Unattended Baggage Monitoring in Public Stations

Ngo Le The Bach
School of Computing and Innovative Technologies
British University Vietnam
Hanoi, Vietnam
bach.nlt@st.buv.edu.vn

Hamza Mutaher
Collage of Computing, Engineering, and Built Environment
Birmingham City University
Birmingham, UK
hamza.mutaher@bcu.ac.uk

ABSTRACT

Any unattended baggage in public stations is believed to be considered as a threat to citizens. Fortunately, identification of such objects is one of computer vision problem which can be solved by AI and ML algorithm. Therefore, the need for accuracy and real-time application to increase safety in public places such as airports, train stations is mandatory. The project explores an automated framework of unattended baggage detection for video surveillance, and the stages of developing an algorithm. The various methods of performing data processing and feature selection are highlighted in this literature review. These are background subtraction, and few-shot learning applied to network classifier. First, for background subtraction, the project utilizes an algorithm that is ViBe (Video Background extractor). The interest foreground object will be extracted by comparison of centroids with the background frame over a specific period of time. he project's performance is assessed by carrying out the experiments on Abandoned Objects Dataset. Moreover, the extracted objects will go through a few-shot classifier to correctly categorize objects. The project evaluates different algorithm evaluation and scoring techniques. These are basic accuracy benchmarks for classification such as f1 score include true positive, true negative, false positive, false negative. In this project, there is also a literature review that compares previous works and identify a lack of research in background subtraction and traditional network classification. Secondly, for image classification, this project uses few-shot learning and explore its strength compared to traditional CNN-based image classification. Fewshot learning is particularly used in scenarios where data is insufficient due to expensive collecting and annotating. This approach not only makes an application cost-effective but also reduces the computational resources.

Keywords: Unattended baggage detection, Few-shot learning, Image Classification, ViBe algorithm, Background subtraction

1. Introduction

In contemporary settings, computer vision emerges as a fundamental domain within Artificial Intelligence, encapsulating processes such as acquisition, analysis, and recognition of digital images and videos, including the crucial aspect of object detection. This component is notably critical due to its extensive applications in surveillance and security systems, as highlighted by Dwivedi et al (2020). Traditional surveillance methodologies, which rely on multiple personnel to monitor many video feeds, are prone to human error. However, the advent of advancements in Artificial Intelligence and Machine Learning has revolutionized these systems by automating the visual inspection processes, thereby enhancing both the efficiency and accuracy of such systems. The necessity for robust video surveillance systems has been accentuated by the increasing prevalence of security threats and breaches, along with the intrinsic limitations of human monitoring. The safety of public spaces such as airports, shopping centers, and public squares is compromised by various risks including theft, physical confrontations, or the concealment of hazardous materials within items like suitcases or backpacks. These items must be swiftly identified and reported to security personnel. A predominant challenge within existing surveillance systems is their often-delayed detection capabilities, which can result in incidents that lead to public distress and numerous casualties.

Effective surveillance should focus on continuous monitoring and the prompt and precise assessment of potential threats. It is critical that no unattended item is overlooked; however, the response to such detections should be proportionate to the visual observations and any additional information known about the item. While a zero-tolerance policy towards unattended objects is advisable, evaluations must be balanced, considering both the visual characteristics and any pertinent contextual information. For instance, while an unattended piece of luggage may require merely locating the owner nearby, more urgent action is warranted if the item is deliberately hidden and exhibits signs of potential danger such as exposed wires, circuit boards, batteries, adhesive tapes, or unusual substances.

Furthermore, if an unattended item is discovered following a report of suspicious activity, an escalated security response is imperative. To enhance the capabilities of surveillance systems effectively, it is crucial to integrate innovative visual detection technologies and modern electronic devices. These technologies not only aid in quickly identifying suspicious behaviors but also ensure the rapid communication of such findings to security teams for immediate action. The modernization of security systems through advanced computer vision techniques represents a proactive approach to managing security risks in public spaces. By automating the detection and analysis of potential threats, these systems significantly reduce the dependency on human monitoring, diminish the likelihood of human error, and enhance the overall safety and security of public environments. This paper is organized as, section 2 investigates the literature review, section 3 discusses the methodology of the work, section 4 explores the implementation and section 5 discusses the work result.

2. Literature Review

The increasing need for robust security measures in densely populated public areas such as airports, train stations, and shopping centers has become a critical concern. As urban environments grow and become more crowded, the limitations of traditional security systems, which often rely heavily on manual monitoring by security personnel, are becoming evident. These traditional methods are resource-intensive and prone to human errors, which can be particularly problematic in extensive, crowded settings. Thus, there is an urgent need for innovative solutions that can enhance the management and detection of potential security threats, such as unattended baggage, which poses significant safety risks.

This proposal introduces a sophisticated surveillance method designed to improve the rapid identification and management of unattended items to enhance public safety effectively. By utilizing advanced technological solutions, this system aims to transform traditional reactive security measures into proactive security management.

The proposed system operates through a two-stage approach:

- Background Subtraction: This initial phase employs algorithms to differentiate stationary objects from their moving backgrounds. This capability is crucial as it helps to spotlight objects like unattended baggage that remain static amid the bustling activity of a public space. Identifying these objects as regions of interest is the first step toward assessing potential threats (Piccardi, 2004).
- Object Classification: Following the isolation of these objects, the system examines their nature. Advanced classification algorithms analyze each item's physical characteristics to categorize them as either benign personal belongings or potential security threats. This classification relies on various visual indicators such as size, shape, and other distinctive features that might suggest a risk.

