矩阵计算器：https://matrix.reshish.com/multiplication.php

科学计算器：https://www.desmos.com/scientific?lang=zh-CN

Week1 Introduction

1.1 Ill-posed and Well-posed

- Well-posed: Mapping from world to image (3D to 2D) is **unique** -> forward problem
- Ill-posed: Mapping from image to world (2D to 3D) is **NOT unique** -> inverse problem
- To recover depth, we need extra information
    - another image or
    - pior knowledge about the structure of the scene

Week2 Image Formation

2.1 Physics of image formation

2.1.1 Ingredients of image formation

- Radiometric parameters (intensity/colour) --> 决定图片亮度与颜色
    - Illumination
    - Surface reflectance properties
    - Sensor properties
- Geometric parameters -->决定位置
    - Camera position
    - Camera optics
    - projection geometry

2.1.2 Pinhole camera

- Focus焦点：all rays coming from a scene point converge into a single image point
- Exposure曝光时间：is the time needed to allow enough light through to form an image
    - The longer the exposure the more blurred an image is likely to be
- Aperture胶圈: the smaller the aperture, the longer the exposure time
- 小pinhole会导致对焦清晰，但图像模糊; 大pinhole，更亮的图像但模糊；
    - 需要Lens才能产生清晰，在焦点，明亮的照片
- pinhole camera所有的东西都在焦点上，无论image plan length是多少
    - pinhole camera has an infinite focal range, image will be in focus no matter waht distance the image plan is from the plane

2.1.3 Lens

$$\frac{1}{f} = \frac{1}{z} + \frac{1}{z'}$$

f = focal length

z = distance of object from the lens

z' = distance of image from the lens

2.2 Geometry of Image Formation

2.2.1 Intrinsic Parameters

Maps points on the **image plan** into **pixel image coordinates**

P' = M P 把相机中的某点P通过投影矩阵M变换到了像素坐标中

2.2.2 Extrinsic Parameters

2.2.3 Digital image

TOP Left is Origin

- Pixelisation：每个小区域取平均值（Mincraft中的像素风）
- Quantization：用一个finite num of discrete value代替intensity value
  - 比如1bit/pixel = 2^1Gray levels （**BINARY IMAGE**），表示1bit可以表示两种不同的pixel, 即为0和1
  - 2bits/pixel = 2^2Gray levels，表示2bits可以表示4种不同的pixel，一张图用4种不一样的pixel来表示, 一张图中可以映射到[0,63], [64, 127], [128,191], [192, 255]这四个区间内。
  - 8bits/pixel = 2^8Gray levels，表示8bits可以表示256种不同的pixel，一张图用256种不一样的pixel来表示，即为**原图**

2.3 Camera Sensors

2.3.1 Demosaicing

Demosaicing is a process that computes the colour(RGB values) at every pixel based on the local red, green and blue values.

Demosaicing is a method of interpolation

2.3.1.1 Nearest Neighbour Interpolation

拷贝最近邻相同色的pixel value

2.3.1.2 Bilinear Interpolation

取最近邻相同色两个或四个pixel的平均值

2.3.1.3 Smooth Hue Transition Interpolation

- 对于绿色pixel：与Bilinear Interpolation相同
- 对于红和蓝色pixel：Bilinear Interpolation of the ratio between red/blue and green

2.3.1.4 Edge-Directed Interpolation

- calculate horizontal H and vertical gradients V
- If H < V, Gx = averay of 水平邻近的G
- If H > V, Gx = averay of 垂直邻近的G
- else，Gx = averay of邻近的G

2.4 Eye

2.4.1 Photoreceptor

光感受器可以把光信号转换为电信号

1. Photoreceptor包括：

- Rods杆体细胞：high sensitivity（可在光线昏暗的情况下工作）
- Cones圆锥细胞：low sensitvity（需要明亮的光线），有三种不同种类的cones对不同波长敏感
    - Blue
    - Green
    - Red

2. Photoreceptor不是均匀分布在视网膜上的

- blind spot：无Photoreceptor
- fovea：no rods, high density of cones. **#Green cones > # Red cones << #Blue cones** 对绿色最敏感
- periphery：high concentration of rods，few cones

2.4.2 Ganglion cells

神经节细胞：视网膜最终段的神经细胞

两种类型：

- ON-centre, off-surround: active if central stimulus is **brighter** than background
- Off-centre, ON-surround: active if central stimulus is **darker** than background

视网膜上的Photoreceptor光感受器(杆体细胞和锥体细胞)通过接受光并将它转换为输出神经信号而来影响许多神经节细胞以及视觉皮层中的神经细胞．反过来，任何一种神经细胞的输出都依赖于视网膜上的Photoreceptor光感受器．我们称直接或间接影响某一特定神经细胞的光感受器细胞的全体为该特定神经细胞的感受野(receptive field), 比如**神经节细胞的感受野就是神经节细胞中的所有Photoreceptor**

在视觉系统中，任何层次或水平上的单个神经细胞均在视网膜上有一特定代表区域，在该区域上的光学刺激能影响该神经细胞的活动，这个区域定义为该细胞的视觉感受野。

Colour Opponent Cells

- Red ON/Green OFF, Red OFF/Green ON
- Green ON/Red OFF, Green OFF/Red ON
- Blue ON/Yellow OFF, Blue OFF/Yellow ON
- Yellow ON/Blue OFF, Yellow OFF/Blue ON

Yellow 是red和green的平均值

The standard way of modelling ganglion cell receptive field is by using a DOG operator

