

第十章 视频分析

-- 目标检测与目标跟踪

1. 概论
2. 目标检测与跟踪技术的典型应用
3. 目标检测方法的基本概念与原理
4. 目标跟踪方法涉及的基本问题
5. 基本模板匹配的跟踪算法

目标检测跟踪的含义

是近年来计算机视觉领域中备受关注的前沿方向，包含运动目标的图像序列中检测、识别、跟踪目标，并对其行为进行理解和描述。

目标分析的研究涉及到模式识别、图像处理、计算机视觉、人工智能等学科的内容。

随着现代信息处理技术的发展，目标检测跟踪在军用、民用及医学等方面具有重要意义，有广泛的应用前景和潜在的经济价值，已成为一项极为重要和基本的技术。然而，由于背景的多样性和复杂性，在工程应用中仍有许多问题有待解决。

目标检测和跟踪的困难点

- **形态变化** - 姿态变化是目标跟踪中常见的干扰问题。运动目标发生姿态变化时, 会导致它的**特征以及外观模型**发生改变, 容易导致跟踪失败。
- **尺度变化** - 尺度的自适应也是目标跟踪中的关键问题。当目标尺度缩小时, 由于**跟踪框不能自适应跟踪**, 会将很多背景信息包含在内, 导致目标模型的更新错误: 当目标尺度增大时, 由于跟踪框不能将目标完全包括在内, 跟踪框内目标信息不全, 也会导致目标模型的更新错误。因此, 实现尺度自适应跟踪是十分必要的。
- **遮挡与消失** - 目标在运动过程中可能出现被遮挡或者短暂的消失情况。当这种情况发生时, 跟踪框容易将**遮挡物以及背景信息包含在跟踪框**内, 若目标被完全遮挡时, 由于找不到目标的对应模型, 会导致跟踪失败。
- **图像模糊** - 光照强度变化, 目标快速运动, 低分辨率等情况会导致图像模型, 尤其是在运动目标与背景相似的情况下更为明显。因此, 选择有效的特征对目标和背景进行区分非常必要。



deformation



illumination variation



blur & fast motion



知乎 @王方浩
background clutter



out-of-plane rotation



in-plane rotation



scale variation



5



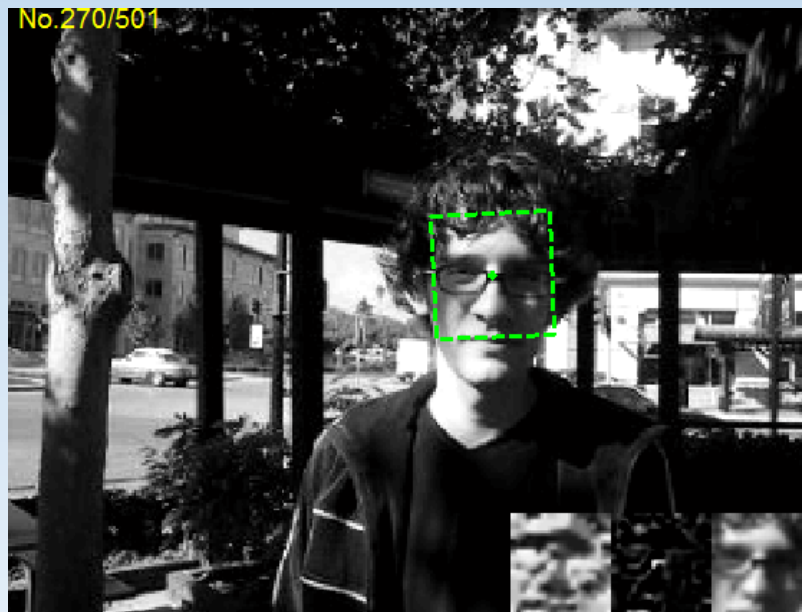
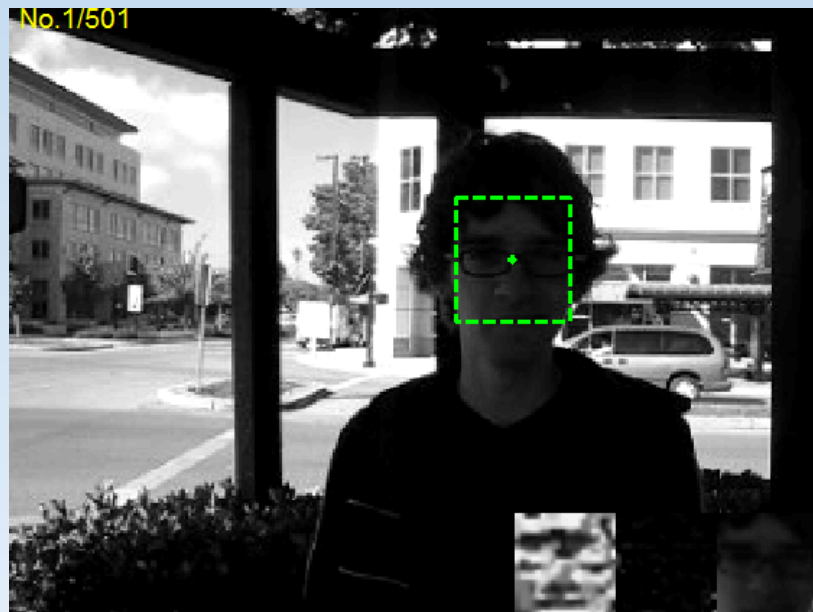
occlusion



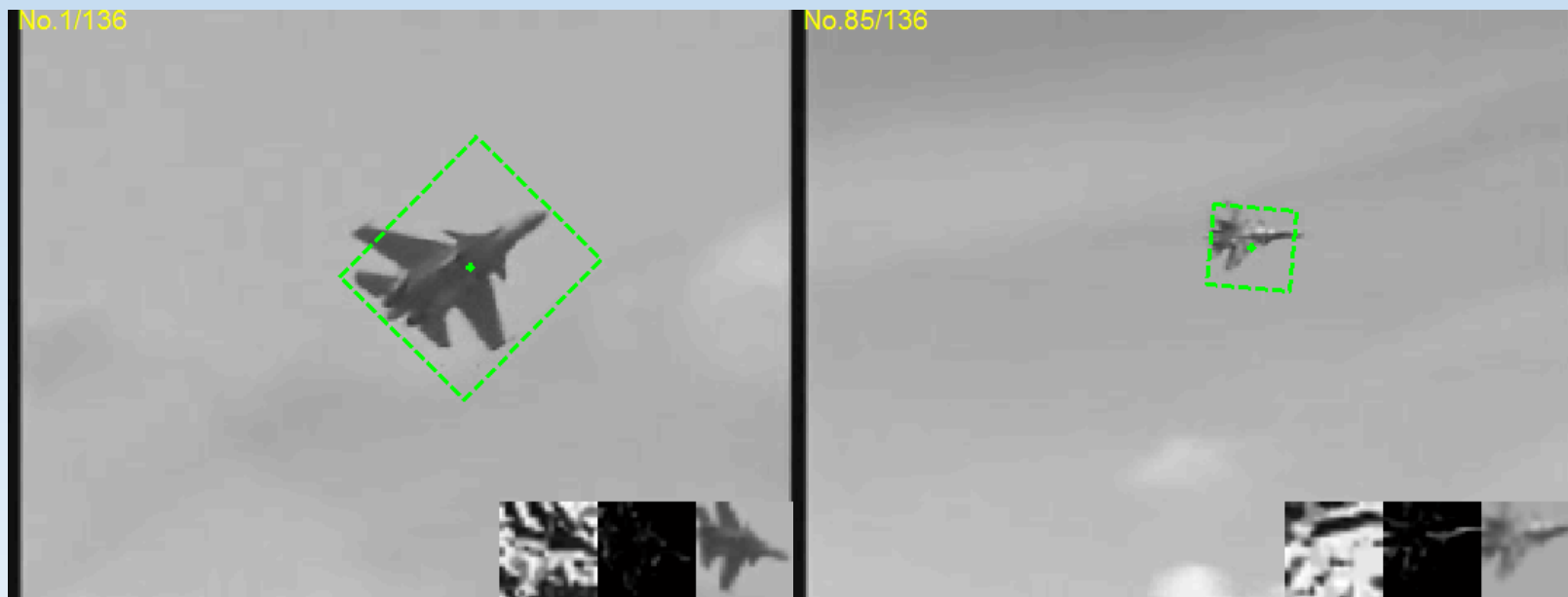
知乎 @王方浩
out-of-view

https://blog.csdn.net/weixin_42407742

光照变化



尺度变化



姿态变化



应用：

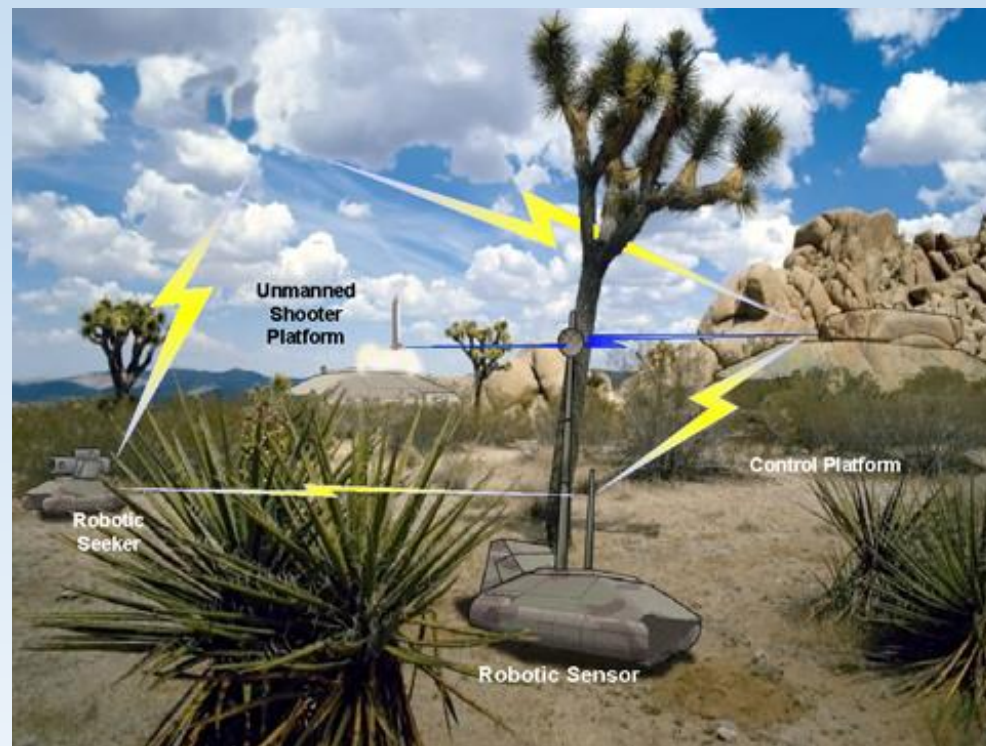
安防系统： 可以实现停车场监视、车辆识别、车牌号识别、探测并跟踪“可疑”目标；正确判断出是否有目标进入警戒区，识别出进入警戒区的目标是否为报警对象，有效跟踪可疑对象，根据面孔、眼底、指纹等特征识别特定人；监测固定区域的人流量等都是安防系统的重要功能。

智能交通： 车牌识别、违章超速、并线、掉头、流量控制与交通疏导、禁行区检测。

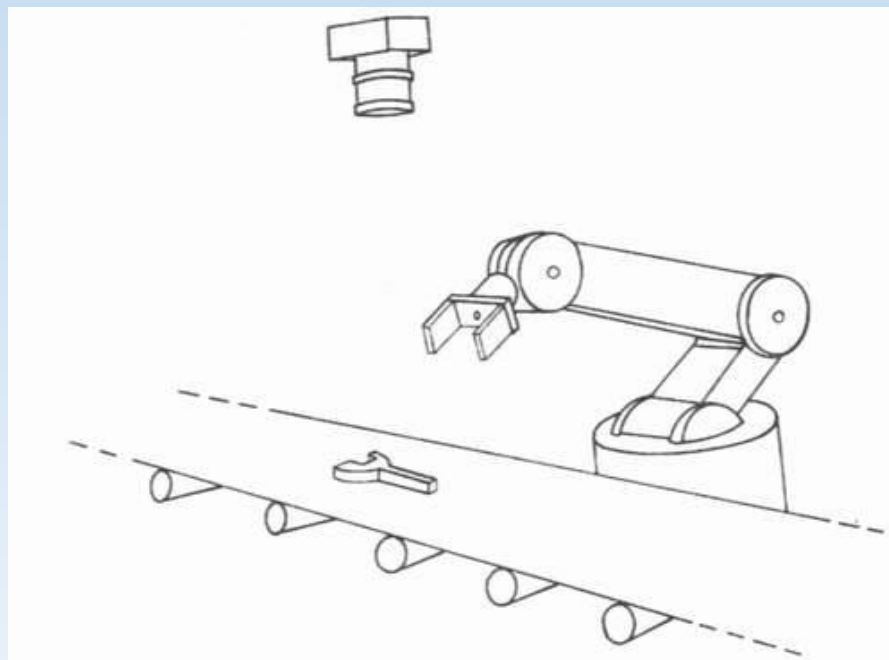
医学图像分析与诊断： 成像方法包括传统的X射线成像、计算机层析(CT)成像、核磁共振成像(MRI)、超声成像等。在医学图像诊断方面有两个应用：一是对图像进行增强、标记、染色等处理来帮助医生诊断疾病，并协助医生对感兴趣的区域进行定量测量和比较；二是利用专家知识系统对图像进行自动分析和解释，检测并识别出可疑区域，给出诊断结果。

