

Matematica, immagini e visione computazionale: gioie, dolori e possibili sfide¹

Giovanni Naldi
Dipartimento di Matematica "F. Enriques"
Università degli Studi di Milano
Via Saldini 50, Milano
giovanni.naldi@unimi.it

"... La ricchezza della tua esperienza pratica con problemi sensati ed interessanti darà alla matematica una nuova direzione e un nuovo impeto..." (da una lettera di Leopold Kronecker a Hermann von Helmholtz)

"... Non esiste una categoria della scienza alla quale noi possiamo dare il nome di scienza applicata. Ci sono la scienza e le applicazioni della scienza legate insieme come il frutto dell'albero che li genera..." (Louis Pasteur)

"...O studenti studiate le matematiche, e non edificate senza fondamenti... Chi biasima la somma certezza delle matematiche si pasce di confusione, e mai porrà silenzio alle contraddizioni delle sofistiche scienze, colle quali s'impara uno eterno gridore." (Leonardo da Vinci, Fogli di Windsor, Royal Library)

1. Introduzione.

"Anche l'occhio vuole la sua parte", certamente questa sentenza suggerisce la necessità di armonia estetica che tutti desideriamo. Bene, ma cosa vuol dire vedere? Come viene codificata, memorizzata, recuperata l'informazione visiva? La domanda è una classica domanda che spalanca un abisso (o una cima tremendamente impervia). Per dirla con l'Amleto di Shakespeare *"E allora dagli il benvenuto, come si fa con gli stranieri. Vi sono più cose in cielo e in terra, Orazio, di quante non ne sogni la tua filosofia"*.

La questione della visione è quindi ardua, ci limiteremo qui a sottolineare che non possiamo ridurre l'atto della visione alla mera percezione. Per rendercene conto facciamo un "esperimento mentale" tratto da un celebre passo *delle Ricerche filosofiche* di L. Wittgenstein e ripreso da GianCarlo Rota in un famoso articolo sull'intelligenza artificiale [1]. Il passo riguarda la scrittura ma è direttamente applicabile anche alla visione. Vogliamo dimostrare che l'atto di leggere un testo è un atto intenzionale che non può essere ridotto a una sequenza di atti mentali e tantomeno di natura fisica e meccanica. Vediamo alcune possibili descrizioni della lettura che possono essere invalidate:

D1) seguo con gli occhi una successione di simboli scritti su un pezzo di carta. Ma può accadere che io segua questa azione senza neppure conoscere i simboli della scrittura in cui il testo è redatto.

D2) Un gruppo di medici, avendomi messo alcuni elettrodi in testa, registra delle onde cerebrali la cui analisi armonica dà risultati sostanzialmente identici a quelli ottenuti su un'altra persona che effettivamente stava leggendo il medesimo testo. Ma non è escluso che, avendolo in precedenza imparato a memoria, io stia ripetendolo mentalmente, pur facendo finta di leggerlo.

D3) Pronuncio correttamente ad alta voce il testo, parola per parola, pur non essendo a conoscenza della lingua nella quale è stato scritto, ma avendone solo imparato la corrispondenza tra parole scritte e pronuncia.

D4) Leggo una parola al giorno. Ogni volta che leggo una parola ho dimenticato le precedenti. Alla fine ho eseguito gli stessi atti elementari di uno che ha effettivamente letto la frase, ma non posso dire che nel mio caso sia accaduto il fenomeno della lettura.

D5) Senza sapere né leggere né scrivere, riproduco meticolosamente lo stesso testo su un foglio di carta.

Si vede che ciascuna delle descrizioni empiriche date non coglie la realtà del fenomeno del leggere. L'atto di osservare qualsiasi oggetto può essere sottoposto allo stesso genere di analisi.

A livello di definizione generale per la visione siamo quindi abbastanza lontani e questo può sembrare scoraggiante. Riprendiamo ancora Rota quando parla dell'intelligenza artificiale. *"Ma allora cosa vede l'occhio quando vede, per riprendere la frase del fisiologo Jerry Lettvin? Le vecchie spiegazioni riduzionistiche presupponevano un oggetto, al quale venivano psicologicamente sovrapposti ruoli e funzioni variabili a seconda del contesto. Ma agli effetti pratici del problema dell'intelligenza artificiale la postulazione di un oggetto diventa totalmente irrilevante, anzi fuorviante. In senso risiede nella funzione, non nell'oggetto ed è la funzione che deve essere descritta, con una descrizione che prescindere completamente dall'appello a meccanismi psicologici o neurofisiologici."* La funzione essenziale della visione però sembra quasi evidente: descrivere e determinare le proprietà del mondo che ci circonda. Notiamo anche che i dati neurofisiologici, anatomici, psicologici, morfologici, ... si stanno accumulando e da questi ipotesi sulla percezione e sulla elaborazione di

¹ Contributo pubblicato in: *Mathesis*, n. 18, 2006, pp. 3-23.

immagini visive si fanno strada. Facciamo qualche cenno del sistema visivo umano. Il nostro sistema visivo è composto dall'occhio, sistema ottico più retina, e quella parte del Sistema Nervoso Centrale che elabora l'informazione visiva, cioè realizza quel processo che dallo stadio iniziale di attivazione dei recettori della retina ad opera della luce porta all'interpretazione, ovvero a dare significato a ciò che vediamo. L'occhio raccoglie una banda di radiazioni elettromagnetiche "rimbalzate" su diverse superfici e provenienti da fonti luminose diverse ed il cervello elabora questa informazione formando il quadro della scena come noi la percepiamo. La banda dello spettro elettromagnetico visibile all'occhio umano varia approssimativamente tra i 400 e i 770 nm (nanometri): per avere solo un'idea il colore rosso corrisponde alla radiazione a 700 nm, il violetto a 400 nm, il verde poco più di 500 nm. Dal punto di vista della quantità di informazione osserviamo che i nostri occhi riescono a raccogliere informazioni dell'ordine di 10^8 bit al secondo. Informazioni poi trasmesse a quelle zone del nostro cervello dedicate a funzioni di più alto livello tra cui l'estrazione di caratteristiche o effettuare il riconoscimento. A dispetto della grande quantità di informazione accumulata, un uomo riesce mediamente a riconoscere un gran numero di oggetti in meno di 200ms. Si intuisce facilmente che la capacità di elaborazione del cervello non può interamente spiegare questa incredibile performance elaborando una tale mole di informazioni in così poco tempo. La presenza di un meccanismo che filtra il flusso di informazione dalla retina al cervello sembra essere la spiegazione più coerente per le prestazioni del nostro sistema visivo (si veda Figura 1).

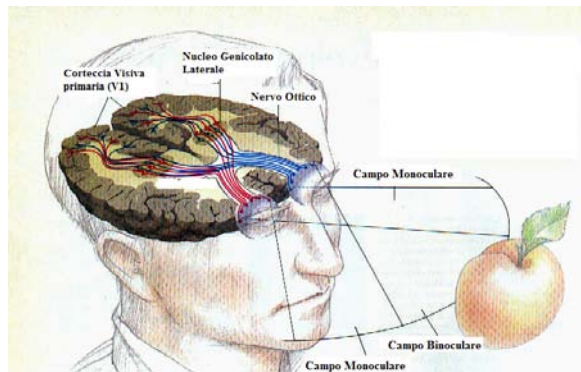


Figura 1. Principali suddivisioni anatomiche del sistema visivo umano.

