ADVANCED DRIVER ASSISTANCE SYSTEMS WITH COMPUTER VISION METHODS

ΠΡΟΗΓΜΕΝΑ ΣΥΣΤΗΜΑΤΑ ΥΠΟΒΟΗΘΗΣΗΣ ΟΔΗΓΟΥ ΜΕ ΜΕΘΟΔΟΥΣ ΥΠΟΛΟΓΙΣΤΙΚΗΣ ΟΡΑΣΗΣ

DOCTORAL DISSERTATION
OF
GEORGIOS SIOGKAS
DIPLOMA IN ELECTRICAL AND COMPUTER ENGINEERING

UNIVERSITY OF PATRAS
DEPARTMENT OF ELECTRICAL AND COMPUTER ENGINEERING

DISSERTATION No: 308
PATRAS - JUNE 2013

ΠΙΣΤΟΠΟΙΗΣΗ

Πιστοποιείται ότι η παρούσα διδακτορική διατριβή με θέμα:

«ΠΡΟΗΓΜΕΝΑ ΣΥΣΤΗΜΑΤΑ ΥΠΟΒΟΗΘΗΣΗΣ ΟΔΗΓΟΥ ΜΕ ΜΕΘΟΔΟΥΣ ΥΠΟΛΟΓΙΣΤΙΚΗΣ ΟΡΑΣΗΣ»

του κ. Γεώργιου Σιόγκα, παρουσιάστηκε δημοσίως την Δευτέρα 17 Ιουνίου 2013 και ώρα 11:00 στην αίθουσα συνεδριάσεων του Τμήματος Ηλεκτρολόγων Μηχανικών και Τεχνολογίας Υπολογιστών του Πανεπιστημίου Πατρών, εξετάστηκε και εγκρίθηκε από την ακόλουθη εξεταστική επιτροπή:

1. **Δερματάς Ευάγγελος**, Αναπληρωτής Καθηγητής (Επιβλέπων, Τμήμα ΗΜ&ΤΥ, Πανεπιστήμιο Πατρών)
2. **Τζές Αντώνιος**, Καθηγητής (Μέλος της Τριμελούς, Τμήμα ΗΜ&ΤΥ, Πανεπιστήμιο Πατρών)
3. **Φακωτάκης Νίκος,** Καθηγητής (Τμήμα ΗΜ&ΤΥ, Πανεπιστήμιο Πατρών)
4. **Μπερμπερίδης Κωνσταντίνος,** Καθηγητής (Τμήμα Μηχανικών Η/Υ & Πληροφορικής, Πανεπιστήμιο Πατρών)
5. **Ψαράκης Εμμανουήλ**, Επίκουρος Καθηγητής (Τμήμα Μηχανικών Η/Υ & Πληροφορικής, Πανεπιστήμιο Πατρών)
6. **Οικονόμου Γεώργιος**, Καθηγητής (Τμήμα Φυσικής, Πανεπιστήμιο Πατρών)

17 Ιουνίου 2013

Ο επιβλέπων Καθηγητής

Ευάγγελος Σ. Δερματάς
Αναπ. Καθηγητής

Ο αναπληρωτής πρόεδρος του Τμήματος

Γαβριήλ Β. Γιαννακόπουλος
Καθηγητής

Abstract

Traffic accidents are one of the main reasons for the loss of human lives worldwide. Their increasing number has led to the realization that the use of advanced technology for manufacturing safer vehicles is imperative for limiting casualties. Since technological breakthroughs allowed the incorporation of cheap, low consumption systems with high processing speeds in vehicles, it became apparent that complex computer vision techniques could be used to assist drivers in navigating their vehicles. In this direction, this thesis focuses on providing novel solutions for different tasks involved in advanced driver assistance systems. More specifically, this thesis proposes novel sub-systems for traffic sign recognition, traffic light recognition, preceding vehicle detection and road detection. The techniques used for developing the proposed solutions are based on color image processing with a focus on illumination invariance, using symmetry information for man-made objects (like traffic signs, traffic lights and vehicles) detection, spatiotemporal tracking of detected results and automated image segmentation for road detection. The proposed systems were implemented with a goal of robustness to changes of illumination and weather conditions, as well as to diverse driving environments. A special focus on the prospect for real-time implementation has also been given. The results presented in this thesis indicate the superiority of the proposed methods to their counterparts found in relevant literature in both normal and challenging conditions, especially in the cases of preceding vehicle detection and road detection. Hopefully, parts of this research will provide new insights for future developments in the field of intelligent transportation.

Περίληψη

Τα αυτοκινητιστικά δυστυχήματα αποτελούν μια από τις κυριότερες αιτίες θανάτου παγκοσμίως. Ο αυξανόμενος αριθμός τους οδήγησε στην συνειδητοποίηση ότι η χρήση προηγμένης τεχνολογίας για την κατασκευή ασφαλέστερων οχημάτων είναι απαραίτητη για την μείωση των ατυχημάτων και κατά συνέπεια των θανάτων που οφείλονται σε αυτά. Από τη στιγμή που οι τεχνολογικές εξελίξεις επέτρεψαν την ενσωμάτωση φθηνών, χαμηλής κατανάλωσης συστημάτων με μεγάλη επεξεργαστική ταχύτητα σε οχήματα, κατέστη προφανές ότι περίπλοκες τεχνικές υπολογιστικής όρασης μπορούσαν πλέον να χρησιμοποιηθούν για την υποβοήθηση της οδήγησης. Σε αυτή την κατεύθυνση, η παρούσα διατριβή εστιάζει στην ανάπτυξη καινοτόμων λύσεων για διαφορετικά κομμάτια που εμπλέκονται στα προηγμένα συστήματα υποβοήθησης του οδηγού. Πιο συγκεκριμένα, σε αυτή την διατριβή προτείνονται καινοτόμα υποσυστήματα για την αναγνώριση σημάτων οδικής κυκλοφορίας, την αναγνώριση φωτεινών σηματοδοτών, τον εντοπισμό προπορευόμενου οχήματος και τον εντοπισμό δρόμου. Οι τεχνικές που χρησιμοποιήθηκαν για την ανάπτυξη των προτεινόμενων λύσεων βασίζονται στην χρωματική επεξεργασία εικόνας με έμφαση στην ανεξαρτησία από την φωτεινότητα της σκηνής, στην χρήση πληροφορίας συμμετρίας για τον εντοπισμό χαρακτηριστικών αντικειμένων (όπως σήματα οδικής κυκλοφορίας, φωτεινοί σηματοδότες και οχήματα), στην χωροχρονική παρακολούθηση των εντοπισμένων αντικειμένων και στην αυτόματη κατάτμηση εικόνας για τον εντοπισμό δρόμου. Τα προτεινόμενα συστήματα αναπτύχθηκαν με στόχο την ανθεκτικότητα σε αλλαγές της φωτεινότητας ή τις καιρικές συνθήκες, καθώς και στην οδήγηση σε απαιτητικά περιβάλλοντα. Επίσης, έχει δοθεί ιδιαίτερη έμφαση στην προοπτική υλοποίησης συστημάτων πραγματικού χρόνου. Τα αποτελέσματα που παρουσιάζονται σε αυτή την διατριβή αποδεικνύουν την ανωτερότητα των προτεινόμενων μεθόδων έναντι αντίστοιχων της σχετικής βιβλιογραφίας, ειδικά στις περιπτώσεις του εντοπισμού προπορευόμενου οχήματος και του εντοπισμού δρόμου. Ελπίζουμε ότι μέρη της έρευνας αυτής θα εμπνεύσουν νέες προσεγγίσεις για τις μελλοντικές υλοποιήσεις αντίστοιχων συστημάτων.

Εκτεταμένη Περίληψη στα Ελληνικά

Extended Abstract in Greek

1. Εισαγωγή

Κατά τις περασμένες δεκαετίες, τα αυτοκινητικά δυστυχήματα έχουν εξελιχθεί σε ένα από τα πρώτα αίτια θανάτων παγκοσμίως. Όπως προκύπτει από μια έρευνα του Παγκόσμιου Οργανισμού Υγείας που εκδόθηκε το 2004 [1], τα αυτοκινητικά δυστυχήματα κόστισαν την ζωή σε περίπου 1.2 εκατομμύρια ανθρώπους παγκοσμίως, ενώ περίπου 50 εκατομμύρια άνθρωποι τραυματίστηκαν σοβαρά εξ' αιτίας τους. Η σοβαρότητα της κατάστασης ενισχύεται και από το γεγονός ότι το 2000 τα αυτοκινητικά δυστυχήματα ήταν η κυριότερη αιτία θανάσιμων τραυματισμών παγκοσμίως, συμβάλλοντας στο 23% των θανάσιμων τραυματισμών. Επιπροσθέτως, το 2004 ήταν η πρώτη αιτία θανάτων στις ηλικίες 15-29 και η δεύτερη στις ηλικίες 5-14.

Ακόμα πιο ανησυχητικές είναι οι προβλέψεις ότι ο αριθμός των θανάτων από αυτοκινητικά δυστυχήματα θα υποστεί αύξηση κατά 65% στο χρονικό διάστημα 2000-2020 [2], [3]. Πιο συγκεκριμένα, ο αριθμός των θανάτων από το αίτιο αυτό αναμένεται να διπλασιαστεί στην χρονική περίοδο 2004-2030 [1]. Μιλώντας με ποσοστά, τα αυτοκινητικά δυστυχήματα προβλέπεται να γίνουν η 5^{η} αιτία θανάτων μέχρι το 2030, προκαλώντας το 3.6% από 2.2% των θανάτων.

Τα νοσηρά αυτά στατιστικά για την αύξηση των θανάτων από αυτοκινητικά δυστυχήματα οφείλονται κυρίως στην διαρκή αύξηση του αριθμού αυτοκινήτων που κινούνται παγκοσμίως, οδηγώντας σε μια αντίστοιχη αύξηση των ατυχημάτων. Εντούτοις, η οδική ασφάλεια κατά τα τελευταία χρόνια έχει βελτιωθεί κυρίως εξαιτίας των τεχνολογικών εξελίξεων στον χώρο των συστημάτων παθητικής ασφάλειας τα οποία έχουν επιβληθεί σε όλα τα νέα μοντέλα επιβατηγών οχημάτων. Τέτοια συστήματα, όπως είναι οι ζώνες ασφαλείας, οι αερόσακοι και διάφορες βελτιώσεις στο σασί των οχημάτων, βελτίωσαν σημαντικά την αξιοπιστία τους σε περίπτωση ατυχήματος, τόσο για τους επιβάτες, όσο και για τους πεζούς. Παρόλα αυτά, οι βελτιώσεις των συστημάτων παθητικής ασφάλειας φαίνεται να έχουν φτάσει σε έναν κορεσμό από άποψη μείωσης των δυστυχημάτων. Ο κυριότερος λόγος, είναι ότι αδυνατούν να συμβάλλουν στην αποφυγή ατυχημάτων που οφείλονται σε οδηγικά σφάλματα. Μια πρόσφατη έρευνα του Εθνικού Οργανισμού Διαχείρισης Οδικής Ασφάλειας σε Αυτοκινητοδρόμους στις Η.Π.Α. [4], έδειξε ότι ένα ποσοστό 41% των οδικών ατυχημάτων που σχετίζονται με τον οδηγό οφείλεται σε σφάλματα αναγνώρισης του κινδύνου, ενώ και ένα 34% οφείλεται σε σφάλματα απόφασης.

Από όλα τα παραπάνω, μπορεί να εξαχθεί το συμπέρασμα ότι η περαιτέρω μείωση των θανατηφόρων οδικών ατυχημάτων πρέπει να βασιστεί πλέον σε ενεργητικά συστήματα υψηλής τεχνολογικής στάθμης. Πολλά από αυτά εντάσσονται στην ευρύτερη κατηγορία που ονομάζεται Προηγμένα Συστήματα Υποβοήθησης

Οδήγησης (ΠΣΥΟ). Τα ΠΣΥΟ είναι νέες αναπτυσσόμενες τεχνολογίες που ως πρωταρχικό σκοπό έχουν να παρέχουν αυτόματη υποστηρικτική πληροφορία στους οδηγούς, η οποία μπορεί να συμβάλλει στην μείωση των ατυχημάτων.

Σε αυτή την κατεύθυνση, το Ευρωπαϊκό Πρόγραμμα Αξιολόγησης Καινούργιων Αυτοκινήτων έχει ανακοινώσει [5] ότι από το έτος 2014, κάθε αυτοκινητοβιομηχανία που θέλει να επιτύχει μια τέλεια (5-αστέρων) βαθμολογία για τα οχήματά της θα πρέπει να τα εξοπλίζει με ένα σύστημα Αυτόματης Επείγουσας Πέδησης. Παρόμοιες πρωτοβουλίες του Euro NCAP από την ίδρυσή του το 1996 έχουν προκαλέσει την βελτίωση των συστημάτων παθητικής αλλά και ενεργητικής ασφάλειας που χρησιμοποιούνται στα επιβατηγά οχήματα. Τα κριτήρια που τίθενται από τον οργανισμό είναι μια καλή εκτίμηση των τεχνολογιών αιχμής που θα πρέπει να ενσωματωθούν στα οχήματα καθημερινής χρήσης. Κάποια από τα μελλοντικά σχέδια μπορούν να γίνουν εμφανή με μια μελέτη των επιβραβεύσεων που έχει θεσπίσει από το 2010 το Euro NCAP, σχετικά με την ενσωμάτωση προηγμένων συστημάτων ασφαλείας στα αυτοκίνητα παραγωγής.

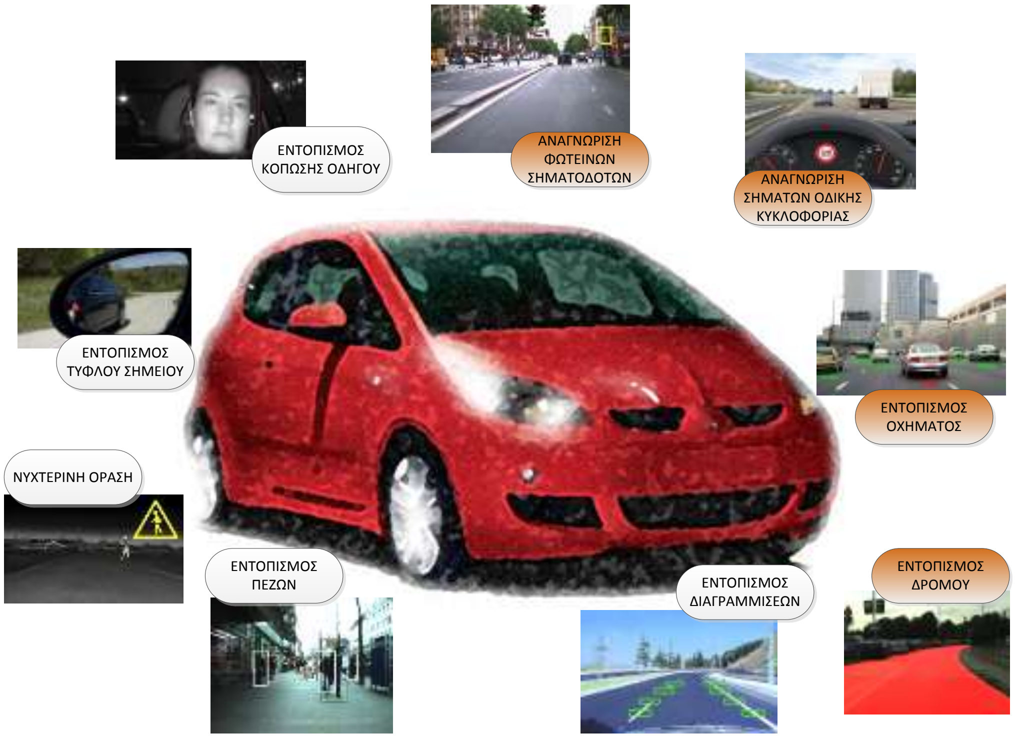
Τα συστήματα που έχουν επιβραβευθεί ως τώρα από το Euro NCAP υποδηλώνουν μια γενική κατεύθυνση προς την χρήση μεθόδων υπολογιστικής όρασης για την ανάπτυξη των σύγχρονων ΠΣΥΟ. Είτε με την μορφή υποστηρικτικής πληροφορίας σε συστήματα συστοιχίας αισθητήρων, είτε ως μοναδική πηγή πληροφορίας, οι κάμερες αποτελούν σημαντικό εργαλείο για τα ΠΣΥΟ. Πιο συγκεκριμένα, από όλα τα συστήματα που έχουν επιβραβευθεί μέχρι σήμερα, μόνο η Αυτόματη Επείγουσα Κλήση δεν χρησιμοποιεί οπτική πληροφορία. Ο κυριότερος λόγος που έγινε αυτή η στροφή προς την οπτική πληροφορία σε σχέση με προηγούμενες υλοποιήσεις των ΠΣΥΟ είναι η διαρκώς αυξανόμενη επεξεργαστική ισχύς των ενσωματωμένων συστημάτων ψηφιακής επεξεργασίας σήματος, που, σε συνδυασμό με την ανάπτυξη φθηνών βιντεοκαμερών με καλή ποιότητα εικόνας, επέτρεψαν την γρήγορη υλοποίηση πολύπλοκων αλγορίθμων υπολογιστικής όρασης σε προσιτές τιμές για τα επιβατηγά οχήματα.

2. Δομή ενός ΠΣΥΟ Βασισμένου σε Οπτική Πληροφορία

Τα μοντέρνα ΠΣΥΟ που έχουν βγει ως τώρα στην αγορά χρησιμοποιούν κυρίως συστοιχίες αισθητήρων και καμερών. Ένα ενδεικτικό σενάριο για ένα ολοκληρωμένο ΠΣΥΟ που θα χρησιμοποιεί μόνο οπτική πληροφορία από συστοιχία καμερών, παρουσιάζεται στην Εικόνα 1.

Στόχος αυτής της διατριβής είναι η έρευνα και ανάπτυξη λύσεων για κάποια από τα πιο απαιτητικά προβλήματα που μπορούν να βασιστούν στην χρήση μιας κάμερας τοποθετημένης στο ταμπλό του οχήματος, όπως είναι:

- η αναγνώριση σημάτων οδικής κυκλοφορίας (ΑΣΟΚ),
- η αναγνώριση φωτεινών σηματοδοτών (ΑΦΣ),
- ο εντοπισμός προπορευόμενων οχημάτων (ΕΠΟ) και
- ο εντοπισμός δρόμου (ΕΔ).


ΕΝΤΟΠΙΣΜΟΣ ΚΟΠΩΣΗΣ ΟΔΗΓΟΥ
ΑΝΑΓΝΩΡΙΣΗ ΦΩΤΕΙΝΩΝ ΣΗΜΑΤΟΔΟΤΩΝ
ΑΝΑΓΝΩΡΙΣΗ ΣΗΜΑΤΩΝ ΟΔΙΚΗΣ ΚΥΚΛΟΦΟΡΙΑΣ
ΕΝΤΟΠΙΣΜΟΣ ΤΥΦΛΟΥ ΣΗΜΕΙΟΥ
ΕΝΤΟΠΙΣΜΟΣ ΟΧΗΜΑΤΟΣ
ΝΥΧΤΕΡΙΝΗ ΟΡΑΣΗ
ΕΝΤΟΠΙΣΜΟΣ ΠΕΖΩΝ
ΕΝΤΟΠΙΣΜΟΣ ΔΙΑΓΡΑΜΜΙΣΕΩΝ
ΕΝΤΟΠΙΣΜΟΣ ΔΡΟΜΟΥ

Εικόνα 1: Ενδεικτική δομή ενός ΠΣΥΟ ολοκληρωτικά βασισμένου σε οπτική πληροφορία. Σε αυτή την διατριβή έχουν μελετηθεί τα σκιασμένα υποσυστήματα .

Ειδική μέριμνα δόθηκε στην ανάπτυξη και την αξιολόγηση συστημάτων που να είναι ανθεκτικά σε διαφορετικές συνθήκες καιρού και φωτισμού, καθώς και να παρέχουν ικανοποιητική απόδοση σε πρωινές και νυχτερινές σκηνές οδήγησης, χωρίς να επηρεάζονται αρνητικά από το περιβάλλον οδήγησης.

3. Βάσεις Δεδομένων για την Ανάπτυξη και Αξιολόγηση των ΠΣΥΟ

Ένα από τα μεγαλύτερα προβλήματα του χώρου της υπολογιστικής όρασης είναι η έλλειψη προσεχτικά επιλεγμένων, διαθέσιμων στο ευρύ κοινό βάσεων δεδομένων με πληθώρα βιντεοσκοπήσεων χαρακτηρισμένων ανά εικονοστοιχείο από ειδικούς για την αξιολόγηση των συστημάτων. Όσο πιο περίπλοκο είναι το πρόβλημα που επιχειρεί να λύσει το σύστημα, τόσο μεγαλύτερο θα πρέπει να είναι και το σύνολο των βιντεοσκοπήσεων που θα χρησιμοποιηθούν στην αξιολόγηση, ώστε να καλύπτουν όλο το εύρος των συνθηκών που μπορεί να προκύψουν. Για τα συστήματα που αναπτύχθηκαν στα πλαίσια αυτής της διατριβής, χρησιμοποιήθηκε ένας αρκετά μεγάλος αριθμός από βιντεοσκοπήσεις, άλλες με ενσωματωμένους χαρακτηρισμούς περιοχών για ποσοτική αξιολόγηση και άλλες χωρίς χαρακτηρισμούς, για ποιοτική αξιολόγηση. Το σύνολο των βιντεοσκοπήσεων που χρησιμοποιήθηκαν παρατίθεται στον Πίνακας 1.

Πίνακας 1: Βιντεοσκοπήσεις που χρησιμοποιήθηκαν για τις ανάγκες της διατριβής.

No	Name	Used for	Annot.	Car	Camera	Resolution	fps	Duration	Environ.	Weather	Daytime	Place
1	TSR1	TSR	No	AX	DCR-TRV60E	720x576	25	8:55	Rural	Sunny	Day	Ioannina
2	TSR2	TSR	No	AX	DCR-HC85	720x576	25	2:10	Rural	Rainy	Day	Patras
3	TSR3	TSR	No	AX	DCR-HC85	720x576	25	1:05	Rural	Good	Night	Patras
4	TSR4	TSR	No	AX	DCR-HC85	720x576	25	3:07	City	Good	Night	Patras
5	Patras1	RD/VD	No	AX	Pinhole	640x480	25	> 2 h	Mixed	Mixed	Mixed	Patras
6	Patras2	RD/VD	No	Colt	PV-GS180	720x576	25	> 1 h	City	Sunny	Day	Patras
7	Ioann1	RD/VD	No	Colt	HDC-SD100	1080i	25	~1 h	Rural	Mixed	Day	Ioannina
8	LARA	TLR/VD	TLs	C3	Marling F-046C	640x480	25	8:49	City	Sunny	Day	Paris
9	DiploDoc	RD	Road	-	MEGA-D	320x240	15	0:56	Mixed	Sunny	Noon	Trento
10	Alvarez1	RD	Road	ZZ	Bumblebee	320x240	15	0:56	Rural	Sunny	Noon	Barcelona
11	Alvarez2	RD	Road	ZZ	Bumblebee	320x240	15	0:32	Rural	Rainy	Morning	Barcelona
12	HRI1	VD	R/V/TS/TL	-	Unknown	800x600	10	16:24	Mixed	Dry	Afternoon	Unknown
13	HRI2	VD	R/V/TS/TL	-	Unknown	800x600	20	18:50	Mixed	Dry	Evening	Unknown
14	HRI3	VD	R/V/TS/TL	-	Unknown	800x600	10	11:12	Mixed	Rainy	Afternoon	Unknown
15	HRI4	VD	R/V/TS/TL	-	Unknown	800x600	10	11:22	Mixed	Dry	Night	Unknown
16	HRI5	VD	R/V/TS/TL	-	Unknown	800x600	20	13:48	Mixed	Snow	Afternoon	Unknown
17	YouTube	TLR/VD	No	-	Various	Various	-	> 2 h	Various	Various	Various	Various
18	Caltech 1999	VD	No	-	Unknown	892x592	-	126 frames	Parking	Sunny	Day	Caltech
19	Caltech 2001	VD	No	-	Unknown	360x240	-	526 frames	Urban	Sunny	Day	S. California

Οι παραπάνω βιντεοσκοπήσεις, χρησιμοποιήθηκαν για διάφορους λόγους στην διάρκεια της διατριβής. Πιο συγκεκριμένα:

i. Τα βίντεο που περιγράφονται στις γραμμές 1 έως και 7 του Πίνακας 1 τραβήχτηκαν αποκλειστικά για τους σκοπούς της διατριβής, με διαφορετικές κάμερες, σε διαφορετικές συνθήκες, εποχές και ώρες της ημέρας. Χρησιμοποιήθηκαν για χρήση κατά την ανάπτυξη των συστημάτων, καθώς και για την ποιοτική τους αξιολόγηση.

ii. Για τους σκοπούς της ποσοτικής αξιολόγησης των συστημάτων που αναπτύχθηκαν στα πλαίσια της διατριβής, χρησιμοποιήθηκαν τα βίντεο που αναλύονται στις γραμμές 8 έως και 11 του Πίνακας 1. Επίσης, μια ποσοτική αξιολόγηση για το σύστημα ΑΣΟΚ βασίστηκε και στις βιντεοσκοπήσεις TSR1 έως TSR4 που παρουσιάζονται στις πρώτες 4 γραμμές του Πίνακας 1. Το σύστημα ΑΦΣ που αναπτύχθηκε αξιολογήθηκε με βάση την βιντεοσκόπηση LARA, που αποτελείται από 11179 καρέ και παρουσιάζεται στο [6]. Η βάση αυτή περιέχει χαρακτηρισμένους 32 φωτεινούς σηματοδότες. Η βιντεοσκόπηση αυτή έχει χρησιμοποιηθεί επίσης και για την αξιολόγηση των συστημάτων ΕΠΟ που αναπτύχθηκαν. Τέλος, για την ποσοτική αξιολόγηση του συστήματος ΕΔ που αναπτύχθηκε στα πλαίσια της διατριβής, χρησιμοποιήθηκαν και οι βιντεοσκοπήσεις που παρουσιάζονται στις γραμμές 9, 10 και 11 του Πίνακας 1. Η βάση DiploDoc αποτελείται από 5 σύντομα βίντεο οδήγησης, αποτελούμενα στο σύνολό τους από 865 καρέ με σημειωμένο το περίγραμμα του δρόμου και έχει χρησιμοποιηθεί στα [7], [8]. Η βάση που χρησιμοποιήθηκε από τον Alvarez στο [9] και παρουσιάζεται στις γραμμές 10 και 11 του Πίνακας 1 αποτελείται από 2 βιντεοσκοπήσεις συνόλου 1335 καρέ με χαρακτηρισμένο τον δρόμο στα 1005 από αυτά.

iii. Τα υπόλοιπα βίντεο που περιγράφονται στον Πίνακας 1 χρησιμοποιήθηκαν για την ποιοτική αξιολόγηση των συστημάτων της διατριβής σε διαφορετικές καιρικές συνθήκες και ώρες της ημέρας, καθώς επίσης και σε διαφορετικά περιβάλλοντα και χώρες.

iv. Τέλος, για την αξιολόγηση των στατικών μεθόδων αναγνώρισης προπορευόμενου οχήματος που αναπτύχθηκαν στην διατριβή χρησιμοποιήθηκαν και δύο βάσεις δεδομένων με φωτογραφίες του πίσω μέρους αυτοκινήτων τραβηγμένες α) μέσα σε χώρο στάθμευσης (Caltech 1999) και β) κατά την διάρκεια οδήγησης (Caltech 2001).

4. Μετρικές Αξιολόγησης της Απόδοσης των Συστημάτων

Για να γίνει εφικτή η αξιολόγηση των συστημάτων που αναπτύχθηκαν στα πλαίσια της διατριβής είναι αναγκαία η χρησιμοποίηση κατάλληλων μεγεθών και μεθόδων ποσοτικής και ποιοτικής αξιολόγησης. Πιο συγκεκριμένα, για την αξιολόγηση των συστημάτων χρησιμοποιήθηκαν:

Ποσοτικές μετρικές που χρησιμοποιούνται συχνά σε ανάλογες αξιολογήσεις και βασίζονται σε τέσσερις διακριτούς αριθμούς που προκύπτουν κατά την κατηγοριοποίηση των εικονοστοιχείων μιας εικόνας σε δύο κατηγορίες (θετικά και αρνητικά): τον αριθμό των Αληθώς Θετικών (True Positive, TP) εικονοστοιχείων, δηλαδή των εικονοστοιχείων που κατηγοριοποιήθηκαν σωστά ως θετικά, τον αριθμό των Ψευδώς Θετικών (False Positive, FP) εικονοστοιχείων, δηλαδή των εικονοστοιχείων που κατηγοριοποιήθηκαν λανθασμένα ως θετικά, τον αριθμό των Αληθώς Αρνητικών (True Negative, TN) εικονοστοιχείων, δηλαδή των εικονοστοιχείων που κατηγοριοποιήθηκαν σωστά ως αρνητικά και τον αριθμό των Ψευδώς Αρνητικών (False Negative, FN) εικονοστοιχείων, δηλαδή των εικονοστοιχείων που κατηγοριοποιήθηκαν λανθασμένα ως αρνητικά. Με βάση τους παραπάνω ορισμούς, μπορούμε να ορίσουμε τις ακόλουθες μετρικές που μπορούν να χαρακτηρίσουν την απόδοση της κατηγοριοποίησης:

- Ορθότητα: $P = TP / (TP + FP)$
- Πληρότητα: $R = TP / (TP + FN)$
- Ποιότητα: $g = TP / (TP + FP + FN)$
- Ακρίβεια: $A = (TP + TN) / (TP + TN + FP + FN)$
- Αποτελεσματικότητα: $F = 2PR / (P + R)$

Όλα τα παραπάνω μεγέθη έχουν εύρος τιμών από το 0 (χειρότερο αποτέλεσμα) έως το 1 (καλύτερο αποτέλεσμα).

Εκτός από την ποσοτική αξιολόγηση, τα συστήματα που βασίζονται σε μεθόδους υπολογιστικής όρασης πρέπει να αξιολογούνται και ποιοτικά, για δύο σημαντικούς λόγους: πρώτον γιατί μια χαρακτηρισμένη βάση δεδομένων που να καλύπτει όλες τις πιθανές εκδοχές ενός πολύπλοκου προβλήματος όπως η οδήγηση είναι πρακτικά αδύνατον να κατασκευαστεί και δεύτερον ο χαρακτηρισμός ανά εικονοστοιχείο περιοχών της εικόνας από ειδικούς είναι πολλές φορές υποκειμενικός. Για τους δύο αυτούς λόγους, η ποιοτική ανάλυση των αποτελεσμάτων ενός συστήματος σε συγκεκριμένα καρέ μπορεί να δώσει επιπρόσθετη χρήσιμη πληροφορία για την απόδοση του συστήματος, αλλά και για την ποιότητα της χρησιμοποιηθείσας βάσης δεδομένων.

Ένα τρίτο και πάρα πολύ σημαντικό στοιχείο ενός ΠΣΥΟ, που πρέπει να αξιολογηθεί ξεχωριστά, είναι η ταχύτητά του. Τα ΠΣΥΟ απαιτείται να λειτουργούν σε πραγματικό χρόνο και να παρέχουν τα αποτελέσματά τους με την μικρότερη δυνατή καθυστέρηση. Με δεδομένο ότι οι κάμερες που χρησιμοποιούνται για τέτοιες εφαρμογές έχουν μια μέση συχνότητα των 25 καρέ το δευτερόλεπτο και ότι τα επιβατηγά οχήματα ταξιδεύουν με ταχύτητες που καλύπτονται από αυτή την συχνότητα, ένας πρώτος στόχος επεξεργαστικής συχνότητας των 25 καρέ το δευτερόλεπτο για ένα ΠΣΥΟ είναι λογικός. Σε συνθήκες οδήγησης σε πόλη με μικρές ταχύτητες, αυτός ο στόχος μπορεί να μειωθεί, ενώ για κάποιους αυτοκινητοδρόμους που επιτρέπουν στους οδηγούς να κινηθούν με πολύ μεγάλες ταχύτητες, ο στόχος αυξάνεται.

5. Αναγνώριση Σημάτων Οδικής Κυκλοφορίας

Το πρώτο σύστημα που αναπτύχθηκε στα πλαίσια της διατριβής είχε στόχο την ΑΣΟΚ μέσα από κινούμενο όχημα, με χρήση πληροφορίας που προέρχεται από βιντεοκάμερα προσαρμοσμένη στο παρμπρίζ του οχήματος.

Το σύστημα ΑΣΟΚ χωρίζεται σε τέσσερα βασικά στάδια: το στάδιο προ-επεξεργασίας της εικόνας με στόχο τον εντοπισμό χρωμάτων που χρησιμοποιούνται σε ΣΟΚ, το στάδιο εντοπισμού των υποψηφίων ΣΟΚ που βασίζεται στην συμμετρία τους, το στάδιο παρακολούθησης των ΣΟΚ σε συνεχόμενα καρέ και το στάδιο ταξινόμησης των ΣΟΚ.

Για την προ-επεξεργασία των καρέ της βιντεοσκόπησης χρησιμοποιείται μια μέθοδος διαχωρισμού των χρωμάτων της εικόνας σε επιμέρους χρωματικούς χάρτες, βασισμένη στον χρωματικό χώρο CIE-L*a*b*. Ο διαχωρισμός ακολουθείται από μια διαδικασία κατωφλίωσης κάθε ενός από τους χρωματικούς χάρτες με χρήση της μεθόδου κατωφλίωσης του Otsu [10].

Στην συνέχεια οι κατωφλιωμένες εικόνες περνάνε από έναν γρήγορο μετασχηματισμό ακτινικής συμμετρίας (ΓΜΑΣ) [11] για να εντοπιστούν τα κέντρα συμμετρικών σχημάτων με έντονα χρώματα, ένα χαρακτηριστικό γνώρισμα των ΣΟΚ. Ο εντοπισμός των ΣΟΚ για κάθε καρέ συμπληρώνεται από μια διαδικασία ταυτοποίησης του σχήματος του ΣΟΚ, καθώς εκτός από αυστηρή ακτινική συμμετρία (για κυκλικά σχήματα), ο ΓΜΑΣ εντοπίζει και άλλα συμμετρικά σχήματα.

Τα εντοπισμένα υποψήφια ΣΟΚ σε ένα καρέ παρακολουθούνται στα επόμενα καρέ της βιντεοσκόπησης, με μια μέθοδο η οποία απομονώνει την περιοχή γύρω από το εντοπισμένο σχήμα και ψάχνει στην ίδια περιοχή για ένα λίγο μεγαλύτερο συμμετρικό, όμοιο σχήμα στο επόμενο καρέ.

Το τελευταίο βήμα είναι η ταξινόμηση των εντοπισμένων ΣΟΚ με βάση έναν απλό έλεγχο της κανονικοποιημένης ετεροσυσχέτισης του εντοπισμένου υποψηφίου ΣΟΚ με κάθε ένα από τα πρωτότυπα μοντέλα ΣΟΚ του κώδικα οδικής κυκλοφορίας. Δύο διαφορετικοί συνδυασμοί των τιμών ετεροσυσχέτισης στα τρία κανάλια του CIE-L*a*b* ορίζουν δύο ταξινομητές με χρήση των οποίων γίνεται η ταξινόμηση των ΣΟΚ. Ο ένας δουλεύει καλύτερα για μπλε σήματα και ο άλλος για κόκκινα.

Το σύστημα ΑΣΟΚ εξετάστηκε και αξιολογήθηκε σε διάφορες συνθήκες, μεταξύ των οποίων οδήγηση κατά τις πρωινές ώρες με ηλιόλουστο καιρό, συννεφιασμένο καιρό, βροχή, αλλά και νυχτερινή οδήγηση. Τα αποτελέσματα εντοπισμού κρίθηκαν πάρα πολύ ικανοποιητικά από πλευράς πληρότητας, αλλά δεν έχουν αρκετά μεγάλη ορθότητα. Το στάδιο παρακολούθησης συμβάλλει θετικά στην αύξηση της ορθότητας των εντοπισμών ΣΟΚ. Το υποσύστημα ταξινόμησης πάσχει σε απόλυτα ποσοστά επιτυχίας σε σχέση με άλλα της βιβλιογραφίας, ειδικά σε αντίξοες συνθήκες, αλλά σημαντικό ρόλο σε αυτό παίζει η μεγάλη βάση ΣΟΚ που χρησιμοποιείται στο σύστημά μας.

6. Εντοπισμός Φωτεινών Σηματοδοτών σε Αντίξοες Συνθήκες με Χρήση Χρώματος, Συμμετρίας και Χωροχρονικής Πληροφορίας

Το δεύτερο σύστημα που αναπτύχθηκε στα πλαίσια της διατριβής είχε ως στόχο τον εντοπισμό ΦΣ από βιντεοσκοπήσεις που λαμβάνονται από βιντεοκάμερα τοποθετημένη στο παρμπρίζ κινούμενου οχήματος.

Το σύστημα που αναπτύχθηκε βασίζεται σε τρία διαδοχικά στάδια επεξεργασίας. Το πρώτο στάδιο είναι αυτό της προ-επεξεργασίας εικόνας, όπου κάθε καρέ μετατρέπεται από RGB σε CIE-L*a*b* και στην συνέχεια ακολουθείται μια διαδικασία ενίσχυσης της διαφοράς μεταξύ του κόκκινου και του πράσινου χρώματος, ειδικά για περιοχές με υψηλή τιμή φωτεινότητας (όπως είναι τα φανάρια). Αφού οι εικόνες αυτές αναμιχθούν με τις αντίστοιχες της ενισχυμένης διαφοράς κίτρινου-μπλε, ακολουθεί μια διαδικασία μορφολογικού γεμίσματος τρυπών για την καταπολέμηση του φαινομένου που είναι γνωστό ως "blooming effect".

Το δεύτερο στάδιο επεξεργασίας αποτελείται από έναν ΓΜΑΣ για τον εντοπισμό ακτινικά συμμετρικών σχημάτων διαφόρων ακτινών μέσα στην εικόνα, το οποίο ακολουθείται από την απομόνωση των μεγίστων/ελαχίστων τιμών του αποτελέσματος του ΓΜΑΣ. Τα μέγιστα/ελάχιστα υποδηλώνουν πιθανή ύπαρξη κόκκινου/πράσινου ΦΣ αντίστοιχα στα συγκεκριμένα σημεία.

Για να επαληθευθεί ή να απορριφθεί η ύπαρξη ΦΣ στα ύποπτα σημεία, ακολουθεί μια διαδικασία ελέγχου της χωροχρονικής επιμονής των υποψηφίων ΦΣ. Πιο συγκεκριμένα, για να εγκριθεί κάποιο υποψήφιο κέντρο ΦΣ, θα πρέπει να εμφανίζεται σε μια μικρή γειτονία του κάδρου για έναν ικανό αριθμό από συνεχόμενα καρέ.

Το προτεινόμενο σύστημα αξιολογήθηκε ποσοτικά στην βιντεοσκόπηση LARA. Τα αποτελέσματα έδειξαν ότι είναι άμεσα συγκρινόμενο με αντίστοιχα συστήματα της βιβλιογραφίας και μάλιστα χωρίς να χρησιμοποιεί κάποιο μορφολογικό μοντέλο των ΦΣ, ούτε κάποιον μηχανισμό εκπαίδευσης. Πιο συγκεκριμένα, η μέθοδος που προτείνεται επιτυγχάνει αποτελέσματα πληρότητας που αγγίζουν το 94%, αλλά με ορθότητα περίπου 61%.

Τα ποιοτικά αποτελέσματα που παρατηρήθηκαν σε πλάνα που συλλέχθηκαν από το Internet (YouTube), σε διάφορες συνθήκες και περιβάλλοντα οδήγησης, έδειξαν ότι το προτεινόμενο σύστημα είναι αρκετά αξιόπιστο και σε συνθήκες οδήγησης υπό βροχή, αλλά και την νύχτα. Το πιο σημαντικό του μειονέκτημα παρατηρείται σε νυχτερινή οδήγηση σε κέντρα μεγαλουπόλεων, όπου υπάρχουν πολλές φωτεινές πηγές που μπορεί να οδηγήσουν σε σύγχυση το σύστημα εντοπισμού υποψηφίων ΦΣ.

7. Εντοπισμός Προπορευόμενου Οχήματος από μια Εικόνα

Το τρίτο σύστημα που αναπτύχθηκε στα πλαίσια της παρούσας διατριβής είχε στόχο τον εντοπισμό προπορευόμενου οχήματος με χρήση πληροφορίας που προέρχεται από μια εικόνα τραβηγμένη από το εσωτερικό ενός κινούμενου οχήματος.

Το προτεινόμενο σύστημα εντοπισμού των προπορευόμενων οχημάτων βασίζεται σε μεθόδους που εκμεταλλεύονται την εκ των προτέρων γνώση για την εξαγωγή των πιο χαρακτηριστικών γνωρισμάτων των οχημάτων. Η βασική πληροφορία εξάγεται από τις υποψήφιες περιοχές πίσω φαναριών, για τον εντοπισμό των οποίων χρησιμοποιείται χρωματική κατάτμηση της εικόνας στον CIE-L*a*b* χρωματικό χώρο, καθώς και πληροφορία ακτινικής συμμετρίας με βάση τον ΓΜΑΣ.

Τα πίσω φανάρια αποτελούν ένα εμφανές χαρακτηριστικό γνώρισμα για τον ΕΠΟ. Ξέχωρα από το γεγονός ότι αποτελούν ένα κοινό γνώρισμα όλων των οχημάτων, λόγω νομοθεσίας, είναι επίσης εμφανή κάτω από διαφορετικές συνθήκες φωτισμού, καιρού, καθώς και από την ώρα της ημέρας. Επιπροσθέτως, μπορούν να χρησιμοποιηθούν για να παράγουν μια πρώιμη ειδοποίηση για πιθανό κίνδυνο, αφού τα αναμμένα πίσω φανάρια υποδηλώνουν ότι το προπορευόμενο όχημα επιβραδύνει.

Το δεύτερο κομμάτι του προτεινόμενου συστήματος περιλαμβάνει μορφολογικό ταίριασμα των ανιχνευμένων υποψήφιων φαναριών με μια ανίχνευση οριζόντιων ακμών για να οριστούν περιοχές της εικόνας που πιθανώς περιέχουν οχήματα. Ακολούθως, ένας έλεγχος αξονικής συμμετρίας κατά την κάθετη μεσοκάθετο χρησιμοποιείται για την επαλήθευση της ύπαρξης οχήματος. Ο έλεγχος αξονικής συμμετρίας έχει βασιστεί στον συνδυασμό δύο μεγεθών σύγκρισης της ομοιότητας μεταξύ δύο εικόνων (στην περίπτωση αυτή οι δύο εικόνες είναι το αριστερό και το δεξί κομμάτι της περιοχής του υποψηφίου οχήματος), της Δομικής Ομοιότητας (Structural SIMilarity - SSIM) [12] και του Μέσου Απόλυτου Σφάλματος (Mean Absolute Error - MAE). Για τα επιτυχώς εντοπισμένα οχήματα πραγματοποιείται στην συνέχεια μια εκτίμηση της απόστασής τους από το όχημα του συστήματος.

Το σύστημα ΕΠΟ από μια εικόνα αξιολογήθηκε σε δύο βάσεις δεδομένων με εικόνες προπορευόμενων οχημάτων, την Caltech 1999 και την Caltech 2001. Επίσης εξετάστηκε η απόδοσή του στα καρέ της βιντεοσκόπησης LARA που περιείχαν κάποιο προπορευόμενο όχημα. Τα ποσοτικά αποτελέσματα σε αυτές τις βάσεις έδειξαν ότι το σύστημα δουλεύει πολύ καλά από άποψης πληρότητας, φτάνοντας ποσοστά που αγγίζουν το 93.6%. Όταν δε το προπορευόμενο όχημα φρενάρει, οπότε τα πίσω φώτα του είναι αναμμένα, το ποσοστό εκτινάσσεται σε τιμές που φτάνουν το 99.2%.

Το σύστημα ΕΠΟ αξιολογήθηκε και ποιοτικά σε πλάνα που συλλέχθηκαν από το Internet (YouTube), σε διάφορες συνθήκες και περιβάλλοντα οδήγησης. Τα ποιοτικά αυτά αποτελέσματα έδειξαν ότι το σύστημα είναι αρκετά ανθεκτικό και σε συνθήκες βροχόπτωσης. Τα μεγαλύτερα προβλήματα τα αντιμετωπίζει σε πολύ δύσκολα περιβάλλοντα οδήγησης, όπως για παράδειγμα την οδήγηση σε νυχτερινά περιβάλλοντα με πολύ κίνηση, όπου τα κόκκινα φώτα πολλαπλασιάζονται στην σκηνή.

8. Εντοπισμός Προπορευόμενου Οχήματος με Χρήση Βίντεο

Το σύστημα εντοπισμού προπορευόμενου οχήματος που παρουσιάστηκε στην προηγούμενη ενότητα πάσχει από το ότι χρησιμοποιεί μόνο πληροφορία από ένα καρέ (ή φωτογραφία). Το γεγονός αυτό περιορίζει τις δυνατότητες του συστήματος και προκαλεί αύξηση των εσφαλμένα θετικών εντοπισμών, με αποτέλεσμα την μείωση του ποσοστού ορθότητας των εντοπισμών.

Μια εξέλιξη του συστήματος ώστε να εκμεταλλεύεται χωροχρονική πληροφορία προερχόμενη από βίντεο έχει επίσης προταθεί στα πλαίσια της διατριβής. Αυτή την φορά το πρώτο υποσύστημα που στοχεύει σε εντοπισμό των πίσω φαναριών του προπορευόμενου οχήματος έχει παραλλαχθεί ελαφρά και χρησιμοποιεί ένα υβριδικό κανάλι χρωματικής πληροφορίας που εξάγεται από μια ανάμιξη των δύο από τα τρία κανάλια του $c_1c_2c_3$ χρωματικού χώρου [13]. Το κανάλι αυτό υφίσταται ένα φιλτράρισμα μεσαίου, μεγέθους 3x3 και εν συνεχεία υπολογίζεται ο ΓΜΑΣ του για διάφορες ακτίνες. Το αποτέλεσμα του ΓΜΑΣ ελέγχεται και εντοπίζονται τα τοπικά του μέγιστα πάνω από ένα κατώφλι, τα οποία ορίζονται ως υποψήφια φανάρια οχημάτων.

Στην συνέχεια όλα τα πιθανά ζεύγη υποψηφίων φαναριών εξετάζονται ως προς την μεταξύ τους απόσταση και γωνία και ανάλογα με το αποτέλεσμα απορρίπτονται, ή εγκρίνονται για περαιτέρω έλεγχο. Ο έλεγχος αυτός έχει δύο στάδια. Το πρώτο στάδιο αφορά στην στατική επαλήθευση των υποψηφίων ζευγών και βασίζεται στην Δομική Ομοιότητα του δεξιού και αριστερού κομματιού μιας ορθογώνιας παραλληλόγραμμης περιοχής που περιλαμβάνει τα δύο υποψήφια φανάρια.

Το στάδιο αυτό πάσχει από το συχνό φαινόμενο της επαλήθευσης πολλών περιοχών ανάμεσα σε δύο φανάρια, λόγω του ελέγχου πολλών ακτινών στον ΓΜΑΣ. Για να περιοριστεί το φαινόμενο αυτό ακολουθεί ένα δεύτερο, δυναμικό, στάδιο επαλήθευσης. Σε αυτό χρησιμοποιούνται τα αποτελέσματα ενός φίλτρου Kalman με εισόδους τις συντεταγμένες x και y του οχήματος καθώς και το πλάτος του, W. Η πρόβλεψη της νέας θέσης και του μεγέθους του υποψηφίου οχήματος συγκρίνεται με το αποτέλεσμα του στατικού εντοπισμού και αν είναι κοντά, το αποτέλεσμα του εντοπισμού γίνεται δεκτό. Ειδάλλως, ακολουθεί μια διαδικασία εντοπισμού υποψηφίου οχήματος εντός της περιοχής που έχει προβλεφθεί από το φίλτρο. Αν δεν βρεθεί ούτε εκεί κάποιο υποψήφιο όχημα, το ΠΟ θεωρείται "χαμένο" και η διαδικασία παρακολούθησης με το φίλτρο Kalman σταματάει μέχρι να εντοπισθεί ξανά κάποιο υποψήφιο όχημα με την στατική διαδικασία.

Το σύστημα ΕΠΟ με χρήση βίντεο αξιολογήθηκε τόσο με τις στατικές σκηνές Caltech 1999 και 2001, όσο και με την βιντεοσκόπηση LARA, αλλά και πέντε ακόμα βιντεοσκοπήσεις οδήγησης σε διάφορες συνθήκες (HRI1-HRI5). Τα ποσοτικά στατιστικά στις στατικές σκηνές (χωρίς την χρήση της παρακολούθησης) φτάσανε σε πληρότητα το 94.2%, με ορθότητα 95.1%. Τα ποσοστά σε εικόνες οχημάτων που φρενάρουν ανήλθαν σε 97,7% και 98% αντίστοιχα. Τα ποσοστά αυτά είναι ανώτερα από αντίστοιχα της βιβλιογραφίας.

Στην βιντεοσκόπηση LARA ωστόσο, το προτεινόμενο σύστημα έδωσε μέτρια αποτελέσματα χωρίς την χρήση του υποσυστήματος παρακολούθησης. Πιο συγκεκριμένα, η πληρότητα δεν ξεπέρασε το 72.5% με ορθότητα 73.6%. Με την χρήση του υποσυστήματος παρακολούθησης, τα ποσοστά αυτά εκτοξεύτηκαν σε 93.1% και 94.4% αντίστοιχα, επιβεβαιώνοντας την επιτυχία της μεθόδου.

Τα ποιοτικά αποτελέσματα στις βιντεοσκοπήσεις HRI1-HRI5 έδειξαν ότι το σύστημα που προτείνεται είναι ανθεκτικό και σε διαφορετικές συνθήκες καιρού και περιβάλλοντα οδήγησης. Πιο συγκεκριμένα, το σύστημα ανταπεξήλθε ικανοποιητικά τόσο σε οδήγηση υπό βροχή και χιόνι, όσο και σε συνθήκες νυχτερινής οδήγησης.

Τέλος, μελετήθηκε και η επίδραση της ανάλυσης των πλάνων στην απόδοση του συστήματος και βρέθηκε ότι στην ανάλυση 160x120, η ταχύτητα επεξεργασίας είναι περίπου 7 καρέ το δευτερόλεπτο σε Matlab, ταχύτητα που θα μπορούσε να βελτιωθεί περαιτέρω αν ο ΓΜΑΣ υλοποιηθεί σε παράλληλη αρχιτεκτονική, όπως είναι εφικτό.

9. Εντοπισμός δρόμου

Το σύστημα εντοπισμού δρόμου που αναπτύχθηκε στα πλαίσια της διδακτορικής διατριβής βασίζεται στην πληροφορία που λαμβάνεται από μία έγχρωμη κάμερα που τοποθετείται εντός του οχήματος. Η κάμερα τοποθετείται στο επάνω μέρος του παρμπρίζ, έτσι ώστε να μην εμποδίζει τη θέα του οδηγού και να περικλείει στο οπτικό της πεδίο όσο περισσότερο γίνεται από τον δρόμο που βρίσκεται ακριβώς μπροστά από το αυτοκίνητο. Η ακριβής θέση και κλίση της κάμερας είναι σημαντική μόνο για την αρχικοποίηση της εκ των προτέρων γνωστής πληροφορίας, δηλαδή του μέρους της εικόνας που θεωρείται πολύ πιθανό να ανήκει στον δρόμο και του μέρους της εικόνας που θεωρείται σχεδόν απίθανο να ανήκει στον δρόμο.

Τα πλάνα της βιντεοσκόπησης εξετάζονται σε ζεύγη συνεχόμενων καρέ, για την εξαγωγή χωροχρονικής πληροφορίας που θα χρησιμοποιηθεί για τον χαρακτηρισμό κάποιων εικονοστοιχείων ως σπόρων του δρόμου και κάποιων ως σπόρων της υπόλοιπης εικόνας. Οι σπόροι που επιλέγονται από αυτό το στάδιο επικυρώνονται με βάση τα αποτελέσματα του εντοπισμού δρόμου από το προηγούμενο καρέ και στην συνέχεια χρησιμοποιούνται σε έναν Αλγόριθμο Τυχαίου Περιπατητή (ΑΤΠ) για την εκτίμηση των εικονοστοιχείων που ανήκουν στον δρόμο στο τρέχον καρέ.

Στην διαδικασία κατάτμησης χρησιμοποιείται για τον ΑΤΠ η λύση του προβλήματος Dirichlet [14], μια μέθοδος που μειώνει σημαντικά τον απαιτούμενο χρόνο υπολογισμού της λύσης, ενώ δίνει τα ίδια αποτελέσματα με τον κλασικό ΑΤΠ [15], [16]. Πιο συγκεκριμένα, ο αλγόριθμος εντοπισμού δρόμου που αναπτύχθηκε, αποτελείται από τα παρακάτω βήματα:

i. Αρχικά συλλέγεται εκ των προτέρων πληροφορία που αφορά την θέση της κάμερας μέσα στο όχημα, καθώς και την κλίση της. Με χρήση αυτών των δεδομένων μπορούν να τοποθετηθούν οι αρχικοί σπόροι μη-δρόμου (στο πάνω κομμάτι του καρέ).
ii. Στην συνέχεια, διαδοχικά ζεύγη από καρέ στους χρόνους t-1 και t υφίστανται μετατροπή από το RGB στο πρώτο κανάλι του $c_1c_2c_3$ χρωματικού χώρου, για την μεγαλύτερη ανθεκτικότητα σε φωτοσκιάσεις, ή απότομες αλλαγές φωτεινότητας.
iii. Τα c_1 κανάλια των δύο καρέ χρησιμοποιούνται για τον υπολογισμό του μέτρου της οπτικής ροής Horn Schunck [17] στο χρονικό σημείο t, το οποίο από εδώ και στο εξής θα ονομάζεται ροή HSC1.
iv. Η ροή HSC1 κατωφλιώνεται χρησιμοποιώντας την τεχνική Otsu [10] για να διαχωριστούν τα εικονοστοιχεία που πιθανά ανήκουν στον δρόμο (χαμηλή τιμή ροής HSC1) από τα εικονοστοιχεία που δεν ανήκουν στον δρόμο (εμπόδια, με υψηλή τιμή ροής HSC1).
v. Τα εικονοστοιχεία της κατηγορίας με υψηλή τιμή ροής HSC1 που επιλέχθηκαν κατά την κατωφλίωση συνδυάζονται με τα εικονοστοιχεία του εκ των προτέρων γνωστού κομματιού του καρέ που δεν ανήκει στον δρόμο (από το βήμα (i)) και το αποτέλεσμα μας δίνει το σύνολο των σπόρων μη-δρόμου.
vi. Το αποτέλεσμα του εντοπισμού δρόμου από το καρέ t-1 χρησιμοποιείται για να οριστεί ένα τραπέζιο, στα εικονοστοιχεία της περιμέτρου του οποίου τοποθετούνται οι σπόροι της κατηγορίας δρόμος. Αν κάποιο από τα εικονοστοιχεία της περιμέτρου περιέχει ήδη σπόρο μη-δρόμου, τότε κρατάει τον ήδη υπάρχοντα χαρακτηρισμό.
vii. Η εικόνα RGB του καρέ σε χρόνο t υποδειγματοληπτείται για να χρησιμοποιηθεί στον ΑΤΠ με μικρότερη ανάλυση, οδηγώντας σε γρηγορότερο υπολογισμό του αποτελέσματος.
viii. Η εικόνα που προκύπτει από το βήμα (vii), καθώς και τα σύνολα σπόρων δρόμου και μη-δρόμου που προέκυψαν από τα βήματα (vi) και (v) αντίστοιχα, χρησιμοποιούνται ως είσοδοι στον ΑΤΠ, για την εκτίμηση των πιθανοτήτων κάθε εικονοστοιχείου του καρέ της χρονικής στιγμής t που δεν περιέχει σπόρο κάποιας κατηγορίας, να ανήκει στον δρόμο.

ix. Ο χάρτης πιθανοτήτων που προκύπτει από το βήμα (viii) κατωφλιώνεται και το τελικό αποτέλεσμα υπερδειγματοληπτείται για να επιστρέψει στην αρχική ανάλυση.

Το σύστημα ΕΔ που προτείνεται εδώ αξιολογήθηκε ποσοτικά με διάφορες βιντεοσκοπήσεις, όπως είναι τα πέντε βίντεο που αποτελούν την DiploDoc, καθώς και τα δύο βίντεο Alvarez1 και Alvarez2. Επίσης, ποιοτική αξιολόγηση της μεθόδου έγινε και σε βιντεοσκοπήσεις που ελήφθησαν στα πλαίσια της διατριβής σε διάφορες συνθήκες και περιβάλλοντα οδήγησης (Patras1, Patras2, Ioann1).

