

Orientation Optimization Based 3D Video Stabilization for Portable Cameras

Chengdu Yang¹, Hongli Xu¹, Hui Yin^{1,2}, Hua Huang^{1,2}

¹College of Computer Science and Technology, Beijing Jiaotong University, Beijing

²Beijing Key Lab of Transportation Data Analysis and Mining, Beijing Jiaotong University, Beijing

Email: 13120440@bjtu.edu.cn

Received: Dec. 5th, 2015; accepted: Dec. 25th, 2015; published: Dec. 28th, 2015

Copyright © 2015 by authors and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

We present a novel view optimization algorithm in the paper for video stabilization. Most existing 2D video stabilization methods may fail to estimate the motion parameters of cameras. In order to solve above problems, we propose a video stabilization method based on 3D technology. Firstly, recovers the original 3D camera motion using the structure-from-motion system. Then, a virtual camera path is computed by polynomial curve fitting. And then, smoothes the view orientation by vectorial interpolation. Finally, corrects the frames which selected in conformity with the smoothing location and orientation of cameras. Our experiments on stabilizing challenging videos of real scenes demonstrate the effectiveness of our technique.

Keywords

Video Stabilization, Rotation Optimization, Polynomial Curve-Fitting

基于视角优化的便携像机视频去抖算法

杨诚笃¹, 许宏丽¹, 尹辉^{1,2}, 黄华^{1,2}

¹北京交通大学计算机学院, 北京

²北京交通大学交通数据分析与挖掘北京市重点实验室, 北京

Email: 13120440@bjtu.edu.cn

收稿日期: 2015年12月5日; 录用日期: 2015年12月25日; 发布日期: 2015年12月28日

摘要

针对便携式摄像机的视频去抖问题,本文提出一种基于摄像机视角优化的视频去抖算法。首先,通过structure-from-motion算法估计摄像机连续运动位姿。其次,基于摄像机的空间位置进行多项式曲线拟合,得到摄像机运动的虚拟路径。然后,通过向量插值平滑视角朝向。最后,基于相机运动平滑曲线与平滑视角对视频帧进行筛选和修正实现视频去抖。实验证明,该算法很好地解决了2D去抖算法对摄像机运动参数估计不足的问题。在真实数据集上的实验结果验证了本文算法在针对便携式摄像机视频数据的去抖有效性。

关键词

视频去抖, 视角优化, 多项式拟合

1. 引言

随着家用便携摄像设备的普及,非专业拍摄的视频大量涌现。由于缺乏良好的固定措施和拍摄技巧,这些非专业视频往往存在视频视角晃动大、可观赏性不高的问题。一般可以采用去除抖动的方法对视频进行处理,提升视频的观赏质量。因此改善视频的可观赏性就成为非专业视频优化的一个主要挑战。

视频去抖是视频处理的中心问题之一,常见的视频去抖动主要是基于2D方法。

基于2D的方法主要通过估计单应性矩阵[1],如基于运动矢量的视频去抖动算法,通过块运动估计来估算摄像机的全局运动参数[2]-[4]。利用RANSAC算法[5]或图像的SIFT特征匹配[6]增强全局运动参数估计的鲁棒性。目前最先进的基于2D的去抖算法是由Shuaicheng Liu提出的对同一时段多特征点轨迹平滑[7]。基于2D的方法虽然鲁棒性较强、计算代价较小,但不能合理描述相机的运动模型,无法对其运动路径和视角进行优化[8]。

基于3D的视频去抖技术以多视角几何为基础[9]-[11],通过构建相机运动模型并平滑相机运动参数以达到视频去抖的目的。相机抖动是产生视频抖动的根本原因,相机运动参数可直观反应相机运动,因此通过平滑相机运动参数去抖更加合理。

便携式相机的拍摄内容带有拍摄者的主观随意性,其视频图像往往不满足一般的图像处理技术所必需应用条件,一般的去抖方法在处理此类视频时的效果欠佳。鉴于便携式摄像机视频中造成抖动的主要原因在于摄像机拍摄视角的非均匀变化,本文基于对视角优化进行运动补偿的思路,提出一种基于视角优化的3D视频去抖算法。主要贡献有:通过摄像机运动估计,对视频帧进行筛选和视角优化。实验结果表明,本文提出的方法能够解决2D去抖算法对摄像机运动估计不足的问题。

