Embedded Machine Vision and Intelligent Automation

Exercise 1

Sarthak Jain

June 14, 2019

Performed on Jetson Nano

Question 1

1. Components of video analytics:
   Although a broad definition with many disciplines, video analytics can be said to be composed of at least the following, image acquisition and encoding, Computer Vision, Machine Vision, Image Processing, Machine Learning, Real-time interactive systems and storage, networking, databasing and computing.
2. Basic structure of video analytics applications:
   Video analytics applications can eventually be broken down into two sections, one where the processing is performed on the embedded device itself and the other where analysis is performed on a remote cloud/server. The reasons for the same are that it is not always practical to have image and computer vision processing being performed on an embedded device which probably has limited resources.
3. Two use cases of video analytics are provided:
   Namely, Augmented Reality views and Skeletal transforms. Augmented Reality views can be used anywhere in normal life, from improving a consumer's shopping experience to a better filtering of the environment around, leading to a sharpened focus on required objects. Similarly, skeletal transforms can be useful wherever tracking the movements of an animal or human is required.

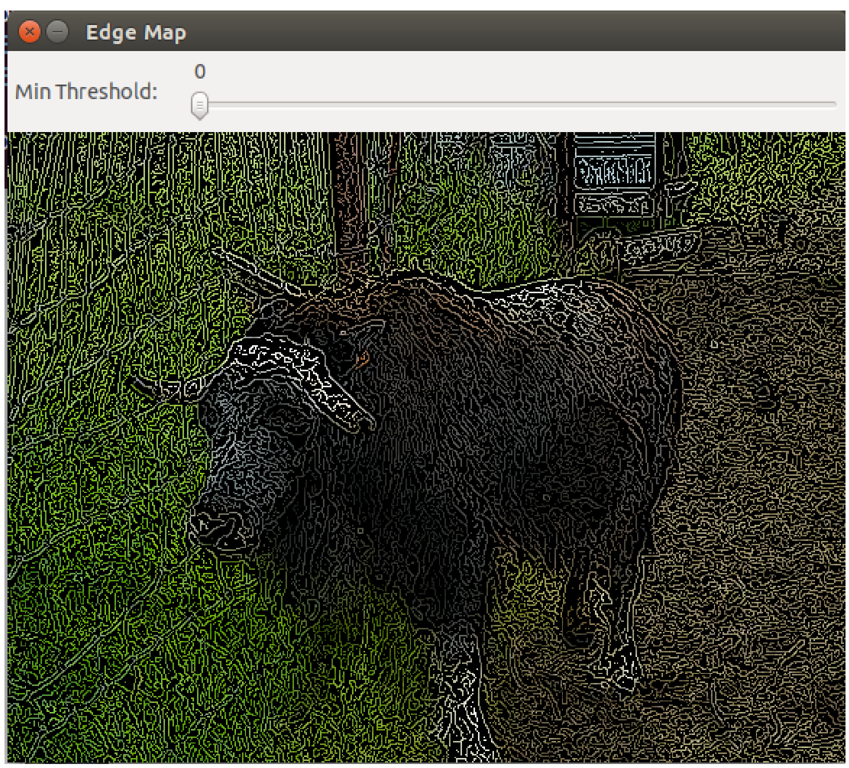
To summarize, I would say video analytics has a lot of scope in the future. Machine vision has already been affecting society in subtle ways, but video analytics has the ability to affect almost everything in our environment. The reach of this field extends to medical domains, defense systems, marketing as well as entertainment systems. Video analytics can be broken down into a few main sections as listed above, and can be broadly implemented with two techniques, either on-chip itself or remotely via cloud. Augmented reality is a massive possible use case of this technology, with applications ranging from self-driving cars to self-aware environments, to traffic safety. Skeletal transform systems are already being used in gesture recognition systems, and can potentially be used to identify rudimentary reactions of animals or humans.

