Audio-visual low power system for endangered waterbirds monitoring

Aya Sakhri, Oussama Hadji, Chakir Bouarrouguen, Moufida Maimour, Nasreddine Kouadria, Abderrezak Benyahia, Eric Rondeau, Noureddine Doghmane, Saliha Harize

▶ **To cite this version:**

Aya Sakhri, Oussama Hadji, Chakir Bouarrouguen, Moufida Maimour, Nasreddine Kouadria, et al.. Audio-visual low power system for endangered waterbirds monitoring. 2nd IFAC Workshop on Integrated Assessment Modelling for Environmental Systems, IAMES 2022, Jun 2022, Tarbes, France. hal-03699799

HAL Id: hal-03699799
https://hal.univ-lorraine.fr/hal-03699799

Submitted on 20 Jun 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Audio-Visual Low Power System for Endangered Waterbirds Monitoring ⋆

Aya Sakhri [*,**] **Oussama Hadji** [***] **Chakir Bouarrouguen** [***]
Moufida Maimour [*] **Nasreddine Kouadria** [**]
Abderrezak Benyahia [***] **Eric Rondeau** [*]
Noureddine Doghmane [**] **Saliha Harize** [**]

[*] *Lorraine University, CNRS, UMR 7039, Campus Sciences, BP 270239, F-54000 Nancy, France (e-mail:aya.sakhri@univ-lorraine.fr).*
[**] *LASA laboratoy, Badji Mokhtar Annaba University, Algeria*
[***] *LASTIC laboratory, University of Batna 2, Batna, Algeria*

Abstract: Recently, more attention is being payed to the ecological environment due to the rapid decline of animal species, especially birds. Wireless Multimedia Sensor Networks (WMSN) can be leveraged to monitor, protect and assess changes in bird populations by periodically capturing and sending images to a collect station. In this paper, we propose a three-tier architecture following the *Cloud-Fog-Edge* model with an adequate placement of the surveillance system tasks. To address the sensor network limited resources, we propose to reduce the amount of visual data to be reported by sending compressed region of interest (ROI) of only target species images. Endangered birds are identified based on their calls before triggering the camera. However, audio recognition is likely to fail due to the ambient noise that can be encountered in a natural environment. As a result, we augmented our automatic bird song recognition system with an appropriate noise reduction technique. Thanks to the efficient cooperation of its three tiers, the proposed system, based on low-complexity audiovisual information processing, results in a better use of network resources while maximizing the amount of pertinent information.

Keywords: Birds Monitoring, WMSN, Audio Denoising, ROI Extraction, Energy Efficiency.

1. INTRODUCTION

Wetlands (marshes, lakes, wet meadows, ...) cover about six percent of the earth's surface. Characterized by a unique biodiversity, they represent one of the richest and most diverse ecosystems on our planet. Wetlands are home to a very wide variety of animal and plant species. They play a major role in biodiversity, in particular, they are privileged places for tens of thousands of waterbirds of different species to winter or make a temporary halt. Birds respond to possible changes in the environment and are good indicators of biodiversity. They also have a great role in environmental security. The decline of some avian species can cause serious problems in the food chain and affect environmental safety or even public health.

The State of the World's Birds Allinson (2018) revealed that we lost more than 161 species, which represents an extinction rate much higher than natural. Although some species can still be considered as breeding, others are already considered extinct in France. This gives great importance to wetlands as more than half of the wetlands have been destroyed over the last century due to urbanization, agricultural intensification and pollution. The survival of our environmental system from various dangers is a primary responsibility which can be achieved through endangered birds monitoring. Fauna and flora monitoring in general and birds in their natural habitat in particular, has several interests Archaux (2011). It mainly provide reliable indicators for regular assessment of the number of threatened birds, which serves environmental and even public health. Most of the time, environmental monitoring is done manually in a limited space and time. The presence of human observers in the field may disturb the birds behavior. Wireless sensor networks and more recently multimedia ones (WMSN Akyildiz et al. (2007)) appear as a good alternative to set up an automatic environmental monitoring system Dyo et al. (2012); Kizilkaya et al. (2022). For birds monitoring purposes, Stattner et al. (2011) proposed to use acoustic sensors for automatic counting of songbirds.

