

计算机视觉导论

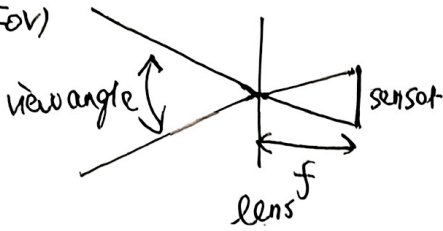
Lecture 2: Image formation.

一. 相机原理:

Pinhole camera: 小孔成像 aperture. 光圈

lens: $\frac{1}{i} + \frac{1}{o} = \frac{1}{f}$. $m = \frac{h_i}{h_o} = -\frac{i}{o}$

Field of View (FOV)



FOV $\propto$ 底片大小, $\frac{1}{f}$

F-数: $N = \frac{f}{D}$, D : 光圈大小

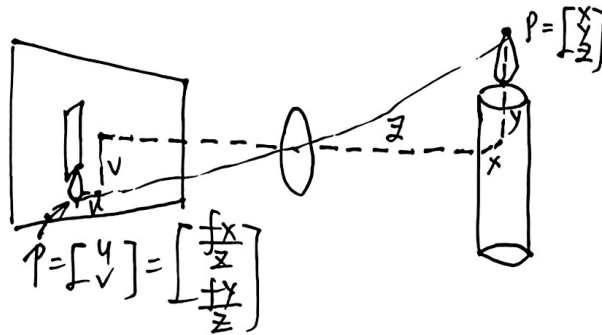
失焦 (点 $\rightarrow$ 线) 像的大小 $b \propto D \propto \frac{1}{N}$

Depth of Field (DoF 景深): 不失焦的距离 ($b < \text{pixel size}$) $DoF \propto o, N, \frac{1}{f}$

背景虚化的方法: $D \uparrow, f \uparrow, o \downarrow$, 远离背景 长减小景深 DoF.
大光圈 长焦距 贴近人

二. 何成像:

小孔成像的透视投影



齐次坐标:
$$\begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} = \begin{bmatrix} fx \\ fy \\ fz \\ 1 \end{bmatrix} \approx \begin{bmatrix} fx \\ fy \\ fz \\ 1 \end{bmatrix}$$

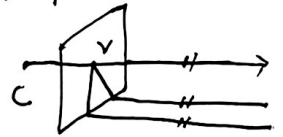
透视投影的性质: 保持平行, 角度改变.

灭点 vanishing point: 3D 平行线在 2D 图的交点, 代表 3D 平行线, 位置说明光线朝向, 可能在图像外或无穷远

vanishing line: vanishing point 的连线 平面朝向.

透视畸变 distortion: 角度变化

radial distortion: 由镜头造成的枕形畸变或桶形畸变 (直线变曲)
广角镜头 长焦镜头



三. 光学成像 (颜色, 亮度).

快门: 控制曝光时间. $\text{pixel value} = \int \text{intensity} \cdot \text{exposure time}$ 长快门小光圈.

Rolling shutter (滚动快门): 逐行曝光. 快速移动相机或物体变化迅速会导致顶部与底部错位.
解决方法: 稳定相机. 提高快门速度

颜色: RGB (0-255) 字节

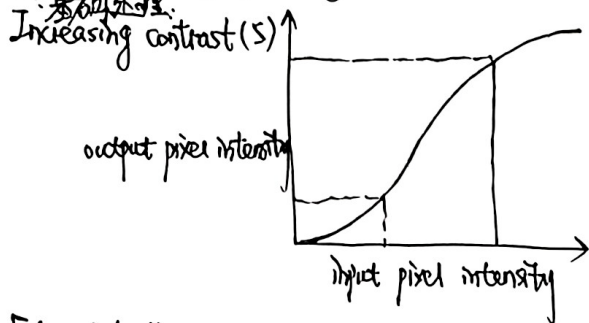
HSV (色调, 饱和度, 强度)

Bayer 滤波: 用一块传感器记录 RGB 颜色强度 (50% 绿, 25% 红, 25% 蓝)



Lecture 3: Image processing.

