KISA Dataset

1. Dataset Description

Korea Internet & Security Agency (KISA) is a specialized organization in the information security field and provides performance test dataset and evaluation certification to strengthen the technological competitiveness of intelligent CCTV, which is an important element of physical security. The video DB is divided into distribution/testing/certification purposes. Only distribution and testing purposes are made public, while authentication purposes are kept private. Table 1 describes the dataset.

Table 1. KISA Dataset Description

Distribution Set	Testing Set	Certificate Set
• Available for download when visiting KISA • Having minimum number of videos in each category • Using for online distribution in the future	• Available through streaming when visiting KISA • Having minimal footages in each scenario • More clips than distribution set	• Private access • Having more clips than testing and distribution set

The dataset is divided into six main categories: **Loitering, Intrusion, Abandonment, Falldown, Violence (Fighting) and Fire Detection.** The other minority categories are Queueing and People Counting. The evaluation and certification standards for intelligent CCTV require that 90% or more accurate detection is given in each of the six main categories. Recorded videos are captured at resolution 1280x720 and 30 FPS. Table 2 summarizes the KISA dataset.

Table 2. KISA Dataset Summary

No	Scenarios	Description	# of videos	# of frames	Duration (s)
1	**Loitering**	Wandering and standing in certain area	30	331,390	11,046
2	**Intrusion**	Trespassing, jumping into the water, jumping over the fence, step into the railway	30	297,420	9,914
3	**Abandonment**	Bag left unattended	10	94,289	3,143
4	**Falldown**	Falling to the ground or a lower surface	10	98,644	3,288
5	**Violence**	Fighting between two people	10	108,793	3,626
6	**Fire Detection**	Flame and smoke detection	10	109,166	3,639
7	Queueing	Waiting in line	5	131,610	4,387
8	People Counting	Counting human instances in specific areas	5	72,519	2,417

1.1 Loitering

Loitering refers to the act of lingering or remaining in a particular place without any apparent purpose or activity. It often involves staying in a location for an extended period, often longer than what is considered normal or acceptable, without a clear reason. Loitering can sometimes be associated with suspicious behavior, as it may be perceived as unusual or potentially indicative of unlawful activities. In certain contexts, such as public spaces or businesses, loitering rules or regulations may be in place to maintain order and security. The KISA dataset includes 30 videos of loitering, along with corresponding labels in different locations. Fig. 1 shows some sample images in this scenario.

Figure 1. Samples image in loitering scenario.

1.2 Intrusion

Intrusion refers to the act of entering a place, system, or space without permission or authorization. It involves the unauthorized or unwanted entry into an area, often with the intent to access, disrupt, or compromise something within that space. The KISA dataset includes 30 videos of intrusion, along with corresponding labels in different locations. Fig. 2 shows some sample images in this scenario.

Figure 2. Samples image in intrusion scenario.

1.3 Abandonment

The abandonment refers to a scenario in which a bag or package is found in a public place or a location where it is not under the immediate supervision of its owner, and there is no one present who appears to be responsible for or in control of the bag. This scenario can raise security and safety concerns, particularly in crowded or sensitive areas, as it may suggest the possibility of an unattended item containing hazardous materials or posing a security threat. The KISA dataset includes 10 videos of abandonment, along with corresponding labels in different locations. Fig. 3 shows some sample images in this scenario.

Figure 3. Samples image in abandonment scenario.

1.4 Falldown

Falldown scenario typically refers to situations where individuals experience falls or accidents that result in them falling to the ground or a lower surface. These scenarios can vary widely in their nature and can occur in various contexts, such as home environments, workplaces, public spaces, and healthcare settings. Understanding and addressing fall scenarios is essential for injury prevention and the safety and well-being of individuals. The KISA dataset includes 10 videos of falldown, along with corresponding labels in different locations. Fig. 4 shows some sample images in this scenario.

