

视感知激励——多视觉线索集成的贝叶斯方法与应用

薛建儒*, 郑南宁*, 钟小品, 平林江

西安交通大学人工智能与机器人研究所, 西安 710049

* 同等贡献. 联系人, E-mail: jrxue@mail.xjtu.edu.cn

2007-08-31 收稿, 2007-12-25 接受

国家自然科学基金(批准号: 60405004, 60635050)、国家高技术研究发展计划(批准号: 2006AA01Z318, 2006AA01Z192)资助项目

摘要 视觉线索可看作视感知的一类冲激或称为激励. 在贝叶斯估计理论估计的框架内, 通过以概率形式描述视觉线索来研究视感知过程的认知机理, 探索一种多视觉线索信息统计集成方法. 这种方法在计算机视觉尤其是视频分析领域得到了广泛应用. 介绍多视觉线索贝叶斯集成的基本思想和一些最新的研究成果, 结合作者的研究工作讨论在贝叶斯概率模型框架下实现多视觉线索信息统计集成方法及相应的算法, 同时针对复杂背景下视频目标跟踪应用实例和多目标跟踪中的目标交错问题, 探讨了多视觉线索信息统计集成与视感知过程的关系, 并提出了多视觉线索贝叶斯集成方法及其应用中需要进一步探讨的新问题.

关键词

感知激励
视觉线索集成
贝叶斯分析
视频分析
目标跟踪

近年来, 视觉信息的分析与理解, 特别是一些基本问题如图像分割、感知组合、目标识别与跟踪、立体视觉与运动视觉等已成为数学、神经物理学与心理学等学科进行算法验证与科学实验的重要内容. 在解决视觉计算理论的基本问题过程中, 从不同角度出发的主要视觉计算理论与方法如调和分析与偏微分方程理论、贝叶斯理论与马尔科夫随机场、统计计算理论、基于图论的不同层次上图匹配与图语法、机器学习等都取得了长足的进展, 并在多图像特征或目标属性表达及其综合的集成研究中呈现出大一统的新型视觉计算理论框架的趋势. 另一方面, 视神经科学领域的研究者在单神经元与集群神经元水平上, 在视觉皮层的神经结构与功能、视觉神经信息编码与处理层次结构上的研究中取得了重要进展; 在生物视觉系统的注意转移机制与主动视觉感知的研究中也有了初步研究成果和发现. 得益于这两方面发展的强力驱动, 以认知机理为基础的新型视觉计算理论已成为计算机视觉学科发展的主要方向, 尤其是其中的多种视觉线索集成(或融合)已逐渐成为十分活跃的研究方向之一.

视神经科学中的视觉皮层功能专门化理论^[1,2]认为: 在对某一感兴趣目标感知过程中, 人类视觉系统

对目标的颜色、形状、纹理、运动及其他属性的信息处理由视觉皮层的各特定区域完成, 而整个认知任务的完成需要经不同传播途径分发和信息融合机制来集成这些分散区域的感知信息. 视觉线索就是指反映这些属性的信息, 它可看作视感知的一类冲激或称为激励. 由于视觉线索的不确定性以及海量的视觉数据, 各视觉线索的感知方法与融合机制引起了研究者的兴趣, 并且出现了 Dempster-Shafer、模糊逻辑以及神经网络、贝叶斯等多视觉线索集成方法. 作者认为以贝叶斯估计理论为基础, 用统计推理的方法研究以条件概率分布形式描述的视觉线索在完成视觉认知任务过程中的角色与贡献, 必然会促进视觉计算理论的发展. 大量的研究与实践结果也表明^[3-6], 多视觉线索的贝叶斯集成可以有效扩展机器视觉系统的应用范围, 提高系统精度, 增强系统适应性.

相对于如 Dempster-Shafer、模糊逻辑以及神经网络等信息集成方法, 视觉线索贝叶斯集成具有以下优势: (1) 联合概率密度可以有效表征各线索间的相关性, 并且为各线索先验知识的引入提供了有效途径; (2) 便于对时间序列信号建模, 如隐马尔可夫模型(HMM)、动态贝叶斯网络(DBN)等模型对于时间序列的时-空相关性具有较强的描述能力; (3) 可以概率形式方

便引入信息论、模式识别理论中的模型与算法。

视觉线索贝叶斯集成目的在于: (1) 充分利用各视觉线索的先验知识以及它们间的互信息降低估计的不确定性; (2) 利用线索之间的互补性和相关性提高系统的鲁棒性。基于单个视觉线索的算法或模型在实际应用中具有局限性, 无法保证实际应用时该线索的可靠性。相反, 互补的多种视觉线索可以有效提高视觉系统的鲁棒性和适应能力。因此多视觉线索集成的两个关键问题分别是根据视觉线索的不确定性选择要参与集成的线索集合和对所选定各视觉线索间的相关性进行统计建模。

文献[2,3]对线索不确定性的研究发现, 如果图 1 中某一视觉线索 z 的概率分布的方差较小, 那么基于该线索的估计 x (假设要估计变量为 1D 变量) 可靠性高, 因为小方差概率分布会降低估计的不确定, 因此集成时需要给小方差分布的视觉线索较大的权重。另一方面, 视觉线索之间也存在着协作或竞争关系, 因此在视觉线索集成时需要考虑视觉线索间的相关性。

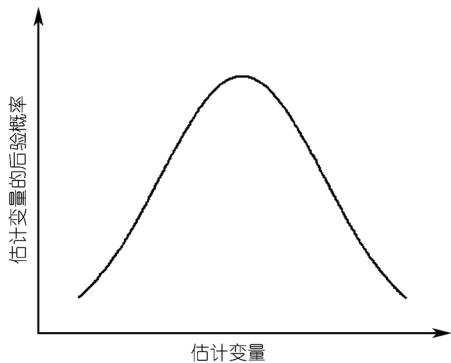


图 1 采用某线索时估计变量的条件概率分布

1 问题描述

视觉线索反映了感兴趣目标的某种属性, 在视觉计算理论中, 视觉线索的表现形式多样: 图像传感器输出的原始数据; 原始数据经过特定滤波器的滤波响应; 描述视觉模式的模型或方法。每条视觉线索具有不同的适用范围、在不同成像条件下对目标具有不同的分辨能力, 并且视觉线索间相关程度各异。因此设计基于多视觉线索集成的视觉系统时, 必须考虑两个相耦合的基本问题: (1) 线索选择即如何选择线索以形成互补且具有较强分辨力的集合; (2) 线索集成策略(或方法), 即如何最大程度地有效利用所选择的视觉线索。

