



UNIVERSITÀ DEGLI STUDI DI PARMA
DIPARTIMENTO DI INGEGNERIA DELL'INFORMAZIONE

Dottorato di Ricerca in Tecnologie dell'Informazione
XX Ciclo

Rean Isabella Fedriga

**APPLICAZIONE DI TECNICHE DI VISIONE
ARTIFICIALE IN DIVERSI DOMINI
SPETTRALI**

DISSERTAZIONE PRESENTATA PER IL CONSEGUIMENTO
DEL TITOLO DI DOTTORE DI RICERCA

GENNAIO 2008

*To mamitta and pa
and to all the ships that pass in the night*

Indice

Premessa	1
Introduzione	9
I Individuazione di pedoni in attraversamento su passaggi pedonali	11
1 Introduzione	13
2 Considerazioni preliminari	15
3 Procedimento	19
3.1 In breve	19
3.2 In dettaglio	20
3.2.1 Rimozione della distorsione e della prospettiva	20
3.2.2 Differenza stereo e <i>background subtraction</i>	23
3.2.3 Estrazione degli oggetti	23
3.2.4 Classificazione di pedoni e veicoli	26
3.2.5 Impiego della correlazione temporale	27
4 Risultati sperimentali	31

II Localizzazione di pedoni in ambientazioni notturne da un veicolo autonomo in movimento	35
5 Introduzione	37
6 Considerazioni preliminari	41
7 Procedimento	47
7.1 In breve	47
7.2 In dettaglio	49
7.2.1 Elaborazione preliminare dell'immagine	49
7.2.2 Valutazione dei bounding box	49
7.2.3 Confronto con modelli	58
8 Risultati sperimentali	65
 III Individuazione di oggetti in galleggiamento sulla rotta di imbarcazioni autonome	 71
9 Introduzione	73
10 Considerazioni preliminari	75
11 Procedimento	81
11.1 In breve	81
11.2 In dettaglio	82
11.2.1 Calibrazione della telecamera	82
11.2.2 Ricerca del mare	83
11.2.3 Ricerca della linea di orizzonte	85
11.2.4 Ricerca degli ostacoli	85
11.2.5 Tracking degli ostacoli	87
11.2.6 Invio dei risultati via rete	89

INDICE	v
12 Risultati sperimentali	91
Conclusioni	95
Ringraziamenti	97
Bibliografia	98

Premessa

Tecniche di visione: movimento, forma, colore

Quando osserviamo il mondo esterno, l'evento primario è rappresentato dalla messa a fuoco dei raggi di luce in corrispondenza dei recettori - coni e bastoncelli - collocati nella retina di ciascun occhio. I recettori, colpiti dalla luce, emettono segnali elettrici diretti alla retina ed al cervello, che ne estraggono le informazioni utili dal punto di vista biologico. Il risultato è la nostra percezione della scena che stiamo osservando, in tutti i suoi dettagli di forma, profondità, movimento, colore e struttura.

L'occhio è stato spesso paragonato ad una macchina fotografica, ma sarebbe forse più appropriato il paragone con una telecamera, fissata ad un cavalletto, che insegue automaticamente le immagini: un sistema che si metta a fuoco da solo, si regoli automaticamente per l'intensità luminosa, abbia un obiettivo autopulente e sia collegato ad un computer di eccezionali capacità di analisi parallela. Il fenomeno della visione, quella capacità di catturare la luce che colpisce le retine e trasformarla in una scena visiva che abbia senso, è estremamente complessa ed ancora oscura, per molti aspetti. La retina, trasformando la luce in segnali nervosi, dà il via al processo di estrazione dall'ambiente di ciò che possiede un contenuto informativo utile dal punto di vista biologico, sopprimendo ciò che invece è ridondante. Quello che si ottiene come output dell'occhio dopo questi processi cerebrali, un'informazione molto più sofisti-

cata della rappresentazione puntiforme del mondo esistente a livello di coni e bastoncelli.

Nella corteccia visiva primaria le informazioni passano attraverso diversi stadi, per ognuno dei quali esistono cellule dedicate alla percezione di informazioni molto dettagliate, come il movimento. Per tutti gli animali, esseri umani compresi, i cambiamenti che si verificano nell'ambiente esterno sono molto più importanti rispetto a ciò che rimane stazionario. Tuttavia, per la visione umana, il movimento è anche una modalità di esplorazione di scenari statici. Nell'esplorazione visiva dell'ambiente, gli occhi scattano da un punto di interesse all'altro: non è possibile esplorare una scena statica facendo scorrere su di essa gli occhi con movimenti continui. Il sistema visivo si occupa di tenere fissa sulla retina l'immagine che si sta guardando, impedendole di svanire. Se la scena si muove, gli occhi la seguono fissando un punto e mantenendo un contatto visivo con esso muovendo gli occhi finché l'oggetto inquadrato esce dal campo visivo; a questo punto viene scelto un altro oggetto [1].



Figura 1: *Dinamismo di un cane al guinzaglio*, Giacomo Balla (1912): il concetto di movimento in pittura.

Accanto alle cellule della corteccia visiva primaria dedicate alla percezione del movimento, ne esistono altre specializzate nel rilevamento delle forme degli oggetti attraverso l'individuazione dei loro contorni. Sembra che i contorni luce-buio siano la componente più importante della percezione umana. L'ombreggiatura degli oggetti, costituita da passaggi gradualmente luce-buio, e l'aspetto della loro superficie, danno importanti indicazioni sulla loro forma e spessore [2].



Figura 2: *Cattura di Cristo*, Caravaggio (1602): la definizione delle forme come contrasto tra luce e buio.

Infine, la percezione del colore. Anche se nella specie umana, rispetto ad altri animali e vegetali, la pressione evolutiva a mantenere o a migliorare la visione del colore sembra non essere così intensa, il colore rimane uno degli aspetti più attraenti di ciò che vediamo. Tuttavia, ai fini dell'estrazione delle informazioni contenute in un'immagine da parte della retina, la scala cromatica non è di primaria importanza. Una scena che contenga i colori della luce spettrale o solo

toni di una scala monocromatica (bianco e nero), perderà la sua componente di piacevolezza visiva, ma manterrà per il cervello lo stesso contenuto informativo. Un po' come succede per i film in bianco e nero: pur non essendo a colori, non perdono il loro significato ai nostri occhi di spettatori.

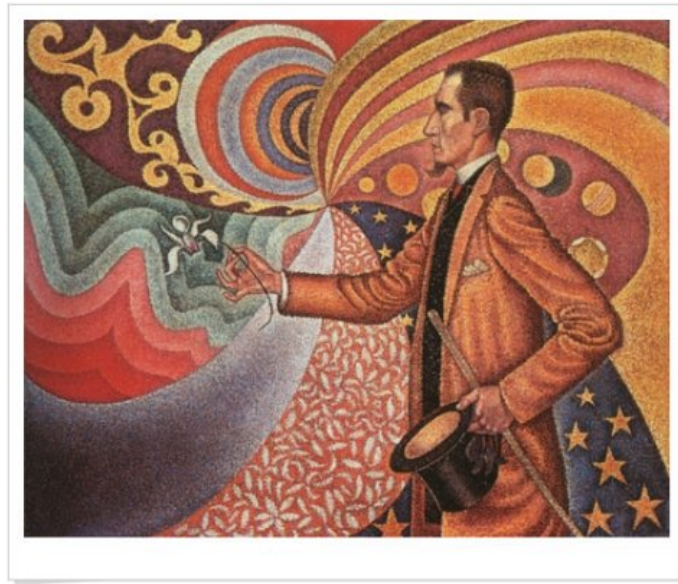


Figura 3: *Portrait de M. Félix Fénéon*, Paul Signac (1890): la tecnica del puntinismo per lo studio del colore.

In visione artificiale sono state create molte tecniche che traggono ispirazione dalle caratteristiche percettive della visione umana. Tra queste, molte sono indirizzate alla rilevazione degli oggetti in movimento - tecniche *motion-based* [3, 4, 5, 6] -, al riconoscimento di oggetti attraverso la forma - tecniche *shape-based* [4, 6, 7] - ed all'analisi di caratteristiche delle superfici, come il loro colore - tecniche *texture-based* [8].

Domini spettrali: visibile, NIR e FIR

Lo spettro visibile - anche chiamato spettro ottico - è la porzione dello spettro elettromagnetico che può essere percepita dall'occhio umano. La radiazione elettromagnetica in questo intervallo di lunghezze d'onda viene chiamata luce visibile o, semplicemente, luce. Un occhio umano normale reagisce a lunghezze d'onda nell'aria tra i 380 ed i 750 nm. Le lunghezze d'onda corrispondenti in acqua o altri mezzi sono ridotte di un fattore uguale all'indice refrattivo del mezzo stesso. In termini di frequenza questo corrisponde ad una banda intorno tra i 400 ed i 790 THz. Un occhio adattato alla luce possiede generalmente la sua massima sensibilità intorno ai 555 nm (540 THz), nella regione verde dello spettro ottico. Lo spettro non contiene però tutti i colori che gli occhi ed il cervello umano possono distinguere. Marrone, rosa e magenta sono assenti, per esempio, perchè necessitano della combinazione di più lunghezze d'onda, preferibilmente sfumature del rosso.

Il vicino infrarosso (Near InfraRed - NIR) è la regione di lunghezze d'onda più vicina alla radiazione percepibile dall'occhio umano. Il confine tra luce visibile ed infrarossa non è definito con precisione. L'occhio umano è marcatamente meno sensibile alla luce sopra i 700 nm di lunghezza d'onda, perciò frequenze più corte non danno contributi significativi a scene illuminate da sorgenti di luce comuni. Luci molto intense invece - come quelle emesse dai laser per esempio - possono essere rilevate fino a circa 780 nm e vengono percepite come luce rossa. L'inizio della regione infrarossa è definito (stando a diversi standard) da vari valori compresi tra 700 e 800 nm.

Il NIR è definito dalla capacità di assorbimento dell'acqua e viene anche indicato come infrarosso *riflesso*.

Il lontano infrarosso (Far InfraRed - FIR) si colloca nella regione di lunghezze d'onda tra i 15 ed i 1,000µm. Questa è la regione in cui i sensori possono

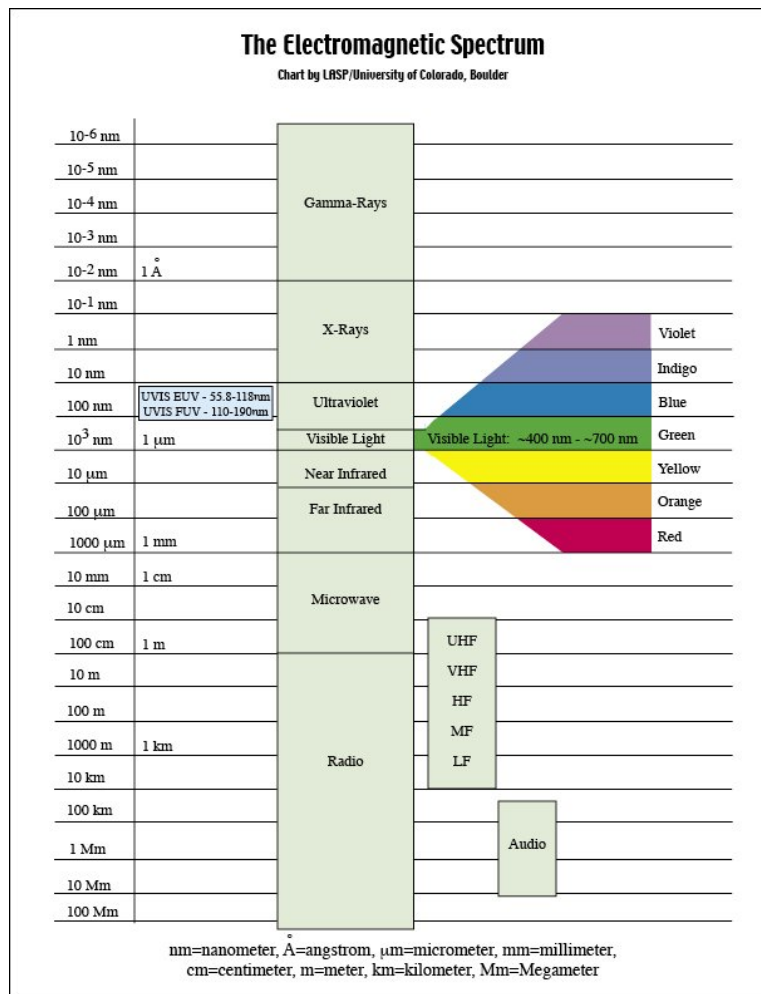


Figura 4: Spettro delle radiazioni elettromagnetiche

percepire una rappresentazione completamente passiva del mondo basata sulle sole emissioni termiche, senza bisogno di sorgenti di luce o calore esterne come il sole, la luna o illuminatori termici. Il FIR è anche noto come infrarosso *termico*.

La radiazione infrarossa viene usata in apparecchi di visione notturna, quando la luce visibile non basta. I sensori infrarossi convertono le radiazioni percepite in un'immagine che può essere monocromatica - gli oggetti più caldi risulteranno più chiari - oppure può essere usato un sistema di falsi colori per rappresentare le diverse temperature.

Lo spettro e la quantità di radiazione termica emesse da un oggetto dipendono fortemente dalla superficie dell'oggetto. Questo fenomeno rende possibile la visualizzazione della temperatura di un oggetto con una telecamera termica. Ci sono però altri fattori che influenzano la radiazione rilevabile di un oggetto, come la sua emissività e la radiazione prodotta dai suoi dintorni, anche questa riflessa dall'oggetto. Tutti gli oggetti ad una temperatura superiore allo zero assoluto emettono radiazioni nel campo dell'infrarosso.

Introduzione

*Alice once remarked, "I see nobody on the road"
and the King replied in a fretful tone
"I only wish I had such eyes, to be able to see Nobody
and at that distance too".
Lewis Carrol, *Through the Looking-Glass**

L'applicazione di tecniche di visione artificiale in diversi domini spettrali è argomento molto trattato in letteratura. L'approccio che abbiamo deciso di seguire riguarda l'applicazione di alcune tra tali tecniche in situazioni poco affrontate dai lavori precedenti e con uno sguardo rivolto a quanto avviene nel sistema visivo umano, cercando di sopperire alle carenze valutative tipiche di un sistema artificiale che tenta di emularne uno naturale, anche con l'impiego di sensori che forniscono immagini dal contenuto informativo diverso rispetto a quello che caratterizza le immagini percepibili dall'occhio.

Sono state effettuate tre sperimentazioni, impiegando sensori operanti nelle tre regioni dello spettro elettromagnetico che risultano abitualmente utilizzate nell'ambito della visione artificiale: visibile, NIR e FIR. Obiettivo comune delle sperimentazioni è stato individuare oggetti, anche se di natura diversa, nelle immagini acquisite: pedoni e veicoli, nel caso della prima indagine; solo pedoni, nella seconda, ed oggetti generici, nella terza. Per ogni sperimenta-

zione, le ricerche sono state condotte concentrando l'attenzione sull'aspetto che, dall'osservazione delle immagini, abbiamo ritenuto più significativo per la discriminazione degli oggetti all'interno delle immagini stesse da parte del nostro sistema visivo di osservatori umani. Così, ci è sembrato che il movimento potesse essere l'aspetto distintivo degli oggetti nelle immagini acquisite per la prima sperimentazione, mentre la forma è stata ritenuta la caratteristica più idonea alla trattazione della seconda, e per la terza, il colore degli oggetti rispetto a quello di sfondo.

L'esposizione è suddivisa in tre parti, ognuna dedicata ad una delle sperimentazioni, ed in essa vengono presentate le considerazioni preliminari avvenute, i dettagli implementativi ed i risultati ottenuti.

Al termine dell'esposizione, alcune considerazioni conclusive.

Parte I

Individuazione di pedoni in attraversamento su passaggi pedonali

Capitolo 1

Introduzione

Il sistema sviluppato è un passaggio pedonale intelligente in grado di ricavare informazioni sullo stato di occupazione delle strisce grazie alla visione artificiale.

Le statistiche riportano che una percentuale significativa degli incidenti che coinvolgono pedoni avviene in corrispondenza degli incroci segnalati [9].

Il rilevamento di pedoni presso i passaggi pedonali può contribuire all'aumento della sicurezza per i pedoni che necessitano di più tempo per attraversare la strada, come persone disabili o anziane, prolungando il tempo di permanenza del verde sul semaforo pedonale, quando necessario, ed inviando avvertimenti agli automobilisti. Allo stesso tempo comporterebbe una riduzione dei rallentamenti del traffico, poiché l'intervallo di tempo con semaforo verde potrebbe essere ridotto qualora l'attraversamento dei pedoni finisse prima della scomparsa del segnale di "via libera" sul semaforo pedonale, permettendo di evitare inutili attese per i veicoli.

Le ricerche precedenti relative al controllo semaforico erano principalmente focalizzate sul monitoraggio dei veicoli, mentre scarsa attenzione è stata riservata ai pedoni. La maggior parte dei sistemi sviluppati per il rilevamento di pedoni agli incroci si basa sull'impiego di una sola telecamera fissa con il

punto di osservazione piuttosto distante dalla zona controllata, per esempio su un edificio [10, 11, 12, 13]. Questi sistemi sfruttano generalmente algoritmi di rilevazione del movimento per segmentare¹ gli oggetti basati sulla costruzione di un'immagine di riferimento dello sfondo in continuo aggiornamento. In alcuni casi, per incrementare la robustezza del rilevamento nel caso avvengano quelle variazioni di illuminazione che tipicamente caratterizzano gli ambienti esterni, sono stati utilizzati modelli dello sfondo ricavati statisticamente [10] o un'immagine di riferimento composta dai bordi fissi della scena [13].

Uno dei problemi principali relativi alla tecnica di *background subtraction*² - diffusamente utilizzata per l'individuazione di oggetti in movimento da sistemi a telecamera fissa - è che anche le ombre vengono rilevate come parti degli oggetti in movimento, diventando un punto critico se la classificazione dell'oggetto rilevato si basa sulle dimensioni o sulla forma. Per questo motivo vengono spesso adottate tecniche specifiche per la rimozione delle ombre [11, 14].

¹Procedimento di partizione di un'immagine in regioni multiple allo scopo di semplificare e/o cambiare la rappresentazione di un'immagine in parti di essa che siano più significative e facili da analizzare.

²Tecnica di elaborazione delle immagini che permette di identificare oggetti in movimento nella porzione di un'immagine che differisce in modo significativo da un'immagine dello sfondo.

Capitolo 2

Considerazioni preliminari

Abbiamo studiato per il sistema un allestimento molto generale che potesse essere adattato a tipi diversi di passaggi pedonali. Una coppia di telecamere è stata installata sopra un supporto quadrato, all'estremità di un palo a L, affacciato sulle strisce pedonali dell'attraversamento (Figura 2.1). Sono state selezionate telecamere operanti nello spettro del visibile sensibile anche all'intervallo di radiazioni del vicino infrarosso ed illuminatori nel vicino infrarosso per permettere, in eventuali sviluppi successivi, la rilevazione notturna dei pedoni.

