

图象工程（中）

图 象 分 析

（第4版）

章毓晋

清华大学电子工程系 100084 北京

第3单元 特性分析

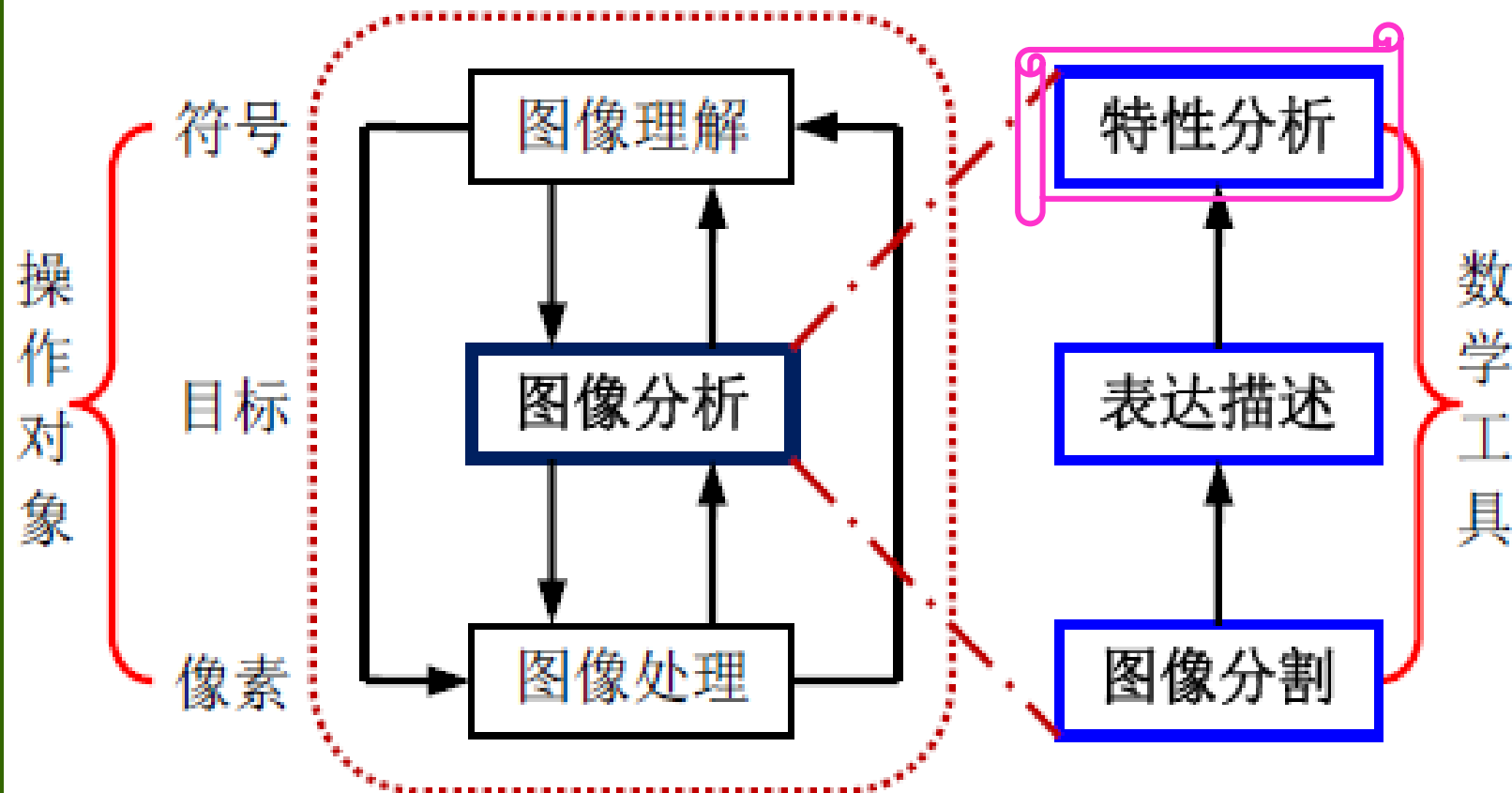
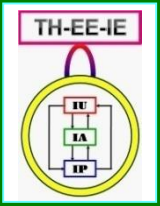
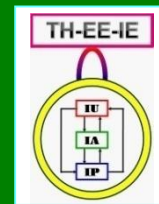


图 1.2.2 图像分析主要功能模块



第3单元 特性分析

- 第9章 纹理分析：对纹理的表达和描述，纹理图象分割，分类和合成
- 第10章 形状分析：特征的方法，形状变换的方法，基于关系的方法
- 第11章 运动分析：景的变化（外在照明改变、视场改变）和物的变化（目标位置朝向改变）
- 第12章 显著性和属性
显著性：与主观感知相关联
属性：人指定名称并可观察到的特性

