A Real-time Portable Bioacoustics Species Identification Design Concepts

Naufal Alee[#1], Phaklen Ehkan[#2], L.M. Kamarudin[#3], R. Badlishah Ahmad[#]
[#] School of Computer and Communication Engineering
Universiti Malaysia Perlis, Pauh Putra Campus, 02600 Arau, Perlis, Malaysia
[1]thepsin@gmail.com
[2]phaklen@unimap.edu.my
[3]3latifahmunirah@unimap.edu.my

Abstract— Recently there has been an increasing demand of an automated system for animal species identification, where it needs a perfect good knowledge, understanding of the nature under vision and proper efficient system design. Embedded systems nowadays are offering a brilliant solution. Based on nature of economic and feasibility of advanced, embedded technology is chosen. This paper proposes a design of real-time portable bioacoustics species identification system. It contains two major correlated modules apart, the identification module and the system control module. The identification module is to be implemented in FPGA hardware to achieve species identification process while the system control module will manage and control the entire system. The proposed system is a combination of hardware, software development and operating system customization. It is designed to be decentralize, therefore the need of any server is eliminated. It can be placed anywhere, can be viewed and accessed from anywhere through a web server built-in.

Keyword- Species Identification; Biodiversity; Embedded System; Single Board Computer; Field Programmable Gate Array; Internet of Things; Wireless Sensor Network

I. INTRODUCTION

Nowadays, biodiversity becomes very important and urgent task. It is going to be more important because of world climate change and the rise of human population that cause animals emigrations and high rates of species loss. In the need of conservation, biologists and ecologist need more information. Unfortunately, biodiversity assessment is a difficult task for them. Conventional biodiversity assessment is insufficient in terms of time and cost, assessing with humans seems impossible not to make any interfere to habitants and sometime it is too risky. The conservative biodiversity assessment requires the assumption of expertise because data are often acquired by animal sound production. For example, the sounds of each animal scratching with trees, drinking water or howling can be differentiated by experts. However, few methods have been developed by researchers to reduce those disadvantages. Biodiversity assessment by analysing recorded acoustic signal has been introduced where an array of microphones has to be deployed in the field to record acoustic signal during periods of time. Recorded data will then be collected either manually or automatically and transferred to the server for species identification. Thus, a server with high performance processing of complex classification algorithm is required.

While the technology of embedded system and data-acquisition (DAQ) have been significantly advancing every year. The opportunity to improve the technology of observing and monitoring wildlife is arise. DAQ systems are widely used in industry and consumer applications to achieve remote data collection and monitoring. The DAQ system is designed to collect data in the simplest form, and then it will be sent to server. An embedded DAQ system with remote accessibility has been introduced by A. Z. Alkar and M. A. Karaca [1]. It was built to monitor road traffic. An embedded board is integrated with a camera and a global system for mobile communications (GSM) modules. It acquires bulk image and saves into flash memory on the embedded unit. The acquired image will be transferred to a server via general packet radio service (GPRS). Users can access to embedded device directly for real-time or to the server for logged data. Later in 2013, [2] have developed a system called automated remote biodiversity monitoring network (ARBIMON) but instead of acquiring images, ARBIMON acquires an audio. It consists of 3 parts called acoustic permanent station, field base station and ARBIMON server. The main devices of the acoustic permanent station include a microphone, a router and iPod (2G) control recording devices. It used to send recorded audio to MacMini computer on the field base station. The field base station receives data from the acoustic permanent station via wireless communication. Other than MacMini, field base station also has 1 Tb external hard drive and ability to access the internet. This station is used to store data, compress the recording file and forward to the ARBIMON server at University of Puerto Rico for data processing and data management. General diagram of ARBIMON is shown in Figure1.

Fig. 1. General diagram of ARBIMON system

Several DAQ systems are similar to [1], such as [3], were designed for low cost and decentralize. These applications are able to collect data, but incapable of doing complex computational methods (e.g. artificial neural network (ANN), Gaussian mixture model (GMM), hidden Markov model (HMM)) establishing in the classification process. Those systems that similar to [2], such as [4] – [6], are focused for computational performance. The needed data to be transferred to server is very powerful. These systems are not considered about cost or decentralization.

