UNIVERSIDAD CARLOS III DE MADRID

TESIS DOCTORAL

DATA FUSION ARCHITECTURE FOR INTELLIGENT VEHICLES

Autor:

Fernando García Fernández

Directores:

Dr. José María Armingol Moreno

Dr. Arturo de la Escalera Hueso

**DEPARTAMENTO DE
INGENIERÍA DE SISTEMAS Y AUTOMÁTICA**

Leganés, Noviembre 2012

TESIS DOCTORAL

DATA FUSION ARCHITECTURE FOR INTELLIGENT VEHICLES

Autor: **Fernando García Fernández**

Directores: **Dr. José María Armingol Moreno**

Dr. Arturo de la Escalera Hueso

Firma del Tribunal Calificador:

Firma

Presidente:

Secretario:

Vocal:

Calificación:

Leganés, de de

This thesis is submitted to the Departamento de Ingeniería de Sistemas y Automática of the Escuela Politécnica Superior of the Universidad Carlos III de Madrid, for the degree of Doctor of Philosophy. This thesis is entirely my own work, and, except where otherwise indicated, describes my own research.

Dedicado a mí hermano Pauli, con todo mi corazón

Dedicated to my brother Pauli, with all my heart

“Escoge un trabajo que te guste, y nunca tendrás que trabajar ni un solo día de tu vida.”

Confucio.

“Choose a job you love and you will never have to work a day in your life.”

Confucius

ACKNOLEDGEMENTS

Jamás me pude imaginar, hace unos años, cuando acudía a mi entrevista con Arturo, que esa inocente entrevista de trabajo me abriría las puertas de lo que ha sido una de las etapas más interesantes de mi vida. Una de las cosas más agradables de este trabajo es la cantidad de gente que conoces, gente que te aporta, en mayor o menor medida, experiencias y vivencias que me han ayudado a crecer, como persona, investigador y docente. A todas esas personas que me han ayudado a llegar hasta este punto van dedicadas estas líneas. A todos, gracias por haber compartido conmigo tantos momentos increíbles.

En primer lugar, tengo que agradecer a mis directores de tesis, José María y Arturo, su dedicación y apoyo incondicional. Gracias, sobre todo por permitirme trabajar en lo que más me gusta.

A mis padres, sin ellos esto no sería posible. Gracias por estar siempre ahí, por ser mi apoyo, por aguantarme en los momentos malos y por disfrutar los buenos. Gracias por haberme ayudado a llegar hasta aquí y haberme permitido conseguir este proyecto, pero sobre todo, gracias por ser como sois.

A todos esos amigos y compañeros que durante años han sido tan importantes en mi vida:

A Antonio, le conozco desde hace tanto, que ni lo recordamos y por supuesto a Sara, ¡muchas gracias amigos!

A mis compañeros de laboratorio, a los que están, y a los que se fueron, especialmente a Gustavo, María, Gema y Juan Pablo, ellos llegaron a ser, no solo compañeros, sino buenos amigos, con los que he compartido frustraciones y alegrías. Gracias chicos, habéis grabado vuestro recuerdo en este doctorado.

También a Raúl y los demás miembros del GUTI, con vosotros descubrí la investigación y puedo asegurar que disfruté y disfruto de ella con vosotros.

A mis amigos de Colombia, Carlos, Fercho, Lukas y Nati. Ellos han sido una parte importante de estos años, me han abierto un mundo totalmente nuevo y con ellos he creado un vínculo que espero dure muchos años. ¡Los quiero, los adoro y los extraño!

A mis alumnos, a todos, pero sobre todo a los tres gamberros que fueron el principio de mi etapa docente y ahora se han convertido en compañeros y amigos: Chechu, Gerar y JJ.

A todos los que han hecho de los congresos y las estancias aventuras únicas, miembros de Vislab en Italia y a Jesús y sus chicos de Colmenarejo.

A mis amigos del pueblo, crecimos juntos y luchamos cada año por mantener lo que tantos años nos costó crear.

To my beloved Marlena, it is amazing feeling somebody that close despite the distance. You make me so happy.

A mi familia, sentirles cerca, su apoyo y ánimo incondicional me hacen más fuerte, ayudándome en los momentos más duros. Especialmente a mis primos Alberto y Jorge, y a Nani, los mejores ayudantes que un investigador puede tener. Y a mis abuelas, os quiero, sois una parte importante de mi vida y además representáis la parte viva de los abuelos que tanto quiero y querré siempre.

Finalmente, esta tesis está dedicada a Pauli, mi hermano, y por supuesto a Carmen y Lucía. Creo que sobran las palabras y las descripciones, quien me conoce y ha vivido cerca de mí, sabe lo importante que sois en mi vida y lo que valoro teneros cerca cada día, os quiero con todo mi corazón.

RESUMEN

Los accidentes de tráfico son un grave problema social y económico, cada año el coste tanto en vidas humanas como económico es incontable, por lo que cualquier acción que conlleve la reducción o eliminación de esta lacra es importante. Durante los últimos años se han hecho avances para mitigar el número de accidentes y reducir sus consecuencias. Estos esfuerzos han dado sus frutos, reduciendo el número de accidentes y sus víctimas. Sin embargo el número de heridos y muertos en accidentes de este tipo es aún muy alto, por lo que no hay que rebajar los esfuerzos encaminados a hacer desaparecer tan importante problema.

Los recientes avances en tecnologías de la información han permitido la creación de sistemas de ayuda a la conducción cada vez más complejos, capaces de ayudar e incluso sustituir al conductor, permitiendo una conducción más segura y eficiente. Pero estos complejos sistemas requieren de los sensores más fiables, capaces de permitir reconstruir el entorno, identificar los distintos objetos que se encuentran en él e identificar los potenciales peligros. Los sensores disponibles en la actualidad han demostrado ser insuficientes para tan ardua tarea, debido a los enormes requerimientos que conlleva una aplicación de seguridad en carretera. Por lo tanto, combinar los diferentes sensores disponibles se antoja necesario para llegar a los niveles de eficiencia y confianza que requieren este tipo de aplicaciones. De esta forma, las limitaciones de cada sensor pueden ser superadas, gracias al uso combinado de los diferentes sensores, cada uno de ellos proporcionando información que complementa la obtenida por otros sistemas. Este tipo de aplicaciones se denomina aplicaciones de Fusión Sensorial.

El presente trabajo busca aportar soluciones en el entorno de los vehículos inteligentes, mediante técnicas de fusión sensorial, a clásicos problemas relacionados con la seguridad vial. Se buscará combinar diferentes sensores y otras fuentes de información, para obtener un sistema fiable, capaz de satisfacer las exigentes demandas de este tipo de aplicaciones.

Los estudios realizados y algoritmos propuestos están enmarcados en dos campos de investigación bien conocidos y populares. Los Sistemas Inteligentes de Transporte (ITS- por sus siglas en ingles- Intelligent Transportation Systems), marco en el que se centra la presente tesis, que engloba las diferentes tecnologías que durante los últimos años han permitido dotar

a los sistemas de transporte de mejoras que aumentan la seguridad y eficiencia de los sistemas de transporte tradicionales, gracias a las novedades en el campo de las tecnologías de la información. Por otro lado las técnicas de Fusión Sensorial (DF -por sus siglas en ingles- Data Fusión) engloban las diferentes técnicas y procesos necesarios para combinar diferentes fuentes de información, permitiendo mejorar las prestaciones y dando fiabilidad a los sistemas finales. La presente tesis buscará el empleo de las técnicas de Fusión Sensorial para dar solución a problemas relacionados con Sistemas Inteligentes de Transporte.

Los sensores escogidos para esta aplicación son un escáner láser y visión por computador. El primero es un sensor ampliamente conocido, que durante los últimos años ha comenzado a emplearse en el mundo de los ITS con unos excelentes resultados. El segundo de este conjunto de sensores es uno de los sistemas más empleados durante los últimos años, para dotar de cada vez más complejos y versátiles aplicaciones en el mundo de los ITS. Gracias a la visión por computador, aplicaciones tan necesarias para la seguridad como detección de señales de tráfico, líneas de la carreta, peatones, etcétera, que hace unos años parecía ciencia ficción, están cada vez más cerca.

La aplicación que se presenta pretende dar solución al problema de reconstrucción de entornos viales, identificando a los principales usuarios de la carretera (vehículos y peatones) mediante técnicas de Fusión Sensorial. La solución implementada busca dar una completa solución a todos los niveles del proceso de fusión sensorial, proveyendo de las diferentes herramientas, no solo para detectar los otros usuarios, sino para dar una estimación del peligro que cada una de estas detecciones implica. Para lograr este propósito, además de los sensores ya comentados han sido necesarias otras fuentes de información, como sensores GPS, inerciales e información contextual.

Los algoritmos presentados pretenden ser un importante paso adelante en el mundo de los Sistemas Inteligentes de Transporte, proporcionando novedosos algoritmos basados en tecnologías de Fusión Sensorial que permitirán detectar y estimar el movimiento de los peatones y vehículos de forma fiable y robusta.

Finalmente hay que remarcar que en el marco de la presente tesis, la falta de sistemas de detección e identificación de obstáculos basados en radar láser provocó la necesidad de implementar novedosos algoritmos que detectasen e identificasen, en la medida de lo posible y pese a las limitaciones de la tecnología, los diferentes obstáculos que se pueden encontrar en la carretera basándose en este sensor.

ABSTRACT

Traffic accidents are an important socio-economic problem. Every year, the cost in human lives and the economic consequences are inestimable. During the latest years, efforts to reduce or mitigate this problem have lead to a reduction in casualties. But, the death toll in road accidents is still a problem, which means that there is still much work to be done.

Recent advances in information technology have lead to more complex applications, which have the ability to help or even substitute the driver in case of hazardous situations, allowing more secure and efficient driving. But these complex systems require more trustable and accurate sensing technology that allows detecting and identifying the surrounding environment as well as identifying the different objects and users. However, the sensing technology available nowadays is insufficient itself, and thus combining the different available technologies is mandatory in order to fulfill the exigent requirements of safety road applications. In this way, the limitations of every system are overcome. More dependable and reliable information can be thus obtained. These kinds of applications are called Data Fusion (DF) applications.

The present document tries to provide a solution for the Data Fusion problem in the Intelligent Transport System (ITS) field by providing a set of techniques and algorithms that allow the combination of information from different sensors. By combining these sensors the basic performances of the classical approaches in ITS can be enhanced, satisfying the demands of safety applications.

The works presented are related with two researching fields. Intelligent Transport System is the researching field where this thesis was established. ITS tries to use the recent advances in Information Technology to increase the security and efficiency of the transport systems. Data Fusion techniques, on the other hand, try to give solution to the process related with the combination of information from different sources, enhancing the basic capacities of the systems and adding trustability to the inferences. This work attempts to use the Data Fusion algorithms and techniques to provide solution to classic ITS applications.

The sensors used in the present application include a laser scanner and computer vision. First is a well known sensor, widely used, and during more recent years have started to be applied in different ITS applications, showing advanced performance mainly related to its trustability. Second is a recent sensor in automotive applications widely used in all recent ITS advances in the last decade. Thanks to computer vision road security applications (e.g. traffic sign detection, driver monitoring, lane detection, pedestrian detection, etc.) advancements are becoming possible.

The present thesis tries to solve the environment reconstruction problem, identifying users of the roads (i.e. pedestrians and vehicles) by the use of Data Fusion techniques. The solution delivers a complete level based solution to the Data Fusion problem. It provides different tools for detecting as well as estimates the degree of danger that involve any detection. Presented algorithms represents a step forward in the ITS world, providing novel Data Fusion based algorithms that allow the detection and estimation of movement of pedestrians and vehicles in a robust and trustable way. To perform such a demanding task other information sources were needed: GPS, inertial systems and context information.

Finally, it is important to remark that in the frame of the present thesis, the lack of detection and identification techniques based in radar laser resulted in the need to research and provide more innovative approaches, based in the use of laser scanner, able to detect and identify the different actors involved in the road environment.

INDEX

RESUMEN .... **III**

ABSTRACT .... **V**

INDEX .... **VII**

INDEX OF FIGURES .... **XIII**

INDEX OF EQUATIONS .... **XIX**

CHAPTER 1. INTRODUCTION **1**

1.1 Fusion **3**

1.2 Proposal **4**

CHAPTER 2. STATE OF THE ART **7**

2.1 Introduction **7**

2.2 Data Fusion **9**

2.2.1 Data Fusion Definition 9
2.2.2 Architectures 10
a) Division according to the abstraction level 10
b) Centralized vs. decentralized Data Fusion 13
2.2.3 Data Fusion Models 14
a) The JDL model 14
b) Other models 16
2.2.4 Data Architecture and Models Conclusions 19
2.2.5 Fata Fusion Processes and Tasks 20
2.2.6 Data Fusion in Automotive Applications 22
a) Low-level fusion 22
b) Medium-level 22
c) High-level fusion 23

d) Other approaches 24

2.3 Sensing Technologies in Intelligent Vehicles 27

2.3.1 Lidar Environment Detection 27

2.3.2 Vision Environment Detection 29

a) Data acquisition 30

b) Region of Interest detection 31

c) Classification and verification 33

2.3.3 Other Sensing Technologies 35

2.3.4 Context Information 35

2.4 Conclusion 36

CHAPTER 3. GENERAL DESCRIPTION 39

3.1 The Platform 40

3.2 Data Sources 43

3.2.1 Laser Scanner 43

3.2.2 Computer Vision 44

3.2.3 Inertial Sensor 44

3.2.4 Context Information 45

3.3 Fusion Levels 46

3.3.1 Level 0 (Sub – Object Refinement) 46

3.3.2 Level 1 (Object Refinement) 47

3.3.3 Level 2 & 3 (Situation and Threat Refinement) 48

3.3.4 Level 4 (Process Refinement) 49

3.4 Proposal Phases 49

3.4.1 Obstacle Detection 49

3.4.2 Multiple Target Tracking 50

3.4.3 Danger Estimation 51

3.4.4 Test and Conclusions 51

CHAPTER 4. OBSTACLE DETECTION AND CLASSIFICATION 53

4.1 Laser Scanner Obstacle Detection and Classification 53

4.1.1 Stage 1. Low Level Detection 53

a) Movement compensation 54

b) Polyline creation 56
c) Low level classification 57
d) Other obstacles 62
4.1.2 Stage 2. Higher Level Classification 62
a) Previous obstacle movement correction 62
b) Obstacle comparison 63
c) Higher level classification 64
d) Final classification 64

4.2 Data Alignment 65
4.2.1 Alignment Problem 65
a) Time alignment 65
b) Coordinate space alignment 68
4.2.2 Solution Proposed 68
a) Coordinate change 69
b) Extrinsic calibration 70
c) Time alignment 73

4.3 Vision Based Classification 74
4.3.1 Pedestrian Classification 74
a) Preprocessing stage 75
b) Gradient computation 75
c) Orientation binning 76
d) Description blocks 76
e) Blocks normalization 76
f) SVM classification 76
4.3.2 Vehicle Classification 78
a) Feature creation and selection 78
b) Adaboost classification 79

CHAPTER 5. TRACKING AND DATA ASSOCIATION 81

5.1 Position Estimation 82
5.1.1 Classic Estimators 83
5.1.2 Kalman Filter 84
5.1.3 Extended Kalman Filter 86
5.1.4 Unscented Kalman Filter 87
5.1.5 Bayesian Estimators 90

5.2 Association Methods 92
5.2.1 Gating Techniques 93
5.2.2 Global Nearest Neighbor (GNN) 94
5.2.3 Multiple Hypothesis Tracker (MHT) 95
5.2.4 Probabilistic Data Association (PDA) 96
5.2.5 Joint Probabilistic Data Association (JPDA) 97

5.3 Tracking and Data Association Fusion Procedures for Road Environments 100
5.3.1 Target Model 100
5.3.2 Estimation Methods 102
5.3.3 Global Nearest Neighbor Solution for Data Fusion in Road Environments 102
a) Gating and data association 102
b) Track creation and deletion logic 103
c) Estimation 104
5.3.4 Multiple Hypothesis Tracker approach for Data Fusion in Road Environment 104
a) Gating and data association 105
b) Track management 105
c) Estimation 105
5.3.5 Joint Probabilistic Data Fusion Association Filter for Road Environments 106
a) Assignation process 106
b) Filter update 107
c) Track management 107

5.4 Danger Estimation 108
5.4.1 Pedestrian 108
a) Response distance 109
b) Braking distance 109
c) Danger zones 110
d) Danger estimation 112
5.4.2 Vehicles 113
a) Relevant distances 113
b) Braking distance 113
c) Safety distance 113

5.5 Conclusions 114

CHAPTER 6. PRACTICAL RESULTS 117

6.1 Calibration Tests 118

6.2 Pedestrians 122
6.2.1 Low Level Algorithms Performance for Pedestrian Detection 122
6.2.2 Tracking Algorithms Performance 126
6.2.3 Association Methods Performance 128
a) MHT vs. GNN tests 128
b) JPDA vs. GNN tests 130
6.2.4 Fusion System Performance 137

6.3 Vehicles 138
6.3.1 Low Level Algorithms Performance for Vehicle Detection 138
6.3.2 Tracking Algorithm Performance 140
6.3.3 Association Methods Performance 140
a) MHT vs. GNN test 141
b) JPDA vs. GNN test 142
6.3.4 Fusion System Performance 146

CHAPTER 7. CONCLUSIONS 147

7.1 Conclusions 147

7.2 Contributions 147

7.3 Future Works 149

REFERENCES 151

INDEX OF FIGURES

Figure 1.1. Test platform IVVI 2.0. 4
Figure 1.2. Detection examples for vehicle (right) and pedestrian (left). Blue boxes represent laser based pedestrian detection, yellow boxes laser vehicle detection, in red the visual based detections. 5
Figure 2.1. Google autonomous vehicle 8
Figure 2.2. Low level Fusion processing diagram. Fusion is performed from the raw data creating a new set of more complex information to be processed. 11
Figure 2.3. Medium Level approach diagram, preprocessing is performed for each sensor and a combined feature set of data is used combining features from all the sensors. 11
Figure 2.4. High level approaches perform an inference for each sensor and combines them according to the certainty degree of each one. 12
Figure 2.5. JDL Model defined in 1985 14
Figure 2.6 The new definition of the JDL model, given in [6] and [5]. 15
Figure 2.7 Waterfall model [9] 17
Figure 2.8. Boyd Model 17
Figure 2.9. Omnibus model [12] 18
Figure 2.10. Thomopoulos architecture. 19
Figure 2.11. Pedestrian ROI detection using disparity map. (a) stereo system used.(b) image of one of the sensors. (c) binarized image. (d) disparity map. (e) close obstacle detection. (f) Bounding boxes representing the pedestrians. This image was taken from [18]. 22
Figure 2.12. Image and laser scanner data set representation, from [19] . 23
Figure 2.13. High level fusion based pedestrian detection in [22]. 24
Figure 2.14. vehicle detection based on laser scanner ROI detection and SVM vision based classification [24]. 25
Figure 2.15. Cooperative fusion between vision and laser scanner detection in real road situation, from [28]. 25
Figure 2.16. Pedestrian correctly found in danger zones from [29] 26
Figure 2.17. Three sensor detections given by the software interface in [30]. 26
Figure 2.18. Grid based detection and tracking in [31]. 27
Figure 2.19. Processes of computer vision detection and classification algorithms. 29
Figure 2.20. Sensing technologies used by Vislab, laboratory of Parma (Italy), during the Urban Challenge in 2007, with several configurations of cameras, including the novel trinocular system. Left the overall field of view, right the images of the truck with the sensing devices highlighted [71]. 31
Figure 2.21. ROI detection using temperature with fir image [76]. 33

FIGURE 2.22. MODELS OF THE PEDESTRIANS CREATED IN THE APPROACH [86]. (A) IS THE GENERAL MODEL, (B) THE MODEL FOR CLOSED LEGS, (C) ALMOST CLOSED, (D) ALMOST OPEN AND (E) OPEN 35
FIGURE 3.1. OVERALL SYSTEM DIAGRAM WITH THE CORRESPONDING FUSION LEVELS IMPLEMENTATION. 39
FIGURE 3.2. TEST PLATFORM IVVI 2.0 WITH THE LASER SCANNER MOUNTED IN THE FRONT BUMPER. 40
FIGURE 3.3. SENSOR DEVICES INSTALLED IN IVVI 2.0 41
FIGURE 3.4. PROCESSING UNITS AND OTHER DEVICES INSTALLED IN THE PLATFORM IVVI 2.0. 42
FIGURE 3.5.TECHNOLOGIES BEING DEVELOPED AND RESEARCHED IN THE PLATFORM IVVI 2.0. 43
FIGURE 3.6. LASER SCANNER ENVIRONMENT RECONSTRUCTION EXAMPLE. DISTANCES FROM 5,10, 15 AND 20 ARE HIGHLIGHTED IN YELLOW, AS WELL AS SOME BOUNDING BOXES WITH RELEVANT OBSTACLES. 44
FIGURE 3.7. HOG FEATURE BASED PEDESTRIAN DETECTION (LEFT). HAAR CASCADE CLASSIFIER BASED VEHICLE DETECTION (RIGTH). 44
FIGURE 3.8. MTI-G GPS AIDED INERTIAL MEASUREMENT... 45
FIGURE 3.9. BOUNDING BOXES FOR VEHICLE DETECTION CREATED AFTER LASER SCANNED OBSTACLE DETECTION. 47
FIGURE 3.10. LASER SCANNER BASED PEDESTRIAN DETECTION EXAMPLE. 50
FIGURE 4.1. LASER SCANNER BASED ROAD ENVIRONMENT RECONSTRUCTION SCHEME. 54
FIGURE 4.2. VEHICLE MOVEMENT COMPENSATION AND DATA ALIGNMENT. FIGURE A SHOWS THE DETECTION POINTS, IN BLUE BEFORE THE VEHICLE MOVEMENT COMPENSATION AND IN RED THE COMPENSATED. FIGURE B SHOWS THE SHAPE RECONSTRUCTED AFTER THE MOVEMENT COMPENSATION. FIGURE C SHOWS THE FINAL ALIGNMENT OF THE LASER SCANNER DATA AND THE IMAGE. 55
FIGURE 4.3. SEGMENT CREATION EXAMPLE BASED IN THE DISTANCE AMONG DETECTIONS... 56
FIGURE 4.4.POLYLINE CREATION EXAMPLE. 57
FIGURE 4.5. ROAD BORDER DETECTION BASED IN LITTLE OBSTACLE HISTOGRAMS. WHEN THE HISTOGRAM IS GREATER THAN A GIVEN VALUE ROAD BORDERS ARE DETECTED. 58
FIGURE 4.6. LASER BEHAVIOR FOR A COMPLETE SCAN. 59
FIGURE 4.7. MOVING VEHICLE PATTERN. 59
FIGURE 4.8. EXAMPLES OF DIFFERENT PEDESTRIAN'S PATTERNS USED TO STUDY THE COMMON PATTERN. 61
FIGURE 4.9. (A) PATTERN FOR PEDESTRIAN DETECTION. (B) DIFFERENT EXAMPLES OF DIFFERENT PATTERNS GIVEN BY PEDESTRIAN WITH DIFFERENT LEG POSITIONS WITH THE INFORMATION FROM LASER SCANNER TRANSLATED TO THE IMAGE. 61
FIGURE 4.10. LASER SCANNER LOW LEVEL DETECTION EXAMPLE (LEFT) AND REAL IMAGE OF THE ENVIRONMENT (RIGHT). 62
FIGURE 4.11. SHAPE CHARACTERISTICS USED TO COMPARE DETECTED OBSTACLES. 64
FIGURE 4.12. SENSORS' BEHAVIOR WHERE ALL SENSORS ARE PERIODIC. PERIODS DO NOT HAVE TO BE EQUAL. 66
FIGURE 4.13. SENSORS BEHAVIOR WHEN SOME SENSORS ARE EVENT-ORIENTED SOME OTHER ARE PERIODICAL. 66
FIGURE 4.14. SENSORS BEHAVIOR WHEN ALL SENSORS ARE EVENT-ORIENTED.. 67
FIGURE 4.15. TWO REFERENCE POINTS FOR IMAGE AND LASER SCANNER. LASER SCANNER DETECTIONS ARE SHOWN, WITH PEDESTRIAN DETECTION IN HORIZONTAL RECTANGLES. VERTICAL RECTANGLES SHOW THE IMAGE DETECTIONS. 68

FIGURE 4.16. DIFFERENT COORDINATE SYSTEMS: (X_C,Y_C,Z_C) ARE CAMERA COORDINATE SYSTEM, (X_L,Y_L,Z_L) IS THE LASER COORDINATE SYSTEM, AND (X_V,Y_V,Z_V) IS THE COORDINATE SYSTEM OF THE VEHICLE. COORDINATES (U,V) ARE THE IMAGE COORDINATES. 69

FIGURE 4.17. BOUNDING BOXES (LEFT) AND LASER SCANNER DETECTION (RIGHT). FOR VEHICLES, SMALL OBSTACLES ARE FILTERED, THUS ONLY BIG OBSTACLES ARE CONSIDERED INTERESTING TO PERFORM VEHICLE DETECTION. 70

FIGURE 4.18. EXAMPLE OF TWO CALIBRATION SEQUENCES WHERE THE EULER ANGLES ARE CORRECTED, VARYING THE PARAMETERS SHOWN IN THE RED CIRCLE (YAW,PITCH,ROLL ANGLES AND X,Y,Z DISTANCES BETWEEN COORDINATES AXES). 71

FIGURE 4.19. MAXIMUM PITCHING ERROR ANGLE ALLOWED REPRESENTATION. ACCORDING TO THE DISTANCE AND THE PEDESTRIAN HEIGHT. 72

FIGURE 4.20. LASER SCANNER DETAIL WITH THE INERTIAL MEASUREMENT SYSTEM MOUNTED IN THE TOP. FRONTAL VIEW (LEFT) AND TOP VIEW (RIGHT). 72

FIGURE 4.21. TEMPORAL REPRESENTATION OF A SEQUENCE, THE LASER SCANNER HAS A FIXED TEMPORIZED THANKS TO THE FASTER ALGORITHM AND THE FASTER PERIOD. EVERY LECTURE FROM THE CAMERA (IN RED) IS EXTRAPOLATED TO THE NEXT LECTURE OF THE LASER SCANNER (IN BLUE). 73

FIGURE 4.22. EXTREME EXAMPLE OF THE NON-SYNCHRONIZATION GIVEN BY THE LASER AND THE IMAGE. LASER SCANNER PERFORMS THE DETECTION WITH SOME DELAY WITH THE CAMERA DETECTION. 74

FIGURE 4.23. FIGURE TAKEN FROM THE ORIGINAL PAPER OF HOG FEATURES FOR PEDESTRIAN DETECTION [83]. (A) THE AVERAGE GRADIENT IMAGE OVER THE TRAINING EXAMPLES. (B) EACH "PIXEL" SHOWS THE MAXIMUM POSITIVE SVM WEIGHT IN THE BLOCK CENTERED ON THE PIXEL. (C) LIKEWISE FOR THE NEGATIVE SVM WEIGHTS. (D) A TEST IMAGE. (E) IT'S COMPUTED R-HOG DESCRIPTOR. (F,G) THE R-HOG DESCRIPTOR WEIGHTED BY RESPECTIVELY THE POSITIVE AND THE NEGATIVE SVM WEIGHTS. 77

FIGURE 4.24. EXAMPLES OF THE PEDESTRIAN DETECTION (RED BOXES) IN IMAGES USING HOG FEATURES. 77

FIGURE 4.25. HAAR-LIKE FEATURES DEFINED BY VIOLA AND JONES [78]. 79

FIGURE 4.26. CASCADE CLASSIFIER DEFINITION GIVEN IN [78]. 79

FIGURE 4.27. EXAMPLE OF HAAR-LIKE FEATURE BASED VEHICLE DETECTION (RED BOXES). 80

FIGURE 5.1. BLOCK DIAGRAM OF THE MODEL OF THE GENERAL SYSTEM. 83

FIGURE 5.2.GENERAL BLOCK DIAGRAM FOR TARGET TRACKING. 84

FIGURE 5.3. GNN UNCERTAINTY PROBLEM. CROSSES REPRESENT NEW OBSERVATIONS, AND POINTS THE OLD TRACKS, CIRCLES ARE THE GATING. THIS SITUATION SOME OF THE NEW DETECTIONS ARE DIFFICULT TO ASSIGN TO THE OLD TRACKS BECAUSE THEY FALL IN THE GATING OF MORE THAN ONE TRACK. 94

FIGURE 5.4. EXAMPLE REPRESENTATION. TWO TRACKS AND THREE OBSERVATIONS. 98

FIGURE 5.5 PROCESSES FOR MHT DATA FUSION APPROACH FOR ROAD ENVIRONMENTS. 104

FIGURE 5.6. DANGER ZONES REPRESENTED WITH THE RELEVANT DISTANCES AND THE FIELDS OF VIEW OF THE DIFFERENT DEVICES. 111

FIGURE 5.7. ESTIMATION DANGER EXAMPLE FOR A PEDESTRIAN AT 40 KM/H, DEPENDING ON THE DISTANCE IN METERS. 112

FIGURE 6.1. EXPERIMENTS TO MEASURE THE MEASUREMENT ERRORS. (A) X POSITION IS FIXED. (B) Y POSITION IS FIXED. 118

FIGURE 6.2 . RESULTS OF THE EXPERIMENTS TO MEASURE THE MEASUREMENT OF ERRORS Y (IN METERS) VS X (IN METERS). (A) TOTAL MEASUREMENTS OF THE EXPERIMENT FOR LATERAL MOVEMENT. (B) TOTAL MEASUREMENT OF THE EXPERIMENTS FOR VERTICAL MOVEMENT. (C) Y ERROR VS X COORDINATES IN LATERAL EXPERIMENTS. (D) X ERROR VS Y COORDINATE IN VERTICAL MOVEMENT EXPERIMENTS. 119

FIGURE 6.3. TEST PERFORMED TO CHECK THE VIABILITY OF THE LASER SCANNER SYSTEM........ 120

FIGURE 6.4. VEHICLES USED FOR THE TEST. GREEN VEHICLE HAD MOUNTED THE LASER SCANNER AND METAL GREY AND BACK VEHICLE WERE USED TO MEASURE THE REFLEXION OF THE LASER SCANNER. 120

FIGURE 6.5. ERROR OF THE MEASUREMENT OF THE WIDTH OF THE VEHICLE, OBTAINED BY LASER SCANNER ACCORDING TO THE DISTANCE. 121

FIGURE 6.6. EXAMPLE FRAMES OF THE DIFFERENT TESTS PERFORMED FOR PEDESTRIAN DETECTION. TEST SEQUENCES WERE PERFORMED IN A PARKING LOT (A), INTER-URBAN SCENARIOS (B) AND URBAN SCENARIOS (C) AND (D). 123

FIGURE 6.7. EXAMPLE OF MISDETECTIONS OF THE CAMERA DUE TO MULTIPLE ROIS POINTING TO A SINGLE PEDESTRIAN. TWO LASER SCANNERS POINT TO THE PEDESTRIAN, THUS BOTH OF THEM RETURN POSITIVE DETECTIONS. IN THESE SITUATIONS IT COULD BE CONSIDERED A POSITIVE DETECTION BY THE BASIC HOG FEATURE APPROACH, ALTHOUGH THEY WERE CONSIDERED MISDETECTIONS OF THE VISION SYSTEM IN THE PRESENT APPROACH THAT USES LASER SCANNER ROI DETECTION. 124

FIGURE 6.8. POSITIVE DETECTIONS EXAMPLES. BLUE BOXES REPRESENT LASER SCANNER POSITIVE DETECTIONS. RED BOXES THE IMAGE POSITIVES. ALSO LASER SCANNER POLYLINE RECONSTRUCTIONS ARE SHOWED IN THE IMAGE. 125

FIGURE 6.9. EXAMPLE OF ERRORS IN THE TESTS. BLUE BOXES REPRESENT LASER SCANNER POSITIVE DETECTIONS. RED BOXES THE IMAGE POSITIVES. ALSO LASER SCANNER POLYLINE RECONSTRUCTIONS ARE SHOWED IN THE IMAGE. 126

FIGURE 6.10. BASIC TEST SEQUENCE WITH A PEDESTRIAN PERFORMING LATERAL MOVEMENT. LEFT IS THE GNN APPROACH, RIGHT MHT. 129

FIGURE 6.11 INTER-URBAN SCENARIO WITH TWO PEDESTRIANS MOVING FROM LEFT TO RIGHT CROSSING THE ROAD. LEFT GNN APPROACH; RIGHT MHT. THE CIRCLE HIGHLIGHTS PART OF THE SEQUENCES WHERE THE FALSE POSITIVES ARE PROPAGATED. 130

FIGURE 6.12. FRAMES OF THE SEQUENCE FOR TESTING GNN VS JDPA WITH TWO PEDESTRIANS CROSSING THE ROAD WITH TRAJECTORIES THAT CROSS........ 131

FIGURE 6.13. RESULTS OF THE SEQUENCE WITH TWO PEDESTRIANS CROSSING THE ROAD. A IS THE OVERALL RESULT FOR GNN AND B FOR THE JPDA. C AND D GIVE DETAILS OF THE MOMENT WHERE BOTH PEDESTRIANS CROSS. E AND F FINALLY GIVE DETAILS OF THE CLOSEST PEDESTRIAN IN THE MOMENT WHERE BOTH PEDESTRIAN CROSS. AT RIGHT RESULTS FOR JPDA AND LEFT FOR GNN. 132

FIGURE 6.14. FRAMES OF THE TEST 2 FOR TESTING GNN VS JDPA IN URBAN ENVIRONMENTS WITH THREE PEDESTRIANS INVOLVED........ 133

FIGURE 6.15. RESULTS OF THE SEQUENCE THREE PEDESTRIANS IN INTERURBAN SCENARIO. AT RIGHT RESULTS FOR JPDA AND LEFT FOR GNN. FIRST, THE RESULTS SEQUENCES ARE DEPICTED (UP). THE SAME RESULTS ARE SHOWN DOWN WITH THE MAIN DIFFERENCES HIGHLIGHTED IN (A) AND (B). 134

FIGURE 6.16. FRAMES OF THE SEQUENCE THAT WAS USED TO COMPARE THE BEHAVIOR OF THE GNN AND THE JDPA WITH TWO CLOSE PEDESTRIANS. AS SHOWN IN THE FIGURE, IT IS DIFFICULT FOR THE LASER SCANNER TO DIFFERENTIATE AMONG THEM. 135
FIGURE 6.17. RESULTS OF THE SEQUENCE WITH TWO PEDESTRIANS WALKING CLOSE TO EACH OTHER. AT RIGHT RESULTS FOR JPDA AND LEFT FOR GNN. FIRST THE RESULTS SEQUENCES ARE DEPICTED (ABOVE). THE SAME RESULTS ARE SHOWN DOWN WITH THE MAIN DIFFERENCES HIGHLIGHTED WITH A RED CIRCLE. 136
FIGURE 6.18. MHT (RIGHT) PERFORMANCE VS. GNN (LEFT). IN GREEN THE ESTIMATION WITH NO OBSERVATION MATCHES AND IN BLACK MATCHES. 141
FIGURE 6.19. FRAME EXAMPLES OF THE TEST PERFORMED WITH THE IVVI 2.0 FOLLOWING A VEHICLE. THE MOVEMENT INVOLVED TURNINGS INSIDE A ROUNDABOUT 143
FIGURE 6.20. RESULTS OF THE TRACKING OF A SINGLE VEHICLE AFTER SEVERAL MOVEMENTS, INCLUDING A ROUNDABOUT. JDPA RESULTS ARE SHOWN IN RIGHT AND GNN IN LEFT IMAGE. GREEN DETECTIONS REPRESENT ESTIMATION WITH NO MATCH, BLACK ARE MATCHING OBSERVATIONS. 143
FIGURE 6.21. FRAMES OF THE TEST SEQUENCE WITH THE VEHICLE (IVVI 2.0) PERFORMING AN OVERTAKING MANEUVER. . 144
FIGURE 6.22. RESULTS OF THE TEST SEQUENCE WITH THE VEHICLE (IVVI 2.0) PERFORMING AN OVERTAKING MANEUVER. JDPA AND GNN RESULTS ARE SHOWN (RIGHT AND LEFT RESPECTIVELY). GREEN DETECTIONS REPRESENT ESTIMATION WITH NO MATCH, BLACK ARE MATCHING OBSERVATIONS 144
FIGURE 6.23. FRAMES OF THE TEST SEQUENCE WITH TWO VEHICLES CROSSING 145
FIGURE 6.24. RESULTS OF THE TEST SEQUENCE WITH TWO VEHICLES IN A CROSSING. JDPA RESULTS ARE SHOWN AT RIGHT AND GNN AT LEFT IN THE IMAGE. GREEN DETECTIONS REPRESENT ESTIMATION WITH NO MATCH, BLACK ARE MATCHING OBSERVATIONS. 145

INDEX OF EQUATIONS

(1) 54
(2) 56
(3) 56
(4) 59
(5) 60
(6) 60
(7) 60
(8) 61
(9) 63
(10) 63
(11) 63
(12) 63
(13) 64
(14) 69
(15) 70
(16) 80
(17) 82
(18) 82
(19) 82
(20) 82
(21) 84
(22) 85
(23) 85
(24) 85
(25) 85
(26) 85
(27) 85
(28) 86
(29) 86
(30) 86
(31) 86
(32) 86
(33) 86

(34) 86
(35) 87
(36) 87
(37) 87
(38) 87
(39) 88
(40) 88
(41) 88
(42) 88
(43) 89
(44) 89
(45) 89
(46) 89
(47) 89
(48) 89
(49) 90
(50) 90
(51) 90
(52) 90
(53) 90
(54) 90
(55) 90
(56) 91
(57) 91
(58) 91
(59) 91
(60) 93
(61) 93
(62) 94
(63) 96
(64) 96
(65) 98
(66) 98
(67) 100
(68) 101
(69) 101
(70) 101
(71) 101

(72) 101
(73) 101
(74) 102
(75) 102
(76) 109
(77) 110
(78) 110
(79) 112
(80) 112
(81) 113
(82) 114
(83) 114
(84) 114

CHAPTER 1. INTRODUCTION

In the second decade of the 21st century driving is assumed to be a common factor in everyday life. Working, holidays, socializing are activities that exist in everyone's life, and most of these actions require road mobility. Mobility is essential in all developed countries for both economical and social reasons. Thus, any road related problem has an important impact in a citizens' life. Traffic congestions, road accidents, oil prizes, etc. are some of the common problems every road user has to deal with. They represent a problem that affects everyone's life, whether economic or personal.

Among all the problems that are related with transportation, traffic accidents are the most dramatic since they deal with human lives. The efforts during more recent years, such as an increase in the security measures in roads and vehicles or the enhancement of traffic laws to decrease drivers' misbehaviors, have lead to reduced death tolls in road accidents. The International Traffic Safety Data and Analysis Group (IRTAD) on its annual report in 2010 [1] expressed that the safety of roads in the member groups increased, reducing the death tolls. In 2009 the United States reached its lowest fatality rate of the last 50 years (Table 1). Although the report remarks the fact that the recent global economical crisis may have lead to a decrease in the volume of road traffic, hence decreasing the traffic accidents, the reduction of the number of accidents is higher than expected. It also informs that the death average during the last decade has considerably decreased, presenting a decreasing rate higher than the three preceding decades. In the European union, according to European Information Society [2], each year more than 1.3 million road accidents occur in which over 41 thousand people die. Thus, even though the efforts are helping to mitigate this number, there is still a considerable amount of work to be done.

The new information and communication technologies developed in the last decade allow the creation of more complex and reliable safety applications. These new applications are able to reduce the number of accidents and deaths in the road by both preventing them and abating the harm caused by accidents.

Road Fatalities[1]								
Recent data					Long-term trends – Average annual change			
Country	2009	2008	Change 2009-2008	Change 2009-2000 [3]	2000-2009[3]	1990-1999	1980-1989	1970-1979
Argentina[4)]	7,364	7,552	-2.5%	12%		-	-	-
Australia	1,492	1,442	3.5%	-18%	-2.2%	-3.0%	-1.7%	-0.9%
Austria	633	679	-6.8%	-35%	-4.7%	-4.0%	-2.7%	-1.8%
Belgium[2]	955	944	1.2%	-35%	-4.7%	-3.8%	-2.0%	-3.0%
Cambodia[4]	1,717	1,638	4.8%	328%	17.5%	-	-	-
Canada[2]	2130	2,419	-11.9%	-27%	-3.4%	-3.1%	-2.8%	1.6%
Czech Republic	901	1,076	-16.3%	-39%	-5.4%	1.3%	-1.7%	-4.0%
Denmark	303	406	-25.4%	-39%	-5.4%	-2.3%	-0.3%	-5.4%
Finland	279	344	-18.9%	-30%	-3.8%	-4.4%	3.2%	-5.2%
France	4,273	4,275	-0.05%	-48%	-6.9%	-0.7%	-1.8%	-2.1%
Germany	4,152	4,477	-7.3%	-45%	-6.4%	-3.8%	-4.7%	-3.4%
Greece	1,456	1,553	-6.2%	-29%	-3.7%	0.4%	3.7%	3.4%
Hungary	822	996	-17.5%	-32%	-4.1%	-6.7%	3.2%	0.8%
Iceland	17	12	41.7%	-47%	-6.8%	-1.5%	1.3%	3.4%
Ireland	238	279	-14.7%	-43%	-6.0%	-1.6%	-2.2%	1.4%
Israel	314	412	-23.8%	-31%	-4.0%	1.2%	1.0%	0.8%
Italy	4,237	4,725	-10.3%	-40%	-5.5%	-0.7%	-3.1%	-2.3%
Japan	5,772	6,023	-4.2%	-45%	-6.3%	-3.7%	2.7%	-7.3%
Korea	5,838	5,870	-0.5%	-43%	-6.0%	-3.0%	9.4%	7.7%
Lithunia[4]	370	499	-25.9%	-42%	-5.9%	--3.2%	2.9%	2.0%
Luxembourg	48	36	33.3%	-37%	-5.0%	-2.2%	-4.1%	-4.2%
Malaysia[4]	6,745	6,527	3.3%	12%	1.2%	-	-	-
Netherlands	644	677	-4.9%	-40%	-5.6%	-2.6%	-3.4%	-5.1%
New Zealand	384	365	5.2%	-17%	-2.0%	-3.9%	2.7%	-1.8%
Norway	212	255	-16.9%	-38%	-5.1%	-1.0%	0.6%	-2.7%
Poland	4,572	5,437	-15.9%	-27%	-3.5%	-0.9%	1.3%	5.9%
Portugal	840	885	-5.1%	-55%	-8.5%	-3.1%	0.5%	4.9%
Slovenia	171	214	-20.1%	-46%	-6.5%	-4.7%	-0.1%	1.9%
Spain	2,714	3,100	-12.5%	-53%	-8.0%	-4.9%	4.1%	2.4%
Sweden	358	397	-9.8%	-39%	-5.4%	-3.1%	0.7%	-3.8%
Switzerland	349	357	-2.2%	-41%	-5.7%	-5.0%	-3.3%	-3.2%
United Kingdom	2,337	2645	-11.6%	-35%	-4.6%	-4.5%	-1.2%	-1.7%
United States	33,808	37,423	-9.7%	-19%	-2.4%	-0.7%	-1.3%	-0.3%

Source: IRTAD, see www.irtad.net

1. Police-recorded fatalities. Death within 30 days. Lithuania: death within 7 days for 1990.
2. Provisional data for 2009.
3. 2004-2009 for Argentina.
4. Accession countries. Data are under review

Table 1. Road Fatalities in countries inscribed in The International Traffic Safety Data and Analysis Group (IRTAD) [1].

Most of traffic accidents are related to human error. Carelessness and erroneous decisions by the driver are the two main factors that cause traffic accidents. These kinds of errors, related with human nature, are impossible to be eliminated, although efforts can be made to decrease them. Recent researches in Intelligent Vehicles have focused on using advances in information technologies to prevent these errors. Advance Driver Assistance Systems (ADAS) try to warn the driver and preventing driver in case of hazardous situations.

Sensor trustability is one of the main issues when dealing with road Safety applications. ADAS requires the most reliable set of sensors to fulfill the requirements of such demanding applications. Thus, to accomplish such a difficult task, the combination of different information sources is mandatory to overcome the limitations of each sensor independently.

1.1 Fusion

Road Safety applications require the most reliable sensor systems. During recent years, the advances in information technology have lead to more complex road safety applications, which are able to cope with a high variety of situations. But a single sensor is not enough to provide reliable results necessary to fulfill the demanding requirements that these applications need. Here is where Data Fusion (DF) presents a key point in road safety applications. Recent research in Intelligent Transport Systems (ITS) research field tries to overcome the limitations of the sensors by combining them. Also contextual information has a key role for robust safety applications to provide reliable detection and complete situation assessment.

The present thesis aim is to prove that DF techniques can deliver more robust and reliable road safety applications, by combining the capacities of different sensors. The sensors used are laser scanner, computer vision, and inertial system. The application also takes advantage of contextual information, which is a very recent issue in the Data Fusion researching field. In this way, all levels of the fusion process are fulfilled. Contextual information can be helpful for both increasing the accuracy of each sensor independently and providing new information sources to improve the performance of the Fusion process. The contextual information definition and its relevance in the fusion process are provided in chapter 2.

Three sensors were used in the scope of the present thesis to prove the above assumptions:

1- Laser scanner. Recent researches have focused on the use of this well-known sensor in automotive applications. Its robustness and reliability have been proved in different tests and contests (e.g. DARPA Grand and Urban Challenge).

2- Computer vision. This is a common sensor in ITS researches and nowadays can be found in commercial systems for automotive applications.

3- Inertial sensor. This is an improved GPS with inertial correction that allows the user to accurately measure and estimate not only GPS position and velocity, but also euler angles, acceleration, etcetera. Furthermore, this sensor system allows obtaining contextual information about vehicle state (i.e. velocity, euler angles and angular speed) that is added to a prior knowledge of the nature of the matter.

1.2 Proposal

As mentioned before, the present thesis assumes that DF procedures can help to enhance the possibility of classic ADAS systems, which usually are based in single sensor approaches. In this paper, taking advantage of the knowledge of the Intelligent System Lab [3], and its platform IVVI (Intelligent Vehicle based in Visual Information) 2.0 (Figure 1.1), these assumptions will be proved by implementing and testing algorithms that complete the classical approaches for road environment detection and classification. Part of the present work is common among ITS community: visual based approaches that try to deal with ADAS common problems, such as vehicle detection and pedestrian detection. Therefore, by adding new sensors available in the platform, such as laser scanner and inertial system, the capabilities of the detection and classification algorithms are enhanced, providing more reliable and robust algorithms able to fulfill the strong demands of these applications.

Figure 1.1. Test platform IVVI 2.0.

The work presented focuses on the detections and protection of the typical road users, as in pedestrians and other vehicles (Figure 1.2). The danger estimation that involves these detections will be estimated, thus providing a complete solution to road safety applications based in Data Fusion theory. Therefore, the present work provides a novel approach to enhance the classic pedestrian and vehicle detection systems with Data Fusion and Contextual Information, providing a complete tool to detect and classify the different road users and estimate the danger that are involved in these detections.