若视觉线索之间相关性较弱, 集成时需要定义测度以有效衡量各线索的分辨能力, 然后在经过排序后的线索中选择具有较强分辨力的视觉线索形成线索集合; 若视觉线索之间的相关性较强, 选择视觉线索时必须同时考虑候选视觉线索的分辨能力及它与已选视觉线索之间的相关性。对此问题常用的方法是贪婪迭代算法^[4,5], 该算法的基本思想为: 假设经过 k 次迭代后已得到含有 k 个线索的集合 $Z_k = \{z_1, \dots, z_k\}$, z_i 表示集合中第 i 个线索, 则在第 $k+1$ 次循环中选择线索 z_{k+1} 使得在增加了该线索的集合 $Z_{k+1} = \{z_1, \dots, z_{k+1}\}$ 条件下, 待估计量 x 的概率分布 $p(x|Z_{k+1})$ 与估计量 x 实际概率分布 $q(x)$ 之间的距离最小, 即

$$z_{k+1} = \arg \min_z D(p(x|Z_k, z), q(x)), \quad (1)$$

其中 $D(\cdot)$ 是两个概率分布之间的距离函数, 一般使用 Kullback Leibler(KL)距离; 由于实际分布 $q(x)$ 未知, 一般用 $\bar{q}_k(x) = p(x|Z_k)$ 代替。

显然, 贪婪迭代算法也适用于线索弱耦合情况。然而由于视觉线索通常携带大量信息, 并且各视觉线索间相关性强弱未知, 因此联合条件分布 $p(x|Z_k)$ 的解析形式通常难以得到。现有的视觉线索贝叶斯集成方法均立足于对此分布以不同方式进行最大程度地逼近。

给定 n 个视觉线索的集合 $Z = \{z_1, \dots, z_n\}$, 待估计量 x 的最优估计由最大后验准则(MAP)得到

$$\begin{aligned} \hat{x}^{\text{MAP}} &= \arg \max_x p(x|z_1, \dots, z_n) \\ &= \arg \max_x \frac{p(z_1, \dots, z_n|x)p(x)}{\int p(z_1, \dots, z_n|x)p(x)dx}, \end{aligned} \quad (2)$$

其中 $p(z_1, \dots, z_n|x)$ 为视觉线索的联合条件似然, $p(x)$ 为待估计量的先验分布。(2)式表示所有线索需汇聚在一起获得联合似然 $p(z_1, \dots, z_n|x)$, 然后由融合中心采用 MAP 准则作最终决策, 这一结构属于集中式融合, 能提供最优的集成性能, 因此可以作为比较各种集成方法性能的基准。

视觉皮层功能专门化理论认为, 视觉线索集成实质上是一个分布式融合过程: 先对各视觉线索单独处理以得到各线索条件似然 $p_i(z_i|x)$, 然后利用集成策略函数 $f(\cdot)$ 进行综合,

$$\begin{aligned} p(x|z_1, \dots, z_n) &\approx f(p_1(z_1|x)p(x), \\ &\quad \dots, p_n(z_n|x)p(x); w) \end{aligned} \quad (3)$$

或

$$p(x|z_1, \dots, z_n) \approx f[p_1(z_1|x), \dots, p_n(z_n|x); w, p(x)], \quad (4)$$

其中集成策略函数 $f: [0, 1]^n \mapsto [0, 1]$, w 是函数 f 的参数向量.

基于上述讨论, 可将视觉线索贝叶斯集成问题描述为: 假设根据给定视觉任务选择了 n 条视觉线索, 并且各视觉线索的似然密度为 $p_i(z_i|x)$, 集成策略函数为由参数变量 w 确定的 $f(\cdot)$, 则估计量的后验概率分布具有(3)或(4)式的形式, 并可根据 MAP 或 MMSE 准则得到最优估计及相应的参数 w . (2)式给出了待估计量的 MAP 估计. 值得指出的是, 若 Z 中各视觉线索为原始传感数据或滤波响应时, 此时的线索集成实质为数据融合; 当 Z 中各视觉线索为不同模型时, 此时的视觉线索集成为多模型融合.

不难发现, 视觉线索的贝叶斯集成与融合估计密切相关. 首先, 估计融合的研究基础是经典的线性估计理论, 所取得较为成熟的理论结果大多基于线性模型与高斯噪声假设; 而基于贝叶斯的多视觉线索集成对先验知识、似然函数和后验估计均使用概率分布表示, 已具备了突破线性高斯假设的约束条件.

其次, 从实现结构上看, 多线索贝叶斯集成采用分布式结构, 遵循视觉皮层功能专门化理论, 各视觉线索独立工作, 集成在融合中心进行; 而融合估计方法具有集中式和分布式两种实现结构. 从应用范围看, 融合估计在目标跟踪领域应用较为广泛, 因此也常称为跟踪融合; 而视觉线索贝叶斯集成可广泛应用于分割、视觉模式表征、跟踪等基本视觉问题^[6].

2 线索贝叶斯集成方法

近年来, 出现了大量视觉线索集成方法^[7~13]. 作者根据视觉线索和集成策略之间的异同, 将视觉线索贝叶斯集成方法分为四类: 线性意见汇聚、对数意见汇聚、模型融合以及概率图模型集成, 本节将对这四类方法进行分析比较. 为简化起见, 本文后续部分中的视觉线索简称为线索.

2.1 线性意见汇聚

文献[14]中, 对基于各线索的意见进行线性汇聚 (linear opinion pooling, LOP), 集成函数为线性加权函数,

$$p(x|z_1, \dots, z_n) = \sum_{i=1}^n w_i p_i(x|z_i), \quad (5)$$

其中权重 $w_1, \dots, w_n$ 非负且 $\sum_{i=1}^n w_i = 1$, 并有

$p_i(x|z_i) \propto p_i(z_i|x) p_i(x)$, LOP 方法如图 2(a)所示.

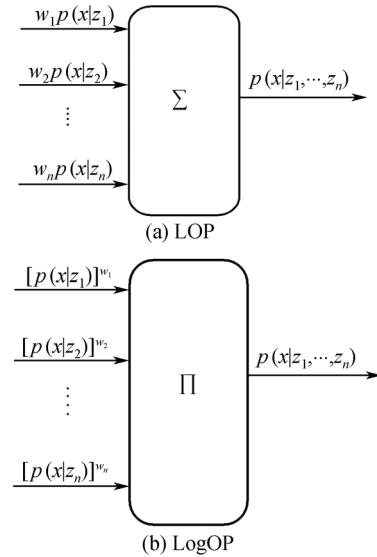


图 2

(a) 线性意见汇聚; (b) 对数意见汇聚

LOP中各线索以概率分布加权和形式实现集成, 并满足边缘属性(marginalization property)^[15]. 其中权重 w_i 反映第 i 个线索的对集成意见的可靠性. LOP可对连续状态进行估计. 如何决定权重 w_i 成为一个影响决策质量的关键问题, 不合适的权值将导致出现某一线索“独裁”的现象. 常用的确定线索权值的方法有以下几种^[16]:

(i) 均等权值, 即 $w_i = 1/n$, $i = 1, \dots, n$. 均等权值说明意见汇聚系统对各线索的可靠性没有先验知识, 即对所有线索平等对待.

