

Camera based Smart Surveillance System

Ishan Kokadwar

Dept. of Computer Engineering
PCCOE, Pune

Anurag Kukarni

Dept. of Computer Engineering
PCCOE, Pune

Vaibhav Limbhore

Dept. of Computer Engineering
PCCOE, Pune

Sayali Khare

Dept. of Computer Engineering
PCCOE, Pune

Swati Chandurkar

Dept. of Computer Engineering
PCCOE, Pune

ABSTRACT

One of the key sectors which is now concern worldwide is security and privacy. Due to the emergence of protecting premises, providing security is one of the most important tasks. Thus, to provide security, the video surveillance system was introduced. But video surveillance system has few drawbacks such as human intervention to observe footage, storage, and expertise to analyze multiple camera system. To overcome it, in their project authors have made a system that automatically tracks each individual under surveillance, maintains logs instead of video recordings, analyzes frames and generates helpful insights. Convolution Neural Network (CNN) based deep learning model is used to build the system.

General Terms

Artificial Intelligence, Convolution neural network, Data Analysis, Deep learning, Face Recognition, Image Video Surveillance

Keywords

Artificial Intelligence, Convolution neural network, Data Analysis, Deep learning, Face Recognition, Image Video Surveillance

1. INTRODUCTION

Nowadays, security is a major concern in every organization. To satisfy it the organizations use surveillance cameras. The limitation in using them is that there must be an operator to watch the stream from the cameras and take respective decisions. The use of camera based surveillance has been extended from security to tracking, environment analysis and threat analysis. By using the power of modern computing and hardware it is possible to automate the process. In our project, we have used deep learning Convolution Neural Network to analyse the video stream, take needy actions, present the data, and maintain logs for future work. The organization using this system is freed from handling surveillance team and efficiently tracks each activity on campus premises. The dashboard of our system helps to get useful insight from multiple CCTV cameras in a single window. Here the video frame from the camera is greyscaled, padded for resolution matching, then interpreted by MTCNN based model where faces are identified and indexed. Further, these indexed faces are passed to custom-designed CNN model for face recognition.

2. RELATED LITERATURE

Greyscale images have two channels describing the colour and coloured images have three channels i.e. RGB. When dealing with the coloured image one has to compute for 3d

matrix and for greyscale 2d matrix hence to reduce complexity for convolution computation greyscale images are used. Image padding adds new pixels along the edges of an image. The border provides space for annotations or acts as a boundary while using advanced filtering techniques and makes it compatible with further processing. Convolution neural network (CNN)[1] is a type of Artificial Neural Network which is used to manipulate imagery data. It works on each pixel of the image and processes it. CNN's basic building blocks are the neurons that compute together and extract features from the image. Over the process, it assigns weights to each neuron based on the reaction of the environment to the outcome generated by CNN model. In this way it gets trained over certain features. Multi-task Convolution neural Network is an algorithm consisting of 3 stages, which detects the bounding boxes of faces in an image along with their 5 Point Face Landmarks. Each Stage gradually improves the output by passing it through CNN. It returns the candidate bounding boxes of the face and then these are indexed for further recognition.

3. PROJECT APPROACH

Project is implemented by considering educational institution under consideration. While students are being registered for the course in college, data gathering is done by taking video of a student's face. And from this video, multiple frames are extracted and stored in the database. For every student, 4000 images are stored in greyscaled and cropped format. Further, this database is used to train the CNN[2] model for face identification. The outcome of this training is the CNN model for classifying multiple students on campus at the same time. The other side of the project is collecting the real-time video stream from the CCTV cameras and tracking each frame from it. For each frame following tasks are performed: a) Grey scaling the frame. b) Padding for respective dimensions. c) Passing it to MTCNN for boxing and indexing faces in it. d) Passing these indexed faces to our CNN for face N identification. e) Logging it in the database. f) Analyzing the logs for the insights

4. DATASET GENERATION

Dataset is created by taking a 1-minute video of each individual and then extracting frames from a video and adding to a dataset. Images for each individual are increased to 4000 by copying the same images until we get 4000 images. Then these images are gray scaled and passed to the MTCNN model for cropping out faces from an image. Then these cropped images are passed on to custom CNN for training.

