

# 图 象 理 解

（第4版）

章毓晋

清华大学电子工程系 100084 北京

## 第4单元 研究示例

- 第13章 多传感器图象信息融合
- 第14章 基于内容的图象和视频检索
- 第15章 时空行为理解
  - 一些得到较多关注的研究领域
  - 结合利用不同传感器所获得的数据
  - 检索是各类视觉信息在全球得到广泛采集、传输和应用背景下一个新的研究领域
  - 图象理解需要充分掌握时空信息，分析人物行为，解释场景含义

第15讲

章毓晋 (TH-EE-4E) ZHANG YU JIN

第2页

## 第15章 时空行为理解

### 图象理解内容

判断场景中有哪些景物、它们随时间如何改变其在空间的位置、姿态、速度、关系等，并进而解释整个场景的态势氛围、发展趋势等含义

简言之，要在时空中把握景物的动作、确定动作的目的，并进而理解它们所传递的语义信息

动作检测和分类近期已得到很多关注和研究

高抽象层次的行为识别与解释（与语义和智能相关）研究还开展不多，技术在不停地发展更新中

第15讲

章毓晋 (TH-EE-4E) ZHANG YU JIN

第3页

## 第15章 时空行为理解

### 理解时空行为

#### ✓ 时空：客观

- 位置、轨迹、速度、外观、姿态、关系、…
- 对群体目标活动中的聚合、消散、分化、合并等动态演变现象，…

#### ✓ 行为：主观

- 举止、动向、态势、情感、…

第15讲

章毓晋 (TH-EE-4E) ZHANG YU JIN

第4页

## 第15章 时空行为理解

- 15.1 时空技术
- 15.2 时空兴趣点
- 15.3 动态轨迹学习和分析
- 15.4 动作分类和识别
- 15.5 活动和行为建模
- 15.6 主体与动作联合建模

第15讲

章毓晋 (TH-EE-4E) ZHANG YU JIN

第5页

## 15.1 时空技术

### 1. 新的领域

从1996年开始的图象工程综述

五大类：图象处理，图象分析，图象理解，技术应用，综述评论

图象理解：匹配融合，场景恢复，图象感知和解释，基于内容的图象和视频检索（2000年）

进入第二个十年时，图象理解大类中增加了一个新的小类 —— C5：时空技术（3-D运动分析，姿态检测，对象跟踪，行为判断和理解）

第15讲

章毓晋 (TH-EE-4E) ZHANG YU JIN

第6页

## 15.1 时空技术

### 从动作到行为的五个层次

- (1) 动作基元 (action primitives)
- (2) 动作 (action)
- (3) 活动 (activity)
- (4) 事件 (events)
- (5) 行为 (behavior)

第15讲

章毓晋 (TH-EE-IE) ZHANG YU JIN

第7页

## 15.2 时空兴趣点

### 1. 空间兴趣点的检测

使用线性尺度空间表达, 对图象建模

$$L^{\text{sp}}(x, y; \sigma_f^2) = g^{\text{sp}}(x, y; \sigma_f^2) \otimes f^{\text{sp}}(x, y)$$

高斯核

$$g^{\text{sp}}(x, y; \sigma_f^2) = \frac{1}{2\pi\sigma_f^2} \exp[-(x^2 + y^2) / 2\sigma_f^2]$$

Harris兴趣点 (角点) 检测

$$\begin{aligned} I^{\text{sp}}(\bullet; \sigma_f^2, \sigma_f^2) &= g^{\text{sp}}(\bullet; \sigma_f^2) \otimes \{[\nabla L(\bullet; \sigma_f^2)][\nabla L(\bullet; \sigma_f^2)]^T\} \\ &= g^{\text{sp}}(\bullet; \sigma_f^2) \otimes \begin{bmatrix} (L_x^{\text{sp}})^2 & L_x^{\text{sp}} L_y^{\text{sp}} \\ L_x^{\text{sp}} L_y^{\text{sp}} & (L_y^{\text{sp}})^2 \end{bmatrix} \end{aligned}$$

一个点的局部  
邻域的朝向分  
布协方差矩阵

第15讲

章毓晋 (TH-EE-IE) ZHANG YU JIN

第8页

## 15.2 时空兴趣点

### 2. 时空兴趣点的检测

检测在局部时空体中具有沿时和空轴都有图  
象值较大变化的位置

$$L(\bullet; \sigma_f^2, \tau_f^2) = g(\bullet; \sigma_f^2, \tau_f^2) \otimes f(\bullet)$$