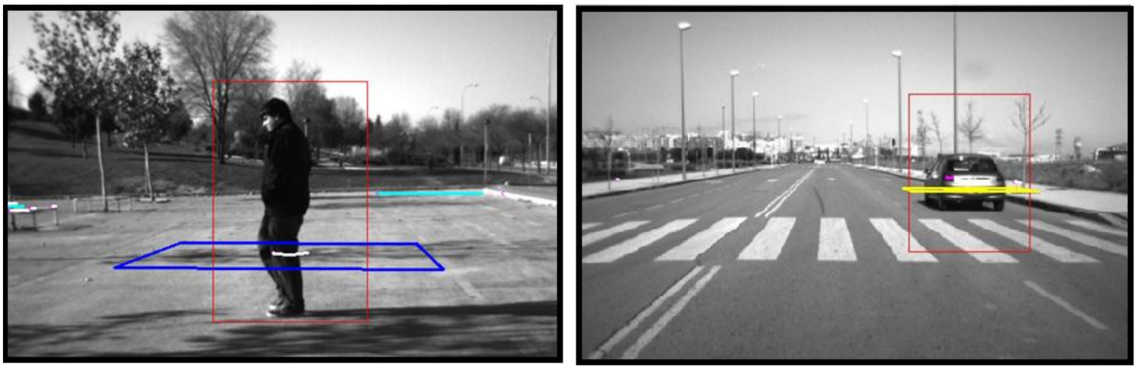


Figure 1.2. Detection examples for vehicle (right) and pedestrian (left). Blue boxes represent laser based pedestrian detection, yellow boxes laser vehicle detection, in red the visual based detections.

Several phases were accomplished in the present thesis:

First, research in the laser scanned based road environment reconstruction was necessary, in order to provide additional information source able to detect and classify road obstacles. Laser scanner is frequently used in recent automotive applications thanks to the reliability of its detections. But the different approaches lack reliable classification algorithms due to the limited information provided by these sensors. Hence, an innovative algorithm needed to be provided to complete the detections given by the cameras capable of detecting the obstacles in the environment and providing a classification of them.

The algorithms used for vision-based detection are well known and several possibilities are available. The approaches used in the present work, although these algorithms are state of the art algorithms that represents the latest advances in the field, they don't represent novel approaches. Therefore, the explanation of the algorithms used is limited to the basic procedures, sources and references in case of further information required are provided.

The next stage dealt with classic Multiple Target Tracking (MTT) and fusion problems i.e. data association and tracking. In this part of the thesis, several approaches were developed and tested to give a complete solution with all of the different configurations possible.

Finally, in order to provide complete fusion architecture, interactions of the detections were studied, estimating the danger involving the detections. This way the system fulfills all levels that involve DF procedures. In this stage, context information was used to complete the sensors information. Context information is a recent issue in DF that tries to enhance the information given by the sensors with some a prior knowledge of the problem.

Tests were performed in the scope of the present thesis to substantiate the performances of the algorithms as well as to provide conclusions regarding them, which are detailed in the last chapters.

In the next chapter, before providing details of the proposal, a complete description of the Data Fusion and Intelligent Transport System state of the art is given. First, Data Fusion introduction is mandatory to give context to the scope of the solution. The basic concepts of the Data Fusion researching field are provided. Second, Intelligent Transport Systems state of the art description with the different available sensors and algorithms is mandatory to provide to the reader an overall outlook of the technologies available. This information is to be regarded in the presented information fusion approach allowing to identify the contributions of the present work.

CHAPTER 2.
STATE OF THE ART

2.1 Introduction

Advances in Intelligent Transport Systems have lead to more complex and challenging applications, which deal with a great variety of situations. Such demanding tasks are impossible to accomplish by a single sensor, thus fusion of different sensor technologies has become mandatory to fulfill these requirements.

Latest advances related with ITS have proved that laser scanner is one of the key sensors when dealing with the most demanding automotive applications, as it was in DARPA World and Urban Challenge [4] and Google autonomous vehicle (Figure 2.1). But laser scanner is not the only sensor that is necessary to accomplish such demanding applications. Other sensors already in use in road applications have to be used (i.e. short and long distance radars, computer vision, sonar sensors, etc.) to complete the information provided. Each one of these sensors have different limitations inherent to their technology such as limitations in reliability of information, weather sensitivity, lighting conditions, and distance. In order to overcome the limitations of the different sensors, fusion has to be performed to combine the different sources of information. Current vehicles already incorporate new and useful Advance Driver Assistance Systems, but the next step is the cooperation among them to provide a full ADAS application capable of dealing with the most demanding situations. Here is where fusion has an important role. The aim of this thesis is to provide an architecture for Data Fusion in road environments, enabling the detection and classification of the different obstacles involved in road scenarios, by the use of laser scanner and computer vision technologies.

Before going deeper into the thesis, it is necessary to give a review of the various researching lines and works that different researchers are following all over the world and to provide a theoretical context to the present work. Present thesis deals with two scientific fields:

- **Intelligent Transport System** , represented by the *Intelligent Transport System Society* (ITS Society) with two main conferences every year, *IEEE Intelligent vehicle Symposium*

(IV) and *IEEE International Conference on Intelligent Transportation Systems* (TSC), and two main journals, *IEEE Transactions on Intelligent Transportation Systems* and *IEEE Intelligent Transportation Systems Magazine*.

- **Data Fusion**. Data Fusion technologies and algorithms research are mainly represented by the nonprofit organization dedicated to advancing the knowledge, theory, and applications of information fusion: *International Society of Information Fusion* (ISIF). Its main forums are the annual *International Conference on Information Fusion* and journals including *Journal of Advances in Information Fusion* (JAIF) and *Information Fusion*.

Figure 2.1. Google autonomous vehicle.

In the current chapter, a brief introduction to Data Fusion is given, including the key points necessary in each data fusion application, to help understand the working line presented in the thesis. Later, a more specific view of Data Fusion focused on road applications is presented, introducing the most remarkable works involving Data Fusion and Intelligent Transport System. Finally, the different sensing technologies for automotive applications are presented, with the available configurations of each one. This way the reader can have all the necessary background to understand the scope of the present work and the contributions that it provides.

At the end of the chapter some conclusions are given, summarizing the present situation of data fusion in Intelligent Transport System. Also, some answers are offered, which are the next steps to be performed in the related topics, focusing on the problems that the present thesis attempts to solve and the contributions that it intends to provide.

2.2 Data Fusion

The Data Fusion modern concept dates from the period between World Wars I and II. It is a concept adopted by the United States Department of Defense (DoD) with the aim of improving Command and Control (C2) decision-making. The intention was to create technology and scientific base that could help in C2 tasks by adding information from several sources. From that time, Data Fusion has been one of the key elements in defense and intelligence researches. Fusion has become ubiquitous in more recent decades. Now it has a key role in more than defense issues, thanks to advances in information technology. Data Fusion is no longer a term associated only with military and intelligence applications. In recent times, Data Fusion is a key element in many everyday applications. Robotics, vehicles, industry, and communications are some examples of fields where data fusion has a key role nowadays.

One of the main problems related with Data Fusion is the ambiguity of the terminology. Usually, the basic concept of Data Fusion itself is hard to define and has a wide field of action. As a consequence of this, several definitions of Data Fusion have been given as well as several models for Data Fusion procedures. The United States DoD created the Joint Directors of Laboratories (JDL) Data Fusion Group in mid 80s with the aim of improving communications among military researchers and system developers. The JDL tried to create a common model for data Fusion processing as well as a new lexicon. Over the years these definitions and models, despite much criticism, has become the basis of the Data Fusion. In this section, different models and definitions are going to be described, focusing on JDL model, which currently is the basis of most Data Fusion procedures.

2.2.1 Data Fusion Definition

Data Fusion (DF) is sometimes referred to as sensor fusion. JDL defined DF in the 80s as:

"*A process dealing with the association, correlation, and combination of data and information from single and multiple sources to achieve refined position and identity estimates, and complete and timely assessments of situations and threats, and their significance. The process is characterized by continuous refinements of its estimates and assessments, and the evaluation of the need for additional sources, or modification of the process itself, to achieve improved results*." [5].

Authors have often pointed out that the definition is very restrictive. Consequently, several other definitions have been proposed. [5] Hall and Llinas give a wider definition, describing DF as the set of techniques and procedures that *“seeks to combine information from multiple sources to achieve inferences that cannot be obtained from a single sensor or source, or wise quantity exceeds that of an inference drawn from any single source*”.

A similar definition is given [6] by A. Steinberg and C. Bowman: “*Data fusion is the process of combining data or information to estimate or predict entity states*”. As it can be discerned new definitions, much less restrictive, tend to give Data Fusion an open dimension, allowing it to be used in any discipline.

2.2.2 **Architectures**

Different models and architectures are defined. The variety of definitions and models are usually set due to the open definition of the Data Fusion and the wide varieties of situations possible to use Data Fusion processes. In this section, some common divisions among the architectures will be presented.

a) Division according to the abstraction level

Data Fusion architectures are typically divided according to the abstraction level in which the fusion is performed. The simplicity and utility of this division makes it one of the broadest:

- **Low level** fusion is usually referred as direct fusion. It combines unprocessed information from different sources to create a more complex set of data to be processed.

  Low level fusion procedures can be directly applied when both sensors are measuring the same physical phenomena (e.g. two images of the same targets). Classic estimators such as Kalman Filters (KF) are very common for raw data fusion.

- **Medium level** or feature level fusion is utilized when each system separately performs a preprocessing level where some information is extracted, and inference is provided accordingly. Each system provides a feature vector that is combined to give a higher-level estimation. This estimation is typically performed by machine learning algorithms such as Neural Networks, State Vector Machines, etc.

- **High level** or decision level fusion is a procedure that combines the decisions performed regarding a given target or inference using the information of a single sensor. Thus, the information or final decision is performed, taking into account each subsystem's decisions and its trustability. Typically, these approaches are based on basic voting schemes, decision trees, Bayesian inferences, etc.

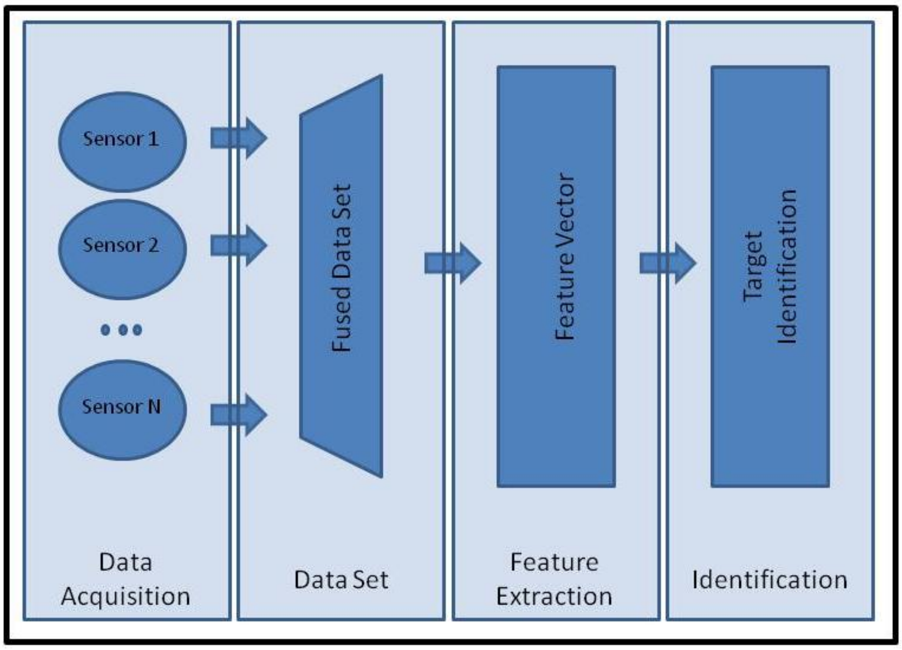


Figure 2.2. Low level Fusion processing diagram. Fusion is performed from the raw data creating a new set of more complex information to be processed.

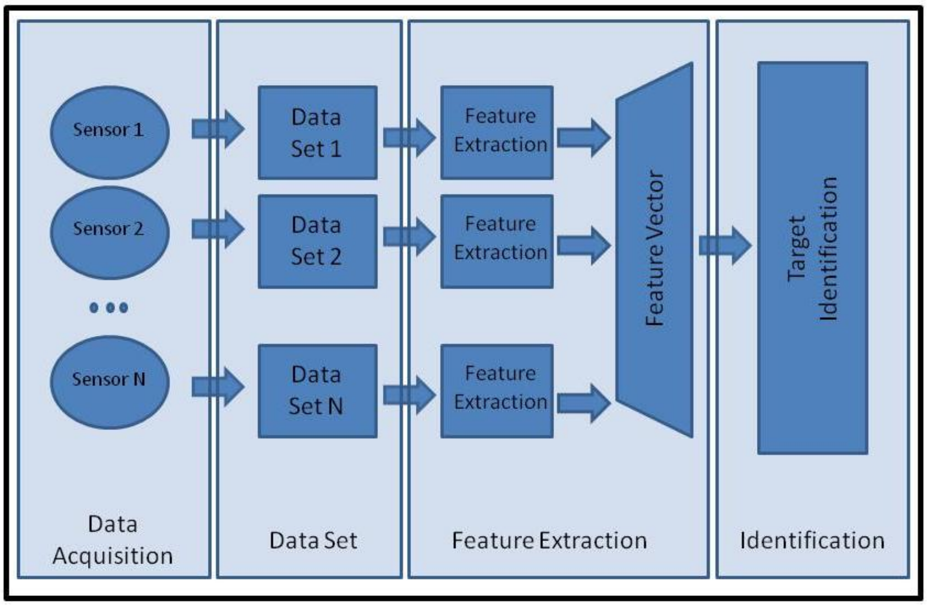


Figure 2.3. Medium Level approach diagram, preprocessing is performed for each sensor and a combined feature set of data is used combining features from all the sensors.

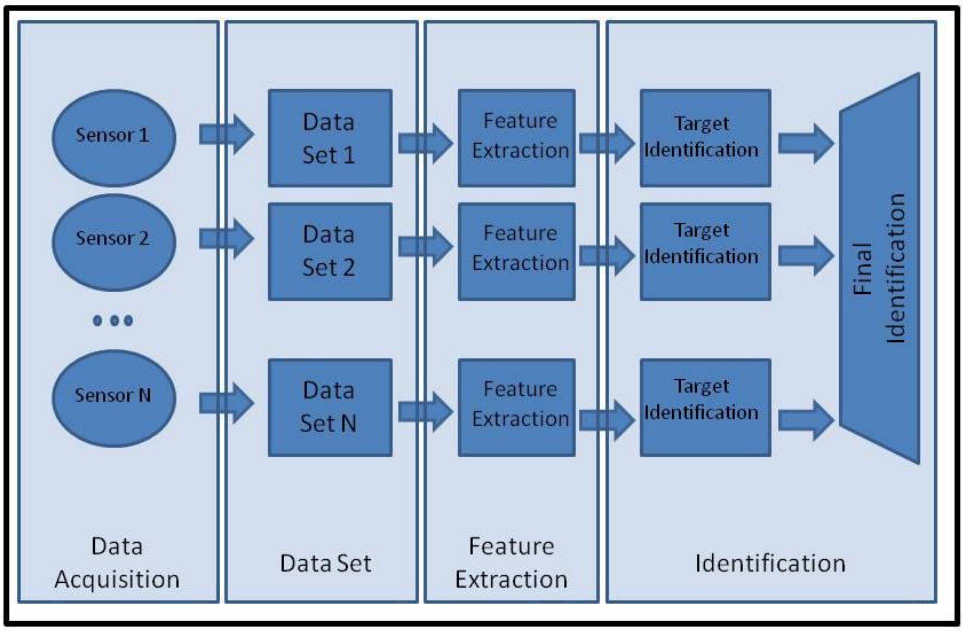


Figure 2.4. High level approaches perform an inference for each sensor and combines them according to the certainty degree of each one.

The use of any of these approaches is connected with nature of the data fusion procedure. Each of them have some advantages and disadvantages that must be considered when making a choice:

- Low-level procedures are complex and abstract. The new set of data created is intended to give more complex information; it may lead to more accurate estimations but with high effort costs, mainly related with data alignment. Thus, it is assumed that low-level fusion adds more information as well as more complexity to the system. Since these systems are completely dependent on the sensors used, adding new sensors to the system requires a complete revision of the procedure creating the worst possible choice when a scalable system is needed.

- Feature level fusion has the advantage of being in an intermediate level, allowing the systems to have extra information as a result of the different sources and maintaining a medium level of complexity. The use of different features from various sensors allows the system to take advantage of the possibilities that each sensor can provide independently. On the other hand, the training processes, usually found in these approaches, make the addition of new sensors to the system a difficult task since it must to be trained using the features of the new sensors.

- High-level fusion requires less complexity since it is based in different pre-established subsystems. The mission of the fusion process here is to add reliability and certainty to the detections and estimations, given by the different subsystems, through combining the final information provided by each subsystem. This way, high-level fusion is easy to implement even though it lacks most of the advantages of the information fusion since the information is only fused al the end of the process. On the other hand, scalability is easy. A new sensor would add more confidence and certainty to the detections in an immediate way, generally without adding complexity.

b) *Centralized vs. decentralized Data Fusion*

Mobility is a key factor in modern applications thanks to the advances in communications and information technologies. These new challenges require different topologies, according to the way in which communications are performed. These applications are divided in nodes, each one of them is formed by one or more sensors, connected to other nodes and with a processing unit. These architectures are basically divided according to where the DF is performed:

- **Centralized Systems** are those systems where each node sends all information to a central node where fusion is performed. This way the central node has all of the information from different sensors, which means that Fusion can be accomplished with more certainty.

- **Decentralized Fusion schemes** are those schemes where each node performs fusion locally, with the information from the local node and, sometimes, from the adjacent nodes. There is no central node that performs global fusion. Nodes do not have information of the global topology. These schemes easily enable the scalability of the systems since a new node can be easily added or removed. On the other hand the lack of global information suggests that the real fusion procedure is usually not as effective as centralized schemes.

Some authors tend to associate these two topology schemes for Data Fusion with the previously presented Data Fusion differentiation, according to the abstraction level where fusion is performed. This is due to the fact that decentralized Data Fusion usually involves high-level Fusion since each subsystem performs a decision independently. Low-level as well as medium-level fusion schemes require unique estimator, thus it is usually associated with centralized schemes.

2.2.3 Data Fusion Models

a) The JDL model

It is commonly remarked that one of the main barriers in the technological transfer in data fusion is the lack of unified terminology. In order to overcome this limitation JDL created a model in 1985 (Figure 2.5) that intended to give a common frame that every data fusion application should share. The model attempted to be very general, given the multiple types of applications and fields where DF can be used.

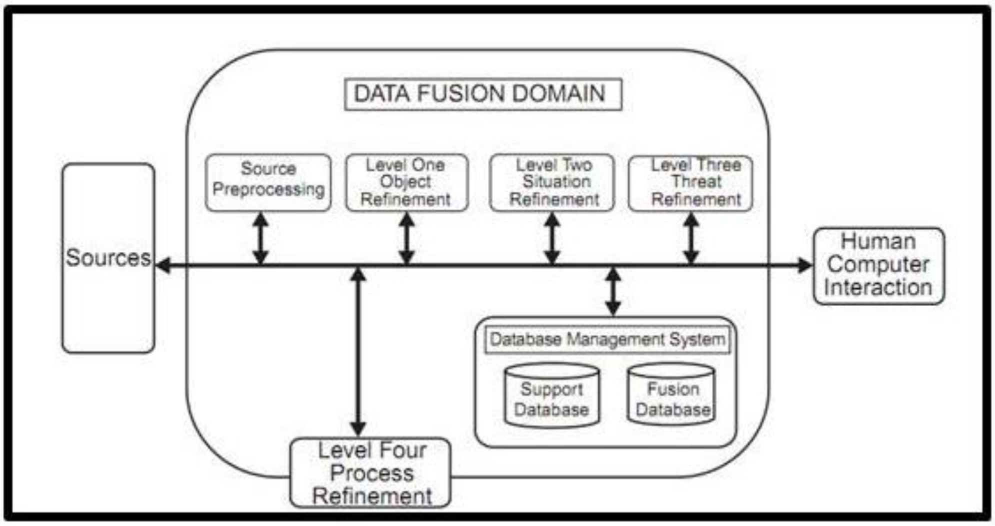


Figure 2.5. JDL Model defined in 1985.

The model is divided in 2 layers and 4 sub-processes. As explained in [5], the 4 sub-processes are as follows:

Level 1 processing (Object Refinement) is aimed at combining sensor data in order to obtain the most reliable and accurate estimate of an entity's position, velocity, attributes, and identity;

Level 2 processing (Situation Refinement) dynamically attempts to develop a description of current relationships among entities and events in the context of their environment;

Level 3 processing (Threat Refinement) projects the current situation into the future to draw inferences about enemy threats, friend and foe vulnerabilities, and opportunities for operations;

Level 4 processing (Process Refinement) is a meta-process that monitors the overall data fusion process to assess and improve real-time system performance.

In 1999, Bowman and White [6] provided a revision of the model (Figure 2.6) and level redefinitions in order to provide more useful categorization of the different types of problems, generally solved by different techniques, while at the same time trying to maintain a degree of consistency.

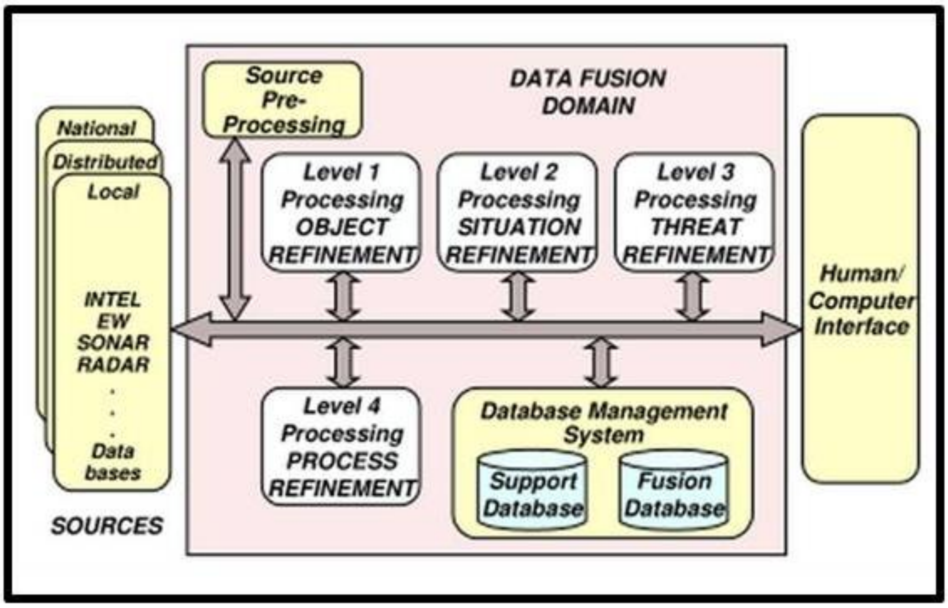


Figure 2.6 The new definition of the JDL model, given in [6] and [5].

• Level 0 — Sub-Object Data Assessment: estimation and prediction of signal- or object-observable states on the basis of pixel/signal-level data association and characterization.

• Level 1 — Object Assessment: estimation and prediction of entity states on the basis of inferences from observations.

• Level 2 — Situation Assessment: estimation and prediction of entity states on the basis of inferred relations among entities.

• Level 3 — Impact Assessment: estimation and prediction of effects on situations of planned or estimated/predicted actions by the participants (e.g., assessing susceptibilities and vulnerabilities to estimated/predicted threat actions, given one's own planned actions).

• Level 4 — Process Refinement (an element of Resource Management): adaptive data acquisition and processing to support mission objectives.

b) Other models

JDL was designed to be a functional model. It is intended to be a set of definitions of the functions that could compromise any data fusion system. It has sometimes been misinterpreted as a process model, meaning that developers should follow the levels in strict order. Also, tactical military definitions are sometimes hard to be applied to some fields.

There are other drawbacks commonly associated with the JDL model:

- The abstraction of the model makes it difficult to appropriately interpret the levels for real DF applications.
- Because it is a data or information centered model, it can be difficult to extend or reuse applications built with this model [7].
- It does not depict the processes to follow or perform the data fusion, since it is a functional model.

All these issues cause a creation of different, more process-oriented, models that specify the steps to follow in the Data Fusion applications.

The Waterfall Fusion model [8] and [9] is a process-oriented model, mainly focused in the lower levels, as seen in Figure 2.7. The model depicts the process to follow and perform the DF, and it can be easily associated with JDL. Although this model is more process-oriented than JDL, it is very similar given that the correspondence between both models is easy to find. Thus, this model suffers most of the drawbacks of the JDL. Its major limitation is the lack of feedback. It is also limited by communication between levels.

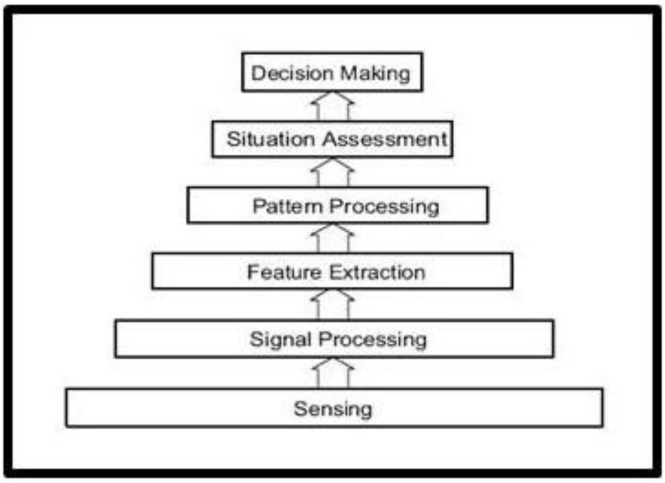


Figure 2.7 Waterfall model [9]

Boyd proposed a cycle containing four stages ([10] and [11]), representing the classic decision support mechanism for information operation, that is also widely used for business decision-support. Since decision-support systems are coupled with fusion systems [11], this model has also been used in DF. The Boyd model represents a closed loop of stages that are very useful, allowing a general overview of the general tasks of the system. But, the vision that it depicts is rather general and the structure is not helpful when trying to identify the steps that the fusion procedure should follow and perform.

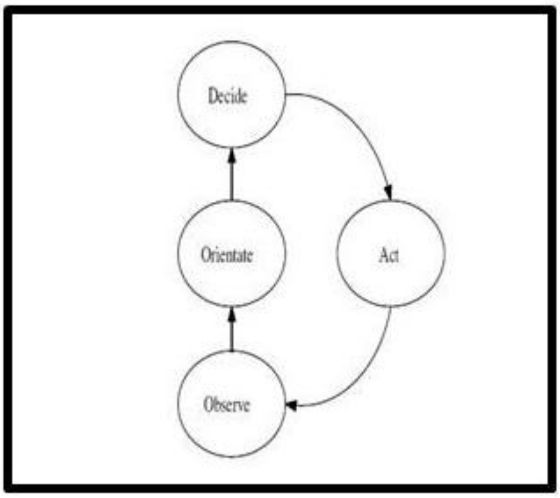


Figure 2.8. Boyd Model.

Bedworth and O'Brien analyzed different models, and used their strengths and weaknesses to create a new model that overcomes the limitations of all the previous presented models [12], called the Omnibus model (Figure 2.9). The model is based on a cycle structure, as the

Boyd model, but this model is more detailed and process-oriented. The aim is to be recursively used in the same application in two different levels of abstraction. First, it is used to characterize and structure the overall system. Second, the same structures are used to model the single subtasks of the proposed system. The separation of the DF into different tasks is very sophisticated and makes Omnibus a widely spread model. The main drawback of the model is the lack of modularity that allows systems to be horizontally separated in different processed nodes, which could be tested and used by different applications. Advances in network and information technologies that require this modularity make it inconvenient when developing DF applications using this model.

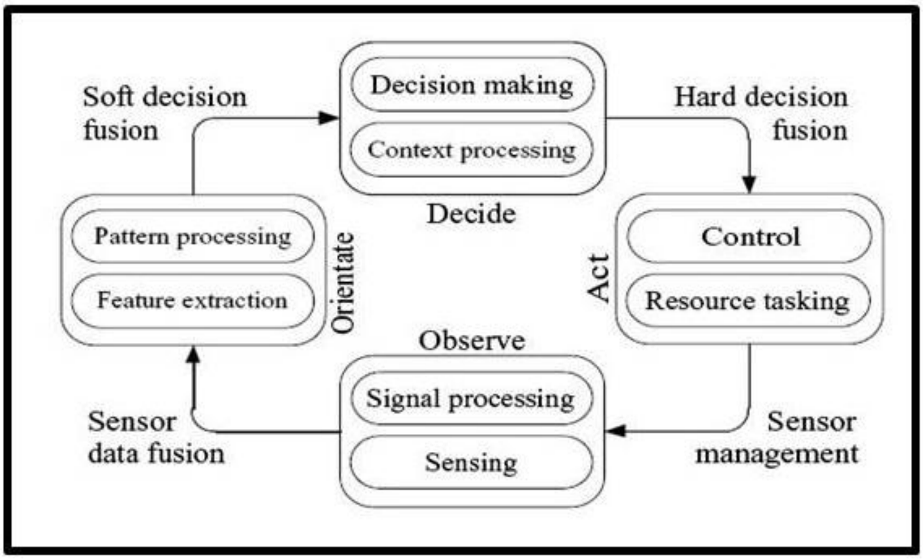


Figure 2.9. Omnibus model [12]

Thomopolulos proposed a different model [13], not strictly connected with military applications. The work was focused in the field of robotics. His model proposed a generic architecture and analytical framework to address DF in three data processing levels: signal level, level of evidence and level of dynamics.

Besides the model, Thomopoulos architecture also depicts the way in which function has to be performed in each level. *Signal level* fusion has to be performed using correlation and learning. *Evidence level* fusion uses statistical models to infer locally the information. And finally *Dynamics level* performs data fusion by either centralized or decentralized fashion, assuming that the mathematical model is known.

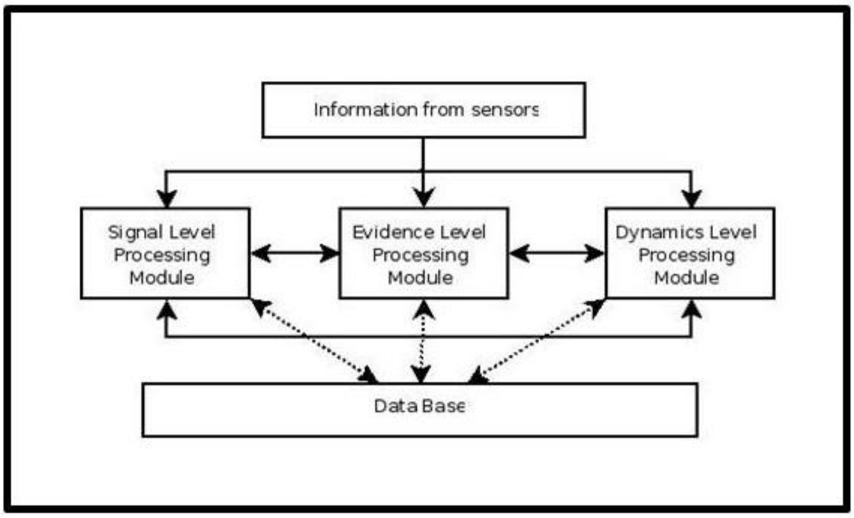


Figure 2.10. Thomopoulos architecture.

This model is very useful for horizontal modularity, since it can be divided in modules that are easily tested and reused by other applications. The main drawback is the robotic-application oriented that makes its generalization difficult.

2.2.4 Data Architecture and Models Conclusions

Data Fusion is a wide concept that makes generalizations difficult. DF is a concept that is becoming omnipresent in many engineering applications due to advances in Information Technologies. Hence, the definition of a model that generalizes the concepts is a tough task. JDL model is the most commonly used model, since it was one of the first definitions of Information Fusion, and the subsequent actualizations of the model helped to adapt the old fashion military-oriented fusion terminology to the recent applications.

Since the JDL model as well as its terminology is widely accepted in DF community, the present thesis follows this model. Processes and Tasks that follow and perform DF will be presented in next section of the chapter. In [5] and [14] D.Hall and J. Llinas give a deeper explanation of the processes using the JDL terminology.

The military oriented DF vision given by the JDL model is very easy to adapt to the field of this thesis. Road Security applications, as military applications, are oriented to threat detection, risk identification and decision-making. From a road safety point of view, this target terminology is related with the pedestrians, vehicles, and other objects that may compromise the safety of road users. Risk identification can be easily associated with the danger involved in these detections, according to the movement of the vehicle where the detection is performed.

Finally, decision-making is connected with the actions performed to avoid dangerous situations, or in case of imminent danger, which actions should be performed to mitigate the damages or injuries that this situation may produce (e.g. avoid or mitigate pedestrian injuries).

2.2.5 Fata Fusion Processes and Tasks

D. Hall and J. Llinas in [5] and [14] provide a content table that depicts the following processes and tasks associated with JDL levels (Table 2).

JDL Process	Processing function	Techniques
Level 1:Object Refinement	Data alignment	• Coordinate transforms • Units Adjustments
	Data/object correlation	• Gating techniques • Multiple hypothesis association probabilistic data association • Nearest neighbor
	Position/kinematic and attribute estimation	• Sequential estimation - Kalman filter - $\alpha\beta$ filter - Multiple hypothesis • Batch estimation • Maximum likelihood • Hybrid methods
	Object identity estimation	• Physical models • Feature-based techniques - Neural networks - Cluster algorithms - Pattern recognition • Syntactic models
Level 2:Situation Refinement	Object aggregation Event/activity interpretation Contextual interpretation	• Knowledge-based systems(KBS) - Rule-based expert systems - Fuzzy logic - Frame-based • Logical templating • Neural networks - Blackboard systems
Level 3:Threat Refinement	Aggregate force estimation Intent prediction Multi-perspective assessment	• Neural networks - Blackboard systems • Fast-time engagement models
Level 4:Process Refinement	Performance evaluation	• Measure of evaluation • Measures of performance • Utility theory
	Process control	• Multi-objective optimization - Linear programming - Goal programming
	Source requirement determination	• Sensor models
	Mission management	• Knowledge-based systems

Table 2. Processes and task to follow associated to the different fusion levels. Source [5] and [14].

Data Fusion in road safety technologies is a recent issue that tries to enhance the detection capacity of the available sensors by combining several sensor technologies. As will be detailed in the next section of this chapter, these researches are state of the art and focus mainly on the detection and classification of the various objects that may be found in a road environment. Thus, the processes that are generally detailed in Intelligent Transport System, which use Data Fusion for detection purposes, are mainly addressed in levels 0 and 1. In particular, this thesis is intended to create fusion architecture for vehicle and pedestrian detection, thus the main processes detailed are the located in the mentioned levels. Although to provide a complete Fusion application, multilevel solution will be provided at chapter 3. Providing solution for processes addressed in Fusion levels 2 and 3, and providing key elements to solve level 4. Accordingly, for the present approach, MTT with DF approach is intended, processes focused mainly on solving levels 0 and 1.The tasks related with these levels are:

- **Data Alignment**. Sensors do not usually share a common coordinate system; it means different coordinate axis, as well as, different units and transitions. These units have to be reordered to a common frame that both sensors must share in order to combine the information provided by them. Section 4.2 details the different possibilities and options that can be used for this problem, detailing the solution for the present approach.

- **Data/object correlation**. Older detections have to be associated with new detections in order to keep track of them and to provide information to upper layers about the behavior of the targets. Also, detections from different sensors and subsystems have to be associated. Sections 5.2 and 5.3 illustrate the different solutions available in MTT and the solutions used in this thesis for Data Fusion in road environments.

- **Position/movement estimation**. Estimation techniques are used to predict the movement of targets in future detections. This prediction can be used for future associations or to help upper levels to estimate the behavior of the targets. Estimation methods are detailed in section 5.1.

- **Object/identity estimation**. Final decisions about the detections are provided, and, usually, a confidence level about the estimation is also provided. Chapter 5 details track management.

The first is a key element of DF applications. The following three processes are not only typical processes of DF, but these processes are usually included in MTT applications. Chapter 5 deals with the different procedures related with MTT techniques, applied to the DF problem of the present approach. Deeper explanations of the utilized MTT procedures are given in [11] and [12].

2.2.6 Data Fusion in Automotive Applications

Different approaches in the Intelligent Transport System related with Data Fusion are generally divided according to the abstraction level. As previously mentioned, this division is sometimes differentiated among three levels (low, medium, and high) or centralized/decentralized schemes. In the present approach, three-abstraction level differentiation is applied, although it can be easily extrapolated to the centralized/decentralized scheme differentiation.

a) Low-level fusion

Low-level approaches try to create a new set of information from different sources. In this topic, stereo system is a well-known fusion system that uses low-level fusion since it receives information from two different cameras and fuses them to create a new set of information that includes depth information (Figure 2.11). This new set of information is called disparity map. In [17] and [18] stereovision is used to perform pedestrian detection over this new set of information, while computer vision based pattern matching methods are employed to give a final decision, such as active contours or probabilistic models.

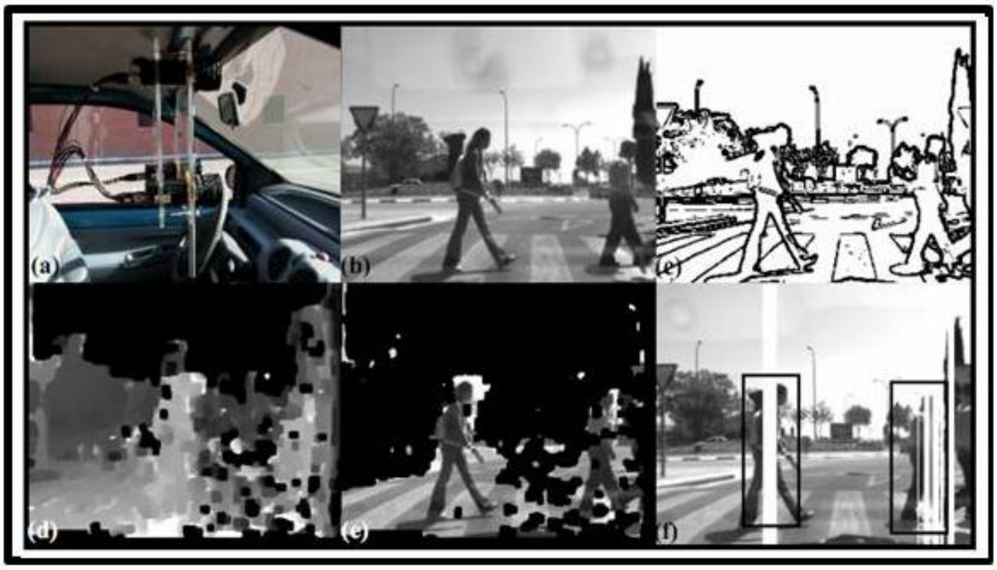


Figure 2.11. Pedestrian ROI detection using disparity map. (a) stereo system used.(b) image of one of the sensors. (c) binarized image. (d) disparity map. (e) close obstacle detection. (f) Bounding boxes representing the pedestrians. This image was taken from [18].

b) Medium-level

In medium-level approaches, some pre-processing is performed for each sensor to create a set of features for each one. These sets are combined to create a combined set that is utilized to perform the obstacle detection and classification (Figure 2.12). In [19] and [20], features are extracted for each sensor independently and a new data set is created, and authors present different approaches whether combining or not the different features of the different sensors comparing results. The final classification is performed by five different methods: Naïve Bayes, GMMC, NN, FLDA, and SVM.

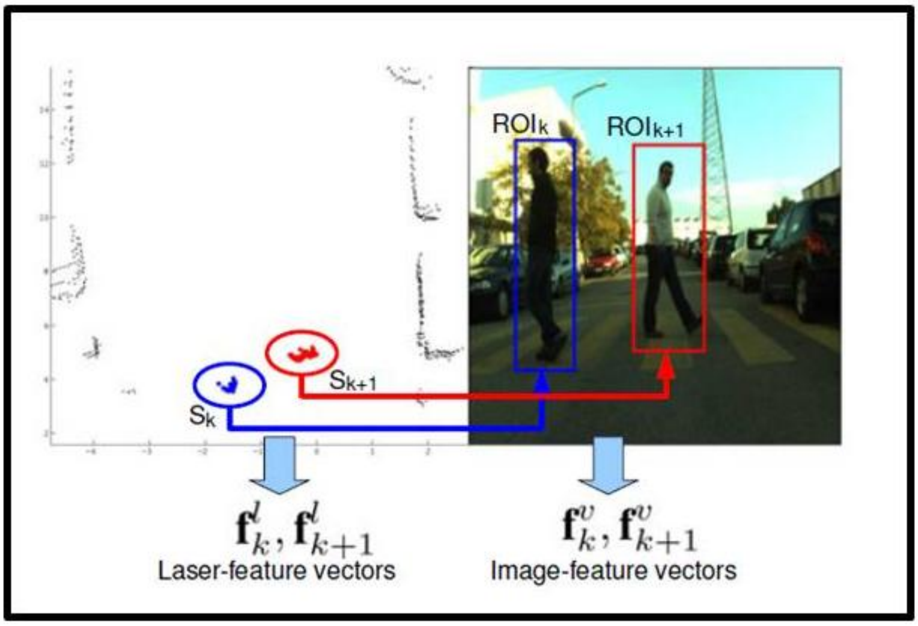


Figure 2.12. Image and laser scanner data set representation, from [19] .

c) High-level fusion

High-level fusion approaches perform detections and classifications for each sensor independently, and a final stage combines the detections according to the certainty of the detections and sensors accuracy. [21] uses Adaboost vision based pedestrian detection and Gaussian Mixture Model classifier (GMM) for laser scanner based pedestrian detections. Finally, a Bayesian decisor is used to combine detections of both subsystems. In [22], pedestrians are detected by a laser scanner using multi-dimensional features, which describes the geometric properties of the detections and Histograms of Oriented Gradients (HOG) features and Support Vector Machine (SVM) for pedestrian detection using computer vision. Final fusion is performed by a Bayesian modeling approach; similar approach is presented in [20], among some other medium level approaches. In [23], track to track fusion of obstacles, with no obstacle classification, is performed, using laser scanner, vision, and long and short range radars.

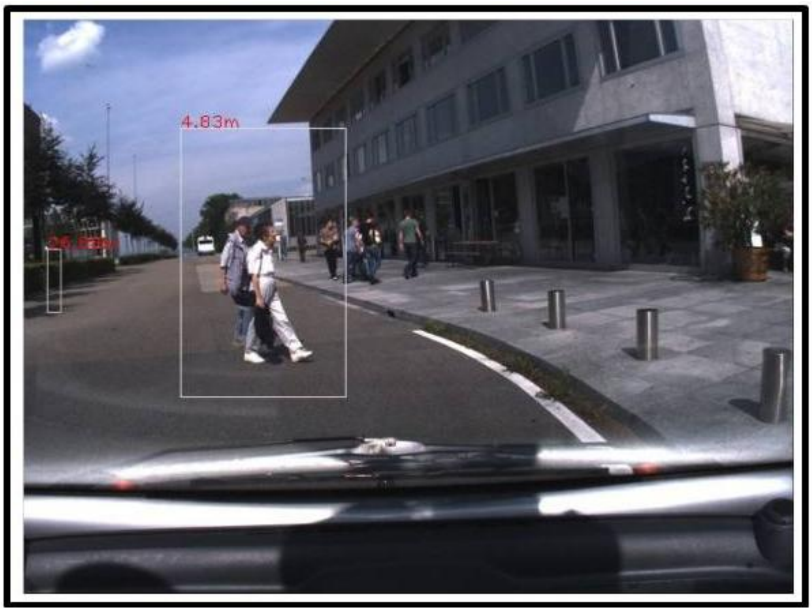


Figure 2.13. High level fusion based pedestrian detection in [22].

d) Other approaches

Other Data Fusion approaches, among Intelligent Vehicles research, use data from a laser scanner to detect regions of interest (ROI) in images, and computer vision to classify among different obstacles that are included in these ROIs. In [24] (Figure 2.14), raw images with SVM machine learning method is utilized. [25] uses Convolutional Neural Networks. [26] uses HOG features and SVM classification approach. Finally, [27] uses Invariant features and SVM to perform the vision-based pedestrian detections. These approaches take advantage of the trustability of the laser scanner for obstacle detection, however, fusion is limited to speed up the process by detecting robust ROIs. Consequently, the information added by the fusion process is limited and can hardly be considered real Data Fusion.

Some fusion approaches take advantage of the properties from various sensors in a different way, which does not fit with any of the previously presented configurations:

[28] combines information from a stereovision camera and a laser scanner. First, the application uses stereovision information to locate the road. Then it uses this information to remove those obstacles that are irrelevant for the application (i.e. outside the road). Finally, It constructes a set of obstacles using the information from both sensors. Tracking is performed using a Kalman Filter approach (Figure 2.15).

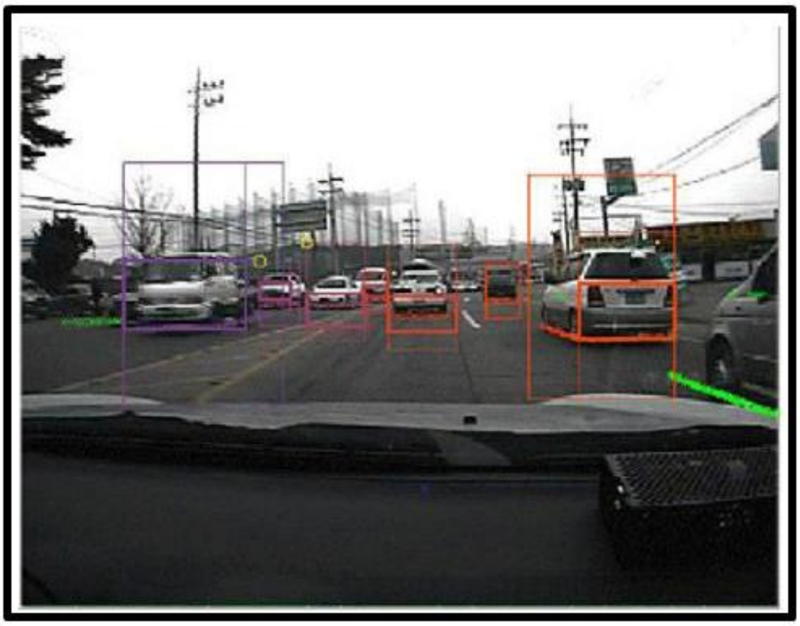


Figure 2.14. vehicle detection based on laser scanner ROI detection and SVM vision based classification [24].

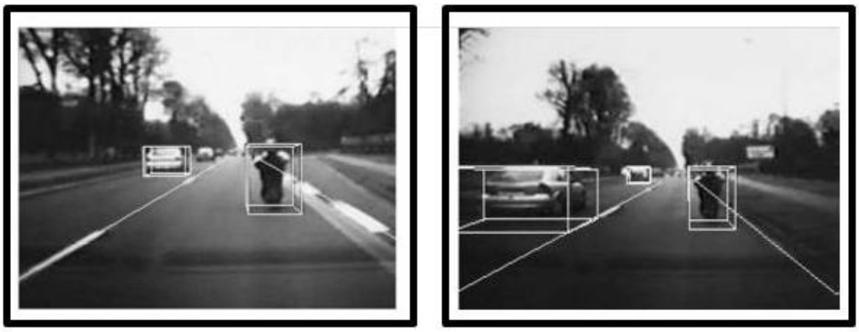


Figure 2.15. Cooperative fusion between vision and laser scanner detection in real road situation, from [28].

