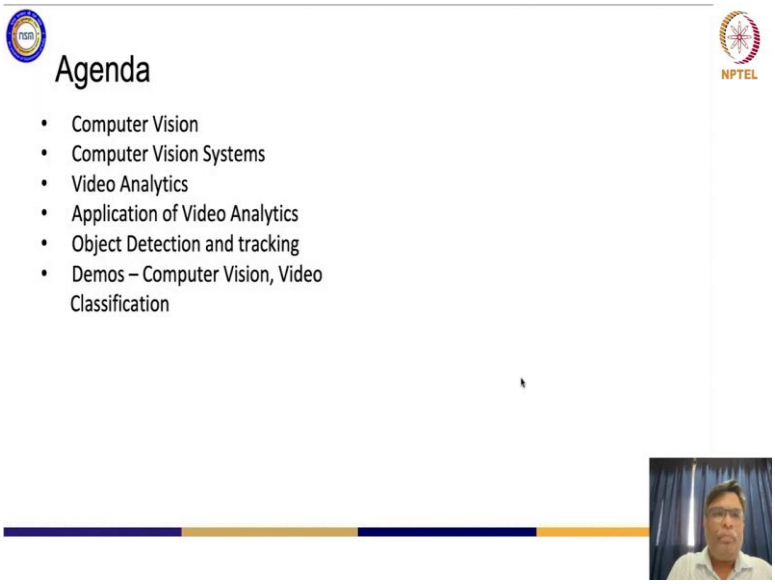
Applied Accelerated Artificial Intelligence
Prof. Satyadhyan Chickerur
Department of Computer Science and Engineering
Indian Institute of Technology, Palakkad

Lecture - 54
Applied AI: Smart City (Intelligent Video Analytics)
Session 1 - Part 1

So, today we are having this session on Intelligent Video Analytics and this particular thing is going to be linked to a use case of smart city and we have split this into two specific sessions, wherein in session 1, we are trying to actually start with how basically video analytics has been going on ok without GPUs and with GPUs.

In today's session and tomorrow, or the next class, we are going to discuss about how basically certain specific streaming modules right, which are basically termed as deep stream and certain other video analytic libraries and frameworks could be used effectively for faster training as well as inferencing. So, these are two specific sessions which we would be handling ok.

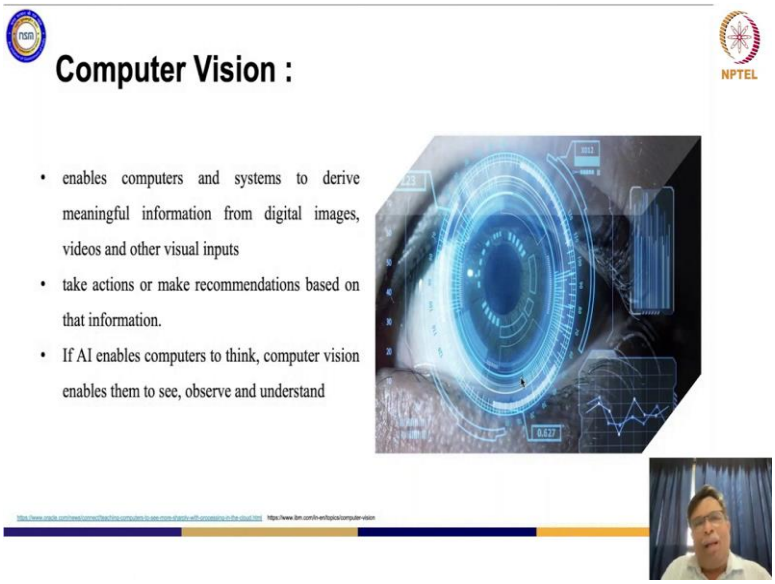
(Refer Slide Time: 01:28)



So, to start with, today's agenda let us start with a very very basic cursory introduction to what a computer vision is, what would comprise of computer vision systems, what is video analytics, what are the various applications of video analytics and what is object detection, what is object tracking and we will have certain demos.

And these demos today, we will have it which are simple CPU based programs; then, going up to the GPU based programs and those are on colab right. So, we will also share the colab links. So, which you can work on ok and then, we will try to do some computer vision applications and video classification applications to end today's class session.

(Refer Slide Time: 02:20)



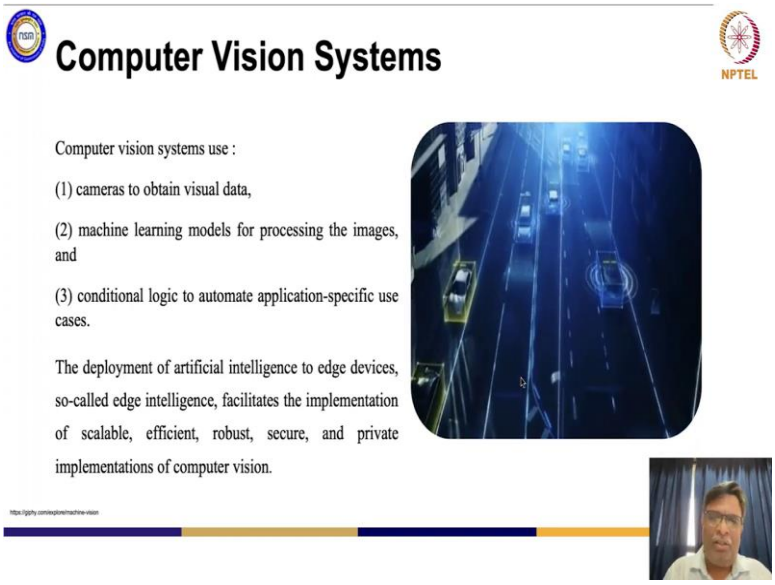
So, what basically is computer vision right? As we all know it is going to enable computers and systems to derive meaningful information from digital images, videos and other visual inputs. So, when we talk of digital images and videos, now to this when we say other visual inputs, we can add certain things like point cloud data or 3D models from Kinect camera and all of this right.

One of the examples which we will be showing is inflated 3D today and this basically uses the inputs from the Kinect camera right. So, we will try to see that as well and the basic idea of using computer vision as you would have seen in case of COVID times right, a lot of computer vision applications were being used for so many use cases and so many applications right.

Starting from trying to detect temperature, to trying to find out the distance between two people right whether they are maintaining social distance or not and now, there are the applications wherein even if you wear a mask you can be detected and so on and so forth right. So, the basic idea of such computer vision systems is to take actions or make recommendations based on the information which you are collecting.

Now, this information can be in real time, it can be dynamic or it can be actually recorded and based on that, you can take certain decisions right. Now, the idea is if we are trying to develop AI applications and if AI is enabling computers to think, computer vision is something which enables the AI module right to see, observe, understand and then, take a decision right. So, this is how basically you would link the computer vision thing with the existing artificial intelligence applications right.

