Crowd Analysis System for Images of CCTV Camera

Vishakha L. Bansod, Asha Ambhaikar

Abstract: It is important to monitor and analyze the crowd for peaceful event organization and minimum number of persons killed or injured in a stampede, war or accident in public or religious places. Ultimately, crowd management strategies should be adopted for crowd safety. Detecting the abnormal crowd behaviour is the interesting research area. Traditionally to monitor the crowd closed circuit television is used. Video analyses are used in many application areas. This work is a paper of traditional approach and is a study of different attributes of crowd which include crowd counting and estimation of density, detection of crowd motion. It also presents the study about behaviour understanding of crowd and crowd tracking. We have also presents crowd analysis using different deep learning methods. This paper showed results of crowd analysis using convolutional neural network and different recurrent neural network training model.

Keywords: Abnormal Behaviour, Convolution Neural Network (CNN), Crowd Analysis, Deep Learning, Recurrent Neural network (RNN),

I. INTRODUCTION

A large group of individuals who have gathered closely together is referred as a Crowd. It is defined in different ways according to the situations. For example, when more than 100 people gathered together in India (or such dense populated countries) may recognize a crowd. But recognition of crowd in the countries where population is thin like Canada a crowd is define as a group of twenty people. Hence a crowd is differing from community to community and thus become difficult to define uniquely.

People may gathered in a places such as markets, temples, stadiums, subways, religious festivals, public demonstrations, procession, concerts, football/cricket matches, races, sport events, etc. It is important to monitor and analyze the crowd for peaceful event organization and minimum number of persons killed or injured in a stampede, war or accident in public or religious places. Ultimately, crowd management strategies should be adopted for crowd safety. Different attributes are used to analyze the crowd. These attributes are crowd behavior detection, crowd density estimation, crowd counting and crowd scene analysis [32][10]. The abnormal behaviour of crowd and its analysis is now a research area of the researcher [8]. Crowd analysis can be carried out in three steps: i) pre-processing ii) object tracking and iii) event and behavior recognition.

Segmentation and background subtraction are used for pre-processing. Object tracking is categorized as i) tracking individual object and ii) tracking group of object. Histogram of oriented gradient are used to analyze the behavior of tracked object. Optical flow combines with hidden Markov Models to recognize object behavior. The respective result from above traditional analysis concept takes much time. Deep learning method, especially CNN [32][10] and RNN for crowd behavior analysis is used. There are a lot of algorithms that have been used for image classification before CNN. Users used to create features from images and then feed those features into some classification algorithm like SVM. Automatic features can be extracted from the image using CNN. While if the algorithm is used by a pixel vector, it loses most of the spatial interaction between the pixels, but CNN effectively uses the adjacent pixel information to accurately down sample image by convolution and then finally uses the prediction layer. The great advantage of RNN is that through this connection the network is able to refer to the end states and can therefore process arbitrary sequences of inputs. The disadvantages of RNN are the amount of training required which is comparatively much more than other kinds of networks and the issue of vanishing gradients where gradients tend to slowly disappear as back-propagation across multiple unfolding of the RNN layer. To analyze the crowd behavior using Different object detection model of CNN such as InceptionV4 and Inception ResNets [6] with different RNN training model [21] can be used.

II. CCTV VIDEO ANALYSIS

Closed Circuit Television (CCTV) is used as a traditional method of monitoring the crowd. CCTV surveillance is useful technique to monitor the crowd, but its required human visual ability to monitor the crowd. To estimate the size and density of crowd some system is needed that is Crowd Analysis. To determine the number and type of people crowd analysis is very useful. Using cameras we can analyze and understand the density of the crowd and behaviour patterns of the crowd. Uninterrupted watch through cameras is adapted by safety department. The great amount of video can be capture using CCTV cameras. A person cannot watch whole video data. Sometime the important part of video is never watch or gone over again, because of, in relation to existence without of time or useable materials. Sometimes suspicious behaviour will be missed and accident may occur

A. Video analysis: Application areas identified
Video analytics are used in many application areas.

1. People Counting.
2. Heat Map
3. Dwell time
4. Crowd detection
Crowd Analysis System for Images of CCTV Camera

5. Queue Management 6. Vehicle counting
7. Forbidden Direction 8. Object Counting
9. Motion Detection.

The above listed area the crowd analysis is not easy part. All types of movement, behaviour and action are recognized.