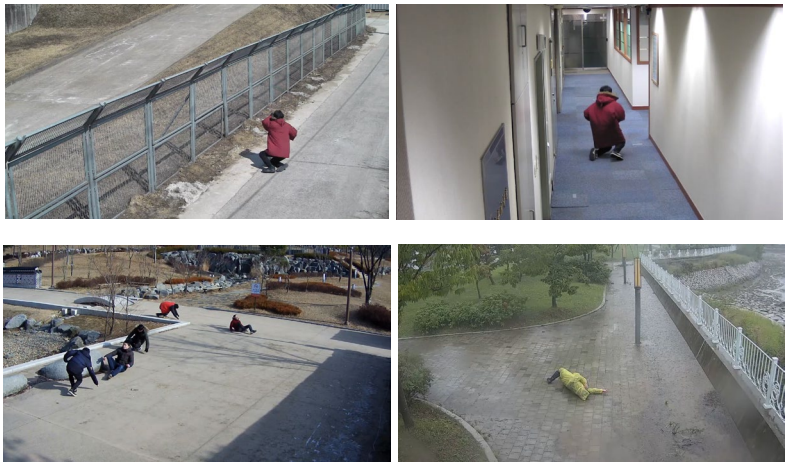

Figure 4. Samples image in falldown scenario.

1.5 Violence

Violence scenarios refers to situations or contexts in which violent or aggressive behavior occurs. Violence can take many forms, ranging from physical altercations to verbal abuse and may occur in various settings. Understanding and addressing violence scenarios are essential for maintaining safety, security, and well-being in different environments. The KISA dataset includes 10 videos of violence, along with corresponding labels in different locations. Fig. 5 shows some sample images in this scenario.

Figure 5. Samples image in violence scenario.

1.6 Fire Detection

Fire detection scenario refers to situations and contexts where the detection of fires or smoke is crucial for safety and protection. Detecting fires in their early stages is essential for preventing the spread of flames, minimizing property damage, and ensuring the safety of individuals. Various fire detection scenarios exist, and the methods used for detection may vary depending on the environment and potential sources of fire. The KISA dataset includes 10 videos of fire detection, along with corresponding labels in different locations. Fig. 6 shows some sample images in this scenario.

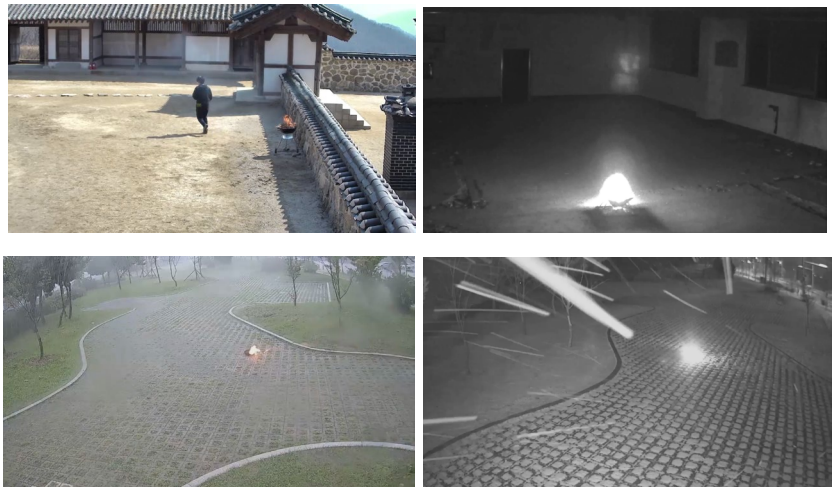

Figure 6. Samples image in fire detection scenario.

1.7 Queueing and People Counting

Queueing and people counting scenarios involve the monitoring and management of individuals' movements in various settings. These scenarios are essential for optimizing operational efficiency, ensuring crowd control, and enhancing safety and security. The KISA dataset includes 5 videos for each scenario, along with corresponding labels in different locations. Fig. 7 shows some sample images in these scenarios.

Figure 7. Samples image in queueing and people counting scenario. Top row: queueing, Bottom row: people counting.

