ISSN: 2278-4632
Vol-13, Issue-04, No.06, April : 2023

TRACE MORTAL PURSUIT FRAMEWORK USING AI

[1]**Dr.V.Shanmukha Rao**, Associate Professor, Department of Information Technology, Andhra Loyola Institute of Engineering and Technology, Vijayawada, Andhra Pradesh, Email: shanmukharao.v@gmail.com
[2]**Md.Imran**, Assistant Professors, Department of Information Technology, Andhra Loyola Institute of Engineering and Technology, Vijayawada, Andhra Pradesh, Email: imran02.md@gmail.com
[3]**D.S.Srinivas**, Assistant Professors, Department of Information Technology, Andhra Loyola Institute of Engineering and Technology, Vijayawada, Andhra Pradesh, Email: seshasrinivas.divi@gmail.com
[4]**D.Varun Prasad**, Associate Professor, Dept of CSE, DVR & Dr. HS MIC College of Technology, Kanchikacherla, Andhra Pradesh, India, Email: varunprasad@mictech.ac.in

ABSTRACT –
Computer interfaces novelties are noticeable to expand AI based technology. Human actions and movements are recorded, tracked, and noticed with the help of AI based expert systems. NLP, ML, and its techniques are owned for tracking of human behaviours, according to the cautions that are generated in crucial time of situations will assist the mortals. To restrain emergency situations, AI based set-up and its mode are popular in human way of life. Surveillances gadgets are required around-the-clock to track record in the form of facts that are analysed using computer visionary techniques, NLP is utilized to recognize the human behaviour speech to diagnosis the circumstances and act according to posies. RNN with LSTM techniques are exercised and examine to execute the framework.
Keywords: NLP, facts, RNN, LSTM, examine, framework.

INTRODUCTION

Is it feasible to accurately understand what is happening at a certain location when we are not physically present without watching video footage? Since we are all currently busy with other tasks, we do not have a lot of time to dedicate to watching the entire film to understand what is happening. But there is another choice for this, namely an audio clip with a person narrating the scene. The main advantage of this is we can simultaneously save time and multi-task i.e. doing our work by listening to the audio clip that is generated by getting the up-to-date information and if any person suddenly falls which may cause heavy injuries that may lead to a major medical issue for elderly people. Therefore, to prevent such emergencies, it will also feature an alarm system to detect human falls. This is made feasible by utilizing cutting-edge technology like computer vision and image processing to record live events, RNN with LSTMs to process and analyse the recorded ones, and natural language processing to provide a description of what is happening. Users receive audio clips that are created using the Google Text to Speech API. We frequently encounter CC cameras installed in and around our surroundings in daily life. They film everything that is happening in the location 24/7 but we do not have enough time to watch everything that is being recorded. Is it possible to accurately comprehend what is happening at a certain location while we are not physically present there without watching video clips? There may be occasions when you must leave for work, leaving elderly family members alone at home. In some circumstances, it can be difficult for us to monitor their situation, and they themselves might be unable to contact with us. Then there is our application called "Tracking of Human Activities," where we can monitor a variety of human behaviours, create an audio clip from the input recording, and even recognize a person's unexpected fall and send out an alert. We can listen to something that has been captured and converted to audio while driving or even while seated in a conference. This helps us save a tone of time while also letting us know what is going on in the area. Early approaches for creating picture descriptions assemble image data using the image's static object class libraries, which are then characterized using statistical language models. The query expansion method, which pulls similar images from a huge dataset and uses the distribution stated in association with the obtained photos, are some indirect approaches to solving the problem of image description that have also been presented. The common drawback of all the brainstorming techniques mentioned

