

信号处理  
*Journal of Signal Processing*  
ISSN 1003-0530, CN 11-2406/TN

## 《信号处理》网络首发论文

题目: Savitzky-Golay 滤波和 L1 范数优化相结合的视频稳像算法  
作者: 杨雨辰, 毋立芳, 简萌, 赵博煊, 刘旭  
网络首发日期: 2020-09-15  
引用格式: 杨雨辰, 毋立芳, 简萌, 赵博煊, 刘旭. Savitzky-Golay 滤波和 L1 范数优化相结合的视频稳像算法. 信号处理.  
<https://kns.cnki.net/kcms/detail/11.2406.TN.20200915.1456.013.html>



**网络首发:** 在编辑部工作流程中, 稿件从录用到出版要经历录用定稿、排版定稿、整期汇编定稿等阶段。录用定稿指内容已经确定, 且通过同行评议、主编终审同意刊用的稿件。排版定稿指录用定稿按照期刊特定版式 (包括网络呈现版式) 排版后的稿件, 可暂不确定出版年、卷、期和页码。整期汇编定稿指出版年、卷、期、页码均已确定的印刷或数字出版的整期汇编稿件。录用定稿网络首发稿件内容必须符合《出版管理条例》和《期刊出版管理规定》的有关规定; 学术研究成果具有创新性、科学性和先进性, 符合编辑部对刊文的录用要求, 不存在学术不端行为及其他侵权行为; 稿件内容应基本符合国家有关书刊编辑、出版的技术标准, 正确使用和统一规范语言文字、符号、数字、外文字母、法定计量单位及地图标注等。为确保录用定稿网络首发的严肃性, 录用定稿一经发布, 不得修改论文题目、作者、机构名称和学术内容, 只可基于编辑规范进行少量文字的修改。

**出版确认:** 纸质期刊编辑部通过与《中国学术期刊 (光盘版)》电子杂志社有限公司签约, 在《中国学术期刊 (网络版)》出版传播平台上创办与纸质期刊内容一致的网络版, 以单篇或整期出版形式, 在印刷出版之前刊发论文的录用定稿、排版定稿、整期汇编定稿。因为《中国学术期刊 (网络版)》是国家新闻出版广电总局批准的网络连续型出版物 (ISSN 2096-4188, CN 11-6037/Z), 所以签约期刊的网络版上网络首发论文视为正式出版。

# Savitzky-Golay 滤波和 L1 范数优化相结合的视频稳像算法

杨雨辰<sup>1</sup> 毋立芳<sup>1</sup> 简萌<sup>1</sup> 赵博煊<sup>1</sup> 刘旭<sup>1</sup>  
(1.北京工业大学信息学部, 北京 100124)

**摘要：**近年来,越来越多的人利用手持设备拍摄视频表达自己或社会问题等,这类视频常存在抖动现象,如何消除视频序列中的随机抖动成为近年来的研究热点。本文面向视频去抖动应用场景,提出了一种 Savitzky-Golay 滤波和 L1 范数优化相结合的视频稳像算法。首先基于帧间配对特征点进行运动估计;然后利用 Savitzky-Golay 滤波算法进行路径平滑处理,去除运动路径中的高频抖动分量,避免了噪声对后续生成稳定路径的影响;进一步,对平滑后的有效路径信息进行基于优化的 L1 范数优化算法拟合,得到最终的稳定路径;最后对原始视频图像序列进行运动补偿,生成稳定的图像序列。对比实验结果表明,本文方法稳像效果优于原始 L1 范数优化算法和最近较好的方法,算法速度快于目前效果较好的二维稳像算法。

**关键词：**视频稳像; Savitzky-Golay 滤波; L1 范数优化; 运动补偿

**中图分类号：**TP391.4

**文献标识码：**A

## Video stabilization algorithm based on Savitzky-Golay filtering and L1 norm optimization

Yang Yuchen<sup>1</sup> Wu Lifang<sup>1</sup> Jian Meng<sup>1</sup> Zhao Boxuan<sup>1</sup> Liu Xu<sup>1</sup>

(1.Department of Information, Beijing University of Technology, Beijing 100124, China)

**Abstract:** In recent years, more and more people use hand-held devices to shoot videos to express themselves or social problems, etc. This kind of video often has the phenomenon of jitter, how to eliminate the random jitter in the video sequence has become a research focus in recent years. In this paper a video stabilization algorithm based on Savitzky-Golay filtering and L1 norm optimization was proposed for real time application, which is used to de-jitter video captured by hand-held cameras. Firstly, the motion is estimated based on the feature points of inter-frame pairing. Next, Savitzky-Golay filtering algorithm is used for path smoothing to remove the high frequency jitter components in the moving path, so as to avoid the influence of noise on the subsequent generation of stable paths. Then, the L1 norm optimization algorithm is used to optimize the smooth effective path information to fit the final stable path. Finally, the motion compensation of the original video image sequence is carried out to generate a stable image sequence. The compared experimental results demonstrate that the proposed method is better than the original L1 norm optimization algorithm and a popular method in video stabilization and faster than the current two-dimensional image stabilization algorithm.

