

13题左右

1 引言

Gestalt Laws 6条

-理解每条意思，能简单解释

1. 接近：人们自然而然的对靠近的事物进行分类
2. 相似：人们容易把相似的物体分为一组
3. 共同命运：对象容易被视为行进在光滑路径上的线条
4. 对称：人们容易将对象视为对称的且围绕一个中心
5. 连续：对象的元素容易群聚在一起且能形成整体
6. 闭合：人们在观察诸如形状、信封、照片时会自动补全它们

Marr视觉表示框架的三个阶段

-Primal Sketch

对输入的原始图像抽取基本特征（角点、边缘、纹理、线条、边界等，即基元图）

-2.5D Sketch

在以观测者为中心的坐标系中，由输入图像和基元图恢复场景可见部分的一些深度信息（深度、法线方向、轮廓等）

-3D Sketch

在以物体为中心的坐标系下，由输入图像、基元图、二维半图，恢复表示识别三维物体

2 二值图像

几何特性

-能举例说出有哪些几何特性

1. 面积（0阶距）
2. 区域中心（1阶距）
3. 方向（最小二乘法）
4. 伸长率 $E = \frac{X_{max}}{X_{min}}$
5. 密集度 $C = \frac{A}{P^2}$ ，P为周长，A为面积
6. 形态比（区域最小外接矩形的长宽比）
7. 欧拉数（连通分量数-洞数）

投影计算

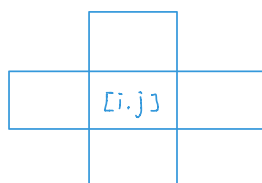
-水平、垂直

水平投影：计算每一列像素为1的个数

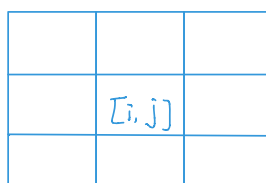
垂直投影：计算每一行像素为1的个数

-知道定义与基本原理，给例子会计算

连通区域



四连通



八连通

-连通分量标记算法（贯序）

1. 从左到右，从上到下扫描图像

2. 若像素点值为1:

1) 若上与左有且仅有一个标记，则复制这一标记

2) 若两点标记相同，则复制这一标记

3) 若两点标记不同，则复制上点标记，并在等价表中将这两个标记记为等价

价

4) 若两点均无标记，则给该像素点分配一个新的标记并记入等价表

3. 在等价表的每一等价集中找到最低的标记

4. 扫描图像，将该最低标记取代与其等价的每一个标记

-区域边界跟踪算法

1. 从左到右，从上到下扫描图像，求区域S的起始点

2. 用c表示当前边界上被跟踪的像素点，使 $c=S(k)$ ，记其左邻点为b

3. 按照逆时针方向记从b开始的c的8个八邻点 $n1 \rightarrow n8$

4. 从b开始，沿逆时针方向找到第一个区域S中的 n_i

5. 置 $c=S(k)=n_i$ ， $b=n_{i-1}$ ——保证b不属于S

6. 重复步骤345，直到 $S(k)=S(0)$

3 边缘

模版卷积

-给一个图像与一个模版，会计算卷积结果

若有 3×3 的像素区域R与卷积核G:

卷积运算: $R_5 = \sum_{i=1}^9 R_i G_i$

Origin of Edges

- 四种最主要的不连续性 (discontinuity)

1. surfaces normal discontinuity
2. depth discontinuity
3. surface color discontinuity
4. illumination discontinuity

边缘检测的基本思想

先确定图像中的边缘像素，然后再把这些像素连接在一起就构成所需的区域边界。函数导数反映图像灰度变化的显著程度，边缘所在位置即一阶导数的局部最大值或二阶导数的过零点。

基于一阶的边缘检测 (有哪些)

* 梯度:

* 幅值: $|G(x,y)| = \sqrt{G_x^2 + G_y^2} \approx |G_x| + |G_y| \approx \max(|G_x|, |G_y|)$

* 方向: $\alpha(x,y) = \arctan(G_y/G_x)$

- 差分近似: $\begin{cases} G_x = f[x+1, y] - f[x, y] \\ G_y = f[x, y] - f[x, y+1] \end{cases} \Rightarrow \begin{cases} G_x = \begin{bmatrix} -1 & 1 \end{bmatrix} \\ G_y = \begin{bmatrix} 1 \\ -1 \end{bmatrix} \end{cases}$

