Mine Automation and Data Analytics

Prof. Radhakanta Koner

Department of Mining Engineering

IIT (ISM) Dhanbad

Week-6

Lecture-26

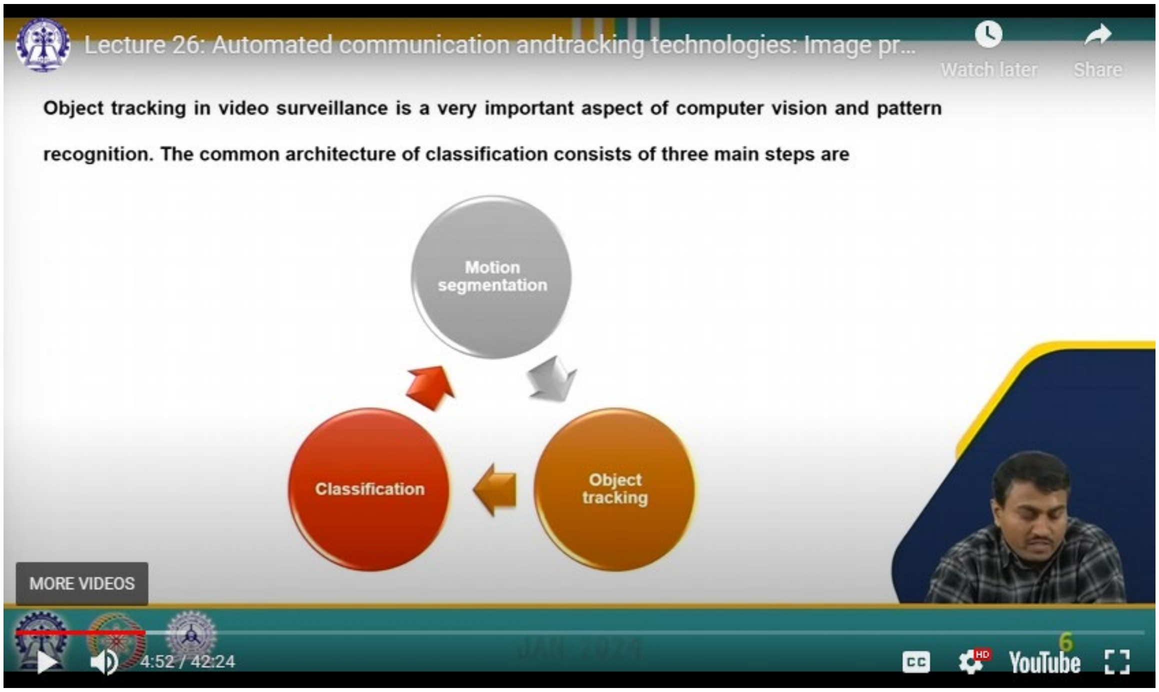
Automated communication and tracking technologies: Image processing

Welcome back to my course mine automation and data analytics. Today, in this lesson, we will discuss on image tracking and communication. This is very important topic in the context of automation in the mining and we are looking into the application of this module for tracking the miners posture, miners movement into the sensitivity area so that we can maintain safety. So, these things can be done automatically using the image and image processing and that is what we are going to discuss in this module in details.

So, in this lesson, we will discuss the following. We will introduce about the automated communication and tracking technology. Basically, it is the different parts of this technology, and we will discuss with a case study the real-time object detection and tracking for unmanned aerial vehicles based on convolution neural networks.

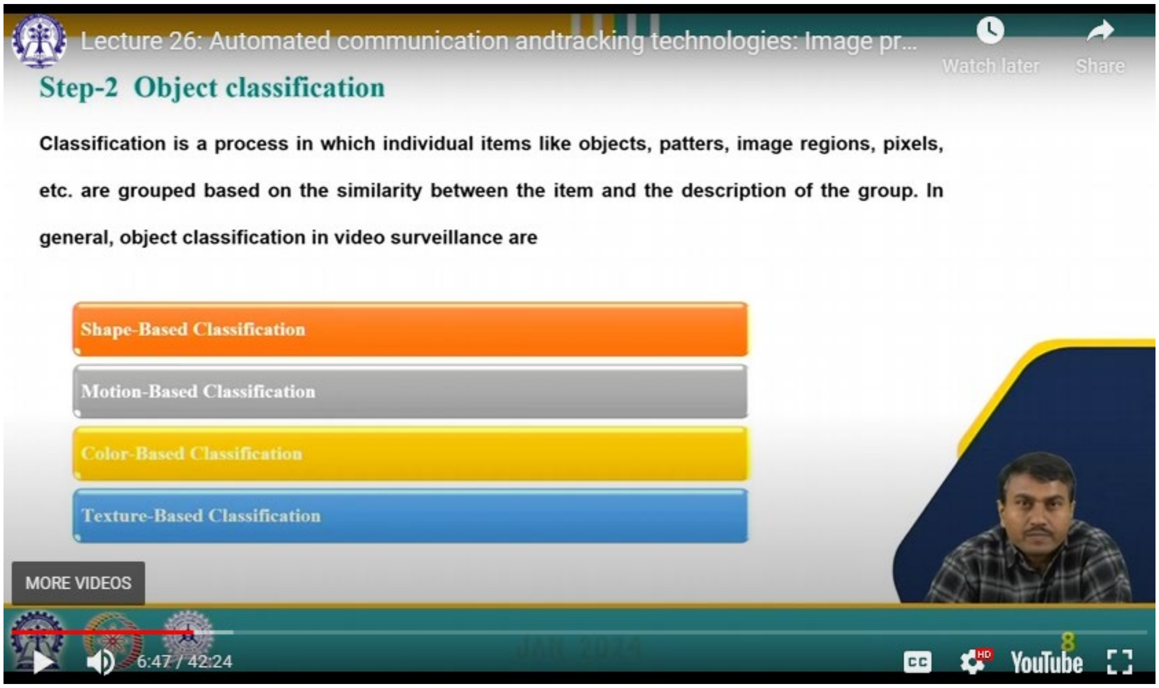
So, what is automated communication and tracking technology? Image processing is a form of processing with input as image such as photograph or video frame and output can be characteristics or parameter related to the image. So, for example if video image was there and miners is working so we are basically interested to know the posture of working, the speed of working and the miners movement into the sensitive area. So, the, computer vision is an area that consists of methods for incorporating analysing and visualizing images. So, surveillance stands for monitoring the behaviour activities and other changing information usually of people for the purpose of influencing, directing and protecting them, and this is very essential into the context of automated mining and, where we require very efficient supervision and control over the miners movement. So, the process of locating moving object using a camera is basically called video tracking. So, in simple terms, tracking means associated target objects in consecutive video frame, and difficulty arise especially when objects are moving rapidly as compared to the frame rate. Frame rate is basically the hardware specification of the camera of the of the hardware sensor or when the track object change direction over time. A sequential flow of object detection object tracking object identification and its behavior completes the process framework of tracking.

So here we have four parts: object detection, object tracking, object identification, and its behavior. So, object tracking in video surveillance is a very important aspect of computer vision and pattern recognition. The common architecture of classification consists of three main steps. One is the motion segmentation very important then the object tracking which object we want to track accurately we are pinpointing we are locating that object and then classification of that particular object. So, these three are run in one by one and that makes surveillance more effective.