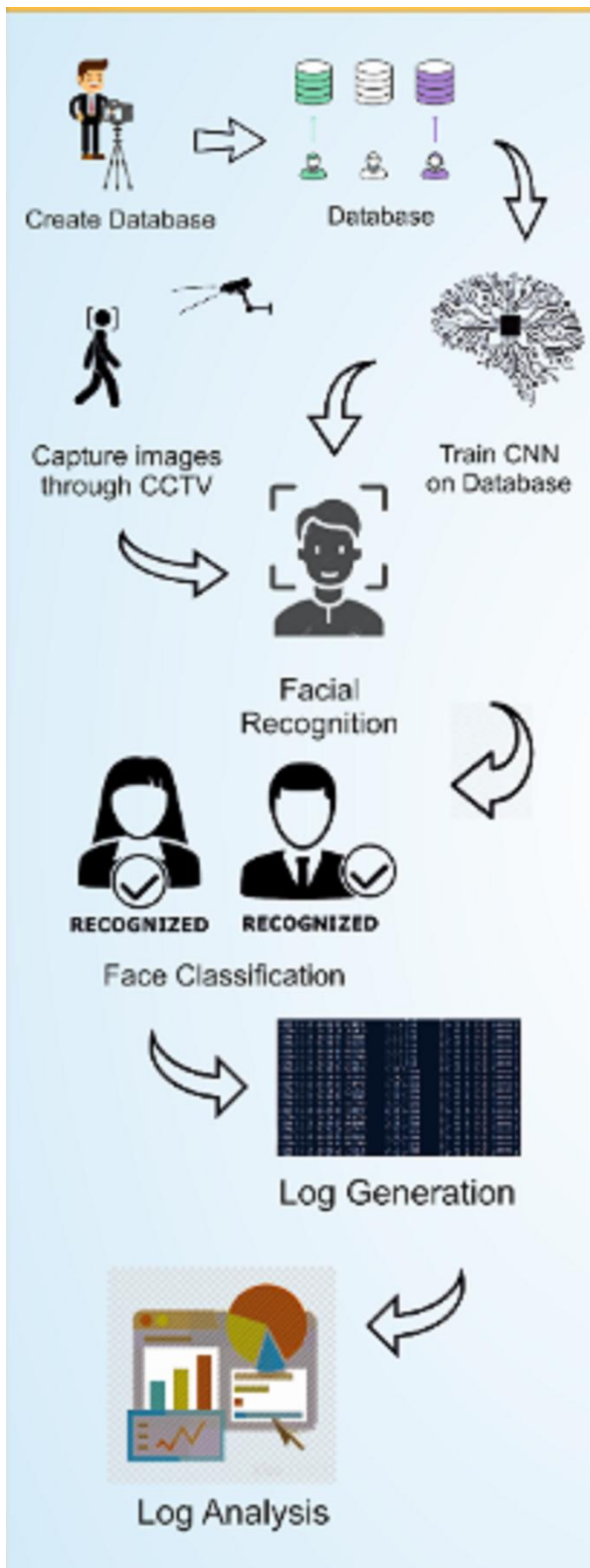


Fig. 1: Project Flow

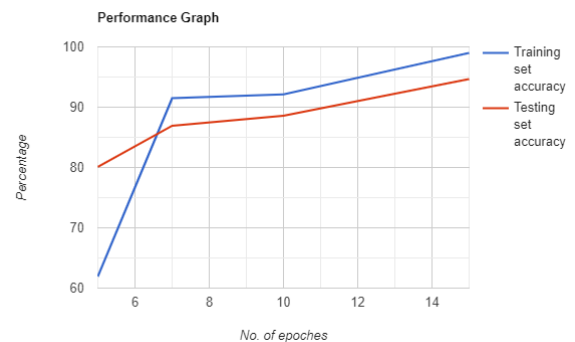
The authors have divided each individual's dataset into 80 % for training and 20% for validation. This process is carried out for each new entry to the system. By using the feature of continuous training, we increase the accuracy of the model for each individual further by adding specified images from CCTV Cameras. Here, authors have trained the model on 4 people as 4 different classes.

5. CUSTOM BUILD CNN FOR RECOGNITION

Once the images are indexed by the MTCNN module then it is passed to the recognition module. • Recognition module consists of a deep learning model of 5 layered Convolution Neural network architecture. • The 1st layer consists of 16 filters of convolution, second consists of 32 filters of convolution, 3rd one with 64 filters of convolution, 4th with 512 units, and dropout of 20 percent to avoid overfitting to the data. Lastly, the output is sent to the softmax layer to categorize the faces. • The optimizer used here is Adam with a learning rate set to 0.0001 and the loss function used is sparse categorical cross-entropy • Here, authors have trained the model for 4 faces as 4 classes in their project. The authors got an accuracy of 89.63 percent for this layer.

During experimental analysis on custom CNN, it was tuned on various hyperparameters and repeated the training process until CNN performed well on the testing dataset. The major parameters that contributed the most in improving the performance of the model were the number of epochs, batch size, the learning rate of Adam optimizer.

The number of epochs was gradually increased in subsequent runs of optimizing CNN and authors observe that as they increase the number of epochs the accuracy of CNN increases but after some values of epochs they got to the point of diminishing returns. The following chart describes the performance trend observed by tuning the hyperparameters



6. RESULTS OF APPLICATION

6.1 Logs

The logs are the entry for every activity that is tracked on the campus premises. As log files take less disk space than the video file so one can maintain records for a long time. Here is the snapshot from the log file.

