



La visione

La vista è quello dei nostri sensi che è stato maggiormente oggetto di studio nel campo dell'A.I.; queste note serviranno ad evidenziare le tecniche d'approccio al problema, i risultati ottenuti, e le prospettive future.

Studiare il processo della visione è, in altri termini, ricercare le modalità di trasduzione di fenomeni fisici in rappresentazioni simboliche, proprie del ricettore, sia un cervello umano o macchina. Pertanto lo studio del fenomeno può ricondursi a quello di una fonte di immagini (l'oggetto da osservare), di un mezzo destinato a ricevere ed analizzare le immagini, emesse (se così si può dire) dall'oggetto stesso, e di tutte le fenomenologie e le operazioni comprese tra questi due poli.

Generalizziamo il problema: la visione dipende da un occhio, un apparecchio destinato a catturare e mettere a fuoco la luce proveniente da un oggetto. Ogni immagine emette raggi luminosi in tutte le direzioni. Un occhio (con ciò intenderemo, d'ora in poi, sia il nostro organo che un qualsiasi generico complesso meccanico assolvante, più o meno completamente, la stessa funzione) è un mezzo, fornito di una lente e di un foro attraverso cui il raggio proveniente da un punto può essere indirizzato e messo a fuoco su una superficie.

L'insieme dei punti luminosi messi insieme in tal modo sulla superficie ricevente, sia essa la superficie interna del globo oculare, una pellicola fotografica o l'apparecchiatura di ricezione di una telecamera, rappresenta l'immagine. L'immagine così prodotta (lo ricorderemo, fin dai nostri studi di liceo) si presenterà capovolta e rovesciata ma ciò, per il nostro discorso, è almeno per il momento irrilevante. È invece più importante il fatto che l'immagine è bidimensionale, mentre l'oggetto è tridimensionale: inevitabilmente, l'immagine risulterà ambigua. Ciononostante, per la descrizione generica del problema che effettueremo in queste pagine, l'errore proveniente dalla generalizzazione del concetto può essere ritenuto trascurabile, e sottaciuto.

Contrariamente a molte problematiche della A.I., il meccanismo della visione può essere generalizzato in ma-

niera abbastanza ristretta. Vedere, per una macchina come per l'uomo, in una immagine, è individuare gli oggetti che la producono; se escludiamo, per semplicità, la visione a colori, il problema può essere ricondotto al riconoscimento della forma, della posizione e delle dimensioni degli oggetti che compongono l'immagine stessa. Non sempre tale definizione è, comunque, esauriente: ad esempio l'immagine di una spiaggia non è certo funzione del numero, della posizione o della forma dei granelli di sabbia.

Una immagine, secondo la definizione non a colori, può essere definita da una funzione rappresentata da livelli di grigio. Il livello varia da 1, in cui l'immagine è del tutto bianca ad un massimo di 0, con corrispondente immagine nera, con tutti gli innumerevoli livelli intermedi. Essa può essere intera inserita in un sistema di coordinate cartesiane, su una superficie bidimensionale; l'immagine stessa potrà essere pertanto inserita in una array bidimensionale, ad esempio [immagine (n1,n2)] in cui ciascuna cella [immagine (x,y)] conterrà la più piccola unità di registrazione, il «pixel» (che sta per «picture element»), costituita da un numero. Questo numero è la rappresentazione quantitativa numerica del livello di grigio del singolo pixel; ovviamente, poiché ogni cella della matrice contiene il livello medio di grigio del pixel, la scansione dell'immagine e la sua rappresentazione, anche simbolica, sarà più precisa quanto più piccole saranno le dimensioni del pixel di base. L'esempio in figura B è abbastanza semplificato; l'occhio umano ed una telecamera eseguono scansioni ben più precise e particolareggiate, ma il principio di base è lo stesso. L'unica differenza è che l'occhio è un analizzatore ben più raffinato, complesso ed efficiente di qualsiasi macchina; ma anche la membrana retinica, con i suoi ricettori e bastoncelli, ha un suo proprio limite di scanning e di parsing. Punti della immagine più piccoli dei bastoncelli non sono più visibili dall'occhio umano; l'immagine diviene microscopica. Questo strumento non fa altro che aumentare la grandezza del punto da vedere, in modo che esso possa essere intercettato

dai bastoncelli e fornire quella determinata informazione (e, in senso informatico, riempire la cella dell'array). È quanto abbiamo detto prima a proposito della sabbia della spiaggia: finché il granulo, anzi l'immagine di esso è più piccolo del ricettore stesso l'occhio non distinguerà il particolare; al contrario, avvicinando l'occhio alla sabbia od utilizzando un cannocchiale l'immagine virtuale sulla retina risulta allargata in modo tale che il granulo risulta intercettabile al ricettore e, quindi, visibile.