L'occhio è un sistema biologico complesso che ha l'obiettivo fondamentale di ottenere immagini a fuoco sulla retina, superficie dove sono situati i fotorecettori: recettori sensibili alla luce. L'occhio è un organo di forma approssimativamente sferica il cui raggio è di circa 12 mm. Nell'involucro e protettivo in cui è inserito e che ne preserva la forma, sono situate tre coppie di muscoli che muovono l'occhio nell'orbita. Ogni cambiamento di posizione dell'occhio è causato dalla contrazione ben coordinata di questi muscoli. L'occhio è costituito da un sistema ottico (cornea, cristallino, umor acqueo, corpo vitreo) e da un sistema sensoriale, la retina. La cornea e il cristallino sono responsabili della rifrazione della luce che entra nell'occhio. L'umor acqueo è un liquido, continuamente generato e assorbito, controlla la pressione intraoculare, che è più grande della pressione atmosferica, per mantenere l'integrità strutturale dell'occhio. Il cristallino è una lente biconvessa a curvatura regolabile ed è proprio la curvatura variabile del cristallino che ci consente di mettere a fuoco oggetti a distanze diverse. La cornea è lo strato più esterno, lo strato attraverso il quale la luce entra nell'occhio. La cornea è costituita da una complessa struttura lamellare ma risulta trasparente. Dal punto di vista ottico la cornea è un diotro convergente (interfaccia tra due mezzi con indici di rifrazione diversi). Il campo visivo di ogni occhio è la regione dello spazio esterno in cui l'occhio, quando si tiene in posizione fissa e non ruota nell'orbita, può vedere gli oggetti. La nostra visione è stereoscopica o binoculare: la decodifica della tridimensionalità avviene dalle disparità binoculari (cioè dall'integrazione delle informazioni ottenute dai differenti campi visivi degli occhi). Questo processo che potrebbe sembrare banale, lo compiamo senza sforzo e spontaneamente, comporta invece sofisticati metodi computazionali. La complessa struttura del bulbo oculare ha il compito di formare una immagine del mondo esterno sulla retina. La retina poi è in grado di trasformare segnali luminosi in segnali "nervosi". La retina ha la forma di una lamina dello spessore di circa 1/4 di millimetro, essa ci permette di vedere in condizioni luminose differenti e a discriminare le lunghezze d'onda per farci vedere i colori e ci assicura una precisione sufficiente ad individuare un capello o un granello di polvere ad alcuni metri di distanza. La connessione con il cervello propriamente detto avviene attraverso un fascio di fibre: il nervo ottico. La fila di cellule nella parte posteriore della retina è costituita dai fotorecettori: i bastoncelli e i coni.

I bastoncelli sono più numerosi dei coni e sono responsabili della visione in condizione di scarsa illuminazione. I coni non rispondono alle basse intensità luminose ma sono responsabili della capacità di distinguere i dettagli fini e o della visione dei colori. Un fatto interessante: pare che i fotorecettori in fase di decodifica dei colori utilizzino una codifica RGB (Red-Green-Blue). Infatti prove sperimentali sull'attivazione dei fotorecettori ha evidenziato famiglie di fotorecettori più sensibili al rosso, altre più sensibili al verde ed altre più sensibili al blue. Il colore sembra quindi decomposto in queste tre componenti fondamentali, come avviene per i monitor solitamente utilizzati dai nostri calcolatori. La retina si collega a una parte ben distinta del cervello: la corteccia visiva primaria o striata, denominata anche area V1. Retina e area V1 sono collegate tramite una struttura subcorticale, il nucleo genicolato laterale che comprende sei strati di cellule. È stato dimostrata l'esistenza e la funzionalità di altre aree nella corteccia visiva separate dalla V1: aree V2, V3, V3A, V4, V5. Pare che diverse aree hanno diverse specializzazioni rispetto agli stimoli: colori, linee con varie orientazioni, punti in movimento, ... Quindi l'informazione e la percezione visiva viene decomposta ed integrata all'interno della corteccia visiva: come definire il progetto complessivo? Quali potrebbero essere i contributi dei Matematici? Sviluppare una descrizione matematica per i fenomeni fisiologici e cognitivi importanti scoperti nell'ambito della percezione visiva. Estrarre le leggi fondamentali, dedurre le proprietà computazionali che permettono l'efficienza del sistema visivo ed integrare queste conoscenze in algoritmi. Sviluppare nuove metodologie matematiche per questi studi (eventualmente utili per altre applicazioni scientifiche).

Se la visione naturale, cioè degli esseri viventi, presenta già difficoltà in fase di definizione cosa si propone la visione computazionale? Certamente la visione computazionale non tenta di replicare la visione umana: il tentativo pare sia destinato al fallimento anche solo per la intrinseca differenza tra i due "hardware". Anche il sogno di Leonardo di replicare per il volo animale è fallito nella forma primitiva mentre gli aerei di oggi (lontani dall'aver una struttura simile ad un volatile) hanno egregiamente risolto il problema. Se volessimo dare una definizione, si potrebbe dire che la visione computazionale, nella lettura nota anche come Computer Vision, studia tecniche per l'analisi automatica di immagini finalizzata ad acquisire informazione sul mondo esterno. Le tecniche studiate devono supplire diversi livelli cognitivi da quello più basso (acquisizione dell'immagine) a quelli più elevati (interpretazione della scena).

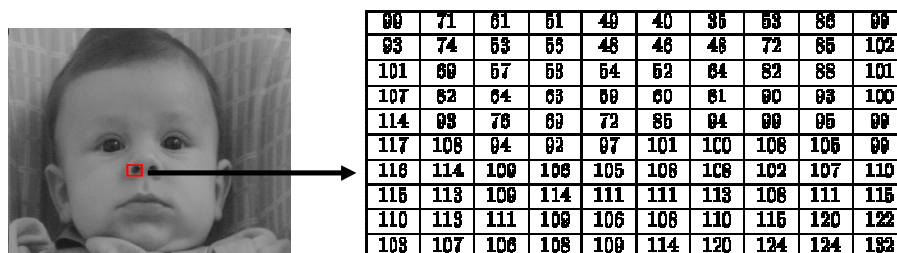


Figura 2. Un'immagine digitale a livelli di grigio, la tabella a destra rappresenta il valore dei singoli pixel dell'area incorniciata.

In visione computazionale un'immagine è semplicemente una funzione f in due variabili. Il valore $f(x,y)$ rappresenta una misura di qualche caratteristica (luminanza, colori, ...) in una "posizione" (x,y) di una scena (ovviamente si pensa di aver definito un sistema di riferimento per poter parlare senza ambiguità di posizione). Per esempio pensiamo ad una antica e suggestiva fotografia in bianco e nero (B/N): in ogni punto la funzione f rappresenta la tonalità di grigio presente nel punto stesso. Il processo di formazione di una immagine da analizzare attraverso un computer può essere molto diverso da quello essenzialmente di natura ottica che abbiamo sommariamente descritto in precedenza (per esempio, nelle immagini mediche spesso viene misurato il coefficiente di assorbimento dei tessuti rispetto a particolari radiazioni elettromagnetiche). Torniamo all'esempio della fotografia B/N, il "piano immagine" I , ovvero il dominio della funzione f , è un sottoinsieme limitato del piano \mathbb{R}^2 (un rettangolo tipicamente), un codominio della f è un intervallo $[0, F_{\max}]$. Una funzione associata ad un'immagine raramente può essere espressa mediante una semplice funzione analitica per cui, per effettuare su di essa elaborazioni mediante il calcolatore, è necessario che la funzione f venga campionata e quantizzata. Campionare una immagine significa prendere sul piano immagine un numero finito di punti (i Teoremi del campionamento stabiliscono le condizioni necessarie per non avere perdita di informazione o per non perderne "troppa"). Quantizzare significa prendere un numero finito di valori interi nell'intervallo $[0, F_{\max}]$. Naturalmente un campionamento ed una quantizzazione non adatte o inadeguate comportano una grave perdita di informazione con relativo deterioramento dell'immagine. In particolare il numero di livelli di una eventuale quantizzazione deve essere tale da non far percepire "salti di livello di grigio" all'osservatore. Nella nostra fotografia ciò può accadere se i livelli di grigio (numero di suddivisioni dell'intervallo $[0,$

F_{\max}], sono meno di 5. In genere si usano 256 livelli di grigio, ma per alcune applicazioni questo valore può cambiare: per esempio in ambito medico le immagini NMR (Nuclear Magnetic Resonance) utilizzano 4096 livelli di grigio o più.

In Figura 2 si mostra un esempio di immagine digitale e la rappresentazione di una sua piccola parte come matrice di valori interi. Ogni valore intero rappresenta un certo livello di valore di grigio. Il fatto che l'immagine si possa rappresentare attraverso una matrice è conseguenza del fatto che la fotografia è stata suddivisa in un certo numero di celle elementari attraverso un numero opportuno numero di righe e di colonne. Cambiare numero di righe e di colonne implica cambiare l'area della cella elementare e quindi la risoluzione con cui si decompone l'immagine. I valori di grigio, cioè i valori nella tabella che rappresenta l'immagine, si possono pensare come medie locali dei valori di grigio nella cella (rappresentazione vicina a come operano, per esempio, le fotocamere digitali). Un'immagine digitale ottenuta e descritta come appena descritto è un'immagine raster. Le celle elementari dette si chiamano anche *pixel* (picture elements). In certe applicazioni l'immagine potrebbe non essere descritta pixel per pixel ma attraverso un insieme di primitive geometriche (linee, archi, rettangoli, ...), abbiamo un'immagine vettoriale.



