**CV (Mark)**

**Lecture 0**

1. What make up an image? #像素

Pixels – contain the location and brightness(color) information (2D) / besides, contain depth and time information (3D or 4D)

1. 2D images are matrices of numbers #二维图片用矩阵表示

**Lecture 1 Eye and Human Vision**

1. Is human vision good for computer vision?

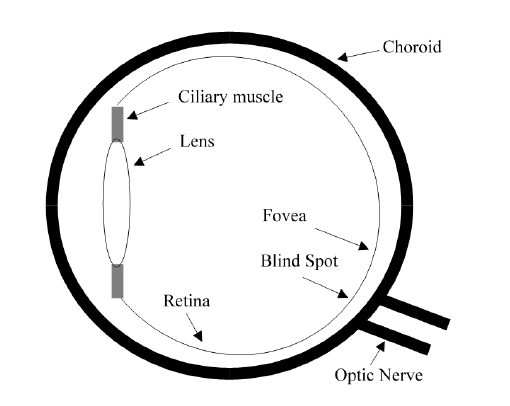
They are functionally similar. There are some suitable properties for developing computer vision techniques, while some are undesirable.

1. Any differences between human vision and computer vision?

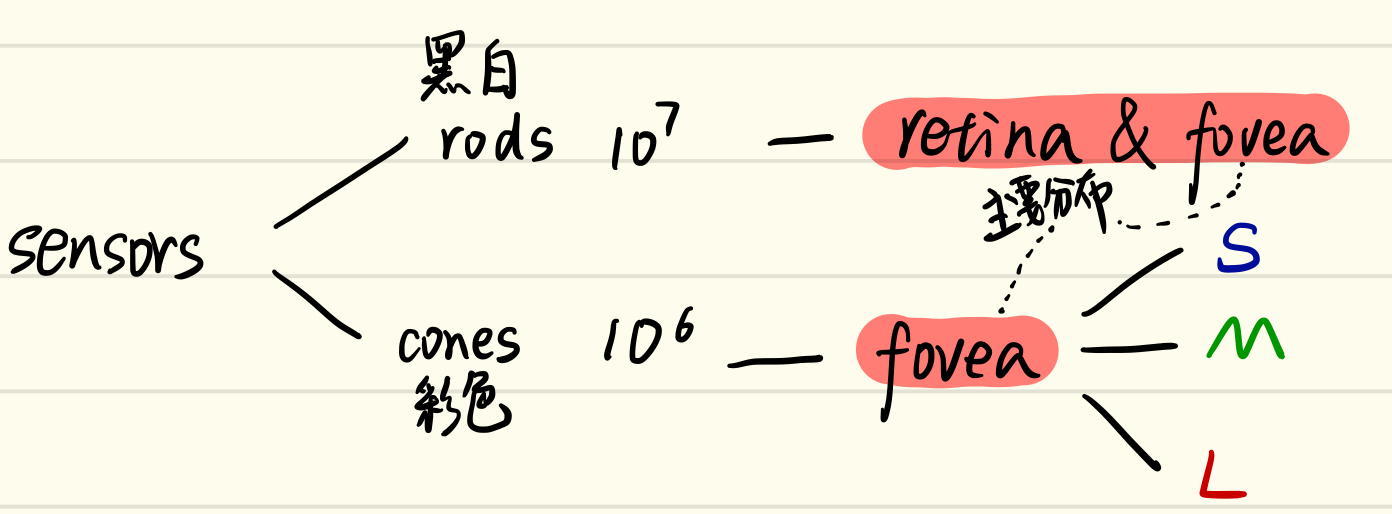
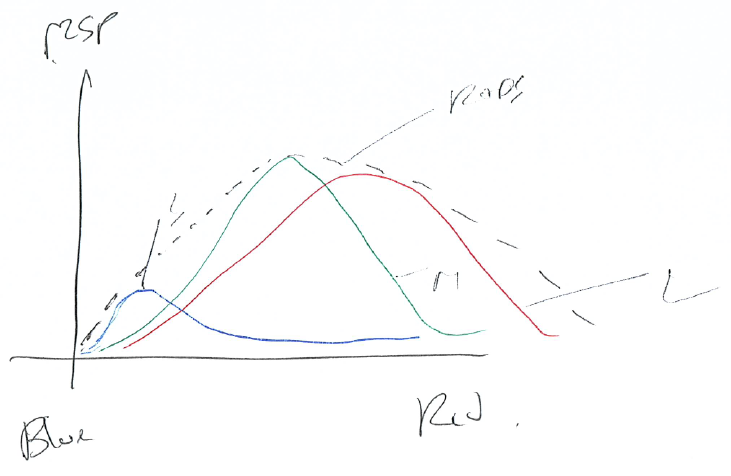
Human vision can distinguish relative distance well but is poor for absolute distance, while computer vision is the other way around.

#第1和2点可以理解为对人类视觉和计算机视觉的整体把握

1. Eye – physical model – evolve for defense and survival
   1. the function of human eye is to form an image, specifically muscles rotate the eye, and shape the lens, to form an image on the fovea (focal point) where the majority of sensors are situated, the blind spot is where the optic nerve starts, there are no sensors there #人眼功能的实现



* 1. sensors – rods and cones #光感元

rods are sensitive to light, so the vision response is actually logarithmic and depends on brightness adaption from dark conditions, where the image is formed on the rods, to brighter conditions, where images are formed on the cones.

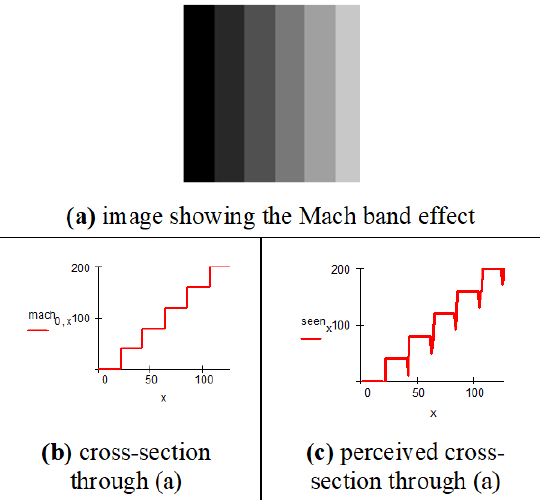
意思就是说rods细胞对光敏感，具有暗视觉的单色能力，所以当光线较暗时眼睛利用rods在retina上成像；而光线强时利用cones在fovea上成像

* 1. Mach bands #马赫带效应

These Mach bands do not really exist but are introduced by your eye. The bands arise from overshoot in the eyes’ response at boundaries of regions of different

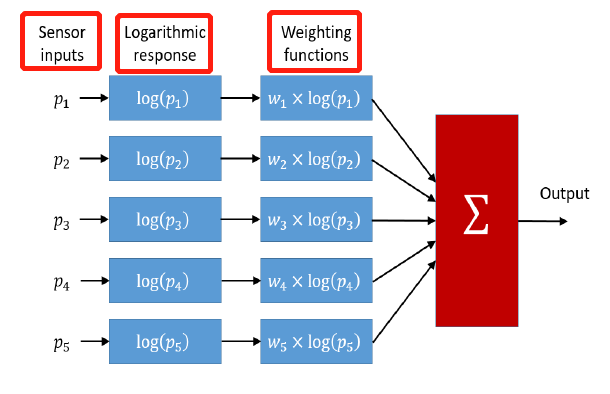
intensity (this aids us to differentiate between objects in our field of view).

意思是不同亮度变化人眼会产生过冲响应



#3.2和3.3是人眼的两个固有性质，这两个性质不利于计算机构建视觉，所以属于undesirable

1. Neural processing – experimental model #combine signals



The response is transformed by a logarithmic function，then multiplied by a weighting factor that controls the contribution of a particular sensor.

怎么通过这个实现edge detection？待讨论，书本原句：If the weights in one half are chosen to be negative, while those in the other half are positive, then the output will show detection of contrast (change in brightness), given by the differencing action of the weighting functions.

1. Processing – psychology #process signals
   1. There are 2 areas in brain to process neural signals – occipital cortex (patterns) and associative cortex (links).
   2. The processing in brain is more than just physical response but also knowledge (prior knowledge of solid geometry).
   3. Human eyes are easily deluded if it’s not been trained.

**Lecture 2 Image Formation**

1. 图片的组成有两个指标m-bit和N-size #m和N
   1. m controls the number of brightness values, gives a range of 2^m values (8-bit is 0 to 255), each bit is 6 dB

m越大，能描述的亮度值越多，图片的对比度越大，噪点越少 (less noise)，携带信息越多，图像变化减缓 (change less rapidly and carry more information)

