

# 基于时空关联度的视频车祸识别算法

赵 阳<sup>1</sup>, 李翼超<sup>2</sup>, 李 雪<sup>2</sup>

(1. 江苏省交通工程建设局, 南京 210004; 2. 长安大学 电控学院, 西安 710000)

**摘要:** 为了从一大段车祸视频中快速获取车祸发生时的那一小段视频, 从中准确识别车祸视频中车辆碰撞的局部信息, 从而尽快通知相关部门采取营救措施, 避免二次事故的发生, 实现快速辅助救援, 提出了时空结合的车祸识别方法; 通过对车祸视频进行局部特征点的提取和描述, 以及对车祸视频进行局部特征时空角点检测, 从而实现对车祸视频进行时空特征的提取; 经特征提取后可以获得车祸视频的 3DSIFT、STIP 和 3DHOG 三种特征, 对这 3 种特征分别进行不同组合的串联融合, 融合后可以得到 5 种串联结果, 对其串联特征进行 K-MEANS 聚类 and KNN 识别, 从而实现车祸视频的识别; 实验结果表明, 利用时空关联的算法进行识别, 能够准确找到车祸发生的片段, 可以提高车祸识别精度; 对比传统算法所得到的 33% 识别率, 基于时空关联度的视频车祸识别算法可提高车祸识别精度至 62%。

**关键词:** 车祸识别; 视觉图像; 时空关联; 特征提取; 识别精度

## Video Car Accident Recognition Algorithm Based on Spatio-temporal Correlation Degree

Zhao Yang<sup>1</sup>, Li Yichao<sup>2</sup>, Li Xue<sup>2</sup>

(1. Jiangsu Provincial Traffic Engineering Construction Bureau, Nanjing 210004, China;

2. School of Electronic Control, Chang'an University, Xi'an 710000, China)

**Abstract:** In order to quickly obtain a small video of a car accident from a large section of car accident video, the local information of the vehicle collision in the car accident video is accurately identified, so that the relevant department is notified as soon as possible to take rescue measures to avoid the occurrence of a second accident and achieve rapid assistance. Rescue, put forward a combination of space-time car accident identification method. Through the extraction and description of the local feature points of the car accident video, and the detection of the temporal and spatial corners of the local features of the car accident video, the temporal and spatial features of the car accident video are extracted. After the feature extraction, 3DSIFT, STIP and 3DHOG features of the car accident video can be obtained. These three features are respectively combined by different combinations of tandem fusion. After merging, 5 tandem results can be obtained, and K-MEANS clustering of the tandem features is performed. And KNN recognition to realize the recognition of car accident video. The experimental results show that using the algorithm of spatial-temporal correlation to identify the car accident can accurately find the fragments of car accidents, which can improve the accuracy of car accident recognition. Compared with the recognition rate of 33% obtained by the traditional algorithm, the video traffic accident recognition algorithm based on spatio-temporal correlation can improve the recognition accuracy of the crash to 62%.

**Keywords:** car accident recognition; visual image; spatiotemporal correlation; feature extraction; recognition accuracy

## 0 引言

我国是一个交通事故多发的国家, 据统计, 2010~2015 年, 因车祸事故死亡的人数都在 20 万以上<sup>[1]</sup>。车祸识别可以快速地为车祸现场管理提供准确的依据, 提高车祸现场快速处理的能力, 减少因车祸处理不及时而造成的人员伤亡, 有效避免二次事故发生。近年来, 车祸识别方法已成为交通安全领域研究的热门问题, 受到了很多专家的

