Appunti su alcune parti del corso

di

Elaborazioni delle Immagini

Prof.ssa Paola Campadelli

Anno 1998/99

Indice

Capitolo 1

Riflettiamo su che cosa significa vedere

1.1 Introduzione

Nel corso di elaborazione delle immagini studieremo le operazioni che a partire da una immagine portano a rappresentazioni via via più astratte sino ad ottenere una immagine suddivisa in regioni (immagine segmentata) che dovrebbero corrispondere ad *oggetti distinti* di una scena. Le regioni possono essere descritte mediante la definizione di opportune caratteristiche e, giunti a questo stadio, è possibile tentare una classificazione degli oggetti o fornire un'interpretazione della scena.

Per capire le difficoltà della interpretazione dobbiamo riconoscere che "elaborare immagini" nel senso di arrivare alla interpretazione di una scena equivale a vedere.

Rifletteremo pertanto su cosa significa vedere e lo faremo utilizzando il lavoro di psicologi, fisiologi e teorici della visione. Accenneremo al problema perché una volta capita l'enorme complessità del processo visivo sapremo meglio valutare i limiti degli strumenti e dei metodi che consentono di operare sulle immagini in modo utile.

Il compito del sistema visivo (e dei sistemi di percezione) è quello di fornire all'animale le informazioni ad esso necessarie per interagire con l'ambiente.

J.J. Gibson, uno psicologo che ha dedicato 50 anni della sua vita a capire che cosa sia la visione, propone come livello adeguato di studio quello che egli definisce un "approccio ecologico" al problema della percezione e della percezione visiva in particolare. Egli enfatizza il fatto che per capire quanto complesso sia e cosa significhi vedere occorre esaminare l'ambiente in cui l'animale (l'uomo) vive. L'ambiente è ciò che l'animale percepisce; il sistema nervoso degli organismi viventi si è evoluto per interagire con l'ambiente, i sistemi visivi di animali che vivono in ambienti differenti hanno caratteristiche differenti. Per inciso, sottolineo che i sistemi di visione automatica costruiti sino ad ora operano in ambienti molto semplici e ben definiti.

Un accennno all'analisi della percezione visiva fatta da Gibson ci aiuterà a capire quanto complesso sia il compito; accenneremo poi alla "teoria computazionale della visione" sviluppata da P. Marr per vedere come il problema è stato posto dal punto di vista dell'elaborazione dell'informazione. Presenteremo il punto di vista generale della teoria, il modo in cui il problema della visione viene posto, non analizzeremo che pochi elementi delle soluzioni proposte, soluzioni che sono tutt'oggi oggetto di sperimentazione.

Cercherò ora di sintetizzare alcune osservazioni effettuate di Gibson, osservazioni che, anche se apparentemente semplici, mi sembrano estremamente interessanti. L'osservazione fondamentale da cui parte Gibson è che la *visione naturale* si ha quando un organismo è *libero di muoversi* nell'ambiente; essere libero di muoversi, di muovere la testa, gli occhi, consente infatti ad un organismo di cambiare il punto di vista e cambiare il punto di vista permette di acquisire una informazione molto più ricca della scena che si sta osservando.

Poiché la visione è fondamentale per un organismo che deve agire in un certo ambiente (pensate ad un animale ma anche ad un robot) è importante chiedersi che cos'è l'ambiente.

1.2 Da J.J.Gibson - The ecological approach to visual perception -

L'animale e l'ambiente

L'ambiente è ciò che circonda gli organismi che hanno sistemi percettivi e che sono in grado di agire. In altre parole l'ambiente è ciò che circonda gli animali; esso include altri animali, piante e cose non viventi.

Gli animali necessitano di distinguere non solo le sostanze e gli oggetti dell'ambiente ma anche gli altri animali e le differenze tra essi. Non possono permettersi di confondere la preda con il predatore, la propria specie con un'altra specie, il maschio con la femmina.

Si può asserire che le parole *animale e ambiente* sono una coppia non separabile; ognuno dei due termini implica l'altro. Nessun animale potrebbe esistere senza un ambiente che lo circonda, allo stesso modo la parola ambiente implica un animale. La superficie della terra milioni di anni prima che la vita si sviluppasse non era un ambiente, era una realtà fisica, un ambiente potenziale, prerequisito all'evoluzione della vita; potremmo chiamarla mondo ma non era un ambiente.

Ogni animale è un soggetto che *percepisce* l'ambiente e *agisce* sull'ambiente. Ciò non significa che esso percepisca il mondo fisico (quello descritto dalla fisica) e che si comporti nello spazio e nel tempo descritti dalla fisica ...

Il mondo fisico comprende ogni cosa dagli atomi, agli oggetti terrestri, alle galassie. Queste cose esistono a differenti livelli di grandezza che vanno dai quasi inimmaginabili estremi costituiti dagli atomi e dalle galassie. Nessuno di questi estremi è un ambiente. Il livello a cui un ambiente esiste è un livello intermedio misurato in millimetri e metri; anche le dimensioni degli animali si situano in questo livello intermedio. In breve le dimensioni e masse delle cose che compongono un ambiente sono comparabili a quelle degli animali. Inoltre le cose più piccole sono nidificate in quelle più grandi ...

I livelli di descrizione del mondo forniti dalla fisica moderna vanno dal livello atomico a quello cosmico; questi livelli sono inappropriati per lo studio della percezione. Chi studia la percezione può essere interessato solo a cose che fanno parte dell'habitat dell'uomo o dell'animale perché noi ci comportiamo rispetto alle cose che possiamo guardare, ascoltare, toccare, gustare, odorare. I sistemi percettivi degli animali non sono in grado di cogliere atomi e galassie ...

La spiegazione di come noi osservatori umani o almeno alcuni di noi possono *visualizzare* un atomo o una galassia anche se non lo possono *vedere* non è tanto un problema di percezione quanto di ragionamento, conoscenza. E' importante capire prima come avviene che noi percepiamo l'ambiente, come apprendiamo le stesse cose che i nostri antenati percepivano prima di imparare a conoscere atomi e galassie. Ci interessa ciò che chiameremo *percezione diretta* e non la *percezione indiretta* che si può apprendere (comprendere) attraverso microscopi, telescopi, fotografie, films e ancora meno il tipo di comprensione che si ha attraverso il linguaggio o la scrittura ...

La base dell'ambiente terrestre è il suolo ... *strutturato in unità di varie dimensioni metriche* nidificate le une nelle altre. Queste unità tendono a ripetersi su tutta la superficie della terra, i granelli di sabbia tendono ad essere della stessa grandezza ovunque. I fili d'erba sono più o meno simili l'uno all'altro; queste unità naturali non sono naturalmente perfettamente uniformi come le piastrelle di un pavimento, esse sono stocasticamente regolari. In altre parole le unità che compongono il suolo non diventano più piccole quando uno si muove per esempio verso il nord ...

Un' altra differenza tra un ambiente e il mondo della fisica è la *scala temporale* dei processi e degli eventi che noi scegliamo di considerare; la durata dei processi nell'ambiente è misurata solo in anni e in secondi; la durata della vita degli animali si misura in questo ambito; le variazioni che possiamo percepire, quelle da cui dipendono le nostre azioni, non sono né estremamente lente né estremamente rapide. Gli osservatori umani non possono percepire l'erosione di una montagna, ma possono riconoscere la caduta di un masso; possono notare lo spostamento di una sedia in una stanza ma non quello di un elettrone in un atomo ... L'astratto flusso del tempo privo di eventi, per quanto sia un concetto utile in fisica, non ha realtà per un animale. Noi non percepiamo il tempo ma processi, sequenze, cambiamenti, o così io assumo. La consapevolezza degli uomini per il tempo dell'orologio, il tempo socializzato è un'altra cosa ...

