



# 光学影像序列中基于多视角聚类的群组行为分析

李学龙<sup>1\*</sup>, 陈穆林<sup>2</sup>, 王琦<sup>2,3</sup>

1. 中国科学院西安光学精密机械研究所, 西安 710119

2. 西北工业大学计算机学院与光学影像分析与学习中心, 西安 710072

3. 西北工业大学无人系统研究院, 西安 710072

\* 通信作者. E-mail: xuelong.li@opt.ac.cn

收稿日期: 2017-12-25; 接受日期: 2018-02-22; 网络出版日期: 2018-08-23

**摘要** 群组行为分析是光学影像序列分析中的一项重要课题, 在近年来引起了人工智能领域研究人员的广泛关注. 与行人个体相比, 群组能提供更高层的语义表示, 为分析人群运动模式提供基础. 本文将人群影像序列中的影像块作为研究对象, 提出了一种基于多视角聚类的群组行为分析方法, 对运动模式不同的群组进行区分, 主要研究内容有: (1) 提出了基于特征点的影像块构图方法, 从交互关系、空间位置、运动方向分布, 以及运动规律等方面衡量影像块之间的关系; (2) 提出了一种多视角聚类方法, 通过融合多种特征对每个影像块分配类标签, 并引入图多样性正则项以避免特征冗余; (3) 提出了一种类合并方法, 根据类内特征点的运动方向和类中心位置坐标, 对关联度较高的类别进行合并, 自动确定最终群组数目. CUHK 人群数据集上的实验结果证明了该方法能够准确划分出影像数据中的群组. 另外, 与现有方法相比, 本文提出的多视角聚类方法也在不同数据集上取得了较好的实验结果.

**关键词** 人群分析, 群组行为分析, 聚类算法, 图聚类, 多视角聚类

## 1 引言

近几十年来人口快速增长, 人群聚集活动日益频繁, 随之而来的安全问题给公共场所管理带来了严峻挑战, 使得公众对智能视频监控的需求更加迫切. 光学影像序列分析利用成像系统, 结合计算机视觉技术对影像数据进行解析, 成为当前的研究热点. 群组行为分析是光学影像序列分析领域的重要课题. 在拥挤场景中, 人群普遍以群组形式进行运动. 对影像数据中的群组进行分析, 可以获取人群运动的高层语义信息. 因此, 群组行为分析成为一项研究重点, 并具有广泛的实际应用价值, 在行人跟踪<sup>[1]</sup>、人群计数<sup>[2]</sup>、场景语义分割<sup>[3]</sup>、异常检测<sup>[4]</sup>等领域中扮演了关键角色. 目前, 国内外涌现了大量群组行为分析方法. 然而, 由于人群场景中存在严重的遮挡与背景噪声, 这些算法仍存在研究对象局部化, 特征利用不充分等问题.

**引用格式:** 李学龙, 陈穆林, 王琦. 光学影像序列中基于多视角聚类的群组行为分析. 中国科学: 信息科学, 2018, 48: 1227–1241, doi: 10.1360/N112017-00284  
Li X L, Chen M L, Wang Q. Multiview-based group behavior analysis in optical image sequence (in Chinese). Sci Sin Inform, 2018, 48: 1227–1241, doi: 10.1360/N112017-00284

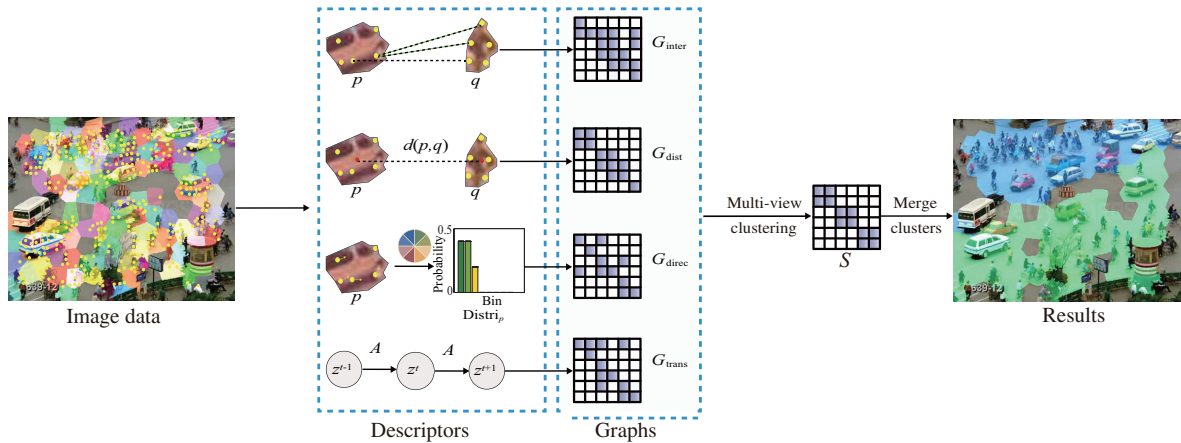


图 1 (网络版彩图) 算法流程图. 首先, 将影像数据分割成块. 然后, 对从交互关系、空间距离、运动方向分布及运动规律 4 个方面对影像块构造相似图. 进而通过多视角聚类学习到最优图, 完成影像块类别的划分. 最后, 对关联度高的影像块进行合并, 得到结果.

Figure 1 (Color online) The pipeline of the proposed method. First, the image data is segmented into patches. Then, we construct the similarity graphs for the image patches from four aspects, including interaction, spatial distance, motion direction, and motion transition. And a multiview clustering is proposed to cluster the patches. Finally, the clusters with high consistency are merged into final groups.

对个体运动的研究, 是分析群组行为的基础前提. 在大规模人群场景中, 严重的遮挡使得对个体的准确检测与追踪难以实现, 因此现有的人群分析方法大多以运动特征点 [5~11] 或固定粒子 [3, 4, 12] 为研究对象, 从而避免对行人个体的直接提取. 然而, 在同一个体上可能会存在多个运动特征点或固定粒子, 这些点的运动模式可能会存在很大差异. 例如, 行人个体运动时, 头部点和脚部点的运动方向可能是相反的. 此外, 由于过于局部化, 特征点或粒子的运动并不稳定, 往往会出现运动方向抖动的现象. 因此, 运动特征点和固定粒子并不能很好地反映个体的运动规律. 针对此问题, Li 等 [13] 提出利用邻近点的速度方向分布来弥补单个特征点的局部性, 但时序线索的丢失限制了该方法的效果.

群组行为分析的另一个难点在于特征的利用. 人群场景中个体的密集分布, 使得颜色、纹理等常用特征难以利用. 现有的方法 [5, 6, 8~11, 14] 大多以空间位置、运动速度为线索, 将距离近、运动方向相似的点视为同一群组, 缺少对运动规律、结构上下文等特征的挖掘. Sharma 等 [15] 在划分群组时使用了轨迹形状、多尺度密度等特征, 但其直接将多种特征合并为同一个向量, 未根据重要性对特征进行区分. Li 等 [13] 利用多视角聚类算法, 分别从运动方向、结构上下文两个方面构建相似图, 自动学习不同相似图的权重, 取得了较好结果. 但是, 在聚类过程中未考虑特征的多样性, 存在特征冗余问题. 因此, 如何设计更有力的特征, 并提出更合理的特征利用方法, 是群组行为分析中亟待解决的问题.

