# A Video Stabilization Algorithm for Train-Mounted Intrusion Detection System Based on ALP Keypoints

Ling Guan, Xiaofeng Li, Han Yang
School of Traffic and Transportation, Beijing Jiaotong University
Beijing, China
e-mail: 15120755@bjtu.edu.cn, xfengli@bjtu.edu.cn, 14120801@bjtu.edu.cn

Limin Jia
State Key Laboratory of Rail Traffic Control and Safety, Beijing Jiaotong University
Being, China
e-mail: lmjia@bjtu.edu.cn

***Abstract*—With the rapid increasing of railway mileage, automatic railway intrusion detection by analyzing the video from train-mounted camera is becoming very meaningful. But there exits serious jitter in the video since the camera always vibrates with train when it is running, a video stabilization reprocessing procedure is the prerequisite before the intrusion detection analysis. To solve this question, a robust video stabilization algorithm based on low degree polynomial detector is presented in this paper. Generally, this algorithm consists of three stages. Firstly, keypoints are extracted and tracked by ALP detector and KLT tracker respectively between adjoining frame of the video. Secondly, the local motion vector is obtained by image affine transformation model. Finally, the global motion vectors are corrected and smoothed by using RANSAC and Kalman filter respectively. At the end of this paper, experiments based on PSNR and ITF values are given to show that the proposed algorithm is more effective than Harris corner detector which is often used in the traditional stabilization algorithm.**

***Keywords-video stabilization; ALP detector; KLT tracker; motion estimation; Kalman filter***

## I. INTRODUCTION

With the rapid increasing of railway mileage, automatic railway intrusion detection by analyzing the video from train-mounted camera is becoming more and more meaningful since it can avoid serious traffic accident by detecting obstacle in time. But there exits serious jitter in the video since the camera always vibrates with train when it is running. So a video stabilization reprocessing procedure is the prerequisite before the intrusion detection analysis. Hence, video stabilization techniques have gained attention, as this method can obtain high quality and stable video sequences.

Motion estimation, the process of estimating motion offset of image sequences, is the prerequisite of video stabilization. The most often used motion estimation algorithms are block matching and feature matching. Feature based matching algorithm have gained larger consensus for their better performance. Features matching stabilization algorithm like SIFT [1], SURF [2], [3], specific optical flow [4], FAST and BRIEF [5], MSER [6], Harris [7] and points detected by machine learning [8] have been discussed. Each method has its own advantage and disadvantage.

A Low-degree Polynomial (ALP) keypoints detector was proposed by Telecom Italia in 2013 and was adopted by CDVS (Compact Descriptors for Visual Search) ISO/IEC MPEG-7 Part 13 MPEG international standard in 2014 [9,10]. The points obtained by ALP are invariant to image scale, rotation and robust to changing viewpoints and illumination. Comparing to the SIFT descriptor, ALP performs better on the keypoints location and running speed [11]. Meanwhile, the Harris corner detector has higher speed but it is not scale invariant and affine invariant [12] so the ALP keypoints are more satisfied for video stabilization.

For the better performance of ALP keypoints, this paper presents an effective video stabilization algorithm based on ALP detector. The rest of this paper is organized as follows. In section II, the flowchart of the proposed algorithm is described. In section III, the main stages of ALP detector are presented. In section IV, motion estimation and filtering methods are introduced. Finally, the experiment results and conclusion are given in section V and section VI respectively.

## II. PROPOSED FRAMEWORK

The main flow of proposed stabilization algorithm in this paper is shown in figure 1.

The algorithm first extracts ALP feature from the first frame and the KLT tracker is applied to track ALP keypoints for next frame. Then the feature points are selected using RANSAC by removing the outliners and the global motion vector will also be obtained. Next, using Kalman filter to smooth the motion. Finally, the next frame is compensated by the motion vector.

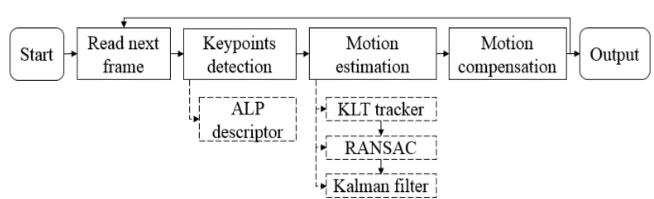



Figure 1. Flowchart of the proposed algorithm.

