ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com

International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 8, Special Issue 1, August 2021

A Deep Learning Based Model for Detecting a Person from Surveillance Video for Crime Investigation

Sunandha Rajagopal[1], Dhannya J[2], Soumya Koshy[3], Cini Joseph[4]

Assistant Professor, Kristu Jyoti College of Management and Technology, Kottayam, Kerala, India[1]
Assistant Professor, Kristu Jyoti College of Management and Technology, Kottayam, Kerala, India[2]
Assistant Professor, Kristu Jyoti College of Management and Technology, Kottayam, Kerala, India[3]
Assistant Professor, Kristu Jyoti College of Management and Technology, Kottayam, Kerala, India[4]

Abstract— In crime investigations, there could be cases where the authorities would need to track the movements of a suspected person or want to see whether a particular person is involved in a crime. The video stream form surveillance cameras play very important role in such situations. CCTV cameras have been implemented in most of the cities, wherever security is most important. Manually searching for a particular person in a surveillance video is a very tedious task which requires a lot of time and man power. It will be useful if we could automatically scan through the video frames and locate a person in a particular frame. In this paper we present a model for automatically detect the presence of a person in a video stream. In this model, the process of finding a person in a video stream is divided into a number of phases including frame extraction from the video, training the deep neural network using the data set, filtering the video frames for improving quality and tracking the person using trained Convolutional Neural Networks.

Keywords: Crime Investigation, Surveillance, Pattern Matching, Convolutional Neural Networks.

I. INTRODUCTION

Nowadays surveillance cameras have been installed in most of the public places as well as private areas like home and offices. The video streams from the surveillance cameras play very important role in various applications like crowd detection, suspicious activity detection, person tracking etc. With the growing demands for automated video analysis, there is a great interest in objects detection and tracking of moving objects. In analysing video data, the primary focus is on tracking moving objects such as people or vehicles.

Recent years, detecting humans in a surveillance video scene has gotten more attention due to its wide variety of aapplications, including abnormal event detection, characterization of human manner, counting the number of people in a dense crowd, and identifying individuals. The visuals that we get from a surveillance video could be with low resolution. In the case of scenes collected form static cameras, there will be minimal change in background.

Objects in the outdoor surveillance are often detected in far field. Most existing digital video surveillance systems rely on human observers for detecting specific activities in a real-time video scene. However, there are limitations in the human capability to monitor simultaneous events in surveillance displays[4]. Thus, human motion analysis is one of the most exciting research topics in computer vision and pattern recognition.

In this study our focus is on detecting people rather than recognizing their complex activities. Sensing humans in a video stream is an extremely difficult task from a machine vision perspective since the appearance of human beings can vary widely with changes in clothing, lighting, and even posture.

In this paper, we are presenting a conceptual model for finding a particular person from surveillance video. First, we have to extract Key Frames using the method of “Global Comparison between frames”. Then the quality of the KFs is to be improved using filtering techniques and a trained CNN is used to recognize the person from this KFs.

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com

International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 8, Special Issue 1, August 2021

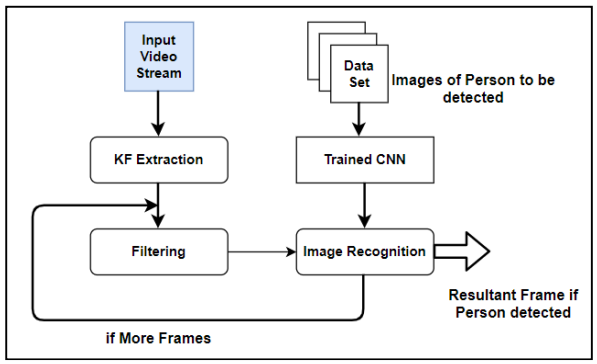


Fig.1: Block Diagram of the phases in the prescribed model

The paper is organized as follows. A brief review on different methods used for object tracking from videos are described in Section II and Section III explains the methods and materials for the working phases of this model. Finally, Section IV concludes the paper.