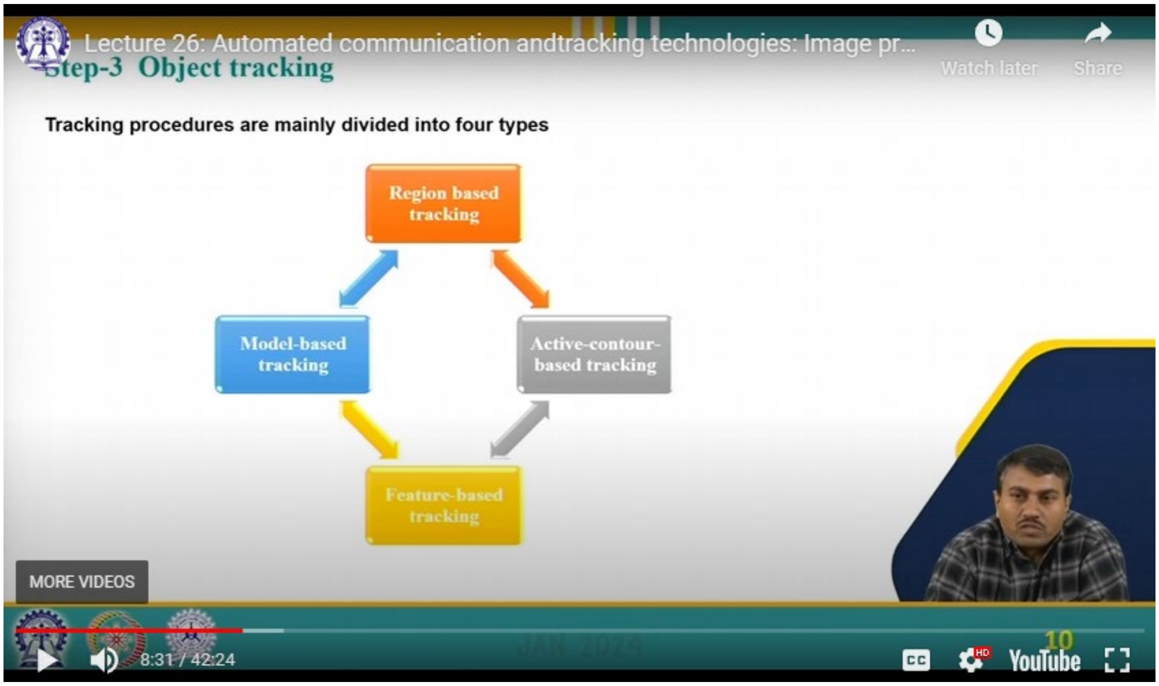
So, what is motion segmentation? Object detection is a computer vision technology that deals with identifying instances of object such as human, vehicle, animals or, bird or other moving objects. So, object detection is one of the initial steps of object tracking. So, the object we want to track we have to first detect the object. So, a video surveillance system for stationary camera generally includes some part of the motion detection.

So here involving the background subtraction because in the background too many other information might be there so we want to segregate that to identify yes this is the particular object this is the particular object motion I want to track. Then, the temporal differencing over time the change and the optical flow of that particular picture and the object. Step 2: object classification. Classification is a process in which individual item like object, patterns, object region, pixels are grouped based on the similarity between the item and the description of the group. So, in general objects classification in video surveillance are shape based classification, second is the motion based classification and third is color based classification and fourth is texture based classification.



Object tracking. In simple terms, tracking is the problem of estimating the trajectory of an object in the image plane as it moves around the scene. Depending upon tracking domain methods and algorithms, a tracker can also provide object-centric information like the area, orientation and shape of an object. So once objects are detected, the next task in the video surveillance process is to track the object from one frame to another frame. So, tracking objects can be complex due to complex object shape, object motion, non-rigid nature of the object, scene illumination changes over the path and the time partial or full object contour occultation, etc.

Object tracking. So, tracking procedure are mainly subdivided into four component. One is the region-based tracking, second is active contour-based tracking and fourth is the feature based tracking and the last one is the model based tracking.



So, these fours are comprises of the object tracking. So, there are many algorithms and modules available as open source. So those modules are incorporated these features that we have already discussed. So now we are going to discuss and a case study that utilize these algorithms and specific to tracking the movement of people in indoor scene as well as outdoor scenes. So, this experiment was conducted to see the efficiency of this kind of system how efficiently we can track the object and very important thing this case study has highlighted the requirement of a computational power. We can understand that in a mine environment or, a workshop environment or, a factory environment or a processing plant environment, we may not install we may not be able to install the high-level workstation. A low-level computational power might be available there.

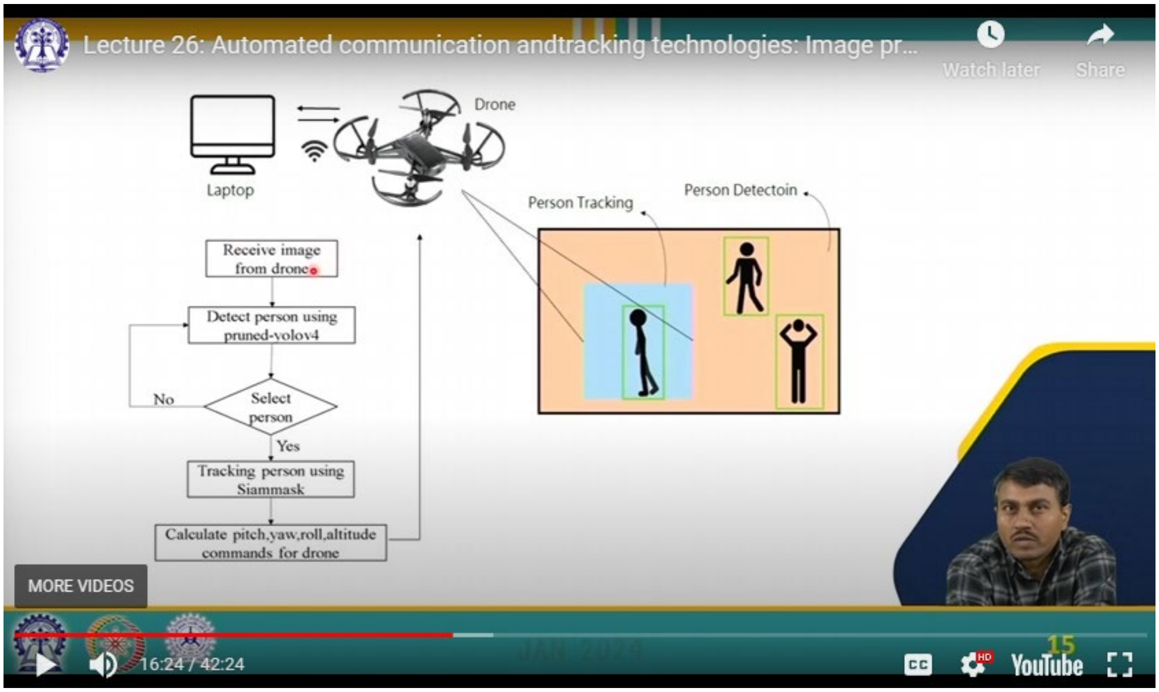
So, this particular case study has been highlighted that with a low computational power computer the tracking can be done very efficiently with very high level or desired level of accuracy that is basically the intention and target of this case study. So, let us deliver on this case study in a brief. So, this particular study is concentrated on object the object is here person human person. So, this particular study has utilised the open source-based framework and software that is a robot operating system to implement image detection and tracking for controlling the UAV, and here I want to mention that these ROS is very much compatible to integrate with different environment and different sensing system that basically requires certification of these UAV maneuverance and so and so forth. The hardware required here is laptop a very lightweight laptop and low computational power laptop that has been used and all the models are lightweight, and this particular model have also benchmark and compare with high-level models and the prune model that this particular module has used and seen then the efficiency between these two and based on different parameters four parameters it has identified it has found that yes with less computational power this kind of tracking can be efficient and can efficiently be maneuver in indoor and outdoor environment.

