

# 基于四元数模型的密集人群视频特征提取

殷梦娇 俞亚萍 张之江

(上海大学通信与信息工程学院 上海 200072)

**摘要:** 密集人群场景下的视频异常事件检测是当今智能监控技术研究中的一个热点。本文针对如何合理提取面向密集人群场景视频的时空特征、以及提高密集人群异常检测的效率进行研究,结合人类视觉感知系统相关知识,分析了将视频的时间特征和空间特征相融合的四元数傅里叶变换,提出了一种新的适用于密集人群场景的特征提取方法。最后通过实验证明,本文所提出的特征能够较为全面准确地描述密集人群视频场景中的特征,并取得了良好的异常检测效果。

**关键词:** 密集人群视频;四元数傅里叶变换;特征提取

**中图分类号:** TP751 **文献标识码:** A **国家标准学科分类代码:** 510.4050

## Feature extraction of dense crowd video based on quaternion model

Yin Mengjiao Yu Yaping Zhang Zhijiang

(School of Communication and Information Engineering, Shanghai University, Shanghai 200072, China)

**Abstract:** Abnormal crowd behavior detection is a highly focused research area of the intelligent monitoring. The paper mainly aims at studying how to extract spatio-temporal characteristics of dense crowd video, and how to improve the efficiency of anomaly detection. Combining with human visual system(HVS), we proposed a novel method based on the analysis of the quaternion Fourier transform which is a fusion of spatio-temporal characteristics in order to extract features of dense crowd scene. It is proved by the experiment that the proposed method can describe the dense crowd scene from different aspects and achieve good effect of anomaly detection as well.

**Keywords:** dense crowd video; quaternion Fourier transform; feature extraction

## 1 引言

随着城市人口激增、人群活动日益频繁,以及视频监控设备价格的不断下降,大量摄像头被安装在各公共场所中作为城市安全保障的重要手段。由此产生的大量视频需要耗费大量人力物力进行监控处理,且容易发生误报漏报的现象,因此需要通过智能监控技术对视频进行监控和分析,及时捕捉动态人群相关信息,识别异常情况。

近年来,密集人群场景的分析成为智能监控研究的热点,越来越多的研究转向密集场景下的特征提取和模式区分<sup>[1]</sup>。目前对于密集人群场景的特征提取主要有3类方法:1)基于流的特征提取,适用于视野开阔且运动呈整体趋势的室外场景,Ali等人<sup>[2]</sup>提出了拉格朗日粒子动态方法用于密集人流检测与分割,Mehran<sup>[3]</sup>提出了通过社会力模型建模,对基于粒子流提取的特征进行异常检测;2)基于粒子运动轨迹的特征提取,适用于密度不高的室内场

景或视野有限的室外场景,如Bolei Zhou等人<sup>[4]</sup>提出的基于运动轨迹的特征提取算法,能够通过轨迹的路径相似度对密集人群进行模式聚类与异常检测;3)基于局部时空特征的提取方法,适用于人群密度极大、异常拥挤的场景,典型的如Kratz等人<sup>[5]</sup>提出的时空梯度模型特征提取,以该特征为基础的局部时空运动模式通过聚类,可以将相似的运动合并,相差较大的运动分开<sup>[6]</sup>,与隐马尔科夫模型相结合能够在密集场景中实现多目标的稳定预测与异常检测。

针对密集人群场景的特点,结合人类视觉感知系统,对视频特性进行分析,提出了一种新的底层特征提取方法:即采用四元数傅里叶变换来融合视频在空间域和时间域的几个不同特征,然后结合人类视觉系统感知理论将提取出的傅里叶变换频谱用于计算密集人群场景视频的显著性特征。该特征从多个维度描述了密集人群场景,能够提高密集人群场景下异常事件检测的效率。

