

# Visione delle macchine: una sfida anche per i matematici

Massimo Ferri

Dip. di Matematica, Piazza di P.ta S. Donato, 5 40126 Bologna

<http://vis.dm.unibo.it/>

## 1 Introduzione

Come mai siamo nel 2001 eppure l'intelligente — benché folle — calcolatore HAL di “Odissea nello spazio”<sup>1</sup> è ancora fantascienza? Come mai compiti faticosi ma apparentemente semplici come la pulizia della casa sono ancora strettamente pertinenza di operatori umani? Come mai un programma per calcolatore ha già potuto battere il campione mondiale di scacchi [7], ma un programma che trovi in Internet tutte le immagini che contengono, per esempio, un cavallo è ancora di là da venire [9, 18]?

Una risposta semplice, ma insoddisfacente, potrebbe essere la necessità di ancora maggiore potenza di calcolo. In effetti, il travolgente progresso nella velocità e nella disponibilità di memoria è stato senz'altro un ingrediente del successo di *Deep Blue* contro il campione di scacchi Kasparov. Gian Carlo Rota soleva ricordare che negli anni '60 il progetto di intelligenza artificiale che appariva più ambizioso era proprio quello di un programma che giocasse a scacchi, mentre sembrava più modesto il compito di capire ciò che viene percepito dagli occhi o da una telecamera, compito tranquillamente assolto da un bambino normale di pochi mesi. Invece, per usare le sue parole, “*quando viene richiesta un'ampia capacità di memoria, un computer avrà risultati migliori di un essere umano; tuttavia qualunque compito che coinvolga anche in minima parte la comprensione effettiva risulta tremendamente difficile per un computer. Provate a costruire un computer che riconosca se un uomo indossa un'uniforme!*” [28].

---

<sup>1</sup>“2001: odissea nello spazio”, regia di S. Kubrick, MGM Inc., 1968. Secondo la sceneggiatura del film, l'elaboratore HAL 9000 sarebbe entrato in servizio nel 1992.

Quando spunta la locuzione “intelligenza artificiale” è difficile non farsi catturare dalla domanda “ma può un calcolatore realmente pensare?” [29, 6, 27, 11]. È una domanda fondamentale ma insidiosa: i dibattiti che ne nascono appassionano ed arricchiscono chi li segue, ma spesso rischiano di paralizzare un effettivo progresso. Per questo, oltre che per la mia più limitata competenza, confinerò questa rassegna alla sola visione artificiale: cosa significa “vedere” per un calcolatore? Quali sono le difficoltà che s’incontrano nell’estrarre informazioni da immagini? In particolare: quali informazioni conviene estrarre? Una volta estratte delle informazioni, come possiamo organizzarle in una effettiva comprensione delle immagini? Come può una comprensione anche limitata essere utile per scopi pratici circoscritti?

In mezzo a tante domande, spero che emerga una possibile risposta alle mie domande iniziali: forse occorrono strutture. Intendo strutture simboliche manipolabili da un calcolatore, ma tanto varie e versatili da fornire modelli per gli innumerevoli aspetti della percezione della realtà. Infatti solo un’adeguata percezione della realtà, per quanto parziale e mirata, costituisce la base necessaria per interagire con essa. Confido, fra l’altro, di condividere col lettore l’emozione che, come insegnante, provo nel vedere utilizzi imprevisti degli strumenti più classici e consueti, nonché lo stimolo che, come ricercatore, ricevo da questo studio: la robotica, e in particolare la visione delle macchine, offre problemi difficili e multiformi che attendono non solo una soluzione, ma ancor prima una formalizzazione e forse lo sviluppo di nuovi ambiti matematici.

## 2 Scomporre un’immagine

Anni fa, dovendo descrivere il processo della visione, mi sarei soffermato sulla formazione delle immagini sulla retina, passando per l’iride e il cristallino, ma non sarei andato più in là. Non mi sarei reso conto di descrivere l’*ottica* invece della *visione*. La differenza può essere illustrata dalla Tabella 1 e dalla Figura 1.

Un’immagine *digitale* (cioè numerica, discreta) consiste usualmente in una matrice di interi come quella della Tabella 1: i numeri rappresentano i valori, approssimati in questo caso ad interi compresi fra 0 e 255, dell’intensità luminosa percepita da una matrice di sensori, l’analogo della retina in una telecamera digitale <sup>2</sup>. Come me, molti probabilmente non riescono a trarre dalla Tabella 1 informazioni interessanti sull’oggetto ritratto dalla telecame-

---

<sup>2</sup>Il colore invece viene codificato mediante una terna di numeri.

117	125	133	127	130	130	133	121	116	115	100	91	93	94	99	103	112	105	109	106
134	133	138	138	132	134	130	133	128	123	121	113	106	102	99	106	113	109	109	113
146	147	138	140	125	134	124	115	102	96	93	94	99	96	99	100	103	110	109	110
144	141	136	130	120	108	88	74	53	37	31	37	35	39	53	79	93	100	109	116
139	136	129	119	102	85	58	31	41	77	51	53	53	33	37	41	69	94	105	108
132	127	117	102	87	57	49	77	42	28	17	15	13	13	17	41	53	69	88	100
124	120	108	94	72	74	72	31	35	31	15	13	15	11	15	13	46	75	83	96
125	115	102	93	88	82	42	79	113	41	19	100	82	11	11	17	31	91	99	100
124	116	109	99	91	113	99	140	144	57	20	20	15	11	15	17	63	87	119	124
136	133	133	135	138	133	132	144	150	120	24	17	15	15	17	20	115	113	88	150
158	157	157	154	149	145	133	127	146	150	116	35	20	19	28	105	124	128	141	171
155	154	156	155	146	155	154	154	147	139	148	150	138	120	128	129	130	151	156	165
150	151	154	162	166	167	169	174	172	167	177	166	164	140	134	120	121	120	127	172
145	149	151	157	165	169	173	179	176	166	166	157	145	136	129	124	120	136	163	168
144	148	153	160	159	158	165	172	165	169	157	151	149	141	130	140	151	162	169	167
144	141	147	155	154	149	156	151	157	157	151	144	147	147	149	159	158	159	166	165
139	140	140	150	153	151	150	146	140	139	138	140	145	151	149	156	156	162	162	161
136	134	138	146	156	164	153	146	145	136	139	139	140	141	149	157	159	161	169	166
136	133	136	135	144	159	168	159	151	142	141	145	139	146	153	156	164	167	172	168
133	129	140	142	146	159	167	165	154	151	146	141	147	154	156	160	161	157	153	154

Tabella 1: La matrice di toni di grigio di un'immagine digitale.

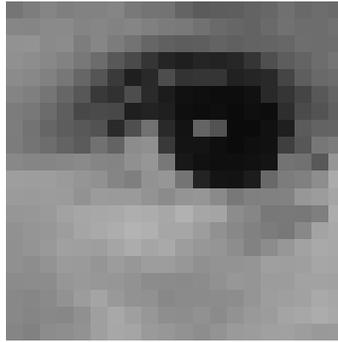


Figura 1: L'immagine corrispondente alla matrice della Tabella 1. Riprodotta da [32] per gentile concessione di Prentice-Hall, Inc. Upper Saddle River, New Jersey 07458.

ra. Eppure questo è il prodotto dell'ottica. Se invece rappresentiamo ogni numero con un rettangolino (*pixel*) uniformemente colorato con un grigio di chiarezza proporzionale al numero, otteniamo la Figura 1 che, soprattutto se osservata da una certa distanza, fa subito scattare nel nostro sistema visivo un complesso meccanismo di interpretazione: questa è la vera e propria visione. Si noti che la Tabella 1 e la Figura 1 contengono esattamente la stessa informazione, ed anzi nella figura l'informazione è meno certa, in quanto osservando un pixel non riesco a valutare esattamente il numero a cui corrisponde.

Il primo problema che si pone è quello di raggruppare pixel (o, equivalentemente, elementi della matrice) corrispondenti a parti in qualche senso omogenee; dualmente, possiamo invece cercare linee di demarcazione. Tutto ciò va

condotto in presenza di *rumore*, cioè inevitabili segnali spurî che inquinano l'immagine (per esempio macchie puntiformi).

## 2.1 Contorni

Una parete illuminata ci appare omogenea, ma se guardiamo attentamente vediamo che la sua luminosità presenta variazioni anche cospicue, ma gradualmente. Istintivamente poniamo il confine della parete dove la variazione è più netta. Individuare questi confini, o *contorni*, non è facile. Molta ricerca è stata svolta — e molta è tuttora in corso — su algoritmi di rilevamento dei contorni. Illustro ora il principio di uno dei primi modelli computazionali della detezione di contorni, basato sul sistema visivo umano [19].