为减少上述问题带来的影响, 本文提出一种基于多视角聚类的群组行为分析方法 (multiview-based group behavior analysis, MGBA), 以影像块为研究对象, 根据内部特征点的运动模式差异, 将其分为不同群组. 首先, 影像数据被分割成块. 与特征点相比, 影像块具有更强的稳定性与全局性, 能够更好地反映人群的运动模式. 同时, 提出 4 种构图方法, 分别从交互关系、空间位置、运动方向分布及运动规律等方面对不同影像块进行比较. 然后, 提出一种权重自学习的多视角聚类方法, 在融合 4 种相似图的同时, 引入图多样性正则项, 强调了相似图的多样性. 最后, 由于假设的类别数目并不准确, 被分为不同类别的影像块可能属于同一群组, 因此提出一种基于运动方向和类中心位置的类合并方法, 得到最终的群组. 本文算法的流程图如图 1 所示.

本文的章节安排如下. 第 2 节简要回顾了群组行为分析和多视角聚类的相关工作; 第 3 节介绍从运动方向分布、空间位置、交互关系及运动规律等方面对影像块构建相似图的方法; 第 4 节提出了基于图多样性的多视角聚类方法, 并设计了求解算法; 第 5 节介绍基于运动方向和类中心位置的类合并方法; 第 6 节展示并讨论群组行为分析方法和多视角聚类方法的实验结果; 第 7 节总结全文.

## 2 相关工作

本文主要研究人群场景中的群组划分问题, 并提出一种多视角聚类方法. 本节对群组划分及多视角聚类的相关工作进行回顾.

群组划分是人群行为分析中的重要研究课题, 在近年来出现了大量群组划分算法, 将行为模式不同的个体划分为不同群组. Zhou 等<sup>[8]</sup>利用特征点空间位置关系及运动方向相似度, 计算点之间的拓扑关系, 以处理具有复杂结构的人群. Wu 等<sup>[7]</sup>设计了点密度特征, 对特征点进行聚类. 上述方法只能处理单帧影像, 无法利用影像序列中的时序线索. Zhou 等<sup>[5]</sup>检测并跟踪场景中的运动特征点, 如果两个特征点在连续一段时间内保持临近关系, 且运动方向一致, 则被认为属于同一群组. Shao 等<sup>[6]</sup>在群组划分中加入了轨迹变化规律先验, 将群组中的离群点去除. Wang 等<sup>[10]</sup>根据运动轨迹计算特征点相似度, 并计算特征点间拓扑关系, 完成群组划分. Ali 等<sup>[12]</sup>利用 Lyapunov 算子对人群场景中的粒子光流进行建模, 对场景进行划分. Wang 等<sup>[3]</sup>利用热量传播原理, 对粒子光流进行处理, 得到平滑的运动向量. 这 5 种方法仅利用了特征点或粒子的空间位置与运动方向信息, 特征利用不充分. 且以上所有方法均将特征点作为个体, 研究对象过于局部化, 易受轨迹抖动问题的干扰. Li 等<sup>[13]</sup>在对特征点进行处理时引入了结构上下文信息, 缓解了研究对象局部化的问题, 但该方法仍缺少对时序线索的利用. 同时, 该方法未考虑不同视角间可能存在的重合信息, 存在特征冗余问题.

多视角聚类是近年来机器学习领域的研究热点, 其目的在于整合多视角信息, 挖掘数据内在结构, 完成聚类任务. Kumar 等<sup>[16]</sup>假设所有视角权重相似, 将协同规范思想与谱聚类结合, 使各视角获得一致的聚类结果. Cai 等<sup>[17]</sup>提出一种多模态谱聚类方法, 利用 Laplace 矩阵对多视角数据进行整合, 并寻找各视角相似图的最优线性组合, 作为最优图. Li 等<sup>[18]</sup>提出选取具有代表性的数据点进行构图, 并利用二值图拟合各视角相似图, 以处理大规模数据. Xia 等<sup>[19]</sup>对每个视角的相似图进行去噪处理, 利用 Markov 链学习一个低秩转换矩阵, 并使用 Markov 链得到聚类结果. Liu 等<sup>[20]</sup>提出在多视角相似图线性组合的邻域内, 找到最能代表数据内在结构的图, 以扩大最优图的搜索空间. Liu 等<sup>[21]</sup>与 Cao 等<sup>[22]</sup>在整合各视角相似图时, 引入多样性正则项, 避免对重合度高的相似图分配相似权重. 以上方法均不能直接得到聚类结果, 需要使用  $k$  均值方法作为后处理步骤, 因此结果易受初始条件影响. Li 等<sup>[13]</sup>与 Nie 等<sup>[23]</sup>对 Laplace 矩阵的秩进行限制, 寻找最优相似图, 所得相似图可直接作为类别指示矩阵. 但以上两种方法未考虑图多样性, 当相似图重合度较高时, 难以挖掘各视角间的差异性信息.

## 3 相似图构建

群组由密切联系的个体构成, 对个体运动的分析是实现群组研究的关键. 在复杂的人群场景中, 行人个体难以提取, 因此本文采用运动特征点代替个体. 考虑到特征点过于局部化, 本方法将影像数据分割为影像块, 如图 1 所示. 同一影像块中的特征点之间距离较近, 且运动模式基本一致, 可看作统一整体. 因此, 将影像块作为研究对象, 能够有效缓解特征点局部化带来的问题. 具体地, 本工作利用 gKLT 方法<sup>[8]</sup>检测并跟踪运动特征点, 得到特征点的运动轨迹, 并使用 SLIC 方法<sup>[24]</sup>将每一帧影像数据分

割为  $N_{\text{patch}}$  块, 移除不含任何特征点的影像块. 其他跟踪方法<sup>[25]</sup> 和图像分割方法<sup>[26, 27]</sup> 同样适用. 下面对 4 种相似图构建方法进行阐述.

**交互关系.** 人群场景中的个体之间存在密切的交互关系. 行为学研究<sup>[28]</sup> 表明, 个体间的频繁交互促使了群组的产生. 因此, 在计算影像块相似度时, 需考虑其内部特征点的交互关系. 首先, 对特征点间的交互关系进行判断. 对于任意特征点  $i$ , 根据在当前影像帧中的空间距离, 利用  $k$  近邻方法找到其  $k$  个近邻. 假设特征点  $j$  是  $i$  的近邻, 记  $i$  和  $j$  的运动方向分别为  $v_i$  与  $v_j$ , 如果两者运动方向相似度大于阈值  $\text{th}_1$ , 则认为两者之间存在交互, 将两者交互度  $\text{Inter}(i, j)$  设为 1:

$$\text{Inter}(i, j) = \begin{cases} 1, & \text{if } \cos(v_i, v_j) \geq \text{th}_1, \\ 0, & \text{else.} \end{cases} \quad (1)$$

对于影像块  $p$  和  $q$ , 如果两者之间有 3 对以上特征点交互度为 1, 则将  $p$  与  $q$  的交互关系相似度  $G_{\text{inter}}(p, q)$  设为 1:

$$G_{\text{inter}}(p, q) = \begin{cases} 1, & \text{if } \sum_{i, j} \text{Inter}(i, j) \geq 3, (i \in p, j \in q), \\ 0, & \text{else,} \end{cases} \quad (2)$$

对每一对影像块计算相似度, 即得到交互关系相似图  $G_{\text{inter}}$ . 如果两个影像块中的特征点存在密切交互, 则其相似度更大. 因此, 本方法能够充分挖掘个体间的交互关系.

