International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 07 Issue: 06 | June 2020 www.irjet.net p-ISSN: 2395-0072

PixPy – An Application for Image Processing

Abhishek Kashyap[1], Ajinkya Khanzode[2], Hrishikesh Kulkarni[3], Vaibhav Jajoo[4], Prof. Nital Adikane[5]

[1,2,3,4]Student MIT College of Engineering, Pune,
[5]Professor, Dept of MIT College of Engineering, Pune, Maharashtra, India.

Abstract – *Computer Vision is one of the fastest growing and most important technologies in current times. This requires improvements in deep learning and information processing concepts. Our aim in building this system is to extract maximum amount of information as possible from an image. This information will be processed and analyzed in order to be able to provide a helping hand for the computer vision systems. This project will consist of 4 underlying modules:*
1) Image to Text Conversion
2) Text to Speech
3) Object Detection
4) Human Emotion Detection

Key Words: Optical Character Recognition, Image Processing, Convolutional Neural Network.

1. INTRODUCTION

This project strives for maximum efficiency also trying increase the speed by using appropriate datasets. Based on general observations an image consists of three main groups (Texts, Objects and Humans) which form the basis this project. The Applications of the individual modules are:

1) Optical Character Recognition:
- Conversion of files from one format to the other
- Archiving and Retrieval of data that is necessary in print-media, law, government records, libraries, insurance firms and banks.
- With the help of OCR, what could take hours to type, edit and organize could take just minutes or seconds.

2) Text to speech Conversion:
- Communication aid for the blind.
- Vocal Synthesizers for Educational and Telecommunication applications
- Human-Computer interaction.

3) Object Detection:
- Auto-pilot Systems
- Traffic Monitoring
- Anti-theft Systems
- Digital Watermarking

4) Facial Emotion Recognition:
- Digital Media Feedback Systems.
- Video phone and teleconferencing.
- Forensic applications.
- Cosmetology.

2. PixPy Modules:

2.1 Image to text Conversion

This module operates on the concept of Optical Character Recognition (OCR). It is the conversion of images containing typed or handwritten text. The system assumes that the image has simple background and the image itself to be horizontally aligned.

The steps involved are:
Step 1: Image Preprocessing
Step 2: Detection of edges
Step 3: Text Region Identification
Step 4: Modification and Smoothening
Step 5: Alphabet Identification
Step 6: Word Formation
Step 7: Copying to a Text file

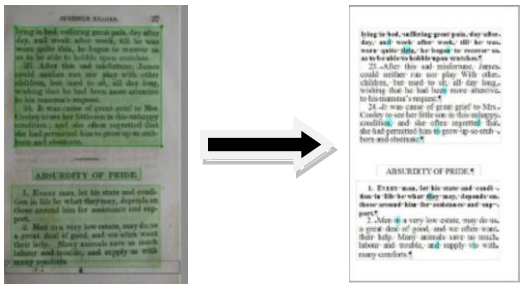


Figure 1: Image to Text

2.2 Text-to-Speech Conversion

A Text to Speech synthesizer utilizes a voice synthesizer that is able to convert text input into spoken words. The basis of this module is the Digital Signal Processing (DSP) that represents text into electrical signals which are converted into synthesized voice.

Making computers talk can improve Human-Computer interaction drastically. The reason behind this is that, voice communication is most familiar mode of information exchange and it is useful for people that are differently abled. The working of Text-to-Speech module is fairly simple. Initially the text obtained is accepted word by word identified by spaces and punctuation marks. Then these words are analyzed and understood by the TTS engine. Next, these words are rendered into electrical signals that are converted to recognizable voice by the transducer present in the speakers.

The Text-to-Speech technology can be implemented by the following methods:

- On demand recorded voice playing.
- Use of Phonemes by assembling into a speaking pattern.
- Use of diaphones for a modern and natural synthesis of voice.

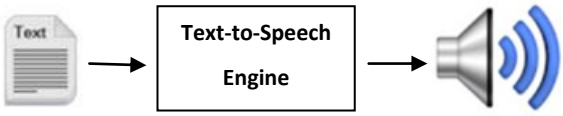


Figure 2: Text to Speech

2.3 OBJECT DETECTION

Object Detection is the technology that deals with localizing and identifying various objects present in an image or in a frame.

In modern methodologies of computer vision, objects are identified using CNN (Computational Neural Network). The CNN works in the following stages:

STAGE 1: Identifying Regions
STAGE 2: Assigning labels to the identified objects
STAGE 3: Localizing and adding description

Convolutional Neural Network:
It is a Deep Learning algorithm that works with the pixels present in an input image to assign weights/biases to obtained information. It requires very little pre-processing as compared to other algorithms, hence provides more speed and accuracy over a dataset. It is basically a collection of ConvNet layers joined to a neural network.

