Large Area IOT-Acoustic Effective Surveillance System

Ezron Ember[1], Siti Zaleha Abdul Hamid[1*], Herlina Abdul Rahim[1]

[1] Faculty of Electrical Engineering, Universiti Teknologi Malaysia, 81310, Skudai, Johor, Malaysia

Abstract

This paper presents a novel large scale SM systems that identifies and restricts acoustic variations from the norm to address the difficulties of IoT acoustic surveillance systems in students residential area in a disseminated communitarian way. The project discusses the potential of this system in communitarian control and SM, quality, automated self-sorted out conduct, automatic abnormal event detection, environment security and indoor restriction. It expects to limit vitality utilization and correspondence overhead while keeping up high following quality and versatility. The proposed approach disseminates vitality four times slower than that of the famous unified approach while giving 6% higher following quality.

Keywords: Sensor Management, Internet of Thing, Support Vector Machines (SVM), Linear Discriminate Analysis (LDA)

1. Introduction

The pervasive Internet of Things (IoT) is a rising worldwide framework where inserted savvy objects are combined with unavoidable correspondence to empower nitty gritty portrayal of the genuine physical applications. Perceiving that cutting edges society faces new kinds of dangers; safety efforts require a worldview move from incorporated and investigative answers for circulated what's more, precaution ones. In that capacity, the many-sided quality and spread of security dangers have made it clear that unavoidable IoT observation is the best arrangement. Such IoT frameworks contain a heterogeneous gathering of sensor hubs spread over an extensive zone where every sensor has a fractional perspective of the earth, and the system all in all screens the whole territory under reconnaissance. In such extensive scale asset obliged systems, the utilization of visual sensors is jumped by various difficulties, including; a higher handling multifaceted nature, affectability to brightening conditions, constraints of the field of view, protection concerns, data transfer capacity constraints, limited power supply, and high level of centralization. Nonetheless, sound signs in the dynamic scene give imperative data. For instance in student's residential area, loud student's screams, or explosion or detecting abnormal environment like theft cases that can be caught by receivers.

**corresponding author: szaleha@utm.my*

The acoustic data from dynamic scene can give rich data about the condition of the earth which bring down handling, stockpiling, and transmission costs. Although acoustic observation began getting more consideration from the research group around the world lately, most state of the craftsmanship concentrate just on anomalous occasion acknowledgment utilizing brought together frameworks (Yusuf, S.A., Brown, D.J., & Mackinnon, A., 2015, Sharan, R.V., & Moir, T.J., 2014, Moon, S. et al., 2014, Łopatka, K., Kotus, J., & Czyżewski, A., 2014, Kotus, J., Lopatka, K., & Czyzewski, A., 2014) in this manner, disregarding the requirement for disseminated activity to empower large scale IoT acoustic surveillance systems that works under vitality and asset imperatives.

Sensor Management (SM) is fundamental in IoT acoustic surveillance frameworks as there is a need to oversee and organize the detecting assets to gather the most total and precise information from a dynamic scene under asset limitations. Accomplishing effective appropriated SM is trying because of the spatially circulated nature and the constrained assets of sensor systems, exceptionally in indoor conditions where circumstances are portrayed by a high thickness of targets, stochastic situations, and dynamic dangers. To-date appropriated substantial scale SM is as yet a test (Hilal, A.R. & Basir, O., 2013). Huge research endeavours concentrated on brought together methodologies (Atia, G.K., Veeravalli, V.V., & Fuemmeler, J.A., 2011, Kolba, M.P., Scott, W.R., & Collins, L.M., 2011, Kolba, M.P., & Collins, L.M., 2009, Li, Y. et al., 2009). Despite the fact that the examination group proposed various decentralized SM, constrained research exertion has been coordinated in applying of such decentralized SM approaches in unavoidable IoT acoustic surveillance applications.

This paper presents a novel large scale SM systems that identifies and restricts acoustic variations from the norm to address the difficulties of IoT acoustic surveillance systems in students residential area in a disseminated communitarian way. The project discusses the potential of this system in communitarian control and SM, quality, automated self-sorted out conduct, automatic abnormal event detection, environment security and indoor restriction. It expects to limit vitality utilization and correspondence overhead while keeping up high following quality and versatility.

2.1 Background of Study

This section will discussed about IoT acoustic surveillance system and sensor management system.

2.1.1 IoT Acoustic Surveillance System

Visual sensors are the most received methodology in brilliant unavoidable observation frameworks. However, utilization of visual observation is highly costing and increases the worries about security issues. These problems can be solved using sound-based surveillance frameworks. Besides, it also helps give imperative data that is troublesome or relatively incomprehensible for visual signs to identify. For instance, an uproarious student's shout or explosion can be effortlessly caught by receivers. As of late, the examination on human acoustic elucidation has pulled in a lot of consideration.

