航拍视频中运动目标检测算法研究

唐佳林1 郑杰锋1 李熙莹2,3 苏秉华1

(北京理工大学珠海学院信息学院 珠海 519088)¹ (中山大学工学院智能交通研究中心 广州 510006)² (广东省智能交通系统重点实验室 广州 510006)³

摘 要 针对复杂背景下航拍视频中的运动目标检测问题,提出一种基于改进的特征匹配算法与全局运动补偿的防抖方法,以及结合多帧能量累积的运动目标检测算法。首先,采取局部区域匹配法加快该算法的处理速度,避免运动目标对背景补偿的影响;其次,利用尺度不变的 SURF 算法,结合快速近似最邻近搜索算法得到匹配点对,并通过双向匹配和 K-近邻算法筛选优秀匹配点;然后,建立仿射变换模型,求解运动参数,并进行运动补偿;最后,通过多帧能量累积进行目标检测。仿真结果表明,该方法具有良好的运动目标检测效果。

关键词 航拍视频,特征匹配,快速近似最邻近搜索算法,运动补偿,运动目标检测

中图法分类号 TP391.41 文献标识码 A

Research on Detecting Algorithm of Moving Target in Aerial Video

TANG Jia-lin¹ ZHENG Jie-feng¹ LI Xi-ying^{2,3} SU Bing-hua¹

(School of Information Technology, Beijing Institute of Technology, Zhuhai 519088, China)¹

(Research Center of Intelligent Transportation System, School of Engineering, Sun Yat-sen University, Guangzhou 510006, China)²
(Guangdong Provincial Key Laboratory of Intelligent Transportation System, Guangzhou 510006, China)³

Abstract A detecting algorithm of moving target based on the combination between multiple frame energy accumulation and a stabilization method which is combined between improved feature matching algorithm and global motion compensation according to the moving target detection in the aerial video under complex background was proposed. Firstly, it uses the matching method of local area to increase the processing speed of the algorithm, and avoids the influence of the moving target on the background compensation. Secondly, it gets the matching points using the scale-invariant SURF algorithm in combination with the fast approximate nearest neighbor search method, and screens the superior matching points with bilateral matching and k-nearest Neighbor algorithm. Thirdly, it builds the affine transformation model, solves the motion parameters, and makes motion compensation. Finally, it detects moving target through multiple frame difference accumulation. The simulation results show that the method has good effect on moving target detection.

Keywords Aerial video, Feature matching, Fast approximate nearest neighbor search algorithm, Motion compensation, Moving target detection

运动目标检测是计算机视觉领域的重要内容,摄像系统有固定平台和移动平台两种情况。在固定平台下,背景处于静止状态,对运动目标进行检测时比较简单,在国内已经取得了较好的研究成果。而对于移动平台,摄像机的抖动性使得场景内存在两个运动矢量,且航拍视频拍摄的场景较为复杂,对运动目标的检测较为困难,因此其成为了国内外的研究热点[1-4]。

运动背景下的目标检测主要是利用运动参数估计和运动补偿的方法检测目标^[5],目前已提出利用 Harris 角点和 SU-SAN 角点进行匹配,再进行运动估计,进而进行图片配准^[6-7]。文献[8]通过提取 SIFT 特征点进行匹配,进而对运动目标进行检测。Harris 和 SUSAN 只能对较为稳定的视频

图像进行匹配,不宜检测有复杂背景下的运动目标。SIFT 算法具有良好的性能,能适应多种复杂情况,但其计算量大,不具有实用性^[9-15]。SURF 算法是 SIFT 算法的改进,其处理速度快于 SIFT 算法,但仍然存在计算量大、匹配精度不高等缺点。因此本文提出了一种基于局部特征匹配的运动补偿与多帧能量累积相结合的运动目标检测算法。实验结果表明,该方法具有良好的检测效果。

1 基于改进的局部区域特征匹配与运动补偿的防 抖算法

1.1 SURF 算法

Hessian 矩阵是 SURF 算法的核心,像素点矩阵如下:

本文受广东省科技计划项目(2013B051000044),广东大学生科技创新培育专项资金(pdjh2017b0927),广东省青年创新人才项目(2016KQNCX204)资助。