**空间距离.** 群组往往由空间距离接近的个体组成. 同样地, 对于两个影像块, 其空间距离越小, 则属于同一群组的概率越大. 假设影像块  $p$  的中心位置为  $(\hat{x}_p, \hat{y}_p)$ , 则  $p$  与  $q$  的空间距离为

$$d(p, q) = \sqrt{(\hat{x}_p - \hat{x}_q)^2 + (\hat{y}_p - \hat{y}_q)^2}, \quad (3)$$

进而将  $p$  与  $q$  的空间距离相似度定义为

$$G_{\text{dist}}(p, q) = \exp\left(-\frac{d(p, q)}{h}\right), \quad (4)$$

式中  $h$  为影像高度. 影像块距离越大, 相似度越小, 被分为相同群组的可能性越小. 因此本方法能够掌握个体间的空间位置关系.

**运动方向分布.** 运动方向是判断个体行为是否相似的重要依据. 为衡量影像块相似度, 需掌握其内部特征点的运动方向分布. 如图 1 所示, 将方向空间分为 8 个通道, 则影像块  $p$  的运动方向分布  $\text{Distri}_p$  即为其内部特征点在 8 个通道上的分布,  $\text{Distri}_p$  的第  $a$  个元素为

$$\text{Distri}_p(a) = \text{pro}(v_i \in \text{bin}_a | i \in p), \quad (5)$$

其中  $\text{pro}()$  为概率,  $v_i \in \text{bin}_a$  表示特征点  $i$  的速度属于第  $a$  个通道 ( $a = [1, \dots, 8]$ ). 对于影像块  $p$  和  $q$ , 其运动方向分布相似度被定义为

$$G_{\text{direc}}(p, q) = \exp\{-\text{KLD}(\text{Distri}_p || \text{Distri}_q)\}, \quad (6)$$

其中  $\text{KLD}(\text{Distri}_p || \text{Distri}_q)$  为  $\text{Distri}_p$  与  $\text{Distri}_q$  间的 KL 散度 (Kullback-Leibler divergence)<sup>[29]</sup>, 用于计算不同分布间的差异.

**运动规律.** 同一群组中的个体, 运动规律通常相似. 运动规律是个体空间位置在时间序列上的变化规律, 反映了个体的行为模式, 因此这里对影像块内特征点的运动规律进行研究. 对于影像块  $p$ , 计算其内部特征点在每一帧中的平均空间位置, 为减少计算复杂度, 仅取前 6 帧到当前帧的值, 可得到长度为 6 的轨迹  $\{(\tilde{x}_p^t, \tilde{y}_p^t)\}_{t=1}^6$ , 将其称为  $p$  的运动轨迹. 个体空间位置的变化在时间序列上是连续的, 因此可假设  $p$  在  $t$  时刻的位置可由  $t-1$  时刻的位置变换得到. 记  $(\tilde{x}_p^t, \tilde{y}_p^t, 1)^T$  为  $z_p^t$  ( $1$  为偏置,  $T$  为转置符号), 则可建立如下模型:

$$z_p^t = A_p z_p^{t-1} + \delta^t, \quad (7)$$

其中  $A_p \in \mathbb{R}^{n \times n}$  为空间位置变换矩阵,  $\delta^t$  为高斯噪声. 已知  $\{z_p^t\}_{t=1}^6$ , 可利用 Kalman 平滑器<sup>[30]</sup> 得到变换矩阵  $A_p$ ,  $A_p$  中元素大小即体现了影像块  $p$  内所有特征点的运动规律. 然后, 根据空间位置变换矩阵, 将影像块运动规律相似度定义为

$$G_{\text{trans}}(p, q) = \frac{\text{tr}(A_p^T A_q)}{\|A_p\|_F \times \|A_q\|_F}, \quad (8)$$

式中  $\|\cdot\|_F$  代表 Frobenius 范数. 若  $p$  和  $q$  的变换矩阵相似, 则  $\text{tr}(A_p^T A_q)$  较大, 而分母项  $\|A_p\|_F \times \|A_q\|_F$  的作用是将相似度控制在 0 到 1 之间. 通过挖掘空间位置变换特性, 本方法对影像块内特征点的运动规律进行了合理探究, 并充分利用了时序线索.

## 4 基于图多样性的多视角聚类

本文的研究目标是将影像块划分为不同群组, 实质上是对影像块进行聚类. 本节提出一种基于图多样性的多视角聚类方法 (multiview clustering with diversity regularization, MCD), 结合上文所提的 4 种相似图, 对影像块分配类标签. 并针对所提聚类方法, 设计了一种迭代求解算法.

### 4.1 聚类方法

多视角聚类是机器学习领域近年来的研究热点. 随着信息搜集速度的提高, 同一目标可用来自多种视角的特征来表示. 传统的聚类方法通常只利用单一视角的特征, 缺乏对多视角信息的整合, 因此难以充分挖掘数据的内在结构. 多视角聚类方法对多个视角的特征进行有机结合, 利用不同视角的互补性, 对数据进行全面分析, 在近年来被广泛应用于场景分类、图像分类, 及文本分类等实际任务中, 并取得了较好结果.

在群组的划分中, 可利用交互关系、运动方向分布、空间距离、运动规律等 4 个视角的相似图, 对影像块进行多视角聚类. 现有的多视角聚类方法大多将  $k$  均值聚类作为后处理步骤, 由于  $k$  均值方法依赖于初始值, 因此这些方法难以获得稳定的结果. Li 等<sup>[13]</sup> 提出一种基于 Laplace 矩阵秩限制的方法, 结合多种相似图, 学习一个具有  $c$  个连通区域的最优相似图 ( $c$  为类别数目), 根据最优相似图内数据的连通关系, 得到最终类别. 该方法目标函数如下:

$$\begin{aligned} \min_{w, S} & \left\| S - \sum_{v=1}^{n_v} w_v G_v \right\|_F^2 \\ \text{s.t.} & w \geq 0, \sum_v w_v = 1, S \geq 0, \\ & \sum_q S_{pq} = 1, \text{rank}(L_S) = n - c, \end{aligned} \quad (9)$$

其中  $G_v$  为第  $v$  个视角的相似图,  $w_v$  为对应权重,  $S$  为所求最优相似图,  $n_v$  为视角数目,  $n$  为数据量,  $L_S$  为  $S$  的 Laplace 矩阵. 根据文献 [31] 的证明, 限制  $L_S$  的秩为  $n - c$ , 则最终得到的  $S$  具有  $c$  个连通区域. 属于同类别数据, 在  $S$  中应属于同一个连通区域. 因此,  $S$  可被看作类别指示矩阵, 得到  $S$  即完成聚类. 然而, 该方法存在特征冗余问题. 具体地, 假设相似图  $G_1$  权重较大, 则与  $G_1$  重合度高的其他相似图也会被分配较大权重, 造成特征利用的过度冗余, 以及互补信息的缺失. 为充分利用相似图互补性, 本文提出一种基于图多样性的多视角聚类方法. 首先, 定义视角相关矩阵  $H \in \mathbb{R}^{n_v \times n_v}$  为

$$H(u, v) = \frac{\text{Tr}(G_u^T G_v)}{\|G_u\|_F \times \|G_v\|_F}, \quad (10)$$

若第  $u$  和  $v$  个视角相似图重合度高, 则  $H(u, v)$  较大. 为避免关联度高的图被分配相似权重, 提出以下多样性正则项:

$$\min_w \sum_{u,v} w_u H(u, v) w_v = w^T H w. \quad (11)$$

$H(u, v)$  越大, 则  $w_u w_v$  越小, 从而使整合后的相似图更具多样性. 结合式 (9), 得到如下目标函数:

$$\begin{aligned} \min_{w,S} & \left\| S - \sum_{v=1}^{n_v} w_v G_v \right\|_F^2 + \beta w^T H w \\ \text{s.t.} & \quad w \geq 0, \sum_v w_v = 1, S \geq 0, \\ & \quad \sum_q S_{pq} = 1, \text{rank}(L_S) = n - c, \end{aligned} \quad (12)$$

