Artificial Intelligence-Driven Video Indexing for Rapid Surveillance Footage Summarization and Review

Jaemin Jung[1], **Soonyong Park**[1], **Harim Kim**[1], **Changha Lee**[2] and **Charmgil Hong**[1]
[1]School of Computer Science and Electrical Engineering, Handong Global University
[2]GMDSOFT
{21900675, parksy3978, hrkim, charmgil}@handong.ac.kr
yielding@gmdsoft.co.kr

Abstract

This paper introduces VIDEX, an advanced tool designed to streamline the analysis of surveillance video through a user-friendly interface. VIDEX achieves high development efficiency and maintainability utilizing the Model-View-ViewModel (MVVM) design pattern. The core feature of VIDEX is a footage summary using object detection and anomaly detection. Its architecture ensures efficient data management by organizing detected objects and anomalies within an indexed database, thus facilitating a more rapid review process. Additionally, multi-threading was used to shorten the processing time. VIDEX provides a video summarization that can be used primarily in the criminal investigation stage using the information stored in a database. Discover more about VIDEX and access its resources at https://github.com/nth221/videx.

1 Introduction

In recent decades, the proliferation of CCTV (closed-circuit television) cameras for surveillance and security has been on a continuous rise. This increase plays a crucial role in preventing accidents and aiding in crime investigations across various settings, including public areas, proprietary spaces, and residential buildings. However, the sheer volume of footage generated poses a significant challenge in terms of effective analysis and management. Traditional methods of reviewing lengthy CCTV recordings demand substantial human labor and are often inefficient. This has led to a growing need for enhanced analytical tools that leverage artificial intelligence (AI) technology and reduce manual efforts. Recognizing this, we have developed an easy-to-use surveillance video analysis tool that harnesses the power of advanced object detection [Redmon *et al.*, 2016] and anomaly detection [Aggarwal, 2017] techniques. The characteristics of objects and abnormal behaviors related to unexpected or criminal situations are categorized and used for criminal investigations. Our approach, utilizing such AI-driven automated detection methods, seeks to transform surveillance video analysis and make it more efficient and accessible to users.

VIDEX facilitates automatic object detection and anomaly detection, to enable rapid access to critical events within the footage. Employing object detection allows us to identify and classify various entities within the footage that could be important in video analysis; Whereas, adopting anomaly detection lets us detect irregular patterns or activities that deviate from the norm and could spot potential security threats. By cataloging information about detected objects and anomalies in an indexed database, VIDEX ensures efficient retrieval and analysis of the video frames that require attention for review. This significantly cuts down the time and resources needed for a thorough examination of the surveillance footage.

The development of VIDEX represents a convergence of software development and relevant AI technologies. By leveraging a YOLO (You Only Look Once) [Redmon *et al.*, 2016] model for object detection and non-parametric outlier detection methods with 3D-CNN (3D-Convolutional Neural Network) [Carreira and Zisserman, 2017] for anomaly detection, VIDEX simplifies the bookkeeping and management of detected objects and minimizes the need for repeated manual review and analysis. Additionally, the detection methods are integrated with an intuitive Windows-based interface, developed with C# and the WPF (Windows Presentation Foundation). This flexibility, combined with the ability to distinguish anomaly scenes in the video, significantly increases the efficiency and effectiveness of surveillance operations. Consequently, VIDEX not only enhances access to surveillance footage but also significantly improves the analytical workflow and insights derived from surveillance video analysis.

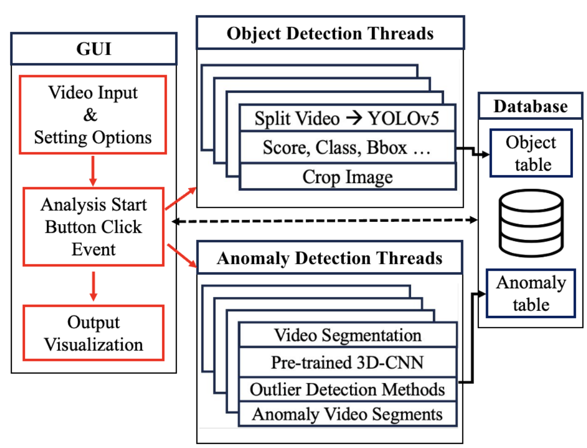

Figure 1: VIDEX system design overview

This demo showcases the near future of surveillance footage analysis and highlights the importance and usefulness of advancing security monitoring capabilities. The paper is structured as follows: Section 2 presents the System Design of VIDEX; Section 3 explores the specific components and use cases of the object detection and anomaly detection models; Section 4 concludes the paper and discusses future work.

