

3 L'algoritmo "VIL"

3.1 Introduzione

ITS (Intelligent Transportation System) è il nome di un consorzio di ricerca mondiale (in Europa [ERT00]) che si propone lo scopo di integrare le informazioni e le tecnologie di comunicazione tra i veicoli e le reti di trasporto di persone e merci. Si definiscono sistemi "intelligenti" poiché portano maggiori conoscenze ai viaggiatori ed agli operatori. Nelle automobili la ricerca ITS aiuta i conducenti alla scelta del percorso migliore evitando i tratti maggiormente congestionati e diminuisce il rischio di collisioni grazie ad una serie di tecnologie per l'innalzamento della sicurezza in condizioni critiche. Su treni e autobus l'installazione di sistemi ITS aiuta il gestore all'ottimizzazione delle risorse di flotta e l'utente offrendo emissione automatica di biglietti e informazioni in tempo reale sui tempi di percorrenza. Sulle reti stradali la tecnologia dei sistemi ITS coordina i segnali di traffico (semafori, display informativi,...) rileva e gestisce le emergenze causate da incidenti e sovraccarichi istruendo i guidatori.

In molte grandi città (anche a Bologna) sono installati sistemi ITS [UTO00] per il controllo del traffico e l'ottimizzazione dei flussi. Conseguenze immediate sono la diminuzione delle emissioni di gas di scarico e del carico di stress per gli automobilisti e, più in generale, un miglioramento della qualità della vita dei cittadini. La visione artificiale gioca un ruolo di spicco nella tecnologia dei sistemi ITS per il controllo del traffico. Questi necessitano infatti di un elevato numero di sensori per la raccolta dei dati di flusso. Fino ad ora sono stati utilizzati sensori con trasduttore induttivo detti "spire induttive": una volta

poste sotto il manto stradale per mezzo di tagli, le spire percorse da corrente alternata rilevano il passaggio del veicolo grazie alla variazione dell'induttanza che questo genera passando sulla spira. L'elevata affidabilità del sensore a spira ed il costo relativamente basso dell'elettronica necessaria hanno per contro alcuni svantaggi, primo tra tutti il costo dell'installazione e della manutenzione. Per installarli infatti occorre interdire il traffico sulla corsia di lavoro ed utilizzare squadre e mezzi specializzati in lavori stradali. Accade spesso poi che interventi e scavi successivi o la semplice dilatazione del bitume dovuta alle variazioni di temperatura che esso sopporta durante il ciclo stagionale, causino la rottura del sensore richiedendo costosi interventi di manutenzione. Si è così pensato di sostituire i sensori a spira con telecamere collegate a sistemi di image processing che siano in grado di fornire le stesse informazioni rese dai sensori induttivi. L'approccio più diretto mutua la sua filosofia proprio dai sensori induttivi: si definiscono sull'immagine piccole finestre in corrispondenza delle corsie la cui area assomiglia alla superficie utile del corrispondente sensore induttivo. Inoltre con una sola telecamera è possibile sostituire più sensori a spira abbattendo ulteriormente i costi del sistema. Le telecamere richiedono un minor numero di interventi di manutenzione rispetto alle spire e non richiedono l'interruzione del traffico durante installazione e manutenzione. L'obiettivo finale di questa ricerca è la realizzazione di una telecamera intelligente (dotata di algoritmi di elaborazione di immagini) in grado di sostituire i sensori induttivi interfacciandosi direttamente alle connessioni preesistenti che in tali erano state predisposte per le spire.