(ii) 权值正比于排序结果, 即 $w_i = s_i / \sum_{j=1}^n s_j$, $i = 1, \dots, n$, s_i 为线索 i 的重要性排序序号.

(iii) 权值正比于自评估值, 即 $w_i = r_i / \sum_{j=1}^n r_j$, $i = 1, \dots, n$, r_i 为线索 i 的自评估值. 这一方法的合理性在于各线索具有各自的“专业领域”, 自评估反映了线索与该领域的相关程度.

(iv) 从训练样本中学习权值. 以权值为参数, 在样本集上构造似然函数, 使用最大似然等准则通过学习可得到权值.

实际应用中, Spengler等人^[11]使用颜色、对比度和运动等视觉线索将LOP的成功应用于嘈杂背景下头部跟踪.

2.2 对数意见汇聚

文献[15~17]的研究提出意见对数汇聚(logarithmic opinion pooling, LogOP)相对于 LOP 可能更符合人类的认知过程的观点. 与 LOP 的加权平均集成方式不同, LogOP 采用乘性集成函数, 如图 2(b) 示.

$$p(\mathbf{x} | \mathbf{z}_1, \dots, \mathbf{z}_n) = \frac{\prod_{i=1}^n [p(\mathbf{x} | \mathbf{z}_i)]^{w_i}}{\int_{\mathbf{x}} \prod_{i=1}^n [p(\mathbf{x} | \mathbf{z}_i)]^{w_i} d\mathbf{x}} = \kappa_1 \prod_{i=1}^n [p(\mathbf{x} | \mathbf{z}_i)]^{w_i}, \quad (6)$$

$$p(\mathbf{x} | \mathbf{z}_1, \dots, \mathbf{z}_n) = \frac{\prod_{i=1}^n [p(\mathbf{z}_i | \mathbf{x})]^{w_i} p(\mathbf{x})}{\int_{\mathbf{x}} \prod_{i=1}^n [p(\mathbf{z}_i | \mathbf{x})]^{w_i} p(\mathbf{x}) d\mathbf{x}} = \kappa_2 \prod_{i=1}^n [p(\mathbf{z}_i | \mathbf{x})]^{w_i} p(\mathbf{x}). \quad (7)$$

其中 κ_1, κ_2 为归一化常量, w_i 为第 i 个线索的权重指数, 不难发现, 对(6)式两边取对数则得到与(5)式一致的形式, 因此称为意见对数汇聚.

权值选择影响 LogOP 集成的效果. 一般地, $w_1, \dots, w_n$ 的取值需满足特定条件, 典型条件是 $\sum_{i=1}^n w_i = 1$. 当 $w_i = 1/n, i = 1, \dots, n$, (6)式意味着集成意见为各线索意见的乘积, 此时意见汇聚又称为独立意见汇聚(independent opinion pooling)^[17]; 若某线索的权值为零值, 则表示该线索对集成无贡献; 若某线索 z_j 对估计 $\tilde{x}$ 的意见为 $p(\tilde{x} | z_j) = 0$, 则 $\tilde{x}$ 被彻底否决, 这是在许多应用场合中不希望发生的情况. 特别地当 $w_i = 1, i = 1, \dots, n$ 时, 采用(7)式的似然集成方式既简便又实用, 被称为独立似然汇聚(independent likelihood pooling)^[17]或朴素贝叶斯集成. 不难发现独立似然汇聚要求线索间相互独立. 若各线索似然 $p(z_i | \mathbf{x})$ 以及先验 $p(\mathbf{x})$ 均为高斯分布, 则该集成策略的 MAP 估计等价于各线索 MAP 估计的线性加权^[18,19], 即加权选举法(weighted voting).

文献[20]对加权和集成策略(或 LOP)与乘性策略(或 LogOP)对误差的敏感性进行了分析, 研究结果表明 LogOP 对噪声较为敏感, 而 LogOP 能有效提高估计的可靠性. 图 4(b) 对此给出了解释. 根据这一发现, 作者提出了一种集成策略自适应方案^[6], 即根据各视觉线索的可靠性, 集成策略在 LOP 与 LogOP 间实现自适应切换^[6]. 该方法较好地解决了复杂场景中目标跟踪问题, 本文第 3 节的应用部分对此方法给出了较为

详细的说明.

2.3 模型融合

当视觉线索为描述现象的模型或方法时, 多视觉线索的集成又可称作模型融合. 以 $z_1, \dots, z_n$ 表示模型标号, y 为观测数据, 则模型融合可表示如下:

$$p(\mathbf{x} | y) = \sum_i p(\mathbf{x} | y, z_i) p(z_i | y), \quad (8)$$

其中 $p(\mathbf{x} | y, z_i)$ 为给定观测 y 时从模型 z_i 获得的意见; 模型概率 $p(z_i | y)$ 为给定观测 y 时模型 z_i 发生的概率. 显然若 $p(z_i | y)$ 定义为与 y 无关的常量, 则模型融合将退化为 LOP.

模型融合被广泛应用于鲁棒控制问题, 比较典型应用是目标跟踪的多模型方法^[21,22]. 对于跟踪问题, 以 t 表示离散时间标记, 以 $y_i(t)$ 表示模型 $z_i(t)$ 所产生的观测, 则全部观测为 $\mathbf{y}(t) = \{y_1(t), \dots, y_n(t)\}$, 历史观测 $Y^t = \{\mathbf{y}(1), \dots, \mathbf{y}(t)\}$, 多模型跟踪问题转化为估计后验密度 $p(\mathbf{x}(t) | Y^t)$ 且可分解为

$$p(\mathbf{x}(t) | Y^t) = \sum_i p(\mathbf{x}(t) | Y^t, z_i(t)) p(z_i(t) | Y^t) = v \sum_i p(\mathbf{y}(t) | z_i(t), \mathbf{x}(t)) p(\mathbf{x}(t) | z_i(t), Y^{t-1}) p(z_i(t) | Y^t), \quad (9)$$

其中模型条件预测 $p(\mathbf{x}(t) | z_i(t), Y^{t-1})$ 可进一步分解

$$p(\mathbf{x}(t) | z_i(t), Y^{t-1}) = \sum_j p(\mathbf{x}(t) | z_i(t), z_j(t-1), Y^{t-1}) \times p(z_j(t-1) | z_i(t), Y^{t-1}) \approx \sum_j p(\mathbf{x}(t) | z_i(t), z_j(t-1), \hat{\mathbf{x}}(t-1)) p(z_j(t-1) | z_i(t), Y^{t-1}), \quad (10)$$

其中 v 为归一化常量, $\hat{\mathbf{x}}(t-1)$ 为 $t-1$ 时刻的估计值, $p(\mathbf{x}(t) | z_i(t), z_j(t-1), \hat{\mathbf{x}}(t-1))$ 可为任意递推滤波过程, $p(z_j(t-1) | z_i(t), Y^{t-1})$ 称为混合概率.

