

遥感图像与视频处理技术

第8讲—目标检测跟踪

郑 锦

Email: JinZheng@buaa.edu.cn



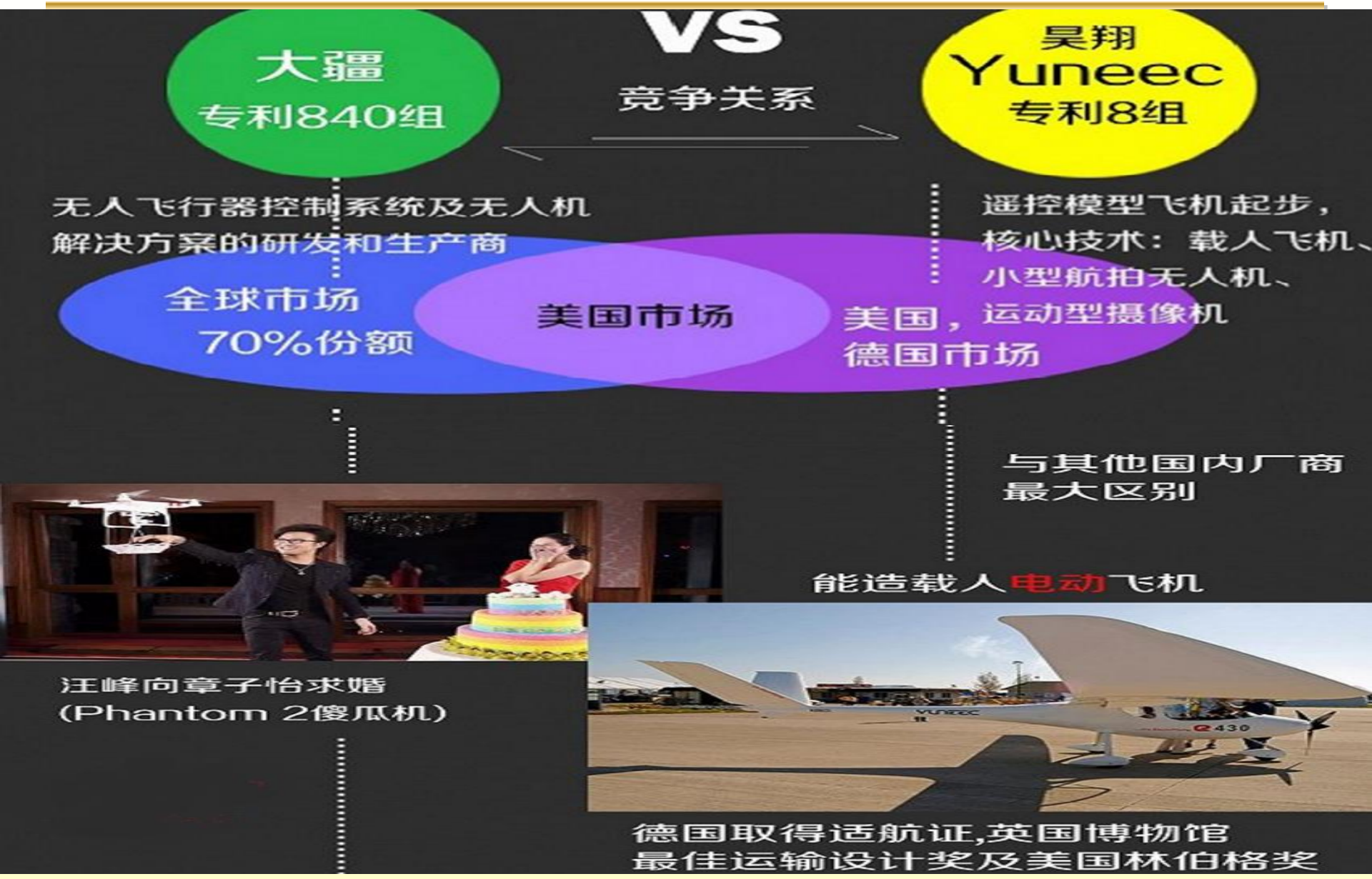
目录

1、引言 一个例子

2、运动目标检测

3、运动目标跟踪

大疆、昊翔专利战 (2016年4月)



03 大疆被侵犯的美国专利

目标跟踪系统和方法 (US9164506)

——视觉跟踪和避障功能上

权利要求30个，索赔项8个

权利里：任何一种可以被图像设备找到的目标信息都被视为保护范围，

可更换云台 (US9280038)

——为影像“供稳”的支撑设备上

权利要求30个，索赔项19个

申请

2014.08

授权

2015.10.20

2016.03.08

大疆、昊翔专利战 (2016年4月)

精灵4 ¥8999.00

实现自主避障绕行、
视觉识别自动跟拍移动物体、
指点方向自主飞行、
云台设计



台风H ¥8888.00

智能避障、
智能选择规划路线、
4K视频拍摄、
360度自由旋转



PK

高端市场

大疆、昊翔专利战（2016年4月）

- **避障：**

- **大疆**——双目摄像头+超声波技术
- **昊翔**——单目摄像头+结构光

- **跟踪：**

- **大疆**——视觉跟踪（主摄像头）
- **昊翔**——GPS跟踪

大疆、昊翔专利战（2016年4月）

- **GPS跟踪：**是指在被跟踪目标上佩戴GPS信号器，无人机跟踪GPS信号，因此实现对目标的跟踪，这个过程就是以跟踪A而实现跟踪B。
- **视觉跟踪：**对连续的图像序列进行运动目标检测、提取特征、分类识别、跟踪滤波、行为识别，以获得目标准确的运动信息参数（如位置、速度等），并对其进行相应的处理分析，实现对目标的行为理解。

大疆、昊翔专利战（2016年4月）

- **大疆**：诉讼状提及的506专利中对于目标跟踪的权利主张，从权利要求书中对target的限定，任何一种可以被图像设备找到的目标信息都被视为保护范围，而这其中就包括了GPS信号跟踪。
- 只要是无无人机去追踪一个被拍摄的目标，就落入了该专利的保护范围？

无人机视觉跟踪现状

- 现有无人机的视觉跟踪技术：

- 一种是索尼等专业相机企业所使用的**动态像素识别**。索尼的相机可以实时追踪动态的目标如飞翔的小鸟，而且一直是对焦清晰的状态

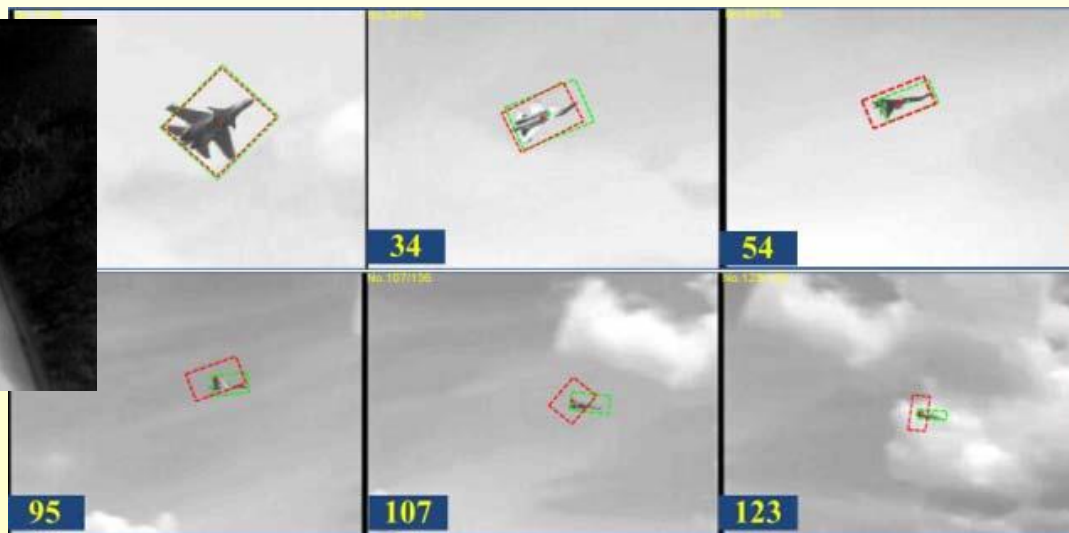
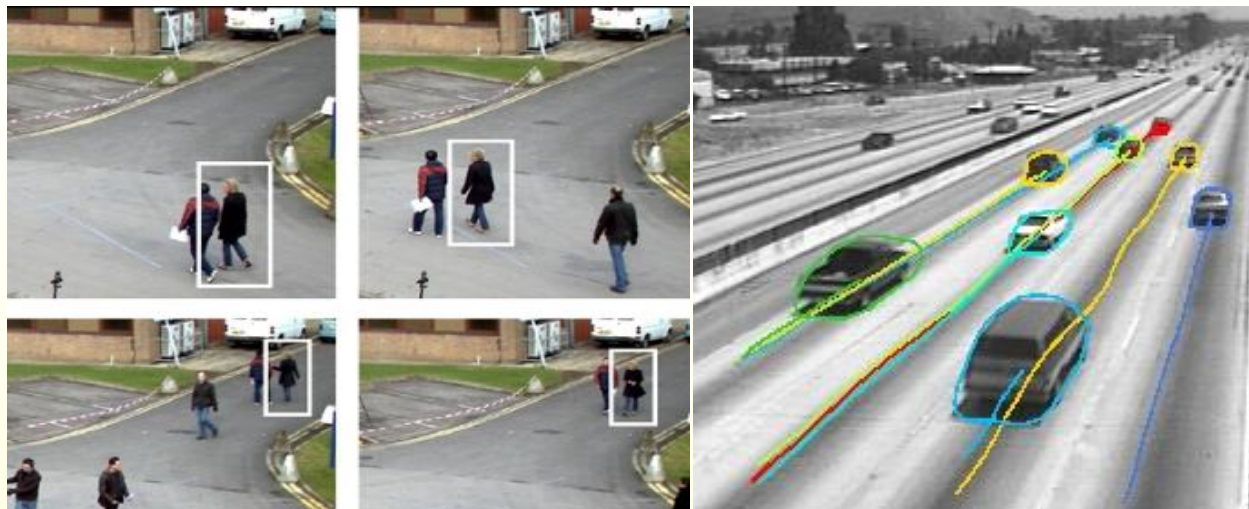
运动特征跟踪？