3.2 Loop emulation e VIL

L'acronimo "VIL" significa Virtual Inductive Loop. Tale nome è stato scelto per mettere in risalto l'intento di riprodurre e sostituire i sensori a spira (loop) induttiva di cui si è parlato nel Par. 3.1. Per le ragioni giuste dette, in letteratura si parla di "loop emulation" quando si intende descrivere un sensore per il rilevamento dei parametri di un flusso di autoveicoli. Inoltre la loop emulation

realizzata con telecamere fornisce l'ulteriore vantaggio di poter trasmettere immagini che documentano visivamente la scena ad una eventuale centrale operativa. Il progetto VIL ha subito diverse modifiche nel corso del suo sviluppo. Sono state prodotte diverse versioni dell'algoritmo di Computer Vision implementate da principio "off-line" processando sequenze di immagini precedentemente salvate su disco rigido. In seguito si è passati ad una ingegnerizzazione del sistema su personal computer di modeste capacità, facendo uso di un frame grabber di basso costo ed ottimizzando il software tradotto in C++ in modo da raggiungere un frame rate di 25 immagini al secondo, che corrisponde ad una elaborazione real-time per il sistema televisivo PAL.

Esistono già in commercio alcuni sistemi di loop-emulation. Tra i più noti vi sono:

"Autoscope", realizzato dal Image Sensing-Econolite [AUT98]

"Video Track 900", prodotto da Peek Traffic [VID98]

"Vantage" di Odetics [VAN98]

"Traficon" sviluppato da Traficon [TRA98]

Tali sistemi fanno uso di normali telecamere collegate alla unità di processing attraverso un cavo video. Uno o più frame grabber provvedono a digitalizzare le immagini così che queste possano essere processate dagli algoritmi di Computer Vision.

3.3 Parametro di confronto della spira virtuale

I sistemi di loop emulation fanno uso di una immagine di riferimento (comunemente detta background o sfondo) con cui viene confrontata l'immagine contenuta nella finestra predefinita [MIC91]. In questo modo si può stimare se la differenza tra le due debba essere interpretata come segno della presenza di un veicolo in corrispondenza della finestra stessa. Per questo motivo vi sono due fattori che rivestono una notevole importanza per ottenere il buon funzionamento di un algoritmo di loop emulation: il primo è costituito dal modo in cui si

aggiorna l'immagine di riferimento ed il secondo dal modo attraverso il quale questa è confrontata con lo stato istantaneo della finestra.

Durante il normale funzionamento è normale prevedere che le condizioni di illuminazione della scena ripresa cambino in continuazione. Tali variazioni sono dovute al normale ciclo solare così come ad eventi metereologici in genere o al contributo della illuminazione artificiale. Per questo è impensabile che un sistema funzioni correttamente con l'immagine di riferimento acquisita una sola volta in fase di installazione ed è quindi necessario prevedere un suo aggiornamento. Esso deve avvenire idealmente quando nella finestra non sono contenuti veicoli e comunque quando essa non risente di fattori di disturbo che ne modificano lo stato di "quiete". Per contro, anziché utilizzare un criterio di aggiornamento selettivo i cui malfunzionamenti possono dipendere da errori sulla stime della assenza di veicoli, esiste la possibilità di aggiornare il riferimento in modo continuo evitando quindi di preoccuparsi dello stato della finestra. L'ipotesi sotto cui questo è possibile è che nessun oggetto si fermi per un periodo prolungato di tempo in corrispondenza della finestra stessa. In tale caso infatti l'immagine della sua presenza verrebbe assunta come riferimento causando la scomparsa del segnale sulla sua presenza e generando malfunzionamenti sulla rivelazione dei veicoli successivi. Per preservare il contenuto del background si utilizza di norma un aggiornamento lento attraverso l'uso di coefficienti detti anche "pesi".

