A Novel AOSLO Video Image Stabilization Algorithm

Guangqiu Chen, Yu Chen

School of Electronic and Information Engineering, Changchun University of Science and Technology, Changchun, 130022, China

Abstract

Confocal scanning laser ophthalmoscope based on adaptive optics can correct the aberrations, which can achieve high resolution imaging in vivo retinal. But because of the micro tremor of the patient and the jitter of the device, the video image of the adaptive optics scanning laser ophthalmoscopy (AOSLO) output is blurred. To solve this problem, an AOSLO image stabilization algorithm based on Harris-SIFT is proposed. Firstly, Harris-SIFT algorithm is used to match feature vector, then the RANSAC algorithm is used to verify the match accuracy, the motion vector value estimated is obtained. Finally, the fixed frame compensation method is used to realize the motion compensation. Experimental results show that the proposed algorithm can effectively remove jitter and enhance the contrast of video image.

Keywords

AOSLO, RANSAC, SIFT.

1. Introduction

AOSLO is that adaptive optics is used to correct the aberrations of the eye and high resolution retinal cell video image is realized by combined with confocal scanning laser ophthalmoscope (CSLO) technology. During the doctor checking patient's condition by AOSLO, the micro tremor of the patient and the jitter of the device make the observed video image blur. So it is necessary to stabilize video image by stabilization algorithm, so that better video image can be obtained to assist the doctor to make an accurate judgment. In this paper, AOSLO video image stabilization algorithm based on Harris-SIFT is proposed. For improving the matching speed, the Harris-SIFT algorithm is used to match the feature vector of the video image. For improving the matching accuracy and stability, the RANSAC algorithm is used to verify the matching accuracy to obtain the motion vector estimation. The motion compensation is realized by the fixed frame compensation method to output stable video image.

2. Scale-invariant feature transform (SIFT)

Scale-invariant feature transform (SIFT) algorithm is a new image local feature description algorithm based on scale space, image scaling, rotation and affine transform on the basis of the summary on the existing feature detection method by the invariant technology, which can be divided into 4 steps[1][2].

2.1 DETECTION OF EXTREME POINTS IN THE SCALE SPACE

The scale space theory first appeared in the field of computer vision, and whose purpose is to simulate the multi-scale characteristics of image data. In reference [3], It is proved that the Gauss convolution kernel is the only transform kernel in realizing scale transform, and the only linear kernel. Given a 2D image, the scale space representation at different scales can be obtained from the image and Gauss kernel convolution

$$L(x,y,\sigma)=G(x,y,\sigma)*I(x,y) \tag{1}$$

Where (x,y) represents spatial coordinate, σ represents scale coordinate, $G(x,y,\sigma)$ represents Gaussian function with scale variable

$$G(x,y,\sigma)=\frac{1}{2\pi\sigma^2}e^{-(x^2+y^2)/2\sigma^2} \tag{2}$$

For effectively detect the stable feature points in the scale space, the difference of Gaussian (DOG) scale space is proposed, which is the approximation of the LOG operator with scale normalization. DOG operator is simple, which can be generated by Gauss difference kernel in different scales and image convolution.

$$\begin{aligned} D(x,y,\sigma)&=\left(G(x,y,k\sigma)-G(x,y,\sigma)\right)*I(x,y)\\ &=L(x,y,k\sigma)-L(x,y,\sigma)\end{aligned} \tag{3}$$

For finding the extreme points in the scale space, each sampling point should be compared with its adjacent points to see whether it is larger or smaller than the adjacent points in the image domain and the scale domain.

2.2 Precise location of the extreme points

The position and scale of the feature points can be accurately located and the characteristic points with the low contrast and the unstable edge response points (the DOG operator can generate strong edge response) can be removed by fitting 3D quadratic function, which can enhance the matching stability and improve the ability of anti noise. The accurate location of the feature points position and scale coordinates (sub-pixel accuracy) can be obtained by two order Taylor expansion interpolation of DOG function $D(X)$.

$$D(X)=D+\frac{\partial D^T}{\partial X}X+\frac{1}{2}X^T\frac{\partial^2 D}{\partial X^2}X \tag{4}$$

Where $X=(x,y,\sigma)$ represents the position and scale offset between the sampling point and the feature point. The first derivative of the formula (4) is set to 0, the offset vector for the accurate position of the feature points can be obtained

$$\hat{X}=-\frac{\partial^2 D^{-1}}{\partial X^2}\frac{\partial D}{\partial X} \tag{5}$$