Non sono tanto importanti i concetti di spazio e di tempo quanto quelli di *permanenza e cambiamento*. La configurazione dell'ambiente è permanente per certi aspetti e variabile per altri. Un osservatore può riconoscere la stessa stanza in diverse occasioni pur percependo le variazioni, o un bambino che cresce pur notandone la crescita... Nulla è permanente per sempre, pertanto è meglio parlare di persistenza nel cambiamento.

La persistenza delle configurazioni geometriche dell'ambiente dipende in parte dal tipo di sostanze che lo compongono, dalla loro rigidità, resistenza alla deformazione ... Naturalmente quando si parla di ambiente permanente ci si riferisce alle sostanze solide ... Quando una sostanza solida di forma costante fonde diciamo che l'*oggetto ha cessato di esistere*. Questo modo di parlare è ecologico non fisico ... L'ecologia chiama questa non persistenza una distruzione dell'oggetto, mentre la fisica parla di cambiamento di stato. Entrambe le asserzioni sono corrette ma la prima è più adeguata rispetto al comportamento di bambini o animali ... La cessazione dell'esistenza, la distruzione di un oggetto sono *eventi ambientali* estremamente importanti da percepire ... Un ambiente normalmente manifesta alcune cose che persistono ed altre che non persistono; alcune caratteristiche sono invarianti, altre variano. Un ambiente completamente invariante, che non cambia in nessuna delle sue parti e privo di movimento, sarebbe completamente rigido e ovviamente non potrebbe più essere un ambiente ... Un mondo che sia privo di cambiamento in molti dei suoi aspetti e modificabile in altri, ma che non sia né morto ad un estremo né caotico all'altro estremo, è di grande importanza per la nostra indagine ...

Riassumendo: l'ambiente degli animali e dell'uomo è ciò che percepiamo. L'ambiente non è il mondo descritto dalla fisica. L'osservatore e l'ambiente sono complementari. I componenti e gli eventi dell'ambiente sono costituiti da unità nidificate. L'ambiente persiste per certi aspetti e cambia per altri. Le variazioni più radicali sono il cessare di esistere e il cominciare ad esistere.

Mezzo, sostanze, superfici

L'ambiente terrestre è descrivibile in termini di *mezzo (medium), sostanze e superfici* che le separano ... Il nostro pianeta consiste soprattutto di terra, acqua aria: un solido, un liquido e un gas ... *L'interfaccia* tra due qualsiasi di questi stati della materia costituisce una *superficie* ... La più importante delle superfici per gli animali terrestri è il suolo, esso è la superficie di supporto. L'aria e l'acqua sono mezzi per la locomozione animale. L'aria è un mezzo migliore per la locomozione, offre minor resistenza, non richiede l'anatomia dei pesci per i movimenti rapidi. Un altra caratteristica dell'aria e dell'acqua è che esse sono generalmente *trasparenti*, trasmettono la luce, mentre un solido è generalmente opaco, assorbe o riflette la luce. Un mezzo omogeneo perciò *consente (offre)* la visione. La luce non solo è trasmessa da un mezzo ma riverbera cioè colpisce le superfici ed è riflessa da esse ad enorme velocità ... Questo riverberante flusso di luce è la condizione che chiamiamo *illuminazione*. La illuminazione riempie il mezzo, c'è una luce ambientale in ogni punto cioè luce che giunge in ogni punto da tutte le direzioni ... Una terza caratteristica dell'aria e dell'acqua è che essere trasmettono vibrazioni o onde di pressione generate da eventi meccanici. Rendono perciò possibile udire ciò che chiamiamo suoni, più esattamente permettono l'ascolto di eventi vibratori. Inoltre esse consentono la rapida diffusione di sostanze chimiche, offrono pertanto la possibilità di odorare una sorgente cioè individuare una sostanza da lontano.

Osserviamo che la locomozione animale di solito non è senza un obiettivo ma è guidata o controllata dalla vista se un animale può vedere, dal suono se un animale può sentire, dall'odore se può odorare ... Il mezzo in cui gli animali si muovono (ed in cui gli oggetti possono essere mossi) è allo stesso tempo il mezzo che consente la diffusione della luce, del suono, dell'odore provenienti da sorgenti dell'ambiente. Ogni punto del mezzo è un possibile punto di osservazione per ogni osservatore che può vedere, sentire, odorare. E questi punti di osservazione sono connessi da percorsi di possibile locomozione. Via via che l'osservatore si sposta da punto a punto l'informazione ottica, acustica, luminosa cambia in accordo con lo spostamento ... Tutti questi fatti circa corpi che si muovono e circa la trasmissione di luce, suono, odore in un mezzo sono consistenti con la fisica, la meccanica, l'ottica, l'acustica, la chimica; ma si pongono ad un livello più elevato.

Un'altra importante caratteristica del mezzo è che esso consente la respirazione; gli animali devono respirare pertanto il mezzo deve essere relativamente costante e relativamente omogeneo ... Non ci sono discontinuità improvvise in un mezzo, non ci sono barriere tra un volume ed un altro, in altre parole non ci sono superfici. Questa omogeneità è cruciale, è ciò che permette alle onde elettromagnetiche e sonore di viaggiare in fronti sferici a partire da una sorgente. Infine una ulteriore caratteristica di un mezzo è la sua intrinseca polarità su-giù. La forza di gravità tira verso il basso non verso l'alto, la luce viene dall'alto non dal basso, dal cielo non dal substrato. A causa della gravità la pressione dell'acqua e dell'aria aumentano verso il basso e decrescono verso l'alto. Il mezzo non è isotropo, ha un asse di riferimento assoluto: l'asse verticale.

Per riassumere, le caratteristiche fondamentali di un ambiente sono che esso offre la respirazione, permette la locomozione, può essere riempito di illuminazione e perciò offre la visione, consente di distinguere vibrazioni ed emanazioni che diffondono e infine ha un asse di riferimento assoluto. Tutte queste *offerte* della natura queste possibilità o opportunità sono invarianti. Sono state straordinariamente costanti attraverso l'evoluzione della vita animale.

Consideriamo ora la porzione dell'ambiente che non trasmette liberamente la luce o gli odori, che non permette il movimento dei corpi o la locomozione degli animali. La materia nello stato solido o semisolido è detta sostanziale. Le sostanze in questa accezione del termine sono più o meno rigide, più o meno resistenti alla deformazione, più o meno penetrabili da corpi solidi e più o meno permanenti nella loro forma. Esse sono usulmente opache alla luce. Inoltre la porzione sostanziale dell'ambiente è, al contrario del mezzo, eterogenea. Le sostanze differiscono nella composizione chimica ... Il suolo, le rocce, la sabbia, il fango, l'olio, il legno, i minerali, i metalli e soprattutto i vari tessuti di animali e piante sono esempi di sostanze dell'ambiente.

La ragione per cui gli animali necessitano di distinguere tra le varie sostanze presenti nell'ambiente è ovvia. Alcune sostanze sono nutritive altre non nutritive, alcune sono tossiche; ed è molto importante per un animale affamato essere in grado di distinguere le sostanze commestibili da quelle non commestibili per mezzo della visione o dell'odorato piuttosto che per mezzo del tatto.

Le sostanze dell'ambiente cambiano naturalmente sia strutturalmente che chimicamente. Alcuni solidi si dissolvono e le loro superfici cessano di esistere; le foglie cadono e le piante si decompongono ma la gran parte delle sostanze non cambiano né strutturalmente né chimicamente. E' soprattutto per questo che l'ambiente è persistente.