Figura 3. Esempio di immagine digitale (A); con una diversa risoluzione, più grossolana con la formazione di quadretti più grandi (B); con diversa quantizzazione dei livelli di grigio (C) con meno livelli.

In Figura 3 si mostra l'effetto del cambiamento di risoluzione spaziale e di quantizzazione per i valori di livello di grigio (l'immagine è Lena un'immagine famosa spesso utilizzata come immagine test in esperimenti di Computer Vision). Nel seguito utilizzeremo immagini in bianco e nero con livelli di grigio. Per le immagini a colori ricordiamo che possiamo pensarle come sovrapposizione di tre immagini monocromatiche (verde, rosso e blue): la funzione f diventa una funzione a valori vettoriali. Tale scelta dei colori fondamentali non è certo l'unica possibile ma la più utilizzata (codifica RGB). Ogni singola immagine monocromatica può essere pensata come un'immagine a livelli di valori di verde, di rosso e di blue e quindi formalmente come tre immagini del tipo livello di grigio (un codominio è un intervallo limitato). Che elaborazioni sono presenti in Computer Vision? Elenchiamo di seguito alcuni aspetti per una classificazione generalmente accettata. Solitamente si distingue tra visione a basso livello, *low level vision*, che produce una nuova immagine; visione a medio livello, *mid level vision*, per estrarre informazioni di tipo strutturale; visione ad alto livello, *high level vision*, che produce una interpretazione della scena (livello semantico). Più in dettaglio abbiamo la "*Low-level vision*" ovvero la prima elaborazione delle immagini per l'estrazione di caratteristiche: calcolo del movimento degli oggetti sull'immagine (optical flow), estrazione della profondità, riconoscimento di tessiture. "*Low-level vision*" come rappresentazione intermedia: calcolo delle sorgenti di illuminazione, determinazione delle forme dai contorni (shape from edges), determinazione di forme da tessitura (shape from texture), determinazione di forme da ombreggiatura (shape from shading), determinazione della struttura 3D e del movimento 3D di oggetti da sequenze di immagini (Structure from Motion), ricostruzione 3D da stereo di oggetti della scena, parametri geometrici del sistema di visione (calibrazione). Infine per la *High level vision*: interpretazione e movimento degli oggetti nella scena. L'elaborazione di immagini, o di sequenze di immagini, costituisce il primo livello di un sistema di visione computazionale. Rimandiamo per i molti temi presenti in questo ambito ad alcuni testi ampiamente utilizzati in corsi specialistici [2,3,4]. Il testo di Rosenfeld e KaK non è recente ma rappresenta ormai un classico sugli argomenti di elaborazione e rappresentazione di immagini con una organizzazione ed esposizione degli argomenti chiara ed ancora valida.

Nel seguito tratteremo due soli esempi: un esempio di visione a basso livello (estrazione di bordi) ed un esempio per il livello medio (segmentazione). Non tratteremo il livello più alto anche se è il livello più interessante e su cui la ricerca necessita maggiormente di nuove idee. Non potendolo trattare nemmeno in modo estremamente superficiale rimandiamo ad almeno ad un sito importante per i matematici essendo il sito di David Mumford, medaglia Fields (oltre che vincitori di altri numerosi e prestigiosi riconoscimenti) e attualmente impegnato nella ricerca in Pattern Theory e nei processi del sistema visivo: <http://www.dam.brown.edu/people/mumford/>

Concluderemo con una breve escursione nell'ambito della rappresentazione efficiente dei dati.

2. Alla ricerca del bordo perduto.

Negli anni 1969, 1970 e 1971, David Marr pubblicò una serie di lavori che ambivano a spiegare il funzionamento di gran parte del sistema nervoso: il primo lavoro riguardava il cervelletto, il secondo la neocorteccia, il terzo l'ippocampo. L'influsso di Marr è stato notevole. Per quanto riguarda la visione un'osservazione centrale di Marr è la seguente: "... *al centro della percezione visiva è l'inferenza della struttura del mondo reale, derivata dalla struttura dell'immagine. La teoria della visione è esattamente la teoria di come ciò può essere fatto, e il suo interesse principale è nei confronti dei limiti fisici e delle assunzioni che rendono possibile tale inferenza.*" Il maggior contributo di Marr [5] è consistito nella sua proposta metodologica che espone i tre differenti livelli ai quali un sistema di elaborazione di informazione deve essere compreso per poter affermare che esso è stato compreso completamente. Questi tre livelli sono: teoria computazionale, rappresentazione e algoritmi, realizzazione fisica. Diversi lavori successivi, basati sull'impostazione di Marr, hanno evidenziato che molti dei problemi che devono essere risolti per effettuare la ricostruzione di scene, come la determinazione del movimento, la ricostruzione delle forme dall'ombreggiatura, l'interpolazione di superfici, l'estrazione dei bordi, ecc., rientrano nella classe di problemi mal posti o mal condizionati. Vediamo una delle primitive utili all'analisi dell'immagine: il calcolo dei bordi. Nei punti dove vi sono variazioni fisiche di una superficie l'immagine della stessa superficie presenta solitamente evidenti variazioni di intensità: illuminazione non uniforme, variazioni nella geometria (orientazione e profondità), riflessione. Questi punti possono essere utili per definire i bordi, i contorni di oggetti presenti in un'immagine. È possibile identificare le variazioni locali di intensità in un'immagine confrontando i valori di pixel adiacenti. Per questa operazione lo strumento matematico adeguato è una stima numerica della derivata prima: in questo caso la derivata dà una indicazione della rapidità di variazione dell'intensità lungo un cammino attraverso la matrice dei pixel. La posizione dei punti di minimo e di massimo della derivata prima possono identificare la posizione di un bordo.

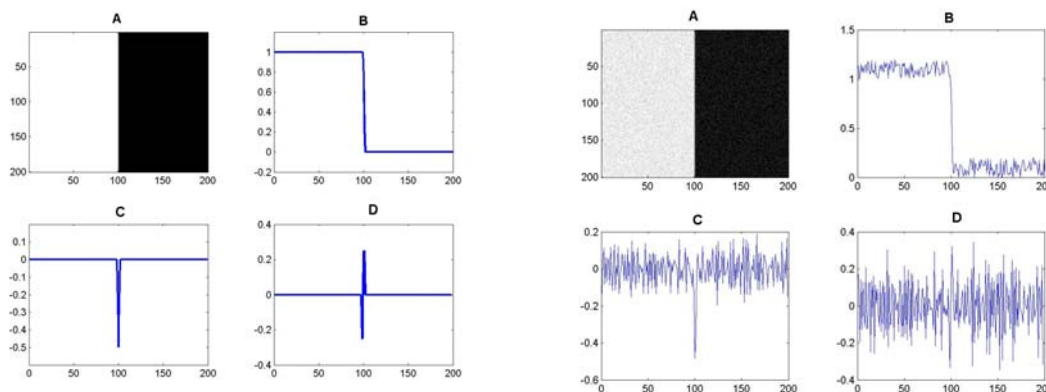


Figura 4. La stima numerica delle derivate di un'immagine servono per individuare le variazioni di intensità nello spazio dei pixel. A sinistra si mostra un'immagine test con due regioni ben distinte (A), in (B) si mostra una tipica riga dell'immagine, in (C) una stima dell'andamento della derivata prima della riga, in (D) una stima della derivata seconda. A destra si mostrano figure analoghe ma per un'immagine in presenza di rumore (situazione tipica nelle immagini reali).