重视。

目前, 车祸识别主要有基于射频、物联网, 以及基于视觉图像等识别方法<sup>[2-4]</sup>。相比之下, 基于视觉图像的车祸识别方法因具有直观、省时高效、设备要求低、获取信息准确等优势<sup>[4]</sup>而得到了广泛的研究。在基于时间的视频识别方法研究中, 时间受限的聚类法<sup>[5]</sup>将镜头相似的镜头聚类到一个固定的时间窗口中, 而位于窗口外的镜头相似性不予考虑, 因此聚类结果不够完全; 为了克服时间受限镜头聚类算法的不足, YongRui 等提出了时间自适应算法<sup>[6,7]</sup>, 即镜头的相似度随着它们之间的时间距离的改变而变化, 距离越大, 相似度越小。在基于空间的视频识别方法研究中, 镜头聚类的方法将镜头的关键帧合并为一个图像, 然后对合并后的图像分块, 求出镜头关键帧图像之间距离最小(相似度最大)的块; 镜头分类<sup>[8]</sup>与镜头聚类不

收稿日期: 2018-04-20; 修回日期: 2018-05-23。

**作者简介:** 赵 阳(1976-), 男, 江苏东台人, 副处长, 硕士, 高级工程师, 主要从事交通工程项目管理方向的研究。

**通讯作者:** 李翼超(1994-), 男, 陕西韩城人, 硕士研究生, 主要从事图像识别与处理方向的研究。

李 雪(1980-), 陕西咸阳人, 女, 副教授, 硕士生导师, 交通信息工程及控制专业工学博士。

同, 它是对镜头与镜头之间特征的相似性进行比较, 将特征不同的镜头分成不同的类。以上研究均仅在时间维或空间维的单一维度上进行, 对车祸现场细小碎片的可识别特征难以准确分割, 导致识别虚警率较高。

基于时空关联度的视频识别方法可以有效的避免维度利用过低的问题<sup>[9-10]</sup>, Mehmet C 等将时空关联度应用到无线传感器网络中<sup>[11]</sup>, Sadiq Ali 等将时空关联度应用到多天线频谱感知中<sup>[12]</sup>, Matoba O 等将时空关联度应用到快速安全通信中<sup>[13]</sup>。而对于基于时空关联度进行视频车祸识别, 目前鲜有研究。

本文将利用时空关联的方法实现车祸识别。首先为了将视频转化为更易表达的静态序列, 利用镜头分割算法<sup>[14]</sup>将视频在时间维上分为视频帧图片, 然后为了更准确地识别车祸, 对获得的时间维上的帧序列提取多种空间特征, 将这些特征串联融合, 从中筛选出需要的最优特征。实验结果表明, 利用本文算法进行车祸识别, 能够有效提高车祸识别精度, 为车祸现场管理提供了准确依据。

## 1 时空特征提取算法

### 1.1 车祸局部特征点提取

视频中的目标物体存在旋转、缩放、平移、视角变化、局部遮挡等情况, 这会使得视频识别出现误识别的现象<sup>[15]</sup>。为了解决这些问题, 利用 3DSIFT 算法提取局部特征点。SIFT 描述子可以在关键点的临近区域内进行梯度直方图统计, 从而可以构建出方向唯一的特征向量, 具有较好的方向稳定性。通过对车祸视频进行 3DSIFT 特征提取, 可得出车祸视频的特征直方图, 特征直方图可以描述车祸发生前后的特征变化。

在局部特征点提取算法中, 首先在高斯差分 (DoG) 尺度空间内, 检测图像的关键点。为了确保关键点的准确性, SIFT 描述子需要对 DoG 空间相邻图像进行采用和比较。通过比较临近区间内的点, 可得出该区间内的极大值点或极小值点, 并且将不符合条件的或符合程度低的点删除, 得出的极值点即所需关键点。如图 1 所示。

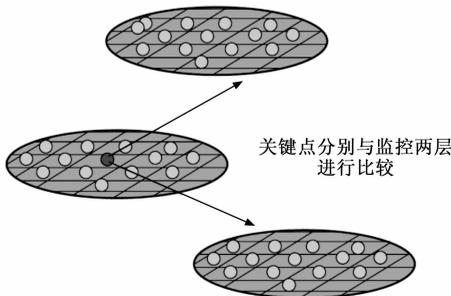


图 1 基于 DoG 空间内的极值点检测

得出图像的关键点后, 为了保证 SIFT 特征的方向稳定性, 需要对 SIFT 描述子在关键点临近区域内进行尺度方向直方图统计, 然后求取特征点的主方向。这个过程将图