2. 基于视角优化的去抖算法

利用便携式相机进行视频拍摄时,相机的运动路径一般较为简单、往复运动少,因此摄像机位置的短时震动对画面的可观赏性影响影响较小,使得视角的非均匀变化成为导致视频抖动的主要原因。因此,可以在估计摄像机位置的基础上,对相机位置进行曲线拟合,得到相机的主运动趋势后对相机运动参数进行筛选,并对相机朝向进行平滑处理。最后,根据相机视角平滑处理的结果对所选的帧进行视角变换,从而弱化或消除视角剧烈变化造成的视频抖动。

2.1. 算法原理

首先,通过运动结构重建(structure-from-motion, SFM)[12]恢复摄像机运动参数,并估算其姿态。运

动结构重建由特征检测、KLT 跟踪重构帧间关系，估算相机的运动参数，主要是相机位置与朝向。

其次，由最小二乘法对相机实际位置拟合，得到相机运动的平滑曲线。最小二乘法通过各个时刻相机实际位置拟合出与所有相机位置距离偏差最小的高次平滑曲线。此曲线反映相机平滑的整体运动的趋势。

再次，通过球面线性插值平滑相机视角。对任一时刻相机朝向和相机初始朝向进行球面线性插值，在不改变相机朝向运动趋势的同时，平滑相机朝向运动幅度。

最后，基于相机运动平滑曲线与平滑视角对视频帧进行筛选和修正。

其原理如图 1 所示。

2.2. 算法流程

视频去抖算法：

输入： I_1, I_2, \dots, I_N (视频中的连续帧序列)

输出： I'_1, I'_2, \dots, I'_M (去抖后的视频帧序列)

1. 基于 SFM 算法估算相机在不同时刻的旋转矩阵 $R = \{R_1, R_2, \dots, R_N\}$ 和平移向量 $t = \{t_1, t_2, \dots, t_N\}$ ，计算摄像机各个时刻的位置 $P = \{P_1, P_2, \dots, P_N\}$ 和朝向 $D = \{D_1, D_2, \dots, D_N\}$ ；对相机的位置 P 进行曲线拟合(2)，得到摄像机主运动趋势的近似路径 $C = \{C_1, C_2, \dots, C_M\}$ ；
2. 针对每一帧相机朝向 D_i ，通过向量插值公式(6)，估算路径中 C_i 处朝向平滑的相机朝向 \hat{Q}_1 。
3. 对路径 C 中的每个位置，执行以下处理过程：
 - 1) 做相机到拟合曲线 C 的投影，找到在每个路径位置点 C_i 有投影的摄像机位置 $\{P_i^j\}$ 。
 - 2) 若某位置点不存在相机投影，则补充相邻路径位置的投影相机位置。
 - 3) 由公式(7)在摄像机位置集合 $\{P_i^j\}$ 中选取距离路径位置点最近，且朝向与平滑朝向 O_i 偏差最小的帧为待变换帧 N_i 。
 - 4) 由公式(8)对输入视频帧做视角变换。

