Optical Flow estimation & Global motion estimation in the image plane with RANSAC algorithm

Mohammad Mohaiminul Islam

I. Introduction

This report is generated as an analytical description for the project optical flow and global motion estimation in the image plane with RANSAC algorithm. The objective of the work is two fold. First, Estimating the optical flow and global motion with RANSAC algorithm. Second, Measure the information content derived from the video sequences with different metrics such as Mean squared error(MSE), Entropy and Peak signal to noise ratio (PSNR) for study. Later a comparative analysis was carried out between the response of these each metrics for set of video sequences containing different kinds of motion in the sequences. Also the effects of hyper-parameter deltaT has been explored.

II. Background

In this section theoretical background have been discussed for all the topics under consideration for this project.

A. Motion Estimation

Motion is an intrinsic property of the world and an integral part of our visual experience.Motion estimation is the process of obtaining the motion vectors that describe the displacement of the each pixel from one two dimensional image to another. Usually for video sequences motion are estimated between two consecutive frames. Estimating motion is an inherently difficult problem as the motion is in three dimension but the images are just projection of three dimensional scene onto a two dimensional image plane [1]. Here we will estimate the the global motion with the Optical Flow.

B. Optical Flow Estimation

Optical flow is the pattern of apparent motion of objects, surfaces, and edges in a visual scene. It is the motion of objects between consecutive frames of sequence, caused by the relative movement between the object and camera [2]. We can describe the problem of optical flow by fig. 1 . Here the image intensity *I* can be expressed as the function of spatial dimension (x,y) and time *t*. That is if we take the first frame *I(x,y,t)* and shifts its pixel spatially by *(dx,dy)* over time *t* we have our new image *I(x+dx,y+dy,t)*. One of the main important assumption to be made is pixel intensity remains same for between two frames as shown in equation (1) [3].

$$I(x + dx, y + dy, t + dt) = I(x, y, t) \tag{1}$$

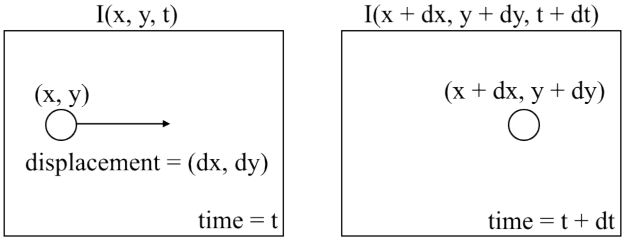

Fig. 1. Optical Flow problem illustration

Source : https://nanonets.com/blog/optical-flow/

Now if we take the Taylor series approximation of the RHS we get equation (2) and after eliminating common factors we get equation (3)

$$I(x+dx, y+dy, t+dt) = I(x, y, t) + dx\frac{\partial I}{\partial x} + dy\frac{\partial I}{\partial y} + dt\frac{\partial I}{\partial t} \tag{2}$$

$$dx\frac{\partial I}{\partial x} + dy\frac{\partial I}{\partial y} + dt\frac{\partial I}{\partial t} = 0 \tag{3}$$

Finally we divide (3) by *dt* to derive the Optical Flow equation as following in the equation (4)

$$\frac{\partial I}{\partial x}u + \frac{\partial I}{\partial y}v + \frac{\partial I}{\partial t} = 0 \tag{4}$$

where *dx/dt=u* and *dy/dt=v*. From each term of equation 4 we get the image gradients along the horizontal axis, the vertical axis,and time. Hence,the problem of optical flow is solving *u* and *v* to determine movement over time.

C. Measurement Metrics

in the subsection all the metrics for measuring the information content of the motion video sequence have been discussed.

- **Mean Squared Error (MSE).** Measures the avg. of squares of distances or errors. Which is the avg. squared difference between the actual and estimated values. MSE is a risk function, corresponding to the expected value of the squared error loss[4]. That MSE is almost always strictly positive can be attributed to either randomness, or the fact that the estimator does not account for information that could produce a more accurate estimate.[3] The MSE is a measure of the quality of an estimator—it is always non-negative, and values closer to zero are better[5]. In the context of measuring the motion information content MSE characterizes contrasts and motion between frames/sequences under consideration. Equation 5 shows MSE considering *p* is the spatial location of the an individual pixel and *t* is time.

