

Fish 计算机视觉知识点整理

Gestalt Laws (格式塔法则)

Law of Proximity (接近原则)

物体越接近，那么它们更容易被感知为同一组的。

Law of Similarity (相似原则)

若物体具有相似特征，那么它们更容易被感知为同一组的。

Law of Common Fate (共方向原则)

若物体向共同方向运动，那么它们更容易被感知为同一组的。

Law of Symmetry (对称原则)

我们倾向于把不对称，不完全，复杂的图形感知成对称、完全、简单的图形。

Law of Continuity (连续原则)

我们倾向于感知连续，而不是零散。也就是我们会把一些看起来零碎的东西看做是连续的。

Law of Closure (封闭原则)

哪怕物体不完整（不存在），我们也能根据认知脑补出缺失的部分。

Marr视觉表示框架的三个阶段？

Primal Sketch

进行图像处理。提取角度、边缘、纹理、线条、边界等基本特征。

2.5D Sketch

以观测点为中心的坐标系中，恢复场景可见部分深度、法线、轮廓等信息。

3D Model

以物体为中心的坐标系中，恢复、表示和识别三维物体。

二值图像

几何特性

面积:

$$A = \sum_{i=0}^{n-1} \sum_{j=0}^{m-1} B[i, j]$$

区域中心:

$$\bar{x} = \frac{\sum_{i=0}^{n-1} \sum_{j=0}^{m-1} jB[i, j]}{A}$$
$$\bar{y} = -\frac{\sum_{i=0}^{n-1} \sum_{j=0}^{m-1} iB[i, j]}{A}$$

方向:

有些形状无方向，其余以长轴方向为形状的方向。

方向的直线方程:

$$\rho = x \cos \theta + y \sin \theta$$

求出夹角:

$$\sin 2\theta = \pm \frac{b}{\sqrt{b^2 + (a-c)^2}}$$
$$\cos 2\theta = \pm \frac{a-c}{\sqrt{b^2 + (a-c)^2}}$$
$$a = \sum_{i=0}^{n-1} \sum_{j=0}^{m-1} (x_{ij} - \bar{x})^2 B[i, j]$$
$$b = 2 \sum_{i=0}^{n-1} \sum_{j=0}^{m-1} (x_{ij} - \bar{x})(y_{ij} - \bar{y}) B[i, j]$$
$$c = \sum_{i=0}^{n-1} \sum_{j=0}^{m-1} (y_{ij} - \bar{y})^2 B[i, j]$$

伸长率:

$$E = \frac{\chi_{\max}}{\chi_{\min}}$$

密集度:

其中 **p**: 周长，**A**: 面积

$$C = \frac{A}{p^2}$$

形态比:

区域的最小外界矩形长宽比。

欧拉数: 连通分量数 - 洞数

投影计算

水平投影：计算每一列像素为1的个数。

垂直投影：计算每一行像素为1的个数。

对角线投影：从左下到右上，计算每一个对角线像素为1的个数。

连通区域

连通分量标记算法（贯序）

第一个方法是递归，也就是找到没有标记的像素值为1的点，递归标记邻接点。

第二个方法是贯序，就是从左到右，从上到下扫描。对于当前扫描到的像素点，如果它的像素值为1：

- 1) 上面点和左面点仅有一个有标记，复制
- 2) 上面点和左面点标记相同，复制
- 3) 上面点和左面点标记不同，复制上面点的标记，并记录为等价标记。
- 4) 上面点和左面点都无标记，分配一个新的标记。

区域边界跟踪算法

- (1) 从左到右，从上到下扫描图像，求得起始点。
- (2) c为当前跟踪的像素点：

n₈	n₇	n₆
n₁	c	n₅
n₂	n₃	n₄

从n₁~n₈，找到第一个边界点。则找到的点为新的c。

- (3) 重复上述操作。直到回到第一个边界点。

边缘

模板卷积（能计算卷积）

Origin of Edges

surface normal discontinuity（表面法线不连续）

depth discontinuity（深度不连续）

surface color discontinuity（表面颜色不连续）

illumination discontinuity（光照不连续）

边缘检测的基本思想

边缘检测的目标是找到图像中突变（不连续）的地方。大多数语义和形状的信息都能从边缘信息中获取。

基于一阶的边缘检测（梯度）

(1) Roberts交叉算子

$$G_x = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \quad G_y = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$$

