

# 人群异常状态检测的图分析方法

朱海龙<sup>1</sup> 刘鹏<sup>1</sup> 刘家锋<sup>1</sup> 唐降龙<sup>1</sup>

**摘要** 提出一种图分析方法用于动态人群场景异常状态检测. 使用自适应 Mean shift 算法对场景速度场进行非参数概率密度估计聚类, 聚类结果构成以聚类中心为顶点、各聚类中心之间距离为边权重的无向图. 通过分析图顶点的空间分布及边权重矩阵动态系统的预测值与观测值之间的离散程度, 对动态场景中的异常事件进行检测和定位. 使用多个典型动态场景视频数据库进行对比实验, 结果表明图分析方法适应性强、可有效监控动态人群场景中的异常状态.

**关键词** 非参数密度估计, 自适应 Mean shift, 图分析, 人群异常检测, 动态场景

**引用格式** 朱海龙, 刘鹏, 刘家锋, 唐降龙. 人群异常状态检测的图分析方法. 自动化学报, 2012, 38(5): 742–750

**DOI** 10.3724/SP.J.1004.2012.00742

## A Graph Analysis Method for Abnormal Crowd State Detection

ZHU Hai-Long<sup>1</sup> LIU Peng<sup>1</sup> LIU Jia-Feng<sup>1</sup> TANG Xiang-Long<sup>1</sup>

**Abstract** An abnormality detection method for a dynamic crowd scene is proposed based on graph analysis. After the non-parametric clustering in velocity space via an adaptive mean shift algorithm, we get the clustering results containing some cluster centers and Euclidean distances between them, and they can form a graph whose vertexes are the cluster centers and edge weights are the distances. Through analyzing the vertexes' distribution in feature space and the state transform of a dynamic system made by the sequence of the edge weight matrix, we can detect and locate the abnormal events in the scenario. To testify the method's effectiveness, we conducted experiments on several well-known datasets and obtained good performance in both abnormal events detection and location. The results show that the graph analysis method has strong adaptability and can efficiently detect the abnormal states in dynamic crowd scene.

**Key words** Non-parametric probability density estimation, adaptive mean shift algorithm, graph analysis, crowd abnormality detection, dynamic scene

**Citation** Zhu Hai-Long, Liu Peng, Liu Jia-Feng, Tang Xiang-Long. A graph analysis method for abnormal crowd state detection. *Acta Automatica Sinica*, 2012, 38(5): 742–750

人群状态检测是智能视频监控和智能环境的主要内容, 也是实现自动人群管理的基本要求. 动态人群场景的异常状态常伴随着各种危险, 使得异常检测成为计算机视觉分析领域的一个重要组成部分. 人们通过常识所能认知的非正常情况称为异常事件, 是发生概率很小的随机事件, 如道路拥堵、人群紧急疏散、机动车辆驶入人行道等. 发生了异常事件并持续一定时间的人群场景处于异常状态, 反之处于正常状态, 都是人群行为的结果. 场景中的各类群体事件最终都以运动特征表现出来, 所以人群异常状态检测实质上是对感兴趣目标的运动特征加以分类, 根据分类结果判断当前场景状态. 目前有两类方法,

1) 提取场景底层特征并使用统计学分析方法检测异常<sup>[1–3]</sup>; 2) 建立场景模型, 用学习算法估计模型参数后用该模型进行异常检测<sup>[4–7]</sup>.

使用底层图像特征, 如颜色、角点、轮廓、纹理等进行人群事件检测, 算法简单、易于实现. 如 Albiol 等<sup>[1]</sup> 提取图像角点特征计算速度场, 以速度期望为标准检测目标的高速运动事件. 当人群中有较严重遮挡时提取的角点特征并不准确, 影响了检测效果. Srivastava 等<sup>[2]</sup> 通过实验证明指定区域内通过的人群数量与累积的前景像素数量有关, 选择纹理特征描述场景拥挤程度的比例因子, 根据比例因子的大小确定场景是否危险. 该方法适用于光照条件不变的室内环境以保证纹理特征的稳定性, 如地铁站、廊道等. Chan 等<sup>[3]</sup> 将视频看作是对不同纹理抽样的结果, 把纹理方法发展为类似混合高斯的混合动态纹理系统, 使用期望最大化算法获取线性系统参数. 在时间上根据视频纹理的联合分布将视频段聚类, 在空间上将具有相同纹理特征的区域聚类达到运动分割的目的. 该方法对场景做粗略分类, 适用于室外大场景的粗估计, 如在什么时间段内道路上人/车是稠密、顺畅还是稀疏等. 此类方法<sup>[1–3]</sup>

收稿日期 2011-08-23 录用日期 2012-01-05  
Manuscript received August 23, 2011; accepted January 5, 2012  
国家自然科学基金 (60702032, 61171184), 黑龙江省自然科学基金 (F201021) 资助  
Supported by National Natural Science Foundation of China (60702032, 61171184) and Natural Science Foundation of Heilongjiang Province (F201021)  
本文责任编辑 贾云得  
Recommended by Associate Editor JIA Yun-De  
1. 哈尔滨工业大学计算机科学与技术学院 哈尔滨 150001  
1. School of Computer Science and Technology, Harbin Institute of Technology, Harbin 150001

直接使用图像低层特征, 在光照发生变化或有较严重遮挡时会直接影响特征提取的准确性, 对场景变化的适应性较差.

