

N L P R



特征点提取、描述与匹配

董秋雷

中国科学院自动化研究所

qldong@nlpr.ia.ac.cn

提纲

- 图像特征点提取
 - SIFT (I) - 特征点
- 图像特征点描述与匹配
 - SIFT (II) - 描述子
 - ORB→BRIEF
- 基于DNN的特征描述子学习
- 空间点匹配：ICP
- 鲁棒估计：RNASAC

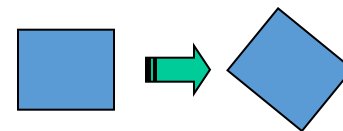
SIFT特征点检测 (I)

- David G. Lowe, Distinctive Image Features from Scale-Invariant Keypoints, International Journal of Computer Vision, Vol. 60, Page 91-110, 2004.

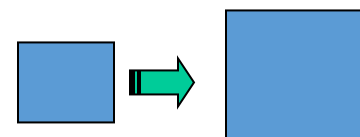
图像变化的类型

- 几何变化

- 旋转

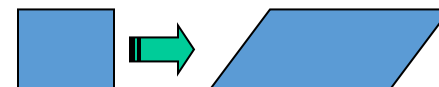


- 相似 (旋转 + 各向相同的尺度缩放)



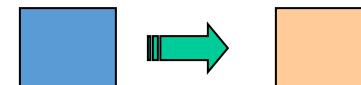
- 仿射 (非各向相同的尺度缩放)

适用于：物体局部为平面



- 灰度变化

- 仿射灰度变化 ($I \rightarrow aI + b$)



SIFT

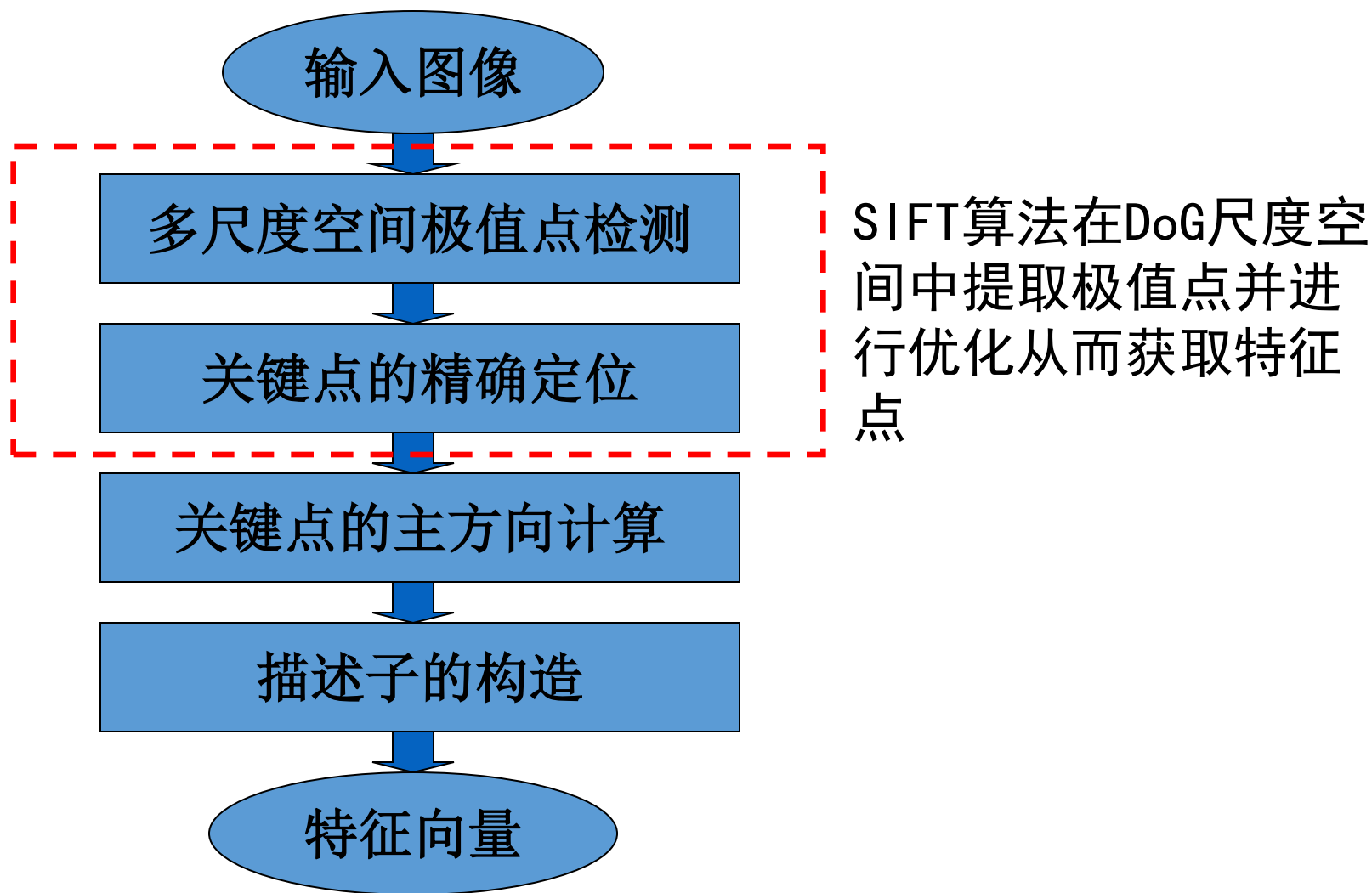
SIFT — Scale Invariant Feature Transform

- 一种特征提取算法 (Lowe 1999, 2004)
- DoG 特征检测 + SIFT 描述子
- SIFT描述子性能评价最佳 (Mikolajczyk&Schmid, 2005)
- 迄今使用最为广泛的一种特征

SIFT特征的性质

- 不变性
 - 对图像的旋转和尺度变化具有不变性
 - 对三维视角变化和光照变化具有很强的适应性
 - 局部特征，在遮挡和场景杂乱时仍保持不变性
- 辨别力强
 - 特征之间相互区分的能力强，有利于匹配
- 数量较多
 - 一般 500×500 的图像能提取出约2000个特征点
- 扩展性强

SIFT特征提取算法流程



尺度空间的定义

- 目的：检测在尺度变化时仍然稳定的特征
- 定义图像 $I(x, y)$ 的尺度空间：

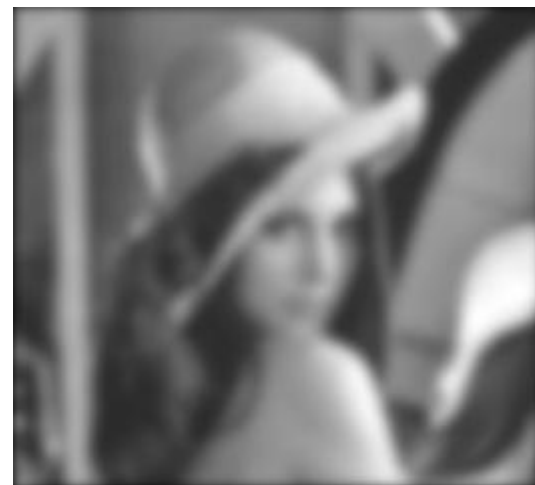
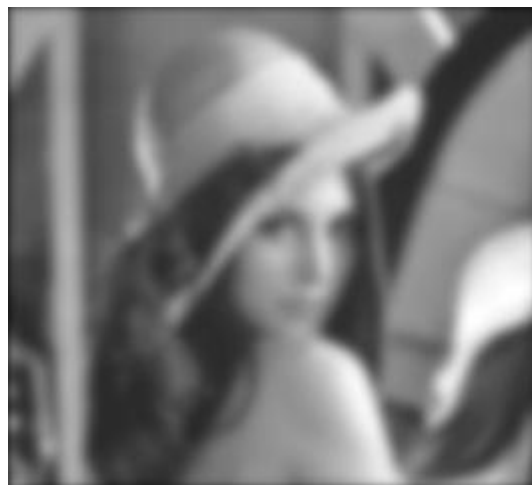
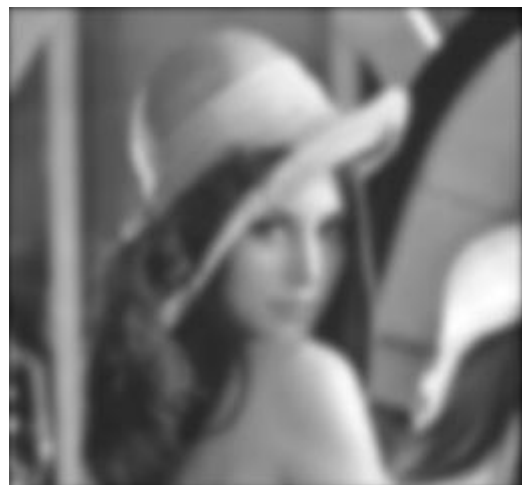
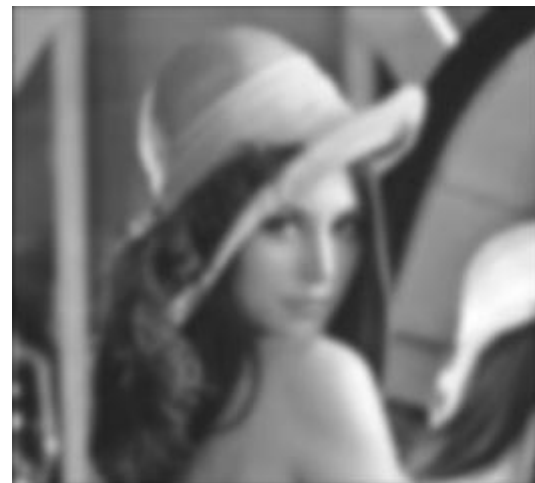
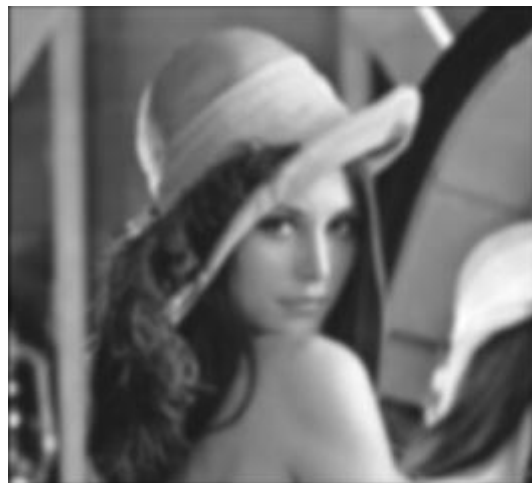
$$L(x, y, \sigma) = G(x, y, \sigma) * I(x, y)$$

其中：

$$G(x, y, \sigma) = \frac{1}{2\pi\sigma^2} e^{-(x^2+y^2)/2\sigma^2}$$

尺度参数 σ ，当 σ 连续变化， $G(x, y, \sigma)$ 构成图像的尺度空间

图像尺度空间



高斯差分尺度空间

- 为了在尺度空间中检测稳定的关键点，构造高斯差分尺度空间
- 高斯差分尺度空间（高斯差分DoG: Difference of Gaussian）

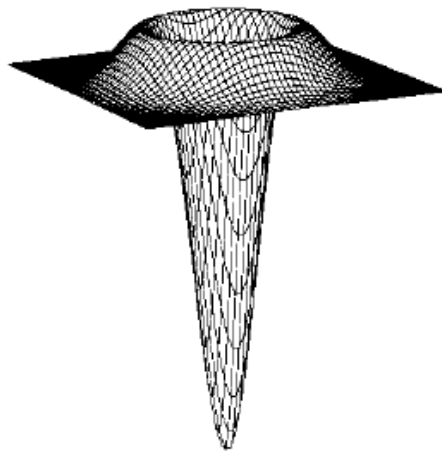
$$\begin{aligned} D(x, y, \sigma) &= (G(x, y, k\sigma) - G(x, y, \sigma)) * I(x, y) \\ &= L(x, y, k\sigma) - L(x, y, \sigma) \end{aligned}$$

几个理由：

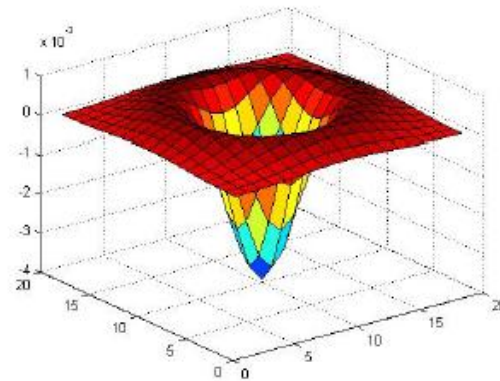
- (1) 计算效率高：高斯卷积，减法
- (2) 高斯差分是对尺度归一化LoG的一个很好的近似，而尺度归一化的LoG空间具有真正的尺度不变性 (Lindeberg 1994)
- (3) 实验比较表明，从尺度归一化LoG空间中提取的图像特征的尺度稳定性最好，优于梯度、Hessian或Harris角点函数。

高斯差分尺度空间

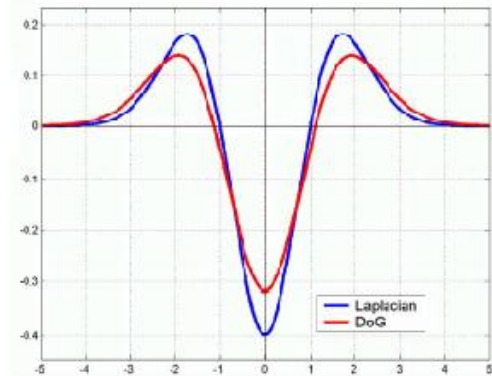
- LoG VS DoG



LoG



DoG

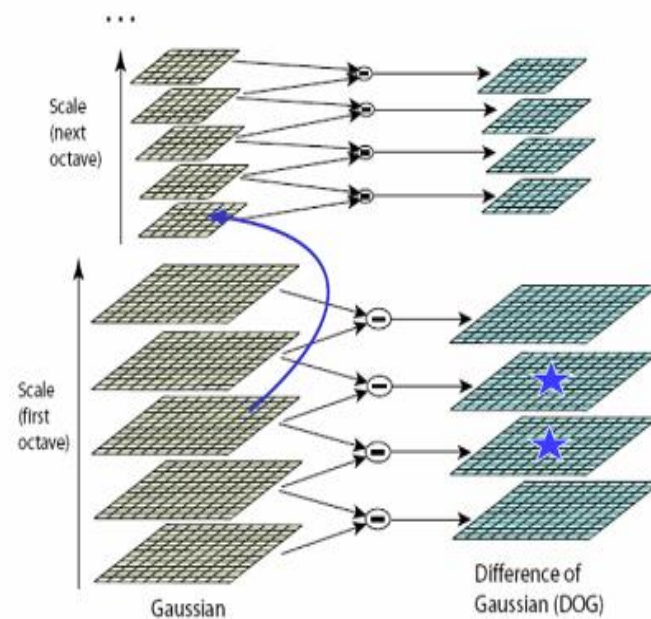
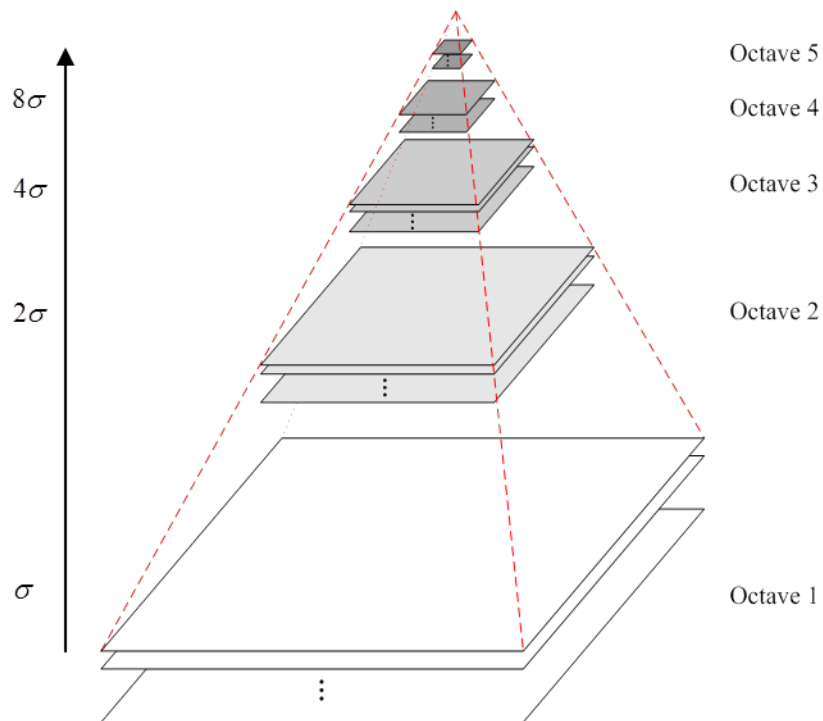


(b)