Secondo quanto ci è stato possibile apprendere attraverso le fonti letterarie la funzione SAD (Sum of Absolute Differences, somma dei valori assoluti delle differenze) è la scelta tipica per confrontare il background alla finestra corrente [INI85], [INI89], [WAT94], [HIL78] e [VIA99]. A causa della estrema variabilità delle condizioni ambientali nelle aree urbane il background cambia frequentemente il suo aspetto ed è quindi necessario dotare l'algoritmo di rivelazione di un opportuno stadio di aggiornamento. Le ombre, le nubi e le variazioni luminose dovute a cause artificiali assieme ai sistemi di diaframma automatico ("auto iris" o "auto gain") sono tra le più comuni fonti di rumore per il background. L'idea proposta in questa tesi è di eseguire il confronto con il background della finestra sensibile per mezzo della funzione di Cross

Correlazione Normalizzata (NCCF). Osservando la Figura 3-1 se si considerano i pixel di una finestra sensibile $M \times N$ come le componenti di un vettore $M \times N$ -dimensionale, la funzione SAD corrisponde alla norma L_1 del vettore differenza \vec{D} tra i due rappresentanti rispettivamente la finestra corrente \vec{C} ed il background \vec{B} . Essa dipende significativamente sia dalla luminosità che dalla trama (texture) di entrambe le immagini.

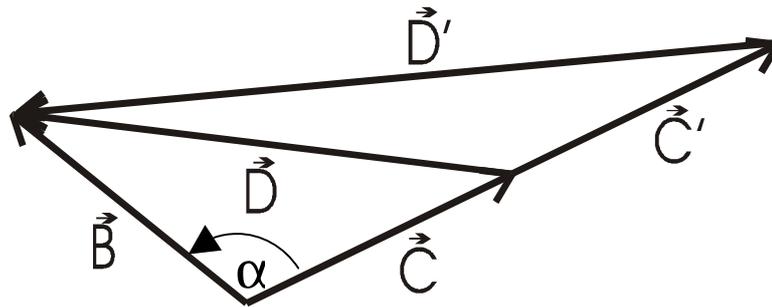


Figura 3-1 Comparazione img-bkg con SAD ed NCC in presenza di variazioni di luminosità

Come detto, la luminosità cambia frequentemente a causa delle fonti di rumore menzionate in precedenza. Nella Figura 3-1, a causa di un incremento di luminosità \vec{C} cambia in \vec{C}' e la differenza \vec{D} diventa \vec{D}' . Al contrario la funzione NCC è proporzionale all'angolo α tra i due vettori immagine corrente e background. Per questo motivo può considerarsi più immune di SAD rispetto alle variazioni di luminosità dovute al rumore (cioè è potenzialmente sensibile solo alle variazioni della texture).

3.4 Metodi di confronto finestra-sfondo

In una prima versione dell'algoritmo è stata implementata la regola descritta in Figura 3-1. Tale soluzione consente all'algoritmo di funzionare correttamente a meno che non avvengano variazioni di illuminazione della scena, come documentato in [DIS00].

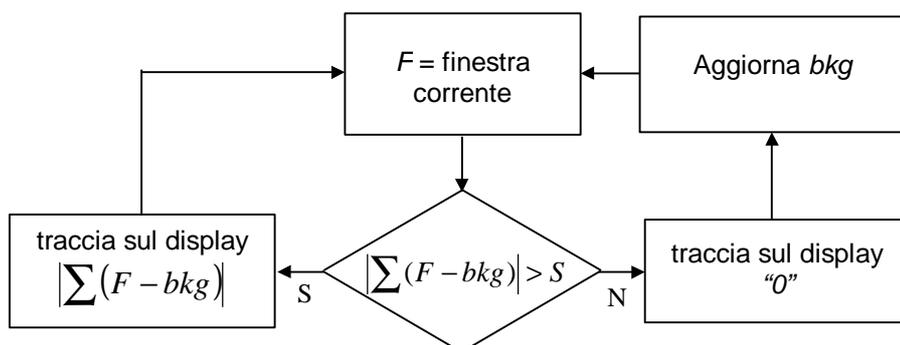


Figura 3-1 schema a blocchi dell'aggiornamento dello sfondo

Ad ogni iterazione l'algoritmo aggiorna una variabile matrice F con i pixel della finestra sensibile; poi calcola la somma dei valori assoluti delle differenze pixel per pixel (SAD sum of absolute differences) con la finestra di riferimento bkg e, se questo supera una soglia S , viene tracciato sul display un bin proporzionale al valore della somma stessa. Il risultato ottenuto è mostrato in Figura 3-2. Al centro della immagine si nota in bianco la finestra sensibile (loop virtuale) che dà origine all'istogramma. I diversi colori per i bin dell'istogramma significano un passaggio per lo zero e quindi un nuovo veicolo transitato.