另一类方法的通用做法是构建一个场景模型, 使用学习的方法对模型参数进行估计. Wu 等<sup>[4]</sup> 计算光流并进行质点平流传送, 对质点轨迹进行聚类得到人群流轨迹, 然后利用混乱不变性 (最大李雅普诺夫指数和相关元) 提取轨迹的混乱动态特征并建立概率模型. 学习正样本得到模型参数并采用最大似然估计准则来识别正常和非正常视频, 使用基于混合高斯的算法以确定异常区域位置和尺寸. 该方法借鉴了流体力学思想, 需要较高密度人群建立人群流轨迹, 不适用于中、低密度人群. Helbing 等<sup>[5]</sup> 工作的最具特色之处在于将牛顿力学定律引入人群动力学分析中, 称之为“社会力” (Social force). Mehran 等<sup>[6]</sup> 对社会力模型进行了拓展, 将其应用到相对稀疏的人群行为建模上. 为避免对每个局部目标跟踪而产生大量计算, 使用了质点平流传送的整体分析方法. 在平流传送的基础上建立力学模型, 对社会力流构建语料库后用隐含狄利克雷分配 (Latent Dirichlet allocation, LDA) 训练并发现人群行为标题, 再利用词袋法 (Word bag, WB) 进行模式分析以检测人群事件. LDA 训练和词袋模型的 EM 近似算法导致该方法的计算量偏大. Wang 等<sup>[7]</sup> 利用高频和时空特征在全局和局部两种尺度下检测人群异常: 使用 LDA 对全局正常模态建模, 使用多重隐马尔科夫模型对局部正常模态建模. 文献 [7] 认为高频表达了人群场景的动态特性, 但仅凭高频特征无法判断人群行为是否异常, 还需要使用词袋和 LDA 方法识别全局异常、使用有竞争机制的多重隐马尔科夫模型识别局部异常, 导致处理过程繁琐、效

率较低. 因此, 建立复杂模型并使用训练方法进行参数估计的方法虽然取得了明显的进展<sup>[4-7]</sup>, 但建立训练集本身就是一件费时费力的工作, 当变换场景时需要重建训练集并重新学习模型参数. 再者, 训练样本和训练方法选择得是否恰当直接影响模型参数估计精度, 容易发生训练不足或者过拟合现象.

无论采用低层视觉特征还是基于学习的场景模型异常检测方法, 都欠缺对环境的适应和泛化能力. 一种特征或一组参数并不能对所有的视频中人群异常状态进行有效的检测, 因此, 研究非参数的、具有良好适应性和泛化能力的人群异常状态检测方法是十分必要的.

本文提出一种用于人群场景异常状态检测的图分析方法, 核心内容分为运动特征聚类 and 图分析两部分. 其中第一部分是参数估计的运动特征聚类: 计算感兴趣区域的速度场构成特征空间, 提出自适应 Mean shift 值域带宽计算方法, 对速度空间进行非参数聚类. 第二部分是对速度场的聚类结果进行图分析, 将图论与聚类中心分布和动态系统模型结合起来分析动态场景、检测人群异常状态. 非参数估计的运动特征聚类方法不需要训练样本, 也不需要分割前景目标; 图顶点的空间分布能够反映人群的速度分布, 融合了人群流量信息的边权重矩阵序列构成对动态系统的描述, 可反映特征空间的状态变化趋势. 两者结合用于检测、定位动态人群场景的异常状态, 使该方法适应性较强, 具有较好的泛化能力, 处理流程如图 1 所示.

### 1 特征空间与 Mean shift 聚类算法

有许多特征可以用来描述视频场景, 如颜色、几

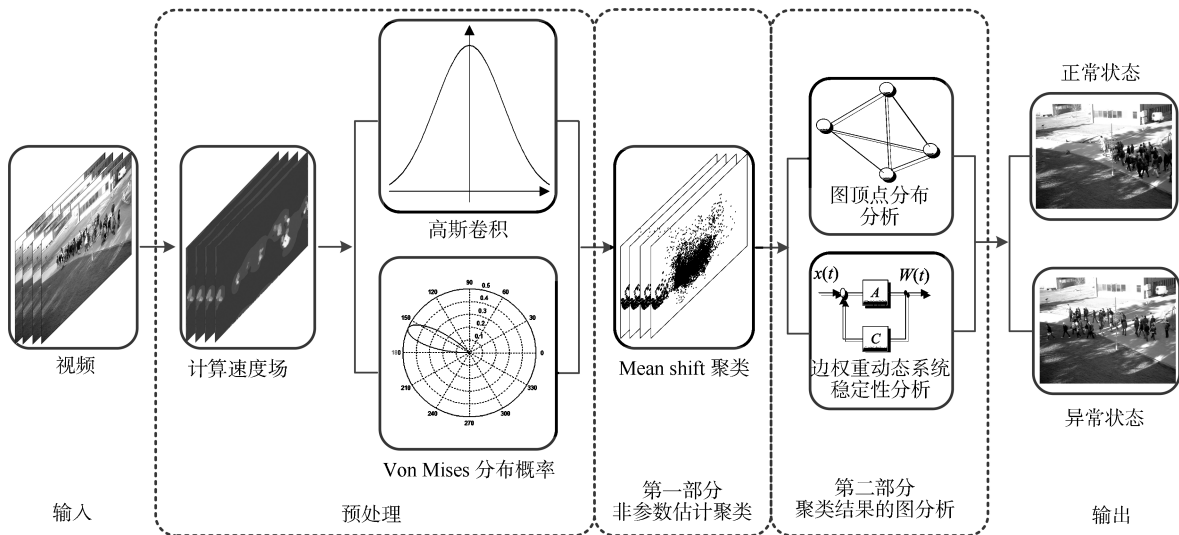


图 1 处理流程

Fig. 1 Processing flow

何形状、纹理、频率等. 考虑到人群场景的动态特性, 本文使用速度场建立特征空间. 设有图像序列  $\{I'\}^{\tau'}$ ,  $\tau'$  为图像数量. 将连续 2 帧图像  $I'_t, I'_{t+1}$  叠加得到新图像序列  $\{I\}^{\tau}$  以降低噪声, 其中  $\tau = \lceil \tau'/2 \rceil$ ,  $t < \tau'$  且  $t = 1, 3, 5, \dots$ . 在新序列  $\{I\}^{\tau}$  的连续两帧间计算光流<sup>[8]</sup> 得到速度场序列  $\{(u, v)\}^{\tau-1}$ , 其中  $u, v$  分别为水平和垂直方向的速度分量. 为抑制速度场噪声和简化运动方向聚类: 1) 使用高斯卷积平滑速度幅值; 2) 使用 Von Mises 分布<sup>[9]</sup> 将运动方向由  $[0, 2\pi]$  弧度量化为 8 个方向值  $0, \dots, 7$ , 分别对应弧度区间  $[\frac{k\pi}{4}, \frac{(k+1)\pi}{4}]$ ,  $k = 0, \dots, 7$ .