Different acoustic research concentrated on distinguishing natural sound occasions, for example, human shouts, shots, breaking glass, entryway pummels, blasts, puppy barks, telephone rings, youngsters' voice, and machine sounds. Different investigations performed in building sound–based for precise recognition frameworks. In that capacity, investigating the condition of human empowers the detections of student's emotions connected to condition of risk, e.g., fear, anger, panic, and stress. These exploration studies can be connected to infer circumstance mindful learning of unusual or undermining circumstances. Source confinement of sound in nearness of ambient noise and indiscreet sounds is a testing errand that is required of numerous IoT acoustic applications. Sound source confinement techniques can be created in view of a variety of sound sensors (amplifiers) that scattered in condition in dispersed or reduced directional ways (Rex, J., 2011).

A basic nearness technique in light based of signal quality level can be used to find sound source. In directional exhibit, time-of-entry and angle of arrival can be removed from the received pattern. In the two plans, position of sensors is known to centre calculation and framework execution relies upon exactness of microphones positions and also, their affectability level.

2.2 Sensor Management (SM)

The term "sensor management," is refers to control of the degrees of freedom in an agile sensor system to satisfy operational constraints and achieve operational objectives. A key test in the outline of unavoidable surveillance systems is dealing with the expansive number of tangible assets and making choices that determines the performance of the entire framework.

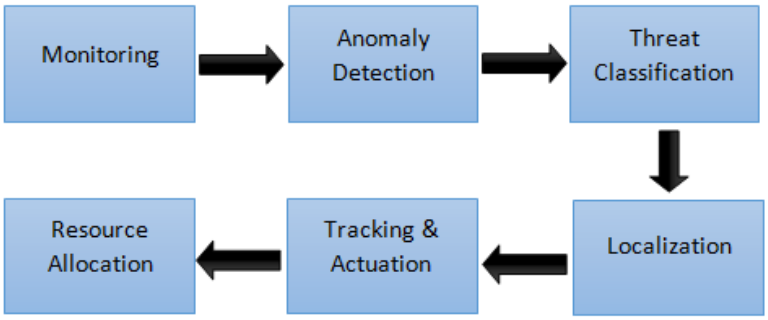

Figure 1. The surveillance Systems Components

The research group have committed critical endeavours to address the SM issue. The vast majority of these endeavours have focused on concentrated methodologies (Atia, G.K., Veeravalli, V.V., & Fuemmeler, J.A., 2011, Kolba, M.P., Scott, W.R., & Collins, L.M., 2011, Kolba, M.P., & Collins, L.M., 2009, Li, Y. et al., 2009). Despite the fact that the unified approach gives a solitary control point and organized order stream which is appropriate for mission basic applications, it restrains the adaptability of the framework and channels the framework assets. Data theoretic methodologies were additionally utilized as a part of the SM issue to assess the future data pick up of sensor estimation. In any case, the data – theoretic methodologies can give over – accentuation on the nature of data to the detriment of different system parameters, e.g., bandwidth, network life – time, and energy.

2.2 Proposed System

The activity of sensor hubs assumes an essential part in the adequacy of the general framework execution. Sensor hubs have four methods of task; dynamic/detecting, sit still/tuning in, transmitting/accepting, and dozing. A sensor hub must be in one of these four modes at any given time. On the off chance that there is no confinement on the vitality save, the sensor system would gather each conceivable estimation to augment the circumstance mindfulness. In any case, the constrained vitality spending manages that the sensor should just gather the estimations that contribute in accomplishing higher situation-awareness. Accordingly, sensor hubs are required to work under the most minimal vitality utilization expected to accomplish the required execution. The sensor vitality show embraced in this paper is in view of the work in (Hilal, A.R. et al., 2018) in which every sensor segment devours diverse vitality in different task states and in addition amid the state progress.

This study is the outline and improvement of an independent SM for acoustic IoT surveillance frameworks. The proposed system screens the Volume of Interest (VoI) in vitality productive way through a conveyed system of independent sensors that utilizations sound signals to recognize oddities, groups these inconsistencies by utilizing prepared models, at that point confines them, and distributes the required assets to track such inconsistencies through the VoI.

Figure 2 shows the main operation of the sensor to enable large – scale IoT acoustic surveillance system. The proposed systems characterizes an appropriated variation from the norm recognition, cooperative basic leadership, and an indoor confinement approaches on-leading group of an arrangement of keen sensors hubs and an arrangement of delegate hubs appropriated inside a VoI.