2. Annotation Description

The provided XML serves as an example of a label file and is structured as follows:

```xml
<?xml version="1.0" encoding="utf-8"?>
<KisaLibraryIndex>
    <Library>
        <Scenario>Intrusion</Scenario>
        <Dataset>KISA2016</Dataset>
        <Libversion>1.0</Libversion>
        <Clip>
            <Header>
                <Filename>C014102_003.mp4</Filename>
                <Stage>1</Stage>
                <Duration>00:06:34</Duration>
                <AlarmEvents>1</AlarmEvents>
                <Location>Busstation</Location>
                <Weather>
                    <TimeOfDay>Day</TimeOfDay>
                    <Clouds>None</Clouds>
                    <Windy>None</Windy>
                    <Rain>No</Rain>
                    <Snow>No</Snow>
                    <Fog>No</Fog>
```

```
                </Weather>
                <Distraction>No</Distraction>
                <DetectionAreas>2</DetectionAreas>
                <DetectArea>
                    <Point>1,55</Point>
                    <Point>1280,55</Point>
                    <Point>1280,720</Point>
                    <Point>1,720</Point>
                </DetectArea>
                <Intrusion>
                    <Point>1274,145</Point>
                    <Point>631,68</Point>
                    <Point>328,132</Point>
                    <Point>1274,507</Point>
                </Intrusion>
            </Header>
            <Alarms>
                <Alarm>
                    <StartTime>00:03:18</StartTime>
                    <AlarmDescription>Intrusion</AlarmDescription>
                    <AlarmDuration>00:00:12</AlarmDuration>
                </Alarm>
            </Alarms>
        </Clip>
    </Library>
</KisaLibraryIndex>
```

- The root element is <KisaLibraryIndex>.
- Within this root element, there is a child element <Library> which contains information related to a specific scenario, dataset, and library version.
- Under the <Library> element, there is a <Clip> element representing a video clip in the library.
- Within the <Clip> element, there is a <Header> section that contains metadata related to the video clip, including the filename, stage, duration, alarm events, location, weather conditions, distractions, detection areas, and specific detection areas.
- The <Header> section also includes a <Loitering> or <Intrusion> section, which defines a specific loitering and intrusion area using poly points.
- Under the <Clip> element, there is an <Alarms> section that provides information about alarms triggered during the video clip.
- Within the <Alarms> section, there is an <Alarm> element that specifies the start time, alarm description, and alarm duration.

3. Assignment: Abnormal Behavior Recognition Model Development

Task Overview: The objective of this assignment is to develop a deep learning model that can detect and recognize abnormal behaviors in surveillance videos. Each student is responsible for creating and validating their own abnormal behavior recognition model using the provided video stream dataset.

Data Allocation: Students will receive a 90% subset of the KISA dataset, which includes video files, corresponding XML labels, and maps with polygon points defining detection areas. The remaining 10% of the data will be reserved for evaluating the model's performance upon project completion. For example, in the case of the "Loitering" class, comprising 30 total video files, each student will be allocated 27 video files for model development, training, and validation, while the remaining 3 video files will be retained for evaluation purposes.

Model Development: Each student is required to develop a single deep learning model capable of detecting and classifying abnormalities in video streams into six predefined categories. The model must trigger an alarm and provide the corresponding label when it identifies any abnormal event in the videos. Be sure to follow the seven-step approach for building a deep learning model, as discussed in class.

Training and Validation: Students are expected to train and validate their recognition models using the provided data.

Submission: Students are required to submit their project outcomes, which should include:

- A comprehensive project report
- The source code
- Result videos

Performance Evaluation: After successful model development and validation by all students, Cuong will assess each student's model performance using a set of test videos, representing 10% of the video files across the six abnormal behavior categories. The evaluation will be based on the following evaluation metrics for each abnormal behavior category, and a grade will be assigned according to the assessment results. (KISA requires a minimum accuracy of over 90% in all six abnormal behavior categories to issue a certificate.)

Abnormal Behavior	Accuracy	Precision	Recall	F1-Score
Loitering				
Intrusion				
Abandonment				
Falldown				
Violence				
Fire				

3.1 Loitering

The loitering scenario has two detection areas as shown on the below example map file:

```
<?xml version='1.0' encoding='utf-8'?>
```

```
<DA>
    <DetectionAreas>2</DetectionAreas>
    <DetectArea>
        <Point>11,4</Point>
        <Point>771,45</Point>
        <Point>1127,89</Point>
        <Point>1271,100</Point>
        <Point>1273,707</Point>
        <Point>7,714</Point>
    </DetectArea>
    <Loitering>
        <Point>655,317</Point>
        <Point>769,264</Point>
        <Point>842,183</Point>
        <Point>960,75</Point>
        <Point>1114,149</Point>
        <Point>1269,211</Point>
        <Point>1269,601</Point>
        <Point>1149,601</Point>
        <Point>666,477</Point>
    </Loitering>
</DA>
```