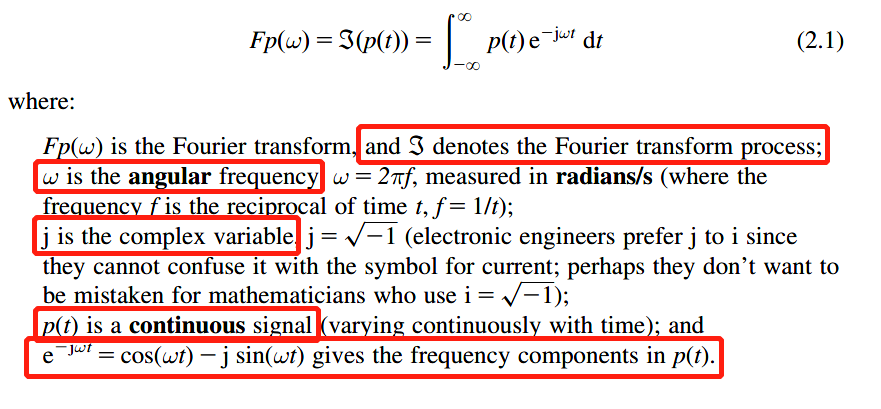
* 1. N is the resolution of an image

N越大，分辨率越大，携带的细节信息越多，但是占用电脑内存增大（通常来说图像的物理尺寸也会变大）

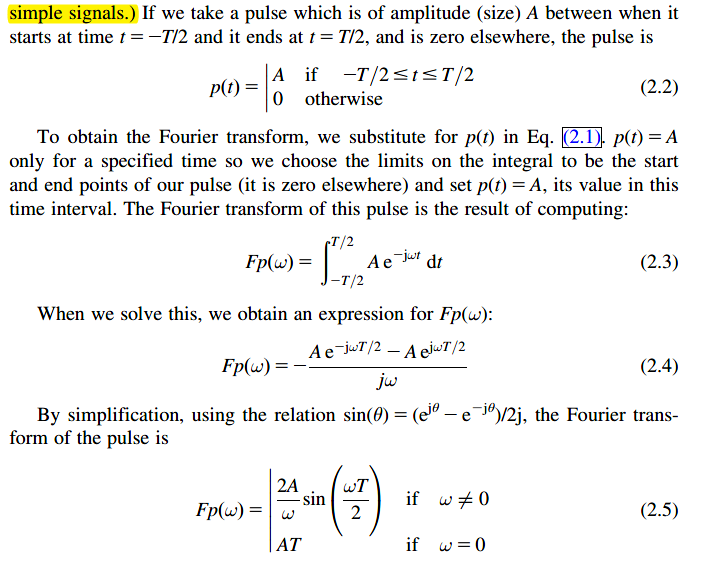
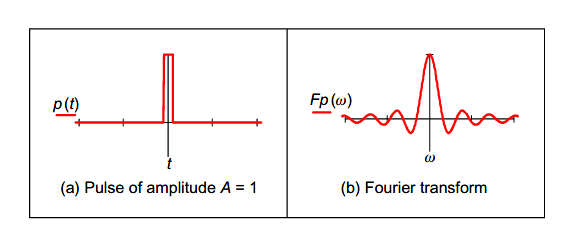
引出问题->Are there theoretical guidelines for choosing it? sampling criterion

（通过信号分析，在频域看采样标准，所以引入傅里叶变换）

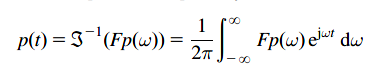
1. Continuous Fourier Transform #连续傅里叶
   1. definition: The Fourier transform is a way of mapping a signal into its component frequencies.
   2. formula:



* 1. a pulse signal: (pp43)

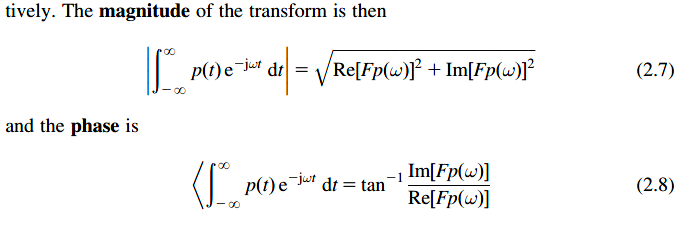
* 1. inverse CFT



Particularly, the zero-frequency component is called D.C. component (直流分量) which represents the average value of the samples.

* 1. magnitude and phase

The outcome of FT is a complex number which is usually represented in terms of its magnitude and phase.



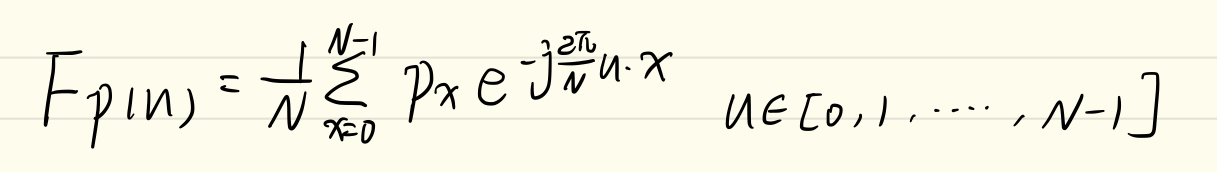
**Lecture 3 Image Sampling**

1. the sampling criterion specifies the condition for the correct choice of sampling frequency. – bad sampling leads to aliasing #why?
2. #how?
   1. Nyquist sampling theory is 1D (the sampling frequency must be at least twice the highest frequency of the sampled signal)
   2. 2D for image sampling (take two pixels for every pixel of interest)

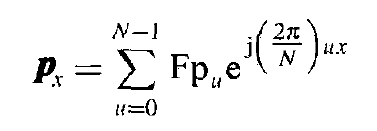
If we obey the sampling criterion, it means we sample signals (or points) at a proper frequency which avoids spectra colliding. If not, we will reconstruct an image with lower resolution where something different can be seen – an alias of the true information which appears by the result of sampling.

Sample points lead to sample frequencies. (引入离散傅里叶)

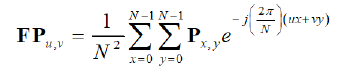
1. Discrete Fourier Transform #离散傅里叶
   1. 1D DFT #一维离散傅里叶
      1. formula:

(与书本有出入)

* + 1. inverse 1D DFT

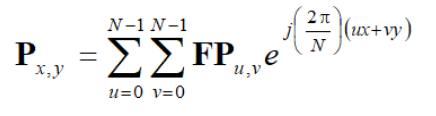


* 1. 2D DFT #二维离散傅里叶
     1. definition: This is a transform of pixels (sampled picture points) with a 2D spatial location indexed by coordinates x and y. This implies that we have two dimensions of frequency, u and v, which are the horizontal and vertical spatial frequencies, respectively.
     2. formula:

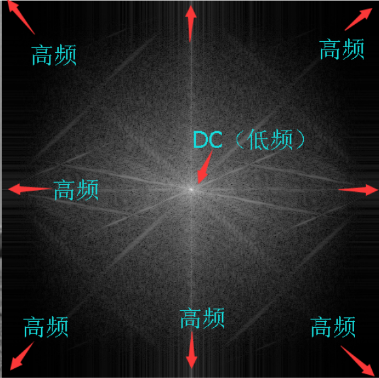


Px,y is original image point

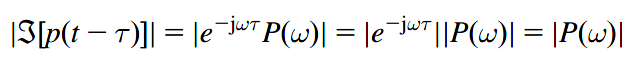
* + 1. inverse 2D DFT

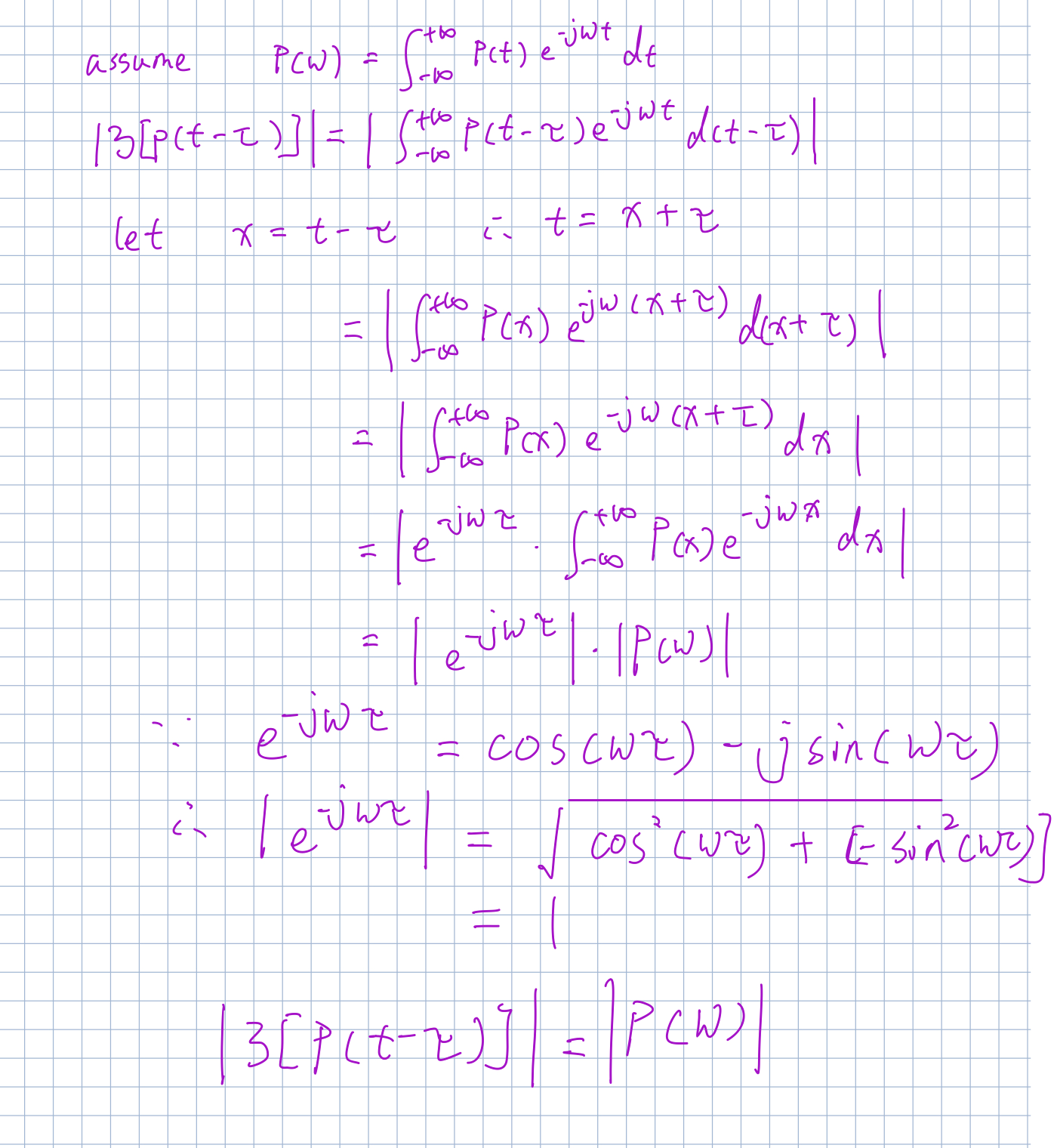


* + 1. analysis: 傅立叶变换是将图像的灰度分布函数变换为图像的频率分布函数，低频对应原图灰度变化平坦的部分（多），高频对于灰度变化剧烈的部分（少），所以低频描述的是图片的轮廓（contour），而高频描述的是细节和边界

（频谱中心化后）

* 1. Properties #离散傅里叶的属性
     1. shift invariance - magnitude of FT does not depend on position #位移不变





advantages: It will allow us to give the same description to different instances of the same shape, but a different description to a different shape. But the phase will change.

application: If the Fourier transform is used to analyze an image of a human face or one of cloth, to describe it by its spatial frequency, we do not need to control the position of the camera, or the object, precisely.

