



JOURNAL OF IMAGE AND GRAPHICS

主办: 中国科学院遥感与数字地球研究所
中国图象图形学学会
北京应用物理与计算数学研究所

中国图象图形学报

2013
08
VOL.18

ISSN1006-8961
CN11-3758/TB



北魏



隋



初唐

数字文化遗产研究 P968

综述

尺度不变特征变换算子综述

刘立, 詹茵茵, 罗扬, 刘朝晖, 彭复员 885

图像处理和编码

改进的H.264帧间模式选择算法

任克强, 张旭光, 罗会兰 893

均质雾天下摄像机动态标定的车速估计

宋洪军, 陈阳舟, 郜园园 901

基于人脸五官特征的空域差错掩盖算法

张江鑫, 谢晋, 邝万坤 913

采用圆谐-傅里叶矩的图像区域复制粘贴篡改检测

秦娟, 李峰, 向凌云, 殷荻茗 919

图像分析和识别

基于视频感知哈希的视频篡改检测与多粒度定位

朱映映, 文振焜, 杜以华, 邓良太 924

视觉注意模型的道路监控视频关键帧提取

刘云鹏, 张三元, 王仁芳, 张引 933

格拉斯曼流形上的半监督判别分析

姜伟, 陆瑶, 杨炳儒 944

精神疲劳实时监测中多面部特征时序分类模型

陈云华, 张灵, 丁伍洋, 严明玉 953

图像分类中的概率乘积核函数

杨赛, 赵春霞 961

轮廓整体结构约束的壁画图像相似性度量

唐大伟, 鲁东明, 杨冰, 许端清 968

应用色彩空间聚类方法实现道路建模

向宸薇, 王拓, 于舰 976

图像理解和计算机视觉

杂乱背景和摄像机移动下的时空兴趣点检测

刘长红, 陈勇, 王明文 982

形状和颜色混合不变量在图像检索中的应用

公明, 曹伟国, 李华 990

计算机图形学

结合统计分类和边缘特征的最优视图提取

潘翔, 陈敖, 章国栋 1004

医学图像处理

局部熵驱动下的脑MR图像分割与偏移场恢复耦合模型

张建伟, 杨红, 陈允杰, 方林, 詹天明 1011

改进的模糊聚类实质肺结节3维分割

李翠芳, 聂生东, 王远军, 孙希文, 郑斌 1019

遥感图像处理

高光谱海岸带区域分割的活动轮廓模型

王相海, 金弋博 1031

改进的基于非局部均值的极化SAR相干斑抑制

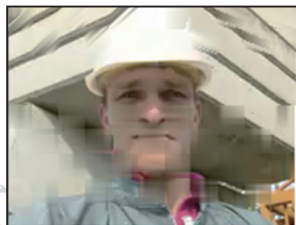
赵忠民, 赵拥军, 牛朝阳 1038

小波变换和稀疏表示相结合的遥感图像融合

刘婷, 程建 1045

《中国图象图形学报》论文摘要写作要求

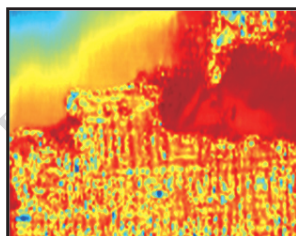
1054



基于人脸五官特征的空域差错掩盖算法 (第913页)



应用色彩空间聚类方法实现道路建模 (第976页)



改进的基于非局部均值的极化SAR相干斑抑制 (第1038页)

CONTENTS

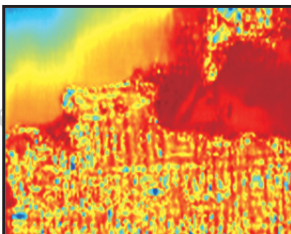
JOURNAL OF IMAGE AND GRAPHICS



Spatial error concealment algorithm based on the facial features (P913)



Road modeling using color spatial clustering (P976)



Improved polarimetric SAR speckle filter based on non-local means (P1038)

Review

Summarization of the scale invariant feature transform

Liu Li , Zhan Yinyin, Luo Yang, Liu Chaohui, Peng Fuyuan 885

Image Processing and Coding

Improved inter frame mode selection algorithm for H.264

Ren Keqiang, Zhang Xuguang, Luo Huilan 893

Vehicle velocity estimation based on camera calibration under homogenous fog

Song Hongjun, Chen Yangzhou, Gao Yuanyuan 901

Spatial error concealment algorithm based on the facial features

Zhang Jiangxin, Xie Jin, Kuang Wankun 913

Detection of image region copy-move forgery using radial harmonic Fourier moments

Qin Juan, Li Feng, Xiang Lingyun, Yin Changming 919

Image Analysis and Recognition

Video forgery detection and multi-granularity location based on video perceptual hashing

Zhu Yingying, Wen Zhenkun, Du Yihua, Deng Liangtai 924

Key frame extraction based on the visual attention model for lane surveillance video

Liu Yunpeng, Zhang Sanyuan, Wang Renfang, Zhang Yin 933

Semi-supervised discriminant analysis on Grassmannian manifold

Jiang Wei, Lu Yao, Yang Bingru 944

Time-series classification model based on multiple facial feature for real-time mental fatigue monitoring

Chen Yunhua, Zhang Ling, Ding Wuyang, Yan Mingyu 953

Probability product kernels in image classification

Yang Sai, Zhao Chunxia 961

Similarity metrics between mural images with constraints of the overall structure of contours

Tang Dawei, Lu Dongming, Yang Bing, Xu Duanqing 968

Road modeling using color spatial clustering

Xiang Chenwei, Wang Tuo, Yu Jian 976

Image Understanding and Computer Vision

Spatio-temporal interest point detection in cluttered backgrounds with camera movements