Così come per il sistema prototipo, prevedendone estensioni future, è stato scelto un incrocio pedonale e ciclabile, anche se i contrassegni riservati alle biciclette non sono stati presi in considerazione in questa fase, per non perdere di generalità.

Le telecamere sono state collocate a 6 m dal terreno, rendendo necessario l'impiego di ottiche *fish-eye* che permettessero di inquadrare tutto il passaggio pedonale.

Sono state inizialmente installate sul supporto 3 telecamere, allo scopo di valutare quale tra le due configurazioni - con *baseline*¹ delle telecamere per-

¹Distanza tra gli assi ottici delle due telecamere che formano il sistema stereoscopico.

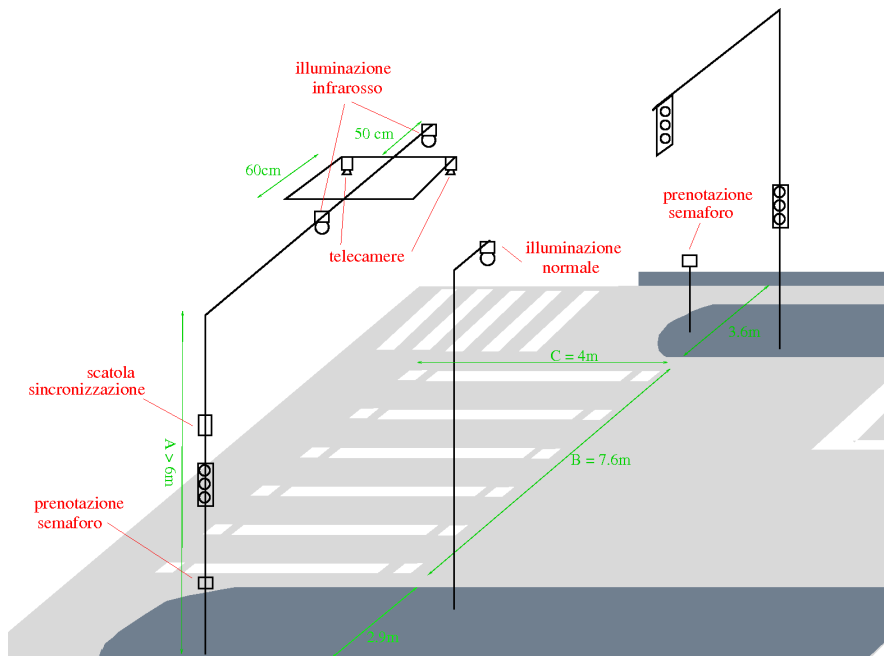


Figura 2.1: Configurazione finale del sistema: 2 telecamere e 2 illuminatori che si affacciano sul passaggio pedonale.



Figura 2.2: Particolare del sistema di acquisizione nella sua configurazione iniziale.



Figura 2.3: Fotografie del sistema installato - nella sua configurazione iniziale.

pendicolare o parallela alle strisce pedonali - risultasse la migliore ai fini del rilevamento.

Dall'osservazione delle immagini, quelle acquisite con configurazione perpendicolare (Figura 2.4) sono risultate più deformate a causa della disposizione del palo: l'asse longitudinale del braccio affacciato sulle strisce non era disposto in direzione perpendicolare a queste, ma sensibilmente ruotato. Presentando un'inquadratura molto distorta dell'attraversamento, sono state ritenute inadatte all'applicazione poiché avrebbero reso la fase di rimozione della distorsione - preliminare all'elaborazione - meno efficace.

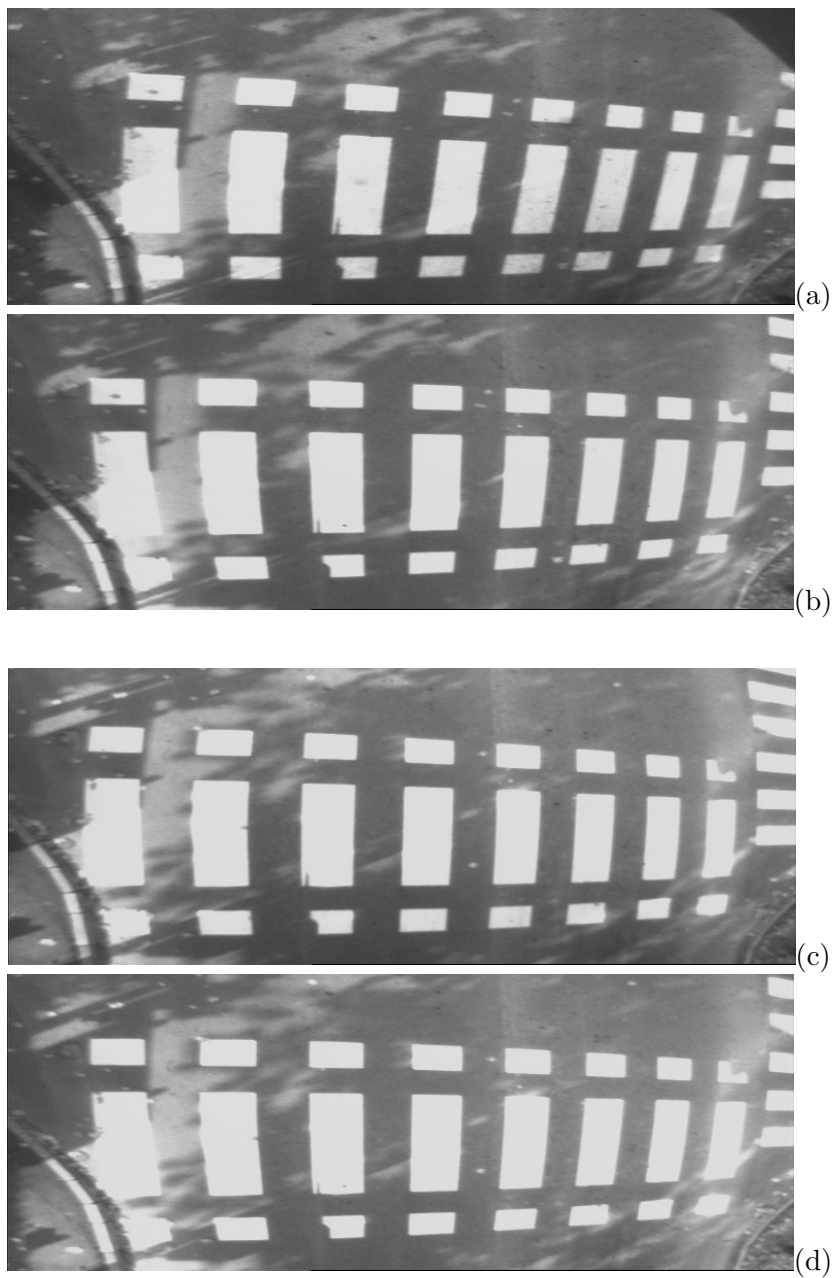


Figura 2.4: Immagini acquisite dalle telecamere di sinistra - (a),(c) - e di destra - (b), (d) - nelle due configurazioni di test: ortogonale e parallela alle strisce del passaggio pedonale.

Capitolo 3

Procedimento

3.1 In breve

Per la segmentazione degli oggetti, è stato adottato un approccio di tipo *stereo matching*¹. Dal momento che le due immagini stereo devono essere confrontate, la distorsione introdotta dalle lenti deve essere compensata utilizzando un modello appropriato della lente stessa. Questo è il primo passo da affrontare. L'idea principale su cui si basa l'algoritmo consiste nell'applicazione della Inverse Perspective Mapping² alle immagini di sinistra e destra e l'analisi della differenza tra le due viste dall'alto ottenute per estrarre gli oggetti emergenti dal piano stradale. La differenza stereo può essere affetta da disturbi dovuti a piccoli errori di calibrazione del sistema stereo ed alla rimozione non perfetta della distorsione della lente. Per affrontare i problemi dovuti a questi disturbi, un'immagine di riferimento dello sfondo viene sottratta dalla differenza stereo. L'immagine di un oggetto nella differenza stereo risulta frammentata ed incompleta poiché deriva dalla discrepanza tra ogni oggetto ed il terreno, essendo

¹Tecnica che cerca le corrispondenze tra i pixel di due immagini stereoscopiche.

²Tecnica di trasformazione geometrica che proietta ogni pixel della vista prospettica 2D di un oggetto 3D rimappandolo in una nuova posizione e costruisce una nuova immagine su un piano 2D [15].

quest'ultimo caratterizzato dall'asfalto e dalle strisce bianche tracciate su di esso. In effetti gli oggetti appaiono come agglomerati di aree di colore grigio chiaro. Si effettua quindi una ricerca di gruppi di pixel chiari adiacenti e ciascuno viene racchiuso in un *bounding box*³. I bounding box adiacenti o sovrapposti appartengono, con molta probabilità, allo stesso oggetto e vengono dunque fusi in un unico bounding box. I bounding box ricavati da questa fase vengono poi espansi per aderire meglio alla forma dell'oggetto utilizzando punti specifici del piano stradale come sensori di presenza.

I veicoli e gli altri oggetti in moto lungo le corsie della strada vengono distinti dai pedoni grazie a considerazioni relative alla direzione lungo la quale avviene il loro movimento in fotogrammi successivi.

Non viene effettuato un *tracking*⁴ completo degli oggetti, ma si applica una correlazione temporale per eliminare i falsi positivi non persistenti e compensare gli occasionali falsi negativi.

3.2 In dettaglio

3.2.1 Rimozione della distorsione e della prospettiva

Distorsione e prospettiva vengono rimosse dalle immagini acquisite tramite una rapida mappatura dei pixel basata su una Look-Up Table⁵. La mappatura da applicare dipende dai parametri specifici delle lenti e della calibrazione. Per la caratterizzazione delle lenti e la calibrazione delle telecamere è stata sviluppata un'apposita procedura automatica [16] basata sull'osservazione delle

³Rettangoli che delimitano zone ritenute di interesse all'interno di un'immagine.

⁴Tecnica di inseguimento degli oggetti rilevati in un'immagine che utilizza dati relativi ad ogni oggetto in immagini precedenti per predire la posizione dello stesso oggetto nell'immagine attuale.

⁵Struttura dati usata per sostituire operazioni di calcolo a tempo di esecuzione di un programma con una più semplice operazione di consultazione; nell'elaborazione delle immagini sono tabelle che collegano indici numerici (generalmente da 0 a 255) con un valori di output.

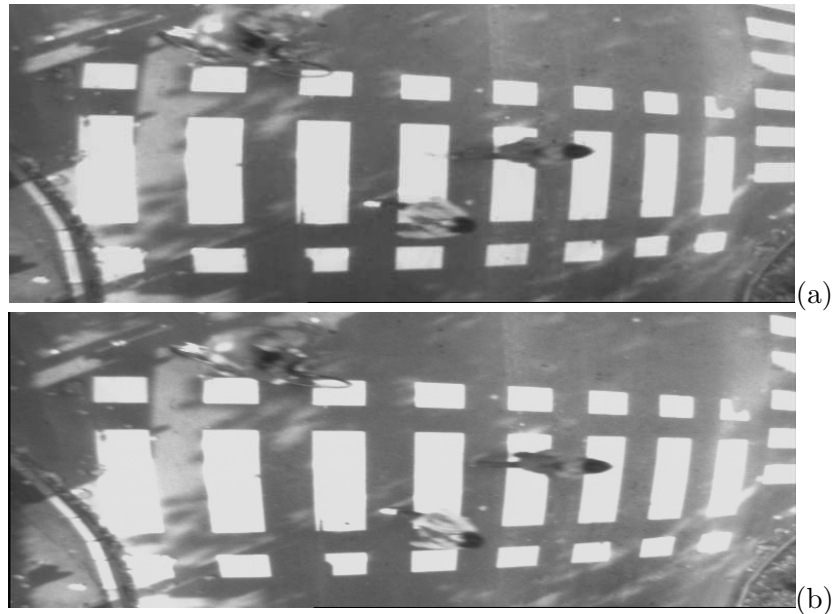


Figura 3.1: Immagini acquisite dalle telecamere sinistra (a) e destra (b).

strisce pedonali nelle immagini, in particolare assumendo che queste siano parallele e delle stesse dimensioni nel mondo.

Questa procedura di calibrazione è di grande importanza per rendere possibile il confronto tra le due immagini stereo. La configurazione del sito infatti rende difficile ottenere in modo preciso l'allineamento e l'orientazione corretta delle telecamere. In certi casi può anche risultare impossibile posizionare l'apparato stereo in corrispondenza del centro esatto della carreggiata. Inoltre la calibrazione automatica permette al sistema di funzionare con diversi tipi di telecamere e lenti, la cui scelta dipende dalle caratteristiche dell'incrocio.

La ricalibrazione automatica non richiede l'intervento di un operatore umano e può avvenire ad intervalli regolari per compensare le deviazioni relative a orientazione e posizione delle telecamere che possono verificarsi durante il funzionamento del sistema a causa di vibrazioni, variazioni di temperatura e

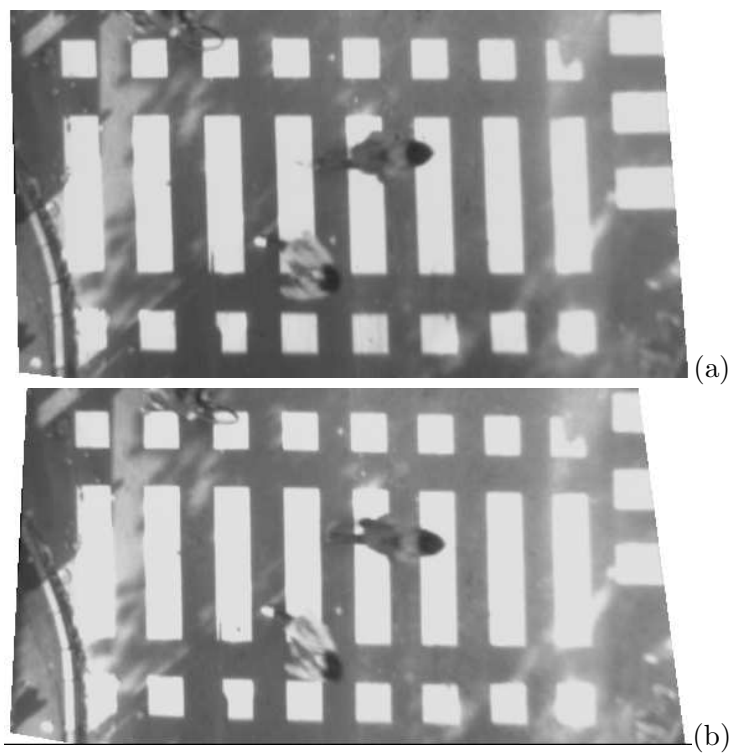


Figura 3.2: Immagini acquisite dalle telecamere dopo la rimozione della distorsione e dell'effetto prospettico.

fenomeni atmosferici.

3.2.2 Differenza stereo e *background subtraction*

Le due viste dall'alto ottenute dopo aver applicato la Inverse Perspective Mapping (Figura 3.2.a e 3.2.b) differiscono significativamente solo in corrispondenza degli oggetti che si ergono dal piano stradale; una volta ricavata la loro differenza, questa può essere utilizzata per le fasi di ricerca successive. L'operazione elimina tutti quegli oggetti che appartengono al terreno come ombre, foglie e segnaletica orizzontale. A causa delle piccole variazioni che si ottengono in seguito al processo di calibrazione della coppia stereo e della non perfetta rimozione della distorsione delle immagini dovuta alle lenti, le immagini ottenute dopo la rimappatura differiscono ancora leggermente, in modo tale che l'immagine della loro differenza contiene alcuni elementi appartenenti allo sfondo, come mostrato in Figura 3.3.a. Per affrontare i problemi causati da questi disturbi, si esegue la sottrazione tra un'immagine di riferimento dello sfondo e quella della differenza stereo. Questa immagine di riferimento viene calcolata come media pesata di un numero prefissato delle immagini di differenza stereo precedenti. Ai valori più recenti che contribuiscono alla media viene assegnato un peso piccolo rispetto ai valori più vecchi, in modo tale che l'immagine di riferimento risulta contenere solo elementi statici, mentre elementi che appaiono occasionalmente risultano avere scarsa influenza.

3.2.3 Estrazione degli oggetti

L'immagine della differenza stereo dopo la rimozione dello sfondo mostrata in Figura 3.3.b risulta caratterizzata da poche zone chiare su sfondo nero. Questi raggruppamenti di pixel chiari - li indicheremo come *blob* - vengono identificati tramite una procedura che calcola in modo alternato gli istogrammi relativi alle righe ed alle colonne dell'immagine. Tali istogrammi evidenziano quali sono le strisce orizzontali e verticali più interessanti dell'immagine. Intersecando

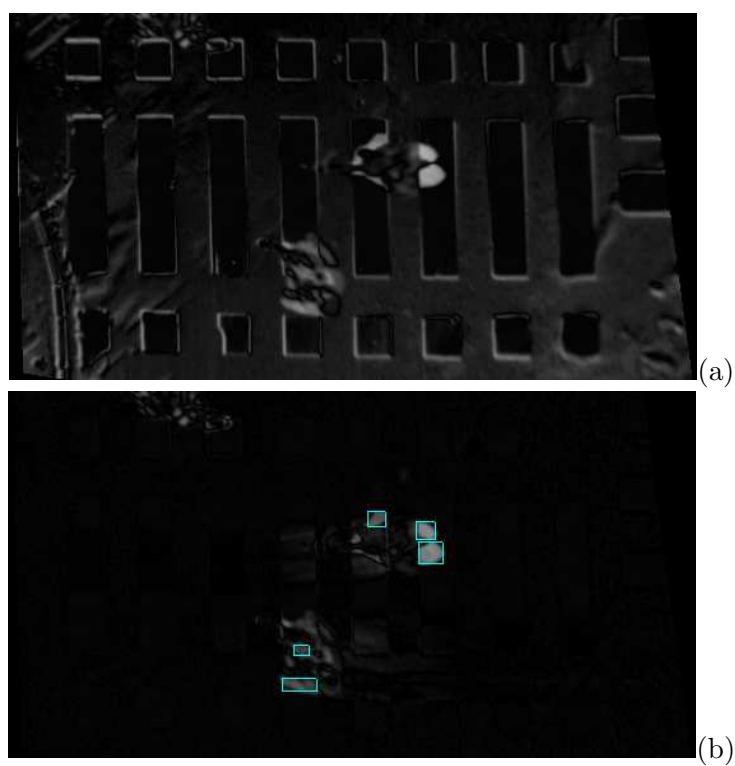


Figura 3.3: Immagine della differenza stereo (a) prima e (b) dopo la rimozione dello sfondo. I blob chiari vengono delimitati da bounding box di colore azzurro.