高斯差分尺度空间

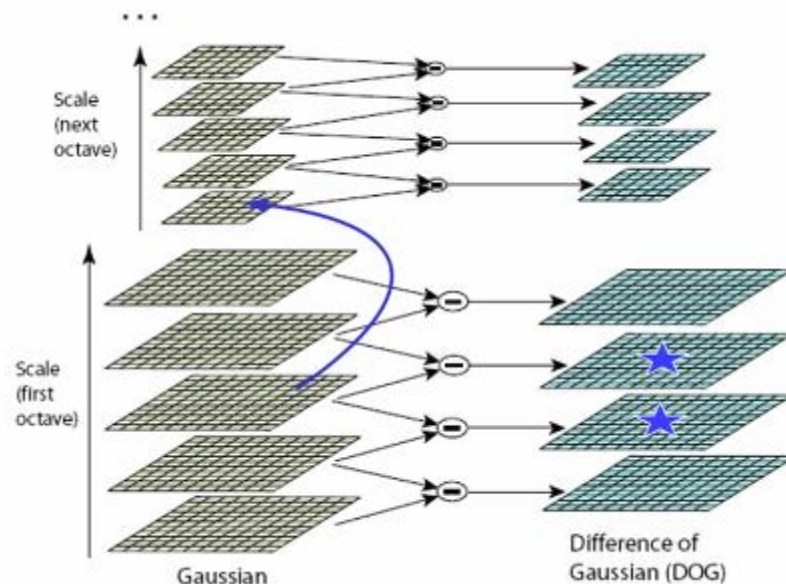


高斯金字塔



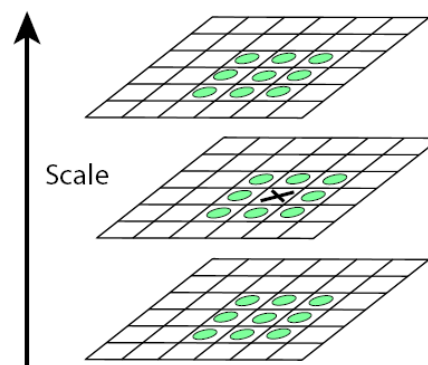
DoG尺度空间极值点检测

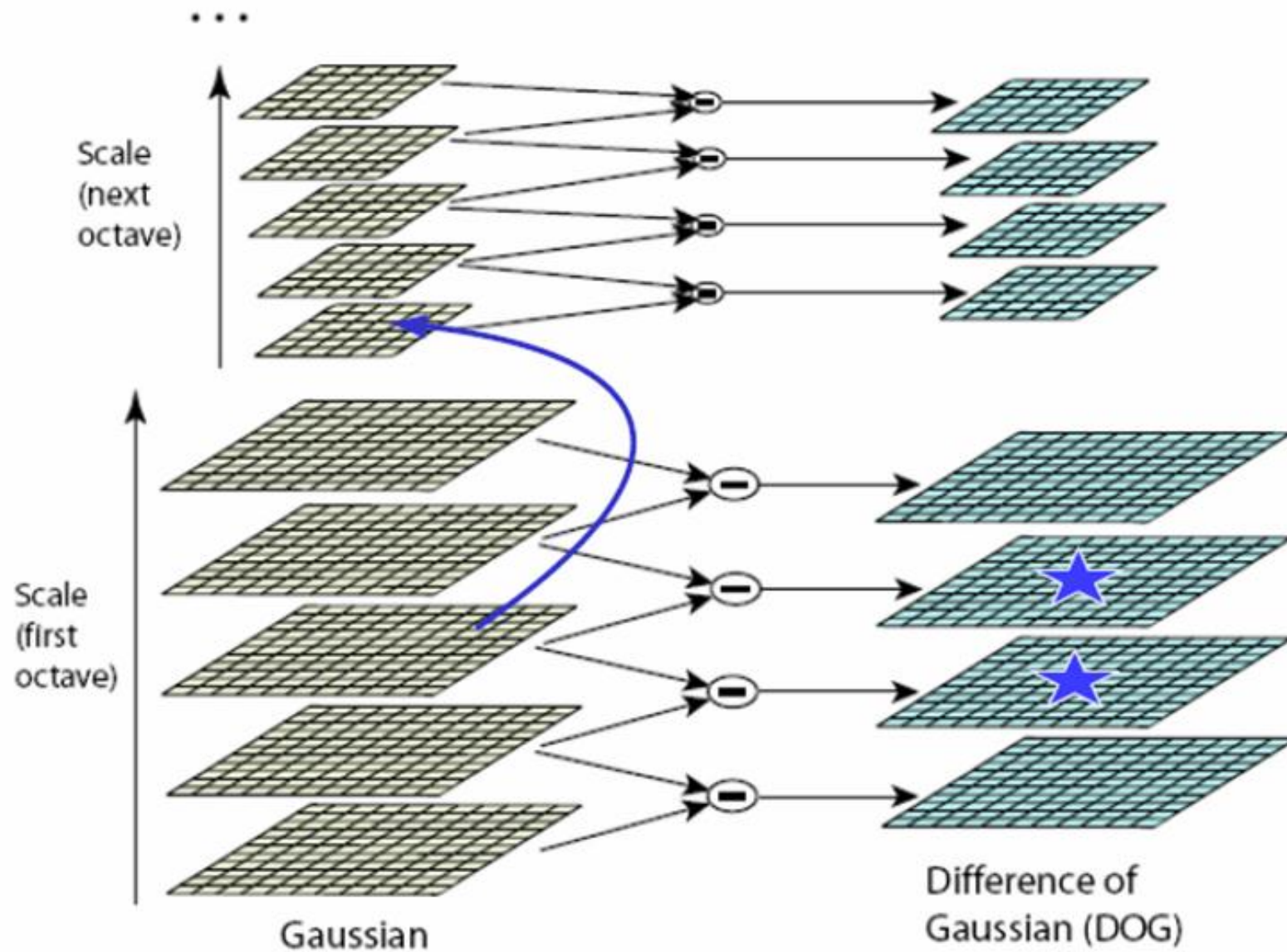
建立的三维尺度空间形式



极值点检测：

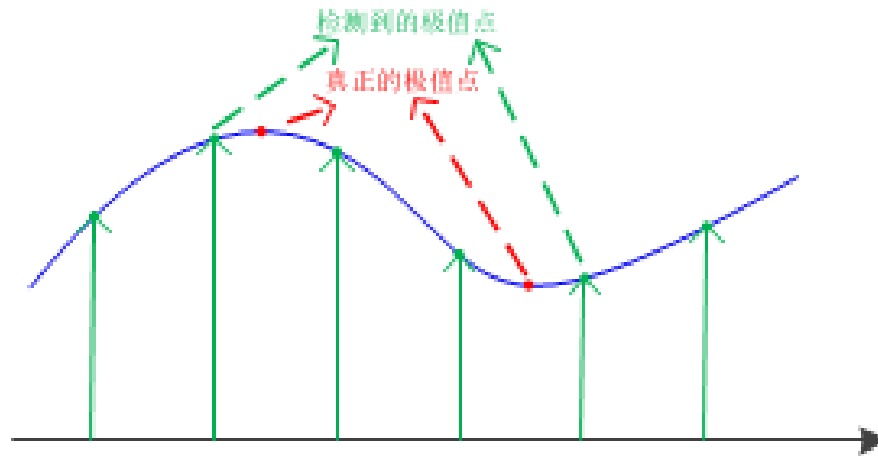
在三维尺度空间 $D(x, y, \sigma)$ 中，搜索每个点的26邻域，若该点为局部极值点，则保存为候选关键点。





关键点的精确定位

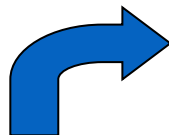
- 动机：在离散采样中搜索到的极值点不一定是真实空间的极值点。
- 基本原理：对尺度空间DoG函数进行曲线拟合，计算其极值点，从而实现关键点的精确定位。



关键点的精确定位

- 三元二次函数拟合 $D(x) = D + \frac{\partial D^T}{\partial x} x + \frac{1}{2} x^T \frac{\partial^2 D}{\partial x^2} x$

计算精细偏移量

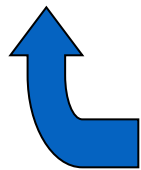


$$\hat{x} = -\frac{\partial^2 D^{-1}}{\partial x^2} \frac{\partial D}{\partial x}$$



改变关键点的
初始位置 x ，
重复计算精细
偏移量

若 $\hat{x} = (\hat{x}, \hat{y}, \hat{\sigma})^T$
中的三个变量任意
一个偏移量大于0.5
(精确极值点更接近
于另一个邻点)



去除不稳定的关键点

❖ 去除对比度低的点 $D(\hat{x}) = D + \frac{1}{2} \frac{\partial D^T}{\partial x} \hat{x}$

设定阈值 $|D(\hat{x})| < 0.03$

❖ 去除边缘上的点

利用Hessian矩阵 H 判断关键点是否位于边缘

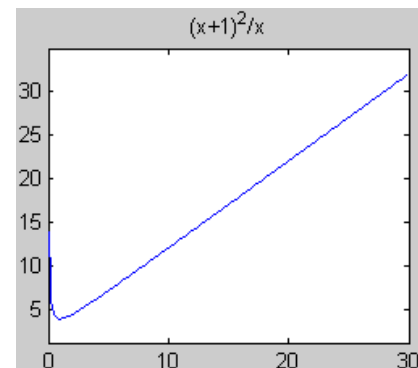
$$H = \begin{bmatrix} D_{xx} & D_{xy} \\ D_{xy} & D_{yy} \end{bmatrix}$$

H 的两个特征值对应两个方向上的主曲率大小。
在边缘位置，两个特征值的比值

设两个特征值之间的比值为 r ： $\lambda_1 = r\lambda_2$

$$\frac{\text{Tr}^2(H)}{\text{Det}(H)} = \frac{(\lambda_1 + \lambda_2)^2}{\lambda_1 \lambda_2} = \frac{(r\lambda_2 + \lambda_2)^2}{r\lambda_2 \lambda_2} = \frac{(r + 1)^2}{r} > \frac{(10 + 1)^2}{10}$$

通过计算 H 的迹和行列式避免直接求特征值，减少了计算量.



例子



原始图像
大小233*189像素

多尺度DoG空间中的极值点
832个

例子



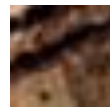
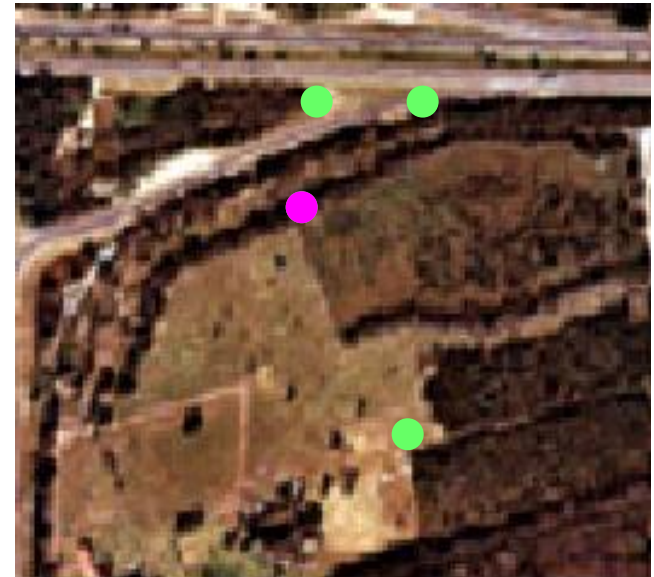
对比度阈值处理
832 \rightarrow 729



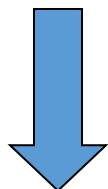
去除边缘上的点
729 \rightarrow 536

特征点描述与匹配

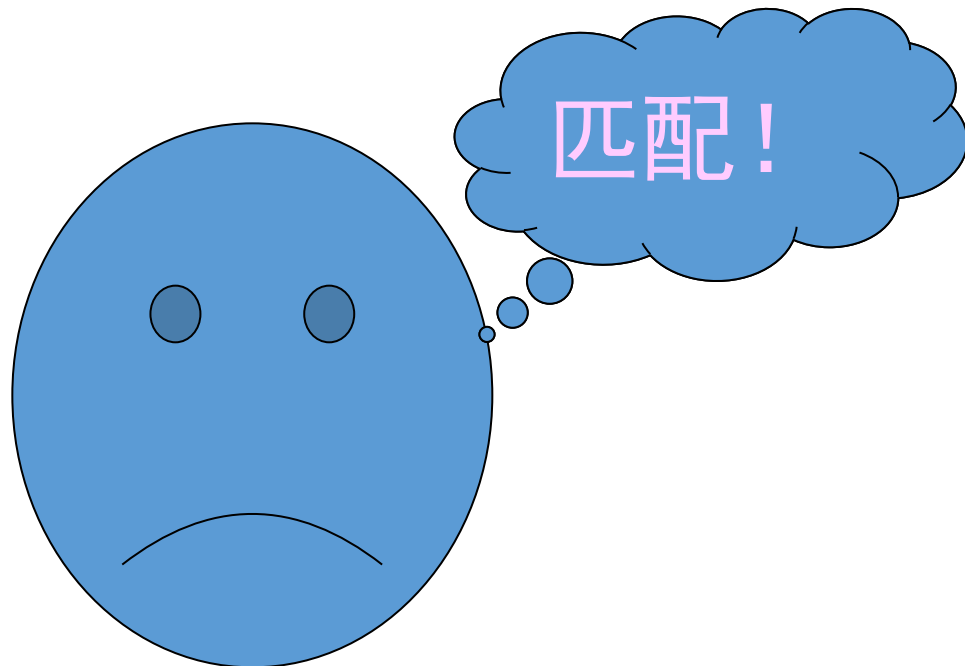
特征点描述与匹配示例



数字图像



特征提取



匹配：确定不同图像中对应空间同一物体的投影的过程

匹配是基于多幅图像视觉问题的的基本步骤

特征匹配



```
graph TD; A[特征匹配] --> B[点匹配]; A --> C[直线匹配]; A --> D[曲线匹配]; A --> E[区域匹配];
```

点匹配

直线匹配

曲线匹配

区域匹配

点匹配的基本原则

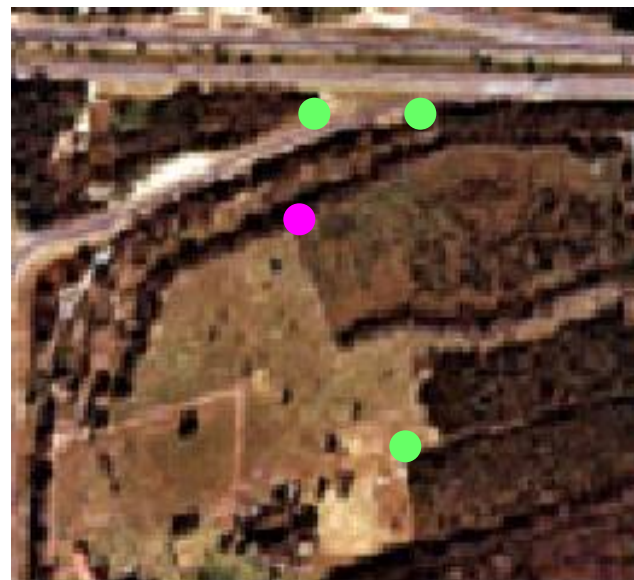
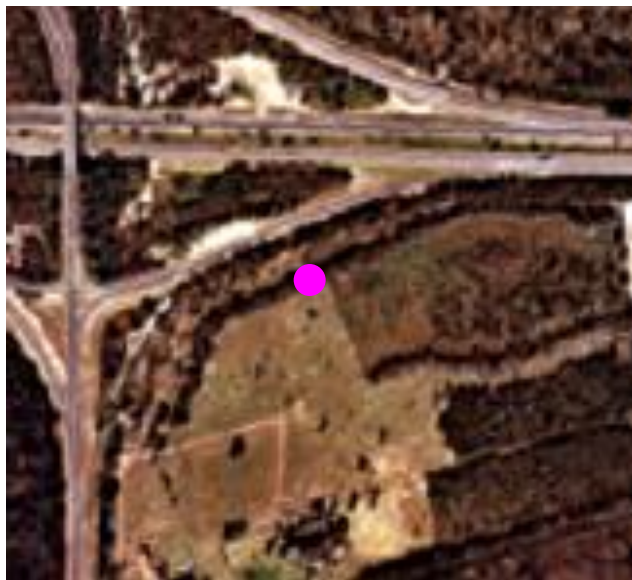
利用图像点周围的信息来描述点，如灰度信息，颜色信息，梯度信息等，然后进行相似性度量