Τα ποσοτικά αποτελέσματα που επιτεύχθηκαν με την χρήση των προαναφερθέντων βιντεοσκοπήσεων απέδειξαν την ανωτερότητα της προτεινόμενης μεθόδου σε σχέση με αντίστοιχες της βιβλιογραφίας. Πιο συγκεκριμένα, στις βιντεοσκοπήσεις DiploDoc, η προτεινόμενη μέθοδος πετυχαίνει ποσοστά ποιότητας της τάξης του 93%, ενώ άλλες μέθοδοι περιορίζονται κάτω από το 90%. Επίσης, στις βιντεοσκοπήσεις Alvarez1 και Alvarez2, η προτεινόμενη μέθοδος φτάνει με μικρές προσαρμογές σε αποτελεσματικότητα της τάξης του 0.92, έναντι 0.89 άλλων μεθόδων.

Η επιτυχία της μεθόδου γίνεται ορατή και από την ποιοτική αξιολόγησή της σε βιντεοσκοπήσεις κάτω από αντίξοες συνθήκες, όπως βροχόπτωση, έντονες φωτοσκιάσεις, νυχτερινή οδήγηση κ.α. Το μεγαλύτερο πλεονέκτημά της είναι ότι προσαρμόζεται σε απότομες αλλαγές φωτεινότητας σχετικά γρήγορα, χωρίς να χρησιμοποιεί κάποιο μοντέλο δρόμου, ή κάποια διαδικασία εκπαίδευσης.

Τέλος, το σύστημα ΕΔ που αναπτύχθηκε εξετάστηκε και ως προς την ιδανική ανάλυση βιντεοσκόπησης που πρέπει να χρησιμοποιεί, τόσο για την καλύτερη απόδοση του συστήματος, όσο και για την ταχύτερη επεξεργασία των καρέ. Η πειραματική διαδικασία ανέδειξε ως καλύτερη επιλογή την ανάλυση 160x120 που δίνει ταχύτητα επεξεργασίας περίπου 10 καρέ το δευτερόλεπτο, σε Matlab, ένα νούμερο που είναι αρκετά ελπιδοφόρο για πιθανή υλοποίηση σε πραγματικό χρόνο με χρήση DSP.

10. Συμπεράσματα

Η παρούσα διδακτορική διατριβή έχει συμβάλλει ποικιλοτρόπως στην έρευνα των ΠΣΥΟ που βασίζονται αποκλειστικά σε οπτική πληροφορία. Ξεκινώντας από βασικές μεθόδους επεξεργασίας χρωματικής πληροφορίας, κατάτμησης εικόνας, εντοπισμού συμμετριών και παρακολούθησης αντικειμένων, αναπτύχθηκαν νέες τεχνικές που εξειδικεύονται σε προβλήματα που συναντώνται σε σκηνές οδήγησης και προσαρμόστηκαν έτσι ώστε να βελτιστοποιήσουν τον συνδυασμό απόδοσης-ταχύτητας που είναι απαραίτητος σε τέτοια συστήματα.

Στα πλαίσια της διατριβής προτάθηκαν νέες τεχνικές χρωματικής κατάτμησης εικόνων, βασισμένες σε ήδη υπάρχοντες, ή υβριδικούς χρωματικούς χώρους και σε κλασικές μεθόδους κατωφλίωσης. Επίσης, μελετήθηκαν και προτάθηκαν τεχνικές αξιοποίησης της ακτινικής συμμετρίας για τον εντοπισμό περιοχών της εικόνας που μπορεί να ανήκουν σε ΣΟΚ, ΦΣ, ή φώτα οχημάτων. Η ακτινική συμμετρία συμπληρώθηκε με πρωτότυπες μεθόδους ελέγχου αξονικής συμμετρίας που εκμεταλλεύονται τεχνικές ελέγχου ομοιότητας εικόνων. Επίσης, εξετάστηκαν και αναπτύχθηκαν τεχνικές επαλήθευσης των στατικών αποτελεσμάτων εντοπισμού ΣΟΚ, ΦΣ, ή προπορευόμενων οχημάτων με χρήση χωροχρονικής πληροφορίας, είτε με πρωτότυπες μεθόδους, είτε με παραλλαγές υπαρχόντων μεθόδων, όπως τα φίλτρα Kalman. Τέλος, αναπτύχθηκε παραλλαγμένη, αυτόματη μέθοδος κατάτμησης εικόνας βασισμένη στον ΑΤΠ, για τον αξιόπιστο ΕΔ. Η μέθοδος χρησιμοποιεί χωροχρονική, καθώς και εκ των προτέρων πληροφορία για την επιτυχημένη επιλογή σπόρων για την αρχικοποίηση του ΑΤΠ.

Σε πρακτικό επίπεδο, μελετήθηκαν τα προβλήματα των ΠΣΥΟ σε αντίξοες συνθήκες και απαιτητικά περιβάλλοντα οδήγησης και εξήχθησαν χρήσιμα συμπεράσματα για την φύση των πιο σοβαρών αντιξοοτήτων και για πιθανούς τρόπους αντιμετώπισής τους. Επίσης, έγινε εκτενής μελέτη για την επίδραση της ανάλυσης των εικόνων που χρησιμοποιούνται σε τέτοιες εφαρμογές, τόσο για την βέλτιστη επίδοση από πλευράς ποιότητας, όσο και για την ταχύτερη δυνατή επεξεργασία των καρέ. Στα πλαίσια αυτής της μελέτης εξήχθησαν και κάποια μη-διαισθητικά αποτελέσματα, όπως το ότι κάποιες φορές η μείωση της ανάλυσης οδηγεί στην αύξηση της ποιότητας των αποτελεσμάτων (π.χ. στον ΕΔ), μέχρις ενός σημείου. Τέλος, συζητήθηκε και αναλύθηκε η ανάγκη ανάπτυξης βάσεων δεδομένων με επισημασμένες βιντεοσκοπήσεις σε διάφορες συνθήκες, για την καλύτερη και δικαιότερη σύγκριση των αναπτυσσόμενων ΠΣΥΟ. Για αυτό τον λόγο και προτιμήθηκε να γίνει ποσοτική ανάλυση των αποτελεσμάτων της διατριβής μόνο σε τέτοιες βάσεις, όπου βέβαια ήταν αυτό εφικτό.

Μελλοντικές κατευθύνσεις της έρευνας που διενεργήθηκε στα πλαίσια αυτής της διατριβής περιλαμβάνουν την σύνδεση διαφόρων από τις μεθόδους που αναπτύχθηκαν, με στόχο την ανάπτυξη ενός πλήρους ΠΣΥΟ, αλλά και την βελτίωση της απόδοσης των επιμέρους συστημάτων με χρήση της πληροφορίας που θα λαμβάνεται από τα υπόλοιπα συστήματα. Επίσης θα μελετηθεί η πιθανότητα υλοποίησης κάποιων από τις μεθόδους σε αρχιτεκτονικές που επιτρέπουν την λειτουργία τους σε πραγματικό χρόνο. Τέλος, θα επιχειρηθεί η χρήση συστοιχιών καμερών και ο συνδυασμός της πληροφορίας τους για βελτίωση των αποτελεσμάτων των ΠΣΥΟ που έχουν ήδη αναπτυχθεί.

Βιβλιογραφία

[1] M. Peden, R. Scurfield, D. Sleet, D. Mohan, A.A. Hyder, E. Jarawan, and C. Mathers, “WHO | World report on road traffic injury prevention,” *WHO*, 2004. [Online]. Available: http://www.who.int/violence_injury_prevention/publications/road_traffic/world_report/en/index.html. [Accessed: 04-Nov-2012].

[2] E. Kopits and M. Cropper, “Traffic fatalities and economic growth,” *Accident Analysis & Prevention*, vol. 37, no. 1, pp. 169–178, 2005.

[3] C. J. L. Murray and A. D. Lopez, “The global burden of disease and injury series, volume 1: a comprehensive assessment of mortality and disability from diseases, injuries, and risk factors in 1990 and projected to 2020,” *Cambridge. MA*, 1996.

[4] U.S. Departmetnt of Transportation, “National Motor Vehicle Crash Causation Survey: Report to Congress,” National Highway Traffic Safety Administration, Jul. 2008.

[5] “Autonomous Emergency Braking - AEB | Euro NCAP - For safer cars crash test safety rating.” [Online]. Available: http://www.euroncap.com/results/aeb.aspx. [Accessed: 17-Nov-2012].

[6] Robotics Centre of Mines ParisTech, “Traffic Lights Recognition (TLR) public benchmarks [La Route Automatisée],” 01-May-2010. [Online]. Available: http://www.lara.prd.fr/benchmarks/trafficlightsrecognition. [Accessed: 06-Oct-2011].

[7] P. Lombardi, M. Zanin, and S. Messelodi, “Unified stereovision for ground, road, and obstacle detection,” in *IEEE Intelligent Vehicles Symposium, 2005. Proceedings*, 2005, pp. 783 – 788.

[8] P. Lombardi, M. Zanin, and S. Messelodi, “Switching models for vision-based on-board road detection,” in *2005 IEEE Intelligent Transportation Systems, 2005. Proceedings*, 2005, pp. 67 – 72.

[9] J. M. Á. Alvarez and A. M. Lopez, “Road Detection Based on Illuminant Invariance,” *IEEE Trans. Intell. Transport. Syst.*, vol. 12, no. 1, pp. 184–193, Mar. 2011.

[10] N. Otsu and others, “A threshold selection method from gray-level histograms,” *IEEE Transactions on systems, Man, and Cybernetics*, vol. 9, no. 1, pp. 62–66, 1979.

[11] G. Loy and A. Zelinsky, “Fast radial symmetry for detecting points of interest,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no. 8, pp. 959 – 973, Aug. 2003.

[12] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, “Image quality assessment: from error visibility to structural similarity,” *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600 –612, Apr. 2004.

[13] T. Gevers and W. M. Smeulders, “Color based object recognition,” *Pattern recognition*, vol. 32, no. 3, pp. 453–464, 1999.

[14] L. Grady, “Random walks for image segmentation,” *IEEE transactions on pattern analysis and machine intelligence*, pp. 1768–1783, 2006.

[15] S. Kakutani, “Markov processes and the Dirichlet problem,” 1945, vol. 21, pp. 227–233.

[16] P. G. Doyle and J. L. Snell, “Random walks and electric networks,” *Washington D.C.: Mathematical Association of America*, vol. 22, 1984.

[17] B. K. Horn and B. G. Schunck, “Determining optical flow,” *Artificial intelligence*, vol. 17, no. 1–3, pp. 185–203, 1981.

CERTIFICATION

This is to certify that GEORGIOS SIOGKAS, who received his Diploma in Electrical and Computer Engineering from the University of Patras, defended his Doctoral dissertation, entitled

"ADVANCED DRIVER ASSISTANCE SYSTEMS WITH COMPUTER VISION METHODS"

in public on the 17th of June, 2013, at 11:00, in the Department of Electrical and Computer Engineering, at the University of Patras. The Doctoral dissertation was examined and approved by the following examination committee:

Kostantinos Berberidis, Professor in the Computer Engineering and Informatics Department of the University of Patras

Evangelos Dermatas, Associate Professor in the Electrical and Computer Engineering Department of the University of Patras, principal supervisor, head of the 3-supervising committee

Georgios Economou, Associate Professor in the Physics Department of the University of Patras

Nikolaos Fakotakis, Professor in the Electrical and Computer Engineering Department of the University of Patras

Emmanouil Psarakis, Assistant Professor in the Computer Engineering and Informatics Department of the University of Patras

Antonios Tzes, Professor in the Electrical and Computer Engineering Department of the University of Patras, member of the 3-supervising committee

Patras, 17 of June, 2013

Principal Supervisor	Deputy Head of the Electrical and Computer Engineering Department
E. Dermatas Assoc. Professor	G. Giannakopoulos Professor

ACKNOWLEDGEMENTS

This PhD thesis has turned out to be much more demanding than I had first imagined it to be, both in terms of the hard work involved and in terms of the perseverance needed. Having successfully completed it, I feel the need to thank certain people for their contribution to this important project.

First of all, I want to thank my supervisor, Associate Professor Evangelos Dermatas, who believed in me and supported my application for a PhD position, even at times when there were not so many people that thought I could make it. Without his support, this PhD would never have started in the first place.

I would also like to thank all the members of the examination committee, and especially to the members of the supervising committee Professors Anthony Tzes and Thanos Stouraitis, for contributing to the enhancement of the thesis with their insightful comments and recommendations. The comments from all of them on my work at the end of the thesis defense were also very inspiring and encouraging and will guide me in my future endeavors.

A special thanks goes to all my colleagues in the Pattern Recognition Lab and more notably to Giannis Chatzis, Marios Vlachos and Giorgos Apostolopoulos, who were my "brothers in arms" during the first few years of my PhD. A special reference should be made to Evangelos Skodras, who proved to be a very important research collaborator, both in his undergraduate and his postgraduate years. His contribution was invaluable and I would like to sincerely thank him for this.

It would be a shame not to include in my acknowledgements two fellow PhD students from a different field, namely Aris Bloutsos and Dimitris Baros. Together, we formed a "postgraduate support group", which provided us all with the necessary endurance to complete our studies. Two of us made it to the end; the third will soon follow.

I also have to mention the important contribution of two people that will soon be officially considered my family; Kate Makri and Grigoris Chiotakis. They opened up their home, showed great care and furthermore they truly believed in me. If only they were accompanied by two people that left early, before I got to know them; they were Nikos and George Makris, may they rest in peace.

It goes without saying that this demanding project would not start at all if it weren't for my parents, who provided me with the necessary physical and mental supplies throughout my life. They have always stood by me and continue to do so. Their never-ending love and support mean the world to me. In this category I would like to include my younger brother, Panagiotis who has been my friend and roommate for a large part of my life and provided great moral support throughout my PhD years.

Finally, I am lost for words when it comes to thanking my soon to be wife, Maro. She is my continuous source of love, inspiration and encouragement. It would be doubtful if this adventurous journey of knowledge would reach a happy ending without her.

Contents

Abstract .... **V**

Περίληψη .... **VII**

Εκτεταμένη Περίληψη στα Ελληνικά .... **IX**

Contents .... **i**

List of Figures .... **v**

List of Tables .... **xi**

Glossary of Acronyms .... **xiii**

Glossary of Mathematical Symbols .... ***xv***

Chapter 1 .... ***1***

Advanced Driver Assistance Systems (ADAS) **1**

- 1.1 Introduction 1
- 1.2 The need for ADAS development 1
- 1.3 Vision based subsystems used in commercial solutions 3
  - 1.3.1 Traffic Sign Recognition 4
  - 1.3.2 Road Lane Recognition 5
  - 1.3.3 Vehicle Detection 6
  - 1.3.4 Driver Drowsiness Detection 7
  - 1.3.5 Pedestrian Detection Systems 8
  - 1.3.6 Night Vision Systems 8
- 1.4 Vision based subsystems still in research stage 9
  - 1.4.1 Traffic Lights Recognition 9
  - 1.4.2 Driver Gaze Detection 9
  - 1.4.3 Road Detection 9
- 1.5 General structure of a computer vision based ADAS 10
- 1.6 Datasets used for system development and evaluation 10
  - 1.6.1 Video streams acquired for qualitative evaluation of thesis methods 11
  - 1.6.2 Video streams used for quantitative evaluation of thesis methods 11
  - 1.6.3 Publicly available video streams used for qualitative evaluation 12
  - 1.6.4 Publicly available image sequences used for quantitative evaluation 13
- 1.7 Performance evaluation measures 13
  - 1.7.1 Quantitative measures for quality evaluation 13
  - 1.7.2 Qualitative performance evaluation 14
  - 1.7.3 Processing speed performance assessment 14
- 1.8 Conclusions 14

Chapter 2 .... ***15***

Computer Vision Algorithms 15
2.1 Introduction 15
2.2 Color spaces and illumination invariance 15
2.2.1 Comparison of color spaces for real world scenes 15
2.3 Fast Radial Symmetry for efficient detection of symmetrical objects 17
2.4 Otsu thresholding for efficient bimodal image segmentation 19
2.5 Optical Flow for moving objects detection 21
2.6 Discrete Time Kalman Filter for object tracking 23
2.6.1 Definition of the discrete process model 24
2.6.2 Algorithm of the Discrete Kalman Filter 24
2.7 Random Walker for efficient image segmentation 26
2.8 Measures for object similarity assessment 28
2.8.1 Normalized Cross-Correlation (NCC) 28
2.8.2 Mean Absolute Error (MAE) 29
2.8.3 Structural Similarity Index (SSIM) 29
2.9 Conclusions 29

Chapter 3* *31

Traffic Sign Recognition System 31
3.1 Introduction 31
3.2 The Traffic Sign Recognition problem 31
3.3 Related work 33
3.4 TSD system structure 33
3.4.1 Color Space Selection - Color Segmentation 34
3.4.2 Symmetry Detection 35
3.4.3 Center Localization – Shape Determination 38
3.5 Traffic Sign Tracking 39
3.6 Traffic Sign Classification (TSC) 41
3.7 Experimental results 42
3.7.1 Traffic Sign Detection results 42
3.7.2 Traffic Sign Classification results 44
3.8 Conclusions 45

Chapter 4* *47

Traffic Light Recognition 47
4.1 Introduction 47
4.2 Literature review 47
4.3 TLR system structure 49
4.3.1 Image pre-processing 50
4.3.2 Radial Symmetry Detection 53
4.3.3 TL candidates verification 54
4.4 Experimental results 55
4.4.1 Quantitative results 56
4.4.2 Qualitative results 57

        4.4.2.1 Driving under rainy conditions 57
        4.4.2.2 Driving at night 58
        4.4.2.3 Known limitations of the Traffic Lights Recognition system 59
    4.5 Conclusions 59

Chapter 5* *61

Single Image Detection of Preceding Vehicles 61

    5.1 Introduction 61
    5.2 Literature Review 61
    5.3 Preceding Vehicle Detection 62
      5.3.1 Red lights detection 63
      5.3.2 Morphological lights pairing 65
      5.3.3 Horizontal edge boundaries 65
      5.3.4 Axial Symmetry check 66
      5.3.5 Preceding Vehicle distance estimation 66
    5.4 Experimental results 66
    5.5 Conclusions 68

Chapter 6* *69

Video-based Detection of Preceding Vehicles 69

    6.1 Introduction 69
    6.2 Literature review 69
    6.3 Proposed system overview 70
    6.4 Modules analysis 72
      6.4.1 Hypothesis Generation stage 72
        6.4.1.1 Red color segmentation 72
        6.4.1.2 Radial Symmetry Detection 73
        6.4.1.3 Light candidates pairing 75
      6.4.2 Hypothesis Verification stage 76
        6.4.2.1 Static verification 76
        6.4.2.2 Dynamic verification 79
    6.5 Experimental results 80
      6.5.1 Quantitative results using single images 81
      6.5.2 Quantitative results using videos 82
      6.5.3 Qualitative results in adverse conditions 83
      6.5.4 Performance issues 85
    6.6 Conclusions 86

Chapter 7* *87

Road Detection 87

    7.1 Introduction 87
    7.2 Literature review 87
    7.3 System overview 89
      7.3.1 Spatiotemporal seeds definition 91

7.3.2 Image Resolution Reduction 95
7.4 Experimental results 95
7.4.1 Qualitative Road Detection results in normal conditions 96
7.4.2 Qualitative Road Detection results in adverse conditions 96
7.4.3 Quantitative Road Detection results in mixed conditions 99
7.4.4 Sensitivity analysis of proposed RD system in mixed conditions 101
7.4.5 Quantitative Road Detection results in challenging conditions 103
7.4.6 Selection of optimal resolution for the RD system 104
7.5 Conclusions 105

Chapter 8* *107

Conclusions and Future Work 107

8.1 Introduction 107
8.2 Traffic Sign Recognition 107
8.2.1 Proposed TSR system 107
8.2.2 Performance evaluation 107
8.2.3 Real time implementation potential 108
8.2.4 Prospects of the proposed system 108
8.3 Traffic Lights Recognition 108
8.3.1 Proposed TLR system 108
8.3.2 Performance evaluation 109
8.3.3 Real time implementation potential 109
8.3.4 Prospects of the proposed system 109
8.4 Preceding Vehicle Detection 109
8.4.1 Proposed PVD system 109
8.4.2 Performance evaluation 110
8.4.3 Real time implementation potential 110
8.4.4 Prospects of the proposed systems 111
8.5 Road Detection 111
8.5.1 Proposed RD system 111
8.5.2 Performance evaluation 111
8.5.3 Real time implementation potential 111
8.5.4 Prospects of the proposed system 112
8.6 Final words and future work 112
8.7 Dissertation Publications 113
8.7.1 Journal Publications 113
8.7.2 Conference Publications 113

Bibliography 114

List of Figures

Figure 1: Distribution of global injury mortality by cause (2002)................................1

Figure 2: Projected global death for selected causes (2004-2030) 2

Figure 3: Trends in road traffic fatality rates in selected high-income countries 2

Figure 4: (a) Opel Eye system uses a camera to recognize speed signs and warn the driver. (b) Mobileye's traffic sign recognition system. It recognizes speed signs and warns the driver if speed limit is exceeded. 5

Figure 5 : (a) Lane Departure Warning from Opel Eye, (b) Infiniti Lane Departure Prevention, (c) Volkswagen Lane Assist and (d) Ford Lane Keeping Aid. 6

Figure 6 : (a) Audi Pre Sense Front Plus uses two long range radars and a camera to achieve AEB, (b) Mobileye uses just one monocular camera to warn drivers for impending collisions. 7

Figure 7 : Pedestrian Detection Systems by (a) Volvo, (b) Subaru and (c) Mobileye. .8

Figure 8 : (a) Near-IR night vision system by Mercedes-Benz and (b) FIR Night vision system by BMW...8

Figure 9 : Structure of a complete, entirely vision-based ADAS. The highlighted systems are covered in this thesis. 10

Figure 10 : (a) The original RGB image, (b) its R (red) channel and (c) its a* (red-green difference) channel...16

Figure 11 : The same scene in different color spaces. Each row contains the original image and the separate channels intensity: first row shows the RGB image, second the L*a*b*, third the HSV, fourth the YCbCr and fifth the $c_1c_2c_3$ image. 16

Figure 12 : Pixels affected by the gradient element g(p) for a range n=2. 17

Figure 13 : (a) Original image, (b) a* channel, (c) FRST result for N = {3, 6, 9} and α=1, (d) detected ROIs. 19

Figure 14 : (a) Original image and (b) Otsu's thresholding on channel a* of the image. 20

Figure 15 : (a) Original image and (b) result of Otsu's thresholding on the FRST of channel a*. 21

Figure 16 : Two consecutive frames with dense shadows in (a) and (b) produce the pseudo-colored optical flows of (c) for Horn-Schunck with a=3 and k=1, (d) for Horn-Schunck with a=15 and k=1500, (e) C. Liu for iterations 1,1,1 and (f) C. Liu for iterations 15,15,15. 23

Figure 17: Analysis of the operation of the Kalman Filter as presented in [46]..........25

Figure 18 : Seeds definition (left column) and segmentation result (right column). Road seeds are denoted with the red line, while non-road seeds with the blue one. In the segmented images, red color annotates the road pixels and black annotates the background. 28

Figure 19 : Traffic Signs under different environmental conditions: (a) excellent weather and illumination angle, (b) occlusions, (c) shaky camera, (d) driving against the sun, (e) rainy conditions and (f) night driving. 31

Figure 20 : Complete set of traffic signs. 32

Figure 21 : Structure of the TSR system 34

Figure 22 : The top row contains the original image on the left and the thresholded luminosity, L_{bo} on the right. The middle row shows the normalized red and green channels derived from a^* (left and right respectively). The bottom row demonstrates the normalized blue and yellow channels derived from b^* (left and right respectively). 35

Figure 23: Original image (a) and detected symmetrical shapes from all 4 channels for radii of size (b) $n = 5$, (c) $n = 8$ and (d) $n = 11$. Red circles show the two most prominent symmetrical shapes for each radius. 36

Figure 24 : Original image with triangular signs present (a) and detected symmetrical shapes from all 4 channels for radii of size (b) n = 5, (c) n = 8 and (d) n = 11. Red circles denote the two most prominent red results. 37

Figure 25: (a) Original image with a blue rectangular TS present and (b) detected FRST centers from all 4 channels for radii of size n = 5, n = 8 and n = 11. Blue circles denote the two most prominent blue results. 37

Figure 26 : (a) The original image with the results of non-maxima suppression for the total FRST, (b) cropped candidate TSs, (c) binarized TSs, (d) shape templates and (e) classified shapes overlaid on candidates. 38

Figure 27: (a) The original image with the results of non-maxima suppression for the total FRST, (b) cropped candidate TSs, (c) binarized TSs, (d) shape templates and (e) classified shape overlaid on candidate. 39

Figure 28: (a) First frame and detected TS candidates of 10px radial, (b) second frame and detected TS candidates of 10px radial, (c) third frame and detected TS candidates of 10px radial, (d) third frame and detected TS candidates of 10px radial, plus detected TS candidates of 11px radial around the areas of previously detected TS candidates (yellow rectangles in (d)). 40

Figure 29 : The TS dataset used for the evaluation of the TSR system. 42

Figure 30 : Typical TSD examples for a single frame using radii in {10, 15, 20}. 43

Figure 31 : Typical TSD and TST synergy results. 43

Figure 32 : Typical TSD & TST synergy results in challenging scenes. 44

Figure 33 : Examples of correct (top four rows) and incorrect (bottom two rows) classifications using the frame-by-frame method. 44

Figure 34 : Structural diagram of the proposed TLR system.....................................49

Figure 35 : (a) Original image, (b) L* channel, (c) a* channel and (d) *RG* channel. ..50

Figure 36 : (a) The original RGB image and (b) its RG channel, where the red TL on the right suffers from the "blooming effect". (c) The YB channel and finally (d) is the RGYB channel, with the "blooming effect" corrected. 51

Figure 37 : Handling the “blooming effect” at night. (a) RGB image, (b) RGYB channel, (c) RGYB channel after filling process. 52

Figure 38: Original driving scenes at day and night time (left column) and their corresponding FRST (right column). Dark spots denote green TLs, light spots denote red TLs. Local maxima/minima detection is performed above the orange line (which depends on the camera placement). 53

Figure 39 : TL annotation based on radial *n*. 54

Figure 40 : Spatiotemporal persistency check. Notice the trails left by the centers of two green TLs. 55

Figure 41 : Four consecutive frames, TL candidates annotated by rectangles. Non persistent candidates are dismissed. Candidates persistent for 3 out of 4 consecutive frames get verified. 55

Figure 42 : TL detection results in daytime driving. A false positive red color TL example is show in the bottom left image and a false negative is shown in the bottom right. 57

Figure 43 : Successful TL detection results in rainy conditions. 58

Figure 44 : TL detection results in urban night driving. 58

Figure 45 : Examples of temporary or permanent failures of the proposed system. ...59

Figure 46 : Structure of the proposed PVD system. 63

Figure 47 : (a) Original image, (b) red subspace of the L*a*b* color space visible and (c) FRST of the red subspace. 64

Figure 48 : (a) Original image, (b) red subspace of the L*a*b* color space where the “blooming effect” is visible and (c) FRST of the red subspace, a*. 64

Figure 49 : (a) Search region on the original image, (b) edge map of the search region, (c) its horizontal projection and (d) bounding box containing the candidate vehicle. .65

Figure 50 : Detection results for the data sets used; Left Column: Caltech DB (cars 1999), middle column: Caltech DB (cars 2001) and right column: Lara Urban Sequence. 67

Figure 51 : Detected vehicles in adverse weather conditions. 68

Figure 52: Flowchart of the proposed system. 71

Figure 53: Highlighting of rear red lights using the new Red color channel. (a), (b) Preceding vehicle images without and with illuminated rear brake lights and (c), (d) their respective new Red color channel. 73

Figure 54: Various examples of FRST based detection of preceding vehicle rear lights. The original images appear in odd rows and the respective FRST results in even rows. The results have been estimated for one radius (8 pixels for top rows and 6 pixels for bottom rows). 74

Figure 55: Cancellation of the “blooming effect” using the FRST. (a) Original image, (b) Red channel with presence of the blooming effect and (c) FRST result localizes rear lights. 74

Figure 56: A pair of candidate rear lights with detected centers denoted as CL and CR. The criteria for accepting a candidate pair depend on their relative distance, d, and their relative angle, θ, which should be smaller than two predefined thresholds. 75

Figure 57: First stage of static verification process. (a) The detected ROI containing the pair of candidate lights. (b) Cropped region of interest with vertical bisector highlighted in yellow. (c) Left sub-image, (d) mirrored right sub-image. 76

Figure 58: Multiple ROIs generated by the HV module 77

Figure 59: Successful detection results from Caltech 2001 (first row) and Caltech 1999 (second row). Two false positive results between two closely moving vehicles are shown in the third row. The problem of spatial inconsistency of the detection result in consecutive frames is demonstrated in the bottom row. 78

Figure 60: Example of Kalman tracking in the LARA video dataset. The vehicle detections are shown in yellow rectangles while the red rectangles denote the target being tracked. 83

Figure 61: Examples of vehicle detections from the proposed system in adverse conditions. Top row: night driving. Second row: driving in the rain. Third row: driving in snow. Fourth row: Typical false positive examples. 84

Figure 62 : Flowchart of the proposed Road Detection System. 90

Figure 63 : Geometrical properties used for road (red trapezoid) and non-road seeds definition. 92

Figure 64 : (a), (b) Two consecutive frames in the presence of shadows. (c) Normalized optical flow magnitude for grayscale images, (d) normalized HSC1 flow magnitude. Both flow magnitude images were estimated for a=15 and k=5 iterations. Bright pixels denote high optical flow magnitudes. 94

Figure 65 : Non-road (blue pixels) and road (red pixels) seeds using the proposed method in various conditions. 95

Figure 66 : Road detection results in normal driving conditions using the proposed system. 96

Figure 67 : Successful road detection results in adverse conditions. (a)-(d) Various scenes with heavy shadows, (e) passing through a tunnel, (f) passing under a bridge, (g), (h) rainy conditions and (i) night driving. 98

Figure 68 : Less successful road detection in particularly challenging conditions. (a) Approaching shadowed road, (b) entering tunnel, (c) heavy rainfall with camera above the effective area of the wipers, (d) wiper passing in front of camera, (e) night driving in badly lit road, (f) misclassification due to high rate video compression.....99

Figure 69 : Qualitative results in frames with low quality rate. Yellow: TP, Green: FN, Red: FP. 101

List of Tables

Table 1: Advanced Safety Technologies rewarded by Euro NCAP since 2010............4

Table 2 : Video streams used for the purposes of the thesis..11

Table 3: TSR rates using only one frame...45

Table 4 : Detection accuracy of the PVD System using a single image......................66

Table 5: Vehicle detection rates in both Caltech datasets..81

Table 6: Vehicle detection rates in Caltech 2001 dataset 82

Table 7: Vehicle detection rates before and after tracking 82

Table 8: Absolute and relative processing times and frame rates analysis for different resolutions. (AS: Axial Similarity, FRST: Fast Radial Symmetry Transform, PM: Pair Matching, MF: Median Filtering, Others: Other modules)...85

Table 9: Overall performance of proposed RD algorithm compared to [123] and [128]...100

Table 10 : Description of DIPLODOC video streams conditions 100

Table 11 : Performance of proposed RD System in DIPLODOC video streams 100

Table 12 : Segmentation rate of the proposed RD system in DIPLODOC sequences for different setups 102

Table 13 : Overall performance achieved by proposed RD algorithm compared to [24] and [131], [133]. 104

Table 14: Results achieved for various resolutions of DIPLODOC video streams...105

Glossary of Acronyms

ADAS	Advanced Driver Assistance Systems
AEB	Autonomous Emergency Braking
$c_1c_2c_3$	Illumination Invariant color space
Caltech 1999	Dataset of pictures of vehicles rear views taken in a parking lot
Caltech 2001	Dataset of pictures of vehicles rear views taken in freeways
CCD	Charge-Coupled Device
CIE-L*a*b*	CIE 1976 (L*, a*, b*) color space
DARPA	Defense Advanced Research Projects Agency
DIPLODOC	Project with annotated video streams used for road detection
E.U.	European Union
Euro NCAP	European New Car Assessment Programme
FCW	Forward Collision Warning
FIR	Far Infrared
Fps	Frames per second
FRST	Fast Radial Symmetry Transform
GPS	Global Positioning System
HG	Hypothesis Generation
HSC1	Horn-Schunck optical flow using c_1 color channel
HRI1 ~ HRI5	Datasets with annotated video streams used for vehicle detection
HSI	Hue-Saturation-Intensity color space
HSV	Hue-Saturation-Value color space
HV	Hypothesis Verification
ISA	Intelligent Speed Assistance
LED	Light Emitting Diode
LiDAR	Light Detection And Ranging
MAE	Mean Absolute Error
NCC	Normalized Cross-Correlation
NHTSA	National Highway Traffic Safety Administration
NIR	Near Infrared
PAL	Phase Alternating Line video
PVD	Preceding Vehicle Detection
RD	Road Detection
RGB	Red-Green-Blue color space
ROI	Region Of Interest
RWA	Random Walker Algorithm
SSIM	Structural Similarity metric
TL	Traffic Light
TLC	Traffic Light Classification
TLD	Traffic Light Detection
TLR	Traffic Light Recognition
TS	Traffic Sign

TSC	Traffic Sign Classification
TSD	Traffic Sign Detection
TSR	Traffic Sign Recognition
TSR1 ~ TSR4	Video streams used for Traffic Sign Recognition
TST	Traffic Sign Tracking
U.S.A.	United States of America
YCbCr	Luminance, Blue and Red Chroma difference color channel

Glossary of Mathematical Symbols

A	Accuracy measure
a*	Red-green opponency channel
b*	Yellow-blue opponency channel
C	Candidate image
D	Dirichlet integral
F	Effectiveness measure
FN	False Negative
FP	False Positive
g	Quality measure
g	Gradient
I	Image
i	Index of row
I_t	Derivative of image along the time dimension
I_x	Derivative of image along the horizontal dimension
I_y	Derivative of image along the vertical dimension
j	Index of column
k	Number of iterations
l	Index of frame
L	Laplacian matrix
L*	Lightness channel
L_{bo}	Thresholded lightness channel
Ls	Left sub-image
M_n	Magnitude projection image
n	Radius of symmetrical shape
N	Range of radii
O_n	Orientation projection image
P	Precision (also called Correctness)
p	Pixel
R	Recall (also called Completeness)
RG	Lightness boosted red-green color opponency channel
$RGYB$	Color channel derived by adding RG, YB
Rs	Right sub-image
$round(.)$	Rounding operator
S_n	Radial symmetry contribution
T	Template image
t	Time
TN	True Negative
TP	True Positive
V_B	Background seeds for RWA
V_R	Road seeds for RWA
V_x	Horizontal optical flow

V_y	Vertical optical flow
YB	Lightness boosted yellow-blue color opponency channel
α	Optical flow regularization constant
β	User-defined parameter of RWA

Chapter 1

Advanced Driver Assistance Systems (ADAS)

1.1 Introduction

This chapter contains an introduction to Advanced Driver Assistance Systems (ADAS). It begins with an overview of the road accident statistics worldwide that dictate the need for further predictive action towards safer vehicles, moving on to a demonstration of state-of-the-art systems that are already commercially available. Then, an overview of vision-based systems already implemented by vehicle manufacturers and after-market vendors is presented, followed by a presentation of vision-based ADAS still in research stage. An exposition of the parts of a complete vision-only ADAS is next, wrapped up by a short summary of the systems developed for this thesis. The chapter closes with a presentation of the datasets used for the purposes of the thesis, the performance evaluation methods for quantitative and qualitative assessment of the proposed systems and some final conclusions.

1.2 The need for ADAS development

During the past few decades, road accidents have proven to be one of the most common causes for the loss of human lives. According to a study by the World Health Organization issued in 2004 [1], road accidents were estimated as the cause for 1.2 million people killed and as many as 50 million people injured worldwide. The seriousness of the situation can be reflected by the fact that in 2000, road accidents were the main cause of death inflicting injuries, resulting to approximately 23% of worldwide injury related deaths (see Figure 1), while in 2004 they were the second overall cause of death for people of ages 5-14 and the first cause of death for ages 15-29 years [2].

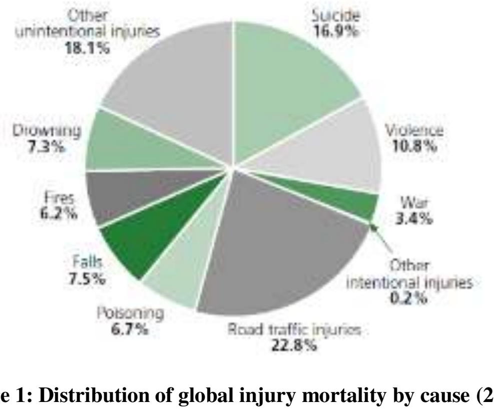


Figure 1: Distribution of global injury mortality by cause (2002).

More importantly, the forecasts state that road accident related deaths will raise by 65% between 2000 and 2020 [3], [4]. In fact, the absolute number of road-accident inflicted deaths is estimated to double in the period 2004-2030 [1], as shown in Figure 2. In terms of percentages, road accidents are predicted to become the 5th leading death cause by 2030, increasing by a factor of 0.54, from 2.2% to 3.6% of the world's deaths.

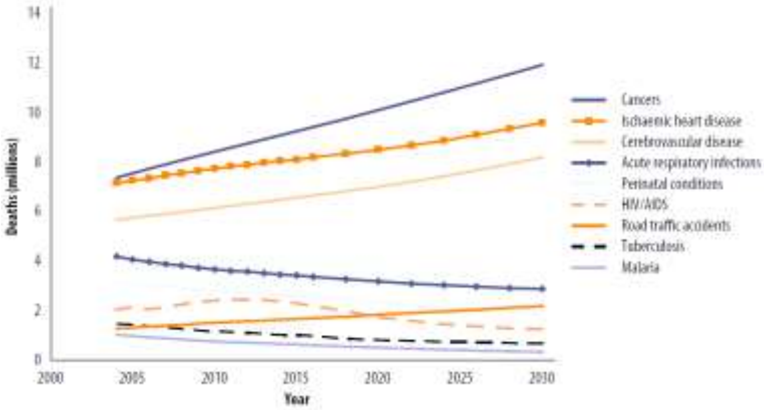


Figure 2: Projected global death for selected causes (2004-2030)

These dire predictions concerning the rise of road accident related deaths are mainly based on the fact that the total number of vehicles worldwide rises, resulting in a subsequent increase of accidents. However, vehicle safety has improved over the years, mainly because of the advances of passive safety systems that have been included in almost all commercial vehicles. Such systems include seat belts, airbags and various improvements made on the vehicle bodies, improving their crashworthiness for passengers and pedestrians. Improvements in passive safety may have offered a lot so far, but now relative technologies seem to have reached a peak, since they cannot offer solutions to further mitigate deaths caused by accidents inflicted by drivers' mistakes. As demonstrated in Figure 3, the reduction in road accident inflicted death has been slowed down over the past decade (except from France). In fact, a recent study by the National Highway Traffic Safety Administration (NHTSA) of the U.S.A. revealed that from all the driver-related causes of road accidents, a 41% was due to recognition errors and a 34% due to decision errors [5].

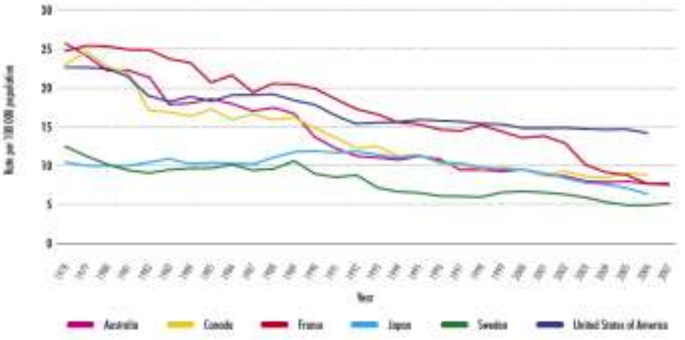


Figure 3: Trends in road traffic fatality rates in selected high-income countries

A conclusion that can be drawn from the analysis of road accident related data is that the further reduction of accidents has to rely more on active, state-of-the-art solutions generally known as Advanced Driver Assistance Systems. ADAS are new emerging technologies, primarily developed as automated advisory systems for drivers with a goal to enhance driving safety.

In this context, the European New Car Assessment Programme (Euro NCAP) has announced [6] that starting in year 2014, every car manufacturer hoping to achieve a perfect, five-star, score on the Euro NCAP safety ratings will have to equip their vehicle with an Autonomous Emergency Braking (AEB) system. Similar efforts from Euro NCAP since it was first established in 1996, led to the improvement in passive and active safety systems used in vehicles. The corresponding criteria are a good indication of state-of-the-art systems that are supposed to be embedded into commercial vehicles. Some of the future trends are revealed when studying the rewards that Euro NCAP has established since 2010, which give incentives to manufacturers to include advanced safety systems in their vehicles. The advanced safety technologies that have been rewarded since 2010 until today are included in Table 1.

1.3 Vision based subsystems used in commercial solutions

The summary of state-of-the-art ADAS recently included in commercial vehicles denotes a turn towards heavy usage of computer vision methods. An increasing number of ADAS rely either on the information fusion of several sensors, including cameras, or on the utilization of information derived solely from monocular, or stereoscopic cameras. In fact, from the ADAS presented in Table 1 only Automatic Emergency Call does not include a computer vision based possible subsystem. The use of cameras was somehow neglected in the past mainly because computer vision algorithms required very powerful and expensive processors which, combined with the high prices of good quality cameras, raised the price of camera-based ADAS to a level that did not allow their commercialization.

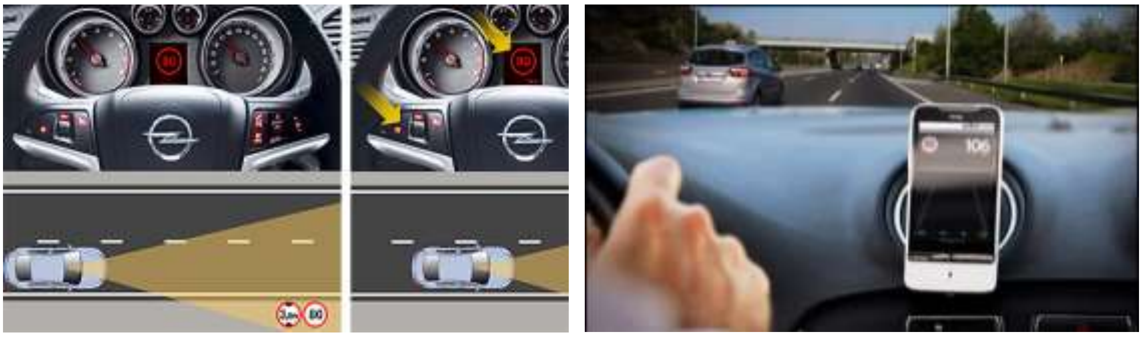
Recently, the increasing processing power of embedded digital signal processing systems combined with the development of cheap, small video cameras with good image quality led ADAS research towards implementation of systems that operate based on information coming only from cameras. Such systems usually concentrate only on one of the problems of driver assistance, since the inclusion of all different problems in one single system is still a quite challenging issue. In this section, the most prominent examples of vision based subsystems of commercially available ADAS are presented.

Table 1: Advanced Safety Technologies rewarded by Euro NCAP since 2010.

Advanced Safety Technology	Description	Rewarded Manufacturers
Blind Spot Monitoring	Radar or camera based systems to warn a driver when changing lanes about possible car approaching from the "blind spot".	Mazda (2011 - Mazda Rear Vehicle Monitoring system, RVM) Audi (2010 - Audi Side Assist)
Lane Support Systems	Lane Departure Warning: camera based systems to warn drivers in case of steering into another lane. Lane Keep Assist: camera based systems that steer the car back to the driving lane	Ford (2011 – Lane Keeping Aid) Infiniti (2011 - Lane Departure Prevention, LDP) Opel (2010 – Opel eye) Volkswagen (2010 - Lane Assist)
Speed Alert Systems	Also known as Intelligent Speed Assistance (**ISA**), it is a system that warns drivers of exceeding the speed limit of the road, or the manually set speed limit	No rewards until the time of writing.
Autonomous Emergency Braking	Radar or lidar based systems that act independently of the driver (**Autonomous**) to avoid or mitigate an accident in critical situations (**Emergency**) by applying the brakes (**Braking**)	Audi (2012 – Pre Sense Front Plus) Ford (2011 – Forward Alert, Active City Stop) Mercedes-Benz (2011 - Collision Prevention Assist) Volkswagen (2011 - City Emergency Brake) Honda (2010 - Collision Mitigation Brake System) Mercedes-Benz (2010 - PRE-SAFE® Brake) Volvo (2010 - Volvo City Safety)
Attention Assist	Drowsiness detection systems that warn drivers to prevent them from sleeping on the wheel by prompting for breaks.	Ford (2011 – Driver Alert) Mercedes-Benz (2011 – Attention Assist)
Automatic Emergency Call	System that sends an automatic message to an emergency call center in case the vehicle has been in a crash.	BMW (2010 - Assist Advanced eCall) Citroen (2010 - Localized Emergency Call) Peugeot (2010 - Connect SOS)
Pre-crash Systems	Systems that predict an accident to optimize the functionality of protection systems such as restraints (seatbelts) or airbags.	Honda (2010 - Collision Mitigation Brake System) Mercedes-Benz (2010 - PRE-SAFE®) Mercedes-Benz (2010 - PRE-SAFE® Brake)
Vision Enhancement Systems	Systems improving the visibility of the driver, even in adverse lighting or weather conditions, like night driving, or driving through fog.	Opel (2011 - Adaptive Forward Lighting, AFL)

1.3.1 Traffic Sign Recognition

Traffic signs carry important information about the driving environment and they assist drivers in making correct decisions for the safe navigation of their vehicles. Automatic recognition of traffic signs has been the most widely researched ADAS during the past decades. In this scientific field, a high number of papers, laboratory systems and reports have been already presented using generic image processing and machine learning algorithms. Recently, traffic sign recognition (TSR) systems have been installed in Speed Alert Systems, like the one introduced by Opel and included in the 2008 Insignia model. The system is called "Opel Eye" [7] and among other things, it detects speed signs and no-overtake signs and informs the driver of their existence as demonstrated in Figure 4(a). After-market systems using TSR can also be found, as the one developed by Mobileye and used in the BMW 7 series [8]. It is a system similar to the Opel Eye, but it also claims to detect LED speed signs as well (Figure 4(b)).



(a) **(b)**

Figure 4: (a) Opel Eye system uses a camera to recognize speed signs and warn the driver. (b) Mobileye's traffic sign recognition system. It recognizes speed signs and warns the driver if speed limit is exceeded.

Even though some versions of TSR are already available commercially, the technology will have reached maturity only when it will have tackled some rather challenging problems. More specifically:

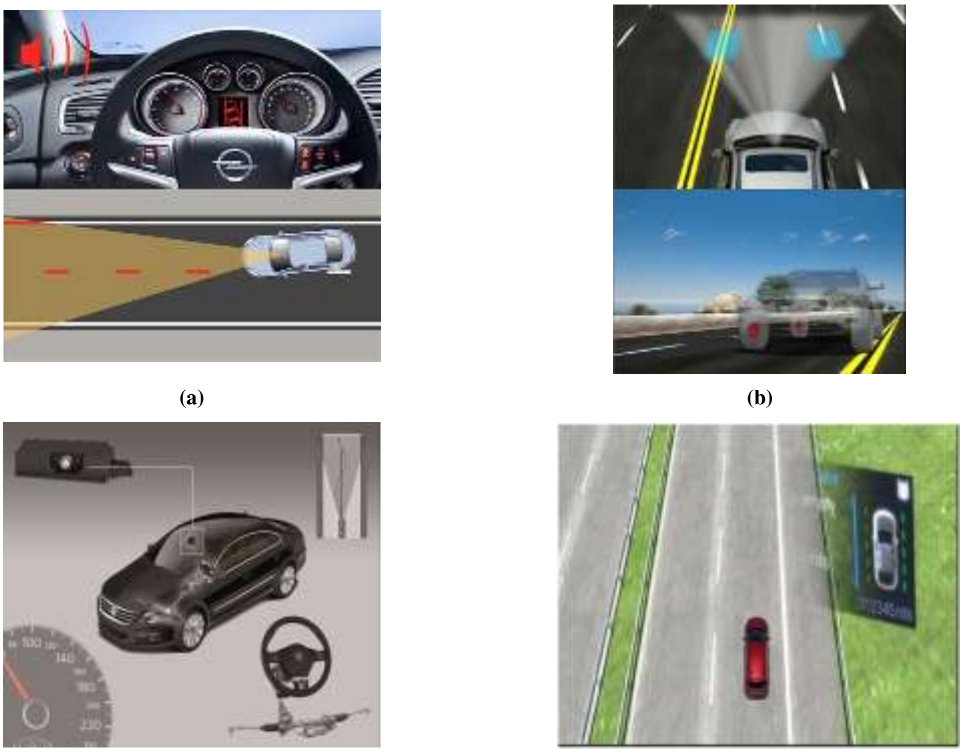
a) The TSR systems detect and recognize only speed signs and no-overtake signs. While they are very useful for ADAS, they are still oversimplified versions of a complete traffic sign recognizer which would include all possible traffic signs.

b) The systems that are commercially available are not always functional, depending on the weather and lighting conditions. Even though this problem is well-known in the computer vision society, it still poses a big obstacle in the usage of such systems for autonomous driving, or even reliable driver warning systems.

These challenges are the main reasons why computer vision based TSR are not reliable for commercial applications i.e. for a great variety of driving scenarios and at least the most important road signs. This is why research in the area of traffic sign recognition is still very active [9]. Moreover TSR is a task that can be solved using GPS based technologies in collaboration with precise and dynamically changing maps. This simple and efficient solution leads the major automotive companies to minimize their effort in TSR technologies.

1.3.2 Road Lane Recognition

One of the most mature vision-based technologies for an ADAS is road lane recognition. Its presence is vital in Lane Support Systems, as it provides the basic information needed for such a task. Road lane recognition is targeted mainly on highway driving, or on driving in well-maintained city or country roads. The technology behind it is fairly simple, since there are several robust algorithms for line finding in an image, the most widely used being Hough transform [10]. Commercially available solutions include "Opel Eye", which has a module for lane departure warning for the driver, Infiniti with its Lane Departure Prevention module, Volkswagen with the Lane Assist System and Ford with the Lane Keeping Aid system. Some indicative pictures of the aforementioned systems are shown in Figure 5.



(a) (b)

(c) (d)

Figure 5 : (a) Lane Departure Warning from Opel Eye, (b) Infiniti Lane Departure Prevention, (c) Volkswagen Lane Assist and (d) Ford Lane Keeping Aid.

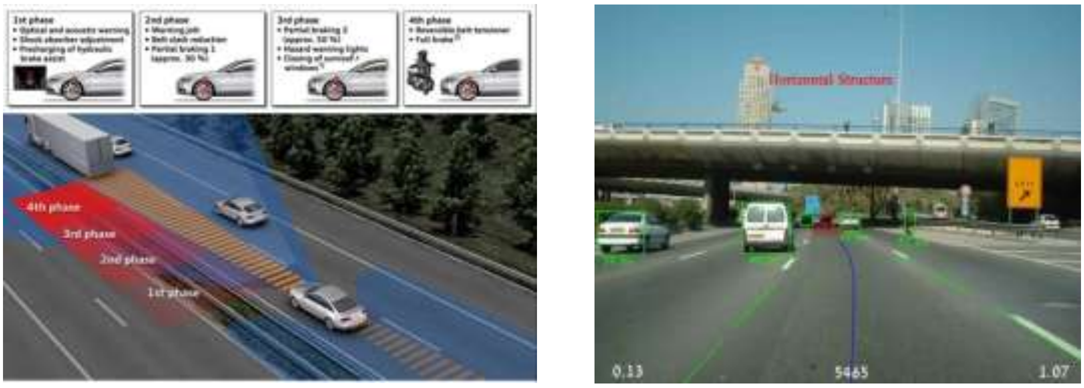
Lane assist systems are invaluable for highway driving, but they still fall short of what is expected by an ADAS in situations of driving in unstructured or badly maintained roads, where road lanes are not visible. In these scenarios such systems do not have an alternative solution to the problem of lane keeping, so they cannot offer useful services to the driver. In such scenarios, a road detection module would deliver much more reliable information.

1.3.3 Vehicle Detection

Vehicle detection is the cornerstone of several ADAS, since it provides information about possible dangers like impending collisions. It can therefore be used in Autonomous Emergency Braking systems, in pre-crash systems and also in blind-spot detection.

Several systems use alternative techniques to achieve vehicle detection, like radars (Mercedes Benz Pre-SAFE® Brake and Collision Prevention Assist, Honda Collision Mitigation Brake System, Ford Forward Alert) or Light Detection And Ranging (LIDAR) technology (Volvo City Safety, Volkswagen City Emergency Brake, Ford Active City Stop). Radars are chosen due to their good distance measurement when a sufficient radar reflectance of objects in front of the vehicle is present. However, they can be negatively affected by mud, snow or leaves blocking its "view" and appear problematic when other cars cut in to the lane of the ego-vehicle, or when the ego-vehicle makes a small radius corner. On the other hand LIDARs offer distance measurement both in daytime and nighttime, but their sensors are compromised when stained by mud or snow and they fail to operate in adverse conditions such as fog or heavy rain.

The aforementioned disadvantages of those two technologies are the reason vehicle manufacturers have started using them in synergy with cameras. One such system is Audi's Pre Sense Front Plus, which combines information from two long range radars that detect obstacles in front of the vehicle, with data from a windscreen-mounted camera to assess the probability of an impending forward collision and warn the driver, or apply the brakes if the threat is imminent. Vision has also been used in Mobileye's vehicle detection after-market solution [11], which is used for several modules, like Forward Collision Warning, Headway Monitoring and Warning, etc. The difference is that Mobileye has implemented vision-only systems, something that is particularly hard to achieve for all possible driving conditions. The two aforementioned camera-based systems are shown in Figure 6.


1st phase
• Optical and acoustic warning
• Shock absorber adjustment
• Precharging of hydraulic brake assist
2nd phase
• Warning jolt
• Belt slack reduction
• Partial braking 1 (approx. 30 %)
3rd phase
• Partial braking 2 (approx. 50 %)
• Hazard warning lights
• Closing of sunroof + windows
4th phase
• Reversible belt tensioner
• Full brake
4th phase
3rd phase
2nd phase
1st phase
Horizontal Structure
0.13
5465
1.07

(a) **(b)**

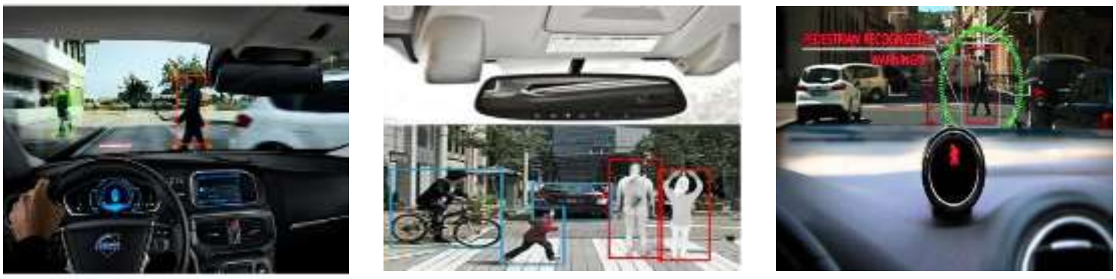
Figure 6 : (a) Audi Pre Sense Front Plus uses two long range radars and a camera to achieve AEB, (b) Mobileye uses just one monocular camera to warn drivers for impending collisions.

1.3.4 Driver Drowsiness Detection

The use of camera positioned inside the vehicle and facing the driver is among the most reliable methods to evaluate the state of mind the driver is in. This can be used for detecting possible driver drowsiness and issue an audible warning sign that warns him/her to stop driving and take a break. However, commercial systems in vehicles have not used this approach, but a rather more indirect one, i.e. the use of information about the driving style to determine if the driver has low vigilance. Ford Driver Alert uses a front-facing camera to detect sudden and exaggerated corrections to the vehicle's motion, which are characteristics of sleepy or inattentive drivers. Mercedes-Benz Attention Assist uses the same idea, but without a camera; information about steering angle comes from a sensitive sensor and is used for assessing potential drowsiness of the driver. Methods in modern literature tend to focus on solving the problem in the more direct way, monitoring the gaze and head pose of the driver [12]. However, this is still an ongoing research with results not robust enough for commercialization, since there are many challenges still to be faced, like inconsistent lighting, occlusion of the eyes from glasses, etc.