像均分为多个小块, 将不同小块内的方向向量提取出来, 然后在将这些小块以及它们的方向向量合并到一起, 从而构建出唯一的, 具有代表性的特征向量。如图 2 所示。

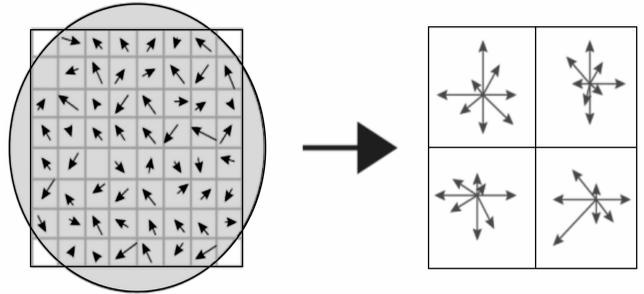


图 2 SIFT 特征描述子构建

通过尺度方向直方图统计得出特征点的主方向后, 需要根据求取出来的这个主方向构建梯度直方图, 即生成关键点描述子。计算关键点  $p$  的  $k$  邻域点  $p_i$  ( $i=1, 2, \dots, k$ ), 为了保证旋转不变性, 将关键点  $p$  的  $k$  邻域点旋转至关键点主方位, 邻域点坐标变换公式如式 (1) 所示。

$$p_i' = \begin{cases} \cos\alpha_p \cos\beta_p & -\sin\alpha_p & -\cos\alpha_p \sin\beta_p \\ \sin\alpha_p \cos\beta_p & \cos\alpha_p & -\sin\alpha_p \sin\beta_p \\ \sin\beta_p & 0 & \cos\beta_p \end{cases} p_i \quad (1)$$

其中:  $p_i'$  为旋转后对应的点坐标。对于关键点  $p$  及其邻域点  $p_i'$  可以用得到的 4 元组  $(m, \theta, \varphi, \delta)$  表示, 将关键点邻域分为  $n * n$  的子区域, 分别统计子区域中所有邻域点在各个方向的角度分布, 得到关键点  $p$  处的特征描述子。从而生成完整的 3DSIFT 特征描述子。

### 1.2 车祸局部特征的时空角点检测

为了检测车祸视频中撞车细节的局部特征, 采用 STIP (space-time interest point) 时空角点检测算法, 需要先将图像兴趣点的检测从 2D 空间兴趣点转化为 3D 空-时兴趣点。采用 Harris-Corner 兴趣点检测, 在图像的  $X$  轴和  $Y$  轴的基础上添加时间  $T$  轴, 并消除视频图像结构中存在的加速度, 从而将二维特征转化为三维特征。

在进行 Harris-Corner 兴趣点检测前, 需要先创建一个空-时兴趣点特征描述器, 求得兴趣点的特征描述向量。对于每一个样本点  $(x, y, t, \sigma, \tau)$ , 特征描述器描述的是关于中心点  $(x, y, t)$  周围区域的 3D 空间, 空间大小  $\Delta_x(\sigma)$  和  $\Delta_y(\sigma)$  是  $\sigma$  的函数, 时间长度  $\Delta_t(\tau)$  是  $\tau$  的函数。在这里描述器的大小将由式 (2) (3) 所示。

$$\Delta_x(\sigma) = \Delta_y(\sigma) = 2 \cdot \text{ceil}(3\sigma) + 1 \quad (2)$$

$$\Delta_t(\tau) = 2 \cdot \text{ceil}(3\tau) + 1 \quad (3)$$

根据这些描述器求得兴趣点的特征描述向量。得到空-时兴趣点特征描述器后再进行 Harris-Corner 兴趣点检测, 如式 (4) 所示。

$$L(\cdot, \sigma_i^2, \tau_i^2) = \partial_{\mu} M(g^{\mu}(\cdot, \sigma_i^2, \tau_i^2)) * f(\cdot) \quad (4)$$

