A Video Stabilization Method based on Inter-Frame Image Matching Score

Qinghe Zheng[1]

[1] Shandong University

Received: 9 December 2016 Accepted: 4 January 2017 Published: 15 January 2017

Abstract

Video stabilization is an important video enhancement technology which aims at removing annoying shaky motion from videos. In this paper, we propose an robust and efficient video stabilization algorithm based on inter-frame image matching score. Firstly, image matching is performed by a method combining Maximally Stable Extremal Regions (MSERs) detection algorithm and Features from Accelerated Segment Test (FAST) corner detection algorithm, which can get the matching score and the motion parameters of the frame image. Then, the matching score is filtered to filter out the high frequency component and keep the low frequency component. And the motion compensation is performed on the current frame image according to the ratio of the matching score before and after the filtering to retain the global motion and remove the local jitter. Various classical corner detection operators and region matching operators are compared in experiments. And experimental results illustrate that the proposed method is effective to stabilize translational, rotational, and zooming jitter and robust to local motions, and has the state-of-the-art processing speed to meet the needs of real-time equipment.

Index terms— video stabilization, video warping, motion estimation, motion compensation, partial compensation.

1 I. Introduction

ideo enhancement is getting more and more attention with the increasing popularity of digital visual media. As one of the most important ways of enhancement, video stabilization is a technique for removing abnormal image offsets such as jitter and rotation, etc., by digital image processing. One of the most obvious differences between professional and amateur level video is the quality of camera motion; hand-held amateur video is typically shaky and undirected while professionals use careful planning. Given the unstable video, the video stabilization is designed to synthesize new image sequences seen from the new stable camera trajectory.

The typically algorithm mainly consists of the following parts: feature point extraction, feature point matching, motion estimation, motion compensation, synthesis of new video sequences. Prior techniques for software video stabilization follow two main approaches, providing either high quality or robustness and efficiency. At present, the most commonly used video image stabilization method is 2D stabilization [1], which is widely used in commercial software and military. This method is suitable for the 2D motion models, which is very effective for the affine or projection transformation of the current frame. However, due to the inability to simulate the camera movement caused by the disparity and other issues, the two-dimensional motion model is very fragile and poor stability. Then, 3D video stabilization technique was proposed by Buehler in 2001 [2] and developed by Liu in 2009 [3], which shows a strong stability and has the ability to simulate the camera's 3D trajectory. In this method, a new structurefrom-motion (SFM) technique [4] is used to construct the 3D model of the background and camera motion, and then various new filtering ideas are started around the new 3D trajectory model [5,6].

But SFM is a fundamentally difficult problem, and the generality of current solutions is limited when applied to the diverse camera motions of amateur-level video. The problem with 3D stabilization and 2D stabilization is opposite: the 3D model is too complex to be calculated in real time and the robustness is too poor. So it is difficult to use the 3D image stabilization technology in daily business and medical treatment. In general, requiring 3D reconstruction hinders the practicality of the 3D stabilization pipeline.

In this paper, we introduce a robust and efficient method for software video stabilization. In spite of the image stabilization platform has been widely used in professional equipment and achieved good results, it still requires additional hardware support, and isn't suitable for amateur consumers. For example, video quality will be severely reduced due to camera vibration in situations like taking pictures by a tourist enthusiast on a bumpy car.

2 II. Video Stabilization Algorithm

Video stabilization mainly includes four stages: image pre-processing, image matching, motion estimation and motion compensation (see Fig. 1). Image pre-processing is to eliminate the interference of fuzzy, gray shift and geometric distortion caused by the inconsistency of the light in the process of obtaining the video, which is able to reduce the difficulty of image matching and improve the accuracy of image matching. Image matching is the key step of video stabilization, which directly determines the quality of the final video. The purpose of image matching is to find a spatial transformation, so that the coordinates of the overlapping parts in the image can be accurately matched. Image matching algorithm needs not only to ensure the accuracy of image matching, but also to minimize the amount of computation. Motion estimation is a complete set of techniques for extracting motion information from video sequences. The main content of motion estimation is how to get enough motion vectors quickly and effectively according to the coordinates of matching feature points. Motion compensation is to predict and compensate the current image by the previous image, and to compensate the corresponding motion information of the previous frame according to the motion vector. The key of motion compensation is to distinguish local jitter and global motion effectively, which makes the final video get a good visual effect.