The <DetectArea> element indicates the area where an abnormal event may occur using polygon points. The <Loitering> element presents the area where the loitering event happens. If human objects enter this loitering area for a prolonged duration, they are identified as exhibiting loitering behavior. As part of their training model, students are required to establish a duration threshold for detecting loitering behavior based on the XML labels file, which includes an <Alarm> element specifying the start time, alarm description, and alarm duration.

```
<Alarms>
   <Alarm>
      <StartTime>00:03:18</StartTime>
      <AlarmDescription>Loitering</AlarmDescription>
      <AlarmDuration>00:00:12</AlarmDuration>
   </Alarm>
</Alarms>
```

For reference, Fig. 8 illustrates the map area for the loitering scenario, with pink polygon is the <DetectArea> and blue polygon is the <Loitering>.

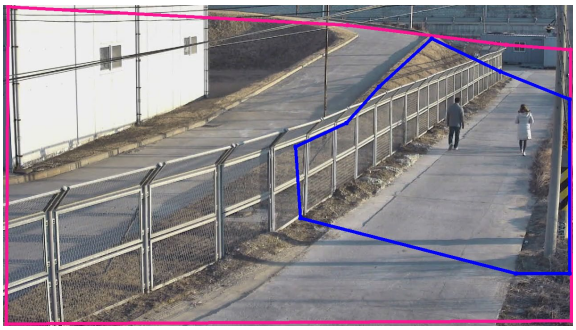

Figure 8. Polygon of detection area (pink color) and loitering area (blue color).

3.2 Intrusion

The example map file for intrusion scenario as shown below:

```
<?xml version="1.0" encoding="utf-8"?>
<DA>
  <DetectionAreas>2</DetectionAreas>
  <DetectArea>
    <Point>0,0</Point>
    <Point>1280,0</Point>
    <Point>1280,720</Point>
    <Point>0,720</Point>
  </DetectArea>
  <Intrusion>
    <Point>430,160</Point>
    <Point>673,156</Point>
    <Point>1087,354</Point>
    <Point>1065,449</Point>
    <Point>408,471</Point>
    <Point>425,160</Point>
  </Intrusion>
</DA>
```

The <DetectArea> element indicates the area where an abnormal event may occur using polygon points. The <Intrusion> element presents the area where the intrusion event happens, if human objects enter this intrusion area, they are identified as exhibiting intrusion behavior. Students are tasked with providing the <Alarm> information for this scenario specifying the start time, description, and duration.

```
<Alarms>
   <Alarm>
      <StartTime>00:01:50</StartTime>
      <AlarmDescription>Intrusion</AlarmDescription>
      <AlarmDuration>00:00:05</AlarmDuration>
   </Alarm>
</Alarms>
```

Fig 9 illustrates the map area for intrusion scenario.

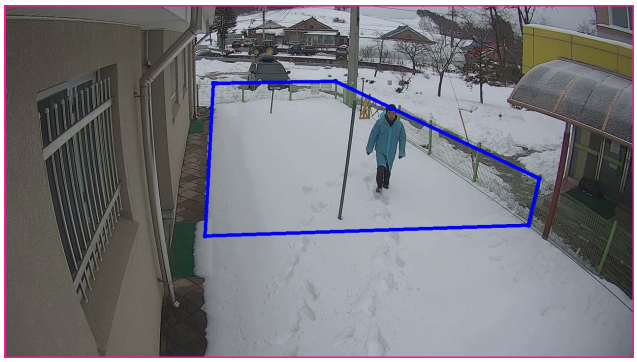

Figure 9. Polygon of detection area (pink color) and intrusion area (blue color).

3.3 Abandonment, Falldown, Violence, and Fire Detection

These scenarios utilize the same map file structure, incorporating only the <DetectArea> element to identify the region where an abnormal event might take place. Students are responsible for providing the <Alarm> information for these scenarios, which includes the start time, description, and duration. The example file map is shown below:

```
<?xml version='1.0' encoding='utf-8'?>
<DA>
    <DetectionAreas>1</DetectionAreas>
    <DetectArea>
        <Point>345,1</Point>
        <Point>1278,1</Point>
        <Point>1278,718</Point>
        <Point>345,718</Point>
    </DetectArea>
</DA>
```

For reference, Fig. 10 illustrates the map area for the violence scenario, with pink polygon is the <DetectArea>

Figure 10. Polygon of detection area (pink color) for violence scenario.