

文章编号:1671-6833(2024)05-0077-09

基于相机运动轨迹的鲁棒无人机航拍稳像算法

余松森, 龙嘉豪, 周 诺, 梁 军

(华南师范大学 软件学院, 广东 佛山 528225)

摘 要:针对高空颠簸环境影响无人机采集延时稳定图像问题,提出了一种适用于悬停拍摄及移动拍摄的航拍视频防抖算法。通过无人机摄像头获取延时摄影视频,提取全局范围内的部分视频帧进行直方图分布比较,进而确定视频是否是在无人机移动状态下拍摄的,并将视频划分为包含与不包含摄像机主动运动两类。对于包含摄像机主动运动的一类,首先,采用 FAST 角点检测加光流法进行特征点提取配对;其次,利用 RANSAC 算法剔除误匹配特征点并进行摄像机运动轨迹估计;最后,利用高斯滤波对运动估计参数进行平滑以得到稳定的摄像机运动轨迹。对于不包含摄像机主动运动的一类,首先,对首帧进行网格划分并基于 Harris 矩阵提取各网格的特征点;其次,在后续帧对这些特征点进行光流追踪;再次,通过反向光流及 Harris 矩阵计算,增加特征点约束,完成特征点提取及匹配;最后,利用保留的特征点估计后续帧至首帧的稳定变换。利用该算法对视频进行场景分类及画面稳定,实验结果表明:视频分类模块能正确区分两类视频;对比其他两种方法,稳定后的视频图像平均峰值信噪比提升更大;对于不含摄像机主动运动的视频,可实现画面绝对稳定,图像平均峰值信噪比提升超过 39%,而其他两种方法仅提升 10%~12%。

关键词:无人机; 视频稳定; 特征提取; 运动估计; 峰值信噪比

中图分类号:TP391

文献标志码:A

doi:10.13705/j.issn.1671-6833.2024.02.004

近年来随着无人机市场的蓬勃发展,使用消费级无人机的用户人数也愈来愈多,而消费级无人机的一个重要功能——延时摄影,尤其受到用户青睐。在摄像头固定状态下,延时摄影可分为两种:一种是在无人机悬停时以高阔、宽广的视野进行长视频拍摄;另一种则是在无人机飞行时进行长视频拍摄,拍摄结束后延时摄影功能将输出高帧率视频,达到每秒播放数秒原视频画面的效果。然而在无人机延时摄影中,由于每秒画面帧数的增加,会凸显画面抖动,进而导致视频抖动的问题,解决该问题需要出色的电子防抖技术,这是本文研究的重点。电子防抖可以减少视频中画面抖动,有效提高视频质量;同时大量的视觉处理任务如目标检测、目标跟踪等^[1]都需要依靠该技术获得稳定的画面。

传统的电子防抖方法可以按处理的维度分为二维运动方法和三维运动方法。其中,二维运动方法利用跟踪的特征进行二维变换估计(如仿射变换、

单应变换),并将这些变换应用到视频帧中合成稳定视频;而三维运动方法则获取三维信息,在三维空间中利用运动结构或附加传感器(例如惯性测量单元和深度传感器)对帧运动进行建模。虽然基于三维的方法效果更佳,但需要额外的器件来获取图像深度信息,因此会增加无人机重量和硬件成本,不利于轻量化无人机量产。相比之下,基于二维的方法具有计算快、鲁棒性高、应用成本低等优点,更加适用于无人机航拍视频防抖,因此,本文研究的是基于二维方法的视频稳定算法。

此前已有一些基于二维运动模型的方法表现出了不错的视频稳定质量,Grundmann 等^[2]应用电影拍摄思想,模拟专业摄影师使用的摄像机运动,提出了一种自动定向最优 L1 优化的摄像机路径,将摄像机路径定义为由恒定线段、线性线段和抛物线段组成的长线段,获得了更为可靠的摄像机运动路径。Lee 等^[3]则直接估计帧间的鲁棒特征运动轨迹并对

收稿日期:2023-10-20;**修订日期:**2023-11-29

基金项目:广东省基础与应用基础研究基金区域联合基金项目(2022A1515140110);广东省基础与应用基础研究基金区域联合基金项目(重点项目)(2020B1515120089);国家自然科学基金资助项目(61172156)

通信作者:梁军(1983—),男,江西高安人,华南师范大学讲师,博士,主要从事图论、机器学习、算法设计方面的研究, E-mail: liangjun@m.scnu.edu.cn。

引用本文:余松森,龙嘉豪,周诺,等.基于相机运动轨迹的鲁棒无人机航拍稳像算法[J].郑州大学学报(工学版),2024,45(5):77-85.(YU S S, LONG J H, ZHOU N, et al. Robust algorithm for aerial video stabilization of UAV based on camera motion trajectory[J]. Journal of Zhengzhou University (Engineering Science), 2024, 45(5): 77-85.)

其进行拼接、修剪,而不去估计相机路径,有效避免了运动模型的不足及运动分类不准确等问题,但该方法会受到前景运动物体的干扰。上述方法均不能很好地解决无人机航拍视频中摄像机运动轨迹集中在某一圆形区域的情况。

针对无人机航拍视频,唐佳林等^[4]假定前景运动物体位于图像中心,提出了基于特征匹配与运动补偿的视频稳像算法,该算法选取场景内 4 个角落区域向内缩进一定距离的 4 个矩形框作为特征匹配区域,由此来避免前景运动物体对帧间变换矩阵估计的影响。对于前景运动物体带来的问题,谢亚晋等^[5]提出一种电子稳像方法,基于最小生成树提取图像背景特征点,避免因提取到前景运动物体而造成错误的摄像机运动估计。然后采用改进的卡尔曼滤波法,引入两个卡尔曼滤波器对摄像机的运动进行分离,其中一个滤波器去除摄像机随机运动带来的抖动噪声,而另一个滤波器则保留摄像机的主动运动,从而实现摄像机全局运动轨迹的平滑操作。唐佳林等^[4]提出的方法能较好处理摄像机运动轨迹集中在某一圆形区域的情况,但不能很好地处理摄像机运动轨迹为曲线或直线的情况,谢亚晋等^[5]提出的方法则与之相反。为了保证特征点的均匀分布,使帧间运动估计更为准确,王杰等^[6]改进了 Harris 角点的提取方式,将每帧图像进行等比例分割,分割成同等尺寸子区域,然后根据不同子区域的灰度均方差设置影响角点选择的阈值,解决了局部角点过于聚簇的问题,但该方法与谢亚晋等^[5]提出的方法一样,无法很好地处理摄像机运动轨迹集中在某一圆形区域的情况。

