Deep Learning in Computer Vision Lab 1: Optical Flow estimation & Global motion estimation in the image plane with RANSAC algorithm

26/09/2019

Tara Vanhatalo
M2 Modélisation Mathématique pour le Signal et l'Image
Université de Bordeaux

Overview

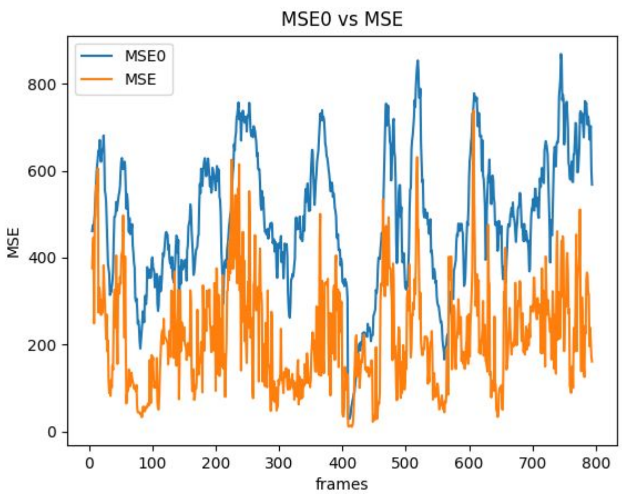
The first goal of this lab is to implement (using the Python programming language) and analyse the results of certain metrics describing information provided by motion in different videos (MSE, PSNR, entropy).

The second goal is to familiarise ourselves with functions from the OpenCV library and compute the global motion estimation and visualize the distribution of the energy of residual motion. To do this, we had to complete a skeleton code given to us.

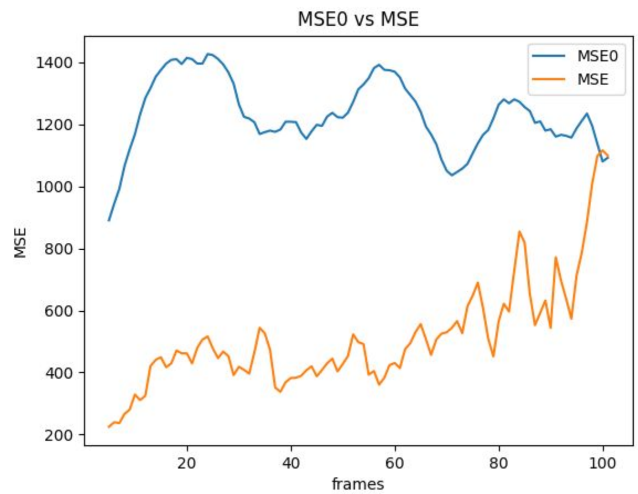
For this first lab, we'll be studying the videos vtest.avi, PersonConvergence_702.MOV and Birds_720.MOV. We observed the results every five frames (i.e. deltaT = 5) on each of these videos.

Metrics Computation

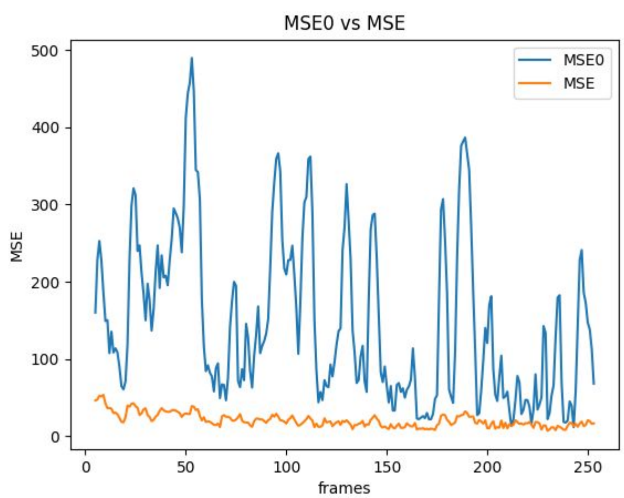
I. MSE (Mean Squared Error)