$$MSE = \frac{1}{m * n} . \sum I(p,t) - I(p, t - \Delta t) \qquad (5)$$

- **Peak Signal to Noise Ratio (PSNR).** The ratio between the maximum possible power of a signal and the power of corrupting noise that affects the fidelity of its representation. It is also a measurement of distortion used in digital images , especially in image compression . It makes it possible to quantify the performance of the encoders by measuring the quality of reconstruction of the compressed image compared to the original image[6]. Considering the the height of the image is *m* and width of the image is *n* equation 6 describe the formula for calculation.

$$PSNR = 10 \log_{10} \frac{255^2}{MSE} \qquad (6)$$

- **Entropy.** the entropy of a random variable is the average level of "information", "surprise", or "uncertainty" inherent in the variable's possible outcomes [7]. In this context it measures the amount of information in a system. We study the entropy of an original motion picture sequence and of the error sequence which is the difference between consecutive frames of a motion image sequence. Equation 7 describe the Entropy considering *p (xi)* is the probability to have the certain gray value x$_i$

$$Entropy = -\sum p(x_i) \log_2 p(x_i) \qquad (7)$$

III. Dataset

The testing dataset are composed with the real-world video sequence in six different scenario with wide variety for situation , camera position , camera motion and object motion inside of them. In some of the video sequence sudden motion of object or camera is also present. All the different situation poses different kinds of challenges in term of optical flow and motion estimation.

IV. Implementation

In this project for implementing motion estimation and measurement of information content in motion video sequence Python has been as Programming Language.Also Open CV2 and Python have been used for developing all the functionalities. Optical flow has been calculated using algorithm described by Gunnar Farneback [3] which is available in Open CV2.

V. Result & Analysis

In this section results generated from the motion video sequence have been presented and discussed. Also effects of the controllable parameter of the system *deltaT* has been discussed.

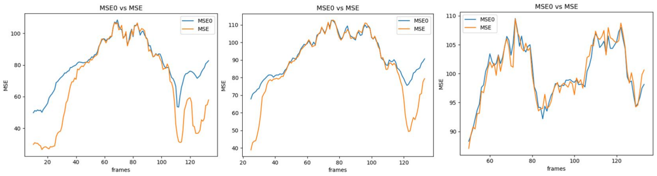

Fig. 2. MSE at deltaT value 10, 25 and 50 respectively (left to right) for video sequence LongJump. *MSE0* denotes error before compensation of global motion and*MSE* denotes error after compensation of global motion

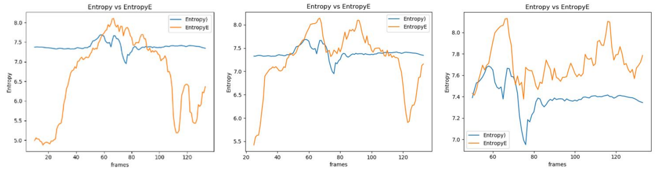

Fig. 3. Entropy at deltaT value 10, 25 and 50 respectively (left to right) for video sequence LongJump. *Entropy* denote entropy of current frame and *EntropyE* error image

The value of the *deltaT* determine the gap between the previous frame and current frame, in other words reference frame and current frame. if we analyze the measurement metrics graph to understand the effect, it can be seen in the fig. 2 that as we increase the deltaT value, the upper bound of mean squared error increases.When the deltaT value is 10, the MSE range is 0 to 100 but when we increase the deltaT value to 25 or 50 the MSE range goes to 0 to 110. Also as the deltaT value increased the shape of the plot changes. This can be explained as, if the deltaT values is small , reference frame and current frame is not very apart from each other hence the variation in motion and pixel intensity tend to be less. As the variation in motion is less the mean squared error is tend to be less. This phenomenon is also inversely true in case of our peak signal to noise ratio (PSNR). In *PSNR* plot of fig. 4 it can be seen that as the *deltaT* value increases the upper limit of value decreases. This is true for most of the video sequence in general.

Next if we look into the MSE plots in fig. 2 it is observable that for *longJump* video sequence that, there are some minor fluctuation with two peak at around frame 60 and 80 for *deltaT* value 10 and 25. But if we increase the deltaT (in this case 50) by a large margin we can see rapid fluctuation in the whole video sequence (right most plot of fig. 2).