Cross-correlation基本原理

特征点匹
配

利用**相关函数**，评价两幅图像特征点邻域的灰度**相似性**以确定**对应点**。

示意图



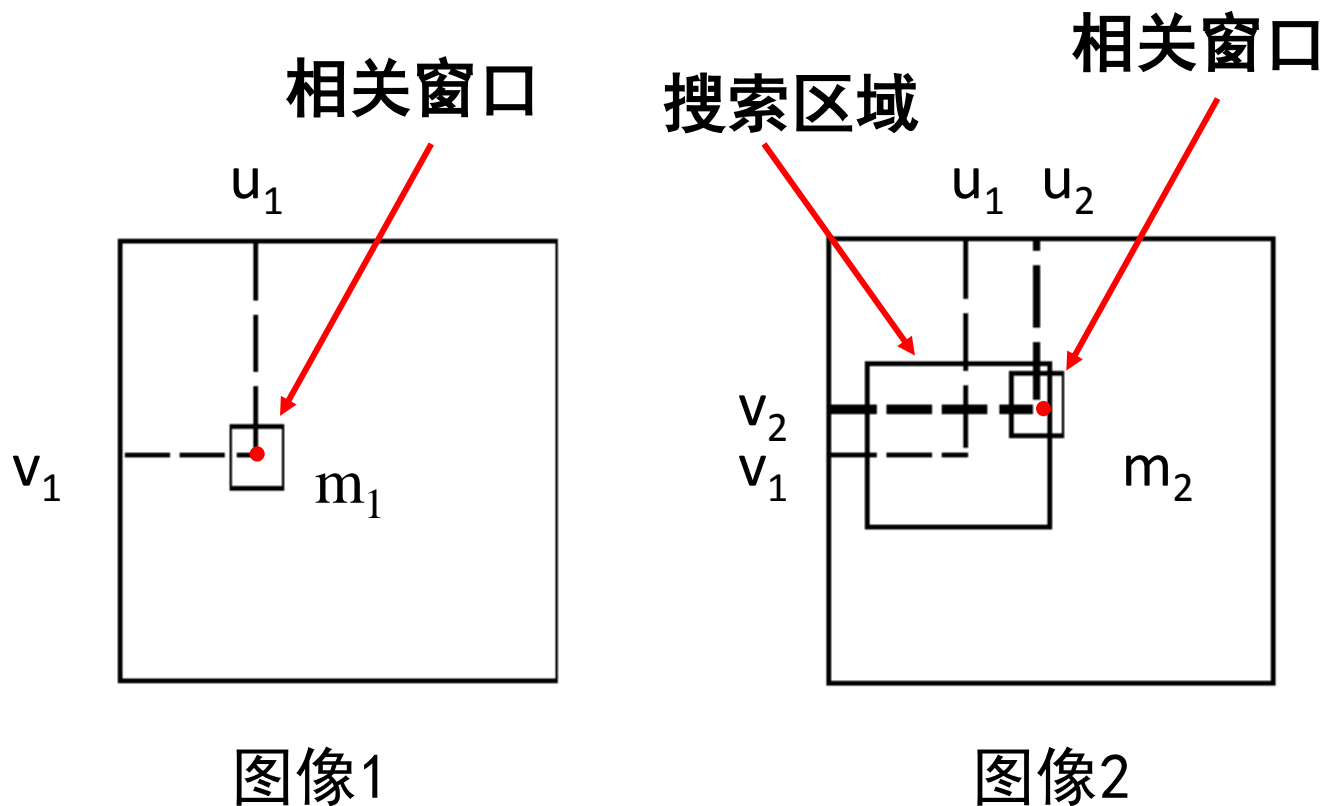
相关函数



对应点

相似程度

相关窗口和搜索区域



相关函数

$$\begin{aligned} & \text{Score}(m_1, m_2) \\ &= \frac{\sum_{i=-n}^n \sum_{j=-m}^m [I_1(u_1 + i, v_1 + j) - \bar{I}_1(u_1, v_1)] \times [I_2(u_2 + i, v_2 + j) - \bar{I}_2(u_2, v_2)]}{(2n + 1)(2m + 1)\sqrt{\sigma^2(I_1) \times \sigma^2(I_2)}} \end{aligned}$$

$$\bar{I}_k(u, v) = \frac{\sum_{i=-n}^n \sum_{j=-m}^m I_k(u + i, v + j)}{(2n + 1)(2m + 1)}$$

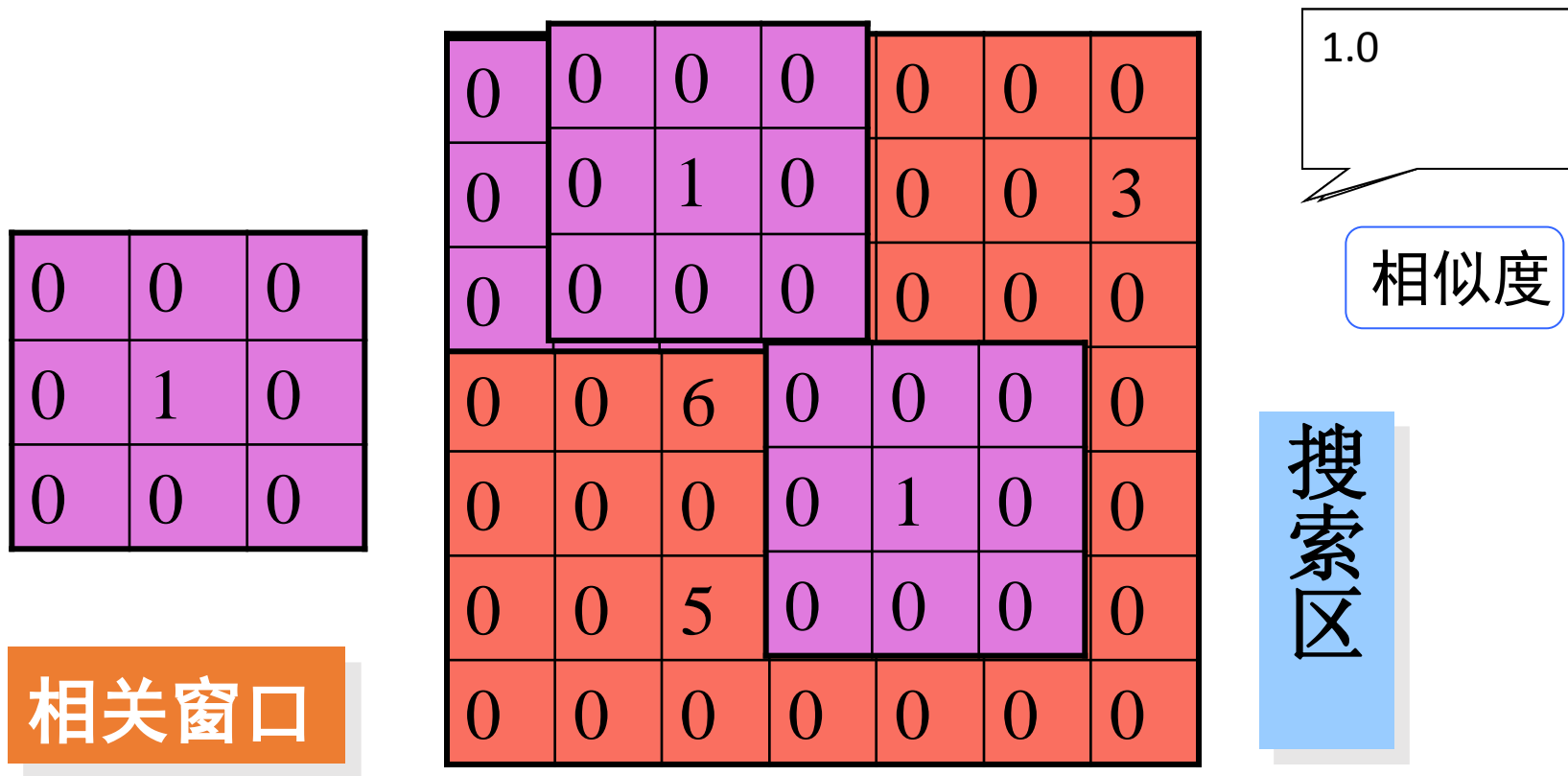
均值

$$\sigma(I_k) = \sqrt{\frac{\sum_{i=-n}^n \sum_{j=-m}^m [I_k(u + i, v + j) - \bar{I}_k(u, v)]^2}{(2n + 1)(2m + 1)}}$$

标准差

搜索过程

利用计算机对数字影像进行数值计算完成匹配



匹配

左图每个特征点对应右图的一个候选匹配集

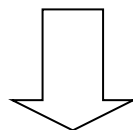
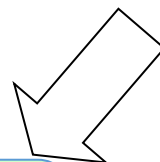
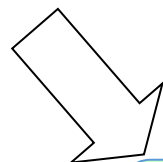
右图每个特征点对应左图的一个候选匹配集

左图特征点集

右图特征点集

匹配方法

对应点集



Cross-correlation特性

- 基于图像灰度
- 如何确定窗口大小和形状是最大的问题
- 没有旋转不变性
- 对光照变化敏感
- 计算代价大

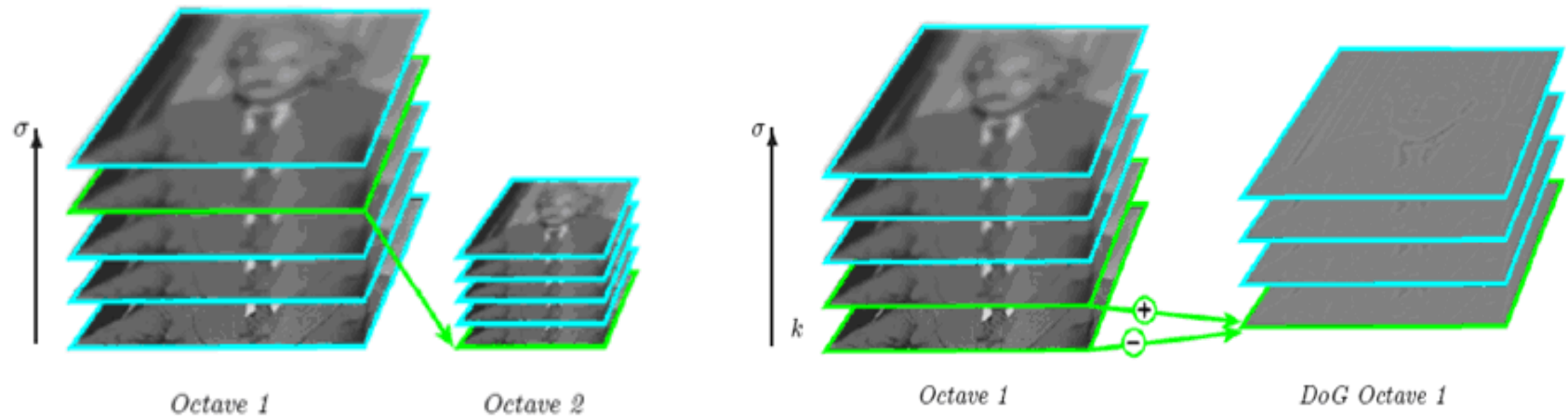
SIFT特征点检测 (II)

- David G. Lowe, Distinctive Image Features from Scale-Invariant Keypoints, International Journal of Computer Vision, Vol. 60, Page 91-110, 2004.

SIFT特征匹配算法

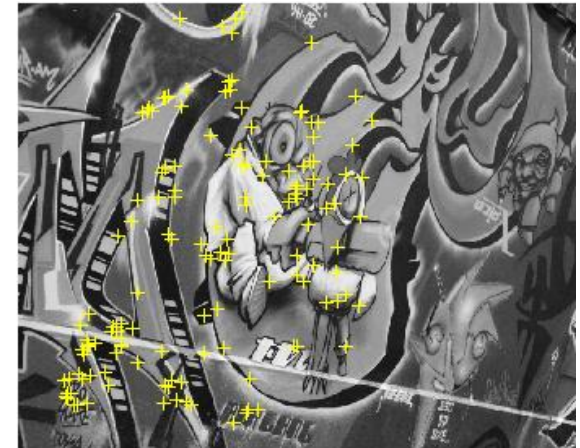
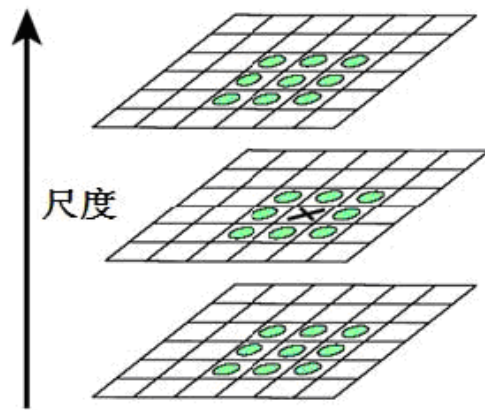
- SIFT特征的提取 (DoG特征检测算子)
- SIFT描述子的构造
- 匹配方法

SIFT特征提取再回顾



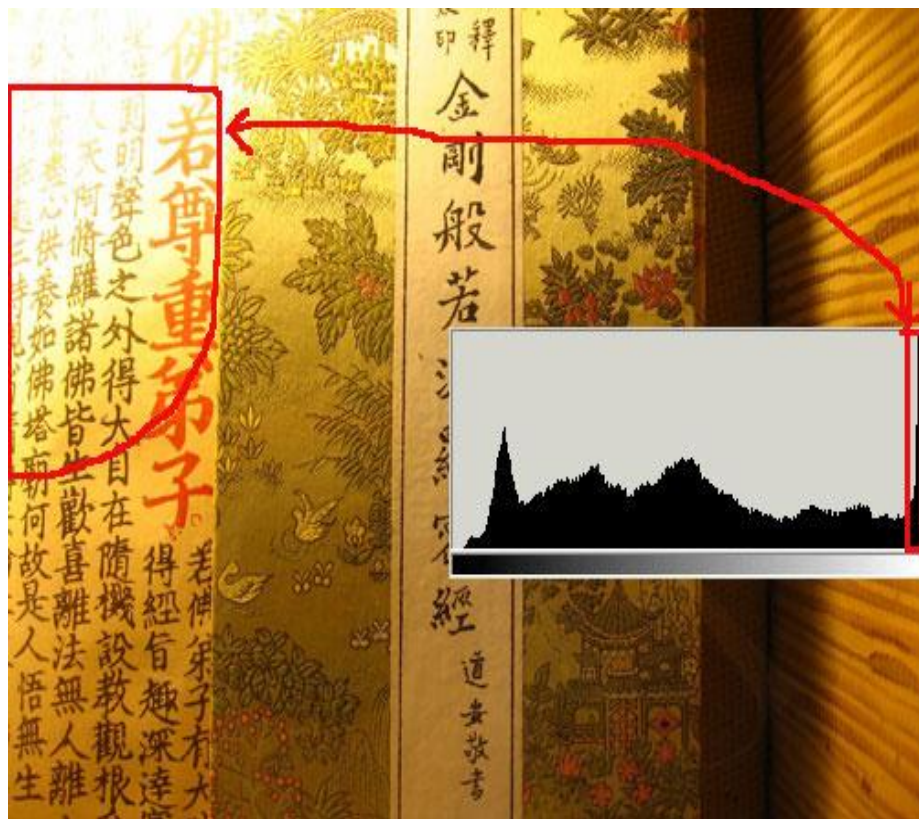
建立高斯差分金字塔

SIFT特征提取再回顾



局部极值检测 . . .

直方图(Histogram)



直方图表示

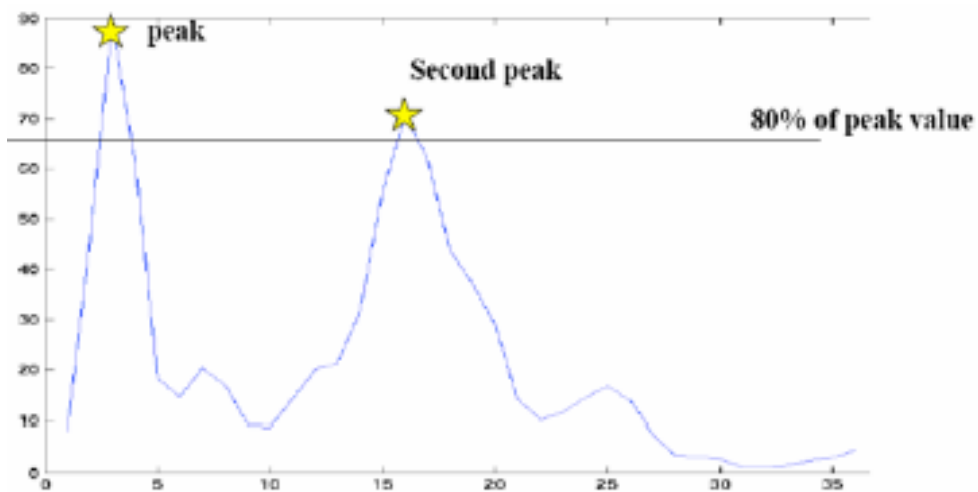
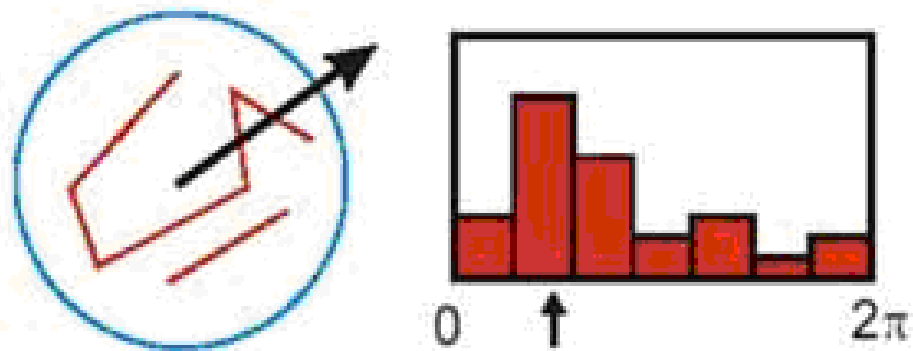
中每
计
示
示频数