MSE for vtest.avi



MSE for Birds_702.MOV



MSE for PersonConvergence_702.MOV

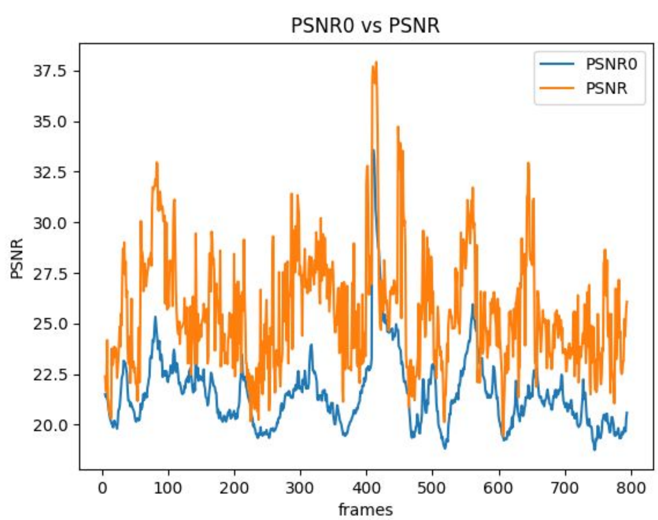
MSE is between the compensated frame and the current frame (the compensated frame being the current frame with the pixels from the previous frame projected onto it). MSE0 is between the previous frame and the current frame.

The MSE0 is greater than the MSE for both of these videos which is to be expected because we just have information from the previous frame in the MSE0 whereas the MSE uses the current frame and the current frame with a projection of the pixels from the previous frame so it makes sense that these two frames would be more similar and have less of an error between them.

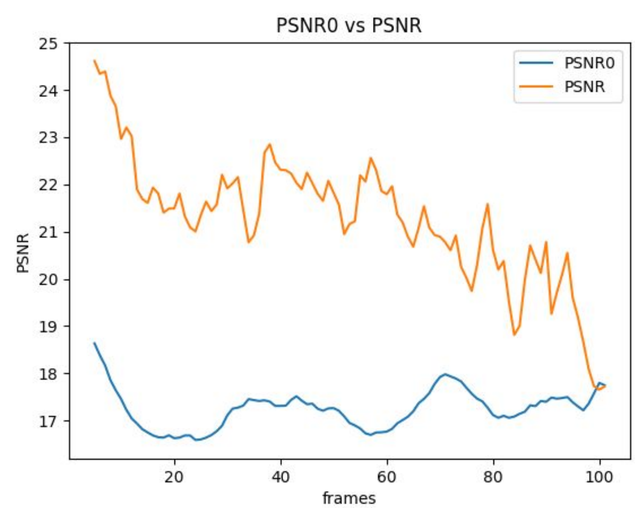
The MSE for Birds_702.MOV is greater than the MSE for vtest.avi and PersonConvergence_702.MOV because there is more general movement (the camera pans) whereas in PersonConvergence_702.MOV and vtest.avi there is only a bit of movement in the middle of the frame (and the top left corner for the vtest video), otherwise the camera is stable and there is generally little movement.

There's one problem with using the MSE, this is that it depends strongly on image intensity scaling, e.g. the MSE is smaller on a 10-bit image than on an 8-bit image, so it's difficult to compare between two different images, we use the PSNR to remedy this.

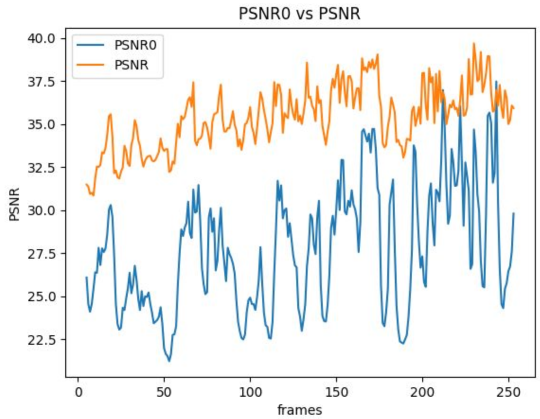
II. PSNR(Peak Signal to Noise Ratio)



PSNR for vtest.avi



PSNR for Brids_702.MOV



PSNR for PersonConvergence_702.MOV

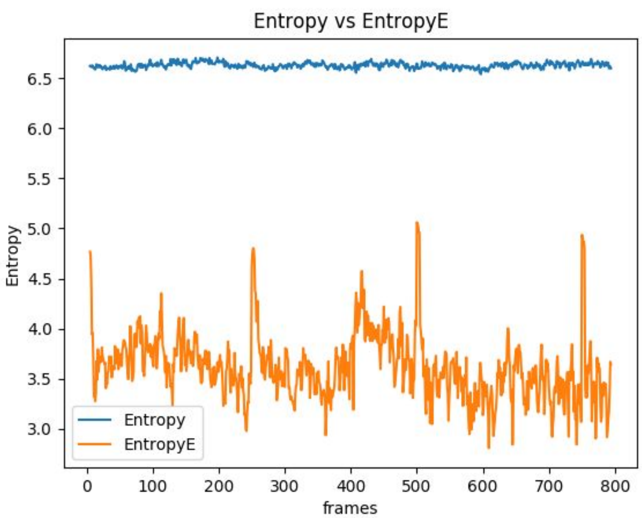
PSNR is calculated with MSE (between compensated frame and current frame) whereas PSNR0 is calculated with MSE0 (between previous frame and current frame) so that the PSNR can scale the MSE according to the range of each of our images.