Liu Changhong, Chen Yong, Wang Mingwen 982

Application of combined shape and color invariants to image retrieval

Gong Ming, Cao Weiguo, Li Hua 990

Computer Graphics

Best view selection based on statistical classification and edge features

Pan Xiang, Chen Ao, Zhang Guodong 1004

Medical Image Processing

Brain MR image segmentation and bias correction model based on local entropy

Zhang Jianwei, Yang Hong, Chen Yunjie, Fang Ling, Zhan Tianming 1011

Segmentation of sub-solid pulmonary nodules based on improved fuzzy

C-means clustering

Li Cuifang, Nie Shengdong, Wang Yuanjun, Sun Xiwen, Zheng Bin 1019

Remote Sensing Image Processing

The active contour model for segmentation of coastal hyperspectral remote sensing image

Wang Xianghai, Yin Yibo 1031

Improved polarimetric SAR speckle filter based on non-local means

Zhao Zhongmin, Zhao Yongjun, Niu Chaoyang 1038

Remote sensing image fusion with wavelet transform and sparse representation

Liu Ting, Cheng Jian 1045

中图法分类号: TP391.4 文献标识码: A 文章编号: 1006-8961(2013)08-0982-08

论文引用格式: 刘长红,陈勇,王明文. 杂乱背景和摄像机移动下的时空兴趣点检测[J]. 中国图象图形学报, 2013, 18(8): 982-989. [DOI: 10.11843/jig.20130813]

杂乱背景和摄像机移动下的时空兴趣点检测

刘长红¹, 陈勇², 王明文¹

1. 江西师范大学计算机信息工程学院, 南昌 330022; 2. 南昌工程学院工商管理学院, 南昌 330029

摘要: 提出了一种杂乱背景和摄像机移动场景下的时空兴趣点检测方法。结合空间域上的平滑和时间域上的平滑;在空间域上利用非线性各向异性扩散进行平滑,从而有效平滑掉噪声和杂乱背景中的细小结构,同时保持了人体动作的边缘信息;在时间域上利用了一个与梯度成正比的抑制函数对摄像机晃动/移动所导致的时间关系上的不确定性进行建模。本文方法有效地解决了复杂场景下时空兴趣点检测面临的杂乱背景和摄像机移动问题。

关键词: 时空兴趣点检测;非线性各向异性扩散;杂乱背景;摄像机移动

Spatio-temporal interest point detection in cluttered backgrounds with camera movements

Liu Changhong¹, Chen Yong², Wang Mingwen¹

1. College of Computer and Information Engineering, Jiangxi Normal University, Nanchang 330022, China;

2. Business Administration College, Nanchang Institute of Technology, Nanchang 330029, China

Abstract: A spatio-temporal interest point detection approach in the environment of clutter backgrounds with camera shaking/movements is proposed. It incorporates smoothing functions in the spatial and temporal domain. Non-linear anisotropic diffusion is used in spatial direction to efficiently smooth noises and small structures in cluttered backgrounds and to preserve edge information in human action. The uncertainty of the temporal correlation due to camera shaking/movements is modeled by a suppressing function in proportion to the image gradients. This method efficiently solves the problem of cluttered backgrounds and camera shaking/movements in complex scenes.

Key words: spatio-temporal interest point detection; nonlinear anisotropic diffusion; clutter background; camera movement

0 引言

在复杂场景下的人体动作识别中,存在人体表现的变化、遮挡、杂乱背景、光照变化以及摄像机晃动/移动这些因素的影响,使得全局表示方法难以准确地提取出侧影或光流信息,侧影和光流特征受这些因素影响很大,因此本文采用局部表示的方法来进行复杂场景下的人体动作识别,局部表示的方法

不需要背景减除,而且能对少量的人体表现、光照和视角变化以及部分遮挡保持不变。局部表示方法通常采用基于 BoF 的识别框架,基于 BoF 的人体动作识别模型一般包括:时空兴趣点检测和表示、矢量量化、直方图计算和分类几个步骤,而时空兴趣点检测在该识别模型中起着首要作用,时空兴趣点检测的性能直接影响到后面一系列的处理过程,如果兴趣点检测错误,那么后面的方法再好都得不到很好的识别性能,也就是局部表示中后面其他的一系列处

收稿日期:2012-10-12;修回日期:2013-01-14

基金项目:国家自然科学基金项目(61272212,61262036);江西师范大学博士启动基金项目

第一作者简介:刘长红(1977—),女,讲师,2011年于北京科技大学获计算机应用技术博士学位,主要研究方向图像处理、模式识别、计算机视觉。E-mail:ahong286@126.com

理过程都依赖检测到的兴趣点的鲁棒性和可靠性。根据时空兴趣点提取到的特征编码了视频中的形状和运动信息,是视频中内容的一种紧致表示。在过去的十多年,研究者们提出了很多 2D 空间和 3D 时空兴趣点检测方法以及描述方法^[1-5]。

Laptev 提出的 Harris3D 检测方法^[1]是 Harris 角点检测子 3D 的扩展形式,分别对时空维上使用高斯平滑函数和时空梯度计算每个点的时空二阶矩矩阵 $u(\cdot; \sigma, \tau) = g(\cdot; s\sigma, s\tau) * (\nabla L(\cdot; \sigma, \tau)(\nabla L(\cdot; \sigma, \tau))^T)$, 时空兴趣点为时空邻域中有重大变化的点,通过 $H = \det(u) - k \operatorname{tr}^3(u)$ 的局部极大值给定。该方法的时空兴趣点为包含空间角的区域和速度变化的方向,因此检测到的兴趣点非常稀疏,大致对应运动的起点和终点。该方法的缺点是:兴趣点非常稀疏,检测不到匀速运动的人体动作变化;容易受噪声、杂乱背景和摄像机晃动/移动的影响,即便是非常微小的摄像机晃动都将检测到很多的背景点。