原始定义

推广

一般意义

主方向计算



分圆周为角度区间

统计梯度幅值和

抛物线插值

精确定位主方向

特征点主方向

$$\Omega_r = \{X_j: ||X_j - X|| < r\}$$

以特征点X为中心的圆形邻域

$$\theta_{main}(X) = \arg \max_{\theta \in [0, 2\pi)} \left\{ \sum ||\nabla f(X_i)|| : X_i \in \Omega_r, \text{Arg}(\nabla f(X_i)) = \theta \right\}$$

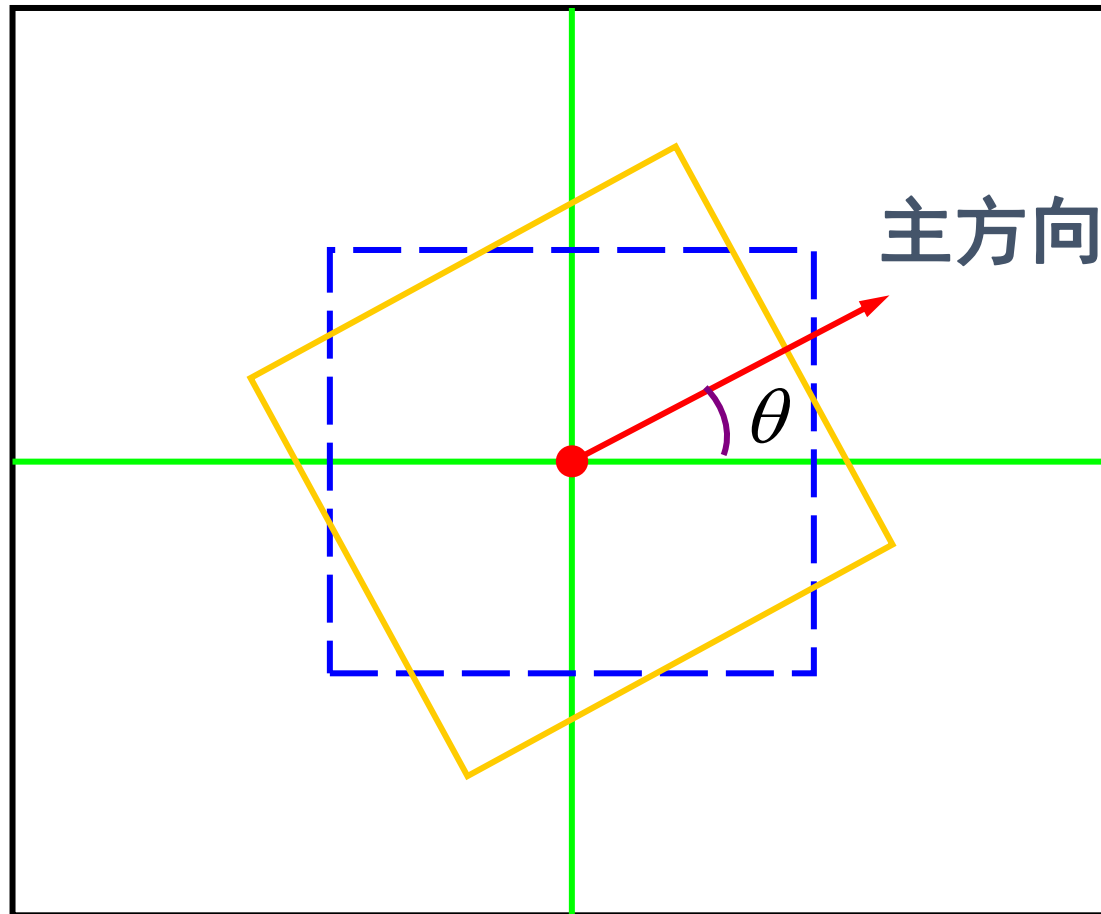
$$\nabla f(X_i)$$

梯度

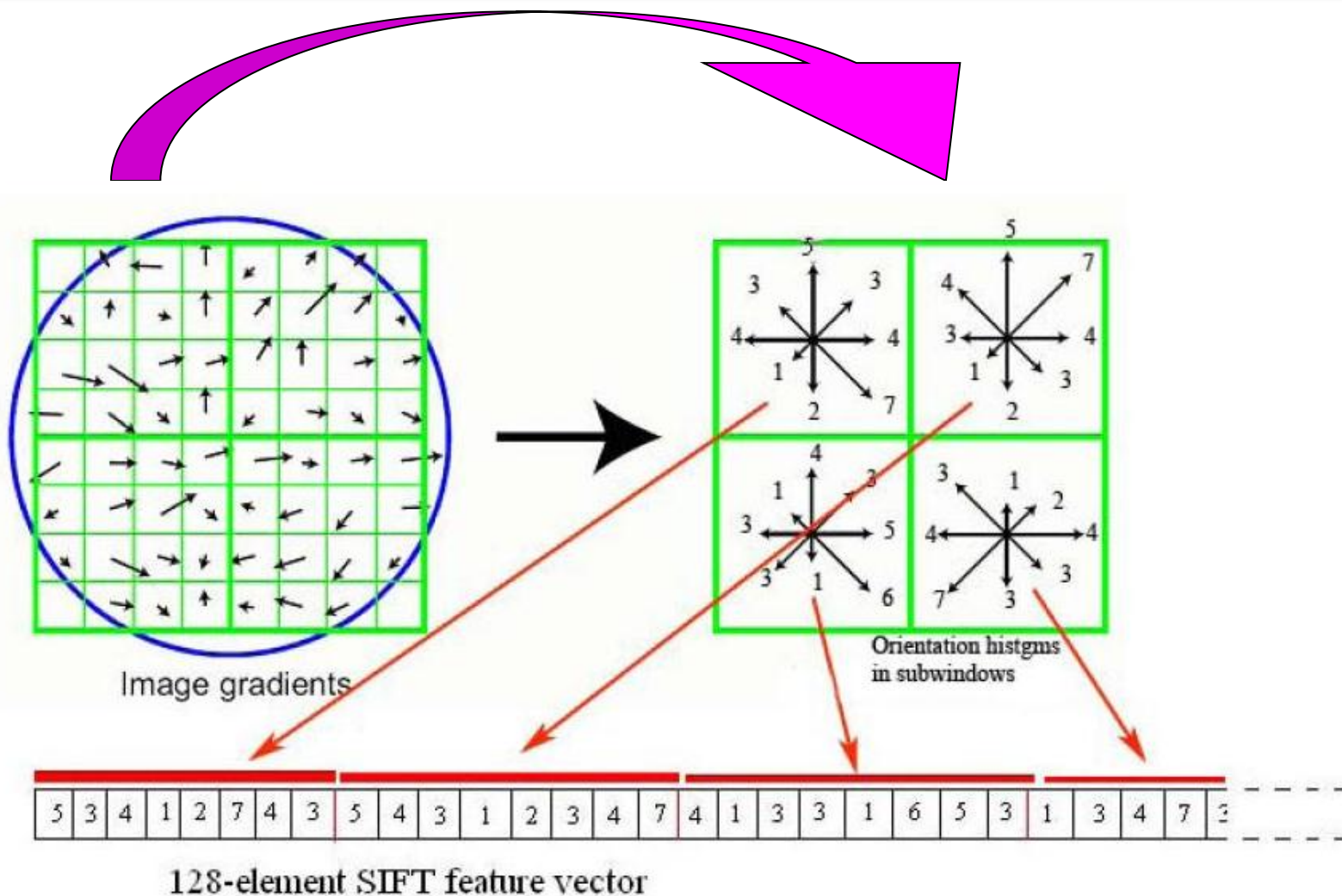
$$\text{Arg}(\nabla f(X_i))$$

梯度幅角

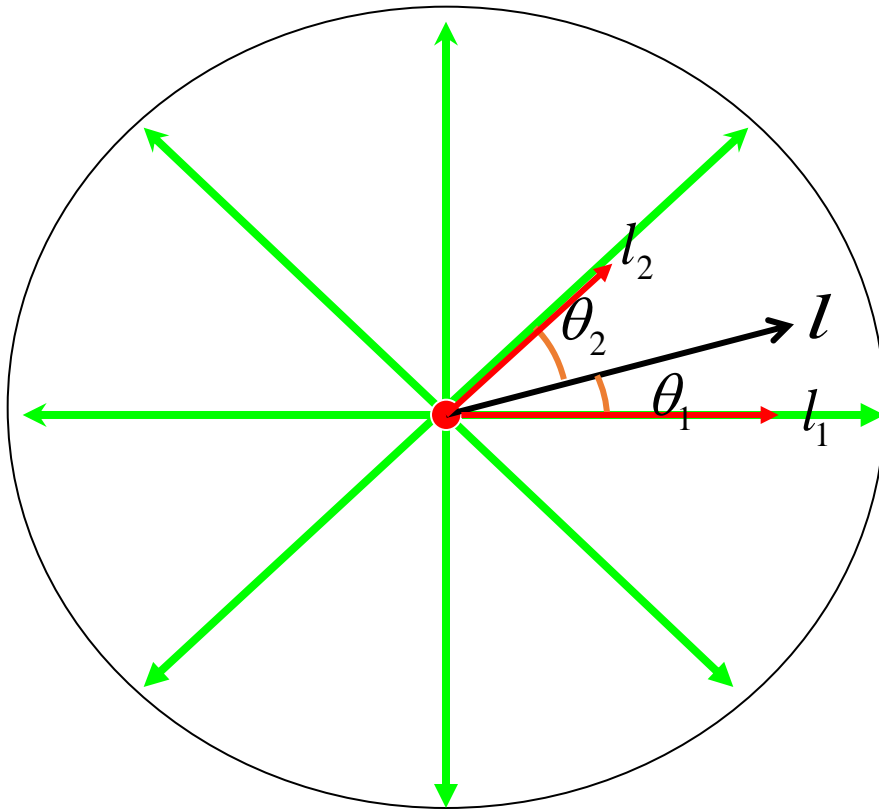
主方向校正



SIFT描述子的构造



梯度幅值的插值运算

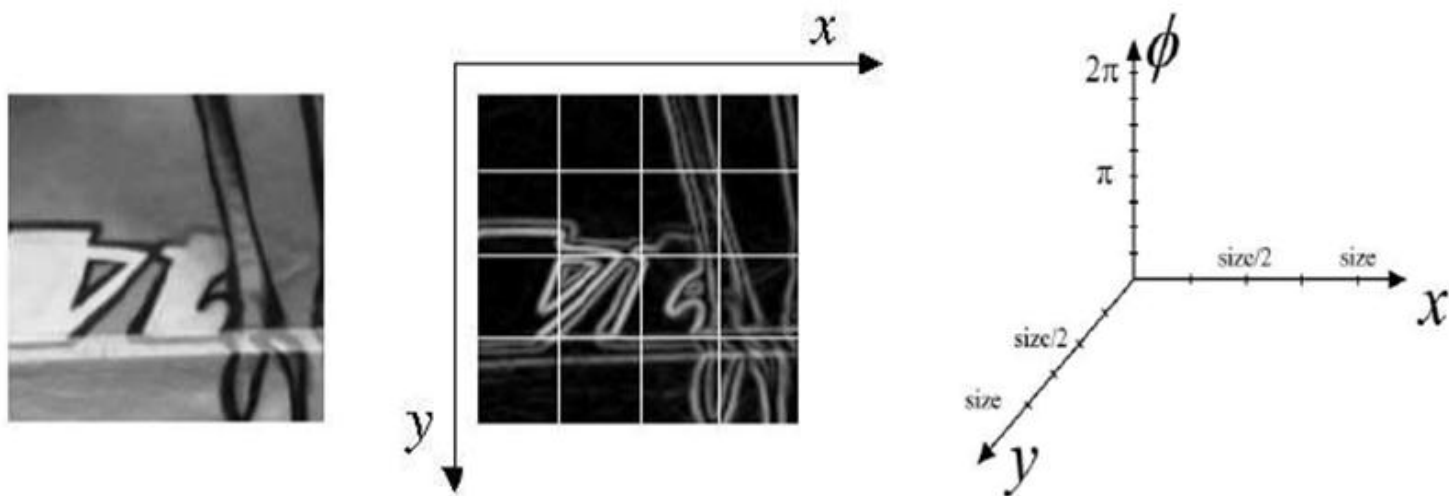


$$l_1 = \frac{\theta_2}{\theta_1 + \theta_2} l$$

$$l_2 = \frac{\theta_1}{\theta_1 + \theta_2} l$$

SIFT描述子的构造

- 三维直方图的计算
 - 将关键点邻域划分为 4×4 个子区域
 - 梯度方向划分为8个方向
 - 直方图的价值为梯度幅值的累加
 - 得到一个 $4 \times 4 \times 8 = 128$ 维向量



SIFT描述子的特点

- SIFT描述子的特点
 - 直方图统计：提高了对图像局部形变的适应能力
 - 子区域划分：弥补了丢失的位置信息，增强辨别力
 - 16×16 的邻域和 4×4 的子区域都进行了类似于高斯函数的加权处理，强化中心区域，淡化边缘区域的影响

SIFT描述子的不变性

- 尺度不变
 - 根据关键点的尺度选取高斯图像和邻域大小
- 旋转不变
 - 将邻域内每点的位置和梯度方向根据关键点的主方向进行旋转
- 适应复杂几何变形
 - 采用分块直方图统计、高斯加权等细节处理
- 适应复杂光照变化
 - 线性光照：归一化128维向量
 - 非线性光照：将128维中所有大于0.2的元素赋值为0.2

SIFT特征匹配

- 匹配准则

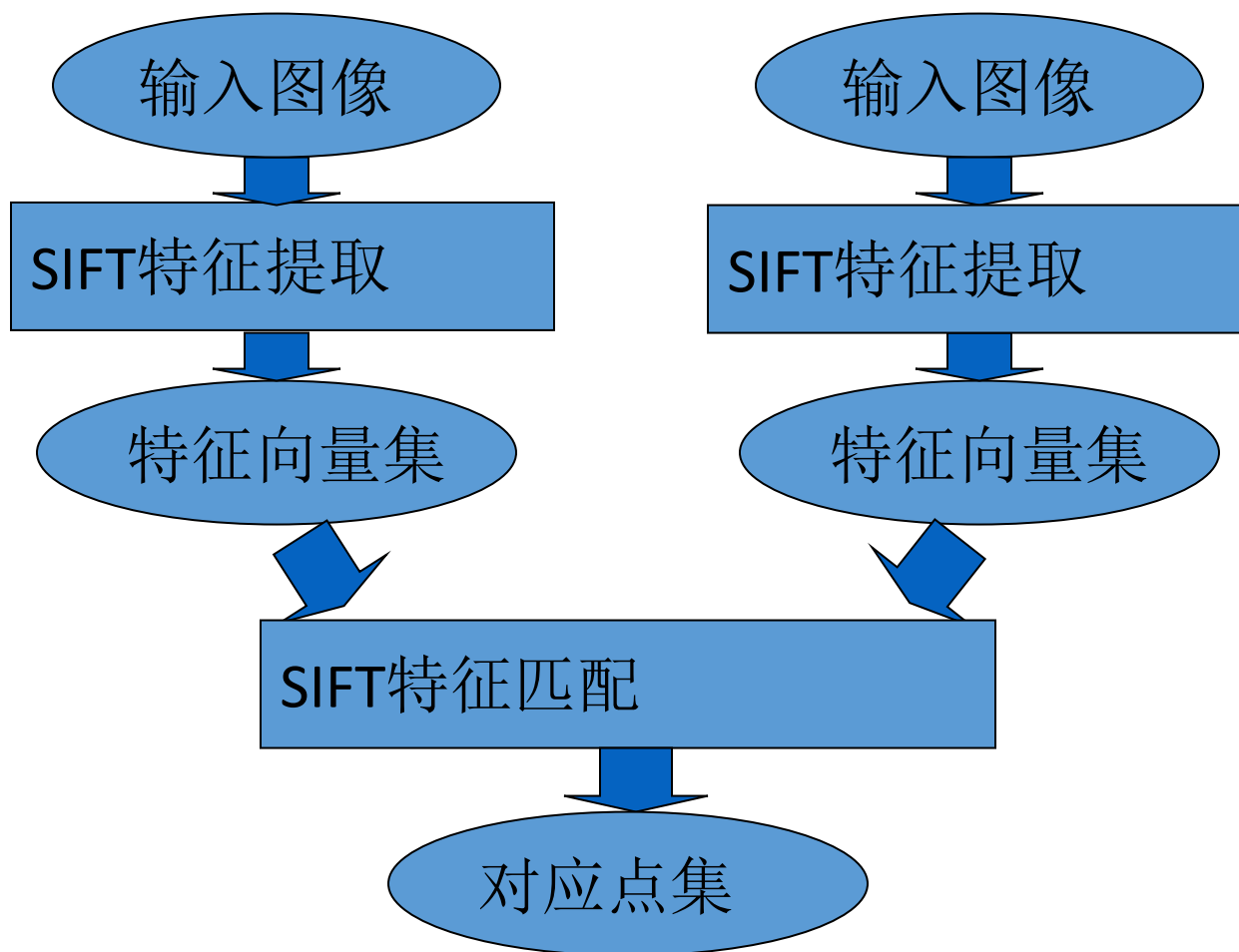
- NNDR (Nearest Neighbor Distance Ratio)

$$NNDR = \frac{d(m + m_1)}{d(m + m_2)}$$

- 搜索策略

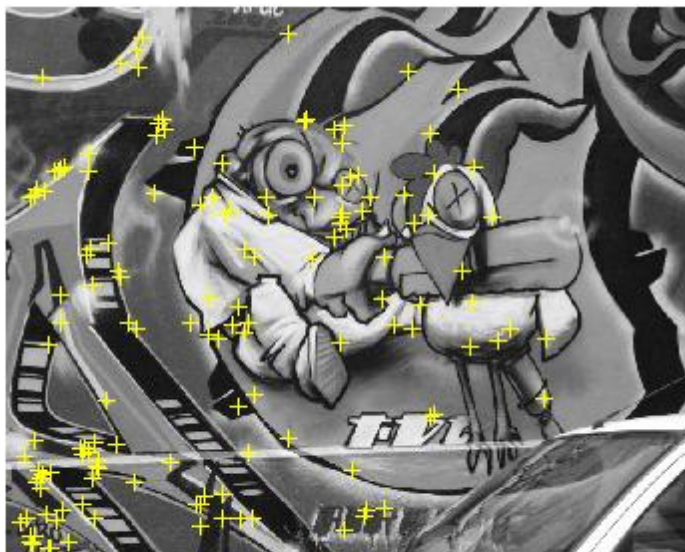
- BBF算法 (Best Bin First) (Beis&Lowe, 1997)