We can see that on all of these graphs our PSNR0 is smaller than our PSNR, this is to be expected considering we have to divide by MSE (MSE0 for PSNR0) in our calculation for the PSNR. We can also see that our values for both videos are closer together than those for our MSE, so the scaling we obtained from the using the PSNR instead worked properly.

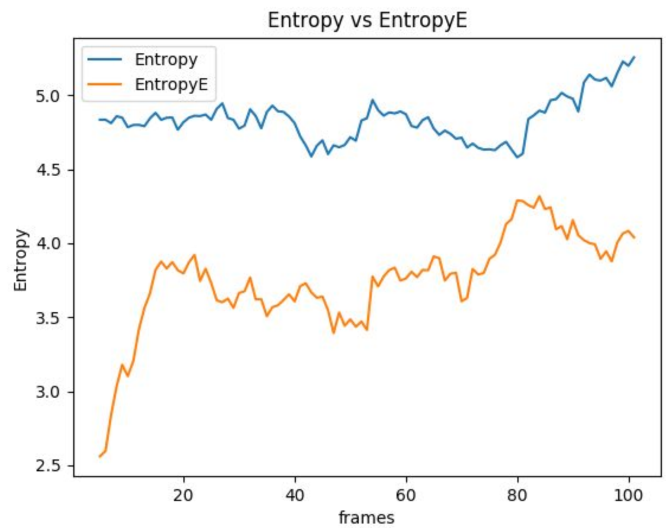
The results read on these two graphs reaffirm what we concluded from the MSE/MSE0 graphs in the previous section. This is that vtest.avi has more details than Birds_702.MOV and Birds_702.MOV has more homogenous regions in each frame than vtest.avi and also there is the movement of the camera that adds noise in the Birds_702.MOV video.

In both vtest.avi and PersonConvergence_702.MOV the camera remains stable, however there is still a difference in their respective PSNRs. This is because there is more detail in PersonConvergence_702.MOV.

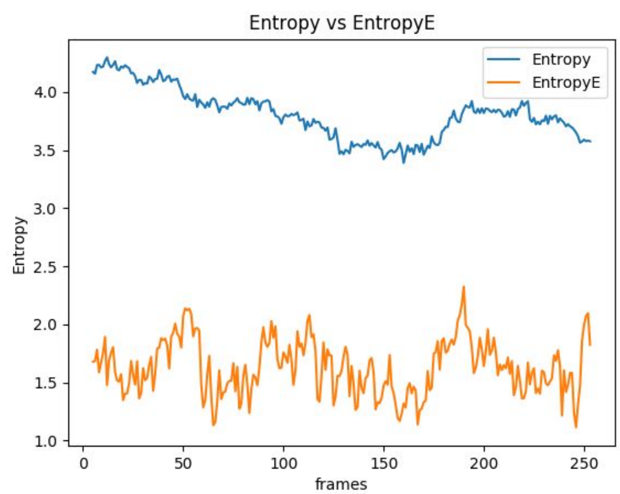
III. Entropy



Entropy for vtest.avi



Entropy for Birds_702.MOV



Entropy for PersonConvergence_702.MOV

We can see that Entropy (i.e. the entropy of our base video in grayscale) remains relatively stable around a value of just above 6.5. This is a relatively high entropy value so that means our images contain quite a lot of information, that is that this video is less predictable (less homogenous) than our other video Birds_702.MOV that has a smaller entropy value.

The EntropyE (i.e. the entropy of our error image between the compensated frame and the current frame) is a lot smaller than our other entropy. This makes a lot of sense when we observe the displayable error image (in our next section). Seeing as the error between these two images is null for vtest.avi in the regions where there is no movement, the error image will be very homogenous in these regions and therefore the entropy will be smaller.

In Birds_702.MOV however, seeing as the camera pans, there will always be some kind of error between two frames (even if it is small), so EntropyE will be greater.

The entropies we have for PersonConvergence_702.MOV are the smallest. This means that the placement of the pixels for upcoming frames will be the most predictable. This makes sense because a lot of sense because the only movement we have in this video is small, localised and in one direction.

IV. Displayable Error Image

Displayable Error Image for vtest.avi

When running the code, a video of the error image should start to play. In this video we can clearly see the differences in movement: The plain gray areas are where the movement between frames is minimal or null, the white areas are where we have movement in one direction and the black areas are where the movement is in the opposite direction.

