

Gennaro Percannella^(*)

Rilevare oggetti in movimento: come funziona un algoritmo?

Rilevamento automatico delle persone che entrano in un'area non autorizzata, del rilascio di un bagaglio incustodito in un'area pubblica, di immondizia sul ciglio della strada; rimozione di oggetti in un museo; segnalazione della presenza di un'auto in panne nella corsia di emergenza; calcolo della velocità puntuale e media dei veicoli in ingresso in un Comune; conteggio di persone presenti in un negozio; determinazione dei percorsi preferenziali dei clienti all'interno di un supermercato... di cosa stiamo parlando? La risposta è immediata per chi ha già sentito parlare di *Analisi Video Intelligente* (spesso anche detta più brevemente "video analisi"), ovvero quell'insieme di modelli matematici, ancora oggi alla frontiera della ricerca scientifica, implementati attraverso tecnologie software, che consentono di analizzare un flusso video acquisito da una telecamera per riconoscere automaticamente ed in tempo reale quanto sta avvenendo nella scena.

^(*) Docente del corso di Visione Artificiale presso la Università degli Studi di Salerno e socio fondatore di A.I. Tech srl, spin-off accademico specializzato in progettazione e sviluppo di sistemi di audio e video analisi (www.aitech-solutions.eu)

Le operazioni di riconoscimento citate nel sommario rappresentano solo un piccolo sottoinsieme delle possibili funzioni che la video analisi mette oggi a disposizione e con il passare del tempo, grazie alla maggiore diffusione di tali tecnologie ed ai progressi della ricerca scientifica condotta nelle Università e nei Centri di ricerca di tutto il pianeta, tale insieme è destinato a crescere significativamente con funzioni sempre più complesse e affidabili. Bisogna anche dire che negli anni si è abusato del termine “video analisi”, attribuendo a tale tecnologia un potere *taumaturgico*. Troppo spesso si promettono capacità di analisi e affidabilità che sono ben oltre lo stato dell’arte non solo dei prodotti commerciali, ma anche della stessa ricerca scientifica: è bene quindi fare un po’ di chiarezza.

RILEVAMENTO OGGETTI

In questo articolo intendiamo fare luce sulle potenzialità - ma anche i limiti - dell’Analisi Video Intelligente, soffermandoci sulla prima fase delle elaborazioni effettuate dagli algoritmi di video analisi al fine di comprenderne i principi di funzionamento. Ci riferiamo alla cosiddetta fase di rilevamento oggetti, cioè a quell’insieme di operazioni finalizzate ad individuare in ogni fotogramma acquisito dalla telecamera i pixel associati agli oggetti di interesse (pixel di “foreground”) che stanno interagendo nella scena, separandoli dai pixel che rappresentano lo sfondo della scena (“background”). Il rilevamento degli oggetti è di fondamentale importanza in quanto tutte le fasi successive per la rilevazione di specifici comportamenti degli oggetti di interesse o per la misurazione di quantità fisiche, quali la velocità di un veicolo, partono dal presupposto che gli oggetti siano stati preventivamente rilevati in maniera precisa. Infatti, gli errori commessi durante la fase di rilevazione degli oggetti si ripercuotono sulle fasi successive, compromettendone il risultato finale. Nella nostra disamina faremo riferimento a funzioni di video analisi in grado di operare con telecamere fisse soffermandoci sui tre stadi di elaborazione (vedi **Figura 1**) che consentono il rilevamento degli oggetti: rilevamento foreground, aggiornamento background e post-elaborazione. Ci soffermeremo anche sulle principali cause di errore e le contromisure adottate per cercare di limitarne l’incidenza.

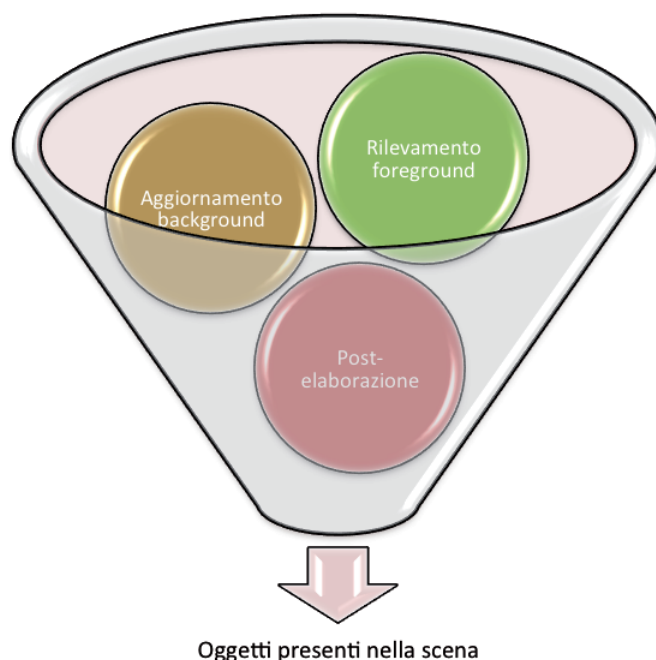


Figura 1: Operazioni svolte da un algoritmo di video analisi per la rilevazione degli oggetti presenti in una scena.