Week3 Low Level (Artificial)

3.1 Convolution and Correlation

- If mask is symmetrical, then cross-correlation = convolution
- Convolution: 先顺时针旋转180后再相乘
- cross-correlation不满足结合律和交换律

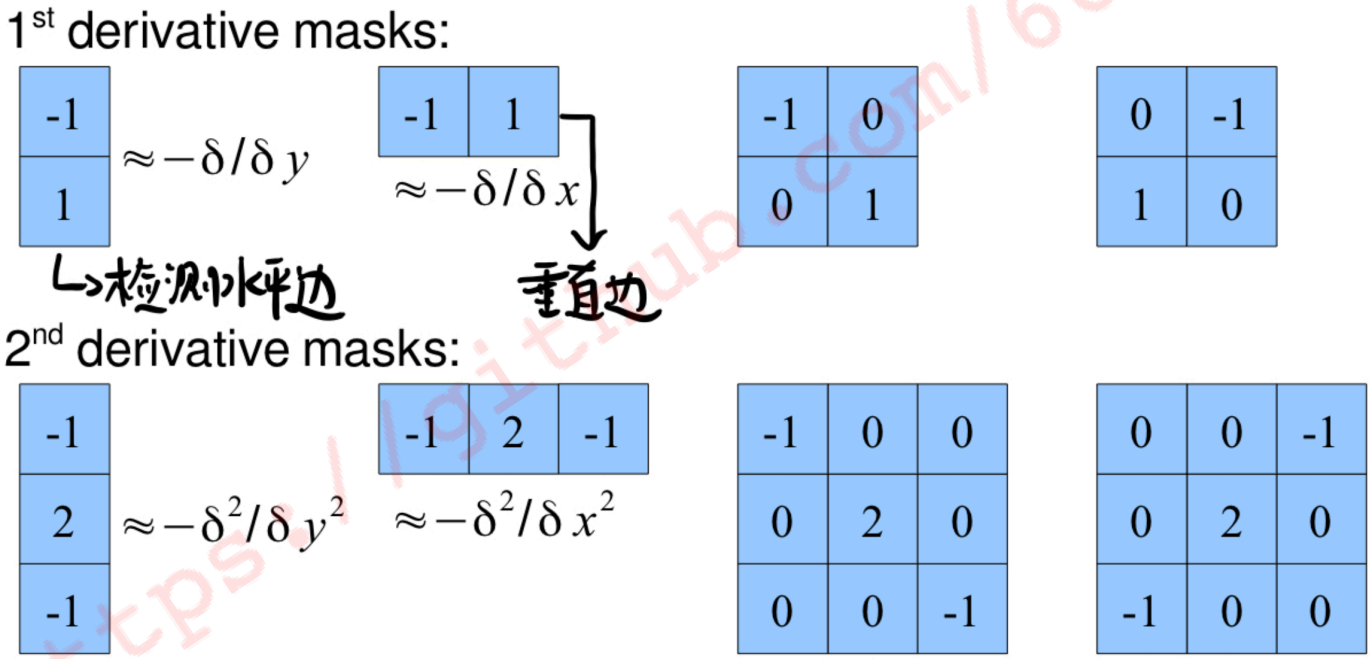
3.2 Smoothing

平滑高频率

- Gaussian mask
  - standard deviation越大，卷积完越模糊，因为更多的pixel被平滑
  - mask width起码要大于等于6倍的standard deviation
  - 2D高斯核可以拆解为两个1D高斯核的乘积

3.3 Differencing mask

The difference between pixel values measures the gradient of the intensity values



3.3.1 Laplacian mask

```
-1 -1 -1
-1  8 -1          or
-1 -1 -1

-1/8 -1/8 -1/8            -1  -1  -1
-1/8   1  -1/8     or     -1   8  -1
-1/8 -1/8 -1/8            -1  -1  -1
```

可以检测水平，垂直，对角边, 和为1

Laplacian mask and Difference masks are sensitive to noise.

3.4 Edge Detection

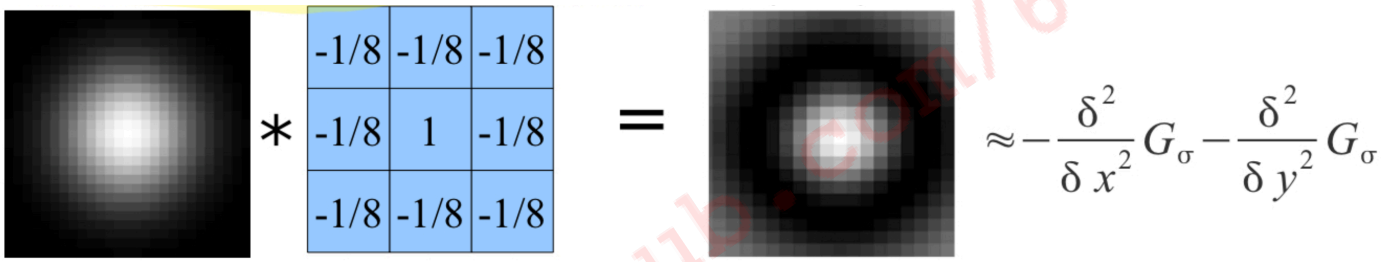
3.4.1 Intensity discontinuities

- Depth discontinuity
- Orientation discontinuity
- Reflectance discontinuity
- Illumination discontinuity