Assumiamo temporaneamente un modello continuo dell'immagine, concepandola come un'applicazione  $h$  (almeno  $\mathcal{C}^2$ ) da un rettangolo  $I \times I'$  ad  $\mathbb{R}$ , che ad ogni punto associa il tono di grigio, inteso come un numero reale: l'immagine digitale è una discretizzazione di  $h$ . La Figura 2 mostra in alto una possibile sezione monodimensionale di  $h$ , intorno ad una transizione fra due zone piuttosto omogenee. È chiaro che il flesso di questa curva è il punto dove possiamo ragionevolmente porre il confine fra le due zone. Purtroppo, nelle situazioni reali il rumore e soprattutto la discretizzazione non rendono agevole la determinazione diretta del flesso. Osserviamo la derivata prima della restrizione considerata (illustrata subito sotto); naturalmente il minimo locale della derivata corrisponde al flesso cercato; neanche il minimo, però, si determina facilmente. Per individuare il punto, possiamo tracciare la derivata seconda (in basso): il minimo locale della derivata prima corrisponde ad un passaggio repentino della derivata seconda da valori negativi a valori positivi. Questo "attraversamento dello zero" della derivata seconda, finalmente, è facilmente reperibile anche nella discretizzazione ed in presenza di rumore.

Questo giustifica l'idea che, tornando alla dimensione due del dominio, i contorni vengano individuati come *attraversamenti dello zero del laplaciano* di  $h$ . Ma come calcolarli in un'immagine discreta? E come eliminare o almeno attutire il rumore, che tende a creare una grande quantità di falsi contorni?

Dato che il rumore corrisponde normalmente ad alte frequenze spaziali, si procede a sfocare l'immagine. Questo procedimento fisico equivale ad un'operazione tipica dell'analisi matematica: un prodotto di convoluzione. Si convolve  $h$  con una gaussiana  $g(x, y) = e^{-\frac{x^2+y^2}{2\pi\sigma^2}}$  con  $\sigma$  relativamente piccolo. L'algoritmo di ricerca dei contorni si dovrebbe a questo punto applicare

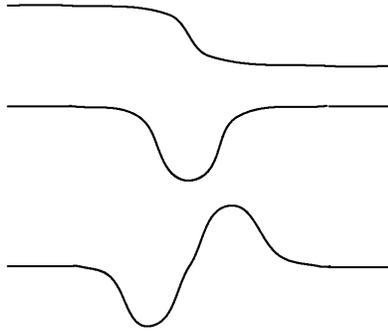


Figura 2: Una sezione di  $h$  e le sue derivate prima e seconda.

all'immagine sfocata. Ma qui interviene un teorema classico: il laplaciano della convoluzione, cioè  $\nabla^2(g * h)$ , è uguale alla convoluzione di  $h$  con il laplaciano della gaussiana, cioè  $(\nabla^2 g) * h$ . Quindi una sola convoluzione permette contemporaneamente di attutire il rumore e determinare i contorni!

Tornando alla realtà discreta dell'immagine digitale, la convoluzione viene effettuata sostituendo ad ogni elemento della matrice una media pesata degli elementi di un suo intorno quadrato, dove i pesi sono valori approssimati del laplaciano della gaussiana. La Figura 3 mostra un'immagine e il risultato del procedimento descritto con due diversi valori di  $\sigma$ .



Figura 3: Un'immagine (a sinistra) e i suoi contorni con  $\sigma = 1$  pixel e  $\sigma = 2$  pixel (a destra). Riprodotta da [32] per gentile concessione di Prentice-Hall, Inc. Upper Saddle River, New Jersey 07458.

## 2.2 Segmentazione per regioni

Con un atteggiamento in qualche senso complementare, si può cercare di raggruppare pixel in regioni omogenee. In questo tipo di *segmentazione*, l'algoritmo di gran lunga migliore utilizza il *funzionale di Mumford-Shah*. Si veda la Figura 4 per l'applicazione di questo metodo all'immagine della Figura 3 con diversi numeri di regioni.



Figura 4: Segmentazione con 3, 20 e 100 regioni.

Questa volta il modello continuo è un po' più realistico, in quanto la funzione  $h$  che rappresenta l'immagine è considerata costante a tratti. Il dominio è ancora un rettangolo  $R = I \times I'$  e i pixel ne formano una partizione (con opportune convenzioni per attribuire i segmenti di frontiera comune all'uno o all'altro pixel);  $h$  è costante su ogni pixel. Certo, però, non vogliamo identificare le regioni della segmentazione con le regioni su cui  $h$  è costante: l'immagine risulterebbe eccessivamente "sbriciolata". L'idea è allora di cercare una nuova funzione  $f$  e l'unione  $\Gamma$  di un insieme finito di curve (che saranno i contorni delle regioni) che minimizzino il funzionale

$$E(h, \Gamma) = \mu^2 \int \int_R (f - h)^2 dx dy + \int \int_{R-\Gamma} \|\nabla f\|^2 dx dy + \nu |\Gamma|.$$

Giocando sui parametri  $\mu$  e  $\nu$  si può dosare il rilievo dei tre termini, correlati rispettivamente alla distanza dalla funzione  $h$  (il dato in ingresso), alla ripidità della funzione soluzione  $f$  e alla lunghezza totale delle curve di contorno  $\Gamma$ .

Questo problema a discontinuità libere suscita tuttora ricerche di carattere teorico e numerico: gli spazi funzionali in cui determinare le soluzioni, l'analisi delle giunzioni e delle terminazioni delle curve di contorno, e natu-

ralmente i metodi di risoluzione costituiscono alcune delle sfide scaturite da questo problema eminentemente applicativo. Un vanto italiano è il lavoro, in quest’ambito, di De Giorgi [2] e della sua scuola.

Fra gli sviluppi di questo metodo, voglio accennare ad un funzionale, sviluppato da Mumford e Nitzberg [24], la cui minimizzazione conduce addirittura alla determinazione di regioni “nascoste”, cioè corrispondenti ad oggetti della scena occlusi da altri (come nella Figura 5 a sinistra) o addirittura illusorî (come il “Triangolo di Kanisza” della Figura 5 a destra).



Figura 5: Sbarre occluse (a sinistra) e il Triangolo di Kanisza (a destra).

### 3 Capire un disegno

Nel periodo in cui ci si rendeva conto della difficoltà di estrarre linee significative da un’immagine, ci si cominciava anche a chiedere come utilizzare le linee stesse per una effettiva comprensione dell’immagine. Una corrente di ricerca si occupò, allora, dell’interpretazione automatica dei *line drawing*, cioè dei disegni geometrici, normalmente in assonometria, rappresentanti oggetti poliedrici tridimensionali [12].

#### 3.1 Unicità

È il momento di accennare ad un problema generale della visione. Salvo il caso particolare delle *range image*, di cui parlerò più avanti, i dati che vengono trattati sono costituiti da una o più immagini bidimensionali, mentre la *scena*, cioè l’insieme degli oggetti raffigurati, è tridimensionale. Nel caso di visione monoculare, o di un disegno in assonometria, sono infinite le possibili scene che hanno originato una stessa immagine. L’unicità dell’interpretazione viene forzata dall’euristica o, in termini matematici, dall’ipotesi di vincoli aggiuntivi.

Un tipico esempio è l'interpretazione degli angoli. In effetti, il nostro ambiente abbonda di angoli retti, e anche se un angolo non appare retto in assonometria, siamo spesso portati ad interpretarlo come tale. Questo meccanismo, di cui siamo normalmente inconsapevoli, risalta in modo sorprendente nell'illusione ottica della *camera di Ames*, una grande scatola irregolare che viene però interpretata dal sistema visivo come parallelepipedo retto, anche se ciò comporta conseguenze paradossali [17, pag.124]. Questa euristica, opportunamente condizionata, fornisce soluzioni univoche e corrette nella stragrande maggioranza dei casi di applicazione concreta.

### 3.2 Esistenza

Anche imponendo tutti i vincoli che possano garantire l'unicità, gli studiosi dei line drawing si imbatterono presto in un grave problema: i “disegni impossibili”. Nobilitati artisticamente da Reutersvärd e soprattutto da Escher, questi disegni che non ammettono un'interpretazione come reali oggetti tridimensionali sono stati studiati anche da matematici, per esempio dai Penrose (v. Fig. 6) [26].