**Key words:** video stabilization; Savitzky-Golay filtering; L1 norm optimization; motion compensation

### 1 引言

随着手持设备的发展与普及,人们可以利用手机、数码相机等设备进行简单的视频拍摄。与专业摄像机相比,手持设备拍摄出

的视频在视觉上常常会存在不希望出现的抖动情况。因此,如何消除视频序列中的随机抖动,成为了近年来研究学者们的研究热点。

视频稳像算法包括二维稳像方法、三维

稳像方法和基于深度学习的稳像方法。相比于三维稳像方法和基于深度学习的稳像方法,二维稳像方法具有处理速度快的优点,容易做到实时稳像,近年来受到关注。二维视频稳像算法包括以下三个步骤:(1)估计原始摄像机路径(即全局运动);(2)对摄像机路径进行平滑或者拟合;(3)对原始路径进行运动补偿,生成稳定的视频。其中摄像机路径估计是关键环节。

目前常用的全局路径估计方法包括 Liu 等人<sup>[1]</sup>和 Yu 等人<sup>[18]</sup>的基于网格的图像局部的全局运动估计和 Grundmann 等人<sup>[2]</sup>的基于特征点跟踪的全局运动估计,其中 Bundled Paths 算法<sup>[1]</sup>虽然精度高但是计算复杂度高,耗时长,而 L1 范数优化算法<sup>[2]</sup>利用单一的运动模型,计算复杂度小,更容易做到实时稳像。但是由于原始路径存在高频抖动噪声,导致 L1 范数优化算法<sup>[2]</sup>路径拟合结果产生误差,从而使得稳定后的视频存在残余抖动,稳像效果不佳。如果能通过路径平滑将原始视频路径中的高频抖动噪声去除,使路径拟合结果更准确,则有望彻底去除抖动。另外,L1 范数优化算法中未考虑视频缩放尺度的路径拟合问题,导致稳定后的视频在缩放层面的平滑度不够。

针对以上问题,本文受<sup>[23][24]</sup>启发,提出了路径平滑步骤:通过 Savitzky-Golay 滤波(简称 S-G 滤波)算法去除相机原始路径的高频抖动噪声。经过处理后的相机运动路径去除了高频噪声,使得后续的路径拟合结果更准确。S-G 滤波算法<sup>[16][23]</sup>是一种信号处理领域常见的平滑算法,可以滤除信号中的高频分量,实现降噪,同时保留原信号形状,而且算法计算速度很快,适用于实时处理应用。

本文提出了一种 S-G 滤波和 L1 范数优化相结合的数字稳像方法。首先对原始视频图像序列进行基于特征点跟踪的全局运动估计;其次通过 S-G 滤波算法对提取的全局运动路径进行平滑处理,去除相机全局运动高频噪声;最后采用优化的 L1 范数优化算法进行稳定路径的拟合。通过对不稳定视频帧进行运动补偿,合成了摄像机的稳定视频序列。在视频稳定效果上,超过了原有的

L1 范数优化算法和 Wang 等人<sup>[21]</sup>的深度学习方法,略低于 Bundled Paths 算法,算法运行时间接近于 Bundled Paths 算法的 1/5。

## 2 相关工作

目前,已经有很多视频稳像的方法,包括二维、三维和基于深度学习的视频稳像方法都对视频有一定的稳像效果。

**二维稳像方法**的模型简单,计算复杂度低。首先估计相机的运动,提取运动轨迹。全局运动估计的方法包括下列方法:Chen 等人<sup>[3]</sup>利用块匹配运动估计得到参数运动模型;Erturk<sup>[4]</sup>使用相位相关算法进行全局运动估计;Xu 等人<sup>[23]</sup>人使用 Horn-Schunck 算法进行全局运动估计。Lucas 与 Kanade<sup>[5]</sup>的金字塔光流为很多稳像算法的全局运动估计提供了基础;基于 Harris 算法<sup>[6]</sup>、FAST 算法<sup>[7]</sup>、KLT 算法<sup>[8]</sup>等的特征点法构建变换矩阵描述相机的运动。其次,对相机的运动路径进行运动补偿,总体上有两种处理方式,运动路径平滑和运动路径拟合。运动路径平滑方法是对抖动相机的运动路径进行噪声的平滑,以消除高频和低频抖动。Yasuyuki 等人<sup>[9]</sup>使用高斯滤波器进行平滑;Erturk 等人<sup>[10]</sup>使用卡尔曼滤波器进行运动平滑,实现稳像;Xu 等人<sup>[23]</sup>和 Tsoligkas 等人<sup>[24]</sup>对相机路径进行基于 S-G 滤波器的路径平滑实现视频稳像。Jiang 等人<sup>[25]</sup>将利用原始视频序列生成广角视图,通过抽取视频帧合成稳定视频。Yang 等人将粒子滤波器首先引入视频稳像领域进行路径平滑。运动路径拟合方法模仿专业电影摄像技巧,将相机路径由静止、直线和抛物线组合而成,不仅能够去除抖动视频的高低频抖动,更可以使稳定后的视频更符合人观看视频的主观习惯。Chen 等人<sup>[11]</sup>将相机路径拟合成折线路径。Grundmann 等人<sup>[12]</sup>利用 L1 范数优化处理算法将相机路径用静止、直线、抛物线三种线条拟合,并分配不同的权重,取得了不错的效果。新兴的二维算法如 Liu 等人<sup>[1]</sup>将视频帧图像分为大小相等的网格进行基于“as-similar-as-possible”思想的多路径稳像算法等方法取得了优越

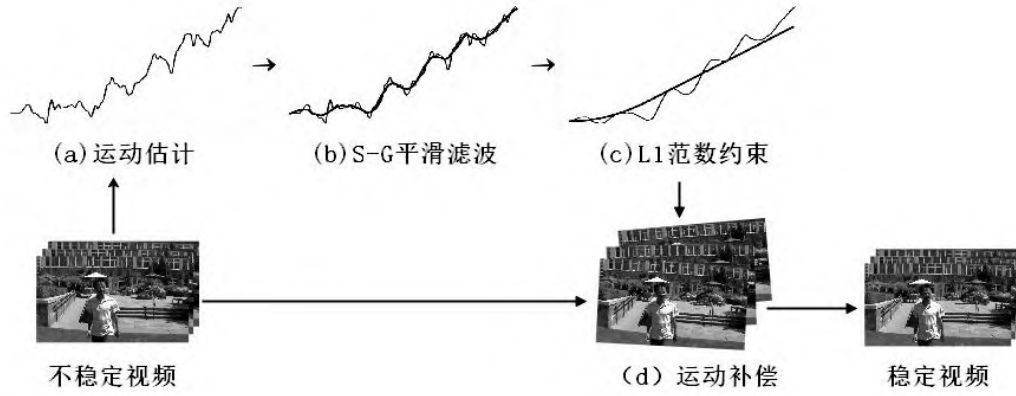


图1 算法流程图

Fig.1 algorithm flow chart

的效果,但是算法复杂性较高,处理速度慢于传统二维算法。

**三维稳像方法**和二维方法具有同样的处理步骤,估计相机的三位运动模型,进行运动路径平滑和生成稳定视频。Liu 等人<sup>[12]</sup>利用 SFM 算法进行相机路径估计,并进行运动平滑。最后利用 3D 场景点云和参考帧图像进行稳定视频的生成。三维稳像的运动路径平滑基于非线性滤波, Lee 等人<sup>[13]</sup>将运动数据投射到向量空间进行滤波,最后将数据投射回运动空间。