Question 2

Example Test #1: Canny Transform

Original Image

Transformed Image



Example Test #2: Hough-Line transform

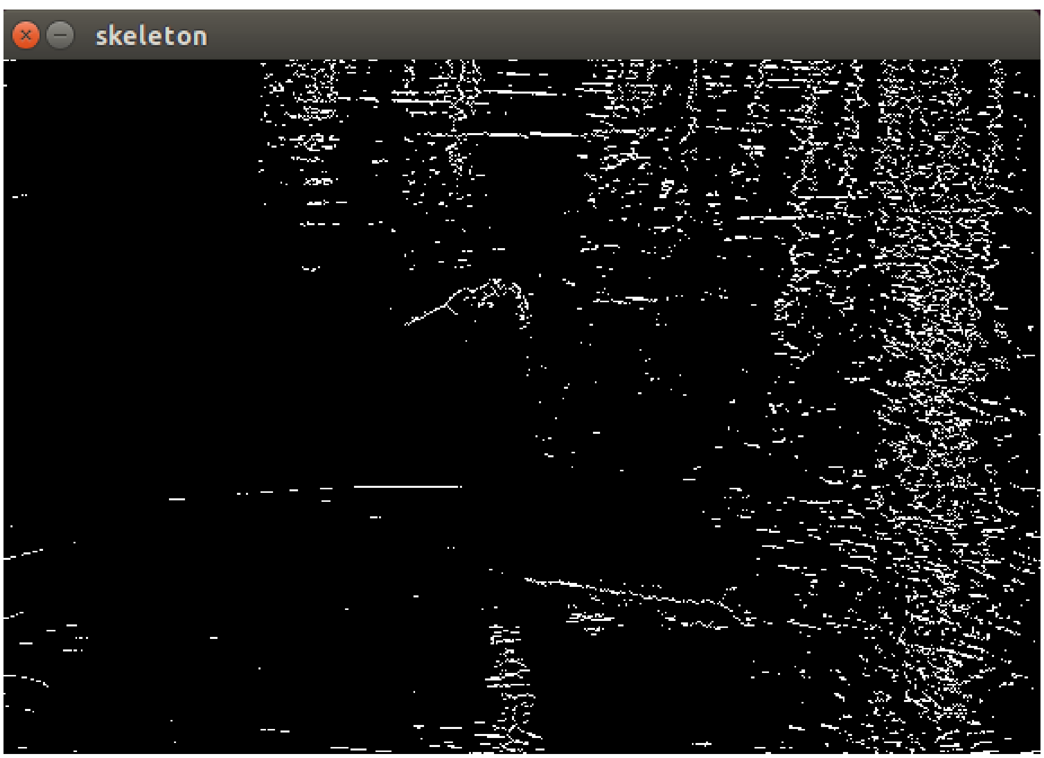
Original Image

Transformed Image

Example Test #3: Skeletal Transform

Original Image

Transformed Image



Sobel Transform:

Original Image

Transformed Image

Sobel Transforms are a very common method of finding any sharp change in the intensity or color of an image. They work by approximating the derivative of an image. In other words, any such change in an image can be found where the pixel value varies vastly from that of its neighbor pixel.

The Sobel transform mask of an image can be calculated as:

a. Horizontal changes:

$$G_x = \begin{bmatrix} -1 & 0 & +1 \\ -2 & 0 & +2 \\ -1 & 0 & +1 \end{bmatrix} * I$$

b. Vertical changes:

$$G_y = \begin{bmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ +1 & +2 & +1 \end{bmatrix} * I$$

c. And lastly, approximation of gradient

$$G = \sqrt{G_x^2 + G_y^2}$$

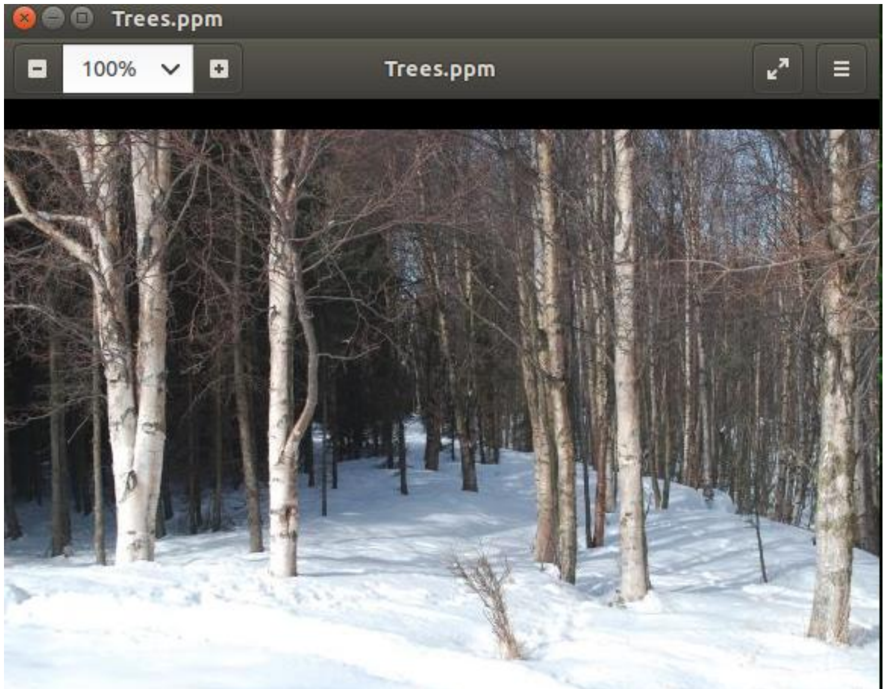
If we consider the horizontal Sobel mask, it will make prominent the horizontal changes in an image. It will calculate difference in pixel intensities along the rows, and since the center row consists of zeroes, the mask will, by moving through the entire image, make prominent the differences between adjacent rows, or vertical changes in pixel intensities.

