Facultatea de Electronică,Telecomunicaţii şi Tehnologia Informaţiei

PhD Thesis

Audio-Video Techniques for the Recognition of Intruders and Fires in Secure Environments

PhD Candidate: **Scientific Supervisor:**
Ing. Elena-Roxana BUHUS **Prof. Dr. Ing. Corneliu RUSU**

Doctoral Thesis Evaluation Committe:

President: Prof. Dr. Ing. Sorin Hintea - Technical University of Cluj-Napoca

Scientific Supervisor: Prof. Dr. Ing. Corneliu Rusu - Technical University of Cluj-Napoca

Referees:

- Prof. Dr. Eng. Liviu Goras - "Gh. Asachi" Technical University of Iasi;
- Prof. Dr. Eng. Alexandru Isar - Politehnica University of Timisoara;
- Prof. Dr. Eng. Eugen Lupu - Technical University of Cluj-Napoca

– Cluj-Napoca –
2023

List of Abbreviations

ARFF	Attribute-Relation File Format
ARTMAP	Adaptive Resonance Theory Map
CIE	Commission Internationale de l'Eclairage
CIE L*A*B*	Luminance, Red-green axis, Blue-yellow axis
CIELAB	Luminance, Red-green axis, Blue-yellow axis
CIE LUV	L coordinate, U coordinate, V coordinate
CIE XYZ	X coordinate, Y coordinate, Z coordinate
CMYK	Cyan Magenta Yellow black
CRT	Cathode Ray Tube
CSV	Comma-Separated Value
DFT	Discrete Fourier Transform
DCT	The Discrete Cosine Transform
DTFT	Discrete-Time Fourier Transform
FAM	Fuzzy Adaptive Resonance Theory Map
FIS	Fuzzy Inference System
FN	False Negative
FP	False Positive
GMM	Generalized Method of Moments
GNU	General Public License
HSL	Hue Saturation Lightness
HSV	Hue Saturation Value
IDTF	Invers Discrete Fourier Transform
IBk	Instance based algorithm: K-nearest neighbours regression
LBP	Local Binary Patterns
LCD	Liquid Crystal Display
LPC	Linear Predictive Coding
LPCC	Linear Predictive Cepstral Coefficient
LMT	Logistic Model Tree
MB	Moving Block
MISO	Multiple Input Sigle-Output
NCS	Natural Colour System
RGB	Red Green Blue
RGBA	Red Green Blue Alpha
ROI	Region of Interest
ROI -A	Region of Interes Absent
ROI -P	Region of Interest Present

REP	Reduced Error Pruning
SB	Smoke Block
SFAM	Simplified Fuzzy Adaptive Resonance Theory Map
SMO	Sequential Minimal Optimization
SVM	Support Vectors Machine
KSTAR	K-nearest neighbour
JRIP	Java Repeated Incremental Pruning
J48	Java 48
TESPAR	Time Encoded Signal Processing and Recognition
TN	True Negative
TP	True Positive
Hz	Hertz
C1	Logistic Classifier
C2	Multilayer Perceptron Classifier
C3	Simple Logistic Classifier
C4	SMO Classifier
C5	Ibk Classifier
C6	Kstar Classifier
C7	Attribute Selected Classifier
C8	Bagging Classifier
C9	Classification Via Regression
C10	Filtered Classifier
C11	Iterative Classifier Optimizer
C12	Logit Boost Classifier
C13	Multi Class Classifier
C14	Random Committee Classifier
C15	Random Sub Space Classifier
C16	Decision Table Classifier
C17	Jrip Classifier
C18	PART Classifier
C19	J48 Classifier
C20	LMT Classifier
C21	Random Forest Classifier
C22	REP Tree Classifier

Introduction

In this doctoral thesis, the study and development of intelligent methods for monitoring and detecting intruders in protected open spaces have been proposed. The aim was to identify solutions that enable timely response to unforeseen situations, such as poaching activities (through the detection of gunshot sounds or unauthorized vehicles), illegal deforestation (through the detection of chainsaw sounds), and the early identification of fires. Timely identification of these situations provides significant support to authorities in the preservation and protection of the environment.

Audio-video monitoring of these protected areas has become one of the primary means of detecting various unwanted events. This method can detect the presence of fires, intruders, poaching activities, as well as illegal deforestation in forested areas.

Environmental protection organizations are proactive in identifying events involving illegal deforestation and hunting in natural reserves and forests through various support actions for local authorities. They aim to frequently identify and report such situations through social media platforms (Facebook) and radio and television broadcasts (e.g., Agent Green, Green Peace Romania).

