Patient Monitoring and Unauthorized Person Alarm System

R. Venkadesh[1], Dhiyanu Priya.S[2], Karpagam.R[3], Madhumithra.G[4]

[1] *Assistant Professor, Computer science and engineering, Mahendra Engineering College, Tamilnadu, India*

[2,3,4] *UG Students, Computer science and engineering, Mahendra Engineering College, Tamilnadu, India*

Abstract - **Camera video is a video collected by multi-channel fisheye lens and obtained by splicing and fusion. Compared with traditional video, video has better integrity, immersion and presence. Therefore, Camera video has been widely used in the video field. In order to meet the client's personalized tracking needs in the Camera broadcast system and monitoring system, this system designs and implements a people-tracking system based on personalized distribution in the Camera and provides personalized services for the client to select target objects. Compared with the traditional manual sliding screen for viewing, this automatic live broadcast mode with the target person as the video center greatly improves the user experience. This system uses the method of deep learning to process the video. After the video is processed by the detection unit based on YOLOv3 and the recognition unit based on ResNet50, the coordinates of the target person in the Camera video will be obtained. According to the different objects selected by each client, the local video of the corresponding target object is distributed to each client through a streaming media server to meet the personalized tracking needs of different clients.**

Index Terms - **Alarm system, camera & Machine**

1.INTRODUCTION

Aside from a few technical giants that live and breathe machine learning, most industries are only starting up. There is limited experience with real-life machine learning applications. Companies are overwhelmed with sorting many things out for the first time and rushing to deploy. Data scientists do everything from data cleaning to the A/B test setup. Model operations, maintenance, and support are often only an afterthought. One of the critical, but often overlooked components of this machine learning afterlife is monitoring. With the learning techniques we use these days, a model is never final. In training, it studies the past examples. Once released into the wild, it works with new data: this can be user clickstream, product sales, or credit applications. With time, this data deviates from what the model has seen in training. Sooner or later, even the most accurate and carefully tested solution starts to degrade.

Since entering the 21st century, with the vigorous development of computer graphics and network communication technology, panoramic technology has made great progress, but due to the limitation of the Internet transmission bandwidth, the processing and transmission of panoramic video has encountered considerable difficulties. Nowadays, with the rise of digital image processing technology and network communication technology, the bottleneck of panoramic video has been solved. Especially in the current 5G boom, the transmission rate has been dramatically improved, which has solved the problem of huge amount of data transmission and greatly promoted Development of panoramic video technology. Panoramic video has been widely used in monitoring, broadcasting, video conferencing and other fields. In China, iQiyi, Douyu, and Huya are headed. Overseas, Facebook and Youtobe as representatives, have successively launched 360 panoramic viewing experience services. Taking iQiyi as an example, 5G technology is used to support 4K high-definition panoramic live broadcast. However, there are still blackouts and other phenomena in terms of user interaction, and the multi-client resource allocation and transmission strategy is not well solved, which seriously affects the user experience. Research on resource allocation and transmission strategies has been continuing.

1.1 PRINCIPAL COMPONENT ANALYSIS

The main idea of principal component analysis (PCA) is to reduce the dimensionality of a data set consisting of many variables correlated with each other, either heavily or lightly, while retaining the variation present in the dataset, up to the maximum extent. The same is done by transforming the variables to a new set of variables, which are known as the principal components (or simply, the PCs) and are orthogonal, ordered such that the retention of variation present in the original variables decreases as we move down in the order. So, in this way, the 1st principal component retains maximum variation that was present in the original components. The principal components are the eigenvectors of a covariance matrix, and hence they are orthogonal. Importantly, the dataset on which PCA technique is to be used must be scaled. The results are also sensitive to the relative scaling. As a layman, it is a method of summarizing data. Imagine some wine bottles on a dining table. Each wine is described by its attributes like colour, strength, age, etc. But redundancy will arise because many of them will measure related properties. So what PCA will do in this case is summarize each wine in the stock with less characteristics. Intuitively, Principal Component Analysis can supply the user with a lower-dimensional picture, a projection or "shadow" of this object when viewed from its most informative viewpoint.

Data science is a "concept to unify statistics, data analysis, machine learning and their related methods" in order to "understand and analyze actual phenomena" with data It employs techniques and theories drawn from many fields within the context of mathematics, statistics, information science, and computer science. Data analysis is a process of inspecting, cleansing, transforming,
and modeling data with the goal of discovering useful information, informing conclusions, and supporting decision-making. Data analysis has multiple facets and approaches, encompassing diverse techniques under a variety of names, while being used in different business, science, and social science domains. In today's business, data analysis is playing a role in making decisions more scientific and helping the business achieve effective operation.

2. EXISTING SYSTEM

Driver inattention is among major contributing factors to traffic accidents. There have been and continue to be efforts by governing bodies, car manufacturers, and researchers to prevent driver inattention or, failing that, to mitigate its effects. Many vehicles nowadays come equipped with driver monitoring systems that can alert the driver to, or compensate for, inattention. Moreover, the research community continues to explore and investigate more robust approaches to deal with inattention. Meanwhile, vehicle automation, to various degrees, is becoming more prevalent, with the human's role in the driving task changing depending on the level of autonomy. This necessitates that inattention detection, moving forward, be studied and designed in view of automation and in the context of a specific level of vehicle autonomy. Driver inattention and vehicle automation interact in a complex way, and that needs to be taken into account in the design of future vehicles. We explore this interaction in this paper in light of research findings, and survey inattention detection systems and attempt to contextualize them within popular frameworks for next-generation autonomous vehicles.

