Multi-Modal Context Aware Monitoring System for Smart Classrooms

Demonstration using Reference Solutions

Sangeeta Ghangam
sangeeta.ghangam@intel.com
Intel Corporation
Chandler, Arizona, USA

Sanjay V Addicam
addicam.v.sanjay@intel.com
Intel Corporation
Chandler, Arizona, USA

Gabriel L Silva
gabriel.l.silva@intel.com
Intel Corporation
Chandler, Arizona, USA

Wendy Chin
wendy.chin@intel.com
Intel Technology Sdn.Bhd.
Bayan Lepas, Pulau Pinang, Malaysia

Michelle Lai
mee.sim.lai@intel.com
Intel Technology Sdn.Bhd.
Bayan Lepas, Pulau Pinang, Malaysia

ABSTRACT
Given the proliferation of IoT sensors, every location is now monitored by several different sensory endpoints. IoT sensors enable context awareness for a given system or environment by allowing automated reactions to sensor readings or events. Combining audio, video and environmental sensors provides an effective multi-modal monitoring approach resulting in less uncertainty with respect to the event, better response times and improved safety quotient. In this project showcase abstract we explain the rationale and details of the end-to-end demonstration, to prove the efficacy of a multi-modal system in enhancing safety for classrooms.

CCS CONCEPTS
• Hardware → Emerging architectures; Sensor devices and platforms; Sensors and actuators; Beamforming; Sound-based input / output; • Theory of computation → Machine learning theory.

KEYWORDS
Multi-Modal, Environmental sensors, Audio, Video Sensors, Smart Campus, Safety, Context Awareness

1 INTRODUCTION
According to the University of Washington [1], 3.85 deaths in the United States per 100,000 people are due to gun deaths, the number does not include the huge amount of the unreported firearm crimes. The incidents, especially in the crowded public places for instance schools and event venues, are devastating. Schools have become one of the targets since it is difficult to keep the campuses closed completely. Additionally, most of the campuses are lacking of the gunshot and criminal detection technology that can trigger the emergency responders with real-time data during the active shooter event to mitigate casualties in the disaster. Response times after an incident are normally predicated on how fast the law enforcement agencies are notified. Immediate reaction by the people impacted by the incident varies based on the severity. When panicked most people will try to move towards safety. If there is a monitoring system made of different sensors at the location of the incident it provides a mechanism to automatically alert both the response teams and people in the vicinity as to the actions they can take towards safety. Measurable reduction in emergency response time has been noticed when sensors in a crash recorder automatically send notification about the road accident to response teams [2]. While the monitoring techniques demonstrated as part of this showcase can be applied to a varied number of scenarios, we picked smart classrooms, as safety is one of the key value proposition for IoT in classrooms [3]. In the following sections we describe the different modules that constitute the monitoring system used in the demonstration followed by setup considerations and results using a simulated incident in a classroom environment.

2 PROJECT PARTNERS
The sensor hub includes sensors selected per United States Environmental Protection Agency (EPA) guidelines and from Bosch and SGX. The sensor application cloud software has been developed using technology from Amazon. The video analytics software uses emotion recognition algorithm based on AffectNet which was licensed from University of Denver. The media and display pipeline has been built using technology from Microsoft and custom Media Foundation transforms.
Table 1: List of sensors included with the sensor hub

<table>
<thead>
<tr>
<th>Sensor</th>
<th>Technology</th>
<th>Threshold</th>
</tr>
</thead>
<tbody>
<tr>
<td>BME280-Temperature</td>
<td>MetalOxide</td>
<td>22°C</td>
</tr>
<tr>
<td>MICS6814-NO2</td>
<td>Electrochemical</td>
<td>High (151-200ppm)</td>
</tr>
<tr>
<td>ParticulatePMS5003</td>
<td>Optical</td>
<td>High (70-499)</td>
</tr>
</tbody>
</table>

3 PROJECT DESCRIPTION

The demonstration scenario includes a video and audio recording from a classroom session in progress. The recording consists of regular teacher student interaction for the first 8-10 minutes after which we simulate an air quality event followed by multiple alarms towards the end of 15 minutes. This recording will be used to generate real time inference in our analytics application, sensor and audio application from client device to cloud.

