International Journal of Research and Development in Applied Science and Engineering (IJRDASE)
ISSN: 2454-6844

Implementing Object Detection in Video Frames through Kalman Filtering

Divya Paswan, Krishna Kumar, Garima Singh, Aman Singh, Abhishek Yadav
Institute of Technology and Management, Gida, Gorakhpur,
divyapasi22@gmail.com, kk_gkp@rediffrail.com, garimasingh15082004@gmail.com, rajputsahilsingh9009@gmail.com, ay400573@gmail.com

Abstract: Traditional object detection is done using basic features and handcrafted architecture which eventually doesn't give effective results. So, to overcome this, the use of advanced technology based on machine learning comes into play as it has a wide hand in this field. Thereby this work brings an effective object detection model from video frames in which initially data collection from video is performed for background subtraction for extraction of moving object is performrd using a smart estimation and prediction approach that performs perfectly using integration of object features. The study shows that the proposed method performs well with high accuracy when compared with other state-of-art models.

Keywords: Object Detection, Kalman filter, Data Mining, AI.

1. Introduction:

Many applications are utilizing video, for detecting pedestrians, recognizing unusual behavior in parking lots and so on. Nowadays, retrieving moving objects and automating the analysis of video is more frequently used. An example of multimedia data is video, which combines a variety of types of data like text, picture, metadata, visual, and audio. The primary aim of video data analysis is to recognize and track moving objects such as people across frames. Security, surveillance, entertainment, medical and legal applications, as well as medical education and sports make use of Video Data Mining. Video data mining is based on finding and analyzing patterns in massive amounts of video data. The video is made up of a series of images. The video material may be divided into two categories: i) low-level feature data, such as colour, texture, and form and ii) high-level feature details, such as audio and video. Syntactic information like video material contains conspicuous objects, their spatial-temporal [2] location and their spatialtemporal connection. Semantic data, explain what is happening in the video, such as spatial features offered by a video frame, position characters moved in the screen etc. The aspects of time characterize a succession of video frames such as the actions of actors and the movement of objects in sequence. The first step in extracting information about the objects in a video is to detect moving objects in video streams. In many computer vision applications, including video surveillance and people annotation, this is the first step. Figure 1(a) shows the overall use of emerging technology over object detection. Figure 1(b) explains most of the research works that happened over the object detection areas.

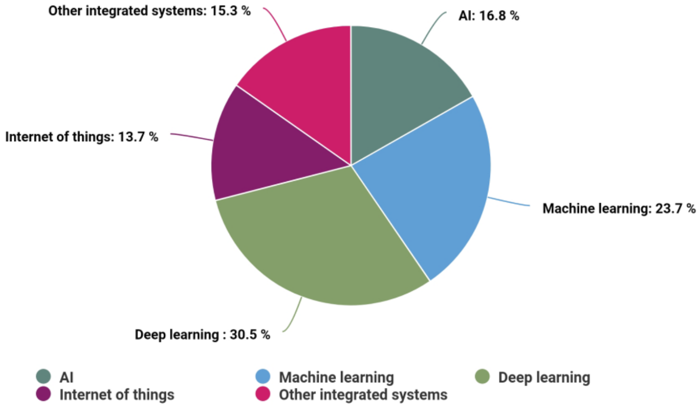

Figure 1(a): Most used emerging technology for the Object detection

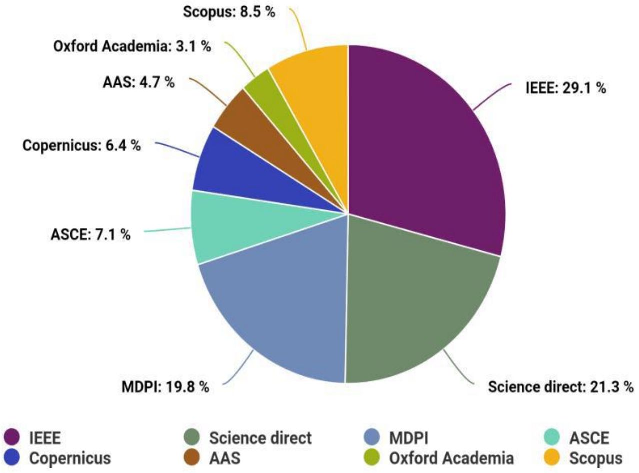

Figure 1(b): Papers published regarding object detection

2. Related Work:

The journal (Schmidhuber, 2015) explains the working of deep learning model and also distinguishes the difference between the deep and shallow learners. Deep learning algorithm allows computational models that are comprised with many layers for processing the requirement with multiple abstraction layers for filtering the data that are needed. The deep learning model has many state of the art applications(Lecun, Bengio and Hinton, 2015) object detection, voice processing, speech recognition and various other scrutinizing application. Deep learning recognizes the convolute or complex structure in the large datasets and shows the model to learn their own internal parameters which are observed from the previous layer which can be used in the current layer, this can be done using the backpropagation method. Convolution nets have brought advance development in processing audio, video, speech and image when compared to recurrent nets.