II. LITERATURE REVIEW

This section gives a short outline of various tracking A comprehensive survey of different tracking methods can be found in previous studies [6,7,8].

In traditional methods, object detection could be finished by using background subtraction, optical float and spatio-temporal filtering strategies. Once detected, a shifting object could be labelled as a human being through the usage of form-based, texture-based or movement-based features. These models typically based on based on motion- and observation-based models.

The motion model is comprised of the recognition and prediction of object position in successive frames,[9,10], but the experimental model focused on the appearance and position of the tracked object across the frame.[11] Some researchers used the template-based method for object tracking. Many researchers applied machine learning–based methods for object tracking, which classifies [13] the tracked object such as boosting,[14] using random forest,[15] Hough forest,[16] structural learning,[14] and support vector machine (SVM).[26] Some proposed feature-based tracking methods such as Haar-like features,[17] local binary pattern (LBP),[18] histogram of oriented gradient (HoG),[19,20] scale-invariant feature transform (SFIT),[21] discrete cosine transform (DCT),[22,23] and shape features.[24] Other techniques employed Kalman filters or Hungarian algorithm.[25]

Many researchers used combination of multiple cues information and presented methods for object tracking which combines feature-based detector with the probabilistic segmentation method. Majority of those methods are especially advanced for frontal view records set which may suffer from occlusion problems. Some scholars addressed occlusion problems by using overhead cameras for surveillance. [27]

This paper describes a deep learning-based model, in which the key frames are extracted from the video streams, their quality is improved using filtering technique and a trained CNN is used to identify a person from the extracted KF.

III. METHODS AND MATERIALS

The prescribed model is organized in four phases. This section describes the working of each phase in detail.

A. KEY FRAME EXTRACTION

In order to detect a person from a video, the best method would be comparing each frame with a required image. But it wouldn't be practical, since a surveillance video stream may contain enormous frames. So, in this method we propose to extract key frames and this KFs can be compared with the images of the particular person to be detected.

Keyframes are the essential frames which incorporate data of a start or end point of a movement. It is a shot that defines the starting and ending points of a smooth transition. Key frame extraction is a prevailing tool that outfits video content by selecting only a set of key frames to represent video streams.

In this model, we use the "Global Comparison between frames" to extract keyframe from a video stream. The algorithm determines the shot obstacles by optimizing a predefined goal feature that relies upon at the software. The objective function could be one of the following four measures.

1) Even temporal variance: The shot segment, or a key frame, having equal temporal variance are selected
2) Maximum coverage: Maximize the representation coverage of each key frame.
3) Minimum correlation: Minimize the sum of correlations between key frames, so that the selected key frames as uncorrelated with each other.
4) Minimum reconstruction error: Minimize the sum of the differences between each frame and its corresponding predicted frame reconstructed from the set of key frames using interpolation.

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com

International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 8, Special Issue 1, August 2021

There are different methods for extracting KFs from video streams. The reason for choosing this particular method is that, here the global characteristics are reflected into the extracted KF, number of the extracted KFs is manageable and, the KF set are considered denser. But when compared to other sequential methods, it has more computational complexity.

B. FILTERING

Even if surveillance cameras are very common in public places, the scenes obtained from these are ordinarily with low resolution. The quality of the video frames may affect the process of detecting the presence of a person in the frame. Even a trained deep neural network may fail in recognizing a particular person from the low-quality video frames. So, before, image recognition, we need to improve the quality of the video frame by the method of image filtering.

Each filtering method has its limitations. In this model, we propose to use the median filter for enhancing the quality of KFs generated in the first phase. The median filter is a non-linear digital filtering technique, regularly used to eliminate noise from an image. Here, the output samples are calculated as the median of the input samples in a particular window. In the case of median filter, the centre pixel of a M × M window is substituted by the median value of the corresponding window, considering the noise signals to be different from the median. Under specific conditions, it conserves edges while removing noise.