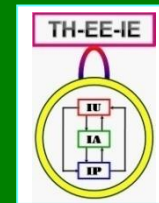


第11章 运动分析

图象序列和运动信息

序列图象（也称动态图象）

- 有一定次序（给定或假设）的一系列图象
- 视频图象是（时间上）等间隔的序列图象
- 一类特殊的3-D图象， $f(x, y, t)$
- 序列图象分析：建立在对静止图象中目标的分析基础之上，还需要在技术上有所扩展、在手段上有所变化、在目的上有所拓宽



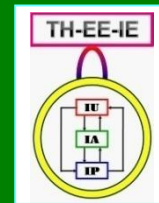
第11章 运动分析

11.1 运动研究内容

11.2 运动目标检测

11.3 运动目标分割

11.4 运动目标跟踪



11.1 运动研究内容

运动分析的研究目的和工作内容

(1) 运动（信息）检测{上册}

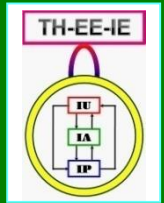
检测场景/视场中的（运动）变化

(2) 运动目标定位和跟踪

确定场景中的运动目标（位置，姿态，轨迹，特点，进一步运动的方向和路线，等等）

运动目标定位 \Leftrightarrow 运动目标检测

- 定位更关注位置，常对运动情况有一定假设
- 检测还常考虑目标自身特性



11.1 运动研究内容

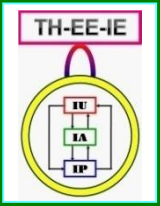
运动分析的研究目的和工作内容

(3) 运动目标分割和分析

需要精确检测目标的运动情况，提取目标区域
常需要利用视频摄像机获取序列图象，并区分场景中的全局运动和局部运动。还常需要获取目标的3-D特性，运动参数，分析运动规律，或进一步识别运动物体的类别

(4) 立体景物重建和行动/场景理解 {下册}

计算立体景物（各部分或之间）的深度/距离、确定其表面的朝向以及遮盖情况等以理解场景



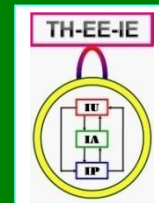
11.2 运动目标检测

检测局部运动及其模式

11.2.1 背景建模

11.2.2 光流场

11.2.3 特定运动模式的检测



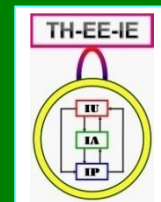
11.2.1 背景建模

1. 基本原理

并不将背景看成是完全不变的，而是计算和保持一个动态（满足某种模型）的背景帧

一种具体算法主要包括如下步骤

- ① 先获取前 N 帧图象，在每个像素处，确定这 N 帧的中值，作为当前的背景值
- ② 获取第 $N+1$ 帧，计算该帧与当前背景在各像素处的差
- ③ 消除差图象中非常小的区域并填充大区域中的孔。保留下来的区域代表了场景中运动的目标
- ④ 结合第 $N+1$ 帧更新中值 ⑤ 返回到步骤②



11.2.1 背景建模

2. 典型实用方法

(1) 基于单高斯模型的方法

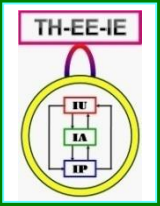
认为像素点的值在视频序列中服从高斯分布

针对每个固定的像素位置，计算 N 帧训练图像序列中该位置像素值的均值 μ 和方差 σ ，并从而唯一地确定出一个单高斯背景模型

运动检测时用背景相减的方法， $|\mu_T - \mu| \leq 3\sigma$

?对应用条件要求较严，对光照强度变化敏感

?运动前景与静止背景不易分离开



11.2.1 背景建模

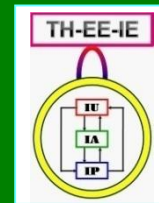
2. 典型实用方法

(2) 基于视频初始化的方法

在有运动前景的情况下，将各像素点上的背景值先提取出来，将静止背景与运动前景分离开来，然后再进行背景建模

对每个像素位置的长度为 N 的序列进行截取

通过这个初始化，把背景静止但有运动前景的情况转化成背景静止也没有运动前景的情况，滤除了运动前景对背景建模的影响



11.2.1 背景建模

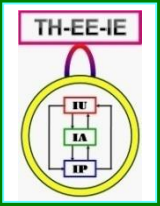
2. 典型实用方法

(3) 基于高斯混合模型的方法

引入高斯混合模型，解决运动背景下的建模
对各个像素分别用混合的高斯分布来建模，
有些高斯分布代表背景而有些高斯分布代表前景

对背景的多个状态分别建模，根据数据属于
哪个状态来更新该状态的模型参数

$$P[f(t)] = \sum_{k=1}^K w_k(t) \frac{1}{\sqrt{2\pi}} \exp \left[\frac{-[f(t) - \mu_k(t)]^2}{\sigma_k^2(t)} \right]$$