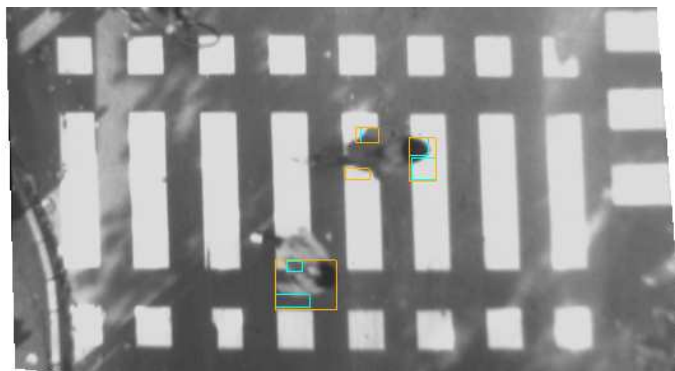


Figura 3.4: I bounding box azzurri adiacenti o sovrappoventesi vengono fusi insieme: il risultato è un bounding box di colore arancio. Questo box può essere più largo del più piccolo box che include tutti i box azzurri (come avviene per il pedone nella parte inferiore dell'immagine) perché anche i box derivanti dal fotogramma precedente (non mostrati in quest'immagine) sono coinvolti nel processo di fusione.

ripetutamente queste strisce e ricalcolando gli istogrammi nelle aree risultanti porta a trovare i bounding box che racchiudono le regioni chiare (Figura 3.3.b). Ogni bounding box viene preso in considerazione nelle fasi di elaborazione successive solo se ritenuto sufficientemente grande.

Poiché i blob chiari appartenenti ad un oggetto possono essere molto piccoli e frammentati, i bounding box vicini tra loro o sovrapposti vengono iterativamente raggruppati insieme in un unico bounding box che li racchiude tutti. Inoltre, per poter affrontare l'elevata variabilità dei blob corrispondenti ad uno stesso oggetto da fotogramma a fotogramma, viene tenuta in considerazione la storia dei bounding box. Più precisamente, i bounding box etichettati come pedoni nel fotogramma precedente vengono aggiunti al gruppo dei box che devono essere fusi insieme nel fotogramma corrente. In questo modo si ottengono risultati più stabili. Un esempio viene mostrato in Figura 3.4.

La fase successiva dell'algoritmo ha lo scopo di definire con ulteriori e più precisi adattamenti, i bordi di delimitazione dei bounding box. In particolare,

i bounding box vengono espansi per aderire meglio alla forma dell'oggetto che contengono utilizzando punti specifici sul terreno che abbiamo denominato *sensori di presenza*. Questi sensori sono punti equispaziati disposti sui bordi delle strisce (punti verdi e rossi in Figura 3.5).

Questi punti sono caratterizzati da un intorno di due aree uniformi, una grigia (l'asfalto) e l'altra bianca (la striscia). Un oggetto (pedone o veicolo) che passi sopra un sensore - o meglio, che lo copra - con grande probabilità ne cambierà il colore e la configurazione delle aree nell'intorno specifico. Questa variazione viene interpretata come un cambiamento da 'spento' ad 'acceso' nello stato del sensore.

I bounding box vengono espansi su ogni lato fino a raggiungere il più vicino tra i sensori esterni il cui stato sia cambiato da spento ad acceso. Dopo questo adattamento, la procedura iterativa che fonde i bounding box vicini e sovrapposti viene ripetuta (Figura 3.5).

3.2.4 Classificazione di pedoni e veicoli

I veicoli non possono essere distinti dai pedoni basandosi solo sulle dimensioni dei bounding box che li circondano poiché quando solo una porzione dei veicoli è presente nell'immagine, le dimensioni del bounding box di tale porzione può essere simile a quella dei pedoni. Inoltre, pedoni adiacenti fra loro possono generare un unico largo bounding box e venire quindi classificati erroneamente. La discriminazione tra pedoni e veicoli avviene considerando la direzione del loro movimento. Si assume che gli oggetti presenti nella scena osservata siano pedoni se il loro movimento nell'immagine è prevalentemente orizzontale; altrimenti vengono classificati come veicoli.

Le posizioni dei bounding box - specificatamente dei loro baricentri - rilevate in un numero prefissato di fotogrammi precedenti, vengono conservate in un buffer. Quando la direzione di un bounding box del fotogramma corrente deve essere valutata, vengono esaminate due aree triangolari adiacenti ad esso (Figura 3.6).

L'idea è quella di verificare se il box ha lasciato dietro di sé (ovvero nei fotogrammi precedenti) una scia prevalentemente orizzontale alla sua destra o alla sua sinistra. Se viene rilevato un insieme adeguato di tracce nella zona triangolare sinistra/destra del box, questo viene classificato come box contenente un pedone diretto a destra/sinistra. Altrimenti si etichetta il contenuto del box come non-pedone (per es. veicolo).

Questa operazione introduce un piccolo ritardo nell'operazione di riconoscimento dei pedoni dal momento in cui compaiono nell'immagine. Poiché il sistema funziona a 23 Hz, questo ritardo può essere ritenuto trascurabile.

Durante la valutazione della qualità di una scia contenuta nel triangolo di sinistra, vengono esaminate solo le tracce lasciate dai pedoni e non-pedoni diretti verso destra, valutando come maggiore il contributo delle prime. Considerazioni analoghe vengono applicate al triangolo di destra.

3.2.5 Impiego della correlazione temporale

Poiché i bounding box variano molto da fotogramma a fotogramma, effettuare il tracking degli oggetti rilevati non è un problema triviale. In effetti, a causa della prospettiva, le dimensioni della figura del pedone variano nelle diverse regioni dell'immagine. Inoltre la forma del pedone cambia molto mentre cammina, particolarmente dal punto di vista dall'alto, come quello scelto in questa sperimentazione.

Per rendere più stabili i risultati si è dunque fatto ricorso alla correlazione temporale.

Per ogni pedone precedentemente rilevato ci si assicura che un nuovo pedone che procede nella stessa direzione sia presente nell'immagine corrente in posizione compatibile con il suo movimento. Altrimenti si è verificata la presenza di un falso negativo che deve essere compensata. Viene aggiunto un finto pedone con la stessa direzione in una posizione stimata dalla sua traiettoria recente. In caso di rilevamento sbagliato in più fotogrammi successivi, questa

correzione può avvenire per un numero determinato di volte, finché il pedone non viene “agganciato” di nuovo o perso definitivamente.

In più, per eliminare la comparsa isolata di falsi positivi, viene richiesta una certa persistenza in fotogrammi consecutivi prima che un pedone appena comparso possa essere incluso nel risultato finale. La breve latenza introdotta nel rilevamento è, anche in questo caso, trascurabile.

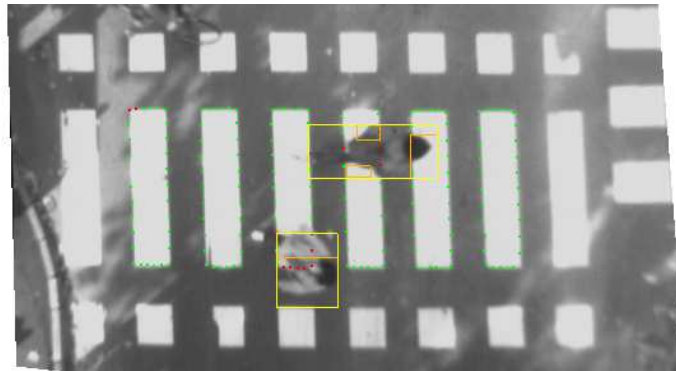


Figura 3.5: I punti verdi rappresentano i sensori spenti (nessun oggetto in sosta o transito sopra di loro), mentre i punti rossi rappresentano i sensori accesi (coperti da un oggetto che si trova sopra di essi). I bounding box di colore arancio vengono espansi fino al sensore esterno rosso più vicino e quindi fusi se vicini o sovrapposti. I bounding box risultanti sono mostrati in giallo.

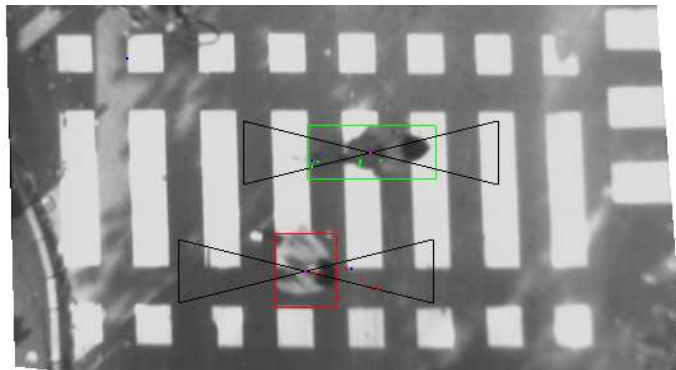


Figura 3.6: I pedoni vengono indicati con un bounding box verde, se procedono verso destra; da uno rosso se procedono verso sinistra. I loro baricentri vengono indicati con un pallino di color fucsia, mentre le due aree triangolari sono delimitate da semirette tracciate in nero. Le tracce vengono mostrate con punti verdi se lasciate da pedoni diretti a destra, con punti rossi, se lasciate invece da pedoni diretti a sinistra.

Capitolo 4

Risultati sperimentali

Il sistema descritto è stato sviluppato per migliorare la sicurezza dei pedoni ed ottimizzare il flusso veicolare in corrispondenza degli attraversamenti pedonali. Dopo la calibrazione, eseguita con procedure automatiche, viene svolta la fase di rilevamento dei pedoni.

Allo scopo di effettuare un rilevamento corretto dei pedoni, la fase preliminare di calibrazione risulta estremamente importante. Il sistema è stato ideato per ricalibrarsi automaticamente in orari prefissati (gli orari più adatti sono quelli in cui risulta minore la probabilità di trovare pedoni o veicoli in attraversamento sulle strisce, come il mattino molto presto) oppure manualmente da un operatore umano.

Le immagini elaborate sono state acquisite in condizioni luminose diurne, in diversi orari e diverse stagioni dell'anno per testare l'efficacia dei sensori. Pur essendo piuttosto flessibile dal punto di vista del funzionamento, le migliori prestazioni dell'algoritmo sono state ottenute nel caso di condizioni luminose diurne. Il sistema di rilevamento dei pedoni può effettuare l'intera computazione in 44 ms su un processore Pentium IV @ 2.8 GHz.

Le figure seguenti 4.1 presentano alcuni dei risultati ottenuti in diverse situazioni. I pedoni rilevati sono evidenziati da un pallino di colore blu. Quando

il sistema trova le strisce occupate da almeno una persona, nella zona in alto a destra dell'immagine compare anche il segnale stradale che avverte della presenza di pedoni in attraversamento.

La figura 4.1.a mostra una situazione di rilevamento semplice: due pedoni separati che camminano sulle strisce.

Anche se il sistema è stato progettato per monitorare solamente passaggi pedonali ordinari, quello scelto per la sperimentazione era sia per pedoni che per ciclisti.

La figura 4.1.b mostra il corretto rilevamento di due pedoni ed un ciclista passanti sulle strisce.

In figura 4.1.c la presenza di molte persone, anche molto ravvicinate tra loro, rende il rilevamento di ogni singolo pedone molto difficile. Ciononostante l'algoritmo è in grado di rilevare la presenza di alcuni dei pedoni presenti sulle strisce, fatto sufficiente ad indicare che il passaggio pedonale risulta occupato da pedoni e quindi non libero per il passaggio dei veicoli.

La figura 4.1.d mostra un'altra situazione complessa: uno dei pedoni sta utilizzando un ombrello ed attraversa al di fuori delle strisce. In questo caso, a causa della mancanza della fase di adattamento sui sensori, i bounding box appartenenti all'ombrello non possono essere aggregati tra loro, generando rilevamenti multipli.

Il sistema proposto pu essere utilizzato per migliorare il controllo semaforico del traffico: pu aumentare la sicurezza per gli utenti della strada più vulnerabili incrementando la disciplina dei pedoni presso gli incroci semaforici, mantenendo alti livelli di circolazione del traffico.

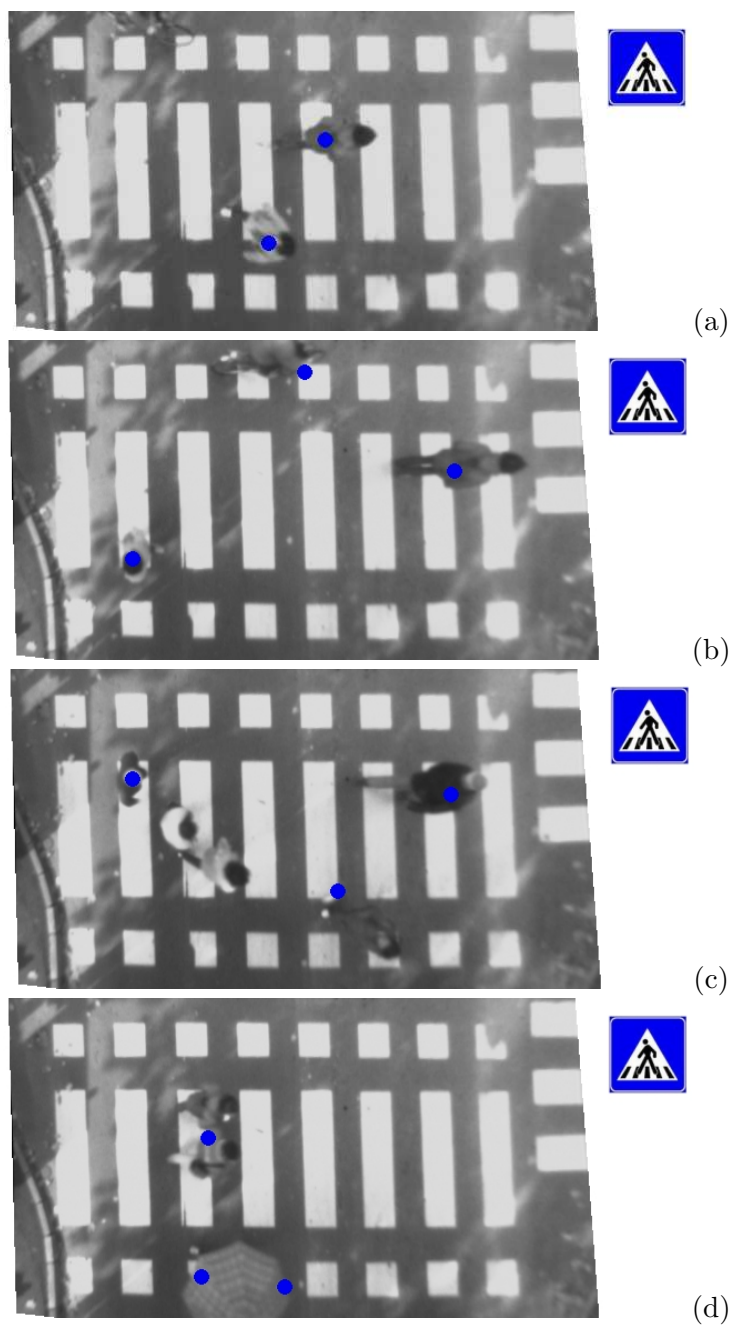


Figura 4.1: Risultati in diverse situazioni.

Parte II

Localizzazione di pedoni in ambientazioni notturne da un veicolo autonomo in movimento

Capitolo 5

Introduzione

Scopo di questa indagine è stato valutare l'impiego di una telecamera con sensore NIR per la rilevazione di pedoni di fronte ad un veicolo in movimento, ricorrendo ai soli fari abbaglianti del veicolo come fonte di illuminazione della zona ad esso antistante. Abbiamo ritenuto la radiazione NIR una banda di frequenze adatta per un dispositivo a sensore singolo.

Il rilevamento di pedoni in scenari urbani ed extraurbani è uno dei compiti più impegnativi nel campo della visione artificiale. Negli ultimi anni, a causa del grande impatto nell'ambito della sicurezza stradale, molte ricerche si sono concentrate nell'ambito del rilevamento dei pedoni. Alcuni approcci proposti in merito si basano su considerazioni relative alla morfologia umana (tecniche *shape-based*, [4, 6, 7]), sull'analisi delle tessiture (tecniche *texture-based*), sul confronto con modelli (tecniche *template-based*), sull'individuazione del movimento (tecniche *motion-based* [3, 4, 5, 6]) e sull'impiego della stereoscopia o di reti neurali [17, 18, 19]. Nel caso della rilevazione dei pedoni in ambito automobilistico, essendo il sistema di acquisizione in movimento, la distinzione degli oggetti di interesse dallo sfondo risulta più complessa, rispetto al caso di sistemi di acquisizione fissi. Inoltre i pedoni possiedono sagome estremamente variabili, dovute anche all'abbigliamento (colori, tessuti, ecc.) ed agli accessori

(borse, zaini, ecc.) che ne rendono più difficile l'identificazione.

I lavori più recenti relativi all'individuazione dei pedoni in immagini acquisite nel lontano infrarosso (FIR) hanno prodotto buoni risultati, grazie all'aspetto caratteristico degli esseri umani in questo dominio: sono stati proposti efficaci algoritmi di rilevazione dei pedoni [20], [21], [22], [23]. Tuttavia, insieme alle argomentazioni positive appena citate, rimane un certo numero di svantaggi nell'utilizzo della radiazione FIR in applicazioni automotive. Le immagini termiche infatti hanno una risoluzione spaziale¹ inferiore rispetto a quelle ottenute con la radiazione NIR, riducendo la sensibilità alle caratteristiche più dettagliate della trama dell'immagine. Queste hanno generalmente una dinamica inferiore per quanto riguarda i livelli di grigio rispetto alle NIR (Figura 5.1).



Figura 5.1: La stessa scena notturna, acquisita nello spettro del FIR (a) e del NIR (b) - con scarsa illuminazione esterna.

L'energia emessa dagli oggetti nella banda del FIR dipende dalla loro temperatura. In ambientazioni esterne c'è un numero imprevedibile di fattori che

¹Nelle tecniche di riproduzione dell'immagine, livello di dettaglio della riproduzione, che si misura in base al numero di punti impiegati per rappresentarla; ha corrispondenza con la qualità dell'immagine.

possono influenzare la temperatura e quindi il contrasto termico, come nuvole, umidità, precipitazioni e vento. Tutti questi effetti alterano e cambiano costantemente il contrasto termico rendendo gli scenari stradali visti nel dominio del lontano infrarosso a volte difficili da interpretare. Inoltre il contrasto termico variabile può avvicinarsi allo zero [24], rendendo, a volte, oggetti larghi come le auto, difficili da vedere.

Le immagini NIR invece possiedono un contrasto nei livelli di grigio migliore: la radiazione del vicino infrarosso viene riflessa dagli ostacoli in modo simile a quanto accade con le radiazioni dello spettro del visibile. Ed è infatti per questa necessità di fonti di illuminazione per poter acquisire le radiazioni riflesse dagli oggetti circostanti, che i sistemi NIR sono anche detti *attivi*.

Per fornire questa illuminazione in scenari notturni, molti sistemi utilizzano illuminatori NIR specifici [25], [26], [27] che coprono un campo visivo più ampio (fino a 150 m) rispetto a quello caratteristico dei fari anabbaglianti o abbaglianti in dotazione ai veicoli. Tuttavia questi illuminatori hanno un notevole aspetto negativo: se due veicoli, equipaggiati con illuminatori NIR, dovessero incrociarsi, questi si accecherebbero vicendevolmente, poiché la retina dell'occhio umano è vulnerabile alle scottature termiche emesse dalla radiazione NIR. Per questi motivi nel sistema sviluppato si è voluto investigare un sistema di visione NIR che utilizza solo l'equipaggiamento standard dei veicoli, fari anabbaglianti o abbaglianti.