So, for the object detector, train a convolution neural network based on the yellow four version architecture. So here this particular algorithms is based on the yellow version four and further in this case study they pruned it so they have used the pruned yellow version four. So, this study employ the pruned version of the yellow four version four object detector and the CM mask monocular object tracker to detect and track the target person captured by the camera of the drone. So, the object detection system consists of four main components. One is the object detection identifying the object as human yes it is a human and this is the window where human is detected.

Second is the target tracking there was a movement and there was a movement of the object so this particular different frame that compare and attract the movement in direction over a spatial plane and the proportional integral derivative control we will basically deal with in details about this particular control in SCADA framework. So, this is basically this control have proportional means linear then derivative change rate of change and then integral so all three controls together it is basically the very high level of control that is PID control on the drone movement and the UAV driver package. So, these are the four component of this particular case study. So, this study utilize a very consumer grade drone that is Tello-drone for implementing the object detection and tracking system and this is easily available in the market and during the tracking process the UAV control parameters include the roll that is lateral pitch this clockwise anticlause yaw that is forward backward and the altitude the vertical movement. So, all of which are control using the PID control that we have already dealt with.

So, this PID controller takes the position and distance of the target object as input then based on the next image is identified center of the image and the tracked. So, is there any difference? So, they ensure the difference is minimal or nears to zero that basically the target of this particular model. So, the position and distance are calculated using the monocular front facing camera of the UAV. So, in this particular framework, we will discuss the object detection model pruning and visual tracking. So here I want to discuss in a little that the model pruning concept. So, you can understand that when drone is basically sending data and here in this particular case study we have used 30 Hertz data sending rate.

So, there may be some unnecessary information that we are basically transferring through the Wi-Fi on the network, and this YOLO algorithm prune version YOLO algorithm have identified the channel number of channels in the image and the object and the highlighted channel that is required more importance and the low highlighted channels are to be pruned so that data bandwidth is reduced computational power required is reduced. So, this particular model is very much highlighted on the pruned concept or pruned model application for identifying object and tracking the object. System setup so a laptop computer communicates with the tallo drone via the Wi-Fi network and drone transmit image at a constant frequency of 30 Hertz process using a pruned version of the YOLO 4 algorithm for object detection. So, this is basically the framework and the hardware part of this particular system. So, you can see it is a drone that is connected with the laptop okay and a drone sending the image at 30 Hertz rate okay.



So now laptop receive the image from the drone okay. Now it will try to detect the person using the pruned YOLO version 4. So, if it is detected, then select a person, and if it is not able to, then again redo the run the model YOLO 4, then again select the person if it is yes. Now, tracking the person using the CM mask and based on that now, calculate the pitch your role and altitude and command the drone accordingly. So once the object is detected once the object is detected now we will further precise the center of the object and the object detected whether it is matched precisely and the error is zero so that basically the work of the PID controller.

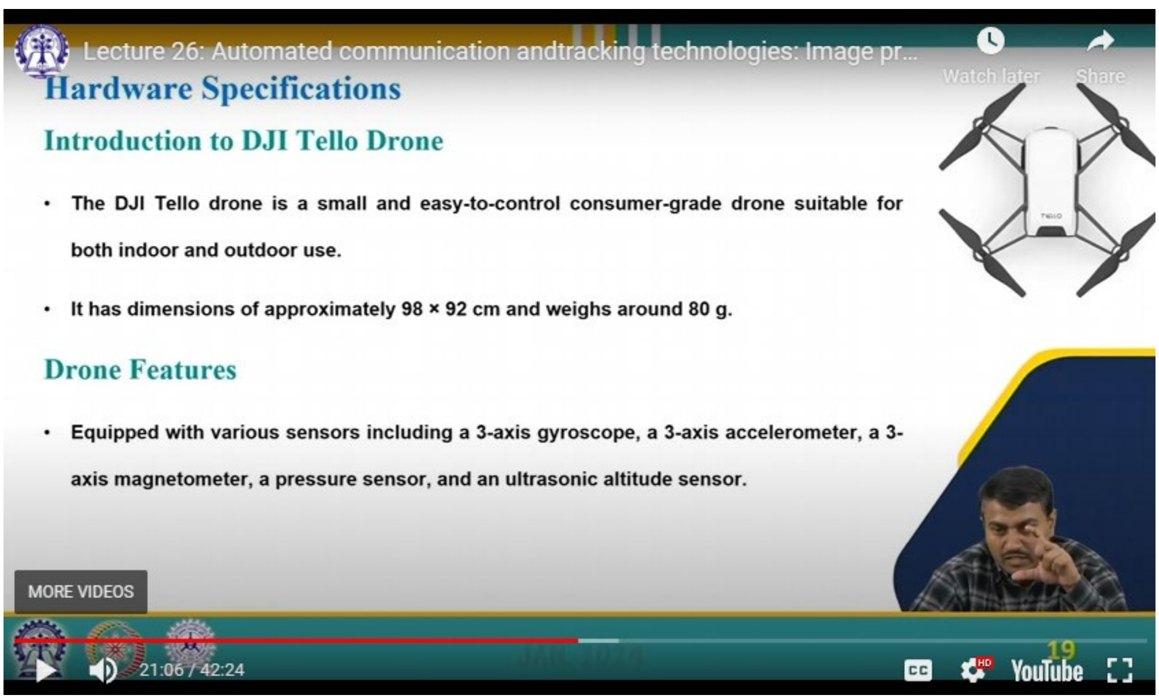
So, on that drone forward, backward movement, lateral movement, vertical up movement, and the clockwise anticlockwise movement are basically controlled to make object detection and tracking more and more and more accurate. So here is it between basically the interest area this is the initially this is the box now it has further concentrated on this area that is the green box object is identified person has been tracked so person been after detected person been tracked the movement okay. So, first of all, we have to detect the person, then the question of tracking. So, for that, during the tracking, you need the PID controller. So, user can select the bounding box that green bounding box I have already shown based on their requirement for object detection.