[29] uses information from laser scanner to search particular zones of the environment where pedestrians could be located and visibility is reduced, such as the space between two vehicles, and performs detections using vision approach (Figure 2.16).

[30] uses laser scanner and radar approach for obstacle detection and tracking as well as a camera to show the results. Obstacle classification only differentiates among moving and non-moving obstacles through computing Mahalanobis among the clusters given by the laser scanner (Figure 2.17).

Figure 2.16. Pedestrian correctly found in danger zones from [29] .

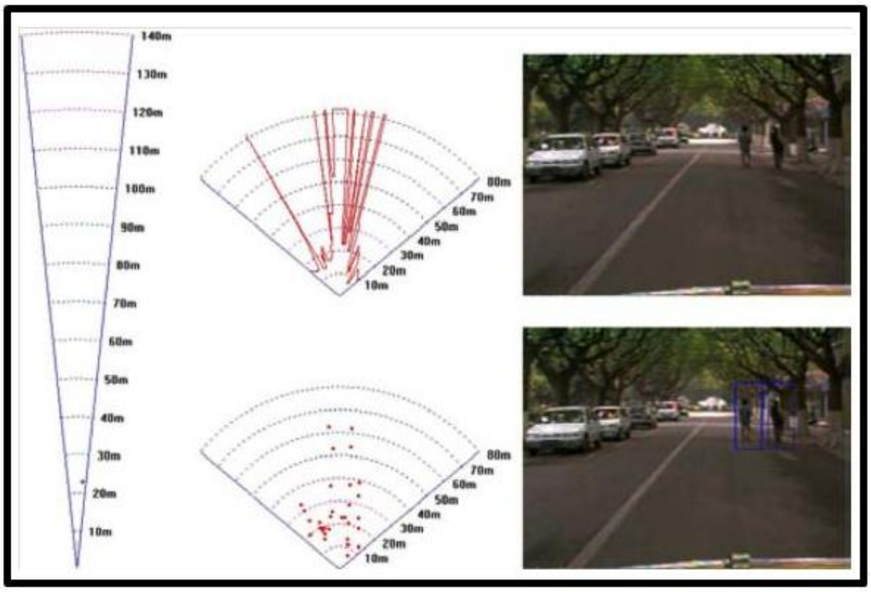


Figure 2.17. Three sensor detections given by the software interface in [30].

Some authors also presented grid-based fusion. Fusion is performed dividing the detection space in a grid space, fusing the information based on the sensor’s accuracy and detection’s certainty. [23] and [31] (Figure 2.18).

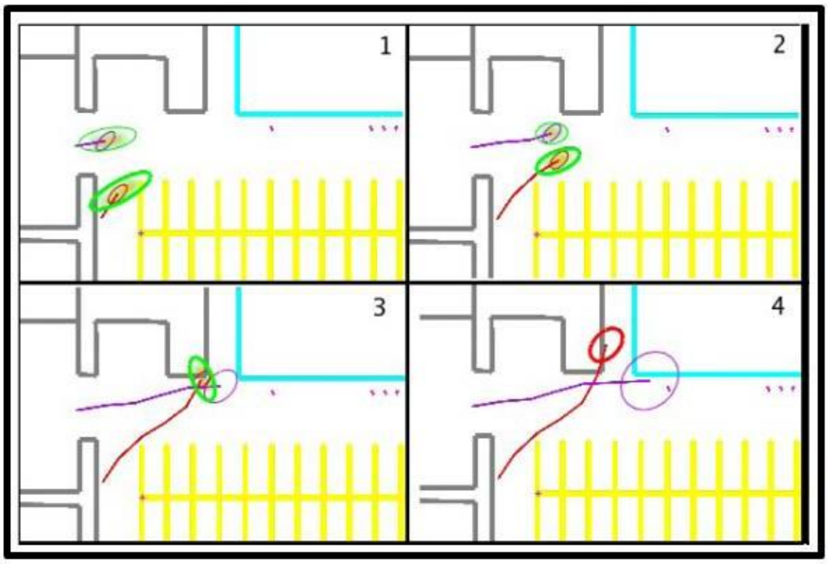


Figure 2.18. Grid based detection and tracking in [31].

2.3 Sensing Technologies in Intelligent Vehicles

The previous section relates the importance of Fusion to provide reliable detections, thus able to fulfill the safety requirements that any road application would need. Now that the importance of the Data Fusion and the key points of the Data Fusion problem are known, the next section will discuss the set of sensors available to use in Intelligent Transport Systems applications. A complete study of the different sensing technologies was mandatory before selecting the correct sensors to be used for the approach presented in this work.

Laser scanner, computer vision, and radar are the most common sensors in Intelligent Transport System technologies. Each one of them presents some advantages and disadvantages as well as different configurations.

2.3.1 Lidar Environment Detection

Laser scanner technology is a well-known sensor in robotics research. The most recent events based on driver technologies, such as DARPA Grand and Urban Challenge ([32], [33], [34], [35] and [36]), have demonstrated that laser scanners are trustable and versatile sensors. They are capable for use with modern ADAS applications [37], [38] and [39]. The main drawback of laser scanners is the lack of information that makes obstacle identification an easier task. As already remarked, the use of a laser scanner in road environments is a recent advance, thus, research is focused on this technology to enhance the performances of the road applications. Taking this into account, the first steps on this thesis were taken in the research and develop of a tool that allows detecting and identifying the different obstacles that are found in road environments using laser scanner. In this section, different applications that use laser scanners with this purpose are presented. Chapter 4 supplies a detailed explanation of the algorithms utilized to detect and identify the different obstacles using a single layer laser scanner.

Laser scanner applications in road environments cover different requirements for safety purposes. Laser scanner can be used to map the surrounding scenario or can even be useful for Simultaneous Localization And Mapping (SLAM) applications [40], [41], [42], [43], [44], [45] and [46].

Mapping applications can be performed both in 2D or 3D, depending on the sensors and their capabilities. 3D reconstruction has a problem of dealing with a high amount of data. This leads to laborious tasks with highly computational costs, mainly when algorithms are based on pattern matching [47] and [48]. Occupation grids can cope better with these costs, but they are free space oriented, thus having more problems with obstacle classification [49] and [50].

Model based laser scanner classification uses a model to arrange the segments founded by the sensor. [51] classifies clouds of points according to certain constraints between points. Lately, it uses Kalman Filter to predict their movements. [52], [53], [54] and [55] classify segments according to an historic record of the movement from different obstacles and those obstacles' behaviors. [56] integrates the patterns detected over time to give an estimation of the real shape of the obstacle; it also performs the detections according to obstacle and certain constraints. [40] uses the Localization algorithm to improve the detection and classification based on pattern matching. Finally, [57] performs pattern classification according to the morphology and occlusion. Next it uses a voting scheme; classification is improved along time, therefore giving an estimation of the certainly of the classification.

In other works, classification is performed according to feature's vectors. [58] uses reflectivity and size of the objects to create the feature's vector. [59] uses these features to create a voting scheme with a weight associated to each feature, to classify the obstacles. [60] uses the features to create a probability density function for each class, which is used in a Bayesian filter to determine the type of the obstacle. [61] creates the features vector using size information from the laser detection and video data.

Once the obstacle is detected, a tracking stage helps to avoid false positives, keeping the detection among time, and improving the accuracy of the system. Kalman Filters are usually used as in [62] and [63], while other tracking schemes can be used as particle filters as with [64] or others [65], [66], [67], [68] and [15].

One of the problems related to laser scanner detection and classification is occlusion due to unforeseen obstacles in the road. Laser Scanners beam is a ray that can be easily occluded by any other obstacle within road environments, especially in 2D laser scanner where information is limited to a single detection. [57] and [69] try to solve this problem.

2.3.2 Vision Environment Detection

Computer vision algorithms try to identify known patterns in digital images. During the last decade, advances in the processing power of computers made it possible to create more complex and robust algorithms able to infer the variety of objects that can be found in images.

The present section presents different algorithms available for computer vision detection and classification in road environments. Efforts in providing new algorithms for computer vision in ITS are numerous. This thesis presents only the most relevant algorithms presented and tested in recent years.

Computer vision techniques can be divided according to the acquisition device used e.g. single camera, stereo-cameras, and infra-red cameras. The processes necessary to perform object detection and classification are usually the same (Figure 2.19), changing the techniques used according to the information provided by the device and the object to identify.

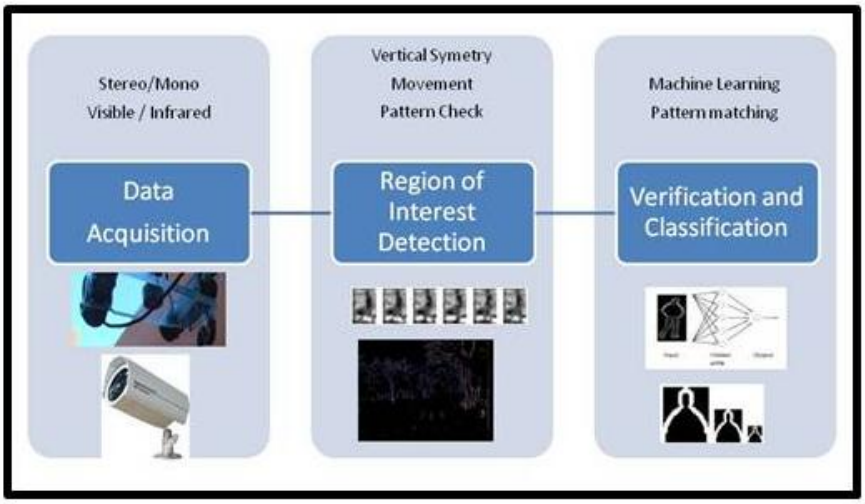


Figure 2.19. Processes of computer vision detection and classification algorithms.

The common processes to follow are depicted in Figure 2.19 and are detailed in next section:

a) Data acquisition

The Data Acquisition process is the procedure of acquiring the image from the device. The various devices available acquire different information. Therefore, further processes are determined by this step, according to the information retrieved from the camera and represented in the image. Several devices can be found:

Mono cameras

The advantages of a mono camera, compared to other devices, are the simplicity of the approach and the low cost. These advantages help the existence of numerous algorithms, which use a mono camera that can be easily extrapolated to the specific problem of road safety. Within this category, several options are available (b/w, color, etc.).

Stereo-cameras

With two cameras located at known intervals, it is possible to add depth to the information. Depth information is provided by detecting the difference in the pixels of the images of both devices. This difference is represented in the disparity map, which represents the disparity of the pixels for the same feature between the two images. With this new information, more robust algorithms can be developed. The extremely high computational cost required by this process for this new information is a drawback. New approaches try to add the recent parallel programming algorithms to the classic stereovision based vehicle and pedestrian detection algorithms. This parallel processing helps to perform the detections in real time. It is important to remark that all these applications consist of safety application where real time is a key issue [70].

Some approaches enhance the capabilities of these stereovision based algorithms by adding more cameras that provide different fields of vision. [71] presents an original configuration of stereo system with three cameras; the system selects according to the region in the image to cover which is the best pair of cameras to use. Figure 2.20 shows the different sensing devices used by vislab in the DARPA Urban Challenge In [71], including the trinocular system.

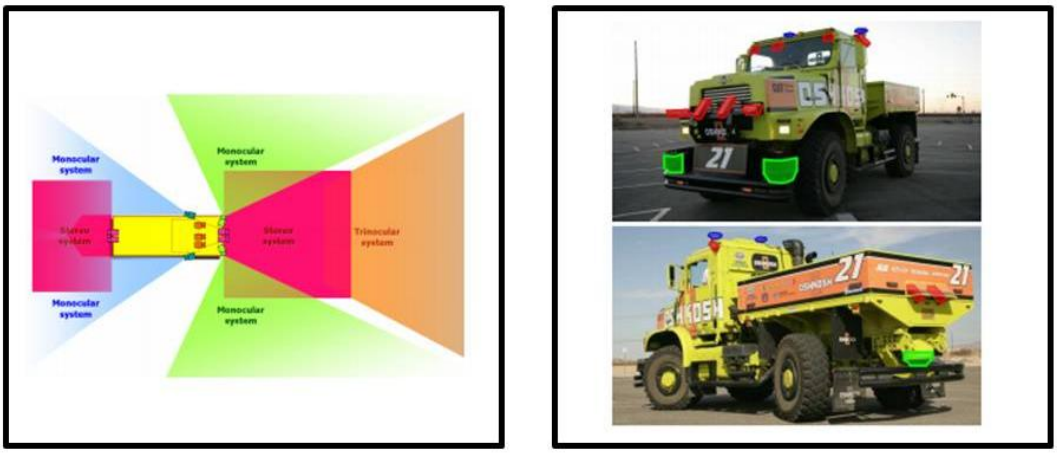


Figure 2.20. Sensing technologies used by Vislab, laboratory of Parma (Italy), during the Urban Challenge in 2007, with several configurations of cameras, including the novel trinocular system. Left the overall field of view, right the images of the truck with the sensing devices highlighted [71].

Infrared cameras

Night vision is a common issue when dealing with Intelligent Transport System since the absence of light makes driving at night more difficult. These situations make Advance Driver Assistance Systems very useful.

Classic computer cameras use the visible spectrum to capture images. Infrared cameras perform detection in the infrared spectrum, which allows them to perform detections, even in situations with no light available. Far Infra Red (FIR) cameras are common among ITS works because these cameras detect heat. Thus, for pedestrian detection, as well as other vehicle detection, the heating information is very useful for detection in the absence of light.

There are some drawbacks in this technology. In the case of FIR, when the temperature of the background is similar to the obstacle’s temperature cameras (daylight situations), it is impossible to differentiate them. Moreover, the absence of color limits some of the capabilities of the cameras.

Some approaches also use a stereovision system based on FIR images, so the information provided by the FIR cameras is enhanced with depth information [72].

b) Region of Interest detection

After data acquisition, it is important to reduce the parts of the image where the classification algorithm is performed. As it was previously remarked, computer vision is a task that requires a high computational cost; therefore, it is important to reduce the region of the image where the classification algorithm is applied.

Region of interest detection consists of identifying the parts of the image where obstacles are more likely to be found. The algorithms are based on the nature of the device or on fast and simple algorithms that do not consume resources. Accordingly, they can quickly reduce the amount of data to search.

Region of interest is also important to reduce false positives. For example, in Fusion procedures, some authors use the reliability of the laser scanner to create a region of interest. In this way, parts of the image where obstacles are impossible to be found are eliminated from the algorithm. Hence, checking only regions where there is an obstacle reduces false positives.

The Region of interest detection is closely related with the device used:

Mono camera:

In mono camera approaches, region of interest detection is a difficult task; no special feature of the acquisition device can be used to help create these Regions of Interest.

Some authors proposed vertical symmetry, which can be quickly and easily calculated, to create these regions. This takes advantage of the vertical symmetry of the pedestrians' and vehicles' shapes. Consequently, only those regions in the image that have a high vertical symmetry are used in the classification stage [73].

Other approaches used the movement detection, such as subtraction of consecutive images, to detect those parts of the image where there are changes. These approaches, although fast and robust, have limitations, including the fact that non-moving objects are discarded in this stage[74].

Stereovision:

Stereovision approaches provide depth, which can be used to detect obstacles as well as to sort obstacles with similar shapes to pedestrians [75]. After calculating the disparity map and detecting the obstacles in the image, it is possible to use this disparity map to detect vertical symmetry, as it is used in monocular approaches, but with higher reliability, as explained in [18].

Far Infra Red

In FIR Cameras, ROI detection approaches usually take advantage of the temperature difference between the interest obstacles and the environment, like a pedestrian's body temperature difference from the background. Thus, thanks to the special behavior of the device, it is possible to detect these regions in an efficient way [76].

Figure 2.21. ROI detection using temperature with fir image [76].

c) Classification and verification

The final stage consists of the verification that the algorithm has identified the search obstacle. In computer vision, classification is typically performed by two different approaches: machine learning algorithms or pattern recognition. The first requires a stage of processing information from the image, creating feature vectors that represent the information given by the image. This way the image is simplified, reducing the amount of data to process and allowing the system to focus on relevant information. Second approaches try to match the obstacles in the image with models that represent the obstacles.

Machine Learning

The machine learning method's goal is to automatically learn to recognize complex patterns and make intelligent decisions based on training data used to "set up" the system and the input data. The set of all possible solutions is too large to be covered by the set of solutions used to train the algorithms, causing difficulties. As a result, selecting the applied training set to teach the system to "learn" how to infer the results is vital and should be accomplished with special care. Classical algorithms used for vision based detection in road environments are:

Decision Tree learning algorithms, such as Adaboost, are typical for face detection. In [77], Adaboost approach was used to detect the driver's face, analyzing the driver's behavior and the physical conditions, to create a warning in the event of carelessness, such as inattention or drowsiness. This Adaboost scheme for face detection was proposed by [78] and combined with

Haar-Like features to perform facial detection. Haar-Like features have also been used to detect vehicles in [79] and pedestrians in [79] and [80]. Deeper Explanation of Haar-Like features is given in chapter 4. It is the method used in the present approach for visual based vehicle detection.

Neural Networks (NN) is a method that is based in the biological neural networks function. These NN are based on interconnected neurons. They are used to represent nonlinear statistical data, modeling complex relations between inputs and outputs, patterns in data, or to capture the statistical structure of an unknown joint probabilistic distribution between observed variables. NN have been used to perform pedestrian detection in [81] and have utilized stereo cameras in [25]. NN have also been employed in traffic sign detection approaches [82].

Support Vector Machines (SVM) is one of the most common machine learning algorithms in recent approaches. SVM construct a hyper-plane in a high or an infinite-dimensional space, which can be used for classification, regression, or other tasks. They are a set of supervised learning algorithms used for classification and regression. The SVM algorithms automatically create a decisor that labels a given input according to the training set that was provided. SVM is used with HOG features to perform pedestrian detection in the most common approach in the ITS community [83]. Further explanation of this approach is given in chapter 4.

It is important in any machine learning approach to choose the appropriate features to perform an accurate classification.

Pattern recognition

Pattern based detections consist of the correlation between the region of interest and a given pattern. The algorithms to perform this correlation are numerous (e.g. normalized correlation and model based).

In [84], geometric models are used, checking if the given ROI fits the constraints that represent the vehicle. In [85] and [86], probability templates are used. These probabilistic models, where each pixel’s intensity represents the frequency of discovery of the pixel in a pedestrian detection, are created by a learning algorithm that uses several test images with pedestrians (Figure 2.22). In [86], different models are created according the pose of the legs of the pedestrians. Subsequently, the authors can infer and track the pedestrian movement.

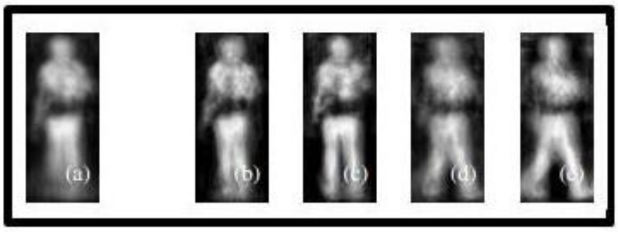


Figure 2.22. Models of the pedestrians created in the approach [86]. (a) Is the general model, (b) the model for closed legs, (c) almost closed, (d) almost open and (e) open.

2.3.3 Other Sensing Technologies

Among the sensors used for automotive application, several sensors can provide different information useful for Intelligent Transport System.

Commercial Systems based on radar application are already available, as with Stop and Go and cruise control systems. But the scope of these applications is limited because radars require a highly reflective surface for detecting, thus non-metallic objects like pedestrians and cyclists are difficult to be detected by radar.

Other sensors used in commercial applications include ultrasound sensors and infrared sensors. The scope of these sensors is very limited. In the case of ultrasound sensors, once the vehicle reaches a certain velocity the sensors malfunction due to high air pressure. In addition, the sensor's field of view is limited to only a few meters. Automotive applications use infrared sensors to detect involuntary line crossing, among other applications. Although all these sensors have a limited scope, the low cost and the reliability of them make them very useful in specific situations, such as parking maneuvers and line crossing detection.

2.3.4 Context Information

Context information is a modern concept in Data Fusion scientific field.

One definition of context information in Data Fusion field can be found in [87]:

"Contextual Information is that information that can be said to "surround" a situation of interest in the world. It is information that aids in understanding the (estimated) situation and also aids in reacting to the situation, if a reaction is required."

Although ITS applications have never used the concept of context information, it is important to take into account the prior knowledge of the application's nature in any Fusion application. This includes information about the detectable objects, physical information, behavior, etc., as well as about the environment, e.g. road safety regulation, driver response time, and more. All this information is useful to perform an accurate detection and also to infer the danger that is involved in all of these detections.

2.4 Conclusion

The state of the art chapter was aimed as an introduction to DF and ITS worlds. First, it was a classical researching field that has just entered into a newer world of ITS, creating a whole researching field thanks to the necessity of accurate and reliable sensing technologies. Second, it portrayed a recent concept that links modern information technologies with transport technologies, helping to improve the efficiency and safety of the classical transport systems. The chapter proposed an overview of the two fields, qualifying the reader with a proper context to understand the contributions of the present thesis and the background where it is situated.

As DF is a innovative method in the ITS field (in the scope of the present thesis), the first part of the chapter focused on introducing the reader the DF world, providing the basic knowledge and processes needed to follow in any DF application. Later, the second part of the chapter introduced cutting edge sensing technologies for Intelligent Vehicles. An Explanation of the most popular algorithms that make use of these technologies was provided.

All these advances represent important steps forward in the field of ITS in providing a reliable tool able to assist the driver by preventing dangerous situations. However, these applications all have limitations due to the nature of the sensors used. This thesis represents a step forward by merging two of the most cutting edge technologies in the ITS field. Through this fusion, the problems inherent in each technology can be overcome.

The proposal provides novel approaches that enhance the capacities of the basic vision algorithms already available. First, it must be stated that at the beginning of the work, the laser scanner technology was starting to be applied to road applications. Thus, the first part of the proposal consisted of creating novel and reliable laser scanner based algorithms that could help to improve the reliability of the vision-based algorithms. Later, the fusion procedures attempted to combine the capacities of both sensors to provide better performances. Again, it must be related that the visual approaches used in the present thesis are the most reliable approaches of those presented. HOG feature based pedestrian detection represents the most reliable approach among the wide variety of works available in the ITS world. Finally, Haar-like vehicle detection is a robust and fast approach that provides reliable vehicle detection with low requirements. Furthermore, the use of the original concept of context information helps to provide a complete fusion application that fulfills the requirements of classical Data Fusion approaches.

CHAPTER 3. GENERAL DESCRIPTION

As previously explained, the present thesis proposes an innovative system that combines different information sources (i.e. laser scanner, computer vision, inertial system and context information) to provide a reliable system, capable of providing road environment detection and classification. This detection is focused on the most vulnerable road users i.e. pedestrians and other vehicles. The proposal also tries to solve a typical problem when dealing with traffic safety applications: the danger associated with the detections.

In the present chapter, the overall description of the proposal is given, including data sources, data flows, and the processes that involve the fusion process (Figure 3.1).

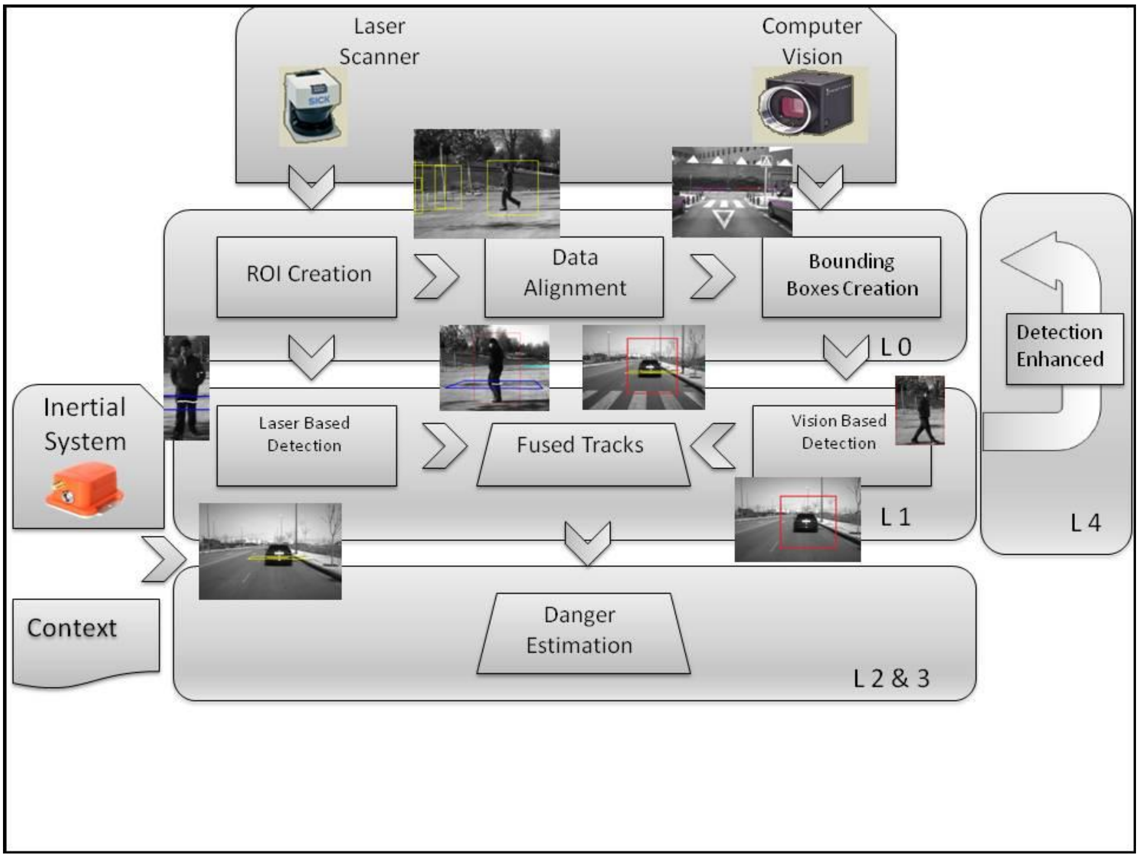


Figure 3.1. Overall System diagram with the corresponding fusion levels implementation.

Figure 3.1 summarizes the overall system and indicates the processes that are necessary to fulfill all levels of the fusion procedure. As shown by the figure, there are four well-differentiated data sources. Two sensors perform target identification and classification; these sensors are the laser scanner and the computer camera. Context Data and intertial system complete the information, adding contextual information allowing Data Fusion at all levels.

3.1 The Platform

Intelligent Vehicle based on Visual Information (Figure 3.2), or IVVI 2.0, is the second research platform for the implementation of laser technology and computer based systems, with the goal of building ADASs. The purpose of the IVVI platform is to test perception algorithms under real conditions. Different sensing capabilities are being researched, such as road lanes, pedestrians, vehicles and traffic signs. They can be taken as inputs for some ADASs like Lane Keeping System, Adaptive Cruise Control, Pedestrian Protector, and Traffic Sign Recognition.

Figure 3.2. Test Platform IVVI 2.0 with the laser scanner mounted in the front bumper.

In the scope of the IVVI 2.0 project, several technologies have been researched and developed, and some of them are already available in the platform. The different sensing devices already available for testing and developing the different sensing technologies are depicted in Figure 3.3.

(a) A color camera, pointing to the inside of the vehicle, used to monitor the driver, warning in case of misbehavior or somnolence.

(b) A small monitor embedded in the dashboard of the vehicle, which allows checking the results of the different algorithms in real time.

(c) Laser scanner, used in present work.

(d) A far infrared camera that allows detection in absence of light.

(e) A color camera, pointing to the exterior, used for traffic sign detection and classification.

(f) Bumblebee stereo system, used for obstacle detection and classification.

(g) GPS-inertial device that is used to acquire information about movement of the vehicle.

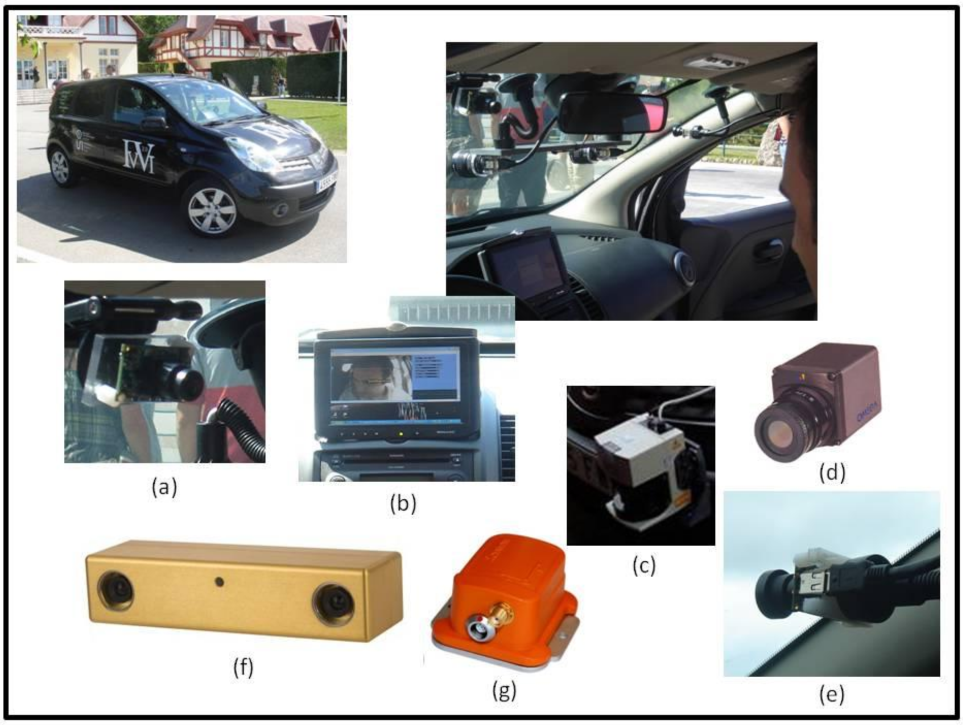


Figure 3.3. Sensor devices installed in IVVI 2.0

These devices allow research and development of different algorithms intended for use in ADAS applications. But to process such demanding applications, several processing units are necessary. Figure 3.4 shows some of the available processing units.

(a) Wi-Fi router installed in the bumper of the vehicle, which allows network connection within the computers and with an exterior network through wireless Internet connection.

(b) Different PDAs and Smartphones are connected to the processing units to allow direct contact with occupants.

(c) Batteries, inverters, and a backup system supply different devices with power.

(d) Computers are installed in the vehicle's trunk. These computers are continuously being updated to have the most effective processing capacity.

(e) Can-Bus readers allow the vehicle's status to be monitored.

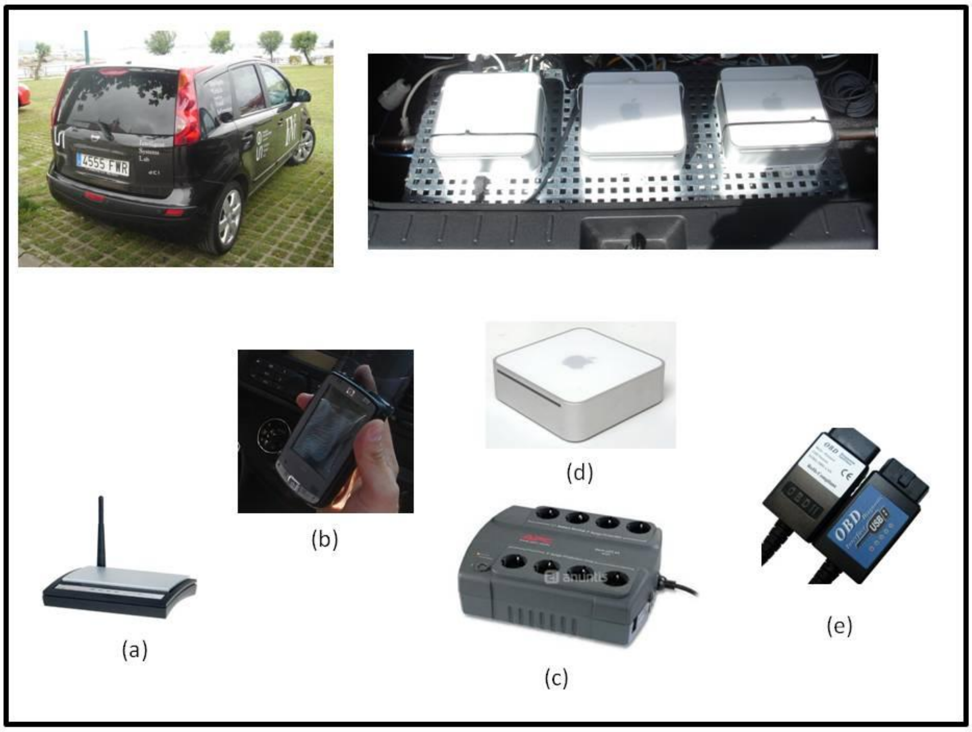


Figure 3.4. Processing units and other devices installed in the platform IVVI 2.0.

Some examples of the technologies researched and developed using the IVVI 2.0 platform are depicted in Figure 3.5:

(a) Pedestrian detection[88],

(b) Driver monitoring [77] under daylight.

(c) Driver monitoring [77] in night conditions.

(d) Lane detection and classification system [89]

(e) Traffic sign recognition [82]

(f) Stereovision based obstacle detection and ego-motion [90]

(g) Fusion procedures, using laser scanner and vision, system presented in this work.

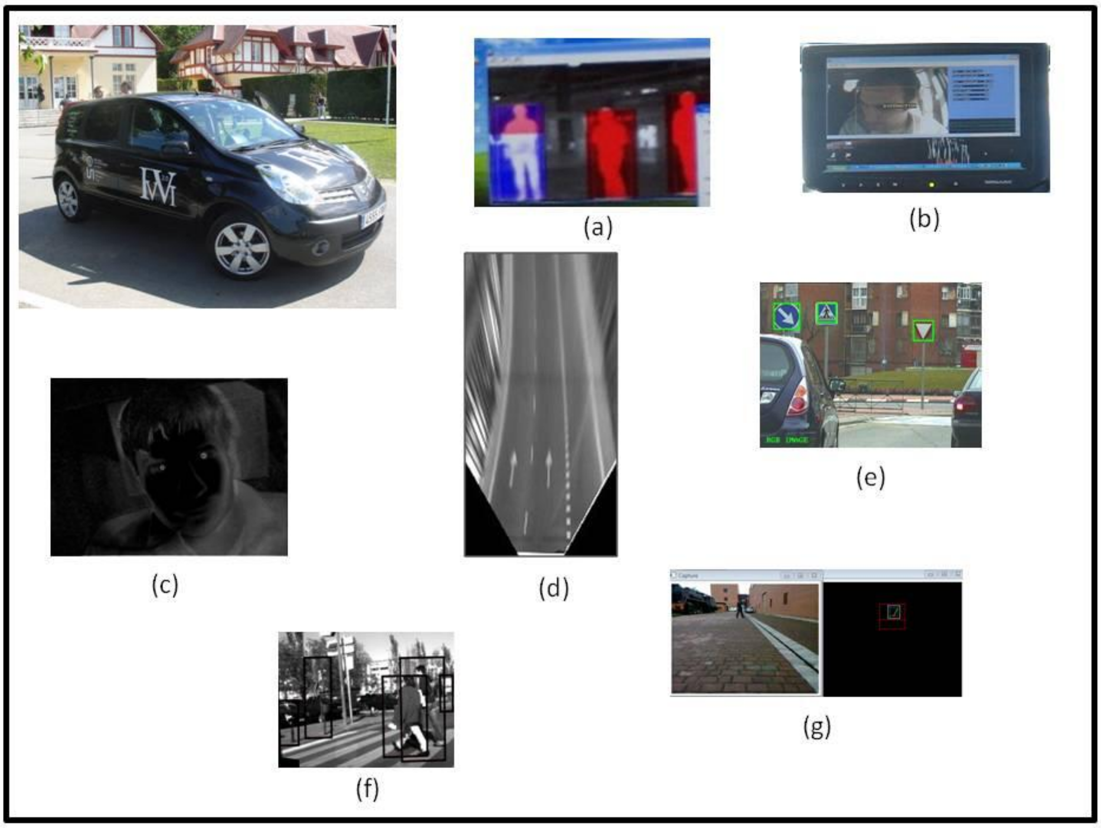


Figure 3.5.Technologies being developed and researched in the platform IVVI 2.0.

3.2 Data Sources

Four different data sources are used in the present approach: laser scanner, computer vision, inertial sensor, and context information.

3.2.1 Laser Scanner

The laser scanner model put in operation was a LMS 291 S-05 from SICK. The laser scanner configuration was 100° of field of view, a maximum distance of 82 meters, and an angular resolution of 0.25°. There is a detailed description of the system and its configuration is given in chapter 4.

The laser scanner employed provides detections based in time of flight operating mode, conveying the distance to the closest obstacles for a given angle. This way, after a full rotation, a 2D reconstruction of the environment is possible (Figure 3.6).

Figure 3.6. Laser scanner environment reconstruction example. Distances from 5,10, 15 and 20 are highlighted in yellow, as well as some bounding boxes with relevant obstacles.

3.2.2 Computer Vision

Computer vision is a common tool in the most recent advances in road safety applications. This thesis uses the most recent advances in visual processing applications for road safety, such as Haar Cascade Classifier [78] or HOG descriptor [83].

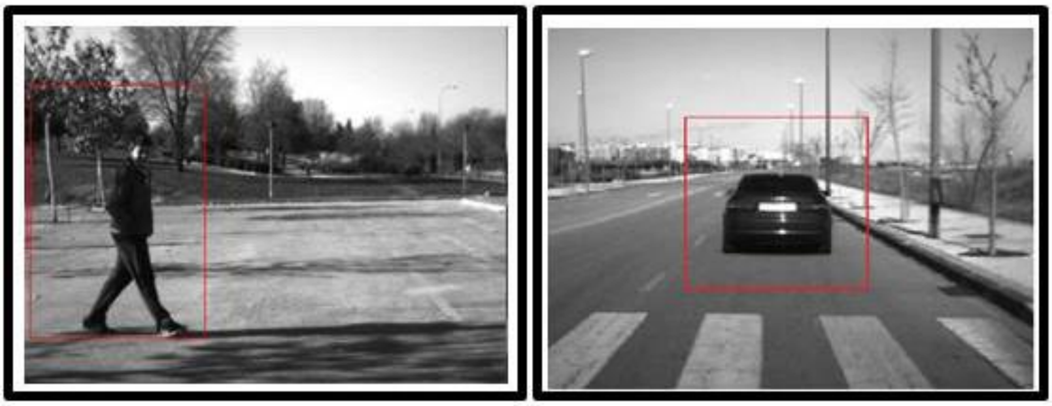


Figure 3.7. HOG Feature based pedestrian Detection (Left). Haar cascade Classifier based Vehicle detection (Rigth).

3.2.3 Inertial Sensor

MTi-G [91] is a compact and low-weight measurement unit for control and navigation of (un)manned systems and other objects. It is an inertial system aided by GPS. The size and weight of the system make it very flexible tool. It is provided with several working scenarios that adapt the unit to aerospace, automotive or general applications. Additionally, MTi-G provides inertial and barometric enhanced 3D position and velocity data at a higher update rate than possible with a typical GPS receiver.

Figure 3.8. MTI-G GPS aided inertial measurement.

MTI-G proved to be a very useful tool for the present application by providing typical GPS information such as velocity and position. This online information was valuable for estimating danger. Besides, Euler angles and angular velocities provided were important for laser scanner detection, as detailed in chapter 4.

3.2.4 Context Information

In present proposal context information was added to complete the understanding of the interactions of the detections with the test platform. This helped estimate the danger involved with detections. In the proposal, the situations where context fusion proved to be useful were the following:

- **Physical constraints** This is used for low-level detection. By applying some constraints to the low level detections false positives were avoided. These include pedestrian sizes, vehicles size, etc.
- **Reaction times**. The time that any driver needs to respond to a stimulus is especially important when trying to calculate the response time for the driver in danger situations.
- **Braking distance**. The braking distance is the distance that it is needed to completely stop the vehicle. This is one of the most important distances when

dealing with road safety applications, since it represents the space where it is impossible to avoid any collision by stopping the vehicle.

- **Safety distances**. This is the distance between vehicles travelling on the road. It is important in safety applications because this distance assures a safe driving.

The relevance of the provided contextual information will be referred to in various chapters of the proposal.

3.3 Fusion Levels

From the Data Fusion researching field point of view, this thesis attempts to solve the problem of road environment detection and classification. Formulating the limits and scope of the different levels of data fusion are significant. This section illustrates a solution to the problem-oriented levels of data infusion. It also provides a detailed explanation of the processes utilized to solve the problems related to the different levels involved in Data Fusion, as based mainly in the scheme of Figure 3.1. Later, the document also presents a detailed explanation of the different processes exposed.

3.3.1 Level 0 (Sub – Object Refinement)

These processes involve methods to estimate the existence of targets and the features of interest related to them. In this case, the level is related to the detection of obstacles either by laser scanner or by camera. Later, level 1 processes should classify these obstacles according to the information available.

In order to reduce false positives and to provide reliable detections, the laser scanner was considered the main sensor for this purpose. This was accomplished because the reliability of the laser scanner was established in the process of providing distances to all the non-occluded obstacles with high certainty. Consequently, in order to provide reliable detections, laser scanner was used to detect obstacles in the surroundings for both, laser scanner detections and visual based detections. The process used was a clustering process that merges detection points according to the distance between them. These obstacles represent the Regions Of Interest (ROIs) that both sensors later classify.

In order to provide accurate positions of the obstacles detections in the laser scanner, data alignment is mandatory. Data alignment coordinates, specified by the laser scanner, will be transferred, first to the camera coordinate system to provide the regions of interest in the image, and later to the vehicle coordinate system. The processes followed to perform this data alignment will be described in chapter 4.

After data alignment, the vision zones, where the algorithms perform the obstacles classification, are created according to the sizes of the investigated obstacles and the algorithm to use, such as when looking for vehicles obstacles with size similar to a vehicle are selected. These zones are called Regions of Interest (ROIs) and are represented by bounding boxes.

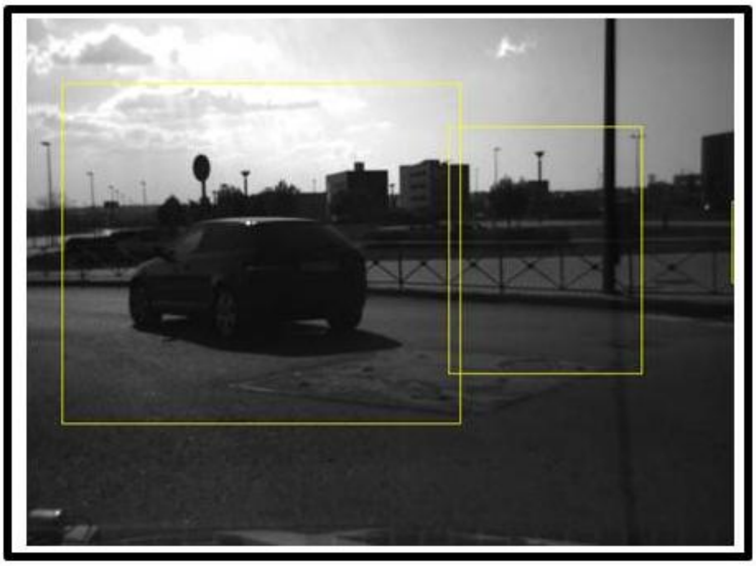


Figure 3.9. Bounding boxes for vehicle detection created after laser scanned obstacle detection.

3.3.2 Level 1 (Object Refinement)

In this level, the location, parameterization and identification of the obstacles are preformed. In the present proposal, these processes are performed first by each system independently, and then both detections are combined to create and manage the detections' tracks obtained from the detections from both sensors. Thus, the aforementioned ROIs are checked by both subsystems, taking into account the information obtained by each sensor, which obtain different classifications. As both regions of interest are acquired from the same set of obstacles detected by the laser scanner, the inter-sensor association is inherent for the algorithm.

Laser Based Detection uses the exceptional behavior of the laser scanner and the robustness of the system to classify among different kind of obstacles. Limited information and some common errors due to strong pitching movements, dust, etc., make this classification difficult, and here is when the classifications provided by the camera could help to avoid mistakes.

Vision Based Detection. For the present proposal two kinds of obstacles were taken into account: vehicles and single pedestrians. These two road users represent the most important users of the roads, and several vision approaches have been created to detect them. As previously mentioned, the algorithms used for these detections are the most well-known and proved algorithms that have presented the best results for the detection of each kind of obstacle.

Finally, the fusion procedure should deal with the detections of each sensor creating **Fused Tracks,** which provide better performance than the sensors independently. These processes involve track management (creation, deletion and update), tracking procedures and data association.

3.3.3 Level 2 & 3 (Situation and Threat Refinement)

Previous levels focused on road users' detection and classification, as well as the tracking the given detections. Levels 2 and 3 deal with safety issues. In level 2, interactions between detections and the test platform are studied. Finally, in level 3, the thread level that any of the detections involves is analyzed.

In order to give an estimation of the danger involved in any detection, important information is added to the lower level information of the targets. First, information relative to the vehicle where the detections are performed is mandatory to help the system evaluate interactions amongst the different users of the roads. Furthermore, traffic safety information is necessary to estimate the danger level involved in all interactions. Contextual information has an important role in evaluating the situation of the detections and interactions among them. Traffic accident reconstruction mathematics and other constraints, such as driver reaction times and pedestrian's sizes, are among the contextual information taken into account to define the security involved in detections. Different defined distances evaluate the danger of detections:

- **Response distance**. The distance the vehicle covers in the time that the driver needs to respond to a stimulus and perform a maneuver.
- **Braking distance**. The distance the vehicle would cover before completely stopping.

- **Safety distance**. The distance that should be maintained with the preceding vehicle to ensure safety.

Full fusion procedure is obtained, providing solutions for all the fusion levels. Chapter 5 provides detailed mathematics used to calculate these distances.

3.3.4 Level 4 (Process Refinement)

Information of the detections performed by the different subsystems is included in low-level detection, such as vehicle or pedestrian sizes, velocities, etc. This information is used to enhance the low-level detection in the subsequent scans.

This level 4 solution does not represent a sophisticated process refinement solution, rather it is useful to improve future detections performed by the lowest level, implementing all fusion levels.

3.4 Proposal Phases

The phases for a complete Data Fusion safety application will be detailed in this thesis. The proposal will have to deal with all the levels of fusion applications. These phases are based on classic Multiple Target Tracking applications phases, adding danger estimation. The phases include the following:

- Obstacle detection
- Multiple Target Tracking
- Danger Estimation
- Test and conclusions

The subsequent chapters detail the different processes followed to accomplish these tasks.

3.4.1 Obstacle Detection

In this section, the description of the algorithms utilized for obstacle detection and classification are provided, focusing on laser scanner algorithms. Laser scanner algorithms used in the proposal are innovative methods that represent an important contribution to the pedestrian detection algorithms available in the ITS research field. Conversely, vision based algorithms represent state of the art pedestrian detection, even though they are not original contributions to the field. As a result, explanations of these approaches will be minor.

