Machine Vision

W3 Image formation and representation

- A digital image is an array, or a matrix, of pixels in columns and rows.

电荷耦合器件(高感光度的半导体材料)

- CCD (charged-coupled device) 光→电荷，电信号→数字信号

CMOS (complementary metal-oxide semiconductor)

→低… 互补金属氧化物半导体，感光元件

高端摄像头

- SNR (signal-to-noise ratio) 信噪比

指MRI信号和噪声比值，SNR越低，图像颗粒感越明显

- Pixelcount
  - More pixels → higher detail
  - More pixels → smaller pixels?

  → lower SNR → difficult for lens to resolve

- Depth of field (DOF) 景深

在聚焦完成后，焦点前后的范围内所呈现的清晰图像的距离，这一前一后的范围，称为景深.

- 透视投影：从某个投射中心将物体投射到单一投影面上得到的图形，和人的视觉效果接近.

- 消失点 → 平行线的视觉相交点

- 全息投影：利用干涉和衍射原理记录并再现物体的三维图像

· 分类：{ · Binary images (Black and white images) 二值图像
· Greyscale images 灰度图像
· Colour images

· Binary images
0 — 黑
1 — 白

· Greyscale images (8 bits. 2^8 256 scales)
0 - 255
黑 → 白

· Colour images (e.g. RGB) (3 layer → 24 bits. 16777216 scales)
red blue green
3×8

· Colour separation - Bayer filter
每个发光二极管只感应红色/绿/蓝信息，分别记录R.G.B.

· 人眼对绿色信息更敏感.

· Color space - HSI
Hue 色调 — 表示人对不同颜色的感受
Saturation 饱和度 — 表示颜色纯度
Intensity 亮度 — 表示颜色明亮程度

· RGB to Greyscale
$I = a*R + b*G + c*B$
e.g. matlab $I = 0.2989*R + 0.5870*G + 0.1140*B$

- Grayscale → Binary ? 利用阈值

W4 : Image thresholding and binary morphology

- Image Segmentation
  involves isolating features of interest (e.g. an object) from other parts of the images in order to analyse individual features

- Thresholding
  - histogram based method
  - clustering
  - mixture model

- Grey level histogram analysis
  - show the nature of an image.
  - separate an object from the background
  - analysis of a frequency distribution of the image intensity data.

- Gray level to Binary Thresholding
  possible methods for establishing thresh:
  - Human adjustment
  - "goodness measure" (e.g. empirical observations)
  - use mean or median of ~~inten~~ intensity distribution
  - use a statistical method.