By integrating these two processes, the surveillance system is not only capable of detecting but also classifying unattended objects with speed and accuracy. This functionality enables security personnel to prioritize their responses more effectively, focusing on genuine threats while reducing the risk of oversight. Therefore, this also enhanced surveillance system, with its stateof-the-art detection and classification capabilities, promises a substantial improvement in the security of public spaces. Offering a more precise, efficient, and responsive method for managing potential threats, it meets the modern needs for public safety and marks a crucial advancement in proactive security planning. This system is not merely a technological enhancement but an essential component of future-oriented security strategies that will make public environments safer for everyone.

2.1. ABODA Dataset

The Abandoned Baggage Object DAtaset dataset (ABODA, 2025), specifically designed for the task of Abandoned Object Detection (AOD), serves as a primary resource for this project. This dataset leverages video surveillance footage from diverse sources, intentionally chosen to reflect a wide range of scenarios related to AOD challenges (Choudhry et al., 2023). These scenarios feature diverse camera angles, crowded situations, and environmental conditions across various public spaces, which makes exceptionally suitable for this project. It contains eleven video sequences including indoor and outdoor scenes that capture a diverse range of real-world scenarios, presenting significant challenges for detection algorithms. These scenarios include complex scenes with crowded environments, varied lighting conditions (including night-time situations), and both indoor and outdoor settings. This rich tapestry of real-world complexities is crucial for training robust detection models that can effectively operate under diverse conditions encountered in practical surveillance and security applications. The open-source nature of ABODA presents a significant advantage. Its public availability simplifies access and eliminates the complexities associated with licensing restrictions. Nevertheless, this dataset remains unannotated, and dedicating time to label the data is not a viable option within the scope of this project.

2.2. Background Subtraction

Background subtraction, also known as foreground segmentation, typically initiates video processing workflows, identifying moving objects and foreground entities for subsequent examination. This widely used approach is to detect moving objects in videos from static cameras. Despite being a fundamental task in computer vision, background subtraction remains a complex challenge due to the diverse array of scenarios it encounters (Garcia-Garcia et al ,. 2020). Factors such as weather conditions, fluctuations in illumination, dynamic backgrounds, and low frame rates contribute to the intricacies of this task, making it an ongoing area of research and development within the field (Bou et al., 2022).

- Background Model Creation: An initial model that represents the background scene needs to be established. This model can be a simple statistical representation (e.g., average pixel values) or a more complex model based on learning algorithms.
- Foreground Segmentation: Each incoming frame is compared against the background model. Significant deviations from the model are classified as belonging to the foreground (potential object).
- Background Model Update: To adapt to gradual scene changes (e.g., illumination variations), the background model needs to be updated over time. This ensures the model remains representative of the current background scene.

In the past, background subtraction primarily employed unsupervised approaches, which utilizes algorithm of computer vision techniques. This trend has shifted in recent years with the introduction of supervised methods that capitalize on the capabilities of deep neural networks for more effective background subtraction.

2.2.1. Unsupervised Method

Unsupervised methods present a straightforward and efficient technique for background subtraction, particularly advantageous in environments where labelled data are unavailable or impractical to obtain. These methods are distinguished by their reliance on algorithmic processes rather than pre-trained models, facilitating significant computational efficiency.

- Construction of a Background Model: Initially, the algorithm constructs a baseline model that encapsulates the typical background scene. This model serves as a benchmark against which new frames are evaluated. There are several methods for background modelling include:
  - Frame Differencing: This simplest approach involves subtracting the current frame from a reference background image. The resulting difference image highlights areas with significant intensity changes, potentially indicating foreground objects. However, it is highly susceptible to noise and illumination variations.
  - Median Filtering: This technique utilizes the median value within a local neighborhood of pixels in the video sequence to represent the background. It offers better noise resistance than frame differencing but struggles with fast-moving objects.
  - Background Modelling with Running Average: This method maintains a running average of pixel intensities over time, establishing a dynamic background model. Pixels deviating significantly from the average are classified as foreground. While effective for static backgrounds, it can struggle with gradual scene changes.
  - Gaussian Mixture Models (GMM): It represents the background as a weighted sum of multiple Gaussian distributions. This allows for modelling background variations like shadows or flickering lights. However, it requires careful parameter tuning and can be computationally expensive.
  - Codebook Models: These methods partition the background into representative codewords (e.g., groups of pixels sharing similar intensity values). The current frame pixels are compared against the codebook to determine foreground or background classification. This approach offers efficiency but might struggle with complex or dynamic backgrounds.
- Foreground Segmentation: As new video frames are introduced; they are compared against the established background model. This comparison aims to pinpoint pixels that exhibit substantial deviations, which are presumed to represent foreground elements or moving objects.
- Adaptive Model Updates: After the frame analysis, the background model is dynamically updated to integrate changes observed in the scene. This adaptation is crucial for maintaining the accuracy of the model in the face of gradual environmental shifts, such as lighting changes or the introduction of new static objects into the scene.

This approach not only eliminates the need for manually label data source but also ensures that the system remains robust and responsive to evolving conditions within the observed environment.

2.2.2. Supervised Method

Recent years have witnessed a transformative shift in computer vision (Park et al ,. 2021) through the integration of Convolutional Neural Networks (CNNs). These networks have revolutionized tasks like image classification, semantic segmentation, and image reconstruction. Their ability to learn complex, high-level representations has solidified their position as a cornerstone for advancements in machine learning research. This technological prowess extends significantly to background subtraction, where innovative CNN-based methods have been developed to refine or entirely replace traditional workflows.

