

Understanding ISP Pipeline - Noise Reduction

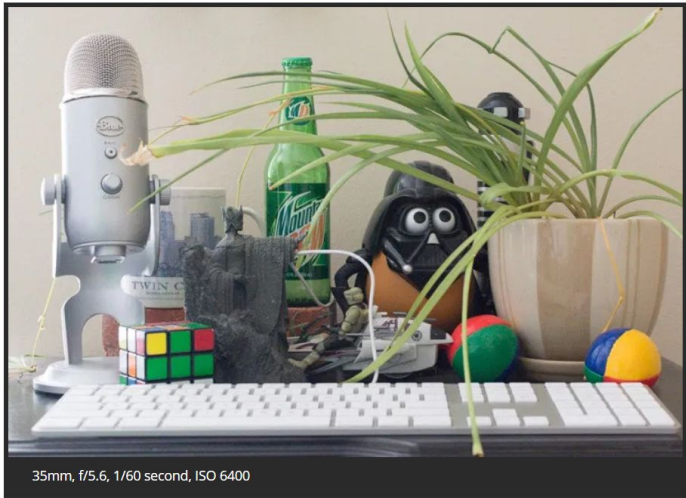
刘斯宁
Camera技术专家

已关注

100 人赞同了该文章

背景

下图是使用单反相机在ISO6400条件下拍摄24MP分辨率图片的预览效果。



看起来似乎还不错，但是如果将图像放大到100%显示，则真实效果是这样的，



原生画面上实际充满了各种颜色杂乱的斑点和斑块，几何线条也不太清晰。这些破坏图像质量的东西就是噪声。

噪声 (Noise)

噪声是图像中不请自来的信号。当相机拍摄一个亮度十分均匀的区域时，相机输出的结果会不可避免地叠加一部分噪声。

相机的输出信号=真实信号+噪声信号。由于真实信号的具体值是未知的，所以只能用多次测量的平均值来代替。用数学公式表示就是，

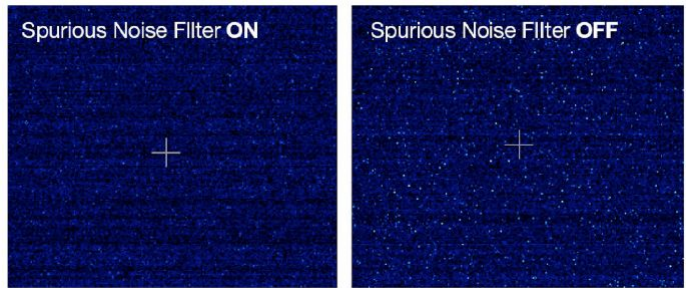
$$x = x_{mean} \pm \sigma_x$$

在数码相机和摄影机产品中，总的规律是光圈越小、（电子）快门越短、ISO越高，图像噪声就越大。图像中的噪声有多种来源，下图列举了几种常见的噪声模式，即

- 固定模式噪声 (FPN)，由于图像传感器的空间不均匀性引起的，硅片本身的瑕疵，失效像素（坏点），像素参数的随机分布，均可构成固定模式噪声

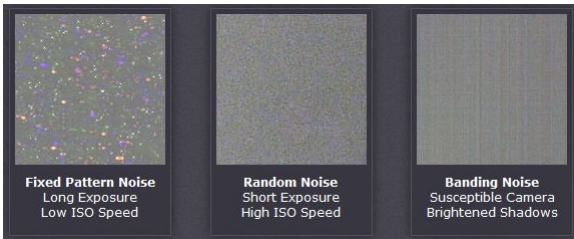


- 随机噪声，由于电子的随机热运动、光信号的统计涨落等多种原因引起的噪声，在芯片中普遍存在的1/f噪声也是一种随机噪声

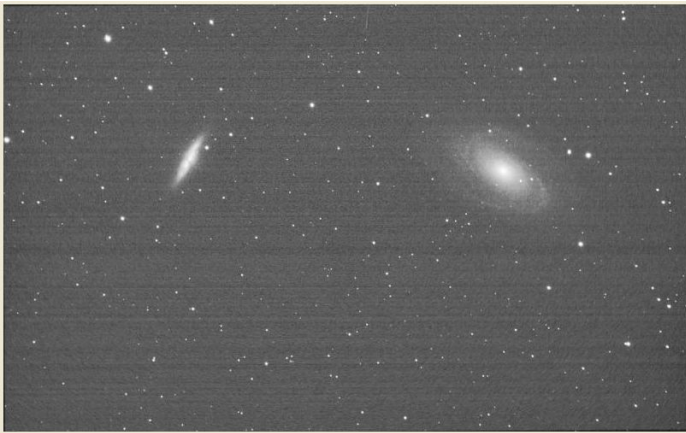


椒盐噪声

- 条带噪声（Banding），有水平条带和垂直条带两种表现形式，有固定模式和随机模式两种来源，其中固定模式来源包括传感器中各晶体管工作参数相对理论值的漂移；垂直条带往往是由于传感器中存在两个ADC器件，其参数不完全一致；随机模式往往由于电源纹波引起的，更多细节可参考本专栏的主题文章