In a similar manner, the Sobel vertical mask will make prominent the vertical changes in an image.

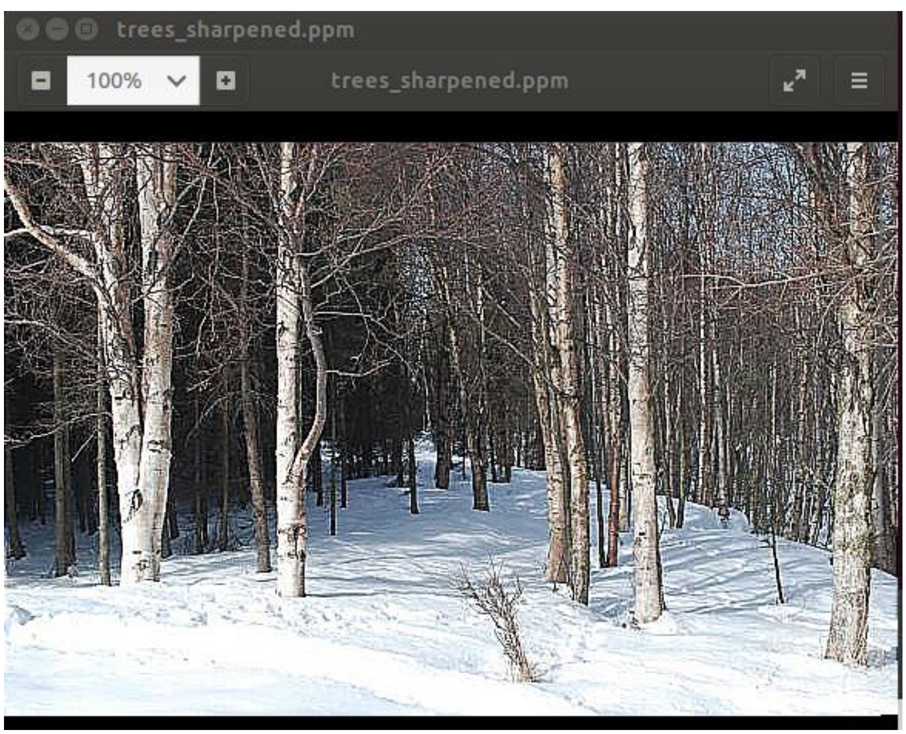
The root of sum of squares gives the Sobel transform over the entire image.