(2) Sobel算子

$$G_x = \begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix} \quad G_y = \begin{bmatrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{bmatrix}$$

(3) Prewitt算子

$$G_x = \begin{bmatrix} -1 & 0 & 1 \\ -1 & 0 & 1 \\ -1 & 0 & 1 \end{bmatrix} \quad G_y = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 0 & 0 \\ -1 & -1 & -1 \end{bmatrix}$$

基于二阶的边缘检测

原理：二阶导数过零点为边缘

-Laplacian算子

$$\frac{\partial^2 f}{\partial y^2} = (f[i+1, j] - 2f[i, j]) + f[i-1, j]$$

$$\nabla^2 \approx \begin{bmatrix} 0 & 1 & 0 \\ 1 & -4 & 1 \\ 0 & 1 & 0 \end{bmatrix}$$

$$\nabla^2 \approx \begin{bmatrix} 1 & 4 & 1 \\ 4 & -20 & 4 \\ 1 & 4 & 1 \end{bmatrix}$$

-LoG算子

先与高斯函数卷积，再求拉普拉斯微分；或先求拉普拉斯微分，再与图像卷积。

0	0	-1	0	0
0	-1	-2	-1	0
-1	-2	16	-2	-1
0	-1	-2	-1	0
0	0	-1	0	0

(拉普拉斯)

为什么高斯：平滑去噪和边缘检测是一对矛盾，应用高斯函数的一阶导数，在二者之间获得最佳的平衡

Canny边缘检测

1.高斯滤波

$$S[i, j] = G[i, j; \sigma] * I[i, j]$$

2.一阶偏导计算梯度和方向

$$G_x[i, j] \approx (S[i, j+1] - S[i, j] + S[i+1, j+1] - S[i+1, j]) / 2$$

$$G_y[i, j] \approx (S[i, j] - S[i+1, j] + S[i, j+1] - S[i+1, j+1]) / 2$$

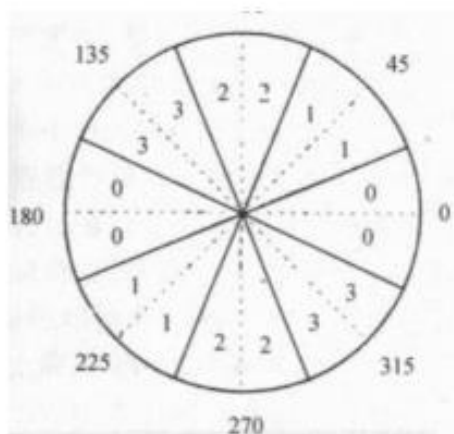
$$M[i, j] = \sqrt{G_x[i, j]^2 + G_y[i, j]^2}$$

$$\theta[i, j] = \arctan(G_y[i, j] / G_x[i, j])$$

3.梯度幅值非极大值抑制

$$\zeta[i, j] = \text{Sector}(\theta[i, j])$$

将方向角离散到四个扇区之一。



$$N[i, j] = \text{NMS}(M[i, j], \zeta[i, j]) \quad (\text{如果 } M[i, j] \text{ 不比沿梯度线方向上的两个相邻点幅值大, 则 } N[i, j] =$$

0)

4.用双阈值检测和连接边缘

(1) 设高低两个阈值

(2) 大于两个阈值：是边缘；大于低阈值小于高阈值，可能是边缘。

在高阈值边缘图中出现断点后，在低阈值边缘图中搜索边缘点。

局部特征

Harris角点检测

$$E(u, v) = \sum_{x, y} w(x, y) [I(x+u, y+v) - I(x, y)]^2$$

$w(x, y)$ 为窗口, I 为函数, u, v 为偏移。

以上公式可近似为:

$$E(u, v) \cong [u, v] M \begin{bmatrix} u \\ v \end{bmatrix}$$

其中:

$$M = \sum_{x, y} w(x, y) \begin{bmatrix} I_x^2 & I_x I_y \\ I_x I_y & I_y^2 \end{bmatrix}$$