· Learning based Image Segmentation

· Semantic Segmentation 语义分割：让计算机根据图像的语义来进行分割

· Instance Segmentation

实例分割：是语义分割的更进一步，对于图像中每个物体都要进行区分和标注

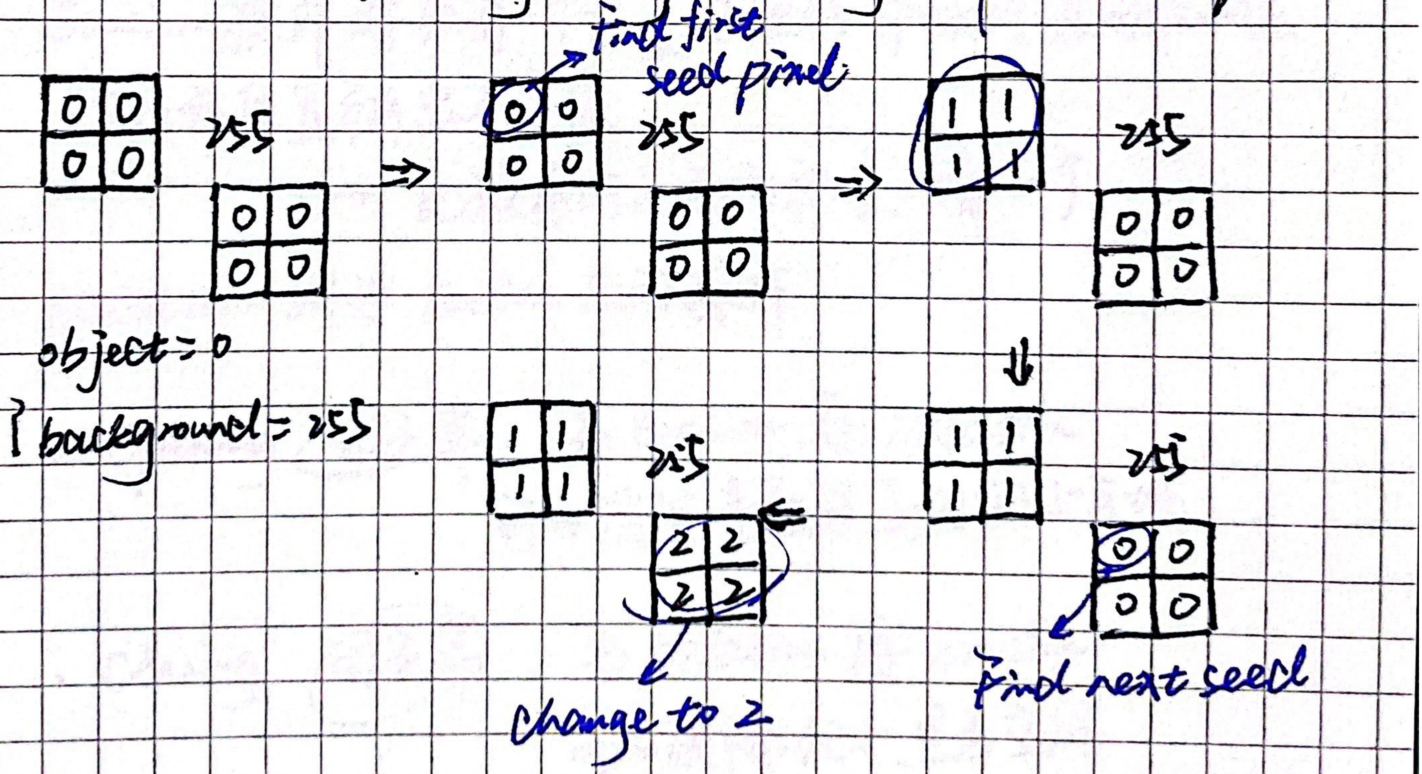
· Morphological operations

To find a connected region

~~Produce~~ Procedure:

1. Find a pixel in the region (seed)
2. Grow the region by locating all connected pixels.

Grow region by changing all connected pixels to 1



- The concept of 4 or 8 connected

$\longrightarrow j$, $\downarrow i$

8 $(i-1, j-1)$	1 $(i-1, j)$	2 $(i+1, j+1)$
7 $(i, j-1)$	0 (i, j)	3 $(i, j+1)$
6 $(i+1, j-1)$	5 $(i+1, j)$	4 $(i+1, j+1)$

- 4-adjacent point/pixel 1.3.5.7
- 8-adjacent Pixel 1~8. 如何解决touching?

腐蚀
- Erosion: remove pixels that are adjacent to the background (4 or 8 adjacent) (去掉毛刺)
- Dilation: add new pixels around boundary of region (4 or 8 adjacent). (补全缺点)

膨胀

Erosion操作则取每一个位置的矩形邻域内值的最小值作为该位置输出灰度值

$\longrightarrow$ 比较亮的区域面积↓，暗↑

Dilation 则为 Erosion 反向操作.

- Opening 开运算 为 Erosion 再 Dilation
  可以消除亮度较高的细小区域
- Closing 闭运算 为 Dilation 再 Erosion
  可以消除细小黑色空洞

Fit: 结构元中所有像素都和图像区域重合

Hit: … 存在像素和图像区域有重合

W5 Image Convolution

- Traditional five stages of machine vision

light
W4

Scene constraints → Image acquisition → Pre-processing → Segmentation (W8)

four-in-one

Segmentation → Feature extraction (W5, W6) → Classification / interpretation → Interaction → Scene constraints

Feature (a piece of information that can be used to solve a problem)

type of feature

- Edges (Contours)
- Corners (interest points)

Common types of noise

椒盐噪声 表现为在图像上引起较强视觉效果的孤立像素点或像素块.