Molti studi presentati in letteratura descrivono sistemi per l'individuazione di pedoni in sequenze video prese da telecamere NIR installate su un veicolo. Alcuni di questi fanno uso di una sola telecamera [28] o di un sistema ibrido con una telecamera NIR ed una FIR [29]. Anche gli illuminatori NIR vengono spesso impiegati [30], [31].

Per ridurre i costi di equipaggiamento ed evitare possibili danni agli occhi causati dalle radiazioni NIR, è stato studiato un sistema di acquisizione con una sola telecamera che fa uso solo dei fari del veicolo per la visione notturna.

Capitolo 6

Considerazioni preliminari

Abbiamo utilizzato un equipaggiamento formato da una telecamera con sensore NIR e dai fari anabbaglianti ed abbaglianti del veicolo, come sorgenti di illuminazione. Abbiamo quindi analizzato le immagini acquisite con il nostro sistema in differenti condizioni luminose, per trovare quali potessero essere le più adatte all'individuazione di pedoni.

Condizioni diurne

La radiazione NIR viene emessa insieme a quella visibile e possiede la stessa distribuzione spaziale. Per questi motivi, a causa della contiguità tra lo spettro NIR e quello visibile, gli oggetti naturali possono riflettere entrambe le radiazioni in modo molto simile. Le immagini acquisite in condizioni luminose diurne sono risultate povere dal punto di vista del contenuto informativo (Figura 6.1).

Condizioni notturne utilizzando i fari anabbaglianti

Come accennato, essendo il NIR una radiazione riflessa, di notte, per poter acquisire immagini significative, è necessario l'impiego di fonti di illuminazione esterne. Avendo scartato gli illuminatori NIR specifici, abbiamo inizialmente



Figura 6.1: La stessa scena acquisita nello spettro visibile (a) e NIR (b)

utilizzato i fari anabbaglianti, dotazione standard dei veicoli. L'idea sottesa era di sviluppare un algoritmo di rilevamento che fosse in grado di rilevare i pedoni localizzandone le sole gambe. I fari anabbaglianti infatti illuminano solo la parte bassa di fronte al veicolo, lasciando quella superiore più oscurata. Il vantaggio di questo tipo di immagini, oltre ad evitare che i veicoli incrociati ed i pedoni vengano abbagliati, è che, essendo più piccola l'area illuminata, altri tipi di disturbi, come i segnali stradali, non compaiono nelle immagini NIR o, in caso lo facciano, il loro tono di grigio è inferiore rispetto a quello delle gambe dei pedoni, agevolando le procedure per la loro individuazione. Purtroppo la zona illuminata è fortemente asimmetrica (il faro di sinistra è orientato verso il basso, mentre nella zona di destra, il faro punta più verso l'alto), compromettendo l'individuazione dei pedoni che arrivano dal lato sinistro del veicolo, che appaiono nell'ombra (Figura 6.2.a).

Condizioni notturne utilizzando i fari abbaglianti

Gli esperimenti successivi sono stati dedicati a valutare l'impiego dei fari abbaglianti del veicolo. Il problema principale sorto in questo caso è stato che il

cono di luce proiettato dai fari non risultava abbastanza alto da illuminare le teste dei pedoni in modo adeguato, così come invece accadeva per i loro busti e le gambe (Figura 6.2.b). A causa della particolare proporzione esistente tra la testa e le spalle negli esseri umani, molto spesso questo elemento risulta di grande importanza per discriminare i pedoni da altri oggetti appartenenti all'ambiente urbano, come i cartelli stradali e gli alberi. Tuttavia, l'area illuminata risultante si è dimostrata sufficientemente estesa ed uniforme, permettendo ad ostacoli generici e pedoni di apparire in buon contrasto con lo sfondo.

Valutate le prove precedenti, abbiamo scelto di adottare la terza delle situazioni analizzate (Figura 6.3). La configurazione finale del sistema impiegato è rappresentata graficamente in Figura 6.4. La telecamera utilizzata per le acquisizioni è stata installata nell'abitacolo della vettura (Figura 6.5.b), vicino allo specchietto retrovisore (Figura 6.5.a). L'intervallo di rilevazione delle distanze è compreso tra 9 m e 35 m, poiché i pedoni che si trovano al di fuori di questo intervallo non vengono illuminati in modo adeguato per essere localizzati.



Figura 6.2: Uso dei fari anabbaglianti (a) ed abbaglianti (b)



Figura 6.3: Pedoni visti dal sistema NIR

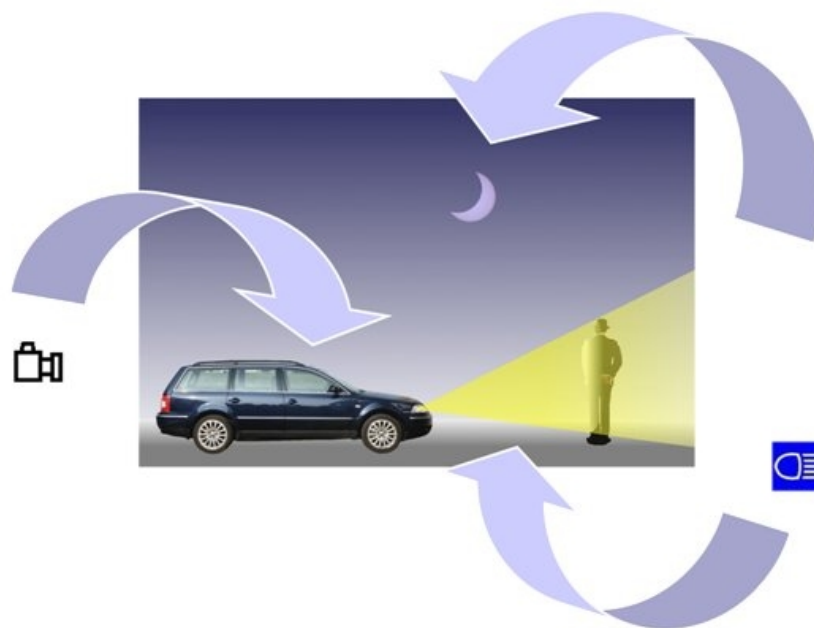


Figura 6.4: Riassunto grafico del sistema: telecamera con sensore NIR a bordo di un veicolo, ambientazioni notturne, illuminazione della zona antistante il veicolo ottenuta dai suoi fari abbaglianti.



(a)



(b)

Figura 6.5: La telecamera NIR installata vicino allo specchietto retrovisore del veicolo (a) ed un'immagine del veicolo utilizzato (b).

Capitolo 7

Procedimento

7.1 In breve

Una caratteristica rilevante dell'algoritmo è che è stato sviluppato senza l'utilizzo dei parametri di calibrazione della telecamera. Questa scelta ha comportato la necessità di gestire alcune difficoltà in più dal punto di vista dello sviluppo dell'algoritmo, ma ha comportato grandi semplificazioni della fase di preparazione del sistema. I dati di calibrazione potranno essere introdotti in fase successiva per il raffinamento dei risultati. L'algoritmo (tracciato in Figura 7.1), prevede un'elaborazione preliminare dell'immagine, in modo da aumentare il contrasto tra le zone molto chiare, con forte riflessione della radiazione NIR. Successivamente vengono individuate le aree contenenti possibili pedoni, attraverso la ricerca delle regioni più chiare dell'immagine, isolate da opportuni bounding box. Si effettua una prima valutazione dei bounding box trovati, in base a criteri di forma e proporzione, dopodiché si passa al confronto, secondo due metodi diversi, con modelli. La combinazione pesata dei contributi ottenuti dai confronti precedenti, permette di ottenere la validazione o lo scarto dei bounding box.

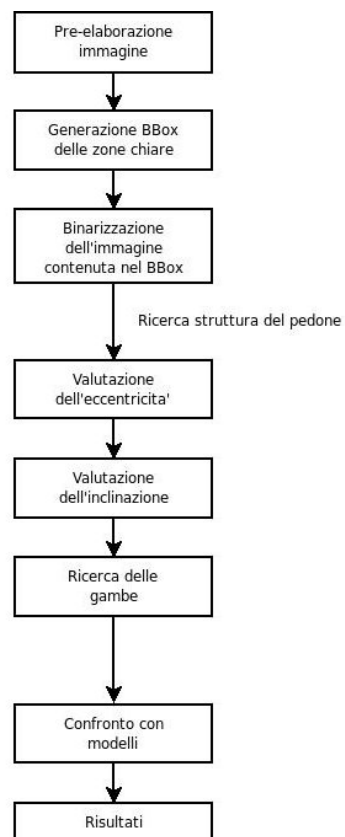


Figura 7.1: Diagramma dell'algoritmo.

7.2 In dettaglio

7.2.1 Elaborazione preliminare dell'immagine

La prima operazione di basso livello consiste nell'aumentare il contrasto delle zone più chiare dell'immagine. Alti valori di luminosità identificano oggetti, come i cartelli stradali, la pelle dei pedoni ed alcuni tessuti utilizzati nell'abbigliamento che riflettono fortemente la radiazione NIR. L'introduzione di questo risalto, migliora il contrasto tra queste zone e lo sfondo (Figura 7.2.b). Le zone rilevate vengono quindi circondate da bounding box (Figura 7.2.c) che da questo punto in avanti diventeranno le aree del cui contenuto valutare le caratteristiche, anziché operare sull'immagine intera rendendo l'algoritmo computazionalmente meno efficiente. Sui bounding box trovati viene effettuata una prima operazione di filtraggio mirata ad eliminare quelli di proporzioni non plausibili per un pedone (come bounding box molto alti e stretti o bassi e larghi, Figura 7.2.d). Poiché il perimetro dei bounding box può risultare impreciso rispetto alla sagoma completa del pedone, viene applicata un'espansione del bounding box stesso valutando, per ogni riga ed ogni colonna di un pixel che circondano il bounding box di partenza, se vi sono contenuti un numero sufficiente di pixel che superino una soglia prefissata, quest'ultima ottenuta empiricamente come metà del valor medio di colore dei pixel contenuti nel bounding box di partenza (Figura 7.3).

7.2.2 Valutazione dei bounding box

Utilizzando la soglia calcolata precedentemente, i bounding box vengono binarizzati. La valutazione di tutto il corpo e delle gambe avviene ricorrendo ai momenti geometrici, un gruppo di descrittori statistici che permettono di misurare alcuni parametri dell'immagine ritenuti invarianti. Questi descrittori consentono di stimare alcune proprietà di una figura (il corpo umano, nel nostro caso) e di conseguenza effettuare una normalizzazione della figura, rendendo il suo riconoscimento insensibile alla posizione in cui si trova, alle sue

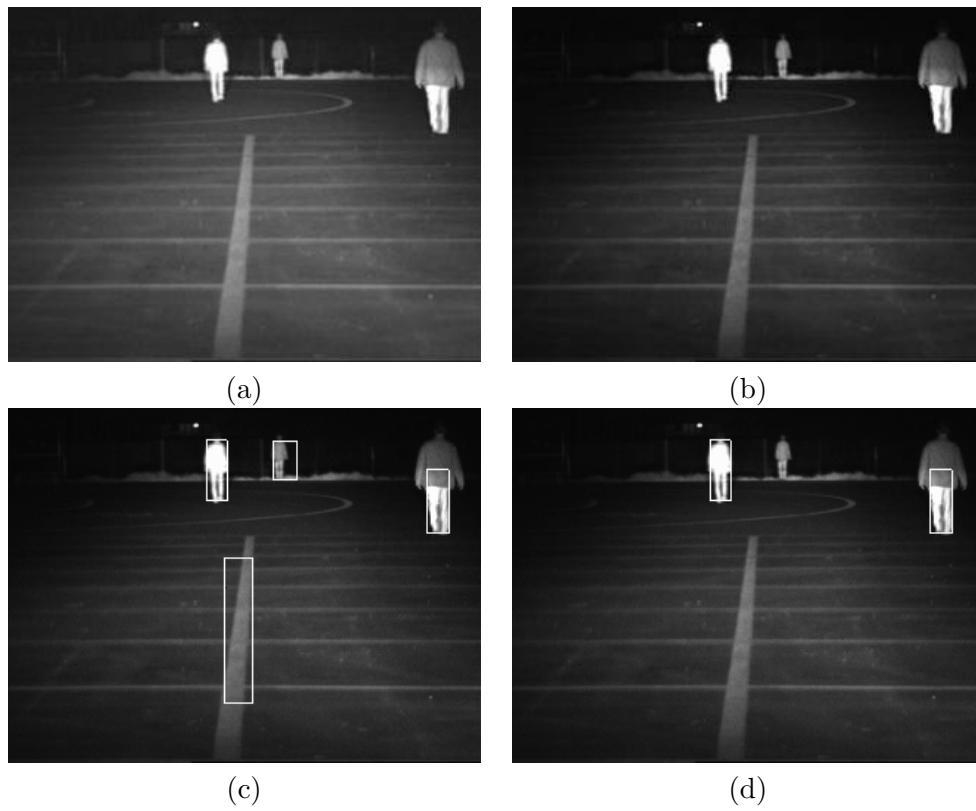


Figura 7.2: Passi dell'elaborazione preliminare operata sull'immagine: immagine originale (a), dopo la correzione del contrasto(b), dopo il tracciamento dei bounding box (c) ed in seguito al primo filtraggio di questi (d).

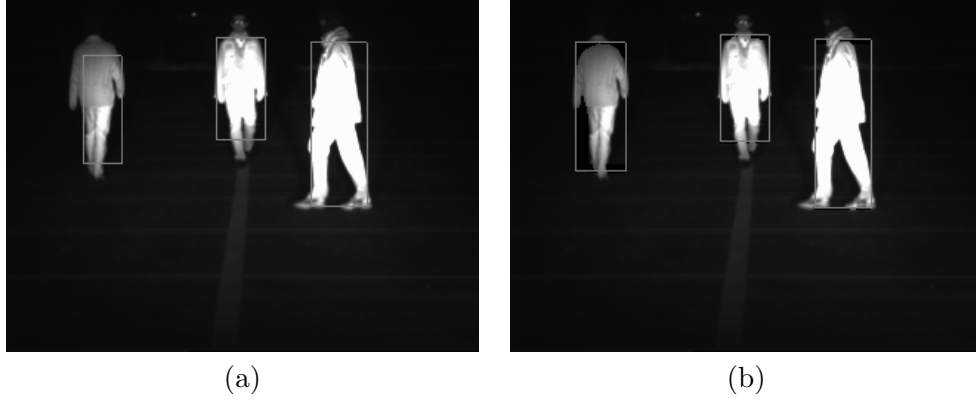


Figura 7.3: Espansione dei bounding box: originali (a) ed espansi (b).

dimensioni, ad eventuali deformazioni o rotazioni ed a piccole quantità di rumore. Per una generica funzione discreta in due variabili, i momenti vengono definiti come:

$$M_{jk} \doteq \int \int X^j Y^k f(X, Y) dX dY \quad (7.1)$$

in cui j e k identificano *l'ordine* del momento rispettivamente nelle dimensioni X ed Y . Nel caso particolare dell'uso dei momenti per l'analisi di una funzione discreta in due dimensioni, quale può essere un'immagine, l'equazione 7.1 si riscrive come:

$$M_{jk} \doteq \sum \sum X^j Y^k I(X, Y) \quad (7.2)$$

dove la funzione $I(X, Y)$ rappresenta il valore binarizzato del pixel alle coordinate (X, Y) . Di particolare interesse sono i momenti di ordine zero e primo: analizzando infatti l'equazione 7.2 si trova che:

$$M_{0,0} = \sum \sum I(X, Y) \equiv Area \quad (7.3)$$

$$M_{1,0} = \sum \sum X \cdot I(X, Y) \quad (7.4)$$

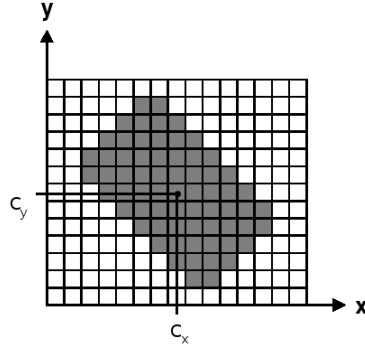


Figura 7.4: Baricentro di un gruppo di pixel.

$$M_{0,1} = \sum \sum Y \cdot I(X, Y) \quad (7.5)$$

da cui si può ottenere:

$$C_x = \frac{M_{1,0}}{Area} \quad (7.6)$$

$$C_y = \frac{M_{0,1}}{Area} \quad (7.7)$$

in cui C_x e C_y rappresentano le coordinate del baricentro, o centro di massa, dell'immagine binarizzata.

È quindi possibile calcolare i momenti di ordine superiore avendo come punto di riferimento le coordinate del baricentro, per cui l'espressione dell'equazione 7.2 diventa:

$$M_{jk} \doteq \sum \sum (X - C_X)^j \cdot (Y - C_Y)^k \cdot I(X, Y) \quad (7.8)$$

invariante alle traslazioni dell'immagine. Dividendo inoltre l'equazione 7.8 per $Area$, si ottiene l'invarianza alla scalatura dell'immagine nel calcolo dei

momenti di ordine superiore:

$$M_{jk} \doteq \frac{1}{Area} \cdot \sum \sum (X - C_X)^j \cdot (Y - C_Y)^k \cdot I(X, Y) \quad (7.9)$$

Da quest'espressione si possono calcolare i momenti di secondo ordine di riga, colonna e misti, molto importanti per capire com'è orientata l'immagine nel piano (Figura 7.5):

$$M_X = M_{2,0} \doteq \frac{1}{Area} \cdot \sum \sum (X - C_X)^2 \cdot I(X, Y) \quad (7.10)$$

$$M_Y = M_{0,2} \doteq \frac{1}{Area} \cdot \sum \sum (Y - C_Y)^2 \cdot I(X, Y) \quad (7.11)$$

$$M_{XY} = M_{1,1} \doteq \frac{1}{Area} \cdot \sum \sum (X - C_X) \cdot (Y - C_Y) \cdot I(X, Y) \quad (7.12)$$

in cui:

- M_X rappresenta il momento di inerzia, relativo all'asse Y, rispetto al centro di massa;
- M_Y rappresenta il momento di inerzia, relativo all'asse X, rispetto al centro di massa;
- M_{XY} rappresenta il momento di inerzia misto, relativo agli assi X ed Y, rispetto al centro di massa. Vale 0 se X o Y sono gli assi *principali* della figura.

Utilizzando i momenti M_X, M_Y e M_{XY} è possibile infatti calcolare esattamente l'angolo secondo il quale è ruotata una figura rispetto al proprio *asse principale* o *asse ad inerzia minima*:¹

$$\tan 2 \cdot \alpha = \frac{2 \cdot M_{XY}}{M_X - M_Y} \quad \Longrightarrow \quad \alpha = \frac{1}{2} \arctan \frac{2 \cdot M_{XY}}{M_X - M_Y} \quad (7.13)$$

¹Inerzia relativa ad un asse: somma delle distanze quadratiche tra ogni pixel dell'immagine e l'asse.