11.2.1 背景建模

2. 典型实用方法

(4) 基于码本的方法

将每个像素点用一个码本表示，一个码本可包含一个或多个码字，每个码字代表一个状态

码本最初借助对训练帧图象进行学习而生成
用时域滤波器滤除码本中代表运动前景的码字，保留代表背景的码字

用空域滤波器将那些被时域滤波器错误滤除的码字（代表较少出现的背景）恢复到码本中

11.2.1 背景建模

3. 效果示例

(1) 静止背景中无运动前景时的结果

要检测的是其进入场景的人 单高斯模型



(a)



(b)



(c)

图 11.2.1 静止背景中无运动前景时的结果

11.2.1 背景建模

3. 效果示例

(2) 静止背景中有运动前景时的结果

初始场景有人，后离去，要检测离去的人

参考

视频初始化

基于码本



(a)



(b)



(c)



(d)

图 11.2.2 静止背景中有运动前景时的结果

{表11.2.1}

11.2.1 背景建模

3. 效果示例

(3) 运动背景中无运动前景时的结果

初始场景里树在晃动，要检测进入场景的人

参考

高斯混合模型

基于码本



(a)



(b)

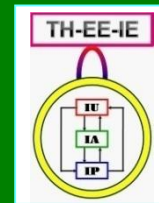


(c)



(d)

图 11.2.3 运动背景中无运动前景时的结果



11.2.2 光流场

1. 光流方程

时刻 t : 一个图象点在 (x, y) 处; 时刻 $t + dt$: 该图象点移动到 $(x + dx, y + dy)$; 设 dt 很小, 图象点灰度未变

$$f(x, y, t) = f(x + dx, y + dy, t + dt)$$

用泰勒级数展开, 令 $dt \rightarrow 0$, 取极限略去高阶项

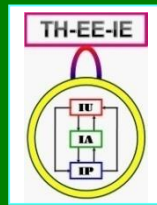
$$\frac{\partial f}{\partial x} \frac{dx}{dt} + \frac{\partial f}{\partial y} \frac{dy}{dt} = \frac{\partial f}{\partial x} u + \frac{\partial f}{\partial y} v = - \frac{\partial f}{\partial t}$$

$$u = dx/dt$$

$$v = dy/dt$$

令: $f_x = \partial f / \partial x$ $f_y = \partial f / \partial y$ $f_t = \partial f / \partial t$

$$[f_x, f_y][u, v]^T = -f_t$$



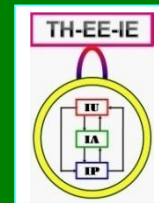
11.2.2 光流场

1. 光流方程

灰度时间变化率是灰度空间变化率与该点空间运动速度的乘积。均用一阶差分平均值来估计

$$f_t \approx \frac{1}{4} [f(x, y, t+1) + f(x+1, y, t+1) + f(x, y+1, t+1) + f(x+1, y+1, t+1)] \\ - \frac{1}{4} [f(x, y, t) + f(x+1, y, t) + f(x, y+1, t) + f(x+1, y+1, t)]$$

$$f_x \approx \frac{1}{4} [f(x+1, y, t) + f(x+1, y+1, t) + f(x+1, y, t+1) + f(x+1, y+1, t+1)] \\ - \frac{1}{4} [f(x, y, t) + f(x, y+1, t) + f(x, y, t+1) + f(x, y+1, t+1)]$$



11.2.2 光流场

2. 最小二乘法光流估计

在连续两幅图象 $f(x, y, t)$ 和 $f(x, y, t+1)$ 上取具有相同 u 和 v 的同一个目标上的 N 个不同位置的像素

$$\mathbf{f}_t = \begin{bmatrix} -\hat{f}_t^{(1)} \\ -\hat{f}_t^{(2)} \\ \vdots \\ -\hat{f}_t^{(N)} \end{bmatrix} \quad F_{xy} = \begin{bmatrix} \hat{f}_x^{(1)} & \hat{f}_y^{(1)} \\ \hat{f}_x^{(2)} & \hat{f}_y^{(2)} \\ \vdots & \vdots \\ \hat{f}_x^{(N)} & \hat{f}_y^{(N)} \end{bmatrix}$$

$$\begin{bmatrix} u & v \end{bmatrix}^T = \left(F_{xy}^T F_{xy} \right)^{-1} F_{xy}^T \mathbf{f}_t$$