1.3.5 Pedestrian Detection Systems

Another feature of ADAS that has been extensively researched is pedestrian detection [13], [14], [15]. However, this technology was late-blossomed in commercial systems, as only Volvo has included such functionality in one of its cars and more specifically the Volvo S-60 (starting from the 2010 model). The task of pedestrian detection is accomplished with the use of a front-facing camera in combination with a radar sensor that measures the distances to the pedestrians (Figure 7(a)). A vision-only approach has been scheduled to be used by Subaru in its EyeSight™ system announced for 2013 (Figure 7(b)). The system will be based on a stereoscopic pair of cameras placed on both sides of the rearview mirror of the vehicle [16]. The vision-only approach has been followed by Mobileye as well, in their after-market Pedestrian Collision Warning System which is demonstrated in Figure 7(c).

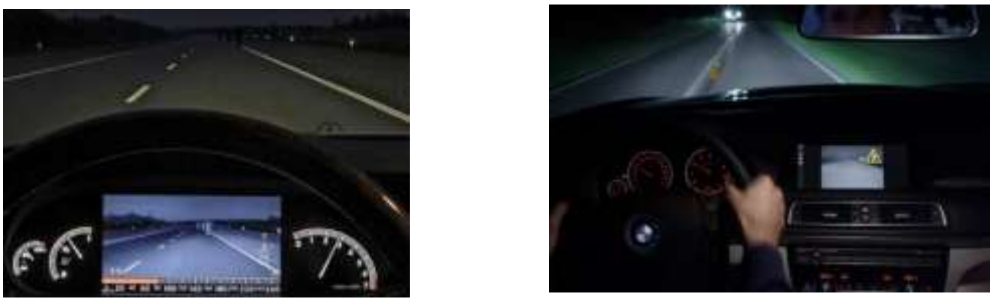


(a) **(b)** **(c)**

Figure 7 : Pedestrian Detection Systems by (a) Volvo, (b) Subaru and (c) Mobileye.

1.3.6 Night Vision Systems

Recently [17], BMW and Mercedes-Benz offered night vision as an extra feature of their 7-series and S-Class models respectively, using different approaches. BMW used a passive Far-InfraRed (FIR) sensor to stream night vision images in the car monitor, while Mercedes-Benz used a near-IR system for the same reason. Both systems are really expensive and this is why they are offered only in top-class vehicles of the two brands. When combined with a module like pedestrian detection, these systems can prove truly life-saving for the driver, because they provide extra visibility. The two systems are shown in Figure 8.



(a) **(b)**

Figure 8 : (a) Near-IR night vision system by Mercedes-Benz and (b) FIR Night vision system by BMW.

1.4 Vision based subsystems still in research stage

The majority of the subsystems presented in the previous section have been included in commercial vehicles after exhaustive research of their robustness and trustworthiness. As the paradigm of pedestrian detection shows, a system that has been extensively covered by scientific research might take a long time until it reaches commercial maturity. Thus, there are still some systems not mature enough to be considered for inclusion in commercial vehicles. In addition, even the commercially available systems have their limitations; bad weather conditions, limited visibility, sudden illumination changes, dense traffic etc. are only some of the factors that can deteriorate the performance of a vision-based ADAS. In this section we will present some of the open research areas for vision based ADAS.

1.4.1 Traffic Lights Recognition

One of the most neglected research areas in ADAS technology is the recognition of traffic lights. Taking into account that the number of serious accidents caused by traffic light violations is significantly greater than the number caused by violations of other types of traffic signs, not many researchers have treated them as essential information for an ADAS.

Traffic light detection is not a trivial problem due to the very dense presence of red and green light sources in cities. This effect increases the false positive error rate of traffic light recognition systems significantly.

1.4.2 Driver Gaze Detection

In the previous section driver inattention was mentioned; however commercial vehicles have not used computer vision methods for this task. Current research has been moving towards using in-vehicle cameras monitoring the driver, so that not only his/her potential drowsy state is detected, but also in order to follow his/her gaze so that it can be correlated to the driving scene and warn for potential lack of focus to a critical point. However, these attempts are still in a very early stage and they are still far from being mature enough to be deployed in commercial vehicles.

1.4.3 Road Detection

Road lane detection is, as mentioned already, one of the most mature technologies in commercially available ADAS. However, they provide useful assistance only in specific situations such as highway driving, or driving on well-preserved roads. This is not the general case though; many roads lack distinct markings, while others have no markings at all. Driving in unstructured and even unpaved rural roads is also a great challenge, as shown in recent DARPA challenges [18]. Such challenges require a more general drivable path detection algorithm and this is where researchers have focused lately; robust road detection in all kinds of environments, at any time of the day and under different weather conditions.

1.5 General structure of a computer vision based ADAS

Currently commercial vehicular technology has only included vision systems in conjunction with other sensors, or specific subsystems that are robust enough to be used in real world driving scenarios, like lane detection systems. A robust setup of ADAS that would only use vision information is presented in Figure 9.

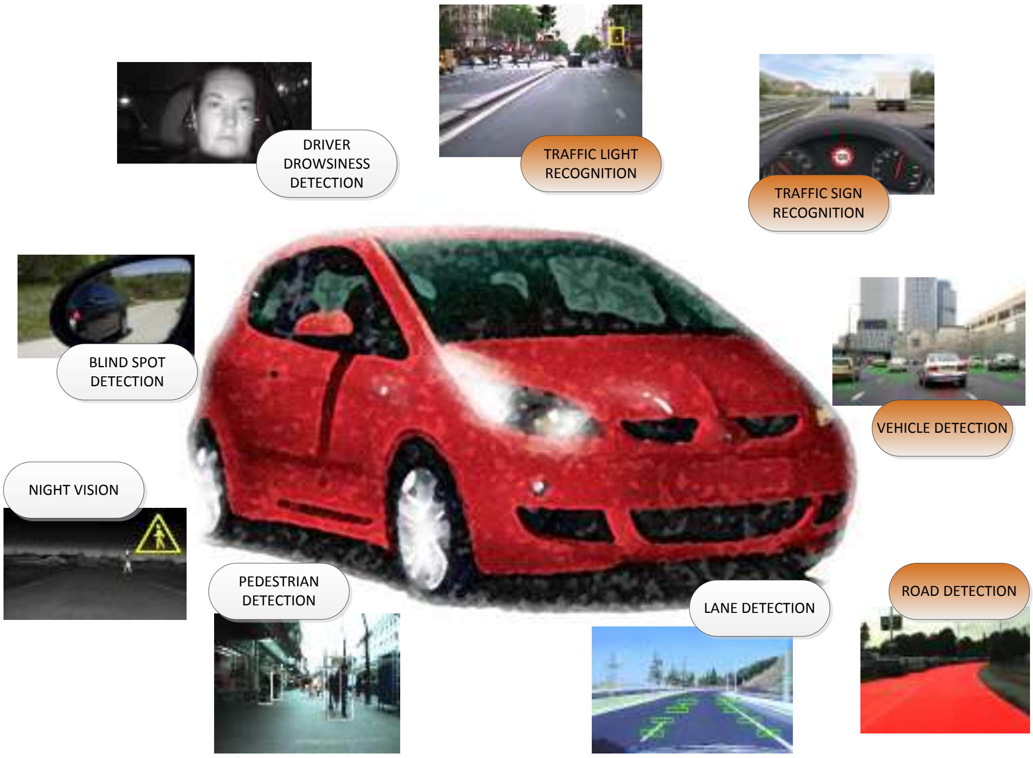

DRIVER DROWSINESS DETECTION
TRAFFIC LIGHT RECOGNITION
TRAFFIC SIGN RECOGNITION
BLIND SPOT DETECTION
VEHICLE DETECTION
NIGHT VISION
PEDESTRIAN DETECTION
LANE DETECTION
ROAD DETECTION

Figure 9 : Structure of a complete, entirely vision-based ADAS. The highlighted systems are covered in this thesis.

The continuous progress in the area of automatic assistance and driver warning methods indicates that we are not very far from developing very reliable ADAS in typical driving conditions. In such systems, the majority of sub-systems will be implemented using vision processing methods.

The scope of this thesis is to provide novel solutions for some of the most challenging problems using frontal-faced cameras, i.e. traffic sign recognition, traffic light recognition, vehicle detection and road detection. Specific attention is given to developing and testing systems that are robust to weather and illumination changes, providing acceptable accuracy in both daytime and nighttime driving scenarios, while not being affected by the driving environment. All systems are designed for monocular cameras mounted on the windscreen of the car.

1.6 Datasets used for system development and evaluation

One of the great problems faced during the evaluation of a computer vision system is usually the lack of carefully selected, publicly available, manually annotated datasets with a variety of examples. When the system deals with a real-world problem and demands a great mixture of videos taken under different conditions and frame-based annotation information is required, then the number of man hours that have to be spent on the construction of the dataset is extremely high. The availability of very small publicly available annotated datasets is the most important reason for the absence of common benchmarks for ADAS. The systems developed in thesis are no exception; they have been tested on several different video streams, spanning from videos shot for the purposes of the thesis to publicly available video datasets containing manually annotated results. The datasets used, an overview of which is given in Table 2, will be presented in the following sections.

Table 2 : Video streams used for the purposes of the thesis

No	Name	Used for	Annot.	Car	Camera	Resolution	fps	Duration	Environ.	Weather	Daytime	Place
1	TSR1	TSR	No	AX	DCR-TRV60E	720x576	25	8:55	Rural	Sunny	Day	Ioannina
2	TSR2	TSR	No	AX	DCR-HC85	720x576	25	2:10	Rural	Rainy	Day	Patras
3	TSR3	TSR	No	AX	DCR-HC85	720x576	25	1:05	Rural	Good	Night	Patras
4	TSR4	TSR	No	AX	DCR-HC85	720x576	25	3:07	City	Good	Night	Patras
5	Patras1	RD/VD	No	AX	Pinhole	640x480	25	> 2 h	Mixed	Mixed	Mixed	Patras
6	Patras2	RD/VD	No	Colt	PV-GS180	720x576	25	> 1 h	City	Sunny	Day	Patras
7	Ioann1	RD/VD	No	Colt	HDC-SD100	1080i	25	~1 h	Rural	Mixed	Day	Ioannina
8	LARA	TLR/VD	TLs	C3	Marling F-046C	640x480	25	8:49	City	Sunny	Day	Paris
9	DiploDoc	RD	Road	-	MEGA-D	320x240	15	0:56	Mixed	Sunny	Noon	Trento
10	Alvarez1	RD	Road	ZZ	Bumblebee	320x240	15	0:56	Rural	Sunny	Noon	Barcelona
11	Alvarez2	RD	Road	ZZ	Bumblebee	320x240	15	0:32	Rural	Rainy	Morning	Barcelona
12	HRI1	VD	R/V/TS/TL	-	Unknown	800x600	10	16:24	Mixed	Dry	Afternoon	Unknown
13	HRI2	VD	R/V/TS/TL	-	Unknown	800x600	20	18:50	Mixed	Dry	Evening	Unknown
14	HRI3	VD	R/V/TS/TL	-	Unknown	800x600	10	11:12	Mixed	Rainy	Afternoon	Unknown
15	HRI4	VD	R/V/TS/TL	-	Unknown	800x600	10	11:22	Mixed	Dry	Night	Unknown
16	HRI5	VD	R/V/TS/TL	-	Unknown	800x600	20	13:48	Mixed	Snow	Afternoon	Unknown
17	YouTube	TLR/VD	No	-	Various	Various	-	> 2 h	Various	Various	Various	Various
18	Caltech 1999	VD	No	-	Unknown	892x592	-	126 frames	Parking	Sunny	Day	Caltech
19	Caltech 2001	VD	No	-	Unknown	360x240	-	526 frames	Urban	Sunny	Day	S. California

1.6.1 Video streams acquired for qualitative evaluation of thesis methods

For the purposes of this thesis, several video streams of driving in different environments and conditions were acquired. The videos were shot using different cameras with various resolutions which were mounted on the windscreen of the vehicle. The video streams include various weather conditions, different driving environments and were shot at different hours of the day. The details of the video streams are presented in lines 1 through 7 of Table 2, along with information on the subsystem they were used for. These video streams were used mostly in the development and the qualitative evaluation of the ADAS presented in the thesis.

1.6.2 Video streams used for quantitative evaluation of thesis methods

For quantitative evaluation of the proposed ADAS, video streams have to be frame based manually annotated with traffic signs, traffic lights and road area information. As previously mentioned, this is a very basic problem of the field of real-world computer vision systems, since the complexity and diversity of the scenes combined with the vast number of frames that have to be manually annotated make such efforts extremely costly and time consuming. The quantitative analysis of the proposed ADAS methods was based on a few publicly available video datasets.

The process of traffic signs detection was probably the most difficult to assess, since there isn't any sufficient publicly available annotated datasets to use at the time of development and evaluation of the thesis systems. Instead, a frame by frame inspection and characterization of the results on selected parts of the video streams TSR1 to TSR4 (lines 1 to 4 in Table 2) is used. Recently, traffic signs recognition has been more thoroughly covered by researchers [19]-[20], but the TS dataset used was substantially smaller than the one used for the implementation of our system.

Traffic lights recognition also is a challenging process to evaluate. A recent work by de Charette et al [21] provided a useful video stream found in [22] called LARA, which comprises 11179 frames of a daytime city driving video, with ground truth manual annotations of the 32 traffic lights met, in all the frames of their appearance.

Road segmentation is probably the most challenging scenario for manual annotation. The few publicly available video streams with manually annotated road area that can be found online typically comprise very few frames, like the DiploDoc sequence used by Lombardi et al in [23] (number 9 in the list of Table 2) and the two sequences used by Alvarez et al in [24] (numbers 10 and 11 in Table 2). The aforementioned video streams contain manual annotations of the road area in stereo frame pairs taken from driving scenes. The one used in [23] has 865 annotated frames and the one used in [24] comprises 1335 frames in total, of which 1005 are annotated with road information.

1.6.3 Publicly available video streams used for qualitative evaluation

For the qualitative assessment of the systems presented in this thesis in extremely adverse conditions, we have also used over 2 hours of video streams shot with on-board video cameras downloaded from YouTube. The purpose of using these videos were mainly to add more representative examples of driving scenarios that could not be covered using the videos mentioned in the previous sections, e.g. driving in snow, driving at night in big city roads (New York, Los Angeles), driving in fog, etc. These videos cannot be extensively described, since they are taken with different video cameras, in diverse environments and they are used just for qualitative purposes.

Finally, at the late stages of this research, access to a large dataset of driving scene videos acquired by Honda Research Laboratories was provided [25], [26], [27], [28], [29], [30]. These videos were acquired in a diversity of environments, weather and illumination conditions and they were extensively annotated (free road area, traffic lights, traffic signs and vehicles were among the annotated categories). These videos were used for qualitative result analysis purposes in this thesis. The details of the five video streams comprising the dataset are thoroughly described in [27] and have been summarized in lines 12-16 of Table 2. Driving in both highway and urban roads in dry weather conditions has been covered for daytime in HRI1 and HRI2. The former has a frame rate of 10 fps, while the latter was shot at 20 fps. Heavy rain conditions have been covered in HRI3, where the vehicle faces also urban and highway scenes in the afternoon. Night driving in dry conditions has been included in HRI4, once again in both urban and highway roads. Both HRI3 and HRI4 have been shot at a frame rate of 10 fps. Finally, driving after heavy snowfall has been included in HRI5. This stream was acquired at 20 fps in the afternoon, while driving in both urban and highway roads.

1.6.4 Publicly available image sequences used for quantitative evaluation

For quantitative evaluation of the static performance of the preceding vehicle detection algorithms developed in this thesis, two datasets of images of the rear part of vehicles provided by Caltech [31], were also used. The first dataset is called Caltech 1999 and comprises 126 images, sized 892x592 pixels, of cars from the rear taken at the parking lots of Caltech. The second dataset, called Caltech 2001, contains 526 pictures of 60 different cars, with resolution 360x240, that have been taken in freeways of southern California. Both datasets are ideal for the evaluation of preceding vehicle algorithms, since they contain a variety of vehicles in front of the ego-vehicle.

1.7 Performance evaluation measures

To achieve an efficient evaluation of the systems developed in the scope of this thesis, it is imperative that we use a mixture of quantitative and qualitative methods that are designed for this purpose. These methods are shortly presented here and will be used throughout this dissertation.

1.7.1 Quantitative measures for quality evaluation

The measures that are most commonly used for assessing the detection quality of all computer vision systems are based on the definition of four basic metrics. When the problem at hand is classifying the pixels into two classes (i.e. positive and negative), let *True Positive* (*TP*) be the number of pixels classified correctly as belonging to the positive class, *False Positive* (*FP*) the number of pixels misclassified as belonging to the positive class, *False Negative* (*FN*) the number of pixels misclassified as belonging to the negative class and finally *True Negative* (*TN*) the number of pixels classified correctly as belonging to the negative class. Then, the following metrics can be defined:

- *Correctness* or *Precision:* $P = TP / (TP + FP)$
- *Completeness* or *Recall :* $R = TP / (TP + FN)$
- *Quality:* $g = TP / (TP + FP + FN)$
- *Accuracy:* $A = (TP + TN) / (TP + TN + FP + FN)$
- *Effectiveness:* $F = 2PR / (P + R)$

All the aforementioned measures range from 0 to 1, with 1 denoting the best result (i.e. zero value for at least one type of classification error) and 0 denoting the worst possible result.

1.7.2 Qualitative performance evaluation

Apart from quantitative performance appraisal, it is very important that a real-world computer vision system is also assessed qualitatively. The reason for this is twofold; first it is impossible to construct an annotated dataset that includes all possible variations of a complex process like driving, and a frame-by-frame qualitative analysis of the results can offer valuable insight on possible improvements of a system. Furthermore, manual annotation is often subjective therefore the qualitative analysis could pinpoint frames where the result of a computer vision based system is appraised very strictly using the quantitative method. For these reasons, the systems presented in this thesis have undergone an extensive qualitative assessment process, with a special attention to adverse conditions, which is described separately in each chapter.

1.7.3 Processing speed performance assessment

Quality is not the only factor that is critical for ADAS. Obviously it is the most important factor, since such systems are used for the safety of human lives, but processing speed also plays a very important role in this context. More specifically, ADAS have to operate in real time; hence their processing speed must be faster than the frame acquisition rate of video cameras. Given the up-to-date typical camera frame rates, a goal of approximately 25 frames per second is what ADAS aim for. Depending on the specific details of each application this goal might be either a little lower (i.e. for city driving with low speeds), or even higher when highway driving must be handled. For all these reasons, a special section is included where applicable, covering the speed performance assessment for each system developed as part of this thesis.

1.8 Conclusions

This chapter presented and justified the current worldwide trend for the development of reliable Intelligent Transportation Systems and more specifically Advanced Driver Assistance Systems. Most of the modern commercially available ADAS based on computer vision methods have been reported. Furthermore, the state-of-the-art systems that utilize only visual information have been briefly examined and a general structure of vision-only ADAS has been proposed. Finally, the datasets used for the purposes of this thesis, as well as the metrics for the evaluation of the implemented systems have been described. In the next chapter, the core computer vision algorithms used for the development of the systems in this thesis will be presented.

Chapter 2

Computer Vision Algorithms

2.1 Introduction

This chapter contains a short presentation of the image processing concepts and algorithms that were used in the development of the systems described in this thesis. More specifically, the concept of shadow robustness and its importance in real world computer vision applications are analyzed and several color spaces that promise robustness to illumination changes are demonstrated. Consequently, the usage of image thresholding algorithms is described with a focus on the widely used Otsu algorithm. An analysis on the importance of symmetry for the detection of traffic signs and lights follows, concentrating a fast radial transform solution that provides an efficient way to tackle such problems. Then, the importance of motion information in problems related to ADAS is given and various optical flow algorithms are presented. A thorough discussion on the significance of image segmentation with a special mention to the Random Walker algorithm in ADAS is presented, including also its usage and implementation issues. Finally, an overview of popular methods for measuring image similarity is provided.

2.2 Color spaces and illumination invariance

The notion of illumination invariance is especially important for computer vision applications in the domain of real world problems. A system that can be used under all weather and illumination conditions should be robust to changes in environmental conditions, since the visual characteristics of a real world scene greatly vary depending on whether, for example, the picture is taken in a sunny day or a gloomy afternoon. Several approaches for different applications have been proposed in relevant literature [32], [33], with most of them suggesting the use of color spaces more robust to illumination changes than the normally used RGB color space. For the development of ADAS in this thesis, the CIE L*a*b* color space, the recently proposed $c_1c_2c_3$ color space and hybrid color spaces based on the two aforementioned ones have been used.

2.2.1 Comparison of color spaces for real world scenes

The main problem of the RGB color space when used for real world scenes is mainly the high correlation between its channels, which makes it vulnerable to illumination changes. On the contrary the CIE L*a*b* color space [34] is less affected by illumination changes and furthermore, it is characterized by a perceptual uniformity of colors, which is very useful for pattern recognition applications of man-made objects like traffic signs or traffic lights. The superiority of the CIE L*a*b* color space over RGB is demonstrated in Figure 10, where the red signs stand out much more in channel a* of the L*a*b* image, than the R coefficient of the RGB image.

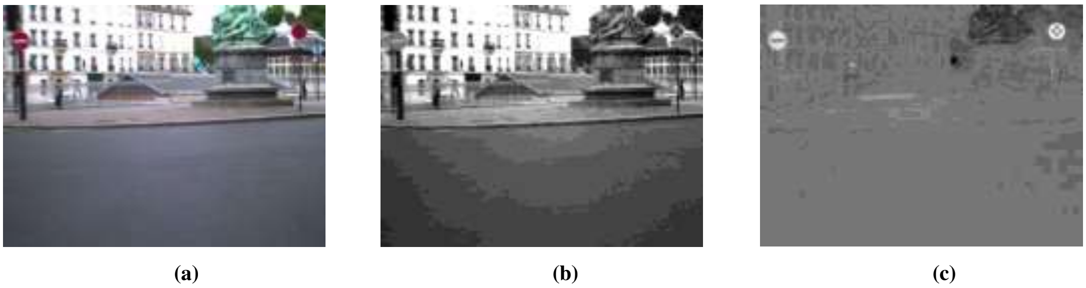

(a) (b) (c)

Figure 10 : (a) The original RGB image, (b) its R (red) channel and (c) its a* (red-green difference) channel.

Other color spaces that promise illumination intensity invariance are the HSV, YCbCr, and generally every color space that separates luminosity information from color information. The increasing interest in illumination intensity invariant color spaces due to the use of computer vision methods for real world applications, has led to a series of novel color spaces such as $c_1c_2c_3$ [35]. All the aforementioned color spaces are presented in Figure 11 in a scene with dense shadows.

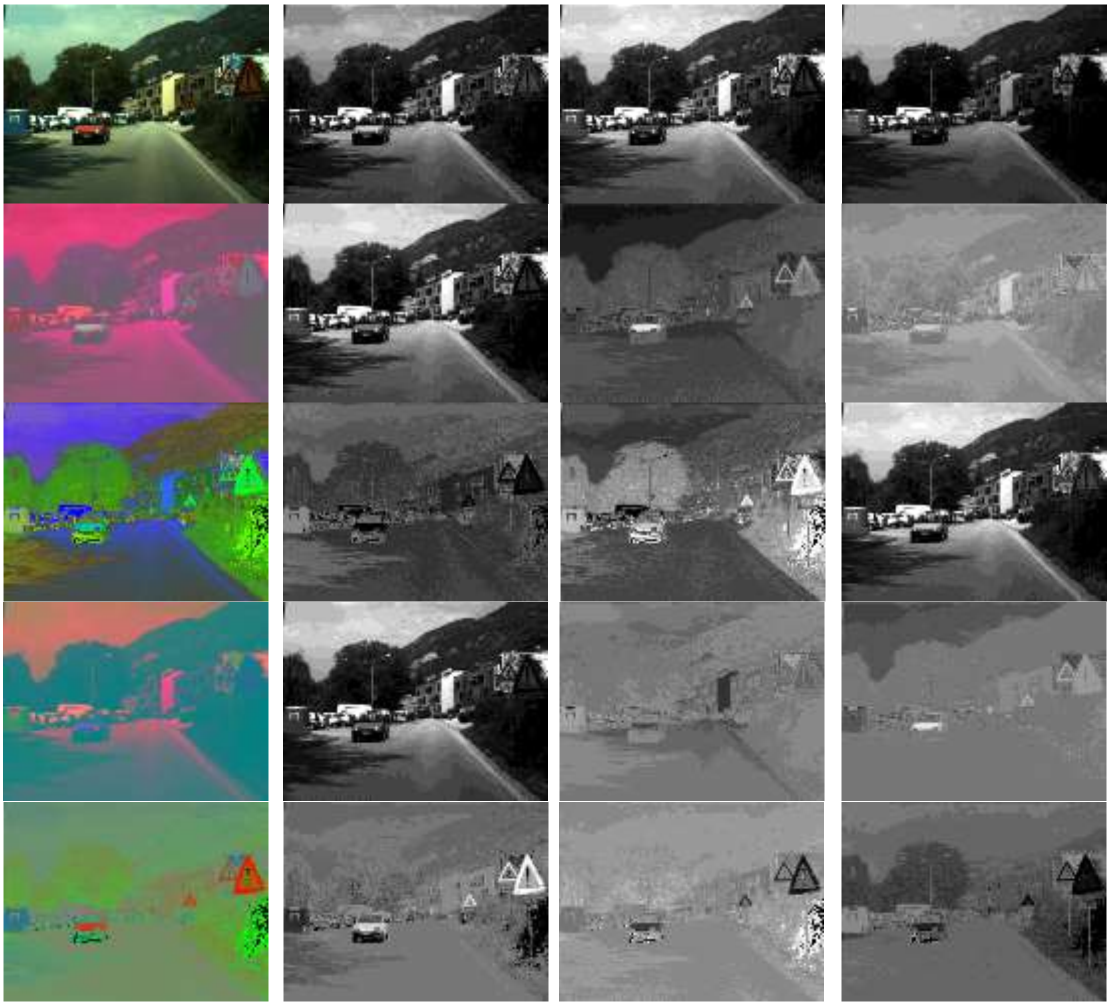


Figure 11 : The same scene in different color spaces. Each row contains the original image and the separate channels intensity: first row shows the RGB image, second the L*a*b*, third the HSV, fourth the YCbCr and fifth the $c_1c_2c_3$ image.

The different color channels represented in Figure 11 show that the color spaces that separate the intensity/brightness/luminosity channel from the chromaticity channels appear vulnerable in shadows only in the first channel. Another conclusion that can be drawn is that $c_1c_2c_3$ appears to separate colors more successfully than other color spaces. The worst results in color discrimination are drawn from the RGB color space, due to strong correlation of its channels.

2.3 Fast Radial Symmetry for efficient detection of symmetrical objects

The detection of human-made traffic-related objects appearing in a real-world scene relies greatly on the localization of symmetrical shapes. For systems aiming at detecting traffic signs, traffic lights or vehicle lights, a synergy of color and symmetry information provides a solid foundation for efficient detection. Since a short discussion about the color space influence to the object detection has already been covered, this section will concentrate on describing how symmetry can be used as a useful cue.

The selected algorithm in this thesis for efficient symmetry detection is the Fast Radial Symmetry Transform (FRST). This method is based on a simple and computationally efficient voting process which is described in [36]. The algorithm is designed for grayscale images which have to be examined for symmetrical shapes of various radii in a user-defined set, *N*. The value of the transform at range $n \in N$ indicates the contribution of the image gradients to the radial symmetry at a distance of *n* pixels from a pixel *p*, as shown in Figure 12 (taken from [36]).

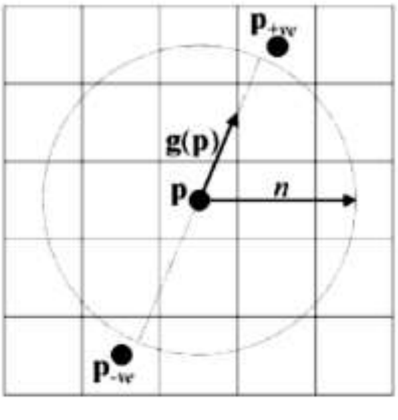


Figure 12 : Pixels affected by the gradient element g(p) for a range n=2.

The FRST is based on the examination of the intensity gradient *g* (using a 3x3 Sobel operator) at each image pixel *p*, which determines a positively affected pixel $p_{+ve}(p)$ and a negatively affected pixel $p_{-ve}(p)$ at a distance *n* away from *p*. The coordinates of the two affected pixels are defined by the orientation of the gradient; $p_{+ve}(p)$ is the pixel that the gradient is pointing to while $p_{-ve}(p)$ is the pixel that the gradient is pointing directly away from. More specifically, the positively affected pixel coordinates are estimated by

$$p_{+ve}(p) = p + round\left(\frac{g(p)}{\|g(p)\|}n\right), \tag{2.1}$$

while the negatively affected pixel coordinates are

$$p_{-ve}(p) = p - round\left(\frac{g(p)}{\|g(p)\|}n\right), \tag{2.2}$$

where "round" is a mathematical operation that rounds each element of the to the nearest integer.

Using the information of the positively and negatively affected pixels, two new images can be formed; the orientation projection image O_n and the magnitude projection image M_n. These two images are initially zero; for each pair of positively and negatively affected pixels that are calculated, the corresponding pixels in the orientation projection image are incremented and decremented respectively by 1. The same process is performed for the magnitude projection image, only this time the corresponding pixels are incremented and decremented respectively by $\|g(p)\|$. This process is summarized in the following equations:

$$O_n(p_{+ve}(p)) \leftarrow O_n(p_{+ve}(p)) + 1, \tag{2.3}$$

$$O_n(p_{-ve}(p)) \leftarrow O_n(p_{-ve}(p)) - 1, \tag{2.4}$$

$$M_n(p_{+ve}(p)) \leftarrow M_n(p_{+ve}(p)) + \|g(p)\|, \tag{2.5}$$

$$M_n(p_{-ve}(p)) \leftarrow M_n(p_{-ve}(p)) - \|g(p)\|. \tag{2.6}$$

Next, two more matrices are formed, using the following formulas:

$$F_n(p) = \frac{M_n(p)}{k_n}\left(\frac{|\tilde{O}_n(p)|}{k_n}\right)^a, \tag{2.7}$$

$$\tilde{O}_n(p) = \begin{cases} O_n(p), \text{ if } O_n(p) < k_n \\ k_n, \; otherwise \end{cases}, \tag{2.8}$$

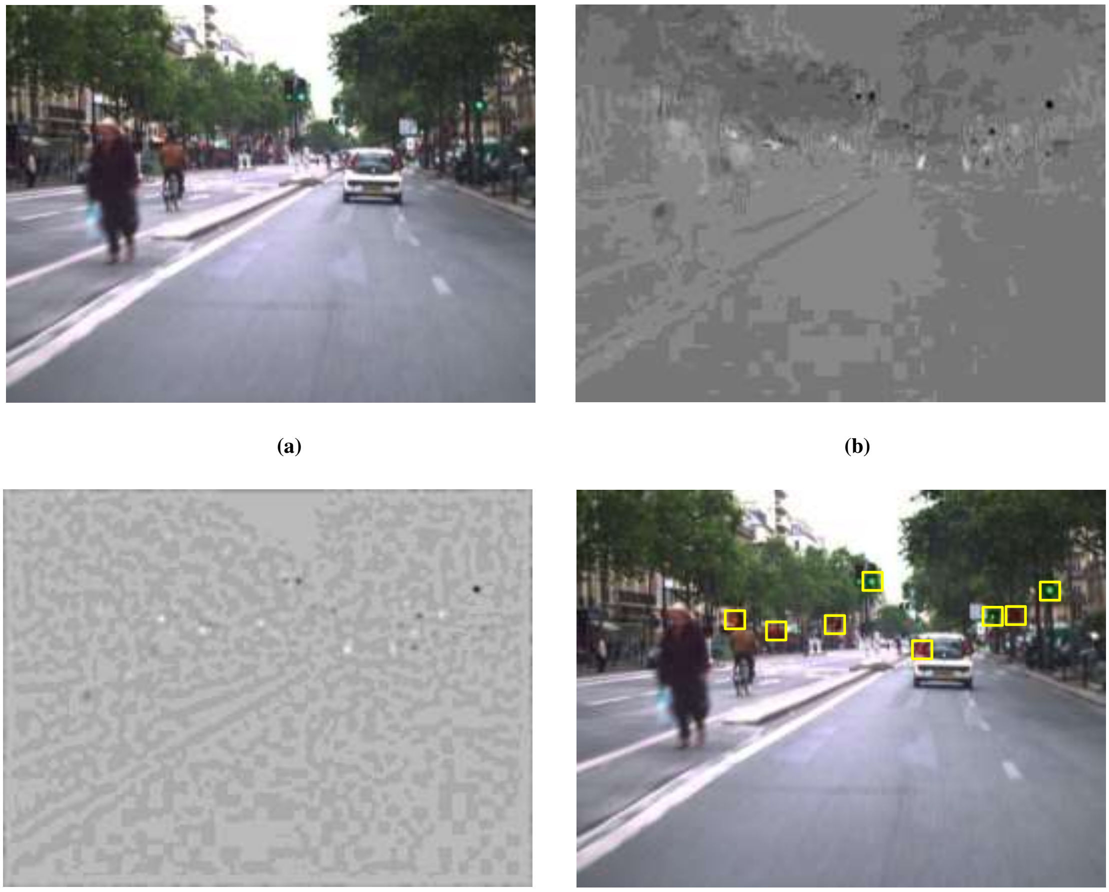
where a is a parameter that denotes radial strictness (the higher it is, the more "strict" the transform for non-radial symmetrical shapes) and k_n is a scaling factor for O_n and M_n across different radii, as defined in [36]. Then, for one specific radial n, the radial symmetry contribution is defined as

$$S_n = F_n * A_n, \tag{2.9}$$

where (*) denotes the convolution of the two matrices and A_n is a 2-dimensional Gaussian probability density function. Finally, when more than one radius is considered, the total symmetry transform is defined as the average of all the symmetry contributions over the radii range. If N denotes the radii considered, then the final symmetry transform matrix is

$$S = \frac{1}{N}\sum_{n \in N} S_n. \tag{2.10}$$

The resulting images acquired from the FRST when it is performed in chromaticity channels is particularly interesting for applications related to traffic signs or traffic lights detection, or even with the detection of the rear lights of vehicles. One typical example of all these attributes is shown in Figure 13, where the FRST in channel a* of the original image leads to a result that detects 3 green lights, one traffic sign and one vehicle rear light, while producing 3 false alarms. The FRST was estimated for radii of 3, 6 and 9 pixels with a radial strictness of 1 and then linearly normalized in the range [0, 1]. Values of the normalized FRST below 0.4 (dark spots) and over 0.9 (light spots) were used to extract the Regions Of Interest (ROIs) as shown in Figure 13(d).



(a) **(b)**

(c) **(d)**

Figure 13 : (a) Original image, (b) a* channel, (c) FRST result for N = {3, 6, 9} and α=1, (d) detected ROIs.

2.4 Otsu thresholding for efficient bimodal image segmentation

The problem of object detection in real world scenes often simplified into a bimodal segmentation task, or in more complex cases into multimodal segmentation using histogram information.

An example of this type of problems is met when the FRST contains objects that stand out from the background, so their efficient detection relies greatly on the usage of a thresholding method for images with bimodal histograms. The most popular fully automatic method for estimating the intensity threshold is proposed by Otsu's algorithm [37], which is based on the iterative splitting of the image pixels into two classes, with an ultimate goal to minimize their intra-class variance. As Otsu proved, this optimization criterion is equivalent to maximizing the inter-class variance. The method begins by splitting the image pixels into two classes, $Class_1$ containing the pixels with value less than or equal than a threshold th and $Class_2$ containing the pixels with value larger than th, thus making their probability distributions equal to:

$$\begin{aligned} Class_1: &\quad p_1/\omega_1(th),\ldots,p_{th}/\omega_1(th) \\ Class_2: &\quad p_{th+1}/\omega_2(th),\ldots,p_L/\omega_2(th) \end{aligned}, \qquad 2.11$$

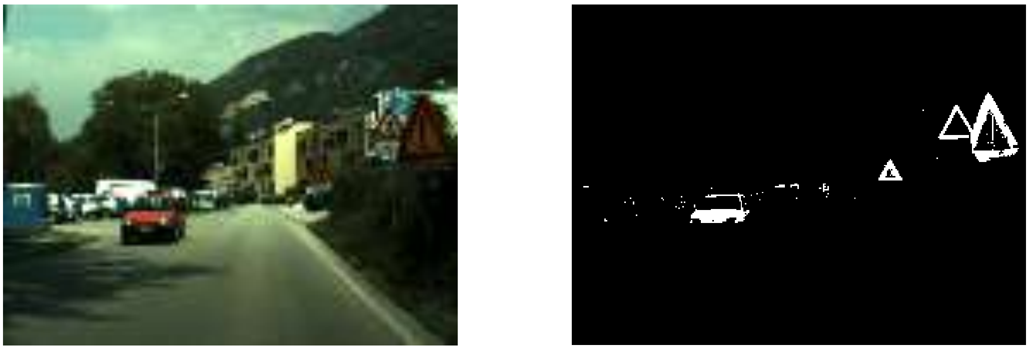
where $\omega_1(t)=\sum_{i=1}^{th}p_i$ and $\omega_2(t)=\sum_{i=th+1}^{L}p_i$.

The intra-class variance is defined as the weighted sum of variances of the two classes, which is given by

$$\sigma_\omega^2(th)=\omega_1(th)\sigma_1^2(th)+\omega_2(th)\sigma_2^2(th) \qquad 2.12$$

where $\sigma_\omega^2(th)$ is the intra-class variance and $\sigma_1^2(th)$, $\sigma_2^2(th)$ are the variances of the pixels below and above the threshold th, respectively. Otsu's method iteratively estimates the threshold th until the intra-class variance is minimized (or the inter-class variance is maximized).

Otsu’s algorithm was originally proposed for the segmentation of grayscale images. However, since color is a valuable cue for traffic signs, or lights detection, it is important to investigate the usage of the algorithm for the color coefficients. In this direction, the effect of Otsu’s algorithm in the detection of traffic signs is demonstrated in Figure 14, where the original image has first been converted to L*a*b* and its a* channel has been thresholded.

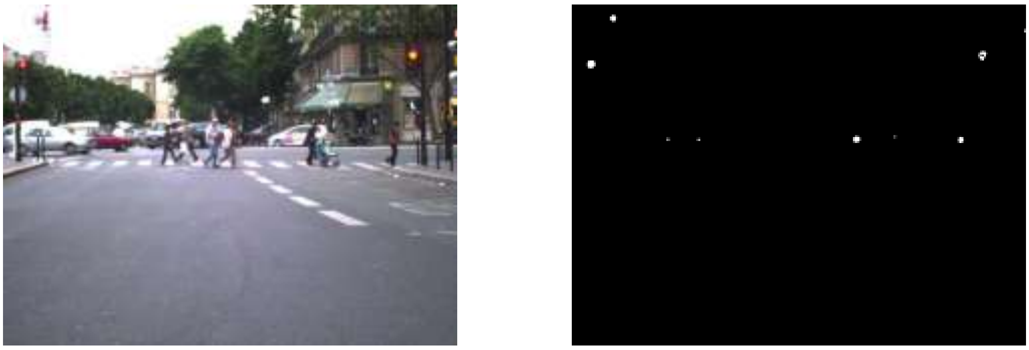


(a) **(b)**

Figure 14 : (a) Original image and (b) Otsu's thresholding on channel a* of the image.

However, color alone cannot always be enough for the detection of signs or lights, since many types of objects in real world scenes chromatically resemble traffic lights. This is why alternative approaches could be used, like the utilization of Otsu's thresholding algorithm for a more efficient detection of symmetrical shapes within an image. As was shown in the previous section, the FRST result needs to be segmented through histogram thresholding in order to select the most prominent results. Such a process is shown in Figure 15, where the original image has been converted to

L*a*b* and the FRST is applied to channel a*. The transformed image is then segmented using Otsu's threshold estimation algorithm to produce the result presented in Figure 15(b) denoting the possible ROIs for red lights.



(a) (b)

Figure 15 : (a) Original image and (b) result of Otsu's thresholding on the FRST of channel a*.

2.5 Optical Flow for moving objects detection

The methods described so far are especially useful for the detection of human-made symmetrical objects in static images, but fail to take advantage of the most useful cue for such applications which is motion information. In driving scenes acquired from on-board cameras, the images contain a great number of moving objects, since there are not only absolute motions to be considered, but also relative motions to the velocity of the ego-vehicle.

A very common methodology to derive motion information of an object is to estimate its optical flow, which is defined in [38] as the *"approximation to image motion defined as the projection of velocities of 3-D surface points onto the imaging plane of a visual sensor"*. Various algorithms have been proposed for such an approximation, each with its pros and cons. A general rule that could be drawn by the relevant literature [39] would be that there is a trade-off between processing speed of such algorithms and their accuracy. Optical flow algorithms that offer dense, accurate results usually tend to have prohibiting computational complexity for real-time applications. On the other hand, faster algorithms that produce sparse optical flow results cannot be trusted to provide an accurate approximation that could be used for tasks like motion-based segmentation of objects.

In this thesis, the sparse optical flow proposed by Horn and Schunck in [40] is adopted; the algorithm is based on the assumption that optical flow is smooth over the whole image. Consequently, the method iteratively aims to minimize flow fluctuations favoring solutions that produce a more smooth optical flow result. Let I_x, I_y and I_t be the derivatives of the image intensity values of a grayscale image along the horizontal, vertical and time dimensions respectively, estimated from

$$I_x = \frac{1}{4}\left(I_{(i,j+1,l)} - I_{(i,j,l)} + I_{(i+1,j+1,l)} - I_{(i+1,j,l)} + I_{(i,j+1,l+1)} - I_{(i,j,l+1)} + I_{(i+1,j+1,l+1)} - I_{(i+1,j,l+1)}\right), \quad 2.13$$

$$I_y = \frac{1}{4}\left(I_{(i+1,j,l)} - I_{(i,j,l)} + I_{(i+1,j+1,l)} - I_{(i,j+1,l)} + I_{(i+1,j,l+1)} - I_{(i,j,l+1)} + I_{(i+1,j+1,l+1)} - I_{(i,j+1,l+1)}\right), \quad 2.14$$

$$I_t = \frac{1}{4}\left(I_{(i,j,l+1)} - I_{(i,j,l)} + I_{(i+1,j,l+1)} - I_{(i+1,j,l)} + I_{(i,j+1,l+1)} - I_{(i,j+1,l)} + I_{(i+1,j+1,l+1)} - I_{(i+1,j+1,l)} \right), \qquad 2.15$$

where i, j, l are the row, column and frame number respectively. Then, the horizontal and vertical optical flow values at each pixel with coordinates i and j at time $k+1$, namely V_x^{k+1} and V_y^{k+1} are iteratively derived by

$$V_x^{k+1} = \bar{V}_x^{k} - \frac{I_x(I_x\bar{V}_x^{k} + I_y\bar{V}_y^{k} + I_t)}{\alpha^2 + I_x^2 + I_y^2}, \qquad 2.16$$

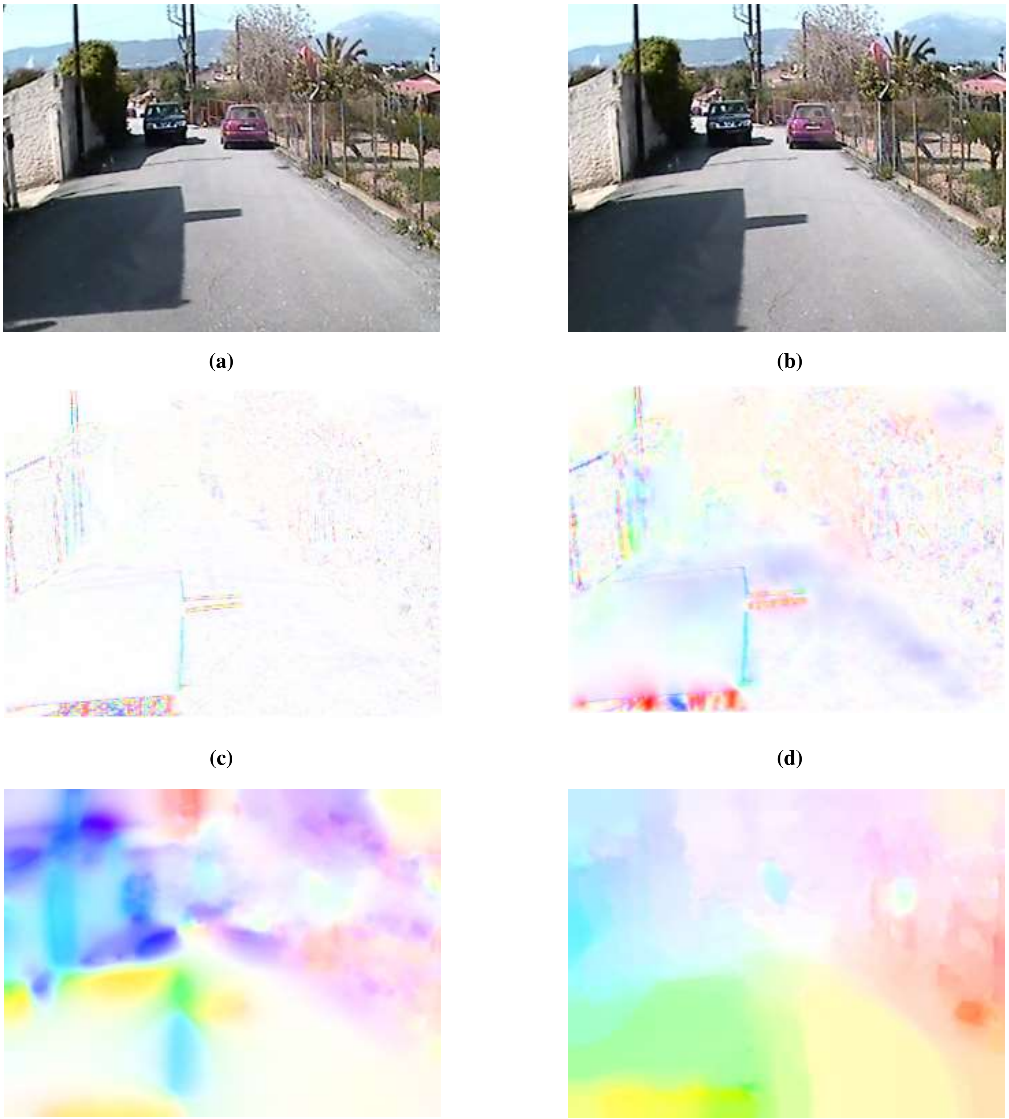
and

$$V_y^{k+1} = \bar{V}_y^{k} - \frac{I_y(I_x\bar{V}_x^{k} + I_y\bar{V}_y^{k} + I_t)}{\alpha^2 + I_x^2 + I_y^2}, \qquad 2.17$$

where $\bar{V}_x^k$ and $\bar{V}_y^k$ are the average values of the previous velocity estimates for axes x and y respectively, α is a regularization constant which leads to smoother flow results when increased and k is the number of iterations. For small values of k and a, the algorithm produces quite sparse results, which get denser as those values increase.

Even though the Horn-Schunck algorithm produces denser results than other local optical flow methods, it is not as successful as more recent methods, like the one in [41]. In his work, C. Liu proposes a mixture of local and global methods first introduced in [42] and [43], with the difference of using conjugate gradients instead of Gauss-Seidel or over-relaxation methods for solving large linear systems. This type of optical flow algorithms produce more accurate results than the Horn-Schunck approach, but the implementation of such algorithms requires more processing power. However, when real life applications are considered, even the dense optical flow algorithms have problems discriminating between real objects and shadows, as shown in Figure 16. To demonstrate this problem for the two consecutive frames of Figure 16 (a) and (b), the algorithm of Horn and Schunck as described in [40] was applied, first using a=3 and only one iteration (k=1) to acquire the sparse result of Figure 16(c) and then using a larger value of a=15 and many iterations (k=1500) to acquire the result of Figure 16(d). Then, the dense algorithm of [41] was applied, for few outer/middle/inner loop iterations (1,1,1) that led to the result of Figure 16(e) and then for more iterations (15,15,15) that led to the result of Figure 16(f).

By examining the results the first conclusion derived is that the method of Horn-Schunck does indeed produce a more sparse result, with a tendency of estimating larger flow values near the boundaries of objects. The sparseness is reduced when the number of iterations rises, but the algorithm still produces flow values at the edges of shadows that are significantly different than those inside the shadows. The algorithm of Liu does improve the situation providing denser flow results that make more physical sense. However, even this algorithm fails to produce meaningful results in the presence of shadows, even if the number of iterations used is so large that the algorithm cannot be used for real-time applications using typical embedded systems.



(a) (b)

(c) (d)

(e) (f)

Figure 16 : Two consecutive frames with dense shadows in (a) and (b) produce the pseudo-colored optical flows of (c) for Horn-Schunck with a=3 and k=1, (d) for Horn-Schunck with a=15 and k=1500, (e) C. Liu for iterations 1,1,1 and (f) C. Liu for iterations 15,15,15.

2.6 Discrete Time Kalman Filter for object tracking

Another method that is very popular in computer vision applications is Kalman filtering. A Kalman filter is essentially an estimator that, in this context, can be used to predict the position of an object in future frames of a video. Kalman tracking can provide multiple advantages for any video-based computer vision application, since it can reduce the area of the frame that is scanned for the existence of an object thus minimizing false positive detections and reducing the time needed for the process. An added advantage offered by the very nature of the Kalman filter, is the smoothing effect that can improve the robustness of the tracking result to the presence of noise that is often observed in videos.

2.6.1 Definition of the discrete process model

For the purposes of this thesis, only the discrete time Kalman filter [44], [45], [46] will be considered. This filter provides a solution to the problem of estimating the state $x \in \Re^n$ of a linear discrete time process that abides by the stochastic difference equation 2.18, given a measurement $z \in \Re^m$ described by equation 2.19:

$$x_k = Ax_{k-1} + Bu_k + w_{k-1}, \tag{2.18}$$

$$z_k = Hx_k + v_k, \tag{2.19}$$

where x_k and z_k are the vectors representing the model state and measurement at the discrete time step *k*, respectively, while A, B and H are the transition, control and measurement matrices of the model. The transition matrix *A*, is sized *n* x *n* and connects the previous process state (at time *k* - 1) to the current state (at time *k*). Matrix *B* is *n* x *l* and it is optional. Its function is to relate the control input $u \in \Re^l$ to state *x*. Finally, *H* is a *m* x *n* matrix that relates the process state to the measurement, z_k. All these matrices may change at each time step, but they are generally assumed to be constant.

Typically, the noises of the state and measurement are considered to be white, zero mean Gaussian, statistically independent to each other and represented by w_k and v_k respectively. Their probability distributions can be denoted as

$$p(w) \sim N(0,Q), \tag{2.20}$$

$$p(v) \sim N(0,R), \tag{2.21}$$

where the *process noise covariance matrix, Q* and the *measurement covariance matrix, R* can also change over time in applications where the noise parameters are changes through time. In most studies they are assumed constant.

2.6.2 Algorithm of the Discrete Kalman Filter

As already mentioned, the Kalman filter is an estimator of a process that relies on a recursive, feedback-based algorithm. Once a prediction has been provided by the filter at a certain point in time, feedback is given by measurements which could contain noise. Apparently, this divides the Kalman filtering process into two stages, namely the ***time update*** and the ***measurement update***. The former projects the current process state and error covariance (*a-priori*) estimates to the next step, while the latter provides the feedback from measurements in order to improve the new (*a-posteriori*) estimates.

Let $\hat{x}_k^- \in \Re^n$ be the *a-priori* state estimate and $\hat{x}_k \in \Re^n$ the *a-posteriori* estimate at time *k*, given measurement z_k; the *a-priori* and *a-posteriori* estimate errors are then defined as

$$e_k^- \equiv x_k - \hat{x}_k^- \quad \text{and} \tag{2.22}$$

$$e_k \equiv x_k - \hat{x}_k, \tag{2.23}$$

leading to the *a-priori* and *a-posteriori* estimate error covariance matrices that follow

$$P_k^- = E[e_k^- (e_k^-)^T] \quad \text{and} \tag{2.24}$$

$$P_k = E[e_k (e_k)^T], \tag{2.25}$$

where *E* denotes the variance.

The ***time update equations*** of the discrete Kalman filter can be defined as

$$\hat{x}_k^- = A\hat{x}_{k-1} + Bu_k + w_{k-1}, \tag{2.26}$$

$$P_k^- = AP_{k-1}A^T + Q. \tag{2.27}$$

The ***measurement update equations*** will be

$$K_k \equiv P_k^- H^T (HP_k^- H^T + R)^{-1}, \tag{2.28}$$

$$\hat{x}_k = \hat{x}_k^- + K_k(z_k - H\hat{x}_k^-), \tag{2.29}$$

$$P_k = (I - K_k H)P_k^-. \tag{2.30}$$

The matrix K_k is called the Kalman gain and it is the first parameter of the filter estimated in the measurement update phase.

The complete Kalman filtering algorithm is a recursive succession of the time and measurement update steps. The order of the steps involved in the Kalman filtering process and the interaction between the two phases are described more closely in Figure 17, as presented in [46].

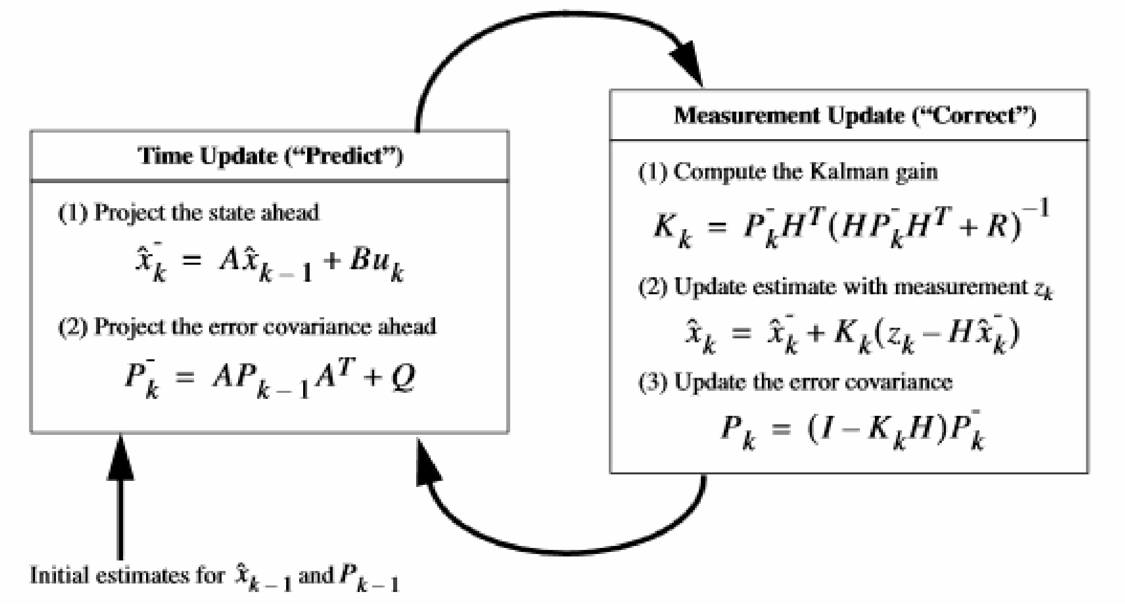

Time Update ("Predict")
(1) Project the state ahead
$\hat{x}_k^- = A\hat{x}_{k-1} + Bu_k$
(2) Project the error covariance ahead
$P_k^- = AP_{k-1}A^T + Q$
Measurement Update ("Correct")
(1) Compute the Kalman gain
$K_k = P_k^- H^T(HP_k^- H^T + R)^{-1}$
(2) Update estimate with measurement z_k
$\hat{x}_k = \hat{x}_k^- + K_k(z_k - H\hat{x}_k^-)$
(3) Update the error covariance
$P_k = (I - K_k H)P_k^-$
Initial estimates for $\hat{x}_{k-1}$ and P_{k-1}

Figure 17: Analysis of the operation of the Kalman Filter as presented in [46].

2.7 Random Walker for efficient image segmentation

The Random Walker Algorithm (RWA) as an image segmentation tool is presented in [47]. It is based on a random process, in which a walker starting from a pixel is moving to neighbor pixels according to a probability estimated using the intensity difference between the two pixels. If a set of pixels is associated to M segmentation classes, denoted also as seed pixels, the RWA estimates at each pixel the probabilities of a walker starting moving towards pixel positions to arrive at a seed. In most applications the departure pixel is annotated to the class with the maximum probability of first-arrival. Seeds are the pixels that are manually defined to certainly belong in one of the segmentation classes. The Random Walker segmentation process that is proposed in [47] is based on a closed form solution of the Dirichlet problem, yielding a faster solution while producing the same accuracy results as the classic RWA of [48], [49]. The main advantages of the RWA [47] for real life problems like road detection are the following:

- it is robust to noise, therefore suitable for real life applications,
- it locates weak or missing boundaries that are common in road scenes,
- it is quite fast in the implementation presented in [47],
- it can be used for multi-object segmentation problems, thus enabling its generalization to problems like moving obstacles detection and
- it provides a road probability matrix instead of a road mask, allowing further post-processing of the probabilities matrix in different applications, i.e. segmenting the road surface to safe or unsafe areas.

