Homework Assignment 3
CS696, Applied Computer Vision

Template Matching with Fixed Scale

Template matching is a method of finding the location of a template image in a same or different larger image.
In the template matching with fixed scale, the template image is not scaled and only the original template size is searched throughout the original image using the sliding window method.
A few of the several techniques with which the template matching to find the exact location of the template image in the larger image can be implemented are:

- Correlation
- Zero-mean correlation
- Sum Square Difference (SSD)
- Normalized Cross Correlation (NCC)

In the homework, I have created 2 functions (zmc_func.m and ssd_func.m) for Sum Square Differences and Zero-Mean Correlation respectively and used the inbuilt function for the Normalized Cross Correlation.

Algorithms:

Zero-Mean Correlation and Sum Square Differences

- First we get the template and the original image and then convert them into the gray scale.
- We find the zero mean correlation using the function defined by me and then choose the maximum value out of all the ZMC/SSD values.
- We then get the index coordinates of the maximum values using the ind2sub function.
- Finally we get the zmc/ssd error using the Euclidean distance between the estimated location and the original (ground-truth) locations.

Normalized Cross Correlation

- We use the normxcorr2 function to compute the normalized cross-correlation of the grayscale matrices of template and the original image.
- This function outputs a matrix containing the correlation coefficients of the 2 images.
- We calculate the normalization correlation of all the elements in the matrix, starting from the middle of template matrix, and ending with the sum of the middle of template matrix and the image size similar to the way of the sliding window.

- Finally we get the NCC error using the Euclidean distance between the estimated location and the original (ground-truth) locations.

zmc_func.m

- As per the equation provided below, where f is the original image and g is the filter provided, we take the summation of product of all the coordinates from the starting image until the template image coordinates end.
- We update the filter image by subtracting the mean value of the filter image from the original filter image.
- To do the following function, we need to pad the original image with the half the size of the filter image in each size.
- We then apply the following equation manually in matlab and return the output.

$$h[m,n] = \sum_{\{k,l\}} f(m+k, n+l) - \bar{f})(g[k,l])$$

ncc_func.m

- Normalized Cross Correlation is a little bit different than Zero-Mean Correlation.
- As per the equation provided below, where f is the original image and g is the filter provided, we take the summation of square of all the coordinates of the original image subtracted from all the coordinates of the template image until the template image coordinates end.
- We update the filter image by subtracting the mean value of the filter image from the original filter image.
- To do the following function, we need to pad the original image with the half the size of the filter image in each size.

$$h[m,n] = \sum_{k,l} (g[k,l] - f[m+k, n+l])^2$$

Template Matching with Varied Scale

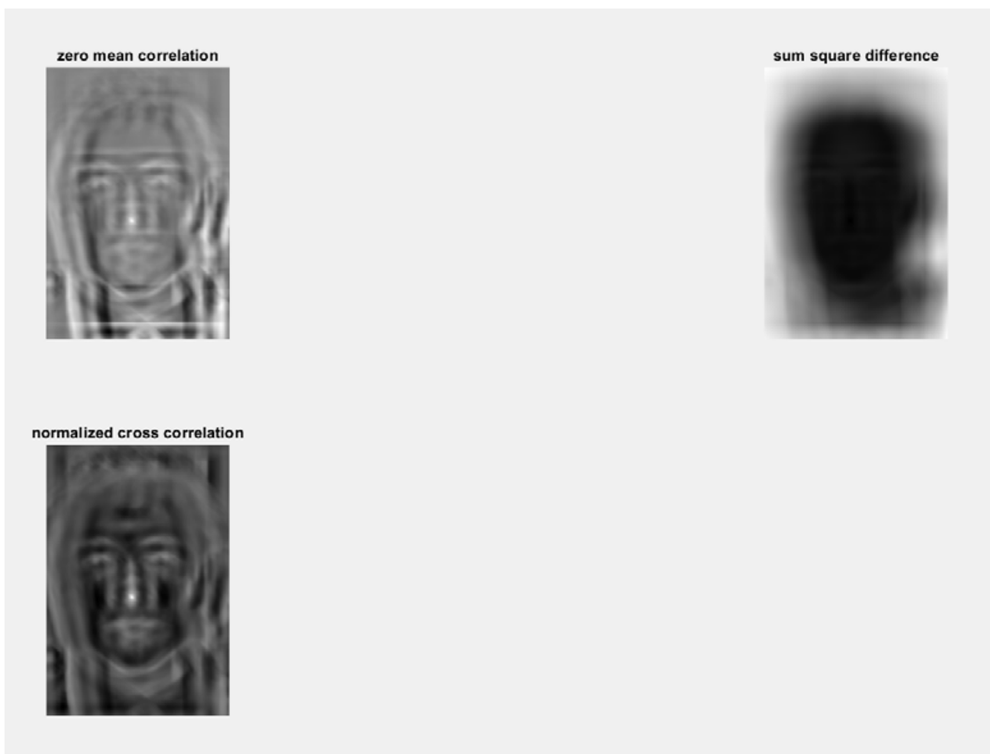
In this type of template matching, the template image is scaled and is either larger or smaller than the original template image size. This scaled template image is searched throughout the original image using the sliding window method.

- We use 3 different scales i.e. 2, 1, and 0.5
- Starting with converting both the images to grayscale, we resize the original image with the respective scale in a for loop
- Within the for loop, we perform all the operations as performed in the above steps for the template matching with fixed scale
- Similarly we compute the localization errors for all the three methods (ZMC, NCC and SSD) and do it 3 times for 3 different sizes respectively.

