IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Convolutional Neural Network Based use Surveillance Videos for Recognizing Human Actions Based on Machine Learning.

Dipak Daitkar[1], Divyesh Patil[2], Akshay Desai[3], Prasad Kawade[4], Prof. M.R. Bendre[5]

[1,2,3,4] Students & [5]Asst. Prof. of Department of Computer Engineering,

Savitribai Phule Pune University, Pune, Maharashtra, India

Abstract: Data analytics is the method of processing an image, gathering data, and analyzing the data for getting domain-specific information. In the current trend, besides analyzing any image for information retrieval, analyzing live surveillance videos for detecting activities that take place in its coverage area has become more important. Such systems will be implemented in real-time. Automated face recognition from surveillance videos becomes easier while using a training model such as Artificial Neural Network. Hand detection is assisted by skin color estimation. This research work aims to detect suspicious activities such as object exchange, entry of a new person, peeping into another's answer sheet, and person exchange from the video captured by a surveillance camera during examinations. Nowadays, people pay more attention to the fairness of surveillance, so it is meaningful to detect abnormal behavior to ensure the order of criminal identification. Most current methods propose models for particular abnormal behavior. In this system, we extract the optical flow of video data and propose a 3D convolution neural networks model to deal with the problem. This requires the process of face recognition, criminal recognition, and detecting the contact between the face of the same person and that among different persons. Automation of "criminal identification & detection" will help decrease the error rate due to manual monitoring.

Index Terms - Video Surveillance, Anomaly detection, Artificial neural network based sparsity learning, Criminal Identification, etc

I. INTRODUCTION

In Human face and human behavioral pattern play an important role in person identification? Visual information is a key source for such identifications. Surveillance videos provide such visual information which can be viewed as live videos, or it can be played back for future references. The recent trend of 'automation' has its impact even in the field of video analytics. Video analytics can be used for a wide variety of applications like motion detection, human activity prediction, person identification, abnormal activity recognition, vehicle counting, people counting at crowded places, etc. In this domain, the two factors which are used for person identification are technically termed as face recognition and gait recognition respectively. Among these two techniques, face recognition is more versatile for automated person identification through surveillance videos. Face recognition can be used to predict the orientation of a person's head, which in turn will help to predict a person's behavior. Motion recognition with face recognition is very useful in many applications such as verification of a person, identification of a person and detecting presence or absence of a person at a specific place and time. In addition, human interactions such as subtle contact among two individuals, head motion detection, hand gesture recognition and estimation are used to devise a system that can identify and recognize suspicious behavior among pupil in an examination hall successfully.

This system provides a methodology for suspicious human activity detection through face recognition. Video processing is used in two main domains such as security and research. Such a technology uses intelligent algorithms to monitor live videos. Computational complexities and time complexities are some of the key factors while designing a real-time system. The system which uses an algorithm with a relatively lower time complexity, using less hardware resources and which produces good results will be more useful for time-critical applications like bank robbery detection, patient monitoring system, detecting and reporting suspicious activities at the railway station, exam holes etc.

II. RELATED WORK

- Bin Zhou, Li Fei-Fei, Eric P. Xing "Online Detection of Unusual Events in Videos via Dynamic Sparse Coding". Author propose an improved Real-time unusual event detection in video stream has been a difficult challenge due to the lack of sufficient training information, volatility of the definitions for both normality and abnormality, time constraints, and statistical limitation of the fitness of any parametric models. They propose a fully unsupervised dynamic sparse coding approach for detecting unusual events in videos based on online sparse reconstructibility of query signals from an atomically learned event dictionary, which forms sparse coding bases. Based on an intuition that usual events in a video are more likely to be constructible from an event dictionary, whereas unusual events are not, our algorithm employs a principled convex optimization formulation that allows both a sparse reconstruction code, and an online dictionary to be jointly inferred and updated.
- Mohammad Sabokrou, Mahmood Fathy, Mojtaba Hoseini, Reinhard Klette, "Real-Time Anomaly Detection and Localization in Crowded Scenes". In this paper, we propose a method for real-time anomaly detection and localization in crowded scenes. Each video is defined as a set of non-overlapping cubic patches, and is described using two local and global descriptors. These descriptors capture the video properties from different aspects. By incorporating simple and cost-effective Gaussian classifiers, we can distinguish normal activities and anomalies in videos. The local and global features are based on structure similarity between adjacent patches and the features learned in an unsupervised way, using a sparse auto encoder.
- Cewu Lu. Jianping Shi, Jiaya Jia, "Abnormal Event Detection at 150 FPS in MATLAB". Speedy abnormal event detection meets the growing demand to process an enormous number of surveillance videos. Based on inherent redundancy of video structures, we propose an efficient sparse combination learning framework. It achieves decent performance in the detection phase without compromising result quality. The short running time is guaranteed because the new method effectively turns the original complicated problem to one in which only a few costless small-scale least square optimization steps are involved. Our method reaches high detection rates on benchmark datasets at a speed of 140150 frames per second on average when computing on an ordinary desktop PC using MATLAB.