The evolution of voice technology also has been advanced. Technology of artificial intelligence and forensic science give machines the human-like ability to distinguish voice's identity from many sources such as an acoustic signal produced by the animal. Bioacoustics focuses on animal acoustic communication. It is highly associated with the technique of feature extraction and pattern classification. Feature extraction is used to process the acoustic signal waveform into a representation of lower level information. Few varieties of feature extraction techniques are practically used for recognition task, such as linear predictive coding (LPC) [7], Mel frequency cepstral coefficient (MFCC) [8], perceptual linear predictive (PLP) [9] and others. The MFCC and PLP are the most popular acoustic features used.

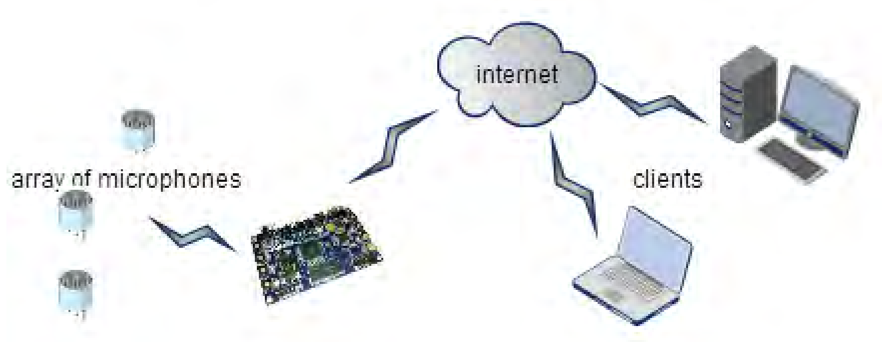

Fig. 2. Diagram of the proposed system

The pattern classification is an important part in acoustic modeling. Animal acoustic signal carries much information about each species. The results of it strongly affecting the identification engine. Many techniques have been developed earlier such as dynamic time warping (DTW) [10] [11], and HMM. These techniques are not efficient for real-time application [12]. There are alternative methods have been introduced such as vector quantization (VQ) [13], GMM and support vector machine (SVM) [14].The GMM is one of the most commonly used types of pattern classifier. The use of GMM is most common due to its capability of performing a complete text independent situation. The performance of this technique has been verified in text independent identification [15]. The GMM technique of pattern classification in previous studies appeared to have several advantages. GMM is based on probabilistic framework and it provides high-accuracy recognition. However, the process practically does not always produce satisfied result due to the long computational time [16]. Consequently, alternative methods must be sought in order to reduce processing time problem for GMM technique. GMM implemented very well performance, but its training process requires a lot of time and they get numerically unstable when trained with small amount of data. The main problem is the inversion of the covariance matrices. Therefore, hardware solution has been proven to solve the speed problem [17].

There are several hardware applications that can be used for better performance such as application specific integrated circuits (ASICs) or digital signal processing (DSP) applications. Even though ASICs provide the highest performance, but they are customized for a specific application and very expensive. DSP-based are cost efficient, low in power consumption and heat-emission, but they only provide a limited speed for data processing. FPGAs are usually slower than ASICs, but have the advantage of shorter time to market, ability to be re-programmed in the field for errors correction as well as upgrades, flexibility, and reducing-cost. Therefore, they combine many advantages of ASICs and DSPs [18]. The use of hardware description languages (HDLs) allows FPGAs to be more suitable for different types of designs. Due to the exponential increase of technologies, designers faced with many problems which require the advent of systems that can be fast, flexible, and re-programmable. With the advantages of real-time and in-circuit re-configurability make the FPGA-based system flexible, programmable and reliable. R. Ramos-Lara, et al. [19] have investigated the performance of FPGA in comparison with Pentium IV computer for a single voice stream. Results shows that FPGA achieved the same performance with 24% lower cost. In [20], FPGA implementation based on GMM for speaker identification has been implemented. It shows that classification in hardware is about ninety times faster. The hardware systemis capable of processing more audio streams in real time than could be done in standard computer.