Dollar 提出的 Cuboid 检测方法^[2]指明了一定领域内人体动作表示的时空角的弱点,分别在时间和空间维上采用线性滤波器,在空域上使用 2D Gaussian 滤波器和时域上使用 1D 的 Gabor 滤波器,然后这些滤波器被组合产生最后的响应,兴趣点对应局部响应最大的位置。该兴趣点检测方法提供了更密集的兴趣点,在人体动作识别得到广泛应用。但该方法还是在平移运动(translational motions)上检测到的兴趣点非常少,很容易误检测到噪声,对于复杂场景下的人体动作,检测的兴趣点很多位于背景上,而不是人体动作上,对慢移动物体、微量摄像机运动无效,特别对于大些幅度的摄像机晃动和移动,几乎检测不到人体动作,检测到的兴趣点全是背景物体,这些结果可能是由于时间域上的 1D Gabor 滤波器对噪声和前景纹理很敏感。如图 1 所示,由于骑马运动中摄像机移动比较快,目标和背景同时都发生

运动,而且是户外,背景非常杂乱,因而检测到的兴趣点基本上为背景点。

Oikonomopoulos 提出的基于 Saliency 的检测方法^[3]扩展了 Kadir-Brady 的基于互信息的 Saliency 检测用于 3D 的时空兴趣点检测,在各种尺度上估计各个像素上邻域内的熵,尺度之间熵变换的局部极值位置为兴趣点。这种兴趣点检测方法没有考虑摄像机晃动/移动所导致的时间关系上的不确定性,得到不准确的熵率,从而检测到背景中很多错误的兴趣点。

Willems 提出的 Hessian 检测方法^[4]是图像 blob 检测中 Hessian 突性测量的时空扩展形式,利用 3D Hessian 矩阵的行列式测量突性(saliency)来确定兴趣点的位置和尺度。然而该方法同样对动态复杂的场景很敏感,容易受到背景物体移动的影响,特别是摄像机的晃动和移动,导致兴趣点大量为背景中的点。

从以上的分析可知,现有典型的时空兴趣点检测方法都存在以下的问题:只能对相对简单背景下的人体动作进行检测,对复杂动态场景下的人体动作难以检测到,容易受到噪声、杂乱背景中的物体和摄像机晃动/移动所带来的影响。然而现实场景的视频都存在这些因素的影响。因此需要寻求新的解决方法来解决这些问题。

最近 Shabani 等人提出了一种非线性滤波器的方法来提取时空序列中鲁棒的时空特征^[6]。该方法采用了非线性各向异性扩散滤波器(nonlinear anisotropic diffusion filtering)来进行现实视频中时空兴趣点检测,结合了两个自适应的非线性平滑函数对低梯度区域强平滑而高梯度区域弱平滑。该方法很好地解决了以上存在的一些问题,能够抑制一些噪声、杂乱背景所带来的干扰,而且对摄像机的晃动和移动所导致的时间关系上的不确定性进行了建模,但是该方法在空间域上采用了 Perona-Malik 各向异性扩散及其停止函数,在目标的边缘还是会产生一些模糊和错位,而且只对微小的摄像机晃动和移动的场景有效,对于像“骑车”、“骑马”、“户外行走”等背景非常杂乱和摄像机晃动/移动比较大的情况下效果还是不好,检测的兴趣点仅有少量在人体上或马的各部位上。

针对杂乱背景和摄像机移动场景下的时空兴趣点检测问题,提出了一种新的时空兴趣点检测方法,采用与 Shabani 相似的方法,同时对空间域上和时间域上进行平滑,对空间域上采用了各向异性扩散,扩散过程是多向(multi-lateral)的,时间域上建模了

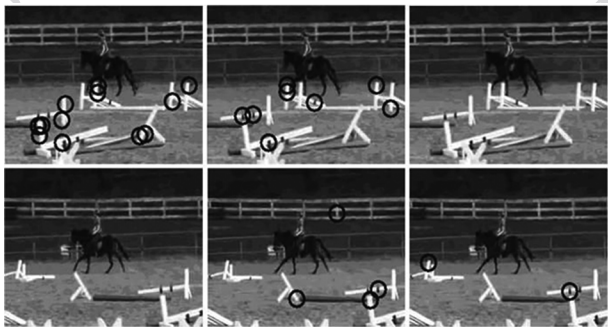


图 1 兴趣点检测方法^[2]检测结果

Fig. 1 The results of interest point detection approach^[2]

摄像机晃动/移动导致的不确定性,扩散过程是单向(unilateral)的。但在空间域上采用 Weickert 提出的边缘增强的非线性各向异性扩散滤波器,而时间上采用与梯度成正比的抑制函数,来抑制摄像机晃动/移动所导致的影响。

1 非线性各向异性扩散滤波

结构张量(structure tensor)是图像分析的有力工具,能检测边、角、线等图像局部结构,常用来估计图像结构方向场和分析图像局部几何结构,已成功地应用到图像结构方向场计算、特征检测、图像去噪等领域。结构张量的定义为

$$J_{\rho}(\nabla I_{\sigma}) = G_{\rho} * (\nabla I_{\sigma} \otimes \nabla I_{\sigma}) = \begin{pmatrix} J_{11} & J_{12} \\ J_{12} & J_{22} \end{pmatrix} \quad (1)$$

