Touchless Typing using CNNs

Under the guidance of
Dr. Sushama Nagpal

Shivam Kantival 364/CO/13
Tanay Jha 376/CO/13
Ujjwal Peshin 380/CO/13

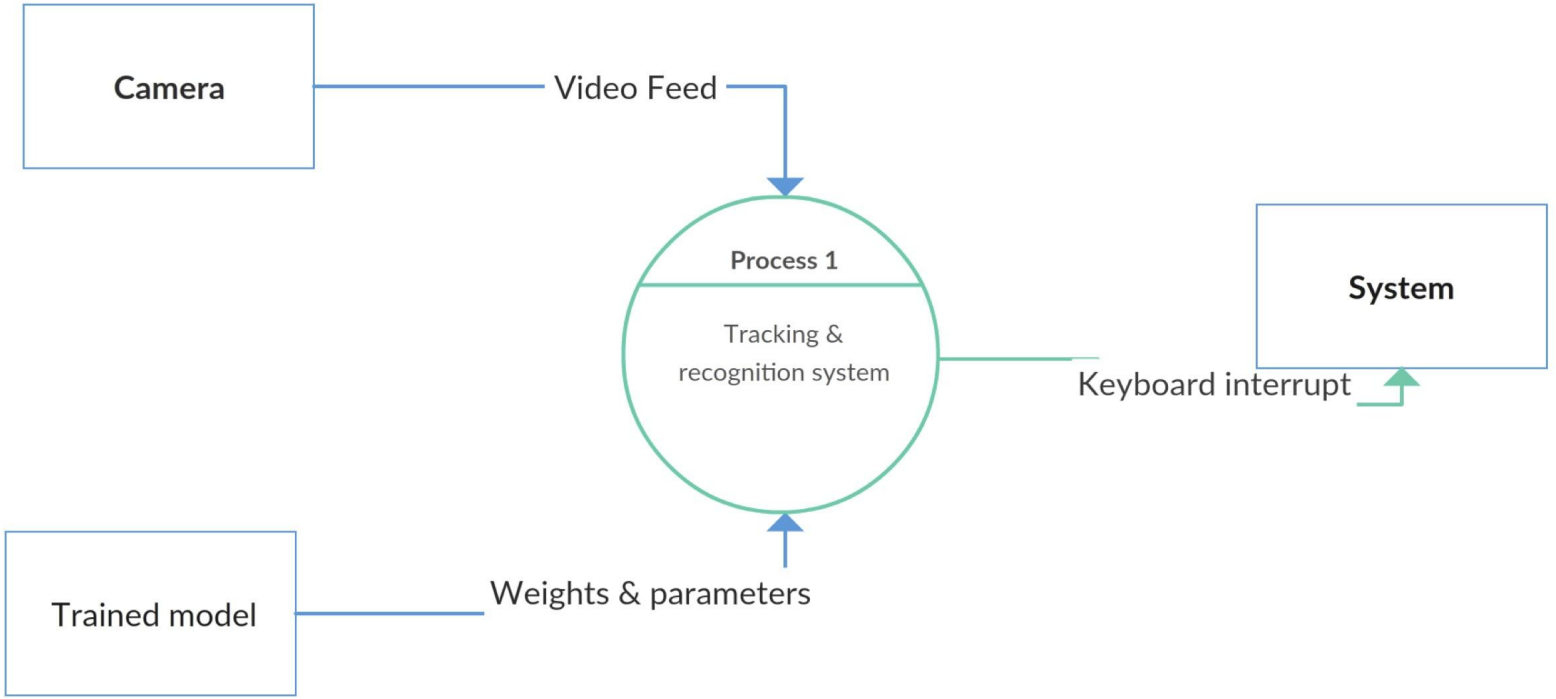
Introduction

- Humans have developed several ways of HCI.
- Gesture recognition allows communication with computers without any mechanical devices.
- A machine has the ability to learn patterns using artificial neural networks.

Motivation

- Intuitive on larger screens
- Beneficial for dirty hands
- Easier to use in the winter with gloves
- Gameplay can be made more natural
- Combined with future 3D displays and projections for immersive experience

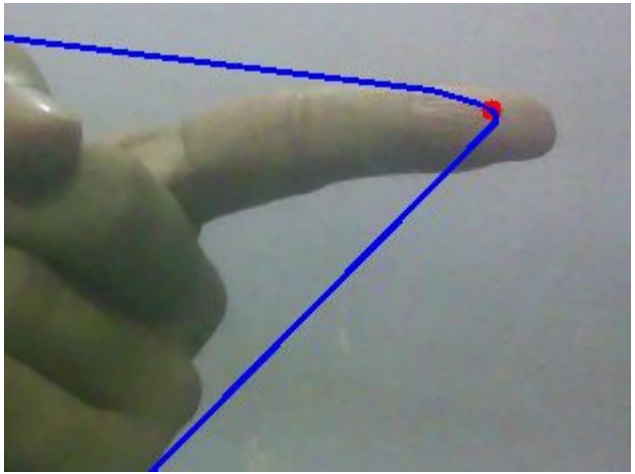
Goal


Camera
Video Feed
Process 1
Tracking & recognition system
System
Keyboard interrupt
Trained model
Weights & parameters

What All We Tried

Using Convex Hull

- Histogram subtracts the background from image, to leave parts of the image that contain skin.
- We can detect the contours having the color of the skin.
- <u>Disadvantages</u>
  - Awkward finger position required.
  - Heavily dependent on the surrounding environment.