## III. ALP KEYPOINTS EXTRACTION

The first step in video stabilization algorithm is to identify keypoints from subsequent frames.

In order to find keypoints, ALP approximates the result of the LOG filtering by means of polynomials, used to find extrema in the scale space. The ALP keypoint detection procedures consist of two main stages:

*1)* Scale-space extrema detection

The scale space shall be structured in some octaves with 4 increasing positive scale parameter $\sigma$. Each image in each octave shall be produced by scale-normalized Laplacian filtering of the Gaussian-filtered images, followed by SIFT descriptor. Then the image's pyramid with octaves is generated.

For each pixel $(x,y)$ in the image, a polynomial approximation to the scale-space function shall be searched for a local extremum. The function is shown in equation 1:

$$p(x,y,\sigma)=\alpha_3(x,y)\sigma^3+\alpha_2(x,y)\sigma^2+\alpha_1(x,y)\sigma+\alpha_0(x,y) \quad (1)$$

where $\alpha_0(x,y)\sim\alpha_3(x,y)$ are obtained by computing weighted sums of the octave images.

Candidates $\{x,y,\sigma^*(x,y)\}$ shall be obtained by means of $p(x,y,\sigma)$. To detect the scale-space extrema, ALP first locates the local extrema in the $\sigma$ direction by setting its first derivative to zero, and then it compares the point to its 8 neighbors in the image.

*2)* Local feature selection

Candidates in the polynomial shall be refine to subpixel and some duplicates will be eliminated. Then candidates will be assigned a dominant orientation for rotation invariance. The remaining candidates $\{x,y,\sigma^*(x,y)\}$ are input to the next processing step.

For smaller feature data sizes (512 bytes to 4KB) to accelerate the computation, selecting a subset of keypoints is critical. A relevance measure is computed for each keypoint. The relevance measure has been statistically learned based on six characteristics of interest points: the scale in scale-space $\sigma$; the orientation $\theta$; the LoG response value $D$; the distance $d$ to the image center; the ratio $\rho$ of the squared trace of the Hessian and the second derivative $p_{\delta\delta}$ of the scale space function with regard to $\sigma$. The relevance score $r$ for a point is obtained by multiplying the conditional probabilities of each characteristic, as shown in equation 2:

$$\begin{aligned} &r(\sigma,\theta,d,D,\rho,p_{\delta\delta}) \\ &\quad=f_1(\sigma)\cdot f_2(\theta)\cdot f_3(d)\cdot f_4(D)\cdot f_5(\rho)\cdot f_6(p_{\delta\delta}) \end{aligned} \quad (2)$$

where the factors $f_1 \sim f_5$ are taken from the normative tables of the learned conditional distributions according to the interest point characteristics.

In this paper, the keypoints is ranked and selected based on the relevance measure $r$.

## IV. Motion Estimation and Filtering

The second stages in video stabilization is motion estimation and image compensation. Traditional keypoints matching method such as Hamming distance and nearest neighbour algorithm cost much time. To overcome this shortage, this paper adopts KLT tracker to track keypoints between frames. Then the affine transformation matrix is obtained according to the model of camera motion and the stabilized frame is outputted by motion compensation.

### *A. KLT Tracker*

KLT tracker, also called LK tracker, is an efficient point-tracking algorithm. KLT is a kind of implementation for optical flow and can match keypoints in image sequences with high speed and accuracy [13].

Let $I(x,y)$ and $J(x,y)$ be two 2D grayscale images. $I(x,y)$ is referenced as the first image, and $J(x,y)$ as the second image. The motion changing of pixel can be described by a translation model:

$$I(x,y,t)=J(x+dx,y+dy,t+dt) \quad (3)$$

where $t$ is time, and $d=(dx,dy)^T$ is the image velocity or the optical flow.

Matching problem can be transformed into calculating minimum problem. Assuming a feature window $W$ containing the texture information. Let $u=\left[u_x,u_y\right]^T$ be a keypoint on the first image. The residual function $\varepsilon$ is defined as follows:

$$\varepsilon(d)=\varepsilon(dx,dy)=\sum_{x=u_x-w_x}^{u_x+w_x}\sum_{y=u_y-w_y}^{u_y+w_y}(I(x,y)-J(x+d_x,y+d_y))^2 \quad (4)$$

where $w_x$ and $w_y$ are integration window $W$ size parameter.