So, the pruned YOLO version 4 is utilised for person detection with a detected bounding box displayed on the screen. Object tracking with CM mask so this system you utilize the CM maze network CM mask for object tracking. A tracking algorithm based on a PID controller estimate the roll pitch yaw and altitude based on the track object position and distance. User introduction and object selection so user can select a specific object of interest by clicking on its bounding box as when required. So, the system extract the person within the bounding box as a template frame for the CM mask network to enable subsequent tracking.

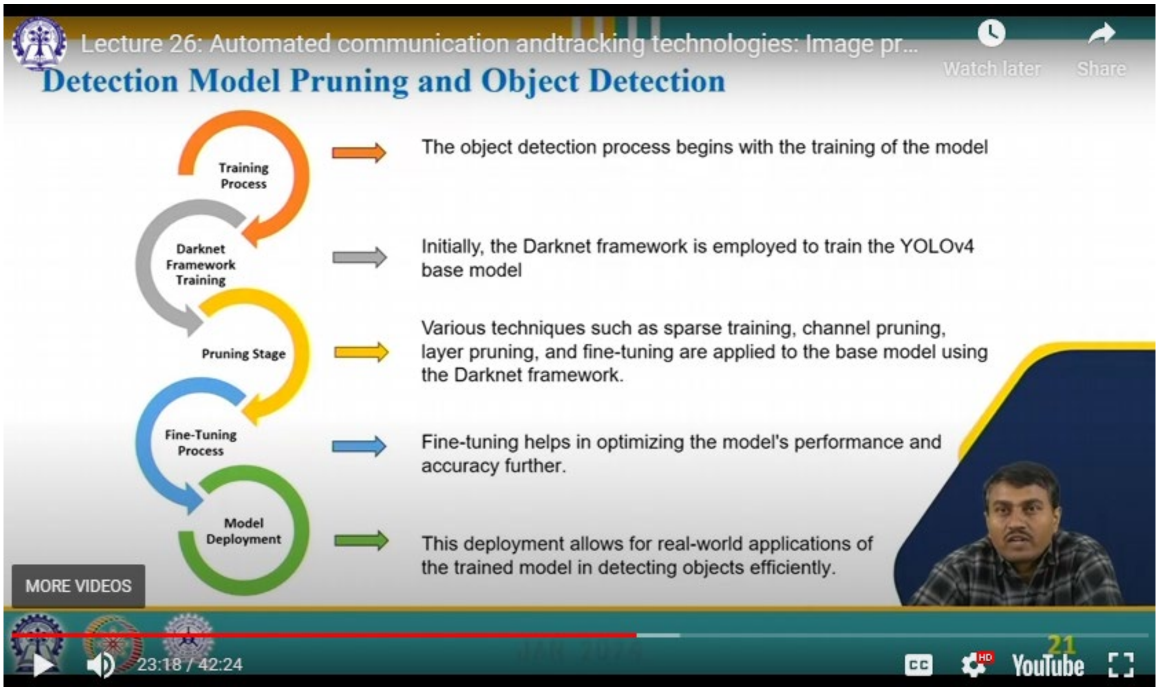
Flight command generation so the tracking algorithm now calculates the error between the target and the center of the frame target and center of the frame target and the center of frame. So now, here is the work of the PID. So, the errors are if the error is more now the PID has to be fast so the it basically works on the error so these errors are as the input of the PID controller to generate flight commands for the yaw roll pitch and altitude adjustment.

Handling no target detection and very interesting if no target is detected drone maintain its own position until a target appears in the image field it will maintain the same position it will not change PID controller is not active now because no object is detected. So, this ensures stability and prevent unnecessary movements when no object of interest is present.

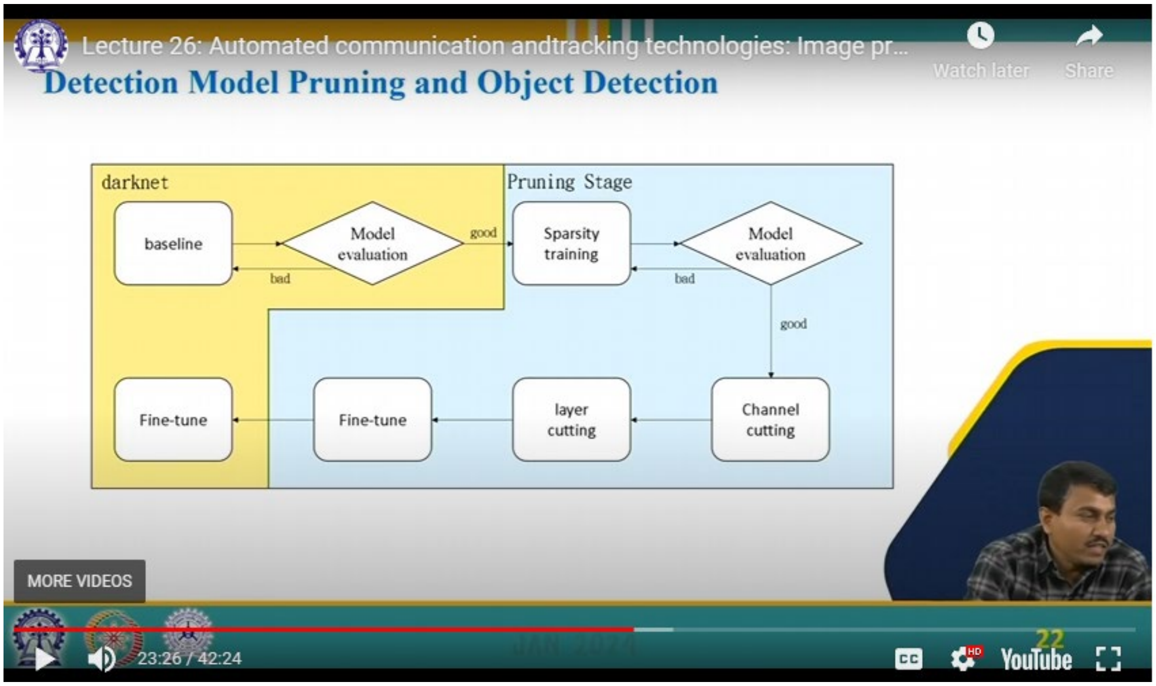
Harder specification so this is the trailer drone this is the trailer drone very lightweight and consumer grade drone available easily in the market manufactured by DJI and this is a small easy to control consumer grade suitable for both indoor and outdoor environment and dimension is 98 by 92 centimeter and weight around 80 grams sorry. Drone features equipped with various sensors including a three-axis gyroscope a three axis accelerometer a three axis magnetometer a pressure sensor and ultrasonic altitude sensor these basically maintains a calm measure the height very precisely that is why these ultrasonic altitude sensor is used. Features a front-facing camera with a resolution of 1280*720 capable of capturing video at 30 frames per second FPS is 30. Communication and connectivity the Taylor drone can communicate with other device, such as smartphone or laptop via a Wi-Fi network so in this particular study a PC was used for communication with the Tello drone and this setup allow for data exchange and control of the drone function during experiment or applications. Detection model pruning and object detection so first there is a training process it involves the object detection process begin with the training of the model.