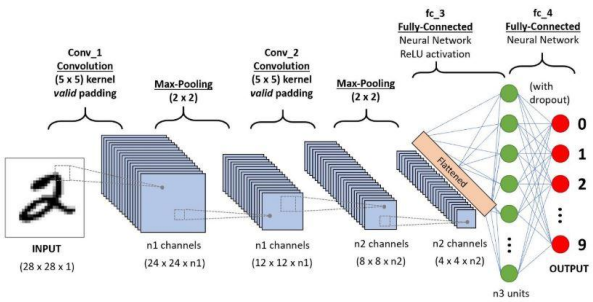


Figure 3: Convolutional Neural Network

The steps of algorithm are:

Step 1: Take the input image/frame.
Step 2: Divide the image in pre-defined number of regions.
Step 3: Consider each segment as an input.
Step 4: Pass these segments to ConvNet layers.
Step 5: Process each pixel in the input using the neural network to assign a weight according to the dataset.
Step 6: Localize the identified region/object.
Step 7: Output the result with a representational box along with tag (as the name of the object).

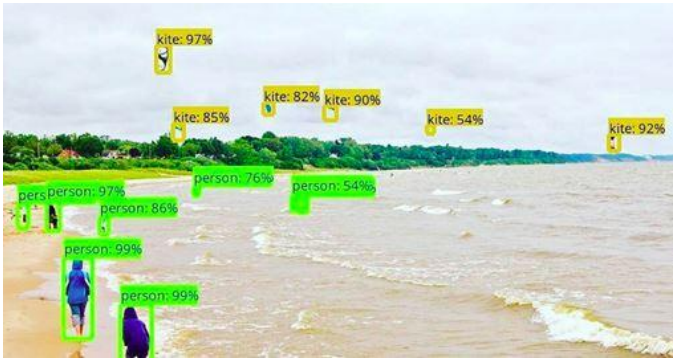


Figure 4: Object Detection

2.4 HUMAN EMOTION DETECTION

This module will be able to recognize facial expressions by using RCNN and 3rd order Bezier curve that works on the identification and analysis of characteristic points of human face.

In this module, the algorithm divides the face structure into three regions of utmost importance:

- Left eye region
- Right eye region
- Mouth region

The pixel movement and density changes present in these regions can be analyzed to identify the emotion.

The working of the algorithm can be as follows:

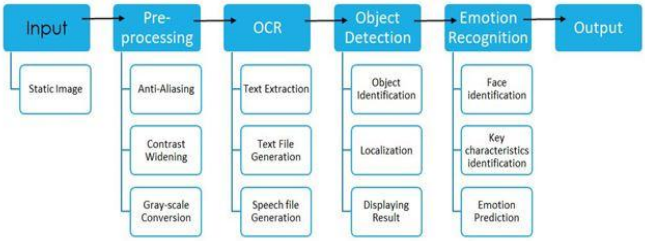
STEP 1: Input image
STEP 2: Contrast Stretching
STEP 3: Color Scheme Conversion into YCrCb scheme
STEP 4: Detecting Connected Regions
STEP 5: Binary Conversion of pixels
STEP 6: Identifying eye and mouth region
STEP 7: Calculating pixel distances and densities
STEP 8: Plotting Bezier Curve
STEP 9: Feeding pixel information to RCNN
STEP 10: Detecting Emotion
STEP 11: Localizing and output

Figure 5: Emotion Recognition

3. TECHNOLOGIES USED

- **Python**
- **Convolutional Neural Network (CNN)**
- **Regional CNN**
- **Tensorflow**
- **Django**
- **Tesseract Engine**
- **Text to Speech Engine**

4. SYSTEM FLOW DIAGRAM



5. CONCLUSION

A system for extracting information present in images has been explained. This system can be improved in terms of speed and accuracy by using specialized datasets, processing engines and advanced hardware. The proposed method is based on the learnings from the previously implemented approaches mentioned in the references. It provides a simple solution to a complex problem and focuses on feasibility in terms of real-life situations. Our future objective would be to test and compare various other approaches along with improved hardware to put forth an even faster and reliable system.

6. REFERENCES

[1] Zhang, S., Zhao, X. and Lei, B. (2012). Facial expression recognition based on local binary patterns and local
fisher discriminant analysis.

[2] Chen, J., Chen, Z., Chi, Z., & Fu, H. (2014, August). Facial expression recognition based on facial components
detection and hog features.

[3] I. Foster and C. Kesselman. The Grid: Blueprint for a New Computing Infrastructure. Morgan-Kaufmann, 2016.

[4] Dynamics of facial expression: Recognition of facial actions and their temporal Segments from face profile image sequences;, Man, Cybern. B, vol. 36, no. 2, pp. 433 449,2006.

[5] 3-D facial expression recognition based on primitive surface feature Distribution", J.Wang, L. Yin, X. Wei, and Y. Su, June 2006, pp.1399-1406.