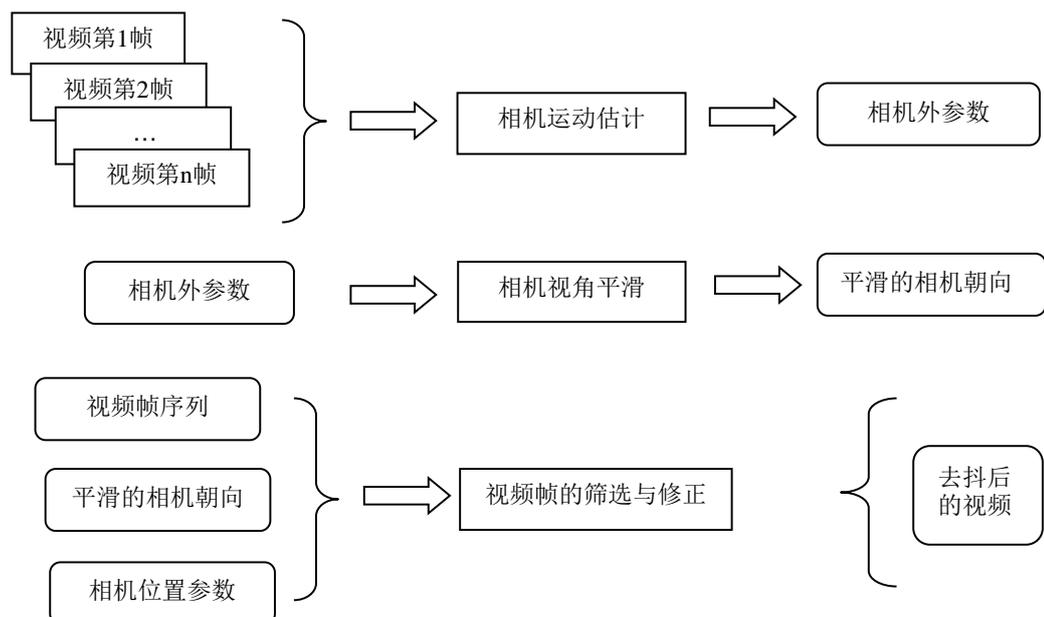


Figure 1. The fundamental of the proposed algorithm

图 1. 算法原理图

3. 摄像机运动估计

估计摄像机的运动趋势可以分离摄像机的运动趋势和抖动。视频拍摄过程中,摄像机的物理位置是连续变化的,高质量的专业视频拍摄中通常不存在无规则的随机抖动。对于非专业的便携式相机视频,缺乏专业的去抖处理,因而需要一定的手段估计摄像机的运动趋势,以平滑抖动。本文通过运动结构重建估算相机的位姿,并分别通过最小二乘法 and 球面线性插值法对位姿做平滑处理。

3.1. 运动结构重建

运动结构重建(SFM)[12]算法基于多视角几何,通过多幅图像的两两特征匹配,生成摄像机的运动参数和特征点的三维点云数据。如图2所示。

首先,对视频序列帧 Image1、Image2、Image3 中的特征点 (x_1, y_1) 、 (x_2, y_2) 、 (x_3, y_3) 进行两两匹配。

其次,寻找基线最大、匹配最好的两帧图像,并且基于其特征匹配估算本征矩阵,由本征矩阵还原相机外参并三角化特征点三维位置。

最后,逐步加入新的视频帧进行外参与特征点三维位置的计算,调整世界坐标系原点,直到所有视频帧都完成三维位置计算。

3.2. 摄像机位姿估算

最小二乘法是一种数学回归技术。其算法通过最小化均方误差的思想,寻找离散数据的最佳匹配曲线。

对数学上一条曲线,可通过高次多项式进行拟合。通过最小二乘法,可由一组离散坐标点拟合出一条连续运动曲线的高次多项式表达式,以预测曲线上某点的坐标值。

由于非专业的便携式相机视频一般是帮助拍摄者回顾浏览路径的,一般不会有针对统一内容的反复拍摄。因此这类视频一般浏览路径简单,且主运动方向上的往复运动时间持续短暂,使得通过曲线拟合摄像机运动路径成为可能。如图3所示,由最小二乘拟合,基于每一时刻的摄像机真实位置估算出主运动趋势的平滑曲线,可以反映真实摄像机的主运动。

记 i 时刻相机坐标系与世界坐标系之间的旋转矩阵和平移向量为 R_i 和 t_i , 则相机在世界坐标系下的位置 P_i 表示为:

$$P_i = -R_i^T t_i \quad (1)$$

式中 R_i^T 是 R_i 的转置矩阵。

对摄像机位置 P 做最小二乘的多项式拟合得路径位置 C :