(Refer Slide Time: 04:24)



Now, when you say computer vision systems right on the right side, you are seeing certain autonomous vehicles right and trying to basically do things right which a human driver generally does for example, changing the lane, trying to detect and this is all actually involving a lot of things. It is just not computer vision right; it has got communications; it has got information exchange right.

So, you have the software defined networks, 5G, all of this is connected right. But the basic idea is we are trying to concentrate on imaging applications right at present and computer vision systems for that matter would be using cameras to obtain the visual data. So, these cameras right which are sensors can be of various types right, we are not going into the details; but you have got various types of cameras right, you can get the heat map ok from a camera.

You can similarly get 3D point cloud information from a camera. So, there are various types of cameras available and each of these cameras or modalities will give us a visual

data which is represented in some specific form right and these vision systems will basically use this type of data obtained from various sensors which are cameras and would use machine learning models for processing these images.

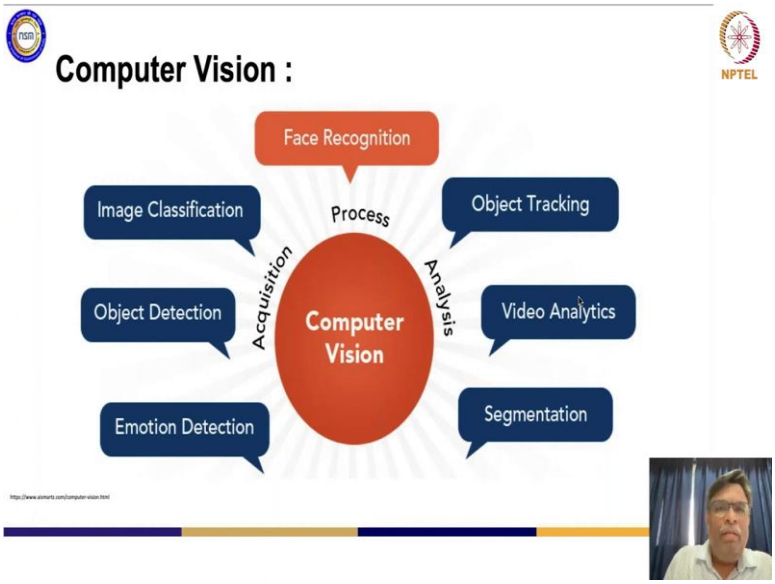
So, these can be images, these can be 3D models, these can be a point cloud, these can be so on and so forth ok. So, this is how you are going to actually develop your system. Now, these systems will also use right conditional logic to automate application specific use cases.

So, this basically means that you will have to actually develop a automated application and for each of this right, you will have to have a specific logic or a business logic which you need to develop and then, come up with your own computer vision system doing x, y, z application right.

So, now, one of the things is when you want to deploy this right, the deployment of artificial intelligence or deployment of these applications on edge devices or you call it as edge intelligence is going to actually facilitate the implementation of scalable robust secure and private implementations of computer vision. So, this is what it means.

So, now, nowadays people are talking of how do you secure these computer vision systems using block chain and so on and so forth; we are not going into those details. But the idea is that these systems need to be secure right and so, how is that security aspect also comes into in computer vision systems is also a good area of research nowadays.

(Refer Slide Time: 07:28)



So, generally, if you see computer vision applications from the user point of view or from programmer point of view, you have got so many things which people talk about right. But the effective idea is from the sensors, you do something called as acquisition and then, you process them and then, you do some analysis.

So, this is how broadly a computer vision application actually is developed. So, you acquire, you process and then you do the analysis right. So, process can be preprocessing or post processing after acquisition; but analysis has to be done on that particular information.

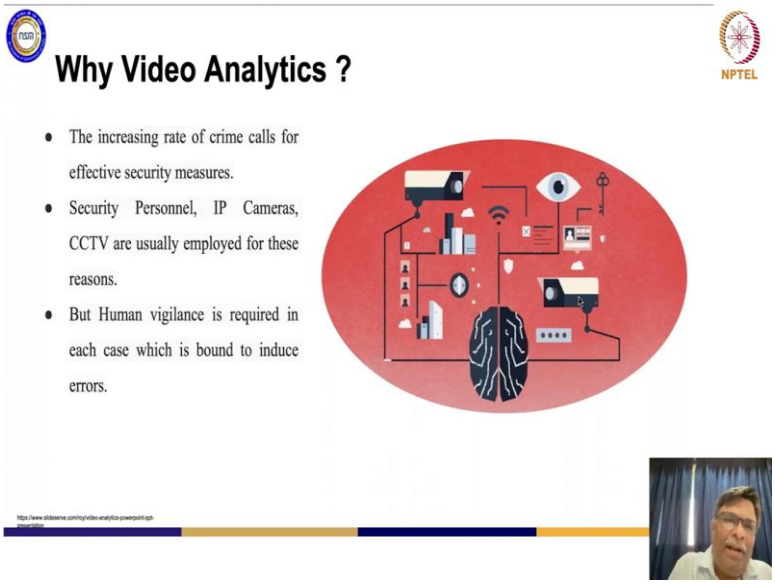
Now, you have got applications of emotion, detection, object detection, image classification and face recognition, then you have object tracking, you have video analytics and we have segmentation. So, these are various applications ok which computer vision experts try to work on and one of the things which we would be concentrating today is on video analytics right. So, this is what would be the concentration of today's work. Excuse me. Right.

(Refer Slide Time: 08:42)

So, what actually is video analytics? So, video content analysis or video content analytics is also known as video analysis and video analytics, which basically is the capability of automatically analyzing videos to detect and determine the temporal and spatial elements. So, the basic idea here is even if you try to assume a video to be a set of static frames, but ultimately the results which have to be coming to you in real time is to be actually given to you in the form of a video itself right.

So, instead of just trying to have a static image ok and then, trying to do a bounding box on that, you will have to actually do a video analytics or a bounding box inside a video, when the video frames are moving right. So, this is what basically you are trying to do and you are trying to actually determine the temporal and spatial events right.

