Capturing Unknown Person using AI Techniques

Pinnu Yasaswini

Department of Computer Science, Sri Venkateswara University, Tirupati

Abstract— *Security is the top most priority for everybody. A camera which uses modern AI techniques to solve the one of the major security issues "UNKNOWN PERSON" in India. Camera would act as a guard to your house. And all of these ensure safety and reduces the risk of being robbed. One thing that is common in every person is the sense of security. Today people are finding ways to secure their properties and/or possessions. We planned to use the latest technologies and expertise in the best way to make a project that would try to solve one of the biggest security problems of the society. Humans are very good at recognizing faces and complex patterns. The face plays a major role in our social intercourse in conveying identity. The human ability to recognize faces is remarkable. We can recognize thousands of faces learned throughout our lifetime and identify familiar faces at a glance even after years of separation. Our proposed system machine will also be learned faces as well as human. Camera will detect unwanted person in the home with the help of Artificial Intelligence using image processing and will notify the owner in their absence. Or any misbehave activities will done to detect using sensor by sense vibration. Haar cascade Algorithm has been used to do image comparison and match with the database and determine intruders to prevent any Unknown Person or disasters.*

Keywords: Haar cascade Algorithm, Face detection, Image comparison, Motion Detection.

I. INTRODUCTION

CCTV (Closed Circuit Television) is a visual surveillance technology designed for monitoring a variety of environments and activities. CCTV systems typically involve a fixed communication link between cameras and monitors. The main drawback of this is someone should be monitoring the system continuously. When a security camera is installed, a human being has to be on alert all the time while the camera is on since any little movement can require attention of the supervisor. But most of small objects do not need the supervisor's attention since they could be birds, cats, dogs etc.

The proposing new system can capture an image periodically with a predefined time interval and compare this image with a pre-captured image to detect a movement and to calculate the size of moving object through the image. The size of moving object is calculated and additionally the decision whether the image containing a certain moving objects which seems to be worth of human attention should be registered or not is made automatically. When motion occurs, captured image is compared with the authorized faces in the database, if any match is found then the system won't generate alerts, else it starts alerting the owner. The system alerts as fast as possible by sending mail to the authenticated users.

1.1 Artificial Intelligence

Artificial intelligence (AI) is the ability of a computer program or a machine to think and learn. It is also a field of study which tries to make computers "smart". As machines become increasingly capable, mental facilities once thought to require intelligence are removed from the definition. AI is an area of computer sciences that emphasizes the creation of intelligent machines that work and reacts like humans. Some of the activities computers with artificial intelligence are designed for include: Face recognition, Learning, Planning, Decision making etc. Artificial intelligence is the use of computer science programming to imitate human thought and action by analysing data and surroundings, solving or anticipating problems and learning or self-teaching to adapt to a variety of tasks.

1.2 Object Detection and Tracking

There is a wide range of computer vision tasks benefiting society such as object classification, detection, tracking, counting, Semantic Segmentation, Captioning image, etc. Process of identifying objects in an image and finding its position is known as object detection. Various object detection tasks. With advancements in field of computer vision assisted by AI, realization of tasks was realizable along t time scale. Semantic segmentation task of clustering pixels based on similarities. Classification + Localization and object detection method of identifying class of object and drawing a bounding box around it to make it distinct. Instance segmentation is semantic segmentation applied to multi objects. The general intuition to perform the task is to apply CNN over the image. CNN works on image patches to carry out the task many such salient regions can be obtained by Region-Proposal Networks like Region Convolution Neural network (RCNN), Fast- Region

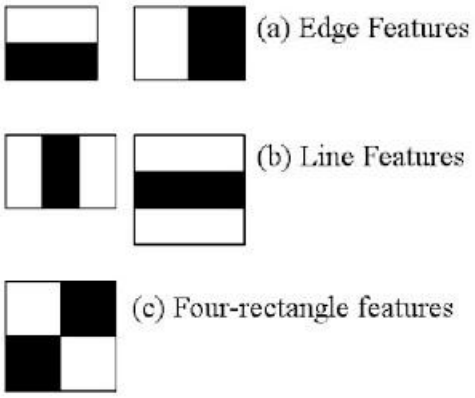
Convolutional Neural Network (Fast-RCNN), Faster- Region Convolutional Neural Network (Faster-RCNN). To perform selective search for object recognition Hierarchal Grouping Algorithm is used. Few bottlenecks by these approaches are mitigated by state-of the-art algorithms like You Only Look Once (YOLO), Single shot Detector (SSD). The efficient object detection algorithm is one which assures to give bounding box to all objects of vivid size to be recognized, with great computational capabilities, faster processing. YOLO and SSD assure to render promising results, but have a tradeoff between speed and accuracy. Hence, selection of algorithm is application specific.

