



## 第七章 运动估计和运动目标检测

---



# 概述

---

- 视频运动目标检测指在视频图像序列中判断是否有前景目标的运动，如果有则对目标进行初始定位的检测过程。
- 运动目标检测是实现目标识别跟踪的基础。
- 运动目标检测是视频监控、机器视觉，以及武器制导等领域的重要研究课题。
- 在许多场合，如交通流量的监测、汽车的自动驾驶或辅助驾驶、重要场所的安保、航空和军用飞行器的制导等，都需要准确地对运动目标进行检测。



# 概述

---

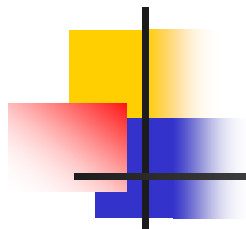
- 在交通运输领域，智能交通系统通过分析交通视频序列来对车辆、行人等目标的运动进行检测、定位、识别和跟踪，完成交通监视、交通控制、出入控制、救援管理等功能，从而增加交通的机动性，提高道路网的通行能力，并提高交通的安全水平，降低交通事故的可能性，减轻交通事故的危害。
- 在视频监控领域，银行、金库、机场等重要场所的安全保卫工作依赖于运动目标分割技术。通过对监控视频序列中的运动目标进行检测、分割、跟踪与记录，从而发现和识别具有异常行为的可疑目标，以预防犯罪行为的发生，实现完全自动化的监控功能。



## 概述

---

- 由于在视频序列中，人们感兴趣的主体一般是具有自主运动的物体，所以现有大多数的视频分割算法处理的是运动目标的分割。

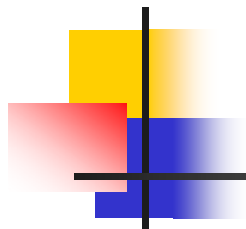


# 本章内容

---

一、运动估计

二、运动目标检测



# 本章内容

---

## 一、运动估计

**1. 块匹配法**

**2. 光流场法**

## 二、运动目标检测



# 1. 块匹配法

---

- 视频动态变化可能由摄像机运动、物体运动或光照改变引起，也可能由物体结构、大小或形状变化引起。
- 为了简化分析，通常我们假设场景变化是由摄像机运动和物体运动引起的，并假设物体是刚性的。



# 1. 块匹配法

---

- 块匹配运动估计因算法简单、便于硬件实现得到广泛应用。
- 比如在数字视频压缩技术中，**MPEG**系列标准和**H.264**标准都采用了块匹配运动估计和运动补偿算法。
- 块匹配运动估计与光流场方法不同，它不需要计算每一个像素的运动，而只是计算由若干像素组成的像素块的运动。对于许多视频应用来说，块运动分析是一种较好的近似。

