

**MULTIMODAL DETECTION AND PREEMPTION SYSTEM FOR
EMERGENCY VEHICLES USING COMPUTER VISION AND ACOUSTIC
ANALYSIS**

By

Indra Kristiawan
22452012

MASTER'S DEGREE
in

MECHANICAL ENGINEERING - MECHATRONIC
FACULTY OF ENGINEERING & INFORMATION TECHNOLOGY



SWISS GERMAN UNIVERSITY
The Prominence Tower
Jalan Jalur Sutera Barat No. 15, Alam Sutera
Tangerang, Banten 15143 - Indonesia
Revision after the Thesis Defense on 15 August 2025

STATEMENT BY THE AUTHOR

I hereby declare that this submission is my own work and to the best of my knowledge, it contains no material previously published or written by another person, nor material which to a substantial extent has been accepted for the award of any other degree or diploma at any educational institution, except where due acknowledgement is made in the thesis.



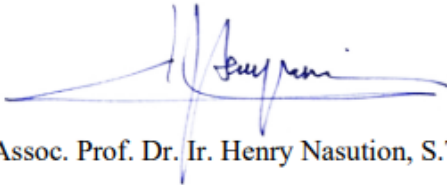
Indra Kristiawan

Student



Date

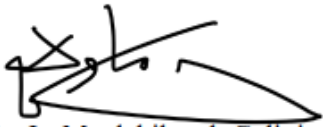
Approved by:



Assoc. Prof. Dr. Ir. Henry Nasution, S.T., M.T., IPP.

Thesis Advisor

Date

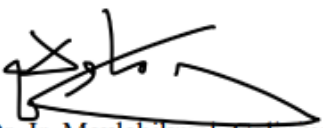


Dr. Ir. Maulahikmah Galinium, S.Kom., M.Sc.

27 August 2025

Thesis Co-Advisor

Date



Dr. Ir. Maulahikmah Galinium, S.Kom., M.Sc.

27 August 2025

Dean

Date

Indra Kristiawan

ABSTRACT

MULTIMODAL DETECTION AND PREEMPTION SYSTEM FOR EMERGENCY VEHICLES USING COMPUTER VISION AND ACOUSTIC ANALYSIS

By

Indra Kristiawan

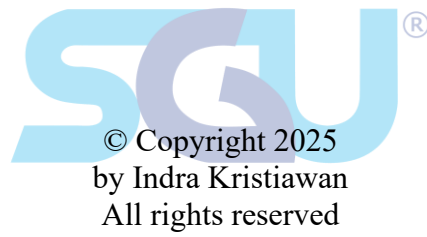
Assoc. Prof. Dr. Ir. Henry Nasution, S.T., M.T., IPP.

Dr. Ir. Maulahikmah Galinium, S.Kom., M.Sc.

SWISS GERMAN UNIVERSITY

Emergency vehicle detection in intelligent traffic systems often suffers from high false positive rates when relying solely on vision-based approaches. To address this limitation, a multimodal framework combining computer vision and acoustic analysis is proposed. The system integrates a YOLOv5-based vision module deployed on Jetson Nano with an ESP32-S3 acoustic verifier extracting Mel-Filterbank Energy (MFE) features, with detections confirmed only when both modalities agree through an AND-gate fusion strategy. Over 11 days, 1,380 timestamped traffic events were recorded under operational conditions. The vision module alone produced 1,321 detections, of which only 421 were true positives, resulting in a preemption accuracy of 31.8%. With multimodal fusion, 61 of 70 true emergency events were successfully detected, raising accuracy to 87.1% while eliminating 900 false positives entirely. No ghost detections were observed, and no emergency events with active sirens were missed by the acoustic verifier, affirming the robustness of the fusion logic. These findings demonstrate that multimodal integration achieves 100% precision, substantially improves reliability, and enables practical edge deployment in intelligent traffic management systems.

Keywords: emergency vehicle detection, multimodal system, computer vision, acoustic analysis, false positive reduction



DEDICATION

This thesis is dedicated to my beloved family, whose unwavering support, encouragement, and prayers have been the foundation of my journey.

To my mentors and peers, thank you for your guidance, inspiration, and collaboration throughout this academic pursuit.

And finally, to the future of intelligent traffic management systems and innovation—may this work contribute, even in a small way, to the advancement of mechatronics and technology for the betterment of society.



ACKNOWLEDGEMENTS

First and foremost, I would like to express my deepest gratitude to Dr. Ir. Henry Nasution, M.T., IPP, my thesis advisor, for his invaluable guidance, constructive feedback, and continuous encouragement throughout the course of this research. His expertise and dedication have been instrumental in shaping the direction and outcome of this work.

I would also like to extend my sincere appreciation to my co-advisor, Dr. Maulahikmah Galinium, S.Kom, M.Sc, for her insightful suggestions, technical advice, and consistent support that enriched the quality of this thesis.