## 2 图像中的四元数傅里叶变换

四元数的概念由 Hamilton 于 1843 年首次提出<sup>[6]</sup>,四元数也称超复数,是由 1 个实部和 3 个虚部构成,形如:

$$q = a + bi + cj + dk \quad (1)$$

式中:  $a$ 、 $b$ 、 $c$ 、 $d$  均为实数,  $i$ 、 $j$ 、 $k$  为虚数单位,并遵循以下规则:

$$i^2 = j^2 = k^2 = ijk = -1 \quad (2)$$

$$ij = k \quad jk = i \quad ki = j \quad (3)$$

$$ji = -k \quad kj = -i \quad ik = -j \quad (4)$$

四元数  $q$  的模为:

$$|q| = \sqrt{a^2 + b^2 + c^2 + d^2} \quad (5)$$

傅里叶变换是图像处理中的常用手段,被广泛用于计算机视觉的研究中。大量研究发现傅里叶相位谱携带了图像的局部信息,而振幅谱携带的是图像的表象和方位信息。本文利用四元数傅里叶变换(QFT)而非基础的傅里叶变换来获取每个图像块的振幅谱。与单个特征的简单加成不同,四元数傅里叶变换允许将图像的多个特征作为一个整体进行空间变换,即把一个图像在向量域进行变换,能同时有机表示图像中的亮度、密度、纹理、方向分布等各种信息,对于图像中的较小区域的突变更加敏感。

## 3 基于 HVS 的密集人群视频场景四元数模型

### 3.1 模型框架

文献[8]利用四元傅里叶变换的相位谱进行图像显著性检测,文献[9]结合人类视觉系统(human visual system, HVS)理论和傅里叶变换的幅度谱用于图像的显著性信息检测。然而视频场景不同于静态图像,其场景是不断变换的,包含了时间轴角度的运动信息,具有很强的空时相关性。将视频帧划分为  $M$  个小的图像块,对每个图像块综合空域特征(亮度、边缘信息等)和时域特征(运动特征)进行四元数傅里叶变换,并将变换后的幅度谱与 HVS 理论结合获取密集人群视频的显著性特征,该显著性特征可结合其他辅助特征用于密集人群场景视频的异常事件检测。

密集人群场景四元数模型的显著性特征图谱提取算法框架流程如图 1 所示。

图 1 中最终计算得出的密集人群场景的特征图谱  $s_i$  由如下表达式表示:

$$s_i = \sum_{i \neq j} \omega_{i,j} D_{i,j} \quad (6)$$

式中:  $i$ 、 $j$  分别表示每帧划分出的第  $i$ 、 $j$  个小图像块,  $D_{i,j}$  表示两图像块之间的谱差异,  $\omega_{i,j}$  表示权重。

### 3.2 四元数模型的特征构建

现有的密集人群场景视频的特征提取往往只注重运动信息的提取,而忽略空域信息。事实上,一段正常的运动视频帧在空间的轮廓和纹理等特征上也具有一定规律性,人眼对于异常的认识包括了空间结构和时域运动两个方面。

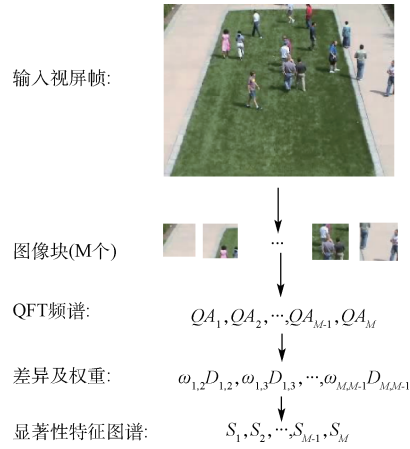


图 1 基于 HVS 理论的密集人群场景四元数模型特征提取算法框架

经过大量实验验证,本文从空域和时域选取下文所述的 4 个特征来构建四元数模型。模型中空域信息构成四元数的 3 个部分,时域运动信息构成四元数的另一个部分。

#### 1) 空域特征

对于密集人群视频场景,现有的空域特征往往需要结合复杂的物理学模型,虽然在特定模型下具有较好检测效果,但并不具有普适性。本文在构造密集人群视频时,分别提取视频帧的零阶、一阶以及二阶的边缘梯度信息作为四元数模型中的 3 个参数。通过不同阶的梯度边缘信息,可以多维度地表示视频帧的亮度、轮廓、纹理等局部信息,能较为简明准确地表示密集人群场景的空域特征。