作者提出一种观测预测的方法^[22], 将模型融合推广到每个模型对应观测 $y_1(t), \dots, y_n(t)$ 各不相同的情况:

$$\tilde{\mathbf{y}}(t) = \sum_i p(z_i(t) | Y^{t-1}) \tilde{\mathbf{y}}_i(k), \quad (11)$$

$$\tilde{\mathbf{y}}(t) = \sum_i p(z_i(t) | Y^{t-1}) \tilde{\mathbf{y}}_i(k), \quad (12)$$

其中 $\tilde{\mathbf{y}}_i(k)$ 表示模型 z_i 的观测预测, $p(z_i(t) | Y^{t-1})$ 为模型 z_i 的预测概率. 该方法已被成功应用于易混淆运动员的跟踪问题.

2.4 概率图模型

前面已经指出, 影响线索集成性能的一个重要因素是线索间的相关性. 近十多年来, 概率图模型 (Graphical Model) 如贝叶斯网络、马尔可夫随机场等的研究^[23~28]取得一定进展并在视觉计算中得到了大量应用. 概率图模型适合于线索相关性建模体现在三方面: 第一, 易于描述各视觉线索之间的相关性; 第二, 概率推理可通过线索间信息交互实现, 不受线索独立性假设的局限; 第三, 概率图模型从建模到推理均具有比较成熟的理论与技术^[26].

给定图 3 所示的无向图 $G=(V, E)$, 其顶点集 V 和边集 E 由变量节点 x_i 和证据或线索节点 z_j 组成. 仅考虑由点对的簇组成的邻域系, 由 Hammersley-Clifford 定理可知, 变量节点联合概率分布可分解如下:

$$\begin{cases} p(x) = \frac{1}{R} \cdot \prod_{s \in V} \psi_s(x_s) \cdot \prod_{(s,t) \in E} \psi_{s,t}(x_s, x_t), \\ R = \sum_x \prod_{s \in V} \psi_s(x_s) \cdot \prod_{(s,t) \in E} \psi_{s,t}(x_s, x_t). \end{cases} \quad (13)$$

其中 ψ_s , $\psi_{s,t}$ 分别为一元簇和二元簇兼容函数, 它们刻画了簇内节点间的相关性, R 为拆分函数. 定义各兼容函数为

$$\begin{cases} \psi_s(x_s) = p_s(x_s) \cdot p(z_s | x_s), \\ \psi_{s,t}(x_s, x_t) = p(x_s, x_t) / [p(x_s)p(x_t)]. \end{cases} \quad (14)$$

其中 $p_s(x_s)$ 为节点 s 的边缘概率.

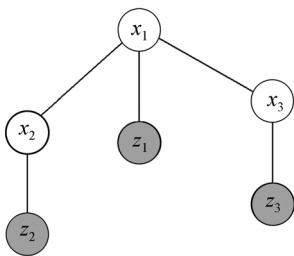


图 3 概率图模型

信任传播算法(belief propagation, BP)^[26,29,30]是一种常用的求解无环图中各节点概率的方法. BP算法中, 消息更新与节点信任计算公式为

$$M_{us}^{(k)}(x_s) \propto \int_{x_u} \psi_{us}(x_u, x_s) \cdot \psi_s(x_s) \cdot \prod_{v \in \Gamma(u) \setminus s} M_{vs}^{(k-1)}(x_u), \quad (15)$$

$$p_s(x_s) \propto \psi_s(x_s) \cdot \prod_{u \in \Gamma(s)} M_{us}^{(k)}(x_s), \quad (16)$$

其中 $\Gamma(s)$ 是节点 s 的邻域节点集, $\Gamma(u) \setminus s$ 为除节点 s 以外节点 u 的邻域节点集; $M_{us}^{(k)}$ 表示利用和积(sum-product)算法经过 k 次迭代计算得到的从节点 u 向节点 s 传递的消息. BP 算法也可用于环形图的推理, 即环形信任传播(Loopy BP). 除 BP 算法外, 最大积(max-product)算法^[25]也可用于图模型 MAP 估计的计算.

标准的 BP 算法要求离散变量, 且状态有限. 但在视觉任务中, 要估计的变量通常为连续变量并且维数较高, 而对高维连续变量的状态空间完全离散化不可行, 因此(15)式的积分计算成为阻止 BP 应用的一个障碍. 近年来研究者在标准 BP 算法中引入重要性抽样技术, 使用带权值的粒子集合替代(15)式的消息, 代表性研究工作有文献[24,31]. 但是, 这些方法中的 Monte Carlo 计算严重依赖于样本的抽样位置. 为了弥补节点信任信息缺失的不足和提高抽样样本的有效性, 作者提出了分层抽样信任传播算法^[32,33]求解状态估计, 并成功应用于多目标跟踪问题. Roth 等人^[34]提出基于极大熵原理的 Gibbs 模型^[4]集成滤波器组的响应以构造用于贝叶斯目标跟踪的似然模型. Ren 等人^[35]使用条件随机场在底层、中级以及高级 3 个层次上集成多种视觉线索. 这些方法为基于概率图模型求解视觉线索集成问题提供了新思路.

2.5 其他方法

除了以上提到的几种方法以外, 研究者还在信息论基础上提出最大化互信息的多线索贝叶斯集成方法^[13]; Pan 等人^[36]在最大互信息框架下提出一种通过选取辅助函数来获取线索相关信息的结构; 在能量模型的基础上, Chakraborty 等人^[37]利用博弈原理集成区域线索和边界线索用以分割特定图像, Patras 等人^[38]构造了一个类似于 EM 的迭代推理算法集成区域标记线索和运动分割线索并完成视频场景分割问题.