- 一种是大疆精灵4采用的是**大面积色块识别**技术。精灵4一旦被跟踪目标的颜色发生了巨大变化，就无法继续跟踪

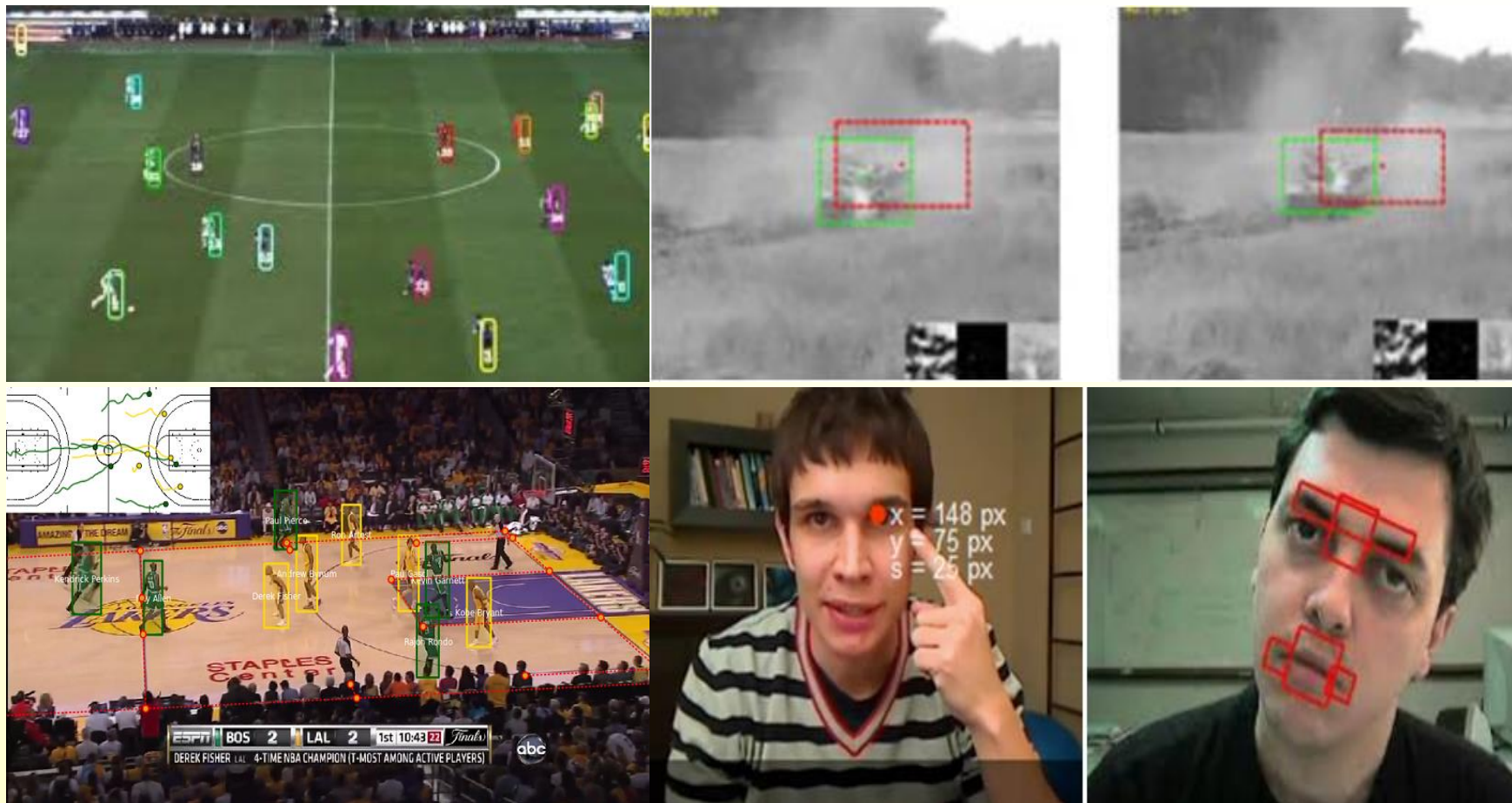
颜色特征跟踪？

运动目标检测、运动目标跟踪应用

应用领域：**安防**
监控，**交通管理**，
体育分析，
军事识别，人机
交互……



运动目标检测、运动目标跟踪应用

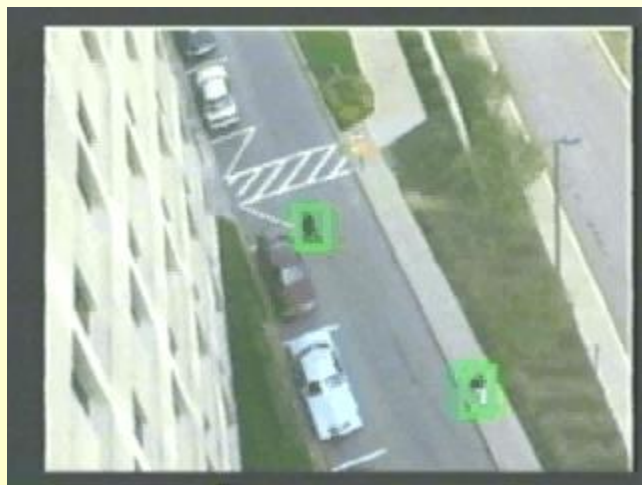


航拍视频主要以可见光、红外视频为主

1、VSAM

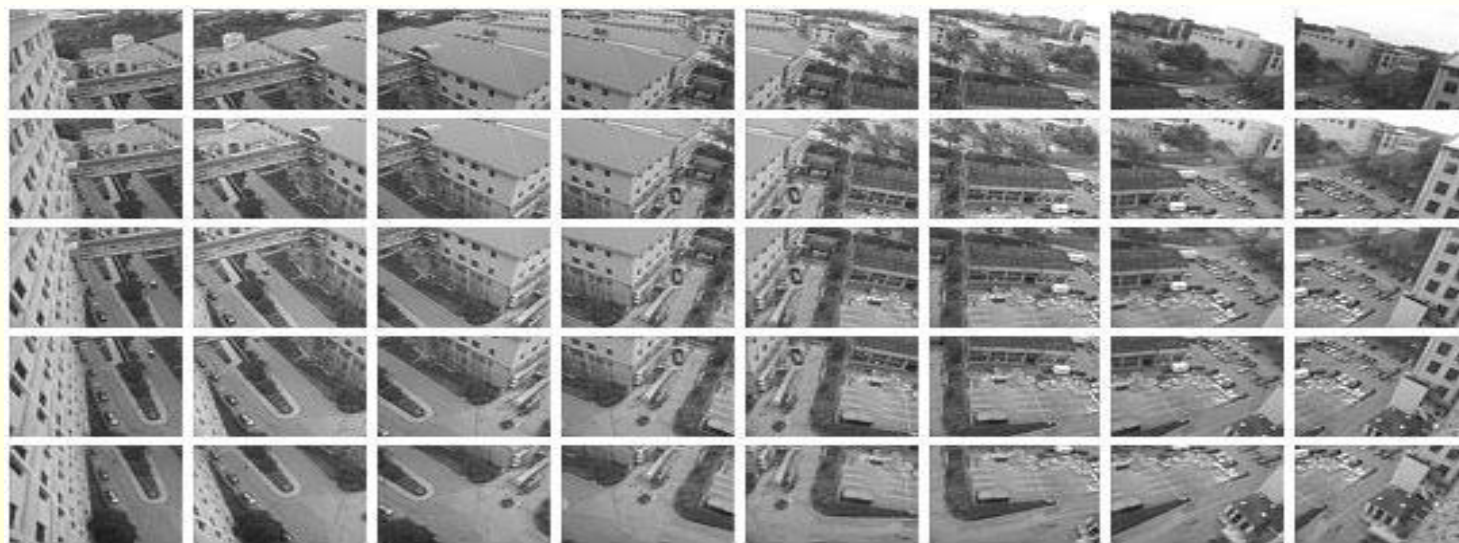
已知的一些系统

- 1997年至1999年间，美国国防高级研究计划局资助卡内基梅隆大学和萨尔诺夫戴维研究中心等著名大学和公司联合研制的视频监视与监控系统。
- 目标是开发**自动视频理解技术**，并用于实现未来战争、人力费用昂贵、非常危险或者人力无法实现等场合的监控。
- 融合了**数字摄像机、音频采集头、红外和微波报警探测器、温度探测器等多种类型的传感器**，可以对监控地区进行全方位的昼夜监控。
- 在机载航空摄像机方面，不需要经常性的人工操纵，就能**自动对准地面监视目标，实现对重要目标的长时间监视。**



1、VSAM

使用架设在高处多方位旋转云台上的单个摄像机，全方位地实施视频监控。系统首先有规律地初始化一系列背景图像，然后利用基于特征区域的方法将实际摄录的视频图与相应的背景图作匹配，再利用**背景减除法**检测运动目标。



1、VSAM

针对机载航空摄像机所拍摄的视频图像，该技术的关键在于对航空**摄像机**的**自运动补偿**，对经过补偿的图像，利用**三帧差减**的方法检测目标。



1、VSAM

对传统的**卡尔曼滤波**思想进行了扩展，并使用了**带目标模板更新的相关匹配算法**实现了多目标的跟踪。



2、美国康奈尔大学：航拍视频目标检测与持续跟踪系统

能够对多运动目标实现较长时间的准确跟踪，即使在短时间内目标被遮挡，或移出视场，或目标时静时动。系统在运动背景估计与补偿中所涉及的主要技术是基于**Kanade-Lucas-Tomasi(KLT)算法的特征点跟踪**和**基于M估计的鲁棒性仿射参数估计**。利用**三帧差减**的方法检测目标运动，利用**形态学操作**分割图像并定位运动目标。对多目标进行标记之后，利用**Hausdorff距离匹配和模板更新**的方法对目标进行长时间的跟踪。



3、COCOA

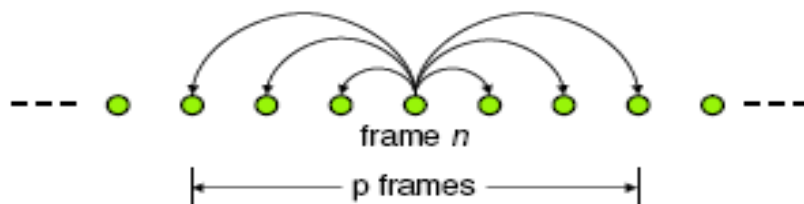
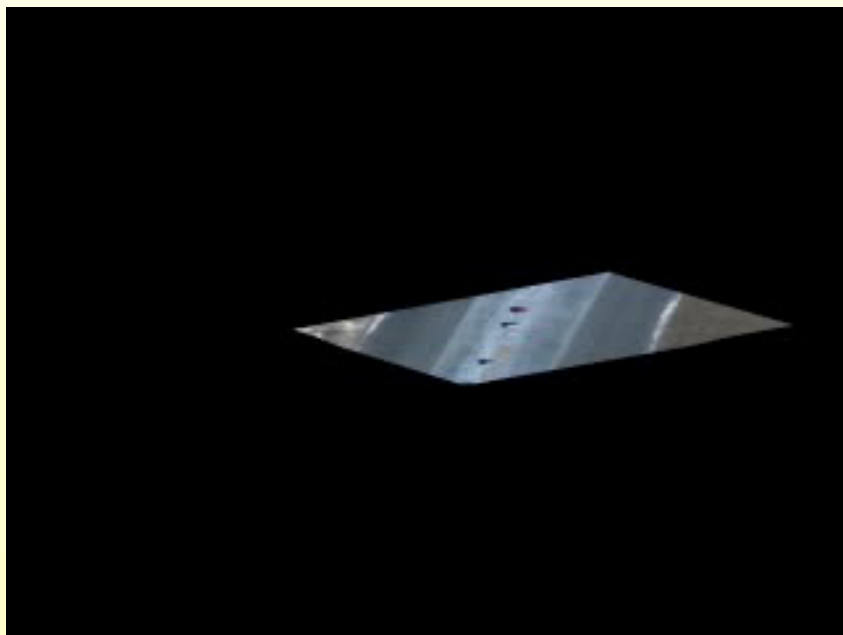
COCOA系统是一种无人机航拍图像目标跟踪系统。该系统针对一段视频图像序列，通过三大技术环节，即**背景运动补偿、运动检测与目标跟踪**，来完成对目标的检测与跟踪。该系统基于MATLAB平台，可以适用于不同的光传感器（可见光或红外），最小的可跟踪目标约为100像素大小。

该系统对**机载光电传感器或红外传感器**所摄视频图像进行鲁棒性背景运动补偿，并可生成全景图，利于更高层次的应用。对图像中多种运动目标（如汽车、坦克、摩托车等）进行可靠性检测并进行持久地跟踪。



3、COCOA

在背景运动补偿方面，利用**Harris角点检测算法**分别提取相邻两帧图像的特征点，对每一特征点先进行简单的粗匹配，而后利用**RANSAC鲁棒估计算法**完成特征点的筛选，通过筛选后的特征点的运动矢量集合来估计全局运动矢量。在运动检测方面，通过**累积帧差法**对汽车、卡车、坦克、摩托车等独立运动进行检测，再利用数字图像**形态学操作**去除噪声、捕捉大致的目标区域，再利用**几何活动轮廓的水平集方法**提取目标轮廓，最后利用**基于核函数的方法**（如Mean Shift方法）与模板更新实现目标跟踪。



3、COCOA



视频序列

背景运动补偿

基于特征 +
基于灰度梯度

运动检测

累积帧差法 +
形态学操作

目标跟踪

Level Set方法 +
Mean Shift方法

COCOA系统的基本技术环节

目录

1、引言

2、运动目标检测

- 摄像机静止下的运动目标检测
- 摄像机运动下的运动目标检测

3、运动目标跟踪

目标检测方法的基本概念与原理

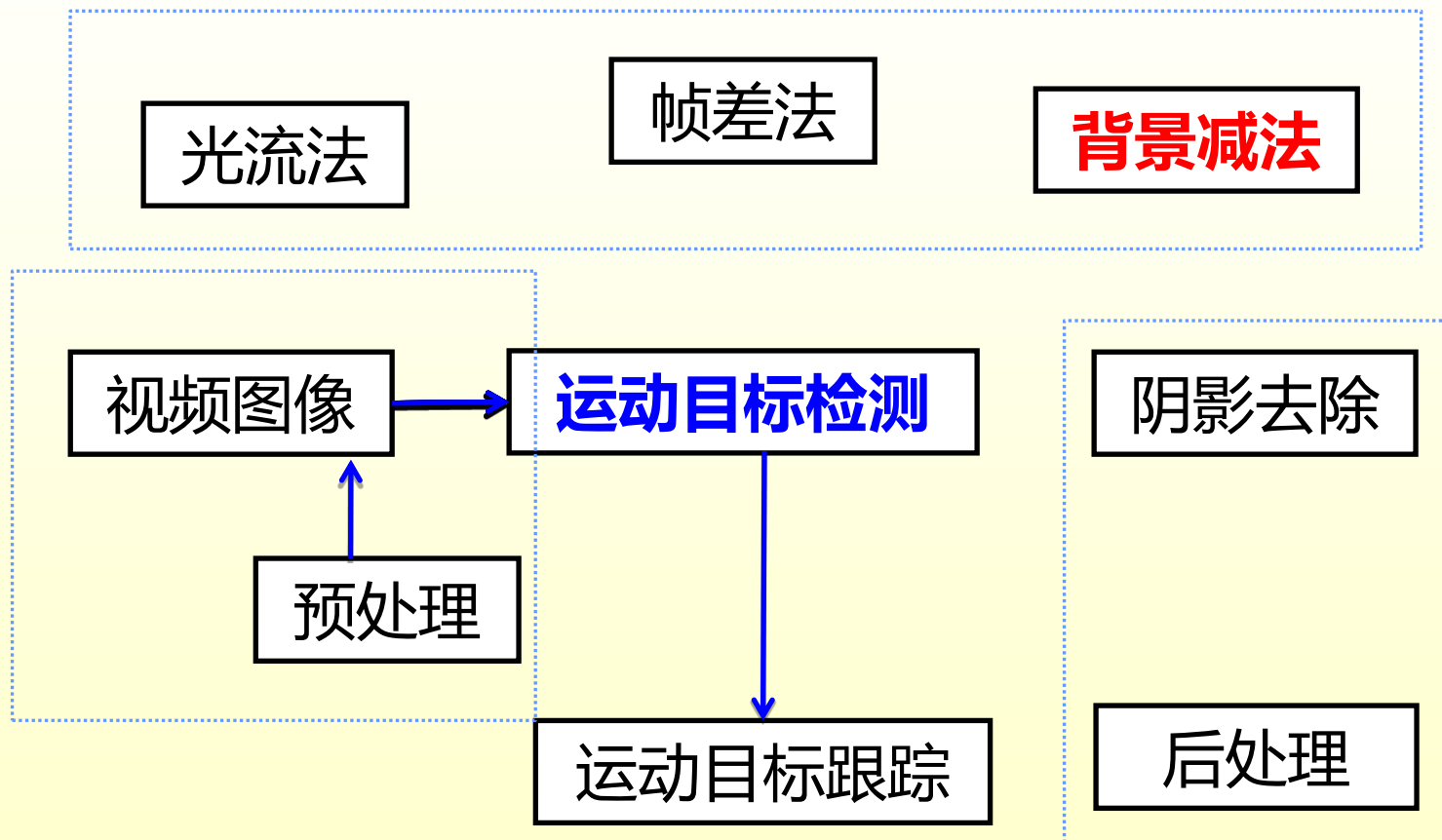
显著性检测？深度学习目标识别？

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；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。
- 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。

运动特征

运动目标检测基本方法及流程



- 彩色图像灰度化

$$GRAY = 0.114 \bullet R + 0.587 \bullet G + 0.299 \bullet B$$

- 图像的平滑滤波

- (1)中值滤波

- (2)均值滤波

- (3)高斯滤波

- 光流法

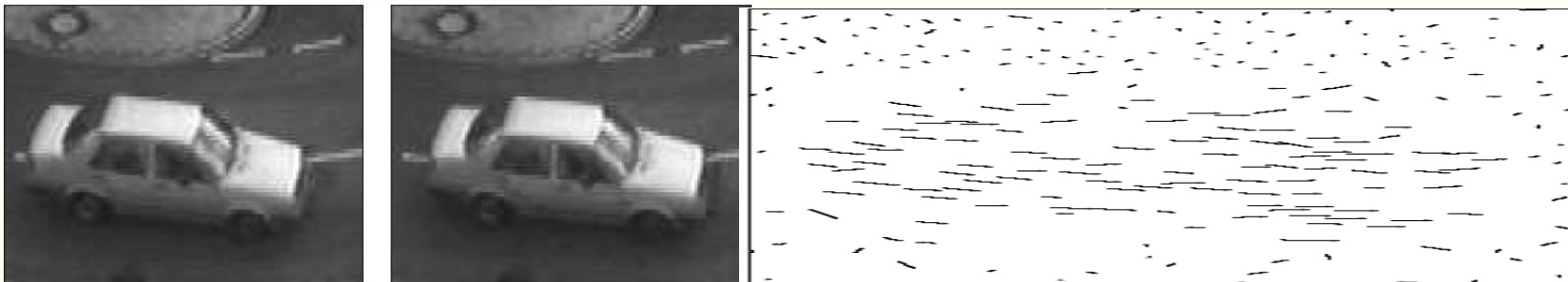
- 相邻帧差法

- 背景减法

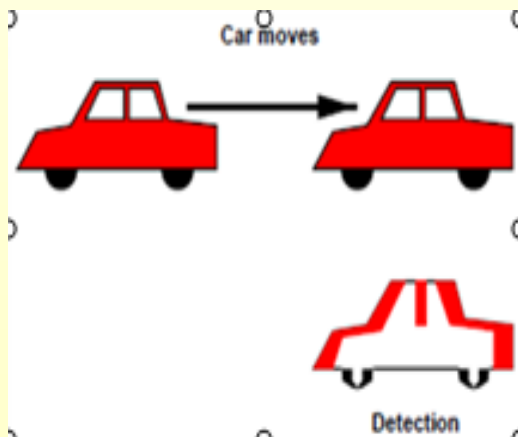
- 背景相减法
- 均值滤波法
- W4模型
- 自适应背景模型
- 单高斯模型
- 混合高斯模型

运动目标检测

□**光流法**：二维速度场，包含了物体表面结构和动态行为的重要信息



□**相邻帧差法**：将连续两帧或三帧图像对应像素点的灰度值相减



□背景减法

原理： 建立一个无运动目标的背景图像(第一帧无运动目标的图像或前N帧无运动目标的图像的均值或中值)，然后将当前图像的像素值与背景图像的像素值相减，通过设置一定的阈值，分割运动目标。

