

2 ANALISI DEL CONTESTO TRAFFICO

2.1 Descrizione dell'analisi

Questo capitolo descrive il contesto di immagini e sequenze aventi per oggetto scene di traffico. L'analisi è eseguita dal punto di vista delle applicazioni di Computer Vision a partire da esempi narrati in letteratura. Si passano in rassegna i possibili scenari, con attenzione particolare alle rispettive peculiarità, mettendone in evidenza l'ampia variabilità. Le soluzioni adottate sono classificate in due superclassi ("tripwire" e "tracking") prendendo spunto da [CHA95]

2.2 Approcci possibili

In generale esistono due tipi di approcci di Computer Vision per l'analisi di immagini al fine di ottenere dati utili al controllo dei flussi di traffico. Nel primo caso, con i cosiddetti sistemi "tripwire" (Autoscope [MIC91], "CCATS" [TRA98], Vantage [VAN98], Videotrack [VID98], ...) i veicoli sono rilevati analizzando i pixel contenuti in piccole finestre o linee selezionate nell'immagine; la filosofia di questo approccio trae spunto dai sensori a spira induttiva ancor oggi usati a questo scopo mutuandone il concetto di area sensibile localizzata al centro della corsia stradale. Nel secondo caso i cosiddetti sistemi di tipo "tracking", (EVA e IDET [CHA95]) in cui si considera una sottoparte

(maschera) consistente dell'immagine data, tale da rendere possibile l'inseguimento (tracking) del veicolo all'interno di tale area.

I sistemi tripwire sono più adatti a raccogliere dati lungo arterie stradali ove è dato per normale un flusso unidirezionale, il numero degli accessi laterali e delle intersezioni è ridotto e la situazione di congestione è un evento raro. Essi consentono di effettuare il conteggio dei veicoli, di misurarne la velocità media su tratti di qualche decina di metri e di effettuarne una classificazione sommaria, sulla base della loro lunghezza, tra un paio di misure possibili. Trovano facilmente applicazione in autostrade e superstrade.

Questi sistemi richiedono un semplice setup con pochi parametri da inserire, sono poco sensibili nei confronti delle oscillazioni delle telecamere, implicano un basso carico computazionale grazie al limitato numero di pixel in esame. Essi sono tuttavia vulnerabili nei confronti di riflessi ed ombre, necessitano di complessi algoritmi per l'esercizio notturno e di accuratezza nella scelta delle dimensioni dei traguardi sensibili, al fine di evitare errate associazioni tra fari e numero di veicoli transitati e alee riguardo l'appartenenza a più corsie dello stesso veicolo.

Tra gli esempi citati esistono anche sistemi tripwire dove lo sviluppo di un software di supporto ed interpretazione dei dati rende possibile la loro applicazione in aree urbane per il controllo delle intersezioni. L'analisi dei dati è ampliata in modo da gestire modelli comprendenti più incroci, fino a contemplare intere reti stradali urbane. Questi applicativi [UTO00] usano modelli che spesso si basano sulla definizione di coda di veicoli: agli incroci stradali nasce una coda quando i veicoli si fermano alla linea di stop con semaforo rosso; la fine della coda si sposta indietro ad ogni nuovo veicolo che sopraggiunge. Essa comincia poi a dissolversi (gestione FIFO) quando i veicoli in testa ripartono (il concetto di coda di automezzi è stato studiato in maniera esaustiva, ad esempio, da Montgomery e Quinn [MON91] che lo analizzano in termini di "onde d'urto di traffico" generate dalla cadenza dei semafori; il modello non considera solo le ripercussioni locali, ma predice gli effetti su più intersezioni).

I sistemi tripwire caratterizzano la coda di un incrocio segmentando la lunghezza delle corsie per mezzo dei traguardi e calcolando il numero dei veicoli

presenti in un tronco come differenza tra quelli entrati e quelli usciti. Questi sistemi prescindono dalla descrizione del moto dei veicoli, il cui studio non è richiesto, se non nei casi in cui sia necessario il calcolo della loro velocità media che si misura come segue: il veicolo che transita sul traguardo produce una variazione dei pixel che supera la soglia di background, avviando così un cronometro; al traguardo successivo (posto sufficientemente vicino in modo da non creare ambiguità) si calcola la velocità media tra i due traguardi sulla base del tempo trascorso. Per la gestione delle code di veicoli, i sistemi tripwire implementano uno “stack” che le rappresenta virtualmente ed alle cui locazioni sono associati i veicoli. La “coda” ha una collocazione spaziale generica all'interno del segmento di carreggiata. Essa esiste dal momento in cui il bilancio in/out non è in pareggio e sul traguardo normalmente situato sulla linea di stop c'è un veicolo fermo (o non è segnalato transito). Coi sistemi tripwire si possono anche descrivere gli stati di “congestione” o “rallentamento” del traffico: essi dipendono dalla media delle velocità di ogni veicolo.