This fluctuation again can be explained by the fact that if we have large gap between reference frame and current frame , there is a high probability that the pixel position has already been shifted a lot from reference frame to current frame for every point of our measurement. Also for deltaT

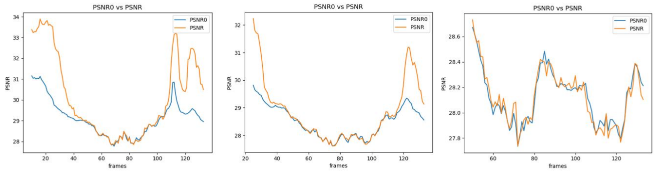

Fig. 4. PSNR at deltaT value 10, 25 and 50 respectively (left to right) for video sequence LongJump .*PSNR0* denotes PSNR before compensation of global motion and*PSNR0* denotes PSNR after compensation of global motion

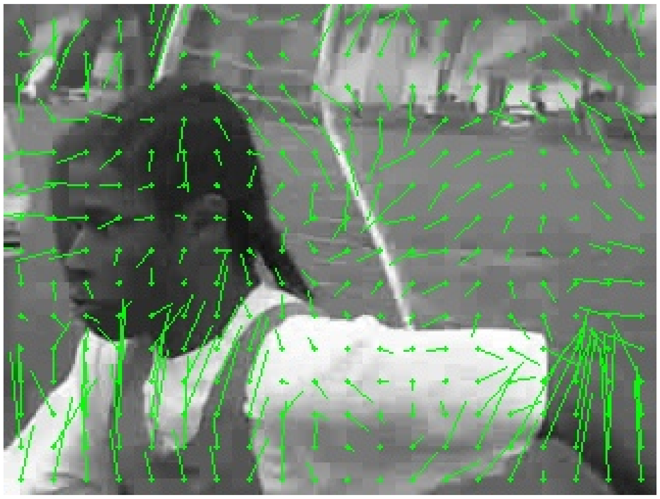

Fig. 5. Frame 66 of video sequence *longJump* taken at deltaT=10. Shows strong optical flow in the frame.

value 10 and 25, the system were able to compensate the the motion and hence the reduced *MSE* except around frame 60 and 80. There is very sudden large motion for which the system fails to estimate the actual motion in 60 to 80 frames.If we look into the actual frame of the video sequence where the error has increased greatly we observe that the camera zooms in into the person covering almost the entire frame (see fig. 5) as a result movement of that person was creating a strong and rapidly changing response in terms of optical flow. Also the camera rotates more than 100 degrees leading a massive change in the background pixels . As a result *MSE* and *Entropy* is higher in this particular area of the curve.

However the *Entropy* plot at fig. 3 has some interesting features. There is a peak point where the entropy of error image exceeds the entropy of current image for deltaT value 10 and 25. On the other hand Entropy is always higher for the error image for deltaT value 50. This suggests that around this peak region the information content of error image is higher than current frame. Error image can usually be higher when there is strong global motion in between the reference frame and current frame (see fig. 6) . If we investigate in the video sequence (at deltaT=10) we can see that, there is very strong and sudden motion around frame 66 and beyond compared to frame 55 and beyond.(see figure 5)

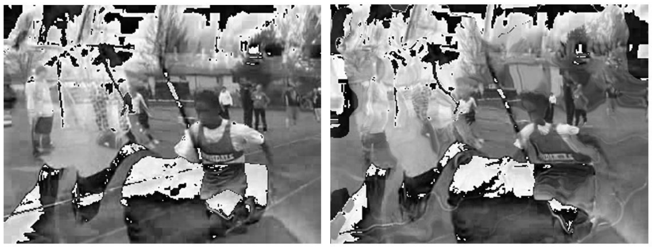

Fig. 6. Uncompensated & compensated error image for frame 66 in video sequence *longJump*

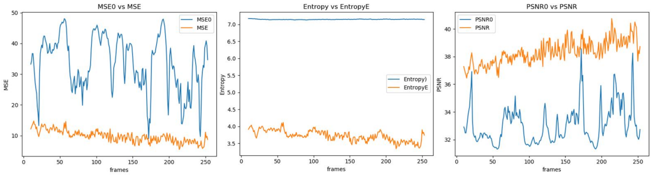

Fig. 7. MSE , Entropy and PSNR plot for *PersonConvergence* video sequence at deltaT=10. *MSE0, PSNR0* denotes error before compensation of global motion and *MSE, PSNR* denotes error after compensation of global motion. *Entropy* denote entropy of current frame and *EntropyE* error image

