到 BT. 2020 标准下的数字信号。选择在模拟信号域进行色域转换而不是直接在数字信号域进行色域转换的原因是为了保留更多的信息与细节。由于量化过程会存在信息的丢弃，并且 BT. 709 的色域比 BT. 2020 小，如果直接用数字信号进行转换，会使得 BT. 2020 色域下的图像效果变差。因此，通过逆量化过程，恢复图像的相关细节，然后进行模拟信号的色域转换，就可以解决这个问题。

BT. 709 色域到 BT. 2020 色域的转换标准遵循 BT. 2087 标准 [6]。标准中根据信号源的亮度是否恒定，将转换方式分为以下两种：

2. 非恒定亮度信号的转换



逆量化 色域转换 量化

图 8-8. BT. 709 到 BT. 2020 在非恒定亮度信号格式下的转换过程

上图是非恒定亮度信号的转换流程图。由于 BT. 709 与 BT. 2020 均支持两种编码方式： $R'G'B'$ 和 $Y'C_bC_r$ ，因此逆量化过程的输入与量化过程的输出均有两条路径。根据需求的不同，可以选择不同的输入输出模式。

每一个区块的具体实现过程以及相关函数定义，可以参考 [6]。

2. 恒定亮度信号转换

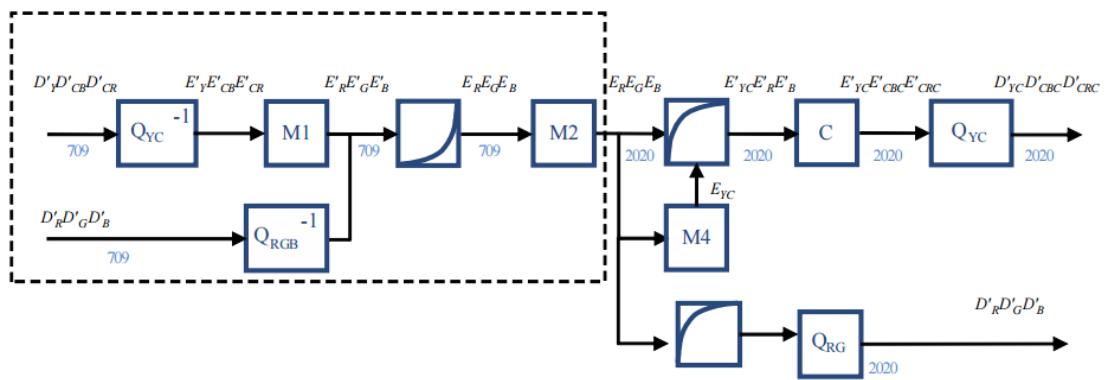


图 8-9. BT. 709 到 BT. 2020 在恒定亮度信号格式下的转换过程

上图是恒定亮度信号的转换流程图。由于 BT. 709 与 BT. 2020 均支持两种编码方式： $R'G'B'$ 和 $Y'C_bC_r$ ，因此逆量化过程的输入与量化过程的输出均有两条路径。根据需求的不同，可以选择不同的输入输出模式。

每一个区块的具体实现过程以及相关函数定义，可以参考 [6]。

8.3 色域映射算法

由于 BT.2407 标准只是提供了标准，指导思想与设计原则，因此有部分色域映射算法的映射效果要优于标准中推荐的算法。下面，将介绍一种近年来新提出的，并且效果优秀的色域映射算法。

8.3.1 HDR 颜色精确重构色域映射框架

[14]这种色域映射框架的优势在于，它能够将 HDR 图像的颜色精确地映射到 SDR 图像的颜色范围内。同时，它的运算量与其他色域映射算法相比要小，也可以很好地与现有的色调映射（Tone mapping）算法兼容。这种算法选择的映射颜色空间是 CIE LCh 表示法。将颜色转换为了 CIE LCh 表示法后，将色度通道进行压缩，使得色域能够与输出色域一致。这样的好处是避免了色相以及明度的移动，从而避免了一些严重的人工痕迹。

这种算法的流程图如图 10 所示。算法的输入是 XYZ 表示法下的 HDR 图像 I 。首先，图像的亮度通道 $I(L)$ 被已有的色调映射算法压缩。色度通道 $I(C)$ 通过下文将要介绍的色度压缩算法进行校正，以消除色调映射过程中产生的不需要的饱和度。最后，图像的色度与亮度信息 $I(C,L)$ 通过色域截断步骤进行处理，以保证每个像素点的像素值都在目标色域之内，记做 $G(C,L)$ ，以减小图像的形态改变。下面将对其中的每一步都进行详细的介绍。

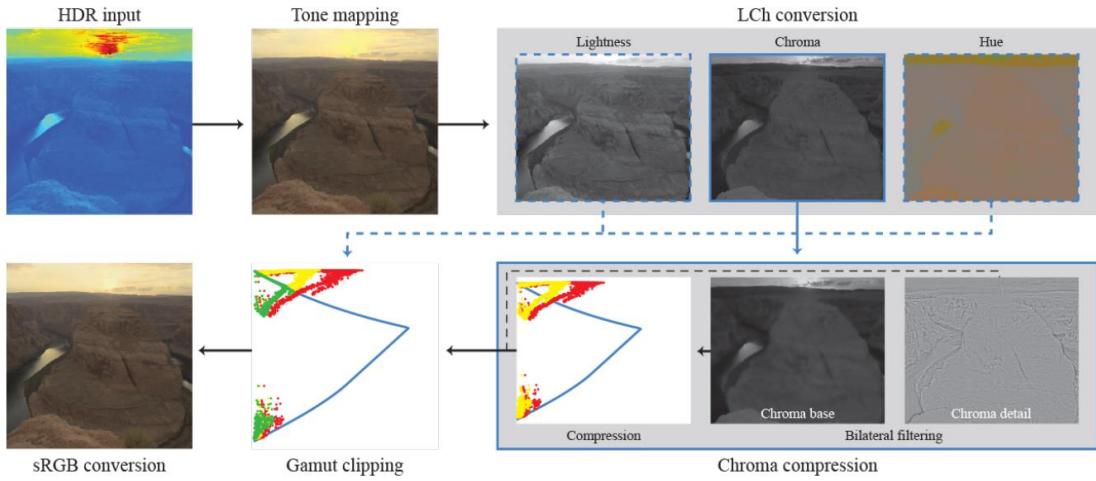


图 8-10 算法流程图

亮度压缩

由于算法希望通过色域压缩以及随后的色域截断步骤校正色调映射过程中产生的所有失真，因此，算法的第一步，就是利用色调映射，对亮度进行压缩。任何现存的色调映射算法，理论上都可以直接集成到这个色域映射的框架中，作为第一步亮度压缩中使用的色调映射算法。这里以 Photographic TMO [7]为例，介绍亮度压缩的整个流程与步骤。

1. 输入的 HDR 图像被转换到 TMO 所需的表示法下。(本例中，转换到 CIE xyY)
2. 利用色调曲线，对通道 Y 进行映射，得到压缩后的亮度信道 Y_c 。这里并不限制色调映射算法的类别，全局算法与局部算法都可以在此处应用。
3. 压缩后的亮度重新添加到图像当中，并且转换为 XYZ 表示法。
4. 最后，色调映射后的图像 I 被正则化，使得 Y 的最大值设置为 100，以供后续处理。在 Photographic TMO 的例子中，由于输出的亮度值在 0-1 之间，因此需要进行比例放大，将亮度值进行改变，以供后续处理。

这种亮度压缩的流程，提供了非常高的色调映射的灵活性，使用者可以使用任何想要使用的色调映射算法，但是，这种灵活性的代价是后续需要进行大量的计算来校正色调映射算法带来的色域改变，因为很多色调映射算法在设计的时候并没有考虑色域的问题，仅仅考虑了亮度的压缩。

色域边界计算

由于这种算法的设计目的是避免不需要的色度与色相偏移，并且避免不可控的色域截断（即使被截断的色域在目标色域之外），因此，想要实现这些目标，就需要将这些

颜色参数分开。因为如果这些颜色参数或者称之为颜色信息混合在一起的话，调整某一参数时，另一个参数也会被同时改变，这不利于算法的设计与优化。因此，该算法将图像的颜色使用 CIE $L^*C^*h^*$ 表示，也就是 CIE $L^*a^*b^*$ 的圆柱形表示法。这两种表示法在传统的色域映射算法 [1] 中也经常被使用。

为了确定正确的压缩量以及评估一个给定的像素点是否处在输出色域中，算法需要首先确定色域的边界。在这种算法中，目标色域被设置为 sRGB，在转换到 Lab 时使用 D65 白点。同时，目标色域与白点也可以根据需求设置为其他值。输入色域则是输入的 HDR 图像所包括的颜色范围。计算输入与输出色域的边界的方法基于 [1] 提供的获取 sRGB 色域在 LCh 表示法下的色域边界。图 11 显示了输入色域与输出色域的极端差异。在明度与色度上，两个色域的极端差异都在 10 倍以上。

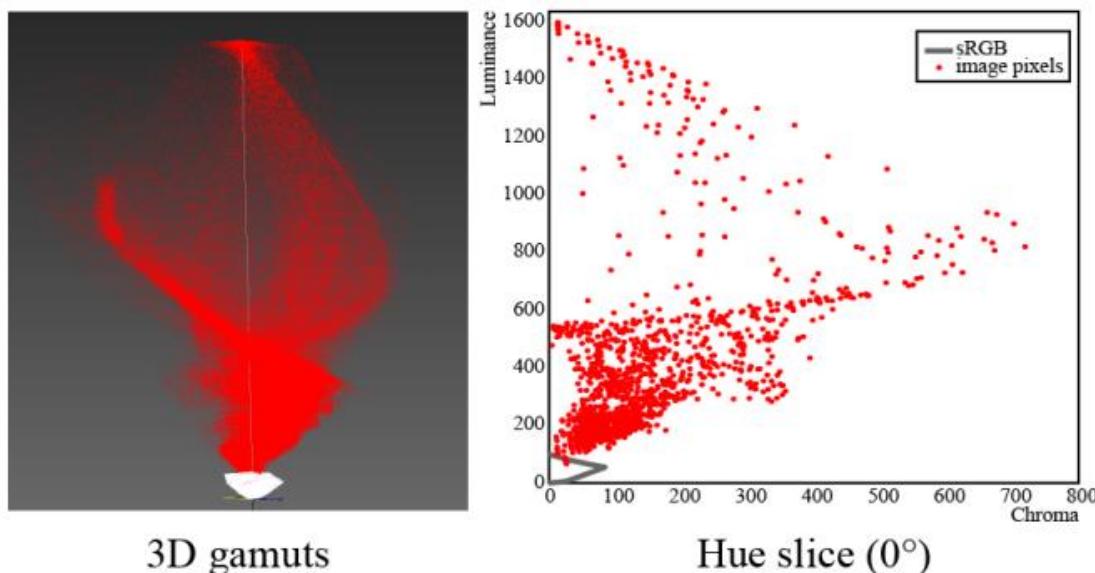


图 8-11 输入与输出色域的极端差异

色度压缩

对于 HDR 图像来说，图像中的大量像素点的色度值都在目标色域之外。如果对色域的压缩进行不恰当处理的话，会造成这些像素点的色相发生不好的偏移。并且，亮度域的色调压缩会造成图像的饱和度增大。为了校正这些现象，有两种算法可以被用来压缩色度 $I(C)$ 。

色相特性方法

这种方法将色度值 $I(C)$ 通过两步进行压缩。第一步是为每个色相 $h \in \{1^\circ \dots 360^\circ\}$ 确定一个比例因子 R_h ，这样就可以构成一个向量 \mathbf{R} 。为了实现这一目标，算法通过按比例放缩色域边界 $G_h(C, L)$ ，直到这个色域边界能够完全地包括住所有这一色相 h 下的像素点。

接下来正式地定义这种算法：对于一个色相，定义它的初始色相比例 $R_{h,0} = 1$ 。如果存在这一色相下的点在色域范围之外，在每个迭代步骤 i 处，增加比例 $R_{h,i}$ 并且按比例放缩色域边界：

$$R_{h,i} = R_{h,i-1} + d$$

$$G_{h,i}(C, L) = \begin{bmatrix} R_{h,i} & 0 \\ 0 & R_{h,i} \end{bmatrix} G_h(C, L) \quad (6)$$

这里的增量 d 设置为一个较小的量即可（通常设置为 0.1）。

尽管使用这里的比例系数可以直接对源图像进行处理，从而将源图像的色度压缩到目标色域之中，但是在实际操作过程中，如果利用源图像的色域整体来得到 \mathbf{R} ，然后进行压缩，这时会产生一种非常差的效果，因为源图像的色域非常大，因此如果 \mathbf{R} 来自于源图像的整体，由于一些及其明亮的像素点的存在，整幅图像的平均色度会被压缩得过低。因此，在实际应用当中，通常不会利用源图像的整个色域来得到 \mathbf{R} ，而是取其中的一部分色域。

在实际操作过程中，选取哪部分色域，以及选取多大的色域，是一个重要的考量。色域选取的比例取决于图像内容分布的。如果整幅图像中都存在需要被压缩的像素点（即存在色度超出目标色域的像素点），那么选取的比例就可以很大。反之，如果只有一小片区域的像素需要被压缩，那么就选取很小的比例，从而避免图像的效果变差。在实际应用过程中，判断需要被压缩的像素是否分布与全图的方式是计算出需要被压缩的像素构成的连通分支的数量，然后再计算出其中包含的像素的数量，两者的比值就是判断的标准。通常如果这个比值小于 0.01，那么选取的比例就不宜过大。

为了保持图像细节的良好，而不会由于很多边缘被平滑操作而产生细节丢失，图像的色度通道 $I(C)$ 需要首先通过一个双边滤波器 ($\sigma_s = 0.2 \max(I_{width}, I_{height})$, $\sigma_r = 0.05 \max(I(C))$) 来获取 base layer, $I(C)_{base}$ ，因此实际上用于 (6) 中的 $I(C)$ 实际上是 $I(C)_{base}$ 。而获取 detail layer 的方式可以是减法，也可以是除法，两者的效果基本一致。算法中实际应用的，是通过除法方式来获取 detail layer: $I(C)_{detail} = I(C)/I(C)_{base}$ 。

尽管上述的方法已经能够修正大多数的失真，如果直接使用不同色相下的 R_h ，将会在最后的图像中间歇地产生不连续，原因是相邻的色相之间，可能存在一些细微的改变，而这种改变将会影响最后的结果。因此，在最后的比例放缩之前，需要对 \mathbf{R} 进行平滑操作，从而得到 \mathbf{R}' 。这里使用的平滑函数是 lbox (averaging box) 函数，因为这种平滑函数在性能优秀并且运算效率较高。

最终，在每个输出图像的每个色相 h 进行比例缩放：

$$I(C)'_{base,h} = \frac{1}{R'} I(C)_{base,h} \quad (7)$$

接着，再使用 detail layer 来重构最终的色域压缩后的图像 $I(C)' = I(C)'_{base} \times I(C)_{detail}$ ，其中， \times 表示元素相乘。

全局方法

前面描述的色相特性方法可以很好地压缩色度，并且可以最大化地利用色域空间。而这种方法的代价是计算复杂度很高。同时也可发现，色度通道的动态范围与显示色域比较起来实际上并不是很大（这里的动态范围并不是指 HDR 的动态范围，只是表示色度通道的最大值与最小值的比值）。因此，一个线性的压缩算法可能也能提供不错的效果。因此，在这种前提下，有另一种更简单，低计算复杂度的方法，也可以投入使用。前面提到的色相特性方法，可以用于色调映射的后续处理，因为在这种情况下，精确度是首要的考量；而在其他一些对精确性要求不高，而计算能力有限的情况下，例如显示设备的色域映射，就可以使用这里提到的线性映射方法。

在全局方法中，仍然将像素按照 0-359 的色相角进行分类，对每一类进行压缩。但是与前面方法不同的是，色相特性方法针对每个不同的色相角使用不同的压缩系数，而在全局方法中，为了减少计算量，使用所有色相角中最小的系数，对所有的色相角进行压缩。

类似于色相特性方法，这里的全局压缩方法也只对 base layer 进行压缩：

$$I(C)'_{base,h} = I(C)_{base,h} \min_{h \in [0^\circ, 359^\circ]} \left[\frac{Cusp_{d,h}}{Cusp_{s,h}} \right] \quad (8)$$

这种方法只在 $\min[Cusp_{d,h}/Cusp_{s,h}] < 1$ 时，对色度进行压缩。

此时，全局算法与色相特性算法一样，也会由于运用这个源色域而导致色度压缩的时候产生过度压缩，进而影响图像质量。为了避免这种影响，全局算法中也使用了相同的利用部分源色域的方式来避免过度压缩。但是，全局算法还需要额外进行一步操作，以将部分仍然处在色域边界之外的像素点转换到目标色域内。这种方法将在下一部分进行介绍。

色域截断 (Gamut Clipping)

由于算法到目前为止，色度与亮度的运算与处理都是分开的，因此不能保证将两者合并起来之后，颜色不会处在色域边界之外。这种现象可以见图 12。在色调映射过程中，图像的亮度通道的值会被压缩。尽管色调映射后的图像的最大亮度值会被保证为 100nit，但是可能仍然处在色域边界之外（红色点）。利用色度压缩能够有效地减轻这个问题（黄色点），使得像素点向色域边界移动，但是这并没能完全解决这个问题。为了将所有像素点变换到色域边界之内，同时不影响已经处在色域边界中的像素点，算法在最后一步，使用了色域截断处理。由于像素点的色度与亮度可能都在色域边界之外，色域截断处理需要在色度与亮度两个通道上进行，因此在实际处理过程中，需要考虑两方面的因素，并且可能两者的优化不可兼得。

尽管现在有很多算法可以实现色域截断，并且能够给每个色域边界之外的像素点一条移动路径，由此可以将这个像素点移动到色域边界内，但是这些算法的应用场景都是只考虑色度一个方面，而不考虑亮度。在本算法的特殊应用场景里，色域边界之外的点都是高亮度与高色度的点。

由于高亮区域的点难以处理，因此在调整色度与亮度时，需要在两者之间保持微妙的平衡以保证结果图像的外观不会发生很大的改变。这里使用的方法是，在同一个色相角的系列像素中，对于一个像素点 $p \in I'$ ，在色域边界上可以确定一个点 P_{clipC} 使得 $P_{clipC}(C) \in G(C)$, $P_{clipC}(L) \in p(L)$ 。类似的，再确定一个值 P_{clipL} , $P_{clipL}(C)$ 的值不变，并且 $p(L)_{clipL}$ 移动到色域边界上。计算出这两个点后，利用线性插入将色域外的像素映射到目标色域的色域边界上。

