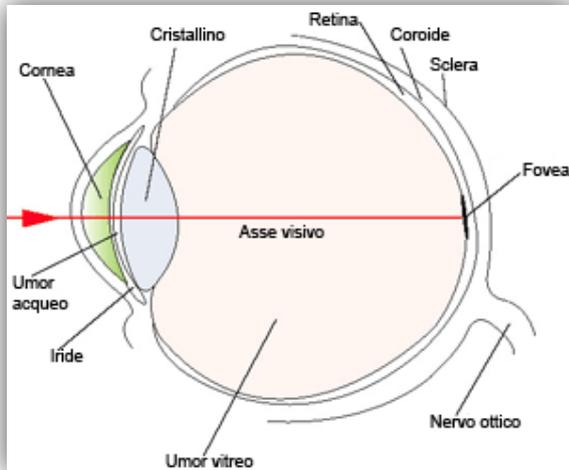


Capitolo 2

FONDAMENTI DELLE IMMAGINI DIGITALI

Occhio Umano

L'occhio umano ha una forma pressoché sferica con un diametro di circa 20 mm. E' circondato da tre membrane: *cornea e sclera* (membrane esterne); *coroide; retina* (membrana interna).



Per la nostra trattazione ci interessano le seguenti informazioni:

La **coroide** ha uno strato di colore scuro che riduce la quantità di luce che penetra nel bulbo oculare. Inoltre nella parte anteriore si divide in *corona ciliare e iride*. Quest'ultima si contrae e dilata (con un diametro da 2 a 8 mm) controllando la quantità di luce che entra nell'occhio.

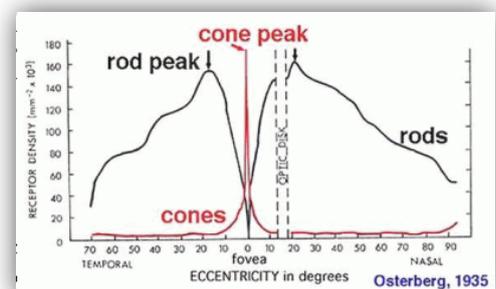
La **retina** è una membrana interna sulla quale viene impressa la luce che proviene da un oggetto. Sulla retina sono disposte due classi di recettori di luce e la loro distribuzione determina la visione vera e propria. Nella retina è presente anche un punto cieco dovuto all'assenza di recettori.

Le *due classi di recettori* sono:

Coni: sono posizionati nella parte centrale della retina chiamata fovea e sono molto sensibili ai colori. Sono circa 7 milioni. I dettagli fini colti dagli esseri umani sono dovuti ai coni ed al fatto che ognuno di essi è collegato al proprio terminale nervoso.

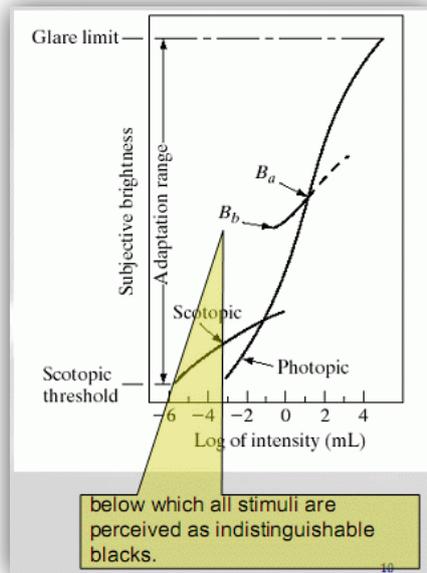
La visione legata ai coni viene chiamata *fotopica*.

Bastoncelli: sono distribuiti in tutta la superficie della retina; variano da 75 a 150 milioni. I dettagli percepiti dai bastoncelli sono meno dettagliati in quanto sono tutti collegati ad un solo terminale nervoso. Inoltre sono sensibili fino a bassi livelli d'illuminazione e non sono coinvolti nella percezione dei colori. La visione legata ai bastoncelli viene chiamata *scotopica*.



Legato alla corona ciliare grazie a delle fibre c'è il **cristallino** formato per il 70% da acqua. Il cristallino è importante perché la lunghezza focale necessaria è raggiunta variando la forma proprio del cristallino, in quanto la distanza tra cristallino e retina (cioè la nostra regione di imaging) è fissa a circa 17 mm. Le fibre della corona ciliare ispessiscono o assottigliano il cristallino, rispettivamente per oggetti vicini e lontani. L'immagine della retina è focalizzata principalmente nella regione della fovea. La percezione poi utilizza la sollecitazione dei recettori luminosi, che trasformano l'energia radiante in impulsi elettrici, che infine vengono decodificati dal cervello.

La *luminosità soggettiva* (intensità percepita dal sistema visivo umano) è una funzione logaritmica dell'intensità della luce incidente sull'occhio.