Anche la derivata seconda, o meglio una sua stima numerica, ci può aiutare: un bordo nella matrice dei pixel si traduce in un cambiamento di segno della derivata seconda (e quindi come si dice nella terminologia tecnica corrispondente: in "passaggio attraverso zero"). Le derivate sono utili ma per avere buoni risultati per quanto riguarda immagini reali occorre procedere con cautela: le variazioni dei valori dei pixel sono "mascherate" dall'equivalente visivo del rumore. In Figura 4 si mostra un esempio in cui l'approssimazione delle derivate, prima e seconda, è stata effettuata attraverso opportuni rapporti incrementali (*differenze finite*). Da quest'ultimo punto di vista ci si ricorda che la matrice dei pixel rappresenta il campionamento di una certa funzione f per la quale vogliamo calcolare un'approssimazione (F_x, F_y) del gradiente,

$$F_x \approx \frac{\partial f}{\partial x}, \quad F_y \approx \frac{\partial f}{\partial y}, \quad \nabla f = \left(\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y} \right).$$

Un'idea "naturale" per i valori approssimati F_x e F_y consiste nel considerare i rapporti incrementali

$$F_x = \frac{f(x + \Delta x, y) - f(x, y)}{\Delta x}, \quad F_y = \frac{f(x, y + \Delta y) - f(x, y)}{\Delta y}$$

con $\Delta x, \Delta y$ incrementi da scegliere. Nel caso delle immagini risulta $\Delta x=1, \Delta y=1$, ovvero ci si sposta al minimo di un pixel, siamo comunque interessati all'ordine di grandezza della variazione piuttosto che all'accuratezza numerica. La scelta fatta non è certamente la sola possibile. Indicando con F la matrice dei valori dei pixel le formule precedenti si possono scrivere come (precedentemente non avevamo esplicitato il punto rispetto a cui si calcolava il rapporto incrementale, adesso puntualizziamo che i rapporti incrementali si possono riferire al pixel corrispondente alla riga i -esima e colonna j -esima)

$$F_x(i, j) = F(i + 1, j) - F(i, j), \quad F_y(i, j) = F(i, j + 1) - F(i, j).$$

Vediamo altre proposte utilizzate nell'ambito dell'elaborazione di immagini (si tratta di medie pesate di rapporti incrementali basati su differenze centrali che hanno una maggiore accuratezza)

Formule di Prewitt

$$F_x(i, j) = \frac{[F(i + 1, j + 1) - F(i - 1, j + 1)] + [F(i + 1, j) - F(i - 1, j)] + [F(i + 1, j - 1) - F(i - 1, j - 1)]}{6}$$

$$F_y(i, j) = \frac{[F(i, j + 1) - F(i, j - 1)] + [F(i + 1, j + 1) - F(i + 1, j - 1)] + [F(i - 1, j + 1) - F(i - 1, j - 1)]}{6}$$

Formule di Sobel

$$F_x(i, j) = \frac{[F(i + 1, j + 1) - F(i - 1, j + 1)] + 2[F(i + 1, j) - F(i - 1, j)] + [F(i + 1, j - 1) - F(i - 1, j - 1)]}{8}$$

$$F_y(i, j) = \frac{2[F(i, j + 1) - F(i, j - 1)] + [F(i + 1, j + 1) - F(i + 1, j - 1)] + [F(i - 1, j + 1) - F(i - 1, j - 1)]}{8}$$

Anche la rilevazione dei bordi, come molti altri problemi in visione computazionale, è un problema mal posto. Consideriamo un esempio 1D, la funzione

$$f(x) = g(x) + \varepsilon \sin(\omega x) \Rightarrow f'(x) = g'(x) + (\varepsilon \omega) \cos(\omega x)$$

dove g è una funzione regolare e ε, ω parametri positivi $0 < \varepsilon \ll 1$. Le alte frequenze, ω grande, sono quelle che recano più disturbo. Per attenuare il problema dovuto al rumore solitamente si cerca di regolarizzare l'immagine filtrando opportunamente i valori dei pixel contigui ed "attenuando" le alte frequenze presenti nell'immagine. Dal punto di vista matematico l'operazione di regolarizzazione avviene attraverso opportune operazioni di convoluzione. In particolare si utilizza una convoluzione con una funzione Gaussiana di assegnata media e varianza. Operazioni di questo tipo sono dette di *smoothing* e precedono altre elaborazioni. Inoltre nella pratica dell'elaborazione di immagini anche le formule di Prewitt piuttosto che quelle di Sobel, e formule simili, si implementano utilizzando convoluzioni con opportuni nuclei di convoluzione. Per la decisione se un punto è un punto di bordo dalle componenti del gradiente si calcola un valore del modulo del vettore gradiente. Quindi pixel per pixel si genera una nuova immagine in cui i nuovi valori rappresentano i moduli dei gradienti. Stabilita una soglia si identificherà nella nuova immagine un punto come punto di bordo se il modulo corrispondente supera la soglia prefissata. Il valore di soglia è un parametro piuttosto "delicato" ed e, solitamente, valori ottimali si possono stabilire dopo una buona sperimentazione numerica riguardante classi di immagini. In Figura 5 si mostra come la variazione della soglia per il calcolo dei bordi attraverso un operatore di Sobel possa produrre bordi differenti. I vettori gradiente sono perpendicolari alle curve di livello dell'intensità dei valori dei pixel, la loro determinazione non solo fornisce i punti di bordo ma può anche dare indicazioni sull'orientazione dei bordi stessi. I bordi potrebbero anche essere determinati attraverso gli zeri del Laplaciano della funzione immagine f ,

$$\nabla^2 f = \frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2}$$

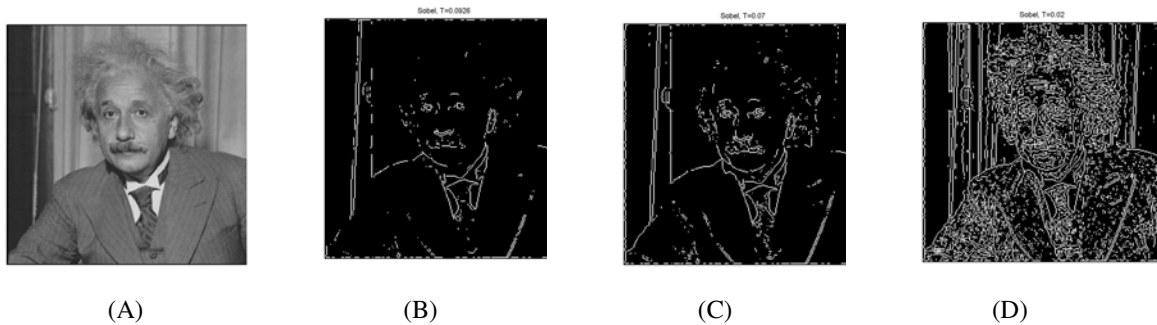


Figura 5. Applicazione dell'operatore di Sobel per diverse scelte della tolleranza di discriminazione dei punti di bordo, la scelta è fatta in base al valore approssimato del modulo del gradiente. In (A) l'immagine originale e da (B) a (D) valori decrescenti del valore di soglia.

Grazie alle proprietà degli operatori di convoluzione e del Laplaciano è possibile dimostrare che le operazioni di regolarizzazione e di calcolo del Laplaciano possono essere scambiate: il Laplaciano dell'immagine regolarizzata è uguale alla regolarizzazione del Laplaciano dell'immagine. Questa osservazione permette un'implementazione molto efficiente, il filtro globale che si ottiene è detto filtro LOG (Laplacian of Gaussian). Il metodo LOG è stato proposto per la prima volta da Hildreth e Marr [6]. L'Algoritmo di Hildreth e Marr è essenzialmente il seguente:

- Sia f_0 l'immagine originale;
- Regolarizzare (attraverso la convoluzione) l'immagine f_0 con Gaussiane G_t di varianza crescente

$$G_t(x, y) = \frac{1}{4\pi t} e^{-(x^2+y^2)/(4t)}$$

ottenendo la famiglia di immagini $f(x,y,t)$;

- Ad ogni valore di scala t , si calcolano i punti dove $\nabla f \neq 0$ e $\nabla^2 f$ cambia segno, questi punti sono detti zero-crossing;
- (opzionale) si eliminano gli zero crossing nei quali il gradiente piccolo.

Nella pratica bastano alcuni valori t di una scala diadica: 2,4,8,16,... Un Algoritmo collegato con quello di Hildereth e Marr è il metodo di Canny in cui ancora ci sono le convoluzioni con una famiglia di Gaussiane ma cambia la definizione di punto di bordo per ogni scala (zero crossing) selezionando in sostanza quelli in cui il gradiente non è nullo mentre $g'(0)=0$, dove

$$g(t) = |\nabla f| \left(\underline{x} + t \frac{\nabla f}{|\nabla f|} \right)$$

e \underline{x} rappresenta il vettore (x,y) mentre $|\cdot|$ il modulo. L'obiettivo è di eliminare dall'immagine i pixel che non sono massimi locali rispetto all'orientazione del gradiente. Inoltre vengono "seguiti" solo i punti di bordo significativi attraverso una opportuna scelta di soglie multiple. In Figura 6 si mostra un esempio relativo alla fotografia di Figura 5 con identificazione dei bordi attraverso il metodo LOG e il metodo di Canny [7].