In the case of median filter, it scans through the signal entry one by one, and replaces each entry with the median of neighboring entries. The outline of neighbors is called the "window". We can compute the median by first sorting the pixel values from the window into numerical order, and then replacing the pixel being considered with the middle (median) pixel value. For one-dimensional signals, the most apparent window is the first few prior and succeeding entries. Figure 2 gives an illustration about the calculation.

119	123	118	120	117
125	121	119	114	115
119	124	123	145	125
120	123	118	117	126
126	119	122	127	125

Neighborhood Values:

114, 115, 117, 118, 119, 145, 123, 125, 126

Median Value= 119

Fig2: Calculation of median in the selected window

In the example, the central pixel value of 145 is deceiving when compared to the surrounding pixels and is replaced with the median value: 119. The window used here is a 3×3 square neighborhood.

The main advantage of using a median filter is that a median value is a more robust average than a mean. So not even a single unrepresentative pixel in a window will not noticeably affect the median value. Also, since the median value must be the value of one of the pixels in the selected window, it could not bring any new value causing a drastic change in the image.

C. IMAGE RECOGNITION USING TRAINED CNN

Convolutional Neural Networks are type of artificial neural networks, mainly used for image processing. It utilizes profound figuring out how to perform both generative and descriptive errands, regularly utilizing machine vison that incorporates picture and video recognition. The CNNs had already been proved to the best for image recognition.

The CNN is organized into different layers. The input layer, output layer, and a hidden layer which contains series of convolution and pooling layers and fully connected layer. The input layer takes the input image as grey scale. The convolution layer is responsible for identifying various features of the image. The fully connected layer predicts the class of image based on the output from convolution process.

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com

International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 8, Special Issue 1, August 2021

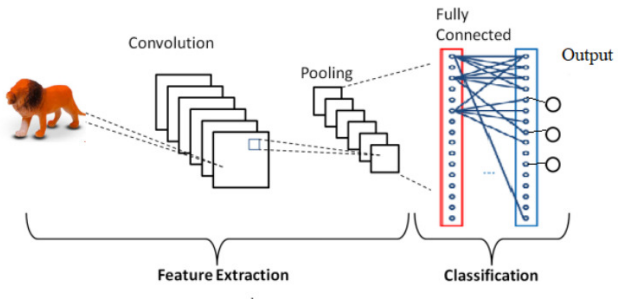


Fig 3: Architecture of CNN

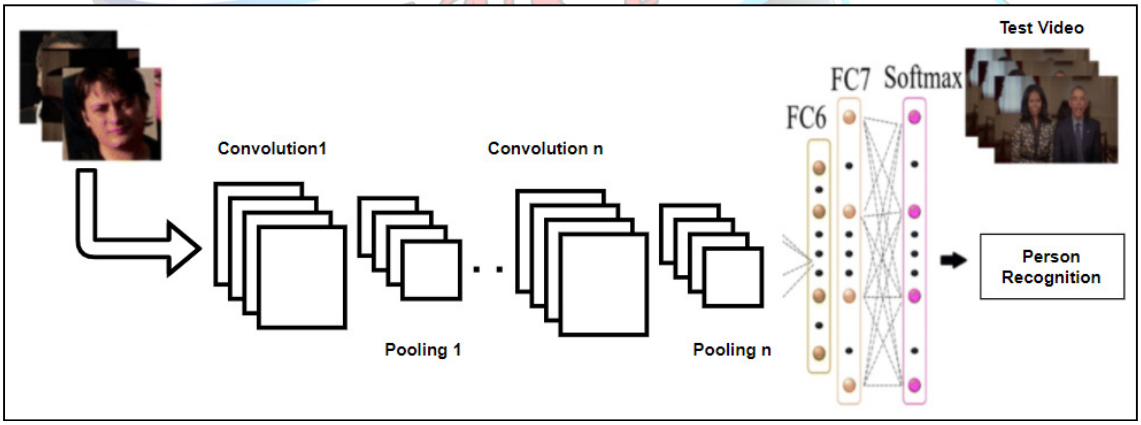
The convolution layer performs the mathematical operation called 'convolution', which involves the multiplication between an array of input data and a two-dimensional array of M×M weights, called a filter or a kernel. The filter is slid over the input image and the dot product is taken between the filter and the parts of the input image with respect to the size of the filter. The output is a feature map which will give us details about the image such as the corners and edges.