La linea curva rappresenta le intensità a cui la vista può adattarsi. Il sistema visivo umano non riesce a interessare simultaneamente tutti i valori del range, ma risolve il problema variando la propria sensibilità; questa variazione viene definita *adattamento alla luminosità*. Il passaggio dalla visione scotopica a quella fotopica avviene gradualmente. Il range dei livelli distinti di intensità che l'occhio può discriminare simultaneamente è inferiore rispetto a quella totale di adattamento.

Il **rapporto di Weber** e' definito:

$$\frac{\Delta I_c}{I}$$

dove ΔI_c è l'incremento dell'illuminazione che risulta discriminabile rispetto all'illuminazione I dello sfondo.

Un valore basso del rapporto indica che anche solo una piccola percentuale di variazione è discriminabile. Un valore alto del

rapporto indica che è necessaria una maggiore percentuale di variazione di intensità e si ha quindi una scarsa discriminazione della luminosità.

Quindi la capacità discriminativa rispetto alla luminosità è scarsa (rapporto di Weber aumenta) ai livelli più bassi di illuminazione, e aumenta (rapporto di Weber diminuisce) al crescere dell'illuminazione. Questo avviene perché con poca illuminazione la vista utilizza i bastoncelli, mentre con molta illuminazione vengono utilizzati i coni.

Acquisizione delle immagini

L'*acquisizione delle immagini* avviene seguendo il seguente processo:

l'energia in entrata è trasformata in tensione elettrica grazie alla combinazione dell'elettricità in input e del materiale del sensore sensibile a un particolare tipo dell'energia rilevata. La forma d'onda della tensione di uscita è la risposta del sensore, e una quantità digitale viene ottenuta da ciascun sensore attraverso la digitalizzazione di questa risposta. La risposta di ogni singolo sensore è proporzionale all'integrale dell'energia luminosa proiettata sulla superficie del sensore.

Nell'ambito delle macchine fotografiche digitali vengono utilizzati i CCD (Charge-Coupled Device) e i CMOS (Complementary Metal Oxide Semiconductor).

I processi di fabbricazione dei due sensori sono differenti, così come lo è la disposizione dei circuiti su di essi, fermo restando che si tratta sempre di piastrine piene di fotodiodi che raccolgono la luce e la convogliano.

Ai fini fotografici la differenza più rilevante è quella relativa proprio alla *raccolta di luce*:

- nei **CCD** la carica elettrica immagazzinata dai singoli fotodiodi viene trasferita, accumulandosi man mano lungo le file di fotodiodi, fino ai bordi del sensore, dove poi viene amplificata ed infine convertita in un segnale digitale (da un apposito ADC, Analog-to-Digital Converter). In pratica la carica elettrica viene letta una riga alla volta, e poi il parziale (di ogni riga) viene riportato alla riga successiva e così via, in sequenza, fino a coprire l'intero sensore, quindi viene trasportata della carica elettrica.
- i sensori fabbricati con un processo di tipo **CMOS**, invece, lavorano diversamente: ogni fotodiodo dispone di un amplificatore e di un convertitore, quindi la carica elettrica accumulata viene convertita in differenza di potenziale – il cui trasporto richiede molta meno energia, quindi a parità di altre condizioni, un sensore CMOS consuma meno di un sensore CCD.

Punti fermi riguardo i sensori di tipo CMOS:

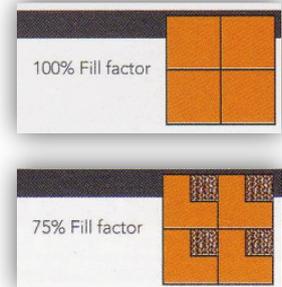
- tendono ad essere di più facile fabbricazione e più economici
- consentono di implementare sul sensore dei componenti che i sensori CCD non ospitano (l'amplificatore, l'ADC), e questo porta alla realizzazione di chip più piccoli: ciò spiega perché i sensori di tipo CMOS sono la norma sui cellulari, sulle fotocamere compatte, e così via
- a parità di altre condizioni consumano (e scaldano) meno

A questo punto verrebbe da chiedersi perché si trovino ancora in giro dei CCD e la risposta è: qualità d'immagine. I sensori CCD hanno le potenzialità per offrire una maggiore gamma dinamica, meno rumore e maggiore sensibilità.

In generale i sensori CMOS hanno due limiti: il rumore e la sensibilità. Poiché c'è un amplificatore per ogni fotodiodo, basta una minima disuniformità nel funzionamento di uno o più di questi amplificatori per generare pixel irregolari e/o disturbati.

Inoltre, proprio la maggiore presenza di circuiteria sul sensore genera più rumore rispetto ad un CCD perché un po' di rumore accompagna inevitabilmente ogni componente elettronico.

