



Terza parte

La visione

Perché la visione divenga un nuovo mezzo attraverso cui il computer possa, in modo ancora più efficace, contribuire a risolvere i nostri problemi ed a sollevarci da lavoro inutile e stressante, occorre che essa sia davvero efficace, ed esca dal limbo degli studi teorici e delle esercitazioni su modelli. Il grande problema cui va incontro, infatti, l'organizzazione della visione è la capacità di far percepire alla macchina non l'immagine (cosa questa di non grande difficoltà), ma il significato, il valore, il senso dell'immagine da essa percepita. Il vero guaio, alla fin fine, è che alla visione partecipano certe sfumature che ben difficilmente sarebbero codificabili da favore di una macchina. Ad esempio, una cosa è il riconoscimento di una immagine come il volto di una persona, ma ben diversa sarebbe la complessità del problema se, dallo stesso volto, la macchina dovesse estrarre il significato delle emozioni che lo percorrono: amore, gioia, serenità, cruccio, dolore, ecc. In effetti si dovrebbe assegnare, alle capacità sensorie del calcolatore, una finezza tale, una raffinatezza di algoritmi di riconoscimento, ben difficilmente realizzabile allo stato odierno delle conoscenze. Il vero problema, inoltre, è che quasi mai la comprensione del valore e del significato di un oggetto è legato alla sua semplice immagine, ma sovente essa ha senso solo se inserita in un contesto rappresentato da altre immagini, che concorrono a dar significato ad una scena. Tanto per intenderci, una macchina destinata all'apertura della porta di casa, dovrebbe interpretare non solo l'immagine del padrone, ma anche, in maniera per quanto rozza, le sue intenzioni, senza continuare ad aprire e chiudere la porta senza senso, se il proprietario continua a passeggiare davanti alla porta di casa (cosa che avviene oggi passeggiando davanti ad una porta automatica di una stazione o di un supermercato). A questo punto bisognerebbe inserire un algoritmo del tipo: «Apri la porta se vedi il padrone con le chiavi in mano», ma ci sarebbe anche l'eccezione delle chiavi tratte di

tasca per cercare quella della macchina; allora bisognerebbe aggiungere una nuova condizione, tipo: «Deve avere le chiavi in mano ed essere rivolto verso la porta, né troppo lontano, (perché potrebbe, la cosa non avere senso), né troppo vicino, per non costringere il padrone ad aspettare come un salame che la porta decida di aprirsi». La macchina dovrebbe essere, allora, abbastanza intelligente da interpretare anche il significato della scena che sta guardando, e non solo da riconoscere gli oggetti.

La strada della interpretazione delle scene a mezzo delle immagini è lunga e magari ne parleremo qualche volta più a lungo: adesso invece siamo ancora alle fasi iniziali, dove ci interessa ancora parlare del riconoscimento intrinseco delle immagini, inteso come capacità di correlazione non dubbia (o almeno il meno possibile dubbia) tra immagine ed oggetto, ancorché semplice.

Il vocabolario illustrato della lingua italiana di Devoto-Oli definisce la visione come «Osservazione accurata fatta allo scopo di trarre utili informazioni». Queste «informazioni» di cui alla definizione possono essere generalizzate come luminosità dell'oggetto, distanza, colore, caratteristiche della superficie dell'oggetto stesso. Tutte hanno la loro importanza, ma poiché non è possibile combattere diversi avversari contemporaneamente, occorrerà fare un distinguo posticipando od eliminando alcune esigenze, a favore di altre.

Ad esempio è possibile tralasciare il problema del colore ed indirizzare tutti gli sforzi sulla brillantezza, intensità, luminosità dell'oggetto. D'altro canto il colore solo di rado contribuisce in maniera determinante e irrinunciabile al riconoscimento di un oggetto; in ultima analisi siamo andati avanti con i televisori in bianco e nero per tanti anni senza aver avuto grossi problemi di comprensione delle immagini che guardavamo. Una foto in bianco e nero, come generico caso, costituisce se vogliamo, una banale rappresentazio-

ne su carta della brillantezza di una immagine: essa manca di colore, di profondità (che ai nostri occhi viene fornita dalla sensazione stereoscopica data dalla sovrapposizione, come è noto, di due immagini leggermente dissimili), di moto e di informazioni circa la natura fisica degli oggetti che rappresenta; in altri termini manca di tutti gli attributi essenziali che avevamo descritti precedentemente, tranne quello della luminosità. Ciononostante è una immagine facilmente interpretabile.