- Roberts交叉算子: $G_x = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \quad G_y = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$

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

- Prewitt算子 (运算较快): $G_x = \begin{bmatrix} -1 & 0 & 1 \\ -1 & 0 & 1 \\ -1 & 0 & 1 \end{bmatrix} \quad G_y = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 0 & 0 \\ -1 & -1 & -1 \end{bmatrix}$

- 均值差分: 对应prewitt, sobel, sethi算子

基于二阶的边缘检测 (有哪些)

- Laplacian算子

$$\nabla^2 f = \frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2} + \frac{\partial^2 f}{\partial z^2}$$

$$\text{如 } \frac{\partial^2 f}{\partial x^2} = f[i, j+1] - 2f[i, j] + f[i, j-1]$$

- LoG算子 (Marr&Hildreth算子): 为什么+G

即高斯平滑+Laplace

$$\Rightarrow h(x,y) = \nabla^2 [g(x,y) * f(x,y)] = [\nabla^2 g(x,y)] * f(x,y)$$

高斯平滑滤波器对图像进行平滑处理，以降低Laplacian操作对于噪声的敏感性

Canny边缘检测

-理解Canny边缘检测方法，能写出该方法的关键步骤，能说出其中两个阈值的意义

1. 高斯滤波平滑图像
 2. 一阶偏导有限差分计算梯度
 3. 对梯度幅值进行NMS非极大值抑制
 4. 双阈值法连接边缘
- * 连接高阈值图边缘，出现断点时在低阈值边缘途中的八邻点中搜索边缘点

双阈值的意义：

- * 高阈值把物体与背景区分开，若过高会导致部分轮廓丢失
- * 低阈值用来平滑边缘轮廓，若太低会产生假边缘

4 局部特征Local Feature

Harris角点检测

-知道基本思想

1. 使用一个固定窗口在图像上进行任意方向上的滑动
2. 比较滑动前与滑动后窗口中的像素灰度变化程度
3. 如果任意方向上的滑动都有较大灰度变化，则可认定该窗口中存在角点

-会推导公式 $E(u, v) \cong [u, v] M \begin{bmatrix} u \\ v \end{bmatrix}$

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

由泰勒公式： $I(x+u, y+v) = I(x, y) + I_x u + I_y v + O(u^2, v^2)$

代入得： $E(u, v) = \sum_{x, y} w(x, y) [I_x u + I_y v + O(u^2, v^2)]^2$

省略高次： $E(u, v) = \sum_{x, y} w(x, y) [I_x u + I_y v]^2$

由： $(I_x u + I_y v)^2 = [u \ v] \begin{bmatrix} I_x^2 & I_x I_y \\ I_x I_y & I_y^2 \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix}$

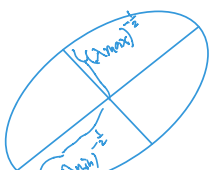
代入得 $E(u, v) = [u \ v] \left(\sum_{x, y} w(x, y) \begin{bmatrix} I_x^2 & I_x I_y \\ I_x I_y & I_y^2 \end{bmatrix} \right) \begin{bmatrix} u \\ v \end{bmatrix}$

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

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

-理解 λ_{\max} , λ_{\min} 两个值的含义，与harris角点的关系

最后得到的Harris公式从几何意义上是一个椭圆，长短轴分别沿着矩阵的两个特征向量的方向



均较大且相当时，图像窗口在所有方向上移动都会产生明显灰度变化，即为角点

-论述对旋转不变性、灰度仿射不变性、尺度不变性的情况

✓

✗

✗

SIFT描述子的计算

-Full version的基本计算步骤

1. 构建尺度空间，建立图像金字塔
2. 在相邻的26个点中寻找极值点
3. 使用近似的harris角点检测去除不好的特征点，检测关键点的位置和尺度，并且去除边缘响应点
4. 用16×16的窗口放在特征点附近，并分成16个4×4的窗口
5. 计算窗口中每个像素的边方向
6. 去掉方向能量小的边
7. 将每个小窗口中的所有方向里三成8个方向，共128个