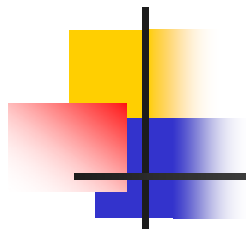




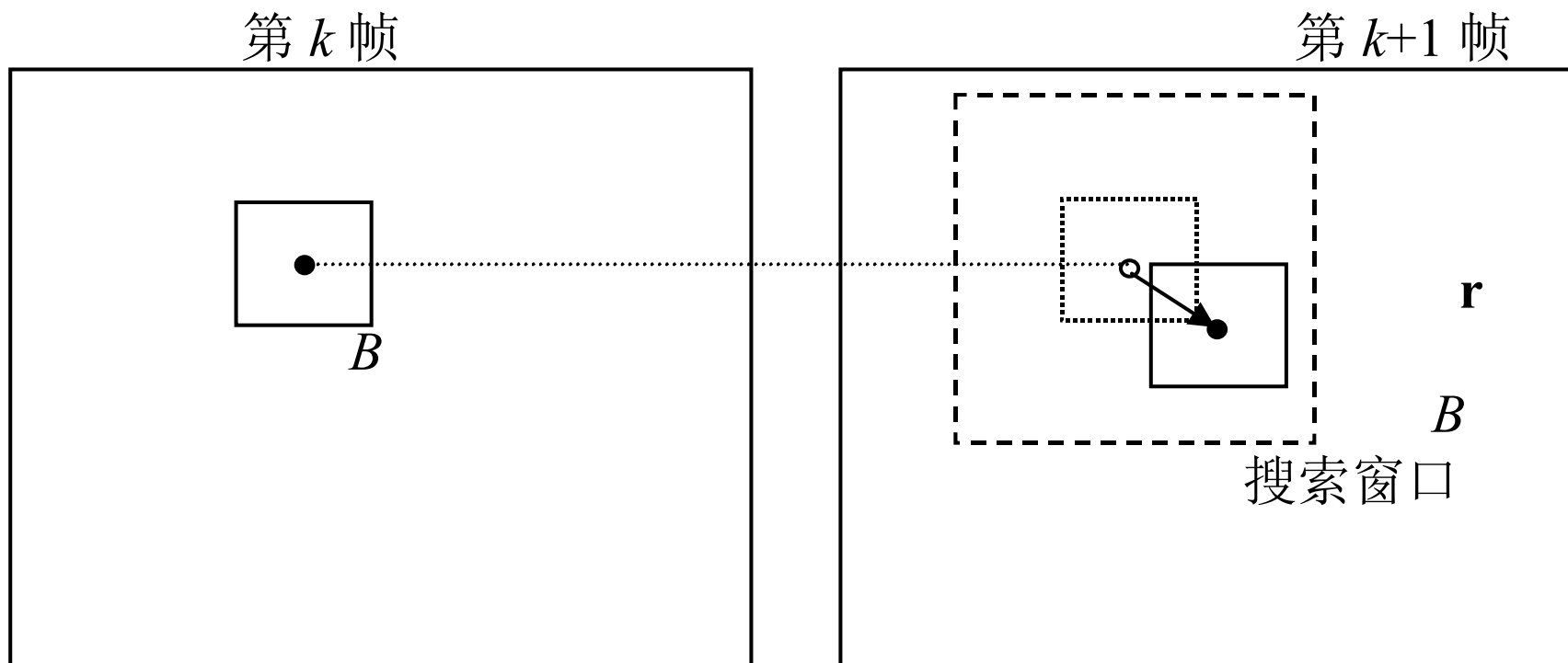
# 1. 块匹配法

---

- 其基本思想是将视频序列的每一帧分成互不重叠的宏块，并认为宏块内所有像素的位移量都相同，然后对每个宏块到参考帧在某一给定搜索范围内根据一定的匹配准则找出与当前块最相似的块，即匹配块。匹配块与当前块的相对位移即为运动矢量。
- 块匹配运动估计的几个因素：匹配准则、搜索策略、块尺寸选择、搜索范围



# 1. 块匹配法





# 1. 块匹配法 --- 匹配准则

- 常见的运动估计匹配准则有三种：**MAD**、**MSE**和**NCCF**（最大互相关准则）。由于**MAD**没有乘除操作，不需做乘法运算，实现简单方便，所以使用较多。通常使用求和绝对误差（**SAD**）代替**MAD**。

$$MAD(i, j) = \frac{1}{MN} \sum_{m=0}^M \sum_{n=0}^N |f_k(m, n) - f_{k-1}(m+i, n+j)|$$

$$MSE(i, j) = \frac{1}{MN} \sum_{m=0}^M \sum_{n=0}^N [f_k(m, n) - f_{k-1}(m+i, n+j)]^2$$

$$NCCF(i, j) = \frac{\sum_{m=0}^M \sum_{n=0}^N f_k(m, n) f_{k-1}(m+i, n+j)}{\left[ \sum_{m=0}^M \sum_{n=0}^N f_k^2(m, n) \right]^{1/2} \left[ \sum_{m=0}^M \sum_{n=0}^N f_{k-1}^2(m, n) \right]^{1/2}}$$

$$SAD(i, j) = \sum_{m=0}^M \sum_{n=0}^N |f_k(m, n) - f_{k-1}(m+i, n+j)|$$



# 1. 块匹配法 --- 搜索策略

---

- 全搜索法对搜索区域的所有位置进行穷尽搜索；精度最高；计算复杂，难以实时处理。
- 快速搜索策略得到的可能是局部最优值，但由于其快速计算的实用性，在实际中得到了广泛的应用。