III. CROWD ANALYSIS USING TRADITIONAL APPROACH

The crowds can be analyzed through different attributes which include crowd counting and estimation of density, detection of crowd motion [32]. The common process of crowd analysis consists of three steps:

- **Preprocessing**: The features extraction, object-detection and classification can be done in this steps. Segmentation, background subtraction are used for preprocessing.
- **Object Tracking**: It can be divided into four types a) Region-based object tracking b) Active contour-based object tracking c) Feature-based object tracking d) Model-based object tracking. Model-based object tracking can be done using mean shift function.
- **Event/Behavior Recognition**: Tracking using HOG (Histogram of oriented gradients) are used to analyze the behaviour of tracked object. Optical flow with HMM (Hidden Markov models) is used to recognize the object behaviour.

A. Crowd behaviour analysis

In this paper [5] latest progress of algorithms are discussed in which the behaviour recognition for transition view monitoring applications. Here video classified into four classes: 1) People which not interacting with other 2) People which interacting with other 3) People to vehicle interaction 4) people to place interaction.

This paper presented preprocessing steps as new techniques. This preprocessing steps increase the consistency of behaviour detection. For identification of behaviour no standard methods are used, but used the evaluation tools for identification.

KLT tracker has been used for feature extraction, detection and tracking on multiple frame and real crowd scenes video. This model detects the behaviour of crowd of individual camera and gives better results of detection. [23]

This paper [27] identifies behaviour of crowd using optical flows technique with local analysis and without used of object detection and tracking. This approach classified real video from web and classified into five classes using Jacobian matrix. The disadvantage of this method is that it is not applicable for that video contains overlapping motion patterns.

B. Crowd density estimation

Crowd density estimation is used to analyze the crowd. Using optical flow with backgrounds subtraction methods to removed the background, then new features is removed from foreground image using texture analysis and finally neural network is used classify the video[17].

Discrete cosine transformation (DCT) is used to transform crowd model into frequency domain [13]. Background differencing method is adapted for feature extraction. Then using support vector machine (SVM) to classify the video into five classes: i) very low, ii) low, iii) moderate, iv) high, v) very high density.

A new method of high density crowd segmentation and density estimation is explained in this paper [39]. In this paper, three concepts are discussed: intra crowd motion, accumulated mosaic image difference method and statistical analysis method. To detect the density of crowd a geometry module with AMID is used. This approach gives better result for any crowded area, but low computational complexity for real-time monitoring.

Higher order singular value decomposition (HOSVD) combine with support vector machine (SVM) method is discussed in this paper [40] to estimate the crowd density. HOSVD method consist of different steps: i) Gaussian bluring approach as the noise removal from image ii) HOSVD approach for feature extraction iii) SVM classifier to classify the crowd density.

Clustering motion cues approach to estimates the crowd density gives accurate results [22]. Using the above approach the mean error was found. Many existing methods for density estimation use the training which is supervised.

C. Crowd counting

In this paper [28], crowd counting using head detection. Interested points can be calculated by using gradient orientation, and then background subtraction applied. Finally that simple morphological operation used to fill the gaps after foreground segmentation. The adaboost classifier used as a head detector. This approach is efficient to estimate the crowd from single frame. To minimize the false detection this approach is not efficient.

Crowd counting can be done by different strategies i) detection method ii) clustering method ii) Regression method [18]. The existing regression method for crowd counting is differing than the detection and clustering method in both data evaluation and experimental setting. It is very difficult to draw a conclusion that the regression method is more effective than other two crowd counting strategies.