3. PROPOSED SYSTEM

Widely used face detection model, based on HoG features and SVM. HOG is a simple and Powerful feature descriptor. This is based on HOG (Histogram of Oriented Gradients) with linear SVM machine learning algorithm to perform Face detection. HOG is a robust object detection because object shape is characterized using the local intensity gradient distribution and edge direction.

4. SOFTWARE REQUIREMENTS

Operating System: Windows 7 ,8, 8.1, 10 64 bit
Software: Python 3.6.8
Tools: Python Tool kit
4.1 MODULES DESCRIPTION
Data Gathering
Train the Recognizer
Recognization
Motion Detection
Alarm
4.1.1 DATA GATHERING
Data Gathering Modules involves process of capturing the facial details of the persons.

Capturing Face Data and make it as a Data set to identify the corresponding person on when it comes to the camera.

Through the camera all of the persons face data was collected and stored as a data set.

4.1.2 TRAIN THE RECOGNIZER

In this module phase all the collected face data sets are trained to the machine.

CNN algorithm is used in this module to train the face recognizer

By using the CNN Algorithm Data sets are trained for the next phase of Recognition Process.

4.1.3 RECOGNITION

Recognition Module finds out the person who are coming through the camera by analysing the collected and trained data sets

When the authenticated person comes through the camera it will allows the person for further process.

In this process camera collects the persons face details by frame rate and compare the frames with the collected data sets.

4.1.4 MOTION DETECTION

Motion detection process involves any unwanted movement in the particular specified area.

It detects the motions and movement in the specific area and it will find out the unwanted moments or unauthenticated persons in the specific zone

4.1.5 ALARM

Alarm is a sound buzzer when an unwanted motions detected in the restriction area, Sounds will make for security purposes.

4.2 CNN ALGORITHM

A convolutional neural network, or CNN, is a deep learning neural network sketched for processing structured arrays of data such as portrayals.

CNN are very satisfactory at picking up on design in the input image, such as lines, gradients, circles, or even eyes and faces.

This characteristic that makes convolutional neural network so robust for computer vision.

CNN can run directly on an underdone image and do not need any pre-processing.

A convolutional neural network is a feed forward neural network, seldom with up to 20.

The strength of a convolutional neural network comes from a particular kind of layer called the convolutional layer.

CNN contains many convolutional layers assembled on top of each other, each one competent of recognizing more sophisticated shapes.

With three or four convolutional layers it is viable to recognize handwritten digits and with 25 layers it is possible to differentiate human faces.

The agenda for this sphere is to activate machines to view the world as humans do, perceive it in an alike fashion and even use the knowledge for a multitude of duty such as image and video recognition, image inspection and classification, media recreation, recommendation systems, natural language processing, etc.

4.3 Convolutional Neural Network Design:

The construction of a convolutional neural network is a multi-layered feed-forward neural network, made by assembling many unseen layers on top of each other in a particular order.

It is the sequential design that give permission to CNN to learn hierarchical attributes.

In CNN, some of them followed by grouping layers and hidden layers are typically convolutional layers followed by activation layers.

The pre-processing needed in a Convent is kindred to that of the related pattern of neurons in the human brain and was motivated by the organization of the Visual Cortex.

4.4 Case Study of CNN for Diabetic retinopathy:

Diabetic retinopathy also known as diabetic eye disease, is a medical state in which destruction occurs to the retina due to diabetes mellitus, it is a major cause of blindness in advance countries.

Diabetic retinopathy influences up to 80 percent of those who have had diabetes for 20 years or more.

The overlong a person has diabetes, the higher his or her chances of growing diabetic retinopathy.

It is also the main cause of blindness in people of age group 20-64.

Diabetic retinopathy is the outcome of destruction to the small blood vessels and neurons of the retina.

Artificial Intelligence has been witnessing a monumental growth in bridging the gap between the capabilities of humans and machines. Researchers and enthusiasts alike, work on numerous aspects of the

field to make amazing things happen. One of many such areas is the domain of Computer Vision.

The agenda for this field is to enable machines to view the world as humans do, perceive it in a similar manner and even use the knowledge for a multitude of tasks such as Image & Video recognition, Image Analysis & Classification, Media Recreation, Recommendation Systems, Natural Language Processing, etc. The advancements in Computer Vision with Deep Learning has been constructed and perfected with time, primarily over one particular algorithm —

a. Convolutional Neural Network.
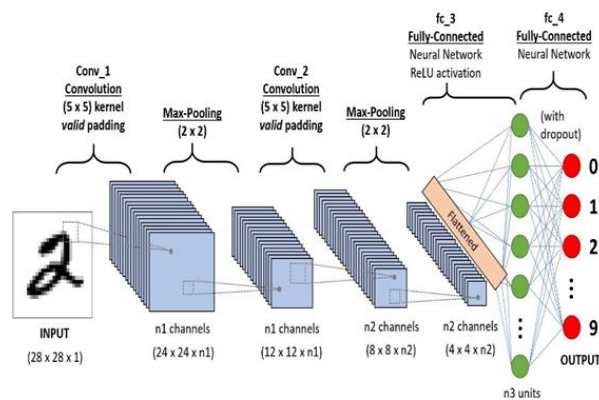
A CNN sequence to classify handwritten digits

A Convolutional Neural Network (ConvNet/CNN) is a Deep Learning algorithm which can take in an input image, assign importance (learnable weights and biases) to various aspects/objects in the image and be able to differentiate one from the other. The pre-processing required in a ConvNet is much lower as compared to other classification algorithms. While in primitive methods filters are hand-engineered, with enough training, ConvNets have the ability to learn these filters/characteristics.