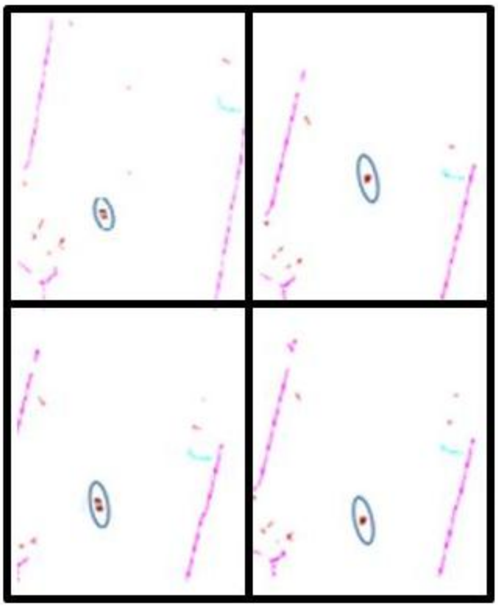


Figure 3.10. Laser Scanner based pedestrian detection example.

3.4.2 Multiple Target Tracking

Different methods used for Multiple Target Tracking (MTT) are detailed in chapter 5. These methods deal with fusion of information from both sensors and also with tracking and data association.

Several configurations were tried combining the different possibilities available in multiple target tracking and adapting them to pedestrian detection with data fusion. Tracking procedures used include:

- Kalman Filtering
- Unscented Kalman Filtering
- Particle Filter

Data association plays an important role in Multiple Target Tracking. Several approaches are provided, focusing on the implementation of the fusion system, thus providing a complete and extensive comparison of the different possible configurations. The Association methods used were:

- Nearest Neighbors (NN)
- Multiple Hypothesis Tracker (MHT)

- Joint Probabilistic Data Association (JPDA)

Finally, chapter 5 deals with other problems related with the MTT, such as distance definition, gating, track logic (creation, deletion and update of tracks), etc.

3.4.3 Danger Estimation

Although the main purpose of this thesis is not to provide an accurate danger estimation of the detections, all Data Fusion applications have to deal with situation assessment and threat estimation. For the present proposal danger, estimation was created both for vehicles and pedestrians. As previously mentioned, context information is crucial at this point. A detailed explanation of the different distances and constrains relevant for danger estimation are given in chapter 5. The present work provides a novel approach able to enhance the classic pedestrian detection systems with contextual information. The result is a complete Data Fusion tool that detects and classifies the different road users and estimates the danger that is involved in these detections, providing a solution to all fusion levels.

The actions to perform in case of a dangerous situation are out of the scope of the present work. It is not the purpose of the present thesis for the vehicle to perform emergency stops, avoiding maneuvers, triggering and alarm, etc. Thus, no further information concerning the necessary actions to avoid a detected danger will be provided.

3.4.4 Test and Conclusions

Several tests were performed to check the viability of the different approaches introduced in the present thesis. The different algorithms were checked both in controlled environments and real conditions. Chapter 6 depicts the different tests performed and presents the results. Chapter 7 provides some conclusions obtained from the different approaches proposed in the present thesis, providing a general overview of the contributions of the thesis including the future steps to follow.

CHAPTER 4. OBSTACLE DETECTION AND CLASSIFICATION

In this chapter, low-level detection and classification are detailed. First, the chapter presents the algorithms used to detect and classify different obstacles found in the road by the laser scanner. Later, data alignment is detailed, explaining the process followed to synchronize the detection of both subsystems. Finally, visual classification algorithms used to classify pedestrians and vehicles based in computer vision are explained.

4.1 Laser Scanner Obstacle Detection and Classification

The main disadvantage with the laser scanner is the relatively small amount of information provided. However, it is sufficient to provide a first estimation of the obstacle's detected shape, and it even provides classification results that later visual approaches will attempt to confirm.

The algorithm is composed of two stages (Figure 4.1). In low-level stage, the data is received and a low-level identification is performed. In the higher-level stage, the data is integrated over a specific time period, resulting in a higher classification level. This integration is performed by correlating obstacles in subsequent scans.

4.1.1 Stage 1. Low Level Detection

Low-level detection is composed of four subsystems, each performing a different task that depends on the results from the previous stage.

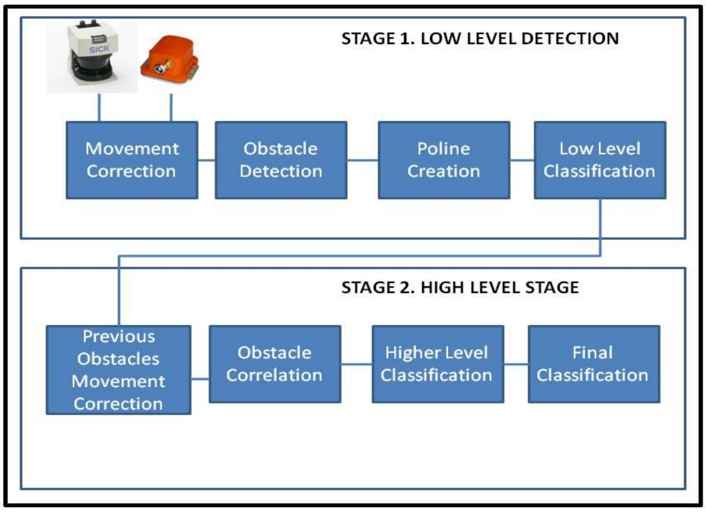


Figure 4.1. Laser Scanner based road environment reconstruction scheme.

a) Movement compensation

The data received by the laser is corrected according to the movement of the vehicle. This is done to avoid misdetection due to time differences between the spots that are part of the scan.

Errors due to strong pitching movements are avoided using the inertial system. Pitch movement is checked. When there is a strong pitching movement, laser scanner detection is disabled to avoid misdetections and errors. In the next section, data alignment is detailed, providing an explanation of the errors that pitching movements and mistakes in the calibration process can cause.

Euler angles detected by the inertial system are used to correct the displacement of the measurements caused by the movement of the vehicle. Equation (1) depicts the compensation with the rotation and translation matrixes needed to correct this movement. This way the points are referenced to the position of the last point received (Figure 4.2 (a)).

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = R(\begin{bmatrix} x_0 \\ y_0 \\ z_0 \end{bmatrix} + T_v + T_0)$$

$$R =$$

$$\begin{bmatrix} \cos(\Delta\delta) & 0 & sin(\Delta\delta) \\ 0 & 1 & 0 \\ -sin(\Delta\delta) & 0 & cos(\Delta\delta) \end{bmatrix} \begin{bmatrix} 1 & 0 & 1 \\ 0 & cos(\Delta\varphi) & -sin(\Delta\varphi) \\ 0 & sin(\Delta\varphi) & cos(\Delta\varphi) \end{bmatrix} \begin{bmatrix} \cos(\Delta\theta) & -\sin(\Delta\theta) & 0 \\ sin(\Delta\theta) & cos(\Delta\theta) & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad (1)$$

$$T_v = \begin{bmatrix} vT_i \cdot cos(\Delta\theta) \\ vT_i \cdot sin(\Delta\theta) \\ 0 \end{bmatrix}, T_0 = \begin{bmatrix} x_t \\ y_t \\ z_t \end{bmatrix}$$

where $\Delta\delta$, $\Delta\varphi$ and $\Delta\theta$ corresponds to the increment of the auler angles roll, pitch and yaw respectively for a given period of time T_i . Coordinates (x,y,z) and (x_0,y_0,z_0) are the Cartesian coordinates of a given point after and before respectively to the vehicle movement compensation. R is the rotation matrix, T_v the translation matrix according to the velocity of the vehicle, T_0 the translation matrix according to the position of the laser and the inertial sensor. v is the velocity of the car, T_i the time between the given point and the first one in a given scan. Finally, (x_t,y_t,z_t) is the distance from the laser scanner coordinate system to the inertial measurement system.

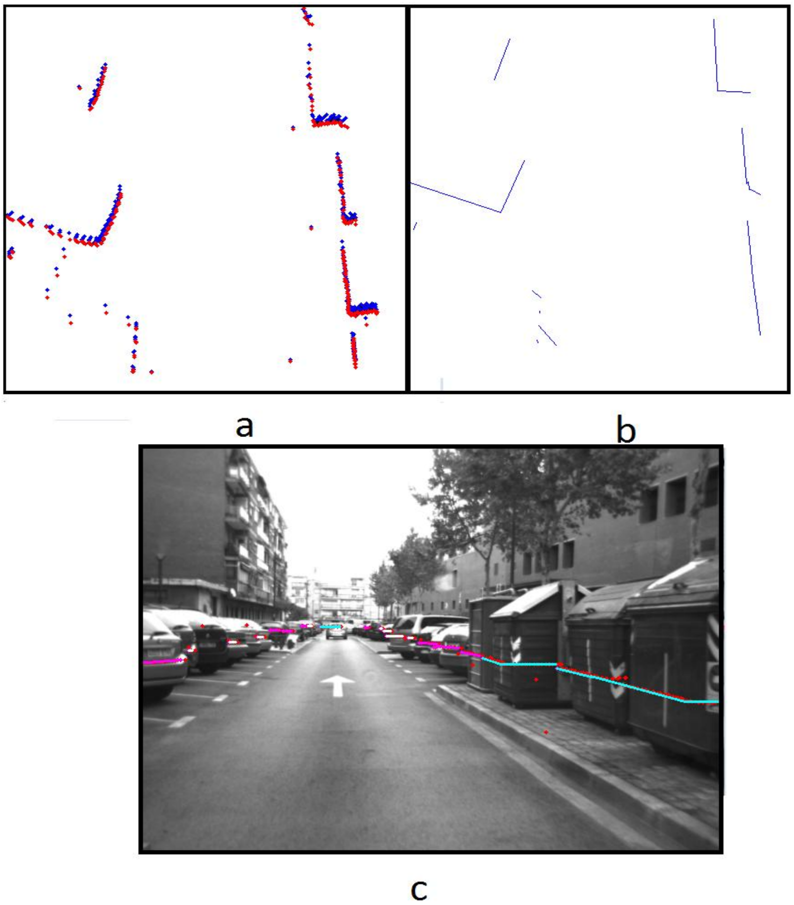


Figure 4.2. Vehicle movement compensation and data alignment. Figure a shows the detection points, in blue before the vehicle movement compensation and in red the compensated. Figure b shows the shape reconstructed after the movement compensation. Figure c shows the final alignment of the laser scanner data and the image.

The clouds of points are clustered using Euclidean distance and a threshold that is distance dependant (equation (2)).

$$\text{th} = th_0 + K \cdot dist \qquad (2)$$

were th_0 is the threshold base and K is a proportional constant which is multiplied by the distance.

Therefore, for a given point $p(x_i,y_i)$ it may be treated as belonging to a segment S_j if it satisfies:

$$p_i(x_i, y_i) \epsilon\, S_j \rightarrow \{\exists\, [\, p_j(x_j, y_j)\ \epsilon\, S_j\,]\text{: } d(p_j, p_i) < th\,\} \qquad (3)$$

The algorithm checks for all of the points, if the case arises where a point is not included within a segment, a new segment is then created. After all the points have been verified, the algorithm searches for segments containing only one point, these are then removed as they are considered false detection points.

In Figure 4.3 an example of segment creation is presented. Here segment A is created, but after the verification process, where only one point is found, this point is removed from the final segments.

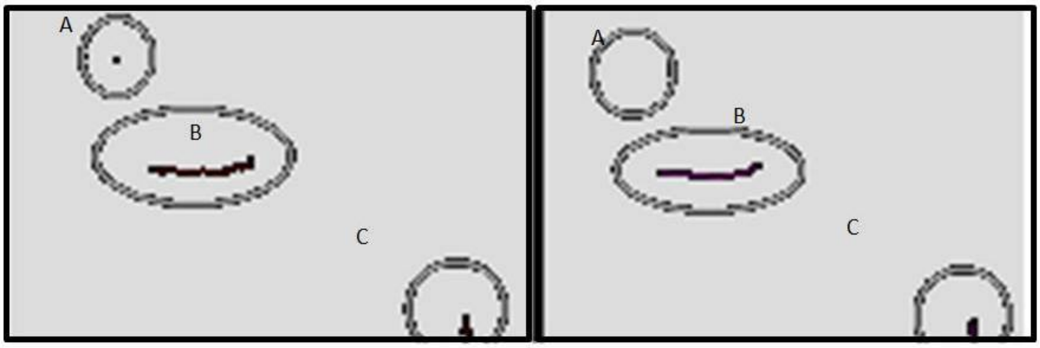


Figure 4.3. Segment creation example based in the distance among detections.

b) Polyline creation

Once the segments are created, the points contained within each segment are merged using lines known as polylines. These lines are merged together according to the distance between the points included. The first and last points are merged using a line. For each point contained within this segment, the distance to the line is computed, and if it is higher than a given threshold, two new lines are created merging these three points. This process is repeated for every point within these new lines.

An example of polyline creation is shown in Figure 4.4. As $dis\big(p_3, Rect(p_1, p_2)\big) < th$□ two new lines are created, $Rect$ (p1,p3) and $Rect$ (p3,p4). This process is repeated for p4. Finally, the resulting form of the polylines is shown. The Point p5 has a smaller distance than the threshold, thus, this point is considered to be a spare point. The final result of obstacle reconstruction is shown in Figure 4.2(b).

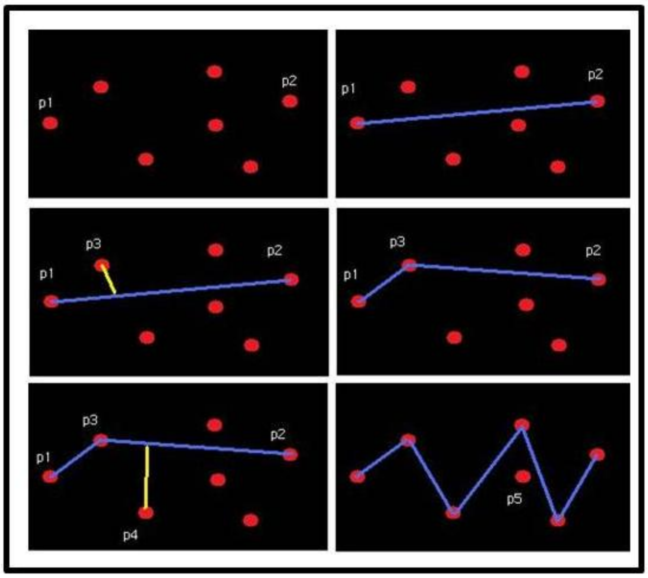


Figure 4.4.Polyline creation example.

c) Low level classification

Low-level classification is performed with the information provided by the previously described stage. Here obstacles can be differentiated:

a) Little obstacles

These are regarded as obstacles in which size is not compatible with that of a vehicle, buildings or any other large obstacles that are commonly found in the road environment. Typically, these kinds of obstacles can be pedestrians, lampposts, milestones, trees, traffic sings, and other small obstacles that can be found in road environments. With pedestrians, a further algorithm will attempt to determine, by comparing the resulting polyline with a pattern, if the obstacle is a pedestrian or not.

b) Road limits

Two different possibilities exist for these obstacles. These two classifications are performed according to the following procedures:

<u>Big obstacles</u>. If an obstacle bigger than a given threshold th is found, it is considered as a candidate for a road limit. The position is checked, and if it is located parallel to the trajectory of the car, it is finally labeled as a road limit.

Other road limits detection. In the first stage, little obstacles are found and labeled. After this first classification, two histograms of little obstacles, which represent the frequency of little obstacles along the x and y axis of the road, are created. If the frequency for a given axis is sufficient, it may be considered that the obstacles found on the road borders can be considered as road limits. If a curve is detected using yaw angle measurements, this detection is disregarded, as road limits are not parallel or perpendicular to the movement of the car. This method, although straightforward, is demonstrated as being fast and reliable for road border obstacle detection in the majority of scenarios.

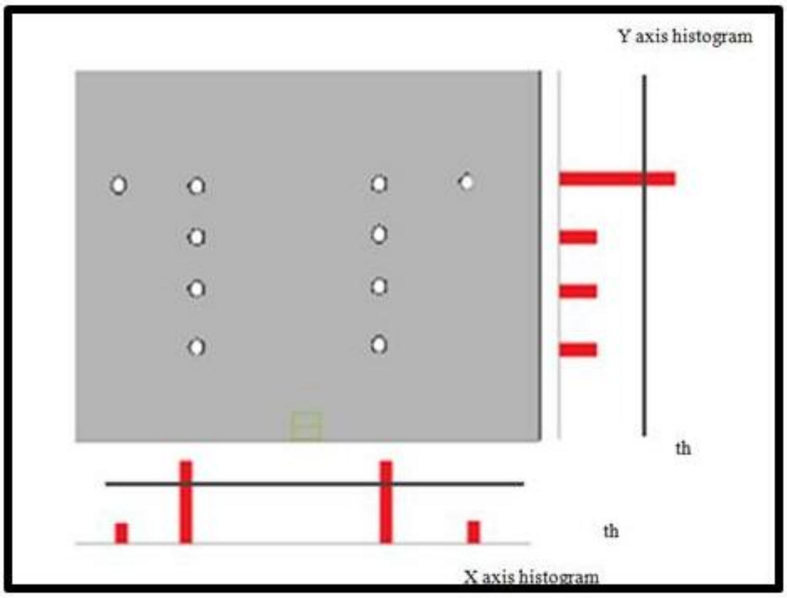


Figure 4.5. Road border detection based in little obstacle histograms. When the histogram is greater than a given value road borders are detected.

c) Possible vehicles

The pattern provided by moving obstacles can be differentiated and used to perform vehicle obstacle classification. These obstacles can be detected using the special behavior of laser scanners. In the case of the model used in this approach with a configuration of 0.25° resolution, it performs 4 independent scans which provide 4 sets of spots with a 1° resolution. Each scan is separated 0.25° with respect to the previous one (Figure 4.6). Thus, after 4 scans, the laser scanner returns a complete set of spots separated by 0.25°. When a moving obstacle is found, the four scans performed by the laser scanner for a single detection appear with a variation that is proportional to the speed and direction of the detected object and the test vehicle.

Once this particular pattern with a serrated shape is found, the velocity of the vehicle can be calculated by measuring the distance between two consecutive points (Figure 4.7) (equations (4) to (7)).

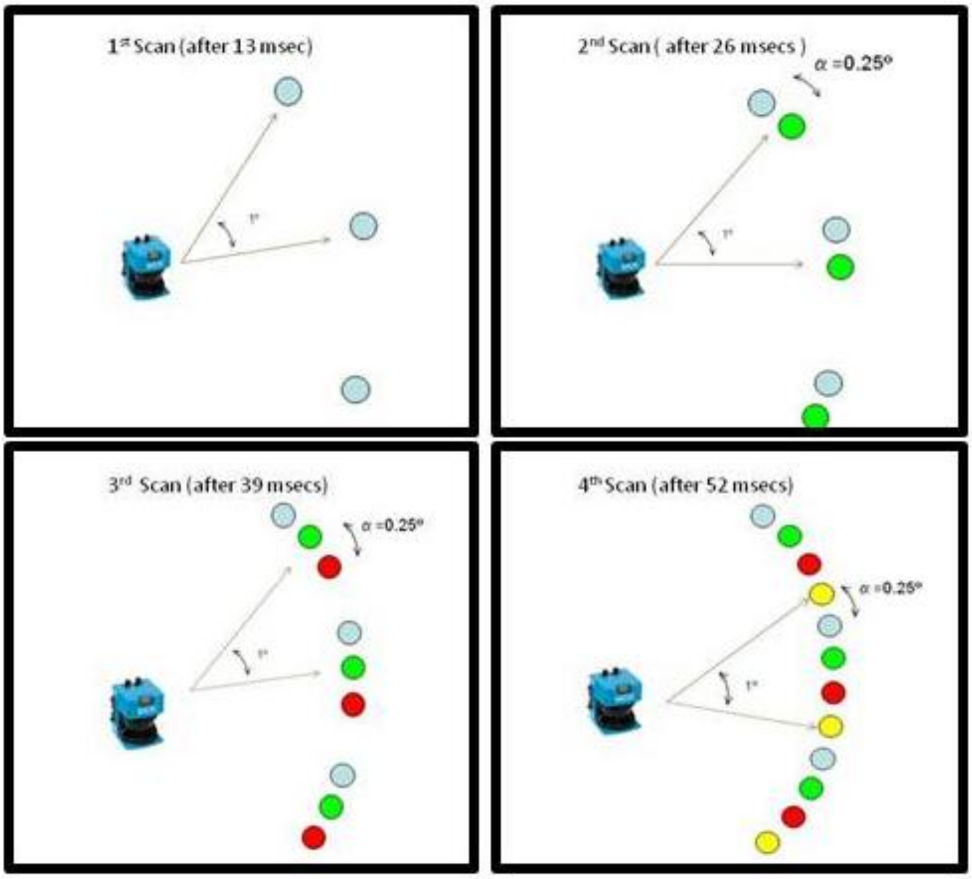


Figure 4.6. Laser behavior for a complete scan.

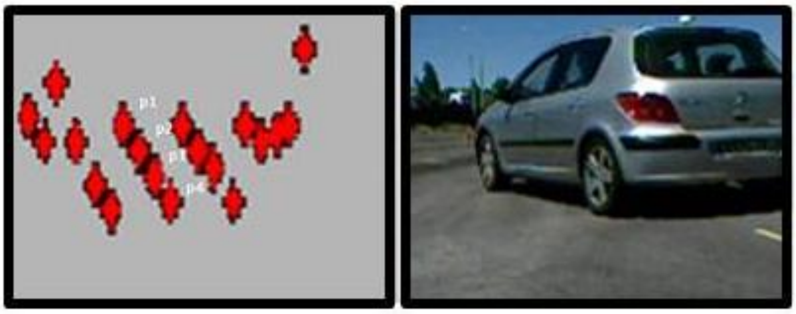


Figure 4.7. Moving vehicle pattern.

$$v = \frac{\sqrt[2]{(x_1 - x_2)^2 + (y_1 - y_2)^2}}{T} \tag{4}$$

where T is the rotation period which is T = 13msecs. Since there are 4 scans, three different speeds can be measured in order to provide a more reliable measurement.

$$Vy = \frac{\sum_{n=N}^{N-2} \frac{y_n - y_{n-1}}{t}}{3} \tag{5}$$

$$Vx = \frac{\sum_{n=N}^{N-2} \frac{x_n - x_{n-1}}{t}}{3} \tag{6}$$

$$v = \sqrt{{v_x}^2 + {v_y}^2} \tag{7}$$

where t =13 msegs and v is in m/s.

False positives can be avoided through detecting impossible velocities using these values represented above.

d) Pedestrians

A pedestrian classification is performed in two steps, taking into account the prior contextual information. First, obstacles with a size proportional to a pedestrian are selected among the different obstacles, and the shape of the polyline is checked with a typical pedestrian pattern.

Context information

The size of pedestrian used in the present application to select possible pedestrians is based on the model of human body ([92] and [93]), which models the human body as an ellipse. [92] details a study of the physical dimension of human being. It is generally accepted that physical dimensions for pedestrians was given in the early 70's in [93]. This value corresponds to an ellipse with two main axes with values (57.9cm x 33cm). The ellipse includes the body of a dressed human being. Other researches [94] use anthropological studies to conclude that this ellipse is (45.58 cm x 28.20 cm). Finally, both [92] and [93] conclude that this dimension can be approximate to an ellipse, which main axes are 0.6 and 0.5 meters. This last assumption is the model used to perform the pedestrian detection in this thesis.

Finally, a study of the different patterns given by pedestrians was performed (in Figure 4.8) giving the conclusions found in Figure 4.9 (a).

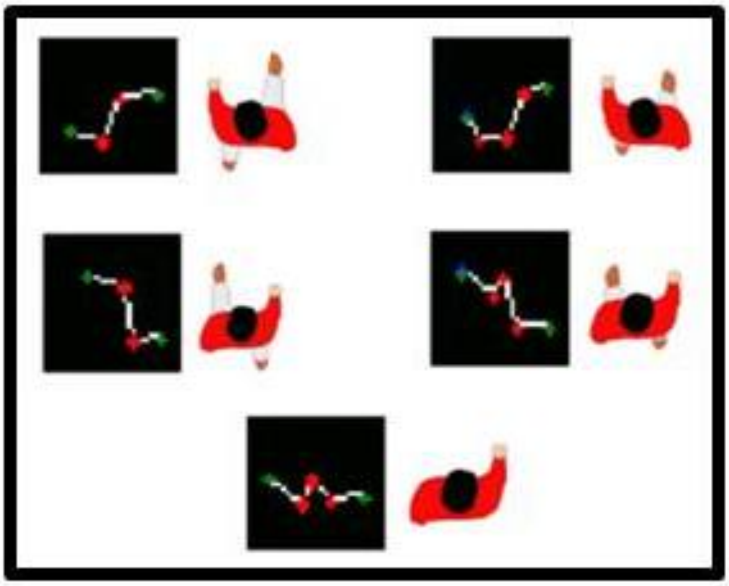
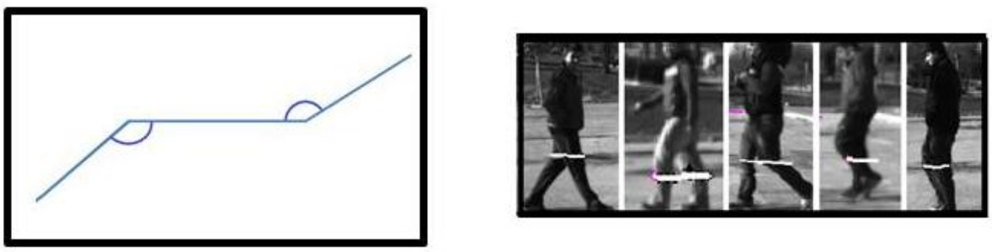


Figure 4.8. Examples of different pedestrian's patterns used to study the common pattern.

Pattern Matching

In this pattern, three polylines are presented, and the angles that connect the polylines are included within the limits of$[0, \frac{\pi}{2}]$.



(a) (b)

Figure 4.9. (a) Pattern for pedestrian detection. (b) Different examples of different patterns given by pedestrian with different leg positions with the information from laser scanner translated to the image.

A pattern matching process computes the two angles and gives a similarity score where 1 means 100% match. Equation (8) depicts this similarity score:

$$Similarity = \frac{2\theta_1}{\pi} \cdot \frac{2\theta_2}{\pi} \tag{8}$$

where θ_1 and θ_2 are the angles that connect two consecutive lines.

Similarity is computed between any two consecutive lines, in every obstacle in which the dimensions are within the limits of the human being described before. If the similarity is bigger than a given threshold, the obstacle is considered to be a pedestrian.

It is assumed that previous patterns are very common when dealing with a laser scanner, thus, false positives are expected. Fusion is important at this point to overcome the limitation.

d) Other obstacles

These obstacles are those that can not be fixed with any other patterns previously presented.

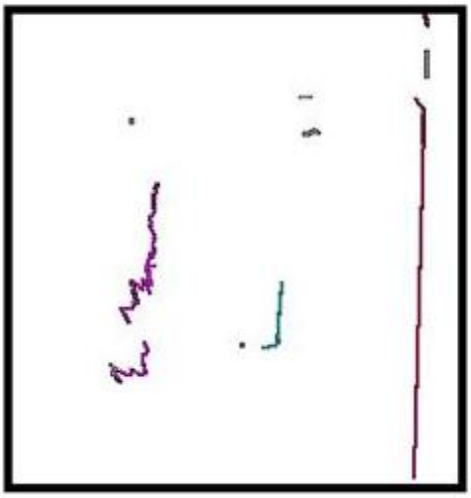


Figure 4.10. Laser scanner low level detection example (left) and real image of the environment (right).

4.1.2 Stage 2. Higher Level Classification

A higher-level stage is required to observe the behavior of different obstacles during a specific time period. Previously scanned obstacles are stored and verified using the new low-level detection.

a) Previous obstacle movement correction

Movement correction is performed according to the movement of the vehicle in the same way as the low-level detection, explained in the previous section. Because of this correction, all the previously detected obstacles can be referenced to the current vehicle position.

For obstacles labeled as *possible vehicles*, the velocity of the detected vehicle, according to previous scans, is computed, and the next position is calculated taking the movement of the car into account. If there is not enough information available from previous scans, the low-level speed detected from the previous scans (eq. (4) to (7)) is used to calculate this position. Once the vehicle has been detected during several scans, the velocity is corrected using this high level information. High-level velocity information is considered to be more accurate because it eliminates the laser rotation displacement, which leads to possible measurement errors.

Once the movement of the vehicle has been corrected and the movement of possible vehicles is computed, obstacles are searched within a window, according to the size of the obstacles from previous scans. If an obstacle is found within this window of the current scan, a comparison algorithm is used to verify if the obstacle is the same. If several obstacles have been found, the obstacle with the most amount of similarities according to several parameters, is considered to be the same obstacle.

b) Obstacle comparison

The comparison process is carried out according to shape characteristics, such as the width and position (Figure 4.11) (eq. (9) to (11)). If all of the comparisons remain within certain values, they are considered to be the same obstacle. If a circumstance arises where there are several possible candidates, the one with the closest value is considered to be the same obstacle.

$$\mathrm{x_{med}} = \mathrm{x_{min}} + (\mathrm{x_{max}} - \mathrm{x_{min}})/2 \tag{9}$$

$$\mathrm{y_{min}} \tag{10}$$

$$\mathrm{width} = \mathrm{x_{max}} - \mathrm{x_{min}} \tag{11}$$

where all the values represent the values of a obstacle given in the Figure 4.11.

These three parameters resulted in the most representatives. Ymax was also considered, but it resulted a parameter with excessive variation due to the nature of the laser scanner information (e.g. occlusions); therefore, it was discarded.

If more than one obstacle is found as a previous scan obstacle, the one lower correlating factor is considered (12) to correspond to this obstacle.

$$Corr = \gamma_1 \cdot d_{width} + \gamma_2 \cdot d_{xmed} + \gamma_3 \cdot d_{ymin} \tag{12}$$

where $d_i$is the distance of the new obstacle's parameter to the previous one, and γ_i is the weight applied to each distance of the different parameters that have been considered.

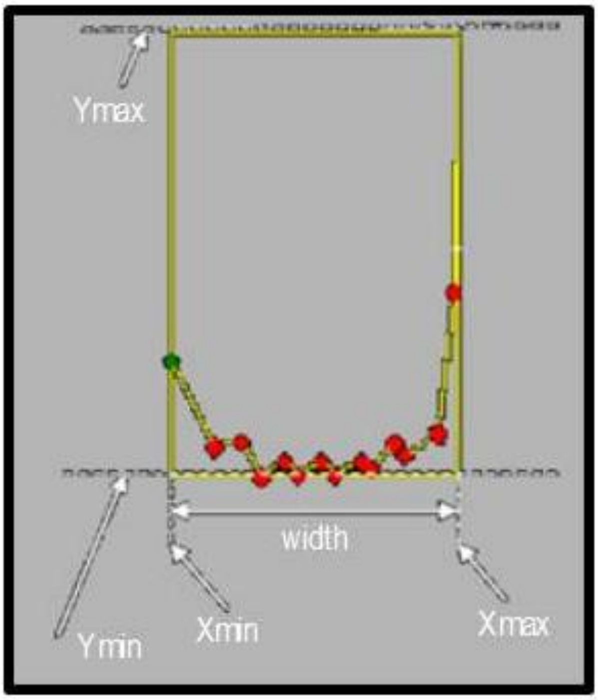


Figure 4.11. Shape characteristics used to compare detected obstacles.

c) Higher level classification

The higher-level classification algorithm is based on a voting scheme (13) that uses the ten last movements and low-level classification to perform the final decision:

$$V_i = \delta_i N_i \tag{13}$$

where V_i represents the number of votes for each type of obstacle, δ_i is a gain factor associated with each obstacle, and N_i is the number of times that an obstacle has been considered as being this type of obstacle during the low-level detection. The biggest V_i value represents the type of obstacle that has been detected.

d) Final classification

Before the final classification is carried out, several higher-level filters are used to correct possible false positives in the event of possible vehicle detection or pedestrians. These false detections can be avoided by computing the last ten movements stored. Some of the filters are associated with impossible velocity for vehicles or pedestrians. Also high lateral movements are considered false detections for vehicles.

4.2 Data Alignment

The present section deals with data alignment problem. First, an introduction to the data alignment problem is mandatory to both provide theoretical formulation of the problem and explain the different possible solutions. Later, the solutions used in the present proposal are given.

4.2.1 Alignment Problem

Sensors involved in DF applications usually do not measure the same physical phenomena; they are not located in the same place nor do they measure using the same coordinate systems. Data alignment function provides a general frame of referencing. It can be applied whether to commensurate or non-commensurate information. In the second example it is a mandatory process before proceeding to DF. In this general frame of referencing, a common coordinate system (space and time) is found where the data from the different sensors, as well as the global knowledge, can be presented.

Data Alignment is a sensor specific problem. It is very difficult and usually connected with the application, so no general mathematics or techniques can be found.

Usually the most common data alignment problems related with DF applications are the different coordinate systems used by the sensors and time synchronization. First are regarding the sensors and that they are allocated in different positions. Consequently, coordinate system translations are mandatory to work with the detections of the different sensors at the same coordinate frame. Later, they are related with the time when the detections are performed. Usually the different sensors perform the detections at different times, so special care should be taken when fusing data, given the time delays.

There are other problems related to data alignment, as with measurements in different units. But for the present work, coordinate alignment and time synchronization are the main problems that should be taken into account. The next section depicts the different possibilities and solutions for these two problems.

a) Time alignment

Time alignment is very specific for the sensors used in each application, and usually this problem is very difficult to solve. Although there are no general techniques, some typical examples will be presented, as well as their typical solutions:

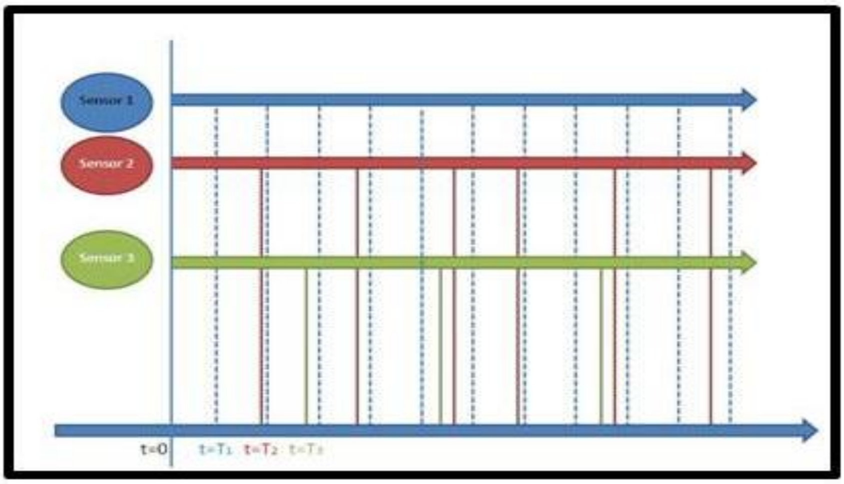


Figure 4.12. Sensors' behavior where all sensors are periodic. Periods do not have to be equal.

Two kinds of nodes or sensors can be differentiated according to the time scheduling, synchronous and asynchronous (or event-oriented). Periodic sensors give a detection every time. Asynchronous sensors, so called event-oriented, are those sensors in which it is not possible to know in advance when the detections are performed. Thus, the system is waiting for a detection or "event" at any time. According to this, the basic scenarios found are represented in figures Figure 4.13 and Figure 4.14:

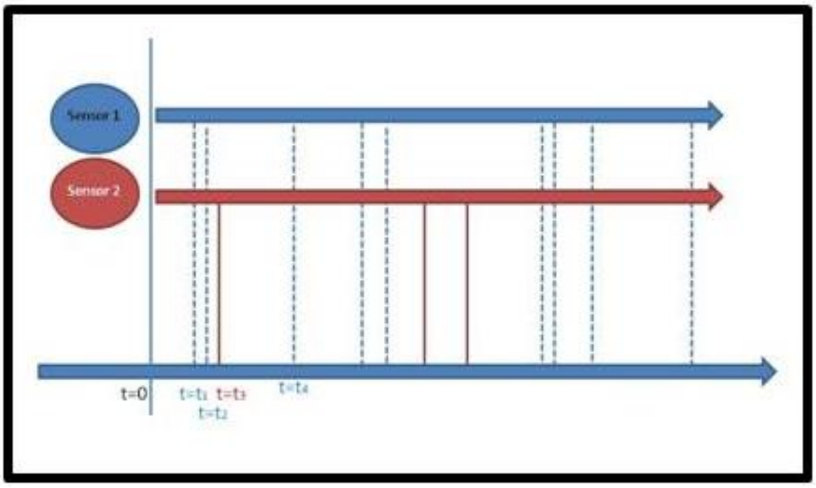


Figure 4.13. Sensors behavior when some sensors are event-oriented some other are periodical.

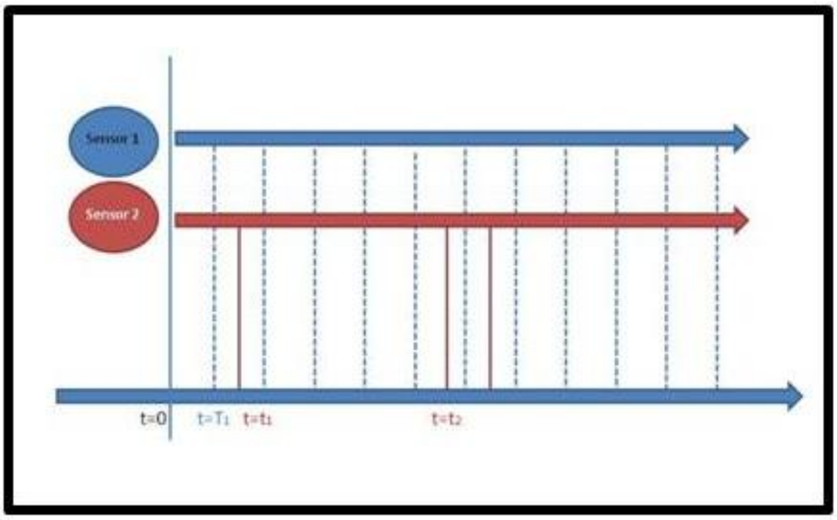


Figure 4.14. Sensors behavior when all sensors are event-oriented.

When all sensors have synchronous behavior (Figure 2.14), depending on the time requirements (real time applications, delays, etc.) different procedures can be accomplished:

- When real time performance is important, the highest acquisition frequency sensor is usually the main sensor and lower frequency ones are usually used for back up. This occurs because of the importance of the fast response time with these applications.
- When the performance of detections is the key point, information from all the sensors is sensitive, highest sensor period can be used as an estimation point. In this case, detections performed by the fastest sensor can be translated to the time where lowest detection is performed (e.g. to use movement of a vehicle to infer where the delayed detection is in the current time).

In systems where asynchronous detections are present, time alignment is less intuitive. The nature of the applications, and the sensors involved, subordinate the time alignment. When there is a sensor that periodically receives information (Figure 4.13), it is intuitive to use the period of the sensor as a time base for estimation purposes, since the time of the subsequent detections is known, and then use the events based sensors to corroborate or discard the detections. When no periodic sensors are available (Figure 4.14), a pre-established period can be utilized in case periodical updates are necessary or estimation methods that do not need periodical updates (e. g. Kalman Filters) can be used to estimate the position of targets in future situations.

b) Coordinate space alignment

Sensors are located in different positions, angles, etc. Therefore, measurements are usually referenced to different coordinate systems. Figure 4.15 shows the different coordinate systems when dealing with two sensors, image and laser scanner. Image coordinate system is based on pixels (u,v), and laser scanner coordinate is a 3D Cartesian coordinate system given in meters. Its origin and units are totally different (pixels vs meters).

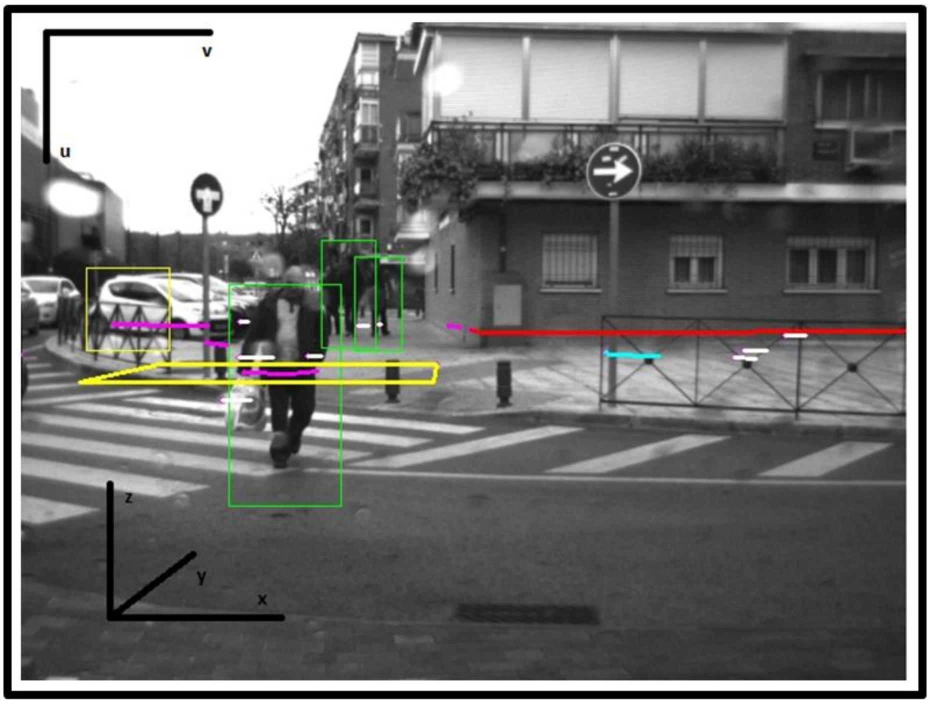


Figure 4.15. Two reference points for image and laser scanner. Laser scanner detections are shown, with pedestrian detection in horizontal rectangles. Vertical rectangles show the image detections.

Coordinate changes are dependent on the situation of every sensor and the applications (e. g. sensors located in a different position in a vehicle), as well as the physical phenomenon that every sensor measures. Thus, the space alignment is application-oriented and no general algorithms can be found.

4.2.2 **Solution Proposed**

The obstacles detected by the laser scanner were translated to the camera coordinate system, creating bounding boxes, where the visual based classification is performed. To perform this coordinate change some processes were necessary. After obstacle clustering, performed by the laser scanner, the main properties of the obstacles were checked and a set of obstacles was created. Later, coordinate transformation was mandatory to translate these points to the visual coordinate system, this was a change of coordinates from 2D plane to another 2D coordinate system, but the coordinates to transform are not the same, since the planes refer to different coordinates (x,y) vs (x,z). Figure 4.16 shows the different coordinate systems. All detections must be referred to the central part of the front bumper of the vehicle.

Figure 4.16. Different coordinate systems: (x_c,y_c,z_c) are camera coordinate system, (x_l,y_l,z_l) is the laser coordinate system, and (x_v,y_v,z_v) is the coordinate system of the vehicle. Coordinates (u,v) are the image coordinates.

a) Coordinate change

The coordinate changes are similar to the rotation and translation performed by the movement compensation of the vehicle in the laser scanner shown in (1), but with some changes, as is depicted in (14) . In this case, the rotation angles correspond to the angles of rotation between the different coordinate systems. The translation vector T corresponds to the distance from the laser scanner to the camera.

$$\begin{bmatrix} x_c \\ y_c \\ z_c \end{bmatrix} = R(\begin{bmatrix} x_l \\ y_l \\ z_l \end{bmatrix} + T) \tag{14}$$

where R is the rotation matrix shown in (1) corresponding to the Euler angles that represent rotation between the different coordinate systems. T is the translation vector $T = \begin{bmatrix} x_t \\ y_t \\ z_t \end{bmatrix}$ that corresponds to the distance between the coordinate systems, shown in Figure 4.16 , (x_c,y_c,z_c) are the camera coordinates and (x_l,y_l,z_l) the laser scanner coordinates.

After changing the coordinate system to the camera coordinates, the ROIs have to be transformed to the camera coordinate system (u,v) through pin-hole model.

$$\lambda \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} f & 0 & u_0 \\ 0 & f & v_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_c \\ z_c \\ y_c \end{bmatrix} \tag{15}$$

where u_0 and v_0 are the center coordinates of the camera coordinate system in pixels. (u,v) are the coordinates in the camera coordinate system in pixels. x_c y_c and z_c are the Cartesian coordinates from camera. And f is the focal length.

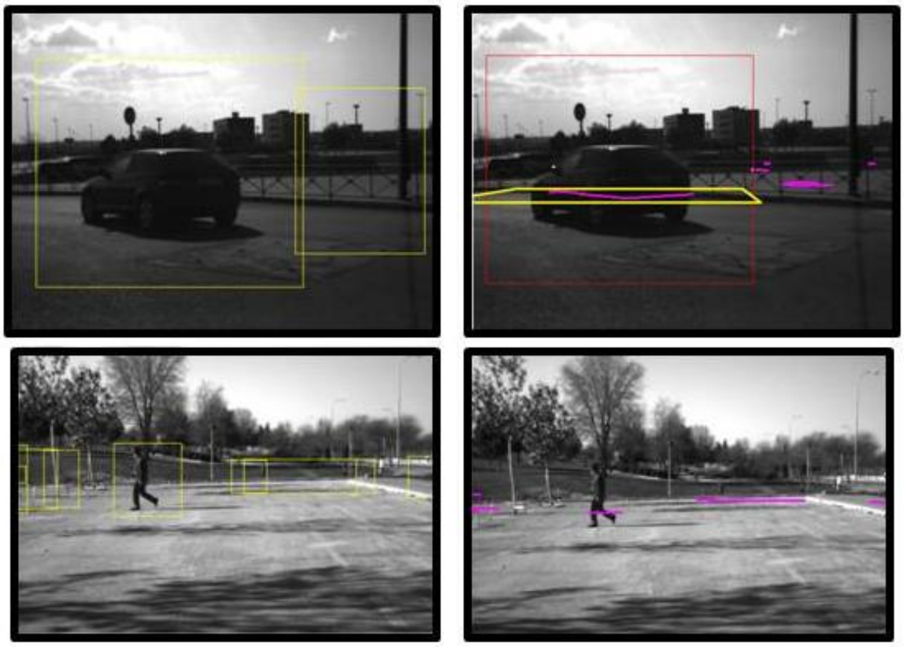


Figure 4.17. Bounding boxes (left) and laser scanner detection (right). For vehicles, small obstacles are filtered, thus only big obstacles are considered interesting to perform vehicle detection.

After the coordinate changes, two different kinds of bounding box sets were created. First, bounding boxes of low and medium size obstacles were created. These bounding boxes were used to search for pedestrians using visual procedures. Later, higher sizes bounding boxes were used to perform vehicle classification based on visual procedures. These visual procedures will be explained in section 4.3.

b) Extrinsic calibration

Extrinsic parameters of both sensors had to be measured to allow an accurate coordinate translation. Intrinsic calibration was not necessary since both subsystems were provided with these parameters.

Before proceeding to an accurate extrinsic calculation, important extrinsic parameters had to be identified. According to equation (14) and Figure 4.16, the important extrinsic parameters are:

- (x_t,y_t,z_t) These correspond to the distances from the laser scanner to the camera in the coordinate system. The distances from the sensors to the center of the vehicle's front bumper should also be measured. These distances could be measured manually.