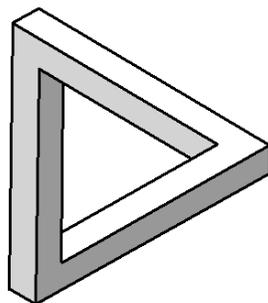


Figura 6: La “tribarra” dei Penrose.

Naturalmente, un qualunque programma per l'interpretazione dei line drawing avrebbe dovuto prima di tutto escludere i disegni impossibili. La ricerca che ne scaturì fu molto interessante. La maggior parte dei ricercatori tentò una classificazione di situazioni locali (convessità, occlusione, ecc.) puntando ad identificare l'esistenza di un'interpretazione con la possibilità di un'estensione globale di queste etichettature. Alcuni tentarono la traduzione delle relazioni locali in termini di logica proposizionale, mirando a riconoscere i disegni impossibili come contraddizioni.

Erano veramente numerose e varie le impostazioni matematiche con cui si

tentò di risolvere il problema. Infine, K. Sugihara escogitò la soluzione definitiva esprimendo le relazioni locali in termini di equazioni e disequazioni lineari [31]. Questo è uno dei rari casi in cui un problema di visione è stato completamente risolto. Si tratta pur sempre di un risultato eminentemente teorico, dato che nessun procedimento pratico di elaborazione di un'immagine reale produce un line drawing perfetto.

## 4 Forma da X

Anche se i problemi relativi ai line drawing sono sostanzialmente risolti, resta il fatto che ancora non si riesce ad estrarre da un'immagine qualcosa che assomigli ad un disegno geometrico. Inoltre sono molti gli oggetti non poliedrici di cui, comunque, il nostro sistema visivo riesce ad interpretare la forma. Quali sono gli indizi che sfruttiamo? Quali possiamo far utilizzare da una macchina?

C'è tutta una corrente di ricerca, in visione artificiale, chiamata *forma da X* ("shape from X"), che in realtà è formata da molti rivioli: l'obiettivo è quello di determinare la forma di un oggetto da particolari caratteristiche della (o delle) immagini. Ecco una lista, sicuramente incompleta:

- forma dalla stereoscopia,
- forma dal moto,
- forma dall'ombreggiatura,
- forma da messa a fuoco – sfocatura,
- forma dalla tessitura,
- forma da luce strutturata.

Alcuni di questi metodi non richiedono attività da parte dell'operatore durante la ripresa (per esempio la forma dalla stereoscopia), altri richiedono un intervento sui parametri del sistema ottico (per esempio la forma da messa a fuoco – sfocatura), altri ancora impongono un rilevante controllo sulla scena (in particolare la forma da luce strutturata).

Obiettivo comune di questi metodi è determinare un'applicazione che ad ogni punto (in realtà ad ogni pixel) dell'immagine associi o il valore di profondità (cioè distanza dall'osservatore) del punto  $P$  corrispondente nella scena, o la direzione della retta normale alla superficie dell'oggetto in  $P$ . Accennerò qui solo a tre temi.

## 4.1 Stereoscopia

Il primo metodo di rilevamento della forma che viene in mente è senz'altro la stereoscopia, cioè la fusione delle informazioni contenute in due immagini riprese da posizioni leggermente diverse. È meraviglioso come questo metodo ci fornisca un'impressione di immagine unica — e non doppia — vividamente tridimensionale. Vi sono diversi algoritmi che permettono ad un calcolatore di determinare una mappa di profondità a partire da una coppia di immagini.

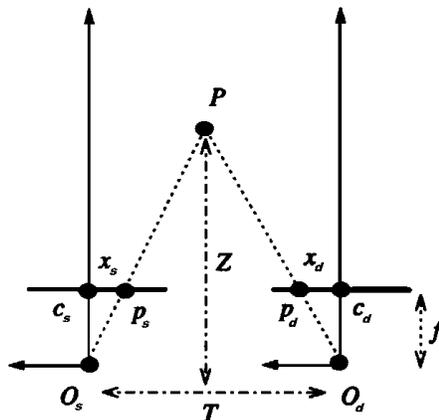


Figura 7: Un semplice apparato di stereoscopia. Riprodotta da [32] per gentile concessione di Prentice-Hall, Inc. Upper Saddle River, New Jersey 07458.

Il problema della stereoscopia può apparire ingannevolmente semplice. Immaginiamo un singolo punto luminoso nella scena, ripreso da due macchine fotografiche; per semplicità assumiamo che le pellicole delle due macchine fotografiche giacciono su uno stesso piano verticale, ad una stessa altezza (si veda la Figura 7, che rappresenta il sistema visto dall'alto). Se sovrapponiamo le due pellicole sviluppate, troveremo che le due immagini del punto si trovano su una stessa retta orizzontale, distanziate di una certa quantità  $D = x_d - x_s$ , detta *disparità*. Ora, se sono note la distanza  $T$  fra gli assi ottici e la lunghezza focale  $f$ , il calcolo della profondità  $Z$  è effettivamente molto facile:  $Z = f \frac{T}{D}$ .

Il guaio è che questo risolve solo uno dei due problemi della stereoscopia: l'altro è il difficile problema della *corrispondenza* (v. la Figura 8). Nelle scene reali i punti luminosi non sono uno, ma migliaia; inoltre, alcuni punti visti da una macchina fotografica non sono visibili dall'altra e viceversa.

Non mi soffermo sui diversi metodi statistici, variazionali o basati sui contorni, che si utilizzano per determinare la corrispondenza. Voglio invece

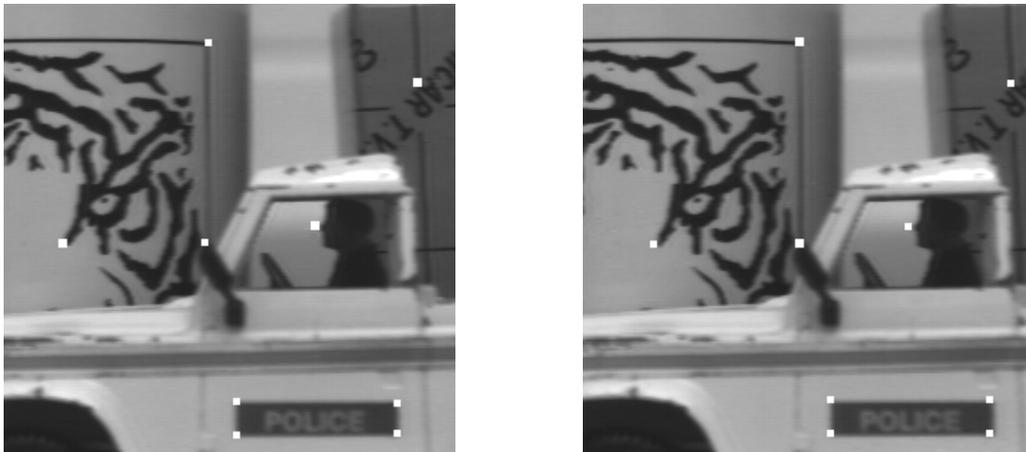


Figura 8: Una coppia di immagini stereoscopiche ed alcuni punti in corrispondenza. Riprodotta da [32] per gentile concessione di Prentice-Hall, Inc. Upper Saddle River, New Jersey 07458.

segnalare come la geometria permetta di ridurre la dimensione del problema da due ad uno mediante il *vincolo epipolare*. Si osservi la Figura 9, dove sono schematizzati i piani di proiezione (*piani immagine*)  $\pi_s, \pi_d$  di due macchine fotografiche<sup>3</sup>; gli assi ottici, per comodità, sono questa volta convergenti. La retta congiungente i due punti di vista  $O_s$  ed  $O_d$ , detta *asse epipolare*, interseca i due piani immagine in due punti detti *epipoli*. Per ogni punto  $P$  della scena e per l'asse epipolare passa allora esattamente un piano; questo sega i piani immagine in due rette, dette *epipolari*. La ricerca della corrispondenza si può dunque restringere a queste due rette. Le rette epipolari si corrispondono in una trasformazione proiettiva, caratteristica della configurazione, fra i due fasci centrati negli epipoli.

La robotica prende spunto dalla natura, ma non è obbligata a copiarla. Alcuni gruppi di ricerca hanno scelto di affidare completamente ai vincoli epipolari la determinazione della corrispondenza: usano la visione *trinoculare*, compiuta mediante tre macchine fotografiche disposte in modo che gli assi epipolari non coincidano.