优点： 算法简单、实时性较高

缺点： 对背景的依赖性较高



□背景减法—均值滤波法



t-3帧

t-2帧

t-1帧

当前帧t

$$b(x, y, t) = \frac{1}{L} \sum_{i=0}^{L-1} f(x, y, t - i)$$

优点：运算量低，速度快，在有实时性要求且对准确性要求不高时得到广泛的应用；

缺点：对存储空间需求较高，速度慢的物体时可能出现空洞。

□背景减法—W4模型

最大灰度值 $MAX(x, y, t)$

最小灰度值 $MIN(x, y, t)$

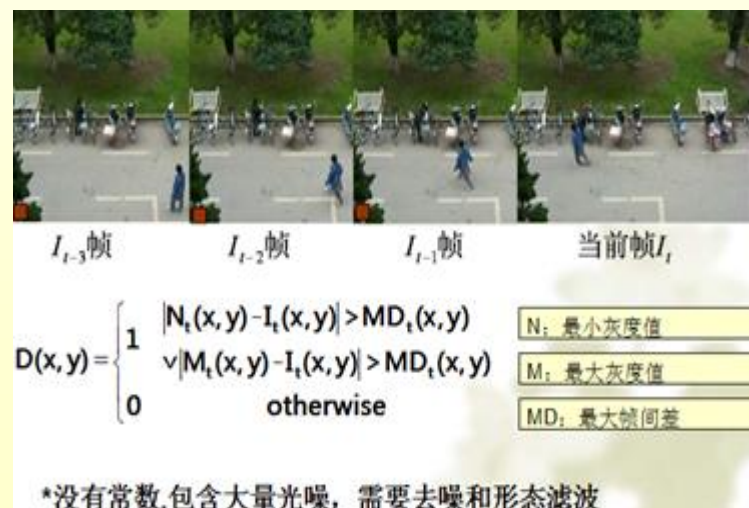
最大邻间差分 $DMAX(x, y, t)$ 描述, 相邻帧
对应位置像素灰度差的最大值

图像序列的
前 L 帧估计

新观测值 $f(x, y, t)$

优点: 同均值滤波法相比检测效果较好, 运算量也不大, 能够满足实时性要求, 而且对目标比较敏感。

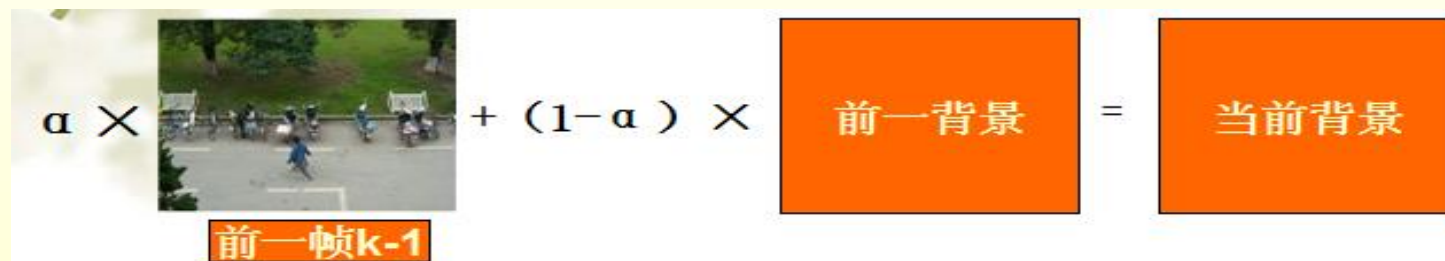
缺点: 当运动目标面积小且与背景对比度低时效果较差。



□背景减法—自适应背景模型

$$f_t = \begin{cases} 1, & |I_t - B_t| > T \\ 0, & |I_t - B_t| \leq T \end{cases}$$

$$B_{t+1}(x, y) = \begin{cases} \alpha \cdot B_t(x, y) + (1 - \alpha)I_t, & f_t = 0 \\ B_t(x, y), & f_t = 1 \end{cases}$$



α 是任意选择的适应参数

$$B_1(x, y) = I_1(x, y)$$

$$B_t(x, y) = \alpha I_t(x, y) + (1 - \alpha)B_{t-1}(x, y)$$

$$D_t(x, y) = \begin{cases} 1 & |I_t(x, y) - B_t(x, y)| > T \\ 0 & \text{otherwise} \end{cases}$$

α 为自适应参数，其取值直接影响背景的更新质量

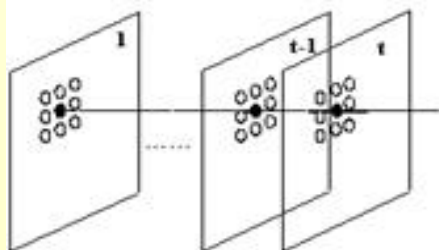
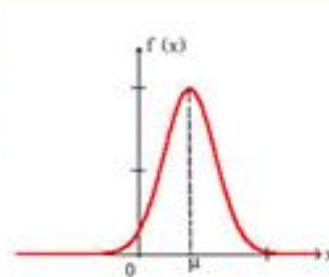
背景减法—单高斯法

- 对每一个像素利用高斯函数建模，每一像素点都认为服从均值和标准方差的分布，且每一点的高斯分布是独立的。

$$\{X_1, X_2, \dots, X_t\} = \{I(x, y, i) \mid 1 \leq i \leq t\}$$

t 表示第 t 帧，从时间轴上看，这些点的集合符合高斯分布，即，

$$P(X_t) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(X-\mu)^2}{2\sigma^2}}$$



单高斯模型：

假设每个像素的灰度在时间域上满足正态分布：
$$D(x, y) = \begin{cases} 1 & |I_t(x, y) - \mu(x, y)| > \lambda \delta_t(x, y) \\ 0 & \text{otherwise} \end{cases}$$

Default: $\lambda = 4.5$

$$\mu_t = \alpha \mu_{t-1} + (1 - \alpha) I_t$$

$$\delta_t^2 = \alpha \delta_{t-1}^2 + (1 - \alpha) (I_t - \mu_t)^2$$

□背景减法—单高斯法

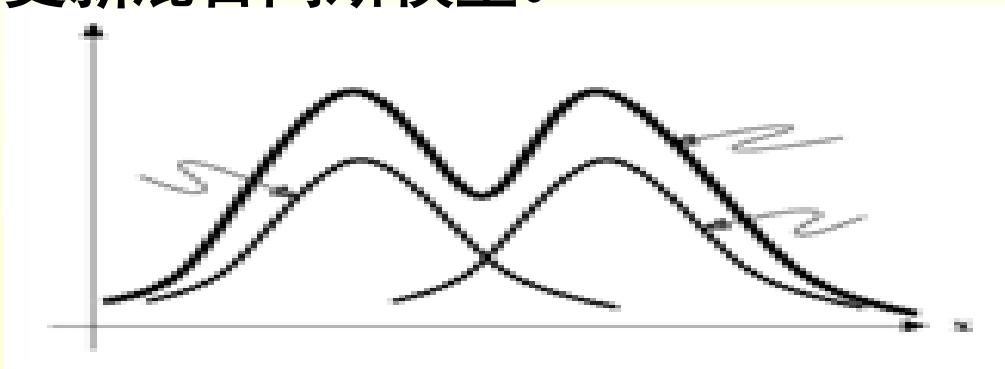


优点：单高斯分布背景模型在室内(或其他简单场景)进行运动目标检测可以得到较好的效果，由于运算量小，处理速度非常快，而且检测到的目标比较完整。

缺点：当场景比较复杂时，模型会变得不稳定，而且抗噪声干扰的能力较差。

□背景减法—混合高斯法

- 混合高斯模型使用K（基本为3到5个）个高斯模型来表征图像中各个像素点的特征；
- 用当前图像中的每个像素点与混合高斯模型匹配,如果成功则判定该点为背景点, 否则为前景点；
- 获得后更新混合高斯模型。



The parameters μ_i and σ_i for the matched distribution are adjusted by the following equations:

$$\mu_{k,t} = (1 - \eta_k) \mu_{k,t-1} + \eta_k x_t$$

$$\sigma_{k,t}^2 = (1 - \eta_k) \sigma_{k,t-1}^2 + \eta_k (x_t - \mu_{k,t})^T (x_t - \mu_{k,t})$$

□背景减法—混合高斯法

优点：

- (1) 混合高斯模型可以模拟复杂的多峰背景(如摇动的树枝, 摆动的旗帜等)
- (2) 不仅能准确的检测出大面积目标还能检测出小面积目标
- (3) 运算量不是非常大, 能够满足实时性需要

缺点：

- (1) 能够有效的解决光线渐变的问题, 但是对于光线突变非常敏感



(a) 原始图像

(b) 背景图像

(c) 单高斯法检测结果

(d) 混合高斯法检测结果

形态学处理：腐蚀、膨胀

□运动目标由于颜色和背景接近或者其他噪声的干扰，可能导致检测的结果出现空洞、孤立噪声点等现象。

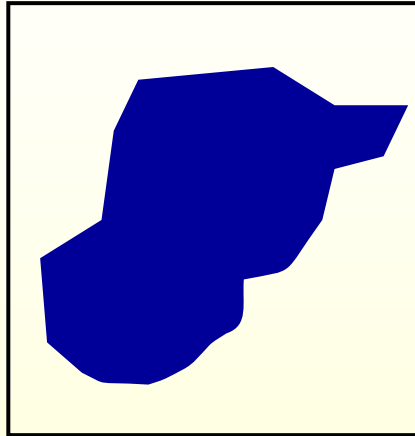
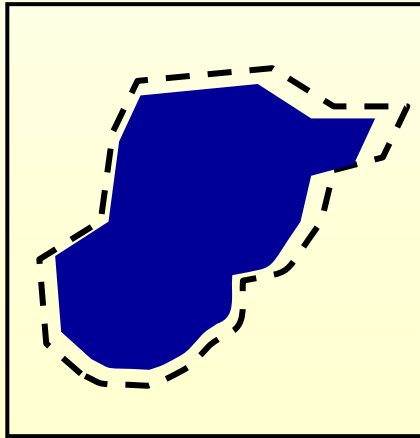
□基本思想：采用一些具有特定形态的结构元素提取图像中的对应形状、去除孤立噪声点，从而实现图像分析的目的。

连通域处理、孔洞填充

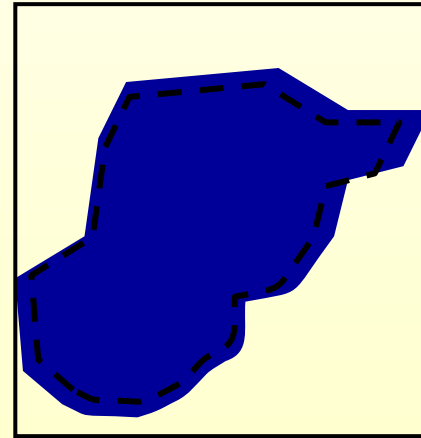
在使用形态学滤波处理完图像后，一些小的干扰区域已经被除去，小的间隙被连接上，小的孔洞被填充上，但是仍然会有相对较大的孔洞存在于检测目标内部。因此，我们使用连通性检测的方法进行处理，来去除检测到的运动目标内部的孔洞。

• 腐蚀与膨胀

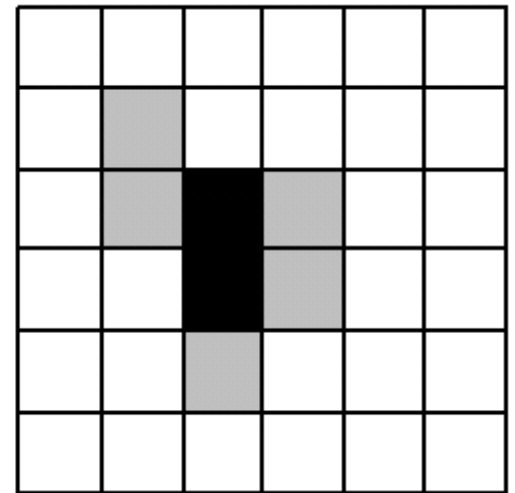
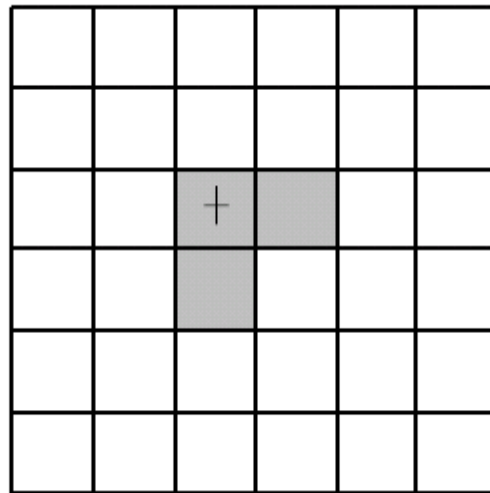
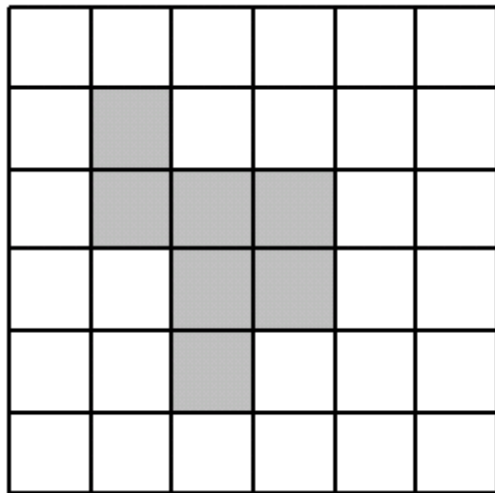
腐蚀