其中  $\alpha$  为参数. 参考文献 [32] 的理论, 限制  $L_S$  的秩为  $n - c$ , 等同于求解如下问题:

$$\min_{F \in \mathbb{R}^{n \times c}, F^T F = I} \text{Tr}(F^T L_S F), \quad (13)$$

其中  $I \in \mathbb{R}^{n \times n}$  为单位矩阵. 因此得出最终目标函数为

$$\begin{aligned} \min_{w,S,F} & \left\| S - \sum_{v=1}^{n_v} w_v G_v \right\|_F^2 + \beta w^T H w + \lambda \text{Tr}(F^T L_S F) \\ \text{s.t.} & \quad w \geq 0, \sum_v w_v = 1, S \geq 0, \sum_q S_{pq} = 1, F^T F = I, \end{aligned} \quad (14)$$

式中  $\lambda$  为参数. 多样性正则项的引入, 使所提聚类方法能够充分挖掘不同视角间的互补信息. 同时, 对  $L_S$  的秩进行限制, 可以使  $S$  直接作为类别指示矩阵, 无需使用  $k$  均值聚类方法作为后处理步骤. 因此本方法只需一次求解即可得到聚类结果.

需要注意的是, 在通常的聚类任务中, 类别数目  $c$  是给定的. 而在群组行为分析中, 很难预测群组数目. 因此, 本文在划分群组时, 将  $c$  设为一个较大的值, 完成聚类后, 再对关联度高的类别进行合并, 如图 1 所示.

#### 4.2 求解算法

目标函数 (14) 中涉及 3 个需要优化的变量, 因此提出一种迭代求解方法, 在优化一个变量时, 固定其他变量, 迭代至收敛后得到最优解. 首先, 将  $w$  的每个元素初始化为  $\frac{1}{n_v}$ , 然后用文献 [32] 中的方法初始化  $S$ .

优化  $F$ . 当对  $F$  求解时, 固定其他变量, 式 (14) 即转化为式 (13). 最优  $F$  即为  $L_S$  的前  $c$  个最小特征值对应的特征向量集合.

优化  $S$ . 求解  $S$  时, 式 (14) 转化为

$$\begin{aligned} \min_S \quad & \left\| S - \sum_{v=1}^{n_v} w_v G_v \right\|_F^2 + \lambda \text{Tr}(F^T L_S F) \\ \text{s.t.} \quad & S \geq 0, \sum_q S_{pq} = 1, \end{aligned} \tag{15}$$

对于问题 (15) 的求解方法参见文献 [32].

优化  $w$ . 固定  $F$  和  $S$ , 优化  $w$  时, 目标函数转化为

$$\begin{aligned} \min_w \quad & \left\| S - \sum_{v=1}^{n_v} w_v G_v \right\|_F^2 + \beta w^T H w \\ \text{s.t.} \quad & w \geq 0, \sum_v w_v = 1, \end{aligned} \tag{16}$$

为将问题 (16) 分解到每一行, 需将其转换为向量形式. 将矩阵  $S$  按列拉伸为列向量  $\tilde{s} \in \mathbb{R}^{n^2 \times 1}$ , 同理将各视角相似图  $G_1, \dots, G_{n_v}$  拉伸为列向量  $\tilde{g}_1, \dots, \tilde{g}_{n_v} \in \mathbb{R}^{n^2 \times 1}$ , 并合并为矩阵  $\tilde{G} \in \mathbb{R}^{n^2 \times n_v}$ . 则式 (16) 转换为向量求解问题:

$$\begin{aligned} \min_w \quad & \left\| \tilde{s} - \tilde{G} w \right\|_2^2 + \beta w^T H w \\ \text{s.t.} \quad & w \geq 0, \sum_v w_v = 1, \end{aligned} \tag{17}$$

展开范数, 去掉无关项, 得到

$$\begin{aligned} \min_w \quad & w^T (\tilde{G}^T \tilde{G} + \beta H) w - 2w^T \tilde{G}^T \tilde{s} \\ \text{s.t.} \quad & w \geq 0, \sum_v w_v = 1. \end{aligned} \tag{18}$$

由于以上问题不能转化为闭式解, 因此本文使用增广 Lagrange 方法<sup>[33]</sup> 对其进行优化. 引入变量  $\tilde{w} = w$ , 记  $\tilde{G}^T \tilde{G} + \beta H$  为  $A$ , 记  $\tilde{G}^T \tilde{s}$  为  $b$ , 则上述问题可转化为

$$\begin{aligned} \min_w \quad & w^T A \tilde{w} - 2w^T b \\ \text{s.t.} \quad & w \geq 0, \sum_v w_v = 1, \tilde{w} = w. \end{aligned} \tag{19}$$

根据增广 Lagrange 方法, 求解问题 (19) 等同于优化以下问题:

$$\begin{aligned} \min_{w, \tilde{w}} \quad & w^T A \tilde{w} - 2w^T b + \frac{\mu}{2} \left\| w - \tilde{w} + \frac{\Lambda}{\mu} \right\|_2^2 \\ \text{s.t.} \quad & w \geq 0, \sum_v w_v = 1, \end{aligned} \tag{20}$$

其中  $\mu$  和  $\Lambda$  为增广 Lagrange 参数. 算法 1 介绍了问题 (20) 的求解方法. 在迭代中, 随着  $\mu$  的不断增长,  $\|w - \tilde{w} + \frac{\Lambda}{\mu}\|_2^2$  逐渐接近 0, 最终  $\tilde{w}$  等于  $w$ , 因此问题 (20) 的最优解即为问题 (19) 的解. 对目标函数 (14) 的整体求解过程如算法 2 所示.

**算法 1** 问题 (20) 的求解算法

- 
- 1: Set  $1 < \rho < 2$ , initial  $\mu$  and  $\Lambda$ ;
  - 2: **repeat**
  - 3:   Update  $\tilde{w}$  with  $\tilde{w} = w - \frac{1}{\mu}(A^T w + \Lambda)$ ;
  - 4:   Update  $w$  by solving  $\min_{w \geq 0, \sum_v w_v = 1} \|w - \tilde{w} + \frac{1}{\mu}\Lambda + \frac{Av-2b}{\mu}\|_2^2$  with an efficient optimization method [32];
  - 5:   Update  $\mu$  by  $\mu = \rho\mu$ ;
  - 6:   Update  $\Lambda$  by  $\Lambda = \Lambda + \mu(w - \tilde{w})$ ;
  - 7: **until** Converge.
- 

**算法 2** 目标函数 (14) 的求解算法**Require:** Graphs  $\{G_v\}_{v=1}^n$ , parameter  $c, \beta$  and  $\lambda$ ;**Ensure:** Optimal Graph  $S$ ;

- 
- 1: Initialize  $S$  and  $w$ ;
  - 2: **repeat**
  - 3:   Update  $F$  with (13);
  - 4:   Update  $S$  by solving problem (15);
  - 5:   Update  $w$  by solving problem (20);
  - 6: **until** Converge.
- 

## 5 基于运动方向和类中心位置的类合并方法

上文提到, 由于群组数目难以预测, 因此将  $c$  设为较大值, 聚类得到的类别数目大于实际群组数目. 本节提出一种类合并方法, 根据类内特征点的运动方向以及类中心位置坐标, 对关联程度较高的类别进行合并, 得到最终群组.