对 RGB 图像而言,其零阶梯度就是其本身。考虑到单独的 RGB 混合了色度和亮度信息,并不是一种好的特征提取方式,且人眼对于亮度的敏感性高于色度,故直接选取原图像的亮度信息作为零阶梯度信息。假设 R, G, B 表示图像块的红、绿、蓝 3 个颜色分量,则图像块的零阶边缘信息——即亮度特征表示为:

$$I(n, m) = (R(n, m) + G(n, m) + B(n, m)) / 3 \quad (7)$$

一阶梯度信息就是图像的边缘信息,本文选择 Canny 算子求得。Canny 分割算法采用一阶偏导的有限差分来计算梯度幅值和方向,该算子功能较其他边缘检测算子更具准确性,能有效获取图像的一阶边缘轮廓信息,计算结果下文用  $E(n, m)$  表示。

二阶边缘梯度信息其实与条状的区域块相关,能有效表示一些纹理上的灰度突变信息。一方面,人类视觉系统可观测范围类似条状<sup>[10-11]</sup>;另一方面,密集人群场景中可以忽略每个个体的局部特征,而将其简单认为是条状。因此,获取二阶梯度边缘特征对于后续的检测是有意义的,其具体计算方法参见文献[12],计算结果下文用  $B(n, m)$  表示。

#### 2) 时域特征

考虑到运动信息会极大地影响人眼视觉注意力,因此

视频的运动信息强度非常重要,当人眼开始关注视频时,首先会受到场景中色彩、对比度的影响,但很快适应后,会更加关注场景内的运动物体,特别是一些与其他对象运动差异很大的部分。

光流场用于表示图片中各个能够观察到点的运动,是图像帧中三维运动速度在二维图像上的投影,十分适用于人群密集的场景视频分析以及对前景运动对象的特征检测。对图像块求得的光流场信息下文用  $M(n, m)$  表示。

将光流场的运动信息与上文得到的空域信息结合,作为四元数傅里叶变换的 4 个参数,可以有效地度量密集人群场景视频帧的特征。

考虑到四元数傅里叶变换的对称性质<sup>[13]</sup>,实部虚部具体对应何种特征不会影响最终变换的结果,故此处只确定四个参数,无需将其与实部虚部一一对应。综上,密集人群场景视频帧中的每个图像块可以表示为如下四元数表达式:

$$q(n, m) = M(n, m) + I(n, m)\mu_1 + E(n, m)\mu_2 + B(n, m)\mu_3 \quad (8)$$

式中:  $i, j, k$  为单元纯四元数。

### 3.3 四元数傅里叶变换及显著性特征图谱计算

根据式(2)~(4)所列四元数的性质。式(8)能表示为以下形式:

$$q(n, m) = f_1(n, m) + f_2(n, m)\mu_2 \quad (9)$$

$$f_1(n, m) = M(n, m) + I(n, m)\mu_1 \quad (10)$$

$$f_2(n, m) = E(n, m) + B(n, m)\mu_1 \quad (11)$$

研究表明,四元数傅里叶变换可以根据两个标准的快速傅里叶变换计算得到<sup>[13]</sup>,因此式(9)中的  $q(n, m)$  可以根据以下计算得到:

$$Q[u, v] = F_1[u, v] + F_2[u, v]\mu_2 \quad (12)$$

$$F_i[u, v] = \frac{1}{\sqrt{MN}} \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} e^{-j_1 2\pi((nm)/M + (nu)/N)} f_i(n, m) \quad (13)$$

式中:  $i \in \{1, 2\}$ ,  $[u, v]$  是图像块在频域中的相应位置,  $N$  和  $M$  分别是图像块的高和宽。

式(12)的四元数傅里叶变换结果可以写成如下的形式:

$$Q(n, m) = Ae^{j\varphi} \quad (14)$$

幅度谱  $A$  可由下式计算得到:

$$A = |Q[u, v]| \quad (15)$$

由式(15)可以得到每个图像块四元数傅里叶变换的频谱,用于表示该图像块的整体特征。

视频帧中任意两个图像块  $i, j$  的差  $D_{i,j}$  为:

$$D_{i,j} = \sqrt{\sum_m (\log(A_m^i + 1) - \log(A_m^j + 1))^2} \quad (16)$$

将上述结果与 HVS 模型中相结合<sup>[9]</sup>,计算图像块之间的差异性权重  $w_{i,j}$ ,根据式(6)能将整个视频帧根据所分图像块计算出显著性特征图谱  $S_l(a, b)$ ,其中  $l$  表示帧数,  $a,$