Figura 3-2 VIL, tracciamento della funzione SAD

3.5 Il "VIL Analiser"

Su un personal computer dotato di scheda frame grabber sono state implementate successivamente sia la rivelazione di veicoli ottenuta con SAD che quella ottenuta per mezzo di NCC. Il frame grabber è stato collegato ad un videoregistratore contenente un nastro con sequenze di traffico riprese da una telecamera preposta alla sorveglianza di nodi stradali critici. Il frame grabber di tipo economico utilizza un circuito integrato molto comune in questo tipo di applicazioni (Brookthree "BT848" [BRO00]) che integra sia la funzione di digitalizzazione che l'interfaccia con il bus PCI. I driver del frame grabber sono stati interfacciati al sistema operativo con il kit di sviluppo Microsoft Vision SDK [VIS00]. Questo strumento permette di accedere ai puntatori alle aree di memoria contenenti le immagini così da ottenerle dal frame grabber sotto forma di array alla velocità massima consentita dalla elaborazione e visualizzarle sulla finestra dell'applicazione. L'elaborazione dell'immagine è realizzata da codice C++.

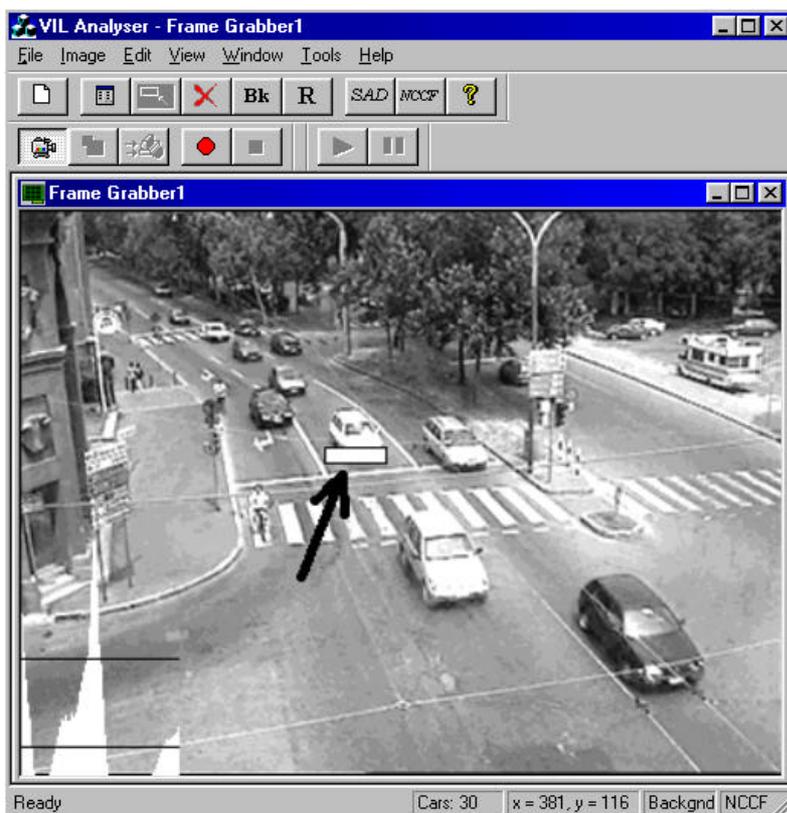


Figura 3-1 layout del “VIL analiser”