Il problema della visione, in analogia a quanto appena detto, è stato inizialmente legato al problema della lettura della brillantezza di una immagine. Data per scontata la presenza e l'utilizzo, da parte della macchina, di attrezzature capaci di inviarle, in forma digitale, rappresentazioni analogiche estratte da un dispositivo di lettura ottica, tipicamente una telecamera, stabiliamo, come assioma, e parzialmente come stessa tesi, che il calcolatore sarà capace di vedere quando sarà in grado di descrivere il contenuto negli stessi termini con cui lo farebbe un essere umano e, ancora, sarà capace di utilizzare questa conoscenza nella successiva soluzione di nuovi problemi, siano essi di visione o non.

La risoluzione di un problema è, nelle metodologie umane, l'assimilazione del problema stesso a tecniche più facili, di cui sia già nota la tecnica di soluzione o, addirittura, la soluzione stessa. Ad esempio, in geometria, la ricerca dell'area di un poligono qualsiasi è trasformata nel calcolo delle aree di triangoli in cui il poligono è scindibile. Allo stesso modo, in meccanica, il progetto di un cambio d'automobile è niente altro che la soluzione di una serie di più piccoli problemi relativi ad ingranaggi collegati tra di loro. Ogni volta quindi che ad un problema è possibile sostituirne un altro (od una serie di altri), con caratteristiche risolutive o difficoltà di sviluppo più semplici del precedente, si è fatto un passo avanti nella risoluzione del problema. La tecnica di dissociazione del problema in altri più minuti ha inoltre il vantaggio innegabile di creare rami del problema stesso che possono essere trascurati o, viceversa, che pur nella loro semplicità possono portare un contributo essenziale alla risoluzione del problema stesso. Tanto per intenderci, nel riconoscimento della immagine di un'automobile può avere scarso significato il riconoscimento che ci siano due o quattro sportelli. E viceversa l'esame di una mano guantata, con sole quattro dita, assegna pressoché univocamente l'immagine analizzata al personaggio di Topolino.

Il riconoscimento di una immagine è legato, comunque, innanzi tutto alla

qualità ed al tipo della immagine stessa.

In questo caso complessità, qualità, definizione della immagine stessa giocano ruoli importanti nella facilità di interpretazione. Allo stato attuale della tecnologia è possibile spingere in maniera elevata l'accuratezza di alcune di queste qualità, prima tra tutte la definizione della immagine (anche se questa tecnica cozza in maniera pesante con l'esigenza di tenere bassa la quantità di memoria utilizzata). Ma ci sono altri problemi, non risolvibili solo attraverso l'aumento della definizione della immagine. Ad esempio è inutile aumentare il potere risolutivo del mezzo se l'immagine che si osserva è, ad esempio, uno schizzo, o magari la telecamera non è a fuoco. Ma non basta: i criteri appena elencati sono legati alla struttura singola dell'oggetto, nella sua interezza ed individualità, ma non è detto, anzi quasi mai è vero che venga affidato alla macchina il semplice riconoscimento di oggetti isolati: molto più frequentemente, invece, occorre tener conto di immagini di oggetti sovrapposti, parzialmente in ombra, o magari visti secondo prospettive strane (avete mai guardato una bottiglia dal fondo, od una macrofotografia di un semplice foglio di carta?). Ancora, occorre tener conto della tipologia delle immagini destinate al riconoscimento: infatti mentre questo non crea alcun problema per l'occhio (ed il cervello) umano, quali sono i criteri da definire ed assegnare ad una macchina per il riconoscimento di un oggetto che resta in parte fuori dalla visuale?

«Blade Runner», il capolavoro di Ridley Scott, ha lasciato intendere, tra l'altro, come i problemi del riconoscimento degli oggetti nelle immagini sia tanto difficile da non poter essere dato per scontato neppure in un futuro talmente avanzato da aver creato i «replicanti». Il protagonista, un Harrison Ford un po' più truce del solito, benché utilizzi una sofisticatissima macchina (obbedisce ad ordini a voce, cosa già ben difficile, come vedremo, da mettere a punto) capace di analizzare una fotografia (prevedibilmente olografica) e di mostrarne anche le parti nascoste dalla sovrapposizione di altri oggetti (una colonna), è chiamato sempre in causa dalla macchina stessa per il riconoscimento finale dell'oggetto nascosto nella stessa immagine.