Per descrivere l'ambiente abbiamo utilizzato la triade di *mezzo, sostanze, superfici*, che consentono sia la persistenza che il cambiamento. Il mezzo è separato dalle sostanze dell'ambiente dalle superfici. Fino a quando le sostanze persistono le superfici persistono.

Tutte le superfici hanno una certa *disposizione* ed anche la *disposizione* tende a persistere ...

Ci sono superfici, quali quelle dei corpi incandescenti che emettono luce, o superfici piatte di sostanze trasparenti quali il vetro che trasmettono la luce con rifrazione, o superfici piatte lucide come gli specchi che riflettono la luce, o superfici che assorbono la luce.

Perché nella triade mezzo, sostanze, superfici le superfici sono così importanti? La superficie è il luogo in cui la maggior parte delle azioni si svolge. La superficie, non l'interno di una sostanza, è il luogo in cui la luce è riflessa o assorbita. La superficie non l'interno è ciò che l'animale tocca. La superficie è dove avvengono le variazioni chimiche.

Sarebbe opportuno formulare quelle che potrebbero essere chiamate *leggi ecologiche delle superfici*; propongo le seguenti senza ambizioni di completezza:

1. Tutte le sostanze che persistono hanno superfici, e tutte le superfici hanno una *disposizione*.

2. Ogni superficie ha una resistenza alla deformazione, in dipendenza dalla viscosità della sostanza.

3. Ogni superficie ha una resistenza alla disintegrazione, in dipendenza dalla coesione della sostanza.

4. Ogni superficie ha una tessitura (texture) caratteristica, in dipendenza dalla composizione della sostanza; generalmente ha sia una *tessitura di disposizione* sia una *tessitura di pigmento*.

5. Ogni superficie ha una forma caratteristica, una configurazione di larga scala.

6. Una superficie può essere fortemente o debolmente illuminata, in luce o in ombra.

7. Una superficie illuminata può assorbire molta o poca della luce che cade su di essa.

8. Una superficie ha caratteristiche di riflessione che dipendono dalla sostanza (nessuna sostanza ad esclusione del mezzo trasmette perfettamente la luce ed è pertanto trasparente).

9. Una superficie ha una distribuzione caratteristica di rapporti di riflessione per diverse lunghezze d'onda della luce in dipendenza dalla sostanza. Questa proprietà è ciò che determina il colore.

Consideriamo la legge 4., essa afferma che ogni superficie ha una tessitura caratteristica; la tessitura può essere pensata come la *struttura della superficie* che è distinta dalla struttura della sostanza sottostante la superficie. Naturalmente stiamo parlando di struttura relativamente fine dell'ambiente, le dimensioni sono dell'ordine di centimetri e millimetri.

Le superfici delle rocce, del suolo arato, dell'erba sono aggregati di differenti unità: cristalli, ammassi, e steli di erba rispettivamente, ma queste unità sono nidificate in unità più grandi.

La tessitura di una superficie origina da due fattori principali: primo, le sostanze naturali sono raramente omogeneee, sono più spesso aggregati di differenti sostanze omogenee; e secondo, esse sono raramente amorfe ma piuttosto aggregati di cristalli e pezzi della stessa sostanza. Perciò, le superfici delle sostanze naturali non sono mai né omogenee né amorfe ma hanno sia una tessitura chimica sia una tessitura fisica; sono generalmente sia conglomerate che corrugate; parliamo pertanto di una tessitura di pigmento e di una tessitura di disposizione. Ciò dice che superfici perfettamente omogenee e perfettamente liscie costituiscono un astratto caso limite. Gli specchi sono rari in natura.

Quando le unità chimiche e geometriche di una superficie sono relativamente piccole si ha una tessitura fine; quando sono relativamente larghe si ha una tessitura grossolana. Se le unità sono sufficientemente distinte da essere contate, la densità della tessitura può essere misurata come numero di unità in una arbitraria unità di area. Ma ciò è molto spesso difficile da fare perché le unità di tessitura sono generalmente nidificate le une nelle altre a differenti livelli di grandezza.

Le rocce, il suolo, l'humus, il fango, la sabbia, il ghiaccio, la neve, hanno diverse tessiture; le bacche, le foglie, i frutti di un albero hanno diverse tessiture; le superfici degli animali sono "tessute" in modo diverso dal pelo, dalle piume, dalla pelle. Le superfici con cui gli uomini primitivi fabbricavano i loro utensili hanno diverse tessiture: il legno, l'osso, le fibre. Le superfici degli ambienti artificiali - mattoni, plastica, carta, pietre - hanno diverse tessiture.

Le tessiture in ogni caso specificano di che sostanza si tratta, perciò le tessiture sono qualcosa di molto importante.

Consideriamo ora la struttura di una superficie su di una scala che è relativamente larga. La disposizione di una superficie può spesso essere analizzata in facce; facce rivolte oppure no ad una sorgente di illuminazione o ad un punto di osservazione.

Una superficie rivolta nella direzione di illuminazione prevalente sarà più illuminata di una superficie rivolta in una direzione diversa (la luce incidente non è mai unidirezionale come sarebbe nel vuoto, è più o meno omnidirezionale; spesso comunque si individua una direzione prevalente). Pertanto diverse facce adiacenti nell'ambiente saranno diversamente illuminate ad ogni data ora del giorno.

Riassumendo: le sostanze dell'ambiente devono essere distinte; un modo potente per distinguerle è vedere le loro superfici.

Una superficie ha proprietà caratteristiche che possono persistere o modificarsi: la sua disposizione, la sua tessitura, la proprietà di essere illuminata o in ombra, la proprietà di riflettere una certa frazione della luce che cade su di essa.

Consideriamo la differenza tra i termini usati per descrivere ciò che ho chiamato la disposizione di un habitat e i termini usati in geometria. *Superfici* e *mezzo* sono termini ecologici; *piani* e *spazio* sono i più vicini termini geometrici ad essi equivalenti, ma osserviamo le differenze. I piani sono privi di colore; le superfici sono colorate. I piani sono fantasmi trasparenti; le superfici sono generalmente opache e sostanziali. La intersezione di due piani, una linea, non è la stessa cosa della giunzione di due superfici piatte, un bordo. *Una superficie può essere vista, un piano può solo essere visualizzato.* Una superficie ha la proprietà di essere situata di fronte ad una sorgente di illuminazione o ad un punto di osservazione; un piano non ha queste proprietà.

L'ambiente di ogni animale contiene sostanze, superfici con la loro disposizione, oggetti, luoghi, eventi e altri animali. Questi differiscono dagli oggetti inanimati per molti aspetti ma soprattutto per il fatto che essi si muovono spontaneamente. I loro movimenti sono sempre *deformazioni della superficie.* Inoltre lo stile del movimento, il modo di deformazione è unico per ogni animale. Gli animali sono per la percezione gli oggetti più complessi che un ambiente può presentare ad un osservatore. Un altro animale può essere preda o predatore, potenziale compagno o rivale, adulto o giovane. Inoltre può essere temporaneamente sveglio o addormentato, ricettivo o non ricettivo, affamato oppure no. Ciò che un animale offre è specificato dalle sue caratteristiche permanenti e dal suo stato temporaneo ed esso può offrire di nutrire o essere nutrito, di accoppiarsi o combattere. Ciò che gli animali offrono ad un osservatore non è solo il comportamento ma le interazioni sociali.

L'informazione per la percezione visiva

Avendo descritto l'ambiente, descriverò ora quale informazione è disponibile ad un osservatore per percepire l'ambiente. Per la percezione visiva, l'informazione è ovviamente la luce. Cerchiamo di distinguere la luce come energia fisica, la luce come stimolo per la visione e la luce come informazione per la percezione.