式中, $\otimes$ 表示张量积运算, I_{σ} 为高斯滤波后的图像, σ 是高斯滤波器 G_{σ} 的方差,用来减少噪声对求导运算的影响, ρ 是高斯滤波器 G_{ρ} 的方差,张量积 $\nabla I_{\sigma} \otimes \nabla I_{\sigma}$ 与高斯滤波器 G_{ρ} 的卷积主要用来增强对线、角、断裂边缘等几何结构的刻画能力。

结构张量 J_{ρ} 是对称的半正定阵,其两个正交的特征向量 V_1 和 V_2 分别为

$$V_1 = \begin{pmatrix} 2J_{12} \\ J_{22} - J_{11} + \sqrt{(J_{11} - J_{22})^2 + 4J_{12}^2} \end{pmatrix} \quad (2)$$

$$V_2 = V_1^{\perp}$$

式中,“ $\perp$ ”表示 V_2 与 V_1 垂直。

对应的特征值 U_1 和 U_2 为

$$U_1 = \frac{1}{2}(J_{11} + J_{22} + \sqrt{(J_{11} - J_{22})^2 + 4J_{12}^2}) \quad (3)$$

$$U_2 = \frac{1}{2}(J_{11} + J_{22} - \sqrt{(J_{11} - J_{22})^2 + 4J_{12}^2})$$

式中,特征向量 V_1 为几何结构的最大对比度方向,对应几何结构的垂直方向;特征向量 V_2 为几何结构的最小对比度方向,对应几何结构的方向。特征值 $U_1 \geq U_2 \geq 0$ 刻画了图像的局部几何结构,当 $U_1 = U_2 = 0$ 表示常量平滑区域,当 $U_1 \gg U_2 = 0$ 表示直边区域,当 $U_1 \geq U_2 \gg 0$ 表示角区域。

为了使扩散过程适用于结构张量,Weickert 提出了基于以下方程的非线性各向异性的扩散滤波器^[7],即

$$\frac{\partial I}{\partial t} = \text{div}(D \cdot \nabla I) \quad (4)$$

$$I(x, y, 0) = I_0(x, y)$$

和 Perona-Malik 扩散模型中一样, $I_0(x, y)$ 表示原图像, $I(x, y, t)$ 表示随时间演变平滑后的图像, t 是扩散时间。 D 为扩散张量,是一个正定的对称矩阵,是结构张量的函数,使用和结构张量 J_{ρ} 中一样的特征向量 V_1 和 V_2 ,平滑的方向沿着图像的局部结构,平滑强度由 D 中的特征值 λ_1 和 λ_2 给定,其表示为

$$D = [V_1 \ V_2] \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} [V_1 \ V_2]^T \quad (5)$$

$\lambda_1 \geq \lambda_2 \geq 0$, λ_1 和 λ_2 的值位于0和1之间,可通过设置 λ_1 和 λ_2 合适的值来控制对图像中几何结构的平滑强度。当 $\lambda_1 = 1$ 表示在特征向量 V_1 方向进行强平滑, $\lambda_1 = 0$ 则在特征向量 V_1 方向没有平滑;同理, $\lambda_2 = 1$ 则表示在特征向量 V_2 方向进行强平滑, $\lambda_2 = 0$ 表示在特征向量 V_2 方向没有平滑。

对于边缘增强,则 $\{U_i\}_{i=1}^2$ 中越小的值对应 $\{\lambda_i\}_{i=1}^2$ 中的值应该取越大的值,而 $\{U_i\}_{i=1}^2$ 中越大的值对应 $\{\lambda_i\}_{i=1}^2$ 中的值应该取越小的值。因为 $\{U_i\}_{i=1}^2$ 中越小的值,对应越平滑的方向(图像边缘方向),那么应该在该方向进行强平滑,所以对应的 λ_i 应该越大,而 $\{U_i\}_{i=1}^2$ 中越大的值,对应图像边缘的法线方向,所以对应的 λ_i 应该越小,保持边缘。因此,对于边缘增强,为了对图像局部区域内进行平滑而边缘禁止扩散,Weickert 设置 $\{\lambda_i\}_{i=1}^2$ 的值为

$$\begin{cases} \lambda_1 = g(|\nabla I|^2) \\ \lambda_2 = 1 \end{cases} \quad (6)$$

$$g(s) = \begin{cases} 1 & s \leq 0 \\ 1 - \exp\left(\frac{-C_m}{(s/\lambda)^m}\right) & s > 0 \end{cases} \quad (7)$$

式中, C_m 是一个常量。该方法能够平滑掉一些噪声和弱纹理区域,而且能够很好地保持图像结构边缘。

2 杂乱背景和摄像机移动下的时空兴趣点检测

2.1 非线性各向异性扩散滤波的时空兴趣点检测算法

受 Shabani 兴趣点检测方法^[6]启发,本文时空兴趣点检测方法也包括两部分的平滑:空间域上的平滑和时间域上的平滑。空间域上的平滑是多向的,采用了各向异性扩散及其停止函数。而时间域上的平滑主要是考虑摄像机移动晃动/移动所导致的时间关系上的不确定性,因此当前帧仅仅受前1