产品检验： 在工业领域，目前已经用于产品外形检验、表面缺陷检验，比如，滑块及滑槽的外形检验以及装配后的位置检验，以决定它们能否装配在一起，并且准确无误地完成装配任务；发动机内壁麻点、刻痕和裂缝等缺陷检查以决定产品的质量。通过X射线照相或超声探测获取物体内部的图像，可以实现内部缺陷检验，如钢梁内部裂纹和气孔等缺陷检验。

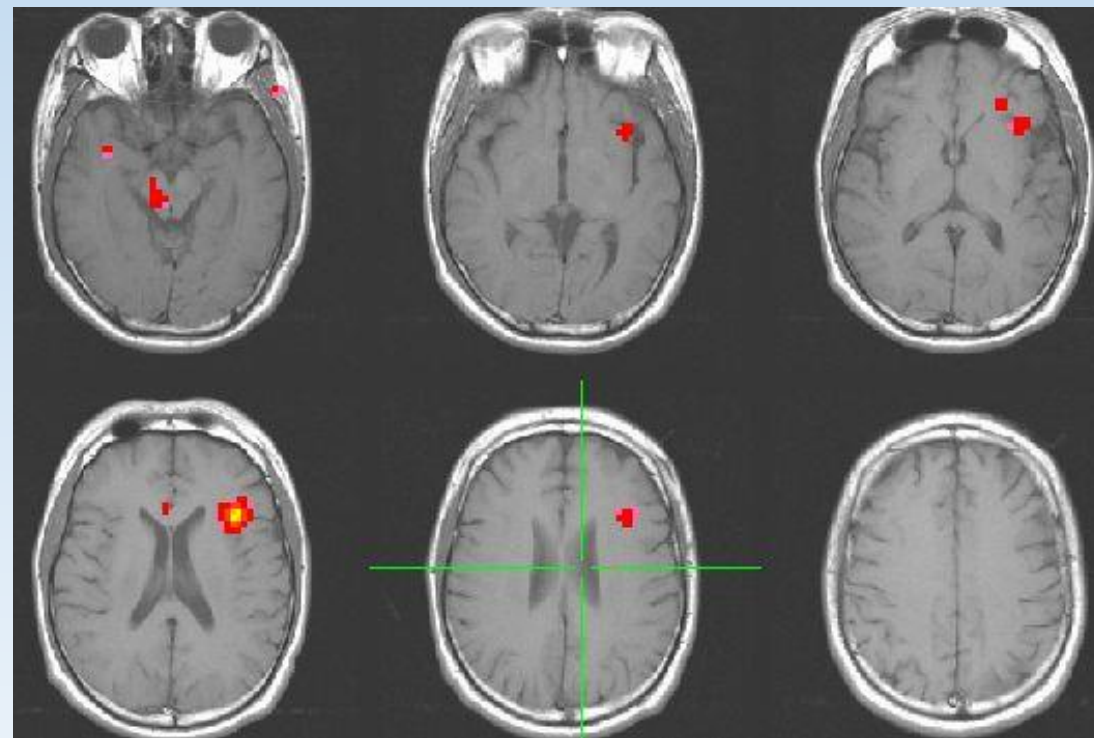
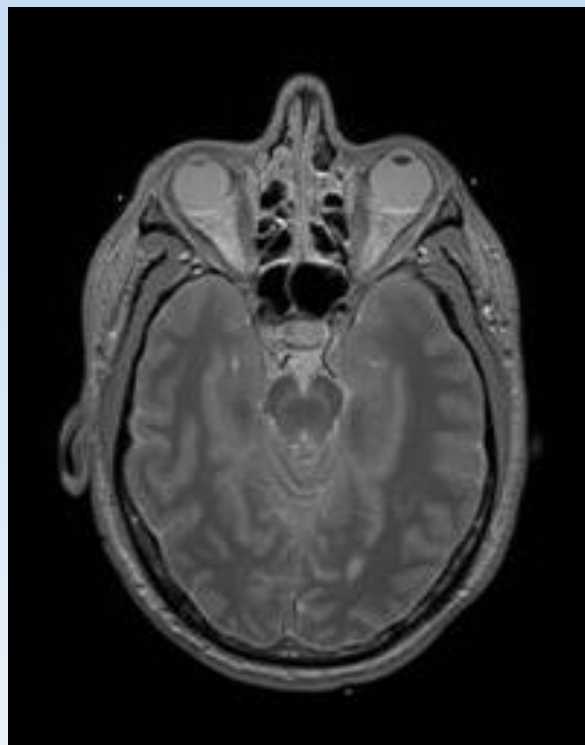
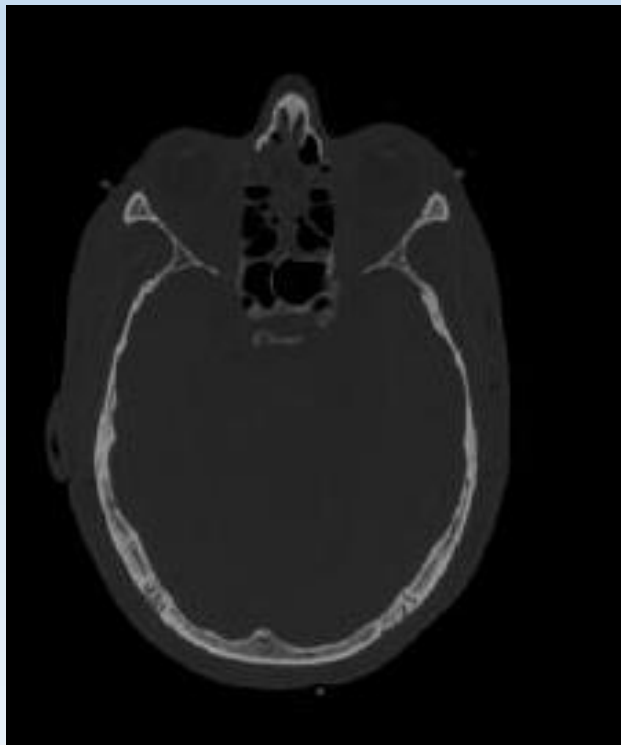
在军事应用示例-无人驾驶小车



生产线上具有简单视觉系统的工业机器人系统



可视化人体CT MRI FMRI 图像



1. 概论
2. 目标检测与跟踪技术的典型应用
3. 目标检测方法的基本概念与原理
4. 目标跟踪方法涉及的基本问题
5. 基本模板匹配的跟踪算法

目标检测方法的基本概念与原理

目标检测的分类：

目标检测从目标特性角度分为单幅**静止图像**检测和**运动目标**检测；

- 静止目标检测通常是利用**单帧图像信息**，对于大目标，可以利用**图像分割**或**特征匹配**等方法提取出目标，但对于低对比度、低信噪比的小目标，利用单帧信息很难检测出有效目标。
- 运动目标检测又分为**静止背景**下的运动目标检测和**运动背景**下的运动目标检测。运动目标可以利用图像的**运动序列信息**，与单幅图像不同，连续采集的图像序列能反映场景中目标的运动和场景的变化情况，更有利于小目标的探测。

目标检测的基本思想：

用摄像机采集的图像可以认为是由三个分量组成的：目标图像、背景图像和噪声图像。

目的：

抑制背景

消除噪声

突出目标



抑制背景和消除噪声都是为了更好地**突出目标**服务。

如何突出目标：

灰度奇异性特征、几何形状特征、运动特征、频谱特征等。

不同情况采用不同的方法：

- 对于目标与背景亮度差异较大、且背景比较简单的情况，通常会采用利用阈值分割的方法来提取目标；
- 对于复杂背景下的目标，需要利用目标的几何形状特征或运动特征等来提取目标。静止背景下常用的方法有背景差分法、帧间差分法等，运动背景下常用的有背景配准法、光流法、运动估计法等；
- 低信噪比下小目标检测是难点也是重点，通常采用滤波的方法。常用的有三维匹配滤波、动态规划法、多级假设检验等。