Quanto alla sensibilità, sempre perché i sensori CMOS ospitano più circuiti (rispetto ai CCD), ne deriva che una parte della loro superficie non è destinata alla raccolta di luce (la percentuale di un punto realmente utilizzata per raccogliere luce si chiama "fill factor") ma appunto ad ospitare tali circuiti, mentre un CCD utilizza tutta la superficie per raccogliere la luce.



Abbiamo detto che un'immagine viene definita come una funzione bidimensionale $f(x, y)$ con valori discreti. $f(x, y)$ deve essere finita e diversa da zero; solitamente ciò si indica con l'intervallo $[0, L-1]$ dove 0 indica il nero, e $L-1$ indica il bianco (ovviamente si parla di un'immagine monocromatica).

La convenzione vuole che l'*origine* di un'immagine digitale si trovi nella parte in alto a sinistra, con l'asse positivo delle x che si estende verso il basso, l'asse positivo delle y verso destra.

Inoltre è caratterizzata da due componenti:

- **illuminazione**: quantità di luce diretta incidente sulla scena. La sua natura è determinata dalla fonte di illuminazione.
- **riflettanza**: quantità di luce riflessa dagli oggetti presenti nella scena. La sua natura è determinata dalle caratteristiche degli oggetti.

Queste due componenti danno origine alla $f(x, y)$:

$$f(x, y) = i(x, y) r(x, y)$$

$$\text{con } 0 < i(x, y) < \infty$$

$$\text{e } 0 < r(x, y) < 1$$

Campionamento e quantizzazione di immagini

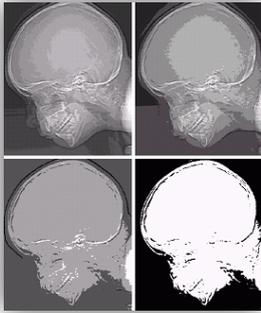
Se un'immagine digitale è una funzione discreta allora dobbiamo convertire i dati continui in forma digitale tramite i due processi di:

- **campionamento**: operazione di digitalizzazione delle coordinate. L'esattezza del campionamento è determinata dalla qualità delle componenti ottiche del sistema.
- **quantizzazione**: operazione di digitalizzazione dell'ampiezza. Quando la funzione bidimensionale discreta ha un valore di intensità intero per ogni coppia di coordinate (x, y) . L'esattezza della quantizzazione è determinata dal livello di rumore del segnale campionato. La qualità di un'immagine digitale dipende in larga misura dal numero di campioni e dai livelli di intensità discreti utilizzati per la quantizzazione e il campionamento.

Il *range dinamico* di un sistema di imaging viene definito come il rapporto tra l'intensità massima misurabile e il livello minimo di intensità rilevabile nel sistema.

Il limite più alto viene indicato come *saturazione* , quello più basso come *rumore* .

Il *contrasto* di un'immagine è la differenza di intensità tra il più alto e il più basso livello di intensità di un'immagine.



La quantizzazione può portare a degli effetti come il *falso contorno* che si crea quando nelle zone omogenee di un'immagine digitale non si utilizza un numero sufficiente di livelli di intensità.

Il numero di bit *b* richiesti per memorizzare un'immagine digitale è:

$$b = N \times M \times k$$

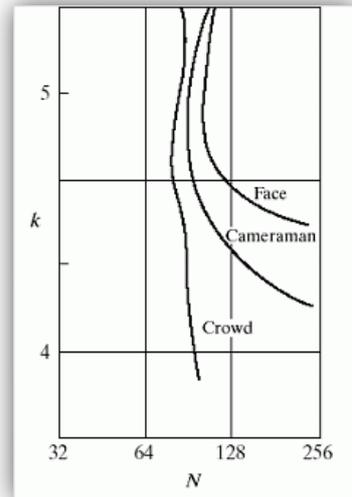
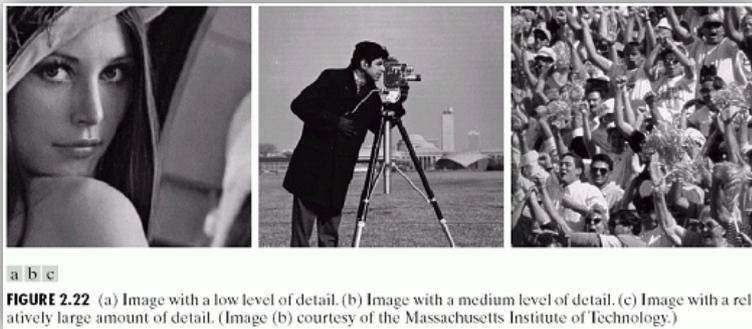
dove *k* è l'esponente che determina il numero di livelli di intensità:

$$L = 2^k$$

Le *curve di isopreferenza* nel piano *Nk* , introdotte da Huang, quantificano in modo sperimentale gli effetti prodotti sulla qualità di un'immagine variando contemporaneamente *N* e *k* (quindi nel caso in cui l'immagine è quadrata e la formula di prima diventa $b = N^2 \times k$). Ogni punto nel piano *Nk* rappresenta un'immagine avente i valori di *N* e *k* uguali alle coordinate di quel punto.