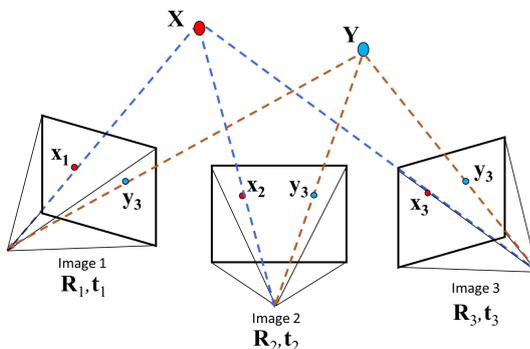


Figure 2. The reconstruction of feature points' positions and camera's motion model by SFM

图2. SFM 算法对特征点的空间位置和摄像机运动模型的恢复

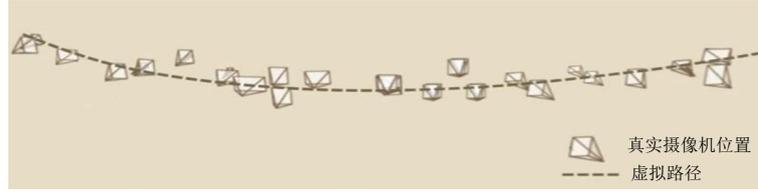


Figure 3. The real positions and the planning path of the camera
图 3. 摄像机实际运动位置与优化后的虚拟摄像机路径

$$\min_C \sum_{i=1}^N |\delta_i| = \sum_{i=1}^N |C_i - P_i| \quad (2)$$

式中 C_i 表示时刻 i 空间三维曲线的位置， δ_i 表示在 i 时刻曲线 C 与摄像机真实位置 P_i 的偏差。

4. 相机视角平滑算法

球面线性插值是一种可以保持向量长度不变的线性插值方法，用于计算向量间角度变化的插值。以四元数为工具的球面线性插值算法[13]，可以计算等角速度变化的向量角度线性插值，以达到使向量角度均匀变化的目的。

令相机朝向为相机坐标系中 Z 轴的正方向，记 i 时刻相机的旋转矩阵为 R_i ，则相机在世界坐标系下的朝向 D_i 可表示为：

$$D_i = R_i^T \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \quad (3)$$

式中 R_i^T 是 R_i 的转置矩阵。

由四元组 $Q_i = \langle w_i, v_i \rangle$ 表示相机朝向旋转，其中：

$$v_i = D_i \times D_{i+1} \quad (4)$$

表示由方向 D_i 转到方向 D_{i+1} 的旋转轴。

$$w_i = D_i \cdot D_{i+1} \quad (5)$$

表示由方向 D_i 转到方向 D_{i+1} 的旋转角。

设 θ 是两四元组 Q_i 与 Q_j 间的转角中值， α 为插值 Q_m 与四元组 Q_j 的相似程度，则 Q_m 可由下式求得：

$$Q_m = \text{slerp}(Q_i, Q_j, t) = \frac{Q_i \sin((1-t)\theta) + Q_j \sin(t\theta)}{\sin(\theta)} \quad (6)$$

设首帧视角旋转为稳定旋转，平滑算法为：

-
- 输入： Q_1, \dots, Q_n (原始旋转)
 输出： $\hat{Q}_1, \dots, \hat{Q}_n$ (平滑后旋转)
1. For $k=1$ to α
 2. For $i=2$ to n
 3. $\hat{Q}_i = \text{slerp}(Q_i, \hat{Q}_{i-1}, t)$
 4. End
 5. End
-

其中 α 为迭代次数。

5. 视频帧的筛选与修正

本文算法避免对视频帧做较大变换，以防止视频内容发生较大扭曲，因此选取距离虚拟路径最近且视角与虚拟视角最接近的视频帧做视角变换与修正。

5.1. 视频帧筛选算法

基于视角优化的视频帧筛选策略如下：

- 1) 首先选取摄像机位置距离虚拟路径较近的视频帧为待选变换帧。做各个时刻摄像机位置与虚拟路径的投影距离。路径上每个位置点可能有 0 个至几个摄像机与其距离是最近的，有的虚拟路径上也可能没有邻近的摄像机。
- 2) 如果路径上某位置没有邻近摄像机，则添加此位置的前后路径位置的摄像机位置为待选。
- 3) 根据待选摄像机，选取距离虚拟路径最近且视角与优化估计视角偏差 ψ 最小的视频帧为最优变换帧。

$$N_i = \arg \min_j \left\{ \lambda_1 (P_j - C_i) + \lambda_2 \cos^{-1} (D_j \cdot O_i) \right\} \quad (7)$$