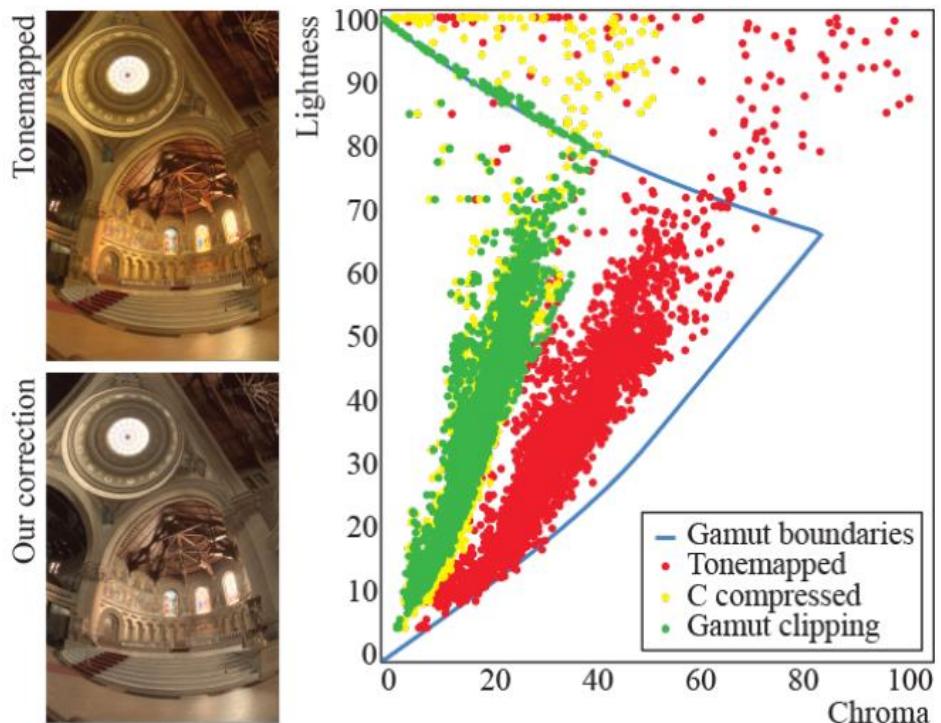


图 8-12 色调映射后，很多点实际上都落在了色域边界之外（如图中红色点所示），接着，随着色域压缩算法的使用，像素点不断地逼近色域边界，但是仍然有一部分像素点处在色域边界之外

第九章-HDR 质量评价技术

本章聚焦 HDR 质量评价技术，对于编解码、色调映射以及逆色调映射等不同任务，通常会采取不同的评价方法。本章将从主观评价和客观评价两个角度对常用的 HDR 视觉质量评价技术做整体介绍。

9.1 HDR 主观评价方法

国际电信联盟组织 ITU 根据不同使用场景给出不同的 HDR 主观视频质量测试方案。本文主要参考 ITU-R Rec. BT.500[1]和 Rec. ITU-R BT.1788[2]建议书中的测试方法，列出几种具有代表性的、在实际研究应用过程中使用频率较高的测试方法：单刺激法（ITU-R Rec. BT.500）、双刺激连续质量标度方法（ITU-R Rec. BT.500）、SAMVIQ 方法（ITU-R Rec. BT.1788）；并对不同的 EOTF 曲线的评估质量和 CSM 测试模式做了介绍。

9.1.1 单刺激法 SSM

在单刺激法（Single Stimulus Methods）中，测试素材只包含测试序列，不包含相应的原始序列。测试前，会使用模拟演示序列为每位测试者讲解测试流程。测试时，对每一位测试者采用不同的随机顺序。测试者则对观看的每一个视频分别打分。测试流程如下图 1 所示。

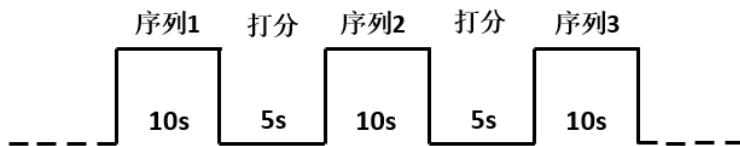


图 9-1 单刺激法测试流程

单刺激法通常采用五级质量量表，如下表 1 所示。该量表中分值为 MOS（Mean Opinion Score）分值，表示是对“视频质量”进行打分，即，认为该序列质量越高，则分数越高。有时，该方法也会采用九级或十一级质量量表，这两种情况时，即为最高分值为 9 分或 11 分。

表 1 单刺激法质量和损伤量表

MOS 分值	描述	损伤
5	优	不可察觉
4	良	可察觉，但不讨厌
3	中	稍微讨厌
2	差	讨厌
1	劣	很讨厌

该种测试方法的优点是首先较为符合测试者的实际日常使用场景，即，通常人们在看视频时，并不会事先观看无损源参考视频。其次，这种方法最大化了测试者每分钟可观看视频数量，节约了测试成本。最后，在严谨地进行测试设计和指导演示的情况下，即使是由不同批次的测试者进行测试，它依然具有稳定的可重复性。单刺激法的缺点也源自于它不与无损源参考视频比较这一优点，当需要检测某些特定细微差别如颜色变化的视频时则无法用此种方法进行测试。

9.1.2 双刺激连续质量标度法 DSCQS

在双刺激连续质量标度方法（Double Stimulus Continuous Quality Scale）中，需要对每个测试图像的两种状态 A&B 进行评分。其中一个是来自信号源的图像，即基准图像 A；另一个可能是经过被测系统输出的图像，即被测图像 B。基准和被测图像交替显示两次或多次之后（通常是两次）进行评分。

不同测试图像的一连串显示评分过程中，基准和被测图像呈现的先后次序以伪随机方式变动(观看员事先并不知道哪一个是基准图像)，要求观看员只简单地对每对图像的总体质量进行评分，并在评分表上作出标记。测试流程见下图 2。

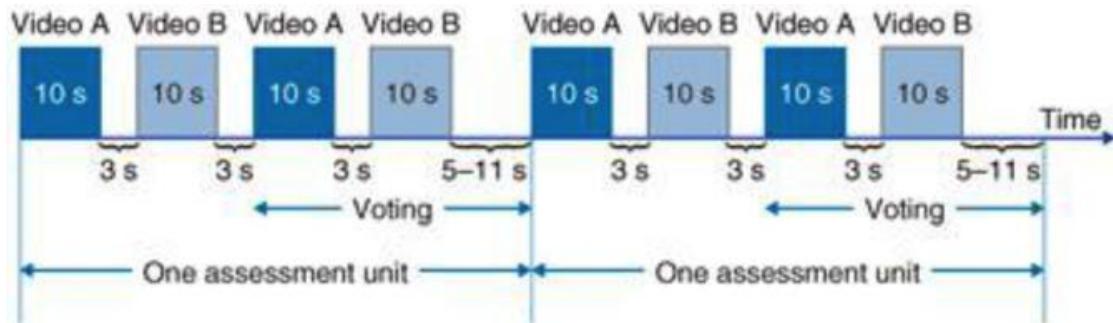


图 9-2 双刺激连续质量标度法测试流程

评分表由若干对纵向标度线组成，以适应对每个测试图像两种状态的评分。与单刺激法为避免量化误差，标度线提供连续标度，且被分成 5 个等级，相当于标准的 5 级质量标度范围。一个典型的 DSCQS 评分量表见下图 3。

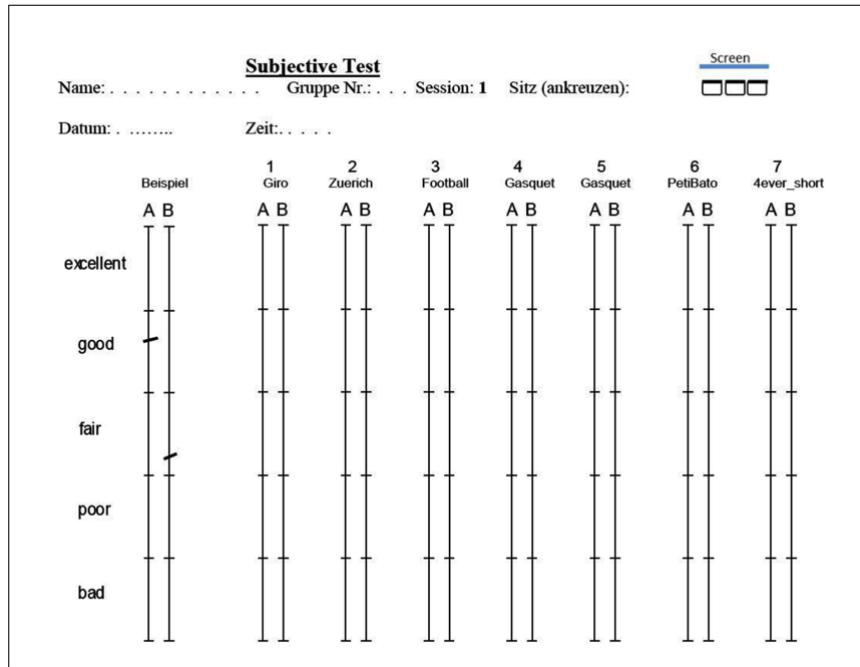


图 9-3 双刺激连续质量标度法测试样表

9.1.3 SAMVIQ 方法

在 SAMVIQ (Subjective Assessment Method for Video Image Quality) 方法中，观测者准许观看一个片段的若干个版本。当所有版本都经观测者评定后，可对之后的片段内容进行评估。不同版本可由观测者通过计算机图形接口随机选择。根据需要，观测者可以停止、评审并修改某个片段各个版本的评分，也可以反复观看。

SAMVIQ 质量评估方法使用标度，以提供对视频片段质量的精准测量。各个观测者在从 0 到 100 评分的连续标尺上移动滑条，该连续标尺有 5 个线性排列的质量描述(优、良、中、差、劣)。其用户界面见下图 4。

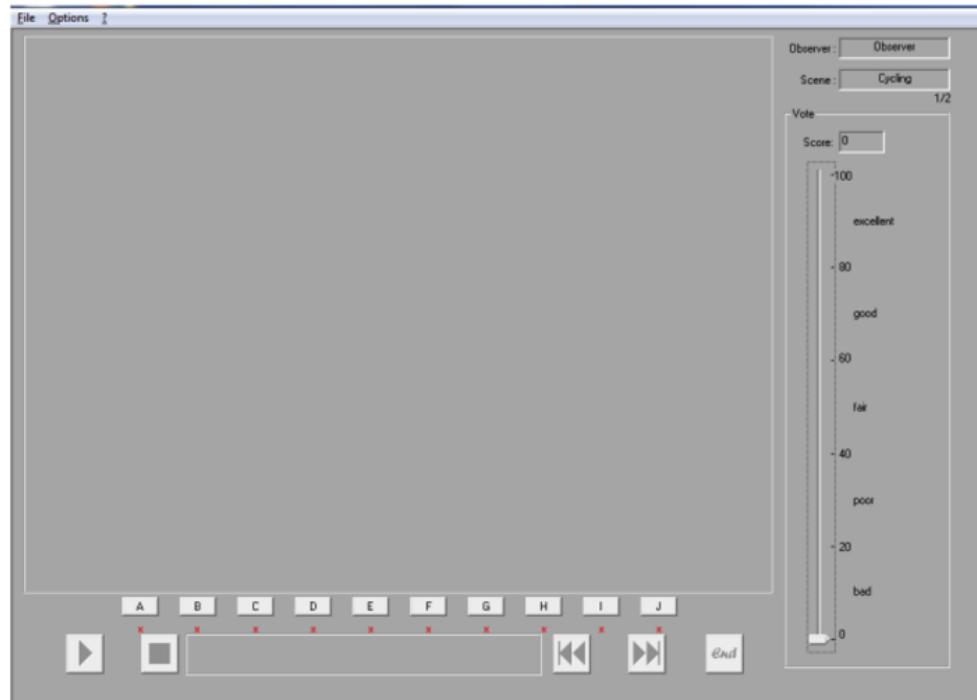


图 9-4 SAMVIQ 方法程序用户界面

SAMVIQ 包括一个显性基准（即未经处理的）原始片段，其评分通常为 100；以及相同片段的若干个版本，这些版本包括经处理的和未经处理的（即隐含基准）片段。

经测试表明，可以使用显性基准来最大限度地缩小分值的标准差，尤其是对多媒体数字信号编解码器性能的评估。为了评估基准的内在质量，也可加上隐含基准分。值得注意的是，当没有可用的基准时，测试仍有可能进行，但标准的偏差会显著增大。

SAMVIQ 测试流程

- a) 逐个视频地进行测试。
- b) 对当前场景，可能以任何次序来播放任何片段，并为其打分。每个片段都可以多次播放和打分。
- c) 从一个场景到另一个场景，对片段的访问是随机的，防止观测者试图根据已排好的次序、以完全相同的方式做出判定。实际上，在一个测试中，算法的次序仍保持相同，以便简化对结果的分析和陈述。只有来自相同按钮的相应访问是随机的。
- d) 对第一次观测，当前的片段必须在打分之前全部播放过；否则，可能立即打分和停止。
- e) 为测试下一个场景，必须为当前场景的所有片段打分。
- f) 为完成测试，必须为所有场景的所有片段打分。

典型的测试方式见下图 5。

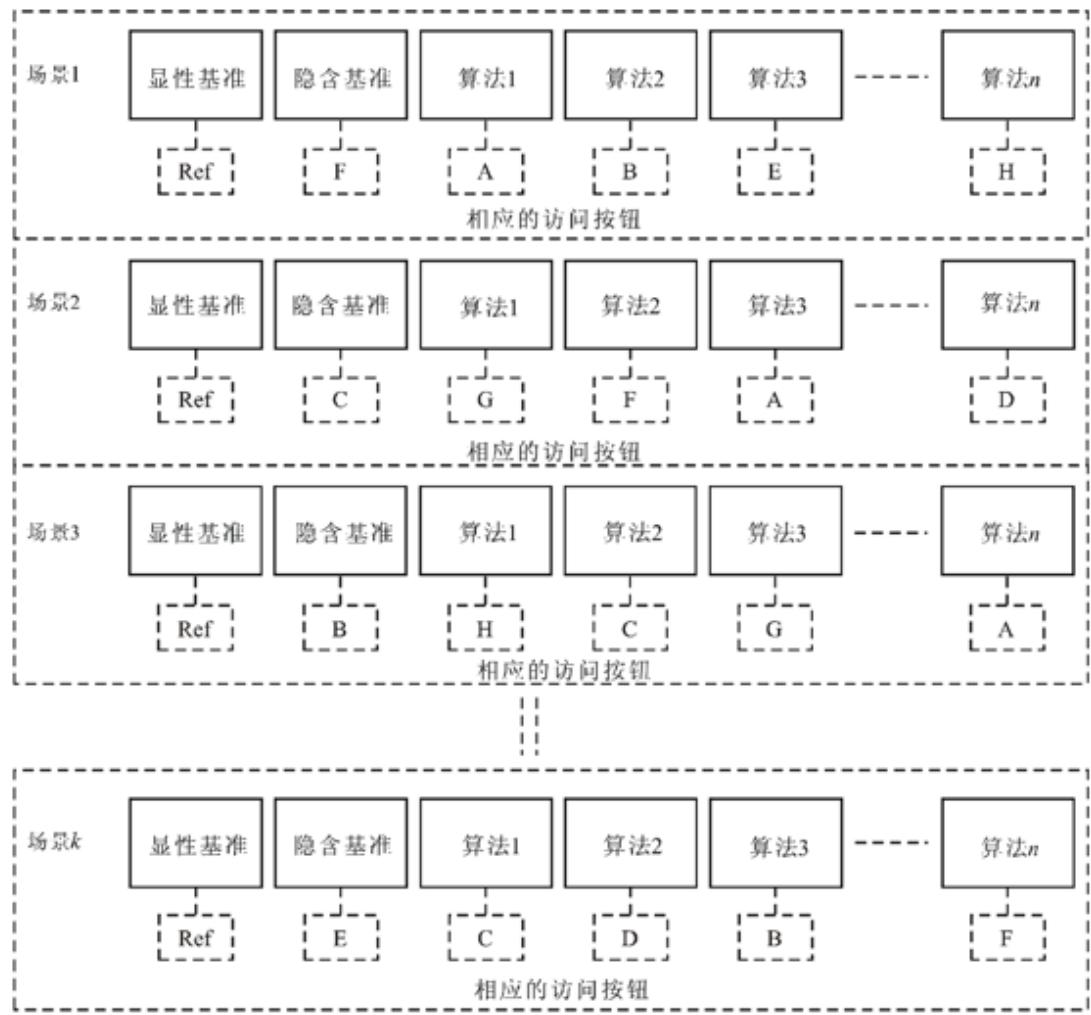


图 9-5 SAMVIQ 方法测试流程[2]

SAMVIQ 方法适用于多媒体内容，原因是它可能结合图像处理的不同特点，例如多媒体数字信号编解码器类型、图像格式、比特率、图像缩放等。

9.1.4 对于不同 EOTF 的主观评测

HDR 技术中常用的光电转换函数曲线有 PQ 曲线和 HLG (Hybrid Log Gamma 曲线) 对于同样条件下拍摄和编码解码的 HDR 图像就产生了差别。

EBU 报告中采取 DSCQS 方法和 SAMVIQ 方法对于两种方案的质量进行了评测，其采用了不同比特率、不同编码方式的 HDR 视频进行测试，其结果见下图 6、图 7。

