Journal of Interdisciplinary Mathematics
ISSN: 0972-0502 (Print), ISSN: 2169-012X (Online)
Vol. 26 (2023), No. 3, pp. 341–350
DOI : 10.47974/JIM-1665

A novel approach of surveillance video indexing and retrieval using object detection and tracking

Shailesh D. Kamble *
Department of Artificial Intelligence and Data Science
Indira Gandhi Delhi Technical University for Women
Delhi
India

Dilip Kumar Jang Bahadur Saini †
Department of Computer Science and Engineering
Himalayan School of Science & Technology
Swami Rama Himalayan University
Uttarakhand
India

Sachin Jain §
Department of Computer Science and Engineering
Oklahoma State University Stillwater
U.S.A.

Kapil Kumar ‡
Department of Computer Application
Meerut Institute of Engineering & Technology
Meerut
Uttar Pradesh
India

* *E-mail:* **shailesh_2kin@rediffmail.com (Corresponding Author)**
† *E-mail:* **dilipsaini@gmail.com**
§ *E-mail:* **sachin.jain@okstate.edu**
‡ *E-mail:* **kapilkumarvermaji@gmail.com**

Sunil Kumar [@]
Department of Computer Science and Engineering
Meerut Institute of Engineering & Technology
Meerut
Uttar Pradesh
India

Dharmesh Dhabliya [#]
Department of Information Technology
Vishwakarma Institute of Information Technology
Pune
Maharashtra
India

Abstract

The problem of searching videos in large databases i.e. multimedia applications is a major challenge. Therefore, video indexing is used to search the location of the particular video in a large database quickly. Quickly locating the video in the large database is the good quality of indexing. Still, there is a scope of improvement in quickly searching a video in a large database in terms of assigning labels to video. In computer vision, real-time object detection and tracking is a gigantic, vibrant yet indecisive and intricate area. You only look once (YOLO) algorithm is used to detect the object and background subtraction is used to track the object. In this paper, video indexing using object detection / tracking can be performed on single object in a video. In future, video indexing can be performed on multiple objects in a video.

Subject Classification: *97R20, 97R60.*

Keywords: *YOLO, Real-time object detection, Multi-objects tracking, Video indexing.*

1. Introduction

The utilization of multimedia information is suitable in wide range of applications like telemedicine, production and distribution systems, distance learning, digital libraries, tourism, disseminated sources like CAD/ CAM, GIS and the document summarization in World Wide Web (WWW) [1]. Huge amounts of CCTV content present many challenges in the effective and efficient management, indexing and access to useful information [2]. Videos contains rich basis of information. It offers visual

[@] *E-mail:* **sunilymca2k5@gmail.com**
[#] *E-mail:* **dharmesh.dhabliya@viit.ac.in**

material about the scenes. Nevertheless, this information is obliquely hidden in the raw video and is offered at the price of a very high level of temporal redundancy. The advent of video as a source of data and information on the computer opens up possibilities for accessing, viewing and managing video content [3]. It won't support the way you interact with the video information defined. At present, the mode to access information of interest is to scan the videos one by one. The only way to edit, comment, or edit video is to process the video frame by frame. This process is slow and long [4].

A series of frames/images is called as video, which are shown to the viewer according to the concept of "Perception of Vision". As technology advances, video becomes an important source of data, such as text and graphics. For example, CCTV cameras are used as evidence in many criminal cases [5]. However, the duration of the video may be different and the viewer may lose sight of everything if the video is confusing or too long. Video indexing plays an important role for long videos [6]. Object-based video indexing is closely related to object detection / tracking. Detecting an object in video consists of confirming the existence of the object in a series of images and determining its detection position. [7-8]. Indexing of multimedia information is essential to define, organize and store multimedia data and to assist users for finding multimedia resources fast and conveniently. The important form of multimedia information is a Dynamic video [9].

Visual e-commerce scrutiny (analysis of trends of interest in user selection and ordering), correlation research in digital institutions, instant viewing of video directories, analysis of news, useful video search (intelligent management of videos) and video inspection which have aroused researchers' interests [10].

The method presents the utmost difficulties in designing a system that bridges the research gap among low and high-level semantic functions. [11]. The foremost contribution of this research work is to detect and track objects in real time and to index an object of video. To create an environment where all essential libraries and data to be trained on 80 classes of objects will be stored and to detect objects in real time using the web camera present in the laptop [12].