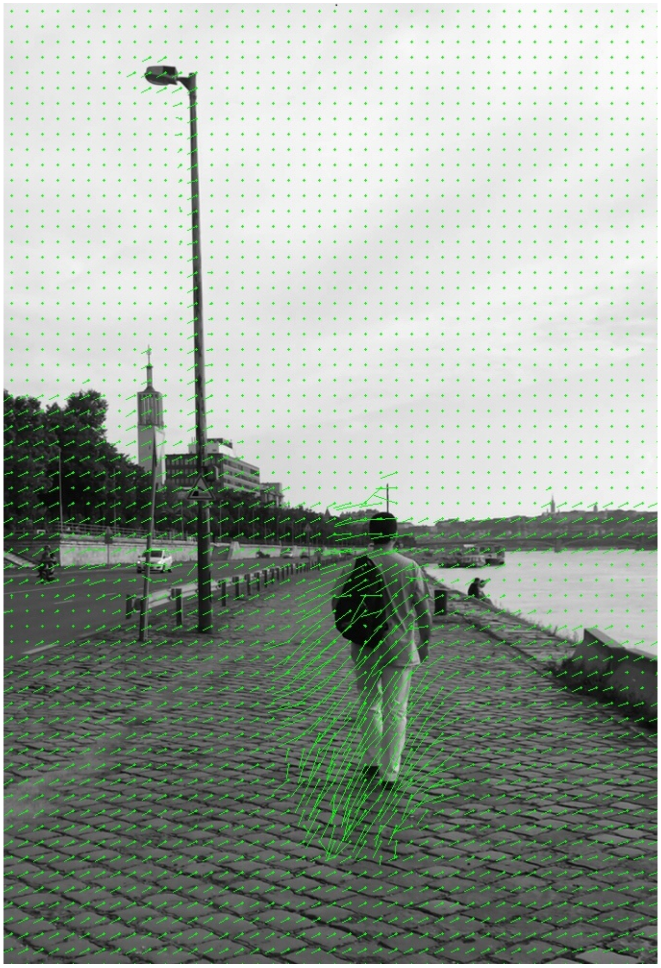

Fig. 8. Frame 58 of video sequence *personConvergence*

Next, in the fig. 7 plot of *MSE* , *Entropy* and *PSNR*

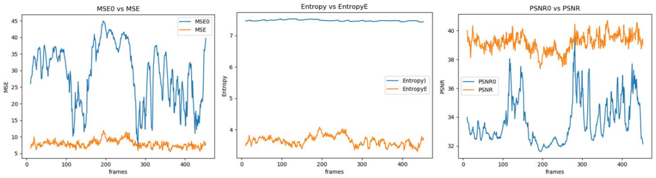

Fig. 9. MSE , Entropy and PSNR plot for *SmallCar* video sequence at deltaT=10. *MSE0, PSNR0* denotes error before compensation of global motion and *MSE, PSNR* denotes error after compensation of global motion. *Entropy* denote entropy of current frame and *EntropyE* denotes entropy of difference of reference frame and current frame

for video sequence *personConvergence* are shown. In *MSE0* curve, it can be observed that the error is fluctuating a lot. which can be explained by large amount of camera movement present in the video sequence. On other hand ,very small error can be observed after global motion compensation, also the curve for after motion compensation looks more stable and consistent.Besides there are two sudden peaks around frame 55 and 185. If we investigate the actual video frame it can be found that there is large motion being created by camera motion and person's motion (see fig.8).

Now let us look at the Entropy plot, it can be observed that Entropy of the current frame is almost flat (blue curve) but lot higher than entropy of error image (orange curve). As entropy is a measure of amount of information, if we take the entropy of current frame entropy should be higher as there is high uncertainty , in other words more information in all the pixels. Also it has not been changed over time. On the other hand , the error image surface is more homogeneous and most of the pixels are empty as they hold only the frame difference and most of the pixels are similar.