$\hat{X}$ is added to the original rough feature points coordinate X and the accurate sub pixel interpolation estimates of the feature points can be obtained. The substitution of equation (5) in equation (4) can be obtained:

$$D(\hat{X})=D+\frac{1}{2}\frac{\partial D}{\partial X}\hat{X} \tag{6}$$

When $|D(X)|$ is less than a certain threshold, the feature points can be rounded, usually this feature is sensitive to noise, therefore unstable. Furthermore, the edge response points with unstable should be eliminated. The tangential principal curvature at the edge is larger and that at vertical direction vertical is smaller in this kind of extreme point in DOG function. For detecting whether the principal curvature is less than a certain threshold, it is only necessary to detect whether the following inequality is satisfied

$$\frac{Tr(H)^2}{Det(H)} < \frac{(r+1)^2}{r} \tag{7}$$

Where *H* represents Hessian matrix of DOG function, $H = \begin{pmatrix} D_{xx} & D_{xy} \\ D_{xy} & D_{yy} \end{pmatrix}$, In the course of the experiment, $r = 10$

2.3 Select Main Direction For Each Feature Point

By the gradient direction distribution characteristic of the neighborhood pixels of the feature points, the parameters of each feature point are specified, which makes the operator rotation invariance. Firstly, Gradient mode and orientation of feature points are computed in Gauss space by the following equation

$$\begin{cases} m(x,y) = \sqrt{\left(L(x+1,y) - L(x-1,y)\right)^2 + \left(L(x,y+1) - L(x,y-1)\right)^2} \\ \theta(x,y) = tan^{-1}\left(L(x,y+1) - L(x,y-1)\right) / \left(L(x+1,y) - L(x-1,y)\right) \end{cases} \tag{8}$$

Then a gradient direction histogram is created by the sampling in the neighborhood of the feature points in Gauss space. Histogram of each 10 degree as a column, a total of 36 columns, and then each sampling point in the neighborhood is placed into the appropriate column according to the gradient direction θ, gradient modulus *m* is used as contribution weight. Finally, the main peak value is regarded as the main direction of the feature point, the local peak of more than 80% main peak value is selected as auxiliary direction. So a feature point may be specified with multiple directions, which can enhance matching robust.

2.4 Generating feature point description operator

The coordinate axis is rotated into the direction of the feature points to ensure rotation invariance. The feature points are taken as the center to select a window ,the size is 8×8, then the gradient direction histogram in 8 directions is calculated on each 4×4 block. The cumulative value in each gradient direction can be computed to form a seed. The neighborhood directional information association algorithm can enhance the ability of anti noise, and also provide a better fault tolerance for the feature matching with localization error. In the actual calculation process, for enhancing the matching robustness, each feature point is described by 4×4, a total of 16 seeds. So a feature point can produce 128 data, which can generate 128 dimensional SIFT feature vector. Now the SIFT feature vector has removed the influence of scale change, rotation and other geometric distortion factors, furthermore, the length of the feature vector is normalized, the influence of the illumination change can be removed.

3. HARRIS-SIFT

When SIFT feature extraction and matching algorithms are applied in the AOSLO video image stabilization system with high real-time requirement, it exits three main problems[4][5]

(1) The complexity of Feature extraction is too high and the computation time is too long.

(2) The feature points generated are too much, which can affect the matching and searching speed.

(3) The overall significance of the feature set is poor.

For solving the real-time problem of SIFT algorithm, the other more effective feature point detection operator should be considered to replace the extreme point extraction algorithm in SIFT. Harris operator is an effective point feature extraction operator, as following

$$R = det(C) - ktr^2(C) \tag{9}$$

Where *det* represents the determinant of a matrix, *tr* represents matrix trace, *C* represents correlation matrix

$$C(x) = \begin{bmatrix} I_u^2(x) & I_{uv}(x) \\ I_{uv}(x) & I_v^2(x) \end{bmatrix} \tag{10}$$

Where $I_u(x), I_v(x)$ and $I_{uv}(x)$ represent partial derivative and two order mixed partial derivative, k is a empirical value, usually take 0.04 -0.06.