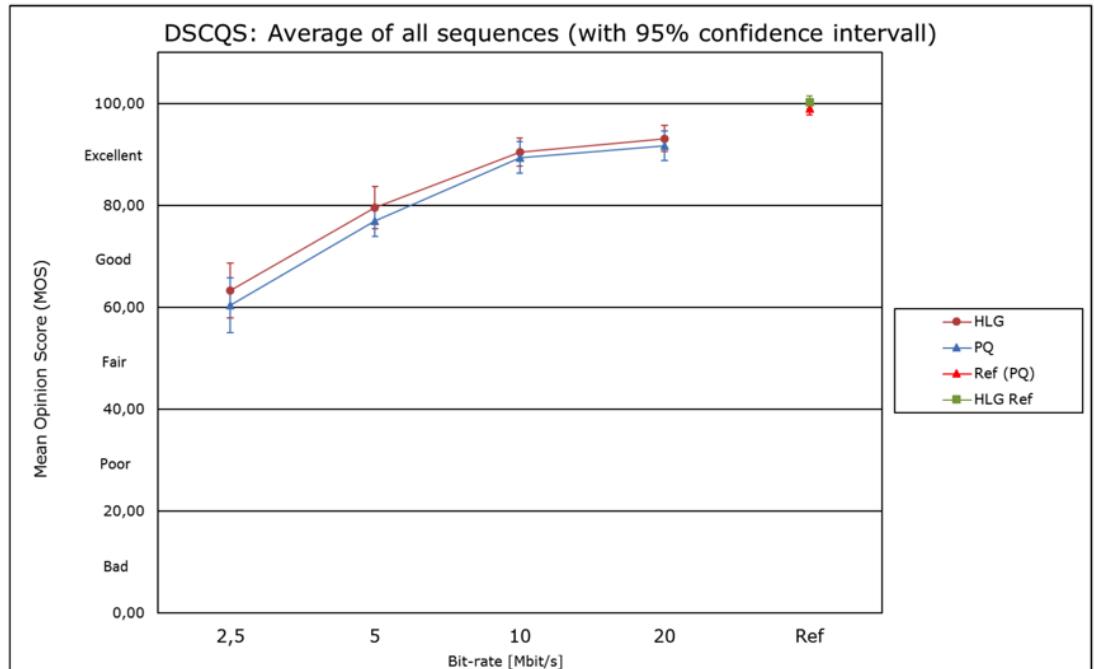


图 9-6 DSCQS 方法测试 PQ&HLG

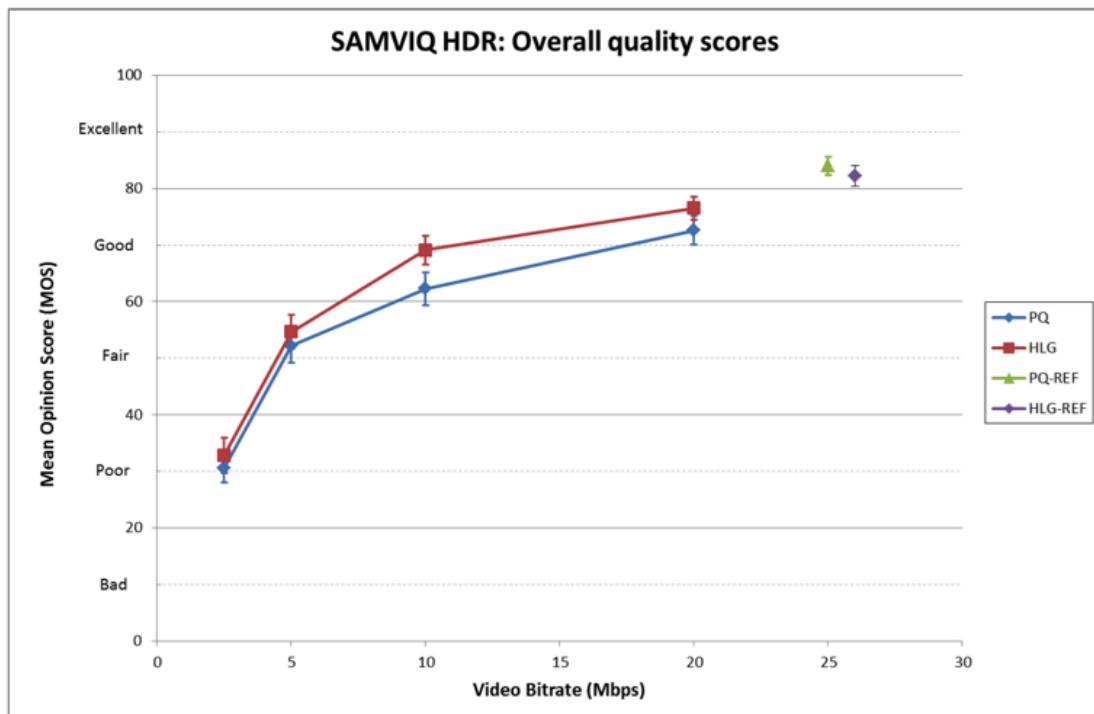


图 9-7 SAMVIQ 方法测试 PQ & HLG

由此可见，DSCQS 下的两种方法的主观评测质量几近相同，而 SAMVIQ 下 HLG 整体稍高于 PQ。两种方法在高码率下显示出高的 MOS 分。

9.1.5 Color Space and Monitor (CSM) 测试模式

HDR 图像相较于 SDR 具有更高的亮度和色度，因此对比 SDR 只需要对于屏幕图像做直接的观察，HDR 的质量评价需要对颜色空间、显示器、亮度标准做更多的模式设计。 SMPTE 设立的 CSM 测试模式[3]包含了十几种针对性的测试模式。下面将介绍常用的两种模式。

多空间：

通过展示八个颜色空间的同一特征图像，通过呈现效果可以辨别出其来源于哪种颜色空间，如下图即属于 BT.709；并且可以做到颜色空间的转换，如下图将 PQ2020 转换到 BT.709 并在 BT.709 的显示器上播放，则可以正确显示。

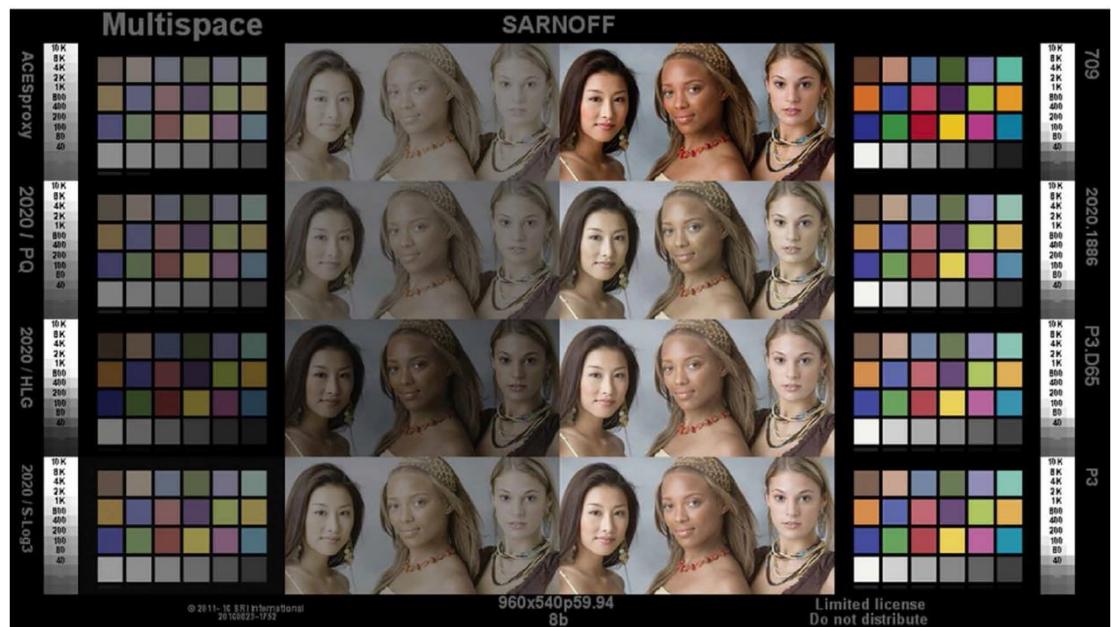


图 9-8 多空间模式[3]

三色标：

不同的颜色空间又不同的色标，因此可以通过把常用的 PQ2020、PQP3D65、BT.709 的色标组合到同一颜色容器（PQ2020）中，就可以直接用矢量转换做颜色空间的转换。见下图 9。



图 9-9 颜色空间的转换[3]

其他的模式简单介绍见下图 10。

TABLE 1. Some of the patterns included in the Color Space and Monitor (CSM) Suite.

	Multispace: Faces, Macbeth Chart and Stair step rendered in eight color spaces in one image. Default spaces: ACESproxy, PQ/2020, HLG/2020, S-Log3, 709, 2020/1886, P3/D65, and P3. Custom spaces available.
	Jacob's Ladder: 103 calibrated steps covering 8 decades of brightness as 13 steps per decade. Shows tone mapping and clipping. Each decade strip contains a rising step to help verify the darkest visible level.
	Color Ramps: 28 rotating gradients in 7 colors covering 0-15 nits, 15-60, 60-240, and 240-960 nits. Gray background contains noise to challenge encoders. Reveals quantization banding issues of monitors and systems.
	Code Values (Dark): 10-bit labeled neutral patches covering code values 0 to 127. Allows determination of "Full" or "Limited/SMPTE" range interpretation of code values. Clearly shows groups of four patches when truncated to 8-bits.
	Dark Moons: (contrast enhanced here for printing). Rotating shallow gradients each starting at 0 nits and rising to a specified number of stops below 100 nits. Reveals how well monitors come out of black.
	Dark Chips: 10 blinking chips on a black background. Indicates the smallest step from black that can be seen on a given monitor in a given viewing environment.
	Color Chips: ST 303M "Macbeth" chip chart. Also includes Triple Gamut strip of saturated colors in three color spaces, and luma sweep. Use to verify correctness of color space conversion or interpretation.
	Nit Chips: A set of a dozen small-area brightness references. Useful for quick characterization of monitor brightness and tone mapping.
	Zone Plate: Two-dimensional frequency sweep of luma values. Reveals image scaling and resolution conversions.
	Pixel Strips: alternating rows of black and white lines (black/white/black/white) and colored lines (red/green/red/green) reveal whether a monitor's color resolution is as good as its black and white resolution. Also checks gamma matching and Y'CC matrix matching.
	Triple Color Bars: Bars from 2020, P3, and 709 in a 2020 container. Allows verification of color space conversion using a vectorscope.
	Triple Gamut: 18 color samples around the edge of each of three gamuts. With on-screen annotations of xyY values for each color. Scaled to 100 nits to avoid blown out colors.
	Inter-Gamut: Top row is colors around the 2020 gamut and bottom row is colors around the 709 gamut. Intermediate colors are evenly spaced (in x, y) between color pairs.
	Full Gamut: Top row is colors around the 2020 gamut and bottom row is neutral D65. Intermediate colors are evenly spaced in x, y between each 2020 color and D65.

图 9-10 CSM 模式列表[3]

9.1.6 主观评测方法总结

HDR 图像具有高亮度和宽色域的特点，因此其主观评测一般采取双刺激法，对不同的片段进行多次评估可以提高其准确性，但因此需要耗费大量的人力和测试成本。工业界提出多种客观评价方法，下一节将对其作详细介绍。

9.2 HDR 客观评价方法

HDR 客观质量评价方法与主观方法区别在于不需要观察者做主观评分，而是根据图像特征做直接的数据处理，具有更高的效率，根据失真图像其对于参考图像数据的需求程度可以分为三类：

全参考 FR (Full Reference): 将失真视频图像和无失真的源参考图像进行对比，评价失真图像相对源参考图像的质量损失

半参考 RR (Reduced reference): 从失真视频图像和无失真原始图像中分别提取图像层的某些有效特征，得出对失真视频帧的质量评价（不常用）

无参考 NR (No Reference): 不参考源图像

下面对 FR 方法和 NR 方法进行详细的分类。

9.2.1 全参考质量评价方法

根据[4]传统的全参考质量评价方法可分为以下几类：

面向统计的度量：

主要通过计算像素点值的不同来比较参考图像和失真图像的差距，常用的有 MSE 和 PSNR。但其对于 HDR 适用度较差，改进方案有：

1) mPSNR[5]通过对于不同曝光度独立计算 MSE 值再取平均；

2) PU-PSNR 通过在计算 MSE 前先做 Perceptually Uniform (PU) 编码[6]从而考虑到了人类视觉系统对于亮度的非线性反应，从而更准确地评估了 HDR 图像的质量。

结构相似性度量：

通过比较亮度、对比度、结构等差异性来衡量失真性，常用的有 UQI[7]、SSIM[8]、MS-SSIM[9]、M-SVD[10]、QILV[11]等，PU-SSIM 也类似 PU-PSNR 做了编码来提高适用性。

视觉信息度量：

通过测量 HVS 的一些视觉特征或者视觉保真度来衡量失真性，常用的方法有 IFC[12]、VIF[13]、VIFp[13]、FSIM[14]等

信息权重度量：

通过对不同的区域设置局部权重代表其对失真的感知程度，来更符合感知标准地衡量失真性，常见的有 IW-MSE[15]、IW-PSNR[15]、IW-SSIM[15]等。

基于 HVS 的度量：

尝试模拟人类对于自然场景的感知，常见地有 JND_st[16]、WSNR[17]、DN[18]等。颜色差异度量：

尝试补偿 CIE1976 的颜色梯度和感知的颜色的非线性，常用的有 CIE1976[19]、CIE94[20]、CMC[21]、CIEDE2000[22]等。

限于篇幅，下面详细介绍两种最常用、准确性最高的适用于 HDR 图像的质量评价算法 HDR-VDP-2[23]以及 HDR-VQM[24]。

HDR-VDP-2

HDR-VDP-2[23]是一个用于比较参考图片和测试图片的视觉度量，并提供可见性和质量两个方面的预测信息：可见度——参考图片和测试图片之间的差异性能够被普通的观看者看得到的可能性；质量——测试图像相对于参考图片的质量退化，以平均意见得分 MOS 进行表达。

下图是 HDR-VDP-2 的质量评测标准的使用流程。其输入分别是测试图片和有质量损失的参考图片（一般是两个 HDR 图像或者两个 LDR 图像），然后经过 HDR-VDP-2 处理后产生一个检测概率图：整个概率值使用 0-1 的 P 表示，P 越大表示检测概率越大；质量预测使用 0-100 的参数 Q_{MOS}，Q 值越大表示质量越好。检测概率图告诉我们有多大的可能会感受到两个图的差异性，红色表示可能性较高，绿色表示可能性较低。由于失真是噪声和模糊的共同造成的，因此在平滑区域（噪声）和高对比度区域有最大的可能性检测到失真。

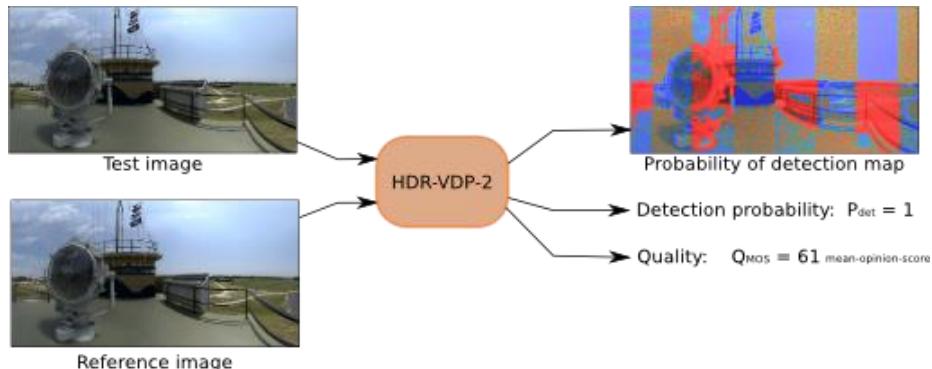


图 9-11 HDR-VDP-2 的输入及输出

尽管视觉差异性度量有很多，但相比于其他可见差异性度量，HDR-VDP-2 度量有其独特特点。首先，它可以应用到真实世界全范围亮度，即可以进行 HDR 的质量评价；其次，它对可见性和质量进行分别预测，这两个标准适用于不同目的且不相关；HDR-VDP-2 经过了严格测试和校准用于保证高精确度；最后，该方法代码开源：

<http://hdrvdp.sourceforge.net/wiki/>

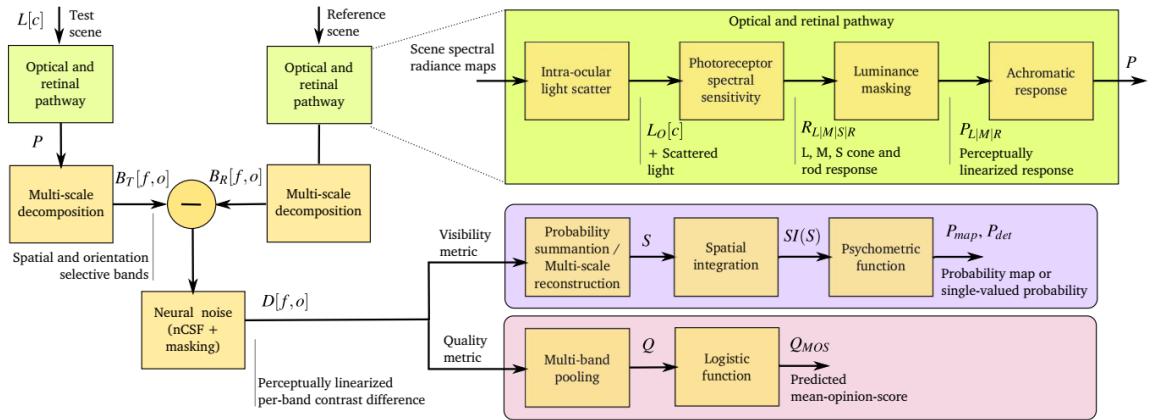


图 9-12 HDR-VDP-2 算法流程框图（包含可见性预测与质量预测）

HDR-VDP-2 主要包含的模块如图 12 所示，下面对各模块的主要功能和关键公式做简要介绍。

光学及视网膜通道模型 (Optical and retinal pathway)

1) 眼内光分散 (Intra-ocular light scatter)

透过眼睛的光少部分分散在角膜、晶状体及视网膜上，这种分散放大了图像高频信息的同时减弱了投射到视网膜上光的对比度。这种现象在观看含有强光 HDR 场景时更明显。光色散通过如下调制转移函数 (Modulation Transfer Function, MTF) 进行建模，该模型的输入是光谱图 $L[c]$:

$$F\{L_O\} = F\{L\}[c] \cdot \text{MTF} \quad (1) \text{Equation Section (Next)} \text{Equation Section (Next)}$$

其中 $F\{\cdot\}$ 表示傅里叶变换， \cdot 表示输入光谱图的标号。MTF 函数如下所示，其中 p 表示图像域以度为单位的频率周期。

$$\text{MTF} = \sum_{k=1..4} a_k e^{-b_k p} \quad (2)$$

2) 感光光谱灵敏度 (Photoreceptor spectral sensitivity)

感光光谱灵敏度曲线表示光感受细胞可以感知特定波长光的概率，如下图所示对于 S/M/L 锥细胞及视杆细胞的敏感度曲线。

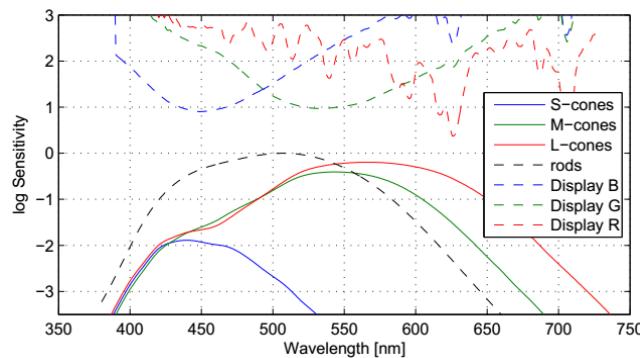


图 9-13 视锥及视杆细胞的感光光谱灵敏度曲线

当观测到具有频谱函数 $f[c]$ 的光时，不同感光细胞可以感知光的比例表示为

$$v_{L|M|S|R}[c] = \int_{\lambda} \sigma_{L|M|S|R}(\lambda) \cdot f[c](\lambda) d\lambda \quad (3)$$

其中 σ 表示视锥及视杆细胞的频谱敏感度， c 代表散射函数为 $f[c]$ 的输入频谱图标。当给定 N 个输入频谱图时，每种感光细胞感受的总光量为

$$R_{L|M|S|R} = \sum_{c=1}^N L_c[c] \cdot v_{L|M|S|R}[c] \quad (4)$$

3) 亮度掩盖 (Luminance masking)