推导过程:

$$\begin{aligned} E(u, v) &= \sum_{x, y} w(x, y) [I(x+u, y+v) - I(x, y)]^2 \\ &= \sum_{x, y} w(x, y) [I(x, y) + uI_x(x, y) + vI_y(x, y) - I(x, y)]^2 \\ &= \sum_{x, y} w(x, y) [uI_x(x, y) + vI_y(x, y)]^2 \\ &= [u, v] \left(\sum_{x, y} w(x, y) \begin{bmatrix} I_x^2 & I_x I_y \\ I_x I_y & I_y^2 \end{bmatrix} \right) \begin{bmatrix} u \\ v \end{bmatrix} \end{aligned}$$

我们可以求出M的特征值, 如果两个特征值都很大, 并且比较接近, 那么是角点; 如果一个远大于另一个, 那么是边; 否则是平面区域。

用响应函数R来衡量是否是角点, 计算方法如下 (k 取0.04-0.06) :

$$R = \det M - k (\text{trace } M)^2$$

$$\begin{aligned} \det M &= \lambda_1 \lambda_2 \\ \text{trace } M &= \lambda_1 + \lambda_2 \end{aligned}$$

若 $R > 0$ (大于某一阈值), 则为角点; $R < 0$, 则为边; R 绝对值很小, 则为平面区域。
选取R得到的符合条件点的局部最大值作为结果。

SIFT描述子

1.构建尺度空间，建立图像金字塔。

2.寻找极值点（相邻的26个点中最大/最小值）

3.去除不好的特征点：使用近似的harris corner，检测关键点的位置和尺度，并且去除边缘响应点。

4.16X16的窗口来检测特征。

对每个像素计算边的朝向（梯度角度-90），然后转换到8个方向上，并分割到4X4的网格中。（共128维数据）

剔除不好的边（梯度阈值）。

创建存留边朝向的直方图。

利用关键点邻域像素的梯度方向分布特性为每个关键点指定方向参数，使算子具备旋转不变性。

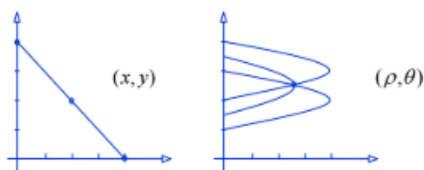
尺度不变的原理

哪怕处在不同的尺度，它们对应的区域是相同的。

Hough变换直线检测

图像中每一点对参数组合进行表决，赢得多数票的参数组合为胜者（结果）。

用极坐标来表示直线，从(x,y)转换到(p,0)空间。



1.量化参数空间（到合适的精度）

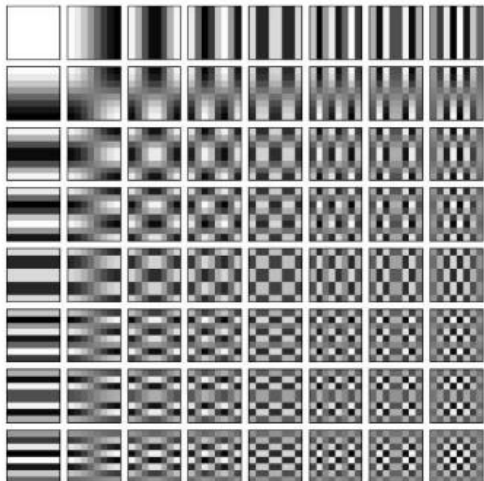
2.初始化累加器为零。

3.对每一点，在其满足的参数方程对应的累加器加1。

4.得到累加器最大值对应模型的参数

图像的傅立叶变换

变换：用正弦来表示，对于二维图像而言，由以下的基图像表示：



低频与高频：亮度灰度剧烈变化的地方是高频（图像边缘和轮廓的度量），对应边缘；变化不大的是低频（图像强度的综合度量），对应大片色块。近处看到的是高频分量，远处观察到的是低频分量。