Il programma chiamato “VIL analiser” consente all’utente di posizionare le finestre sensibili nella immagine e valutare le prestazioni inerenti il conteggio riportate nell’apposita finestra. Il layout del “VIL analiser” è mostrato nella Figura 3-1. Il loop virtuale definito dall’utente è indicato dalla freccia nera mentre il contatore dei veicoli compare nella barra di stato dell’applicazione; il segnale di uscita generato dal “VIL analiser” comparando l’immagine corrente con il background è simile a quello prodotto da un sensore a spira tradizionale [ELS00] come si può notare osservando la forma d’onda tracciata nell’angolo inferiore sinistro della finestra dell’applicazione. Il contatore di veicoli è incrementato tramite l’uso di una soglia a isteresi sul segnale di uscita: per far progredire il contatore il segnale deve prima superare la soglia superiore e quindi scendere sotto quella inferiore. “VIL analiser” consente anche all’utente di scegliere tra un background fisso (acquisito una sola volta, quando non sono presenti veicoli in corrispondenza della finestra) e un aggiornamento pesato del

background, tramite cui esso è aggiornato con continuità dal software in accordo con due pesi: $New_bkg = Old_bkg \cdot w_1 + Actual_img \cdot w_2$, dove $w_1 + w_2 = 1$. I due pesi controllano la velocità con la quale l'algoritmo aggiorna il background: tanto più w_1 tende all'unità tanto più il processo di aggiornamento sarà lento.

3.6 Risultati sperimentali

Per valutare le prestazioni delle rivelazioni basate su SAD e su NCC abbiamo individuato alcune sequenze campione (di benchmark) caratterizzate da andamenti della luminosità ben definiti. Abbiamo scorso una registrazione di 4 ore individuando le tre sequenze descritte nella Tabella 3-1.

| | Seq1 | Seq2 | Seq3 |
|-------------------------------|-----------------------------------|--|---|
| Trend luminoso | Luce costante per l'intera durata | Lieve incremento della luminosità genera rumore al di sopra della soglia della funzione SAD per il 4% degli eventi | Rapido incremento della luminosità dopo che si è verificato il 43% degli eventi |
| Durata | 6' 05" | 7' 06" | 6' 49" |
| Eventi veri | 99 | 93 | 74 |
| Occlusioni sugli eventi veri% | 15% | 4% | 22% |

Tabella 3-1 Sequenze di benchmark

“Eventi veri” esprime il numero di veicoli che un loop virtuale avrebbe dovuto rivelare. “Occlusioni sugli eventi veri %” è la percentuale di eventi veri nei quali il background non compare tra due veicoli consecutivi. Nella tabella Tabella 3-2 sono riportati i risultati sperimentali del processo di valutazione. L'aggiornamento continuo del background è ottenuto ponendo $w_1 = 0.95$ e $w_2 = 0.05$. Le dimensioni della spira virtuale sono 7 pixel in altezza e 26 pixel in larghezza. I valori delle soglie per l'isteresi sono stati mantenuti costanti per l'intero insieme di misure.

Il parametro utilizzato per la valutazione delle prestazioni è definito come:

$$Accuratezza\% = \left(1 - \left| \frac{Ev_veri - Ev_misurati}{Ev_veri} \right| \right) \cdot 100$$

La stima dell'*Accuratezza%* non tiene conto di possibili sovra/sotto-compensazioni del conteggio. Questo è il modo più comune per stimare l'accuratezza degli algoritmi nei sistemi UTC (Urban Traffic Control) [COT94] dal momento che essi devono fornire conteggi in varie condizioni di illuminazione, nell'arco dell'intera giornata, piuttosto che dare informazioni esatte sulla presenza di veicoli in un istante preciso. Le sovrastime nei conteggi accadono frequentemente a causa della presenza del parabrezza (o a causa di fronti multipli nella texture di veicoli lunghi) che generano più di un picco nel segnale prodotto dalla finzione di confronto. Questo problema può essere attenuato aumentando la dimensione della spira virtuale. Sfortunatamente questo rimedio riduce la risoluzione aumentando il numero delle sottostime quando i veicoli viaggiano mantenendo ridotte le distanze, condizione che riduce la porzione di background visibile tra due veicoli consecutivi. Questo inconveniente aumenta se le immagini sono riprese da una angolazione che generi una forte distorsione prospettica. Per questo si è applicata alle immagini una trasformazione tale da ridurre l'effetto della prospettiva con lo scopo di incrementare l'area dello sfondo tra i veicoli che decresce progressivamente all'aumentare della distanza degli oggetti dall'obiettivo. I risultati della trasformazione sono mostrati in Figura 3-1.