---

<sup>3</sup>In effetti, i piani immagine dovrebbero essere *dietro* i punti  $O_s$  ed  $O_d$ ; quella schematizzata è piuttosto la configurazione del “velo” del Dürer [25]

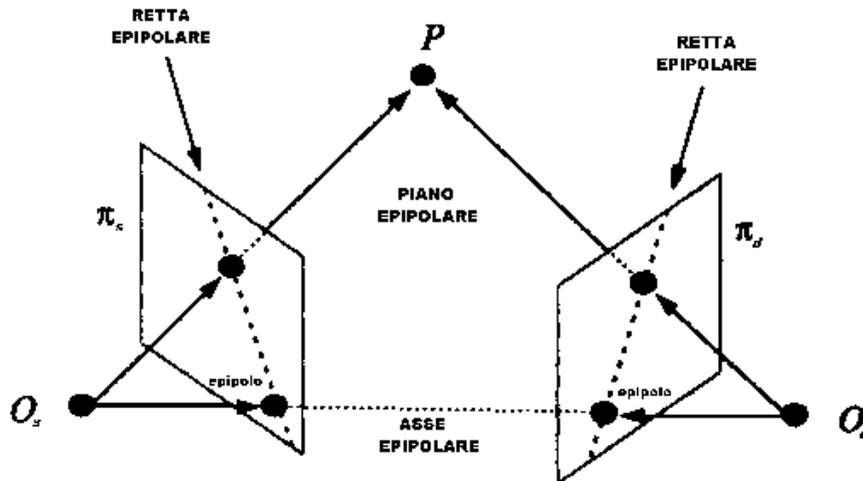


Figura 9: La geometria del vincolo epipolare. Riprodotta da [32] per gentile concessione di Prentice-Hall, Inc. Upper Saddle River, New Jersey 07458.

## 4.2 Ombreggiatura

Ben sapevano i pittori del Barocco (in particolare i bolognesi Carracci) che il chiaroscuro può produrre un'impressione di rilievo spettacolare. In effetti, il nostro sistema visivo può facilmente dedurre la forma di una superficie dall'ombreggiatura anche in una singola immagine, e perfino senza conoscere la direzione da cui proviene la luce<sup>4</sup>. Come si può utilizzare questa informazione in un sistema automatico?

La *radianza* di una superficie è la quantità di energia luminosa da essa emessa. Supponiamo che la superficie del solido ritratto sia *lambertiana*, cioè tale che diffonda ma non rifletta la luce. La radianza di una superficie lambertiana dipende solo dall'angolo d'incidenza della luce che la colpisce.

Ciò che viene registrato nei singoli punti dell'immagine è invece l'*irradianza*, cioè la quantità di energia luminosa ricevuta (anzi la sua discretizzazione, l'*intensità*, cioè il livello di grigio del singolo pixel). Con buona approssimazione l'irradianza può essere considerata proporzionale alla radianza. Dunque, per una fissata e nota sorgente luminosa, l'irradianza di un punto  $p$  dell'immagine pone un vincolo alle possibili rette normali alla superficie nel

<sup>4</sup>In realtà, se non conosciamo tale direzione la ricostruzione ha un'ambiguità intrinseca: il più delle volte essa è risolta euristicamente dal fatto che, riconoscendo l'oggetto, sappiamo se è concavo o convesso.

suo punto  $P$  di cui  $p$  è la proiezione: esse formano un cono quadrico.

Se venisse osservato il singolo punto  $P$ , la normale sarebbe indeterminata. Tuttavia possiamo ricavare i coni analoghi per tutti i punti di una regione (individuata dalla segmentazione) che possiamo supporre di classe  $\mathcal{C}^2$ . Se inoltre si rileva un punto in cui la normale sia nota (per esempio un punto interno alla regione, in cui la radianza sia massima), allora è possibile ottenere una mappa univoca delle normali ed anche la profondità relativa di ogni punto (come nella Figura 10). Infatti, data la profondità  $z$  di un punto di coordinate  $(x_0, y_0)$ , la profondità del generico punto di coordinate  $(x_1, y_1)$  si ottiene come  $z(x_1, y_1) = z(x_0, y_0) + \int_{(x_0, y_0)}^{(x_1, y_1)} (pdx + qdy)$ , dove l'integrale è calcolato lungo una qualunque curva da  $(x_0, y_0)$  ad  $(x_1, y_1)$ , e dove  $p$  e  $q$  sono, in ogni punto, i coefficienti dell'equazione  $z = px + qy + c$  del piano tangente. Risulta più semplice, comunque, risolvere le ambiguità illuminando la scena da tre direzioni non complanari.

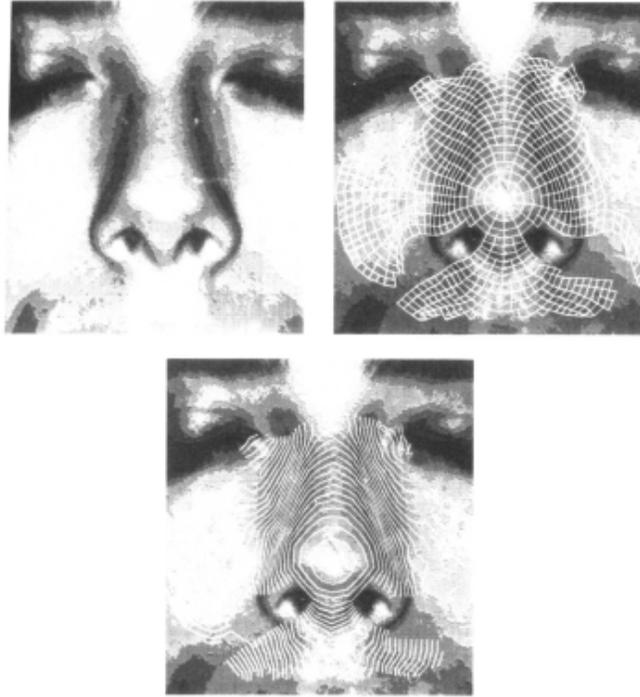


Figura 10: Forma dall'ombreggiatura. Riprodotta da B.K.P. Horn "Robot vision" per concessione di MIT Press, Cambridge, Massachusetts 02142.

### 4.3 Luce strutturata

Abbiamo appena visto un esempio di metodo *attivo*, in cui l'operatore interviene sui parametri, in questo caso le direzioni di illuminazione. Chiaramente, l'uso di metodi attivi ci allontana dal sogno di un sistema visivo di uso generale, che imiti in qualche modo le caratteristiche del sistema visivo umano. Peraltro, molto spesso le esigenze pratiche — per esempio in ambito medico, industriale o militare — richiedono sistemi di visione con capacità circoscritte a situazioni ben definite. Per esempio, se la superficie d'interesse viene esplorata in un laboratorio, nulla vieta che l'illuminazione venga totalmente controllata dall'operatore.

Questo è l'elemento essenziale dei tanti e vari metodi di *luce strutturata*. Tutti consistono nella proiezione di strutture note: spesso si proietta semplicemente una lama di luce con cui si spazza la superficie d'interesse. Più interessante è la proiezione di una griglia a maglie quadrate, come nella Figura 11 a sinistra. Anche in questo caso, come nella forma dall'ombreggiatura (e nella forma dalla tessitura, che non viene trattata in quest'articolo) si ottiene informazione sulla direzione della normale alla superficie in ogni nodo della griglia proiettata.

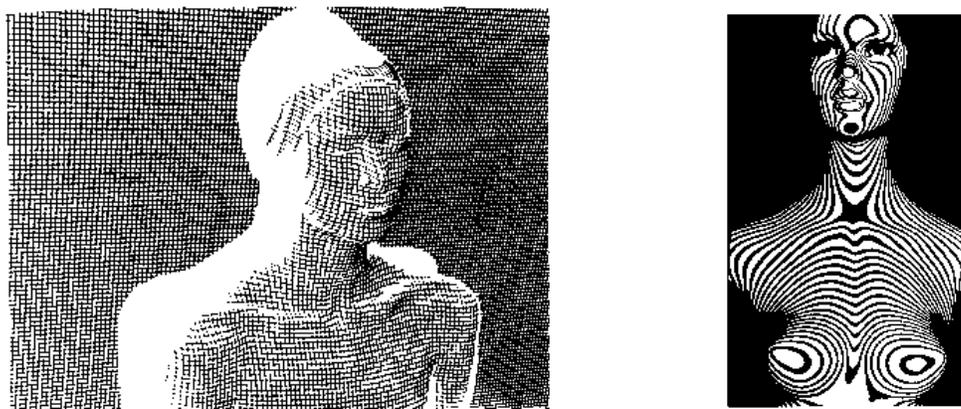


Figura 11: Proiezione di una griglia (a sinistra) e frange *moiré* (a destra)

Riprodotta da [14] per concessione di MIT Press, Cambridge, Massachusetts 02142.