经过上述预处理的速度场构成特征空间  $\mathcal{O} = \{(u, v)\}$ , 可在此空间上进行人群状态聚类. 然而客观环境的复杂性和群体中个体主观意识的差异, 使人群在场景中表现出复杂的行为模式. 图 2 显示了一个视频场景, 其中图 2(a) 为视频, 图 2(b) 和图 2(c) 分别为当前速度场在水平方向和垂直方向的分量, 图 2(d) 是对这些速度样本的聚类情况. 图 2(d) 表明: 1) 速度场是描述视频场景的一种有效特征; 2) 由于人群行为的多样性导致速度空间的聚类结果具有任意形状, 如果直接使用参数估计模型, 将迫使人们预先对模型做出各种约束性假设以便估计模型参数. 不但做出准确的假设较为困难, 处理结果还包含较严重的人为干预的痕迹. 所以, 使用参数估计的方法描述视频场景并不恰当, 而选择对数据分布不做任何假设的非参数方法对速度场进行聚类是一种有效手段.

本文采用 Mean shift<sup>[10]</sup> 算法实现这一目标, 并根据人群运动的特点在值域计算和空域选择两方面对原有方法进行了改进. 为便于叙述, 下面简要复述 Mean shift 算法的主要内容. 给定特征空间的  $m$  个样本点  $\mathbf{x}_i = (u_i, v_i)$ ,  $i = 1, \dots, m$ , 当前点  $\mathbf{x}$  的核密度估计为

$$\hat{f}(\mathbf{x}) = \frac{m_{K,h}}{m} \sum_{i=1}^m K\left(\frac{d^2(\mathbf{x}, \mathbf{x}_i)}{h}\right) \quad (1)$$

其中,  $m_{K,h} > 0$  为归一化系数,  $K(z) > 0$  ( $z \geq 0$ ) 为剖面函数 (本文使用高斯函数),  $d(\mathbf{x}, \mathbf{x}_i)$  是定义在特征空间的距离度量,  $h$  为值域宽度. 令  $Q(z) = -K'(z)$  并求式 (1) 的梯度, 可计算出指向最大梯度方向的均值漂移向量为

$$\mathbf{m}_h(\mathbf{x}) = m'_{K,h} \frac{\nabla \hat{f}(\mathbf{x})}{\hat{f}(\mathbf{x})} = \frac{\sum_{i=1}^m \mathbf{x}_i Q(\frac{d^2(\mathbf{x}, \mathbf{x}_i)}{h})}{\sum_{i=1}^m Q(\frac{d^2(\mathbf{x}, \mathbf{x}_i)}{h})} - \mathbf{x} \quad (2)$$

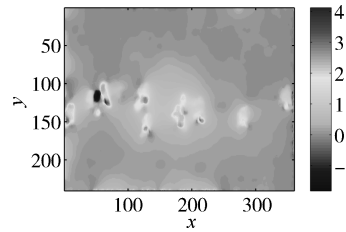
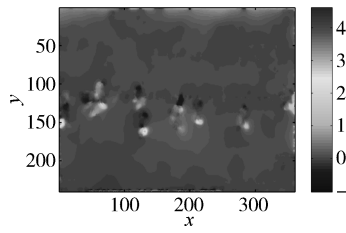
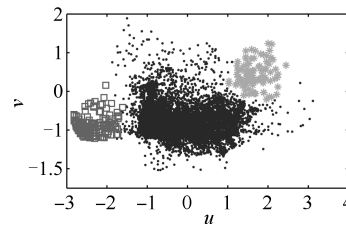
以  $\mathbf{m}_h(\mathbf{x})$  为增量, 利用梯度上升法由当前点  $\mathbf{x}_j$  漂移到下一点  $\mathbf{x}_{j+1}$ :

$$\mathbf{x}_{j+1} = \mathbf{x}_j + \mathbf{m}_h(\mathbf{x}_j), \quad j = 1, \dots, m-1 \quad (3)$$



(a) 视频

(a) Video

(b) 速度场  $u$  分量(b)  $u$  of the velocity field(c) 速度场  $v$  分量(c)  $v$  of the velocity field

(d) 对速度场聚类

(d) Clustering the field

图 2 速度场聚类的多样性

Fig. 2 Multi-appearance of velocity field clustering

随机选择特征空间中的一个样本点作为聚类种子, 迭代执行式 (2) 和式 (3) 直至满足  $\|\mathbf{m}_h(\mathbf{x}_j)\| < \varepsilon$  ( $\varepsilon > 0$ ) 时停止, 到达聚类中心位置, 得到该类所属范围内概率密度的最大值  $\hat{f}_{\max}$ , 完成一次聚类. 然后另选其他样本点再次作为聚类种子, 重复上述迭代过程可完成全部聚类工作.

## 2 非参数人群状态聚类

均值漂移聚类算法需要确定值域带宽  $h$  和聚类空间  $\mathcal{O}$ , 其中值域带宽  $h$  决定着概率密度函数  $f(s)$  的平滑解析度. 原有方法对所有聚类取同一个值域带宽  $H = \text{diag}\{h, h, \dots, h\}$ , 而图 2(d) 表明各个

聚类的样本值域和样本数量差别都很大, 在整个特征空间上采用相同的全局最优带宽  $H$  显然是不恰当的. 另一方面, 人群场景的背景区域不包含感兴趣目标的运动信息, Mean shift 算法的有效聚类空间应为去除背景后的  $\mathcal{O}' \subset \mathcal{O}$ . 针对以上两点, 本文分别对 Mean shift 算法的值域计算和空域选择进行了改进, 以提高聚类准确性和聚类效率.

### 2.1 自适应 Mean shift 值域带宽的计算方法

确定值域带宽的关键在于既要准确估计概率密度函数, 又不能大幅增加计算成本. 考虑到值域带宽与各聚类的离散程度有关, 本文使用各聚类速度分布的方差作为该类值域宽度的参数. 设人群场景速度空间有  $c$  个聚类中心, 第  $k \in \{1, \dots, c\}$  个聚类的值域带宽  $h_k$  是该类中样本方差  $D_k$  的函数  $h_k = \lambda D_k$ , 得到新的值域对角阵:

$$H = \text{diag}\{h_1, \dots, h_c\} = \text{diag}\{\lambda D_1, \dots, \lambda D_c\} \quad (4)$$

其中,  $\lambda \in [0, 1]$  为调整系数, 只是在运算初始时设置一次. 这样, 可在聚类过程中通过式 (5) 自适应计算各聚类的值域带宽:

$$h_{k,t} = (1 - \alpha)h_{k,t-1} + \alpha(\Delta D_k), \quad t = 1, \dots, \tau - 1 \quad (5)$$

其中,  $\alpha \in (0, 0.1)$  为学习率, 也是在运算初始时设置一次.  $\Delta D_k = D_k^{t+1} - D_k^t$  为在时间上连续的两个速度空间中第  $k$  个聚类的方差增量.