11.2.2 光流场

3. 运动分析中的光流

动态图象中的运动是下面4种基本运动的组合

(a) 与摄像机的距离为常数的平动

(b) 相对于摄像机在深度方向上沿视线的平动

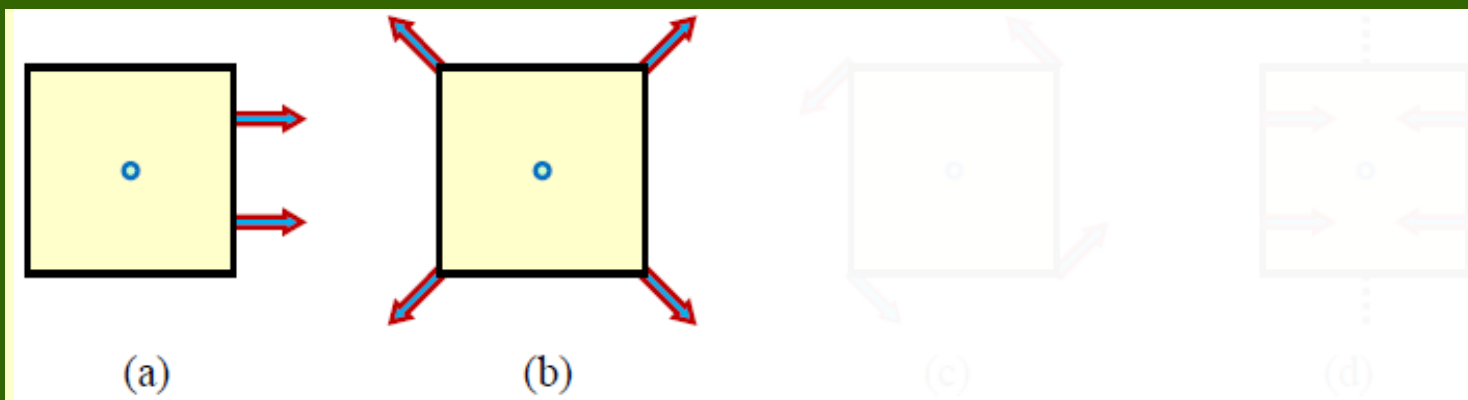


图 11.2.5 运动形式的识别

11.2.2 光流场

3. 运动分析中的光流

动态图象中的运动是下面4种基本运动的组合

(c) 围绕视线等距离的转动

(d) 与视线正交的平面目标的转动

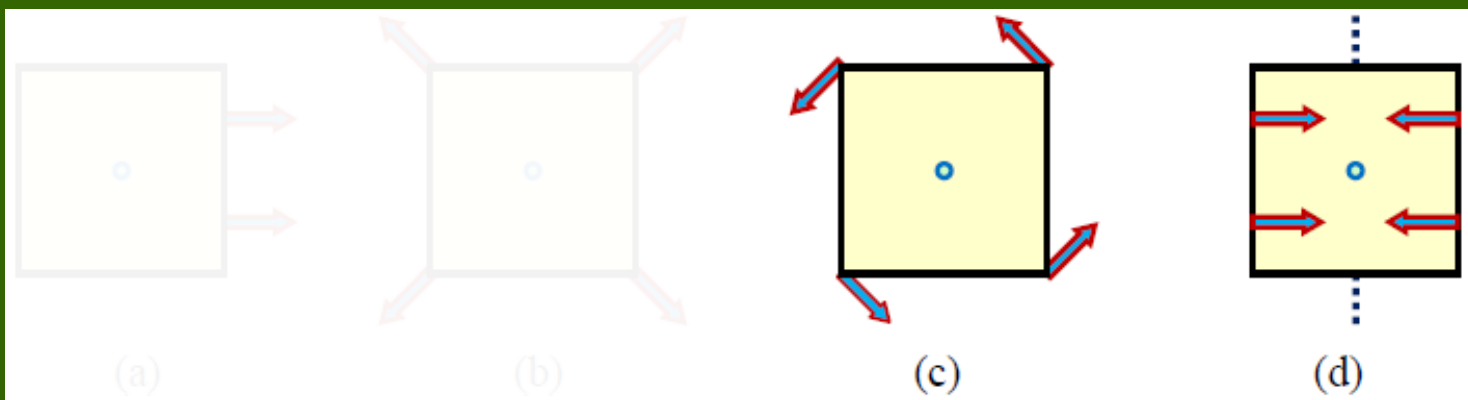


图 11.2.5 运动形式的识别

11.2.2 光流场

对光流场的解释

- (a) 其中仅有一个目标向右移动
- (b) 对应摄像机向前（进入纸内）运动 + (a)
- (c) 对应有一个目标向固定着的摄像机方向运动
- (d) 对应有一个目标绕摄像机视线旋转的情况
- (e) 对应有一个目标绕与视线正交的一根水平轴旋转