DOI: 10.36893.JK.2023.V13I04N16.0017-0022

is that they neither provide an end-to-end mature general model to address this issue nor do they make intuitive feature observations on items or actions in the image. Deep neural networks with encoder-decoder components are used to generate natural language descriptions of images on-the-fly. The use of an attention-based technique to determine where to focus on the image or video has been attempted in some later works On the other hand, they continue to ignore the difference between sentences that describe low-level video elements and those that clearly express high-level video concepts. Recent works include explicit high-level semantic ideas of the input image/video to overcome the issues. The most likely nouns, verbs, situations, and prepositions that make up the sentence can also be used to predict the visual description. Because to the convolution kernel size constraint, 3D-CNN can only capture data over a brief period. Unfortunately, it was discovered that using this strategy led to an exponential accumulation of grammatical errors and decreased word association as video length increased. A "discriminator" module is introduced to the system architecture to address this LSTM flaw, acting as an opponent to the sentence generator. A huge interest in photographs and videos has been sparked via caption generating. In high-level vision tasks, it is challenging for the models to choose appropriate subjects in a complicated context and produce desirable captions.

PROPOSED METHODOLOGY

The proposed methodology isa unique image captioning model based on high-level image attributes considering current efforts. The senior monitoring system's most crucial component is automatic human fall detection. Several fall detection strategies have recently been proposed. To assess a person's movement using vibration and pressure- based systems, which place sensors on the floor, comes first. Here, computer vision-based technologies offer promising and practical solutions. Second, wearable technology based onaccelerometers was demonstrated. However, because they are wearable devices, they must be wornconstantly, which can be painful for elderly persons. The third group includes radar-based systems thatemploy "Doppler effects" from backscattered waves. Although reliant on radar signals, this technology frequently generates false alarms since falls are mistaken for other human actions, such as lying down or sitting up. The final one is a vision-based system, which has become extremely important in the past ten years for a variety of reasons, including the fact that it does not need to be worn, can cover large spaces, and can employ various camera sensors.

LITERATURE SURVEY

Ding, G., Chen, M., Zhao, S. et al. automatically generating captions for an image. They have used the encoder-decoder framework to generate a more descriptive sentence for the given image. They have used different weights for the words. The correlation between words and images are taken during the training phase. They maximized the consensus core between the captions generated by the captioning model and the reference information from the neighboring images of the target image, which can reduce them, is recognition problem. They mainly used computer vision and natural language processing domains. Executed on the datasets namely MSCOCO, Flickr30k.Cao,P., Yang, Z., Sun,L. etal. Image Captioning with Bidirectional Semantic Attention-Based Guiding of Long Short-Term memoryend-to-end approach is proposing a bidirectional semantic attention-based guiding of longshort-termmemory (Bag-model for image captioning. The proposed model consciously refines image features from the previously generated text. By fine-tuning the parameters ofconvolution neural networks, Bag-LSTM obtains more text - Model is dynamically lever aging more text-conditional image features, acquired by the semantic attention mechanism, as guidance information. They haveused bidirectional LSTM as the caption generator,which is capable enough of learning long term relations between visual features and semantic information by making use of both historical andfutureinformation.HaoranWang, YueZhang, Xia sheng Yu. etal. Image Caption Generation Methods",Computational Intelligence and Neuroscience, statistical probability language model to generate handcraft features and a neural network model based on anencoder-decoder language model to extract deep features. They used several attention mechanisms to improve the effect of image captioning. In thismodel, they used MSCOCO, Flickr8k, Flickr30k, PASCAL 1K, AI Challenger Dataset, and STAIR Captions datasets. BLEU and METEOR are for machine translations similarly

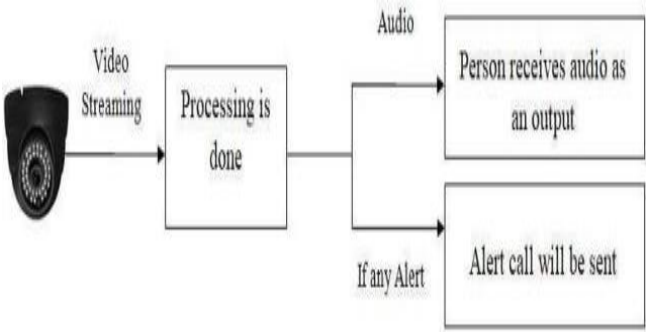
 DOI: 10.36893.JK.2023.V13I04N16.0017-0022