感光细胞对光波长敏感，对光强呈现高度非线性响应。人眼可观察很宽的物理光亮度范围，主要是由于感光细胞具有增益控制能力，可以根据接收光强调整感光敏感度。这种过程通常称为亮度掩盖。使用如下非线性感光函数 $t_{L|M|R}$ 描述上述调整机制。

$$P_{L|M|R} = t_{L|M|R}(R_{L|M|S|R}) \quad (5)$$

上述函数未考虑 S 锥细胞（因为此类细胞不承担亮度感知功能）。非线性感光函数如下：

$$t_{L|M|R}(r) = s_{peak} \int_{r_{min}}^r \frac{1}{\Delta r_{L|M|R}} d\mu = s_{peak} \int_{r_{min}}^r \frac{s_{L|M|R}(\mu)}{\mu} d\mu \quad (6)$$

其中 r 表示感光细胞的接收光 ($RL|M|R$)， r_{min} 是最小的可感知光强($10-6 cd/m^2$)， $\Delta r(r)$ 是检测阈值， $s_{L|M|R}$ 是感光细胞光敏感度， s_{peak} 是调整视觉系统的敏感度峰值（需要针对不同的数据集调整）。 $s_{L|M|R}$ 通过如下方式计算，首先计算不同感光细胞敏感度的和：

$$s_A(l) = s_L(r_L) + s_M(r_M) + s_R(r_R) \quad (7)$$

其中 l 是适应光强度 (Adapting Luminance, $l=L_a+R_m$)， s_A 包含在 CSF 函数中，通过在每个亮度级别下最大化对比度敏感度求得

$$s_A(l) = \max_{\rho} (\text{CSF}(\rho, l)) \quad (8)$$

由于不存在可以单独分析 L -和 M -锥细胞光敏感度的数据，因此假设 $s_L=s_M$ ，但是对于一些色盲患者的测定可以得到 R -杆细胞的光敏感度 s_R ，因此视锥细胞敏感度函数可以视为正常人与色盲患者的光敏感度函数之差决定：

$$s_{L|M}(r) = 0.5(s_A(2r) - s_R(2r)) \quad (9)$$

对比度敏感函数 (Contrast Sensitivity Function, CSF) 函数模型如下：

$$\text{CSF}(\rho) = p_4 s_A(l) \frac{\text{MTF}(\rho)}{\sqrt{(1 + (p_1 \cdot \rho) p_2) \cdot (1 - e^{-(\rho/7)^2})^{-p_3}}} \quad (10)$$

其中 ρ 是以周期每角度 (cycles-per-degree) 为单位的空间域频率， $p_1..4$ 是在为不同的适应亮度 l 下的进行参数拟合得到的参数，其他亮度值对应参数通过使用对数亮度值作为插值系数对测定亮度值下的对比度函数值进行插值得到。 $s_A(l)$ 是视锥与视杆细胞的联合亮度敏感度（定义如式 7），通过如下计算模型得到：

$$s_A(l) = p_5 \left(\left(\frac{p_6}{l} \right)^{p_7} + 1 \right)^{-p_8} \quad (11)$$

通过调整 p_4 及 p_5 参数使得 CSF 与 $sA(l)$ 的比值峰值为 1. 故 $sA(l)$ 可以计算得到。

4) 消色差相应 (Achromatic response)

为了计算视锥及视杆细胞的消色差响应，将它们对应相加：

$$P = P_L + P_M + P_R \quad (12)$$

相等权重是基于 L/M 细胞对于亮度感知的贡献相同。而视杆细胞对亮度的感知贡献 PR 由非线性感光函数 $tL|M|R$ 控制，因此不需要额外的权重。

多尺度分解 (Multi-scale decomposition)

一些研究表明视觉中枢中存在一种机制，该机制对于图像域频率和方向的某个特定范围具有选择性。为了模拟这种机制，视觉模型通常使用多尺度图像分解，一般利用小波变换或金字塔结构。本模型中使用了方向金字塔结构以对不同的图像域频率和方向进行分隔。与其他视觉分解方式相同，每个频率带的带宽随着频率的减少而减半。每幅图像分解为四个方向带以及该图像分辨率对应的最大空间域频率带数目。

神经噪声 (Neural noise)

HDR-VDP 将对比度检测中的差异归于不同来源的噪声，影响不同频率带中的对比度检测的总体噪声等于信号独立噪声（即 **neural contrast sensitivity function**）以及信号依赖噪声（即 **visual masking**）。若参考图片和待检测图片的第 f 个频率带及第 o 个方向金字塔表示为 $BT|R[f,o]$ ，噪声归一化的信号差异为

$$D[f,o] = \frac{|B_T[f,o] - B_R[f,o]|^p}{\sqrt{N_{nCSF}^2[f,o] + N_{mask}^2[f,o]}} \quad (13)$$

指指数经试验测定 p 取 3.5，控制掩盖函数的形状。

1) 神经对比度敏感函数 (Neural contrast sensitivity function)

信号独立噪声可以通过测定 CSF 的实验测定。然而，为了使用 CSF 函数，需要抵消其中已经被考虑在 MTF 函数（公式 2）的光流信息以及被考虑在光感受函数内的亮度独立信息，因此神经对比度敏感函数通过人眼光学的 MTF 函数与联合亮度敏感度函数 SA（公式 8）对 CSF 归一化得到的，又因噪声幅度与敏感度幅度成反比，因此噪声幅度为

$$N_{nCSF}[f,o] = \frac{1}{nCSF[f,o]} = \frac{MTF(\rho, L_a) s_A(L_a)}{CSF(\rho, L_a)} \quad (14)$$

ρ 是空间域频率带 f 的峰值敏感度，通过如下式计算

$$\rho = \frac{n_{ppd}}{2^f} \quad (15)$$

其中 n_{ppd} 是输入图像给定视角像素的数目，对于最高频率带 $f=1$ 。

2) 对比掩盖 (Contrast masking)

信号独立分量 N_{mask} 主要刻画对比度掩盖现象，该现象使得不均匀背景中的细微差别可见性降低。如果一个模式叠加到另一个具有类似空间域频率及方向的模式之上，

则前者的可见性降低，该现象被称为对比掩盖现象，通过如下三元素求和进行建模。

$$N_{mask}[f, o] = \frac{k_{self}}{n_f} (n_f B_M[f, o])^q + \frac{k_{xo}}{n_f} \left(n_f \sum_{i \in O \setminus \{o\}} B_M[f, i] \right)^q + \frac{k_{xn}}{n_f} (n_{f+1} B_M[f+1, o] + n_{f-1} B_M[f-1, o])^q$$

(16)

其中第一行对应于自掩盖，第二行对应于跨方向掩盖，第三行对应于两个临近频率带的掩盖。kself, kxo, 和 kxn 是控制不同来源掩盖的权重。第二行的 O 表示方向集合，指数 q 控制掩盖函数的斜率。nf=2-(f-1)用来使用 q 指数化前归一化项，BM[f, o]是频带 f 及方向 o 上的活动

$$B_M[f, o] = \min\{|B_T[f, o]|, |B_R[f, o]|\} nCSF[f, o]$$

(17)

可见度指标 (Visibility metric)

1) 心理测量函数 (Psychometric function)

公式 (13) 中将每个频带中的信号进行了归一化，D=1 对应于某个特定频率和方向选择机制的检测阈值。D[f,o] 通过如下心理测量函数映射为概率值 P

$$P[f, o] = 1 - \exp(\log(0.5) D^\beta[f, o])$$

(18)

其中是测量函数的斜率，取值为 1，引入常数项 log(0.5) 的目的是使得 D=1 时 P = 0.5。

2) 概率集成 (Probability summation)

Pmap 对应于不同频率带及方向的概率集成，表现为一张空间图，图中的每个像素点表示

$$\begin{aligned} P_{map} &= 1 - \prod_{(f, o)} (1 - P[f, o]) \\ &= 1 - \prod_{(f, o)} \exp(\log(0.5) D^\beta[f, o]) \\ &= 1 - \exp\left(\log(0.5) \sum_{(f, o)} D^\beta[f, o]\right) \end{aligned} \quad (19)$$

通过将乘积运算变为加法运算，可以使用可控金字塔 (steerable pyramid) 的重建变换将不同频带的概率相加，这样的重建变换包含每个频带内的重建滤波、上采样及将不同频带所得结果求和，与 (19) 中对不同频带间差异 D 的求和等价，因此在实际操作中使用可操控金字塔重建函数 F⁻¹ 进行如下操作，SI 表示空间域集成函数：

$$P_{map} = 1 - \exp\left(\log(0.5) SI(F^{-1}(D^\beta))\right) \quad (20)$$

3) 空间域集成 (Spatial integration)

使用空间域集成可以使得大型模式更容易被检测到，其数学表达如下

$$SI(S) = \frac{\sum S}{\max(S)} \cdot S \quad (21)$$

其中 $S = F^{-1}(D^\beta)$ 表示对比度差异图。

质量指标 (Quality metric)

一般情况下，重要的是视觉差异对图像质量的影响而非视觉差异本身。

HDR-VDP-2 设计目的主要用于预测视觉差异而不是质量，但是经过如下方式可转换为质量分数。

1) 池化策略 (Pooling strategy)

质量指标的主要目的是将一对待对比图片的进行感知层面上的区分，因此损伤的幅度值对应于视觉可见性而不是像素值之间的数学差值。HDR-VDP-2 通过对每一频带计算 $D[f, o]$ 实现这一目标，然而此值对应于不同频带内的像素值差异，需要将不同频带内的信息聚类为预测图像质量的单一数值。

为了找到最佳的聚类策略，作者在 LIVE 及 TID2008 数据库上（由于 HDR-DVP-2 操作在物理亮度值上，因此各数据库中的图片首先被转换为三色 XYZ 值，假设标准 LCD 显示器，CCFL 背光，sRGB 色域坐标， $\text{gamma} = 2.2$ ， 180 cd/m^2 峰值亮度及 1 cd/m^2 的黑值）测试了 20 种不同聚类方法，例如：最大值聚类，百分比聚类，以及 Minkowski 聚类，最终选择了如下聚类方式：

$$Q = \frac{1}{F \cdot O} \sum_{f=1}^F \sum_{o=1}^O w_f \log \left(\frac{1}{I} \sum_{i=1}^I D^2[f, o](i) + \varepsilon \right)$$

(22)

其中 i 表示像素下标， ε 表示避免当 D 值为 0 时运算结果溢出添加的常量(10-5)， I 表示像素总数， w_f 表示逐频带权重，初始值为 1，在 LIVE 数据库上使用褪火法进行优化。

2) 逻辑回归映射 (Logistics function)

客观预测值不是直接对应主观 MOS 值，需要如下的 logistic 非线性函数映射为主观分值：

$$Q_{MOS} = \frac{100}{1 + \exp(q_1(Q + q_2))}$$

(23)

上述聚类以及映射过程使用 LIVE 数据库进行，使用 TID2008 数据库进行检验。

需要注意的是，目前 HDR-VDP-2 只进行亮度值的比较，而忽略色度信息，不对色度差异性进行检测。另外，它只能检测到参考图片和测试图片的差异性，给出基于差异的有参质量得分。

HDR-VQM

HDR-VQM[24]质量评价方法具有动态范围独立的特征，并且考虑到了时域变化通过 HVS 的时域模型，由于人类视觉系统倾向于在特定时间内关注特定区域，从而可在时空界内分析质量变化。其算法流程图见下图 14。

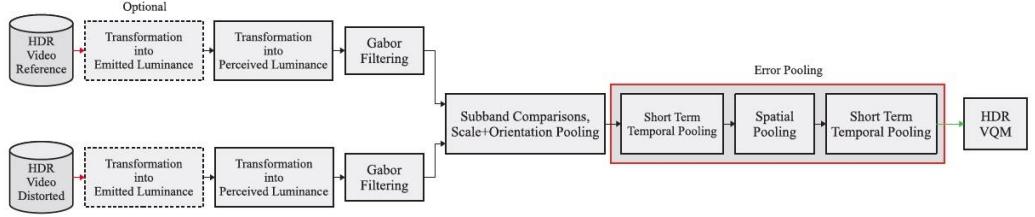


图 9-14 HDR-VQM 算法流程图[24]

第一阶段将信号的范围转换到给定的显示范围。这与 HDR 一样重要，因为不同的 HDR 显示具有不同的峰值亮度和动态范围。若源信号已参考给定范围，则不需要这一操作。

第二阶段，根据人类视觉系统对亮度的反应，利用 PU 编码将信号转换为基于感知的表示，从而保持视觉感知上具有均匀的间距。

子带误差阶段，在参考信号和失真信号之间产生了空间误差，并使用 log-Gabor 滤波器在不同尺度和方向上找到不同。子带对滤波后的信号取逆 DFT 来获得。其空间误差计算公式为：

$$Err_{t,s,o} = \frac{2l_{t,s,o}^{src} l_{t,s,o}^{dis} + \epsilon}{(l_{t,s,o}^{src})^2 + (l_{t,s,o}^{dis})^2 + \epsilon} \quad (24)$$

其中 $l_{t,s,o}^{src}$ 和 $l_{t,s,o}^{dis}$ 分别指参考视频和对应失真视频在第 t 帧，尺度 s，方向 o 上的子带， ϵ 是个小的正数，来避免非连续性。则每帧的误差公式为：

$$Err_t = \frac{1}{N_{scale} \times N_{ori}} \sum_{s=1}^{N_{scale}} \sum_{o=1}^{N_{ori}} Err_{t,s,o} \quad (25)$$

考虑到短时记忆效应，最后阶段在临近非重叠时空区域将误差池化，并计算整体值其公式为：

$$HDR-VQM = \frac{1}{|L_p|} \times \left| v \in L_p \right| \sum_{t_s \in L_p} \sum_{v \in L_p} ST_{v,t_s} \quad (26)$$

(26)

其中 L_p 表示最低 p% 值的集合， $v & t_s$ 分别空间和时间的下标， ST_{v,t_s} 为对应的误差帧。

FR 质量评价总结

Francesco 等人对于各类 FR 方法做了评测[25]，其评测标准是与主观测试评分做相似度估计，常用的标准吻合程度通常用 LCC 线性相关系数、SROCC 秩相关系数、RMSE 均方根误差、OR 背离率等参数来衡量。其中客观评分要先进行非线性压缩，再与主观评分做拟合。吻合度最高的方法为 puPSNR、puSSIM、HDR-VDP、HDR-VQM，这也是业内广泛接受并使用的方法。

9. 2. 2 无参考质量评价方法

尽管学术界对于无参考质量评价的方法很多，但是尚缺乏专门针对 HDR 进行优化，并得到业界广泛认可的方法。为完整起见，这里仅做简要介绍。随着 HDR 主观测试数据的逐渐丰富，NR 评价方法会逐步成熟。传统上，无参考（NR）质量评价有两大类方法：

A. 有特定失真

即基于一些指标如模糊度、块效应、噪声、对比度等进行计算，常用的有 JND: just noticeable distortion[26], VAR: variance[27], LAP: laplacian[28], GRAD: gradient[28], FTM: frequency threshold metric[29], HPM: HP metric[30], Marziliano: Marziliano blurring metric[29], KurtZhang:kurtosisbased metric[31], KurtWav: kurtosis of wavelet coefficients [32], AutoCorr: auto correlation [28], RTBM: Riemannian tensor-based metric[33];

由于 HDR 图像具有更大色域和更高亮度，指标评价上会考虑更多，如色彩丰富度、亮度等。

B. 无特定失真，则有如下几类：

1) 基于支持向量机

先用 SVM 进行失真类型识别，从而对特定失真类型进行 SVR 回归模型分析。如基于失真辨识的图像真实性和完整性评价(DIIVINE)算法[34]和盲/无参考图像空域质量评价(BRISQUE)[35] 算法。

2) 基于概率模型的算法

这类方法首先建立图像特征与图像质量之间的统计概率模型，对待评价图像，提取特征后根据概率模型计算最大后验概率的图像质量来估计图像质量。如基于 DCT 统计信息盲图像完整性指数(BLIINDS)算法[36]和 ILNIQE 算法[37]。

3) 基于码本的方法

这类方法通过聚类分析根据图像特征生成码本(词典)，并通过某种方式建立码本和图像质量之间的映射关系，对待评价图像，提取特征后通过匹配码本来估计图像质量。如无参考图像质量评价码本表示(CORNIA) [38] 、BLISS[39]、dipIQ[40]、HOSA[41]等

4) 基于神经网络的方法

与基于支持向量机的方法类似，这类方法先提取一定的图像变换域或空间特征，再基于已知质量数据训练一个神经网络回归分析模型，由图像特征预测图像质量。如 deepIQA[42] 和多任务端到端优化深度网络（MEON）[43]，其中 MEON 的性能已经逼近主观评测方法，这也是未来研究的主要方向。

第三部分 HDR 编码方案

第十章-MPEG 及 VCEG 的 HDR 编码优化

与传统标准动态范围（SDR）视频相比，高动态范围（HDR）视频由于比特深度的增加提供了更加丰富的亮区细节和暗区细节。最新的显示技术通过清晰地再现 HDR 视频内容使得为用户提供身临其境的观看体验成为可能。面对目前日益增长的 HDR 视频消费需求，研究现有的压缩工具或引入新的技术来高效压缩 HDR 视频变得十分迫切。本文将介绍有关 MPEG 及 VCEG 针对 HDR 视频压缩的一些工作。

10.1 背景

传统电视和显示器上显示的视频信号为标准动态范围视频，设计的亮度范围为 $0.01\text{cd}/\text{m}^2$ 到 $100\text{cd}/\text{m}^2$ ，对比度为 $10000: 1$ 。而在现实世界中，亮度范围要大得多，例如阳光照射的云层具有 $100000\text{ cd}/\text{m}^2$ 的亮度，而夜空的亮度仅为 $0.0001\text{ cd}/\text{m}^2$ 。近年来，显示器的亮度峰值持续在增加，市面上已经出现了亮度峰值可以达到 $1000\text{cd}/\text{m}^2$ 的电视机产品，而一些专业的监视器则可以达到 $10000\text{cd}/\text{m}^2$ ，这些显示器大多基于 LCD 技术以及 LED 背光，使得亮度峰值能够达到较高的水平。OLED 显示器也可以达到一个很高的亮度级别，但峰值略低于 LCD & LED 技术，不过 OLED 的优点是其暗度可以达到 $0.0005\text{cd}/\text{m}^2$ 。

超高清电视（UHDTV）的关键目标之一是增强用户观看视频时的体验，也就是所谓的“沉浸感”和“真实感”。仅仅依靠提高分辨率不足以完全达到这一目的，也无法创建、捕捉和显示比今天的电视机具有更高峰值亮度和更大对比度的视频内容。高动态范围（HDR）视频的目的是提高每个像素还原视频真实场景内容的能力，特别是，最亮像素的亮度增加，同时最暗像素可以变得更暗。另外，呈现更强的真实感需要提供更加丰富的色彩，因此必须使用具有较大色域范围的颜色空间（宽色域 WCG）来表示更加饱满的色彩。因此，高动态范围（HDR）与宽色域（WCG）相结合，可以使视频看起来更加清晰、逼真，与传统的标准动态范围视频相比，能够有效改善用户的视频观看体验。