Alcuni corpi materiali emettono luce, altri no ... Gli oggetti non luminosi riflettono parte della luce che cade su di essi da una sorgente ... Al livello ecologico delle sostanze, superfici, e mezzo, siamo interessati solo ad alcune delle leggi fisiche della luce; principalmente la diffusione, la riflessione, l'assorbimento.

Alla luce del giorno, parte della luce del sole raggiunge la terra in raggi paralleli ma un'altra parte è dispersa nella trasmissione attraverso l'atmosfera che non è mai perfettamente trasparente. La luce è dispersa (diffusa) ancora di più quando colpisce superfici tessute. (Si può parlare di riflessione diffusa che non deve essere confusa con la riflessione degli specchi che è governata dalla legge di angoli uguali di incidenza e riflessione).

E giungiamo ora al punto più importante: *la luce ambientale ha una struttura.*

Solo in quanto la luce ambientale ha una struttura essa è in grado di specificare l'ambiente. Io intendo con ciò che la luce ad un punto di osservazione deve essere diversa nelle diverse direzioni (o devono esserci differenze nelle diverse direzioni) per poter contenere informazione. Le differenze sono principalmente differenze di intensità. Il termine che userò per descrivere la luce ambientale con struttura è *vettore ottico ambientale* (ambient optic array) ...

Quale sarebbe il caso limite di luce ambientale senza struttura? Si avrebbe se l'aria fosse piena di una nebbia tanto fitta che la luce non potrebbe riverberare tra le superfici ma solo tra le goccioline del mezzo ... Ad ogni punto di osservazione ci sarebbe radiazione ma senza differenza nelle diverse direzioni; *senza transizioni o gradazioni di intensità non ci sarabbe struttura*, non ci sarebbe vettore ... Nel caso di luce ambientale non strutturata, un ambiente non è specificato e nessuna informazione circa l'ambiente è disponibile ... In questo rispetto la luce ambientale non sarebbe diversa dal buio ambientale ... La stimolazione dei recettori della retina non può essere vista. La stimolazione è una condizione necessaria per vedere ma non è sufficiente. Nello stimolo deve esserci informazione per il sistema percettivo non solo stimolazione per i recettori ... La visione dipende dalla luce ma ciò non significa necessariamente che noi vediamo la luce o abbiamo sensazioni di luce ... E' stato osservato che animali privati per un certo periodo della *struttura dello stimolo luminoso* ma non dello *stimolo luminoso* restavano parzialmente ciechi. Crudamente parlando, essi non potevano usare il loro sistema propriamente. Non si aveva degenerazione anatomica dei fotorecettori, come si ha negli animali tenuti al buio, ma gli aggiustamenti esploratori del sistema visivo non si erano sviluppati normalmente.

Riassumendo: la luce ambientale ha struttura, essa rende disponibile l'informazione circa le superfici riflettenti. Se la luce ambientale fossa non strutturata o indifferenziata essa non fornirebbe alcuna informazione circa l'ambiente, ma stimolerebbe comunque i fotorecettori dell'occhio. Pertanto c'è una chiara distinzione tra informazione dello stimolo e stimolazione.

Esperimento con occhiali traslucidi

L'eliminazione, per mezzo di occhiali traslucidi, della struttura della luce che colpisce l'occhio è un esperimento che è stato ripetuto molte volte. L'osservatore è cieco, non alla luce, perché i fotorecettori sono stimolati, ma all'ambiente perché il sistema visivo non riesce ad essere attivo, i suoi aggiustamenti sono frustrati. L'osservatore non può guardare o guardare intorno.

Esperimento con una lastra di vetro

E' ben noto che una lastra pulita di vetro che non riflette nulla all'occhio di un osservatore è, come noi diciamo, invisibile. Questo fatto è molto interessante poiché significa che una persona percepisce aria quando invece c'è una superficie. E' capitato di vedere persone che cercano di attraversare una porta di vetro. Una lastra di vetro perfettamente pulita trasmette sia la luce considerata come energia che la luce considerata come informazione. Una lastra di vetro brinata o zigrinata non trasmette informazione. Attraverso la lastra pulita si può vedere, attraverso quella zigrinata no. L'ultima può essere vista, la prima no.

Un'impercettibile lastra di vetro può essere resa via via pi percepibile facendo cadere polvere su di essa. In questo caso intermedio, la lastra trasmette sia il vettore generato dall'ambiente dietro il vetro sia il vettore generato dal vetro stesso. Vediamo le altre superfici attraverso il vetro.

Eventi e informazione per percepire gli eventi

Tre tipi di eventi si possono distinguere: cambiamenti nella disposizione delle superfici, cambiamenti nella tessitura o nel colore delle superfici, cambiamenti nella esistenza di una superficie. Esempi del primo tipo sono traslazioni e rotazioni di un oggetto, collisioni, deformazioni. Esempi del secondo tipo sono le alterazioni, spesso prive di nome ma ricche di significato, delle superfici di piante (es. maturazione di frutti) e di animali. Esempi del terzo tipo sono le transizioni di evaporazione, dissipazione, decadimento, dissolvimento.

L'informazione ottica per distinguere i vari eventi può essere solo quella che deriva dai diversi disturbi nella struttura del vettore ottico ambientale.

L'informazione ottica per la percezione di sé

L'informazione ottica che specifica il sé, intendendo con ciò la testa, il corpo, le braccia, le mani, accompagna l'informazione ottica che specifica l'ambiente. Le due sorgenti di informazione coesistono ... La percezione del sé accompagna la percezione dell'ambiente. Uno percepisce l'ambiente e al tempo stesso percepisce se stesso. I limiti del campo visivo occludono l'ambiente rimanente e, quando la testa ruota, le occlusioni si modificano rivelando ciò che era nascosto e nascondendo ciò che era visibile. Anche con la locomozione accade ciò ... L'ambiente stazionario è specificato insieme al movimento dell'osservatore.

La teoria delle offerte

Il mezzo, le sostanze, le superfici, gli oggetti, i luoghi, gli altri animali hanno offerte per un dato animale. Essi *offrono* benefici o ingiurie, vita o morte. Questo è ciò che deve essere percepito.

Le possibilità di un ambiente e il modo in cui si svolge la vita di un animale sono inseparabilmente legati. L'ambiente vincola ciò che l'animale può fare e il concetto di nicchia ecologica riflette queso fatto. Entro certi limiti l'uomo può alterare le offerte dell'ambiente.

Le offerte sono proprietà che hanno senso rispetto ad un osservatore.

La ipotesi che la informazione nella luce ambientale specifica le offerte è il culmine dell' ottica ecologica.

[J.J.Gibson -The ecological approach to visual perception- Lawrence Erlbaum Associates, Publishers 1986].

Abbiamo visto che esiste una informazione nella struttura della intensità luminosa; la luce è strutturata dall'ambiente.

Compito della visione è pertanto estrarre questa informazione che è informazione sulle superfici, sulla loro disposizione, sulla loro tessitura, sul loro colore.

Capire il processo visivo completamente significa, "capire come descrizioni del mondo possono essere ottenute, in modo efficiente e corretto, da immagini di esso" (Marr pag 99).

La proposta di Marr è quella di analizzare il problema della percezione visiva come un compito di elaborazione dell'informazione.