由于影像块内部所有特征点可视为统一整体, 对影像块进行多视角聚类后, 影像块的类别即为对应特征点的类别. 对于类别  $\text{cluster}_1$  与  $\text{cluster}_2$ , 如果两者类中心距离较近且内部特征点运动方向一致, 则可认为两者关联度较高. 因此, 参照式 (4) 与 (6), 计算  $\text{cluster}_1$  与  $\text{cluster}_2$  的空间距离及运动方向分布相似度  $G_l(\text{cluster}_1, \text{cluster}_2)$  与  $G_t(\text{cluster}_1, \text{cluster}_2)$ . 若所得相似度均大于阈值  $\text{th}_2$ , 则认为  $\text{cluster}_1$  与  $\text{cluster}_2$  关联度高, 应属于同一群组, 对其进行合并. 通过迭代地对群组进行合并, 直至剩余类别不满足合并条件, 即得到最终群组. 为避免合并顺序对最终结果造成的影响, 在每次迭代中, 只挑选空间距离相似度最大的一组进行合并. 通过对类别关联度高的影像块进行合并, 本方法能够自动确定最终群组数目.

## 6 实验

本文实验主要分为两部分. 首先, 在人群数据集上对群组划分结果进行检验, 并讨论实验结果. 然后, 在多视角数据集上展示了基于图多样性的多视角聚类方法的效果.

**评价标准.** 在群组划分与多视角聚类实验中, 采用聚类实验中常用的分类准确度 (accuracy, ACC) 与 F-值 (F-score) [13] 作为算法评价标准. ACC 与 F-score 越大, 则算法性能越好.

**参数选择.** 在实验中, 参数  $N_{\text{patch}}, k, c, \beta$  分别被设置为 200, 15, 15,  $2^6$ . 参数  $\lambda$  被初始化为 1, 在迭代求解中, 如果  $L_S$  的秩大于  $n - c$ , 则增大  $\lambda$ , 反之减小  $\lambda$ . 阈值  $\text{th}_1, \text{th}_2$  均被设为 0.7. 由于在多视角聚类后进行了类合并操作, 因此  $\text{th}_1, k$  与  $c$  对最终实验结果影响不大. 为确定  $\text{th}_2$  的值, 本文进行了参数分析, 随机在 CUHK 数据集中选取 100 个影像序列, 将前 30 帧影像作为训练数据, 用于选



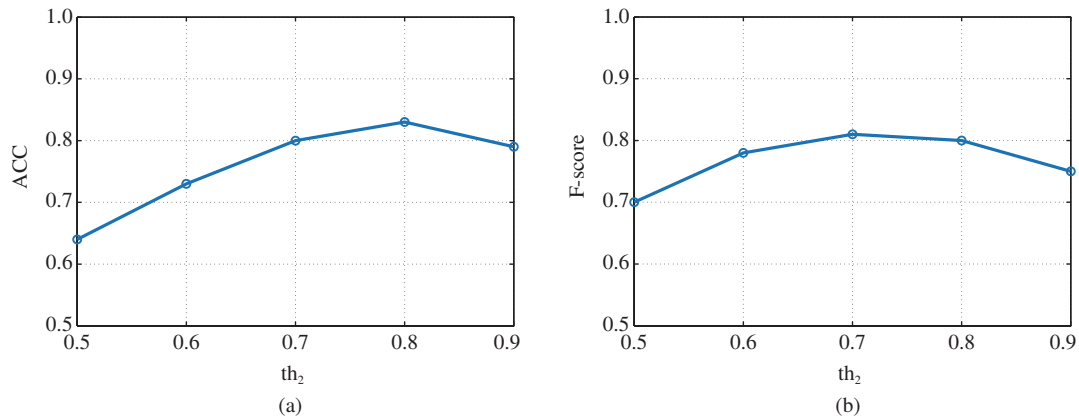
图 2 (网络版彩图) 在训练集上, 参数  $th_2$  对群组检测结果的影响Figure 2 (Color online) The influence of  $th_2$  on the group detection (a) ACC and (b) F-score with the training data

表 1 不同算法在 CUHK Crowd 数据集上的实验结果

Table 1 Experimental results on CUHK Crowd dataset<sup>a)</sup>

	CF	CT	MCC	CDC	MPF	MGBA	V-inter	V-dist	V-direc	V-trans
ACC	0.70	0.75	0.68	0.67	0.80	<b>0.85</b>	0.74	0.75	0.73	0.76
F-score	0.67	0.74	0.67	0.67	0.79	<b>0.83</b>	0.74	0.73	0.72	0.75

a) Best results are in bold face.

择参数. 而其余所有影像帧被用作测试集. 图 2 中展示了  $th_2$  的值对群组检测结果的影响. 可看出, 当  $th_2$  在 0.7 到 0.9 的范围内时, 算法效果较好. 因此, 本文将  $th_2$  的值设置为 0.7. 而参数  $\beta$  对聚类结果的影响, 将在多视角聚类实验部分进行讨论.

### 6.1 群组划分实验

为检验本文所提的 MGBA 算法的效果, 本小节在 CUHK Crowd 数据集<sup>[6]</sup>上进行实验, 并将代表最先进水平的 5 种群组划分方法作为对比算法.

**数据集.** CUHK Crowd 数据集包括 474 个影像序列, 其中 300 个序列具有人工标注. 被标注的序列给出了每个特征点的群组标签. 本文选用已标注的 300 个影像序列进行实验, 在划分的测试集上进行群组划分, 并取 ACC 和 F-score 的平均值作为实验结果. 由于本文方法的研究对象是影像块, 因此每个特征点的群组标签即为对应影像块的标签.

**对比算法.** 实验中采用的对比算法为 coherent filtering (CF)<sup>[5]</sup>, collective transition (CT)<sup>[6]</sup>, measuring crowd collectiveness (MCC)<sup>[8]</sup>, coherent density clustering (CDC)<sup>[7]</sup>, 以及 multiview-based parameter free (MPF)<sup>[13]</sup>. 为保证实验的公平性, 每种方法使用其最优参数.

**实验结果.** 各算法的群组划分结果如表 1 所示. 可看出, MGBA 算法取得了最高的 ACC 与 F-score. CF 将一段时间内保持近邻关系且运动方向一致的点划分为同一群组, 即使临近特征点关系发生变化, 也会被分入相同群组. CT 利用特征点轨迹的变化规律, 对群组内的特征点进行筛选, 但难以处理特征点的轨迹抖动问题. MCC 与 CDC 将每一帧影像进行单独处理, 利用特征点在当前帧中的相互关系进行群组划分, 缺少对时序线索的利用. 此外, 以上方法都将单个特征点视为个体, 存在研究对象过于局部化的问题. MPF 建立上下文相似图与运动相似图, 对特征点进行多视角聚类, 利用近邻点的运动方向分布弥补特征点局部化的问题, 但同样未使用时序线索, 且未考虑视角间的重合度, 易造成

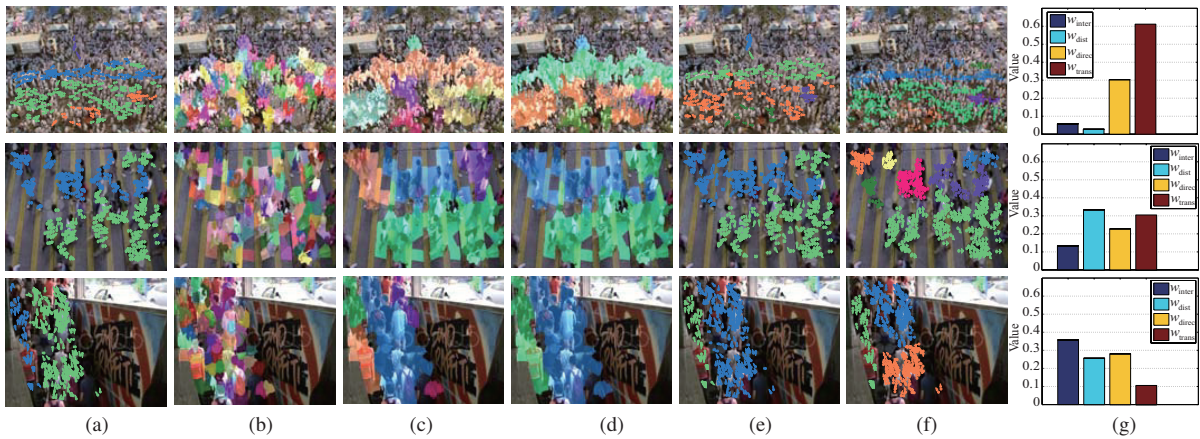


图 3 CUHK Crowd 数据集上的群组划分结果

Figure 3 Representation results on CUHK Crowd dataset. (a) Ground truth; (b) image patches; (c) multiview clustering results; (d) classes after merging; (e) groups detected by the proposed method; (f) groups detected by CF; (g) weight distribution of different views