唐佳林(1982一),男,硕士,讲师,主要研究方向为人工智能与模式识别,E-mail;thong03@qq.com;**郑杰锋**(1995一),男,主要研究方向为人工智能与模式识别,E-mail;1210423695@qq.com(通信作者);**李熙莹**(1972一),女,博士,副教授,硕士生导师,主要研究方向为图像处理、视频图像交通信息获取技术研究及应用;**苏秉华**(1962一),男,博士,教授,硕士生导师,主要研究方向为光电技术、数字图像处理、机器视觉。

$$H(f(x,y)) = \begin{bmatrix} \frac{\partial^2 f}{\partial x^2} & \frac{\partial^2 f}{\partial x \partial y} \\ \frac{\partial^2 f}{\partial x \partial y} & \frac{\partial^2 f}{\partial y^2} \end{bmatrix}$$
(1)

设某图像的点为 I(x,y),尺度 σ 的 Hessian 矩阵为:

$$H(x,\sigma) = \begin{bmatrix} L_{xx}(x,\sigma) & L_{xy}(x,\sigma) \\ L_{xy}(x,\sigma) & L_{yy}(x,\sigma) \end{bmatrix}$$
 (2)

其中,

$$L(x,t) = G(t) \cdot I(x,t) \tag{3}$$

高斯核 G(t)为:

$$G(t) = \frac{\partial^2 g(t)}{\partial x^2} \tag{4}$$

采用近似值代替 L(x,t), Hessian 判别式为:

$$\det(H) = D_{xx}D_{yy} - (0.9D_{xy})^2$$
 (5) 可用 $\det(H)$ 来判别极值点。

通过上式可以确定极值点,利用非极大值抑制确定特征点,再选取特征点方向,构造 SURF 特征点描述符。

1.2 选取局部匹配区域

对全局的特征点进行提取时存在计算量大的缺点,由于拍摄的场景内有大量的运动物体,如果进行全局特征点匹配将导致对摄像机移动参数的错误估计,无法求取正确的变换参数,从而无法进行运动补偿。针对该问题,本文提出局部区域特征匹配。大部分情况下,运动物体出现在中间区域,为避免选取到运动物体,本文将区域设定在4个角落。摄像机的抖动将导致4个边角区域信息发生变化。综上所述,本文选取场景内4个角落区域向内缩进一定距离的4个矩形框,如图1所示。

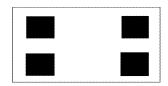


图 1 区域图

1.3 SURF 结合 FLANN 的快速近似最邻近算法

当前提出的许多近似最近邻搜索算法都以牺牲一定精度来加快速度,在 FLANN 算法中分析了大量的近似最近邻搜索算法,对于高维空间中的最邻近搜索问题,选取了性能最好的 K-means tree 分层和随机 KD-tree。

FLANN实现了根据使用者输入的数据情况自动确定最佳算法和最优参数值。本文引入FLANN结合 SURF进行快速高效匹配。本文对匹配过程加以改进和优化,特征点的匹配流程如下。

- 1)首先使用 SURF 提取关键点,计算特征描述符。
- 2)为了加快检测速度,先训练一个匹配器,根据数据选取 最优的索引类型随机 KD-tree 树和层次 K-means tree 为特征 描述子建立索引树。
- 3)计算两帧图像对应特征点的欧氏距离,然后选取最小的距离 l,设定一个阈值 $L=u \times l$,设 u=2。当对应特征点距离小于 L 时,保留匹配点,否则剔除。
- 4)将两帧图像进行双向匹配,重复步骤 3),若匹配的特征 点对与上一步骤得到的结果一样,此时的特征点对才是最终得 到的匹配点。

为了得到优秀的匹配点,本文引入 K-近邻算法。K-近邻

算法对每个描述子寻找 K 个最邻近的匹配点,当查询描述子的总数少于 K 时,取总数量,设 K=2。

其中两个最邻近的点分别是最优匹配点和次优匹配点。 当最优点的测量距离非常小(远远小于次优点)时,保留最优 匹配点。当两个匹配点的距离比较相近时,该最优匹配点可 能出错,因此将其剔除。最终得到的匹配图如图 2 所示。



图 2 匹配图

1.4 匹配算法的比较

将改进前后的匹配算法进行比较,并且与另一种改进的 SURF 算法进行对比。匹配算法的比较结果如表 1 所列,本 文算法通过有效筛选,确保了匹配点的正确性。同时对算法 进行优化,以及通过局部区域进行匹配,使得处理时间大大减 少,确保了算法的实时性与有效性。仿真结果表明,本文改进 的匹配算法具有优越性。