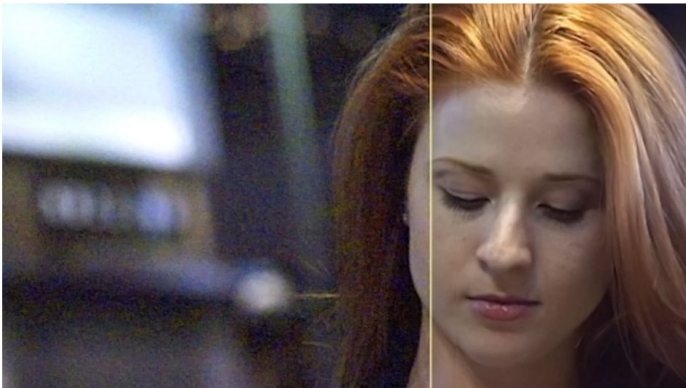


下图是一个条带噪声的例子。



噪声特性

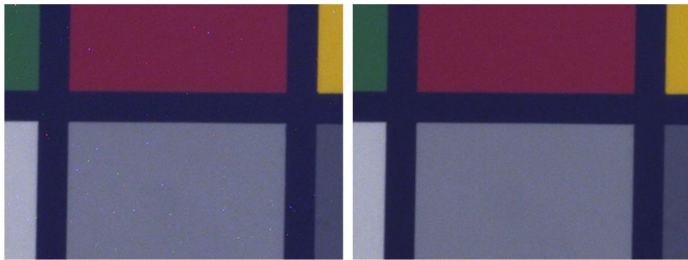
高斯噪声 Gaussian，也称热噪声或约翰逊-奈奎斯特噪声。在电子系统中，自由电子的热运动是一种典型的高斯噪声源，其特点是在工作频段内，噪声的功率谱密度符合均匀分布，而瞬时值符合高斯分布。



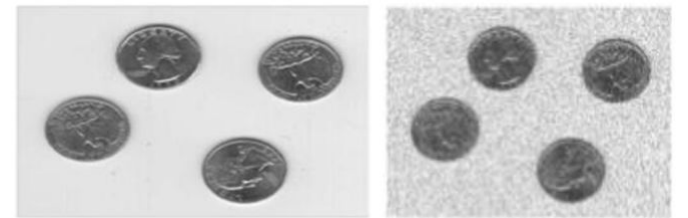
散粒噪声 Shot noise，源于光信号（或电信号）本身的统计涨落，符合泊松分布。



椒盐噪声 Salt and Pepper，取盐和胡椒的颜色，salt指代暗区中的亮点，pepper指代亮区中的暗点。CMOS sensor 中的hot pixel在图像中即表现为椒盐噪声。



斑点噪声 Speckle，根据相干原理进行成像的系统中普遍存在的一类噪声，如超声、雷达、SAR 等。噪声原理是反射表面在波长尺度上是粗糙的，因此反射波存在明显的散射效应（scattering），导致所成的像出现斑点。



数字噪声 Digital，由sensor和ISP的数字增益引起，原因与数字系统的有限精度有关。



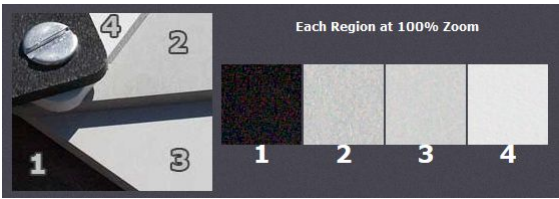
量化噪声 Quantization noise，数字系统对模拟信号进行采样时，由于采样精度不足导致的一种失真形式。目前主流的图像都是采用8比特量化，多数的时候效果是不错的，但是在描述单色平坦区域时（如白墙）容易出现明显的台阶效应。未来的趋势是采用10比特量化，每个（R,G,B）数据能够描述超过10亿种颜色，基本可以消除台阶效应。

Noise Profile

Noise profile 似乎可以翻译成“噪声档案”，其含义是某个噪声源（sensor）在不同工作条件下所表现出来的噪声特性的集合，一般将其表示成增益（gain）的函数，如果这个函数关系可以用某个数学公式描述，则可以进一步用该数学公式的系数定义。

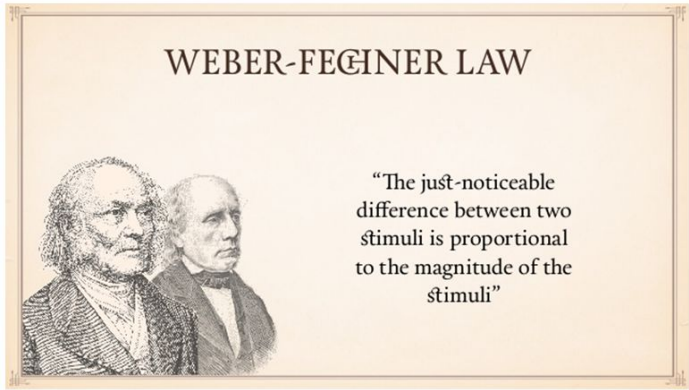
噪声的知觉特性

在一幅图像内部，当噪声的信号幅值相同时，对人类视觉来说暗部的噪声会比亮部的噪声更明显，如下图所示。



P.S. 这是由人眼的生理特性决定的。德国生理学家Ernest Heinrich Weber(1795-1878) 和他的学生Gustav Theodor Fechner 于1834年提出，最小可分辨的刺激变化值与刺激值本身之比是一个常数，用公式表示为