与 SDR 视频相比，可以从以下几个方面了解 HDR 视频的效果：

- 即使分辨率没有改变，图像也可能看起来更加清晰。人眼将清晰度视为分辨率和对比度的组合。因此，当对比度增加时，人眼感知的清晰度也会增加。因此，HD 和 4K 视频都可以通过高动态范围（HDR）增强视觉效果。

- 明亮的物体看起来会更加逼真。许多物体包含明亮区域，即使很小，也可能会影响观看效果，因为它们比周围环境更亮。HDR 视频会以较高的保真度地表现明亮区域的细节，增加水面、玻璃表面、金属表面等物体的真实感。

- 阴影区域中的细节可能会增加。在 SDR 视频中，较暗区域看起来是均匀和平坦的，而在 HDR 视频相同场景下可以清楚地看到阴影中的细节。

- HDR 和 WCG 的组合还会产生更加饱满的色彩。对于诸如天空之类的明亮物体，HDR 和 WGC 允许出现明亮且同时饱满的颜色。

10.2 HDR 发展历程

10.2.1 ITU-R for SDR and HDR

在 ITU-R 的建议书中描述了用于表示 SDR 和 HDR 视频格式的基本技术规范。其中 BT.709 “Parameter values for the HDTV standards for production and international programme exchange” 包括了如何将 8 比特线性的 RGB 图像信号转换为非线性的 R'G'B' 信号，这个转换通常称为光电转换函数 (opto-electronic transfer function, OETF)。BT.709 还包含了 CIE1931 色度坐标的定义，以及从 R'G'B' 向 Y'CbCr 颜色空间转换方法的介绍。实际上，相比于色度的变化，人眼对于亮度的变化更为敏感，因此通过这种转换之后就可以降低色度分量的空间采样率。但应该注意的是，亮度分量并不直接对应于实际的亮度信息，一部分亮度信息包含在色度分量中，这就会导致在对色度分量进行下采样时也会对亮度信息产生影响，这种现象称为色度泄漏 (chroma leakage)，这对 HDR 视频的影响会比 SDR 视频更严重，下文会讨论关于色度泄漏的解决方法。BT.1886 “Reference electro-optical transfer function for flat panel display used in HDTV studio production” 中介绍了从非线性 R'G'B' 信号向线性 RGB 信号转换的操作方法，该转换操作在视频内容显示前进行，称为电光转换函数 (electro-optical transfer function, EOTF)。这里提出的电光转换方程 EOTF 与 BT.709 标准中的光电转换方程 OETF 并不是互逆的关系，此 EOTF 和逆 OETF 转换函数的差异可以用一个光电光学转换函数 (opto-optical transfer function) 或系统参数 gamma 来表示，该 gamma 参数能够使得图像看起来更加自然真切。BT.2035 “A reference viewing environment for evaluation of HDTV programme material or completed programmes” 中给出了视频观看环境的参考设置，其中规定了亮度范围为 0.01cd/m² 到 100cd/m²。

BT.2020 “Parameter values for ultra-high definition television systems for production and international programme exchange” 中引入了 4K (3840x2160) 分辨率和宽色域 (WCG) 的概念。宽色域意味着其基色范围相对于传统视频更广，因而可以显示出更加饱满的颜色，例如深蓝、品红、荧光绿等。但需要注意的是，没有 HDR 的 WCG 其效果是很有限的，同样地，不引入 WCG 的 HDR 视频也是没有太大意义的，因此 HDR 和 WCG 可以很自然地结合起来用于改善每个像素点的重建效果，呈现出更为生动的视频内容。此外，如前文所述，HDR 和 WCG 产生的效果几乎完全独立于空间分辨率，因此可以灵活应用在 4K 分辨率及 1080p 分辨率视频中。与此同时，BT.2020 包括了与 BT.709 相同的非恒定亮度颜色空间转换，还包括可以直接应用在 RGB 信号上的恒定亮度转换。在恒定亮度下，OETF 在使用前要先进行颜色空间转换，但这种方法没有非恒定亮度那样应用广泛。

BT.2100 “Image parameter values for high dynamic range television for use in production and international programme exchange” 中介绍了两种不同的 HDR 转换函数：HLG (Hybrid Log-Gamma) 和 PQ (perceptual quantization)。HLG 的光电转换函数 OETF 与 ARIB STD-B67 中的 OETF 是相同的，并且对支持 WCG (BT.2020) 的 SDR 显示器有一定的兼容性，但需要注意的是 HLG 本身不兼容只支持 BT.709 色域的显示器。PQ 电光转换函数 EOTF 与 SMPTE ST.2084 中的 EOTF 相同，并具有与人类视觉系统中分布比较一致的亮度级。SMPTE ST.2084 于 2013 年发布，近年来被广泛应用于工业界，并被多个不同的标准组织采纳。MPEG 和 JCT-VC 相关的工作已经将 PQ 电光转换函数 EOTF 作为研究和测试的基准。此外，BT.2100 中的颜色基色与颜色空间转换和 BT.2020 中给出的一致。

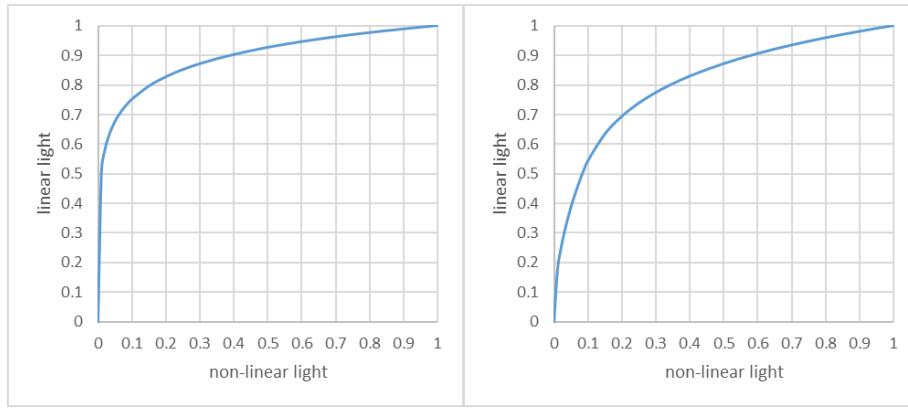


图 10-1 PQ（左）和 HLG（右）转换曲线

10.2.2 MPEG 和 VCEG 关于 HDR 的工作

MPEG 在 2013 年开始研究 HDR 视频压缩时进行了多项探索性的研究测试，主要思路是将 HDR 视频压缩和处理过程的各部分分开来研究，例如在此期间研究了不同的电光转换函数，并对不同的颜色格式转换和采样方法进行了测试和评估，与此同时还研究了一些主观视频质量测试方法和客观视频质量评价指标。可以预计的是，传统的基于像素误差的评价指标如 PSNR 在高精度 HDR 视频的评价中无法提供太多有价值的参考信息。目前已提出了一些专门针对 HDR 视频的质量指标，例如 wPSNR、tPSNR-Y、deltaE100、PSNR-L100 等，但仍然没有一个主客观评估结果较为一致的评价指标。

2015 年初，MPEG 决定发布针对 HDR 和 WCG 视频压缩技术标准的 CfE[1]，目的是为了确定当前是否存在能够改善 HDR 视频压缩性能的新技术。CfE 中给出了一个参考配置作为对比基准 Anchor 1.0，其中大部分技术方案与 HDR10（在下一节介绍）完全兼容。Anchor 1.0 使用 PQ 电光转换函数、Y'CbCr 颜色空间、HEVC Main10 压缩配置，基于未修改的 HEVC 参考软件 HM 进行编码，并使用 HDR Tools 软件包[2]来实现必要的预处理和后处理转换步骤。HDR Tools 本身是开源免费的，其在 MPEG 和 VCEG 进行的多种评估测试中都有所应用。

CfE 提出后，收到了来自 Apple、Arris、BBC、Dolby、Ericsson、FastVDO、InterDigital、MovieLabs、NGCodec、Philips、Qualcomm、Technicolor、University of Warwick/goHDR 等总共 9 种不同的技术提案，所有提案的编码结果在两个独立的实验室分别进行了主观评测，得到的结果与 Anchor 1.0 进行了对比。主观评测结果显示，在很多序列上部分提案的结果相比 Anchor 1.0 有明显的提升[3]。

与此同时，CfE 中的一些序列测试结果表明，可以通过 chroma QP offset [4]（第 3 节介绍）将少量比特从亮度分量转移到色度分量，从而可以减少色度失真。此外也发现在 Anchor 1.0 中分配给图像暗区的比特过多，而分配给亮区的比特太少[5]，为此引入了亮度调整（Luma Adjustment，第 3 节介绍）方法。最后综合多种新的改进技术形成了另一个基准，Anchor 3.2。相对于 Anchor 1.0，Anchor 3.2 在各个方面都有很大程度的改善。与 Anchor 1.0 中的方法相比，同等视觉质量条件下 Anchor 3.2 可以实现大约平均 27% 的比特率降低（使用 MOS BD-Rate 计算）。不过应该注意的是，Anchor 3.2 中的改变都发生在 HEVC 编码端，而在解码端没有作任何改变，也就是说 Anchor 3.2 与 HDR10 完全兼容。Anchor 3.2 与 Anchor 1.0 的编码结果对比如图 2 所示。



图 10-2 Anchor 1.0（左）和 Anchor 3.2（右）在相同码率下的压缩效果

在 2016 年 2 月的 MPEG 会议上，对已有工作进行了分析后得出结论：暂时没有找到一套明确的技术方案能够依此建立针对 HDR 视频的新的 HEVC 规范，因此研究重点转向了技术报告。MPEG 和 VCEG 随后的研究工作明确了与 HDR 视频转换和压缩相关的技术特性，并形成了一份记录这些特性的技术报告，用来指导 HDR 视频的处理过程，即对支持 PQ 转换函数、4:2:0 Y'CbCr 并使用 HEVC Main10 Profile 的 HDR 编码方案，也就是 Anchor 3.2 中所包含的编码 HDR 视频的方案提供技术参考。该技术报告从 2015 年就开始着手准备了，第一个正式的版本在 2016 年 2 月发布，随后被确定为 MPEG-H 的一个新的部分，也就是 MPEG-H Part 14。该技术报告的第一版为“Conversion and coding practices for HDR/WCG video”[6]，后来标题变得更加具体，即为“Conversion and Coding Practices for HDR/WCG Y'CbCr 4:2:0 Video with PQ Transfer Characteristics”[7]。

这份技术报告所描述的 HDR 视频处理流程如图 3 所示，其中包括 BT.2020 中所规定的线性 RGB 图像信号向 10 比特、窄范围、PQ 转换函数、4:2:0 采样率、非恒定亮度 Y'CbCr 图像信号进行转换时所必需的预处理步骤，具体包括 EOTF 的应用、颜色空间转换、量化和色度下采样等转换操作。相应地，该报告还介绍了在 HDR 视频解码之后执行的后处理转换步骤，这些步骤用于将视频信号转换回线性的 RGB 表示，从而能够在显示器上显示。此外，中间编码部分的介绍并不是对视频编码的完整描述，只是突出说明了针对 HDR 视频压缩的两种有效的方法：基于平均亮度水平的 QP 自适应选择和基于内容色域的色度 QP 偏移，这两种方法将在本文第三部分详细介绍。

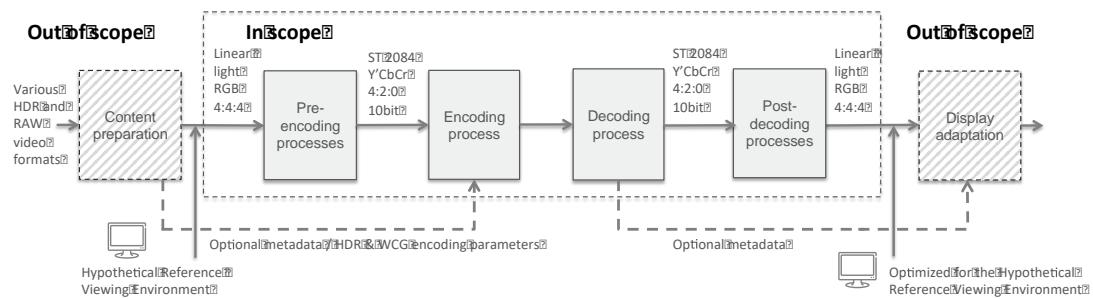


图 10-3 HDR 视频处理流程^[7]

10.2.3 基于 HEVC 的 HDR 方案

HEVC 具有较高的压缩效率，支持 10 比特精度、HDR 转换函数和 WCG，是 HDR/WCG 视频数据的理想载体。[8]中给出了将 HEVC 编码器及其 Main10 配置文件用于消费者 HDR/WCG 应用的可行性。该报告记录了使用 HEVC Main10 配置文件和 BT.2100 PQ 转换函数在不同比特率下编码 HDR 后所进行的一些主观测试结果。该报告的结论是，HEVC Main 10 配置文件确实能够以行业常见的比特率高质量地编码 HDR/WCG 视频，而无需对 HEVC 标准进行任何进一步的修改。[7]中具体给出了关于如何使用 HEVC 和 PQ 函数转换 HDR 视频以进行高效压缩，而[9]中给出了一些使用 HEVC 和 HLG 函数进行 HDR 视频编码的指导性方案。

HEVC 编码器、10 比特精度、BT.2100 色彩基色和 PQ 转换函数的组合通常称为 PQ10，同样，HEVC 编码器、10 比特精度、BT.2100 色彩基色和 HLG 转换函数通常称为 HLG10。PQ10 和 HLG10 编码 HDR 的方案目前已被许多标准化机构采用和部署，其中包括：

- ETSI / DVB

2017 年 2 月发布的技术规范 2.3.1 版本中，UHD 第 2 阶段部署采用 HEVC 编码，包括使用 PQ10 或 HLG10 方案。

- BD ROM

BDA（Blu-ray Disc Association）是最早支持 HDR 视频的行业组织之一，编码方案中采用了多个不同的 HDR 关键技术，其中主要使用的是 HDR10。HDR10 技术涵盖：PQ 转换函数、BT.2020 色域、非恒定亮度 Y'CbCr 颜色空间、4:2:0 采样格式、采用 HEVC Main10 编码配置。在实际应用时，HDR10 中还包括一些常用的元数据，比如 Mastering Display Colour Information 及 Content Light Level Information 等。

2015 年 7 月发布的用于蓝光光盘的技术规范（3.0 版本）支持 HEVC Main 10 Profile Level 5.1，最高码率 100 Mbps，分辨率为 HD 或 4K，帧率 24-60fps，兼容 PQ10 方案。

- Ultra HD Forum

2017 年 4 月 1 日发布的技术规范“UHD Phase A 内容参数”指定使用 HEVC Main10 编码配置，并采用 PQ10 或 HLG10 中的颜色表示方案。

- 3GPP

基于感知量化器 PQ 并遵循 ITU-R BT.2100 颜色格式的 HDR 方案被添加到了 3GPP TS 26.116 电视视频配置集中，具体包括 HEVC Main10 编码方案，转换函数、颜色基色和相关元数据的规定，以及与 PSS（分组交换流）和 MBMS（多媒体广播和多播服务）的集成。

- ATSC 3.0

使用 PQ10 或 HLG10 方案编码 HDR 视频。

除了标准化机构采用 PQ10 和 HLG10 之外，还有许多商业机构直接使用 HEVC 标准进行 HDR/WCG 编码、传输和播放，其中包括行业领先的流媒体服务商、主流的手机和电视制造商。目前基本上市场上所有销售的具有 HDR 功能的电视机都内置了 10 比特 HEVC 解码器。

10.3 HDR 视频编码端的关键技术

10.3.1 Luma adjustment

之前已经提到，在 Y'CbCr 表示中，一些亮度信息携带在色度分量中，当色度分量进行下采样时，可能出现色度泄漏。对于 HDR 视频，由于 PQ 电光转换函数 EOTF 具有高度非线性

特性以及相邻像素之间的亮度可能显著不同，这种影响可能非常大。图 4 给出了色度泄漏的示例，并在之后介绍了如何使用亮度调整方法解决此问题。应当注意，该例子使用了 BT.709 容器中所有可用的色度代码值，并在 BT.709 显示器上分级显示。这是因为在 MPEG 中暂时没有可用的具有完整 BT.2020 色域的测试材料，但是对于 BT.2020 容器中的测试材料预期也会具有类似的效果。因此，图 4 中的示例经过了色调映射，以便能够在 SDR 显示器中呈现。



图 10-4 色度泄漏示例 (a.原图 (4:4:4) b.色度下采样图 (4:2:0) c. Luma adjustment)

如图 5 所示，亮度调整方法[10]增加了色度上采样环节，使得 Cb 和 Cr 信息参与到亮度分量的调整中来，以确保解码后在对色度进行上采样时保留有价值的高度信息，如图 6 所示。

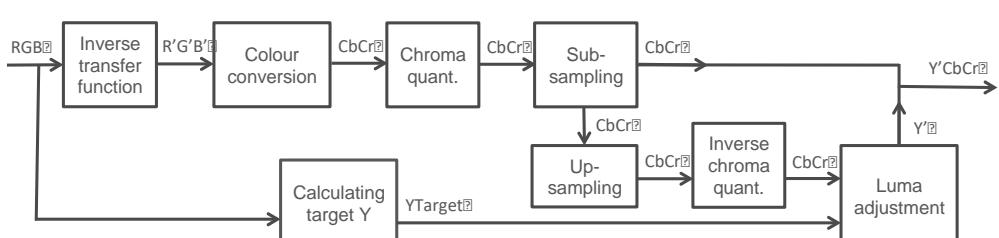


图 10-5 编码预处理过程亮度调整方法示意图

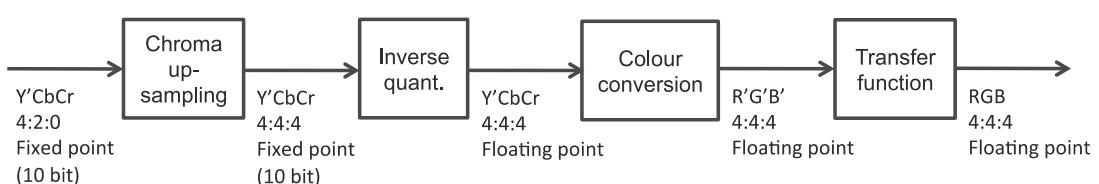


图 10-6 解码后处理过程示意图

在后处理阶段，根据亮度调整值 Y'_{adjust} 和经过上采样反量化得到的 $\widetilde{\text{Cb}}$ 和 $\widetilde{\text{Cr}}$ 经过一系列转换可以得到重建值 $\{\widetilde{R}, \widetilde{G}, \widetilde{B}\}$ ，确保 $\widetilde{\text{RGB}}$ 能够接近最初的RGB是优化的目标。 $\widetilde{\text{RGB}}$ 和RGB之间的差异可以表示如下：

$$D = \text{Diff}(\text{RGB}, \widetilde{\text{RGB}}) \quad (1)$$

一般采用基于权重的误差和来表示 D，权重 w_R, w_G, w_B 可以设定为各分量对亮度的贡献比例。

$$D = (w_R * (\widetilde{R} - R) + w_G * (\widetilde{G} - G) + w_B * (\widetilde{B} - B))^2 \quad (2)$$