- Euler angles among the different coordinate system. The most important issue was the change from laser scanner to the camera coordinate systems, since this was a difficult task to measure accurately. To perform this, a supervised method was created, which allows performing this calibration by an online procedure. Images were taken with several obstacles that allow measurement of all angles and to check changes in the projection. By projecting the points of the image online, the difference between the points and the image could be measured. The results were stored in a configuration file to be loaded automatically in future executions (Figure 4.18).

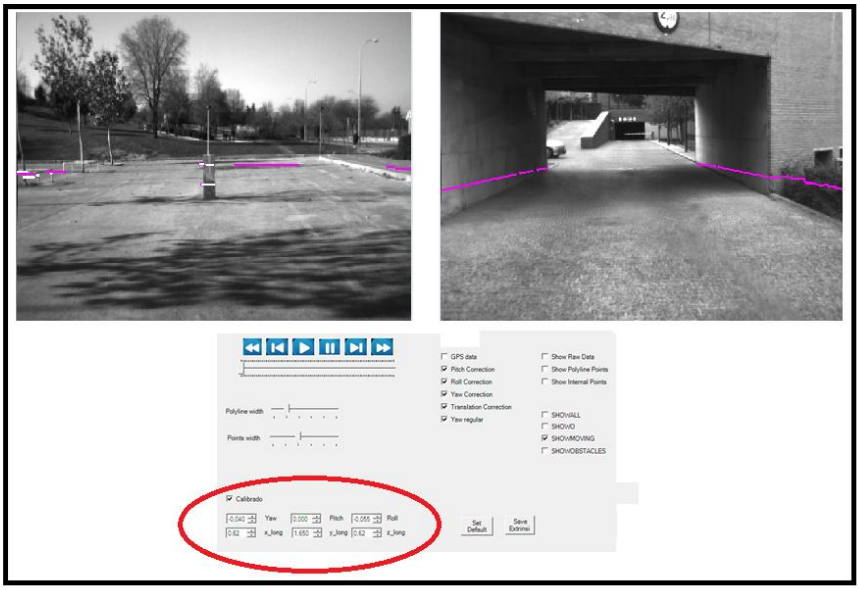


Figure 4.18. Example of two calibration sequences where the Euler angles are corrected, varying the parameters shown in the red circle (yaw,pitch,roll angles and x,y,z distances between coordinates axes).

- Euler angles of the isolated laser scanner coordinate system should be accurately measured. The single layer scanner property creates a limitation not only due to the limited information, but also for the extreme dependence on the pitch angle. To avoid this inconvenience, first an accurate measurement of the Euler angles should be accomplished. Later, pitching movement should be monitored to detect extreme changes caused by pitching variations that could lead to misdetections. Figure 4.19 shows how accuracy in this pitching calibration should be accomplished. For a pedestrian of 1.80 meters at a distance of 50 meters (the laser scanner maximum distance configuration is at 82 meters), a pitching angle higher than 1.37 degrees means that the beam of the angle would go over the head of the pedestrian. As a consequence, no detection by any or both sensors would be performed. Manual calibration was performed to assure that the pedestrian was detected at maximum distance.

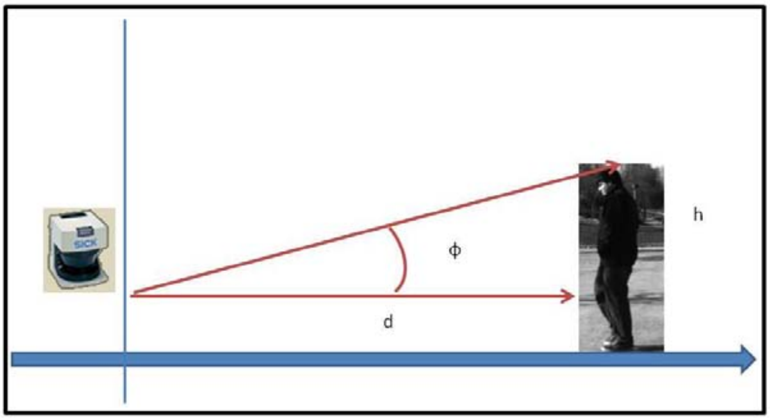


Figure 4.19. Maximum pitching error angle allowed representation. According to the distance and the pedestrian height.

Although other Euler angles' error were considered irrelevant, inertial system mounted in the top of the laser was used both to check the other angles and to detect strong pitching movements. Since the inertial system returned angular movement, strong pitching movement could be detected, and in these situations, lasers scanner detection was discarded.

Figure 4.20. Laser scanner detail with the inertial measurement system mounted in the top. Frontal view (left) and top view (right).

c) Time alignment

Taking into account time constraints, the laser scanner period was used as a reference period since the scanning frequency is higher, giving a full scan every 52 milliseconds. The time consumed to perform the classification is insignificant enough to not add any delays in the process. The selected camera with the subsequent processing time was more time consuming, giving a variable time depending on the amount of ROIs. Detections given by the camera were translated to the next laser scanner detection as in Figure 4.21.

Figure 4.21. depicts what happened every time an image or a laser scanner is received and processed. The fastest algorithm from the laser scanner allows it to be processed without delay. The camera, on the other hand, has more computational cost that could produce delays in the acquisition, giving a lower acquisition frequency than the expected (aprox. 100 milliseconds). Besides, the divergence in the amount of bounding boxes and the computational cost algorithm causes a problem. Any vision detection that has a variable delay makes it more unreliable for synchronizing purposes than the laser scanner period.

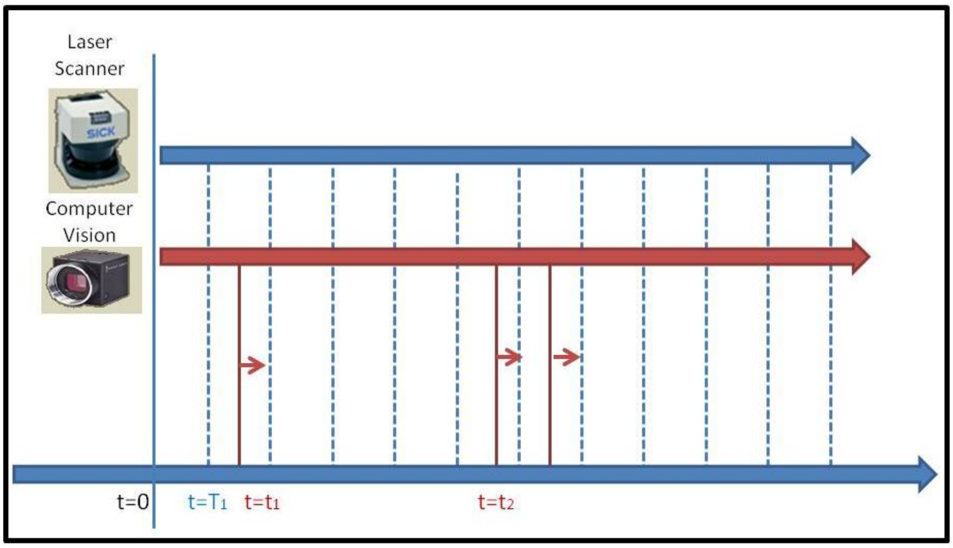


Figure 4.21. Temporal representation of a sequence, the laser scanner has a fixed temporized thanks to the faster algorithm and the faster period. Every lecture from the camera (in red) is extrapolated to the next lecture of the laser scanner (in blue).

One extreme example of this non-synchronization is given in Figure 4.22. The size of the bounding boxes was designed to be big enough to take these delays into account.

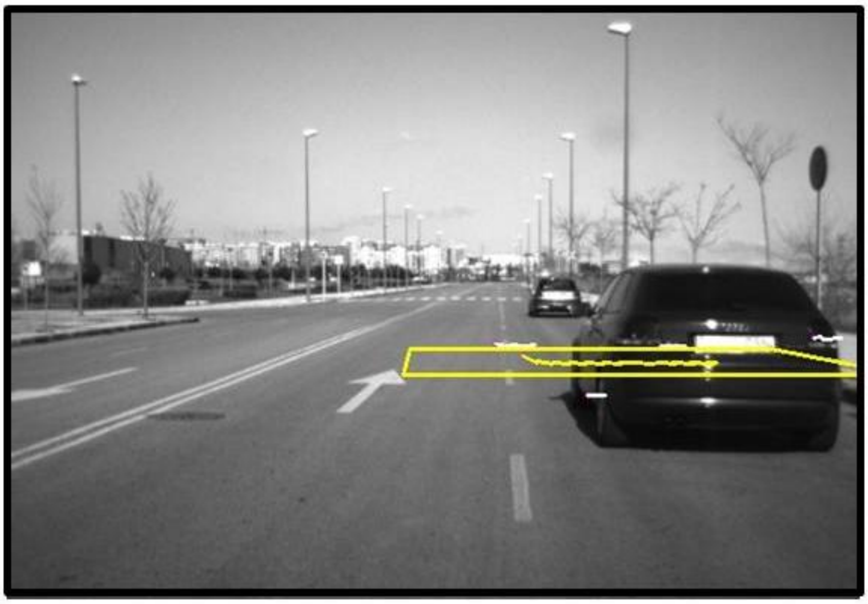


Figure 4.22. Extreme example of the non-synchronization given by the laser and the image. Laser scanner performs the detection with some delay with the camera detection.

4.3 Vision Based Classification

Two kinds of obstacles were taken into account for vision-based classification: vehicles and pedestrians. As remarked in previous chapters, the scope of the present thesis is to prove that laser scanner and fusion can help overcome some of the limitations given by the classic vision based procedure. In the present section, these approaches used for vehicle and pedestrian classification are described. The basic explanation is provided, as well as the references for further information. Again it should be remarked that it is out of the scope of the present thesis to provide a novel approach based in vision, thus the present chapter is limited to an explanation of the algorithms used.

Previous sections explained the laser scanned based bounding boxes creation process. Once these bounding boxes are provided for the visual module, each one of them is checked with the subsequent algorithms to check if the obstacles inside the box represent pedestrians or vehicles.

4.3.1 Pedestrian Classification

Pedestrian classification was based in the Histrograms of Oriented Gradients (HOG) that were originally used for human detection in [83]. The final classification is performed with Support Vector Machine classifier.

Histogram of Oriented Gradients

The theory after the HOG feature description is based on local appearance and shape of all objects in an image, which can be described by the distribution of intensity gradients or edge directions. The implementation divides the image into small-connected regions (cells) that can have different shapes (circles or squares). For each cell a histogram of gradient directions (or edge orientations) for the pixels within the cell is compiled. The combination of all these histograms represents the descriptor of the image.

Some improvements can be performed through the original algorithms to improve the accuracy of the method, such as changing the size or shape of the cells. Another typical improvement is to divide the image in blocks bigger than the cells where the intensity is measured to perform local histogram normalization within the cells of the block. This normalization results in better performance against illumination changes and shadowing.

The main advantage of HOG features over other descriptor methods is that it operates in local cells, which means that it is invariant to geometric and photometric variations. On the other hand, the main disadvantage is the object orientation is dependency. With pedestrians, the different orientations of the pedestrian could lead to some misdetections. To solve this problem authors (Dalai and Triggs) explained that fine orientation sampling, coarse special sampling and the previously explained photometric normalization could help to avoid these misdetections due to the movement of the pedestrians.

The algorithms steps followed to perform the pedestrian detection based in HOG features is composed of a **preprocessing stage**, **gradient computation**, **orientation binning**, **description blocks**, **blocks normalization** and **SVM classification** .

a) Preprocessing stage

Any visual procedure first needs a preprocessing stage. In this case, this preprocessing is not such an important step. Dalal and Triggs pointed out that this preprocessing stage is not such an important step, achieving the same results even avoiding this stage. Authors tested Gaussian, smoothing the preprocessing stage, but no improvements were obtained.

b) Gradient computation

Since the preprocessing stage could be avoided, the first step usually consists of calculating the gradients. This is usually performed with a Sobel mask of 1-D centered point discrete mask in one or both of the horizontal and vertical directions. Other masks could be used, but for the specific case of human classification those masks did not show better performance.

c) Orientation binning

The next step contains the calculations of the cell histogram. Each pixel within this cells computes with a vote that is weighted (whether by the gradient magnitude itself or some function of the magnitude). These weighted votes compute for an orientation-based histogram channel consisting of the values found in the gradient computation performed before.

As previously stated, these cells can have different shapes: rectangular or radial. The histograms channels can be spread over 0 to 180 degrees (unsigned) or 0 to 360 degrees (signed).

Dalai and Triggs used unsigned gradients and 9 histogram channels for their approach.

d) Description blocks

In order to avoid errors due to illumination and contrasts, the gradient strengths should be locally normalized. The final descriptor is the vector of all components of the normalized cells from all the blocks in the detection windows. These blocks usually overlap; thus, cells contribute to more than one block to the descriptor.

Two kinds of geometry blocks are used: R-HOG (rectangular) and C-HOG (circular). The first is square grids represented by three parameters: number of cells per block, number of pixels per cell and number of channels per cell histogram. Dalal and Trigger optimal parameters in their experiments were 3x3 cellblocks of 6x6 pixels cells with 9 histograms. Later, this can be found in two variants: central cell and angularly divided central cell. Four parameters are used to describe these blocks: number of angular and radial bins, radius of the center bin, and the expansion factor for the radius of additional radial bins.

e) Blocks normalization

Up to four different approaches for the normalization process are given in [83], which all show the same performance more or less.

f) SVM classification

The final step is the classification process using Histogram Orient Gradient descriptors. The previously described descriptors were first trained for the Support Vector Machine (SVM) classifier. SVM is a binary classifier that looks for the optimal hyperplane for a decision function. It is a machine-learning algorithm very spread in computer vision approaches.

Figure 4.23 shows examples given by the author of the HOG features.

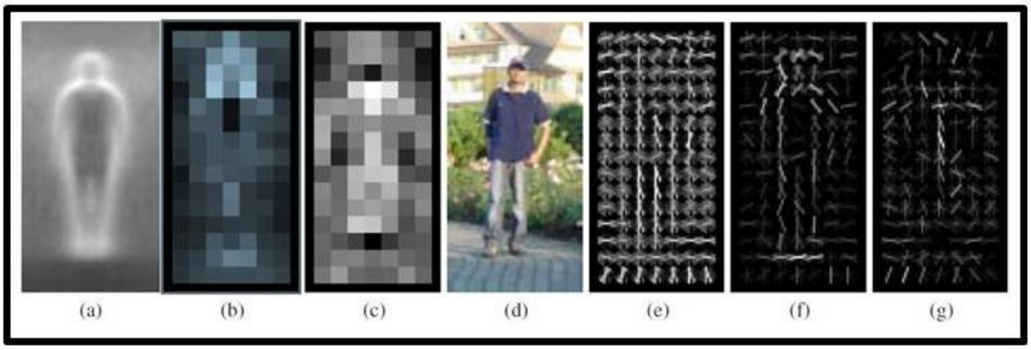


Figure 4.23. Figure taken from the original paper of HOG features for pedestrian detection [83]. (a) The average gradient image over the training examples. (b) Each "pixel" shows the maximum positive SVM weight in the block centered on the pixel. (c) Likewise for the negative SVM weights. (d) A test image. (e) It's computed R-HOG descriptor. (f,g) The R-HOG descriptor weighted by respectively the positive and the negative SVM weights.

The approach used for the present work was based on the OpenCV 2.0 [95] (Open Computer Vision library) pedestrian detection Approach. Examples of pedestrian detections are found in Figure 4.24.

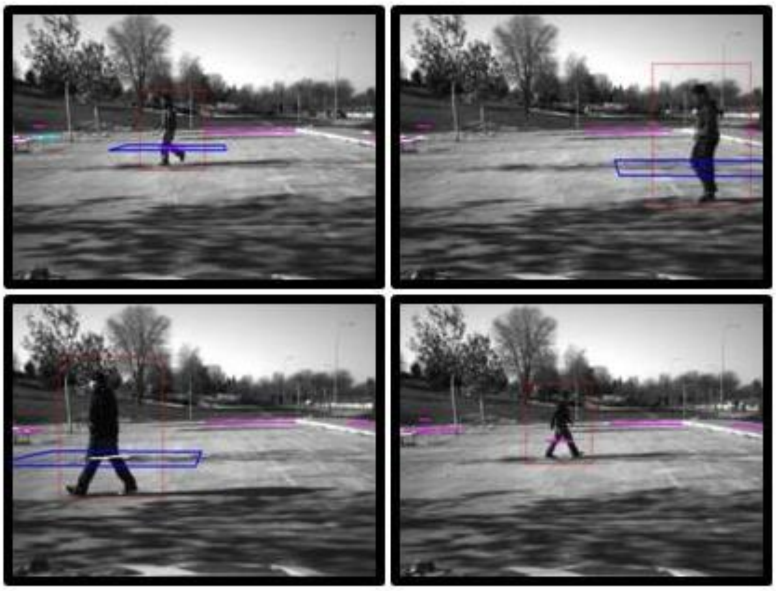


Figure 4.24. Examples of the pedestrian detection (red boxes) in images using HOG features.

4.3.2 **Vehicle Classification**

Vehicle classification system, used for the present approach, was based in the Haar-Like features presented by Viola and Jones [78]. This approach is based on the use of fast Adaboost classifiers and simple features that allow to classifying obstacles in a fast and reliable way. By checking a high amount of features in different training images, the algorithm can select those relevant features that provide reliable detection in a fast way, according to the obstacles to classify.

Viola and Jones presented definitions of simple image based features. These features are obtained in a fast and sequential way using sums and subtractions of the values of pixels in certain regions. Finally, Adaboost selected the best features for machine-learning schemes [96]. Cascade classifiers, based on these simple features, are created, which result in a sequential classification.

a) Feature creation and selection

Viola and Jones remark in their work that even though the definition of the features used in their work is simple and primitive, compared to other more sophisticated approaches, it is one of the more positive points of theirs, since they can be calculated in a fast way, providing similar or even better results. Furthermore, these features offer a rich image representation. In the example application given by the authors, a facial detection is provided. Because of these features, regions of the image, important for facial detection such as eyes, mouth, etc., are identified in an efficient way.

Four kinds of features were defined. Figure 4.25 presents these features. The pixels in the squares are summed. Then the values of the sums of different squares (with different colors in Figure 4.25) are subtracted. Two *two-square* features were created, a *three-square* feature and a *four-square* feature.

Numerous features were created in the fixed window. For a simple image, the number of features is far greater than the number of pixels. Thus, to reduce the amount of features, a training process based on Adaboost reduces the number of features leaving only the ones that proved to be relevant in the training sequence.

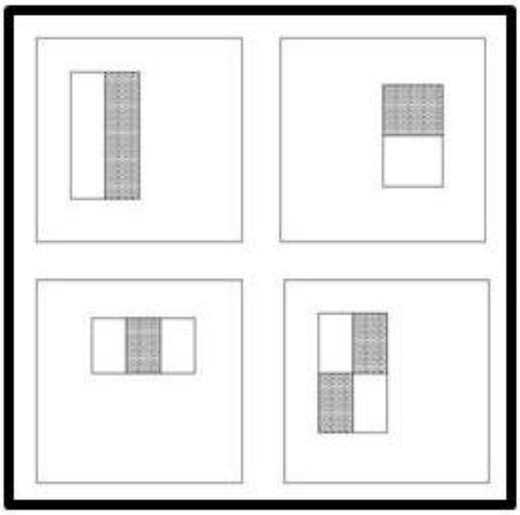


Figure 4.25. Haar-like features defined by Viola and Jones [78].

The selection reduces the number of features and creates an Adaboost approach. This approach is based on small, very simple decisors, which separate detections into basic stages that check a simple feature within the limits. Therefore, a cascade classifier, which uses these simple decisors in a sequential way, checks for these features one by one. In each stage, if one feature fits the threshold, the next feature is checked; otherwise the candidate is rejected. This way the hierarchy of features can combine the recurrence of features to provide a fast and reliable classification. Figure 4.26 gives a scheme of one of these classifiers.

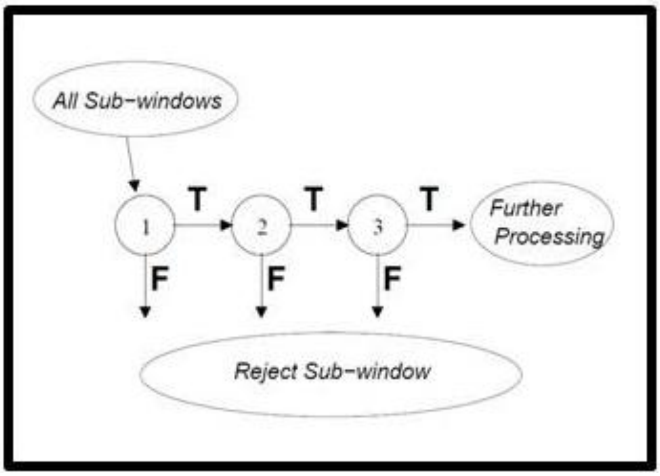


Figure 4.26. Cascade classifier definition given in [78].

b) Adaboost classification

Adaboost machine-learning algorithms is a method that divides the classify problem in small classifiers, weak classifiers. These weak classifiers commonly check the features separately. Thus by combining several layers of weak classifiers in different ways, a more powerful and robust classifier is created.

Viola and Jones proposed a weak classifier for a given feature:

$$h_j(x) = \begin{cases} 1 \; if \; p_i f_i(x) < \; p_i \theta_i \\ 0 \quad\quad otherwise \end{cases} \tag{16}$$

where $h_j(x)$ is the classifier, $f_i(x)$ is a given feature, $\theta_i$is the threshold and p_i is the parity indicating the direction of the inequity sign.

By combining these weak classifiers in a cascade, the classifier can select using these simple features in a robust way (Figure 4.26).

Haar-like feature classifier was utilized to perform the vehicle classification. Here again OpenCV provides the tools to perform this classification. OpenCV was used both to train the classifier and to perform the classification. Figure 4.27 shows some examples of vehicle detection based on the Haar-like features.

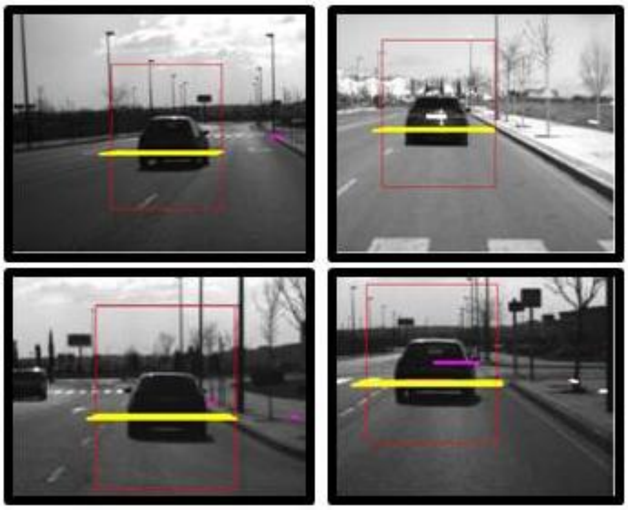


Figure 4.27. Example of Haar-Like feature based vehicle detection (red boxes).

CHAPTER 5. TRACKING AND DATA ASSOCIATION

The present chapter deals with a common problem in Data Fusion application, Multiple Target Tracking (MTT). As explained in chapter 2, two main processes are related to this task.

- Position estimation. This is one of the main issues when dealing with MTT applications since this estimation allows analyzing and inferring the movements of the different targets found. In the present chapter several approaches for estimation tracking are detailed. Chapter 6 provides results with comparison of the different approaches presented in this chapter.

- Data association is mandatory. Usually data association has two processes: first, data association from the different sensors is necessary to recognize which detections correspond to the same obstacles. Second, the association of the new detection with past detections (so called tracks) helps to determine the trajectories of the different obstacles found along time in the environment.

  For the present approach, data association from different sensors is inherent to the algorithms explained in previous chapters. Both sensors perform the classification over the same set of obstacles, detected using laser scanners. Hence no special process is necessary for this purpose.

  Second association problem is going to be explained in present chapter, paying attention to the different possibilities implemented. Chapter 6 provides the comparison of the different association algorithms.

Some other issues related with MTT will be described in the present chapter, target management, Gating procedures, distances, etc. First a theoretical explanation of the algorithms used is given, and later the implementation necessary to add Data Fusion to the basic approaches is detailed.

Finally, threat assessment is explained in the last section of the chapter. The danger that any of the detections represent is calculated using context information and inertial system, solving Fusion levels 2 and 3.

5.1 Position Estimation

Position estimation is a key point in target detection and tracking. The idea is to be able to estimate, according to measurements where the target is going to be after a given time, whether to predict the movement or to keep a track of the movement of the target.

Formulation:

Let it be a system in which the state is described by the state vector X. It evolves according to the following model:

$$\mathbf{X_{k+1}} = \mathbf{f_k}(\mathbf{X_k}, \mathbf{U_k}, \mathbf{W_k}) \tag{17}$$

where X_k is the state vector at time t_k, U_k is a known input at time t_k, W_k is a random system noise. Function f is called system transfer function and it is assumed to be known.

At an instant k+1 a vector Z_{k+1} of measurements is received. These measurements are related to the state vector via the measurements equation:

$$\mathbf{Z_{k+1}} = \mathbf{h_k}(\mathbf{X_k}, \mathbf{V_k}) \tag{18}$$

where X_k is the state vector at time t_k and V_k is a white noise that affects the measures. It is assumed to be an uncorrelated random variables vector. The system model would be represented as follows:

$$\mathbf{X_{k+1}} = \mathbf{F_k X_k} + \mathbf{B_k U_k} + \mathbf{W_k} \tag{19}$$

$$\mathbf{Z_{k+1}} = \mathbf{H_k X_k} + \mathbf{V_k} \tag{20}$$

where matrixes F is the state transition model matrix and H is the observation model matrix that matches the true state space into the observed space. Figure 5.1 shows the block diagram of the model.

It has to be remarked that the discrete system presented in previous equations has some assumptions:

- It is considered that the future state of the system only depends on the current state and the new inputs. This is called a Markov process.
- Transition and observation matrixes are considered to be known.

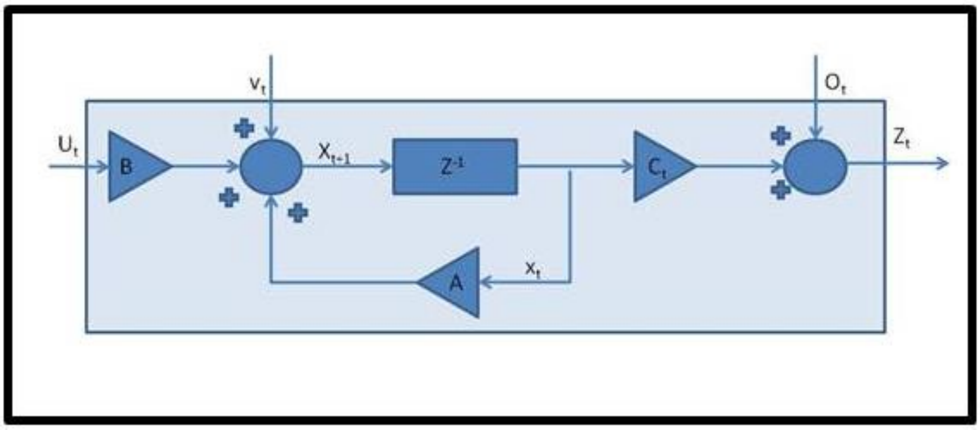


Figure 5.1. Block diagram of the model of the general system.

5.1.1 Classic Estimators

In target tracking the model is usually adopted with input U=0. Since the models for targets have the information limited to the positions, and sometimes some extra information related to velocity is added [97]. This block diagram is represented in Figure 5.2.

The classic estimation process is a recursive process, as it is shown in Figure 5.2. This recursive processing is also used in other estimator methods that will be explained later.

1. Prediction. The state $X_{k|k-1}$is predicted according to the model.
2. Correction. Here the predicted output $Y_{k|k-1}$ is compared to the real measurements and the estimation matrix (K_k) is corrected according to this difference. The method depends on the algorithm used to perform this correction. Although all of them try to minimize the estimation error.

The main disadvantage of the model is that it is weakly modeled against the uncertainty of the model, since both system and measurement noise are not taken into account in the estimation of the state. Therefore, classic estimators are only useful for observation or short time period tracking that avoids noise to diverge the estimations.

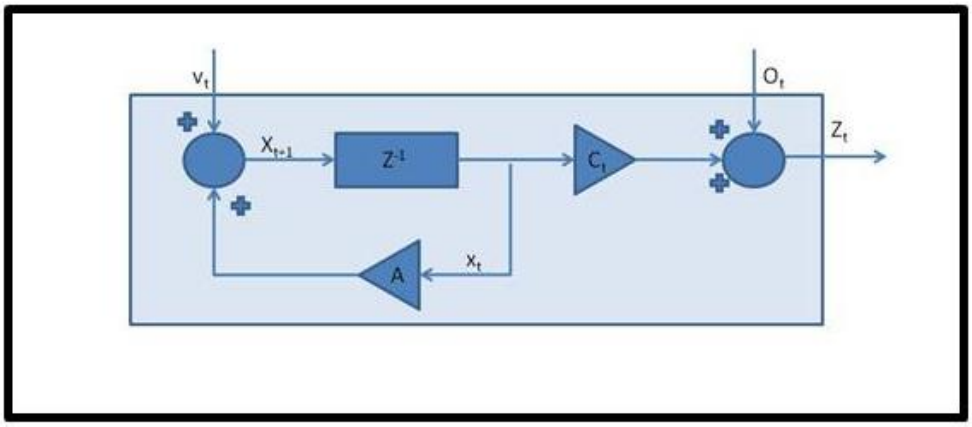


Figure 5.2. General block diagram for target tracking.

5.1.2 **Kalman Filter**

Kalman Filter (KF) was proposed in 1960 [98]. It is used to obtain the solution that minimized the estimation error by a recursive formulation of the classic estimator.

Some restrictions are assumed:

- The system model is assumed to be lineal. In case of no lineal model, some extension to the Kalman Filter has been proposed such as the Extended Kalman Filter [97] or Unscented Kalman Filtering [99], which are explained in this section.
- The system and measurement noises are considered to be zero-mean, white, Gaussian processes with known covariance Q_k and R_k respectively. Thus in this way, KF incorporates the uncertainty into the model, which lacked the classic estimators presented before.

The two stages of the estimator for Kalman Filtering are:

1. Prediction.

   In this stage, the state vector X_k at time t_k is estimated according to the model and previous state at time t_{k-1}, that is: $\hat{X}_{k|k-1}$. Also at this stage the covariance of the error is predicted ($P_{k|k-1}$). Equation (21) and (22) depicts this prediction process.

$$\hat{X}_{k|k-1} = F_k\hat{X}_{k-1|k-1} + B_kU_k \tag{21}$$

$$P_{k|k-1} = F_k P_{k-1|k-1} F_k^T + B_k U_k \tag{22}$$

2. Update stage.

According to the measure vector Z_k , the model and the previously calculated $P_{k|k-1}$, the estimation matrix K_k, also called Optimal Kalman Gain, is calculated. This Gain is used to actualize the state vector and the covariance matrix for the time t_k. As it is described in equation (23) to (27):

Innovation or measurement residual

$$\hat{Y}_k = Z_k - H_k \hat{X}_{k|k-1} \tag{23}$$

Innovation (or residual) covariance

$$S_k = H_k P_{k|k-1} H_k^T + R_k \tag{24}$$

Optimal Kalman gain

$$H_k = P_{k|k-1} H_k^T + S^{-1} \tag{25}$$

Updated (a posteriori) state estimate

$$\hat{X}_{k|k} = \hat{X}_{k|k-1} + K_k \hat{Y}_k \tag{26}$$

Updated (a posteriori) estimate covariance

$$P_{k|k} = (I - K_k H_k) P_k \tag{27}$$

Although KF is a very well known tool that has proved through the years its utility, the above restrictions make that the accuracy obtained by the Kalman Filter is only obtained when the model fits these constraints. When dealing with non-lineal models, Extended Kalman Filter (EKF) and Unscented Kalman Filter (UKF) are the two main proposals that can solve the problem of dealing with linear filters tracking problem.

Another problem related with Kalman Filtering is that it is an unimodal filter. It means it can only track a single target for each filter. Thus for MTT it is necessary to create a Kalman Filter for every target to track.

In the next sections the different methodologies that try to solve the non-linearity problem related to the estimation filters are described i.e. EKF and UKF.

5.1.3 **Extended Kalman Filter**

The extended Kalman Filter is the nonlinear version of the Kalman Filter which linealizes about an estimation of the current mean and covariance. In the case of well defined models it is considered the de-facto standard for nonlinear sate estimators, navigation systems and GPS.

In the Extended Kalman Filters, state transition and observation models (f and h in equations (17) and (18) of the system models and repeated in (28) and (29) for self content) don't need to be lineal functions, but differentiable functions.

$$\mathbf{X_{k+1}} = \mathbf{f_k}(\mathbf{X_k}, \mathbf{U_k}, \mathbf{W_k}) \tag{28}$$

$$\mathbf{Z_k} = \mathbf{h_k}(\mathbf{X_k}, \mathbf{V_k}) \tag{29}$$

The functions f and h can be used to predict the state in the next time k+1 and to compute the predicted measurement from the predicted state. But these functions cannot be applied to the covariance matrix directly; instead a matrix of partial derivatives (the Jacobian) is computed. In this way a linearization about the estimated mean and covariance is accomplished by performing multivariable Taylor Series expansion method (37) and (38).

Thus the equations (19) to (27) of the Kalman Filter finally are calculated as follows:

Prediction:

$$\widehat{X}_{k|k-1} = f_k(\widehat{X}_{k-1|k-1}, U_k) \tag{30}$$

$$P_{k|k-1} = \grave{F}_{k-1} P_{k-1|k-1} \grave{F}_{k-1}^{T} + B_k U_k \tag{31}$$

Update stage

Innovation or measurement residual

$$\widehat{Y}_k = Z_k - h_k(\widehat{X}_{k|k-1}) \tag{32}$$

Innovation (or residual) covariance

$$S_k = \grave{H}_k P_{k|k-1} \grave{H}_k^{T} + R_k \tag{33}$$

Optimal Kalman gain

$$H_k = P_{k|k-1} \grave{H}_k^{T} + S^{-1} \tag{34}$$

Updated (a posteriori) state estimate

$$\hat{X}_{k|k} = \hat{X}_{k|k-1} + K_k \hat{Y}_k \quad (35)$$

Updated (a posteriori) estimate covariance

$$P_{k|k} = \left(I - K_k \grave{H}_k\right) P_k \quad (36)$$

where transition and observation matrixes are defined to be the following Jacobians:

$$\grave{F}_{k-1} = \left.\frac{\partial f}{\partial x}\right|_{\hat{X}_{k-1|k-1,U_k}} \quad (37)$$

$$\grave{H}_k = \left.\frac{\partial h}{\partial x}\right|_{\hat{X}_{k|k-1}} \quad (38)$$

Although it has been proved to be the best and *de facto* standard for well modeled system with nonlinearities, several drawbacks are commonly accepted:

- If the initial estimation of the state is wrong, or if the process is incorrectly modeled, the filter may quickly diverge.
- Also, the estimated covariance matrix tends to underestimate the true covariance matrix; therefore, there are risks to become inconsistent in the statistical sense.

Julier S.J. and Uhlmann [100] in the proposal of Unscented Kalman Filtering explained the drawbacks of the EKF as follows:

“The extended Kalman filter (EKF) is probably the most widely used estimation algorithm for nonlinear systems. However, more than 35 years of experience in the estimation community has shown that is difficult to implement, difficult to tune, and only reliable for systems that are almost linear on the time scale of the updates. Many of these difficulties arise from its use of linearization."

5.1.4 **Unscented Kalman Filter**

When the previously presented functions f and h (state transition function and observational model) are highly non lineal, the EKF usually gives poor results. The reason for this bad performance is the propagation of the covariance through the linearization of a strong non-lineal model. The unscented Kalman Filter [100] is a deterministic based approximation method that is becoming popular in the latest years, providing more accurate results than EKF for strong non-linear models.

The UKF uses a deterministic sampling technique, the Unscented Transform, to create a minimal set of sample points (sigma points) around the mean of the state estimated. The sigma points are then propagated through the non-lineal functions from which the mean and the covariance of the estimate are then recovered. Besides the better performance obtained, the process followed has less computational cost since the calculation of the Jacobians is avoided.

As with the EKF, the UKF prediction can be used independently from the UKF update, in combination with a linear (or indeed EKF) update, or vice versa. For the present approach the model used is a lineal model although the real movement of the pedestrians could present non-linealities. Thus, this last assumption was utilized, using a lineal model for the predictions, by means of the UKF the non-linealities under the model trying to be solved.

Formulation of UKF

Prediction:

The estimated state and covariance are augmented with the mean and covariance of the process noise.

$$X^{a}{}_{k-1|k-1} = \left[\widehat{X}^{T}{}_{k-1|k-1}\ E\left[W_k^T\right]\right]^T \tag{39}$$

$$P^{a}{}_{k-1|k-1} = \begin{bmatrix} P_{k-1|k-1} & 0 \\ 0 & Q_k \end{bmatrix} \tag{40}$$

A set of 2*L*+1 sigma points is derived from the augmented state and covariance where *L* is the dimension of the augmented state.

$$X^{0}{}_{k-1|k-1} = X^{a}{}_{k-1|k-1}$$

$$X^{i}{}_{k-1|k-1} = X^{a}{}_{k-1|k-1} + \left(\sqrt{(L+\lambda)P^{a}{}_{k-1|k-1}}\right)_i ,\ i = 1 \ldots L \tag{41}$$

$$X^{i}{}_{k-1|k-1} = X^{a}{}_{k-1|k-1} - \left(\sqrt{(L+\lambda)P^{a}{}_{k-1|k-1}}\right)_{i-L} ,\ i = L+1, \ldots 2L$$

where $\left(\sqrt{(L+\lambda)P^{a}{}_{k-1|k-1}}\right)_i$ is the ith column of the matrix square root of $(L+\lambda)P^{a}{}_{k-1|k-1}$.

The sigma points are propagated through the transition function *f*

$$X^{i}{}_{k|k-1} = f(X^{i}{}_{k-1|k-1})\ i = 0 \ldots 2L \tag{42}$$

The weighted sigma points are recombined to produce the predicted state and covariance.

$$\widehat{X}_{k|k-1} = \sum_{i=0}^{2L} W_s^i X^i{}_{k|k-1} \tag{43}$$

$$P_{k|k-1} = \sum_{i=0}^{2L} W_c^i [X^i{}_{k|k-1} - \widehat{X}_{k|k-1}][X^i{}_{k|k-1} - \widehat{X}_{k|k-1}]^T \tag{44}$$

where the weights for the state and covariance are given by:

$$\begin{gathered} W_s^0 = \frac{\lambda}{L+\lambda} \\ W_c^0 = \frac{\lambda}{L+\lambda} + 1 - \alpha^2 + \beta \\ W_c^i = W_s^i = \frac{1}{2(L+\lambda)} \\ \lambda = \alpha^2(L+k) - L \end{gathered} \tag{45}$$

α and k control the spread of the sigma points. β is related to the distribution of X. Normal values are $\alpha = 10^{-3}$, k=0 and $\beta = 2$ if the true distribution of X is Gaussian , $\beta = 2$ is optimal.

For the present approach these normal values were used.

Update

The predicted state and covariance are augmented as before but this time with the mean and the covariance or the measurement error.

$$X^a{}_{k|k-1} = \left[\widehat{X}^T{}_{k|k-1}\, E\left[V_k^T\right]\right]^T \tag{46}$$

$$P^a{}_{k|k-1} = \begin{bmatrix} P_{k|k-1} & 0 \\ 0 & R_k \end{bmatrix} \tag{47}$$

As before, a set of 2L + 1 sigma points is derived from the augmented state and covariance where L is the dimension of the augmented state.

$$X^0{}_{k-1|k-1} = X^a{}_{k-1|k-1}$$

$$X^i{}_{k-1|k-1} = X^a{}_{k-1|k-1} + \left(\sqrt{(L+\lambda)P^a{}_{k-1|k-1}}\right)_i, i = 1 \ldots L \tag{48}$$

$$X^i{}_{k-1|k-1} = X^a{}_{k-1|k-1} - \left(\sqrt{(L+\lambda)P^a{}_{k-1|k-1}}\right)_{i-L}, i = L+1, \ldots 2L$$

Alternatively, if the UKF prediction has been used, the sigma points themselves can be augmented along the following lines

$$X_{k|k-1} := \left[X^{T}{}_{k|k-1}\ E\left[V_k^T\right]\right]^T \pm \sqrt{(L+\lambda)R^{a}{}_{k}} \tag{49}$$

with $R_k^a = \begin{bmatrix} 0 & 0 \\ 0 & R_k \end{bmatrix}$

The sigma points are projected through the observation function:

$$\gamma^{i}{}_{k} = h(X^{i}{}_{k|k-1})\ i = 0 \ldots 2L \tag{50}$$

The weighted sigma points are recombined to produce the predicted measurement and predicted measurement covariance.

$$\hat{Z}_k = \sum_{i=0}^{2L} W_s^i \gamma^{i}{}_{k} \tag{51}$$

$$P_{\approx k \approx k} = \sum_{i=0}^{2L} W_c^i [\gamma^{i}{}_{k} - \hat{Z}_k][\gamma^{i}{}_{k} - \hat{Z}_k]^T \tag{52}$$

The state-measurement cross-covariance matrix,

$$P_{x_{k\approx k}} = \sum_{i=0}^{2L} W_c^i [X^{i}{}_{k|k-1} - \hat{X}_{k|k-1}][\gamma^{i}{}_{k} - \hat{Z}_k]^T \tag{53}$$

is used to compute the UKF Kalman gain.

$$K_k = P_{x_{k\approx k}} P_{\approx k \approx k}{}^{-1} \tag{54}$$

And the updated covariance is the predicted covariance, minus the predicted measurement covariance, weighted by the Kalman gain.

$$P_{k|k} = P_{k|k-1} - K_k P_{\approx k \approx k}{}^{-1} K_k{}^T \tag{55}$$

5.1.5 **Bayesian Estimators**

The main characteristics of these methods are that, as KF, they can deal with the uncertainty associated to the system and the measurements. Furthermore, the main advantage is that they are not as restricted as KF. With KF the output of the filter is estimation with a Probability Density Function (PDF). The main difference with respect to KF is that in this case this PDF is a generic solution not specific for a parameterized Gaussian.

The Bayesian methods are based in a different model than the presented before in equations (17) to (20) . In this case, the systems are defined by a PDF that characterizes the evolution of the state $p(X_t|X_{1:t-1})$ vector and another for the measures vector $p(Y_t|X_t)$. Also the output of the estimator is not a deterministic value but a PDF associated to the estimation of the estate vector, based only on the observed data $p(X_k|Y_{1:k})$, It is also called the posterior distribution. Where k is the actual time t_k and $Y_{1:k}$ are the measurements from the initial instant t=0 to the last measure t_k. This value is calculated using the Bayes rule in the recursive Bayesian Filter.

$$p_{(X_k|Y_{1:k})} = \frac{p_{(Y_k|X_k,Y_{1:k-1})}p_{(X_k|Y_{1:k-1})}}{p_{(Y_k|Y_{1:k-1})}} \tag{56}$$

Using the assumptions of the Markov model (state X_k is only dependant on the previous state and the input) and the fact that $p_{(Y_k|Y_{1:k-1})}$ is the normalization factor that usually is represented as ŋ. Equation (56) can be converted to:

$$p_{(X_k|Y_{1:k})} = ŋp_{(Y_k|X_k)}p_{(X_k|Y_{1:k-1})} \tag{57}$$

Thus the recursive formulation would be as follows:

$$p_{(X_k|Y_{1:k})} = ŋp_{(Y_k|X_k)} \int p_{(Y_k|X_{k-1})}p_{(X_{k-1}|Y_{1:k-1})}dX \tag{58}$$

In equation (58) it is represented both stages prediction and correction:

- Prediction

  Where the a priori probability of the state vector is obtained for the time t_k this is obtained from the previous estimation $p(X_{k-1}|Y_{1:k-1})$

$$p_{(X_k|Y_{1:k-1})} = \int p_{(Y_k|X_{k-1})}p_{(X_{k-1}|Y_{1:k-1})}dX \tag{59}$$

- Correction

  The *a priori* probability in (59) and the observation model given in $p_{(Y_k|X_k)}$ are used to obtain the estimation value in the time t_k.

Once the Bayes filter has been presented, it has to be remarked that there is no implementation of the Bayes Filter in the continuous space. This is because the integral that it includes makes it necessary to take into account all the possible states. Thus, there are different versions of the Bayes Filters according to the different assumptions done. Kalman Filter explained before it is the only implementation of the Bayes Filter in the continuous space and it is done thanks to the assumptions related to the nature of the model and the noise involved.

All the other implementations of the Bayes Filter are obtained by discretizing any of the PDF involved in the filter. Depending on the variable discretized, different algorithms that have been used to solve different problems can be found.

One of the typical problems solved with the Bayes filter is Mapping, localization or Simultanous Mapping and Localization (SLAM), by using the discretization of the state space using cells in the mapping procedure. These cell maps or grid maps [101] are commonly found also for automotive and fusion applications [102] [31] [103][104] and [105].

Particle Filter is a discretized version of the Bayes Estimator that makes it more computationally efficient. It uses a discretized a posteriori probability using a space including only the most probable values of it. Thus, it is possible to give a robust and accurate estimation using Bayes Filter.

The estimation given by the particle filter, using the classical approaches, is based in the uni-modal approach [106]; it means that a particle filter would be necessary for each target. This classic model was used in the present approach, test results are presented in chapter 6. Recent approaches try to give multimodal solution to the estimation problem, thus a single particle filter can be used to track all the targets at once [107].

5.2 Association Methods

New observations provided by the sensors should be correlated with the estimated from previous detections in order to track the movements of the targets. Association methods are crucial in Multiple Target Tracking mainly in clouded environments where it is difficult to differentiate among detections that correspond to the different targets. In this thesis, especially taking into account pedestrians and vehicle detections, the process in eventual situations can become very cloudy, thus this step has to be carefully taken into account. For this reason, several association methods were tested and they are detailed in the present chapter, providing several possible configurations. These configurations and the different estimation techniques that were tested provide a set of solutions based in Data Fusion, able to adapt to a wide variety of situations in road environments.

The first step in any association method is to reduce the candidates for each association. This technique, called Gating, consists of the reduction of the association possibilities by reducing the candidates for a given association to the most likely. Not likely associations, such as distant or low probability ones, are discarded. As a result, the computational cost of the subsequent association algorithm is reduced.

In this section, the theoretical definition of the association technique is provided. The next section depicts the Data Fusion implementations proposed for road environments.

5.2.1 Gating Techniques

The idea of gating is to reduce the amount of measurements taken into account when performing the association with old tracks. Using gating techniques, unlikely pairs are eliminated. The algorithm can then focus on the most likely pairs, reducing the computational costs.

Basic tracking algorithms, mainly based on the Kalman Filter, use the measurement and standard deviations to create some of the simplest gating techniques determined by the accuracy of the measurements. These basic techniques are Rectangular Gates and Ellipsoidal Gates [15] and [16].

Rectangular gates utilize a rectangle using the mentioned standard deviation. Hence, only new detections within the limits of the rectangles, calculated using equations (60) and (61), are taken into account for the association method.

$$|y_l - \hat{y}_l| \leq K_{Gl}\sigma_r \tag{60}$$

where y_l is the new observation and $\hat{y}_l$ the estimated position of the tracks, thus only the observation which difference from the previous track match this inequity are taken into account in the data association process. And σ_r is defined, taking into account the errors of the observation and the predictions (equation (61)). K_{Gl} is a constant value that is defined according to the application. Typically it's chosen $K_{Gl} \geq 3$.

$$\sigma_r = \sqrt{{\sigma_o}^2 + {\sigma_p}^2} \tag{61}$$

where σ_o is the measurement error and $\sigma_p$the appropriate diagonal element taken from the KF covariance matrix.