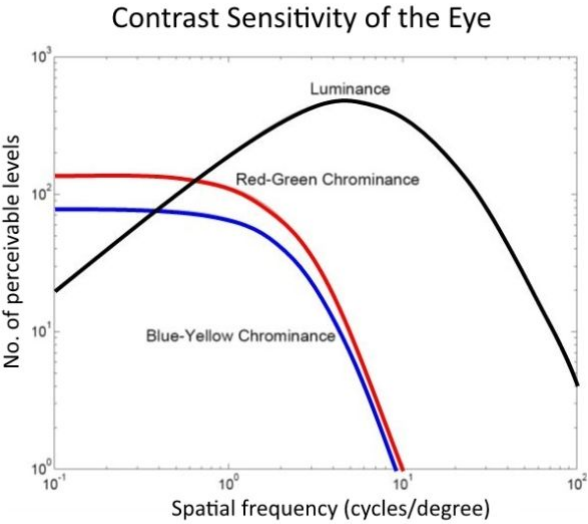
这就是著名的韦伯-费希纳定律。



举例来说，如果50%的受试者能够分辨从100支蜡烛中移走1支蜡烛导致的亮度差别，那么对于200支蜡烛，需要移走2支才能被同样比例的受试者分辨出来。

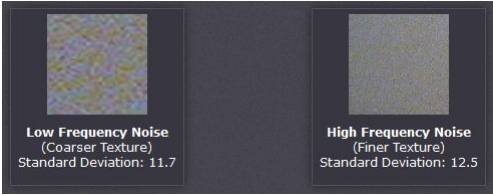
噪声一方面会表现为像素亮度的随机波动，对于彩色图像还会造成像素颜色的随机波动，给图像引入并不存在的颜色，这对图像质量造成的破坏往往更加严重。

关于人类视觉系统（Human Vision System）的研究表明，人眼对亮度和色度的空间分辨能力是不同的，在低频段色度的分辨能力大于亮度，随着频率提高缓慢衰减，频率高于一定阈值后色度的空间分辨能力迅速衰减至零，而亮度空间分辨能力则是先增大再衰减。

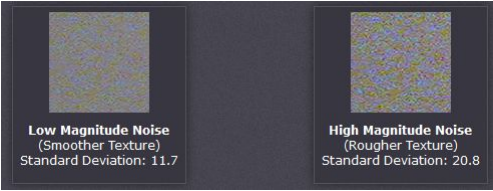




噪声的空间频率高低会影响关于图像质量的主观感受，而且这种影响与人的直觉不一定相符。在下图例子中，右图主要包含高频噪声，在人眼看来其图像质量比左图要好很多，其实右图的方差是12.5，大于左图的方差11.7，左图感觉噪声更大的主要原因是噪声的频率更低一些。



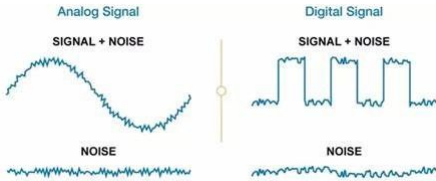
因此，关于噪声的一般规律是，幅度小的高频噪声对主观图像质量影响较小，而幅度大的低频噪声对主观图像质量影响较大。



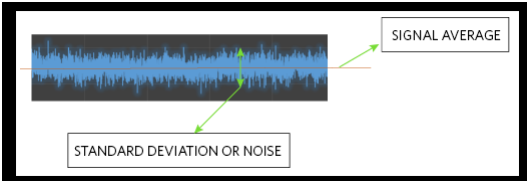
噪声对图像质量造成的影响并不仅限于人眼观察时引起的主观不适，它对后续的图像使用环节也会带来很多负面的影响，比如会引起视频编码的码率上升、人脸识别准确率下降等问题。

信噪比 (SNR)

任何实际测量到的信号都无可避免地含有噪声成分，而评价一幅图像成像质量的标准是信号幅度与噪声幅度的比值，即信噪比。

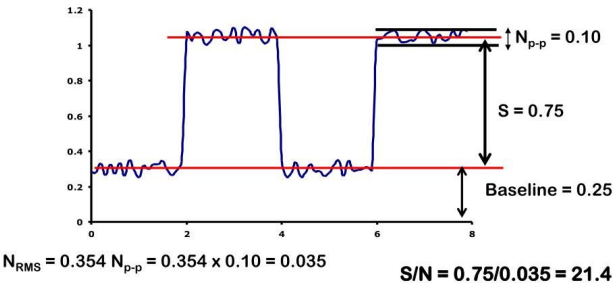


对于一组测量数据样本，一般用均值作为信号的幅值，而该组样本的标准差反映噪声的大小。



Signal-to-Noise Ratio

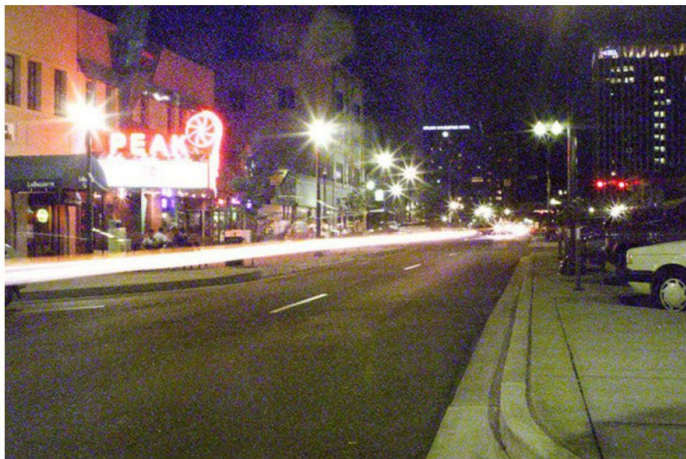
Total signal level nor noise level determine an experiment's ability to accurately detect an analyte. Rather it is the ratio of the two that is critical. The S/N Ratio.