### 2.2 有效 Mean shift 聚类空间的计算方法

场景背景区域的理论速度为零, 但由于存在各种干扰, 很多背景区域像素点的速度幅值大于零. 为去除这些样本点, 设置全局速度阈值  $T_V > 0$ , 小于  $T_V$  的速度样本不参与聚类过程, 得到有效样本空间:

$$\mathcal{O}' = \{(u, v) | u^2 + v^2 > T_V\} \subset \mathcal{O} \quad (6)$$

另外, 传统 Mean shift 算法随机选择种子点. 若该种子点处于样本稀疏区域, 在均值漂移过程中将沿着较为平缓的坡度 (梯度相对较小方向) 向聚类中心逼近, 迭代次数较多. 本文先估计种子点值域半径区域内样本点的密度, 若小于预设值则另选其他点, 直至满足要求再开始聚类.

## 3 聚类结果的图分析方法

对速度场聚类后, 本文提出图分析方法对聚类结果进行分析以检测人群的异常状态.

### 3.1 建立聚类中心无向图

如图 3 所示, 将各聚类中心作为顶点连接起来, 以欧氏距离为边权重形成带权无向图  $G = (V, E,$

$W)$ ,  $V$ 、 $E$ 、 $W$  分别为顶点集合、边集合和边权重实对称矩阵. 当场景中人群运动处于稳定状态时各个聚类中心的位置在小范围内波动, 见图 3(a) 和 3(b); 当人群运动发生较大变化时聚类中心也随之产生较大变动, 见图 3(c). 所以, 图顶点的空间分布反映了动态场景中速度的分布情况. 实际上, 图 3 中每个特征点除了速度特征  $(u, v)$  外还包括该样本点在视频中的空间坐标分量  $(x, y)$ , 将二者合成一个扩展的特征向量  $\mathbf{x} = (x, y, u, v)^T$ , 其中  $u = \frac{dx}{dt}$ ,  $v = \frac{dy}{dt}$ . 前文在聚类时只是显式用到了速度特征  $(u, v)$ , 位置特征  $(x, y)$  则隐含在每个样本之中. 所以, 图顶点所代表的聚类也隐式包含了以该类速度运动的人群位置信息.

边权重矩阵元素为顶点间距离的 2-范数, 故  $W$  为正实对称阵  $\{w_{ij}\}^+$ ,  $w_{ij} = w_{ji} \geq 0$ ,  $w_{ii} = 0$ ,  $i, j \in \{1, \dots, N\}$ ,  $N = \text{length}(V) \in \mathbf{Z}^+$  为  $G$  的顶点数量. 各顶点表示各类运动的平均速度, 顶点间距离为各类运动间的速度差. 所以, 边权重矩阵描述了动态场景中各聚类速度的离散程度. 为了全面考查人群异常状态, 仅使用速度和位置特征是不够的, 还应考虑各类人群的密度、流量及具有相似速度的人群规模等信息. 设第  $i$  个聚类的样本数量为  $\beta_i$ , 则第  $i$  类的人群数量可用该聚类的样本数量  $\beta_i$  表示. 人群密度为人群数量与所占面积的商, 在场景固定的前提下, 为简化计算, 各类人群所占面积统一使用场景面积常量  $S$ , 有  $\rho_i = \beta_i/S = \beta_i S'$ . 所以, 使用密度信息将第  $i, j$  两顶点间边权重扩展为

$$w'_{ij} = (\rho_i + \rho_j)w_{ij} = S'(\beta_i + \beta_j)w_{ij} \quad (7)$$

得到扩展边权重矩阵  $W = \{w'_{ij}\}^+$ . 上式中  $w_{ij}$  为两类运动速度差, 而人群流量是人群密度与速度的积, 所以式 (7) 具有流量差公式  $\rho(v^1 - v^2)$  的形式, 表示视频中人群流量的增量. 因此, 序列  $W$  的动态变化刻画了场景中人群密度、人群流量的变化情况.

### 3.2 图顶点的空间分布

将特征空间等分为如图 3(d) 所示的  $R$  个局部区域, 区域的大小取决于人群处于稳定状态时顶点在速度空间中波动的最大范围. 在  $t$  时刻, 当前图  $G_t$  的顶点  $o_i \in V^t$  在第  $r \in \{1, \dots, R\}$  个区域  $l_r$  出现的概率为  $p_{l_r}(o_i, t)$ . 那么, 对于整个图序列  $\{G\}^{\tau-1}$  而言, 顶点  $o_i$  在  $l_r$  中发生的联合概率为

$$p_{l_r}(o_i) = \prod_{t=1}^{\tau-1} p_{l_r}(o_i, t) \quad (8)$$

通过上式可知各顶点在特征空间中的分布情况, 间接可知场景中感兴趣目标的速度大小和方向. 例如, 场景中水平方向较大速度分布在图 3(d) 中右侧

阴影区域, 根据顶点在该区域发生的联合概率可检测场景中是否有快速物体通过.

