多源异构感知数据融合方法及其在目标定位跟踪中的应用

胡永利, 朴星霖, 孙艳丰*, 尹宝才

北京工业大学计算机学院多媒体与智能软件北京市重点实验室, 北京 100124

* 通信作者. E-mail: yfsun@bjut.edu.cn

收稿日期: 2013-04-01; 接受日期: 2013-07-26

国家重点基础研究发展计划(973计划)(批准号: 2011CB302703)、国家自然科学基金(批准号: 61133003, 61171169)、北京市自然科学基金(批准号: 4132013, KZ201310005006)资助项目

摘要 物联网的全面感知产生了海量的感知数据, 并且感知数据呈现为显著的多源异构性. 因此, 如何实现海量多源异构感知数据的智能处理是一个具有挑战性的课题. 数据融合是处理多模态数据并挖掘提取有价值信息的有效手段, 但针对多源异构数据, 特别是非结构化的视频多媒体信息, 如何实现高效的融合计算还面临许多问题需要解决. 本文针对物联网多源异构感知信息的处理问题, 提出多层次的多源异构数据融合方法, 并以基于无线信号、视频和深度感知数据的目标定位跟踪应用为切入点, 重点研究多源异构数据的处理、特征表示和数据融合方法. 根据不同类型数据的特性采用不同的数据融合方法, 通过挖掘无线信号、视频和深度等多源异构数据内在的关联性, 实现多源异构数据有价值信息的有效利用. 实际复杂场景的实验表明, 本文提出的基于多源异构数据融合的目标跟踪和定位方法, 能够解决传统依赖单源同质数据的目标跟踪方法所面临的光照变化和遮挡交错等难点问题, 并且可以获得较为准确的运动目标三维位置, 具有良好的跟踪定位效果.

关键词 多源异构 信息融合 跟踪定位 特征表示 物联网

1 引言

作为物联网的基本功能, 信息感知为物联网应用提供了信息来源, 是物联网应用的基础, 也是实现物联网信息“全面感知”应用目标的技术保证. 但物联网原始感知信息表现为显著的多源异构特性, 其中多源性体现在物联网感知数据来源于多个分布式的网络感知节点, 不稳定的网络传输和节点的不同步往往会造成感知信息的缺失和断续, 以及信息的时空关系不一致等问题; 异构性体现在物联网感知数据无论在类型、性质、形式和内容上均存在极大的差异, 例如除了通常的温湿度、光照度、气体浓度等标量数据外, 还有视频和声音等多媒体数据, 同时由于传感器采样和量化方式的不同, 各种不同的感知数据的采样频率和测量精度也存在很大差异. 因此, 针对物联网多源异构的感知信息, 需要研究有效的数据清洗、数据融合、数据压缩等数据处理方法, 将多源异构感知信息整合为物联网应用所需要的确定性信息. 特别是近年来受到极大关注的数据融合处理方法, 通过挖掘数据内在的关联性,

引用格式: 胡永利, 朴星霖, 孙艳丰, 等. 多源异构感知数据融合方法及其在目标定位跟踪中的应用. 中国科学: 信息科学, 2013, 43: 1288-1306, doi: 10.1360/112013-120

从原始数据和数据特征等多个层次对数据进行综合处理, 获得应用需求的有价值信息, 在多源异构数据处理方面得到广泛的应用。

在物联网应用环境中, 数据融合是根据多源异构感知数据的时空相关性去除错误和冗余数据, 获取确定数据, 并在特征层和语义层提取有效信息, 以满足为物联网应用的需求。文献 [1] 对无线传感器网络中的数据融合技术做了系统综述。按照数据处理的层次, 数据融合可分为数据层融合、特征层融合和决策层融合。数据层融合通常采用概率统计方法、回归分析和卡尔曼滤波等方法, 消除冗余信息, 去除噪声和异常值。文献 [2] 提出的分布式融合方法, 采用极大似然估计实现了局部感知数据的稳定估计, 消除了数据异常, 解决了不同步数据的融合问题。文献 [3] 在研究基于簇结构的数据融合时, 为了解决数据收集中簇头节点的数据冲突问题, 采用 Bayes 方法估计发送数据节点的数量。为了提高 Bayes 数据融合的计算效率, 文献 [4] 实现了后验概率的分布式计算。文献 [5,6] 建立了感知数据的回归模型, 通过模型的回归计算大幅减少了数据量。文献 [7] 将漂移均值滤波用于观测值的数据处理, 以及事件和事件边界的估计。卡尔曼滤波在观测值预测 [8]、上下文信息预测 [9], 甚至 MAC 层数据帧大小预测 [10] 等方面均有应用。但卡尔曼滤波不能很好地处理非高斯噪声和低采样率的数据, 因此粒子滤波方法被引入无线传感器网络的数据分析和处理中, 特别是节点定位和跟踪方面。文献 [11] 采用粒子滤波方法实现了基于网络几何属性的目标跟踪。在特征层或决策层采用 D-S 证据理论、模糊逻辑、神经网络及语义融合等方法, 对来自多源异质网络节点的多类型不确定性数据进行融合, 可以实现事件检测、状态评估、语义分析等高层的决策和判别。文献 [12] 采用 D-S 证据理论实现了网络路由状态的分析, 给出了路由是否需要重建的判别。模糊逻辑能很好地处理推理和决策中的不确定性因素, 因此在节点定位跟踪 [13]、簇头选择 [14]、路由构建 [15], 以及 MAC 存取控制 [16] 等方面得到了应用。神经网络能够将不确定的数据通过学习转化为系统理解的形式, 且适合大规模并行处理, 因此可应用于物联网多节点感知数据的融合。文献 [17] 将神经网络用于目标识别系统的多传感器数据的融合。语义融合技术是基于数据语义描述的高层数据融合方法。该方法一般从感知数据提取抽象语义, 通过语义的模式匹配实现状态监测或分类识别。文献 [18] 最早提出了语义融合方法, 并将其应用于传感机器人的状态识别。

数据融合能消除错误和冗余数据, 有效减少数据量, 获取确定数据和有价值的语义信息。因此, 数据融合已成为物联网信息感知的关键技术和研究热点。但现有工作大部分是面向网络数据传输量和能耗的减少, 研究数据的高效传输和聚集的方法, 而且往往只针对单一类型的标量数据, 例如温湿度和光照等数据的融合, 而对于音视频多媒体数据等异构数据的融合方法研究还不够深入。而在视频监控、目标定位和跟踪、行为分析等研究领域, 基于多源异构数据融合的感知定位和跟踪方法得到了初步的研究, 并取得了优于单一感知数据的效果。例如, 文献 [19] 使用二维与三维相结合的方式对路上行人进行检测。首先使用摄像机标定的方法获取道路的三维场景信息, 并得到场景中的感兴趣区域将其投影在二维图像上。然后使用图像检测器在目标区域进行行人检测。最后根据正结果的尺度与场景尺度的比值对检测结果进行细化, 从而得到更加准确的检测结果。文献 [20] 完全依靠激光扫描仪来获取行人可能出现的区域, 对每个可能区域使用传统的神经网络方法进行检测。文献 [21] 提出了一种实时行人分类系统, 首先使用激光扫描仪来确定车辆间的区域, 然后使用图像分类器对该区域进行检测, 该系统的目的是用于检测街道上突然出现的行人。文献 [22] 提出在激光成像空间使用高斯混合模型进行行人检测。文献 [23] 提出了一种融合不同感知器的方法。在激光空间使用 AdaBoost 方法获取特征并训练分类器, 然后将分类结果送入 Bayes 概率网络进行分析, 实现行人和车辆的分类。在图像空间, 使用模版匹配方法来匹配场景中的行人和车辆。最后根据跟踪时两者对应的目标将这两个感知器的结果融合在一起。文献 [24] 提出了一种基于语义的融合方法。根据 Markov 逻辑网络将二者结合在

图 1 基于多源异构感知数据融合的定位和跟踪研究方案

Figure 1 Objects positioning and tracking framework based on the fusion of multi-source heterogeneous data

一起, 首先使用原有的机制对人体的各个部分分别训练各自的检测器, 然后根据人体的组合条件对部分检测器的结果进行分析和组合, 最后得出检测结果. 该方法可以在简单场景下有效处理遮挡的问题.