(Refer Slide Time: 09:50)

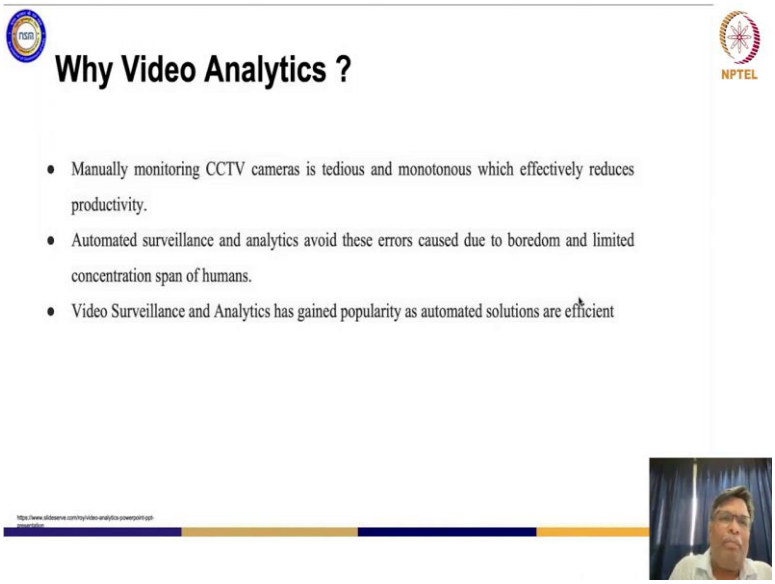


Yeah, so, why video analytics right; why video analytics? So, the use cases if you see the increasing rate of crime calls for effective security measures, security personnel IP cameras, CCTVs are usually employed for these reasons. So, the data is tremendously getting generated right and you have got all of this data with you.

But how is that you are going to actually come out with certain specific informative decision making results right. So, that is why this video analytics ok is gaining a lot of importance and human vigilance is required in each case which is bound to induce errors.

Now, the idea is this that assuming that you automate everything right, your IP cameras, your CCTVs, everything is foolproof right; but still you will have to have dependency on human vigilance which tries or wherein, certain errors might be induced right or introduced. So, the idea is this that we should try to do in such a way our application development that the errors which are induced need to be either reduced or has to be totally removed right. So, this is the basic idea.

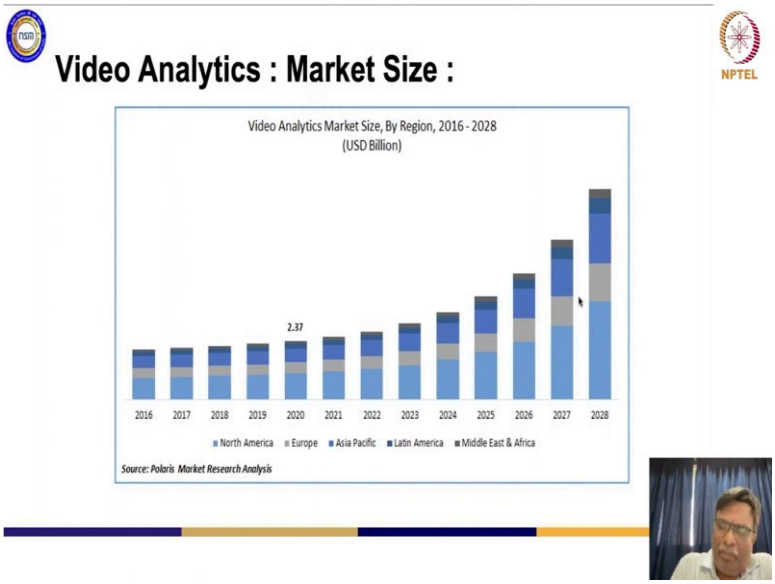
(Refer Slide Time: 11:28)



Now, when you are trying to manually monitor CCTV cameras, it is a very tedious job and monotonous job right and since it is monotonous and it is very very tedious cumbersome, it effectively reduces productivity.

And when you talk of automated surveillance and analytics, these are surely going to avoid errors which are caused by the reasons above and this is actually going to become an automated solution, which is to be efficient and there are ways by which we are supposed to be making these applications efficient right. So, this is the basic gist of why video analytics is gaining popularity today ok.

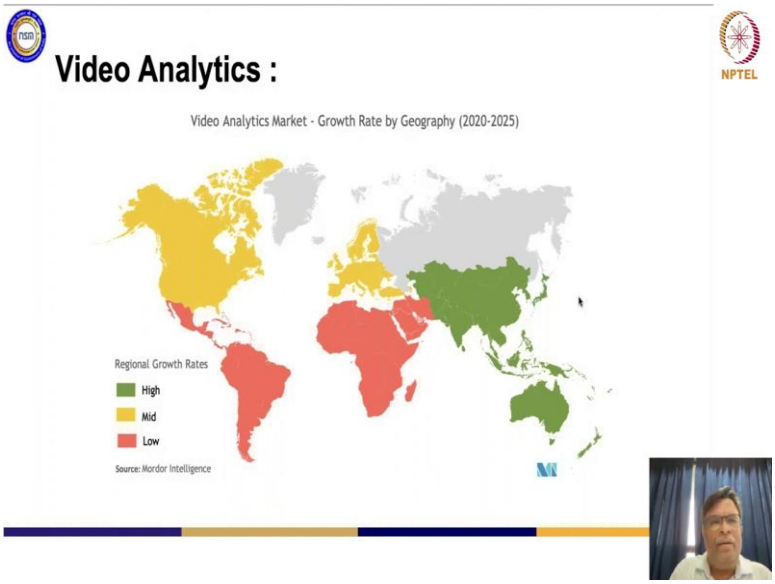
(Refer Slide Time: 12:26)



If you see the market size for people who want to do research in this area, this is basically a Polaris market research analysis report which talks of how video analytics right business is going to actually be growing right in terms of USD from 2016 to 2028. If you see this across various geographies from North America to Middle East and Africa, if you see by 2028, it is actually going to be right very very huge.

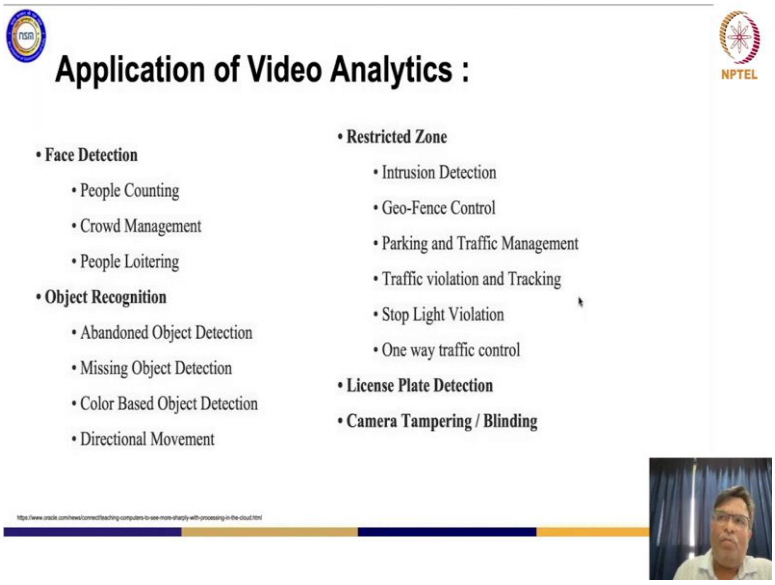
Now, at present, the market is somewhere around 2.37 us billion dollars right. But as it comes to 2028, it will be actually 10 times or 20 times more than this right. So, this is the actual area, where in a lot of concentration needs to be actually given ok.