膨胀



腐蚀



静止背景下运动目标的检测实验结果：



(a) 第1帧图像



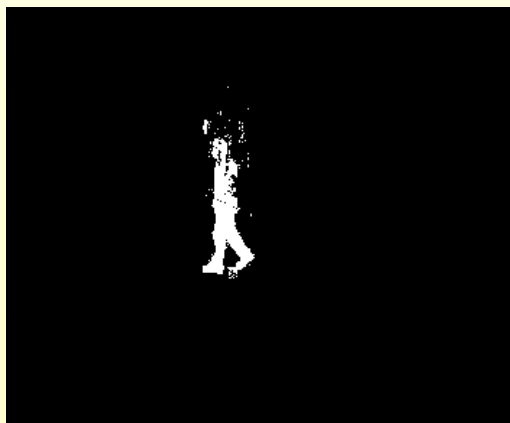
(b) 第2帧图像



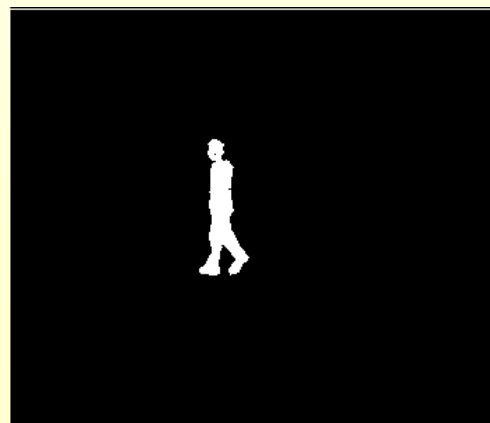
(c) 变化区域图像



(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

运动目标的检测

动态背景下的运动目标

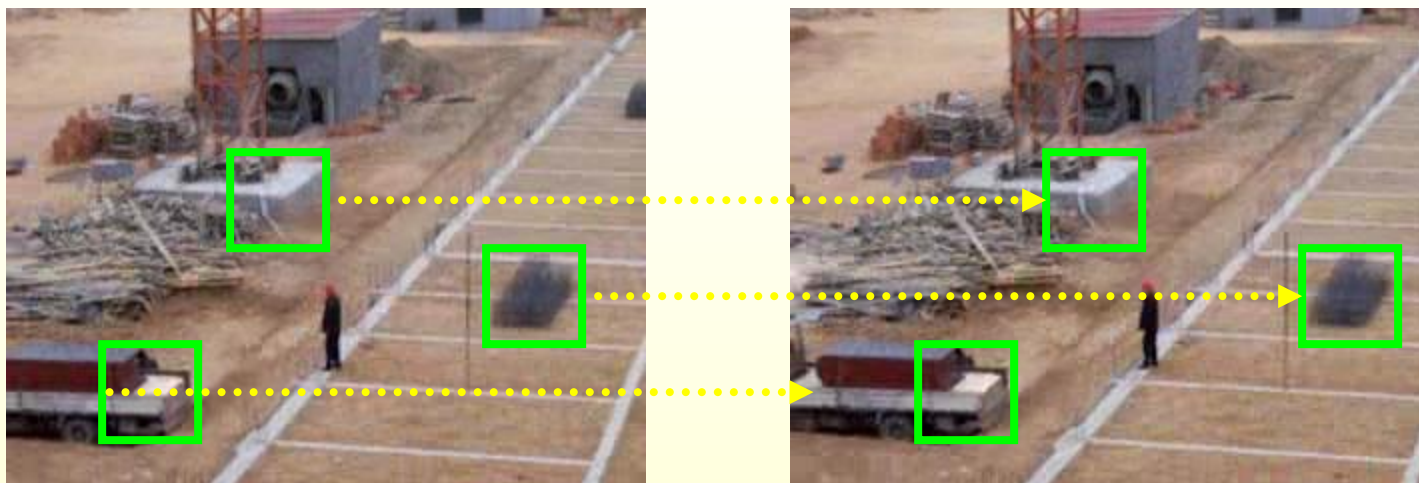
检测方法：在目标运动的同时，摄像机也由于运载平台的姿态或位置改变而发生运动

- 光流法
- 背景补偿法：块匹配法、贝叶斯法、像素递归法……

检测背景运动并补偿



块匹配补偿背景运动



子块划分

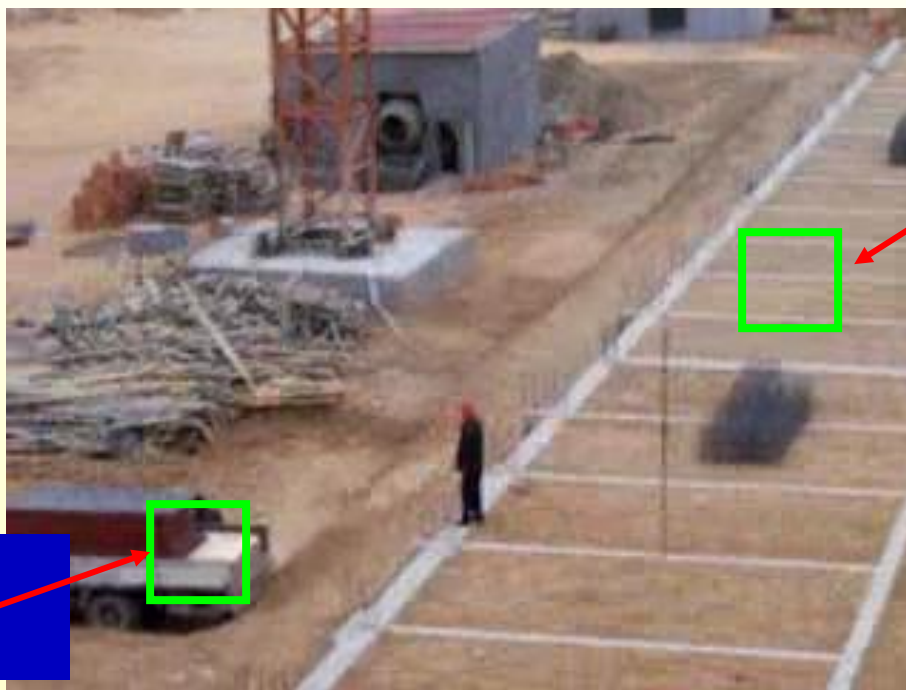


块匹配



背景运动补偿

- 特征丰富的区域



- Good

- Bad

背景运动补偿技术

背景运动补偿技术的核心问题是背景运动矢量的估计。

传统的运动估计采用平动来建模运动场（效果差）

目前，常采用具有丰富描述参数的**仿射变换**和**透射变换**来建模运动场



背景运动补偿技术 - 全局运动参数的鲁棒估计

最小二乘估计算法的缺陷：无法分辨和排除错误的特征点匹配。如果将这些错误的对应特征点代入最小二乘估计中，势必造成很大的偏差。

鲁棒参数估计(robust estimation)研究在观测值出现粗差(gross error, 即错误和异常)的情况下, 如何求得最优的参数估计。鲁棒估计既可以克服野点数据的影响, 又具备一般估计方法的统计特性, 是更广义的最优估计。

C.V. Stewart, “**Robust Parameter Estimation in Computer Vision**”
SIAM Rev., vol.41, no. 3, pp.513-537, 1999

背景运动补偿技术 - M估计

- 对于含有外点的数据，如果对所有样本点使用一样的权重，在拟合模型时外点对模型会有较大的干扰，由此为出发点想到，如果降低外点的权重，则可以降低外点对模型的影响，这也就是M估计的一个思想。
- 但问题是我们怎么知道哪些是外点呢，M估计中将与所估计模型偏差大的点视为外点，降低与模型偏差越大的点的权重。

背景运动补偿技术 - M估计

M估计是一种广义的最大似然估计方法。参数 a 的M估计定义为：

$$\hat{a} = \arg \min_a \sum_{x_i \in X} \rho(r_{i,a} / \sigma_i)$$

$\rho(u)$ 、 $u = r_{i,a} / \sigma_i$ 是一个鲁棒损失函数 (robust loss function) , 且是一个关于 $|u|$ 的单调非减函数, 常用的函数有Tukey, Cauchy, Huber函数等。 $r_{i,a}$ 是第 i 个数据相对于估计值 a 的偏差。可以将其定义为特征点经参数变换后的值与匹配点的距离, σ_i 为 $r_{i,a}$ 的标准方差。

对上式求解可得：

$$\sum_{x_i \in X} \psi(r_{i,a} / \sigma_i) \frac{dr_{i,a}}{da} \frac{1}{\sigma_i} = 0 \quad \psi(u) = \rho'(u)$$

背景运动补偿技术 - M估计

若引入权重函数: $\omega(u) \cdot u = \psi(u)$, 则有:

$$\sum_{x_i \in X} \omega(r_{i,a} / \sigma_i) \frac{1}{\sigma_i^2} \frac{dr_{i,a}}{da} r_{i,a} = 0$$

其实该式表示的就是**迭代加权最小二乘法**(Iterative Re-weighted Least Squares, IRLS)

该方法在每一步迭代中, 根据相对于当前估计值 a 的偏差和方差, 求每个数据的权重 $\omega_i = \omega(r_{i,a} / \sigma_i)$, 进而在下一步迭代中, 利用刚刚求得的权重值, 使用加权最小二乘法(Weighted Least Squares)得到一个新的估计值 a , 以及每个数据的偏差和方差。照此循环, 若干次迭代后IRLS收敛到最终的估计值 a 。 a 的初始值通常由最小二乘法估算。

背景运动补偿技术 - RANSAC算法

随机样本一致算法(Random Sample Consensus algorithm)
可以处理含有很高比例野点的情形。RANSAC通过随机选取一定量的样本，分别估计 a ，再从中选取一个最优的作为最终的估计值。RANSAC用内点个数来量化 a 的质量。

M. A. Fischler, R. C. Bolles, “Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography” Comm. of the ACM, Vol 24, pp 381-395, 1981

背景运动补偿技术 - RANSAC算法

寻找一条直线，使得它可以拟合一个2维的点集。

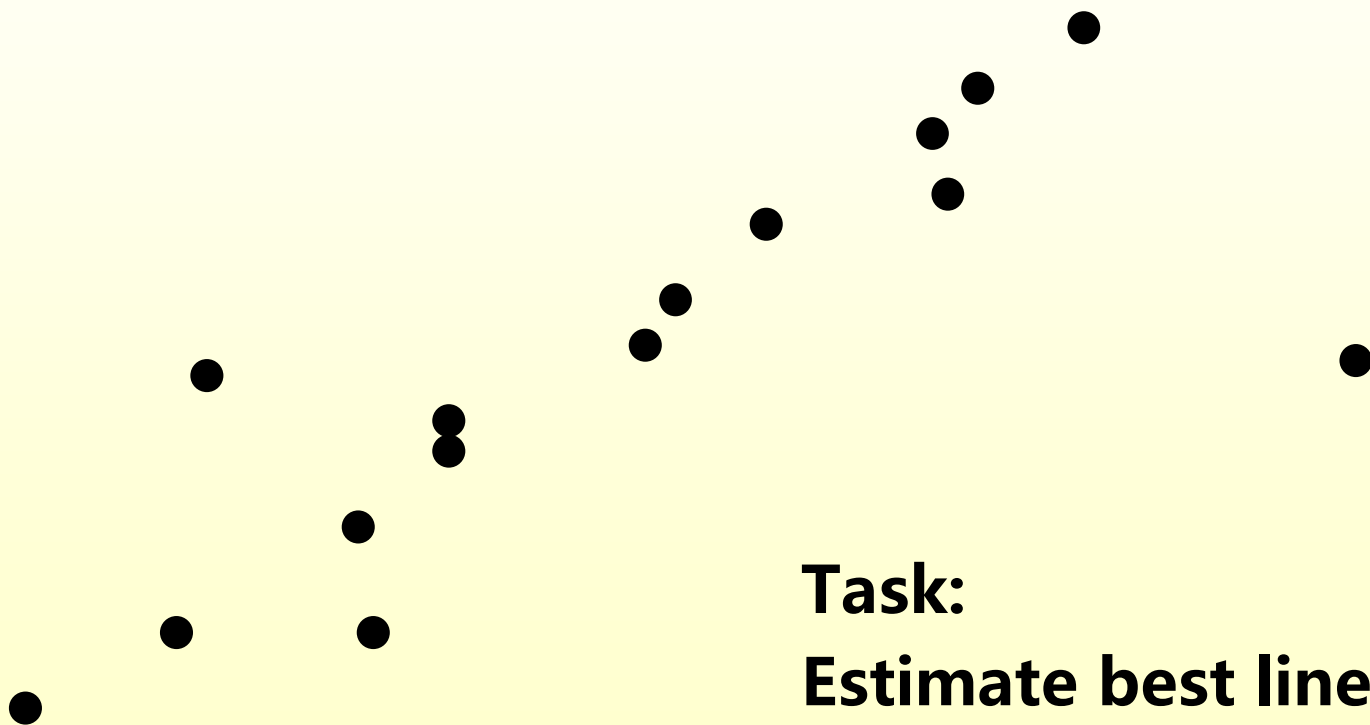
这可以进一步叙述成：寻找一个1维仿射变换 $y=ax+b$ ，拟合一组2维平面中的点。

两个问题：1.一条拟合数据的直线；2.对有效点和无效点的分类。

RANSAC算法：一个在无效点比例较大的情况下仍然有效的估计算法。

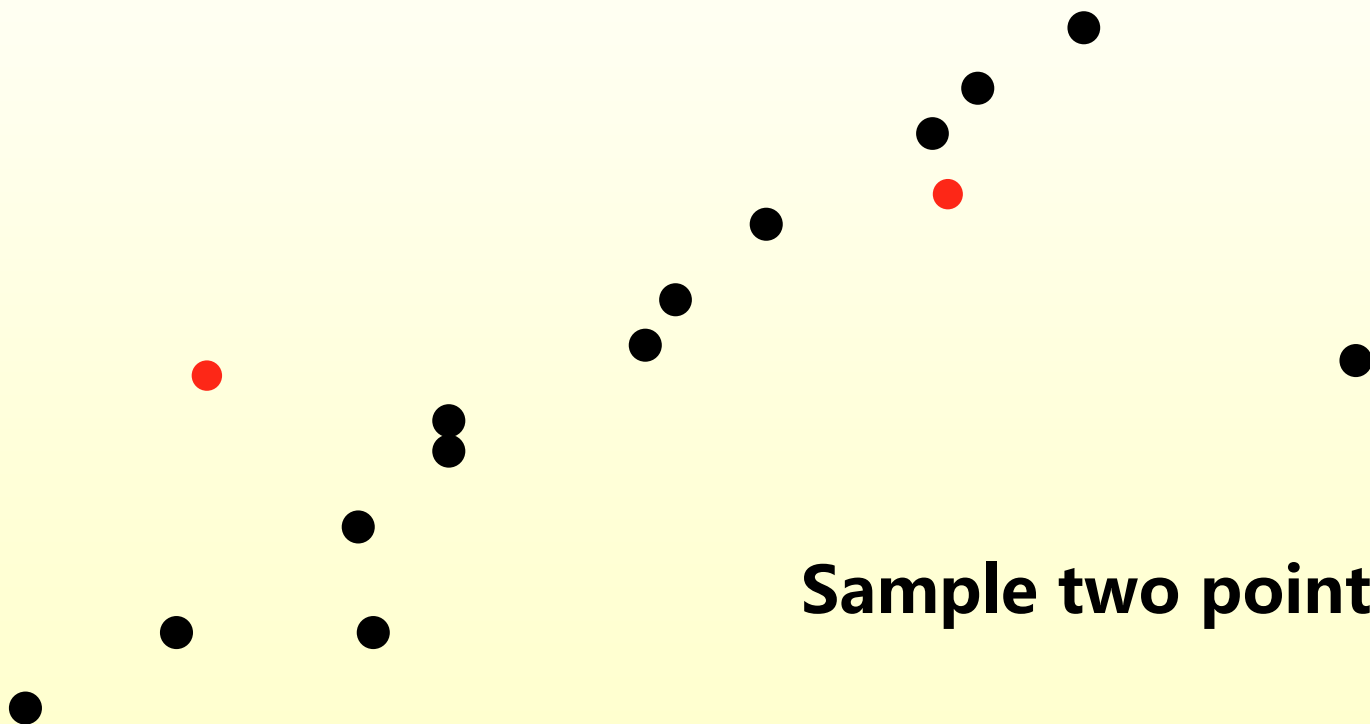
背景运动补偿技术 - RANSAC算法

RANSAC算法的思想：首先随机地选择两个点，这两个点确定了一条直线。我们称直线的支撑为在这条直线的一定距离范围内的点的数目。这样的随机选择重复数次，然后具有最大支撑的直线被确认为是点集的拟合。在拟合的误差距离范围内的点被认为是有效点，它们构成所谓的一致集，反之则为无效点。