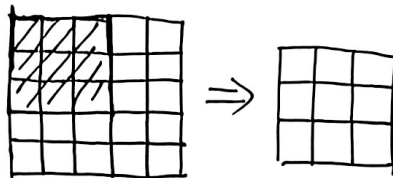
一. 基础处理:



Edge detection:

convolution: dot product for each window

padding: zero/edge value/symmetric



$$\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \rightarrow \begin{bmatrix} 4 & 3 & 4 & 3 \\ 2 & 1 & 2 & 1 \\ 4 & 3 & 4 & 3 \\ 2 & 1 & 2 & 1 \end{bmatrix}$$

Gaussian blur: $f(x,y) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}}$

sharpen: $\begin{bmatrix} 0 & -1 & 0 \\ -1 & 5 & -1 \\ 0 & -1 & 0 \end{bmatrix}$ edge detection: $\begin{bmatrix} 1 & 0 & -1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix}$ $\begin{bmatrix} 1 & -2 & 1 \\ 0 & 0 & 0 \\ -1 & 2 & -1 \end{bmatrix}$

垂直边缘 水平边缘

二. 采样:

resize: 改变分辨率/大小

降采样 $\Rightarrow$ Aliasing 走样 (Moire 纹) 原因: 采样频率太低导致信号在频域的混叠. Fourier 变换: $\sin, \cos$ 正负权重逼近

卷积定理: Spatial $\Leftrightarrow$ Freq

$$g(x) = f(x) * h(x) \Leftrightarrow G(u) = F(u) H(u)$$

$$g(x) = f(x) \otimes h(x) \Leftrightarrow G(u) = F(u) * H(u)$$

low-pass filter 低通滤波器: 让低频信号通过 例: (模糊) 高斯, 均值

高通 边缘检测.

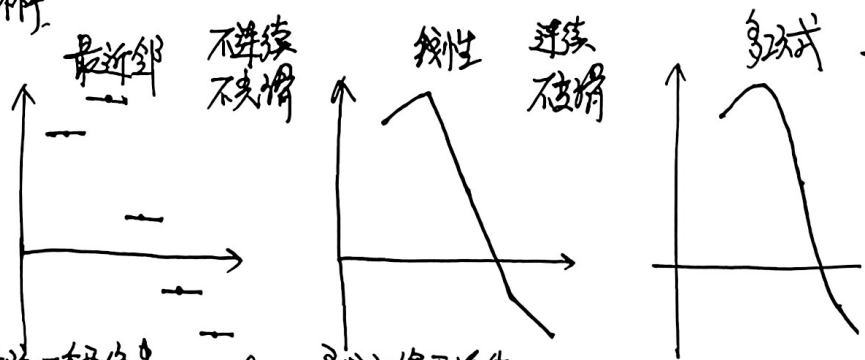
抗走样: 先通过低通滤波器去除高频成分, 再采样

与(如高斯滤波器)卷积

三. 上采样:

插值: 最近邻, 线性, 多项式(三次)

二维(二次线性插值)



改变长宽比: seam-carving: 将图像寻找边缘, 去除不重要像素 enlarge: 寻找边缘再插值.



Lecture 4: Model fitting and optimization.

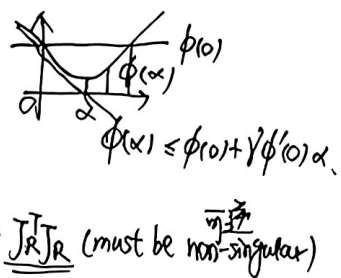
MSE: $\hat{x} = \arg \min_x \sum_i (b_i - a_i^T x)^2$. $x = (A^T A)^{-1} A^T b$. $F(x_k + \Delta x) = F(x_k) + \underbrace{J_F^T \Delta x}_{= \nabla F} + \frac{1}{2} \Delta x^T \underbrace{H_F}_{= \nabla^2 F} \Delta x$.