建立一个空-时图像序列模型, 如式 (5) 所示。

$$g(x, y, t; \sigma_x^2, \tau_x^2) = \frac{\exp(-(x^2 + y^2)/(2\sigma_x^2 - \tau_x^2/2\tau_x^2))}{\sqrt{(2\pi)^2 \sigma_x^4 \tau_x^2}} \quad (5)$$

得出 3 维 Harris—Corner 检测的三阶矩阵, 如式 (6) 所示。

$$\mu = g(\cdot, \cdot, \cdot; \sigma_x^2, \tau_x^2) * \begin{bmatrix} L_x^2 & L_x L_y & L_x L_t \\ L_x L_y & L_y^2 & L_y L_t \\ L_x L_t & L_y L_t & L_t^2 \end{bmatrix} \quad (6)$$

为了检测兴趣点, 需要搜索区域中具有现在特征值  $\lambda_1, \lambda_2, \lambda_3$  的  $\mu$ , 并扩展式 (5) 中空间 STIP 兴趣点, 检测成为空一时 STIP 兴趣点检测, 如式 (7) 所示。

$$H = \det(\mu) - k \text{trace}^3(\mu) = \lambda_1, \lambda_2, \lambda_3 - k(\lambda_1, \lambda_2, \lambda_3)^3 \quad (7)$$

得出  $H$  的极大值点, 然后检测、选择及排除得到需要的时空角点。

### 1.3 车祸局部特征点描述

为了将图像空间中的三维兴趣点特征转化为描述符向量, 采用局部特征点描述 HOG3D 算法。HOG3D 三维梯度方向直方图描述器将视频空间看做一个正  $N$  面体, 然后对该正  $N$  面体进行均匀分割, 将它切成均匀的小块, 然后在这些小块里面进行统计梯度直方图。一般先在视频空间或图像空间里进行区域检测, 然后对其感兴趣的区域里进行兴趣点采集, 从而得到其三维兴趣点, 具体流程如图 3 所示。

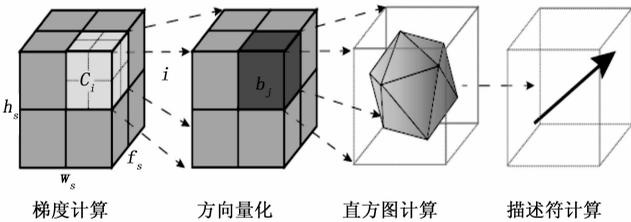


图 3 HOG3D 三维梯度方向直方图描述器

在计算 HOG3D 特征时, 首先从图像中定位出一个大的目标区域 Block 进行图像分块, 将图像按照它的大小尺寸等进行均匀的分割。然后, 均分 Block 至较小的块 Cell, 直至均分 Cell 形成最小的块 Pixel, 也就是像素点。

对图像进行打块后需要在 Block 内进行归一化直方图。对每个小块进行直方图提取, 再将这些直方图经过串联的方式组合为一个完整的特征向量。算法通过局部打块的方式逐一获取整个系统的图像特征, 不仅获取了整体特征, 同时获得了局部细节特征; 由于进行了直方图归一化的操作, 可有效避免因外界光线变化所带来的系统干扰。