此外,一些基于学习的方法^[7-9]利用彩色视频隐式地学习帧间运动、利用光流进行运动分析并学习帧稳定;基于三维信息的方法^[10-13]利用多传感器如陀螺仪、深度相机等获取其他维度的信息辅助三维重建及轨迹估计;基于深度学习的方法^[14-18]在特征空间中学习帧稳定的信息,这些视

频稳定的算法有着不俗的表现。然而,上述这些方法大多以视频为手持移动拍摄作为前提,并没有考虑到相机抖动是围绕同一空间位置产生的,因而鲁棒性受限。

综上所述,现有的方法往往只能对单一类别相机运动轨迹的视频进行防抖处理。因此,本文提出了一种综合处理方法,能满足不同的运动轨迹的视频的防抖需求。通过视频自处理分类技术,将不同轨迹的视频自动分为两类进行处理,一类为包含摄像机主动运动(无人机在遥控指令下飞行带来的摄像头位移)的视频,另一类为不包含摄像机主动运动的视频,定向防抖有效地提高了画面的稳定性。总的来说,本文有以下三方面创新点。

(1)利用基于直方图分布的方法将视频分为含摄像机运动及不含摄像机运动,随后根据类别进行防抖处理,可适应不同场景下的防抖需求。

(2)对于类别为不含摄像机主动运动的视频,提出了一种基于 Harris 矩阵^[19]与图像区域划分的特征点提取方案并添加特征点置信度约束,然后通过匹配的特征点估计并应用当前帧至首帧的单应性变换来实现画面稳定,消除了该类别视频用全局运动估计法防抖后仍包含画面抖动的问题。

(3)对于包含摄像机主动运动的视频,采用基于 FAST^[20]角点检测和 RANSAC^[21]的全局二维运动估计方法,以获得摄像机运动轨迹,随后采用高斯滤波对摄像机轨迹进行平滑处理,利用滤波后得到抖动运动估计量,对前后帧图像进行仿射变换,最后利用插值算法^[22-23]对变换后图像中未定义区域进行校正。对比其他基于特征匹配^[24]的运动估计法,该方法有了更大的提升。

1 航拍视频稳定方法

本文在防抖处理前,提前对视频进行分类,区分出视频类别,随后对两类视频应用不同的处理方案。本文的算法框架如图 1 所示。

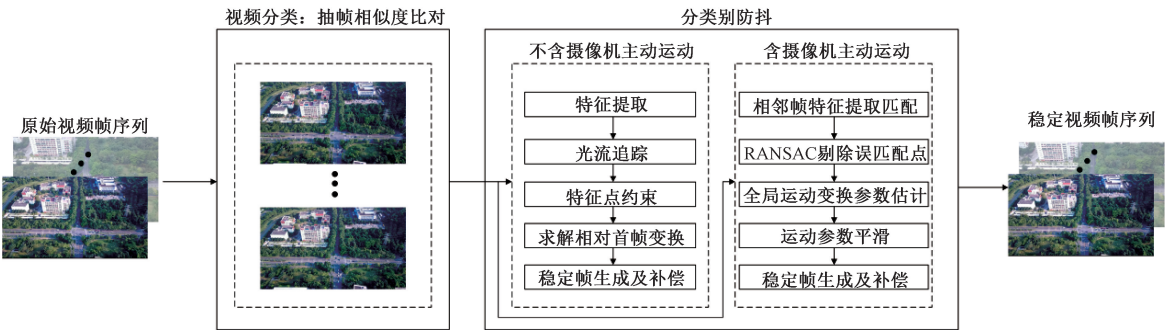


图 1 视频稳定算法框架

Figure 1 Video stabilization algorithm framework

1.1 视频分类

本文对视频进行分类的依据如下:等间距抽取视频序列中的帧,确保在视频的前中后段都有选定的帧,给定一个相似度阈值 S_v ,随后两两比对这些帧画面的色调及饱和度直方图分布,若某一次比对结果低于设定的阈值,则说明画面发生了非相机随机抖动造成的变化,从而视该视频为包含相机主动运动;若全部比对结果均不低于阈值,则把视频认作不包含相机主动运动,只包含相机随机运动。经实验分析,阈值 S_v 的值设为 0.85 较为合适。图 2 为该算法的流程图。

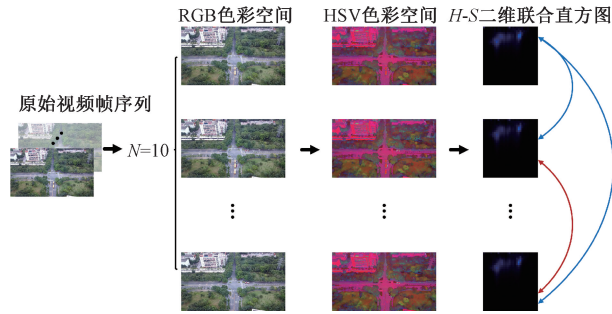


图2 视频分类算法示意图

Figure 2 Video classification algorithm diagram

给定一个视频序列 $I = \{I_1, I_2, \dots, I_N\}$, $N > 30$, $k = N/10$, 其中 I_N 代表视频中的第 N 帧。从 I 中抽取第 $I_{i \times k}$ ($i = 1, 2, \dots, 10$) 帧共 10 幅图片,将 10 幅图片从 RGB 色彩空间转换至 HSV 色彩空间,随后统计每一张图片的 H (色调) 维度和 S (饱和度) 维度的灰度值直方图并对其进行归一化处理得到联合直方图,然后比较两幅图片的二维联合直方图分布,若当前得到的两幅图片的二维联合直方图分别为 H_1 、 H_2 , 那么图片相似度的算法如下。