Since images are usually nonlinear, we apply an activation function to the feature map to increase the non-linearity of the network. It removes the negative values from activation map by setting them to zero. This is another step in Convolution layer, known as rectified linear unit (ReLU)

Usually, the convolutional layer will be followed by a pooling layer, which reduces the size of the feature map resulted from convolution operation, so as to reduce the computational cost. With respect to the method used, there can be different types of pooling namely max-pooling, average-pooling, sum-pooling etc.

The Fully Connected (FC) layer is just before the output layer, consists of the weights and biases and is used to connect the neurons between two different layers. In this, the input image from the previous layers is flattened and fed to the FC layer. After this, the flattened vector goes through few more FC layers where, we apply mathematical functions to this. In this stage, the classification process begins to take place. At this step, the error is calculated and then backpropagated so that the weights and feature detectors are adjusted to optimize the performance of the model. Then the process is repeated. This is how our network trains on the data.

In our model, we have to train the CNN with different images of person, whose presence is to be detected from the video. Better data set we use for training the CNN will increase the chance of better identification of the person. Figure 4 demonstrates the working of CNN.



IV. CONCLUSIION

The video streams from surveillance cameras play a vital role in various applications, like object/person detection, traffic management, crowd analysis etc. Especially in crime investigation, it may be frequently required to find the presence of a particular person in surveillance video. It would be tiresome for someone to search the entire video streams manually for detecting the

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com

International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 8, Special Issue 1, August 2021

person. Through this paper, we suggest a conceptual model for person identification from surveillance video. This model could be an advantage to the authorities to track a person through surveillance for crime investigations. The overall working of the model is summarized in the following flow chart.

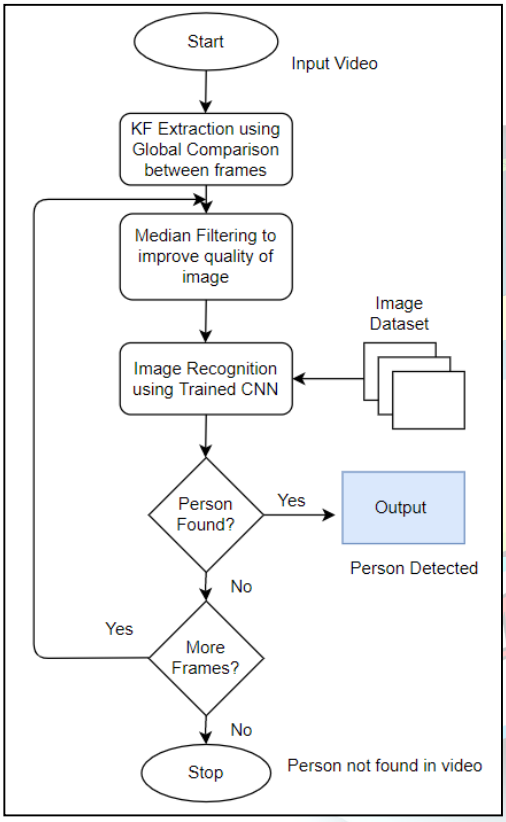


Fig: Flow chart demonstrating working of the prescribed model

REFERENCES

[1] Nuno Guimarães INESC/IST, R. Alves Redol, 9, 6o, 1000 Lisboa *email: {Ines.Oliveira,Nuno.Correia,Nuno.Guimaraes}@inesc.pt*, Image Processing Techniques for Video Content Extraction Inês Oliveira, *Nuno Correia,*

[2] 1D.P. Kucherov, 2R.G. Katsalap, 3L.V. Zbrozhek *1Faculty of Computer Science, National aviation university, Kiev, Ukraine 2Institute of Encyclopedic Research, National Academy of Sciences, Kiev, Ukraine 3Faculty of Computer Science,* , Improving Images Quality by Combination of Filtering Methods *National aviation university, Kiev, Ukraine*