怎么理解拉普拉斯金字塔的每一层是带通滤波？

拉普拉斯金字塔是将图像下采样后再上采样得到的差值图像。

相减 保留细节 高通

下采样 降噪 低通

相机模型

景深：

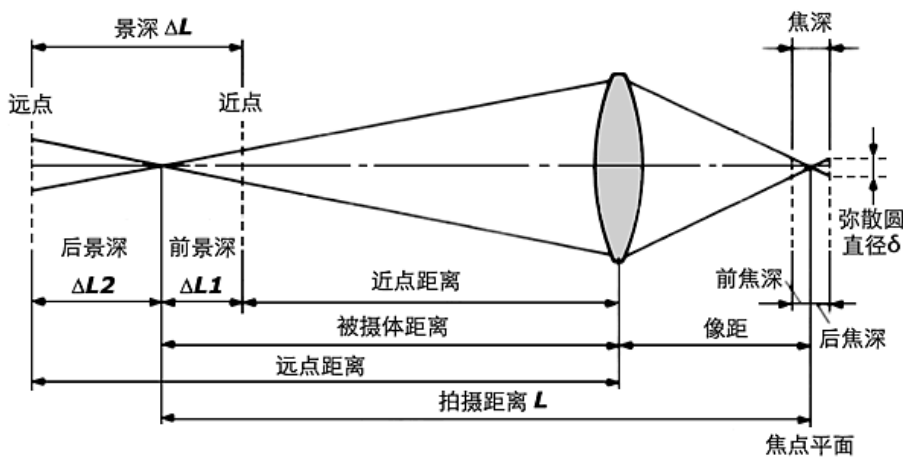
摄像机镜头能够取得清晰图像的成像所测定的被摄物体前后距离范围。

景深大，背景和物体都很清楚；景深小，物体清楚，但背景虚化。

景深随镜头的焦距、光圈值、拍摄距离而变化。对于固定焦距和拍摄距离，使用光圈越小，景深越大。

焦距越短，镜头的视场角越宽。

$$AFOV(^{\circ}) = 2 \times \tan^{-1} \left(\frac{h}{2f} \right)$$



理想的针孔相机 (pinhole camera) 模型

投影公式：

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & -1/d & 0 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} = \begin{bmatrix} x \\ y \\ -z/d \end{bmatrix} \Rightarrow \left(-\frac{x}{z}, -\frac{y}{z} \right)$$

参数：

内参数

Called **intrinsic parameters**

$$(f_x, f_y, c_x, c_y)$$

f_x, f_y 焦距, c_x, c_y 主点坐标

The units

• F : mm

• s_x, s_y : pixel/mm

• f_x, f_y : pixel

其中， dx 和 dy 是 x 和 y 方向一个像素的长度； r 是； u_0 和 v_0 代表图像中心和原点坐标之间横纵相差的像素个数。

人脸识别

主元分析(PCA)

用于数据集降维。

选择一个新的坐标系进行线性降维，使得第一轴上是最大投影方向，第二轴上是第二大投影方向.....以此类推。

$$z = \mathbf{a}^T \mathbf{x} = \sum_{i=1}^d a_i x_i$$

假设原样本数据 \mathbf{x} 投影到一个单位向量 \mathbf{a} 上，投影结果是

(向量的投影: $\mathbf{u}_v = \frac{\mathbf{u} \cdot \mathbf{v}}{|\mathbf{v}|}$ ，由于投影方向为单位向量，则 $\mathbf{u}_v = \mathbf{u} \cdot \mathbf{v}$ ，即 $\mathbf{u}_v = \sum u_i v_i$)

最终得到的投影结果的方差最大，即使得 $\text{var}(z)$ 最大。

$$\begin{aligned} \text{var}(z_a^T) &= E((z_a^T)^2) - [E(z_a^T)]^2 \\ &= E[(\sum_{i=1}^d a_i x_i)^2] - [E(\sum_{i=1}^d a_i x_i)]^2 \\ &= \sum_{i=1}^d a_i a_j E(x_i x_j) - \sum_{i=1}^d a_i a_j E(x_i) E(x_j) \\ &= \sum_{i=1}^d a_i a_j [E(x_i x_j) - E(x_i) E(x_j)] \\ &= \sum_{i=1}^d a_i a_j S_{i,j} \end{aligned}$$