Convolution neural networks are iterative neural network where convolution layers turn out in turn with subsampling layers(Cires et al., 2003)(Schmidhuber, 2015), which is a resemblance of simple and complex cells present in the neuron. This depends on way the neural nets getting trained as well as identification of convolution and subsampling layers. Image processing layer is an optional layer where all the pre-processing of the data is done and convolutional layer is parametrized by way of the dimensions and the wide variety of the maps, kernel sizes, skipping elements and connection table.

Convolution networks being a subset of deep learning model acts as a best and strong model for understanding the extraction of features and visual models. (Krizhevsky, Sutskever and Hinton) used convolution neural networks prosperously for grasp detection as a classifier in the detection pipeline using sliding window concept. The problem addressed here is coincidentally merges with the problem arises here, but the only difference is processing pipeline and different usage of network architecture, this increases the accuracy at greater speeds comparatively.

Contemporary work on grasp detection mainly spotlights the issues in detecting grasps solely from RGB-D data(Saxena, Driemeyer and Ng, 2008). These algorithms depend on the machine learning techniques to detect a good grasp from the data. Grasp visual models are the state of the work objects and it is well known for single object view, now not a complete bodily model.

The core problem to be addressed in computer vision is object detection(Nowozin and Lampert, 2010). To start with detection of pipelines usually start with decoction of available robust features from input images (SIFT, HOG, Convolution features). Then, in the available feature space classifiers or localizers are passed to detect objects. After that the concept of sliding window is implemented either throughout the image or at a part of the image with the help of classifier or localizers.

3. Methodology:

Artificial intelligence (AI) has revolutionized countless industries and aspects of daily life, permeating everything from virtual assistants to advanced robotics. At its core, AI involves the development of computer systems capable of performing tasks that typically require human intelligence, such as learning, problem-solving, perception, and language understanding. Through machine learning algorithms, AI systems can analyze vast amounts of data to identify patterns, make predictions, and continuously improve their performance over time. Natural language processing enables AI to comprehend and generate human language, facilitating communication between machines and humans. From personalized recommendations on streaming platforms to autonomous vehicles navigating complex environments, AI is reshaping how we work, live, and interact with technology, with its potential for further innovation and societal impact only continuing to grow. However, ethical considerations and concerns about the societal implications of AI deployment remain significant, prompting ongoing discussions about responsible development, transparency, and accountability in the use of artificial intelligence.

Proposed Work:

A suggested architecture for object detection is shown in Figure 2, in which a video is taken as an input and a single frame is selected for object detection. A video frame is first broken into image frames, and then a single frame is used to detect the object of interest [10]. During the first stage of processing, the images are analyzed for their characteristics. The structure and form of the pixels provide enough information to identify the significant elements in an image, so not every pixel is sent to the neural network. The edges of the picture may be recognised using the feature extraction procedure. Then it will be passed for dimensionality reduction where with the help of SE block those features will be selected and finally given to classification for required results.

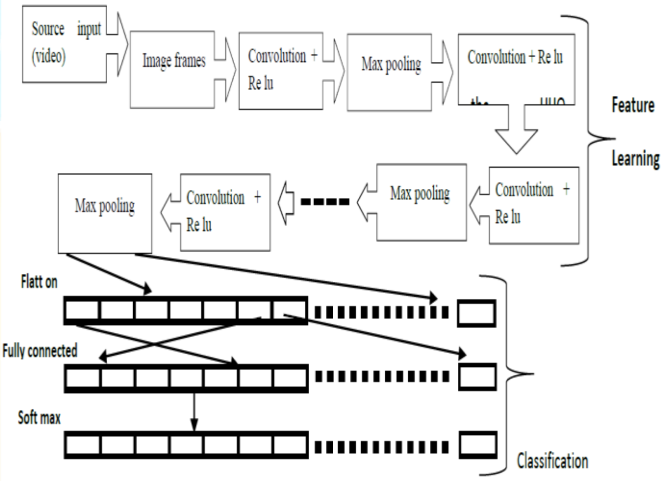

Fig 2: Flow of objective detection

4. Result and Discussion:

This work shows the use the computer vision in object detection by configuring Kalman Filter approach to track objects by image differencing. These results are discussed with its main steps of algorithm at the specified video of moving objects and describes performance in the form of detected and actual track of records. The object detection algorithm has many uses, including applications in control, navigation, computer vision, and time series econometrics. This section illustrates how to use the Kalman filter for tracking objects and focuses on three important features:

(a) Prediction of object's future location

(b) Reduction of noise introduced by inaccurate detections

(c) Facilitating the process of association of multiple objects to their tracks

The main challenges of object tracking are due to movement of background or camera. Without the use of Kalman filter, let us first demonstrate the challenges of tracking an object in a video. The following video shows a green ball moving from left to right on the floor.