SIFT特征匹配

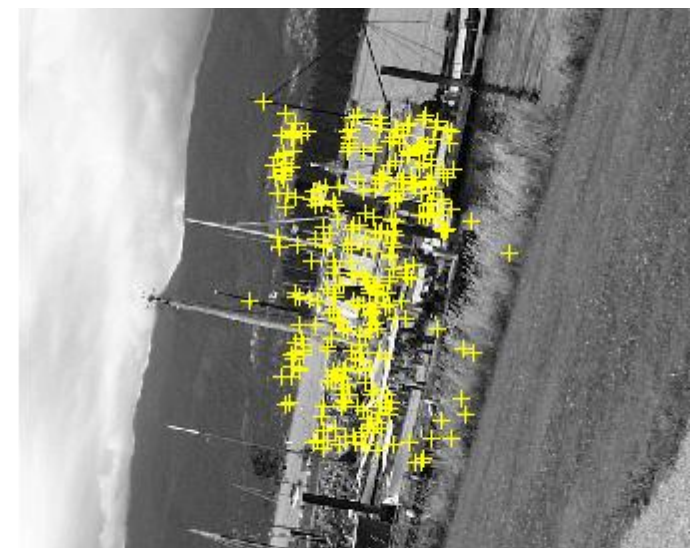


实验结果一视角和旋转变化

视角变化

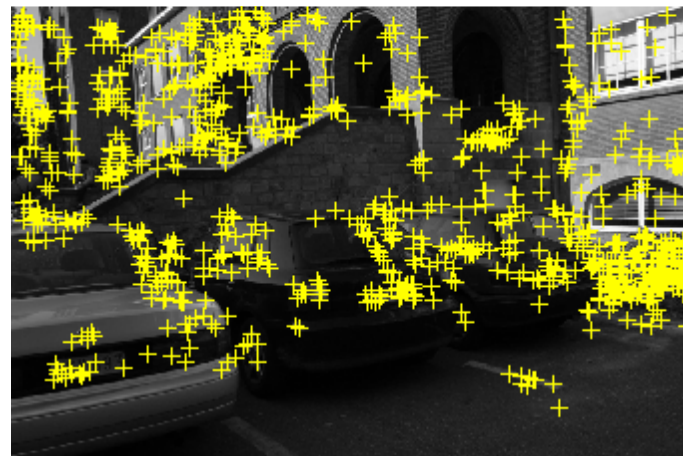


旋转 +
尺度变化

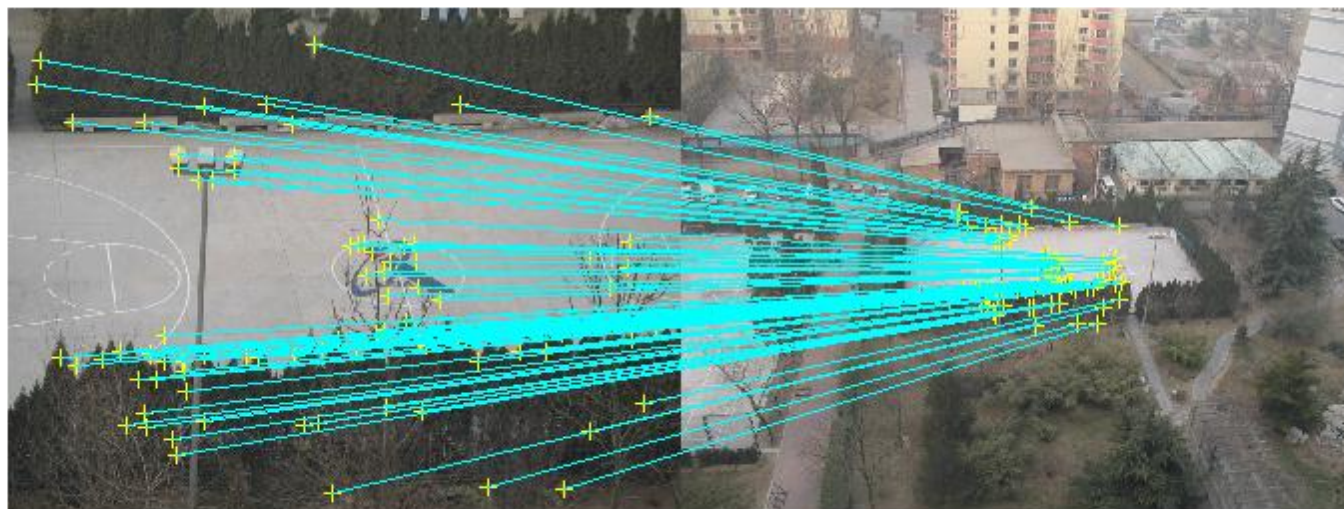


实验结果—光照和尺度变化

光照变化



尺度变化



ORB→BRIEF描述子

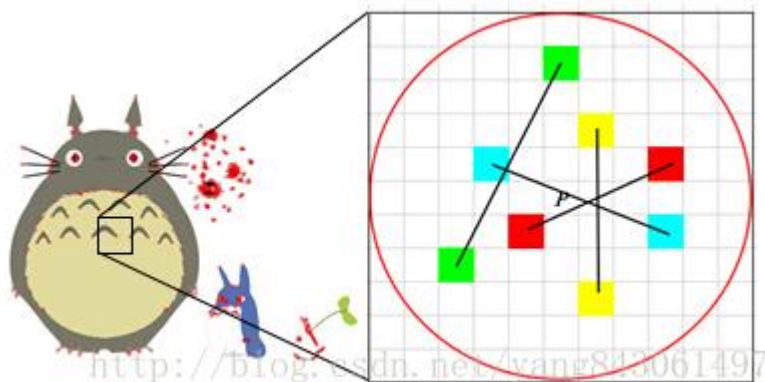
- BRIEF: Binary Robust Independent Elementary Features, ECCV2010.

ORB特征再回顾

- ORB (Oriented FAST and Rotated BRIEF)
 - $ORB = FAST + BRIEF$ (改进的版本)
- 研究动机：快速性、兼顾准确性
- BRIEF: Binary Robust Independent Elementary Features
 - 二进制描述子
 - 高速、低存储

ORB特征再回顾

- BRIEF: Binary Robust Independent Elementary Features
 - 描述子构造方法：在已提取出的特征点的邻域patch中，随机选取若干对像素点，分别比较每对像素点的灰度大小，赋值为1或0，从而构建描述子。



$$\tau(\mathbf{p}; \mathbf{x}, \mathbf{y}) := \begin{cases} 1 & \text{if } \mathbf{p}(\mathbf{x}) < \mathbf{p}(\mathbf{y}) \\ 0 & \text{otherwise} \end{cases}$$

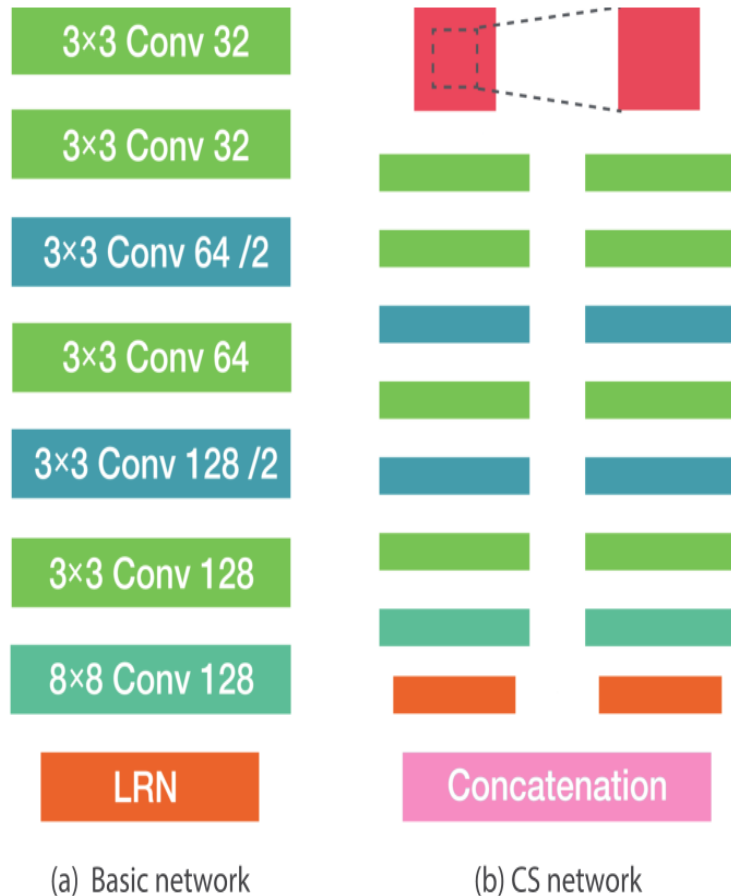
$$f_{n_d}(\mathbf{p}) := \sum_{1 \leq i \leq n_d} 2^{i-1} \tau(\mathbf{p}; \mathbf{x}_i, \mathbf{y}_i)$$

ORB特征再回顾

- 匹配：
 - 汉明距离（Hamming Distance）：在信息论中，两个等长字符串之间的汉明距离是两个字符串对应位置的不同字符的个数。
 - 实例：“1011101”与“1001001”之间的汉明距离是2
- BRIEF存在的不足？
- ORB = FAST + BRIEF（改进的版本）
 - 特征点的处理
 - 多尺度
 - 主方向
 - 课后练习

基于DNN的特征描述子学习

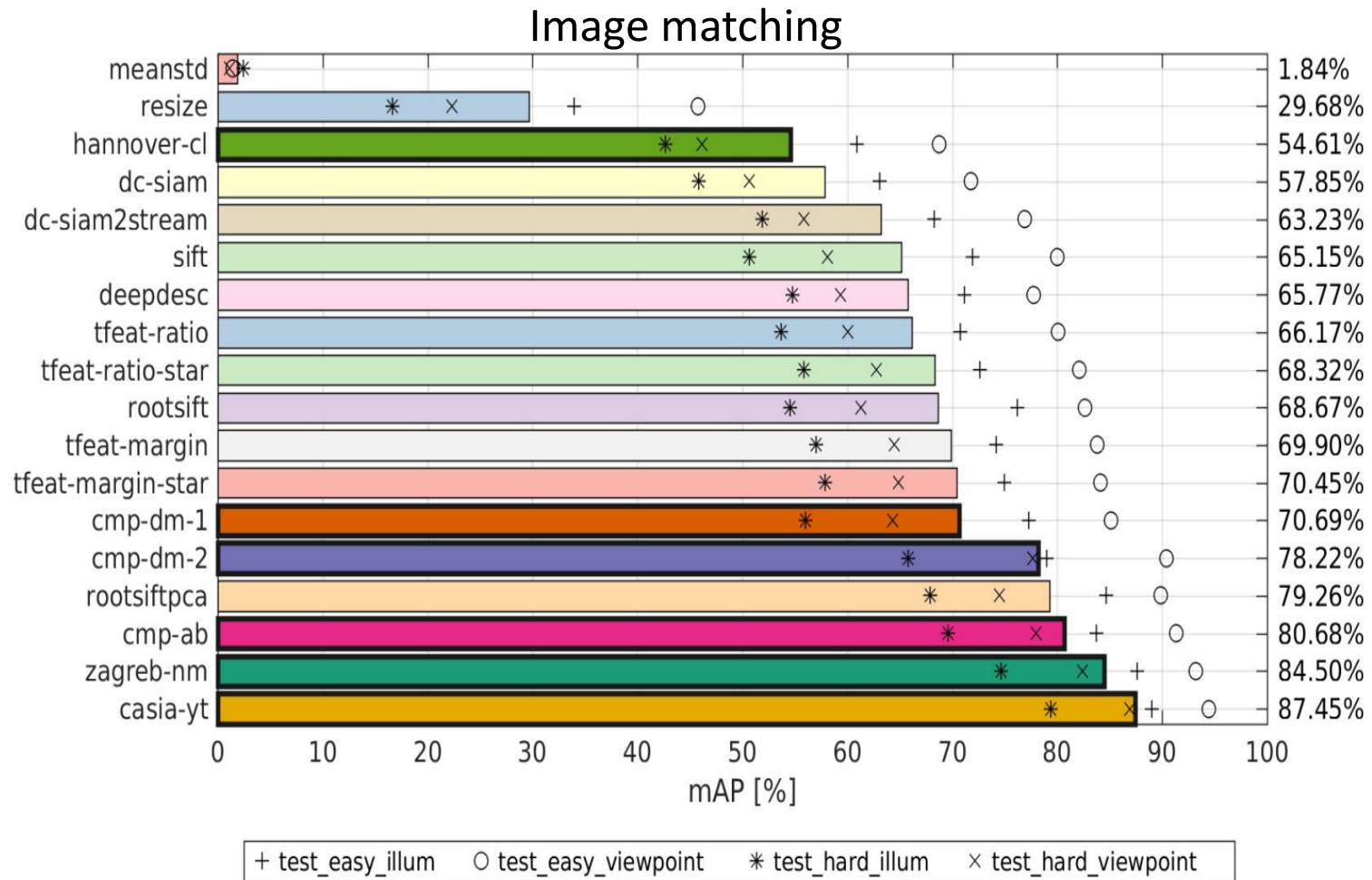
L2-Net: Network architecture



L2-Net takes an all convolution structure

- ◆ Batch normalization (BN) is used after each convolutional layer, but with minor modifications, i.e., we do not update the weighting and bias parameters of the BN layers and fix them to be 1 and respectively.
- ◆ L2-Net converts 32×32 input patches to 128 dimensional descriptors .
- ◆ The central-surround (CS) L2-Net is the concatenation of two separate L2-Nets with a two tower structure.