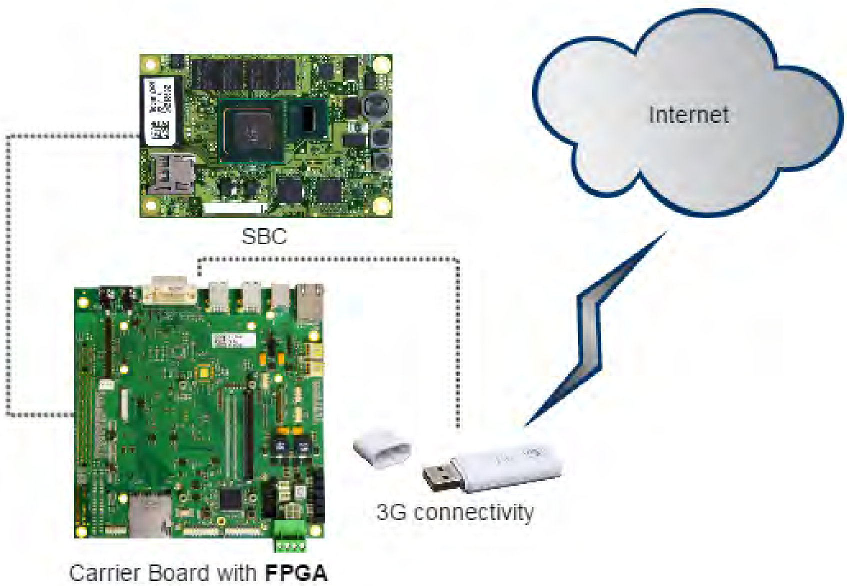

Fig. 3. Hardware components of proposed system

In this research, a proposed bioacoustics species identification system is constructed to provide an efficient method for biodiversity assessment. The system is composed of identification and system control modules. The identification module operates on the FPGA which offers a parallel processing scheme that can help with complex species identification, while system control module is operated by SBC which offers flexibility of installation, mobility, cost and size. The system control module is composed of overall system, managing and viewing package.

II. BIOACOUSTICS SPECIES IDENTIFICATION SYSTEM

It has always been difficult to develop a robust identification system not only because bioacoustics signal is dynamically varies by many factors and the complexity of biometric algorithms, but also the requirement of real-time monitoring. The problem of identification system belongs to pattern classification. The goal of pattern classification is to classify objects of interest into a number of categories or classes [21]. The categories or classes here are referred to the individual species. One proposed solution is to ascertain and enhance GMM pattern classification approach via reconfigurable hardware implementation. This pattern classification approach should be able to handle large database in short time limit, whereas the accuracy rate is still maintained or even higher than the conventional software based GMM pattern classification technique. This paper categorizes the proposed system mainly into two modules, the identification and system control module. Figure 4 shows the block diagram of the proposed system. The system integrates various platforms to be successful solution and to overcome previous research lacks.

A. Identification Module

There are various hardware components to be used for developing bioacoustics species identification. In general, the hardware components are FPGA, SBC and 3G wireless connectivity module. SBC is a low-cost, low-power consumption, small-size, robust and fully integrated PC compatible, which employs advanced embedded technology to develop compact computing system. The SBC is a complete computer build on a single circuit board. The primary advantages of SBC versus Desktop PC are cost and size. Nowadays, it has been widely used in consumer electronics, industrial control, navigation equipment, medical instrument, network communication, automobile, military and national defence [22] and etc.