彩噪 (Chroma Noise)



ISP内部会使用YUV空间对像素进行处理，其中Y表示像素的亮度，UV分别表示蓝色和红色的色度。当环境照度不足时，Y、U、V信号的期望值（绝对值）都很小，此时UV分量的随机噪声就变得相当显著，在画面上表现为彩色的噪点或噪斑，如下图所示。



目前主流的CMOS sensor都是使用硅基材料制造的，而硅材料在蓝光波段的光电转换效率较低，所以在低照度条件下拍摄的图像几乎都是蓝色通道的噪声最为显著。

降噪原理

- 1. 对于画面中的平坦区域（homogeneous area），认为像素变化主要是由噪声引起的，可以加大降噪力度；
- 2. 对于画面中的纹理区域（textured area），认为像素变化主要是由纹理引起的，噪声只占较小部分，需要控制降噪力度，尽量保持图像的纹理特征。

为了区分纹理与平坦区域，我们需要设计一种算法，给定一个像素以及以该像素为中心的一个邻域，算法需要计算出该像素的纹理指标（权重）。意法半导体的算法专家们提出了一种基于频域变换的方法，基本思想是将使用8x8大小的DCT变换将小块图像变换到频域，然后对频率分量进行分析，

- 1. 如果DCT变换后非零系数很少，且能量集中在低频，说明是平坦区域，应加强降噪力度；
- 2. 如果两个相邻的DCT单元具备相同或相近的直流分量，则说明是平坦区域，应加强降噪力度；
- 3. 如果某个方向（水平、垂直、对角）上存在明显占优的系数，则说明存在强边缘，应避免降噪；
- 4. 如果某两个方向上存在占优的系数，则说明存在弱边缘，应进行中等强度的降噪；
- 5. 如果很多方向上都有非零的系数，但又没有明显的优势系数，此时可能是纹理与噪声并存的情况，需要一定强度的滤波，削弱高频成分；

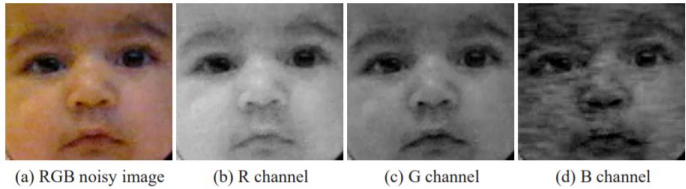
Chroma Noise Reduction in DCT Domain Using Soft-Thresholding

Antonio Buemi, Arcangelo Bruna, Massimo Mancuso, Alessandro Capra, and Giuseppe Spampinato

STMicroelectronics, Advanced System Technology, Stradale Primosole 50, 95121 Catania, Italy

其它研究表明，降彩噪的力度不仅需要考虑纹理特征，还需要考虑画面的亮度。因为彩噪在画面的暗区最明显，所以应将噪范围局限在暗区。

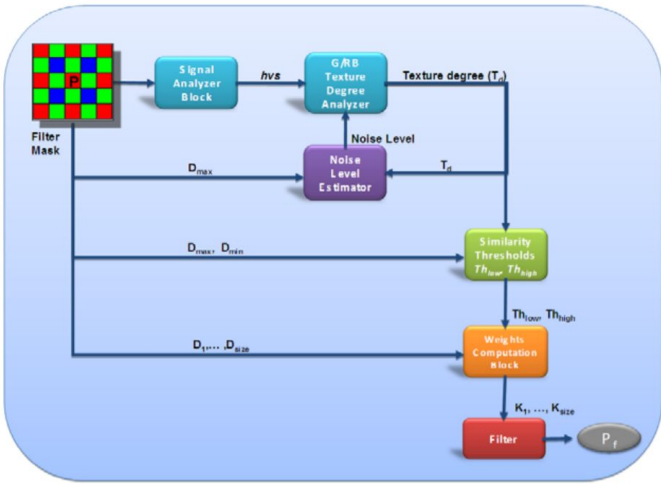
另外，一般的规律是蓝像素的噪声最大（如下图所示），所以可以对蓝色通道使用更强的降噪力度。



综合来看，针对彩噪的降噪算法需要包含下列子算法，

- 纹理强度分析（texture degree / edgeness level analysis）
- 噪声强度估计（noise level estimation）
- 噪声阈值判定（noise threshold definition），区分信号与噪声的边界
- 权重生成（weight generation），决定对降噪和不降噪的倾向因子