其中 S 代表协方差矩阵。

求最大投影方向：在限制条件下 $\mathbf{a}^T \mathbf{a} = 1$ ，使 $\text{var}(z_a^T) = \mathbf{a}^T S \mathbf{a}$ 最大（拉格朗日乘法）

$$\begin{aligned} L &= \mathbf{a}^T S \mathbf{a} - \lambda (\mathbf{a}^T \mathbf{a} - 1) \\ \frac{\partial L}{\partial \mathbf{a}} &= 0 \Rightarrow S \mathbf{a} - \lambda \mathbf{a} = 0 \end{aligned}$$

经过计算，得到投影方向 \mathbf{a} 是矩阵 S 最大特征根对应的特征向量。

Eigenface

- 1.预处理：根据人眼位置进行裁剪，进行灰度均衡化。
- 2.将二维人脸图像按一行行向量拼成一列，得到列图像；并把所有列图像拼起来，并求出平均人脸。
- 3.求图像的协方差矩阵。
- 4.求协方差矩阵的特征值，以及归一化的特征向量，即为特征人脸。

识别

将两张图像都投影到人脸空间，比较投影向量的欧氏距离。

重构

将图像投影到人脸空间，通过左乘特征人脸空间矩阵恢复。

- 识别

$$\mathbf{y}_f = A^T \mathbf{f}$$

- 重构

$$\hat{\mathbf{f}} = A \mathbf{y}_f$$

光流

光流解决的是什么问题？

评估从H到I的像素运动，给出图像H中的一个像素，找到图像I中相同颜色的相近像素。解决的是像素对应问题。

光流三个基本假设是什么？

亮度恒定性

空间相干性

细微运动

一个点的约束公式会推导

$$0 = I_t + \nabla I \cdot [u \ v]$$

$$\begin{aligned} 0 &= I(x+u, y+v) - H(x, y) \\ &\approx I(x, y) + I_x u + I_y v - H(x, y) \quad \text{shorthand: } I_x = \frac{\partial I}{\partial x} \\ &\approx (I(x, y) - H(x, y)) + I_x u + I_y v \\ &\approx I_t + I_x u + I_y v \\ &\approx I_t + \nabla I \cdot [u \ v] \end{aligned}$$

哪些位置光流比较可靠？为什么？

high textured? corner.

图像分割

基于k-means聚类的图像分割

第一步：任意选择k个sift特征点作为初始聚类质心。

第二步：对于每个sift特征点，计算它们与k个聚类质心的欧式距离，找到最小的那个聚类质心，将该特征点放入此聚类质心集合中。

第三步：对于每个聚类质心集合，用所有元素均值来更新质心。

第四步：比较更新前后聚类质心集合是否相近（距离小于某阈值），相近则完成聚类，否则返回步骤2，如果迭代次数太多聚类失败。

基于Mean Shift的图像分割基本原理与基本思路

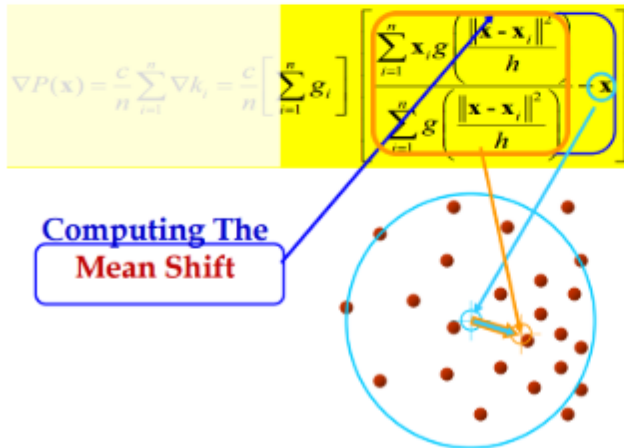
原理：特征值的加权概率分布来描述目标模型

第一步：选择一点 x 为圆心， h 为半径，画圆，得到落在圆中的所有点。

第二步：对于所有点，将其与点 x 相连得到多条向量，求向量的和得到质心：

$$y_{j+1} = \frac{\sum_{i=1}^n x_i g\left(\left\|\frac{x-x_i}{h}\right\|^2\right)}{\sum_{i=1}^n g\left(\left\|\frac{x-x_i}{h}\right\|^2\right)}$$