The architecture of a ConvNet is analogous to that of the connectivity pattern of Neurons in the Human Brain and was inspired by the organization of the Visual Cortex. Individual neurons respond to stimuli only in a restricted region of the visual field known as the Receptive Field. A collection of such fields overlap to cover the entire visual area.

Why ConvNets over Feed-Forward Neural Nets?
Flattening of a 3x3 image matrix into a 9x1 vector

An image is nothing but a matrix of pixel values, right? So why not just flatten the image (e.g. 3x3 image matrix into a 9x1 vector) and feed it to a Multi-Level Perceptron for classification purposes? Uh.. not really.

In cases of extremely basic binary images, the method might show an average precision score while performing prediction of classes but would have little to no accuracy when it comes to complex images having pixel dependencies throughout.

A ConvNet is able to successfully capture the Spatial and Temporal dependencies in an image through the application of relevant filters. The architecture performs a better fitting to the image dataset due to the reduction in the number of parameters involved and reusability of weights. In other words, the network can be trained to understand the sophistication of the image better.
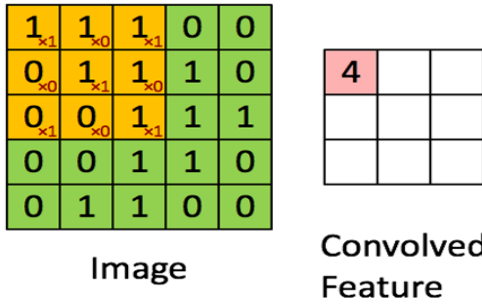
Input Image
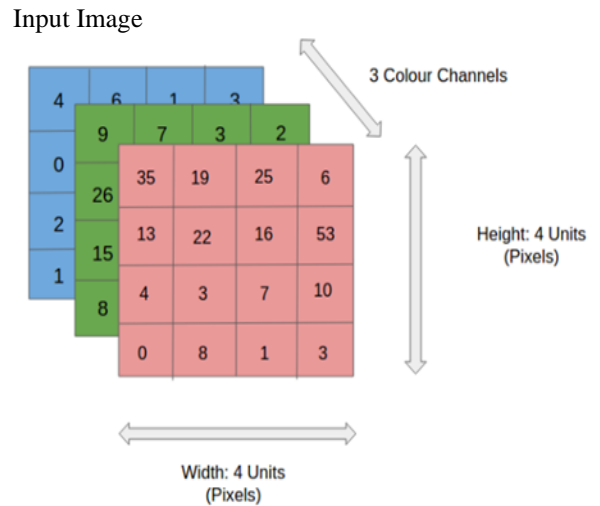
4x4x3 RGB Image

In the figure, we have an RGB image which has been separated by its three color planes — Red, Green, and Blue. There are a number of such color spaces in which images exist — Grayscale, RGB, HSV, CMYK, etc.

You can imagine how computationally intensive things would get once the images reach dimensions, say 8K (7680×4320). The role of the Convent is to reduce the images into a form which is easier to process, without losing features which are critical for getting a good prediction. This is important when we are to design an architecture which is not only good at learning features but also is scalable to massive datasets.

Convolution Layer — The Kernel

Image   Convolved Feature

Convoluting a 5x5x1 image with a 3x3x1 kernel to get a 3x3x1 convolved feature

Image Dimensions = 5 (Height) x 5 (Breadth) x 1 (Number of channels, eg. RGB)

In the above demonstration, the green section resembles our 5x5x1 input image, I. The element involved in carrying out the convolution operation in the first part of a Convolutional Layer is called the Kernel/Filter, K, represented in the color yellow. We have selected K as a 3x3x1 matrix.

Kernel/Filter, K = 1 0 1
0 1 0
1 0 1

The Kernel shifts 9 times because of Stride Length = 1 (Non-Strided), every time performing a matrix multiplication operation between K and the portion P of the image over which the kernel is hovering.
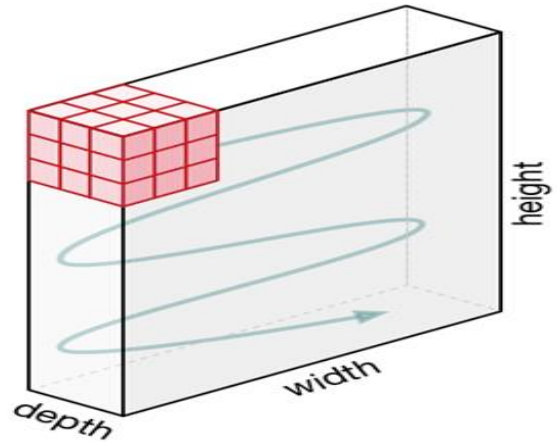
Movement of the Kernel