背景运动补偿技术 - RANSAC算法

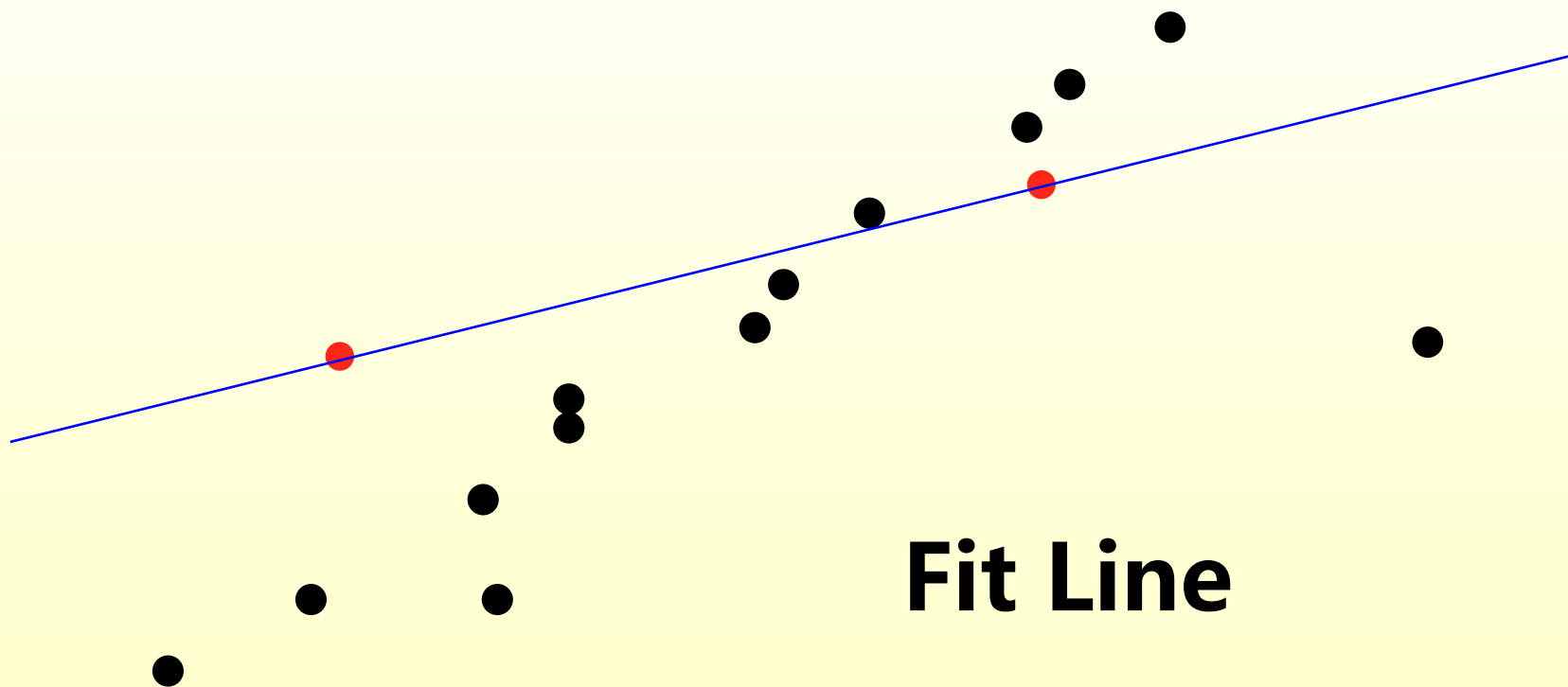
RANSAC算法的思想：**首先随机地选择两个点**，这两个点确定了一条直线。我们称直线的支撑为在这条直线的一定距离范围内的点的数目。这样的随机选择重复数次，然后具有最大支撑的直线被确认为是点集的拟合。在拟合的误差距离范围内的点被认为是有效点，它们构成所谓的一致集，反之则为无效点。



Sample two points

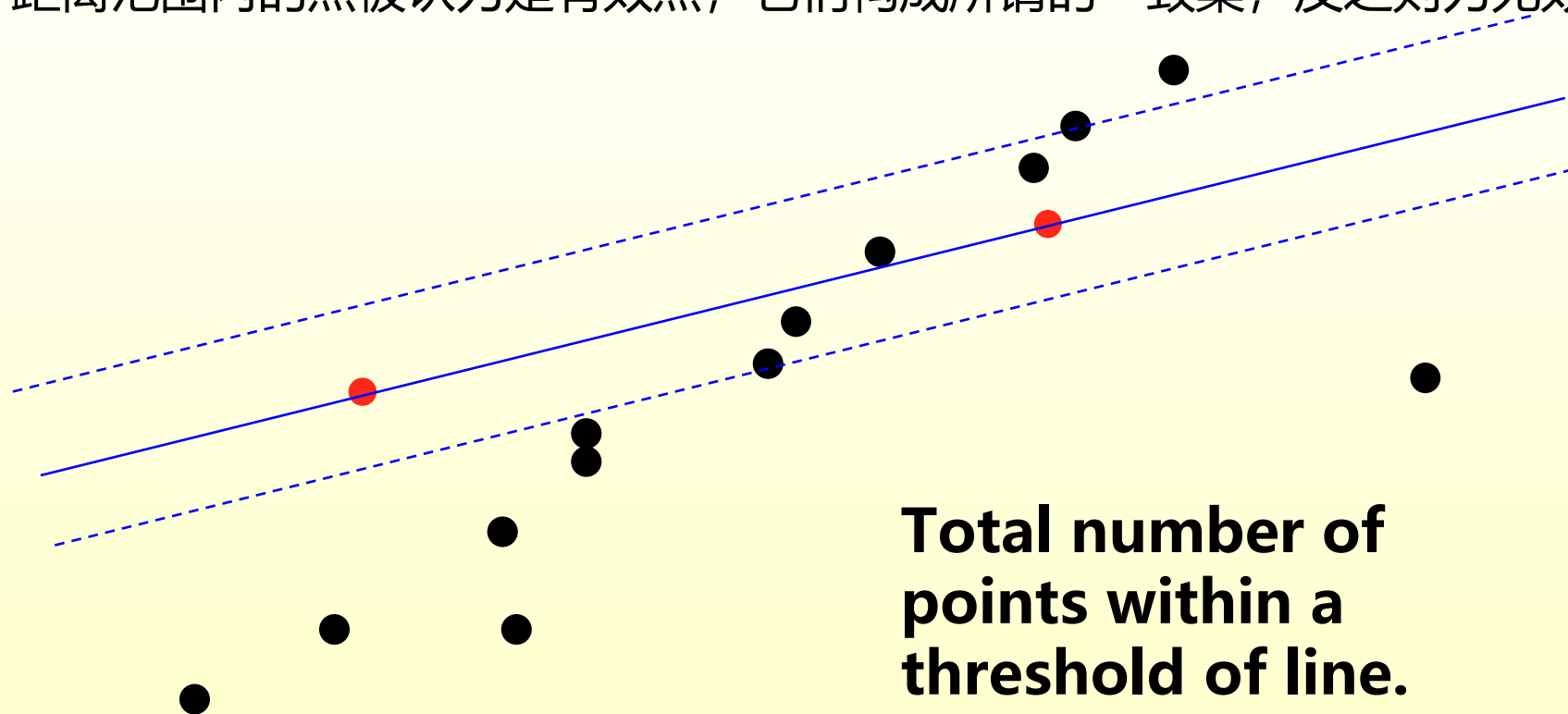
背景运动补偿技术 - RANSAC算法

RANSAC算法的思想：首先随机地选择两个点，**这两个点确定了一条直线**。我们称直线的支撑为在这条直线的一定距离范围内的点的数目。这样的随机选择重复数次，然后具有最大支撑的直线被确认为是点集的拟合。在拟合的误差距离范围内的点被认为是有效点，它们构成所谓的一致集，反之则为无效点。



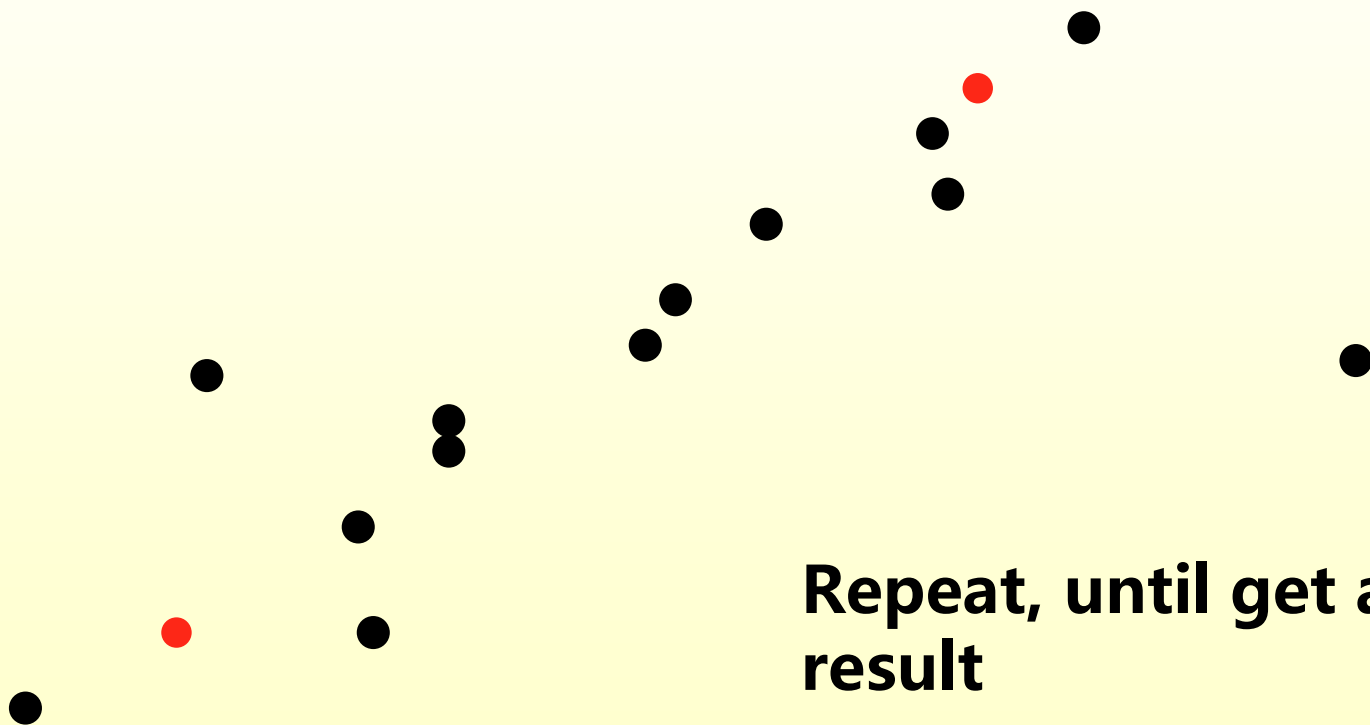
背景运动补偿技术 - RANSAC算法

RANSAC算法的思想：首先随机地选择两个点，这两个点确定了一条直线。
我们称直线的支撑为在这条直线的一定距离范围内的点的数目。这样的随机选择重复数次，然后具有最大支撑的直线被确认为是点集的拟合。在拟合的误差距离范围内的点被认为是有效点，它们构成所谓的一致集，反之则为无效点。



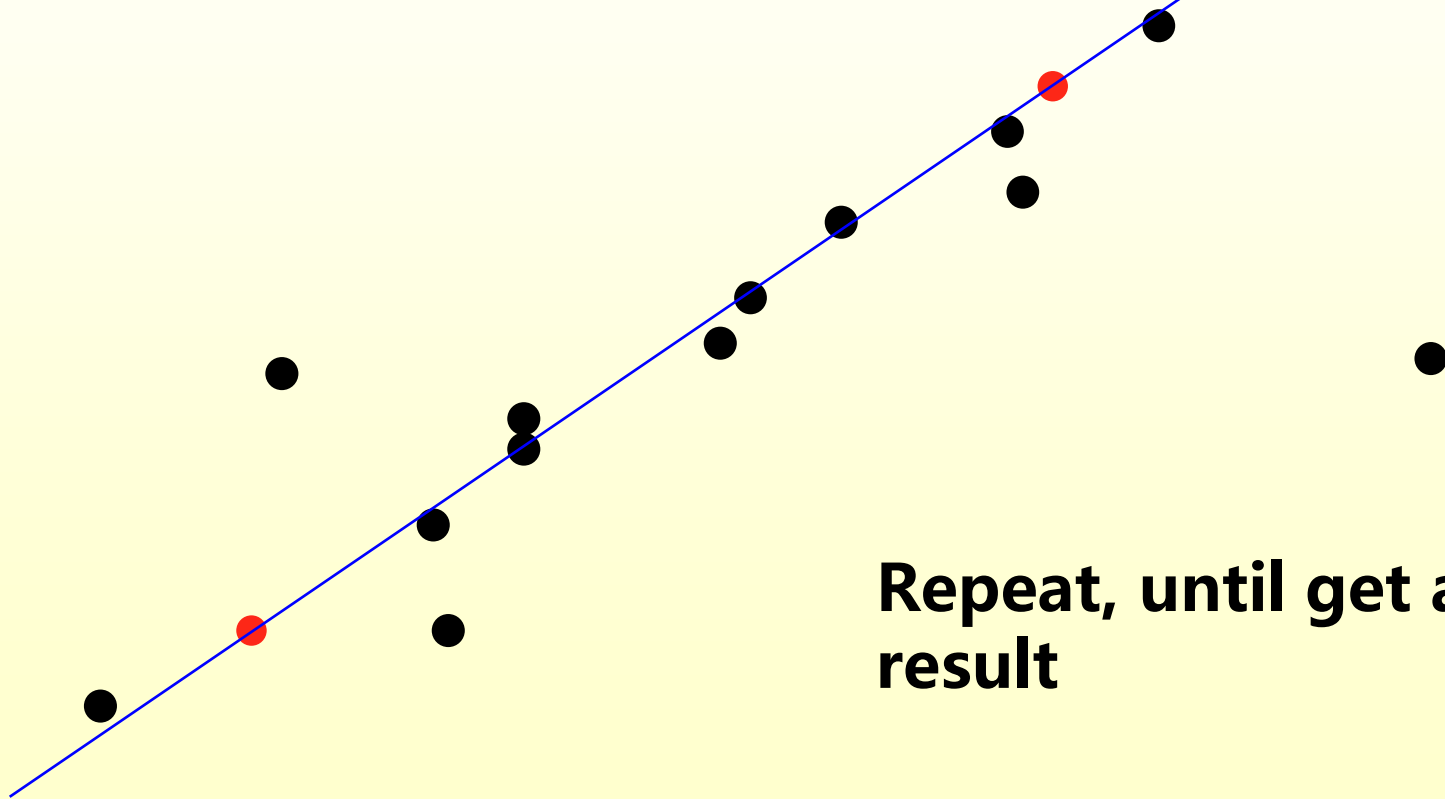
背景运动补偿技术 - RANSAC算法

RANSAC算法的思想：首先随机地选择两个点，这两个点确定了一条直线。我们称直线的支撑为在这条直线的一定距离范围内的点的数目。**这样的随机选择重复数次，然后具有最大支撑的直线被确认为是点集的拟合。**在拟合的误差距离范围内的点被认为是有效点，它们构成所谓的一致集，反之则为无效点。