3 线索贝叶斯集成的应用

近年来视觉线索集成在计算视觉领域得到了高度重视. 视觉目标跟踪技术得到了广泛地应用, 视频中丰富的视觉线索, 迫切需要引入视觉线索集成以提高目标跟踪的精度与稳健性. 许多研究者对此进行了大量的研究, 作者^[6]提出了一种具有四层融合框架的目标鲁棒跟踪系统; Sigal 等人^[39]使用概率图模

型融合来自关节链目标各个部件的信息,实现跟踪具有高维状态的关节链目标;Beal等人^[40]使用概率图模型把音频线索与视频线索集成起来,能更准确地在会议室定位讲话者;Khan等人^[41]使用LogOP技术集成颜色、运动以及空间线索用于视频场景分割;Wang等人^[42]在像素级构建条件随机场来集成颜色线索和运动线索,分割运动目标以及阴影。

本节将从3个具有挑战性的视觉目标跟踪实例出发,进一步对前述的视觉线索贝叶斯集成方法进行分析比较以及讨论如何根据视觉任务合理选择。作者按照视频分析问题难点将视觉目标跟踪应用分为强干扰目标跟踪、强机动目标跟踪、以及多目标跟踪3大类,以下各小节将分析这3类跟踪问题难点以及相应的多视觉线索贝叶斯集成解决方案。

3.1 强干扰目标跟踪

场景中的强干扰是对视频目标跟踪方法的一大挑战,如相似目标靠近或遮挡、背景嘈杂、以及光照强烈变化所致的目标外观变化等。不同的干扰会使得不同的视觉线索失效,因此可靠、稳定的目标跟踪要求各种视觉线索协同工作。

不同的视觉线索贝叶斯集成方法具有不同适用场合和特点。如图4(b)示,LogOP可提高集成可靠性但对噪声敏感,因而容易丢失目标;LOP对噪声不敏感但集成精度不高。实际应用时需要根据环境选择合适的集成技术。作者提出一种基于粒子滤波器的集成策略自适应切换机制^[61]:以样本二阶中心矩的

Frobenius 范数来度量线索的不确定性,若该值过大则切换到强鲁棒性的LOP,否则使用高可靠性的LogOP。

为验证此观点,用一段双手交叉运动视频和一段足球场运动员运动的视频分别对基于LOP和LogOP集成策略的目标跟踪方法进行测试。在双手运动视频中,两只非常相似的手作上下交错运动,选取被遮挡的手为跟踪目标,其中3帧跟踪结果见**错误!未找到引用源。**(a),从上至下分别使用LogOP集成、LOP集成以及作者提出的集成策略自适应方法,从左至右分别为第12/28/49帧。

图5中左上角显示视频帧号,左下角为当前帧所使用的集成方式,底部黑色突起区域反映了粒子在横坐标系投影的分布情况。对于足球运动员视频,由于存在外观相似队友相互靠近的情况,因此稳定跟踪运动员非常困难,图5(b)从左至右分别显示第61/79/111帧的集成策略自适应跟踪结果。靠近的队友成为跟踪器的强烈干扰,此时把集成方式切换到强鲁棒性的LOP方式,如图中第61帧;离开干扰源后,再切换到高可靠性的LogOP方式,如第111帧,最终实现了目标稳定跟踪。

不同线索具有不同的适用条件。例如在光照较弱的环境下目标的颜色信息的分辨能力比轮廓信息下降要快。为此,作者提出了一种应用概率图模型建立线索相关模型并以序贯均值场进行推理的线索集成方案^[33]。图中节点间的兼容函数为混合高斯函数^[43],且兼容函数非对称。第一个实验使用一段长时

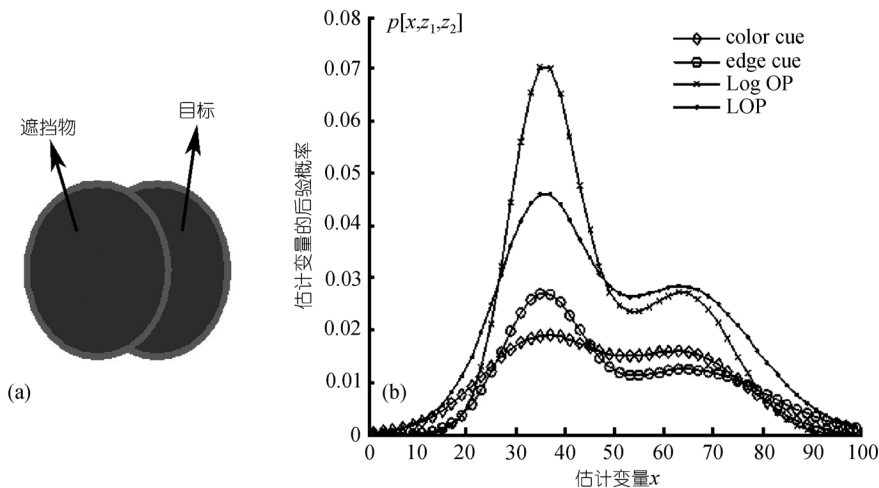
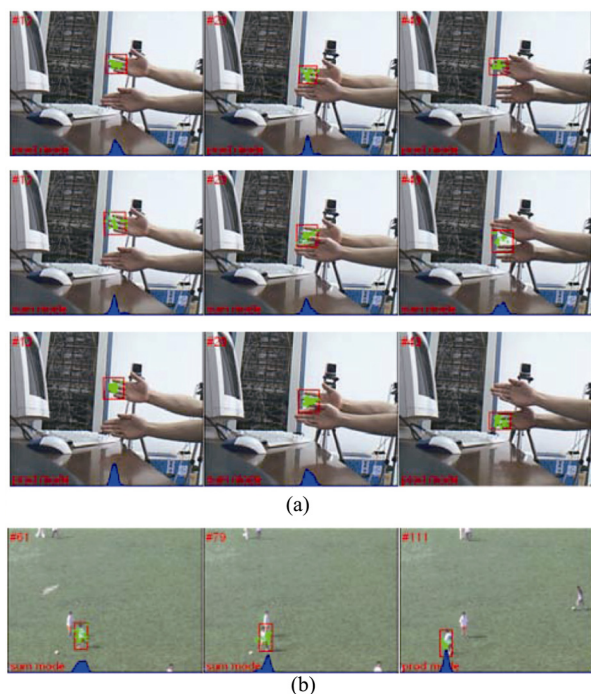


图4

(a) 目标被遮挡示意图; (b) 独立使用颜色、边缘线索得到的估计结果,以及分别使用LOP及LogOP集成策略对颜色与边缘线索进行集成得到的估计结果

图 5 相似目标干扰跟踪效果¹⁾

(a) 双手视频跟踪结果. 第 1, 2 行分别为使用 LOP 和 LogOP 集成规则的跟踪器跟踪结果, 第 3 行为使用本文提出的自适应集成规则的跟踪器跟踪结果; (b) 使用自适应集成规则跟踪足球运动员的跟踪结果

间视频序列, 其中设计了许多跟踪的难点, 如背景嘈杂、背景光照变化、目标大小和速度都在缓慢变化等. 图 6(a)中白色椭圆给出了 3 种集成方法的跟踪结果, 每行从左至右分别为序列的第 159/201/221 帧, 3 种策略分别是: LOP(第 1 行)、LogOP(第 2 行)与概率图模型集成算法(第 3 行). 实验中线索参数设置详见文献[33]. 实验发现, 噪声被乘性放大因此 LogOP 集成方法对噪声敏感; 而 LOP 对噪声具有更强的容忍性, 但损失了精度因此在目标加速运动时将偏离真实状态, 最终这两种方法都无法可靠跟踪. 图模型集成方法跟踪效果如图 6(a)第 3 行示, 其中短虚线、长虚线和实线分别为形状、颜色和亮度变化线索的跟踪情况.