针对物联网多源异构感知数据的融合问题, 本文以物联网典型的目标定位和跟踪应用为切入点, 研究无线、视频和深度等感知信息的有效融合方法, 特别是研究无线信号强度及其定位结果的标量数据与视频和深度序列等非结构化数据的融合方法. 实际上, 目标定位与跟踪一直是物联网应用领域的研究热点, 文献 [25] 全面综述了无线网络的定位和跟踪研究进展. 虽然目前基于场景无线指纹库的无线定位方法可以获得较为理想的定位效果, 并且效率高可实时应用, 但是无线定位易受无线信号强度不稳定等因素的影响, 其捕捉到的信息只包含目标的位置和轨迹信息, 无法提供视频等视觉细节信息, 例如目标的纹理、姿态和动作等信息. 虽然激光测距传感器可以获得精度很高的三维信息, 但其扫描过程很费时, 无法用来获取动态场景的三维信息, 难以适应快速分析与理解的需求, 而且当场景中出现遮挡时, 激光测距传感器的检测效果也会大打折扣. 基于视频的目标检测和跟踪一直是计算机视觉领域的重要研究课题之一, 研究者先后提出了许多性能良好的检测跟踪算法. 例如, 常见的基于时间序列分析的卡尔曼滤波 [26] 和粒子滤波 [27] 等方法. 最近, 稀疏表示和压缩感知等新理论也被引入到视频跟踪中用以解决人体的特征表示问题 [28,29]. 但现有的视频人体跟踪和检测方法还存在许多难点问题需要解决, 例如大幅度的人体姿态变化、衣着变化、背景变化、环境光变化、目标遮挡等都会对跟踪结果产生极大的影响. 因此, 不同的目标感知定位方法都有各自的特点, 面向复杂的物联网应用场景, 单独使用任何一种感知手段往往难以获得令人满意的结果.

针对复杂场景下实时精确的多目标定位与跟踪问题, 本文利用更多、更有效的数据源, 包括无线、视频和深度信息, 挖掘传感器之间的互补性, 提出多源异构信息的有效融合方法, 以突破现有技术的瓶颈, 实现鲁棒的目标定位与跟踪. 本文提出的多源异构数据融合的研究方案如图 1 所示, 对于获取的无线、视频和深度感知数据, 首先进行单源数据的处理, 包括基于无线信号强度数据的目标定位, 基于视频和深度序列的目标定位跟踪等; 然后面向数据融合进行多源异构数据的时空对齐; 在此基础上, 研究面向目标跟踪定位的多源异构数据融合方法及数据融合表达模型, 并通过实际场景的实验, 验证上述方法的有效性. 本文的结构如下: 第 2 节给出多源异构数据的处理方法; 第 3 节解决多源异构数据的时空对齐问题; 第 4 节重点叙述提出的数据融合方法; 第 5 节给出实验结果; 最后总结全文.

2 无线、视频和深度数据处理

在进行数据融合之前, 需要对多源异构数据分别进行相应的处理, 获得便于融合的数据表示形式. 本文中的无线数据利用若干个 Wi-Fi 无线 AP 和移动终端实现采集, 视频采用多路高清摄像头获取, 深度信息采用可同时获取视频和深度 (称之为 RGB-D) 序列数据的 MS Kinect 感知器采集. 针对上述无线信号强度观测数据、视频和深度序列分别提出相应的数据处理方法, 包括基于低秩矩阵重构的无线指纹库构建、基于压缩采样的视频跟踪和基于深度特征的跟踪方法.

2.1 基于低秩矩阵重构的无线指纹库构建和无线定位方法

指纹法^[30]是目前无线定位的主流方法, 具有较理想的定位精度. 该方法不是直接将信号强度的测量值映射为信号传播距离, 而是利用预先观测到的场景信号强度作为场景信号特征来推断观察者的位置, 其实质是对信号强度与位置间的内在规律进行学习, 并通过新测量值与学习过的样本点的匹配实现定位. 预先建立指纹库是指纹法实现定位的基础, 通常为了获取全面准确的指纹数据, 往往需要大量的信号强度测量工作. 为了减少指纹获取的工作量, 并实现实用的定位方法, 本文利用无线信号强度分布的连续性和局部平滑性, 提出一种平滑约束的低秩重构模型, 基于少量信号强度采样点, 通过优化方法重构指纹库, 并实现目标的无线定位.

设无线定位的目标区域需要采样的某个无线信号源的观测为矩阵 X , 其元素个数为 N , 即需要在 N 个位置进行观测. 为了减少观测, 我们只从目标区域随机选择 $n(n < N)$ 个采样点进行观测, 记这个部分观测矩阵为 B , B 与 X 的关系可以用一个掩膜算子来表示, 即 $A(X) = B$, 其中 A 为模板矩阵, 其元素 $A(i, j)$ 为 1 表示在该位置有观测, 为 0 表示该位置没有观测, 即 $B(i, j) = A(i, j)X(i, j)$. 如果将 X 看作是一个具有低秩特性的矩阵, 则基于少量观测的指纹库构建问题可通过下面的低秩矩阵填充模型来求解:

$$\min \text{rank}(X), \quad \text{s.t.} \quad A(X) = B. \quad (1)$$

无线信号通常除了一些噪音和奇异点, 在空间中的变化是连续的, 然而在传统低秩矩阵填充模型^[31]中, 虽然重构结果在已知采样点的重构误差很小, 但整个重构矩阵并不连续. 因此, 为了获得符合实际情况的信号强度重构矩阵, 本文在上述基本低秩矩阵填充模型中引入平滑性约束, 即利用带平滑化的低秩矩阵填充模型对原始指纹库进行重构, 带平滑化的低秩矩阵填充模型为

$$\min \text{rank}(X) + \lambda S(X), \quad \text{s.t.} \quad A(X) = B, \quad (2)$$

其中 $S(X)$ 是一个表示 X 连续性的平滑因子, λ 是由实验获得的平衡系数. 具体地, $S(X)$ 通过矩阵水平和垂直方向的差分来定义, 即 $S(X) = \|D_x(X)\|_F^2 + \|D_y(X)\|_F^2$, 其中 $D_x(X)$ 是一个大小为 $N_1 \times (N_2 - 1)$ 的矩阵, 代表矩阵 X 中各元素水平方向的差分 $D_x(X) = X(i, j+1) - X(i, j)$, $D_y(X)$ 是一个大小为 $(N_1 - 1) \times N_2$ 的矩阵, 代表矩阵中各元素垂直方向的差分 $D_y(X) = X(i+1, j) - X(i, j)$, 运算符 $\|\cdot\|_F^2$ 代表矩阵的 Frobenius 范数. 为了求解模型, 本文采用文献^[32]的方法将低秩约束转化为矩阵的 F 范式, 并将 X 进行 SVD 分解, 通过交替迭代实现优化求解. 具体地, X 的 SVD 分解表示为 $X = U\Sigma V^T$, 其中 U 是一个大小为 $N_1 \times N_1$ 的正交矩阵, V 是一个大小为 $N_2 \times N_2$ 的正交矩阵, Σ 是一个大小为 $N_1 \times N_2$ 的对角矩阵, 包含降序排列的奇异值 σ_k . 进一步 $X = U\Sigma V^T = LR^T$, 其中 $L = U\Sigma^{1/2}$, $R = V\Sigma^{1/2}$. 设 L 为一个大小为 $N_1 \times K$ 的矩阵, R 为一个大小为 $N_2 \times K$ 的矩阵, 这里 K 是对矩阵 X 秩的估计值. 考虑到移动终端接收到的信号强度值通常不够精确, 因此场景中的矩阵

图 2 基于低秩矩阵填充的指纹库重构

Figure 2 RSS fingerprint reconstruction based on low rank approximation. (a) The simulated RSS distribution; (b) the reconstructed RSS distribution from 20% measurements

不完全符合低秩特性, 这里放宽 $A(LR^T) = B$ 的约束条件, 从而将上述约束优化模型转化为下面的无约束优化模型:

$$\min \|L\|_F^2 + \|R\|_F^2 + \eta \|A(X) - B\|_F^2 + \lambda (\|D_x(X)\|_F^2 + \|D_y(X)\|_F^2), \quad (3)$$

其中 $\|A(X) - B\|_F^2$ 表示在采样子集 B 上的重构误差, η 是权系数. 通过 L 和 R 的交替迭代可实现上述模型的求解, 即给定 L 和 R 的初始值, 固定其中一个并迭代更新另一个, 然后交换, 最终得到 L 和 R , 并通过 $X = LR^T$ 实现指纹库的重构. 图 2 是在仿真实验中对 50×100 大小的原始信号强度使用 20% 随机采样点的重构结果.