**基于深度学习的方法:** 随着卷积神经网络(CNN)的发展,近年来一些人利用预训练好的网络进行视频稳像工作。Wang 等人<sup>[21]</sup>提出了 Stabnet 模型,首次将卷积神经网络引入视频稳像任务,通过历史帧直接计算当前帧的单应性矩阵。Huang 等人<sup>[19]</sup>利用 Siamese 网络和之前的帧序列预测当前帧的运动矩阵。Yu 等人<sup>[20]</sup>直接将帧间的外观变化建模为连续帧的密集光流场。Lin 等人<sup>[22]</sup>提出了基于变形网格的算法,提高了相邻帧的连续性。端到端的算法无需设计复杂的平滑算法和拟合算法,只需通过对网络进行训练实现稳像,但是模型训练要求大量的稳定-不稳定配对的数据集。而且算法计算速度较慢,很难应用于实时计算。

### 3 本文方法

#### 3.1 方法概述

本文的方法主要分为四个部分,分别是运动估计、S-G 平滑滤波、L1 范数优化和运动补偿,如图 1 所示:首先对输入的视频

进行全局运动估计,分别检测出视频每帧图像的角点,每两帧之间检测出的角点进行配对,根据配对成功的角点计算每两帧之间的仿射变换矩阵。从仿射变换矩阵中提取出后一帧相对于前一帧的全局运动变量,即水平方向位移量、垂直方向位移量、倾斜角度、缩放尺度。将视频图形帧序列所有帧之间的全局运动变量分别叠加后,得到摄像机的全局运动路径。

其次,对全局运动路径进行基于 S-G 滤波算法的平滑处理,去除全局运动路径中的高频分量,消除高频抖动对后续运动路径拟合的影响。

进一步,对经过平滑的运动路径进行基于 L1 范数优化算法的路径拟合。本文对原算法进行了优化,在 Grundmann 等人<sup>[2]</sup>的基础上加入了缩放尺度的路径拟合,并改变了权重。得到了平滑的运动路径。

最后,计算稳定后每一帧的运动路径相对于原始运动路径的补偿量,对每一帧的图像进行运动补偿,得到稳定的视频。

#### 3.2 全局运动估计

首先,对输入的视频进行全局运动估计,计算相机的运动路径。两帧图像在二维空间内存在平移、倾斜、缩放等变换。本文采用了四参数仿射变换模型进行全局运动的估计。仿射变换矩阵模型如下所示:

$$\mathbf{A} = \begin{bmatrix} s \cdot \cos \theta & -s \cdot \sin \theta & x \\ s \cdot \sin \theta & s \cdot \cos \theta & y \end{bmatrix} \quad (1)$$

$x$  和  $y$  分别代表后一帧图像相对于前一帧图像的水平 and 垂直位移量,  $s$  代表缩放尺





图2 视频中第  $t$  帧和第  $t+1$  帧之间的匹配角点

Fig.2 matched corner in video frame  $t$  and video frame  $t+1$

度,  $\theta$  代表倾斜角度。利用两帧之间匹配的角点计算仿射变换矩阵, 本文采用 Shi-Tomasi 算法<sup>[14]</sup>进行角点检测。

角点是图像中的极值点。在同一场景内, 角点具有稳定的性质。Shi-Tomasi 算法<sup>[14]</sup>利用固定尺寸的滑动窗口在帧图像的灰度图上移动, 对于窗口在任意方向的移动都有明显灰度变换的情况, 定义该位置为角点。每帧图像检测 100 个角点, 得到第  $t$  帧的角点序列  $f_t$ 。

其次, 利用 Lucas-Kanade 稀疏光流法跟踪角点  $f_t$ , 得到角点在  $t+1$  帧的位置  $f_{t+1}$ 。

Lucas-Kanade 稀疏光流法计算两帧之间的稀疏光流, 并基于亮度恒定假设跟踪两帧之间相匹配的角点。两帧之间的角点跟踪结果如图2所示, 匹配角点用黑色直线相连。为方便观察, 图中仅展示部分角点的跟踪结果。

对匹配的角点利用 RANSAC 算法<sup>[15]</sup>进行随机抽样和多次迭代, 每次抽取 10 对匹配角点并迭代 20 次计算两帧之间的最优仿射变换矩阵  $A$ 。

### 3.3 基于 S-G 滤波算法的运动路径平滑

#### 3.3.1 S-G 滤波算法

S-G 滤波器 (也称卷积平滑) 被广泛地应用于数据流平滑去噪处理, 是一种在时域内基于区域性多项式最小二乘法拟合的滤波方法。最大的特点是去除信号噪声的同时保持信号的形状、宽度不变。

定义一个宽度为  $n=2m+1$  的窗口, 对原始曲线进行自左至右的测量。考虑一组以数据点  $y_t$  为中心的  $2m+1$  个数据, 对这组数据点进行基于  $k-1$  次多项式的拟合, 如式(2)所示。

$$y = a_0 + a_1 \cdot x + a_2 \cdot x^2 + \dots + a_{k-1} \cdot x^{k-1} \quad (2)$$