表 1 匹配算法的比较

算法	处理后的 匹配点数	匹配 正确率/%	匹配 时间/ms
SURF 算法	600	78.0	1880
改进的 SURF 匹配法	400	93.0	550
本文算法	350	98.8	320

1.5 变换参数求解与运动补偿

根据仿射变换模型,假设前一帧图像和后一帧图像对应的像素点位置分别为 $X_n = \begin{bmatrix} x_n \\ y_n \end{bmatrix}$ 和 $X_{n+1} = \begin{bmatrix} x_{n+1} \\ y_{n+1} \end{bmatrix}$,通常用

2*3的矩阵来表示仿射变换。设 $A=\begin{bmatrix}a_1&a_2\\a_3&a_4\end{bmatrix},B=\begin{bmatrix}b_1\\b_2\end{bmatrix}$,则

 X_n 和 X_{n+1} 为:

$$X_{n+1} = A \cdot X_n + B \tag{6}$$

其中,A表示水平和垂直尺度,B为平移参数。式(6)为:

$$\begin{bmatrix} x_{n+1} \\ y_{n+1} \end{bmatrix} = \begin{bmatrix} a_1 x_n + a_2 y_n + b_1 \\ a_3 x_n + a_4 y_n + b_2 \end{bmatrix}$$
 (7)

基于上述方法,可以通过相邻帧图像来求解变换参数。由于有多个特征匹配点,因此本文采用最小二乘法来求解最优解,再采用双线性插值法来对当前帧进行全局运动补偿。

2 基于多帧能量累积的运动目标检测

对配准后的视频帧进行处理,通过对两帧图像进行差分 可以分割出运动目标,但效果不佳,因此本文使用多帧差分能 量累积的方法。

一般而言,通过上述方法得到的差分图还会存在噪点,并 且分割的运动物体还存在黑洞等情况,因此本文对多帧能量 累积后的图片做进一步的处理。首先进行均值滤波,对滤波 后的图像再进行二值化处理。

基于多帧差分能量累积与后处理进行目标分割的方法流程图如图 3 所示。

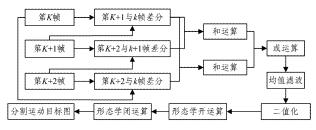


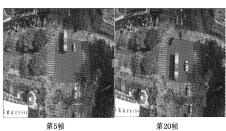
图 3 目标分割流程

3 实验结果及分析

为验证本文算法的性能,在 VS2015 环境下实现本文的 算法,采用大小为 1080 * 1920 的航拍视频进行验证。

3.1 防抖算法效果

使用本文算法对航拍视频进行处理,截取视频中对应的 多帧图像,结果如图 4 所示。以左下角楼房为参考,可以看出 防抖前后的效果。4 帧稳像图像如图 5 所示。



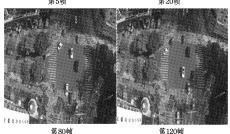
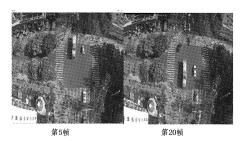


图 4 4 帧原始图像



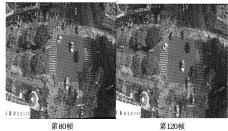


图 5 4 帧稳像图像

为了更直观地显示消抖效果,本文将图像进行了差分。 如图 6 所示,航拍视频的抖动使得未处理的视频帧与前一帧的差分图不仅有运动车辆、行人、光照等自身背景的变化,还造成了场景内的大部分区域发生变化。而通过本文的视频防抖算法进行处理后,由视频抖动造成的影响则大大减 小,只存在车辆和行人等运动物体,如图 7 所示。





图 6 原始视频分割

图 7 补偿后的目标分割

3.2 目标检测效果

由于图 7 分割出的运动目标不够清晰,检测效果达不到 要求,因此通过基于多帧能量累积的运动目标检测方法进行 运动目标累积,结果如图 8 所示。由图 8 可知,该方法对运动 目标检测具有很好的效果。

由于图像中仍存在噪点且运动目标存在黑洞等情况,因此根据图 3 所示流程对图像做进一步处理,结果如图 10 所示。本文算法可以较好地检测出视频中的运动车辆和行人,且测试表明本文算法有较好的实时性。