背景运动补偿技术 - RANSAC算法

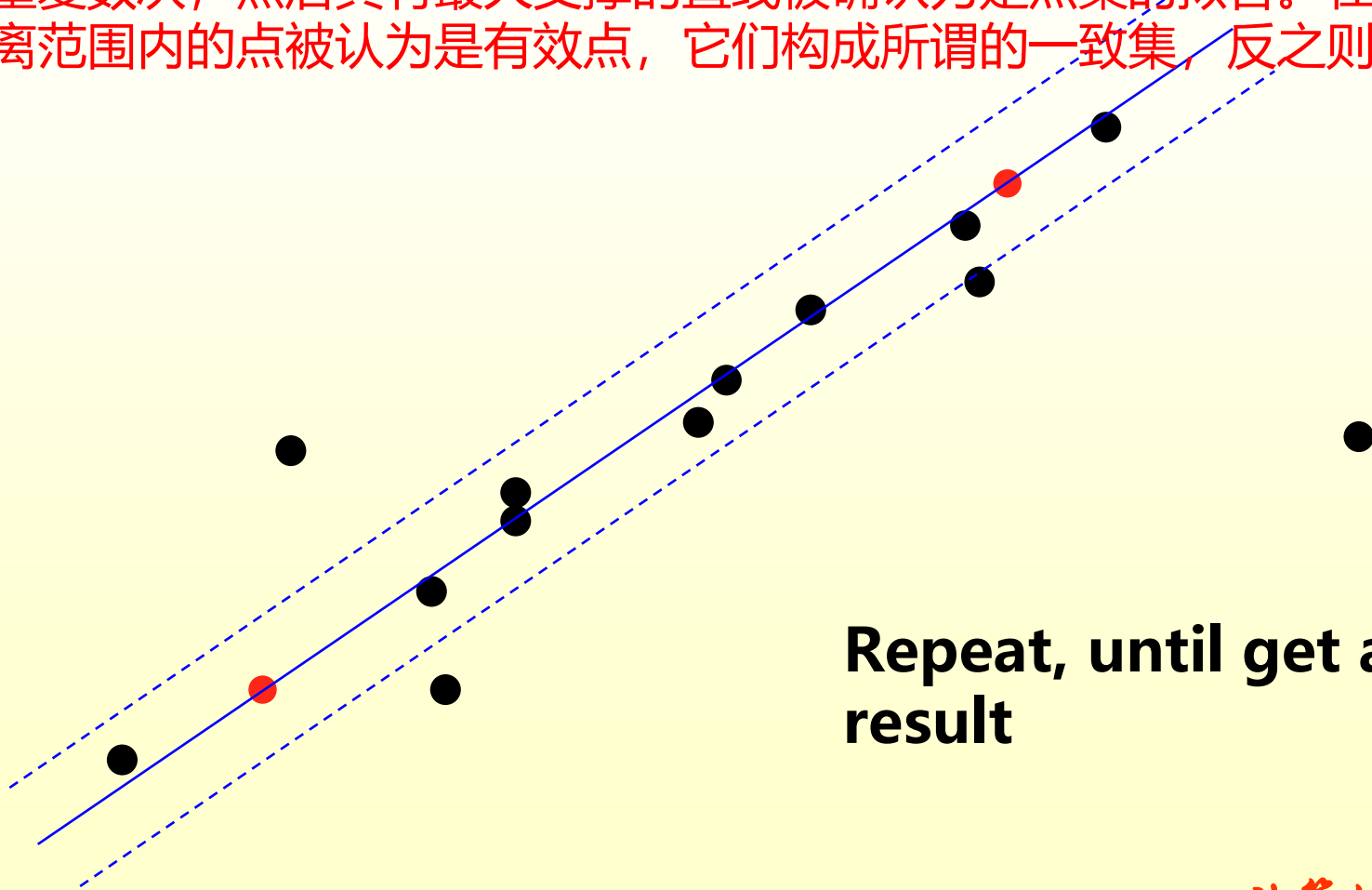
RANSAC算法的思想：首先随机地选择两个点，这两个点确定了一条直线。我们称直线的支撑为在这条直线的一定距离范围内的点的数目。**这样的随机选择重复数次，然后具有最大支撑的直线被确认为是点集的拟合。**在拟合的误差距离范围内的点被认为是有效点，它们构成所谓的一致集，反之则为无效点。



Repeat, until get a good result

背景运动补偿技术 - RANSAC算法

RANSAC算法的思想：首先随机地选择两个点，这两个点确定了一条直线。我们称直线的支撑为在这条直线的一定距离范围内的点的数目。这样的随机选择重复数次，然后具有最大支撑的直线被确认为是点集的拟合。在拟合的误差距离范围内的点被认为是有效点，它们构成所谓的一致集，反之则为无效点。



Repeat, until get a good result

目录

- 1、引言**
- 2、运动目标检测**
- 3、运动目标跟踪**

抓到羚羊需要几步？

• 猎手之眼

- 广阔的草原上，一只落单的羚羊正在进食，它长在两侧的双眼警惕的观察着环境。
- 草食动物的双眼一般长在头部两侧，双眼各自观测一片区域，这样可以带来更大视野范围，以便减小视野的盲区，预防可能的袭击——来自猎手的偷袭。
- 然而羚羊并没有发现，在它不远处的石堆后，一只猎豹伏低了身子，正在悄无声息的靠近。
- 与草食动物不同，肉食动物的双眼往往集中在面部前方，双眼同时观测同一区域，以便获得立体视觉，从而测量出目标的距离。
- 猎豹通过双目视觉测量出目标大概在十米外，前方五米处有一堆石块，猎豹推算好了目标距离和袭击路线，站定，呼吸减缓，身姿伏低，肌肉绷紧，双眼紧盯，大自然花费数十亿年打造的这台杀戮机器现在犹如一架绞满的弩，只需要一个扣动扳机的时机？
- 时机很快到了，抬头张望了一圈的羚羊仍然没有察觉到危机，它重又低下头，准备进食，在它低头的那一刻？

• 扳机扣响！

- 蓄势已久的猎豹如离弦之箭射了出去，快到石堆的时候猎豹敏捷地拐了一个弯绕过障碍，目标一直牢牢的锁定在视野范围中，距离随着剧烈的心跳越来越近，越来越近？
- 上面这个猎手与猎物的故事就是目标追踪最简略的原理，实现这样的目标追踪的效果，大自然用了几十亿年的时间来进化打磨。

抓到羚羊需要几步？

- 1, 从茫茫无垠的草原上发现羚羊并牢牢盯住。——**检测识别**
- 2, 观测羚羊周围的环境与路线上的障碍。——**识别避障**
- 3, 思考行动路径以及奔跑!——**跟踪、路径规划**

无人机跟踪进化之路

空中机器人（无人机）上的目标追踪功能，经历的进化历程：



其实不同的外表之下，都有着一颗小小的GPS接收机。

1、GPS：Lily无人机，众筹的Airdog，甚至老牌的无人机厂商3DR。

必须要满足某种前提条件：
携带某种设备，这种设备或者是手环，或者是“魔棒”，或者是遥控器等。

所说的抓到羚羊的3步，使用这种方式的目标跟踪，事实上跳过1，2步，直接控制无人机往目标飞行就可以了。

无人机跟踪进化之路

如果还是使用猎豹和羚羊的故事举例的话，使用GPS进行目标跟踪的原理大概就是：

猎豹瞎了，羚羊一副生无可恋的样子大声朝猎豹喊：你来抓我呀你来抓我呀我在你的前方偏右32度，距离12.6米！你有本事你来抓我呀抓我呀！

然后猎豹就循着声音所说的方向和距离猛地冲向了目标，然后吧唧一声，撞死在了岩石上？

Sad story.

GPS最致命的问题：

- ❑ 没有办法感知环境，也就没有办法知道前面有障碍物。
- ❑ 需要被跟踪目标配合，安装GPS设备。
- ❑ GPS定位精度较低（2.5米？），并且更新频率较低，存在干扰。故而无无人机并不知道目标所在的确切位置，可能目标难以出现在画面中，或者无人机拍摄的位置距离目标非常远。

无人机跟踪进化之路

2、机器视觉

猎豹捕猎的过程如果由机器视觉来完成这个任务：

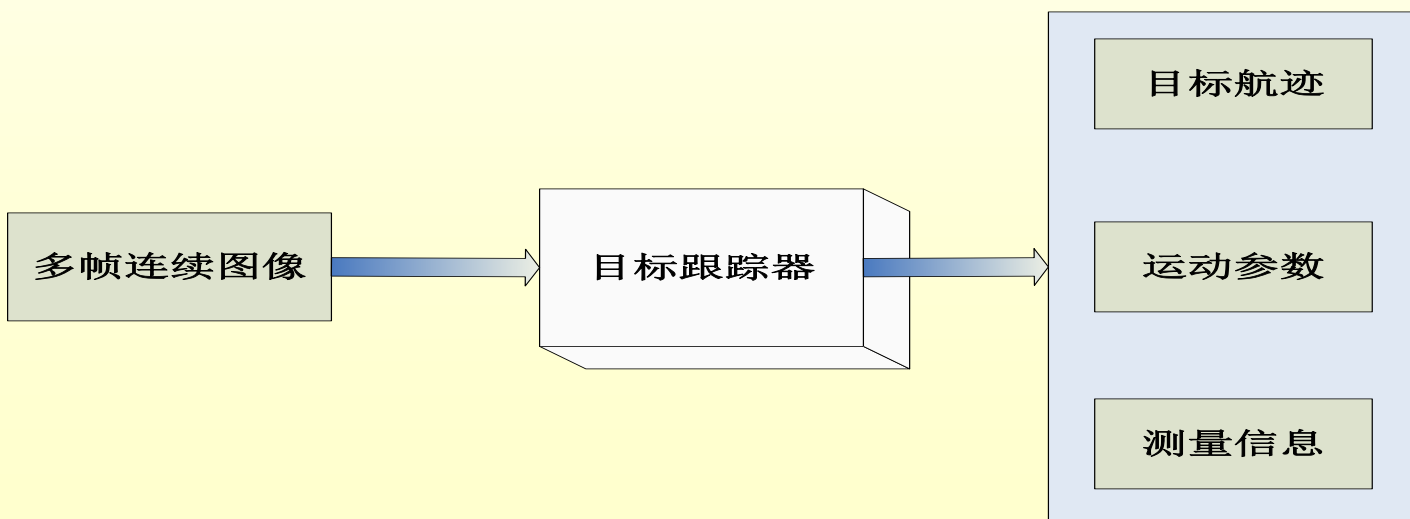
- 首先解决的是对环境的感知(双目观测到前方有石堆)——**检测**
- 其次需要知道自己的速度与位置，(观测到自己周围环境在视野中后退，得到自己前进的速度)——**运动估计**
- 然后需要从复杂的背景中将目标稳定准确的检测出来，(羚羊尝试发现披着伪装色的猎豹，可惜失败了，不过猎豹做到了。)——**检测**
- 最后在视觉信息的基础上，飞行器进行航迹规划，控制飞行器以平滑的运动轨迹躲避障碍并跟上目标。(猎豹躲避障碍并追上目标)——**跟踪**



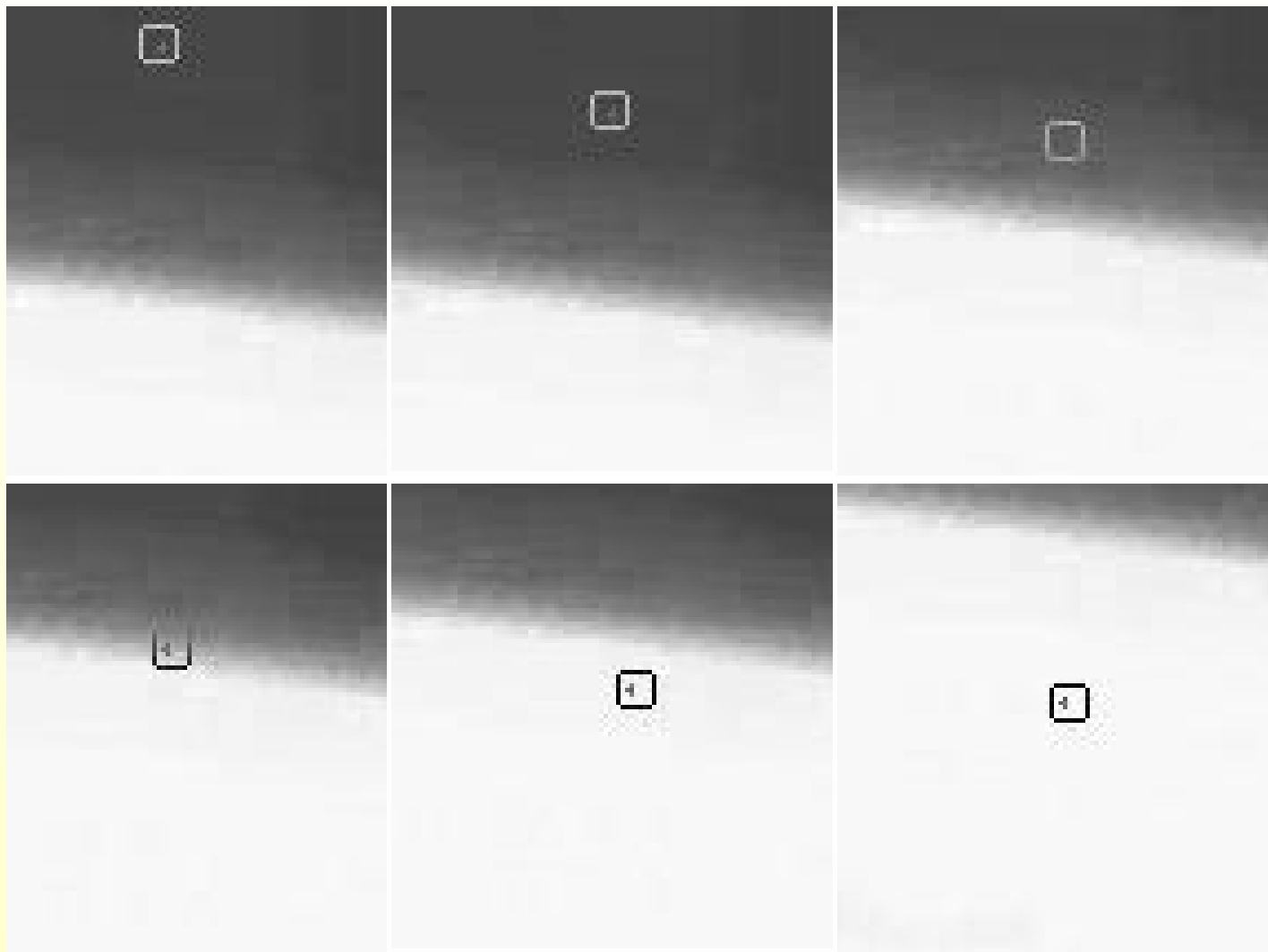
目标跟踪方法涉及的基本问题

目标跟踪的定义：

- 所谓目标跟踪，可以简单地定义为对连续的视频序列中的目标维持一条**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**。



实例1：弱小点目标的跟踪实例



弱小点目标跟踪实例

实例2：F-22战斗机的目标跟踪系统



实例3：直升机、无人机着降光电引导技术

■ 舰载直升机在反潜、救生、登陆、布雷与扫雷、火力校正、预警、侦察中广泛应用。

■ 由于中型舰船甲板较小，因风浪使甲板处于不规则的运动，直升机着舰的事故高。



跟踪识别的难点

- 视频图像呈现出来的是有限分辨率的数值矩阵，且像素值在不停变化。光照变化、阴影变换、物体重叠、遮挡、形变、旋转、角度变换，任何一个都可以让这些信总变得面目全非。
- 反应要快！

Challenges 3: occlusion



Magritte, 1937

无人机视频中目标检测跟踪的难点

- 平台运动、背景变化大且干扰大、目标尺寸不定
- 不同视角下目标表观差异大
- 远距离观察下目标弱小、模糊
- 实时性
- 机载嵌入式实现

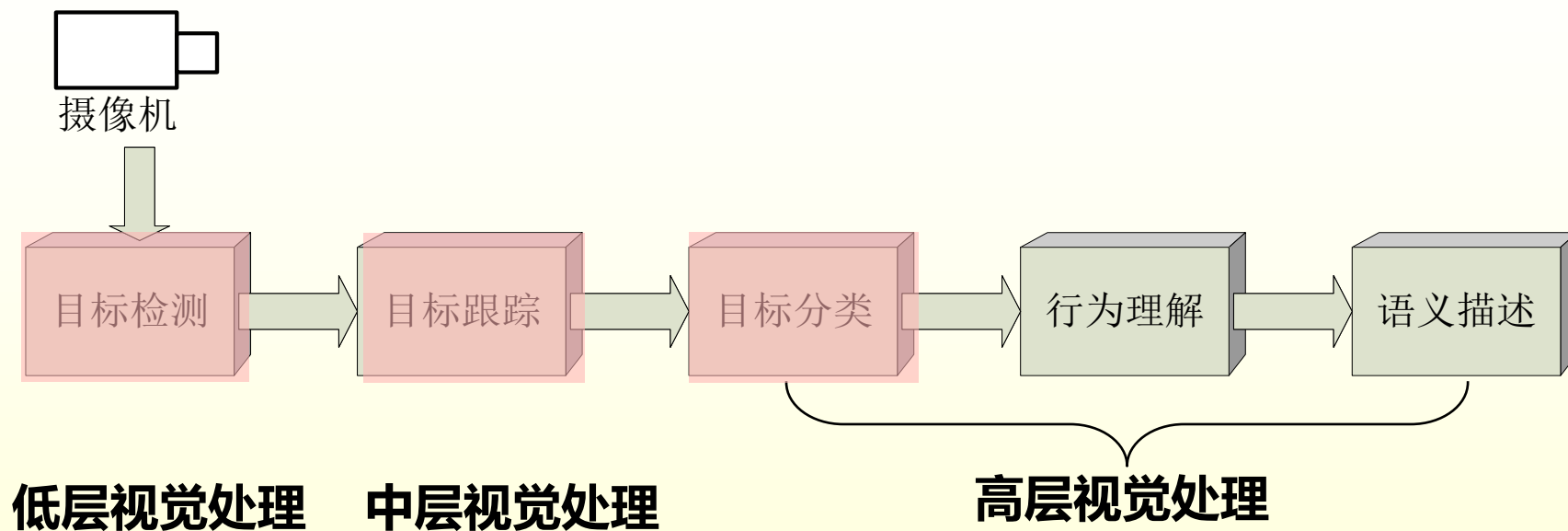
技术现状(国际)

