

review

- Gestalt Laws （格式塔法则）

格式塔法则

1. 图像与背景关系原则（figure-ground）：物体/图形比背景更突出
2. 接近原则（proximity）：接近/邻近的物体会被认为是一个整体
3. 相似原则（similarity）：刺激物的形状/大小/颜色/强度等物理属性方面相似时，刺激物被认为是一个整体
4. 连续性原则（continuity）：若图形的某些部分可以被看作连接在一起的，则这些部分很可能被认为是一个整体
5. 封闭/闭合原则（closure）：对于有些没有闭合的图形，主体能自行填补缺口使之被认为是一个整体
6. 蕴含律：对复杂对象进行感知时，人们倾向于把对象看作对称的，简单的，规则的图形

- Marr 视觉表示框架的三个阶段？

1. Primal Sketch：处理输入的原始图像，抽取**基本特征**（角点、边缘、纹理、线条、边界etc.）——特征的集合称为**基元图**
2. 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etc.）——不是真正的物体三维表示
3. 3D Model：在以**物体为中心**的坐标系中，由输入图像、基元图、二维半图，恢复表示识别三维物体

边缘

1. 模板卷积

给一个图像和一个模板，会计算卷积结果

2. Origin of edges

四种最主要的不连续（discontinuity）

- 图像深度不连续处
- 图像（梯度）朝向不连续处
- 图像光照不连续处
- 纹理变化处

3. 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，一阶导数的局部极大值，或二阶导数的过零点。

4. 基于一阶的边缘检测（有哪些）

Roberts交叉算子，Sobel算子，Prewitt算子（运算较快）

5. 基于二阶的边缘检测（有哪些）

Laplacian 算子

LoG算子（Marr&Hildreth算子）：为什么要加G

对噪声敏感，因此先高斯滤波去噪，再用拉氏算子进行图像的边缘检测。

6. Canny边缘检测

理解Canny边缘检测方法，能写出该方法关键步骤，能说出其中两个阈值的含义

- a. 高斯滤波器平滑图像
- b. 用一阶偏导有限差分计算梯度幅值和方向
- c. 对梯度幅值进行非极大值抑制（NMS）
- d. 用双阈值算法检测和连接边缘
 - 得到高阈值图 $N[i, j] > T_2$ ，低阈值图 $N[i, j] < T_1$
 - 连接高阈值边缘；出现断点时在低阈值边缘图中的八邻点域搜索边缘点

7. 非极大值抑制如何起作用（NMS）

- 去掉幅值局部变化非极大的点。
 - 将梯度角离散为圆周的四个扇区之一，作抑制运算
 - 方向角离散化： $\xi[i, j] = \text{Sector}(\theta[i, j])$ ，按照扇区进行离散化。
 - 抑制，得到新幅值图： $N[i, j] = \text{NMS}(M[i, j], \xi[i, j])$
 - 若 $M[i, j]$ 不必沿梯度线方向上的两个相邻点的幅值大，则 $N[i, j] = 0$ 。（梯度就没有了）

局部特征

1. Harris 角点检测

- a. 知道basic idea/基本思想

原理：在图像对指定大小窗口进行各个方向的平移，观测窗口内图像的相似程度

- Flat: 窗口内图像基本无变化
- Edge: 沿Edge平移窗口时，窗口内图像无变化
- Corner: 各方向平移时都有较大变化

b. 推导

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

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

$$E \cong [u \quad v] M \begin{bmatrix} u \\ v \end{bmatrix}$$

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

$$\det M = \lambda_1 \lambda_2$$

$$\text{trace} M = \lambda_1 + \lambda_2$$

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

- $R > 0$ (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。

c. 理解 λ_{max} 和 λ_{min} 两个值的含义，与harris角点关系？

$E(u, v)$ 事实上是一个二次型。在水平方向上切片 (slice)，会得到椭圆。

我们可以对矩阵 M 进行特征值分解，得到两个特征向量 λ_1, λ_2 (或称 $\lambda_{max}, \lambda_{min}$)，其中大的相当于椭圆长轴，小的相当于短轴。

(在角点的位置上，长轴和短轴差别不大；在非角点，则椭圆会很扁。决定Harris Corner成功的因素在 λ_{min} 上， λ_{min} 要比较大才能够是一个合格的Harris Corner)

