MASKS AND FILTERS

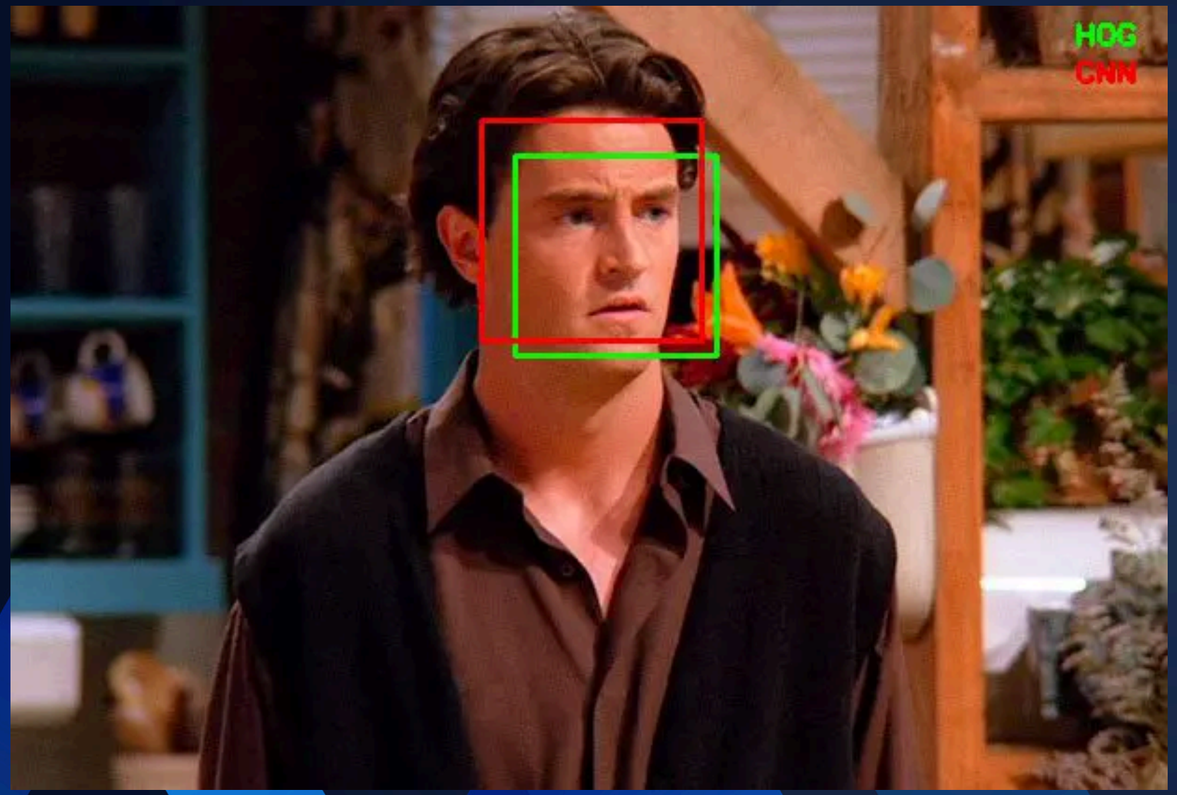
Presented by Seitzhanova Altyn

THE TECHNOLOGY

Their augmented reality filters tap into the large and rapidly growing field of computer vision. Computer Vision is starting to be utilized more and more in our society. It is how you scan your checks and the data is extracted from the lines. It is how you can deposit checks with your phone. It is how Facebook knows who's in your photos, how self-driving cars can avoid running over people and how you can give yourself a dodgy nose.

HOW SNAPCHAT WORKS?

The specific area of Computer Vision that Snapchat filters use is called Image processing. Image processing is the transformation of an image by performing mathematical operations on each individual pixel on the provided picture.



1 — FACE DETECTION:

The first step works like this: Given an input image or video frame, find out all present human faces and output their bounding box (i.e. The rectangle coordinates in the form: X, Y, Width & Height).