-为什么使用梯度信息？好处？

可以显示边缘信息，并在光照变化时有抵抗能力

-如何实现旋转不变？

旋转时每一个关键点周围的点也在跟着旋转，不会影响SIFT向量，在计算grid里面的梯度bin时需要旋转到主方向

尺度不变的原理

金字塔模型对每一种尺度都能实现检测，通过进一步算LoG算子得到的尺度来计算feature的范围，因此不同尺度都能得到类似的feature

5 曲线

Hough变换

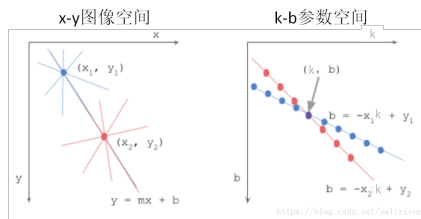
-用来解决什么问题

基于投票原理的参数估计方法，一种重要的形状检验技术

-基本思想

图像中的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

-会用图示解释Hough变换做直线检测的具体原理



如果x-y图像空间中有很多点在k-b空间中相交于一点，那么这个交点就是我们要检测的直线。

-对于直线检测或圆的检测，能写出算法的基本步骤

1. 适当量化参数空间
2. 对每个待拟合的点，其满足的参数方程对应的累加器+1
3. 所有累加器中最大值对应的参数组合即为模型参数

直线：参数组合 (r, θ)

圆弧：参数组合 (a, b)

$$\begin{cases} x = a + r \cos \theta \\ y = b + r \sin \theta \end{cases} \quad \theta \text{ 为边缘点梯度}$$

-参数空间离散化，精度低或高分别有什么不好？

精度过低：检测不准

精度过高：参数空间的峰值扩散，增大计算量

6 主元分析与人脸识别

主元分析 (PCA)

-PCA方法的基本思想、主要作用

用于数据集降维，选择一个新的坐标系进行线性降维，使得第一轴上时最大投影方向，第二轴上是次大投影方向……

-什么样的数据用PCA会比较有效？多元高斯——重构误差比较小

统计意义上有强相关性的数据

-优化目标函数的推导

有d维空间 $x = (x_1, \dots, x_d)$ ，投影方向 $a_i = (a_{i1}, a_{i2}, \dots, a_{id})^T$ 且 $a_i a_i^T = 1$ ，投影值为 $z_i = a_i^T x = \sum a_{ij} x_j$

目标即最大化 $\text{var}(z_i) = a_i^T S a_i$ ，其中 $S = E(X_i Y_i) - E(X_i) E(Y_i) = \text{Cov}(X_i, Y_i)$

-关于选取多少个特征向量构建子空间，常用什么方法？

1. 运用统计学方法，衡量单个特征与响应变量之间的关系
2. 基于机器学习的特征选择

Eigenface

-“Eigenface”是什么

利用特征脸进行人脸识别的方法

-Eigenface人脸识别方法的基本步骤

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行，得到列图像。并把所有列图像拼起来，并求出平均人脸
3. 求出图像的协方差矩阵
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

-将重构用于人脸识别的原理

- 识别：将两张图像都投影到人脸空间比较投影向量的欧式距离
- 重构：将图像投影到人脸空间，通过左乘特征人脸矩阵恢复

7 图像频域与图像分析

图像的傅立叶变换

-理解变换的基本含义

傅立叶变换是一种函数从空间域到频率域的变换，任何信号包括图像信号都可以变成一系列正弦信号的叠加，在图像领域就是将图像brightness variance作为正弦变量。傅立叶变换储存每一个频率的magnitude和phase信息，前者代表这个频率上有多少信号，后者表示空间信息