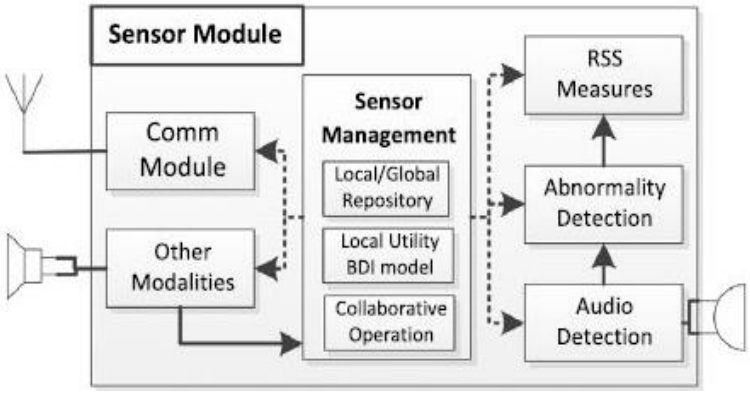

Figure 2. Sensor Software Modules High-Level Overview

These sensor hubs perform natural observing for potential undermining exercises, while the delegate hubs are more able sensors that help in the limited basic leadership process. When the proposed system is implemented, it is expected to produce output such as detect abnormal event in the residential area (students taking drug in dark sites, bully, noisy act disturbing other students, fights, gunshot, fire, the act of stealing and so on).

2.3.1 System Components

The hidden sensor system includes a number of stationary heterogeneous battery-worked sensors outfitted with sound and radio finder. Such sensors are dependable for gathering and investigating significant information to accomplish the framework targets in a vitality proficient way. In functional surveillance situations, the strange occasions must be reasoned progressively as the operational circumstance changes after some time. In like manner, the

systems sensor parts are intended to be skilled of self-ruling perceiving the basic zones/focuses inside the VoI. In this manner, these segments are intended to have a neighbourhood control what's more, administration unit that can initiate the neighbourhood irregular occasion recognition and assessments modules. These modules break down the information gathered from the VoI, and gauge its level of criticality. The level of criticality is last bolstered back to the administration module to settle on a choice to prompt the neighbouring representative assuming further examination of the circumstance is required.

2.3 Abnormal Event Recognition Module (AERM)

The proposed systems join audio data and centres around the acknowledgment of human shouts and human discourse under high worry to distinguish anomalous circumstance requiring the consideration of the security master. The AERM module comprises of three stages which are pre-processing, extraction, and classification.

In pre-processing stage, the sound foundation clamour is portrayed by the higher level of changeability contrasted with video foundation which is regularly static or gradually fluctuating. Due to inconsistencies in the recorded sound signal, the pre-processing step is performed on input motions before separating the highlights. Channel activities are performed on the signs to decrease clamour because of natural conditions and dispense with other unessential segments that are caught amid the chronicle of the sound signal. The info discourse signal is separated by a pre-accentuation channel with a coefficient of 0.97. This channel is connected to decrease the distinctions in control of the different parts in the signal to enhance the SNR. After the signal is separated, it is sectioned into measure up to covering outlines with a 25 ms length and a 10 ms move, as appeared in Figure 3. Each casing is then duplicated by a Hamming window previously separating highlights from the signal to maintain a strategic distance from issues because of signals discontinuities, and to extricate a smoother form and decrease the swell in the range.

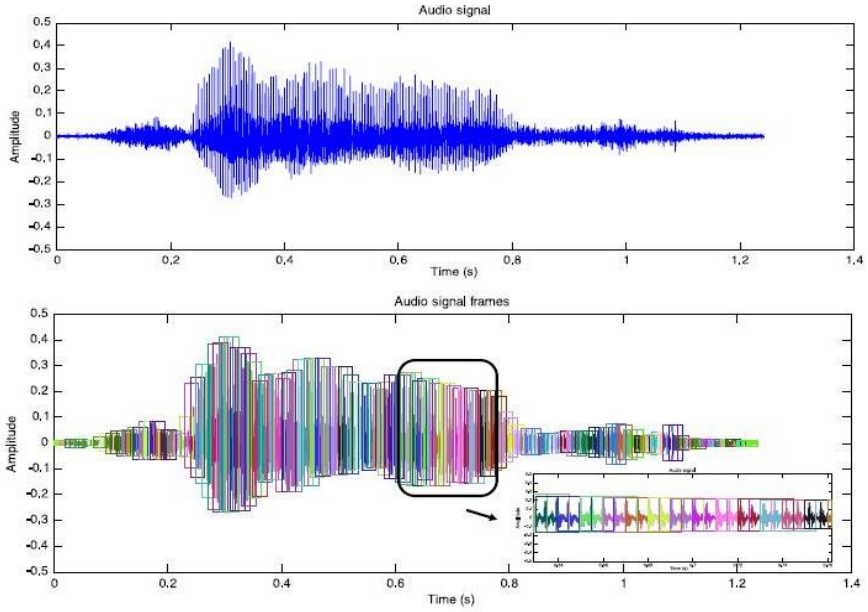

Figure 3. Time Space Sound Signals (best) and Sound Example Outlines (base)