FACE DETECTION

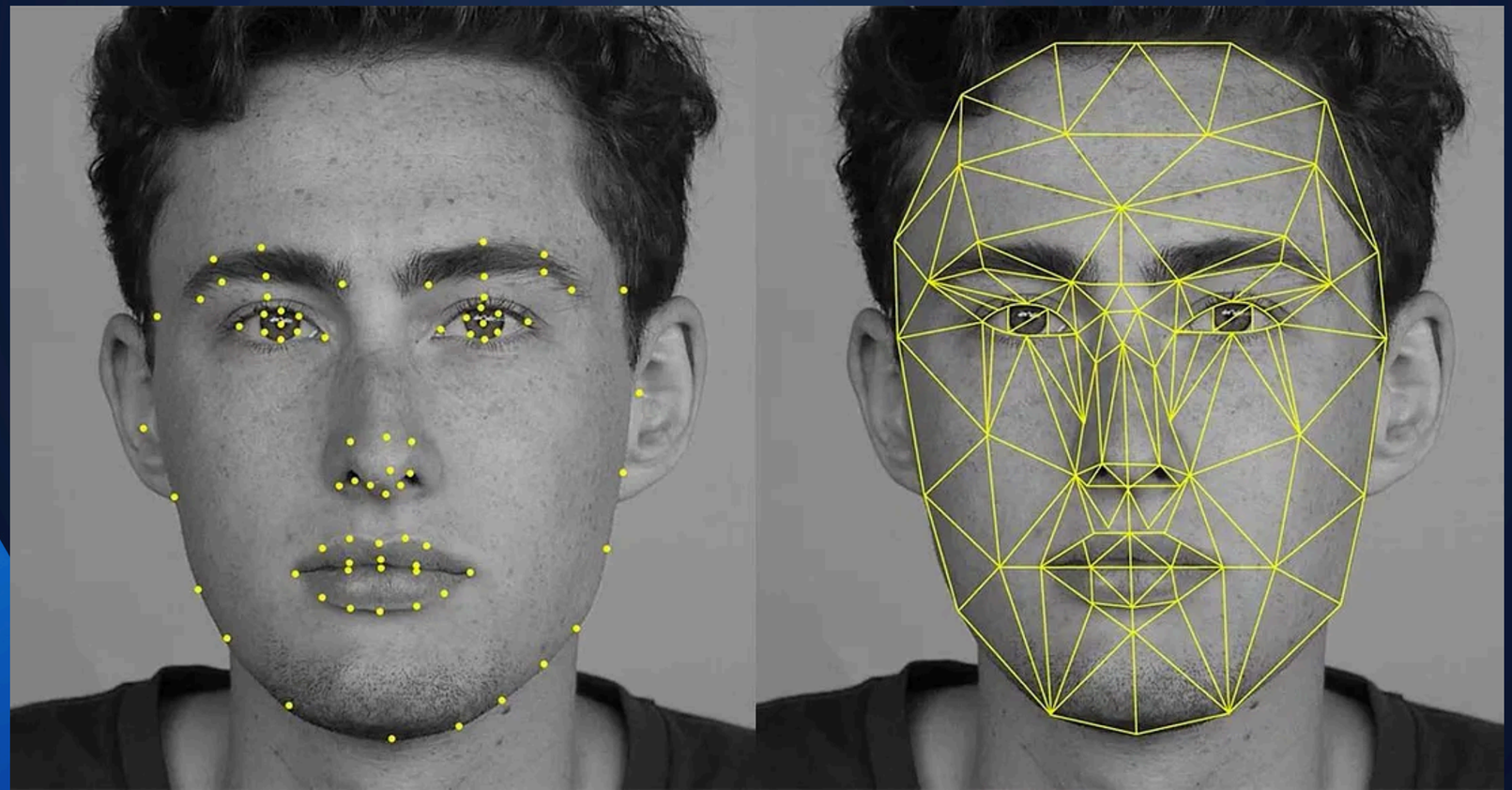
Face detection has been a solved problem since the early 2000s but faces some challenges including detecting tiny, partial & non-frontal faces. The most widely used technique is a combination of Histogram of Oriented Gradients (HOG for short) and Support Vector Machine (SVM) that achieve mediocre to relatively good detection ratios given a good quality image but this method is not capable of real-time detection at least on the CPU.

HOG/SVM

Given an input image, compute the pyramidal representation of that image which is a pyramid of multi scaled downed version of the original image. For each entry on the pyramid, a sliding window approach is used. The sliding window concept is quite simple. By looping over an image with a constant step size, small image patches typically of size 64 x 128 pixels are extracted at different scales. For each patch, the algorithm makes a decision if it contains a face or not. The HOG is computed for the current window and passed to the SVM classifier (Linear or not) for the decision to take place (i.e. Face or not). When done with the pyramid, a non-maxima suppression (NMS for short) operation usually take place in order to discard stacked rectangles.

2 — FACIAL LANDMARKS:

This is the next step in our analysis phase and works as follows: For each detected face, output the local region coordinates for each member or facial feature of that face. This includes the eyes, bone, lips, nose, mouth,... coordinates usually in the form of points (X,Y).



2 — IMAGE PROCESSING

The Active Shape Model is a facial model that has been trained by the manual marking of the borders of facial features on hundreds to thousands of images. Through machine learning, an “average face” is created and aligns this with the image that is provided. This average face, of course, does not fit exactly with the user’s face (we all have diverse faces), so after fitting the face, pixels around the edge of the “average face” are examined to look for differences in shading.

Because of the training that the algorithm went through, (the Machine Learning process), it has a basic skeleton of how certain facial features should look, so it looks for a similar pattern in the given image. Even if some of the initial changes are wrong, by taking into account the position of other points that it has fixed, the algorithm will correct errors it made regarding where it thought certain aspects of your face are. The model then adjusts and creates a mesh (a 3D model that can shift aniuyiy d scale with your face).

CONCLUSION

This implementation demonstrates the effectiveness of HOG+SVM for face detection and Active Shape Models (ASM) for precise facial landmark tracking. By leveraging these techniques, we can achieve accurate real-time detection and seamless overlay of visual elements. The combination of computer vision, affine transformations, and alpha blending enables dynamic modifications while maintaining natural alignment. lity applications.

THANK YOU