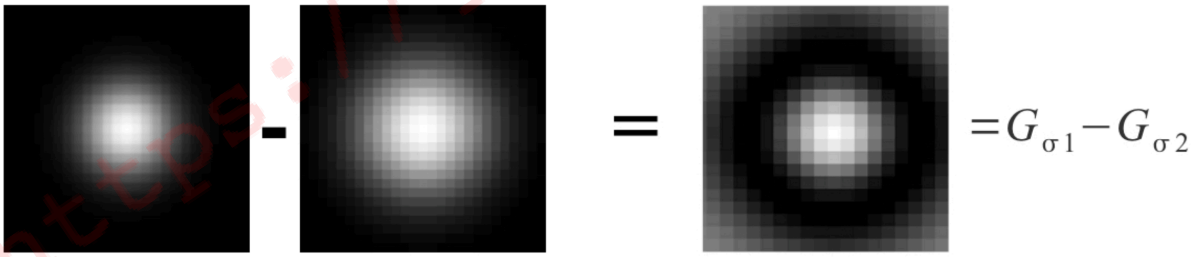
3.4.2 LoG / DoG Mask

边缘检测：先高斯平滑再与Laplacian mask卷积检测边缘，相当于高斯核*Laplacian mask

LoG = Gaussian mask * Laplacian mask

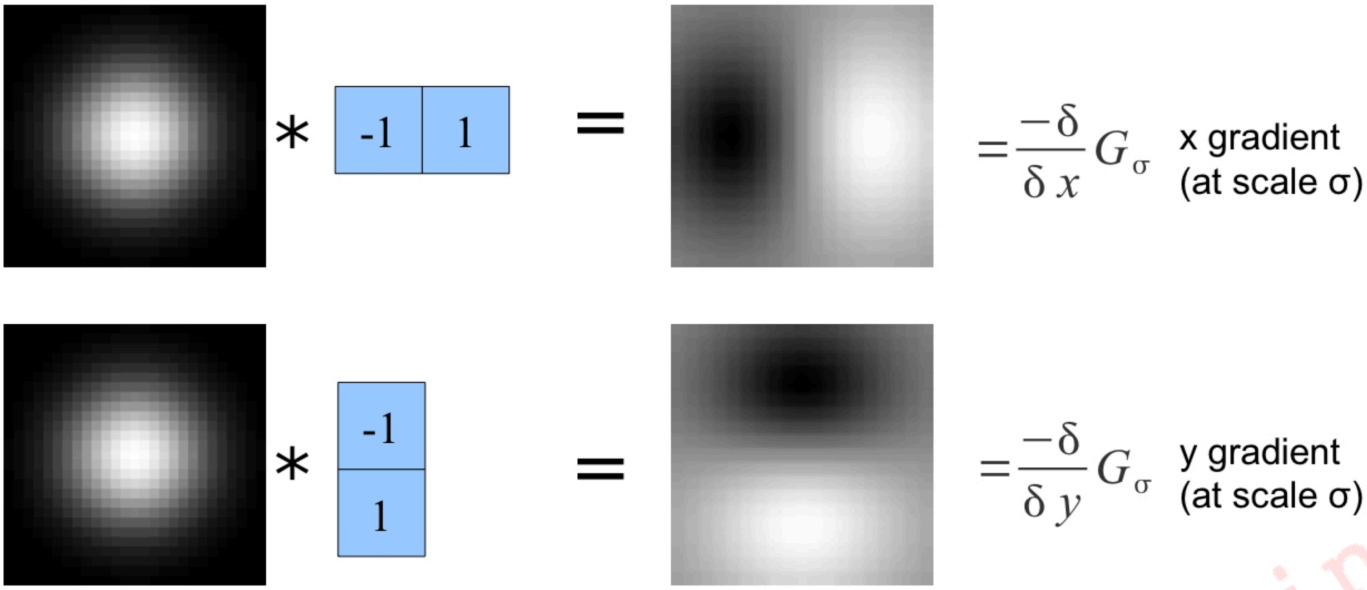


It is similar to the Difference of Gaussians or DoG mask (or "centre-surround" or "Mexican hat").



可以用DoG来近似LoG，DoG是用两个不同尺度的高斯核相减，得到LoG

除LoG之外，还可以使用一阶高斯偏导核来检测水平和垂直边缘：



右边的一阶高斯偏导核 = 高斯核 * 一维卷积

3.5 Canny Edge detector

- Filter image with derivative of Gaussian
- Calculate the gradient magnitude and orientation
- Non-maximum suppression(细化粗边)
- Linking and thresholding(hysteresis):
  - i. Define high and low thresholds
  - ii. Apply a high threshold on the magnitude to initialize a contours and continue tracing the contour until the magnitude falls below a low threshold

3.6 Image Pyramids

3.6.1 Gaussian Pyramid

a Gaussian image pyramid is a multiscale representation of a single image at different resolutions obtained by iteratively convolving an image with a Gaussian filter and down-sampling.

3.6.2 Laplacian Pyramid

a Laplacian image pyramid is a multiscale representation of a single image that highlights intensity discontinuities at multiple scales. It is obtained by iteratively convolving an image with a Gaussian filter, subtracting the smoothed image from the previous one, and down-sampling the smoothed image.