In this work, we propose a three-tier monitoring architecture that leverages both audio and visual sensors to identify and estimate the population of endangered birds species. To cope with WMSN limited resources especially in terms of bandwidth and energy, we suggest to reduce the amount of visual data to be processed as images captures, encoding and delivery are the main energy consumers in the whole system. A sensor network may evolve hundreds of nodes deployed in a harsh environment, which makes battery replacement a difficult task. Therefore, increasing the lifetime of the sensors to enable efficient operation of the network is a key issue. Multiple low-cost visual data compression algorithms have been proposed

⋆ This work was supported in part by the PHC TASSILI 21MDU323.

ZainEldin et al. (2015). The extraction of image features have been investigated by Civelek and Yazici (2016) and others limited the image capture using motion detection as Magno et al. (2013). However, all these techniques are not sufficient since camera nodes continue to capture and process images including irrelevant ones.

In our system, we propose to only capture relevant images of target species. This can be achieved through a low-complexity audio-based recognition prior to triggering the camera. There are very few works that combine audio-visual data processing in a sensor node to decrease the energy consumption as proposed by Koyuncu et al. (2018). However, they do not consider the overlapping of various environmental noises, a major constraint in our application since they degrade the sound quality which may result in false identification of species. To ensure an accurate and robust sound recognition, we additionally implement a noise reduction method prior to the audio-based recognition. The proposed overall system implements an efficient cooperative architecture with low-complexity audio-visual information processing with the ultimate objective of efficiently utilizing network resources while maximizing the amount of relevant information delivered to the end user.

The remainder of this paper is organized as follows. Section 2 describes our WMSN based bird surveillance architecture. The performance of our proposal with respect to other strategies is discussed in Section 3 before concluding.

2. WMSN-BASED BIRDS SURVEILLANCE SYSTEM

We propose the use of a surveillance infrastructure based on WMSN to monitor migratory birds in their natural habitat (wetlands). Acoustic sensors equipped with microphones and image sensors are to be deployed in conjunction with other types of sensors for presence detection, for example. The objective of this monitoring is to identify, recognize and count the number of a species of birds (considered threatened) on the basis of their vocalizations and photographs.

2.1 Network Model

We propose to adopt a three-tier architecture following the *Cloud-Fog-Edge* model as shown in Figure 1. The core of the network (edge) is composed of multimedia and scalar wireless sensors. In addition to capturing the data of interest, the sensors are capable of performing a number of processing tasks of moderate complexity with limited memory requirements. They transmit the raw or processed data to the base station or gateway (fog level) where moderate complexity processing can be performed. This level allows for greater computing resources while still being close to the data source and the end user. The cloud level allows for the most powerful analysis algorithms to be run, requiring more resources. This three-tiered processing also optimizes network bandwidth compared to single cloud-level processing. The cloud will also host a message broker based on protocols such as MQTT Standard (2014); Hunkeler et al. (2008) which, through its publish/subscribe paradigm, allows a monitoring agent (subscriber) to be informed of any detected critical situation.

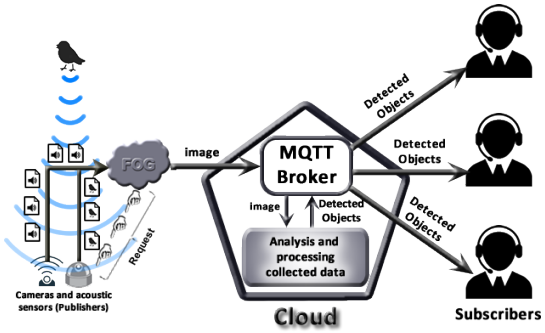


Fig. 1. Energy-efficient audio-visual surveillance strategy

With respect to our target application, instead of repeatedly sending images despite the absence of relevant information in the area being monitored, audio sensors are leveraged to allow for target detection before triggering the camera. The captured image is then compressed and transmitted to the cloud to confirm the presence of an event of interest as well as to perform further visual analysis such as counting and localization. The audio data recorded by the audio sensors are sent to the base station or gateway (fog level) to perform the audio based identification of the targeted birds. If a target species is identified, a request is sent back to a visual sensor node that covers the area of the audio origin to capture an image to be sent to the cloud.

