PAPER REPORT

VIDEO STABILISATION ON HANDHELD DEVICES

Name: Shubham Kiran Ingale

NetID: ski227

N No: N19148465

Sub: DSP Lab

Table of Contents

Video Stabilisation on Handheld Devices

I have always been fascinated by the way handheld devices like the smart phones and digital cameras like DSLRs and mirrorless have improved upon there camera performance, Especially smartphones. Initially when smartphones came with camera sensors, their capabilities were very limited and lacked in various aspects like picture quality, stabilisation, night photography. But today we are at a stage where a smartphone like the iPhone 12 Pro Max comes with a feature like the sensor shift stabilisation technology, which is seen in professional grade video and photography cameras. While there are a lot of technologies that interest me that smartphones have to offer these days, but particularly in this paper we will take a look at various video stabilisation techniques that allow users to film without any stabilisation gear using these nifty devices.

Types of Video Stabilisation

A lot of cameras and smartphones today use various techniques to achieve video stabilisation. These are used either separately or more often as a combination of two or more. Stabilisation is achieved using both electromechanical systems as well as real time video processing. Some techniques also are employed in post processing to achieve stabilisation. Broadly the techniques can be classified into hardware and software techniques.

Hardware Techniques

The hardware techniques consist of optical image stabilisation and Sensor shift stabilisation. In these techniques the sensor or the lens is moved to compensate for the high frequency inter-frame jitter. These systems work in real time. These are a bit expensive due to their complex construction and packing in small areas as in a smartphone. These are very effective in reducing jitter while walking with the device.

Software Techniques

On the software front there are namely two approaches, real-time stabilisation and stabilisation in post processing. Software stabilisation is also referred to as EIS or Electronic Image Stabilisation (Digital Video Stabilisation). We will take a look at both techniques. We will also take a look at the issues of a portable sensors and how is stabilisation achieved in it. And how zooming is stabilised in a video.

Real Time Stabilisation:

In digital video stabilisation there are three main steps. Step one is called motion estimation, followed by step two which is motion smoothing and then the step three is called motion compensation. Motion estimation is carried out between a reference frame and acquired frames. Reference frame is the first frame when EIS is turn on, on a device. All the frames that follow are called acquired frames. After a particular interval the reference frame changes to some new frame in the future and whole process continues. Between the reference and acquired frames we try to compute the shift. For example, if we are shooting a video, usually there are instances when we have to pan the camera to point it to the subject or zoom. The panning motion is the motion that should be estimated by the technique so that the acquired frames are shifted accordingly so that the intent of the video (the motions carried out with the camera, panning or zooming) is preserved. Once we compute the shift, we apply this shifting to the acquired frames.

Now the way we find out shift is very interesting. The motion estimation is carried out by tracking features. Features are roughly categorized into two, local and global features. Local features are features like points, faces, edges, colours and textures. The local features have higher tolerance to occlusion but are susceptible to noise. The global features are like lane lines and road exit points in images.

So, the positions of these local and global features are tracked over a number of frames. The position of features tracked across the frames is called feature motion vector. Since the tracking of these features in subsequent frames may result is an unexpected movement at first, but when done over a certain number of frames, feature trajectories can be achieved. We use filters to optimise trajectories and form smooth path. Optimisation in simple sense could be considered like an average motion trajectory.

Feature Extraction:

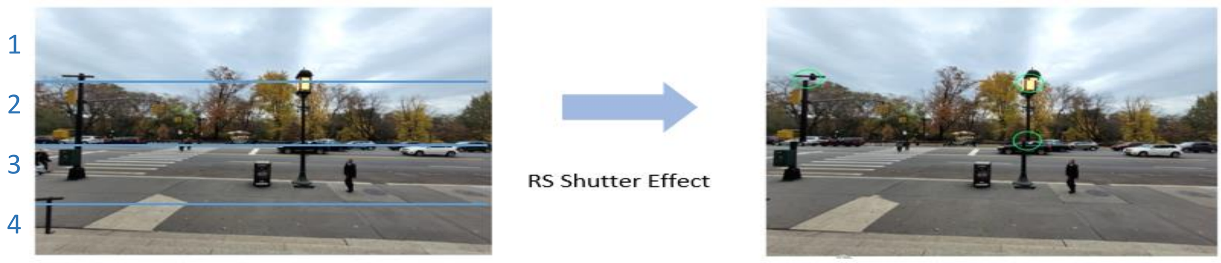
Features are obtained from reference frame. They are found out by searching and mapping algorithms applied to the reference frames. Before the searching and mapping, the reference frame is divided into regions. This is because if there is a moving object in the video, that can add a lot of noise in the motion vectors of the features, dividing allows us to look at these regions in isolation and filter accordingly later on, so that we get a better sense of actual resultant motion or path of the camera movement.

Reference Frame | Reference Frame with features being marked | Acquired Frame with new position of features

Motion Compensation is the last step in the process. In this process we make sure that undefined areas are not included in the final output. After motion smoothing, the path formed can lead to undefined areas at the border. These areas should be addressed so that the video output of the stabilisation is does not include undefined regions. There are two techniques that are used to solve this problem. The first one is called mosaicking. In this technique the information from previous frames is used to fill the undefined area. Another technique is called motion inpainting. Here data from the previous frames (acquired frames) is used to create motion field to filled in the undefined area.