-理解图像的低频成分与高频成分

低频分量：图像变化平缓的部分，也就是图像轮廓信息

高频分量：图像突变部分，某些情况下指图像边缘信息/噪声

图像分解

-从图像分解角度，理解傅立叶变换的意义

-怎么样理解拉普拉斯图像金字塔的每一层都是带通滤波

拉普拉斯金字塔是通过源图像减去先缩小再放大的一系列图像构成的，下采样的时候丢失了高频滤波，相邻金字塔相减时丢失低频信息，因此只有中间频段的信息保留了下来

8 图像拼接Image Stitching

RANSAC

-Generally speaking, 可以解决什么样的问题

3D重建中的位置确定问题

-理解其过程的核心思想

1. 内群数据可以通过几组模型的参数来叙述其分布，离群数据则是不适合模型化的数据

2. 数据会受到噪声的影响，噪声指离群数据

3. RANSAC假定：给定一组很小的内存，存在一个程序，可以估算最适用于这一组数据模型的参数

-优点？

适用性强，易于实现

-基本步骤（迭代Loop）

1. 在数据中随机选择几个点设为内群

2. 拟合内群的模型

3. 把刚才没选到的点带入模型中，计算是否为内群

4. 记录内群数量并重复上述步骤

5. 哪次计算中内群数量最多，就是所要求的解

-outlier点比例给定的情况下，则k次采样（迭代）后计算成功的概率是？

$$1 - (1 - w^n)^k$$

图像拼接

-实现两张图像自动拼接的基本步骤

找出两张图像中相似的点（至少四个，homography矩阵的计算至少需要4个点），然后计算一张图片可以变换到另一张图片的变换矩阵，用这个矩阵把那张图片变换后放到另一张图的相应位置，并重新计算图像重叠部分的像素值

9 物体识别

Visual Recognition

-基本任务大概可以分为哪几大类

1. 图片分类

2. 检测和定位物体/图像分割

3. 估计语义和几何属性

4. 对人类活动和事件进行分类

-都有哪些挑战因素，简单举例

视角变化、光线变化、尺度变化、物体形变、物体遮挡、背景凌乱、内部类别多样

基于词袋（BoW）的物体分类

-图像的BoW (bag-of-words) 是指什么意思?

图像中的单词被定义为一个图像块的特征向量, 图像的Bow模型即所有图像块的特征向量得到的直方图

-基本步骤

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

基于卷积全局优化的物体分类

-会简单推导并理解含义

每张图是1024个像素的, 因此我们需要求得一个 1024×10 的矩阵 W , 相乘后得到 $y' = Wx$, 表示每个类别的概率。因此我们希望 y_i 都在0-1之间, 且十个之和为1。有softmax函数满足这个特点,

$$a(x)_i = \frac{e^{x_i}}{\sum_j e^{x_j}}$$

因此修正 $y' = SM(Wx)$ 。

我们还希望自己预测的结果 y' 尽可能接近真实的分类结果 y , 因此就有

$$C_{pred} = \arg \max (y')$$

此处增加一个loss函数 $L = -\log(p_c)$

其中 p_c 是我们对正确类别的预测。

取 W 为:

$$W^* = \arg \min_W (-\sum_{xy} \log(p_c))$$

对所有 x 原图与 y 相应分类的 W 的loss求和, 去一个loss最小时候的 W

10 深度学习

深度学习

-怎么理解被称为end-to-end的学习?

直接输入原始数据, 让模型自己去学习特征, 最后输出结果。

-神经网络的学习, 数学本质上是求解什么?

通过参数求最佳解的过程

-会写出基于梯度下降法的学习框架

1. 给定待优化连续可微函数 $J(\theta)$, 学习率 α 与一组初始值 $\theta_0 = (\theta_{01}, \theta_{02}, \dots, \theta_{0n})$
2. 计算待优化函数梯度 $\nabla J(\theta_0)$
3. 更新迭代公式 $\theta^{(1)} = \theta_0 - \alpha \nabla J(\theta_0)$
4. 计算 $\theta^{(1)}$ 处函数梯度 $\nabla J(\theta_1)$
5. 由 $\|\nabla J(\theta)\| \leq \epsilon$ 判断算法收敛
6. 若收敛则停止, 否则继续迭代.

CNN

-理解卷积层与Pooling层的作用

卷积层：构建卷积神经网络的核心层，产生了大部分的计算量

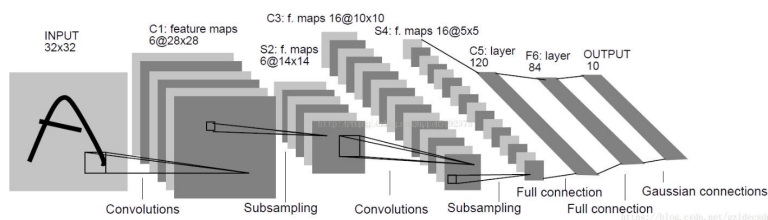
Pooling层：在连续的卷积层之间周期性插入一个pooling层，可以逐渐降低数据体的空间尺寸，从而减少网络中的参数数量，减少计算资源耗费，有效控制过拟合