Poiché, sia per una macchina che per una persona, il problema va affrontato, all'inizio, in una più semplice dimensione, per poi passare a fasi più sofisticate, il fatto che il problema abbia tante sfaccettature diverse (e in altre parole, tanti modi di essere affrontato), ha portato diversi ricercatori ad affrontare il problema in maniera e sotto ottiche diverse. Il risultato è sta-

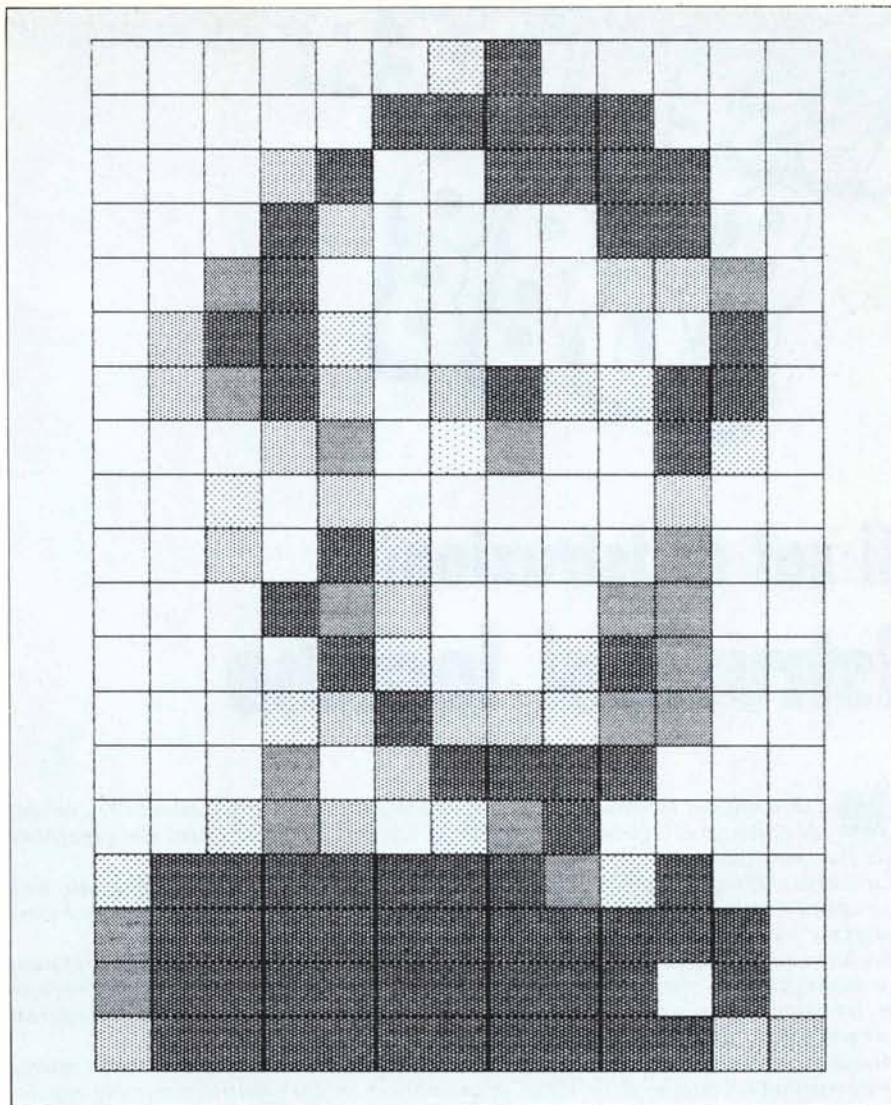


Figura a - Esempio di immagine di volto a bassa risoluzione, ma ancora riconoscibile. Da «Bertram Raphael»: Il Computer che pensa» a sua volta ricavato da alcuni esperimenti di Leon Harmon, Bell Telephone Laboratories.

to che si sono create delle «scuole», delle «correnti», a tutto svantaggio della omogeneità e della confrontabilità dei risultati.

D'altro canto lo stesso fatto di affrontare il problema secondo strade diverse ha portato allo spezzettamento delle energie anche nel campo della progettazione dei tool complementari, destinati a mettere a disposizione del computer i dati destinati a riconoscere l'immagine.

In altre parole, il ricercatore che punti le sue energie sulla definizione dell'immagine sarà interessato principalmente a sviluppare hardware dotato di memoria potente e facilmente manipolabile; al contrario chi sta studiando il problema dell'accuratezza avrà ovviamente a cuore macchine che rendano, per quanto possibile, l'immagine netta e precisa (negli studi iniziali

secondo tale tendenza furono utilizzati disegni tracciati con penne a china). Ancora, a chi interessa la resa (si intende come resa la definizione dell'immagine come rappresentazione più o meno accurata dei particolari dell'oggetto rappresentato) starà a cuore la formazione di «tipi» di confronto, da riferire all'immagine principale (l'applicazione più comune di tale problema è il riconoscimento della parola scritta, con la ricerca univoca dei tipi dattilografici e tipografici da riferire inequivocabilmente alla lettera stabilita). Come si vede, di rebus da risolvere ce n'è parecchi; dalla prossima volta vedremo come è possibile affrontare il problema, da un punto di vista generale, almeno nelle sue linee più semplici, e sperare di ottenere soluzioni soddisfacenti.