常用的目标检测方法：

第一类:是**基于像素分析**的方法，主要有基于图像分割的方法、帧间差分方法、相关算法、光流法等；

第二类:是**基于特征匹配**的方法，主要利用的特征有：角点、直边缘、曲边缘等局部特征和形心、表面积、周长、投影特征等全局特征；

第三类:是**基于频域**的方法，较典型的是基于傅立叶变换和基于小波变换的方法。

第四类:是**基于识别的检测**方法，较典型的是基于边缘碎片模型的目标检测识别方法；

本节主要介绍：

利用图像分割技术、图像匹配技术的目标检测

基于图像分割技术的目标检测方法

图像分割的目的：

把图像分解成构成它的部件和对象；

有选择性地定位感兴趣对象在图像中的位置和范围。

图像分割的基本思路：

从简到难，逐级分割

控制背景环境，降低分割难度

注意力集中在感兴趣的对象，缩小不相干图像成分的干扰。

图像分割的基本方法：

基于阈值的图像分割

基于形态学的图像分割

基于边缘的图像分割

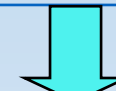
基于区域的图像分割



提取轮廓



车牌定位



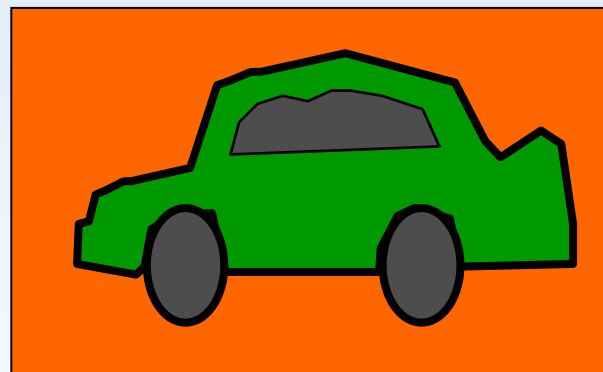
车牌识别

图像分割的基本策略：

把像素按灰度划分到各个物体对应的区域中去；

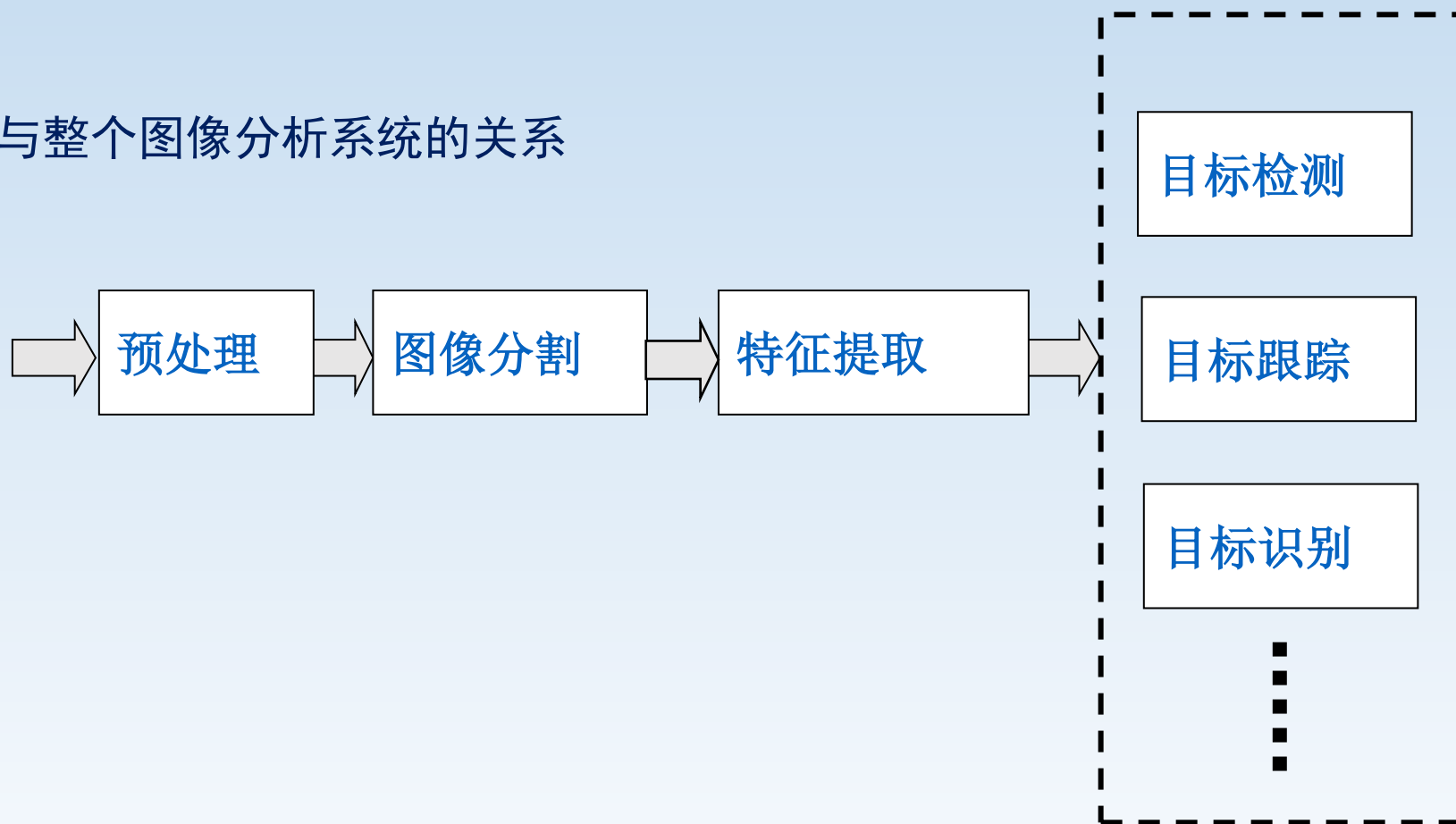
先确定边缘像素点，然后将它们连接起来构成所需的边界；

确定存在于区域间的边界；



图像分割：将数字图像划分成与实际目标或区域紧密相关的若干区域的过程。

图像分割与整个图像分析系统的关系



基于图像匹配技术的目标检测算法

基于图像灰度的匹配算法

- 绝对平衡搜索法
- 归一化互相关匹配
- 图像匹配的加速算法

基于图像特征的匹配算法

- 目标像素数:符合目标灰度分布的像素点总数;
- 目标均值:目标像素点的灰度均值;
- 复杂度:边界像素点数与总目标像素点数的比值;
- 长宽比:目标像素数最多一行与像素数最多一列比值;
- 紧凑度:目标像素数与目标长、宽之比值。
- 点、线等几何形状特征
- 不变矩特征
- 边缘特征
- 幅度、直方图、频率系数
- 基于相似性判据最优化的方法

基于核密度估计的mean shift方法

基于局部不变特征的匹配方法（SIFT、SURF等）

基于神经网络的方法

什么是图像匹配？

也叫**图像配准**，是图像处理的基本任务之一，用于将不同时间、不同传感器、不同视角及不同拍摄条件下获取的**两幅或多幅**图像进行匹配，在军事、遥感、医学、计算机视觉等领域有广泛的应用。

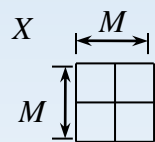
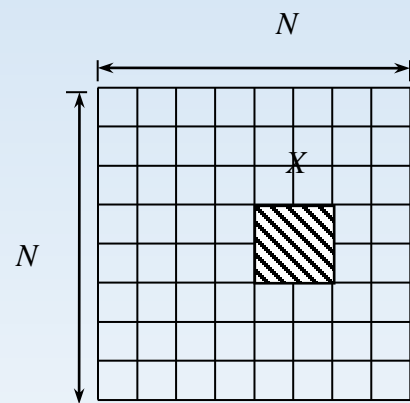
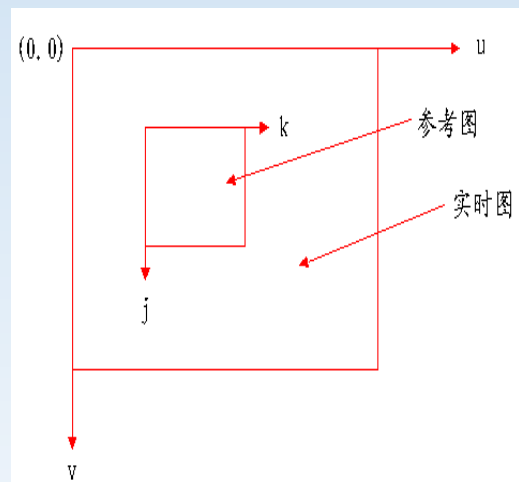
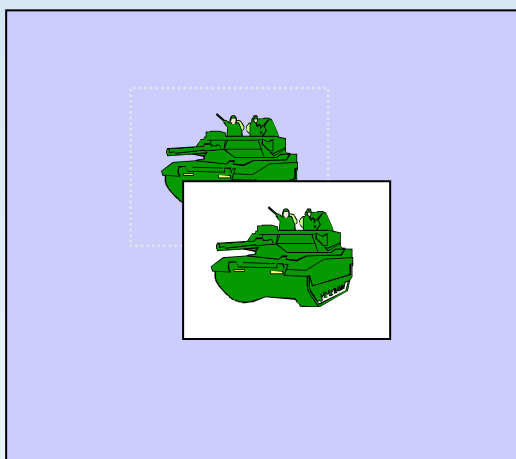
图像配准技术可分为**手工配准**和**自动配准**两种方式。自动配准是在整个配准过程中不需要人干预的一种配准技术，它也是图像配准技术的最终发展目标。

在配准窗口范围内，同一目标的两幅图像完全相同，完成图像配准并不困难。

实际中，同一目标的两幅图像常常是在不同条件下获得的（如不同的成像时间、不同的成像位置，甚至不同的成像系统等），再加上各种噪声的影响，使同一目标的两幅图像不可能完全相同，只能做到某种程度的相似，因此图像配准是一个相当复杂的技术过程。

配准的方法：

图像配准的一般做法是，首先在参考图上选取以某一目标点为中心的图像子块，并称它为图像配准的目标窗口，然后让目标窗口在待配准图上有秩序地移动，每移到一个位置，就把目标窗口与待配准图的对应部分进行相关比较，直到找到配准位置为止。



归一化直接互相关匹配结果：



原始图像(第0帧)



第10帧跟踪结果



第20帧跟踪结果

先边缘处理在归一化互相关匹配结果：



原始图像(第0帧)



第10帧跟踪结果



第20帧跟踪结果

运动目标的检测

静止背景下的运动目标检测方法

- 帧间差分法
- 背景差分法
- 运动能量积累法

动态背景下的运动目标检测方法

- 光流法
- 运动参数估计法：块匹配法、贝叶斯法、像素递归法
- 带全局运动补偿的算法

运动图像序列的分类：

运动图像序列可以为低信噪比情况下的目标检测提供更多的有用信息，利用序列图像能够检测出单帧图像中很难检测出的目标。由运动目标所形成的图像序列可分为两种情况：一种是静止背景，另一种是变化背景。

前一种情况通常发生在摄像机处于相对静止的状态下，产生背景静止的运动图像序列，这时可以采用帧间差分或自适应背景对消的方法来去除背景干扰，计算方法简单；

后一种情况发生在目标运动同时，摄像机也处于相对运动状态，产生背景变化的运动图像序列，从而导致运动目标的变化与背景本身的变化混淆在一起。这种情况比较复杂，若采用消除背景的方法，则需要进行图像配准；若采用突出目标的方法，则需要在配准的前提下进行多帧能量积累和噪声抑制。

帧间差分法

将前后两帧图像对应像素点的灰度值相减，在环境亮度变化不大的情况下，若对应像素灰度相差很小，可以认为此处景物是静止的，若图像区域某处的灰度变化很大，可以认为这是由于图像中运动物体引起的，将这些区域标记下来，利用这些标记的像素区域，就可以求出运动目标在图像中的位置。

第 j 帧 $f_{m,n,j}(x,y)$ 与第 k 帧 $f_{m,n,k}(x,y)$ 间的变化可用一个二值差分图像来表示：

$$DP_{jk} f_{m,n}(x,y) = \begin{cases} 1 & |f_{m,n,j}(x,y) - f_{m,n,k}(x,y)| > T \\ 0 & \text{其它} \end{cases}$$

帧间差分算法的优缺点：

优点：

- 对动态背景适应快
- 计算速度快
- 占用系统资源少
- 特别适用于噪声较少的红外视频

缺点：

- 对噪声较大的可见光视频效果不理想
- 该算法假定目标是不断运动的。如果物体静止一段时间，那么该算法将检测不到目标
- 如果背景的某一部分移动，那么将会在检测结果中出现“空洞”
- 该算法一般检测不到移动目标的全部

背景差分法

是采用图像序列中的当前帧和背景参考模型比较来检测运动物体的一种方法, 其性能依赖于所使用的背景建模技术。

在基于背景差分方法的运动目标检测中, 背景图像的建模和模拟的准确程度, 直接影响到检测的效果。不论任何运动目标检测算法, 都要尽可能的满足任何图像场景的处理要求, 但是由于场景的复杂性、不可预知性、以及各种环境干扰和噪声的存在, 如光照的突然变化、实际背景图像中有些物体的波动、摄像机的抖动、运动物体进出场景对原场景的影响等, 使得背景的建模和模拟变得比较困难。

背景差分算法的优缺点：

优点：

- 背景差分法检测运动目标速度快，检测准确，易于实现。
其关键是背景图像的获取

缺点：

- 在实际应用中，静止背景是不易直接获得的，同时，由于背景图像的动态变化，需要通过视频序列的帧间信息来估计和恢复背景，即背景重建，所以要选择性的更新背景。

帧间差分法和背景差分法的异同：

- **背景差分法**的原理是通过数学模型建立监控场景的背景图像，再将当前帧图像和背景图像做差分，得到差分图像，最后通过阈值化差分图像得到运动掩模。
- **帧间差分法**是一种通过对视频图像序列中相邻几帧（主要是两帧或三帧）作差分运算，并阈值化差分图像来获得运动目标轮廓的方法。
- **主要相同点：**都是运动目标的提取方法，都具有计算简单、速度快的优点，都可以通过对图像进行差分运算、阈值运算获取到运动掩模。
- **主要区别：**帧间差分法差分的两幅图像的时间间隔较短，受光线变化影响小，适合于动态变化场景；简单背景差分法适用于固定摄像机的室内监控场景，参考背景和实际背景的相似性直接关系到运动检测结果的正确与否。如果没有一个好的自适应的参考背景，容易受到光线变化等环境变化的影响。