D. Crowd motion Estimation

The min cut/max flow algorithm is used for crowd motion segmentation. Here, correlation among neighboring blocks is computed and subsequently used for multi-labeling optimization [34]. The segmentation allows for an efficient identification of anomalies, detected as deviations from the regular flow stored in the crowd motion model. This method performs effectively for crowd motion segmentation as well as for anomaly detection.

To estimate the motion of the crowded scene [33] corner features method is used. In this method, corner features on the scene initialize first then apply optical flow method. Using enthalpy measure approximating the moving corner features, interaction forces are computed, then corner features of interest is obtaining.

These features are exploited to extract the orientation patterns, used as input for training a random forest.

Crowd behavior analysis has been briefly in the areas of computer vision and pattern recognition. Many surveys have been conducted on behaviour and characteristics of crowds.
The following is a summary of the surveyed papers for crowd analysis using conventional methods and deep neural networks [32].

Table I: Different CNN Model for Crowd Analysis

<table>
<thead>
<tr>
<th>Ref.</th>
<th>Year</th>
<th>Areas discussed</th>
<th>Shortfalls</th>
</tr>
</thead>
<tbody>
<tr>
<td>[14]</td>
<td>2004</td>
<td>Focus on the tasks such as detection, tracking, understands the behaviour, identification of person from multiple cameras.</td>
<td>Occlusion handling, fusion of 2D tracking, anomaly detection and prediction of behaviour, a fusion of data from multiple cameras</td>
</tr>
<tr>
<td>[24]</td>
<td>2011</td>
<td>Discusses the pre-processing, motion detection, tracking methodologies using temporal segmentation and semantic description.</td>
<td>Interaction between group of individual are lacking</td>
</tr>
<tr>
<td>[26]</td>
<td>2012</td>
<td>Detect the crowd behaviour by measuring stillness level of crowd.</td>
<td>Multiple resolutions, and their methods are lacking</td>
</tr>
<tr>
<td>[1]</td>
<td>2015</td>
<td>Automatic human behaviour detection from video. Computational detection techniques and public datasets and its application.</td>
<td>Multiple features fusion and dynamic scenes feature vectors fusion through various methods are lacking</td>
</tr>
<tr>
<td>[16]</td>
<td>2015</td>
<td>Discusses crowd scene analysis using three aspects: segmentation of motion pattern, recognize the behaviour of crowd and anomaly detection.</td>
<td>Multi-sensor feature fusion (MSFF) is missing in the discussion</td>
</tr>
<tr>
<td>[29]</td>
<td>2016</td>
<td>Discusses crowd behaviour on density-based analysis.</td>
<td>Real-time crowd analysis, social force models, and static crowd models are lacking</td>
</tr>
<tr>
<td>[15]</td>
<td>2016</td>
<td>Detecting the crowd behaviour using the the concepts of physics and biology. Physics concepts like optimization is used to describe the crowd behaviour.</td>
<td>Deep learning methods for crowd behaviour analysis are lacking</td>
</tr>
</tbody>
</table>

IV. CROWD ANALYSIS USING DEEP NEURAL NETWORK

But the crowd analysis using traditional methods are not so efficient in selecting good features. The above traditional crowd analysis concept takes lot of time to present their respective results. Deep learning concepts can be solved these issues. The evolution of artificial intelligence from rule based system to automatically feature identification passes machine learning, representation and finally deep learning. Object detection can be achieved using two approaches, machine learning approaches and deep learning approaches.