帧的影响,而与后面的帧和间接前面的帧是无关的,扩散是单向的,因此时间上的平滑仅考虑和直接的前 1 帧有关。对空间域上的平滑采用 Weickert 提出的边缘增强的非线性各向异性扩散滤波器^[7],保持和增强了边缘信息,同时又平滑掉噪声和图像结构内部;而对时间域上摄像机所带来的背景移动,采用了一个抑制函数,与 Shabani 方法^[6]中的时间域上的平滑函数相反,采用的抑制函数与时间域上的梯度成正比,也就是梯度越大,平滑强度越强,梯度越弱,平滑强度越弱。因为人体的区域一般都比较大,时间域上的平滑仅仅会导致人体边界区域的一些扩散,而对于比较小的区域(如线状物体)的运动变化则能得到比较好的平滑,减弱了它的运动变化强度。

兴趣点检测方法的非线性各向异性的滤波方程为

$$\frac{\partial \mathbf{I}(x, y, t_i, s)}{\partial s} = \text{div}(\mathbf{D} \cdot \nabla \mathbf{I}_{\text{sp}}) + g_{\text{tp}}(|\nabla \mathbf{I}_{\text{tp}}|) \cdot \nabla \mathbf{I}_{\text{tp}} \quad (8)$$

$$\mathbf{I}(x, y, t_i, 0) = \mathbf{I}(x, y, t_i)$$

式中, $\mathbf{I}(x, y, t_i, s)$ 表示第 t_i 帧在尺度 s (相当于扩散时间) 上平滑后的图像; $\mathbf{I}(x, y, t_i)$ 表示第 t_i 帧的原图像, div 表示扩散算子; $\nabla \mathbf{I}_{\text{sp}}$ 和 $\nabla \mathbf{I}_{\text{tp}}$ 分别表示空间上的梯度和时间上的梯度; g_{tp} 表示时间上的抑制函数,与时间域上的梯度成正比,抑制摄像机晃动/移动所带来的背景移动的影响,即

$$g_{\text{tp}}(|\nabla \mathbf{I}_{\text{tp}}|) = 1/(1 + (\lambda_{\text{tp}}/|\nabla \mathbf{I}_{\text{tp}}|)^2) \quad (9)$$

式中, λ_{tp} 控制时间梯度对平滑的程度。 $\mathbf{D}$ 为空间域上的扩散张量,是结构张量 $\mathbf{J}_\rho$ 的函数,采用 Weickert 提出的边缘增强的非线性各向异性扩散^[7],因此扩散张量为

$$\mathbf{D} = [\mathbf{V}_1 \quad \mathbf{V}_2] \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} [\mathbf{V}_1 \quad \mathbf{V}_2]^T = \begin{bmatrix} D_{11} & D_{12} \\ D_{12} & D_{22} \end{bmatrix} \quad (10)$$

$$\begin{cases} \lambda_1 = g_{\text{sp}}(|\nabla \mathbf{I}_{\text{sp}}|^2) \\ \lambda_2 = 1 \end{cases} \quad (11)$$

$$g_{\text{sp}}(|\nabla \mathbf{I}_{\text{sp}}|^2) = 1/(1 + (|\nabla \mathbf{I}_{\text{sp}}|/\lambda_{\text{sp}})^2) \quad (12)$$

式中, λ_{sp} 控制空间域上的梯度对平滑的程度。 $\mathbf{V}_1$ 和 $\mathbf{V}_2$ 为结构张量 $\mathbf{J}_\rho$ 的两个正交向量。结合式(1)、式(2)和式(11)可以计算得到扩散张量中各个元素为

$$D_{11} = \frac{1}{2} \{ \lambda_1 + \lambda_2 + [(\lambda_2 - \lambda_1) \times (J_{11} - J_{22})/\alpha] \}$$

$$D_{12} = (\lambda_2 - \lambda_1) \times J_{12}/\alpha$$

$$D_{22} = \frac{1}{2} \{ \lambda_1 + \lambda_2 - [(\lambda_2 - \lambda_1) \times (J_{11} - J_{22})/\alpha] \} \quad (13)$$

式中

$$\alpha = \sqrt{(J_{11} - J_{22})^2 + 4J_{12}^2} \quad (14)$$

给定人体动作视频图像 $\{\mathbf{I}(x, y, t_i)\}_{i=1}^N$, N 表示视频帧数,采用式(8)对每帧图像进行平滑,将平滑后的图像与前 1 帧的原图像进行差分,得到的结果为突出的轮廓或点,然后对差分后的整个视频图像在一定时空窗口内采用非极大值抑制求解局部极大值,兴趣点即为局部极大值的位置。基于非线性各向异性扩散的时空兴趣点检测算法如下:

$\{\mathbf{I}(x, y, t_i)\}_{i=1}^N$: 输入的人体动作视频, N : 总共的帧数, s : 尺度(扩散时间), λ_{sp} : 空间域上控制梯度对平滑的程度, λ_{tp} : 时间域上控制梯度对平滑的程度, σ 是结构张量中高斯滤波器 $\mathbf{G}_\sigma$ 的方差, $\sigma_x, \sigma_y, \sigma_t$: 分别为局部非极大值抑制窗口的空间和时间上的尺度。

1) 利用式(9)对每帧进行平滑计算 $\mathbf{I}(x, y, t_i, s)$, 并得到每帧的突图。

for $i = 1, \dots, N$ do

for $j = 1, \dots, s$ do

$$\text{利用 } \frac{\partial \mathbf{I}(x, y, t_i, s)}{\partial s} = \text{div}(\mathbf{D} \cdot \nabla \mathbf{I}_{\text{sp}}) + g_{\text{tp}}(|\nabla \mathbf{I}_{\text{tp}}|) \cdot \nabla \mathbf{I}_{\text{tp}}$$

计算 $\mathbf{I}(x, y, t_i, s)$;