- d. 论述对旋转不变性、灰度仿射不变性、尺度不变性的情况
Harris Corner的角点能保持旋转、平移、亮度变化等的情况，但不能解决尺度带来的影响，
因为窗口的大小是固定的。因此其对缩放非常敏感。

2. SIFT描述子的计算

a. Full version 计算的步骤

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. 用 16x16 的窗口放在特征点附近
5. 将 16x16 分成 16 个 4x4 的窗口
6. 计算窗口中每个像素的边的方向（梯度角减去 90° ）
7. 丢掉方向能量小的边（使用阈值）用直方图描述结果
8. 将每个小窗口中的所有方向离散成 8 个方向，一共 $16 \times 8 = 128$ 个

Full version

- Divide the 16x16 window into a 4x4 grid of cells (2x2 case shown below)
- Compute an orientation histogram for each cell
- 16 cells * 8 orientations = 128 dimensional descriptor

b. 为什么使用梯度信息？好处？

因为梯度信息可以表示边缘信息，并且在光照变化时有抵抗能力

c. 如何实现旋转不变的？

旋转不变：旋转的时候每一个关键点周围的点也会跟着旋转，不会影响SIFT向量。所以SIFT对旋转不敏感（在计算grid里面的梯度bin前需要旋转到主方向，因此有了一定的旋转不变性）

d. 尺度不变的原理

尺度不变：金字塔模型，对每一种尺度都能进行检测，所以具有尺度不变性（通过前一步算LoG得到的尺度来确定计算feature的范围，所以不同尺度能得到类似的feature。）

曲线

1. Hough 变换

a. 用来解决什么问题？

形状检测，相比较前面的方法，解决一张图片中有多个形状，比如多个直线、多个圆的形状检测问题

b. 基本思想

对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。

c. 会用图示解释Hough变换做直线检测的具体原理

直线检测的Hough变换： $y = mx + c \Rightarrow c = -mx + y$ ，以 (x, y) 为自变量， (m, c) 为因变量，每个点 (x, y) 对应空间 (m, c) 上的一条直线 —— 这些经过同一个点，这个点对应一个 (m, c) 的组合。

避免垂直直线所带来的问题，可以采用极坐标表示。将 (x, y) 变换到 (ρ, θ) 。

d. 对于直线检测或圆的检测，能写出算法的基本步骤

算法步骤：

1. 适当地量化参数空间（合适的精度即可）；
2. 假定参数空间的每一个单元都是一个累加器，把累加器初始化为零；
3. 对图像空间的每一点，在其所满足的参数方程对应的累加器上加1；
4. 累加器阵列的最大值对应模型的参数。

圆弧检测：

$x = a + r \cos \theta, y = b + r \sin \theta \Rightarrow b = a \tan \theta - x \tan \theta + y$ ，得到一个关于 a, b 的参数空间。

人脸识别

1. 主成分分析（PCA）

a. PCA方法的基本思想、主要作用

最大化投影后所有点的方差

b. 什么样的数据使用PCA会比较有效？

随机变量分布接近于高斯分布的，PCA降维效果更好

c. 优化目标函数推导

PCA定义

- d-维空间 $\mathbf{x} = (x_1, x_2, \dots, x_d)$
- 投影方向 $\mathbf{a}_1 = (a_1^1, a_1^2, \dots, a_1^d)^T$ where $\mathbf{a}_1^T \mathbf{a}_1 = 1$
- 投影值 $z_1 = \mathbf{a}_1^T \mathbf{x} = \sum_{i=1}^d a_1^i x_i$
- 问题
 - 最大化 $\text{var}(z_1)$
 - 求投影方向 $\arg \max_{\mathbf{a}_1} \text{var}(z_1)$

PCA求解

$$\begin{aligned}
 \text{var}(z_1) &= E(z_1^2) - [E(z_1)]^2 = E\left[\left(\sum_{i=1}^d a_1^i x_i\right)^2\right] - \left(E\left(\sum_{i=1}^d a_1^i x_i\right)\right)^2 \\
 &= \sum_{i,j=1}^d a_1^i a_1^j E(x_i x_j) - \sum_{i,j=1}^d a_1^i a_1^j E(x_i) E(x_j) \\
 &= \sum_{i,j=1}^d a_1^i a_1^j [E(x_i x_j) - E(x_i) E(x_j)] \\
 &= \sum_{i,j=1}^d a_1^i a_1^j S_{ij} \quad \text{令 } S_{ij} = E(x_i x_j) - E(x_i) E(x_j) \\
 &= \mathbf{a}_1^T \mathbf{S} \mathbf{a}_1 \quad \mathbf{S} ? \quad \text{cov}(x_i, x_j) = E(x_i x_j) - E(x_i) E(x_j)
 \end{aligned}$$