算法 1 二维联合直方图相似度比对。

输入: 两张二维联合直方图;

输出: 相似度值。

- ① foreach $h_1[j], h_2[j]$ in H
- ② $a \leftarrow h_1[j]$; // a 为 H_1 中下标为 j 处像素值;
- ③ $b \leftarrow h_2[j]$; // b 为 H_2 中下标为 j 处像素值;
- ④ $S_1 \leftarrow S_1 + a$;
- ⑤ $S_2 \leftarrow S_2 + b$;
- ⑥ $S_{11} \leftarrow S_{11} + a \cdot a$;
- ⑦ $S_{12} \leftarrow S_{12} + a \cdot b$;
- ⑧ $S_{22} \leftarrow S_{22} + b \cdot b$;
- ⑨ $t \leftarrow H_1 \cdot \text{total}()$;
- ⑩ $s \leftarrow 1.0/t$;
- ⑪ $r \leftarrow S_{12} - S_1 \cdot S_2 \cdot s$;
- ⑫ $d \leftarrow (S_{11} - S_1 \cdot S_1 \cdot s) \cdot (S_{22} - S_2 \cdot S_2 \cdot s)$;

⑬ $\text{result} \leftarrow \text{abs}(d) > 2.22 \cdot 10^{-16} ? r/\text{sqrt}(d) : 1.0$ 。

其中, h_1, h_2 分别为这两张图上同一点的灰度值; S_{12} 为图上每一点 h_1 与 h_2 的乘积之和; S_1 为每一点的 h_1 之和; S_{11} 为每一点 h_1 自乘之和; S_2 为每一点的 h_2 之和; S_{22} 为每一点 h_2 自乘之和; s 为 1 除以 H_1 所有点 h_1 之和; r 为两个二维联合直方图之间的协方差; d 衡量的是 H_1 和 H_2 之间的方差乘积, 反映了两个直方图统计分布的一致性; 最终根据 d 值调整得到用于比较的相似度 result , 将 result 的值与阈值 S_v 进行比较即可对视频进行分类。

1.2 不含摄像机主动运动的视频防抖算法

由于该类无人机航拍视频具有如下特点: 仅存在画面边缘的抖动, 画面背景几乎无变化以及无大面积前景运动物体, 本文将视频第一帧作为参考帧, 并使后续所有帧最大程度与参考帧重合来消除画面抖动。图 3 为不含摄像机主动运动的视频防抖算法流程。

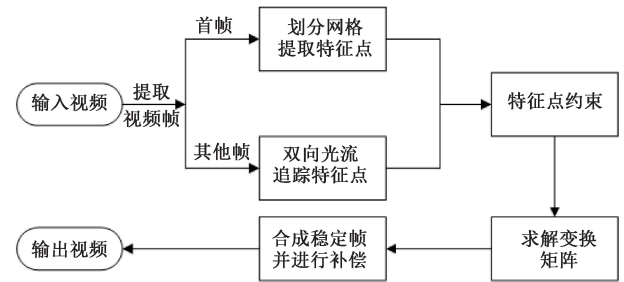


图3 不含摄像机主动运动的视频防抖算法流程图

Figure 3 Flow chart of video stabilization algorithm without active camera motion

1.2.1 特征提取

受王杰等^[6]研究的启发, 为了确保图像特征点数目充足且分布均匀, 本文提出了一种基于 Harris 矩阵的特征点提取方式, 将图片划分成网格进行特征提取。本文只对视频首帧进行特征点提取, 步骤如下。

步骤 1 考虑到原视频分辨率, 本文保持一定的比例进行图片分割, 将图片划分为 20×15 的网格。

步骤 2 计算该网格内每一点的 Harris 矩阵, 对于点 (x, y) , 该点 Harris 矩阵的计算方式^[19]为

$$M(X, Y) = \sum_w \begin{bmatrix} I_x(x, y)^2 & I_x(x, y)I_y(x, y) \\ I_x(x, y)I_y(x, y) & I_y(x, y)^2 \end{bmatrix} = \begin{bmatrix} \sum_w I_x(x, y)^2 & \sum_w I_x(x, y)I_y(x, y) \\ \sum_w I_x(x, y)I_y(x, y) & \sum_w I_y(x, y)^2 \end{bmatrix} = \begin{bmatrix} A & C \\ C & B \end{bmatrix} \quad (1)$$

式中: w 为以点 (x, y) 为中心的窗口函数, 通常取高斯函数; I_x 为图像在 x 方向上的梯度; I_y 为图像在 y

方向上的梯度。 A 的值越大表明该点 x 方向上像素值变化越大, B 的值越大表明该点 y 方向上像素值变化越大, 因此 $A+B$ 的值越大, 则该点为特征点的可能性就越大, 这里直接将 $A+B$ 值最大的点设为特征点, 并记录其坐标信息。

1.2.2 特征匹配及特征点约束

首帧特征点提取完成后, 进行后续帧与首帧的特征匹配。利用金字塔光流法^[25]在后续帧中追踪首帧特征点的位置, 得到当前帧特征点坐标, 随后再反向在首帧中追踪当前帧特征点。为了保证参与帧稳定变换估计的特征点尽可能地出现在所有帧中, 以及避免某个特征点因其 Harris 矩阵的 A 值或 B 值偏大导致特征点估计错误, 这里为每个网格提取到的特征点设置一个置信度 *Reliability* 来对其进行约束, 算法如下。

算法 2 特征点约束。

输入: 各网格 Harris 矩阵信息、特征点坐标;