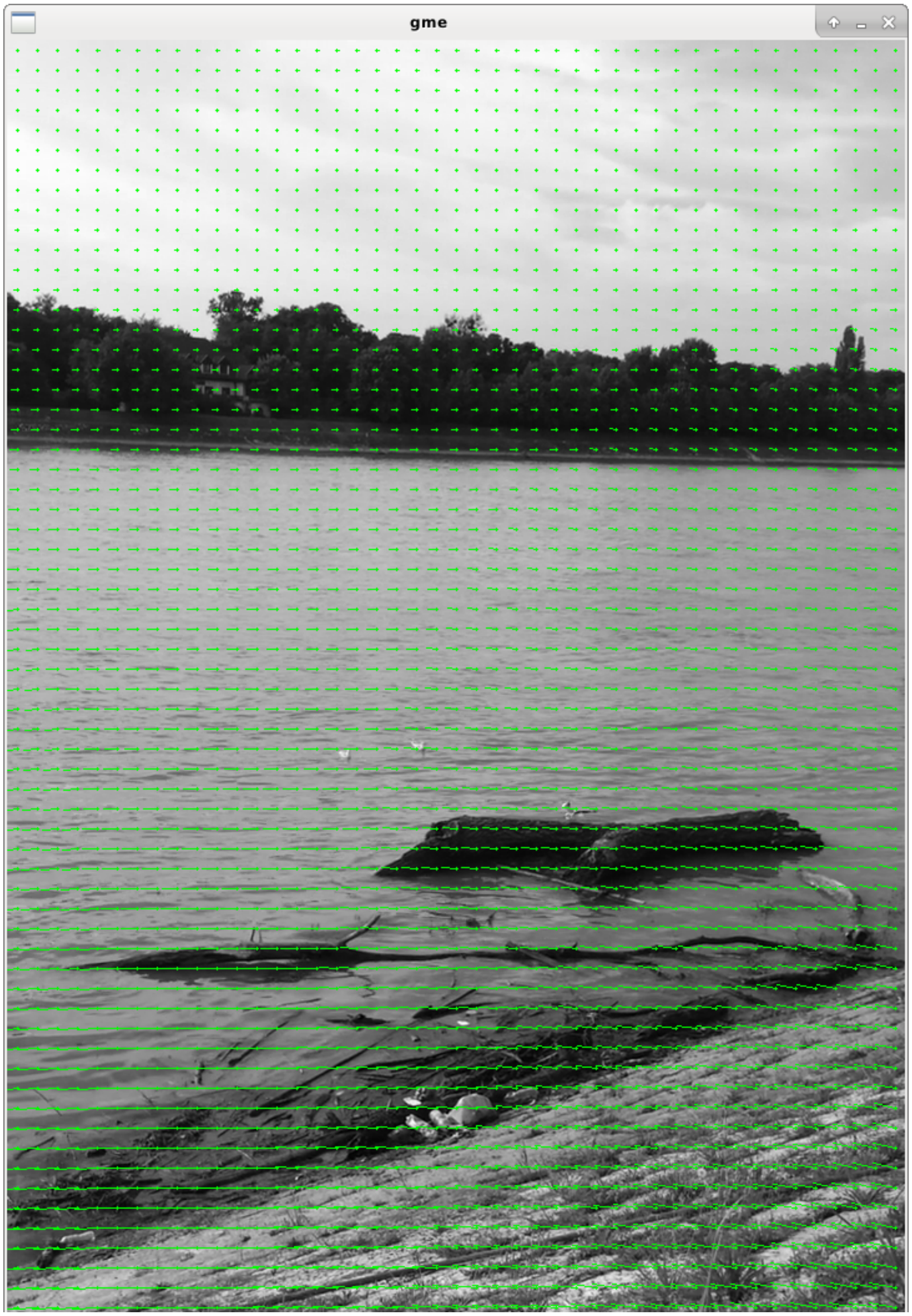
Displayable Error Image for Birds_702.MOV

Displayable Error Image for PersonConvergence_702.MOV

Global Motion Estimation (GME)

GME for the vtest.avi video

First of all, we can observe a big difference in the GME of Birds_702.MOV compared to those of PersonConvergence_702.MOV and vtest.avi (the GME image for the second and third video can be seen below). The difference being that the GME of the vtest.avi and PersonConvergence.MOV video, we can see no vectors, only points, whereas on the GME of the Birds_720.MOV we can see vectors that are all more or less in the same direction. This is because in the first two videos mentioned we have localised movement but the camera stays in the same place so the global motion is practically null. In the Birds_702.MOV video however, the camera is moving so we have quite strong global motion in the opposite direction of the camera movement.



GME for the Birds_702.MOV video

GME for the PersonConvergence_702.MOV