Figura 3-1-a)imm. originale

-b)imm. con correzione della prospettiva

Lievi miglioramenti sono stati notati nelle prestazioni di entrambi le funzioni misuranti, ma questi sono limitati ai casi in cui l'asfalto compaia davvero tra due veicoli mentre non c'è nulla che possa sostituire una migliore angolazione di ripresa nel caso in cui lo sfondo sia totalmente occluso. A fronte dell'elevato costo computazionale introdotto da una trasformazione di questo tipo, in grado di produrre risultati appena soddisfacenti, si è deciso di escluderla dalla elaborazione runtime e dalle misure effettuate, assumendo di poter contare su una installazione opportuna dei sensori.

Osservando la Tabella 3-2 da sinistra a destra e partendo alla seq 1 possiamo osservare nella seconda colonna che la NCC ha prestazioni migliori di SAD per un 18%. Questo è dovuto alla maggior sensibilità della funzione SAD ai toni di grigio dei veicoli: se i veicoli hanno una tonalità simile a quella del background la funzione SAD non riesce a rilevare una differenza sufficiente. Al contrario la NCC riesce a rilevare i veicoli in modo molto più indipendente dalla loro tonalità di grigio. Nella colonna 3 sia SAD che NCC migliorano l'accuratezza sotto l'effetto dello stadio di aggiornamento pesato del background. Il miglioramento delle prestazioni dovuto all'aggiornamento del background potrebbe sorprendere in una sequenza senza cambiamenti di luminosità. In effetti ciò può essere compreso se consideriamo le prolungate occlusioni dovute ai "treni" di veicoli. Se si attiva il modulo di aggiornamento continuo, il background tende ad assomigliare alla texture dei veicoli che sono passati fino a quel momento. Accade così che il rivelatore può meglio riconoscere le differenze tra i veicoli stessi. Difatti la seq 1 contiene una percentuale di occlusioni non trascurabile. In altre parole lo stadio di aggiornamento pesato migliora le prestazioni del VIL rispetto alla occlusione reciproca dei veicoli.

| | Luce Costante (seq1) | | Lieve variazione luminosa (seq2) (coinvolge il 4% degli eventi) | | Discontinuità luminosa (seq3) (dopo il 43% degli eventi) | |
|------|--|---|--|---|---|---|
| | Bkg acquisito una volta: Accuratezza% | Bkg aggiornato continuamente: Accuratezza% | Bkg acquisito una volta: Accuratezza% | Bkg aggiornato continuamente: Accuratezza% | Bkg acquisito una volta: Accuratezza% | Bkg aggiornato continuamente: Accuratezza% |
| SAD | 72.7% | 81.5% | 78.5% | 79.3% | 36.5% | 70.3% |
| NCC | 90.9% | 97.8% | 100.0% | 98.9% | 86.5% | 89.2% |
| (Δ%) | (18%) | (16%) | (22%) | (20%) | (50%) | (19%) |