- Salt and pepper noise - Contains random occurrences of black and white pixels. (黑白相间亮暗噪点)
- Gaussian noise - variations in intensity drawn from a Gaussian distribution. 概率密度函数服从高斯分布的一类噪声

往往由图像切割引起.

- 图像传感器在拍摄时不够明亮. 亮度不够均匀
- 电路各元件自身噪声和相互影响
- 图传长时间工作. 温度过高.

- Image filters
  - a technique for modifying or enhancing an image
  - filter an image to emphasize certain features or remove other features
  - Smoothing, sharpening and edge enhancement.
  - Applications: Remove ~~....~~ noise, detect edge.

(Image filter)
- for noise reduction
  - 根据周边像素点, e.g.

$$\begin{matrix} 0 & 20 & 30 \\ 12 & 255 & 15 \\ 17 & 10 & 14 \end{matrix} \xrightarrow{\text{(some function)}} \begin{matrix} \cdots & & \cdots \\ & 40 & \\ \cdots & & \cdots \end{matrix}$$

- Linear filtering
  - one simple version: linear filtering (convolution) 卷积
    - replace each pixels by a linear combination of its neighbours
    - the function of the linear combination is called "kernel" (or "mask", "filter")

$$\begin{matrix} 0 & 20 & 30 \\ 12 & 255 & 15 \\ 17 & 10 & 14 \end{matrix} \otimes \begin{matrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{matrix} \rightarrow \begin{matrix} 0 & 20 & 30 \\ 12 & 40 & 15 \\ 17 & 10 & 14 \end{matrix}$$

$\otimes$: 卷积运算符号

kernel (核心)

· A convolution is done by multiplying pixels and its neighbouring pixels intensity value by a matrix

· Kernel: a usually small matrix of numbers that is used in image convolution.
→ 3×3

· 2D convolution

· 加权 → $\begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix} \otimes \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix}$

$= a*1 + b*2 + c*3 + \cdots + i*9$

· Example of average filter 均值滤波 适用于 Gaussian noise.

原理 · $x(i,j) = 1/m \sum_{k=i-1}^{i+1} \sum_{l=j-1}^{j+1} f(k,l)$

(中心点像素为全部点像素值平均值) 服从正态分布

$\begin{matrix} 36 & 36 & 36 & \cdots \\ 36 & 45 & 36 & \cdots \\ 36 & 45 & 45 & \cdots \\ \cdots & & & \cdots \\ \cdots & & & \cdots \end{matrix} * \frac{1}{9} \begin{matrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{matrix} \rightarrow \begin{matrix} \cdots & \cdots & \cdots \\ \boxed{39} & \boxed{} & \\ \cdots & & \\ \cdots & & \cdots \\ \cdots & & \cdots \end{matrix}$

· kernel size 过大会导致模糊

· Average filter 不合适 S&P noise removal.
会模糊图像. 幅度不同. 幅值近似相等且随机分布.
位
⇓
· 采用 median filter for S&P noise removal.
中值滤波

对于一连串数字 {1. 4. 6. 8. 9} 来说.
中值为 6. 推广至图像

3×3矩阵区域 → 9个像素点 排序 → 中心点赋值为这9个像素中值

· Padding 填充问题

· 不做边界处理

· 0 填充 → 在矩阵边界填充0.

0	0	0	0	0	0
0	1	2	3	4	0
0	5	6	7	8	0
0	9	10	11	12	0
0	13	14	15	16	0
0	0	0	0	0	0

· 填充最近像素值

1	1	2	3	4	4
1	1	2	3	4	4
5	5	6	7	8	8
9	9	10	11	12	12
13	13	14	15	16	16
13	13	14	15	16	16

· Edge detection (Using image gradients & Convolution)