输出: 特征点置信度。

- ① Foreach i in $grids$; // i 为网格序号
- ② $\Delta x \leftarrow r_points[i].x - c_points[i].x$;
- ③ $\Delta y \leftarrow r_points[i].y - c_points[i].y$;
- ④ $\alpha = HarrisMatrices[i].A$;
- ⑤ $\beta = HarrisMatrices[i].B$;
- ⑥ $\gamma = HarrisMatrices[i].C$;
- ⑦ $v \leftarrow (\alpha + \beta + \sqrt{(\alpha - \beta)}) \cdot (\alpha - \beta) + 4 \cdot \gamma \cdot \gamma$; // 进行特征值计算
- ⑧ $E[i] \leftarrow 0.5f \cdot v$; // 带权偏移计算
- ⑨ $s \leftarrow (\alpha \cdot \Delta x \cdot \Delta x + \beta \cdot \Delta y \cdot \Delta y + 2 \cdot \gamma \cdot \Delta x \cdot \Delta y) / E[i]$;
- ⑩ $isValid[i] \leftarrow is_forward_matches[i] \&\& is_backward_matches[i] \&\& s < 0.1 \&\& (\Delta x \cdot \Delta x + \Delta y \cdot \Delta y) < 1\ 000$;
- ⑪ $Reliability[i] \leftarrow 0.95 \cdot Reliability[i]$;
- ⑫ If $isValid[i]$
- ⑬ $Reliability[i] \leftarrow Reliability[i] + 0.05$ 。

其中, i 为网格序号; r_points 为首帧特征点坐标; c_points 为当前帧特征点坐标; $HarrisMatrices$ 为特征点的 Harris 矩阵信息; $is_forward_matches$ 表示首帧特征点是否在当前帧追踪成功; $is_backward_matches$ 表示当前帧特征点是否在首帧追踪成功; $isValid$ 表示当前帧该特征点是否有效。

1.2.3 变换矩阵求解

根据单应变换模型, 假设参考帧及当前帧对应的特征点的齐次坐标分别为 $L_r = [u_1, v_1, 1]^T$, $L_c =$

$[u_2, v_2, 1]^T$, 利用 3×3 的矩阵表示单应变换, 设 $H =$

$$\begin{bmatrix} h_1 & h_2 & h_3 \\ h_4 & h_5 & h_6 \\ h_7 & h_8 & h_9 \end{bmatrix}, \text{ 那么 } L_r \text{ 与 } L_c \text{ 的关系可以表示为}$$

$$L_c = H \times L_r. \quad (2)$$

式(2)中 H 即为单应矩阵, 由于单应矩阵自身的约束, h_9 可设为 1, 则 $H =$

$$\begin{bmatrix} h_1 & h_2 & h_3 \\ h_4 & h_5 & h_6 \\ h_7 & h_8 & 1 \end{bmatrix}. \text{ 将 } H, L_r, L_c$$

代入式(2), 可得

$$\begin{cases} u_2 = \frac{h_1 \cdot u_1 + h_2 \cdot v_1 + h_3}{h_7 \cdot u_1 + h_8 \cdot v_1 + 1}; \\ v_2 = \frac{h_4 \cdot u_1 + h_5 \cdot v_1 + h_6}{h_7 \cdot u_1 + h_8 \cdot v_1 + 1}. \end{cases} \quad (3)$$

式(3)进一步整理可得

$$\begin{cases} h_1 \cdot u_1 + h_2 \cdot v_1 + h_3 - h_7 \cdot u_1 \cdot u_2 - h_8 \cdot v_1 \cdot u_2 = u_2; \\ h_4 \cdot u_1 + h_5 \cdot v_1 + h_6 - h_7 \cdot u_1 \cdot v_2 - h_8 \cdot v_1 \cdot v_2 = v_2. \end{cases} \quad (4)$$

式(4)的矩阵形式为

$$\begin{bmatrix} u_1 & v_1 & 1 & 0 & 0 & 0 & -u_1 u_2 & -v_1 u_2 \\ 0 & 0 & 0 & u_1 & v_1 & 1 & -u_1 v_2 & -v_1 v_2 \end{bmatrix} \begin{bmatrix} h_1 \\ h_2 \\ h_3 \\ h_4 \\ h_5 \\ h_6 \\ h_7 \\ h_8 \end{bmatrix} = \begin{bmatrix} u_2 \\ v_2 \end{bmatrix}. \quad (5)$$

由此可知一组匹配点能得到两个方程, 那么利用 4 组匹配点(无 3 点共线)即可求得单应变换矩阵 H 。该步将 $isValid$ 值设为 1 且 $Reliability > 0.5$ 的特征点代入计算, 由于存在多个特征匹配点, 本文采用最小二乘法来求得最优解。

求得当前帧与首帧的单应变换矩阵 H 后, 将其应用于当前帧即可得到稳定的视频帧, 再采用双线性插值进行伪影修复即可得到最终防抖后的视频帧。

1.3 含摄像机主动运动的视频防抖算法

针对包含摄像机主动运动的视频, 防抖算法的主要思想是通过相邻帧的特征匹配来计算帧间的相对平移、旋转和缩放, 从而得到相邻帧之间的变换关系, 并累计这些变换关系以计算摄像机的运动轨迹。为了滤除高频抖动分量, 本文使用滤波算法平滑摄像机的运动轨迹, 然后根据摄像机运动轨迹, 对画面进行补偿, 以合成稳定的视频帧。

本文采用基于光流和 RANSAC 的二维运动估计方法估计摄像机运动轨迹,而在轨迹平滑方面采用高斯滤波方法,然后通过双线性插值对画面进行补偿。算法整体流程如图 4 所示。

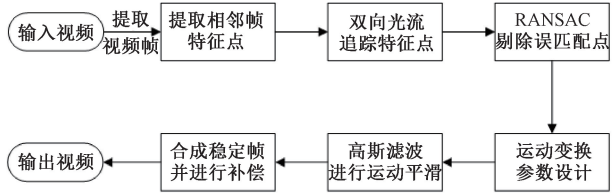


图 4 含摄像机主动运动的视频防抖算法流程图
Figure 4 Flow chart of video stabilization algorithm with active camera motion