静止背景下运动目标的检测实验结果：



(a) 第1帧图像



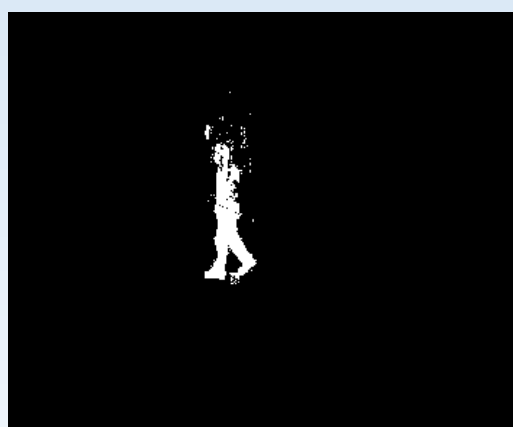
(b) 第2帧图像



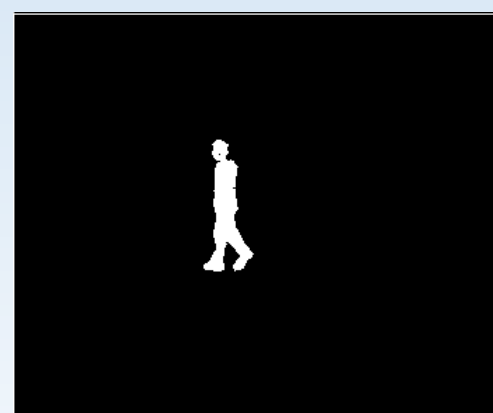
(c) 变化区域图像



(d) 提取出的背景图像



(e) 变化区域与背景差分图像

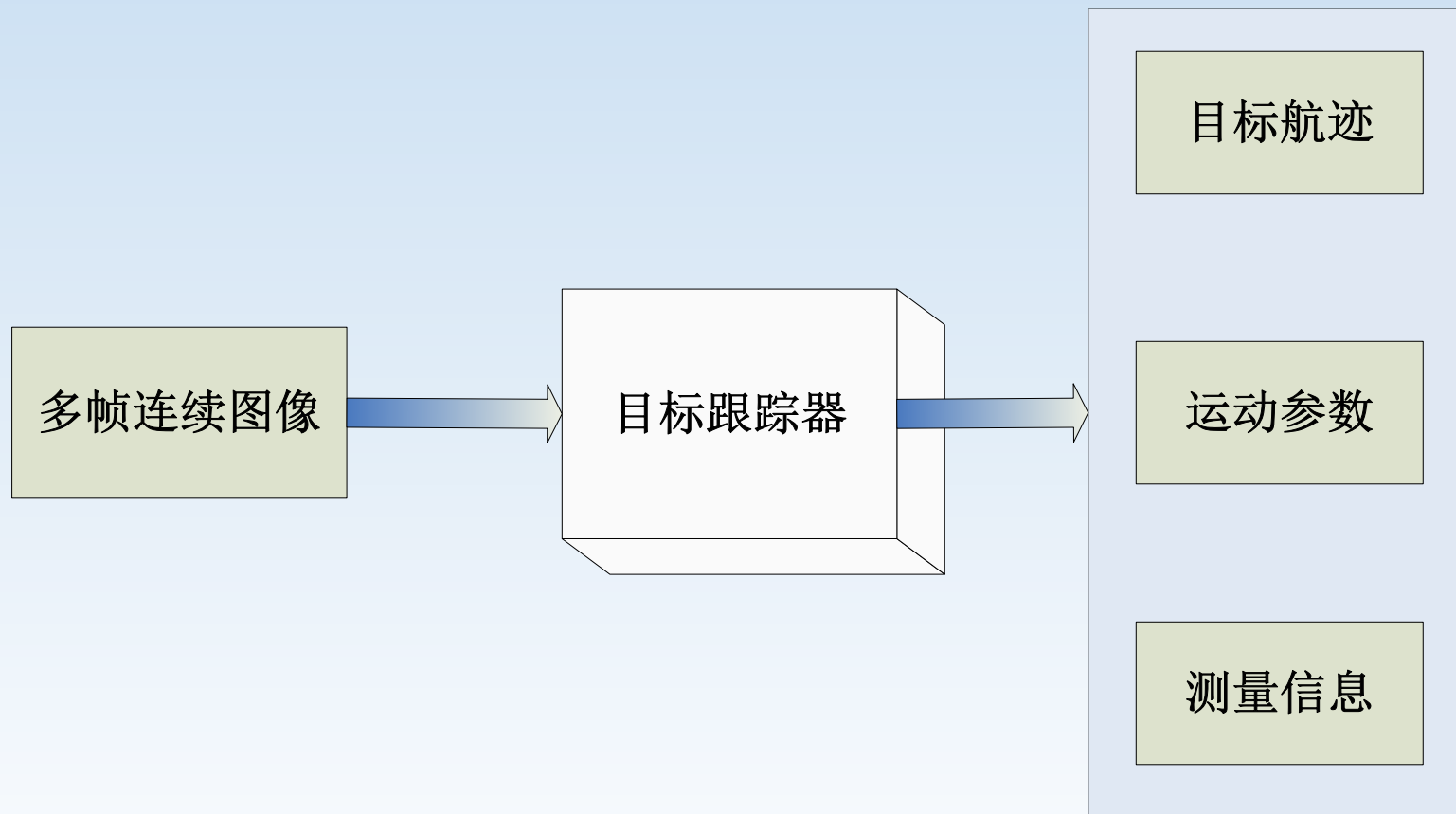


(f) 运动目标检测结果

1. 概论
2. 目标检测与跟踪技术的典型应用
3. 目标检测方法的基本概念与原理
- 4. 目标跟踪方法涉及的基本问题**
5. 基本模板匹配的跟踪算法

目标跟踪方法涉及的基本问题

定义：所谓目标跟踪，可以简单地定义为对连续的视频序列中的某个特定目标给出一条航迹，进而获得目标的**位置、速度等运动参数**，以及**形状、大小、颜色**等对后续目标分析与理解非常重要的**测量信息**。



应用：

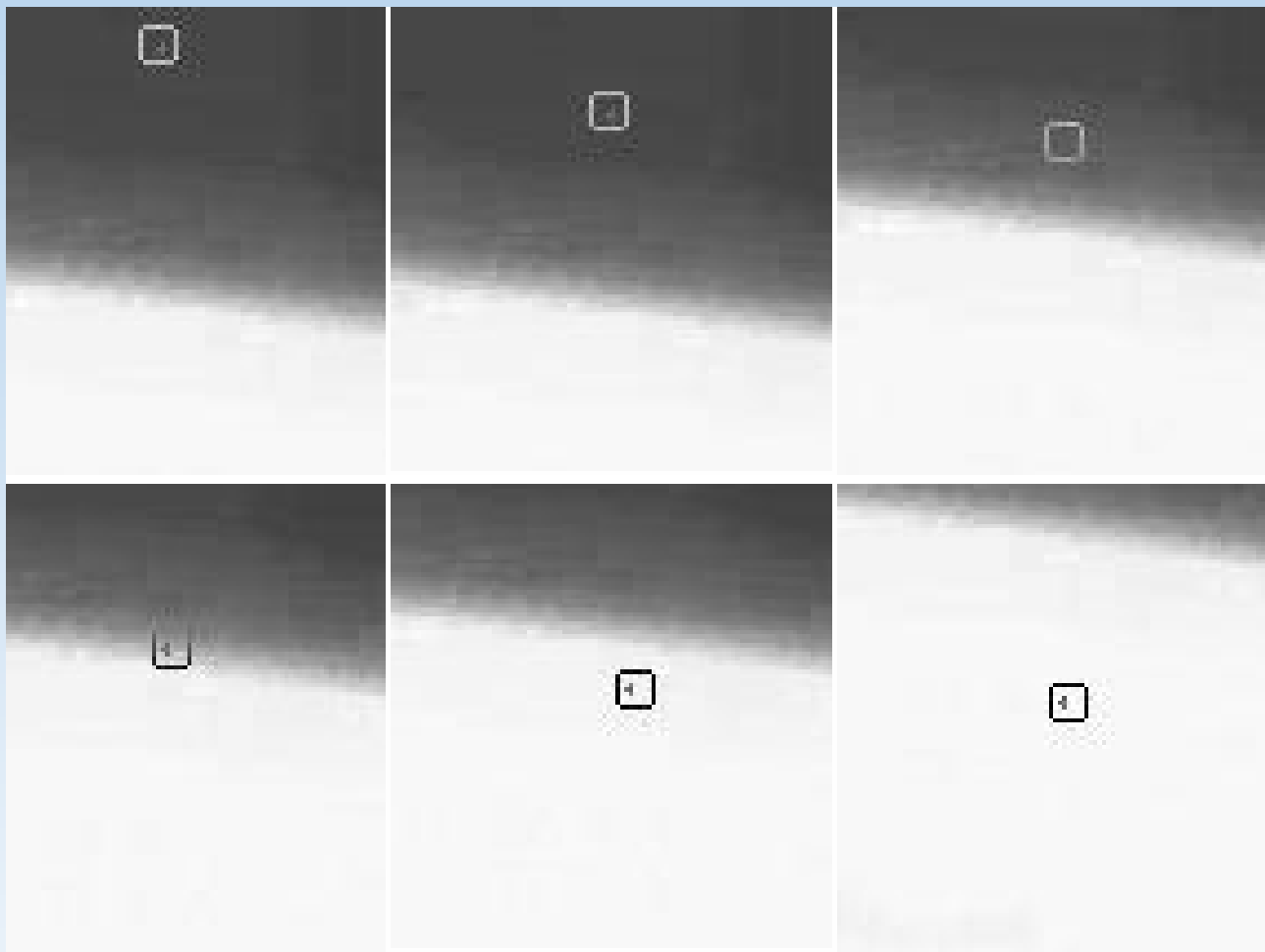
在工业过程控制、医学研究、交通监视、自动导航、天文观测等领域有重要的实用价值。尤其在军事上，目标跟踪技术已被成功地用于武器的成像制导、军事侦察和监视方面。