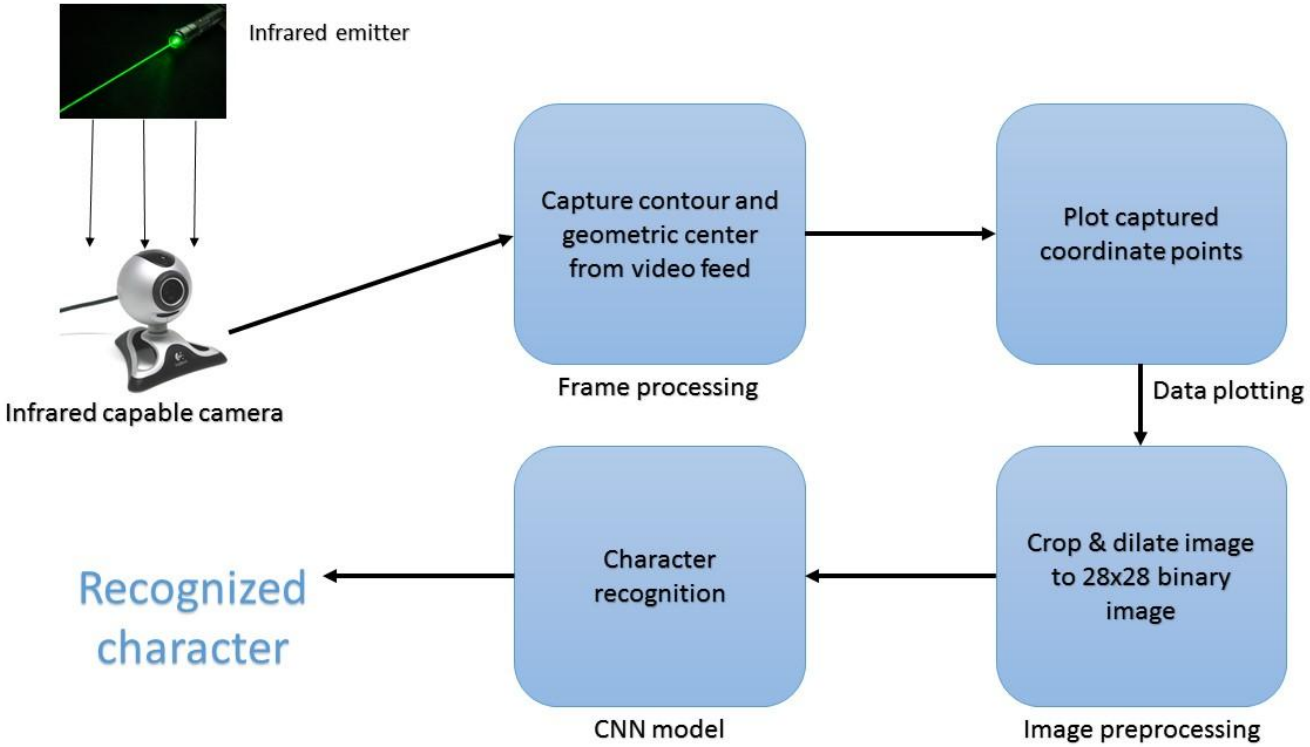


Colored Object Tracking

- Find contours in the video feed having HSV value in a range.
- Choose the contour having the largest area.
- Find and track the center of the circular contour.
- <u>Disadvantages</u>
  - Dependent on the surrounding environment.
  - The object should not be placed very far.

The Present System

Working


Infrared emitter
Infrared capable camera
Capture contour and geometric center from video feed
Frame processing
Plot captured coordinate points
Data plotting
Crop & dilate image to 28x28 binary image
Image preprocessing
Character recognition
CNN model
Recognized character

IR Tracking

- Modify the webcam to detect the IR rays.
- Find contour and the corresponding geometric center.
- The centers of contours found are added to a deque.
- Advantages
  - No dependance on lighting conditions.
  - Can work from large distances.

Plotting Function

- Use the ‘pts’ deque passed to it by the previous module.
- Plot the tracked points having 800x600 dimensions.
- This plot is saved in a temporary file named “foo1.jpeg”.