# 1. 块匹配法 --- 块尺寸选择

---

- 块尺寸常常根据先验知识或经验来确定。
- 视频处理中可采用 $8*8$ 或 $16*16$ 的块



## 2. 光流场法 --- 基本概念

- 当人的眼睛观察运动物体时，物体的景象在人眼的视网膜上形成一系列连续变化的图像，这一系列连续变化的信息不断“流过”视网膜（即图像平面），好像一种光的“流”，故称之为**光流**（**optical flow**）。
- 光流是空间运动物体在观测成像面上的像素运动的瞬时速度。
- 光流的研究是利用视频序列中的像素强度数据的**时域变化**和**相关性**来确定每个像素位置的“运动”。



## 2. 光流场法 --- 基本概念

- 光流场法和块匹配运动估计都可以反映视频相邻帧之间的运动信息。
- 将二维图像平面特定坐标点上的灰度瞬时变化率定义为**光流矢量**。
- **光流场**是指图像灰度模式的表观运动，它是一个二维矢量场，所包含的信息就是各个像素点的瞬时运动速度矢量信息。
- **光流场中每个像素都有一个运动矢量**，因此可以较为准确地反映相邻帧之间的运动。



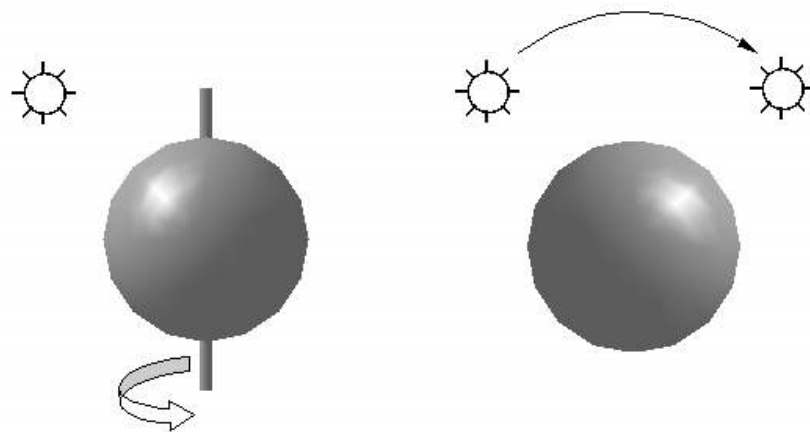
## 2. 光流场法 --- 基本概念

### ■ 运动场与光流场

- ✓ 运动场(motion field): 一个运动物体在空间产生一个三维的速度场，运动前后空间对应点在图象上的投影形成一个二维运动场。
- ✓ 光流场(optical flow field): 是指图像亮度模式的表观（或视在）运动，是二维矢量场。它包含的信息即是各像素点的瞬时运动速度矢量信息。



## 2. 光流场法 --- 基本概念



光流在任意地方都  
等于零，然而，运  
动场却不等于零

明暗模式运动将随着光  
源运动。此时光流不等  
于零，但运动场为零。



## 2. 光流场法 --- 基本原理

- 设在时刻 $t$ 时，像素点 $(x, y)$ 处的灰度值为  $I(x, y, t)$  ；  
在时刻  $(t + \Delta t)$  时，该点运动到新的位置，它在图像上的位置变为  $(x + \Delta x, y + \Delta y)$ ，灰度值为  $I(x + \Delta x, y + \Delta y, t + \Delta t)$
- 根据一致性假设，即沿着运动轨迹像素的灰度保持不变，也就是

$$\frac{dI(x, y, t)}{dt} = 0$$

则

$$I(x, y, t) = I(x + \Delta x, y + \Delta y, t + \Delta t)$$

泰勒展开



## 2. 光流场法 --- 基本原理

- 设 $\mathbf{u}$ 和 $\mathbf{v}$ 分别为该点的光流矢量沿 $x$ 和 $y$ 方向的分量，且  $u = \frac{dx}{dt}$ ,  $v = \frac{dy}{dt}$ 。将上式右边用泰勒公式展

开，得到  $I(x + \Delta x, y + \Delta y, t + \Delta t)$

$$= I(x, y, t) + \frac{\partial I}{\partial x} \Delta x + \frac{\partial I}{\partial y} \Delta y + \frac{\partial I}{\partial t} \Delta t + \varepsilon$$

- 忽略二阶以上的高次项，则有

$$\frac{\partial I}{\partial x} \Delta x + \frac{\partial I}{\partial y} \Delta y + \frac{\partial I}{\partial t} \Delta t = 0$$