3 III. Image Matching

In this part, we will introduce two classical image matching algorithms which are used in the fourth part: MSERs algorithm [7] proposed in 2014 and FAST corner detection algorithm [8] proposed in 2012 used for video stabilization.

4 a) Region-based matching algorithm

MSERs use the concept of a watershed in the terrain to find a stable local area. Previous watershed transforms were mainly used for image segmentation. The algorithm focused on the water level at the time of regional merging. At this time, the small water puddles and ponds were unstable and the connected water volume changed drastically. Strictly defined from the mathematical point of view, MSER is a region which has the smallest change in the number of pixels at a given threshold. MSERs is currently recognized as the best performance of the affine invariant region.

5 Algorithm Steps

6 ?

The pixels of a given image are ordered in gray scale values.

7 ?

Add the pixels into images in accordance with ascending or descending and link the area.

8 ?

Define Q as an arbitrary connected region in the binary image corresponding to the threshold value. When threshold changes in (i-?, i+?), connected regions corresponding to Q i+? and Q i-? . Within this range of variation, the region q(i) with minimal change rate is considered to be MSERs.

9 b) Feature-based matching algorithm

FAST is a corner detection method, which can be used for the extraction of feature points and the completion of tracking and mapping objects. The most prominent advantage of this algorithm is its computational efficiency and good repeatability. The basic principle of the algorithm is to use a circumference of 16 pixels (a circle with a radius of 3 pixels drawn by the Bresenham algorithm) to determine whether the center pixel P is the corner point. Then the center pixel is called the corner point: If the brightness of N pixels on the circumference are larger than the sum of center pixel and a threshold T, or smaller than the difference between the center and the threshold T. In an image, the non corner points are more easily measured and accounted for the majority of the pixels. Therefore, the first elimination of non corner points will greatly improve the detection rate of corner points.

10 Algorithm steps

11 ?

Detect the non corner points on the circle. ? Determine whether the center point is a corner point and make a corner detection for each point on the circle if it is true.

12 ?

Remove the non-maximum corner and get the output corner point.

13 IV. Motion Estimation

2D parametric motion model is used for the motion of the camera (see Fig. **??**). The moving camera is attached to the coordinate system O-XYZ and the corresponding projection onto the image plane is attached to the system O-PQ. The camera motion consists of two components: a translation (T x , T y , T z) T and a rotation (?, ?, ?) T , which represent roll, pitch and yaw of the motion. A point with an image coordinate (p, q) in the space (x, y, z) will move to another location (x', y', z') with an image coordinate (p', q') and the focal length f c will become f' c through inter frame motion. The relationship of corner points in space and the image plane is defined by Eq. (1) and Eq. (2), respectively. a, b, c, d, e, f, g, h, i among the equation is the parameters of motion matrix.

14 Image plane and the coordinate plane

? ? ? ? ? ? ? ? ? + + ? + + = ? + + ? + + = z T f if hq gp z T f ff eq dp f q z T f if hq gp z T f cf bq ap f p z c c y c c c z c c x c c c / / / / ' ' ' '**(1)**

If the rotation angle of the collected video sequence frame image in the camera motion process is less than 5 ? , Eq. (2) can be approximated as:? ? ? ? ? ? ? ? + ? ? + + ? = ? + ? ? + = z T f f q p z T f f q p f q z T f f q p z T f f q p f p T f f q p s ? + ? = ? ? (4)

Then Eq. (2) can be expressed as:? ? ? ? ? ? + + ? = ? ? ? + = ? z T f f q p q s z T f f q p p s y c c x c c / / ' ' ? ? ? ? (5)

Two equations are provided by each set of matching corner points, thus 2N equations will be provided by N pairs, and subsequently, the motion parameters can be obtained by the least square solution.