Question 3

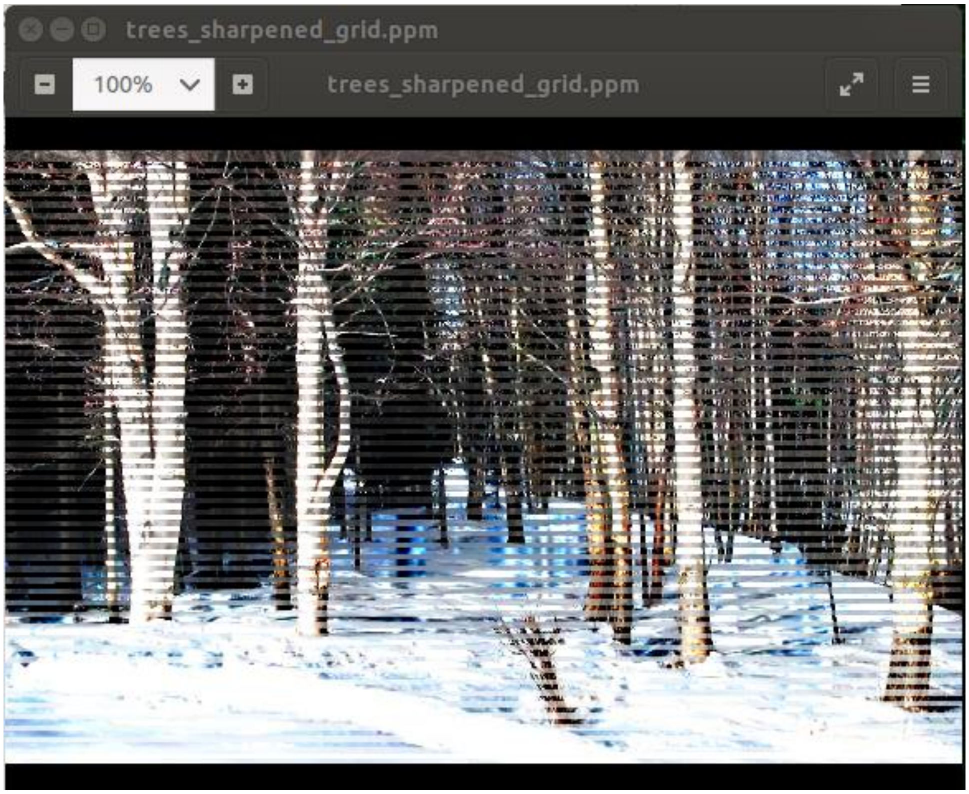
Original Image



Sharpened Image



Sharpened Grid Image



The only code modifications required in the sharpen.c and sharpen_grid.c codes were to change the height and width dimensions to that of the Trees.jpg image, namely 372x580.

PSF explanation:

The Point spread function is basically used to blur images. Mathematically speaking, an impulse function when passed through any system does not result in a delta function, but a slightly enhanced response of the system when convolved with the impulse function.

In scientific terms, the point spread function can be defined as the three-dimensional pattern generated by an ideal source of light. A single point source of light cannot be replicated in 3-D space, and can only be replicated as a diffracted version of the same.

This 3-D effect is the very same thing that eventually proves the PSF as an edge sharpening function. An enhanced version of the function spoken about above is used in coding to remove contrast between spatial points in the image, making it seem more three-dimensional, and providing the effect of having sharpened the images.

Differences between sharpen.c and sharpen_grid.c

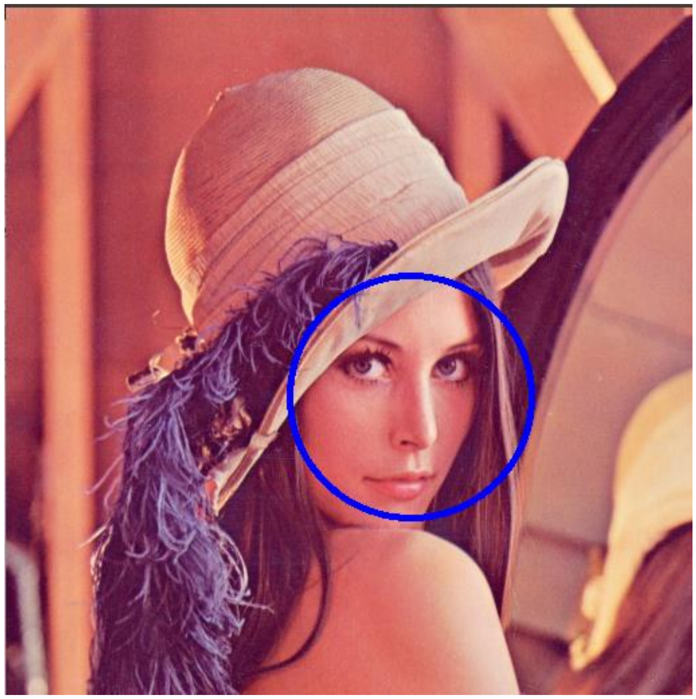
The main difference in the two codes is that of the size of input image that can be provided to each. Sharpen.c can take an input image of size 300x400, whereas sharpen_grid.c can take an image of 3000x4000.

Sharpen.c works by dividing the entire image into its RGB components, and storing them in different arrays, one for each component. The Point spread function is distinctly applied over each channel, and the corresponding output is stored in a combined array, giving the sharpened image.

Sharpen_grid.c, by virtue of its larger inout image size, divides the input image into smaller blocks, or grids. Each block is then processed by a similar process to sharpen.c, but run on a different thread, allowing the code to run faster than it would have otherwise. Each frame has the point spread function applied on it 1000 times, leading to better sharpening before recombining. This leads to better sharpening by the sharpen_grid.c code.