获得了场景的无线信号指纹信息, 本文采用最近邻方法实现观测样本的定位. 设未知位置 $\hat{P}$ 获得的信号强度为 $\text{RSS}(\hat{P}) = (\text{RSS}_{AP1}, \dots, \text{RSS}_{APm})^T$, 其中 m 为 AP 的数量, 则 $\hat{P}$ 可以通过指纹库中的近邻信号强度所在位置的加权来获得

$$\hat{P} = \sum_{i=1}^{N_b} w_i P_i, \quad (4)$$

其中 N_b 是近邻指纹采样点的数量, $w_i = \exp(-\frac{\|\text{RSS}(\hat{P}) - \text{RSS}(P_i)\|^2}{2\sigma}) / \sum_{i=1}^{N_b} \exp(-\frac{\|\text{RSS}(\hat{P}) - \text{RSS}(P_i)\|^2}{2\sigma})$ 是基于高斯核函数的信号强度相似度确定的权系数, 其中 σ 是信号强度误差的一个估计值.

实验表明, 本文提出的基于低秩矩阵填充的无线指纹库重构方法可以有效降低无线指纹数据采样的工作量, 并获得较高精度的定位结果.

2.2 深度数据处理和三维定位

现有的深度采集设备获取的原始深度图像通常质量较差, 尤其在光滑物体表面和存在遮挡的边缘区域往往存在噪声或深度数据缺失, 因此需要对深度数据进行相应处理, 修复缺失值, 以便于利用深度数据进行定位和目标跟踪. 目前, 针对上述问题已经提出了一些深度数据修复算法, 并获得了较为理想的修复效果, 例如基于 QuickShift 的修复方法^[33] 和利用深度对应视频的边缘保持滤波方法^[34] 等. 考虑到人体检测跟踪对深度信息处理的实时性要求, 本文提出了一种基于视频图像邻域信息匹配的深度图像快速恢复方法, 该方法的基本思想是利用视频图像局部像素的一致性查找深度图像缺失像素的有效近邻像素, 利用其有效像素值填充缺失值完成修复. 具体地, 首先通过中值滤波对深度图像进行处理, 去除局部孤立的缺失值和噪声点, 然后针对较大范围的缺失值设定搜索邻域的半径, 在该邻域内随机选择一个像素点, 如果该像素点的深度值有效, 并且其对应的视频图像灰度值与缺失像素对应

图 3 原始深度图像和修复后的深度图像

Figure 3 The original depth image and its restored result

的视频图像灰度值的误差在给定的阈值范围内, 则使用该像素深度值修复缺失像素的深度值, 否则选择邻域内下一个随机像素点进行探测. 如果在给定的探测次数内没有实现修复, 则增加搜索领域的半径, 在更大的范围继续尝试修复, 直到得到满足要求的像素. 在实际采集的 RGB-D 序列数据上的实验表明, 本文提出的方法通常只需要探测几次即可完成缺失值的修复, 算法具有很高的效率. 图 3 是深度图像修复前后的对比效果.

获得了修复后的深度图像, 为了实现目标的定位, 还需要将深度图像的像素值转换为三维空间的坐标, 即实现基于深度数据的定位. 根据 Kinect 的工作原理^[35], 给定深度图像中 (i, j) 位置的深度值 d_{ij} , 则其相对于 Kinect 坐标系的三维位置的齐次坐标 $X_{ij} = (x, y, z, 1)^T$ 由下式来计算:

$$X_{ij} = A(W_{ij}^T, Z_{ij}, 1)^T, \quad (5)$$

其中 $W_{ij} = f(Z_{ij}(i, j)^T + c)$ 是一个二维向量, $Z_{ij} = \frac{1}{d_0 d_{ij}}$, A 是一个 4×4 的变换矩阵. 这里 d_0, f, c 和 A 是通过实验获得的参数, 由 Kinect 自身来确定.

2.3 基于压缩采样的目标视频跟踪

对于视频数据, 本文采用实时性非常好的基于压缩感知理论的压缩跟踪算法^[36]实现目标跟踪. 压缩跟踪算法的基本思想是构造目标区域的稀疏观测, 并将目标区域投影到一个稀疏域获得特征, 通过模板和目标的特征相似度实现目标跟踪. 根据压缩感知理论^[37], 目标区域的特征表示为

$$V = RB, \quad (6)$$

其中 $B \in R^m$ 为基于图像块定义的目标区域的 Haar-like 特征^[38,39]集合, $R \in R^{n \times m}$ 是观测投影矩阵.

要获得理想的跟踪结果, 观测投影矩阵的构造非常关键. 一般地, 可采用随机高斯矩阵产生相应的观测矩阵, 但随机高斯矩阵过于密集, 当 m 太大时, 内存和计算量过大. 因此, 本文利用随机选择的少量特征实现观测采样, 具体如图 4 所示, 在每一次观测时 (即观测矩阵的一行) 只随机选择 s 个采样点, 为了实现特征的快速计算设 $2 \leq s \leq 4$, 并且给相应的特征随机给定权值, 通过随机加权获得压缩的观测特征, 即第 i 次观测的特征 v_i 定义为

$$v_i = \sum_{j=1}^n r_{ij} B_j, \quad (7)$$

图 4 压缩特征的生成

Figure 4 The generation of compressive feature

其中 $r(i, j)$ 是取值为 $\{-1, 0, 1\}$ 的随机数, 且 $P(r_{ij} = -1) = P(r_{ij} = 1) = \frac{1}{2s}$, $P(r_{ij} = 0) = 1 - \frac{1}{s}$, B_j 是图像块的 Haar-like 特征, 这里采用图像块内的像素之和进行快速计算. 通过上述计算得到的压缩特征 v_i 可以看作一种广义的 Haar-like 特征. 在实际中, 每次观测的随机像素数 s 在 $\{2, 3, 4\}$ 中等概率选取, 这样使得每次观测的特征计算非常高效.

基于上述压缩特征的目标跟踪, 可以看作在图像序列中寻找与目标特征最相似的候选区域. 候选区域的选取方式是遍历上一次目标区域的邻近区域, 通过目标区域是否包含前景和背景信息的 Bayes 准则实现相似性判别:

$$H(V) = \log \left(\frac{\prod_{i=1}^n p(v_i|y=1)p(y=1)}{\prod_{i=1}^n p(v_i|y=0)p(y=0)} \right) = \sum_{i=1}^n \log \left(\frac{p(v_i|y=1)}{p(v_i|y=0)} \right), \quad (8)$$

其中 $p(v_i|y=1) \sim N(\mu_i^1, \sigma_i^1)$, $p(v_i|y=0) \sim N(\mu_i^0, \sigma_i^0)$. μ^1, σ^1 表示目标样本的均值和标准差, μ^0, σ^0 表示背景样本的均值和标准差. 通常目标样本在靠近目标区域的位置选取, 背景样本在远离目标区域的位置选取. 使得 $H(V)$ 最大的候选区域为最终的目标位置. 同时根据选取的当前目标位置, 对相应的前后景参数进行更新, 以便适应目标因为环境光照和姿态等变化引起的目标模板的变化, 具体通过下式实现参数更新:

$$\mu_i^1 \leftarrow \lambda \mu_i^1 + (1 - \lambda) \mu^1, \quad (9)$$

$$\sigma_i^1 \leftarrow \sqrt{\lambda(\sigma_i^1)^2 + (1 - \lambda)(\sigma^1)^2 + \lambda(1 - \lambda)(\mu_i^1 - \mu^1)^2}, \quad (10)$$

其中参数 λ 越小表示系数更新的速度越快, 保留之前目标的特征越少.

