Facial Liveness Testing For The Web

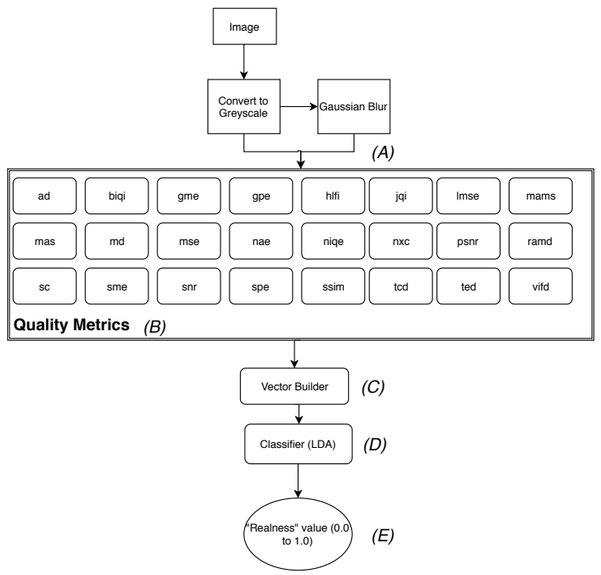
Ryan Collins

Department of Computer Science, Durham University

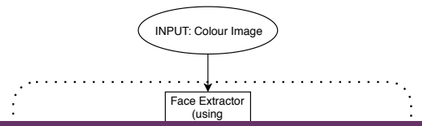
Introduction

The goal of this project was to investigate facial liveness tests,
and propose a small set of liveness tests to include in a facial liveness system. This project aimed to investigate existing literature and produce liveness tests that are more designed for existing real-life applications, therefore not requiring any extra hardware (aside from what's available within a modern day smartphone or computer). Furthermore, the liveness tests must also be fairly fast to compute, since a real world web system would be limited by users' attention spans. It's important to explain the difference between recognition and liveness: recognition is the process of identifying whether someone is who they say they are, but liveness is about detecting potential acts of spoofing within a given input (i.e. is the person alive, or are they simply a screen/piece of paper). The first, and most common type of attack, is a 2D presentation attack. This occurs where an image, or video, is displayed in front of the camera (using either a screen or a piece of paper). To combat these types of attacks, two different methods were developed: the first being an image quality based method called WIQA, and another being a Residual Network based classifier to classify realness based on facial structure. The other type of attack is a 3D based mask attack. This occurs where a mask is created and presented as input. Masks might have eye cutouts, and mouth cutouts. A new 3D-based liveness test was proposed based on a two part approach: (i) VRN based 3D reconstruction (ii) VoxNet based 3D classification. However, results with this method weren't within desired parameters, and it therefore wasn't suitable for a production grade system.

Image Quality Assessment



Residual Networks for Facial Liveness



Evaluation

Quality Test

The Quality Test performed well, and is
suitable for a web based liveness service due to the high 87% accuracy achieved. True positives and true negatives (that is, correct predictions for both real and fake) were high which means that the classifier is classifying correctly. However, the high false positive rate shows a slight cause for concern, since the model in 12.5% of cases classified a fake image as real, which isn't ideal for our security focused solution. This could potentially be solved by adjusting output variable of the classifier (using a figure representing 'fakeness' rather than realness). This would hopefully lead to more false negatives, which are inconvenient but more secure for the system.
In terms of computational
performance, a 1.40 second time for classifying a single image is within the limits of the expected 2 seconds, and could be improved further with parallel based methods, and not relying on libsvm for the BIQI metric.

2D CNN

The 2D CNN Test performed adequately. While the accuracy was lower than expected (at 71%), the classifier itself still performed better than random. Furthermore, the model was very good at classifying true negatives, with a total percentage of 71.5% being true negatives. While the model itself had a higher than expected false negative percentage, this is just inconvenient for the user rather than a security problem. The model showed it could classify true positives, but this figure was fairly low. This could be improved by improving the training process: ensuring each input image has a correctly identified face (as some images without a detectable face would have been left to classify the entire image, just resized to the input size). The less noise in the input data, the better the potential results in the future. Furthermore, using a larger ResNet model, such as ResNet-101 could lead to better results.
In terms of performance, classifying
an image is very quick, but the time taken to load the model is what took the most time (due to memory needing to be allocated and written to). In production, providing a model was preloaded and ready to accept input, the computation time of this metric would be very fast, and therefore be ideal for inclusion in a liveness web service.

3D VoxNet Liveness Test

As seen from the results, this method of liveness test isn't feasible. With extra training, while the accuracy of the model could be improved, the real time memory requirements don't seem feasible. Based on the VoxNet paper, the accuracy achieved with a 32x32 VoxNet classifier on the SUOD dataset (a 3D dataset of objects) is 69%. While this accuracy could be justified with reasonable computational and memory characteristics, this further proves that this method isn't feasible in the current state. In the future, a new approach of detecting 3D attacks is necessary.

Conclusion

This project showed that creating a facial liveness service for the web is a feasible
idea, and performs fairly well for 2D attacks, with adequate accuracy and computational requirements. The image quality liveness test is accurate and fast, while the ResNet based method is a feasible idea and performs adequately, and could be improved further to improve the accuracy. For 3D attacks however, the proposed VoxNet based model performed badly and is not recommended for inclusion in a liveness test web service.