Jack Landers
Professor S. Wood
ECEN 640 - Digital Image Processing
2 December 2025

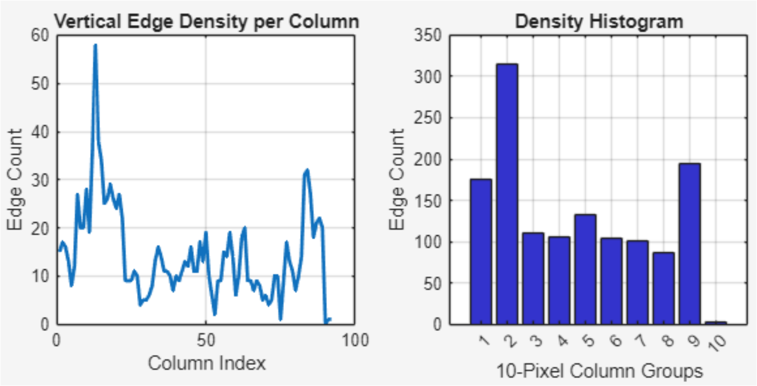
Eye Segmentation with Low-Level Vision Methods

Implementation Description

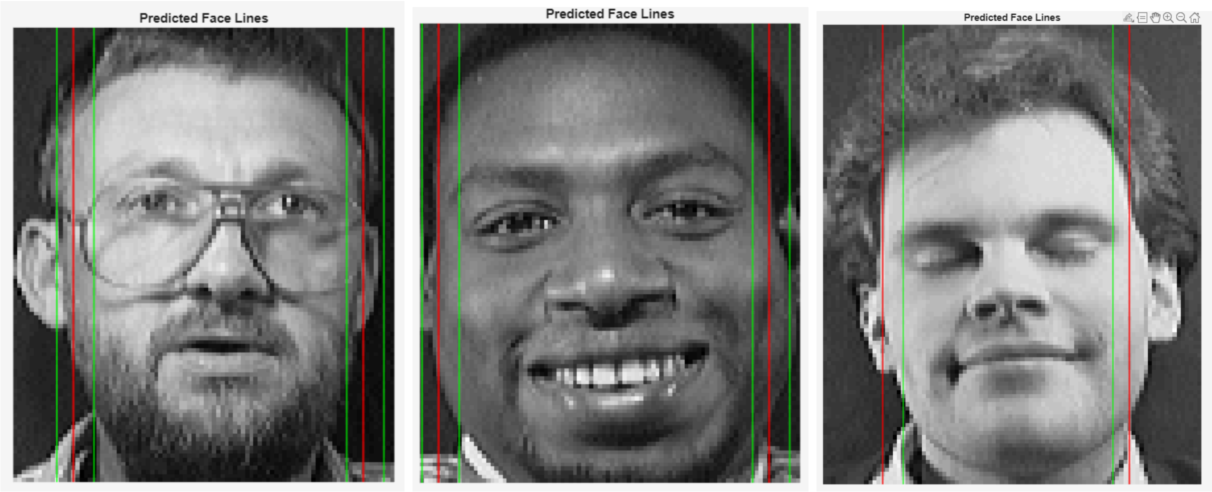
This project aims to explore the effectiveness of a variety of classical digital image processing techniques, in order to segment the pixels representing the eyes of subjects in a portrait photograph. The goal is to generalize across subjects and conditions, such that the precise location of their eyes can be recognized as a subfeature on the face. A final pipeline was structured to combine three methods in sequence: FFT edge evaluation, equalized intensity thresholding, and region approximation filtering. This was built upon developing both a bottom up approach and a top down approach, in order to isolate what steps of the planned process worked best, what was unexpectedly ineffective, and where the useful information from both methods could be extracted.

What was most effective about a bottom up approach was utilizing the combination of Canny and Sobel edges. This combination was found most reliably by preprocessing the image with FFT sharpening. Next, to most effectively combine the edges was not to simply take their intersection, but to locate neighborhoods where they are both detected. A weighted combination also enables the favoring of the Canny filter, such that both glasses and the eyebrows do not have too pronounced of an effect. But in order to take advantage of this, the edges must first be smoothed with a gaussian blur filter, creating a probabilistic overlap map of the edges that can be thresholded to find high frequency neighborhoods. With this, eye regions became more apparent, but there was still a significant quantity of artifacts remaining from the nose, ears, and mouth. A further approach was to find regions on the face that were darkest. A comparison of thresholding various preprocessed versions of the image, found that equalization would give the most accurate results for thresholding to isolate the pupils and eye sockets. This could then be dilated significantly to create a mask over the edge map, removing some of the remaining artifacts while retaining the eye shape. What failed from our proposed bottom up approach was to find correspondence between a sampled region of the image, with a generalized kernel. Testing various ovals, circles, and even cropped images from the dataset, all failed to reliably isolate similar structures, without also highlighting areas that were just naturally curved, such as the sides of the face, ears, and jaws. Instead, the most successful approach was to combine a top-down strategy with a bottom up approach.

