



La visione reale del mondo reale

Tutto quanto abbiamo visto nelle precedenti puntate relative alla possibilità di «comprensione» e «lettura» del mondo reale da parte di una macchina si è basato essenzialmente su procedure di analisi di figure poliedriche ideali (cubi, prismi) o, in mancanza, su tentativi di riconoscere figure mettendo insieme spigoli e vertici, oltre tutto rispondenti a caratteristiche particolari. Il mondo reale ben raramente è così semplice ed ordinato e questo per due motivi essenziali; l'osservatore (uomo) riconosce un oggetto anche a dispetto di notevole mancanza dei particolari (un chiodo lo si riconosce anche se completamente infisso nel legno); e riconosce con facilità la maggior parte degli oggetti anche se gravati di notevole imperfezioni di input, se così si può dire (ad esempio, una nave od un aereo è riconoscibile anche se non si vedono oblò, eliche o fumaio- li). E, in quest'ultimo caso, non si vede neppure il comandante, anche se, ovviamente, si sa che c'è. Tutto ciò avviene, nel primo e nel secondo caso, perché avviene una perfetta combinazione tra conoscenza ottenuta da un semplice input dal mondo esterno, e conoscenze pregresse cui si attinge per completare la totale conoscenza delle cose osservate; vale a dire che il mondo reale va osservato non solo con gli occhi della testa ma anche con quelli della mente; è possibile simulare questo processo tramite una macchina? Lo vedremo in queste puntate.

L'interpretazione delle immagini

Nella puntata precedente abbiamo evidenziato come sia necessaria una perfetta illuminazione dell'oggetto perché questo sia ben riconoscibile; ma non sempre sono queste le condi-

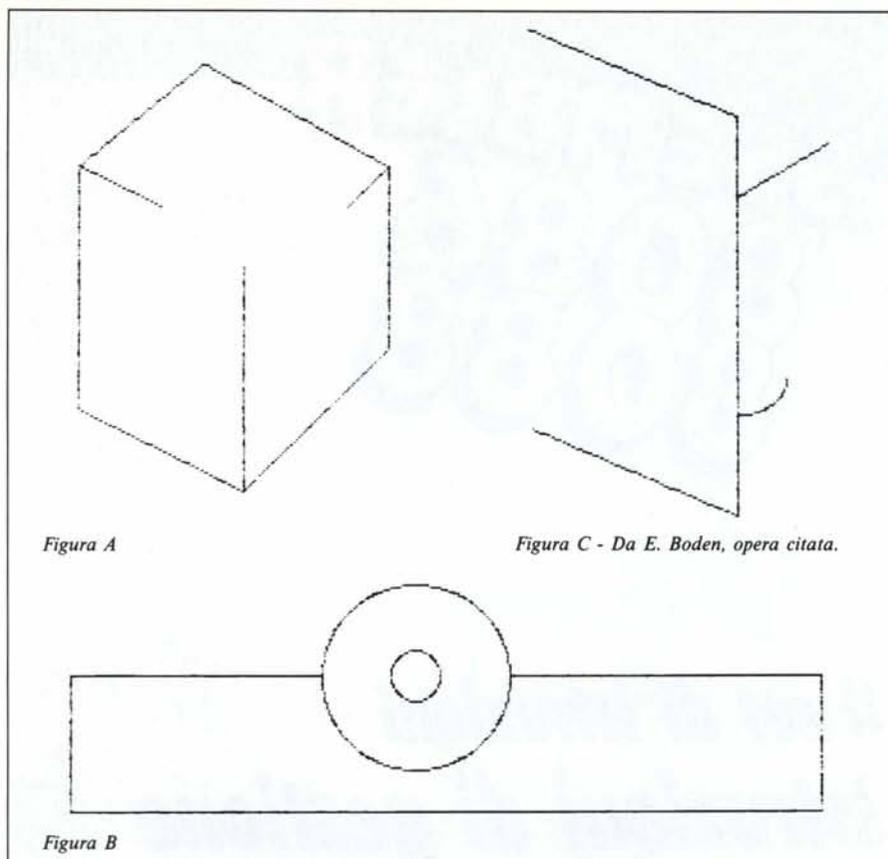
zioni ideali per la lettura di una immagine. La luce, come, con un esempio molto ironico riferisce Margaret Boden nel suo «Paradox of explanation», Proc. Aristot. Soc., NS, 62 (1962), pag. 158, può essere crudele non solo nel rivelare l'età delle star del cinema, ma anche per i rigidi programmi e processi di riconoscimento degli oggetti descritti nelle pagine precedenti. Una fotografia (od una immagine letta da uno scanner, ancorché perfezionato) di un dado bianco su uno sfondo altrettanto bianco, il tutto illuminato da una forte luce diffusa, molto probabilmente, anche in funzione delle poche ombre presenti, sarà difficilmente interpretabile, specie se non saranno visibili alcune linee fondamentali. E queste stesse linee, lette da una telecamera o da uno scanner, saranno altrettanto ben poco identificabili.

Consideriamo, ad esempio, la figura A, già analizzata in una puntata scorsa, vista stavolta nell'ottica della precedente premessa; l'angolo del cubo, al centro, non è visibile a causa di una illuminazione molto forte. Il fatto che la mente «aggiunge» la parte mancante in funzione di conoscenze, non solo geometriche, pregresse; tutto ciò lascia, tra l'altro, prevedere che la differenza tra visione reale e fittizia è molto meno definita di quanto non si immagini.