2 System Design

VIDEX's architecture, illustrated in Figure 1, comprises three main components: a user-facing GUI, object and anomaly detection modules that run in parallel, and an indexed database storage. This section provides a detailed overview of each component one by one.

2.1 Graphical User Interface

Employing the Model-View-ViewModel (MVVM) pattern, which is popular in modern software development, VIDEX not only separates the UI (user interface) from the non-UI code to improve maintainability, but also yields a code structure that is efficiently organized with diverse functionalities. The Model in VIDEX is responsible for data import and storage; it defines and instantiates data classes for effective database management. The View manages the user interface and processes user inputs. The VIDEX interface includes `SettingView` for importing videos and configuring settings, along with `ObjectDetectionView` and `AnomalyDetectionView`, which display the results of object detection and anomaly detection, respectively. The ViewModel responds to events in the View and executes operations designed for those events.

Figure 2 displays the screenshots of the VIDEX interface. Figure 2 (a) presents the `SettingView`, which includes a Video Player, a Video Option Panel, and an Object Filter Panel. The Video Player is where users can upload and play videos. The Video Option Panel lets users select and analyze specific sections of the video. The Object Filter Panel allows users to choose which objects to detect. Figure 2 (b) shows the `ObjecDetectionView` that contains the output video with detection bounding boxes, a Detected Object Gallery, and an Object Graph visualizing object appearances. 2 (c) shows the `AnomalyDetectionView`, with an Outlier Graph displays the scores of outliers detected in the video.

2.2 Multi-threaded Detection Flows

Considering that video data typically comprise extensive sets of continuously changing image frames, processing these data can be highly time-consuming. VIDEX addresses this challenge by employing a multi-threading strategy. This approach involves dividing a single process into several execution units, referred to as threads, which promote resource sharing and boost system utilization.

VIDEX's threading pipeline solves the race condition and deadlock problems that may arise when multi-threading is applied to object detection, using the following approach. First, the entire video is segmented according to the number of threads allocated to object detection. Each thread is assigned a portion of these segments, and independent operation is possible since no frames are shared among them. Every thread processes its assigned video frames using the YOLOv5 [Jocher, 2020] model, via ONNX (Open Neural Network eXchange) [Bai *et al.*, 2019], and captures relevant information such as object class, frame number, bounding box coordinates, and size, which are then stored in the database. Concurrently, this information is retrieved from the database to update the user interface.

For anomaly detection, VIDEX assigns a separate thread to each detection task to effectively parallelize the process. This begins by dividing the input video into smaller segments that serve as the basic units for anomaly detection. Spatio-temporal feature embeddings are then extracted using a pre-trained 3D-CNN model (also accessed via ONNX), trained on large-scale action recognition datasets. Each thread then analyzes these embeddings to pinpoint anomalies, utilizing specified non-parametric outlier detection techniques. This strategic use of multi-threading enables VIDEX to achieve parallelism across processes and enhance processing speed.

2.3 Index Database

VIDEX stores and manages the results of object detection and anomaly detection in a database. This design provides multiple benefits. First, it enables the consistent organization of information about detected objects and anomalies, including frames and object types. Second, utilizing a database optimizes search speeds through database indexing.

Indexes in a database are auxiliary data structures designed to improve the efficiency of table searches. Without indexes, a database search would require scanning the entire table, and the process could be increasingly time-consuming as the volume of data grows. Therefore, VIDEX utilizes indexing to enable rapid access to information about detected objects and anomalies stored in the database. This is particularly beneficial in a multi-threaded environment where data might not be entered sequentially. That is, by employing database indexing we not only accelerate search speed but also ensure data is organized in an effective manner.