目的：通过对传感器拍摄到的图像序列进行分析，计算出目标在每帧图像上的位置，给出目标速度的估计。

指标：可靠性和精度。

远距离，目标面积较小、机动性不强，通常采用**滤波方法**跟踪目标以提高跟踪精度。

近距离，目标具有一定面积其帧间抖动较大时，一般采用**窗口质心跟踪**或**匹配跟踪方法**以保持跟踪的稳定性和精度。



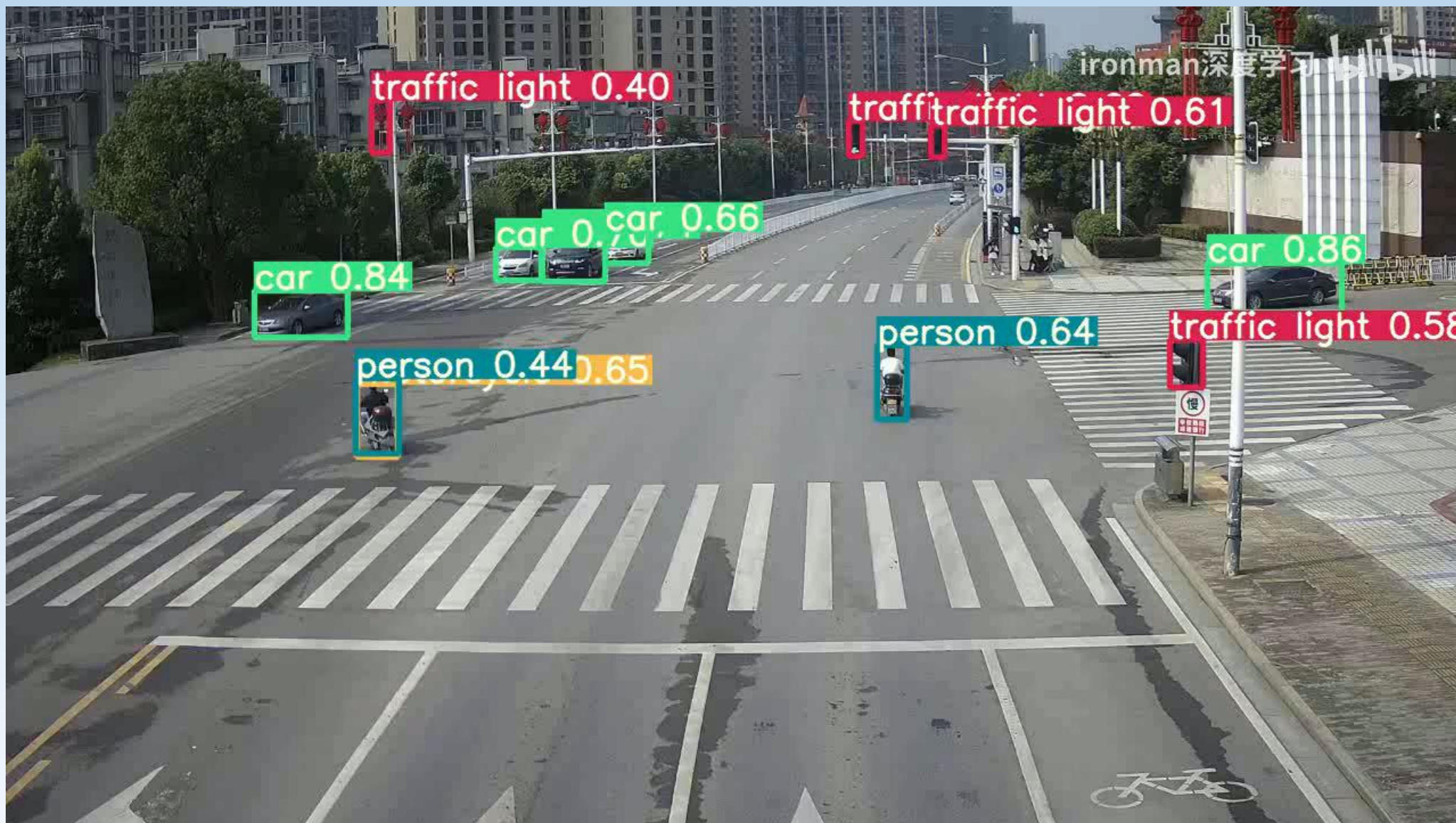
弱小点目标跟踪实例



人脸跟踪实例



F-22战斗机的目标跟踪系统



traffic light 0.40

traffic light 0.61

ironman 深度学习

car 0.79 car 0.66

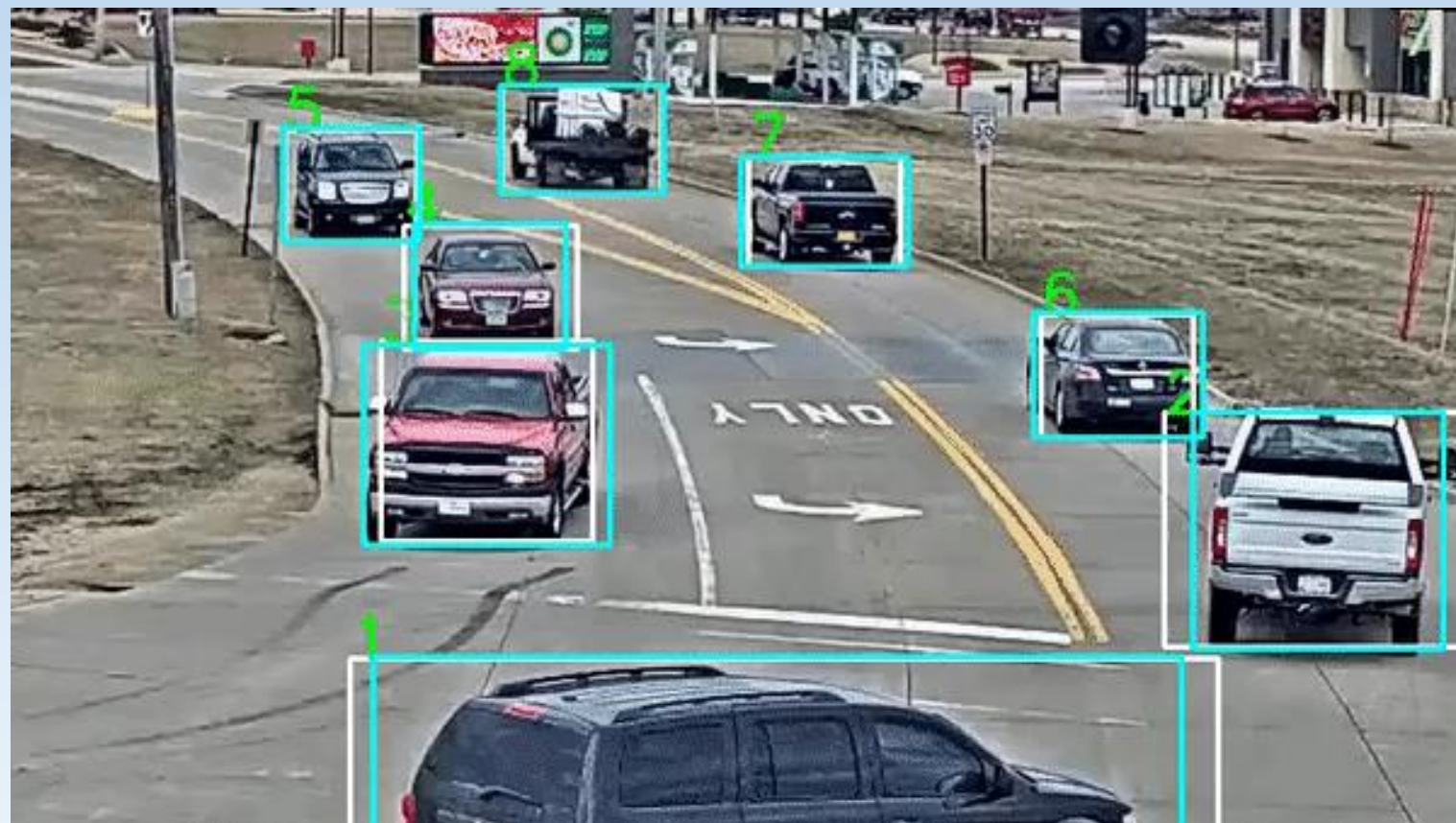
car 0.84

car 0.86

person 0.64

traffic light 0.58

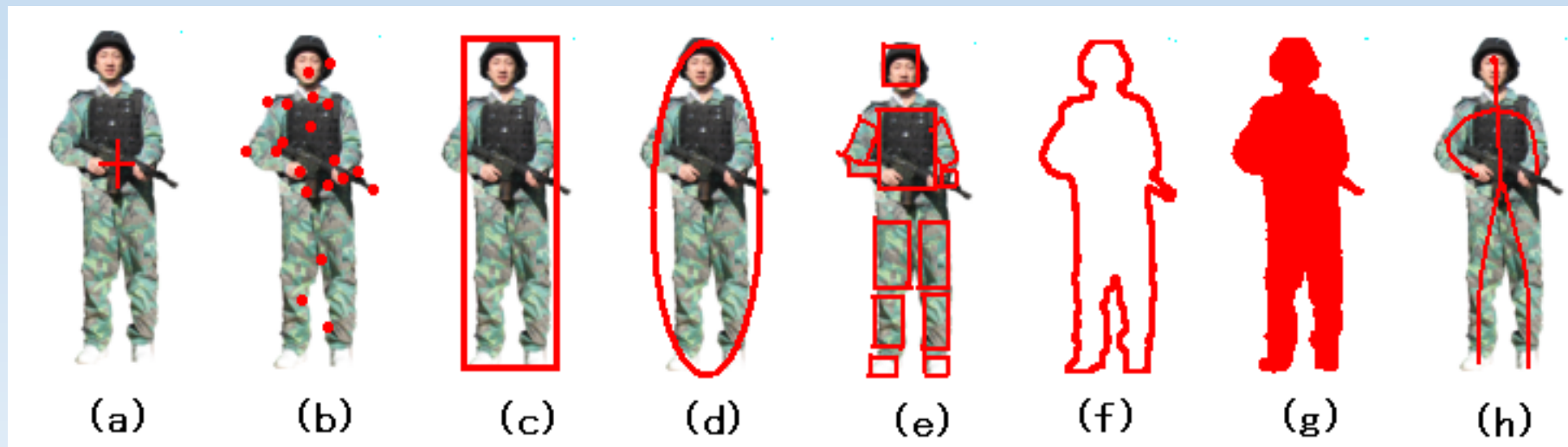
person 0.44



目标跟踪问题的描述

- 怎样表示 (Representation) 被跟踪的目标?
- 选择哪些特征 (Feature) 作为跟踪输入?
- 使用什么模型 (Model) 对目标的运动、特征及形状进行跟踪?

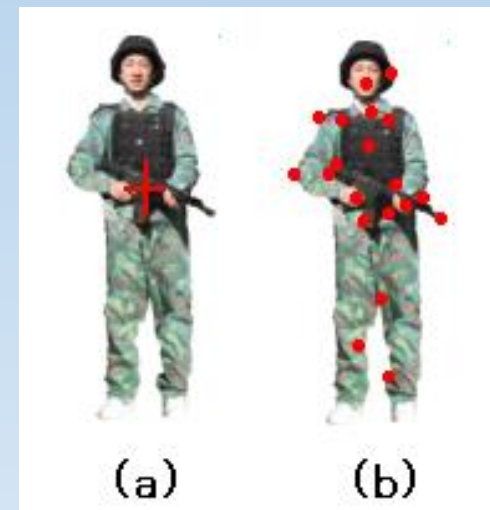
(1) 目标跟踪的表示方法:



(a) 质心表示, (b) 特征点集表示, (c) 矩形框表示, (d) 椭圆框表示, (e) 关节模型表示, (f) 轮廓表示, (g) 侧影表示, (h) 骨架模型表示

点表示：

在图像中用一个点（图a）或点集（图b）来表示被跟踪目标，这些点一般取为目标的质心、角点等。这种表示方式通常使用在目标区域相对整个图像比较小或者目标为只有几个到十几个像素的点目标的情况下。



基本几何形状表示：

目标被表示为一个矩形区域（图c）或椭圆区域（图d），这种表示法可以描述目标的平移、仿射变换或者投影变换，并且对刚体和非刚体目标都可以表示。



关节模型表示：

当目标可以认为是由各个部位通过结点衔接到一起的时候，就可以用关节模型来表示（图e），而各个部位则可以用几何形状来表示。



轮廓表示：

目标的外边界称为轮廓（图f），而轮廓内的区域被称为侧影（图g），它们都可以用来表示复杂的非刚体目标。



骨架模型表示：

目标骨架（图h）可以通过对侧影的中轴变换得到。骨架模型通常可以表示关节模型目标和非刚体目标。



目标跟踪任务分类

- **单目标跟踪** - 给定一个目标，追踪这个目标的位置。
- **多目标跟踪** - 追踪多个目标的位置
- **Person Re-ID** - 行人重识别，是利用计算机视觉技术判断图像或者视频序列中是否存在特定行人的技术。广泛被认为是一个图像检索的子问题。给定一个监控行人图像，检索跨设备下的该行人图像。旨在弥补固定的摄像头的视觉局限，并可与行人检测/行人跟踪技术相结合。
- **MTMCT** - 多目标多摄像头跟踪（Multi-target Multi-camera Tracking），跟踪多个摄像头拍摄的多个人
- **姿态跟踪** - 追踪人的姿态

目标跟踪方法

按照时间顺序，目标跟踪的方法经历了从经典算法到基于核相关滤波算法，再到基于深度学习的跟踪算法的过程。

- 经典跟踪算法（帧差法）
 - 模板匹配 (Template Match)
 - 均值漂移 (Mean Shift)
- 基于核相关滤波的跟踪算法
 - 卡尔曼滤波器 (Kalman Filter)
 - 粒子滤波器 (Particle Filter)
- 基于深度学习的跟踪算法

帧差法 ---最常用的运动目标检测和分割方法之一。

基本原理：在图像序列**相邻两帧或三帧**间采用基于像素的**时间差分**来提取出图像中的运动区域。

将相邻帧图像对应像素值相减得到差分图像，然后对差分图像二值化。

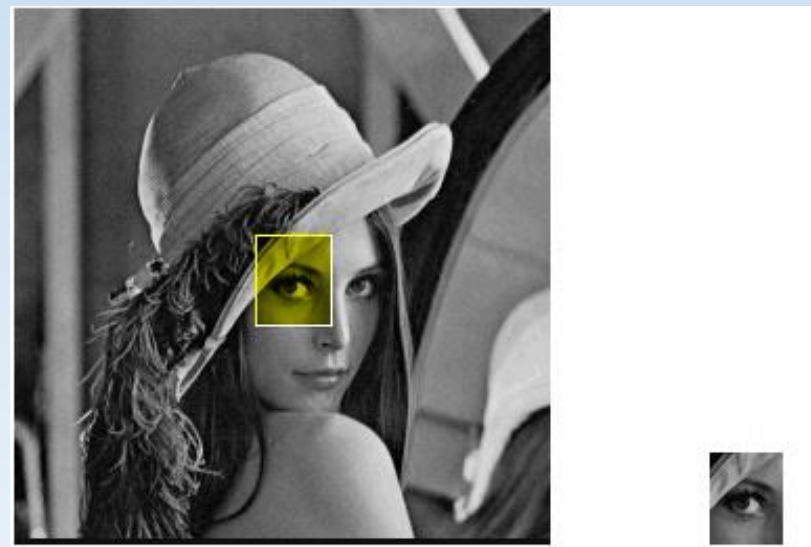
在环境亮度变化不大的情况下，如果对应像素值变化小于事先确定的阈值时，认为此处为背景像素；如果图像区域的像素值变化很大，认为这是由于图像中运动物体引起的，将这些区域标记为前景像素，利用标记的像素区域可以确定运动目标在图像中的位置。