显然, 这 3 条线索在图模型集成方法指导下相互合作从而集成结果更鲁棒、更精确. 第 2 个实验使用了一段部分遮挡视频序列²⁾, 应用概率图模型线索



图 6 环境干扰目标跟踪效果

(a) 白色椭圆给出了 3 种集成方法的跟踪结果, 每行从左至右分别为序列的第 159/201/221 帧, 3 种策略分别是: LOP(第 1 行)、LogOP(第 2 行)与概率图模型集成算法(第 3 行); (b) 部分遮挡情况下概率图模型线索集成方法的跟踪结果

集成方法跟踪头部的效果如图 6(b)示, 实验表明该集成方式对部分遮挡目标同样有效.

3.2 强机动目标跟踪

视频目标的强机动性也是跟踪算法的一大挑战, 通过多个运动模型集成来处理视频目标运动模型不确定性的例子很多^[44]. 作者把线索集成方法应用到运动员跟踪问题中, 结果如图 7(a)所示. 在实验视频序列中, 以相同队服出现的队友会干扰观测的正确分配, 此时运动信息的重要性高于目标的形态信息. 然而运动员的运动复杂使运动建模困难, 我们选取 3 种线性运动模型: 常速运动(CV)、低噪声的常加速运动(LowCA)以及高噪声的常加速运动(HighCA), 然后采用贝叶斯集成方法来逼近运动员真实的非线性运动.

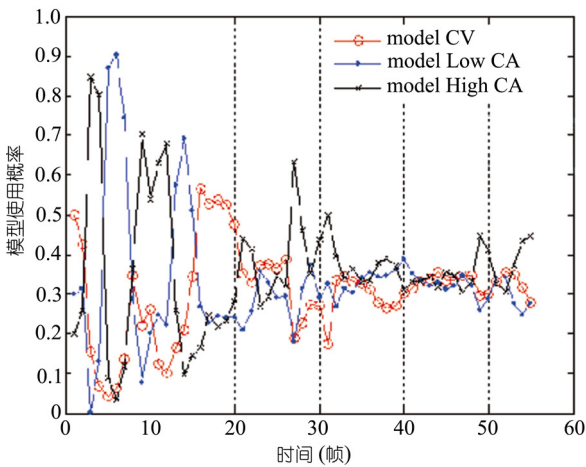
1) 由于篇幅关系, 此处只使用其中 3 帧图像作为对比, 相关的跟踪效果视频请利用以下链接下载: <http://aiar.xjtu.edu.cn/~zhong/demo/demo-dbhand.rar>; <http://aiar.xjtu.edu.cn/~zhong/demo/demo-fbplayers.rar>

2) 该序列为标准测试序列: <http://robotics.stanford.edu/birch/headtracker/seq/>

图 7(a)中, 第 1、第 2 行分别是使用 Mean Shift 方法和 Mean Shift + Kalman 方法跟踪的结果, 不难看出 Mean Shift 方法在遇到相似目标时会出现误跟, 而 Mean Shift + Kalman 也由于运动模型过于单一导致最终的跟踪失败, 只有多模型线索集成方法能始终跟踪预定目标, 如第 3 行所示, 其中 3 个线索的概率变化如图 7(b)所示, 可见不同模型线索在目标运动的不同阶段贡献不同。



(a)



(b)

图 7 多模型跟踪实验结果各运动模型使用概率
随时间变化曲线

(a) 第 1、第 2 行分别为使用 Mean Shift 方法和 Mean Shift + Kalman 方法跟踪的结果, 第 3 行为使用多模型方法跟踪的结果; (b) 3 种运动模型使用概率随时间变化的曲线

3.3 视频多目标跟踪

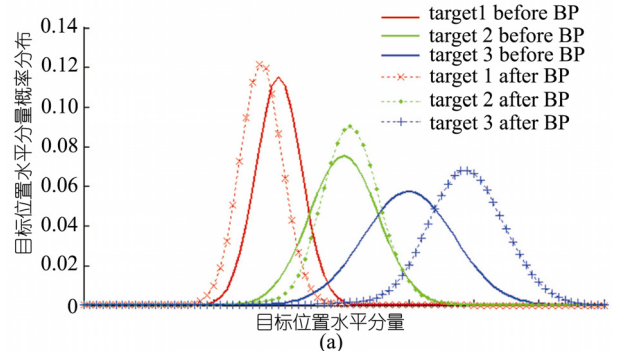
视频中多目标跟踪问题是近几年最受关注的难点之一^[45], 传统的雷达多目标跟踪方法通常把跟踪

问题划分为数据关联和状态估计两个部分, 然而数据关联中假设个数会随目标数目的增加呈指数级增长, 因此可能出现组合爆炸问题. 与传统雷达多目标跟踪相比, 视频多目标跟踪更是充斥着复杂背景干扰、遮挡等多种不确定因素. 常用的求解方法是序贯贝叶斯方法, 其中代表性工作有: MacCormick 等人^[46]提出的概率唯一性原则解决多目标遮挡问题, Tao 等人^[47]使用一种高效的分级算法在高维状态空间进行逼近推理.

作者认为概率图模型能集成多目标的检测信息以及交互信息, 进而实现多目标的鲁棒跟踪^[32,48]. 所提出方法的基本思想是: 使用耦合的马尔可夫随机场描述多目标状态以及多目标间的相互作用, 并应用序贯分层抽样信任传播算法^[32,33,48]估计多目标的状态.

使用 3 个观测为高斯分布的目标进行仿真实验, 选择兼容函数

, 其中 α 和 β 为可调节参数, 经过信任传播推理后结果如图 8(a)所示, 容易发现目标相互靠近时, 该集成方法产生互斥作用, 这对多目标跟踪是合理的. 将该集成方法应用于冰球运动员跟踪结果如图 8



(a)



(b)



(c)

图 8 视频多目标跟踪效果

示,其中(b)为传统的多独立跟踪器跟踪效果,(c)是概率图模型线索集成方法跟踪结果.不难看出该方法不但有利于解决多目标遮挡问题,而且有助于集成目标检测线索从而能跟踪新出现目标.

