Xiaowei Yuan

ECE 332

15 November 2022

MP 7 Report

INTRODUCTION

In this MP, I use the CC/NCC/SSD method to achieve visual target tracking. There are 500 frames of a girl's face in a video as a sample test. And there are two rectangle boundaries in each image. One is the range of detection while the other is the cropped area of our tracked target. The new frames are combined into new videos again after training.

CONTENTS

Window extraction:

Firstly, we need to manually extract the girl's face area, the cropped image will be saved. A mouse down and mouse up subprogram will be needed to do the image extraction. The extract area will be saved later for building detected areas, they will have the same size.

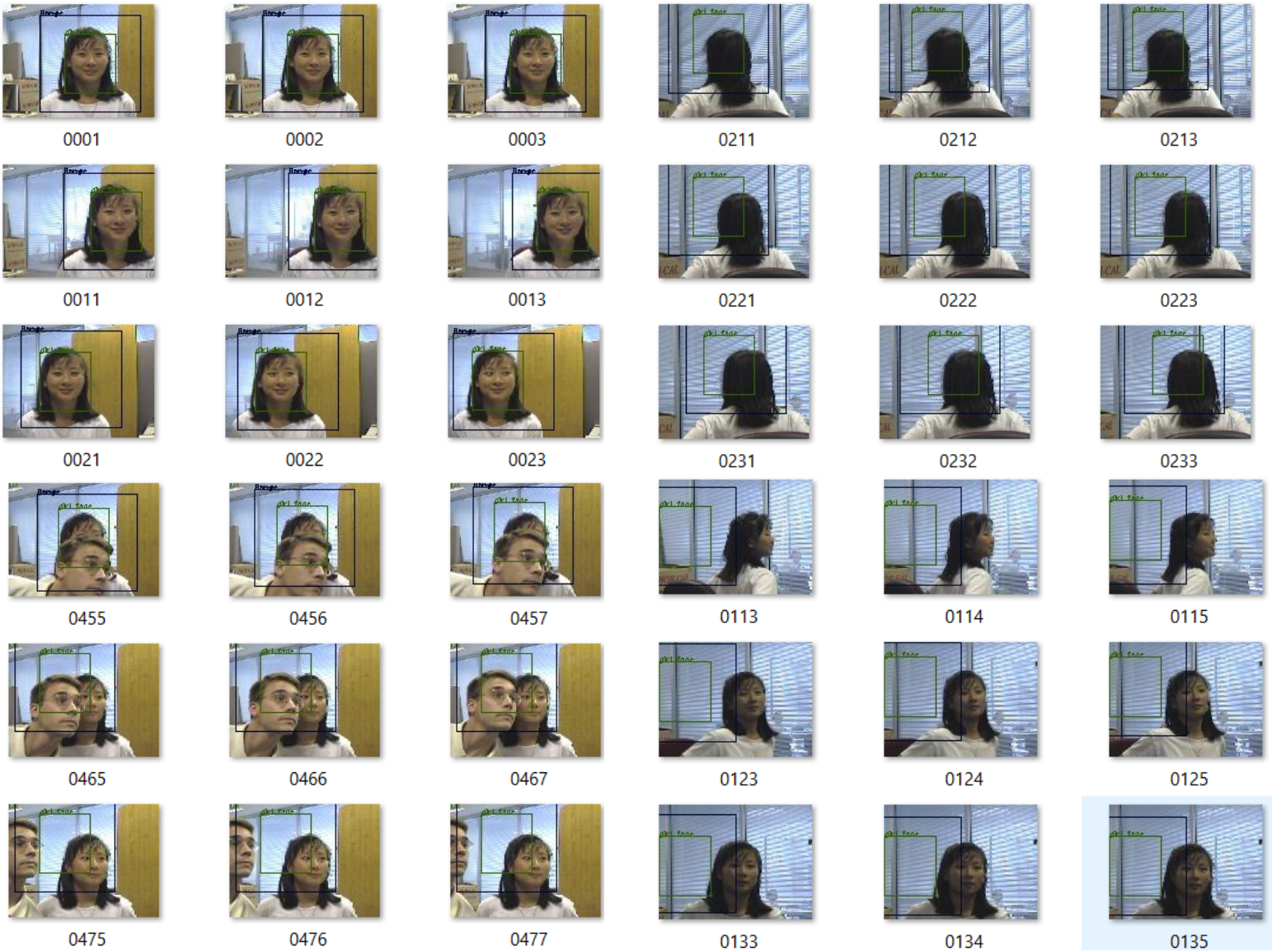
Two rectangular boundaries and Image matching:

Below are the steps I did for each frame.