· 图像梯度

· 图像是一个离散的二维函数 $f(x,y)$

$$\Rightarrow \frac{\partial f(x,y)}{\partial x} = \lim_{\epsilon \to 0} \frac{f(x+\epsilon, y) - f(x,y)}{\epsilon}$$

$$\frac{\partial f(x,y)}{\partial y} = \lim_{\epsilon \to 0} \frac{f(x, y+\epsilon) - f(x,y)}{\epsilon}$$

· * 最小的 ϵ 为 1 像素.　　$\epsilon = 1$

$$\frac{\partial f(x,y)}{\partial x} = f(x+1,y) - f(x,y) = gx$$

$$\frac{\partial f(x,y)}{\partial y} = f(x,y+1) - f(x,y) = gy$$

这分别为在 (x,y) 点处 x 方向和 y 方向上的梯度,

相当于 2个 相邻像素间的差值.

$$\rightarrow M(x,y) = |gx| + |gy|$$

· Sobel operator 索贝尔算子 (一阶微分算子)

把图像中每个像素的上下左右四领域的灰度值加权差，在边缘处达到极值从而检测边缘.

· 算子模板

$$\begin{matrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{matrix}$$

检测水平方向边缘

$$\begin{matrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{matrix}$$

检测垂直方向边缘.

· Prewitt operator (算子) (一阶微分算子) 利用像素点上下、左右邻点的灰度差，在边缘处达到极值检测边缘

· 模板

$$\begin{matrix} 1 & 1 & 1 \\ 0 & 0 & 0 \\ -1 & -1 & -1 \end{matrix} \quad 和 \quad \begin{matrix} -1 & 0 & 1 \\ -1 & 0 & 1 \\ -1 & 0 & 1 \end{matrix}$$

· Laplacian operator (二阶微分算子)

· 模板 4邻域

$$\begin{matrix} 0 & 1 & 0 \\ 1 & -4 & 1 \\ 0 & 1 & 0 \end{matrix}$$

8邻域

$$\begin{matrix} 1 & 1 & 1 \\ 1 & -8 & 1 \\ 1 & 1 & 1 \end{matrix}$$

Given a camera and a still scene. how can you reduce noise?

- Improve lighting
- Increase exposure time
- Use a better camera sensor, lens, etc.
- Take multiple image and calculate the average of them.

w6. Features

How does the brain solve visual object recognition?

- Detect feature points
- Describe feature points
- Matching

Feature detection and des~~c~~ri~~p~~tion

feature vector

· 口 ⟶ description 0 1 2 5 4 …

日 ↗

detection

↕ matching

a b $\underbrace{\text{c d e}}_{\text{prior knowledge}}$ …

△ 对于机器视觉来说，高维数据图像需要具有一些

Feature invariance

- Greyscale invariance
- Rotation invariance
- Illumination invariance
  照明
- Scale invariance

- Greyscale invariance
- Texture Spectrum

· Texture 是分析多种类图像的重要特征

· A texture image 可以分解为一個基本小单元，称为纹理单元 units.

$$\begin{matrix} 60 & 20 & 65 \\ 80 & 40 & 30 \\ 60 & 40 & 20 \end{matrix} \longrightarrow \begin{matrix} 2 & 0 & 2 \\ 2 & & 0 \\ 2 & 1 & 0 \end{matrix} \qquad P = \begin{cases} 2 & \text{if } p > \text{centre\_val} \\ 1 & \text{if } p = \text{centre\_val} \\ 0 & \text{if } p < \text{centre\_val} \end{cases}$$

$3^8 = 6561$ units to characterise the texture aspect in all the eight directions from the central pixel.

· The occurrence distribution of Texture units is called the Texture Spectrum.

升级

· Local binary patterns (LBP) 局部二值模式

· $2^8 = 256$ instead of $3^8 = 6561$ compared to Texture Spectrum.

· Transform an image into an array or image of integer labels describing small-scale appearance (Texture) of the image.