下图显示了一种典型的算法流程。

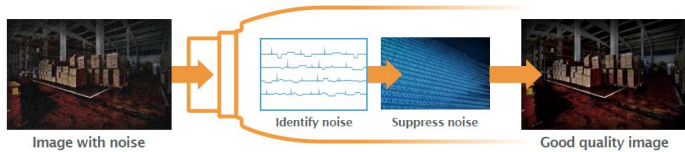


下图给出了一个纹理强度分析算法的例子，左边是原图，右边是纹理检测算法输出的权重图像，白色代表纹理，黑色代表平坦。



空域降噪 (Spatial)

空域降噪是一种2D降噪方法，它只处理一帧图像内部的噪声。降噪算法根据实现原理不同可以分成很多种类型，比如线性/非线性、空域/频域，频域又包括小波变换、傅里叶变换或其他变换。



在camera ISP 中使用的降噪算法需要足够简单、快速、节省内存、适合硬件实现。经典的低通滤波器如中值滤波或高斯滤波虽然足够简单但是容易破坏图像中的边缘，所以主流ISP产品一般会使用某个加强了边缘保持特性的改进版本，如引导滤波（guided filter），双边滤波（bilateral filter）等。

关于双边滤波的原理介绍可参考知友的文章

高斯滤波的基本原理是，在处理当前像素的过程中会用到当前像素附近的像素做加权平均，权重取决于当前像素和周围像素的距离，距离越远则权重越小。双边滤波是在高斯滤波原理的基础上做了扩展，它除了考虑像素距离之外，还考虑当前像素和周围像素亮度的差异，对距离近但颜色差异大的像素会分配小的权重，这样就实现了边缘保持。

小波变换的原理是将信号分解到不同的频带，每个频带保留了一定的空域信息。通过对这些小波系数做阈值处理、滤波或者基于统计建模的处理，再反变换回空域，可以实现有效的降噪效果。

小波变换和双边滤波结合使用可以取得更好的效果。对低频信号做双边滤波，对高频信号做阈值处理，结合之后生成低频信号再做双边滤波，如此反复。这个方法的优势是可以针对不同的频带调节降噪的强度，同时保持了边界，在很多 ISP 中已有应用。

时域降噪 (Temporal)

时域降噪是一种3D降噪方法，它的主要思想是利用多帧图像在时间上的相关性实现降噪。



一种最简单的实现方法是时域均值滤波，即将相邻几帧图像做加权平均。由于累加后噪声的增长速度（根号关系）小于信号的增长速度（线性关系），所以图像的信噪比会提高。这种方法的主要问题在于只适合处理静态图像，如果画面中存在运动的物体则会出现伪影（ghost effect），如下图所示。

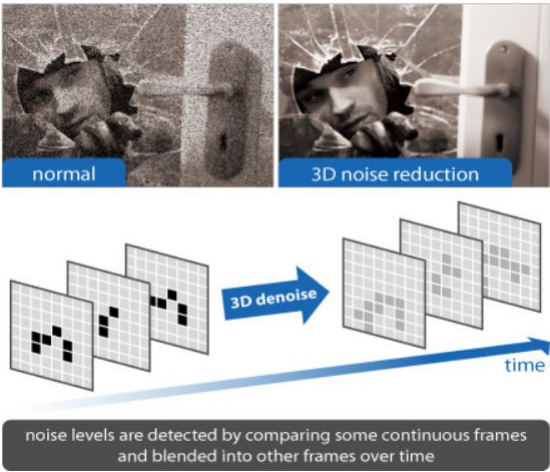


时域降噪技术常伴随运动伪影问题

注意：网上似乎有些人将时域降噪定义为2DNR，遇到时应注意鉴别。

运动适应降噪 Motion Adaptive Noise Filter

对基本的时域降噪进行一些改造就可以得到一种自适应降噪算法。假设图像中坐标(x,y)处的像素值为P(x,y)，新一帧中同位置像素值为P'(x,y)，如果两个像素值的差异小于某阈值，即|P'-P| < threshold，则可以用P代替P'。这种方法对静止的图像效果非常明显，尤其是可以显著降低H.264/265编码码率。

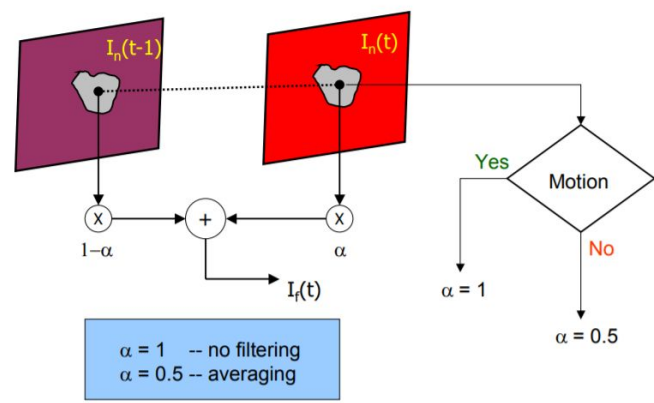


假设 $g(i,j,k)$ 代表有噪声的视频信号序列，其中 (i,j) 代表像素坐标， k 代表时间序列，则 $g(i,j,k)$ 可以表示为真实信号 $f(i,j,k)$ 和噪声信号 $n(i,j,k)$ 的叠加形式