The filter moves to the right with a certain Stride Value till it parses the complete width. Moving on, it hops down to the beginning (left) of the image with the same Stride Value and repeats the process until the entire image is traversed.
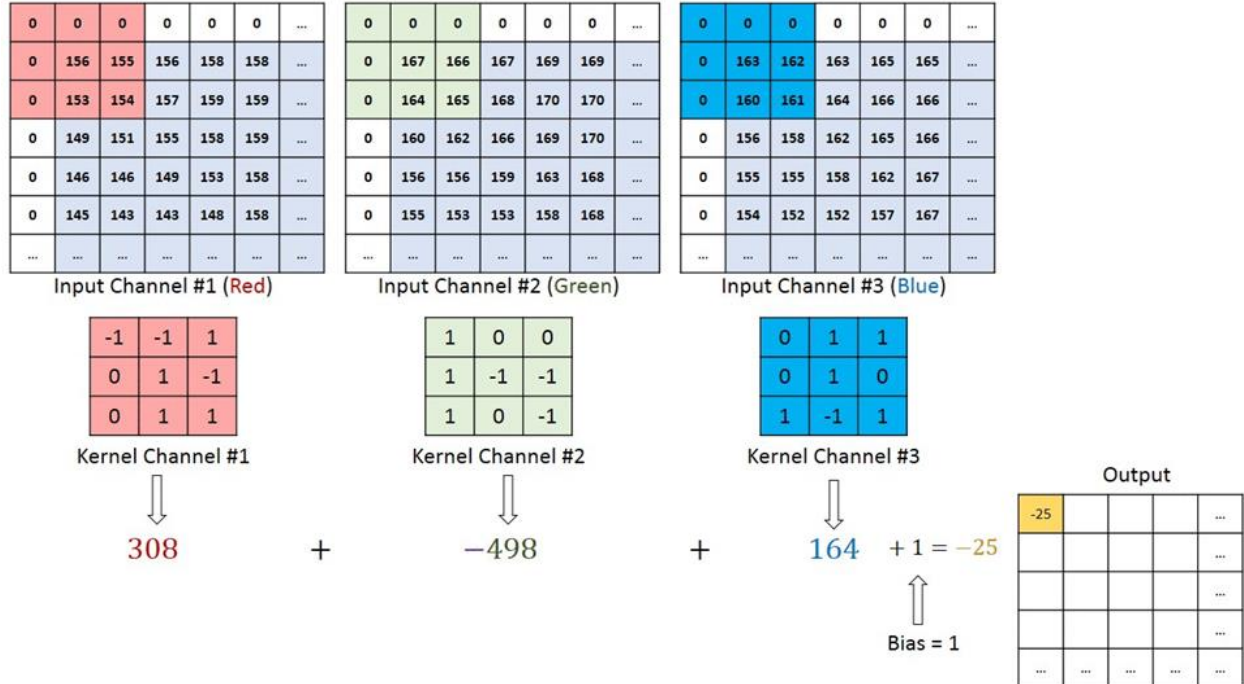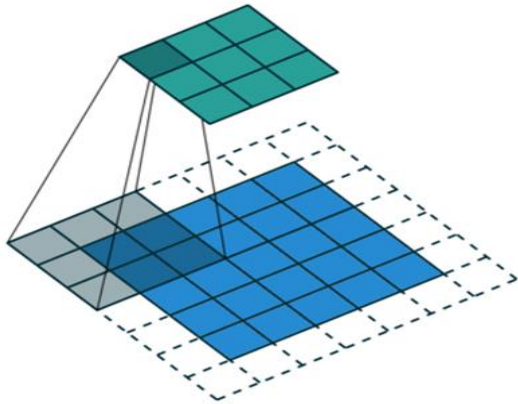

Convolution operation on a MxNx3 image matrix with a 3x3x3 Kernel

In the case of images with multiple channels (e.g. RGB), the Kernel has the same depth as that of the input image. Matrix Multiplication is performed between Kn and in stack ([K1, I1]; [K2, I2]; [K3, I3]) and all the results are summed with the bias to give us

a squashed one-depth channel Convoluted Feature Output.

Convolution Operation with Stride Length = 2

The objective of the Convolution Operation is to extract the high-level features such as edges, from the input image. ConvNets need not be limited to only one Convolutional Layer. Conventionally, the first ConvLayer is responsible for capturing the Low-Level features such as edges, color, gradient orientation, etc. With added layers, the architecture adapts to the High-Level features as well, giving us a network which has the wholesome understanding of images in the dataset, similar to how we would.

There are two types of results to the operation — one in which the convolved feature is reduced in dimensionality as compared to the input, and the other in which the dimensionality is either increased or remains the same. This is done by applying Valid Padding in case of the former, or Same Padding in the case of the latter.
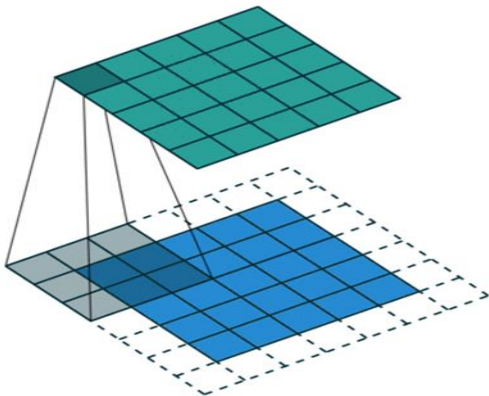
SAME padding: 5x5x1 image is padded with 0s to create a 6x6x1 image

When we augment the 5x5x1 image into a 6x6x1 image and then apply the 3x3x1 kernel over it, we find that the convolved matrix turns out to be of dimensions 5x5x1. Hence the name — Same Padding.

On the other hand, if we perform the same operation without padding, we are presented with a matrix which has dimensions of the Kernel (3x3x1) itself — Valid Padding.