In-situ image is coming continuously so first there were with training process second is the darknet framing framework training so it is initially the darknet framework is employed to maintain and train the yellow stable model so that unnecessary parameter is highlighted hyper parameter is fine-tuned and an image is sent to the next level of processing. Then pruning stage, various techniques such as sparse training channel pruning, layer pruning, and fine tuning are applied to the base model using the darknet framework. Fine-tuning process fine-tuning helps to optimise the model performance and further increase the accuracy and model deployment so this model is now ready for deployment in a real-world applications for detecting object efficiently.

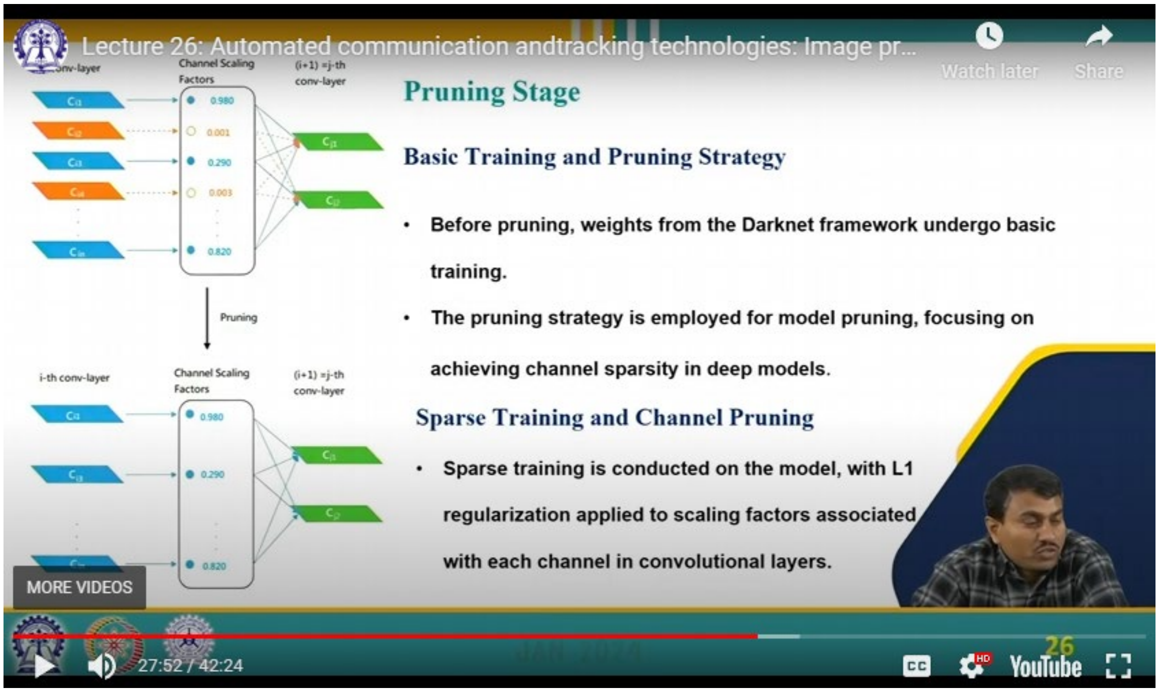


So, this is basically the framework here. initially the image is gone through the darknet framework, so based on the bad, they will reject, and then model evaluation is done for the good image, it will go to the pruning stage so first, it will do the sparsity training then again model will evaluate for the good image it will prune the channel number of channels unimportant channel will be pruned then again unimportant layers will be pruned further it would be fine-tuned hyper parameter will be fine-tuned, and again it will be fine-tuning using the darknet framework.



Darknet training so this darknet framework is utilised for training the yellow version 4 model with adjustment made to various hyper parameter to enhance accuracy and the performance. Adjusting input size is one crucial hyper parameter adjusted in the input size of the network impacting the model ability to detect small object increasing the input size aid in detecting small object but may slow down inference speed and consume more GPU memory. So yellow version 4 network down sample the input size by a factor of 32 necessary input width and height to be multiplied of 32. So, in this study, the input size is set to 416 by 416 to ensure compatibility with the network. Batch size and subdivision the batch size and subdivision hyper parameter are adjusted based on the GPU performance. So here batch size represent the number of image loading during training typically set to 64 and if GPU memory is insufficient each batch is subdivided into smaller sub batch to fit into the memory.

In this study the batch size set to 64 and subdivision set to 8 optimizing the GPU memory usage. Number of iteration training in the darknet framework is measured in iteration not in epoch so each object class should ideally have at least 2,000 iteration for effective training. So, with only one class the number of iteration is set to 2200 to ensure sufficient training for higher accuracy and performance. Pruning stage, so here you can see in the convolution layer, the channel scale factor is given Cij, Ci2, Ci1, Ci2, Ci3, Ci4 and Cin, and there is a channel scaling factor of 0.98, 0.001, 0.29, 0.003 and 0.820. So, these are less important channels because for scaling factor is very less, so it basically pruned that now on the convolution i plus 1th layer it is Cij as j1 and j2, so after pruning it kept only these three 1 2 then the convolution layer Cj1 and j2 so requirement of the computational power is reduced by removing these channels. So before pruning wait from the darknet framework undergo basic training the pruning strategy is employed for model pruning focusing on achieving channel sparsity in deep models sparse training and channel pruning sparse training is conducted on the model with L1 regularization applied to scaling factor associated with each channel in convolution neural network and layers.



This regularization helps identify unimportant channel which are subsequently pruned based on their scaling factor values compact model generation. After pruning a compact model is obtained potentially sacrificing some less important channel for reduced model size and complexity. So, this compact model is then fine-tuned to achieved comparable or even higher accuracy compared to the fully trained network.

Sparsity training here, a batch normalization BN layer is added after each convolution layer in yellow version 4 to expedite convergence and enhance generalization. Normalization process the BN layer normalizes convolution features using batch statistics represented by the equation y is equal to gamma into x minus x bar divided by sigma square epsilon plus beta. Here x bar and the sigma square is represent the mean and variance of the input feature in mini batch respectively. Gamma and beta represent the trainable scale factor and bias in the batch normalization program. Indicator of channel importance so scale factor gamma in the BN layer is utilized as an indicator of channel performance and L1 regularization is applied to gamma to facilitate channel level sparse training distinguishing between important and unimportant channel efficiently.