根据 R 和 R' 之间的 EOTF 转换关系，上式等价于：

$$D = (w_R * (\text{EOTF}(\widetilde{R}') - \text{EOTF}(R')) + w_G * (\text{EOTF}(\widetilde{G}') - \text{EOTF}(G')) + w_B * (\text{EOTF}(\widetilde{B}') - \text{EOTF}(B')))^2 \quad (3)$$

其中， $\text{EOTF}(x)$ 为 PQ EOTF。为了得到可以收敛的结果，对 $\text{EOTF}(x)$ 进行泰勒级数展开：

$$\text{EOTF}(x_i + \Delta) \approx \text{EOTF}(x_i) + \text{EOTF}'(x_i) * \Delta_x \quad (4)$$

$\text{EOTF}'(x_i)$ 是 $\text{EOTF}(x)$ 关于 x 在 x_i 处的一阶导数。将(4)式带入(3)式可得：

$$D = (w_R * \text{EOTF}'(R') * \Delta_R + w_G * \text{EOTF}'(G') * \Delta_G + w_B * \text{EOTF}'(B') * \Delta_B)^2 \quad (5)$$

由于 $R'G'B'$ 与 $Y'\widetilde{\text{Cb}}\widetilde{\text{Cr}}$ 之间可以进行如下转换：

$$\begin{bmatrix} R' \\ G' \\ B' \end{bmatrix} = A \begin{bmatrix} Y' \\ \text{Cb} \\ \text{Cr} \end{bmatrix} = W^{-1} \begin{bmatrix} Y' \\ \text{Cb} \\ \text{Cr} \end{bmatrix} = \frac{1}{\det(W)} \text{adj}(W) \begin{bmatrix} Y' \\ \text{Cb} \\ \text{Cr} \end{bmatrix} = \begin{bmatrix} a_{RY} & a_{RCb} & a_{RCr} \\ a_{GY} & a_{Gcb} & a_{Gcr} \\ a_{BY} & a_{BCb} & a_{BCr} \end{bmatrix} \begin{bmatrix} Y' \\ \text{Cb} \\ \text{Cr} \end{bmatrix} \quad (6)$$

将上式代入(5)式替换掉变化量，D 可表示为如下所示，记为(7)式。

$$\begin{aligned} & (w_R * \text{EOTF}'(R') * (a_{RY} * Y'_{\text{adjust}} + a_{RCb} * \widetilde{\text{Cb}} + a_{RCr} * \widetilde{\text{Cr}} \\ & \quad - (a_{RY} * Y' + a_{RCb} * \text{Cb} + a_{RCr} * \text{Cr})) + \\ & w_G * \text{EOTF}'(G') * (a_{GY} * Y'_{\text{adjust}} + a_{Gcb} * \widetilde{\text{Cb}} + a_{Gcr} * \widetilde{\text{Cr}} \\ & \quad - (a_{GY} * Y' + a_{Gcb} * \text{Cb} + a_{Gcr} * \text{Cr})) + \\ & w_B * \text{EOTF}'(B') * (a_{BY} * Y'_{\text{adjust}} + a_{BCb} * \widetilde{\text{Cb}} + a_{BCr} * \widetilde{\text{Cr}} \\ & \quad - (a_{BY} * Y' + a_{BCb} * \text{Cb} + a_{BCr} * \text{Cr})) \big)^2 \end{aligned}$$

经过矩阵 A 的系数替换之后，可简化为：

$$D = (w_R * \text{EOTF}'(R') * (Y'_{\text{adjust}} - e_R) + w_G * \text{EOTF}'(G') * (Y'_{\text{adjust}} - e_G) + w_B * \text{EOTF}'(B') *$$

$$(Y'_{\text{adjust}} - e_B)^2$$

(8)

其中部分系数如下所示：

$$e_R = Y' - (\widetilde{Cr} - Cr) * a_{RCr},$$

(9)

$$e_G = Y' - (\widetilde{Cb} - Cb) * a_{GCb} - (\widetilde{Cr} - Cr) * a_{GCr},$$

(10)

$$e_B = Y' - (\widetilde{Cb} - Cb) * a_{BCb}.$$

(11)

为了得到(8)式的局部最优解，对其关于 Y'_{adjust} 求导并令 D 为 0 可得：

$$Y'_{\text{adjust}} = \frac{w_R * \text{EOTF}'(R) * e_R + w_G * \text{EOTF}'(G) * e_G + w_B * \text{EOTF}'(B) * e_B}{w_R * \text{EOTF}'(R) + w_G * \text{EOTF}'(G) + w_B * \text{EOTF}'(B)} \quad (12)$$

计算(12)式即可求导调整后的亮度值 Y'_{adjust} 。

10.3.2 QP selection based on average luma

对于 SDR 数据，使用的是 BT.709 中的幂律转换函数，而 HDR/WCG 数据则一般使用在 SMPTE ST 2084 和 BT.2100 中定义的 PQ 转换函数，因此经过处理后数据特性会有很大差别，可以使用 BT.709 转换函数和 BT.2100 PQ 转换函数处理相同的 SDR 数据来观察这一差异。对于 10 比特的数据，如果源数据的亮度峰值为 $100\text{cd}/\text{m}^2$ 并使用 BT.709 转换函数，亮度分量分布在 64-940，而如果使用 BT.2100 PQ 转换函数，亮度分量分布在 64-590，可以看到两者范围差异较大。如果使用 PQ 时在 code level 为 509 ($100\text{cd}/\text{m}^2$) 附近出现一些扰动 (+/- 1)，等价于使用 BT.709 在 code level 为 940 ($100\text{cd}/\text{m}^2$) 出现 +/- 4 的扰动。与此同时，使用 PQ 时在 code level 为 80 ($0.01\text{cd}/\text{m}^2$) 附近出现一些扰动 (+/- 1)，等价于使用 BT.709 在 code level 为 80 ($0.01\text{cd}/\text{m}^2$) 出现 +/- 1 的扰动。因此，可以推断如果编码器在亮区和暗区附近使用相同的方式处理 1 个 code level 的误差，那么相对 BT.709 使用 PQ 时将会在亮区 ($100\text{cd}/\text{m}^2$) 产生 4 倍左右的误差。也就是说，从 BT.709 切换到 PQ 将会导致许多比特从亮区重新分配到暗区。因此，原先针对 BT.709 在亮区和暗区取得较好平衡的编码器，如果不做修改而直接使用 PQ，则分配给亮区的比特过少，暗区的比特过多，所以亮区就会出现失真，对于 HDR/WCG 数据会表现得更加明显。一种常见的做法是基于局部块的平均亮度值自适应地调整块的 QP。

许多编码器都采用了自适应 QP 选择的方法。在 HEVC 中，可以单独为每个编码单元设置 QP，以便使得比特分配与人类视觉系统特性保持一致。通常，对于平坦区域（低方差）使用较低的 QP 并且对于纹理密集的区域（高方差）使用较高的 QP。此外在其他一些系统中，还可能会考虑亮度、边缘、运动等特性。但这些方法都是针对 SDR 内容特性的。对于使用 PQ 电光转换函数的 HDR 视频，需要考虑上述转换函数带来的数据特性上的改变。具体地，通过对具有较低平均亮度值的块（暗区域）使用较高的 QP，而对具有较高平均亮度值的块（亮区域）使用较低的 QP，可以有效改善图像的主观质量。下面给出一个基于平均亮度进行自适应量化的示例。

$$QP_{PQ} = QP_{709} + dQP(int_L)$$

(13)

int_L 表示局部 64×64 CTU 编码块的平均亮度值 L_{average} ， $int_L = \text{Round}(L_{\text{average}})$ 。dQP 与 int_L 的对应关系通过如下查找表和图 7 给出。

表 1 dQP 与 int_L 之间的对应关系

luma int_L range	dQP
$\text{int}_L < 301$	3
$301 \leq \text{int}_L < 367$	2
$367 \leq \text{int}_L < 434$	1
$434 \leq \text{int}_L < 501$	0
$501 \leq \text{int}_L < 567$	-1
$567 \leq \text{int}_L < 634$	-2
$634 \leq \text{int}_L < 701$	-3
$701 \leq \text{int}_L < 767$	-4
$767 \leq \text{int}_L < 834$	-5
$\text{int}_L \geq 834$	-6

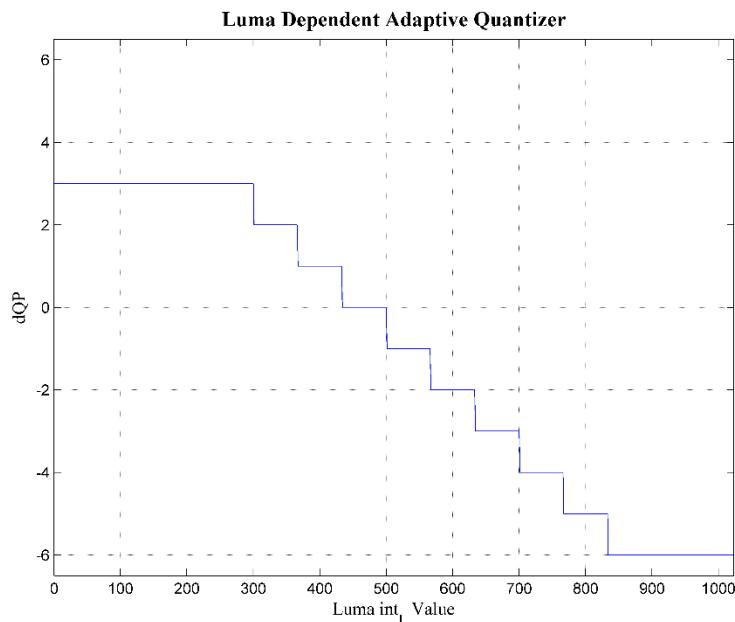


图 10-7 dQP 与 int_L 之间的对应关系

10.3.3 Chroma QP offset based on content colour gamut

针对使用 BT.709 色域和光电转换函数的 SDR 内容，现有的编码器配置已经可以在亮度和色度之间取得较好的平衡。对于 HDR 视频，与 SDR 视频相比，PQ EOTF 和 BT.2020 色彩基色的组合会使得 Cb 和 Cr 的取值更加紧密地聚集在零附近。因此，使用同样的编码配置编码 HDR 视频可能不会获得与 SDR 类似的性能。具体在编码 HDR 视频时，过于集中的色度分量在量化过程中会出现丢失，此时比特分配会从色度分量向亮度分量偏移。如果没有为色度分量分配足够多的比特，就会产生色度失真，例如，白色区域中会出现青色和品红色，如图 2 左图所示，窗户和伞内区域发散出了多种颜色。

HEVC 中的 Chroma QP offset，调节着亮度 QP 与色度 QP 之间的平衡关系。通过对 Cb 和 Cr 单独设置色度 QP 偏移，可以确保在亮度使用较高的 QP 值时，色度 QP 仍然维持在一个正常的水平，使得在量化期间不会发生较为严重的色度信息丢失。例如，在失真较为严重的低码率情况，对 Chroma QP offset 取一个较大的负值能够有效减少色度失真，在码率增加到某个点时，色度分量的质量可能已经足够好了，所以在高码率情况下可以采取较小的偏移值，

甚至设为 0。另外，如果在编码过程中已知视频内容没有应用全色域（通过分析视频或视频附带的元数据可知），则可以进一步优化色度 QP 偏移以提供更好的编码结果。

以下给出了设置 chroma QP offset 的一个例子。假定已知 HDR 捕捉和显示设备的颜色基色，chroma QP offset 可以基于亮度 QP 值和基色因子 c_{cb}, c_{cr} 采用下式给出的模型进行计算：

$$QPoffsetCb = \text{Clip3}(-12, 0, \text{Round}(c_{cb} * (k * QP + l))) \quad (14)$$

$$QPoffsetCr = \text{Clip3}(-12, 0, \text{Round}(c_{cr} * (k * QP + l))) \quad (15)$$

如果 HDR 捕捉和显示设备的颜色基色相同，则 $c_{cb} = 1$, $c_{cr} = 1$ 。如果捕捉设备采用 P3D65 基色，显示设备采用 BT.2020，则 $c_{cb} = 1.04$, $c_{cr} = 1.39$ 。如果捕捉设备采用 BT.709，显示设备采用 BT.2020，则 $c_{cb} = 1.14$, $c_{cr} = 1.78$ 。最后， $k = -0.46$, $l = 0.26$ 。

第十一章-HEVC/H. 265 的 HDR 编码方案

前文我们对 HEVC 的 HDR 编码优化技术做了介绍，侧重编码性能的提升。本章主要阐述 HEVC 中 HDR/WCG 相关的整体编码方案，包括不同应用场景下的 HEVC 扩展编码技术。

11.1 背景

HDR 信号一般意味着使用更多比特，一般的 HDR 信号倾向于使用 10 比特。这代表着更多的量化阶数，不同阶之间更平滑的过渡，但这也意味着更高的带宽和存储需求。未压缩的单个 HDR 像素需要 12 个字节的内存来存储 RGB 值的三个单精度浮点数。未压缩的 4K UHD 分辨率（3840×2160 像素）的单个 HDR 帧需要大约 94.92M 字节的存储，并且每秒 30 帧的一分钟数据需要 166G 字节。如果没有压缩，会给存储和传输带来巨大的压力。虽然地面宽带的带宽经常增加，但即使这样也不足以在未来播放任何未压缩的 HDR 电影，因此对数字信号进行压缩是进一步的需求。

HDR 视频的标准中，最常用的是由 BDA(Blu-ray Disc Association)支持的 HDR10 标准，这是最早支持 HDR 视频的几种标准之一。HDR10 标准包含的内容有：使用 PQ 电光转换函数和 BT.2020 标准中的色度空间，10 比特的非恒定亮度 Y'CbCr 空间，4:2:0 的色度抽样方法，压缩方法采用 HEVC Main 10 配置文件。在实际应用时，HDR10 中还应该包括一些常用的元数据，比如 Mastering Display Colour Information 或者 Content Light Level Information 等。

2013 年 7 月，MPEG 开始针对 HDR 和 WCG 内容分发的需求，展开探索性的实验，打破了视频压缩和处理链的各部分，转换函数被分离研究，不同的颜色表示格式测试各种抽样方法，并且做了一些主观视频质量测试方法和客观视频质量测试指标。MPEG 最初的重点是高质量的应用程序，但后来扩大到包括广播，流媒体和存储应用程序。随后，MPEG 于 2015 年 2 月发布了证据征集（CfE），并于 2015 年 6 月报告了对技术响应的评估。CfE 的目标是评估当前 HEVC 标准的 HDR 和 WCG 视频的压缩性能，并确定是否应考虑其他有前景的技术以用于 HEVC 的未来扩展。

11.2 HEVC 编码对 HDR/WCG 的支持

在 H.265/HEVC 针对 HDR 和 WCG 的 CfE 中，得到了两种类型的解决方案：一种是单层的 HDR 编码方案，包括元数据的使用，但并不需要对落后的 SDR 设备的向后兼容性；第二种明确的关注对 SDR 设备的兼容性，使用单层或多层的编码方案。这里指明的对 SDR 的向后兼容性不仅包括对现有编码器，也就是 HEVC Main 10 的兼容，也包括了对现有的 SDR 显示器的兼容。对于单层 HDR 编码，作为对比的参考技术是 HEVC Main 10；对于多层 HDR 编码，参考技术是 SHVC Scalable Main 10。HDR 和 WCG 技术需要的位深至少为 10bit，在电视制造业中，HEVC Main 10 已经广泛应用在 UHD 视频中。因此，HEVC Main 10 已经作为由 CfE 提出的技术中的单层案例的方法和可分级案例中的基础层。

第一和第二版本的 HEVC 标准已经包含了几种可应用于 HDR 和 WCG 视频的工具，包括

支持 HDR 信号传输和 BT.2020 色域的 HDR 视频信号格式、用于支持 HDR 信号在接收端转换显示的补充增强信号（SEI）信息的定义和支持更高位深视频编码的应用。除此之外，HEVC 的可分级扩展（SHVC）还包括色域可扩展性（CGS）和位深可扩展性（BDS）工具来解决针对 SDR BT.709 的向后兼容性。表 1 概括了 HEVC 编码器中为 HDR 和 WCG 提供的一些工具，而这些工具将在接下来的几个部分分别介绍。

表 11-1 HEVC 中 HDR/WCG 工具

Metadata / Tools	Purpose
VUI	Provides information related to the coded video signal “container”: color primaries, transfer function, and color space to map video sample code levels to display intensities. Besides the traditional ITU-R BT.709 combination required by DVD, ATSC, DVB, and Blu-ray 1.0, some camera log transfers and the new transfer function SMPTE ST 2084 [33], among others can be indicated. BT.2020 primaries were added to AVC and HEVC in 2013, followed by XYZ in 2014.
tone mapping information SEI message	Post-decode downloadable tone mapping curve applied to RGB primary signal. Includes multiple methods to transmit one or more curves compactly within the bitstream, where each curve targets a different mapping scenario. Note: originally developed for AVC in 2005.
mastering display color volume SEI message	Signals brightness range, primaries, and white point of the display monitor used during the grading of video content (SMPTE ST 2086 [34]).
color remapping information SEI message	Post-decode downloadable intensity map LUTs and 3x3 matrix for bit depth (dynamic range) and color space conversion.
knee function information SEI message	Post-decode downloadable tone mapping curve applied to linear (brightness) signal on all three RGB channels.
CGS/BDS	Defines color mapping between base layer and SHVC enhancement layer (e.g. mapping from BT.709 base layer to BT.2020 enhancement layer), for bit-depth and color gamut scalability.

