K Pranath Reddy

2016B5A30572H

PHY F213

Object detection and tracking using adaptive correlation filters

November 25, 2017

ABSTRACT

This project is based on the research paper “Visual object tracking using adaptive correlation filters” by David S. Bolme, J. Ross Beveridge, Bruce A. Draper and Yui Man Lui of Colorado State University. The main objective of this project is to use correlation filter based tracking to detect and track objects in a video. This project is done using OpenCV and the main problems addressed are related to the drawbacks of traditional object detection methods.

INTRODUCTION :

Traditional object detection methods include correlating the template with a target frame and checking for a strong 2D Gaussian shaped peak centred on the target in the output image. This method works in most cases but it fails when there is a change in lighting, scale, out of plane rotation and occlusion. The method used in this project eliminates these problems by updating the filter in realtime, it does so by finding a filter that minimises the sum of squared error between the actual output of the correlation and the desired output of the correlation. The object is detected in one frame of the video and is then tracked in the subsequent frames, this process is much more efficient than detecting the object in every frame of the video. Tracking starts from the first frame itself after assigning the tracking window without any need for a database or training. The performance is measured by finding the peak-to-sidelobe ratio (PSR) of the output which also helps in detecting occlusion.

MATHEMATICAL EQUATIONS :

Let f be the input image and h be the filter , we now take Fourier transform of the input image and the filter.

$$F = \mathscr{F}(f)$$

$$H = \mathscr{F}(h)$$

Let g be the output , the correlation of input image and filter gives us the desired output.

$$\mathscr{F}(g) = G = F \odot H^* \tag{1}$$

Let the set of training images be f_i and the set of training outputs be g_i ,their respective Fourier transforms become F_i and G_i .

$$H_i^* = \frac{G_i}{F_i} \tag{2}$$

We now find a filter H that minimises the sum of squared error between the actual output of the correlation and the desired output of the correlation. Hence this filter is also known as MOSSE filter (minimum output sum of squared error).

$$\min_{H^*}\left(\sum_i |F_i \times H^* - G_i|^2 \right) \tag{3}$$

$$\frac{\partial}{\partial H^*}\left(\sum_i |F_i \times H^* - G_i|^2 \right) = 0 \tag{4}$$

$$\frac{\partial}{\partial H^*}\left(\sum_i \left(F_i \times H^* - G_i\right)\left(F_i \times H^* - G_i\right)^* \right) = 0 \tag{5}$$

$$\frac{\partial}{\partial H^*}\left(\sum_i (F_i F_i^* H H^* - \ F_i G_i^* H^* - F_i^* G_i H + G_i G_i^*) \right) = 0 \tag{6}$$

$$\sum_i (F_i F_i^* H - F_i G_i^*) = 0 \tag{7}$$

$$H = \frac{\sum_i F_i G_i^*}{\sum_i F_i F_i^*} \equiv \frac{\sum_i F_i G_i^*}{\sum_i I_i} \tag{8}$$

$$H^* = \frac{\sum_i F_i^* G_i}{\sum_i F_i F_i^*} \tag{9}$$

The numerator in Equation 9 is the correlation between the input and the desired output and the denominator is the intensity of the input.

We now introduce a new factor called Learning rate (η) , learning rate puts more weight on recent frames and lets the effect of previous frames decay exponentially over time. It helps the filter adapt to the changes of the object.

$$H_i^* = \frac{A_i}{B_i} \tag{10}$$

$$A_i = \eta F_i^* G_i + (1-\eta) A_{i-1} \tag{11}$$

$$B_i = \eta F_i^* F_i + (1-\eta) B_{i-1} \tag{12}$$

The Peak-to-sidelobe ratio (PSR) is used as a performance measure and we can detect occlusion by setting a threshold value which helps in freezing the tracker when the object is covered and resume tracking when the object reappears.

$$PSR = \frac{g_{max} - \mu}{\sigma} \tag{13}$$

Where , g_{max} is the peak value and μ and σ are the mean and standard deviation of the sidelobe.

PROCEDURE :

The program for MOSSE filter is written in python and OpenCV and the template matching is done in MATLAB. The threshold value of PSR is set at 8.0 and the learning rate is set at 0.125 (The program files are in the google drive link provided at the end of this report)

RESULTS :