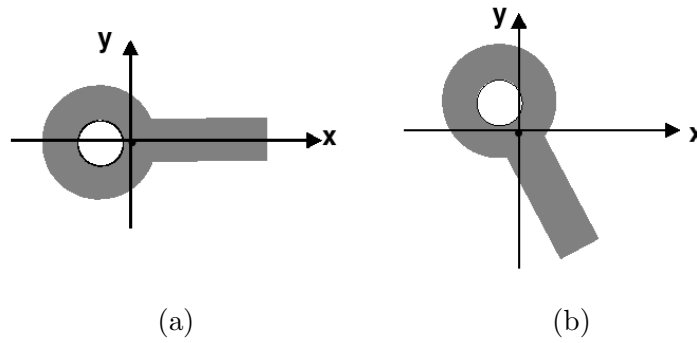


Figura 7.5: Momenti di secondo ordine: in (a) $M_X > M_Y$ e $M_{XY} = 0$, mentre in (b) $M_X < M_Y$ e $M_{XY} \neq 0$.

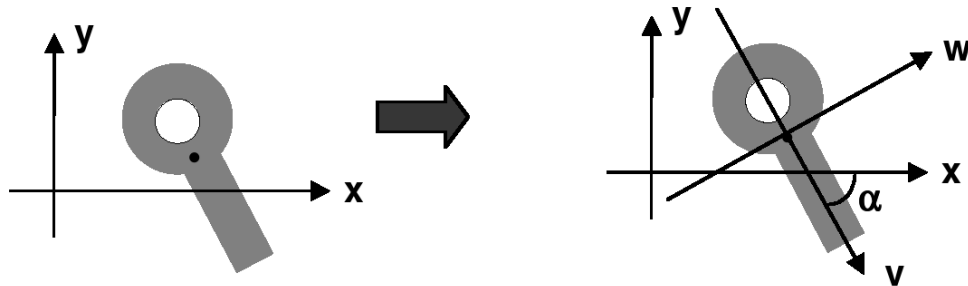


Figura 7.6: Asse ad inerzia minima di un'immagine.

Un altro importante parametro che è possibile valutare attraverso i momenti di secondo ordine, è l'*eccentricità*, attraverso la quale è possibile quantificare la somiglianza di una certa figura ad un cerchio. L'eccentricità è definita come

$$Ec = \frac{1}{Area} \cdot [(M_Y - M_X)^2 + 4 \cdot M_{XY}] \quad (7.14)$$

e assume valori crescenti con il livello di dissomiglianza da un cerchio.

Attraverso questi parametri è possibile effettuare una prima valutazione qua-

litativa di un'immagine binaria, senza bisogno di considerare la sua forma.

Per ogni fase di valutazione è stato fissato un intervallo di validità dei parametri, in modo da poter eliminare dalle fasi di elaborazione successive i bounding box con parametri non appartenenti agli intervalli identificati. Dopo le precedenti considerazioni relative ai momenti geometrici, i parametri che abbiamo ritenuto essere discriminanti nella fase di valutazione sono i seguenti:

1) *Eccentricità*: dall'osservazione di molte sequenze di immagini, abbiamo rilevato che era possibile fissare un intervallo di valori entro il quale l'eccentricità dei pedoni è compresa, consentendo all'algoritmo di scartare piuttosto facilmente molti oggetti appartenenti all'ambiente urbano. Per esempio, i cartelli stradali circolari hanno bassi valori di eccentricità, mentre altri oggetti della segnaletica urbana, come i pali, hanno elevati valori di eccentricità (Figura 7.7).

2) *Inclinazione degli oggetti*: per ogni bounding box, dopo aver localizzato l'asse principale degli oggetti inclusi in esso, si considera l'angolo di inclinazione di questo asse. Come nel caso del parametro precedente, anche l'inclinazione dell'asse permette un'operazione di filtraggio piuttosto semplice di altri oggetti diversi dai pedoni (come gli alberi, per esempio) (Figura 7.8).

3) *Inclinazione delle gambe*: come ultimo parametro da valutare, abbiamo cercato le gambe dei pedoni, limitando le operazioni di valutazione all'area inferiore di ogni bounding box sopravvissuto alle operazioni di filtraggio precedenti, suddivisa in due sotto-aree - di altezza stimata su base empirica - per la fase di individuazione delle gambe (Figura 7.9).

Per ognuna delle due sotto-aree di ricerca delle gambe, sono stati calcolati i valori di inclinazione locali degli assi, tramite il calcolo dei momenti geometrici (Figura 7.10).

Abbiamo così potuto isolare alcuni intervalli di valori di inclinazione caratteristici relativi alle posture assunte da pedoni in movimento (Figura 7.10).

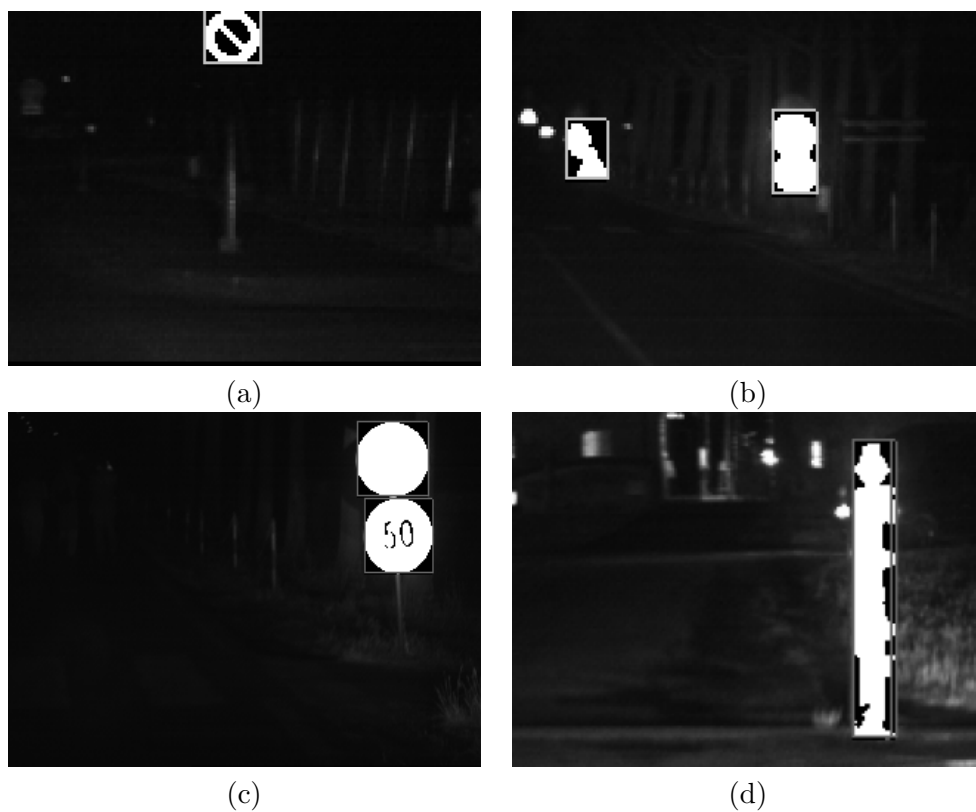


Figura 7.7: Alcuni oggetti eliminati utilizzando il loro valore di eccentricità: nelle immagini (a), (b) e (c) quelli caratterizzati da valori di eccentricità molto bassi, mentre l'oggetto (d) da un valore molto elevato.



Figura 7.8: Alcuni oggetti eliminati dalla stima dell'inclinazione del loro asse principale: dispositivi di illuminazione (a) e cartelli stradali raggruppati (b).



Figura 7.9: Suddivisione del bounding-box per la ricerca delle gambe.

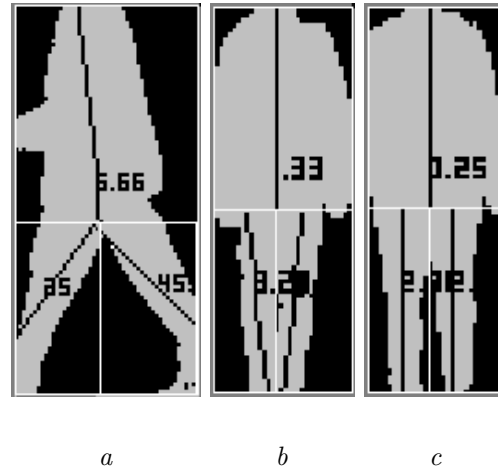


Figura 7.10: Esempi di ricerca delle gambe su pedoni in diverse posture.

7.2.3 Confronto con modelli

Una volta ottenute la struttura di base degli oggetti ed un'indicazione qualitativa sulla postura dei pedoni, è necessario effettuare un controllo più accurato del contenuto dei bounding box. Abbiamo quindi introdotto un'operazione di confronto tra ogni bounding box rimasto ed un gruppo di modelli precalcolati. Nella scelta dell'algoritmo di confronto, abbiamo dedicato molta attenzione ai costi computazionali, trattandosi di un'applicazione real-time. Per diminuire il numero di confronti necessari tra ogni bounding box individuato ed i modelli precalcolati, si è scelto di classificare preventivamente le posture degli oggetti, in base all'angolo assunto dagli assi trovati nelle sotto-aree. In particolare, è stata calcolata la distanza Δ tra gli angoli δ_1 e δ_2 delle due gambe, come rappresentato in Figura 7.11.

I modelli precalcolati vengono suddivisi in 3 possibili classi utilizzando come discriminante l'angolo Δ mentre si trovano in diverse posizioni (in cammino, fermi). Ogni classe è caratterizzata da un intervallo di valori di Δ (Figura 7.12).

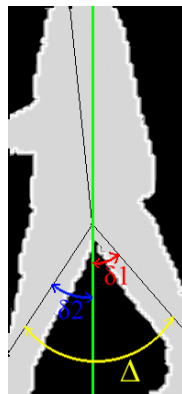


Figura 7.11: Distanza angolare Δ delle gambe.



I^a classe

II^a classe

III^a classe

Figura 7.12: Esempi di bounding box classificati secondo lo schema proposto.

Per la fase di confronto successiva, viene considerato solo il contorno dei pedoni.

Il processo di confronto completo avviene tramite l'applicazione di due metodi differenti, combinando i risultati di questi metodi in una somma pesata.

I° metodo: i modelli per il confronto vengono calcolati come mappe di probabilità dei pixel appartenenti ai contorni dei pedoni, in diverse posture. L'insieme dei modelli si ottiene come segue:

- i pedoni vengono estratti da un insieme di immagini annotate² manualmente;
- i pedoni vengono classificati in tre gruppi, in base al valore dell'angolo Δ formato dalle loro gambe (Figura 7.13);
- vengono estratti i bordi dei pedoni al punto precedente;
- per ogni gruppo, viene calcolato un modello probabilistico [20].

L'algoritmo di confronto funziona nel modo seguente: ad ogni pixel appartenente al contorno dell'oggetto che si sta valutando, viene assegnato il valore del pixel corrispondente nel modello selezionato - quello che rappresenta la classe di pedoni aventi lo stesso angolo Δ . Questi valori vengono sommati: se la somma risultante supera una soglia prefissata, calcolata come percentuale dell'intero contorno dell'oggetto, il confronto restituisce un risultato positivo.

II° metodo: poiché il metodo precedente si è rivelato semplice ma non sempre efficace, abbiamo sviluppato un secondo metodo di confronto. In questo caso - una sorta di estensione del *I° metodo* - consideriamo gli oggetti ed i contorni dei modelli nella loro configurazione in coordinate polari³. L'insieme

²L'annotazione, in visione artificiale, è un procedimento di estrazione manuale di parti di un'immagine con le quali confrontare i risultati ottenuti da un'algoritmo che svolga la stessa operazione di estrazione in modo automatico.

³

$$R = \sqrt{X^2 + Y^2} \quad e \quad \theta = \arctan \frac{Y}{X} \quad (7.15)$$

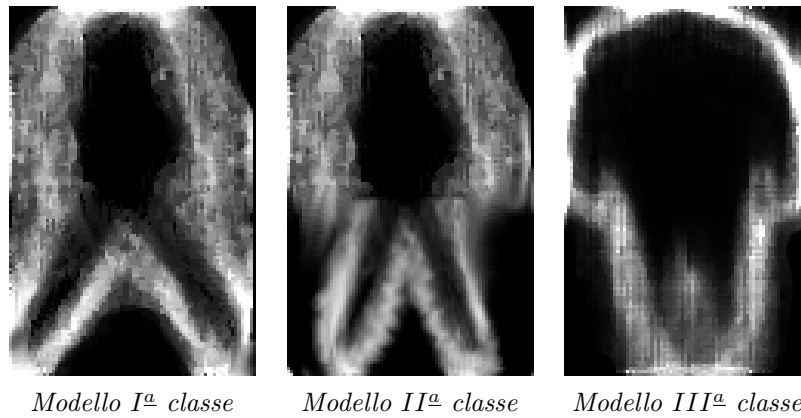
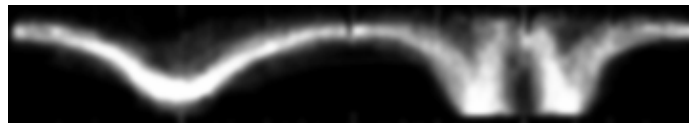


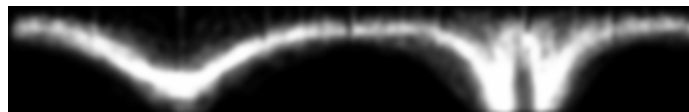
Figura 7.13: Modelli per il confronto.

dei modelli si ottiene utilizzando lo stesso procedimento applicato nel caso del I^o metodo. Si ricavano le classi di modelli di Figura 7.14.

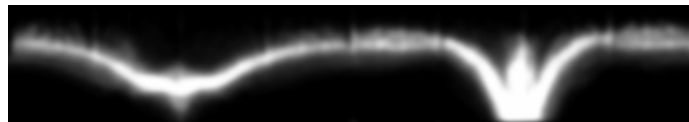
Questo sistema di valutazione si è dimostrato molto efficace in diverse situazioni (Figura 7.15). Funziona infatti come buon metodo complementare al primo; da qui l'idea di utilizzare entrambi i metodi combinati in una somma pesata (Figura 7.16).



Modello polare I^a classe

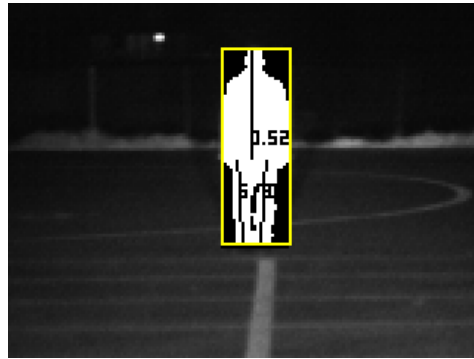


Modello polare II^a classe

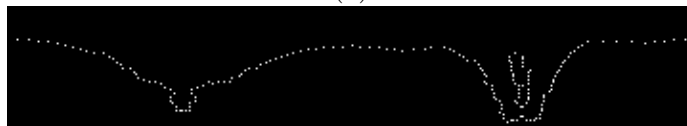


Modello polare III^a classe

Figura 7.14: Modelli per il confronto in coordinate polari.



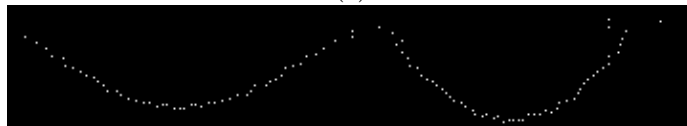
(a)



(b)



(c)



(d)

Figura 7.15: Diagrammi polari del contorno degli oggetti contenuti nei bounding box trovati: (b) del contorno del pedone in (a); (d) del segnale stradale in (c). L'asse delle ascisse rappresenta angoli compresi tra 0° e 360° , mentre l'asse delle ordinate rappresenta i valori del raggio - per un valore massimo pari a 60 pixel.

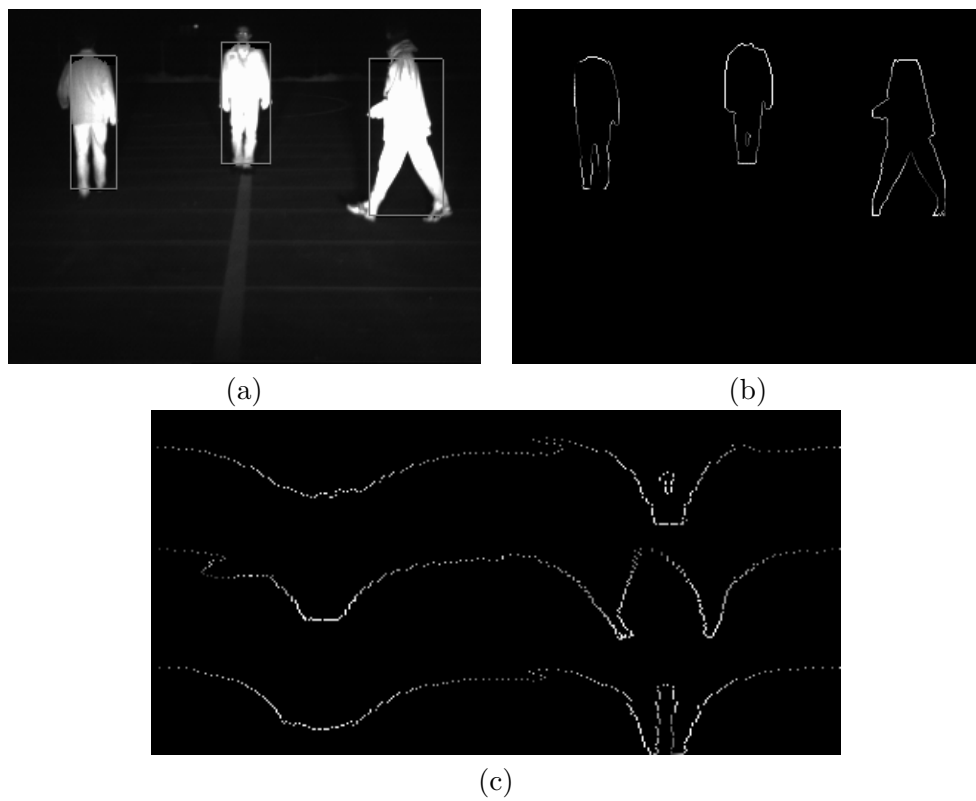


Figura 7.16: Risultati del confronto con entrambi i metodi: I^o (b) e II^o (c). Più chiari sono i pixel nelle immagini (b) e (c), migliori sono i risultati del confronto.