在实际环境拍摄的包含运动人体的视频上的跟踪实验表明, 本文采用的跟踪方法具有较好的鲁棒性, 并且能满足实时应用的需求.

3 数据时空对齐

由于不同感知节点采集的数据在时间空间上存在不同步、不关联和形式不一致等问题, 因此在数据融合之前需要实现多源异构感知数据的目标关联与时序对齐. 这里通过多源异构数据的关键帧协同检测和关键帧对齐等方法, 实现多源异构感知数据的时空一致表达.

对于视频、深度、无线数据, 本文提出了利用时空上下文信息进行多源异构数据的协同关键帧分析方法, 以找出具有区分能力的关键帧数据, 便于下一步的数据对齐和数据融合. 设输入的原始视频序列、深度序列和无线观测数据分别为 $S_v = \{f_i^v | i = 0, \dots, N_v\}$, $S_d = \{f_i^d | i = 0, \dots, N_d\}$, $S_w =$

图 5 基于联合 HMM 的关键帧对齐

Figure 5 Key frames alignment based on the jointed HMM

$\{f_i^w | i = 0, \dots, N_w\}$. 在特定时刻 t , 观察窗口 $[t-k, t-1]$ 的上下文信息为 $\sigma = [f_{t-k}, f_{t-1}]$, 则当前时刻异构数据的可区分性可定义为当前时刻状态 f_t 的观测概率 $P(f_t|\sigma)$ 的函数 $D(t) = g(P(f_t|\sigma))$. 函数 $g(\cdot)$ 取为递减函数, 即根据上下文观测数据以及其它源数据可预测的概率越大, 其数据的可区分性越小. 对于深度和无线信号, 其上下文采用空间位置信息表示. 对于视频数据, 采用基于视觉特征相似度的方法, 此时后验概率可以使用视觉相似度的倒数来近似表示. 有了各个数据序列的后验概率, 则可以定义下面的多源数据协同可区分性:

$$D(t) = g(P(f_t^v|\sigma) \cdot P(f_t^d|\sigma) \cdot P(f_t^w|\sigma)), \quad (11)$$

基于上述定义, 给定可作为关键帧数据的阈值 D_0 , 则满足 $D(t) > D_0$ 的数据帧即可确定为多源异构时序数据的关键帧. 通过上面的协同关键帧检测最终形成多源异构数据的关键帧序列 $K_v = \{k_i^v | i = 0, \dots, N\}$, $K_d = \{k_i^d | i = 0, \dots, N\}$, $K_w = \{k_i^w | i = 0, \dots, N\}$.

在协同检测得到关键帧的基础上, 需要解决多源数据的关键帧目标对齐问题. 对于输入的关键帧序列 K_v, K_d, K_w , 本文采用联合 HMM 模型进行目标关联. 对于多源异构数据序列, 构建相似度矩阵 S 及状态转移矩阵 M , 例如对图 5 所示的转移过程可建立相应的优化目标函数:

$$V_{tm} = P(k_t^i|m) \cdot \max_{y \in Y} \{\alpha_y^{k_j} V_{(t-1)y}\}, \quad (12)$$

并设定初始状态分布为 $V_{0m} = P(k_0^i|m) \cdot \pi_m$. 为了进行快速的对齐, 这里采用 Viterbi 算法进行计算. 同时, 由于异构数据的帧率不同, 对某个局部状态 t , 计算 t 的一个邻域 $(t-\Delta t, t+\Delta t)$ 的统计量, 以消除帧率不同带来的影响. 为了完成有效的动态匹配, 可通过求解下列目标函数:

$$\text{sim}_t(i, j) = \max_{t' \in (t-\Delta t, t+\Delta t)} \{\text{sim}(k_t^j, k_{t'}^i)\}, \quad (13)$$

其中 $\text{sim}(\cdot)$ 是不同数据关键帧的相似度度量, 这里根据 $\text{sim}_t(i, j)$ 构建 Viterbi 过程的相似度和转移概率矩阵, 实现多源异构数据精确的对齐.

4 多源异构数据的多层次融合方法

由于无线、视频和深度的数据性质和形式存在很大差异, 而且就移动目标定位跟踪的应用而言, 基于不同数据获得的目标位置的有效范围和感知精度也有很大差异, 例如无线定位可以在较大范围实现目标的定位, 但受环境因素的影响无线信号强度的波动很大, 精度不理想; 深度信息的定位精度相对较高且不受光照等外部因素的影响, 但目前的深度采集设备的空间采样分辨率不高, 且感知范围非常

有限, 通常只适合局部范围的目标定位; 而视频跟踪只能获得图像空间的位置信息, 且其跟踪效果易受环境光照以及目标遮挡等因素的影响. 因此, 面向目标跟踪定位的应用, 如何有效融合多源异构数据, 发挥每种数据的优势, 利用异构信息消除基于单一数据跟踪定位错误和异常, 是一项具有挑战性的工作. 本文深入分析和研究每种数据在定位跟踪方面的特点, 提出了不同数据融合的策略和方法, 通过多源异构数据在数据、特征和决策等多个层次的融合, 实现了鲁棒的目标跟踪定位. 下面以无线定位与视频跟踪的融合, 视频与深度的融合, 以及无线、视频和深度多源异构数据的融合为例, 给出相应的融合方法.

4.1 无线定位与视频跟踪的融合

在光照条件较好, 目标不存在大部分遮挡或目标纹理区分性较大的情况下, 目前主流的基于视觉特征的视频跟踪方法具有理想的跟踪效果, 然而在视觉特征受到光照变化或目标遮挡的影响时, 视频跟踪经常会发生跟丢或误跟等情况, 而且此时很难进行判断, 因此会导致目标跟踪的失败. 与视频跟踪相比, 虽然无线定位的误差较大, 精度有限, 但无线定位通常具有一定连续性, 通常情况下不会出现跟丢和误跟的现象. 因此, 本文提出无线定位和视频跟踪融合的跟踪方法, 通过无线的连续定位发现视频跟踪中目标丢失和误跟的事件, 并及时给出视频重新跟踪的起始点, 实现连续跟踪.

无线定位和视频跟踪融合的关键是要对视频跟踪位置和无线定位位置的一致性进行建模评价. 设时刻 t 视频图像平面的目标位置为 $X_{vt} = (x_{vt}, y_{vt})^T \in R^2$, 而无线定位的目标位置为 $X_{wt} = (x_{wt}, y_{wt}, z_{wt})^T \in R^3$, 这两个位置都可以看作实际位置的一个随机变量, 即 $P(X_{vt}) \sim N(\mu_{vt}, \Sigma_{vt})$, $P(X_{wt}) \sim N(\mu_{wt}, \Sigma_{wt})$. 通常在视频跟踪正常的情况下, 视频的跟踪定位比较准确, 而无线定位的误差较大, 因此可以近似将无线定位位置在图像平面的投影点看作以视频跟踪点为中心的随机变量, 从而可以通过其分布来定义视频和无线定位的一致性度量:

$$\text{sim}(X_{vt}, X_{wt}) = \exp \left[-\frac{1}{2} (X'_{wt} - X_{vt})^T (\Sigma'_{wt})^{-1} (X'_{wt} - X_{vt}) \right], \quad (14)$$

其中 X'_{wt} 是 X_{wt} 在图像视频图像上的投影点, Σ'_{wt} 是投影点的样本协方差阵.