I sistemi con approccio di tipo tracking sono utilizzati per risolvere situazioni ove interessi sapere quanti veicoli siano presenti e dove si trovino nella scena. I sistemi di tipo tracking forniscono direttamente più informazioni rispetto ai sistemi tripwire poiché sono sensibili ad un maggiore numero di fattori percepibili visivamente. Per questo motivo richiedono codice specifico che riduca il rumore inevitabilmente presente in un flusso di dati così elevato. I sistemi tracking misurano i parametri caratteristici della coda a partire dal moto che deve necessariamente essere rilevato tra due fotogrammi e dal quale non possono prescindere: grazie al moto si possono rilevare i veicoli che saranno etichettati e dunque collocati in punti precisi nell'immagine della scena. L'operazione appena menzionata ne implica il conteggio. In questo caso la gestione parallela della “coda virtuale” non è indispensabile ma, in pratica, aiuta la gestione di quella serie di errori come la ‘fusione’ (merging) di due veicoli contigui con le stesse caratteristiche di moto o l'assimilazione di un veicolo fermo allo sfondo.

Utilizzando la loro capacità di compiere misure dimensionali i sistemi “tracking” risolvono e associano correttamente le ombre ai rispettivi veicoli e

riconoscono i cambi di corsia; durante le riprese notturne sono in grado di associare gli spot dei fari ai rispettivi veicoli che possono classificare al più in tre differenti categorie; hanno anche una buona precisione sulla misura della velocità (mediamente migliore del 95%).

I sistemi tracking sono sensibili alle condizioni di illuminazione: fondi stradali illuminati da luce scarsa o non uniforme rendono più difficile a questi sistemi il riconoscimento dei veicoli.

Gli algoritmi di tracking sono sensibili alle oscillazioni della telecamera il cui angolo di ripresa è di fondamentale importanza soprattutto per le conseguenze che esso ha sulla occlusione della vista dei veicoli retrostanti. Le camere dovrebbero essere posizionate alla massima altezza possibile (almeno 7.5m, meglio 9m). Nelle installazioni esistenti le telecamere possono essere montate su ponti o su pali, meglio se tra le corsie, perché ciò consente di controllare anche il flusso opposto, rende l'occlusione meno evidente e si presta meno a subire atti vandalici. Sui sistemi tracking si possono integrare optional come la decodifica delle targhe automobilistiche (che esula dallo spettro delle nostre applicazioni), il rilevamento di incidenti, il computo dei cambi di corsia.

I dati così raccolti sono codificati e memorizzati localmente e costituiscono un archivio 'storico' a cui il singolo sistema attinge per auto-reinizializzarsi.

2.3 Sistemi "tripwire"

Questo tipo di sistemi trae spunto dal sensore di rivelazione passaggio veicoli di tipo "spira induttiva". I sistemi a basati sull'induzione elettromagnetica e quelli basati su Computer Vision hanno in comune il modello e differiscono per il tipo di trasduttore: spire induttive nei primi, telecamera a CCD nei secondi. Quindi le variabili di sistema utilizzate dai software dei due sistemi sono le medesime.

Il sistema tripwire collegato alla camera esamina solo alcuni traguardi posti longitudinalmente e trasversalmente alle corsie. Il primo vantaggio che se

ne trae è quello di un limitato costo computazionale dovuto a image-processing. Il semplice criterio con il quale si stabilisce la presenza di un veicolo sul traguardo è quello della ‘differenza entro soglia’ dei valori dei pixel dallo sfondo corrente. Con opportune correlazioni sui traguardi si possono estrapolare la velocità media (correlandone due successivi) e la lunghezza del veicolo (misurando il tempo di transito a una data velocità). Il numero limitato di pixel esaminati in questo approccio garantisce velocità di elaborazione e risultati soddisfacenti, se non sono richieste accurate misure cinematiche. Con riferimento a [SOT96] la misura real-time del flusso di auto su una carreggiata basata su edge detection con comparazione a soglia dinamica dà risultati con errore inferiore al 10%, ma non è possibile sfruttare il sistema per il rilevamento di code (e quindi applicarlo in ambito urbano) perché predisposto per operare su autostrade piuttosto che in prossimità di incroci.