Capitolo 8

Risultati sperimentali

Con questa sperimentazione abbiamo valutato un sistema di visione NIR per il rilevamento di pedoni in scenari notturni da un veicolo in movimento, utilizzando come sorgenti di illuminazione i fari abbaglianti del veicolo stesso. Nel corso dell'analisi del problema, abbiamo dedicato molta attenzione all'individuazione delle caratteristiche più distintive delle immagini NIR. L'algoritmo sviluppato evidenzia innanzitutto le zone chiare delle immagini, circondandole in seguito con bounding box in modo da ridurre la dimensione delle aree da valutare che passeranno alla successiva fase di valutazione. Quindi valuta il contenuto di ogni bounding box con una somma pesata di due metodi di confronto. Le migliori prestazioni dell'algoritmo sono state ottenute nei casi di buone condizioni luminose ed in situazioni in cui l'illuminazione dovuta all'ambiente circostante o al traffico non hanno creato disturbi nelle immagini acquisite. I tempi di computazione richiesti dall'algoritmo dipendono fortemente dal numero di bounding box rilevati nell'immagine, determinando il numero di comparazioni necessarie nella fase di confronto. Il tempo medio richiesto è di circa 20 - 25 ms su di un processore AMD Athlon XP2200+ @ 1.8 GHz. Per misurare le prestazioni del sistema sviluppato, abbiamo confrontato i bounding box validi ottenuti dall'algoritmo con quelli tracciati da

un operatore umano tramite l'annotazione manuale delle sequenze acquisite. Poiché i pedoni vengono localizzati tramite bounding box, è stato necessario valutare la percentuale di sovrapposizione dei bounding box generati dall'algoritmo con quelli annotati manualmente. Attraverso questo confronto è stato possibile rilevare in modo oggettivo le prestazioni dell'algoritmo. Per quanto riguarda i rilevamenti corretti, le migliori prestazioni sono state ottenute in immagini dove fosse alto il contrasto tra pedoni e sfondo (Figure 8.1 e 8.2), con percentuali di riconoscimento attorno al 90%. In immagini molto rumorose invece, con molte sorgenti di luce oltre ai fari del veicolo, la qualità dei risultati decade significativamente fino ad ottenere, come le più scarse percentuali di riconoscimento, cifre attorno al 30%.



Figura 8.1: Immagine originale (a) e rispettivo risultato dell'elaborazione (b).

Altri risultati sono mostrati in Figura 8.3: i bounding box correttamente rilevati sono quelli di colore verde. La Figura 8.3.b riporta il risultato di rilevamento in un caso semplice, dove i pedoni risultano in forte contrasto rispetto allo sfondo. In Figura 8.3.d è riportata invece una situazione più complessa, con molti elementi di disturbo come cartelli stradali ed alberi. L'algoritmo rileva correttamente solo uno dei due pedoni presenti nella scena, poiché uno

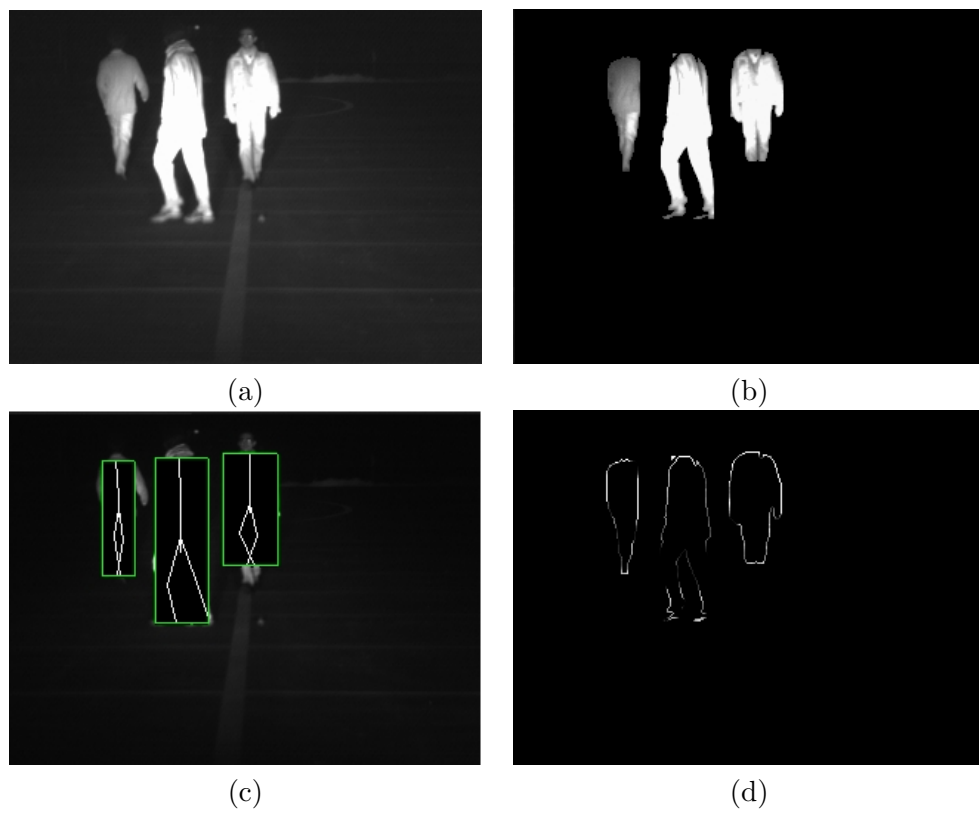


Figura 8.2: Immagine originale (a) e rispettivo risultato dell'elaborazione (b); in (c) sono raffigurati gli assi di minima inerzia delle sagome all'interno dei bounding box, mentre in (d) sono tracciati i contorni dei pedoni individuati.

dei due porta con se una borsa che ne altera la sagoma, invalidando le fasi di confronto con i modelli. La Figura 8.3.f mostra una situazione molto complessa: viene effettuata la rilevazione corretta di uno solo dei due pedoni. I diffusi riflessi luminosi dello scenario urbano interferiscono con quello prodotto dai pedoni, alterandone la sagoma che pertanto, come nel caso di Figura 8.3.d, non supera le fasi di confronto.

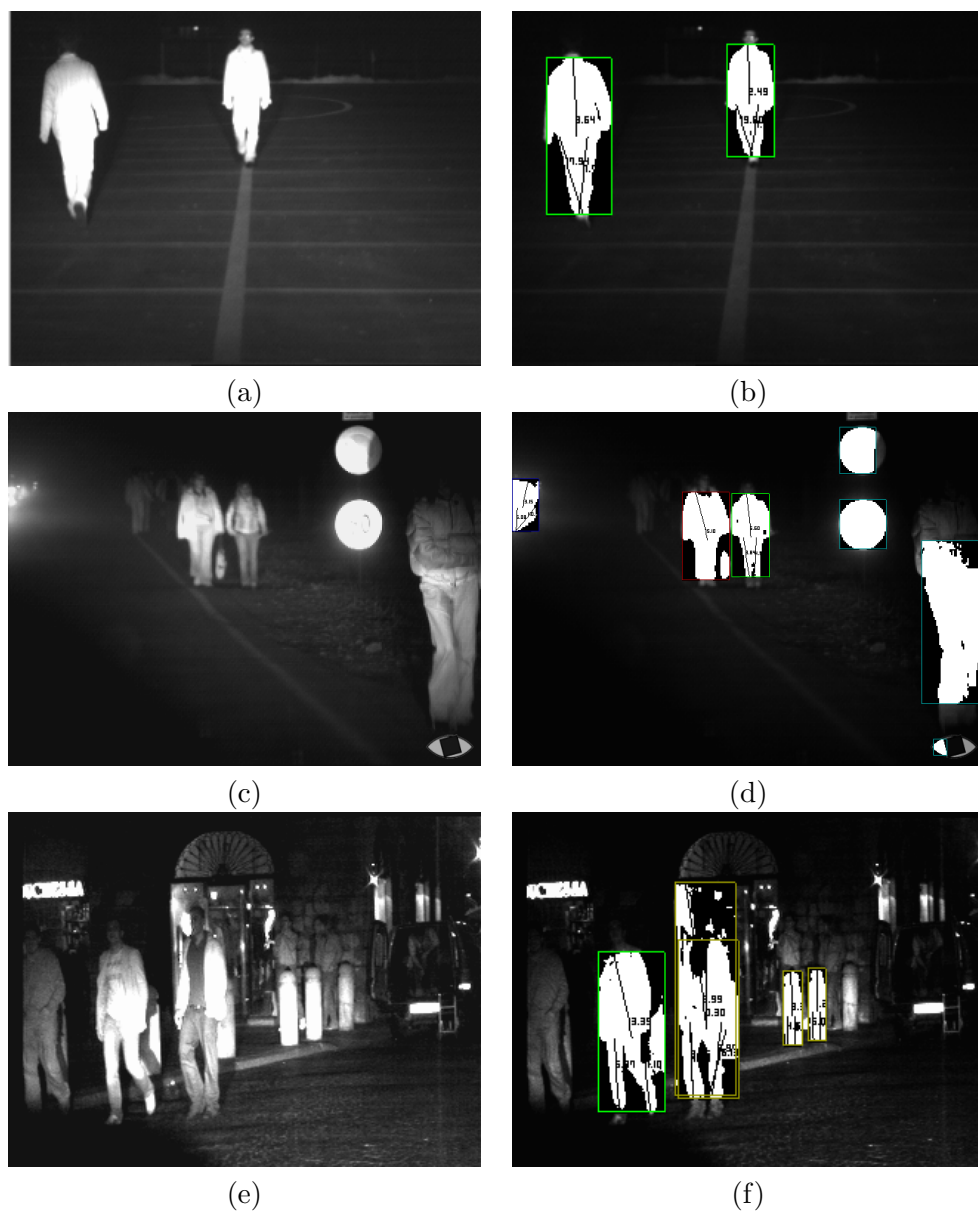


Figura 8.3: Immagini originali affiancate dai relativi risultati. Gli oggetti rilevati come pedoni vengono indicati dal bounding-box di colore verde.

Parte III

Individuazione di oggetti in galleggiamento sulla rotta di imbarcazioni autonome

Capitolo 9

Introduzione

In questa sperimentazione è stato sviluppato un sistema per il rilevamento, mediante tecniche di visione artificiale, di imbarcazioni lontane, natanti ed altri ostacoli galleggianti potenzialmente in grado di intralciare la rotta dell'imbarcazione autonoma recante a bordo il sistema stesso.

In ambito nautico la navigazione di imbarcazioni comandate a distanza da un operatore viene utilizzata per l'espletamento di operazioni ritenute pericolose per gli uomini, come lo scandaglio di zone per la localizzazione di mine inesplose ed il successivo brillamento di queste, per l'esplorazione e la videosorveglianza delle zone costiere o portuali. L'impiego del solo operatore per il controllo visuale delle immagini risulta spesso inadeguato: brevi distrazioni possono comportare l'incapacità di far virare l'imbarcazione, comandata a distanza, in tempo per evitare ostacoli galleggianti o altre imbarcazioni comparse all'improvviso sulla sua rotta. Un sistema di visione installato a bordo dell'imbarcazione che le permettesse di localizzare ostacoli lungo la sua rotta, potrebbe coadiuvare in modo significativo l'operatore umano nella sua attività di monitoraggio.

In letteratura non abbiamo trovato lavori che riportino questo genere di indagini relative a imbarcazioni autonome né in navigazione controllata a distanza.

Capitolo 10

Considerazioni preliminari

Durante la prima sessione di acquisizioni è stato installato a bordo del motoscafo un sistema per l'acquisizione di immagini composto da 4 telecamere installate, tramite degli snodi regolabili che permettessero di variare l'orientazione di ogni telecamera, su una barra a sua volta fissata sul tetto della cabina del motoscafo (Figura 10.1). La disposizione delle 4 telecamere sulla barra ha permesso di acquisire simultaneamente 4 immagini della stessa scena reale con 2 sensori differenti (infrarosso termico/FIR e visibile) e 2 modalità: monoscopica e stereoscopica. Le acquisizioni sono avvenute sia in condizioni di luminosità diurne (Figura 10.2), che in condizioni notturne (Figura 10.3), per valutare anche un eventuale utilizzo del sistema durante la notte.

Dalla fase di studio e analisi dei flussi di immagini abbiamo rilevato che delle 4 telecamere utilizzate nelle sessioni di acquisizione la più idonea all'applicazione risultava essere la telecamera termica singola. Le altre telecamere FIR hanno dimostrato di avere minore sensitività termica¹ ed una più ridotta risoluzione spaziale (Figura 10.2), mentre la telecamera con sensore nello spettro del visibile non si è dimostrata utile per la prima fase di sviluppo dell'algoritmo.

¹Capacità di risposta ad uno stimolo fisico - la temperatura in questo caso - oppure di rilevarne piccole quantità o variazioni.



(a)



(b)

Figura 10.1: La barra con le 4 telecamere installate (a) ed il motoscafo con la barra fissata sul tetto della cabina (b).

Le informazioni sul colore infatti, che si riteneva in un primo momento potessero essere utili nell'individuazione, si sono dimostrate invece insufficienti per effettuare una robusta discriminazione degli ostacoli dalla superficie del mare. Oltre alla variabilità intrinseca del colore della superficie del mare, verniciature di imbarcazioni dai colori ad esso molto simili ed il parziale occultamento di ostacoli dovuto alle onde, sono state le osservazioni che hanno portato ad abbandonare l'impiego di questo sensore. Il contenuto informativo delle immagini acquisite nello spettro del visibile potrebbe tuttavia risultare vantaggioso per eventuali raffinamenti dei risultati ottenuti dopo l'elaborazione delle immagini appartenenti all'infrarosso termico. Abbiamo scelto di utilizzare un sistema monoscopico e non uno stereoscopico per l'estrema difficoltà che avrebbe comportato la calibrazione di quest'ultimo, a causa del costante movimento dell'imbarcazione nell'acqua. Per quanto riguarda le condizioni di luminosità invece, pur ritenendo la telecamera scelta la più idonea tra quelle valutate, abbiamo riscontrato che in condizioni notturne il contrasto termico tra la superficie del mare e gli oggetti immersi in esso risultava molto meno significativo rispetto a quello ottenuto in condizioni diurne, fino a rendere quasi impercettibile anche la distinzione tra il mare e la zona costiera, in certi casi (Figura 10.3).

Nella sua configurazione finale, il sistema risulta quindi composto da un'unica telecamera termica fissata su una barra installata sul tetto della cabina del motoscafo. (Figura 10.5).

Dopo l'analisi in laboratorio delle sequenze acquisite sono stati valutati diversi scenari. Quello ritenuto come più promettente ai fini dell'efficacia dell'algoritmo prevede il rilevamento di imbarcazioni vicine (entro i 500 m) e oggetti molto vicini semi-sommersi che si trovino nella zona di fronte all'imbarcazione in mare aperto o zone costiere, in condizioni luminose diurne, visibilità buona o ottima e mare calmo o poco mosso.

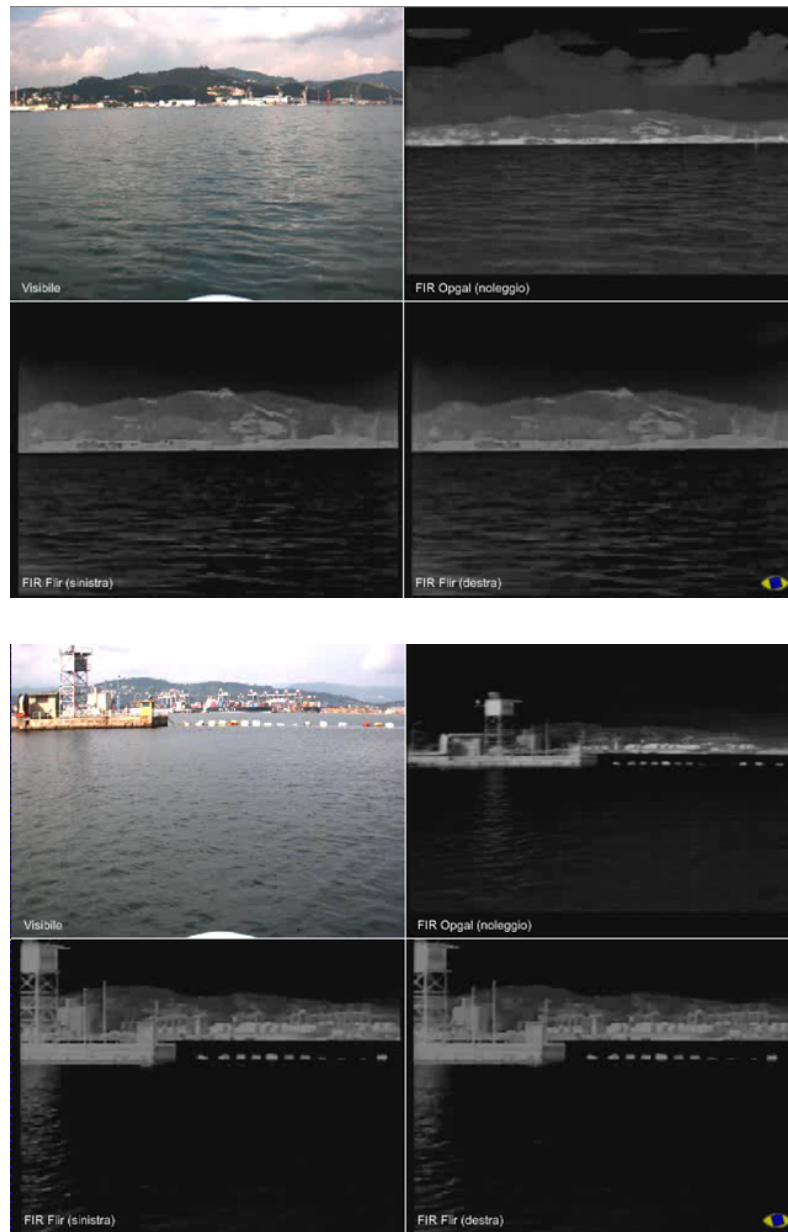


Figura 10.2: Esempi di immagini acquisite di giorno dal sistema a 4 telecamere.

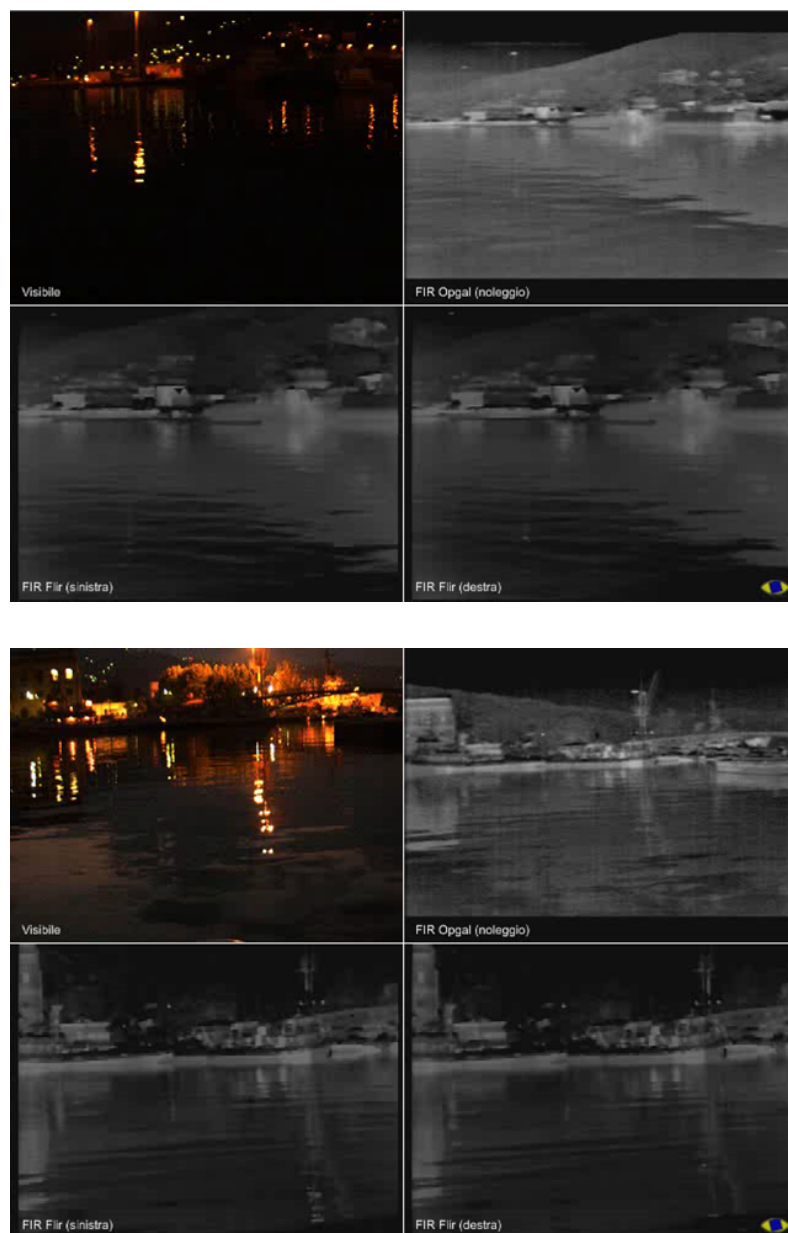


Figura 10.3: Esempi di immagini acquisite di notte dal sistema a 4 telecamere.