Figura 6. Il metodo di Canny, immagine a sinistra, ed il metodo LOG, immagine a destra, all'opera per scovare i bordi.

Sorprendentemente pare che anche nel nostro sistema visivo sia presente tutto ciò che occorre per un filtraggio LOG. Già nel 1865 pare che Ernst Mach avesse osservato che la percezione visiva sembra rinforzare le variazioni spaziali

dell'intensità luminosa. Vari dati fanno pensare che anche la retina compia qualcosa di simile. La comunicazione della retina con il resto del cervello è fatto attraverso le cellule gangliari. La cellula ottiene il suo ingresso, mediato da alcuni neuroni intermedi, da un gruppo di fotorecettori. Quello che i dati sperimentali suggeriscono è che per certe cellule gangliari il campo ricettivo abbia una organizzazione centro-periferia che si avvicina molto al Laplaciano di una Gaussiana. In particolare la luminosità al centro del campo recettivo eccita la cellula gangliare, la luminosità in un anello periferico la inibisce. Infine osserviamo che da un'analisi di metodi come quello di Hildreth-Marr o di Canny sono apparse evidenti connessioni con l'ambito delle equazioni alle derivate parziali. Per esempio: la convoluzione di un dato iniziale con differenti Gaussiane a scala crescente corrisponde esattamente allo studio dell'evoluzione della soluzione dell'equazione di diffusione lineare. L'introduzione di opportune equazioni alle derivate parziali non lineari, il loro legame con la geometria delle immagini e delle caratteristiche che si desiderano estrarre, ha portato rapidamente ad una vasta letteratura ed anche a problemi matematici molto difficili ma anche molto interessanti. Anche lo studio dell'evoluzione di funzioni (immagini) al variare di un parametro di scala (metodi scale-space in Computer Vision) ha portato a feconde ricerche sia dal punto di vista dell'elaborazione di immagini che dal punto di vista della Matematica coinvolta. Rimandiamo solo come esempio ai libri di Sapiro [8] e Guichard e Morel [9] oltre che, ancora a titolo di esempio, all'articolo di fondazione assiomatica di P.L. Lions, Morel e collaboratori [10].

3. Affettiamo un'immagine: la segmentazione.

Per interpretare una scena un primo passo ovviamente consiste nel comprendere che oggi sono presenti. Abbiamo accennato nel paragrafo precedente ad un primo tentativo per individuare alcune caratteristiche degli oggetti: determinare i bordi. Un problema collegato a questo ma su un livello più "sostanzioso", consiste in una visione globale della scena e nella suddivisione della stessa in regioni (insiemi di pixel) costituite da punti omogenei rispetto ad una data caratteristica visiva. Tale problema si chiama *segmentazione* dell'immagine e si pone come obiettivo principale la separazione degli oggetti dallo sfondo della scena e la distinzione tra oggetti differenti per qualche caratteristica. Se indichiamo con R l'intero dominio dell'immagine (pensata ancora come funzione f , consideriamo inoltre ancora immagini a livelli di grigio) possiamo pensare la segmentazione come quel processo che ripartisce R in N parti, sottoinsiemi, R_1, R_2, \dots, R_N tali che valgano le seguenti proprietà:

- (i) $\bigcup_{i=1}^N R_i = R$;
- (ii) R_i è una regione connessa per ogni $i=1, \dots, N$;
- (iii) $R_i \cap R_j = \emptyset$ per ogni i e j con $i \neq j$;
- (iv) $P(R_i) = V$ per ogni $i=1, 2, \dots, N$
- (v) $P(R_i \cup R_j) = F$ per ogni $i \neq j$;

dove P indica un predicato logico sui punti di una certa sottoregione R_i ed il cui valore è V , vero, se i pixel di R_i hanno un valore di livello di grigio che soddisfi certe caratteristiche, F , falso, altrimenti. Ovviamente la definizione andrebbe resa più precisa specificando meglio le proprietà geometriche e topologiche dei possibili sottoinsiemi R_i e dei loro bordi. Quanto detto rende comunque l'idea di segmentazione. Vediamo qualche algoritmo per la separazione di oggetti dallo sfondo. L'approccio più semplice utilizza l'istogramma dell'immagine. L'istogramma di un'immagine digitale a livelli di grigio indica il numero di pixel dell'immagine per ciascun livello di grigio. Dall'istogramma possono essere estratte informazioni interessanti sull'immagine: se la maggior parte dei pixel assume valori di grigio "vicini" ciò significa che l'immagine ha uno scarso contrasto (range dinamico); se nell'istogramma sono predominanti le basse intensità, l'immagine è molto scura e viceversa; un istogramma bimodale denota spesso la presenza di un oggetto abbastanza omogeneo su uno sfondo di luminosità pressoché costante. In Figura 7 si mostra un esempio di immagine che rappresenta alcuni batteri ed il corrispondente istogramma dei livelli di grigio. Operare su un istogramma significa stabilire una mappa h dei livelli di grigio, $h: [0, 255] \rightarrow [0, 255]$ e nella sostituzione del valore $F(i, j)$ con il valore $h(F(i, j))$ per ogni pixel dell'immagine. Nel caso dell'isolamento di un solo tipo di oggetto dall'analisi dell'istogramma è possibile determinare un valore di soglia T ragionevole, singola soglia globale T , e cambiare i pixel in una nuova immagine attraverso la mappa

$$h(p) = \begin{cases} 0 & \text{se } p \leq T \\ 255 & \text{altrimenti} \end{cases}$$

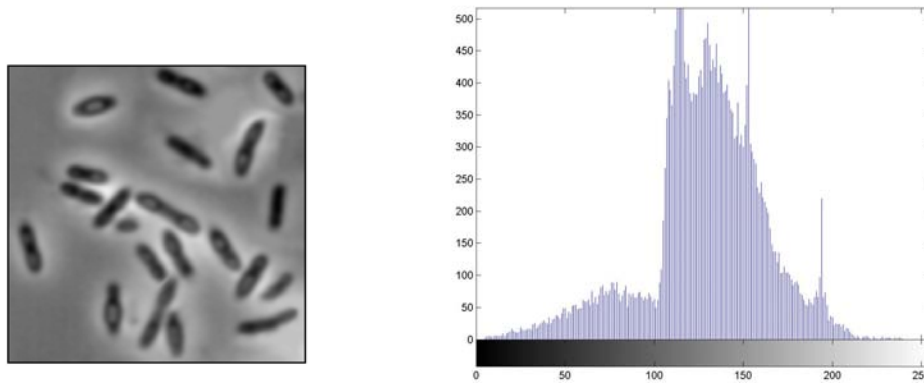


Figura 7. Un'immagine a livelli di grigio ed il suo istogramma.

O scambiando il ruolo di 0 e 255. Applicando questa regola all'immagine in Figura 7 per valori soglia $T_1=50$, $T_2=100$ si ottengono le nuove immagini in Figura 8.

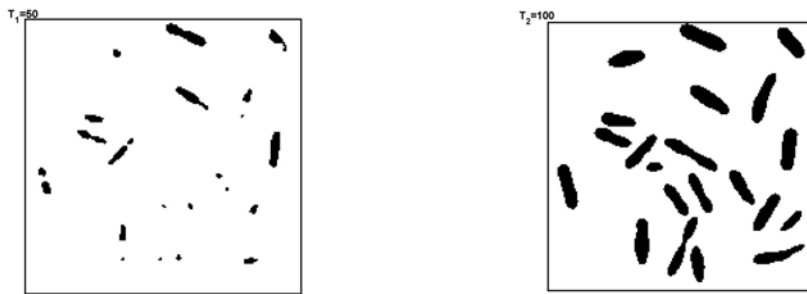


Figura 8. Esempio di semplice segmentazione con soglia globale, $T_1=50$ a sinistra, $T_2=100$ a destra.