Gradient Descent: $x_{k+1} = x_k - \alpha J_F^T$. $\phi(\alpha) = F(x_0 + \alpha h)$. $\Delta \phi(\alpha) \downarrow, \alpha \uparrow$. $\phi(\alpha) > \phi(0)$. $\alpha \downarrow$

Newton: $x_{k+1} = x_k - \frac{H_F^T}{J_F^T}$. fast convergence, high computation

Gauss-Newton: $R(x) = b - f(x; a)$. $\Delta x = -(J_R^T J_R)^{-1} J_R^T R(x_k)$ $x_{k+1} = x_k - \underbrace{(J_R^T J_R)^{-1} J_F^T}_{\text{estimate } H_F \text{ as } J_R^T J_R \text{ (must be non-singular)}}$

LM: $x_{k+1} = x_k - \underbrace{(J_R^T J_R + \lambda I)}_{\text{正则化}} J_F^T$



Robust estimate, 去除 outlier.

1. 鲁棒的 loss func: L_1 loss / Huber loss

2. RANSAC: 拟合直线:

Iterate:

1. 随机选择两个点拟合一条直线
2. 投票 (统计直线一定宽度内的点数)
3. 选择票数最高的模型.

过拟合与欠拟合: L_1 / L_2 正则化.

Lecture 5: Image Matching and Motion Estimation.

一. 图像匹配.

1. 检测: feature points.

Harris corner detector: (1). 计算各点的协方差矩阵.

(PCA) (2). 计算特征值: λ_1, λ_2

(3). 分类: λ_1, λ_2 都小 $\rightarrow$ 平坦; λ_1 远大于 $\lambda_2 \rightarrow$ 边缘; λ_1, λ_2 都大 $\rightarrow$ 角点. $f = \frac{\lambda_1 \lambda_2}{\lambda_1 + \lambda_2}$

性质: 亮度 偏移不变, 缩放改变. 平移, 旋转: 不变, 缩放: 改变 (解决: 选择合适尺度)

尝试不同的窗口大小, 选择响应最大的窗口.

Blob detector: 与 Harris 有同样的缩放变化问题

$\log: \nabla f(x) = f(x) \nabla \log f(x)$, $\log$ (在近似): $\nabla^2 \log f \approx \nabla^2 f - \frac{\nabla f \nabla f^T}{f}$

2. 描述子: 特征点的描述向量

SIFT (Scale-Invariant Feature Transform): 梯度方向的直方图

旋转不变性: 方向归一化 尺度不变性: 解决同 Harris detector.

3. 匹配:

最近邻: $\|f_1 - f_2\|$.

重纹理 (消除歧义匹配): ratio test: $\frac{\|f_1 - f_2\|}{\|f_1 - f_3\|}$ best match second best 越大, 歧义性越强.

mutual nearest neighbor: f_2 is the nearest neighbor of f_1 in I_2 . f_1 is the nearest neighbor of f_2 in I_1 .



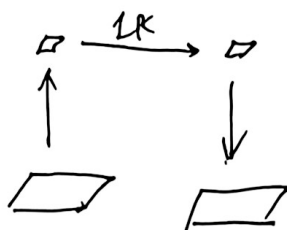
二. 运动估计

LK算法: 假设 (1) 小运动 (2) 亮度一致 (3) 空间一致

缺点: 边缘问题 边缘无法计算

非小运动 (降低分辨率) coarse-to-fine
其它违反假设的情况

feature tracking: 跟踪每帧的特征点
optical flow: 帧与帧像素的变化



Lecture 6: 图像拼接

求射变换: (1) 射变换: $\begin{pmatrix} x' \\ y' \\ 1 \end{pmatrix} \approx \begin{pmatrix} a & b & t_x \\ c & d & t_y \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ 1 \end{pmatrix}$

自由度: 6

(2) 单应变换: $\begin{pmatrix} x_i \\ y_i \\ 1 \end{pmatrix} \approx \begin{pmatrix} h_{00} & h_{01} & h_{02} \\ h_{10} & h_{11} & h_{12} \\ h_{20} & h_{21} & h_{22} \end{pmatrix} \begin{pmatrix} x_i \\ y_i \\ 1 \end{pmatrix}, [h_{00} \dots h_{22}] = 1$

自由度: 8

需要 4 对匹配

拼接过程: 1. 输入 随机选一组匹配

2. 特征匹配 计算匹配对数

3. 用 RANSAC 计算变换 选择匹配最多的变换

4. inverse warping

Inverse warping: $(x, y) = T^{-1}(x', y')$ in $f(x, y)$. 利用变换逆矩阵, 求目标图像在原图的对应位置, 插值取颜色 $\rightarrow$ 处理非整数映射 高质量
柱面投影: 全景图像拼接 (相机旋转 $\Rightarrow$ 圆柱上的平移)