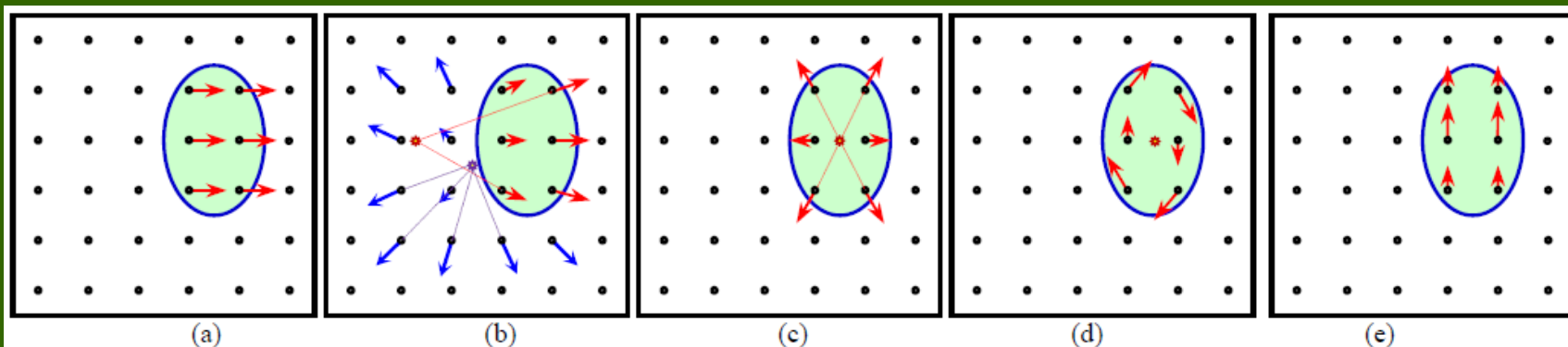
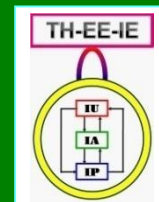


图 11.2.6 对光流场的解释

{P.272}



11.2.3 特定运动模式的检测

确定某些特定的运动模式

- 结合使用基于图象的信息和基于运动的信息
- 运动信息可通过确定先后采集的图象之间的特定差异来获得
- 运动的方向可借助下面4种差图象计算

上
下
左
右

$$\begin{aligned}U &= |f_t - f_{t+\delta t \uparrow}| \\D &= |f_t - f_{t+\delta t \downarrow}| \\L &= |f_t - f_{t+\delta t \leftarrow}| \\R &= |f_t - f_{t+\delta t \rightarrow}|\end{aligned}$$

图 象
移 动
方 向

11.2.3 特定运动模式的检测

✓ 运动的幅度可用对图块区域的求和来得到

(1) 令 $s(x, y)$ 代表对一行像素的累积和, $s(x, -1)=0$

(2) 令 $I(x, y)$ 是一幅**积分图**, $I(-1, y)=0$

(3) 计算行的累积和 $s(x, y)$

$$s(x, y) = s(x, y-1) + f(x, y)$$

(4) 逐行扫描构建积分图 $I(x, y)$

$$I(x, y) = I(x-1, y) + s(x, y)$$

✓ 矩形和的计算

$$D_{\text{sum}} = I(\delta) + I(\alpha) - [I(\beta) + I(\gamma)]$$

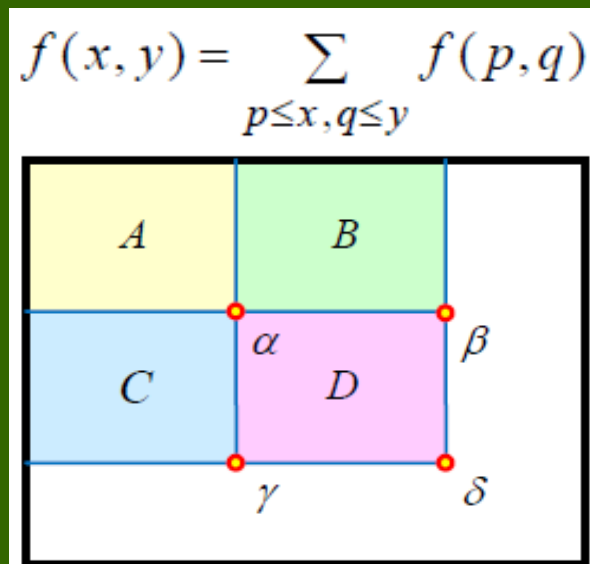
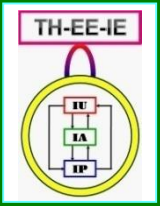


图 11.2.7 积分图像计算示意



11.3 运动目标分割

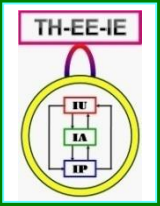
视频图象是3-D图象: $f(x, y, t)$

- 时间分割: 分解为时间片段 (常为镜头)
- 空间分割: 把独立运动的目标逐帧检测出来

11.3.1 目标分割和运动信息提取

11.3.2 稠密光流算法

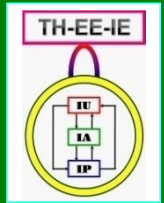
11.3.3 基于参数和模型的分割



11.3.1 目标分割和运动信息提取

三种策略

- (1) 先分割之后再计算运动信息（直接法）
先获得区域边界，再估计区域运动模型
- (2) 先计算运动信息再分割（间接法）
先计算全图运动矢量场，再分割运动矢量场
- (3) 同时计算运动信息和进行分割
比较复杂，需要相当大的计算量