· These labels directly or their statistics are used for further analysis. e.g. histograms

LBP calculation

$$\begin{matrix} 60 & 20 & 65 \\ 80 & 60 & 30 \\ 60 & 40 & 20 \end{matrix} \longrightarrow \begin{matrix} 1 & 0 & 1 \\ 1 & 0 & 0 \\ 1 & 1 & 0 \end{matrix}$$

$$\begin{matrix} 1 & 2 & 4 \\ 128 & & 8 \\ 64 & 32 & 16 \end{matrix} \quad \text{weight}$$

$$\begin{matrix} (128 & 64 & 32 & 16 & 8 & 4 & 2 & 1) \\ \times & \times & \times & \times & \times & \times & \times & \times \\ 1 & 1 & 1 & 0 & 0 & 1 & 0 & 1 \end{matrix}$$

$$\downarrow$$

229

describing a texture unit.

- LBP-Uniform patterns
  - Uniformity measure U("pattern") is the number of bitwise transitions from 0 to 1 or vice versa.
  - A local binary pattern is called uniform if its uniformity measure is most 2.

transition

1 1 1 0 0 1 0 1

transition

Example

0	0	0	0	0	0	0	0	0 transitions	uniform.
0	1	1	1	0	0	0	0	2	
1	1	0	0	1	1	1	1	2	
1	1	0	0	1	0	0	1	4	nonuniform.
0	1	0	1	0	0	1	1	5	

Why omit non-uniform patterns?

- 物理/现实世界中获取八邻域的所有模式中有 90% 是 uniform patterns.
- 和总共 256 个模式相比，uniform patterns 只有 58 个.

⇒ 这意味着可以扔掉近 20% non-uniform patterns.

但依然保留了来自现实世界 90% 的关键信息.

LBP-features

- Binary pattern values ← 299.
- Histogram of the above

- mainly for monochrome but have been extended for color.

LBP - Advantages

- High discriminative power

↓

- good performance
- Computational Simplicity
- Invariance to grayscale changes

Disadvantages

- Not invariant to rotations
- Size increases exponentially 指数地 as the number of neighbourhoods increases.
- Dose not use magnitude information of pixel difference

Rotation invariance solution

→ rotate weight with the raw image

1	0	1		1	2	4
1	0	0	←	128		8
1	1	0		64	32	16

1	1	1		64	128	1
1	0	0	←	32		2
0	0	1		16	8	4

how to achieve illumination invariance (to some extent)

One solution:

- Divide image into cells
- Compute LBP histogram per cell
- Concatenate all histogram (with or without normalisation)

Image content is transformed into local feature coordinates that are invariant to translation, rotation, scale, and other imaging parameters.

↓ How. (之前用了 average filter to blur an image).

Gaussian kernel. (越近中心权重越高)

$$\frac{1}{273}\begin{matrix} 1 & 4 & 7 & 4 & 1 \\ 4 & 16 & 26 & 16 & 4 \\ 7 & 26 & 41 & 26 & 7 \\ 4 & 16 & 26 & 16 & 4 \\ 1 & 4 & 7 & 4 & 1 \end{matrix}$$ ← A 2D Gaussian kernel.

- Gaussian Convolution

  Step 1. Image subsequently blurred using a Gaussian Convolution

  Step 2. Scale space constructed to simulate different scale of observation.

  Step 3. candidates for the key points

· Visualisation of SIFT keypoints

SIFT (Scale-invariant feature transform) 尺度不变特征变换

可在图像中检测出关键点.

SIFT 算法特点.

1、对旋转.尺度缩放.亮度变化保持不变性.

2、区分性好.信息量丰富

...

· SIFT applications

· object recognition

· SLAM 同步定位和地图构建.

· panorama stitching 全景拼接

· Tracking

w7 Machine learning 略

What? Algorithms and statistical models that can perform a task without using explicit instructions, and that can improve their performance using training data.

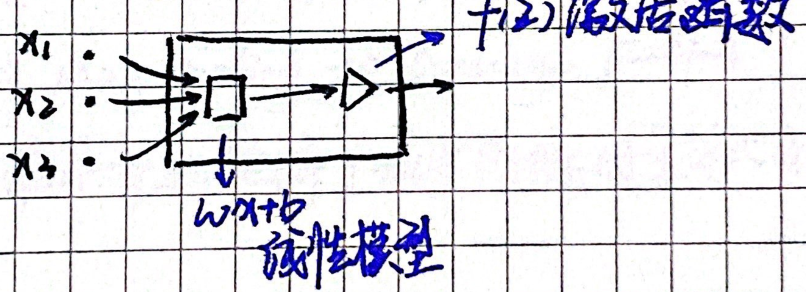
w8. Object detection

· image classification

Classify an image according to its visual content. outputting a class or probability that the input is a particular class.