Il principio appena discusso, che assomiglia una retina ad una array bidimensionale, per quanto estesa e particolareggiata, è comunque limitativo del concetto di visione, non solo dal punto di vista dell'analisi dei particolari contribuenti alla immagine stessa. Altre informazioni, non solamente di scanning puro e semplice, concorrono alla visione; ne parleremo ancora nel corso di queste note. È importante, comunque, che per adesso, sia ben chiara la semplificazione generica tramite l'array.

Che l'immagine sia estremamente limitativa, comunque, come rappresentazione, nei confronti dell'originale (si consideri, quindi, quali sono i problemi impliciti nella ancora maggiore approssimazione connessa con il parsing, l'analisi della immagine, da parte di una macchina) è dimostrato dal fatto che non sempre la relazione oggetto-immagine è biunivoca. La rappresentazione in immagine manca sovente di informazioni preziose nei confronti dell'oggetto d'origine. Come appare in figura C, ogni punto della immagine corrisponde ad uno dell'oggetto. Sfortunatamente ci sono infiniti oggetti che possono produrre la stessa immagine. In figura, le coordinate del punto A, sono, nel mondo reale (x, y e z). Nella immagine (bidimensionale) permangono le corrispondenti coordinate x' ed y'; z' il parametro corrispondente al valore 2, la profondità all'oggetto, sparisce. La distanza dalla lente all'oggetto è la lunghezza focale. Per effetto della similitudine tra triangoli il rapporto tra x' ed y' è pari al rapporto tra i corrispondenti valori, nel mondo reale, ed il parametro z.

Pertanto dall'oggetto, poiché x, y e z sono sempre conoscibili, si perviene univocamente ad x' ed y'. Viceversa x' ed y' possono invece essere determinati da un infinito numero di valori di x e y. L'ambiguità può essere resa, ancora di più evidente, dall'esame della figura D, che mostra come una immagine, anche proveniente da un oggetto bidimensionale, dipenda non solo dalla grandezza, ma dalla sua profondità (distanza dal centro della lente, punto di ribaltamento dell'immagine), e dalla sua orientazione (angolo con uno

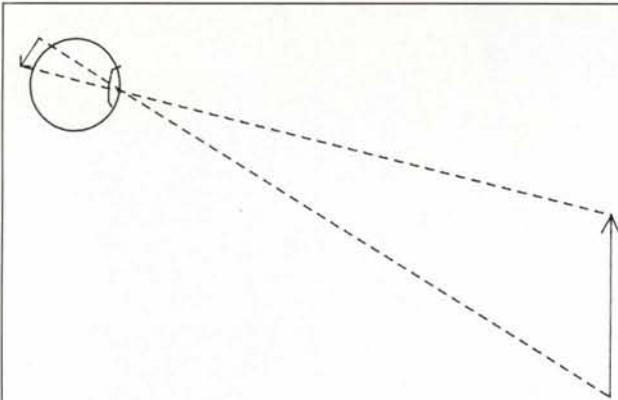


Figura A

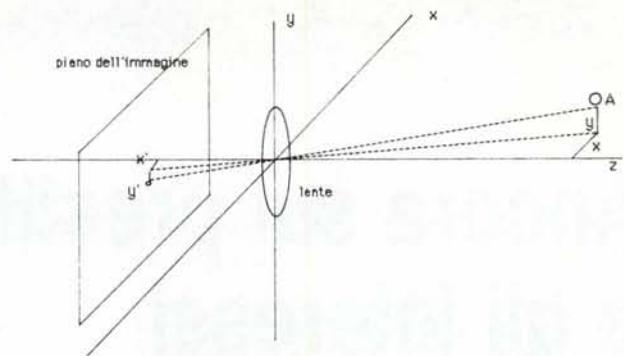
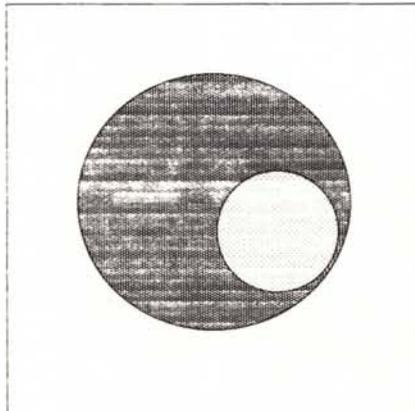


Figura C

Figura A - Meccanismo semplificato della formazione di una immagine.