Per giungere ad un certo avvicinamento tra tecniche di visione reale e «meccanica» è necessario definire, almeno in larga massima, le modalità di acquisizione dell'immagine da parte della mente. La mente umana cerca di ricostruire immagini parziali secondo processi logici più o meno complicati che, all'atto pratico, possono risultare

non sempre veritieri. Tutto dipende dalla quantità di «materiale» presente nella figura e da quello «disponibile» nella mente per completare l'immagine stessa. Così si va da processi logici abbastanza semplici e, oltre tutto, univoci, come quelli della figura A, dove è indiscutibile che l'oggetto sia un parallelepipedo, a processi estremamente complessi, dove non è sempre possibile riconoscere immediatamente l'oggetto rappresentato (un esempio è quello della figura B, nota agli appassionati di enigmistica, dove viene rappresentato, visto dall'alto, un messicano, col sombrero, seduto su una panca). Non solo, ma può accadere che certe figure, come quelle della figura C, siano interpretabili solo presupponendo la maggior parte della scena (per la cronaca il disegno, tratto dalla stessa opera della Boden, può essere interpretato come un soldato armato di lancia portata in spalla, seguito da un cane, ambedue per la maggior parte nascosti da una porta); ciononostante non è possibile vedere né cane né padrone. Ma il principio è valido per poter automatizzare la visione, o ci sarà sempre la possibilità di interpretare in maniera differente la figura?

Di fronte a tali domande tutte le belle teorie circa i solidi regolari, ben illuminati e visibili, di Guzman, Roberts e soci vanno educatamente a farsi benedire. Il problema sta, forse un poco semplicisticamente, nel fatto che i programmi finora esaminati sono poco interattivi. In effetti un programma destinato alla visione non deve solo aggiungere linee, dove sono invisibili, in base a schemi verificati e precisi, ma deve altresì funzionare da propositore, all'utente od ad un sistema esper-



to ben costruito, di serie di probabili linee, spigoli, vertici, anche in funzione di uno specifico «tema» del soggetto o della scena osservata. Come è possibile organizzare un programma del genere? A dire la verità un programma del genere già esiste e può essere considerato, a tutti gli effetti, l'insieme dei migliori algoritmi di riconoscimento propri dei programmi di Guzman e Roberts. Si tratta di un programma, definito Interpret, sviluppato da Gilbert Falhs (G. Falhs, «Interpretation of imperfect Line Data as Three-Dimensional Scene», Artificial Intelligence, 3 (1972), pag. 101-104) come parte di un grosso progetto finalizzato alla realizzazione di un occhio artificiale allo Stanford Institute. Lo spirito con cui fu progettato, molto ambizioso e solo in parte realizzato, fu quello di costruire una mappa tridimensionale efficace di una immagine (ovviamente bidimensionale), eliminando gli inconvenienti già visti negli algoritmi di «line finding» che avevano afflitto i lavori precedenti.

«Interpret» è formato da sei moduli, che funzionano sequenzialmente. Si parte da Segment, versione personalizzata del programma di Guzman, perfezionato nella ricerca ed ipotizzazione delle linee nascoste. Il secondo modulo, Support, molto specializzato (affronta un importante argomento che i programmi predecessori avevano completamente trascurato) decide, per ogni corpo così individuato se e come esso è appoggiato su un altro, partendo dal presupposto, comunque poco limitativo, che nessuno dei corpi visibili sia sospeso, o attaccato alla faccia di qualche altro corpo.

Il modulo successivo, Complete (questo modulo fu quello più soggetto

a successive modifiche ed aggiornamenti), completa le linee nascoste se incontrovertibili, e trasferisce il controllo al modulo successivo, Recognize, che verifica la rispondenza di ogni corpo separato con nove modelli standard (definiti dall'autore «Prototipi») e che, inoltre, localizza i corpi nello spazio, individuando la loro posizione reciproca. Il tutto viene passato ad un modulo successivo, Predict, che ricostruisce, in base ai dati messi a disposizione dai programmi precedenti, una figura immaginaria perfetta, cancellando le linee nascoste.

C'è da evidenziare una cosa; Recognize è un modulo abbastanza intelligente, in quanto, pur non tollerando grosse imperfezioni, riesce a riconoscere linee non perfettamente delineate o parzialmente in ombra o confuse.

Infine l'ultimo modulo, Verify, riunisce il lavoro eseguito e cerca di dare un senso alle figure analizzate, verificando la completezza e la logicità delle stesse. Se ciò non avviene, si torna al modulo Recognize, per una nuova interpretazione della scena stessa.

Vediamo come, all'atto pratico, funziona il programma, applicandolo alla figura A. Giunti al terzo stadio, Complete possiede tre tool per rendere logica la figura: Join, Addcorner e Addline.

Come gli stessi nomi fanno capire,

Join unisce due segmenti giacenti sulla stessa linea, per completare lo spigolo tra due facce. Il secondo modulo cerca di ricostruire una confluenza logica tra tre linee, mentre il terzo esegue una analisi dei vertici per verificare se tra essi è possibile inserire una intera linea, destinata a completare una figura. Nell'esempio citato, Addline non ha ovviamente senso, né serve Join, che ha bisogno di due linee che si fronteggiano. Resta, ovviamente, Addcorner, che permette il completamento della figura nel modo più efficace.

Il punto più debole della catena, all'atto pratico, si è rivelato il modulo Recognize, per l'implicita debolezza della comparazione di profili virtuali (in tutto, neppure numerosi, essendo come abbiamo visto solo nove) con situazioni generali non sempre rigidamente inquadrabili. Interpret, quindi, ricade, ancora una volta, nelle problematiche di Obscene, pur se gli algoritmi di riconoscimento sono ben più raffinati di quelli dei programmi precedenti. Il metodo di analisi per analogia ha, quindi, dei suoi limiti insuperabili. È arrivato il momento di rovesciare il problema, vale a dire di far confluire dati nella rappresentazione prima di eseguire una interpretazione. Vedremo nella prossima puntata come ciò è possibile.