为使方程有解, 一般令  $n$  大于  $k$ 。对于  $2m+1$  个数据, 分别有  $2m+1$  个方程, 因此可以将拟合方程表示为矩阵形式。设数据值向量为  $Y$ , 系数矩阵为  $A$ , 自变量矩阵为  $X$ , 残差为  $\varepsilon$ 。方程可写为如下形式。

$$Y = X \cdot A + \varepsilon \quad (3)$$

经过最小二乘法解得的稀疏矩阵为  $A^\wedge$ , 其值如式(4)所示。拟合后的数据预测值为  $Y^\wedge$ , 其值如式(5)所示。

$$A^\wedge = (X^T \cdot X)^{-1} \cdot X^T \cdot Y \quad (4)$$

$$Y^\wedge = X \cdot A^\wedge = X \cdot (X^T \cdot X)^{-1} \cdot X^T \cdot Y \quad (5)$$

窗口自左至右滑动, 直至拟合完所有数据点。拟合后的曲线即被去除了原有的高频分量。

#### 3.3.2 运动路径平滑

原始路径中存在大量低高频分量, 其中大部分来源于拍摄者手部的快速抖动造成的摄像机镜头抖动。本文对相机的原始路径进行拟合, 从而生成稳定的相机路径。而原始运动路径中的高频分量噪声对于路径拟合是无益的, 因此需要首先将其去除。对高频分量的平滑操作可以选取上述 S-G 滤波算法, 具体过程如下。

将每两帧之间的运动分量分别累加, 得到原始视频的运动路径, 如式(6)所示,  $t$  为帧数,  $k$  是视频帧数。同理可求出  $C_y$ 、 $C_\theta$ 、 $C_s$ , 分别代表四个维度的相机运动轨迹。

$$C_x = C_x^t + \sum_{t=1}^{k-1} x^t, \quad C_x^t = 0 \quad (6)$$

将得到的运动轨迹分别进行平滑。通过实验发现, 缩放维度的运动路径不存在高频分量, 经过滤波处理后会增大误差。因此将路

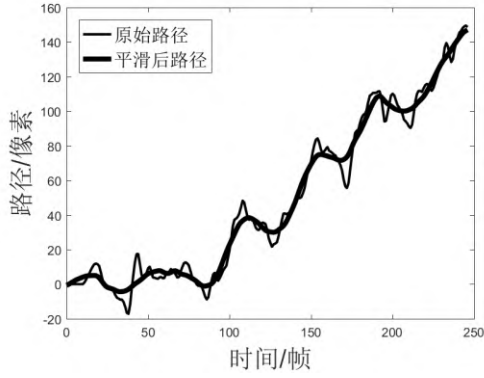


图3 S-G 平滑效果图

Fig.3 S-G smooth renderings

径的值  $C_x$ 、 $C_y$ 、 $C_\theta$  分别通过 S-G 滤波器进行平滑处理。平滑效果如下图所示。图中细曲线为相机原始的水平维度运动路径  $C_x$ ，粗曲线为平滑后的运动路径，横坐标为时间（帧），纵坐标为运动路径（像素）。经过平滑后，原始运动路径的高频分量被成功去除，同时保留了包含有效信息的低频分量，更有利于后续的路径拟合处理。

### 3.4 基于优化的 L1 范数优化的路径拟合

为了得到稳定的相机运动路径，同时让相机路径更符合专业相机的拍摄风格，本文采用了 Grundmann 等人<sup>[2]</sup>的 L1 范数优化算法，并对算法进行了优化，增加了稳像的效果。

将经过平滑后的相机运动路径定义为  $C_t^s$ 。每两帧之间的全局运动变量定义为运动矢量为  $F_t^s$ 。 $C_t^s$  可以如式(7)所示化简。

$$C_t^s = C_{t-1} F_t^s = F_t^s F_{t-1}^s \dots F_1^s \quad (7)$$

为了得到稳定的视频帧序列，需要分别对每帧图像进行运动补偿，定义补偿量为  $B_t^L$ ， $B_t^L$  中包含水平、垂直、倾斜角度、缩放尺度四个维度的补偿值。定义稳定后的相机运动路径为  $P_t$ ，那么  $P_t$  可以表示为

$$P_t = C_t B_t^L \quad (8)$$

稳定后的相机路径由下面三种线条进行拟合：

(1) 相机静止，即相机运动路径的一阶导数  $D(P)$  为零。使用前向差分公式计算  $D(P)$ ，并将式(7)(8)代入公式并化简，即

$$|D(P)| = \sum_t |P_{t+1} - P_t| = \sum_t |C_{t+1} B_{t+1}^L - C_t B_t^L| =$$

$$\sum_t |C_t F_{t+1} B_{t+1}^L - C_t B_t^L| \leq \sum_t |C_t| |F_{t+1} B_{t+1}^L - B_t^L|,$$

$$F_{t+1} B_{t+1}^L - B_t^L = R_t \quad (9)$$

因此，最小化  $D(P)$  即最小化残差  $R_t$ 。

(2) 相机匀速移动，即相机运动路径的二阶导数  $D^2(P)$  为零，即

$$|D^2(P)| = \sum_t |DP_{t+2} - DP_{t+1}| = \sum_t |P_{t+2} - 2P_{t+1} + P_t| \leq |R_{t+1} + R_t| \quad (10)$$