时间	事件	机构
1974	智能实时电视跟踪系统，以提高光电跟踪测量仪器的性能，并适应多目标跟踪测量的需要。	美国白沙导弹靶场、新墨西哥州立大学和亚利桑那大学科学中心
1980s	为空军研制的机载火控系统的多功能红外相干光学传感器，具有多目标捕获、跟踪、分类、武器投掷和地形回避等功能。	美国福特公司和沃特公司
1985	掩护船防御系统能跟踪处理200多个目标。	美国海军
1989-90	公布了宽视场望远镜实时跟踪系统(WFOV)和多目标跟踪宽视场摄像机系统。	美国Larrence Livermore国家实验室
1991	红外图像识别跟踪系统，采用灰度、形状等多个特征，其Exoatmospheric Re-Entry Interceptor Subsystem(ERIS)拦截器可快速识别真假来袭的导弹。	美国陆军

技术现状(国际)

时间	事件	机构
1992	提出的“有识别能力的拦截弹”也是一种具有智能的对多特征进行信息融合处理的识别跟踪系统，它要求实现边识别边跟踪的功能。	美国国防部
1997	视觉监控(Visual Surveillance and Monitoring ,VSAM)重大项目，主要研究用于战场及普通民用场景监控的自动视频理解技术。	美国国防部，联合卡内基梅隆大学、麻省理工学院等
1999	设计了一套航拍视频检测与持续跟踪系统，该系统能够对多运动目标实现长时间的准确跟踪，即使发生短时间内目标被遮挡或目标时静时动的情况。	美国康奈尔大学计算机系
2005	开发出了基于MATLAB的COCOA系统，用于无人机低空航拍视频图像的目标检测与跟踪处理。	美国中佛罗里达大学计算机视觉实验室

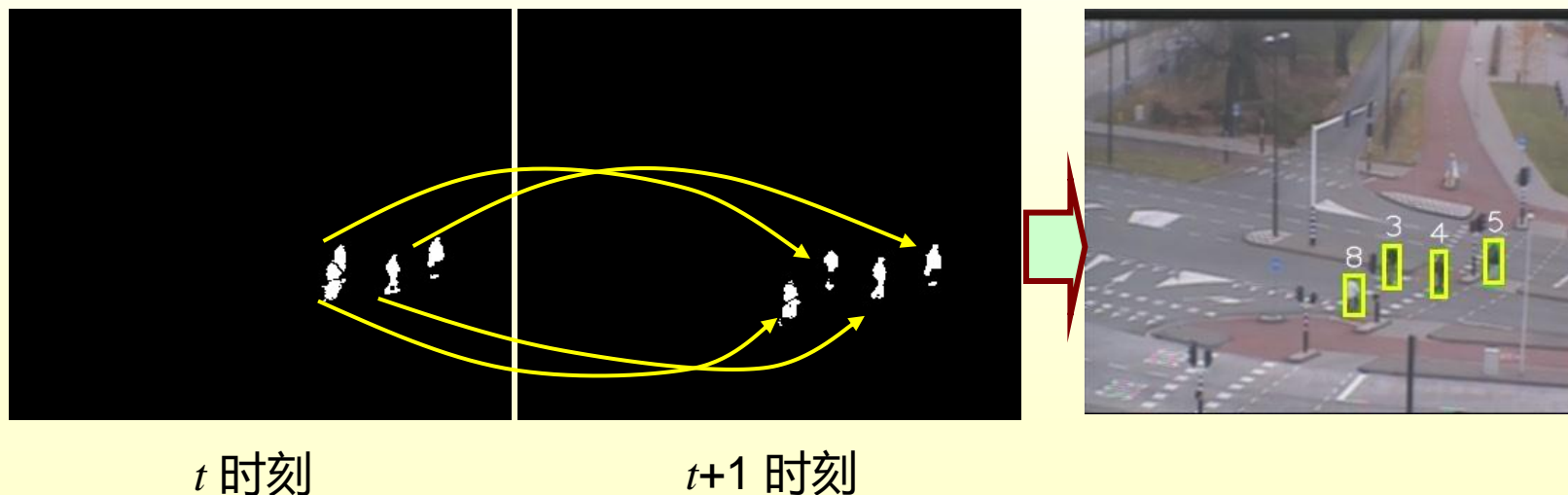




目标检测与跟踪的关系

① 先检测后跟踪(DBT)

先检测出每帧中的运动目标，然后匹配前后帧中的目标以实现轨迹关联。

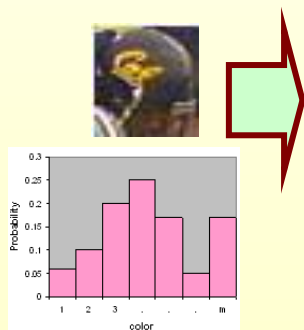


目标检测与跟踪的关系

② 先跟踪后检测(TBD)

将目标的检测与跟踪相结合，利用跟踪结果来确定检测所要处理的区域范围；跟踪时，则利用检测来获得目标状态的观测。

首先建立描述目标的特征模型，在起始帧初始化后，不断在后续帧进行匹配搜索。



(a)

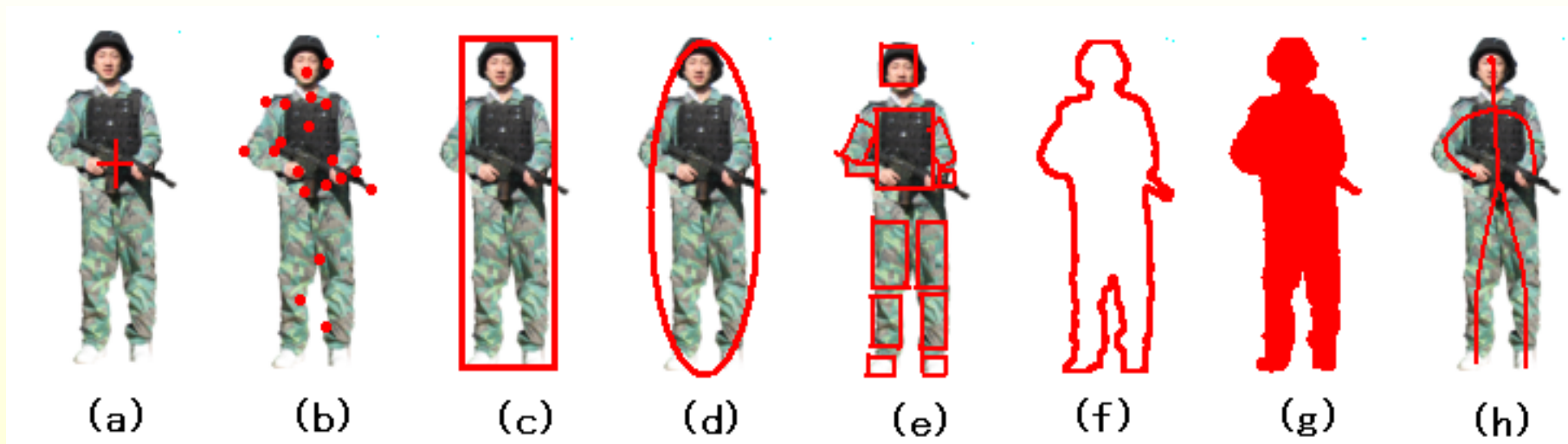
(b)

(c)

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?

目标跟踪的表示方法

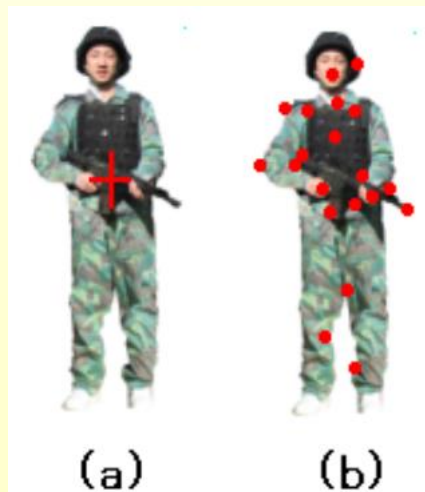


(a)质心表示, (b)特征点集表示, (c)矩形框表示, (d)椭圆框表示, (e)关节模型表示, (f) 轮廓表示, (g) 侧影表示, (h) 骨架模型表示

目标跟踪的表示方法

➤ 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。



目标跟踪的表示方法

➤ 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：

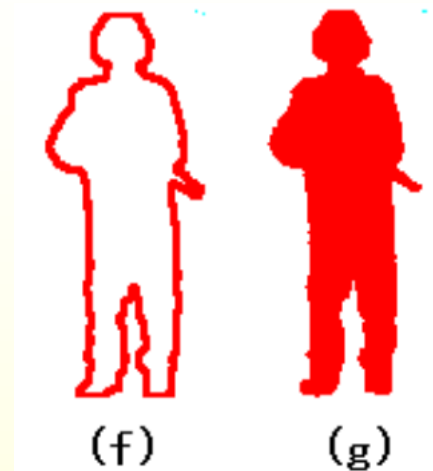
当目标可以认为是由各个部位通过结点衔接到一起的时候，就可以用关节模型来表示（图e），而各个部位则可以用几何形状来表示。



目标跟踪的表示方法

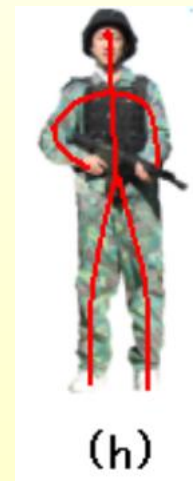
➤ 轮廓表示：

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。



目标特征的选择

颜色特征：

颜色特征是一种全局特征，描述了目标的表面性质。一般颜色特征是基于像素点的特征，此时所有属于目标区域的像素都有各自的贡献。

- **优点：**颜色直方图是最常用的表达颜色特征的方法，不受图像旋转和平移变化的影响，进一步借助归一化还可不受尺度变化的影响。
- **缺点：**对光照变化非常敏感。**没有表达出颜色空间分布的信息**，而且由于颜色对图像或图像区域的方向、大小等变化不敏感，所以不能很好地捕捉目标的局部特征。
- **常用方法：**主要的颜色特征包括**颜色直方图**、**颜色矩**、**颜色聚合向量**等。

目标特征的选择

边缘特征：

目标边缘通常代表强度变化较大的像素点。

- **优点：** 边缘特征的抗光照变化能力强。
- **缺点：** 具有较大形变的目标的边缘特征不稳定；边缘特征空间的相似性与人视觉系统感受到的相似性往往有差别；而且，从2D图像中反映出来的形状常不是3D物体真实的形状；由于视点的变化可能会产生各种失真等等，这些问题还需要进一步研究解决。
- **常用方法：** Canny算子、Sobel算子。

目标特征的选择

纹理特征：

与颜色特征不同，纹理特征不是基于像素点的特征，它需要一个额外的在包含多个像素点的区域中进行统计计算的过程。在模式匹配中，这种区域性的特征具有较大的优越性，不会由于局部的偏差而无法匹配成功。

- **优点：**作为一种统计特征，纹理特征常具有旋转不变性，并且对于噪声有较强的抵抗能力。
- **缺点：**当图像的分辨率变化的时候，所计算出来的纹理可能会有较大偏差；另外，由于有可能受到光照、反射情况的影响，从2D图像中反映出来的纹理不一定是3D物体表面的真实纹理。
- **常用方法：**HOG、共生矩阵、SIFT(Scale-Invariant Feature Transform)算子、小波特征等。

目标特征的选择

光流特征：

光流特征是一个区域内的像素帧间偏移量的密集场，它是基于亮度恒常这个基本假设的。

- **优点：**对基于运动的跟踪效果好。
- **缺点：**计算速度慢、对光照变化敏感的缺点。
- **常用方法：**Lucas-Kanada法等。

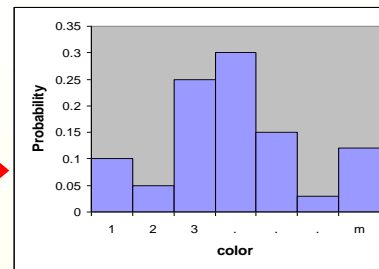
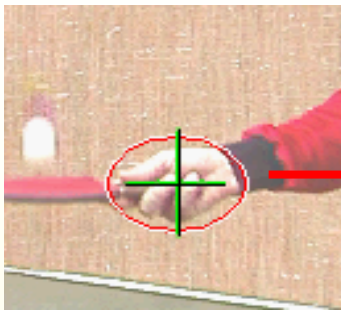
目标特征的选择

- ❖ 从不同的特征角度：颜色特征、边缘特征、纹理特征……
- ❖ 从不同的描述角度
 - ❖ 点、线等几何形状特征
 - ❖ 不变矩特征
 - ❖ 幅度、直方图、频率系数
 - ❖ 拓扑结构特征
 - ❖ ……

常见目标表示

■ 颜色直方图

忽略了目标的空间结构信息。



■ Snake轮廓模型

仅考虑目标的边界，在目标与背景容易混淆的情况下，目标模型通常容易退化。

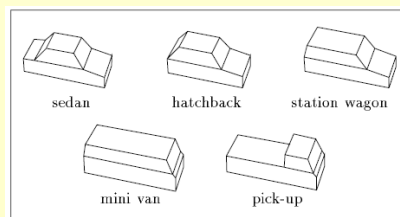
■ 局部特征模型

一组局部特征来表示，缺乏目标的全局结构信息，稳定性较弱，易受噪声、目标姿态以及光照条件的影响。但是对遮挡等有鲁棒性。



■ 2D/3D空间模型

模型参数估计的运算量很大，并且难以保证模型的精度。



图像匹配技术

- 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。

成像目标跟踪技术

■ 常见的目标跟踪方法

- 波门跟踪；
- 边缘跟踪、峰值（对比度）跟踪等；
- 质心跟踪；
- 相关跟踪（模板匹配）；
- 轨迹预测和记忆跟踪算法（Kalman、粒子滤波等）；
- 多目标跟踪及编号维持；
- 基于模式识别及特征分类的目标跟踪
—— 在线、离线学习的跟踪方法。