- Deep Learning-based Methods: One noteworthy application of deep learning in background subtraction involves streamlining the background modelling process.
  - For example, Braham & Van Droogenbroeck (2016) have pioneered a method where a CNN is specifically trained to remove backgrounds. This approach utilizes scene-specific data, eliminating the need for intricate background model constructions commonly used in traditional methods.
  - Alternatively, Babaee et al. introduced a universal CNN model that can learn from multiple scenes simultaneously. This model leverages binary masks from the SuBSENSE algorithm as a foundation for learning, enhancing its generalizability for background subtraction tasks. Furthermore, Sakkos et al. have proposed a comprehensive technique that capitalizes on temporal information through 3D convolutions. This method integrates insights from previous frames to foster a more profound understanding of scene dynamics.

- Random Forests: These ensemble learning methods combine multiple decision trees trained on different subsets of the training data. This offers robustness to noise and outliers but requires careful parameter tuning.
- Deep Learning-based Methods: Convolutional Neural Networks (CNNs) are increasingly being used for background subtraction. Trained on labelled video datasets, CNNs can learn complex features that distinguish foreground objects from even complex backgrounds. However, this approach offers high accuracy but requires significant training data and computational resources.

While supervised CNN methods demonstrate superior accuracy compared to traditional background subtraction techniques, they are not without limitations. These methods often require substantial computational resources and extensive volumes of annotated data for effective training. This allows for more robust and adaptable models compared to unsupervised methods. However, it can pose a challenge for real-time deployment scenarios.

2.3. Few-shot Learning

As previously mentioned, a significant constraint for this project is the limited time available for extensive data collection and labelling. Traditional deep learning approaches often necessitate vast amounts of labelled data to achieve optimal performance. However, Few-Shot Learning (FSL) presents a compelling solution for limited data scenario. FSL is believed to perform well in such situations where labelled data is scarce. By leveraging its ability to learn effectively from a limited number of labelled examples per class, FSL can potentially address the time constraints associated with data collection and labelling for the classification task within this project (Hu et al., 2020).

2.3.1. Few-shot Variations

The realm of Few-Shot Learning encompasses a spectrum of scenarios based on the number of labelled examples available per class. Here is a breakdown of the different variations:

- Few-Shot Learning: This term is often used as a broader umbrella term encompassing all scenarios where the number of labeled samples per class is limited (typically K <= 5).
- One-Shot Learning (OSL): This is an extreme case of FSL where the model is trained with only one libeled example per class (K = 1). While challenging, OSL algorithms aim to leverage prior knowledge or external information to achieve classification on unseen examples.
- Zero-Shot Learning (ZSL): This scenario pushes the boundaries of FSL by attempting to classify objects from completely unseen classes during training (K = 0). ZSL models rely heavily on semantic relationships between classes and external knowledge sources like word embeddings to perform classification. While still an active research area, ZSL holds promise for situations where libeled data is entirely unavailable for new classes.

2.3.2. Approaches to Few-Shot Learning

Within the realm of Few-Shot Learning (FSL), the most prevalent formulation is N-way, Kshot image classification. This paradigm necessitates the model to distinguish between N distinct classes, with K labelled samples available per class for training. The labelled samples constitute the support set, which serves as the foundation for the model's learning process.

For example, imagine a scenario where the task aims to classify various dog breeds using FSL. If N equals 3, the model would be tasked with differentiating between three dog classes, such as Labrador, Saint-Bernard, and Pug. Furthermore, suppose K is 2, signifying that the support set contains only two libeled images per dog class. The FSL challenge lies in classifying a separate query set of unlabeled dog images (denoted by Q) into one of the three classes using the limited information gleaned from the support set. While this task might seem intuitive for humans, even with limited exposure to specific dog breeds, replicating this capability in AI necessitates advanced techniques like Meta-learning. Meta-learning empowers AI models to leverage their learning experience to effectively tackle new classification tasks with minimal data, making it a crucial tool for FSL applications.