PCA求解

最大化 $\text{var}(z_1) = \mathbf{a}_1^T \mathbf{S} \mathbf{a}_1$ *subject to* $\mathbf{a}_1^T \mathbf{a}_1 = 1$

典型的等式约束下的最优化问题，可以用Lagrange乘子法
设 λ 为Lagrange乘子，则转为最大化

$$\mathbf{a}_1^T \mathbf{S} \mathbf{a}_1 - \lambda(\mathbf{a}_1^T \mathbf{a}_1 - 1)$$

求微分，得必要条件：

$$\mathbf{S} \mathbf{a}_1 - \lambda \mathbf{a}_1 = 0 \implies \mathbf{S} \mathbf{a}_1 = \lambda \mathbf{a}_1$$

PCA求解

最大化 $\text{var}(z_1) = \mathbf{a}_1^T \mathbf{S} \mathbf{a}_1$ *subject to* $\mathbf{a}_1^T \mathbf{a}_1 = 1$

$$\implies \text{var}(z_1) = \mathbf{a}_1^T \lambda \mathbf{a}_1 = \lambda \mathbf{a}_1^T \mathbf{a}_1 = \lambda$$

为了使 $\text{var}(z_1)$ 取得最大值，必须用最大特征值对应的特征向量！

2. Eigenface

a. “Eigenface”是什么？

将人脸图像进行预处理、归一化后，用PCA计算得到的特征向量，也叫特征人脸 Eigenface

用这些Eigenface构成图像空间变换的基

b. Eigenface人脸识别方法的基本步骤

Eigenface步骤

1. 获得人脸图像的训练集，通常为整个人脸数据库；
2. 对所有人脸图像作归一化处理；
3. 通过**PCA**计算获得一组特征向量(特征脸)。通常一百个特征向量就足够；
4. 将每幅人脸图像都投影到由该组特征脸张成的子空间中，得到在该子空间坐标；
5. 对输入的一幅待测图像，归一化后，将其映射到特征脸子空间中。然后用某种距离度量来描述两幅人脸图像的相似性，如欧氏距离。

c. 将重构用于人脸检测的原理

- 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。

图像拼接 Image Stitching

1. RANSAC

a. Generally speaking，可以解决什么样的问题？

从一组含有Outlier的数据中正确估计数学模型参数的算法

b. 理解其过程的核心思想

inlier数据可以通过一组合适的模型参数描述其分布，但是outlier数据偏离正常范围

c. 基本步骤（迭代Loop）

RANSAC loop:

1. 先随机选择一组种子点，开始做最基础的transformation的估计
2. 假设该transformation是正确的，找落在该transformation的inlier
3. 根据已有的inlier，重新计算transformation，使得变换得到更新
4. 重复2-3步，直到inliers最多

d. outlier点比例给定的情况下，则次采样（迭代）后计算成功的概率是？

RANSAC: How many samples?

- How many samples are needed?
 - Suppose w is fraction of inliers (points from line).
 - n points needed to define hypothesis (2 for lines)
 - k samples chosen.
- Prob. that a single sample of n points is correct: w^n
- Prob. that all k samples fail is: $(1 - w^n)^k$

⇒ Choose k high enough to keep this below desired failure rate.

e. 思考：与Hough变换有什么共同之处？

2. 图像拼接

a. 实现两张图片自动拼接的主要步骤

1. 检测关键点
2. 建立SIFT描述子
3. SIFT特征匹配
4. 根据匹配的特征点对计算变换矩阵

- 变换矩阵 $T = \begin{bmatrix} t_{11} & t_{12} & t_{13} \\ t_{21} & t_{22} & t_{23} \\ 0 & 0 & 1 \end{bmatrix}$ —— 一组点对提供两个方程 ⇒ 需要至少三个点对
- 使用RANSAC提高求解准确度