Per questo, al peso modesto dell’elaborazione di image-processing di questi sistemi, si può abbinare un modello fisico raffinato e ottenere informazioni più complete come nel caso del noto sistema tripwire “Autoscope” [MIC91] [AUT98] operante in real-time. Esso si adatta automaticamente a numerosi tipi di sfondo stradale senza la necessità di reference-marks allo startup o durante il funzionamento. I suoi progettisti ne garantiscono il funzionamento in presenza di ombre, riflessi, cambi d’illuminazione; ciò fa supporre che siano stati messi a punto filtri per questo tipo di disturbi basati sull’inseguimento e sul labelling del veicolo. Esso funziona in condizioni di traffico congestionato o rallentato, consente un posizionamento arbitrario delle camere, immune, entro certi limiti, dagli angoli di vista e richiede un processore Intel-386 con l’ausilio di un preprocessore per immagini.

In fase di installazione si configurano i traguardi trasversalmente e longitudinalmente alle corsie (3 o 4 cross-arrays “EDCLA” per ogni camera, come fossero spire lungo la corsia) che permettono la corretta interpretazione dei sorpassi e forniscono informazioni sui cambi d’illuminazione. La riduzione dei disturbi derivanti dalle ombre in movimento avviene attraverso la regolazione dinamica della soglia e la modifica del gain (calcolata sulla base di stime statistiche). L’implementazione real-time degli algoritmi di processing comporta

in una prima fase la riduzione dei dati ad una serie di caratteristiche e valori misurati. Successivamente si ricavano le informazioni più dettagliate nello stadio di calcolo a più alto livello.

Per rendere un'idea della filosofia tripwire, nella Tabella 2-1 sono elencate e descritte le variabili utilizzate dal modello di Autoscope.

s	numero stazione (linea di rilevamento)
S	numero totale di stazioni
n	numero di veicoli transitanti su una stazione
l	numero della corsia
L	numero di corsie
Dx_s	spazio tra le stazioni s ed $s+1$
T	numero di intervalli di scanning aventi durata Δt
t_{sln}	tempo al quale la n -esima macchina nella l -esima corsia giunge alla stazione s
n_{sln}	velocità della n -esima macchina alla stazione s , linea di rilevamento l
Q_{sl}	volume alla stazione s , linea l , sul periodo $T\Delta t$
Q_s	volume totale su tutte le corsie della stazione s

Tabella 2-1

Esiste un'ampia possibilità di scelta per i sensori adatti a sistemi basati su Computer Vision: telecamere a colori, B/N o all'infrarosso, normali o progressive scan, a seconda dell'applicazione.

In Autoscope appare un cenno riguardo a prove effettuate con un trasduttore di tipo radar in grado di identificare le masse (veicoli) che riflettano le radioonde elettromagnetiche con lunghezza d'onda di qualche cm. Tale tipo di trasduttore fornirà in ogni modo un'immagine video che sarà però scevra dal rumore dovuto alle condizioni d'illuminazione dell'ambiente, dei fari delle auto e della pubblica illuminazione. Risultati di questo test si sono rivelati migliori di un 5% rispetto a quelli ottenibili con normali telecamere. Sono ovviamente da tenere in considerazione un maggior costo per trasduttore ed il problema dell'inquinamento elettromagnetico al giorno d'oggi non più trascurabile.

2.4 Sistemi “tracking”

Un sistema “tracking” analizza una sottoparte consistente dell’immagine proveniente dal sensore la cui forma è definita in funzione della scena e delle regioni d’interesse. Si definisce “regione di interesse” di un sistema tracking quel sottoinsieme di pixel di una immagine ove si sviluppa il moto che si intende studiare.

In più di un caso sono stati raggiunti buoni risultati in termini di velocità di elaborazione, costo ed affidabilità. Per esempio in [AUB96] si è ottenuta una elaborazione con frame rate = 6 frames/sec facendo uso di un microprocessore Intel-486 e di un frame grabber. Il software di questo sistema si compone di quattro blocchi: 1-setup, 2-acquisizione, 3-pattern-detection, rimozione dei disturbi, 4-calcolo della lunghezza della coda. La parte di acquisizione consiste nell’estrazione dei contorni degli oggetti e nel calcolo delle primitive della scena invarianti per cambiamenti globali d’illuminazione. L’immagine risultante da questa elaborazione, chiamata “Ict”, è utilizzata per aggiornare l’immagine di background “Ibt”. La velocità di aggiornamento dipende sia dal periodo del rosso al semaforo che dalle condizioni di congestione ed ha lo scopo di escludere dallo “sfondo” (o riferimento) i veicoli fermi in coda durante l’attesa del verde o a causa di congestione. Tutte le forme che non fanno parte del background sono estratte per differenza tra l’immagine corrente e quella di riferimento $|Ict - Ibt|$: queste corrispondono a veicoli, ma anche fonti di rumore come riflessi, luci, ombre. Nella parte di pattern-detection, tre moduli rimuovono il rumore e il modulo di tracking rileva e registra tutte le forme in movimento come primitive corrispondenti a veicoli. In questo modo si eliminano le forme non in movimento come luci, ombre e riflessi fermi. Il modulo di variazione toglie le forme introdotte dai fari sulla strada. Queste sono caratterizzate da un falso moto apparente introdotto dai fari (che quindi rende impossibile al tracking la loro eliminazione) e da una stabilità delle primitive che appaiono. Quindi, per rimuoverle, bisogna valutare la stabilità delle primitive tra immagini successive. Il modulo “texture” rimuove le ombre dei veicoli. Il quarto stadio calcola la lunghezza della coda. Ogni forma “sopravvissuta” al passo precedente è