3 Operationalization of VIDEX

VIDEX utilizes object detection and anomaly detection methods for surveillance video analysis. Object detection pinpoints the spatial coordinates and classes of specific objects within frames. This paper assumes that object detection methods can spot objects that are potentially linked to criminal activities and leverages this capability to summarize surveillance footage. On the other hand, anomaly detection aims to identify unusual activities within video data by treating these anomalies as statistical outliers. By employing anomaly detection methods, we seek to identify significant abnormal events within the video and, consequently, aid investigators in their review and evaluations. Subsequent sections will delve into the specifics of each detection method.

3.1 Object Detection

The integration of object detection methods within VIDEX opens up a large spectrum of possibilities for customizable surveillance analysis. This adaptability enables the system to identify and categorize various objects and, thus, allows law

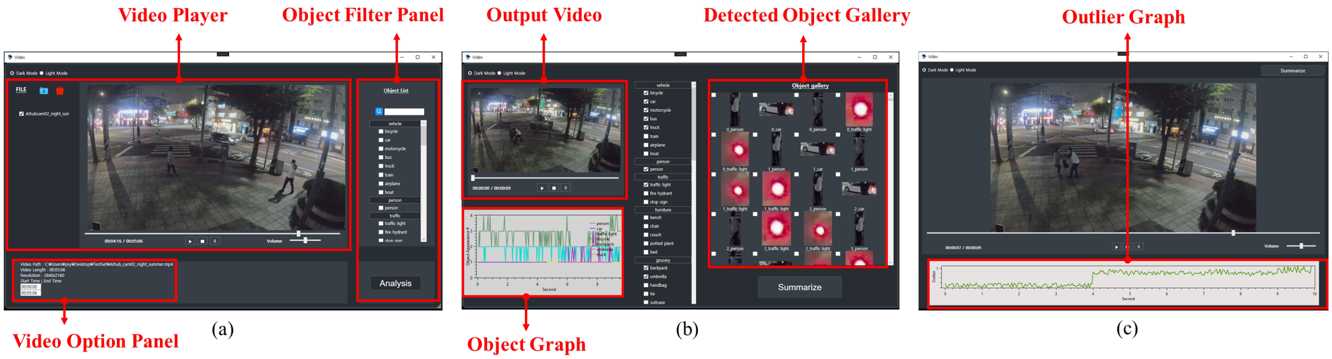

Figure 2: (a): GUI of VIDEX video file SettingView, (b): ObjectDetectionView, (c): AnomalyDetectionView .

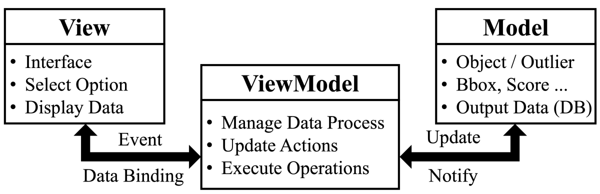

Figure 3: The Model-View-ViewModel Design of VIDEX

enforcement and security professionals to focus only on objects of interest relevant to criminal investigations. By setting parameters to recognize specific classes of objects, VIDEX becomes an invaluable tool for improving the efficiency and efficacy of surveillance operations.

In realizing this capability, VIDEX harnesses the power of YOLOv5 [Jocher, 2020], a model celebrated for its exceptional detection accuracy and rapid computational performance. This choice significantly reduces training times compared to its earlier versions. YOLOv5's strength in recognizing small objects and its minimal capacity requirement further streamline VIDEX's operational footprint, which ensures the system remains lightweight and agile. To demonstrate the system's capabilities, we employ a YOLOv5 model pretrained on the expansive MS COCO dataset, which is rich in diversity with 80 identifiable object types, including people, cars, and buses.

To integrate YOLOv5 into VIDEX, which is implemented in C#, we used ONNX for model conversion. The detected objects in the user-uploaded videos are identified with the bounding box coordinates, class labels, and confidence scores, and this information is stored in the database. VIDEX displays these details through thumbnails and graphs in the results window for easy visualization and supports image summary generation. Through this, investigators can easily and quickly proceed with the analysis without having to see what objects are in individual video frames.