$b$  分别表示图像帧每列、每行所包含图像块的个数。该特征图谱描述了每一视频帧在时域和空域上的多维度特征。

## 4 实验结果与分析

本文选取了可公开获得的明尼苏达大学提出的 UMN 群体异常数据集以及网络上搜集到的广场、人行道等常见室内外监控视频进行实验。

### 4.1 密集人群视频显著性特征图谱

对数据集中截取的视频帧采用本文的方法进行实验的结果显示如图 2 所示。其中图 2(a)为原视频帧,图 2(b)为基于四元数的傅里叶变换频谱图(为了更清晰显示此处将频谱图做了对数处理),图 2(c)为根据 HVS 模型获取的显著性特征图谱。第一行和第二行为 UMN 数据集中的视频场景,第三行为搜集到的地铁口密集人群视频场景。显然,除了空间域比较显著的亮度和边缘信息,该模型也能较好地检测出运动信息。

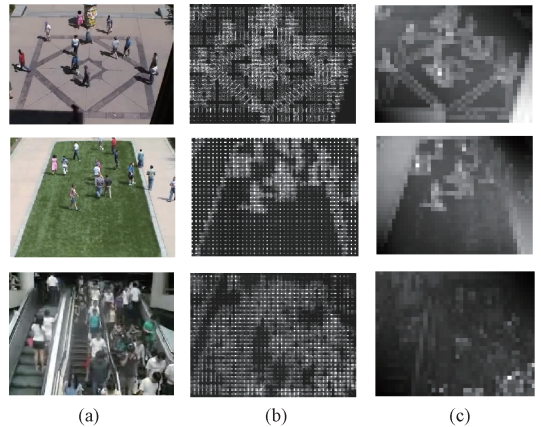


图 2 密集人群显著性特征图谱提取示例

### 4.2 异常检测应用

为检验本文特征提取方法的有效性,采用机器学习的方法进行异常检测。支持向量机(support vector machine, SVM)主要靠超平面进行划分分类,是一种被广泛采用的监督式学习方法。本文将计算所得密集人群场景特征图谱与光流方向直方图 HOOF<sup>[14]</sup>级联,作为 SVM 的特征输入。

本文验证特征异常检测效率的算法框架如图 3 所示。

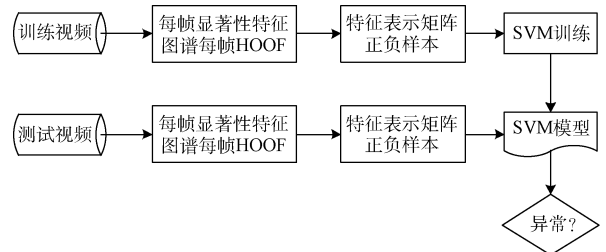


图 3 异常检测算法流程

每个视频帧按  $20 \times 20$  的尺寸被分为  $12 \times 16$  的图像块做光流直方图,每块的光流直方图为 9 维;每帧以  $8 \times 8$  的块大小做四元数傅里叶变换并获得  $30 \times 40$  的显著性特征矩阵,单帧视频有 2 928 维的特征,但是考虑到很多块是不包含任何有效运动信息的,因此在特征提取之前以有运动像素个数大于 60% 则保留数据否则置零来做预处理,故得到的特征其实非常稀疏的,通过 PCA 降维可大大提高检测效率。