Output:

For the following image, the errors predicted are:

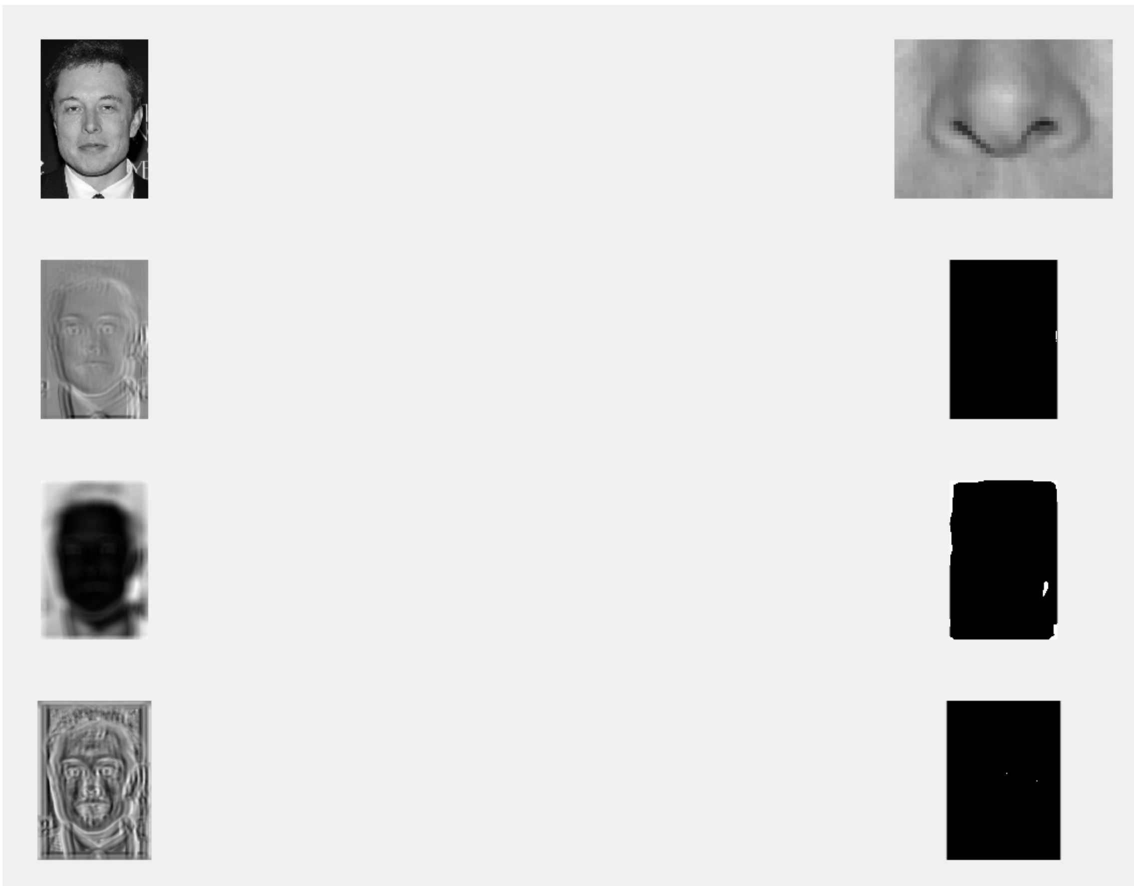
Error in ZMC is: 60.000000
Error in SSD is: 309.195731
Error in NCC is: 59.405387



For the respective Scaling Factor, the Errors predicted are:

Scaling Factor is: 2.000000
Error in ZMC is: 927.414686
Error in SSD is: 927.414686
Error in NCC is: 607.664381

Scaling Factor is: 1.000000
Error in ZMC is: 247.729691
Error in SSD is: 247.729691
Error in NCC is: 96.254870



Scaling Factor is: 0.500000
Error in ZMC is: 81.841310
Error in SSD is: 81.841310
Error in NCC is: 59.405387

It can be noted from the output that only the NCC (normalized cross correlation) was able to detect the template images from the original images scaled to 1 and 0.5.
Thus the NCC gave the least localization error compared to the zero-mean correlation and sum squared differences.
The error was too high for the other methods for the model to be able to detect the template images from the original images.

According to me, the possible issues for the zero-mean correlation was that the output was sensitive to the gain/contrast and the pixels in the filter image, which are near the mean have very little effect on the output values.

According to me, the possible issues for the Sum Square Differences was that the output might be too sensitive to the average intensity of the image which might cause fluctuations and errors in the output values.

Along with it, I have even tried to identify the template image from a separate larger image, for e.g. a person's nose, with other person's face.
I was able to get the following readings:

Scaling Factor is: 2.000000
Error in ZMC is: 567.381706
Error in SSD is: 567.381706
Error in NCC is: 369.476657

Scaling Factor is: 1.000000
Error in ZMC is: 221.720545
Error in SSD is: 221.720545
Error in NCC is: 193.662077

Scaling Factor is: 0.500000
Error in ZMC is: 176.527618
Error in SSD is: 176.527618
Error in NCC is: 136.132289

I can notice it here as well, that the least error is found in NCC with the scaling factor being 0.5
The rest of the images having NCC with the least errors with the scaling factor of 1 being better than the scaling factor of 2.

The nose can be visible in the figure 2 of the output images (fig 1, fig 2, fig 3 – respectively).

I have even calculated the histogram of the localization errors of the output images.