5. 图像混合：高斯金字塔 → 拉普拉斯金字塔 → 左右各一半 → 上采样恢复图像

物体识别

1. Visual Recognition

a. 基本任务可以分为哪几大类？

Visual Recognition

- Algorithms that have the capability to:
 - Classify images or videos
 - Detect and localize objects
 - Estimate semantic and geometrical attributes
 - Classify human activities and events

b. 都有哪些挑战因素？

视角变化、光照变化、尺度变化、形变、遮挡、背景凌乱、内容类别多样

2. 基于词袋 (BoW) 的物体分类

a. 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，图的Bow模型即图像中所有图像块的特征向量得到的直方图

b. 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 (codewords dictionary)

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

深度学习

1. 深度学习

a. 怎么理解被称为端到端 (end-to-end) 的学习？

在学习过程中不进行分模块或分阶段进行训练，直接优化任务的总体目标。

b. 神经网络的学习，数学本质上是求解什么？

求解一组label使得交叉熵损失最小

c. 常用的基本方法？

反向传播、梯度下降、pooling

2. CNN

a. 理解卷积层与Pooling层的作用

卷积层：用于提取图像中的特征

池化层：筛选特征，逐渐降低数据体的空间尺寸，这样的话就能减少网络中参数的数量，使得计算资源耗费变少，也能有效控制过拟合。

b. 会计算第一个卷积层的各种weight个数

▪ 输出图像大小计算

记输出为 $M \times M$ 的feature map，输入为 $N \times N$ ，卷积核为 $K \times K$ ，步长为 S ，padding为 P ，则有 $M = (N - K + 2P) / S + 1$

▪ 权重个数和神经元数目计算：

1. 每一次 $K \times K$ 区域内的卷积都对应一个神经元，因此每个神经元有 K^2 个权重，1个偏置——多通道时权重要**乘上通道数**
2. 总神经元数为 M （输出图像大小），连接数为 $(K^2 + 1) \times M$ （此处未考虑多通道）——一般使用多个卷积核，则神经元个数乘上卷积核数

- 更新的参数个数：CNN中一个卷积核下对应的所有神经元**共享参数**，因此需要更新的参数个数为 $(K^2 + 1) \times \text{通道数} \times \text{卷积核数}$

3. BP算法

a. 知道BP算法的作用

计算偏导即梯度

b. 理解“梯度下降法”与BP算法的关系

BP算法是用来计算损失函数相对于神经网络参数的梯度。而梯度下降法是一种优化算法，用于寻找最小化损失函数的参数。

c. 给一个具体例子，会计算梯度反向传播的过程

Backpropagation: a simple example

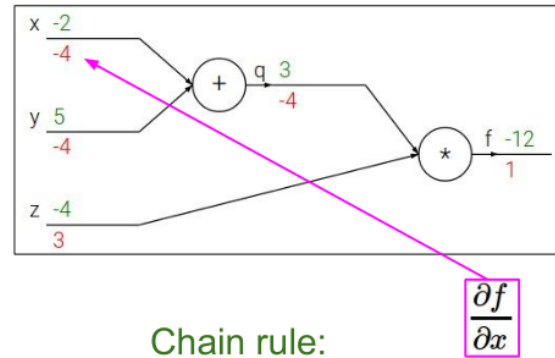
$$f(x, y, z) = (x + y)z$$

e.g. $x = -2, y = 5, z = -4$

$$q = x + y \quad \frac{\partial q}{\partial x} = 1, \frac{\partial q}{\partial y} = 1$$

$$f = qz \quad \frac{\partial f}{\partial q} = z, \frac{\partial f}{\partial z} = q$$

Want: $\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}, \frac{\partial f}{\partial z}$



光流

1. 光流解决的是什么问题？

运动跟踪

2. 光流的三个基本假设是什么？

- 亮度恒常性：目标像素强度在相邻帧不发生变化—— $I(x + u, y + v, t + 1) = I(x, y, t)$
- 空间一致性：相邻像素拥有相似运动
- 时间规律：相邻帧的时间间隔足够短

3. 一个点的约束公式会推导：

1. 假设一个目标像素在 t 时刻亮度为 $I(x, y, t)$ ， $t + \delta t$ 时刻亮度为 $I(x + u, y + v, t + \delta t)$ ，则两者相等
2. Taylor展开得 $I_x u + I_y v + I_t = 0, I_x = \frac{\partial I}{\partial x}, I_y = \frac{\partial I}{\partial y}, I_t = \frac{\partial I}{\partial t}$ ，即 $-I_t = \nabla I \cdot \begin{pmatrix} u \\ v \end{pmatrix}$

