

1 INTRODUZIONE

1.1 Computer Vision e Pattern Recognition

Gli studi condotti nel corso del mio Dottorato di Ricerca si possono maggiormente ricondurre a quella branca della scienza che va sotto il nome di Computer Vision (o Machine Vision, o Visione Artificiale). Una descrizione intuitiva di questa disciplina può essere la seguente: la Computer Vision riproduce sui calcolatori elettronici il percorso cognitivo compiuto dall'essere umano nell'interpretazione della realtà che lo circonda attraverso le immagini che esso percepisce per mezzo degli occhi. In ogni essere umano tale percorso coinvolge anche la mente e, con questa, tutta la serie di "esperienze" bagaglio culturale della sua storia che, in molti casi, contribuiscono in modo determinante a completare ciò che egli "vede" ed a generare una interpretazione esaustiva della realtà circostante. Questa caratteristica rende più che ambizioso lo scopo enunciato nella definizione. Ciò diventa maggiormente vero dato che ancor oggi si preferisce, in molti casi, rinunciare al colore ed alla profondità, analizzando immagini in bianco e nero e provenienti da una singola vista. Poche sono infatti quelle applicazioni dove si considerano immagini a colori o "stereo" ed ancor meno quelle dove si analizzano entrambi questi aspetti. Il motivo di queste semplificazioni risiede nella inadeguatezza della potenza di calcolo rispetto alle informazioni da acquisire e processare.

Molti dei problemi di Computer Vision sono classificati come aspetti di Pattern Recognition, la cui traduzione letterale corrispondente non rende giustizia del concetto che identifica questo tipo di problema. Oltretutto la Pattern

Recognition non è tipica solo della Computer Vision. Si può pensare di avere a che fare con un problema di Pattern Recognition ogni volta che si intende classificare (applicare una tassonomia a) un oggetto o -in Computer Vision- ad una immagine o un suo dettaglio. Secondo una definizione che ritengo sufficientemente rigorosa, si può asserire che tutti i problemi relazionati con tassonomie si riducono a trovare uno o più “invarianti” possibilmente complementari fra di loro. La comparazione avviene attraverso una conveniente misura di distanza. Risolvere un problema di “Pattern recognition” significa allora, dato un modello e una serie di azioni, trovare una invariante per le trasformazioni che le azioni generano sul modello.

In genere si dice che risolvere qualsiasi problema si riduce a trovare l’invariante, poiché questo è il vero scoglio. Ad esempio, si immagini di voler trovare l’immagine del mio volto in un database di immagini. L’immagine di me che si ha a disposizione per la ricerca rappresenta il mio profilo destro, è quindi ruotata e magari anche traslata rispetto a quella contenuta nel database (modello). Chiameremo la rotazione e la traslazione “azioni” sul modello di “me” contenuto nel database. Il problema di Pattern Recognition consiste nel riconoscere il modello a partire dal profilo a disposizione; per risolverlo bisogna fornire al sistema una invariante. Analizzando il problema, dunque la base di dati, mi rendo conto che tutte le altre immagini ritraggono volti di persone di colore mentre la mia pelle è chiara. Quindi è facile individuare l’invariante per riconoscermi: il colore. Posso dichiararmi soddisfatto perché sono in grado di trovare sempre e comunque la mia immagine, indipendentemente dal tipo di azione (traslazione, rotazione, ecc.) applicata al prototipo. I problemi cominciano quando nel database c’è più di un uomo bianco. In questo caso si dice che l’invariante non è completa, ovvero manca qualcosa. Molti degli algoritmi e dei sistemi basati su Computer Vision descritti nella presente tesi contengono un problema di Pattern Recognition.

1.2 Utilità e applicazioni della Computer Vision

La visione artificiale sta sempre più assumendo il ruolo di supporto nell'analisi di immagini e filmati. La quantità di informazione contenuta in una immagine, e ancor di più in un filmato (= sequenza di immagini), fa sì che il compito di analizzarle possa, in molti contesti, risultare pesante per l'uomo, per la noiosa attesa di eventi rari o per l'elevato sforzo di concentrazione richiesto dal verificarsi di una rapida successione di eventi.