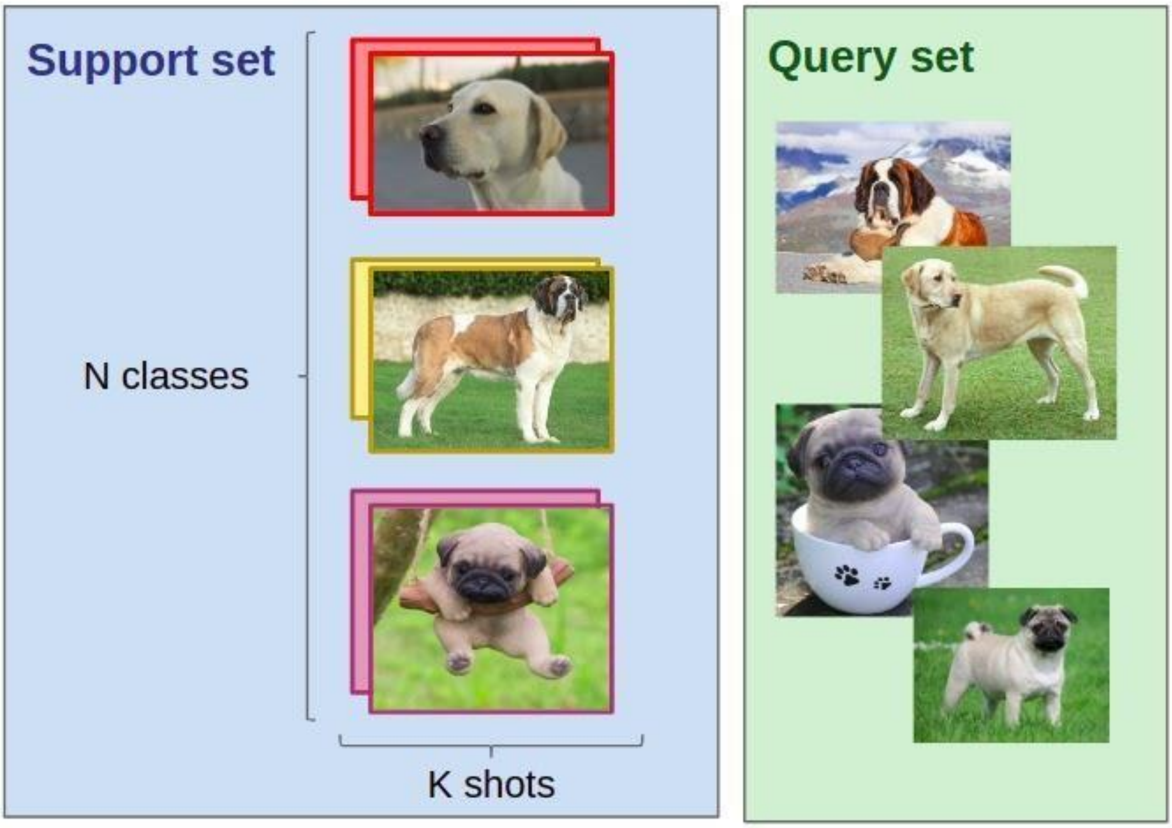


Figure 1: Few-shot Image Classification Task (Bennequin, 2022)

Few-Shot Learning (FSL) hinges on a crucial concept: acquiring knowledge from similar problems encountered previously. This connection aligns FSL with the principles of Meta-Learning (Dhillon et al,. 2019).

In traditional classification tasks, a model is trained on a large dataset to distinguish between categories. This trained model is then evaluated on a separate dataset to assess its performance. However, Meta-Learning takes a different approach. Instead of focusing solely on a single dataset, it aims to learn how to learn in general. This is achieved by analyzing a wide range of classification problems. By examining these problems, the model can develop the ability to solve new, unseen problems even with limited data.

3. Methodology

This research employs the Saunders onion research methodology cited by (Phair & Warren, 2021). The methodology is introduced in “Research Method for Business Students” which has five layers and a core as described below.

The outermost layer defines the research philosophy, with choices including Positivism, which focuses on objective realities, Interpretivism, highlighting subjective ideas, and Pragmatism, employing the most effective methodologies. The subsequent layer, the research approach, can follow an inductive path where research leads to theory formation, or a deductive path where existing theories are tested. This approach is intrinsically linked to the qualitative research type, which probes into experiences and meanings, or quantitative, which concentrates on quantifiable data.

The third layer addresses the research strategies available, such as experimental research, which tests theories under controlled conditions, and case studies, which involve detailed examinations of instances. Additional methods encompass grounded theory, which allows data to shape the research questions, ethnography, studying subjects within their natural settings. The following layer concerns data type choices, ranging from mono-method, which uses solely qualitative or quantitative data, to mixed-method, which integrates both, and multi-method, employing various techniques within each type.

The final consideration is the time horizon for data collection. It is an important step because gathering data is often the beginning of the project. It includes distinguishing between crosssectional studies, where data is collected at one time, and longitudinal studies, where data is gathered over several points in time. This framework by (Phair & Warren, 2021) helps ensure that research methodology is conducted with a structured approach, which shortens researching time and facilitates systematic investigation and credible results. And makes the project stronger. It is also important to have correctly done research architectures to not waste time.

4. Implementation

The project has a brief overview pipeline for an automated unattended baggage detection. This includes two-stage approach that utilizes advanced background subtraction algorithm ViBe and few-shot learning classification techniques.