Figura 10.4: La telecamera termica installata sulla cabina dell'imbarcazione.



Figura 10.5: Un'immagine d'esempio rilevata dalla telecamera.

Capitolo 11

Procedimento

11.1 In breve

L'algoritmo sviluppato sfrutta le immagini acquisite dalla telecamera termica nelle quali gli ostacoli (dove per ostacolo si intende qualsiasi oggetto attorniato dall'acqua) sono riconducibili a raggruppamenti di punti caldi, isolati. Dopo una fase preliminare di calibrazione della telecamera, viene effettuata una ricerca delle caratteristiche distintive della superficie del mare. Si passa quindi alla ricerca della linea dell'orizzonte, intesa come separazione tra mare e cielo, in caso di acquisizione di immagini in mare aperto, come separazione tra mare e terra invece, in caso di scenari portuali o costieri. La ricerca degli ostacoli avviene considerando tali i *cluster* (raggruppamenti) di pixel sottostanti la linea di orizzonte e con caratteristiche diverse da quelle della superficie del mare. Per stabilizzare i risultati è stato introdotto un algoritmo di tracking degli ostacoli. I dati relativi ad ogni ostacolo vengono in seguito inviati a terra via rete per essere conservati in una banca dati o visualizzati da un operatore su mappa nautica virtuale.

11.2 In dettaglio

11.2.1 Calibrazione della telecamera

Il sistema di rilevamento sviluppato prevede una prima fase di calibrazione della telecamera, necessaria per la ricostruzione della corrispondenza tra i valori delle distanze degli ostacoli dalla telecamera a bordo dell'imbarcazione ottenuti nel sistema di riferimento dell'immagine con quelli finali appartenenti al sistema di riferimento del mondo reale (Figura 11.1).

La calibrazione avviene impiegando il metodo della Inverse Perspective Mapping ed è estremamente importante che le operazioni di rilevamento ed impostazione dei parametri di calibrazione avvengano in modo accurato poiché da queste dipende la precisione delle distanze degli ostacoli dall'imbarcazione nel mondo reale.

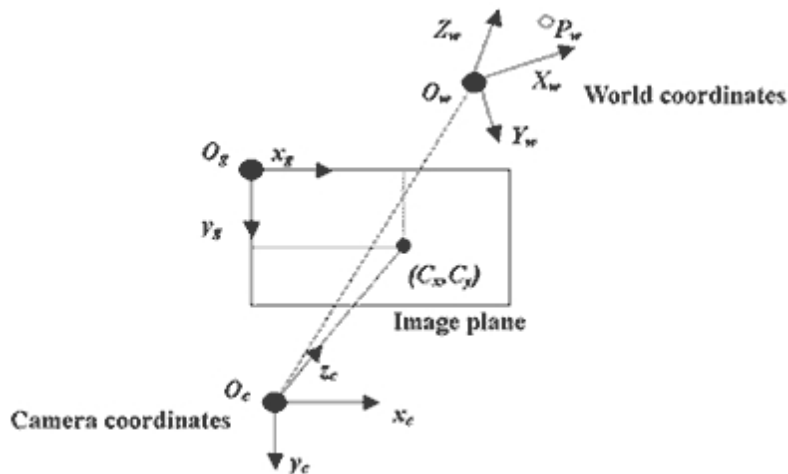


Figura 11.1: I 3 sistemi di riferimento coinvolti nella calibrazione: quello della telecamera, quello dell'immagine e quello del mondo esterno.

Oltre ai parametri di calibrazione citati, devono essere considerati anche i 3 an-

goli di orientazione della telecamera nello spazio: imbardata (*Yaw*), beccheggio (*Pitch*) e rollio (*Roll*) (Figura 11.2).

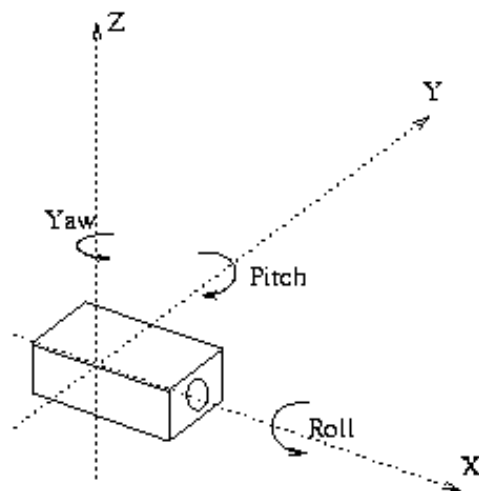


Figura 11.2: Rappresentazione dei 3 parametri di orientazione.

11.2.2 Ricerca del mare

Dopo le operazioni di calibrazione si passa a quella di caratterizzazione del mare. Viene considerata un'area rettangolare nella parte inferiore dell'immagine totalmente occupata dalla superficie del mare e senza presenza di ostacoli (Figura 11.4). Relativamente a quest'area vengono calcolati alcuni parametri che possano rendere distintiva la superficie del mare da altri ostacoli immersi in questa come il valor medio, la varianza ed il contrasto di colore dei pixel (Figure 11.3.c, 11.3.d e 11.3.e). Si passa quindi a binarizzare l'immagine utilizzando come soglia il valor medio del mare ricavato in precedenza. Questa operazione permette di mantenere solo le porzioni dell'immagine di colore distinto da quello del mare, come gli oggetti in galleggiamento e, se presente, la zona costiera (Figura 11.3.b).

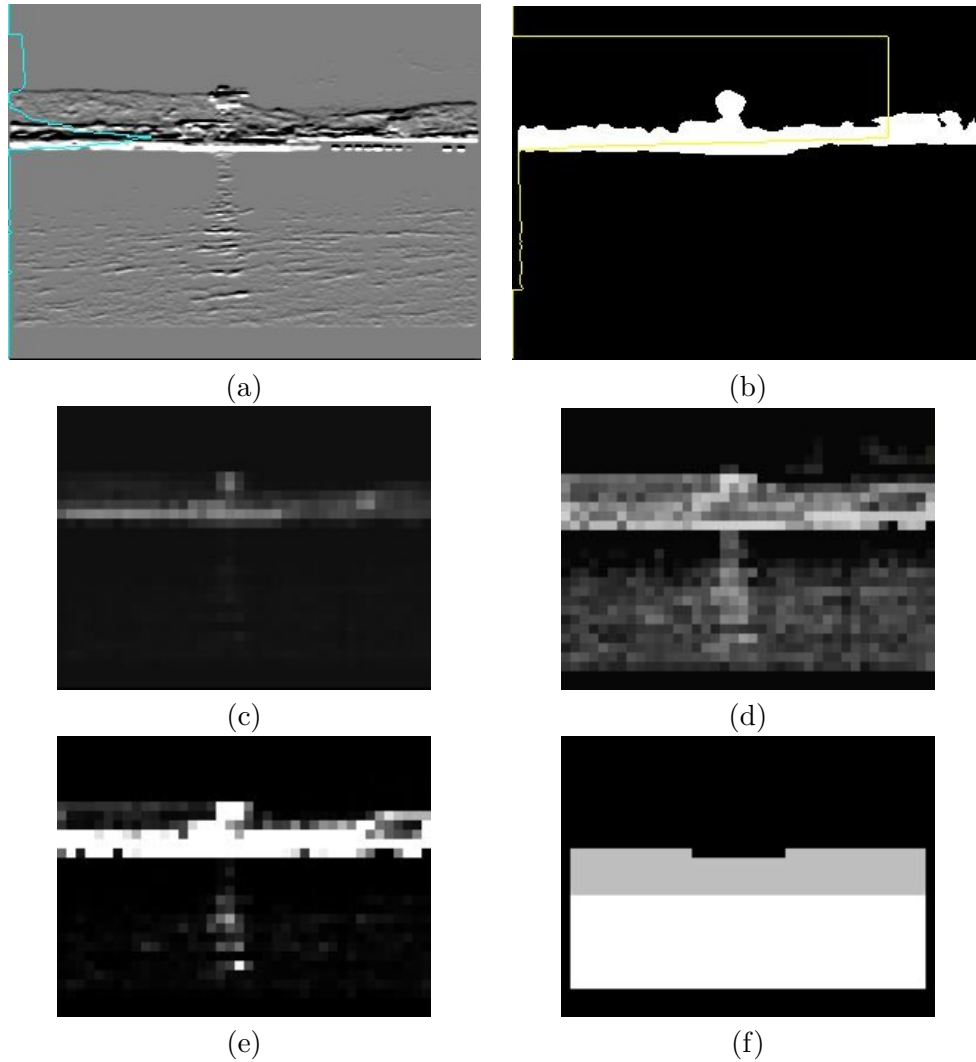


Figura 11.3: Le fasi dell'algoritmo: un istogramma dei valori dei pixel per ogni riga tracciato in azzurro sull'immagine dei bordi orizzontali (a) e lo stesso tipo di istogramma tracciato sull'immagine binarizzata dopo l'espansione (b); seguono le visualizzazioni dei parametri di caratterizzazione del mare: valor medio (c), contrasto (d), varianza (e) ed infine la rappresentazione delle zone espanse di mare e tutto ciò che supera la linea di orizzonte (f): la zona rimanente (in nero) rappresenta l'ostacolo trovato.

11.2.3 Ricerca della linea di orizzonte

Analogamente al metodo di ricerca del mare, la zona di ricerca dell'orizzonte viene delimitata da un'area rettangolare (Figura 11.4). La definizione di orizzonte prevederebbe di considerare come tale la linea apparente di separazione tra cielo e terra o tra cielo e mare. Ai fini del nostro algoritmo invece abbiamo indicato con questo termine la linea che separa la superficie del mare dal cielo, in caso di riprese in mare aperto, o dalla costa, in caso di navigazione verso terra (Figura 11.3.f). Tuttavia viene anche effettuata una stima della posizione dell'orizzonte secondo la sua definizione corretta, utilizzando i parametri di calibrazione iniziali della telecamera. Per il continuo movimento dell'imbarcazione - soprattutto quello di beccheggio - a causa del moto ondoso però la linea ottenuta con i parametri di calibrazione risulta spesso inesatta e perciò poco affidabile, come vedremo in seguito, per la stima della distanza degli ostacoli dall'imbarcazione. L'immagine binarizzata, ottenuta al passo precedente, viene scorsa per colonne all'interno del rettangolo di ricerca. Ogni colonna di pixel il cui numero superi una soglia prefissata, viene conservata in un vettore sul quale si applica in seguito il metodo dei minimi quadrati per ottenere la retta orizzontale che meglio approssimi le colonne di punti conservate nel vettore stesso. Sulla retta orizzontale ottenuta viene quindi applicato un fattore correttivo utilizzando il valore dell'ordinata del punto centrale della retta stessa.

11.2.4 Ricerca degli ostacoli

Una volta delimitata l'area di ricerca degli ostacoli alla zona sottostante l'orizzonte (viene ritenuto ostacolo qualsiasi oggetto attorniato dall'acqua), si procede alla loro localizzazione. Verranno rilevati come ostacoli tutti i cluster di pixel caratterizzati da posizione sottostante la linea di orizzonte e parametri di colore diversi da quelli che caratterizzano la superficie del mare trovati in precedenza (Figura 11.5).

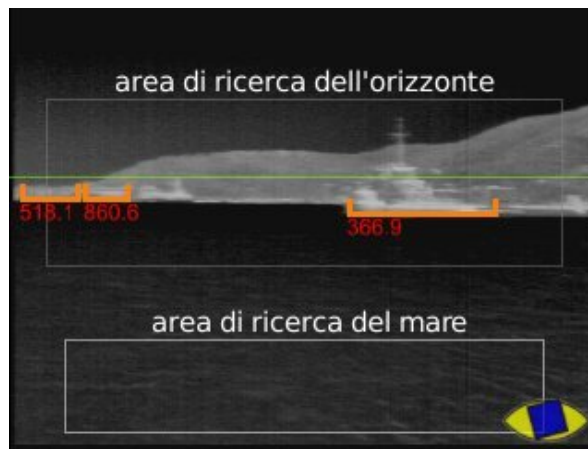


Figura 11.4: In evidenza le aree di ricerca del mare e dell'orizzonte. La linea verde rappresenta l'orizzonte stimato dai parametri di calibrazione e non risulta coincidere (correttamente, secondo la definizione corretta di orizzonte) con quella di separazione tra mare e zona costiera. Le spezzate arancioni a forma di “u” delimitano gli ostacoli rilevati; sotto vengono indicate in rosso le distanze in metri di questi dall'imbarcazione - stimate utilizzando i parametri di calibrazione iniziali della telecamera.

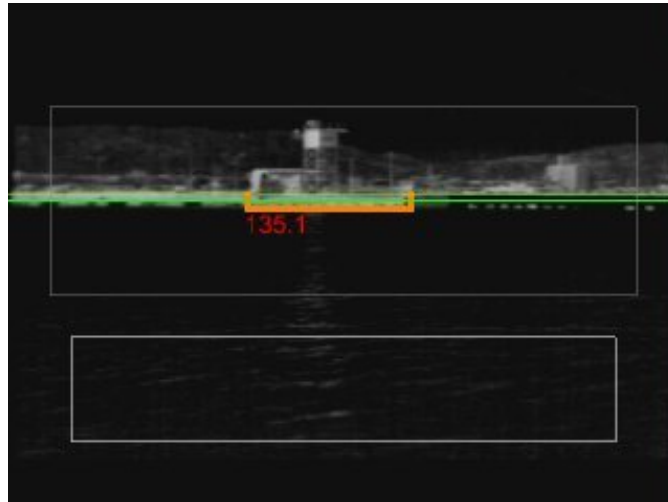


Figura 11.5: Risultato finale relativo alle immagini di Figura 11.3: l'immagine originale con gli ostacoli rilevati.

11.2.5 Tracking degli ostacoli

La fase di localizzazione degli ostacoli viene completata da una di tracking degli stessi allo scopo di irrobustire e stabilizzare il rilevamento. Per la presenza del moto ondoso infatti, si può verificare spesso che ostacoli in galleggiamento, anche se ancorati ai fondali e non in moto come possono essere altre imbarcazioni, vengano periodicamente occultati dall'acqua. Senza il contributo del tracking dunque la rilevazione degli ostacoli risulterebbe molto intermittente, e pertanto fastidiosa sia per un operatore che dovesse controllare su display le immagini acquisite dall'imbarcazione, sia per una rappresentazione su mappa nautica virtuale. L'output ottenuto è consiste di un'immagine avente come sfondo l'immagine originale acquisita dalla telecamera e come risultato dell'elaborazione il tracciamento in colore arancione delle zone occupate dagli ostacoli rilevati insieme ad una stima della loro distanza dall'imbarcazione (Figura 11.5).

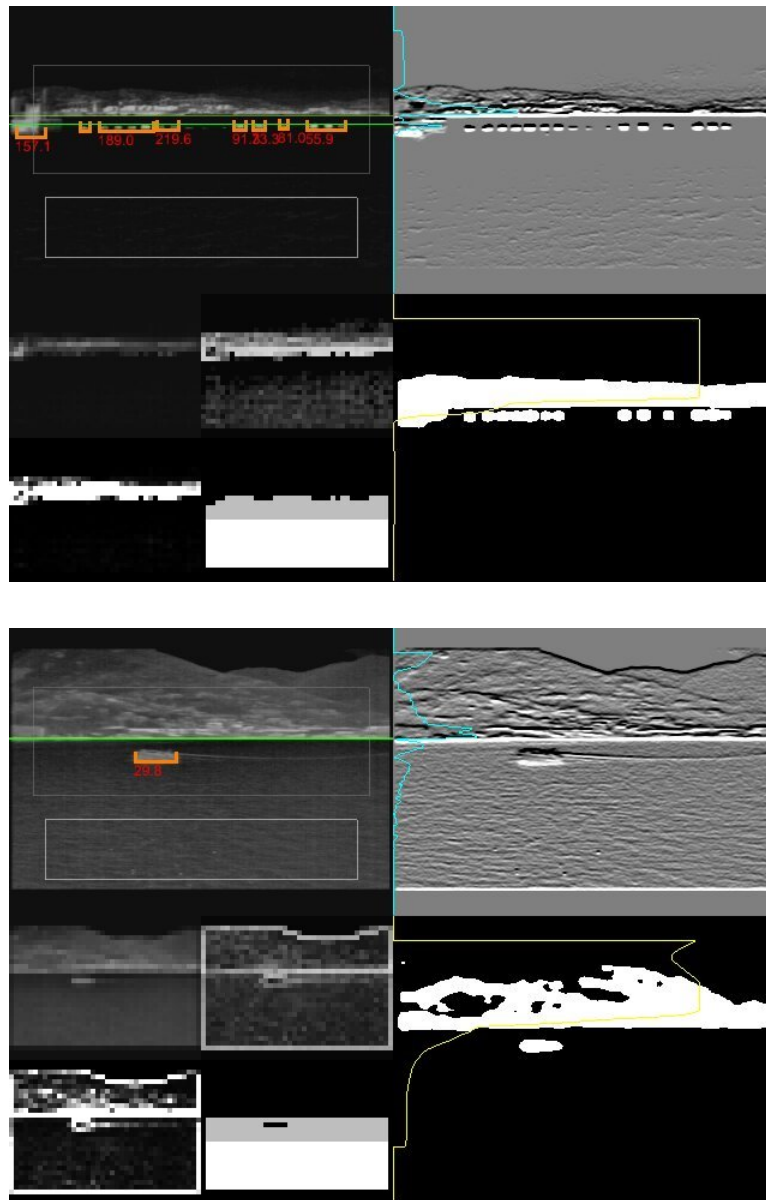


Figura 11.6: Altri esempi di risultati con le relative fasi intermedie del procedimento.