## 2. 光流场法 --- 基本原理

- 等式两边同除以  $\Delta t$ ，而  $\Delta t \rightarrow 0$ ，则

$$\frac{\partial I}{\partial x} \frac{dx}{dt} + \frac{\partial I}{\partial y} \frac{dy}{dt} + \frac{\partial I}{\partial t} = 0$$

$\Rightarrow$

$$I_x u + I_y v + I_t = 0$$

其中  $I_x = \frac{\partial I}{\partial x}$ ， $I_y = \frac{\partial I}{\partial y}$ ， $I_t = \frac{\partial I}{\partial t}$ 。分别为参考像素点的灰度值沿  $x$ ， $y$ ， $t$  这三个方向的偏导数。



## 2. 光流场法 --- 基本原理

---

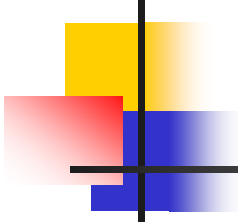
- 上述光流方程中，有 $u$ 和 $v$ 两个变量，但光流的基本等式只有一个方程，因此解是非唯一的。必须附加另外的约束条件才能求解。



## 2. 光流场法 --- 光流场计算

- 为了寻求另外一个约束条件以求解光流方程，光流场的计算有两种典型的方法：Horn-Schunck法和Lucas-Kanade法
- Horn-Schunck法的基本思想是光流在整个图像上光滑变化，运动场既满足光流约束方程又满足全局平滑性。
- 所谓平滑，就是在给定的邻域内  $\nabla^2 u + \nabla^2 v$  应尽可能地小。其中

$$\nabla^2 u = \left[ \frac{\partial u}{\partial x} \right]^2 + \left[ \frac{\partial u}{\partial y} \right]^2 \quad \nabla^2 v = \left[ \frac{\partial v}{\partial x} \right]^2 + \left[ \frac{\partial v}{\partial y} \right]^2$$



## 2. 光流场法 --- 光流场计算

$$e^2(\mathbf{x}) = (I_x u + I_y v + I_t)^2$$

$$s^2(\mathbf{x}) = \iint \left[ \left( \frac{\partial u}{\partial x} \right)^2 + \left( \frac{\partial u}{\partial y} \right)^2 + \left( \frac{\partial v}{\partial x} \right)^2 + \left( \frac{\partial v}{\partial y} \right)^2 \right] dx dy$$

$$E = \iint \{ e^2(\mathbf{x}) + \alpha s^2(\mathbf{x}) \} dx dy$$



## 2. 光流场法 --- 光流场计算

离散表示

$$E = \sum_i \sum_j (e^2(i, j) + \alpha s^2(i, j))$$

$$\frac{\partial E}{\partial u} = 2(I_x u + I_y v + I_t)I_x + 2\alpha(u - \bar{u})$$

$$\frac{\partial E}{\partial v} = 2(I_x u + I_y v + I_t)I_y + 2\alpha(v - \bar{v})$$

当上两式为零时，E取极小值。

用有限差分方法将每个方程中的拉普拉斯算子换成局部邻域图像流矢量的加权和，并使用迭代方法求解这两个差分方程。

$$u^{n+1} = \bar{u}^n - I_x \frac{I_x \bar{u}^n + I_y \bar{v}^n + I_t}{\alpha + I_x^2 + I_y^2} \quad v^{n+1} = \bar{v}^n - I_y \frac{I_x \bar{u}^n + I_y \bar{v}^n + I_t}{\alpha + I_x^2 + I_y^2}$$





## 2. 光流场法 --- 光流场计算

- Lucas-Kanade法：假设在一个小的空间邻域  $\Omega$  上运动矢量保持恒定，然后使用加权最小二乘方（Weighted Least-Squares）估计光流。