11.3.2 稠密光流算法

1. 求解光流方程

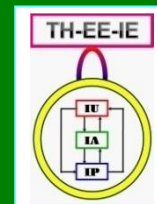
$$(f_x, f_y) \cdot (u, v) = -f_t$$

？：一个方程，两个未知量 (u, v)

- 光流误差 e_{of} ：运动矢量场中不符合光流方程的部分

$$e_{\text{of}} = \frac{\partial f}{\partial x} u + \frac{\partial f}{\partial y} v + \frac{\partial f}{\partial t}$$

要使 e_{of} 在整个帧内的平方和达到最小



11.3.2 稠密光流算法

1. 求解光流方程

- （光流）速度场梯度误差 e_s^2

$$e_s^2 = \left(\frac{\partial u}{\partial x} \right)^2 + \left(\frac{\partial u}{\partial y} \right)^2 + \left(\frac{\partial v}{\partial x} \right)^2 + \left(\frac{\partial v}{\partial y} \right)^2$$

要使整个运动矢量场尽可能趋于平滑

- Horn-Schunck法同时考虑两种约束

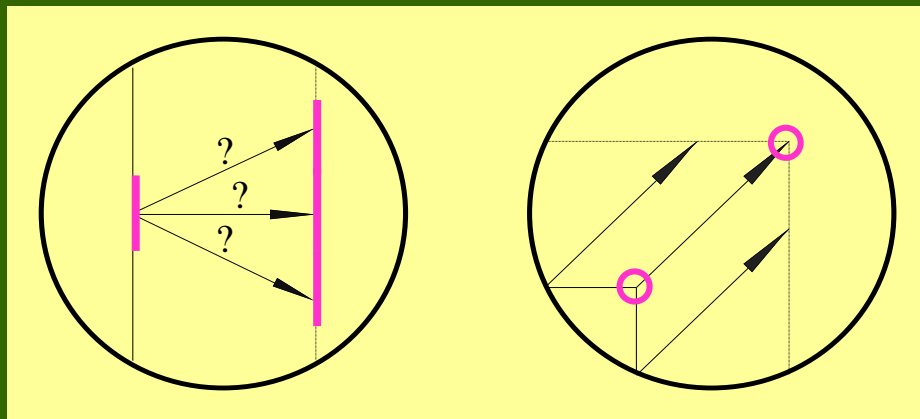
$$\min_{u,v(x,y)} \int_A \left[e_{\text{of}}^2(u,v) + \alpha^2 e_s^2(u,v) \right] dx dy$$

11.3.2 稠密光流算法

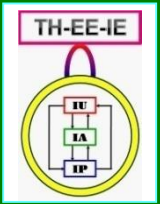
2. 基本算法的问题

(1) 孔径问题：模板尺寸有限而产生的问题

例：只能确定法线分量（与边缘垂直）而无法确定与边缘平行的分量



? 算法的计算量常很大，一般基于该算法的运动信息提取通常只能做到离线的非实时处理



11.3.2 稠密光流算法

2. 基本算法的问题

(2) 逼近误差问题：当运动矢量的幅度较大时，逼近结果的误差也会较大

- 改进 {
- a) 用位移帧差项 $f(x + \bar{u}_n, y + \bar{v}_n, t + 1) - f(x, y, t)$ 代替光流误差项 e_{of}
 - b) 用平均梯度代替偏导 $\partial f / \partial x$, $\partial f / \partial y$

$$f_x = \frac{1}{2} \left[\frac{\partial f}{\partial x}(x + \bar{u}_n, y + \bar{v}_n, t + 1) + \frac{\partial f}{\partial x}(x, y, t) \right]$$

11.3.2 稠密光流算法

3. 全局运动补偿 {P.277}

用块匹配法计算出的运动矢量场

对全局运动估计出的运动矢量场

带全局运动补偿的稠密光流迭代算法计算出的局部运动矢量场



(a)

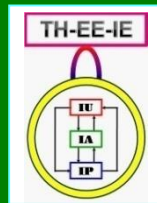


(b)



(c)

图 11.3.2 块匹配法和改进稠密光流迭代算法计算结果的比较

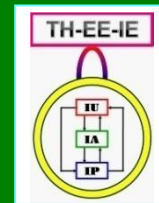


11.3.3 基于参数和模型的分割

假设图象中有 K 个独立的运动目标，则由此计算出的每个光流矢量对应一个不透明目标在3-D空间进行刚体运动所得到的投影。这样每个独立的运动目标都可准确地用一组映射参数来描述