A volte migliori risultati sono ottenuti partizionando l'immagine in regioni (per esempio rettangolari) più piccole ed applicando un'elaborazione dell'istogramma locale in ogni sottoregione. Un approccio differente consiste nella scansione dell'immagine pixel per pixel, un pixel P è assegnato ad una regione R adiacente se il criterio di appartenenza è soddisfatto in tale regione in modo 'ottimale'. Questo approccio è detto *Region growing*, si veda Figura 9. Un criterio utilizzato, per esempio, consiste nel confrontare il valore del pixel con il valore medio dei valori di livello di grigio dei pixel della regione. Tale criterio comporta una stima della varianza dei pixel della regione selezionata per un'analisi statistica del procedimento.

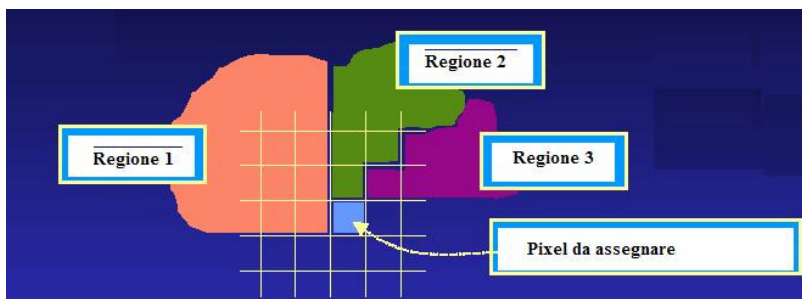


Figura 9. Un passo dell'algoritmo di segmentazione del tipo *Region growing*.

Il risultato dell'algoritmo di tipo Region growing dipende dall'ordine in cui si scansiona l'immagine. Per questo motivo dopo una prima segmentazione si analizzano le regioni individuate e si fa una eventuale fusione tra regioni (*merging*). Anche in questo passo occorre stabilire un criterio per la decisione: solitamente si utilizzano non solo le informazioni statistiche delle regioni ma anche le informazioni relative al loro perimetro e ad eventuali bordi in comune (per una esposizione di qualche criterio si veda per esempio [11]). Come è stato osservato da da Jean-Michel Morel [12], la maggior parte degli algoritmi di segmentazione cercano in realtà di minimizzare qualche energia legata alla segmentazione stessa.

L'idea di introdurre una funzione energia è, in un certo senso, naturale in quanto permette il confronto tra due segmentazioni (partizioni) differenti. Il primo modello riconosciuto di energia di segmentazione è certamente il modello di Mumford-Shah [13] che si basa su seminali idee di David Mumford [14]. Il modello M-S si basa su una approssimazione dell'immagine I con una funzione u regolare a tratti cioè regolare in ogni sottoregione R_i della segmentazione. Le variabili sono rappresentate quindi dalla coppia (u, K) dell'approssimante e dei bordi (edge) K della segmentazione di I . La migliore segmentazione minimizza il funzionale dell'energia $E(u, K)$ che contiene tre termini:

1. un termine che 'controlla' la variazione e la regolarità della funzione u sulle componenti aperte connesse R_i del dominio R dell'immagine e con bordi K ;
2. un termine per controllare la qualità dell'approssimazione dell'immagine I con la funzione u ;
3. un termine per il controllo della lunghezza, la regolarità, la posizione dei bordi K , inibendo fenomeni spuri di sovra-segmentazione.

In generale, dato un sottoinsieme Ω aperto e limitato di \mathbf{R}^2 (nella pratica Ω sarà il dominio dell'immagine I) il funzionale $E(u, K)$ Mumford e Shah si scrive come

$$E(u, K) = \iint_{\Omega \setminus K} \left(\|Du\|^2 + \mu |u - I|^2 \right) dx_1 dx_2 + \alpha l(K \cap \Omega),$$

dove K rappresenta l'unione della famiglia di bordi (incognita a priori), $x=(x_1, x_2)$ denota un punto di Ω , $l(K \cap \Omega)$ indica la somma delle lunghezze dei bordi K mentre u è una funzione scalare differenziabile in $\Omega \setminus K$; con $\|Du\|$ si indica invece la norma Euclidea del gradiente di u . Il dato I è la funzione limitata che definisce l'immagine definita su Ω . La lunghezza dei bordi è misurata attraverso la misura unidimensionale di Hausdorff. Dal punto di vista dell'elaborazione di immagini la scelta dei parametri μ e α corrisponde alla scelta di una data scala e di una soglia di contrasto dell'immagine. Il problema della minimizzazione di $E(u, K)$ e la dimostrazioni di alcune congetture sul funzionale di Mumford e Shah è stato particolarmente fertile per la Matematica coinvolgendo i campi del Calcolo delle variazioni, della teoria geometrica della misura, delle equazioni differenziali alle derivate parziali, dell'Analisi funzionale (solo per citare alcuni ambiti). Molti lavori dedicati al modello M-S provengono dalla scuola italiana (solo per citare qualche studioso, Ennio De Giorgi, Luigi Ambrosio, Gianni Dal Maso, Sergio Solimini, Antonio Leaci, Massimo Gobbino, Franco Tomarelli, Alessandro Sarti, Giovanna Citti, ecc...), e dalla scuola francese (in particolare Jean-Michel Morel, Alexis Bonnet e Guy David). Non tenteremo nemmeno di dare qui una descrizione dei risultati ottenuti, osserviamo solo che anche per il modello M-S si è ipotizzata una certa plausibilità dal punto di vista della dinamica di reti neurali. Infine va almeno citata la nozione innovativa di problemi a discontinuità libere (*free discontinuity problems*) introdotta da Ennio De Giorgi [15] che inserisce anche il modello M-S in un quadro funzionale più generale. Anche i problemi di analisi numerica legati all'effettiva minimizzazione del funzionale M-S hanno portato contributi alla Matematica oltre a permettere un utilizzo efficace dell'approccio con il funzionale dell'energia. In Figura 10 si mostra un esempio di segmentazione di immagine tramite il funzionale di Mumford e Shah.

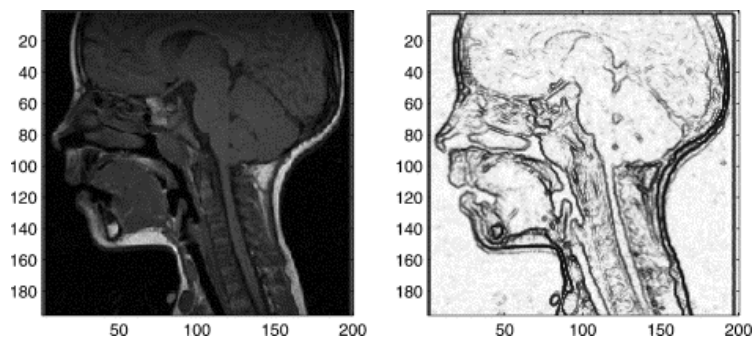


Figura 10. Un esempio di segmentazione con il funzionale di Mumford e Shah (tratto da *Calculus of variations and image segmentation*, M. Carriero, A. Leaci, F. Tomarelli, *Journal of Physiology-Paris*, Vol. 97 (2-3), pp. 343-353, 2003).

4. Proviamo a concludere?

Siamo partiti dal problema della percezione visiva, in seguito abbiamo fatto cenno a due possibili operazioni che hanno probabilmente un corrispondente nel sistema visivo. È infatti ipotesi ampiamente accettata che i primi stadi della visione codifichino alcune proprietà semplici e utili in una scena e per la interpretazione. L'attenzione poi focalizza, integra, e seleziona le caratteristiche presenti in alcune posizioni particolari. Abbiamo brevemente trattato, ambientandolo tra le immagini digitali a livelli di grigio, il problema dell'estrazione dei bordi ed il problema collegato della segmentazione.

Ovviamente abbiamo saltato e nemmeno accennato a molti altri problemi e completamente ignorato le immagini a colori: questo breve articolo vuole solo essere un'introduzione ai molteplici aspetti della visione computazionale e delle sue relazioni con la l'elaborazione delle immagini e con la Matematica (con condiviso vantaggio). Varie aree della corteccia visiva operano in modo specializzato rispetto all'informazione visiva. Le connessioni tra le aree differenti permettono il flusso bidirezionale tra le varie aree e permettono anche di risolvere eventuali conflitti tra cellule nervose che hanno capacità differenti ma che reagiscono allo stesso stimolo. Un tipico esempio di questa situazione sono le figure con contorni illusori come quelle introdotte da Gaetano Kanizsa (noto psicologo sperimentale triestino), si veda Figura 11.