Lecture 7: SFM

一. 相机模型

1. 坐标变换 (世界 $\rightarrow$ 相机): $X_w = \begin{bmatrix} x_w \\ y_w \\ z_w \end{bmatrix} \Rightarrow X_c = \begin{bmatrix} x_c \\ y_c \\ z_c \end{bmatrix}$

相机的外参: 相机在世界坐标的位置 C_w , 旋转 $R = \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{bmatrix} = \begin{bmatrix} \hat{x}_c \\ \hat{y}_c \\ \hat{z}_c \end{bmatrix}$

$$X_c = R(X_w - C_w) = RX_w + t \Rightarrow \begin{bmatrix} x_c \\ y_c \\ z_c \end{bmatrix} = \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{bmatrix} \begin{bmatrix} x_w \\ y_w \\ z_w \end{bmatrix} + \begin{bmatrix} t_x \\ t_y \\ t_z \end{bmatrix}$$

$$\tilde{X}_c = \begin{bmatrix} x_c \\ y_c \\ z_c \\ 1 \end{bmatrix} = \begin{bmatrix} r_{11} & r_{12} & r_{13} & t_x \\ r_{21} & r_{22} & r_{23} & t_y \\ r_{31} & r_{32} & r_{33} & t_z \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_w \\ y_w \\ z_w \\ 1 \end{bmatrix} = \text{Mext} \begin{bmatrix} x_w \\ y_w \\ z_w \\ 1 \end{bmatrix}$$

相机的内参矩阵

2. 透视投影

$$X_c = \begin{bmatrix} x_c \\ y_c \\ z_c \end{bmatrix} \Rightarrow X_i = f \cdot \begin{bmatrix} x_c/z_c \\ y_c/z_c \\ 1 \end{bmatrix} = f \cdot \begin{bmatrix} x_i \\ y_i \\ 1 \end{bmatrix}$$

3. 2D 到像素的映射

m_x, m_y 分别是 x, y 方向的像素密度 (pixels/mm). 则像素坐标 $u = m_x x_i = m_x f \frac{x_c}{z_c}, v = m_y y_i = m_y f \frac{y_c}{z_c}$

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} \approx \begin{bmatrix} f_x & 0 & c_x & 0 \\ 0 & f_y & c_y & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} x_c \\ y_c \\ z_c \\ 1 \end{bmatrix}, f_x = m_x f, f_y = m_y f.$$

$$= \text{Mint} \begin{bmatrix} x_c \\ y_c \\ z_c \\ 1 \end{bmatrix}$$

相机的内参矩阵

$$P = \text{Mint} \cdot \text{Mext}$$

4.



一. 相机标定 (求解内外参).

标定板匹配: $X_w = \begin{bmatrix} x_w \\ y_w \\ z_w \end{bmatrix} \Rightarrow u_i = \begin{bmatrix} u \\ v \end{bmatrix}$. 列方程 $u = P X_w$ 求解 P . 再分解 $P = M_{int} \cdot M_{ext}$.

视觉定位问题: 通过图像定位相机位置

利用2D点与3D点的对应求解相机参数.

若已知 M_{int} 求解 M_{ext} . 则是 P_nP 问题. (Perspective-n-Point)

P3P算法: 4对对应点.

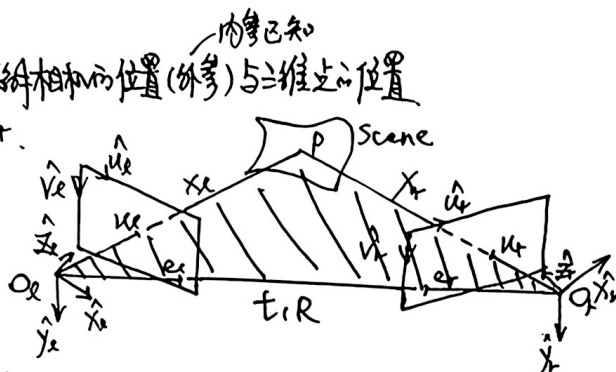
三. SfM.