The RWA consists of the following successive steps:

1) First, the image is transformed to a graph by transforming the brightness differences of each pixel and its neighboring pixels to the edges of a fully connected graph. When color or multimodal images are concerned, the entropy maximization rule using the Gaussian weighting function leads to:

$$w_{ij} = \exp(-\beta \| I_i - I_j \|_2), \qquad 2.31$$

where I_i and I_j are the color channel vector values of pixels i and j respectively. The square gradients are linearly normalized to [0, 1] before applying 2.31. The only user-defined parameter of the RWA is β.

2) The next step is the construction of a linear system of equations the solution of which derives, for each unlabeled pixel, the road and non-road probabilities. To construct the system of equations, first the graph Laplacian matrix is defined as

$$L_{ij} = \begin{cases} d_i & \text{if } i = j, \\ -w_{ij} & \text{if } u_i \text{ and } u_j \text{ are adjacent nodes}, \\ 0 & \text{otherwise}, \end{cases} \qquad 2.32$$

where L_{ij} is the value of the pixel indexed by vertices u_i, u_j and $d_i = \sum w(e_{ij})$ is the degree of a vertex for all edges e_{ij} incident on u_i.

3) The Dirichlet integral is

$$D[x]=\frac{1}{2}x^T Lx=\frac{1}{2}\sum_{e_{ij}\in E} w_{ij}(x_i - x_j)^2, \qquad 2.33$$

where x is a combinatorial harmonic, i.e. a function that minimizes (2.33) and e_{ij} is an edge of the graph spanning vertices u_i and u_j .

4) The graph vertices are split into two classes, V_S (seeded nodes) and V_U (unseeded nodes). Without loss of generality, the pixels in vectors L and x are sorted into two sets, the seeds and the unseeded pixels.

5) Then, decomposition of equation 2.33 is performed:

$$D[x_U]=\frac{1}{2}\left(x_S^T x_U^T\right)\begin{bmatrix} L_S & B \\ B^T & L_U \end{bmatrix}\begin{bmatrix} x_S \\ x_U \end{bmatrix}, \qquad 2.34$$

where L_S are the seeded and L_U the unseeded pixels of L, respectively, B is a result of the decomposition and x_U, x_S are the sets of road probabilities corresponding to unseeded and seeded pixels respectively. Differentiating $D[x_U]$ with respect to x_U, yields

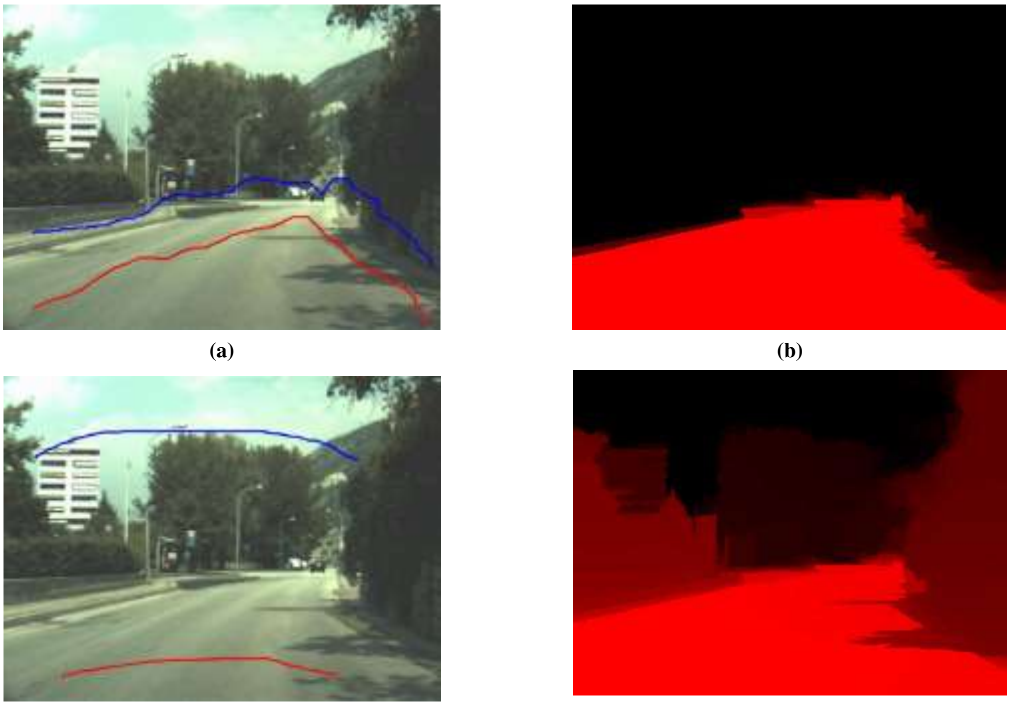
$$L_U x_U = -B^T x_S. \qquad 2.35$$

6) Assuming a two-class segmentation problem, if the road seeds are denoted by V_R and the non-road (background) seeds by V_B, where $V_R \cap V_B = \varnothing,\ V_R \cup V_B = V_S$, then the probabilities x_i^R that a random walker leaving pixel u_i arrives at a pixel in V_R before arriving at a pixel in V_B can be computed by

$$L_U x^R = -B^T m^R, \qquad 2.36$$

where m_R = 1 for road seeds and m_R = 0 for non-road seeds. The annotation rule assigns each pixel to the class with the maximum probability. In the two-class road segmentation problem this is equivalent to assigning a pixel i to the road class, if $x_i^R > 0.5$.

Since in most applications reliable definition of a sufficient number of seeds that are properly placed leads the RWA to produce accurate segmentation results, similar results would be expected in a two-class segmentation problem like road detection. A typical example of the importance of seeds placement is shown in Fig. 1, where two different manual seed placements lead the RWA to significantly different results. When the road and non-road seeds are placed close to the road borders, the segmentation result is superior and can be easily thresholded (Figure 18(a), (b)). On the contrary, a less precise seed placement leads to a result with raised ambiguity that is not easy to threshold and appears problematic in shadows as shown in Figure 18(c), (d).



(a) (b)

(c) (d)

Figure 18 : Seeds definition (left column) and segmentation result (right column). Road seeds are denoted with the red line, while non-road seeds with the blue one. In the segmented images, red color annotates the road pixels and black annotates the background.

2.8 Measures for object similarity assessment

So far we have presented and analyzed algorithms for image segmentation or object detection. Often, the detected objects in an image have to be compared to some templates so that their similarity is verified. In this section, the selected similarity matching techniques for the purposes of this thesis are presented.

2.8.1 Normalized Cross-Correlation (NCC)

Given two images, a template T and a candidate C, their similarity can be assessed using the normalized cross-correlation measure. This measure is estimated as described in [50], [51], by

$$\gamma(i,j)=\frac{\sum_{x,y}[C(x,y)-\overline{C}_{i,j}][t(x-i,y-j)-\overline{T}]}{\sqrt{\sum_{x,y}[C(x,y)-\overline{C}_{i,j}]^2\sum_{x,y}[T(x-i,y-j)-\overline{T}]^2}}, \qquad 2.37$$

where $\overline{T}$ is the mean of the template and $\overline{C}_{i,j}$ is the mean of image $C(x, y)$ in the area under the template. The NCC of two identical images will produce a result with maximum value of 1. NCC can also be used for template matching, i.e. the detection of a specific object in an image. In that case, the coordinates of the maximum value of the NCC will be localized on the centre of the region where the template lies in the image.

2.8.2 *Mean Absolute Error (MAE)*

The metric of MAE is also a popular measure to determine if two images *T* and *C* are similar. The metric is given by

$$MAE = \frac{1}{M \times N} \sum_{i=1}^{M} \sum_{j=1}^{N} \left| C_{ij} - T_{ij} \right| , \qquad 2.38$$

where *M, N* are the number of rows and columns of the two images respectively. Two identical images will result in an error that is equal to zero.

2.8.3 *Structural Similarity Index (SSIM)*

A more complicated metric that combines luminance, contrast and structural information in one single index, is the SSIM introduced in [52]. The SSIM is estimated by

$$SSIM(C,T) = \frac{(2\mu_C \mu_T + K_1)(2\sigma_{CT} + K_2)}{({\mu_C}^2 + {\mu_T}^2 + K_1)({\sigma_C}^2 + {\sigma_T}^2 + K_2)}, \qquad 2.39$$

where μ_C, μ_T are the average intensity values of the two images, σ_C, σ_T are the standard deviations of the two images, σ_{CT} is the covariance of the two images and K_1, K_2 are two constants that ensure that the denominator will never be zero. K_1 and K_2 are calculated using: $K_1 = (k_1 L)^2$ and $K_2 = (k_2 L)^2$, where L is the dynamic range of the pixel values (255 for 8-bit images) and k_1, k_2 are very small constants. The value of SSIM for identical images is 1.

2.9 Conclusions

In this chapter, a presentation of all computer vision algorithms that are used throughout this thesis is given. Color space selection and illumination invariance were discussed, bearing in mind that the overall goal is to develop ADASs that perform efficiently under diverse weather and illumination conditions. Fast symmetry detection was then discussed, using the FRST as selected algorithm, because of its speed and effectiveness in detecting bright and dark symmetrical shapes in images. Histogram-based bimodal image segmentation using the popular Otsu's method was also presented and connected to both color segmentation and symmetrical shapes detection problems. Motion also plays an important role in ADAS, so the incorporation of optical flow algorithms for motion estimation was discussed, with a focus on the issue of robustness to shadows and processing complexity, which are a common problem in applications of this nature. The most important algorithm utilized in this thesis was described in the next section; the RWA is a fast and noise-tolerant image segmentation algorithm and it is used in the road detection sub-system. Finally, measures used for the comparison of two images are described, as they are used in template matching and candidate verification purposes in several sub-systems.

Chapter 3

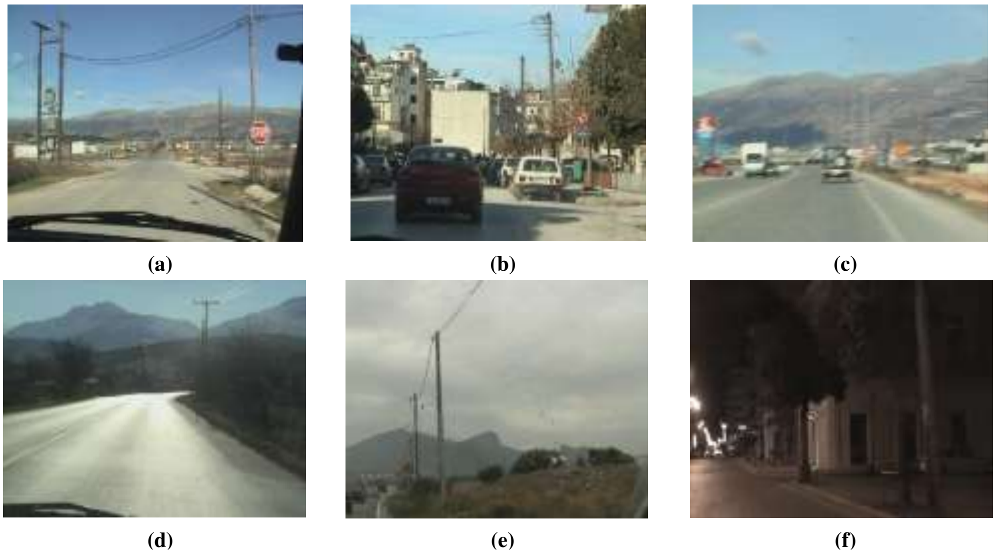
Traffic Sign Recognition System

3.1 Introduction

In this chapter we present the first system developed for the purposes of this thesis, which is the Traffic Sign Recognition (TSR) system published in [53]. A small analysis on the importance and challenges of such a system in a vehicle is presented followed by the state-of-the-art algorithms that provide solutions to this problem. The next section presents the structure of the proposed system and then its modules are analyzed and explained. The system is then evaluated in terms of accuracy and the conclusions and future work suggestions follow.

3.2 The Traffic Sign Recognition problem

The need for an efficient TSR system is very closely related with the high numbers of road accidents that are caused by the driver's inability to focus on the information conveyed by traffic signs. Especially signs that inform about dangers, or speed limits, as well as the signs that prohibit some action are extremely necessary for secure driving. However, the vast amount of visual and other kinds of information that is brought into the attention of drivers at any given time often distracts their attention from what is really important for their driving reactions. Furthermore, the great diversity of environmental conditions and driving scenes can deteriorate drivers' visibility and lead to neglecting to focus on Traffic Signs (TS). Some typical examples of the diversity of conditions met are shown in Figure 19.


(a) (b) (c)

(d) (e) (f)

Figure 19 : Traffic Signs under different environmental conditions: (a) excellent weather and illumination angle, (b) occlusions, (c) shaky camera, (d) driving against the sun, (e) rainy conditions and (f) night driving.

TSR can be divided into two discrete problems; namely, the Traffic Sign Detection (TSD) and the Traffic Sign Classification (TSC). The very useful Tracking of Traffic Signs (TST) can be used to enhance the performance of the two aforementioned problems both in accuracy and in computational complexity. What is needed in a trustworthy TSR system is for it to be accurate, minimizing the false alarms rate and furthermore be able to perform efficiently even in challenging driving scenarios and adverse conditions.

The full set of traffic signs can be split into three main groups: danger proclamation signs, traffic regulation signs and informational signs. Signs that belong to the first group are placed to warn drivers of the dangers that exist ahead on the road, so they can anticipate them. The second group comprises signs that inform the drivers of the special obligations, restrictions or prohibitions they should conform to. The signs of the third group provide information that assists the driver in the navigation task, such as junctions, distances etc. As one can understand by examining the full set of traffic signs in Figure 20, signs are designed to assist drivers in spotting them easily in natural scenes. This is achieved by selecting colors and shapes that differentiate the signs from the background. Consequently, the main colors that are used are red, blue, yellow and green, with black or white ideograms. The shapes of the signs are symmetrical. Triangles, circles, octagons, diamonds and rectangles are used. The shape and colors of a TS, along with the ideogram that it contains define its significance.

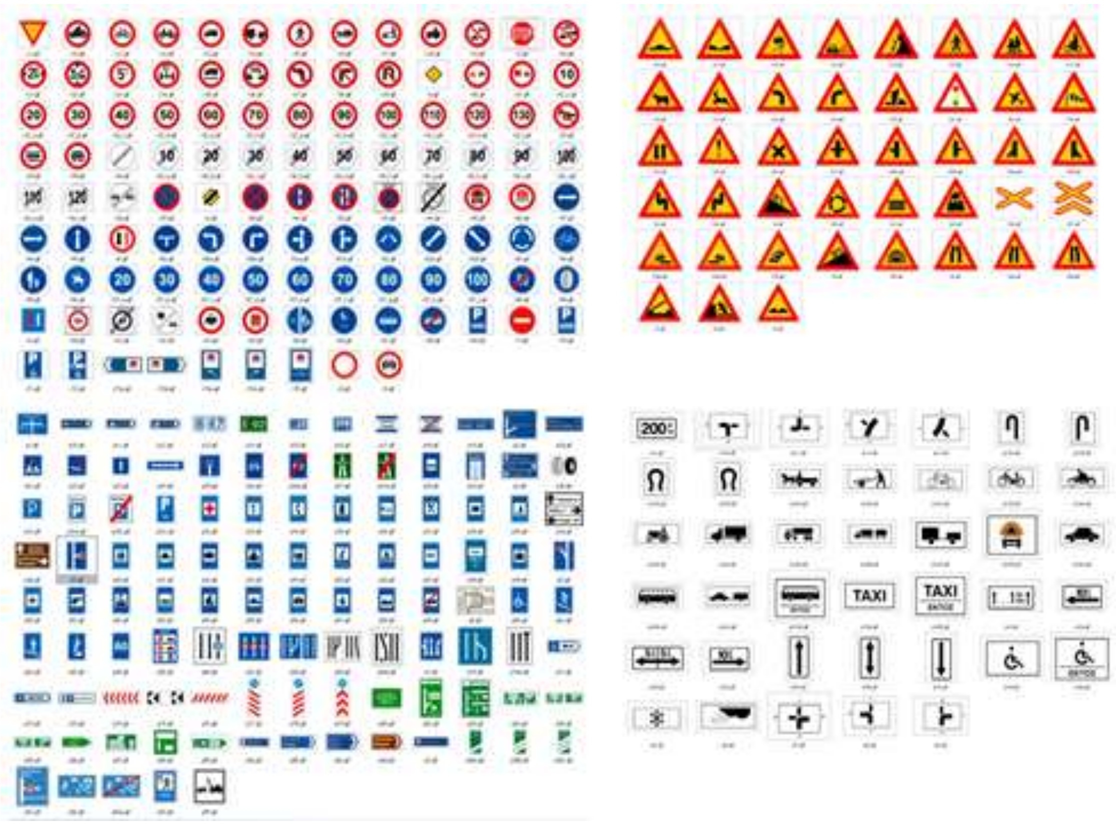


Figure 20 : Complete set of traffic signs.

3.3 Related work

The TSR problem has attracted the attention of many researchers over the past decade. The development of a system that can robustly detect and classify TSs in real time, has twofold benefits; it can be used in ADAS, assisting the driver to focus more on the navigation of the vehicle by providing the information given by the signs but also, such systems can be embedded in fully autonomous vehicles that travel in the existing road infrastructure. Naturally, in order for these systems to be functional, they must have a number of advantages, such as:

1. They must be resilient to any change in lighting or weather conditions.
2. They need to be able to recognize partially occluded signs, as well as signs that are either rotated, or not exactly perpendicular to the camera axis.
3. These systems must be robust to the deterioration of the color of some signs, usually due to their age and bad weather conditions.

A very important cue for efficient TSR is the TS color information. However there are some researchers who prefer not to use it. In those implementations where grayscale images are preferred, the TRD is based mainly on morphology features, such as symmetry [54], [55], [56], distance transformations from offline generated templates [57] and pyramidal structures for border detection [58], [59], [60]. A machine learning approach to TSR using genetic algorithms and neural networks is proposed in [61]. In color based TSR methods, proper selection of color space plays a vital role. In [62], [63], [64], the standard RGB color space, or the ratios between its color coefficients are used. However, RGB is not the ideal color space for real-world problems such as TSR, because it is very susceptible to lighting changes due to the high correlation of its channels. Thus, color spaces that are less sensitive to such changes are preferred; such channels are the HSI, which is used in [65], [66], [67], [68], [69], [70], [71], or the LUV space in [72]. After color segmentation is performed, road signs are detected using circular and triangular shape matching [65], neural networks [66], [71] and genetic algorithms [68], [69]. Simulated annealing is used together with genetic algorithms [70].

Among the most important tracking algorithms used in TST, the authors of [71] use Kalman filters, while Kalman-Bucy filtering is utilized in [57] and a motion-model plus temporal information propagation is proposed in [63]. For TSC, the most popular methods are based on template matching, either by cross-correlation [65], [70] or by matching pursuit filters [67]. Various types of neural networks for TSC have been used in [64], [66], [69], [72] and a Bayesian generative modeling is described in [63].

3.4 TSD system structure

The proposed system performs road-sign detection, tracking, and classification. It is founded upon a mixture of widely used methods, like the FRST of [36], and partly modified ones, such as a modified Otsu threshold of [37], used here for color segmentation. Moreover, a number of novel features are introduced, such as the use of CIE-L*a*b color space for the color image processing stage, and sign tracking in multiple frames by examining previously selected sub-windows of each frame. The structure of the proposed TSD system is shown in Figure 21.

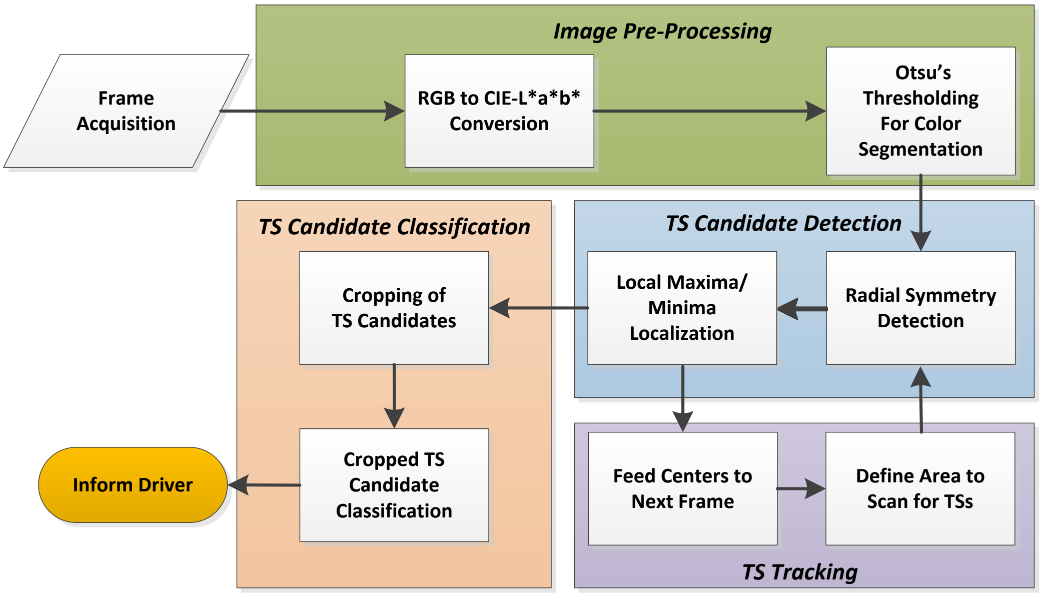


Image Pre-Processing
Frame Acquisition
RGB to CIE-L*a*b* Conversion
Otsu's Thresholding For Color Segmentation
TS Candidate Classification
TS Candidate Detection
Cropping of TS Candidates
Local Maxima/ Minima Localization
Radial Symmetry Detection
Inform Driver
Cropped TS Candidate Classification
Feed Centers to Next Frame
Define Area to Scan for TSs
TS Tracking

Figure 21 : Structure of the TSR system

The system processes each image by converting it to L*a*b format and then by performing color segmentation. The segmented frame is further processed by the TSD and TST modules. The centers and radii of all the detected road signs are used in the tracking stage for the next frame and are also used to crop the signs for classification.

3.4.1 Color Space Selection - Color Segmentation

For the purposes of the TSR system, the CIE L*a*b color space is chosen because of its property of having two color-opponency channels (a* and b*) that represent perceptual color differences which are also linearly spaced. The illumination alterations in an image affect mostly the L^* (Lightness) channel of the L*a*b space. Thus, an examination of the L^* channel of an image provides information on the brightness conditions in which it was acquired. This property can prove beneficial for the segmentation process, as the borders of traffic signs typically appear lighter than the mean luminosity in dark scenes, and darker than the mean luminosity in well-lit scenes. This can be especially useful in night driving conditions, as it filters out much of the background scene.

The image thresholding algorithm proposed by Otsu et al in [37] and described in Section 2.4, is used to transform the Lightness channel to a binary image denoted as L_{bo}. Also, four binary images are estimated by bisecting the positive and negative part of a^* and b^* channels, to acquire four chromatic subspaces; the negative subspaces are then multiplied by -1 to ensure positive definition for all channels. This process is summarized in Figure 22.

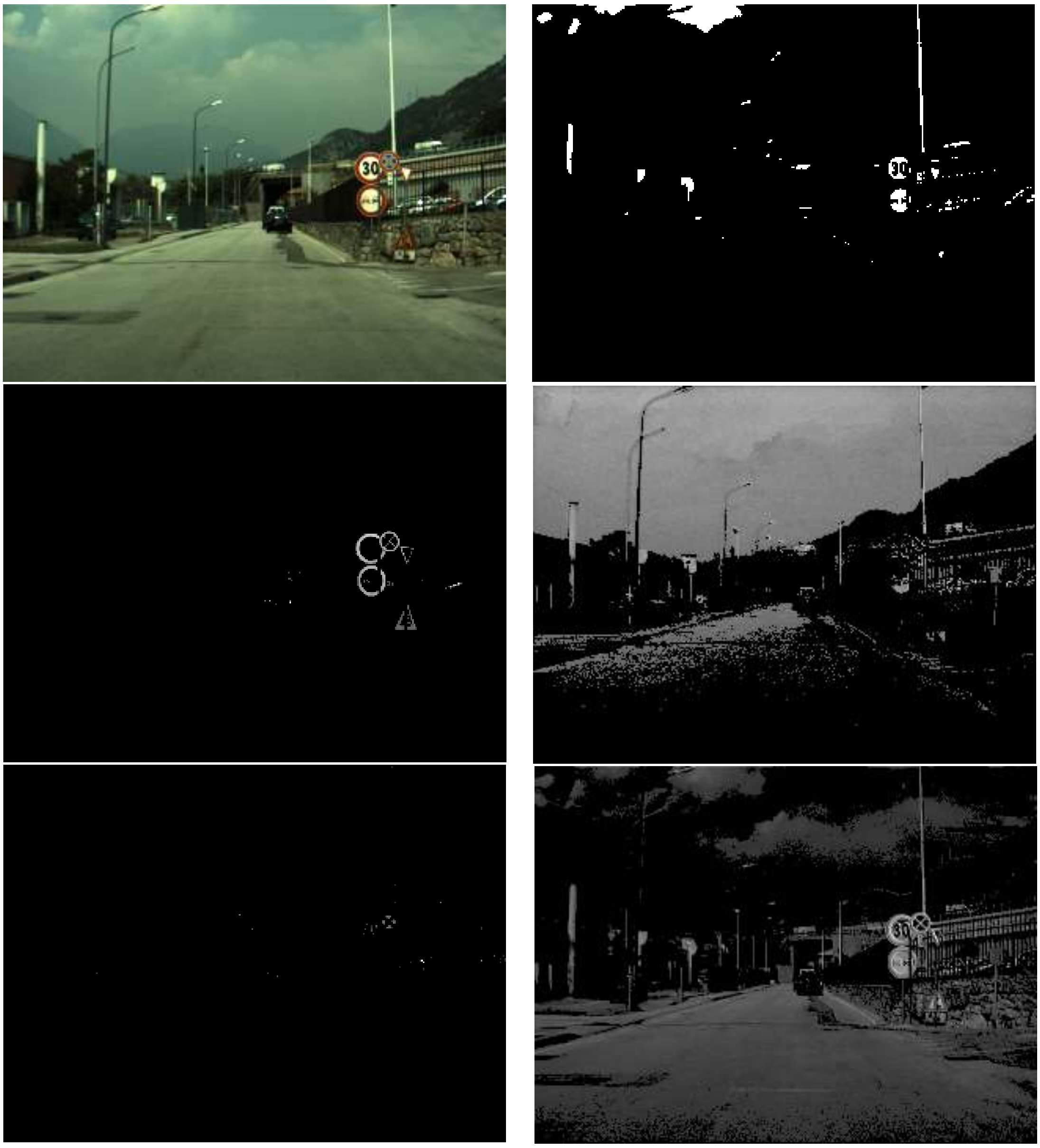


Figure 22 : The top row contains the original image on the left and the thresholded luminosity, L_{bo} on the right. The middle row shows the normalized red and green channels derived from a^* (left and right respectively). The bottom row demonstrates the normalized blue and yellow channels derived from b^* (left and right respectively).

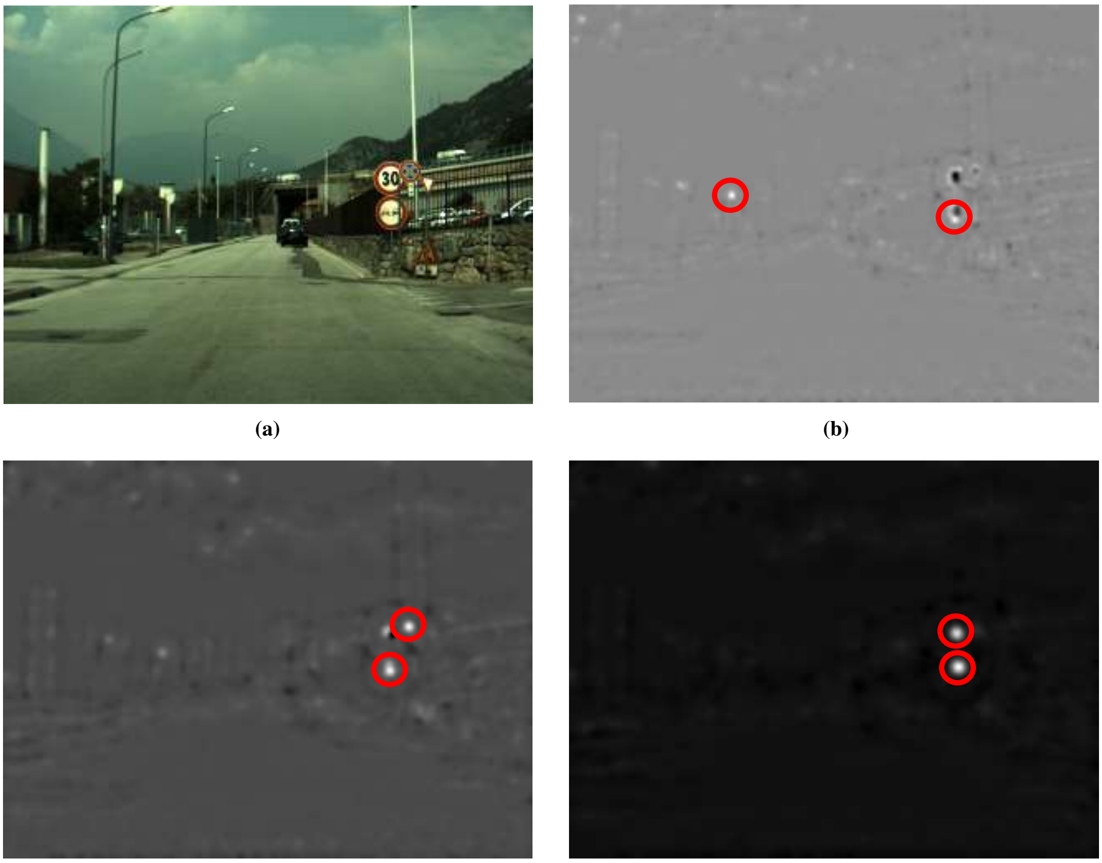
The four chromatic subspaces are then transformed to binary images using Otsu's thresholding algorithm [37]. The four color-based ROIs are estimated by the intersection of the corresponding binary image and the L_{bo}, if the mean luminosity of the frame is lower than 40 (a heuristic threshold discriminating dark from bright scenes), otherwise the four ROIs are identical to the corresponding binary images. The total segmented area is defined by the union of the four ROIs.

3.4.2 Symmetry Detection

The four chromatic coefficients produced in the previous section are scanned for symmetrical shapes, using the FRST proposed in [36]. This method has also been used in [55] due to its computational efficiency and its relevance to the TSD problem.

Depending on the prominent colors of TSs that must detected, some or all chromatic coefficients of the segmented image are used.

In the proposed system the symmetry detection method is optimized for circular shapes, but different symmetrical shapes can also be detected by adjusting the radial strictness factor a to a low value, e.g. 1. This means that every road sign in the image can be detected by this method, as long as it remains in the image frame after the color segmentation. The symmetry detection algorithm scans for shapes of one or more given radii n, belonging in a range N. A result of this process is demonstrated in Figure 23, where TSs of multiple radii are detected. In this example, the radii n in the range $N = [5, 8, 11]$ are used. The results acquired for these radii in each of the four channels were added to get the separate results of Figure 23 (b)-(d).



(a) (b)

(c) (d)

Figure 23: Original image (a) and detected symmetrical shapes from all 4 channels for radii of size (b) $n = 5$, (c) $n = 8$ and (d) $n = 11$. Red circles show the two most prominent symmetrical shapes for each radius.

In order to blindly detect every existing road sign, a large set of radii has to be used, increasing the processing time significantly. This is apparent in the example of Figure 23, since the two large circular signs were best detected by using $n = 11$, while the smaller circular red-blue sign was detected for $n = 8$. The two triangular signs did not produce large symmetry values at their centers, mainly because of the occlusion of the red-yellow sign and the size of the white-yellow one that does not allow for its red frame to be very visible. The synergy of these two facts with the presence of other, more prominent, TSs in the scene leads to a possible miss of the two triangular signs. However, this is not the general case, since most triangular signs can also be detected with the aforementioned method. Such detection is demonstrated in Figure 24. The fusion of all 3 radii FRST results leads to the detection of both triangular signs, with two false positive results on the surface of the red vehicle.

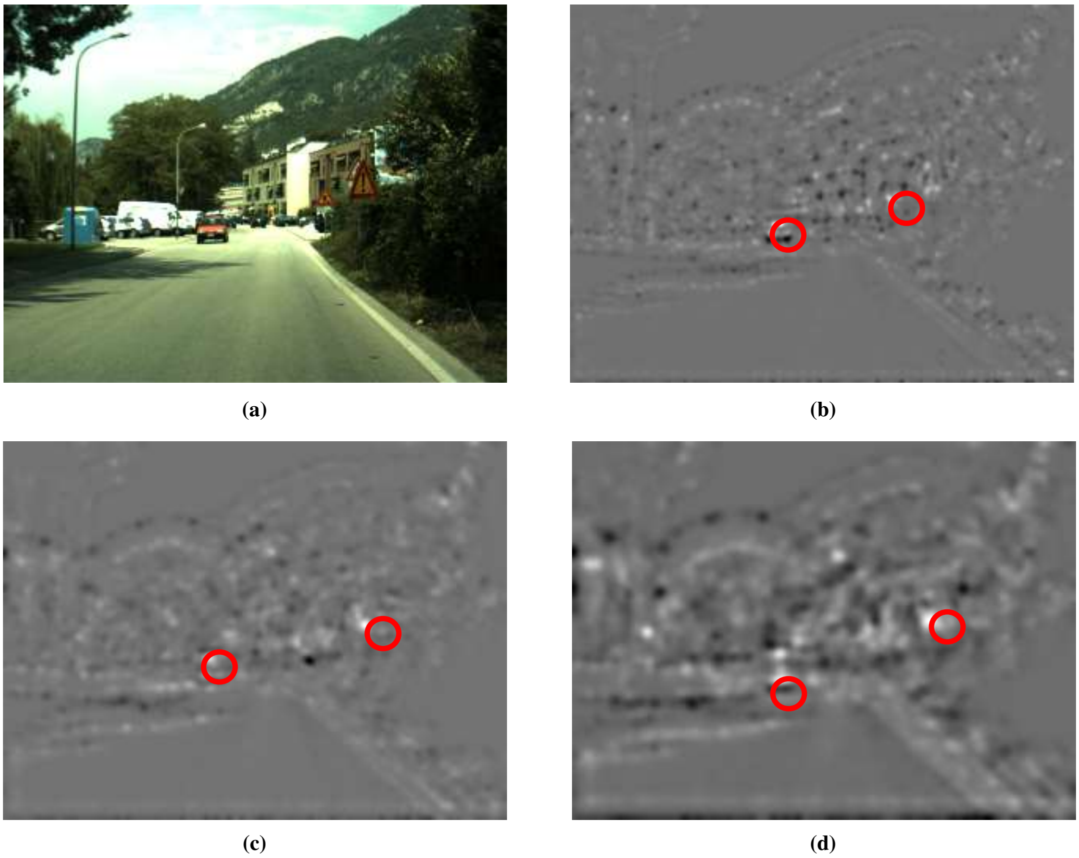

(a) (b) (c) (d)

Figure 24 : Original image with triangular signs present (a) and detected symmetrical shapes from all 4 channels for radii of size (b) n = 5, (c) n = 8 and (d) n = 11. Red circles denote the two most prominent red results.

The FRST also produces good results in detecting blue rectangular signs when using a low value for the radial strictness factor. However, there is a good chance that the detected center will be different than the actual center of mass of the detected TS, but this deficiency can be dealt with in the post-processing phase. Such an example is shown in Figure 25.

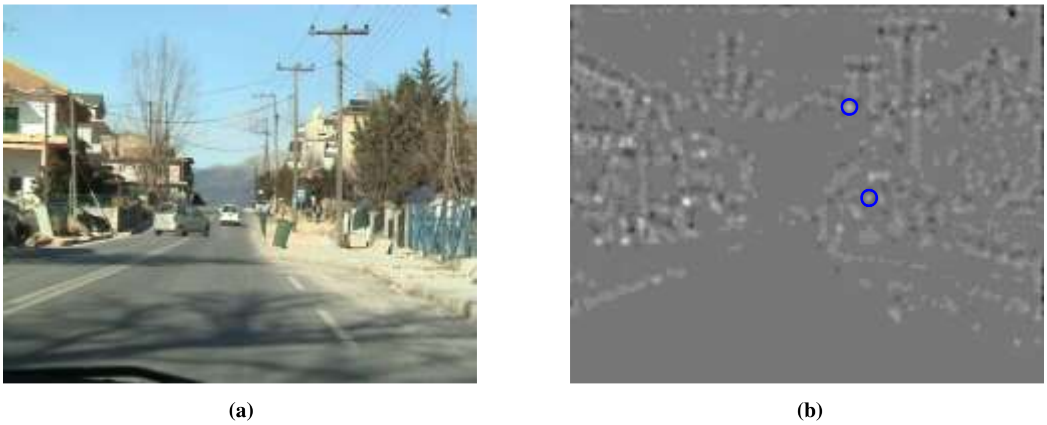

(a) (b)

Figure 25: (a) Original image with a blue rectangular TS present and (b) detected FRST centers from all 4 channels for radii of size n = 5, n = 8 and n = 11. Blue circles denote the two most prominent blue results.

The required computations when tackling a problem with multiple radii can be reduced by taking into account two inherent properties of the TSD problem:

(a) Computations can be performed in parallel architecture, as the calculations needed for the symmetry detection can be performed independently for each radius.

(b) Using detected TS positions from previous frames, the regions of the image that have to be scanned for larger radii are reduced in size and number.

The second approach has been implemented in the proposed system reducing both the number of radii that have to be used and the size of the scanned regions. Before this process can commence, the detected candidate TSs from the aforementioned process must be cropped and identification of their shape must be carried out.

3.4.3 Center Localization – Shape Determination

The sum of the FRST result for all radii and all color channels is filtered by a non-maxima suppression method using a window of size n_{max} x n_{max}, where n_{max} is the maximum radius used in the FRST. After the candidate TS centers have been located, they are cropped using their radius length. More specifically, a square area with dimensions 2.5n x 2.5n with center of mass the detected symmetrical shape center, is cropped and used as a candidate TS.

Once the TS candidate is cropped, it is transformed to binary by thresholding the channel of its prominent color (a* for red, b* for yellow or blue), then undergoes a binary closing to fix possible partial occlusion issues and the holes in the image are filled, in order to perform NCC based template matching with the different road sign shapes templates (i.e. circle, triangle, octagon and rectangle). At the end of this stage, the approximate location, radius and exact shape of the candidate TS have been estimated leading to a reduction of misclassifications for the TS, since it only has to be compared to the template TSs of the same shape type. The process described above is summarized in Figure 26. The NCC values between each binarized candidate in Figure 26(c) and each of the shape templates are compared and the highest value for each candidate denotes its shape class. When the binarized candidate is completely black, as in the fourth case, the candidate is rejected.

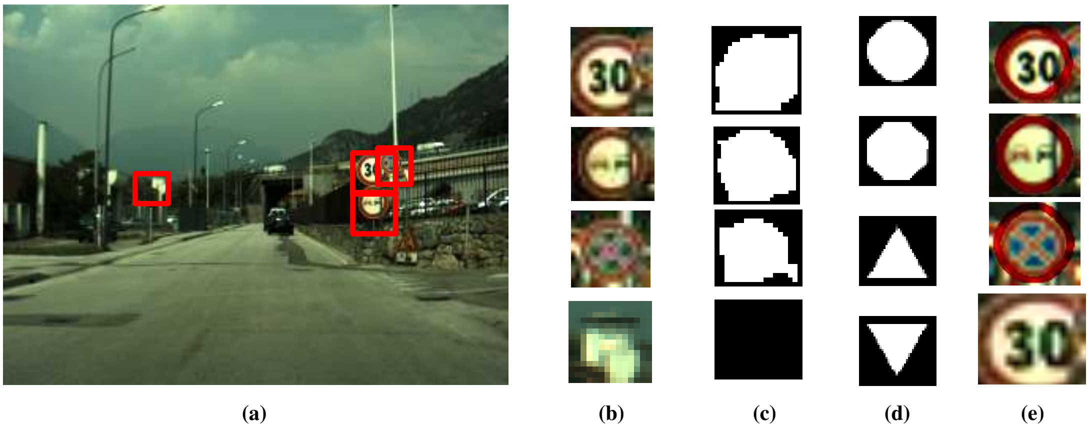


Figure 26 : (a) The original image with the results of non-maxima suppression for the total FRST, (b) cropped candidate TSs, (c) binarized TSs, (d) shape templates and (e) classified shapes overlaid on candidates.

The results demonstrated in the previous figure bring forward two valuable conclusions: a) the circular red signs were all detected and their shape was correctly classified and b) the white object detected as a candidate sign produces a black binarized image after the pre-processing stage, hence leading to the rejection of the candidate.

The same process is followed in the case of blue TSs, since the only parts of the system that change are the color channel involved in the localization process (b* negative values instead of a* positive values) and the shape templates (rectangles instead of octagons and triangles). Such a case is demonstrated in Figure 27, where only one center location remains after the suppression process and the NCC result for the circle template is higher than that of the rectangular template (approximately 0.7 vs. 0.6 respectively).

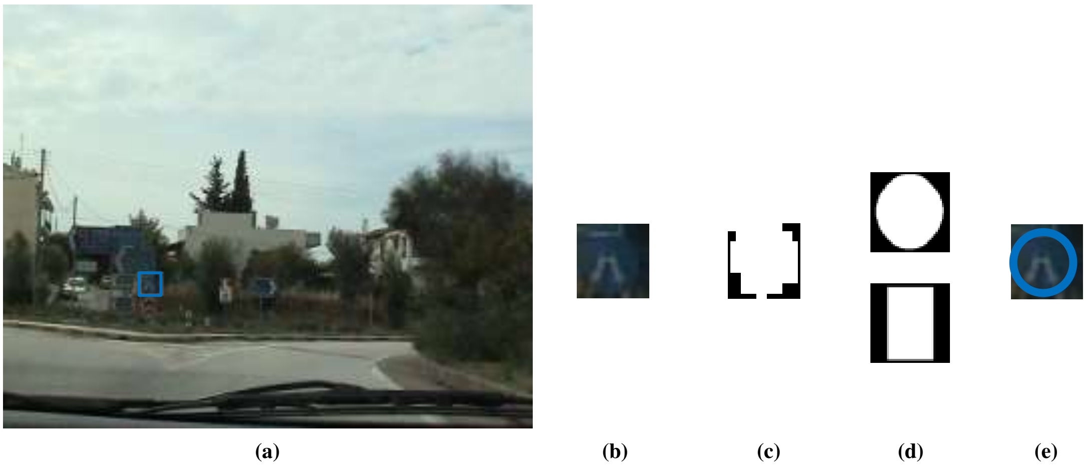

(a) (b) (c) (d) (e)

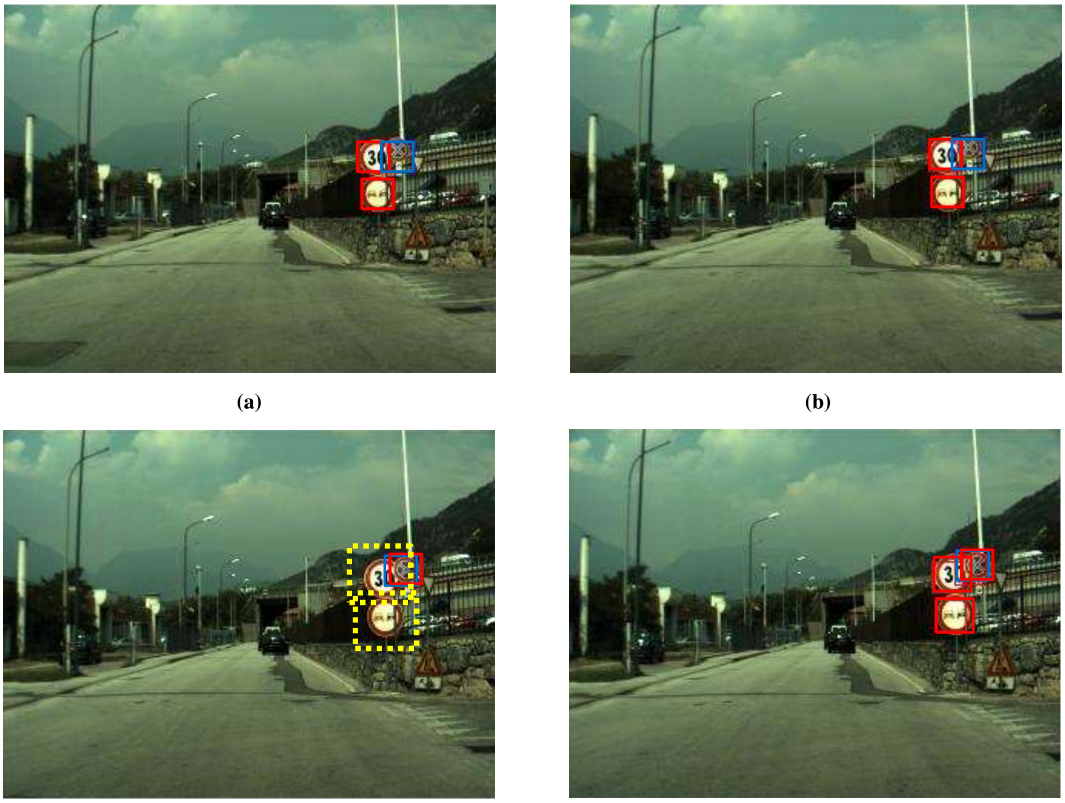
Figure 27: (a) The original image with the results of non-maxima suppression for the total FRST, (b) cropped candidate TSs, (c) binarized TSs, (d) shape templates and (e) classified shape overlaid on candidate.

3.5 Traffic Sign Tracking

The tracking module has been designed to minimize the computational burden and tracking errors of the system. Aiming at this direction, the center coordinates of all detected candidate TSs, of a chosen small radius (e.g. 10 pixels), are passed to the module that processes the next frame. This module performs symmetry detection identical to that of the TSD module for a sub-window of specified size and centered in the coordinates given by the localization procedure of the previous frame. Thus, once having detected a road sign centered in (i, j), the algorithm scans the next frame in an area around (i, j), for a symmetrical shape with a radius of the nearest integer value greater than the one estimated in the previous frame.

This procedure is repeated for every sign detected in the previous frame, regardless if it was its first appearance or if it was tracked from an earlier frame. An obvious flaw of this method is the possibility of temporary loss of visibility of the tracked sign (either total, or partial) in one or more consecutive frames, which results in tracking failure. An efficient parallel process method is used to resolve this problem, by choosing more than one radii in the initial candidate TS detection process (e.g. 10, 15 and 20 pixels), thus the sign is localized again in the following frames. This technique slightly increases the computational effort but recovers missing road signs by including more results in the basic detection process.

A situation where this method is successfully applied is demonstrated in Figure 28. In the first two frames, the circular red signs have been detected (the no parking sign has been detected as blue), when the FRST has been set to detect symmetrical shapes with radials of 10 pixels. In the third frame, the two larger circular signs cannot be detected for a radial of 10 pixels. However, when the regions around the previously detected TS candidates (yellow dash rectangles) are scanned for symmetrical red shapes of 11 pixels radial, the two circular signs are detected again. One important detail that should be noted, is that the process described is limited to the area in the yellow rectangles and possible center results that lie inside the regions that already contain a detected candidate are discarded (to avoid overlapping results for the same TS).



(a) (b) (c) (d)

Figure 28: (a) First frame and detected TS candidates of 10px radial, (b) second frame and detected TS candidates of 10px radial, (c) third frame and detected TS candidates of 10px radial, (d) third frame and detected TS candidates of 10px radial, plus detected TS candidates of 11px radial around the areas of previously detected TS candidates (yellow rectangles in (d)).

Another thing worth noting is the fact that the no parking sign gets detected as both red and blue in the third frame. This is an early indication about the exact class of the aforementioned sign, without having to check it against all circular TS templates in the classification stage.

3.6 Traffic Sign Classification (TSC)

In the previous sections, the candidate TSs detection and tracking modules of the proposed system in the frames sequences shot from a moving vehicle have been presented. With a successful detection and tracking procedure, every TS in the driver's visual field is cropped, resized to 64x64 pixels and passed on to the TSC module. The number of potential template matches has already been significantly reduced by processing the information of the color and shape of the TS, as explained in previous sections.

More specifically, four categories of road-signs can be defined; circular red, triangular red, circular blue and rectangular blue. Two more categories, i.e. the octagonal STOP sign and the inverse triangle shaped yield sign, can be classified by the shape classification module, since they are unique. The classification rule is based on the maximum value of the NCC defined in Equation 2.37, of the cropped TS candidates with the reference template TSs. The classification process for the circular and rectangular blue signs uses only the NCC for the b* channel, while for all the other categories the sum of NCC is used for classification. More specifically, two classifiers are defined; one for red circular or rectangular TSs, and one for blue circular and rectangular signs. The classifiers T_1 and T_2 are estimated by

$$T_1 : Index = \underset{1 \le k \le S}{ArgMax}\{\max(NCC(C_{\mathrm{L}*},T_{\mathrm{L}*}(k))) + \max(NCC(C_{\mathrm{a}*},T_{\mathrm{a}*}(k))) + \max(NCC(C_{\mathrm{b}*},T_{\mathrm{b}*}(k)))\}\,, \tag{3.1}$$

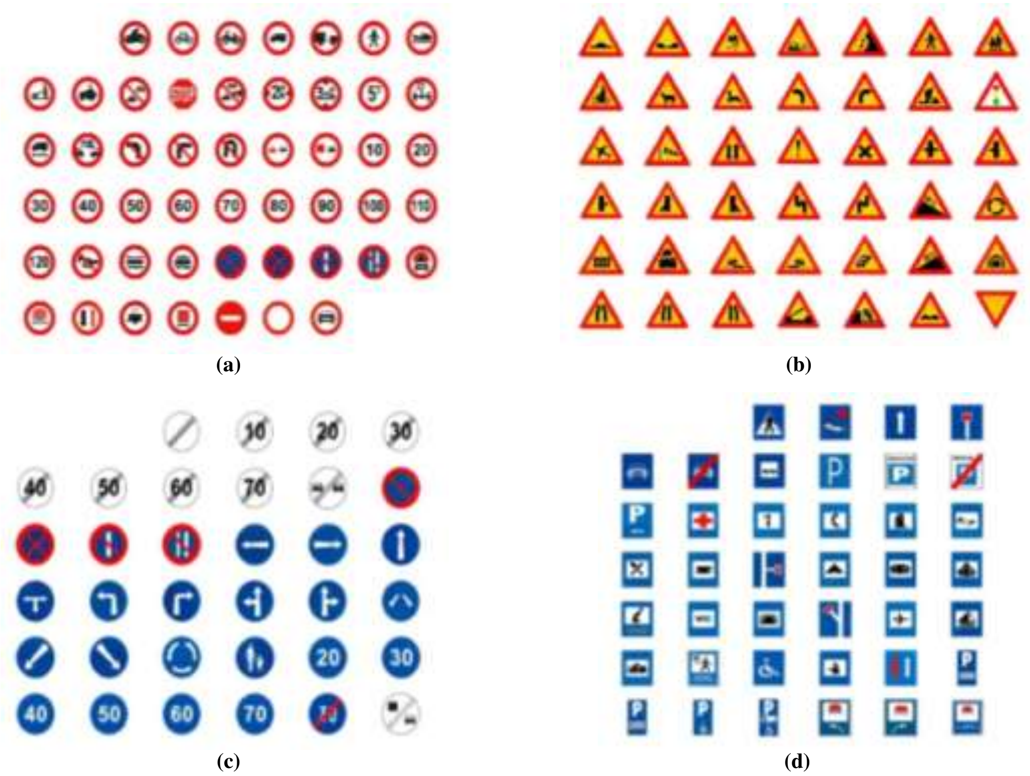
$$T_2 : Index = \begin{cases} \underset{1 \le k \le S}{ArgMax}\left\{\max\left(|\,NCC(C_{\mathrm{a}*},T_{\mathrm{a}*}(k))\,|\right)\right\}, & \text{if } C \in \{\text{red signs}\} \\ \underset{1 \le k \le S}{ArgMax}\left\{\max\left(|\,NCC(C_{\mathrm{b}*},T_{\mathrm{b}*}(k))\,|\right)\right\}, & \text{if } C \in \{\text{blue signs}\}, \end{cases} \tag{3.2}$$

where C is the cropped candidate TS image, $T(k)$ is the k^{th} template TS from the selected category and $NCC(C_{\mathrm{L}*},T_{\mathrm{L}*}(k))$, $NCC(C_{\mathrm{a}*},T_{\mathrm{a}*}(k))$, $NCC(C_{\mathrm{b}*},T_{\mathrm{b}*}(k))$ are the normalized cross-correlation results for the 3 color channels between the candidate TS and the k^{th} template TS as estimated using equation 2.37. Finally, S is the total number of signs in the shape/color category that the candidate TS has been classified by the shape classifier of the previous step.

The four categories of signs that are included in our system are presented in Figure 29. Two important observations are:

(i) the STOP sign has been included in the circular red signs category as a fail-safe measure for cases where the shape of a detected STOP sign is classified as circular. The same goes for the inverse triangle yield TS.

(ii) The no parking / no stop signs have been included in both red and blue circular categories to improve their detection rate.



(a) (b) (c) (d)

Figure 29 : The TS dataset used for the evaluation of the TSR system.

3.7 Experimental results

Our quantitative experimental data consisted of the 4 TSR video clips presented in sub-section 1.6.1, acquired from two different video cameras, the SONY HC85 and the SONY DCR-TRV60E using PAL non-interlaced video (frames of 720x576 pixels) at 25 fps. The system was also qualitatively assessed using parts of the other video streams presented in Section 1.6 as well.

3.7.1 Traffic Sign Detection results

The strongest part of the proposed system is its TSD module, which is robust and precise, even in challenging situations. A very high percentage of the traffic signs in the sequences used for quantitative evaluations were detected successfully, reaching a rate of approximately 95.3% when using three possible radii (10, 15 and 20 pixels). A TS is considered successfully detected when the TS has been located by the TSD module at least once in the frames of its appearance in the video stream. This means that even without the TST module, the FRST based detection is a very effective means of accomplishing this task. However, the big disadvantage of not using the tracking module is the large number of false positive results and overlapping regions. In Figure 30, some typical examples of the TSD process using the aforementioned radii are shown.

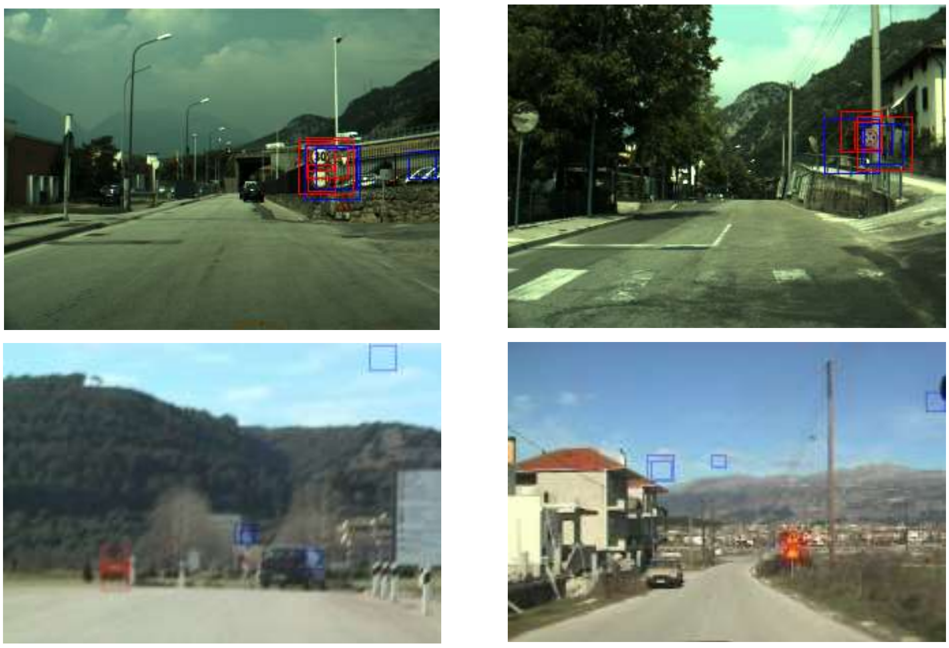


Figure 30 : Typical TSD examples for a single frame using radii in {10, 15, 20}.

What becomes obvious from the analysis of Figure 30, is that despite its very high detection rate, the TSD module alone would be a poor choice in real-life applications since its precision rates are very low (approximately 53%). This is the reason why the TST module combined with the shape classifier are imperative for the system to be considered usable. When using those two modules, the final TSD result is much more immune to false positives and ROI overlaps, as demonstrated in Figure 31.