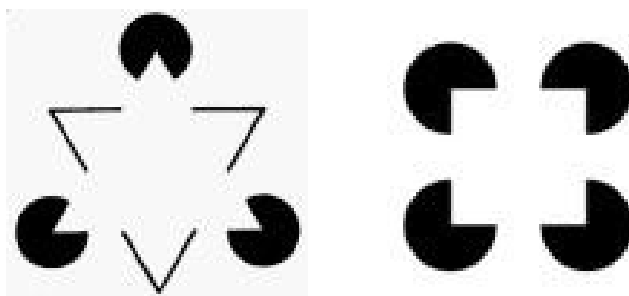


Figura 11. Esempi di immagini di Kanizsa con contorni illusori.

In questo caso, le cellule sensibili alla forma (area V1), non rispondono all'illusione e non segnalano la presenza dei bordi. Altre cellule della corteccia (area V2) ricevono l'ingresso dalle cellule precedenti ma, avendo un campo recettivo più ampio e funzioni differenti, sostanzialmente "ipotizzano" l'esistenza di una linea. Per risolvere il conflitto le due famiglie di cellule riinviano i segnali e si forma l'illusione dei bordi. Un'altra osservazione riguardo alle illusioni di Kanizsa consiste nel fatto che il sistema visivo non si ferma ad elaborazioni locali ma si basa anche sul contesto generale (altrimenti si vedrebbero solo bordi come regioni di transizione tra livelli di luminosità differenti). La percezione visiva rimane ancora in larga parte affascinante ma misteriosa. Allo stato attuale, almeno nella conoscenza di chi scrive, una teoria scientifica che tenta di dedurre leggi generali per la visione e che appare abbastanza generale oltre che promettere una buona formalizzazione appare essere la Gestalt theory. Il programma della Gestalt risale a Max Wertheimer con un lavoro del 1923 [16]. Nel programma di Wertheimer ci sono due tipi di leggi. Nella prima famiglia di leggi si parte da singoli dati locali e si costruiscono ricorsivamente gruppi più ampi nell'immagine percepita. Ogni gruppo riguarda una particolare qualità (colore, forma, direzione, ...). Una seconda classe di principi riguarda la cooperazione e i conflitti tra le leggi precedenti. Nel 1975, Wolfgang Metzger [17] fornisce una ampia panoramica di cinquant'anni di ricerche nella Gestalt. Lo sforzo attuale consiste nel tradurre il programma della Gestalt in un programma di Visione Computazionale attraverso la Matematica: il progetto sembra avvincente e promettente [18].

Concludiamo con un ultimo aspetto che riguarda ogni sistema visivo ma che, al solito, considereremo nel caso di immagini digitali a livelli di grigio. Il problema consiste nel 'comprimere' le informazioni contenute in un'immagine al fine di selezionare solo quelle significative e al fine di ottimizzare l'occupazione di memoria e la velocità di trasmissione. Il tipo di tecniche coinvolte sono, probabilmente in modo inconsapevole, utilizzate da chiunque utilizzi un computer, la rete o semplicemente una macchina fotografica digitale (si pensi ai formati jpeg per le immagini mpef per i video, mp3 per i file audio). Facciamo un solo esempio per capire il problema. Consideriamo per semplicità un vettore s in una dimensione e non un'immagine (matrice di pixel). Supponiamo che i valori di s siano,

$$s = [10 \ 20 \ -5 \ 15 \ 12 \ 12 \ 30 \ 25]$$

invece del segnale originale consideriamo una sua rappresentazione equivalente: ad ogni coppia di valori a, b sostituiamo i valori della media e della loro differenza,

$$a, b \Rightarrow m = (a+b)/2 ; d = (a-b)/2.$$

Operando, come detto per ogni coppia di valori contigui si ottiene,

$$s = [10 \ 20 \ -5 \ 15 \ 12 \ 12 \ 30 \ 25] \Rightarrow s^1 = [15 \ -5 \ 5 \ -10 \ 12 \ 0 \ 27.5 \ 2.5]$$

per esempio per la prima coppia $a=10$, $b=20$, quindi $m=(20+10)/2=15$, $d=(10-20)/2=-5$ e analogamente per le altre coppie. A questo punto possiamo raggruppare nelle prime posizioni tutte le medie e nelle restanti posizioni del vettore s^1 le differenze:

$$s^1 = [15 \ -5 \ 5 \ -10 \ 12 \ 0 \ 27.5 \ 2.5] \Rightarrow \underline{s}^1 = [15 \ 5 \ 12 \ 27.5 \ | \ -5 \ -10 \ 0 \ 2.5].$$

A questo punto ripetiamo l'operazione media/differenze sulle prime componenti (le quattro medie delle coppie del vettore originale s), e sostituiamo ancora le coppie di valori con le coppie medie/differenze:

$$\underline{s}^1 = [15 \ 5 \ 12 \ 27.5 \ | \ -5 \ -10 \ 0 \ 2.5] \Rightarrow s^2 = [10 \ 5 \ 19.75 \ -7.75 \ | \ -5 \ -10 \ 0 \ 2.5].$$

Ancora raggruppiamo le medie nella prima parte del vettore e le differenze nella seconda metà,

$$s^2 = [10 \ 5 \ 19.75 \ -7.75 \ | \ -5 \ -10 \ 0 \ 2.5] \Rightarrow \underline{s}^2 = [10 \ 19.75 \ | \ 5 \ -7.75 \ | \ -5 \ -10 \ 0 \ 2.5].$$

Infine possiamo sostituire alla prima coppia del vettore \underline{s}^2 la coppia media/differenza (questa volta la permutazione finale non è necessaria),

$$\underline{s}^2 = [10 \ 19.75 \ | \ 5 \ -7.75 \ | \ -5 \ -10 \ 0 \ 2.5] \Rightarrow \underline{s}^3 = [14.875 \ | \ -4.875 \ | \ 5 \ -7.75 \ | \ -5 \ -10 \ 0 \ 2.5].$$

Il contenuto d'informazione del vettore \underline{s}^3 è il medesimo del vettore di partenza s . Infatti le operazioni fatte sono invertibili, dalla coppia media/differenza m,d possiamo risalire univocamente alla coppia originale a,b :

$$m, d \Rightarrow a = (m+d) ; b = (m-d).$$

Ogni singolo passo è quindi invertibile e dalle componenti dell'ultimo vettore possiamo ricostruire il vettore originale (ovviamente occorre tenere conto delle permutazioni fatte). In Figura 12 si mostra il vettore originale, abbiamo collegato i punti con una spezzata, ed il vettore trasformato finale (in ascissa ci sono gli indici delle componenti).

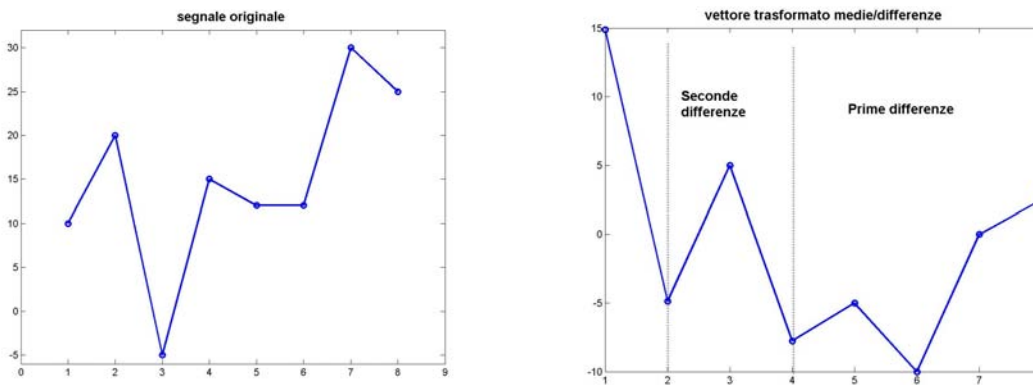


Figura 12. Rappresentazione grafica di un vettore e del vettore trasformato con successive operazioni di media/differenza.