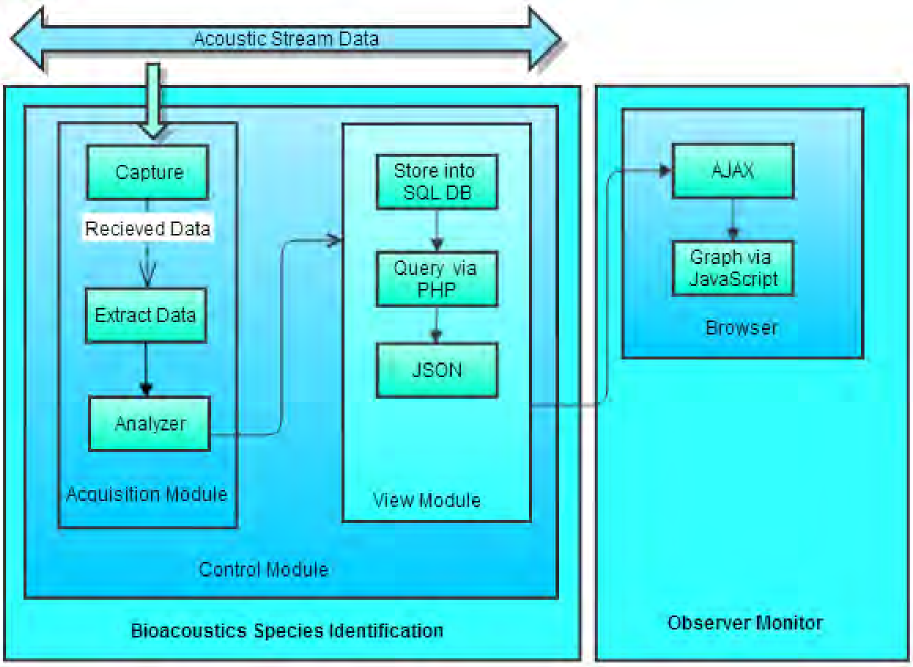

Fig. 4. Bioacoustics identification method

PCs are powerful enough to allow large data streams to perform identification one species for the best results, with very less background noise. Next steps have to be taken in order to reduce the computational complexity, so that the real time identification is achievable. Hardware implementation has the computational resources to achieve real-time scheme. One of the main advantages include FPGA in this system is to obtain processing speed improvements in comparing to software implementation and allow processing of multiple voice streams with an increase of population of species. The training of the acoustic models is to be performed offline and the classification stage is to be implemented in hardware-based using FPGA. Feature vectors from the data will be stored in random access memory (RAM) which is directly connected to the FPGA.

Species identification module is implemented in software platform by using Matlab for troubleshooting and testing before hardware implementation. This enabled variables and factors affecting species identification to be investigated more easily than in hardware implementation. The best methods will be chosen for implementing in hardware upon completion of the testing in software. A software-based system will also be used as a backup during hardware implementation as if there is any changes required in the algorithm for the purpose of hardware optimization for instance can be first tested in the software. The objectives for developing a species identification system in the software include to develop a working system that is easier to manipulate as debugging and to investigate various factors affecting performance in determining the best implementation in hardware. The factors include the number of Gaussian components, data normalization, and the energy component of the speech signal and the length of test utterance.

Then, the implementation of the system will be experimented with software–based Matlab and well-known speech recognition toolkit, hidden Markov model toolkit (HTK) platform. HTK is a portable toolkit for building and manipulating hidden Markov models. It is primarily used for speech recognition research although it has been used for numerous other applications including research into speech synthesis, character recognition and DNA sequencing. The HTK will be used to perform the analysis of front-end processing while Matlab to perform the statistical analysis using GMM approach. It is easier to investigate which criteria or factors affecting the species identification system compared to system in hardware implementation. Based on these experiments the best options are to be implemented in hardware for identification. The two most significant results to be considered are first, relation between the numbers of components used in the Gaussian mixture and the accuracy of the system and second, the relation between the length of test utterance and accuracy. Limitations in the hardware in terms of area and speed will be determined.

B. System Control Module

This subsection describes a design of the Bioacoustics species identification's system control module (SCM). SCM needs to work on the lower layer, closely with the operating system (OS) layer, with native programming language such as C and customization of OS to be authorized to manage the entire system. SCM will be capable of hardware probing, doing routine jobs, status checking and fixing some errors or noticing the error, unmanned.