4. 哪些位置光流比较可靠？为什么？

角点处的光流能够通过角点邻域完全确定下来，因此角点处的运动信息最为可靠；
使用纹理复杂区域，梯度比较大且方向不同，求出来的特征值比较大

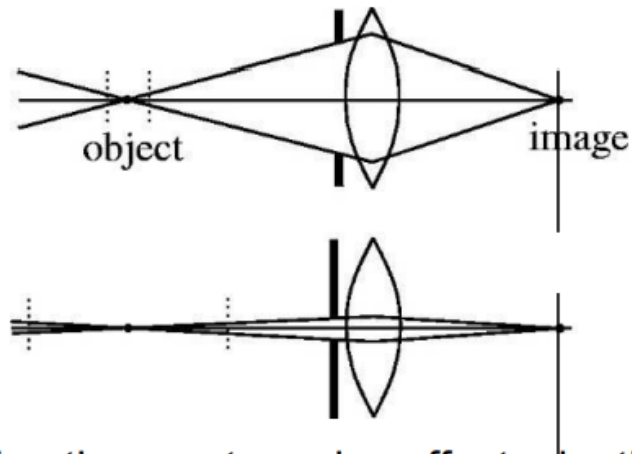
相机模型

1. 理解：景深、光圈、焦距、视场

a. 光圈对景深的影响，理解原理

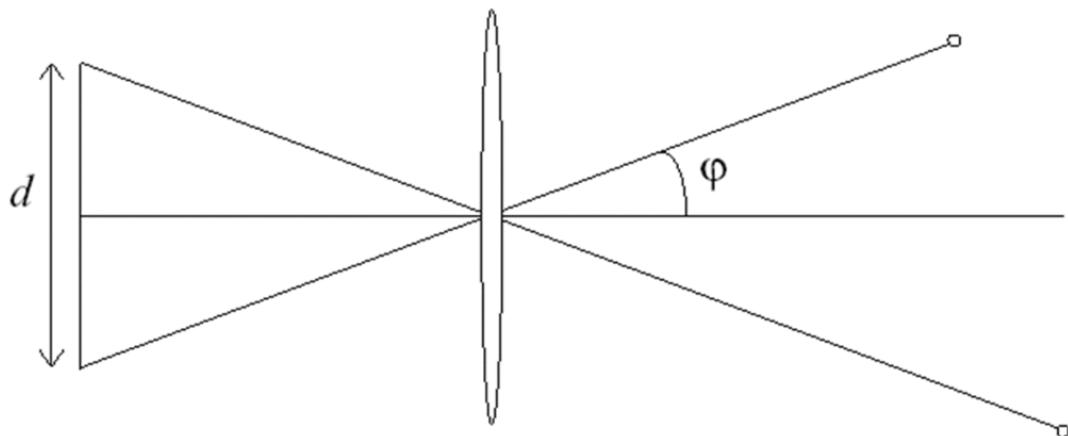
小光圈 大景深

Aperture controls Depth of Field



- Changing the aperture size affects depth of field
 - A **smaller** aperture increases the range in which the object is approximately in focus
 - But small aperture reduces amount of light – need to increase exposure

- b. 焦距对视场的影响，理解原理
短焦距 大视场角



2. 理想的针孔相机 (pinhole camera) 模型

- a. 基本投影公式，并能画图说明，会推导写出齐次定标形式下的透视投影公式是什么（矩阵形式）

■ All the parameters inside

$$\begin{cases} x_{screen} = f_x \left(\frac{Y}{Z} \right) + c_x, \\ y_{screen} = f_y \left(\frac{X}{Z} \right) + c_y \end{cases} \quad f_x = F s_x, f_y = F s_y$$