- Mahmudul Hasan, Jonghyun Choiy, Jan Neumanny, Amit K. Roy-Chowdhury, Larry S. Davisz, "Learning Temporal Regularity in Video Sequences". Perceiving meaningful activities in a long video sequence is a challenging problem due to ambiguous definition of 'meaningfulness' as well as clutters in the scene. We approach this problem by learning a generative model for regular motion patterns (termed as regularity) using multiple sources with very limited supervision. Specifically, we propose two methods that are built upon the auto encoders for their ability to work with little to no supervision. They first leverage the conventional handcrafted spatio temporal local features and learn a fully connected auto encoder on them. Second, we build a fully convolutional feed-forward auto encoder to learn both the local features and the classifiers as an end-to-end learning framework.

III. PROPOSED SYSTEM

In this system, we extract the optical flow of image data and propose a Convolution Neural Networks model to deal with the problem. The proposed system extracts the spatial and temporal features from image data and these features can be directly feed into the classifier for model learning or inference. The experiments on our own made dataset show that the proposed model achieves superior performance in comparison to current methods.

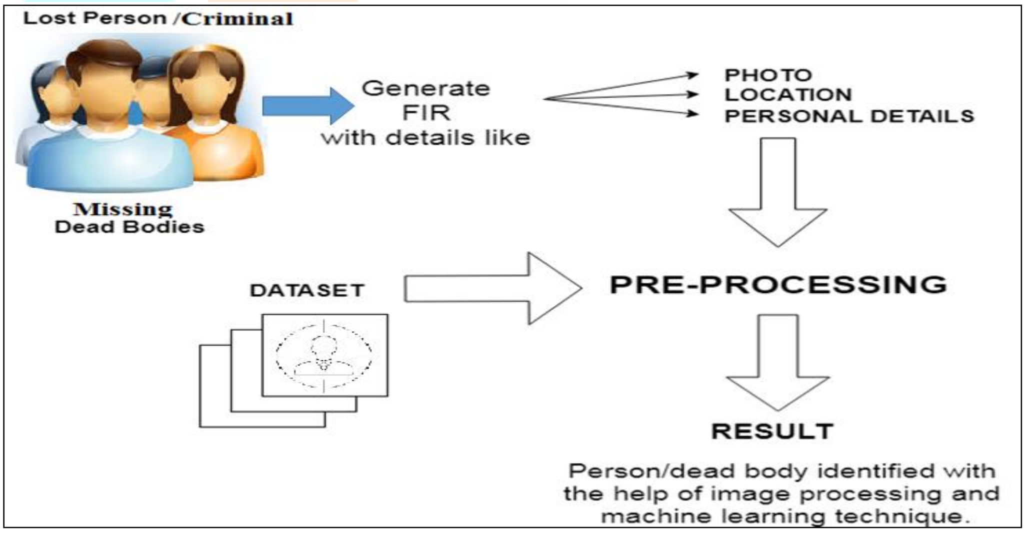

Fig.1: Proposed System Architecture

There is an abnormal increase in the crime rate and also the number of criminals is increasing, this leads towards a great concern about the security issues. Crime preventions and criminal identification are the primary issues before the police personnel, since property and lives protection are the basic concerns of the police but to combat the crime, the availability of police personnel is limited. With the advent of security technology, cameras especially CCTV have been installed in many public and private areas to provide surveillance activities. The footage of the CCTV can be used to identify suspects on scene. In this paper, an automated facial recognition system for criminal database was proposed using known Haar feature-based cascade classifier. This system will be able to detect face and recognize face automatically in real time. An accurate location of the face is still a challenging task. Viola-Jones framework has been widely used by researchers in order to detect the location of faces and objects in a given image. Face detection classifiers are shared by public communities, such as OpenCV

IV. RESEARCH METHODOLOGY

A. Preprocessing:

Given a visual input (image), illumination normalization, registration and alignment between the image sequences, and face detection are typical required preprocessing steps. Other types of signals, such as speech or physiological recordings, may also need preprocessing, such as segmentation. The most popular algorithm for face detection has been proposed by Viola and Jones. Some off-the-shelf facial expression analysis applications have also been used widely as preprocessing tools, enabling researchers to focus on deriving high level information.