When the Harris operator *R* of a certain point is larger than the threshold value *T*, the point is the corner point..The merit s of the Harris operator are that: (1) simple calculation (2) The feature points are uniform and reasonable, which can reflect the structure of the image (3) Feature corner points can be extracted quantitatively (4) Even in the presence of image rotation, gray level changes, noise effects and viewpoint transformation, it is also the most stable point feature extraction algorithm. So the extreme points in SIFT algorithm is replaced by the Harris feature points, then defines the main direction of each feature point and produces feature vector description for each feature points, which is called Harris-SIFT algorithm. When the feature vectors of the two image feature points are generated, the Euclidean distance of the feature points is used as the similarity measurement of the feature points.

4. Ransac matching verification and motion compensation

In AOSLO video image, the target is are similar, which is easy to enhance the similarity of the feature point, which can result in false matching. These false matching points make the processing speed of the algorithm is affected, and the accuracy of the results is reduced.For reducing the error and improve the stability, the Random Sample Consensus (RANSAC) algorithm is used to verify the matching accuracy. RANSAC algorithm steps are as follows

(1) The samples were randomly selected from a sample of RANSAC, namely 4 matching points.

(2) According to these 4 matching points to calculate the transformation matrix M

(3) According to the sample set, the transform matrix M, and the error metric function, the consensus of the current transform matrix is calculated and returns the number of the consensus elements

(4) According to the elements number of the current consensus to determine whether the optimal consensus, if it is true, the current consensus is updated.

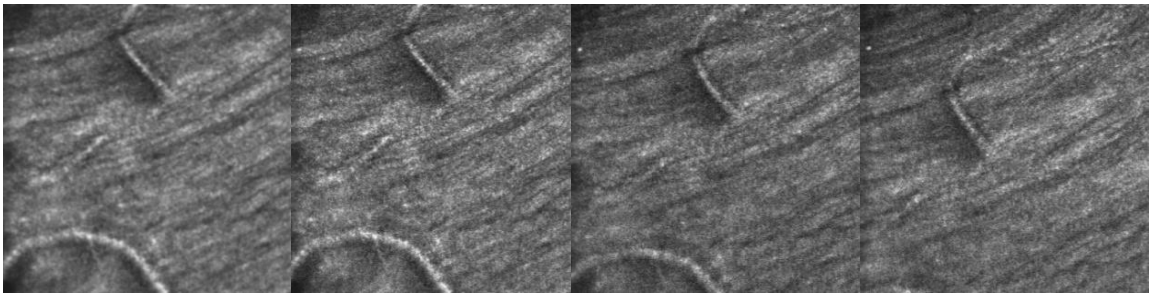
(5) Update the current error probability p, if p is greater than the allowable minimum error probability , repeat (1) to (4) , continue, until the current error probability p is less than the minimum error probability.

The above steps is repeated, image stabilization is outputted.

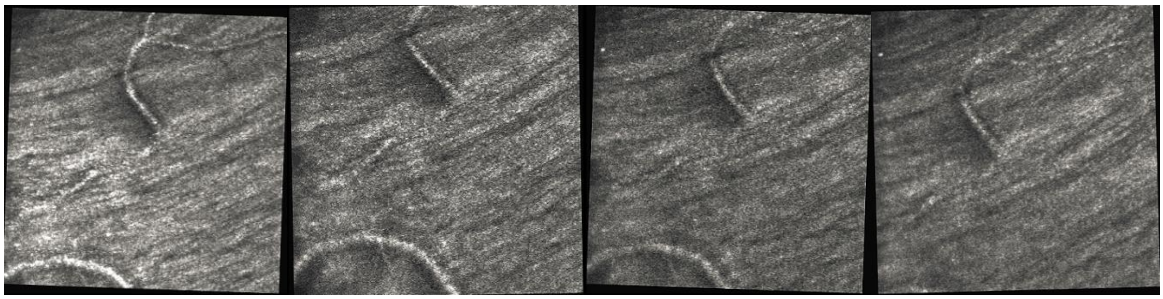
As can be seen from the imaging characteristics of AOSLO, objects mainly move in the form of translational motion, very little rotation and scaling, so the motion angle can be negligible. So the fixed frame compensation method is used to compensate the motion of the video image.

5. Simulation and results analysis

For demonstrating the effectiveness and stability of the proposed method, the proposed algorithm is used to stabilize the retinal vessel layer and the retinal cell layer image sequence obtained by AOSLO. The original image sequence is shown in figure 1(a) and figure 2(a), the stabilization image sequence is shown in figure 1 (b) and figure 2 (b).

