

Il riconoscimento del mondo reale

Prima parte

Come abbiamo fatto notare più volte su queste pagine, tutti i tentativi finora descritti non potevano essere considerati fine a se stessi. Una macchina che legge, interpreta e riconosce un parallelepipedo ha ben poco significato al di fuori di una pura sperimentazione didattica, e non giustificerebbe gli impegni economici e mentali che finora abbiamo descritto, anche nell'ambito della più pura ricerca universitaria. Quanto era finora successo era solo il primo gradino della successiva fase, quella della realizzazione di un dispositivo capace di leggere il mondo reale e di essere con esso indipendentemente interattivo (pensate al solito HAL per avere il top di questa aspirazione, ancorché fantascientifica). Il progetto di riconoscimento di immagini di oggetti non solo irregolari, ma anche «deformabili», vale a dire che non rispettano le loro proporzioni nel tempo è certamente ambizioso, visto che non è possibile, ancora di più in questo caso, affidarsi a tecniche di analogia tra le immagini-oggetto ed un catalogo di precedenti

codifiche cui far riferimento per analogia. D'altro canto programmi tipo «line finder», come quelli presentati su queste pagine fino ad ora, sono altrettanto inefficaci, per le intrinseche limitazioni più volte esposte, e relative, da una parte alla sempre da rispettare proporzionalità tra gli oggetti-bordi-spigoli-linee, e dall'altra alla indeformabilità dell'oggetto stesso. Sarebbe come dire che viene chiesto alla macchina di riconoscere il «David» o «la Pietà», ma viene esclusa, in questa ipotesi, che si possa riconoscere Bo Derek od Enzo Ferrari. Come al solito «in media stat virtus», a dire dei latini, vale a dire che un programma destinato al riconoscimento degli oggetti reali ed animati deve far per forza di cose riferimento al meglio delle due tecniche (riconoscimento induttivo dei particolari-analogia).

Finora, come più volte avevamo accennato ed evidenziato nelle precedenti puntate, i programmi di Guzman & C. avevano consentito il riconoscimento di strutture poligonali e poliedriche ben regolari e riconducibili univocamente a figure. Vedremo tra poco e ancora prossimamente che più perfezionati programmi di analisi di immagini lavorano, per così dire, per analogia, tentando assimilazioni e somiglianze di «percezioni» complete, invece di azzardare similitudini su basi di semplici assimilazioni fondate sul solo profilo o contorno, ancorché in base a tecniche molto sofisticate. In base a quest'ultimo principio, infatti, non è sempre, anzi è quasi mai possibile tentare operazioni di assimilazione complesse (secondo un programma basato su questi principi e pilotato da un algoritmo così orientato, un quadro ed uno specchio sono esattamente la stessa cosa, vale a dire un oggetto rettangolare, contenente una rappresentazione pittorica, una immagine). Generalizzando tale principio ed esemplificando, potremo affermare che una porta è a) una struttura rettangolare con la dimensione maggiore verticale o b) un oggetto assimilabile ad una più complessa struttura presente in memoria sotto forma di immagine fotografica. Si potrebbe (e sono proprio questi i termini che si presentano più spesso in bibliografia) dire che un programma sufficientemente evoluto rappresenta un oggetto immagine tramite una media sufficientemente equilibrata di valori «semantici» ed «iconici».

Una struttura «iconica» conserva tutte le informazioni visive ricavate dalla analisi e dalla campionatura della fotografia (incluso colore, brillantezza, ombre, e così via) così da riconoscere ed

individuare, se necessario, oggetti «a forma» di porta. L'altra struttura-rappresentazione è invece più forzatamente parametrica, vale a dire che in essa vengono conservati certi valori e funzioni proprie di elementi non pittorici (come ad esempio, rapporto lunghezza-larghezza, forma degli angoli, presenza delle famigerate strutture di base come Fork, Arrow, ecc). Generalizzando la cosa potremo dire che la prima struttura, la iconica, è deputata a presiedere al riconoscimento ed alla individuazione di particolari concreti, l'altra ai valori, per così dire, astratti.

In questa ottica un programma funzionante in tal guisa, capace cioè di individuare, al limite, linee significative sia in un poliedro che in una faccia umana deve, alla base, adattarsi a due schemi diversi. Esso deve fornire dimensioni «computazionali» in senso numerico, riferite a valori astratti e concreti, basati, tra l'altro, su una miscellanea di forme generali, in termini di particolari individuati dal computer stesso. Esso deve, inoltre, essere capace di «restringere» le informazioni ricevute ad un range ben individuabile di fattori di base, in modo che, attraverso la individuazione di certe uniche ed univoche caratteristiche dell'oggetto immagine sia possibile, poi, risalire all'oggetto origine ed al suo riconoscimento.