For the *PNSR* graph, we can see that as the motion compensation happens the relative noise due to motion gets compensated *PSNR* increase significantly (orange curve)

Next the error metrics for *smallCar* video sequence has been analyzed. In the fig. 9 if we look into the *MSE* plot we can see there are moderate global motion through out the whole video sequence expect from the frame 120 to 300 (blue curve) . In the particular region plot from frame 120 to 300 we can see a large spike in the error. If we investigate the frame we can see that before frame 120 where were no object or cars in the frame all had is camera motion.After 120 frame few cars gets into the frame. Due to the occlusion created by the trees the cars goes and comes out from the occlusion creating sudden motion. During this period we can also notice increased camera motion.These two factor in the frame explains the large error in this region. Now if we look at the motion compensated curve (orange curve) it does not have the large error region. This because the system we able to estimate and compensate the motion reducing the overall mean squared error.The same fact also been reflected in the *PSNR* plot.

For the *Entropy* plot of fig. 9 followed the same trend as explained earlier.*Entropy* for current frame(blue curve) is higher but almost constant on the other hand for error image it is much lower due to motion noise compensation(orange curve).

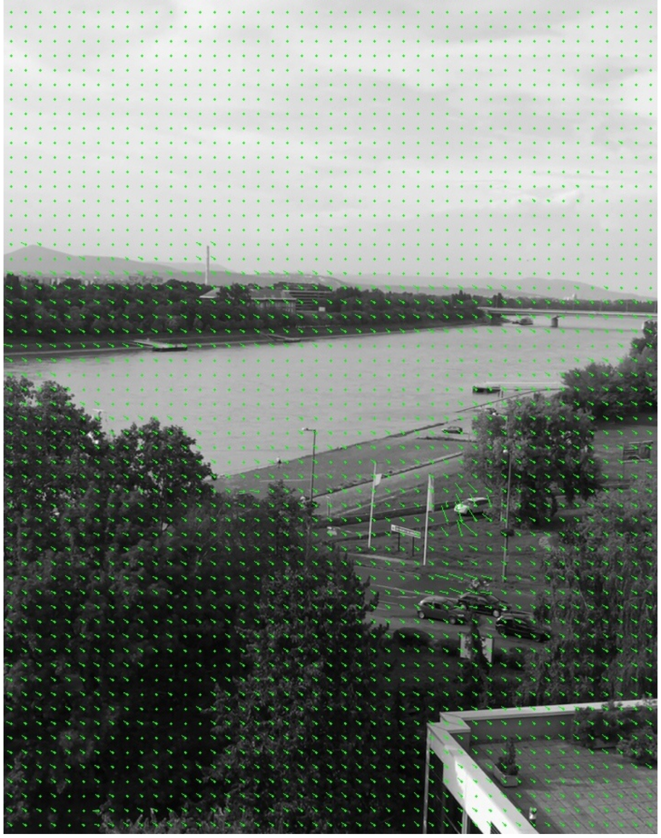

Fig. 10. Frame 245 of video sequence *smallCar*. Car suddenly coming out of occlusion in video sequence

VI. Conclusions

In conclusion , it we can observe that different kind of video sequence and scenario poses different kind of challenges for motion estimation.Precise estimation of motion can be challenging when there is sudden motion of object or camera is present , also occlusion and relative distance between subject and camera needs to be handled. Our controllable parameter *deltaT* has significant effect on our error matrices. This parameter needs to adjusted for each video sequence depending on the motion that is present in the sequence.

References

[1] John X. Liu (2006). Computer Vision and Robotics. Nova Publishers. ISBN 978-1-59454-357-9.

[2] https://en.wikipedia.org/wiki/Opticalflow accessed at 24 September 2020, 11:41 am.

[3] Aires, K. R., Santana, A. M., Medeiros, A. A. (2008, March). Optical flow using color information: preliminary results. In Proceedings of the 2008 ACM symposium on Applied computing (pp. 1607-1611).

[4] Lehmann, E. L.; Casella, George (1998). Theory of Point Estimation (2nd ed.). New York: Springer. ISBN 978-0-387-98502-2. MR 1639875

[5] https://en.wikipedia.org/wiki/Mean_squared_error accessed at 25 September 2020 10:42am
[6] https://en.wikipedia.org/wiki/Peak-signal-to-noise-ratio accessed at 25 September 2020 10:42am
[7] https://en.wikipedia.org/wiki/Entropy_(information_theory)