定义: 给定多张图, 求解相机的位置 (外参) 与三维点的位置

1. 对极几何 epipolar.

Given $\Rightarrow$ 匹配,

求解 R 和 t .



相机坐标原点在另一台相机视角下的成像点
Epipole: e_l, e_r . 从相机看去, 另一相机位置的成像点.

Epipolar plane: 由相机原点 (O_l, O_r), 对极点 (e_l, e_r) 与场景点 P 组成的平面

Epipolar constraint: $[x_r \ y_r \ z_r] E \begin{bmatrix} x_l \\ y_l \\ z_l \end{bmatrix} = 0$, $E = \begin{bmatrix} x_r \\ y_r \\ z_r \end{bmatrix} \times \begin{bmatrix} x_l \\ y_l \\ z_l \end{bmatrix}$ essential matrix
 $X_r^T E X_l = 0$ fundamental matrix

$$\Leftrightarrow [u_r \ v_r \ 1] K_r^{-T} E K_l^{-T} \begin{bmatrix} u_l \\ v_l \\ 1 \end{bmatrix} = 0, F = K_r^T E K_l^{-1}$$

2. 相机位姿 (pose) 估计.

步骤: 1. 通过SIFT找到至少8对匹配点

2. 利用对极约束列方程

3. 求解 F

4. 利用内参求解 E

5. 分解得到 R, t .

3. Triangulation: 利用2D点位置和相机位姿通过光线求交重建3D点云.

4. 多图SfM: 1. 先使用两张图

2. 对每张增加的图:

利用已有3D点计算新相机位置

利用三角化求解更多3D点

3. Bundle adjustment

5. Bundle adjustment: 对相机参数和3D点坐标优化. 重建3D点重投影到相机图像上.

目标函数: $E(P_{proj}, P) = \sum_i \sum_j \|u_j^{(i)} - p_{proj}^{(i)}\|^2$. 所有点的重投影误差.
相机 3D点 相机图像上点的原始位置



Lecture 8: 3D Reconstruction.

一. 深度估计

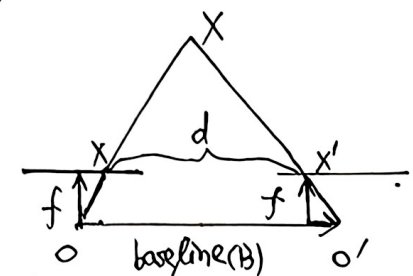
结构光: 投影仪发出红外光图案, 便于匹配

深度图

active: LiDAR (based on time of flight ToF: 发射接收差)

passive: 双目立体视觉 (视差)

depth from disparity:



步骤: 1. 相机标定

2. 特征图像

3. 特征匹配并计算视差

4. 求解深度.

disparity = $X - X' = \frac{B \cdot f}{Z}$, depth = Z .

rectification 矫正: 调整图像使两张图像的 epipolar line 对齐 (将图像平面重投影到一个公共平面)

baseline: 大 baseline 较好, 但过大会导致搜索困难, 小 baseline 会导致较大的深度误差.

二. Multi-View Stereo (MVS).

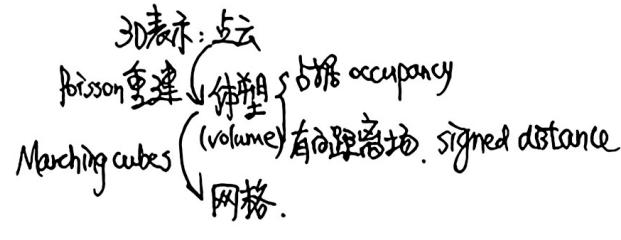
目标: 利用多张图像计算某图像的 深度图 error

方法: 对每个点尝试不同的深度值, 找到误差最小的深度.

PatchMatch: 高效匹配

三. 3D 重建:

1. SfM 得到各个相机的参数
2. 利用 MVS 计算每张图的深度图.
3. 反投影得到 3D 点云.
4. 利用泊松重建得到体素表达
5. Marching cubes 提取网格.



Lecture 9: Deep Learning.