Una variante notevole è costituita dalla produzione di frange *moiré* : Si proietta una griglia regolare, ma a maglie molto fini; un'analogo reticolo viene posto davanti all'obiettivo. Le frange d'interferenza che ne risultano sono linee di livello di una funzione di Morse (cioè una funzione a valori reali i cui punti critici sono non degeneri) definita sulla superficie (v. Figura

11, a destra) [14]. È sorprendente come, nonostante l'assoluta artificialità dell'apparato, il nostro sistema visivo riesca ad interpretare perfettamente la forma dei due manichini della Figura 11.

In un certo senso, rientra fra i metodi di luce strutturata anche l'uso del profonditàmetro laser (*laser range finder*), che fornisce direttamente la mappa di profondità (*range image*) della scena osservata con un metodo mutuato dal sonar.

## 5 Riconoscimento e stima dell'assetto

Può sembrare che, una volta ottenuta una mappa delle profondità o delle normali, il gioco sia fatto, e il riconoscimento di un oggetto sia banale. Purtroppo le cose non stanno così. Ci sono almeno due ragioni per cui una mappa di profondità, per esempio, richiede ulteriore elaborazione.

Una prima ragione è che la mappa di profondità che si ottiene da un'immagine è comunque imprecisa, parziale, e campionata in un modo che dipende strettamente dalla posizione dell'oggetto rispetto all'osservatore; perciò una sovrapposizione punto per punto della mappa ottenuta dall'immagine e del modello conservato in archivio è quanto meno ardua. Inoltre, l'oggetto d'interesse potrebbe non essere del tutto rigido, e la sovrapposizione potrebbe risultare semplicemente impossibile. Una seconda ragione è in termini di economia e di semplicità: i dati di profondità occupano memoria e sono poco significativi. Si preferisce, allora, schematizzare in qualche modo i dati desunti dall'immagine.

### 5.1 Scomposizione

Come primo esempio, si può operare una scomposizione (ulteriore rispetto alla segmentazione prodotta, per esempio, dal funzionale di Mumford–Shah) in regioni a curvatura media  $H$  e gaussiana  $K$  di segno costante (v. Figura 12). Ciò viene, per esempio, in modo naturale dall'analisi delle frange *moiré*, ma si ottiene abbastanza facilmente anche da una *range image*. Il grafo che ha un vertice per ogni regione (etichettato con il tipo di regione) e in cui vertici adiacenti rappresentano regioni confinanti è una descrizione sintetica ma significativa e resistente ad un ampio insieme di deformazioni.

	$K -$	$0$	$+$
$H -$			
$0$			
$+$			

Figura 12: Classificazione locale dalle curvatures  $H$  e  $K$ . Riprodotta da [32] per gentile concessione di Prentice-Hall, Inc. Upper Saddle River, New Jersey 07458.

## 5.2 Immagine gaussiana estesa

La *mappa di Gauss* di una superficie  $S$  bordo di un solido è l'applicazione che ad ogni punto ordinario di  $S$  associa il punto della sfera unitaria  $\mathbb{S}^2$  avente lo stesso versore normale esterno.

Nel caso del bordo di un poliedro, la mappa di Gauss ha come immagine un insieme finito di punti corrispondenti alle facce 2-dimensionali del poliedro. Chiamiamo *immagine gaussiana estesa* l'insieme dei punti di  $\mathbb{S}^2$  che formano l'immagine della mappa di Gauss, ognuno etichettato con l'area della faccia del poliedro corrispondente (nel caso non convesso, potremo avere più facce con la stessa immagine; useremo allora la somma delle aree). Un teorema di Minkowski [20] garantisce che, per un poliedro convesso, l'immagine gaussiana estesa identifica il poliedro a meno di traslazioni (anche senza informazioni sul numero di lati delle facce).

Da un'immagine discreta, ottenuta per esempio mediante forma dall'ombreggiatura, è possibile ricavare un'immagine gaussiana estesa (necessariamente discreta). Questa risulta più significativa della stessa mappa di profondità. Il confronto fra immagini gaussiane estese di un oggetto ritratto nell'immagine e un oggetto di cui si conservi il prototipo in archivio risulta allora agevole, veloce e sufficientemente sicuro. Inoltre, la rotazione che si deve imporre alla sfera per effettuare la sovrapposizione ci dà informazione sull'assetto dell'oggetto nello spazio. La conoscenza dell'assetto è indispensabile, per esempio, se vogliamo manipolare l'oggetto osservato.

Il lettore matematico può giustamente essere turbato dal fatto che si appli-

chino teoremi relativi a poliedri ad oggetti che tali non sono. Per fortuna, vale un'estensione del teorema di Minkowski, dovuta ad Alexandroff [1], al caso di bordi lisci di solidi convessi. In questo caso, nell'immagine gaussiana estesa l'area di una faccia è sostituita dalla curvatura gaussiana in un punto; in effetti, il procedimento descritto prima risulta una discretizzazione del procedimento ideale basato sul teorema di Alexandroff. La necessità di una discretizzazione il più possibile uniforme ha stimolato l'interesse dei ricercatori verso le tassellazioni della sfera.

### 5.3 Grafo e mappa delle viste

Qualunque sia il nostro punto di osservazione, avremo sempre una vista parziale dell'oggetto d'interesse. Per esempio, del cubo della Figura 13, da un certo punto di osservazione abbiamo solo la vista delle facce  $\bar{A}$ ,  $B$ ,  $C$ .

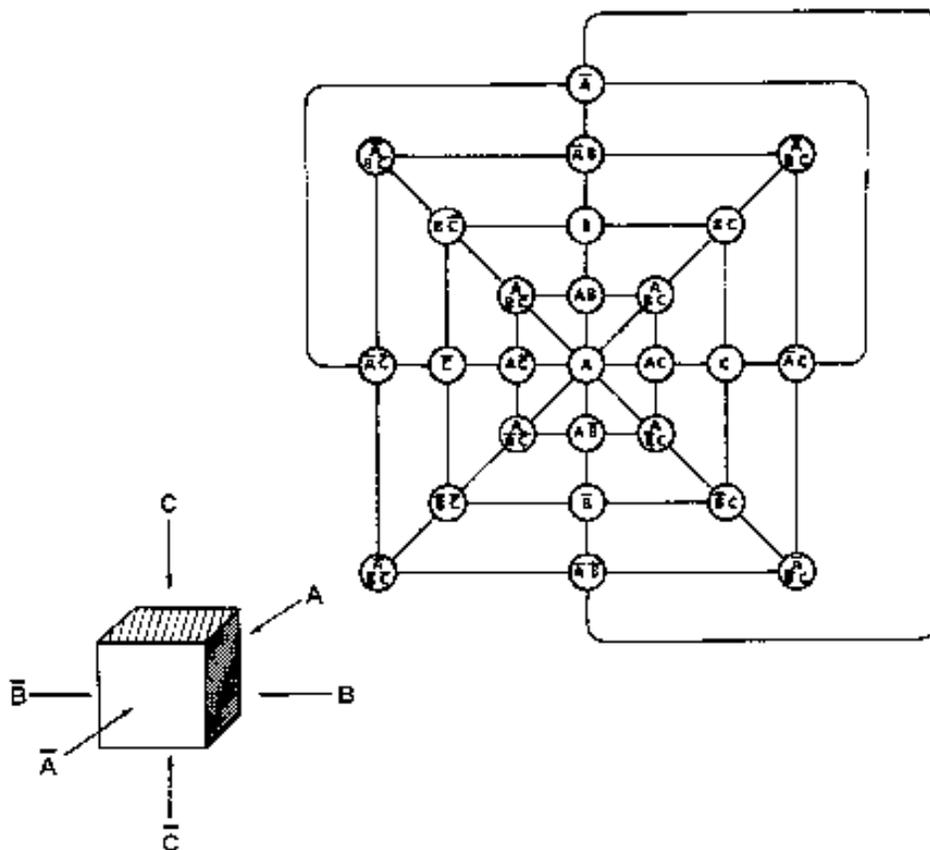


Figura 13: Il grafo delle viste di un cubo.