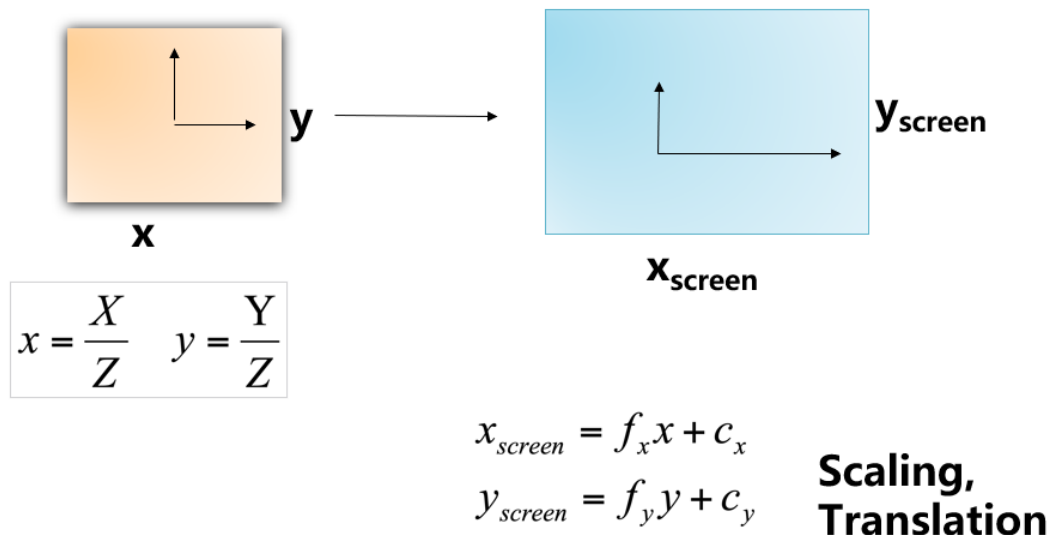
The units

- F : mm
- s_x, s_y : pixel/mm
- f_x, f_y : pixel

• Called **intrinsic parameters**

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

The way in OpenCV



$$\begin{bmatrix} x \\ y \\ w \end{bmatrix} \Rightarrow (x/w, y/w) \quad \begin{bmatrix} x \\ y \\ z \\ w \end{bmatrix} \Rightarrow (x/w, y/w, z/w)$$

$$q = MQ, \text{ where } q = \begin{bmatrix} x \\ y \\ w \end{bmatrix}, M = \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix}, Q = \begin{bmatrix} X \\ Y \\ Z \end{bmatrix}$$

b. 齐次坐标的好处？

非线性转为线性，简便运算

c. 相机模型有哪几个内参（不包括畸变参数）。会写内参矩阵。

3. 畸变

a. 径向畸变与切向畸变各是什么原因引起的？

b. 径向畸变常见的有哪两种？

1. 径向畸变

- 原因：由于透镜的几何形状不完美或安装位置引起的畸变

- 分类：枕形畸变/桶形畸变

- 校正模型：
$$\begin{cases} x_{corrected} = x(1 + k_1r^2 + k_2r^4 + k_3r^6) \\ y_{corrected} = y(1 + k_1r^2 + k_2r^4 + k_3r^6) \end{cases}$$