1.3 Deep Learning in HAAR Cascade Algorithm

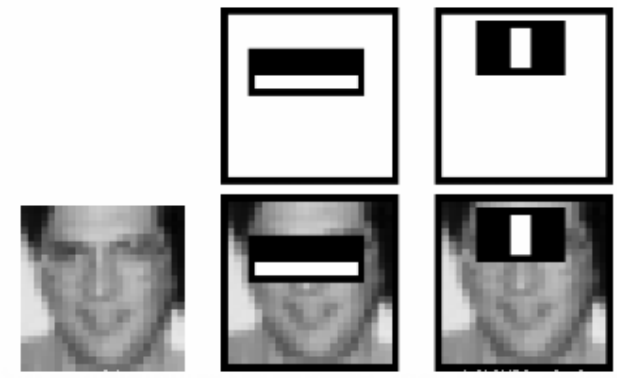
Haar Cascade is a machine learning object detection algorithm used to identify objects in an image or videos. It is a machine learning based approach where a cascade function is trained from a lot of positive and negative images. It is then used to detect objects in other images. The algorithm has four stages:

- Haar Feature Selection
- Creating Integral Images
- Adaboost Training
- Cascading Classifiers

It is well known for being able to detect faces and body parts in an image, but can be trained to identify almost any object. Let's take face detection as an example. Initially, the algorithm needs a lot of positive images of faces and negative images without faces to train the classifier. Then we need to extract features from it. First step is to collect the Haar Features. A Haar feature considers adjacent rectangular regions at a specific location in a detection window, sums up the pixel intensities in each region and calculates the difference between these sums.

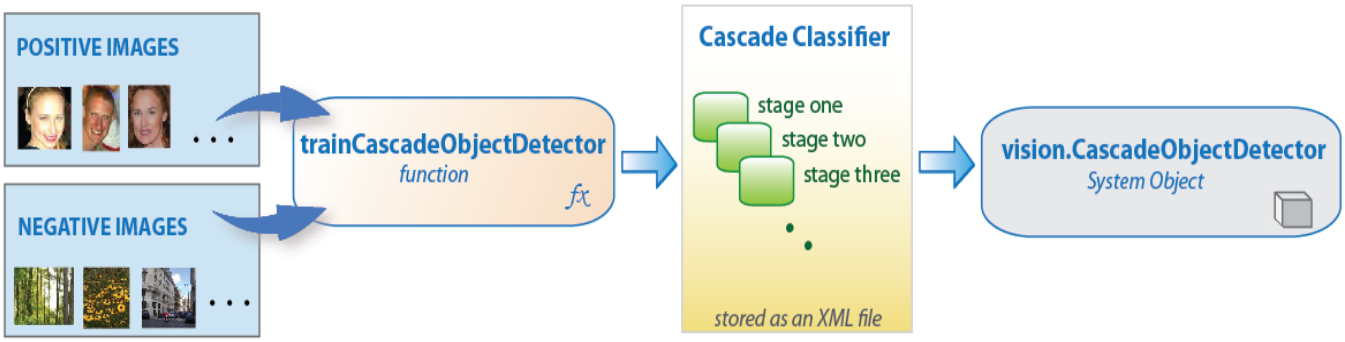


Integral Images are used to make this super-fast. But among all these features we calculated, most of them are irrelevant. For example, consider the image below. Top row shows two good features. The first feature selected seems to focus on the property that the region of the eyes is often darker than the region of the nose and cheeks. The second feature selected relies on the property that the eyes are darker than the bridge of the nose. But the same windows applying on cheeks or any other place is irrelevant.



So how do we select the best features out of 160000+ features? This is accomplished using a concept called Adaboost which both selects the best features and trains the classifiers that use them. This algorithm constructs a “strong” classifier as a linear combination of weighted simple “weak” classifiers. The process is as follows. During the detection phase, a window of the target size is moved over the input image, and for each subsection of the image and Haar features are calculated. You can see this in action in the video below. This difference is then compared to a learned threshold that separates non-objects from objects. Because each Haar feature is only a "weak classifier" (its detection quality is slightly better than random guessing) a large number of Haar features are necessary to describe an object with sufficient accuracy and are therefore organized into cascade classifiers to form a strong classifier.