优点：由于相邻两帧间的时间间隔非常短，用前一帧图像作为当前帧的背景模型具有较好的实时性，其背景不积累，且更新速度快、算法简单、计算量小。

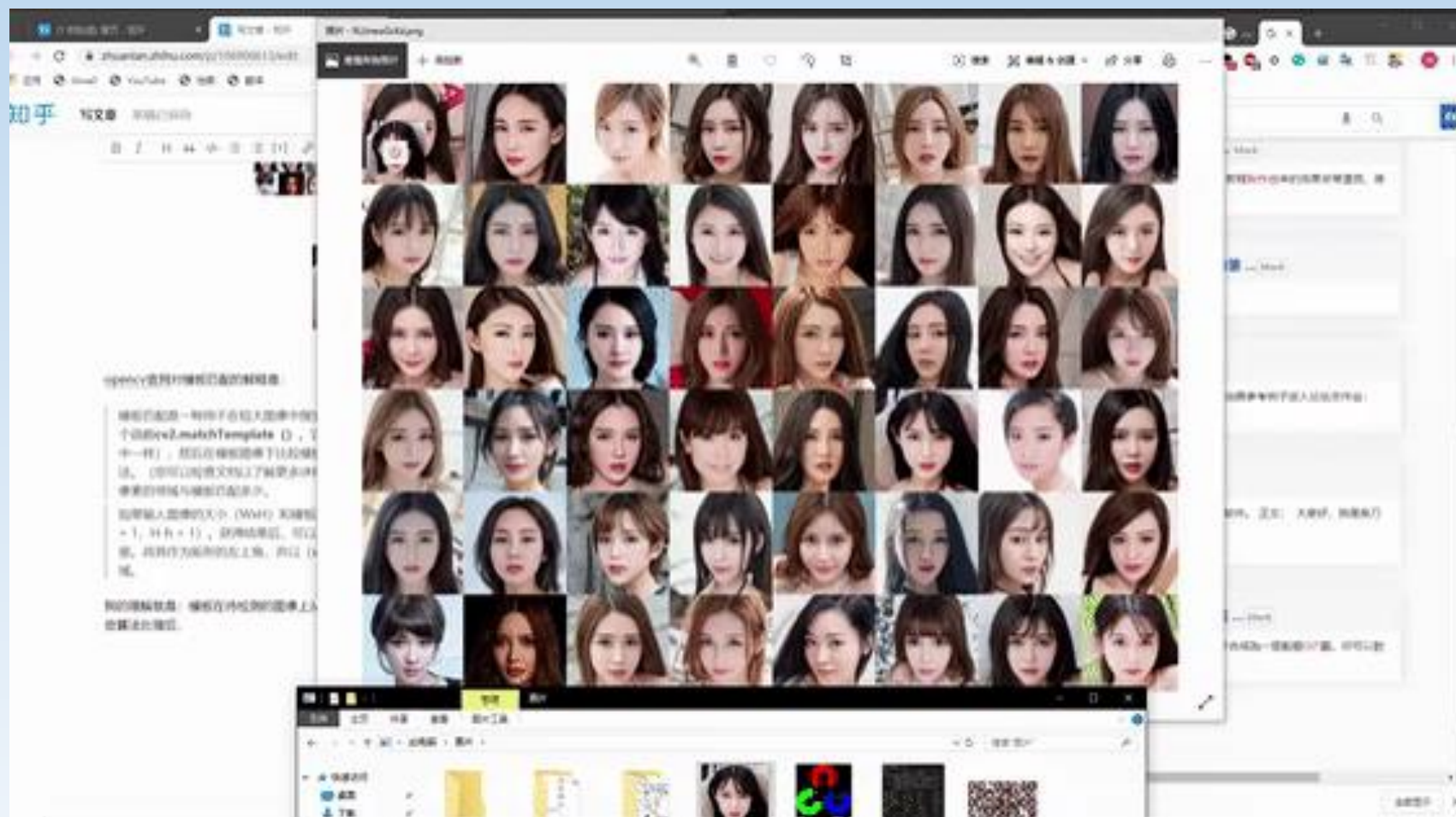
不足：在于对环境噪声较为敏感，阈值的选择很关键，选择过低不足以抑制图像中的噪声，过高则忽略了图像中有用的变化。对于比较大的、颜色一致的运动目标，有可能在目标内部产生空洞，无法完整地提取运动目标。

1、模板匹配法 (Template Matching)

在一幅图像中寻找一个特定目标，这种方法的原理非常简单，遍历图像中的每一个可能的位置，比较各处与模板是否“相似”，当相似度足够高时，就认为找到目标。



概念上比较简单，进行穷尽的搜索计算量非常大。



算法实现:

1. 平方差匹配法
2. 归一化平方差匹配法
3. 相关匹配法
4. 归一化相关匹配法
5. 相关系数匹配法
6. 归一化相关系数匹配法

设得到的矩阵为 $R(x, y)$,模板图像矩阵为 $T(x', y')$,源图像矩阵为 $I(x, y)$

(1) 平方差匹配法 CV_TM_SQDIFF

误差平方和:

$$R(x, y) = \sum_{x', y'} (T(x', y') - I(x + x', y + y'))^2$$

匹配测度

| | | |
|---|---|---|
| 0 | 0 | 5 |
| 0 | 5 | 0 |
| 5 | 0 | 0 |

模板

| | | |
|---|---|---|
| 0 | 1 | 4 |
| 0 | 3 | 0 |
| 6 | 0 | 1 |

源图像

| | | |
|---|---|---|
| 0 | 1 | 1 |
| 0 | 4 | 0 |
| 1 | 0 | 1 |

误差平方和

模板图像像素减去覆盖的源图像像素的差的平方和为对应矩阵的点的值。
误差平方和越接近0，说明匹配程度越高。

模板灰度匹配的特点

- (1) 图像局部匹配。
- (2) 当原图像发生旋转或缩放时，不能实现匹配。



直方图匹配： 根据颜色统计特性进行相似性匹配。

概念： 以一幅图的直方图为基础，找到**原图像**的灰度分布与**模板图像**的灰度分布情况相近的部分。从算法设计上是寻找一个灰度级变换的函数。一般是建立两个图像均值和方差之间的线性关系。

用途： 基于颜色的视频跟踪

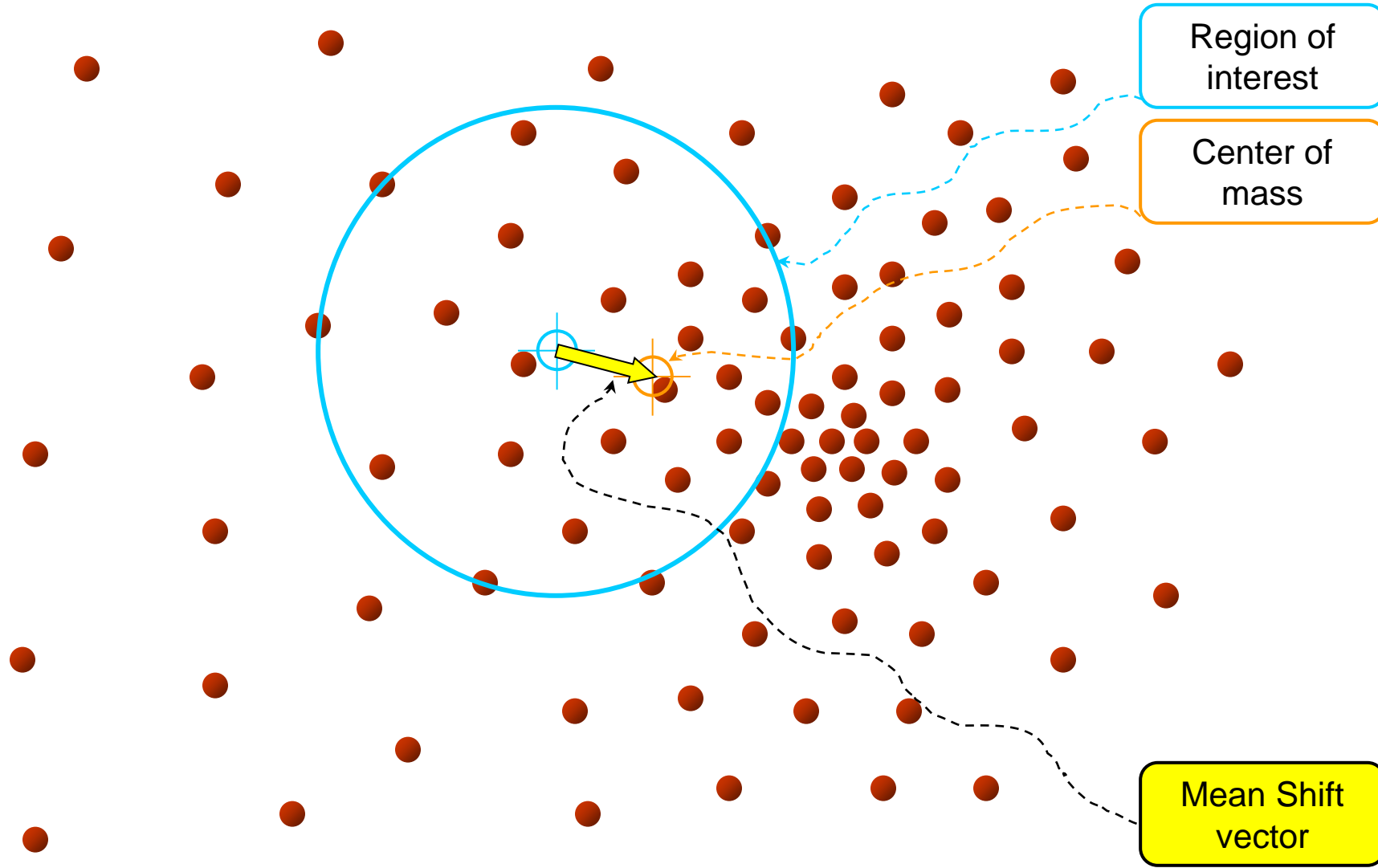
方法： 根据图像直方图进行距离测量：直方图相交法、欧几里得距离法、中心矩法、街坊距离法等。

Mean Shift

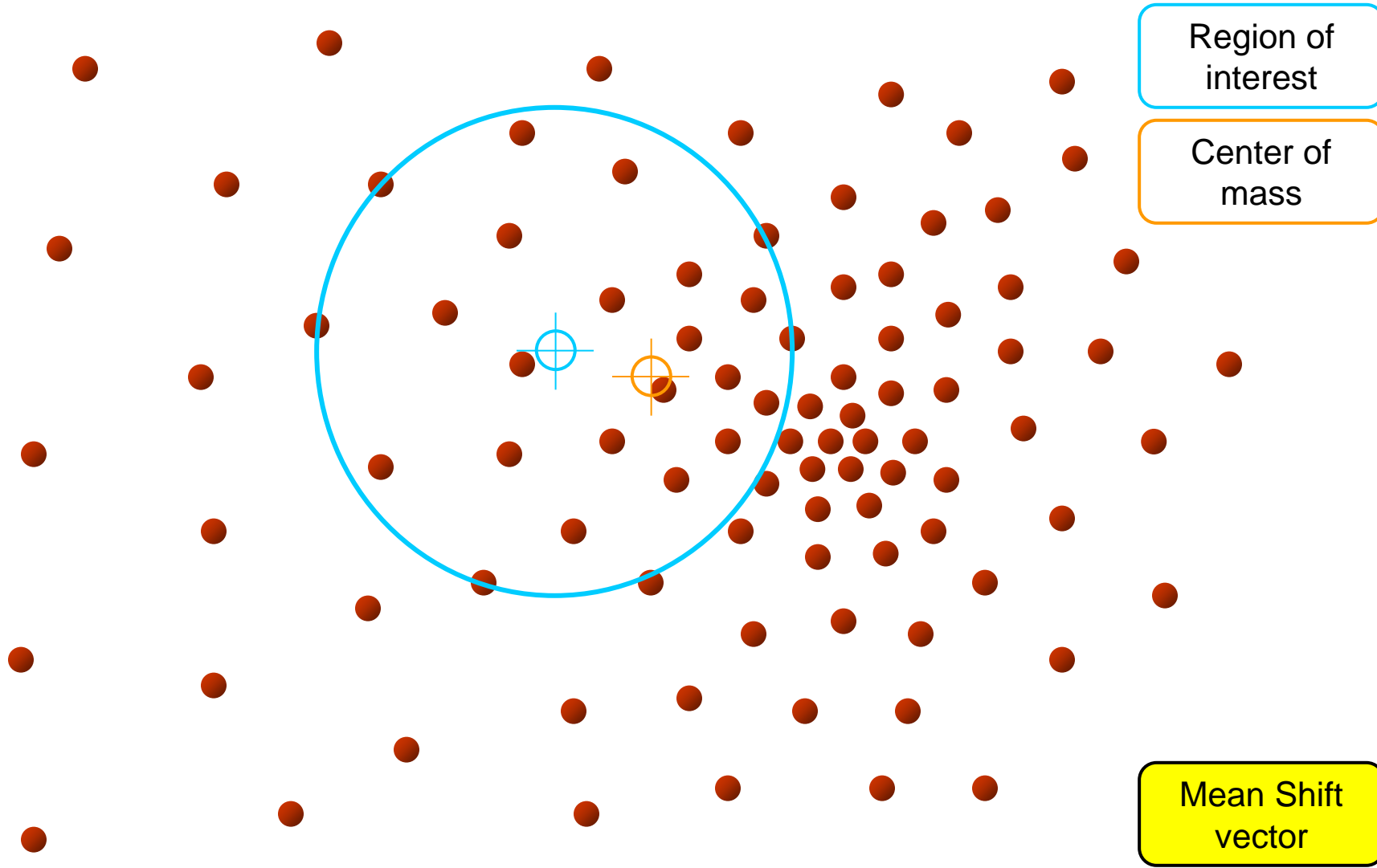
Mean shift 方法是一种基于概率密度分布的跟踪方法，使目标的搜索一直沿着概率梯度上升的方向，迭代收敛到概率密度分布的局部峰值上。