ROUGE, CIDEr,and SPICE are used for several other evaluation criteria.SujinLee, IncheolKim Etal. Multi modal Feature Learning for Video Captioning, visual features of the input video are extracted using C3D and ResNet, and semantic features are obtained using RNN such as LSTM. Semantic feature learning issued to identifyactions, objects, persons, and background in the input video where as Attention - based caption generation is used for effective caption generation using multimodal features. Part-Of-Speech (POS)tag function in Natural Language Toolkit (NLTK) was used to separate nouns and verbs, while plural nouns and tenses of verbs, past, continuous, andsoon, were converted back to their root forms using the lemmatize function in NLTK.They have trained their model using MSRVTT, MSVD datasets.Y.Yangetal.Video Captioning by Adversarial STM,novel approach for video captioning and adopted standard generative adversarial network(GAN) architecture, characterized by interplay of two competing processes. For generator module, they took an existing video captioning concept using LSTM network. For discriminator, they proposed a novel realization specifically tuned for the video captioning problem fall detection algorithm for the elderly based on human posture estimation.G.Sunand Z.Wang etal, Human Posture estimation algorithm. It is based on Open Pose human key point detection,combined with SSD Mobile Net object detection framework. It used SVDD classification algorithm to classify them.K.Sehairi,Chouireband J.Meunier, etal. Elderly fall detection system based on multiple shape features and motion analysis. Sil houette of a person is extracted using a background subtraction technique to estimate the head position, and a finite state machine (FSM) to compute the vertical velocity of the head. They tested on the L2ei dataset. It contains more than 2700 frames have been labeled to train 3 different classifiers.

METHODOLOGY

The procedure has two modules.They are Audio Generation Module and Alert System Module. Audio Generation module will be running and generates audio clip for every 1 hour throughout the life cycle of this software. Alert system module monitors the actions and if there is any unusual activity then it alerts the user.

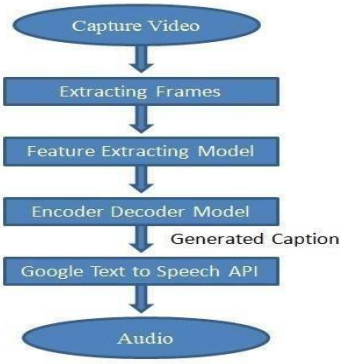
Architecture



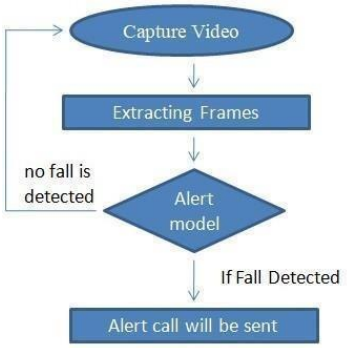
Audio generating model is developed by using CNN for extracting the features and LSTM or generating the captions. And then generated captions are converted to audio by using Google Text to Speech API. This output is sent to the respective users. The process flow of the methodology is show in the figures given below. The Capturing video using cv2 module video streaming coming from cctv is captured. The Extracting frames from the captured video VideoCapture()function of cv2 module is used. The Feature Extracting model is extracting frames are given input to trained model, which is used to extract the features in the given input frame. Feature Extraction Model is constructed by convolution neural network(Inception Resnet V2 architecture).keras API is used.

 DOI: 10.36893.JK.2023.V13I04N16.0017-0022

ISSN: 2278-4632
Vol-13, Issue-04, No.06, April : 2023



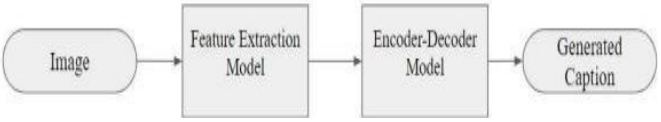
Encoder-Decoder Model generates description for the givenframe. Encoder-Decoder Model is constructed by using RNN with LSTM. keras API is used. For Audio generation, the generated caption is sent to Google Text to Speech API which will convert text to audio. An alert model is developed using CNN. The model generates and classifies with the different instances, then a particular alert call will be generated to intimate to the user. The sample of operation of the model is shown in the above figures. Capturing video is done using cv2 module video streaming. The Extracting frames from the captured video Capture () function of cv2 module isused. The Alert model classifies the frames, for the inputtrained and generates the alerts.