My heartfelt thanks go to all the faculty members and staff of the Mechatronics Department at Swiss German University, for providing a stimulating academic environment and the resources necessary for my research.

I am deeply grateful to my family and friends for their unwavering love, patience, and support during this academic journey. Their belief in me has been a constant source of motivation.

Last but not least, I would like to thank all individuals and institutions who have contributed, directly or indirectly, to the completion of this thesis.

TABLE OF CONTENTS

	Page
STATEMENT BY THE AUTHOR.....	2
ABSTRACT.....	3
DEDICATION.....	5
ACKNOWLEDGEMENTS.....	6
TABLE OF CONTENTS.....	7
LIST OF FIGURES.....	9
LIST OF TABLES.....	10
CHAPTER 1 - INTRODUCTION.....	11
1.1 Background.....	11
1.2 Research Problems.....	14
1.3 Research Objectives.....	15
1.4 Significance of Study.....	16
1.5 Research Questions.....	17
1.6 Hypothesis.....	18
1.7 Research Scope.....	18
CHAPTER 2 - LITERATURE REVIEW.....	20
2.1 Evolution of the Adaptive Traffic Signal Controller (ATSC).....	20
2.2 Computer Vision in ATSC for Emergency Vehicle Detection.....	20
2.3 Audio-Based Detection for Emergency Vehicles.....	21
2.4 Multimodal Approaches: Integrating Vision and Audio.....	22
2.5 Addressing the Research Gap: Proposed Framework.....	22
2.6 Previous Studies Summary.....	23
CHAPTER 3 – RESEARCH METHODS.....	25
3.1 Research Framework.....	25
3.2 Root Cause Analysis.....	26
3.3 System Architecture Design.....	31
CHAPTER 4 – RESULTS AND DISCUSSIONS.....	38
4.1 Research Overview.....	38
4.2 Existing Design Improvement.....	39

4.3	Final Design.....	47
4.4	Computer Vision Ambulance Preemption.....	50
4.5	Acoustic as Confirmatory for Preemption.....	51
4.6	DSP Feature Extraction for Siren Verification.....	51
4.7	Comparative Study of Feature Extractions for Acoustic Performance Improvement	56
4.8	DSP Feature Extraction for Siren Verification.....	61
4.9	Implementation and Deployment	63
4.10	Advanced Deployment and Analysis.....	67
CHAPTER 5 – CONCLUSIONS AND RECCOMENDATIONS		77
5.1	Conclusions	77
5.2	Recommendations	78
GLOSSARY		81
REFERENCES		83
CURRICULUM VITAE		86
SUPPLEMENTARY MATERIAL.....		89



LIST OF FIGURES

Figures	Page
Figure 1.1 Existing System with False Positive.....	12
Figure 1.2 Existing System Performances	13
Figure 3.1 Research Framework	26
Figure 3.2 Fishbone Root Cause Analysis	26
Figure 3.3 System Architecture Design	33
Figure 4.1 Existing System Hardware Schematic.....	40
Figure 4.2 Vision-Based System Flowchart	41
Figure 4.3 Multimodal System with AND Gate Fusion Architecture	42
Figure 4.4 False Positive Cases	43
Figure 4.5 Residual false-positive examples after the fully optimised visual pipeline (v4q). Frames illustrate (a) reflective van look-alike, (b) night-time glare, (c) long-range low-resolution ambiguity, and (d) high-contrast clutter occlusion. Each still is mis-l.....	46
Figure 4.6 Full node-level schematic (camera, Jetson, ESP32, signal controller).....	48
Figure 4.7 DSP Results Visualization.....	55
Figure 4.8 MFCC Data Explorations	58
Figure 4.9 Log-Mel Spectrogram Data Explorer	59
Figure 4.10 MFE Data Explorer	59
Figure 4.11 Field Documentation Experiment.....	64
Figure 4.12 Field deployment documentation showing sensor enclosure positioned to face incoming ambulance trajectory	65
Figure 4.13 Console Trace while Ambulance Siren Detected	66
Figure 5.1 Future Research and Product Roadmap.....	79

LIST OF TABLES

Table	Page
Table 2.1 Comparative Study of Existing Emergency Vehicle Detection Methods.....	23
Table 4.1 Ablation Study of Vision-Only Pipeline Iterations.....	45
Table 4.2 Summarises acoustic subsystem specifications.	49
Table 4.3 Stages Justification	52
Table 4.4 Summary Comparison of FE	57
Table 4.5 Summary of Detection Events (11 Days, 07:00–21:00)	69
Table 4.6 Comprehensive Performance Results (1,380 Detection Events)	70
Table 4.7 Comparative Analysis of Emergency Vehicle Detection Systems	74