Monitoring these environments using an audio detection system offers several advantages compared to simple video monitoring. Sounds can provide information that cannot be extracted from video alone, such as concealed subjects, monitoring dark areas, detecting abnormal sounds in seemingly normal scenes, and providing a higher level of privacy compared to video systems [1].

In this thesis, a series of experiments were conducted to analyze the response of acoustic classifiers in the context of detecting sounds that can be categorized as poaching or illegal deforestation events (e.g., chainsaw sounds, gunshot sounds). Based on these experiments, the most suitable classifiers were identified for the development of potential audio detection systems.

The occurrence of fires in forested areas, and beyond, represents a disruption with a negative impact on both the economy and the environment. Therefore, early fire detection is necessary to minimize its unwanted effects.

Compared to traditional implementations based on smoke detectors [2][3], this thesis proposes a solution based on fire detection from video sequences to overcome the disadvantages of delayed heat sensor responses and reduce false alarms. Additionally, an intelligent video application called "RiskCam" is proposed for detecting various events within the same video sequence or from parallel cameras using specific detection algorithms.

The need to identify solutions to support the design and implementation of audio-video monitoring systems is a current interest. This process involves both the development and analysis of scenarios capable of addressing a range of challenges encountered in protected wilderness environments.

The thesis is divided into two research directions: one focusing on audio and the other on video. The structure of the thesis is as follows:

Chapter 1 presents a series of theoretical concepts related to audio techniques for intruder detection in protected environments.

Chapter 2 illustrates various video techniques currently used for detecting fires in open spaces, along with their evaluation based on databases defined by the author. For fire detection, a separate analysis of fire pixels is considered, using RGB feature classifiers [4][5], RGB and HSV [6], RGB and YCBCR [7], as well as CIE L*A*B YCBCR and iterative K-means clustering [8]. An analysis of smoke pixels is also included, using static and dynamic smoke components with the help of a fuzzy interference system [9][10][11].

The following section outlines the author's personal contributions, presenting the research objectives and the working hypothesis. Chapter 3 presents the general methodology adopted in this thesis.

Chapter 4 presents a study conducted by the author focusing on the detection of intruders, poaching activities, and illegal deforestation, highlighting the experimental results and relevant conclusions.

Chapter 5 is dedicated to a study conducted by the author regarding fire detection.

Chapter 6 is dedicated to a study involving the presentation of an intelligent application concept that encompasses various functionalities. Such an application would be of interest for monitoring open spaces with a high level of security. The proposed algorithms for each functionality are presented, along with the motivation behind this concept based on a case study. A series of experiments and tests conducted for each proposed functionality are also presented.

Chapter 7 includes conclusions and future developments, as well as personal contributions made through this research.

The final part of the thesis consists of the bibliography, corresponding annexes, a list of publications, and research awards obtained.

1 Obiectivele tezei

The current need to define advanced open-space monitoring systems presents a significant interest in protecting flora and fauna. In this paper, it was considered that based on the analysis of current implementations, new possibilities can be evaluated and experimented with to support the design of such systems.

Two research directions were explored in this thesis. One direction was in the context of audio, where techniques for detecting intruders, poaching situations, and illegal deforestation were analyzed. The other research direction was in the context of video, where the detection of fire situations, abnormal motion, and facial authentication were of interest.

The general objectives of this work were as follows:

• Identifying new classification methods for a class of sounds specific to wilderness environments and detecting sounds belonging to possible subjects categorized as intruders or poaching situations.

• Identifying and developing an integrated video solution for the detection of potential risks in protected or high-security wilderness environments.

Both the investigative and developmental aspects for the two aforementioned directions can be viewed as a division of objectives. Thus, we can define, on one hand, a series of theoretical objectives and, on the other hand, a series of practical objectives.

Among the theoretical objectives, we can mention:

• Identifying and understanding methods of capturing acoustic audio signals.

• Digital processing of audio signals.

• Methods for characterizing digital audio signals.

• Methods for classifying digital audio signals.

• Methods for building audio monitoring systems.

• Understanding methods of capturing video sequences.

• Digital image processing.

• Methods for characterizing fire and smoke pixels in images.

• Video fire detection methods.

• Methods for characterizing faces in images.

• Facial recognition methods.

• Methods for motion detection in video.

• Design methods for video monitoring systems.

In addition to the theoretical part, an experimental analysis of the above-mentioned objectives was also desired so that a series of findings could be developed based on the practical results obtained.

Therefore, practical objectives included:

• Identifying a database with relevant sound classes for research purposes: chainsaw, gunshots, bird chirping, human voice, and tractor.

• Characterizing sounds using linear predictive coding.

• Characterizing sounds using cepstral coefficients of linear predictive coding.

• Identifying suitable classifiers for sound classification in WEKA.

• Defining experiments in WEKA and conducting them.