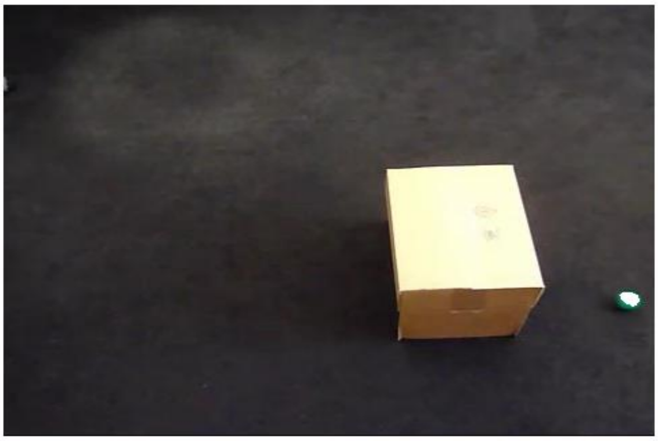

Fig 3: Detection of moving ball by simple image difference method.

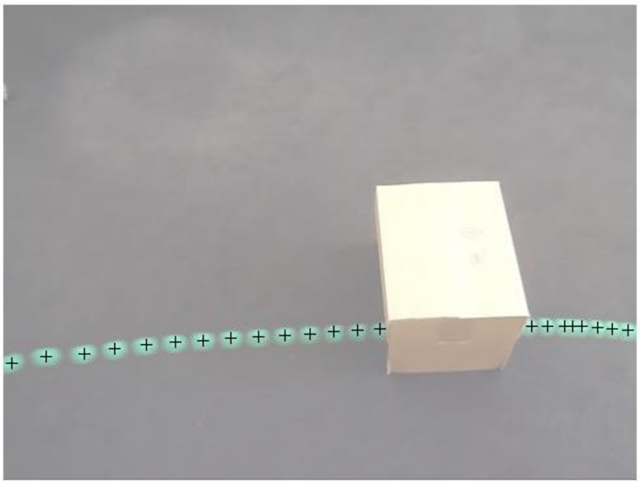

Fig 4: Overlaid view of image frames of video to show the detected object

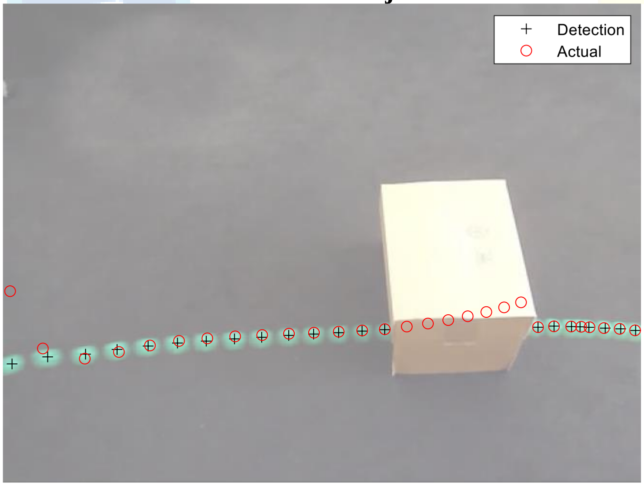

Fig 5: Object actual and detected location without kalman filter

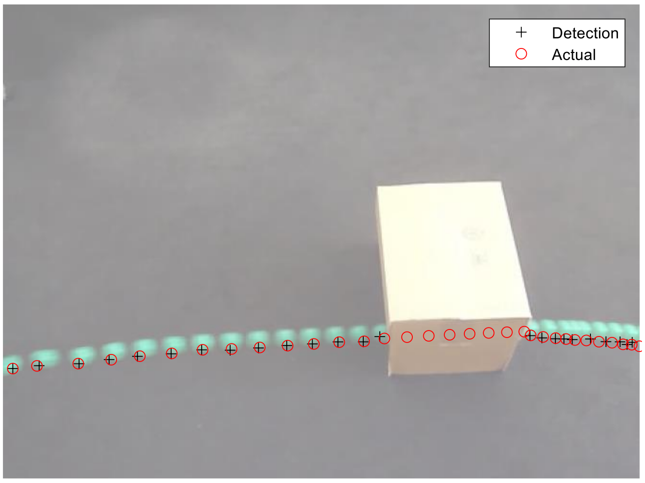

Fig 6: Object actual and detected location with kalman filter

5. Conclusion:

In this work, a visual surveillance system with moving object detection and tracking capability has been presented. Object tracking of any single and multiple moving object has been successfully implemented on standard datasets using Kalman filter with image differencing. The system works on videos of indoor as well as outdoor environment taken using static camera under moderate to complex background condition. This implemented module can be applied to any computer vision application for moving object detection and tracking.