空间方差 $\sigma_f^2$ 和时间方差 $\tau_f^2$

为了检测感兴趣点,  
在 $f$ 中搜索具有 $\mu$ 的显著  
本征值 $\lambda_1, \lambda_2, \lambda_3$ 的区域

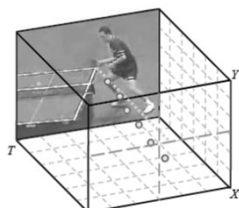


图 15.2.1 时空兴趣点示例

第15讲

章毓晋 (TH-EE-IE) ZHANG YU JIN

第9页

## 15.3 动态轨迹学习和分析

试图通过对场景中各个运动目标行为的描述  
和刻画来提供对监控场景状态的把握

先对目标进行检测并跟踪, 接着用所获得的  
轨迹自动地构建场景模型, 最后用该模型描述监  
控的状况并  
提供对活动  
的标注



图 15.3.1 动态轨迹学习和分析的流程图

第15讲

章毓晋 (TH-EE-IE) ZHANG YU JIN

第10页

## 15.3 动态轨迹学习和分析

在场景建模中, 先将有关事件发生的图象区域  
定义为**兴趣点** (POI), 然后在接下来的学习步骤  
中定义**活动路径** (AP)

在POI/AP学习中的主要工作包括:

- (1) **活动学习**: 通过比较**轨迹**来进行
- (2) **适应**: 研究管理POI/AP模型的技术
- (3) **特征选择**: 确定对特定任务正确的  
动力学表达层次

第15讲

章毓晋 (TH-EE-IE) ZHANG YU JIN

第11页

## 15.3 动态轨迹学习和分析

### 15.3.1 自动场景建模

### 15.3.2 学习路径

### 15.3.3 自动活动分析

第15讲

章毓晋 (TH-EE-IE) ZHANG YU JIN

第12页

### 15.3.1 自动场景建模

#### 1. 目标跟踪

在  $T$  帧视频中被跟踪的目标会生成一系列可推断出来的跟踪状态： $S_T = \{s_1, s_2, \dots, s_T\}$

#### 2. 兴趣点检测

感兴趣区域对应场景（地形）图中的结点  
入/出区域是目标进入或离开视场的位置

停止区域源于场景地标点，即目标在一段时期内趋于固定的位置

第15讲

章毓晋 (TH-EE-IE) ZHANG YU JIN

第13页

### 15.3.1 自动场景建模

#### 3. 活动路径学习

活动定义在开始和结束的两个感兴趣点之间

3种基本结构的主要区别包括输入的种类，运动矢量，轨迹/视频片段，以及运动抽象的方式

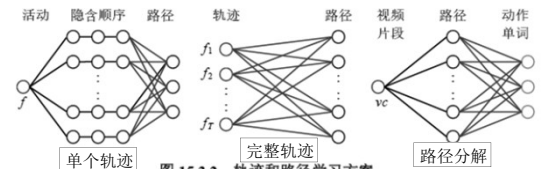


图 15.3.2 轨迹和路径学习方案

第15讲

章毓晋 (TH-EE-IE) ZHANG YU JIN

第14页

### 15.3.2 学习路径

#### 1. 轨迹预处理

- (1) 归一化：保证所有轨迹有相同的长度
- (2) 降维：将轨迹映射到新的低维空间

#### 2. 轨迹聚类

- ① 定义一个距离（对应相似性）测度
- ② 确定聚类更新的策略（非监督组合轨迹）
- ③ 进行聚类验证（根据判断准则优化）

第15讲

章毓晋 (TH-EE-IE) ZHANG YU JIN

第15页

### 15.3.2 学习路径

#### 3. 路径建模

路径模型是对聚类的紧凑表达（图模型推理）

(a) 考虑完整的路径，有平均的中心线，两边还有包络指示路径范围

(b) 将路径分解为子路径（表示成子路径的树）

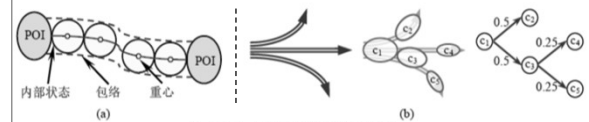


图 15.3.4 两种对路径建模的方式

第15讲

章毓晋 (TH-EE-IE) ZHANG YU JIN

第16页

### 15.3.3 自动活动分析

进行活动分析，以学习感兴趣的事件 {P.368}

- (1) 虚拟篱笆：一旦有入侵（位置信息）就触发分析
- (2) 速度分析：基于当前速度（动态信息）的预警
- (3) 路径分类：利用历史运动模式获得活动路径
- (4) 异常检测
- (5) 在线活动分析
- (6) 目标交互刻画