Week4 Low & Middle Level (Biological)

4.1 The Bio Visual System

- LGN（外侧膝状体核）
- Cerebral Cortex（大脑皮层）

4.2 Primary Visual Cortex

V1 Cell have RF selective for：

- colour
- orientation
- direction of motion
- spatial frenquency
- eye of origin
- binocular disparity
- position

V1 RFs Orientation：

- Simple Cells：只回应特定方向，常被用作edge and bar detectors
- Complex Cells：对位置更invariance，所以常用作edge and bar detectors with some tolerance to location
- Hyper Complex Cells：不仅能回应特定方向，还能回应特定长度的

4.3 Gabor Functions

Energy Model

4.4 Non-Classical RFs

- Classical Receptive Field（cRF）= the region

4.5 Gestalt Laws

Image segmentation in the brain

- 通过Top-down的方式分割图像
- 通过Bottom-up的方式分割图像(Gestalt laws)

Gestalt laws:

1. Proximity
2. Similarity
3. Closure
4. Continuity
5. Common Fate
6. Symmetry
7. Common Region
8. Connectivity

4.6 Border Ownership

某边到底属于谁

V2 Border-ownership cells

Border ownership refers to the fact that the boundary between two regions in an image is perceived as part of one region (the foreground) and not the other region (the background). This means that foreground objects have a defined shape (delineated by the border), whereas background objects appear shapeless.

Week5 Mid-Level Segmentation (Segmentation)

5.1 Thresholding

1. Average intensity
2. Intensity Histogram
3. Hysteresis thresholding

5.1.1 Morphological Operations

1. Dilation: Fill holes
    1. Full match => 1
    2. Some match => 1
    3. no match =>0
2. Erosion: Remove bridges, branches, noise
    1. Full match => 1
    2. Some match => 0
    3. no match =>0

5.2 Region-based

具体请看tutorial5 answer

5.2.1 Region Growing

5.2.2 Region Merging

5.2.3 Region Splitting and Merging

5.3 Clustering Methods

5.3.1 K-means

- Randomly choose k points to act as cluster centres
    - For each data point
        - Calculate its similarity to each cluster centre
        - Allocate data point to closest cluster centre
    - Repeat from step 2 for each data point
    - For each cluster centre
        - Calculate its new position as the mean position its elements

    - Repeat from step 6 for each cluster centre
- Repeat from step 2 until the cluster centres are unchanged

5.3.2 Agglomerative clustering

- single-link clustering
    - Distance between clusters is shortest distance between elements (MIN distance)
- complete-link clustering
    - distance between clusters is longest distance between elements (MAX distance)
- group-average clustering
    - distance between clusters is average distance between all pairs of elements (AVERAGE distance)
- centroid clustering
    - distance between clusters is distance between their averages.

5.3.3 Group Cutting

$$Ncut(A,B) = \frac{cut(A,B)}{assoc(A,V)} + \frac{cut(A,B)}{assis(B,V)}$$

分子：A和B中所有连接的edge的weights之和（红线）

分母：A或B集合所有的edge之和

要避免把A切成一个像素，B是一个整体；或者A是一个整体，B是一个像素

如果A集合有很多点，则分母会变大，整体会变小，整体变小就会被分割，切割后就保证了A几何中有很多点而不是一个像素

当A和B集合里都有很多像素点时，NCut达到最小，会被安全切割

5.4 Fitting

5.4.1 Hough Transform

5.4.2 Active Contours

Energy = Internal energy + External energy

- Internal energy is a function of the shape of the contour, it is reduced if the curve is short and smooth.
- External energy is a function of the image features near the contour, it is reduced if the intensity gradient is high.

Week6 Mid-Level Correspondence

找两张图片中的对应点存在的问题：

- 遮挡
- False match
- 大的搜索空间

6.1 Correlation-based methods

使用滑窗，对比每一个像素的相似度来找对应点

需要大量计算，难以确定Windows大小，物体遮挡，发光不能很好的计算出对应点

6.2 Feature-based methods

先提取特征（interest Points），再对比特征相似度来找对应点

6.2.1 Interest Points

角点corner提供了很好的信息，来找到两张图片的对应点

在角点的周围的intensity gradient在x和y方向很高，因为如果只有x方向则没有x方向的变化，只有y方向同理；角点既有x又有y，所以变化很大，因此intensity gradient很高

6.2.2 Harris

1. Compute Gaussian derivatives at each pixel,
2. Compute second moment matrix M in a Gaussian window around each pixel,
3. Compute corner response function R,
4. Threshold R,
5. Find local maxima of response function （nonmaximum suppression）