B. Manipulation:

This Subsection describes processes involved in feature extraction, dimensionality reduction, and fusion. The output of this processing stage generates the input to the machine learning stage, where no further manipulation of features is taking place.

C. Feature Extraction:

Feature Extraction is an important step in the processing workflow, since subsequent steps entirely depend on it. The approaches reviewed employ a wide range of feature extraction algorithms which, according to the well established taxonomy in, can be classified as a) geometry-based, or b) appearance - based. In the field of depression assessment, several features are derived from the time-series of both (a) and (b) in the form of dynamic features.

D. Face Recognition:

Features related to the face are classified here into features from full face, AUs, facial landmarks, and mouth/eyes.

V. CONVOLUTIONL NEURAL NETWORK

A simple ConvNet is a sequence of layers, and every layer of a ConvNet transforms one volume of activations to another through a differentiable function. We use three main types of layers to build ConvNet architectures: Convolutional Layer, Pooling Layer, and Fully-Connected Layer (exactly as seen in regular Neural Networks). We will stack these layers to form a full ConvNet architecture.

- INPUT [32x32x3] will hold the raw pixel values of the image, in this case an image of width 32, height 32, and with three color channels R,G,B.
- CONV layer will compute the output of neurons that are connected to local regions in the input, each computing a dot product between their weights and a small region they are connected to in the input volume. This may result in volume such as [32x32x12] if we decided to use 12 filters.
- RELU layer will apply an element wise activation function, such as the max(0,x) thresholding at zero. This leaves the size of the volume unchanged ([32x32x12]).
- POOL layer will perform a down sampling operation along the spatial dimensions (width, height), resulting in volume such as [16x16x12].
- FC (i.e. fully-connected) layer will compute the class scores, resulting in volume of size [1x1x10], where each of the 10 numbers correspond to a class score, such as among the 10 categories of CIFAR-10. As with

ordinary Neural Networks and as the name implies, each neuron in this layer will be connected to all the numbers in the previous volume.

VI. RESULTS AND DISCUSSION

The result for proposed system is to identify criminal using image processing (CNN), detect & identification of crime scene section and study extract the optical flow of video data and propose a CNN model to deal with the problem.

Methods	Accuracy	Precision	Recall Rate
Motion Blob[4]	-	-	82%
Template Matching[5]	-	86%	-
Skin+SVM[7]	84%	-	-
Our Method1	87.6%	80.4%	84.3%
Our Method2	89.8%	86.5%	83.2%

Fig.2: Performance of Different Methods

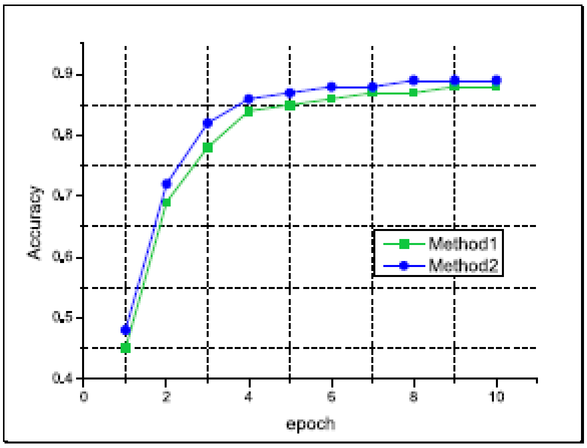

Fig.3: The performance of two "flow image" on CNN (ROC)