[3] *kalaivani.R1, Manicha chezhian.R2 Research Scholar, Computer Science, NGM College, Coimbatore, India1 Associate Professor, omputer Science, NGM College, Coimbatore, India 2*, Object Detection in Video Frames Using Various Approaches

[4] N Sulman, T Sanocki, D Goldgof, R Kasturi, *How effective is human video surveillance performance?* in 19th International Conference on Pattern Recognition, (ICPR 2008) (*IEEE, Piscataway, 2008), pp. 1–3*

1 ISRAA HADI ALI, 2TALIB T. AL - FATLAWI 1 IT College. Babylon University, Iraq 2 Ph.D. student - IT College. Babylon University, Iraq E-mail: 1 israa_hadi1968@yahoo.com, 2 talib.turkey@qu.edu.iq, *KEY FRAME EXTRACTION METHODS*

[5] Zhou S, Ke M, Qiu J, et al. *A survey of multi-object video tracking algorithms.* In: Abawajy JH, Choo KKR, Islam R, et al. (eds) *International conference on applications and techniques in cyber security and intelligence.* Berlin: Springer, 2018, pp.351–369.

[6] Ahmad M, Ahmed I, Ullah K, et al. *Person detection from overhead view: a survey*. Int J Adv Comput Sci Appl 2019; 10(4): 567–577.

[7] Smeulders AW, Chu DM, Cucchiara R, et al. *Visual tracking: an experimental survey*. IEEE Trans Pattern Anal Mach Intell 2013; 36(7): 1442–1468.

[8] Comaniciu D, Ramesh V and Meer P. *Kernel-based object tracking. IEEE Trans Pattern Anal Mach Intell 2003*; 24(5): 564–575.

[9] Li Y, Ai H, Yamashita T, et al. *Tracking in low frame rate video: a cascade particle filter with discriminative observers of different life spans*. IEEE Trans Pattern Anal Mach Intell 2008; 30(10): 1728–1740.

[10] Kwon J and Lee KM. *Tracking by sampling and integrating multiple trackers.* IEEE Trans Pattern Anal Mach Intell 2013; 36(7): 1428–1441.

[11] Wang D, Lu H and Yang MH. *Online object tracking with sparse prototypes.* IEEE Trans Image Process 2012; 22(1): 314–325.

[12] Ahmad M, Khan AM, Mazzara M, et al. *Multi-layer extreme learning machine-based autoencoder for hyperspectral image classification. In*: Proceedings of the 14th international conference on computer vision theory and applications (VISAPP'19), *Prague, Czech Republic, February 2019, pp.25–27.*

[13] Yao R, Shi Q, Shen C, et al. *Part-based visual tracking with online latent structural learning.* In: 2013 IEEE conference on computer vision and pattern recognition, Portland, OR, 23–28 June 2013, pp.2363–2370. New York: IEEE.

[14] Santner J, Leistner C, Saffari A, et al. *Prost: parallel robust online simple tracking*. In: 2010 IEEE conference on computer vision and pattern recognition, San Francisco, *CA, 13–18 June 2010, pp.723–730. New York: IEEE.*

[15] Gall J, Yao A, Razavi N, et al. *Hough forests for object detection, tracking, and action recognition*. IEEE Trans Pattern Anal Mach Intell 2011; 33(11): 2188–2202.

[16] Hare S, Golodetz S, Saffari A, et al. *Struck: structured output tracking with kernels*. IEEE Trans Pattern Anal Mach Intell 2015; 38(10): 2096–2109.

[17] Yang F, Lu H, Zhang W, et al. *Visual tracking via bag of features. IET Image Process* 2012; 6(2): 115–128.

[18] Dalal N and Triggs B. *Histograms of oriented gradients for human detection.* In: 2005 IEEE conference on computer vision and pattern recognition (CVPR'05), vol. 1, San Diego, CA, 20–25 June 2005, pp.886–893. New York: IEEE.