The following repository houses many such GIFs which would help you get a better understanding of how Padding and Stride Length work together to achieve results relevant to our needs.

vdumoulin/conv_arithmetic

A technical report on convolution arithmetic in the context of deep learning - vdumoulin/conv_arithmetic github.com

Pooling Layer

3x3 pooling over 5x5 convolved feature

Similar to the Convolutional Layer, the Pooling layer is responsible for reducing the spatial size of the Convolved Feature. This is to decrease the computational power required to process the data through dimensionality reduction. Furthermore, it is useful for extracting dominant features which are rotational and positional invariant, thus maintaining the process of effectively training of the model.

There are two types of Pooling: Max Pooling and Average Pooling. Max Pooling returns the maximum value from the portion of the image covered by the Kernel. On the other hand, Average Pooling returns the average of all the values from the portion of the image covered by the Kernel.

Max Pooling also performs as a Noise Suppressant. It discards the noisy activations altogether and also performs de-noising along with dimensionality reduction. On the other hand, Average Pooling simply performs dimensionality reduction as a noise suppressing mechanism. Hence, we can say that Max Pooling performs a lot better than Average Pooling.
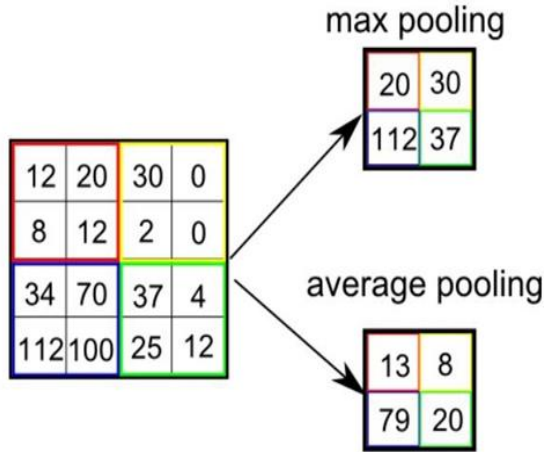
Types of Pooling

The Convolutional Layer and the Pooling Layer, together form the i-th layer of a Convolutional Neural Network. Depending on the complexities in the images, the number of such layers may be increased for capturing low-levels details even further, but at the cost of more computational power.

After going through the above process, we have successfully enabled the model to understand the features. Moving on, we are going to flatten the final output and feed it to a regular Neural Network for classification purposes.

Classification — Fully Connected Layer (FC Layer)
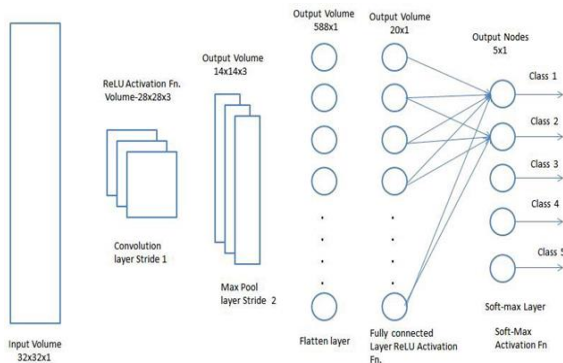
Adding a Fully-Connected layer is a (usually) cheap way of learning non-linear combinations of the high-level features as represented by the output of the convolutional layer. The Fully-Connected layer is learning a possibly non-linear function in that space.

Now that we have converted our input image into a suitable form for our Multi-Level Perceptron, we shall flatten the image into a column vector. The flattened output is fed to a feed-forward neural network and backpropagation applied to every iteration of training. Over a series of epochs, the model is able to distinguish between dominating and certain low-level

features in images and classify them using the Softmax Classification technique.

5. FACIAL RECOGNITION

A fairly simple way is mentioned to implement facial recognition system using Python and OpenCV module along with the explanation of the code step by step in the comments.

Before starting we need to install some libraries in order to implement the code. Below you will see the usage of the library along with the code to install it:

OpenCV:

OpenCV (Open Source Computer Vision Library) is an open-source computer vision and machine learning software library. which is built to provide a common infrastructure for machine learning algorithms and computer vision. It has thousands of optimized algorithms which can be used different purposes like detecting and recognizing faces, identifying objects and many more. We need it to take pictures using our webcam and some manipulation needed to be done in the image.

To install the library you need to install pip in your system after that you can follow the steps in command prompt:

Step 1: pip install opencv-python

Step 2: pip install opencv-contrib-python

5.1 NumPy:

NumPy is the fundamental package for scientific computing in Python which provides a multidimensional array object other mathematical operations can be performed using this but simply speaking we just need it to convert our images into some form of an array so that we can store the model that has been trained. To install the library you can type a simple line of code in your command shell:

pip install numpy

5.2 Haar Cascade

Haar Cascade is basically a classifier which is used to detect the objects for which it has been trained for, from the source. The result is an XML file which stores the trained result. If said simply the Haar Cascade is trained by superimposing the positive image over a set of negative images. The training requires a high spec system and a good internet connection and thousands of training images that is

why it is carried out in the server. For increasing the efficiency of the results they use high-quality images and increase the number of stages for which the classifier is trained. We need haar cascade frontal face recognizer to detect the face from our webcam. To download the haar casade files of different objects you can go the below link: GitHub: HaarCascades

5.3 Python GUI (tkinter)

Tkinter is a simple GUI module used for implementing fairly simple GUI and helps us to interact with code in a simple way. Though for understanding the code its not important for you to know how it works.