A top-down approach showed great success in identifying the regions of the face where the eyes could be. Applying a vertical sobel filter gives the vertical edges of the face. Taking these edges and developing a vertical edge density graph gives insight into which columns define the most gradients from left to right. Using a histogram with a bin size of 10 to collect these values, a reliable approximation for where the sides of the face are can be defined by the two greatest columns. The inner limit of these most frequent columns gives a certain bound for face in almost every single case in the dataset. However, the assumption that the eyes would lie at exactly 1/3rd of the way across the face was almost never the case, due to the variety in rotation. This system is especially reliable for edge case tests, as the gradients caused by artifacts from glasses do not affect the vertical edge distributions significantly, and the cases where eyes are closed and the skin is dark still have sufficient face edges. In the event that this strategy does fail, it is most commonly caused by frequent horizontal edges at the ear region, mainly for images where the ears fall into multiple histogram bin regions. To handle this case, we ensure that the range between the predicted face edges is great enough such that each side of the face is represented, and predict the edges of the face to be within the range 20 to 7-0 if it is not.



The vertical Sobel edges above a threshold are summed across each row to give a vertical edge density plot, that is then collected into bins of size 10 to create a density histogram. (s14/6)



Red lines define the center of the two greatest histogram bins, green lines represent the limits. (s14/6, s22/5, 36/5)

For the vertical limits of the eye region, there was also a high density of horizontal edges at the hairline, the eyebrows, and the chin. This method was very susceptible to misinterpreting the glasses as well, so we use a simpler approach restricting the y-values as the upper middle quartile of pixels. This takes advantage of the reliability of the dataset, and the structure of the images to contain the face proportionally in full, but fails in that it does not take advantage of features detected directly in the image.

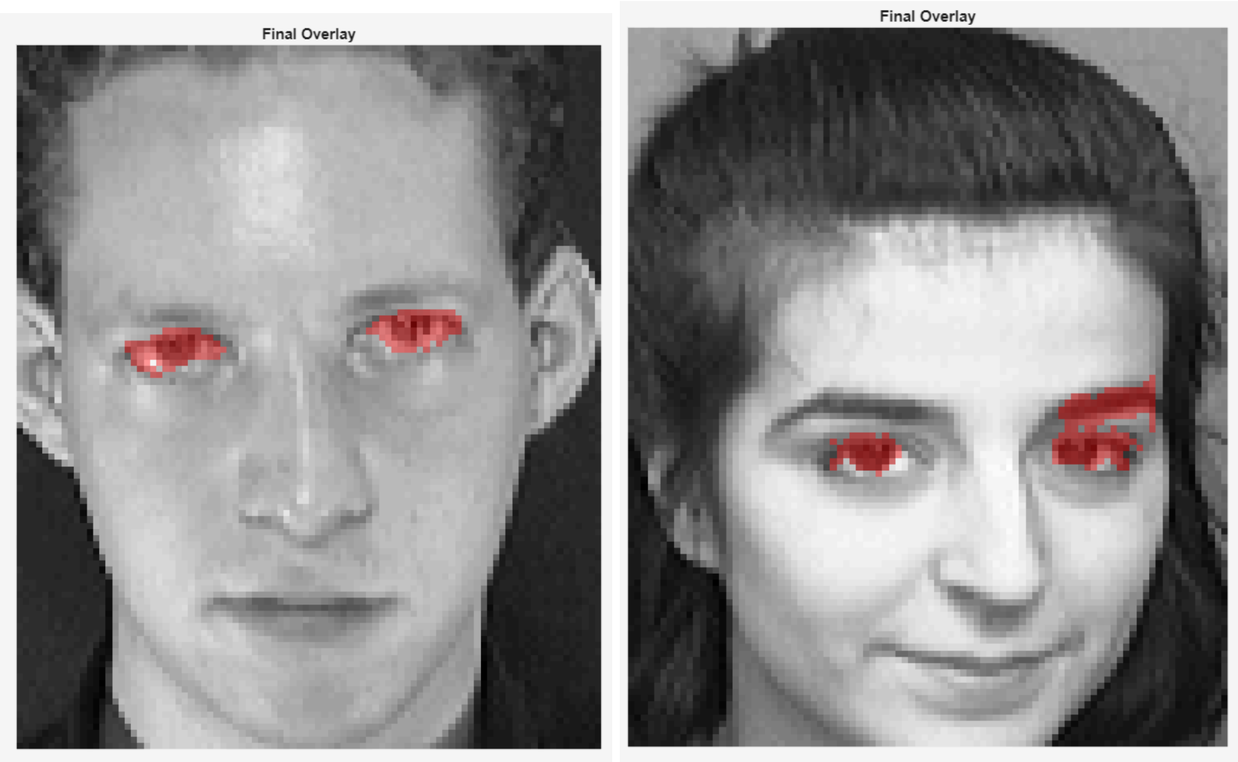
Altogether, this means that the model is built on finding parts of an image that are near very dark points, have a high density of edges, and are located within the region of the image that would most likely be where the eyes lie. Because of the reliance on predicting the vertical limits based on the region in the image, it would not work well for generalized prediction, as other datasets with different image sizes and proportions could not be directly ported without adjusting the parameters.