首先 Meanshift 会对目标进行建模，比如利用目标的颜色分布来描述目标，然后计算目标在下一帧图像上的概率分布，从而迭代得到局部最密集的区域。

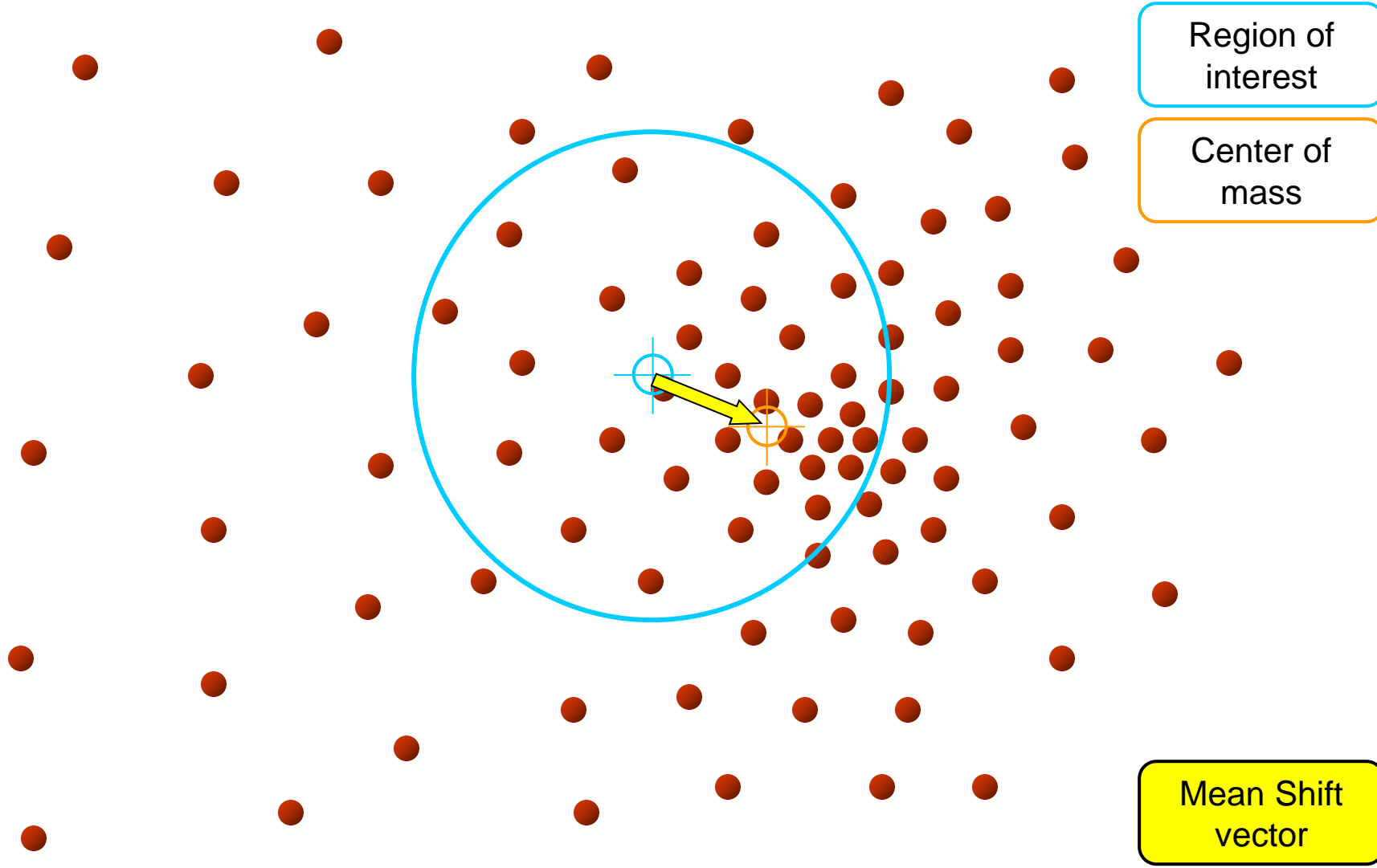
Meanshift 适用于目标的色彩模型和背景差异比较大的情形，早期也用于人脸跟踪。由于 Meanshift 方法的快速计算，它的很多改进方法也一直适用至今。



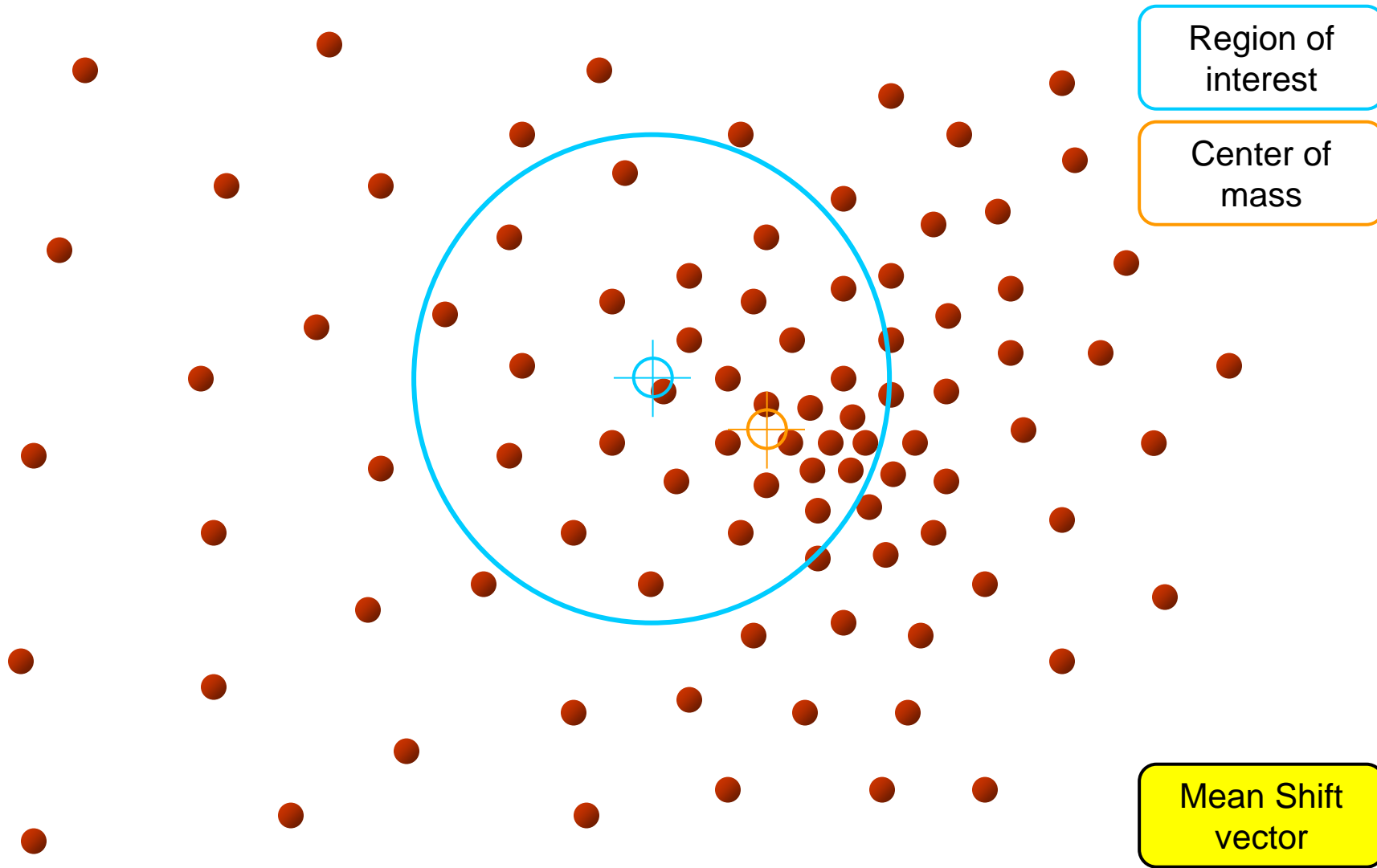
Objective : Find the densest region
Distribution of identical billiard balls



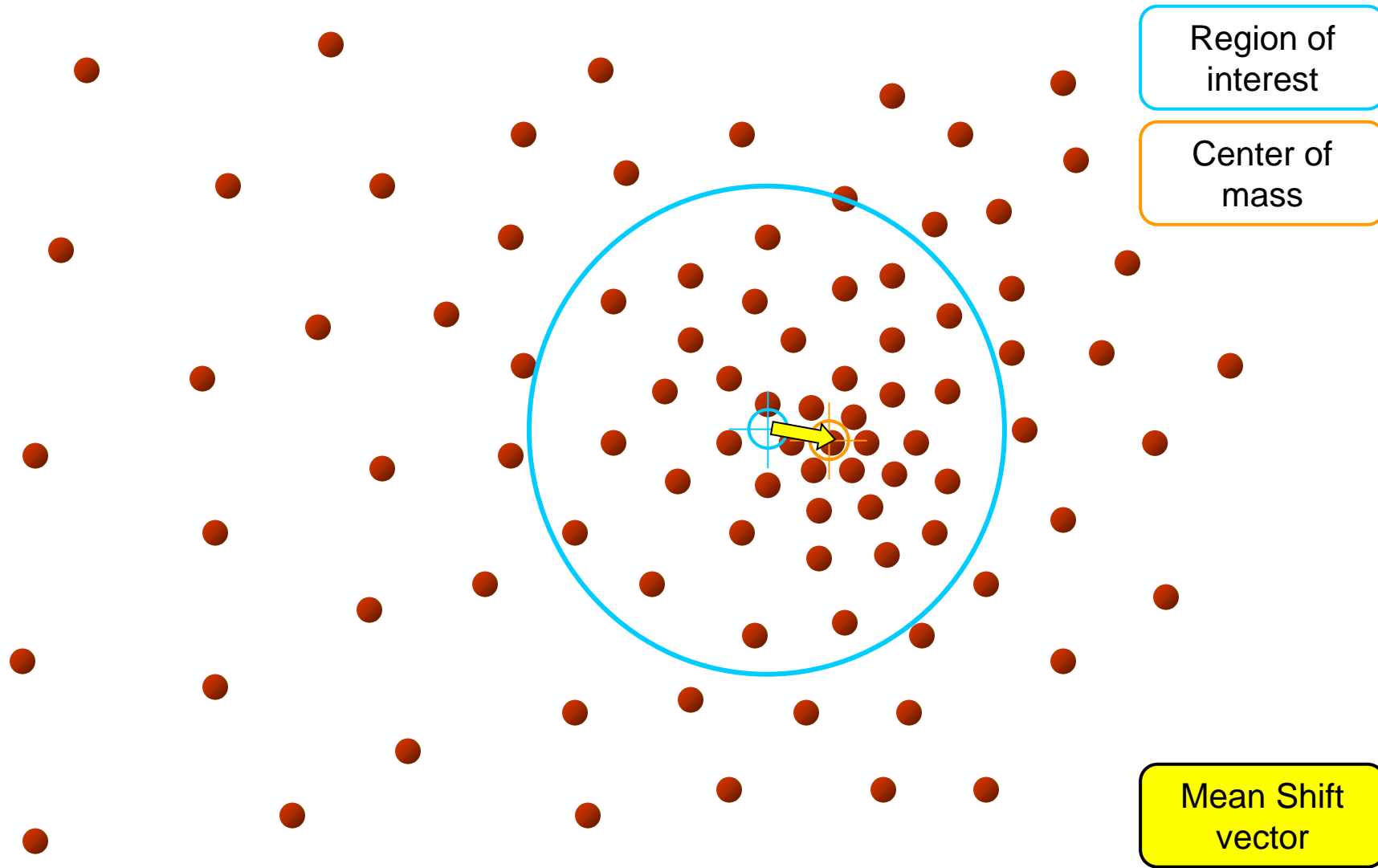
Objective : Find the densest region
Distribution of identical billiard balls