In order to find the optimal $d$, the first derivative of $\varepsilon(d)$ shall be zero. Thus, the solution of this problem is below:

$$Zd=e \quad (5)$$

where: $Z=\iint_W g^T g(x)w(x)dx$, $e=\iint_W\left[I(x)-J(x)\right]g(x)w(x)dx$ and $g=\left[\frac{\partial}{\partial x}\left(\frac{I+J}{2}\right),\frac{\partial}{\partial y}\left(\frac{I+J}{2}\right)\right]^T$.

Standard KLT algorithm can only deal with small pixel displacement. This paper uses pyramid KLT algorithm to achieve pixel tracking for large displacement. The pyramid representation is built in a recursive fashion. The optical flow $d$ first computed at the deepest pyramid level. Then, the result of that computation is propagated to the upper level in a form of an initial guess for the pixel displacement and so on up to the top level (the original image).

### *B. Motion Vector Estimation*

The result obtained by pyramid KLT tracker is a list of successful matching keypoints pairs which can be easily used in the motion estimation algorithm. Global motion can be described by the 2D affine transformation:

$$\begin{bmatrix} x_f \\ y_f \end{bmatrix} = \lambda \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} + \begin{bmatrix} d_x \\ d_y \end{bmatrix} \tag{6}$$

where $\lambda$ is the zoom parameter, $\theta$ the rotation angle, $d_x$ and $d_y$ respectively X-axis and Y-axis shifts.

Each keypoints pair can get a local motion vector by equation 6. But the whole set of matching pairs may contain some wrong pairs, which can affect the result of the global motion vector estimation. This paper adopts RANSAC algorithm to discard outliers that do not match the estimated homography and the global motion vector will also be obtained. Finally, the Kalman filter is adopted to smooth the motion and the stabilized image can be obtained by motion compensation.

## V. EXPERIMENTAL RESULTS

In order to evaluate the performance of proposed algorithm, PSNR (peak signal to noise ratio), as shown in equation 7, is adopted to test the interframe stability after video stabilization. A greater PSNR indicates the smaller interframe deviation and better stabilization effect.

$$PSNR(I_1, I_0) = 10\log\frac{255^2}{MSE(I_1, I_0)} \tag{7}$$

where $MSE(I_1, I_0)$ denotes mean-square error between frames. The definition of $MSE(I_1, I_0)$ is shown in equation 8:

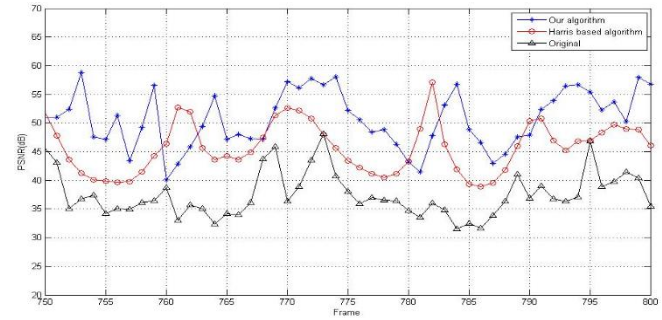
$$MSE(I_1, I_0) = \frac{1}{NM} \times \sum_{n=1}^{N}\sum_{m=1}^{M}(I_1(x_n, y_m) - I_0(x_n, y_m))^2 \tag{8}$$

In the experiments, three railway image sequences and common used Harris corner detector based stabilizaiton algorithm are used as the evaluation tools. The comparing results of PSNR value are shown in figure 2.

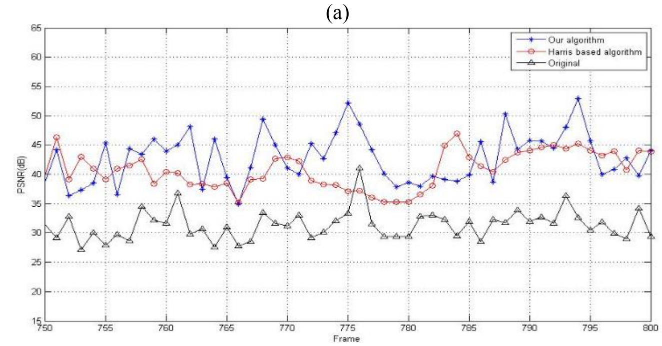
Figure 2 shows that proposed algorithm increases the PSNR values for most of frames in the different sequences.