L2-Net:Experiments



SOSNet

Motivation:

Second Order Similarity (SOS) (二阶相似性)在在图匹配等问题中被广泛应用，相比于**First Order Similarity (FOS)**一阶相似性(一阶相似性)，其优点在与它对各种形状畸变的鲁棒性.。

由于**SOS**计算复杂度高，因此大规模匹配任务仍然基于**FOX**。

为了考虑发挥其作用，考虑在描述子的训练阶段加入**SOS**约束，来提升描述子基于**FOS**的匹配能力。

SOSNet

Implementation:

网络训练过程中的误差信息同时来自两方面：一阶相似性与二阶相似性。

优化FOS的误差

采用了triplet loss 与负样本的mining

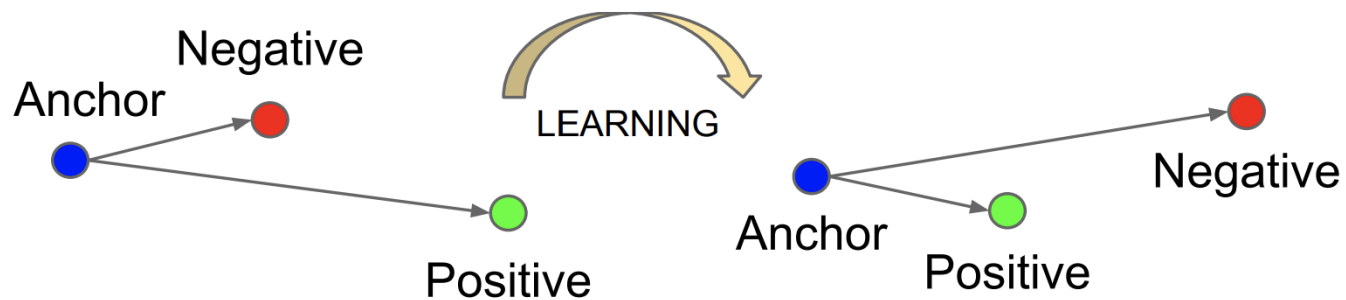
$$\mathcal{L}_{FOS} = \frac{1}{N} \sum_{i=1}^N \max(0, t + d_i^{\text{pos}} - d_i^{\text{neg}})^2,$$

$$d_i^{\text{pos}} = d(\mathbf{x}_i, \mathbf{x}_i^+),$$

$$d_i^{\text{neg}} = \min_{i \neq j} (d(\mathbf{x}_i, \mathbf{x}_j), d(\mathbf{x}_i, \mathbf{x}_j^+), d(\mathbf{x}_i^+, \mathbf{x}_j), d(\mathbf{x}_i^+, \mathbf{x}_j^+))$$

SOSNet

triplet loss



$$\mathcal{L}_{FOS} = \frac{1}{N} \sum_{i=1}^N \max(0, t + d_i^{\text{pos}} - d_i^{\text{neg}})^2,$$

$$d_i^{\text{pos}} = d(\mathbf{x}_i, \mathbf{x}_i^+),$$

$$d_i^{\text{neg}} = \min_{i \neq j} (d(\mathbf{x}_i, \mathbf{x}_j), d(\mathbf{x}_i, \mathbf{x}_j^+), d(\mathbf{x}_i^+, \mathbf{x}_j), d(\mathbf{x}_i^+, \mathbf{x}_j^+))$$

SOSNet

Implementation:

优化SOS的误差

两个描述子间的二阶相似性可表示为‘距离的距离’：

$$d^{(2)}(\mathbf{x}_i, \mathbf{x}_i^+) = \sqrt{\sum_j^N (d(\mathbf{x}_i, \mathbf{x}_j) - d(\mathbf{x}_i^+, \mathbf{x}_j^+))^2}$$

基于二阶相似性的正则项：

$$\mathcal{R}_{SOS} = \frac{1}{N} \sum_i d^{(2)}(\mathbf{x}_i, \mathbf{x}_i^+)$$

总误差：

$$\mathcal{L}_T = \mathcal{L}_{FOS} + \mathcal{R}_{SOS}$$

SOSNet: Experimental Results

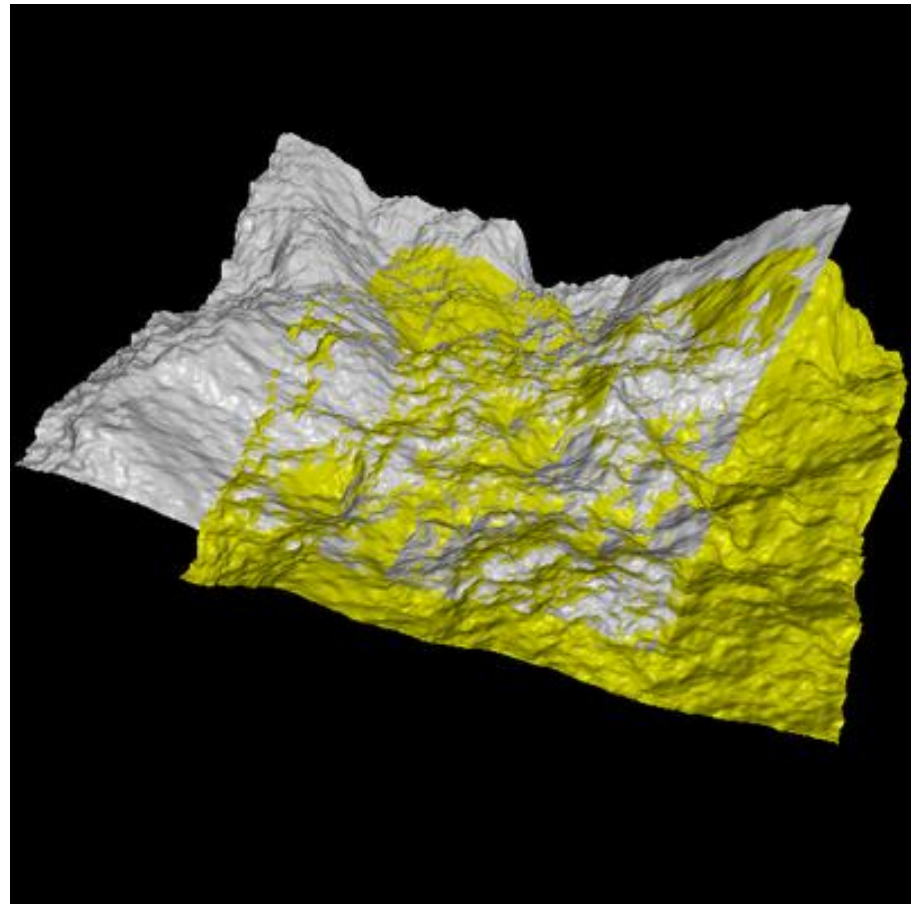
Train	Notredame	Yosemite	Liberty	Yosemite	Liberty	Notredame	Mean
Test	Liberty		Notredame		Yosemite		
SIFT [20]	29.84		22.53		27.29		26.55
DeepDesc [32]	10.9		4.40		5.69		6.99
MatchNet [2]	7.04	11.47	3.82	5.65	11.6	8.70	8.05
L2Net [34]	3.64	5.29	1.15	1.62	4.43	3.30	3.24
CS L2Net [34]	2.55	4.24	0.87	1.39	3.81	2.84	2.61
HardNet [25]	1.47	2.67	0.62	0.88	2.14	1.65	1.57
HardNet-GOR [25, 39]	1.72	2.89	0.63	0.91	2.10	1.59	1.64
Michel <i>et al.</i> [16]	1.79	2.96	0.68	1.02	2.51	1.64	1.77
SOSNet	1.25	2.84	0.58	0.87	1.95	1.25	1.46
TFeat+ [2]	7.39	10.13	3.06	3.80	8.06	7.24	6.64
L2Net+ [34]	2.36	4.70	0.72	1.29	2.57	1.71	2.23
CS L2Net+ [34]	1.71	3.87	0.56	1.09	2.07	1.3	1.76
HardNet+ [25]	1.49	2.51	0.53	0.78	1.96	1.84	1.51
HardNet-GOR+ [25, 39]	1.48	2.43	0.51	0.78	1.76	1.53	1.41
DOAP+ [14]	1.54	2.62	0.43	0.87	2.00	1.21	1.45
DOAP-ST+ [14] [15]	1.47	2.29	0.39	0.78	1.98	1.35	1.38
SOSNet+	1.08	2.12	0.35	0.67	1.03	0.95	1.03
GeoDesc+ [21]	5.47		1.94		4.72		4.05
SOSNet-HP+	2.10		0.79		1.39		1.42

ICP(Iterative Closest Point)

两组点集之间的匹配

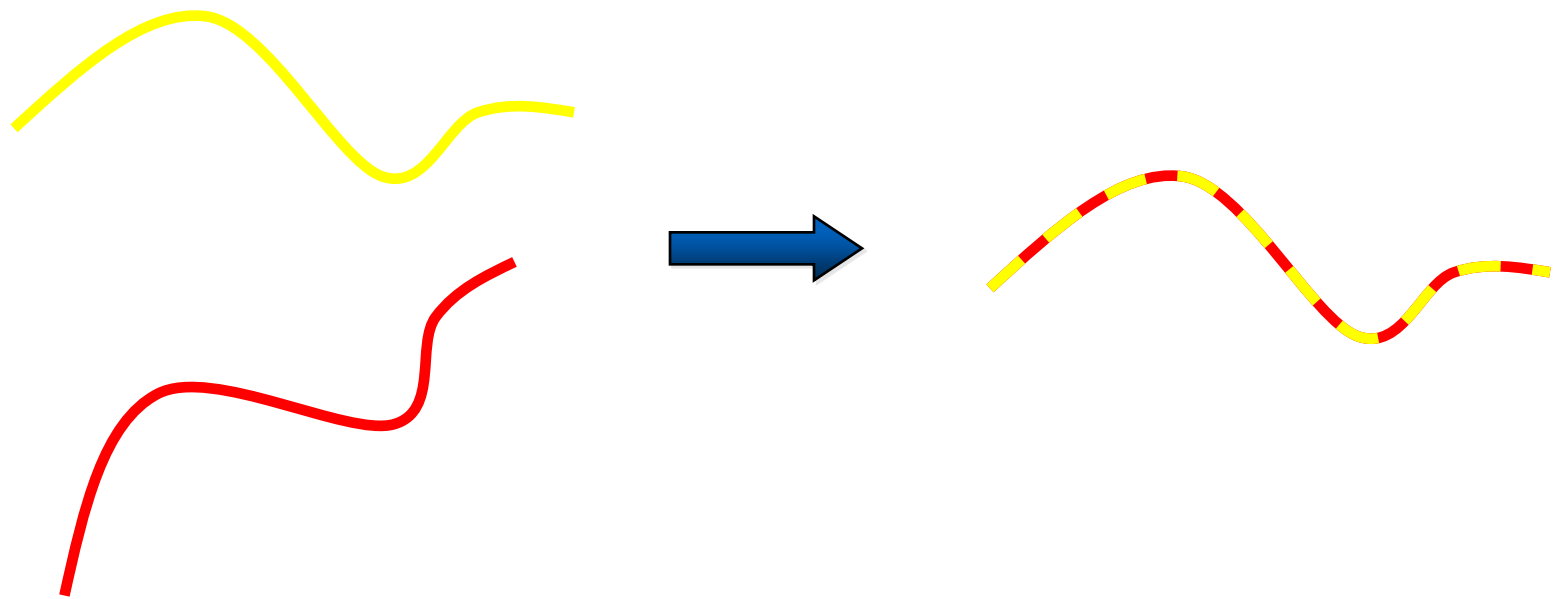
研究背景

- 对齐两个相互交叠的曲面



研究背景

- 对齐两个相互交叠的曲线



问题

- 已知：两个对应点集合

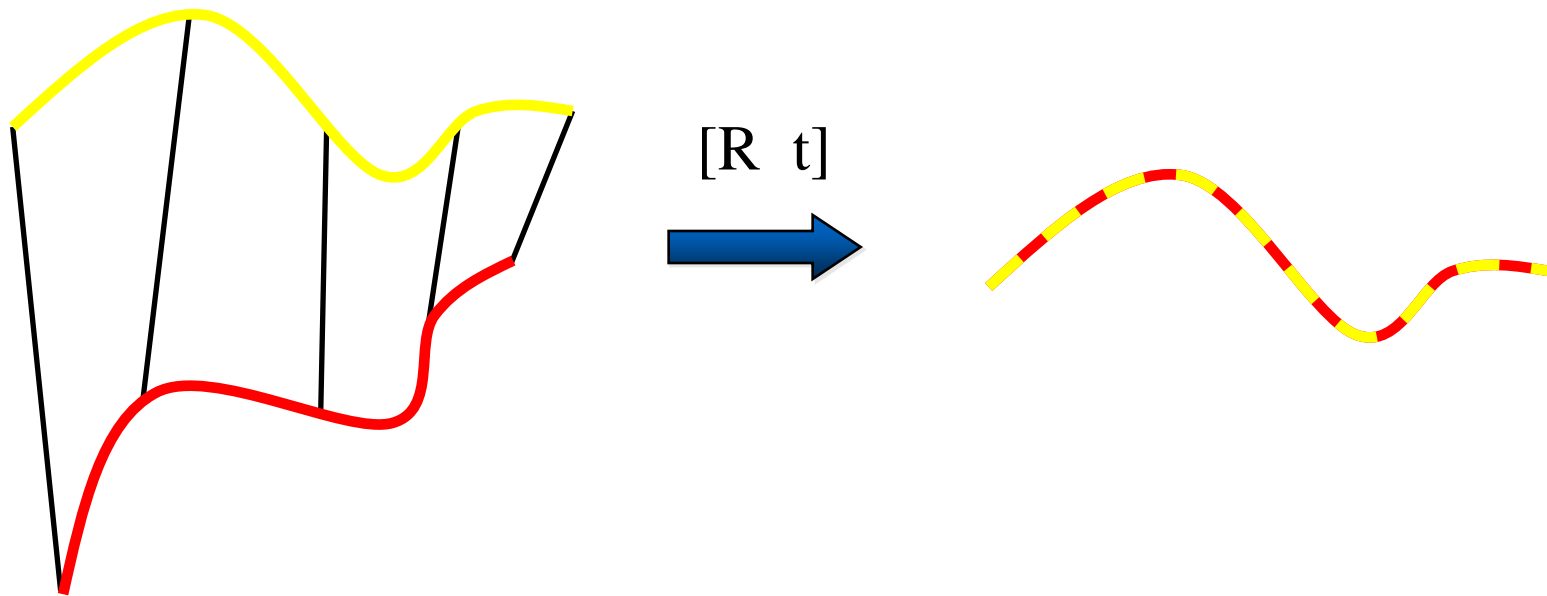
$$X = \{x_1, x_2, x_3, \dots, x_n\} \quad P = \{p_1, p_2, p_3, \dots, p_n\}$$

- 求解：旋转矩阵 R 和平移向量 t ，使得

$$E(R, t) = \arg \min_{R, t} \frac{1}{n} \sum_{i=1}^n \|x_i - Rp_i - t\|^2$$

基本思路

- 如果知道正确的点对应，那么两个点集之间的相对旋转和平移有闭合解



求解最优变换

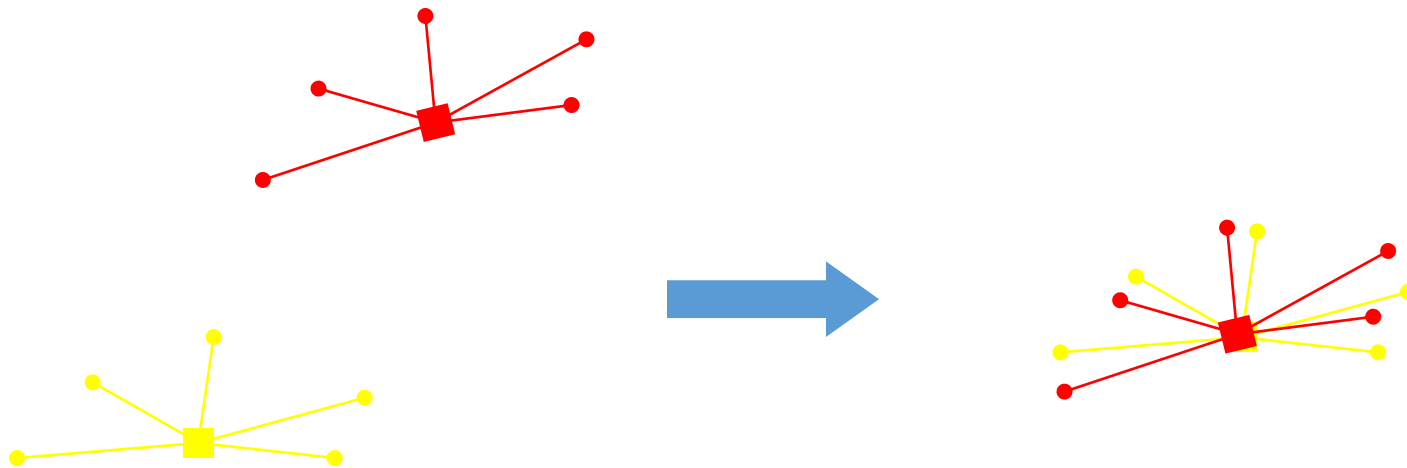
- 计算两个点集 X, P 的质心

$$\mu_x = \frac{1}{n} \sum_{i=1}^n x_i \quad \mu_p = \frac{1}{n} \sum_{i=1}^n p_i$$

- 从两个点集中分别减去对应的质心

$$X' = \{x_i - \mu_x\} = \{x'_i\} \quad P' = \{p_i - \mu_p\} = \{p'_i\}$$

求解最优变换



求解最优变换

- 利用SVD分解求最优变换

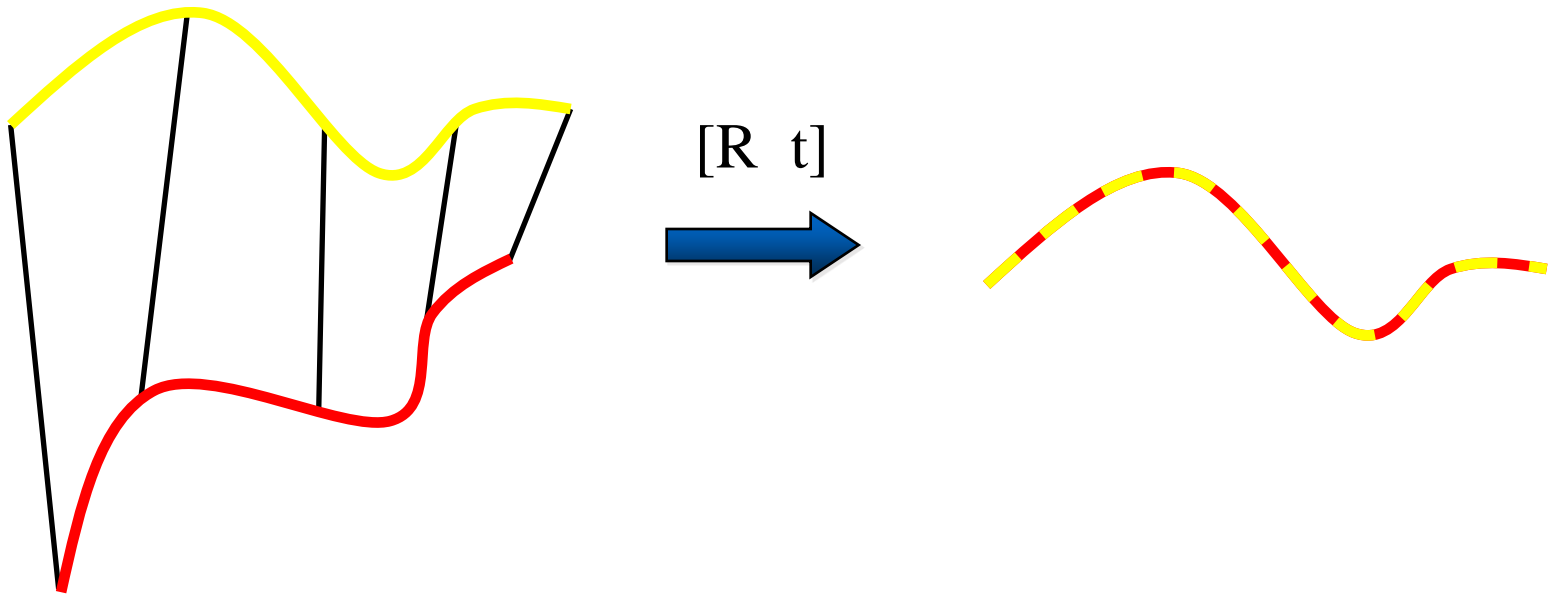
$$W = \sum_{i=1}^n x_i' p_i'^T \quad \longrightarrow \quad W = U \begin{bmatrix} \sigma_1 & 0 & 0 \\ 0 & \sigma_2 & 0 \\ 0 & 0 & \sigma_3 \end{bmatrix} V$$