Tabella 3-2 Dati confronto SAD-NCC

Nella seq 2 il 4% degli eventi è affetto da un cambiamento di luminosità, tanto che la funzione SAD genera un output che supera la soglia inferiore anche negli istanti in cui non ci sono veicoli sulla “spira”: in quel frangente il contatore si arresta. Nella Tabella 3-2, colonna 4 si trova che la differenza tra l’accuratezza di SAD ed NCC è 22%, cioè esattamente il 4% maggiore del caso di seq 1 (colonna 2) dove non sono presenti cambiamenti di illuminazione. Se attiviamo l’aggiornamento pesato sulla seq 2, la differenza tra SAD ed NCC scende al 20% (colonna 5) perché esso consente a SAD il conteggio (dopo qualche istante, a seconda dei pesi impostati) anche dopo le variazioni luminose. Si noti che il 100% di accuratezza della NCC in colonna 4 è raggiunto grazie alla compensazione di sovra/sottostime. Infine nella seq 3, dopo che sono accaduti il 43% degli eventi il valore di SAD sale sopra la soglia a causa del brusco cambiamento luminoso così da arrestare il contatore di veicoli. Per questa ragione l’accuratezza di SAD è molto bassa nella colonna 6. Anche la NCC si comporta peggio nello stessa colonna ma ciò è dovuto alla elevata percentuale di occlusioni (22%) presenti in questa sequenza. Se attiviamo il modulo di aggiornamento pesato (colonna 7) la NCC migliora come al solito mentre la SAD migliora di molto le sue prestazioni (da 36.5% a 70.3%) dato che ora l’aggiornamento del background rende possibile il conteggio anche a questa funzione anche dopo la discontinuità luminosa.

3.7 Conclusioni

La funzione SAD è quella tradizionalmente usata per la misura della differenza tra immagine corrente e background. Si è proposto l’uso della funzione NCC quale criterio di comparazione per la rivelazione di veicoli in sistemi UTC basati sulla Computer Vision.

Il confronto tra un algoritmo di loop emulation basato su NCC ed uno basato su SAD mostra che le prestazioni fornite con l’uso della NCC sono significativamente migliori di quelle di SAD (l’accuratezza di NCC è sempre migliore di un 16% rispetto a quella di SAD) [DIS00]. I risultati ottenuti con

SAD sono in linea con quelli riportati in letteratura [COT94]. L'uso dell'aggiornamento pesato del background garantisce migliori prestazioni per SAD e per NCC, sia in condizioni di luce costante e sia in presenza di variazioni luminose. La NCC è insensibile alle più comuni sorgenti di rumore (generalmente cambiamenti della illuminazione) e consente al sistema di rilevazione di operare anche in assenza di un filtro di rumore. L'uso della NCC quale funzione di detection su piccole finestre non genera carichi computazionali significativi rispetto a quelli richiesti da SAD. Installando "VIL analyser" su un processore Pentium 200 MHz si è raggiunto il processing rate di 25 immagini al secondo.

A seguito di questi risultati sperimentali la funzione NCC sarà utilizzata quale criterio di comparazione per l'implementazione del sensore compatto descritto nella introduzione di questo capitolo, il cui studio di fattibilità è stato condotto presso il laboratorio di Visione Artificiale [VIT99] della facoltà di Ingegneria Elettronica dell'Università di Bologna. Tale sensore avrà per obiettivo la sostituzione di quelli tradizionali di tipo induttivo. Per il suo completamento si renderà necessario dotare l'algoritmo dello stadio di elaborazione notturna durante la quale lo scenario cambia la sua natura. Il suddetto sensore potrà essere applicato con facilità sulle strutture preesistenti (come quelle per l'illuminazione pubblica) e sarà di facile impostazione tanto più le corsie saranno delimitate ed i percorsi dei veicoli vincolati. Sia i loop induttivi che un sistema di loop emulation infatti risentono delle assunzioni poste in fase progettuale, secondo le quali i veicoli transitano per lo più lungo traiettorie predefinite. Se ci trovassimo ad esempio in un ampio parcheggio, senza nessun tipo di barriera sarebbe impossibile richiedere la caratterizzazione del moto dei veicoli al suo interno ed ancor più lo sarebbe se gli oggetti in movimento non fossero più appartenenti alla sola classe degli autoveicoli. Anche per queste ragioni si è deciso di affrontare lo studio descritto nel prossimo capitolo.