Meanwhile, Interframe Transformation Fidelity (ITF) value is then used to objectively assess the stabilization brought by an algorithm. The definition of ITF is shown in equation 9. The higher ITF value, the more stable video sequence. The results of ITF values for different sequence are shown in table 1.

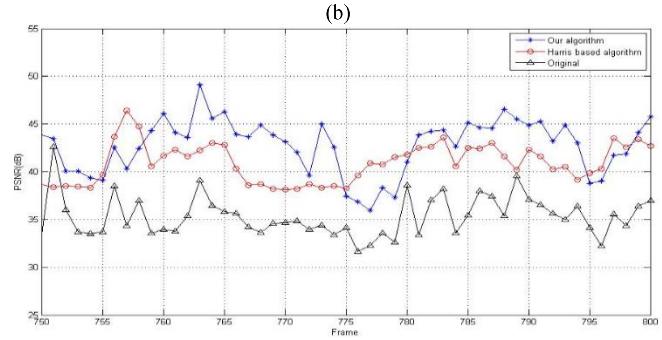
$$ITF = \frac{1}{N_{frame} - 1}\sum_{k=1}^{N_{frame}-1} PSNR(k) \tag{9}$$




(a)




(b)




(c)

Figure 2. PSNR values comparsion, only show the frames from 750~800 for each sequence. (a) sequence 1; (b) sequence 2; (c) sequence 3.

TABLE I. COMPRASION OF ITF FOR DIFFERENT SEQUENCE

| Sequence | Original ITF(dB) | Harris based algorithm ITF(dB) | Our algorithm ITF(dB) |
|---|---|---|---|
| Seq1 | 37.65 | 46.10 | 47.64 |
| Seq2 | 31.04 | 40.80 | 43.07 |
| Seq3 | 38.26 | 50.02 | 52.15 |

Table I shows that our algorihm has a good improvement in the ITF value about 3%~5% comparing to the Harris based stabilization program. The results confirm the effectiveness of the proposed algorithm.