- 如果 $\text{rank}(W) = 3$, $E(R, t)$ 的最优解是唯一的

$$W = UV^T \quad t = \mu_x - R\mu_p$$

三维数据点对应

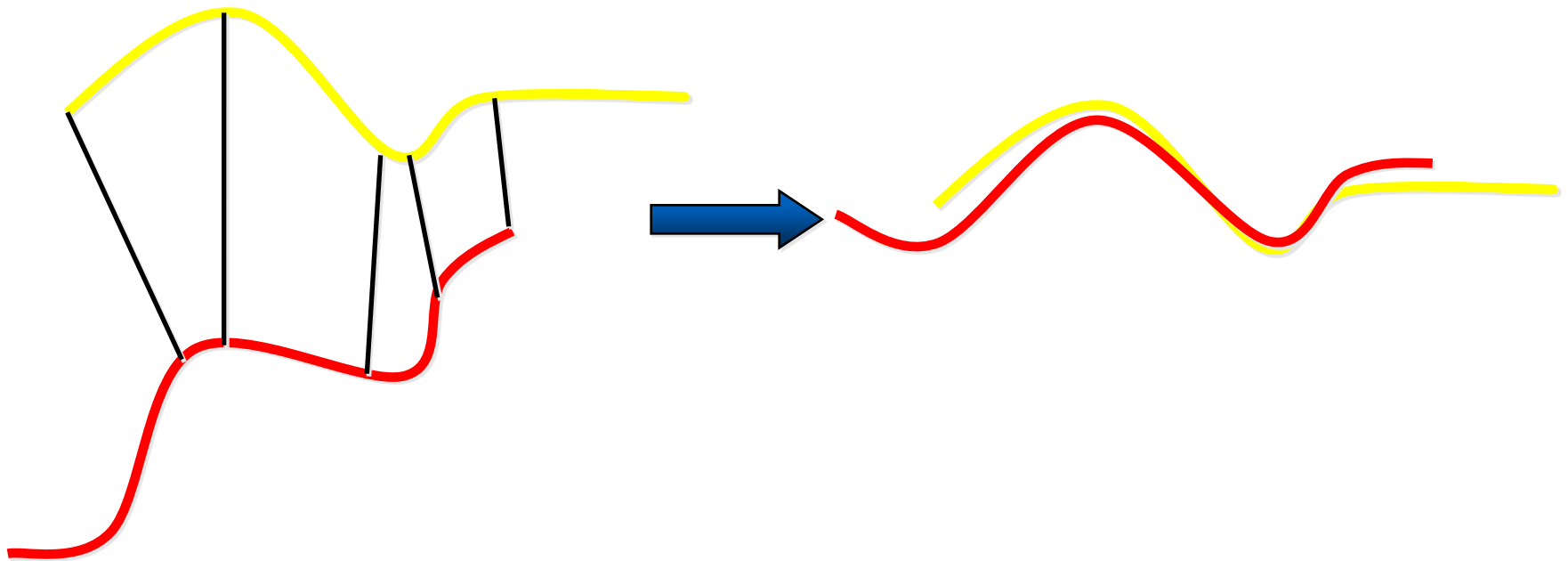
- 如果知道正确的点对应，那么两个点集之间的相对旋转和平移有闭合解



三维数据点对应

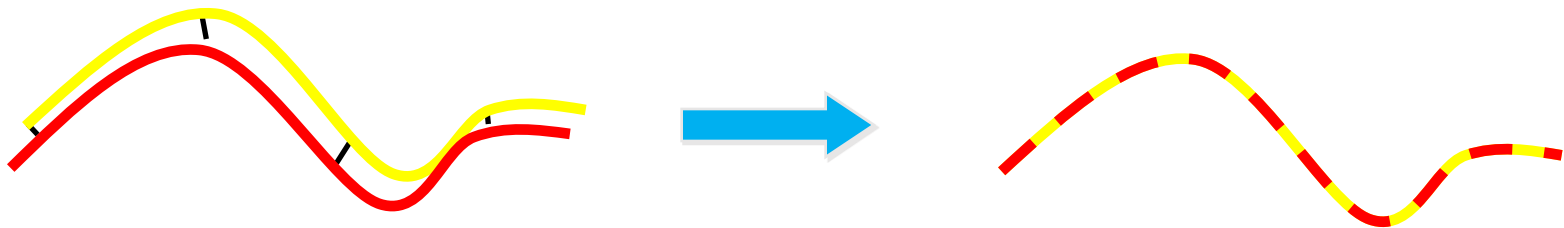
- 如何寻找对应点？
 - 用户输入？
 - 特征检测？
- 在ICP中，假设最近的点为对应点

三维数据点对应

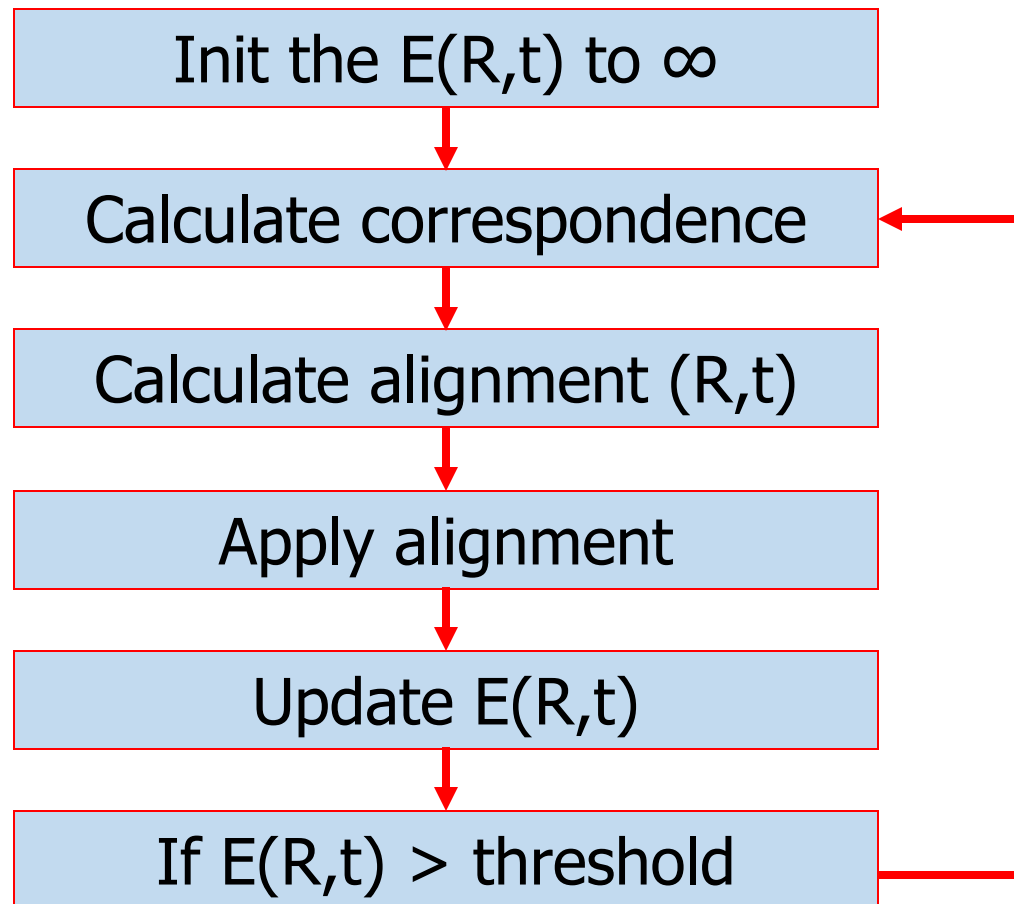


三维数据点对应

- Iterated Closest Points (ICP) [Besl & McKay 92]
- 如果初始点“足够的近”，可以保证收敛性



ICP算法



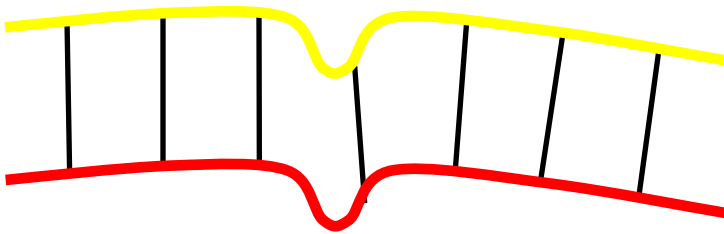
ICP算法

- 点集选取 (selecting)
- 点集匹配 (matching)
- 点集对应权重 (weighting)

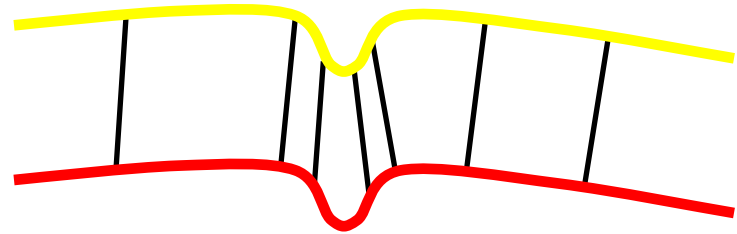
点集选取

- 选择所有点
- 均匀采样 (Uniform Sampling)
- 随机采样 (Random Sampling)
- 法方向空间均匀采样 (Normal-space Uniform Sampling)

点集选取



Uniform Sampling



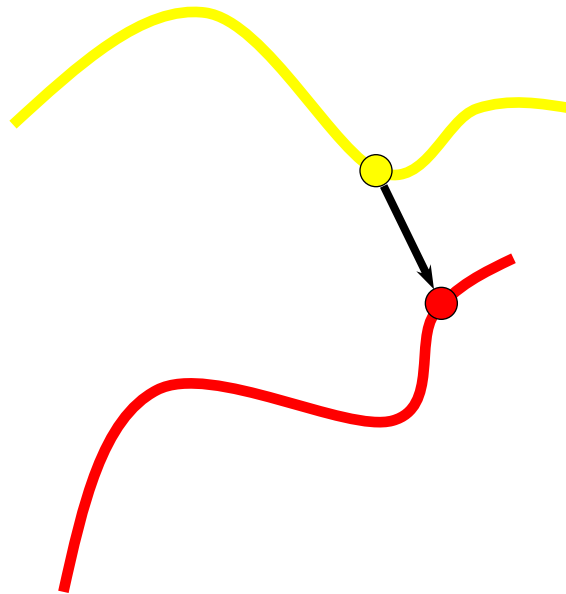
Normal-Space Uniform Sampling

点集匹配

- 最近邻点 (Closest point)
- 法方向最近邻点 (Normal shooting)
- 投影法 (Projection)

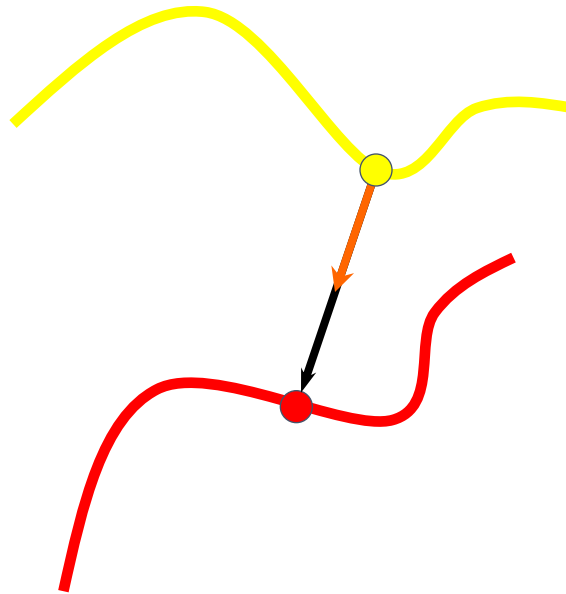
最近邻点

- 稳定，但速度慢



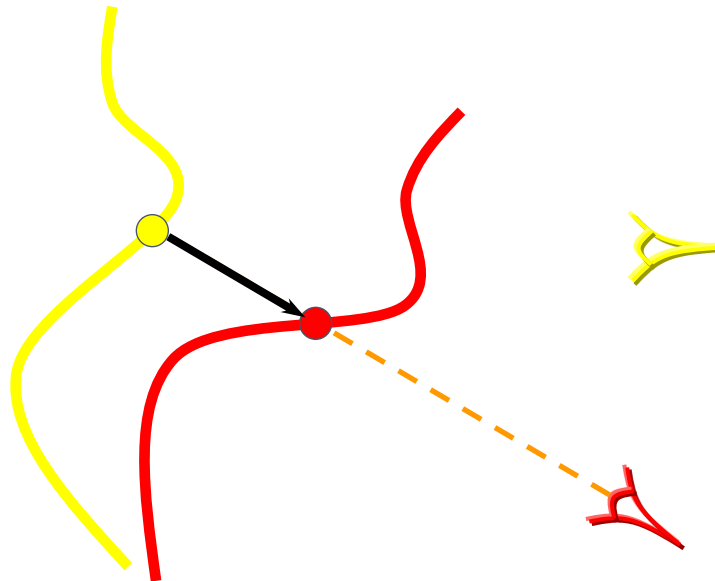
法方向最近邻点

- 平滑曲面效果好，但对噪音敏感



投影法

- 搜索对应点速度快



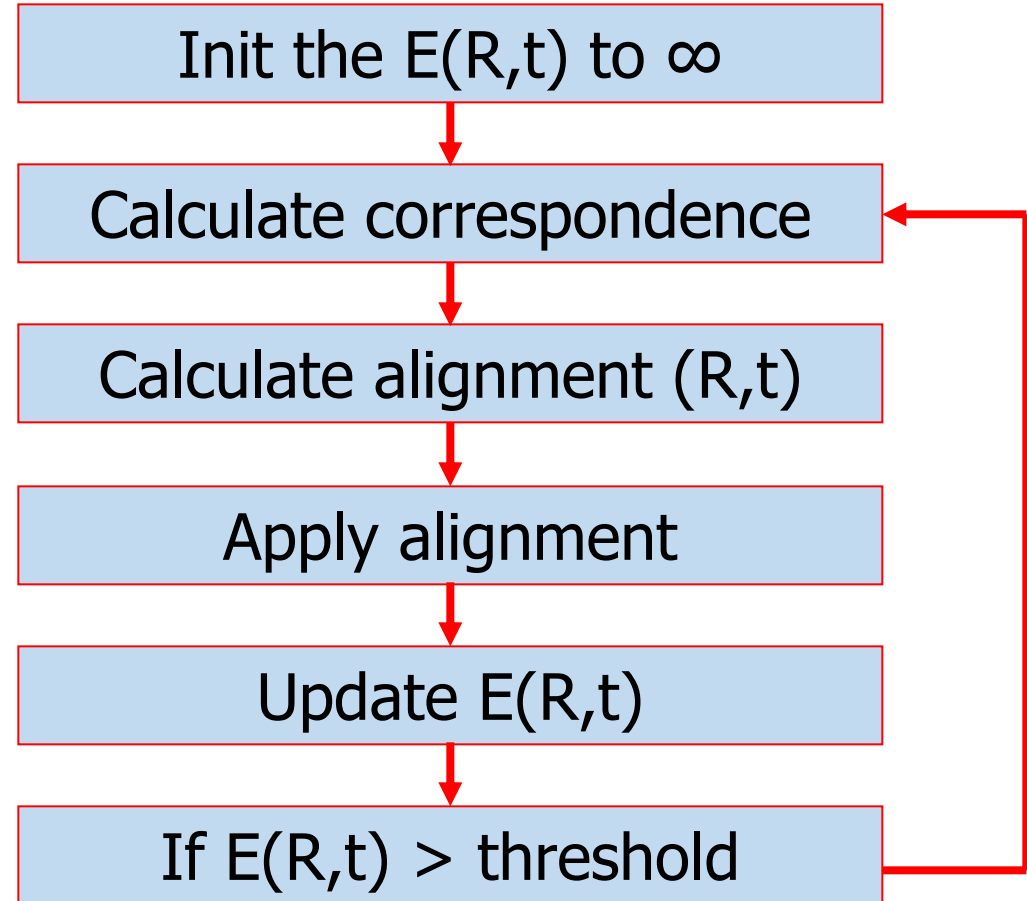
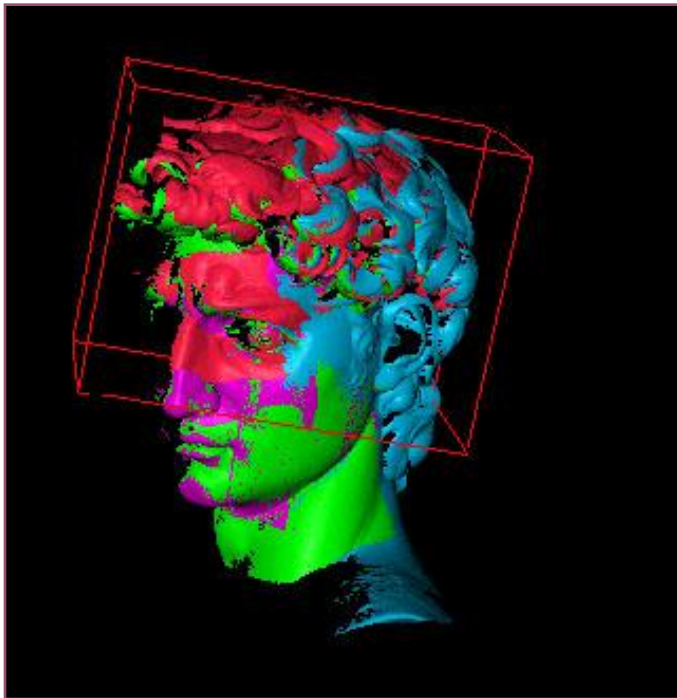
点集对应权重

$$E(R, t) = \arg \min_{R, t} \frac{1}{n} \sum_{i=1}^n ||x_i - Rp_i - t||^2$$

$$E(R, t) = \arg \min_{R, t} \frac{1}{n} \sum_{i=1}^n weight_i ||x_i - Rp_i - t||^2$$