1.4 Cascade Classifier



The cascade classifier consists of a collection of stages, where each stage is an ensemble of weak learners. The weak learners are simple classifiers called decision stumps. Each stage is trained using a technique called boosting. Boosting provides the ability to train a highly accurate classifier by taking a weighted average of the decisions made by the weak learners.

Each stage of the classifier labels the region defined by the current location of the sliding window as either positive or negative. Positive indicates that an object was found and *negative* indicates no objects were found. If the label is negative, the classification of this region is complete, and the detector slides the window to the next location. If the label is positive, the classifier passes the region to the next stage. The detector reports an object found at the current window location when the final stage classifies the region as positive.

The stages are designed to reject negative samples as fast as possible. The assumption is that the vast majority of windows do not contain the object of interest. Conversely, true positives are rare and worth taking the time to verify.

- A true positive occurs when a positive sample is correctly classified.
- A false positive occurs when a negative sample is mistakenly classified as positive.
- A false negative occurs when a positive sample is mistakenly classified as negative.

To work well, each stage in the cascade must have a low false negative rate. If a stage incorrectly labels an object as negative, the classification stops, and you cannot correct the mistake. However, each stage can have a high false positive rate. Even if the detector incorrectly labels a non-object as positive, you can correct the mistake in subsequent stages. Adding more stages reduces the overall false positive rate, but it also reduces the overall true positive rate.

Cascade classifier training requires a set of positive samples and a set of negative images. You must provide a set of positive images with regions of interest specified to be used as positive samples. You can use the Image Labeller to label objects of interest with bounding boxes. The Image Labeller outputs a table to use for positive samples. You also must provide a set of negative images from which the function generates negative samples automatically. To achieve acceptable detector accuracy, set the number of stages, feature type, and other function parameters.

II. LITERATURE REVIEW

Integrating Face Recognition Security System with the Internet of Things

Mohammad Sanaullah Chowdhury 2018

A perennial need for safety in the community depends on country, city, and district. In some instances, feeling safe is required on a 24/7 basis. A popular and cost-effective solution based on the Raspberry Pi has the promise of being both user-

friendly and cost-effective. It pairs the Raspberry Pi to a camera module for face recognition. It learns to detect those with granted access to the specified area under protection. Such stored faces are the subject of system training. If during operation the system recognizes the face in the dataset, then the camera shows the matching name with a confidence level possibly granting access, but alternatively it takes a photo of the subject and sends it as an email notifications warning. The proposed system can implement face recognition even from poor quality images performing well over both known and unknown datasets. Face recognition leverages techniques from the OpenCV library and is written in the Python language.

Attendance System Using Machine Learning-Based Face Detection for Meeting Room Application

Eueung Mulyana 2020

In a modern meeting room, a smart system to make attendance quickly is mandatory. Most of the existing systems perform manual attendance, such as registration and fingerprint. Despite the fingerprint method can reject the Unknown person and give the grant access to the Known person, it will take time to register first a person one-by-one. Moreover, it is possible to create long queues for fingerprint checking before entering the meeting room. Machine learning, along with the Internet of Things (IoT) technology is the best solution; it offers many advantages when applied in the meeting rooms. Generally, the method used is to create a presence by detecting faces. In this paper, we present a facial recognition authentication based on machine learning technology for connection to the meeting rooms. Furthermore, specific website to display the detection result and data storage design testing is developed. The method uses 1) the Dlib library for deep learning purposes, 2) OpenCV for video camera processing, and 3) Face Recognition for Dlib processing. The proposed system allows placing the multiple cameras in a meeting room as needed. However, in this work, we only used one camera as the main system. Tests conducted include identification of one Known person, identification of one Unknown person, identification of two people, and three people. The parameter to be focused is the required time in detecting the number of faces recorded by the camera. The results reveal that the face can be recognized or not recognized, then it will be displayed on the website.

Intelligent Decision Support Systems Based on Machine Learning and Multicriteria Decision-Making

Mohammed Alghaili 2020

Although significant advances have been made recently in the field of face recognition, these have some limitations, especially when faces are in different poses or have different levels of illumination, or when the face is blurred. In this study, we present a system that can directly identify an individual under all conditions by extracting the most important features and using them to identify a person. Our method uses a deep convolutional network that is trained to extract the most important features. A filter is then used to select the most significant of these features by finding features greater than zero, storing their indices, and comparing the features of other identities with the same indices as the original image. Finally, the selected features of each identity in the dataset are subtracted from features of the original image to find the minimum number that refers to that identity. This method gives good results, as we only extract the most important features using the filter to recognize the face in different poses. We achieve state-of-the-art face recognition performance using only half of the 128 bytes per face. The system has an accuracy of 99.7% on the Labeled Faces in the Wild dataset and 94.02% on YouTube Faces DB.