真实信号 $f(i,j,k)$ 当然是不可能知道的，只能用某种方法对其进行估计。一种经典的估计方法是采用以下非线性滤波器，

其中 $f(k)$ 代表第 k 帧图像，它的来源是在第 $k-1$ 帧图像的基础上叠加了 $g(k)$ 与 $f(k-1)$ 的差值成分，即 $\Delta = \alpha(k)[g(k) - f(k-1)]$ ，其中 $\alpha(k)$ 是与 Δ 相关联的阻尼系数。

- 当 $\Delta < \text{thres0}$ 时判断为噪声，采用 $\alpha(k)=0$ ；
- 当 $\Delta > \text{thres1}$ 时判断为真实信号，采用取 $\alpha(k)=1$ ；
- 当 $\text{thres0} < \Delta < \text{thres1}$ 时 $\alpha(k)$ 介于0和1之间。



当然，这种方法的局限也是很明显的，一方面阈值的选取需要能够有效地区分噪声与画面变化，这本身不是一件很容易的事情；另一方面当像素变化超过阈值后此方法就失效了，因此画面前景部分的噪声处于逍遥法外的状态，这也是一个问题。

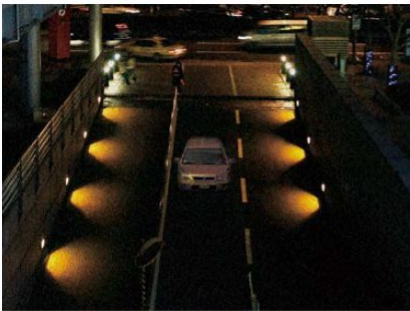
因此，一个理想的滤波器应该同时满足几个特征，即

- 能够有效地区分画面的前景（运动目标）和背景（不动目标）
- 对画面中的前景像素进行空域降噪
- 对画面中的背景像素进行时域降噪

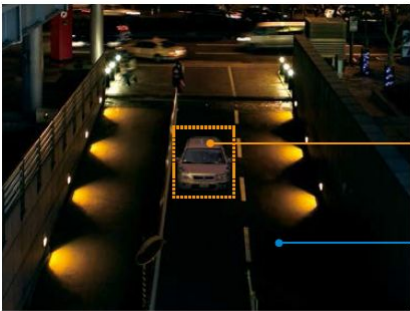
于是就应运而生了STNR技术。

空域时域降噪（STNR）

STNR是一种2D+3D降噪方法，它通过一套算法判别一个像素是属于前景还是背景，被判决为背景的像素将会参与时域平滑，被判决为前景的像素将会参与空域平滑，而判决条件则每一帧都在动态更新，以尽可能保证判决准确性。



Without 3DNR



With 3DNR

注意：网上似乎有些人将时STNR定义为3DNR，遇到时应注意鉴别。

STNR 的核心算法是在对图像中的运动物体进行检测并估计出运动矢量，因此这种技术也称为 Motion Compensated Noise Reduction, MCNR，它依赖于视频编解码中常用的运动估计和运动补偿技术(Motion Estimation & Motion Compensation, MEMC)。

光流法 Optical Flow

假设前一帧时间为 t ，后一帧时间为 $t+\delta t$ 。在理想情况下（ δt 很小，在 δt 时间内图像亮度保持稳定，物体运动速度不大，并且不会突然消失或被遮挡），则前一帧的像素点 $I(x, y, z, t)$ 在后一帧中的位置为 $I(x+\delta x, y+\delta y, z+\delta z, t+\delta t)$ 。

如果假定图像亮度恒定，即可得到光流法的约束方程（constraint equation）

$$I(x, y, z, t) = I(x + \delta x, y + \delta y, z + \delta z, t + \delta t)$$



进一步假定图像中的物体运动速度不大，则后一帧图像可以用前一帧图像的泰勒展开形式表示（只保留一阶近似）

$$I(x+\delta x,y+\delta y,z+\delta z,t+\delta t)=I(x,y,z,t)+\frac{\partial I}{\partial x}\delta x+\frac{\partial I}{\partial y}\delta y+\frac{\partial I}{\partial z}\delta z+\frac{\partial I}{\partial t}\delta t$$

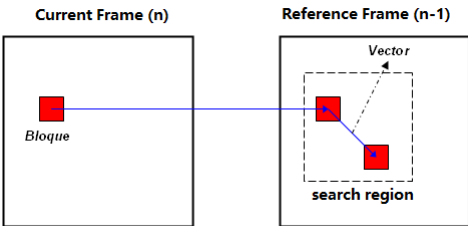
根据约束方程，上式中偏导数之和应为0，即

对于二维图像可忽略z分量，约束方程的最终形式为

通常在一个小窗口内进行光流计算，采用最小二乘法求得该窗口内的平均运动向量。理想情况下，小窗口内的像素应具有一致的运动向量，因此求得的平均向量就代表着物体真实的运动方向。

块匹配法 Block Matching

块匹配的思想是在参考帧（通常是上一帧）中的一个小范围内搜索与当前块(block)最匹配的块，如果确实能够找到则计算出运动向量v。



所谓“最匹配”的块其实可以有很多种不同的评判标准。常用的标准有 Mean Squared Error (MSE) 以及 Mean Absolute Error (MAE) 等算法。