最小化  $D^2(P)$  即最小化残差差分

$R_{t+1} - R_t$ 。

(3) 相机匀加速，即机运动路径的三阶导数  $D^3(P)$  为零，即

$$|D^3(P)| \leq |R_{t+2} - 2R_{t+1} + R_t| \quad (11)$$

最小化  $D^3(P)$  即最小化  $R_{t+2} - 2R_{t+1} + R_t$ 。

因此，路径拟合的目标函数为：

$$O(P) = \omega_1 |D(P)| + \omega_2 |D^2(P)| + \omega_3 |D^3(P)| \quad (12)$$

其中， $\omega_1$ 、 $\omega_2$ 、 $\omega_3$  是三种线条的权重，代表拟合后曲线中相应线条所占的比例。原始算法对相机水平、垂直、倾斜角度进行了基于公式(12)的路径拟合，本文在此基础上将缩放尺度考虑在内，对缩放尺度的路径同样进行了拟合，并基于本文的任务设置了稳像效果更好的权重。最小化目标函数后，便可得到补偿量  $B_t^L$  的最优解。

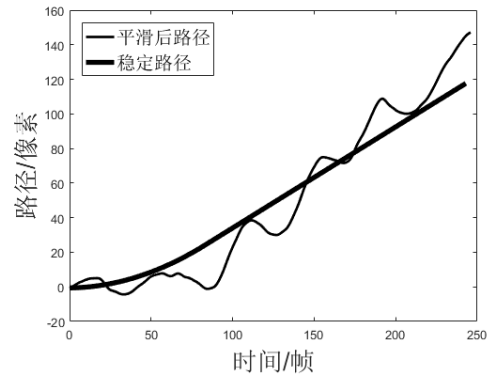


图4 路径拟合效果图

Fig.4 Path fitting renderings

拟合效果如图4所示。图中粗曲线是平滑后的相机水平方向路径  $C_t^s$ ，细曲线是利用  $C_t^s$  拟合的水平方向稳定路径。

### 3.5 运动补偿

经过路径平滑和路径拟合步骤后得到

了稳定的相机运动路径  $P_t$ 。需要根据稳定的



图5 裁剪范围示例

Fig.5 Clipping range example

相机运动路径  $P_t$  和原始相机运动路径  $C_t$  之间的关系（即补偿量  $B_t$ ）得到稳定视频的图像帧序列，生成稳定的视频。由于原始相机路径一共经过了两次处理（路径平滑和路径拟合），因此补偿量由两部分组成，定义如下：

$$B_t = B_t^L + B_t^S \quad (13)$$

$B_t^L$  已经通过 3.4 节中介绍的路径拟合算法求得，还要求解  $B_t^S$ 。 $B_t^S$  可以由平滑后的路径和原始路径求解，数学表示为：

$$B_t^S = C_t^S - C_t \quad (14)$$

原始相机运动路径  $C_t$  经过运动补偿量  $B_t$  进行补偿后得到稳定视频路径  $P_t$ ，如公式(15)所示。

$$P_t = C_t B_t \quad (15)$$

将公式(14)代入到公式(15)，得到最终的运动补偿公式：

$$P_t = C_t (B_t^L + C_t^S - C_t) \quad (16)$$

对公式(16)的求解即可得到稳定后的视频帧图像序列。

进行运动补偿后，图像的位置和姿态发

生了改变，会在视频边界处产生空白区域，影响视觉效果。因此需要对图像进行裁剪。裁剪值为运动补偿量  $B_t$  中的水平和垂直方向的最大补偿值  $Bx(max)$  和  $By(max)$ 。将经过裁剪的图像序列恢复成视频后，即可得到稳定视频。

## 4 实验

### 4.1 实验细节

借鉴文献[17]的方法，视频的稳定效果可以通过峰值信噪比（PSNR）评估。

一段绝对稳定的视频图像序列的运动轨迹十分平滑，每两个连续帧之间的像素误差很小。因此，一段视频图像序列的每两个连续帧之间的 PSNR 值越大，序列就越稳定。本文测量了每段视频的平均 PSNR 值作为评估指标。

$$PSNR_{(avg)} = \frac{1}{N-1} \sum_{i=1}^{N-1} 10 \log \frac{255^2}{MSE(I_i, I_{i+1})} \quad (17)$$

在 S-G 滤波平滑部分，进行了对比实验，对窗口宽度  $n$  和拟合多项式的阶次  $k$  两个参数进行了不同的取值，观察同一段视频的稳像结果，如图 6 所示。纵坐标是多项式的阶次



$k$ , 横坐标是窗口宽度  $n$ , 图中颜色越深表示

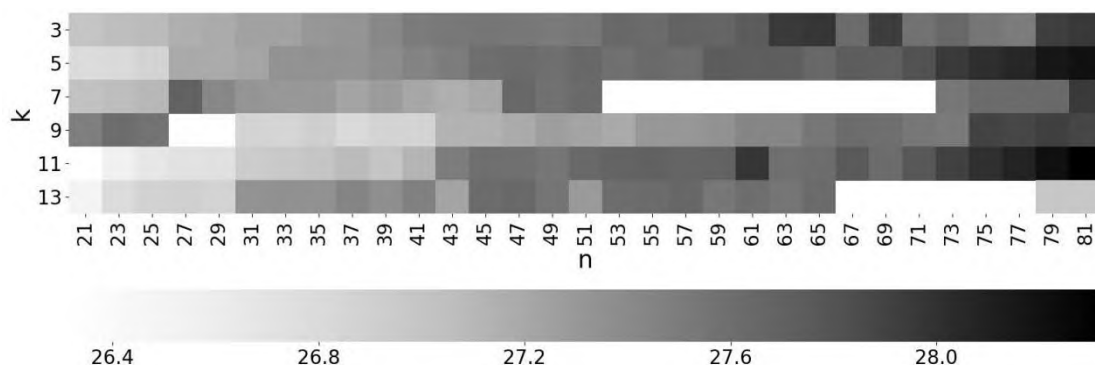


图6 S\_G 滤波对比实验结果

Fig.6 contrast experiment result of S\_G filter

PSNR 值越大。从图中看出,  $k$  取 7、9、13 时出现了稳像失败的情况, 随着  $n$  值的升高, PSNR 值越大。同时观察  $n$  在高取值范围的稳像视频效果, 发现某些视频的裁剪率过大, 导致视频不完整。因此, 为了保持稳像效果、视频完整性以及算法实时性之间的平衡, 本文将窗口宽度  $n$  设置为 63, 将多项式阶次  $k$  设置为 3。