* + 1. rotation - Fourier transform of an image rotates when the source image rotates - the FT rotates as the same amount of as the image #旋转改变

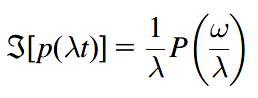
理解：因为二维离散傅里叶保存了每个正弦波的方向，所以原图旋转傅里叶变换也会旋转

application: if the frequency domain properties are to be used in image

analysis, via the Fourier transform, the orientation of the original image needs to be known or fixed

* + 1. frequency scaling - if an image is compressed, equivalent to reducing time, its frequency components will spread, corresponding to increasing frequency #频率伸缩

意义是说当我们增大相机跟物体的距离时，相当于对图片进行缩小，这时候频谱会“往外挪”，就是频率变高（线性变化关系）



* + 1. superposition - we can separate images by looking at their frequency domain components #叠加性

application: given the image of a fingerprint in blood on cloth, by translation to the frequency domain, the Fourier transform of the combined image shows strong components due to the texture and weaker, more scattered, components due to the fingerprint, then we can remove the strong frequencies from spectrum, and conduct inverse Fourier Transform, the fingerprint can now be seen in the resulting image

* + 1. filtering – Fourier gives access to frequency components, so we can use the low frequency components to reconstruct a low-pass filtered image, vice versa. #对图片进行滤波

1. Other Transforms
   1. Discrete cosine transform
   2. Hartley transform

（课堂并没有提及）

1. Applications of 2D FT #应用
   1. separate combined images based on the superposition property of FT
   2. be suitable for recognizing fingerprints and human face based on the shift invariance property of FT
   3. we can reconstructed a low-pass filtered image and high-pass filtered image
   4. we can easily find the Sinusoidal interference through frequency domain and then use filter to eliminate interference
2. Significance of Fourier Transform in Image Analysis: #重要性
   1. 2D discrete Fourier Transform can give the access to frequency components of original images thus we can analyze images and conduct some operations to images.
   2. An image can be reconstructed by its frequencies components through inverse Fourier Transform.
   3. There are many practical applications of Fourier Transform in Image Analysis. (回到5应用)

**Lecture 4 Point Operators**

definition: each pixel value is replaced with a new value obtained from the old one

1. Histogram - the intensity histogram shows how individual brightness levels are occupied in an image #亮度直方图

直方图越往左越暗，越瘦对比度越低，覆盖越广图片对比度越高，图片越清晰

1. Point operators
   1. brighten an image #提升亮度

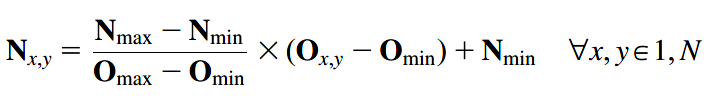


（本质上一个mapping函数）

l represents the brightness of global light, is the minimum value of the output picture brightness range is controlled by a gain, k

k大于1则提升亮度表示范围，增大对比度，l增大则增大图片亮度

* 1. stretch the range of brightness #提升亮度范围
     1. Intensity normalization #直方图亮度正规化

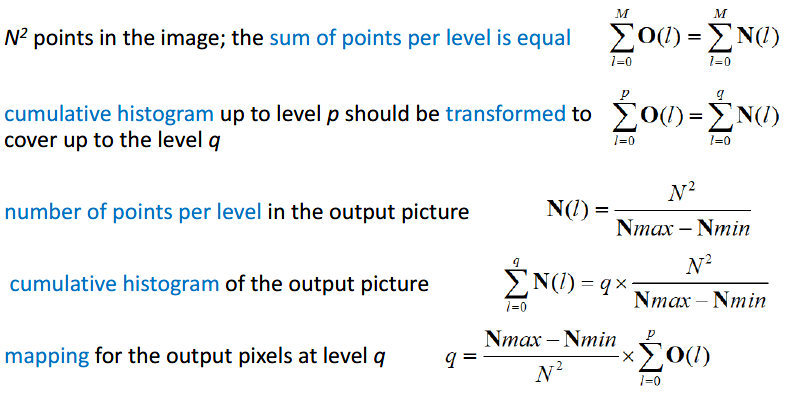


the brightness range changes from [Omin,Omax] to [Nmin(0),Nmax(255)]

（本质上一个mapping函数）

* + 1. Histogram equalization – use for displaying images good for human vision #直方图均衡化

nonlinear and irreversible, histogram equalization aims to change a picture in such a way as to produce a picture with a uniformly flat histogram, where all levels are equiprobable （待讨论）



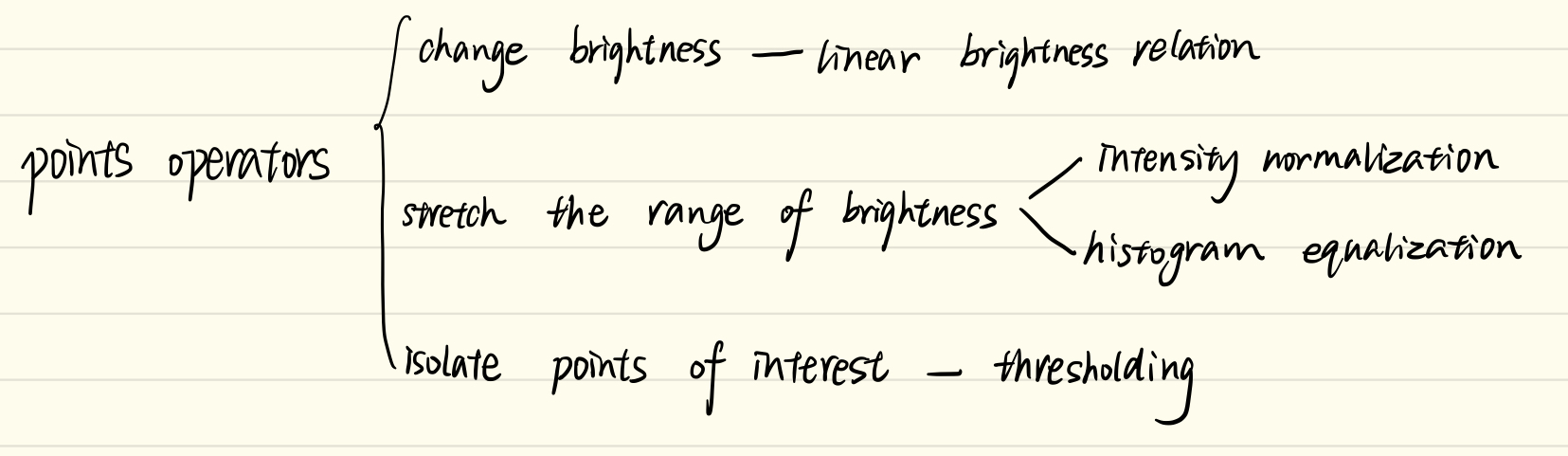
用p跟q实现了一种对应

对图片进行线性亮度变换后它的亮度均衡直方图不会改变，因为线性亮度变化只改变原图亮度直方图的位置跟大小，没有改变形状 （2.1跟2.2.1）

当原图掺杂噪点时（无法完全避免），会改变原图亮度直方图的形状从而影响亮度均衡直方图

* + 1. Comparisons between Intensity normalization and Histogram equalization #两者比较
       1. They both stretch the range of brightness values to has better contrast.
       2. They are both the pre-processing step before image processing.
       3. Histogram equalization is a way particularly suited to human visual analysis.
       4. However, Histogram equalization is a nonlinear and irreversible process which cannot return to the original picture after equalization.
       5. Noise(unavoidable) in the image acquisition process will affect the shape of the original histogram, and hence the equalized version, so histogram equalization finds little use in generic image processing systems though it can be potent in specialized applications.
  1. Thresholding - this operator selects pixels that have a particular value or are within a specified range - provide a way of isolating points of interest #阈值处理

作用

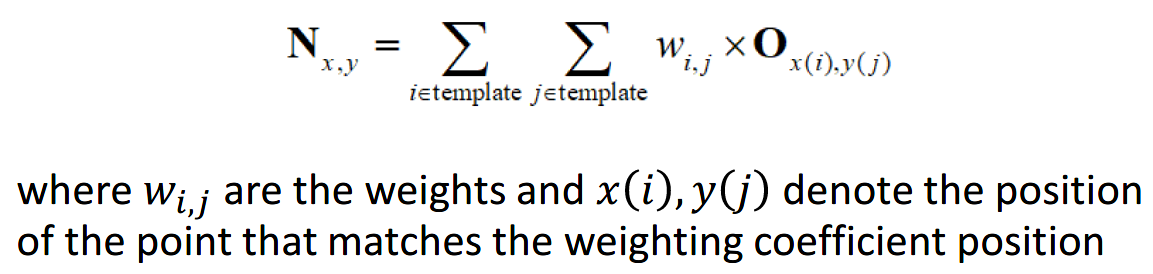


**Lecture 5 Group Operators**

definition: calculate new pixel values from a pixel’s neighborhood by using a “grouping” process

