Video Stabilization System Based on Speeded-up Robust Features

Xie Zheng, Cui Shaohui, Wang Gang, Li Jinlun

Mechanical Engineering College, Shijiazhuang, 050003, China

Email: 971997256@qq.com

Keywords: Video Stabilization, Feature Extraction, Motion Estimation, SURF, RANSAC

Abstract. In this paper,a fast and efficient video stabilization method based on speeded-up robust features (SURF) is presented .We adopted speeded-up robust features as feature descriptor,which are extracted and tracked in each frame .After that, we further refined the matching features through RANSAC, estimating the motion parameters through least squares method and computed the integrated motion. Experimental results illustrate superior performance of the SURF based video stabilization in terms of accuracy and speed when compared with the Scale Invariant Feature Transform (SIFT) based stabilization method.

Introduction

Video sequences acquired by a camera mounted on a mobile platform are affected by unwanted shakes and jitters. Several digital video stabilization approaches have been proposed to overcome this problem such as block matching[1],optical flow[2],phase correlation[3]and feature matching[4]. Block-matching methods divide a frame into blocks,and compute a motion vector for each one through the searching of the more similar block in the next frame.However,the motion estimation could be biased in low-textured image regions due to the aperture problem.Feature-based methods overcome this problem by computing the motion only in regions that stand out according to a specific image feature.At present,with the feature point extraction technology development,based on feature points matching algorithm to stabilize the image has become the electronic technology to stabilize the image of the main development direction.Commonly used in electronic image Feature points are: Harris corner[5],Smallest Univalue Segment Assimilating Nucleu（SUSAN)[6],Scale Invariant Feature Transform（SIFT)[7],Speed Up Robust Feature（SURF)[8].In this context,a novel digital video stabilization algorithm is proposed, which overcomes the previous problems by computing a robust motion estimation through a variation of the SURF algorithm adapted to video sequences to bediscriminative to scale and orientation.

SURF feature extraction and matching

In this paper, using the algorithm of SURF as feature point extraction algorithm of image registration.SURF is a translation, rotation and scale invariant feature detector, which is based on Hessian matrix for its good performance in accuracy[8].Compared with SIFT which uses Difference of Gaussian (DoG) to approximate Laplacian of Gaussian (LoG), SURF pushes the approximation even further. It approximates Laplacian of Gaussian by using a box filter to represent the corresponding kernel. The kernel approximation is efficient in computation by using the integral images and hence the time consumption is independent of the filter size.After building the image pyramid, the process continues by traversing the pyramid to remove those points with low contrast and then searching extrema on neighboring scale images.

Finally,the points are localized to sub-pixel accuracy through scale space interpolation. SURF descriptor uses Haar wavelet in x and y directions to compute a reproducible orientation. To achieve rotation invariance, a square descriptor region is constructed along the dominant orientation and is divided into 4×4 sub-regions and the descriptor is extracted from it.In addition,SURF computes Haar wavelet through integral images, which decreases the computational complexity.Each wavelet requires only six operations to perform the computation.

Since SURF divides all the feature points into two types by the sign of Laplacian,we can boost the matching speed by comparing the sign of Laplacian.In addition,we drop the unreliable matching by comparing the ratio of distances from the closest neighbor to the distance of the next closest neighbor with a predetermined threshold.

Motion estimation

The real camera motion between frames is a 3D motion.As a trade off between the complexity and efficiency, we adopt a 2D affine model to describe the motion between frames:

$$\begin{bmatrix} X_i \\ Y_i \end{bmatrix} = \begin{bmatrix} a_1 & a_2 \\ a_4 & a_5 \end{bmatrix} \begin{bmatrix} x_i \\ y_i \end{bmatrix} + \begin{bmatrix} a_3 \\ a_6 \end{bmatrix} \quad (1)$$

This model describes the pixel displacement between two frames,where X_i and Y_i represent the pixel position in the current frame,and and y_i represent the pixel position in the next frame. It includes 6 parameters: a_2 and a_4 are the rotation factor, a_1 and a_5 are the zoom factor, a_3 and a_6 are the shift in x and y directions。

To estimate these parameters,we need at least 2 pairs of matching features. After we extract the SURF features from the two consecutive frames, we can put these pairs of features into the affine model. We can solve this equation through least squares estimation method. Though we have roughly eliminated those unreliable matched features through comparing the ratio of distances with preset threshold, the local motion vectors still contain some mismatched features. The local motion vectors may also contain the matched features belonging to the ego-moving objects which cannot reflect to the camera motion. Since the least squares method is sensitive to outliers, it would introduce estimation error if we estimate the motion parameters directly. To solve this problem and get the exact motion parameters, we adopt the Random Sample Consensus (RANSAC)[9]to refine the matched features.

This idea is to iteratively guess the model parameters using minimal subsets of points randomly drawn from the input features. Figure.1 illustrates the comparis between original feature set matching and the refined feature set matching. It has totally 68 pairs of matched features in the left image. The mismatched features are also inclued in the illustration. In the right image, all the mismatched features and some of the matched features are removed and the number of matched features is reduced to 13.