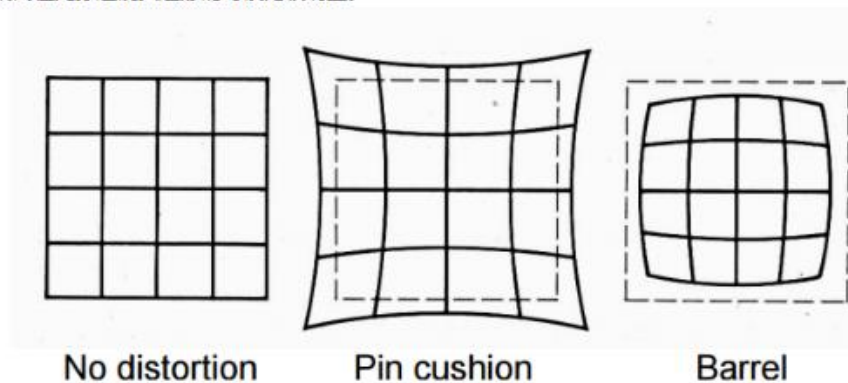
第三步：相加后得到了新的质心转移向量，以新向量的终点为圆心，重复一二步，直到收敛（转移向量接近0）。



相机模型

径向畸变：

由不完美的镜头导致（镜头的形状）；孔的位置
偏离在光穿过镜头边缘的时候比较明显。



切向畸变：

光学元件的偏移

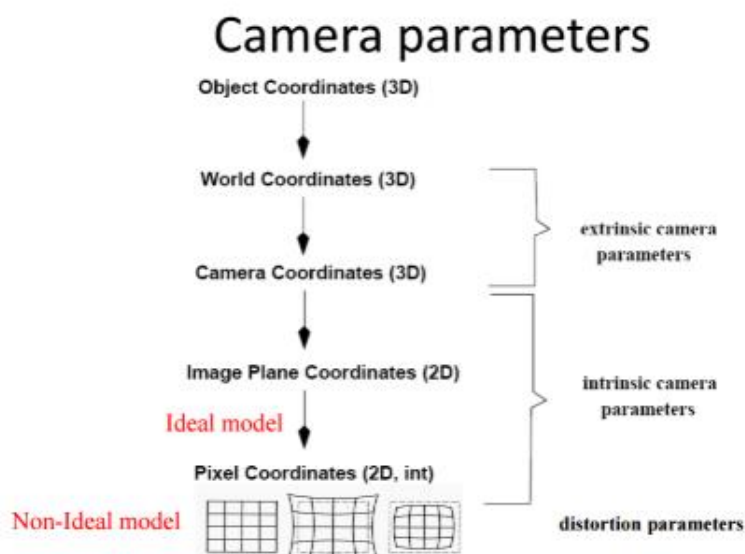
外参有哪几个？ 分别代表什么含义？

Called **extrinsic parameters**

$$(\theta, \varphi, \psi, t_x, t_y, t_z)$$

旋转和平移

内参、外参、畸变参数在成像各阶段中的角色（从三维物体到真实图像的过程）



相机定标

相机定标需要求解哪些参数？

Distortion coefficients, intrinsic para., extrinsic para.