If you wanna learn more about Tkinter click the link below

Python GUI – tkinter

5.4 DATA SET PREPARATION

The current technology amazes people with amazing innovations that not only make life simple but also bearable. Face recognition has over time proven to be the least intrusive and fastest form of biometric verification. Facial Recognition is a category of biometric software that maps an individual's facial features and stores the data as a face print. The software uses deep learning algorithms to compare a live captured image to the stored face print to verify one's identity. Image processing and machine learning are the backbones of this technology. Face recognition has received substantial attention from researchers due to human activities found in various applications of security like an airport, criminal detection, face tracking, forensic, etc. Compared to other biometric traits like palm print, iris, fingerprint, etc., face biometrics can be non-intrusive.

They can be taken even without the user's knowledge and further can be used for security-based applications like criminal detection, face tracking, airport security, and forensic surveillance systems. Face recognition involves capturing face images from a video or a surveillance camera. They are compared with the stored database. Face recognition involves training known images, classify them with known classes, and then they are stored in the database. When a test image is given to the system it is classified and compared with the stored database.

5.5 Image Processing and Machine learning

Image processing by computers involves the process of Computer Vision. It deals with the high-level understanding of digital images or videos. The requirement is to automate tasks that the human visual systems can do. So, a computer should be able to recognize objects such as that of a face of a human being or a lamppost or even a statue.

Image reading

The computer reads any image as a range of values between 0 and 255. For any color image, there are 3 primary colors – Red, green, and blue. A matrix is formed for every primary color and later these matrices combine to provide a Pixel value for the individual R, G, B colors. Each element of the matrices provide data about the intensity of the brightness of the pixel.

OpenCV is a Python library that is designed to solve computer vision problems. OpenCV was originally developed in 1999 by Intel but later supported by Willow Garage. OpenCV supports a variety of programming languages such as C++, Python, Java, etc. Support for multiple platforms including Windows, Linux, and macOS. OpenCV Python is a wrapper class for the original C++ library to be used with Python. Using this, all of the OpenCV array structures get converted to/from NumPy arrays. This makes it easier to integrate it with other libraries that use NumPy. For example, libraries such as SciPy and Matplotlib.

5.6 Machine learning

Every Machine Learning algorithm takes a dataset as input and learns from the data it basically means to learn the algorithm from the provided input and output as data. It identifies the patterns in the data and provides the desired algorithm. For instance, to identify whose face is present in a given image, multiple things can be looked at as a pattern:

Height/width of the face.

Height and width may not be reliable since the image could be rescaled to a smaller face or grid. However, even after rescaling, what remains unchanged are the ratios – the ratio of the height of the face to the width of the face won't change.

Color of the face.

Width of other parts of the face like lips, nose, etc.

There is a pattern involved – different faces have different dimensions like the ones above. Similar faces

have similar dimensions. Machine Learning algorithms only understand numbers so it is quite challenging. This numerical representation of a "face" (or an element in the training set) is termed as a feature vector. A feature vector comprises of various numbers in a specific order.

As a simple example, we can map a "face" into a feature vector which can comprise various features like:
Height of face (cm)
Width of the face (cm)
Average color of face (R, G, B)
Width of lips (cm)
Height of nose (cm)

Essentially, given an image, we can convert them into a feature vector like:
Height of face (cm) Width of the face (cm) Average color of face (RGB) Width of lips (cm) Height of nose (cm) 23.1 15.8 (255, 224, 189) 5.2 4.4. So, the image is now a vector that could be represented as (23.1, 15.8, 255, 224, 189, 5.2, 4.4). There could be countless other features that could be derived from the image,, for instance, hair color, facial hair, spectacles, etc.
Machine Learning does two major functions in face recognition technology. These are given below:
Deriving the feature vector: it is difficult to manually list down all of the features because there are just so many. A Machine Learning algorithm can intelligently label out many of such features. For instance, a complex feature could be the ratio of the height of the nose and width of the forehead.
Matching algorithms: Once the feature vectors have been obtained, a Machine Learning algorithm needs to match a new image with the set of feature vectors present in the corpus.

5.7 Face Recognition Operations
The technology system may vary when it comes to facial recognition. Different software applies different methods and means to achieve face recognition. The stepwise method is as follows:

Face Detection: To begin with, the camera will detect and recognize a face. The face can be best detected when the person is looking directly at the camera as it makes it easy for facial recognition. With the advancements in the technology, this is improved where the face can be detected with slight variation in their posture of face facing to the camera.

Face Analysis: Then the photo of the face is captured and analyzed. Most facial recognition relies on 2D images rather than 3D because it is more convenient to match to the database. Facial recognition software will analyze the distance between your eyes or the shape of your cheekbones.

Image to Data Conversion: Now it is converted to a mathematical formula and these facial features become numbers. This numerical code is known a face print. The way every person has a unique fingerprint, in the same way, they have unique face print.

Match Finding: Then the code is compared against a database of other face prints. This database has photos with identification that can be compared. The technology then identifies a match for your exact features in the provided database. It returns with the match and attached information such as name and addresses or it depends on the information saved in the database of an individual.