### 1.4 算法流程

本文算法以车祸视频帧分割算法的时间特征为基础, 结合局部特征点提取、检测和描述算法, 实现对车祸视频的准确识别。算法流程如图 4 所示。

步骤 1: 视频帧提取。为了将视频在时间维上依次转化为对应的时间帧序列, 采用镜头分割算法将视频转化为系统帧图片。

步骤 2: 空间特征提取。为了解决车祸视频中因目标物

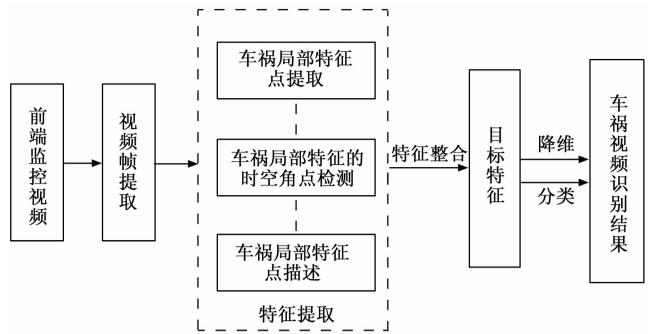


图 4 基于时空关联度的车祸识别流程图

体存在的旋转、缩放、平移、视角变化和局部遮挡等造成的特征方向不稳定的问题, 提出局部特征点提取算法; 为了在从车祸视频中具有加速度的车辆上获取其撞车细节的局部特征, 提出时空角点检测算法; 为了将图像空间中获取的三维时空兴趣点转化为描述符向量, 提出局部特征点描述算法。

步骤 3: 特征融合。串联融合车祸视频中提取的多种特征, 如图 5 所示。

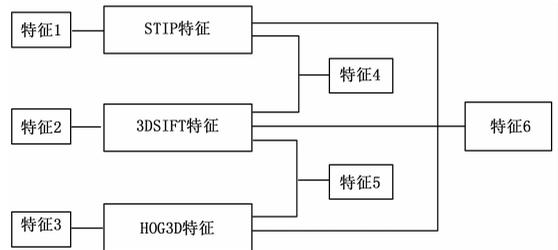


图 5 串联特征融合方案

步骤 4: 降维处理。为了解决视频特征数量过多的问题, 对融合后的视频特征进行降维聚类。

步骤 5: 车祸识别分类。将车祸视频识别的结果分为发生车祸与未发生车祸两类, 得出识别结果。

## 2 算例

### 2.1 视频帧提取

镜头分割即检测镜头的边界, 然后将视频从检测到的边界处分割开, 形成独立镜头的过程。经过分割转化后的结果如图 6 所示。



图 6 经分割转化后的视频帧

将动态的车祸视频转化为镜头帧序列对其进行转存。结果如图 7 所示。



图 7 转存后的视频帧文件

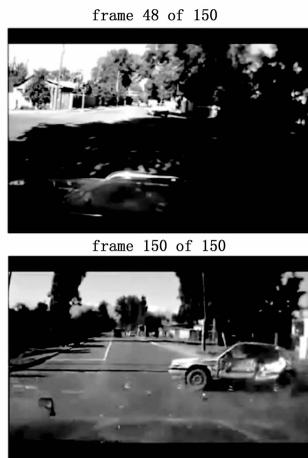


图 9 统计视频时空特性

### 2.2 车祸局部特征点提取

通过对车祸视频进行 3DSIFT 特征提取后, 可以得出车祸视频的特征直方图, 如图 8 所示。

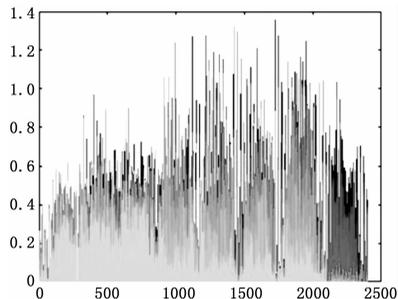


图 8 车祸视频的特征图

从图中可以看出, 在横坐标 2 000 之前特征图基本都是黄色, 表明此时视频处于平缓的状态, 此段视频中的对象并没有发生剧烈的颜色、形状及位置突变; 而在 2 000~2 500 区域时, 特征颜色明显的发生了突变, 表明在这个时刻视频内现场场景发生了剧烈变动, 从而得出在 2 000 时为车祸异常的时刻。