-会理解第一个卷积层的各种weight个数（注意：权重数与连接数的差别）

特征图大小计算： $feature_map = (input_size + padding * 2 - 卷积核边长 + 步长) / 步长$

参数个数 = (卷积核大小 + 步长) * 该层卷积核个数

链接数量 = 【(卷积核大小 + 步长) * 该层卷积核的个数】 * $feature_map$



BP算法

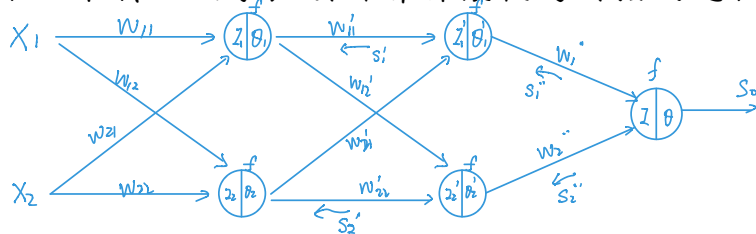
-BP算法作用是求解什么

使用梯度下降法，通过反向传播来不断调整网络的权重和阈值，使网络误差平方和最小，求解局部最好的权重 w

-理解“梯度下降法”与BP算法的关系

BP算法中需要用到梯度下降法，用来配合反向传播，BP算法就是提供了给梯度下降法所需要的所有值

-给一个具体例子，会计算梯度反向传播的过程



$$\begin{aligned} z_1 &= w_{11}x_1 + w_{12}x_2 \\ z_2 &= w_{21}x_1 + w_{22}x_2 \\ z'_1 &= w'_{11}\theta_1 + w'_{12}\theta_2 \\ z'_2 &= w'_{21}\theta_1 + w'_{22}\theta_2 \end{aligned} \Rightarrow \begin{cases} \theta_1 = \text{sig}(z_1) = \frac{1}{1 + e^{-z_1}} \\ \theta_2 = \frac{1}{1 + e^{-z_2}} \\ \theta'_1 = \frac{1}{1 + e^{-z'_1}} \\ \theta'_2 = \frac{1}{1 + e^{-z'_2}} \\ \theta = \frac{1}{1 + e^{-z}} \end{cases} \Rightarrow \begin{aligned} s'_1 &= -\frac{\partial e}{\partial \theta_1} \\ &\vdots \\ s'_2 &= -\frac{\partial e}{\partial \theta_2} \\ s'_0 &= \frac{\partial e}{\partial \theta} = (\theta - y) \end{aligned}$$

步长为 1。

$$\text{如 } \frac{\partial e}{\partial w_i} = \delta \circ f'(t) \cdot \theta_i' \Rightarrow w_i' = w_i - \eta \delta \circ f'(t) \cdot \theta_i'$$

11 光流

光流解决的什么问题

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

$$0 = I(x+u, y+v) - I(x, y)$$

$$\approx [I(x, y) - H(x, y)] + uI_x + vI_y$$

$$\approx I_t + uI_x + vI_y$$

$$\approx I_t + \nabla I[u, v]$$

哪些位置光流比较可靠？为什么？

纹理复杂区域，因为梯度比较大且方向不同，求出来的特征值比较大

12 相机模型

理解：景深/光圈/焦距/视场

-景深？视场？

景深：在镜头前能取得清晰图像的成像所测定的被摄物体前后距离

视场：摄像头能够观察到的最大范围

-光圈对景深的影响？理解原理

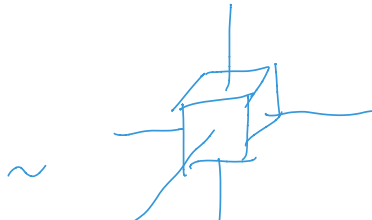
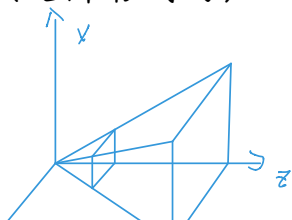
光圈越大，景深越小。光圈改变夹角大小的同时，改变了弥散圆的大小

-焦距对视场的影响？理解原理

焦距越大，视场越小。焦距越大离得越近，整个场景被缩短了，远处的东西被拉到近处而且很大，但被虚化了

理想的针孔相机模型

-基本投影公式，并能画图说明，会推导写出齐次坐标式下的透视投影公式（矩阵形式的）



$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & -\frac{1}{d} & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix} = \begin{pmatrix} X \\ Y \\ -\frac{Z}{d} \\ 1 \end{pmatrix} \quad \left(-d\frac{X}{Z}, -d\frac{Y}{Z}\right)$$