• Collecting and interpreting the results of experiments conducted in WEKA.

• Collecting and defining a set of images and video recordings relevant to the research purpose.

• Evaluating methods for characterizing fire pixels using RGB, HSV, YCBCR, and CIE Lab color models.

• Developing an improved method for characterizing fire pixels using color models.

• Evaluating video smoke detection methods based on fuzzy modeling and Weber contrast analysis.

• Implementing a new facial recognition method using supervised learning.

• Testing and evaluating the performance of the new facial recognition method.

• Designing an intelligent video application for monitoring potential risks in the protected environment.

1.1 General Methodology

Since the purpose of this thesis combined two research areas, both in audio and video, a specific approach was considered for each study conducted. It was assumed that there should be current literature presenting various methods used in both audio and video, as elaborated in Chapters 1 and 2. In this regard, an investigation was carried out to identify scientific works proposing methods and techniques used in both audio and video, as described in Chapters 1 and 2. For the selected methods, tests and experiments were defined, along with a specific working method as presented in Chapters 4 and 5. Following the defined experiments, their results were analyzed and proposed for use where deemed useful. Additionally, improved methods compared to the initial ones were developed, as well as new methods, as outlined in Chapters 5 and 6.

2 Audio Techniques for Intruder Detection

The reduced complexity of audio monitoring systems holds significant interest for all manufacturers of such systems in the current market. This attribute aims to optimize both implementation costs and the response speed of the detection algorithm.

An important aspect for designing such a system is the prior knowledge of the wildlife present in the monitored natural environment. Another aspect is the selection of an algorithm suitable for the application's purpose.

In this regard, it is assumed that the area to be monitored is an open space where various animals, birds, as well as motorized vehicles, machinery, and authorized individuals may be present. The sounds to be subjected to detection need to be extracted from the background noise. Background noise varies from one natural environment to another and typically consists of sounds from birds, animals, wind, rain, vehicles, and human voices.

As a basic schematic of such a system, several blocks with fundamental functions can be identified (see Figure 1.1), such as:

• Acquiring and processing acoustic signals

• Intruder modeling

• Intruder detection

• Alarm

• Alarm response.

A series of works conducted at the Technical University of Cluj-Napoca, investigating such possibilities, can be found in [12], [13], [14], [15], [16], [17], and [18]. The authors in [12] explore the performance of the TESPAR (Time Encoded Signal Processing and Recognition) algorithm, while the authors in [13] present a method that utilizes artificial neural networks.

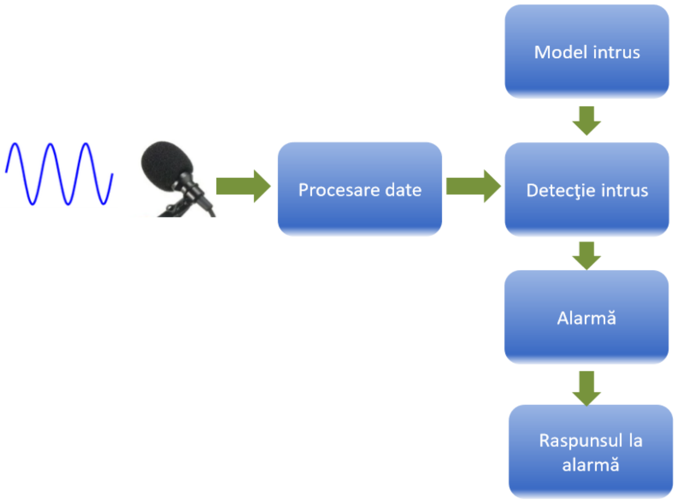

Figura 2.1 A block diagram for an acoustic intruder monitoring system [114]

The method in [14] proposes the use of Mel-Cepstral coefficients for defining the feature set, with Gaussian models and support vector machines as classifiers. [15] presents a series of issues that can arise when designing such audio systems. In this regard, the authors in [16] introduce a range of algorithms used in a sensor-based implementation, along with the constraints that can emerge based on the chosen approach. [17] suggests using sparsograms to define the feature set for a specific sound of interest, with TESPAR, Gaussian GMM models, and Support Vector Machines (SVM) as classifiers.

In [18], the authors propose the use of a network of acoustic sensors for audio signal acquisition, while they recommend the TESPAR method for processing and classification. For all the mentioned approaches, classification rates exceeding 90% were achieved. The processing speed varies significantly depending on the classifier used, owing to the algorithm's complexity.

In the case of the neural network classifier [13], the classification rate exceeds 98%, but the processing time is quite high. This last aspect suggests that this type of classifier may not be suitable for developing a system where a simpler implementation and acceptable accuracy in sound classification are desired.