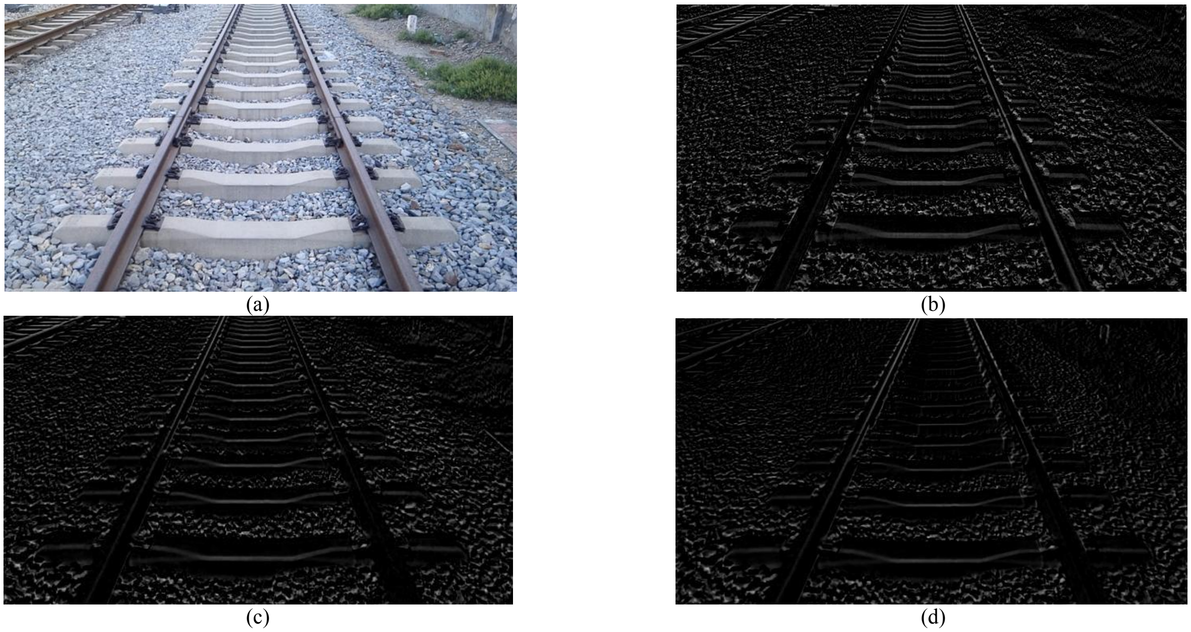



Figure 3. Difference images example between frame 58 and 57 in Sequence 3. (a) the orginial image; (b) the original difference image; (c) stabilize by Harris corner detector based algorithm; (d) stabilize by our algorithm.

Difference image is very fit for visually comparing of consecutive images. This paper also adapts calculating difference image to evaluate the proposed algorithm. The experimental results denote the proposed algorithm has better performance, as shown in figure 3.

## VI. CONCLUSION

Reliable video stabilization is a challenging issue for its application in the train-mounted condition. This paper proposes a new video stabilization algorithm. In this algorithm, ALP detector based feature extraction is used for video frames and the KLT tracker is adopted to track ALP keypoints. Global motion vector estimation is obtained by means of RANSAC and Kalman filter. The experimental results show that the proposed algorithm is more robust and effective than classical Harris corner detector based video stabilization method. The future research is to further optimize the algorithm, and transplant the algorithm into the embedded platform.

## ACKNOWLEDGMENT

This work is supported by the National Key Research and Development Program of China (No. 2016YFB1200402).

## REFERENCES

[1] S.Hong, T.Hong, W.Yang. Multi-resolution unmanned aerial vehicle video stabilization. Proceedings of the IEEE 2010 National Aerospace & Electronics Conference:126-131, 2010.

[2] Tahiyah Nou Shene, K.Sridharan, N.Sudha. Real-Time SURF-Based Video Stabilization System for an FPGA-Driven Mobile Robot. IEEE Transactions on Industrial Electronics 63(8):5012-5021, 2016.

[3] G.Chunxian, Z.Zhe, L.Hui. Hybrid video stabilization for mobile vehicle detection on SURF in aerial surveillance. Discrete Dynamics in Nature and Society, 2015.

[4] S.Liu, L.Yuan..Tan et al. SteadyFlow: Spatially smooth optical flow for video stabilization. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition:4209-4216, 2014.

[5] S.Jeon, I.Yoon, J.Jang et al. Robust Video Stabilization Using Particle Keypoint Update and l1-Optimized Camera Path. Sensors 17(2):337-355, 2017.

[6] M. Okade, and P. K. Biswas. Video stabilization using maximally sta- ble extremal region features. Multimedia Tools Appl 68(3):947-968, 2014.

[7] Li W., Hu J., Li Z., Tang L., Li C. Image Stabilization Based on Harris Corners and Optical Flow. Engineering and Management. KSEM 2011. Lecture Notes in Computer Science, vol 7091. Springer, Berlin, Heidelberg, 2011.

[8] J.Dong, H.Liu. Video Stabilization for Strict Real-Time Applications. IEEE Transactions on Circuits and Systems for Video Technology 27(4):716-724, 2017.

[9] "Compact Descriptors for Visual Search: Evaluation Framework," ISO/IEC JTC1 SC29 WG11 output document N12202, 2011.

[10] S.Paschalakis et al. "Information Technology - Multimedia content descriptors interface - Part 13: Compact Descriptors for Visual Search"in ISO/IEC 15938-13, 2015.

[11] K Cordes. Localization Accuracy of Interest Point Detectors with Different Scale Space Representations. AVSS:247-252, 2013.

[12] Ehab Salahat, Murad Qasaimeh. Recent Advances in Features Extraction and Description Algorithms: A Comprehensive Survey. Annual IEEE Industrial Electronics Society's 18th International Conf. on Industrial Technology (ICIT): 22-25, 2017.

[13] S.Baker, I.Matthews. Lucas-Kanade 20 years on: A unifying framework. International Journal of Computer Vision 56(3):221-255, 2004.