特征冗余. 本文所提的 MGBA 方法对影像块进行研究, 对单个特征点的轨迹抖动不敏感, 减少了研究对象局部化带来的影响. 同时, 本方法从交互关系、空间距离、运动方向分布、运动规律方面充分利用了影像信息, 并在进行多视角聚类时引入了多样性正则项, 因此缓解了时序信息利用不充分、特征冗余等问题, 效果优于对比算法.

为证明整合多视角信息的必要性, 表 1 中给出了分别使用交互关系 (V-inter)、空间距离 (V-dist)、运动方向分布 (V-direc) 及运动规律 (V-trans) 相似图进行单视角群组划分时, 群组划分算法的效果. 与利用单个视角相比, 多视角群组划分 (MGBA) 的效果更好. 这说明多视角方法可以有机结合不同视角的特征, 挖掘其中的互补信息. 图 3 给出了在不同场景中, 各视角的权重分布. 可看出, 在第 1 行的场景中, 运动规律视角对最终结果的作用较大, 而在第 3 行场景中, 交互关系视角起到主导作用. 由此可见, 对于不同场景, 各视角起到的作用不同, 因此对多视角信息的综合利用十分有必要. 此外, 图 3 中展示了部分实验结果. 从第 3 列可看出, 进行多视角聚类后, 影像块被分为多个类别, 类别数目的不准确估计, 造成了类别过多的现象. 而在进行类合并操作后, 关联度高的类别被分入相同群组. 相比于对比算法 CF, 本文方法得到的最终结果与人工标注更为一致.

## 6.2 多视角聚类实验

本小节在 4 个多视角数据集上对基于多样性正则项的多视角聚类方法 MCD 进行检验, 并采用 5 种代表先进水平的多视角聚类方法作为对比算法.

**数据集.** 本实验使用的数据集包括 MSRC-v1<sup>[34]</sup>, Digit<sup>[35]</sup>, Caltech101-7 及 Caltech101-20<sup>[36]</sup>. 数据集的详细信息如下.

- MSRC-v1. 该数据集包括来自 8 个类别的 240 帧影像数据. 参照文献 [23], 本文使用其中 7 类进行实验, 包括 210 帧影像数据. 并提取 5 种特征 (CMT, HOG, LBP, GIST 和 CENT) 作为聚类依据.
- Digit. 该数据集由 2000 帧手写字形影像数据组成, 可被分为 10 类. 在实验时, 选用 6 种特征 (FOU, FAC, KAI, PIX, MOR 和 ZER) 构建相似图.
- Caltech101-7. 该数据集包括来自 7 个类别的 1474 帧物体影像数据, 实验中选用 CMT, HOG, LBP, SIFT, GIST 和 CENT 等 6 种特征进行聚类.

表 2 不同算法在多视角数据集上的实验效果

Table 2 Performance of different methods on multiview clustering<sup>a)</sup>

	ACC					F-score				
	Co-reg	RMSC	MMSC	SMC	MCD	Co-reg	RMSC	MMSC	SMC	MCD
MSRC-v1	0.70	0.67	0.71	0.70	<b>0.75</b>	0.59	0.59	0.61	0.60	<b>0.73</b>
Digits	0.79	0.77	0.84	0.88	<b>0.90</b>	0.72	0.69	0.79	0.86	<b>0.88</b>
Caltech101-7	0.43	0.59	0.68	0.68	<b>0.81</b>	0.45	0.56	0.69	0.64	<b>0.76</b>
Caltech101-20	0.48	0.51	0.51	0.60	<b>0.64</b>	0.39	0.46	0.41	0.42	<b>0.51</b>

a) Best results are in bold face.

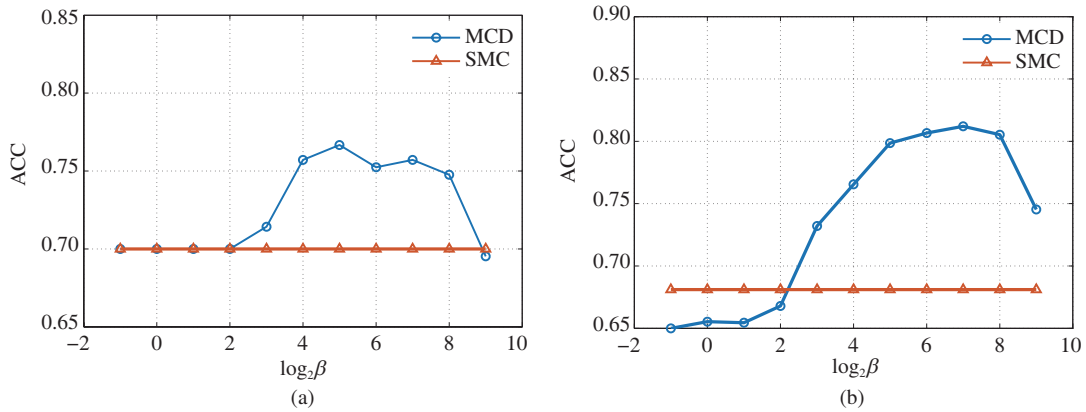


图 4 (网络版彩图) 在不同数据集上, 参数  $\beta$  对多视角聚类 ACC 值的影响

Figure 4 (Color online) Clustering accuracy with different  $\beta$ . (a) MSRC-v1; (b) Caltech101-7

• Caltech101-20. 该数据集包括 2386 帧物体影像数据, 来自 20 类. 实验中使用的特征与 Caltech101-7 中的特征相同.

**对比算法.** 实验中参与比较的算法包括 co-regularized spectral clustering (Co-reg)<sup>[16]</sup>, robust multiview spectral clustering (RMSC)<sup>[19]</sup>, multi-modal spectral clustering (MMSC)<sup>[17]</sup>, 以及 self-weighted multiview clustering (SMC)<sup>[13]</sup>. 为避免初始值对结果的影响, 将利用  $k$  均值聚类作为后处理步骤的 Coreg, RMSC 与 MMSC 重复运行 100 次, 取平均 ACC 与 F-score 作为最终结果.

**实验结果.** 在不同数据集上的多视角聚类实验结果如表 2 所示. 本文所提的 MCD 取得了最优结果. Co-reg 忽略了权重的自动学习, 对每个视角分配同样权重, 因此实验结果较差. RMSC 效果优于 Co-reg, 但当某一视角起主导作用而其他视角作用较低时, 该方法难以发挥作用. MMSC 利用各视角相似图的线性组合学习最优相似图, 因此最优相似图的搜索空间受到限制. SMC 在相似图线性组合的邻域空间内搜索最优相似图, 并对相似图权重进行自主学习, 但由于未考虑相似图间的重合性, 易造成互补信息缺失. MCD 在聚类过程中引入了多样性正则项, 在权重学习过程中强调了相似图的多样性, 有效避免了上述的一系列问题, 提高了多视角聚类的准确度. 另外, 由于 MCD 学习到的相似图可直接作为类别知识矩阵, 无需后处理, 因此只运行一次即可得到最终结果.