The idea of experimentally identifying a series of classifiers in the audio context that could be used in the development of audio monitoring systems in protected environments remains open. In this regard, based on the approach used in [19], where experiments in WEKA were defined to evaluate classifiers in bird sound detection for specific species, we have devised our own methodology within the context of interest. This study is presented in Chapter 4.

3 Video Techniques for Fire Detection

Fires represent a significant and alarming hazard, with a negative impact on both the economy and the environment. The ability to detect a fire in its early stages is a contemporary challenge. Traditional smoke detectors have several disadvantages, such as delayed response of heat sensors and false alarms due to the high sensitivity of smoke detectors. Several techniques used for smoke/fire detection are based on histogram determination, temporal analysis of objects of interest, and event rule determination [49].

Histogram-based techniques essentially involve calculating the captured image and then using the information from it to detect the object of interest. Typically, a decision is made by comparing the calculated histogram with reference histograms for smoke/fire scenes. Other approaches use statistical calculations such as mean and standard deviation to determine the probability of the existence of the event of interest in the current scene [50]. In the category of temporal techniques, the approach is to use differences between image sequences to generate the model for smoke/fire objects [51][52][53]. The number of image sequences used to calculate differences varies depending on the approach used. Once the differences between image sequences are calculated, various methods are employed to segment and identify smoke/fire objects [54][55][56][57].

In some approaches, a number of statistical parameters are calculated from the identified differences, and based on this information, the presence of objects of interest is determined [59]. Color information from the generated models is also used to classify and identify objects of interest in the image. Such techniques are called color matching filters [57][58].

For the segmentation of fire features, color characteristics are used to segment flame regions. The most commonly used color model for flame recognition in image processing is HIS [60], as it is suitable for providing a more human-like approach to color perception. Implementations have also been attempted using RGB [61][62], HSV [62][63], CIE XYZ [60], CIE LUV [60][64], YCbCr [61], CIE L*a*b*, L*a*b*, as well as combinations of these models to reduce the detection of similar objects and thereby decrease false alarms [65].

4 Study 1 – Classifiers in the Audio Context Using LPC/LPCC Prediction Coefficients

4.1 Introduction

A work environment facilitating the development of a series of experiments based on LPC coefficients is WEKA. As presented in Chapter 2, a set of classifiers available in WEKA was selected to evaluate their performance. The results obtained from these experiments aim to identify the most effective classifiers and the optimal number of LPC coefficients to be used in the characterization phase.

4.2 Hypothesis/Objectives

One of the objectives of this investigation is to classify classes of sounds that could be present in protected wilderness environments and detect sounds associated with potential subjects classified as intruders. In this regard, a sound database containing five classes of sounds related to birds, chainsaws, firearms, humans, and tractors was defined. For each sound class, specific features were extracted using linear prediction coefficients (LPC).

Another objective of this investigation was to determine an ideal number of LPC coefficients to achieve an optimal balance between processing speed, resource utilization, and high classification accuracy of the algorithm for certain sound classes.

4.3 Conclusions

After a brief analysis of the previously presented results, a series of experiments were conducted with the objective of reducing the number of LPC coefficients to enhance processing speed during the classification phase, and to evaluate the effectiveness of LPC cepstral coefficients when used with various classifiers. Based on the obtained results, it was observed that the utilization of LPC cepstral coefficients generally leads to improved classification accuracy. Furthermore, it effectively addresses classification errors introduced by gunshot sounds.

As a general observation derived from the aforementioned testing scenarios, it was evident that classification errors occurred more frequently when the database included the class of gunshot sounds, as discussed in the previous chapter. In light of these findings, it can be concluded that the utilization of LPC cepstral coefficients represents a significantly superior option for characterizing audio signals within the specified range, particularly in applications for monitoring intruders in protected wilderness environments.

5 Study 2 – Video Methods for Fire Detection

5.1 Introduction

For this study, it was considered that based on the video techniques analyzed in Chapter 2, a series of experiments could be defined to verify their performance. Accordingly, a database consisting of both images and video sequences containing fire situations was defined. The method used and the results obtained are presented in the following sections.

5.2 Hypothesis/Objectives

In this phase, a series of articles proposing methods and techniques for smoke/fire detection were identified. Details about each method can be found in Chapter 2. These methods were analyzed and implemented in MATLAB. An evaluation of their performance was conducted, focusing on both speed and accuracy in detecting objects of interest.

One objective was to compare the selected methods to identify the most accurate and stable method with the lowest error rate. Additionally, another goal of this study was to improve the methods that allow for this.

5.3 Conclusions

Based on the results highlighted in the previous chapter, it can be concluded that a reliable smoke/fire detection algorithm with a low error rate involves a sufficiently complex implementation to ensure correct and timely detection.