图 15.3.5 利用路径进行碰撞评估

第15讲

章毓晋 (TH-EE-IE) ZHANG YU JIN

第17页

### 15.4 动作分类和识别

基于视觉的人体动作分类和识别是对图象序列（视频）用动作（类）标号进行标记的过程

#### 15.4.1 动作分类

#### 15.4.2 动作识别

第15讲

章毓晋 (TH-EE-IE) ZHANG YU JIN

第18页

## 15.4.1 动作分类

### 1. 直接分类

将观察序列中所有帧的信息都加到单个表达中或对各帧分别进行动作的识别和分类（不特别关注时间域）

### 2. 时间状态模型

- ① 生成模型：学习观察和动作之间的联合分布，对每个动作类建模（考虑所有变化）
- ② 鉴别模型：学习在观察条件下动作类别的概率，并不对类别建模但关注类间的差别

第15讲

章毓晋 (TH-EE-IE) ZHANG YU JIN

第19页

## 15.4.1 动作分类

### 3. 动作检测

不显式地对图象中目标表达建模，也不对动作建模。它将观察序列与编号的视频序列联系起来，以直接检测（已定义的）动作

- ① 基于表观的方法：直接利用对图象中的前景、背景、轮廓、光流等的描述
- ② 基于人体模型的方法：利用人体模型表达行为人的结构特征，如将动作作用人体（部分）关节点序列来描述

第15讲

章毓晋 (TH-EE-IE) ZHANG YU JIN

第20页

## 15.4.2 动作识别

### 1. 整体识别

强调对整个人体目标或单个人体的各个部分进行识别（如基于人体的剪影）

### 2. 姿态建模

- (1) 基于表观的方法（动作姿态）
- (2) 基于人体模型的方法（人体姿态）
- (3) 基于3-D重构的方法（人体姿态）

- 稀疏型：基于时空Harris角点的时空兴趣点
- 稠密型：借助运动强度提取大量时空兴趣点

第15讲

章毓晋 (TH-EE-IE) ZHANG YU JIN

第21页

## 15.4.2 动作识别

### 3. 活动重建

动作导致姿态改变，将姿态定义为一个状态，借助状态空间法将状态之间通过转移概率来切换一个活动序列的构建可通过在对应姿态的状态之间进行一次遍历而得到

### 4. 交互活动

- ① 人与环境的交互
- ② 人际交互：将单人活动结合起来而得到。其中对单人活动可借助概率图模型来描述

第15讲

章毓晋 (TH-EE-IE) ZHANG YU JIN

第22页

## 15.4.2 动作识别

### 5. 群体活动

群体目标运动分析主要以人流、交通流以及自然界的密集生物群体为对象，研究群体目标运动的表达与描述方法，分析群体目标的运动特征以及边界约束对群体目标运动的影响

图 15.4.2 人流监控中对人数的统计

第15讲

章毓晋 (TH-EE-IE) ZHANG YU JIN

第23页

## 15.5 活动和行为建模

- (1) 获取输入视频或序列图象
- (2) 提取精练的底层图象特征
- (3) 从底层特征上升到中层动作描述
- (4) 从基本的动作出发进行高层语义解释

活动的层次要高于动作

图 14.5.1 动作和活动建模识别技术的分类图

第15讲

章毓晋 (TH-EE-IE) ZHANG YU JIN

第24页

## 15.5 活动和行为建模

对动作和活动的建模和识别常采用不同的技术

### 15.5.1 动作建模

### 15.5.2 活动建模和识别

第15讲

章毓晋 (TH-EE-IE) ZHANG YU JIN

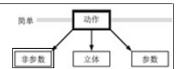
第25页

### 15.5.1 动作建模

#### 1. 非参数建模方法

从每帧中提取特征，并与存储的模板匹配

- (1) 2-D模板：运动检测，跟踪目标，对周期性强的动作进行动作识别
- (2) 3-D目标模型：对时空目标建立的模型
- (3) 流形学习方法：利用学习数据所在的流形可确定数据的固有维数，该固有维数的自由度较小，可帮助在低维空间设计有效的模型



