A NEW HYBRID DEEP LEARNING MODEL FOR HUMAN ACTION RECOGNITION

KATRA GNANESWARI [1], M. RADHIKA [2] PROF J.RAVISANKAR[3]

[1] PG SCHOLAR, DEPT OF CSE, KRISHNAVENI ENGINEERING COLLEGE FOR WOMEN, NARASARAOPET, AP, INDIA.
[2]ASST. PROFESSOR, DEPARTMENT OF CSE, KRISHNAVENI ENGINEERING COLLEGE FOR WOMEN, NARASARAOPET, AP, INDIA
[3]PROFESSOR, DEPARTMENT OF ECE, KRISHNAVENI ENGINEERING COLLEGE FOR WOMEN, NARASARAOPET, AP, INDIA

ABSTRACT: The aim of this project is to develop a model for human actions such as running, jogging, walking, clapping, hand waving and boxing. A series of videos is given for the layout, where an individual executes an event in each video. The action performed on that particular video will be the label of a video. This relationship must be learned by the model, and the label of an input (video) which he never saw can then be predicted. Technically, despite descriptions of these acts, the model would need to learn to distinguish between various human behaviors. There may be many content identification programs which can work on following jobs like Active object tracking for identifying an item such as a vehicle or a human from a CCTV picture and learning the patterns in the movement of humans when we are able to create a pattern that will guide us (humans) to perform a variety of activities.

1. INTRODUCTION

Due to progress on computer vision, computers improve on the resolution of some very difficult problems (such as understanding an image). Models are made where the model can predict what the image is or can detect whether or not a specific object is present in the image if an image is given to the model. These models are known as neural networks (or artificial neural networks) inspired by a human brain structure and function. Deep learning, a subfield of machine learning is thestudy of these neural networks, which over time have introduced several variations of these networks for various problems. For Video Recognition, this approach utilizes deep learning - in the context of a number of labeled images, a model is built so that it can generate a prediction label for a new video. Steps have been taken for execution are downloading, extracting and pre-processing a video dataset then dividing the dataset into training and testing data then creation of a neural network and train it on the training data finally testing the model on the test data.

2. LITERATURE SURVEY

1) A Large Video Database for Human Motion Recognition by H. Jhuang E. Garrote T . Poggio

With nearly one billion online videos viewed everyday, an emerging ne w frontier in computer vision researc h is recognition and searc h in video. While muc h effort has been devoted t o the collection and annotation of larg e scalable static image datasets containing thousands of image categories, human action datasets la g far behind. Cur - rent action recognition databases contain on the order of ten different action categories collected under fairly controlled conditions. State-of-the-art performance on these datasets is now near ceiling and thus there is a need for the

design and creation of ne w benchmarks. To addres s this issue we collected the largest action video database to-date with 51 action categories, whic h in total contain around 7,000 manually annotated clips extracted from a variety of sources ranging from digitized movies to YouTube . We use this database to evaluate the performance of two representative computer vision systems for action recognition and explore the robustness of these methods under various conditions such as camera motion, viewpoint, video quality and occlusion.

2) Real-world Anomaly Detection in Surveillance Videos by Waqas Sultani, Chen Chen, Mubarak Shah

Surveillance videos are able to capture a variety of realistic anomalies. In this paper, we propose to learn anomalies by exploiting both normal and anomalous videos. To avoid annotating the anomalous segments or clips in training videos, which is very time consuming, we propose to learn anomaly through the deep multiple instance ranking framework by leveraging weakly labeled training videos, i.e. the training labels (anomalous or normal) are at video-level instead of clip-level. In our approach, we consider normal and anomalous videos as bags and video segments as instances in multiple instance learning (MIL), and automatically learn a deep anomaly ranking model that predicts high anomaly scores for anomalous video segments. Furthermore, we introduce sparsity and temporal smoothness constraints in the ranking loss function to better localize anomaly during training. We also introduce a new large-scale first of its kind dataset of 128 hours of videos. It consists of 1900 long and untrimmed real-world surveillance videos, with 13 realistic anomalies such as fighting, road accident, burglary, robbery, etc. as well as normal activities. This dataset can be used for two tasks. First, general anomaly detection considering all anomalies in one group and all normal activities in another group. Second, for recognizing each of 13 anomalous activities. Our experimental results show that our MIL method for anomaly detection achieves significant improvement on anomaly detection performance as compared to the state-of-the-art approaches. We provide the results of several recent deep learning baselines on anomalous activity recognition. The low recognition performance of these baselines reveals that our dataset is very challenging and opens more opportunities for future work.