(Refer Slide Time: 13:36)



If you see the video analytics market growth by geography from 2020 to 2025 right, one of the good things is India and Asian countries along with Australia right, all of this will have high regional growth rates ok as compared to some of the other geographies. So, this is where a lot of work would be done up till 2025 to 2026 or whatever; yeah. So, this is why we would be concentrating more on media analytics and the next thing is federated learning in the next two sessions right ok.

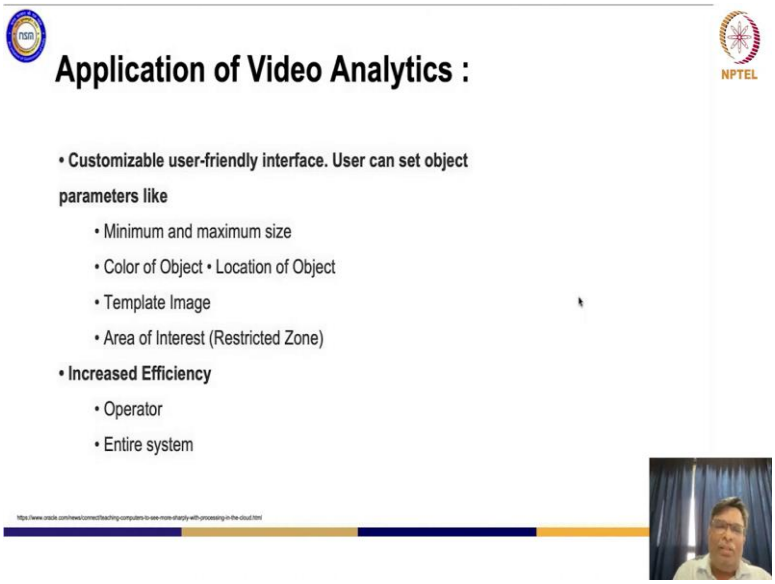
(Refer Slide Time: 14:24)



So, specifically, to go into the details of what type of applications are being developed using videos and then, it basically translates to something like video analytics, we are talking of face detection, people counting, crowd management right, people loitering; object recognition, abundant object detection, missing object detection, color based object detection, directional movements; restricted zones, analysis, intrusion detection, geo-fence control, parking traffic management, traffic violation and tracking, stop light violation, one way traffic control; license plate detection, a lot of work has been done on license plate detection; camera tampering and blinding.

So, these are certain things which we are trying to basically be developing applications for ok.

(Refer Slide Time: 15:27)

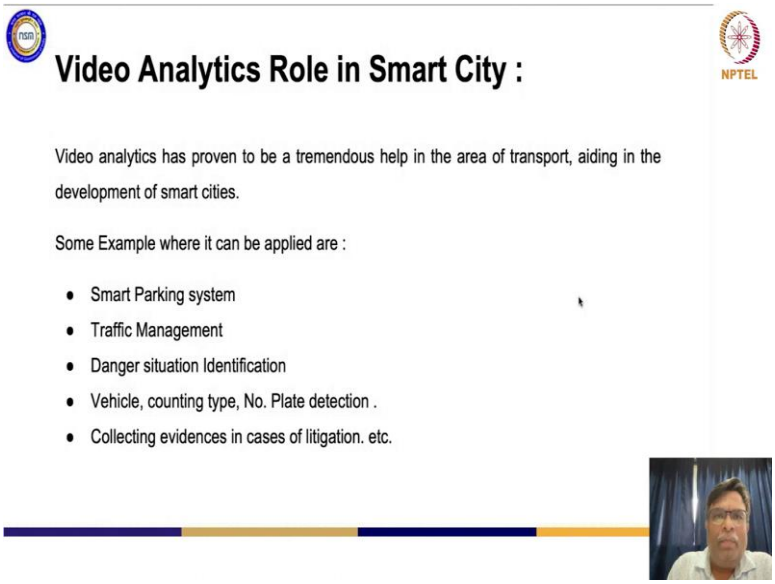


So, the one of the things is like when you are trying to develop certain video analytics application, you are trying to actually come up with a full-fledged application right and when you say it is basically having a full-fledged application which any person can use it, you need to have right customizable user friendly interface right. And user should be in a position to set object parameters right. We will see what are those. We will manually do one particular example and show it to you.

So, there is an option of user setting the parameters like minimum and maximum size, color of the object, location of the object, template image, areas of interest right as in case of developing applications for restricted zone ok and then, increased efficiency like

operators efficiency, how is it going to improve this and how is the entire system ok going to be more efficient in case you are going to give them a application ok. So, these are certain applications and certain best practices, which we should keep in mind, when we are trying to develop a video analytics application ok.

(Refer Slide Time: 16:52)

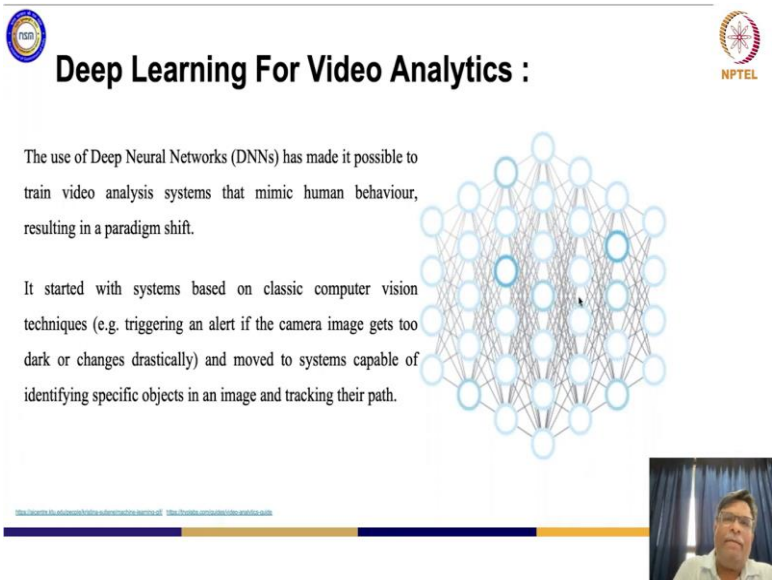


So, now coming down to the video analytics role in smart city which would be concentrating on, video analytics is of tremendous help right in basically transport management ok. In the sense not logistics, but transportation right and aiding in development of smart cities.

So, some of the examples where it can be applied are smart parking system, traffic management, vehicle counting, number plate detection and collecting evidences in case of litigation etc and so forth. So, if you see to some extent, if we go on our roads nowadays and you have fast tag compulsory basically implemented to be used right.