其中 N_i 表示虚拟路径位置 i 处的最优帧号。 O_i 为虚拟路径位置 i 处的摄像机优化朝向。 C_i 表示虚拟路径位置 i 的坐标。 P_j 与 D_j 表示第 j 时刻摄像机的实际位置坐标和朝向。 λ_1 与 λ_2 为权重常数， $\lambda_1 \geq 0$ 、 $\lambda_2 \geq 0$ ，分别表示位置偏差与朝向偏差在目标函数中所占的比重。

实际中可选 $\lambda_1 = \frac{1}{2 \max_j (P_j - C_i)}$ ， $\lambda_2 = \frac{1}{2\pi}$ ，在分别对位置偏差与朝向归一化的基础上使两者权重相等。

5.2. 纸型

视角变换是通过适当的矩阵映射使一视角的图像矩阵变换为从另一视角看去的图像矩阵。通过变换视角，过滤掉连续的视频帧在视角上较大的晃动，使其在视觉上表现平滑。首先，通过内参投影矩阵使相平面坐标变换为对应摄像机坐标系下的坐标向量，然后通过旋转矩阵使此坐标向量变换为虚拟视角的摄像机坐标系下的坐标向量，最后再通过内参矩阵使坐标向量投影到虚拟像平面。

设摄像机在实际位置 i 具有旋转矩阵 R_i ，虚拟路径点 j 的摄像机旋转矩阵为 R'_j 。当摄像机内参矩阵为 K 时，虚拟像平面坐标点 x' 与实际相平面坐标点 x 之间的关系为：

$$x' = KR'_j R_i^T K^{-1} x \quad (8)$$

6. 实验分析

实验表明本文算法在无精确内参的条件下，通过 3D 摄像机运动估计可以达到视频去抖的目的。

实验配置为：intel(R)Xeon(R) E5620 2.40 Hz，内存 12.0 GB，操作系统 Windows Server 2008 R2 Standard。

本文数据集采用了 Feng Liu 提供的抖动视频数据集 0018AF[7]和 shakycar[7]。

视角平滑迭代参数 α 试验中取 3~8 就已经能产生很好的效果。

实验结果如图 4 所示。其中，射线表示相机视角朝向。图 4(a)是视角平滑前数据集 0018AF 的摄像机真实朝向。图 4(b)显示了对视角做平滑估计后的结果。本文算法在不改变相机真实运动轨迹的基础上，对相机视角进行平滑处理。图 4(c)是 Shuaicheng Liu 的 2d 去抖算法对视角的处理情况。Shuaicheng Liu 的算法在优化视角的同时改变了相机原始轨迹，没有真实反映出拍摄者的浏览路径。图 4(d)~(f)分别表示了 shakycar 数据集视角平滑前后的视角变化和 Shuaicheng Liu 算法的视角变化。可以看出，本文算法在不改变相机原始轨迹的基础上，能更好的平滑视角朝向。

本文算法和 Shuaicheng Liu 的算法对视角平滑的对比结果如表 1 所示。其中平均角速度表示了相机视角变化的幅度，平均角速度越大说明视频抖动幅度越大。平均角加速度衡量视角变化的突然程度，平均角加速度越大所产生的视频抖动越剧烈。

表 1 的实验结果显示，本文的视角变换策略是有效的，并且对比 Shuaicheng Liu 算法有更好的视角平滑效果。

本文算法与 Shuaicheng Liu 的 2D 去抖算法的效果对比如图 5 所示。图 5(a)~(c)分别是 0018AF 数据集原始数据、Shuaicheng Liu 的 2D 去抖算法和本文算法对 0018AF 数据集第 12 至 16 帧数据的处理