2. Literature review

Ray et. al [13] provided a new approach for tracking and detection of objects which are present in videos that was captured by movable camera without additional sensors. Here, the moving objects which were generated by a spatiotemporal analysis of the video by utilizing a 3D Gabor filter and then combined by the Minimum Spanning Tree were recognized as clusters of spatiotemporal blobs. At the time of tracking, Linear Assignment Problem solves the data connection problem and then the application of Kalman filter handled the occlusion. There was no need of training or initialization to perform on the sample data was one of the main advantages of the projected system. Tong et. al [14] suggested a low-computational process to detect movable objects in the AVC / H.264 encoder, that encompasses mostly in three steps to achieve multi-granularity such as boundary refinement, spatiotemporal, and background modeling. The data which was generated through encoding of video, e.g. RD Cost was used to detect moving objects to avoid the decoding further. Unlike the existing work, the detection of moving objects was built into the video encoder.

Wang et. al [15] described a scale-aware coupling-structure network (SCNet) of end-to-end, directing on main problems on how to detect video objects with affordable computations and high accuracy. In SCNet, the module lightweight scale-aware was known to be a flexible for the model of variation object scale. In addition, the feature aggregation function has been simplified to improve efficiency. With that of ImageNet VID dataset, experiments were carried out. SCNet achieves better detection efficiency, accurately 79.5% of the mAP, with a point of 5.9 of improvement over the baseline of single-frame. Qiong et. al [16] presented an efficient way to detect salient objects in video.

Wang et. al [17] suggested a novel approach to segmentation of unsupervised video object using aware DOA. This models the spatio-temporal video sequences consistency by detecting the dependencies of background of the neighboring images. To support the hard negatives of high-quality, the algorithm of block matching was then applied to the previous frameworks for the tracking of hard negatives which were associated. Negatives which are general were also presented if the sequence does not contain hard negatives. The experimental outcomes showed that the two types of negatives were corresponding. Lastly, conducted DOA with hard negative, negative, and positive masks to update segmentation of both the background and foreground.

3. Proposed Methodology

The first step in creating an object detection model is to create an environment with all the appropriate and necessary libraries. First, we need to train the image data using a coco data set. It's a huge 26GB dataset, but we only need 300MB for fast processing. The last step is to create a yolo file with trained weights for 80 classes of objects like a person, car, truck, horse, etc. We need a pre-trained model for the process, and this model is then incorporated into the current working dictionary. Next, we need to define the Keras model. For detection purposes, need to know that the object of the image must already be trained in the model, and only then it will correctly detect the object. There is another process named YOLO Algorithm that can be used to detect objects. Figure 1 shows the proposed system/model.

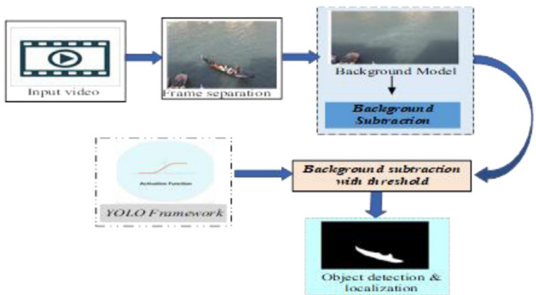


Figure 1

Proposed System/model

3.1 *Yolo Algorithm*

YOLO is one of the fastest algorithms for detecting objects. It is more accurate algorithm, used in real-time object detection without losing precision. YOLO utilizes a single convolutional network to simultaneously predict multiple constraint frames and category probabilities for these boxes [25-26]. It frames the problem of object detection as a regression problem. YOLO trains in full images. The flowchart of YOLO framework is displayed as follows.

The above given table refers to the Summary of YOLO Algorithm Versions. The YOLO Algorithm has 4 versions, Version 1 has totally of 26 layers in which consists of 2 Fully Connected layers and of 24 Convolution layers, and thus has some drawbacks like as incapability and so Version 2 has been introduced. Version 2 contains totally of 30 layers in which this