I due grafici in Figura 12 sono poco significativi, occorre interpretare il vettore trasformato. Le medie successive indicano l'andamento locale medio dove, però, la località cambia nel senso che si mediano valori prima adiacenti poi coppie di medie e via dicendo. In sostanza intervengono valori in intorni sempre più ampi, in un certo senso stiamo guardando il segnale a scale differenti e ci spostiamo da una scala piccola (valori adiacenti) ad una scala grande (ultime medie). Allo stesso modo le differenze indicano la variazione locale e su scale differenti. A questo punto: come selezionare le informazioni? Un primo metodo consiste nello stabilire una soglia e nel porre uguali a zero le differenze in modulo sotto la soglia prescelta. In altri termini: se la differenza è considerata piccola possiamo rappresentare i valori di una coppia attraverso il loro valore medio. Potremmo commettere un errore, anzi quasi certamente lo commetteremo, ma stiamo accettando di fare un'approssimazione (controllata). In Figura 13 si mostra un esempio di tale procedimento con soglia $T_1=5$ e, rispettivamente $T_2=8$, quindi ponendo uguale a zero le componenti di tipo differenza in modulo minore od uguale alla soglia e ricostruendo il segnale dal

vettore trasformato e con differenze alterate. Nei casi reali molte componenti differenza saranno azzerate e si mettono in atto tecniche opportune di memorizzazione di vettori con molti zeri (vettori o matrici sparse) con un guadagno in termini di occupazione di memoria (attenzione, senza un trattamento particolare del vettore modificato non ci sarebbe nessun vantaggio). Le domande che si pongono sono varie, tra cui: possiamo stimare l'errore tra il vettore originale e quello modificato? Vi è una alterazione ragionevole del vettore originale in modo da mantenere proprietà qualitative essenziali? È possibile generalizzare il procedimento? È possibile utilizzare operazioni più sofisticate di media e differenza con un maggiore guadagno?

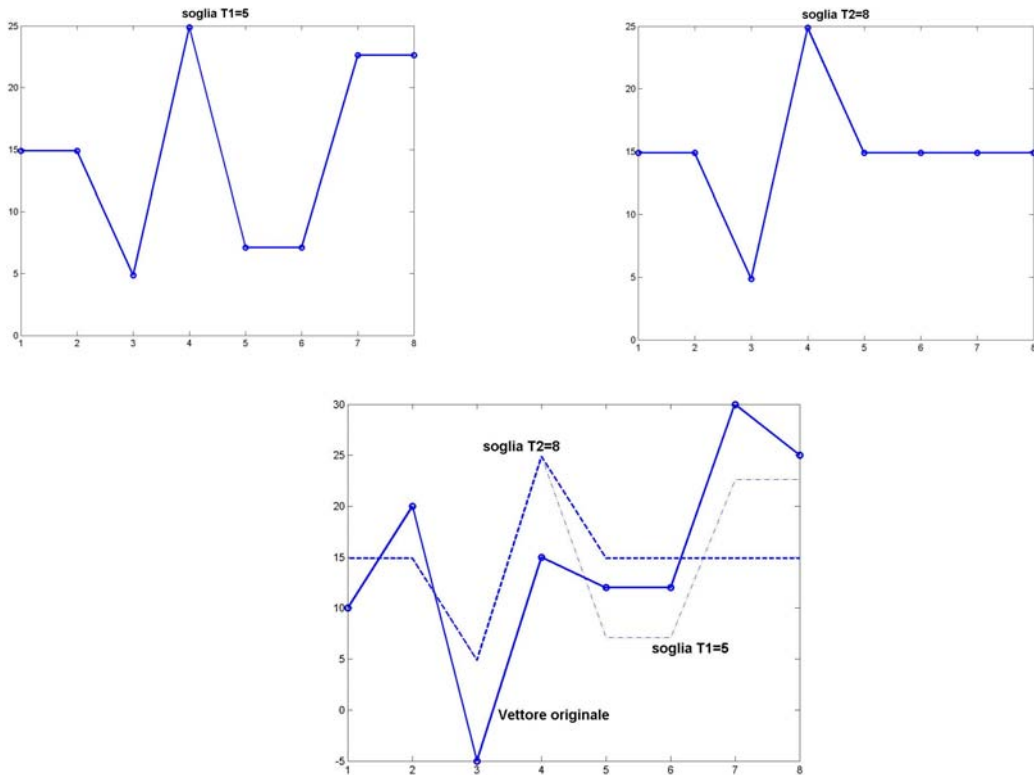


Figura 13. Segnali ricostruiti dopo operazione di azzeramento in base a soglie differenti delle componenti di tipo differenza.

Il procedimento visto può anche essere interpretato utilizzando l'algebra lineare, le diverse rappresentazioni corrispondono a diverse basi. Nell'esempio fatto si passa dalla base canonica di vettori $e_i = [0 \ 0 \ \dots \ 1 \ \dots \ 0]$, con un solo 1 nella posizione i -esima, ad una base detta base di Haar. Cambiando punto di vista e passando da vettori a funzioni e spazi funzionali si costruiscono nuove basi per vari spazi: le basi di Wavelet. A questo punto il racconto diventa di estremo interesse sia Matematico che applicativo ma, come si suol dire, è un'altra storia: speriamo di poterla raccontare in una prossima occasione. Citiamo solo alcuni lavori sulle Wavelet [19,20,21] e rimandiamo alle risorse disponibili in rete per una ricerca personale che lasciamo al lettore interessato (per esempio partendo dal link www.wavelet.org).

Riferimenti Bibliografici

- [1] G.C. Rota, *Osservazioni sull'Intelligenza artificiale*, Bollettino UMI, Conferenza tenuta all'Università di Bologna nel settembre 1985 nell'ambito delle manifestazioni per IX centenario della sua fondazione.
- [2] R. C. Gonzalez, R. E. Woods, *Digital Image Processing*, Addison-Wesley, Reading MA, 1992.
- [3] K.R. Castleman, *Digital Image Processing*, Prentice Hall, 1996.
- [4] A. Rosenfeld, A.C. KaK, *Digital Picture Processing*, Academic Press, New York, 1982.
- [5] D. Marr, *Vision*, W.H. Freeman and Company, 1982.
- [6] D. Marr, E. Hildreth. *Theory of edge detection. Proc. Roy. Soc. Lond.*, B207:187-217, 1980.
- [7] J. Canny, *A computational approach to edge detection*, IEEE Trans. On Pattern Analysis and Machine Intelligence, 8(6),pp679-698, 1986.
- [8] G. Sapiro, *Geometric Partial Differential Equations and Image Analysis*, Cambridge University Press, 2001.
- [9] F. Guichard, J.-M. Morel, *Image Analysis and PDE's*, 2001.
- [10] L. Alvarez, F. Guichard, P.-L. Lions, and J.-M. Morel. *Axioms and fundamental equations of image processing: multiscale analysis and pde*. Arch. for Rat. Mech., 16(9):200-257, 1993.
- [11] A.K. Jain, *Fundamentals of digital image processing*. Prentice Hall, Englewood Cliffs, NJ, 1989.
- [12] J.-M. Morel, S. Solimini, *Variational Methods in Image Segmentation*. Birkhäuser, Berlino, 1995.
- [13] D. Mumford, J. Shah, Optimal approximations by piecewise smooth functions and associated variational problems. *Commun. Pure Appl. Math.* **XLII** (1989), pp. 577–685.
- [14] D. Mumford, Bayesian rationale for the variational formulation. In: *Geometry-Driven Diffusion in Computer Vision*, Kluwer, Dordrecht, 1994.
- [15] E. De Giorgi, Free discontinuity problems in calculus of variations. In: R. Dautray, Editor, *Frontiers in Pure and Applied Mathematics*, North-Holland, Amsterdam, pp. 55–61, 1991.
- [16] M. Wertheimer. Untersuchungen zur lehre der gestalt, II. *Psychologische Forschung*, 4:301, 350, 1923.
- [17] W. Metzger. *Gesetze des Sehens*. Waldemar Kramer, 1975.
- [18] A. Desolneux, L. Moisan and J.-M. Morel, *A theory of digital image analysis*, Lecture Notes, Ecole Normale Supérieure de Cachan, 2004.
- [19] Y. Meyer. *Ondelettes et operatrurs I: Ondelettes*. Herman Paris, 1990.
- [20] S. Mallat, *A Wavelet tour of signal processing*. Academic Press, 1998.
- [21] W. Sweldens, The lifting scheme: a construction of second generation Wavelets. *SIAM J. Math. Anal.*, 1997.