1.3.1 FAST 角点检测

FAST^[20]角点检测步骤如下。

步骤 1 将图像转换为灰度图像。

步骤 2 初始化一个空的角点列表。

步骤 3 对于图像中的每个像素,如果该像素与周围 16 个像素的灰度差异中,有超过 12 个像素的灰度差异值高于或低于设定的阈值,则将该像素视为角点。

步骤 4 将检测到的角点添加到角点列表中。

步骤 5 返回角点列表作为结果。

FAST 算法计算速度快、可检测出的特征点多,虽然其检测出的角点不具备尺度不变性及旋转不变性,但由于该类无人机航拍视频具有如下特点:画面无大幅度旋转、视角无缩放、抖动分量主要集中于画面左右两端,因此可采用检测效率极高的 FAST 角点。

1.3.2 特征匹配及 RANSAC 算法

本文在得到图像特征点后,进行前后帧图像的特征匹配。与 1.2.2 节一致,本文采用金字塔光流法在后续帧追踪前一帧的特征点位置,并反向追踪后续帧的特征点来进行特征匹配,其具体步骤如下。

步骤 1 在当前帧中的每个特征点周围创建一个小的窗口。

步骤 2 在后续帧中寻找与当前窗口中的特征点最匹配的像素,使用像素点的位移来计算特征点的运动矢量。

步骤 3 若是计算出的像素点太远离窗口中心即特征点的初始位置,或者没有匹配的特征点,则视为追踪失败。

由于大量特征点的存在,可能导致误匹配点,本文采用 RANSAC^[21]算法剔除错误匹配的特征点,随后利用最小二乘法估计仿射变换运动参数,该过程具体步骤如下。

步骤 1 随机抽样:从所有的特征点对中随机

选择一个最小样本集(通常是 3 个对应的点),作为当前的候选内点集。

步骤 2 拟合模型:使用最小样本集中的点来拟合一个模型。对于每个最小样本集,通过最小二乘法计算出一个初始的仿射变换矩阵。

步骤 3 内点选择:对于其他未被选中的特征点对,计算它们与当前估计的模型之间的误差。

步骤 4 判断内点:根据预先设定的阈值,将与模型拟合误差小于阈值的特征点对标记为内点。

步骤 5 内点数目判断:如果内点数目大于预先设定的阈值(大于总特征点数的 75%),则继续执行下面的步骤;否则返回第(1)步重新随机抽样。

步骤 6 重新估计模型:使用所有内点重新估计一个更好的模型。

步骤 7 模型评估:根据内点数目评估当前模型的拟合程度或性能。

步骤 8 重复迭代:重复执行上述步骤多次,以选择产生内点数最多的模型作为最终的运动模型。根据这个最终模型,可以确定仿射变换参数,例如平移、旋转、缩放和裁切。

1.3.3 仿射变换模型

假设视频序列当前帧及后一帧对应的特征点的齐次坐标分别为 $L_n = [X_n, Y_n, 1]^T$, $L_{n+1} = [X_{n+1}, Y_{n+1}, 1]^T$, 利用 3×3 的矩阵表示仿射变换,设 $M =$

$$\begin{bmatrix} a_1 & a_2 & \Delta x \\ a_3 & a_4 & \Delta y \\ 0 & 0 & 1 \end{bmatrix}, \text{ 那么 } L_{n+1} \text{ 与 } L_n \text{ 的关系可以表示为}$$

$$L_{n+1} = M \times L_n. \quad (6)$$

$$A = \begin{bmatrix} a_1 & a_2 \\ a_3 & a_4 \end{bmatrix} \text{ 表示旋转及缩放尺度, } t = [\Delta x, \Delta y]^T \text{ 表示平移尺度, 将 } L_{n+1} \text{ 与 } L_n \text{ 代入式(5)有}$$

$$\begin{cases} X_{n+1} = a_1 \cdot X_n + a_2 \cdot Y_n + \Delta x; \\ Y_{n+1} = a_3 \cdot X_n + a_4 \cdot Y_n + \Delta y. \end{cases} \quad (7)$$

通过式(7),可利用相邻帧的匹配特征点求解仿射变换参数,这里同样采用最小二乘法求得最优解。

1.3.4 运动平滑

当求得最佳仿射变换矩阵数据的平均方向后,可通过滤波的方式抹平波峰波谷以得到稳定的摄像机运动轨迹。常用于轨迹平滑的有卡尔曼滤波^[5-6],然而卡尔曼滤波只对满足高斯分布的抖动干扰滤波效果较好,但无人机航拍视频的抖动干扰多数不满足高斯分布,因此卡尔曼滤波方案具有一定的局限性。均值滤波可以有效地去除高频的抖动分量,但滤波后的曲线仍不够平滑,效果欠佳。同

时,由于均值滤波仅考虑当前帧与附近帧的绝对位移,容易造成运动滞后及运动超前现象。高斯滤波与均值滤波类似,但考虑到了不同帧的权重,能有效去除高频抖动分量的同时保证曲线平滑,滤波效果理想,因此本文采用高斯滤波进行运动平滑。

本文将平滑后的仿射变换矩阵数据应用于原视频帧后即得到防抖后的图像,最后采用双线性插值进行运动补偿,填充画面黑色区域。

2 实验结果及分析

为了评估本文算法的性能,本文挑选了两个效果较好的电子稳像算法^[5-6]进行比对,共选取了 7 个视频作为实验数据集,实验视频均为利用无人机延时摄影功能拍摄的视频,其中包含与不包含摄像机主动运动的视频个数分别为 3 和 4。实验采用视频稳定性评价常用的峰值信噪比 $PSNR$ 作为评价指标评估算法的有效性,一般来说, $PSNR$ 越大,说明视频前后两帧的相似度越高,画面稳定效果越好;平均 $PSNR$ 越大,说明视频整体稳定效果越好。