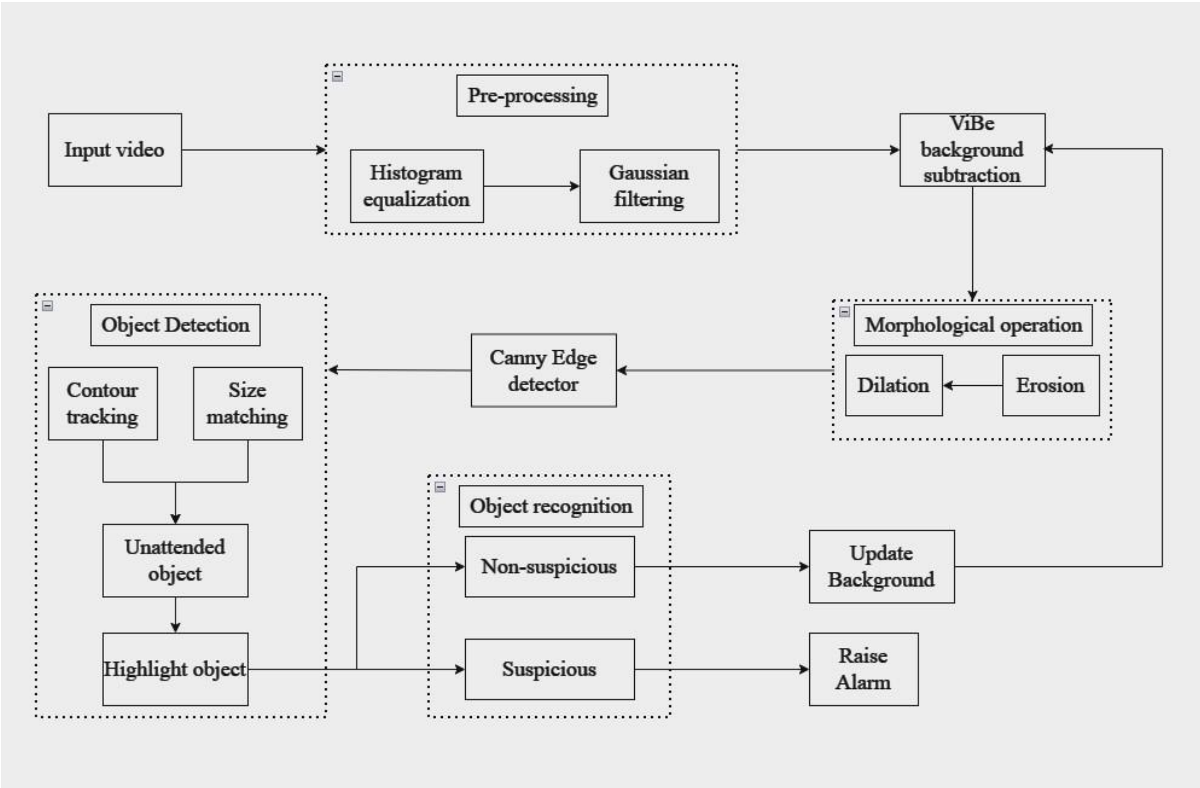


Figure 2: System workflow (Dwivedi et al., 2020)

Firstly, the video input goes through a pre-processing step, where histogram equalization and Gaussian filtering are applied to enhance image quality, adjust contrast, and reduce noise for each frame. This is crucial for the subsequent application of the ViBe background subtraction algorithm, which effectively separates moving objects from static backgrounds by maintaining dynamic background models. Despite being over a decade old, the ViBe algorithm remains popular due to its simplicity and speed. ViBe boosts processing speeds of up to 200 frames per second, making it ideal for real-time applications. However, its disadvantages are that a combination of different techniques must be done to get rid of redundant objects as well as more efficiency the algorithm (Droogenbroeck & Paquot, 2012). Compared to some other statistical background subtraction algorithm, it may have less accuracy. The segmentation process allows us to focus on the isolation of potential objects after comparing with the background initialization. After background subtraction, a binary map consists of background and foreground in the output.

Secondly, after the foreground segmentation, the system moves to object localization phase. Here, a Canny edge detector algorithm defines the contours of foreground segmentation, which are then further refined through morphological operations. They are dilation and erosion to get a well-shaped, improving the accuracy of the contour outputs. And the segmentation map of the contour will be shown after. The resulting object contours go through size matching section to filter out inappropriate detections based on predefined size parameters, which focuses only on objects that match the criteria for potential unattended baggage. After a period of time, in this project, it is designed for the staying static objects for more than 30 seconds, then the centroids of the objects will be saved.

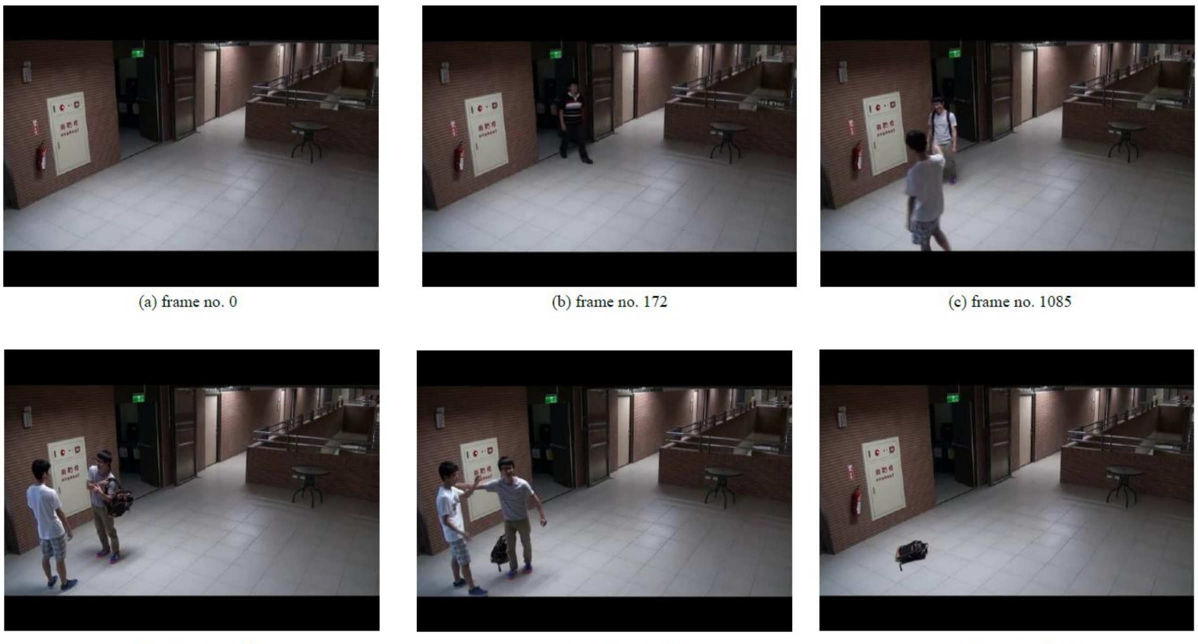
The final stage of the framework involves object recognition, where a few-shot classifier categorizes each extracted Roi as suspicious or non-suspicious. This classification is crucial for real-time security applications, as it determines whether to highlight an object for further inspection or update the background model to ignore non-threatening items. If an object is classified as suspicious, the system raises an alarm, prompting immediate security responses. This artifact not only automates the traditional monitoring process but also enhances the reliability and efficiency of surveillance systems in identifying potential security threats.

Normally the output of object detection models will have a lot of bounding boxes. With such output, there will be a phenomenon of having many bounding boxes for the same object, which causes redundancy of information when our purpose only needs one bounding box for one object. Therefore, in this project, non-max suppression algorithm is used to remove the redundant bounding box (Zou et al ,. 2023).