[19] Lu Y, Wu T and Chun Zhu S. *Online object tracking, learning and parsing with and-or graphs.* In: 2014 IEEE conference on computer vision and pattern recognition, Columbus, OH, 23–28 June 2014, pp.3462–3469. New York: IEEE.

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com

International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 8, Special Issue 1, August 2021

[20] Fan J, Shen X and Wu Y. *Scribble tracker: a mattingbased approach for robust tracking.* IEEE Trans Pattern Anal Mach Intell 2011; 34(8): 1633–1644.

[21] Li X, Dick A, Shen C, et al. *Incremental learning of 3DDCT compact representations for robust visual tracking*. IEEE Trans Pattern Anal Mach Intell 2012; 35(4): 863–881.

[22] Khan FA, Shaheen S, Asif M, et al. *Towards reliable and trustful personal health record systems: a case of clouddew architecture based provenance framework*. J Amb Intel Hum Comp 2019; 10(10): 3795–3808.

[23] Ahmad M, Protasov S, Khan AM, et al. *Fuzziness-based active learning framework to enhance hyperspectral image classification performance for discriminative and generative classifiers*. PLoS One 2018; 13(1): e0188996.

[24] Bewley A, Ge Z, Ott L, et al. *Simple online and realtime tracking. In: 2016 IEEE international conference on image processing (ICIP),* Phoenix, AZ, 25–28 September 2016, pp.3464–3468. New York: IEEE.

[25] Ahmed I, Ahmad A, Piccialli F, et al. *A robust featuresbased person tracker for overhead views in industrial environment*. IEEE Internet Things J 2017; 5(3): 1598–1605.

[26] Misbah Ahmad1 , Imran Ahmed1, Fakhri Alam Khan1 , Fawad Qayum2 and Hanan Aljuaid3., *Convolutional neural network–based person tracking using overhead views* International Journal of Distributed Sensor Networks, 2020, Vol. 16(6), The Author(s) 2020 DOI:10.1177/1550147720934738

[27] Susheel George Joseph, *"Co-Operative Multiple Replica Provable Data Possession for Integrity Verification in Multi-Cloud Storage",* Research Inventy: International Journal of Engineering And Science Vol.4, Issue 5 (May 2014), PP 26-31 ISSN (e): 2278-4721, ISSN (p):2319-6483, http://www.researchinventy.com/papers/v4i5/E045026031.pdf

[28] *Susheel George Joseph, "The Usage of Machine Learning Evolutionary Algorithms in Medical Images Formed by Computer Tomography* (CT) or X-Rays to Detect the Infections due to COVID 19", PENSEE (penseeresearch.com) ISSN: 0031-4773. Volume 51, Issue 4, Page No:1512-1518, April 2021. Available at: https://app.box.com/s/n6nsv8myosb0wb16psvtb8conekpy ohj

[29] Susheel George Joseph, *"A Machine Learning (ML) Modelling Approach in Monitoring and Controlling the Viral Pandemic-COVID 19",* International Journal of Emerging Technologies and Innovative Research (www.jetir.org), ISSN:2349-5162, Vol.7, Issue 6, page no.1709-1717, June 2020: http://www.jetir.org/papers/JETIR2006575.pdf

[30] Susheel George Joseph, Dr. Vijay Pal Singh, *"Denoising of Images using Deep Convolutional Neural Networks (DCNN)",* International Journal of Engineering Development and Research (IJEDR), ISSN:2321-9939, Volume.7, Issue 3, pp.826-832, September 2019, http://www.ijedr.org/papers/IJEDR1903143.pdf.

[31] Susheel George Joseph, "*Lossless Compression of Medical Images using Huffman Algorithm with 3D Predictors and Masking",* International Journal of Emerging Technologies and Innovative Research (www.jetir.org), ISSN:2349-5162, Vol.5, Issue 6, page no.536-540, June-2018, Available : http://www.jetir.org/papers/JETIR1806468.pdf