Per quasi tutti i punti di osservazione, se ci scostiamo di abbastanza poco, la vista non cambia, nel senso che vedremo le stesse facce, benché in modo un po' diverso. La trasformazione che lega due immagini di questo tipo è proiettiva a tratti. Possiamo allora scomporre lo spazio in classi di equivalenza entro le quali le facce viste sono le stesse. Per semplicità, possiamo considerare proiezioni ortogonali anziché centrali; le trasformazioni saranno affini a tratti e la scomposizione sarà effettuata su una superficie sferica di riferimento (in cui ogni punto rappresenta una direzione con verso). Ogni classe si può allora rappresentare come vertice di un grafo (etichettato dalla vista corrispondente), collegando due vertici se si passa con continuità da una classe all'altra: il *grafo delle viste* (v. la Figura 13). Il riconoscimento di un oggetto si può eseguire esplorandone alcune viste e formando un sottografo del grafo delle viste: si cercherà in archivio un oggetto che ammetta proprio quel sottografo.

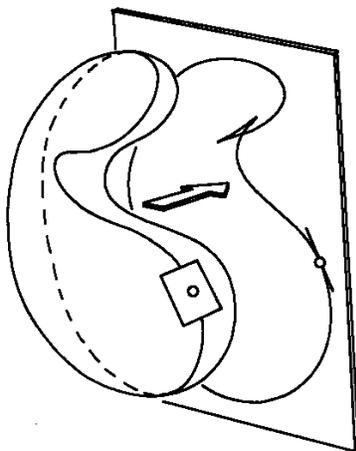


Figura 14: Il contorno apparente di una superficie. Riprodotta da [14] per concessione di MIT Press, Cambridge, Massachusetts 02142.

Questa volta il passaggio da poliedri a superfici lisce è drammatico. Infatti in un poliedro gli spigoli sono elementi essenziali nella scomposizione: è proprio la stabilità dei contorni apparenti dell'oggetto, che garantisce una schematizzazione combinatoria. Finché non sparisce, nascosto da altre parti dell'oggetto, è uno stesso spigolo a fornirci un segmento del contorno apparente anche se ci spostiamo. Nel caso di una superficie liscia, invece, il contorno apparente è la proiezione di una curva sulla superficie che cambia con continuità mentre ci spostiamo: tale curva è il luogo dei punti critici della proiezione stessa come nella Figura 14 [14].

Eppure ci sono “viste” stabili anche in questo caso (v. la Figura 15). La relazione di equivalenza non sarà più definita da trasformazioni proiettive o affini, ma dall’esistenza di un’omotopia liscia fra i contorni apparenti. La schematizzazione sarà data da una suddivisione della sfera di riferimento. Le curve di transizione fra le regioni bidimensionali si possono etichettare mediante una classificazione fornita dalla *teoria delle catastrofi* [34]: un’inaspettata rivincita di questa teoria geometrica nata per rappresentare fenomeni discontinui, apprezzata per le applicazioni fisiche, malvista per le applicazioni mediche, ostracizzata per le applicazioni sociali.

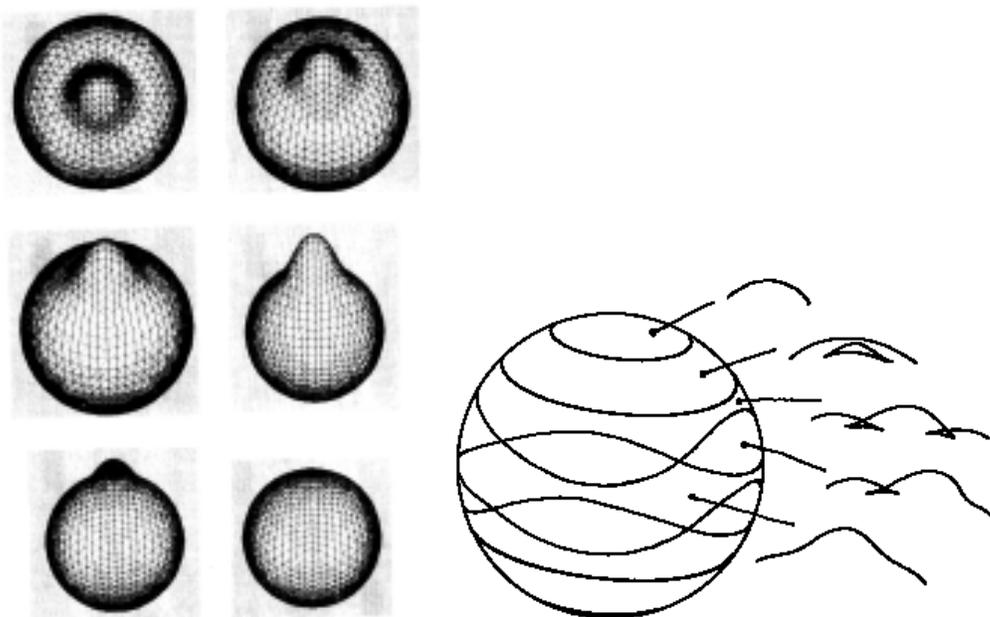


Figura 15: La mappa delle viste di una superficie liscia. Riprodotta da [14] per concessione di MIT Press, Cambridge, Massachusetts 02142.

## 5.4 Gruppi di trasformazioni

Può darsi che non disponiamo di una mappa delle profondità o delle normali. Allora siamo costretti a tentare una sovrapposizione diretta dell’immagine di un oggetto con quelle in archivio. Le situazioni che si presentano sembrano fatte apposta per esemplificare il concetto di geometria come studio di figure a meno di un gruppo di trasformazioni! Vediamone alcune.

Supponiamo di dover riconoscere un oggetto sostanzialmente bidimensionale (per esempio un componente elettronico). Questa situazione è frequente in

ambito industriale, dove un sistema automatico deve provvedere allo smistamento e magari al montaggio di pezzi che si presentano su un nastro trasportatore. La telecamera sarà fissata con l'asse ottico perpendicolare al piano del nastro. Supponiamo che il pezzo da riconoscere venga fermato in modo che il suo baricentro giaccia proprio sull'asse ottico, ma con un assetto sconosciuto. Il sistema ha in archivio immagini di tutti i pezzi possibili, ma ognuno in un assetto standard. A questo punto deve operare una trasformazione, basandosi su punti caratteristici dell'immagine, prima di tentare la sovrapposizione. Il gruppo in cui deve scegliere è semplicemente quello delle rotazioni attorno al centro dell'immagine: è una geometria molto elementare, ma utile. La trasformazione utilizzata per il confronto ci fornirà immediatamente anche l'assetto del pezzo.

In realtà, il più delle volte l'oggetto non viene fermato, e la trasformazione va cercata nel più usuale (e più ampio) gruppo delle uguaglianze dirette. Se c'è la possibilità che l'oggetto sia ribaltato, dobbiamo estendere la ricerca al gruppo di tutte le uguaglianze: siamo nella geometria euclidea. In altre situazioni, potremmo non conoscere con precisione la distanza della telecamera dall'oggetto, e si può rendere necessaria una variazione di scala: abbiamo il gruppo delle similitudini.

Non sempre l'asse ottico è perpendicolare al piano dell'oggetto; è facile immaginare condizioni (per esempio belliche) in cui l'osservazione avviene da una direzione fortemente inclinata. Se la distanza è tale che l'effetto prospettico sia trascurabile, possiamo limitare la ricerca della trasformazione al gruppo delle affinità; altrimenti dobbiamo arrivare al gruppo delle omografie.

Ci sono anche problemi che richiedono gruppi di trasformazioni non usuali. Per esempio, un robot mobile ha necessità di riconoscere e di utilizzare la direzione verticale. Il suo gruppo di trasformazioni è un sottogruppo proprio del gruppo delle uguaglianze.

Nella comunità della visione artificiale c'è attualmente una netta consapevolezza dell'importanza di questi temi, e si è assistito alla valorizzazione (e anche all'estensione!) di invarianti, di costruzioni, di teoremi che paradossalmente stanno scomparendo dagli insegnamenti universitari...

## 5.5 Trasformate

Le trasformate giocano un ruolo essenziale in certi tipi di elaborazione di immagini; cito solo la trasformata di Radon per la tomografia assiale computerizzata (TAC) [3] e quella di Fourier [5] per la correzione di immagini

sfocate o mosse.