The view module is a sub-module of the system control module. From the existing system, data were analysed and graphed by the system itself, but this module uses new approach where all presentation data are to be analysed, drawn and shown by monitoring side, only plain text in JavaScript object notation (JSON) format is needed. The rise of Web 2.0 has brought with it more sophisticated user interfaces and client-side browser functionality. It is to be developed using hypertext markup language (HTML), cascading style sheet (CSS) and JavaScript language. Asynchronous JavaScript and XML (AJAX) is used to reduce data transmit between the system and observer devices. AJAX is a new way of design and develop web applications. It is a new technique that uses a set of open standards technologies, with support by cross-browser and cross-platform compatibility. These technologies are widely used and well known. JavaScript becomes a key factor for browser vendors and web application developers [23]. Browser vendors are developing new technology at JavaScript to speed up its execution. This is necessary because to replace the web application with desktop application require a large amount of JavaScript code that also must execute at a reasonable speed. In addition, users often have multiple tabs running web application, such as email, calendars, image viewing and manipulation, and so on .The technologies work together in different levels, each one with specific functionality to create a powerful new development model for web applications [24].

Acquisition module in Figure 4 represents an identification module which has been earlier described. The capture element represents the process when the FPGA receives data from microphones. Then, the next element is extract which represents the process of feature extraction. The analyser element represents the process that reads the extracted information from extract element and makes identification process described. The last element shows that the analyse data is to be stored in a database.

JSON is a lightweight data exchange format. Relatively to the extensible markup language (XML), it is easy for machines to parse and generate. Its format is different from the XML which uses closed tags. JSON uses a text format that is independent of the language. JSON objects are analysed as string arrays. This can be faster than XML which is above characteristic. JSON has higher parsing efficiency and the advantages of easy preparation. As lightweight AJAX applications have less demanding on the security while requirements for efficiency are high and have good support of JavaScript. It will be loaded and processed by the client side. Web-browser in client side, which can be any device, will read data, analyse and plot graphs. The system control module is to be developed by native programing to control entire system and make system being automated.

III. DISCUSSION AND CONCLUSION

Based on the mentioned disadvantages of conventional biodiversity assessment methods, a new proposed solution of a bioacoustics FPGA-based hardware design has been stated in details in this paper. Due to the advantage of parallel processing, classification approaches for species identification; hardware-based solution is required. This proposed solution focuses on construction of embedded FPGA-based system leading to decrease processing time and result in higher accuracy. The reason is that a GMM is effective and provide a stable accuracy while handling large data. Base on the previous work survey in this paper, it has clearly be shown that successful implementation of a real-time portable bioacoustics species identification system need a detailed study and a serious estimation of problem constraints and variety, especially that the platform solution must combine software and hardware.

REFERENCES

[1] A. Z. Alkar and M. A. Karaca, "An internet-based interactive embedded data-acquisition system for real-time applications," IEEE Transactions on Instrumentation and Measurement, vol. 58, no. 3, pp. 522–529, Mar. 2009.

[2] T. M. Aide, C. Corrada-Bravo, M. Campos-Cerqueira, C. Milan, G. Vega, and R. Alvarez, "Real-time bioacoustics monitoring and automated species identification," PeerJ, vol. 1, p. e103, Jul. 2013.

[3] W.-T. Sung and C.-C. Hsu, "Intelligent environment monitoring system based on innovative integration technology via programmable system on chip platform and ZigBee network," IET Communications, vol. 7, no. 16, pp. 1789–1801, Nov. 2013.

[4] C. H. See, K. V. Horoshenkov, R. A. Abd-Alhameed, Y.-F. Hu, and S. J. Tait, "A low power wireless sensor network for gully pot monitoring in urban catchments," IEEE Sensors Journal, vol. 12, no. 5, pp. 1545–1553, May 2012.

[5] J. P. Tello, O. Manjarres, M. Quijano, A. Blanco, F. Varona, and M. Manrique, "Remote monitoring system of ecg and human body temperature signals," Latin America Transactions, IEEE (Revista IEEE America Latina), vol. 11, no. 1, pp. 314–318, Feb. 2013.

[6] J. Gutierrez, J. F. Villa-Medina, A. Nieto-Garibay, and M. A. Porta-Gandara, "automated irrigation system using a wireless sensor network and gprs module," IEEE Transactions on Instrumentation and Measurement, vol. 63, no. 1, pp. 166–176, Jan. 2014.

[7] B. S. Atal and S. L. Hanauer, "Speech Analysis and Synthesis by Linear Prediction of the Speech Wave," The Journal of the Acoustical Society of America, vol. 50, no. 2B, pp. 637–655, 1971.