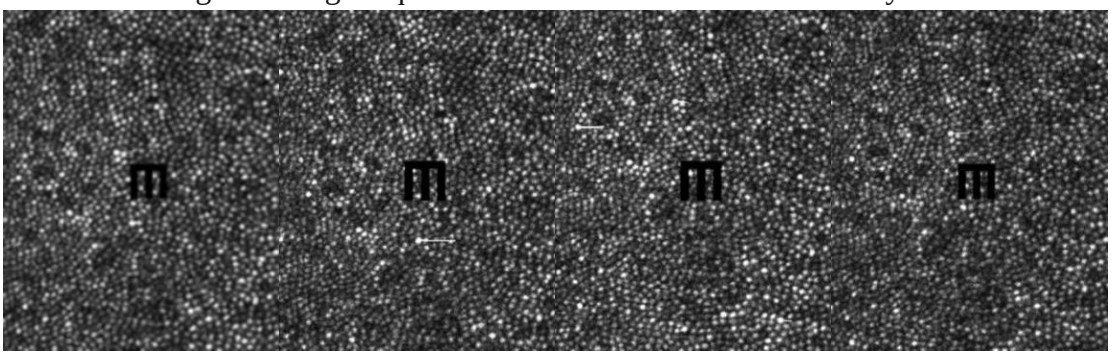


(a) Retinal vessel layer image sequence

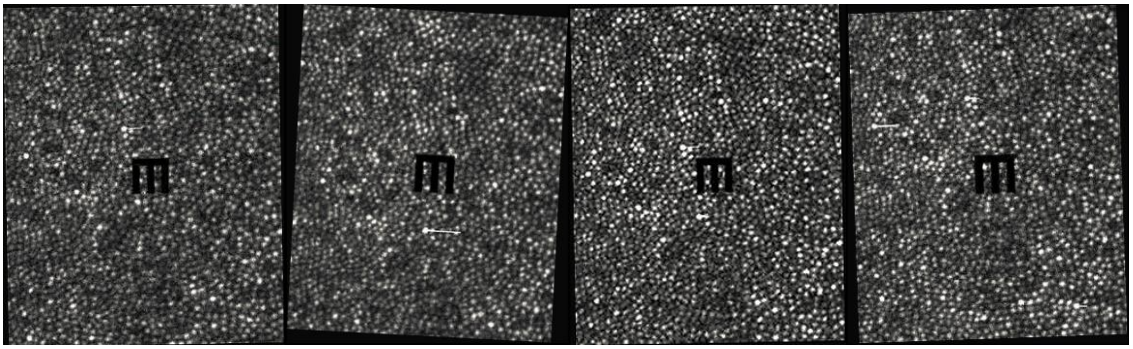


(b) Stabilized retinal vessel layer image sequence

Figure1 Image sequence stabilization of retinal vessel layer



(a) Retinal cell layer image sequence



(b) Stabilized retinal cell layer image sequence

Figure 2 Image sequence stabilization of retinal cell layer

As can be seen in figure 1(b) and figure 2 (b) , the effect of video stabilization is more and more obvious, which shows that the proposed algorithm has better stabilization effect for retinal vascular layer image and retinal cell layer image.

6. Conclusion

Aiming at the defects of existing AOSLO equipment output video image, an AOSLO video image stabilization algorithm based on Harris-SIFT is proposed. Many invariant feature points can be extracted by the SIFT algorithm with good scale, rotation, perspective and illumination invariance, but when it is applied to AOSLO video image with high real-time requirement , the complexity is high and real-time is poor, and the scale invariant feature is difficult to be reflected. The operator proposed Harris-SIFT is a successful improvement for SIFT, which can reduce the complexity of feature extraction and feature matching, greatly improve the real-time.

RANSAC matching verification and motion compensation improve the stability and accuracy of the algorithm. The experimental results show that the proposed image stabilization algorithm has achieved good results in improving the AOSLO video image.

Reference

[1] LOWE D G. Distinctive image features from scale-invariant key points [J]. International Journal of Computer Vision,2004, 60(2): 91-110.

[2]LOWE D G. Object recognition from local scale-invariant feature s[C] Proceedings of the International Conference on Computer Vision, 1999: 1150-1157.

[3]LINDEBERG T. Feature detection with automatic scale selection [J]. International Journal of Computer Vision,1998, 30(2): 79-116.

[4] MIKOLAJCZYK K, SCHMID C. A performance evaluation of local descriptors [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2005, 27(10): 1615-1630.

[5] MIKOLAJCZYK K, TUYTELAARS T, SCHMID C, et al. A comparison of affine region detectors and descriptors [J].International Journal of Computer Vision, 2005, 65(1):43-72.