15 a) Feature point selection

In the traditional method, the motion equation is obtained by detecting and matching the feature points between the frames. Since there are a large number of matching feature points in two adjacent frames to solve a motion equation containing only four parameters, there is a large computational redundancy. At the same time, image feature point matching is prone to mismatch. So the traditional methods need to add a wild point elimination function, which used to remove unreliable feature points that easily lead to false matches. We propose a novel feature point detection method for solving the equations of motion combining the advantages of feature point detection method and region detection method. Firstly, MSERs detection is performed on each image in the video sequence (see To track feature points, a window P*P centered at each selected point is designed and matched using diamond search (DS) method and the sum of absolute difference criterion (SAD) [9]. The searching area is (P+2M)×(P+2N), where M and N are maximum horizontal and vertical displacements, respectively. Thus, the corresponding point is at the center of the matching window. Moreover, two issues are considered in deciding the proper size of the feature window: A large size would cause a dislocation of pixels, but a small size offers less information. In practical use, a feature window with a size 9*9 has a good performance experimentally. Next, the first chapter of the N pictures in each second of the video sequence is set as a reference frame, and the remaining images in each second are matched with the reference frame for feature points. Finally, the least squares method is used to solve the motion equation through the coordinates of matching feature points. In this way, stable and effective feature points can be obtained, which is more robust to noise such as illumination. At the same time, the reduction of the number of feature points can solve the equations of motion more quickly.

16 c) Computing motion parameters

The Eq. (**??**) indicates that the motion includes four parameters: the rotation Î?”?, the translation (Î?”x, Î?”y), and the scaling Î?”?. Given a set of N matched pairs, Î?”? can be defined as:? ? = = × × = ? N i i i N i i i d d d d 1 1 ' ? (6)

where,( ) ( ) 2 2 V v U u d i i i ? + ? = (7) ( ) ( ) 2 ' ' 2 ' ' ' V v U u d i i i ? + ? = (8)

where (U, V) represents the bary center of the points in the current frame, and (U ' , V ') represents the bary center in the reference, respectively. obtained with three unknowns m=[Î?”?, Î?”x, Î?”y] T . The final function B=Am is in the form of a matrix, as shown in Eq. (**??**). Then applying Eq. (**??**) to a set of N pairs of matching feature points, 2N linear functions can be the template matching processing are selected and the initial value of m can be computed by m = (A T A) -1 A T B. Then, the L-M method is used to refine solutions by minimizing the square of coordinate differences. Let (u i , v i) T and (U i , V i) T denote the known feature points and the estimated points, respectively. The object function is defined as Eq. (10).? = ? = 1 0 0 1 ... ...

17 ( ) ( )

[ ]e e V v U u V v U u V v U u E i i i i i i N i i i i i T T 1 2 2 = ? ? ? ? ? ? ? ? ? ? ? ? ? ? ? ? ? = ? + ? = ? =**(10)**

18 V. Motion Compensation

At this stage, it is clear that only the unwanted camera jitter should be removed in the motion of the camera. We suppose that the motion of the camera is usually smooth with slow variation and unwanted camera jitter involves rapid variation. From another point of view, the high frequency component of the motion vector is considered to be an unwanted camera jitter and can be filtered out by a low pass filter. On the basis of this idea, we propose a partial backward compensation method with a novel filtering algorithm. Firstly, the matching scores G i of i-th image and divided by the total number of corners. The higher the approximation degree between the reference frame and the current frame, the higher the matching score. On the next step, one dimensional discrete wavelet transform (DWT) is used to remove the high frequency information, and then one dimensional discrete wavelet reconstruction (IDWT) is performed. The Haar wavelet is used to carry out the wavelet transform. Wavelet transform is the inheritance and development of traditional Fourier transform. Because the multiresolution analysis of wavelet has good localization property in spatial domain and frequency domain, it can be used to analyze the arbitrary details of the object gradually. Retaining the low-frequency information, which is equivalent to retaining the global movement and removing the jitter. Finally, adopt the partial compensation principle: the motion parameters are compensated according to the ratio of the matching scores before and after the filtering, which is given by Eq. (11), where G ib and G ia is the image matching score before and after the wavelet transform. A strategy will be used in the mage matching score calculation phase: remove the current frame with matching scores below a preset threshold N (30 in this paper). In the end, the processed image frame is made into a new video. (11) Compared to the traditional algorithms, which need to filter the horizontal displacement, the vertical displacement and the rotation angle, our method only needs to filter the image matching score, which improves the efficiency and meet the real-time requirement. At the same time, the partial compensation method based on the image matching score can better retain the global motion and avoid the phenomenon of over smoothing.