基于上述一致性度量, 给定相应的阈值, 在无线定位和视频跟踪过程中, 即可持续检查目标跟踪定位的一致性, 如果一致性小于给定的阈值, 则认为视频跟踪发生了目标丢失或误跟, 这时以无线定位位置在视频图像上的投影点为中心, 在其邻域内重新检测目标对象, 并设置视频跟踪的起始点, 从而实现误跟的纠正和持续跟踪.

4.2 基于粒子滤波的融合深度和视频的目标跟踪

由于深度感知器可以同时获得包含视频和深度信息的 RGB-D 序列数据, 因此本文提出了一种基于粒子滤波的融合视频和深度信息的目标跟踪方法, 以提升 RGB-D 序列定位跟踪的准确性. 粒子滤波 (particle filter) 是一种广泛使用的具有良好性能的视频跟踪方法^[40~42], 其核心思想是通过粒子集合匹配可能的目标位置, 并采用顺序重要性采样更新粒子, 实现目标的连续动态跟踪. 对于 RGB-D 序列数据, 我们将目标的位置作为粒子滤波的状态, 并提出一种融合目标区域图像和深度信息的目标表示特征, 将之作为粒子滤波的观测, 从而实现基于粒子滤波的跟踪方法. 这里首先定义粒子滤波的状态变化以实现动态跟踪, 使用一个二阶模型来表示状态变量, 其动态模型如下:

$$S_t = \lambda_1(S_{t-1} - S_0) + \lambda_2(S_{t-2} - S_0) + G_t, \quad (15)$$

图 6 目标区域深度的二值化表示

Figure 6 The binarization of the target block in depth image

其中 S_0 表示初始状态, λ_1 和 λ_2 是由实验得到的权值, G_t 用于调整候选目标区域的位置和大小, 是一个零均值的高斯随机过程向量.

对于观察特征, 我们采用融合纹理图像和深度信息的目标特征表示. 具体地, 对于图像纹理信息, 首先将 RGB 颜色转换到 HSV 颜色空间, 以减小光照变化的影响, 然后采用 HSV 颜色直方图表示目标区域的图像特征^[43], 即目标区域的图像特征可表示为一个向量 $H = [h_1, \dots, h_i, \dots, h_n]^T$, n 表示直方图 bin 的数目, h_i 是颜色落在第 i 个 bin 区域中的频度. 相对于视频图像, 深度图像的局部纹理变化不大, 并考虑应用的实时性要求, 这里将目标区域中的深度值以前一帧目标区域的深度值的平均作为阈值进行二值化, 1 表示目标人体区域, 0 表示背景或其它物体, 获得目标区域的深度特征表示, 如图 6 所示. 为了进一步提高效率, 将该二值图像下采样, 将其表示为向量的形式得到目标最终的深度特征 $D = [d_1, \dots, d_i, \dots, d_m]^T$. m 是下采样后目标区域的像素数.

有了目标区域的特征表示方法, 在当前帧根据特征的相似性匹配候选粒子和模板粒子即可获得最优的粒子区域, 从而实现跟踪. 因此, 需要定义特征的相似性度量. 这里在分别定义图像特征和深度特征的相似性度量的基础上, 定义了两种特征的融合相似性. 对于图像特征, 候选粒子和模板粒子的图像特征 H_O, H_T 的相似性通过 Bhattacharyya distance 距离^[44]来定义:

$$M_v(H_O, H_T) = e^{-\lambda(1 - B_d((H_O, H_T)))}, \quad (16)$$

其中 $B_d(H_O, H_T) = \sum_{i=1}^n \sqrt{h_{O_i} \cdot h_{T_i}}$ 是 H_O 和 H_T 的 Bhattacharyya distance 距离, λ 是用于调整变化率的参数, 它使得 $M_v(\cdot) \in [0, 1]$. 对于深度特征相似性, 同时考虑目标区域的前景 (1 值) 和背景 (0 值) 的相似性, 则候选粒子和模板粒子的深度特征 D_O, D_T 的相似性定义如下:

$$M_d(D_O, D_T) = \frac{\|(D_O \wedge D_T)\|_{l_0}}{\|D_T\|_{l_0}} \cdot \frac{\|(\sim D_O) \wedge (\sim D_T)\|_{l_0}}{\|\sim D_T\|_{l_0}}, \quad (17)$$

其中 $\|\cdot\|_{l_0}$ 表示 l_0 范数, 是非 0 值的个数. $\wedge$ 是 AND 按位操作, $\sim$ 是 NOT 按位操作. 式中第一个乘数表示前景的相似性, 第二个乘数表示背景的相似性. 基于上述图像特征和深度特征的相似性, 我们定义图像和深度特征的融合相似性如下:

$$M(HD_O, HD_T) = M_v(H_O, H_T) \cdot M_d(D_O, D_T), \quad (18)$$

其中 HD_O 和 HD_T 表示候选粒子和模板粒子的融合特征. 定义了上述粒子区域匹配的相似性度量, 则可以通过粒子选择和更新实现目标区域的匹配, 从而实现 RGB-D 序列中目标的连续跟踪.

图 7 基于多源异构数据定位跟踪的决策融合示意图

Figure 7 The illusion of decision fusion of the multi-Source heterogeneous data for positioning and tracking

为了实现更加鲁棒的跟踪和定位, 本文将上述粒子滤波的跟踪方法应用于多个深度摄像机的协同跟踪, 提出如下联合匹配函数来实现多 RGB-D 序列中的粒子匹配:

$$C(\text{HD}_{\text{O1}}, \text{HD}_{\text{O2}}) = \frac{M(\text{HD}_{\text{O1}}, \text{HD}_{\text{T1}}) + M(\text{HD}_{\text{O2}}, \text{HD}_{\text{T2}})}{\omega \|P_1 - R^* \cdot P_2 + T^*\|}, \quad (19)$$

其中 HD_{O1} , HD_{T1} 和 HD_{O2} , HD_{T2} 分别表示摄像机 1 和 2 候选粒子和模板粒子的融合特征. P_1 和 P_2 分别是摄像机 1 和 2 观测粒子 HD_{O1} 和 HD_{O2} 中心的三维坐标. R^* , T^* 是从摄像机 2 到摄像机 1 的变换矩阵, ω 是用于平衡相似性和空间变换的权值. 依据上面的联合匹配准则, 则可以由下面公式得到最优的候选目标位置:

$$(\text{HD}_{\text{O1}}^*, \text{HD}_{\text{O2}}^*) = \arg \max_{\text{HD}_{\text{O1}} \in P_{t1}, \text{HD}_{\text{O2}} \in P_{t2}} C(\text{HD}_{\text{O1}}, \text{HD}_{\text{O2}}) \quad (20)$$

其中 P_{t1} 和 P_{t2} 分别表示摄像机 1 和 2 的粒子集合.

4.3 基于无线定位、多视视频和深度定位的决策级融合

基于无线定位、多视视频和深度定位的决策级融合实际上是在三维空间的一个候选邻域内决策和选择一个最优点, 使得该点投影到各个数据源或部分融合定位空间的投影点与该空间的定位点的总体误差最小. 这里的总体误差可以通过各个源目标特征和模板特征的相似性来计算. 多源异构信息的动态跟踪定位融合过程如图 7 所示, 假设在 $t-1$ 时刻已经获得跟踪目标的三维空间的位置信息 $(X_{t-1}, Y_{t-1}, Z_{t-1})$, 则需要根据目标在视频平面空间、深度空间以及无线信号定位空间等多源数据空间的相似度, 确定跟踪目标在 t 时刻的空间位置信息.