Fig.1. comparis between original feature set matching and the refined feature set matching

The motion vectors between frames can be divided into two parts: undesired jitter and intentional camera motion.Directly approximating the motion parameters with the original motion vectors would cause errors since only the undesired jitter need to be compensated.We cannot just store the current frame and wait until the number of frames reaches certain amount and then process them together.We need a real-time motion separation method to fix this problem. Motion Vector Integration (MVI) with adaptive damping coefficient is a simple and quick method which can both filter the cumulative motion curve and change the damping extend according to the recent two global motion vectors.Actually,the cumulative motion vector at frame n is the summation of previous n-1 global motion vectors plus the global motion vector at frame n.In MVI, the cumulative motion vector at frame n-1 is multiplied by a damping coefficient δ which depends on the value of the latest 2 global motion vectors. The motion vector at frame n can be represented as:

$$IMV(n) = \delta \times IMV(n-1) + GMV(n) \quad (2)$$

Where GMV(n) is the global motion vector between frame n-1 and frame n.If the last two global

motion vectors are small, δ is set to a high value which is close to 1.In this case,the integrated motion vector at frame n could strongly stabilize the video.Correspondingly,if the last two global motion vectors are big, δ is set to a relatively low value to compensate the undesired small jitter and preserve the major camera trajectory.

Experimental Results

We evaluated the performance of the proposed method with several video sequences to observe the processing speed and ensure the number of features per frame.The experiment was carried out with Visual Studio 2010 in Windows Vista Operating System on an Intel Core 2 Duo 2.4GHz CPU system.We adopted PSNR to evaluate our video stabilization performance.PSNR is the corresponding Peak Signal-to-Noise Ratio between frame n-1 and frame n,which is defines as:

$$PSNR = 10\log\frac{(L-1)}{(e_{rms})^2} \tag{3}$$

Where e_{rms} is the root mean square betweenframe n and frame n-1.Figure.2 shows a set of video frames illustrating the effectiveness of our stabilization method. The 18th, 68th and 118th frames of the video are picked up and shown in Figure.2. The upper row shows the original video frames and the bottom row shows the stabilized video frames. Figure.3 shows the PSNR comparison between the original sequence and the stabilized sequence. As we can see, The stabilized video (green curve) has a higher PSNR than the original video (blue curve), and the average PSNR of stabilized video is much higher than that in the original one, which proves that the SURF based video stabilization system has better performance.

Fig.2. Experimental results

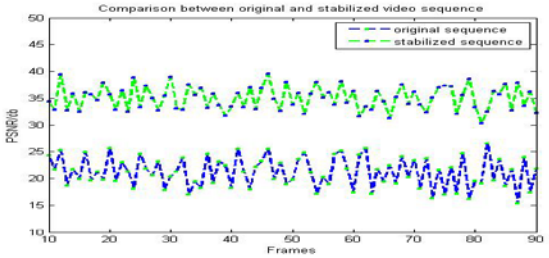

Fig.3. Comparison of PSNR

Conclusions

In this paper, we proposed an efficient approach for video stabilizationbased on SURF. The SURF features are extracted and tracked in each frame and thenfurther refined the matching features through RANSAC, estimating the affine motion parameters.Finally,we compensated the undesired jitter with the pre-computed motion parameters.The obtained results corroborate thesuperior performance of the SURF based video stabilization in terms of accuracy and speed when compared with the SIFTbased stabilization method,which would make real-time video stabilization system possible in larger video frames .

Acknowledgement

In this paper, the research was sponsored by the Nature Science Foundation of Hebei Province (Project No. 201311400380208) and Youth Fund Project of shijiazhuang Institute of Science and Technology (Project No. 2012QZ35).

References

[1] Battiato, S., Puglisi, G., Bruna, A.R. A Robust Video Stabilization System By Adaptive Motion Vectors Filtering. In: IEEE International Conference, pp. 373–376 (2008)

[2] Chang J Y，Hu W F，Cheng M H，et al. Digital image translation and rotation motion stabilization using optical flow technique[J]. IEEE Trans. on Consumer Electronics(S0098-3063)，2002，48(1)：108-115.

[3] Erturk S. Digital image stabilization with sub-image phase correlation based global motion estimation [J]. IEEE Trans. On Consumer Electronics(S0098-3063)，2003，49(4)：1320-1325.

[4].Lowe, D.G. Distinctive Image Features from Scale-Invariant Keypoints. International Journal of Computer Vision 2004

[5].Harris C，Stephens M. A combined corner and edge detector [C]// Fourth Alvey Vision Conference，Manchester，1988，147-151.

[6].Jiang Wentao,Chen Weidong,Li Fuliang.:Electronic image stabilization algorithm based on characteristic point tracking[J].Journal of Appliced Optics,2010,31(1):71-73

[7] Battiato, S., Gallo, G., Puglisi, G., Scellato, S. SIFT Features Tracking for Video Stabilization. In: Proc. ICIAP, pp. 825–830 2007

[8] Bay H, Ess A, Tuytelaars T, et al Speeded-up robust features(SURF). Computer Vision and Image Understanding 2008 110(3):346–359

[9] Fischler, M.A., Bolles, R.C. Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. Communication of ACM 4(6), 381–395 1981