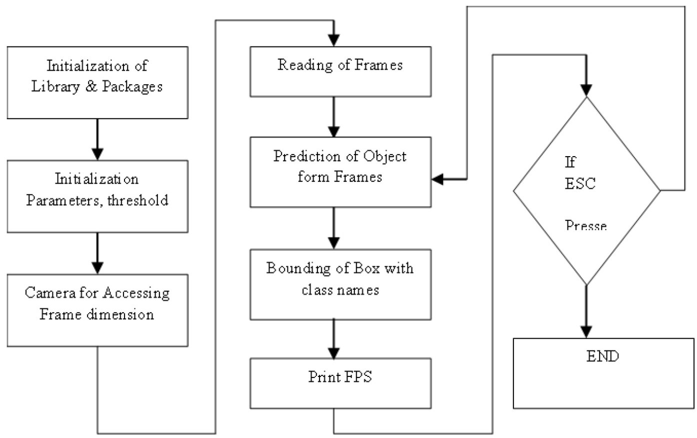


Figure 2

Flowchart describing YOLO Algorithm

Table 1

Summary of YOLO Versions Algorithm

Sr. No	Versions	Salient Features	Drawbacks
1	YOLO-Version-1 (Year:2016)	The first version consists of 26 layers out of which 24 Convolution Layers are present that are followed by 2 Fully Connected layers.	The major issue concerned with Version-1 is its incapability to detect very small objects.
2	YOLO-Version-2 (Year:2017)	The second version consists of 30 layers. Anchor Box is introduced. fully connected layer is absent in this version.	Problem with Version-2 is still is inability to detect very small objects.
3	YOLO-Version-3 (Year:2018)	The third version consists of 106 layers. 9 anchor boxes are introduced in this version.	Quite Good for detecting small objects but still with some limitations. Multi-Class difficulty results in Multi-Label issues are present in this version.
4.	YOLO-Version-4	Fast and most accurate as compared other versions.	It is more complex.

version does not consists of fully connected layer, and still this version was unable to detect very small objects. Version 3 has totally of 106 layers and anchor boxes of 9 are introduced. Version 4 is the most accurate and very fast version as contrasted with that of other 3 versions. Figure 2 shows flowchart of YOLO and Table 1 shows the summary of YOLO algorithm.

To detect the objects moving in videos from cameras, a widely utilized approach is Background subtraction. The reason for this method is to detect the objects moving based on the difference among the key frame and the current frame, frequently mentioned to as a background image or background pattern. The basic background subtraction has various approaches in terms of post processing, background maintenance, and foreground region detection. A pixel at location (x, y) in the current image I_t is foreground image if $|I_t(x,y)- C_t(x,y)|>T$ is satisfied; where T is a threshold. The background image updated by the use of an Infinite Impulse Response (IIR) filter given by equation 1,2 and 3 are as follows,

$$C_{t+1} = \alpha I_t + (1-\alpha)C_i \tag{1}$$

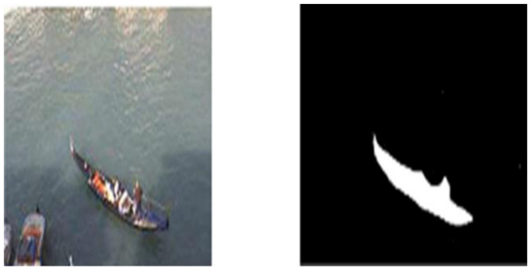
At t + 1time, the frame difference is determined as

$$D(t+1) = |W(a,b,t+1) - W(a,b,t)| \tag{2}$$

The background is the frame supposed at time t. This difference of image only shows a certain intensity for the pixel locations that have changed in two images. The background appears to have been removed because the difference among the intensity levels of the background pixels is zero in an ideal situation. Though, only this method works if all the pixels of the foreground are in motion while all of the background pixels are stationary [29]. A threshold value of "T" could be set for this image difference to advance the background subtraction method.

$$|W(a,b,t+1) - W(a,b,t)| > T \tag{3}$$

The pixel intensity of the image variance is set to a threshold value or filtered based on the T value. This method's accuracy depends on the object speed in the scene. The faster movements of the object in the image may require thresholds of higher values. With the image mentioned in Figure 3 with the white color represents the subtracted image while the black color is the background of the image. And so finally, the output image obtained is the binary image.



(a) Original Image (b) Background Subtraction

Figure 3

Outcome of Background Subtraction

4. Experimental Results and Discussions

The moving object detection in the surveillance area, background subtraction algorithm is utilized in this experiment. The demo has been set up for the proposed technique implementation in the Simulink and coding with MATLAB. At first, initialized reference image in the code is done. We used CCTV footage and the window after processing shows total people count; it also can be programmed to set boundaries to check movement across it. Also speed and people movements are also visible with very good accuracy. This model also works in real time but requires a lot of processing power. Bounding boxes are also setting quite perfectly and stays with movement of objects. The count is also being calculated which the number of persons present in the camera range and then there crossing the line. The implementation is executed in MATLAB and obtained the following results. The following figure shows reference image, and subtracted image.

