Question 3 :

If we try to test our program on the images (s33-s40) which are were not a part of our system, we will end up wrongly identifying these new images. This is because in our current program we only look at point nearest to the current image for identifying it.

To prevent this wrong identification of images, we have to explicitly added threshold values on the confidence so as to mark images as identified or not identified.
On running the program for (s1-s32) we find -
Mean of confidence values which are identified :90.5163%

On running the program for (s33-s40) we find -
Mean of confidence values which are not identified :74.3256%

We define confidence as -

$$\text{Confidence} = \min \left\{ \frac{(\text{alpha}_{p,}\text{o alpha}_i)}{(\| \text{alpha}_p \| \text{ o } \| \text{alpha}_i \|)} \right\}$$

Now, we can clearly see that confidence values are high for images which can be identified and low for those which cannot be identified. Hence we set a threshold parameter (around 82-83%). If the confidence value is above this, we have correctly identified image. Else we have wrongly identified the image as the persons image might not have been initially present in our system.

Now using the above algorithm to identify false negatives/positives -
For s1-s32, if confidence value is less than threshold then it's a false negative. This is because we are marking this image as not-identified even though this image is present in our current system.
Total : 128
False Negative : 14

For s33-s40, if the confidence value is more than the threshold it's false positive. This is because we are marking the image as identified even though this image was not present in our current system.
Total : 32
False Positives : 13