Face Recognition Softwares: Many renowned companies are constantly innovating and improvising to develop face recognition software that is foolproof and dependable. Some prominent software is being discussed below:

a. Deep Vision AI
Deep Vision AI is a front runner company excelling in facial recognition software. The company owns the proprietorship of advanced computer vision technology that can understand images and videos automatically. It then turns the visual content into real-time analytics and provides very valuable insights. Deep Vision AI provides a plug and plays platform to its users worldwide. The users are given real-time alerts and faster response based upon the analysis of camera streams through various AI-based modules. The product offers a highly accurate rate of identification of individuals on a watch list by continuous monitoring of target zones. The software is highly flexible that it can be connected to any existing camera system or can be deployed through the cloud. At present, Deep Vision AI offers the best performance solution in the market supporting real-

time processing at +15 streams per GPU. Business intelligence gathering is helped by providing real-time data of customers, their frequency of visits, or enhancement of security and safety. Further, the output from the software can provide attributes like count, age, gender, etc. that can enhance the understanding of consumer behaviour, changing preferences, shifts with time, and conditions that can guide future marketing efforts and strategies. The users also combine the face recognition capabilities with other AI-based features of Deep Vision AI like vehicle recognition to get more correlated data of the consumers. The company complies with the international data protection laws and applies significant measures for a transparent and secure process of the data generated by its customers. Data privacy and ethics are taken care of. The potential markets include cities, public venues, public transportation, educational institutes, large retailers, etc. Deep Vision AI is a certified partner for NVIDIA's Metropolis, Dell Digital Cities, Amazon AWS, Microsoft, Red Hat, and others.

b. SenseTime

SenseTime is a leading platform developer that has dedicated efforts to create solutions using the innovations in AI and big data analysis. The technology offered by SenseTime is multifunctional. The aspects of this technology are expanding and include the capabilities of facial recognition, image recognition, intelligent video analytics, autonomous driving, and medical image recognition. SenseTime software includes different subparts namely, SensePortrait-S, SensePortrait-D, and SenseFace.

SensePortrait-S is a Static Face Recognition Server. It includes the functionality of face detection from an image source, extraction of features, extraction, and analysis of attributes, and target retrieval from a vast facial image database

SensePortrait D is a Dynamic Face Recognition Server. The capabilities included are face detection, tracking of a face, extraction of features, and comparison and analysis of data from data in multiple surveillance video streams.

SenseFace is a Face Recognition Surveillance Platform. This utility is a Face Recognition technology that uses a deep learning algorithm. SenseFace is very efficient in integrated solutions to intelligent video analysis. It can be extensively used

for target surveillance, analysis of the trajectory of a person, management of population and the associated data analysis, etc

SenseTime has provided its services to many companies and government agencies including Honda, Qualcomm, China Mobile, UnionPay, Huawei, Xiaomi, OPPO, Vivo, and Weibo.

c. Amazon Rekognition

Amazon provides a cloud-based software solution Amazon Rekognition is a service computer vision platform. This solution allows an easy method to add image and video analysis to various applications. It uses a highly scalable and proven deep learning technology. The user is not required to have any machine learning expertise to use this software. The platform can be utilized to identify objects, text, people, activities, and scenes in images and videos. It can also detect any inappropriate content. The user gets a highly accurate facial analysis and facial search capabilities. Hence, the software can be easily used for verification, counting of people, and public safety by detection, analysis, and comparison of faces. Organizations can use Amazon Rekognition Custom Labels to generate data about specific objects and scenes available in images according to their business needs. For example, a model may be easily built to classify specific machine parts on the assembly line or to detect unhealthy plants. The user simply provides the images of objects or scenes he wants to identify, and the service handles the rest.

d. FaceFirst

The FaceFirst software ensures the safety of communities, secure transactions, and great customer experiences. FaceFirst is secure, accurate, private, fast, and scalable software. Plug-and-play solutions are also included for physical security, authentication of identity, access control, and visitor analytics. It can be easily integrated into any system. This computer vision platform has been used for face recognition and automated video analytics by many organizations to prevent crime and improve customer engagement. As a leading provider of effective facial recognition systems, it benefits to retail, transportation, event security, casinos, and other industry and public spaces. FaceFirst ensures the integration of artificial intelligence with existing surveillance systems to prevent theft, fraud, and violence.

e. Trueface

TrueFace is a leading computer vision model that helps people understanding their camera data and convert the data into actionable information. TrueFace is an on-premise computer vision solution that enhances data security and performance speeds. The platform-based solutions are specifically trained as per the requirements of individual deployment and operate effectively in a variety of ecosystems. The software places the utmost priority on the diversity of training data. It ensures equivalent performance for all users irrespective of their widely different requirements. Trueface has developed a suite consisting of SDK's and a dockerized container solution based on the capabilities of machine learning and artificial intelligence. The suite can convert the camera data into actionable intelligence. It can help the organizations to create a safer and smarter environment for its employees, customers, and guests using facial recognition, weapon detection, and age verification technologies.

f. Face++

Face++, is an open platform enabled by Chinese company Megvii. It offers computer vision technologies. It allows users to easily integrate the deep learning-based image analysis recognition technologies into their applications. Face++ uses AI and machine vision in amazing ways to detect and analyze faces, and accurately confirm a person's identity. Face++ is also developer-friendly being an open platform such that any developer can create apps using its algorithms. This feature has resulted in making Face++ the most extensive facial recognition platform in the world, with 300,000 developers from 150 countries using it. The most significant usage of Face++ has been its integration into Alibaba's City Brain platform. This has allowed the analysis of the CCTV network in cities to optimize traffic flows and direct the attention of medics and police by observing incidents.

g. Kairos

Kairos is a state-of-the-art and ethical face recognition solution available to the developers and businesses across the globe. Kairos can be used for Face Recognition via Kairos cloud API, or the user can host Kairos on their servers. The utility can be used for control of data, security, and privacy. The

organizations can ensure a safer and better accessibility experience to their customers. Kairos Face Recognition On-Premises has the added advantage of controlling the data privacy and security, keeping critical data in-house and safe from any potential third-parties/hackers. The speed of face recognition enabled products is highly enhanced because it does not come across the issue of delay and other risks associated with public cloud deployment. Kairos is ultra-scalable architecture such that the search for 10 million faces can be done at approximately the same time as 1 face. It is being accepted by the market with open hands.

h. Cognitec

Cognitec's FaceVACS Engine enables users to develop new applications for face recognition. The engine is very versatile as it allows a clear and logical API for easy integration in other software programs. Cognitec allows the use of the FaceVACS Engine through customized software development kits. The platform can be easily tailored through a set of functions and modules specific to each use case and computing platform. The capabilities of this software include image quality check, secure document issuance, and access control by accurate verification.