All'aumentare del dettaglio le curve di isopreferenza tendono a diventare sempre più verticali, quindi significa che per le immagini con maggiore dettaglio sono necessari pochi livelli di intensità. Nelle immagini con medio - basso dettaglio la qualità rimane invariata in alcuni intervalli in cui il numero di campioni aumenta, ma il numero dei livelli di intensità decresce. Questo avviene perché la diminuzione di *k* fa aumentare il contrasto apparente, che viene percepito come un miglioramento dell'immagine.

Vediamo un esempio:



Nell'immagine di destra sono rappresentate le curve di isopreferenza delle tre immagini di sinistra. Notiamo che hanno le caratteristiche sopra citate. Infatti la curva della folla è praticamente verticale in quanto contiene una grande quantità di dettagli e per rappresentarla sono necessari solo pochi livelli di intensità. Accade l'opposto nel caso della modella.

Interpolazione di un'immagine

L'interpolazione di un'immagine è il processo mediante il quale si utilizzano dei dati noti di segnale per stimarne apparentemente i valori in locazioni sconosciute. Questo strumento viene utilizzato per lo zoom, il rimpicciolimento, la rotazione e le correzioni geometriche.

Vediamo tre tipi di interpolazione:

- *interpolazione nearest neighbor* : assegna a ogni nuova posizione l'intensità del pixel più prossimo nell'immagine originale. Questo metodo è poco utilizzato perché, anche se semplice, introduce

artefatti.

- interpolazione bilineare: l'intensità da assegnare a ciascuna nuova posizione è stimata tramite i quattro pixel più vicini.

L'equazione che regola questo metodo è

$$v(x,y) = ax + by + cxy + d$$

dove v è il valore dell'intensità; mentre i quattro coefficienti sono determinati a partire dalle quattro equazioni nelle quattro incognite ottenibili utilizzando i quattro pixel più vicini al punto (x,y) .

Ovviamente questo tipo di interpolazione offre dei risultati migliori rispetto alla precedente.

- interpolazione bicubica: è come la bilineare con la differenza che questa volta consideriamo i 16 pixel più vicini. Quindi avremo 16 coefficienti determinati con 16 equazioni in 16 incognite.

Relazioni di base tra i pixel

L'intorno di un pixel può essere in due formati:

- 4-intorno o N4(p): quando consideriamo i quattro pixel orizzontali e verticali, vicini al pixel p di coordinate (x,y) , le cui coordinate sono:

$$(x+1,y) (x-1,y) (x,y+1) (x,y-1)$$

Ciascun pixel si trova ad una distanza unitaria da (x,y) . Nel caso in cui il pixel p si trovi vicino ai bordi dell'immagine, l'intorno dei pixel può non essere definito.

Oppure quando consideriamo i vicini diagonali, che hanno coordinate:

$$(x+1,y+1) (x+1,y-1) (x-1,y+1) (x-1,y-1)$$

- 8-intorno o N8(p): quando consideriamo sia i vicini orizzontale e verticali che quelli diagonali.

Definiamo due tipi di **adiacenza**:

- 4-adiacenza: due pixel p e q con valori in V sono 4-adiacenti se q è nell'insieme $N4(p)$
- 8-adiacenza: due pixel p e q con valori in V sono 8-adiacenti se q è nell'insieme $N8(p)$

dove V è l'insieme dei valori di intensità utilizzati per definire l'adiacenza.

Definito S come un sottoinsieme di pixel in un'immagine, due pixel p e q si dicono *connessi* in S se esiste un path tra di essi formato interamente di pixel appartenenti a S . La **componente connessa** di S è l'insieme dei pixel connessi, in S , a ciascun pixel p . S viene detto connesso se all'interno ha un'unica componente connessa.

Definito R come un sottoinsieme di pixel in un'immagine. Chiamiamo R una **regione** dell'immagine se R è un insieme connesso.

L'insieme di tutte le regioni presenti in un'immagine forma il *foreground* (primo piano), mentre il suo complemento forma il *background* (sfondo).

Le regioni possono essere adiacenti, se la loro unione forma un insieme connesso, o disgiunte, quando non sono adiacenti.

Il *bordo* di una regione R è l'insieme dei punti che sono adiacenti a punti dell'insieme complementare di R , cioè l'insieme dei pixel che hanno almeno un vicino nel background.

Quando R coincide con l'immagine intera allora i suoi bordi coincidono con l'insieme dei pixel che si trovano nella prima e ultima riga e nella prima e ultima colonna dell'immagine.