-齐次坐标表示的好处

解决了笛卡尔坐标系无法表示无穷远点的问题

-相机模型有哪几个内参（不包括畸变参数），会写内参矩阵

4个内参

$$M = \begin{pmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{pmatrix}$$

畸变

-径向畸变与切向畸变各是什么原因引起的？

径向畸变：光线在远离透镜中心的地方比靠近中心的地方更加弯曲

切向畸变：透镜不完全平行于图像平面

-径向畸变常见的有哪两种？

桶形畸变

枕形畸变

外参有哪几个？分别表示什么含义？齐次坐标下的外参矩阵会写、会推导

6个，描述了相机在世界坐标系中的位置和朝向

• 3个旋转参数 $R_{3 \times 3}$

• 3个轴位移参数 $t_{3 \times 1}$

$$P = \begin{pmatrix} R_{3 \times 3} & t_{3 \times 1} \\ 0 & 1 \end{pmatrix} P_0$$

画图展示内参、外参、畸变系数在成像各阶段中的角色（从真实的世界坐标到图像坐标的过程）

外参数：世界坐标 \rightarrow 相机坐标

内参数：相机坐标系 \rightarrow 像平面坐标系

畸变系数：非理想模型 \rightarrow 像素

13 相机标定

一般的相机定标

-需要求解哪些参数

畸变参数，内参，外参

-解决这个问题的基本思想是什么?

通过标定板，可以得到 n 个对应的世界坐标三维点 X_i 和对应的图像坐标二维点 x_i ，这些三维点到二维点的转换都可以通过上面提到的相机内参 K ，相机外参 R 和 t ，以及畸变参数 D ，经过一系列的矩阵变换得到。

基于Homography的相机定标

- 有哪些优点?

可以很清楚确定一个点是否在直线/平面上

可以描述无穷远点: $(x, y, 0)$

更简洁的表达欧氏空间变换，可以把平移和旋转写到一个矩阵里

方便表达直线与直线，平面与平面的交点

能够区分一个向量和一个点

-简述基本过程 (4个步骤)

1. calibration target: 2 planes at right angle with checkerboard.

2. we know positions of corners of grid with respect to a coordinate system of the target.

3. obtain from images the corners.

4. using the equations we obtain the camera parameters

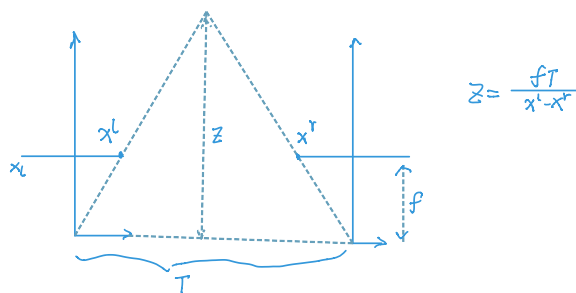
-Homography矩阵有几个自由度? 求解需要至少几个特征点?

8个自由度, 至少4个特征点

14 立体视觉

立体视觉的三角测量基本原理 (Triangulation公式)

-会画视差disparity的那张图，并能推导公式



立体视觉的基本步骤 (review: How to Do Stereo)

1. 消除畸变
2. 矫正相机
3. 两幅图中找到相同特征
4. 三角测量

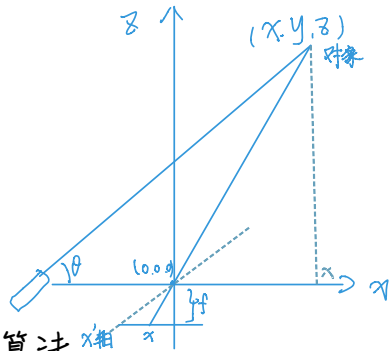
15 结构光三维成像原理

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（不要与disparity图混淆）



$$[x, y, z] = \frac{b}{f \cos \theta - x'} [x', y', f]$$

ICP算法

-要解决什么问题

多个相机配准问题

-基本步骤

1. 建立两个扫描结果之间的对应关系
2. 通过迭代获得一个仿射变换函数能够描述1中的变化关系
3. 对Y应用上一步求得的仿射变换，更新Y
4. 两个结果中距离最近的点作为对应点，计算对应点距离如果大于阈值，重复

23, 否则停止