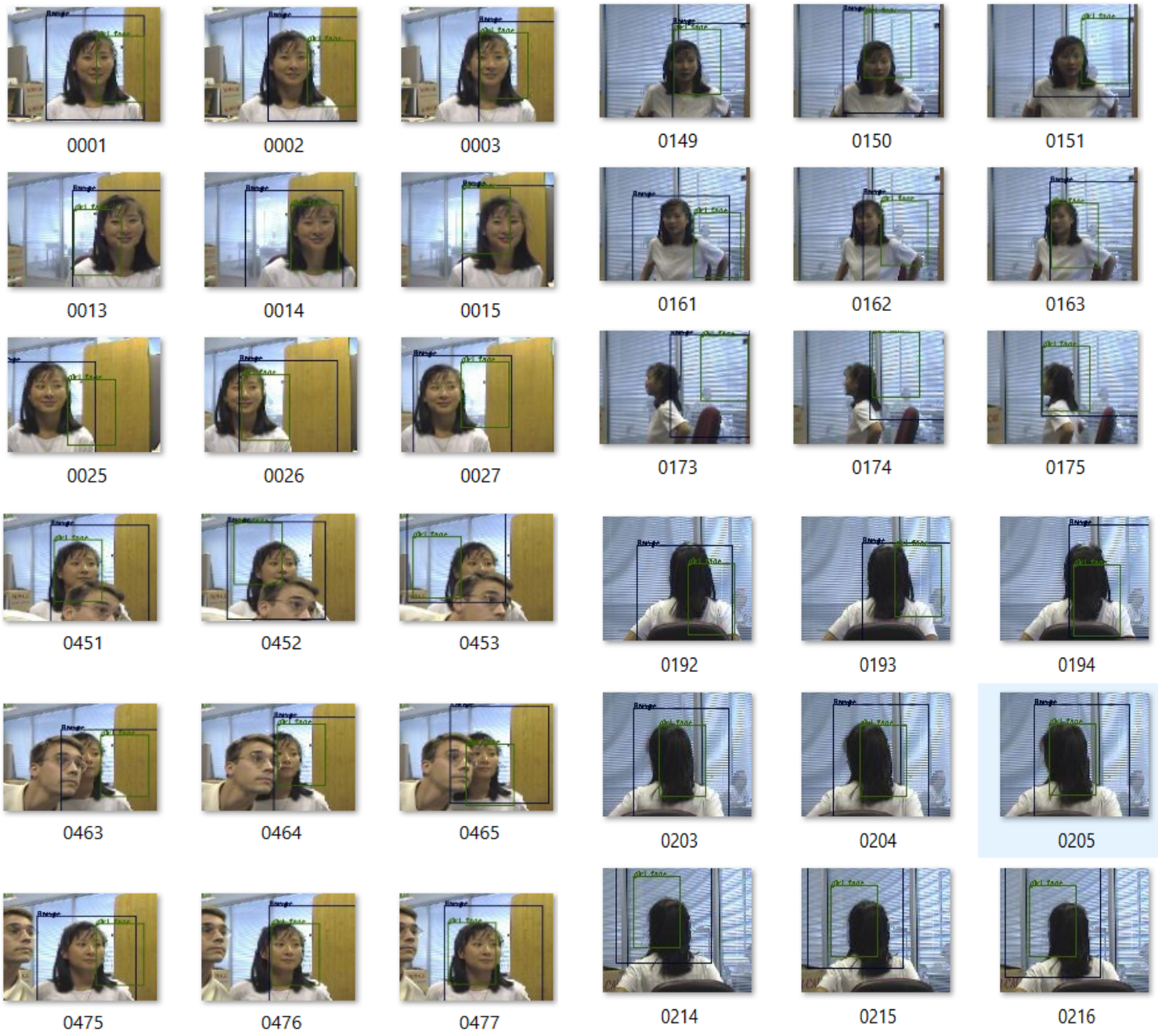
The first boundaries will be built by moving a window either up/down/left/right according to the points location. This boundary is set to black with a name of 'Range'.Then the searching will be inside this area to minimize the searching workload.

We go over each pixel of the area inside the boundaries and do image matching. There are 3 methods to do so.