[8] S. Davis and P. Mermelstein, "Comparison of Parametric Representations for Monosyllabic Word Recognition in Continuously Spoken Sentences," IEEE Transactions on Acoustics, Speech and Signal Processing, vol. 28, no. 4, pp. 357–366, 1980.

[9] H. Hermansky, "Perceptual Linear Predictive (PLP) Analysis of Speech," The Journal of the Acoustical Society of America, vol. 87, no. 4, pp. 1738–1752, 1990.

[10] H. Sakoe and S. Chiba, “Dynamic Programming Algorithm Optimization for Spoken Word Recognition,” IEEE Transactions on Acoustics, Speech and Signal Processing, vol. 26, no. 1, pp. 43–49, 1978.
[11] J. Liu, Q. Cheng, Z. Zheng, and M. Qian, “A DTW-based Probability Model for Speaker Feature Analysis and Data Mining,” Pattern Recognition Letters, vol. 23, no. 11, pp. 1271–1276, Sep. 2002.
[12] L. Rabiner, “A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition,” Proceedings of the IEEE, vol. 77, no. 2, pp. 257–286, 1989.
[13] V. Radová and Z. Švenda, “Speaker Identification Based on Vector Quantization,” in Text, Speech and Dialogue, V. Matousek, P. Mautner, J. Ocelíková, and P. Sojka, Eds. Springer Berlin Heidelberg, 1999, pp. 341–344.
[14] R. Solera-Ureña, D. Martín-Iglesias, A. Gallardo-Antolín, C. Peláez-Moreno, and F. Díaz-de-María, “Robust ASR using Support Vector Machines,” Speech Communication, vol. 49, no. 4, pp. 253–267, Apr. 2007.
[15] D. A. Reynolds and R. C. Rose, “Robust Text-independent Speaker Identification using Gaussian Mixture Speaker Models,” IEEE Transactions on Speech and Audio Processing, vol. 3, no. 1, pp. 72–83, 1995.
[16] Q. Y. Hong, S. Kwong, and H. L. Wang, “optimization of gaussian mixture model parameters for speaker identification,” in Genetic and Evolutionary Computation – GECCO 2004, K. Deb, Ed. Springer Berlin Heidelberg, 2004, pp. 1310–1311.
[17] D. A. R. Dr and W. M. C. Dr, “text-independent speaker recognition,” in Springer Handbook of Speech Processing, P. J. B. Dr, P. M. M. Sondhi, and P. Y. (Arden) H. Dr, Eds. Springer Berlin Heidelberg, 2008, pp. 763–782.
[18] E. Ayeh, K. Agbedanu, Y. Morita, O. Adamo, and P. Guturu, “FPGA implementation of an 8-bit simple processor,” in IEEE Region 5 Conference, 2008, pp. 1–5.
[19] R. Ramos-Lara, M. Lopez-Garcia, E. Canto-Navarro, and L. Puente-Rodriguez, “SVM Speaker Verification System Based on a Low-Cost FPGA,” in International Conference on Field Programmable Logic and Applications (FPL), 2009, pp. 582–586.
[20] P. EhKan, T. Allen, and S. F. Quigley, “FPGA Implementation for GMM-based Speaker Identification,” International Journal of Reconfigurable Computing., vol. 2011, pp. 3:1–3:8, Jan. 2011.
[21] R. O. Duda, P. E. Hart, and D. G. Stork, Pattern Classification. John Wiley & Sons, 2012.
[22] H. Chao, Y. Cao, and Y.-Q. Chen, “Autopilots for small fixed-wing unmanned air vehicles: a survey,” in International Conference on Mechatronics and Automation, 2007. ICMA 2007, 2007, pp. 3144–3149.
[23] H. M. Kienle, “It’s about time to take javascript (more) seriously,” IEEE Software, vol. 27, no. 3, pp. 60–62, May 2010.
[24] J. S. Zepeda and S. V. Chapa, “From desktop applications towards ajax web applications,” in 4th International Conference on Electrical and Electronics Engineering, 2007. ICEEE 2007, 2007, pp. 193–196.