References:

[1] Yiwei Wang, John F. Doherty and Robert E. Van Dyck, "Moving Object Tracking in Video", in proceedings of 29th applied imagery pattern recognition workshop, ISBN 0-7695-0978-9, page 95,2000.

[2] Bhavana C. Bendale, Prof. Anil R. Karwankar, "Moving Object Tracking in Video Using MATLAB", International Journal of Electronics, Communication and Soft Computing Science and Engineering ISSN: 2277-9477, Volume 2, Issue 1.

[3] Marcus A. Brubaker, Leonid Sigal and David J. Fleet, "Video-Based People Tracking", hand book of ambient intelligence under smart environments 2010, pp 57-87.

[4] Emilio Maggio and Andrea Cavallaro, "Video Tracking: Theory and Practice", _rst edition 2011, John Wiley and Sons, Ltd.

[5] Y.Alper, J.Omar, and S.Mubarak. "Object Tracking: A Survey" ACM Computing Surveys, vol. 38, no. 4, Article 13, December 2006.

[6] B. Triggs, P.F. McLauchlan, R.I. Hartley and A.W. "Fitzgibbon. Bundle adjustment – a modern synthesis". In Proceedings of the International Conference on Computer Vision, London, UK, 1999, 298?372.

[7] G.C. Holst and T.S. Lomheim. "CMOS/CCD Sensors and Camera Systems". Bellingham, WA, SPIE Society of Photo-Optical Instrumentation Engineering, 2007.

[8] E. Maggio, M. Taj and A. Cavallaro. "E_cient multi-target visual tracking using random finite sets". IEEE Transactions on Circuits Systems and Video Technology, 18(8), 1016?1027,
2008.

[9] G. David Lowe. Object recognition from local scale-invariant features. Proceedings of the
International Conference on Computer Vision. 2. pp. 1150?1157,1997.

[10] G. David Lowe. Distinctive image features from scale-invariant keypoints. International Journal of Computer Vision, 60(2), pp, 91-110, 2004.

[11] Bruce D. Lucas and Takeo Kanade. An Iterative Image Registration Technique with an Application to Stereo Vision. International Joint Conference on Arti_cial Intelligence, pages 674-679, 1981.

[12] Carlo Tomasi and Takeo Kanade. Detection and Tracking of Point Features. Carnegie Mellon University Technical Report CMU-CS-91-132, April 1991.

[13] Jianbo Shi and Carlo Tomasi. Good Features to Track. IEEE Conference on Computer Vision and Pattern Recognition, pages 593-600, 1994.

[14] Stan Birch_eld. Derivation of Kanade-Lucas-Tomasi Tracking Equation. Unpublished, January 1997.

[15] Y. Cui, S. Samarasekera, Q. Huang. Indoor Monitoring Via the Collaboration Betweena Peripheral Senson and a Foveal Sensor, IEEE Work-shop on Visual Surveillance, Bomba y, India, 2-9, 1998.

[16] G. R. Bradski, Computer Vision Face T racking as a Component of a Perceptual User Interface, IEEE Work. on Applic. Comp. Vis., Princeton, 214-219, 1998.

[17] S.S. Intille, J.W. Davis, A.F. Bobick, Real-Time Closed-World T racking. IEEE Conf. on Comp. Vis. and Pat. Rec., Puerto Rico, 697-703, 1997.

[18] C. Wren, A. Azarbayejani, T. Darrell, A. Pentland, P_nder: Real-Time Tracking of the Human Body, IEEE Trans. Pattern Analysis Machine Intell, 19:780-785, 1997.

[19] A. Eleftheriadis, A. Jacquin. Automatic Face Location Detection and Tracking for Model-Assisted Coding of Video Teleconference Sequences at Low Bit Rates, Signal Processing- Image Communication, 7(3): 231-248, 1995.

[20] D. Fuiorea, V. Gui, D. Pescaru, and C. Toma. Comparative study on RANSAC and Mean shift algorithm, International Symposium on Electronics and Telecommunications Edition 8.
vol. 53(67) Sept. 2008, pp. 80-85.