11.2.1 VUI

VUI（Video Usability Information）是在 H.264/AVC 和 H.265/HEVC 序列参数集中携带的元数据元素的集合，它们共同描述了实际视频信号如何在样本流内存储和映射。术语“视频信号容器”起源于早期的电视和电影工程，早于文件容器概念。VUI 传达视频信号的含义，特别是整数视频采样值如何最终转换为线性显示强度，通常表示为 0 到 1.0 的浮点数信号乘以传递函数的强度范围，例如 SDR 视频的 100nits 和 HDR 视频的 10000nits。普通数字电视信

号是经过量化的一系列样本，它们拥有各自的以下属性：比特深度、非线性（Gamma）映射和导数颜色空间，以及附加属性比如色度亚采样方式。HEVC 视频编码的研究者通常更多的关注视频信号容器的属性，而不是视频编解码器及其参数，编解码器通常被当作“黑箱”来使用，它通过临时压缩，变为用于传输的视频信号，然后由解码器来解码和重新渲染为原始视频信号的表示，最终保持视频编码之前信号容器的原始含义。诸如格式转换和显示处理之类的解码后处理与 VUI 容器属性高度相关，以优化视频信号在目标显示器上的渲染。同样，VUI 容器的属性显著影响了视频信号格式转换的效果，因此为了得到内容的预期效果，对 VUI 中元数据的正确设置显得至关重要。图 1 给出了 VUI 和 SEI 元素在渲染过程即后处理过程中可能有应用的地方。接下来将介绍关键 VUI 比特流元素，其规范中的名称用斜体显示。

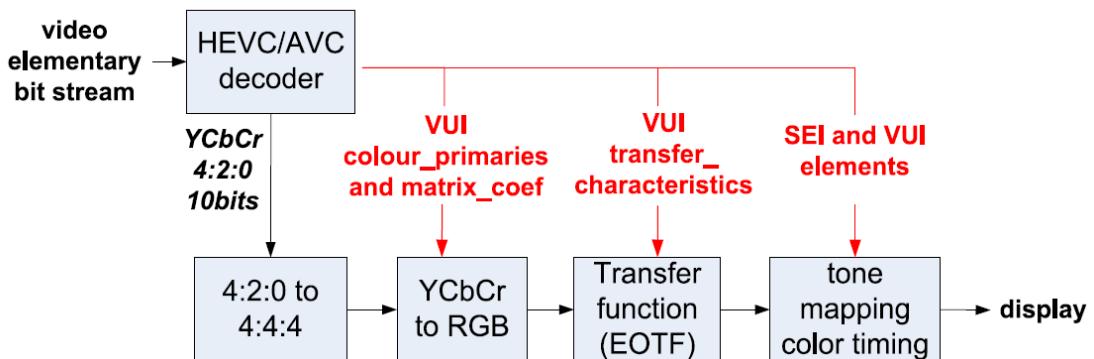


图 11-1 VUI 和 SEI 信息在渲染过程的应用

原色 (colour_primaries)

视频中的原色可以由实际的 CIE 的 xy 坐标表示为 0-1 的数值，这个属性基本上建立了视频信号包含的色域的边界。关于 WCG，HEVC 在 VUI 中包含两个新条目，以支持 BT.2020 和 XYZ 颜色原色。如由 CIE xy 坐标规定的主要三角形轨迹所示，该区域可以在 colour_primaries 的指定值之间自由选取，并且 BT.2020 拥有两倍于 BT.709 的色域范围。

传递函数 (transfor_characteristics)

传递函数建立了样本量化值（0 到 $2^{BitDepth} - 1$ 之间的整数）与亮度显示器上的亮度（以 nit 为单位）的映射关系。VUI 指示符 *transfor_characteristics* 通常对应于参考的 SMPTE 或者 ITU-R 标准的特定方程，其被应用于将原始视频线性信号转换为非线性信号，比如传统的用于 SDR BT.709 的 Gamma 0.45 函数。大多数的传递函数将线性信号转换成更紧凑的非线性信号表示，然后用编码器进行压缩。线性信号一般有均匀的量化步长，因此占用的比特数比较多，不适用于低码率为主导的信号传输，例如，BT.709 中的逆 Gamma 传递函数将非线性 8 位信号扩展到 0.1 到 100 尼特范围内大约 11 位的线性光信号，这也从侧面显示了传递函数带来的码率节省。

新的 PQ 传递函数比通常使用的 Gamma 传递函数更适用于 HDR 视频信号，并且最近被

添加到了 VUI 的传输特性表中。PQ 是 Barten 等人针对人类视觉模型对对比度敏感的特性来设计的，其被收录在 ST.2084 规范中。

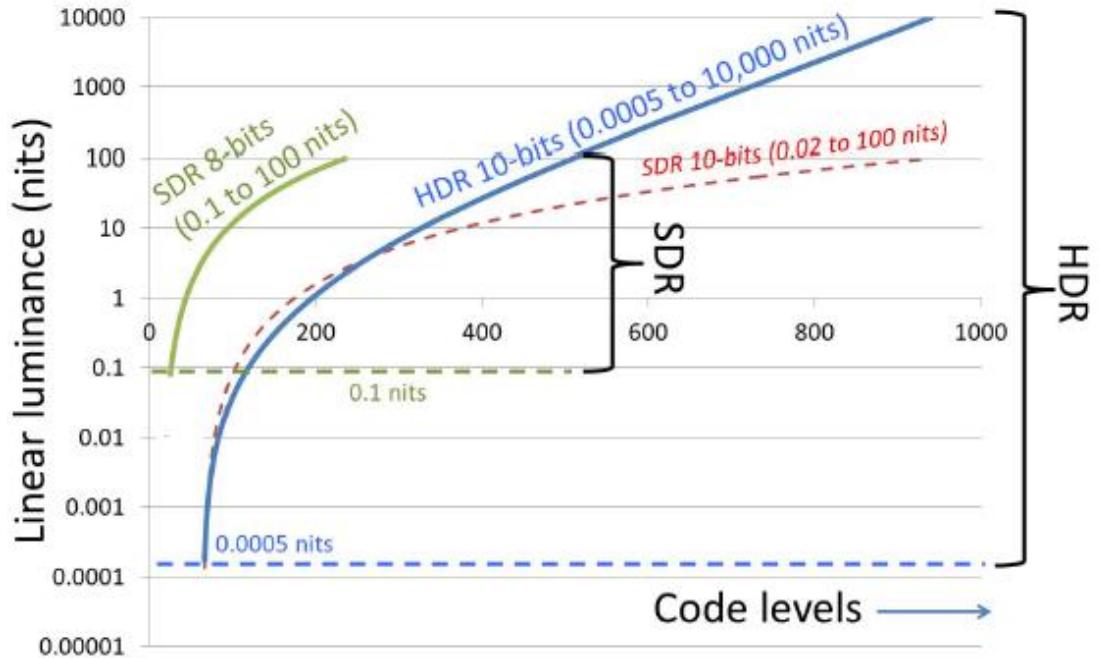


图 11-2 传递函数对比

PQ 传递函数与其他各种传递函数的对比如图 2 所示，该图的纵轴是亮度级别取 \log_{10} 的数值，单位是尼特，横轴是视频的整数样本的编码级别。作为对比，图中的绿色曲线代表了传统的用于 8 比特 SDR 视频的传递函数，蓝色曲线是用于 10 比特 HDR 视频的 PQ 传递函数。如图 5 所示，8 比特 SDR 仅仅覆盖了 0.1-100 尼特的范围，而 10 比特 HDR 仅仅通过额外的两个比特，覆盖了 0.00005~10000 尼特的范围。对于 BT.709 SDR 中能使用 8 比特（16-235）表示的亮度范围 0~100nits，PQ 使用了 9 个比特（120~510），另外一个额外位有效的提供了低于 0.1nits 的较暗处细节，而这些细节映射到 64~119 的量化值范围。而亮度大于传统 SDR 信号限制的 100nits 的部分被映射到 511~940 的量化值范围。需要注意的是，与专业的 10 比特 SDR BT.1886/BT.709 信号容器相比，对于 0.1-10nits 的信号，PQ 曲线密度相对较大，但这是一个权衡，对于 10~100nits 的信号，PQ 传递函数的曲线更加稀疏。

颜色空间 (Matrix Coeffs)

非线性/近似感知统一的编码视频信号的实际域，从由上述传递函数和颜色原色的组合指示的非线性/感知统一的主视频信号导出，最常见的衍生色彩空间类似 Y'CbCr。除此之外，AVC 和 HEVC 认为编码的视频信号是原始信号，例如 4:4:4 RGB 或 XYZ，原始信号的编码在专业应用中是很常见的，例如相机主机或存档视频的夹层压缩。为了能够给出像 Y'CbCr 空间和 4:2:0 的采样比这样更加能量紧凑的色差空间同工的一些比特率上的优化，HEVC 第二个版本添加了交叉分量预测（Cross-component prediction, CCP）工具，基本思路是在解码重建的循环中，将原始信号分解为类似于 Y'CbCr 这样的不相关的几个分量，并用于预测。因此，CCP 是编解码器内部对 AVC 或 HEVC 编码之前应用于主信号的通常外部色差导出的内部替代方案，然后在 AVC 或 HEVC 解码之后通过逆操作处理得到原始信号的近似。

量化范围(*video_full_range_flag*)

量化范围定义了非线性编码的视频信号范围是否利用了从 0 到 $2^{BitDepth} - 1$ 内所有可用的量化代码级别，或者被缩放到某个子集用以提供足够的用于信号负脉冲的空间和用于信号过冲的空间，用于存储通过从滤波器或编解码器量化产生的噪声。消费市场中的视频应用通常需要的是非全范围模式，例如，10 比特下常常将 0 到 1 之间的实数非线性信号值映射到量化值为 64-940 的亮度范围和 64-960 的色度范围。对比之下，传统的 8 比特视频信号的亮度量化范围是 16~235，色度是 16~240。

像所有 VUI 元素一样，*video_full_range_flag* 是在 HEVC 和 AVC 标准范围内的信息元数据，因此不影响任何内部解码操作，解码器仅需要将完整的动态范围内存储的输出视频和图像剪辑到内部解码图像缓冲区，用作后续解码图像的运动预测参考。解码器不会通过语法元素 *video_full_range_flag*，将视频映射到包含的任何较窄的范围，而是将该 VUI 元素传递到最后用于显示器的处理。

在 MPEG HDR 和 WCG 特设组进行的实验中，以及 Ultra HD Blueray 项目的多个调查中显示，10 比特 PQ BT.2020 的 HDR、WCG 信号的编码比特率是与用于 SDR 编码比特率相同的数据级，以 PSNR 为评价标准的 HDR 和 SDR 保真度也很接近。在测试序列中，HDR 技术主要体现在艺术效果中，例如镜面高光、霓虹灯和分散在黑暗场景中的高对比度光和火焰，实际场景的平均亮度级别被设置为与 SDR 视频中的平均亮度接近。除此之外，AVC 和 HEVC 在具有精细细节的镜面高光和高对比度边缘形状会使用更精细的四叉树划分（比如使用 4x4 块）。与进行传统变换量化对比，PQ 方法进行更精细的划分，有更少的残差和边信息，因此带来的性能损失（开销）很少。

11.2.2 SEI 信息

SEI 是 Supplemental Enhancement Information 的缩写，是用来保证 SDR 与 HDR 之间的互用性的机制的方法，用于实现在接收器或者播放器中实现直接显示或转换显示的过程。色调映射是将 HDR 信号转换为较低动态范围的过程，而反色调映射是它的逆过程，旨在增强输入信号的动态范围和色域。这些过程通常由伴随着主视频流的元数据驱动，在创建内容和分级时生成。HEVC 标准包含几种 SEI 信息，可用于 SDR 和 HDR 之间和正常色域与宽色域之间的正向和逆向转换。

HEVC 标准的附录 D 中给出的 Mastering Display Color Volume SEI Message 对应了第一类型的指定表征内容参数的元数据，这些参数是内容描述性的，与解码信号后的具体处理无关。这种元数据表示了内容生成时监视器的实际色域和动态范围，色域和亮度动态范围的乘积空间被称为“颜色容量”。通常情况下，彩色视频的摄影师进行视频调整的创作端的母带监视器旨在代表接收端的现实目标，在实践中，母带制作显示器的颜色容量是编码的视频信号容器颜色容量的子集。例如，如果分级监视器仅呈现 BT.709 色域，但是信号容器指定了更大的 BT.2020 色域，那么掌握显示颜色容量 SEI 信息中的参数就将指示 BT.709 空间的属性，也就是 BT.709 的原色 xy 坐标和白点，SEI 信息还包含了显示器的最大最小亮度。知道母带监视器的子集范围会帮助到优化编码视频在显示设备上的显示效果，这可能与母带制作中使用的不同。另外，在接收端测量编码视频流实际使用的颜色容量也是可行的，但这种分析会导致额

外的延迟，潜在地需要从开始到最后完整地传递内容以建立内容的可信赖的配置文件。

第二种元数据可以更具体的用于自适应过程，这些元数据嵌入了控制色调映射和反色调映射的参数。他们通常针对特定的后处理过程，例如通过查找表组合的信号转换和经由 3×3 矩阵的颜色空间转换。与上述的描述性元数据相反，这种元数据可以认为是处理控制和参数的元数据，例如图 3 中的元数据就可以被用于保证 HDR 转换得到的 SDR 信号符合本来的艺术层面的意义。这一类别的元数据，HEVC 规定了三种 SEI 信息：tone mapping 信息、颜色重映射信息（CRI）和 knee function 信息。

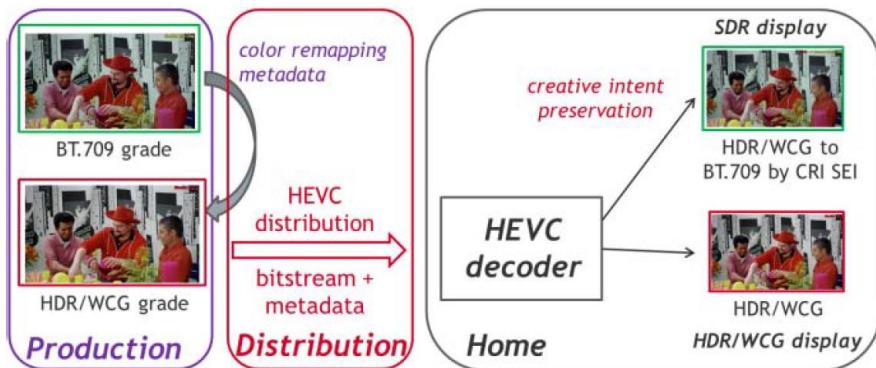


图 11-3 HDR 向 SDR 转换时 CRI 的应用

Tone mapping 的 SEI 信息被设计为携带编码视频序列内的一个或多个色调映射曲线，色调映射曲线用于在显示器上显示解码图像数据，该表达式通过导出在 SEI 信息中转录的一维查找表并将得到的表应用于解码信号的 RGB 分量来进行算术实现。为了提高编码效率，在系统中的一维查找表给出了四种方式：线性模型、S 形模型、分段线性模型和显式查找表模型。选择前两种的方法可以简化通用色调映射行为特征的信令，后两种方法允许更多的任意映射，但是具有更高的比特率上的潜力。在系统的设计中，重要的是要注意通过色调映射信息 SEI 消息标识符标签可以将多于一个色调映射与视频编码序列相关联。这使得系统可以同时支持多个动态范围目标，包括具有比解码的视频数据更大或者更小的动态范围的目标。该信息对应 *tone_map_id* 参数。该参数使得能够存在和维护多个色调映射信息 SEI 消息。而且，应用程序可以使用该参数来识别所包括的色调映射。用于识别色调映射的目的第二种机制是对存在于比特流中的每个色调映射的编码数据位深和目标位深的信令。编码数据比特深度指示 SEI 消息所针对的编码视频序列的比特深度。当该值与比特流中的数据的比特深度相同时，则可以将特定色调图视为使用的候选。或者，当编码数据比特深度值与比特流的比特深度不匹配时，则色调映射图可以被认为是假设的（一种情况是该序列经过了转码或在传输系统内已经进行了色调映射操作）。类似地，色调映射图可以与不同的目标比特深度相关联。该信息指示接收设备或显示器的位深度要求。

颜色重映射（CRI）SEI 信息是用来保证转换的 HDR 和 WCG 与 SDR 信号内容不失真，适用于在生产、后期制作方面制作两个分级版本的内容时。CRI 元数据源自这两个分级版本，并传送到解码器，那么在后处理步骤中可以确保符合艺术意图的转换。重映射模型适用于信号的三种颜色分量信号，例如 RGB 或 Y'CbCr。初始模型基于应用于图像的三个颜色分量样本的三维查找表。后来过程被简化，但是没有明显处理质量上的影响，以适应当前消费电子设备显示能力上的限制。最终版本基于三个一维分段线性输入查找表（LUTin0, LUTin1, LUTin2），随后是一个能够进行颜色空间转换的 3×3 矩阵，紧接着三个一维分段线性输入查找表（LUTout0, LUTout1, LUTout2）。以下等式说明了该模型在颜色样本 RGB 值中的应用：

$$\begin{bmatrix} R_1 \\ G_1 \\ B_1 \end{bmatrix} = \begin{bmatrix} LUT_{in0}[R_{in}] \\ LUT_{in1}[G_{in}] \\ LUT_{in2}[B_{in}] \end{bmatrix}$$

$$\begin{bmatrix} R_2 \\ G_2 \\ B_2 \end{bmatrix} = M_{3 \times 3} \begin{bmatrix} R_1 \\ G_1 \\ B_1 \end{bmatrix}$$

$$\begin{bmatrix} R_{out} \\ G_{out} \\ B_{out} \end{bmatrix} = \begin{bmatrix} LUT_{out0}[R_2] \\ LUT_{out1}[G_2] \\ LUT_{out2}[B_2] \end{bmatrix}$$

这三个步骤的每一步都是非强制的，例如在只有一个输入信号的传递函数对每个颜色分量的应用中，只有第一个一维查找表可以适用。由于输入和输出的位深可以不相同，因此该模型可以支持 SDR 与 HDR 的转换和不同色域之间的转换。

Knee function SEI 信息可以被认为是上面 Tone mapping SEI 信息中第三种可分段线性函数的简化版本，分段点在此定义为拐点。例如图 4 中给出的简单模型，该模型只包含三个拐点，说明了一个典型的 S 形曲线是如何通过拐点函数设计，来实现由一个 1000nits HDR 输入信号映射到预期为 100nits SDR 信号的输出信号。HEVC 提供的 SEI 信息中可以支持最多 999 个拐点的拐点方程。

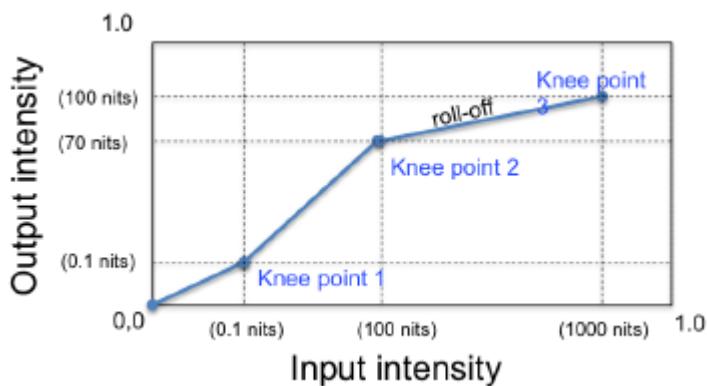


图 11-4 S 形 knee function 示意图

11.2.3 SHVC 中的编码工具——CGS、BDS

SHVC 是 HEVC 编码的可分级扩展功能，为了保持 SHVC 的低执行成本，没有引入新的块级编码工具。为了减少层间冗余，在编码增强层处采用上采样的基本层来重构图像作为层间参考图像。HEVC 的设计通过使用分层时间预测结构的时间子分层特征来支持时间可扩展性，SHVC 提供了额外的可扩展功能，包括位深、空域、质量、隔行扫描到逐行扫描、色域和标准（例如具有 HEVC 增强层的 AVC 基础层）的可扩展性，允许支持具有 SDR 后向兼容性的 HDR 和 WCG 信号。