将最后平滑后的图像 $\mathbf{I}(x, y, t_i, s)$ 与前 1 帧原图像

$\mathbf{I}(x, y, t_{i-1})$ 进行差分得到突图。

end for

end for

2) 计算时空窗口内的响应极大值

对整个视频在时空窗口 $\sigma_x \times \sigma_y \times \sigma_t$ 内利用非极大值抑制方法计算突出的兴趣点位置。

for $k = 1, \dots$, 兴趣点个数

提取兴趣点周围时空体 cuboid 的特征。

end for

2.2 计算方法

采用有限前向差分的方法求解式(8)的非线性各向异性的扩散过程,根据 Weickert 提出的非线性各向异性的扩散方程式(4)的求解方法^[7],采用有限前向差分逼近可得

$$\frac{\partial I}{\partial t} \approx \frac{I^{t+1} - I^t}{\tau} = \sum_{j \in \Omega} A_j^t I^t \Rightarrow I^{t+1} = I^t + \tau \sum_{j \in \Omega} A_j^t I^t \quad (15)$$

式中, Ω 是当前像素点上的空间邻域, τ 是时间戳, I^t 表示在扩散时间 $t\tau$ 上的 $I(x, y)$ 的逼近, 表达式 $\sum_{j \in \Omega} A_j^t I^t$ 是 $\text{div}(\mathbf{D} \cdot \nabla I)$ 的离散化形式, A^t 是空间域上一个可变的模板(mask)。根据 Weickert 的求解方法, 可得到式(8)的有限前向差分逼近为

$$\frac{\partial I_{t_i}^{s+1}}{\partial s} \approx \frac{I_{t_i}^{s+1} - I_{t_i}^s}{\tau} = \sum_{j \in \Omega_{t_i}^s} A_{j, t_i}^s I_{t_i}^s + \nabla I_{t_i}^s g_{t_i}^s (|\nabla I_{t_i}^s|) \Rightarrow I_{t_i}^{s+1} = I_{t_i}^s + \tau \left(\sum_{j \in \Omega_{t_i}^s} A_{j, t_i}^s I_{t_i}^s + \nabla I_{t_i}^s g_{t_i}^s (|\nabla I_{t_i}^s|) \right) \quad (16)$$

式中, Ω_{t_i} 是第 t_i 帧当前像素点上的空间邻域, τ 是时间戳(即尺度步长 Δs), $I_{t_i}^s$ 表示在扩散尺度 $s\tau$ 上的 $I(x, y, t_i)$ 的逼近, 表达式 $\sum_{j \in \Omega_{t_i}^s} A_{j, t_i}^s I_{t_i}^s$ 是 $\text{div}(\mathbf{D} \cdot \nabla I_{\text{sp}})$ 的离散化形式, $A_{t_i}^s$ 是空间域上一个可变的模板。由于时间上平滑仅受前 1 帧的影响, 因此在第 t_i 帧上 $\nabla I_{t_i}^s$ 为

$$\nabla I_{t_i}^s = I(x, y, t_{i-1}) - I(x, y, t_i, s) \quad (17)$$

3 实验与分析

为了验证本文算法的有效性, 在两大公开的人体动作数据集 KTH^[8] 和 YouTube^[9] 上进行了评价, 这两个数据集都是户外现实场景下的人体动作视频, 常被用于评价人体动作的识别性能, 并与现有的时空兴趣点检测方法(Dollar 的方法^[2] 和 Shabani 的方法^[6])进行了比较和分析。

3.1 实验数据集和实验设置

KTH 数据集^[8] 的背景相对简单, 包含轻微摄像机晃动/移动下的 25 个表演者在 4 种不同场景(室内、户外、衣着变化和尺度变化)中的 6 个人体动作(boxing, hand clapping, hand waving, jogging, running, walking)视频, 从中选取了一个有轻微摄像机晃动的“walking”动作视频作为比较样例。YouTube 数据集^[9] 则是完全不受控制的现实环境下的视频, 包含了快速摄像机运动(如骑自行车、骑马和走路的视频)、背景复杂、各种不同的体型大小、表观、姿态、光照变化、视角变化、低分辨率等各种影响, 总共 11 类人体动作, 从中选取了户外场景下快速摄像机移动的

“riding”和“biking”动作视频为评价样例。

实验设置为: 在偏微分方程式(8)的实现中, 采用有限前向差分的方法, 时间间隔设置为 $\Delta s = 0.2$, $\lambda_{\text{sp}} = 5$ 或者 10, $\lambda_{\text{tp}} = 10$, 结构张量中高斯滤波器 G_σ 的方差 $\sigma = 4$ 。由于暂时还没有考虑时空尺度上的兴趣点检测, 所以仅采用了一个时空尺度, KTH 数据集上的时空窗口的尺度分别设置为 $\sigma_x = \sigma_y = 7$ 和 $\sigma_t = 13$, YouTube 数据集中的尺度设置为 $\sigma_x = \sigma_y = 13$ 和 $\sigma_t = 25$ 。

3.2 摄像机轻微晃动下的时空兴趣点检测方法比较

摄像机轻微晃动下的兴趣点检测主要验证了本文方法对噪声、弱纹理区域和弱摄像机晃动的鲁棒性, 并与 Dollar 方法^[2] 和 Shabani 方法^[6] 进行了比较。实验在相对简单的 KTH 数据上进行, 选择摄像机轻微晃动下的“walking”人体动作视频作为样例。图 2 给出了该环境下时空兴趣点的检测结果。