Implementation Analysis

To build a pipeline that was using the most impactful image processing methods, it was important to determine how each step was performing, by visualizing the results on multiple test cases. To do so, preprocessing with equalization and smoothing, as well as post processing with erosion and dilation was shown for all filter cases to develop an understanding of what circumstances the model works best on. This can be done by inspection.

Often, counting pixels or quantitatively determining an intersection over union metric would be excessively burdensome. To evaluate qualitatively it was important to try multiple photographs of the same person for people with different skin colors and both with and without glasses. For overall analysis the final pipeline result is considered successful for the figure below, where the only selected pixels are all over the eyes and only a few have appeared to be missed. The model has not yet been fully qualitatively tested, but one way to do this that was proposed from the plan, would be to count the number of cases where selected regions are on the eye, how many have been missed, and how many have been falsely identified. Having partially tested the dataset, it is apparent that this model satisfies expectations in that for a high number of cases it is detecting the majority of both eye pixels, however it fails in that it also will detect additional pixel, predominantly caused by glasses or at the eyebrows. While restricting the region shows major improvement at this, it still regularly fails. Other cases where the model fails are images where the transition between the skin and eye is less clear, and at the pupils or whites of the eyes. Because the model is built on finding gradients, and correlating to eye shape was unsuccessful, the pipeline is susceptible to failure in that it misses these parts of the eyes where intensity is constant.

Overall, the model works well in a few cases, but for the most part it is unreliable and without some level or correlation to our target feature, it is likely to hallucinate. Nonetheless, it is a notable achievement that a feature can simply be recognized with mathematical properties describing the data without any learned pattern. For the majority of cases it can detect two regions which each overlap either eye, but segmenting the exact pixels without mistaking the glasses or edges at the nose is a further challenge.



Left: A successful prediction. (s1, 3) Right An unsuccessful prediction (s10, 7)

Proposed Improvements

Improvements could be made in the evaluation of this project. Because counting the pixels is laborious, this could be done with machine learning. A highly accurate model would be capable of finding precisely how inaccurate the outputs are, giving further insight into the strengths and weaknesses. It would also be a significant improvement to be able to successfully correlate the shape or structure of a sample to that of an eye without error. This is where machine learning has the greatest advantage because it is learned. The rounded properties of the eyes could not be detected with a circular hough transformation, but if this was detectable in some way, it is a very unique feature on the face to isolate.

Furthermore, in restricting the eye region, only the face sides were used by detection, and in some cases would fail, meaning that the entire width of the image has to be used. If the region could be predicted based on facial symmetry or other geometric properties, the model would be more prepared for generalizing to other datasets.

In class, I asked about how image properties might be used to detect when an image was produced with generative AI rather than a sampled photograph. Since, I found papers describing how this could be achieved by converting RGB to luminance and then finding color gradients. With this, it can be visually apparent that a photograph is generated by AI, because the artificial image is denoised from diffusion. The resulting processed image shows this hidden noise, unlike a pictured image where the color gradient is a natural smooth wave. This can be identified by thresholding the covariance which is very small for random noise. In addition to this, this noise property could give unique insight into the genAI model used to produce the image. It would be an interesting continuation of this project to develop an AI detection algorithm that ensures that the face is real and not falsified. Such a feature would be extremely important for security applications and could be done using the same low level vision methods prepared here.

The code for this project can be found on my Github: [here]