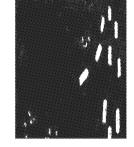


图 8 多帧累积目标分割图

图 9 多帧累积目标分割与 后处理结果图

结束语 本文研究了航拍视频中运动目标的检测,提出一种基于改进的特征匹配算法与全局运动补偿的防抖方法,以及结合多帧能量累积的运动目标检测算法。通过本文方法提高了匹配速率及正确率,解决了飞行器抖动问题,很好地检测出了运动目标,具有较好的实时性和有效性。未来可将此方法应用于智能交通或者侦察中的目标检测与跟踪。

参考文献

- [1] LIU S, XU B, DENG C, et al. A Hybrid Approach for Near-Range Video Stabilization[J]. IEEE Transactions on Circuits & Systems for Video Technology, 2016(99):1.
- [2] YAN F,ILIYASU A M,KHAN A R,et al. Measurements-based Moving Target Detection in Quantum Video[J]. International Journal of Theoretical Physics, 2015, 55(4): 1-12.
- [3] KOH Y J, LEE C, KIM C. Video Stabilization Based on Feature Trajectory Augmentation and Selection and Robust Mesh Grid Warping [J]. IEEE Transactions on Image Processing, 2015, 24(12);5260-5273.
- [4] KOH Y J, LEE C, KIM C. Video Stabilization Based on Feature Trajectory Augmentation and Selection and Robust Mesh Grid Warping[J]. IEEE Transactions on Image Processing, 2015, 24 (12):5260-5273.

(下转第 183 页)

表 3 不同手语者的手语识别结果

_	手语者编号	1	2	3	4	5
	识别准确率	0.9933	0.9777	0.9400	0,8067	0, 7500
	识别所用时间/ ms	985	1073	1129	1079	1089

从整体来说,本文设计的算法的可行性较高,1 秒内即可得到识别结果,实时性可以得到保障。此外,Kinect 本身对使用环境没有特殊要求,仅要求手语者与镜头之间保持0.8~1米的距离,且之间无任何障碍物即可。因此本系统在使用时不受外界环境、光照等影响。

结束语 本文提出了一种对连续的手语语句进行有效识别的新方法,即利用手语者的心理暗示,根据手语轨迹提取关键帧,利用关键帧描述手语词汇,再利用词汇描述连续句子,同时利用关键帧的权值提高识别的准确率。该算法对于不同的手语者(无论是初学还是相对较熟练的人)均取得了较好的效果。该算法的实现依赖于关键帧的提取与识别,因此为了更为稳定地提取关键帧,除了需考虑点密度的变化,还应该加人对手型变化的考虑。此外,在提取关键帧的基础上获得准确、稳定的手语识别结果,依赖于对关键帧的正确识别与分类,因此对关键手型的特征提取以及分类器设计提出了更高的要求。这些都是今后需要进一步深人研究的问题。

参考文献

- [1] STARNER T, PENTLAND A. Real-Time American Sign Language Recognition Using Desk and Wearable Computer Based Video[C]//IEEE Transaction on Pattern Analysis and Machine Intelligence, 1998;1371-1375.
- [2] VON AGRIS U, ZIEREN J, CANZLER U, et al. Recent developments in visual sign language recognition [J]. Universal Access in the Information Society, 2008, 6(4): 323-362.
- [3] GROBEL K, ASSAN M, Isolated sign language recognition using hidden Markov models [C] // IEEE International Conference on Systems, 1997; 162-167.
- [4] RAVIKIRAN J, MAHESH K, MAHISHI S, et al. Finger Detection for Sign Language Recognition [J]. International Association of Engineers, 2009, 2174(1):489-493.
- [5] GUO X L, YANG T T. Gesture recognition based on HMM-FNN model using a Kinect[J]. Journal on Multimodal User Interfaces, 2016; 1-7.