Meanwhile in feature extraction stage, the decision of appropriate highlights that successfully describe the enthusiastic substance in discourse is considered a fundamental piece of the strange occasion acknowledgment module outline. In this work, two kinds of highlights are removed: prosodic highlights, for the most part brief time vitality (E) and crucial recurrence (F0); and unearthly highlights, Mel-Frequency Cepstral Coefficients (MFCCs). The prosodic highlights are demonstrated to give a dependable sign of the human passionate state (Tao, J., Kang, Y., & Li, A., 2006, Samantaray, A.K., et al., 2015). Likewise, the flag vitality contains valuable data that is useful in segregating typical discourse with low vitality from shout signals with higher vitality amplitudes.

The brief span vitality E is acquired by processing E = Σ (Xi) 2 were up until n and i=3, (3) where Xi is the sound flag plenty-fullness of edge I. Furthermore, the essential recurrence F0 is ascertained utilizing time-space approach. Initially, the autocorrelation work is figured for each casing of the sound discourse signals. At that point the crucial recurrence is evaluated by choosing the best top in the deferred interims of the autocorrelation work between 50 Hz and 500 Hz which compares to the typical contribute run discourse. Figure 4 demonstrates the prosodic highlights extricated from the example sound flag appeared in Figure 3. The MFCC highlights exhibit a depiction of the otherworldly state of the sound, broadly utilized as a part

of discourse acknowledgment. It is considered an appropriate element for discourse portrayal since it can adequately show the human discernment to discourse.

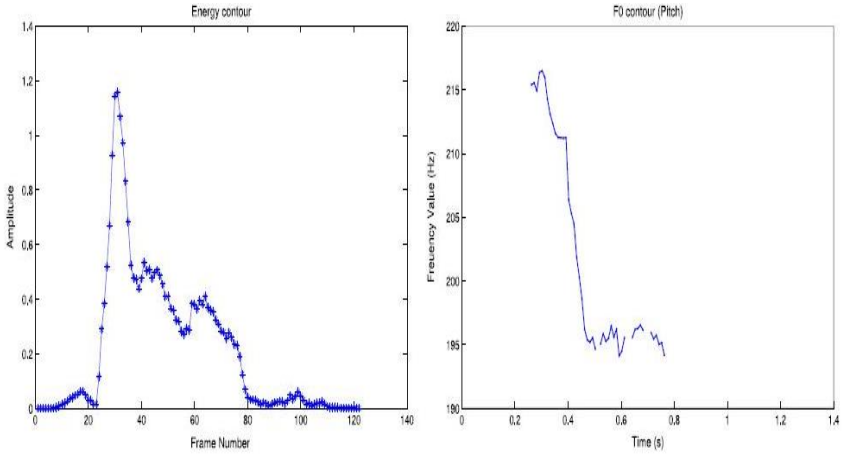

Figure 4. Brief time Energy (left) and Fundamental Recurrence (right) Extricated from Every Sound Edge.

Figure 5 delineates the MFCC highlight extraction for each edge of the sound flag. The MFCC highlights are acquired as takes after: in the first place, the plenty-fullness range of the sound edge is figured utilizing the transient Fast Fourier Transform (FFT), at that point, the recurrence groups are mapped onto the Mel-scale, which gives a superior estimation of the human sound-related framework reaction than the direct scale. The Mel-filter manage an account with 40 channels is connected to the forces of the range.