由于无线定位的精度相对较差, 这里首先采用无线定位方法对目标进行粗定位, 得到在 t 时刻的大致位置 (X_t^w, Y_t^w, Z_t^w) , 然后根据 (X_t^w, Y_t^w, Z_t^w) 相对于 $(X_{t-1}, Y_{t-1}, Z_{t-1})$ 的偏移方向 $\vec{r}$, 在 $(X_{t-1}, Y_{t-1}, Z_{t-1})$ 的邻域内确定候选目标, 将其分别映射到不同数据源的空间中并进行相应的相似度计算. 具体的, 通过摄像机模型可以将候选点投影到二维图像空间, 并利用模板和候选区域的压缩采样特征的比较得到视频图像的候选区域的相似度; 深度匹配也采用类似的方法, 利用融合视频和深度的特征相似度实现模板和候选区域的相似度计算; 无线定位信息则可以通过以候选点为中心的领域的一致性实现相似度的计算. 通过上述多个空间的相似度计算实现定位跟踪的融合, 可以看作一个多目标的优化问

图 8 多源异构时空数据流表示

Figure 8 Multi-source heterogeneous spatial temporal data flow representation

题, 具体形式如下:

$$(X_t, Y_t, Z_t) = \arg \max_{(X_t^c, Y_t^c, Z_t^c) \in N_b((X_{t-1}, Y_{t-1}, Z_{t-1}))} \sum_i \omega_i H_i((X_t^c, Y_t^c, Z_t^c)), \quad (21)$$

其中 (X_t^c, Y_t^c, Z_t^c) 表示候选目标位置, $N_b((X_{t-1}, Y_{t-1}, Z_{t-1}))$ 表示以 $(X_{t-1}, Y_{t-1}, Z_{t-1})$ 为中心, 以 σ 为半径, 同时在 $\vec{r}$ 方向上的邻域, $H_i((X_t^c, Y_t^c, Z_t^c))$ 表示候选点投影到不同数据源空间的模板和候选区域的相似性度量, 例如可以是视频图像空间, 也可以是 RGB-D 序列空间, ω_i 是针对不同源数据对最终定位精度影响的调节参数, 需要根据实验来决定. 上式的优化求解是一个多目标的非线性复杂优化问题, 为了满足实时跟踪和定位的应用, 这里采用粒子群优化方法来解决上述优化问题, 即在候选点空间随机选择粒子点, 通过粒子的迭代实现高效求解.

融合多源异构数据的定位跟踪最终得到的结果是目标的三维空间位置信息、目标在各个数据源提取的目标区域或位置, 以及基于这些区域的特征表示, 是包含多层次的数据和信息时空数据流, 这里按照数据到信息的抽象层次将多源异构数据的跟踪结果表示为目标依时间顺序排列的多层次结构, 以更好地满足对时空数据的管理, 实现时空数据的共享和复用. 具体地, 把时空数据流的表示分为 4 个层次, 如图 8 所示, 一是数据表达层, 该层通过各种感知设备获取目标的不同模态的观测信息; 二是特征表达层, 是对目标的位置属性、时空对齐属性、表观属性和关联属性的特征描述; 三是语义表达层, 该层通过观测目标特征信息组合和抽象, 实现对时空数据的更高层次的表示, 获得语义关联信息, 这里主要是指目标的位置和外观的语义信息; 四是应用表达层, 即通过从底层数据、特征到高层语义的多层次数据流表示, 为不同的后续应用, 例如行为分析、状态评估和事件检查等提供适配的表达. 时空数据流表达模型形式化地描述了多源异构观测信息, 可用于多种不同类型数据的表达, 同时多层次的表示形式有利于不同节点的信息交互, 并且可以满足多样化的应用需求.

5 基于多源异构数据融合的目标跟踪定位实验及分析

为了验证多源异构数据融合的目标跟踪和定位方法的有效性, 我们搭建了由 4 路高清视频摄像

机、6 路深度采集设备以及 15 个 Wi-Fi 无线 AP 和 10 部具有 Wi-Fi 功能的智能移动终端组成的多源异构数据采集平台, 其中高清视频摄像头采用海康 DS-2CD8255F-E 监控摄像头, 该设备数据采集的帧率为 30 fps, 图像分辨率为 1920×1080 ; 深度采集设备采用 MS Kinect 相机, 该设备以 30 fps 帧率同时获取彩色视频 (640×480) 和 16 位深度图像 (320×240); 无线 AP 采用符合 IEEE 802.11b/g/n 标准的工作频率为 2.4 GHz 的 TP-LINK TL-WR740N 无线路由设备; 移动终端采用联想 K860 智能手机, 其 Wi-Fi 信号扫描频率为 1.5 次/秒. 基于上述数据采集平台, 我们在实际场景中采集了大量多源异构数据进行相关的跟踪定位实验. 下面是单源数据以及多源数据融合的目标跟踪试验.

5.1 视频跟踪与融合视频和深度信息的跟踪对比试验

本实验我们将仅使用视频图像信息的跟踪方法与融合图像和深度信息的跟踪方法进行了比较. 实验中使用了两段 RGB-D 视频序列, 在序列 1 中, 目标行人从屋内移动到有较大光照差别的屋外. 如果只使用传统的基于图像特征的方法进行跟踪, 目标走到屋外时跟踪失败, 但融合视频和深度信息的跟踪方法取得了理想的跟踪效果, 实验结果如图 9(a) 所示, 图中第 1 行图像序列表示只用视频图像信息的跟踪效果; 第 2 行图像序列表示融合视频和深度信息的跟踪结果; 第 3 行图像序列表示融合跟踪结果显示在深度图像上的效果. 在序列 2 中, 目标行人逐渐移动到一个与目标有相似外观的人并发生遮挡. 当遮挡发生时, 只使用视频图像信息的跟踪方法不能正确跟踪目标, 但融合视频和深度信息的跟踪方法可以正确跟踪目标, 结果如图 9(b) 所示. 上述实验表明, 基于视频信息的跟踪方法易受环境光照变化和目标相近图像信息的影响, 而融合视频和深度信息的跟踪方法由于深度信息引入, 利用其提供的空间线索, 使得跟踪结果更加稳定.

5.2 融合视频和深度的多路 RGB-D 序列跟踪和定位

融合视频和深度信息的跟踪方法对光照变化和目标近似图像的干扰等具有很好的鲁棒性, 但单路 RGB-D 序列仍然解决不了目标被完全持续地遮挡的情况, 因此我们提出采用多个 RGB-D 摄像机协同的跟踪方法来解决上述遮挡问题, 并实现鲁棒的跟踪和定位. 我们的实验使用两个 RGB-D 摄像机, 通过两路 RGB-D 序列的特征融合, 实现对有持续遮挡的运动目标的跟踪, 实验结果如图 10 所示, 图中第 1 列图像序列是跟踪结果在视频空间上的显示; 第 2 列图像序列是跟踪结果在深度空间上的显示; 第 3 列图像序列是在俯视平面上的定位轨迹和实际路径的对比. 实验中自 338 帧起红色目标在摄像机 1 中被持续遮挡, 直到第 479 帧才又出现, 本文的多视融合跟踪方法可跟踪被持续遮挡的目标, 并且在图像平面跟踪的同时还可以对目标进行三维定位, 获得与实际路径误差较小的俯视平面运动轨迹.