搜索区域一般是根据产品的需求和设计约束而定，常用的有13x13，17x17像素等。搜索区域越大能够检测的运动速度越大，但是所需的算力成本也会呈平方增长。

当图像种存在重复的纹理模式时，光流法和块匹配法都容易失效。在下图的例子中，绿框所示的块状纹理和黄框所示的边缘纹理都容易引起算法失效。而蓝框所示的含直角的块则较容易被算法正确识别。



形态学运算

形态学是基于形状处理图像的一组广泛的图像处理运算。在形态学运算中，图像中的每个像素都基于其领域中其他像素的值进行调整。通过选择邻域的大小和形状可以构造对特定形状敏感的形态学运算。下表是Matlab支持的一些典型形态学运算。

| | |
|-------------------------------|--|
| imerode | Erode image |
| imdilate | Dilate image |
| imopen | Morphologically open image |
| imclose | Morphologically close image |
| imtophat | Top-hat filtering |
| imbothat | Bottom-hat filtering |
| imclearborder | Suppress light structures connected to image border |
| imfill | Fill image regions and holes |
| bwhitmiss | Binary hit-miss operation |
| bwmorph | 针对二值图像的形态学运算 |
| bwmorph3 | Morphological operations on binary volume |
| bwperim | Find perimeter of objects in binary image |
| bwskel | Reduce all objects to lines in 2-D binary image or 3-D binary volume |
| bwulterode | Ultimate erosion |

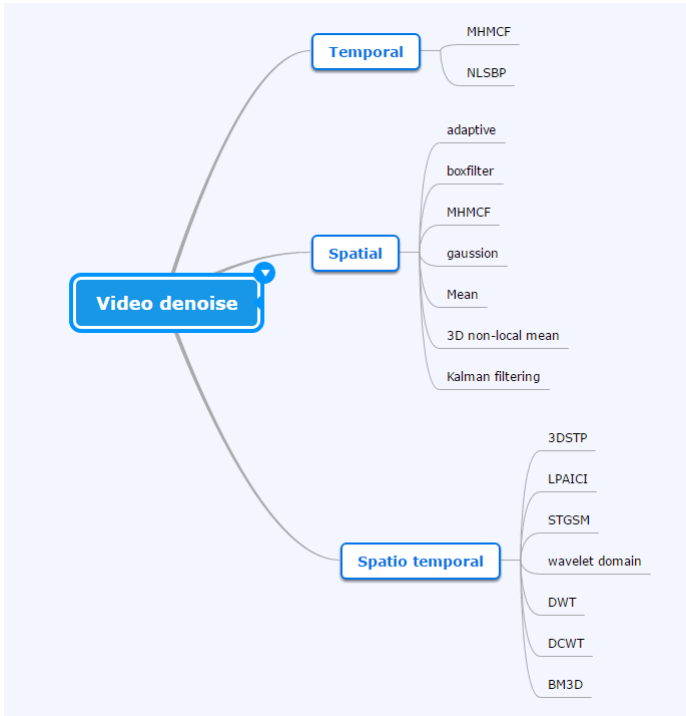
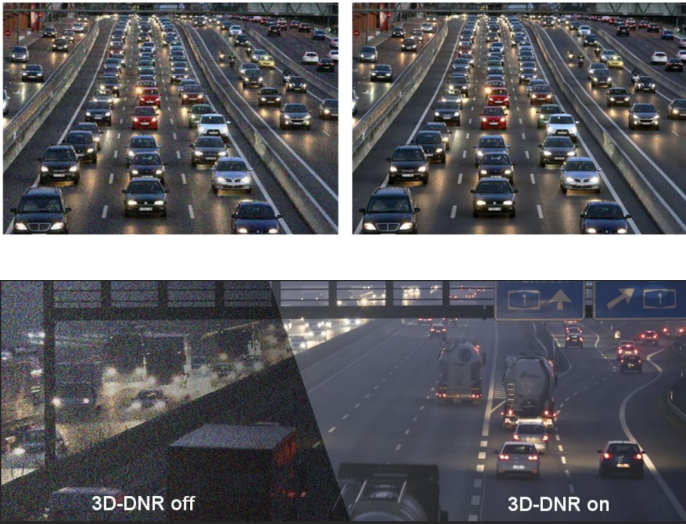


当图像较暗时往往信噪比SNR较低，信噪在噪声中若隐若现。由于噪声的干扰，不论是阈值法还是块匹配法都很容易判断失误，错将背景当作前景，或者反过来。此时，运用形态学方法可以对帮助修正判断，提高判断的准确率，其基本思想是，

- 一旦某个块被判断为前景，则与其相邻的8个块也被无条件判断为前景
- 如果某个块被判断为背景，而与其相邻的8个块中至少存在4个已经被判断为前景，则该块也无条件被判断为前景

这种方法可以显著增加像素被判断为前景的概率，从而增加空域滤波的机会，能够减少运动鬼影的概率。

下面是一些 STNR 效果的例子。



- MHMCF-multi hypothesis motion compensated spatial filter,
- NLSBP- nonlocal sparsity based priori,
- STGSM-spatio temporal Gaussian scale mixture,
- LPAICI-local polynomial approximation and intersection of confidence interval,
- 3DSTP- 3d spatio temporal patches.
- STA3DKF–spatiotemporal adaptive 3d kalman filter,
- MCW – motion compensated wavelet filter,