2.2 Audio-based Target Identification

The proposed real-time audio recognition phase comprises three main steps. The first step consists in denoising the sensed audio signal to eliminate environment noises. To do so, we made use of a new geometric approach (GA) to spectral subtraction proposed by Lu and Loizou (2008) that improves the traditional spectral subtraction used by Weiss et al. (1975). Despite the fact that the latter is less complex, the former solves known shortcomings of the former like the musical noise and signal distortion.

Then, to only consider meaningful information of the audio signal, distinctive features are extracted using Mel Frequency Cepstral Coefficients ($MFCC$), key element of audio recognition. Their role is to identify the relevant components of the audio signal by eliminating irrelevant information. They are characterized by their computational efficiency and high performance under clean conditions Gupta et al. (2013).

Finally, the extracted features are compared to reference ones using Dynamic Time Warping (DTW) distance for classification purposes using a threshold based matching. DTW is one of the known low-complexity algorithms for measuring the similarity between two time sequences. It aligns two feature vector sequences by repeatedly warping the time axis to find the optimal match between the two vectors Müller (2007). If the processed sound is identified as a target, the camera will be activated and an image of the target is captured.

2.3 Image Processing

A captured image has to be sent to the end-user. In order to reduce the amount of data to be transmitted to the

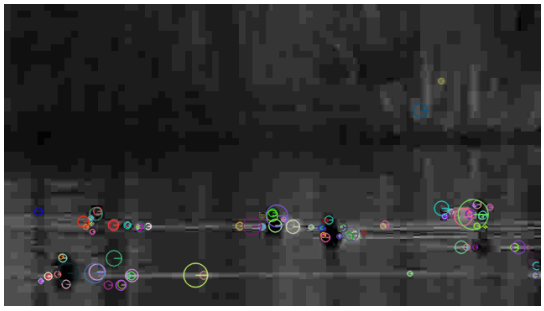


Fig. 2. Example of detected ROI

Sink, we suggest compressing and transmitting only image blocks that contain regions of interest (ROI). Afterwards, we apply a low complexity compression on the selected blocks as to limit the processing resources and save energy.

ROI Detection. Regions of interest are determined based on the features detected in the captured scene. A feature extraction method combines and evaluates variables to extract features, reducing the volume of data that must be processed, while maintaining a high level of accuracy and completeness of the original data. Image local features (also called hot spots and key points) can be defined as a single specific pattern from its immediately nearby pixels, which is usually associated with one or more properties of the image. These properties include edges, corners, regions, etc. These local features are converted into *numerical descriptors* representing a single and compact summary of these local characteristics. Local features (descriptive and invariant) provide a powerful tool that can be used in a wide range of computer vision applications. Many feature detection algorithms exist in the literature but there is no "one size fits all" solution because of the extreme variety of settings and possible scenes Salahat and Qasaimeh (2017).

To select a given algorithm, one has to consider its computation cost, its detected features and its accuracy. Considering the constraints of our edge nodes, the most vital aspect is the computational cost. This is why, we eliminate SIFT Lowe (2004) although it is one of the most used keypoints detection algorithms with a high accuracy. SIFT exhibits a heavy computational burden which does not suit the limited resources at the edge wireless sensors. With respect to our bird surveillance application, we aim to detect moving birds that pose at different angles translation and rotation are much more important than scale Tareen and Saleem (2018). KAZE algorithm Alcantarilla et al. (2012) emerges as a good candidate since it is more accurate with rotation and translation. In KAZE algorithm, a predefined value of the primary scene feature number is stored. When a new scene arrives, the algorithm extracts the new features from the last captured and compares these values with the stored ones if the new value (number of features) is greater than the previous one, which means that the region of interest is detected ; otherwise the scene is empty. An example of detected ROI is shown in Figure 2.

ROI Compression. Blocks identified as containing relevant visual information as a result of the previous paragraph are compressed using a low complex compression proposed in Maimour (2018) where a fast pruned DCT is used with a triangular pattern in which only coefficients

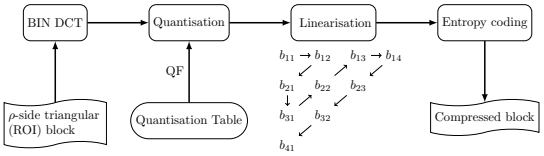


Fig. 3. ROI block encoding sequence ($\rho = 4$).