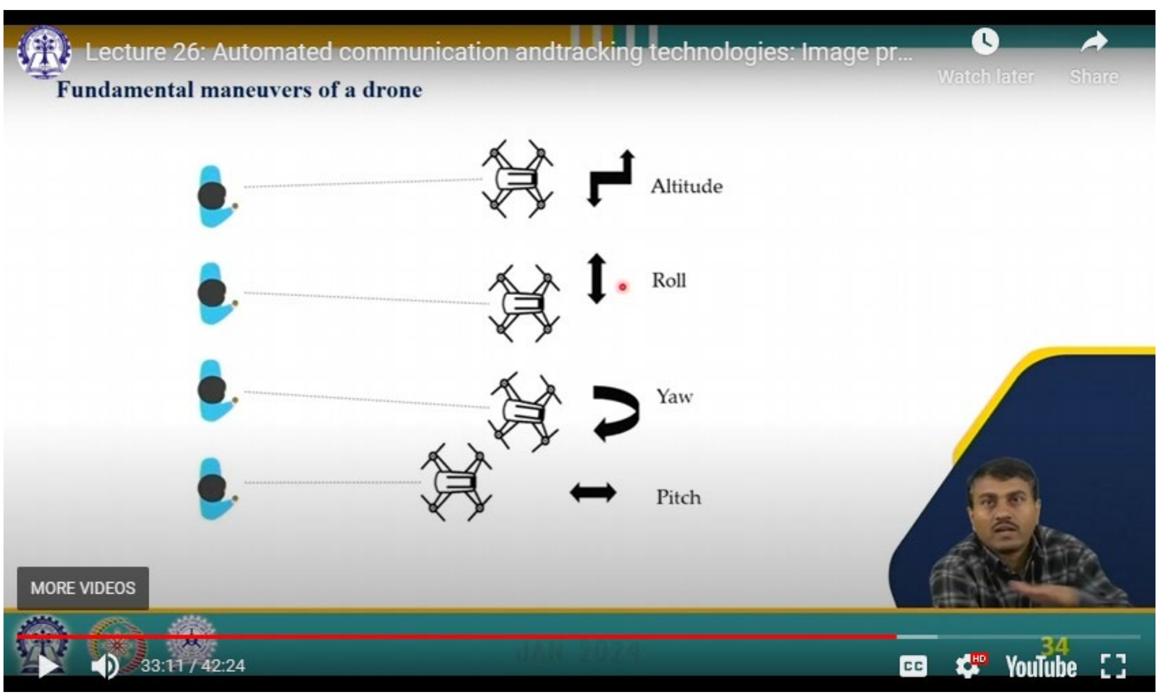
Sparse training loss function the loss function for sparse training incorporates L1 regularization on gamma is expressed as loss yellow plus alpha into summation of function of gamma. So here multiplier penalty function is used alpha and that ensure that minimum loss in the program. So, if gamma represent the L1 normalization applied to gamma and alpha is the penalty factor balancing the two-loss term. Benefits of sparse training. Pruning effectiveness relies on the sparsity of the model.

Sparse training compress most gamma value in the BN layer towards zero leading to two benefits. One is improved model efficiency through network pruning and compression reducing computational complexity. Second identification and pruning of parameters with minimal impact on network performance by sparsifying weights close to zero. Channel cutting following the completion of sparse training the process of channel cutting is initiated to further optimize the model. So total number of channel in the backbone is computed to establish the basis for the model for the channel cutting.

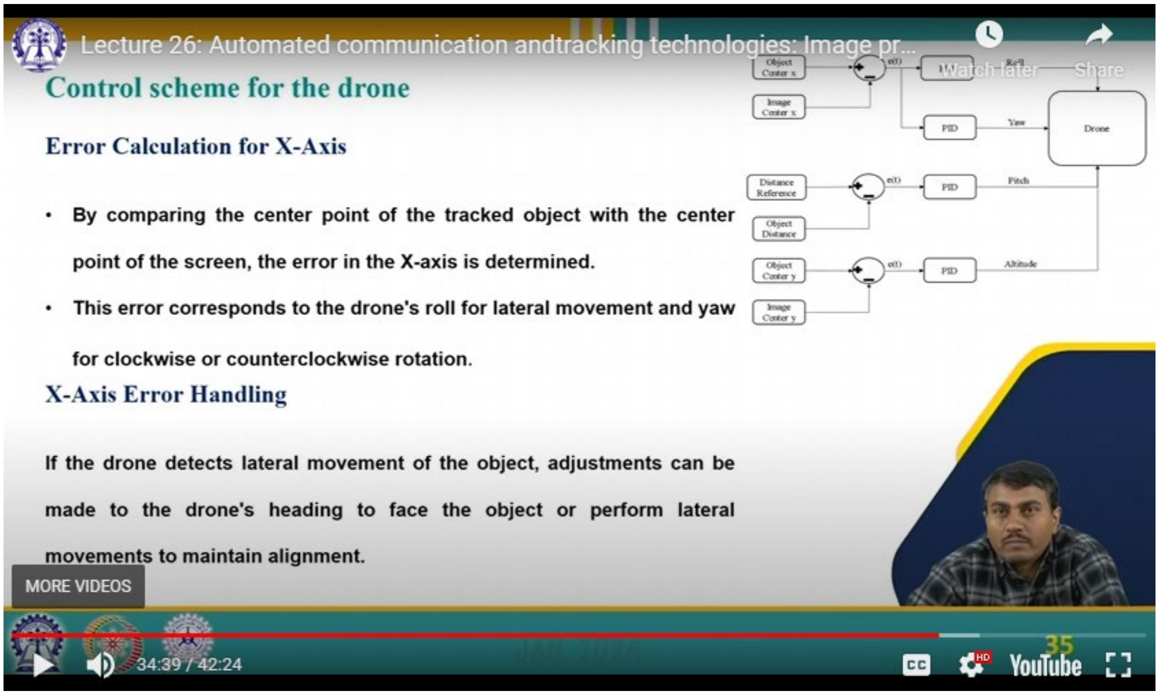
Corresponding gamma value are extracted and stored in a variable then sorted in ascending order. The decision of which channel to retain and which to prune is based on the predefined pruning rate. So, the pruning rate, typically values between 0 and 1, determines the proportion of channels to be pruned. The higher pruning rate signifies a greater degree of pruning. Fine-tuning necessity of fine-tuning in case where pruning adversely affect model accuracy fine tuning becomes essential to restore the prune model accuracy.

So, fine-tuning plays a crucial role in mitigating accuracy loss caused by pruning, thereby enhancing the overall performance of the pruned model. Importance of fine-tuning fine-tuning allow for adjustment of model parameter and optimization of the prune model for improved accuracy. It ensured that the prune model maintains its effectiveness in performing its intended tasks such as object detection experimental approach. So, in the conducted experiment the pruned YOLO version 4 models was retrained using the same training hyper parameter as the normal training process for YOLO version 4. So, this approach ensure consistency and facilitate comparison between the prune and the original model enabling a comprehensive evaluation of their performance.

Fundamental maneuver of the drone. So here are the direction so this is the altitude height change control by the PID. Roll is basically lateral movement, ER is the clockwise or anticlockwise movement, and the pitch is the forward and backward movement of the drone.