Nel riconoscimento, è spesso necessario individuare insiemi di punti allineati. Anche questo può apparire a prima vista un problema banale, ma non lo è affatto, soprattutto per le approssimazioni dovute alla discretizzazione dell'immagine. La *trasformata di Hough* affronta questo problema.

Supponiamo che l'immagine (per il momento concepita nel continuo) sia semplicemente costituita da punti bianchi o neri. La trasformata di Hough associa ad ogni punto nero il fascio di rette di cui esso è il centro. Più precisamente, ogni retta è rappresentata a sua volta da un punto nel piano duale. La corrispondenza non è però l'usuale dualità proiettiva: essa viene effettuata rappresentando la retta nella forma  $(\cos \theta)x + (\sin \theta)y = \rho$ . Alla retta viene associato il punto di coordinate  $(\rho, \theta)$  (con  $\rho \geq 0$ ,  $0 \leq \theta < 2\pi$ ); quindi la trasformata associa ad ogni punto nero dell'immagine una curva (sinusoidale) del duale come nella Figura 16. Punti allineati verranno trasformati in curve del duale passanti per uno stesso punto.

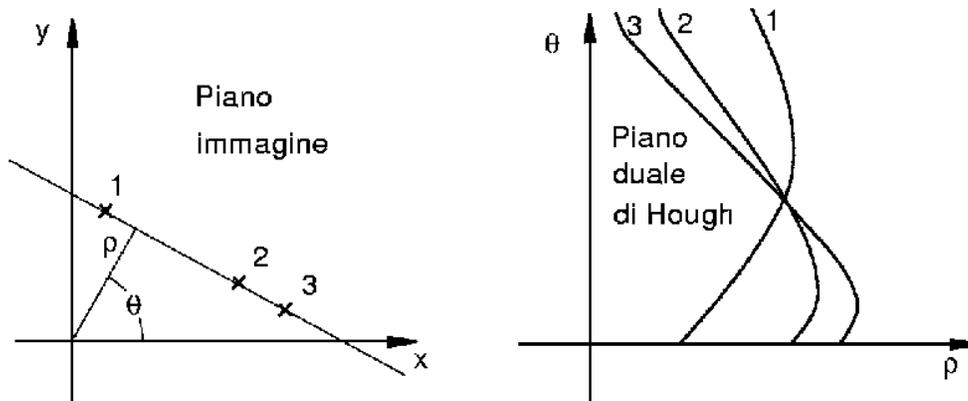


Figura 16: La trasformata di Hough applicata a tre punti allineati.

Niente di sorprendente, dunque. Il vantaggio si nota quando si passa alla realtà discreta delle immagini. Anche il piano duale viene discretizzato: ogni suo elemento rappresenta una retta "discreta" del piano immagine, e riceve un "voto" per ogni pixel dell'immagine che si trova su tale retta. Il risultato della trasformata è dunque un'assegnazione di interi (il numero di voti) ai singoli elementi del piano duale discretizzato. Le rette presenti nell'immagine sono identificate dagli elementi del duale che hanno ricevuto più voti. Il punto essenziale è che, nei casi reali, punti che nel continuo appartenerebbero ad una stessa retta risultano appartenenti a rette diverse nel discreto. Tuttavia queste rette, benché diverse, sono rappresentate da elementi vicini nel duale;

poiché esistono consolidate tecniche per il rilevamento di aggregati (*cluster*) di pixel vicini, questa tecnica risolve efficacemente il problema.

La trasformata di Hough ammette generalizzazioni che consentono la ricerca di altre configurazioni (per esempio ellissi, circonferenze, ecc.): viene semplicemente adottato un diverso codominio.

Nell'ambito delle trasformate vanno annoverate anche le *funzioni di taglia*, dotate di grande flessibilità e particolarmente adatte al riconoscimento di immagini naturali, dove cioè la sovrapposibilità di oggetti equivalenti è scarsa [10]. Le funzioni di taglia sono basate sulla teoria di Morse e sfruttano, per il calcolo, tecniche di teoria dei grafi. Un loro notevole vantaggio è che possono essere costruite in modo da essere invarianti per il gruppo di trasformazioni prescelto.

## 5.6 Classificazione

In realtà, molto spesso il problema del riconoscimento viene spostato su  $n$ -ple di numeri. Ad un'immagine — o meglio ad una parte dell'immagine — viene associata una  $n$ -pla di parametri e i confronti vengono effettuati in  $\mathbb{R}^n$ . Talvolta questi parametri sono i primi coefficienti di uno sviluppo: è il caso dell'analisi di Fourier o mediante ondine (anche dette “onduline”, *wavelets*): queste ultime sono generalmente preferite per la loro migliore resa in presenza di discontinuità a gradino, ciò che è normale nelle immagini [33]. Spesso, invece, la  $n$ -pla è semplicemente una lista (*feature vector*) di parametri scelti euristicamente (per esempio area, perimetro, diametro, ecc.).

Il riconoscimento di un oggetto si riconduce, comunque, a valutare l'appartenenza della corrispondente  $n$ -pla ad uno degli insiemi di  $n$ -ple già presenti, ottenuti preventivamente a partire da un insieme di addestramento (*training set*) costituito da  $n$ -ple corrispondenti ad oggetti per i quali la classificazione è certa e conosciuta.

Un esempio può essere utile. Supponiamo che il sistema di visione debba riconoscere, mediante una telecamera, i quattro componenti di una famiglia. Allora occorrerà addestrare il sistema fornendogli alcune fotografie dei singoli componenti, indicando le persone a cui si riferiscono. Il sistema estrarrà da ogni immagine la  $n$ -pla di parametri predisposta dal programmatore (per esempio alcune distanze fra punti caratteristici: pupille, punta del naso, angoli della bocca, ecc.) e formerà quattro insiemi distinti di  $n$ -ple etichettati ognuno con la persona corrispondente. Se ora presentiamo al sistema una nuova immagine di un volto, esso estrarrà la solita  $n$ -pla e dovrà determi-

nare a quale dei quattro insiemi sia più “ragionevolmente” aggregabile, o stabilire che si tratta di un estraneo.

Il problema dell’assegnazione all’uno o all’altro insieme coinvolge diversi aspetti matematici: un esempio è la scelta di un’opportuna distanza; un altro esempio è l’immersione in uno spazio di dimensione più elevata, quando in  $\mathbb{R}^n$  non sia possibile separare gli insiemi mediante iperpiani.

## 6 Applicazioni

Non dico una descrizione, ma una semplice lista di applicazioni della visione delle macchine richiederebbe da sola lo spazio di un intero articolo. Mi limito qui a segnalare alcuni punti salienti.

- *Manipolazione di oggetti*: qui la stima dell’assetto è il problema principale.
- *Ispezione*: il confronto con il prototipo consente il rilevamento di difetti.
- *Interpretazione di immagini biomediche*: un’efficace automazione renderà possibili screening di massa.
- *Modellazione geometrica*: rilevamento di una superficie, per esempio modellata da un artigiano.
- *Telerilevamento*: l’interpretazione di immagini da satellite.
- *Navigazione robotica*: la pianificazione dei percorsi di un veicolo autonomo [15], nei suoi due grandi capitoli della navigazione *indoor* (all’interno di un edificio) ed *outdoor* (v. Figura 17), molto diversi anche come matematica utilizzata.
- *Supporto alla mobilità*: supporto alla guida umana di veicoli, apparati di ausilio per non vedenti (per esempio il VIDET, sviluppato presso l’Università di Bologna).
- *Ricerca in database di immagini*: la ricerca di un’immagine o di una sequenza video può essere proibitiva se non c’è il supporto di una didascalia esauriente [9, 18].
- *Identificazione personale*: riconoscimento dell’iride, della retina, delle impronte digitali, della firma.
- *Riconoscimento di caratteri*: per lo smistamento postale, per la digitalizzazione di documenti, per la lettura vocale per non vedenti.

- *Riconoscimento di eventi*: controllo del traffico, identificazione di situazioni di pericolo.
- *Riconoscimento di espressioni facciali*: verso un'interfaccia più intuitiva fra umani e calcolatori.