为研究参数  $\beta$  对实验结果的影响, 本文在  $\{2^{-1}, 2^0, \dots, 2^9\}$  范围内调整  $\beta$ , 得到在 MSRC-v1 和 Caltech101-7 上的 ACC 与 F-score 曲线, 并将 SMC 作为基准进行比较, 如图 4 和 5. 可看出, 当  $\beta$  增长时, MCD 的 ACC 值先增长后下降. 造成这种现象的原因是, 当  $\beta$  过小时, 多样性正则项难以发挥作用, 而当  $\beta$  过大时, 最优相似图的拟合误差增大. 而当  $\beta$  在  $\{2^2, \dots, 2^6\}$  范围内时, MCD 的效果均

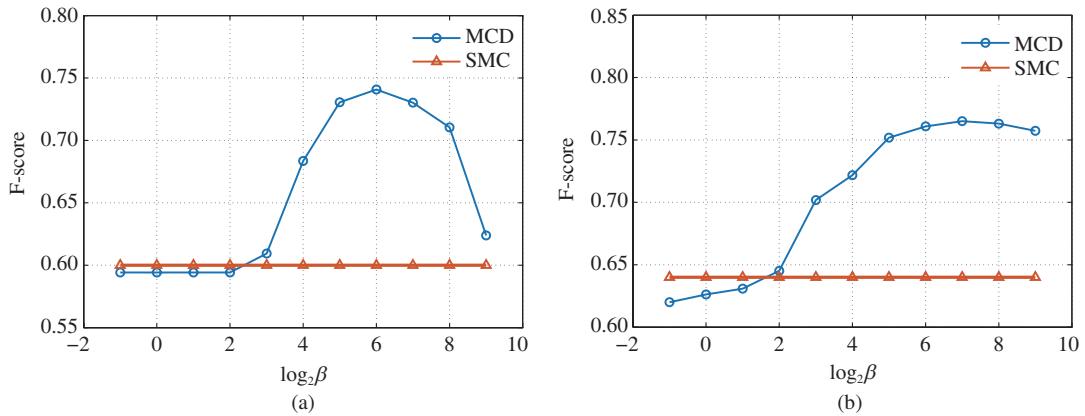


图 5 (网络版彩图) 在不同数据集上, 参数  $\beta$  对多视角聚类 F-score 值的影响  
 Figure 5 (Color online) F-score with different  $\beta$ . (a) MSRC-v1; (b) Caltech101-7

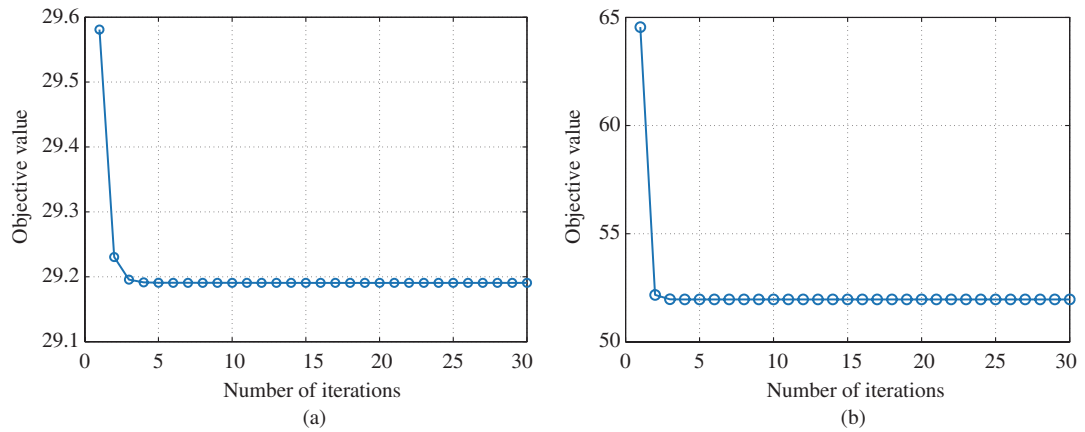


图 6 (网络版彩图) 目标函数在不同数据集上的收敛性曲线  
 Figure 6 (Color online) The convergence curves on different datasets. (a) MSRC-v1; (b) Caltech101-7

优于 SMC. 此外, 图 6 展示了 MCD 在 MSRC-v1 和 Caltech101-7 上的目标函数曲线. 对每个参数的优化, 都使式 (14) 中的目标函数值下降, 最终得到局部最优解. 从图 6 中可看出, MCD 在 5 次迭代内即达到收敛, 证明了求解算法的有效性.

## 7 结论

本文提出了一种基于多视角聚类的群组行为分析方法, 对人群影像序列中的群组进行划分, 缓解了当前人群行为研究中的研究对象局部化和特征利用不充分问题. 该方法首先将人群影像数据分割为影像块, 根据块内特征点的运动模式, 从交互关系、空间距离、运动方向分布及运动规律等 4 个方面对影像块构建相似图. 然后, 提出一种基于多样性正则项的多视角聚类方法, 利用所得相似图对影像块进行聚类, 并设计了相应的求解算法. 最后, 以运动方向分布和空间距离相似度为依据, 合并关联度高的类别, 得到最终群组. 实际数据集上的一系列实验证明了所提群组划分方法与多视角聚类方法的优越性, 以及求解算法的有效性与收敛性.