1. SSD: sum of squared difference. $D = \sum_{u,v} [I(u,v) - T(u,v)]^2$ Here are some result frames of the SSD.



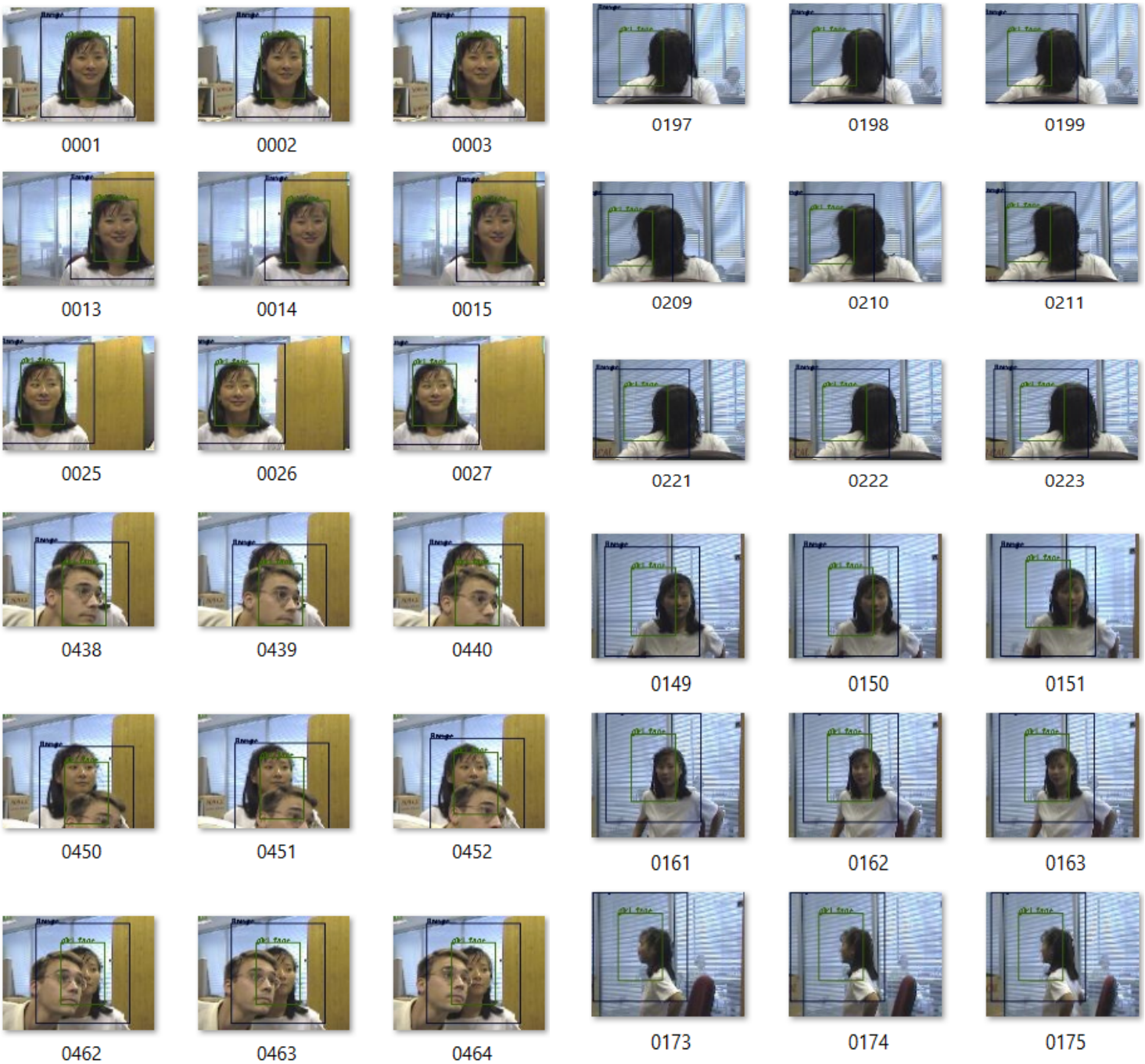
2. CC: cross-correlation. $C = \sum_{u,v} I(u,v)T(u,v)$ Here are some result frames of the CC.



3. NCC: normalized cross-correlation. This is more complicated than CC.

$$N = \frac{\sum_{u,v} \hat{I}(u,v)\hat{T}(u,v)}{\sqrt{\left[\sum_{u,v} \hat{I}(u,v)^2\right]\left[\sum_{u,v} \hat{T}(u,v)^2\right]}}$$

And Here are some result frames of the NCC.



After the searching, the green boundary will appear with the detection of the girl's face.

Frames to Video:

There are 500 images of frames after training, and we use a function to translate them to a video. Here I use FPS = 10 and the final video is 50 seconds long each.

CC

NCC

SSD

ANALYSIS

The three methods of target searching have unique drawbacks and merits.

For CC, the position of the green boundary (target face) changes a lot, but it detects very well when the girl shows her back to us.

For NCC, mostly the green boundary moves smoothly, but it does not detect her back and right face.

For SSD, the green boundary moves smoothly and it also cannot detect the back.

One interesting thing is when the man comes into the camera, all three methods do well in detecting the girl's face. But when the man’s face covers the girl's face, NCC and SSD choose a boundary which is in the middle of the two faces. However, CC will choose one of the man’s face or girl’s face instead of in between them.