So, the basic idea is there it is actually trying to do a lot of things; it is trying to detect the vehicle, how much amount has to be deducted and it gets deducted. So, certain things like this right are going to improve the automation and that is how basically smart city is going to be actually of would be off right. Good, friendly, user friendly or people friendly applications and city ok.

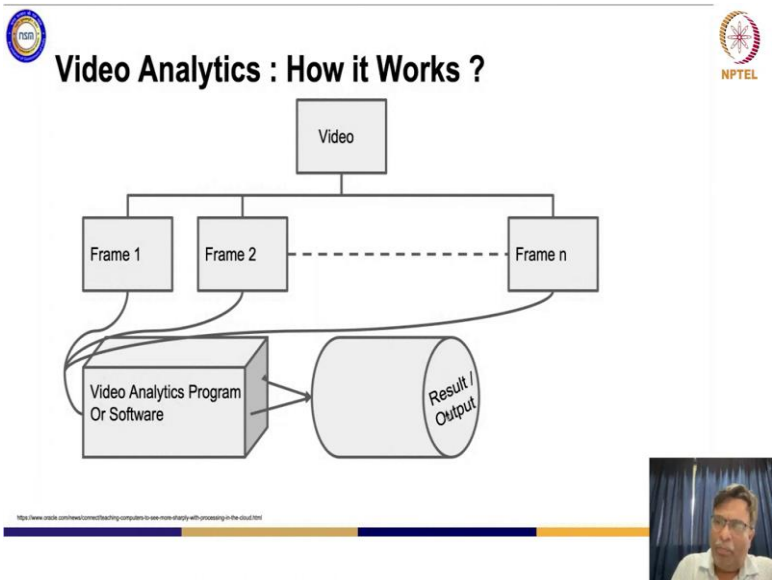
(Refer Slide Time: 18:20)



So, now, let us try to understand when do we use this deep learning for video analytics. Now, the use of deep neural networks has made it possible to train video analysis systems that mimic human behavior because you have got deep neural networks ok and because of this, there is a lot of paradigm shift in decision making process.

Now, if you see this way like when it all started right, it all started with classic computer vision techniques wherein write a camera image; basically image was taken and based on some thresholding and something like that right, it used to trigger an alert. Now, we have come to a stage, wherein we can actually specifically identify specific object and track their paths right, to this level we have reached. Now, for that, we will have to use deep learning neural network models or deep neural networks right ok.

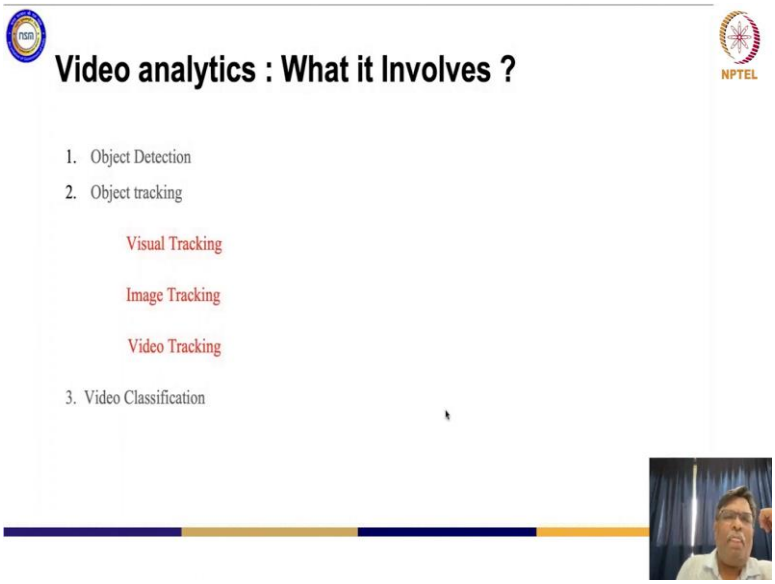
(Refer Slide Time: 19:21)



So, the basic idea is this. You are supposed to be developing a video analytics program or a software and this is a video. Now, what effectively happens is this video will be made up of n frames and all these n frames have to be processed right; maintaining the temporal and spatial connectivities among various frames and they have to be executed are processed in parallel ok.

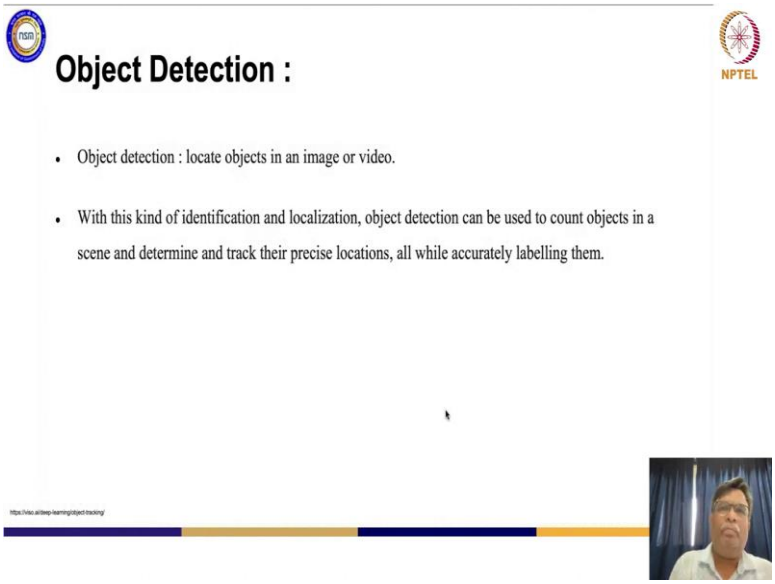
Keeping all these things in mind when you develop a video analytics program or a software, the result ok has to be as expected right. So, this is how the broadly ok; any video analytics program has to be developed ok.

(Refer Slide Time: 20:10)



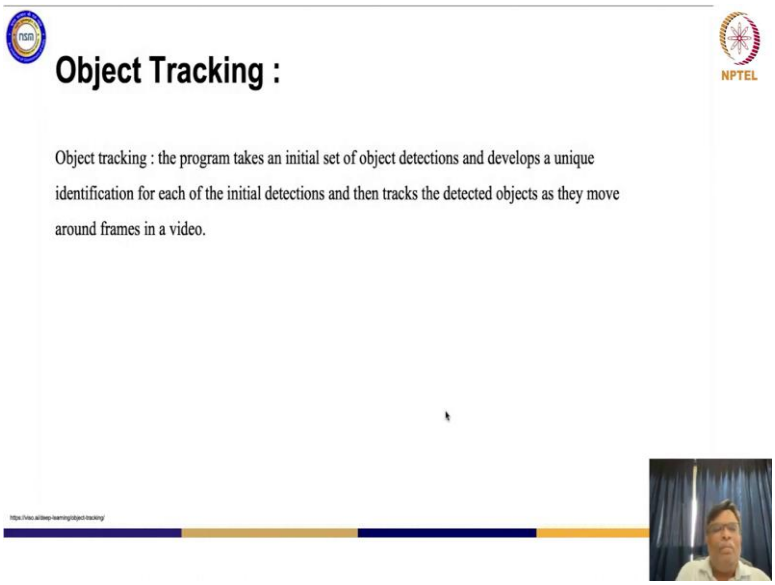
So, video analytics, it involves object detection; when you talk of object tracking, you have visual tracking, you have image tracking, you have video tracking and then, you have got video classification right. So, all of this needs to be done. You have to detect the object, then you will have to track the location of the object; keep on tracking it. There are various ways of doing it; visual tracking, you track the image and you can do video tracking as well and then, you have this video classification which has to be done sometimes ok.