2.1 视频分类准确性验证

视频类别的正确划分是本文算法有效的重要前提。对全部视频进行分类,其中序号为 1~3 的视频为包含摄像机主动运动的视频;序号为 4~7 的视频则为不包含摄像机主动运动的视频,实验结果全部分类正确。为了更直观展示视频分类的过程,这里给出视频 3 及视频 7 用于分类的帧,图 5 为视频 7 中抽取的帧,图 6 为视频 3 中抽取的帧,从左至右、从上至下的图片分别为按时序出现的视频帧。

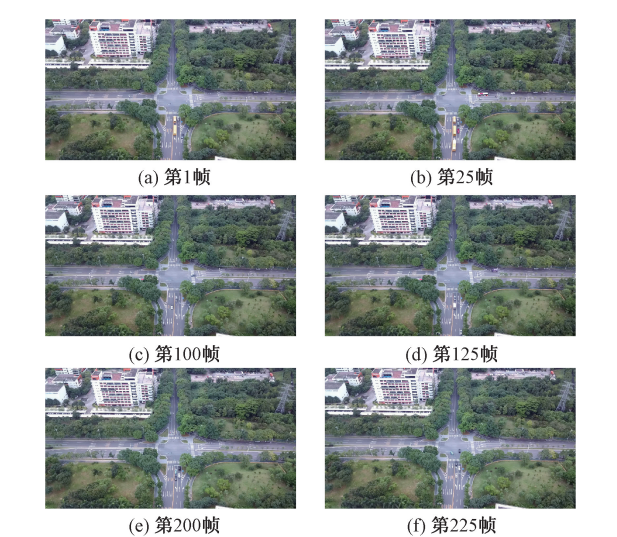


图 5 视频 7 抽取的用于分类的帧

Figure 5 Frames in video seven extracted for classification

图 5 所有图片的背景均保持一致,画面背景几乎无变化,所有图片的相似度值均大于 0.95,而图 6



图 6 视频 3 抽取的用于分类的帧

Figure 6 Frames in video three extracted for classification

中图片画面差异较大,例如抽取出的帧数中的第 1 帧与第 2 帧相似度值为 0.62。

2.2 视频防抖效果及分析

文献[5-6]算法及本文算法在视频数据集上的表现如表 1 所示,表 1 中数值为视频图像平均峰值信噪比。

从表 1 可以看出,本文方法处理抖动无人机航拍视频的能力优于其他比较的方法,稳定后的平均峰值信噪比最高。

表 1 不同算法 $PSNR$ 对比结果

Table 1 Comparison results of $PSNR$ of different algorithms

视频 序号	峰值信噪比 $PSNR/dB$			
	原视频	文献[5]算法	文献[6]算法	本文算法
1	31.010 7	32.097 6	32.123 1	32.602 3
2	30.941 9	31.921 3	31.920 5	32.660 4
3	32.144 1	32.645 8	32.415 8	33.202 5
4	16.529 6	18.182 8	18.257 8	28.994 2
5	15.970 2	17.648 4	17.779 6	29.711 9
6	18.539 0	20.366 9	20.434 7	29.323 2
7	22.636 5	24.124 6	24.192 8	31.549 7

本文的防抖方法在不含摄像机主动运动的航拍视频中表现出色,远超过其他两种方法。以视频 7 为例,从图 7 可以看出,本文方法处理后的视频峰值信噪比为 31.549 7,相比原视频平均峰值信噪比 22.636 5,提升了 39.38%;而文献[5]算法和文献[6]算法分别提升了 6.57%和 6.88%。这是因为本文采用了直接将后续帧投影至首帧的方法,能够有效消除帧间相对位移,实现画面的绝对稳定。而另外两种方法则需要进行轨迹估计和平滑处理,但容易累积帧间偏移误差,导致画面不能达到绝对稳定。图 8 直观展示了不同算法在视频 7 上的防抖效果。

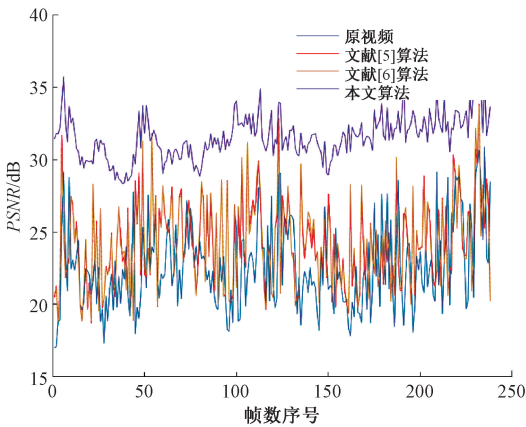


图 7 视频 7 帧间峰值信噪比

Figure 7 Interframe peak signal-to-noise ratio of video seven

以图像左上角区域中较矮的楼房为参考,比较其与边界的距离,可以看出文献[5-6]方法防抖后画面边缘仍是非静止的,而本文方法防抖处理后的画面边缘完全一致,保证了画面的绝对稳定。

对于包含摄像机主动运动的视频,本文方法优于其他两种方法。图 9 为视频 3 上不同算法稳定后的峰值信噪比,视频 3 原视频平均峰值信噪比为 32.144 1,文献[5]算法提升了 1.56%,达到了 32.645 8;文献[6]算法提升了 0.84%,达到了 32.415 8;本文算法则提升了 3.29%,达到了 33.202 5,表现出最优的性能。