## 参考文献

- 1 Wang Q, Fang J W, Yuan Y. Multi-cue based tracking. *Neurocomputing*, 2014, 131: 227–236
- 2 Zhang Y Y, Zhou D S, Chen S Q, et al. Single-image crowd counting via multi-column convolutional neural network. In: *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, Las Vegas, 2016. 589–597
- 3 Wang W Y, Lin W Y, Chen Y Z, et al. Finding coherent motions and semantic regions in crowd scenes: a diffusion and clustering approach. In: *Proceedings of European Conference on Computer Vision*, Zurich, 2014. 756–771
- 4 Yuan Y, Fang J W, Wang Q. Online anomaly detection in crowd scenes via structure analysis. *IEEE Trans Cybern*, 2015, 45: 548–561
- 5 Zhou B L, Tang X O, Wang X G. Coherent filtering: detecting coherent motions from crowd clutters. In: *Proceedings of European Conference on Computer Vision*, Florence, 2012. 857–871
- 6 Shao J, Loy C C, Wang X G. Scene-independent group profiling in crowd. In: *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, Columbus, 2014. 2227–2234
- 7 Wu Y P, Ye Y D, Zhao C Y. Coherent motion detection with collective density clustering. In: *Proceedings of ACM Conference on Multimedia Conference*, Brisbane, 2015. 361–370
- 8 Zhou B L, Tang X O, Zhang H, et al. Measuring crowd collectiveness. *IEEE Trans Pattern Anal Mach Intel*, 2014, 36: 1586–1599
- 9 Zhou B L, Wang X G, Tang X O. Random field topic model for semantic region analysis in crowded scenes from tracklets. In: *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, Colorado Springs, 2011. 3441–3448
- 10 Wang Q, Chen M L, Li X L. Quantifying and detecting collective motion by manifold learning. In: *Proceedings of AAAI Conference on Artificial Intelligence*, San Francisco, 2017. 4292–4298
- 11 Chen M L, Wang Q, Li X L. Anchor-based group detection in crowd scenes. In: *Proceedings of International Conference on Acoustics, Speech and Signal Processing*, New Orleans, 2017. 1378–1382
- 12 Ali S, Shah M. A Lagrangian particle dynamics approach for crowd flow segmentation and stability analysis. In: *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, Minneapolis, 2007
- 13 Li X L, Chen M L, Nie F P, et al. A multiview-based parameter free framework for group detection. In: *Proceedings of AAAI Conference on Artificial Intelligence*, San Francisco, 2017. 4147–4153
- 14 Chen M L, Wang Q, Li X L. Patch-based topic model for group detection. *Sci China Inf Sci*, 2017, 60: 113101
- 15 Sharma R, Guha T. A trajectory clustering approach to crowd flow segmentation in videos. In: *Proceedings of IEEE International Conference on Image Processing*, Phoenix, 2016. 1200–1204
- 16 Kumar A, Rai P, Daum H. Co-regularized multi-view spectral clustering. In: *Proceedings of Advances in Neural Information Processing Systems*, Granada, 2011. 1413–1421
- 17 Cai X, Nie F P, Huang H, et al. Heterogeneous image feature integration via multi-modal spectral clustering. In: *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, Colorado Springs, 2011. 1977–1984
- 18 Li Y Q, Nie F P, Huang H, et al. Large-scale multi-view spectral clustering via bipartite graph. In: *Proceedings of AAAI Conference on Artificial Intelligence*, Texas, 2015. 2750–2756
- 19 Xia R K, Pan Y, Du L, et al. Robust multi-view spectral clustering via low-rank and sparse decomposition. In: *Proceedings of AAAI Conference on Artificial Intelligence*, Quebec, 2014. 2149–2155
- 20 Liu X W, Zhou S H, Wang Y Q, et al. Optimal neighborhood kernel clustering with multiple kernels. In: *Proceedings of AAAI Conference on Artificial Intelligence*, San Francisco, 2017. 2262–2272
- 21 Liu X W, Dou Y, Yin J P, et al. Multiple kernel kmeans clustering with matrix-induced regularization. In: *Proceedings of AAAI Conference on Artificial Intelligence*, Phoenix, 2016. 1888–1894
- 22 Cao X C, Zhang C Q, Fu H Z, et al. Diversity-induced multi-view subspace clustering. In: *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, Boston, 2015. 586–594
- 23 Nie F P, Li J, Li X L. Self-weighted multiview clustering with multiple graphs. In: *Proceedings of International Joint Conference on Artificial Intelligence*, Melbourne, 2017. 2564–2570
- 24 Achanta R, Shaji A, Smith K, et al. SLIC superpixels compared to state-of-the-art superpixel methods. *IEEE Trans*

- Pattern Anal Mach Intel, 2012, 34: 2274–2282
- 25 Senst T, Eiselein V, Sikora T. Robust local optical flow for feature tracking. *IEEE Trans Circ Syst Video Technol*, 2012, 22: 1377–1387
- 26 Geng Q C, Zhou Z, Cao X C. Survey of recent progress in semantic image segmentation with CNNs. *Sci China Inf Sci*, 2018, 61: 051101
- 27 Wang J H, Liu B, Xu K. Semantic segmentation of high-resolution images. *Sci China Inf Sci*, 2017, 60: 123101
- 28 Ballerini M, Cabibbo N, Candelier R, et al. From the cover: interaction ruling animal collective behavior depends on topological rather than metric distance: evidence from a field study. *Proc Natl Acad Sci USA*, 2008, 105: 1232–1237
- 29 Kullback S. On the convergence of discrimination information (corresp.). *IEEE Trans Inf Theory*, 1968, 14: 765–766
- 30 Shumway R H, Stoffer D S. An approach to time series smoothing and forecasting using the EM algorithm. *J Time Ser Anal*, 1982, 3: 253–264
- 31 Mohar B, Alavi Y, Chartrand G, et al. The Laplacian spectrum of graphs. *Graph Theory Combin Appl*, 1991, 18: 871–898
- 32 Nie F P, Wang X Q, Jordan M, et al. The constrained laplacian rank algorithm for graph-based clustering. In: *Proceedings of AAAI Conference on Artificial Intelligence*, Phoenix, 2016. 1969–1976
- 33 Nie F P, Wang H, Huang H, et al. Joint Schatten  $p$ -norm and  $\ell_p$ -norm robust matrix completion for missing value recovery. *Knowl Inf Syst*, 2015, 42: 525–544
- 34 Winn J, Jojic N. LOCUS: learning object classes with unsupervised segmentation. In: *Proceedings of IEEE International Conference on Computer Vision*, Beijing, 2005. 756–763
- 35 Breukelen M, Duin R, Tax D, et al. Handwritten digit recognition by combined classifiers. *Kybernetika*, 1998, 34: 381–386
- 36 Li F F, Fergus R, Perona P. Learning generative visual models from few training examples: an incremental Bayesian approach tested on 101 object categories. *Comput Vision Image Und*, 2007, 106: 59–70

## Multiview-based group behavior analysis in optical image sequence

Xuelong LI<sup>1\*</sup>, Mulin CHEN<sup>2</sup> & Qi WANG<sup>2,3</sup>

1. *Xi'an Institute of Optics and Precision Mechanics, Chinese Academy of Sciences, Xi'an 710119, China;*

2. *School of Computer Science and Center for OPTical IMagery Analysis and Learning (OPTIMAL), Northwestern Polytechnical University, Xi'an 710072, China;*

3. *Unmanned System Research Institute (USRI), Northwestern Polytechnical University, Xi'an 710072, China*

\* Corresponding author. E-mail: xuelong.li@opt.ac.cn

**Abstract** Group behavior analysis is a hot topic in intelligent video surveillance, and has attracted a surge of interest in the field of artificial intelligence. Groups are the basic components of a crowd system, and provide a high-level representation of the crowd phenomenon. By investigating the motion dynamics within each image patch, this paper proposes a multiview-based group behavior analysis method that is able to divide the paths into different groups. The main contributions are threefold: (1) the correlation between image paths is captured from four views (interaction, distance, motion direction, and motion transition), (2) a multiview clustering method with diversity regularization is proposed to perceive the complementary information within the multiview data and alleviate the influence of redundant features, and (3) a cluster merging strategy is designed to combine the highly correlated clusters and determine the final groups automatically. Experimental results on several benchmark datasets validate the good performance of the proposed method.

**Keywords** crowd analysis, group behavior analysis, clustering, graph clustering, multiview clustering



**Xuelong LI** was born in 1976. He received the Ph.D. degree in electronic engineering and information science from the University of Science and Technology of China, Hefei, China, in 2002. He is currently a full professor at the Xi'an Institute of Optics and Precision Mechanics, Chinese Academy of Sciences, Xi'an, China.



**Mulin CHEN** was born in 1991. He received the B.E. degree in software engineering and the M.E. degree in computer application technology from Northwestern Polytechnical University, Xi'an, China, in 2014 and 2016, respectively. He is currently pursuing the Ph.D. degree in the School of Computer Science, and is with the Center for Optical Imagery Analysis and Learning, Northwestern Polytechnical University, Xi'an, China. His current research interests include computer vision and machine learning.



**Qi WANG** was born in 1982. He received the B.E. degree in automation and the Ph.D. degree in pattern recognition and intelligent systems from the University of Science and Technology of China, Hefei, China, in 2005 and 2010, respectively. He is currently a professor in the School of Computer Science, and is with the Unmanned System Research Institute and the Center for OPTical Imagery Analysis and Learning (OPTIMAL), Northwestern Polytechnical University, Xi'an, China. His research interests include computer vision and pattern recognition.