The distinct features include:
A very powerful face localization and face tracking
Efficient algorithms for enrollment, verification, and identification
Accurate checking of age, gender, age, exposure, pose deviation, glasses, eyes closed, uniform lighting detection, unnatural color, image and face geometry
Fulfils the requirements of ePassports by providing ISO 19794-5 full-frontal image type checks and formatting

Utilization of Face Recognition
While facial recognition may seem futuristic, it's currently being used in a variety of ways. Here are some surprising applications of this technology.

Genetic Disorder Identification
There are healthcare apps such as Face2Gene and software like Deep Gestalt that uses facial recognition to detect a genetic disorder. This face is then analyzed and matched with the existing database of disorders.
Airline Industry

Some airlines use facial recognition to identify passengers. This face scanner would help saving time and to prevent the hassle of keeping track of a ticket.

Hospital Security

Facial recognition can be used in hospitals to keep a record of the patients that is far better than keeping records and finding their names, address. It would be easy for the staff to use this app and recognize a patient and get its details within seconds. Secondly, can be used for security purpose where it can detect if the person is genuine or not or is it a patient.

Detection of emotions and sentiments

Real-time emotion detection is yet another valuable application of face recognition in healthcare. It can be used to detect emotions which patients exhibit during their stay in the hospital and analyze the data to determine how they are feeling. The results of the analysis may help to identify if patients need more attention in case they're in pain or sad.

Problems and Challenges

The face recognition technology is facing several challenges. The common problems and challenges that a face recognition system can have while detecting and recognizing faces are discussed in the following paragraphs.

Pose

A Face Recognition System can tolerate cases with small rotation angles, but it becomes difficult to detect if the angle would be large and if the database does not contain all the angles of the face then it can impose a problem.

Expressions

Because of the emotions, human mood varies and results in different expressions. With these facial expressions, the machine could make mistakes to find the correct person identity.

Aging

With time and age face changes it is unique and does not remain rigid due to which it may be difficult to identify a person who is now 60 years old.

Occlusion

Occlusion means blockage. This is due to the presence of various occluding objects such as glasses, beard, moustache, etc. on the face, and when an image is captured, the face lacks some parts. Such a problem can severely affect the classification process of the recognition system.

Illumination

Illumination means light variations. Illumination changes can vary the overall magnitude of light intensity reflected from an object, as well as the pattern of shading and shadows visible in an image. The problem of face recognition over changes in illumination is widely recognized to be difficult for humans and algorithms. The difficulties posed by illumination condition is a challenge for automatic face recognition systems.

Identify similar faces

Different persons may have a similar appearance that sometimes makes it impossible to distinguish.

6. CONCLUSION

With the continuous development of panoramic technology, the application of panoramic technology is getting wider and wider. How to meet the personalized needs of clients under panoramic video has become an issue of increasing concern. According to this hot issue, this article adds personalized video distribution processing technology on the basis of panoramic video and combined with deep learning methods, using detection and re-identification functions to implement a multi-client tracking system for free choice of target objects. It provides a reference basis for the future research on personalized video tracking service of panoramic video.

REFERENCES

[1] Qian F, Ji L, Han B, et al. Optimizing 360 video delivery over cellular networks, Proceedings of the 5th Workshop on All Things Cellular: Operations, Applications and Challenges. ACM, 2016: 1-6.

[2] Yuan Liu, Ruigang Wang. Research and Implementation of P2P Network Content Distribution Centralization, Journal of Xi'an

University of Arts and Science: Natural Science Edition, 17(1): 105-109, 2014.

[3] Xinxiang Wang, Research and Design of Streaming Media Distribution System Based on P2P, Science and Technology Plaza, 2013 (9): 113-116.

[4] ZHANG Zhen, ZHANG Lei. Research on CDN Video Content Delivery Technology in CCN, Software, 36(01): 67-71, 2015.

[5] Lécun Y, Bottou L, Bengio Y, et al. Gradient-based learning applied to document recognitio, Proceedings of the IEEE, 1998, 86(11): 2278-2324.

[6] Krizhevsky A, Sutskever I, Hinton G E. Imagenet classification with deep convolutional neural networks, Advances in neural information processing systems, 2012: 1097-1105.

[7] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition, arXiv preprint arXiv, 2014: 1409.1556.

[8] He K, Zhang X, Ren S, et al. Deep residual learning for image recognition, Proceedings of the IEEE conference on computer vision and pattern recognition, 2016: 770-778.

[9] Girshick R, Donahue J, Darrell T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation, Proceedings of the IEEE conference on computer vision and pattern recognition, 2014: 580-587.

[10] He K, Zhang X, Ren S, et al. Spatial pyramid pooling in deep convolutional networks for visual recognition, IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015,37(9): 1904-1916.