色域可扩展性（CGS）表示基本层和增强层各自具有不同色域的情况下可扩展使用的例

子。例如，DVB Phase II 定义了一种具有 8 比特 BT.709 色域的 HD 内容作为基础层的场景，以及增强层中 10 比特 BT.2020 色域的 UHD 内容。显然，在两层之间不同色域特性的直接层间预测无法在 CGS 场景中工作，因此 SHVC 开发了 CGS 的特殊编码工具，为 SHVC 发布了两个配置文件：Scalable Main profile 和 Scalable Main 10 profile。其中 Scalable Main 10 profile 支持 CSG 相关编码工具。

CGS 的一般过程如下：首先，将配置的基础层重建图像映射到增强层的色域；然后，将映射图像（按需使用上采样）用作当前增强层图像的解码时的层间参考图像，就其本身而论，映射过程就是色调映射，实际上是作为 SHVC 中 CGS 的编码工具引入的唯一新模块。在 CGS 编码工具开发初期，对采样过程和色调映射过程的顺序进行了深入研究，以找出编码性能与编码复杂度之间的最佳权衡。SHVC 采用的阶段顺序在上采样之前进行色调映射，可以大大降低色调映射的计算复杂度，编码性能上只有不到 1% 的损失。

在颜色映射算法设计中有三个主要的元素：基于不对称三维查找表的 CGS、非均匀色度内容分割和亮度和色度分量之间的相位调整。与不使用 CGS 工具的实验对比，这三个元素的应用在不同的测试方案下，可以有 10% 到 15% 的 BD-rate 上的优化。

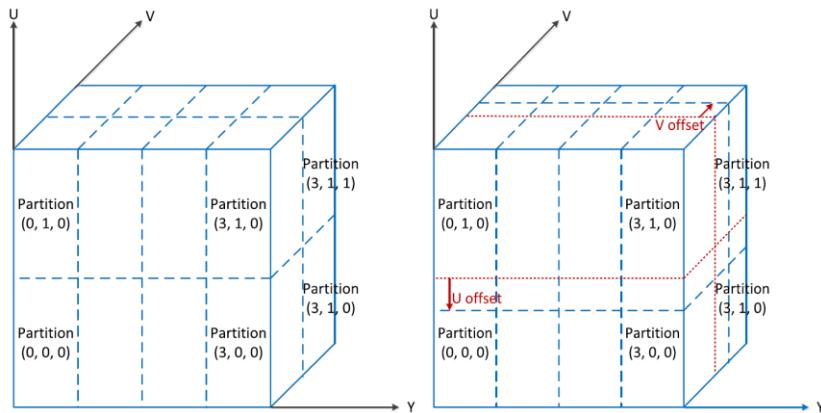


图 11-5 左：非对称 3D LUT；右：非均匀色度分区

不对称三维查找表是在对称 3D LUT 基础上提出的，目的是为了在保持小型 LUT 的同时实现准确的色彩映射，基底层颜色空间被分割成多个相同大小的长方体分区，其中亮度分量比两个色调分量更细，如图 5 中左图所示。对于每个长方体分区，使用以下模型的线性交叉颜色预测，其中 (Y_b, U_b, V_b) 表示基本层色域中的一个像素的三色分量，而 (Y_e, U_e, V_e) 表示相关像素在增强层色域中， a 和 b 表示映射参数。在配置文件 Scalable Main 10 中可以支持多达 $8 \times 2 \times 2$ 长方体分割，由此 LUT 项的最大量可以达到 384 ($8 \times 2 \times 2 \times 4 \times 3$)，相对于对称的 3D LUT 方法来说数量少了很多。

$$\begin{bmatrix} Y_e \\ U_e \\ V_e \end{bmatrix} = \begin{bmatrix} a_{00} & a_{01} & a_{02} \\ a_{10} & a_{11} & a_{12} \\ a_{20} & a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} Y_b \\ U_b \\ V_b \end{bmatrix} + \begin{bmatrix} b_0 \\ b_1 \\ b_2 \end{bmatrix}$$

非均匀速度分区是在了解到色度分量的样本值通常分布不均匀的事实，而提出的。图 5 的右侧图呈现了不均匀色度分割的例子，作为参考，均匀色度分区由虚线表示。为了定位色度分区边界，相对于虚线中的均匀划分的偏移在比特流中用于两个色度分量的信号。

相位调整是为 4:2:0 的色度格式输入视频的特别设计，在这种情况下，色度分量的空间分辨率是在水平和垂直方向上的亮度分量的空间分辨率的一半。此外，色度样本位置通常与亮度样本位置不对齐。HEVC 标准中的 VUI 定义了六个色度位置类型以指定相对于亮度样本位置的色度样本位置。CGS 中的相位调整集中在几种最常见的情况，如图 6 所示。

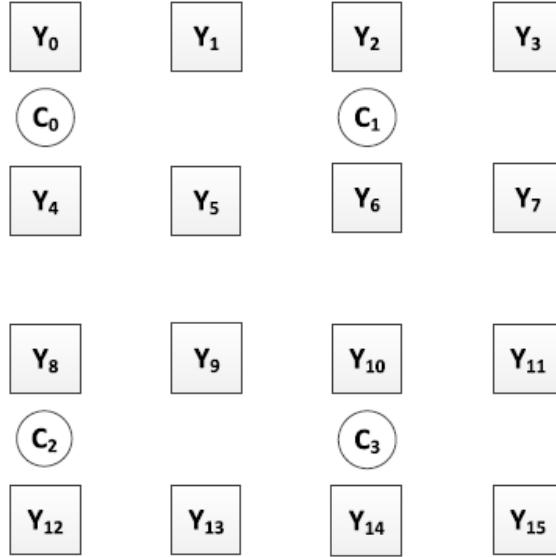


图 11-6 4:2:0 色度模型的常见模式

图 6 给出了一种典型的邻域亮度和色度样本的设定，是由 BT.709 规定的 4:2:0 色度格式。然而，在 BT.2020 中，色度采样位置与亮度样本位置的对应关系是色度格式 4:2:0，亮度样本位置用 Y 表示，色度采样位置用 C 表示。为了将不对称的基于 3D LUT 的颜色映射应用于这三种颜色分量中的任意一种，需要输入一个三元数组(Y, U, V)。为了提高颜色映射处理的精度，在使用交叉颜色分量进行操作之前需要将不同颜色分量的样本位置对齐。具体而言，当计算亮度分量的色彩映射输出时，色度采样值被调整为与它们应用的对应亮度样本位置对齐。将 C 表示为输入色度样本 U 或 V 之一，Y 表示输入亮度信号，根据 Y 所在的位置，使用简单的 2 抽头或 4 抽头滤波器调整 C 的值，方法如下：

$$\begin{aligned} c'(Y_4) &= (c(C_0)*3 + c(C_2) + 2)/4 \\ c'(Y_5) &= ((c(C_0) + c(C_1))*3 + (c(C_2) + c(C_3)) + 4)/8 \\ c'(Y_8) &= (c(C_2)*3 + c(C_0) + 2)/4 \\ c'(Y_9) &= ((c(C_2) + c(C_3))*3 + (c(C_0) + c(C_1)) + 4)/8 \end{aligned}$$

类似的，当计算色度分量的颜色映射输出时，调整亮度样本值以与它们应用的相应色度采样位置对准。也就是说，根据 c 所在的位置，使用简单的 2 抽头滤波器调整 y 的值，方法如下：

$$\begin{aligned}
 y'(C_0) &= (y(Y_0) + y(Y_4) + 1)/2 \\
 y'(C_1) &= (y(Y_2) + y(Y_6) + 1)/2 \\
 y'(C_2) &= (y(Y_8) + y(Y_{12}) + 1)/2 \\
 y'(C_3) &= (y(Y_{10}) + y(Y_{14}) + 1)/2
 \end{aligned}$$

SHVC 有灵活的方式通过有效的信令来支持位深的可扩展性 (BDS)，CGS 中的颜色映射过程以 CGS 参考信号输入和输出的位深，这样也可以支持 BDS。SHVC 中的上采样过程通过调整上采样内的位移数量，根据需要对输入和输出之间的差异进行补偿。支持 BDS 的一种方法是简单地通过相应地制定输入和输出位深来应用组合位深转换和空间分辨率转换的上采样过程，过程很简单，不需要考虑上采样过程的输入和输出之间的颜色特征的变化。这一方法的层间过程在图 7(a)给出，用于 CGS、BDS 和空间组合可扩展性。第二种选择是将位深转换过程和色域转换过程组合在一起。第二种方法的层间过程由图 7(b)给出，与第一种方法相比，第二种方法可以保持足够的转换精度并具有更好的编码性能。

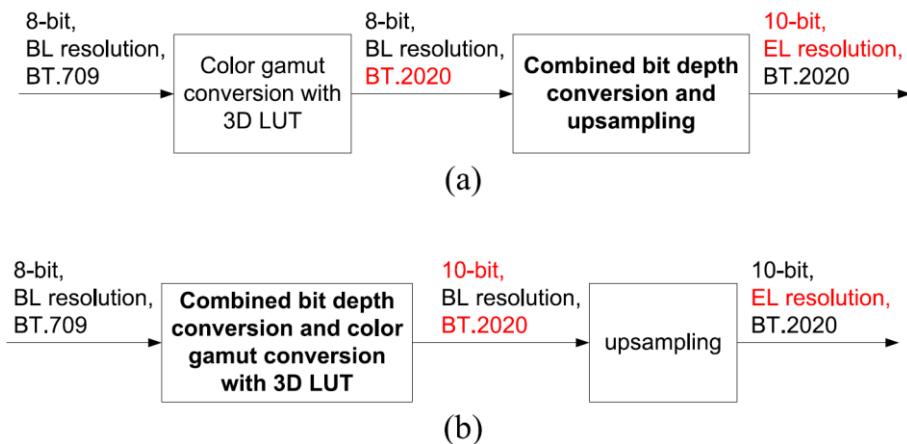


图 11-7 SHVC 层间处理过程

11.3 其他相关编码方案

前面内容介绍了当前 HEVC 标准中包含的 HDR/WCG 相关的编码技术。在本节中，简要介绍其他的一些在 HEVC 标准扩展过程中提出的一些代表性的编码相关技术。

11.3.1 替换颜色空间

在传统方法中，视频编码使用 Y'CbCr 颜色空间，该颜色空间的目的是对 R, G 和 B 颜色分量进行去相关。在 SMPTE ST 2085 中标准化的新 YDzDx 色差空间被添加到 HEVC 版本 3 中的 VUI [1]。YDzDx 通过对 XYZ 原始信号的差分矩阵求导获得，其方式与 Y'CbCr 从 RGB 导出的方式类似。从 XYZ 颜色空间到 YDzDx 空间的变换如下指定 (Y 分量与 XYZ 颜色空间中的 Y 一致)：

$$D_Z = \frac{c_1 \times Z - Y}{2.0} \quad \text{and} \quad D_X = \frac{X - c_2 \times Y}{2.0}$$

c_1 和 c_2 是常数，旨在改善由色度下采样和量化噪声产生的白平衡偏移。在过去十年中，类似的 XYZ 色差衍生空间已被用于数字电影应用中。YDzDx 允许 XYZ 视频通过 4:2:0 设备链，与 4:4:4 色度格式的 XYZ 直接编码相比，提供一点额外的编码性能。电影制片厂 XYZ 主要空间的主要动机是提供更加面向未来的色彩坐标系统，通常用于工作室制作工作流程，根据定义，它完全包含人类视觉的视觉范围，从而避免单独的渲染和分级，也避免了随着时间推移颜色空间的逐步增加。

针对色调线性度优化的其他色彩空间，例如 CIELAB 和 IPT [2], [3]，近年来也在持续针对 HDR 和 WCG 图像和视频编码进行一些研究。

11.3.2 替换编码框架

已经有许多与 HDR 内容的编码有关的工作，主要基于色调映射操作以将 HDR 信号变换为 8 或 10 位整数低动态范围 (LDR) 信号。这些方法可以分为三类。

第一类是基于比特深度的感知域中的 HDR 信号的表示，其高于应用于 SDR 视频的通常的 8 比特（参见[4], [5]）。传递函数的研究属于这一类。该类工作一般使用单层解决方案对得到的转换信号进行编码。

第二类基于调制原理，并使用两个层（参见[6], [7]）。基础层是 SDR 视频，附加层包含调制信息。在解码器侧，将解码的 LDR 内容乘以解码的调制比以重建 HDR 内容。

第三类还使用两个层，其中基于逆色调映射的层间预测来预测来自 LDR 基础层的 HDR 层。预测残差被编码为增强层（参见[8] - [10]）。即使从概念上讲它们可以应用于 HEVC，这些工作最初也是基于 MPEG-2 或 AVC。

除了已经在 HEVC 中指定的技术之外，JCT-VC 也收到几个候选解决方案以用于 HDR 和 WCG 支持。这些提议目标各不相同，比如，比现有的基于 HEVC 的编码链改进编码效率，比特流与传统解码器的向后兼容性，以及解码视频与传统显示器的向后兼容性。大多数提案仅影响编码和解码过程之前和之后的预处理和后处理步骤，因此不影响 HEVC 规范的核心编码工具。然而，其中一些提案可能需要另一个配置文件，而不是 HEVC main 或 main10 配置文件。

结合较小位深度实现的高位深度编码的问题在[11] - [14]中得到解决。常见的想法是重用 HEVC 8（或 10）位设计来编码更高位深度的内容，而无需例如在各种变换级，量化和熵编码级中扩展内部寄存器的位深度。

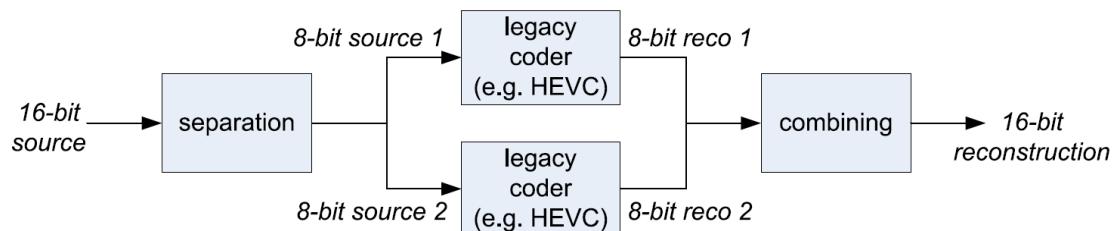


图 11-8 低比特深度编码器进行高比特深度信号编码

如图 8 所示，通过分割高位深度的最高有效位（MSB）和最低有效位（LSB），首先将输入 HDR 图像转换为有限位深度的两个图像（通常每个分量 8 位）样本。考虑到 MSB 和 LSB 之间的差异，编码过程中采用自适应量化调整。MSB 部分被编码为主图像，LSB 图像被编码为辅助图像。该工作使用 SEI 信息以在解码器处组合 MSB 和 LSB 部分。

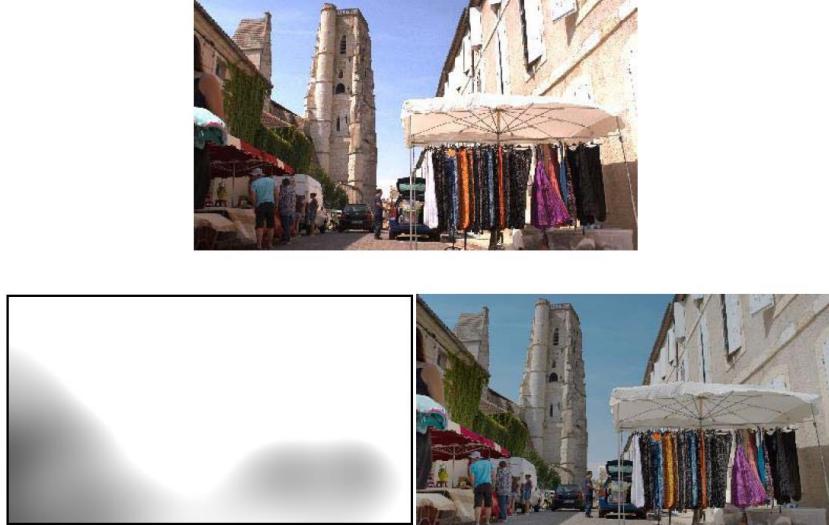


图 11-9 将 HDR 图像（顶部）分解为调制图像（左下）
和有限深度的 SDR 信号（右下）

基于信号调制概念的替代解决方案在[15] - [17]中提交给 JCT-VC。所提出的解决方案旨在提供改进的编码效率和 SDR 后向兼容性。原理是，即使图像全局为高位深度或 HDR，也可以在局部进行低位深度的分解。该解决方案包括编码前的预处理步骤，将输入信号分成两个独立的低位深度信号，一个调制信号（称为调制图像，为输入信号的低通版本）和另一个有限深度信号（SDR 信号）。一旦提取了低频部分，比特深度信号（SDR 图像）对应于剩余部分，如图 9 所示。

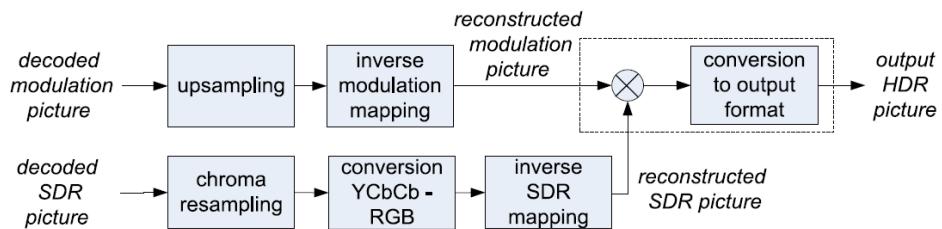


图 11-10 调制方法的解码器端

在解码端，相应的后处理步骤（图 10 中所示）被应用于重构的调制图像和 SDR 图像以生成输出的高位深度图像。一旦解码的 Y'CbCr 4: 2: 0 SDR 信号被解码，则将其重采样为 4: 4: 4 色度格式，然后转换为 RGB 空间。从 SDR 信号（RSDR, BSDR, GSDR）和调制信号 μ 重建 HDR 信号（RHDR, BHDR, GHDR）可以用以下简化的等式表示：

$$R_{\text{HDR}} = f_{\text{Modul}}(\mu) \times f_{\text{SDR}}(R_{\text{SDR}})$$

$$G_{\text{HDR}} = f_{\text{Modul}}(\mu) \times f_{\text{SDR}}(G_{\text{SDR}})$$

$$B_{\text{HDR}} = f_{\text{Modul}}(\mu) \times f_{\text{SDR}}(B_{\text{SDR}})$$

在编码端，调制信号和 SDR 信号都可以（均由线性光输入 HDR 信号产生）通过两个不同的传递函数被映射到非线性表示。必须使用逆映射函数 f_{SDR} 和 f_{Modul} 在解码器侧反转这些操作。组合两个信号的乘法应用于线性光域。完整过程独立地适用于每个像素。

为了在解码端重建 HDR 信号，改提案还提出了新的 SEI 信息（调制信道信息 SEI 信息）以嵌入从解码的调制和解码的 SDR 图像重建 HDR 信号所需的参数，该信息还指示输出重建信号的格式。