19 VI. Experimental Results

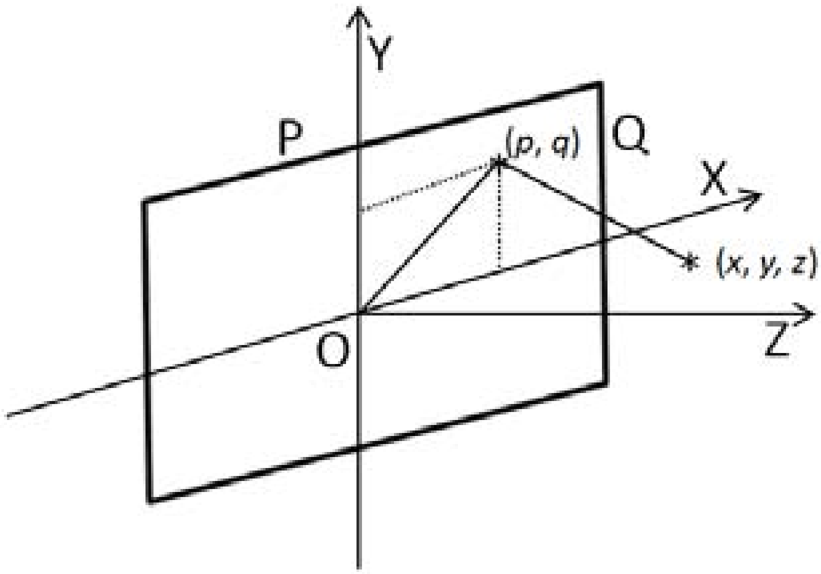
This part presents experimental results obtained from a video sequence, which is widely used by various video processing laboratories. The experiments are carried out on MATLAB R2013a with a i5-4460 CPU. The input video has a resolution of 360×240 and includes 400 frame images in 10s.Firstly, we compare the performance of various classical corner detection operators and region detection operators(see Tab.1). The experiment results show that method combining MSERs and FAST has a faster computing speed than the tradition algorithm by finding and matching the feature points of the entire image. To make an objective evaluation of the video stabilization algorithms, the peak signal-to-noise ratio (PSNR) can be used as a measure. In Fig. **??**, we compare the PSNR of 40 mean images processed by a variety of operators and traditional algorithm. We can see that the performance of the new algorithm is generally better than the traditional algorithm from the experiment results. To make a subjective evaluation of the results, the mean image of first 10 consecutive frame images in the original and stabilized video sequences are given respectively, as shown in Fig. **??**.

20 VII. Conclusions

A robust and fast video stabilization method is proposed, which consists of image matching based on MSERs detection and FAST corner detection, motion estimation and motion compensation based on interframe matching score. The partial compensation method based on inter-frame matching score efficiently removes fluctuations and retains global motion. The speed optimization of algorithm and its low cost and low requirements of equipment hardware makes it possible to be used for non-professional camera enthusiasts and the portable electronic equipment like hand-held visual communication device. The most time-consuming phase of the algorithm is the area detection. A more simple and effective feature region detection method and fast sorting algorithm can make it faster, which needs to be further optimized in future research.

21 VIII. Acknowledgment

This work is partially supported by Shandong Provincial Natural Science Foundation, China (ZR2014 FM030 and ZR2014FM010). [1]



1

Figure 1: Figure 1 :