2. 切向畸变

- 原因：透镜平面和成像平面不平行引起的畸变

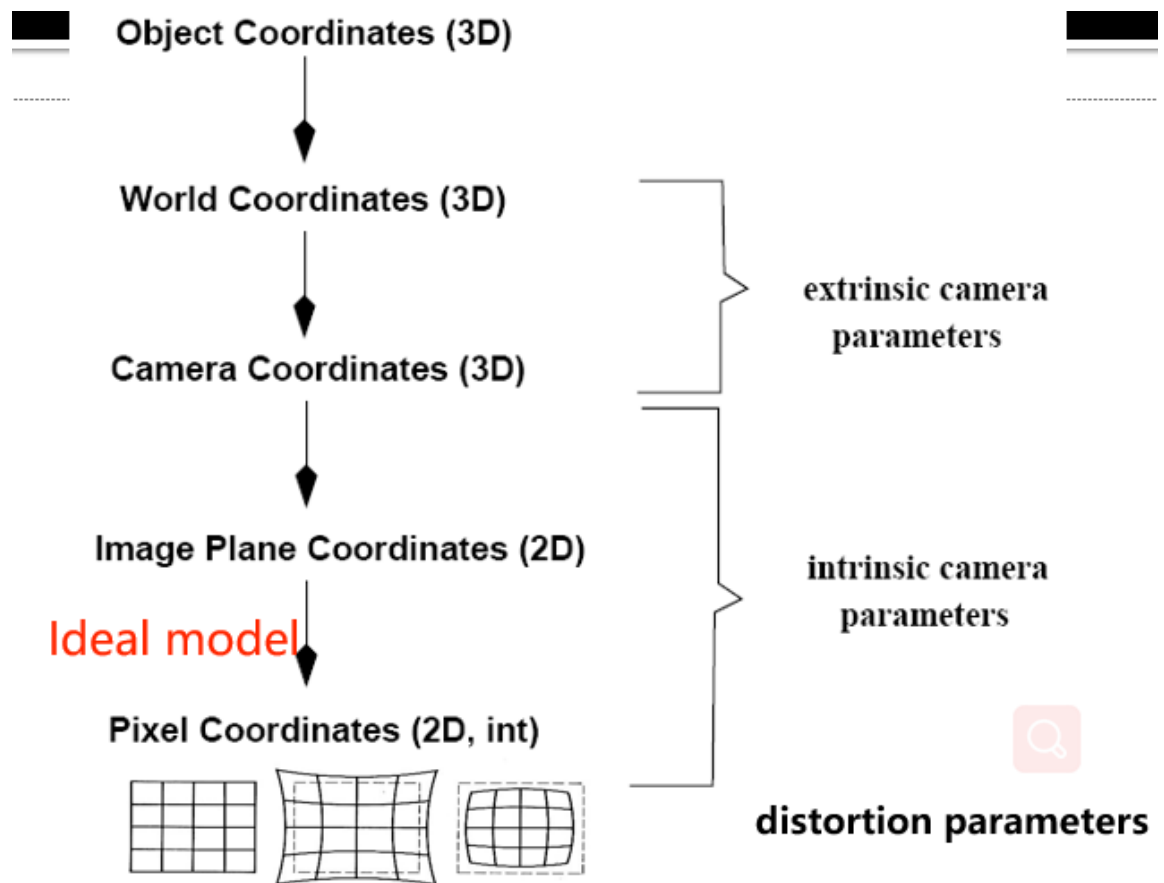
- 校正模型：
$$\begin{cases} x_{corrected} = x + [2p_1y + p_2(r^2 + 2x^2)] \\ y_{corrected} = y[2p_2x + p_1(r^2 + 2y^2)] \end{cases}$$

4. 外参有哪几个？分别代表什么含义？齐次坐标下的外参矩阵会写、会推导。

外参（extrinsic parameters）： $(\theta, \phi, \psi, t_x, t_y, t_z)$ ，或者说：

外参矩阵： $M_{ext} = \begin{bmatrix} \mathbf{R}_{3 \times 3} & \mathbf{t}_{3 \times 1} \\ 0 & 0 & 0 & 1 \end{bmatrix}$ ，有 $Q_{cam} = M_{ext} Q_{obj}$ ，相对的有内参矩阵

5. 画图展示内参、外参、畸变参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）



相机定标 (Camera Calibration)

1. 一般的相机定标
 - a. 需要求解哪些参数？
内参4、外参6、畸变参数5
 - b. 解决这个问题基本思想是什么？
联立线性方程组
2. 基于Homography的相机定标
 - a. 有哪些优点？
 - b. 简述基本过程（4个步骤）

1. 获取标定物体网格的角点在坐标系的位置
2. 找到图片的角点
3. 根据图像空间坐标系到世界坐标系列出等式
4. 求解相机参数

c. Homography矩阵有几个自由度？求解至少需要几个特征点？

- Homogeneous matrix: $H_{33}=1$
 - Degree of freedom for H is 8
- At least 4 points needed.
 - More is helpful.

求解：所有的参数。N 个点 K 个视角可以列出 $2NK$ 个等式，会带来 $6K+4$ 个参数。需要 $2NK > 6K+4$.