从图 2(a)中可以看出 Dollar 方法确实对摄像机晃动很敏感, 而且产生的兴趣点比较稀疏; 摄像机晃动导致了一些轻微的背景运动, 因而就有一些兴趣点位于图像的背景区域, 这可能是由于时间域上的 Gabor 滤波器对噪声和纹理比较敏感, 所以很容易受到摄像机晃动/移动的影响。从图 2(b)和图 2(c)可以看出, Shabani 方法和本文方法获得了比 Dollar 的方法更好的检测效果, 检测到的兴趣点基本都位于人体运动部位, 没有兴趣点在摄像机晃动下的背景区域(见图 2 最后两行)。这证明非线性各向异性扩散滤波确实能够对噪声和弱纹理进行很好的平滑, 而且保持了图像结构边缘。该实验结果也表明 Shabani 方法和本文方法都对摄像机晃动/移动具有很好的鲁棒性。

3.3 杂乱背景和快速摄像机运动下的时空兴趣点检测方法比较

该实验评价了 Dollar 方法、Shabani 方法和本文方法在杂乱背景和快速摄像机移动下的性能。实验样例为复杂场景下 YouTube 数据集中的“biking”和“riding”人体动作, 这两个人体动作由于移动速度快, 存在各种速度的摄像机晃动或移动, 导致了各种不同速度的背景和人体运动, 而且背景杂乱, 有房屋、树、电线杆、围栏、斑马线等的干扰。图 3 给出了“biking”人体动作中 3 种方法的时空兴趣点检测结果, 图 4 给出了“riding”人体动作中 3 种方法的时空兴趣点检测结果。

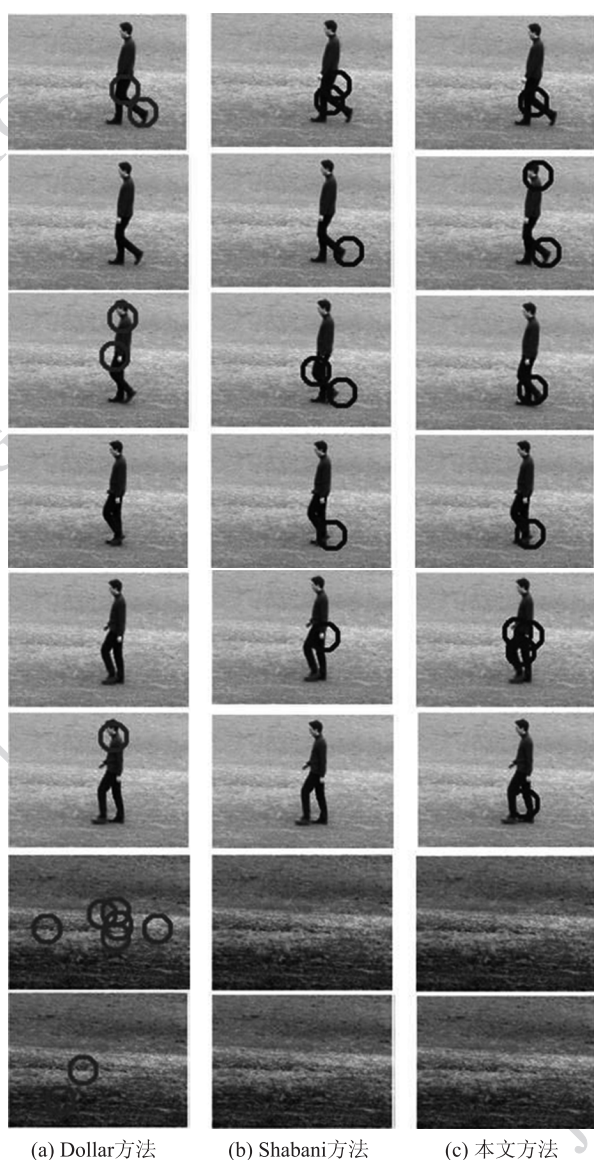


图2 摄像机轻微晃动下检测结果比较

Fig. 2 The comparison of detection results in the camera shaking slightly

从“biking”和“riding”两个人体动作的实验结果可以看出, Dollar 方法对摄像机移动很敏感, 很容易检测到由于摄像机晃动导致的背景运动点, 特别容易检测到线型的结构点, 见图 3(a) 的第 1、2 和最后 1 行所示, 检测的所有点都在背景中的高亮度直线上, 相比较而言, Shabani 方法和本文方法对背景中的细小结构具有更好的平滑性, 大部分兴趣点在结构区域比较大的地方, 这表明 Shabani 方法和本文方法比 Dollar 方法具有更好的鲁棒性, 获得了更好的检测效果。相比较 Shabani 方法而言, 本文方法获得了更好的检测效果, 检测到的兴趣点很多位于人体运动所在的区域(如人、自行车和马运动的



图3 “biking”人体动作的检测结果比较

Fig. 3 The comparison of detection results in “biking” action

部位), 能够捕捉到人体动作的特征, 这也表明了本文方法能够对快速摄像机移动进行更好地建模。

3.4 计算复杂度分析

该实验比较分析了 Dollar、Shabani 和本文方法的计算复杂度, 在 Genuine Intel T2050 1.60 GHz 双核 CPU 和 2 GB 内存的机器上运行, 采用了 YouTube 数据集中的人体动作视频序列, 总共为 1 000 帧, 图像大小为 320×240 , 分别比较了扩散尺度为 $1 \sim 30$ 时各种方法的平均每秒计算的帧数。由于 Dollar 方法没有扩散过程, 而是高斯平滑, 因此在每一个扩散尺度上, Dollar 方法仅用与扩散尺度相同的尺度进行一次高斯平滑。图 5 给出了对比结果, 当扩散尺度为 1 时, Shabani 方法的计算速度最快, 为 11.1 帧/s, Dollar 方法其次, 为 7.4 帧/s, 本文方法最慢, 为 4.8 帧/s; 当扩散尺度为 2 以上时, Dollar 方法计算速度最快, 其次为 Shabani 方法, 本文方法最慢。该结果表明, Dollar 方法的计算复杂度最低, 其次为 Shabani 方法, 然后是本文方法, 这是由于本文方法根据图像几何结构进行不同程度的扩散, 扩散张量的计算耗费了更多的时间, 但在杂乱背景和摄像机移动场景下获得最好的检测结果。然而 Dollar 方法不能进行多尺度表示, 若实现多尺度表示的话, 则计算复杂度远高于 Shabani 方法和本文方法, 而 Shabani