- 固定权重 $weight_i = 1$
- 距离权重 $weight_i = 1 - \frac{Dist(x_i - p_i)}{Dist_{max}}$
- 法方向权重 $weight_i = n_{x_i} n_{p_i}$

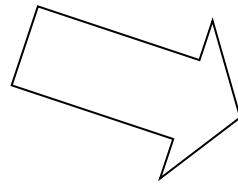
ICP小节



鲁棒匹配的RANSAC框架

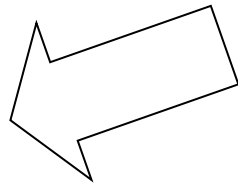
鲁棒匹配问题特点

大量外点存在

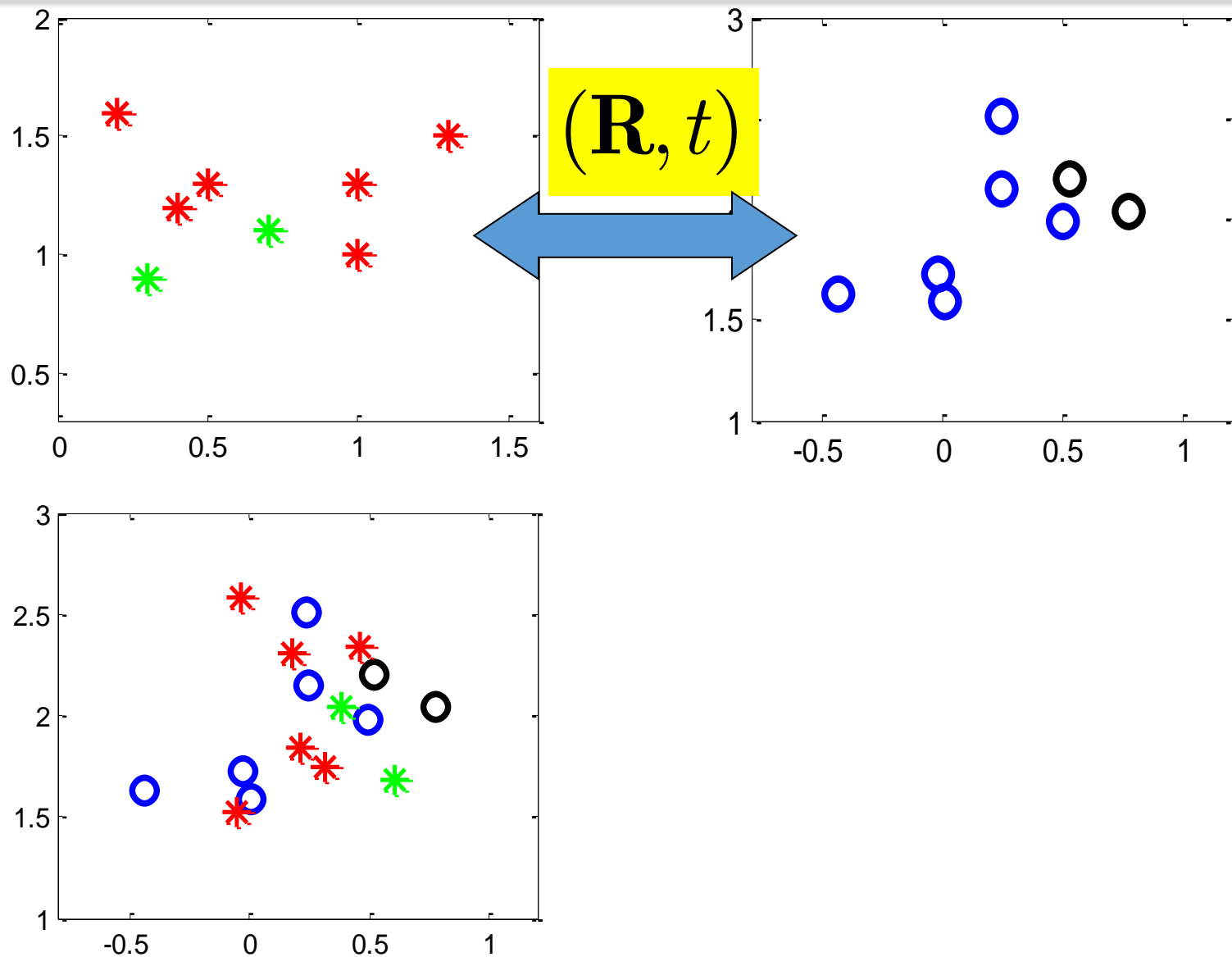


直接拟合模型

结果存在严重偏差



示例问题



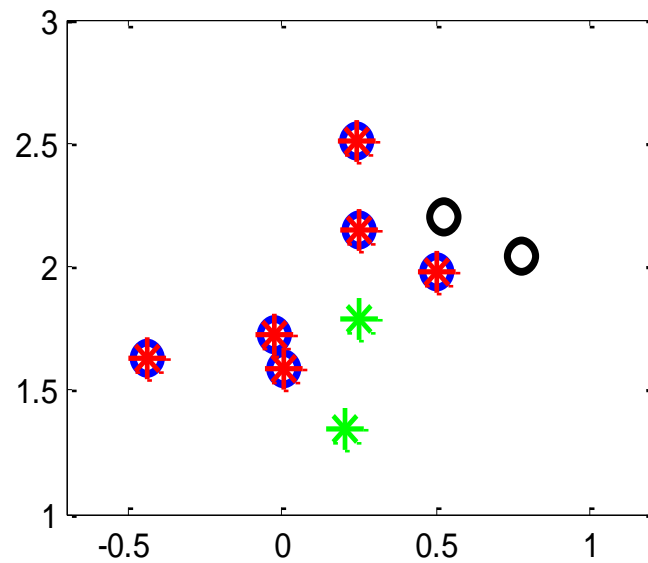
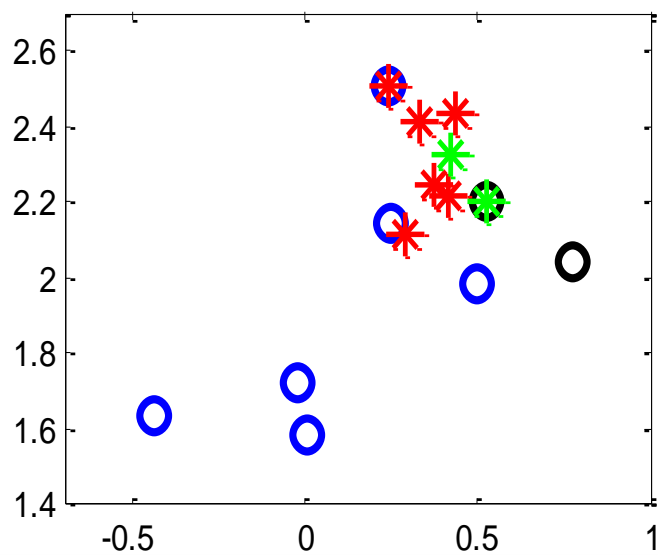
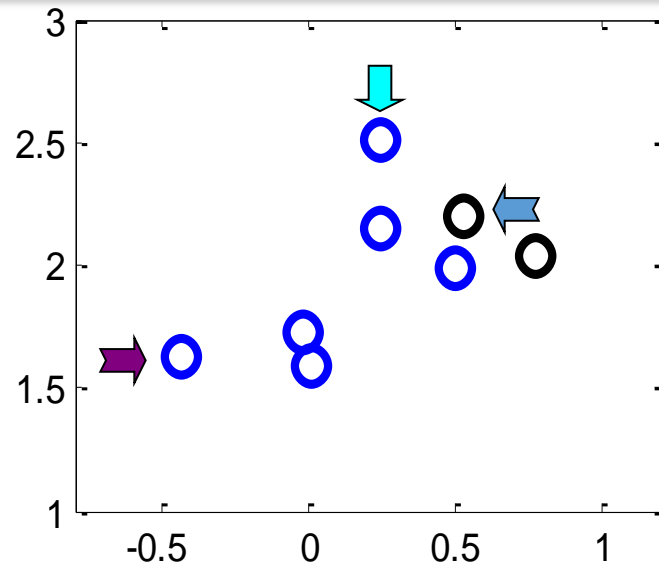
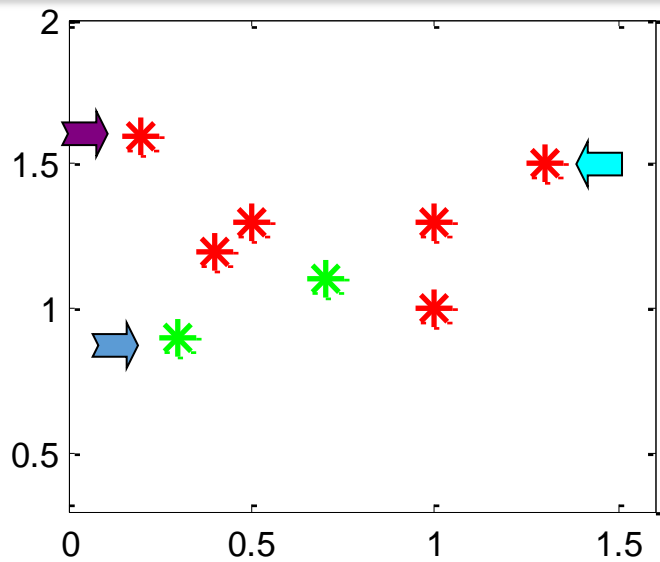
问题如何解决？

- 需要一种算法，能排除外点干扰
- 用排除外点的数据，重新拟合模型，得到最终结果

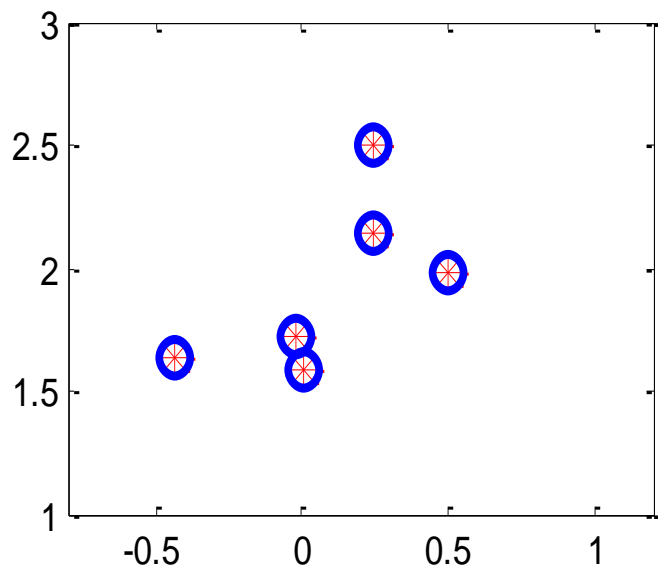
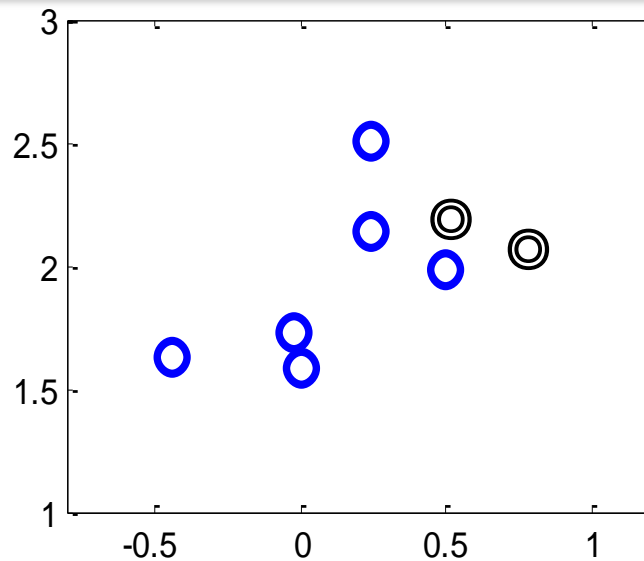
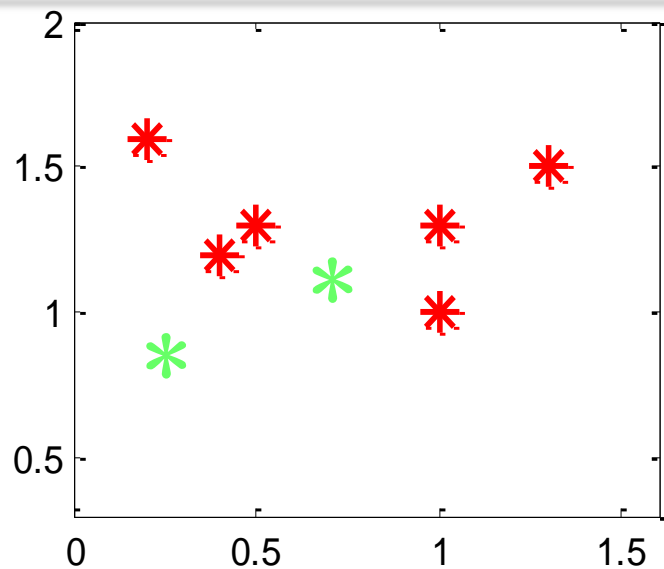
RANSAC估计模型M的一般步骤

- 确定求解模型M，所需要的最小数据点的个数 n 。由 n 个数据点组成的子集称为模型M的一个样本；
- 从数据点集D中随机地抽取一个样本J，由该样本计算模型的一个实例 $M_p(J)$ ，确定与 $M_p(J)$ 之间几何距离 $<$ 阈值 t 的数据点所构成的集合，并记为 $S(M_p(J))$ ，称为实例 $M_p(J)$ 的一致集；
- 如果在一致集 $S(M_p(J))$ 中数据点的个数 $\# S(M_p(J)) >$ 阈值T，则用 $S(M_p(J))$ 重新估计模型M，并输出结果；如果 $\# S(M_p(J)) <$ 阈值T，返回到步骤2；
- 经过K次随机抽样，选择最大的一致集 $S(M_p(J))$ ，用 $S(M_p(J))$ 重新估计模型M，并输出结果。

示例问题



示例问题



课后练习与参考文献

课后练习：

1. 了解ORB算法细节，并试使用C或C++实现完整的ORB特征提取算法。
2. 试使用C或C++实现完整的SIFT特征提取算法。

参考文献：

1. David G. Lowe, Distinctive Image Features from Scale-Invariant Keypoints, International Journal of Computer Vision, Vol. 60, Page 91-110, 2004.
2. Michael Calonder, Vincent Lepetit, Christoph Strecha, Pascal Fua, BRIEF: Binary Robust Independent Elementary Features, ECCV2010.
3. Tian Y. et al., L2-Net: Deep Learning of Discriminative Patch Descriptor in Euclidean Space, CVPR 2017.

小节

- SIFT特征
- ORB→BRIEF
- 基于DNN的特征描述子学习
- 空间点匹配：ICP
- 鲁棒估计：RNASAC

谢谢