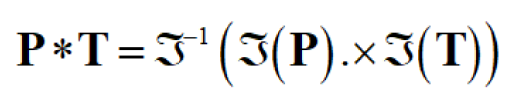
1. Template convolution #模板卷积

border：补0；根据循环位移用另一边界补上（傅里叶变换）；忽略边界



理解的卷积为什么旋转180°是为了对应相乘，如果不旋转的话就是从公式去相乘（图片的左上乘卷积核的右下），老师上课提了不需要记住，一般给的卷积核都是对称的不用旋转

卷积跟傅里叶变换的关系：当卷积核很大时，卷积操作运算速度慢，可以用傅里叶变换去加快速度（因为时域的卷积就是频域的点积）；他们效果等效，只是运算速度的不同



1. Averaging operator #平均算子

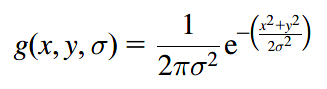
advantage: reduce noise, more smoothing

disadvantage: reduce details that cause blurring

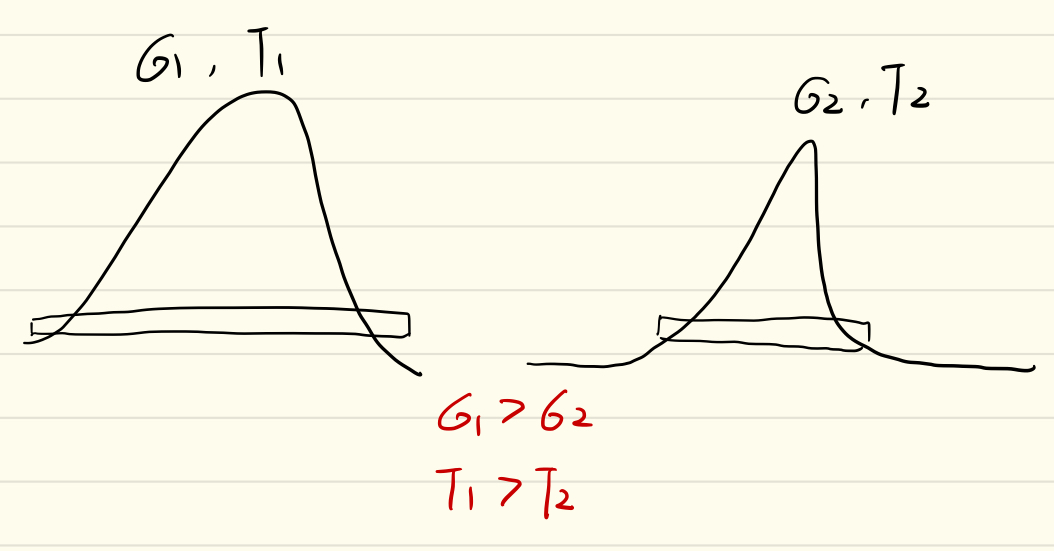
理解：这样对像素值平均化相当于降低图片的灰度差，对应频域里高频信息减少，所以平均算子也是一种对图片的低通滤波，图片的细节信息渐少，保留了大致的轮廓

1. Gaussian averaging operator #高斯平均算子

advantage: remove gaussian distributed noise, optimal for image smoothing

计算高斯卷积核

σ与高斯卷子模板的大小有关，选择适当的σ使得高斯平均模板的边界权重趋于0



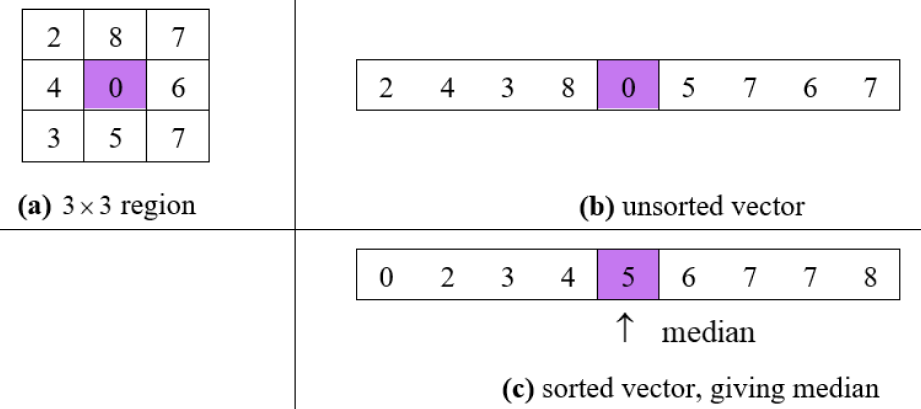
#平均算子和平均高斯算子的比较

两者都是template越大图像越平滑，去除的噪声越多，但是相同size的template，高斯平均算子保留的特征信息比平均算子多

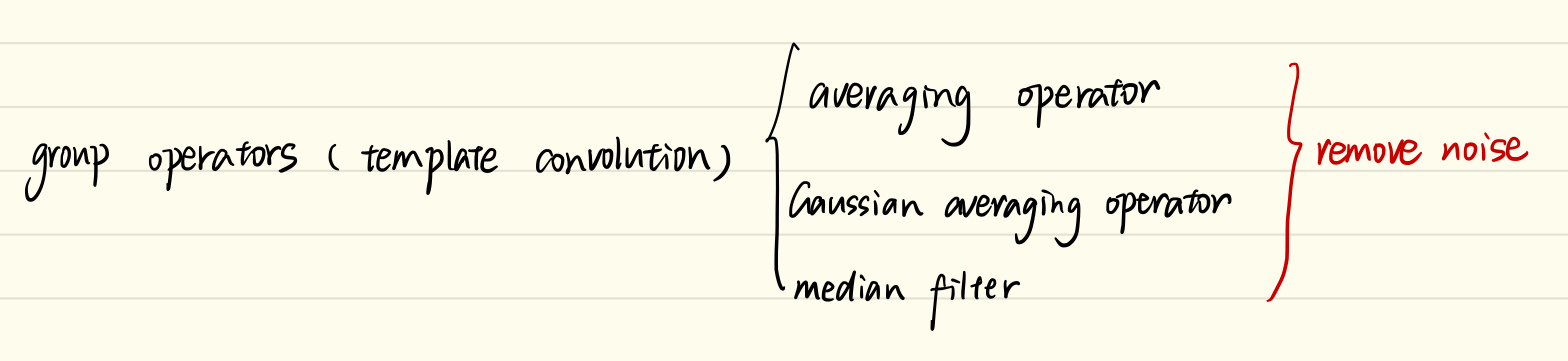
1. Median filter #中值滤波

advantage: remove black and white noise (salt and pepper noise) which come from decoding noise, reserve edge

disadvantage: very slow (no Fourier) 排序



一般的噪声 都是很大或很小，中值跟一般不会受极值的影响



**Lecture 6&7 Edge Detection**

definition: edge detection is a kind of low-level feature extraction

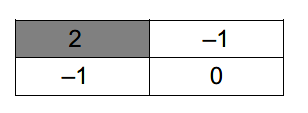
没有考虑位置信息

1. First-order edge-detection operators #一阶边缘检测
   1. Basic operators #基础算子

vertical edge detector

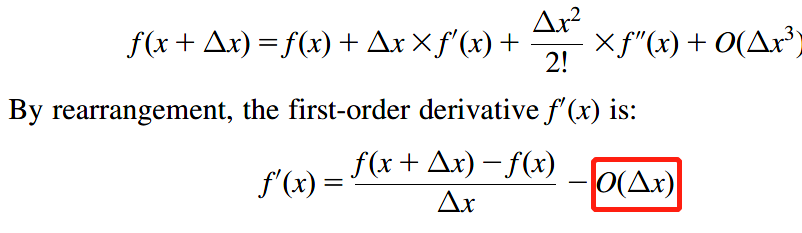
horizontal edge detector



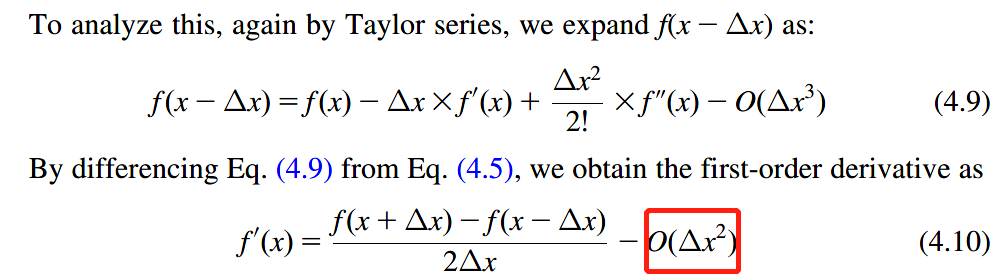


* 1. Improved Basic operators #基础算子改进

泰勒级数的展开证明如果是相邻两个像素点求一阶导的话误差为O(Δx)



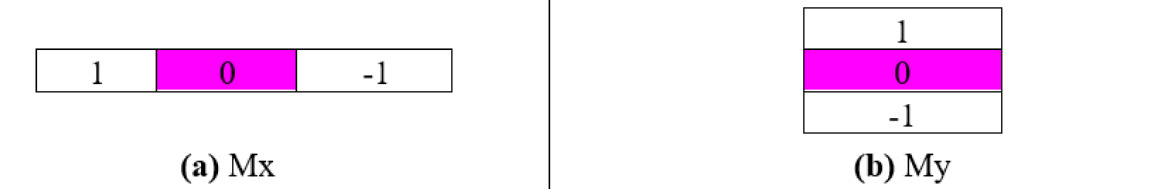
因此可以通过计算相隔一个像素点的两个像素点之间的一阶导，误差变为O(Δx^2)