A proposal would be to conduct an in-depth analysis of methods M3, M4, and M5, and through fusion, develop a new method that will detect both smoke and fire within the same framework (involving the analysis of smoke pixels based on the static and dynamic components of smoke using FUZZY logic). This way, a fusion will be achieved between the RGB, YCbCr, HSV color systems for fire detection and the FUZZY system for smoke detection, and this implementation will be presented in the next chapter.

6 Study 3 – RiskCam Application Concept

6.1 Introduction

In today's world, the development of monitoring systems for high-security areas is of great interest. These systems should be capable of detecting unexpected events such as fires, thefts, poaching, and even unauthorized entry into private spaces. It can be observed that these issues are increasingly highlighted through media and in the virtual environment by organizations dedicated to environmental protection (e.g., AgentGreen, Greenpeace), among others. There is a direct impact when we talk about fires and illegal deforestation, leading to climate change and the extinction of wildlife in the affected areas. Therefore, it disrupts the ecosystem.

Technological advancement has increased the potential for advanced video analysis and the extraction of valuable information, such as motion detection and shape recognition. It can be said that the development of monitoring applications capable of detecting potential risks in a protected area tends to vary in order to effectively address the mentioned issues.

6.2 Hypothesis/Objectives

Intelligent video devices are capable of successfully detecting specific situations from a distance in a video clip. However, when it comes to surveilling significantly large areas that require monitoring, the costs of such a project can be substantial.

Let's assume we have such a situation, and we want to propose a solution that significantly reduces costs. Therefore, we aim to develop a solution for monitoring a large, high-security property. The property consists of more than twenty hectares of forested land. The area contains protected wildlife species, animal shelters, and multiple material storage facilities. The owner has provided us with a map of the area highlighting the most important targets that require surveillance, as shown in Figure 6.1.

Figura 6.1 Possible map with targets requiring surveillance [117]

The requirement would be to create a video solution capable of performing the following tasks:

• Access to the property must be done through facial authentication.

• Fire situations must be detected at all times.

• Intruder detection based on abnormal movements in shelter and storage areas (possible theft).

• Email notification with the identified event and zone (e.g., "Fire detected in Zone C1!").

After conducting a cost analysis, which would involve acquiring the necessary devices for implementing the client's request, we obtained the results in Table 6.1. In the table, we identified the smart camera capable of performing one of the requested tasks. Although there are many manufacturers in this industry, we attempted to select the best compromise between price and the project's goal by choosing suppliers [99], [100], and [101] (as of 2022).

Produs	Preț (euro)	Nr. de camere	Cost Total	Funcționalitate
Wifi IP Light Blub camera	122	4	488	Autentificare facială
SG7200WF Extreme life Wifi night vision detector	324	25	8100	Detecție foc și fum
Aeon OT-1080p POE wifi camera	122	10	1220	Detecție mișcare anormală

Tabel 6.1 Analiza de piață și costurile estimate per dispozitiv smart [117]

According to Table 6.1, the final cost amount would reach 9808 euros (in 2022). If we scale this amount for a project that requires even more cameras, we can imagine how the costs would increase significantly. A simulation of how costs would increase in relation to the number of video cameras is presented in Figure 6.2.

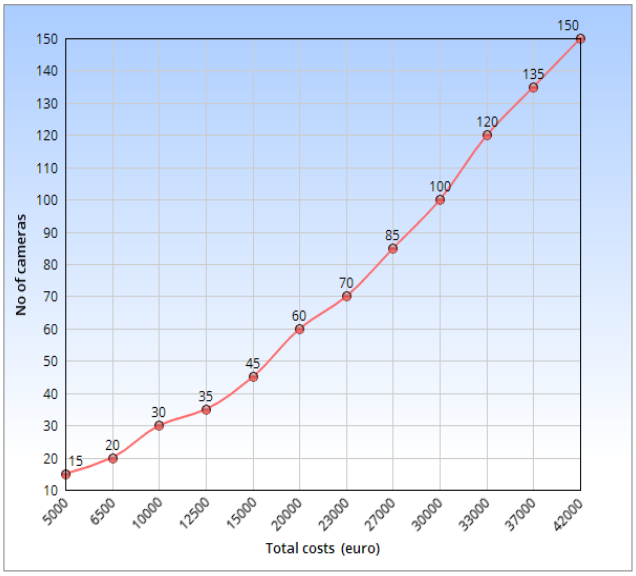

Figura 6.2 Cost Simulation based on sistem complexity [117]

Considering the number of dedicated cameras that would need to be used for implementing a video solution that meets the client's requirements, a question arises: why not use a single type of WiFi camera and develop an application with configurable video detection algorithms for various tasks? If we apply our case study and assume we choose the AeO OT - 1080p POE Wi-Fi camera for monitoring each defined objective, we can reevaluate the initial estimates. In this regard, we achieve a total cost of 3538 euros (in 2022), which means a cost reduction of nearly 64%. Additionally, the practical aspect of remote monitoring is another reason for the need to develop such an application. A significant cost reduction can be observed, which motivated us to develop a new approach for a configurable integrated intelligent video application, as will be presented in the next subsection. We named the application RiskCam [117].