3) Learning realistic human actions from movies by Ivan Laptev; Marcin Marszalek; Cordelia Schmid; Benjamin Rozenfeld

The aim of this paper is to address recognition of natural human actions in diverse and realistic video settings. This challenging but important subject has mostly been ignored in the past due to several problems one of which is the lack of realistic and annotated video datasets. Our first contribution is to address this limitation and to investigate the use of movie scripts for automatic annotation of human actions in videos. We evaluate alternative methods for action retrieval from scripts and show benefits of a text-based classifier. Using the retrieved action samples for visual learning, we next turn to the problem of action classification in video. We present a new method for video classification that builds upon and extends several recent ideas including local space-time features, space-time pyramids and multi-channel non-linear SVMs. The method is shown to improve state-of-the-art results on the standard KTH action dataset by achieving 91.8% accuracy. Given the inherent problem of noisy labels in automatic annotation, we particularly investigate and show high tolerance of our method to annotation errors in the training set. We finally apply the method to learning

and classifying challenging action classes in movies and show promising results.

3. EXISTING SYSTEM:

In the existing work with wearable based or non-wearable based. Wearable based HAR system make use of wearable sensors that are attached on the human body. Wearable based HAR system is intrusive in nature. Non-wearable based HAR system does not require any sensors to attach on the human or to carry any device for activity recognition. Non-wearable based approach can be further categorized into sensor based HAR systems. Sensor based technology use RF signals from sensors, such as RFID, PIR sensors and Wi-Fi signals to detect human activities. Sensor based HAR system are non-intrusive in nature but may not provide high accuracy.

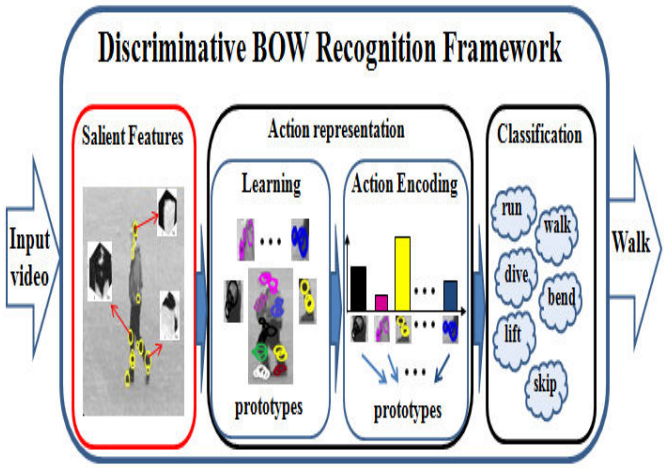
Require the optical sensors to be attached on the human and also demand the need of multiple camera settings. Wearable dives cost are high. .

4. PROPOSED SYSTEM:

The proposed System Vision based technology use videos, image frames from depth cameras or IR cameras to classify human activities. Video-based human activity recognition can be categorized as vision-based according to motion features. The vision based method makes use of RGB or depth image. It does not require the user to carry any devices or to attach any sensors on the human. Therefore, this methodology is getting more consideration nowadays, consequently making the HAR framework simple and easy to be deployed in many applications. The most common type of deep learning method is Convolutional Neural Network (CNN). CNN are largely applied in areas related to computer vision. It consists series of convolution layers through which images are passed for processing.