WF- wiener filter,

GF – Gaussian filter,

MHMCf - multi hypothesis motion compensated spatial filter,

BF –box filter,

STLMMSEF – spatio temporal linear minimization mean square error filter,

OSTA- optimal space time adaptation,

SCBF – Self Cross Bilateral filter,

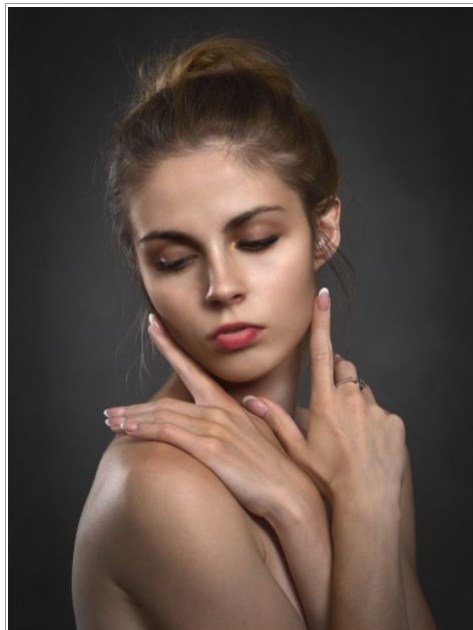
AWAF- adaptive weighted averaging filter.

Keep Natural

如果按照拍摄对象进行分类，摄影可以简单分成风景（landscape）和人像（portrait）两类。

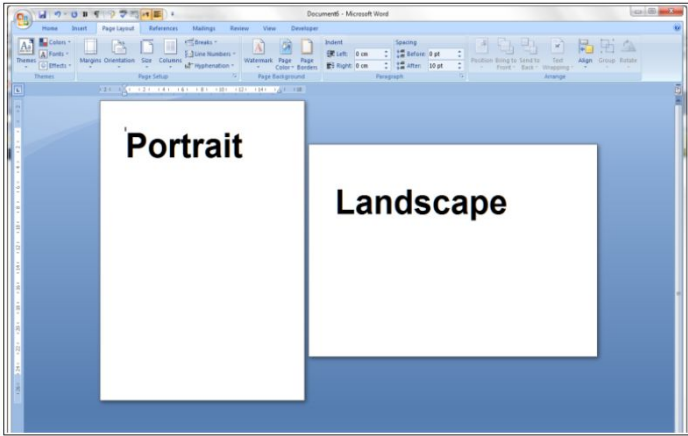


Landscape



Portrait

这个分类是如此经典和形象，以至于人们用Landscape和Portrait的概念来定义文件和照片的布局方向（orientation）。



Landscape vs Portrait

Comparison Table

| Characteristics | Landscape | Portrait |
|-----------------|---|--|
| Meaning | An orientation format where the length of an object, image, or page is longer than the height | An orientation with a longer height than the length. |
| Orientation | It's a horizontal orientation | It's a vertical orientation |
| Used in | Images of a large group of people or things Notices of a horizontal or rectangular shape or design | Images of a single individuals or object Images of tall objects Most documents |

在人像摄影领域，很多人似乎挺喜欢图像中的噪声颗粒感，它们使图像看起来更真实，更有质感。因为没有降噪滤波，图像会保留更多的细节，这也令很多人非常喜欢，所以有些单反相机会特意给出“颗粒感”的模式选项来模拟胶片的图像风格。显然，数码相机上使用CMOS sensor 成像，而胶片使用乳化的银盐颗粒成像，两者的物理特性是完全不同的，所以单反相机的“颗粒感”是使用软件算法添加噪声模拟出来的结果。





Synthetic Grain

After any noise reduction, the image is likely to look a little "off", even if a very light level of noise reduction was used. You may even see some banding in areas like skies and shadow areas, especially if the source was heavily compressed such as H.264 \ AVCHD. However, we can add some "life" back to the image in several ways: mixing back a slight level of the original noise, adding noise via an effect, or overlaying synthetic or scanned film grain over the image, **reduced in opacity until it is barely visible, usually between 15% and 45%**. The intended effect is dithering of the image, which blends together areas that were banding and adds some texture and localized contrast back to skin and other areas with fine details. You can add more grain for creative effect, but only a very small amount of noise \ grain is required for this technique to be effective. For HQ grain, I highly recommend [RGrain](#) (for synthetic grain ideal for invisible dithering) and [CineGrain](#) (for more filmic grain scanned from real film). For the budget conscious, I also created free film grain that you can download [here](#) to experiment with.

| |
|---------------------|
| 35MM LUSH |
| 35MM SMOOTH |
| 35MM EXCITER |
| 35MM COARSE |
| 16MM SOFT |
| 16MM SMOOTH |
| 16MM THICK COATING |
| 16MM OLD STOCK |
| 16MM DAMAGED |
| 8MM CLEAN |
| 8MM NOSTALGIA |
| 8MM DIRT MASTER |
| 8MM INSTANT VINTAGE |
| 8MM BURN IN HELL |



编辑于 2021-07-05 11:15

[图像信号处理器ISP \(Image Signal Processor\)](#) [摄影机](#) [降噪技术](#)