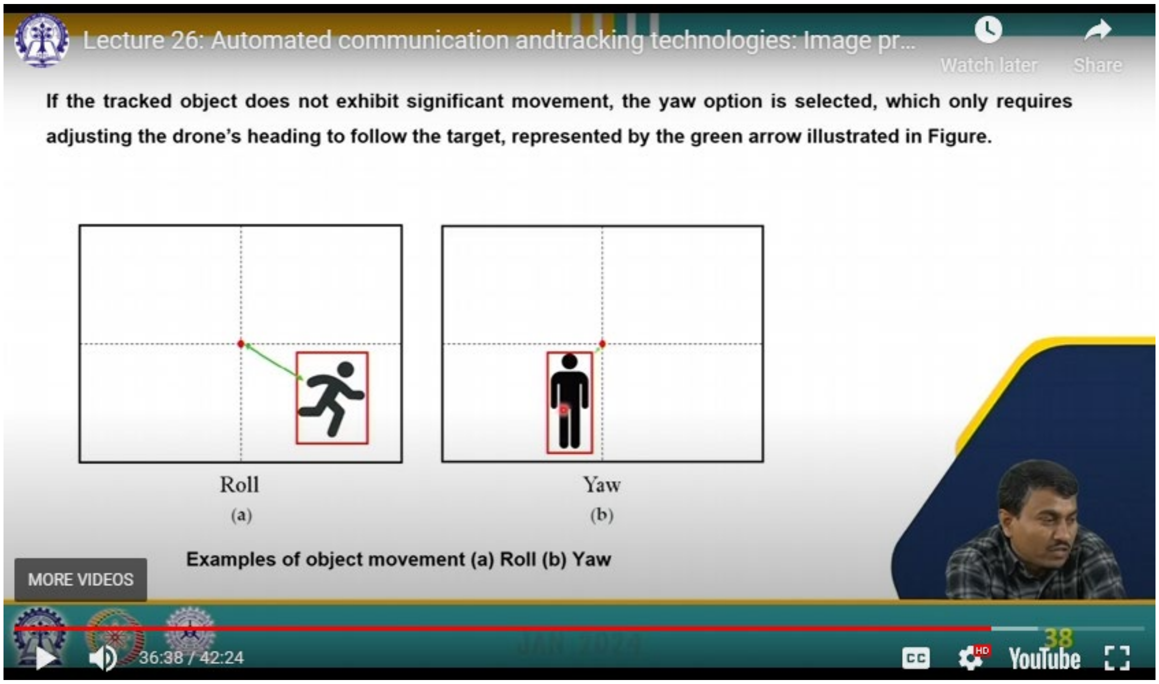


Control scheme for the drone so here you can see PID is for the object center X location object X and the image center so it estimate the error and apply the PID then it is basically apply the PID on roll and ER on the drone. Now the object reference object distance again there are some amount of error it applied PID on the pitch on the drone now it is applied to the drone then object center and the image center Y so it is basically on the altitude there are some error so PID again applied and it is applied to the drone. So, all four is taken together and PID control is now active.



So, error calculation for X axis by comparing the center point of the track object with the center point of the screen the error in the X axis is determined. This error corresponds to the drone roll for lateral movement and ER for the clockwise or counterclockwise rotation. X axis error handling if the drone detect lateral movement of the object adjustment can be made to the drone heading to face the object or perform lateral movement to maintain alignment. Error calculation of the pitch axis: the pitch axis involves forward and backward movement by comparing the distance between the drone and the real object. The desired ideal distance, the distance error, is calculated to control the drone's forward or backward movement accordingly. Pitch error handling so adjustment are made based on the calculated distance error to control the drone forward and backward movement and maintain desired proximity to the track object.

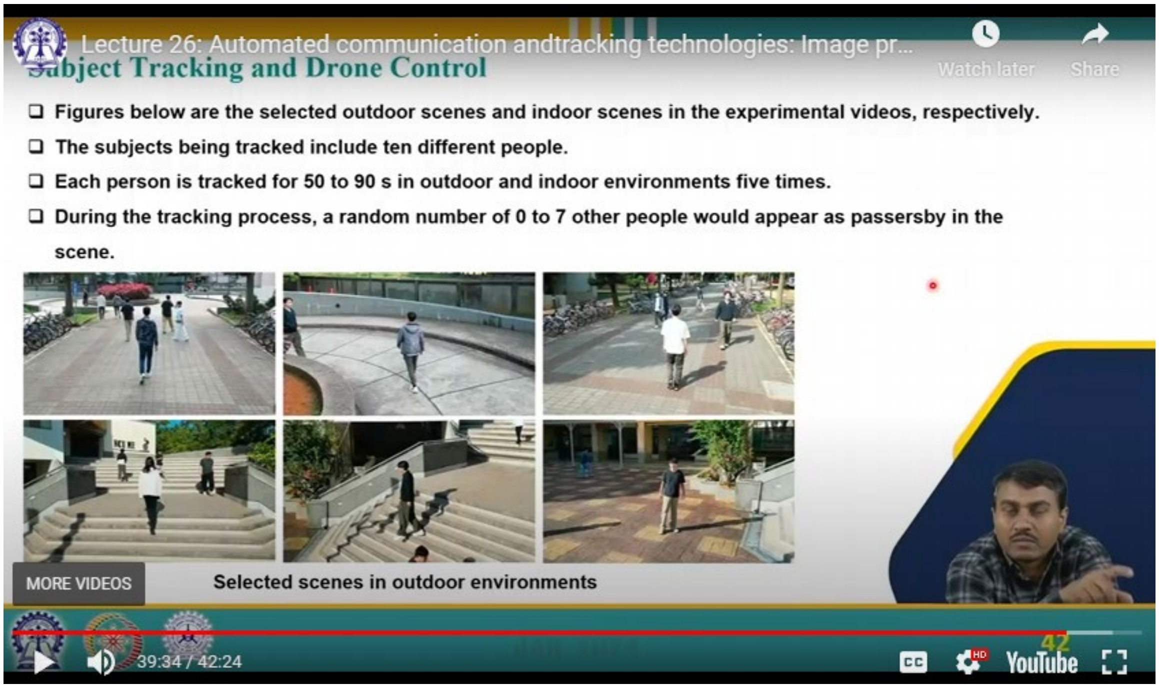
Error calculation for Y axis the error in the Y axis is compared with the Y coordinate of the track object with the Y coordinate of the screen center. So, this error is used to calculate the necessary altitude adjustment for the drone's vertical ascent or descent. Y axis error handling altitude adjustment are made based on the calculated error to control the drone particle movement and maintain the desired altitude relative to the track object. If the track object does not exhibit significant movement, the yaw option is selected, which only requires adjusting the drone heading to follow the target represented by the green aloe illustrated here. So now everything is tracked, so now green is basically in line with the target so that this basically targets the object here also it has been detected that there is not much amount of movement of the user.