5. Result

After background subtraction algorithm works, the static objects known as ROI are extracted from dynamic video. They will go through a few-shot classifier for further researched. The baggage type to be detected is the type of bags that are commonly used by travellers. Therefore, in this project, there are total of 3 classes: 2 main baggage types: backpack, suitcase and people for clear explanation. In the future, for further development, there are more baggage types to be included such as carton box, clutch, duffel, purse.

5.1. Actual frame

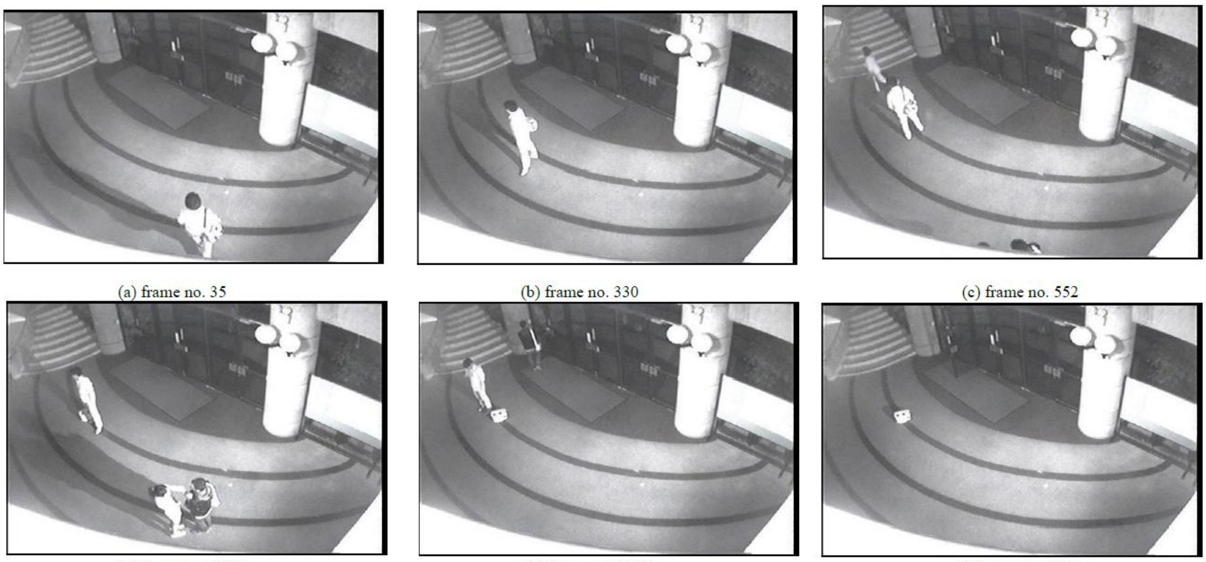


(a) frame no. 0 (b) frame no. 172 (c) frame no. 1085

(d) frame no. 1219 (e) frame no. 1367 (f) frame no. 1613

Figure 3: Normal lightning and condition in an indoor environment

Figure 3 are recorded in natural light with no illumination change and a dropped bag situation



(a) frame no. 35 (b) frame no. 330 (c) frame no. 552

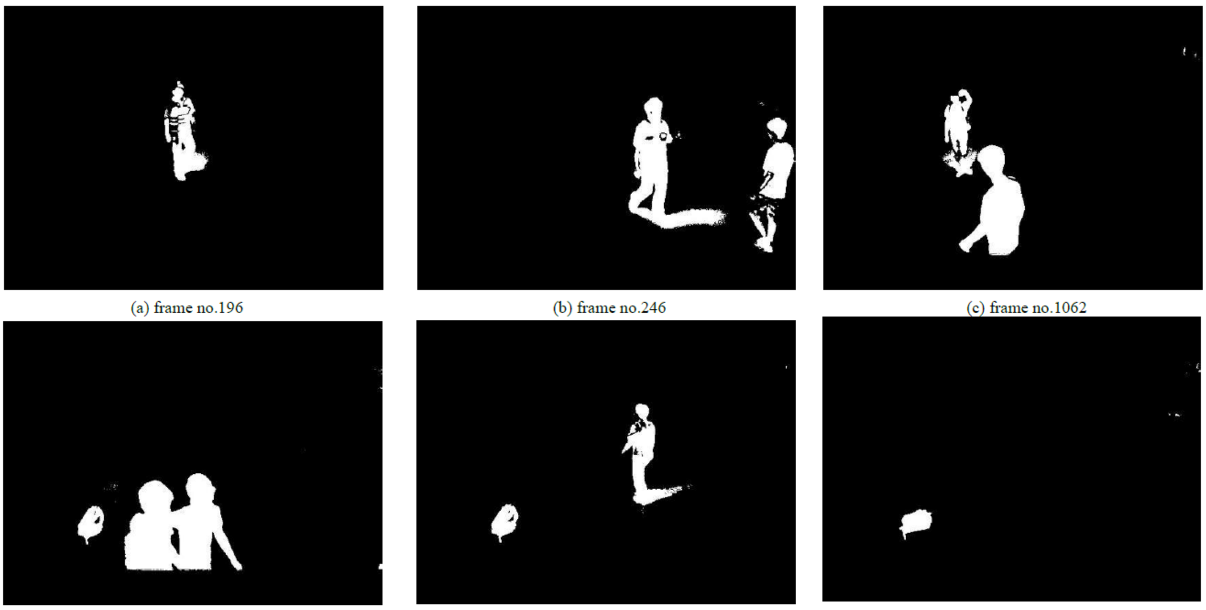
(d) frame no. 970 (e) frame no. 1789 (f) frame no. 2306

Figure 4: Nighttime recorded outdoor

This sequence of frames show original video at outdoor environment and nighttime. There is a challenge when it is at night compared to daylight because the shadow of people and objects can cover it, making the baggage become hard to extract.