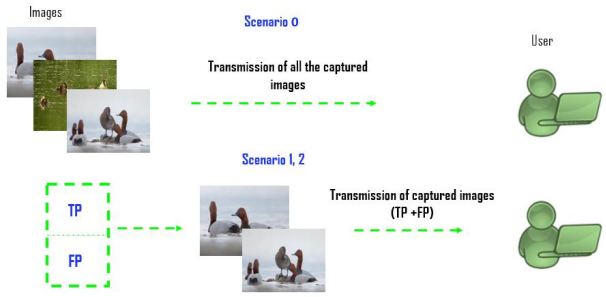


Fig. 4. Scenarios

located at the upper left triangle of side length $\rho \leq 8$ are considered as depicted in Figure 3. The resulting DCT block coefficients are quantized using the JPEG standard quantization matrix. Trade-off between quality level and compression rate can be obtained by choosing an appropriate quality factor (QF). An image visual quality ranges from the poorest ($QF = 1$) to the best quality ($QF = 100$). The obtained block is then zigzag linearized before applying an exponential-Golomb code Teuhola (1978) as a lossless entropy encoding.

3. PERFORMANCE ANALYSIS OF THE PROPOSED SYSTEM

In our surveillance application, we are concerned by monitoring three species of waterbirds threatened of extinction, namely, the *Ferruginous Duck*, the *Common Pochard* and the *White Headed Duck*. We assume that the monitoring is performed periodically N times during a day and compare the following scenarios (Figure 4) :

- $S0$ (resp. $S0$-ROI) : an image (resp. a ROI) is transmitted every period without audio identification ;
- $S1$: an image is only captured and transmitted when the audio identification detects a bird of interest without denoising. In $S1$-ROI, only ROI are transmitted ;
- $S2$ (resp. $S2$-ROI, our proposed system) operates as $S1$ (resp. $S1$-ROI) but with a denoising phase prior to the audio identification.

In order to evaluate the performance of the different scenarios, we made use of iFogSim Gupta et al. (2017) that allows modeling three-tier (edge-fog-cloud) architectures and considers physical, logical and management levels. The physical includes machine parameters at each level that are summarized in Table 1 as well as the edge devices characteristics. In our study, we considered sensor nodes equipped with an ARM Cortex M3 micro-controller cor (2018) and an ATMEL radio interface (Table 2). An audio sensor captures periodically a 2-second audio with a sampling frequency 22050 Hz. A visual sensor captures images of resolution 640×360 and compresses the image (or its ROI) prior to its transmission. The computational cost of the audio identification and the ROI extraction are

Table 1. Physical level parameters

Parameters	Cloud	Fog	Edge
CPU(MIPS)	44800	2800	100
RAM(MB)	40000	16000	128
Uplink bandwidth(Kbps)	10000	10000	250
Downlink bandwidth(Kbps)	10000	250	250

Table 2. IoT-LAB M3 board parameters

Micro-controller	Cortex M3 - 72 MHz, Power = 23 mW;
Cycles count	Add.[1], Mult.[2], Div.[12].
Radio chip	ATMEL AT86RF231
	$Power_{Tx} = 3$ dbm, data rate = 250 kbps
	$I_{Tx} = 14$ mA, $I_{Rx} = 12.3$ mA

Table 3. Computational cost

Step	Mult.	Div.	Add.
Audio denoising (GA)	823,038	411,420	618,128
Audio feature extraction (MFCC)	407,517	203,561	306,671
Classification (DTW)	–	–	313,638
KAZE nonlinear scale-space comp.	3,686,400	3,686,400	7,372,800
KAZE feature detection	–	–	921,600
KAZE feature description	–	30,464	–

summarized in Table 3. We assumed a link latency of 2 ms between the edge and the fog and 100 ms between the fog and the cloud. In iFogSim, an application is a collection of modules (tasks) linked by tuples (data flows) and modeled as a directed acyclic graph. We placed the modules at the appropriate level and set the characteristics of the tuples according to our scenarios.