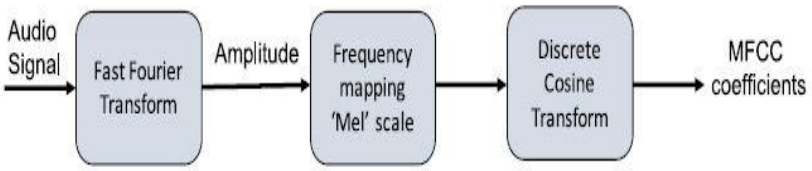

Figure 5. Layout of the MFCC highlight extraction for each casing of the sound flag (13 MFCC coefficients for every edge were chosen)

Then, in classification stage, Bolster Vector Machines (SVM) was broadly utilized as a part of discourse feeling acknowledgment. It was likewise utilized for a few sound occasion grouping applications. The focal points of SVM incorporate the worldwide optimality of the preparation calculations and the presence of information subordinate speculation limits. Thus, SVM classifier is picked as the fundamental classifier in this work. Keeping in mind the end goal to pick the best portion capacity of the SVM classifier for the current issue, the execution of the SVM classifier for various part works are at first explored (e.g., Linear, Polynomial, and Gaussian Radial Basis Function (RBF) pieces). The parameters of the polynomial and the RBF bits are tuned by cross approval approach on the preparation information to locate the best an incentive for every parameter. At that point the SVM classifier is prepared utilizing the upgraded parameters. What's more, the Linear Discriminant Analysis (LDA) classifier is chosen to explore the execution of the module without the impact of complex grouping approaches with the end goal of correlation in this consider.

Communications assumes a vital part in the activity of the modules and parts of proposed systems and it is viewed as a frame of incitation. At first, the sensors filter the VoI to gather sound information which thus is broke down to recognize any peculiarities in the related sound signs. In the event that a sensor distinguishes the nearness of an unusual occasion inside the VoI, the sensor alarms the bunch delegate of the irregular occasion by means of a sign message. At the point when a delegate hub gets a prompt message, the delegate gathers the important sensor estimations by pooling neighbouring sensors and agents, at that point assesses the risk level related to the unusual occasion. At that point, the delegate enacts the restriction module to find the situation of the strange occasion to advise the law authorization faculty. For high-threat level targets, assign hub actuates agreeable following, in which sensors work together to keep up close checking of the danger area. Agreeable following is utilitarian property in the type of prompting or handoff used to build the nature of observation, as appeared in Figure 6.

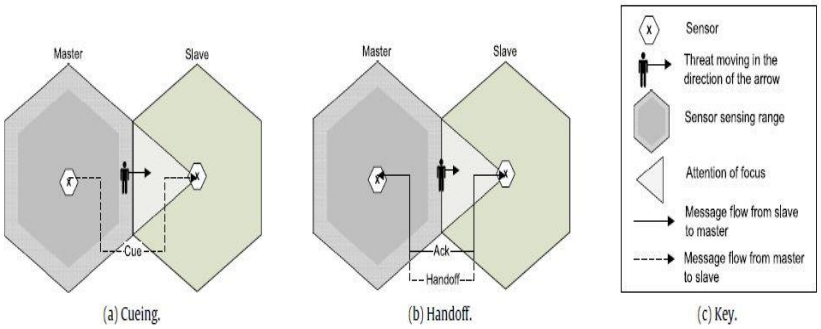

Figure 6. Sensor Participation Utilizing Prompting and Handoff of Targets

2.5 Events Localization modules

Indoor limitation is testing, since radio engendering is more disorganized than in outside settings because of checks. IntelliSurv tends to the occasion restriction issue by utilizing conflicting sign estimations, in the frame of Received-Signal-Strength (RSS) or time-of-entry from grapple sensors with known areas, to evaluate the objective area. The GPS recipient is the most well-known, exceptionally exact restriction sensor for outside. In thick urban regions or inside, structures can cover the got signals and essentially counteract exact restriction. It is important to utilize distinctive information modalities, for instance, sound, radio, and video, to unravel restriction issues, particularly where the odometry is troublesome. Despite the fact that, the signal quality is an available and moderate estimation for assessing the hub to-hub remove, it is uproarious and off base due to blurring channels delays. Most range-based restriction strategies to-date appraises the separation, as per the experimental channel models to gather the area. Along these lines, confinement strategies in light of time-delay-of-entry estimations, which are quick and compelling, are more down to earth. Additionally, the low computational unpredictability of these strategies suits continuous usage inside a sensor organizes.

2.6 System Decision Making

The decision-making problem is testing due to the inalienable trade-off between the estimation of data in the sensor estimations also, the vitality cost, joined with the combinatorial multifaceted nature of the choice space. The Belief–Desire–Intention (BDI) demonstrates is a capable specialist worldview that can give complex thinking, objective