1.3 Da D. Marr - Vision -

Comprendere sistemi complessi che elaborano informazione

Quasi mai un qualsiasi sistema complesso può essere compreso attraverso la semplice estrapolazione delle proprietà dei suoi componenti elementari. Considerate, per esempio, un gas in una bottiglia. Una descrizione degli effetti termodinamici - temperatura, pressione, densità, e le relazioni tra questi fattori - non è formulata usando un grande insieme di equazioni, una per ciascuna delle particelle interessate. Questi effetti sono descritti al loro livello, che è quello di una enorme collezione di particelle; lo sforzo è quello di mostrare che in linea di principio le descrizioni, macroscopiche e microscopiche sono consistenti le une con le altre. Se uno spera di raggiungere la completa comprensione di un sistema così complesso come un sistema nervoso, un embrione che si sviluppa, un insieme di vie metaboliche, una bottiglia di gas od anche un grande programma per computer, allora deve essere preparato a contemplare *differenti tipi di spiegazione a diversi livelli di descrizione* che sono legati, almeno in linea di principio, in un tutto coerente, anche se connettere i livelli in dettaglio è impraticabile. Per il caso specifico di un sistema che risolve un complesso problema di elaborazione dell'informazione ci sono inoltre i due fili gemelli di processo e rappresentazione.

Rappresentazione e descrizione

Una *rappresentazione* è un sistema formale per rendere esplicite certe entità o tipi di informazione, insieme con la specificazione di come il sistema fa ciò. Io chiamerò il risultato di usare una rappresentazione per descrivere una data entità una *descrizione* dell'entità in quella rappresentazione.

Per esempio i sistemi numerici arabico, romano e binario sono tutti sistemi formali per rappresentare i numeri. La rappresentazione arabica consiste di stringhe di simboli presi dall'insieme $(0, 1, 2, 3, 4, 5, 6, 7, 8, 9)$, e la regola per costruire la descrizione di un particolare intero n è quella di decomporre n in una somma di multipli di potenze di 10 e unire questi multipli in una stringa con le potenze maggiori a sinistra e le potenze minori a destra. La descrizione di un numero nel sistema binario rende esplicita la decomposizione del numero in potenze di 2 ...

Ogni particolare rappresentazione rende certe informazioni esplicite alle spese di informazioni che vengono celate e che può essere molto difficile reperire. Il modo in cui l'informazione è rappresentata può influenzare moltissimo la facilità con cui cose differenti possono essere fatte. Questo è già evidente dall'esempio dei numeri: è facile sommare, fare sottrazioni, moltiplicazioni col sistema arabico o binario ma non è per nulla facile farlo con i numerali romani ... Un problema analogo si ha nella realizzazione dei computers. L'elettronica è più adeguata per sistemi binari.

Processi

Il termine di processo è molto ampio ... Voglio restringere l'attenzione al significato che si utilizza con le macchine che eseguono compiti di elaborazione dell'informazione. Esaminiamo in profondità le nozioni che stanno dietro un semplice dispositivo di cassa di un supermercato.

Ci sono diversi livelli a cui è necessario comprendere tale dispositivo e sembra molto utile pensare a tre di essi. Il più astratto è il livello del *che cosa* fa la macchina e *perché*. Ciò che fa è aritmetica, perciò il nostro primo compito è padroneggiare la teoria dell'addizione. L'addizione è una funzione che mappa una coppia di numeri in un numero; l'addizione ha alcune proprietà astratte (commutativa, associativa, esistenza dello 0, esistenza dell'inverso). Queste proprietà sono parte della teoria fondamentale dell'addizione. Esse sono vere indipendentemente dal modo in cui sono scritti i numeri. Perciò parte di questo primo livello è qualcosa che può essere caratterizzato con che cosa è computato.

L'altra metà di questo livello di spiegazione è perché un registratore di cassa esegue l'addizione e non la moltiplicazione quando combina i prezzi delle merci per arrivare ad un prezzo finale. La ragione è che le regole che noi intuitivamente sentiamo appropriate per combinare i prezzi individuali di fatto definiscono l'operazione matematica di addizione. Queste possono essere formulate come vincoli nel seguente modo:

1. Se non compri nulla non devi pagare nulla; e comprare niente e qualcosa dovrebbe costare come comprare qualcosa (regola dello 0).

2. L'ordine in cui le merci sono presentate al cassiere non dovrebbe modificare il totale (commutatività).

3. Porre le merci in pile distinte e pagare per ogni pila separatamente non dovrebbe modificare il totale (associatività).

4. Se compri una merce e poi la rendi chiedendo i soldi, la spesa totale è 0 (legge dell'inverso).

Ciò è quello che io chiamo *teoria computazionale del registratore di cassa*. Le sue caratteristiche importanti sono che (1) essa contiene argomenti separati per ciò che deve essere computato e perché e (2) le risultanti operazioni sono definite unicamente dai vincoli che devono essere soddisfatti. Nella teoria dei processi visivi il compito sottostante è di derivare proprietà del mondo da immagini di esso; il compito è quello di isolare vincoli che siano abbastanza potenti da consentire una definizione del processo; questo è il tema centrale della nostra indagine.

Affinché un processo possa essere realizzato è necessario scegliere una rappresentazione per le entità che il processo manipola. Il secondo livello di analisi di un processo coinvolge pertanto la scelta di due cose (1) una rappresentazione per l'input e l'output del processo e (2) un algoritmo che realizza le trasformazioni ...

Il primo livello specifica che *cosa* e *perché*, il secondo livello specifica *come* ...

Il terzo livello è quello del dispositivo in cui il processo è realizzato fisicamente.

Naturalmente c'è sempre un'ampia scelta per le rappresentazioni; inoltre la scelta degli algoritmi dipende in modo cruciale dalle rappresentazioni. Infine per una data rappresentazione ci sono spesso molti possibili algoritmi.

I tre livelli

Possiamo riassumere la discussione fatta nella maniera rappresentata in figura, che illustra i differenti livelli a cui un dispositivo che elabora informazione deve essere capito prima di poter dire che lo si è capito completamente.

Teoria computazionale	**Rappresentazione e algoritmo**	**Realizzazione hardware**
Q ual è lo scopo della computazione, perché essa è appropriata, e qual'è la logica della strategia con cui raggiungere lo scopo?	Come può essere implementata (realizzata) questa teoria computazionale? In particolare, qual'è la rappresentazione per l'ingresso e l'uscita, e qual è l'algoritmo per la trasformazione?	Come possono essere realizzati fisicamente l'algoritmo e la rappresentazione?

Ad un estremo, il livello superiore, c' è la teoria computazionale astratta del dispositivo, in cui il comportamento del dispositivo è caratterizzato come funzione da un tipo di informazione ad un altro, le proprietà astratte di questa funzione sono definite con precisione e la sua adeguatezza per il compito di interesse sono dimostrate. Al centro c'è la scelta della rappresentazione per l'ingresso e l'uscita e gli algoritmi che devono essere utilizzati per trasformare l'uno nell'altro. All'altro estremo ci sono i dettagli di come gli algoritmi e le rappresentazioni sono realizzati fisicamente, la architettura dettagliata del computer per intenderci. Questi tre livelli sono accoppiati solo in modo lasco. La scelta di un algoritmo è influenzata per esempio, da ciò che deve fare e dall'hardware su cui deve essere eseguito. Ma c'è una vasta scelta disponibile ad ogni livello, e la spiegazione di ogni livello coinvolge argomenti che sono parecchio indipendenti dagli altri due.

Ognuno dei tre livelli di descrizione avrà il suo posto nell'eventuale comprensione dell'elaborazione dell'informazione percettiva, e naturalmente essi sono logicamente e causalmente relati. Ma un punto importante da notare è che poiché i tre livelli sono interconnessi solo lascamente, alcuni fenomeni possono essere spiegati solo ad uno o due di essi.

Importanza della teoria computazionale

Sebbene gli algoritmi e i meccanismi siano empiricamente più accessibili, è il livello più alto che è criticamente importante da un punto di vista di elaborazione dell'informazione. La ragione di ciò è che la natura delle computazioni che sottostanno alla percezione dipende più dal problema che occorre risolvere che dal particolare hardware in cui le soluzioni sono implementate. Per dire le cose in altro modo possiamo dire che un algoritmo si capisce più facilmente comprendendo la natura del problema che deve essere risolto piuttosto che esaminando i meccanismi (e l'hardware) in cui è realizzato.