PROCESS FLOW

IMPLEMENTATION

The implementation of the model is as follows



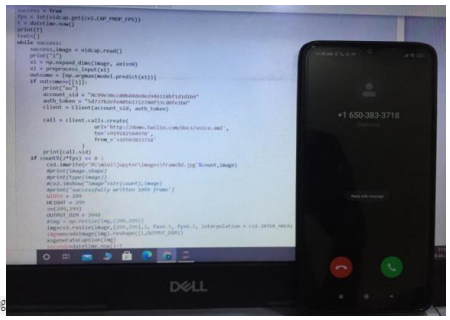
This architecture consists of three models: Feature Extraction Model, Encoder model and encoder – decoder model. The Feature Extraction Model is basically responsible for acquiring features from an image for training. It finally gives a vector as output which consists of features of the input image. This vector is sent to an encoder model as input. Features are parts or patterns of an object in an image that help to identify it. Traditional Computer Vision techniques for feature detection include Harris Corner Detection - Uses a Gaussian window function to detect corners. Shi-Tomasi Corner Detector - The authors modified the scoring function used in Harris Corner Detection to achieve a better corner detection technique. Scale-Invariant Feature Transform (SIFT) -This technique is scale invariant unlike the previous two. Speeded-Up Robust Features(SURF)-This is a faster version of SIFT as the name says. Features from Accelerated Segment Test (FAST) - This is a much faster corner detection technique compared to SURF. Binary Robust Independent Elementary Features(BRIEF)-This is only a feature descriptor that can be used with any other feature detector.This technique reduces the memory usage by converting descriptors in floating point numbers to binary strings. Oriented FAST and Rotated BRIEF(ORB)-SIFT and SURF are patented and this algorithm from OpenCV labs is a free alternative to them,that uses FAST key point detector and BRIEF descriptor.The Encoder-Decoder Model architecture for recurrent neural networks is the standard neural machine translation method that rivals and, in some cases,beats classical statistical machine translation methods. Google's translate service uses recurrent neural network architecture with an encoder-decoder at its heart. Sutskever

DOI: 10.36893.JK.2023.V13I04N16.0017-0022

model is for direct end-to-end machine translation. Cho model - that extends the architecture with GRU units and an attention mechanism. Sutskever NMT Model- It was one of the first neural machine translation systems to outperform a standard statistical machine learning model on a significant translation problem, making it a key model in thefieldof machine translation. In Encoder, Data must be encoded to be in the desired format. In the context of machine learning, transform a list of English words into a two-dimensional vector, sometimes referred to as the hidden state. Recurrent neural networks are stacked to create the encoder (RNN).In the Decoder Model, A message that has been encoded must be decode before it can be understood. The Pictionary team'ssecond member will transcribe the image into a word.The decoder will transform the two-dimensional vector into the output sequence, which is the English phrase, in the machine learning model. To predict the English term, it is also constructed with RNN layers and a thick layer. The possibility of different input and output sequence lengths is one of this model's key advantages. This makes way for some in triguinguses, such question-and-answer sessions or video captioning. In alert generation module, detects fall is an instance occurred of a person is identified and sends an alert call to the concerned user. Here, video is captured first, and then it is pre-processed. The Alert model receives this pre-processed data after that. The user would receive an alert call with a prerecorded voice message of a particular instance generated by the model.

Dataset Collection is taken from the Flickr8k_Dataset. The dataset has a pre-defined training dataset 6,000 images),development dataset(1,000images),and test dataset(1,000images).The model of the output was simulated is as follows, the Alert Generation Module generates captions are displayed in the terminal and generated audio file will be saved at the working directory.

RESULT



The generated caption is then transferred to the google text to speech API which then converts text into an audio file. This will then be sent to the end user. The alert generation module generates an alert call if any fall instance is detected in live monitoring.

CONCLUSION

An intelligent system that monitors human activity automatically sends the alerts the end users. When it was identifies the unexpected activity. There will be enhancements of the model to improve the efficiency and to generate an accurate result for the unidentified human instances.