Harris Corner对尺度变换不是Invariance的！！

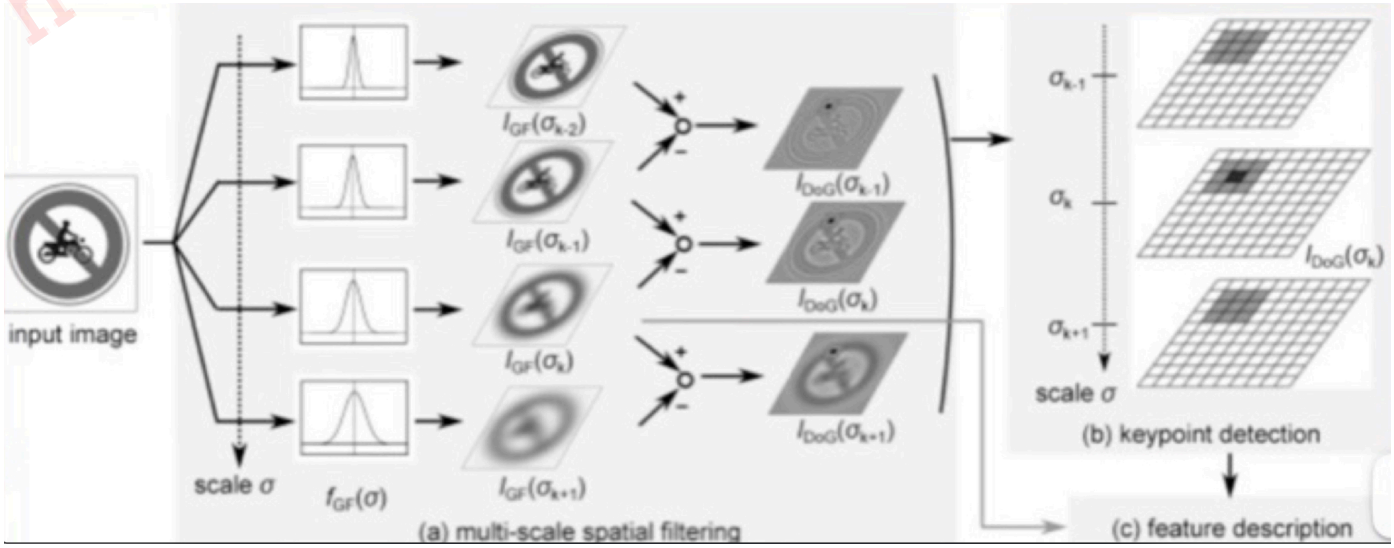
Harris invariant to brightness changing(intensity shift) and covariant to translation and rotation. In the cases image brightness changing(intensity shift), translation and rotation, Harris corners still can be detected by Harris detector.

6.2.3 SIFT

针对HArris角点没有尺度不变性提出的，目标：可以检测出同一图片按照不同尺度缩放的对应区域

Laplacian算子具有尺度选择性，当laplacian与blob尺度匹配时，在blob中心laplacian有最值，就是尺度变的区域

SIFT流程



1. Construct Scale Space(用不同尺度的高斯核卷积图片来获取不同尺度下的图片，来模拟真实世界中由于远近距离造成的尺度和模糊程度)
2. Compute Difference of Gaussian，来近似LoG
3. Find local maxima in scale (Scan each DoG images, look at all neighboring 26 points) <- 同时可实现非最大值抑制
4. Sub Pixel Locate Potential feature points
5. Assign keypoint orientation
6. Build keypoint descriptors

- 如何 Build SIFT Feature descriptor?

1. Determine the affine region
2. Normalize region
3. Remove the rotational ambiguity
4. Form a descriptor from normalized region

6.2.4 Match

如何在两张图片之间找到true correspondence？？

RANSAC流程

利弊

Week7 Mid-Level Stero & Depth

7.1 Stereo Camera

Depth information can be recovered using 2 images

7.1.1 Coplanar Camera(Simple case)

在相机共面的情况下，disparity等于某像素在左右两个相机之差 $d = x_L - x_R$

$depth = f\frac{baseline}{disparity}$

通过公式可以看出，disparity与depth成反比，即距离像面越近的点，在左右相机中的视差越大，反之亦然。我们根据两张图片的disparity就可以计算出depth map。但是如何得到disparity呢？通过计算两张图中的correspondence位置，来获取disparity。所以how to solve stereo correspondence problem成了关键问题。

通过极平面来找correspondence。通过几何约束将搜索范围缩小到对应的极线上。

Stereo Constraints on Correspondence：

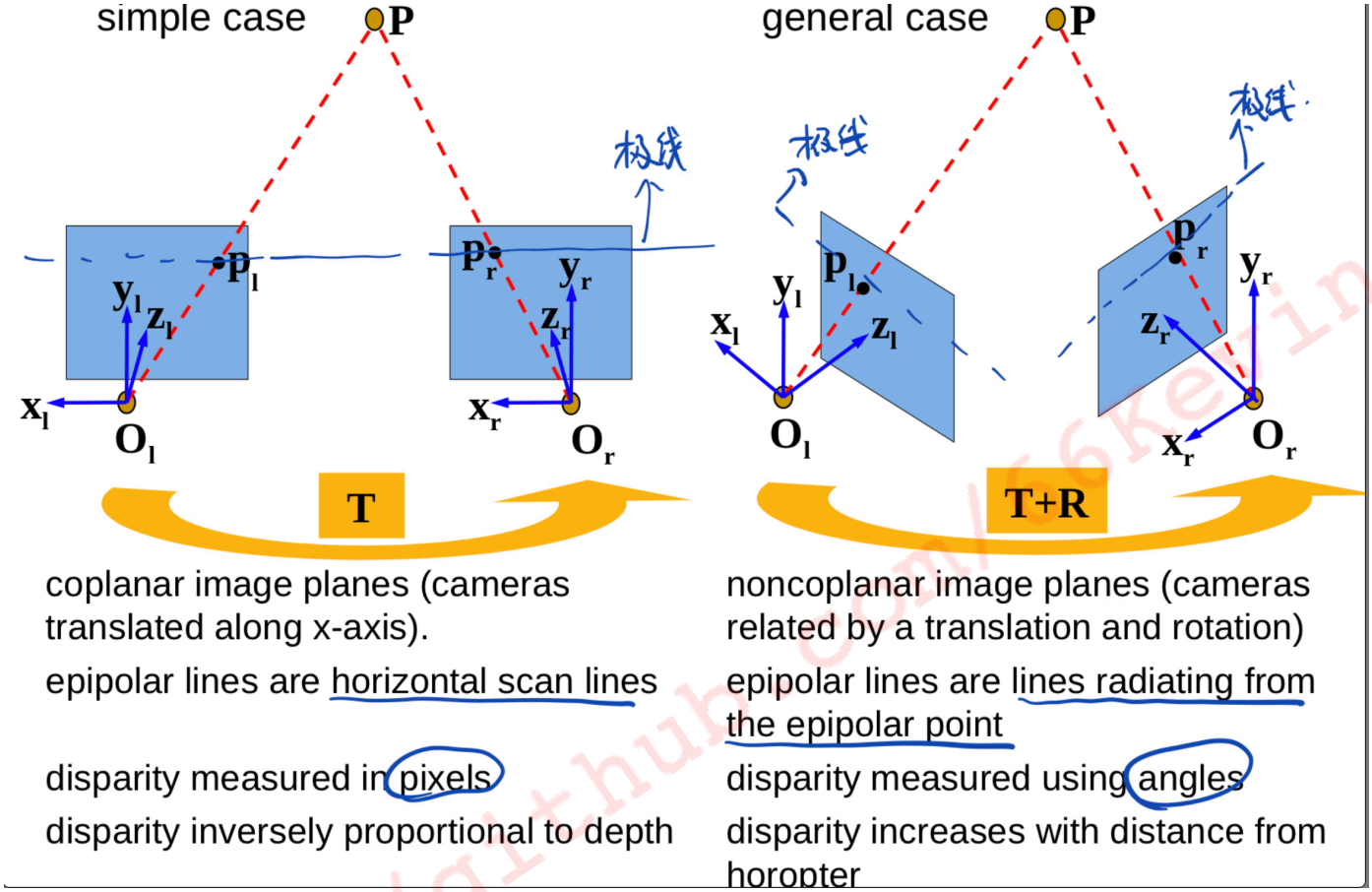
- Epipolar constraints：通过几何约束将搜索范围缩小到对应的极线上
- Maximum disparity：因为物体距离相机的距离要远远大于baseline的长度，如果相机距离物体太近，就没法拍到有共同 部分的照片，也就找不到corresponding points：fails for points closer to cameras than Zmin.
- Continuity
- Uniqueness
- Ordering