5.2. Result

In this section, after conducting on all eleven videos of ABODA dataset and the first frame is set as the background. The segmentation map to localize the objects is presented.



(a) frame no.196 (b) frame no.246 (c) frame no.1062

(d) frame no. 1431 (e) frame no.1497 (f) frame no.1930

Figure 5: Segmentation map under normal conditions

Figure 5 shows the segmentation map of video 1 in ABODA dataset. The unattended bag is correctly identified with well-shaped.

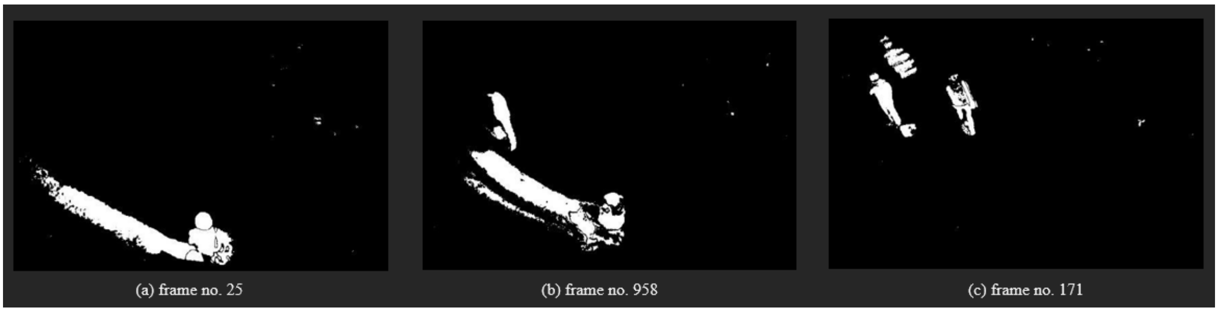


Figure 6: Segmentation map of an outdoor night-time scene

Figure 6 shows the video recorded at outdoor in night-time. Because it is at night, therefore, people and objects will have their shadow and make the background subtraction impossible to detect and extract if the shadow covers the objects.

The results form that in both cases which include the outdoor and indoor environment, if the camera stays static and the illumination effects do not appear then the unattended objects are correctly detected. However, in dynamic background, the project may not accurately localize the object which is likely to mark it as background and ignore it or raise false alarms.

5.1.2. Zero-shot classification

In this section, the zero-shot classification used in this project is presented. It consists of 5 key steps to implement a zero-shot classification.

- Model selection: Select and load the Clip model with the appropriate architecture, for example ViT-B/32. This model is chosen due to its ability to effectively perform zero-shot classification tasks, where it uses natural language understanding to classify images.
- Data preparation: Firstly, with image processing, the project loads the target image using the Python Imaging Library (PIL) and converts the image to RGB format to maintain consistency in colour channel interpretation. Secondly, for preprocessing, applying the preprocessing transformations defined by the CLIP model typically includes resizing the image to the dimensions expected by the model, normalizing pixel values, and converting the image into a tensor. Finally, a pair of image and corresponding class text is created. These prompts should mimic natural language descriptions that CLIP was trained on, enhancing the model's ability to relate text vectors to image vectors. The dataset has 5 classes: backpack, people, handbag, suitcase, non-object.
- Features and similarities computation: Image and text encoding are computed to extract its features. Moreover, in this project, normalization is also added to avoid overfitting. Then, the computation of cosine similarity between the encoded image vector and text vector to map them. Next, SoftMax function is used to convert them into probability values, which helps in categorizing the relevant class for the extracted images.
- Result display: Identify the class with highest SoftMax scores as the predicted category for the image.

Figure 7: Extracted ROI of the system

If the classifier returns one of two baggage classes: backpacks, suitcases then the system will raise an alarm to the security team. However, if the classifier returns people class as predicted output, then the background subtraction model will have an update function to make that object become background.

In this project, benchmark plays a vital section to measure how successful the system can detect. As part of the evaluation section, to measure the performance of the project, the project utilizes the benchmark established by Dwivedi et al. (2020) for unattended object detection using contour formation through background subtraction. There are three objective benchmarks: Correct Object Detection Rate (CODR), Object Success Rate (OSR), False Alarm Rate (FAR). The percentage of successfully recognized unattended objects is known as the CODR. The fraction of incorrectly detected unattended items to all unattended object identifications is known as the false alarm rate, or FAR. Accordingly, the Object Success Rate of CODR and FAR definitions of this approach are provided. Equations 1, 2, and 3 are used to compute CODR, FAR, and OSR as follows.

$$CODR = \frac{TP}{TP + FN} \tag{1}$$

$$FAR = \frac{FP}{TP + FP} \tag{2}$$

$$OSR = \frac{CODR}{CODR + FAR} \tag{3}$$

Figure 8: Benchmark formulas for unattended object detection (Dwivedi et al., 2020)

True Positive (TP) is the total of correctly identified unattended objects. On the other hand, False Positive (FP) is the wrong number. False Negative (FN) is the number of unattended objects that are not detected.

