

Image Formation

成像模型

- sensor 越大, FOV越大; 焦距越大, FOV越小
 - F数: aperture的直径 $D = \frac{f}{N}$, N 是F数
 - lens defocus, 失焦, 弥散圆
 - 景深, 弥散圆的直径小于像素尺寸
 - 如何模糊背景: 大光圈, 长焦, 近的前景, 远的背景
- geometric image formation
- 齐次坐标
 - 灭点: 任何两个平行线有相同的灭点, 相机中心和灭点的连线平行于这些线
 - 灭线: 一堆灭点组成的线, 灭线决定了一个平面
 - distortion: 径向畸变, 由镜头造成, 不是理想的透镜。桶型畸变: 广角镜头; 枕型: 长焦
- photometric image formation
- 卷帘快门效应: 它是通过Sensor逐行曝光的方式实现的。在曝光开始的时候, Sensor逐行扫描逐行进行曝光, 直至所有像素点都被曝光。当然, 所有的动作在极短的时间内完成。不同行像元的曝光时间不同。
 - HSV颜色空间: Hue表示色彩信息, Saturation表示饱和度, 饱和度为0代表纯白, Value明度, 颜色的敏感程度, 明度为0表示纯黑色
 - Bayre filter(拜尔滤镜) 在一块滤镜上设置不同颜色, 人眼对绿色比较敏感, 绿色更多

Image Processing

- 增加对比度: S-curve
- 锐化的图像: $I + (I - blur(I))$
- 提取水平边和竖直边的卷积核

- 下采样，摩尔纹，产生走样的原因：采样过慢，信号改变过快
- 卷积定理，卷积核是低通滤波器，wider kernel=lower frequency
- 香农定理：采样频率大于 $2f_0$
- 如何减少走样：增加采样频率；反走样，再采样前滤掉高频信号(使用低通滤波器)
- 图像放大：上采样，插值的方法
 - 最近邻插值，不连续不光滑
 - 线性插值，连续不光滑
 - 多项式插值：连续光滑
 - 双线性插值：
- 改变图像长宽比例：seam carving, seam insertion
 - 去除掉一些不重要的像素点？
 - 怎么衡量呢？edge energy: $E = \frac{\partial I}{\partial x} + \frac{\partial I}{\partial y}$
 - 从上到下，找到一条路径，使得edge energy最小

Model Fitting and Optimization

- 最速梯度下降法：不一定沿着最小值方向，接近最小值收敛较慢，zigzag
- 牛顿法：加上了二阶导海森矩阵 $\Delta x = -H_F^{-1} J_F^T$ ，虽然最小值附近快速收敛，但是海森矩阵计算量太大
- 高斯牛顿法：useful for solving nonlinear least squares
 - 优化残差向量
 - 泰勒展开残差向量
 - $\Delta x = -(J_R^T J_R)^{-1} J_R^T R(x_k)$
- Levenberg-Marquardt: $\Delta x = -(J_R^T J_R + \lambda I)^{-1} J_R^T R(x_k)$
 - start quickly $\lambda \rightarrow \infty$
 - converge quickly
 - LM=Gradient descent+Gauss-Newton
- Robust Estimation

- outliers: 改变目标函数 Huber
 - RANSAC: use data point to vote
- Overfitting L2正则化, 抑制冗余变量, L1正则化, 稀疏正则化
- GraphCut : image labelling problems
 - neighbouring pixels tend to take the same label
 - 把一张图片的每一个像素看作一个 graph 中的 vertex, 并在像素之间建 edge, 并将 weight 定义为两像素之间的相似性或关联性 (affinity or similarity), 我们将小权的边删去, 最终会形成若干连通分量, 而这些连通分量中的点则被视为一个“分割”。

Image Matching and Motion Estimation

- 特征匹配: 检测特征点(detection),描述特征点, 特征匹配
- detection: 性质: uniqueness, 角点
- 用主成分分析来衡量, Harris 角点检测
 - $f = \frac{\lambda_1 \lambda_2}{\lambda_1 + \lambda_2}$
 - 我们希望相应的像素点是invariant to image transformation
 - 对光强部分不变, 平移不变, 旋转不变, 但是scaling会变
 - find scale that gives local maximum of f
 - 使用图像金字塔实现
- descriptors: SIFT
- 朝向归一化: dominate orientation
- Matching:
 - ratio test: ambiguous matches have large ratio scores.
 - Mutual nearest neighbour:
- Motion estimation: 给出两帧画面, 估计出像素的运动
 - Feature tracking: 离散
 - Optical flow: 连续稠密
- LK method
 - Key assumptions:

- small motion
- 光度一致
- spatial coherence
- analogy to Harris corner detect(方程有解的条件是能被哈里斯角点检测法检测到角点)

Image Stitching

- Image Warping
 - global warping,所有像素点都是同一个映射函数
 - 单应性变换
 - 图像平移：两个自由度；仿射：六个自由度；投影：8个自由度
- 求解warping，使用inverse warping，若不在像素点上则插值
- Image Stitching:
 - RANSAC for Translation，找到之后用所有的inliners再拟合一次，最小二乘算H
- 柱型投影：相机的旋转变成平移

Structure from Motion

- coordinate transformation:Extrinsic Matrix
- perspective projection+image plane to image sensor mapping: Intrinsic Matrix
- Projection Matrix $P = M_{int} M_{ext}$
- Camera Calibration: obtain P
 - Then, QR factorization, obtain K, R, t
- PnP problem: 6DoF problem
 - 最小化重投影误差
- Structure from Motion
 - 对极几何，本质矩阵，基础矩阵
- Traingulation, 解三维坐标

- Bundle Adjustment : 最小化重投影误差

Depth estimation and 3D reconstruction

- 立体匹配: 视差与深度成反比
 - pipeline: calibrate cameras, rectify images, compute disparity, estimate depth
- MVS:
 - Plane Sweep 得到深度图, cost volume
 - PatchMatch 算法
- 3D重建
- 三维表示: 点云、体素、网格
- 从深度图得到三维网格的pipeline
 - 将深度图转化为体素表示(点云), 方便计算机处理
 - 泊松表面重建: 优化梯度与法向的loss, 用泊松方程求解
 - Marching cubes 三维中有256中情况 2^8

Deep learning

- 线性分类问题中不能用最小二乘作为损失函数, 分数是连续的但是标签是离散的, 我们使用交叉熵损失函数。
- 卷积神经网络
- 训练神经网络:
 - data split: validation 上选超参, 在test 上测试
 - 防止过拟合: dropout, regularization, data augmentation
- ResNet:
 - adding a layer increases error? identity map is not easy to learn 梯度消失
 - only make each layer to learn Δx_i

Recognition

- Semantic segmentation
 - 全卷积神经网络, U-Net+skip connection
 - evaluation metric: IoU
- Object detection
 - Region proposals
 - R-CNN
 - evaluation metric: non-max suppression
 - Fast R-CNN: 过卷积神经网络后再小区域proposal得到bbox和class, 其中RoL pooling crop+resize features
 - Faster R-CNN, 用卷积神经网络生成proposals
 - RPN: anchor box, 输出k乘像素个box,带有score, 选top n的作为region proposal
 - two-stage object detector
 - first stage: run once per image(backbone network or RPN)
 - second stage: crop features: RoL pool/align, predict class and offset
 - Single-stage object detection: Yolo
- Instance segmentation: object detection+ 每个bbox里面semantic segmentation
- Human pose estimation:
 - represent the pose of human by locating a set of keypoints
 - Single human: 输出热力图
 - Multiple humans:
 - Top-down: detect in each bbox
 - Bottom-up: group keypoints to form humans(Openpose)
 - Openpose: Link parts based on part affinity fields

3D deep learning

Computational Photography

- HDR 高动态范围成像
 - $\text{Exposure} = \text{Gain} \times \text{Irradiance} \times \text{Time}$
 - dynamic range: 最大值和最小值的比值
- Deblurring
 - reason: 失焦or 运动模糊
 - NBID已知卷积核, 逆卷积, 傅里叶逆变化, Wiener Filter 卷积核是低通, 逆卷积高通, 放大了噪声
 - 优化的方法, 希望图像梯度场稀疏
 - BID 不知道卷积核, 卷积核正的且稀疏
- GAN G: 生成假的图片来糊弄D, D识别假图片
- $\arg \min_G \max_D E_{xy}(\log D(G(x)) + \log(1 - D(y)))$
-