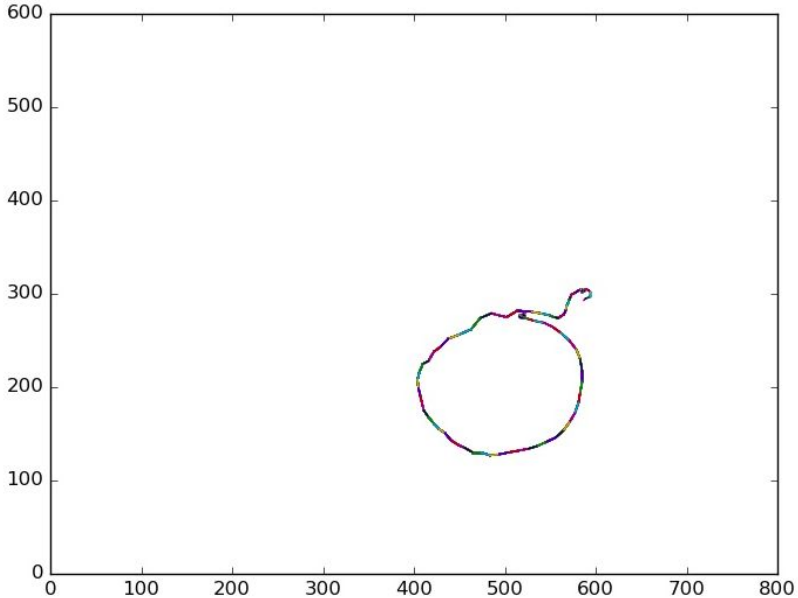
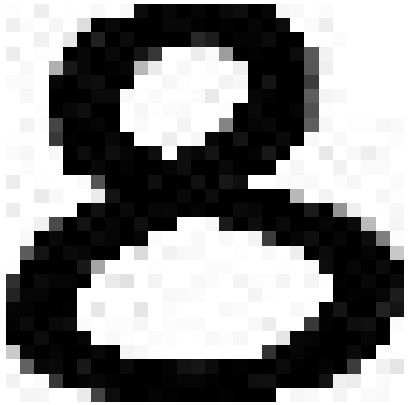


Image Preprocessing

- “foo1.jpeg” is converted to binary image.
- Find AOI by a minimal enclosing rectangle.
- This sub spaced image is then resized to a 28x28 pixel image.
- Necessary dilation is added to the pixels for better prediction.



Predictions and Accuracy

Character Prediction

- LeNet architecture used to predict the character.
- Output is the predicted character and the confidence.
- The CNN used is trained using images from MNIST, and test set is normalized.

Results

```
LENET ARCHITECTURE TRAINED ON MNIST
Number of  0 correctly predicted =  99 /100
Number of  1 correctly predicted =  98 /100
Number of  2 correctly predicted =  99 /100
Number of  3 correctly predicted =  100 /100
Number of  4 correctly predicted =  98 /100
Number of  5 correctly predicted =  100 /100
Number of  6 correctly predicted =  96 /100
Number of  7 correctly predicted =  100 /100
Number of  8 correctly predicted =  99 /100
Number of  9 correctly predicted =  97 /100
Accuracy using MNIST trained LeNet =  98.6
```

Accuracy Achieved by LeNet Architecture trained on MNIST

```
TRANSFER LEARNING USING INCEPTION MODEL
Number of  0 correctly predicted =  20 /20
Number of  1 correctly predicted =  20 /20
Number of  2 correctly predicted =  18 /20
Number of  3 correctly predicted =  19 /20
Number of  4 correctly predicted =  20 /20
Number of  5 correctly predicted =  20 /20
Number of  6 correctly predicted =  20 /20
Number of  7 correctly predicted =  20 /20
Number of  8 correctly predicted =  20 /20
Number of  9 correctly predicted =  19 /20
Accuracy obtained using tranfer learning =  98.0 %
```

Accuracy Achieved by Inception Model trained on custom dataset

```
LENET ARCHITECTURE TRAINED ON CHARACTER DATASET
Number of  a correctly predicted =  506 / 600
Number of  b correctly predicted =  506 / 600
Number of  c correctly predicted =  506 / 600
Number of  d correctly predicted =  506 / 600
Number of  e correctly predicted =  506 / 600
Number of  f correctly predicted =  506 / 600
Number of  g correctly predicted =  0 / 0
Number of  h correctly predicted =  0 / 0
Number of  i correctly predicted =  506 / 600
Number of  j correctly predicted =  0 / 0
Number of  k correctly predicted =  506 / 600
Number of  l correctly predicted =  0 / 0
Number of  m correctly predicted =  400 / 475
Number of  n correctly predicted =  506 / 600
Number of  o correctly predicted =  0 / 0
Number of  p correctly predicted =  506 / 600
Number of  q correctly predicted =  0 / 0
Number of  r correctly predicted =  506 / 600
Number of  s correctly predicted =  0 / 0
Number of  t correctly predicted =  506 / 600
Number of  u correctly predicted =  0 / 0
Number of  v correctly predicted =  0 / 0
Number of  w correctly predicted =  0 / 0
Number of  x correctly predicted =  506 / 600
Number of  y correctly predicted =  506 / 600
Number of  z correctly predicted =  506 / 600
Accuracy using MNIST trained LeNet =  84.32717678100265
```

Accuracy Achieved by LeNet Architecture trained on Character Dataset

Future Scope

- Extended to include other characters.
- Extended to generate complete sentences.
- Use a GPU for better prediction accuracy.
- Improve user experience.

Any Questions ?

Thank You