图4 “riding”人体动作的检测结果比较

Fig. 4 The comparison of detection results in “riding” action

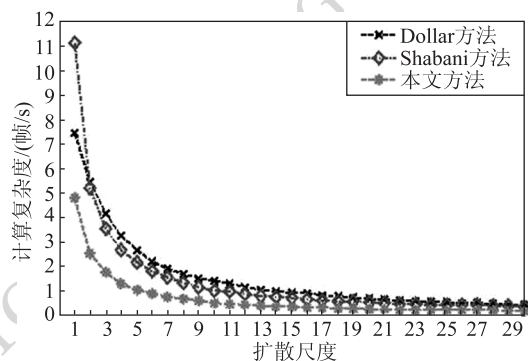


图5 计算复杂度比较

Fig. 5 The comparison of computational complexity

方法和本文方法的扩散过程实质已实现了多尺度表示。

4 结 论

时空兴趣点检测在基于局部表示的人体动作识别方法中起着关键的作用。本文主要研究了杂乱背

景和摄像机移动场景下的时空兴趣点检测问题,特别是针对摄像机晃动/移动幅度比较大的情况下的时空兴趣点检测。因为现有的检测方法大多不能适用于复杂的动态场景下的兴趣点检测,特别是摄像机晃动/移动幅度比较大的情况下,很多兴趣点基本上都在背景物体上,这是因为很多方法是对整个时空域进行滤波,没有对摄像机晃动/移动所导致的时间关系上的不确定性建模。针对这些问题,结合空间域上的平滑和时间域上的平滑,提出了一种新的时空兴趣点检测方法。空间域上的平滑是多向的,采用了 Weickert 提出的边缘增强的非线性各向异性扩散滤波器,保持和增强了边缘信息,同时又平滑掉噪声和图像结构内部。时间域上的平滑考虑了摄像机移动晃动/移动所导致的时间关系上的不确定性,当前帧仅仅受前 1 帧的影响,而与后面的帧和间接前面的帧是无关的,扩散是单向的,时间上的平滑仅和直接的前 1 帧有关,采用了与梯度成正比的抑制函数,来抑制摄像机晃动/移动所导致的影响。

最后在复杂的现实场景下的人体动作视频 YouTube 数据集上进行了验证,实验结果表明采用 Weickert 提出的边缘增强的非线性各向异性扩散滤波不但能够平滑掉人体区域内的常量区域和一些背景噪声,而且能够使得兴趣点位于突出的人体轮廓边缘上,通过结合时域上的抑制函数,能够平滑掉摄像机移动导致的一些线状或细区域的运动,使得兴趣点大部分还是位于人体动作区域上。与现有的 Dollar 方法和 Shabani 方法相比较,本文方法在杂乱背景和摄像机晃动/移动场景下获得了更好的检测结果。

参考文献 (References)

- [1] Laptev I, Lindeberg T. Space-time interest points[C]//Proceedings of IEEE International Conference on Computer Vision. Nice, France: IEEE Computer Society, 2003: 432-439.
- [2] Dollár P, Rabaud V, Cottrell G, et al. Behavior recognition via sparse spatio-temporal features[C]// Proceedings of the International Workshop on Visual Surveillance and Performance Evaluation of Tracking and Surveillance. Beijing, China: IEEE Computer Society, 2005: 65-72.
- [3] Oikonomopoulos A, Patras I, Pantic M. Spatiotemporal salient points for visual recognition of human actions[J]. IEEE Transactions on Systems, Man, and Cybernetics (SMC)-Part B: Cybernetics, 2006, 36(3): 710-719.
- [4] Willems G, Tuytelaars T, Luc J, et al. An efficient dense and

- scale-invariant spatio-temporal interest point detector [C]//Proceedings of the European Conference on Computer Vision. Marseille, France; Springer, 2008: 650-663.
- [5] Wang H, Ullah M M, Kläser A, et al. Evaluation of local spatio-temporal features for action recognition [C]//Proceedings of the British Machine Vision Conference. London, UK; British Machine Vision Association, 2009:1-11.
- [6] Shabani H, Clausi D A, Zelek J S. Towards a robust spatio-temporal interest point detection for human action recognition [C]//Proceedings of the Conference on Computer and Robot Vision. Kelowna, British Columbia, Canada; IEEE Computer Society, 2009: 237-243.
- [7] Weickert J. Anisotropic Diffusion in Image Processing [M]. Stuttgart; B. G. Teubner, 1998.
- [8] Schödl C, Laptev I, Caputo B. Recognizing human actions: a local SVM approach [C]//Proceedings of the International Conference on Pattern Recognition. Cambridge, UK; IEEE Computer Society, 2004: 32-36.
- [9] Liu J, Luo J, Shah M. Recognizing realistic actions from videos "in the wild" [C]//Proceedings of the Conference on Computer Vision and Pattern Recognition. Miami, Florida, USA; IEEE, 2009: 1996-2003.