基于最小生成树的特征提取方式^[5]可以有效避免前景运动物体对于帧间稳定变换估计的影响,

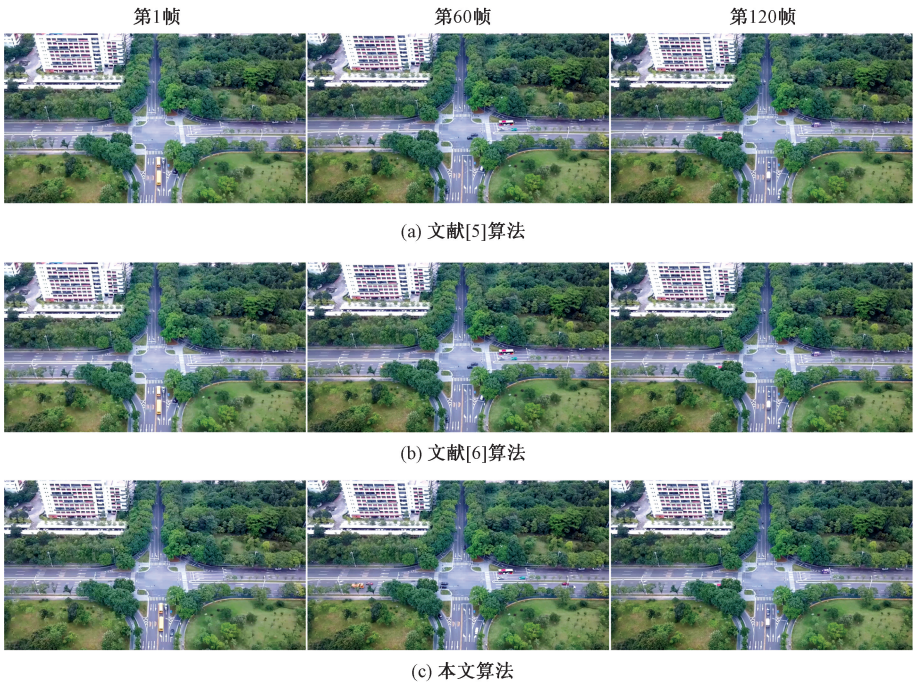


图 8 不同算法防抖效果展示

Figure 8 Display of anti-shake effect of different algorithms

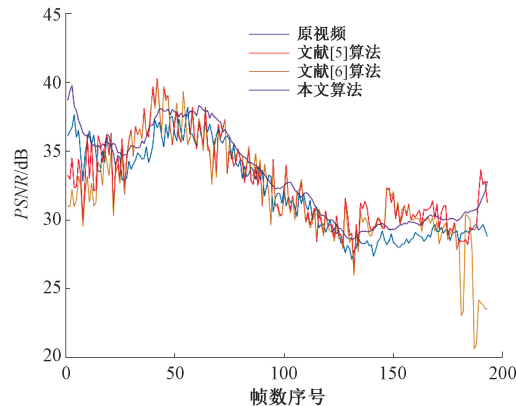


图 9 视频 3 帧间峰值信噪比

Figure 9 Interframe peak signal-to-noise ratio of video three

但由于无人机视角的关系,其优势并不明显,甚至可能错误地剔除了一部分信息,其与卡尔曼滤波相结合的防抖方式更适用于地面移动机器人拍摄视频的稳定。虽然卡尔曼滤波在系统动态建模以及测量噪声的建模和处理上更具优势,但由于在高空无人机运动模式的不稳定导致了相机的运动模式不稳定或难以预测,这可能导致卡尔曼滤波无法准确估计系统的状态。在这种情况下,使用高斯滤波能够通过平滑观测数据来减少运动模式的不稳定性,提供相对较好的平滑效果。

3 结论

本文针对无人机航拍视频中普遍存在的画面抖

动问题,提出了一种基于相机运动轨迹的自动分类视频稳定算法。该方法采用直方图分布法将视频进行分类处理,提高了算法的鲁棒性。实验结果表明,本文所采用的视频分类方法能够保证分类准确性。与基于最小生成树加自适应卡尔曼滤波法及基于 Harris 加自适应卡尔曼滤波法相比,本文方法防抖后的视频图像平均 *PSNR* 提升最高,作为一种低成本、高效率的稳定方法,可以有效地提高无人机航拍视频的质量,提升用户观感体验。

参考文献:

- [1] 苑玉彬,吴一全,赵朗月,等. 基于深度学习的无人机航拍视频多目标检测与跟踪研究进展[J]. 航空学报, 2023, 44(18): 028334.
YUAN Y B, WU Y Q, ZHAO L Y, et al. Research progress of UAV aerial video multi-object detection and tracking based on deep learning[J]. Acta Aeronautica et Astronautica Sinica, 2023, 44(18): 028334.
- [2] GRUNDMANN M, KWATRA V, ESSA I. Auto-directed video stabilization with robust L1 optimal camera paths [C]//Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2011: 225-232.
- [3] LEE K Y, CHUANG Y Y, CHEN B Y, et al. Video stabilization using robust feature trajectories [C]//2009 IEEE 12th International Conference on Computer Vision. Piscataway: IEEE, 2010: 1397-1404.
- [4] 唐佳林,郑杰峰,李熙莹,等. 基于特征匹配与运动补偿的视频稳像算法[J]. 计算机应用研究, 2018, 35(2): 608-610, 614.
TANG J L, ZHENG J F, LI X Y, et al. Video stabilization algorithm based on feature matching and motion compensation [J]. Application Research of Computers, 2018, 35(2): 608-610, 614.
- [5] 谢亚晋,徐之海,冯华君,等. 基于最小生成树与改进卡尔曼滤波器的实时电子稳像方法[J]. 光子学报, 2018, 47(1): 0110002.
XIE Y J, XU Z H, FENG H J, et al. Real-time video stabilization based on minimal spanning tree and modified Kalman filter[J]. Acta Photonica Sinica, 2018, 47(1): 0110002.
- [6] 王杰,经俊森,陈正伟,等. 基于 Harris 和卡尔曼滤波的农业机器人田间稳像算法[J]. 农业机械学报, 2023, 54(1)30-36, 53.
WANG J, JING J S, CHEN Z W, et al. Field image stabilization algorithm for agricultural robot based on Harris and Kalman filter[J]. Transactions of the Chinese Society for Agricultural Machinery, 2023, 54(1)30-36, 53.
- [7] YU J Y, RAMAMOORTHY R. Learning video stabilization using optical flow [C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2020: 8156-8164.
- [8] ALI M K, YU S, KIM T H. Learning deep video stabilization without optical flow [EB/OL]. (2020-11-19) [2023-06-16]. <https://doi.org/10.48550/arXiv.2011.09697>.
- [9] XU Y F, ZHANG J, MAYBANK S J, et al. DUT: learning video stabilization by simply watching unstable videos [J]. IEEE Transactions on Image Processing, 2022, 31: 4306-4320.
- [10] BRADLEY A, KLIVINGTON J, TRISCARI J, et al. Cinematic-L1 video stabilization with a log-homography model [C]//Conference on Applications of Computer Vision (WACV). Piscataway: IEEE, 2021: 1040-1048.
- [11] LEE Y C, TSENG K W, CHEN Y T, et al. 3D video stabilization with depth estimation by CNN-based optimization [C]//Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2021: 10616-10625.
- [12] LIU F, GLEICHER M, JIN H, et al. Content-preserving warps for 3D video stabilization [J]. ACM Transactions on Graphics, 2009, 28(3): 44.
- [13] LIU S, WANG Y, YUAN L, et al. Video stabilization with a depth camera [C]//Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2012: 89-95.
- [14] LIU Y L, LAI W S, YANG M H, et al. Hybrid neural fusion for full-frame video stabilization [C]//International Conference on Computer Vision (ICCV). Piscataway: IEEE, 2022: 2279-2288.
- [15] SHI Z M, SHI F H, LAI W S, et al. Deep online fused video stabilization [C]//Conference on Applications of Computer Vision (WACV). Piscataway: IEEE, 2022: 865-873.
- [16] WANG M, YANG G Y, LIN J K, et al. Deep online video stabilization with multi-grid warping transformation learning [J]. IEEE Transactions on Image Processing, 2018, 28(5): 2283-2292.
- [17] XU S Z, HU J, WANG M, et al. Deep video stabilization using adversarial networks [J]. Computer Graphics Forum, 2018, 37(7): 267-276.
- [18] ZHAO M, LING Q. PWStableNet: learning pixel-wise warping maps for video stabilization [J]. IEEE Transactions on Image Processing, 2020, 29: 3582-3595.
- [19] 吴国楠,周超超,尹文波. 基于 Harris 角点与改进 Hu 矩的电子稳像算法 [J]. 计算机工程, 2013, 39(3): 300-305, 310.
WU G N, ZHOU C C, YIN W. Electronic image stabiliza-

zation algorithm based on Harris corner and modified Hough moments[J]. Computer Engineering, 2013, 39(3): 300-305, 310.

[20] ROSTEN E, DRUMMOND T. Machine learning for high-speed corner detection[C]// Proceedings of the 9th European conference on Computer Vision. Berlin: Springer, 2006: 430-443.

[21] STANKOVIĆ L, BRAJOVIĆ M, STANKOVIĆ I, et al. RANSAC-based signal denoising using compressive sensing[J]. Circuits, Systems, and Signal Processing, 2021, 40(8): 3907-3928.

[22] 孙辉. Bayer 彩色序列图像的电子稳像[J]. 光学精密工程, 2014, 22(2): 481-488.

SUN H. Electronic image stabilization for Bayer color video sequence image[J]. Optics and Precision Engineering, 2014, 22(2): 481-488.

[23] 王金鑫, 秦子龙, 曹泽宁, 等. 基于八叉树的修正克里金空间插值算法[J]. 郑州大学学报(工学版), 2021, 42(6): 21-27.

WANG J X, QIN Z L, CAO Z N, et al. Modified Kriging spatial interpolation algorithm based on octree mechanism[J]. Journal of Zhengzhou University (Engineering Science), 2021, 42(6): 21-27.

[24] 华蓓, 陈前, 黄汝维. 基于特征匹配的图像真伪检测方法的研究[J]. 郑州大学学报(工学版), 2022, 43(2): 22-27.

HUA B, CHEN Q, HUANG R W. Research of image authenticity detection method based on feature matching[J]. Journal of Zhengzhou University (Engineering Science), 2022, 43(2): 22-27.

[25] AHN B, HAN Y, KWEON I S. Real-time facial landmarks tracking using active shape model and LK optical flow[C]//2012 9th International Conference on Ubiquitous Robots and Ambient Intelligence (URAI). Piscataway: IEEE, 2013: 541-543.

Robust Algorithm for Aerial Video Stabilization of UAV Based on Camera Motion Trajectory

YU Songsen, LONG Jiahao, ZHOU Nuo, LIANG Jun

(School of Software, South China Normal University, Foshan 528225, China)

Abstract: Aiming at the problem of high altitude turbulence environment to the time-delay stable image acquisition of UAV, an anti-shaking algorithm for aerial video was proposed for hovering shooting and moving shooting. Firstly from the time-delay photography video captured the UAV camera, some video frames were extracted globally to compare their histogram distributions. This comparison could identify whether the video contained active camera motion or not, and help categorize the video accordingly. For videos with active camera motion, FAST corner detection and optical flow methods were used to extract and match feature points. The RANSAC algorithm could remove all mismatched feature points, and estimate the camera's motion trajectory. The resulting motion estimation parameters were then smoothed using Gaussian filtering, producing a stable camera motion trajectory. For videos without active camera motion, the first frame was divided into grids and feature points were extracted based on Harris matrix. Optical flow tracking was carried out on these feature points in subsequent frames. Reverse optical flow and Harris matrix calculation were used to extract and match feature points, to increase the constraint of feature points. Finally, the retained feature points were used to estimate the stable transformation from subsequent frames to the first frame. Experimental results showed that the video classification module could correctly distinguish between the two types of videos. The algorithm was used to classify the video scene and stabilize the picture. Compared to other methods, this algorithm could improve the average peak signal-to-noise ratio of stabilized video images the most. For videos without active camera motion, the image could be absolutely stable, and the average peak signal-to-noise ratio of the image was increased by more than 39%, while the other two methods only by 10% to 12%.

Keywords: UAV; video stabilization; feature extraction; motion estimation; peak signal-to-noise ratio