Un più recente lavoro di Guzman (A. Guzman, Analysis of curved lines drawings, using context and global information, MI-6, pagg. 325-376) illustra e dimostra esaurientemente la necessità di sviluppare nuovi concetti e tecniche di lettura delle immagini, in una sola parola, un nuovo «linguaggio» atto a descrivere in maniera più specifica nuovi domini della conoscenza. Nella sua trattazione, Guzman affronta una discussione teorica di questo problema, partendo, come al solito, dalla più grande e generale interpretazione di linee di una normale scena, sfrondata della maggior parte dei particolari inutili (così come avviene negli album da disegno dei bambini). Come già avveniva in Gra-

pe, egli evidenzia come qualsiasi rappresentazione pittorica è pur sempre ambigua, visto che, in qualsiasi caso, il puro criterio di analogia e somiglianza di un oggetto immagine con quello presente in memoria sotto forma di un catalogo-libreria di immagini-forme pre-costituite è pur sempre pericoloso e limitativo. In questi termini, d'altro canto, una descrizione semantica del disegno-immagine, in termini di linee rette e strutture poliedriche, è fondamentalmente inadeguato e nuove, più accurate strade nella ricerca e nell'analisi dei dati si impongono anche nel semplice riconoscimento di immagini comunque non standardizzabili.

Il grande problema, nella individuazione di queste immagini, non è, come potrebbe sembrare a prima vista, quello di analizzare curve particolari, non ben inquadrabili in un contesto logico preciso, ma il fatto, come accennavamo prima, che la stessa curva può assumere contorni diversi e, in teoria, innumerevoli. Il problema dei contorni è, inoltre, lo ricordiamo, afflitto da quel grave ed ancora praticamente insoluto dilemma rappresentato dal riconoscimento delle strutture concave.

Facciamo un esempio, tanto per intendere la tecnica di «comprensione» dell'oggetto, in termini «semantici» ed «iconici». L'occasione ci viene fornita dallo stesso Guzman, che esemplifica, nello scritto citato, la tipologia di riconoscimento di un cappello. Da una parte esiste una descrizione accurata della «forma» del cappello stesso sotto forma di contorni di falda e di cocca, dall'altra esistono una serie di descrizioni pittoriche, per forza di cose finita, relativa a tipi particolari di cappello. Il riconoscimento avviene indifferentemente passando prima per l'una e poi per l'altra delle tipologie di riconoscimento, e può accadere, comunque, che si possa avere un riconoscimento in ambedue i casi come in uno solo. Ciò nonostante, come nel caso del cappello visto da sopra di una delle puntate scorse, non è possibile prevedere tutte

le possibilità di visione del cappello stesso senza il possesso, implicito, di una serie, ancorché finita, di conoscenze «non visuali» dell'oggetto cappello, così come l'uomo riconosce come tale un colbacco, un elmetto militare o un copricapo delle ragazze «cacao meraviglioso» anche se questi esulano, come forma, da tutte quelle codificabili in qualche modo. In altri termini manca ancora alla macchina quello che comunemente viene definito come senso comune, che permette di riconoscere come tale una cuffia da notte anche a chi non ne ha mai vista una.

Da questo a passare al riconoscimento di facce il passo è semplice nella forma (anche se complesso nella sostanza). Il problema fu all'inizio semplificato cercando un algoritmo che riconoscesse termini, facce, espressioni, piuttosto semplici (e non a caso si fecero esperimenti su disegni di bambini e fotogrammi di cartoni animati). I primi esperimenti (che comunque, per giungere ad un risultato richiesero uno sforzo elevato) più che riconoscere il protagonista del fotogramma, consistettero nel cercare di isolare, per quanto possibile, in una scena, i personaggi; non a caso, per la semplicità degli oggetti-attori, e per la ridotta modificabilità facciale dei personaggi, fu scelto come oggetto del riconoscimento il mondo di Charlie Brown, inizialmente, ancora, limitato a singoli fotogrammi separati l'uno dall'altro, senza, cioè, tentare di riconoscere, negli stessi, il movimento.

Da qui al riconoscimento di facce e di oggetti in moto il passo era altrettanto breve; e la fascia degli sperimentatori e degli studiosi del problema si espanse notevolmente, con gran impegno di uomini e mezzi.

Proprio in quel periodo nuove possibilità venivano fornite, nel campo pratico, dalla comparsa sul mercato di macchine più potenti e di linguaggi più raffinati. Vedremo la prossima volta i risultati che ne vennero fuori, e che prelusero ad una conquista ancora più ardua; l'analisi del mondo reale. 