Ellipsoidal Gates use ellipsoids based in standard deviation to perform the gating [15] and [16].

Once the Gating procedure is performed, next step in the association process is to perform the association according to the method selected.

5.2.2 Global Nearest Neighbor (GNN)

This approach is also defined as single hypothesis tracking, this is due to the fact that only the most likely hypothesis in each detection is taken into account, and all other options are discarded. This is the simplest and probably the most widely applied method for tracking.

Track association methods also have to deal with the track deletion and track creation. Here GNN can deal with scores that can be computed to creation and deletion, but the most common method used is a simple M/N rule for confirmation and a N_D consecutive misses for track deletion.

GNN's main difficulty is the association of common or conflict situations. It is typically found in a cloudy situation when more than one observation is within the limit of the gating (Figure 5.3).

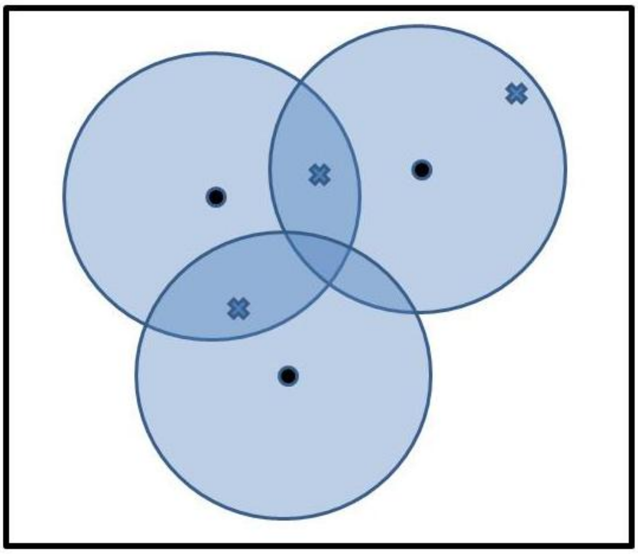


Figure 5.3. GNN uncertainty problem. Crosses represent new observations, and points the old tracks, circles are the Gating. This situation some of the new detections are difficult to assign to the old tracks because they fall in the gating of more than one track.

To solve this problem, the metric, or cost, used to decide which is the most likely association is important. One example of it, presented in equation (62) is the generalized statistical distance that penalizes those tracks with less certainty [108] and [109].

$$d_{G_{ij}}^2 = d_{ij}^2 + ln\left[\left|S_{ij}\right|\right] \tag{62}$$

where d_{ij} is the distance and $ln\left[\left|S_{ij}\right|\right]$ is the logarithm of the residual covariance .

Given this distance shown in equation (62), the typical approach uses a matrix, called assignment matrix, that combines the old tracks, detections and new tracks. The way of solving this assignment for a given matrix can vary and is dependent on the available information and the metric [15].

5.2.3 Multiple Hypothesis Tracker (MHT)

This method is based on the idea of tracking all the possible hypotheses along time [68]. This way each iteration gives a global solution at time t_k.

This association method implicit generates new targets; therefore, it is especially useful for MTT applications. But the main disadvantage is that the amount of tracks that it exponentially generates is impossible to deal with after a few iterations. Here is where simplifications of the basic method allow creating real time applications.

- Pruning methods that allow eliminating or merging different tracks, whether or not likely to be real or to put together more similarly.
- Temporal windows that allow eliminating those tracks that haven't been found during the last iterations. This process can be included in the previously presented pruning process and are very common, not only in MHT but other association methods.
- Gating. The idea is to use the gate to limit the amount of tracks. As a result, only the observation inside the gate (the most likely) will be added to the multiple hypotheses set of tracks.

Kalman Filter is the most typical estimator that is used with this association method [68], [110] and [103].

Once the estate vector is predicted for each track, and the association is done for the given observations, the estate is actualized for each track. When the model allows more than one association for each target, the state correction shown in equation (22) (so called innovation or residual measurement) is corrected through the combination of statistical values (e.g. mean) of all hypotheses. The resulting PDF is a Gaussian mixture because in this case, KF is being used. Where the system is not linear EKF [111] or PF [112]can be used.

5.2.4 **Probabilistic Data Association (PDA)**

These methods calculate the probability that a given observation corresponds to a given object being tracked or does not correspond to it, and use it, not only to select which is the best option, but also to actualize the filter.

Here we should introduce the clutter, it is a fake measurement that is usually added for mathematical integrity and represents a non-association. If a track is associated with the clutter, it means there is no association for a track in that given time.

PDA is a process created to track a single object, where it checks the probability of an observation to be associated or not with the track. Joint Probabilistic Data Association, which will be explained later, is the extension of PDA that takes into account multiple target interactions. Another property of this algorithm is that on its original version it takes into account all the measurements from the beginning to a given time t_k.

Gating is usually introduced to reduce the computational costs as well as other assumptions that help the algorithms to be feasible for real time, even though it compromises the effectiveness of the algorithms. Some of these assumptions are pruning low probability values, or to use only the instantaneous values for a time t_k instead of all of the historical or other values.

Typically PDA is found in target applications with a Kalman Filter. This is called Probabilistic Data Association Filter (PDAF), and it is similar to MHT, described earlier. The main difference is that it is single target oriented and probabilistic, thus the measurements $Y_{1:m}$ (where m is the number of observations available in a time t_k) are combined in the update stage associated by its probability $p_{1:m}$ that is generated by the PDA algorithm [97].

Equations (23) and (26) of the KF becomes equations (63) and (64):

$$R_k = \sum_{i=1}^{m} \left[p_i \left(Z_{i,k} - H_k \hat{X}_{k|k-1} \right) \right] \tag{63}$$

$$\hat{X}_{k|k} = \hat{X}_{k|k-1} + K_k R_k \tag{64}$$

In equation (63) and (64) the innovation or residual R_k is calculated, taking into account all the different possibilities according to their probabilities $p_{1:m}$ given by the PDA algorithm.

5.2.5 Joint Probabilistic Data Association (JPDA)

PDA is object oriented, and thus approaches with multiple targets are not easily implemented. Extension using different PDAF for each target could be done, but the lack of interaction between tracks (e.g. track crossing) makes it an inefficient tool when dealing with multiple tracks.

JPDA is the extension of the PDA, taking into account the interaction among tracks. Each probability value is calculated, conditioned to the rest of measures giving $(n+1)!^m$ solutions to the joined association, where n is the number of observations and m is the number of overall measures. It defines the k+1 association hypotheses to each target through the joining of probability values $p_{1:m,0:k}$.

The computational costs of these approaches are high due to the amount of variables taken into account. As PDA, several approaches have tried to overcome these limitations by gating and pruning, among other procedures. On the other hand, the accuracy of this approach makes it a very interesting option, when accuracy is the main goal over other limitations.

Application of KF with JPDA is called Joint Probabilistic Data Association Filter JPDAF. This procedure is similar to the previously presented PDAF. The main difference is that in this case Kalman Filters are presented, one for each track, and in the update stage the joined probability $p_{1:m_j,0:k}$ is used to obtain the residual measurement of each one.

Blackman, in his book [15], gives an example that helps to describe the method in a fast and easy way:

Given an association problem with three observations and two tracks, the distances of the observations to the predictions of the gates are represented in Table 3 and shown in Figure 5.4.(∞ Refers to observations out of the gates).

Observations \Tracks	1	2
1	1	∞
2	$\sqrt{2}$	$\sqrt{2.5}$
3	2	$\sqrt{3}$

Table 3. Distances from the observations to the predicted positions of the tracks in the JPDA example given by Blackman in [15].

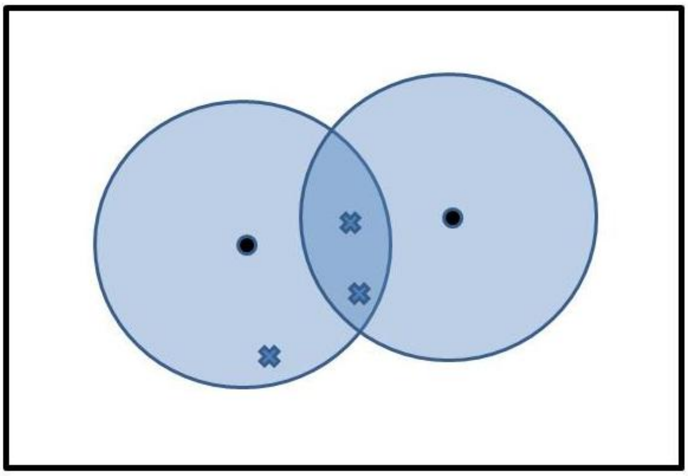


Figure 5.4. Example representation. Two tracks and three observations.

Besides, assuming the likelihood function associated with the assignment observation j to track i:

$$g_{i,j} = \frac{e^{-\frac{{d_{i,j}}^2}{2}}}{(2\pi)^{M/2}\sqrt{|S_{ij}|}} \tag{65}$$

where $d_{i,j}$ is the distance between the prediction and the observation given in Table 3. M is the dimension and $S_{i,j}$ is the residual covariance matrix. In the example used, since a Cartesian example was used (as it is in the approach) $\sqrt{|S_{ij}|} = \sigma_x\sigma_y$ and M=2 .

Besides two probabilities should be defined β and P_D as follows:

- β represents the false positive probability. For the example β=0.003 was considered,
- P_D is the probability of detection. In the example P_D=0.7.

Table 4 depicts the results for the likehood of each of the hypotheses where a 0 corresponds to an unassigned track in a given observation. A common factor of β^N appears in the case (as it is in the example) where the number of observations is higher than the tracks and N is the difference between those numbers. In the case where the number of tracks is higher than the observations, the common factor would be (1- P_D). The normalized probability $P(H_i)$ is obtained from the equation (66).

$$P(H_i) = \frac{P\,(H_i)}{\sum_{i=1}^{N} P\,(H_i)} \tag{66}$$

#Hypothesis	Track 0	Track 1	Likehood of the Hiphotnesis $P'(H_i)$	Normalized probability $P(H_i)$
1	0	0	$(1-P_D)^2\beta^3$	0.011
2	1	0	$g_{11} P_D (1-P_D) \beta^2$	0.086
3	2	0	$g_{12} P_D (1-P_D) \beta^2$	0.053
4	3	0	$g_{13} P_D (1-P_D) \beta^2$	0.019
5	0	2	$g_{22} P_D (1-P_D) \beta^2$	0.041
6	1	2	$g_{11} g_{22} P_D^2 \beta$	0.306
7	3	2	$g_{13} g_{22} P_D^2 \beta$	0.068
8	0	3	$g_{23} P_D (1-P_D) \beta^2$	0.032
9	1	3	$g_{11} g_{23} P_D^2 \beta$	0.239
10	2	3	$g_{12} g_{23} P_D^2 \beta$	0.145

Table 4. Likehood values for the example.

Finally, to compute the probability of a given observation i to be assigned to a track j, all of the hypotheses with this assumption should be taken into account:

$p_{10} = P(H_1) + P(H_5) + P(H_8) = 0.084.$

$p_{11} = P(H_2) + P(H_6) + P(H_9) = 0.631.$

$p_{12} = P(H_1) + P(H_{10}) = 0.198.$

$p_{13} = P(H_4) + P(H_7) = 0.087.$

$p_{20} = P(H_1) + P(H_2) + P(H_3) + P(H_4) = 0.084.$

$p_{21} = 0;$

$p_{22} = P(H_5) + P(H_6) + P(H_7) = 0.415.$

$p_{20} = P(H_8) + P(H_9) + P(H_{10}) = 0.416.$

Finally the assignation would be observation 1 to track 1 and observation 2 or 3 to track 2.

One of the main discussions of the JDPA methods is the track creations, since in the previously presented example two tracks were considered and neither creation nor deletion is taken into account. In the proposal explanation, novel track creation logic is proposed, taking into account the Data Fusion nature of the proposal, using two kinds of tracks consolidated and non consolidated.

Finally, after the association algorithms each one of the association possibilities, within the gate, is used to update the Kalman Filter used for a given track.

$$R_{k,j} = \sum_{i=1}^{m} \left[p_{i,j} \left(Z_{j,k} - H_{j,k} \widehat{X}_{j,k|k-1} \right) \right] \tag{67}$$

where m is the number of observations within the range of the gate and j is the corresponding track. $p_{i,j}$ is the corresponding probability that a given observation i corresponds to a track j, obtained in the association process.

5.3 Tracking and Data Association Fusion Procedures for Road Environments

After the previous explanation of the different methods available for the estimation and association process, the next section depicts the algorithms constructed following the previous explained methods. MTT application based in Data Fusion is intended by using previous algorithms and combining them in different configurations. Chapter 6 will provide an extent relation of the different test performed to the different algorithms that are described in this section.

Before dealing with the research of the best association solution, estimation procedure should be checked. Three methods were developed and tested, KF, UKF and PF. All of these methods are based on the presented in this chapter. Later, using the GNN approach, these algorithms are tested and results are presented in chapter 6. Besides, in the estimation process three association methods were created. The first is based on GNN, which is the simplest, and represents the basis of the subsequent procedures. Later, the MHT and JPDA methods were created and tested to check the advantages or disadvantages of the different procedures by giving an extent comparison of the different possibilities available.

5.3.1 Target Model

Thanks to the high frequency of the laser scanner, a modeling of different obstacles using constant velocity model could be provided. This approach is not entirely efficient, mainly in the case of vehicles when performing lateral movements. But, for the scope of this application where the vehicles detected were mainly in interurban scenarios, lateral movements are not common and the fast detection of the laser scanner made this approach reliable enough despite the limitations. A model to track pedestrians and vehicles using the constant velocity model is given, modeling accelerations as the system errors. Equations (68) and (69) present the system error Q and measurement error R covariance matrixes.

$$Q = \begin{bmatrix} \frac{{a_x}^2 t^3}{3} & \frac{{a_x}^2 t^2}{2} & 0 & 0 \\ \frac{{a_x}^2 t^2}{2} & {a_x}^2 & 0 & 0 \\ 0 & 0 & \frac{{a_y}^2 t^3}{3} & \frac{{a_y}^2 t^2}{2} \\ 0 & 0 & \frac{{a_y}^2 t^2}{2} & {a_y}^2 \end{bmatrix} \tag{68}$$

$$R = \begin{pmatrix} {\sigma^2}_{\epsilon,x} & 0 \\ 0 & {\sigma^2}_{\epsilon,y} \end{pmatrix} \tag{69}$$

where ${\sigma^2}_{\epsilon,x}$ y ${\sigma^2}_{\epsilon,y}$ is the standard deviation for the measures in x, y coordinates. These deviations have been calculated using test sequences. As both systems share the ROI coordinates, the deviation in the measurements is considered equal for both detections.

The values a_x and a_y in equation (69) is the maximum amplitude of the acceleration in every axe, In the case of human as it is pointed [113], it is $11 m/s^2$. For vehicles it was estimated up to $3\ m/s^2$.

Constant velocity model equations are depicted in Equations (70) to (73).

$$\hat{X} = \begin{bmatrix} x \\ y \\ v_x \\ v_y \end{bmatrix} \tag{70}$$

$$\hat{Y} = \begin{bmatrix} x \\ y \end{bmatrix} \tag{71}$$

$$H = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix} \tag{72}$$

$$A = \begin{bmatrix} 1 & 0 & t & 0 \\ 0 & 1 & 0 & t \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{73}$$

where v_x and v_y are the speed of the pedestrian and t is the time elapsed, $\hat{X}$ is the state vector, $\hat{Y}$ the measurements vector, H the transition matrix and A the state transition model.

5.3.2 Estimation Methods

Estimation methods developed for the present thesis included the following three: Kalman Filter, Unscentered Kalman Filter and Particle Filter. The next chapter depicts the results provided for each method in the different test performed.

Kalman Filter approach was created using the lineal model used. Unscented Kalman filter allows modeling the linearity errors thanks to the unscentered solution previously explained. Finally, particle filter formulation allows a more general solution, in exchange of computational costs. These three methods were tested using Global Nearest Neighbors solution that is explained in the following section.

5.3.3 Global Nearest Neighbor Solution for Data Fusion in Road Environments

The simplest solution when dealing with any fusion applications is a GNN approach. It consists, as it was explained before, in the selection of the most suitable solution at a given time and using them in the subsequent scans, discarding the less likely ones.

This approach represented the basis of subsequent approaches that use more complexes solutions (i.e. MHT and JPDA) also created in the scope of the present work. This is the reason that in chapter 6 the different tests that are presented compare the results obtained with these two approaches with the basic GNN approach. Besides, the simplicity of the algorithm for both developing and configuring made it the best solution for testing the estimation filters. Thus, the GNN approach was used to test the estimation filters presented before, for the target model presented.

a) Gating and data association

Gating is performed using a square approach (74):

$$K_{Gl}\sigma_r \tag{74}$$

where $\sigma_r$is the residual standard deviation and K_{Gl} is a constant that was empirically chosen .

After gating, association is performed using normalized distance and a stability factor, giving less priority to less stable measures:

$$d^2 = \frac{(x_i - \overline{x})^2}{\sigma_x^2} + \frac{(y_i - \overline{y})^2}{\sigma_y^2} + \ln(\sigma_x \sigma_y) \tag{75}$$

where d is the computed distance between previous and presented tracks to be associated. And (σ_x, σ_y) the appropriate values of covariance matrix of Kalman Filter.

Assignment Matrix was used to track association, following a least overall cost assignment [16] and [15].

b) Track creation and deletion logic

Track creation and deletion policy follows a logic that was found empirically, and depicted in the next table. Two kinds of tracks were defined, consolidated and non consolidated. First corresponds to track detected by both sensors whether concurrently or in subsequent scans. Later means that a track has been detected by a single sensor, thus it is not trustable enough since the other sensor have not corroborated it.

Track vs New observation	**Single sensor**	**Both sensors**	**No match**
Non consolidated	If sensor 1= detected & sensor 2 =detected them track **consolidated**. Otherwise **non consolidated**. Track **updated**.	Track **consolidated**. Track **updated**.	If #consecutive_no_detections > 4 them Track **eliminated**.
Consolidated	Track **updated**.	Track **updated**.	If #consecutive_no_detections > 5 them Track **eliminated**.
No match	New **non consolidated** track	New **consolidated** track.	

Table 5. Track management logic for GNN solution, according to the sensors that detects it in the updating process.

The use of consolidated and non consolidated tracks helps the system to add reliability. Only consolidated tracks are considered trustable detections. Hence, only they are reported. This way false positives, mainly from the laser scanner, are discarded because detections that are not corroborated by the vision system are not reported. Furthermore, the use of both sensors, once the track is consolidated, to update the tracks, allows that once a pedestrian is detected it can be tracked, even if it is not detected by one of the subsystems, e.g. when it goes out of the camera field of view. These implications of the Data Fusion algorithm are tried in section 6 where the different tests performed are detailed. The results of the fusion algorithm obtained enhance the basic performances of the different sensors independently.

c) Estimation

As it was explained before, the association methods used for the GNN solution were three, KF, UKF and PF. The solution to the different tests performed, and the conclusions are given in Chapter 6 and 7 respectively.

5.3.4 **Multiple Hypothesis Tracker approach for Data Fusion in Road Environment**

It was previously explained that MHT approaches take into account all the possible combinations of the tracks with the new observations.

One of the advantages of this approach is that it tracks all possible combinations. Thus, if in one of the time steps one misassociation is produced, it can be corrected in subsequent steps because this combination is tracked along time. It also provides the best global solution for a given time. Figure 5.5 represents the process that takes part in the MHT fusion procedure.

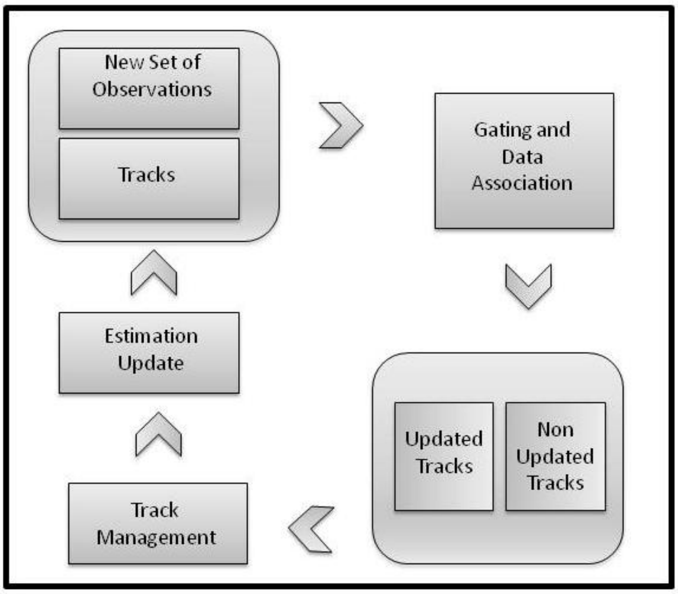


Figure 5.5 Processes for MHT Data Fusion approach for road environments.

But also an important drawback arises: to track all the solutions there is an increase the computational costs and the resources required. Also, misdetections can be tracked during longer time, since all the solutions are taken into account. Thus, a mechanism to allow pruning the unlikely tracks is necessary.

a) Gating and data association

A similar procedure to the GNN Gating presented in equation (74) was used to reduce the number of pairs taken into account, and as a result reducing the computational cost of the algorithm.

A single association per observation was allowed, thus only the selected association would be updated regarding to the fusion information (i.e. detections performed by each sensor). All of the remaining possibilities would not be considered updates, although for estimation purposes there is state actualization. This way even though non-optimal matches would keep track of it after some iterations, the non updated combinations are eliminated.

As in GGN approach, the Assignment Matrix was used to associate a given observation to the tracks. The distance definition remains the same as depicted in equation (75).

The association algorithm presented creates automatically new tracks, since every possibility creates a different track. The previously explained updating policy allows that those tracks, which have a probability of being a false match, are rapidly eliminated after a few iterations because every detection only computes for a single track. Although in the presence of several tracks it can become unstable, as it is explained in chapter 6.

b) Track management

An MHT algorithm implicit creates tracks. Only in those iterations where a given observation does not match any of the tracks a new tack is created.

The track update follows the policy that was used in GNN approach and presented in Table 5. Only consolidated tracks are considered to be real detection. In this approach the number of tracks to follow can be higher mainly in a cloudy environment, although it is an important drawback, the solution obtained at the given time can be considered closer to the global solution since it takes into account combinations of longer time window.

c) Estimation

The estimation method used for this approach was KF, since the results obtained for the different estimation techniques did not differ too much, and the KF approach represented a fast and reliable procedure easy to be adapted to the MHT solution, explained in the previous section. Single association was allowed for each observation, thus no changed should be necessary to the standard KF approach explained.

5.3.5 Joint Probabilistic Data Fusion Association Filter for Road Environments

Data fusion approach using the JPDAF (JPDA Filter) was proposed to overcome the limitations of the basic GNN approach and given the instability of the solution proposed with MHT, obtained in the different test and that is explained in chapter 6. It is based on the previous explanation of JPDA and combined with DF problem for road environments.

Three steps were created in the association algorithm created for the proposal. Following a loop similar to the presented in GNN and MHT approaches the three steps to follow are:

- Assignation process. Where all the joint probabilities are calculated. And a single assignation is performed for each track.
- Filter updating. Following equation (52) each track filter is updated with information of all the observation within the gate of the track.
- Track management. New tracks are created and deleted following the logic given by the fusion procedure.

a) Assignation process

In this step the probabilities for all possible hypotheses are computed and the joined probabilities calculated. Thus, an assignation matrix is created, where each row represents an observation, and each column a track. As a result, probabilities for all the combinations are computed. The assignation is performed according to this matrix. Probabilities are calculated using (65) and (66).

The assignation is performed following a 1/1 assignment. It means that only one track is assigned to a given observation. In this way, an observation only can be assigned to a single track, all tracks with no assignment would increase their counter for non-detection, and if the counter reaches to a given value, they are eliminated.

b) Filter update

According to equation (52) all the possible associations for a given track should be taken into account in the actualization process; thus, the correction is performed using the joined probabilities calculated.

This process of updating stage takes into account all the detections within the gate. It is possible that a given observation is used in the update stage of the filter for more than one track, although it is considered to belong to a single track for the track management process. Thus, for track management a single assignation policy is followed, a given observation is used only to update the track logic of a single track. This dual behavior is one of the main differences to the classic JPDA filter applications.

c) Track management

Track management logic follows logic equal to the logic created in the GNN approach, and a single observation can only update the status of a single track. Consolidated and non consolidated tracks are also used, but only consolidated tracks are reported because they are the only considered reliable enough. Thus only the assigned observation can contribute to change the current status of a single track from non consolidated to consolidated. Table 5 depicts the logic of this updating process.

Track creation logic is different in this case. Only when an observation is out of the gate of any track, a new track is created. This solution is interesting to avoid false positives related with more than one camera positive detection of the same pedestrian as it will be detailed in chapter 6. On the other hand, two very close pedestrians could represent a problem because this algorithm could have difficulties to differentiate among them. But two considerations should be taken in these situations:

- First, the laser scanner itself already needs the pedestrian to be separated; otherwise it would consider it a single obstacle due to the segmentation process explained in chapter 4.
- JPDA association process would use both observations to perform the tracking. Thus, although they both are considered to be a single pedestrian, both detections are used in the KF actualization process. Hence, the error is minimized.

A test demonstrated that a previously presented algorithm could, in certain situations, reach unstable behavior. This happens when several tracks compete for a single observation. In these situations, the cluster is the most powerful option due to the weight of the joined probabilities of the other options different from the track to be calculated in a given time (joint probabilities association process penalizes when there is going to be another track, with high probability, that is not going to be assigned to that observation). To overcome this problem, a special behavior was created, based on the fact that once the probabilities are assigned, a given track is eliminated from the assignation process, and once again all the joined probabilities are calculated with the remaining tracks. This way the problem is avoided by eliminating the weight of the already assigned solutions in subsequent assignations. In the case of several tracks pointing to a single observation, this solution would first assign cluster to the less probable and eliminate its weight in subsequent assignations until one of them is selected as more likely than the cluster. Different tests prove that the computational cost added because of the necessity to recalculate the joining probabilities is negligible and the system proved to be stable.

5.4 Danger Estimation

Previously presented detection and tracking procedures represent a step forward for pedestrian and vehicle detection systems for road safety applications, but to complete the fusion application Fusion Levels 2 and 3 must be detailed. In this section, situation and threat assessment is intended, using context information to augment the capacity of the system, creating a reliable and robust application. Both context information and GP-inertial device provide relevant information to estimate the danger involved in any detection. Later allows the current status of the vehicle e.g. velocity and position to be known. First, is important to know the real danger involved in any detection.

The nature and the behavior of the two road users (vehicles and pedestrians) that the present approach tries to detect makes it mandatory to differentiate the danger and situation assessment calculation for each one. Therefore, the present section is divided in two parts. The first deals with the danger involved with pedestrian detected in the surroundings, and the second with the vehicles detected.

5.4.1 Pedestrian

Before giving an estimation of the danger that involves any pedestrian detection, two distances should be taken into account, breaking distance and response distance. These distances represent context information that should be taken into account to calculate the danger that any detection involves. The first represents the elapsed space before the car is completely stopped, thus collision with any pedestrian farther than this distance can be totally avoided by stopping the car. The second distance represents the distance covered in the time that the driver needs to respond to a given stimulus.

These distances are not the only context information used in this danger estimation. Other information relative to vehicle safety is necessary to calculate this danger. Response time for drivers and some traffic accident reconstruction mathematics are two examples of the context information necessary for danger estimation that will be explained in this section.

a) Response distance

Research generally accepts response times up to 0.66 seconds, as it was shown by Johansson and Rumar [114]. In this article, authors showed that the mean response time for human beings when driving by means of an auditory stimulus is 0.66 seconds. Other authors in recent works have done similar test with similar results [115]. Thus, response distance is the distance that a vehicle with a given velocity would cover during the response time of 0.66 seconds.

b) Braking distance

It is the distance that the car would cover until it completely stops. Many different variables would affect this calculation. The present approach uses basic traffic accident reconstruction mathematics [116] based in the worst case scenario, when the vehicle is fully loaded. Weather conditions may also change the conditions (e.g. road coefficient). Here, some on-line contextual information that the inertial device provides could be useful, such as the temperature measurement.

In a traffic accident reconstruction, worst case scenario means that only front wheels are blocked when braking, this fact displaces the weight of the car to the front of the vehicle. This weight displacement is represented as a change in the friction coefficient; this change is depicted in equation (76).

$$ŋ = \frac{b_2}{L - h\mu} \tag{76}$$

where ŋ is the corrected and μ the real friction coefficient, b_2 is the distance to the rear axis from the mass center, L the longitude of the vehicle and h the height of the mass center. Mass center has to be calculated, but several authors give an approximation of 0.4 the height of the car that was used for this application. Using this approximation, the distance of the vehicle to completely stop is shown in equation (77).

$$d_{stopping} = \frac{v^2}{\eta \mu g} \tag{77}$$

where v is the speed of the vehicle and g the gravity acceleration

But equation (77) is not the braking distance, since the response time presented before should be taken into account because it is the time before the driver starts pressing the brake pedal (equation (78)).

$$d_{braking} = v t_{response} + d_{stopping} \tag{78}$$

c) Danger zones

Danger zones are created for pedestrian safety according to previous relevant distances. These zones help the system to quickly evaluate the degree of danger that any detection involves. Each one of the zones is created according to the different actions performed in case a pedestrian is found in the zone. Table 6 depicts the relation between the zones and the distances.

	From	**to**
Safe zone	Infinite	Braking distance
Danger zone	Braking distance	Response distance
Imminent Collision zone	Response distance	0 meters

Table 6. Relation between danger zones and relevant distances.

Safe zone detections are those pedestrians that are at a distance far enough to warn the driver and completely stop the vehicle before hitting the pedestrian. The danger zone represents the region where it is possible to warn the driver before hitting the pedestrian, but the vehicle is not going to be able to be stopped on time before hitting it. Finally imminent collision zone is the region in the environment where it is impossible warn or stop the car before colliding with the pedestrian.

Figure 5.6 provides a visual representation of the distances, to give a complete picture of the application the relevant distances are also depicted as well, as the field of view of the different devices.

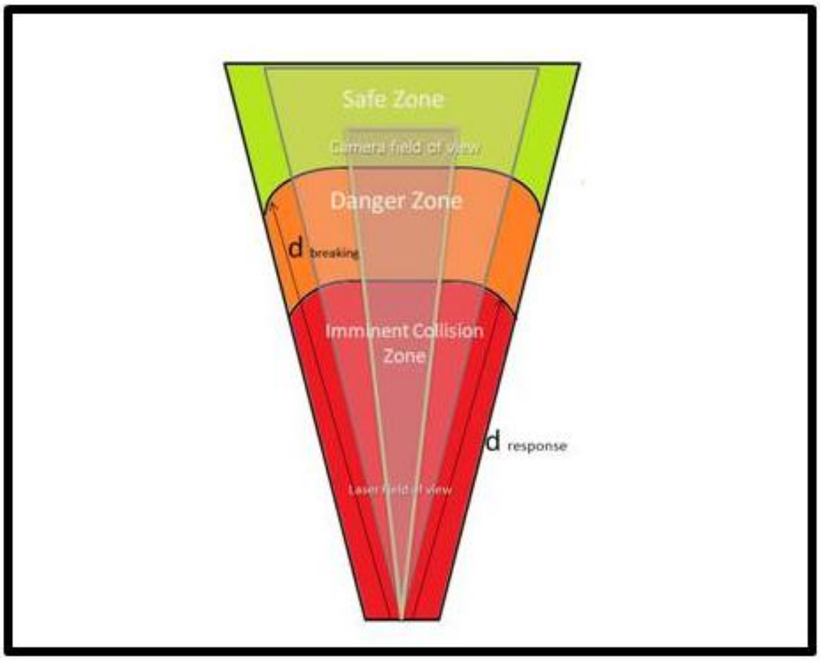


Figure 5.6. Danger zones represented with the relevant distances and the fields of view of the different devices.

The scope of this application is to detect and warn drivers giving an estimation of the danger involved in the detections. Thus, it is out of the scope the actions to perform in any of the detected cases. Further works should deal with this issue, but a first approximation should involve:

- **Safe zone**: These detections involve no imminent danger, thus some visual or acoustic warning may be enough, paying attention to not to saturate the driver with irrelevant information.
- **Danger zone**: Here it is important to warn both, driver and pedestrian to try to avoid the possible collision. Recent works try to determine the safest trajectory to perform automatic avoiding maneuvers, allowing the vehicle to take control over the driver and prevent to harm pedestrians.
- **Imminent collision zone**: Here the only action to take is to trigger any automatic measure to mitigate the harm given to the pedestrian.

d) Danger estimation

Taking previous zones as reference and with the purpose of giving danger estimation to upper layer applications that has to deal with the previously presented situation, a danger estimation function was created. The idea is to give an estimation of the danger involved in any detection. The estimation should grow exponentially the closer the pedestrian is to the vehicle. Furthermore, two values should be taken as reference: In the border between safe zone and danger zone (braking distance), this estimation should be bigger than 0.5 (it was chosen 0.6); second a prior assumption was that at closer distances this function should have a value of 1. Taking all these considerations into account, the following function was created:

$$f(r) = \begin{cases} e^{-\lambda(r-d_r)}, & for \quad 80 \le r \le d_r \\ 1 \quad , & for \quad d_r < r \le 0 \end{cases} \tag{79}$$

where d_r is the response distance, r the distance of the pedestrian to the vehicle, and λ was a value to calculate that would assure the previously presented assumptions (equation (80))

$$e^{-\lambda(d_b-d_r)} = 0.6\text{, thus} \tag{80}$$
$$\lambda = \frac{-\ln 0.6}{(d_b - d_r)}$$

where d_b is the braking distance.

In Figure 5.7 an example for the danger estimation is given for a velocity of 40km/h depending on the distance in meters.

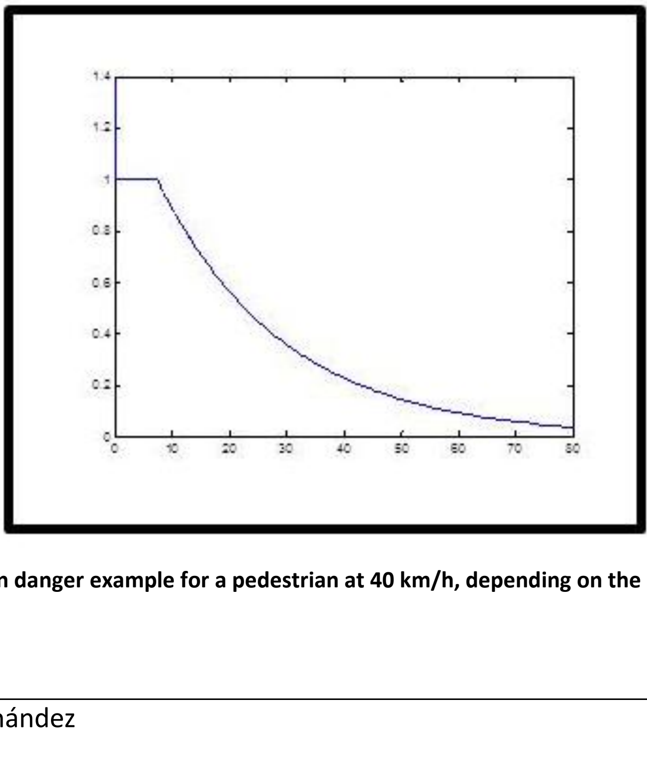


Figure 5.7. Estimation danger example for a pedestrian at 40 km/h, depending on the distance in meters.

5.4.2 Vehicles

Vehicles' distances have some common aspects with pedestrians in danger estimation, mainly related to the physics of vehicles. The main problem related with the vehicles is that the movement of the target vehicle should also be taken into account; therefore, although the problem shares some common physics with the previously explained pedestrian, danger estimation, it represents a completely different solution.

a) Relevant distances

For vehicle safety, two distances are relevant:

Braking distance again is the distance to completely stop the vehicle. It is important to check which of the detections that are in the surroundings of the vehicle should be considered. Other vehicles out of this distance should be tracked but since the interaction level with the vehicle is limited, the danger that involves these detections is limited.

Safety distance is the distance that it is considered safe in relation to another vehicle that is in front of the vehicle. Between these vehicles, a distance lower than this safety distance should be considered a security threat.

b) Braking distance

Braking distance is the same used for pedestrian safety. It is depicted in previous section in equations (76), (77) and (78).

c) Safety distance

Safety distance is usually a subjective measure that depends among other things on the driver situation and road conditions. In [117], [118] and [119] a policy for safety distance in automatic vehicles is presented. This policy can easily be extrapolated to maneuvered vehicles. According to these researches a safe distance S_{di} is defined as following:

$$S_{di} = \lambda_1 (v_1^2 - v_2^2) + \lambda_2 v_1 + \lambda_3 \tag{81}$$

where v_1 is the velocity of the vehicle where the system is mounted (in m/s), v_2 is the velocity of the target vehicle (in m/s). And λ_1, $\lambda_2$and $\lambda_3$are the constant that are defined as follows:

$$\lambda_1 = \frac{1}{2a_{max}} \tag{82}$$

$$\lambda_2 = 2T + \frac{2a_{max}}{J_{max}} \tag{83}$$

$$\lambda_3 = (T + \frac{2a_{max}}{J_{max}} + \frac{2a^2{}_{max}}{J^2{}_{max}})a_{max}T \tag{84}$$

where T is the reaction time presented before, a_{max} is the maximum acceleration and J_{max} the maximum jerk.

5.5 Conclusions

In the present chapter all the different possibilities for tracking and data association were studied. Several solutions were adapted to the fusion problem for road applications, providing different algorithms that were tested in real conditions. Results of the tests are detailed in chapter 6, allowing to understand the advantages and drawbacks of each one. Conclusions of these tests are given in chapter 7.

The tracking algorithms detailed and adapted for the Data Fusion solution for vehicles and pedestrians' detection and tracking were:

- Kalman Filter.
- Unscented Kalman Filter.
- Particle Filter.

For the Data Association problem, three algorithms were adapted for Data Fusion in road applications:

- Global Nearest Neighbor.
- Multiple Hypothesis Tracker.
- Joint Probabilistic Data Association.

In the last part of the chapter, some estimation of the danger that any of the detection involves is given, providing an useful tool to estimate the danger that involves any detection, and as a result providing situation assessment. This way the approach represents a multilevel solution for de Data Fusion problem.

CHAPTER 6. PRACTICAL RESULTS

Different test were performed to check all the algorithms that are presented in the document. Given the nature of the approach, differentiation among vehicles and pedestrians should be done, allowing us to study the performance of the different algorithms in both cases. Thus, a different set of tests were performed for vehicles and pedestrians separately.

It has to be remarked that the proposal presented an important innovation related to the laser scanner detection and classification, thus the present chapter depicts results for each subsystem independently besides the results of the complete system. This way, a complete comparison of the performance of basic approaches and the complete system can be done.

Different tests were performed to check the viability of each of the different algorithms proposed. The different set of tests checked the viability of each algorithm independently and, finally, the viability of the whole fusion system. Four aspects should be checked for each obstacle to detect:

- Each sensor's algorithm performance. First an exhaustive test to the low-level approaches should be made, paying attention to the weakness of each algorithm and the setup of each system in order to provide the proper behavior for the further fusion algorithm.
- Tracking algorithms presented in previous chapters should be tested in real conditions, in order to select the most suitable.
- Once the low-level detection and the tracking algorithms have been tested, the next step is to test the different data association methods presented, in order to provide the one that adapts better in each situation.
- Finally, the whole subsystem should be tested, providing results that prove that the fusion algorithm enhances the basic capacities of the system.

Before performing any test, it is important to calibrate the reliability of the laser scanner and the distances provided by the system, since it is the basis of the coordinate system of both sensor's algorithms. A first test that checks the accuracy of the measurements provided by the laser scanner is performed, this way, standard deviation and correlation between the x,y provided by the laser can be calculated and used in subsequent tests. Besides these measurements, it was important to check the viability of the laser scanner to be used as vehicle detection. To check this viability test involving different kinds of vehicles in performing different movements were performed to check the performance of the laser scanner and proving its usability for vehicle detection.

6.1 Calibration Tests

Two kinds of tests were performed to calibrate the system. The first, regarding pedestrian detection, focuses on the capacity of the laser scanner to give accurate measurement of the pedestrians, providing standard deviation and coordinate correlation. The second test tried to check the viability of the laser scanner to detect moving vehicles. By testing the reflexion behavior under different kinds of vehicles, performing different maneuvers, it could check the viability of the laser scanner used to perform moving vehicle detection.

The first test consisted of test sequences with a single pedestrian performing lateral and vertical movement. In lateral movements y coordinate was fixed, so the system could measure the error in the y coordinate as the pedestrian moves along the x axes. For vertical movements, the pedestrian had x coordinate fixed and moved along the y axis, thus, deviation in x was measured (Figure 6.1). This test was necessary to determine the reliability of the laser scanner system and the capacity of locating pedestrians in the environment, as well as to measure the standard deviations of the measurements for the further estimations test.

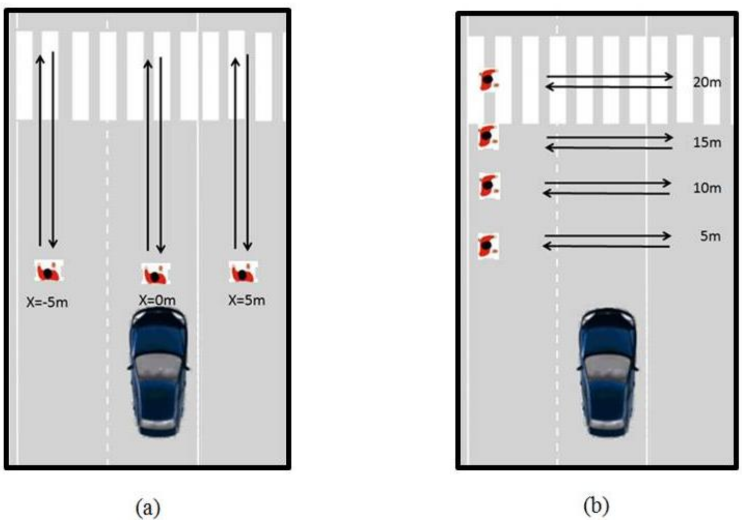


Figure 6.1. Experiments to measure the measurement errors. (a) x position is fixed. (b) y position is fixed.

In Figure 6.2, the results obtained for the test are presented, with the exact location of the laser scanner detections represend in a map, where horizontal axes represents x coordinates in the laser scanner field of view and vertical axes represents the distance to the pedestrian in y coordinate, all in meters. Different colors try to separate the different tests peformed at different distances. c and d are the absolute errors of all the masurements represented also in meters. The assumption of two independent errors for both measurements (x,y) was correct, since the error pattern is similar among the lateral movements and vertical movements, no matter the values in the other axis. Only lateral movements present higher error when a pedestrian is in the center that was negligible for the present application.

Besides the previous conclusions, other information could be obtained from the present test, i.e. standard deviations, which were useful for subsequent tests and for tuning the algorithms. Table 7 depicts the results obtained for the standard deviation.

σ_x[m]	**σ_y[m]**
0.441	0.29

Table 7. Standard deviation measured for the laser scanner detections in meters.

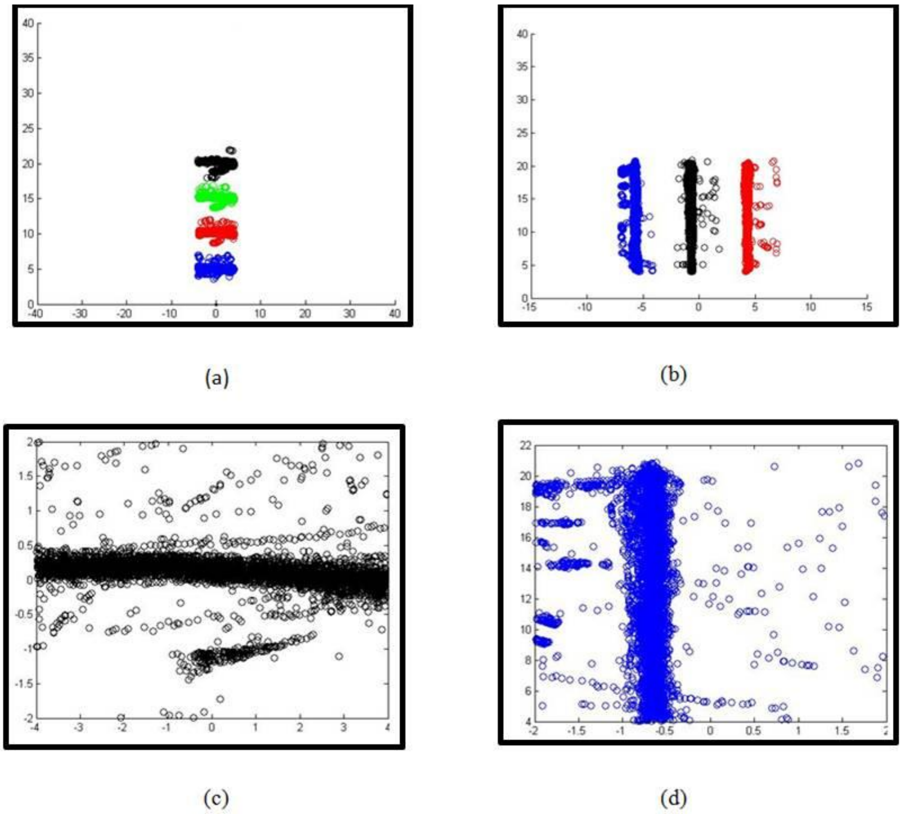


Figure 6.2 . Results of the experiments to measure the measurement of errors y (in meters) vs x (in meters). (a) Total measurements of the experiment for lateral movement. (b) Total measurement of the experiments for vertical movement. (c) y error vs x coordinates in lateral experiments. (d) x error vs y coordinate in vertical movement experiments.

For the case of the vehicles a test was performed to check the viability of the laser scanner to detect vehicles. This test was performed in a close circuit, with vehicles performing different movements that Figure 6.3 depicts.

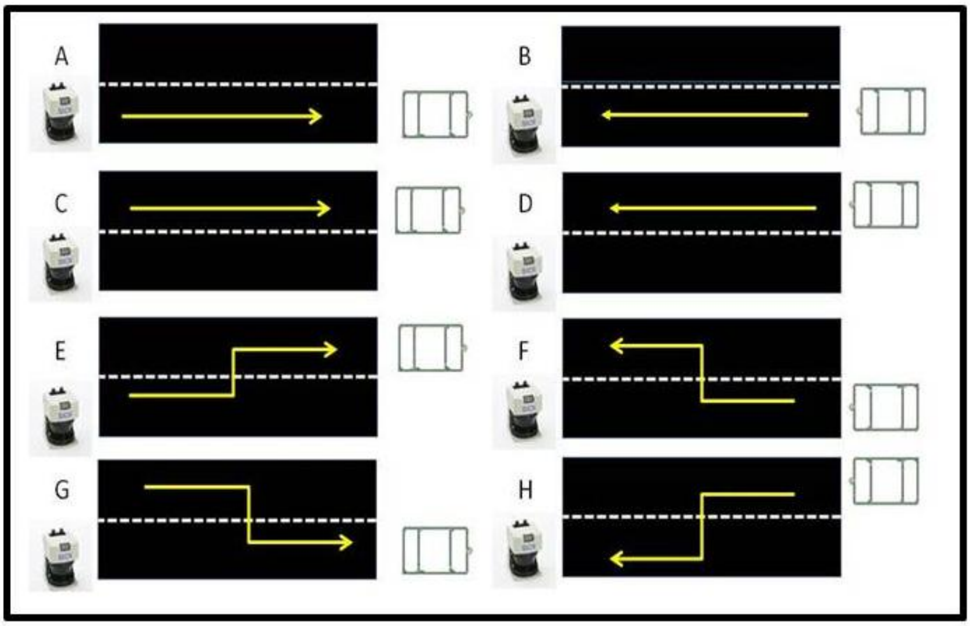


Figure 6.3. Test performed to check the viability of the laser scanner system.