Neural Network-Based Face Detection

Henry A. Rowley

We present a neural network-based upright frontal face detection system. A retinally connected neural network examines small windows of an image and decides whether each window contains a face. The system arbitrates between multiple networks to improve performance over a single network. We present a straightforward procedure for aligning positive face examples for training. To collect negative examples, we use a bootstrap algorithm, which adds false detections into the training set as training progresses. This eliminates the difficult task of manually selecting nonface training examples, which must be chosen to span the entire space of nonface images. Simple heuristics, such as using the fact that faces rarely overlap in images, can further improve the accuracy. Comparisons with several other state-of-the-art face detection systems are presented, showing that our system has comparable performance in terms of detection and false-positive rates.

Deep Learning-Based Gait Recognition Using Smartphones in the Wild

Yanling Wang, Qian Wang, Yi Zhao, Qingquan Li 2020

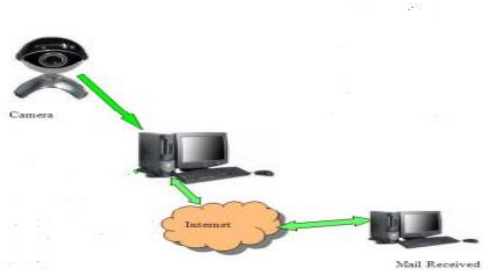
Compared to other biometrics, gait is difficult to conceal and has the advantage of being unobtrusive. Inertial sensors, such as accelerometers and gyroscopes, are often used to capture gait dynamics. These inertial sensors are commonly integrated into smartphones and are widely used by the average person, which makes gait data convenient and inexpensive to collect. In this paper, we study gait recognition using smartphones in the wild. In contrast to traditional methods, which often require a person to walk along a specified road and/or at a normal walking speed, the proposed method collects inertial gait data under unconstrained conditions without knowing when, where, and how the user walks. To obtain good person identification and authentication performance, deep-learning techniques are presented to learn and model the gait biometrics based on walking data. Specifically, a hybrid deep neural network is proposed for robust gait feature representation, where features in the space and time domains are successively abstracted by a convolutional neural network and a recurrent neural network. In the experiments, two datasets collected by smartphones for a total of 118 subjects are used for evaluations. The experiments show that the proposed method achieves higher than 93.5% and 93.7% accuracies in person identification and authentication, respectively.

III. PROPOSED WORK

- The proposing new system can capture an image periodically with a predefined time interval and compare this image with a pre-captured image to detect a movement and to calculate the size of moving object through the image.
- When motion occurs, captured image is compared with the authorized faces in the database by using Haar cascade classifier algorithm, if any match is found then the system won't generate alerts, else it starts alerting the owner. The system alerts as fast as possible by sending mail to the authenticated users and also add the link has given to the below message. If we opened the link, we will see detect the image of the unknown person.
- If anybody doing a misbehave activities like broken the lock or using Weapons to detect the action by using vibration sensor and also send to the alert message to the register mail. It will help to detect the intruder inside the home.

Advantage

- Wherever we go, it will give a high secure to your home.
- We will see the unknown or doing misbehave activities person's image within a second.



The result of image processing is to determine the nature of the moving object. In this implementation, the user can configure the threshold value of minimum movement. If the detected movement is smaller than the threshold value, the system considers that the noise in an image makes a movement, so the movement is negligible. Below are the functions in Haar cascade Algorithm.

Haar Cascade is a machine learning object detection algorithm used to identify objects in an image or videos. It is a machine learning based approach where a cascade function is trained from a lot of positive and negative images. It is then used to detect objects in other images.

Camera will capture the image when movement detection in front of camera, then image will be in the process of data pre-processing. After pre-processing step system will detect the face using face recognition. The detect image will compared to the already trained stored Database images using Haar cascade classifier algorithm, if any match is found then the system won't generate alerts, else it starts alerting the owner. The system alerts as fast as possible by sending mail to the authenticated users and also add the link has given to the below message. If we opened the link, we will see detect the image

of the unknown person and also If anybody doing a misbehave activities like broken the lock or using Weapons to detect the action by using vibration sensor and also send to the alert message to the register mail. It will help to detect the intruder inside the home.