6.3 Conclusions

Based on the results highlighted in this chapter, it can be concluded that the proposed algorithms for implementing the RiskCam application represent a good solution. A future proposal would be to develop more test scenarios for the module that detects smoke regions, as well as attempting to access a live camera feed from a natural reserve, initially for negative tests (fire/smoke regions may be absent).

7 Concluzii finale

A summary of the motivation behind the need for this research topic in this thesis is as follows:

According to a study conducted in 2015 by GreenPeace Romania [112], Romania was losing three hectares of forest per hour. In 2016, Romania's forests covered 6.90 million hectares [112][113]. According to Greenpeace Romania, between 2000 and 2011, approximately 280,108 hectares of forests were deforested or degraded, and the rate of forest degradation remained the same from 2012 to 2014. An estimated area of 361,068 hectares was affected.

Regarding the evolution of illegal logging cases, authorities recorded over 30 cases of illegal logging per day in 2009-2011, over 50 cases in 2012, approximately 62 cases in 2013-2014, and around 96 cases in 2015 [113]. In 2022, it was reported that six hectares were being deforested every hour [122].

As an illustration, a map of illegal deforestation activities for the year 2015, created by Greenpeace Romania, was presented [113].

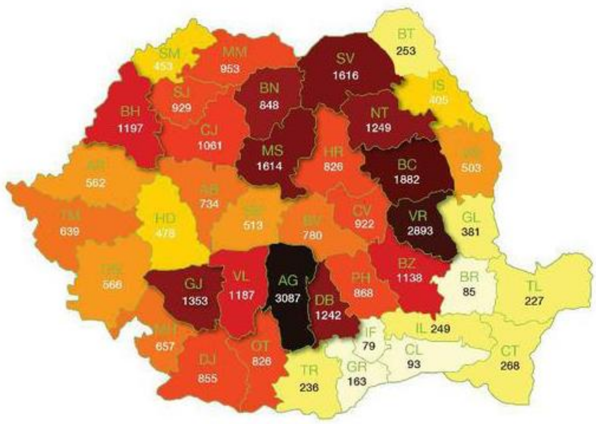

Figura 7.1 Harta pe zone afectate de defrișări ilegale [113]

At the present moment, the situation is not more optimistic in this regard. The only solution is to support the authorities in any way possible to help reduce deforestation at the national level, as proposed in this thesis.

Considering the results obtained and presented in the previous chapters, the following conclusions can be drawn:

1. For characterizing audio signals in the mentioned range, experiments support the idea that the method of LPC cepstral coefficients would be a significantly superior option compared to the classical LPC method.
2. According to the presented experimental results, the accuracy of classifiers is influenced by the number of LPC coefficients used in the classification process: a small number of coefficients increases classification errors, while an optimal number of coefficients decreases classification errors.
3. Among the set of 22 classification algorithms evaluated in this work for the detection of specific sound classes, it was observed that the most performant and stable classifier concerning the number of LPC coefficients used is the LMT classifier (with a minimum of 10 coefficients), achieving an accuracy of 95%.
4. The list of classifiers with low complexity and acceptable accuracy exceeding 95% is composed of: IBK, J48, and Attribute Selected Classifier.
5. For the IBK classifier, the optimal number of LPC coefficients would be 16, for J48, it would be 46, and for the Attribute Selected Classifier, it would be 36.
6. The method of LPC cepstral coefficients effectively addresses the problem of gunshot sound classification because the broad spectrum of the impulse signal is absorbed when calculating the smoothed autoregressive power spectrum, providing a more precise characterization of this type of signal.
7. For the characterization of fire pixels, five methods were evaluated, of which the following three exhibited satisfactory classification accuracy: RGB and HSV feature classifier, RGB and YCbCr classifier, and the fusion of L*a*b*, YCbCr classifier and iterative k-means clustering.
8. For analyzing smoke pixels, a method that refers to both static and dynamic components of smoke using a fuzzy interference system

was employed, yielding satisfactory results and offering the possibility of detecting a fire even in its incipient stage, surpassing the limitations of standard smoke detectors, which are conditioned by a certain distance from the smoke source.