- [6] 谈家谱,徐文胜. 基于 Kinect 的指尖检测与手势识别方法[J]. 计算机应用,2015,35(6):1795-1800.
- [7] HALIM Z, ABBAS G, A Kinect-Based Sign Language Hand Gesture Recognition System for Hearing- and Speech-Impaired: A Pilot Study of Pakistani Sign Language[J]. Assistive Technology, 2015, 27(1):34-43.
- [8] YAN H,ZHANG M,TONG J, et al. Real time robust multi-fin gertips tracking in 3D space using Kinect[J]. Journal of Computer-Aided Design and Computer Graphics, 2013, 25 (12): 1801-1809.
- [9] JIANG F, GAO W, WANG C L, et al. Development in Signer-Independent Sign Language Recognition and the Ideas of Solving Some Key Problems[J]. Journal of Software, 2007, 18(3):477-489.
- [10] NASRI S, BEHRAD A, RAZZAZI F. A novel approach for dynamic hand gesture recognition using contour-based similarity images [J]. International Journal of Computer Mathematics, 2015, 92(4):662-685.
- [11] LIN Y,CHAI X,ZHOU Y,et al. Curve Matching from the View of Manifold for Sign Language Recognition[C]//Workshop on Feature and Similarity Learning for Computer Vision (FSL-CV), ACCV, 2014; 233-246.
- [12] PU J,ZHOU W,ZHANG J,et al. Sign Language Recognition Based on Trajectory Modeling with HMMs[C]//International Conference on Multimedia Modeling, 2016;686-697.
- [13] STARNER T,PENTL A. Visual Recognition of American Sign Language Using Hidden Markov Models [C] // International Workshop on Automatic Face & Gesture Recognition, 1995; 189-194.
- [14] 方高林,高文,陈熙霖,等. 基于 SRN/HMM 的非特定人连续手语识别系统[J]. 软件学报,2002,13(11):2169-2175.
- [15] 中国残疾人联合会教育就业部,中国聋人协会.中国手语[M]. 华夏出版社,2003.
- [16] LISR, HUANGYY, HUZJ, et al. Key Frame Detection Algorithm based on Dynamic Sign Language Video for the Non Specific Population[J]. International Journal of Signal Processing: Image Processing and Pattern Recognition, 2015, 8 (12): 135-148.
- [17] SHI M M, HUANG Y Y, HU Z J. Dynamic Sign Language Recognition Algorithm Using Weighted Gesture Units[J]. Journal of Information and Computational Science, 2015, 12(15): 5611-5621.

(上接第 177 页)

- [5] 刘威,赵文杰,李成,等. 基于改进 ORB 特征匹配的运动小目标 检测[J]. 光电工程,2015,42(10):13-20.
- [6] LIJ, XUT, ZHANG K. Real-Time Feature-Based Video Stabilization on FPGA[J]. IEEE Transactions on Circuits & Systems for Video Technology, 2017, 27(4): 907-919.
- [7] LIANG W, XIE X, WANG J, et al. A SIFT-based mean shift algorithm for moving vehicle tracking [C] // Intelligent Vehicles Symposium Proceedings, IEEE, 2014; 762-767.
- [8] 程德强,郭政,刘洁,等. —种基于改进光流法的电子稳像算法 [J]. 煤炭学院,2015,40(3):707-712.
- [9] KOH Y J, LEE C, KIM C. Video Stabilization Based on Feature Trajectory Augmentation and Selection and Robust Mesh Grid Warping [J]. IEEE Transactions on Image Processing, 2015, 24(12):5260-5273.
- [10] XUE X, CHEN Z, WANG Q, et al. Moving target detection of

- OFDM-MIMO radar based on ST-DFT[J]. Electronics Letters, 2015,52(2):150-152.
- [11] K M,CHANNAPPAYYA S S. An Optical Flow-Based Full Reference Video Quality Assessment Algorithm[J]. IEEE Transactions on Image Processing, 2016, 25(6):2480-2492.
- [12] GU B, HU H, REN Y, et al. Moving Target Detection and Tracking in Complex Background [J]. International Journal of Smart Home, 2015, 9(9):95-102.
- [13] SILVA G, RODRIGUEZ P. Jitter invariant incremental principal component pursuit for video background modeling on the TK1 [C]// Asilomar Conference on Signals, Systems and Computers. IEEE, 2015.
- [14] 闫利,陈林. 一种改进的 SURF 及其在遥感影像匹配中的应用 [J]. 武汉大学学报(信息科学版),2013,38(7):770-773.
- [15] 程德强,郭政,刘洁,等. 一种基于改进光流法的电子稳像算法 [J]. 煤炭学报,2015,40(3):707-712.