(Refer Slide Time: 20:47)



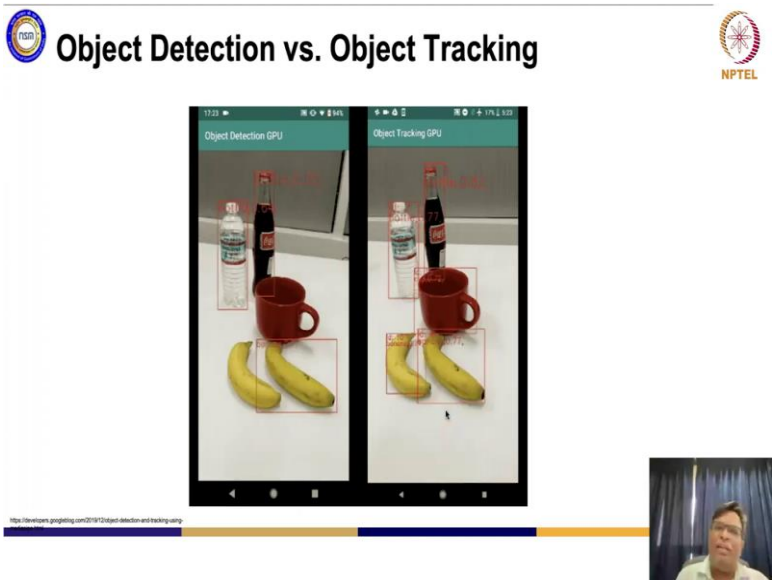
So, let us try to understand object detection. So, object detection basically is to locate objects in an image or a video. With this kind of identification and localization, object detection could be used to count objects in a scene and determine and track their precise location along with labeling right. So, you have to accurately label them along with tracking their precise location.

(Refer Slide Time: 21:17)



Object tracking; the basic idea is that program has to take an initial set of objects which are detected and then, you need to develop a unique identification for each of these initial detections. And then, track the detected objects as they move around the frames in a video right. So, it is something like you know trying to track how the object moves during the whole video right. So, this is how basically it is supposed to be done. So, you are supposed to do the object detection, you are supposed to do object tracking ok.

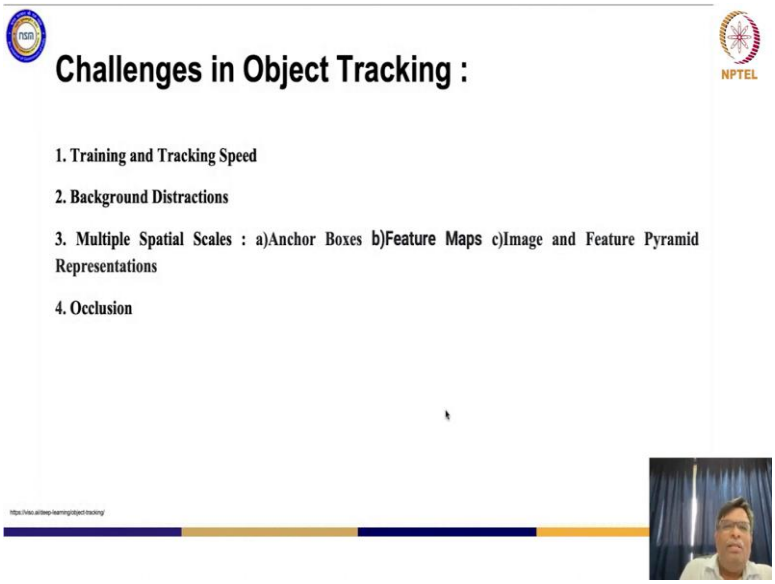
(Refer Slide Time: 21:51)



Now, this is how this is right. So, object detection versus object tracking right; what does it mean? If you see the difference, you are detecting it ok. And then, in the left video, you are just trying to detect it; but when you are trying to do the tracking right, you can see the difference right; it has to be actually trying to actually cover ok, once detected wherever that is going right or wherever is it move its moving right, it has to be tracked.

So, this is the basic difference right between object detection and the object tracking ok in a broader sense right. So, this is what is a basic difference about object detection as well as object tracking ok

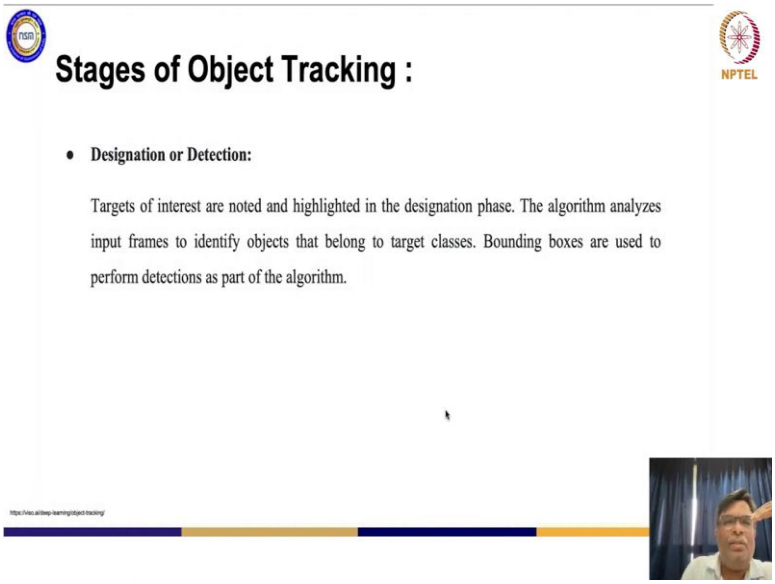
(Refer Slide Time: 22:47)



So, what are the challenges in object tracking actually? So, you are talking of training and tracking speed, at what speed you are supposed to be training and what speed you are supposed to be tracking, how are you going to minimize or deal with the background distractions which are there in the videos and you need to work with multiple spatial scales right.