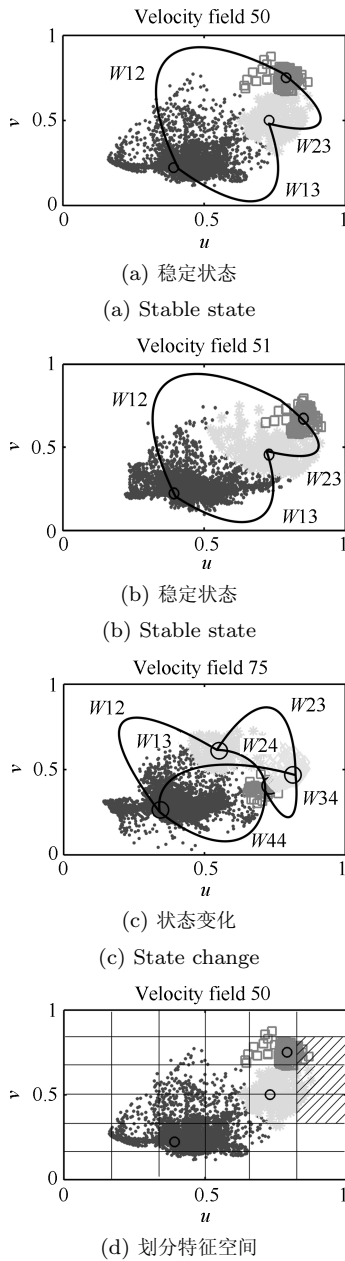


图 3 以聚类中心为顶点的带权无向图  
Fig. 3 Weighted undirected graph with clustering center vertices

### 3.3 边权重矩阵动态模型

边权重包含了丰富的有关图的结构与状态信息, 将顶点所在类的样本规模作为权重系数引入边权重的计算, 可更加合理地描述场景中人群速度、密度和流量的变化情况. 设  $W$  的秩为  $n$ , 图序列  $\{G\}^{\tau-1}$  对应的权重矩阵序列  $\{W\}^{\tau-1}$  在时域上的变化可用动态系统的输出表示:

$$\begin{cases} y(t+1) = Ay(t) \\ W(t) = Cy(t), \quad t = 1, \dots, \tau - 1 \end{cases} \quad (9)$$

其中,  $y(t)$  为隐状态变量,  $W$  为观测值,  $A \in \mathbf{R}^{n \times n}$  为隐状态转换矩阵,  $C \in \mathbf{R}^{N \times n}$  为显状态转换矩阵且  $\text{rank}(C) = n, N \geq n$ . 可利用该模型的预测值与观测值之间的变化判断场景状态. 已知  $W$  为实对称阵, 必定存在正交阵  $P$  使其对角化:

$$P^{-1}WP = \Lambda P^{-1} \Rightarrow W = P\Lambda P^{-1} \quad (10)$$

该动态模型的输出具有马尔科夫性, 递推得:

$$\begin{cases} \hat{C} = P \\ y(t) = \Lambda P^{-1}P \\ \hat{A} = y(t+1)y(t)^{-1} \end{cases} \quad (11)$$

式 (11) 实际上是对  $W$  进行奇异值分解, 将式 (11) 代入式 (9) 得到边权重阵  $W$  的预测值:

$$\hat{W}(t+1) = \hat{C}y(t+1) \quad (12)$$

连续运用上式可以预测若干时间后的系统状态, 比较预测状态趋势与实际观测值之间的偏离程度可判断人群状态的变化是否平稳. 预测值  $\hat{W}$  与观测值  $W$  可用状态转换矩阵表示为两个自回归模型:  $\hat{W} = (\hat{A}, \hat{C})$  和  $W = (A, C)$ , 其谱系数分别为  $c_{\hat{W}}(n)$  和  $c_W(n), n \in \mathbf{Z}$ . 使用 Martin 距离<sup>[11]</sup> 计算  $W$  与  $\hat{W}$  的偏离度:

$$D(\hat{W}, W) = \sqrt{\sum_{n=0}^{\infty} n |c_{\hat{W}}(n) + c_W(n)|^2} \quad (13)$$

如果人群场景处于平稳状态, 则距离  $D(\hat{W}, W)$  接近于零; 若状态发生变化  $D(\hat{W}, W)$  会迅速增加, 大于某个阈值时判断为异常:

$$\begin{cases} \text{正常, } D(\hat{W}, W) < \varepsilon \\ \text{异常, } D(\hat{W}, W) \geq \varepsilon, \quad \varepsilon > 0 \end{cases} \quad (14)$$

当场景状态发生较大变化时,  $G$  的顶点位置及数量也会随之发生变化, 因此  $W$  的秩是个变量. 为保持  $W$  形式的一致性, 设场景中速度最大聚类数量为  $N_{\max}$ , 当前  $W$  的秩为  $N_{\text{cur}}$ , 有  $N_{\text{cur}} \leq N_{\max}$ . 实际的边权重阵  $W$  具有如下形式:

$$W = \begin{bmatrix} W_{\text{cur}} & 0 \\ 0 & 0 \end{bmatrix} \in \mathbf{R}^{N_{\max} \times N_{\max}} \quad (15)$$

其中,  $W_{\text{cur}} \in \mathbf{R}^{N_{\text{cur}} \times N_{\text{cur}}}$  为当前权重矩阵.  $N_{\text{cur}}$  的变化说明图顶点数量的增加或减少, 反映出场景中人群运动状态发生变化, 如由正常状态变为拥堵、紧急疏散、高速物体闯入等异常状态, 或上述异常状态变为正常状态.

综上所述, 通过图  $G$  的顶点分布和边权重矩阵动态模型都可检测人群场景的异常状态, 两种方法各有所侧重. 前者可定位异常, 但不能预测将来状态; 后者可预测场景状态趋势, 却不能确定发生异常状态的场景位置. 将二者结合起来既可检测和定位异常、又可对未来场景状态趋势做出预测, 处理过程如下:

1) 给定  $\lambda, \alpha, \varepsilon, T_V, S', N_{\max}$ , 初始化  $H$ .

2) 从视频中提取位置信息, 计算光流并平滑速度特征形成特征向量  $\mathbf{x} = (x, y, u, v)^T$ , 去除背景区域的速度样本得到  $\mathcal{O}'$ .

3) 对特征向量中的分量  $(u, v)$  进行速度特征 Mean shift 聚类, 更新  $H$ ; 各聚类中心和扩展的边权重矩阵形成无向图  $G$ .

4) 检测异常状态、定位异常发生区域:

if  $((D(\hat{W}, W) \geq \varepsilon) \text{ and } p_{l_r}(o_i) > 0)$  or  $(p_{l_r}(o_i) > 0 \text{ and } T_p > 3)$  (其中  $l_r$  为异常区域,  $T_p$  为动态系统状态改变后持续时长)

发生异常;

预测若干帧后的场景状态;

根据异常聚类元素特征向量的坐标分量  $(x, y)$  在视频中定位异常区域;

end if

5) 重复 2)~4), 直至结束.