本文用 python 实现本文的方法。实验数据集来自于文献[1]提供的稳像数据集, 包含 6 种场景共 144 段视频 (5~20s), 分别是: (1) Crowd, (2) parallax, (3) Quick Rotation, (4) Regular, (5) Zooming, (6) Running。

本文选择了近年来在二维稳像方法中稳像效果较好的 Liu 等人<sup>[1]</sup>的 Bundled Paths 方法、本文的基准模型 Grundmann 等人<sup>[2]</sup>的 L1 范数优化算法和当前结果较好的 Wang 等人<sup>[21]</sup>的 Deep Learning 方法进行对比实验。

## 4.2 结果分析

### 4.2.1 运行时间的比较

表1 运行时间的比较

Fig.1 Run time comparison

视频稳像方法	分辨率/像素	时间/ms
Bundled Paths	1280×720	392
L1 范数优化	1280×720	75
Yu 等人	—	6000
本文方法	1280×720	82
Deep Learning	640×360	3

本文测量了几种方法处理同样分辨率的一帧图像所需要的时间。结果如表 1 所示。

处理同样分辨率的视频, L1 范数优化算法的 75ms/帧比本文方法的 82ms/帧稍快, 而二者均明显快于 Bundled Paths 方法的 392ms/帧。而 Yu 等人<sup>[20]</sup>的基于深度学习算法处理速度明显慢于其他方法, 不利于稳像的实时性。Wang 等人<sup>[21]</sup>的 Deep Learning 方法处理速度快, 可以做到实时稳像。但同时具有网络模型复杂, 参数量大的缺点, 需要 21.3Mb 内存和 GPU 运算, 而本文方法模型仅有 35kb 大小和 CPU 运算, 更适合轻量级的设备。

### 4.2.2 稳像效果对比

视频的稳像效果由 PSNR 值评估。不同方法的稳像视频 PSNR 值如图 7 所示。

通过实验结果可以看出, 大部分情况下 Bundled Paths 方法比本文方法略好, 从第(c)段视频的实验结果可以看出, 本文方法的结果在某些情况下好于 Bundled Paths 方法。从整体来看, 本文的方法优于 L1 范数优化算法结果和 Deep Learning 方法。

本文方法优于 L1 范数优化算法和 Deep Learning 方法的原因是引入了基于 S-G 滤波的路径平滑算法。本文进行了路径平滑实验, 以图 7 第(a)段视频为例进行分析, 实验结果如图 8 所示。(a)(c)中曲线为相机在垂直方向的原始路径; (b)为 Deep Learning 方法对垂直运动路径的稳像效果, 实线为稳像结果。(d)为利用 L1 范数优化算法对相机垂直方向路径直接进行路径拟合的结果, 实线为拟合后的结果; (e)为本文方法中的 S-G 滤波平滑结果, 对原始路径进行去噪处理, 去除原始路径中的高频分量, 平滑后的结果为图