Table II: Deep learning methods

<table>
<thead>
<tr>
<th>Ref.</th>
<th>Year</th>
<th>Title</th>
<th>Deep learning model</th>
</tr>
</thead>
<tbody>
<tr>
<td>[36]</td>
<td>2017</td>
<td>Scene-independent crowd analysis. In: Group and crowd behavior for computer vision</td>
<td>Slicing CNN</td>
</tr>
<tr>
<td>[25]</td>
<td>2017</td>
<td>Deep appearance features for abnormal behavior detection in video.</td>
<td>Pre-trained CNN with one-class SVM</td>
</tr>
<tr>
<td>[12]</td>
<td>2017</td>
<td>Joint detection and recounting of abnormal events by learning deep generic knowledge</td>
<td>Multi-task learning for fast R-CNN</td>
</tr>
</tbody>
</table>

Table II summarizes the various CNN-based models for crowd analysis. Due to big amount of data, feature extraction with a standard fully-connected (FC) network would be very inefficient. Inception network is very deep neural network with high accuracy on image recognition.

A. Crowd analysis using CNN object model and different RNN training model

To classify tremendous amount of videos based on what occur inside, it is not possible to hire the people to sit in front of the computer and do this job. There are both spatial and temporal content to be considered. Video contain lots of images viewed one after other and each frames has order, to capture the order between frames utilization of recurrent neural network for better classification. RNN have problem of long-term dependency, RNN can look back and get information. It can predict the information but not every time because sometimes it is easy to predict and sometime they do require a context to predict a specific word. New type of RNN called Long- short term memory network (LSTM) does not have these problems.

A fully automated deep model learns to classify human actions without using any prior knowledge. The first step is to build 3D convolutional neural network learning based on spatial features. It is then trained to classify each sequence, taking into account the temporal development of the learned features of the recurrent nervous system network. [19].

A deep neural network model is used for addressing the problem of violence detection in videos. The model consists of a convolutional neural network (CNN) for frame level feature extraction to integrate features into the temporal domain.
using convolutional long-term memory [30]. The convLSTM model is capable of producing a better video representation with fewer parameters than LSTM, thereby avoiding over fitting.

Long-term convolution network LRCN [7] models are used both locally and temporally, and are flexible enough to be applied to a variety of vision tasks, including sequential input and output. The LRCN model with deep sequences learns sequential dynamics, and this model represents a fixed view of the input and learns only the dynamics of the output sequences.

B. Crowd video classification methods using CNN & RNN

Different video classification methods using CNN & RNN are:

- Classifying one frame at a time with a Convolution Neural Network.
- Using a time-distributed convolutional network and passing the features to an RNN, in one network
- Using a 3D convolutional network
- Extracting features from each frame with a convolutional neural network and passing the sequence to a separate RNN
- Extracting features from each frame with a convolutional neural Network and passing the sequence to a separate MLP.

1. Classify one frame with CNN at a time

The physical features of the video can be ignored, and trying to categorize each clip by viewing a single frame. Use the Inception V3 model here for feature information that achieves 90% accuracy.

2. In a network, use time-distributed CNNs by sending features to the RNN

The temporal features of video can be consideration. The first neural network is convolution neural network as time distributed wrapper, which allows us to distribute layers of a CNN across an extra dimension time. For the RNN part of the net, use a three-layer GRU, each consisting of 128 nodes, and a 0.2 dropout between each layer that model achieved 41% accuracy.

3. Use a 3D convolutional network

3D Convolution Networks are used for video classification. Here apply convolutions and max pooling in 3D space. Such type of model achieved 51% accuracy in video classification.

4. Extract features with CNN, assign sequence to different RNN

In this method, the first neural network is the Inception network to extract features from our videos, and then pass those to a separate RNN with LSTM achieved 91% accuracy. The RNN easily defeats the CNN-only classification method.

5. Extract the features from each frame with the CNN and assign the sequence to the MLP

In this method apply the same CNN extraction process as in the previous method, but instead of sending each piece of the sequence to an RNN, flatten the sequence and pass into a fully connected network, a multilayer perception (MLP) achieved 88% accuracy in video classification.

V. RESULT

The above video classification methods are implemented in keras & tensorflow with exploring the UCF101 video action dataset. Each video subsample down to 40 frames.