4 讨论

4.1 线索集成理论探索

目前对人类认知机理的研究还处于较低的水平,感知器官或大脑中不同的感知部位在接收外部世界的信息后,对各自的专属信息进行感知,如视觉系统对颜色、纹理以及对比度等敏感,听觉系统对信号频率,而触觉系统对物体的表面和材质敏感,只有综合所有感知信息人类才能恰当地理解外部世界.因此人类对外部世界的认知过程,本质上就是一个多传感信息集成的过程,模式识别过程就是人们依据各种感官所接收的信息把一类事物与其他事物区分开来的过程;对这些不完全可靠、不完备甚至相互矛盾的多通道信息,大脑充分利用它们之间的相关性使之相互监督完成多线索集成,从而获得对外部世界的认识,因此如何正确有效地集成这些不确定线索成为人工智能重要的研究方向.

20世纪50年代以来,人工智能研究领域针对不确定性推理的研究出现了两个主要分支:D-S证据理论与贝叶斯方法^[6].D-S证据理论使用一种称作信任函数的概率赋值来反映信息的不确定性程度,它赋予某假设一个信念值而不必将余下的信念值赋给其余假设,过去大量研究表明它适合于不确定性信息融合,然而D-S证据合成常常会出现组合爆炸问题,因而导致该方法无法适应较为复杂的视觉任务.而将贝叶斯理论引入认知过程来处理不确定性问题已经成为主流.多视觉线索的贝叶斯集成结构允许各个线索知识独立增长,能较好地适应真实环境,具有较强的鲁棒性.新近发展起来的概率图模型能分布式地集成各方面的信息,提供了一种全局最优化(最大后验概率准则)得知识获取途径,并能在信息不完备甚至信息缺失的条件下进行全局推理,这恰好与人类的思维方式相吻合.

4.2 线索贝叶斯集成中有待进一步研究的问题

视觉线索的贝叶斯集成在计算机视觉领域已有不少成功的应用,但在以下几个方面仍然需要进一步的研究工作:

(1) 如何有效地判断多线索间的一致性并检测其中估计失败的线索?过去的研究虽然提出了大量线索集成的方法和工具,却鲜见讨论线索一致性的工作.若线索一致性不能得到保证,大部分集成算法在线索冲突时将失效,可参考目前视觉领域测量一致性初步研究的相关文献^[49,50].

(2) 寻求评价线索间相关性的新方法,或寻找在未知线索相关性条件下进行合理估计的新方法^[51].过去的研究所描述的线索相关性总是以一个既定的先验模型为基础,如概率图模型中的兼容函数,然而单一模型的表示有其自身的局限性,需要寻找一种条件更松弛的线索相关性评价方法.

(3) 研究线索贝叶斯集成的并行处理方法.大部分计算机视觉问题是高维连续空间估计问题,基于离散化或非参数化等解决方法在信息交互阶段占去大部分的计算资源,线索的分布式性质也为并行集成处理提供可能.

(4) 研究快速有效的逼近推理算法.概率图模型的分散式工作机理决定推理过程必然要求大量的计算资源,研究快速逼近推理算法能增大线索集成系统的应用面.

(5) 研究线索集成性能的评估标准.线索集成技术性能评价是任务相关的,但在各种任务下仍然缺乏评估标准以及测试数据.

(6) 研究基于极大熵模型的线索集成方法.当线索规模很大时,极大熵模型可提供线索的对数线性表示,能缓解线索集成模型中学习和推理的困难.

5 结论

视觉线索贝叶斯集成方法是基于贝叶斯统计理论的视觉线索信息融合的方法.其目的是降低估计的不确定性以提高估计的精度、克服线索间的相关性以提高估计的鲁棒性.在实现层面上,有意见线性汇聚、对数汇聚和概率图模型等基本方法;在应用层面上,需要根据具体任务选择合适的视觉线索集合以及集成策略.由于贝叶斯集成框架更符合人类获取与更新知识的方式,因此在计算机视觉尤其是视频分析领域得到了成功应用.大量实验与应用表明视觉线索的贝叶斯集成方法可有效提高视觉系统的精确度和鲁棒性.

参考文献

- 1 Jensen F V. Introduction to bayesian networks. Berlin: Springer-Verlag, 1996
- 2 Jacobs R A, Fine I. Experience-dependent integration of texture and motion cues to depth. *Vision Res*, 1999, 39: 4062—4075 [\[DOI\]](#)
- 3 Jacobs R A. What determines visual cue reliability? *Trends Cogn Sci*, 2002, 6(8): 345—350 [\[DOI\]](#)
- 4 Zhu S C, Wu Y, Mumford D. Filters, random fields and maximum entropy (FRAME), towards a unified theory for texture modeling. *Int J Comput Vis*, 1998, 27(2): 1—20
- 5 Levin A, Weiss Y. Learning to combine bottom-up and top-down segmentation. In: *Proc ECCV*, 2006, Graz, Austria, LNCS3954: 581
- 6 Xue J R, Zheng N N. Robust tracking with and beyond visible spectrum: a four-layer data fusion framework. In: *Proc International Workshop on IWICPAS*, 2006, Xi'an, China, Springer LNCS 4153: 1—16
- 7 Leichter I, Lindenbaum M, Rivlin E. A probabilistic framework for combining tracking algorithms. In: *Proc CVPR*, Washington, DC, 2004. 445—451
- 8 Toyama K, Horvitz E. Bayesian modality fusion, probabilistic integration of multiple vision algorithms for head tracking. In: *Proc ACCV*. Tapei, Taiwan, 2000
- 9 Wang T, Diao Q, Zhang Y, et al. A dynamic bayesian network approach to multi-cue based visual tracking. In: *Proc ICPR*, Cambridge, UK, 2004. 167—170
- 10 Shen C, Anton van den Hengel, Dick A. Probabilistic multiple cue integration for particle filter based tracking. In: *Proc 7th International Conference on DICTA*, Sydney, NSW, Australia, 2003. 399—408
- 11 Spengler M, Schiele B. Towards robust multi-cue integration for visual tracking. *Mach Vis Appl*, 2003, 14: 50—58 [\[DOI\]](#)
- 12 Wu Y, Hua G, Yu T. Switching observation models for contour tracking in clutter. In: *Proc CVPR*, 2003. Madison, Wisconsin
- 13 Mundy J L, Chang C-F. Fusion of intensity, texture, and color in video tracking based on mutual information. In: *Proc 33rd AIPR Workshop*, Washington, DC, USA, 2004. 10—15
- 14 Stone M. The opinion pool. *The Annals of Mathematical Statistics*, 1961, 32(4): 1339—1342
- 15 Clemen R T, Winkler R L. Combining probability distributions from experts in risk analysis. *Risk Anal*, 1999, 19: 187—203
- 16 Benediktsson J A, Swain P H. Consensus theoretic classification methods. *IEEE Trans Syst Man Cybern Part A-Syst Hum*, 1992, 22: 688—704 [\[DOI\]](#)
- 17 Manyika J, Durrant-Whyte H. An information-theoretic approach to management in decentralized data fusion. In: *Proc SPIE* 1992, 1828: 202—213 [\[DOI\]](#)
- 18 Oru I, Maloney L T S M. Weighted linear cue combination with possibly correlated error. *Vision Res*, 2003, 43: 2451—2468 [\[DOI\]](#)
- 19 Li X R, Zhu Y, Wang J, et al. Optimal linear estimation fusion part I: Unified fusion rules. *IEEE Trans Inf Theory*, 2003, 49(9): 2192—2208 [\[DOI\]](#)
- 20 Kittler J, Hatef M, Duin R P W, et al. On combining classifiers. *IEEE Trans Pattern Anal Mach Intell*, 1998, 20(3): 226—239 [\[DOI\]](#)
- 21 Bar-Shalom Y, Li X R. Estimation with applications to tracking and navigation: theory algorithms and software. New York: John Wiley & Sons, 1995
- 22 Zhong X P, Zheng N N, Xue J R. Pseudo measurement based multiple model approach for robust player tracking. In: *Proc ACCV*, Hyderabad, India, 2006. 781—790
- 23 Murphy K P, Weiss Y, Jordan M I. Loopy-belief propagation for approximate inference: an empirical study. In: *Proc 5th Conference on Uncertainty in Artificial Intelligence*, 1999, Sweden, 1999. 467—475
- 24 Sudderth E B, Ihler A T, Freeman W T, et al. Nonparametric belief propagation. In: *Proc CVPR*, Madison, Wisconsin, 2003. 605—612
- 25 Weiss Y, Freeman W T. On the optimality of solutions of the max-product belief-propagation algorithm in arbitrary graphs. *IEEE Trans Inf Theory*, 2001, 47(2): 736—744 [\[DOI\]](#)
- 26 Yedidia J S, Freeman W T, Weiss Y. Constructing free-energy approximations and generalized belief propagation algorithms. *IEEE Trans Inf Theory*, 2005, 51(7): 2282—2312 [\[DOI\]](#)
- 27 Yedidia J S, Freeman W T, Weiss Y. Generalized belief propagation. In: *Proc ANIP*, 2001. 689—695