To perform the test to check the viability of the laser scanner as a vehicle detection system, two vehicles were used (Figure 6.4), a black one and a metal grey one. These two vehicles represent an easy scenario (metal grey has a high reflexivity) and worst case scenario (black). This way the laser performances could be checked in the two most representative scenarios.

Figure 6.4. Vehicles used for the test. Green vehicle had mounted the laser scanner and metal grey and back vehicle were used to measure the reflexion of the laser scanner.

Results depicted give an idea of the error that the laser scanner commits when measuring the width of vehicle, in meters (Figure 6.5). The horizontal axis represents the distance in meters to the laser scanner and the vertical edge the error in absolute values to the real width. This way, the performance of the laser scanner in the task of representing silhouette of the vehicle can be checked. Furthermore, it was proved how even in worst case situations the laser scanner can measure a vehicle approaching with an expected error of 0.5 meters at a distance of 60 meters. It was also proved that the detection was better in moving away movements where the back of the vehicles resulted better for reflexion than the front of the vehicle.

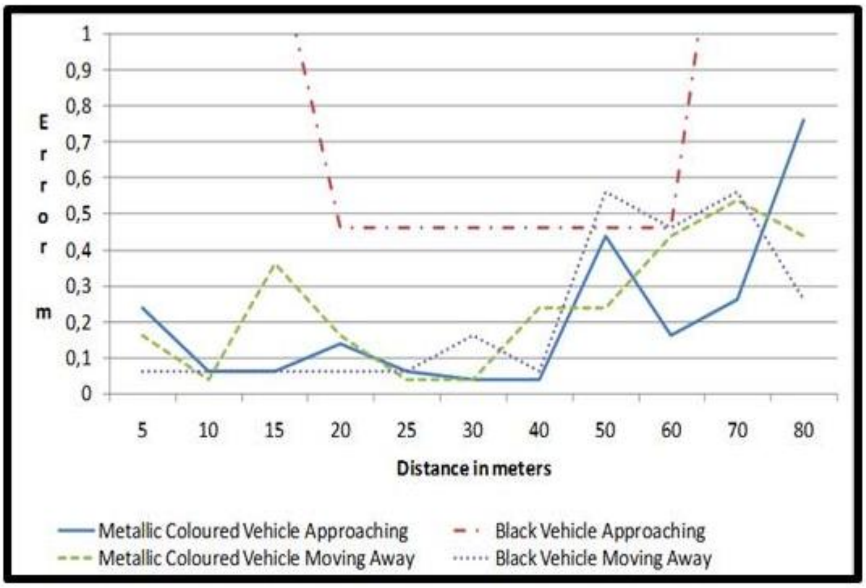


Figure 6.5. Error of the measurement of the width of the vehicle, obtained by laser scanner according to the distance.

A detailed explanation of the test, with all the particulars and more information can be obtained in [120] where all the test performed and the results provided are reported.

All presented tests were used to check the viability of the system as well as to collect some important information, such as standard deviation, which is mandatory in the estimation process. Also some of these tests were used to measure the performances of the laser scanner contributions, by measuring the positive detections in each scenario. The results of these tests, related with the performance of the different algorithms, are detailed in the next section.

In the following sections, the performances of the different algorithms for each sensor independent and for the fusion system are studied. Different tests used to check the different algorithms are depicted, providing comparison for each system independently and for the complete system. To provide complete comparison of each algorithm a division according to the obstacle to find should be performed. The next sections provide results for pedestrians and vehicles respectively.

6.2 Pedestrians

For the pedestrian up to 52 different sequences with more than 10.000 frames were tested. Three sets of sequences were created:

- **Test sequences** were created with single pedestrian performing lateral and vertical movements in an easy environment. The environment, a parking lot, can be described as a sensor friendly scenario, since the number of other obstacles that could represent possible false positives (lampposts, trees…), whether for vision and for laser scanner systems, is limited. These tests helped to setup and check each system independently, and the fusion process itself. They were also very useful for testing the tracking algorithms, since the absence of other obstacles and the limitation of the number of pedestrians helped to avoid misleads, and therefore, check the degree of accuracy of each tracking procedure independently. Figure 6.6 (A).

- **Inter-urban scenarios** were the number of pedestrians is limited, as well as other obstacles that could lead to misdetections. Although they represented real environment tests, the number of misdetections expected is limited. Thus, the level of stress of the test is also limited. Figure 6.6 (B).

- **Urban scenarios** the scenarios represent the worst case scenarios, where the vehicle performs numerous changes in directions as well as the number of obstacles is high, resulting in more misdetections expected by the sensors. Figure 6.6 (C) and (D).

6.2.1 Low Level Algorithms Performance for Pedestrian Detection

During the above detailed tests, more than 4,000 possible pedestrian detections were tested, divided in the different sets of tests, in different conditions. In the tests, any pedestrian in a given frame is considered a positive detection if it is detected, or no detection in case of negative result. On the other hand, if a positive is given for a non-pedestrian obstacle, it is considered a misdetection; the percentage of misdetection is the number of misdetection per 100 frames:

Figure 6.6. Example frames of the different tests performed for pedestrian detection. Test sequences were performed in a parking lot (A), Inter-urban scenarios (B) and urban scenarios (C) and (D).

	% of positive detection	% of misdetections (per frame)
Test	78.01	5.19
Interurban	73.19	3.91
Urban	67.72	6.72
Total:	72.97	5.27

Table 8. Results for the computer vision-based pedestrian detection.

	% of positive detection	% of misdetections (per frame)
Test	79.71	6.23
Interurban	70.35	16.96
Urban	73.61	16.72
Total:	74.56	13.3

Table 9. Laser Scanner pedestrian detection performance.

Given the above results some conclusion can be obtained from the low level approaches:

- Laser Scanner is a novel approach giving very good performance, although the results obtained lack reliability since a percentage of 13.3 % of false positives means that less than every ten frames a false positive is provided. Consequently, false positives are frequent. Here it has to be remarked that these false positives are limited to the space in the image, thus even more false positives appear in the higher field of view of the laser scanner. However, since they are not relevant for the algorithm proposed, they were not taken into account.

- Performance of the computer vision system is high and reliable, mainly in more structured scenarios, but in more complex scenarios such as urban environment this reliability decreases highly due to the abundant information present in this environment that can lead to misdetection.

- It has to be remarked that the setup of the system was performed taking into account these situations, i.e. The threshold for positive detection of the computer vision was increased to add reliability to the detections of the laser scanner during the fusion process.

- False positives provided by the camera were mainly unrelated to the HOG features algorithm presented in chapter 4, but related to the nature of the entire algorithm that was based in the ROIS selected by the laser scanner. As explained before, The ROIs are chosen according to the lectures form the laser; thus, if a spurious observation is returned by the laser, and a pedestrian is included in more than one ROI, all of them are going to be considered pedestrian, and as a result a false positive is returned. Figure 6.7 gives an example of this situation. Although they could be considered positive detection for the computer vision algorithm itself, the approach provided here is based on both sensors, and therefore, they were included in false positive detection.

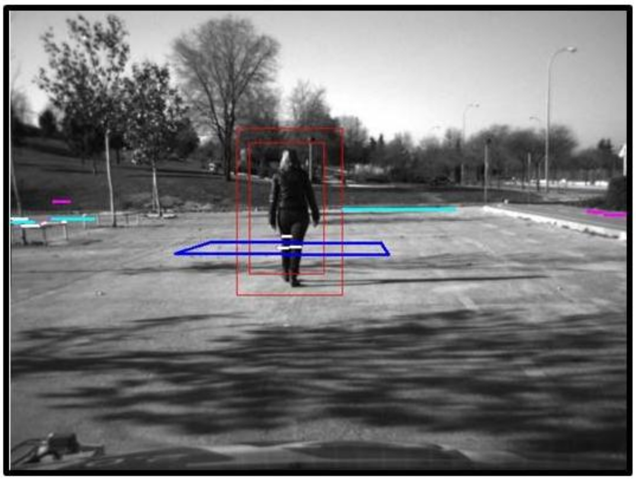


Figure 6.7. Example of misdetections of the camera due to multiple ROIs pointing to a single pedestrian. Two laser scanners point to the pedestrian, thus both of them return positive detections. In these situations it could be considered a positive detection by the basic HOG feature approach, although they were considered misdetections of the vision system in the present approach that uses laser scanner ROI detection.

- It is also clear that the performance of the laser scanner detection is better in easier environments, where the laser scanner suffers fewer problems with the errors due to pitching calibration errors, caused by strong movements. As explained in chapter 4, the beam of the laser has to be accurately measured, and in real road conditions details such as differences in the weight of the vehicle, strong braking maneuvers, etc. provoke that the desired part of the pedestrian to be detected (legs) are not easy to find due to the calibrations errors. Despite this problem, the ratio of positive detections remains in a 70% in the worst situations.

Some examples of typical pedestrian detections and misdetection by each subsystem are provided in subsequent figures. Complete examples of full sequences are given in further sections of the present chapter.

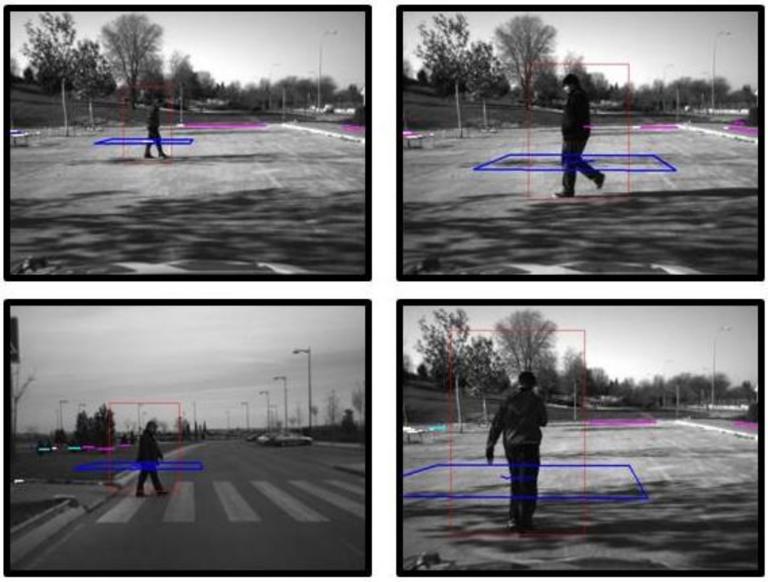


Figure 6.8. Positive detections examples. Blue boxes represent laser scanner positive detections. Red boxes the image positives. Also laser scanner polyline reconstructions are showed in the image.

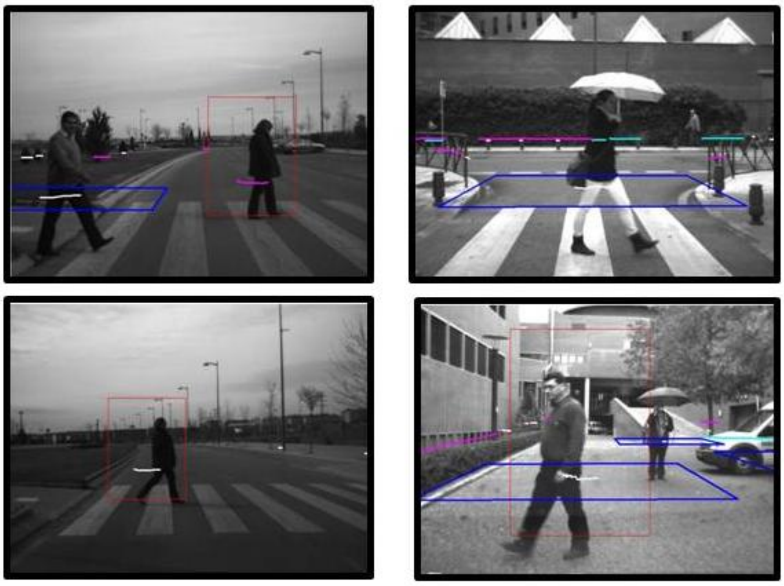


Figure 6.9. Example of errors in the tests. Blue boxes represent laser scanner positive detections. Red boxes the image positives. Also laser scanner polyline reconstructions are showed in the image.

6.2.2 Tracking Algorithms Performance

Tracking algorithms tested are explained in chapter 5, three estimation algorithms were used in the test: Kalman Filter, Unscentered Kalman Filter and Particle Filter. The three of them were tested using the Global Nearest Neighbor approach presented also in chapter 5 and using the test sequences with the movements depicted in Figure 6.1, that represent classic movements of pedestrians. In this way, the model and the performance of the different estimations were tested and the most suitable selected for the subsequent test. Results are depicted in Table 10.

Due to this test it was proved that the visual algorithm performed detections up to 15 meters, farther distances were not covered due to the lack of information. This problem was inherent to the device used and not to the algorithm. A higher performance camera able to provide better quality images should provide longer distance detections.

To measure the accuracy of the tracking algorithm Global Nearest Neighbor (GNN) fusion algorithm presented in chapter 5 was used. The test consisted of situations where consolidated tracks find new observations, and the differences with the predictions of the estimation filters are calculated. The results are depicted in the Table 10. The reason of selecting the GNN algorithm is the simplicity of the algorithm that does not require changes in the classic estimation filter Table 10 depicts the standard deviations of the predictions.

σ KF	σ UKF	σ PF
0.2058 [m]	0.1591 [m]	0.1486 [m]

Table 10. Tracking algorithms' standard deviations.

During the test also the amount of positive detections were checked to see if the selection of different approaches reached different results. The test showed that the results provided by the different algorithms were similar. Thus, the selection of any of the estimation methods did not change the final results related to the ratio of positive detections. Only in the case of the particle filter, with not enough particles the system did not reach to a stable prediction leading to a loss of the performance of the overall system. Finally, it was possible to tune the particle filter with up to 200 particles reaching to same positive results and the performance showed in Table 10.

According to Table 10, the results for the KF are very accurate, therefore, we can consider that the movement model presented in equations (68) to (73) a valid model for pedestrian movement. Since the mean detection error is around 20 centimeters.

It also was proved that the UKF presented a mean improvement of approximately 4 centimeters as well as PF. The results showed that the UKF was useful to overcome the limitations of the lineal model, by correcting the nonlinearities of the movement or the detections inexactitudes. However, it should be remarked that the PF efficiency result was limited because the amount of particles needed to reach to this value results were too big to allow real time performances. UKF on the other hand, proved to be an important improvement given the results to overcome the limitations of the lineal model.

Thus the overall conclusion of the tracking algorithm tested can be summarized in the following points.

- The model presented assumed in chapter 5 resulted a valid model for the present approach since Kalman Filter presented accurate results.
- UKF can help to overcome the nonlinearities of the movement of the pedestrians even though the variations are not significant.
- The PF approach resulted more accurately, but practically, it´s not feasible since the nonlinearities presented were not significant (thus the improvement of the accuracy is limited), and the amount of particles needed to reach to this accuracy was not optimal for a real time application.

- UKF for GNN can be considered the best solution for this problem, since the PF performances resulted to be limited due to the amount of resources consumed with a very limited improvement. However, the linearity of the model used and the simplicity of the Kalman Filter approach make it a suitable solution, easy to apply, with good performances and easy to adapt to other approaches, such as JPDA or MHT, which were tested subsequently using this KF approach.

6.2.3 **Association Methods Performance**

Chapter 5 depicts three association methods that could be used in the present approach. GNN, MHT and JPDA. As it was mentioned before, Kalman Filter is the most suitable solution for MHT and JPDA, and the previous section proves that KF solution for pedestrian tracking is suitable to provide accurate estimation. Therefore, the implementation of both MHT and JPDA was performed using KF approach.

The simplicity of the GNN algorithms makes it one of the most widely used algorithms for a great variety of applications, and therefore, it is considered the base of the present approach. It was the first algorithm tested and used to check both the viability of the different subsystem (laser scanner and vision based detection system) independently and the different tracking possibilities performance. Thus, for the present test it is considered the base algorithm, so it was compared with the MHT and JPDA to check the viability of these new algorithms and the degree of improvement that they supposed.

Test sequences in easy environments, used in previous tests, did not provide an important challenge for any of the algorithms, since the performances were similar. In these tests the absence of false positives and difficult situations such as crosses, changes in the directions, etc., limited the stress degree that the algorithms could suffer. As a result, the core of the test performed to check the viability of the different algorithms was based in the inter-urban and urban scenarios.

a) MHT vs. GNN tests

As mentioned before, test sequences did not create major challenge for any of the two subsystems, providing similar performances. Figure 6.10 shows a pedestrian moving laterally. Both systems give similar results, since in the sequence there are no false positives, and thus there is a single track from the beginning to the end of the sequence. In the figure green circles represent the estimations with no detection (no matches) and black circles represent that the track was found and actualized (the circle is placed in the exact coordinates where the pedestrian is detected). Vertical and horizontal axes correspond to y and x coordinates respectively from the laser scanner in meters. Only consolidated tracks are shown.

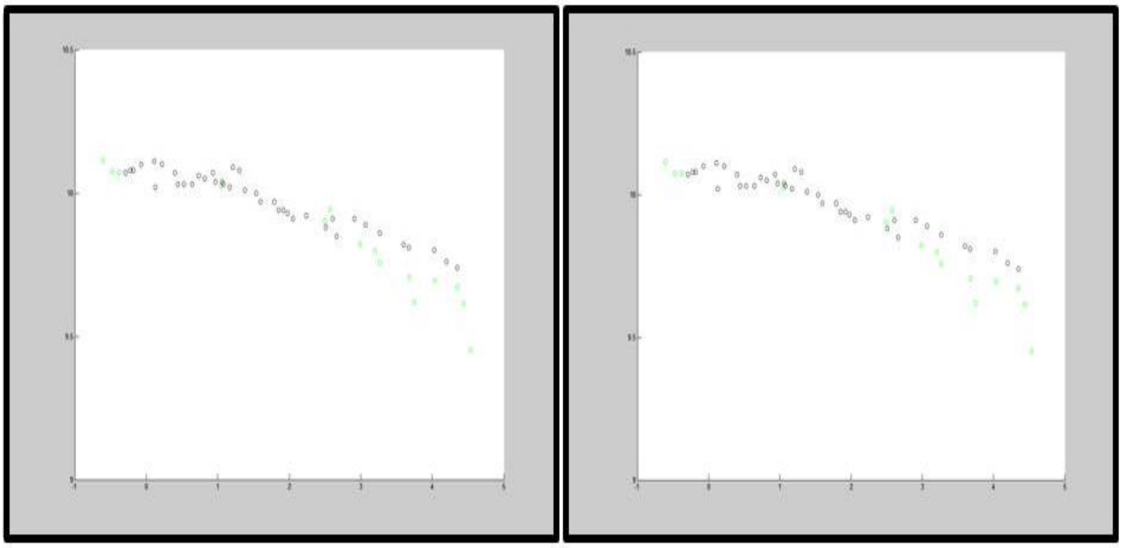


Figure 6.10. Basic test sequence with a pedestrian performing lateral movement. Left is the GNN approach, right MHT.

Problems with MHT arise when a false positive appears when dealing with more complex situations. When false positives appear new tracks are created, and although the false positives are not propagated along time, due to the pruning rules explained in chapter 5, the new tracks created grow exponentially. Consequently, the amount of tracks to track grows exponentially, although they are not considered consolidated. This way the delay that the system needs to give solution grows reaching to several seconds per frame. Figure 6.11 shows a sequence with two pedestrians in an inter-urban scenario walking through a pedestrian crossing. The colors and axis are equal to the depicted in Figure 6.10. Here, in Figure 6.11, we can see how the false positives are propagated through time in the right image (MHT).

In conclusion, through the different test performed, some of which were presented in Figure 6.10 and Figure 6.11, the rules followed to prune and reduce the amount of tracks to follow were not enough to limit this amount. The algorithms, although theoretically would represent the best solution along time, are not suitably in this environment due to the real time requirements. As we can see in Figure 6.11, the track of all possible hypotheses leads to misdetections and the maintenance of false positives for longer time. These results lead to the conclusion that MHT is not a suitable method for the present application, or in case of future implementations. Special care should be taken and modifications made to the pruning algorithms should be carried out.

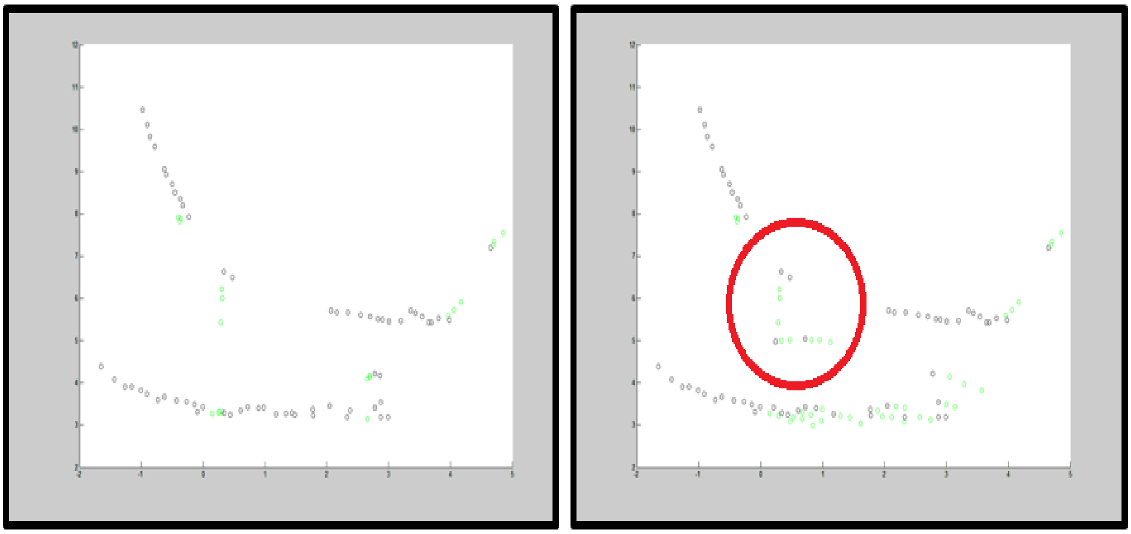


Figure 6.11 Inter-urban scenario with two pedestrians moving from left to right crossing the road. Left GNN approach; right MHT. The circle highlights part of the sequences where the false positives are propagated.

b) JPDA vs. GNN tests

JPDA is a method that requires higher computational costs due to the necessity of implementing the calculation of the joint probabilities. As explained in chapter 5, it has to be performed several times per frame, which can lead to high computations costs, mainly in cloudy environments. Therefore, the prior expectation was a more time consuming approach. But the sophisticated calculation of the association, thanks to the joint probabilities, lead to the expectation of better performances mainly in special situations such as crossing trajectories.

To perform this comparison, several tests were performed, mainly in inter-urban and urban environments, since the test sequences, as with MHT, provided similar performances in all the sequences. Besides, the simplicity of these test sequences with no more than two pedestrians per frame was not helpful to check the performance of the algorithm in environments where the number of tracks could lead to delays. But it was proved that the numbers of false positives were lower in these tests due to the fact that false positives produced by the camera and depicted in Figure 6.7 were avoided with the track management, which will be detailed later.

Three sequences are going to be detailed to give a brief comparison of the performances of the algorithms in the most difficult situations.

- First a pedestrian crossing in inter-urban scenarios is used to provide a real comparison of the performances of the system in one of the most classical situations, two pedestrians crossing the road, with trajectories crossing in the field of view.

- Second an urban complex situation is provided, where up to three pedestrians performing the highest variety of movements are presented. In this sequence the number of false positives expected is higher due the amount of pedestrians involved and the urban scenario.

- Finally the system is checked in a scenario with two pedestrian walking very close (a child and an adult). This situation is especially difficult for the laser scanner because both pedestrians are walking very close to each other, thus in certain moments is difficult to separate them. Although the main purpose of the application was the detection of single pedestrian, this sequence was added to check the performance of the two algorithms in the most difficult situation for the sensors.

Test 1

Figure 6.12 depicts the low-level detections of both subsystems and also depicts the movement of the pedestrians in the pedestrian crossing. The blue boxes represent the laser scanner and the red boxes the visual detection. Figure 6.13 depicts the detection in the environment by the fusion system using a GNN approach left (left) and the JPDA (right). The complete sequence is presented in A and B, and detailed comparison is given in C and D for the farthest pedestrian and in E and F for the closest.

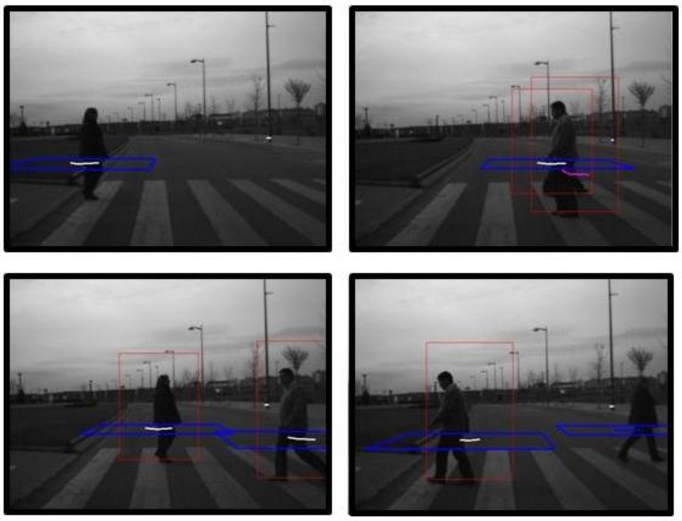


Figure 6.12. Frames of the sequence for testing GNN vs JDPA with two pedestrians crossing the road with trajectories that cross

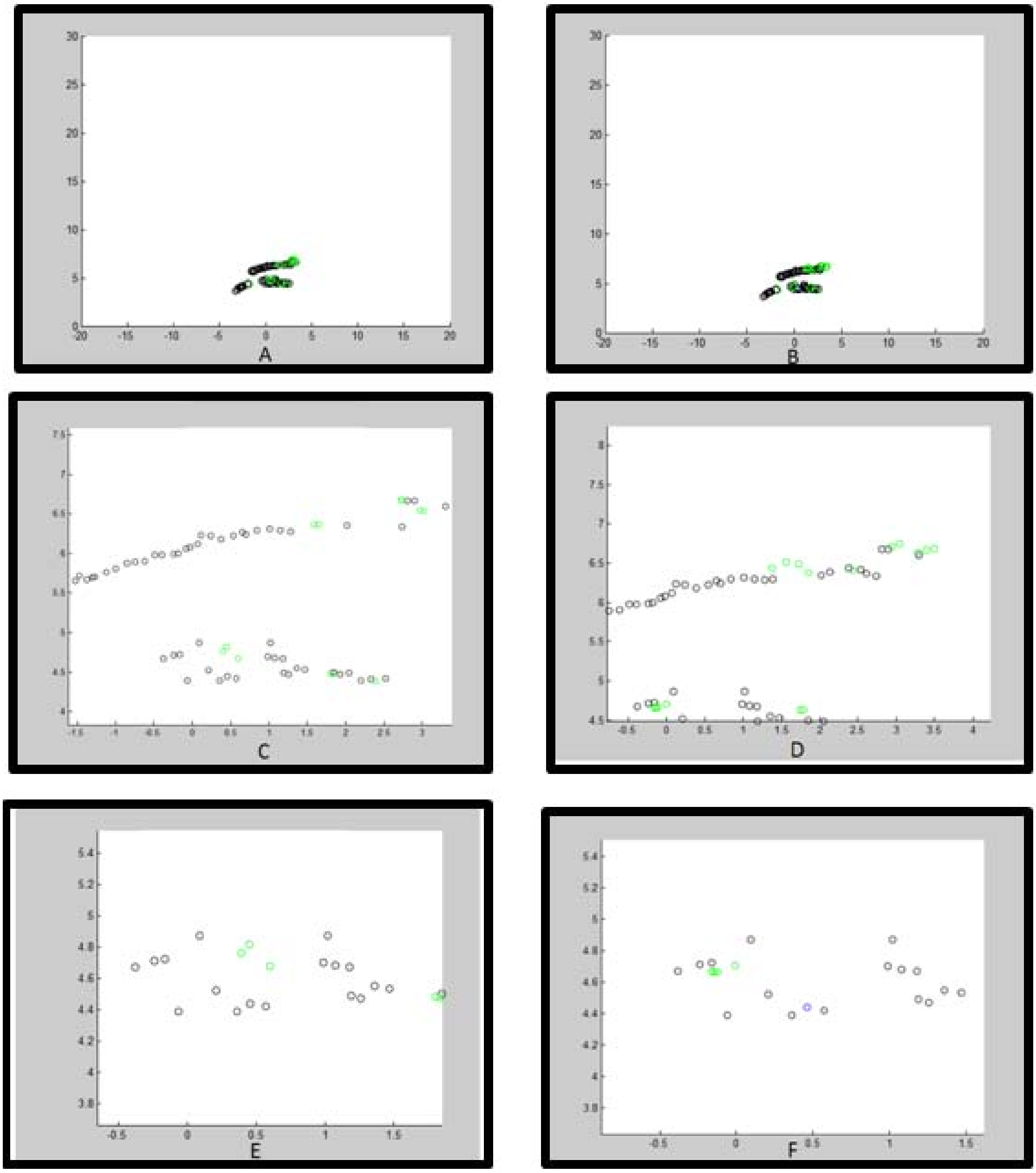


Figure 6.13. Results of the sequence with two pedestrians crossing the road. A is the overall result for GNN and B for the JPDA. C and D give details of the moment where both pedestrians cross. E and F finally give details of the closest pedestrian in the moment where both pedestrian cross. At right results for JPDA and left for GNN.

Detailed study of the sequence depicts that the JPDA approach (Figure 6.13 right) gives a better performance in the case of crossing trajectories, mainly because the estimation of the movement of the pedestrian presents a better behavior in the absence of detection, which leads to tracking the farthest pedestrian better. Although in this case, both algorithms are able to track both pedestrians with certain accuracy giving similar results.

Test 2

In this sequence, up to three different pedestrians were presented. Again the trajectories of both cross at a certain point, and at this time a third static pedestrian is presented. Figure 6.14 depicts the movements of the pedestrians in the sequence and Figure 6.15 the results of the tests.

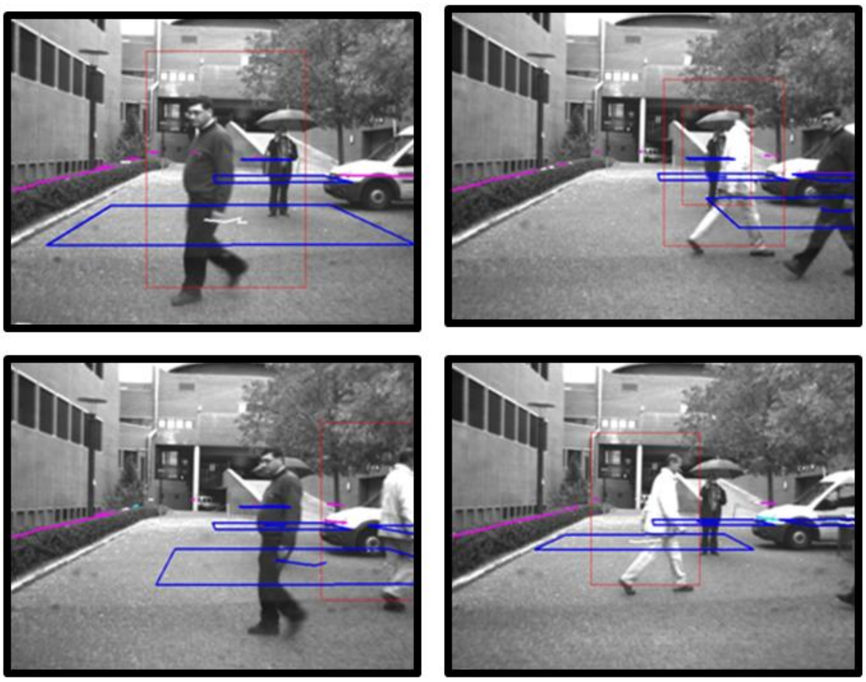


Figure 6.14. Frames of the test 2 for testing GNN vs JDPA in urban environments with three pedestrians involved.

Figure 6.15 (below) highlights the two main points of the sequence. First is the false positive that appears in (a), at this moment a false positive appears in GNN approach while in the JPDA the algorithms is able to overcome the problem. In (b), although both systems have the same false positive, the JPDA stays where the obstacle is, while GNN estimation is worst in comparison to the GNN. Therefore, according to the results given in this test, behavior of the test in the crosses, although similar, resulted in better results in the case of JPDA algorithm.

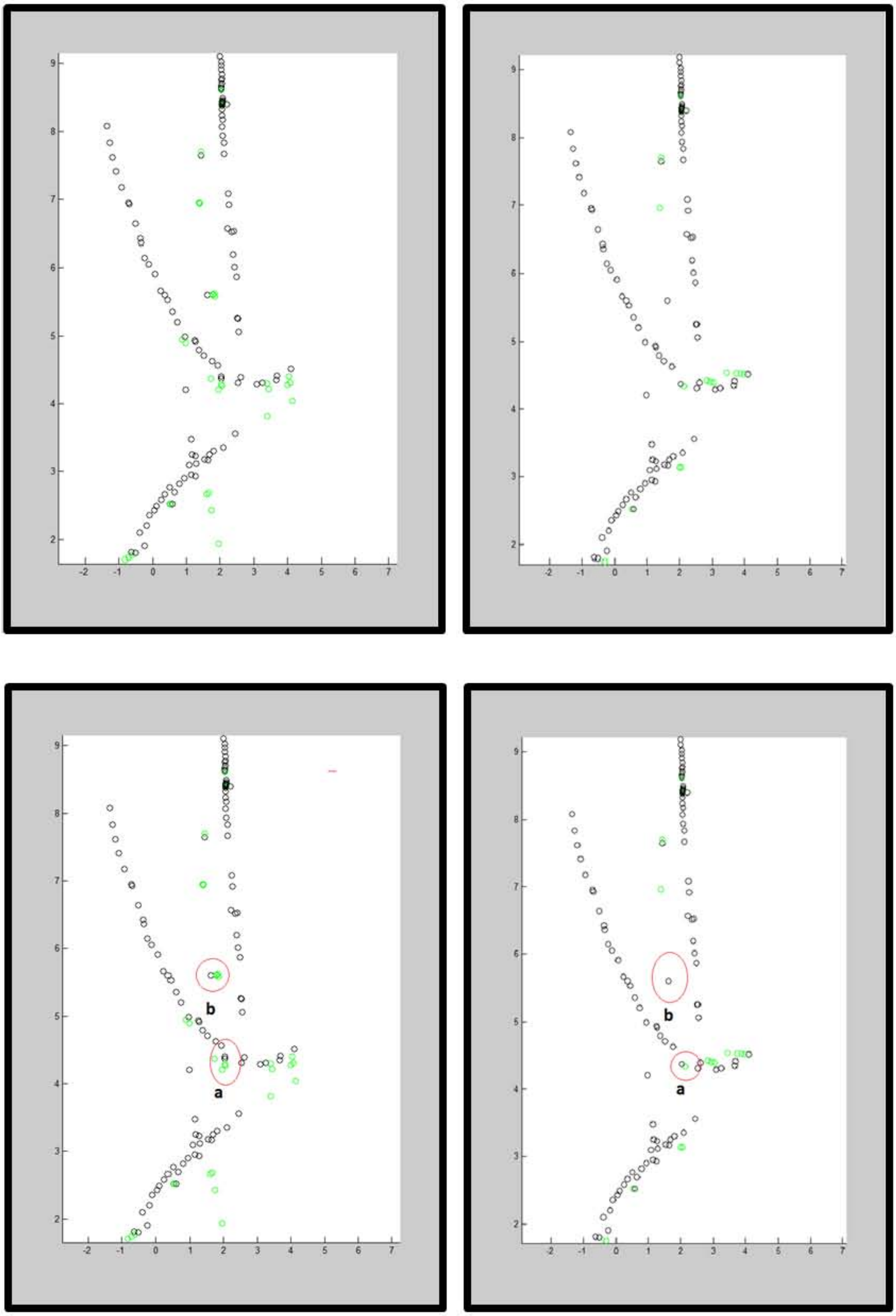


Figure 6.15. Results of the sequence three pedestrians in interurban scenario. At right results for JPDA and left for GNN. First, the results sequences are depicted (up). The same results are shown down with the main differences highlighted in (a) and (b).

Test 3

In this last test, as previously explained, two pedestrians performed vertical movements. These two pedestrians were close, as shown in Figure 6.16, which makes it difficult for the laser to differentiate among them. In this situation, the JPDA is especially useful since it takes into account all the variables plausible to perform the actualization, thus whether the system detects one or two pedestrian, they are always present in the actualization process.

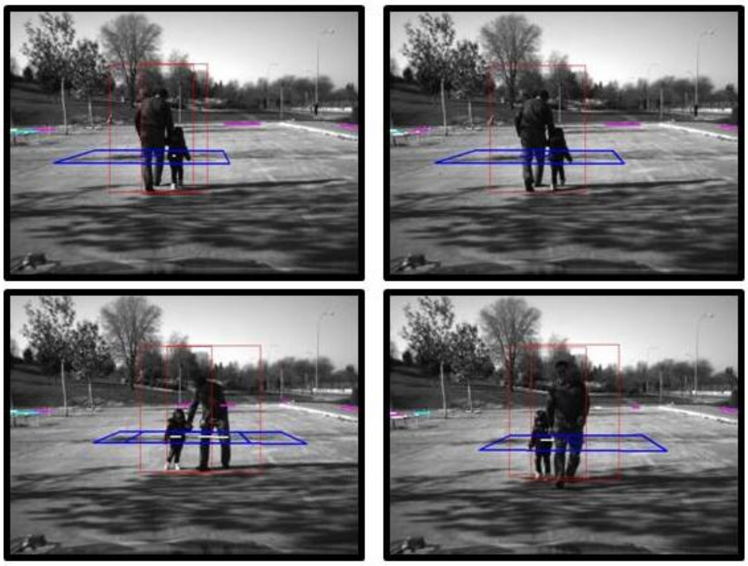


Figure 6.16. Frames of the sequence that was used to compare the behavior of the GNN and the JDPA with two close pedestrians. As shown in the figure, it is difficult for the laser scanner to differentiate among them.

Results depicted in Figure 6.17 (up) show that the performances of the system is similar, although the JDPA approach (right) provide a better solution in special circumstances, as is highlighted in Figure 6.17 (below). The fact that the JPDA takes into account any possible measurement within the gate of the track makes it very suitable for these situations, where even in the case of a single track with several measurements it uses both measurements to update the filter, with the likely function as a weight.

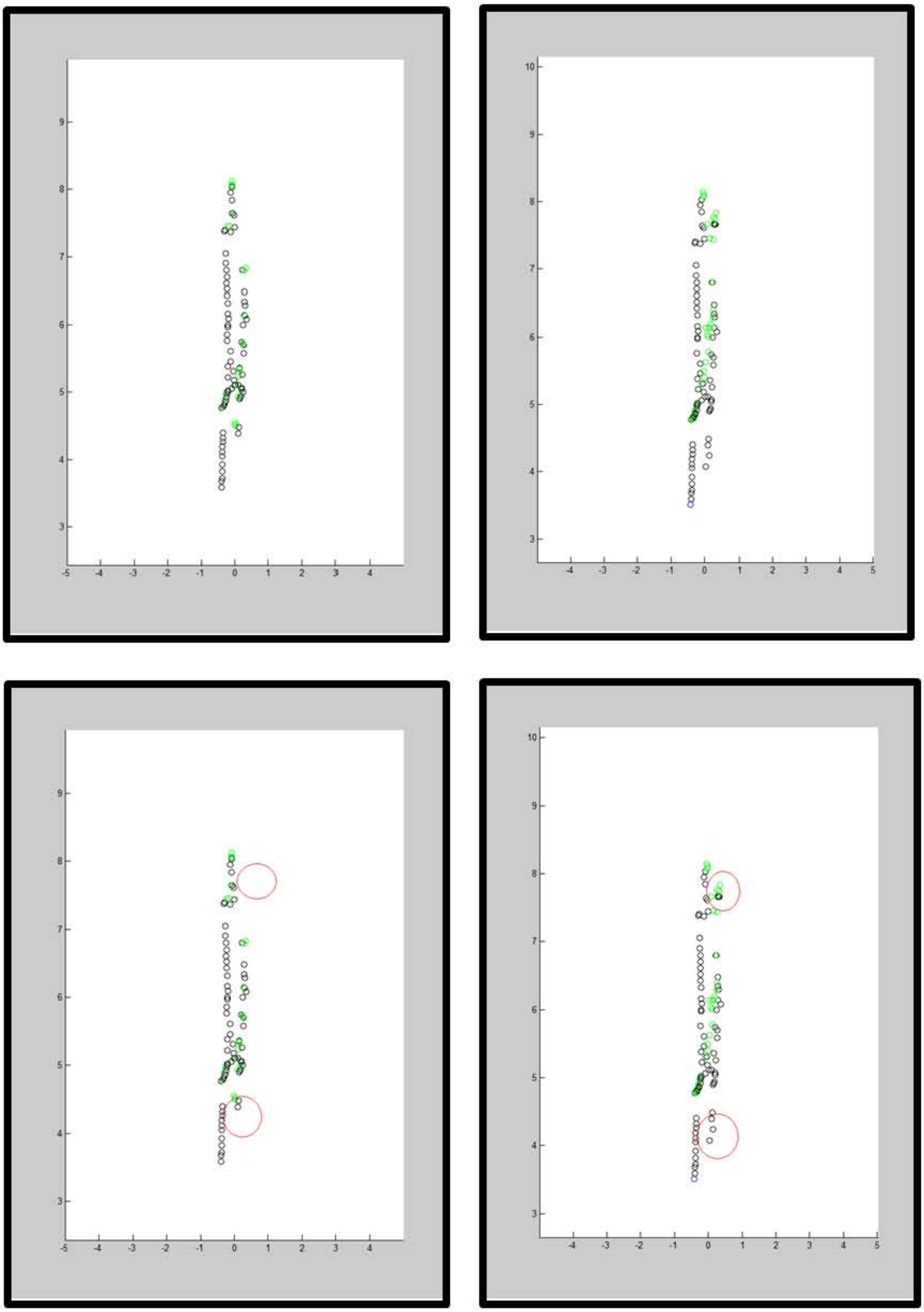


Figure 6.17. Results of the sequence with two pedestrians walking close to each other. At right results for JPDA and left for GNN. First the results sequences are depicted (above). The same results are shown down with the main differences highlighted with a red circle.

The entire test demonstrated that the JDPA approach represented a more accurate approach; the improvements offered to the previously presented GNN approach are summarized in the following points:

- Better estimation, as can be seen in the different tests presented before, the JPDA methods allow tracking the pedestrian, even when the track has been lost for several frames thanks to the improved estimation presented before.

- Improvements against misdetections. Figure 6.7 depicts the typical error of misdetections when dealing with the camera, these errors can be overcome in the JDPA approach since, as it was explained in chapter 5, the new tracks are created only if an observation is outside any gate of the tracks. This way when two pedestrian are detected very close and only one of them corresponds to the same pedestrian, no new track is created and the new track is used according to the respective joint probability in the Kalman Filter. This behavior is very useful in these kinds of approaches given the nature of the laser scanner, as shown in Figure 6.17. It also can be used in future approaches that should try to detect groups of pedestrians where pedestrians can split or merge into groups, in these situations the JDPA association method seems to be perfect since it implicitly uses these detections to actualize the Kalman Filter.

It is also remarkable that all the above improvements were obtained without an increment of the processing time; although the system should calculate the joint probabilities, the system was able to work with the same speed than the GNN approach.

6.2.4 Fusion System Performance

Results obtained in different scenarios, as well as the whole set of tests, are depicted in Table 11. Previously presented performances of the different subsystems are also included independently to allow contrasting of the performance of the whole system and each system independently. The test regarding the fusion system was used applying the KF and the JPDA approaches.

As depicted in Table 11, the results obtained increased and demonstrated the viability of the fusion and how it improves the overall performance of the system. The improvements of the system are summarized in the following points:

- The rate of positive detection is increased in the entire test. This increment is even better in the worst case scenarios for the laser scanner such as urban or inter-urban environments.

- It is also remarkable the improvements in the false positive rate. The results reach to 1%, providing the proof that the system fulfills the main requirements of a safety application: reliability. It is in this point where the improvement of the fusion process is most remarkable.

	Camera		Laser Scanner		**Fusion**	
	% of positive detections	% miss (per frame)	% of positive detections	% miss (per frame)	**% positive detections**	**% miss (per frame)**
Test	78.01	5.19	79.71	6.23	**82.42**	**0.89**
Interurban	73.19	3.91	70.35	16.96	**78.90**	**6.53**
Urban	67.72	6.72	73.61	16.72	**81.76**	**1.95**
Total	72.97	5.27	74.56	13.3	**82.29**	**1.11**

Table 11. Results of the sensors subsystems and overall fusion subsystem, divided in the sets of tests used and overall performance.

6.3 Vehicles

To check the viability of each algorithm independently as well as the fusion system, different tests were performed in real road conditions. These tests included a single vehicle performing all kinds of maneuvers, such as overtaking maneuvers, being overtaken, driving in front of the test vehicle (IVVI 2.0) and performing turnings (including roundabouts). All those tests were mainly performed in inter-urban scenarios, although some of the tests were performed in urban scenarios.

6.3.1 Low Level Algorithms Performance for Vehicle Detection

A first test of the laser scanner based vehicle detection algorithm performance was performed, results are depicted in [84]. Results showed that in the best conditions (no vehicle movement) and with direct vision, the algorithm can reach to a 100% of detection up to 35 meters in approaching movements and 62 meters in moving away movements. As it was explained in the paper, these lower results in approaching movements were due to the lower reflectivity of front parts of vehicles.

The results of the subsequent test performed in different environments for real road conditions are depicted in Table 12 and Table 13 for camera and laser respectively.

Camera

	% of positive detection	**% of misdetections(per frame)**
Total:	47.72	1.13

Table 12. Visual based vehicle detection performance.

Laser Scanner

	% of positive detection	**% of misdetections(per frame)**
Total:	91.03	8.19

Table 13. Laser scanner based vehicle detection performance.

Some details should be highlighted before analyzing the above results:

Laser scanner detection presents a high amount of false positives. Here, only those that are in the camera field of view were taken into account. These errors were frequent in movements involving lateral movements, strong braking or acceleration movements. The inertial device resulted insufficient to avoid these errors. As a result it is in these situations where the fusion process has special importance. Thus, the high positive rate of the laser scanner is important. However, special attention should be taken to the amount of false positives: in one of every ten frames a false positive is returned. Consequently, it is clear that fusion approach is necessary to overcome these problems related to this technology.