5.3 无线定位与视频信息融合的跟踪

将无线定位与视频信息融合进行目标跟踪的目的是要解决遮挡发生时视频跟踪出现的目标丢失或误跟的问题, 主要利用无线定位的与视频跟踪在图像平面上的位置一致性进行判别, 并及时纠正视频跟踪的错误. 实验使用的视频序列是两个目标相向而行, 在发生遮挡后再折返. 只采用视频信息进行跟踪的结果如图 11 所示, 当两个跟踪目标在图 11(b) 处相遇并交错时跟踪发生错误, 目标的标记已经交换如图 11(c) 所示. 而在图 11(d)~(f) 中, 当两个目标返回并重新遮挡交错时跟踪标记则集中在一个目标上. 但在采用无线定位与视频信息融合的跟踪实验中, 在目标进行交错、遮挡以及折返时, 仍然能够正确跟踪目标, 如图 12 所示. 上述实验表明, 利用无线定位信息提供的三维空间信息, 有助于

图 9 视频跟踪与融合视频和深度信息的跟踪的对比

Figure 9 The visual tracking results compared with video-depth fusion tracking results. (a) Illumination variation; (b) interference of similar appearance

图 10 融合视频和深度的多路 RGB-D 序列跟踪和定位

Figure 10 Positioning and tracking based on the fusion of multi-view RGB-D sequences. (a) Tracking result of camera 1; (b) tracking result of camera 2

图 11 基于视频的多目标跟踪结果

Figure 11 Multi-object tracking results of the visual tracking method. (a) At the beginning; (b) occlusion happening; (c) after occlusion; (d) returning; (e) occlusion re-happening; (f) at the end

纠正遮挡发生时视频跟踪的错误.

图 12 融合无线定位与视频信息的多目标跟踪结果

Figure 12 Multi-object tracking results of the fusion method using wireless positioning results. (a) At the beginning; (b) occlusion happening; (c) after occlusion; (d) returning; (e) occlusion re-happening; (f) at the end

图 13 基于多源数据融合的跟踪结果在两路视频中的显示

Figure 13 Multi-object tracking result in two-view video based on multi-source heterogeneous data fusion

5.4 融合无线定位、多视视频和深度信息的跟踪定位

为验证融合多源异构数据的定位跟踪方法的有效性, 我们在一栋大楼的大厅内进行了较大范围的目标定位与跟踪实验, 实验中目标行人手持无线终端沿设定的路径行走, 场景中同时有其他行人出现, 在局部区域常会发生多个目标的遮挡或交错. 采用融合无线定位、多路视频和多路 RGB-D 序列数据的目标定位与跟踪方法, 在遮挡发生或行人交错时能够比较准确地跟踪目标并获得定位信息, 图 13 表示发生遮挡和交错时基于数据融合的跟踪结果在两路视频中的显示. 图 14 给出了基于多源异构数据融合的方法获得的目标在俯视平面的轨迹, 与目标的实际行走路径比较, 其结果表明本文提出的方法具有较高的定位跟踪精度.

6 结论

本文针对物联网多源异构数据的处理和融合问题, 以面向目标跟踪定位的无线信号、视频和深度数据的融合方法展开研究, 提出多源异构数据的处理和特征表示方法, 以及多层次的融合方法. 在数

图 14 不同的跟踪定位方法在俯视平面上的轨迹与真实路径的对比 (单位: 米)

Figure 14 Result comparison between different algorithms and actual path in bird view (unit:m)

据处理和特征表示方面, 提出了利用时空上下文信息进行多源异构数据的协同关键帧分析方法, 实现了多源异构数据时序一致的数据流表达. 在数据融合方面, 提出无线定位和视频跟踪位置一致性约束的融合方法, 视频和深度的特征级融合方法以及无线、视频和深度信息在决策级的融合方法. 实际复杂场景的目标跟踪定位实验表明, 本文提出的基于多源异构数据融合的跟踪定位方法, 能解决基于单源数据的传统跟踪定位方法所面临的遮挡等困难问题, 并且能获得目标较为准确的三维位置.

本文的方法对于无线信号等具有较大观测噪声数据的处理效果还不够理想, 对于视频和深度数据的特征表示方法研究还不够深入, 实验所采用的数据规模相对较小. 下一步的研究工作主要侧重于提高无线信号定位的精度、寻求更加适合多源异构数据特性的数据处理和特征表示方法以及面向大规模多源异构数据融合的高效方法.

致谢 感谢闻征、周薇、马俊、熊林欣、刘洋、徐欢、吴凯在数据采集过程中付出的辛勤劳动.

参考文献

- 1 Nakamura E F, Loureiro A A F, Frery A C. Information fusion for wireless sensor networks: Methods, models, and classifications. *ACM Comput Surv*, 2007, 39: A9/1-55
- 2 Xiao J, Ribeiro A, Luo Z, et al. Distributed compression-estimation using wireless sensor networks. *IEEE Signal Process Mag*, 2006, 23: 27-41
- 3 Yuan Y, Kam M. Distributed decision fusion with a random-access channel for sensor network applications. *IEEE Trans Instrum Meas*, 2004, 53: 1339-1344
- 4 Shah S F A, Ribeiro A, Giannakis G B. Bandwidth-constrained MAP estimation for wireless sensor networks. In: *Proceedings of the 39th Asilomar Conference on Signals, Systems and Computers*. Pacific Grove: IEEE Signals, Systems and Computers, 2005. 215-219
- 5 Guestrin C, Bodik P, Thibaux R, et al. Distributed regression: An efficient framework for modeling sensor network data. In: *Proceedings of the International Symposium on Information Processing in Sensor Networks*. Berkeley: IEEE Computer Society Press, 2004. 1-10

- 6 Willett R, Martin A, Nowak R. Backcasting: Adaptive sampling for sensor networks. In: Proceedings of the International Symposium on Information Processing in Sensor Networks. Berkeley: IEEE Computer Society Press, 2004. 124–133
- 7 Jin G, Nittel S. NED: An efficient noise-tolerant event and event boundary detection algorithm in wireless sensor networks. In: Proceedings of the International Conference on Mobile Data Management. Nara: IEEE Computer Society Press, 2006. 153–161
- 8 Jain A, Chang E, Wang Y. Adaptive stream resource management using Kalman filters. In: Proceedings of the ACM SIGMOD International Conference on Management of Data. Paris: ACM Special Interest Group on Management of Data, 2004. 11–22
- 9 Mascolo C, Musolesi M. SCAR: Context-aware adaptive routing in delay tolerant mobile sensor networks. In: Proceedings of the International Conference on Communications and Mobile Computing. Vancouver: IEEE Computer Society Press, 2006. 533–538
- 10 Raviraj P, Sharif H, Hempel M, et al. MOBMAC: An energy efficient and low latency MAC for mobile wireless sensor networks. In: Proceedings of the Systems Communications. Montreal: IEEE Computer Society Press, 2005. 370–375
- 11 Aslam J, Butler Z, Constantin F, et al. Tracking a moving object with a binary sensor network. In: Proceedings of the 1st International Conference on Embedded Networked Sensor Systems. Los Angeles: IEEE Computer Society Press, 2003. 150–161
- 12 Nakamura E F, Figueiredo C M, Loureiro A A. Information fusion for data dissemination in self-organizing wireless sensor networks. In: Proceedings of the International Conference on Networking. Reunion Island: IEEE Computer Society Press, 2005. 585–593
- 13 Shu H, Liang Q. Fuzzy optimization for distributed sensor deployment. In: Proceedings of IEEE Wireless Communications and Networking Conference. New Orleans: IEEE Computer Society Press, 2005. 1903–1908
- 14 Gupta I, Riordan D, Sampall S. Cluster-head election using fuzzy logic for wireless sensor networks. In: Proceedings of the Annual Communication Networks and Services Research Conference. Halifax: IEEE Computer Society Press, 2005. 255–260
- 15 Yusuf M, Haider T. Energy-aware fuzzy routing for wireless sensor networks. In: Proceedings of the IEEE International Conference on Emerging Technologies. Islamiabad: IEEE Computer Society Press, 2005. 63–69
- 16 Wallace J, Pesch D, Rea S, et al. Fuzzy logic optimization of MAC parameters and sleeping duty-cycles in wireless sensor networks. In: Proceedings of the Vehicular Technology Conference. Dallas: IEEE Computer Society Press, 2005. 1824–1828
- 17 Filillidis A, Jain L C, Martin N. Fusion of intelligent agents for the detection of aircraft in SAR images. *IEEE Trans Pattern Anal Mach Intell*, 2000, 22: 378–384
- 18 Friedlander D S, Phoha S. Semantic information fusion for coordinated signal processing in mobile sensor networks. *Int J High Perform Comput Appl*, 2002, 16: 235–241
- 19 Gerónimo D, Sappa A D, Ponsa D, et al. 2D-3D-based on-board pedestrian detection system. *Comput Vis Image Underst*, 2010, 114: 583–595
- 20 Szarva M, Yoshizawa A, Yamamoto M, et al. Pedestrian detection with convolutional neural networks. In: Proceedings of the IEEE International Conference on Intelligent Vehicles Symposium. Nice: IEEE Computer Society Press, 2005. 224–229
- 21 Broggi A, Bertozzi M, Fascoli A, et al. Shape-based pedestrian detection. In: Proceedings of the IEEE Conference on Intelligent Vehicle Symposium. Dearborn: IEEE Computer Society Press, 2000. 215–220
- 22 Premebida C, Nunes U. A multi-target tracking and GMM classifier for intelligent vehicles. In: IEEE International Conference on Intelligent Transportation Systems. Toronto: IEEE Computer Society Press, 2006. 313–318
- 23 Spinello L, Siegwart R. Human detection using multimodal and multidimensional features. In: IEEE International Conference on Robotics and Automation. California: IEEE Computer Society Press, 2008. 3264–3269
- 24 Oliveira L, Nunes U, Peixoto P, et al. Semantic fusion of laser and vision in pedestrian detection. *Pattern Recognit*. 2010, 43: 3648–3659
- 25 Liu Y, Yang Z, Wang X, et al. Location, localization, and localizability. *J Comput Sci Technol*, 2010, 25: 274–297
- 26 Weng S K, Kuo C M, Tus K. Video object tracking using adaptive Kalman filter. In: International J Visual Communication and Image Representation. Orlando: ACM Academic Press, 2006. 17: 1190–1208