IV. METHODOLOGY

4.1 Motion Detection.

Inputs: Takes the images from the camera instantly.

Output: Compares the Images frequently and gives the result.

Purpose: It takes the images and calls the method to compare these images.

4.2 Image Comparison

Inputs: It takes two images to compare

Output: Comparison of the captured images will be done. If there is a difference the image is saved or else comparison is continued.

Purpose: It will compare the two images taken by the camera.

4.3 Face Detection

Inputs: It takes the captured image and database images

Output: If the image matches with any of the database images, processing will be continued or else it starts alerting.

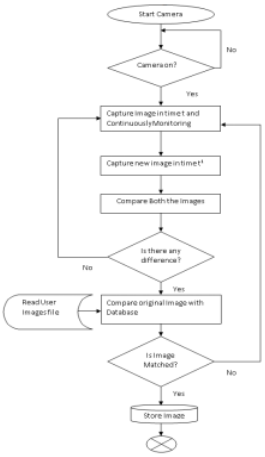
Purpose: To check whether the image is in the database or not

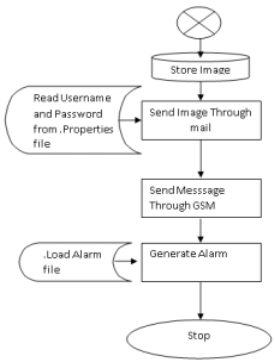
4.4 Sending Mail

Input: It takes the image having difference.

Output: Image will be delivered to the mail.

Purpose: It is to send an image to mailbox





- Start the camera then capture the image in front of camera when movement detection occurs.
- Capturing image send to the database and compare to the stored image.
- If both images are same, it will deny and do same the work.
- Otherwise, image will be sent to the registered Mail.
- And also detect if any misbehave activities may be done using vibration sensor.

V. CONCLUSION

Detecting human beings accurately in a surveillance video is one of the major topics of vision research due to its wide range of applications. It is challenging to process the image obtained from a surveillance video as it has low resolution. A review of the available detection techniques is presented. The detection process occurs in two steps: object detection and object classification. In this paper, all available object detection techniques are categorized into background subtraction, optical flow and spatial-temporal filter methods. The object classification techniques are categorized into shape-based, motion-based and texture-based methods. The characteristics of the benchmark datasets are presented, and major applications of human detection in surveillance video are reviewed.

REFERENCES

[1] Sulman N, Sanocki T, Goldgof D, Kasturi R: How effective is human video surveillance performance? In 19th International Conference on Pattern Recognition, (ICPR 2008). Piscataway: IEEE; 2008:1-3.

[2] Stauffer C, Grimson W: Adaptive background mixture models for real-time tracking. In IEEE Conference on Computer Vision and Pattern Recognition (CVPR 1999). Piscataway: IEEE; 1999:246-252.

[3] Tian YL, Feris RS, Liu H, Hampapur A, Sun M-T: Robust detection of abandoned and removed objects in complex surveillance videos. Syst. Man Cybern. Part C Appl. Rev. IEEE Trans. 2011, 41(5):565-576.

[4] Lee DS: Effective Gaussian mixture learning for video background subtraction. IEEE Trans. Pattern Anal. Mach. Intell. 2005, 27(5):827-835.

[5] Wang J, Bebis G, Miller R: Robust video-based surveillance by integrating target detection with tracking. In IEEE Computer Vision and Pattern Recognition Workshop (CVPRW '06). Piscataway: IEEE; 2006:137.

[6] Stenger B, Ramesh V, Paragios N, Coetzee F, Buhmann JM: Topology free hidden Markov models: application to background modelling. In IEEE International Conference on Computer Vision (ICCV 2001). Piscataway: IEEE; 2001:294-301.

[7] Zhang W, Zhong X, Xu FY: Detection of moving cast shadows using image orthogonal transform. In 18th International Conference on Pattern Recognition (ICPR'06). Piscataway: IEEE; 2006:626-629.

[8] Kim W, Kim C: Background subtraction for dynamic texture scenes using fuzzy color histograms. Signal Process. Lett. IEEE 2012, 19(3):127-130.

[9] Han B, Comaniciu D, Davis L: Sequential kernel density approximation through mode propagation: Applications to background modelling. In Asian Conference on Computer Vision. Jeju Island: Korea; 2004.

[10] Lanza A: Background subtraction by non-parametric probabilistic clustering. In 8th IEEE International Conference on Advanced Video and Signal-Based Surveillance. Piscataway: IEEE; 2011:243-248.