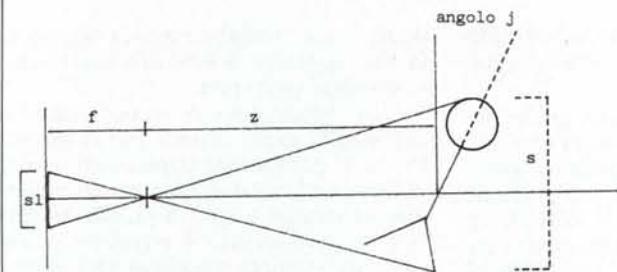
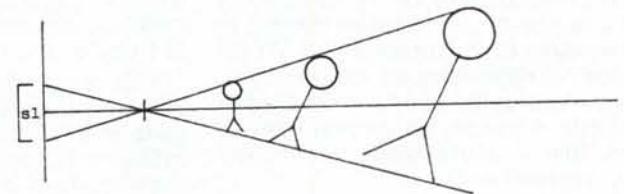
Figura B - Esempio di analisi di una immagine quale array di numeri, ognuno funzione del livello di grigio della immagine stessa.

Figura C - Ambiguità delle dimensioni relative dell'immagine nei confronti dell'oggetto.



0	0	0	0	0	0	0	0	0	0
0	0	0	0	.2	.2	.1	0	0	0
0	0	.2	.7	1	1	.6	.4	0	0
0	0	.7	1	1	1	1	.9	.2	0
0	.3	1	1	1	.8	.5	.7	.5	0
0	.2	1	1	1	.7	.4	.4	.4	0
0	0	.7	1	1	.6	.4	.4	.1	0
0	0	.3	.6	1	1	.7	.3	0	0
0	0	0	0	.1	.1	0	0	0	0
0	0	0	0	0	0	0	0	0	0

Figura B

Figura D₁Figura D₂

degli assi). Le variabili, in questo caso, sono f (la distanza focale), z (la distanza dell'oggetto), S (l'altezza) e l'angolo j (l'orientazione). Se si cerca di analizzare l'oggetto partendo dalla immagine, si ottiene una equazione con più variabili, che ammette infinite soluzioni; la figura D2 mostra, in pratica, alcuni possibili oggetti, tutti determinanti uguali dimensioni dell'immagine. Tanto per fare un esempio provate a fotografare piuttosto da vicino un po' di sottobosco o una superficie ricoperta da ghiaia, e guardatela senza pensare all'originale; non è possibile sapere se state osservando alberi, montagne, dirupi, o semplici sassolini ed arbusti.

Se escludiamo, come abbiamo detto in precedenza il colore, l'immagine è rappresentata solo da toni di grigio diversi. L'albedo, che rappresenta la quantità di luce riflessa dall'oggetto

stesso, è funzione di una serie di variabili rappresentate da:

- coefficiente di riflessione, funzione principalmente del colore dell'oggetto stesso;
- brillantezza della fonte luminosa;
- angolo di incidenza della luce illuminante l'oggetto, che può essere, comunque, sommatoria e funzione di diverse sorgenti luminose;
- orientazione della superficie illuminata nei confronti dell'occhio (o del mezzo destinato a leggere l'immagine stessa).

Alcune superfici, come gli specchi, riflettono praticamente tutta la luce ricevuta. Altre superfici hanno la capacità di riflettere la luce incidente in tutte le direzioni, e vanno sotto il nome di superfici di Lambert. In base a quest'ultima definizione, ed alle variabili precedentemente elencate, una

sfera, anche se uniformemente illuminata, apparirà, all'occhio dell'osservatore, composta di vari toni di grigio (funzione, soprattutto, della orientazione della superficie rispetto all'occhio dell'osservatore).

Ma, poiché l'immagine non tiene conto né evidenzia l'influsso di tali variabili, come è possibile distinguere se l'immagine di una sfera è ottenuta da un originale grigio fortemente illuminato o da un originale bianco in penombra?

Abbiamo, in questa puntata, affrontato ed appena sbizzato il problema; la cosa, come abbiamo appena intravisto, appare ben più complessa di quella che sembrerebbe a prima vista. Vedremo, la prossima volta, come il problema potrà essere affrontato dal punto di vista di una macchina.

MC