这样如果Δx<0的话，O(Δx^2)会小得多

相当于计算：

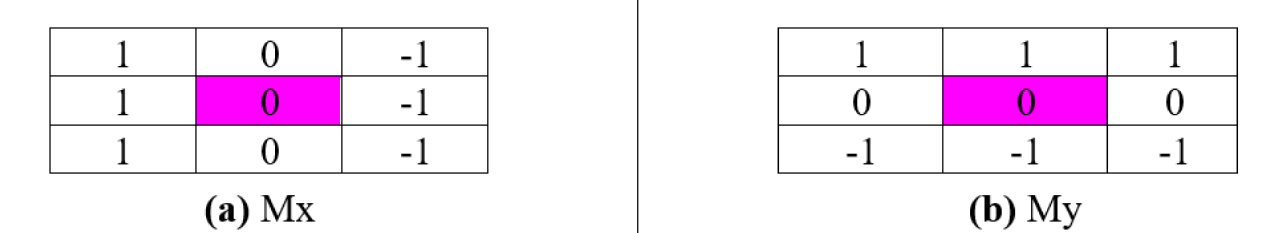




和为1是滤波，平滑 （高斯）， 和为0差分 边缘检测的

* 1. Prewitt edge-detection operator #prewitt

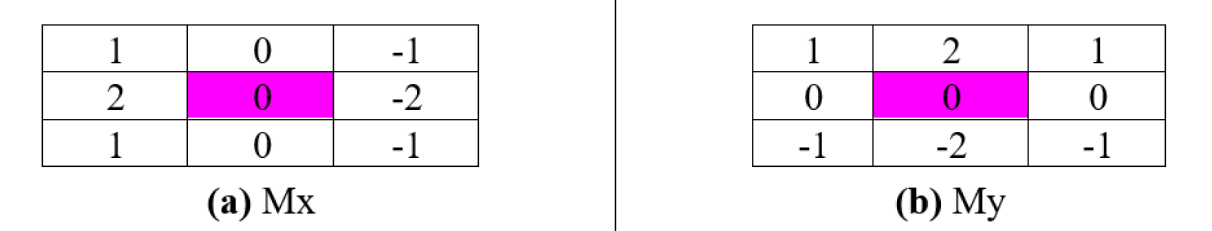
相比于改进版的基础算子，prewitt算子能够在边缘检测的同时平滑图片（因为噪声对边缘检测影响大）



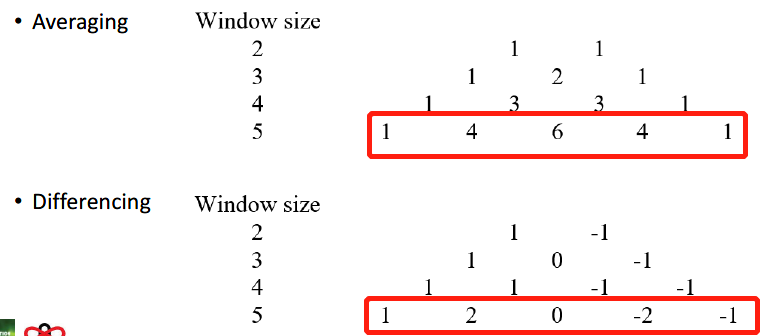
* 1. Sobel edge-detection operator #sobel

相比于prewitt算子，sobel跟它的区别在于平滑操作，prewitt的平滑操作是均值平滑，而sobel类似于高斯平滑（极限情况下）

边界粗 wide edge 高斯，no connective 不连续 (threshold 的缺点)



generalize Sobel – Pascal’s triangle



* 1. Canny edge-detection operator #canny

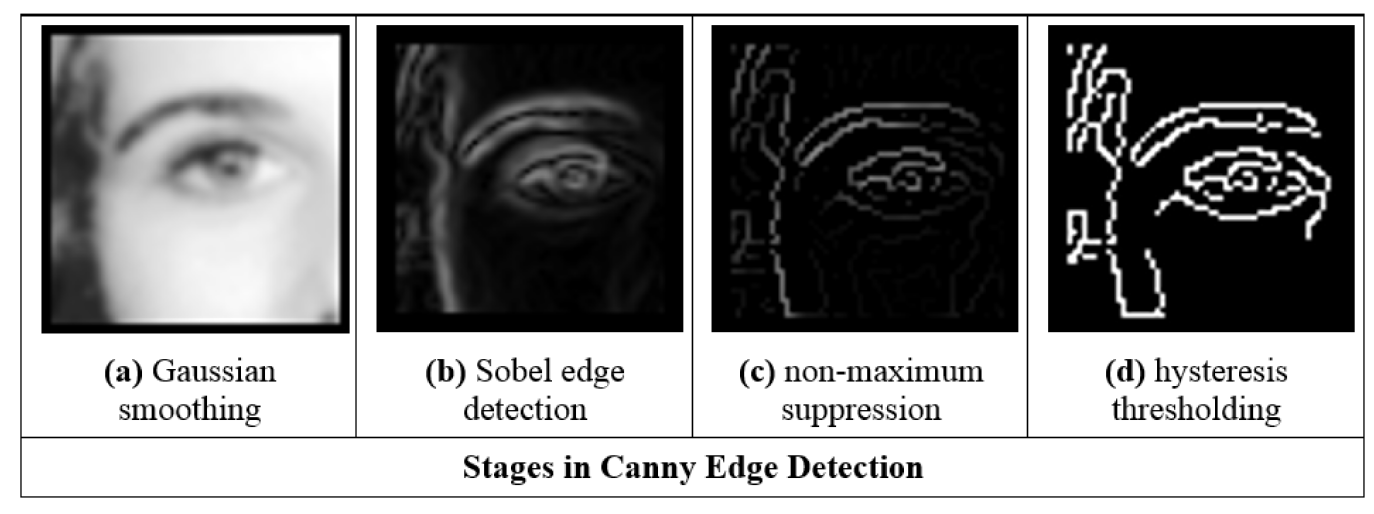
three main objectives:

a) optimal detection with no spurious response – Gaussian averaging

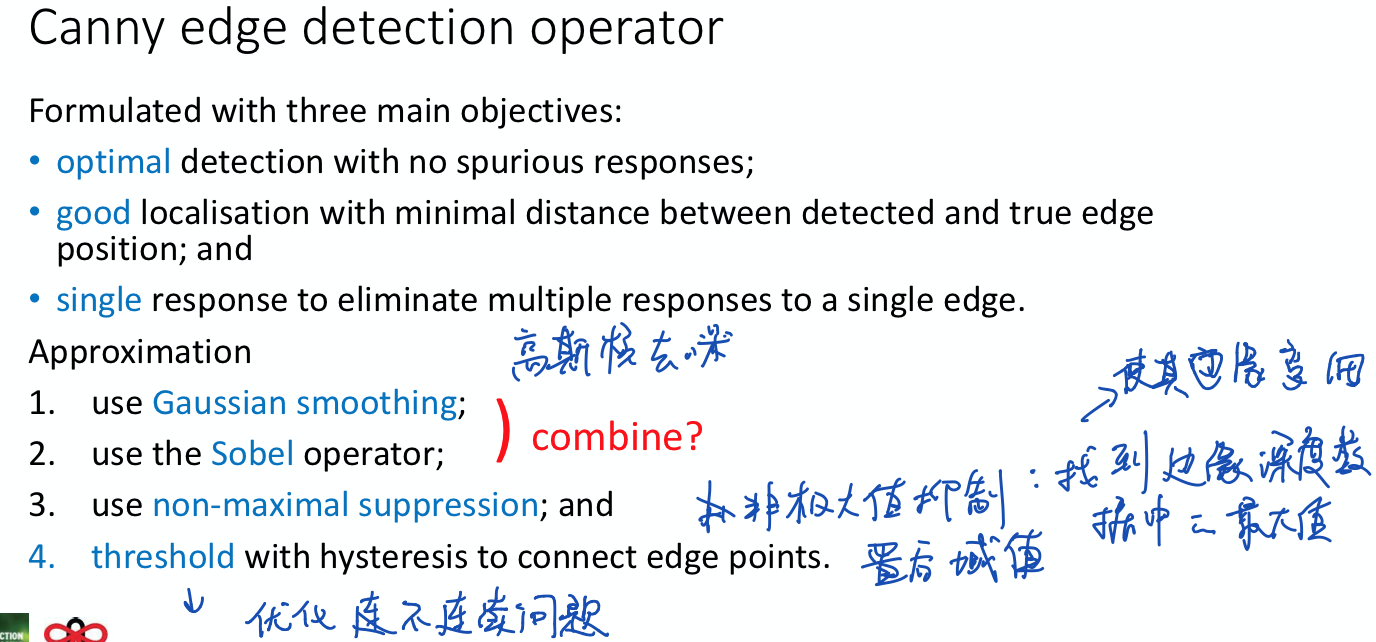
b) good localization with minimal distance between detected and true edge position – non-maximum suppression（非极大值抑制）

c) single response to eliminate multiple responses to a single edge – hysteresis thresholding