第15讲

章毓晋 (TH-EE-IE) ZHANG YU JIN

第26页

### 15.5.1 动作建模

#### 2. 立体建模方法

将视频看作3-D立体像素（体素）的强度表达并将标准的图象特征扩展到3-D

- (1) 时空滤波：采用一组时空滤波器对**视频体**数据滤波。根据响应进一步推出特定的特征
- (2) 基于部件的方法：一个视频体可看作许多局部部件的集合体，各个部件有特殊的运动模式。由于本质上的局部性，对非稳态背景比较鲁棒



第15讲

章毓晋 (TH-EE-IE) ZHANG YU JIN

第27页

### 15.5.1 动作建模

#### 2. 立体建模方法

- (3) 子体匹配：子体是空间上一致的立体区域。它并不需要从尺度空间的极值点提取动作描述符而是检查两个局部时空块之间的相似度（通过比较两个块之间的运动）
- (4) 基于张量的方法：一个3-D时空体可以自然地看作一个有3个独立维的张量，这提供了一种整体匹配视频的直接方法

第15讲

章毓晋 (TH-EE-IE) ZHANG YU JIN

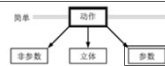
第28页

### 15.5.1 动作建模

#### 3. 参数建模方法

对运动的时间动态建模

- (1) 隐马尔科夫模型：状态空间（有限符号集合）的一种典型模型，对时间序列数据的建模很有效，有很好的推广性和鉴别性
- (2) 线性动态系统：比隐马尔科夫模型更一般化
- (3) 非线性动态系统：切换线性动态系统包括一组线性动态系统和一个切换函数，切换函数通过在模型间的切换来改变模型参数