## 4 实验分析

为验证方法的有效性, 选择了三个视频库进行实验分析, 分别是 PETS'09 Benchmark Data<sup>[12]</sup>, UCSD Dataset<sup>[13]</sup> 和 UMN Dataset<sup>[14]</sup>. 这三个数据集包含了多种典型异常情况, 使用广泛、具有较强的代表性. 实验内容包括: 1) 自适应值域和选择空域对 Mean shift 聚类结果的影响; 2) 使用图分析方法检测异常状态; 3) 效果对比.

1) 自适应值域和选择空域对 Mean Shift 聚类结果的影响

本文认为, 各聚类的值域带宽与自身数据的离散程度相关, 不应使用同一个带宽常量. 再者, 空域应仅包括有效样本数据并且应在一定密度以上的区域内选择起始种子点. 为此, 使用 PETS'09 S2L1 view1 视频数据 (3 通道、576 像素  $\times$  768 像素, 帧数  $\tau' = 200$ ) 进行对比实验, 该视频表现了正常行走的人群突然加速奔跑的场景, 如图 4(a) 所示. 将连续、不重叠的 2 帧图像相加得到新的图像序列  $\{I\}^\tau$ , 新序列的图像数量为原视频的一半, 即  $\tau = 100$ . 在序列  $\{I\}^\tau$  的连续 2 帧间计算光流场得到  $\tau - 1 = 99$  个速度场  $\mathcal{O}$ , 以  $3 \times 3$  固定窗口的高斯核对  $\mathcal{O}$  卷积平滑后, 再将速度方向量化为 8 个方向值. 令  $T_V = 0.1 \times \text{mean}(\mathcal{O})$  为速度阈值去除背景区域得到有

效样本空间  $\mathcal{O}'$ . 取经验值  $\lambda = 0.75$ ,  $\alpha = 0.005$ ,  $h_0 = 0.75$ , 选择概率密度大于 0.01 的样本为种子点对  $\mathcal{O}'$  进行 Mean Shift 聚类并更新  $h_k$ , 结果如图 4 所示. 图 4(b) 为值域带宽  $h = 0.80$ 、空域为  $\mathcal{O}$ 、随机选择种子点的欠聚类结果, 图 4(c) 为值域带宽  $h = 0.55$ 、空域为  $\mathcal{O}$ 、随机选择种子点的过聚类结果, 图 4(d) 为本文方法.

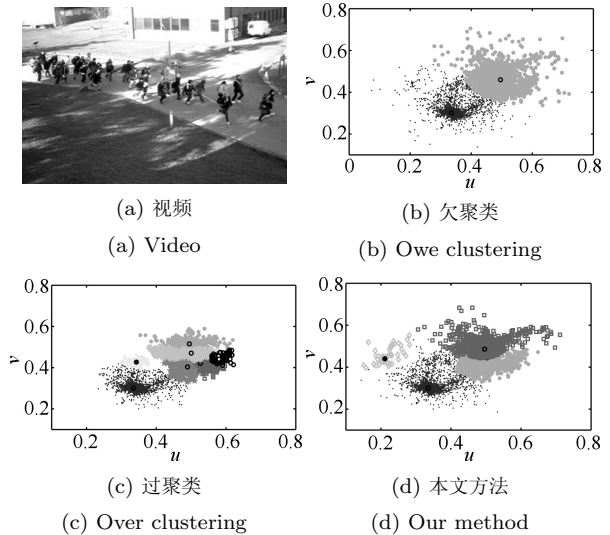


图 4 聚类对比

Fig. 4 Comparison of clustering results

从图 4 可以看出, 传统 Mean shift 方法将相同值域带宽用于所有速度特征聚类, 对某些类别而言恰好适合, 而对其他类别来说却并不适合: 过大导致欠聚类, 过小导致过聚类.

表 1 为使用不同空域和不同的初始种子点选取方法进行聚类时的计算量比较, 说明使用有效样本空间和有条件选择起始漂移种子会在一定程度上改善计算负担.

表 1 均值漂移过程平均迭代次数

Table 1 Average iteration times of the mean shift algorithm

| 所用方法   | 全体样本、随机种子点 | 全体样本、选择种子点 | 有效样本、选择种子点 |
|--------|------------|------------|------------|
| 平均迭代次数 | 14.53      | 13.26      | 10.95      |

为考查不同值域带宽初值对聚类的影响, 设置调整系数  $\lambda$  分别为 1.0, 0.75, 0.55 进行自适应带宽聚类, 结果如图 5 所示. 学习率给定的情况下, 在聚类初始阶段, 当  $\lambda$  的值较大时聚类中心数量要小于  $\lambda$  取较小值时的数量, 经过近 10 次更新后聚类数量稳定于 3. 说明  $\lambda$  对聚类结果的影响限于聚类的初期, 随着值域带宽的不断更新自适应 Mean shift 方法可准确聚类. 因相机抖动等干扰而偶然出现的聚

类数量不准确的情况,可用状态持续性检测  $T_p > 3$  帧滤掉.

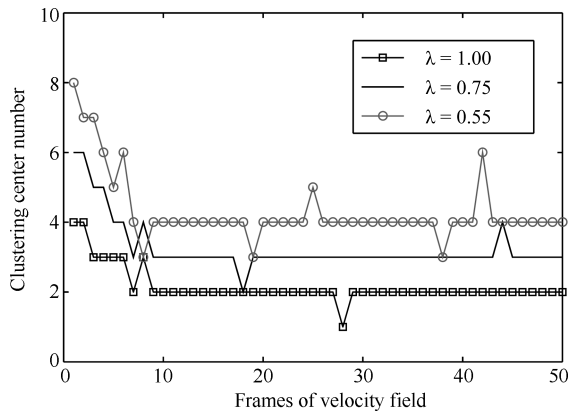
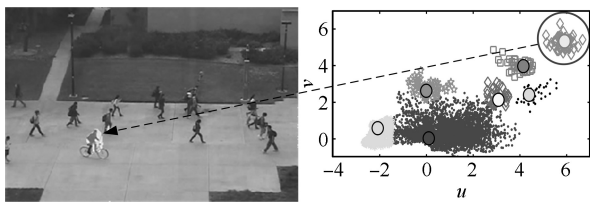


图 5 不同值域带宽初值对聚类结果的影响

Fig. 5 Clustering results with different initialization of the bandwidth

## 2) 使用图分析方法检测异常状态

图顶点分布描述了各聚类所在的特征空间区域,聚类样本隐含的坐标特征可对异常区域定位.下面使用 UCSD Test/Test02 数据(单通道、240 像素  $\times$  360 像素,帧数  $\tau' = 180$ ) 检验异常定位的有效性.该视频为一骑自行车者以较快速度驶入人行道,视为异物闯入.从图 6 可以看出,较高速度的异物闯入平稳运动的人群后,在速度空间的特定区域产生聚类,可利用该区域事件的联合概率判断是否发生异常,并根据该聚类成员隐含的位置信息在视频空间定位异常区域.