Stabilisation on Portable Camera Sensors

In a portable sensor like in a smartphone, the reduced size of sensor creates more issues for stabilisation. The small sensors capture images not at a glance but in rows or in column fashion, which is in rolling order. This is called **Rolling Shutter (RS)**. Other professional grade cameras have sensors with what is called as **Global Shutter (GS)**, which capture image in one go. Now due to rolling Shutter the video captured from camera while in motion from side to side is warped. In other words, the objects in the image or frame appear distorted (distorted here implies a shift in position of the frames). This shift is different for different cameras based on the implementation of the RS Shutter. RS shutter could roll vertically from top to bottom and vis versa, or they could roll horizontally from right to left and vis versa. On a vertical RS camera



with top down roll, the capture of image happens row by row, one after another. That introduces delay in processing the rows. So, in effect we have captured one row of an image frame slightly earlier than the 2^{nd} row, and the 2^{nd} row slightly earlier than the 3^{rd} row and so on. Stabilisation on RS Shutter implementation devices have to take care and resolve this issue of image warping.

Rolling Shutter Compensation

In order to remove RS image warping, we need to compensate of it. So, the way we compensate for it is to find out the position for every point X in first frame what is its location in the second frame. Here First and second frame are subsequent frames. Let us try to find it for a point x. To find out the time a point x was imaged, let us assume few parameters. Let us say that the point **x** in the image was in the $\mathbf{y^{th}}$ row in an image taken from RS shutter camera, and let us assume that frame was taken at time **ti** where i is the frame number. Let **h** be the number of total rows. Let **Td** be the delay in capturing each row in the photo. So, the Time when a point x is imaged is given by following equation

y => the row in which the point x is present

i => the image frame number

ti => time stamp of the frame i

Td => the delay in capturing a row a particular row

Tn => total time taken to capture

T(i,y) => x was imaged at this Time T given frame I and row y

h => Total rows in the RS captured image frame

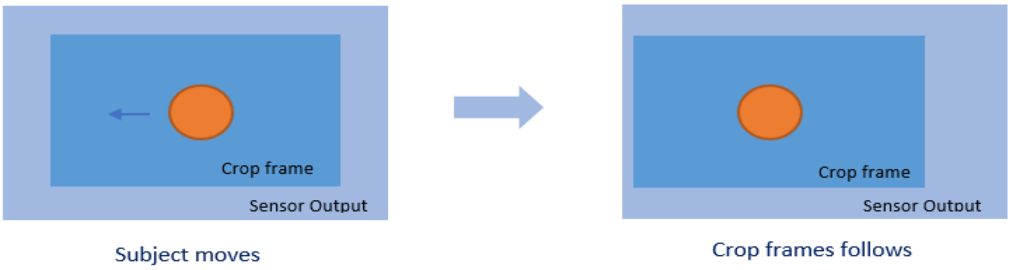
delay	**Td = Tn * y/h**
Capture Time	**T(i,y) = ti + Td**

The above function is used to find out the warping from every point Xi in I^{th} frame to every point Xj in J^{th} frame. J^{th} frame is the very next frame to the I^{th} frame. This allows us to calculate the warp for every point in the I frame which allows us to compensate for the Rolling Shutter. We calculate what is called warp matrix using the above formula. Which enables us to create an artificial camera that appears to posses a Global Shutter and produces stable frames.

Stabilisation while Zooming

In the above section we have already taken a look at how stabilisation is achieved. But can it be applied during zooming in the video? The answer is yes, the same above technique can be used to stabilise video while zooming, with some added nuances to the technique.

When zoomed in, the image frame is very prone to jitter and shakes caused by even the tiniest of movements. To address this problem, most techniques crop in on the image sensor output and use the discarded portion to keep the subject in centre. And then move the cropped window to counter act shakes by tracking the subject features and try keeping it in the centre.



Stabilisation through Post Processing

Post processing stabilisation is not real time technique but computed after the video has been recorded. It uses similar process flow that we follow for real time processing. Here the first step is estimating the camera path in the unstable video. The second step is finding smooth paths for the video. Third step creating stabilised output based on the smooth paths.

In the Second step the algorithm tries to find not a single smooth path but a combination of different smooth paths that are later merged to form a stabilised video. The smooths paths have a defined geometry, some of these paths have constant movement, parabolic or linear movement. Depending on the closest match of a segment in the video these smooth paths are selected.

A crop window is used for movement across the formed path to generate the stabilised output. Cropping improves stability against interframe jitters caused by movements similar to walking, trekking etc. Post processing gives a lot of creative control to the user in terms of the knowledge of the scene and trial and error, to find the desired output which doesn't lose valuable content.

References

Paper Citations

M. Grundmann, V. Kwatra and I. Essa, "Auto-directed video stabilization with robust L1 optimal camera paths," CVPR 2011, Providence, RI, 2011, pp. 225-232, doi: 10.1109/CVPR.2011.5995525.

Y. Zhang, Y. Leng and X. He, "A fast video stabilization algorithm with unexpected motion prediction strategy," 2015 IEEE International Conference on Advanced Intelligent Mechatronics (AIM), Busan, 2015, pp. 571-576, doi: 10.1109/AIM.2015.7222597.

Z. Li, S. Pundlik and G. Luo, "Stabilization of Magnified Videos on a Mobile Device for Visually Impaired," 2013 IEEE Conference on Computer Vision and Pattern Recognition Workshops, Portland, OR, 2013, pp. 54-55, doi: 10.1109/CVPRW.2013.15.

https://graphics.stanford.edu/papers/stabilization/karpenko_gyro.pdf