基于Pattern/Reference Object的相机定标– 已知什么？ 求解什么？ – 简述其基本过程哪几个步骤？

已知：N个角点的标定对象，标定对象的K个视角

求解：相机参数，如内参外参和畸变参数

流程：1.标定对象：知道网格角点的位置

2.从图像中找到角点

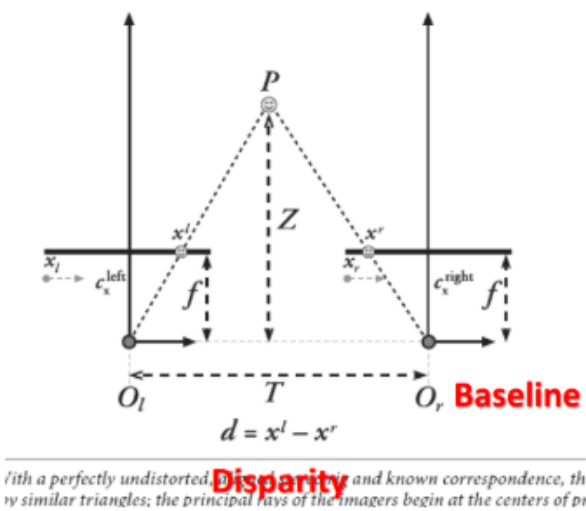
3.建立等式：将图像坐标转换到世界坐标的等式

4.求解公式，得到相机参数

立体视觉

立体视觉的三角测量基本原理(Triangulation 公式)- 会画算 “视差 disparity” 的那张图，并能推导公式

$$\frac{T - (x^l - x^r)}{Z - f} = \frac{T}{Z} \Rightarrow Z = \frac{fT}{x^l - x^r}$$



立体视觉的基本步骤 (review: How toDo Stereo)

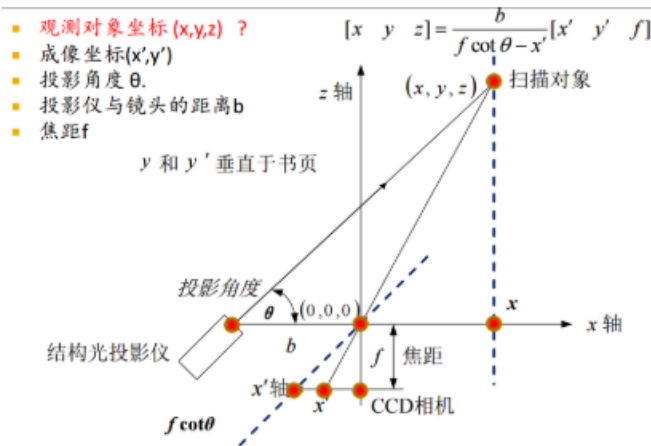
- 1. 标定相机 (消除畸变影响)
- 2. 校正图像
- 3. 计算差距 (disparity)
- 4. 估计深度

三维数据获取

结构光成像系统的构成

结构光投影仪 + CCD相机 + 深度信息重建系统

利用结构光获取三维数据的基本原理- 会画图，会推导公式



ICP算法的作用及其基本步骤

迭代最近点方法 (ICP, Iterative Closest Point)

给定两个三维点集X与Y,将Y配准到X:

- ① 计算Y中的每一个点在X中的对应最近点;
- ② 求使上述对应点的平均距离最小的刚体变换, 获得刚体变换参数 (平移参数与旋转参数);
- ③ 对Y应用上一步求得的刚体变换 (平移与旋转), 更新Y;
- ④ 如果X与Y的对应点对平均距离大于阈值, Goto ①, 否则, 停止计算

物体识别

基于词袋(BoW)的物体分类

- 1.特征提取与表示
- 2.通过训练样本聚类来建立字典
- 3.用字典的直方图来表达一张图像
- 4.根据bag of words来分类未知图像

关于深度学习

基本想法:

- 1.同样被称为representation learning.
- 2.是否存在方法用监督或非监督的方式从数据集中提取有意义的特征
- 3.然后, 通过多层构建使得它有“深度”

一些深度学习方法提出都较早, 但到近几年才兴起, 为什么?

深度学习虽然很早提起, 但由于没有大数据和高性能计算支持, 深度学习无法发挥它的功能; 而这两个特性是随着近几年科技发展才满足条件的。

目前为止, 深度学习特别成功的应用都有哪些?

- 1.微软: 基于深度神经网络的语音识别系统
- 2.Hinton构建的深度神经网络
- 3.word2vector