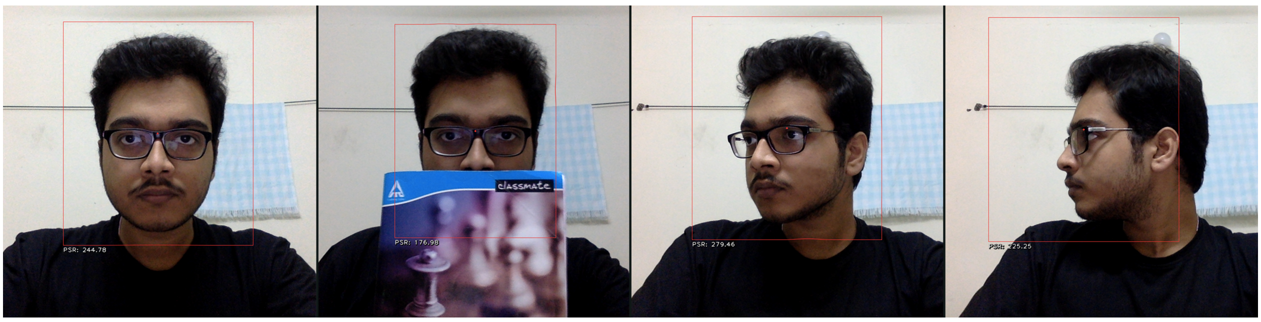
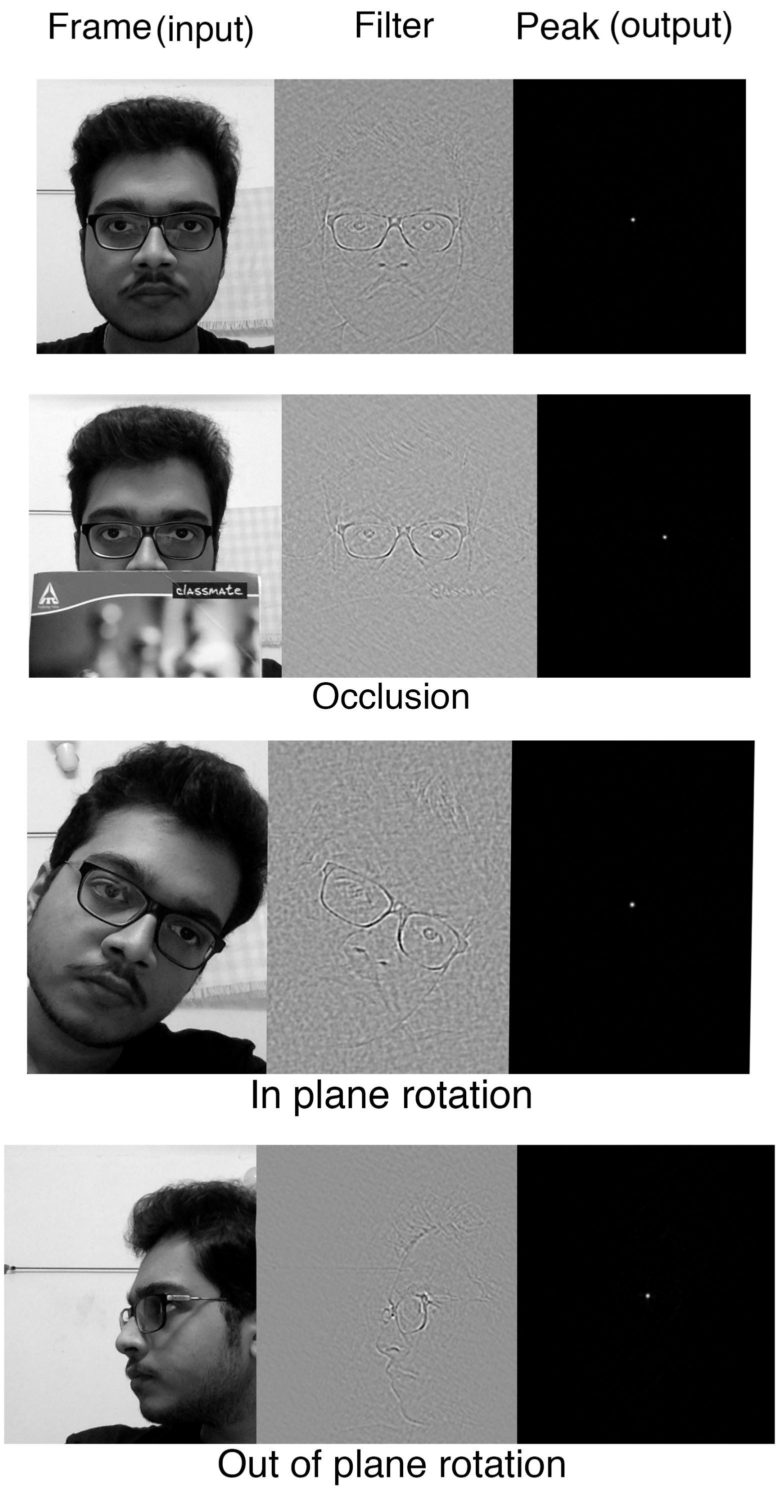
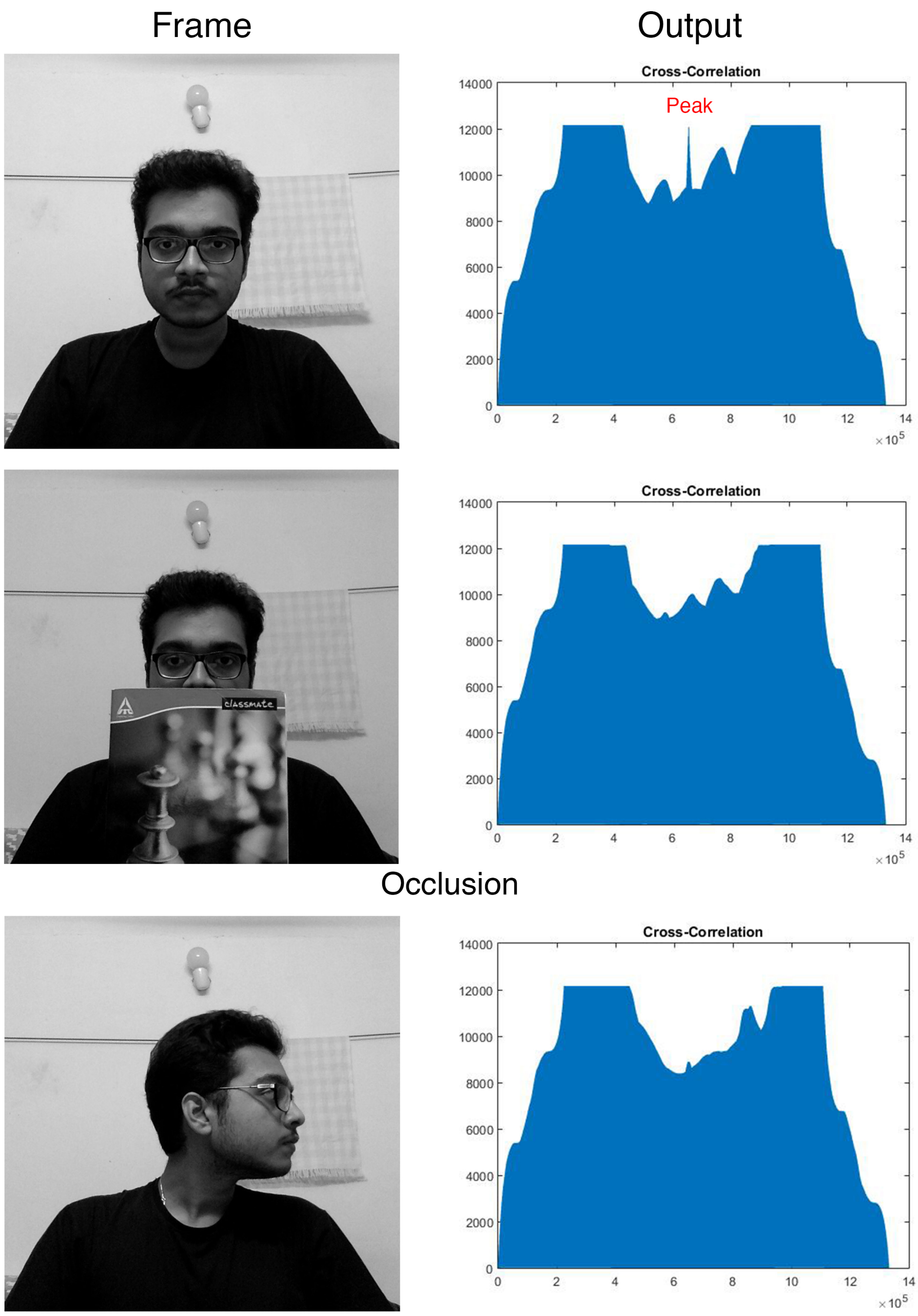