第15讲

章毓晋 (TH-EE-IE) ZHANG YU JIN

第29页

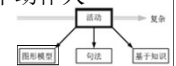
### 15.5.2 活动建模和识别

活动相比于动作，不仅持续时间长，而且大多数人们所关注的活动应用都包括多个动作人

#### 1. 图形模型

- (1) 信念网络：贝叶斯网络先将一组随机变量编码为局部条件概率密度（CPD），再对它们之间的复杂条件依赖性进行编码

对比只能编码一个隐变量的传统HMM，动态信念网络（DBN）可以对若干个随机变量之间的复杂条件依赖关系进行编码



第15讲

章毓晋 (TH-EE-IE) ZHANG YU JIN

第30页

## 15.5.2 活动建模和识别

### 2. 合成方法（句法）

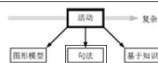
主要借助语法概念和规则来实现

#### (1) 语法

语法利用一组产生式规则描述处理的结构

产生式规则指出如何从词（活动基元）构建句子（活动），以及如何识别句子（视频）满足给定语法（活动模型）的规则

上下文自由语法（CFG）被用来对人体运动和多人交互进行结构化建模和识别



第15讲

章毓晋 (TH-EE-IE) ZHANG YU JIN

第31页

## 15.5.2 活动建模和识别

### 2. 合成方法（句法）

#### (2) 随机语法

随机上下文自由语法（SCFG）对上下文自由语法进行了**概率扩展**，更适合用于将实际中（数据有误差）的视觉模型结合起来

SCFG可用于对活动（其结构假设已知）的语义进行建模（提高了鲁棒性）

SCFG还被用来对多任务的活动（包含多个独立执行线程，断断续续相关交互的活动）建模

第15讲

章毓晋 (TH-EE-IE) ZHANG YU JIN

第32页

## 15.5.2 活动建模和识别

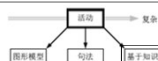
### 3. 基于知识和逻辑的方法

(1) 基于逻辑的方法：依靠严格的逻辑规则来描述一般意义上的领域知识以描述活动

**声明式模型**用场景结构、事件等描述所有期望的活动

活动模型包括场景中目标间的交互

虽然基于逻辑的方法提供了一个结合领域知识的自然方法，它们常包含耗时的对约束条件是否满足的审核



第15讲

章毓晋 (TH-EE-IE) ZHANG YU JIN

第33页

## 15.5.2 活动建模和识别

### 3. 基于知识和逻辑的方法

(2) 本体论的方法：本体可以标准化对活动的定义，允许对特定的配准进行移植，使不同的系统增强互操作性，以及方便地复制和比较系统性能

**视频事件竞赛工作会议**定义了6个视频监控的领域：① 周边和内部的安全；② 铁路交叉的监控；③ 可视银行监控；④ 可视地铁监控；⑤ 仓库安全；⑥ 机场停机坪安全

第15讲

章毓晋 (TH-EE-IE) ZHANG YU JIN

第34页

## 15.6 主体与动作联合建模

图结构 $G = (N, A)$ 表达视频 $f(x, y, t)$

结点集合 $N = (n_1, \dots, n_M)$ ，代表 $M$ 个体素（或 $M$ 个超体素），弧集合 $A(n)$ 代表 $N$ 中在某个 $n$ 的邻域中的体素集合

主体标记集合用 $X$ 表示（随机变量 $\{x\}$ ）

动作标记集合用 $Y$ 表示（随机变量 $\{y\}$ ）

主体-动作理解问题  $(x^*, y^*) = \underset{x, y}{\operatorname{argmax}} P(x, y | M)$  最大后验

第15讲

章毓晋 (TH-EE-IE) ZHANG YU JIN

第35页

## 15.6 主体与动作联合建模

考虑：多个不同类别的主体进行多个不同类别的动作

15.6.1 单标签主体-动作识别

15.6.2 多标签主体-动作识别

15.6.3 主体-动作语义分割

第15讲

章毓晋 (TH-EE-IE) ZHANG YU JIN

第36页



## 15.6.1 单标签主体-动作识别



由单个主体 $x$ 发起单个动作 $y$

### (1) 朴素贝叶斯模型

假设主体和动作是互相独立的，在动作空间中训练一组分类器

### (2) 联合乘积空间模型

利用主体空间 $X$ 和动作空间 $Y$ 生成一个新的标记空间 $Z$ 。乘积关系： $Z = X \times Y$

在联合乘积空间中，可以直接对每个主体-动作元组学习出一个分类器

第15讲

章毓晋 (TH-EE-IE) ZHANG YU JIN

第37页



## 15.6.1 单标签主体-动作识别



由单个主体 $x$ 发起单个动作 $y$

### (3) 三层次模型

统一了朴素贝叶斯模型和联合乘积空间模型

同时在主体空间 $X$ ，动作空间 $Y$ 和联合主体-动作空间 $Z$ 中学习分类器

不仅对主体-动作进行交叉建模，而且也对同一个主体发起不同动作和不同主体发起同一个动作进行建模

第15讲

章毓晋 (TH-EE-IE) ZHANG YU JIN

第38页



## 15.6.2 多标签主体-动作识别



多个主体发起了多个动作

许多主体可以发起同一个动作，但没有一个主体可以发起所有的动作

1个主体发起了2个以上的动作或2个以上的主体发起了1个动作（超过1/3）

表 15.6.2 数据库中主体、动作、主体-动作标签对应的视频段数量

	1	2	3	4	5
主体	2794	936	49	3	0
动作	2639	1037	99	6	1
主体-动作	2503	1051	194	31	3

第15讲

章毓晋 (TH-EE-IE) ZHANG YU JIN

第39页



## 15.6.3 主体-动作语义分割



任务是要在整个视频中为每个体素上的主体-动作寻找标签（最细粒度）

### (1) 朴素贝叶斯模型

分别处理两个类（主体、动作）的标签

### (2) 联合乘积空间模型

利用元组 $[x, y]$ 以联合考虑主体和动作

### (3) 双层次模型

考虑主体和动作变量的联系

第15讲

章毓晋 (TH-EE-IE) ZHANG YU JIN

第40页



## 15.6.3 主体-动作语义分割



任务是要在整个视频中为每个体素上的主体-动作寻找标签

### (4) 三层次模型

同时考虑类别内部的联系以及类别之间的联系  
它将联合乘积空间的结点与主体结点和动作结点全部结合起来

这个三层次模型考虑了在各个主体空间和各个动作空间、以及在联合乘积空间中的所有联系

先前3个基本的模型都是三层次模型的特例

第15讲

章毓晋 (TH-EE-IE) ZHANG YU JIN

第41页



## 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](mailto:zhang-yj@tsinghua.edu.cn)

✉ 个人主页：[oa.ee.tsinghua.edu.cn/~zhangyujin/](http://oa.ee.tsinghua.edu.cn/~zhangyujin/)

第15讲

章毓晋 (TH-EE-IE) ZHANG YU JIN

第42页