etichettata. Si genera così una coda se almeno un'etichetta si trova in prossimità della linea di stop. A partire dalla prima posizione ci si muove indietro di etichetta in etichetta, riempiendo gli spazi vuoti che cadono sotto una soglia il cui valore è scelto durante l'inizializzazione e corrisponde alla massima distanza interveicolare che in una coda -ci si può ragionevolmente aspettare- sia più o meno uguale alla lunghezza di un veicolo. L'esaurimento della coda è determinato dalla impossibilità di iterare il processo. I risultati [AUB96] del collaudo eseguito con immagini da nastri registrati in condizioni di sole/pioggia, nubi passeggere/cielo coperto, hanno stabilito che le condizioni di cielo coperto garantiscono la più alta precisione; il funzionamento notturno ha prestazioni nettamente inferiori a quelle diurne; nelle riprese con sole vi sono molti problemi dovuti al rumore, non totalmente eliminati dal modulo texture.

Una maggiore attenzione al problema notturno è stata dedicata dall'esperimento TITAN [TIT89] per il riconoscimento notturno dei veicoli attraverso immagini. Esse sono analizzate in modo indipendente tra loro ed i veicoli sono identificati a partire dai fanali anteriori con un ciclo suddiviso come segue:

- 1) calcolo dei massimi dell'immagine con soglie adattate. Come risultato si ottengono tutti i fanali, i loro riflessi e i fari pubblici: questo rumore sarà sottratto nei passi successivi;
- 2) applicazione di un gradiente per distinguere i riflessi (che appaiono spesso come diffusi e meno luminosi) dai veri fari delle auto; restano punti che sono massimi locali e del gradiente;
- 3) unificazione dei fari e dei riflessi mediante "dilatazione adattata"; questa trasformazione è basata sul fatto che le intensità dei massimi locali dovuti a riflessi sono minori di quelle dei fari e quindi tendono a sparire dopo la dilatazione;
- 4) i fari di uno stesso veicolo sono confrontati insieme. Questo metodo, che ci permette di porre un marker su ogni veicolo, elide quelli di cui è visibile un solo fanale.

2.5 Conclusioni

E' stato eseguito un attento esame della ricerca sviluppata in Computer Vision sui sistemi per la stima dei parametri di traffico, consultando materiale raccolto a partire dai primi anni '80. Da questo risulta che buona parte delle problematiche inerenti l'applicazione della Computer Vision al traffic-monitoring è stata risolta con successo, anche se implementata in progetti o sistemi commerciali diversi, che risultano insufficienti se considerati singolarmente e la cui integrazione può risultare difficile.

Nonostante siano numerosi i campi ove è possibile introdurre miglioramenti, i sistemi di visione artificiale sembrano pronti per risolvere i problemi di traffic-monitoring. Le misure di velocità eseguite dai sistemi esistenti sono soddisfacenti, mentre quelle di conteggio sono più soggette al rumore tipico di questo tipo di riprese. La classificazione funziona discretamente fino a due/tre classi differenti, in buone condizioni di illuminazione.

Tuttavia, nel materiale raccolto mancano spesso notizie sul carico computazionale delle soluzioni descritte e ciò induce in certi casi il sospetto che sia difficile raggiungere cicli di elaborazione con durata sufficientemente bassa, utili allo scopo di rilevare con efficacia i parametri di traffico, rispetto ai cicli semaforici. In molti casi non si affronta il problema del funzionamento notturno ed in altre condizioni critiche, limitando la verifica della soluzione a poche e selezionate sequenze di fotogrammi. Fortunatamente esistono documenti che si sono occupati del confronto tra vari sistemi e dove emergono dati numerici sulle rispettive prestazioni. Gli algoritmi variano in funzione dei modelli [BRI92], sono influenzati dai vincoli fisici dell'ambiente di lavoro (illuminazione e contrasto, profili e tipi di veicoli, prospettiva, tipi di corsie...) e sfruttano strumenti diversi (matematici, statistici, cinematici, morfologici,...).

Migliorare l'affidabilità del sistema rispetto alle condizioni di illuminazione della scena sia uno dei principali obiettivi da porsi.