9. The need to design an integrated intelligent video application that analyzes and identifies potential risks arose from a case study that promised an estimated 64% reduction in the total system cost, along with the practical and efficient aspect of remote monitoring.
10. Abnormal motion detection in a video sequence through the technique of object segmentation in motion by extracting the background element has proven to be a satisfactory approach and an integrated component of the intelligent application.
11. The novel method of facial authentication using a simplified neural network of fuzzy adaptive theory has shown promising results based on experimental results and can be used both as a standalone component and as an integrated component within the RiskCam application.
12. The fusion of cosine features and binary models within the facial recognition method, as a step in feature extraction, has proven to be robust and efficient in creating the final feature vector.
13. Based on the results and scientific observations highlighted in this work, it can be concluded that the proposed algorithms for implementing the RiskCam application represent a sound solution, as each algorithm achieves the final goal of the application.
14. As future developments, several test scenarios can be elaborated, involving access to live cameras and testing the application under diverse conditions, subsequently identifying the less performing points and thus developing methods for their improvement."

The author's contribution is primarily found in Chapters 3, 4, 5, and 6. As an approach, the study was divided into two research directions: one focused on audio techniques for classifying certain types of sounds, and the other on video techniques for detecting events that can be extracted from video frames.

New methods for characterizing sounds were analyzed and proposed, while a series of existing methods were evaluated based on test data containing sounds of interest for this research. The list of sound classes of interest in the test data includes gunshot sounds, bird sounds, tractor sounds, human voice sounds, and chainsaw sounds.

On the video side, several existing methods for analyzing specific types of objects in video frames were identified and evaluated. New facial recognition methods were also developed, along with the concept of an intelligent application that integrates multiple processing tasks to achieve advanced monitoring, providing authorities with a quick response time. Examples include situations like fires, illegal trespassing on private property, and incorrect facial authentication.

The first phase of the study was dedicated to Chapter 4 (on the audio part) and involved determining test data used to define experiments for evaluating the performance of a number of classifiers available in the WEKA environment. Two methods for creating the feature vector (LPC and LPC cepstral) were also proposed and compared in terms of audio detection system processing speed. From the experiments, it can be concluded that the LPC cepstral coefficient method is a significantly superior option in feature vector creation due to its performance stability across diverse sounds. The results were published in [114][115][118].

In Chapter 5, a series of algorithms in the video domain for fire detection were analyzed. This required the segmentation and identification of fire pixels and smoke pixels in images. Test data were defined for evaluating the methods mentioned in this chapter, and based on the results obtained, the most suitable algorithms were proposed for use in an integrated system, as presented in Chapter 6. Given the desire for a fast response time, the option with satisfactory results was chosen to maintain a low complexity in the processing method. The results were published in [119][121].

Unlike the previous chapter, Chapter 6 proposes a concept of an intelligent video application as an integrated system capable of detecting various events from video frames, involving different algorithms applied to the same video frames. This concept arose as a result of a case study presented at the beginning of the same chapter, which proposed advanced monitoring of a highly structured natural reserve area. The results were published in [117].

For each functionality, a processing method capable of performing that type of detection was proposed. For the fire detection functionality, an improved method was proposed regarding the segmentation of fire pixels. To create a color model for fire pixel segmentation, one or more color spaces can be considered, to which a set of rules characteristic of fire pixels is applied. An implementation that uses a fusion of RGB (red-green-blue), YCbCr (luminance-blue-difference from green-red-difference from green), and HSV (hue-saturation-value) was proposed. The final kernel is obtained by intersecting the results of the rules for each color space. Based on the results obtained, it was concluded that the new method is an improvement compared to what has been implemented so far. The results were published in [117][121].

For the facial authentication functionality, a promising new method was developed based on supervised incremental learning and fusion between feature extraction based on LBP (local binary patterns) and DCT (discrete cosine transform). For abnormal motion detection, an existing classic method was used but subjected to performance tests as presented in the same chapter. Experiments were conducted, and specific datasets were defined for each situation, yielding the mentioned conclusions.

Based on the results highlighted in this work, it can be concluded that the algorithms proposed for the implementation of the RiskCam application represent a promising solution. The results were published in [117][121] and [C1]. Chapters 4 and 5 contain a multitude of conclusions that have helped make strategic decisions for the step-by-step development of this work, providing a solid starting point for anyone interested in technical discussions in the same category.

This doctoral thesis has successfully proposed, from a different perspective, practical solutions to current needs regarding the conservation and protection of wildlife and the environment by adopting an intuitive and experimental research methodology, also supported by theoretical explanations for the obtained results.