ML: learn a model from data.

supervised learning: labelled signal

linear classifier for image: MSE + cross-entropy + softmax (score $\rightarrow$ probability)

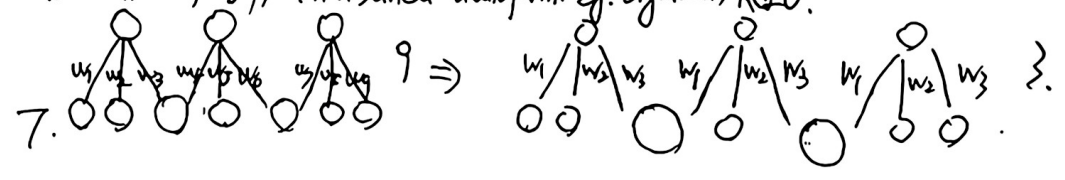
Multi-layer perceptron: $f(x) = \sigma(W_2 \sigma(W_1 x + b_1) + b_2)$, σ : non-linear transform. Eg. sigmoid, ReLU.

CNN: Locally connected layer

convolve weight sharing

$y = \sigma(x \otimes w + b)$

padding, stride. side-channel



$32 \times 32 \times 3$ image + $6.5 \times 5 \times 3$ filter $\xrightarrow{\text{ReLU}}$ $28 \times 28 \times 6$ feature maps + $10.5 \times 5 \times 6$ filters $\xrightarrow{\text{ReLU}}$ $24 \times 24 \times 12 + \dots \rightarrow$

Receptive field 感受野: 输出的每个元素包含 $K \times K$ 个像素的信息 (kernel size = K)

Fully-connected layer:

1层 $\Rightarrow$ 感受野边长: $1 + 2 \times (k-1)$.

Pooling: Max/Average. 聚合信息减小特征图.

Backpropagation, SGD, data split

Regularization, dropout, data augmentation $\Rightarrow$ prevent overfitting. batch normalization $\Rightarrow$ faster convergence, prevent gradient problem

Lecture 10: Recognition.

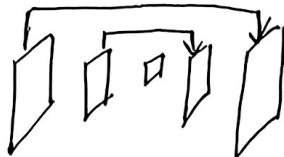
一. 语义分割.

图像像素划分(分类)

Fully CNN with per-pixel cross-entropy.

Pooling, stride then unpooling (interpolation)

U-Net: skip-connection



DeepLab: Fully connected network (FCN) + Conditional random field (CRF), SAM

二. 物体检测/目标检测

输出: 一系列 bounding boxes (label, location, size) 单阶段法: 直接分类每个锚框 (大候选区域)

R-CNN, Fast R-CNN, Faster R-CNN, YOLO (You only look once)

(Region of interest) 先找框再分类 双阶段法 } RPN: region proposal
stop feature + prediction

三. 实例分割.

物体检测 + 像素分割.

Mask-RCNN

全景分割: 对背景也作区分.

④. 人类姿态估计.

17个关节 $\Rightarrow$ 热力图表示

多个人: 自顶向下: 先识别人 (Mask R-CNN) 再寻找关节

自底向上: 先检测关节再组成人 (OpenPose)

Lecture 12: Image Synthesis (Computational Photography)

一. HDR (High Dynamic Range Imaging).

Exposure = Gain $\times$ irradiance $\times$ Time $\leftarrow$ shutter speed, aperture, ISO (感光度) 高感光度可在黑暗环境拍摄但引入噪点.

Dynamic Range 对比度

HDR: 拍摄多张不同曝光下的 LDR 再合并, 合并过程: 1. 寻找合适像素

2. 分配权重

3. weighted average 计算新像素

>. Deblurring.

blurring 成因: 物体不在焦内, 物体移动, 相机不稳.

解决办法: 高聚焦, 短快门 (大光圈, 高ISO), 硬件 (三脚架), 后期

Non-Blind (已知卷积核) image deconvolution (NBID): $G = F \otimes H$ $\Rightarrow F = \text{IFFT}(\text{FFT}(G) \div \text{FFT}(H))$
Captured to be solved

Weiner filter: 控制不收敛

L_1 regularization: $\lambda \|F\|_1 \Rightarrow$ 免滑解.

BID: $\lambda_2 \|H\|_1$