中实线；对平滑后的曲线进行基于优化的 L1 范数优化算法的路径拟合，结果如(f)所

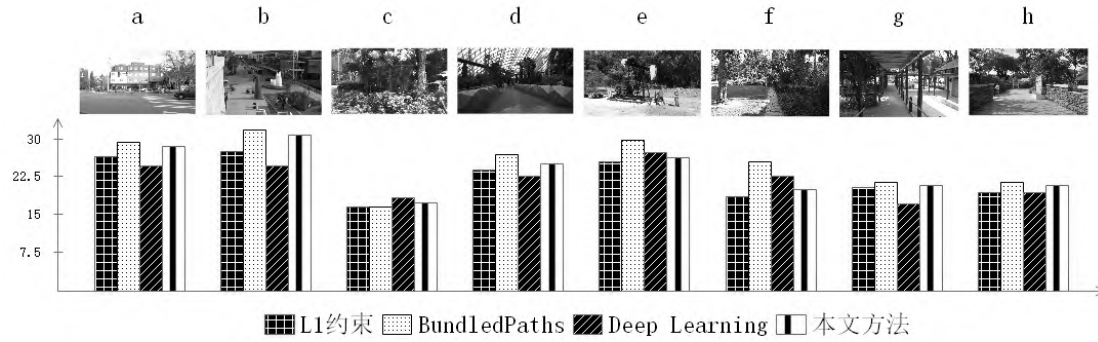


图 7 稳像实验结果

Fig.7 video stabilization experiment results

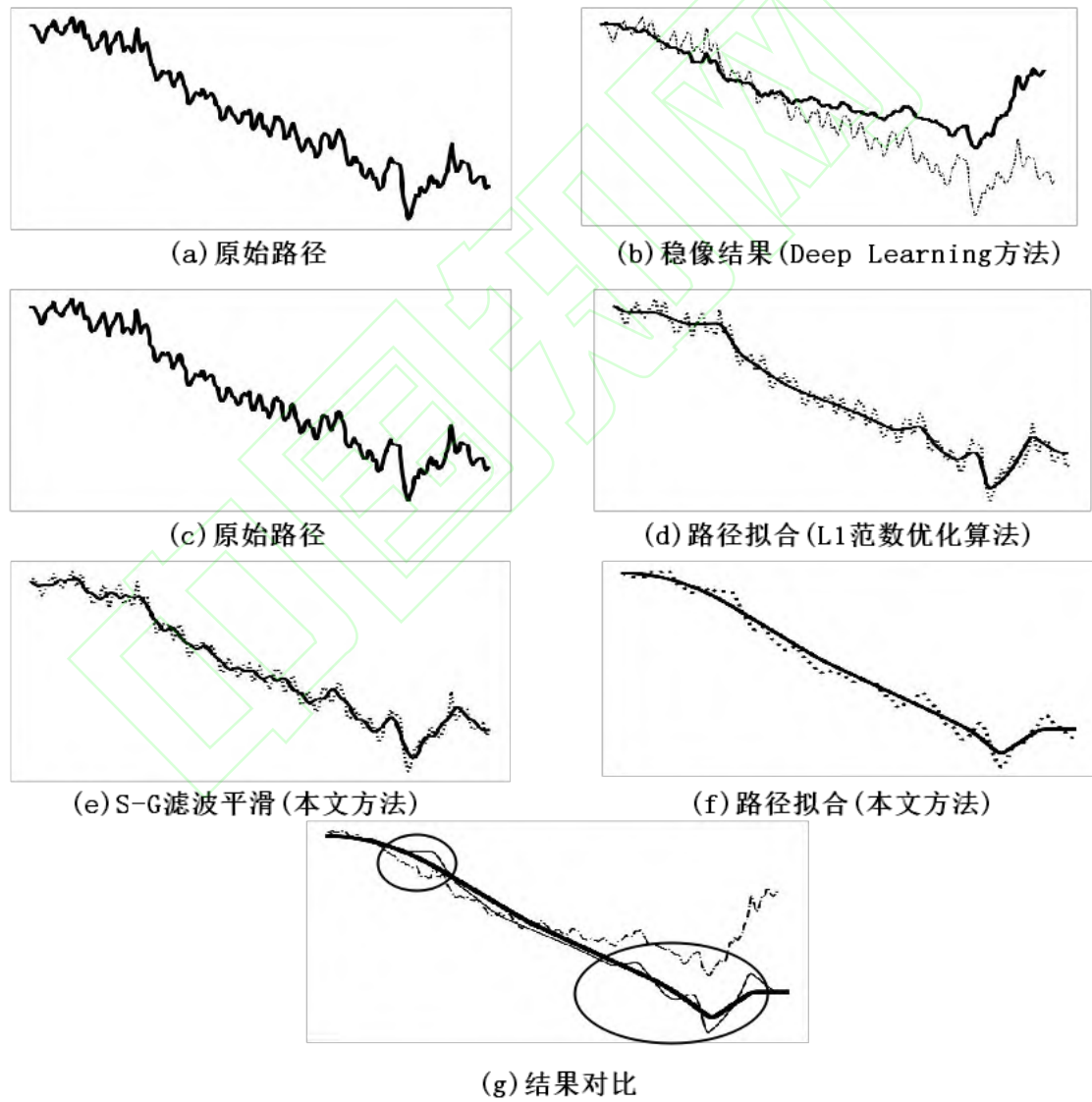


图 8 路径平滑对比实验.(a):视频原始路径.(b):Deep learning 方法的稳像结果.(c):视频原始路径.(d):L1 范数优化算法的路径拟合结果.(e):本文方法的 S-G 滤波平滑结果.(f):本文方法的 L1 范数路径拟合结果.(g):三种方法的输出路径结果对比

Fig.8 Path smoothing contrast experiment.(a):Video original path.(b):Stabilization result of the Deep Learning

method(c):Video original path.(d):Path fitting results of L1 norm optimization algorithm.(e):The s-g filter smooths result in this paper.(f):The L1 norm path fitting results of the method in this paper.(g):comparison of the two output path results of the two methods



图9 本文方法与 Deep Learning 方法及 L1 范数优化方法的稳像效果对比. (a)-(c),(d)-(f),(g)-(i)为三种方法对不同视频的稳像结果对比.

Fig.9 Comparison of stabilization results by the proposed method,Deep Learning method and L1 norm optimization algorithm.(a)-(c),(d)-(f),(g)-(i) are the comparison of the results of three methods for different videos.

示, 实线为稳像后的输出结果。

对比三种方法合成的稳定路径, 如(g)所示。虚线为 Deep Learning 方法的结果, 细实线为 L1 范数优化算法的结果, 粗实线为本文方法得到的结果。可以看出, 虚线存在较多低频与高频噪声, 稳像效果不佳, 稳像后的路径相对于原始路径产生了明显偏移; 细实线在圆圈标记处出现了抖动情况, 粗实线相对于细实线更平滑。由此得出结论, 通过 S-G 滤波平滑算法去除原始路径的高频分量, 得到的有效路径信息能够拟合生成更加稳定的路径。

稳定的路径在视频上体现为画面的稳定。如图 9(a)-(c)所示, 分别代表三种方法对图 7 第(a)段视频的稳像结果, 从左至右分别为时间上连续的视频帧。由于 Deep Learning 方法的稳像效果不佳, 稳像后的视频垂直路径明显偏移于原始视频, (a)中雕塑头部被裁剪掉, 丢失了原始视频的信息。(b)画面中的雕塑随时间的变化在垂直方向出现了抖动, 而(c)中雕塑在垂直方向上没有出现明显的抖情况。图 9(d)-(f)为三种方法对图 7 第(h)段视频的稳像效果对比。可以看出, 图中(d)行 Deep Learning 方法的稳像视频随时间的变化在水平方向产生了较大的位移和抖动, 而(f)行本文方法的稳像视频在水平方向移动更为平滑, 没有产生明显抖动。同时, 相对于(e)行 L1 范数优化算法稳像视频, 本文方法的稳像视频保留了更多地原始视频信息。图 9(g)-(i)为三种方法对图 7 第(d)段视频的稳像效果对比。第(i)行本文方法的稳像结果相对于第(g)行 Deep Learning 方法的稳像结果保留了更多的原始视频信息, 同时在水平方向、垂直方向和倾斜角度三个维度保持了更好的稳定性和平滑度。

实验结果表明, 本文方法的稳像效果超越了原始 L1 范数优化算法和 Wang 等人<sup>[21]</sup>的 Deep Learning 方法, 在计算时间上快于其他二维稳像算法, 在模型复杂度上优于深度学习方法。