Nelle applicazioni industriali per il controllo della qualità, nella raccolta di dati per sistemi automatici o nell'analisi di flussi di immagini da telecamere adibite a videosorveglianza, ed in molti altri campi, la visione artificiale consente di limitare (se non eliminare) la necessità di un controllo umano delle informazioni raccolte.

Il grado di supporto dipende spesso in gran parte da quanto il tipo di ambiente da cui provengono le immagini possa ritenersi "controllato" e cioè dalla misura in cui le variazioni delle condizioni di ripresa possano discostarsi dal modello conosciuto dal sistema di visione che lo analizza. Data, in certi casi, l'impossibilità di operare in condizioni controllate, l'obiettivo principale posto a chi opera ricerca in questo campo è di individuare algoritmi e sistemi il più possibile immuni alle mutazioni rispetto al modello. Tale variabilità è maggiore per immagini provenienti da ambienti all'aperto, mentre è più contenuta per quelle provenienti da ambienti chiusi, come può essere il caso di sistemi di ispezione industriale dei quali, ad esempio, esistono esemplari installati su linee di produzione seriale in grado di fornire ottimi risultati.

1.3 Affidabilità

Mentre si può asserire che buona parte dei compiti "industriali" sono di possibile soluzione, la variabilità delle condizioni "outdoor" pone molti limiti ai sistemi di visione che spesso assolvono un numero limitato di compiti elementari e demandano comunque l'analisi finale all'operatore umano. Il vantaggio

principale portato dai sistemi di visione applicati alla videosorveglianza sta nell'essere sensibili ad alcune situazioni anomale in modo da allertare l'operatore umano risparmiandogli un'attenzione continua al flusso di immagini. Va tuttavia considerata una percentuale non trascurabile di falsi allarmi e di eventi perduti.

Il fronte attualmente dibattuto in ambito di ricerca su sistemi per videosorveglianza si colloca ad un livello più avanzato rispetto ai sistemi in vendita per gli stessi scopi. Esistono infatti alcuni prototipi di sistemi in grado di individuare la presenza di persone in una scena e distinguerle da altri oggetti presenti. Esistono poi sistemi basati su modelli in grado di individuare persone indipendentemente dal loro moto. Il movimento è infatti spesso usato come prima fonte di informazione per definire una regione di interesse in una immagine. Un sistema che non si basa sul moto (ed ha quindi un modello intrinseco più o meno evoluto di ciò che deve cercare) è da considerarsi più potente di uno che usa il moto come "spia" sulle zone di interesse. Infatti non sempre il moto indirizza correttamente "l'attenzione" dell'algoritmo. Esistono fenomeni di moto continuo come il movimento di foglie di alberi o l'intermittenza di lampade o dovuti ad oscillazioni della telecamera, tali per cui il problema immediatamente successivo diventa come separare il moto interessante da quello che rischia di saturare inutilmente le capacità di calcolo del sistema. Il problema appena descritto è meglio conosciuto come l'estrazione del background ovvero la definizione di ciò che non interessa prendere in considerazione dell'immagine. Anche il filtraggio del moto è un argomento su cui sono stati sviluppati numerosi approcci.

1.4 Applicazioni a traffico e videosorveglianza

La visione artificiale conta numerose applicazioni specifiche materia di stima dei parametri del traffico e videosorveglianza che possono essere principalmente divise secondo il livello dell'elaborazione (elementare o complessa), il tipo di ambiente in cui vengono impiegate (outdoor, indoor), il tipo di oggetti verso cui rivolgono la loro attenzione (persone, veicoli o più classi di soggetti).