Per la visione un quadro in termini di rappresentazioni

La visione è un processo che produce da immagini del mondo esterno una descrizione che è utile a chi guarda e non è zeppa di informazione irrilevante. Abbiamo già visto che un processo può essere pensato come una funzione da una rappresentazione ad un'altra, e nel caso della visione umana la rappresentazione iniziale consiste senza dubbio di un vettore di intensità di valori misurati dai fotorecettori della retina.

E' assolutamente adeguato pensare ad una immagine come ad una rappresentazione; gli elementi che sono espliciti nell'immagine sono i valori dell' intensità ad ogni punto del vettore, che convenientemente denotiamo con $I(x, y)$ alla coordinata (x, y) ... Ci riferiremo ad ogni recettore come ad un elemento dell'immagine (pixel) e a tutto il vettore come immagine.

Ma qual'è l'uscita del processo visivo? Abbiamo già convenuto che essa deve consistere di una utile descrizione del mondo, ma questa richiesta è alquanto nebulosa. Possiamo fare meglio? Bene, è perfettamente vero che, al contrario dell'ingresso il risultato della visione è difficile da discernere, e un importante aspetto di questo nuovo approccio è che esso fa proposte concrete su quale può essere la fine ...

Sceglierò come uscita una descrizione delle *forme* e della *posizione delle cose* a partire dalle immagini. Naturalmente ciò non è tutto quanto la visione può fare, mediante la visione noi abbiamo informazioni circa l'illuminazione e le proprietà di riflessione delle superfici che costruiscono le forme, la loro brillanza, il loro colore e tessitura, il loro movimento.

Il desiderabile attraverso il possibile

Per quanto desiderabile possa essere avere un sistema di visione che produce, a partire da una immagine, una descrizione della forma completamente invariante è quasi certamente impossibile ottenere ciò in un solo passo. Noi siamo pertanto giunti all'idea di sequenza di rappresentazioni, partendo da descrizioni che possono essere ottenute direttamente da un'immagine ma che sono accuratamente disegnate per facilitare il successivo recupero di proprietà fisiche gradualmente più obiettive circa la forma di un oggetto. Il passo più importante verso questo obiettivo è la descrizione della geometria delle superfici visibili, poiché l'informazione codificata nelle immagini è dovuta a proprietà locali delle superfici. L'obiettivo di molte computazioni di *early vision* è quello di estrarre questa informazione.

Comunque, questa descrizione delle superfici visibili si rivela inadatta per compiti di riconoscimento; essa dipende criticamente dal punto di vista. Il passo finale consiste pertanto nel trasformare la descrizione di una superficie in un sistema di riferimento che pone al centro l'osservatore, in una rappresentazione della forma tridimensionale di un oggetto e delle sue relazioni spaziali che non dipendano dalla direzione da cui l'oggetto è visto. Questa descrizione finale è realizzata in un sistema di riferimento che ha al centro l'oggetto.

Nome	**Obiettivo**	**Primitive**
Immagine(i)	Rappresenta l'intensità	Valori di intensità ad ogni punto dell'immagine
Primal sketch	Rende esplicite importanti informazioni circa l'immagine a due dimensioni, in prima istanza le variazioni di intensità e la loro distribuzione geometrica e organizzazione.	Zero-crossing Blobs ...
$2^1/_2$-D sketch	Rende esplicito l'orientamento e la profondità delle superfici visibili in un sistema di coordinate definito rispetto a chi vede.	Local surface orientation
3-D model representation	descrive le forme e la loro organizzazione spaziale in un sistema di coordinate indipendente dall'osservatore	...

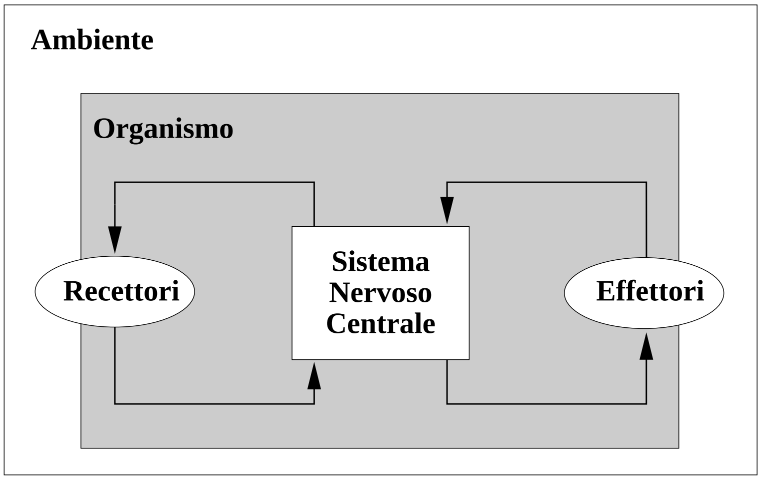
[David Marr -Vision- Freeman and Company 1982.]

1.4 Qualche idea sul Sistema Nervoso Centrale

Che cosa e', come e' strutturalmente organizzato

Il *Sistema nervoso centrale* (SNC), la cui struttura e complessità variano nei diversi gruppi di organismi animali, può essere globalmente definito come il sistema che controlla le interazioni dell'organismo con l'ambiente (il comportamento). E' un complesso elaboratore di informazione che attua una computazione distribuita orientata all'azione sul mondo esterno. Tale azione si manifesta attraverso il movimento (anche il linguaggio a ben guardare può trovare espressione solo mediante il movimento: dell'apparato vocale, delle mani).

Il SNC consente all'organismo di percepire cambiamenti dell'ambiente; esso elabora l'informazione che riceve da sistemi di *recettori* controlla inoltre l'attività degli *effettori*: il sistema muscolare. Può inoltre imparare dalle passate interazioni a modificare il comportamento dell'organismo rendendolo in generale piu' adatto a sopravvivere in un dato ambiente. Schematicamente e astraendo da considerazioni architetturali possiamo così rappresentare il SNC:



Dopo aver parlato del SNC in modo così astratto cerchiamo di dire qualcosa su come è strutturato. Purtroppo le enormi differenze di complessità e struttura che si riscontrano se si osserva il SNC di animali diversi rende possibile questa sola descrizione generale: è una rete di cellule nervose connessa ad apparati di recettori e di effettori. Per dire qualcosa di più si può considerare il SNC dei vertebrati (pesci, anfibi, rettili, uccelli, mammiferi) il cui studio comparato mostra una graduale evoluzione se si passa dagli organismi più semplici (pesci) al più complesso (l'uomo). Le parti più antiche del SNC, quali per esempio il midollo spinale, presentano caratteri quasi immutati, il cervello invece (una parte di SNC situata nella scatola cranica, di cui gli emisferi cerebrali sviluppati soprattutto nei mammiferi) presenta enormi differenze.

L'omogeneità di struttura e organizzazione nelle parti simili fanno sì che gran parte delle informazioni che riteniamo valide anche per il SNC umano derivino in realtà da esperimenti sugli altri mammiferi (gatti, scimmie...). Per dare qualche dato quantitativo consideriamo il SNC dell'uomo. E' costituito da 10^{10} - 10^{11} neuroni ognuno dei quali comunica con migliaia o decine di migliaia di altre cellule.

Questa complessa rete è connessa sia a complessi sistemi di recettori (pensate all'occhio, all'orecchio) per mezzo dei quali riceve dall'ambiente diversi "tipi" di informazione (informazione visiva, uditiva, tattile, olfattiva, gustativa e in generale estero e propriocettiva) sia ad un complesso apparato motore (piu' di 500 muscoli diversi).