Figura 17: Il rover marziano Sojourner, un robot mobile necessariamente autonomo. Fonte <http://mars.jpl.nasa.gov/MPF/rovercom/pix.html>

Importante è anche l'interazione della visione artificiale con la grafica al calcolatore. Questo connubio si applica nelle videoconferenze di nuova concezione: tratti somatici ed espressioni vengono rilevati, trasmessi e riprodotti artificialmente (in “realtà virtuale”) con una mole di dati trasmessi molto ridotta. Lo stesso principio si applica per la *teleoperazione*, cioè l'azionamento da parte di un operatore umano di una macchina remota, per esempio un veicolo per operazioni subacquee a grande profondità. Un altro esempio è la *realtà aumentata*: sovrapposizione di immagini reali con immagini sintetiche; è uno strumento prezioso per il moderno chirurgo (v. Figura 18), ma compare anche sui teleschermi come *grafica virtuale* nelle partite di calcio. Ancora il calcio ha stimolato fortunate applicazioni delle due tecniche congiunte nella costruzione delle *viste sintetiche*, cioè l'uso di immagini reali per creare immagini della stessa scena come se fosse ripresa da altri punti di vista. I nostri studenti, che apprezzano questi accessori del loro svago, ignorano che si basano sulla temuta e vituperata geometria proiettiva!

Vorrei sottolineare che praticamente tutte le applicazioni citate sono strumenti di supporto ad operatori umani; lungi dal togliere lavoro, permette-



Figura 18: Realtà aumentata in chirurgia. Fonte [http://www.ai.mit.edu/projects/medical-vision/surgery/surgical\\_navigation.html](http://www.ai.mit.edu/projects/medical-vision/surgery/surgical_navigation.html)

ranno anzi di estendere e qualificare quelli che adesso sono compiti alienanti, faticosi o pericolosi.

## 7 Conclusioni e letture suggerite

La visione delle macchine è difficile. Un androide che veda e capisca ciò che vede in modo paragonabile ad un umano è ancora fantascienza. Come abbiamo visto in questa sommaria rassegna, le esigenze pratiche impongono progressi in ambiti specifici, in situazioni controllate, con obiettivi ben circoscritti: la ricerca nella visione delle macchine si è allontanata dalla intelligenza artificiale propriamente detta per affrontare queste esigenze. Se da una parte ciò risponde ad un politica della ricerca sempre più utilitaristica, dall'altra si è ottenuto il risultato di spezzare grandi problemi in problemi più piccoli, un modo di agire solitamente fruttuoso.

Comunque sia, è stata sviluppata una miriade di tecniche diversificate sia come obiettivi sia come strumenti; contemporaneamente si sono messi in luce nuovi problemi altrettanto vari. Questa situazione continua a stimolare la fantasia dei ricercatori: occorrono nuovi modelli, nuovi algoritmi che possono essere in gran parte riciclati dalla fisica o da altre discipline, o che possiamo riesumare da vecchi libri in odore di macero. Sempre più, però, viene sollecitata la produzione di matematica del tutto nuova: al compito si accingono alcuni ingegneri; spero che accolga l'invito anche qualche collega matematico, e non solo ai livelli di De Giorgi o di Mumford.

Spero di aver destato la curiosità di approfondire qualcuno degli argomenti toccati. Sono possibili molte letture divulgative ma di buon livello sulle questioni di fondo dell'intelligenza artificiale; cito ad esempio [30, 16] e, per un dibattito di carattere più filosofico, i già citati [11, 27, 29, 6, 28]. Sul futuro della robotica consiglio [13, 21, 22]. Aspetti matematici della visione artificiale sono stati trattati in [4, 8], mentre applicazioni specifiche sono state affrontate su questo stesso Bollettino [3, 17]; un eccellente libro introduttivo è [32]. Un sito di riferimento bibliografico è <http://iris.usc.edu/Vision-Notes/bibliography/contents.html> dove si trovano anche le annuali, esaurienti bibliografie di Azriel Rosenfeld.

## Ringraziamenti

Articolo redatto nell'ambito dell'attività dell'INdAM-GNSAGA e dell'Università di Bologna, finanziamento speciale alle strutture. Desidero ringraziare gli amici del gruppo di Matematica della Visione di Bologna, Francesco Parigi ed Emanuele Trucco per le appassionante discussioni e l'impagabile aiuto.

## Riferimenti bibliografici

- [1] A. Alexandroff, *Existence and uniqueness of a convex surface with a given integral curvature*, Comptes Rendus (Doklady) Acad. Sci. URSS 35 (1942), 131–134.
- [2] L. Ambrosio, *Problemi con discontinuità libere*, Lettera Mat. Pristem 27–28 (1998), XXXIV–XXXVII.
- [3] M. Bertero, *Matematica e immagini: alcuni esempi di applicazioni*, Boll. Un. Mat. Ital. 2–A (1999), 47–67.
- [4] J.W. Bruce, *Seeing – the mathematical viewpoint*, Math. Intell. 6 n.4 (1984), 18–25.
- [5] T.M. Cannon, B.R. Hunt, *Elaborazione di immagini con il calcolatore*, Le Scienze 160 (1981), 120–132.
- [6] P.M. Churchland, P.S. Churchland, *Può una macchina pensare?* Le Scienze 259 (1990), 22–27.

- [7] P. Ciancarini, *Gli scacchi e i matematici*, Boll. Un. Mat. Ital. 2-A (1999), 203–236.
- [8] V. Di Gesù, *Simmetria e visione artificiale*, Lettera Mat. Pristem 24 (1997), 33–41.
- [9] D. Forsyth, J. Malik, R. Wilensky, *In cerca di immagini digitali*, Le Scienze 348 (1997), 62–67.
- [10] P. Frosini, C. Landi, *Size theory as a topological tool for computer vision*, Pattern Recogn. Image Anal., 9 (1999), 596–603.
- [11] D.R. Hofstadter, D. C. Dennett, *L'io della mente*, Adelphi 1985.
- [12] T. Kanade, *Recovery of the three-dimensional shape of an object from a single view*, Artif. Intell. 17 (1981), 409–460.
- [13] H. Kitano, M. Asada, *RoboCup: e ora in campo gli automi*, Le Scienze 372 (1999), 32–42.
- [14] J.J. Koenderink, *Solid shape*, MIT Press, 1990.
- [15] J.-C. Latombe, *Robot Motion Planning*, Kluwer Academic Publ., 1996
- [16] D.B. Lenat, *L'intelligenza artificiale*, Le Scienze 327 (1995), 44–50.
- [17] P. Longo, *La "Flagellazione" di Piero della Francesca fra Talete e Gauss*, Boll. Un. Mat. Ital. 2-A (1999), 121–144.
- [18] C. Lynch, *La ricerca su Internet*, Le Scienze 350 (1997), 60–64.
- [19] D. Marr, E. Hildreth, *Theory of edge detection*, Proc. R. Soc. Lond. B 207 (1980), 187–217.
- [20] H. Minkowski, *Allgemeine Lehrsätze über die konvexen Polyeder*, Nachr. Königl. Gesellsch. Wissensch., math.-phys. Kl., Göttingen (1897), 198–219.
- [21] M. Minsky, *Saranno i robot a ereditare la Terra?* Le Scienze 316 (1994), 96–102.
- [22] H. Moravec, *Quando diventeranno intelligenti i robot?* Le Scienze 376 (1999), 114–122.
- [23] D. Mumford, J. Shah, *Optimal approximations by piecewise smooth functions and associated variational problems*, Commun. Pure Appl. Math. 42 (1989), 577–685.

- [24] M. Nitzberg, D. Mumford, *The 2.1-D sketch*, Proc. 3rd ICCV, Osaka (1990), 138–144.
- [25] C. Pellegrino, E. Barozzi, A. Borrelli, *Prospettiva: il punto di vista della geometria*, Pitagora Ed., 1999.
- [26] L.S. Penrose, R. Penrose, *Impossible objects: a special type of visual illusion*, British J. Psychol. 49 (1958).
- [27] R. Penrose, *La mente nuova dell'imperatore*, Rizzoli, 1992.
- [28] G.C. Rota, D. Sharp, *Matematica, filosofia e intelligenza artificiale*, Lettera Pristem 5 (1992), Dossier 1–16.
- [29] J.R. Searle, *La mente è un programma?* Le Scienze 259 (1990), 16–21.
- [30] L. Stringa, *Il progetto MAIA*, Le Scienze 290 (1992), 86–97.
- [31] K. Sugihara, *A necessary and sufficient condition for a picture to represent a polyhedral scene*, IEEE Trans. PAMI 6 (1984), 578–586.
- [32] E. Trucco, A. Verri, *Introductory techniques for 3-D computer vision*, Prentice Hall, 1998.
- [33] J.S. Walker, *Fourier analysis and wavelet analysis*, Notices Amer. Math. Soc. 44 (1997), 658–670.
- [34] E.C. Zeeman, *Catastrophe theory*, Addison–Wesley 1997.