Le più semplici richiedono che venga generato un allarme (e spesso registrata una sequenza di immagini) al verificarsi di un cambiamento in una parte predefinita dell'immagine ("change detection" ovvero una semplice verifica di variazione rispetto ad una immagine definita "regolare"). Tale cambiamento è spesso da intendersi rispetto al "background" definito in quel momento. L'immagine di (o le informazioni sul) background infatti può essere statica ed acquisita una volta per tutte, quando la scena si trova nello stato "normale" oppure, in applicazioni di maggior complessità, tale riferimento può adattarsi in modo continuo ai mutamenti della scena che non coinvolgono gli oggetti di interesse. In altri sistemi è possibile restringere la detection ad un certo numero di "zone di interesse" ignorando tutto ciò che avviene nel resto dell'immagine. Alcuni sistemi hanno il compito di raccogliere dati (numero di utenti in coda ad uno sportello, di passeggeri presenti su un autobus o di veicoli transitati su una strada...) mentre altri cercano di garantire la sicurezza (controllo di accesso a "zone proibite", individuazione di oggetti abbandonati, assembramenti inconsueti...).

Le applicazioni rivolte al traffico sono generalmente in grado di identificare veicoli, contarli, stabilirne la classe (cioè se trattasi almeno approssimativamente di auto, piccoli furgoni o grandi automezzi) e tenere traccia delle loro traiettorie e del loro moto (velocità, arresto in condizioni di emergenza...). Tali sistemi hanno di norma due campi di applicazione: la raccolta di dati (il numero di veicoli transitati su una carreggiata per la regolazione delle fasi semaforiche, l'analisi del livello di congestione del traffico) e "l'enforcement" ovvero la "verbalizzazione" automatica di sanzioni conseguenti la violazione di alcuni articoli del Codice della Strada (non ci occuperemo del secondo campo di applicazione).

Il genere più vario di applicazioni riguarda l'uomo. Ve ne sono alcune che stimano solo approssimativamente un volume di persone ed altre che inseguono il singolo individuo fino a riconoscerlo nuovamente dopo una occlusione o sono in grado di identificare se questi interagisce con un'altra persona scambiando qualche oggetto. Naturalmente molti sono i limiti a questi tipi di elaborazioni e molte le ipotesi restrittive sotto cui possono funzionare correttamente. Lo stato

attuale dell'arte propone una serie di algoritmi ben lontani da poter essere considerati utili in ogni contesto. Dalla descrizione sin qui condotta si può notare come sia sottile il confine che distingue un sistema per la caratterizzazione di un flusso di traffico da un sistema di videosorveglianza, dal punto di vista dell'analisi richiesta al sistema.

1.5 Percorso della ricerca presentata in questa tesi

In estrema sintesi, il percorso su cui si sviluppa la ricerca condotta ha origine nello studio della letteratura sui vari sistemi per la stima dei parametri dei flussi di traffico sia autostradale che urbano a fronte del quale si è scelto di sviluppare due filoni, l'uno mutuato dalla filosofia dei sensori attualmente in uso e l'altro di tipo più generale. Entrambi hanno portato alla sintesi di due algoritmi distinti. Sul primo sono state condotte misure e confronti sperimentali che hanno dimostrato la sua efficacia in termini di migliori prestazioni rispetto ai sistemi esistenti. Il secondo ha contribuito ad ampliare lo spettro della ricerca: esso ha infatti suggerito (stante l'ultima considerazione del paragrafo precedente) la generalizzazione del contesto: da questo punto in poi ai veicoli si sono aggiunti altri oggetti di interesse, persone incluse, conferendo così al tipo di ricerca un carattere più affine alla videosorveglianza. Le esperienze nient'affatto prive di problematiche, maturate sulle scene relativamente "banali" di traffico, hanno suggerito un approccio alla videosorveglianza che partisse dallo studio dello stadio di basso livello contenuto in ogni sistema di questo tipo. A seguito di una ulteriore ricerca si sono riprodotti molti degli algoritmi presenti in letteratura. Questo ha portato alla sintesi di nuove soluzioni che sono state confrontate con quelle esistenti.