(a) Object detection (b) Object tracking

Figure 4

Indexing Objects

5. Conclusion

Object detection is a basic computerized process of recognizing several objects in a series of images. This document provides an outline of object detection techniques and allows users to index their videos so that they can navigate through the videos more easily. This article utilizes a You Only Look Once algorithm to perform the object detection task and background subtraction to remove the background image and get an object detected accurately. There are different versions of the YOLO object detection algorithms. YOLO version-4 is the most accurate and fastest in terms of real-time object detection speed. By integrating this work with other techniques such as by tracking objects in a live video, we can understand the time of each object in our video. In the future, the YOLO algorithm will play an important role in multi-object detection to detect multiple objects in application of image processing. Video indexing can be tested on multiple objects and applied using motion vector estimation to track objects.

References

[1] Dasiopoulou, Stamatia, Vasileios Mezaris, Ioannis Kompatsiaris, V-K. Papastathis, and Michael G. Strintzis. "Knowledge-assisted semantic video object detection." IEEE Transactions on Circuits and Systems for Video Technology 15, no. 10 : 1210-1224 (2005).

[2] Chamasemani, Fereshteh Falah, Lilly Suriani Affendey, Fatimah Khalid, and Norwati Mustapha. "Object detection and representation method for surveillance video indexing." In 2015 IEEE 3rd International Conference on Smart Instrumentation, Measurement and Applications (ICSIMA), pp. 1-5. IEEE (2015).

[3] Irani, Michal, and Prabu Anandan. "Video indexing based on mosaic representations." Proceedings of the IEEE 86, no. 5 : 905-921 (1998).

[4] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR), vol. 1, pp. 511-518 (Dec. 2001).

[5] M. Weber, M. Welling, and P. Perona, "Towards automatic discovery of object categories," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., vol. 2, pp. 101-108 (Jun. 2000,).

[6] Hatwar Rupali B., Shailesh D. Kamble, Nileshsingh V. Thakur, and Sandeep Kakde. "A review on moving object detection and tracking methods in video." *Int J Pure Appl Mathem* 118, no. 16 : 511-526 (2018).

[7] Shende Priyanka, Mahesh Pawar, and Sandeep Kakde. "A brief review on: mri images reconstruction using GAN." In 2019 International Conference on Communication and Signal Processing (ICCSP), pp. 0139-0142. IEEE (2019).

[8] Nirmalkar Nitish, Shailesh Kamble, and Sandeep Kakde. "A review of image forgery techniques and their detection." In 2015 International Conference on Innovations in Information, Embedded and Communication Systems (ICIIECS), pp. 1-5. IEEE (2015).

[9] Naaz Humerah, Shubhangi Rathkanthiwar, and Sandeep Kakde. "Implementation of hybrid algorithm for image compression and decompression." *International Journal of Engineering Research* 5, no. 5 : 398-403 (2016).

[10] Hu, Weiming, Nianhua Xie, Li Li, Xianglin Zeng, and Stephen Maybank. "A survey on visual content-based video indexing and retrieval." IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews) 41, no. 6 : 797-819 (2011).

[11] Asghar, Muhammad Nabeel, Fiaz Hussain, and Rob Manton. "Video indexing: a survey." *International Journal of Computer and Information Technology* 3, no. 01 (2014).

[12] Naphade, Milind R., T. Kristjansson, B. Frey, and Thomas S. Huang. "Probabilistic multimedia objects (multijects): A novel approach to video indexing and retrieval in multimedia systems." In Proceedings 1998 International Conference on Image Processing. ICIP98 (Cat. No. 98CB36269), pp. 536-540. IEEE (1998).

[13] Ray, Kumar S., and Soma Chakraborty. "Object detection by spatio-temporal analysis and tracking of the detected objects in a video with variable background." *Journal of Visual Communication and Image Representation* 58 : 662-674 (2019).

[14] Tong, Lingling, Feng Dai, Dongming Zhang, Dong Wang, and Yongdong Zhang. "Encoder combined video moving object detection." Neurocomputing 139 : 150-162 (2014).

[15] Wang, Fengchao, Zhewei Xu, Yanfen Gan, Chi-Man Vong, and Qiong Liu. "SCNet: Scale-aware coupling-structure network for efficient video object detection." Neurocomputing 404 : 283-293 (2020).

[16] Wang, Qiong, Lu Zhang, Wenbin Zou, and Kidiyo Kpalma. "Salient video object detection using a virtual border and guided filter." Pattern Recognition 97 : 106998 (2020).

[17] Wang, Ye, Jongmoo Choi, Yueru Chen, Siyang Li, Qin Huang, Kaitai Zhang, Ming-Sui Lee, and C-C. Jay Kuo. "Unsupervised video object segmentation with distractor-aware online adaptation." *Journal of Visual Communication and Image Representation* 74 : 102953 (2021).