## 5 结论

本文提出了一种 S-G 滤波算法和 L1 范

数优化相结合的稳像方法, 利用帧间对应特征点进行运动估计, 有效去除了原始路径中的高频噪声, 利用原始路径中的有效信息通过优化的 L1 范数优化算法生成效果更好的稳定路径。实验证明, 相对于其他二维稳像算法, 本文提出的方法耗时更少, 相对于原始 L1 范数优化算法和基于深度学习的稳像算法, 本文更彻底地去除了原始视频路径的不稳定抖动。

## 参考文献

- [1] Liu S, Yuan L, Tan P, et al. Bundled camera paths for video stabilization[J]. ACM Transactions on Graphics (TOG), 2013, 32(4): 1-10.
- [2] Grundmann M, Kwatra V, Essa I. Auto-directed video stabilization with robust l1 optimal camera paths[C]//CVPR 2011. IEEE, 2011: 225-232.
- [3] Chen T. Video stabilization algorithm using a block-based parametric motion model[J]. Information System Laboratory, Department of Electrical Engineering, Stanford University, 2000.
- [4] Erturk S. Digital image stabilization with sub-image phase correlation based global motion estimation[J]. IEEE transactions on consumer electronics, 2003, 49(4): 1320-1325.
- [5] Lucas B, Kanade T. An iterative image registration technique with an application to stereo vision [C] //Proc of Int Joint Conf on Artificial Intelligence. Menlo Park, CA: AAAI, 1981: 674-679
- [6] Harris C G, Stephens M. A combined corner and edge detector[C]//Alvey vision conference. 1988, 15(50): 10-5244.
- [7] Rosten E, Porter R, Drummond T. Faster and better: A machine learning approach to corner detection[J]. IEEE transactions on pattern analysis and machine



- intelligence, 2008, 32(1): 105-119.
- [8] Shi Jiaobo, Tomasi C. Good features to track [C] //Proc of IEEE Computer Society Conf on Computer Vision and Pattern Recognition Los Alamitos, CA: IEEE Computer Society, 1994:593-600
- [9] Matsushita Y, Ofek E, Ge W, et al. Full-frame video stabilization with motion inpainting[J]. IEEE Transactions on pattern analysis and Machine Intelligence, 2006, 28(7): 1150-1163.
- [10] Erturk S. Real-time digital image stabilization using Kalman filters [J]. Real-Time Imaging, 2002, 8(4): 317-328
- [11] Chen B Y, Lee K Y, Huang W T, et al. Capturing Intention-based Full-Frame Video Stabilization[J]. Computer Graphics Forum, 2008, 27(7):1805-1814.
- [12] Liu F, Gleicher M, Jin H, et al. Content-preserving warps for 3D video stabilization[M]// ACM Transactions on Graphics (TOG). ACM, 2009.
- [13] Lee J, Shin S Y. General construction of time-domain filters for orientation data[J]. IEEE Transactions on Visualization and Computer Graphics, 2002, 8(2): 119-128.
- [14] Shi J. Good features to track[C]//1994 Proceedings of IEEE conference on computer vision and pattern recognition. IEEE, 1994: 593-600.
- [15] Fischler M A, Bolles R C. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography[J]. Communications of the ACM, 1981, 24(6): 381-395.
- [16] Schafer R W. What is a Savitzky-Golay filter?[lecture notes][J]. IEEE Signal processing magazine, 2011, 28(4): 111-117.
- [17] Liu S, Tan P, Yuan L, et al. Meshflow: Minimum latency online video stabilization[C]//European Conference on Computer Vision. Springer, Cham, 2016: 800-815.
- [18] Yu J, Ramamoorthi R. Selfie video stabilization[C]//Proceedings of the European Conference on Computer Vision (ECCV). 2018: 551-566.
- [19] Huang C H, Yin H, Tai Y W, et al. StableNet: Semi-Online, Multi-Scale Deep Video Stabilization[J]. arXiv preprint arXiv:1907.10283, 2019.
- [20] Yu J, Ramamoorthi R. Robust video stabilization by optimization in cnn weight space[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2019: 3800-3808.
- [21] 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.
- [22] Lin K, Jiang N, Liu S, et al. Direct photometric alignment by mesh deformation[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2017: 2405-2413.
- [23] Xu W, Lai X, Xu D, et al. An integrated new scheme for digital video stabilization[J]. Advances in Multimedia, 2013, 2013.
- [24] Tsoligkas N A, Xu D, French I, et al. A motion model based video stabilisation algorithm[C]//2006 World Automation Congress. IEEE, 2006: 1-6.
- [25] 蒋双霜,朱振峰,常冬霞. 利用全景图的视频稳像算法[J]. 信号处理, 2018,34(7): 852-859
- Jiang Shuangshuang, Zhu Zhenfeng, Chang Dongxia. Video Stabilization Algorithm Using Generated Panorama[J]. Journal of Signal Processing, 2018,34(7): 852-859(in Chinese)



## 作者简介



**杨雨辰** 男，1997 年生，北京人。北京工业大学硕士研究生，主要研究方向为视频稳像技术和群体行为识别。  
E-mail: bide333@163.com



**毋立芳**（通信作者） 女，1970 年生，山西人。北京工业大学教授，博士生导师，主要研究方向为图像视频内容分析、社交媒体计算、智能 3D 打印。  
E-mail: lfwu@bjut.edu.cn



**简萌** 女，1987 年生，内蒙古人。北京工业大学讲师，主要研究方向为图像、视频大数据多媒体内容分析，社交多媒体用户个性化推荐，图像分类与检索。  
E-mail: jianmeng648@163.com



**赵博煊** 男，1997 年生，北京人。北京工业大学硕士研究生，主要研究方向为视频目标跟踪。  
E-mail: 871235843@QQ.com



**刘旭** 男，1991 年生，安徽滁州人。北京工业大学博士后研究员，主要研究方向为动作识别、目标检测、分割和深度学习。  
E-mail: liuxu91@bjut.edu.cn