3.1 Audio-based Identification Accuracy

We first, investigate the impact of the denoising step on the identification performance. We built a dataset of 500 noisy calls of our three target specimens that integrates a variety of noises encountered in our wetland. From our dataset, we selected 117 noisy target calls at three different SNR levels : 0 dB , 5 dB and 10 dB. In order to assess the behavior of the system when faced with non-target sounds, we added another 50 recordings. Among them, 30 belong to non-target bird species of the same area as our target ones, retrieved from Xeno-Canto website [1] . The remaining 20 recordings consist in various sounds that can be encountered in the natural habitat of our target birds. We, finally, used 40 clean target audio samples for reference. In total, 167 recordings were considered.

Figure 5 depicts the performance of the audio-based classification in $S1$ and $S2$ based on four metric. Precision is the proportion of relevant audio data among the retrieved sounds in the dataset. Recall (or sensitivity) indicates the ratio of the number of correctly identified birds to the total number of relevant calls in the dataset. F-measure balances between precision and sensitivity. Accuracy denotes the proportion of the total number of predictions that were correct in the dataset. We can clearly notice that $S2$ statistically outperforms $S1$ whatever the metric. The low performance of $S1$ for Recall, F-measure and Accuracy is attributable to the distortion of the audio call corrupted by environmental noises and the system was not able to identify most of the target bird calls. However, it was able to better identify the non-target calls due to the threshold

[1] `https://www.xeno-canto.org/`

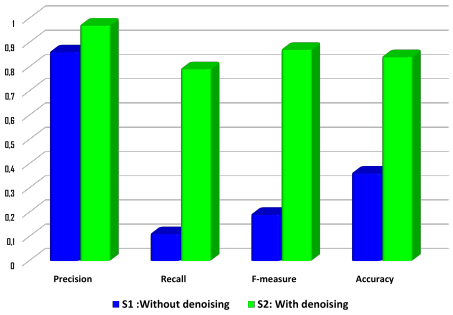


Fig. 5. Classification performance

used for classification. This is what justify the high value of precision of $S1$. In fact, both $S1$ and $S2$ were able to identify the non target calls with equal performance. The only difference is that in $S2$, birds calls were better recognized. To sum up, the $S2$ audio chain behaves better than $S1$ confirming the importance of the denoising phase in our proposal.

3.2 Latency and network usage

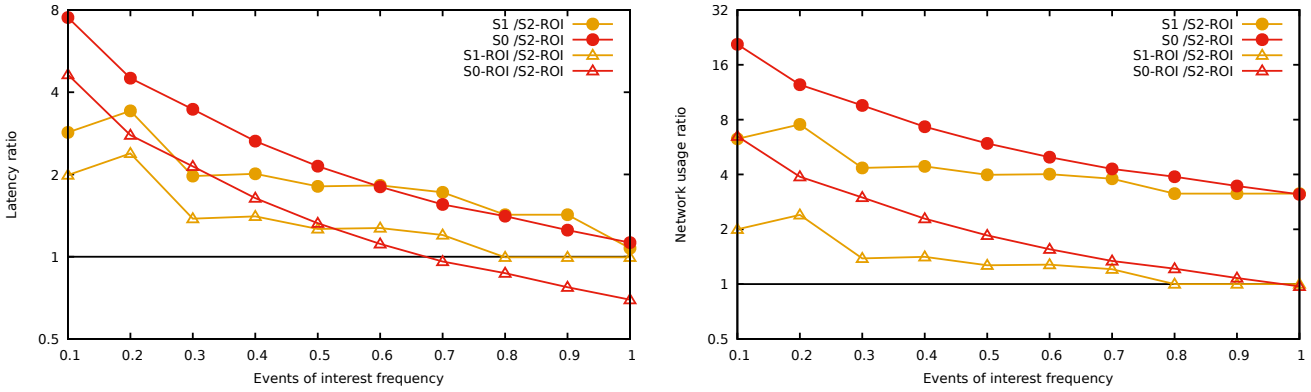
Figure 6 depicts the improvement obtained by our proposal over $S0$ and $S1$ strategies and their ROI variants with based on latency and network usage when considering the relevance of the reported information. We consider here the end to end delay required to obtain a useful information in which an event of interest is detected, from the audio capture until the end-user receives the visual information from the monitored area. Figure 6(a) plots the ratio of the delay achieved by our proposal (S2-ROI) with respect to the other scenarios as a function of the probability p that a target bird occurs in one period during the day. Figure 6(b) shows a similar metric but applied to network usage.