设有 K 组参数矢量，每一组在各个像素处都定义一个光流矢量。由映射参数定义的光流矢量称为基于模型的或合成的光流矢量，这样在每个像素处有 K 个合成光流矢量

对图象光流场分割以获得不同的运动区域 \Leftrightarrow 对像素处的估计光流矢量赋一个合成光流矢量标号



11.3.3 基于参数和模型的分割

1. 借助哈夫变换

6参数仿射流模型

$$v_1 = a_1 + a_2 x_1 + a_3 x_2$$

$$v_2 = a_4 + a_5 x_1 + a_6 x_2$$

如知道在3个或以上点的光流，就可以将系数解出来

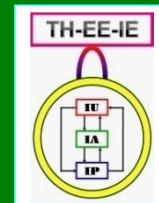
用每个光流矢量 $\mathbf{v}(\mathbf{x}) = [v_1(\mathbf{x}) \ v_2(\mathbf{x})]^T$ 对一组量化的参数投票以最小化：

$$\eta^2(\mathbf{x}) = \eta_1^2(\mathbf{x}) + \eta_2^2(\mathbf{x})$$

$$\eta_1(\mathbf{x}) = v_1(\mathbf{x}) - a_1 - a_2 x_1 - a_3 x_2$$

$$\eta_2(\mathbf{x}) = v_2(\mathbf{x}) - a_4 - a_5 x_1 - a_6 x_2$$

- 计算量很大 \Rightarrow 基于改进哈夫变换的两步算法 {P.278}

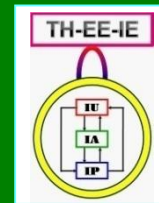


11.3.3 基于参数和模型的分割

2. 借助区域生长

假设场景中的运动物体是由一系列具有不同运动特性的平面区域所构成，不同的区域可以用具有不同仿射参数的区域运动模型来表示

- 种子点的选取：可按照一定的排序准则来完成，如顺序选取、随机选取（都要使种子点处在平滑区域的中部）
- 区域生长准则：用运动矢量是否满足仿射运动模型来作为区域的一致性准则

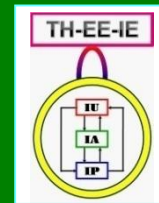


11.4 运动目标跟踪

在每帧视频中都检测和定位出同一个目标

11.4.1 典型技术

11.4.2 子序列决策策略



11.4.1 典型技术

1. 卡尔曼滤波器

在对当前帧内目标跟踪时，常希望能够预测其在后续帧中的位置。为此，需要连续地更新被跟踪目标点的位置和速度

$$\mathbf{x}_i = \mathbf{x}_{i-1} + \mathbf{v}_{i-1}$$

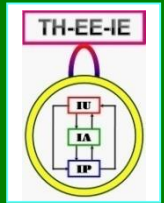
$$\mathbf{v}_i = \mathbf{x}_i - \mathbf{x}_{i-1}$$

三个量：原始位置，观测前对应变量（模型参数）的最优估计值（加上标 $-$ ），观测后对应变量的最优估计值（加上标 $+$ ）

$$\mathbf{x}_i^- = \mathbf{x}_{i-1}^+ + \mathbf{v}_{i-1} + \mathbf{m}_{i-1}$$

$$\mathbf{v}_i^- = \mathbf{v}_{i-1}^+ + \mathbf{n}_{i-1}$$

噪声



11.4.1 典型技术

1. 卡尔曼滤波器

预测方程（最优解）

速度为常数，噪声为高斯

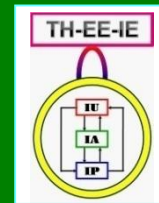
$$\mathbf{x}_i^- = \mathbf{x}_{i-1}^+$$

$$\sigma_i^- = \sigma_{i-1}^+$$

校正方程

$$\mathbf{x}_i^+ = \frac{\mathbf{x}_i / \sigma_i^2 + (\mathbf{x}_i^-) / (\sigma_i^-)^2}{1 / \sigma_i^2 + 1 / (\sigma_i^-)^2}$$
$$\sigma_i^+ = \left[\frac{1}{1 / \sigma_i^2 + 1 / (\sigma_i^-)^2} \right]^{1/2}$$

通过重复测量可以在每次迭代中改进对位置参数的估计以及减少基于它们的误差



11.4.1 典型技术

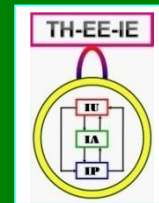
2. 粒子滤波器

用在状态空间传播的随机样本(这些样本被称为“粒子”)来逼近系统状态的后验概率分布(PDF)