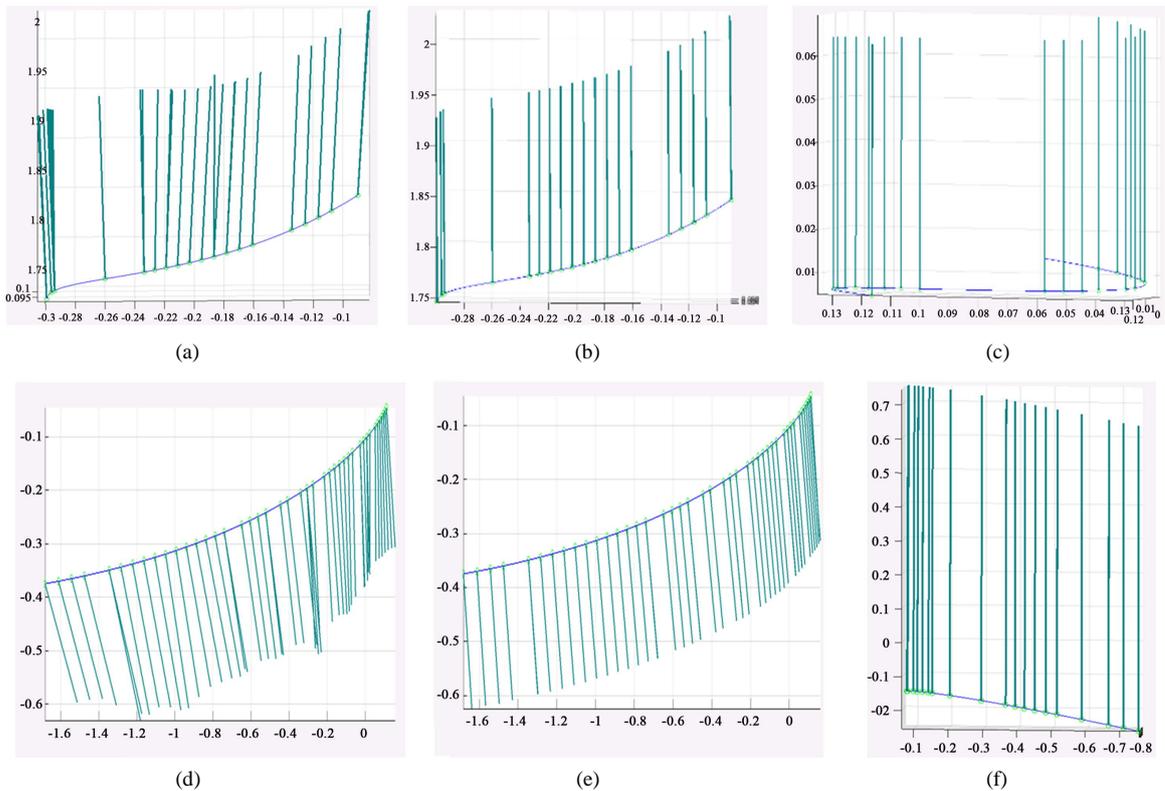


Figure 4. The comparison of before and after the smoothing camera rotation. (a) Original data; (b) The proposed method; (c) Shuaicheng Liu's method; (d) Original data; (e) The proposed method; (f) Shuaicheng Liu's method

图 4. 视角平滑前后的对比。(a) 0018AF 平滑前；(b) 0018AF 平滑后；(c) Shuaicheng Liu 的方法；(d) 0018AF 平滑前；(e) 0018AF 平滑后；(f) Shuaicheng Liu 的方法

Table 1. Smoothing algorithm data contrast

表 1. 平滑算法数据对比

0018AF	原始数据	本文算法	Shuaicheng Liu 算法
平均角速度	0.3633	0.0106	0.0116
平均角加速度	0.0938	0.0012	0.0059

Shakycar	原始数据	本文算法	Shuaicheng Liu 算法
平均角速度	0.1478	0.0174	0.0174
平均角加速度	0.3496	0.0037	0.0087