Question 4

Face detected in Lena image



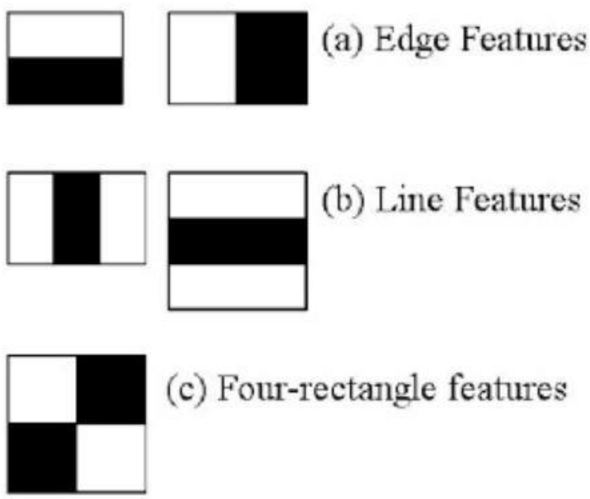
Face detected, Prof. Siewert

Haar cascade classifiers

Haar-based cascade classifiers are an efficient object detection approach wherein a cascade function is trained from a large number of positive and negative images, and then the same classifier is used to detect objects in other images. Face Detection using Haar classifiers uses the simple logic of grouping the features into different stages of classifiers. Each is applied one-by-one. If any group fails the first stage, it is discarded. If it passes, the subsequent stages are applied to it, so on and so forth until the group passes all stages. It is then counted as a face region.

Haar cascade transforms have four stages, Haar feature selection, creation of integral images, Adaboost training and Cascading all the classifiers.

To go into detail, The Haar feature selection process involves the algorithm considering adjacent rectangular frames, summing the pixel intensities in each region and the finding the difference between these sums. The features are shown below:



However, most of these features are almost useless, as the regions must be applied on specific intensity regions. This is where the Adaboost training comes in. It selects the best features and trains the classifiers that combine them.

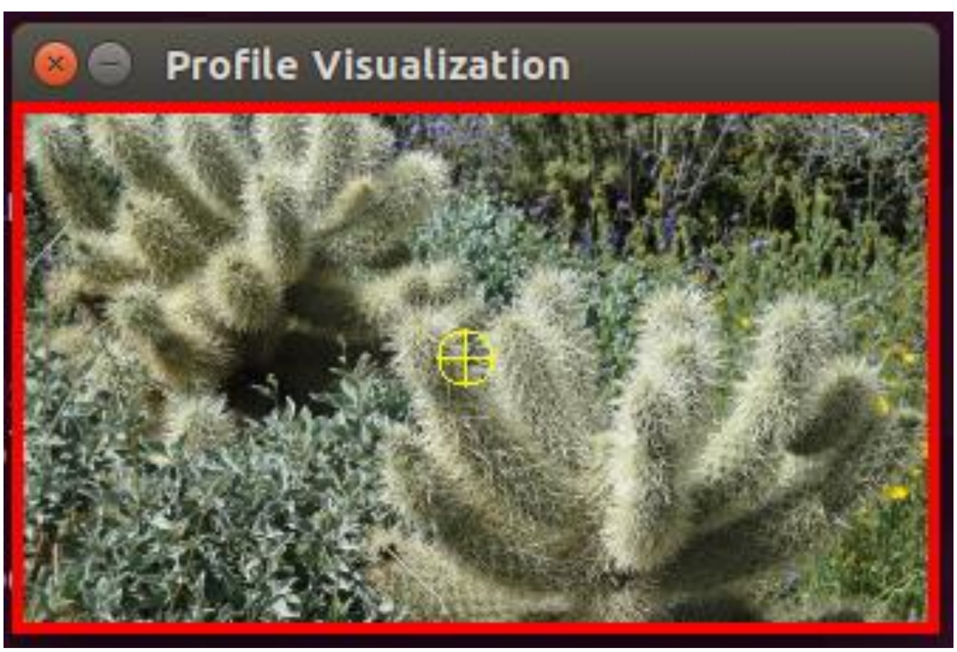
The cascade classifiers consist of a stage of steps, where the weak learning stages are called decision stumps, and each stage is trained by a technique called boosting. The stages are trained to reject negative images and learn the positive ones.

Question 5

Build of simple-opencv

Original cactus image

Modified cactus image with border and crosshairs



References:

1. Explore video analytics in cloud - https://www.ibm.com/developerworks/cloud/library/cl-cloudscaling3-videoanalytics/
2. Sobel Derivatives - https://docs.opencv.org/2.4/doc/tutorials/imgproc/imgtrans/sobel_derivatives/sobel_derivatives.html
3. Understanding Edge Detection - https://medium.com/datadriveninvestor/understanding-edge-detection-sobel-operator-2aada303b900
4. https://docs.opencv.org/3.4.1/d7/d8b/tutorial_py_face_detection.html
5. https://www.olympus-lifescience.com/en/microscope-resource/primer/digitalimaging/deconvolution/deconintro/
6. http://www.willberger.org/cascade-haar-explained/