Our proposal improves in terms of both latency and network usage upon $S0$ and $S1$ where the entire captured images are sent which consumes more network resources. The improvement is more pronounced with respect to $S0$ especially when the occurrence probability of a target is low. Audio identification at the fog leads to a great improvement in network utilization in contrast to $S0$ and $S0$-ROI where we send the images arbitrarily without prior identification.

Compared with $S0$-ROI and $S1$-ROI, our proposed system behaves better in almost cases. In fact, relevant visual data are more likely to be sent. Moreover, sending only ROI allows farther reducing the network resources up to 59% in all scenarios. When the probability of events of interest exceeds 0.8, $S1$-ROI achieves the same performances as our proposal. However, we lose a significant amount of important visual data when compared to $S2$. $S0$-ROI becomes attractive when events of interest are more frequent.

3.3 Energy Consumption

In this section, we investigate the performance of the different scenarios based on their energy requirements. To do so, we estimate the overall energy consumed in one day. This, depending on scenarios, includes the energy



(a) Latency (b) Network usage

Fig. 6. Latency and network usage improvement.

needed to process the audio recording, E_{audio}, and the energy required to capture, encode and transmit visual data, E_{visual}. The total required energy consumed by the network is given by :

$$E = \begin{cases} N\ E_{visual} & \text{for } S0 \\ (TP + FP)\ E_{visual} + N\ E_{audio} & \text{for } S1 \text{ and } S2 \end{cases}$$

Figure 7 plots the ratio of the overall energy consumed in the different scenarios to the one of our proposal ($S2$-ROI) when the sensor node is able to reach the collection station in 5 and 10 hops respectively. First, we note that sending the ROI instead of the entire image reduces energy consumption for all the scenarios. The additional cost introduced by the ROI extraction is compensated by the energy required to transmit visual data. The improvement is higher when the number of crossed hops increases. It is clear that our proposal allows for an improvement when compared to $S0$ and $S0$-ROI that decreases with events of interest likelihood. The low energy consumption of $S1$ compared to $S2$ is explained by the low number of images identified as positive (TP). Without denoising, the system in unable to recognize the target birds which translates in fewer images being processed and transmitted. A significant amount of important data is lost.

To measure the impact of the loss of pertinent information in $S1$ when compared to $S2$, we consider the wasted energy to transmit irrelevant visual data using :

$$E_W = \begin{cases} (FP + TN + FN)E_{visual} & \text{for } S0 \\ FP \times E_{visual} + (FP + TN + FN)E_{audio} & \text{for } S1 \text{ and } S2 \end{cases}$$

Figure 8 plots, for $S1$ and $S2$ and their ROI variants, the percentage of wasted energy to the overall consumed energy. We observe that when only ROIs are compressed and sent, the wasted energy is slightly higher with respect to their corresponding strategy where the whole image is considered. This is due to the fact that the overall energy is lower. We note, as expected, that scenario $S1$ wastes more energy compared to $S2$ especially when a target bird is less likely to show up in the monitored area. The amount of wasted energy is higher for lower number of hops to cross from the visual sensor to the end-user.

As a conclusion, the findings attest that our proposed system ($S2$-ROI) is the most suitable scenario. We do not waste excessive energy as in $S0$ and $S0$-ROI, which consequently prolongs the energy gain of the node's lifetime and improves the network performance. Furthermore, we do not loose important informative data as in $S1$ and $S1$-ROI. As a result, our system offers the best compromise since it preserves network resources in terms of latency, bandwidth and energy while allowing for relevant data to be provided to the end user.

4. CONCLUSION

In this paper, we proposed a three-tier cooperative architecture following the *Cloud-Fog-Edge* model and propose an adequate placement of different surveillance tasks. To limit network resources utilization and hence prolonging its lifetime, we mainly suggested to reduce the amount of visual data to transmit. We implemented and evaluated low-complexity methods for a more accurate audio-based identification as well as an efficient technique for ROI extraction that is more suitable to our targeted species.

