



## Modelli e tecniche d'interpretazione

La figura a) mostra una serie di oggetti; nessuno avrebbe difficoltà ad individuare, in essi, un parallelepipedo, un prisma a base triangolare adagiato su una delle facce laterali (un cuneo), ed un prisma a base concava, anch'esso adagiato su una faccia laterale. Non solo, ma non sfugge, ad una osservazione anche superficiale e frettolosa, che il parallelepipedo è sistemato dietro agli altri due solidi. Ma come riesce l'occhio (e la mente) umana ad individuare che si tratta di parallelepipedi e prismi, e come può stabilirne la prospettiva? Dopo tutto nulla, in un disegno sviluppato su un piano bidimensionale, come quello di un foglio di carta, possiede tali proprietà. E come è possibile stabilire che la figura del cuneo, a destra, è simmetrica e possiede facce opposte eguali?

«Felix monocolus in terra caecorum», dicevano i latini. E l'aforisma è qui valido per la macchina che dovrà leggere ed interpretare l'immagine! Ma chiariamo, innanzitutto, qualche aspetto della questione. Le informazioni visive ricevute da un sistema, o macchina, destinati alle operazioni di percezione, sono ricevute da questo mezzo come forme diverse di pattern (sfondi, retini), di differente colore (si veda la figura di Lincoln di due punte fa) e di differente intensità luminosa (si ricordi che abbiamo esclusa, finora, la visione a colori). Se questo input bidimensionale rappresenta una struttura, un oggetto, una scena tridimensionale, occorrerà che il sistema lettore-visivo stesso sia dotato di una certa, anche se rudimentale, attività e capacità interpretativa della scena stessa.

Tutto ciò comprende anche la capacità di «interpretare», vale a dire di individuare la presenza e le dimensioni, l'orientazione, e la disposizione reciproca degli oggetti nello spazio fisico

tridimensionale, con particolare riguardo (si veda ancora la figura a) precedentemente citata), a parti invisibili dell'oggetto stesso nascoste da altri corpi o da parti e facce dell'oggetto stesso, visibili, più prossime all'occhio. Poiché molte informazioni relative al mondo che ci circonda, e che vediamo, sono da noi derivate, ed a noi note anche se non direttamente presenti nell'immagine, il complesso apparecchiatura di visione-analizzatore dell'immagine deve avere conoscenze molto approfondite degli oggetti che gli vengono presentati per il riconoscimento, per ottenere risposte efficaci e rispondenti alla realtà. Questo principio è valido sia nella visione umana che artificiale: quanti di noi riconoscerrebbero una chiave dinamometrica od un albero a camme se non sono mai entrati in un'officina meccanica? E, ancora, come è possibile ricavarne da una immagine bidimensionale le dimensioni, anche solo relative, degli oggetti rappresentati?

Fin nei più vecchi programmi di trasformazione 2D->3D esistevano già procedure per il riconoscimento degli oggetti anche nelle loro esatte dimensioni e nella loro posizione, assoluta e, soprattutto, relativa. Il programma poteva già costruire ed «immaginare» i profili nascosti; il più efficiente programma di tal genere, prodotto da L.G. Roberts ed implementato come IDEOLOGY MACHINE (descritto da R.P. Abelson in «The structure of belief system», CMTL, pagg. 287-293, e, ancora, da Abelson e Reich in «Implicational Molecules: A Method of Extracting Meaning from Input Sentence», IJCAI-1, issue 37, pagg. 641 e succ.) era capace di tali performance in quanto ricostruiva contorni non dubbi di oggetti in base ad una dettagliata conoscenza della fenomenolo-

gia della geometria prospettica di strutture tridimensionali (soprattutto in relazione alla loro rappresentazione bidimensionale) ed una accurata messa a punto e sorveglianza dell'ottica della camera.

Il programma si serviva estesamente di complesse ed avanzate procedure di manipolazione numerica ed algebrica, in termini di matrici numeriche multidimensionali. L'immagine viene suddivisa in «domini», che mappano le parti dell'oggetto e le relazioni tra esse. Una maggiore definizione dello schema portava, ovviamente, a rendere più preciso, e, parzialmente, più intelligente il processo interpretativo; inoltre il programma, accanto all'elevato livello di dettaglio, introduceva una buona elasticità nella tolleranza delle aree di dubbia interpretazione.

Il programma di Roberts presupponeva che il mondo, al di là della camera, destinato ad essere «visto», è formato da tre classi di oggetti (v. fig. b), presenti nel dominio oggetto in forma semplice e composta; in pratica Roberts ammette e dimostra che qualsiasi oggetto può essere ridotto ad un prisma, rettangolare, cuneiforme, od a base esagonale; se si desidera maggior dettaglio, possono essere introdotti altri solidi, ma l'efficienza del programma si riduce e la necessità di memoria cresce esponenzialmente. Le definizioni di tali solidi (o forme primitive) sono inserite nel programma sotto forma di definizioni geometriche astratte, che descrivono tutte le superfici, gli spigoli, i vertici, senza particolare riferimento a misure ed orientazioni. Alle definizioni così codificate si associano due operatori particolari, «regole», destinati a modificare ed applicare le differenti modalità di rappresentazione degli oggetti presenti nel modo reale alle definizioni schematiche numeri-

che presenti come patrimonio iniziale del programma stesso.