目标跟踪方法

1. 质心跟踪

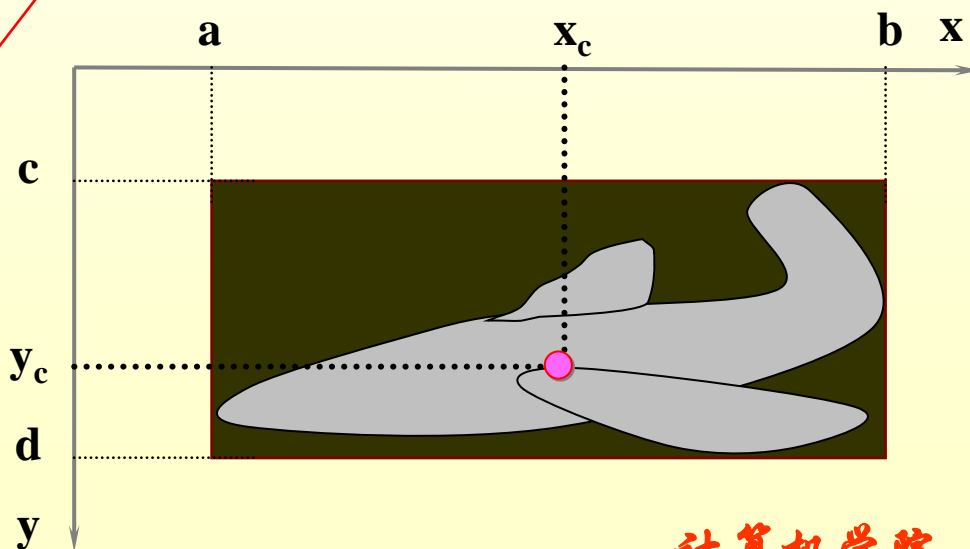
矩心也叫质心或重心，是物体对某轴的静力矩作用中心。如果把目标图像看成是一块质量密度不均匀的薄板，以图像上各像素点的灰度作为各点的质量密度。可以借用矩心的定义式来计算目标图像的矩心。

$$\bar{X} = \frac{\sum_{x=0}^{N-1} \sum_{y=0}^{M-1} xf(x, y)}{\sum_{x=0}^{N-1} \sum_{y=0}^{M-1} f(x, y)}$$

$$\bar{Y} = \frac{\sum_{x=0}^{N-1} \sum_{y=0}^{M-1} yf(x, y)}{\sum_{x=0}^{N-1} \sum_{y=0}^{M-1} f(x, y)}$$

形心

$$\begin{cases} x_c = \frac{1}{MN} \sum_{y=0}^{N-1} \sum_{x=0}^{M-1} x \\ y_c = \frac{1}{MN} \sum_{y=0}^{N-1} \sum_{x=0}^{M-1} y \end{cases}$$



目标跟踪方法

2. 相关跟踪

由于目标运动、姿态发生改变、光照条件改变以及杂波背景的干扰，使得目标图像的分割提取十分困难，计算目标的矩心或形心不准确。在某种情况下，可以采用以图像匹配为基础的跟踪方法，习惯上称之为相关跟踪。

Template



Search image

目标跟踪方法

3、轨迹预测和记忆跟踪算法

Kalman滤波器

利用反馈控制系统估计运动状态，用当前的状态和误差协方差估计下一时刻的状态。

粒子滤波

随机抛洒粒子，利用后验概率确定目标跟踪位置。

目标跟踪方法

■ 问题提出：

线性、高斯系统



卡尔曼滤波 (KF)

非线性、高斯系统



扩展卡尔曼滤波 (EKF)

无极卡尔曼滤波 (UKF)

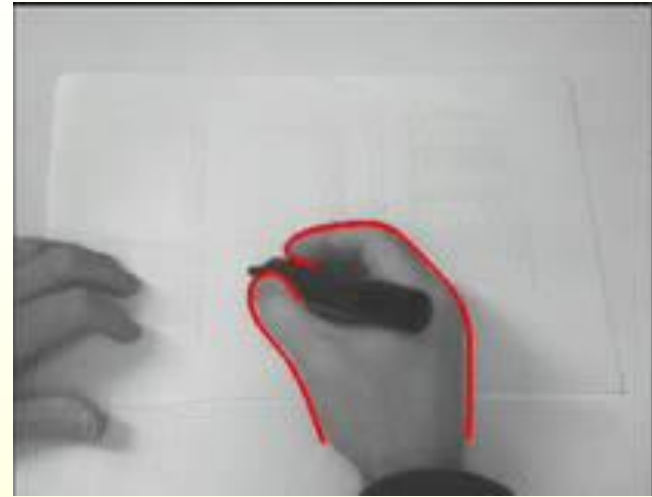
非线性、非高斯系统



粒子滤波 (PF)

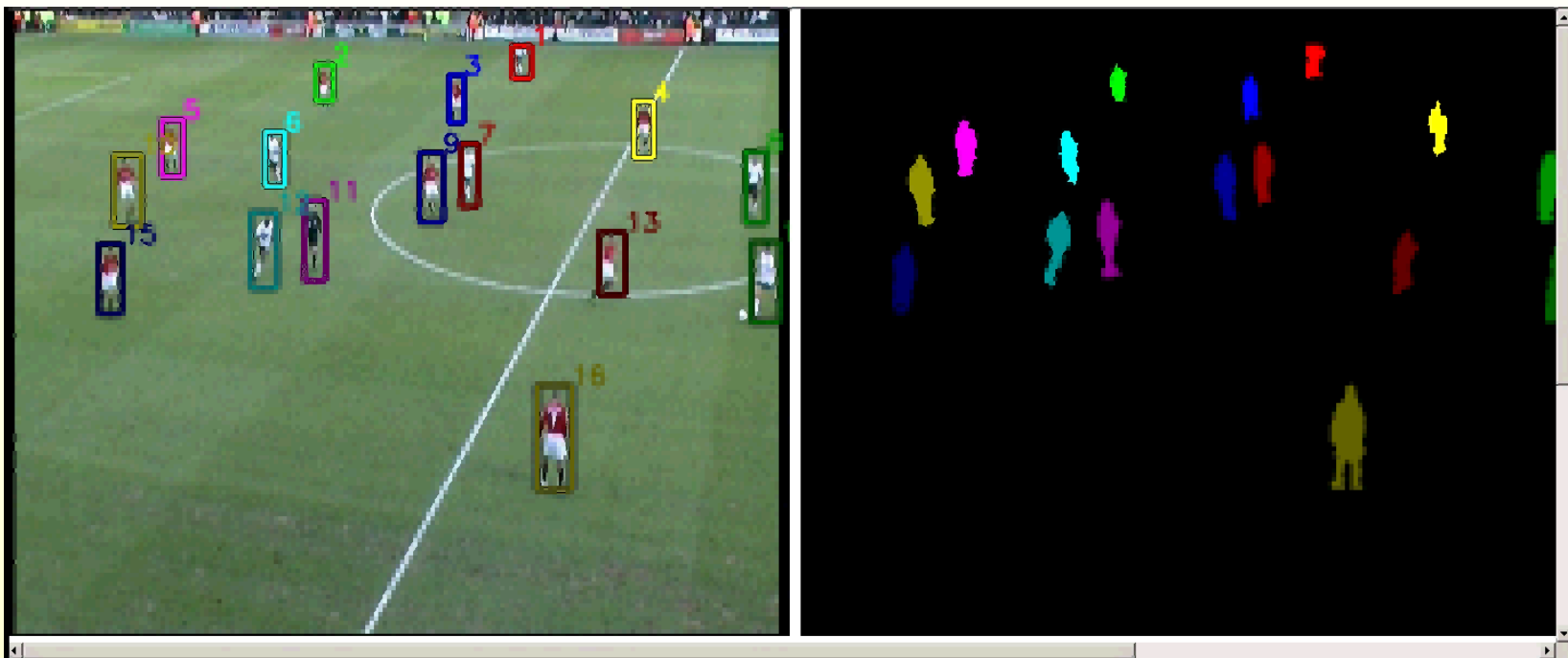
相关滤波 (CF)

Tracking Case



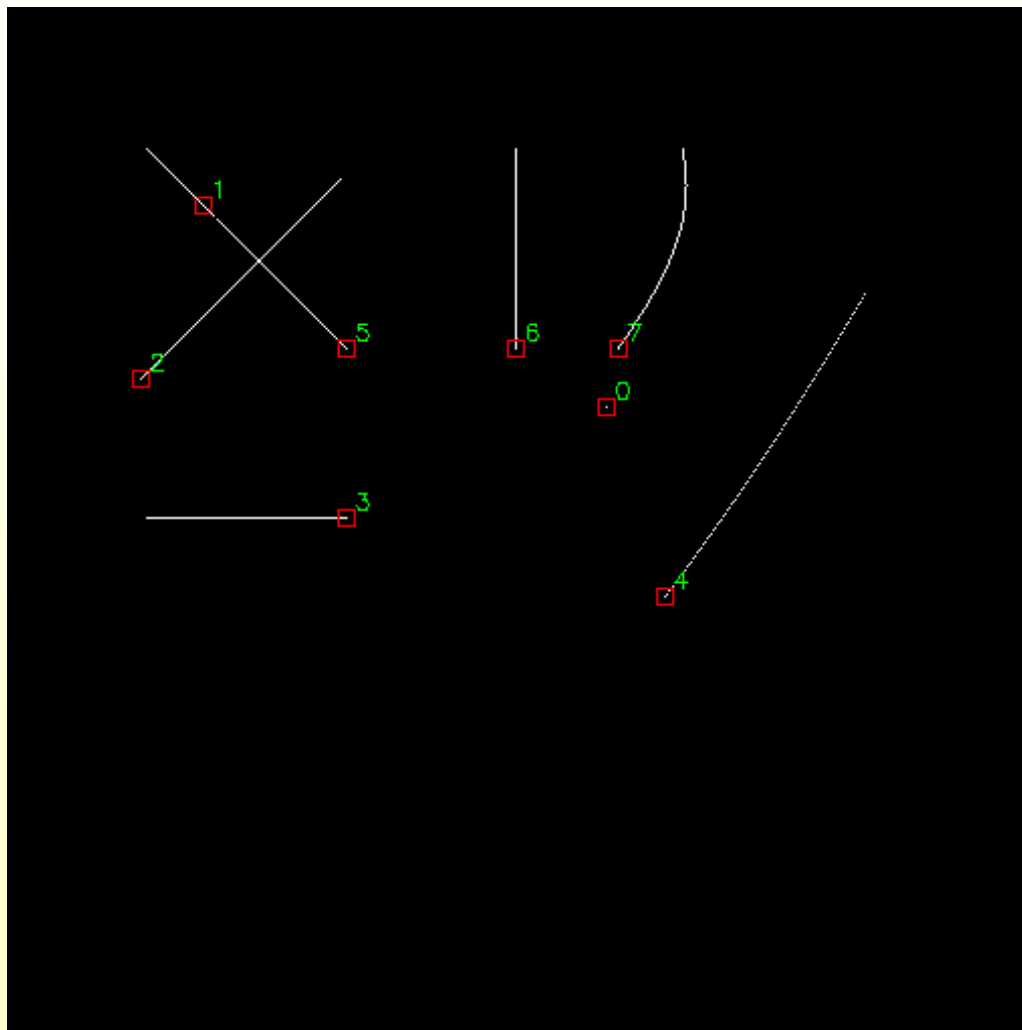
目标跟踪技术

■ 多目标跟踪



多假设跟踪 (Multi-hypothesis tracking , MHT)

多目标跟踪时的编号维持问题



静物遮挡与运动遮挡时的目标跟踪实例



轿车被树遮挡的跟踪结果

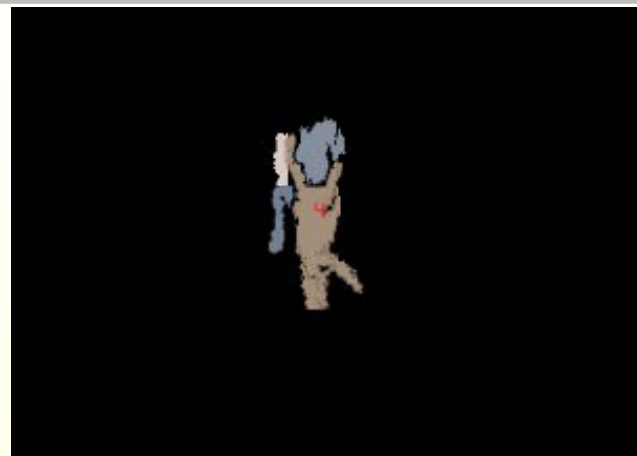


轿车被运动卡车遮挡的跟踪结果

运动遮挡时的目标跟踪实例



(a) 目标发生相互遮挡时的运动区域



(b) 当前区域的主要颜色块



(c) 目标颜色块的匹配结果



(d) 遮挡时目标的跟踪结果

应用颜色信息解决目标发生遮挡时的跟踪问题

多目标跟踪的实验结果

目标 1 和 3 在第三幅所示的图像中发生合并，并将它们作为一个整体跟踪。

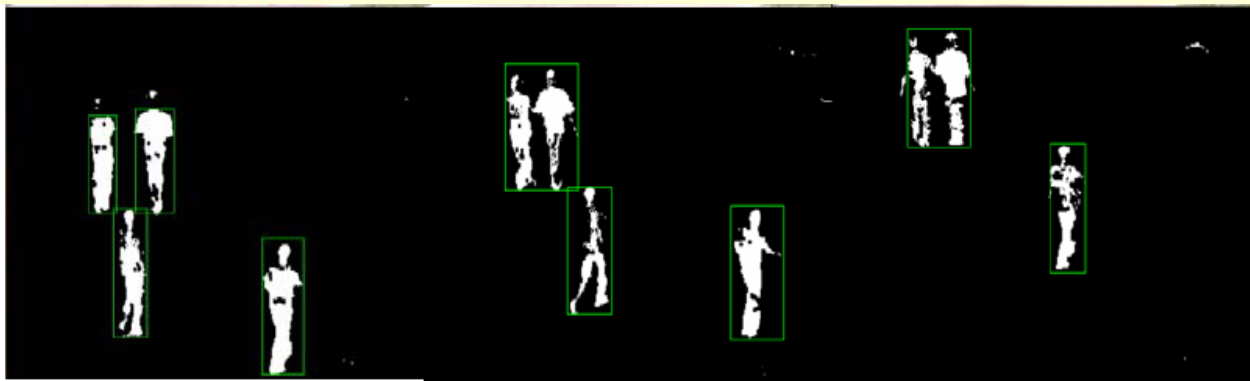
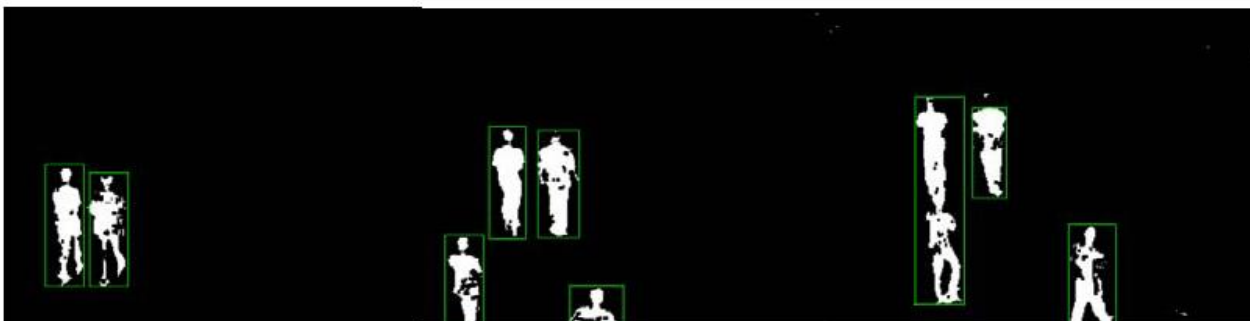
目标 1 和 3 在第四幅发生了分裂。

第五幅目标 1 和目标 2 由于具有共同的运动趋势而合并。

第六幅则是经过时间 T 后，仍保持合并状态的两个目标被最终合并，变为一个新的目标，保持整体跟踪。

目标 4 则由于走出场景而消失。

计算机学院



匹配过程中要注意的关键点：

- **特征空间的选择**

多种特征可供选择，例如：图像本身的灰度、边缘、曲线、角点、直线交点、高曲率点、不变矩、重心等。

- **相似性测度的选择**

相似性测度的选择决定如何确定匹配位置，其配准的程度最后转化为匹配或不匹配。

- **搜索空间与策略的选择**

搜索空间通常是要找到匹配的最佳位置的位置集，很多情况下，减少测量的数量很重要，误匹配位置越多，计算量就越大，问题越严重。可以利用一些可得到的信息去掉不可能匹配的搜索子空间，从而减少计算量。

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

先边缘处理再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