Our performance evaluation show that the proposed system performs well in the identification of bird calls thanks to the denoising phase. Moreover, they attest that the transmission of compressed ROI instead of the entire image allows to reduce the network resources utilization in terms of bandwidth and energy while the amount of pertinent information is maximized. This proved the efficiency of our proposed architecture, in particular, the benefit of the cooperation between its three tiers.

As a future work, we expect to farther reduce the complexity of the different processing phases including the audio-based identification and the ROI extraction methods.

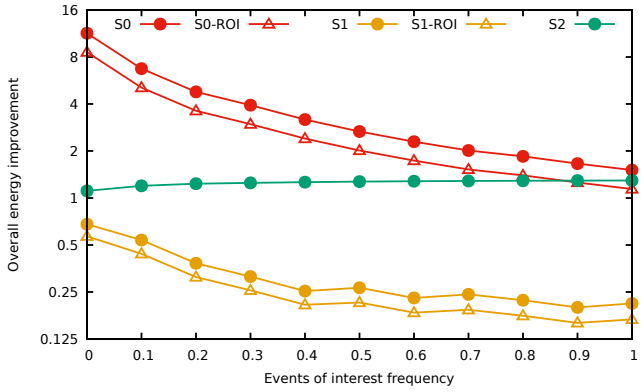
REFERENCES

(2018). Cortex m3 datasheet. URL `iot-lab.github.io/assets/misc/docs/iot-lab-m3/stm32f103re.pdf`.

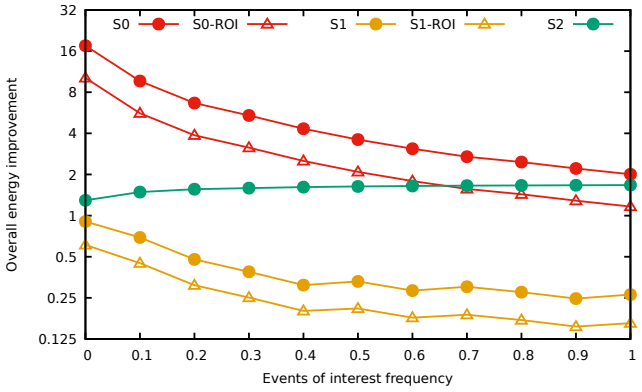
Akyildiz, I.F., Melodia, T., and Chowdhury, K.R. (2007). A survey on wireless multimedia sensor networks. *Computer Networks*, 51(4).

Alcantarilla, P.F., Bartoli, A., and Davison, A.J. (2012). KAZE Features. In A. Fitzgibbon, S. Lazebnik, P. Perona, Y. Sato, and C. Schmid (eds.), *Computer Vision – ECCV 2012*. Springer.

Allinson, T. (2018). State of the world's birds. Technical report, BirdLife International.

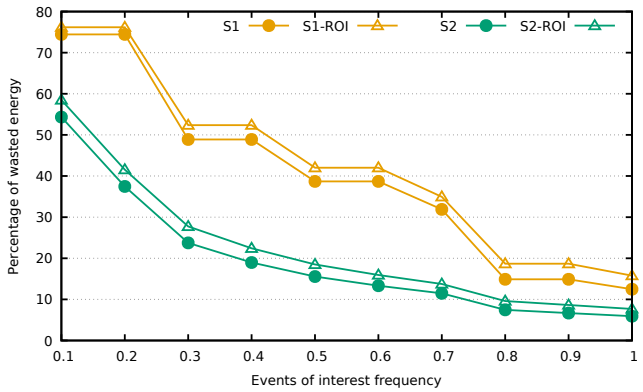


(a) 5 hops

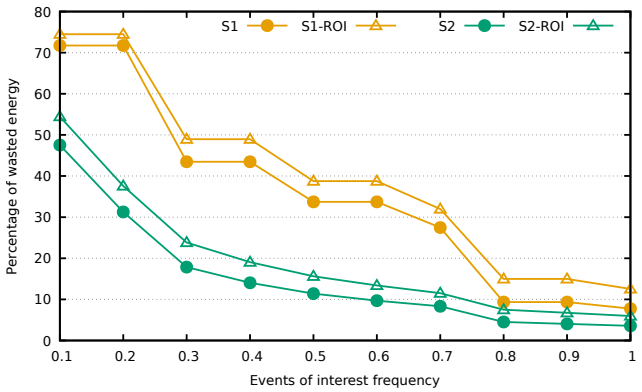


(b) 10 hops

Fig. 7. Improvement of overall consumed energy



(a) 5 hops



(b) 10 hops

Fig. 8. Percentage of wasted energy

Archaux, F. (2011). On methods of biodiversity data collection and monitoring. *Revue Science Eaux & Territoires, Public policy and biodiversity*.