意思是canny算子有3个目标，第一是利用最优的平滑去除噪声实现最优边缘检测，第二个是使边缘在距离最近的地方被检测到，第三个是使单个边缘检测点尽量与相邻的连起来



用这4步去近似上面提到的3个目标，(a)和(b)结合起来为一个步骤对应目标1

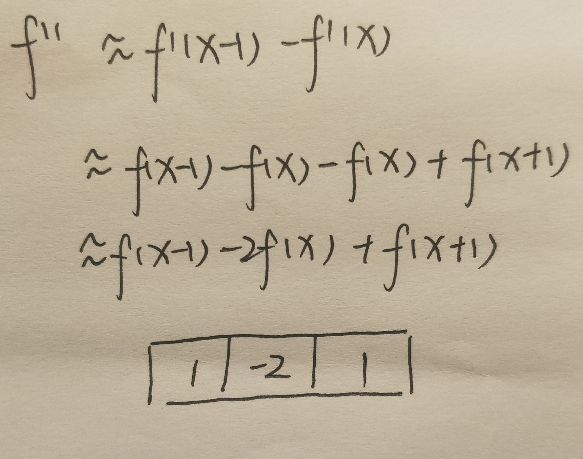


1. Second-order edge-detection operators #二阶边缘检测

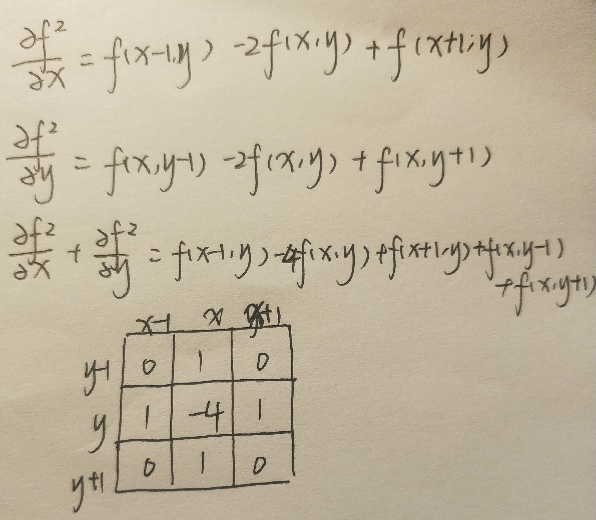
Motivation: differentiate twice to find the zero crossing

* 1. Laplacian operator #laplacian

horizontal dimension: (一次偏导)



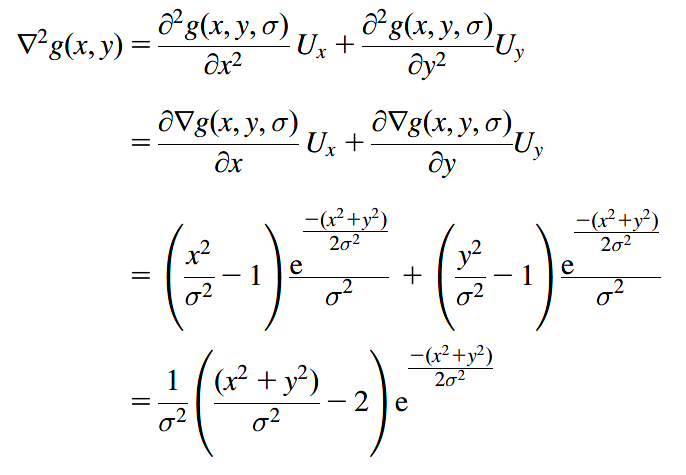
both horizontal and vertical dimension: (两次偏导)



even sensitive to noise ×

* 1. Laplacian of Gaussian operator #LoG

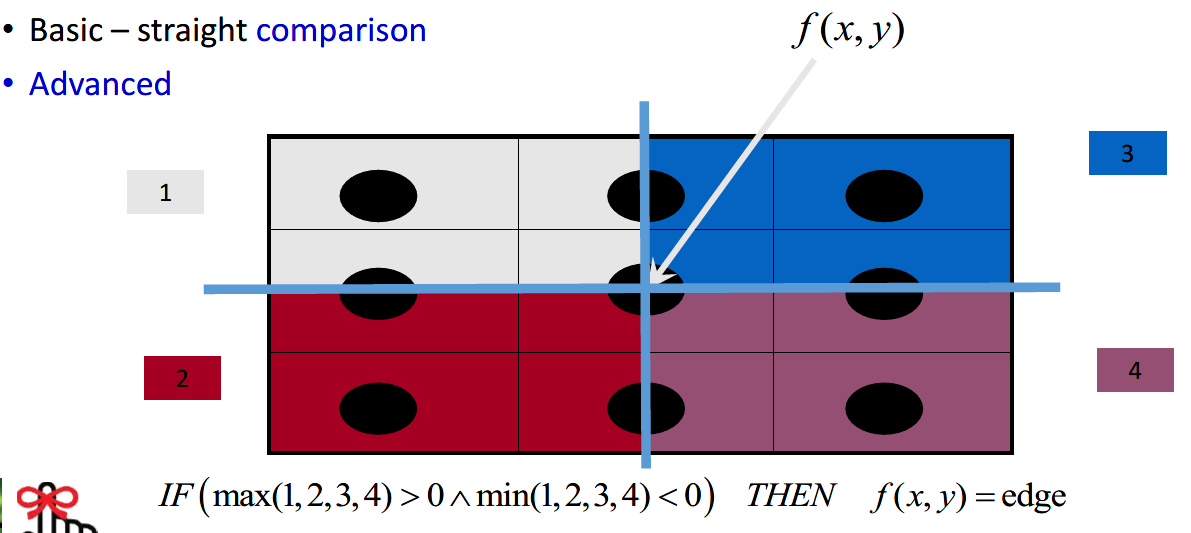
因为拉普拉斯基础算子对噪声十分敏感，一般不采用；LoG是把拉普拉斯二阶边缘检测与高斯平滑相结合的算子



* 1. Marr-Hildreth edge detection #Marr

Marr算子是LoG和过零检测的结合

Zero crossing



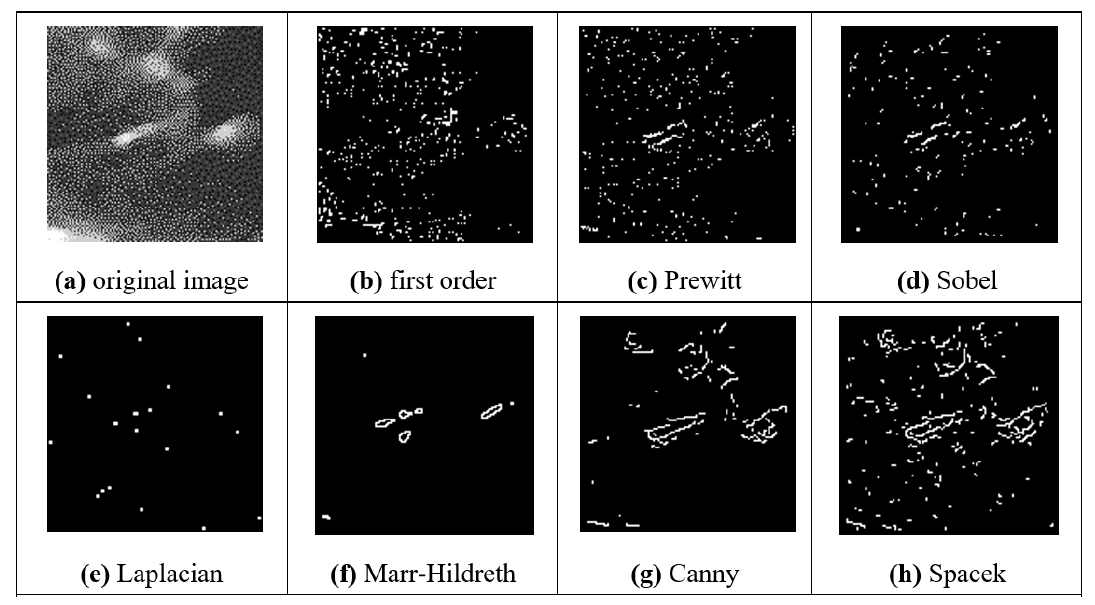
过零检测是因为拉普拉斯不能得出边缘方向，所以用过零检测判断像素是在边缘亮的一边还是暗的一边（？）

讨论：所有边缘检测的卷积核元素和为0，但是平滑的卷积核和边缘检测相结合后元素和为1？

一定可以给出一个封闭的边缘边界

1. Comparisons between two edge-detection operators #一阶和二阶的对比
   1. 灰度均匀变化的地方一阶算子检测不到边缘
   2. 二阶边缘检测算子对噪声敏感，需要对图片进行平滑
   3. 比较b和e，二阶边缘检测出的边缘点个数少，因为二阶边缘强调了亮度变化速度快的点，也就是突出了孤立点和端点，也解释了为什么对噪声敏感 （细）
   4. 比较b，c和d，prewitt采用了均值平滑，识别出来的噪声点数更少；sobel算子的平滑相当于考虑了像素位置的影响（越靠近的像素点影响越大），识别出来的噪声再少一点
   5. 可以看到g的效果最优，因为canny算子有3个步骤，最后结果可以看到检测出来的边缘更细更连续
   6. c，d和g都有水平和垂直两个方向的算子，两个方向梯度平方和开根能得到和梯度的大小，也可以得到梯度方向
   7. 二阶不能得出边缘方向
   8. 比较e跟f，Marr能给出封闭的边缘边界 (ability to provide closed edge boards)而canny不能；

另外一个优点是能避免滞后阈值的递归计算(avoids the recursion associated with hysteresis thresholding)，从而节约了大容量堆栈(massive stack size)