tirelessness, and simple to investigate thinking system. The BDI display is a purposeful model which is enlivened by people mental investigations of psyche and human conduct. The term, purposeful alludes to frameworks whose conduct is ascribed to framework mental states of mind, for example, convictions, inclinations, what's more, goals that assume diverse parts in deciding a specialist's conduct. One of the BDI show alluring highlights is the capacity of operators to ceaselessly reason about convictions, objectives, and goals, furthermore, act in like manner. This model speaks to current vulnerabilities because of impediments in recognition, and future vulnerabilities due to dynamism. BDI specialists can adjust the time spent on thinking about plans and executing them. Be that as it may, the energy of this model may come at the cost of expanded computational overhead, if not intended for effectiveness. In this work, sensor and delegate modules in the systems are intended to show high self-rule by having keen basic leadership abilities that depend on the BDI show. The circulated idea of the proposed framework takes into consideration productive circulation of the workload and for high framework adaptability. Due to the exceptionally circulated nature of IoT observation applications, such plan considers low computational overhead per hub while by and large having high subjective thinking capacities. Utilizing the modular rationales definition, the fundamental framework objective is defined as the association runs in the accompanying. It is accepted that there is a limited arrangement of sensor hubs that cover the VoI consistently. These sensors need to accomplish various clashing targets utilizing various expectation designs.

2.7 Related Works

The received situation is the observation of the Waterloo International Airplane terminal. The air terminal corridors are for all intents and purposes partitioned into work lattice cells utilizing stationary sensors. Every sensor has a detecting scope of 3 × 3 matrix cells. The sensors are stationary having heterogeneous modalities. Every sensor is furnished with a battery of 100 power units. At first, the VoI is completely checked by sensors. The travellers; i.e., targets, enter and leave the airplane terminal haphazardly. Likewise, imprudent blasts of travellers' entries and flights are arbitrarily created to recreate true plane entries or flights. Moreover, the travellers leave the earth at irregular circumstances. The proposed framework is actualized by utilizing the Jadex stage, a java-based thinking motor for sane operators. The stage gives a multi-specialist advancement condition in view of the BDI thinking model and the joint aim hypothesis. In addition, the Jadex thinking motor

encourages a correspondence foundation established on the Agent Communication Dialect (ACL). The proposed framework, IntelliSurv, screens two kinds of dangers: episodes and human dangers. The human dangers are spoken to by insightful portable specialists with sets of convictions, wants, furthermore, expectations. The quantity of dangers fluctuates in the vicinity of one and 20. The dangers don't withdraw from the scene for the span of the reproduction and move around in the airplane terminal. The episodes allude to the anomalous occasions that force or suggest threat to the travellers in the airport, for example, a fire, a loud scream, or a gunshot. These incidents occur within the VoI for a short interval of time or until the law enforcement personnel intervene. The simulation is carried out until all the sensors run out of energy, and the surveillance system fails.

2.7.1 Effect of Signal to Noise Ratio

The SNR is a marker of the level of a coveted flag in correlation to the level of clamour inside the earth. The examinations assess the execution of the IntelliSurv SMF module over shifting SNRs and are led on a 9 × 9 condition, observed by nine sensors and one delegate. The SNR is accepted to shift in the vicinity of 0 and 100 dB. Additionally, it is expected that as it were one danger exists inside the VoI for the length of the recreation, in spite of the fact that occurrences haphazardly happen in a uniform dispersion. Figure 7 exhibits the general system lifetime for both SMF modules, i.e., the lifetime of the system until the point when it totally fizzles what not hubs bite the dust. It is clear that as the level of the coveted flag increments over that of the commotion, the general system lifetime increments. Moreover, the outcomes show that the general system lifetime increments essentially for the IntelliSurv analyzed to that of the brought together framework. Likewise, the general system lifetime for IntelliSurv, in the most pessimistic scenario, is in excess of three times that of the brought together framework. It is important, that for higher SNR values, for instance, SNR=100, the system lifetime for IntelliSurv is relatively triple that where SNR=0. This is expected to the diminished false alert rate and increment estimations exactness, bringing about better on-board occasion characterization and lessened correspondence overhead.

The AERM focuses on breaking down human discourse signs to recognize strange occasions. A key issue for planning the module is the accessibility of open benchmark databases for observation applications. A subset of the database speaking to the real worry in discourse is picked. There are accounts from seven speakers in crazy ride what's without more fall

circumstances, encountering Medium Stress (MS), High Stress (HS), and Screaming (SC). Nonpartisan (N) discourse is incorporated too. The sound signs are disseminated over the four previously mentioned classes.