Steps of video classification methods
1) Pre-processing the given video.
2) Split all the video into train/test folders
3) Extract frame from input video
4) Apply different CNN model.
5) Apply RNN model.
6) Classification result.

The below graph using matlplotlib includes three series:
1. Red color shows the CNN top 1 accuracy.
2. Green color shows top 5 categorical accuracy.
3. Blue color shows the top 1 categorical accuracy.
Fig5: Classification using CNN and MLP

Table III: Accuracy of Different Methods

<table>
<thead>
<tr>
<th>S.N.</th>
<th>Models</th>
<th>Methodology</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Classify one frame with CNN at a time</td>
<td>Inception v3</td>
<td>90%</td>
</tr>
<tr>
<td>2</td>
<td>In a network, use time-distributed CNNs by sending features to the RNN</td>
<td>VGG-16 as a convolutional neural network and three layer GRU as a Recurrent neural network</td>
<td>41%</td>
</tr>
<tr>
<td>3</td>
<td>Use a 3D convolutional network</td>
<td>3D convolutional network with 32,64 and 128 nodes</td>
<td>51%</td>
</tr>
<tr>
<td>4</td>
<td>Extract features with CNN, assign sequence to different RNN</td>
<td>Inception network to extract the features and RNN with LSTM layer followed by dense layer</td>
<td>91%</td>
</tr>
<tr>
<td>5</td>
<td>Extract the features from each frame with the CNN and assign the sequence to the MLP</td>
<td>Inception network and multilayer perceptron with two layer</td>
<td>88%</td>
</tr>
</tbody>
</table>

The result of the above table shows that Inception convolutional networks to extract features followed by a single RNN with LSTM layer, dense layer with some dropout in between achieved great accuracy in video classification.

VI. CONCLUSION

A brief introduction to CCTV video analysis is presented. Firstly, crowd analyses through different attributes using traditional approach are discussed. The CNN based models for crowd analysis are discussed. Finally, we presented crowd analysis using CNN object model and different RNN training models. The Inception Convolution Networks to extract features followed by a single-layer LSTM RNN achieved great accuracy in video classification.

REFERENCES

30. Swathikiran Sudhakaran , Oswald Lanz (2017) Learning to Detect Violent videos using convolutional long short-term memory, Available...
Crowd Analysis System for Images of CCTV Camera


AUTHORS PROFILE

Vishakha Lahu Bansod received her B. E degree in Computer Technology from Nagpur University, India, in 2000, and M. Tech. degree in Computer Science & Engineering from Nagpur University, India, in 2010, and is currently pursuing Ph.D. in Kalinga University, Raipur, India. She is currently working as an Assistant Professor in RCERT, Chandrapur affiliated to DBATU University, Lonere. She has published more than 10 research papers in reputed National and International Journals. She was a member of Board of studies of Computer Science & Engineering, Gondwana University, Gadchiroli. Her research interest includes deep learning, video processing and image processing.

Dr. Prof. Asha Ambhaikar is working as a Professor and Dean Students Welfare at Kalinga University, Naya Raipur. She has also worked as a Principal in G. H. Raisoni college of Engineering and Management, Amravati (Maharashtra). She has published more than 75 research papers in reputed National and International Journals. She was a chairman Board of studies and Member of Academic Council of Information Technology in Chhattisgarh Swami Vivekananda Technical University, Bhilai(C.G.). She is a member of Editorial Board and Reviewer of various Reputed international journal’s and conferences. She has also published two books by Lambert Publication, Germany. She has also received a various Awards like: 1. Best Personality of India 2015 at New Delhi, India. 2. Bharat Excellence Award 2015 at New Delhi, India. ” Her area of research includes Computer Networking, Mobile Adhoc Networking, Sensor Networks, Data Mining, Distributed system, information systems and security and Cloud Computing etc.

Retrieval Number: E6283018520/2020©BEIESP
DOI:10.35940/ijrte.E6283.018520
Published By:
Blue Eyes Intelligence Engineering & Sciences Publication