- 27 Song Y, Li Q L, Sun F C. Shannon Entropy-Based Adaptive Fusion Particle Filter for Visual Tracking. In: IEEE Chinese Conference on Pattern Recognition. Nanjing: IEEE Computer Society Press, 2009. 1–5
- 28 Li H, Shen C, Shi Q. Real-time visual tracking using compressive sensing. In: IEEE Conference on Computer Vision and Pattern Recognition. Providence: IEEE Computer Society Press, 2011. 1305–1312
- 29 Mei X, Ling H. Robust visual tracking and vehicle classification via sparse representation. *IEEE Trans Pattern Anal Mach Intell*, 2011, 33: 2259–2272
- 30 Bahl P, Padmanabhan V N. Radar: an in-building RF-based user location and tracking system. In: Proceedings of the 9th Annual Joint Conference of the IEEE Computer and Communications Societies. Tel Aviv: IEEE Computer Society Press, 2000. 2: 775–784
- 31 Candès E J, Recht B. Exact matrix completion via convex optimization. *International J Foundations of Computational Mathematics*, 2009, 9: 717–772
- 32 Candès E J, Plan Y. Tight oracle inequalities for low-rank matrix recovery from a minimal number of noisy random measurements. *IEEE Trans Inf Theory*, 2011, 57: 2342–2359
- 33 Dakkak A, Husain A. Recovering missing depth information from Microsoft's Kinect. http://www.contrib.andrew.cmu.edu/ammahr/projects/vis_project.pdf
- 34 Liu J Y, Gong X J, Liu J L. Guided inpainting and filtering for Kinect depth maps. In: IEEE International Conference on Pattern Recognition. Tsukuba: IEEE Computer Society Press, 2012. 2055–2058
- 35 Khoshelham K, Elberink S O. Accuracy and resolution of kinect depth data for indoor mapping applications. *J Sensors*, 2012, 12: 1437–1454
- 36 Zhang K H, Zhang L, Yang M H. Real-time compressive tracking. In: Conference of the European Conference on Computer Vision. Firenze: Springer Berlin Heidelberg, 2012. 864–877
- 37 Donoho D L. Compressed sensing. *IEEE Trans Inf Theory*, 2006, 52: 1289–1306
- 38 Viola P, Jones M. Rapid object detection using a boosted cascade of simple features. In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Kauai: IEEE Computer Society Press, 2001. 1: I-511–518
- 39 Li S, Zhang Z. Floatboost learning and statistical face detection. *IEEE Trans Pattern Anal Mach Intell*, 2004, 26: 1–12
- 40 Tang F, Harville M, Tao H, et al. Fusion of local appearance with stereo depth for object tracking. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, Anchorage: IEEE Computer Society Press, 2008. 1–8
- 41 Bai T X, Li Y F. Robust visual tracking with structured sparse representation appearance model. *Pattern Recognit*, 2012, 45: 2390–2404
- 42 Doucet A, De Freitas N, Gordon N. An introduction to sequential Monte Carlo methods. New York: Springer-Verlag, 2001. 3–14
- 43 Swain M J, Ballard D H. Color indexing. *Int J Comput Vis*, 1991, 7: 11–32
- 44 Bhattacharyya, A. On a measure of divergence between two statistical populations defined by their probability distributions. *Bull Calcutta Math Soc*, 1943, 35: 99–109

Multi-source heterogeneous data fusion method and its application in object positioning and tracking

HU YongLi, PIAO XingLin, SUN YanFeng* & YIN BaoCai

Beijing Municipal Key Laboratory of Multimedia and Intelligent Software Technology, College of Computer Science and Technology, Beijing University of Technology, Beijing 100124, China

*E-mail: yfsun@bjut.edu.cn

Abstract The general sensing of Internet of Things (IoT) brings magnanimous sensing data, which shows significant multi-source heterogeneous property. How to process the multi-source heterogeneous sensing data intelligently and efficiently is a challenging problem. Although data fusion is considered an effective approach to processing multi-modal data and extract the hiding valuable information, there are many problems to be solved for multi-source heterogeneous data fusion, especially the unstructured video multimedia information. In this paper, the multi-source heterogeneous data fusion problem is explored and a multi-level fusion method is proposed and applied in object positioning and tracking using wireless signal, video and depth data. In the proposed method, the main opponents of data fusion are deeply studied, including the processing methods of different types of data, the feature representation and different level data fusion methods. For different types of data, from their inherent characteristics, different fusion methods at different levels are adopted here to mine the correlative relations and derive the valuable information of the heterogeneous data. The proposed method is evaluated by the object tracking and positioning experiments in actual complicated scenarios. The results show that the proposed multi-source heterogeneous data fusion based method can solve the difficulties in the traditional single data based tracking method, such as the illumination variation, occlusion and the clutter of background. Additionally, the proposed method can estimate the three-dimensional position of the tracking object with high accuracy.

Keywords multi-source heterogeneous, information fusion, tracking and positioning, feature representation, Internet of Things

HU YongLi was born in 1973. He received his Ph.D. degree from Beijing University of Technology in 2005. He is an associate professor in the College of Computer Science and Technology at Beijing University of Technology. He is a researcher at the Beijing Municipal Key Laboratory of Multimedia and Intelligent Software Technology. His research interest includes computer graphics, pattern recognition and multimedia technology.

PIAO XingLin was born in 1987. He received the bachelor degree from Jilin University in 2010. He is currently a Ph.D. candidate at the College of Computer Science and Technology, Beijing University of Technology. His research interest covers target tracking, WSNs sensing technology, pattern recognition and multimedia technology.

SUN YanFeng was born in 1964. She received her Ph.D. degree from Dalian University of Technology in 1993. She is a professor and doctoral supervisor at the College of Computer Science and Technology, Beijing University of Technology. She is a membership of China Computer Federation. Her research interest includes multi-functional perception and image processing.

YIN BaoCai was born in 1963. He received his Ph.D. degree from Dalian University of Technology in 1993. He is a professor and doctoral supervisor at the College of Computer Science and Technology, Beijing University of Technology. He is a membership of China Computer Federation. His research interest covers multimedia, multifunctional perception, virtual reality and computer graphics.