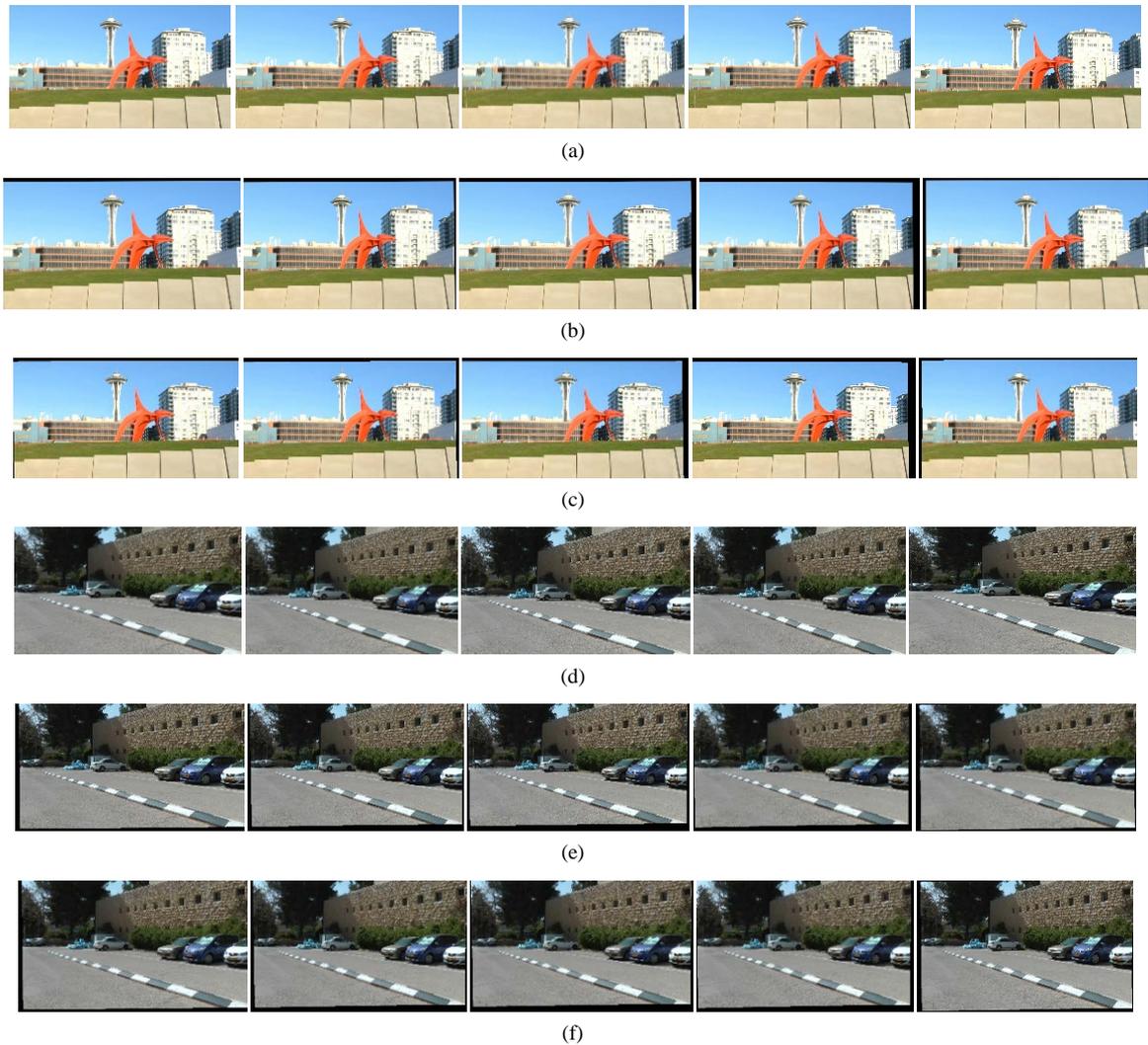


Figure 5. The comparison between the result of ours and Shuaicheng Liu's. (a) The original frames of 0018AF from 12 to 16; (b) The result of the proposed method; (c) The result of Shuaicheng Liu's method; (d) The original frames of shakycar from 22 to 26; (e) The result of the proposed method; (f) The result of Shuaicheng Liu's method