### 2.3 车祸局部特征的时空角点检测

通过 STIP 算法得出车祸视频的统计时空特性, 如图 9 所示。从图中可以看到, 将车祸发生前与发生后的帧图片进行了分隔归离。

该车祸视频的时空特征点检测图如图 10 所示。图中可以看到, 除了传统的  $x$  轴和  $y$  轴外, 增加了时间轴, 这些时空特征点邻域的像素值在时间和空间都有显著的变化。在该算法中, 邻域块的尺度大小能够自适应时间维和空间维。

通过 STIP 算法求得车祸视频的检测兴趣点, 结果如图 11 所示。

图中圆环即 STIP 算法在视频内检测到的点。因为各点在  $x$ 、 $y$  轴的坐标距离都不同, 所以生成的圆环的彼此的直径也不相同。从图中可以看到, 兴趣点能够找到车的车头、车轮、车尾这些重要部位, 准确得出车祸中车辆的局部细节。但也有少数兴趣点的检测位置并不在车身上, 这

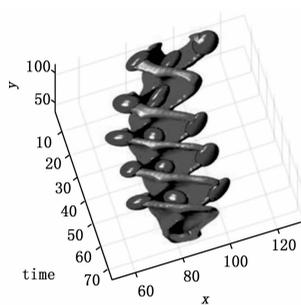


图 10 时空特征点检测图

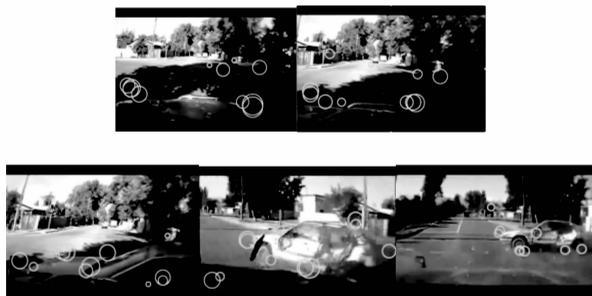


图 11 车祸视频中第 3、59、121、131、150 帧兴趣点提取结果

是因为视频内的车辆在进行平移直线运动或者匀速运动时, 周围的环境没有显著变化, 算法会暂时将周围环境归到车辆本身。但在车祸发生后, 周围环境发生了空间突变, 兴趣点会因此而重新定位, 并正确标记到车辆上。因此 STIP 算法检测到的兴趣点具有很好的准确性。

### 2.4 车祸局部特征点描述

实验中所用到的车祸视频的帧图像大小为  $480 \times 640$ , 把图像里面的每  $16 \times 16$  个元素组成的对象编入为 1 个 Cell 块, 从而生成了  $30 \times 40$  个 Cell 块, 一共有  $(30 - 1) \times (40 - 1) = 1\ 131$  个 Block, 每个 Block 是  $16 \times 4 = 64$  维, 计算得到样本图像 HOG 特征向量维数是  $1\ 131 \times 64 = 72\ 384$ 。最终提取的车祸图片的  $1\ 131 \times 64$  维 HOG 特征的网络图如图 12 所示。

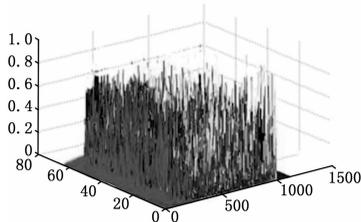


图 12 提取图像 HOG 特征

从图中可以看出，HOG3D 算法提取出了车祸发生的这一瞬间的帧图片，而右边 HOG 特征的网络图中前半部分为深色区域，而后半部分为浅色区域，这表示前后两部分视频内的时空场景发生了剧烈的改变，即在颜色发生改变的位置是视频内容突变前车祸发生碰撞的瞬间。