立体视觉

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

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

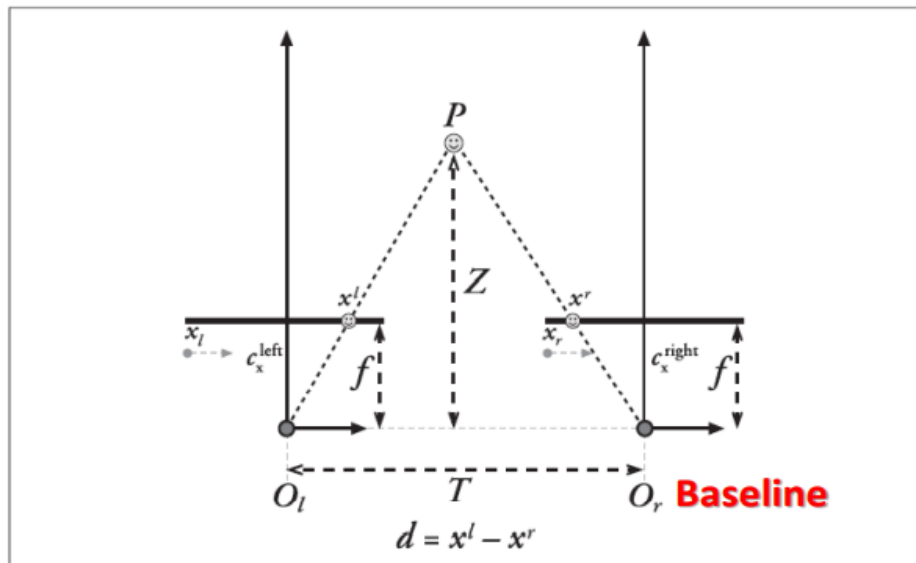


Figure 12-4. With a perfectly undistorted, pinhole camera and known correspondence, the depth Z can be found by similar triangles; the principal rays of the imagers begin at the centers of projection O_l and O_r and extend through the principal points of the two image planes at c_l and c_r .

29

2. 立体视觉的基本步骤

1. 消除畸变
2. 校正相机 (Rectification)
3. 两幅图中找到相同特征
4. 三角测量

结构光三维成像原理

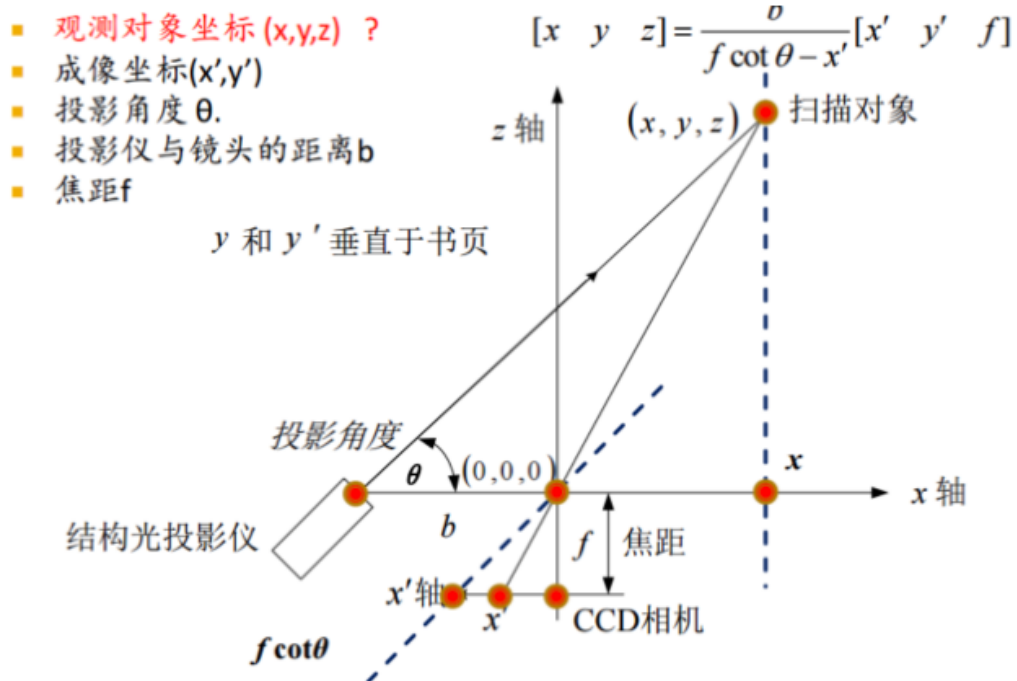
1. 结构光成像系统的构成

结构光投影仪（一台或多台）、CCD相机（一台 或多台）以及深度信息重建系统。

2. 利用结构光获取三维数据的基本原理

a. 会画图，会推导公式

基本原理——公式推导： $[x \ y \ z] = \frac{b}{f \cot \theta - x'} [x' \ y' \ f]$ ， x', y' 为物体在CCD上的投影坐标， b 为投影仪距原点位置， θ 为投影角度



ICP算法

1. 要解决什么问题？

迭代最近点方法，用于多个摄像机的配准问题，即把多个扫描结果拼接在一起形成对扫描对象的完整描述。

2. 基本步骤

基本步骤：给定两个三维点集 X 与 Y ，将 Y 配准到 X

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中对应点之间的变换关系
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，计算对应点的距离如果大于阈值，重复 2,3，否则停止计算