以 UMN 数据库的场景 1,即图 4 中的草坪场景为例,将显著性特征+HOOF 与单独的 HOOF 做对比,检测结果如下图所示。左侧是正常帧示例,正常帧用透明条块表示,右侧为异常帧示例,异常帧用灰色条块表示。图下方的长条,第一个表示人工标定的真实值,第二个为仅仅通过 HOOF 作为特征的检测结果,第三个为综合本文提出的特征图谱和 HOOF 特征的检测结果。实验结果显示,加入基于四元数傅里叶变换的显著性特征图谱的后,误检率和漏检率都有所降低。

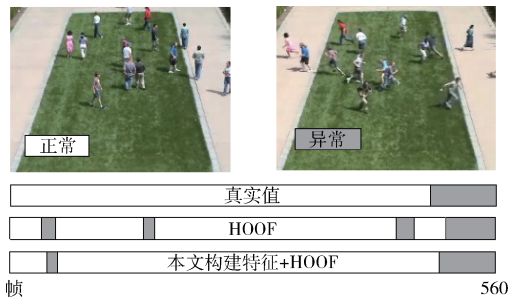


图 4 异常检测结果对比

## 5 结 论

密集场景的人群分析研究是智能监控技术的难点,而目前关于密集人群场景的研究多侧重于群体聚类或异常检测的算法研究,在底层的特征提取方面大都沿用已有的经典方法,或需要结合特定的复杂检测模型<sup>[15]</sup>,不具有普适性。本文给出了一种基于四元数傅里叶变换的特征融合方法,将时域由低阶到高阶的特征与空域的运动特征作为一个整体变换到频域,结合 HVS 理论计算出密集人群场景视频帧的显著性特征图谱。实验证明,该特征能够较为高效地检测出视频帧中的显著性区域,并能独立或与其他辅助特征级联作为异常检测的特征,为后续分析提供重要依据。

## 参考文献

- [1] LI T, CHANG H, WANG M, et al. Crowded scene analysis: a survey[J]. IEEE Transactions on Circuits & Systems for Video Technology, 2015, x(3): 367-386.
- [2] ALI S, SHAH M. A lagrangian particle dynamics

- approach for crowd flow segmentation and stability analysis[C]. IEEE Conference on Computer Vision and Pattern Recognition, 2007:1-6.
- [3] MEHRAN R, OYAMA A, SHAH M. Abnormal crowd behavior detection using social force model[C]. Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2009: 935-942.
- [4] ZHOU B, TANG X, WANG X. Coherent filtering: detecting coherent motions from crowd clutters[J]. Lecture Notes in Computer Science, 2012:857-871.
- [5] KRATZ L, NISHINO K. Anomaly detection in extremely crowded scenes using spatio-temporal motion pattern models [C]. IEEE Conference on Computer Vision & Pattern Recognition. IEEE Computer Society, 2009:1446-1453.
- [6] 谢捷, 焦佳佳, 王丽娜, 等. 密集人群建模中的阈值训练[J]. 电子测量技术, 2013, 36(2):34-38.
- [7] SHWARTZ S, NAMER E, SCHECHNER Y Y. Blind haze separation[C]. IEEE Computer Society, 2006:1984-1991.
- [8] GUO C, ZHANG L. A novel multiresolution spatiotemporal saliency detection model and its applications in image and video compression[J]. IEEE Transactions on Image Processing, 2010, 19(1): 185-198.
- [9] FANG Y, LIN W, LEE B S, et al. Bottom-up saliency detection model based on human visual sensitivity and amplitude spectrum [J]. IEEE Transactions on Multimedia, 2012, 14(1):187-198.
- [10] DALAL N. Finding people in images and videos[J]. Institut National Polytechnique de Grenoble-INPG, 2006.
- [11] HUBEL D H. Eye, brain and vision[J]. Scientific American Library. w. h, 1988, 72(9):1287.
- [12] JIANG Y, MA J. Combination features and models for human detection[C]. 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015.
- [13] SANGWINE S J, ELL T A. Hypercomplex Fourier transforms of color images[J]. IEEE Transactions on Image Processing, 2007, 16(1):22-35.
- [14] WANG T, SNOUSSI H. Histograms of optical flow orientation for visual abnormal events detection[C]. 2012 IEEE Ninth International Conference on Advanced Video and Signal-Based Surveillance (AVSS), 2012:13-18.

(下转第 80 页)