- 状态 $X_t = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_t\}$
- 观察 $Z_t = \{z_1, z_2, \dots, z_t\}$
- 马尔可夫假设: $P(\mathbf{x}_t | \mathbf{x}_{t-1})$, 即 \mathbf{x}_t 概率依赖于 \mathbf{x}_{t-1}

迭代过程: 每一步都保持一组 N 个具有权重 w_i 的采样 s_i :

$$S_t = \{(s_{ti}, w_{ti})\} \quad i=1, 2, \dots, N \quad \sum_i w_i = 1$$



11.4.1 典型技术

(1) 设已知时刻 $t-1$ 的一组加权样本 $S_{t-1} = \{s_{(t-1)i}, w_{(t-1)i}\}$

令权重的累积概率为

$$c_0 = 0$$

$$c_i = c_{i-1} + w_{(t-1)i} \quad i = 1, 2, \dots, N$$

(2) 重要性采样：对最有可能的样本加最大的权重

(3) 预测：使用有关 \mathbf{x}_t 的马尔可夫性质来推导 s_{tn}

(4) 校正：利用观察 Z_t 来获得 $w_{tn} = p(z_t | \mathbf{x}_t = s_{tn})$

(5) 返回到(2)，迭代 N 次

(6) 对 $\{w_{ti}\}$ 归一化，使得 $\sum_i w_i = 1$

(7) 输出对 \mathbf{x}_t 的最优估计

$$\mathbf{x}_t = \sum_{i=1}^N w_{ti} s_{ti}$$

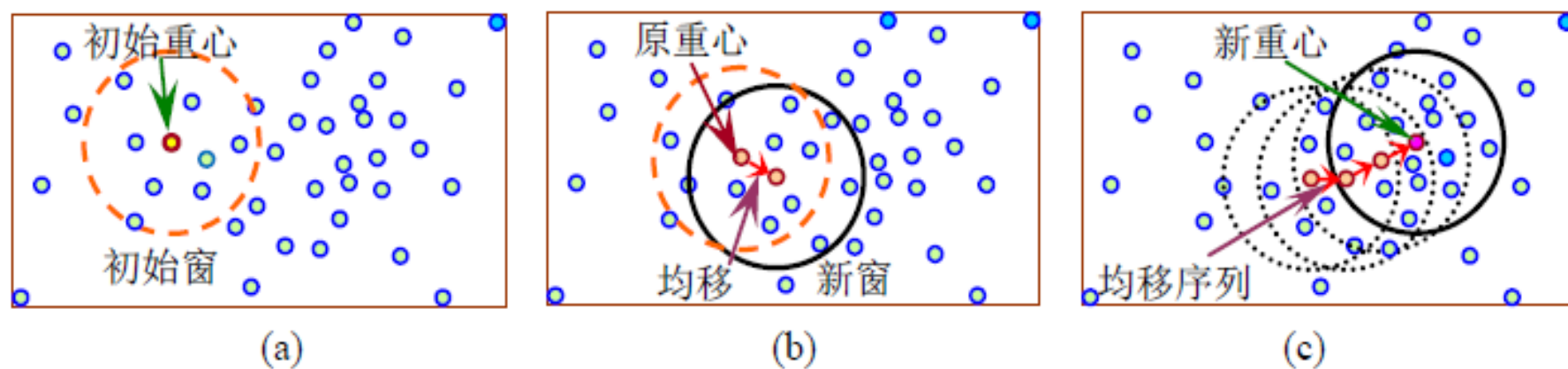
{例11.4.1}

11.4.1 典型技术

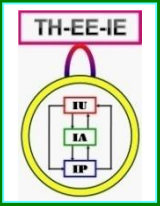
3. 均移和核跟踪

{P.284}

感兴趣区域对应跟踪窗口，不断地将目标模型在跟踪窗口内移动搜索，计算相关值最大的位置。这相当于确定聚类中心时，将窗口移到与重心位置重合（收敛）的工作



{P.74} 图 3.3.5 均移方法的原理示意



11.4.2 子序列决策策略

逐帧进行跟踪，可能的问题是做出决策所利用的信息比较少，且小的误差可能扩散而无法控制

将整个跟踪序列划分为若干个子序列，根据子序列提供的信息对其中的每帧做出全局最优的决策（是对逐帧跟踪的推广）：

- ① 将输入视频分成若干个子序列
- ② 在每个子序列中进行跟踪
- ③ 如果相邻子序列有重叠，则将它们的结果融合起来

{P.286}



联系信息



- ☞ 通信地址：北京清华大学电子工程系
- ☞ 邮政编码：100084
- ☞ 办公地址：清华大学，罗姆楼，6层305室
- ☞ 办公电话：(010) 62798540
- ☞ 传真号码：(010) 62770317
- ☞ 电子邮件：zhang-yj@tsinghua.edu.cn
- ☞ 个人主页：oa.ee.tsinghua.edu.cn/~zhangyujin/