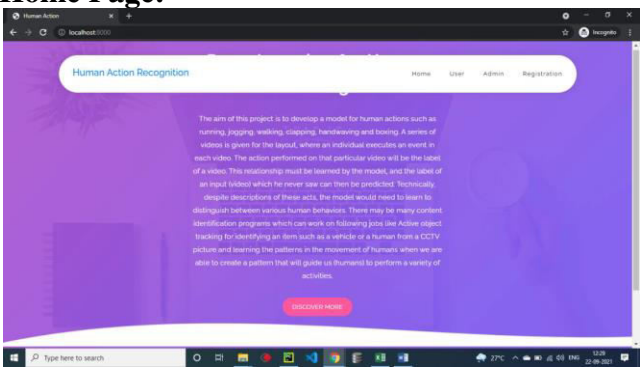
We use CNN to recognize human activities action recognition based on dataset. CNN is an efficient recognition algorithm which is widely used in pattern recognition and image processing. It has many features such as simple structure, less training parameters and adaptability. We use transfer learning to get deep image features and trained machine learning classifiers. Does not require the user to carry any devices or to attach any sensors on the human.

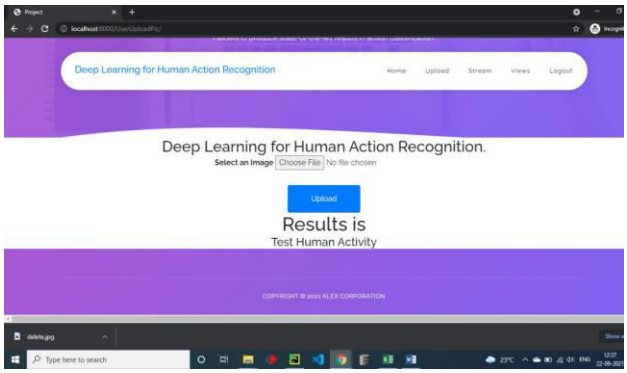
5. SYSTEM ARCHITECTURE



6. SCREEN SHOTS

Home Page:





7.CONCLUSION

High speed processors available for computing are not sufficient for processing deep learning models as the tensors of very large size created after preprocessing datasets. Already datasets used for deep learning video processing are normally inbig size so advance processing demands GPU processing with large memory. Many standard datasets are available for video analysis to validate designed model's accuracy. In recent times it is practically possible to build a good model using very high capacity and complex libraries like Keras, Theano and Torch on Python platform to make machines intelligent, the proposed model achieved nearly 20% more accuracy by preprocessing the dataset as compared to the base model.

8. FURTHER ENHANCEMENT

We carried out a comprehensive study of state-of-the-art methods of human activity recognition and proposed a hierarchical taxonomy for classifying these methods. In future, we aim to extend this study by developing the context-aware recognition system to classify human activities. Also, we will extend our work to recognize complex human activities such as cooking, reading books, and watching TV. A comprehensive review of existing human activity classification benchmarks was also presented and we examined the challenges of data acquisition to the problem of understanding human activity. Finally, we provided the characteristics of building an ideal human activity recognition system.

REFERENCES

[1] Learn Computer Vision Using OpenCV - With Deep Learning CNNs and RNNs | Sunila Gollapudi | Apress. .

[2] “Video Dataset Overview.”: https://www.di.ens.fr/~miech/datasetviz/.

[3] W. Sultani, C. Chen, and M. Shah, “Real-world Anomaly Detection in Surveillance Videos,” ArXiv180104264 Cs, Feb. 2019.

[4] I. Laptev, M. Marszalek, C. Schmid, and B. Rozenfeld, “Learning realistic human actions from movies,” in 2008 IEEE Conference on Computer Vision and Pattern Recognition, 2008, pp. 1–8, doi: 10.1109/CVPR.2008.4587756.

[5] M. Jain, MrinalJain17/Human-Activity-Recognition. 2019.

[6] “Keras vs TensorFlow vs PyTorch | Deep Learning Frameworks,” Edureka, 05-Dec-2018. https://www.edureka.co/blog/keras-vstensorflow- vs-pytorch/.