You have to develop anchor boxes or you have to work with feature maps and then, you have to work with image and feature pyramid representations, along with right occlusion and all of this are various challenges in object tracking right. So, how are you going to actually think of the tracking thing, when the particular object gets secluded or something like that right. So, these are all certain challenges.

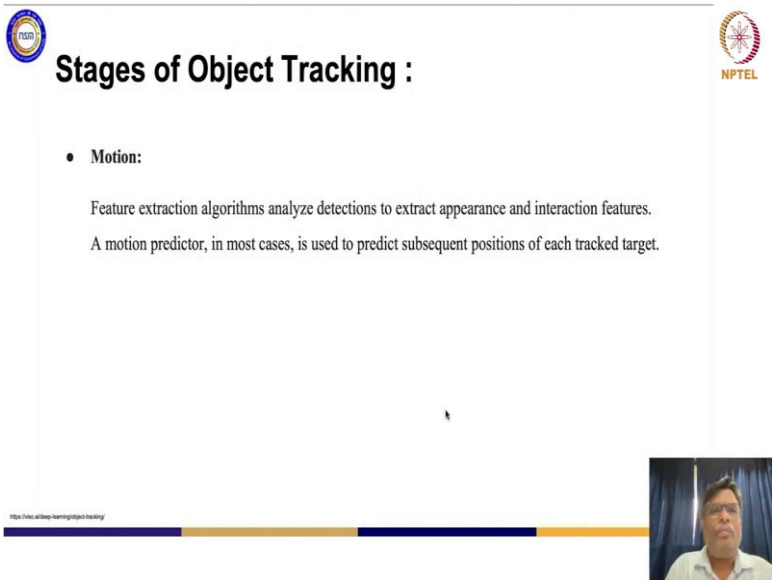
(Refer Slide Time: 23:48)



Now, stages of object detection is basically first is the designation or detection, wherein you talk of targets which are of interest and which basically could be highlighted ok in the designation phase. So, the idea is the algorithms are going to analyze your input frames to identify objects that belong to a target class.

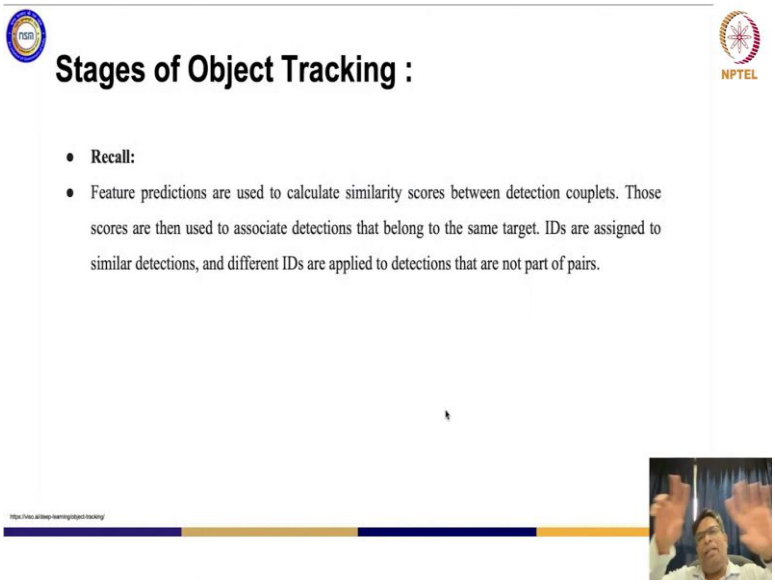
So, bounding boxes are used to perform detection as the part of the algorithm right. So, this is how basically you are going to develop in your algorithms, what we are trying to tell you in this slide is what is the way by which you can start thinking of actually developing such applications right.

(Refer Slide Time: 24:36)



Then, feature extraction algorithms are going to analyze these detected patterns to extract appearances and interaction features and then, you should use a motion predictor ok to predict the subsequent position of each track target right. So, this is what actually you are going to; sorry, when you are going to actually do ok some prediction in terms of a moving object right, you needed to predict ok and then, you need to track.

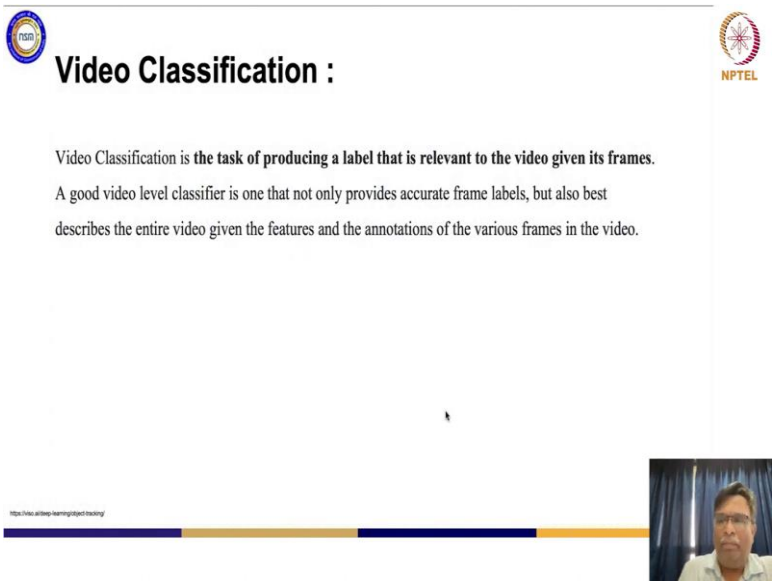
(Refer Slide Time: 25:15)



And then, when you talk of doing all this, there is something which is called as a recall stage. Recall stage basically talks of situations, wherein the features which are predicted

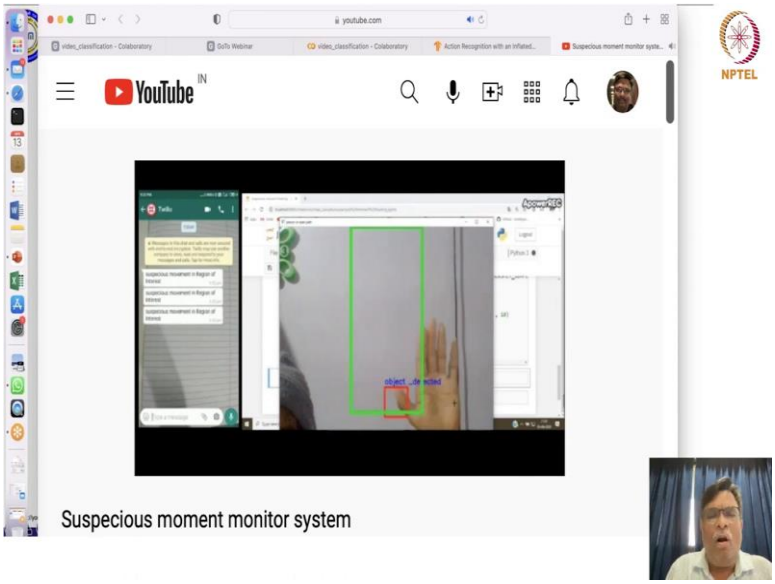
or feature predictions are used to calculate similarity scores right and these scores are then used to associate detections that belong to the same targets and then, you define or assign certain IDs to them and you will also ensure that the different IDs which are applied to those detections if they are different, they are not a part of pairs or they are not of a similar type or whatever right.