Utilizzando, come già abbiamo detto, le leggi e le formule della geometria prospettica, il programma analizza gli oggetti e tenta una assimilazione della immagine ricevuta con i modelli di base originari. Questa operazione, in una con la valutazione delle dimensioni dell'oggetto e della distanza dello sfondo (l'operazione è piuttosto semplice, se si tien conto che la valutazione degli ingombri è diretta funzione degli angoli di visuale, con relativa correlazione con la distanza focale della macchina lettrice stessa) consente di assegnare un oggetto (od una par-

te semplice di esso) ad una delle forme elementari appena descritte. La figura può essere così scissa nei suoi particolari costruttivi ed essere successivamente oggetto di interpretazione.

Il programma di Roberts, ovviamente, doveva accettare alcune restrizioni, la cui assenza avrebbe, ovviamente, complicato notevolmente il già arduo algoritmo destinato al riconoscimento. Proprio perché soggetta alle leggi della prospettiva, la visione tridimensionale presenta notevole complessità, che non può essere del tutto implementata in un programma, ancorché complesso. Tanto per intenderci, la figura d) mostra come le leggi della pro-

spettiva in 3D deformino un prisma in modo tale che solo una delle facce visibili appaia come un parallelogramma. Esasperando il principio, le facce laterali, per un prisma abbastanza lungo, sarebbero dei triangoli, (e questo cozzerebbe con alcuni degli assiomi fondamentali del programma di Roberts, che impone che, ad una figura contenente un triangolo, corrisponda almeno un cuneo). Nella prossima puntata vedremo come è stato possibile aggirare l'ostacolo, utilizzando una tecnica di parsing, d'analisi, semplice ma efficiente, che si basa su due fasi, complementari e successive, di interpretazione dell'immagine. **MC**

disegno ricavato da Tippet & altri: «Optical and electro-optical Information Processing», Cambridge, Mass.; MIT Press. 1965.

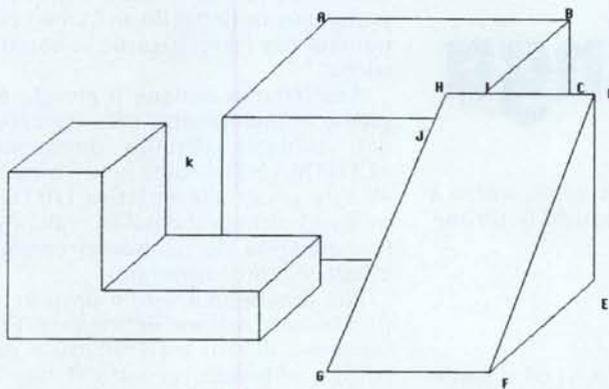


Figura a.

disegno ricavato da Tippet & altri: «Optical and electro-optical Information Processing», ibid.

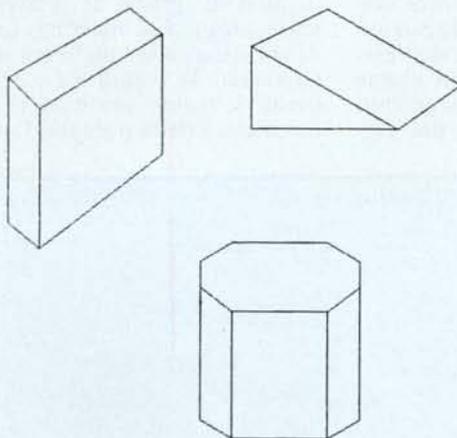


Figura b: i poliedri ammessi dal programma di L.G. Roberts (v. testo).

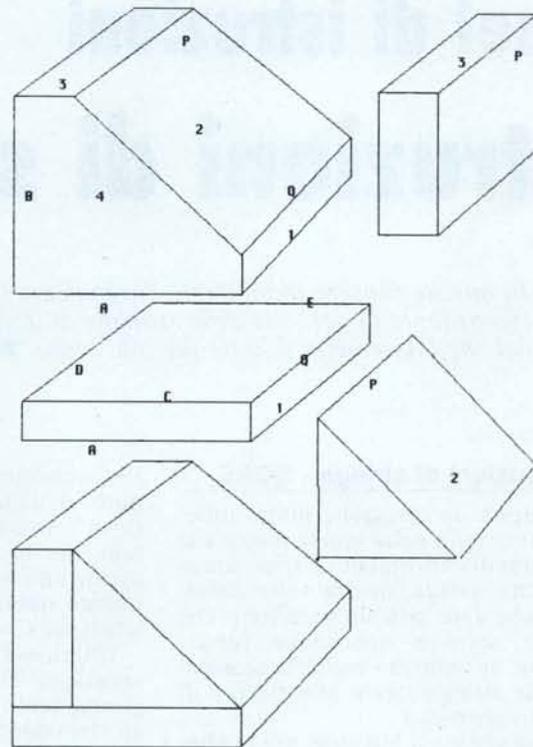


Figura c: modalità di scomposizione di poliedri in altri più semplici per la bibliografia d'origine si vedano le figure precedenti.

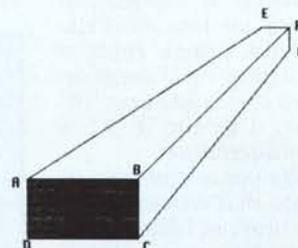


Figura d: la figura è un parallelepipedo, deformato dalla prospettiva. Sola la faccia campita appare come un rettangolo mentre le laterali divengono addirittura triangoli se il parallelepipedo è sufficientemente lungo. In questa ottica occorre modificare l'algoritmo di riconoscimento perché la macchina «veda» un parallelepipedo.