Figure 1 : This figure shows the results of the MOSSE filter in a video sequence involving occlusion and off-plane rotation. The PSR value is shown in the lower left corner of the tracking window.



MOSSE FILTER

Figure 2 : This figure shows the input, filters and correlation output for a video sequence involving occlusion, in-plane rotation and off-plane rotation.



Out of plane rotation

Traditional Template Matching

Figure 3 : This figure shows the results of a traditional template matching technique used for object detection. The images on the left are the frames (input) and the images on the right are the outputs of cross correlation between the frame and the template.

CONCLUSION :

The results obtained in this project clearly shows us the drawbacks of traditional template matching methods, In Figure 3 we can observe that the cross correlation gives us a sharp peak in the case where the object is clearly visible without any constraints but fails to produce a peak in case of occlusion and off-plane rotation. Figures 1 and 2 shows that MOSSE filter is able to track the object perfectly and produce sharp peaks even in cases of occlusion and off-plane rotation. MOSSE filter is a robust and highly efficient method for visual object tracking and can handle wide range of scenarios like in-plane rotation, off-plane rotation, occlusion, illumination and can also be used to track multiple objects at once.

REFERENCES :

David S. Bolme, J. Ross Beveridge, Bruce A. Drapper, and Yui Man Lui. "*Visual Object Tracking using Adaptive Correlation Filters*". CVPR 2010.