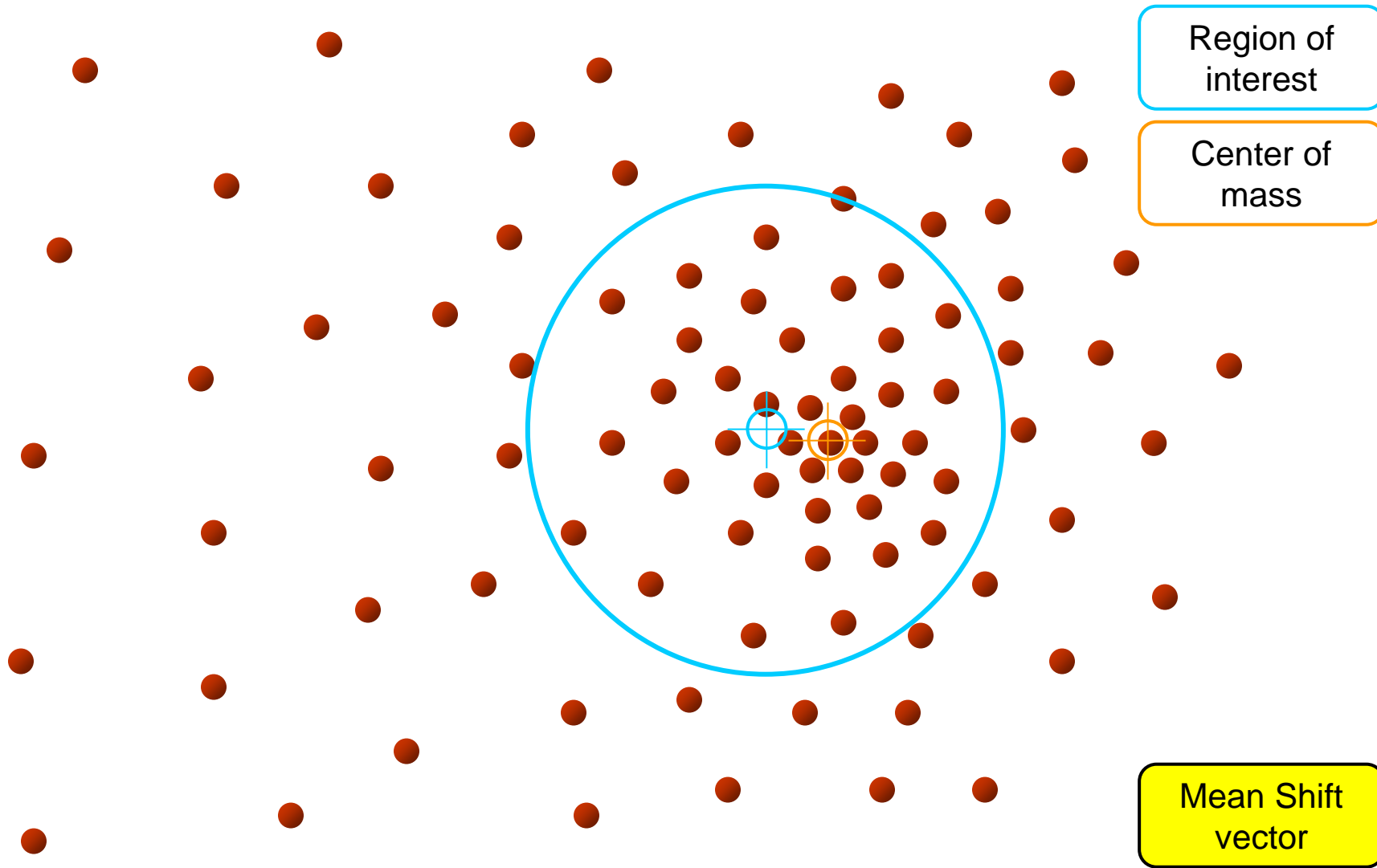
Objective : Find the densest region
Distribution of identical billiard balls



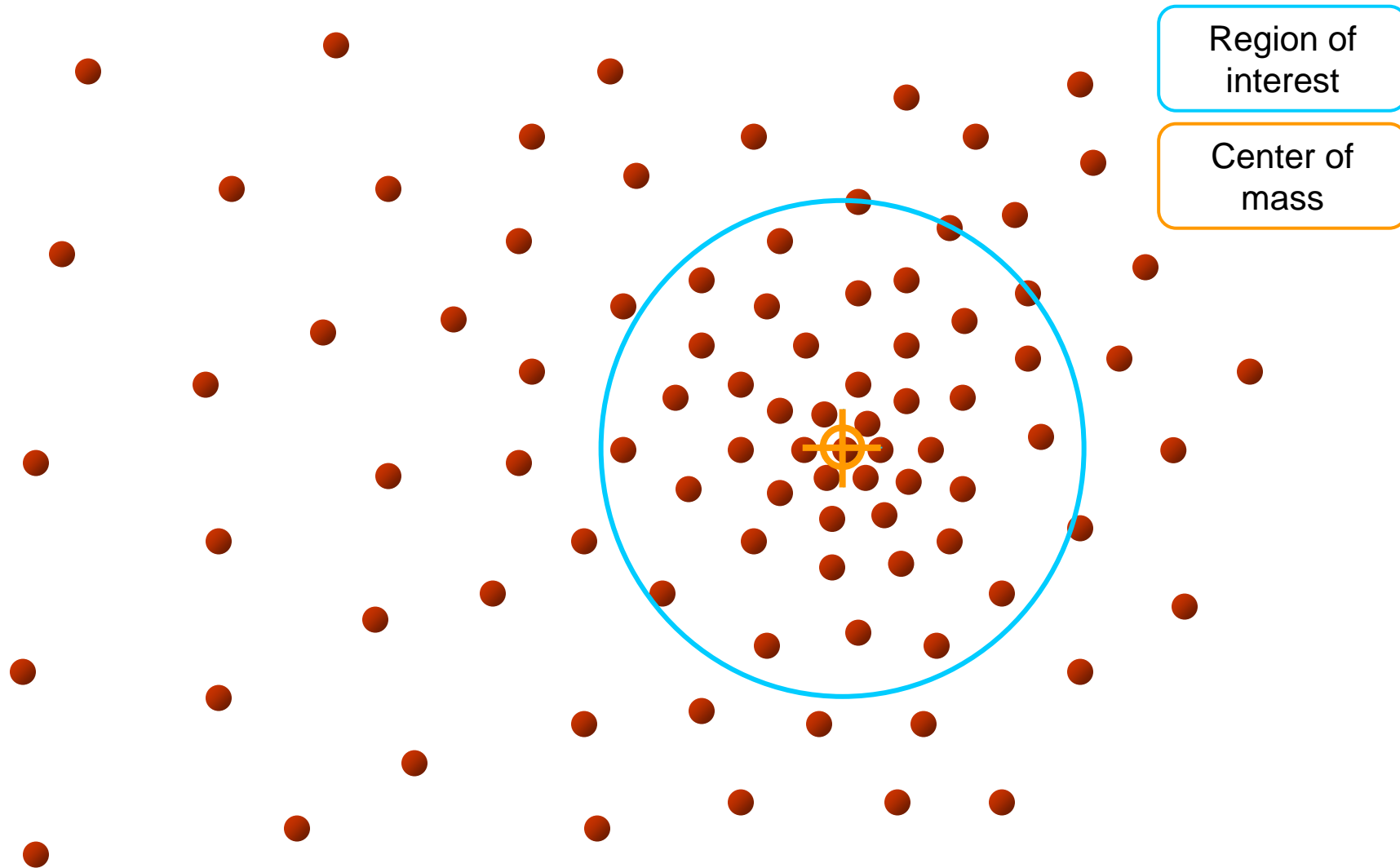
Objective : Find the densest region
Distribution of identical billiard balls



Objective : Find the densest region
Distribution of identical billiard balls



Objective : Find the densest region
Distribution of identical billiard balls



Objective : Find the densest region
Distribution of identical billiard balls

mean shift原始公式如下：表示的是IOR范围内中心点到重心的路程即mean shift，该值是有方向的

$$M(\mathbf{x}) = \frac{1}{n} \sum_{i=1}^n (\mathbf{x} - \mathbf{x}_i)$$

由于不够鲁棒，于是引入了核函数和权重变成了：

基于以上的考虑，对基本的Mean Shift向量形式中增加核函数和样本权重，得到如下的改进的Mean Shift向量形式：

$$M_h(x) = \frac{\sum_{i=1}^n G_H(x_i - x) w(x_i) (x_i - x)}{\sum_{i=1}^n G_H(x_i - x) w(x_i)}$$

其中：

$$G_H(x_i - x) = |H|^{-\frac{1}{2}} G\left(H^{-\frac{1}{2}}(x_i - x)\right)$$

$G(x)$ 是一个单位的核函数。 H 是一个正定的对称 $d \times d$ 矩阵，称为带宽矩阵，其是一个对角阵。 $w(x_i) \geq 0$ 是每一个样本的权重。对角阵 H 的形式为：

$$H = \begin{pmatrix} h_1^2 & 0 & \cdots & 0 \\ 0 & h_2^2 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & h_d^2 \end{pmatrix}_{d \times d}$$

上述的Mean Shift向量可以改写成：

$$M_h(x) = \frac{\sum_{i=1}^n G\left(\frac{x_i - x}{h_i}\right) w(x_i) (x_i - x)}{\sum_{i=1}^n G\left(\frac{x_i - x}{h_i}\right) w(x_i)}$$

通过人工标注的方式在初始帧确定包含跟踪目标的区域。假设其中有 n 个像素用 $\{z_i\}_{i=1\dots n}$ 表示其位置，对选中的区域的灰度颜色空间均匀划分，得到由 m 个相等的区间构成的灰度直方图。目标模型的 q_u 概率密度($u=1,\dots,m$)可表示为：

$$\begin{aligned} q_u &= C \sum_{i=1}^n K\left(\|z_i^*\|^2\right) \delta[b(z_i)-u] \\ C &= 1 / \sum_{i=1}^n K\left(\|z_i^*\|^2\right) \\ z_i^* &= \left(\frac{(x_i - x_0)^2 + (y_i - y_0)^2}{x_0^2 + y_0^2} \right)^{0.5} \end{aligned} \quad (2.11)$$

其中， z_i^* 表示以目标中心为原点的归一化像素位置， (x_0, y_0) 为目标中心坐标。 K 是核函数，常选用 Epanechnikov 核函数， $b(z_i)$ 表示 z_i 处像素属于哪个直方图区间， u 为直方图的颜色索引。 $\delta[b(z_i)-u]$ 函数的作用是判断目标区域中像素 z_i 处的灰度值是否属于直方图中第 u 个单元，等于为 1，否则为 0。 C 是归一化系数。

有了初试的密度函数了，就要在下一帧选取对应位置，计算密度函数。

在第 t 帧时，根据第 $t-1$ 帧的目标中心位置 f_0 ，以 f_0 为搜索窗口的中心，得到候选目标的中心位置坐标 f ，计算当前帧的候选目标区域直方图。该区域的像素用

$$\{z_i\}_{i=1\dots n} \text{ 表示，则候选模型的概率密度为： } p_u(f) = C \sum_{i=1}^n K\left(\left\|\frac{f-z_i}{h}\right\|^2\right) \delta[b(z_i)-u],$$

h 为核函数窗口大小，决定着权重分布，其他参数同目标模型描述。

前后两帧的密度函数都有了，我们需要定义目标函数，即相似性

3、相似性度量

相似性函数用于描述目标模型和候选目标之间的相似程度。本文采用 Bhattacharyya 系数作为相似性函数，其定义为： $\rho(p, q) = \sum_{u=1}^m \sqrt{p_u(f)q_u}$ 。相似函数越大则两个模型越相似。将前一帧中目标的中心位置 f_0 作为搜索窗口的中心，寻找使得相似函数最大的候选区域，即是在本帧中目标的位置。

上边p,q相同的时候该函数最大，即前后两帧选择的框最相近。

有了目标函数就是怎么求解了，最简单的就是迭代了，那迭代步长怎么计算呢。求目标函数的导数了，这里采用先泰勒展开求近似导数。

4、Meanshift 迭代过程

均值漂移的迭代过程，也就是目标位置搜索的过程。为使相似函数最大，对上式进行泰勒展开，得到 Bhattacharyya 系数的近似表达：

$$\rho(p, q) \approx \frac{1}{2} \sum_{u=1}^m \sqrt{p_u(f_0)q_u} + \frac{C}{2} \sum_{i=1}^n w_i K\left(\left\|\frac{f - z_i}{h}\right\|^2\right) \quad (2.12)$$
$$w_i = \sum_{u=1}^m \sqrt{\frac{q_u}{p_u(f)}} \delta[b(z_i) - u]$$

(2.12)中只有第二项随 f 变化，其极大化过程就可以通过候选区域中心向真实区域中心的 Meanshift 迭代方程完成：

$$f_{k+1} = f_k + \frac{\sum_{i=1}^n w_i (f_k - z_i) g\left(\left\|\frac{f_k - z_i}{h}\right\|^2\right)}{\sum_{i=1}^n w_i g\left(\left\|\frac{f_k - z_i}{h}\right\|^2\right)} \quad (2.13)$$

其中， $g(x) = -K'(x)$ ，Meanshift 方法就是从 f_k 起向两个模型相比颜色变化最大的方向不断移动，直到最后两次移动距离小于阈值，即找到当前帧的目标位置，并以此作为下一帧的起始搜索窗口中心，如此重复。

本章结束!