List of publications

Conference Papers:

[C1] E. R. Buhuş, L. Grama, "Leveraging Profiling for Facial Recognition Optimization," 2023 4th International Conference on Advances in Electrical Engineering and Computer Applications, Dalian, China

[C2] E. R. Buhuş, L. Grama and C. Şerbu, "A facial recognition application based on incremental supervised learning," 2017 13th IEEE International Conference on Intelligent Computer Communication and Processing (ICCP), Cluj-Napoca, 2017, pp. 279-286, doi: 10.1109/ICCP.2017.8117017.

[C3] E. R. Buhuş, L. Grama and C. Rusu, "Several classifiers for intruder detection applications," 2017 International Conference on Speech Technology and Human-Computer Dialogue (SpeD), Bucharest, 2017, pp. 1-6, doi: 10.1109/SPED.2017.7990432.

[C4] L. Grama, E. R. Buhuş and C. Rusu, "Acoustic classification using linear predictive coding for wildlife detection systems," 2017 International Symposium on Signals, Circuits and Systems (ISSCS), Iasi, 2017, pp. 1-4, doi: 10.1109/ISSCS.2017.8034944.

[C5] E. R. Buhus, A. Lazar and A. Tavares, "A System-On-Chip Approach in Designing a Dedicated RISC Microcontroller Unit Using the Field-Programmable Gate Array," 2010 Fifth International Conference on Systems, Menuires, 2010, pp. 114-119, doi: 10.1109/ICONS.2010.40.

Scientific Journal Articles:

[R1] Elena Roxana BUHUS, Lacrimioara GRAMA, Alexandra DOBRE, “Riskcam: A Smart Video Surveillance Application Design Concept”, in Acta Technica Napocensis- Electronics and Telecommunications. Cluj-Napoca, Romania, 2018, vol. 59, nr.3, pp1-7, ISSN: 1221-6542.

[R2] Elena Roxana BUHUS, Lacrimioara GRAMA, Corneliu RUSU, “Linear Predictive Cepstral Coefficients In Wildlife Detection Systems”, in Acta Technica Napocensis- Electronics and Telecommunications. Cluj-Napoca, Romania, 2017, vol.58, nr.4, pp.1-5, ISSN: 1221-6542.

[R3] Elena Roxana BUHUS, Anca APĂTEAN , "State of the Art: Methods for Video Based Fire Detection Via Stationary Camera", in Acta Technica Napocensis- Electronics and Telecommunications. Cluj-Napoca, Romania, 2017, vol.58, nr.1, pp.6-11 , ISSN: 1221-6542.

[R4] Elena Roxana BUHUS, Daniel TIMIS, Anca APATEAN, "Automatic Parking Access Using OPENALPR On Raspberry PI3", in Acta Technica Napocensis- Electronics and Telecommunications. Cluj-Napoca, Romania, 2016, vol. 57, nr.3, pp. 10-15, ISSN: 1221-6542.

Doctoral Reports:

[RC1] Elena Roxana BUHUS, "Video Techniques for Fire and Intruder Detection – Analysis and Comparisons" – First scientific research report, scientific coordinator Prof. Dr. Ing. Corneliu Rusu.

[RC2] Elena Roxana BUHUS, "Audio Techniques for Sound Detection and Classification – Analysis and Comparisons" – Second scientific research report, scientific coordinator Prof. Dr. Ing. Corneliu Rusu.

[RC3] Elena Roxana BUHUS, "Implementation and Development of Audio-Video Techniques for Fire, Intruder, and Chainsaw Sound Detection" – Third scientific research report, scientific coordinator Prof. Dr. Ing. Corneliu Rusu.

[RC4] Elena Roxana BUHUS, "Report with the Results Obtained in the Implementation of Audio-Video Techniques for Fire, Intruder, and Anomalous Sound Detection" – Fourth scientific research report, scientific coordinator Prof. Dr. Ing. Corneliu Rusu.

RESEARCH AWARDS

[1] Second Prize at SSET 2018 – Cluj-Napoca, Master/Doctoral Presentation Section, with the paper "Histograms and Supervised Learning for Facial Recognition Applications," Elena Roxana Buhus, L. Grama, C. Serbu

[2] First Prize at SSET 2017– Cluj-Napoca, Master/Doctoral Presentation Section, with the paper "Linear Predictive Cepstral Coefficients in Wildlife Detection Systems," Elena Roxana Buhus, L. Grama

[3] Best Paper Award at the ICONS 2010 conference, The fifth International Conference on Systems, France, Embedded systems and systems-on-the-chip section, with the paper "A System-On-Chip Approach in Designing a Dedicated RISC Microcontroller Unit Using the Field-Programmable Gate Array," Elena Roxana Buhus, Alexandru Lazar, and Adriano Tavares: "https://www.iaria.org/conferences2010/AwardsICONS10.html"