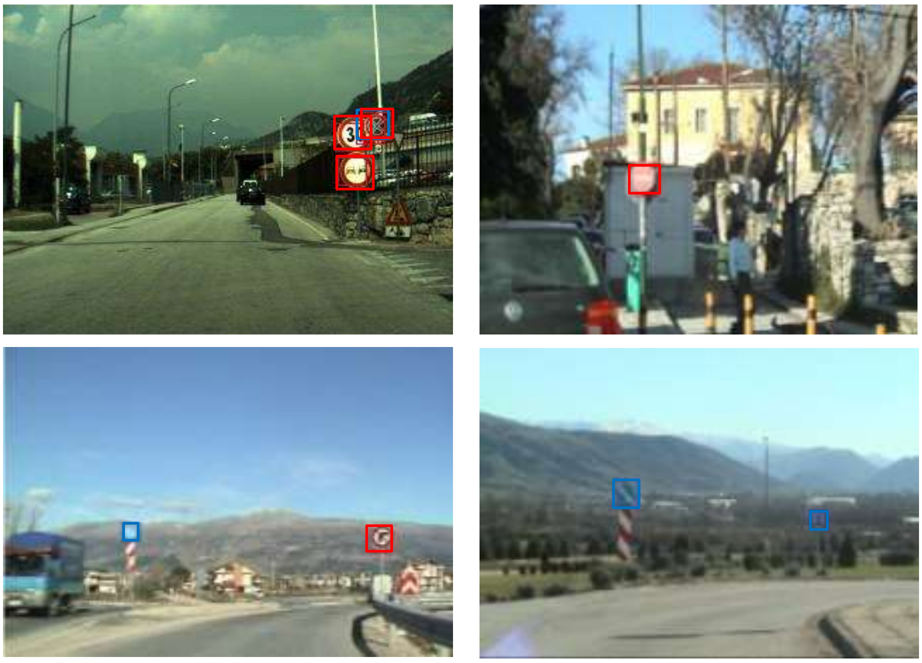


Figure 31 : Typical TSD and TST synergy results.

Using both TSD and TST modules in one system also proves robust in more challenging driving scenarios, as show in Figure 32.

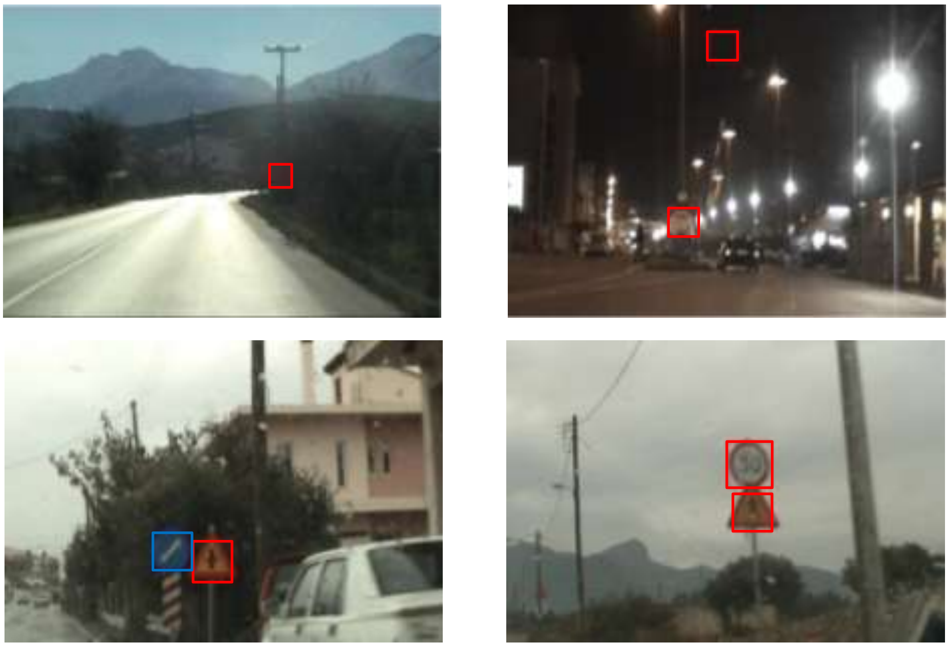


Figure 32 : Typical TSD & TST synergy results in challenging scenes.

3.7.2 Traffic Sign Classification results

The TSC results can also be split in two categories; the frame-by frame TSC results and the TSC results after tracking the sign. In the first case, the correctness of a candidate TS classification is measured in each frame separately, while in the second a TS is considered correctly classified when it has been assigned to the correct class for 5 or more consecutive frames. Some examples of correct and incorrect TS classifications using the first method are shown in Figure 33.

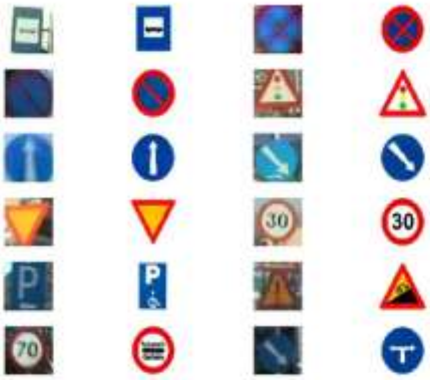


Figure 33 : Examples of correct (top four rows) and incorrect (bottom two rows) classifications using the frame-by-frame method.

In Table 3 the recognition rates for the classification module using the ROIs located by the TSD module for the frame-by-frame case are shown, for normal, night driving and raining conditions. In night recordings, the TSD module tends to locate faulty ROIs due to the presence of a great number of light sources that are often misclassified as TS.

Table 3: TSR rates using only one frame.

TYPE OF ENVIRONMENT		
Normal	**Raining**	**Night driving**
43.92% (343/781)	43.75% (7/16)	6.92% (11/159)

The tracking module does not only improve the detection accuracy of our system, but it also boosts the classification results, since the detected TS candidates can be tracked over time and classified in all frames of their appearance. The consecutive frames TS classification criterion significantly reduces the false positive rate classifying as a sign a ROI that is detected for 5 or more consecutive frames. This approach minimizes the errors involved in a frame-by-frame classification method, where a candidate might have multiple different classifications throughout the time of its appearance in a driving scene.

The proposed tracking method increases the TSC rate to 81.2%, classifying correctly 216 out of 266 signs. In adverse conditions, the proposed system frequently fails to detect the triangular signs especially in low light and rain. Furthermore, a raised number of false positive TS candidates are detected in city night driving.

3.8 Conclusions

This chapter was devoted to the TSR sub-system developed for the purposes of this thesis. The TSR algorithm can be split into three stages, namely the pre-processing stage, the TS candidate detection and the TS candidate verification, which are presented in separate sections. A tracking module that enables a more efficient and robust function of the system is also analyzed. Finally, the system is experimentally evaluated in video streams shot under diverse conditions to prove that, while the detection algorithm is quite efficient, the recognition rate using the NCC method is not very trustworthy when a large subset of TS templates is used.

The sub-systems developed for TSR purposes are very efficient in the TSD part of the process, while introducing a limited number of false positive detections because of the tracking module and the shape classifier that process the results before the final classification. The TSC module is not as successful, but one has to take into account the big number of TS templates that are included in the dataset used for the evaluation of the system, raising the probability of error, especially for signs that are very similar.

Chapter 4

Traffic Light Recognition

4.1 Introduction

The system presented in this chapter performs Traffic Lights Recognition (TLR) and was presented in [73]. The need for the development of such a module for an ADAS is presented in the next section, followed by a literature review. Then, the structure of the TLR system is demonstrated and the different modules comprising the system are explained. An evaluation of the system in terms of accuracy and robustness follows and a comparison to other methods is discussed. Finally, some conclusions about the overall efficiency of the proposed system are drawn and some future work suggestions are given.

4.2 Literature review

TLR is a rather neglected process in ADAS. While a vast amount of all road accidents is caused by drivers who violate Traffic Lights (TLs) indications especially in intersections, the number of systems designed to mitigate this risk is relatively small. TLs might seem like a simple pattern recognition problem for computer vision systems, but the reality lies far from that idea. The driver recognizes TLs by processing morphological information in the scene related to the position and type of the TLs, the presence of intersections, traffic signs, etc. All automatic methods for TLD do not process complex scene information, but detect the presence of TLs only by using their most prominent cues. Therefore, especially in city driving, the sources of light that can be misclassified as TLs very frequently appear, spanning from LED billboards, to the rear lights of other vehicles. The existence of this "noise" can affect TLR systems negatively by increasing the number of false positive detections that may prove disruptive for the driver. When driving at night, this phenomenon is even more obvious, as the number of light sources is significantly increased. The process of TLR is very similar to that of TSR presented in the previous chapter and can be divided into the Traffic Light Detection (TLD) phase and the Traffic Light Classification (TLC) phase. While the former is generally more challenging than TSD due to the smaller size of TLs, the latter is much less complicated than the TSC phase, because the number of classes is very limited.

The idea of using TLR for ADAS in urban environments was first introduced in the late 90's. In [74] a computer vision based Stop & Go algorithm using a color on-board camera is proposed. However, the use of computer vision for ADAS bloomed in the next decade, as computer processor speeds reached a point that enabled real-time implementation of complex algorithms. The work of [75] proposes the fusion of data acquired by color cameras, GPS and the vehicle movement to increase the robustness of their TLR algorithm, which is combined with a tracking module. Their TLD part uses RGB color values, texture and shape features. The approach proposed in [76] uses the HSV space, performing color thresholding followed by a Gaussian filtering process and a verification of TL candidates to achieve TLD in crossroads. A similar approach is followed in [77], based on HSV images and a Gaussian distribution-based model estimated by training images. Shape and temporal consistency information are combined in a post-processing phase to enhance the results. A more straight-forward process is proposed in [78], where the normalized RGB color space is used to locate candidate regions and the results are validated by means of edge and symmetry detection. Color information has been ignored in [79], where grayscale spot detection is followed by an Adaptive Template Matcher, achieving high precision and recall rates in real time. The results of [79] were tested thoroughly and compared to those of an AdaBoost method using the manually annotated videos found at [21]. The problem of TLD in day and night scenes has been addressed in [80] using RGB thresholding and shape dimensions to detect and classify lights in both conditions.

By reviewing related literature to date, the following conclusions can be drawn about vision based TL detection:

- Color information is not always utilized; when it is, either the HSV or the RGB color space is used.
- Many researchers propose the use of heuristic thresholds that cannot be optimal for all possible driving conditions (i.e. shadows, rain, and night). Generally, adverse conditions are not addressed in most papers.
- Symmetry is a widely used cue, either when estimated using novel techniques, or using the well-known Hough transform, but never with the FRST [36].
- Traffic light candidates usually get validated to exclude false positive results. The use of TL models, tracking, or both is the most commonly used solution.
- Apart from [22] that provides a publicly available annotated dataset of on-board video frames taken in Paris, to the best of our knowledge, there are no other annotated datasets for traffic lights detection.

The TLR algorithm developed for this thesis was inspired by the approaches followed for TSR in [81], [53] and [82]. The FRST that was introduced in [36] is employed in all the aforementioned systems, to take advantage of the symmetrical geometry of road signs. The symmetry and color properties are similar in road signs and traffic lights, so these approaches can be a good starting point. The goal of the system is to provide a timely and accurate detection of red and green traffic lights, which will be robust even under adverse illumination or weather conditions.

The proposed system is based on the CIE-L*a*b* color space exploiting the fact that the perceptually uniform a* coefficient is a color opponency channel between red (positive values) and green (negative values). Therefore, it is suitable for distinction between the two prominent classes of TLs. An image processing phase comprising 4-

connected neighborhood image flood-filling on the positive values of a* channel is then applied, to ensure that red traffic lights will appear as filled circles and not as black circles with a light background. The FRST is then utilized to detect symmetrical areas in a*. The proposed system has been tested in various conditions and has been qualitatively and quantitatively assessed, producing very promising results.

4.3 TLR system structure

The hardware setup for the TLR system is similar to most related applications, using a monocular camera mounted on an elevated position on the windscreen of the moving vehicle. The video frames acquired by the camera are processed through three cascade modules. The first module performs image pre-processing, aiming to produce images in which red TLs will appear as very bright circular blobs and green TLs will appear as very dark circular blobs. The output of the pre-processing module is passed on to the traffic light detector, which consists of a FRST for various radii, followed by a detection of multiple local maxima and minima in the top part of the frame. The last module applies a spatiotemporal persistency validation step to exclude those candidates that do not appear in multiple frames, thus minimizing false positives.

The proposed algorithm consists of the following successive steps:

1) Frame acquisition.
2) Image pre-processing:
   a) Conversion of image from RGB to L*a*b*.
   b) Enhancement of red and green color difference.
   c) Holes filling process in enhanced image
3) TL candidate detection:
   a) Radial symmetry detection using the FRST.
   b) Maxima/minima localization.
4) TL candidate verification:
   a) Spatiotemporal persistency check for validation of candidates.

All the aforementioned steps are demonstrated in the flowchart shown in Figure 34.

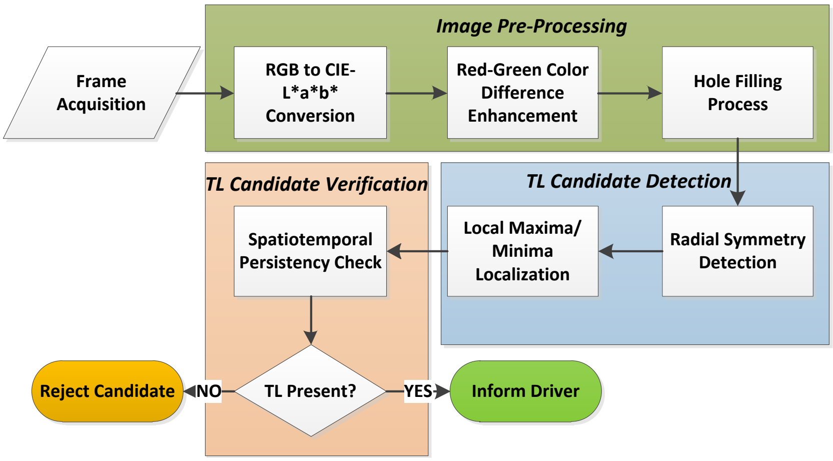


Image Pre-Processing
Frame Acquisition
RGB to CIE-L*a*b* Conversion
Red-Green Color Difference Enhancement
Hole Filling Process
TL Candidate Verification
TL Candidate Detection
Spatiotemporal Persistency Check
Local Maxima/ Minima Localization
Radial Symmetry Detection
Reject Candidate
NO
TL Present?
YES
Inform Driver

Figure 34 : Structural diagram of the proposed TLR system.

4.3.1 Image pre-processing

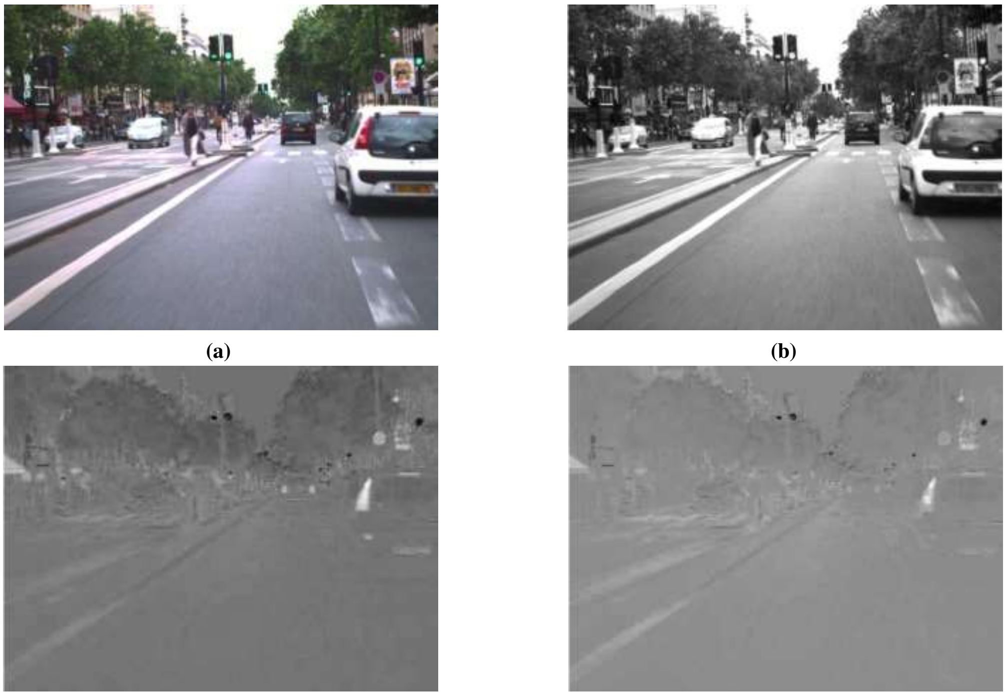
The first stage of the TLR system has the purpose of further enhancing the difference between red and green light sources in the scene using the FRST algorithm that has already been described in Chapter 2 integrating both the TL candidate detection and classification process in the same step. Therefore, it is safe to deduce that the images that will be generated by the pre-processing stage must represent red TLs as very bright circular blobs and green TLs as very dark circular blobs. In the resulting images red TL centers will have large positive values while green TL centers will have large negative values.

To achieve this goal, a new color channel is formed by mixing the two channels of L*a*b* that carry the most important information for TLD. These are the lightness channel L*, which has large values in the TL area and the red-green opponency channel a*, which has large positive values for red TLs and large negative values for green TLs. By multiplying these two channels we produce a channel with enhanced difference of bright red and bright green objects in the scene, called *RG* hereafter:

$$RG(i, j) = \mathrm{L}(i, j) \times \alpha(i, j)\,, \qquad 4.1$$

where i, j are the row and column pixel coordinates.

The aforementioned transformation leads to a further increase of pixels belonging to TLs, while affecting other red and green objects like rooftops, trees etc. to a lesser degree. A typical example is shown in Figure 35, where TLs appear dark both in the *RG* and a* channels, while the tree leaves are less dark in the *RG* images. A similar effect is observed with red signs and non-illuminated rear vehicle lights, which appear less bright in the *RG* channel than in a*.



(a) **(b)**

(c) **(d)**

Figure 35 : (a) Original image, (b) L* channel, (c) a* channel and (d) *RG* channel.

Even though the process described above achieves a better discrimination of TLs from the background, it is not efficient in handling along a common abnormality of real life scenes with bright lights, which is denoted as "blooming effect". This phenomenon describes the appearance of TLs and other bright objects as dark circles with a bright circumference, as shown in Figure 36. Several effects increase the false positive TLs rate, but the most important:

1) the dynamic range of cameras may be unable to capture very bright lights, thus saturating their inner areas and
2) red TLs often include some orange in their hue and green TLs include some blue.

In order to tackle this problem, additional information related to yellow-blue brightness contained in the b* channel is used. The same process followed in 4.1 is applied, i.e. non-linear mixing of channels L* and b* is performed:

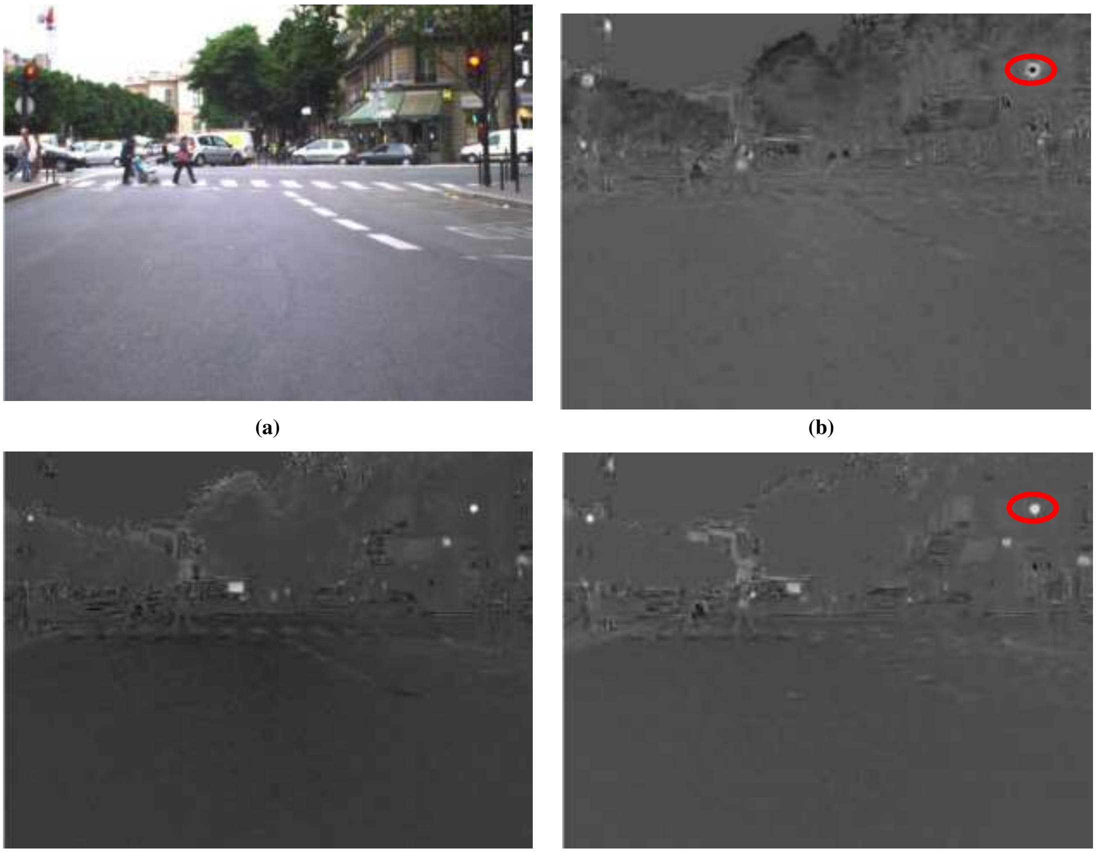
$$YB(i,j) = \mathrm{L}(i,j) \times \mathrm{b}(i,j) \qquad 4.2$$

The summed image of 4.1 and 4.2 is *RGYB*:

$$RGYB(i,j) = RG(i,j) + YB(i,j) = \mathrm{L}(i,j) \times (\mathrm{a}(i,j) + \mathrm{b(i,j)}), \qquad 4.3$$

where i, j are the pixel coordinates.

The *RGYB* image has a positive effect in the reduction of the "blooming effect". An example is given in Figure 36.


(a) (b)

(c) (d)

Figure 36 : (a) The original RGB image and (b) its RG channel, where the red TL on the right suffers from the "blooming effect". (c) The YB channel and finally (d) is the RGYB channel, with the "blooming effect" corrected.

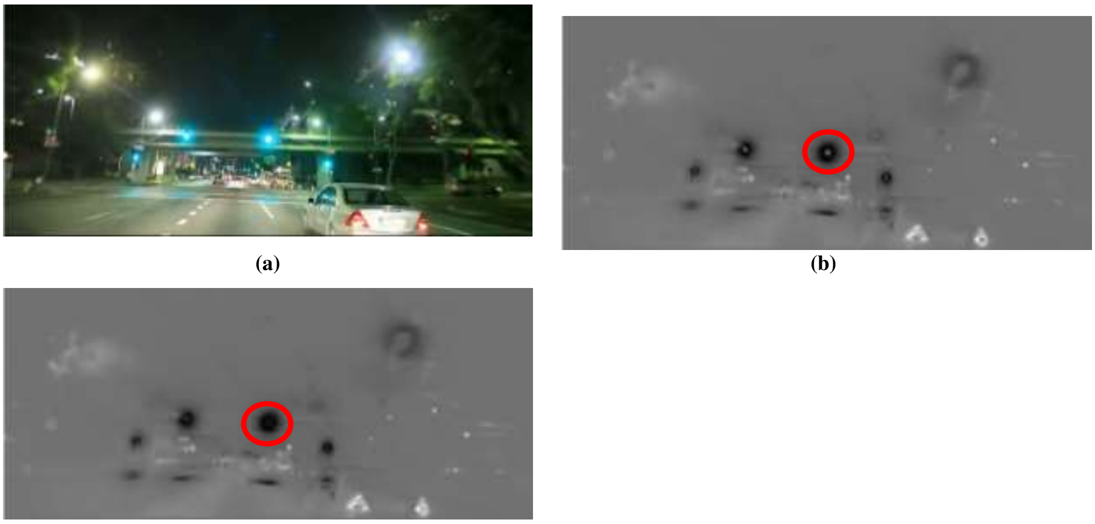
The *RGYB* channel is a reliable input for the forthcoming FRST, since it accomplishes two different goals, by reducing:

1) The prominence of less bright red or green spots in the scene.
2) The "blooming effect".

The first virtue of this transformation will improve the detection rate of the FRST module, since the prominent bright and dark spots will be reduced to light-emitting objects, while the second virtue will enable a simpler classification process, since there will be no correlation between a red light with "blooming effect" and a green light.

An added difficulty is induced to the problem of TLR when more challenging conditions, like night driving, are introduced. More specifically, the appearance of the "blooming effect" cannot be efficiently tackled only by using the *RGYB* channel. Such a scene is demonstrated in Figure 37, where the opposite problem is observed; instead of a dark circle with a bright halo, green TLs appear as small bright circles surrounded by a dark halo. This effect increases the false classification rate, misclassifying green TLs as red ones.

To tackle this problem in night driving, an additional morphological operator processing module is proposed. This is a grayscale 4-connected neighborhood image filling [83] of the holes in both bright and dark areas of the image. The process has two steps; first the *RGYB* image is thresholded to produce two separate channels, one with red and yellow shaded pixels ($RGYB \geq 0$) and one with green and blue ones ($RGYB < 0$). Then, the holes in both images are filled using the method of [83] and the two resulting images are added to form the final result. The process described is demonstrated in Figure 37 where the red ellipses denote areas with a strong blooming effect and the result of our method.



(a)

(b)

(c)

Figure 37 : Handling the "blooming effect" at night. (a) RGB image, (b) RGYB channel, (c) RGYB channel after filling process.

4.3.2 Radial Symmetry Detection

After the *RGYB* image has been estimated at the end of the previous stage, it is scanned by the already presented FRST algorithm to detect symmetrical blobs. As mentioned in Chapter 2,he main feature of the FRST is detecting symmetrical shapes in grayscale images, producing local maxima in the centers of bright blobs and local minima in the centers of dark blobs. The only parameters that have to be defined for this process are a radial strictness factor, a, and the set of radii N that need to be detected.

Since the *RGYB* image includes red and green color opponency, it is appropriate for the FRST. Some examples of the implementation of the FRST (for radii from 2 to 10 pixels with a step of 2 and a=3) to pre-processed frames of various videos are reported in Figure 38.

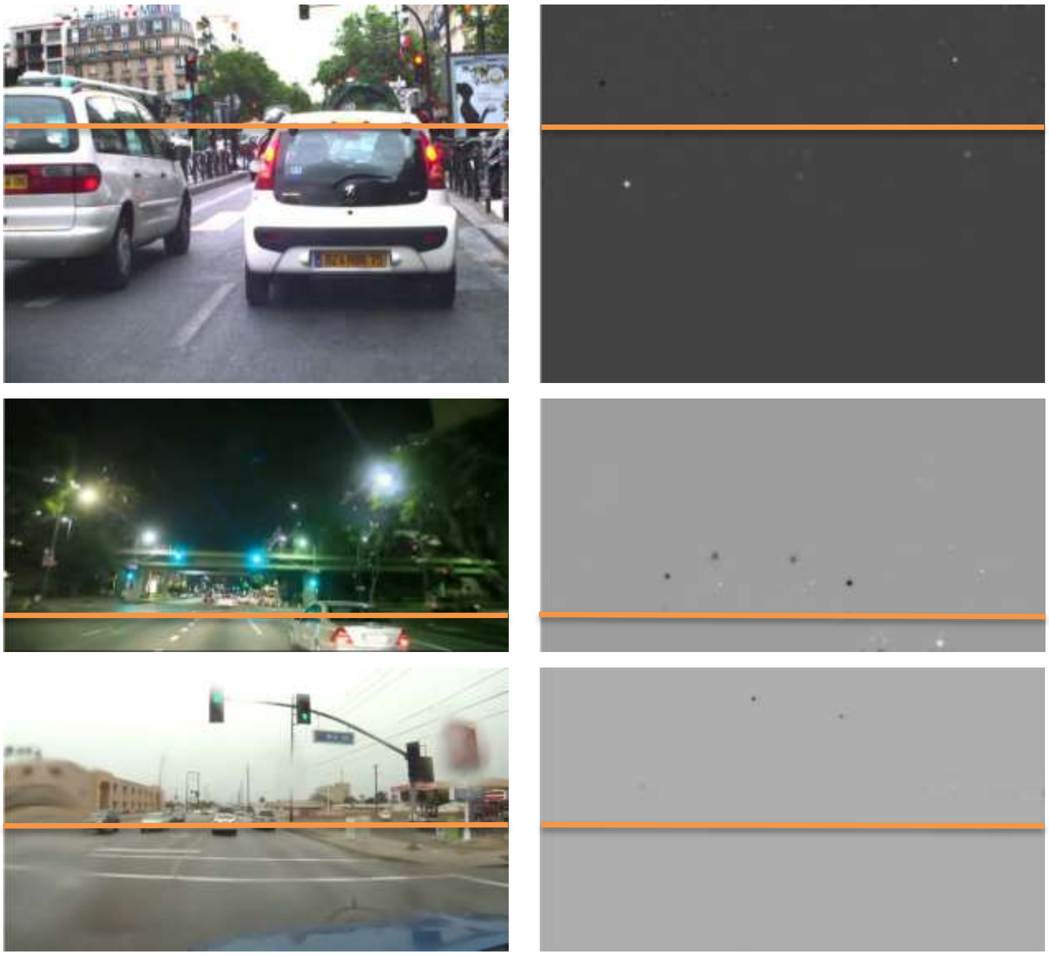


Figure 38: Original driving scenes at day and night time (left column) and their corresponding FRST (right column). Dark spots denote green TLs, light spots denote red TLs. Local maxima/minima detection is performed above the orange line (which depends on the camera placement).

The images presented in Figure 38 show that the centers of the TLs appearing in the scenes are included in the most voted pixels of the frame, provided that the FRST has been applied for a set of radii that includes the TLs radials. Consequently, the result of the FRST undergoes a non-maxima suppression to detect red TLs and a non-minima suppression to detect green TLs must be applied. However, since the nature of the TL detection problem allows attention to be focused on the conspicuous region of the image, i.e. the upper part, only this region will be used for the suppression process. Since the on-board camera was not placed in a constant position in all the videos used, the ROI selection used for suppression must be completed during system calibration.

An automatic ROI selection method using horizon detection, or vanishing point techniques could be utilized [84], [85], but this is not required because the areas where a TL could appear can be easily predicted during the camera setup. Hence, the proposed TL detection system does not include such a method, reducing significantly its computational complexity. In the examples of Figure 38, the suppression took place above the orange line.

The suppression processes were setup to select the greater 5 local maxima and the lower 5 local minima. ROIs with an absolute FRST value lower than half the global absolute maximum were rejected. Each TL candidate ROI is produced from the suppression step as a rectangle with coordinates that are determined by the detected center coordinates, the radius and the color of the TL. More specifically, the annotation of a TL rectangle ROI starts at 6 radii up and 1.5 radii to the left of its FRST center and has a width of 3 radii and a height of 7.5 radii. Similarly, the annotation rectangle for a red TL starts at 1.5 radii up and 1.5 radii to the left of the FRST center and has the same height and width as above. These details are shown in Figure 39.

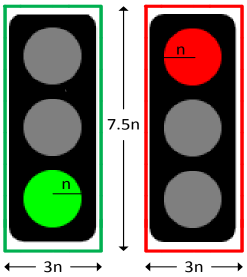


Figure 39 : TL annotation based on radial *n*.

4.3.3 TL candidates verification

The last stage of the proposed TLR system is a TL candidate verification process, which is crucial for the minimization of false positive results that might be caused by bright color blobs that resemble a TL. Road signs, LED billboards etc. can be misinterpreted as TLs, because their shape and color properties are similar to a TL. However, such artifacts usually don’t appear radially symmetrical for more than a few consequent frames, as opposed to the symmetry and color properties of TLs that are more robust to temporal scale variations. Hence, many of the false positive results can be removed by introducing a condition of multiple appearances of a TL in successive frames.

The condition of multiple appearances in successive frames, hereafter called spatiotemporal persistency, states that a TL will appear in the top voted FRST results in a sequence of frames (temporal persistency) and its detected center is expected to leave a track of pixels not far from each other (spatial persistency). Such a result is shown in Figure 40, where the trails left by the centers of two green TLs over a period of 30 frames are denoted by green dots. The trail of sparse dots on the left belong to objects falsely detected as TLs and do not fulfill the persistency criterion.

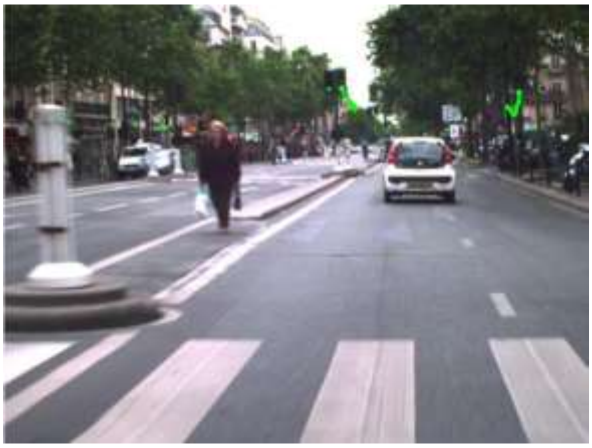


Figure 40 : Spatiotemporal persistency check. Notice the trails left by the centers of two green TLs.

Figure 41, where the first column shows frames 1, and 2 and the second column contains frames 3, and 4. The red rectangles denote a red TL candidate and the yellow ones denote a green TL candidate.

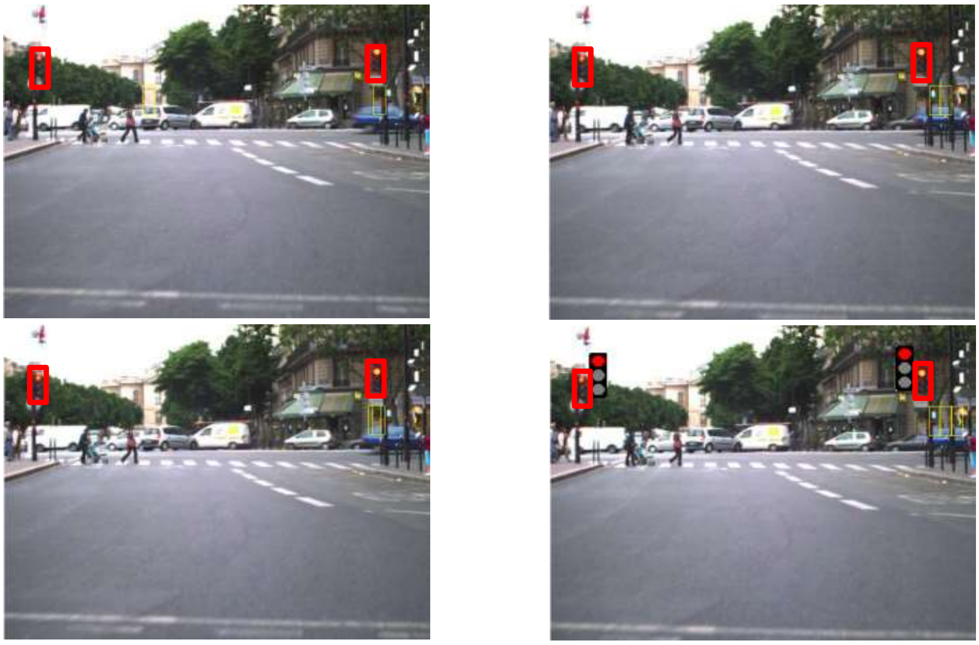


Figure 41 : Four consecutive frames, TL candidates annotated by rectangles. Non persistent candidates are dismissed. Candidates persistent for 3 out of 4 consecutive frames get verified.

4.4 Experimental results

The proposed TLR system has been evaluated using both quantitative and qualitative experimental results. The quantitative assessment has been based on the publicly available annotated video stream of [22], which has already been described in Section

1.6.2. The qualitative evaluation of the system has been based on several videos from the ones that have been described in sub-sections 1.6.1 and 1.6.3, either shot by the authors in Greek roads, or downloaded from YouTube.

The TLR system has been implemented fully in Matlab and tested on a computer with a Core 2 Quad CPU at 2.83GHz, and 4GB of memory. The system implementation code did not use any parallelization. The processing times achieved were directly affected by the resolution of the videos and they fluctuated from 0.1ms to 0.5ms per frame, i.e. near real-time response.

As far as the parameters used for the proposed algorithm are concerned, the radii for the FRST in the experiments were 2,4,6,8, and 10 pixels. The radial strictness was set to $a=3$, and for the spatiotemporal persistency to hold, a candidate must appear in at least three out of four consecutive frames, in an area of a radius of 20 pixels.

4.4.1 Quantitative results

The quantitative results of the TL detector are estimated following the instructions given in [22], for their publicly available, manually annotated video stream. These instructions stated that from the 11179 frames included in the video stream, only 9168 contain at least one TL. However, not all the TLs appearing in these frames are used: TLs with yellow colors (58 instances) are excluded, as well as many lights which are ambiguous due to heavy motion (449 TLs). Apart from these, 745 TLs that are not entirely visible, i.e. are partially outside the frame, are also excluded. Eliminating all these TL instances, a total of 7916 visible red or green TLs remain. The total distinct TLs that comprise these 7916 image instances are 32.

The results scored by the proposed system are estimated following the same rule as [79], which states that if a TL is detected and recognized once in the series of frames where it appears, then it is classified as a true positive. Hence, a false negative result is a TL never detected in its timeline. Using these definitions, the proposed algorithm scores a detection rate (recall) of 93.75%, detecting 30 of the total 32 TLs. The 2 missed TLs were of green color and appeared for 49 and 167 frames in the scene. In both cases, the proposed system detected the paired TLs appearing in the same scene, so the goal of informing the driver was achieved. Some detection examples from the aforementioned dataset are given in Figure 42. The total number of false positive detections is 1559, of which 1245 are false red TL positives and 314 false green TL detections. The red false positive instances concern 12 different objects misclassified as red TLs, while the number of different objects misclassified as green TLs is 7. This means that the precision of the system is 61.22%.

The false positive rate could be significantly reduced by using morphological operations, like the templates utilized by [79] and [21], who report precision rates of up to 97% using also temporal matching. A more direct comparison between our system and the two aforementioned approaches is not feasible, due to the lack of details in their evaluation process. Some characteristic results of successful and unsuccessful recognitions are presented in Figure 42.

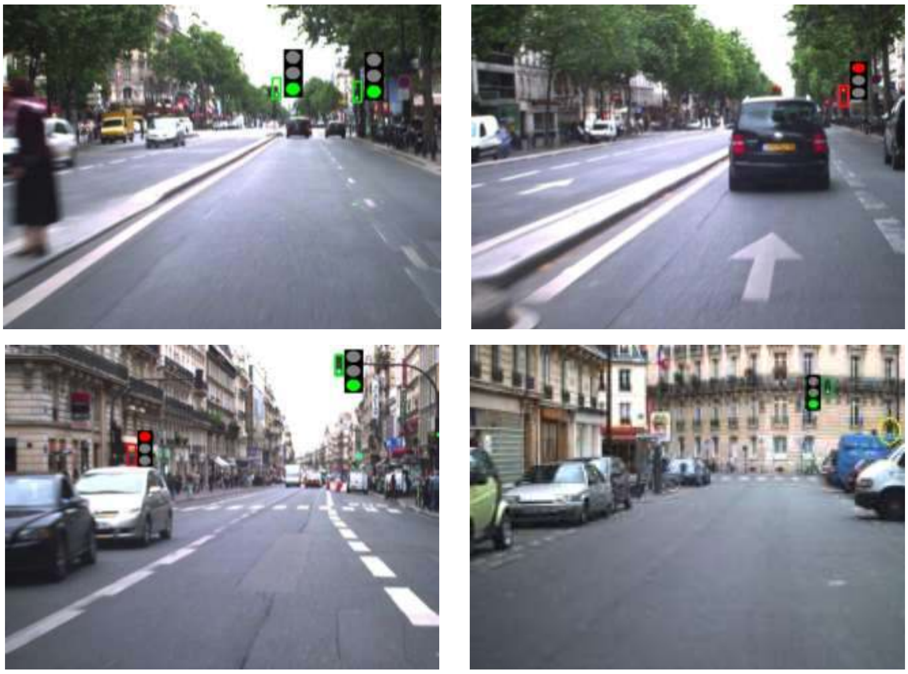


Figure 42 : TL detection results in daytime driving. A false positive red color TL example is show in the bottom left image and a false negative is shown in the bottom right.

4.4.2 Qualitative results

The experimental results reported in sub-section 4.4.1 show that the system appears effective and robust in the case of urban driving in good weather and illumination conditions. The problem of false positives is not so persistent and could be further improved if, as already mentioned, a TL model is used for a final verification and a color constancy module is introduced. However, the great challenge of such systems resides in more demanding driving conditions, including driving under rainy conditions and driving at night time.

For this reason the proposed system is also evaluated in such conditions, so that its robustness and resilience can be examined. More than 2 hours of videos shot from on-board cameras were gathered from the internet and used for the qualitative evaluation of the system. The ultimate goal is to construct a dataset of driving videos shot in challenging conditions, which will be annotated and freely distributed for research purposes in the future.

4.4.2.1 Driving under rainy conditions

Among the most important and challenging cases in ADAS is driving in rainy conditions. The difficulty present is that raindrops on the windshield often distort the driving scene and could cause the various objects to appear disfigured. Another problem in rainy conditions is the partial obstruction by the windshield wiper in various frames. For these reasons, every vision based system should be tested under rainy conditions, as the results produced may vary a lot from the ones achieved in normal weather. Some examples of successful TL detections in rainy conditions are shown in Figure 43.

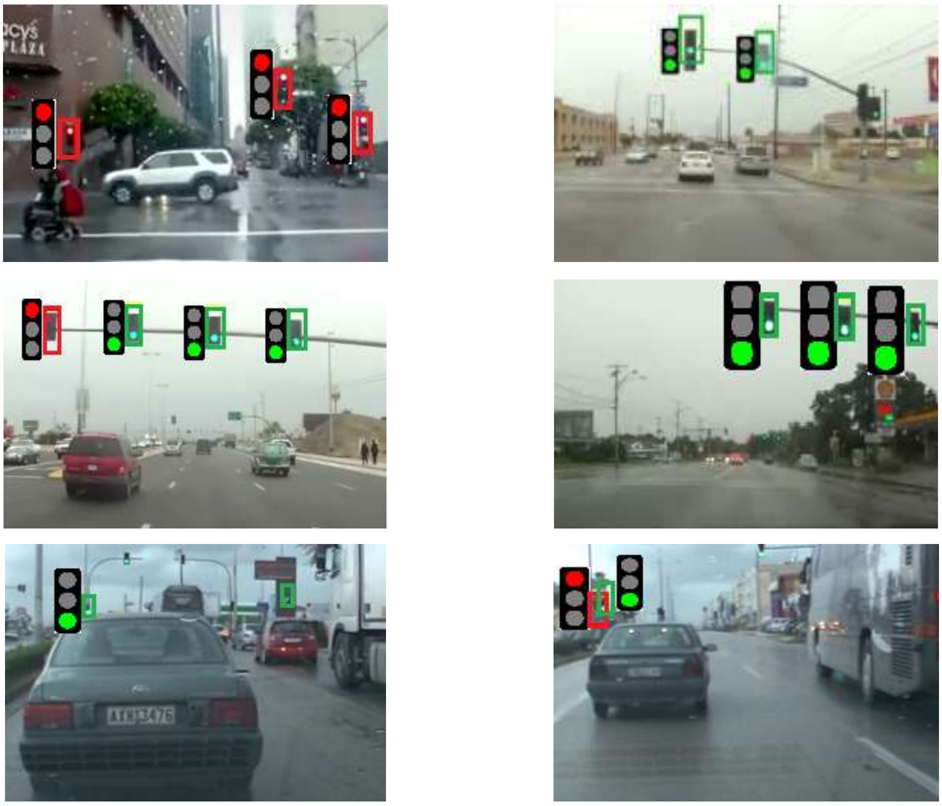


Figure 43 : Successful TL detection results in rainy conditions.

4.4.2.2 Driving at night

The second important category of adverse driving conditions is night driving. The difficulty of the situation relies largely on the environment and the mean luminosity of the scene. If the environment does not include excessive noise like for example dense advertisements or other lighting sources, the proposed system performs satisfactorily, even in urban driving situations. Successful detections in night driving scenarios are presented in Figure 44. Most TLs are successfully detected, even when their glow makes it very difficult to distinguish shape and morphology cues.

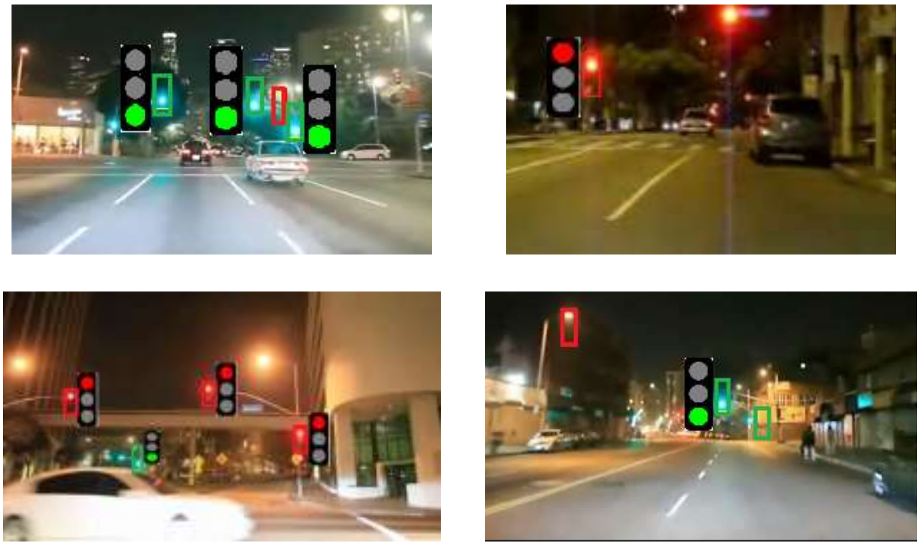


Figure 44 : TL detection results in urban night driving.

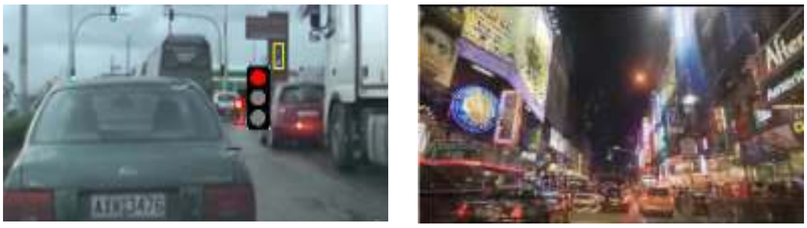
4.4.2.3 Known limitations of the Traffic Lights Recognition system

A very common question when dealing with computer vision applications is whether there are classification rate limitations, i.e. low rates of accuracy derived from processing certain image frames. Compared to the human TLD perception system, a machine based TLD performs poorly. Taking into account that a human driver detects the position of TLs using an extremely accurate object recognition system and by efficiently processing morphological information related to appearance of simultaneous objects positions in the scene (road lines, crossroads, TLs), the accuracy of our TLD system using only TL color information is quite satisfactory.

The proposed method is by no means flawless and can produce persistent errors in some situations. The main problems that can be pinpointed are the following:

(i) The system produces some false positive results that cannot be easily excluded, unless it is used in correlation to other computer vision modules like a vehicle detector or a road detector. An example of such false positives is illuminated vehicle tail lights, or turn lights, as shown in Figure 45(a). This image also includes a false positive result that is eliminated in the following frames.

(ii) The proposed system fails completely in cities like New York (Figure 45(b)), where the visual information is extremely dense and a great number of similar to TLs objects appears in the scene.



(a) (b)

Figure 45 : Examples of temporary or permanent failures of the proposed system.

4.5 Conclusions

In this chapter, a novel automatic algorithm for TLR using a monocular on-board camera has been proposed. The algorithm uses color, symmetry and spatiotemporal information to detect the red and green color of TLs in a fashion resilient to weather, illumination, camera setup and time of day. It utilizes a CIE-L*a*b* based color space with a holes filling process to enhance the separability of red and green traffic lights. A fast radial symmetry transform is then used to detect the most symmetrical red and green regions of the upper part of the image, producing the TL candidates. Finally, a spatiotemporal persistency criterion is applied, to exclude many false positive results. The algorithm has been experimentally assessed in many different scenarios and conditions, producing very high detection rates, even in very adverse conditions.

Future work will be directed towards embedding a tracking module to decrease the false negative results and a color consistency module to further reduce false positives. Furthermore, the combination of the TL detector with other ADAS modules like vehicle, sign and road detection will be explored, so that a complete solution for driver assistance will be proposed.

Chapter 5

Single Image Detection of Preceding Vehicles

5.1 Introduction

In this chapter, an algorithm for Preceding Vehicles Detection (PVD) that can be integrated in ADASs for forward collision avoidance is proposed. It is based upon a combination of multiple cues present on vehicles, such as the red color of rear lights, horizontal edges and symmetry. The system was presented at [86].The rest of the chapter is organized as follows; first, a survey of state-of-the-art methods for PVD is performed, followed by a presentation of the proposed system structure. Then, the separate system modules are presented in detail and an evaluation of the results of the proposed system is performed. Finally, a discussion on future improvements and ideas on the problem of PVD is presented.

5.2 Literature Review

Computer vision based techniques proposed for vehicle detection usually utilize visual features like appearance or motion. An extensive review of the most notable methods until 2006 can be found in [87], while a more recent survey focusing on active learning methods is in [88]. The methods presented in [87] mostly comprise two stages, namely the hypothesis generation (HG) and the hypothesis verification (HV) stage. The HG stage uses information acquired form a monocular or stereoscopic video camera mounted on a moving vehicle to detect areas of interest in a frame where a vehicle might be present. To achieve this, the HG module can use low-level knowledge based cues like color, symmetry, edge or corner detection and texture information. These cues can also be combined to detect higher-level features, like shadows and lights. Other knowledge based methods of HG use motion information, or stereovision in the case of stereo cameras. Once detected, the candidate vehicle areas of interest are passed on to the HV stage, where some systems use representative vehicle patterns to correlate to the detected region, while others use appearance cues to train a classifier which then separates the detected regions to vehicle or not vehicle. Many researchers have also added a third stage to their systems, which tracks the detected vehicles in a sequence of frames, thus improving the detection performance mainly by reducing false positive detections.

A widely used cue for the identification of a preceding vehicle is its rear lights. This approach makes the resulting systems robust in night driving conditions, when other vehicle features are hard to detect due to limited scene illumination. Moreover, the regulations in the U.S.A. [89] and the E.U. [90] state that the rear lights of vehicles should be red, making them a discriminative cue that can also be used at day time, especially for braking vehicles detection. The most common approaches use the RGB color space or its separate channels to segment rear vehicle lights. In [91] rear vehicle lights are segmented based on a combination of grayscale and red color brightness derived from the normalized difference of the R and B color channels; the detected tail lights are then paired and filtered for false positives using an Inverse Perspective Mapping method. The system proposed in [92] computes a “red level” for every pixel in small regions of interest, based on a proportion between RGB channels followed by a symmetry analysis of the color distribution to verify the rear lights and segment the vehicle using edge detection. In [93] the lights detection is based solely on the R channel values and then it is combined with edge and motion information to verify the result. In [94], brake lights are detected at night time using statistically derived thresholds in all three RGB channels. The size of candidate taillights and the speed of the ego-vehicle are then used to assess an impending rear-end collision risk. Even though many systems have used it, RGB is not the optimal color space for real world color segmentation problems mainly due to the high correlation between its channels that makes it difficult to set and manipulate color parameters. Thus, many researchers have proposed the use of alternative color spaces for the task of rear lights detection. In [95], the rear lights are segmented in night driving scenes using a subjectively selected threshold in the Cr channel of the YCbCr color space; however the color information is not used in day scenes, when a horizontal edge detection technique is preferred. O’Malley et al. in [96] used the color distribution of rear-lamp pixels under real-world conditions to derive proper thresholds for the HSV color space, based on the automotive regulations for vehicle appearance. The detected rear lamp candidates are then paired using symmetry analysis and color cross-correlation and finally tracked using Kalman filtering. the CIE-L*a*b* color space is used in [97] for the detection of rear lights, using two fixed thresholds for channels a* and b* . This approach leads to a classification of detected lights in one of three classes: white, orange and red.

The methods described have an important disadvantage as far as benchmarking is concerned; none of them use a publicly available annotated dataset of vehicle images to test their accuracy. The reason for this is twofold: first the number of such datasets is limited and usually is used for the assessment of general image classification algorithms; second, the nature of the problem of vehicle detection allows for a huge variety of different vehicles, environments and conditions that should be considered, making the use of a representative to real-life conditions dataset highly impractical. However, the quantitative assessment of vehicle detection methods in typical situations like the ones included in [31] is useful for the comparison of similar methods.

5.3 Preceding Vehicle Detection

The PVD system proposed is based, as all of the ADASs in this thesis, on the video frames acquired from a monocular camera adjusted on the ego-vehicle windscreen, facing the road. The system uses color, which is a basic cue for detecting the rear part of vehicles due to the existence of red rear lights. Many existing systems utilize static thresholds for red color segmentation of vehicle lights, thus being sensitive to illumination changes and environmental conditions. The main feature of the proposed method lies in the absence of such static thresholds for the detection of rear lights, making our system resilient to illumination conditions alterations, time of day, camera settings and sensor characteristics. Moreover, the use of multiple cues is a viable means for improving the reliability of our system.

The proposed PVD system builds on *a-priori* knowledge on some of most prominent visual features appearing on the rear facade of vehicles. In this direction, the first stage of the system is the extraction of a binary image containing candidate rear vehicle lights, based on color segmentation and radial symmetry. As already mentioned, rear lights represent a conspicuous cue for PVD. The reason for that is twofold; the red rear lights are a common feature among all vehicles according to legislation and they are also visible under different illumination, weather conditions and time of day. Furthermore, they can be used to provide timely warning of an impending collision, as illuminated rear brake lights are an early indication that a vehicle is slowing down.

The second stage of the proposed system involves morphological pairing of candidate lights areas, followed by a horizontal edge detection process and finally an axial symmetry check along the vertical bisector is utilized for vehicle presence verification. For the successfully detected vehicles, a rough estimation of their distance is performed. The architecture of the proposed system is presented in Figure 46.

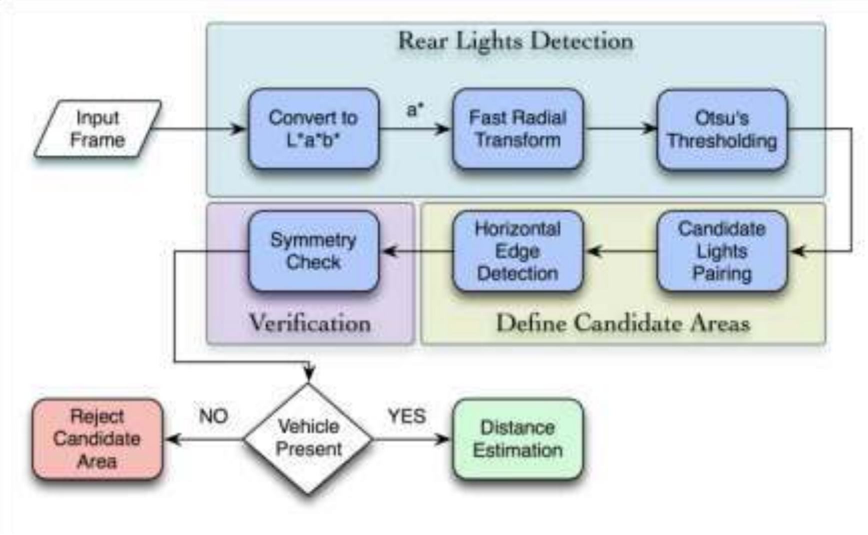

Rear Lights Detection
Input Frame
Convert to L*a*b*
a*
Fast Radial Transform
Otsu's Thresholding
Symmetry Check
Horizontal Edge Detection
Candidate Lights Pairing
Verification
Define Candidate Areas
Reject Candidate Area
NO
Vehicle Present
YES
Distance Estimation

Figure 46 : Structure of the proposed PVD system.

5.3.1 Red lights detection

In the first processing stage of the proposed system, a search for areas with high red chromaticity in the color image is performed. Among various color spaces, CIE-L*a*b* was the most suitable for this application, as it possesses a number of important features. L*a*b* color space has the advantage of being a perceptually uniform color space, mapping equally the perceived color difference into qualitative distance in the color space. In L*a*b*, lightness information (L component) is separated from the chrominance information (a*, b* components), which we utilize for the red color segmentation. As a result, illumination changes have a minimal effect on color information.