### 3 实验结果

经过特征融合会得到 5 种特征，对这 5 种特征进行 K-MEANS 聚类 and KNN 识别，然后从中选出最优的融合方案。如图 13 所示。

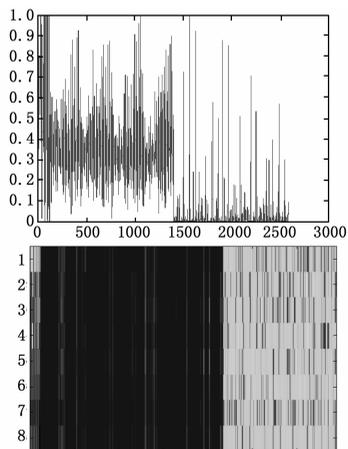


图 13 特征融合后的示意图

图 13 即 3DSIFT 和 STIP 融合以及 3DSIFT、STIP 及 HOG3D 三者融合后的识别结果。从 3DSIFT 与 STIP 的融合结果可以看出，图像的前半部分和后半部分具有显著的差异，即车祸发生前与发生后的特征具有明显的变化。而在 3DSIFT、STIP、HOG3D 种特征融合后可以很清晰的看到车祸前与车祸后的分界点，这表明将 3 种特征融合在一起可以提高车祸识别的准确率。

本文针对不同特征得出的车祸识别率，做了 10 组车祸视频的实验，其中每组实验先仅使用 3DSIFT、STIP 和 HOG3D 中的一种特征进行识别，再使用二维和三维两种特征进行识别，最后使用 3 种特征的融合特征进行识别。实

验所得结果如表 1。

表 1 各类特征在 K-MEANS 聚类与 KNN 识别之后的结果比较

特征/方法	K-MEANS 聚类后 KNN 识别结果
3DSIFT	32%
STIP	29%
HOG3D	33%
3DSIFT+STIP	50%
3DSIFT+HOG3D	51%
3DSIFT+STIP+HOG3D	62%

从表中可以得出，在只使用一种特征时，识别率都低于 50%；进行特征融合后识别率有了明显的提高；而在将 3 种特征全部融合起来后其 KNN 识别率达到最高，为 62%，这对比传统算法所得到的 33%，具有很大的提高。因此可得出结论，将 3DSIFT、STIP 和 HOG3D 三种特征融合后其融合特征能够很大程度上提高车祸识别精度。

### 4 结论

本文以提高车祸视频识别率为目的，针对传统视觉图像技术车祸识别方法的缺点，提出基于时空关联度的车祸视频识别算法，利用视频分割算法将视频分割为时间帧序列，再对视频图像的空间局部特征进行检测、提取和描述后，然后进行融合处理，可以有效利用视频的时空特征。对比实验表明，本文车祸视频算法有效地提高了算法识别精度，对于交通事故实时识别与紧急救援具有重要作用。

#### 参考文献：

- [1] 王 栋, 赵中奇. 基于视觉特征定位的物联网车祸现场识别 [J]. 计算机仿真, 2014, 31 (10): 196-199.
- [2] Lin C L. An approach to adaptive infrared image enhancement for long-range surveillance [J]. Infrared Physics and Technology, 2011, 54 (2): 84-91.
- [3] 顾 明, 郑林涛, 刘中华. 结合暗红色优先和 Gamma 校正的红外交通图像增强算法 [J]. 交通运输工程学报, 2016, 16 (06): 149-158.
- [4] Dixit R S, Gandhe S T. Pedestrian detection system for ADAS using Friendly ARM [A]. IEEE. 2015 International Conference on Energy Systems and Applications [C]. New York: IEEE, 2015: 557-560.
- [5] Rui Yong, Huang Thomas Sharad. Constructing table-of-content for videos [J]. Multimedia Systems, 1999, 7 (5): 359-368.
- [6] Irena Koprinska, Sergio Carrato. Temporal video segmentation: A survey [J]. Signal Processing: Image Communication, 2001, 16 (5).
- [7] Rui Y, Huang T S, Mehrotra S. Construct Table-of-content for Videos [J]. ACM Multimedia System, 1999, 7 (5): 359-368.
- [8] 王学军, 丁洪涛, 陈贺新. 一种基于镜头聚类的视频场景分割方法 [J]. 中国图像图形学报, 2007, 12 (12): 2127-2131.