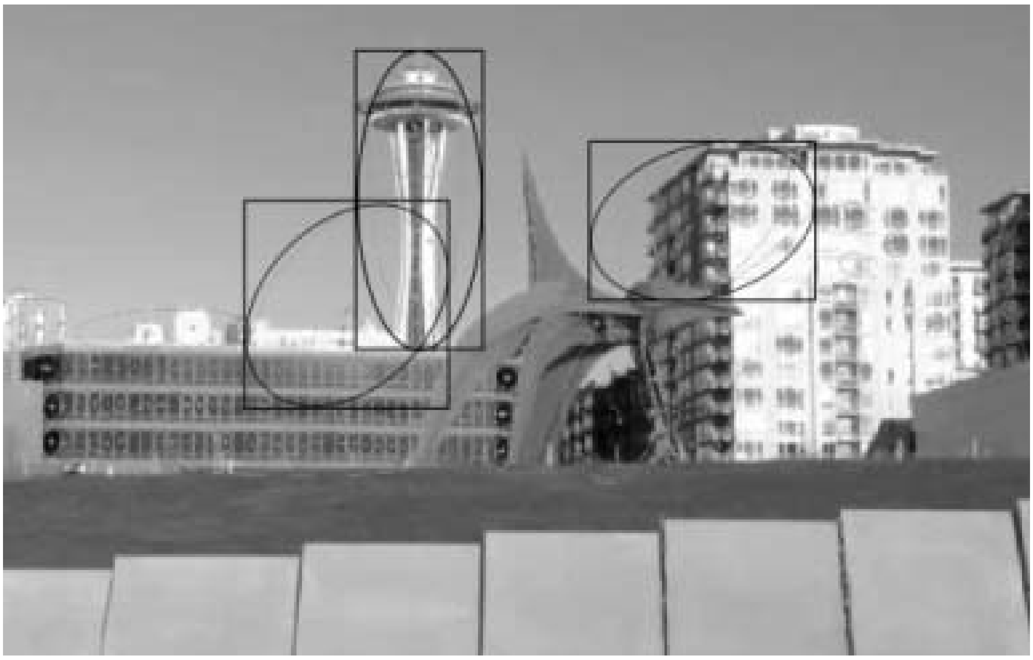


Figure 2: FA

23 

Figure 3: Figure 2 :?Figure 3 :

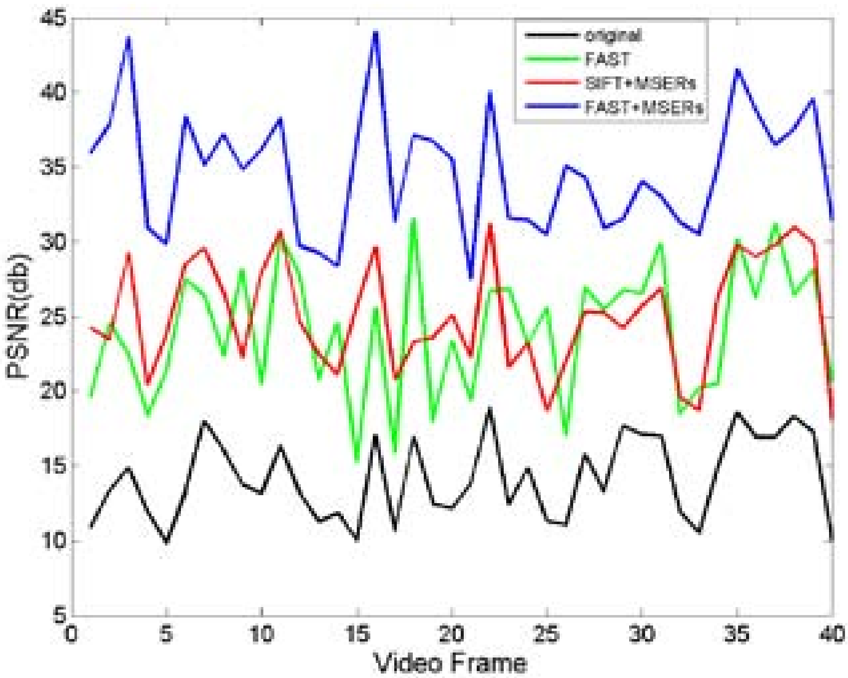
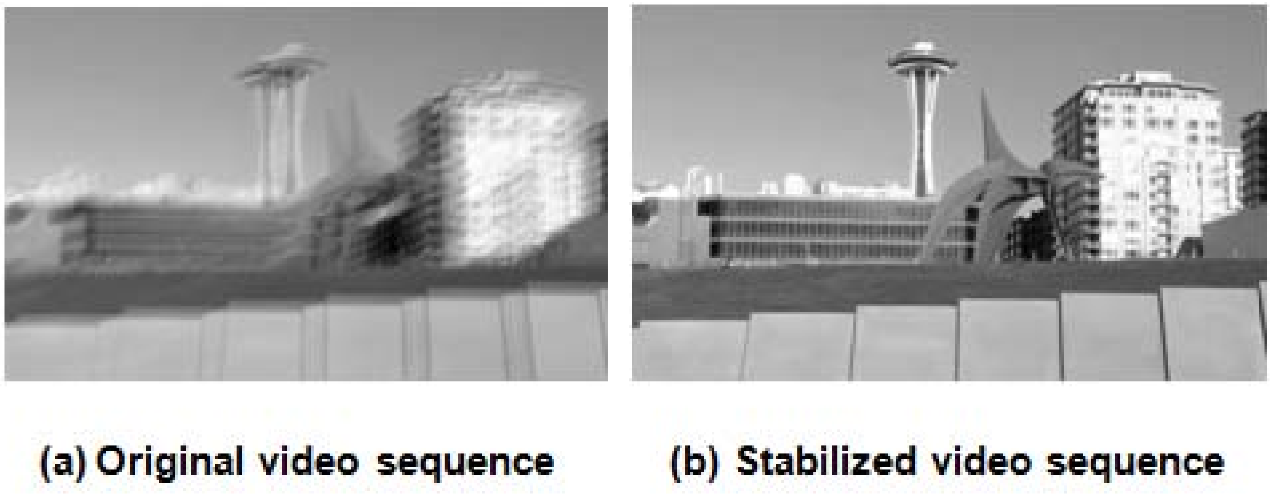
1 