For the segmentation of the possible rear light areas, we utilize component a* (red - green) of L*a*b*, and split the positive from the negative part, in order to acquire only the red subspace. This subspace image is then scanned for symmetrical shapes, using the fast radial symmetry transform presented in [11]. Although there is no constraint in the shape of rear lights, they generally follow a symmetrical pattern. A judicious choice of a low radial-strictness parameter (*a=1*) gives emphasis to non-radially symmetric features [11], thus presenting great values at the positions of rear lights. The results are demonstrated in Figure 47, where channel a* and the FRST appear pseudo-colored.

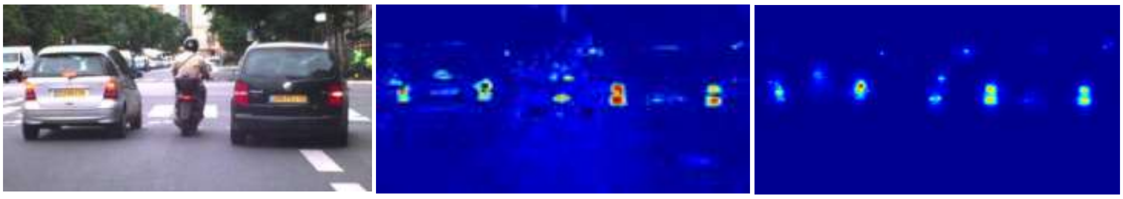

(a) **(b)** **(c)**

Figure 47 : (a) Original image, (b) red subspace of the L*a*b* color space visible and (c) FRST of the red subspace.

The FRST, as already mentioned, scans for shapes in one or more radii in a range called *N*. When the goal is to detect symmetrical shapes of any size, a large size of ranges must be used; however, the computational burden can be greatly reduced by selecting a small sparse set of radii in range *N,* spanning between the extreme sizes of possible rear lights. The result in this case is a very good approximation to the output obtained if all possible radii were examined. Using the FRST approach, the "blooming effect", caused by the saturation of bright pixels in CCD cameras with low dynamic range, is very effectively handled. This is attributed to the fact that saturated lights appear as bright spots with a red halo around, thus yielding large radial symmetry values. This phenomenon is illustrated in Figure 48.

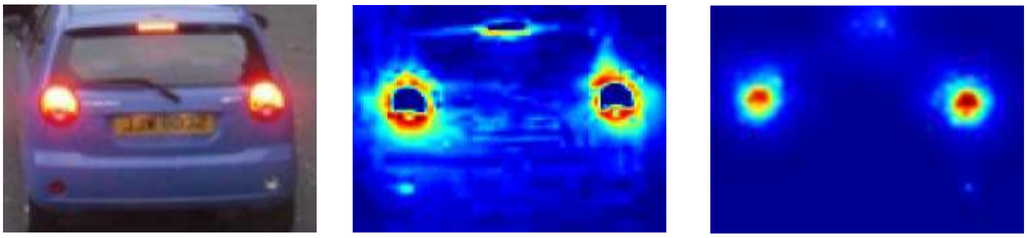

(a) **(b)** **(c)**

Figure 48 : (a) Original image, (b) red subspace of the L*a*b* color space where the "blooming effect" is visible and (c) FRST of the red subspace, a*.

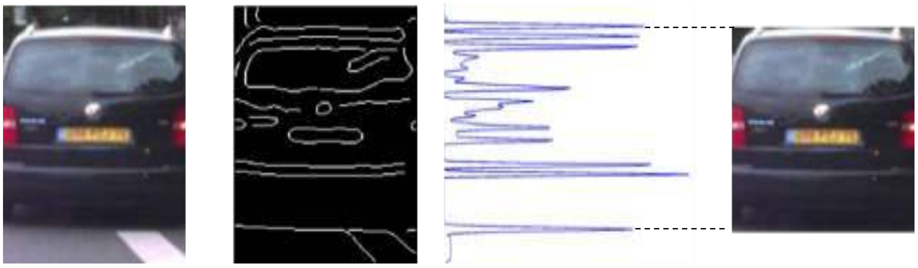
For the final binarization of the image, the fast and efficient Otsu's thresholding algorithm [37], which suggests an optimum threshold by minimizing the weighted sum of variances of brightness for the two classes: the objects and background pixels. The resulting binary image contains the candidate rear vehicle lights.

5.3.2 Morphological lights pairing

In the second stage, a morphological rear lights pairing scheme is applied to the previous binary image to determine vehicle candidates. After connected component labeling, for each region the ellipse that has the same second central moments as the region is estimated, in order to derive its features. The parameters of the ellipse, i.e., the center coordinates, the major and minor axis lengths as well as their area are computed. In order to find pairs of possible lights we consider all the possible $\frac{K\cdot(K-1)}{2}$ two-combinations, where K is the total number of candidates. However, from all these potential pairs only a few meet the prerequisites that we impose, regarding the angle between them and a similarity measure based on their geometrical properties: Assuming that the target vehicle is in the same tilt as the observing vehicle, the candidate pair of lights must be aligned in the horizontal axis (with a permissible inclination of ± 5 degrees). The morphological similarity measure is based on the normalized difference of their major axis length, minor axis length and area.

5.3.3 Horizontal edge boundaries

Given the candidate rear light pairs, we seek the horizontal boundaries of the candidate vehicle. The vertical boundaries are defined by the extreme points of the rear lights. A search region for the upper and lower horizontal boundaries that is proportional to the width of the vehicle is determined, which is assigned as the distance between the extreme points of the rear lights. In Figure 49(a) a search region on the original image is shown. The 'Canny' edge detector is used to detect the edges in the grayscale image of the search region (Figure 49(b)). The horizontal projection of the edge map is then computed (Figure 49(c)), while the peak values indicate pronounced horizontal edges. The upper and lower boundaries of the car are defined as the first and last peak in the projection graph, with value at least equal to the half of the largest value. The outcome of this stage is bounding boxes containing candidate vehicles.



(a) **(b)** **(c)** **(d)**

Figure 49 : (a) Search region on the original image, (b) edge map of the search region, (c) its horizontal projection and (d) bounding box containing the candidate vehicle.

5.3.4 Axial Symmetry check

Symmetry represents a very useful feature for vehicle verification, since it is a characteristic of human-made objects. Vehicles observed from their rear view are in general symmetrical in the vertical direction [87]. The axial symmetry check proposed to verify that a candidate region is indeed symmetrical, is performed by splitting each bounding box image into two sub-images along the vertical bisector and comparing them. The comparison of the sub-images is carried out by utilizing two measures, namely the Mean Absolute Error (MAE) and the Structural SIMilarity (SSIM) measure [52]. Both measures have already been described in Chapter 2.

The verification of a candidate vehicle region is performed heuristically, using two experimentally derived lower bounds for the two aforementioned measures.

5.3.5 Preceding Vehicle distance estimation

Once the system has detected and verified the preceding vehicle, its relative distance to the ego-vehicle can also be estimated. A precise distance calculation is practically infeasible, as the information contained in a single video frame is not sufficient to derive precise depth information. However, a good approximation for the distance of cars of typical size can be achieved; assuming that the average vehicle width is approximately 1.8m and given the width of the detected preceding vehicle in the image (as a proportion of the vehicle's width in pixels to the total image width in pixels) we can estimate the desired distance. A more precise estimation can be computed as in [93], provided the camera characteristics known in advance.

5.4 Experimental results

The performance of the proposed PVD system was tested on the two publicly available datasets named Caltech 1999 and Caltech 2001, which were described in Chapter 1. These test sets contain images of many different cars, under various illumination conditions and cameras. It must be clarified that in Caltech DB (Cars 2001), 22 of the 526 images were excluded because of their red rear lights being modified beyond legislation [96], or because one of the brake lights was blown. For the video stream of [14] only frames that contain a whole visible, preceding vehicle at the same lane and in distance less than 15m were considered (2716 out of 11179 frames). Red vehicles, recognized successfully as large regions in the binary image, were also detected using the same method. The recognition results are summarized in Table 4.

Table 4 : Detection accuracy of the PVD System using a single image.

Dataset	Number of images / frames	Detection Rate	Detection Rate – Braking PV
Caltech DB (Cars 1999)	126	92.1%	-
Caltech DB (Cars 2001)	504	93.6%	99.2%
Lara Urban Sequence 1	2716	92.6%	96.3%

The proposed system scores high detection rates in all test sets (up to 93.6%), and performs outstandingly in cases when the preceding vehicle brake lights are illuminated, as shown in Table 4. This can be attributed to the intensive, highly distinguishable color of brake lights and the ability of our system to handle the “blooming effect” very effectively. This specific feature is of key importance, as accurate recognition at the stage when the preceding vehicle is braking is very crucial for avoiding an impending collision. A fruitful comparison can be made with the system of [98], reporting results at the same datasets (Caltech DBs). The proposed approach performs better (93.3% versus 92% reported in [98]) on both datasets, with the additional advantage of requiring no training. Regarding the false positive detections (7 for the Caltech 1999 and 46 for the Caltech 2001 dataset), these can be further reduced if certain restrictions in the admissible distance between rear lights are included, but in this case the PV detection distance narrows down. A more robust method can be implemented taking into account that most FPs rarely appear in more than one consecutive frame. In this direction, simple spatiotemporal restrictions can be used to detect the FPs not persistent in time. Some representative detection results from all data sets used are illustrated in Figure 50.

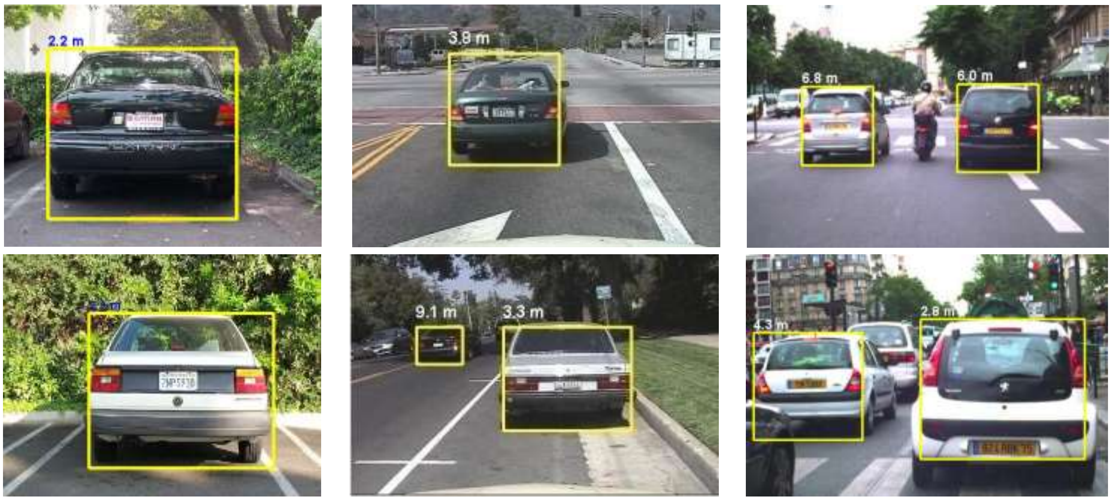


Figure 50 : Detection results for the data sets used; Left Column: Caltech DB (cars 1999), middle column: Caltech DB (cars 2001) and right column: Lara Urban Sequence.

The PVD system proposed was also qualitatively evaluated using images acquired under adverse weather conditions, downloaded from the internet. For these images, although quantitative results are not derived due to different and unknown acquisition systems setup, the proposed single image PVD system performs sufficiently well, yielding promising results as shown in few typical cases of Figure 51.

Further study of the FN errors shown, reveals that the most common cause is the presence of other red artifacts near rear lights, which leads the system to failure of detecting the real pairs of rear lights.

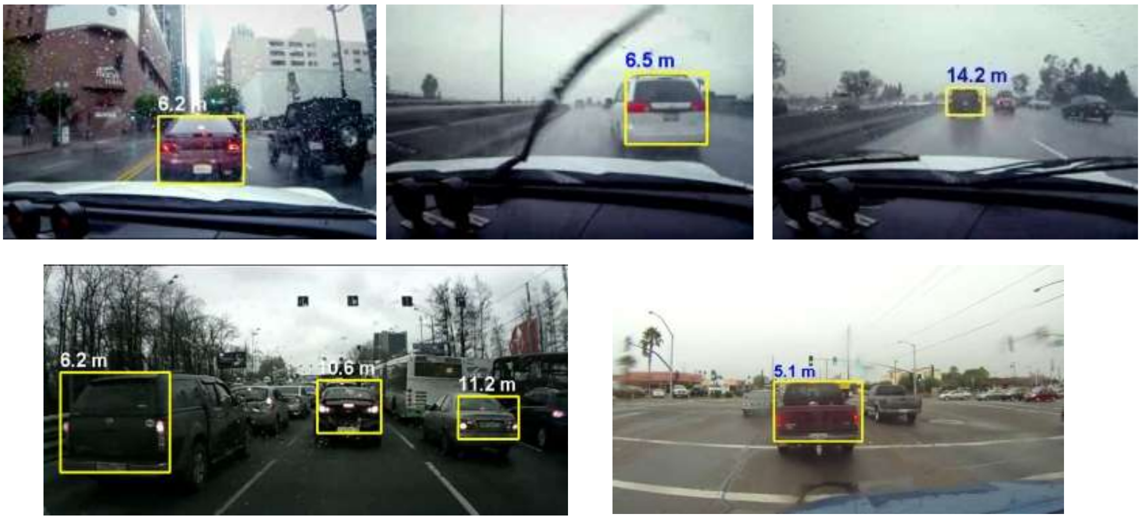


Figure 51 : Detected vehicles in adverse weather conditions.

5.5 Conclusions

The development of a robust and reliable vision-based vehicle detection method is a crucial task for driver assistance systems. In this chapter an automatic, resilient to illumination conditions algorithm for vehicle detection has been presented, using color and radial symmetry information for the segmentation of rear vehicle lights from a single image. After morphological lights pairing and edge boundaries detection, the vehicle presence is verified through symmetry detection in the candidate rectangular ROIs. Experimental results report high detection rates both in typical and in adverse conditions. The proposed algorithm, because of its approach of using rear lights for detection can be easily extended for vehicle detection at night. Future efforts are directed towards vehicle tracking and combining vehicle detection and braking recognition with driver's gaze detection. In this way, the level of attention of the driver can be correlated with the potential danger of an impending collision.

Chapter 6

Video-based Detection of Preceding Vehicles

6.1 Introduction

The vehicle detection method presented in the previous chapter was provided a solution for single images, not taking into account the added information that can be collected by processing successive frames of a video. The information that can be provided by analyzing the motion in the scene may provide valuable in the refinement of the detection results over time, minimizing the false detections and validate the results more efficiently. This chapter will focus on using the static framework presented in the previous method and enhancing it with spatiotemporal information processing to provide more accurate PVD results, even in challenging scenarios. The proposed system is described in [99]. It has been extensively evaluated in both normal and challenging scenarios, proving both effective and robust.

6.2 Literature review

A significant part of literature relevant to the subject of PVD has already been presented in the previous chapter. In this section, some methods that utilize motion information will be presented. O'Malley et al. [96] used the color distribution of rear-lamp pixels under real-world conditions to derive proper thresholds for the HSV color space, based on the automotive regulations for vehicle appearance. The detected rear lamp candidates are then paired using symmetry analysis and color cross-correlation and finally tracked Kalman filtering. Kalman tracking has been utilized again by O'Malley et al. in [100], in a method that also uses modified hardware by means of a polarizing filter fitted onto the camera lens to remove reflections. In their work, after detecting vehicle lights and prior to tracking them, a perspective correction is applied to the light pair areas to enhance their resemblance. Particle filters are used in [101], [102] for vehicle tracking purposes. In [101], different vehicle descriptors are used in combination with data-driven sampling in order to track the detected vehicles. Sivaraman and Trivedi [102] use a training process with an Adaboost classifier which is retrained through an active learning process to improve the vehicle recognition rate. The detected vehicles are then tracked using a particle filter.

In recent papers, numerous methods for vehicle tracking have been proposed. In [103], a SVM classifier of vehicles is combined with a Markov chain Monte Carlo approach for the implementation of the tracking stage. In [104], shape and color properties have been used for a mean shift based tracking method. A probabilistic motion model is constructed for vehicle tracking in driving video streams in [105].

Recently, large datasets of driving scenes video streams have also been made publicly available, proving very useful for benchmarking ADAS algorithms [22], [25].

In this chapter, a vehicle detection system based on the processing of video sequences is presented. It improves and extends the work in [86], which is based on the combined assessment of vehicle descriptors, such as the red color of their tail lights filtered in CIE-L*a*b* and the symmetrical appearance of their rear facade. An efficient reduction of the false negative rated is achieved by the HV stage where a Kalman filtering process handles the spatial coherence of the detections over time by tracking the rear lights area of the detected vehicle. The system has been experimentally evaluated in a variety of publicly available datasets of both still images and video streams, demonstrating superior detection and tracking rates compared to relevant methods.

The main contributions of the proposed system are the following:

(i) the mixture of red color information derived from a novel, robust to illumination fluctuations color channel with the FRST is used to detect the rear lights of the vehicle as symmetrical red blobs.

(ii) The usage of the SSIM index both in color and in symmetry images as a tool to filter out asymmetrical results and reduce the number of rear light candidate areas.

(iii) The usage of the Kalman filter as a sub-system for dynamic vehicle candidate verification, in combination with the SSIM based static verification process.

The rest of this chapter is organized as follows: in Section 6.3 an overview of the proposed system is described, while Section 6.4 discusses its modules in depth using typical examples derived from the evaluation datasets. In Section 6.5, quantitative performance evaluation for the proposed system in static images and video streams is performed and compared to relevant methods. Finally, conclusions about the contribution and future work are presented in Section 6.6.

6.3 Proposed system overview

The proposed system uses *a-priori* knowledge derived from observation of the most prominent vehicles descriptors and more specifically their rear lights and the axial symmetry of their rear facades. The HG stage is based on color segmentation and radial symmetry in order to detect candidate rear vehicle lights in each image frame.

The color and luminosity of rear lights, especially when lit, make them stand out in different weather and illumination conditions, in day and nighttime. Moreover, when rear brake lights are illuminated, they give an advanced warning of potential danger as they denote a decelerating vehicle. The aforementioned color related features of rear vehicle lights inspired the incorporation of a color segmentation process based on the $c_1c_2c_3$ color space, to avoid high sensitivity on shadows and illumination changes. A mixture of two channels is used to produce a single channel which robustly discriminates red colors from other bright spots. A FRST is applied to the estimated "redness". This is the second step of the HG, which ensures that non-symmetrical red blobs will not be considered as rear light candidates. The result of the FRST is searched for local maxima and the most prominent results are counted as candidate rear vehicle lights.

The candidates are passed on to the HV stage, which is based on another very typical specification of vehicles; their axial symmetry along the vertical bisector of the hypothetical line connecting their two rear lights. The axial symmetry detector has two parts; the first involves pairing the candidate lights depending on their inclination and relative distance and the second is the similarity check between the right and the left half of the sub-image containing the two paired candidates, based on the SSIM index [52]. Pairs with small inclination values and large similarity indices are accepted as rear vehicle lights and are used to localize vehicles in the image.

The HG and HV stages are very successful for static images. However, when dealing with video streams, the process of temporal information in the HV stage is beneficial for further reduction of false negative rates. In the proposed approach, a Kalman filter is used to estimate the position of vehicle lights in consecutive frames and consequently check the persistency and accuracy of the HG candidates in time. This has the effect of removing HG candidates that do not appear for more than a couple of frames, thus improving the precision of the system. Furthermore, it improves the spatial accuracy over time, as it provides additional information for the most appropriate candidates in each frame which are also consistent in time.

The functional overview of the proposed system is summarized in Figure 52.

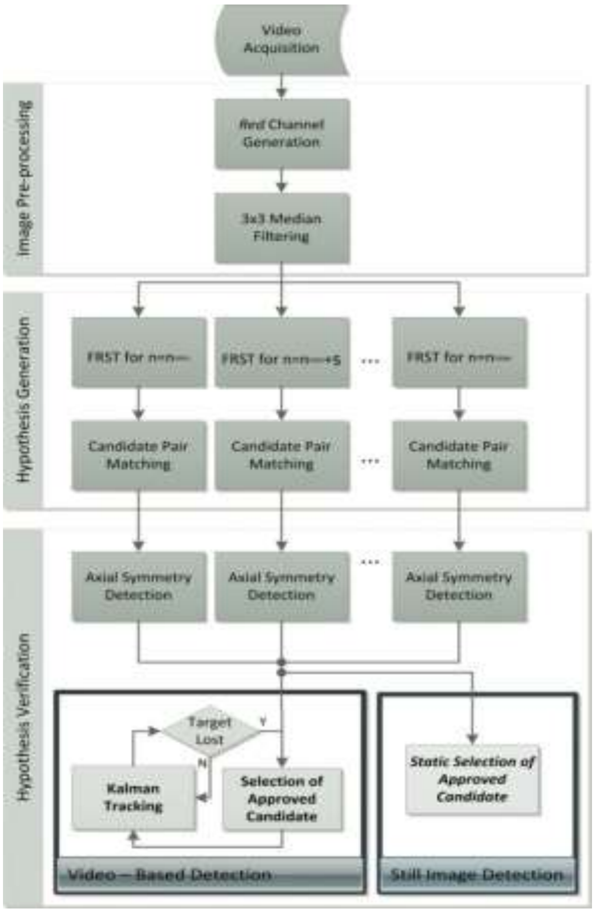

Video Acquisition
Image Pre-processing
Red Channel Generation
3x3 Median Filtering
Hypothesis Generation
FRST for n=n
FRST for n=n+5
...
FRST for n=n
Candidate Pair Matching
Candidate Pair Matching
...
Candidate Pair Matching
Hypothesis Verification
Axial Symmetry Detection
Axial Symmetry Detection
...
Axial Symmetry Detection
Target Lost
Y
N
Kalman Tracking
Selection of Approved Candidate
Static Selection of Approved Candidate
Video – Based Detection
Still Image Detection

Figure 52: Flowchart of the proposed system.

6.4 Modules analysis

6.4.1 Hypothesis Generation stage

In the HG stage, a rough first hypothesis for the locations of possible vehicles in the image is generated. The HG procedure involves red color segmentation of the original image, detection of highly radially symmetric red blobs and their pairing according to their relative topology. The respective stages of HG are analyzed in the following sections.

6.4.1.1 Red color segmentation

The first step of the HG is the color segmentation process that is based on the usage of the $c_1c_2c_3$ color space, which has been mentioned in Chapter 2. This color space is chosen due to its robustness to illumination variations and shadows and its channel values are calculated as

$$c_1(i,j) = \arctan\left(\frac{R(i,j)}{\max\{G(i,j),B(i,j)\}}\right), \qquad 6.1$$

$$c_2(i,j) = \arctan\left(\frac{G(i,j)}{\max\{R(i,j),B(i,j)\}}\right), \qquad 6.2$$

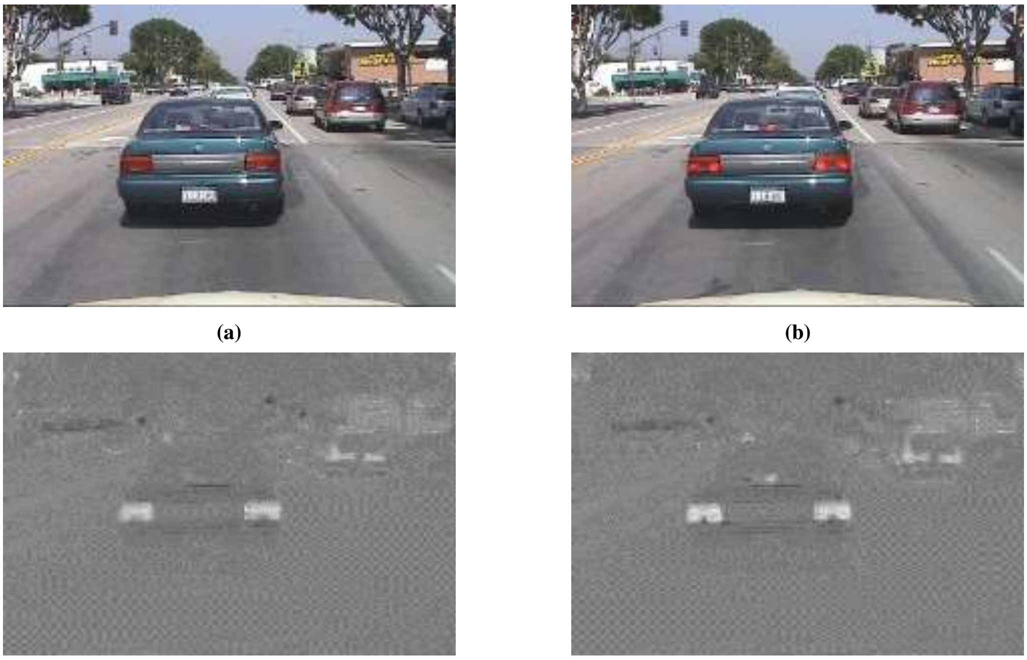
$$c_3(i,j) = \arctan\left(\frac{B(i,j)}{\max\{R(i,j),G(i,j)\}}\right), \qquad 6.3$$

where R, G and B are the red, green and blue channel value of the pixel at the i-th row and j-th column of an image, while c_1, c_2 and c_3 are the three channels of the aforementioned color space. Finally, $\max\{a, b\}$ denotes the maximum value of variables a and b.

Two of its three channels are then mixed to further enhance the color difference of the red colored pixels. More specifically, the proposed color channel used for red color segmentation is formulated as follows

$$Red(i,j) = \frac{c_1(i,j) + (1 - c_2(i,j))}{2}, \qquad 6.4$$

The new color channel is linearly related to c_1 and c_2 color channels and is based on the observation that rear lights are best enhanced by combining high c_1 channel values (high redness) with low c_2 channel values (low greenness). The intensity of the values of red pixels in the new *Red* color channel is even more pronounced when the rear brake lights are illuminated as demonstrated in the examples of Figure 53.



(a) (b)

(c) (d)

Figure 53: Highlighting of rear red lights using the new Red color channel. (a), (b) Preceding vehicle images without and with illuminated rear brake lights and (c), (d) their respective new Red color channel.

6.4.1.2 Radial Symmetry Detection

With the goal of emphasizing rear vehicle lights, the image derived from the red color enhancement is scanned for symmetrical patterns using the fast and efficient FRST presented in Chapter 2. The transform relies on a gradient-based operator which considers the contribution of each pixel to the symmetry, detecting radially symmetrical blobs in an image. Since the rear red lights are depicted in the proposed color channel as bright, symmetrical blobs, they generate large FRST values at their centers and therefore localize the candidate results of the HG stage.

The symmetry candidates are further enhanced if a 3x3 median filtering process is used to smooth the *Red* channel image, providing more accurate center localization. Although the rear lights of vehicles are not always circular like in the car that appears in Figure 53, they generally follow a symmetrical pattern and thus can still be detected by the FRST by using a low radial strictness parameter ($a = 1$). Some examples that demonstrate the successful highlighting of rear vehicle lights using the FRST are shown in Figure 54.

The results described so far have been achieved using just one radius. In order for the system to detect vehicle rear lights of differently sized vehicles at varying distances, the proposed system uses a set of radii that span from the minimum to the maximum expected radius in pixels. Each FRST image result is processed separately in the next step.

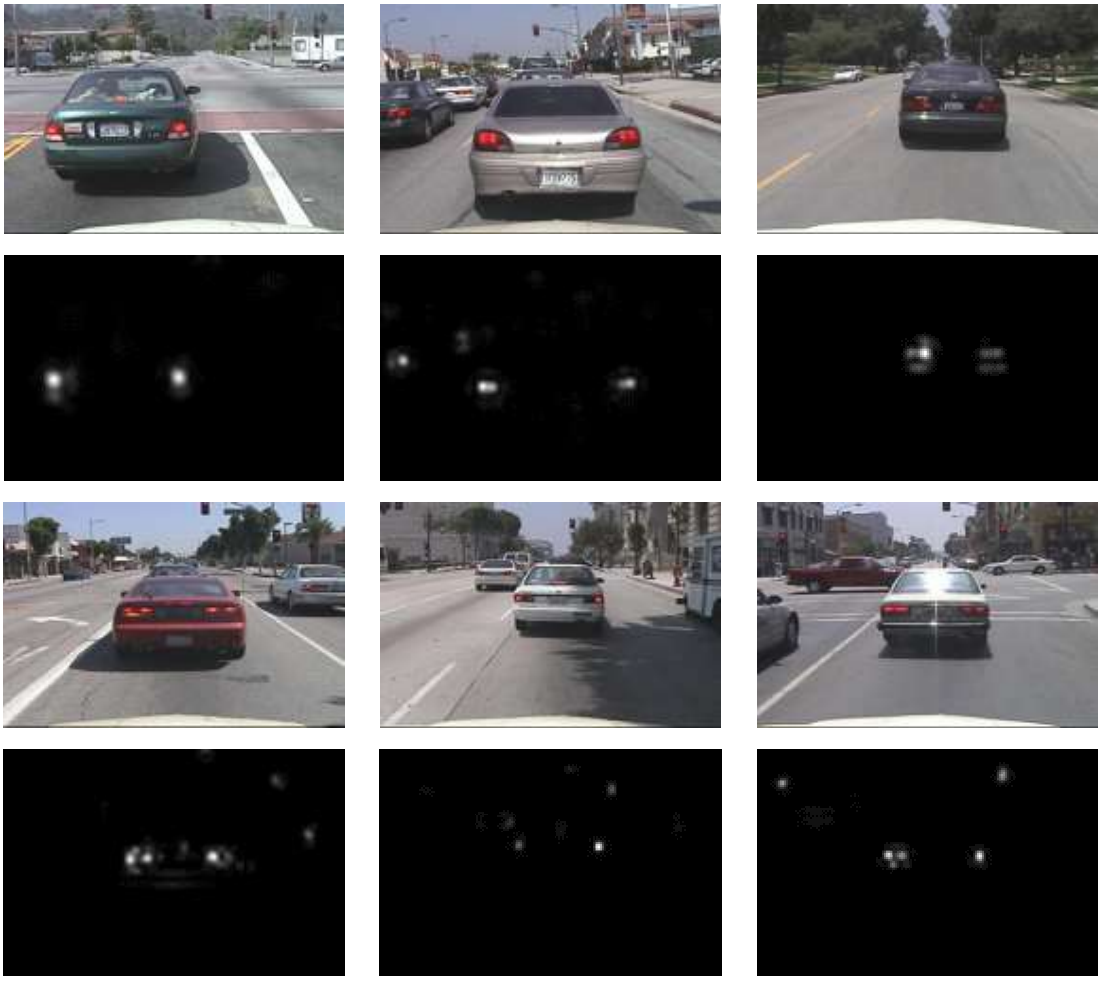
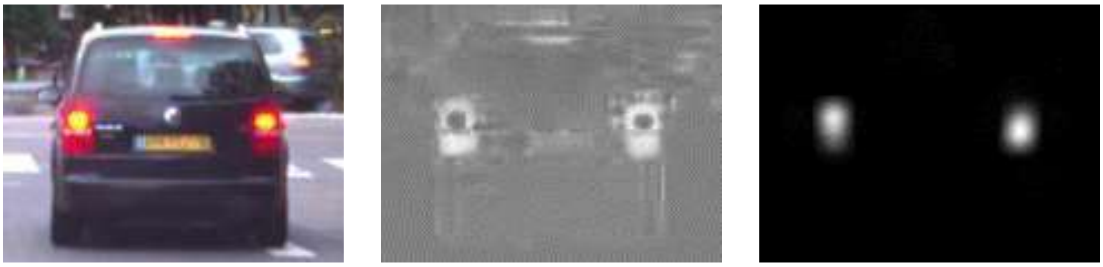


Figure 54: Various examples of FRST based detection of preceding vehicle rear lights. The original images appear in odd rows and the respective FRST results in even rows. The results have been estimated for one radius (8 pixels for top rows and 6 pixels for bottom rows).

Another significant advantage of the FRST usage is that it can handle efficiently the "blooming effect" that appears often in rear lights images, due to the saturation of very bright pixels in CCD cameras with low dynamic range. The robustness of the selected algorithm relies on the fact that saturated lights appear as dark spots surrounded by a bright halo, causing the FRST to treat them as symmetrical dark blobs. Such a result is demonstrated in Figure 55.



(a) **(b)** **(c)**

Figure 55: Cancellation of the "blooming effect" using the FRST. (a) Original image, (b) Red channel with presence of the blooming effect and (c) FRST result localizes rear lights.

Consequently, each image produced by the FRST is scanned for local maxima to pinpoint the centers of candidate rear vehicle lights of a given radius. To reduce the number of candidates and keep false positives low, the local maxima are filtered to include only values larger than a predefined ratio, *r*, of the maximum value.

6.4.1.3 Light candidates pairing

The detected rear lights candidates are subsequently processed so that only true vehicle light pairs are considered for further validation. This procedure excludes unwanted red artifacts from the detection, including two-wheel vehicles, which should be detected with alternative methods. If *N* is the number of detected light candidates, then all the possible pairs will be $(N-1)\cdot N/2$. However, only a few of these pairs can be considered as rear lights of a vehicle.

In order to reduce the number of valid light pair candidates, excluding false positive results, a set of heuristic rules is defined. More specifically, the two candidates must:

(i) be closer to each other than a predefined maximum distance, *d*, which is proportional to the radial of the candidates and

(ii) be closely aligned in the horizontal axis, i.e. have a permissible inclination of an angle, θ, smaller than θ_{max} degrees in absolute value.

The properties mentioned above are graphically depicted in Figure 56. If the predefined conditions hold, then the pair of light candidates is considered as valid and is passed on to the HV stage.

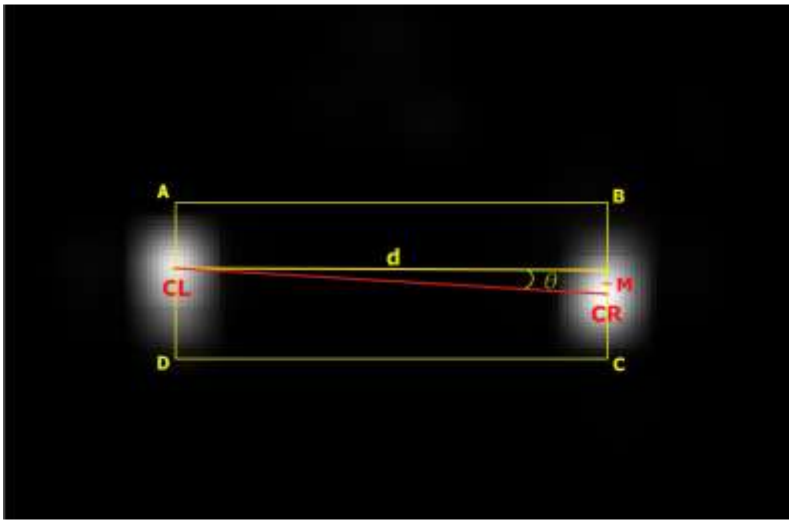


Figure 56: A pair of candidate rear lights with detected centers denoted as CL and CR. The criteria for accepting a candidate pair depend on their relative distance, d, and their relative angle, θ, which should be smaller than two predefined thresholds.

When a pair of symmetrical shape centers has been accepted as a valid candidate pair of rear lights, a rectangular Region Of Interest (ROI) around the two centers is cropped, so that the candidate pair is verified. The rectangular ROI, which is denoted in yellow in Figure 56, is chosen so that it includes the area of the rear of a vehicle that appears most symmetric about the vertical bisector. This area is defined by the rectangle ABCD, with AB=DC=*d*, BM=MC=*d* / 10, and M being the mean of the row indices of CL and CR.

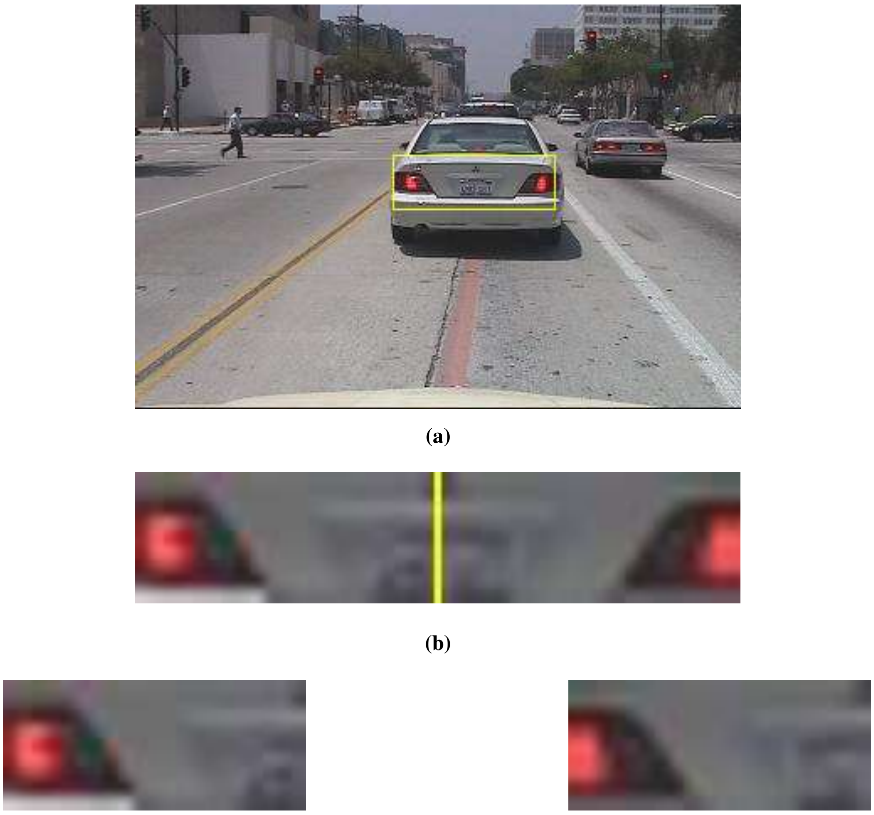
6.4.2 Hypothesis Verification stage

In the HV stage, the vehicle candidate areas defined by the light pairs selected in the previous stage are more closely examined. Each candidate area can be validated as a vehicle, by passing the two verification stages, i.e. the static and the dynamic stage.

6.4.2.1 Static verification

The static verification stage is based on the typical property of all vehicles, which is their rear view symmetry. More specifically, the rear part of vehicles is strongly symmetrical in the vertical direction [87], a property that is even more prominent when the processed area is confined around the detected light pairs.

The axial symmetry assessment is performed in a rather straightforward manner. First, a rectangular area surrounding the pair of candidate lights is cropped and divided into two equal parts (sub-images), along the vertical bisector. Then, the right sub-image is mirrored, i.e. flipped about the vertical axis. This process is summarized in Figure 57.



(a)

(b)

(c)

(d)

Figure 57: First stage of static verification process. (a) The detected ROI containing the pair of candidate lights. (b) Cropped region of interest with vertical bisector highlighted in yellow. (c) Left sub-image, (d) mirrored right sub-image.

The two sub-images are then compared in each channel of the RGB color space using the widely utilized Structural SIMilarity (SSIM) index discussed in Chapter 2. The SSIM measures the similarity of two images based on three comparisons regarding luminance, contrast and structure. The result is one single measure bounded to [0, 1]; the closer its value is to 1, the higher the similarity between the two images. The proposed process compares the mean SSIM index from the three color channels to a user-defined threshold, t_{SSIM}, to verify that the region of interest indeed contains a pair of rear lights.

While the SSIM criterion removes most false positives, the inclusion of many radii in the FRST stage may lead to multiple overlapping rectangular areas around the rear lights of a single vehicle. Furthermore, some false positive examples caused by pairing the rear lights of two different vehicles that are close to each other might still remain. Such examples are given in Figure 58.

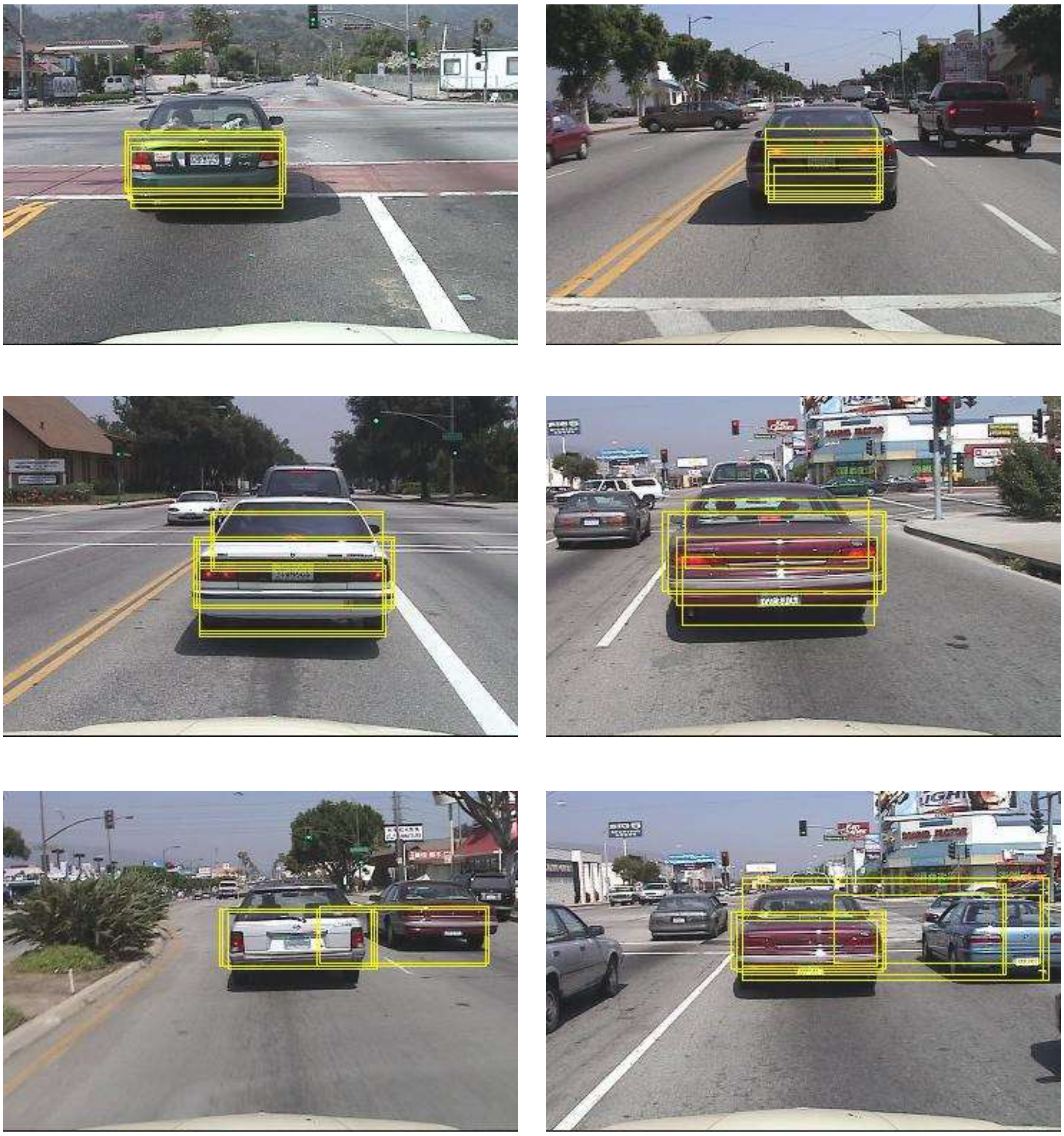


Figure 58: Multiple ROIs generated by the HV module.

To tackle these issues, a rule that limits the preceding vehicle detection results per frame to just one can be applied. To achieve this, the mean color SSIM indices of the detected ROIs are added to the SSIM indices of their mean FRST transform values. The ROI with the largest sum is kept as the accepted vehicle rear lights area. In this approach, the number of false positives is kept to a minimum (at most one per frame), while the detection rate of the method is not affected significantly. Some typical results achieved using this approach, are presented in Figure 59.

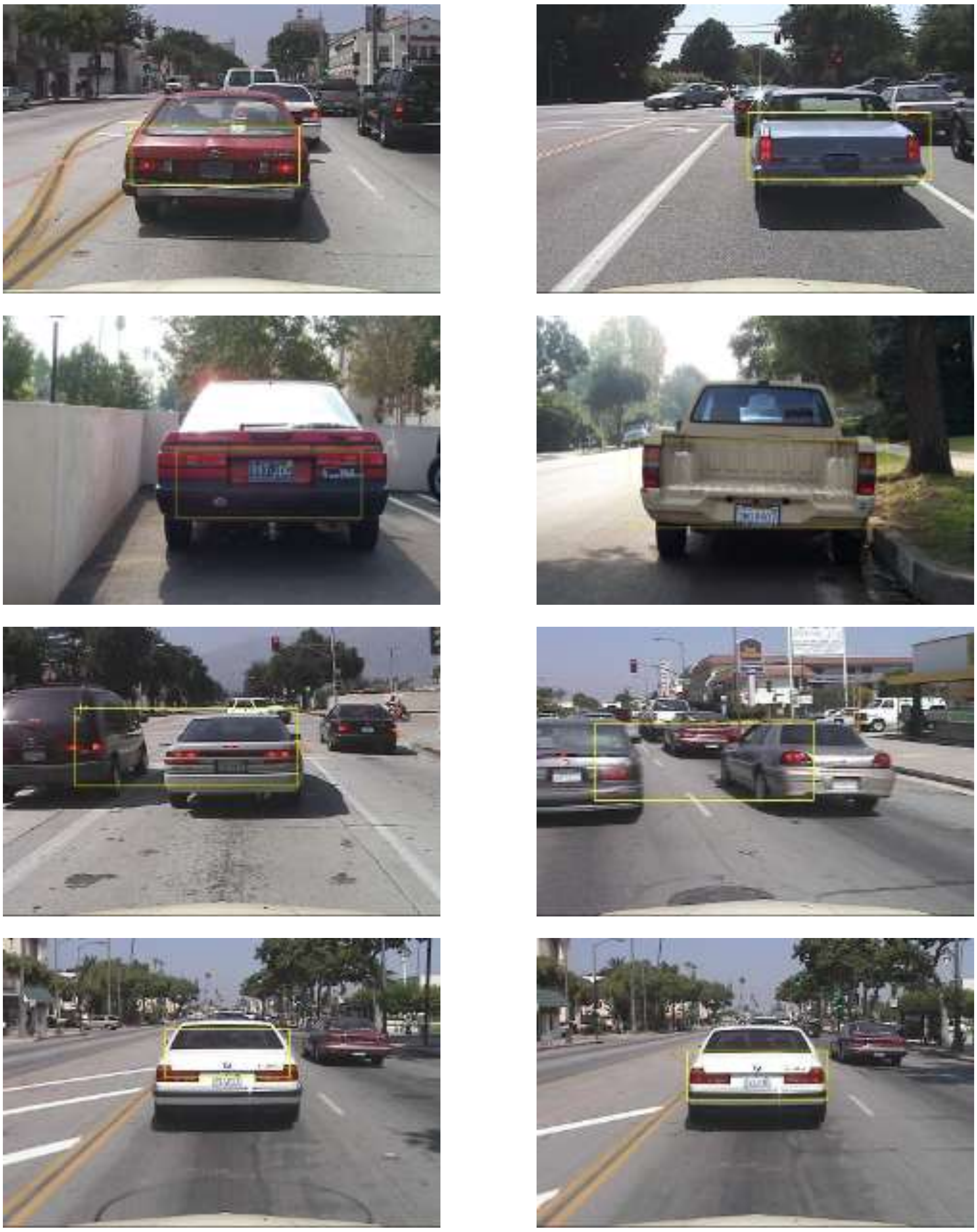


Figure 59: Successful detection results from Caltech 2001 (first row) and Caltech 1999 (second row). Two false positive results between two closely moving vehicles are shown in the third row. The problem of spatial inconsistency of the detection result in consecutive frames is demonstrated in the bottom row.

The two most common problems that may occur when choosing only the most prominent result based on the dual SSIM criterion are:

(i) the verification of an area between two closely moving vehicles in front of the ego-vehicle. This mostly happens when only few detection results lie on the preceding vehicle. Such examples are shown in the third row of Figure 59.

(ii) The verification of a different candidate area lying on the preceding vehicle for each frame, causing the detection results in video streams to appear unreliable in a few cases. While in a sequence of frames the ROIs accuracy may be acceptable since all verified candidates are on the vehicle, the exact location and size of the ROIs play an important role in the following processing steps or estimators, i.e. distance estimation of the preceding vehicle. Such an example is demonstrated in the last row of Figure 59, where the location of the detection has moved to a lower part of the vehicle between two consecutive frames.

These issues are the main reason for including an additional processing module in the verification process, which will also take into account temporal information and consistency rules.

6.4.2.2 Dynamic verification

As already described, the static verification process provides very promising accuracy rates, but it is not always sufficient for dealing with dynamically changing scenes such as a driving environment. Thus, the incorporation of a second verification step, taking advantage of the temporal continuity of video data, is beneficial for a more robust and effective vehicle detection system. The dynamic verification obtained by tracking the target vehicle, suppresses false negative detections, performs smoothing of the detection noise and provides information for the temporal association between frames by interpolating positions during sporadic erroneous or missed detections. Given that the relative movement of the preceding vehicle between two successive frames is small, the modeling of its movement with the linear discrete-time Kalman filter analyzed in Chapter 2 constitutes the optimal choice over other computationally demanding techniques such as the extended Kalman filter or particle filter. As already described, the Kalman filter is a recursive, adaptive technique that estimates the state of a dynamic system from a series of noisy measurements [44], [45]

The implementation of the Kalman filter is fast and straightforward, requiring only the tracking estimate of the previous frame. It estimates the state of a process by recursively updating the system dynamics in two phases - the *time update* phase and the *measurement update* phase. The *time update* equations project forward in time the tracking estimates (state vector and state error covariance matrix) of the previous frame in order to obtain *a-priori* predictions for the current frame. The predicted position is the expected measurement, given all the previous measurements and the mechanics of the system. The state transition matrix, which captures the mechanics of the system, is derived from the theory of motion under constant acceleration. The *measurement update* equations (denoted also as correction equations) are incorporating the new measurement into the *a-priori* prediction in order to obtain an improved, final estimate. A Kalman gain serves at reflecting the importance of the prediction to the current measurement which, in turn, depends on a measurement noise covariance matrix R. The Kalman gain is used to update the predictions, given a current measurement. A more thorough look into the equations and details of the Kalman filter can be found in [45] and [106].

The parameters of the target which are tracked throughout the image sequence are its position and size. Therefore, three observations are used in order to form the state vector; the x-coordinate, y-coordinate of the vehicle and its width W, because the aspect ratio is fixed. A long term observation of the coordinates can yield a description of the trajectory of the target vehicle in an image sequence, while its size and derivative of size can yield a description of the distance and the relative to the ego-vehicle velocity, respectively. For the specific application, the measurement noise covariance parameter R is used to determine the sensitivity of the tracker to updates, by assigning weighting between the current measurements and the previous estimates. For the specific application and given a frame rate of 25 fps, a value of 0.1 ensures a good tradeoff between responsiveness and smoothness during noisy measurements.

In order to initialize the tracking procedure, the position and size of the detected preceding vehicle are stored and monitored for several consecutive frames, to examine the reliability of the detection. Given that the false positives are not persistent in time, i.e. very rarely the same false positive appears in more than one frame, detections with major overlap for several consecutive frames ensure true vehicle detection. The issue of other than preceding vehicle detections like parked vehicles along the side of the road is effortlessly tackled due to their locations' deviation throughout the succession of frames. In the current approach, a consistent detection of the same vehicle for 3 successive frames with a less than 15% change in position and size denotes a 'true preceding vehicle' and becomes the tracking target.

After the initialization of the tracking procedure, for each new frame, the proposed vehicle detection module is executed. A notable difference is that the output detections are not suppressed to just one in this case. Using the Euclidian distance of the Kalman based prediction for the vehicle location and size from all detections derived in the static verification stage, the closest static estimation to the Kalman prediction is kept. If the distance and size difference of the Kalman ROI prediction to the closest static ROI detection are both below 15% of their maximum value, the Kalman uses the chosen detection as the current measurement and performs the *measurement update* stage. Otherwise, the existence of a preceding vehicle is examined within the respective rectangular Kalman ROI prediction, using the same methods with the static detection module. In particular, the two criteria are the existence of two local maxima on either side of the vertical bisector of the rectangle, a strong marker for the presence of rear vehicle lights and a prominent axial symmetry, as defined in the static verification step using the same threshold value. In the case where at least one criterion does not hold true, the target is considered as lost and the tracking procedure is terminated, until a new 'true preceding vehicle' is found.

Along those lines, tracking of the preceding vehicle permits the suppression of false positive detections, the irregularities due to temporary missed or wrong detections and the overall accuracy and robustness of the system is improved.

6.5 Experimental results

The Preceding Vehicle Detection system proposed in this chapter has been experimentally evaluated using both static images found in the Caltech 1999 and 2001 datasets in [31] and video frames from the LARA video stream presented in [22] and the HRI1 through HRI5 frame sequences presented in [25].

6.5.1 Quantitative results using single images

In this section, detection accuracy of the proposed system will be presented using the Caltech 1999 and 2001 datasets already described in Chapter 1. The experimental results were achieved using 1/4th of the original image resolution for the first subset (160x120) and a 1/25th of the original resolution for the second subset (180x119). The radii used for the FRST estimation were 3 to 13 pixels for the first subset and 3 to 16 pixels for the second, with a step of 1 for both. In both cases, radial strictness a was set to 1 and the result of the FRST was estimated for the lower 70% of the images so that only conspicuous areas are included and the computational complexity is reduced by 30%. The ratio used for the non-maximum suppression of the FRST result was set to $r = 1/4$. The rectangular ROIs selected were bounded in $[2n, 20n]$, where n is the detected radial in pixels and the permissible inclination of two detected centers was set to $\theta_{max} = \pm 3$ degrees. Finally, the threshold of the mean SSIM of all color channels was set to $t_{SSIM} = 0.95$.

The aforementioned datasets have been used as a benchmark in several research papers. One of them is a vehicle detection system proposed by Wang and Lien in [98], who follow a statistical approach using PCA and ICE analysis in chosen sub-regions of the image to detect vehicles. Other researchers use the datasets as a benchmark for more generic object detectors. Such attempts were made by Dalal and Triggs [107] who used Histograms of Oriented Gradients to train a linear SVM, or Fergus et al [108] who use a probabilistic, scale-invariant learning scheme to classify objects, so the results acquired from the proposed system can be compared to them for evaluation purposes.

Using these datasets, the proposed system achieves superior results compared to relative works, as shown in Table 5. The recall rate of the proposed method is directly comparable to the best effort by Wang and Lien in [98] (both score 94.2%), scoring at the same time a considerably higher precision rate (95.1% versus 87.5%). Moreover, the proposed method does not require any training process, a significant advantage in applications where a great variety of video scenes can be met, as opposed to those in [98], [107], [108].

Table 5: Vehicle detection rates in both Caltech datasets

	Recall	Precision
Proposed System	**94.2%**	**95.1%**
Wang and Lien [98]	**94.2%**	87.5%
Dalal and Triggs [107]	89.4%	87.5%
Fergus et al [108]	84.8%	87.5%

It should be noted that the second subset (Caltech 1999) is not the typical scenario that has to be handled by a collision warning system. For this reason some more interesting conclusions can be drawn using only the first subset that comprises more typical driving scenes.

More specifically, the images included in the first subset can better characterize the efficiency of the proposed system, since it also includes frames where the brake lights of the preceding vehicle are illuminated. These scenes are of even higher importance, since they can be used for collision warning in cases of sudden braking of the preceding vehicle.

In the Caltech 2001 dataset, the recall-precision values are comparable to the ones mentioned above. However, when the frames containing vehicles with rear lights modified beyond legislation, i.e. 13 frames from 150-158 and 253-256, are removed from the dataset, the recall of the proposed system in the remaining 513 frames rises from 93.2% to 95.3%. Furthermore, the detection rate for braking preceding vehicles reaches 98%. Taking these facts into account, a conclusion that can be drawn is that the proposed method is suitable for collision warning purposes, since it successfully detects the braking preceding vehicles. A detailed evaluation of the proposed method applied to the Caltech 2001 dataset of 526 images taken from driving scenes (358 braking) and the filtered dataset of 513 images (356 braking), is given in Table 6.

Table 6: Vehicle detection rates in Caltech 2001 dataset

	Recall	Precision
Full dataset of 526 images	93.2%	94.2%
Filtered dataset of 513 images	95.3%	95.9%
Braking - Full Dataset	97.7%	98.0%
Braking – Filtered Dataset	98.0%	98.3%

6.5.2 Quantitative results using videos

The proposed system has also been evaluated using the video dataset LARA presented in [22] and discussed in Chapter 1, which comprises images of driving in an urban environment. Only frames that contain unoccluded preceding vehicles at the same lane and in a distance less than 15m were considered for our experiments (2331 out of 11179 frames). The experimental setup involved resizing the images at 1/4th of the original image resolution (320x240), a ratio of r = 1/4 for the non-maximum suppression of the FRST result and radii from 3 to 10 pixels. The threshold of the mean SSIM of all color channels was set to $t_{SSIM} = 0.99$. Table 7 presents the detection rates before and after tracking.