I diversi sitemi di recettori sono connessi mediante distinti fasci di nervi (fibre nervose) a parti diverse della rete in cui è pertanto possibile localizzare delle funzioni.

Che cosa conosciamo e che cosa non conosciamo

Le conoscenze che abbiamo devono essere ascritte a diversi livelli di descrizione del sistema, livelli che sono tra di loro ancora separati.

1. livello psicologico: descrive il comportamento di soggetti in diverse condizioni ambientali

2. livello fisiologico (neurofisiologico): studia come è strutturato e come funziona il sistema, è molto frammentato; viene per esempio analizzata:

   - la morfologia di singoli elementi;

   - la morfologia delle connessioni tra gruppi di elementi e tra singoli elementi;

   - la fisiologia di singole cellule: neuroni centrali, cellule sensoriali, cellule muscolari;

   - la fisiologia delle interazioni tra cellule (sinapsi)
   e andando ad un livello subcellulare:

   - la biofisica delle membrane;

   - la biochimica delle interazioni.

Il problema è sistemare questo insieme di conoscenze in un quadro organico. Immaginate che sia possibile dividere il SNC in moduli che svolgono una particolare funzione (per esempio il sistema visivo): si comincia appena a capire come possono funzionare alcuni di tali sottosistemi. In generale mancano modelli sufficientemente complessi da comporre in modo organico ed, al giusto livello, un numero enorme di dati sperimentali.

Mancano modelli soddisfacenti che compongano il comportamento di singoli neuroni nel comportamento di reti che svolgano funzioni utili all'organismo.

Per dare un'idea della complessità del problema poniamoci una domanda: Come fa un uomo in grado di muoversi liberamente a imparare a raggiungere un determinato oggetto? Nonostante l'apparente semplicità della domanda una risposta completa richiede la comprensione di 3 tipi di problemi problemi quali:

- problemi di percezione (percepire forme, colori...);

- problemi di elaborazione dell'informazione cognitiva (utilizzare le percezioni per riconoscere oggetti, ambienti);

- problemi di controllo motorio.

Nonostante ciò per il sistema visivo e per l'apparato motore cominciano ad esserci tentativi soddisfacenti. Il contributo della *computer science* è stato fondamentale: la necessità di costruire sistemi di visione e robot ha portato a formulare teorie sulla visione e sull'organizzazione del movimento.

1.4.1 Il neurone

La cellula nervosa, o *neurone*, è la componente elementare fondamentale del sistema nervoso. Ci sono circa 10^{10} - 10^{11} neuroni nel sistema nervoso umano.

Nonostante la diversità (esistono molti tipi di neuroni), alcune caratteristiche condivise dai neuroni di organismi anche enormemente distanti nella scala evolutiva, consentono di parlare del neurone come entità astratta.

Il neurone è una cellula specializzata per ricevere, elaborare e trasmettere segnali ad altre cellule. La sua morfologia e struttura sono asservite alla realizzazione di queste funzioni; la descrizione di essenziali elementi morfologici e fisiologici fornirà uno schema del neurone.

I neuroni presentano una grande varietà di forme; nonostante ciò, nella gran parte delle cellule nervose possono essere distinte tre regioni:

1. il corpo cellulare o soma;
2. i dendriti;
3. l'assone o fibra nervosa.

Il *corpo cellulare* contiene il nucleo e gli organuli essenziali al metabolismo cellulare; può avere forma e dimensioni variabili (da circa 10μ a 100μ di diametro).

I *dendriti* sono sottili prolungamenti di forma tubulare che, dipartendosi dal corpo cellulare in varie direzioni, si suddividono più volte formando intorno al corpo cellulare una struttura che può risultare così ramificata da prendere il nome di *albero dendritico.* La estensione dell'albero dendritico varia nei diversi neuroni.

L'*assone* è un prolungamento che percorre generalmente notevoli distanze (anche fino a metri) e si ramifica nella zona di terminazione. Può emettere rami collaterali prima di giungere a destinazione.

A questa suddivisione morfologica corrisponde una distinzione funzionale che, in prima approssimazione, possiamo così descrivere: i dendriti e il corpo cellulare sono strutture atte a *ricevere* ed *elaborare* segnali che provengono da altri neuroni; l'assone *trasmette* segnali ad altri neuroni. I punti di comunicazione tra due neuroni distinti prendono il nome di sinapsi o giunzioni sinaptiche.

Una sinapsi è una struttura specializzata in cui la comunicazione avviene in una sola direzione. Per questa ragione le due entità comunicanti vengono distinte in un *elemento presinaptico*, che invia il segnale, e in un *elemento postsinaptico* che riceve il segnale.

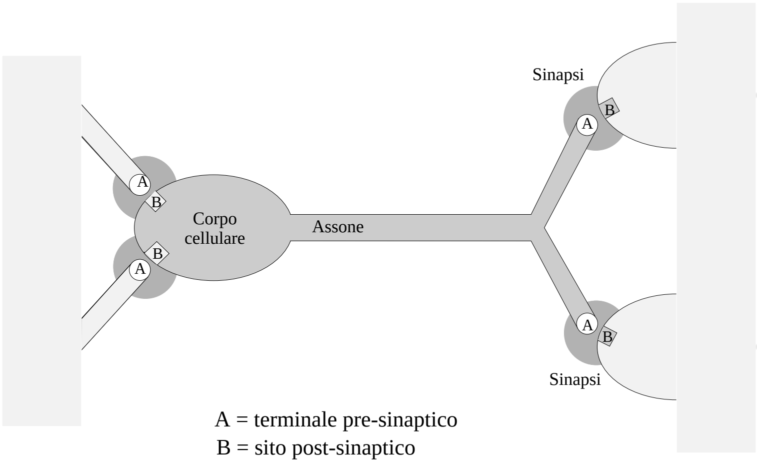
L'elemento presinaptico è generalmente costituito dalla terminazione di un assone; l'elemento postsinaptico è una area della superficie di un dendrite o del corpo cellulare del neurone ricevente.

Il terminale presinaptico e il sito postsinaptico sono separati da uno spazio sinaptico di poche decine di Angstrom (10^{-8} cm).

Il neurone è funzionalmente polarizzato: i siti di ricezione (sinapsi in ingresso) sono distinti da quelli di trasmissione (sinapsi in uscita).

Il numero di sinapsi in ingresso può variare da parecchie centinaia a decine di migliaia (cellule di Purkinje nel cervelletto). Il numero di sinapsi in uscita può variare da poche decine a decine di migliaia. Ogni neurone ha quindi un elevatissimo ventaglio di entrate (*fan-in*) e un elevatissimo ventaglio di uscite (*fan-out*).

Possiamo rappresentare graficamente lo schema descritto che, sebbene non abbia una validità universale, è realizzato nella maggior parte dei neuroni sino ad ora caratterizzati.



A = terminale pre-sinaptico
B = sito post-sinaptico

Esistono eccezioni a questo schema generale: in alcuni neuroni (quali ad esempio i fotorecettori, le cellule bipolari, le cellule amacrine della retina) la distinzione dell'assone tra le ramificazioni che si dipartono dal corpo cellulare non è ovvia. Le sinapsi si formano fra dendriti di neuroni diversi; i dendriti si comportano come strutture che ricevono e trasmettono segnali. In essi si sovrappongono due funzioni che nello schema presentato sono nettamente distinte.

La membrana del neurone

Le funzioni che abbiamo definito di ricezione, elaborazione, trasmissione di segnali, sono rese possibili dalle peculiari proprietà della *membrana neuronale*. La struttura della membrana, elemento di separazione del neurone dall'ambiente, è differenziata nelle parti del neurone che abbiamo indicato come corpo cellulare, dendriti e assone.