CNN 卷积网络神经基本概念

神经元 :



· 全连接 (full connected)

每一层的每个神经元都与下一层的每个一相连

缺点：计算量大.

CNN 参数 (3个)

· kernel 大小

· 填充 · 步长

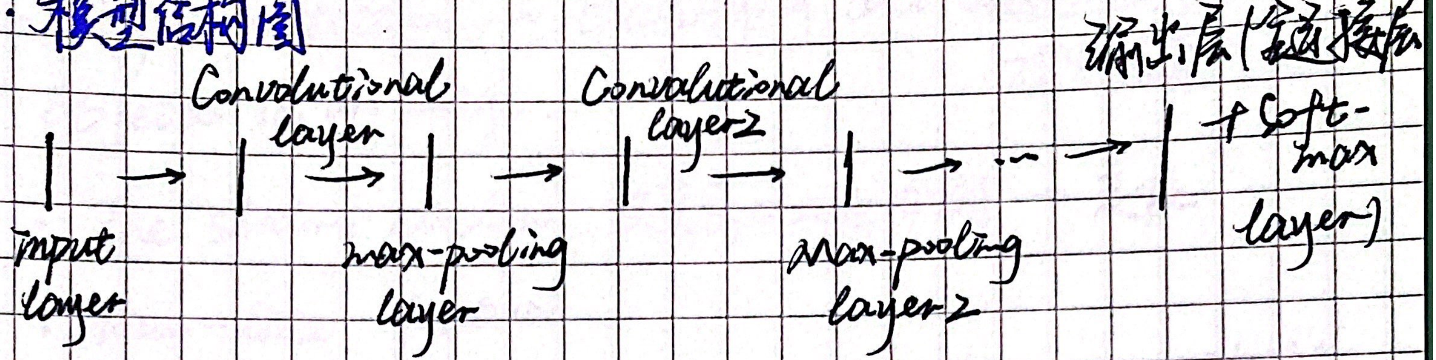
· 卷积核

· 越大，可提取的输入特征越复杂

· 用来检测某一方面的特征，比如垂直边缘，水平边缘…

又称为过滤器

· 模型结构图



· 至少有1个卷积层，用以提取特征。（点积：）

· 卷积层输出→“变胖”：

每个卷积层不单止一个卷积核，特征图则是上一层经过多个卷积核“压缩提炼”，每个卷积核对应一层，多层叠加→“变胖”

· 一般来说，卷积层后面都会加一个池化层

对卷积层输出的特征图进一步特征抽样

主要分为两种 Max pooling 和 Average Pooling

→ 可以有效缩小参数矩阵（减少图像的空间大小）

- object detection → classifying and localising multiple objects in an image.
  找出位置

- The sliding window 改变窗口大小比例 → 定位

- haar-like feature

  e.g. The eye region is darker than the upper-cheeks.

Σ(pixels in white area) − Σ(pixels in black area)

↓ The weak learning algorithm is designed to select the single rectangle feature which best separates the positive and negative examples.

→ 加权输出

- object detection with a CNN
- 在输出层基本上添加一些神经元. 即类标签 class label.
- 框 □. 4个参数: (x,y) 左上角坐标, w 宽度, H 高度

- Evalution
  - Intersection over union (IOU) → 单个检测
  - Precision and Recall
  - Mean average precision (mAP)