The picked subset of the database comprises of an aggregate of 3593 discourse tests. Table 1 demonstrates the appropriation of the sound tests among the four classes and the aggregate length of each class. The normal term of the sound flags in the information is one second. The information is ordered as anomalous and ordinary occasions. The voice signals that demonstrates unusual occasions are spoken to in the arrangement of HS and SC discourse tests, and voice flags that demonstrate ordinary occasions are spoken to in the arrangement of MS and N discourse tests. The information is part arbitrarily into two disjoint sets: 90% for preparing what's more, parameter streamlining for the SVM classifier, and the staying 10% are utilized for testing the framework. The occasion grouping is isolated into two phases. In the first place, the discourse include vector is grouped into one of the four classes (HS, SC, MS, or N) by utilizing the multi-class arrangement approach. Also, an official conclusion is connected to additionally arrange the result of the primary stage into ordinary or strange occasions. The last execution of the framework is computed in light of the yield at the last stage with ground truth marks 1 and 0, to such an extent that 1 demonstrates irregular occasion and 0 shows an ordinary occasion.

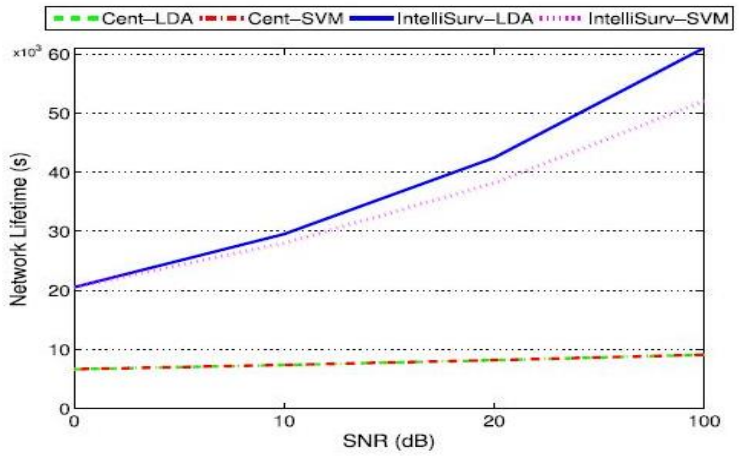

Figure 7. The System Lifetime Over The Expanding SNR

Table 1. Circulation of Size and Length of Sound Classes

Classes	Audio samples	Total audio duration (min)
High Stress (HS)	500	10
Scream (SC)	20	2
Medium Stress (MS)	150	4
Neutral (N)	700	15

3 Conclusion and Future Works

The unpredictability of rising dangers has empowered the advancement of inescapable IoT observation frameworks. A novel sensor administration for inescapable IoT observation, named IntelliSurv, is proposed in this paper to naturally distinguish and confines unusual acoustic occasions in a circulated collective way. The proposed irregularity identification module utilizes Support Vector Machines (SVM) and Linear Discriminate Analysis (LDA) classifiers to recognize and limit human shouts or high-push discourse signals. Unusual occasion limitation is conveyed in light of the trilateration calculation and got flag quality (RSS) data. The outcomes portray that IntelliSurv beats the broadly utilized brought together frameworks regarding vitality, correspondence overhead, organize lifetime, following quality, and identification rate. The proposed approach disseminates vitality four times slower than that of the famous unified approach while giving 6% higher following quality.

As future work, a genuine arrangement of the proposed framework will be explored. Additionally, a choice theoretic approach can be sent rather than the group theoretic one. The variation from the norm acknowledgment module ought to be stretched out to separate between various kinds of shouts, for instance, grown-ups' and kids' shouts and to address the issue of foundation clamour in complex true conditions. Highlight improvement will be performed by researching more highlights to empower the framework to perform well under exceptionally uproarious condition. Moreover, modified bits for the SVM classifier can be examined to upgrade the order precision of the unusual occasion acknowledgment module.

Acknowledgement

The authors would like to thank the supervisor for this project for the guide and University Teknologi Malaysia (UTM) for providing facilities for this study.

REFERENCERS

[1] Yusuf, S. A., Brown, D. J., and Mackinnon, A. (2015). Application of acoustic directional data for audio event recognition via HMM/CRF in perimeter surveillance systems. Robotics and Autonomous Systems, 72, 15-28.

[2] Sharan, R. V., and Moir, T. J. (2014, August). Comparison of multiclass SVM classification techniques in an audio surveillance application under mismatched conditions. In Digital Signal Processing (DSP), 2014 19th International Conference on (pp. 83-88). IEEE.

[3] Moon, S., Shon, S., Kim, W., and Han, D. K. (2014, August). Generalized cross-correlation based noise robust abnormal acoustic event localization utilizing non-negative matrix factorization. In Advanced Video and Signal Based Surveillance (AVSS), 2014 11th IEEE International Conference on (pp. 171-174). IEEE.

[4] Łopatka, K., Kotus, J., and Czyżewski, A. (2014, June). Evaluation of sound event detection, classification and localization in the presence of background noise for acoustic surveillance of hazardous situations. In International Conference on Multimedia Communications, Services and Security (pp. 96-110). Springer, Cham.