$$\sum_{(x,y) \in \Omega} W^2(x) \left( I_x u + I_y v + I_t \right)^2$$

式中  $W(x)$  表示窗口权重函数，它使邻域中心部分对约束产生的影响比外围部分大。



## 2. 光流场法 --- 快速算法

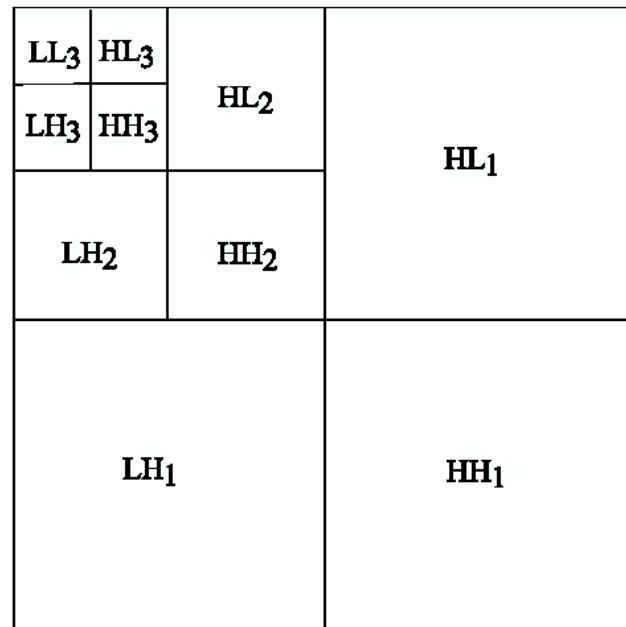
- 快速算法:

采用高斯金字塔将图像分解成不同分辨率的等级，随着级别的增加，分辨率越来越低。

将 **Lucas-Kanade** 方法运用到高斯金字塔中进行光流计算。从金字塔的最粗尺度开始估计光流场，并将在粗尺度下得到的结果作为下一尺度的初始值，最终得到原始图像的光流场。



## 多级分解子带示意图



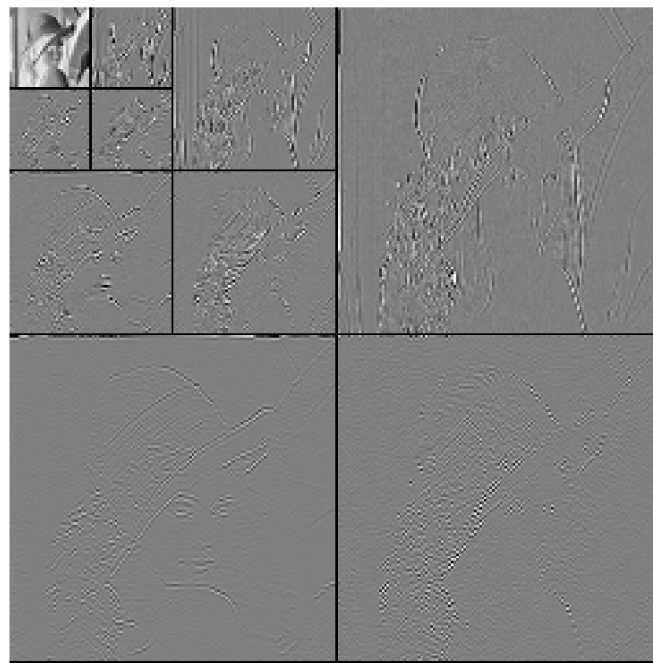
- $LL_i$  为逼近系数构成的子带，是原图像的逼近；
- $HL_i$ 、 $LH_i$  和  $HH_i$  分别为水平方向、垂直方向和对角方向的细节子带



# Lena图像三级分解



**(1) Lena图像**



**(2) Lena图像的三级分解**

- 逼近子带是原图像的低分辨率逼近;
- 细节子带反映图像的边缘和细节信息



## 2. 光流场法 ---存在的问题

---

- 计算光流通常基于以下假设：
  - ✓ 任何物体观察到的亮度随时间恒定不变；
  - ✓ 图像中相邻点以类似的方式进行移动。
- 当上述条件不满足时（例如当光照条件发生变化或运动对象边界处），光流计算可能会产生错误。



## 2. 光流场法 --- 存在的问题

- 对于光线突变的情况，当前帧图像的亮度必然随之急剧变化。利用这个特性，判断视频相邻两帧图像的 **HSI** 空间中的 **H** 空间（色调空间）和 **I** 空间（亮度空间）的变化情况，即当图像 **I** 空间分量变化较大并且 **H** 空间变化比较微弱时，视频中发生了光照突变的情况。

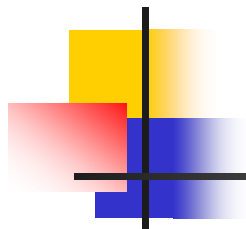


## 2. 光流场法 --- 存在的问题

---

- 局部光流估计:

同样是基于亮度和平滑性假设，但是这种方法将图像分割成若干小区域，在每个小区域内进行光流估计。



# 本章内容

---

## 一、运动估计

## 二、运动目标检测

**1. 时间差分法**

**2. 背景差分法**

**3. 光流计算法**





# 1. 时间差分法

- 时间差分法直接比较视频相邻帧对应像素点发生的相对变化，是一种简单直接的运动目标检测方法。
- 在最简单的形式下，相邻帧之间的变化可用一个二值差分图像表示，即

$$f_d(x, y, t_1, t_2) = \begin{cases} 1 & |f(x, y, t_1) - f(x, y, t_2)| > T \\ 0 & else \end{cases}$$

其中T为阈值。在光照条件不变的前提下，差分图像中取值为1的像素点可被认为是运动物体。

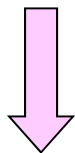


# 1. 时间差分法

---

在序列图像中，通过**逐像素比较**可直接求取前后两帧图像之间的差别

假设**照明条件**在多帧图像间基本**不变化**，那么差图像大于门限值处表明该处的像素发生了移动



也就是说，对时间上相邻的**两幅图像求差**，可以将图像中**目标的位置和形状变化**突出出来



# 1. 时间差分法 --- 阈值的选取

---

时间差分法中阈值的选择相当关键，过低的阈值不能有效抑制视频中的噪声，过高的阈值将遗漏图像中有用的变化。

阈值的选择通常取决于场景光照和摄像机等外界条件。



# 1. 时间差分法 --- 优缺点分析

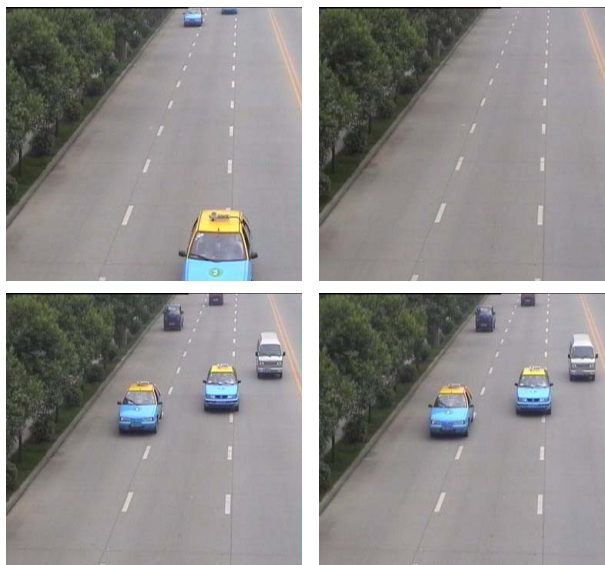
优点：基于差分图像的运动目标检测算法具有较快的检测速度，**适合实时应用**。

缺点：当运动目标内部的灰度分布较为平坦时，往往被误判为背景区域从而在运动目标内部出现**空洞**(即漏警)；另外，由于目标的运动而产生的遮挡和显露的背景区域的分布特性与相邻帧共同的背景区域不同，因此会被误判为运动目标并造成虚警，从而使**运动目标的边界很不精确**；再次，基于差分的运动目标检测算法**易于受到噪声的影响**，在检测结果中出现大量孤立的噪声点。

因此，基于帧差法的运动目标分割算法常常需要进行后续处理，以便分割出具有正确边缘的运动目标。

## 2. 背景差分法

- 检测运动目标常用差分图像的方法，有两种情况：一种是当前连续两帧图像之间的差分称为相邻帧差分法或时间差分法，二是当前图像与固定背景图像之间的差分称为背景差分法。





## 2. 背景差分法

- 背景差分法是目前运动目标检测中最常用的一种方法。利用当前图像与背景图像进行差分，从而检测出运动区域。
- 设时刻 $t$ 背景图像为  $f_b(t)$ ，当前帧图像为  $f_c(t)$ ，则背景差分图像为

$$f_d(x, y, t) = |f_b(x, y, t) - f_c(x, y, t)|$$

若  $f_d(x, y, t) \geq T$ ，则 $(x, y)$ 属于运动目标，反之属于背景。这里 $T$ 是门限阈值。



## 2. 背景差分法

- 基于背景减法的运动目标检测技术，其核心在于背景模型的建立。
- 如果能够获取较好的背景模型，则背景减法不仅能提高目标分割的精度，而且可以发现暂时静止的目标或目标中暂时静止的部分。
- 对视频监控，背景可以事先采集。
- 此外，还可以直接抽取视频图像序列的某一关键帧作为背景，或者计算多帧图像的平均值或中值作为背景。