REFERENCES

[Ding, G., Chen, M., Zhao, S. et al. “Neural Image Caption Generation with Weighted Training and Reference”. Cogn Comput11,763–777 (2019). https://doi.org/10.1007/s12559- 018-9581-x.

Cao,P.,Yang,Z.,Sun,L.etal.“Image Captioning with Bidirectional Semantic Attention- Based Guiding of Long Short-Term Memory”. Neural Process Lett 50,103–119(2019). https://doi.org/10.1007/s11063-018-09973-5.

HaoranWang, YueZhang, Xiaosheng Yu, “An Overview of Image Caption Generation Methods, Computational Intelligence and Neuro science”.vol.2020, Article ID 3062706,

 DOI: 10.36893.JK.2023.V13I04N16.0017-0022

13pages,2020.https://doi.org/10.1155/2020/3062706
SujinLee, IncheolKim, "Multi modal Feature Learning for Video Captioning", Mathematical Problems in Engineering, vol. 2018, Article ID 3125879, 8 pages,2018. https://doi.org/10.1155/2018/3125879
Jeffin Grace well,J.,Pavalarajan,S."Fall detection based on posture classification for smart home environment".J Ambient Intell Human Comput(2019).https://doi.org/10.1007/s12652-019-01600-y
Y.Yangetal.,"Video Captioning by Adversarial LSTM". in IEEE Transactions on Image Processing, vol. 27, no. 11, pp. 5600-5611,Nov.2018, doi: 10.1109/TIP.2018.2855422.
G.SunandZ.Wang,"Fall detection algorithm for the elderly based on human posture estimation,"2020Asia-Pacific Conference on Image Processing, Electronics and Computers (IPEC), Dalian, China, 2020, pp. 172-176, doi:10.1109/IPEC49694.2020.9114962.
K.Sehairi, F.Chouireb and J.Meunier,"Elderly fall detection system based on multiple shape features and motion analysis, "2018 International Conference on Intelligent Systemsand Computer Vision (ISCV), Fez, 2018, pp. 1-,doi:10.1109/ISACV.2018.8354084.
A.Torralba, R.Fergus and W.T.Freeman,"80 Million Tiny Images: A Large DataSet for Non parametric Object and Scene Recognition," in IEEE Transactions on Pattern Analysis and Machine Intelligence,vol.30,no.11,pp.1958-1970,Nov.2008,doi:10.1109/TPAMI.2008.128.
.Ordonez,V.;Kulkarni,G.;Berg,T.L.:Im2text:describing images using 1 million captioned photographs. In:Proceedings of Advances in Neural Information Processing Systems, pp. 1143–1151(2011)
. Dash, S.K.; Saha, S.; Pakray, P.; Gelbukh,A.: Generating image captions through multi modal embedding.J.Intell.FuzzySyst.36(5),4787–4796 (2019).
].Zhou,C.;Mao,Y.;Wang,X.:Topic-pecific image caption generation. In:Proceedings of Chinese Computational Linguistics and Natural Language Processing Based on Naturally Annotated Big Data, pp.321–332(2017).
Ding, S.; Qu, S.; Xi, Y.; Sangaiah, A.K.;Wan,S.:Image caption generation with high-level image features. Proc. Pattern Recognit. Lett. 123,89–95(2019).
.Gan, Z.;Gan,C.;He,X.;Pu,Y.;Tran,K.; Gao,J.;Carin,L.; Deng,L. : Semantic compositional networks for visual captioning. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp.1141–1150(2017).
Simonyan, K.;Zisserman, A Very deep convolutional networks for large-scale image recognition. ArXiv pre printar Xiv:1409.1556(2014).
.Karpathy,A.;Fei-Fei,L.:Deep visual-semantic alignments for generating image descriptions. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 3128– 31
G.Sai Chaitanya Kumar, Dr.Reddi Kiran Kumar, Dr.G.Apparao Naidu, "Noise Removal in Microarray Images using Variational Mode Decomposition Technique " Telecommunication computing Electronics and Control ISSN 1693-6930 Volume 15, Number 4 (2017), pp. 1750-1756

 DOI: 10.36893.JK.2023.V13I04N16.0017-0022