传统图像处理算法

- 图像滤波
  - 线性滤波：方框、均值、高斯 滤波
  - 非线性滤波：中值、双边 滤波
- 形态学 ~~形态学滤波~~

腐蚀、膨胀、开运算、闭运算、形态学梯度、顶帽、黑帽

- 阈值化
- 图像金字塔
  - 高斯金字塔
  - 拉普拉斯金字塔
  - resize 函数
- 边缘检测

Sobel 算子、Laplacian 算子、Canny 算子

直方图均衡化

- 霍夫变换 直线检测、圆检测
- HOG 特征和 SIFT 特征

图像处理中的机器学习

· 无监督学习

k均值聚类算法、谱聚类算法、PCA(主成分分析

· 监督学习

KNN(K最近邻分类器)、贝叶斯分类器(高斯生成模型)

SVM(支持向量机)分类器

目标检测常用算法

· 传统：Cascade + HOG/DPM + Haar/SVM 及改进和优化

· 候选区域/框 + 深度学习分类：

R-CNN (Selective Search + CNN + SVM)

SPP-net (ROI pooling)

fast R-CNN (Selective Search + CNN + ROI)

faster R-CNN (RPN + CNN + ROI)

R-FCN 等系列方法。

· 基于深度学习的回归方法：

YOLO / SSD / DenseBox. etc.

· 分水岭算法（基于拓扑理论的数学形态学 → 拓扑地貌

过程：1. 梯度图像中 → 灰度值分类 + 设定一个测地阈值

2. threshold从灰度值最小值开始增长
起始点

3. 水平面在增长过程中，会碰到周围的邻域像素，测量其到起始点的测地距离，如果小于设定threshold，则将这些像素淹没，否则在这些像素上设置大坝，这样就对这些邻域像素进行了分类。

4. 随着水平面越来越高，会设置更多更高的大坝，直到灰度值最大值，所有区域都在分水岭线上相遇，这些大坝就对整个图像像素进行了分区。

传统分水岭算法存在过度分割的不足，

Opencv 提供了一种改进的分水岭算法，使用一系列预定义标记来引导图像分割的定义方式。

大津算法（OTSU） 对图像进行二值化的高效算法

利用阈值将原图像分为前景、背景两个图象

本质思想是最大化类间方差

设一阈值T将图像中像素分为A（>T）、B（<T）两类

如何求T？

设阈值为k时，像素被分到A概率为$P_A(k)$，

分配到A的像素平均灰度为$m_A(k)$，…$P_B(k)$和$m_B(k)$。

整个图像灰度值m_G，灰度级k的累加值m。

$$P_A(k)\cdot m_A(k)+P_B(k)\cdot m_B(k)=m_G$$

$$P_A(k)+P_B(k)=1$$

代入

方差：$\sigma^2=P_A(k)(m_A(k)-m_G)^2+P_B(k)(m_B(k)-m_G)^2$

$$\Rightarrow \sigma^2=P_A(k)P_B(k)(m_A(k)-m_B(k))^2$$

其中 $P_A(k)=\sum_{i=0}^{k}p_i$，$P_B(k)=\sum_{i=k+1}^{255}p_i$

$m=\sum_{i=0}^{k}ip_i$，$m_G=\sum_{i=0}^{255}ip_i$

$m_A(k)=\dfrac{\sum_{i=0}^{k}ip_i}{P_A(k)}$　$m_B(k)=\dfrac{\sum_{i=k+1}^{255}ip_i}{P_B(k)}$

图像中灰度值为i概率为 $p_i=\dfrac{n_i}{n_0+n_1+\cdots+n_{255}}$

（n_i为灰度为i的像素数量）

$$\Rightarrow \sigma^2=\frac{(m_G\cdot P_A(k)-m)^2}{P_A(k)(1-P_A(k))}$$

使σ^2最大的k即为阈值T。

颜色特征 估计柑橘产量

Input RGB Image
↓
Background Subtraction
↓
Resizing Image
↓
Convert RGB Image to HSV
↓
Split color components:
H, S, and V
↓
Thresholding by H component
↓
orange Color Detection
↓
Background color change
to black
↓
orange color change
to gray

RGB value
determination
for background
→ orange Color Detection

→ Noise Removal
↓
Thresholding
↓
Noise Removal
↓
Make Binary Image
↓
Watershed Segmentation
↓
filling holes
↓
fruit Counting
↓
Comparison between Composed
fruit Counting Algorithm and
Yield Esstimation Result
by flower Counting