Figure 4: 1 FA



45

Figure 5: Figure 4 :Figure 5 :

1

A Video Stabilization Method based on Inter-Frame Image Matching Score
reference image in the video sequence are defined as:
the number of successful matching corner points

Method	Mean value of corners	Computational time(s)	Year 2017
SIFT [11]	81	299.62	
SURF [12] Harris [13] FAST	165 135 59	151.96 96.68 36.14	( ) F
MSERs+SIFT	21	146.88	
MSERs+SURF	47	60.29	
MSERs+Harris	34	29.18	
MSERs+FAST	19	12.82	

Figure 6: Table 1 :

To get the motion parameters, the initial solutions are obtained by pseudo inverset transformation and then refined by Levenberg-Marquardt (LM) method [10]. Firstly, n(n?2) pairs of points with minimal SAD in

.1 Global Journal of Computer Science and Technology

Volume XVII Issue I Version I

[Iete Journal of Research ()] , *Iete Journal of Research* 2014. 60 (5) p. .

[Marquardt ()] 'An Algorithm for Least-Squares Estimation of Nonlinear Parameters'. D Marquardt . *Journal of the Society for Industrial & Applied Mathematics* 2006. 11 (2) p. .

[Liu et al. ()] 'Content-preserving warps for 3d video stabilization'. F Liu , M Gleicher , H Jin , A Agarwala . *Acm Transactions on Graphics* 2009. 28 (3) p. .

[Xu et al. ()] 'Fast feature-based video stabilization without accumulative global motion estimation'. J Xu , H W Chang , S Yang . *IEEE Transactions on Consumer Electronics* 2012. 58 (3) p. .

[Huang et al.] *Feature-based video stabilization for vehicular applications*, K Y Huang , Y M Tsai , C C Tsai .

[Matsushita et al. ()] 'Full-frame video stabilization with motion inpainting'. Y Matsushita , E Ofek , W Ge , X Tang , H Y Shum . *IEEE Transactions on Pattern Analysis & Machine Intelligence* 2006. 28 (7) p. .

[IEEE International Symposium on Consumer Electronics. IEEE] *IEEE International Symposium on Consumer Electronics. IEEE*, 2010 p. .

[Biswas] *Improving Video Stabilization Using Multi-Resolution MSER Features*, P Biswas .

[Buehler et al. ()] 'Non-Metric Image-Based Rendering for Video Stabilization. Computer Vision and Pattern Recognition'. C Buehler , M Bosse , L Mcmillan . *Proceedings of the 2001 IEEE Computer Society Conference on*, (the 2001 IEEE Computer Society Conference on) 2001. 2001. 2001. 2 p. 609.

[Tourapis and Au] *Predictive motion vector field adaptive search technique (PMVFAST): enhancing block-based motion estimation*, A M Tourapis , O C L Au .

[Proceedings of SPIE -The International Society for Optical Engineering ()] *Proceedings of SPIE -The International Society for Optical Engineering*, (SPIE -The International Society for Optical Engineering) 2001. 4310 p. .

[Ratakonda ()] 'Real-time digital video stabilization for multi-media applications'. K Ratakonda . *IEEE International Symposium on Circuits and Systems*, 1998. 4 p. .

[Liu et al. ()] 'Subspace video stabilization'. F Liu , M Gleicher , J Wang , H Jin , A Agarwala . *Acm Transactions on Graphics* 2011. 30 (1) p. .

[Zhang et al. (2011)] 'Video stabilization based on saliency driven SIFT matching and discriminative'. Y Zhang , H Yao , P Xu . *Icimcs 2011, the Third International Conference on Internet Multimedia Computing and Service*, (Chengdu, China) August. 2011. p. .

[Huang et al. ()] 'Video stabilization for vehicular applications using SURFlike descriptor and KD-tree'. K Y Huang , Y M Tsai , C C Tsai . *International Conference on Image Processing*, 2010. p. .