**Lecture 8 Finding shapes**

definition: high-level feature extraction concerns finding shapes and objects in computer

images

1. Thresholding & Substraction #阈值和差分

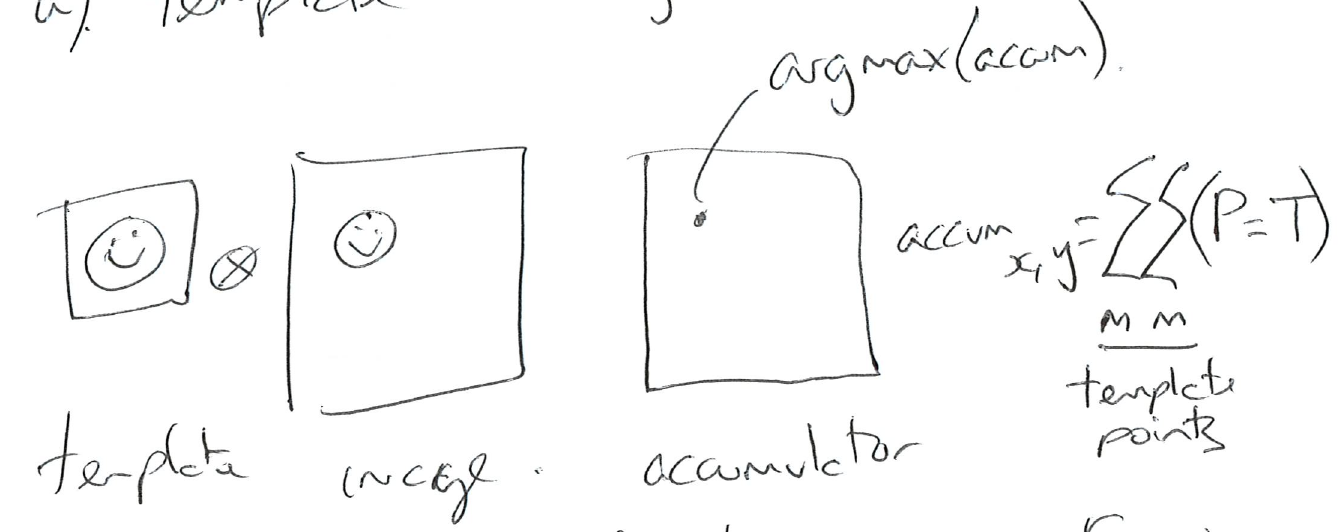
阈值是当光线确定或者识别形状的亮度值确定的情况可以用这种简单快速的方法；差分是识别形状时会受到背景的影响，如果确切知道背景的信息可以提取出物体再识别形状，但是实际中会有噪声；

所以以上两种方法实用性不强

1. Template matching #模板匹配 检测模板

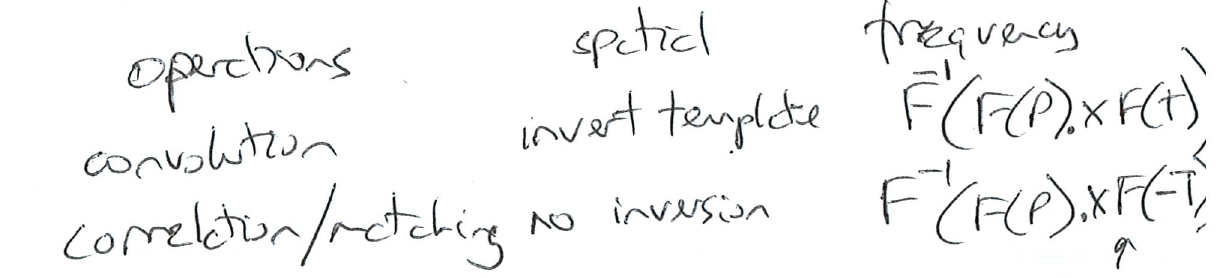
definition: We need to match a template to an image, where the template is a sub-image that contains the shape we are trying to find.

Template 与image 里的图片 有相同的大小 形状 不能旋转



advantage: insensitivity to noise and occlusion 对遮挡与噪声不敏感

disadvantage: when the template becomes larger, it’s slow – Fourier



template需要横纵轴翻转再padding (template 很大 )



disadvantage: 模板匹配具有自身的局限性，主要表现在它只能进行平行移动，若原图像中的匹配目标发生旋转或大小变化，该算法无效

（low-level 不考虑位置 high-level 考虑位置） 边缘检测是检测边缘 low level

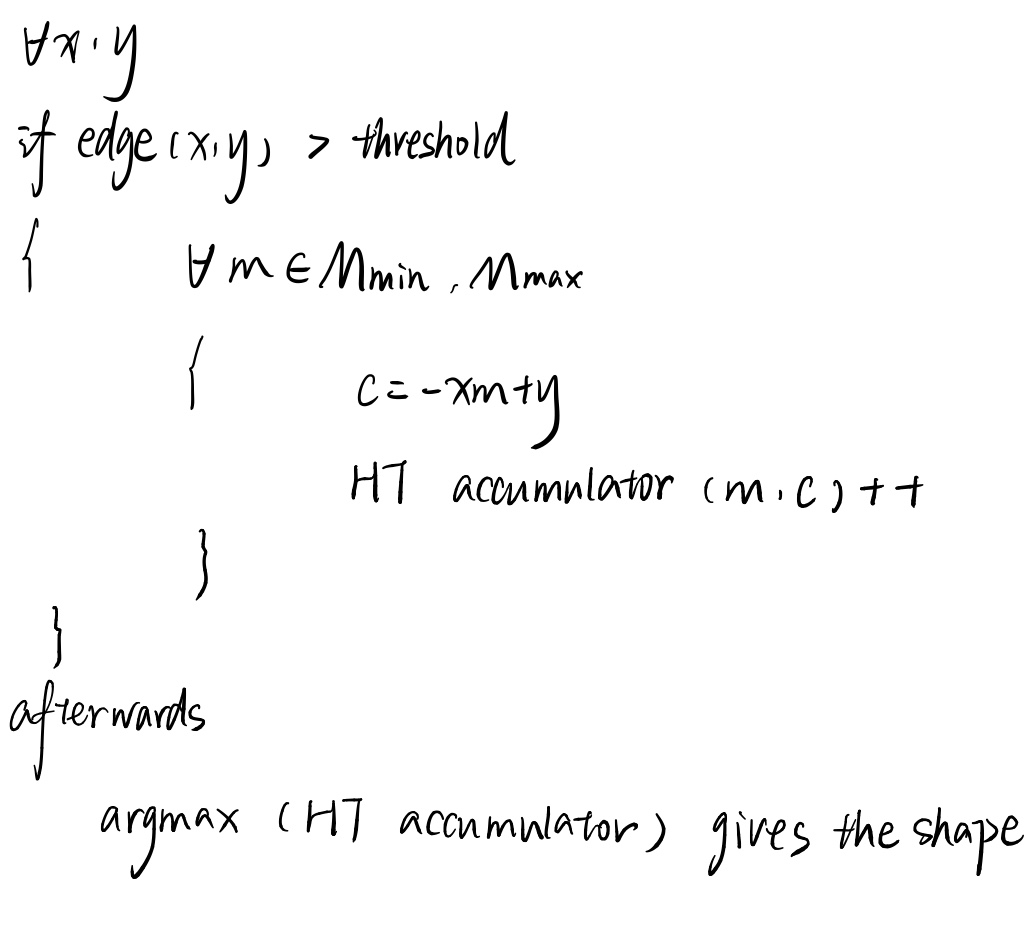
* 1. Hough Transform #霍夫变换（high-level ） 检测直线

definition: an evidence-gathering approach where the evidence is the votes cast in an accumulator array

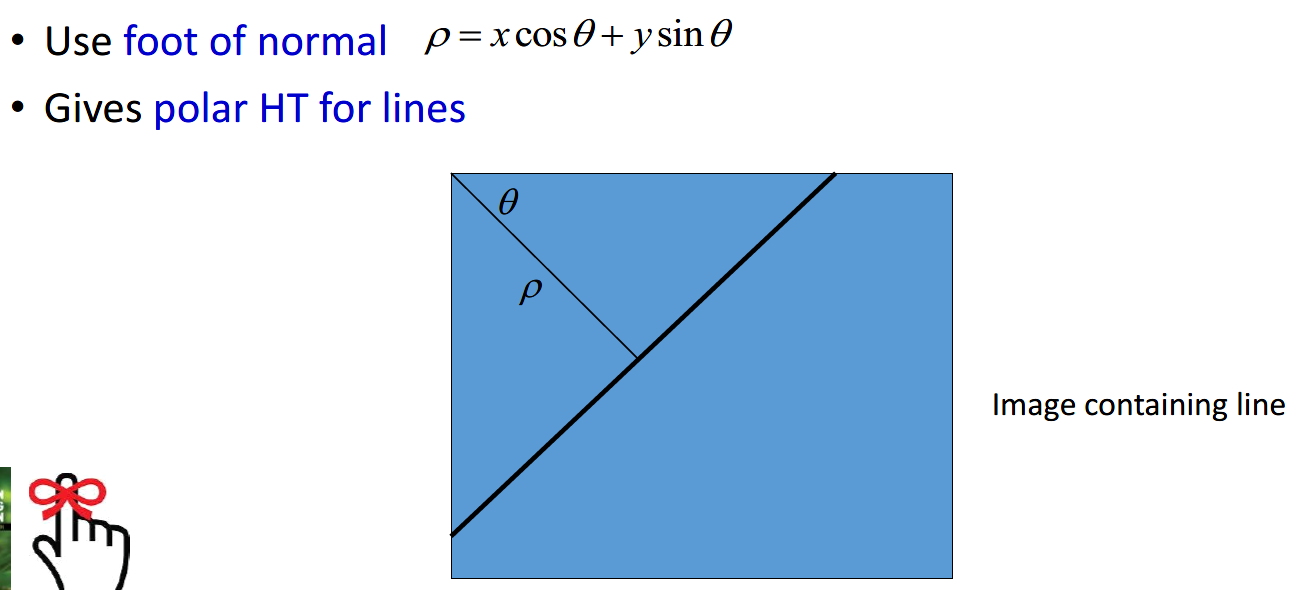
* + 1. finding lines #直线



pseudo code:



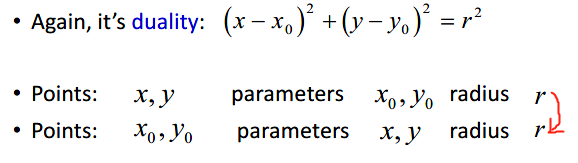
以上是霍夫变换检测直线在笛卡尔坐标系的表示，但是存在一个问题当m趋于无穷时，但是因为点的离散有限的，所以映射到霍夫空间后会有误差，因此采用polar HT

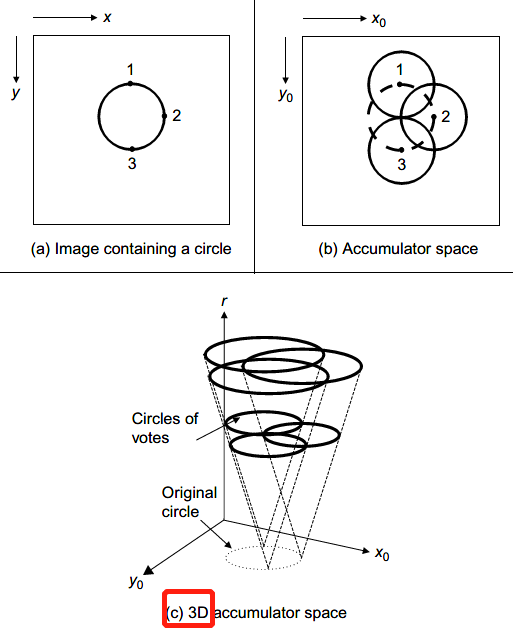


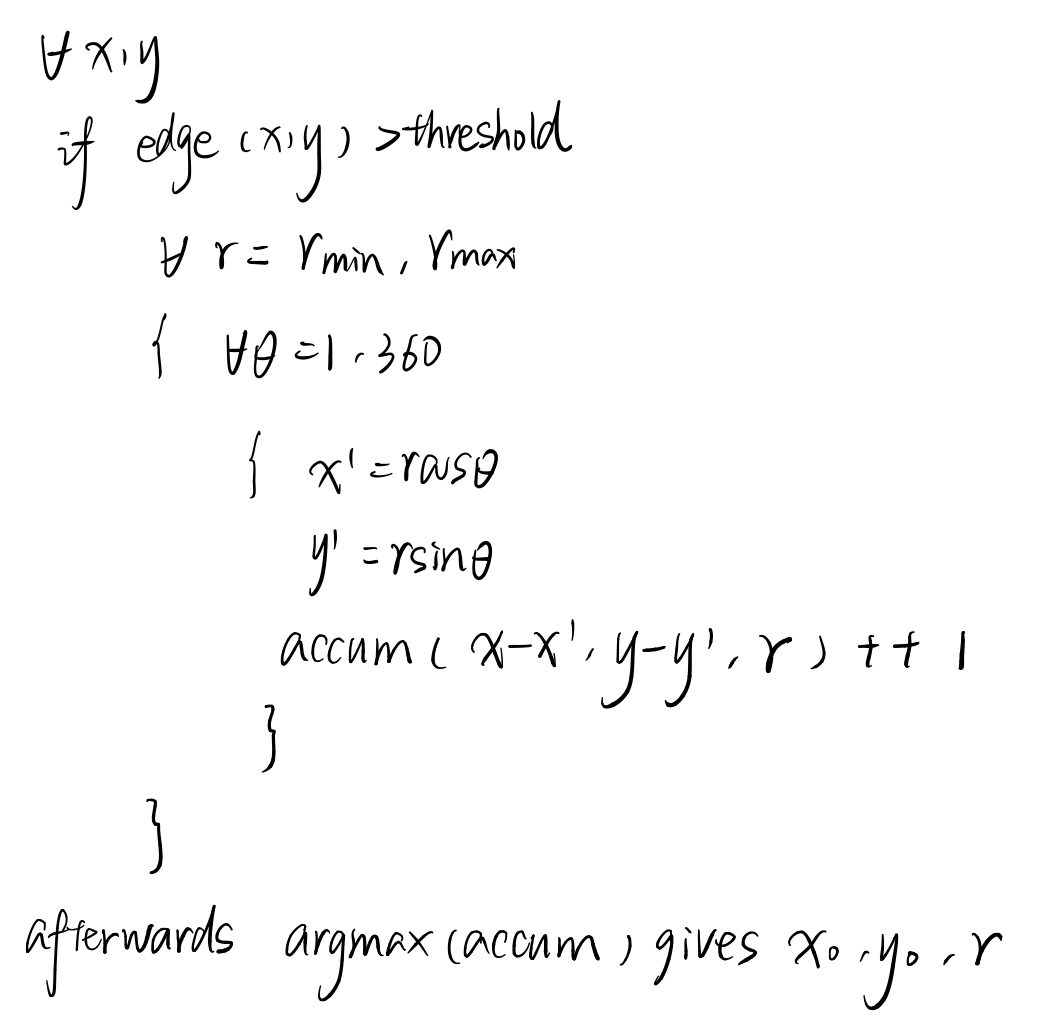
用ρ跟θ就能确定线的位置，且θ范围是[-90°，90°]，ρ的范围为[0，2^1/2]，解决了无穷的问题，并且结果显示极坐标的峰值比笛卡尔坐标更明显，因为离散化效果被削弱

在噪声与遮挡方面， 霍夫与模板检测具备相同的属性

* + 1. finding circles #圆

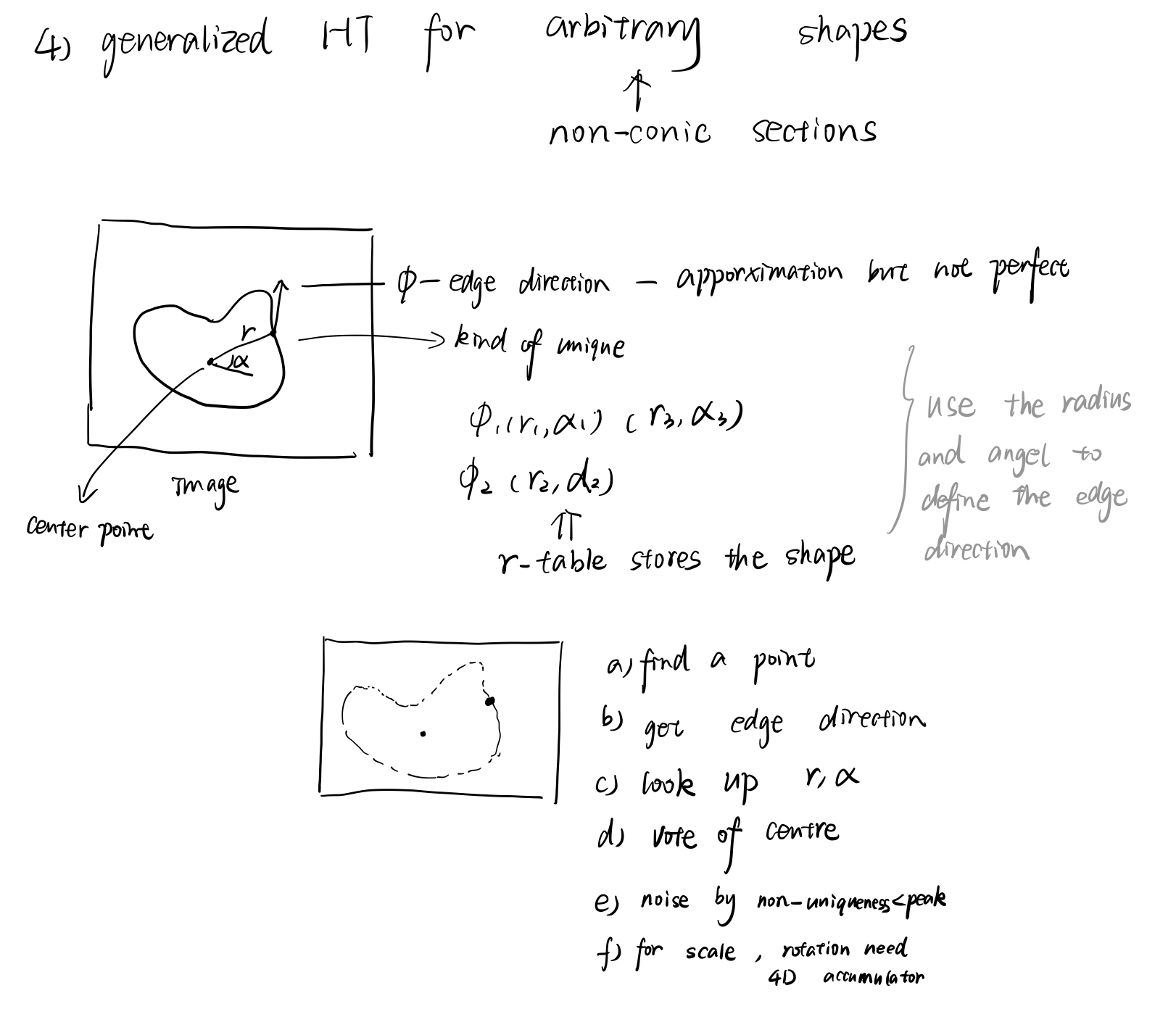






同理可推广到椭圆，为4D空间

* + 1. generalized HT for arbitrary shapes #任意形状



* 1. Comparisons between Template Matching and Hough Transform #两者比较

Template Matching 的时间复杂度O(N^2+m^2)

Hough Transform 的时间复杂度O(number of edge points)

所以霍夫变换大大提升了速度，但是所需空间更多，这两种方法的结果是一样的

霍夫变换跟模板匹配一样能够处理噪声和遮挡问题

优点：handle noise and occlusion, HT is faster than Template Matching

缺点：Template Matching is ineffective when the origin image rotates or scales, HT has to map the points into 4D or higher Hough space, HT needs more storage resources