(a) 异常定位 (b) 图顶点的空间分布

(a) Abnormality location (b) Distribution of graph vertexes

图 6 利用图顶点分布定位异常

Fig. 6 Locating the abnormal state via the distribution of graph vertexes

边权重矩阵动态系统的预测值和观测值间的 Martin 距离反映了场景中人群流量差的变化,仍使用 UCSD Test/Test02 进行异常检测实验,归一化结果如图 7 所示.可以看出,初始阶段(A 段)的误差较大,在没有异常发生的平稳状态下(B 段) Martin 距离维持在较低水平;当场景中有异物闯入时 Martin 距离会迅速增大(C 段),当预测值以一定的学习率适应异常场景后,预测值与观测值的距离会

逐渐下降(D 段)并维持在较低水平,进入异常状态的平稳阶段(E 段).所以,依据预测值与观测值之间距离的变化用来检测人群场景中的状态变化.问题在于,若开始检测时场景中就存在异常(如开始检测时就有自行车或汽车驶过),则预测值与观测值之间的距离也会在初始化阶段过去后维持在较低水平的平稳状态,当异常状态变为正常状态时(如自行车或汽车驶离场景),状态改变也会导致距离增大,对应图 7 中 C 段.表明距离的增大只是表示场景中的状态发生改变,并不能确切地指出是进入了异常状态还是由异常状态转换为正常状态.因此,需要与图顶点分布结合使用来共同检测异常:当图顶点在异常区域持续出现而  $D(\hat{W}, W)$  保持不变(即稳定于当前状态),或者当  $D(\hat{W}, W)$  持续高于阈值并且图顶点在异常区域出现时,可判断进入异常状态.本实验用状态持续时间  $T_p > 3$  帧为标准判断状态持续时间,以消除相机抖动或聚类错误的影响.

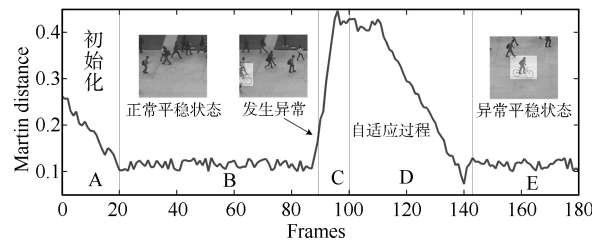


图 7 场景状态变化

Fig. 7 State changing in a dynamic scene

为了解聚类时使用不同的自适应学习率  $\alpha$  和调整系数  $\lambda$  值对检测结果的影响,从 UCSD 库中挑选 10 段视频并手工统计出 37 处异物闯入异常事件,使用本文方法自动检测.检测结果的错误率见表 2 所示.当  $\lambda$  和  $\alpha$  取值较大或较小时错误率较高, $\lambda$  对检测结果的影响小于  $\alpha$ .当  $\alpha$  取值较小时  $h$  达到稳定值的时间较长导致聚类不准确,而当  $\alpha$  较大时  $h$  随场景变化较大导致聚类不稳定.

表 2 不同参数取值下检测错误率(%)

Table 2 The detection errors with different parameters (%)

| $\alpha$ | $\lambda$ |       |       |
|----------|-----------|-------|-------|
|          | 0.55      | 0.75  | 1.00  |
| 0.001    | 16.22     | 13.51 | 18.92 |
| 0.005    | 10.81     | 8.11  | 13.81 |
| 0.050    | 18.92     | 16.22 | 21.63 |

## 3) 检测异常状态效果对比

从 UCSD、PETS'09 和 UMN 数据库中挑选制作了一些异常场景为实验素材,将本文方法与社会力法<sup>[5]</sup>、光流法<sup>[8]</sup>及 Adam 的方法<sup>[15]</sup>进行了对比

实验. 如图 7 所示, 图 7(a)、7(b)、7(c) 分别为高速异物闯入、人群突然加速奔跑、紧急疏散三个异常场景示例, 包括视频和检测结果的 ROC 曲线. 从实验结果看, 光流法对噪声和相机抖动敏感, 当场景光线较暗时表现不稳定; 社会力法在处理远景小目标时因质点的平流传送精度不够造成误差较大; Adam 法在单摄像机视频源时检测结果不理想. 而本文方法综合了场景中速度场的空间分布和人群密度、人群流量信息及状态持续性, 并将图顶点空间分布和边权重矩阵动态系统结合起来对场景状态进行双重判断, 因此在各场景中检测和定位异常时表现稳定、适应性较强.