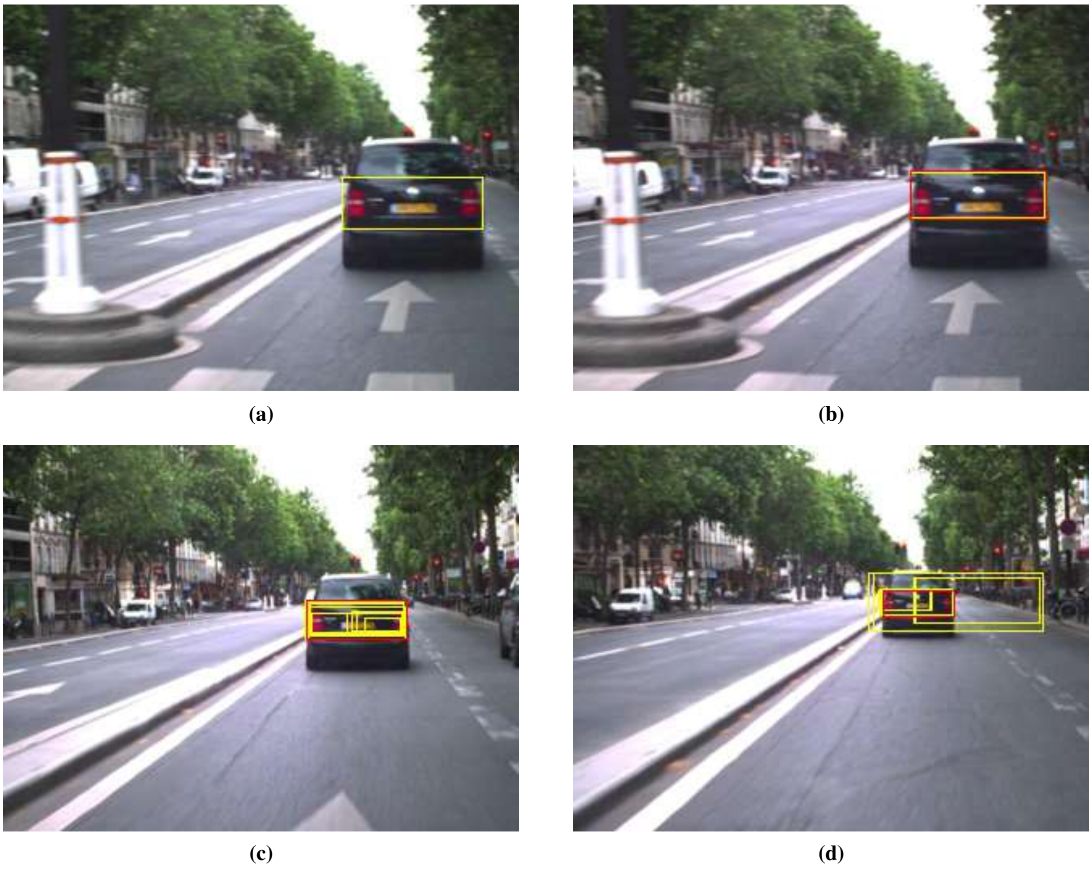
Table 7: Vehicle detection rates before and after tracking

	Recall	Precision
Before tracking	72.5%	73.6%
After tracking	93.1%	94.4%

The results of Table 7 provide supporting evidence that the tracking procedure significantly increases the vehicle precision rate. In most cases false detections are met only for a small number of frames, commonly due to abrupt changes in illumination or frame to frame jitter caused by camera shaking. In these cases, the tracking predictions are used until the target is redetected. Moreover, while the real vehicle might not be in the most prominent detection, in most cases the rule used to limit the detection results to one per frame, reaches the real vehicle position.

The experimental results denote a substantial improvement in terms of both recall and precision, when Kalman filtering verification replaces the combined SSIM verification proposed for still images. The improvement of approximately 20% for both measures is a clear indication of the great influence of the scenes complexity to the static detector and the significance of spatiotemporal processing in such applications. More specifically, the great majority of still images comprising the Caltech 1999 and 2001 datasets include only one vehicle, in a relatively small distance from the camera, while in the LARA dataset, there is a great diversity of scenes, preceding vehicle distances and number. All these factors contribute to the raised importance of a temporally adaptive verification process, such as the Kalman approach proposed here.

The images shown in Figure 60 illustrate instances of tracking results in an image sequence taken from the LARA video dataset. In Figure 60(d) an example of successful elimination of false positives is presented.



(a) **(b)**

(c) **(d)**

Figure 60: Example of Kalman tracking in the LARA video dataset. The vehicle detections are shown in yellow rectangles while the red rectangles denote the target being tracked.

6.5.3 Qualitative results in adverse conditions

The system proposed in this chapter has also been evaluated for robustness to more adverse conditions, using the HRI road traffic video datasets presented in [25], which were also described in chapter 1.

A qualitative assessment reveals that the results achieved by the proposed system in the aforementioned conditions are very promising; therefore it could be used as a starting point for an ADAS that will be robust to any kind of conditions. Some typical examples of vehicle detection using the aforementioned datasets are given in Figure 61.

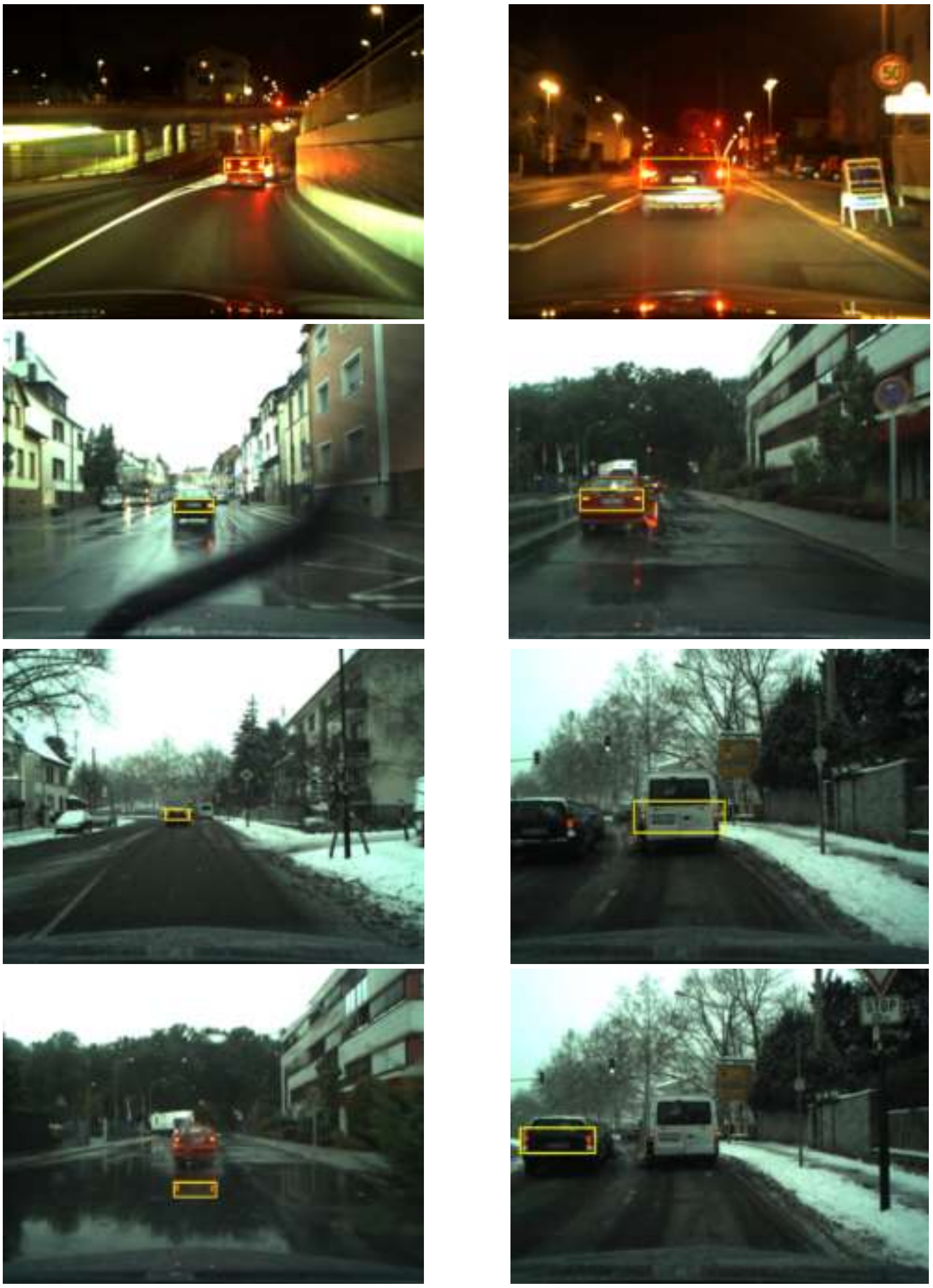


Figure 61: Examples of vehicle detections from the proposed system in adverse conditions. Top row: night driving. Second row: driving in the rain. Third row: driving in snow. Fourth row: Typical false positive examples.

In the first three rows of Figure 61, successful detections from night driving, driving in the rain and driving in snow are presented. In the last row, two typical examples of false detections are depicted. The example on the left shows a false positive caused by the glare of rear lights on the road, while the right image shows an example of vehicle detection that is on the opposite lane. While this is not a strictly false positive error, the preferable result would be the detection of the vehicle on the right. Both results would benefit from an incorporation of a road, or lane detection module that would add useful information for such cases.

6.5.4 Performance issues

As the goal of every ADAS is to be implementable for real time processing purposes, a discussion on the processing speed of each module of the proposed system is of high importance.

For the purposes of this analysis, the mean processing time values required by the system to detect the vehicle position for 100 frames of three different video resolutions have been compared. All tests were conducted for a set of 8 different radii used by the FRST. The aim is to pinpoint the bottlenecks of the system and conclude if an implementation would be feasible for real time applications. All parts of the system have been implemented in Matlab, except a small part of the FRST which has been converted from C++ to Matlab executable. No code parallelization has been used and the tests were run on a 2.8GHz Quad Core processor. The response time performance of various processing modules from these experiments is presented in Table 8.

Table 8: Absolute and relative processing times and frame rates analysis for different resolutions. (AS: Axial Similarity, FRST: Fast Radial Symmetry Transform, PM: Pair Matching, MF: Median Filtering, Others: Other modules).

		AS	**FRST**	**PM**	**MF**	**Others**
160x120	**%**	39	20	15	8	18
	ms	56	29	22	12	26
	T/fps	145ms / 6.9fps				
320x240	**%**	15	33	20	10	22
	ms	49	108	65	33	72
	fps	327ms / 3.1fps				
640x480	**%**	10	41	21	9	19
	ms	138	567	290	125	263
	fps	1383ms / 0.7fps				

The results of Table 8 show that the axial similarity module is the bottleneck of the system when using the smaller possible resolution (1/8th of the original). As the resolution increases, the bottleneck of the system appears in the FRST module, with the pair matching process being the second most time-consuming stage. The fourth place in terms of processing time per frame is steadily occupied by the median filtering process.

A very important implementation issue is the potential increase in processing speed that can be achieved by a proper parallelization of processes. In the proposed system, such a solution is feasible thanks to the lack of correlation between the results of each radius. Most core modules can function separately for each radius, thus multiplying their processing speed by the number of radii used. An approximate 70% of the processing time (spent on the first three modules of Table 8) can be accelerated by a factor defined by the number of radii used, providing real-time performance for smaller resolutions, even in Matlab. Real time processing in higher resolutions can be achieved by DSP implementations.

Another implementation performance issue has to do with the fact that the FRST scans for symmetrical blobs of certain radii, which are experimentally derived. This is an inherent disadvantage of the algorithm, as the precise localization of several symmetrical shapes of unknown, different sizes should require the use of a large number of radii. However, the computational burden can be alleviated by using only a sparse set of radii, spanning from the smallest probable radius of a rear light to the largest one, using a fixed step, or variable step taking into account the expected size. The result of such a process is a sufficient approximation of the output obtained when the whole set of possible radii is used [36], provided that the step is selected accordingly.

6.6 Conclusions

In this chapter, a novel system for vehicle detection based on the localization and tracking of their rear lights has been proposed. The system comprises a static hypothesis generation and verification stage, combined with a Kalman tracking module used for enhancing the spatiotemporal coherence of the detection results and reducing the number of false negative detections. The static stage utilizes a novel, robust to illumination changes color space, which is correlated with a radial symmetry transform to generate a first set of rear lights candidate regions. This set is refined using an axial symmetry criterion based on a similarity metric and the resulting candidates are verified by a Kalman tracker that picks the prominent candidate based on its spatial consistency compared to previous frames.

The system has been extensively evaluated using quantitative and qualitative methods, on publicly available datasets of both still images and video streams. The results show that it outperforms similar methods when compared in still images, while it achieves excellent performance on video streams, even in more adverse environments and conditions, such as driving in rain, snow and at night. Typical detection problems in such cases would benefit from the combination of the proposed system with other ADAS modules, like road, or lane detection.

Chapter 7

Road Detection

7.1 Introduction

In this chapter, an algorithm for Road Detection (RD) that can be integrated in an ADAS for collision avoidance and automatic navigation is proposed. It has been published in [109] and it is based upon the Random Walker Algorithm (RWA) for image segmentation described in Chapter 2, involving also a novel automatic seed generation stage that uses an illumination-invariant optical flow estimation derived from the $c_1c_2c_3$ color space. The information derived from the optical flow are combined with *a-priori* knowledge about the camera setup and segmentation results from previous frames to produce a set of seeds that provide a clear discrimination between road and non-road pixels under a vast diversity of conditions. Methods proposed so far on the field of RD are described and evaluated in the next section and then the proposed system and all its modules are described. An extensive experimental evaluation of the proposed system follows, including qualitative assessment in different conditions and quantitative comparative assessment against other efforts on two publicly available, manually annotated video datasets. Several aspects of the RD system implementation, like resolution and processing complexity, are then studied and proper solutions are proposed. Finally, a discussion on the advantages and disadvantages of the proposed system is carried out.

7.2 Literature review

A popular approach to the problem of road detection is the use of lane markings. Those markings are localized to acquire boundary information which facilitates the road detection process. Methods that rely on lane markings are usually fast and simple, using mainly grayscale images or videos. A pioneer work presents fast mapping from 3D to 2D followed by a horizontal edge detection process [110]. Later work based on edge extraction is presented in [111], [112], [113]. These approaches often fail to perform satisfactorily in more complex scenarios like rural unstructured roads or poorly maintained city roads.

A second popular method in road detection applications is the use of color or brightness information to segment the road, enhanced by some feature extraction process like edge detection to extract the road boundaries. Such efforts are presented in [114], [115] using unsupervised and supervised Bayesian segmentation respectively. Watershed segmentation is used to differentiate the road from the background [116], based on gradient information. Edge information for road boundary detection has also been used in [117-120]. In [118] edge and area information is combined to achieve fast and robust results in urban traffic scenes. Such algorithms are usually fast and accurate in well-controlled environments, however they have not been thoroughly tested in adverse conditions therefore their accuracy in ill-defined road boundaries is not guaranteed. A very challenging issue that often deteriorates the results of such methods is the presence of shadows, which is a very common condition in real-life road scenes.

To improve road detection accuracy and robustness to shadows, many researchers have utilized more complex methods, by processing information related to optical flow [121], [122] and stereo vision acquired from camera pairs [123-127]. Among the most recent stereo vision methods a homographic approach is induced in the road plane and formulated as a maximum a posteriori problem in a Markov Random Field and presented in [127]. This formulation allows for the implementation of an optimization algorithm alternating between road detection and learning the optimal parameters from the current stereo image pair. In [128-130], the *a-priori* knowledge of a road model is used to validate regions of interest as road. Additionally, many researchers propose the use of illumination invariant color spaces, like the well-known HSI [131], or the training-based illumination invariant color space proposed in [24], used along with a likelihood-based classifier to achieve road segmentation. The method presented in [24] has also been applied in [126], to enable stereo camera head pose estimation using a featureless registration between the detected roads from two simultaneous frames acquired from a stereoscopic pair of video cameras. A supervised training algorithm is used to construct a probabilistic road model and combine visual information with a posteriori probability to acquire a "roadness" map in [132].

From the study of the literature, several important conclusions can be drawn:

i) The accuracy of the road detection algorithms depends mainly on the background complexity and the environmental conditions adversity.

ii) Algorithms that oversimplify the problem decrease the response time, but tend to produce poor results in adverse conditions, complex, or unstructured environments.

iii) The use of extended set of features, like motion or depth descriptors, increases robustness but results in slower processing times, especially for large video resolutions.

iv) Many approaches don't take advantage of information that derives from the nature of the problem, i.e. the use of video instead of still frames, or using information about the camera setup in the vehicle to focus the road detection on areas that are expected candidates, i.e. excluding the upper part of the frames.

v) The videos used for assessment are rarely made publicly available and usually tend to exclude adverse conditions like heavy rainfall, fog, snow, night driving etc. In excellent weather conditions and well-structured environments i.e. highways with distinct lane markings, the road detection problem can be solved efficiently with minimal computational complexity effort.

However, as weather conditions deteriorate and more complex environments are met, i.e. narrow city roads in traffic, urban roads with tarmac anomalies, etc., the road detection problem becomes increasingly difficult. Additional effects like reflections on the windshield, large number of shadows, night driving, foreign particles on camera lenses, etc., their synergy and the requirements for real-time response and cost-effectiveness, make the problem of road detection using a single on-board camera even more challenging.

The aim of the proposed system is to perform RD using a monocular color video camera, minimizing the required assumptions and processing the relevant information of the driving scene. Neither road lanes nor road models are used as they are problematic in cases like dense traffic, dead ends, poorly maintained roads etc. Furthermore, the adopted training methods are efficient, avoiding non-parametric models which require a vast amount of training video scenes in the great variety of adverse driving conditions.

7.3 System overview

The road detection system is based on the information captured by an on-board monocular color video camera. The camera is placed on an elevated position on the windshield of the vehicle and it does not disturb the driver's field of view. The hardware setup in the proposed approach is important mainly for the initialization of *a-priori* spatial information, namely the area of high road probability and the area of high non-road probability.

The brightness of the pixels is processed in pairs of consecutive frames to extract spatiotemporal information that will be used to define road and non-road pixel seeds. The seeds are fine-tuned using the segmentation results from the previous frame and used by the RWA for the estimation of the road pixels.

In the segmentation process, the RWA solution of the Dirichlet problem [47] is used. This approach significantly reduces the computations, while providing the same segmentation accuracy as the classic RWA [48], [49]. The implementation details and the rationale of using the RWA have already been discussed in Chapter 2.

However, the inherent supervised nature of the RWA remains a significant problem for its incorporation in a road detection system. To tackle this situation, optical flow information combined with spatial and previous road segmentation information is used to define at each frame a robust set of seeds. Hence, the segmentation process becomes completely automated, relies only on dynamic information extracted from the on-board camera, the a-priori knowledge derived from the camera setup and the previous frame segmentation results. The RWA estimates a probability map for all pixels in a video-frame, and the pixel-based segmentation process is completed after thresholding the probability map.

The functional overview of the proposed system is demonstrated in Figure 62 along with some descriptive test images.

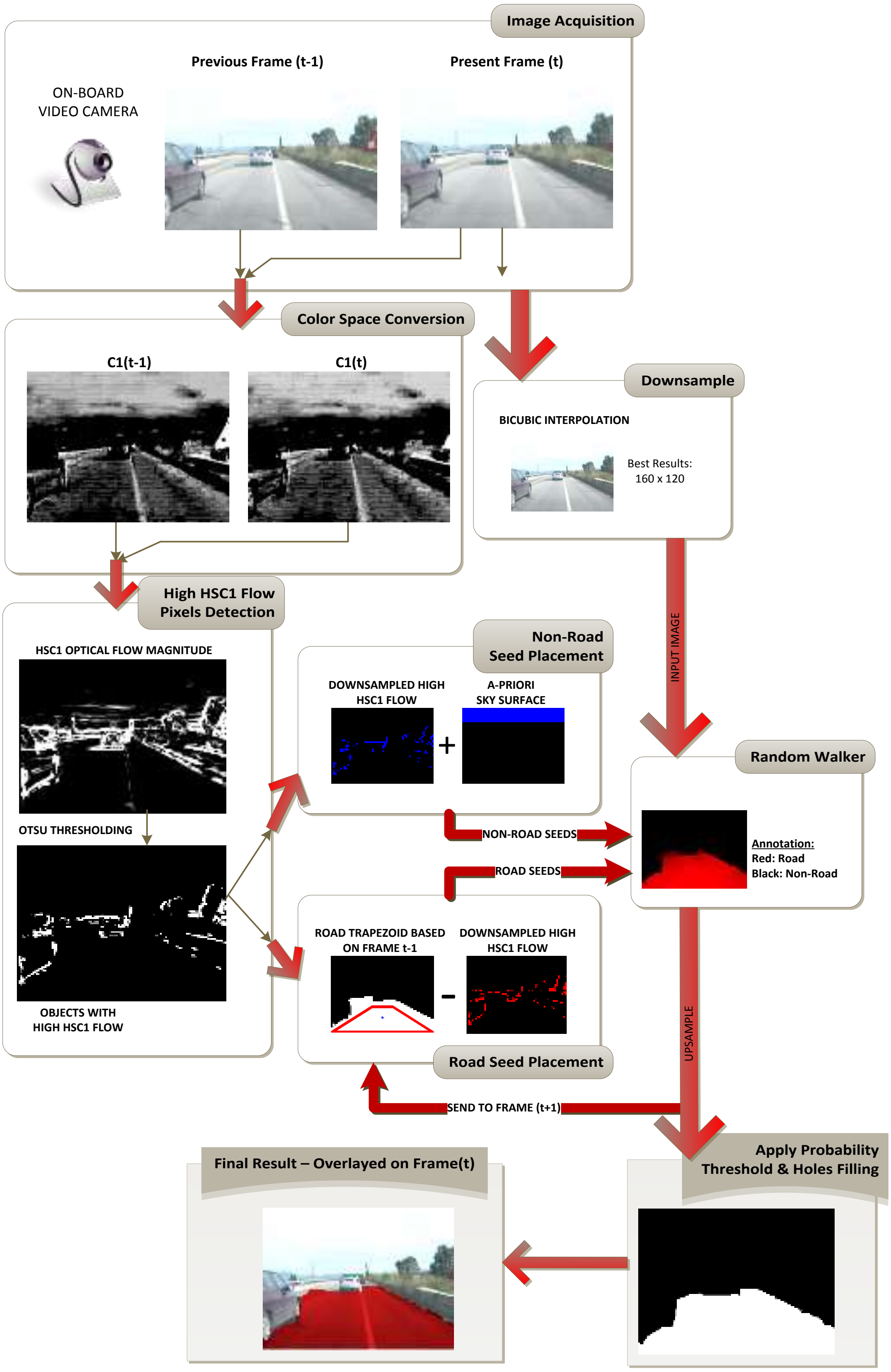


Image Acquisition
Previous Frame (t-1)
Present Frame (t)
ON-BOARD VIDEO CAMERA
Color Space Conversion
C1(t-1)
C1(t)
Downsample
BICUBIC INTERPOLATION
Best Results: 160 x 120
High HSC1 Flow Pixels Detection
HSC1 OPTICAL FLOW MAGNITUDE
OTSU THRESHOLDING
OBJECTS WITH HIGH HSC1 FLOW
Non-Road Seed Placement
DOWNSAMPLED HIGH HSC1 FLOW
A-PRIORI SKY SURFACE
+
INPUT IMAGE
Random Walker
NON-ROAD SEEDS
ROAD SEEDS
Annotation:
Red: Road
Black: Non-Road
ROAD TRAPEZOID BASED ON FRAME t-1
DOWNSAMPLED HIGH HSC1 FLOW
-
Road Seed Placement
SEND TO FRAME (t+1)
UPSAMPLE
Apply Probability Threshold & Holes Filling
Final Result – Overlayed on Frame(t)

Figure 62 : Flowchart of the proposed Road Detection System.

More specifically, the proposed system comprises the following steps:

a) *A-priori* information about the expected position of the sky in the frame is estimated after the camera is installed in the vehicle. This area will be used to assist in discrimination of non-road candidates.

b) Consecutive pairs of frames (*t*, *t*-1) in the $c_1c_2c_3$ color space are processed, to extract motion information.
c) The c_1 channels from both frames are used to estimate the Horn - Schunck optical flow magnitude matrix, from hereon called HSC1 flow.
d) The HSC1 flow is thresholded using the Otsu technique, to discriminate high HSC1 flow non-road areas (obstacles) from low HSC1 flow (textureless) surfaces, therefore road candidates.
e) High HSC1 flow non-road areas are combined with the *a-priori* sky region of step (a) to define the non-road seeds.
f) The road detection result from frame *t*-1 is used to construct the perimeter of a trapezoid. The pixels of this trapezoid are defined as road seeds, unless they already belong to the non-road seeds set.
g) The RGB image of frame *t* is downsampled to a lower resolution to reduce the computational complexity of the RWA.
h) The image from (g) and the seeds defined in (e) and (f) are used by the RWA to estimate the road probability at each unlabeled pixel of frame *t*.
i) The probabilities matrix is thresholded, the resulting image undergoes a hole filling process [83] and then is upsampled to the original resolution.

The proposed feature extraction process can be separated into two independent processes, which can be implemented in parallel:

i) The automatic seed selection process, combining *a-priori* spatial information, previous frame road detection results and spatiotemporal information. It comprises a color space transformation followed by the optical flow estimator and the Otsu thresholding process.
ii) The resolution of the image is reduced using bicubic interpolation and used by the RWA for constructing the L_{ij} matrix of Eq. 2.32.

7.3.1 Spatiotemporal seeds definition

In the proposed system the **road seeds** are located in a predefined area in front of the car and the lower part of the frame. These pixels must exhibit similar mobility in terms of optical flow magnitude and they also must have high probability of belonging to the road surface in the road detection process applied to the previous frame. The **background** or **non-road seeds** are located in a predefined number of rows on the top of the frame, depending on the position and viewing angle of the camera and on pixels that should also have similar optical flow magnitudes, which differ from the ones of the road seeds. Based on the aforementioned definitions, the seed definition can be divided into three parts.

First, the **a-priori information** derived from the nature of video based road detection is applied. The seed selection has been dependent on camera characteristics and the expected seeds positions must be updated every time the camera position, focal length or camera lenses are changed. Depending on the elevation and angle of view, a number of rows in the top part of the frame are arbitrarily defined as non-road seeds.

As shown in Figure 62, these rows typically comprise the sky surface. Furthermore, the road seeds could, based on the camera setup, its characteristics and the nature of the problem, be placed on a trapezoid perimeter in the bottom part of the image. The trapezoid shape, which is initialized manually once the camera is setup, is based on the typical structure of the unobstructed straight road as it appears in images taken from inside a vehicle.

However, the size and coordinates of the trapezoid do not remain fixed, but should be calculated based on the road detection result derived from the **previous frame**, as the vehicle start moving. More specifically, as depicted in Figure 63, let R be the centroid of the detected road in the previous frame and the upper bound defined as the horizontal line passing from the road pixel with the highest vertical coordinate. Also, let N be the pixel of the upper bound line which has the same vertical coordinate (Yc) as the centroid. The bottom base (CD in Fig. 3) of the trapezoid is the same as the base of the road in the previous frame. The upper base of the trapezoid (AB in Figure 63) is equal to the perpendicular distance of the centroid to the upper bound (RN). Finally, the height of the trapezoid is such that its upper base is equidistant to the centroid of the previous road and the upper bound of the previous road. The process described ensures that the road seeds placement is performed dynamically for each frame, taking into account previous results. Furthermore, the placement of the center of the upper base, subject to the location of the centroid (both placed at column Yc, as shown in Figure 63), makes the placement more adaptable for curved roads. Finally, the proportionate length of the upper base of the trapezoid to the perpendicular distance of the centroid to the upper bound, leads the trapezoid to be more adaptable to the shape of the road detected in the previous frame. A typical image with the geometrical properties used to define the road and non-road seeds, is shown in Figure 63.

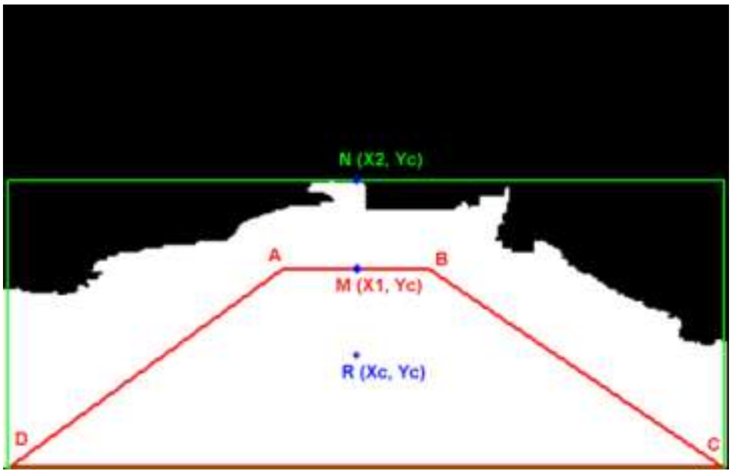


Figure 63 : Geometrical properties used for road (red trapezoid) and non-road seeds definition.

The definition of the seeds is completed by incorporating **spatiotemporal information** to detect non-road pixels belonging to obstacles like vehicles, pedestrians, trees, etc. This type of detection is rather difficult to implement without a stereo vision sensor that could offer structure information. However, in the proposed system a modification of a well-known optical flow algorithm is used to tackle this problem. Taking into account that the road is generally textureless, many sparse optical flow algorithms compute low motion values. Thus effective discrimination of the road from obstacles lane markings and shadow edges can be achieved. This behavior can be characterized as flawed when accurate optical flow estimation is needed; however it proves useful for discriminating road and non-road pixels.

A classic optical flow algorithm that can behave as mentioned, especially when stopped after few iterations, is the one proposed by Horn and Schunck [40]. This method is also fast enough for applications like road segmentation. Its main disadvantage for this application is that it is based on grayscale images, due to the brightness flow smoothness assumption, leading to high sensitivity to shadows and illumination changes. In order to tackle this problem, the proposed system uses the c_1 channel of the $c_1c_2c_3$ color space introduced in [35], which is derived by

$$c_1(i,j) = \arctan\left(\frac{R(i,j)}{\max\{G(i,j), B(i,j)\}}\right), \qquad 7.1$$

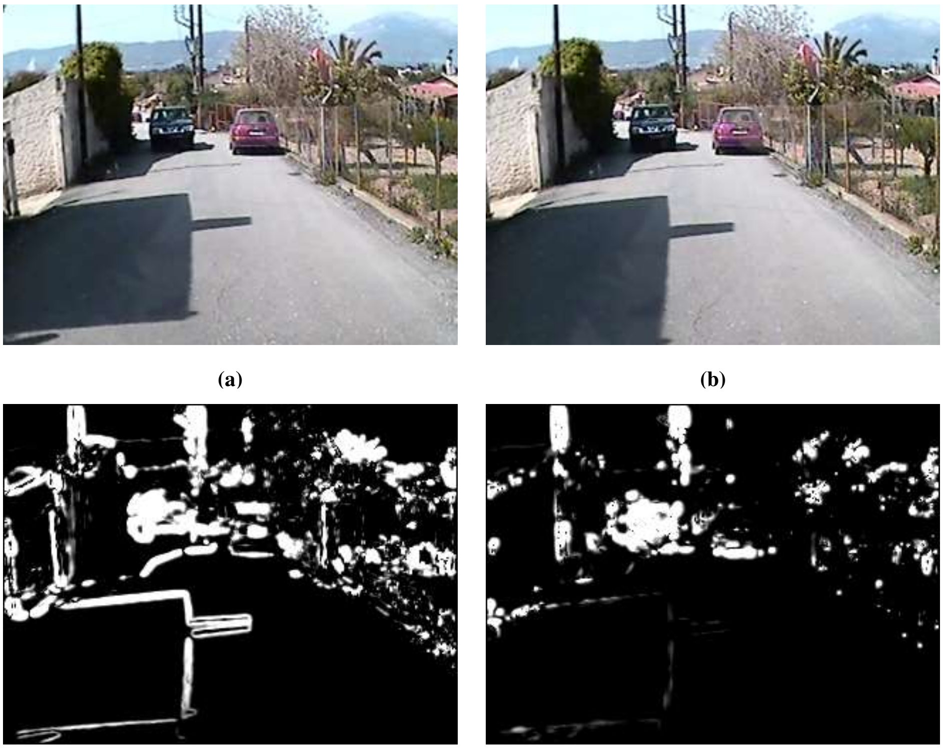
where R, G, B are the red, green and blue brightness value of pixel (i,j). This channel was experimentally found to be the most efficient color channel among the separate channels of CIE-L*a*b*, HSV, YCbCr and RGB color spaces to replace the brightness channel used for robust optical flow estimation in the presence of shadows [40]. Using the c_1 channel for the estimation of the horizontal (u) and vertical (v) optical flow velocities, Equations 2.16 and 2.17 are transformed to

$$V_x^{k+1} = \bar{V}_x^k - \frac{C_{1x}(C_{1x}\bar{V}_x^k + C_{1y}\bar{V}_y^k + C_{1t})}{\alpha^2 + C_{1x}^2 + C_{1y}^2}, \qquad 7.2$$

$$V_y^{k+1} = \bar{V}_y^k - \frac{C_{1y}(C_{1x}\bar{V}_x^k + C_{1y}\bar{V}_y^k + C_{1t})}{\alpha^2 + C_{1x}^2 + C_{1y}^2}, \qquad 7.3$$

where the superscript k denotes the number of iterations, $\bar{V}_x^k$ and $\bar{V}_y^k$ are the average velocities of the last k iterations, α is a regularization constant and C_{1x}, C_{1y}, C_{1t} are the normalized gradients of channel c_1 brightness estimated as described in [40], i.e. it has been linearly transformed to a [0,1] space. The optical flow calculated using the aforementioned color space will hereafter be called **HSC1 flow** (Horn Schunck optical flow using the c_1 color channel). The number of iterations for the HSC1 flow estimation was set to 1, leading to a low computational complexity, spatially sparse result. In the experimental section it is shown that any number of iterations between 1 and 5 is sufficient.

A very descriptive example of the HSC1 effectiveness is shown in Figure 64(c), where a great portion of the road pixels located on the edges of shadows appear to have high optical flow magnitude values when using the method as described in [40]. As shown in Figure 64(d), the HSC1 flow is definitely more reliable than the classic Horn Schunck optical flow for the discrimination between road and non-road seeds.



(a) **(b)**

(c) **(d)**

Figure 64 : (a), (b) Two consecutive frames in the presence of shadows. (c) Normalized optical flow magnitude for grayscale images, (d) normalized HSC1 flow magnitude. Both flow magnitude images were estimated for a=15 and k=5 iterations. Bright pixels denote high optical flow magnitudes.

All the pixels in the image with HSC1 flow magnitude greater than a threshold are defined as non-road seeds. These pixels are removed from the road seeds set if they are located on the trapezoid perimeter that was calculated in the previous step. The adopted thresholding method is the Otsu algorithm described in [37]; a very popular choice in thresholding images with bimodal histograms, i.e. typical two-class segmentation problems. This is apparently the case in the HSC1 flow magnitude thresholding problem, as shown in Figure 64(d), where the high brightness pixels are the non-road class and the low brightness pixels contain both road and non-road pixels. Using Otsu's algorithm to calculate a threshold *th*, that separates the high and low HSC1 magnitude pixels, yields

$$HSC1 = \begin{cases} high, & if\ \sqrt{V_x^2 + V_y^2} \geq th \\ \\ low, & if\ \sqrt{V_x^2 + V_y^2} < th \end{cases}, \qquad 7.4$$

where V_x, V_y are the horizontal and vertical HSC1 flows as estimated using Equations 7.2 and 7.3respectively.

Once the thresholding operation is complete, the non-road seeds are updated to include both the *a-priori* defined "sky region" and the high HSC1 flow class. In addition, the road trapezoid is refined by excluding all non-road seeds that are placed on its perimeter and keeping only long connected segments of the perimeter as the road seeds class, i.e. removing very small line segments that are very close to the non-road seeds and usually belong to an obstacle rather than the road. This refinement is performed in order to incorporate spatiotemporal information in the road seeds definition. As a result, the trapezoid area defined as road could become open instead of closed, thus forcing the RWA to decide what happens when moving obstacles disrupt the drivable area, leading to efficient handling of near obstacles. Some typical results of road and non-road seeds definition in various conditions are presented in Figure 65. The results show that this approach of seed selection is very robust even in challenging conditions. Some problems occur from the large number of falsely placed non-road seeds only in extreme cases of rainfall where visibility is extremely reduced.

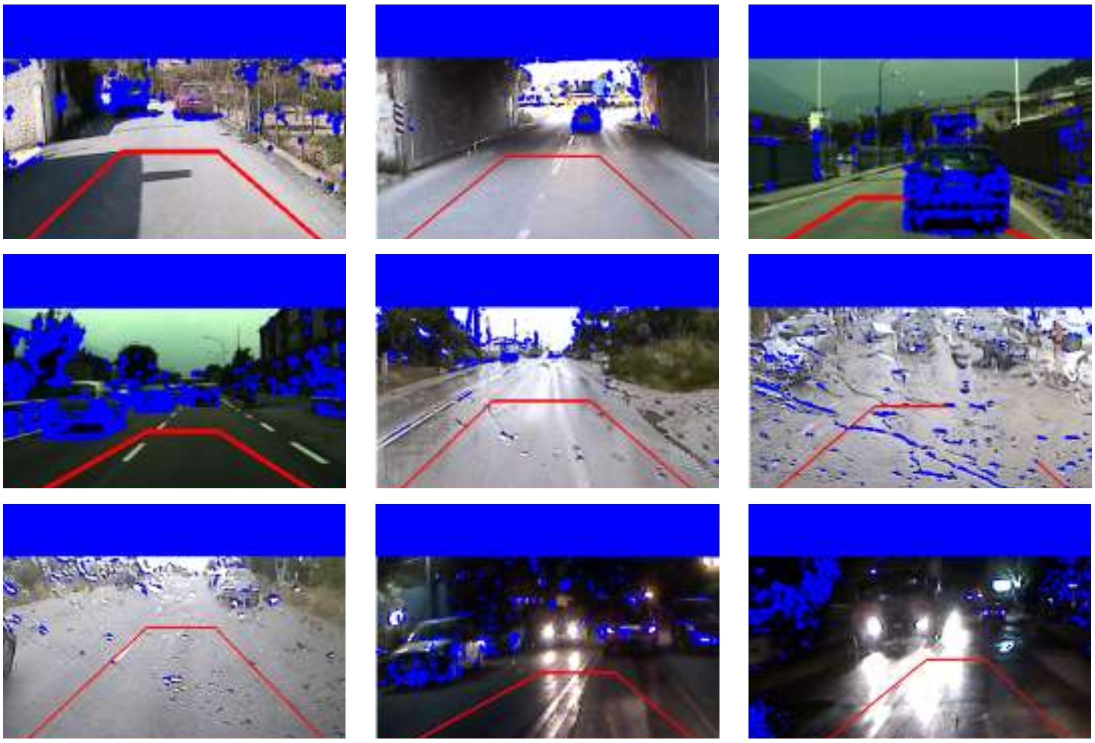


Figure 65 : Non-road (blue pixels) and road (red pixels) seeds using the proposed method in various conditions.

7.3.2 Image Resolution Reduction

Despite being fast, the adopted RWA implementation is still the computational bottleneck of the proposed system and the reduction of the image resolution to be segmented was explored, to further reduce the RWA processing time. A spatial reduction not only does not deteriorate the accuracy of the road detection algorithm but on the contrary, as shown in the next section, increases the road detection quality. The optimal results were observed for images half the size of the originals (160x120 pixels instead of 320x240). The trade-off between road detection accuracy and RWA processing speed is also discussed.

7.4 Experimental results

Three types of video recordings are used in the evaluation experiments:

i) typical road scenes used for qualitative results (datasets Patras1 and DIPLODOC),

ii) fully adverse conditions recordings used for qualitative evaluation in challenging environments (from the datasets Patras1 and Patras2 as described in Chapter 1),

iii) mixed conditions recordings with ground truth annotation (DIPLODOC dataset) used for quantitative evaluation of the proposed road-detection system in comparison to [123], [128] and

iv) challenging conditions recordings with ground truth annotation (Alvarez1 and Alvarez2) used for quantitative evaluation of the proposed road-detection system in comparison to [24].

7.4.1 Qualitative Road Detection results in normal conditions

The system was initially tested in videos shot under normal weather lightly adverse conditions, and mid-day illumination conditions, with the car speed limited to (20-120) Km/Hour. The goal was to evaluate the proposed method in typical conditions, especially in terms of classification correctness. The videos that were classified as having been shot under normal conditions were acquired primarily on daytime without intense sunshine, so that only soft shadows appear in the video scenes. These conditions along with well-structured roads and clear road boundaries produce excellent road detection rates as shown in Figure 66, where a set of road detection results in such conditions is demonstrated. Some temporal inconsistencies might be encountered for image areas with road probabilities near the threshold of 0.5. These can be removed by updating the segmentation probability threshold, instead of the fixed value of 0.5 used in the proposed algorithm. However in the few cases where such inconsistencies are met, they concern scene areas that are not close to the ego vehicle, therefore the significance of the error is lower.

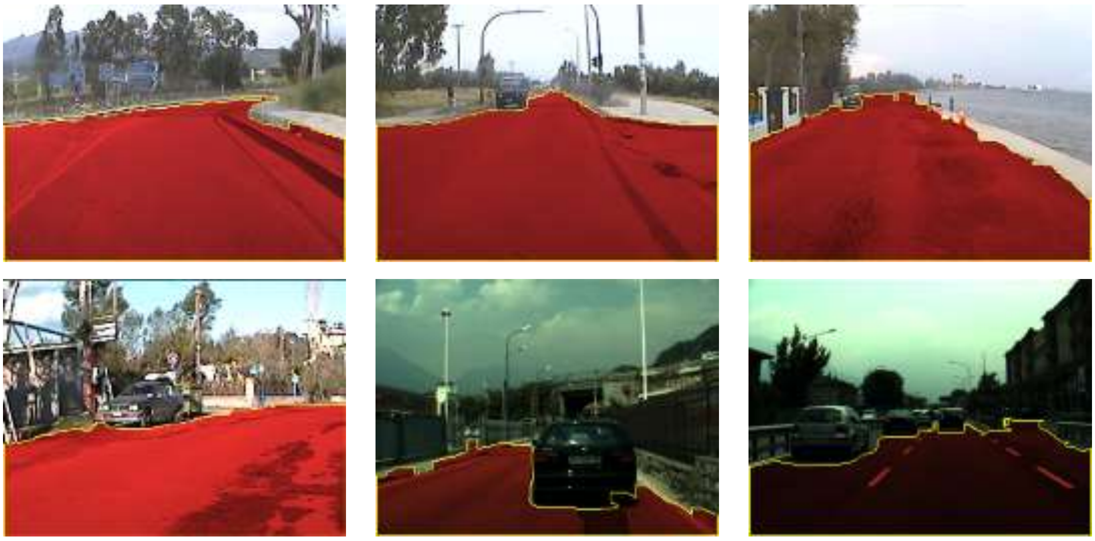


Figure 66 : Road detection results in normal driving conditions using the proposed system.

7.4.2 Qualitative Road Detection results in adverse conditions

The quality of the proposed road detection system is assessed in more adverse conditions with very promising results. Although a quantitative approach is very hard due to the overwhelming variety of adversities in real driving videos, some useful deductions about the advantages and disadvantages of the proposed system can still be drawn by examining the qualitative results in adverse conditions or challenging environments:

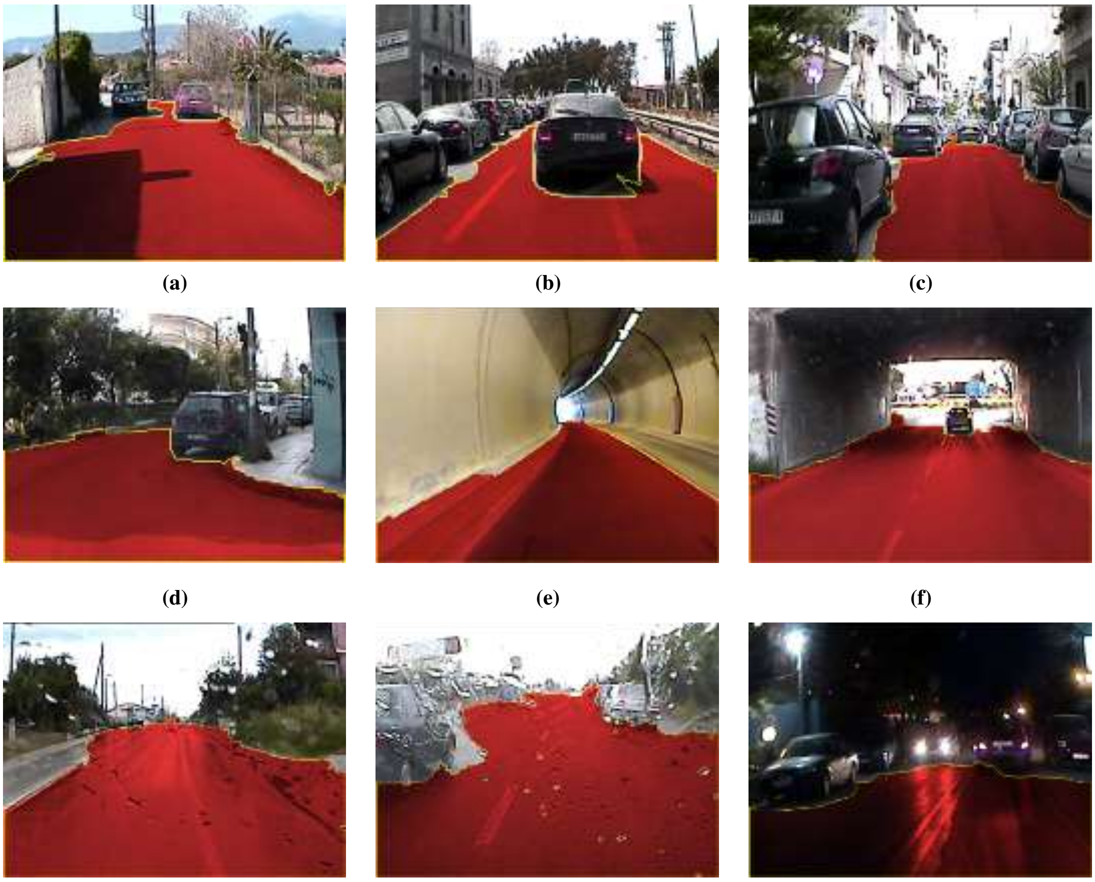
i) The **presence of shadows** is handled effectively, especially in cases where they are close to the ego-vehicle. More specifically, in some cases where the ego-vehicle moves from a sunny area to a shadowed area of the road, or vice-versa, a false detection result resembling the one generated when the ego-vehicle passes over a speed bump is produced. This means that the detected drivable area in front of the ego-vehicle gets smaller, until the trapezoidal seed area forces the RWA to include the more distant part of the road, classifying it correctly as belonging to the road. Examples of successful road detection in scenes with shadows are demonstrated in Figure 67 (a) through (d). One example of the partially problematic result mentioned above is shown in Figure 68(a). However the system quickly adapts and when the ego-vehicle gets closer to the shadow, the shadowed area is correctly recognized as road as shown in Figure 67(d).

ii) The case of **passing through tunnels or bridges** is a typical problem in such systems, due to the sudden illumination changes when entering or exiting. The proposed method appears robust in such scenarios and gives results that are temporally adaptable. More specifically, when entering tunnels, the detected drivable area in front of the ego-vehicle falsely reduces in size (Figure 68(b)), but only for a few frames; when the car enters the tunnel and the camera gain automatically adjusted to the lighting conditions, the proposed system re-estimates the road area using the new frames information producing reliable results as can be seen in Figure 67(e) and Figure 67(f). The opposite effect is met when exiting a tunnel, where the saturated area of the image frame at the exit of the tunnel might be misclassified as road surface. This effect also lasts a couple of frames before the analog circuits of the camera adapt the acquisition gain to the illumination change. After gain correction the proposed method produces reliable segmentation results.

iii) Qualitative results in **rainy conditions** have also been included in the assessment of the proposed method. In cases of light rainfall, the system provides excellent segmentation accuracy. Road detection accuracy is reduced in case of heavy rainfall combined with ineffective camera position, i.e. above the effective area of the windshield wipers (Figure 68(c)). If the field of view lies in the effective area of the windshield wipers, the segmentation accuracy is quite satisfactory except from the frames where the wipers pass in front of the camera (Figure 68(d)). However, this is a temporary error and the algorithm recovers in next frames. A more sophisticated approach would be to post-process the segmentation result taking into account information about the movement of the wipers masking the wiper area, or using previous frame information. Heavy rainfall also can cause some holes in the detected road area, because large raindrops produce high HSC1 flow rates (Figure 67 (h) and Figure 68(c)). This problem can be tackled by the addition of a post processing step of image hole filling, but this could result in misclassification of small obstacles as road.

iv) The scenario of **night driving** is handled efficiently, depending on the lighting conditions of the scene, as well as on the sensitivity of the optical sensor. When the scene is sufficiently lit and the camera depicts night scenes efficiently, the algorithm performs very well (Figure 67(i)); however in the case of badly lit scenes in conjunction with an optical sensor with low quality performance in night scenes, there appears to be some degradation in segmentation quality (Figure 68(e)). This is a problem that might be handled more efficiently with the use of infrared cameras, but the proposed system was not tested with such data.

v) Road detection accuracy is also affected by the blocking effect caused by **video compression** algorithms. This might cause high HSC1 flow values in areas that should normally appear textureless, thus causing the algorithm to include non-road seeds on the road surface. Such an example appears in Figure 68(f).



(a) (b) (c)

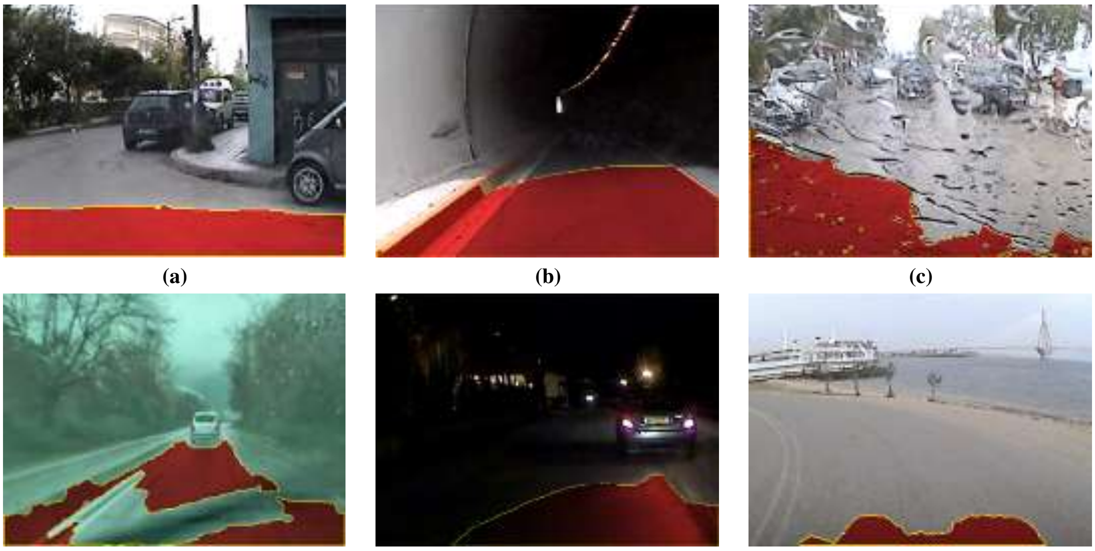
(d) (e) (f)

(g) (h) (i)

Figure 67 : Successful road detection results in adverse conditions. (a)-(d) Various scenes with heavy shadows, (e) passing through a tunnel, (f) passing under a bridge, (g), (h) rainy conditions and (i) night driving.

Some of the less successful detections described in detail above, are shown in Figure 68. In most cases, the system improves the detection result significantly in the next few frames.

Apart from the results discussed in this subsection, the proposed system also showed robust in cases where the camera shakes at times, as well as in cases of driving in roads with many turns. These scenarios, as well as more qualitative results in different video formats can be viewed at: http://www.wcl.ece.upatras.gr/en/gsiogkas.



(a) (b) (c)

(d) (e) (f)

Figure 68 : Less successful road detection in particularly challenging conditions. (a) Approaching shadowed road, (b) entering tunnel, (c) heavy rainfall with camera above the effective area of the wipers, (d) wiper passing in front of camera, (e) night driving in badly lit road, (f) misclassification due to high rate video compression.

7.4.3 Quantitative Road Detection results in mixed conditions

The proposed system was quantitatively evaluated using the five manually annotated video streams from the DIPLODOC dataset provided in [23], including mixed condition scenes. The efficiency of the proposed road detection system was directly compared to [123], [128] giving better segmentation accuracy. The comparison is based on the metrics of quality, recall and precision, presented in sub-section 1.7.3.

Using the aforementioned measures, [123] report several experiments using a method based on stereo vision, achieving an average quality factor over the five videos between 85% and 90%. These results are achieved with neither the correctness nor the completeness ever dropping below 90%. In fact, the trade-off between correctness and completeness almost reached 95% along the diagonal of a plot of these two measures.

In [128], P. Lombardi et al have reported that the monocular road model switching method proposed achieved quality results that fluctuated from approximately 60% to 95% through the entire video stream. Unfortunately, the mean values of the quality measures were not reported, instead three plots demonstrating the aforementioned quality of several experimental setups for all the frames were provided.

When the frame-by-frame performance is measured, the proposed system achieves a quality fluctuation from 74% to 99%. It is clear that the proposed road detection system outperforms the efforts in both [123] and [128]. A comparative view is given in Table 9. The statistics of the proposed method have been estimated using the best system setup and parameter settings, which are: resizing of the frame to be segmented to 160x120 pixels, optical flow estimation for a=3 and iteration number k=1 and β=70 for the RWA.

Table 9: Overall performance of proposed RD algorithm compared to [123] and [128].

		Method in [123]	**Method in** [128]	**Proposed Method**
Quality, $\hat{g}$	***Mean***	85% ~ 90%	N/A	**92.8%**
	Std	N/A	N/A	3.6%
	Min	N/A	60%	**74.0%**
	Max	N/A	95%	**99.1%**
Recall, *R*	***Mean***	N/A	N/A	95.7%
	Std	N/A	N/A	3.5%
Precision, *P*	***Mean***	N/A	N/A	96.9%
	Std	N/A	N/A	2.7%

For a more qualitative view of the results, each of the video streams should be observed and analyzed separately, because the driving conditions among videos differ significantly, as shown in the five columns of Table 10.

Table 10 : Description of DIPLODOC video streams conditions

Sequence Number	Frames	Environment / Traffic	Vehicles	Shadows	Lighting
1	451	Suburban/light	4	Yes	Sunny
2	150	Suburban/none	0	Yes	Sunny
3	100	Suburban/light	2	Yes	Sunny
4	60	Urban/dense	6	No	Gloomy
5	104	Highway/light	2	No	Sunny

In Table 11, the accuracy of the proposed road segmentation method shows robustness in all DIPLODOC sequences, achieving mean segmentation quality above 92%, with a small inefficiency in sequence 3, where the mean quality drops to 89.9% affected by a low completeness rate (93.5%), caused by increased false negatives rate. The standard deviation (Std) of the quality never raised above 3.7%.

Table 11 : Performance of proposed RD System in DIPLODOC video streams

Sequence Number		**1**	**2**	**3**	**4**	**5**
Quality, $\hat{g}$	***Mean***	92.8%	92.9%	89.9%	93.4%	94.8%
	Std	3.7%	3.7%	3.1%	2.0%	2.2%
Recall, *R*	***Mean***	96.1%	97.1%	93.5%	93.7%	94.8%
	Std	3.7%	3.3%	3.0%	2.1%	2.2%
Precision, *P*	***Mean***	96.5%	95.5%	95.9%	99.7%	99.9%
	Std	2.6%	2.5%	2.0%	0.1%	0.1%

A simplified presentation of the proposed method limitations is shown in Figure 69. Low quality scenes from video stream 3 of DIPLODOC dataset are annotated according to three quality measures; true positives are yellow, false positives are red and false negatives are green.