在相机共面的情况下，corresponding points位于两张图片的极线上（极线共面；）

7.1.2 Non-Coplanar Camera(Complex case)

在相机不共面的情况下，disparity等于某像素在左右两个相机角度之差 $d = \alpha_L - \alpha_R$

如果d大于0，则表示outside of the horopter(落在horopter上的点视差都为0)；反之亦然



The steps of Stereo Reconstruction or how to find depth?

1. Calibration(intrinsic extrinsic parameters)
2. Rectification(use Epipole constraints to find correspondence points in 1D space)
3. Calculate the disparity(to find Disparity map use similar triangles in $Z = f\frac{T}{x_r - x_l}$
4. triangulation(to find Depth map)

7.2 Cues to Depth

- Oculomotor（眼球运动）
  - **Accommodation**: The shape of the lens in the eye, or the depth of the image plane in a camera, is related to the depth of objects that will be in focus. Hence, knowledge of these values provides information about the depth of the object being observed.
  - **Convergence**: The rotation of eyes/cameras in a stereo vision system can vary to fixate objects at different depths. Hence, the angle of convergence provides information about the depth of the

object being fixated.

- Monocular (单目视觉)
  - **Interposition**: Nearer objects may occlude more distant objects. Hence occlusion (or interposition) provides information about relative depth.
  - **Size familiarity**: Objects of known size provide depth information, since the smaller the image of the object the greater its depth.
  - **Texture gradients**: For uniformly textured surfaces, the texture elements get smaller and more closely spaced with increasing depth.
  - **Linear perspective**: lines that are parallel in the scene converge towards a vanishing point in the image. As the distance between the lines in the image decreases, so depth increases.
  - **Aerial perspective**: Due to the scattering of light by particles in the atmosphere, distant objects look fuzzier and have lower luminance contrast and colour saturation.
  - **Shading**: The distribution of light and shadow on objects provides a cue for depth.
- Motion
  - **Motion parallax**: As the camera move sideways, objects closer than the fixation point appear to move in a direction opposite to the camera, while objects further away appear to move in the same direction. The speed of movement increases with distance from the fixation point.
  - **Optic Flow**: As a camera moves forward or backward, objects closer to the camera move more quickly across the image plane.
  - **Accretion and deletion**: As a camera moves parts of an object can appear or disappear; these changes in occlusion provides information about relative depth.
  - **Structure from motion (kinetic depth)** : Movement of an object or of the camera can generate different views of an object that can be combined to recover 3D structure.

Week8 Video and Motion

8.1 Optic flow & Motion flow

Optic flow是利用图像序列中的像素在时间域上的变化、相邻帧之间的相关性来找到的上一帧跟当前帧间存在的对应关系，计算出相邻帧之间物体的运动信息的一种方法

Motion flow则是真实世界中，物体在时间域上的变化、相邻帧之间的相关性

为了找到Optic flow就必须找到两frame之间的对应点

- Feature-based methods
- Direct methods

Constraints：

- Small motion： (assume optical flow vectors have small magnitude). Fails if relative motion is fast or frame rate is slow
- Spatial coherence：(assume neighbouring points have similar optical flow). Fails at discontinuities between surfaces at different depths, or surfaces with different motion

找到了对应点就可以计算3d结构与恢复motion:

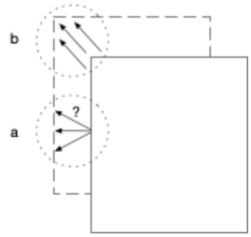
- with knowledge of ego-motion: calculate absolute depth
- Without knowledge of ego-motion:

- calculate relative depths
- Time-to-collision：how long the camera will collapse with object
- direction of ego-motion
- heading of ego-motion

8.2 Aperture problem

孔径问题指无法通过单个算子【计算某个像素值变化的操作，例如：梯度】准确无误地评估物体的运行轨迹。原因是每一个算子只能处理它所负责局部区域的像素值变化，然而同一种像素值变化可能是由物体的多种运行轨迹导致。

> 又例如下面的例子：矩形的实际运行轨迹是自右下向左上。然而在算子a看来，只是一条竖线从右侧实线位置移动到了左侧虚线的位置。因此，算子a并不能准确判断出矩形的实际运行轨迹。



解决方案：

- integrating information from many local motion detectors / image patches, or
- by giving preference to image locations where image structure provides unambiguous information about optic flow (e.g. corners).

Week9 High-Level Vision (Artificial)

9.1 Category hierarchy

上层更抽象，下层更具体

9.2 Template Matching

Template：要被recognized的物体

- 搜索每一块区域
- 计算template与image region的相似度
- 选取超过阈值的最佳区域

Templates 需要与目标物体非常相似才能检测出来，如果物体发生了形变，旋转，就无法检测

解决方案：multi templates for each object

问题：遮挡无法检测，not robust to changes in appearance

9.3 Similarity Measures

We can **maximise** the following measures

- Cross-correlation
- Normalised cross-correlation （cosine of the angle between i and j）
- Correlation coefficient

We can **minimise** the following measures

- Sum of Squared Differences:
- Eculidean distance:
- Sum of Absolute Differences:

9.4 Sliding Window

对于每一块image patch用分类器检测是否包含物体（就不用比较intensity values）

先用image segmentation处理后，会提高速度

9.5 Edge Matching

像template matching一样，只不过先提取边缘

9.6 Model-based object recognition

先假设出物体的形状与姿态，然后在图像中描绘出物体，再比较

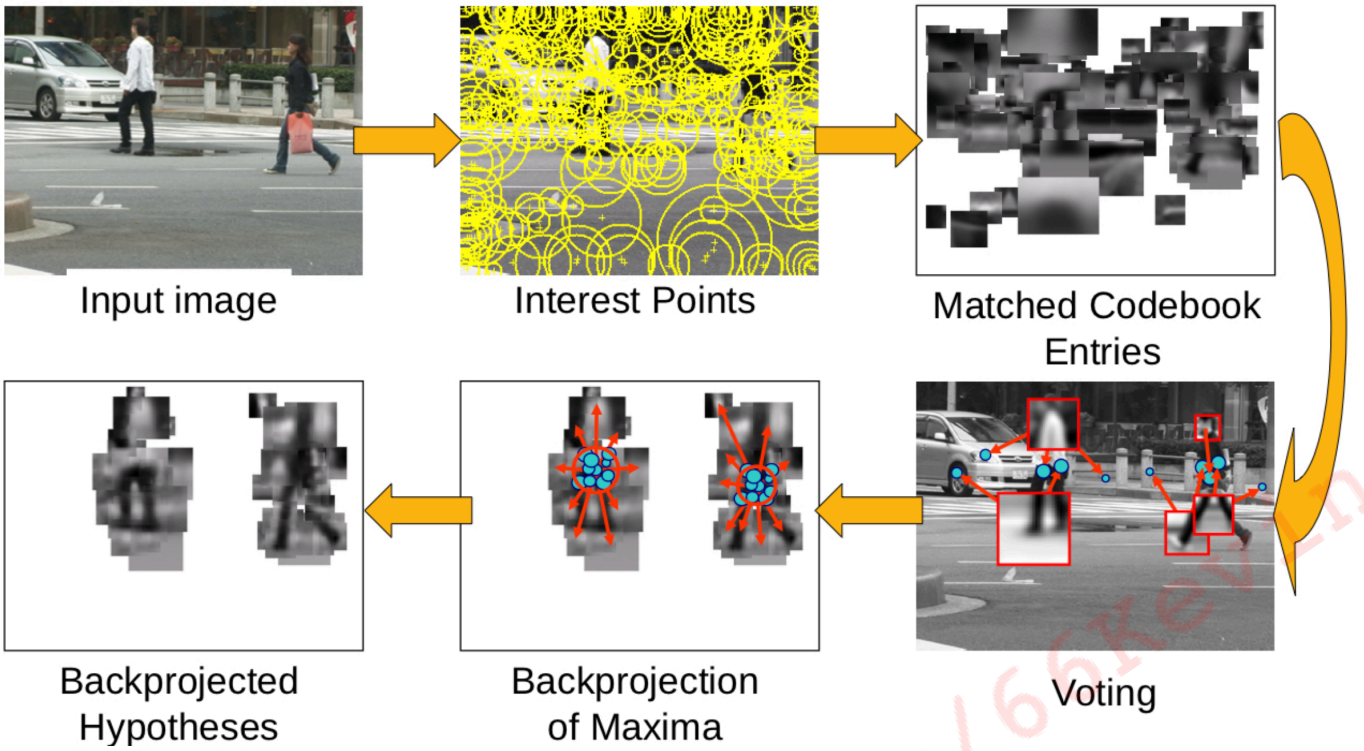
9.7 Intensity histograms

compare histograms to find closes match

Insensitive to small viewpoint changes and spatial configuration

9.8 Implicit Shape Model (ISM)

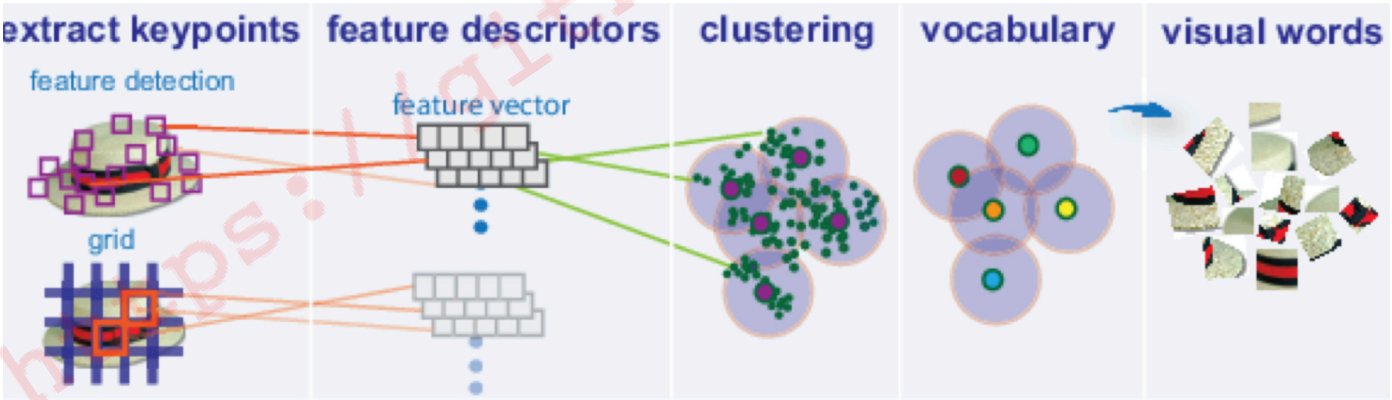
Matching:



9.9 Feature-based object recognition

先提取SIFT特征

9.10 Bag-of-words



9.11 Geometry Invariants

Week10 High-Level Vision (Biological)

10.1 Theories of Object Recognition

10.1.1 Object based：Recognition by Components

主要思路：每个object都用一个3D模型表示，小的geometric components组成了大的整体

structural descriptions：用来表示一个物体的组成部分与内部关系，比如the cube above cylinder

geometrical icons / geons：3D物体，比如cube，sphere，cylinder，wedges，不同的geons组合可以代表不同的obj

使用geons可以达成viewpoint invariance的目的，因为different views of the same obj are represented by the same set of goons, in the same arrangement.

Problems:

- 很难将一幅图片decompose成很多的components
- 有很多自然物体很难用geons表示出来，比如树
- 无法表示细节特征，或者无法区分微小的物体

10.1.2 Image based

3D obj represented by multiple 2D views of the obj

Local(Featural) VS Global(Configural)

Rules VS Prototypes VS Exemplars

- Rules: 所有满足抽象的规则的事物属于一类。比如四条腿会叫的生物：狗；有三条边：三角形；
- Prototypes：计算所有类别的中个体的平均值，新物体与每个类别平均值比较，看属于哪个最近的类别
- Exemplars：每个类别个体用向量表示后保存下来，新物体与每个类别的个体比较，看属于哪个最近的类别

Nearest Mean Classifier（Prototypes）

Nearest/K-Nearest Neighbors Classifier（Exemplars）->无法处理outliers（noise）

10.2 The Cortical Visual System

10.2.1 Pathways

"What" and "Where" pathways: 沿着pathways走，neurons preferred stimuli gets more **complex**, receptive fields become **larger**, and there is greater **invariance** to location, sensitivity to stimulus location **decreased**.

10.2.2 HMAX

Hierarchical Maxpooling Model:

- S-cells(Simple)：sum（and）
- C-cells(Complex)：max（or）

在CNN中卷积层相当于HMAX中的S-cells；池化层相当于HMAX中的C-cells

- **Simple cell**: input is from a number of centre-surround cells which have RFs on a common line. These centre-surround neurons are activated by a bar/edge at the correct orientation, resulting in the simple cell responding to a oriented bar/edge at a specific orientation.
- **Complex cell**: input is from a number of simple cells with the same orientation preference within a

small spatial region. A bar/edge at the correct orientation and location to activate one of these simple cells will result in the complex cell responding, and hence, the complex cell responds to to oriented edges with some tolerance to exact location.

10.3 Bayesian Inference

$$p(H|E) = \frac{p(E|H)p(H)}{p(E)}$$

$p(H|E)$ is the (posterior) probability that hypothesis H is true, given the image evidence E. This is what the vision system needs to evaluate (generally, we want to find the most likely hypothesis that explains the image data.)

$p(E|H)$ is the likelihood that if hypothesis H were true, the image would contain particular evidence E. (Calculating this quantity is based upon our scientific knowledge of the image formation process, e.g. that a certain set of surface properties and illumination conditions would result in a certain image being formed as a result.)

$p(H)$ is the prior assumptions about the likelihood of the hypothesis in the first place. If H is extremely improbable, then stronger evidence is required to support it.

$p(E)$ is the probability that the evidence E would be found in images anyway, regardless of whether or not hypothesis H is true. Thus if p(E) is large (e.g. images contain some bright regions no matter what), then this reduces our confidence in inferring any particular hypothesis H as a result of observing E.