[5] Kotus, J., Lopatka, K., and Czyzewski, A. (2014). Detection and localization of selected acoustic events in acoustic field for smart surveillance applications. Multimedia Tools and Applications, 68(1), 5-21.

[6] A.R. Hilal, O. Basir, Pervasive surveillance system management, in: F. Flammini, R. Setola, G. Franceschetti (Eds.), Effective Surveillance For Homeland Security: Balancing Technology and Social Issues, CRC Press, Taylor & Francis Group, 2013, pp. 199–227.

[7] Atia, G. K., Veeravalli, V. V., and Fuemmeler, J. A. (2011). Sensor scheduling for energy-efficient target tracking in sensor networks. IEEE Transactions on Signal Processing, 59(10), 4923-4937.

[8] Kolba, M. P., Scott, W. R., and Collins, L. M. (2011). A framework for information-based sensor management for the detection of static targets. IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans, 41(1), 105-120.

[9] Kolba, M. P., and Collins, L. M. (2009, May). Sensor management using a new framework for observation modeling. In Detection and Sensing of Mines, Explosive Objects, and Obscured Targets XIV (Vol. 7303, p. 730326). International Society for Optics and Photonics.

[10] Li, Y., Krakow, L. W., Chong, E. K., and Groom, K. N. (2009). Approximate stochastic dynamic programming for sensor scheduling to track multiple targets. Digital Signal Processing, 19(6), 978-989.

[11] Rex, J. (2011). Audio event detection and localisation for directing video surveillance-a survey of potential techniques.

[12] Zajdel, W., Krijnders, J. D., Andringa, T., and Gavrila, D. M. (2007, September). CASSANDRA: audio-video sensor fusion for aggression detection. In Advanced Video and Signal Based Surveillance, 2007. AVSS 2007. IEEE Conference on (pp. 200-205). IEEE.

[13] Menegatti, E., Mumolo, E., Nolich, M., and Pagello, E. (2004, March). A surveillance system based on audio and video sensory agents cooperating with a mobile robot. In Proc. of 8th International Conference on Intelligent Autonomous Systems (IAS-8) (pp. 335-343).

[14] Foggia, P., Petkov, N., Saggese, A., Strisciuglio, N., and Vento, M. (2015). Reliable detection of audio events in highly noisy environments. Pattern Recognition Letters, 65, 22-28.

[15] Huang, W., Chiew, T. K., Li, H., Kok, T. S., and Biswas, J. (2010, June). Scream detection for home applications. In Industrial Electronics and Applications (ICIEA), 2010 the 5th IEEE Conference on (pp. 2115-2120). IEEE.

[16] Kim, K., and Ko, H. (2011, August). Hierarchical approach for abnormal acoustic event classification in an elevator. In Advanced Video and Signal-Based Surveillance (AVSS), 2011 8th IEEE International Conference on (pp. 89-94). IEEE.

[17] Benetos, E., Lafay, G., Lagrange, M., and Plumbley, M. D. (2016, March). Detection of overlapping acoustic events using a temporally-constrained probabilistic model. In Acoustics, Speech and Signal Processing (ICASSP), 2016 IEEE International Conference on (pp. 6450-6454). IEEE.

[18] Cai, R., Lu, L., and Cai, L. H. (2005, March). Unsupervised auditory scene categorization via key audio effects and information-theoretic co-clustering. In Acoustics, Speech, and Signal Processing, 2005. Proceedings.(ICASSP'05). IEEE International Conference on (Vol. 2, pp. ii-1073). IEEE.

[19] Hilal, A. R., Sayedelahl, A., Tabibiazar, A., Kamel, M. S., and Basir, O. A. (2018). A distributed sensor management for large-scale IoT indoor acoustic surveillance. Future Generation Computer Systems.
[20] Tao, J., Kang, Y., and Li, A. (2006). Prosody conversion from neutral speech to emotional speech. IEEE Transactions on Audio, Speech, and Language Processing, 14(4), 1145-1154.
[21] Samantaray, A. K., Mahapatra, K., Kabi, B., and Routray, A. (2015, July). A novel approach of speech emotion recognition with prosody, quality and derived features using SVM classifier for a class of North-Eastern Languages. In Recent Trends in Information Systems (ReTIS), 2015 IEEE 2nd International Conference on (pp. 372-377). IEEE.
[22] Hansen, J. H., and Bou-Ghazale, S. E. (1997). Getting started with SUSAS: A speech under simulated and actual stress database. In Fifth European Conference on Speech Communication and Technology.