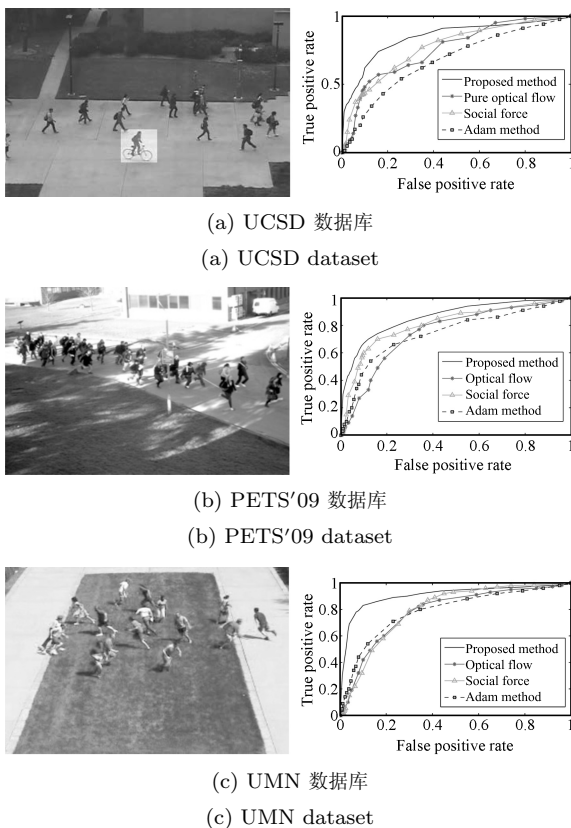


图 8 检测异常状态的 ROC 曲线

Fig. 8 ROCs of four abnormality detection methods in three video datasets

## 5 结论

从人群场景的动态特性而非低层的几何图形和光学测度出发, 提出了一种基于图分析的人群事件检测方法. 为适应动态场景的速度特征空间聚类, 改进了传统 Mean shift 算法的值域宽度和有效空域的计算方法, 可兼顾聚类准确度和聚类效率两个方面. 引入图分析方法对聚类结果进行分析, 将图顶点在特征空间的分布与边权重矩阵动态模型结合起来, 建立了检测和定位异常的图分析方法模型. 该模型

综合考虑了人群场景运动速度、人群密度、人群流量及各聚类人群规模等信息, 可对场景状态做出较准确的界定. 在多个典型视频库上的对比实验表明该方法表现稳定、适应性较强.

## References

- Albiol A, Silla M J, Albiol A, Mossi J M. Video analysis using corner motion statistics. In: Proceedings of 11th IEEE International Workshop on Performance Evaluation of Tracking and Surveillance. Miami, USA: IEEE, 2009. 31–37
- Srivastava S, Ng K K, Delp E J. Crowd flow estimation using multiple visual features for scenes with changing crowd densities. In: Proceedings of 8th IEEE International Conference on Advanced Video and Signal-Based Surveillance. Klagenfurt, Austria: IEEE, 2011. 60–65
- Chan A B, Vasconcelos N. Modeling, clustering, and segmenting video with mixtures of dynamic textures. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2008, **30**(5): 909–926
- Wu S D, Moore B E, Shah M. Chaotic invariants of Lagrangian particle trajectories for anomaly detection in crowded scenes. In: Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition. San Francisco, CA, USA: IEEE, 2010. 2054–2060
- Helbing D, Molnár P. Social force model for pedestrian dynamics. *Physical Review*, 1995, **51**(5): 4282–4286
- Mehran R, Oyama A, Shah M. Abnormal crowd behavior detection using social force model. In: Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Miami, Florida, USA: IEEE, 2009. 935–942
- Wang B, Ye M, Li X, Zhao F J, Ding J. Abnormal crowd behavior detection using high-frequency and spatio-temporal features. *Machine Vision and Applications* (Springer). [Online], available: <http://www.springerlink.com/content/vr38484834416g85/>, January 11, 2012
- Horn B K P, Schunck B G. Determining optical flow. *Artificial Intelligence*, 1981, **17**(1-3): 185–203
- Khatri C G, Mardia K V. The von Mises-Fisher matrix distribution in orientation statistics. *Journal of the Royal Statistical Society: Series B (Methodological)*, 1977, **39**(1): 95–106
- Comaniciu D, Meer P. Mean shift: a robust approach toward feature space analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2002, **24**(5): 603–619
- Martin R J. A metric for ARMA processes. *IEEE Transactions on Signal Processing*, 2000, **48**(4): 1164–1170
- PETS 2009 Benchmark Data [Online], available: <http://www.cvg.rdg.ac.uk/PETS2009/a.html>, January 11, 2012
- Mahadevan V, Li W X, Vasconcelos N. UCSD dataset [Online], available: <http://www.svcl.ucsd.edu/projects/anomaly/>, January 11, 2012
- UMN dataset [Online], available: <http://mha.cs.umn.edu/Movies/Crowd-Activity-All.avi>, January 11, 2012



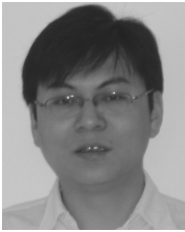
- 15 Adam A, Rivlin E, Shimshoni I, Reinitz D. Robust real-time unusual event detection using multiple fixed-location monitors. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2008, **30**(3): 555–560



**朱海龙** 哈尔滨工业大学计算机科学与技术学院博士研究生. 2008 年获哈尔滨理工大学硕士学位. 主要研究方向为模式识别, 图像处理, 视频处理. 本文通信作者. E-mail: zhl.hit@163.com

**(ZHU Hai-Long** Ph.D. candidate at the School of Computer Science and Technology, Harbin Institute of Technology,

Harbin Institute of Technology. He received his master degree from Harbin University of Science and Technology in 2008. His research interest covers pattern recognition, image processing, and video processing. Corresponding author of this paper.)



**刘 鹏** 哈尔滨工业大学计算机科学与技术学院副教授. 2007 年获哈尔滨工业大学博士学位. 主要研究方向为数字信号处理, 模式识别, 图像处理与机器视觉. E-mail: pengliu@hit.edu.cn

**(LIU Peng** Associate professor at the School of Computer Science and Technology, Harbin Institute of Technology,

Harbin Institute of Technology. He received his Ph.D. degree from Harbin Institute of Technology in 2007. His research interest covers

digital signal processing, pattern recognition, image processing, and computer vision.)



**刘家锋** 哈尔滨工业大学计算机科学与技术学院副教授. 1996 年获哈尔滨工业大学博士学位. 主要研究方向为模式识别, 机器学习, 图像处理, 图像理解与机器视觉. E-mail: jefferyliu@hit.edu.cn

**(LIU Jia-Feng** Associate professor at the School of Computer Science and Technology, Harbin Institute of Technology,

Harbin Institute of Technology. He received his Ph.D. degree from Harbin Institute of Technology in 1996. His research interest covers pattern recognition, machine learning, image processing, image understanding, and computer vision.)



**唐降龙** 哈尔滨工业大学计算机科学与技术学院教授. 1995 年获哈尔滨工业大学博士学位. 主要研究方向为光学字符识别, 生物特征识别, 图像处理及模式识别. E-mail: tangxl@hit.edu.cn

**(TANG Xiang-Long** Professor at the School of Computer Science and Technology, Harbin Institute of Technology,

Harbin Institute of Technology. He received his Ph.D. degree from Harbin Institute of Technology in 1995. His research interest covers optical character recognition (OCR), biometrics, image processing and pattern recognition.)