11.2.6 Invio dei risultati via rete

L'algoritmo prevede la possibilità di inviare via rete i dati relativi alla localizzazione degli ostacoli (in termini di: numero di frame in cui ogni ostacolo è comparso - la storia dell'ostacolo, intervallo di distanze in metri a cui si trova rispetto alla telecamera, lunghezza in radianti rispetto all'asse ottico della telecamera con angoli positivi in senso antiorario) (Figura 11.7).

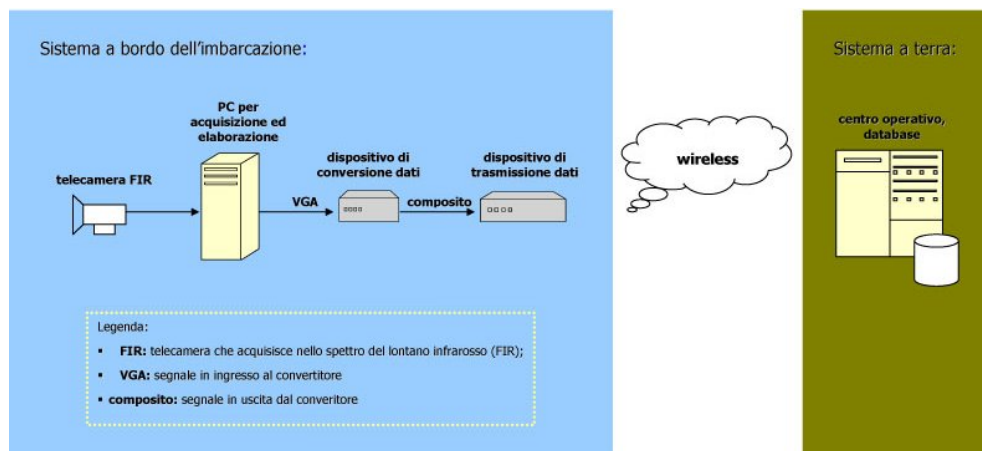


Figura 11.7: Lo schema architetturale per l'invio dei dati dall'imbarcazione ad una postazione di terra.

Capitolo 12

Risultati sperimentali

Il sistema descritto è stato realizzato per valutare la possibilità di dotare un'imbarcazione comandata a distanza di rilevamento automatico, mediante visione artificiale, degli eventuali ostacoli che ne possano intralciare la rotta. Dopo una calibrazione manuale, viene svolta la fase di individuazione degli ostacoli. Per il movimento di beccheggio dell'imbarcazione e l'uso di un sistema di visione monoscopico, la stima delle distanze degli ostacoli dall'imbarcazione rimane inevitabilmente falsata da errore. Le immagini utilizzate sono state acquisite in condizioni luminose diurne, in diversi orari, con diverse condizioni di temperatura esterna e mare tendenzialmente calmo o poco mosso. Il sistema si è dimostrato efficace soprattutto per il rilevamento di ostacoli in mare aperto. Alcuni problemi si sono riscontrati invece nelle zone costiere: in queste situazioni risulta difficile riuscire a rilevare correttamente la linea di orizzonte - separazione tra mare e terra, in questi casi - e di conseguenza tutta la fase successiva di ricerca degli ostacoli che si trovino sotto tale linea diventa più complessa, dando esiti non sempre precisi. Il sistema di rilevamento può eseguire l'intera computazione in 20 ms su un processore Pentium IV HT @ 3.0 GHz.

Nelle figure seguenti vengono riportati alcuni dei risultati ottenuti.

In Figura 12.1.a è rappresentata una situazione di rilevamento semplice: quello di una piattaforma galleggiante in mare aperto, con condizioni di mare calmo, ottimali.

La Figura 12.1.b riporta invece il rilevamento, in mare aperto con condizioni di mare poco mosso, di più ostacoli: navi dislocate a varie distanze dall'imbarcazione.

La Figura 12.1.c mostra una situazione di rilevamento più complessa: ostacoli fissi - una banchina di attracco portuale - e mobili - imbarcazioni - in corrispondenza di un'area costiera con condizioni di mare mosso. L'algoritmo è comunque in grado di rilevare tutti gli ostacoli presenti.

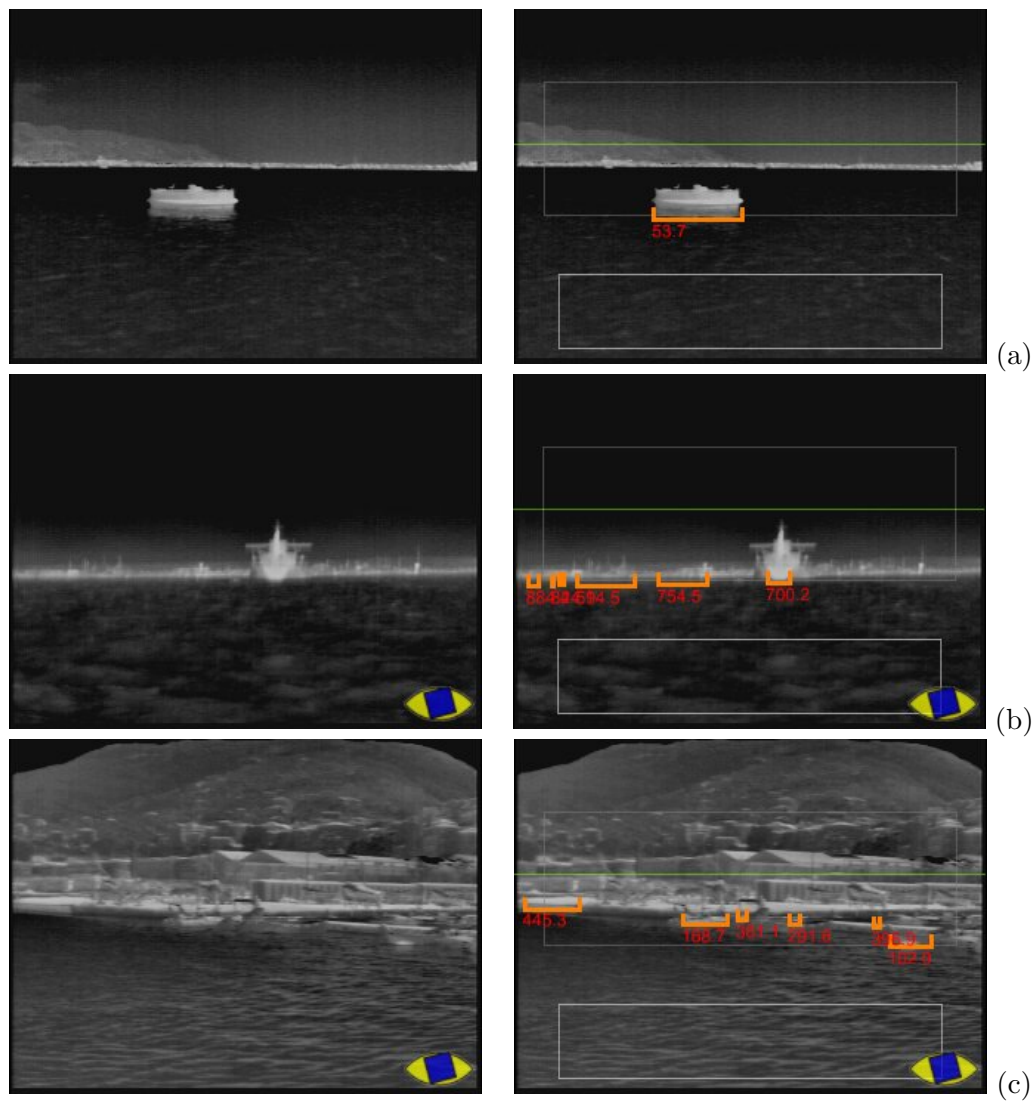


Figura 12.1: Alcune immagini originali affiancate dai relativi risultati.

Conclusioni

Domini spettrali: visibile, NIR o FIR?

Abbiamo valutato l'applicazione di alcune tecniche di visione artificiale ai domini spettrali del visibile, vicino e lontano infrarosso. Benché totalmente differenti tra loro le sperimentazioni ed i rispettivi scopi, ogni applicazione è stata caratterizzata da un approccio, in termini della scelta dei sensori impiegati per affrontare il problema, diverso rispetto a quelli trovati in letteratura. Scopo delle indagini era quello di cercare l'esistenza di un dominio spettrale le cui immagini potessero avere un contributo informativo di più semplice interpretazione tramite l'applicazione di tecniche di visione artificiale rispetto a quelli già impiegati in letteratura per la soluzione di problemi analoghi. Certo l'indagine non può dirsi esaustiva, ma già questi primi risultati danno indicazioni consistenti verso la necessità di utilizzare un complesso di sensori, che abbraccino domini spettrali differenti, in modo da espandere il più possibile il quantitativo di informazioni percepite dal sistema di elaborazione in ogni istante di tempo. A fronte di un carico computazionale indubbiamente maggiore, questa potrebbe essere la direzione futura da prendere per cercare di ottenere risultati di precisione e robustezza paragonabili a quelli acquisiti ed elaborati dal sistema occhio-cervello umano.

O almeno questa ci sembrava la conclusione più idonea, a partire dai risultati della prima sperimentazione. Tuttavia, non era convincente. In letteratura

si possono reperire molti lavori relativi a sperimentazioni di visione artificiale dove vengono impiegati contemporaneamente molti sensori, come sistemi di una o più telecamere operanti in diversi domini spettrali anche affiancate da altri sensori come radar, lidar e laserscanners. Gli ottimi risultati presentati in questi lavori tuttavia rimangono ancora lontani dall'emulare le caratteristiche di precisione e robustezza di interpretazione visiva proprie del cervello umano. Le correnti teoriche più recenti relative al funzionamento del sistema visivo spiegano però che diverse proprietà del mondo visuale, come il movimento, la forma ed il colore, richiederebbero soluzioni computazionali diverse che vengono anatomicamente e funzionalmente implementate in regioni diverse del cervello, operanti indipendentemente. Non solo, ma la percezione di proprietà diverse avverrebbe in modo asincrono [32]. Le informazioni ottenute da questi moduli funzionali indipendenti verrebbero poi convogliate insieme generando l'interpretazione finale della scena osservata [33]. Da queste nuove ipotesi sulla percezione visiva nel cervello è possibile prendere spunto per valutare configurazioni diverse per sistemi di visione artificiale. Sarebbe interessante valutare sistemi composti da più elaboratori paralleli, dedicati ognuno a compiere analisi specializzate sulle immagini acquisite dai sensori di visione. Si potrebbe per esempio considerare un sistema composto di una telecamera e tre elaboratori: uno dedicato alla percezione del movimento, uno all'analisi delle forme ed uno a quella del colore, per poi ricomporre il contenuto informativo dell'immagine una volta ottenuti tutti i risultati delle elaborazioni separate, un po' come sembra avvenire nella corteccia cerebrale primaria del nostro cervello. Il carico computazionale ripartito tra diversi elaboratori dovrebbe garantire che il processo interpretativo si esaurisca entro tempi brevi. Così, anziché una moltiplicazione dei sensori collegati ad un solo elaboratore, potrebbe essere interessante valutare l'inversione della complessità del sistema visivo artificiale, utilizzando un solo tipo di sensore, come avviene nella visione umana, e più elaboratori che operino in modo parallelo e asincrono, per l'interpretazione dell'immagine acquisita.

Ringraziamenti

Ai miei compagni di avventura Stefano Ghidoni e Claudio Caraffi, per la condivisione di estenuanti ed esilaranti ore di redazione in compagnia: vi ringrazio!
E vi auguro una grande, sconfinata fortuna per il futuro...

Bibliografia

- [1] A.L. Yarbus. *Eye Movements and Vision*. New York, USA, 1967.
- [2] D.H. Hubel. *Eye, Vision and Brain*. New York, USA, 1988.
- [3] C. Curio, J. Edelbrunner, T. Kalinke, C. Tzomakas, and W. von Seelen. Walking pedestrian recognition. In *Procs. of the 2000 IEEE International Conference on Intelligent Transportation Systems*, pages 155–162, September 2000.
- [4] F. Bu and C.Y. Chan. Pedestrian detection in transit bus application: sensing technologies and safety solutions. In *Procs. of the 2004 IEEE Intelligent Vehicles Symposium*, pages 100–105, June 2004.
- [5] P. Viola, M. Jones, and D. Snow. Detecting pedestrians using patterns of motion and appearance. In *Procs. of the 2003 IEEE International Conference on Computer Vision*, pages 734–741, October 2003.
- [6] H. Elzein, S. Lakshmanan, and P. Watta. A motion and shape-based pedestrian detection algorithm. In *Procs. of the 2003 IEEE Intelligent Vehicles Symposium*, pages 500–504, June 2003.
- [7] M. Bertozzi, A. Broggi, R. Chapuis, F. Chausse, A. Fascioli, and A. Tibaldi. Shape-based pedestrian detection and localization. In *Procs. of the 2003 IEEE International Conference on Intelligent Transportation Systems*, pages 328–333, October 2003.

-
- [8] M.K. Pietikainen. *Texture Analysis in Machine Vision*. London, UK, 2000.
 - [9] F.N. McLeod, N.B. Hounsell, and B. Rajbhandari. Improving Traffic Signal Control for Pedestrians. In *Procs. of the 2007 IEEE International Conference on Road Transport Information and Control*, pages 268–277, April 2004.
 - [10] H. Vceraraghavan, O. Masoud, and N. Papanikolopoulos. Vision-based monitoring of intersections. In *Procs. of the 2007 IEEE International Conference on Intelligent Transportation Systems*, pages 7–12, October 2002.
 - [11] V. Bhuvaneshwar and P.B. Mirchandani. Real-time detection of crossing pedestrians for traffic-adaptive signal control. In *Procs. of the 2004 IEEE International Conference on Intelligent Transportation Systems*, pages 309–313, October 2004.
 - [12] C. Conde, L.J. Rodríguez-Aragón, Á. Serrano, J. Pérez, and E. Cabello. An Experimental Approach to a Real-Time Controlled Traffic Light Multi-Agent Application. In *Procs. of the 2007 International Conference on Autonomous Agents and Multi Agent Systems*, July 2004.
 - [13] P. Vannoorenberghe, C. Motamed, J.M. Blosseville, , and J.G. Postaire. Monitoring pedestrians in a uncontrolled urban environment by matching low-level features. In *Procs. of the 2007 IEEE International Conference on Systems, Man and Cybernetics*, pages 2259–2264, October 2006.
 - [14] A. Prati, I. Mikic, M.M. Trivedi, and R. Cucchiara. Detecting moving shadows: algorithms and evaluation. 25(7):918–923, July 2003.
 - [15] Massimo Bertozzi, Alberto Broggi, and Alessandra Fascioli. Stereo Inverse Perspective Mapping: Theory and Applications. *Image and Vision Computing Journal*, 8(16):585–590, 1998.

-
- [16] A. Broggi, A. Fascioli, R.I. Fedriga, and S. Ghidoni. Automatic Calibration of a Urban Video Surveillance System through the Observation of Zebra Crossings. In *Procs. of the 2007 International Conference on Computer Vision Systems*, March 2007.
 - [17] H. Nanda, C. Benabdelkedar, and L. Davis. Modelling pedestrian shapes for outlier detection: a neural net based approach. In *Procs. of the 2003 IEEE Intelligent Vehicles Symposium*, pages 428–433, June 2003.
 - [18] Z. Liang and C. Thorpe. Stereo and neural network-based pedestrian detection. In *Procs. of the 2000 IEEE International Conference on Intelligent Transportation Systems*, pages 148–154, September 2000.
 - [19] K. Tabb, S. George, R. Adams, and N. Davey. Human shape recognition from snakes using neural networks. In *Procs. of the 2000 IEEE International Conference on Computational Intelligence and Multimedia Applications*, pages 292–296, September 2000.
 - [20] H. Nanda and L. Davis. Probabilistic Template Based Pedestrian Detection in Infrared Videos. In *Procs. of the 2002 IEEE Intelligent Vehicles Symposium*, pages 15–20, June 2002.
 - [21] Xia Liu and Kikuo Fujimura. Pedestrian Detection using Stereo Night Vision. *IEEE Trans. on Vehicular Technology*, 53(6):1657–1665, November 2004. ISSN 0018-9545.
 - [22] Yajun Fang, Keiichi Yamada, Yoshiki Ninomiya, Berthold K. P. Horn, and Ichiro Masaki. A Shape-independent Method for Pedestrian Detection with Far-infrared Images. *IEEE Trans. on Vehicular Technology*, 53(6):1679–1697, November 2004. ISSN 0018-9545.
 - [23] M. Bertozzi, A. Broggi, R. Chapuis, F. Chausse, A. Fascioli, and A. Tibaldi. Pedestrian Localization and Tracking System with Kalman Filte-

- ring. In *Procs. of the 2004 IEEE Intelligent Vehicles Symposium*, pages 584–589, June 2004.
- [24] W.L. Wolfe and G.J. Zissis. *The Infrared Handbook*. Washington, USA, 1985.
- [25] D. Kooß, F. Bellotti, C. Bellotti, and L. Andreone. EDEL - Enhanced Driver's preception in poor visibility. In *Progress in Automobile Lighting (PAL) Symposium 2003*, Darmstadt, Germany, September 2003.
- [26] D. Kooß, F. Bellotti, C. Bellotti, L. Andreone, and M. Mariani. EDEL - Developing a Second Generation Night-Vision System. In *eBusiness and eWork*, Vienna, Austria, October 2004.
- [27] L. Andreone, P.C. Antonello, F. Bellotti, D. Grimm, M. Mariani, M. Wüller, and J. Heerlein. Development of a new generation of driving support system for night vision enhancement: the EDEL project. In *Intelligent Transport Systems ITS 2005*, Praha, Czech Republic, 2005.
- [28] Hui Sun, Chengying Hua, and Yupin Luo. A Multi-stage Classifier Based Algorithm of Pedestrian Detection in Night with a Near Infrared Camera in a Moving Car. In *Procs. of the 2004 IEEE International Conference on Image and Graphics*, December 2004.
- [29] F. Jahard, D.A. Fish, A.A. Rio, and C.P. Thompson. Far/Near Infrared Adapted Pyramid-Based Fusion for Automotive Night Vision. In *Procs. of the 2005 IEEE International Conference on Image Processing and its Applications*, pages 886–890, July 1997.
- [30] C. Bellotti, F. Bellotti, A. De Gloria, L. Andreone, and M. Mariani. Developing a Near Infrared Based Night Vision System. In *Procs. of the 2004 IEEE Intelligent Vehicles Symposium*, pages 686–691, June 2004.

-
- [31] L. Andreone, F. Bellotti, A. De Gloria, and R. Lauletta. SVM-Based Pedestrian Recognition on Near-InfraRed Images. In *Procs. of the 2005 IEEE International Symposium on Image and Signal Processing and Analysis*, pages 274–278, September 2005.
 - [32] K. Moutoussis and S. Zeki. A direct demonstration of perceptual asynchrony in vision. In *Procs. of Biological Sciences/The Royal Society*, number 264, pages 393–399, 1997.
 - [33] P. Viviani and C. Aymoz. Colour, form, and movement are not perceived simultaneously. In *Vision Research*, number 41, pages 2909–2918, 2001.