图 5. 本文算法与 Shuaicheng Liu 算法效果对比。(a) 0018AF 第 12 至 16 帧原始视频帧; (b) 本文算法效果; (c) Shuaicheng Liu 算法效果; (d) shakycar 第 22 至 26 帧原始视频帧; (e) 本文算法效果; (f) Shuaicheng Liu 算法效果

结果。图 5(d)~(f) 分别是 shakycar 原始数据、Shuaicheng Liu 的 2D 去抖算法和本文算法对数据集第 22 至 26 帧数据的处理结果。图 5(b) 和图 5(e) 显示经本文算法处理的视频帧去除了视频抖动，且不存在图像局部区域的抖动问题。在图 5(c) 和图 5(f) 中，Shuaicheng Liu 的去抖算法由于图像特征点分布不均匀，去抖处理产生了图像局部区域的抖动。本文算法，基于更合理的相机运动模型，避免了帧局部的失真。

7. 结论

本文提出了一种基于对视角优化进行运动补偿的 3D 视频去抖算法，通过估计摄像机 3D 运动轨迹选帧并做视角优化进行视频去抖。我们采用视角优化的方法更合理地修正了摄像机的运动参数，稳定视频内容。实验证明本文算法在缺乏精确摄像机内参数的条件下，可以通过摄像机 3D 运动估计，很好的解决便携设备中普遍存在的视频抖动问题。

参考文献 (References)

- [1] Hartley, R. and Zisserman, A. (2003) *Multiple View Geometry in Computer Vision*. Cambridge University Press, Cambridge.
- [2] 宋利, 周源华, 周军. 基于运动矢量的视频去抖动算法[J]. 上海交通大学学报, 2004(z1): 63-66.
- [3] Song, L., Zhou, Y.H. and Zhou, J. (2005) Robust Video Stabilization Based on Motion Vectors. *Journal of Shanghai University (English Edition)*, **9**, 46-51. <http://dx.doi.org/10.1007/s11741-005-0103-1>
- [4] 徐理东, 林行刚. 视频抖动矫正中全局运动参数的估计[J]. 清华大学学报: 自然科学版, 2007, 47(1): 92-95.
- [5] 赵文华, 姚天翔, 叶秀清, 等. RANSAC 算法在视频去抖动中的应用[J]. 电路与系统学报, 2005, 10(4): 91-94.
- [6] 於俊, 汪增福. 基于 SIFT 特征匹配的实时鲁棒视频去抖系统[J]. 系统工程与电子技术, 2014, 36(2).
- [7] Liu, S., Yuan, L., Tan, P. and Sun, J. (2013) Bundled Camera Paths for Video Stabilization. *ACM Transactions on Graphics (TOG)*, **32**, 78. <http://dx.doi.org/10.1145/2461912.2461995>
- [8] Liu, F., Gleicher, M., Jin, H. and Agarwala, A. (2009) Content-Preserving Warps for 3D Video Stabilization. *ACM Transactions on Graphics (TOG)*, **28**, 44.
- [9] Kopf, J., Cohen, M.F. and Szeliski, R. (2014) First-Person Hyper-Lapse Videos. *ACM Transactions on Graphics (TOG)*, **33**, 78. <http://dx.doi.org/10.1145/2601097.2601195>
- [10] Ryu, Y.G., Roh, H.C. and Chung, M.J. (2010) 3D Video Stabilization for Humanoid Eyes Using Vision and Inertial Sensors Inspired by Human VOR. *IEEE International Conference on Robotics and Biomimetics (ROBIO)*, 1780-1785.
- [11] Buehler, C., Bosse, M. and McMillan, L. (2001) Non-Metric Image-Based Rendering for Video Stabilization. *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, **2**, II-609. <http://dx.doi.org/10.1109/cvpr.2001.991019>
- [12] Snavely, N., Seitz, S.M. and Szeliski, R. (2006) Photo Tourism: Exploring Photo Collections in 3D. *ACM Transactions on Graphics (TOG)*, **25**, 835-846.
- [13] Shoemake, K. (1985) Animating Rotation with Quaternion Curves. *ACM SIGGRAPH Computer Graphics*, **19**, 245-254. <http://dx.doi.org/10.1145/325165.325242>