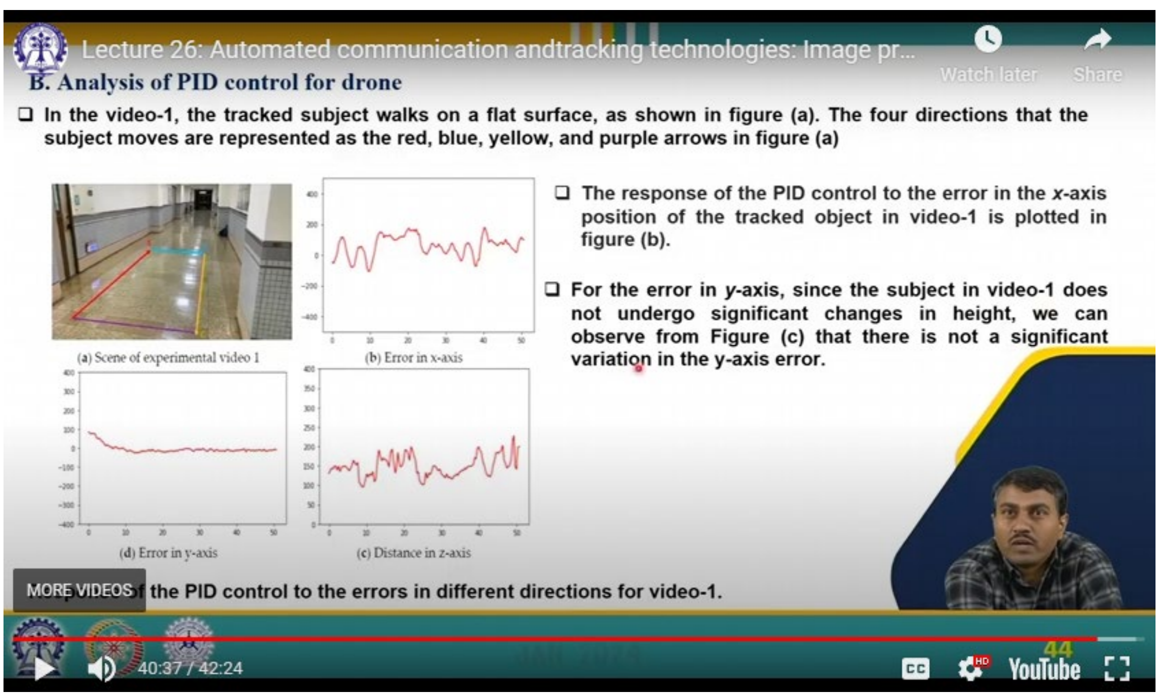
Detection model pruning and object detection, so this study trained baseline YOLO version 4 and the prune version YOLO 4 as well, and it has used the Coco2014 data set. The tiny version of the YOLO version 4 is specifically designed as a lightweight variant for devices with lower computational resources. To achieve this goal, a series of experiments to simulate the exploration needs of drones in a real environment and require a drone to successfully track target objects automatically. So, to evaluate the performance of the object detector, we applied the following four metrics. One is the precision it measures the proportions of true positive among all detection made by the system a higher precision indicates that the system can accurately identify target object reducing the likelihood of false alarm.

Recall it measures the proportion of true positive among all actual target object and a higher recall indicates the system can successfully detect a larger portion of the target object reducing the risk of miss detection. BFLOPs, BFLOPs is a metric used to measure the computational efficiency of a computer system or a machine learning model. It is a commonly used metric for evaluating the computational efficiency and speed of the system or model. Mean average precision at the intersection over Union 0.5, so this is MAP at the rate 0.5 is a commonly used evaluation metric in object detection. It measures the average precision at an intersection over the Union threshold of 0. 5 across different classes.

Subject tracking and drone control so here you can see the figure that basically selected for both outdoor and indoor scene from the video image and the subject being tracked for different ten people's ten different people, and each person is tracked for 50 to 90 seconds in outdoor environment and indoor environment five times. During the tracking process a random number of 0 to 7 number of people would appear as a passerby in the scene to basically create some amount of disturbance in the model and we want to see that yes model is able to track that is here person.

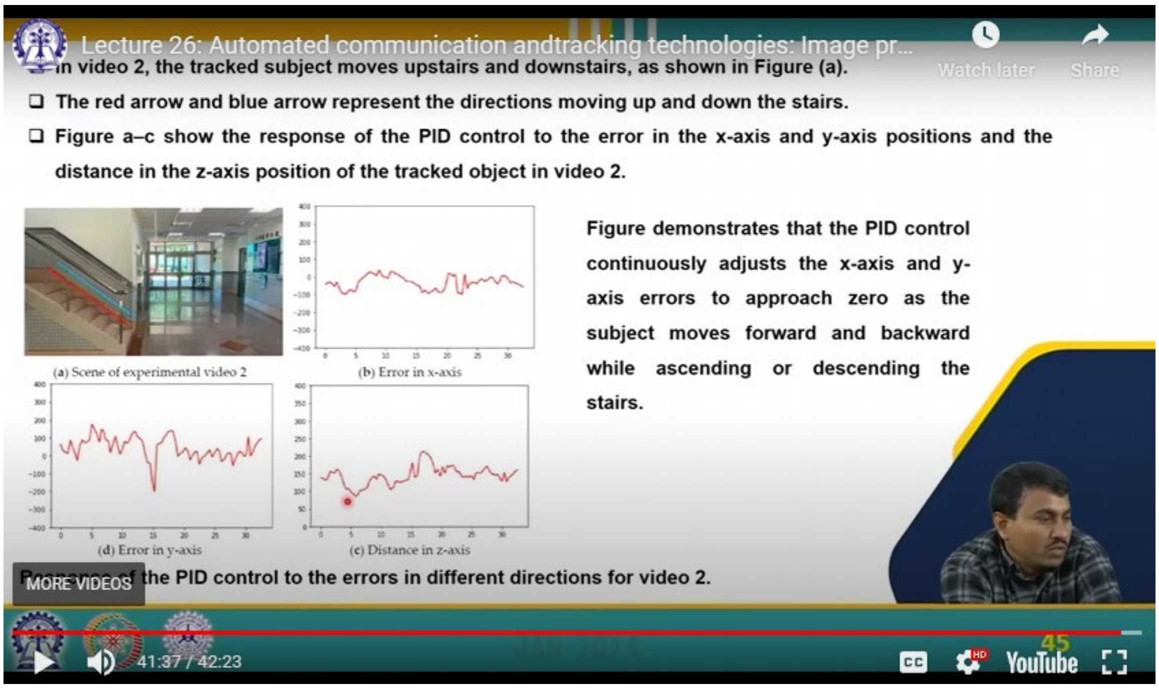


So, this is the indoor scene of the environment of those persons. Analysis of the PID control for drone in the video one-track subject walk on a flat surface here it is the flat surface so here y axis error is very less because it is moving on the same plane y change is very less. So, four figures shown in the subject moves are represented as red, blue, yellow, and purple arrows in the figure here: red, blue yellow and purple movement moves are represented in this line.



So this response of the PID control to the error of the x-axis position of the track object in video one is plotted in figure B here, the x-axis then this is the y-axis error in y-axis since the subject in video one does not undergo significant change in height we can observe on figure C that there is not a significant variation in the y-axis error and figure D plots the response here of the PID control of the error in the z-axis position of the track object in the selected video.

In video two, the track object moves stair and down stair. As you can see okay, the red arrow is basically red and blue arrow represent the direction moving up and down up and down and figure A to C shows the response of the PID control of the error in x-axis y-axis position and the distance in the z-axis position in the track object yellow version two. So, figure demonstrates that PID control continuously adjusts the x-axis and y-axis error to approach zero as the subject moves forward and backwards while ascending or descending the stairs.



So, these are the references.

And let me conclude in a few sentences. So, this study has highlighted that real-time human tracking and detection is possible using the image processing and drone technology and the technology not only improves the quality control measures that can contribute to cost-effectiveness and sustainability in the eastern industrial operation.

Thank you!