- 28 Wu Y, Huang T S. Robust visual tracking by integrating multiple cues based on co-inference learning. *Int J Comput Vis*, 2004, 58(1): 55—71 [\[DOI\]](#)
- 29 Pearl J. Fusion, propagation, and structuring in belief networks. *Arti Intell*, 1986, 29(3): 241—288
- 30 Frey B. Graphical models for machine learning and digital communication. Massachusetts: MIT Press, 1998
- 31 Isard M. Pampas: real-valued graphical models for computer vision. In: *Proc CVPR*, Madison, Wisconsin, 2003. 613—620
- 32 Xue J R, Zheng N N, Zhong X P. Sequential stratified sampling belief propagation for multiple targets tracking. *Sci China Ser F-Inf Sci*, 2006, 49(1): 48—62
- 33 Zhong X P, Xue J R, Zheng N N. Graphical model based cue integration strategy for head tracking. In: *Proc 17th British Conference on Machine Vision*, 2006, Edinburgh
- 34 Roth S, Sigal L, Black M J. Gibbs likelihoods for bayesian tracking. In: *Proc CVPR*, Washington, DC, USA, 2004. 886—893
- 35 Ren X, Fowlkes C C, Malik J. Cue integration in figure/ground labeling. In: *Proc NIPS*, 2005, Vancouver
- 36 Pan H. A bayesian fusion approach and its application to integrating audio and visual signals in HCI. Technique Report of UIUC, 2001
- 37 Chakraborty A, Duncan J S. Game-theoretic integration for image segmentation. *IEEE Trans on PAMI*, 1999, 21(1): 12—30
- 38 Patras I, Hendriks E A, Lagendijk R L. Video segmentation by MAP labeling of watershed segments. *IEEE Trans Pattern Anal Mach Intell*, 2001, 23(3): 326—332 [\[DOI\]](#)
- 39 Sigal L, Bhatia S, Black M J, et al. Tracking loose limbed people. In: *Proc CVPR*, Washington, DC, USA, 2004. 421—428
- 40 Beal M J, Attias H, Jojic N. Audio-video sensor fusion with probabilistic graphical models. In: *Proc ECCV*, Copenhagen, Denmark, 2002. 736—752
- 41 Khan S, Shah M. Object based segmentation of video using color motion and spatial information. In: *Proc CVPR*, Kauai, Hawaii, USA, 2001. 746—751
- 42 Wang Y, Loe K-F, Wu J-K. A Dynamic Conditional Random Field Model for Foreground and Shadow Segmentation. *IEEE Trans Pattern Anal Mach Intell*, 2006, 28(2): 279—289 [\[DOI\]](#)
- 43 Sigal L, Isard M, Sigelman B H, et al. Attractive people: assembling loose-limbed models using non-parametric belief propagation. In: *Advances in NIPS 16*: 2003
- 44 Wu S, Hong L. Hand tracking in a natural conversational environment by the interacting multiple model and probabilistic data association(IMM-PDA) algorithm. *Pattern Recognit*, 2005, 38(11): 2143—2158 [\[DOI\]](#)
- 45 Bar-Shalom Y, Li X R. Multitarget-multisensor tracking: principles and techniques. CT: YBS Publishing, 1995
- 46 MacCormick J, Blake A. Probabilistic exclusion and partitioned sampling for multiple object tracking. *Int J Comput Vis*, 2000, 39(1): 57—71 [\[DOI\]](#)
- 47 Tao H, Sawhney H S, Kumar R. A Sampling Algorithm for Tracking Multiple Objects. In: *Proc Workshop on Vision Algorithms 99*, Corfu, Greece, 1999. 53—68
- 48 薛建儒, 郑南宁, 钟小品. 多目标跟踪的序贯分层抽样信任传播算法. *中国科学 E 辑: 信息科学*, 2005, 35(10): 1049—1063
- 49 Bar-Shalom Y, Birmiwal K. Consistency and robustness of PDAF for target tracking in cluttered environments. *Automatica*, 1983, 19(4): 431—437 [\[DOI\]](#)
- 50 Hua G, Wu Y. Measurement integration under inconsistency for robust tracking. In: *Proc CVPR*, 2006, New York
- 51 Julier S J, Uhlmann J K. Non-divergent estimation algorithm in the presence of unknown correlations. In: *Proc ACCV*, New Mexico, 1997. 2369—2373