Civelek, M. and Yazici, A. (2016). Automated moving object classification in wireless multimedia sensor networks. *IEEE Sensors Journal*, 17(4).

Dyo, V., Ellwood, S.A., Macdonald, D.W., Markham, A., Trigoni, N., Wohlers, R., Mascolo, C., Pásztor, B., Scellato, S., and Yousef, K. (2012). Wildsensing: Design and deployment of a sustainable sensor network for wildlife monitoring. *ACM Transactions on Sensor Networks (TOSN)*, 8(4).

Gupta, H., Vahid Dastjerdi, A., Ghosh, S.K., and Buyya, R. (2017). iFogSim: A toolkit for modeling and simulation of resource management techniques in the Internet of Things, Edge and Fog computing environments. *Software: Practice and Experience*, 47(9).

Gupta, S., Jaafar, J., Ahmad, W.W., and Bansal, A. (2013). Feature extraction using mfcc. *Signal & Image Processing*, 4(4).

Hunkeler, U., Truong, H.L., and Stanford-Clark, A. (2008). Mqtt-s a publish/subscribe protocol for wireless sensor networks. In *IEEE Int. Conf. on Communication Systems Software and Middleware and Workshops (COMSWARE'08)*.

Kizilkaya, B., Ever, E., Yatbaz, H.Y., and Yazici, A. (2022). An effective forest fire detection framework using heterogeneous wireless multimedia sensor networks. *ACM Transactions on Multimedia Computing, Communications, and Applications*, 18(2).

Koyuncu, M., Yazici, A., Civelek, M., Cosar, A., and Sert, M. (2018). Visual and auditory data fusion for energy-efficient and improved object recognition in wireless multimedia sensor networks. *IEEE Sensors Journal*, 19(5).

Lowe, G. (2004). Sift-the scale invariant feature transform. *Int. J*, 2(91-110).

Lu, Y. and Loizou, P.C. (2008). A geometric approach to spectral subtraction. *Speech communication*, 50(6).

Magno, M., Tombari, F., Brunelli, D., Di Stefano, L., and Benini, L. (2013). Multimodal video analysis on self-powered resource-limited wireless smart camera. *IEEE Journal on Emerging and Selected Topics in Circuits and Systems*, 3(2).

Maimour, M. (2018). SenseVid: A traffic trace based tool for QoE video transmission assessment dedicated to wireless video sensor networks. *Simulation Modelling Practice and Theory*, 87.

Müller, M. (2007). *Dynamic Time Warping*, 69–84. Springer Berlin Heidelberg, Berlin, Heidelberg. doi:10.1007/978-3-540-74048-3_4.

Salahat, E. and Qasaimeh, M. (2017). Recent advances in features extraction and description algorithms: A comprehensive survey. In *IEEE Int. Conf. on industrial technology (ICIT)*.

Standard, O. (2014). Mqtt version 3.1.

Stattner, E., Hunel, P., Vidot, N., and Collard, M. (2011). Acoustic scheme to count bird songs with wireless sensor networks. In *2011 IEEE International Symposium on a World of Wireless, Mobile and Multimedia Networks*. IEEE.

Tareen, S.A.K. and Saleem, Z. (2018). A comparative analysis of sift, surf, kaze, akaze, orb, and brisk. In *Int. Conf. on computing, mathematics and engineering technologies (iCoMET)*. IEEE.

Teuhola, J. (1978). A compression method for clustered bit-vectors. *Information processing letters*, 7(6), 308–311.

Weiss, M.R., Aschkenasy, E., and Parsons, T.W. (1975). Study and development of the intel technique for improving speech intelligibility. Technical report, NICOLET SCIENTIFIC CORP NORTHVALE NJ.

ZainEldin, H., Elhosseini, M.A., and Ali, H.A. (2015). Image compression algorithms in wireless multimedia sensor networks: A survey. *Ain Shams engineering journal*, 6(2).