Following is a system block diagram and description of each component used in the demonstration

3.1 Sensor Input

- Cameras – For purposes of this demonstration we have collected video/audio samples which will be passed through the software on host system for real time inference. For generating the video samples we have used 2 cameras setup to function as IP cameras used for surveillance/monitoring. The cameras are positioned to capture the entire room, students, teacher and exits
- Sensor hub with environmental and audio sensors – A sensor hub reference design has been used to connect the list of sensors as captured in Table 1. Threshold value is used to generate events in the software. Choice of the sensors and the thresholds are based on Indoor Air Quality (IAQ) guidelines for schools as published by the EPA [4]

In addition to the environmental sensors, the sensor hub also includes two linear mic arrays for capturing audio events such as alarms, gunshot, glass-break and scream. The mic arrays are designed for 360 degree audio coverage for indoor and outdoor use. It provides Time Division Multiplexed (TDM) data to the sensor hub which includes a codec to covert data into USB 2.0. Below is a snapshot of the sensor hub and sensors used for the demo

3.2 Host Computer

- The audio sensing will occur in real time while the environmental sensor values will be simulated. The Sensor and Audio application manages the setup of the hub per user configuration and compares the incoming data with thresholds for generating events. Events and regularly monitored sensor data are sent to the MQTT broker in the AWS cloud instance.
- The analytics application is an end-to-end media and analytics software reference utilizing Intel MediaSDK, Intel OpenVINO Toolkit, and Microsoft Media Foundation to decode camera feeds and analyze the video frames. The video frames are analyzed against several convolutional neural networks (with some processing concurrently) for
  - emotion recognition – possible use case is to detect any panic emotions in the classroom
  - facial recognition – possible use case is to detect black listed intruders in the classroom or campus
  - student behavior recognition (e.g. raising hands, sitting, and standing) – possible use case is to detect panic stricken behavior
- The inference can be processed via CPU, GPU, or an accelerator such as Movidius via user configuration. The inference data is published via MQTT over websockets to a Dashboard. Below is an example of real time inference using the analytics application
3.3 Cloud Data Management/Dashboard (AWS EC2)

All sensor data and events are uploaded to the cloud to store and process in ‘near-real time’ (latency of <5 mins). Separate EC2 instances have been used to manage data storage in MongoDB and provide metadata search capability. Bulk of the eventing from sensor data is managed by scripts running in elastic map reduce instance which provides latency assurances for ‘near-real time’ processing [5]. The scripts contain classification heuristics for managing events using the given time period and number of events. The events resulting in a severity include – Audio Events (Alarm, Gunshot, Glass Break, Scream) and Sensor Events (Alert)

Severity Classification

- Low: Upto 5 messages in less than 10 minutes
- Medium: Upto 15 messages in 10 or more minutes
- High: Upto 30 messages in 15 or more minutes
- If number of messages and/or the timeline is greater than what is captured above, the status is elevated to ‘Critical’

Event data is stored in database and available for displaying in the dashboard via REST API. In this demo we will show an example dashboard and sensor event stream, in the cloud, via a browser next to the video analytics processing results. All these different pieces of information will be placed in its own ‘zones’ for display on the screen. Below is the result in the cloud which shows the sensor and audio event.

4 RESULTS

- The video analytics accurately detects student behavior, emotions of the participants in the classroom session. With the NO2 exceeding threshold the student reactions changes which is identified by the system.
- The NO2 sensor and Audio sensor accurately detects events as they occur in the classroom session
- Events are processed in the cloud to generate alerts. Participants in the classroom access this data and are able to move in the right direction for safety. Absence of directional data with audio event or a missed sensor event would have resulted in an unsafe result for the classroom participants.

5 CONCLUSION

Multi-modal approach of combing sensors, audio and video results in dramatic reduction in uncertainty when faced with an incident(s). This helps in enabling a safer environment indoors and outdoors.

6 LINK TO DEMO VIDEO

For ease in viewing, videos have been collected for each of the above results

- Video Analytics [6]
- Audio/Sensor Event Detection [7]
- Cloud Dashboard [8]

REFERENCES

[2] Howard R. Champion, Research Professor of Surgery, University of Maryland; Principal Investigator, JS Augenstein, Professor of Surgery, University of Miami; B Cushing, Surgeon, Maine Medical Center; KH Digges, Professor of Engineering, George Washington University; R Hunt, Emergency Physician, State University of NY, Syracuse; R Larkin, Emergency Physician, American College of Emergency Physicians; AC Mulliaris, Mathematician, George Washington University, WJ Sacco, Mathematician, ThinkSharp, Inc.; JH Siegel, Professor of Surgery, University of Medicine and Dentistry of New Jersey. Reducing Highway Deaths and Disabilities with Automatic Wireless Transmission of Serious Injury Probability Ratings from Crash Recorders to Emergency Medical Services Providers NHTSA.