3.2 Anomaly Detection

In the analysis of surveillance footage, behaviors indicative of unexpected or criminal situations – such as assault, vandalism, trespassing, kidnapping, and fainting – can be inferred [Kim *et al.*, 2023]. To identify such abnormal behavior, we employ anomaly detection techniques based on the premise that these behaviors manifested as statistical outliers. The detection process consists of two stages: feature extraction, where spatio-temporal information is extracted from video segments, and detection, where outlier detection methods are applied.

In the feature extraction stage, we use a pre-trained 3D-CNN to effectively capture spatio-temporal information that characterizes activities from segments of fixed-length videos. The input video is divided into segments, with feature embeddings extracted by 3D-CNN for each segment. These embeddings are stored in the database and subsequently analyzed by outlier detection methods in each thread to distinguish between normal and anomalous behavior.

For the detection stage, to conduct a proof-of-concept, we utilize well-known non-parametric outlier detection methods such as LOF (Local Outlier Factor) [Breunig *et al.*, 2000], cbLOF(cluster-based LOF) [He *et al.*, 2003], and iForest (isolation Forest) [Zhao *et al.*, 2018]. Each detection method is modularized and can be interchangeably switched with other generalized outlier detection methods. Feature embeddings identified as anomalies are used for graph generation and summary video creation. To enhance the clarity of detection results and improve the interpretability of summary videos, we introduce padding around the temporal axis of data points marked as anomalies.

4 Conclusion & Future Work

VIDEX is an advanced tool for surveillance video analysis. It delivers outstanding performance through the effective integration of object detection techniques (YOLOv5) and anomaly detection models. Its user-friendly GUI provides visualization of anomaly detection results to increase efficiency and accuracy in video analysis and crime scene investigation. The multi-threaded and parallel processing architecture of VIDEX significantly boosts video indexing and processing speed, while its use of cutting-edge models facilitates real-time processing capabilities and ensures high accuracy. Currently, the scope of detectable objects is confined to those within the COCO dataset. Future updates will expand this range to include detailed classes of weapons and vehicles relevant to criminal activities. This enhancement is anticipated to provide more granular classifications and further improve the investigative efficiency of the tool.

Acknowledgments

This work was partially supported by (1) Institute of Information & Communications Technology Planning & Evaluation (IITP) grant funded by the Korea government (MSIT) (No.2014-3-00123, Development of High Performance Visual BigData Discovery Platform for Large-Scale Realtime Data Analysis) and (2) the MSIT (Ministry of Science and ICT), Korea, under the National Program for Excellence in SW, supervised by the IITP in 2023 (2023-0-00055).

References

[Aggarwal, 2017] Charu C. Aggarwal. *Outlier analysis*. Springer, 2017.

[Bai *et al.*, 2019] Junjie Bai, Fang Lu, Ke Zhang, et al. Onnx: Open neural network exchange. https://github.com/onnx/onnx, 2019.

[Breunig *et al.*, 2000] Markus M Breunig, Hans-Peter Kriegel, Raymond T Ng, and Jörg Sander. Lof: identifying density-based local outliers. In *Proceedings of the 2000 ACM SIGMOD international conference on Management of data*, pages 93–104, 2000.

[Carreira and Zisserman, 2017] Joao Carreira and Andrew Zisserman. Quo vadis, action recognition? a new model and the kinetics dataset. In *proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 6299–6308, 2017.

[He *et al.*, 2003] Zengyou He, Xiaofei Xu, and Shengchun Deng. Discovering cluster-based local outliers. *Pattern recognition letters*, 24(9-10):1641–1650, 2003.

[Jocher, 2020] Glenn Jocher. Ultralytics yolov5, 2020.

[Kim *et al.*, 2023] Harim Kim, Chang Ha Lee, and Charmgil Hong. Crime scene detection in surveillance videos using variational autoencoder-based support vector data description. In *Asian Conference on Intelligent Information and Database Systems*, pages 451–464. Springer, 2023.

[Redmon *et al.*, 2016] Joseph Redmon, Santosh Divvala, Ross Girshick, and Ali Farhadi. You only look once: Unified, real-time object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 779–788, 2016.

[Zhao *et al.*, 2018] Xun Zhao, Yanhong Wu, Dik Lun Lee, and Weiwei Cui. iforest: Interpreting random forests via visual analytics. *IEEE transactions on visualization and computer graphics*, 25(1):407–416, 2018.