Video	Scenario	Illumination	Unattended objects	TP	FP	FN	CODR (%)	FAR (%)	OSR (%)
Video1	Indoor	No	01	01	00	00	100	00	100
Video2	Outdoor	No	01	01	00	00	100	00	100
Video3	Outdoor	No	01	01	00	00	100	00	100
Video4	Outdoor	No	01	01	00	00	100	000	00
Video5	Outdoor	Yes	01	01	00	01	100	00	100
Video6	Outdoor	Yes	02	01	01	01	50	50	00
Video7	Indoor	Yes	01	00	00	01	00	00	00
Video8	Indoor	Yes	01	00	00	01	00	00	00
Video9	Indoor	Yes	01	01	00	00	100	00	100
Video10	Indoor	Yes	01	01	00	00	100	00	100
Video11	Indoor	No	02	00	01	01	00	00	00

Table 1: Results of unattended objects in ABODA dataset

The CODR shows the successful detected baggage and categorize into its classes. However, the few-shot classifier fails to detect some objects that have the small missing edges as the FAR showing in some situations.

Conclusion and Future Work

This project successfully developed an automated detection system for unattended baggage in public areas, integrating advanced techniques such as background subtraction for object detection and few-shot image classification for identifying objects. Extensive testing on the ABODA dataset under a variety of environmental conditions demonstrated the system's robustness, affirming its potential utility in real-world applications. The system exhibited high performance during typical day and night conditions when illumination variations were minimal. However, challenges arose in crowded environments and under intense lighting conditions, where the incidence of blurring and shadows significantly impacted system accuracy. These scenarios highlighted the limitations of the current model, particularly in complex settings characterized by variable lighting and high pedestrian traffic. Looking forward, the project aims to refine the object detection capabilities to better handle diverse and challenging environments. Enhancements will focus on improving resilience to lighting effects and reducing false positives in densely populated areas. Additionally, the utility of few-shot image classification in practical scenarios presents a promising avenue for future research. This technique holds significant promise for expanding the system's adaptability and efficiency, making it a valuable tool for enhancing security measures in public spaces. By advancing these technologies, the project seeks to contribute to safer public environments, reducing risks associated with unattended items through more accurate and timely detection. Apply tracking algorithm to identify who has dropped the objects. Report unattended baggage as well as the one who dropped it. For background subtraction, it has some strategies to adapt to illumination changes. In this project, for lightweight model, each frame is calculated with the additional mean standard deviation formula against a subtraction model. If the deviation is high, the model will trigger a reset of background model for adaptation. After categorizing the object with few-shot classifier, if it is not unattended then the background model should be updated as well.

References

ABODA Dataset. (2015). *ABODA dataset*. GitHub. Retrieved December 10, 2023, from https://github.com/kevinlin311tw/ABODA

Bennequin, E., Tami, M., Toubhans, A., & Hudelot, C. (2022). Few-shot image classification benchmarks are too far from reality: Build back better with semantic task sampling. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 4767-4776). 10.1109/CVPRW56347.2022.00523

Bou, X., Ehret, T., Facciolo, G., Morel, J. M., & von Gioi, R. G. (2022). Reviewing ViBe, a popular background subtraction algorithm for real-time applications. Image Processing On Line, 12, 527-549. https://doi.org/10.5201/ipol.2022.434

Braham, M., & Van Droogenbroeck, M. (2016, May). Deep background subtraction with scene-specific convolutional neural networks. In 2016 international conference on systems, signals and image processing (IWSSIP) (pp. 1-4). IEEE. https://doi.org/10.1109/IWSSIP.2016.7502717

Choudhry, N., Abawajy, J., Huda, S., & Rao, I. (2023). A comprehensive survey of machine learning methods for surveillance videos anomaly detection. IEEE Access, 11, 114680-114713. 10.1109/ACCESS.2023.3321800

Dwivedi, N., Singh, D. K., & Kushwaha, D. S. (2020). An approach for unattended object detection through contour formation using background subtraction. Procedia Computer Science, 171, 1979-1988. https://doi.org/10.1016/j.procs.2020.04.212

Dhillon, G. S., Chaudhari, P., Ravichandran, A., & Soatto, S. (2019). A baseline for few-shot image classification. arXiv preprint arXiv:1909.02729.https://doi.org/10.48550/arXiv.1909.02729

Garcia-Garcia, B., Bouwmans, T., & Silva, A. J. R. (2020). Background subtraction in real applications: Challenges, current models and future directions. Computer Science Review, 35, 100204.https://doi.org/10.1016/j.cosrev.2019.100204

Hu, S. X., Li, D., Stühmer, J., Kim, M., & Hospedales, T. M. (2022). Pushing the limits of simple pipelines for few-shot learning: External data and fine-tuning make a difference. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 9068-9077). https://arxiv.org/abs/2204.07305

Park, S.-S., Tran, V.-T., & Lee, D.-E. (2021). Application of various YOLO models for computer vision-based real-time pothole detection. *Applied Sciences, 11*(23), 11229. https://doi.org/10.3390/app112311229

Phair, D., & Warren, K. (2021). Saunders' research onion: explained simply. Peeling the onion, layer by layer. Johannesburg: GradCoach. https://gradcoach.com/saunders-research-onion

Piccardi, M. (2004, October). Background subtraction techniques: a review. In 2004 IEEE international conference on systems, man and cybernetics (IEEE Cat. No. 04CH37583) (Vol. 4, pp. 3099-3104). IEEE. https://doi.org/10.1109/ICSMC.2004.1400815

Van Droogenbroeck, M., & Paquot, O. (2012, June). Background subtraction: Experiments and improvements for ViBe. In 2012 IEEE computer society conference on computer vision and pattern recognition workshops (pp. 32-37). IEEE. https://doi.org/10.1109/CVPRW.2012.6238924

Zou, Z., Chen, K., Shi, Z., Guo, Y., & Ye, J. (2023). Object detection in 20 years: A survey. Proceedings of the IEEE, 111(3), 257-276. https://doi.org/10.1109/JPROC.2023.3238524