Timestamp	Image No.	detection Vector	Camera-No.	Person id
2020-01-24 10:50:53.835885	1_28_1.jpg	[[1. 0. 0. 0. 0]]	Camera-25	16011012c100000
2020-01-24 10:50:55.891745	1_27_1.jpg	[[0. 0. 1. 0. 0]]	Camera-25	16011012c100002
2020-01-24 10:51:53.835885	1_28_1.jpg	[[1. 0. 0. 0. 0]]	Camera-25	16011012c100000
2020-01-24 10:51:55.891745	1_27_1.jpg	[[0. 0. 1. 0. 0]]	Camera-25	16011012c100002
2020-01-24 10:52:57.953312	1_18_1.jpg	[[0. 0. 1. 0. 0]]	Camera-21	1601012c100000
2020-01-24 10:52:58.975100	1_14_1.jpg	[[0. 0. 1. 0. 0]]	Camera-21	16011012c100002
2020-01-24 10:54:23.835885	1_28_1.jpg	[[1. 0. 0. 0. 0]]	Camera-17	16011012c100000
2020-01-24 10:54:58.835885	1_28_1.jpg	[[1. 0. 0. 0. 0]]	Camera-17	16011012c100002
2020-01-24 10:55:55.891745	1_27_1.jpg	[[0. 0. 1. 0. 0]]	Camera-18	16011012c100000
2020-01-24 10:56:55.891745	1_27_1.jpg	[[0. 0. 1. 0. 0]]	Camera-18	16011012c100002

Fig. 2: Sample logs of the images from the CCTV cameras

It contains

- I. Timestamp
- II. Image No.
- III. Detection Vector
- IV. Camera Number
- V. Person-id

Further analysis operations can be carried out and projected on the dashboard. The access of this database logs is shared with the web dashboard.

6.2 Dashboard

Currently, on the dashboard, authors have given features of continuous live tracking, crowd density map generation and automatic attendance marking. The feature list can be increased according to the requirement. At the Admin login once the student-id is entered along with the date we get path traversed by the user by extracting information from the log data. The live position can also be viewed in the student profile.

Fig. 3: Dashboard Showing Path Traversed by a particular entity using different cameras and logs

Admin can view live/latest location of each student in the student profile section.

Fig. 4: Admin Panel for Entering Camera Details and searching the information related to camera

Admin can also set positions of various cameras on the campus premises. With the help of camera position and log analysis, the crowd density maps are generated. Through these maps, potential crowd places can be analysed. For lecturers, once they enter the timetable of lectures the respective attendance will be automatically marked by interpreting logs. By following such a unique pattern, authors have managed to get system accuracy of 87.42%. For calculating accuracy, authors have trained the system on four faces and tested by using five CCTV cameras.

Fig. 5: Dashboard showing Current location of any entity generated using logs

7. CONCLUSION

This paper proposed a system by which it can record the human moment through face recognition in multiple camera systems. The custom CNN model is tuned in such a way that it gets trained and extracts features that have a high potential of differentiability. We have successfully managed to get an accuracy of 87.42 % of the complete system when trained on four faces and tested on five cameras. The Experimental results are considerable and ensure that this system can be used in environmental conditions. The limitation of the system is that it is less effective in tracking those people who are away from the camera. This system is dependent on CCTV camera conditions and light intensity.

8. REFERENCES

- [1] R. Chauhan, K. K. Ghanshala and R. C. Joshi, "Convolutional Neural Network (CNN) for Image Detection and Recognition," 2018 First International Conference on Secure Cyber Computing and Communication (ICSCCC), Jalandhar, India, 2018, pp. 278-282, doi: 10.1109/ICSCCC.2018.8703316.)Ding, W. and Marchionini, G. 1997 A Study on Video Browsing Strategies. Technical Report. University of Maryland at College Park.
- [2] N. Jmour, S. Zayen and A. Abdelkrim, "Convolutional neural networks for image classification," 2018 International Conference on Advanced Systems and Electric Technologies (IC_ASET), Hammamet, 2018, pp. 397-402, doi: 10.1109/ASET.2018.8379889. Tavel, P. 2007 Modeling and Simulation Design. AK Peters Ltd.
- [3] V. C. Banu, I. M. Costea, F. C. Nemtanu and I. Bădescu, "Intelligent video surveillance system," 2017 IEEE 23rd International Symposium for Design and Technology in

- Electronic Packaging (SIITME), Constanta, 2017, pp. 208-212, doi: 10.1109/SIITME.2017.8259891. Forman, G. 2003. An extensive empirical study of feature selection metrics for text classification. *J. Mach. Learn. Res.* 3 (Mar. 2003), 1289-1305.
- [4] J. G. Dastidar and R. Biswas, "Tracking Human Intrusion through a CCTV," 2015 International Conference on Computational Intelligence and Communication Networks (CICN), Jabalpur, 2015, pp. 461-465, doi: 10.1109/CICN.2015.95.
- [5] W. Yimyam, T. Pinthong, N. Chumuang and M. Ketcham, "Face Detection Criminals through CCTV Cameras," 2018 14th International Conference on Signal-Image Technology & InternetBased Systems (SITIS), Las Palmas de Gran Canaria, Spain, 2018, pp. 351-357,doi:10.1109/SITIS.2018.00061.