Vision approach, on the other hand, has a small positive detection rate (aprox. 50%). It is mainly because the training stage was performed keeping in mind further fusion stages. It was rather interesting having a small positive rate which provides a small false positive rate. It also has to be remarked that this system lacks of tracking that could help to provide better performances.

The results showed that if the camera is able to give a positive detection of a vehicle every 2 or 3 frames, with enough reliability due to the low false positives, it should help to create a trustable and reliable fusion system.

After analyzing the results, the conclusions obtained from these tests are summarized in next points:

- Laser scanner detection algorithms provide a high amount of positives as well as false positives, that later should be avoided in subsequent fusion stage.
- Vision approach can be used to add trustability to these laser scanner detections due to its low false positive rate. However, the low positive rate means that this algorithm is not robust enough, and fusion is necessary to add robustness.

6.3.2 Tracking Algorithm Performance

Again three different tracking algorithms were tested using the basic GNN approach. The tests to check the viability of the system were performed with a single vehicle performing different movements (e.g. overtaking, turning in roundabouts and being overtaken). Three possible estimation filters were tested (KF, UKF and PF) and the results are provided in Table 14. As in the case of the pedestrians, the difference between the predicted position and the observation are checked and the standard deviation calculated.

σ KF	**σ UKF**	**σ PF**
0.4542 [m]	0.4448 [m]	0.4409 [m]

Table 14. Estimation filters performance comparison (Standard deviation measured).

The results obtained with the tracking algorithms were similar to the expected, and they did not differ much to those obtained with pedestrians. The result is that PF is the most accurate but with small improvements in respect to the KF approach. In this case the improvements of the different approaches, such as PF or UKF were negligible, and thus it was considered that the model used for the vehicles as well as the KF, the optimal solution for this approach, was good enough. Besides the similar accuracy, the KF solution is easier to implement and consumes less computational costs. Hence, all these advantages lead to the conclusion that KF was a good approach for the application, so subsequent algorithms such as MHT and JPDAF were implemented using this KF estimation method.

6.3.3 Association Methods Performance

Three association methods were implemented and compared for the vehicle detection approach, (GNN), (MHT) and Joint (JPDA). All of them were implemented using Kalman Filter estimation method as it was explained in chapter 5 and according to the good results provided in the previously presented tests.

As it was done in the pedestrian detection algorithm, several tests were performed to check the differences in the algorithms presented. In the present section each one of them is going to be compared with the classic GNN method. Performances with single vehicle detections, as occurred in the previous pedestrian approach, were similar, but special attention should be made to particularly difficult situations, with several cars.

a) MHT vs. GNN test

Similar to what happened in the pedestrian approach, the procedure for pedestrian detection using MHT provided similar results in the different approaches tested. Problems arise when several obstacles were found in the same gate of a track, the amount of tracks start to grow exponentially, thus the measures created to avoid this problem were insufficient. The consequence of this exponential growth of the number of tracks is an unstable algorithm that consumes a high amount of resources, making it unviable for real time performances.

Figure 6.18 depicts an example of the performances of the system where the performance of the MHT system is clearly lower. These errors are probably due to the fact that a higher amount of tracks consumes the observations faster, not allowing a single track to perform the whole tracking of a single vehicle. Therefore, besides the high computational costs, the track management policy is clearly insufficient and leads to the conclusion that the MHT solution is not suitable for these applications. For this reason, the MHT approach was rejected and the JPDA approach was developed to improve the performance of GNN system.

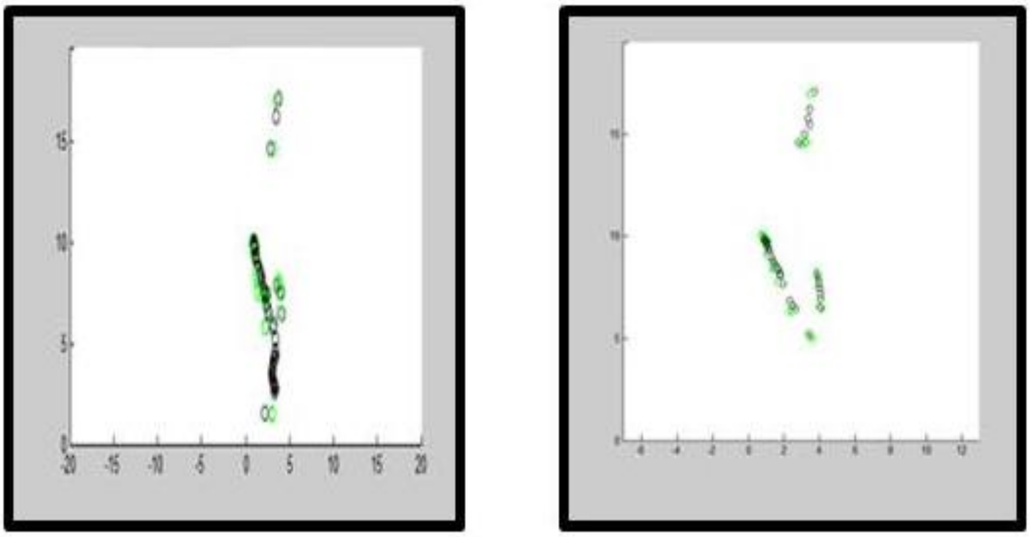


Figure 6.18. MHT (right) performance vs. GNN (left). In green the estimation with no observation matches and in black matches.

b) JPDA vs. GNN test

After the conclusions obtained by the tests with MHT showed in the previous section, the effort focused on a more powerful approach that was able to overcome the limitations of the basic GNN and adapt to the special behavior of the tools. As it happened in the case of pedestrians, and given the good results obtained there, JPDA method seemed to be a good solution for this situation.

In the case of vehicles, the limitations of the bounding boxes to big obstacles that fit into the constraints of the size of a vehicle restrict the visual errors that we found in the pedestrian approach shown in Figure 6.7. Furthermore, in the case of vehicles, the laser scanner errors, due to two very close obstacles, were not as common as with the pedestrians where they converge and separate during the process of walking. Thus, given all these assumptions, the prior expectation of the improvement of the JDPA algorithm over the classic GNN approach was limited.

Several tests were performed, providing results that clarify the advantages and disadvantages of each of the systems. To give an idea of the differences in performances of the algorithms, three different sequences are going to be detailed.

- The first consists of basic movements, where the IVVI 2.0 was moving following another vehicle, performing straight movements, and turning in a roundabout.
- The second sequence involves a more complex movement. IVVI 2.0 performs an overtaking maneuver over two cars.
- The third sequence is the most difficult one, with two vehicles entering in the field of view, one of them performing incorporation to a road, and the other moving inside this road.

Test 1

Behavior in this test, depicted in Figure 6.20, gives an interesting result. As can be observed in the figure, both sequences´ results are identical, as well as the time consumed to perform the detection. This situation was interesting to demonstrate that both algorithms can be considered equivalent when there is a single obstacle in the environment. Some false positives were given from the single scanner algorithms, mainly from laser scanner, but both algorithms were able to eliminate them.

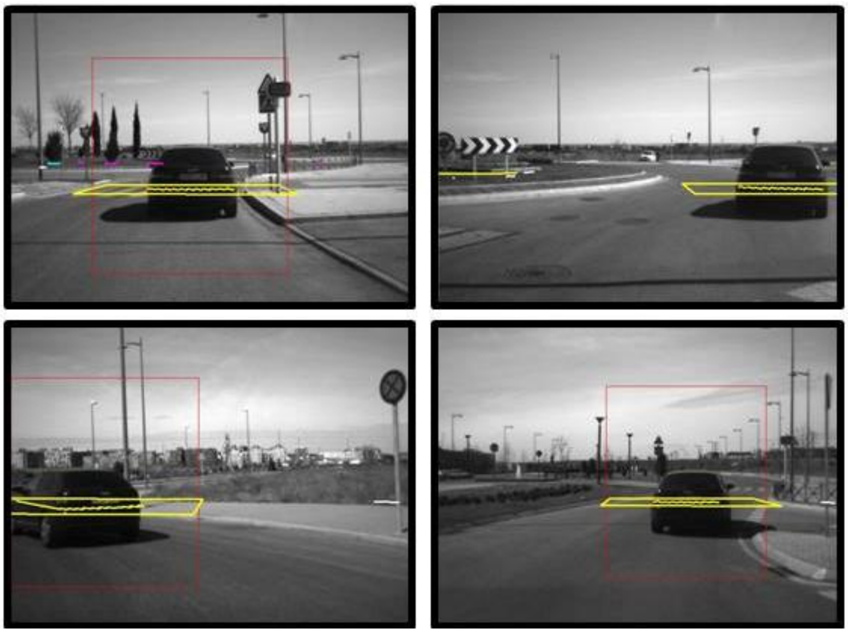


Figure 6.19. Frame examples of the test performed with the IVVI 2.0 following a vehicle. The movement involved turnings inside a roundabout.

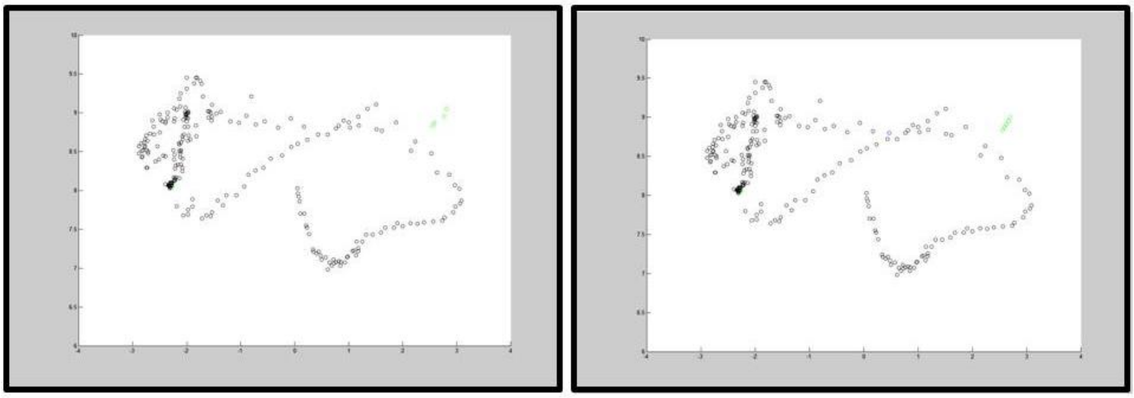


Figure 6.20. Results of the tracking of a single vehicle after several movements, including a roundabout. JDPA results are shown in right and GNN in left image. Green detections represent estimation with no match, black are matching observations.

Test 2

In this sequence, the test platform (IVVI 2.0) performs an overtaking maneuver over two vehicles (Figure 6.21) the results depicted in Figure 6.22 show similar results for both subsystems. Detection performed by the JPDA approach was able to follow the vehicle even outside the field of view of the camera (red circle). It is important to point that the delays to perform the assignations in this test were similar.

Figure 6.21. Frames of the test sequence with the vehicle (IVVI 2.0) performing an overtaking maneuver.

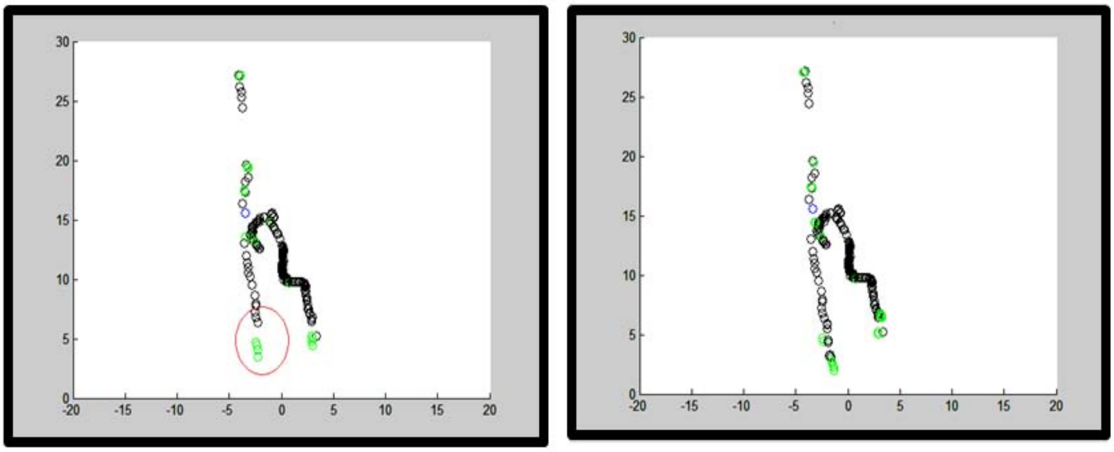


Figure 6.22. Results of the test sequence with the vehicle (IVVI 2.0) performing an overtaking maneuver. JDPA and GNN results are shown (right and left respectively). Green detections represent estimation with no match, black are matching observations.

Test 3

This sequence presents more difficulties, for having more than one vehicle in the sequence, and the fact that one of them occludes the other one. Also, it is important to notice that the IVVI 2.0 platform is performing a high turn that leads to a high number of false positives in the laser scanner algorithm, but thanks to the vision system, these false positives are discarded. However, the problem is that all those false positives introduce new tracks to perform the algorithm, which lead to certain delays in the assignment. Besides this problem, the algorithm, as it is shown in Figure 6.24, gives better results, even though the difference between both algorithms is small. The delays presented could easily be overcome with more powerful computers or a parallel programming approach.

Figure 6.23. Frames of the test sequence with two vehicles crossing.

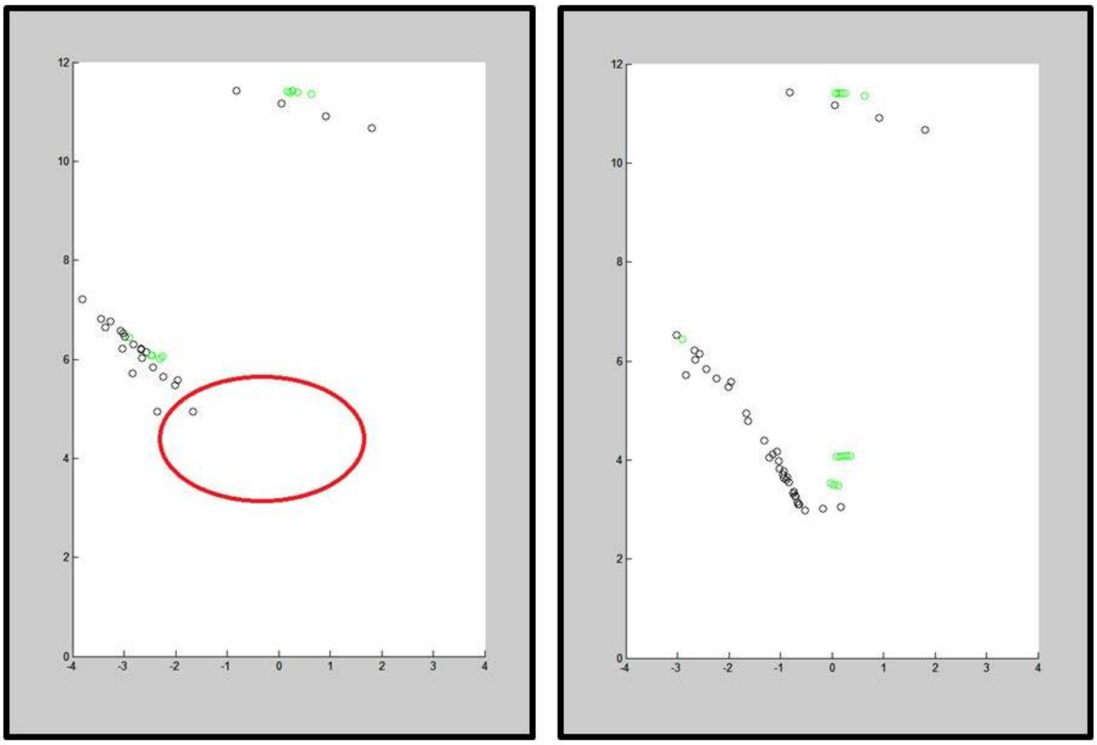


Figure 6.24. Results of the test sequence with two vehicles in a crossing. JDPA results are shown at right and GNN at left in the image. Green detections represent estimation with no match, black are matching observations.

The conclusion of the present test was that the best algorithm is the JPDA in both providing better results and also giving smoother behavior in the tracking process. According to the test, the latter is the cause that allows keeping the track, even in these situations where the GNN approach loses it. On the other hand, it is important to point out that this algorithm has the drawback of the computation costs, so special care should be taken to avoid situations with lots of tracks in case of the use of JDPA approach. It should also be remarked that the GNN approach showed good performances; hence, this algorithm should also be taken into account in case of the necessity of a less cost demanding algorithm with good performances.

6.3.4 Fusion System Performance

Results obtained by the whole set of tests are depicted in Table 15. As was done in pedestrian results, previously presented performance of the different subsystems independently are also included to allow contrasting the performances of the whole system and each system independently. The fusion algorithm consisted of a JPDA approach with KF estimation.

	Camera		Laser Scanner		**Fusion**	
	% of positive detections	% miss (per frame)	% of positive detections	% miss (per frame)	**% positive detections**	**% miss (per frame)**
Total:	47.72	1.13	91.03	8.19	**92.03**	**0.59**

Table 15. Overall results for the test of the vehicle algorithms.

Table 15 depicts the overall results of the complete system in the 28 tests performed over a total of more than 4000 frames. The following points summarize the results obtained.

- The main goal for the fusion system in this approach was to maintain the good results of the laser scanner system providing reliability to the detection by reducing the amount of false positives. As it is depicted in the table, it was possible to accomplish these requirements due to the fusion procedure.

- Given high positive rates of the laser scanner, the task of increasing the positive rate of the overall system in comparison to the laser scanner system was very difficult. Even though it proved extremely difficult, it was slightly increased.

CHAPTER 7. CONCLUSIONS

7.1 Conclusions

As t was depicted in both chapters 1 and 3, the main goal of the present thesis was to provide fusion architecture for intelligent vehicles, which were able to overcome the limitation of each sensor to enhance the basic capacities of each one separately, providing a robust and reliable safety application for road environment. The results provided in the previous chapter show that by fusing the information of the computer camera and a laser scanner, and using other information sources (i.e. context and inertial system) it was possible to accomplish the task. The systems presented also give the possibility to increase the set of sensors thanks to its scalability.

Even though there is still a long way to go to reach the perfect safety application, this thesis represents a step forward in road safety. It provides a complete tool for the safety of the main road users by implementing a complete fusion system that fulfills all the requirements of these kinds of applications. The proposed algorithms use the Data Fusion theory to enhance the capacity of two widely used sensors in the automobile environment.

7.2 Contributions

Among all the works presented in this document it is possible to find several contributions that represent steps forward in different techniques and researching topics. The next points summarize the contributions given in the present work:

- **Laser scanner pedestrian detection**. It was pointed in several occasions during the present document that although the laser scanner provides limited information, it can be used to perform obstacle classification. The present thesis gives a step forward providing a model that combined with a higher level stage can be used to perform pedestrian detection.

- **Laser scanner vehicle detection**. It was demonstrated that the special behavior of the laser can be used to detect vehicles with a high degree of certainty. But, it must be remarked that the nature of the application that requires the laser scanner to be mounted in a moving vehicle, adds complexity to this detection, requiring accurate systems that correct the movement of the vehicle where the device is mounted.

- **Complete fusion solution** for road safety application. This document gives a complete solution of fusion architecture for road safety. Including all the levels that involve any fusion solution and based in the classic JDL model.

- **Danger estimation**. The situation assessment depicted in chapter 5 represents a novel study of the danger that involves any road situation, allowing the system to estimate the degree of danger that represents any detection.

- Novel **GNN fusion** based approach is presented. It uses KF estimation filter and both computer and laser scanner technology to detect and track pedestrians and vehicles. Although it was improved by the JPDA algorithm, this approach proved to be a reliable application able to work in real time in complex environments. The approach represented a novel contribution that subsequent studies improved by the use of JPDA association method. The test proved that this approach provide better performances than the MHT approach that was presented in the thesis

- The studies and tests that were presented also proved that **MHT** approach **is not suitable** for these kinds of applications. Thus, other approaches proposed and tested are more recommendable due to their better performances for both detection performance and computational costs.

- Finally, a novel **JPDA fusion** algorithm for road safety application was presented. It uses both laser scanner and visual technology enhancing their capacities, proving to be the most powerful algorithm for the purpose of the thesis.

All the above contribution proved that the main purpose of the thesis was fulfilled, such as to enhance the capacities of basic sensors in road environment by using data fusion.

7.3 Future Works

Several limitations to the system were found during the research process that finishes with the present thesis. Each one of them opens a new opportunity to perform future researching works, complementing the present project and providing the tools necessary to create the "perfect safety application". These possible improvements or future investigations lines could be:

- **New laser scanner** (already available in the laboratory) that improves information provided by the present laser scanner. New sensors available in the market provide more than a single layer of information, incrementing the possibilities of the system presented here, mainly those related with the necessity of an accurate calibration of the system.
- **Night vision** approaches, such as far infrared cameras could complement the system for situations with low light. Future works should include both daylight and night vision.
- **Stereovision** system (already available in IVVI 2.0) can be used to complete the information given by the laser scanner and the camera with the disparity map.
- The inertial measurements lacks of reliability, mainly when the vehicle is performing turns. A more robust system that monitors the **movement of the vehicle** should be developed to avoid misdetections or errors in the reconstruction of the environment.
- **Estimation models** used for the present approach proved to be useful when performing tracking, but future works should try to anticipate the movement of both pedestrians and other vehicles. To perform such a difficult task, new, more complex and robust models should be used. These models should be able to adapt to the nonlinearities of the movement of the obstacles and predict the movement to give better estimations.
- **Context** is a new issue in Data Fusion. The present approach intended to add context as a new information source but lot of work is still to be done in this field. Future works should take advantage of the context with new information e.g. GPS and map information with relevant information such as zebra crossings and strong corners.

REFERENCES

[1] "International Traffic Safety Data and Analysis Group." [Online]. Available: www.irtad.net. [Accessed: 24-Feb-2012].

[2] European information society, "Intelligent Car Brochure," 2007.

[3] "Intelligent System Lab." [Online]. Available: http://www.uc3m.es/portal/page/portal/dpto_ing_sistemas_automatica/home/research_activities/isl. [Accessed: 01-Apr-2012].

[4] "Darpa Grand Challenge." [Online]. Available: http://archive.darpa.mil/grandchallenge/. [Accessed: 08-Jan-2012].

[5] D. L. Hall and J. Llinas, *Handbook of Multisensor Data Fusion*. CRC Press, 2001, pp. 3-18.

[6] A. N. Steinberg, C. L. Bowman, and F. E. White, "Revisions to the JDL Model," in *Proceedings of the SPIE Conference on Architectures Algorithms and Applications*, 1999.

[7] W. Elmenreich, "A Review on System Architectures for Sensor Fusion Applications," in *Software Technologies for Embedded and Ubiquitous Systems*, vol. 4761, R. Obermaisser, Y. Nah, P. Puschner, and F. Rammig, Eds. Springer Berlin / Heidelberg, 2007, pp. 547-559.

[8] M. Bedworth, "Probability moderation for multilevel information processing," *Personal Communication*, 1992.

[9] M. Markin et al., "Technology Foresight on Data Fusion and Data Processing," *Publication of The Royal Aeronautical Society*, 1997.

[10] J. R. Boyd, "A discourse on winning and losing," *Maxwell AFB Ala Air University Library Document No MU43947 unpublished collection of briefing slides*, no. 43947. 1987.

[11] T. Bass, "Intrusion detection systems and multisensor data fusion," *Communications of the ACM*, vol. 43, no. 4, pp. 99-105, 2000.

[12] M. Bedworth and J. O'Brien, "The Omnibus model: a new model of data fusion?," *Ieee Aerospace And Electronic Systems Magazine*, vol. 15, no. 4, pp. 30-36, 2000.

[13] S. C. A. Thomopoulos, "Sensor integration and data fusion," *Journal of Robotic Systems*, vol. 7, no. 3, pp. 337-372, 1990.

[14] D. L. Hall and J. Llinas, "An introduction to multisensor data fusion," *Proceedings of the IEEE*, vol. 85, no. 1, pp. 6-23, 1997.

[15] S. Blackman and R. Popoli, *Design and analysis of modern tracking systems(Book)*. 1999.

[16] S. S. Blackman, *Multiple-Target Tracking with Radar Application*, vol. 1. Artech House, Norwood, MA, 1986.

[17] M. Bertozzi et al., “Multi Stereo-based Pedestrian Detection by means of Daylight and Far Infrared Cameras,” in *Object Tracking and Classification Beyond the Visible Spectrum*, R. I. Hammoud, Ed. Springer-Verlag, 2009, pp. 371-401.

[18] C. Hilario, J. Collado, J. Armingol, and A. La Escalera, “Pedestrian Detection for Intelligent Vehicles Based on Active Contour Models and Stereo Vision,” *Computer Aided Systems Theory â€“ EUROCAST 2005*. pp. 537-542, 2005.

[19] C. Premebida, O. Ludwig, M. Silva, and U. Nunes, “A Cascade Classifier applied in Pedestrian Detection using Laser and Image-based Features,” *Transportation*, pp. 1153-1159, 2010.

[20] C. Premebida, O. Ludwig, and U. Nunes, “LIDAR and Vision-Based Pedestrian Detection System,” *Journal of Field Robotics*, vol. 26, no. Iv, pp. 696-711, 2009.

[21] C. Premebida, G. Monteiro, U. Nunes, and P. Peixoto, “A Lidar and Vision-based Approach for Pedestrian and Vehicle Detection and Tracking,” *IEEE Intelligent Transportation Systems Conference ITSC*, pp. 1044-1049, 2007.

[22] L. Spinello and R. Siegwart, “Human detection using multimodal and multidimensional features,” *2008 IEEE International Conference on Robotics and Automation*, pp. 3264-3269, 2008.

[23] N. Floudas, A. Polychronopoulos, O. Aycard, J. Burlet, and M. Ahrholdt, “High Level Sensor Data Fusion Approaches For Object Recognition In Road Environment,” *2007 IEEE Intelligent Vehicles Symposium*, pp. 136-141, 2007.

[24] J. P. Hwang, S. E. Cho, K. J. Ryu, S. Park, and E. Kim, “Multi-Classifier Based LIDAR and Camera Fusion,” *2007 IEEE Intelligent Transportation Systems Conference*, pp. 467-472, 2007.

[25] M. Szarvas and U. Sakai, “Real-time Pedestrian Detection Using LIDAR and Convolutional Neural Networks,” in *2006 IEEE Intelligent Vehicles Symposium*, 2006, pp. 213-218.

[26] O. Ludwig, C. Premebida, U. Nunes, and R. Ara, “Evaluation of Boosting-SVM and SRM-SVM Cascade Classifiers in Laser and Vision-based Pedestrian Detection,” in *ITSC*, 2011, pp. 1574-1579.

[27] A. Pérez Grassi, V. Frolov, and F. Puente León, “Information fusion to detect and classify pedestrians using invariant features,” *Information Fusion*, vol. 12, no. 4, pp. 284-292, 2010.

[28] R. Labayrade, C. Royere, D. Gruyer, and D. Aubert, “Cooperative Fusion for Multi-Obstacles Detection With Use of Stereovision and Laser Scanner,” *Autonomous Robots*, vol. 19, no. 2, pp. 117-140, 2005.

[29] A. Broggi, P. Cerri, S. Ghidoni, P. Grisleri, and H. G. Jung, "Localization and analysis of critical areas in urban scenarios," in *Intelligent Vehicles Symposium, 2008 IEEE*, 2008, pp. 1074-1079.

[30] H. Cheng, N. Zheng, X. Zhang, J. Qin, and H. Van De Wetering, "Interactive Road Situation Analysis for Driver Assistance and Safety Warning Systems: Framework and Algorithms," *IEEE Transactions on Intelligent Transportation Systems*, vol. 8, no. 1, pp. 157-167, 2007.

[31] O. Aycard et al., "Grid based Fusion and Tracking," *2006 IEEE Intelligent Transportation Systems Conference*, pp. 2-7, 2006.

[32] M. Buehler et al., "DARPA Grand Challenge 2005," *Industrial Robot An International Journal*, vol. 31, no. 5, pp. 1-43-43, 2007.

[33] T. Defense and P. Agency, "Foreword for Journal of Field Robotics — Special Issue on the DARPA Grand Challenge," *Journal of Field Robotics*, vol. 23, no. 9, pp. 657-658, 2006.

[34] K. Iagnemma and E. M Buehler, "Special issue on the 2007 DARPA urban challenge," *Journal of Field Robotics*, vol. 25, pp. 423-860, 2008.

[35] K. Iagnemma and M. Buehler Eds, "Special Issue on the DARPA Grand Challenge, Part 1," *Journal of Field Robotics*, vol. 23, no. 8. pp. 461-652, 2006.

[36] S. Thrun et al., "Winning the DARPA Grand Challenge," *Journal of Field Robotics*, 2006.

[37] F. Bu and C.-yao Chan, "Pedestrian detection in transit bus application: sensing technologies and safety solutions," *IEEE Proceedings Intelligent Vehicles Symposium 2005*, vol. 2005, pp. 100-105, 2005.

[38] J. Langheim, A. J. Buchanan, U. Lages, and M. Wahl, "CARSENSE - New environment sensing for advanced driver assistance systems," in *Proceedings of the IEEE Intelligent Vehicle Symposium 2001*, 2001, pp. 89-94.

[39] D. M. Gavrila, M. Kunert, and U. Lages, "A multi-sensor approach for the protection of vulnerable traffic participants the PROTECTOR project," *IMTC 2001 Proceedings of the 18th IEEE Instrumentation and Measurement Technology Conference Rediscovering Measurement in the Age of Informatics Cat No01CH 37188*, vol. 3, pp. 2044-2048, 2001.

[40] C.-chih Wang, C. Thorpe, and S. Thrun, "Online simultaneous localization and mapping with detection and tracking of moving objects: theory and results from a ground vehicle in crowded urban areas," *2003 IEEE International Conference on Robotics and Automation Cat No03CH37422*, vol. 1, pp. 842-849, 2003.

[41] P. Moutarlier and R. Chatila, "Stochastic multisensory data fusion for mobile robot location and environment modeling," *5th Int Symposium on Robotics Research*, vol. 1, no. 2, pp. 85-94, 1989.

[42] F. Nashashibi and M. Devy, "3D Incremental Modeling and Robot Localization in a Structured Environment Using a Laser Range Finder," in *ICRA (1)*, 1993, pp. 20-27.

[43] C. Fruh and A. Zakhor, *Fast 3D model generation in urban environments*, vol. 54. VDI/VDE Soc. Meas. & Autom. Control, 2002, pp. 165-170.

[44] H. Zhao and R. Shibasaki, “A vehicle-borne urban 3-D acquisition system using single-row laser range scanners.,” *IEEE transactions on systems man and cybernetics Part B Cybernetics a publication of the IEEE Systems Man and Cybernetics Society*, vol. 33, no. 4, pp. 658-666, 2003.

[45] C.-C. Wang, C. Thorpe, and A. Suppe, “Ladar-Based Detection and Tracking of Moving Objects from a Ground Vehicle at High Speed,” in *Proceedings of the IEEE International Conference on Intelligent Vehicles Symposium IVS*, 2003.

[46] C. Thorpe, “A hierarchical object based representation for simultaneous localization and mapping,” *2004 IEEERSJ International Conference on Intelligent Robots and Systems IROS IEEE Cat No04CH37566*, vol. 1, pp. 412-418, 2004.

[47] T. Rabbani and F. V. D. Heuvel, “EFFICIENT HOUGH TRANSFORM FOR AUTOMATIC DETECTION OF CYLINDERS IN POINT CLOUDS,” *Transform*, vol. 3, pp. 60–65, 2005.

[48] S. Ahn, I. Effenberger, S. Roth-Koch, and E. Westkämper, “Geometric Segmentation and Object Recognition in Unordered and Incomplete Point Cloud,” vol. 2781, B. Michaelis and G. Krell, Eds. Springer Berlin / Heidelberg, 2003, pp. 450-457.

[49] S. Thrun, “Learning Occupancy Grid Maps with Forward Sensor Models,” *Autonomous Robots*, vol. 15, no. 2, pp. 111-127.

[50] M. K. Tay, K. Mekhnacha, C. Chen, M. Yguel, and C. Laugier, “An efficient formulation of the Bayesian occupation filter for target tracking in dynamic environments,” *International Journal Of Vehicle Autonomous Systems*, vol. 6, no. 1/2, p. 155, 2008.

[51] J. Sparbert, K. Dietmayer, and D. Streller, “Lane detection and street type classification using laser range images,” *ITSC 2001 2001 IEEE Intelligent Transportation Systems Proceedings Cat No01TH8585*, no. August, pp. 454-459, 2001.

[52] K. C. Fuerstenberg, K. C. J. Dietmayer, and V. Willhoeft, “Pedestrian Recognition in Urban Traffic using a vehicle based Multilayer Laserscanner,” *Intelligent Vehicle Symposium 2002 IEEE*, vol. 1, pp. 31-35, 2002.

[53] D. Streller, K. Furstenberg, and K. Dietmayer, “Vehicle and object models for robust tracking in traffic scenes using laser range images,” *Proceedings The IEEE 5th International Conference on Intelligent Transportation Systems*, no. September, pp. 118-123, 2002.

[54] K. C. Fuerstenberg, K. C. J. Dietmayer, S. Eisenlauer, and V. Willhoeft, “Multilayer Laserscanner for robust Object Tracking and Classification in Urban Traffic Scenes,” *Proceedings of ITS 2002 9th World Congress on Intelligent Transportation Systems Chicago USA*, pp. 1-10, 2002.

[55] K. C. Furstenberg, J. Hipp, A. Liebram, and K. A, “A Laser Scanner For Detailed Traffic Data Collection And Traffic Control.” 2000.

[56] G. Gate and F. Nashashibi, “Using targets appearance to improve pedestrian classification with a laser scanner,” *IEEE IV*, pp. 571-576, 2008.

[57] F. Nashashibi and A. Bargeton, “Laser-based vehicles tracking and classification using occlusion reasoning and confidence estimation,” *Intelligent Vehicles Symposium 2008 IEEE*, no. 1, pp. 847-852, 2008.

[58] T. Ogawa and K. Takagi, “Lane Recognition Using On-vehicle LIDAR,” *Transformation*, pp. 540-545, 2006.

[59] A. Mendes, L. C. Bento, and U. Nunes, “Multi-target detection and tracking with a laserscanner,” in *IEEE Intelligent Vehicles Symposium 2004*, 2004, pp. 796-801.

[60] C. Premebida and U. Nunes, “A Multi-Target Tracking and GMM-Classifier for Intelligent Vehicles,” in *Proc of the IEEE ITSC 2006*, 2006, pp. 313 - 318.

[61] N. Kaempchen, M. Buehler, and K. Dietmayer, “Feature-level fusion for free-form object tracking using laserscanner and video,” *IEEE Proceedings Intelligent Vehicles Symposium 2005*, pp. 453-458, 2005.

[62] R. MacLachlan and C. Mertz, “Tracking of Moving Objects from a Moving Vehicle Using a Scanning Laser Rangefinder,” *2006 IEEE Intelligent Transportation Systems Conference*, pp. 301-306, 2006.

[63] A. Fod, A. Howard, and M. A. J. Mataric, “A laser-based people tracker,” *Robotics and Automation 2002 Proceedings ICRA 02 IEEE International Conference on*, vol. 3, no. May, pp. 3024-3029, 2002.

[64] D. Schulz, W. Burgard, D. Fox, and A. B. Cremers, “Tracking multiple moving targets with a mobile robot using particle filters and statistical data association,” *Proceedings 2001 ICRA IEEE International Conference on Robotics and Automation Cat No01CH37164*, vol. 2, pp. 1665-1670, 2001.

[65] Y. Bar-Shalom and T. E. Fortmann, *Tracking and Data Association*, vol. 179. Academic Press, 1988, pp. 331-346.

[66] Y. Bar-Shalom, “Tracking methods in a multitarget environment,” *Automatic Control IEEE Transactions on*, vol. 23, no. 4, pp. 618-626, 1978.

[67] Y. Bar-Shalom and X.-R. Li, *Multitarget-Multisensor Tracking: Principles and Techniques*, vol. 16, no. 1. YBS Publishing, 1995, p. 600.

[68] D. Reid, “An algorithm for tracking multiple targets,” *IEEE Transactions on Automatic Control*, vol. 24, no. 6, pp. 843-854, 1979.

[69] R. Multiple, C. Tracking, and O. Reasoning, “California Partners for Advanced Transit and Highways (PATH) Robust Multiple Car Tracking With Occlusion Reasoning Robust Multiple Car Tracking With Occlusion Reasoning,” *Transportation*, 1994.

[70] B. Musleh, A. Escalera, and J. M. Armingol, “Real-Time Pedestrian Recognition in Urban Environments,” in *Advanced Microsystems for Automotive Applications 2011*, G. Meyer and J. Valldorf, Eds. Springer Berlin Heidelberg, 2011, pp. 139-147.

[71] A. Broggi et al., “The passive sensing suite of the TerraMax autonomous vehicle,” in *Intelligent Vehicles Symposium, 2008 IEEE*, 2008, pp. 769-774.

[72] M. Bertozzi, A. Broggi, C. Caraffi, M. Delrose, M. Felisa, and G. Vezzoni, “Pedestrian detection by means of far-infrared stereo vision,” *Computer Vision and Image Understanding*, vol. 106, no. 2-3, pp. 194-204, 2007.

[73] M. Bertozzi, A. Broggi, R. Chapuis, F. Chausse, A. Fascioli, and A. Tibaldi, “Shape-based pedestrian detection and localization,” *Group*, vol. 1, pp. 328-333, 2003.

[74] H. Yue, C. Shao, Y. Zhao, and X. Chen, “Study on Moving Pedestrian Tracking Based on Video Sequences,” *Journal of Transportation Systems Engineering and Information Technology*, vol. 7, no. 4, pp. 47-51, 2007.

[75] B. Musleh, F. García, J. Otamendi, J. M. Armingol, and A. De La Escalera, “Identifying and Tracking Pedestrians Based on Sensor Fusion and Motion Stability Predictions,” *Sensors (Peterboroug*, vol. 10, no. 9, pp. 8028-8053, 2010.

[76] M. Bertozzi, A. Broggi, S. Ghidoni, and M. M. Meinecke, “A Night Vision Module for the Detection of Distant Pedestrians,” *2007 IEEE Intelligent Vehicles Symposium*, pp. 25–30, 2007.

[77] M. J. Flores, J. M. Armingol, and A. Escalera, “Real-Time Warning System for Driver Drowsiness Detection Using Visual Information,” *Journal of Intelligent Robotic Systems*, vol. 59, no. 2, pp. 103-125, 2009.

[78] P. Viola and M. Jones, “Rapid object detection using a boosted cascade of simple features,” *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition CVPR 2001*, vol. 1, no. C, p. I-511-I-518, 2001.

[79] H. Bai, J. Wu, and C. Liu, “Motion and haar-like features based vehicle detection,” in *Multi-Media Modelling Conference Proceedings, 2006 12th International*, 2006, p. 4 pp.

[80] X. Cui, Y. Liu, S. Shan, X. Chen, and W. Gao, “3D Haar-Like Features for Pedestrian Detection,” in *Multimedia and Expo 2007 IEEE International Conference on*, 2007, pp. 1263-1266.

[81] M. Szarvas, A. Yoshizawa, M. Yamamoto, and J. Ogala, “Pedestrian detection with convolutional neural networks,” *IEEE Proceedings Intelligent Vehicles Symposium 2005*, pp. 224-229, 2005.

[82] J. P. Carrasco Pascual, “Advanced Driver Assistance System based on Computer Vision using Detection, Recognition and Tracking of Road Signs,” Universidad Carlos III de Madrid, 2009.

[83] N. Dalal and W. Triggs, “Histograms of Oriented Gradients for Human Detection,” *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition CVPR05*, vol. 1, no. 3, pp. 886-893, 2004.

[84] A. De La Escalera and J. M. Armingol, “Vehicle detection and tracking for visual understanding of road environments,” *Robotica*, vol. 28, no. 6, pp. 847-860, 2010.

[85] H. Nanda and L. Davis, “Probabilistic template based pedestrian detection in infrared videos,” *Intelligent Vehicle Symposium 2002 IEEE*, vol. 1, no. 6, pp. 15-20, 2002.

[86] M. Bertozzi, A. Broggi, C. H. Gomez, R. I. Fedriga, G. Vezzoni, and M. DelRose, “Pedestrian Detection in Far Infrared Images based on the use of Probabilistic Templates,” *2007 IEEE Intelligent Vehicles Symposium*, pp. 327-332, 2007.

[87] J. Gó andmez-Romero et al., “Strategies and techniques for use and exploitation of Contextual Information in high-level fusion architectures,” in *Information Fusion (FUSION), 2010 13th Conference on*, 2010, pp. 1-8.

[88] D. Olmeda, A. De La Escalera, and J. M. Armingol, “Detection and tracking of pedestrians in infrared images,” *2009 3rd International Conference on Signals Circuits and Systems SCS*, pp. 1-6, 2009.

[89] J. M. Collado, C. Hilario, A. De La Escalera, and J. M. Armingol, “Adaptative road lanes detection and classification,” in *Transform*, 2006, vol. 4179, pp. 1151-1162.

[90] B. Musleh, A. D. L. Escalera, and J. M. Armingol, “Obstacle Detection and Localization Using U-V disparity with Accelerated Processing Time by CUDA,” in *IEEE International Conference on Robotics and Automation 2011 ICRA*, 2011.

[91] “MTI-G X-sens.” [Online]. Available: http://www.xsens.com/en/general/mti-g. [Accessed: 05-Mar-2012].

[92] R. J. Skehill, M. Barry, and S. Mcgrath, “Mobility Modelling with Empirical Pedestrian and Vehicular Traffic Characteristics,” *WSEAS TRANSACTIONS on COMMUNICATIONS*, vol. 4, no. 10, 2005.

[93] *Highway Capacity Manual 2000*, no. 193. Transportation Research Board, National Academy of Sciences, 2000.

[94] G. K. Still, “Crowd dynamics,” University of Warwick, 2000.

[95] “Open Computer Vision library(OpenCV).” [Online]. Available: http://opencv.willowgarage.com/wiki/. [Accessed: 11-Mar-2012].

[96] Y. Freund and R. E. Schapire, “A decision theoretic generalization of on-line learning and an application to boosting.,” *Journal of Computer and System Sciences*, vol. 55, no. 1, pp. 119-139, 1997.

[97] Y. Bar-Shalom and X.-R. Li, *Estimation and tracking - Principles, techniques, and software*. Boston: Artech House.

[98] R. E. Kalman, “A new approach to linear filtering and prediction problems,” *Journal Of Basic Engineering*, vol. 82, no. Series D, pp. 35-45, 1960.

[99] S. J. Julier, “New extension of the Kalman filter to nonlinear systems,” *Proceedings of SPIE*, vol. 3, no. 1, pp. 182-193, 1997.

[100] S. Julier and J. Uhlmann, “Unscented filtering and nonlinear estimation,” *Proceedings of the IEEE*, vol. 92, no. 3, pp. 401-422, 2004.

[101] A. Elfes, “Using occupancy grids for mobile robot perception and navigation,” *Computer*, vol. 22, no. 6, pp. 46-57, 1989.

[102] H. Weigel, P. Lindner, and G. Wanielik, “Vehicle Tracking with Lane Assignment by Camera and Lidar Sensor Fusion,” in *IEEE Intelligent Vehicles Symposium*, 2009, pp. 513-520.

[103] T.-D. Vu, J. Burlet, and O. Aycard, “Grid-based localization and local mapping with moving object detection and tracking,” *Information Fusion*, vol. 12, no. 1, pp. 58-69, 2011.

[104] B. Grabe, T. Ike, and M. Hoetter, “Evaluation method of grid based representation from sensor data,” *2009 IEEE Intelligent Vehicles Symposium*, pp. 1245-1250, 2009.

[105] K. Pathak, A. Birk, J. Poppinga, and S. Schwertfeger, “3D forward sensor modeling and application to occupancy grid based sensor fusion,” *2007 IEEERSJ International Conference on Intelligent Robots and Systems*, vol. 2, pp. 2059-2064, 2007.

[106] N. J. Gordon, D. J. Salmond, and A. F. M. Smith, “Novel approach to nonlinear non-Gaussian Bayesian state estimation.,” *IEE PROCEEDINGSF RADAR AND SIGNAL PROCESSING*, vol. 140, no. 2, pp. 107-113, 1993.

[107] C. Hue, J. P. Le Cadre, and P. Perez, “Sequential Monte Carlo methods for multiple target tracking and data fusion,” *IEEE Transactions on Signal Processing*, vol. 50, no. 2, pp. 309-325, 2002.

[108] A. B. Poore, N. Rijavec, M. Liggins, and V. Vannicola, “Data association problems posed as multidimensional assignment problems: problem formulation,” in *Signal and Data Processing of Small Targets 1954*, 1993, pp. 552-561.

[109] A. B. Poore, N. Rijavec, T. Barker, and M. Munger, “Data association problems posed as multidimensional assignment problems: numerical simulations,” in *Signal and Data Processing of Small Targets 1954*, 1993, pp. 564-573.

[110] D. Maksarov and H. Durrant-Whyte, “Mobile vehicle navigation in unknown environments: a multiple hypothesis approach,” *IEE ProcControl Theory Appl*, vol. 142, pp. 385-400, 1995.

[111] I. Coraluppi, Stefano (NATO Undersea Research Centre and I. Carthel, Craig (NATO Undersea Research Centre, “Multi-hypothesis sonar tracking,” in *7 th International Conference on Information Fusion.*, 2004, pp. 33-40.

[112] C. Hue, J. P. Le Cadre, and P. Perez, “A particle filter to track multiple objects,” *Proceedings 2001 IEEE Workshop on MultiObject Tracking*, no. 1, pp. 61-68, 2001.

[113] M. Kohler, “Using the Kalman Filter to track Human Interactive Motion - Modelling and Initialization of the Kalman Filter for Translational Motion -,” 1997.

[114] G. Johansson and K. Rumar, “Drivers’ brake reaction times.,” *Human Factors*, vol. 13, no. 1, pp. 23-27, 1971.

[115] H. Makishita and K. Matsunaga, “Differences of drivers’ reaction times according to age and mental workload.,” *Accident analysis and prevention*, vol. 40, no. 2, pp. 567-575, 2008.

[116] J. C. Collins, *Accident Reconstruction*. Charles C. Thomas Publisher, Limited, 1979.

[117] R. E. Fenton, “A headway safety policy for automated highway operations,” *IEEE Transactions on Vehicular Technology*, vol. 28, no. 1, pp. 22-28, 1979.

[118] C. C. Chien and P. Ioannou, “Automatic vehicle-following,” in *American Control Conference 1992*, 1992, pp. 1748–1752.

[119] P. A. Ioannou and C. C. Chien, “Autonomous intelligent cruise control,” *IEEE Transactions on Vehicular Technology*, vol. 42, no. 4, pp. 657-672, 1993.

[120] “Instituto Universitario de Investigación del Automóvil (INSIA).” [Online]. Available: http://www.insia-upm.es/. [Accessed: 20-May-2012].