These images show that some errors present some level of subjectivity, like in the first row, where a parking space on the left is treated by the human annotator as non-road. A different case is presented in the second and third column, where the region near the motorcyclist is treated as non-road by the algorithm even one frame after his departure from the scene. The non-zero HSC1 flows of that area are known also as ghosting effect. An argument could be made that such a result is useful because it warns the driver that the area on the left is not safe, even 1/15th of a second after the crossing of the motorcycle, which could, depending on its speed, be on the left of the ego-vehicle. A speed of 60km/h by both vehicles would mean that their distance is changed by approximately 2.2m per frame, justifying a warning event. In the last example, a consistent false negative error appears on the right, due to the presence of the thick road line. This is a common deficiency of such algorithms and can be dealt with a post-processing phase that includes detection of road lines in the segmentation map.

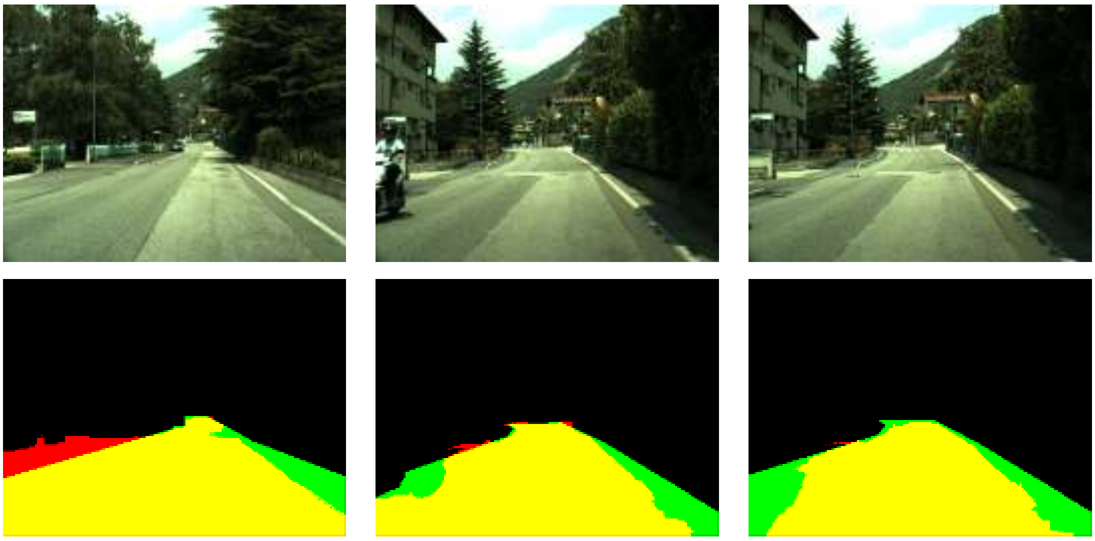


Figure 69 : Qualitative results in frames with low quality rate. Yellow: TP, Green: FN, Red: FP.

7.4.4 Sensitivity analysis of proposed RD system in mixed conditions

The proposed system performs satisfactorily in all the sequences and for all settings used in the experiments, including the parameters a, k for the optical flow, β for the RWA and several reduced image sizes. Extensive experiments were also conducted to assess the influence on the system accuracy from a change in the color space of the frame used in the RWA and the brightness normalization process performed after image acquisition.

More specifically, the experiments carried out on the DIPLODOC sequences concern the following settings using:

1) (a, k) =(3,1), (7,3), (10, 3), (15, 5) for the optical flow estimation,
2) β=70, 90, 110 for the RWA segmentation,
3) the RGB, CIE-L*a*b*, HSV and YCbCr color spaces for the RWA segmentation and
4) resolution of 320x240 (original size), 160x120 (1/2 size), 80x60 (1/4 size), 64x48 (1/5 size) and 40x30 (1/8 size) for the RWA segmentation.

The best experimental results are derived by altering the first three sets of parameters and keeping the spatial resolution at 160x120, where better segmentation accuracies are achieved. In all experiments, the best segmentation accuracy of 93.1%, was achieved using (a, k) = (3, 1) for the optical flow, RGB color space for the RWA, which gave the best results for β=70. The next best results are obtained using the YCbCr color space. The selection of L*a*b* as a color space resulted in significantly worst segmentation accuracy, of approximately 88.6%, affected by the low correctness rate of 90.4%. Similar results were encountered for the HSV color space. Generally, in the experiments with greater values of a, k and β lower segmentation quality is measured. Some of the results described above are summarized in Table 12.

Table 12 : Segmentation rate of the proposed RD system in DIPLODOC sequences for different setups

		$\hat{g}$		***R***		***P***	
Color Space	**(α,k,β)**	**Mean (%)**	**Std (%)**	**Mean (%)**	**Std (%)**	**Mean (%)**	**Std (%)**
RGB	(7,3,70)	92.3	4.4	96.8	3.3	95.2	4.0
	(7,3,110)	92.1	4.7	96.8	3.3	95.0	4.4
	(15,5,70)	92.1	4.6	96.7	3.4	95.1	4.1
	(15,5,110)	91.9	4.7	96.7	3.4	95.0	4.4
CIE-L*a*b*	(7,3,70)	88.6	6.6	97.9	2.9	90.4	7.2
	(7,3,110)	88.5	6.9	98.1	2.8	90.2	7.4
	(15,5,70)	88.3	6.7	97.8	3.1	90.3	7.3
	(15,5,110)	88.2	6.9	97.9	3.0	90.0	7.4
HSV	(7,3,70)	88.6	7.5	98.0	2.9	90.5	8.1
	(7,3,110)	88.7	7.4	98.1	2.8	90.4	8.0
	(15,5,70)	88.4	7.4	98.0	3.0	90.4	8.0
	(15,5,110)	88.4	7.4	97.9	2.9	90.2	8.0
YCbCr	(7,3,70)	92.2	4.5	96.8	3.3	95.2	4.0
	(7,3,110)	92.0	4.6	96.9	3.3	94.9	4.4
	(15,5,70)	92.0	4.6	96.6	3.4	95.1	4.1
	(15,5,110)	91.9	4.6	96.7	3.4	94.9	4.4

Additional experiments were carried out to assess the contribution of each of the seed selection cues to the quality of the road segmentation. More specifically, a run for the entire Diplodoc sequence was performed using the optimum resolution of (160x120), and (a, k, β) = (3, 1, 70) settings, for all possible combinations of: i) the *a-priori* sky region being used or excluded, ii) the road trapezoid being selected based on the result of the previous frame, or being selected by the user in the first frame and remaining stable for all frames and iii) the HSC1 flow being used or not. The experimental results from these experiments led to the following conclusions:

1) the *a-priori* sky region included in the non-road seeds class benefits mainly the segmentation speed of the algorithm, as it can enhance it by almost 10%, when the sky seeds are placed in the top 20% of the frame. In terms of segmentation quality, the sky region does not seem to have any quantitative influence.

2) The segmentation quality is improved when both previous frame road-pixels and HSC1 flow information is processed by the RWA (4% better than static trapezoid). When the RWA derives the road pixels without HSC1 flow data, the static road trapezoid produces a higher correctness rate, influencing the overall quality (both measures are raised by approximately 9%).
3) The use of HSC1 flow information is beneficial in all cases, improving the quality result by a minimum of 4%, affected mainly by the correctness improvement.

A final experiment was carried out using a filled trapezoid for the road seeds instead of its perimeter. In this case, the initialized seeds covered approximately 60 to 70% of the frame (20% for the sky, 30 to 40% for the road and up to 10% for the obstacles), leading to a reduction of the processing time of 10 to 15%. However, this experimental setup is deficient especially in cases where a preceding vehicle is in the trapezoidal area; this has the effect of large portions of the vehicle to be misclassified as road, leading to a deterioration of the correctness.

7.4.5 Quantitative Road Detection results in challenging conditions

The RD system was also assessed under more challenging conditions, using the two video streams Alvarez1 and Alvarez2 provided by Alvarez et al for the purposes of [24] and described in Chapter 1. The two video streams contain driving scenes after rain and in the presence of very dense shadows, two factors that make the RD task quite difficult. The results reported by Alvarez et al were measured in terms of the effectiveness metric, *F*, presented in section 1.7.1, reaching an average value of $\hat{F} = 0.8945 \pm 0.0028$ and an average standard deviation of $\hat{\sigma}_F = 0.1001 \pm 0.0082$. The fluctuations are due to the fact that their method is training-based; therefore the effectiveness was estimated using a cross-validation process.

Using the setup that was described for the DIPLODOC sequence, the proposed RD system achieved directly comparable results, even though it does not include any training phase as the one presented in [24]. More specifically, the proposed RD system reaches an average value of $\hat{F} = 0.8885$, with an average standard deviation of $\hat{\sigma}_F = 0.0084$ when operating with the original image resolution. The aforementioned results were achieved with a very high correctness/precision rate, but a low completeness/recall rate. Another disadvantage derived by using the original resolution is the low processing speed achieved, which is approximately 1.7sec per frame using MATLAB code on a 2.8 GHz Quad Core processor.

When tweaking the setup of our system to balance the precision/recall factor, we can get results that are far superior to the ones reported in [24]. More specifically, if the upper base of the road trapezoid is expanded by a length equal to 20% of the bottom base (10% in each direction), and the probability threshold for the RWA result is dropped to a lower level (e.g. 0.2), the detection efficiency of the proposed RD system rises above 0.9. The first change made, i.e. the expansion of the upper base length, cannot be considered a "tweak", since it is a heuristic variable, based on the camera setup. However, lowering the probability threshold t_p below the value of 0.5 is somewhat counterintuitive, since theoretically for equal *a-priori* probabilities the best RWA segmentation results for two classes' segmentation should be achieved by setting the threshold to the value where the minimum classification error rate is achieved, i.e. by half the maximum probability.

The automated seed selection process of the proposed algorithm significantly influences the optimum threshold value and, hence the better results are achieved for a lower than 0.5 threshold. All results that are reported in this section were obtained using an image hole filling post-processing stage for the thresholded RWA result, so that the proposed method is directly comparable to [24], which also uses such a module.

The final results achieved by three different setups of our system for the original resolution of 640x480, compared to those reported in [24] for their method and the method inspired by [131] and [133]are demonstrated in Table 13.

Table 13 : Overall performance achieved by proposed RD algorithm compared to [24] and [131], [133].

		Method in [24]	**HSI Method inspired by** [131], [133]	**Proposed Method with 6.4.3 setup**	**Proposed Method with longer upper base and t_p = 0.5**	**Proposed Method with longer upper base and t_p = 0.2**
Efficiency, $\hat{F}$	***Mean***	0.8945 ± 0.0028	0.6206 ± 0.0003	0.8885	**0.9070**	**0.9216**
	Std	0.1001 ± 0.0082	0.0343 ± 0.0002	0.0840	**0.0757**	**0.0634**
Quality, $\hat{g}$	***Mean***	N/A	N/A	0.8081	0.8374	0.8603
	Std	N/A	N/A	0.1167	0.1106	0.0963
Recall, *R*	***Mean***	N/A	N/A	0.8162	0.8501	0.9113
	Std	N/A	N/A	0.1220	0.0352	0.0979
Precision, *P*	***Mean***	N/A	N/A	0.9908	0.9859	0.9455
	Std	N/A	N/A	0.0028	0.1171	0.0798

From the results of Table 13, it is evident that the setup used in the proposed system in the previous section is not optimal for every problem. Minor adjustments of the settings can lead to great performance improvements, thus making the most successful setup to outperform the method in [24] by approximately 0.025 in terms of efficiency.

7.4.6 Selection of optimal resolution for the RD system

One of the most important problems in designing road detection applications concerns the selection of the optimal image resolution in the segmentation process. The adopted spatial resolution has a dramatic effect in both the quality and the speed of the system. In very small resolution, the segmentation quality is decreased due to information loss, while using very large images increases the computational complexity, while there is no guarantee of achieving better segmentation results.

Extensive quantitative experiments have been conducted in the DIPLODOC video streams with an original size of 320x240 to study the effects of frames re-sampling to a lower resolution before using them in the RWA. The experimental results for the setup that achieved the best overall quality are given in Table 14, together with the average processing time per frame. The results reported were achieved using MATLAB code on a 2.8 GHz Quad Core processor, Windows 7 OS, without any parallelization. The only part of the code that is not native MATLAB is the Horn Schunck optical flow calculation, which was written in C and compiled as a MATLAB executable file.

The experimental results from the DIPLODOC sequence demonstrate that the best segmentation quality is achieved when the RWA process images half the original size (160x120). The system using this resolution reaches a frame rate of approximately 10fps, giving 101ms processing time per frame. Real-time response at 15 fps is achieved for image resolution of 80x60. In this case the segmentation accuracy is decreased by a 1.5%. Among the most impressive advantages of the proposed RD method are the robustness of the correctness and the completeness rate, which are always greater than 96.5% and 91% respectively.

Table 14: Results achieved for various resolutions of DIPLODOC video streams.

Resolution		**320x240**	**160x120**	**80x60**	**64x48**	**40x30**
Quality, $\hat{g}$	***Mean***	92.5%	**92.8%**	91.3%	90.3%	88.3%
	Std	3.7%	**3.6%**	4.3%	4.5%	5.3%
Recall, *R*	***Mean***	95.5%	**95.7%**	94.3%	93.3%	91.1%
	Std	3.8%	**3.5%**	4.2%	4.5%	5.8%
Precision, *P*	***Mean***	96.8%	**96.9%**	96.7%	96.7%	96.7%
	Std	3.0%	**2.7%**	3.1%	3.2%	3.1%
Processing time per frame (ms)		350	101	62	59	**56**
Frames processed per second (fps)		2.9	9.9	16.0	16.9	**17.8**
Processing time of RWA / HSC1 (ms)		295/4	45/4	11/4	7/4	**3**/4

The RWA module is the computational bottleneck of the proposed system, requiring approximately 1.7s when the original resolution of 640x480 is used. When frame resolutions lower than 320x240 are used, the solution of Equation 2.35 is handled efficiently using Cholesky factorization of the large sparse symmetrical matrix involved and the processing time spent on the RWA gradually stops being problematic for real-time implementations, as we can see in the last row of Table 14. The processing time spent on the optical flow calculation is generally acceptable for the purpose of an ADAS.

7.5 Conclusions

In this chapter, a training-free RD system based on a novel automated version of the RWA utilizing temporal, spatial and *a-priori* information was proposed. The system was extensively tested, both qualitatively and quantitatively, using real on-board video streams shot in various weather, illumination and driving conditions.

The novelty of the proposed system lies on the utilization of shadow-resistant optical flow estimation along *a-priori* spatial information and previous segmentation results to automate the RWA for usage in the RD problem. The system designed showed robustness in many environmental conditions and has the ability to operate in real-time in typical roads, or the presence of lane markings. It also does not require any training, or post-processing module.

The system was evaluated experimentally achieving results superior to other methods tested on the DIPLODOC dataset, [123], [128]. It also proved more efficient than the method presented in [24], compared in the same challenging dataset. Shadows, irregularities of the road surface and vehicles are handled efficiently. The proposed approach can process video signals acquired from different types of video cameras, including low cost ones, at different viewing angles, because all this information is adjusted by calibrating the *a-priori* areas that will be used in the seed placement process.

Chapter 8

Conclusions and Future Work

8.1 Introduction

This thesis has presented and evaluated four computer vision based algorithms that can be used as separate subsystems in a complete ADAS. The problems tackled were traffic signs detection, traffic lights detection, preceding vehicle detection (both in single images and in videos) and road detection. All the methods developed for the purposes of this thesis were published in international conferences and journals, proving their credibility and novelty.

The most special focus of the dissertation was split on preceding vehicles detection and road detection, two of the most popular areas of interest in modern ADAS research. The discussion presented in this chapter will point out the novelties of all the methods developed, pinpoint the issues that are still open and suggest future work on enhancements or evolution of the methods.

8.2 Traffic Sign Recognition

8.2.1 Proposed TSR system

The implementation of a TSR system was the first problem visited during this dissertation. The system was split into two sequential stages; the detection and the classification stage. The detection phase of the system implemented was based on color symmetry information combined with a spatiotemporal technique that reduced the size of the windows to be searched in the video frames, hence improving the processing time performance of the system. The classification stage involved the usage of two empirically selected, cross-correlation based classifiers for comparison of the detected traffic signs with the templates included in a large dataset of the most common signs.

8.2.2 Performance evaluation

The proposed TSD system demonstrated a very high TS detection rate for videos shot in various conditions. The combination of color and radial symmetry proved to be a very good choice for the problem at hand, even for traffic signs that were not circular. The spatiotemporal stage of the algorithm also played an important role in the correctness of the results, rejecting many false positives and focusing the search on more conspicuous areas of the image frames. The detection module provided some very promising results even in cases of partial occlusions, or very badly illuminated scenes, especially due to its usage of the CIE-L*a*b* color space.

The results achieved from the detection stage were fed to a classification module, deploying the two aforementioned classifiers. The cross-correlation classification performed achieves results that are very satisfying given the facts that they use solely visual information and they use almost all the traffic signs of the E.U. However, the classification subsystem is still far from being mature enough for commercial deployment.

Overall, the TSD system proposed in this thesis achieves promising results even in some more challenging situations and could, potentially, be used as a basis for more complicated systems, which probably will have to involve some kind of machine learning approach to improve the classification results.

8.2.3 Real time implementation potential

The proposed system is moderately fast, especially when the input images are of size less than 720x568 pixels (the original resolution of the video streams used). Its image processing stage also has many common parts with the TLD and PVD subsystems, a property that could be exploited in a parallel architecture ADAS to increase the total processing speed of the system without jeopardizing the quality of the results.

8.2.4 Prospects of the proposed system

The TSR system proposed in this thesis is a fairly simplified example of how computer vision can be used to tackle problems that are often met in the ADAS context. While it is promising, it cannot be evenly matched to alternative, more complex solutions in the literature. Furthermore, pure computer vision systems for TSR have been slowly giving their place in the market to either GPS TSR systems, or systems fusing GPS and computer vision. Overall, the fusion of Infrastructure-to-Vehicle (I2V) communication systems with some computer vision method and the GPS signal of the vehicle seems to be a more reliable choice. In this context, the proposed method could be used as a structuring element for such solutions.

8.3 Traffic Lights Recognition

8.3.1 Proposed TLR system

The second subsystem covered in this dissertation is the one performing TLR. Similarly to the TSR module, the TLR uses radial symmetry and color as descriptive features for the generation of candidate regions and then implements a spatiotemporal filtering method to validate the results that belong to TLs. The approach is similar to the one used in TSR, since the basic discriminative cues for both traffic signs and traffic lights are color and shape. In the case of traffic lights, the cues are much stricter since the traffic lights detected must be circular and only have one specific lit color inside them, which can also be used for their classification.

8.3.2 Performance evaluation

The TLR system proposed is successful in terms of completeness, since most of the TLs in the video stream dataset used for evaluation were detected. The problem of the system is its high rate of false positives, which is reduced but not eliminated by the spatiotemporal filtering module used. The issue becomes more evident when the system faces driving in big city roads, especially at nighttime. The biggest advantage of the TLR system is the lack of a training phase or a model describing the TLs. When compared with other methods found in relevant literature, it performs comparably in terms of completeness, but underperforms in terms of correctness. Adverse conditions are also tackled effectively by the system, with a small deterioration of the results (raised number of false positives) in the case of night driving, due to the multitude of light sources.

8.3.3 Real time implementation potential

Our TLR system achieves a near real-time performance, since it achieved a frame rate fluctuating from 2 to 10 fps. The similarities it has with the TSR system could be exploited for a parallel architecture implementation based on a common color filtering process followed by a parallel implementation of the FRST to detect symmetrical objects. Naturally, the processing performance achieved is also directly related with the resolution of the video used and the number of radii that are used in the FRST.

8.3.4 Prospects of the proposed system

Similarly to the case of TSR, the proposed TLR subsystem appears promising for the implementation in real life scenarios, but it still has to be improved in terms of false positive rates. The fusion with additional sources of information from the driving scene, like I2V communications or a GPS signal, could further improve the effectiveness of the system.

8.4 Preceding Vehicle Detection

8.4.1 Proposed PVD system

The PVD system proposed in this thesis has two distinct implementation forms; one subsystem that aims at PVD using a single image and one PVD subsystem for video streams.

The single image method, presented in chapter 5, is based on a mixture of radial and axial symmetry features that define the rear part of vehicles. The radial symmetry is used in conjunction with a color filtering process for detection of red rear lights that are consequently paired to define ROIs where a vehicle could exist. Then, the ROIs are refined by an edge detection algorithm and are checked for axial symmetry to validate the vehicle candidates. The axial symmetry process is based on a combination of two object similarity metrics applied on the two sub-images created when the candidate ROI is split in half along the vertical bisector.

The video-based PVD method, presented in chapter 6, is based on the same initial steps, with three main differences: i) the color filtering process is performed in a different illumination invariant color space, ii) there is no edge detection process involved and iii) the ROI is confined around the area of the rear lights and does not cover the entire rear part of the vehicle. When these steps are completed, the detected ROIs are processed using the results of a Kalman filter, to validate just at most one result as a preceding vehicle.

8.4.2 Performance evaluation

Both versions of the proposed PVD system were evaluated quantitatively using publicly available datasets, providing superior results in comparison to other relevant methods found in the literature. In the case of single image PVD, the results suffered from a high false positive rate, which was minimized in the case of video based PVD, due to the Kalman filtering stage. Both methods were also qualitatively evaluated in more challenging conditions, shown robust accuracy even in cases of rainfall, snow, or night driving.

More specifically, the single image PVD system presented in chapter 5 outperforms similar methods when detection rates are compared. However, it appears to produce a high number of false positives, and also it sometimes fails to verify vehicles that do not have symmetrical rear views. Also, its reliance on the CIE-L*a*b* color space sometimes fails to detect all the vehicle rear lights in the image.

The video based PVD system uses a different color space, derived from a mixture of the channels of the $c_1c_2c_3$ color space. This color space has experimentally been shown to produce better initial detection results. Furthermore, the confinement of the selected ROIs in an area close to the rear lights instead of the whole rear part of the vehicles improves the verification results for vehicles with less symmetry in their rear views. Finally, the inclusion of Kalman filtering for the final verification of the preceding vehicles greatly reduces the false positive rates, without affecting the detection rates of the system.

8.4.3 Real time implementation potential

Both proposed PVD systems are based on the same basic concepts; hence their processing speeds are similar. The bottleneck of the systems is the FRST, except in the case of processing smaller images (160x120) where the bottleneck is the axial symmetry calculation. The frame rate measured in MATLAB almost reached 7 frames per second for the case of small images and without any parallelization. Due to the structure of the proposed algorithm and the nature of the FRST, the overall processing time could be reduced significantly if a parallel architecture was used. The system then could perform in real time (at least at 25 frames per second), even using MATLAB code.

8.4.4 Prospects of the proposed systems

The proposed systems have been thoroughly evaluated even in challenging conditions with very promising results. Especially the video based PVD system can be used as a stand-alone solution, or at least be a part of a more complicated system that uses extra information from radar or lidar systems to provide even better localization results. The system assumes the existence of red symmetrical lights in the rear part of vehicles, so an even stricter legislation when it comes to vehicle manufacturing could improve its robustness. A possible mixture with a road or lane detector can enhance the accuracy of the system even further. Finally, since the system relies greatly on radial symmetry and color filtering, it can be used in conjunction with the previous systems presented (TSR and TLR) to form a complete ADAS solution.

8.5 Road Detection

8.5.1 Proposed RD system

The RD system is the final one proposed in this thesis, but it has been the most successful one in terms of accuracy, speed and robustness.

The first processing step of the RD system is the estimation of the color Horn-Schunck flow of a pair of frames, followed by a thresholding of the result to pinpoint pixels with high flow (which represent non-road objects). Then, these pixels are combined with *a-priori* knowledge about the camera setup and the road pixels detected in the previous frame to automatically initialize a random-walker process. Finally the results of the RWA are then thresholded to locate the road area, which will be used for the RD in the next frame.

8.5.2 Performance evaluation

The RD algorithm proposed in this thesis has been extensively tested and proven to be superior to several of the alternative methods found in recent literature. The assessment of the system was based on two publicly available datasets provided by the researchers whose methods were compared to the one proposed here, therefore the results are as objective as possible.

Extra qualitative assessment of the proposed system was also carried out, using video streams acquired for the purposes of this thesis. The results showed that the system appears robust to particularly challenging scenarios, like heavy rainfall, driving through tunnels and underpasses, driving in presence of dense shadows, or driving at nighttime. Finally, the proposed RD system has a very useful quality; it tends to be adaptive to sudden illumination changes and recover from temporary drops in its detection accuracy.

8.5.3 Real time implementation potential

The proposed RD system is exceptionally fast. It performs in near real-time speeds, fluctuating from approximately 3 to approximately 18 frames per second (using

MATLAB code), depending on the resolution used. It produces its best results at an approximate processing rate of 10 fps, making it a very good candidate for a real time ADAS. A DSP implementation, or even a C++ code optimized for speed, would achieve real time speeds, without jeopardizing the quality of the results.

8.5.4 Prospects of the proposed system

The RD system proposed could, potentially be included in a complete ADAS, to assist with localization of obstacles, or inform the driver for potential dangers. The direct connection of the size of the detected road with impending collisions could prove useful for a collision warning system. More specifically, a potential implementation taking into account the rate of change of the size of drivable road ahead, could inform the driver that an obstacle is approaching, hence extra attention is needed.

The system also can be used in conjunction with a vehicle detection module, to assess the overall dangerousness of detected vehicles in the scene. Possible implementations that could benefit from the RD module include overtaking vehicles detection, preceding vehicle detection, etc. Furthermore, RD results can be used to filter out false positive results in TSR and TLR systems, since the expected positions of traffic lights and signs in a driving scene are constrained and have a close relationship with the position and boundaries of the road.

8.6 Final words and future work

The work carried out and described in this thesis spans through several areas of modern ADAS technology. The usage of information coming only from a monocular on-board color camera for all developed systems incorporates an additional challenge, but adds extra value to the proposed systems, since all of them can be used as the foundations of more complete ADASs which will fuse information from different sources.

In that direction, the future work that can be inspired from this dissertation thesis can potentially focus in fusing several of the proposed systems, or enhancing them with usage of information acquired from additional sources.

A first approach will be to try and merge the three first systems proposed in this thesis into one ADAS that performs TLR, TSR and PVD at the highest possible processing speed. The underlying components of all three subsystems are very compatible, since they are all based on color thresholding (aiming at isolating intense colors) and radial symmetry detection. A complete system that performs these processes in a parallel architecture is a very challenging and rewarding goal.

The aforementioned ADAS can later be enhanced by taking into account information generated from the RD module, leading to an improvement of the detection rates of the individual components, by taking into account spatial information of their results and correlating it to the spatial information of the road area. This might lead to a better overall performance in terms of quality.

A third approach that has great potential, is fusing the results from a complete ADAS like the one just described with the information gathered from a camera pointed at the driver's face. The correlation of the driver's gaze information with potential eminent dangers in the driving scene can prove very effective for early warning in dangerous situations, hence giving the driver a heads-up that could prevent accidents.

Finally, the ultimate goal that is common in all ADAS implementations is to produce systems that are robust in all possible conditions. The systems proposed so far in relevant literature have not been addressing challenging conditions very frequently. Apart from the inherent difficulties to build such a system due to the presence of strong multi-source, non-white and time-variant noise, a reason for this is the lack of benchmarks that can be used to assess the performance of such systems. Therefore, a move towards gathering large datasets of driving videos in challenging, or even adverse conditions could be very valuable to the researchers of this field. Furthermore, the annotation of TLs, TSs, vehicles, pedestrians, other obstacles and the drivable road on a large volume of video frames is a very important goal for vision-based ADAS researchers, so that a common, widely accepted benchmark can be developed.

8.7 Dissertation Publications

The research work described in this dissertation has resulted in the following articles:

8.7.1 Journal Publications

J.1 G. Siogkas and E. Dermatas, "Random Walker Monocular Road Detection in Adverse Conditions using Automated Spatiotemporal Seed Selection," to appear in *Intelligent Transportation Systems, IEEE Transactions on*, vol. 14, no. 2, pp. 527–538, 2013.

J.2 G. Siogkas, E. Skodras, N. Fakotakis, and E. Dermatas, "Rear Lights Vehicle Detection and Tracking," *Integrated Computer Aided Engineering*, invited paper, submitted, pending review, 2013.

8.7.2 Conference Publications

C.1 E. Skodras, G. Siogkas, E. Dermatas, and N. Fakotakis, "Rear lights vehicle detection for collision avoidance," in *Systems, Signals and Image Processing (IWSSIP), 2012 19th International Conference on*, 2012, pp. 134–137.

C.2 G. Siogkas, E. Skodras, and E. Dermatas, "Traffic Lights Detection in Adverse Conditions Using Color, Symmetry and Spatiotemporal Information," in *International Conference on Computer Vision Theory and Applications (VISAPP)*, 2012, pp. 620–627.

C.3 G. K. Siogkas and E. S. Dermatas, "Detection, tracking and classification of road signs in adverse conditions," in *Electrotechnical Conference, 2006. MELECON 2006. IEEE Mediterranean*, 2006, pp. 537–540.

Bibliography

[1] C. Mathers, D. M. Fat, and J. T. Boerma, *The global burden of disease: 2004 update*. World Health Organization, 2008.

[2] “WHO | Global status report on road safety 2009,” *WHO*. [Online]. Available: http://www.who.int/violence_injury_prevention/road_safety_status/2009/en/. [Accessed: 17-Nov-2012].

[3] E. Kopits and M. Cropper, “Traffic fatalities and economic growth,” *Accident Analysis & Prevention*, vol. 37, no. 1, pp. 169–178, 2005.

[4] C. J. L. Murray and A. D. Lopez, “The global burden of disease and injury series, volume 1: a comprehensive assessment of mortality and disability from diseases, injuries, and risk factors in 1990 and projected to 2020,” *Cambridge. MA*, 1996.

[5] U.S. Departmetnt of Transportation, “National Motor Vehicle Crash Causation Survey: Report to Congress,” National Highway Traffic Safety Administration, Jul. 2008.

[6] “Autonomous Emergency Braking - AEB | Euro NCAP - For safer cars crash test safety rating.” [Online]. Available: http://www.euroncap.com/results/aeb.aspx. [Accessed: 17-Nov-2012].

[7] “The INSIGNIA - Safety - Opel Eye.” [Online]. Available: http://www.opelmicrosites.com/insignia/index.html#/safety_opel_eye. [Accessed: 18-Nov-2012].

[8] “Traffic Sign Detection - Mobileye.” [Online]. Available: http://www.mobileye.com/technology/applications/traffic-sign-detection/. [Accessed: 18-Nov-2012].

[9] A. Møgelmose, M. M. Trivedi, and T. B. Moeslund, “Vision-Based Traffic Sign Detection and Analysis for Intelligent Driver Assistance Systems: Perspectives and Survey,” 2012.

[10] R. O. Duda and P. E. Hart, “Use of the Hough transformation to detect lines and curves in pictures,” *Communications of the ACM*, vol. 15, no. 1, pp. 11–15, 1972.

[11] “Vehicle Detection - Mobileye.” [Online]. Available: http://www.mobileye.com/technology/applications/vehicle-detection/. [Accessed: 18-Nov-2012].

[12] Q. Wang, J. Yang, M. Ren, and Y. Zheng, “Driver fatigue detection: a survey,” in *Intelligent Control and Automation, 2006. WCICA 2006. The Sixth World Congress on*, 2006, vol. 2, pp. 8587–8591.

[13] D. Geronimo, A. M. Lopez, A. D. Sappa, and T. Graf, “Survey of pedestrian detection for advanced driver assistance systems,” *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 32, no. 7, pp. 1239–1258, 2010.

[14] M. Enzweiler and D. M. Gavrila, “Monocular pedestrian detection: Survey and experiments,” *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 31, no. 12, pp. 2179–2195, 2009.

[15] P. Dollár, C. Wojek, B. Schiele, and P. Perona, “Pedestrian detection: A benchmark,” in *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, 2009, pp. 304–311.

[16] “Pedestrian-detection & automated platooning,” *CarAdvice.com.au*. [Online]. Available: http://www.caradvice.com.au/67890/2011-brakes-camera-action-pedestrian-detection-automated-platooning/. [Accessed: 18-Nov-2012].

[17] “HowStuffWorks ‘Car Night Vision Systems’,” *HowStuffWorks*. [Online]. Available: http://electronics.howstuffworks.com/gadgets/automotive/in-dash-night-vision-system.htm. [Accessed: 18-Nov-2012].

[18] S. Thrun, M. Montemerlo, H. Dahlkamp, D. Stavens, A. Aron, J. Diebel, P. Fong, J. Gale, M. Halpenny, G. Hoffmann, K. Lau, C. Oakley, M. Palatucci, V. Pratt, P. Stang, S. Strohband, C. Dupont, L.-E. Jendrossek, C. Koelen, C. Markey, C. Rummel, J. van Niekerk, E. Jensen, P. Alessandrini, G. Bradski, B. Davies, S. Ettinger, A. Kaehler, A. Nefian, and P. Mahoney, “Stanley: The Robot That Won the DARPA Grand Challenge,” in *The 2005 DARPA Grand Challenge*, M. Buehler, K. Iagnemma, and S. Singh, Eds. Springer Berlin Heidelberg, 2007, pp. 1–43.

[19] J. Stallkamp, M. Schlipsing, J. Salmen, and C. Igel, “Man vs. computer: Benchmarking machine learning algorithms for traffic sign recognition,” *Neural Networks*, vol. 32, pp. 323–332, Aug. 2012.

[20] X. W. Gao, “A collection of benchmark images for traffic sign research,” in *2011 17th International Conference on Digital Signal Processing (DSP)*, 2011, pp. 1 –6.

[21] R. de Charette and F. Nashashibi, “Traffic light recognition using image processing compared to learning processes,” presented at the IEEE/RSJ International Conference on Intelligent Robots and Systems, 2009. IROS 2009, 2009, pp. 333–338.

[22] Robotics Centre of Mines ParisTech, “Traffic Lights Recognition (TLR) public benchmarks [La Route Automatisée],” 01-May-2010. [Online]. Available: http://www.lara.prd.fr/benchmarks/trafficlightsrecognition. [Accessed: 06-Oct-2011].

[23] “FBK - IT - TeV: Databases,” 27-Sep-2011. [Online]. Available: http://tev.fbk.eu/DATABASES/road.html. [Accessed: 27-Sep-2011].

[24] J. M. Á. Alvarez and A. M. Lopez, “Road Detection Based on Illuminant Invariance,” *IEEE Trans. Intell. Transport. Syst.*, vol. 12, no. 1, pp. 184–193, Mar. 2011.

[25] A. R. T. Gepperth, S. Rebhan, S. Hasler, and J. Fritsch, “Biased competition in visual processing hierarchies: a learning approach using multiple cues,” *Cognitive computation*, vol. 3, no. 1, pp. 146–166, 2011.

[26] A. Gepperth, J. Edelbrunner, and T. Bucher, “Real-time detection and classification of cars in video sequences,” in *Intelligent Vehicles Symposium, 2005. Proceedings. IEEE*, 2005, pp. 625–631.

[27] A. Gepperth, “Technical description of the HRI RoadTraffic dataset.”

[28] A. Gepperth, “Implementation and evaluation of a large-scale object detection system,” 2010.

[29] A. R. Gepperth, “Co-training of context models for real-time vehicle detection,” in *Intelligent Vehicles Symposium (IV), 2012 IEEE*, 2012, pp. 814–820.

[30] A. Gepperth, B. Dittes, and M. G. Ortiz, “The contribution of context information: A case study of object recognition in an intelligent car,” *Neurocomputing*, 2012.

[31] “Computational Vision: Archive.” [Online]. Available: http://www.vision.caltech.edu/html-files/archive.html. [Accessed: 10-Jul-2012].

[32] M. Sridharan and P. Stone, “Color learning and illumination invariance on mobile robots: A survey,” *Robotics and Autonomous Systems*, vol. 57, no. 6, pp. 629–644, 2009.

[33] X. Zou, J. Kittler, and K. Messer, “Illumination invariant face recognition: A survey,” in *Biometrics: Theory, Applications, and Systems, 2007. BTAS 2007. First IEEE International Conference on*, 2007, pp. 1–8.

[34] I. C. on Illumination, *Recommendations on uniform color spaces, color-difference equations, psychometric color terms*. Bureau central de la CIE, 1978.

[35] T. Gevers and W. M. Smeulders, “Color based object recognition,” *Pattern recognition*, vol. 32, no. 3, pp. 453–464, 1999.

[36] G. Loy and A. Zelinsky, “Fast radial symmetry for detecting points of interest,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no. 8, pp. 959 – 973, Aug. 2003.

[37] N. Otsu and others, “A threshold selection method from gray-level histograms,” *IEEE Transactions on systems, Man, and Cybernetics*, vol. 9, no. 1, pp. 62–66, 1979.

[38] S. S. Beauchemin and J. L. Barron, “The computation of optical flow,” *ACM Computing Surveys (CSUR)*, vol. 27, no. 3, pp. 433–466, 1995.

[39] S. Baker, D. Scharstein, J. P. Lewis, S. Roth, M. J. Black, and R. Szeliski, “A database and evaluation methodology for optical flow,” *International journal of computer vision*, pp. 1–31, 2007.

[40] B. K. . Horn and B. G. Schunck, “Determining optical flow,” *Artificial intelligence*, vol. 17, no. 1–3, pp. 185–203, 1981.

[41] C. Liu, “Beyond pixels: exploring new representations and applications for motion analysis,” Massachusetts Institute of Technology, 2009.

[42] T. Brox, A. Bruhn, N. Papenberg, and J. Weickert, “High accuracy optical flow estimation based on a theory for warping,” *Computer Vision-ECCV 2004*, pp. 25–36, 2004.

[43] A. Bruhn, J. Weickert, and C. Schnörr, “Lucas/Kanade meets Horn/Schunck: Combining local and global optic flow methods,” *International Journal of Computer Vision*, vol. 61, no. 3, pp. 211–231, 2005.

[44] R. E. Kalman, “A new approach to linear filtering and prediction problems,” *Journal of basic Engineering*, vol. 82, no. 1, pp. 35–45, 1960.

[45] G. Welch and G. Bishop, “An introduction to the Kalman filter,” 1995.

[46] G. Bishop and G. Welch, “An introduction to the kalman filter,” *Proc of SIGGRAPH, Course*, vol. 8, pp. 27599–3175, 2001.

[47] L. Grady, “Random walks for image segmentation,” *IEEE transactions on pattern analysis and machine intelligence*, pp. 1768–1783, 2006.

[48] S. Kakutani, “Markov processes and the Dirichlet problem,” 1945, vol. 21, pp. 227–233.

[49] P. G. Doyle and J. L. Snell, “Random walks and electric networks,” *Washington D.C.: Mathematical Association of America*, vol. 22, 1984.

[50] J. P. Lewis, “Industrial Light & Magic,” *Fast Normalized Cross-Correlation*, vol. 2011, 1995.

[51] R. M. Haralock and L. G. Shapiro, *Computer and robot vision*. Addison-Wesley Longman Publishing Co., Inc., 1991.

[52] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, “Image quality assessment: from error visibility to structural similarity,” *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600 –612, Apr. 2004.

[53] G. K. Siogkas and E. S. Dermatas, “Detection, tracking and classification of road signs in adverse conditions,” in *Electrotechnical Conference, 2006. MELECON 2006. IEEE Mediterranean*, 2006, pp. 537–540.

[54] S. Escalera and P. Radeva, “Fast greyscale road sign model matching and recognition,” *Recent Advances in Artificial Intelligence Research and Development*, pp. 69–76, 2004.

[55] G. Loy and N. Barnes, “Fast shape-based road sign detection for a driver assistance system,” presented at the 2004 IEEE/RSJ International Conference on Intelligent Robots and Systems, 2004. (IROS 2004). Proceedings, 2004, vol. 1, pp. 70– 75 vol.1.

[56] X. Baró and J. Vitrià, “Fast traffic sign detection on greyscale images,” *Recent Advances in Artificial Intelligence Research and Development*, pp. 69–76, 2004.

[57] D. M. Gavrila and V. Philomin, “Real-time object detection for ‘smart’ vehicles,” in *Computer Vision, 1999. The Proceedings of the Seventh IEEE International Conference on*, 1999, vol. 1, pp. 87–93.

[58] H. Austermeier, U. Büker, B. Mertsching, and S. Zimmermann, “Analysis of Traffic Scenes by Using the Hierarchical Structure Code,” *Advances in Structural and Syntactic Pattern Recognition. Series on Machine Perception and Artificial Intelligence*, vol. 5, pp. 561–570, 1992.

[59] B. Besserer, S. Estable, B. Ulmer, and D. Reichqardt, “Shape classification for traffic sign recognition,” in *First International Workshop on Intelligent Autonomous Vehicles, IFAC April*, 1993.

[60] U. Büker and B. Mertsching, “Parallel evaluation of hierarchical image databases,” *Journal of Parallel and Distributed Computing*, vol. 31, no. 2, pp. 141–152, 1995.

[61] Y. Aoyagi and T. Asakura, “A study on traffic sign recognition in scene image using genetic algorithms and neural networks,” in *Industrial Electronics, Control, and Instrumentation, 1996., Proceedings of the 1996 IEEE IECON 22nd International Conference on*, 1996, vol. 3, pp. 1838–1843.

[62] R. Dahyot, P. Charbonnier, and F. Heitz, “Unsupervised statistical detection of changing objects in camera-in-motion video,” in *Image Processing, 2001. Proceedings. 2001 International Conference on*, 2001, vol. 1, pp. 638–641.

[63] C. Bahlmann, Y. Zhu, V. Ramesh, M. Pellkofer, and T. Koehler, “A system for traffic sign detection, tracking, and recognition using color, shape, and motion information,” in *Intelligent Vehicles Symposium, 2005. Proceedings. IEEE*, 2005, pp. 255–260.

[64] A. de la Escalera, L. E. Moreno, M. A. Salichs, and J. M. Armingol, “Road traffic sign detection and classification,” *IEEE Transactions on Industrial Electronics*, vol. 44, no. 6, pp. 848–859, Dec. 1997.

[65] G. Piccioli, E. De Micheli, P. Parodi, and M. Campani, “Robust method for road sign detection and recognition,” *Image and Vision Computing*, vol. 14, no. 3, pp. 209–223, 1996.

[66] C. Y. Fang, C. S. Fuh, P. S. Yen, S. Cherng, and S. W. Chen, “An automatic road sign recognition system based on a computational model of human recognition processing,” *Computer Vision and Image Understanding*, vol. 96, no. 2, pp. 237–268, 2004.

[67] S. H. Hsu and C. L. Huang, “Road sign detection and recognition using matching pursuit method,” *Image and Vision Computing*, vol. 19, no. 3, pp. 119–129, 2001.

[68] A. De La Escalera, J. M. Armingol, and M. A. Salichs, "Traffic sign detection for driver support systems," in *International conference on field and service robotics*, 2001.

[69] A. de la Escalera, J. M. Armingol, and M. Mata, "Traffic sign recognition and analysis for intelligent vehicles," *Image and vision computing*, vol. 21, no. 3, pp. 247–258, 2003.

[70] A. de la Escalera, J. M. Armingol, J. M. Pastor, and F. J. Rodriguez, "Visual sign information extraction and identification by deformable models for intelligent vehicles," *Intelligent Transportation Systems, IEEE Transactions on*, vol. 5, no. 2, pp. 57–68, 2004.

[71] Chiung-Yao Fang, Sei-Wang Chen, and Chiou-Shann Fuh, "Road-sign detection and tracking," *IEEE Transactions on Vehicular Technology*, vol. 52, no. 5, pp. 1329– 1341, Sep. 2003.

[72] D. S. Kang, N. C. Griswold, and N. Kehtarnavaz, "An invariant traffic sign recognition system based on sequential color processing and geometrical transformation," in *Image Analysis and Interpretation, 1994., Proceedings of the IEEE Southwest Symposium on*, 1994, pp. 88–93.

[73] G. Siogkas, E. Skodras, and E. Dermatas, "Traffic lights detection in adverse conditions using color, symmetry and spatiotemporal information," presented at the International Conference on Computer Vision Theory and Applications (VISAPP 2012), Rome, Italy, Rome, Italy, 2012.

[74] U. Franke, D. Gavrila, S. Gorzig, F. Lindner, F. Puetzold, and C. Wohler, "Autonomous driving goes downtown," *Intelligent Systems and Their Applications, IEEE*, vol. 13, no. 6, pp. 40–48, 1998.

[75] F. Lindner, U. Kressel, and S. Kaelberer, "Robust recognition of traffic signals," presented at the 2004 IEEE Intelligent Vehicles Symposium, 2004, pp. 49–53.

[76] J. In-Hak, C. Seong-Ik, and H. Tae-Hyun, "Detection of Traffic Lights for Vision-Based Car," *Traffic*, vol. 4319/2006, pp. 682–691, 2006.

[77] Y. Shen, U. Ozguner, K. Redmill, and J. Liu, "A robust video based traffic light detection algorithm for intelligent vehicles," pp. 521–526.

[78] M. Omachi and S. Omachi, "Traffic light detection with color and edge information," in *Computer Science and Information Technology, 2009. ICCSIT 2009. 2nd IEEE International Conference on*, 2009, pp. 284–287.

[79] R. de Charette and F. Nashashibi, "Real time visual traffic lights recognition based on Spot Light Detection and adaptive traffic lights templates," presented at the 2009 IEEE Intelligent Vehicles Symposium, 2009, pp. 358–363.

[80] Chunhe Yu, Chuan Huang, and Yao Lang, “Traffic light detection during day and night conditions by a camera,” presented at the 2010 IEEE 10th International Conference on Signal Processing (ICSP), 2010, pp. 821–824.

[81] N. Barnes and A. Zelinsky, “Real-time radial symmetry for speed sign detection,” in *Intelligent Vehicles Symposium, 2004 IEEE*, 2004, pp. 566–571.

[82] N. Barnes, A. Zelinsky, and L. S. Fletcher, “Real-time speed sign detection using the radial symmetry detector,” *Intelligent Transportation Systems, IEEE Transactions on*, vol. 9, no. 2, pp. 322–332, 2008.

[83] P. Soille, *Morphological image analysis : principles and applications with 12 tables*. Berlin [u.a.]: Springer, 1999.

[84] P. Moghadam, J. Starzyk, and W. Wijesoma, “Fast Vanishing Point Detection in Unstructured Environments,” *IEEE Transactions on Image Processing*, vol. PP, no. 99, pp. 1–1, 0.

[85] M. Nieto Doncel, “Detection and tracking of vanishing points in dynamic environments: Detección y seguimiento de puntos de fuga en entornos dinámicos,” Telecomunicacion, 2010.

[86] E. Skodras, G. Siogkas, E. Dermatas, and N. Fakotakis, “Rear lights vehicle detection for collision avoidance,” in *Systems, Signals and Image Processing (IWSSIP), 2012 19th International Conference on*, 2012, pp. 134–137.

[87] Z. Sun, G. Bebis, and R. Miller, “On-road vehicle detection: a review,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 5, pp. 694 – 711, May 2006.

[88] S. Sivaraman and M. Trivedi, “Active learning for on-road vehicle detection: a comparative study,” *Machine Vision and Applications*, pp. 1–13.

[89] Federal Motor Carrier Safety Administration, *Lamps and reflective devices*. .

[90] ECE Regulation 48, *Installation of lighting and light-signalling devices on motor vehicles*. 1958.

[91] Yu-Chen Lin, Che-Chung Lin, Long-Tai Chen, and Ching-Kun Chen, “Adaptive IPM-based lane filtering for night forward vehicle detection,” presented at the 2011 6th IEEE Conference on Industrial Electronics and Applications (ICIEA), 2011, pp. 1568–1573.

[92] L. Gao, C. Li, T. Fang, and Z. Xiong, “Vehicle detection based on color and edge information,” *Image Analysis and Recognition*, pp. 142–150, 2008.

[93] M. Betke, E. Haritaoglu, and L. S. Davis, “Real-time multiple vehicle detection and tracking from a moving vehicle,” *Machine Vision and Applications*, vol. 12, no. 2, pp. 69–83, 2000.

[94] P. Thammakaroon and P. Tangamchit, “Predictive brake warning at night using taillight characteristic,” presented at the IEEE International Symposium on Industrial Electronics, 2009. ISIE 2009, 2009, pp. 217–221.

[95] S. Nagumo, H. Hasegawa, and N. Okamoto, “Extraction of forward vehicles by front-mounted camera using brightness information,” in *Canadian Conference on Electrical and Computer Engineering, 2003. IEEE CCECE 2003*, 2003, vol. 2, pp. 1243 – 1246 vol.2.

[96] R. O’Malley, E. Jones, and M. Glavin, “Rear-Lamp Vehicle Detection and Tracking in Low-Exposure Color Video for Night Conditions,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 11, no. 2, pp. 453–462, Jun. 2010.

[97] I. Cabani, G. Toulminet, and A. Bensrhair, “Color-based detection of vehicle lights,” presented at the IEEE Intelligent Vehicles Symposium, 2005. Proceedings, 2005, pp. 278– 283.

[98] C.-C. R. Wang and J.-J. J. Lien, “Automatic Vehicle Detection Using Local Features-A Statistical Approach,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 9, no. 1, pp. 83 –96, Mar. 2008.

[99] G. Siogkas, E. Skodras, N. Fakotakis, and E. Dermatas, “Rear Lights Vehicle Detection and Tracking,” *ICAE*, vol. submitted, no. pending review, 2013.

[100] R. O’Malley, M. Glavin, and E. Jones, “Vision-based detection and tracking of vehicles to the rear with perspective correction in low-light conditions,” *IET Intelligent Transport Systems*, vol. 5, no. 1, pp. 1 –10, Mar. 2011.

[101] Y.-M. Chan, S.-S. Huang, L.-C. Fu, P.-Y. Hsiao, and M.-F. Lo, “Vehicle detection and tracking under various lighting conditions using a particle filter,” *IET Intelligent Transport Systems*, vol. 6, no. 1, pp. 1 –8, Mar. 2012.

[102] S. Sivaraman and M. M. Trivedi, “A General Active-Learning Framework for On-Road Vehicle Recognition and Tracking,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 11, no. 2, pp. 267 –276, Jun. 2010.

[103] J. Arróspide, L. Salgado, and M. Nieto, “Video analysis-based vehicle detection and tracking using an MCMC sampling framework,” *EURASIP J. Adv. Signal Process.*, vol. 2012, no. 1, pp. 1–20, Dec. 2012.

[104] K. She, G. Bebis, H. Gu, and R. Miller, “Vehicle tracking using on-line fusion of color and shape features,” in *The 7th International IEEE Conference on Intelligent Transportation Systems, 2004. Proceedings*, 2004, pp. 731 – 736.

[105] A. Jazayeri, Hongyuan Cai, Jiang Yu Zheng, and M. Tuceryan, “Vehicle Detection and Tracking in Car Video Based on Motion Model,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 12, no. 2, pp. 583–595, Jun. 2011.

[106] M. Kohler, “Using the Kalman Filter to track Human Interactive Motion – Modelling and Initialization of the Kalman Filter for Translational Motion,” 1997.

[107] N. Dalal and B. Triggs, “Histograms of oriented gradients for human detection,” in *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, 2005, vol. 1, pp. 886–893.

[108] R. Fergus, P. Perona, and A. Zisserman, “Object class recognition by unsupervised scale-invariant learning,” in *Computer Vision and Pattern Recognition, 2003. Proceedings. 2003 IEEE Computer Society Conference on*, 2003, vol. 2, pp. II–264.

[109] G. K. Siogkas and E. S. Dermatas, “Random-Walker Monocular Road Detection in Adverse Conditions Using Automated Spatiotemporal Seed Selection,” IEEE Transactions on Intelligent Transportation Systems, vol. 14, no. 2, pp. 527–538, 2013.

[110] A. Broggi, “Robust real-time lane and road detection in critical shadow conditions,” 1995, pp. 353–358.

[111] J. Calderon, A. Obando, and D. Jaimes, “Road Detection Algorithm for an Autonomous UGV based on Monocular Vision,” in *Electronics, Robotics and Automotive Mechanics Conference, 2007. CERMA 2007*, 2007, pp. 253–259.

[112] Y. Feng, W. Rong-ben, and Z. Rong-hui, “Research on Road Recognition Algorithm Based on structure environment for ITS,” in *Computing, Communication, Control, and Management, 2008. CCCM’08. ISECS International Colloquium on*, 2008, vol. 1, pp. 84–87.

[113] A. López, J. Serrat, C. Ca\ nero, F. Lumbreras, and T. Graf, “Robust lane markings detection and road geometry computation,” *International Journal of Automotive Technology*, vol. 11, no. 3, pp. 395–407, 2010.

[114] J. D. Crisman and C. E. Thorpe, “UNSCARF-a color vision system for the detection of unstructured roads,” in *, 1991 IEEE International Conference on Robotics and Automation, 1991. Proceedings*, 1991, pp. 2496 –2501 vol.3.

[115] J. D. Crisman and C. E. Thorpe, “SCARF: a color vision system that tracks roads and intersections,” *IEEE Transactions on Robotics and Automation*, vol. 9, no. 1, pp. 49 –58, 1993.

[116] S. Beucher and M. Bilodeau, “Road segmentation and obstacle detection by a fast watershed transformation,” in *Intelligent Vehicles’ 94 Symposium, Proceedings of the*, 1994, pp. 296–301.

[117] M.-Y. Chern, “Knowledge-based region classification for rural road area in the color scene image,” in *2004 IEEE International Conference on Networking, Sensing and Control*, 2004, vol. 2, pp. 891 – 896 Vol.2.

[118] Y. He, H. Wang, and B. Zhang, “Color-Based Road Detection in Urban Traffic Scenes,” *IEEE Trans. Intell. Transport. Syst.*, vol. 5, no. 4, pp. 309–318, Dec. 2004.

[119] A. Routray and K. B. Mohanty, “A Fast Edge Detection Algorithm for Road Boundary Extraction Under Non-uniform Light Condition,” in *Information Technology,(ICIT 2007). 10th International Conference on*, 2007, pp. 38–40.

[120] J. Huang, B. Kong, B. Li, and F. Zheng, “A new method of unstructured road detection based on HSV color space and road features,” in *Information Acquisition, 2007. ICIA'07. International Conference on*, 2007, pp. 596–601.

[121] A. Giachetti, M. Campani, and V. Torre, “The use of optical flow for road navigation,” *IEEE Transactions on Robotics and Automation*, vol. 14, no. 1, pp. 34 – 48, 1998.

[122] R. Alix, F. Le Coat, and D. Aubert, “Flat world homography for non-flat world on-road obstacle detection,” in *IEEE Intelligent Vehicles Symposium, 2003. Proceedings*, 2003, pp. 310 – 315.

[123] P. Lombardi, M. Zanin, and S. Messelodi, “Unified stereovision for ground, road, and obstacle detection,” in *IEEE Intelligent Vehicles Symposium, 2005. Proceedings*, 2005, pp. 783 – 788.

[124] A. Seki and M. Okutomi, “Robust Obstacle Detection in General Road Environment Based on Road Extraction and Pose Estimation,” in *2006 IEEE Intelligent Vehicles Symposium*, 2006, pp. 437 –444.

[125] N. Soquet, D. Aubert, and N. Hautiere, “Road Segmentation Supervised by an Extended V-Disparity Algorithm for Autonomous Navigation,” in *2007 IEEE Intelligent Vehicles Symposium*, 2007, pp. 160 –165.

[126] F. Dornaika, J. M. Alvarez, A. D. Sappa, and A. M. Lopez, “A New Framework for Stereo Sensor Pose Through Road Segmentation and Registration,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 12, no. 4, pp. 954 – 966, Dec. 2011.

[127] C. Guo, S. Mita, and D. McAllester, “Robust Road Detection and Tracking in Challenging Scenarios Based on Markov Random Fields With Unsupervised Learning,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 13, no. 3, pp. 1338 –1354, Sep. 2012.

[128] P. Lombardi, M. Zanin, and S. Messelodi, “Switching models for vision-based on-board road detection,” in *2005 IEEE Intelligent Transportation Systems, 2005. Proceedings*, 2005, pp. 67 – 72.

[129] R. Aufrère, R. Chapuis, and F. Chausse, “A dynamic vision algorithm to locate a vehicle on a nonstructured road,” *The International Journal of Robotics Research*, vol. 19, no. 5, pp. 411–423, 2000.

[130] R. Aufrere, F. Marmoiton, R. Chapuis, F. Collange, and J. P. Derutin, “Road detection and vehicle tracking by vision for adaptive cruise control,” *The International Journal of Robotics Research*, vol. 20, no. 4, pp. 267–286, 2001.

[131] C. Rotaru, T. Graf, and J. Zhang, “Color image segmentation in HSI space for automotive applications,” *J Real-Time Image Proc*, vol. 3, no. 4, pp. 311–322, Mar. 2008.

[132] T. T. Son, S. Mita, and A. Takeuchi, “Road detection using segmentation by weighted aggregation based on visual information and a posteriori probability of road regions,” in *IEEE International Conference on Systems, Man and Cybernetics, 2008. SMC 2008*, 2008, pp. 3018 –3025.

[133] M. A. Sotelo, F. J. Rodriguez, L. Magdalena, L. M. Bergasa, and L. Boquete, “A color vision-based lane tracking system for autonomous driving on unmarked roads,” *Autonomous Robots*, vol. 16, no. 1, pp. 95–116, 2004.