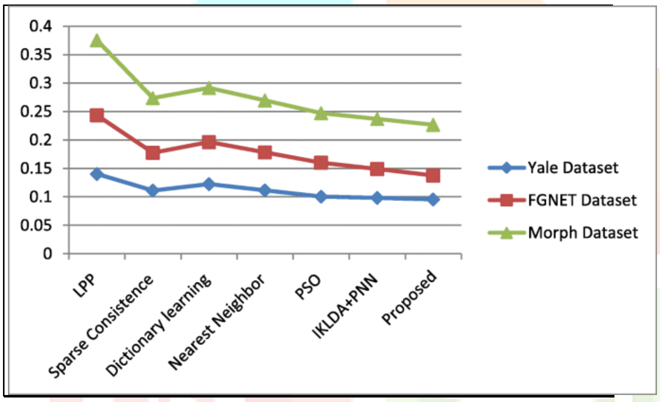

Fig.5: A hybrid feature extraction on face for efficient face recognition

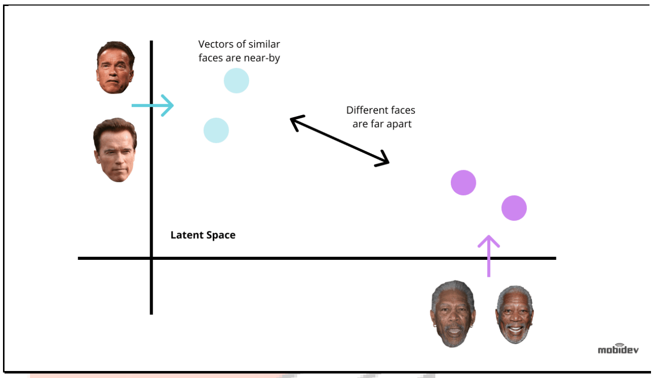
Vectors of similar faces are near-by
Different faces are far apart
Latent Space
mobidev

Fig.4: Face Recognition Vector

VII. CONCLUSION

We propose a unified deep learning based framework for criminal event detection from CCTV surveillance. The proposed system consists of three blocks which are designed to achieve three keys of criminal identification in neural networks. Real-world crime scene events are complicated and diverse. It is difficult to list all of the possible anomalous events. Therefore, it is desirable that the crime scene/Dead body identification detection algorithm does not rely on any prior information about the events. In other words, anomaly detection should be done with minimum supervision. Sparse-coding based approaches are considered as representative methods that achieve state-of-the-art anomaly detection results. These methods assume that only a small initial portion of a video contains normal events, and therefore the initial portion is used to build the normal event dictionary. Then, the main idea for crime scene detection is that anomalous events are not accurately reconstructed able from the normal event dictionary. However, since the environment captured by surveillance cameras can change drastically over the time (e.g. at different times of a day), these approaches produce high false alarm rates for different normal behaviors.

REFERENCES

[1] C. Lu, J. Shi, and J. Jia, “Abnormal event detection at 150 fps in matlab,” in Proceedings of the IEEE international conference on computer vision, 2013, pp. 2720–2727.

[2] M. Sabokrou, M. Fathy, M. Hoseini, and R. Klette, “Real-time anomaly detection and localization in crowded scenes,” in The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops, June 2015

[3] M. Hasan, J. Choi, J. Neumann, A. K. Roy-Chowdhury, and L. S. Davis, “Learning temporal regularity in video sequences,” in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 733–742.

[4] B. Zhao, L. Fei-Fei, and E. P. Xing, “Online detection of unusual events in videos via dynamic sparse coding,” in Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on. IEEE, 2011, pp. 3313– 3320

[5] W. Luo,W. Liu, and S. Gao, “A revisit of sparse coding based anomaly detection in stacked rnn framework,” in The IEEE International Conference on Computer Vision (ICCV), Oct 2017.

[6] S. Wu, B. E. Moore, and M. Shah, “Chaotic invariants of lagrangian particle trajectories for anomaly detection in crowded scenes,” in Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on. IEEE, 2010, pp. 2054–2060.

[7] V. Mahadevan,W. Li, V. Bhalodia, and N. Vasconcelos, “Anomaly detection in crowded scenes,” in Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on. IEEE, 2010, pp. 1975–1981.

[8] N. Dalal and B. Triggs, “Histograms of oriented gradients for human detection,” in Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on, vol.1. IEEE, 2005, pp. 886–893.

[9] L. Kratz and K. Nishino, “Anomaly detection in extremely crowded scenes is using spatio-temporal motion pattern models”, in Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on. IEEE, 2009, pp. 1446–1453.

[10]N. Dalal, B. Triggs, and C. Schmid, “Human detection using oriented histograms of flow and appearance,” in European conference on computer vision Springer, 2006, pp. 428–441. G.