## 2. 背景差分法 --- 背景模型

- 一种经典的方法是**时间平均法**，即对一段时间中的图像序列求和再平均，获得一帧近似的背景图像，但是这种方法容易将前景运动目标混入背景图像当中，产生混合现象。
- **中值滤波**也是最常用的背景建模方法之一。中值滤波的原理认为如果某个像素属于背景区域，那么该像素至少在缓冲区一半的时间长度内均停留在背景区域。基于这个假设，中值滤波算法计算缓冲区内同一位置对应像素的中值，并作为该像素的背景，从而建立背景模型。





## 2. 背景差分法

---

- 上述的背景模型算法在背景静止的条件具有较好的效果，但是当背景包含**噪声或者摄像机运动**的情况下，其检测精度会显著下降。
- 如果背景区域在相邻帧之间**光照、位置**等属性发生变化时，以及在背景区域本身包含**微小运动**如树叶的飘动、水波、雨，以及雪的情况下，如何精确地获取背景模型是一个研究的难点。
- 对背景差分法的改进大多都集中在背景模型的生成。



## 2. 背景差分法

---

- 背景模型可以分为两类：
  - ✓ 参数化背景模型
  - ✓ 非参数化背景模型
- 参数化背景模型，假设背景变化服从某个分布，通过估计相应参数来表示背景模型。
- 非参数化方法不需要假定模型的概率密度分布形式，直接从数据序列中估计得到背景。



## 2. 背景差分法

- 比较著名的利用统计学方法来构建背景模型的算法是**混合高斯模型算法**（**Gaussian Mixture Model, GMM**）。
- 在这个模型中，每个像素都有一个对应的概率密度，它是由几个不同权重的高斯概率密度函数（每个函数的参数不同）之和而成。这些函数的参数和权重则是由观测的历史视频帧的像素统计值决定。
- **GMM**模型能够很好地解决从复杂的自然背景中分割出运动目标的问题，尤其是具有重复运动规律的背景，如随风飘动的树叶，不停流淌的河水等。



## 2. 背景差分法

---

- **GMM**模型有一个很明显的弊端，就是它的算法建立在一个局限性很强的假设下。它假设，背景一定比前景出现的概率大；背景比前景的变化率要低。
- 当视频中有高速运动的前景目标，上述假设将不成立，从而会出现漏检的情况。



## 2. 背景差分法 --- 背景模型

---

- 基于参数化的背景模型的建立具有很多局限性，因此有很多学者提出了基于非参数的背景模型。
- 一种性能较好的非参数背景帧获取方法：**ViBe**算法。
- 该算法为每个像素点建立一个样点集，并通过比较输入像素与样点集的相似度来判断其是否属于背景。
- **ViBe**只要求当前像素满足与样点集中的某些像素相近，而不需要和样点集中的大多数像素相近。这样得到的背景像素滤除了一些极端的情况。



## 2. 背景差分法 --- 背景模型

---

[1] Zivkovic, Z. Improved adaptive Gaussian mixture model for background subtraction [J], Pattern Recognition, 2004.2(17):28-31

[2] O.Barnich and M. Droogenbroeck, ViBe: A Universal Background Subtraction Algorithm for Video Sequences, IEEE Transactions on image Processing, 20(6):1709-1724, June 2011.



### 3. 光流场法

---

- 通常情况下，运动目标具有与背景不一致的运动，因此可以通过分析目标的运动场来分割视频序列。
- 运动场是三维物体的实际运动在成像平面上的投影，然而我们所能得到的只是图像亮度随时间的变化，即所谓的光流场。
- 在一定的条件下，光流场与运动场相互吻合。
- 基于光流场的分割就是通过研究光流场，从视频序列中近似计算运动场，然后根据运动场的运动特征进行视频分割。



### 3. 光流场法

---

- 由于光流场分割算法对噪声极为敏感，光流场的估计精度受限于遮挡问题和孔径问题，因此光流法往往**无法获取准确的目标边界**。
- 为了提高光流分割算法的准确性，可联合目标的颜色、亮度、边缘等空间信息进行多特征的视频分割。
- 另外，光流计算本身运算量较大，耗时较多，**实时应用较为困难**。