RILEVAMENTO FOREGROUND

Nell’ipotesi operativa che prevede l’impiego di telecamere fisse, l’algoritmo di rilevazione degli oggetti determina in ogni fotogramma i pixel che sono variati rispetto ad un fotogramma che rappresenta il background (**Figura 2.b**). Nella versione più semplice, il fotogramma di background (anche detto “reference frame” o più semplicemente “background”) è rappresentato da un’immagine acquisita in un istante di tempo precedente in cui nella scena non erano presenti oggetti. Tuttavia, l’applicabilità di questo approccio in contesti reali è molto limitata, dato che esso non tiene conto delle variazioni che via via intervengono nella scena a causa dei cambi di illuminazione o per il fatto che la scena stessa può modificarsi per l’ingresso, l’uscita o lo spostamento di un elemento dello sfondo (ad esempio una porta che viene aperta), o per un cambiamento ricorrente dovuto al movimento della vegetazione (ad esempio le fronde degli alberi agitate dal vento).

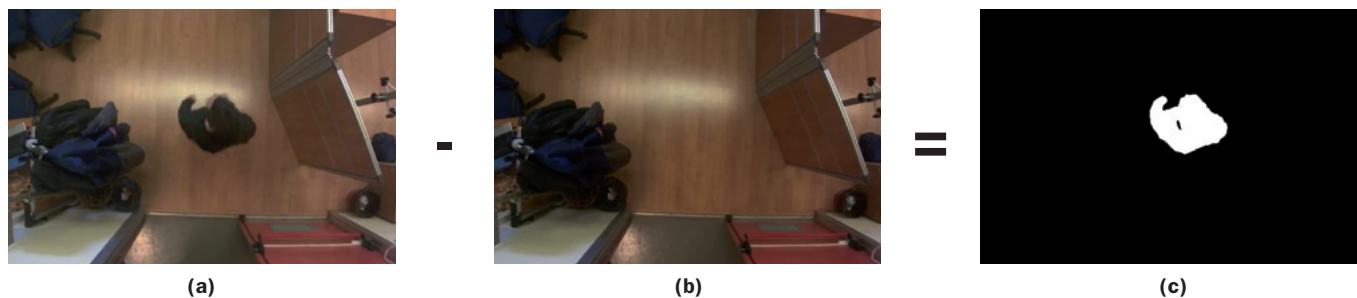


Figura 2: Il rilevamento del foreground all'istante di tempo t è ottenuto per differenza tra (a), il fotogramma acquisito dalla telecamera al tempo t , e (b), il reference frame al tempo t . Il risultato è rappresentato attraverso (c), la maschera di foreground, un fotogramma in cui i pixel bianchi sono associati all'oggetto in movimento, mentre i pixel neri si riferiscono al background.

AGGIORNAMENTO BACKGROUND

Per tenere conto delle variazioni che possono intervenire sullo sfondo, i sistemi di video analisi più avanzati effettuano una modellazione sofisticata del background in grado di aggiornare in maniera automatica la rappresentazione del reference frame per incorporare le modifiche che intervengono nella scena stessa. Tuttavia nell'eseguire tale compito l'algoritmo deve essere abbastanza "intelligente" per far sì che un oggetto di interesse che permanga immobile nella scena per lungo tempo (ad esempio una persona che si fermi per alcuni minuti nell'area inquadrata dalla telecamera) non venga inglobata nello sfondo, ma continui ad esserne segnalata la sua presenza.

POST-ELABORAZIONE

E i problemi non finiscono qui. A rendere ancor più complicata la rilevazione degli oggetti, intervengono altre problematiche, tra cui la presenza di ombre e riflessi che fanno apparire gli oggetti di dimensioni differenti rispetto alla realtà; il camuffamento, dovuto al fatto che taluni oggetti possono avere caratteristiche cromatiche simili al background, facendo sì che porzioni anche grandi di oggetti possano essere non rilevate. In pratica, la fase di post-elaborazione ha lo scopo di "ripulire", ove possibile ed auspicabilmente senza introdurre errori, il risultato del rilevamento del foreground. Le tecniche adottate in questo stadio sono numerose e specifiche del tipo di errore da correggere e la loro trattazione merita un specifico approfondimento⁽¹⁾.

CONCLUSIONI

Il risultato finale della fase di rilevazione degli oggetti descritta finora altro non è che un'immagine (**Figura 2.c**) in cui sono evidenziati con pixel bianchi su sfondo nero tutti gli elementi di interesse del fotogramma correntemente analizzato. Questa rappresentazione degli oggetti presenti nella scena sarà impiegata negli stadi successivi dell'Analisi Video Intelligente, i quali saranno deputati alle elaborazioni di più alto livello finalizzate al loro eventuale riconoscimento (fumo, fiamma, persona, animale, veicolo, ...), oltre all'analisi del movimento degli oggetti stessi per le applicazioni ove è necessario caratterizzare il comportamento (intrusione, loitering, panico, ...). Ottenere affidabilmente anche in condizioni critiche ciò che può apparire una banale immagine in bianco e nero, ma che adesso capiamo essere il risultato della applicazione di algoritmi sofisticati, determina il successo o il fallimento di un sistema di video analisi. L'errore nella rilevazione di pochi pixel in più o in meno può portare un algoritmo di video analisi a "scambiare lucciole per lanterne".



⁽¹⁾ Per un'analisi più dettagliata consultate la rubrica "Pillole di video analisi": <http://www.aitech-solutions.eu/blog/>