(Refer Slide Time: 26:00)



And then, video classification in the end is the task of producing a label that is relevant to the video given its frames. So, good level, good video level classifier for that matter is one that not only provides a accurate frame labels; but also best describes the entire video given the features and the annotation of the various frames in the video right. So, this is how we are going to actually do things ok.

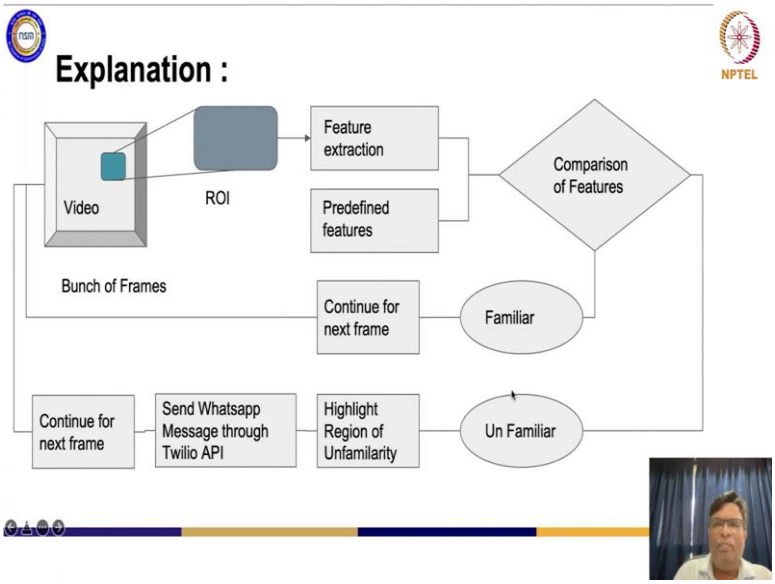
(Refer Slide Time: 26:32)



So, this suspicious movement in the region of interest alert system is one example. So, let me ok. So, this basic thing was actually a suspicious video monitoring system which we of course had not developed; but we will share the link with you and you can basically try to see how basically some activities getting detected and majority of these things are being used nowadays, when you talk of basically like online examination system is being implemented right which detects.

So, all these prometric centers which use these online examination software's right, they are able to detect all this. So, this link is there which we will share and this is just a specious movement monitor system ok. So, this let me just close and yeah.

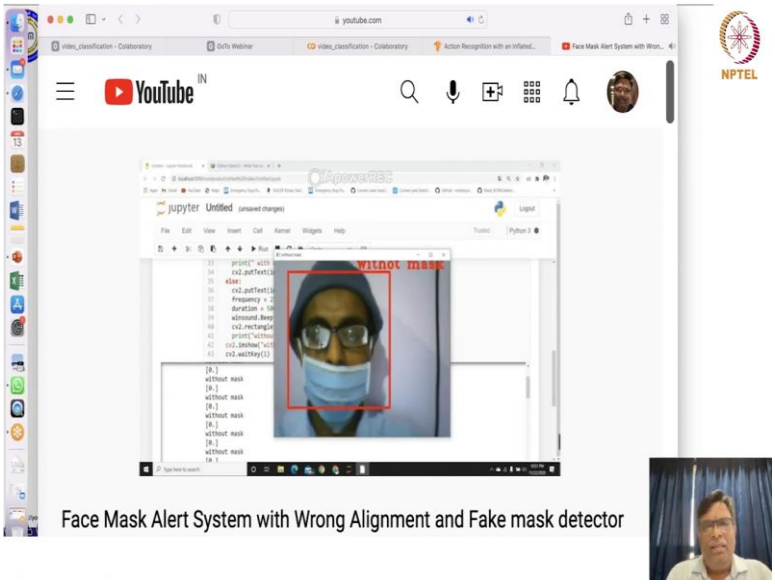
(Refer Slide Time: 27:46)



So, the next thing is trying to understand how we develop such systems. So, the idea is whatever we have discussed till now, we have this explanation of this total thing of total gist of whatever we have discussed, we have this video and inside this video, the bunch of frames. Then, you have to find out the region of interest, you do feature extraction, then you have predetermined features or predefined features.

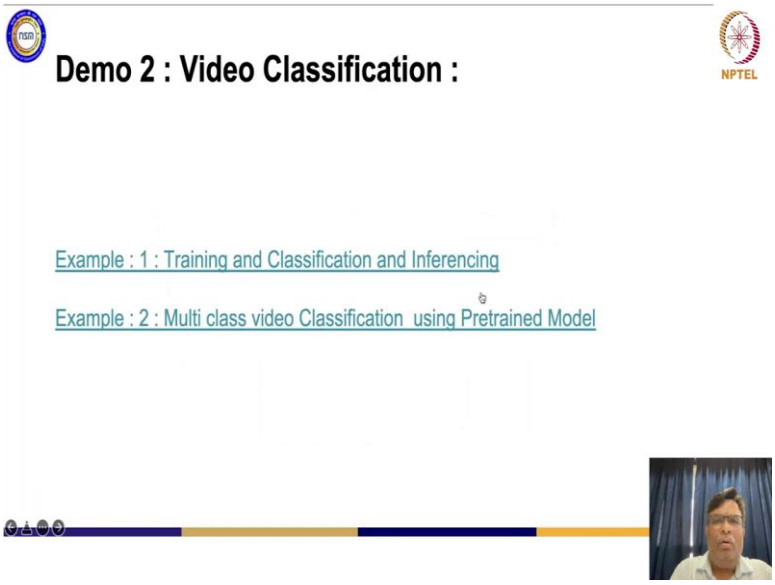
Then, you compare and if they are familiar, then you basically go on doing iterating it. If they are unfamiliar, then how is that it has to be detected and then, you can basically send a WhatsApp message or something right and continue to the next frame or something like this. So, this is how basically the thing is.

(Refer Slide Time: 28:38)



Now, this is again mask alert system which talks of trying to find out how basically right is without a mask or with a mask right. So, this is basically something wherein we have done this and this basically is right sometimes wrong ok; yeah. So, something like this right. So, if you remove the mask your nose. So, it is telling without mask and then with mask or something like that right. So, this is how basically we have been able to do some things on this. So, let me just close this as well; then ok.

(Refer Slide Time: 29:32)



So, let us try to understand.