Una breve descrizione delle caratteristiche della membrana servirà a illustrare quali tipi di segnali vengono ricevuti, elaborati, e trasmessi dai neuroni, e quali tipi di operazioni i neuroni sono in grado di effettuare.

La membrana del neurone è dotata della capacità di mantenere una differenza di concentrazione di alcune specie ioniche (Na^+, K^+, Cl^-) tra l'esterno e l'interno del neurone. Ciò a causa di una permeabilità selettiva ad esse e di un trasferimento attivo (a spese di energia metabolica) di alcune di esse (Na^+, K^+).

La diversa concentrazione di ioni ai lati della membrana genera una differenza di potenziale che prende il nome di *potenziale di membrana*.

In assenza di interazioni del neurone con altri neuroni, il valore del potenziale di membrana è mantenuto costante: l'interno del neurone ha potenziale minore dell'esterno; la differenza di potenziale varia nei diversi neuroni da -60 a -90 mV. Questa differenza di potenziale prende il nome di *potenziale di riposo*.

La situazione di equilibrio (stabilità) che caratterizza la membrana di un neurone in condizioni di riposo può essere modificata da due fattori:

arrivo di un segnale ad un sito postsinaptico;

generazione di un segnale, il *potenziale d'azione*, ad opera della membrana stessa.

Arrivo di un segnale ad un sito postsinaptico

La trasmissione di un segnale attraverso una sinapsi si attua tramite lo spostamento di molecole che, liberate a pacchetti dal terminale presinaptico, attraversano lo spazio sinaptico e si legano a recettori specializzati situati sulla superficie della membrana postsinaptica. Le molecole che fungono da segnale, vengono denominate *mediatori* o *neurotrasmettitori*.

Il legame tra una molecola di mediatore e il recettore della membrana postsinaptica causa l'apertura di *canali* (canali controllati da recettori) nella membrana stessa.

Attraverso questi canali possono fluire gli ioni diversamente distribuiti ai lati della membrana. La corrente ionica che percorre i canali determina una variazione locale del potenziale di membrana. Tale variazione può consistere in un aumento del potenziale di membrana rispetto al valore di riposo (*iperpolarizzazione*) o in una sua diminuzione (*depolarizzazione*).

Per ragioni che diverranno tra poco evidenti, vengono definite *eccitatorie* le sinapsi che provocano una depolarizzazione, *inibitorie* quelle che provocano una iperpolarizzazione.

La variazione locale del potenziale di membrana (sia essa una depolarizzazione o una iperpolarizzazione) si propaga in modo passivo alle zone di membrana adiacenti al sito postsinaptico attivo: propagandosi passivamente, la variazione di potenziale si riduce sino ad estinguersi via via che ci si allontana dal sito di generazione.

L'ordine di grandezza dello spazio in cui si estingue il fenomeno è di qualche centinaio di micron (10^{-6}m). La durata della variazione locale del potenziale dovuta ad un singolo segnale è variabile; l'ordine di grandezza è generalmente pochi ms (10^{-3} secondi).

I siti postsinaptici sui dendriti e sul soma di un neurone possono essere moltissimi. È pertanto logico aspettarsi che le variazioni locali interagiscano. La più semplice modalità d'interazione (evidenziata già agli inizi degli anni 60) è la somma delle depolarizzazioni e iperpolarizzazioni locali. Vengono distinte due modalità di interazione: *sommazione spaziale* le differenze di potenziale generate simultaneamente in diversi siti attivi si sommano; *sommazione temporale* ad ogni sito postsinaptico l'effetto di un secondo segnale si somma all'effetto del precedente se questo non si è ancora estinto.

Occorre sottolineare che, nonostante si parli di somma degli effetti, la integrazione non è sempre un fenomeno lineare: la variazione locale del potenziale postsinaptico può essere diversa della somma dei potenziali generati separatamente.

Un tipo di interazione tra eventi postsinaptici altamente non lineare è stato evidenziato in dendriti di neuroni (cellule gangliari) della retina: sinapsi inibitorie possono "cancellare" l'effetto postsinaptico di sinapsi eccitatorie, gli autori parlano di "operazione di veto".

Questa interazione richiede come condizione la vicinanza delle sinapsi eccitatorie ed inibitorie che devono essere sullo stesso dendrite; inoltre le sinapsi inibitorie devono essere situate tra la sinapsi eccitatoria e il soma dove viene generato il potenziale d'azione.

Generazione di un segnale: il potenziale d'azione

Le variazioni locali del potenziale di membrana che interagiscono secondo le modalità citate vengono passivamente propagate ad una zona della membrana somatica, la parte iniziale dell'assone, che è in grado di esibire un comportamento peculiare.

Se il potenziale raggiunge o supera un valore *soglia* la membrana genera una rapida (1-1.5 ms) ed ampia (60-80mV) variazione di potenziale. Tale potenziale, denominato *potenziale d'azione* (o potenziale a punta o impulso nervoso) origina, come i potenziali postsinaptici, dal flusso di correnti ioniche attraverso la membrana. In questo caso però i canali attraverso cui fluiscono gli ioni sono controllati dal valore del potenziale di membrana (canali controllati dal potenziale); essi si aprono solo se il potenziale raggiunge il valore soglia.

Generato in un punto preciso (il punto a soglia minore situato all'inizio dell'assone), il potenziale d'azione si propaga a tutta la membrana somatica e nell'assone. La *propagazione è attiva*; l'impulso nervoso viene *rigenerato* in punti adiacenti della membrana, in questo modo percorre l'assone senza decremento dal sito di origine ai siti di destinazione (terminali assonici).

La velocità con cui un potenziale d'azione si propaga lungo un assone varia da pochi metri a circa 100 metri/secondo. E' interessante notare che nella gran parte degli assoni dei vertebrati la discontinuità con cui il potenziale di azione procede lungo la fibra è accentuata; i canali responsabili per la rigenerazione del potenziale sono concentrati in piccoli siti specializzati equidistanti. Tra un sito e l'altro, la membrana è avvolta da una guaina (guaina mielinica) che la isola dall'ambiente. La presenza della guaina mielinica aumenta la velocità di conduzione localizzando in precisi punti il fenomeno di generazione.

Un impulso nervoso che ha percorso l'assone, giungendo a depolarizzare una terminazione presinaptica scatena un processo di secrezione (liberazione) di una data quantità di pacchetti di molecole di mediatore. Sono stati identificati parecchi mediatori; essi sono diversi in diversi tipi di neuroni.

È infine necessaria una precisazione: la trasmissione sinaptica attraverso mediatori, caratteristica delle sinapsi, per questo motivo definite, di *tipo chimico*, è la modalità di trasmissione sinaptica più diffusa e studiata, ma non l'unica. E' stata infatti evidenziata, in particolari sinapsi per questo definite *elettriche*, un'altra modalità di trasmissione sinaptica: mediante correnti ioniche.

Scarica ripetitiva

Dopo aver generato un potenziale d'azione la membrana diventa *relativamente refrattaria* alla generazione di un potenziale di azione successivo; la refrattarietà non è assoluta: è semplicemente necessario uno stimolo di intensità maggiore per evocare una risposta nel neurone.

La "refrattarietà" è un fenomeno attivo, la membrana genera infatti correnti che contrastano le correnti depolarizzanti provocate da stimoli eccitatori. La durata di tale fenomeno varia nei diversi neuroni da decine di millisecondi a centinaia di millisecondi. Trascorso tale intervallo di tempo la membrana torna nella condizione di riposo. La refrattarietà è un meccanismo sviluppato dal neurone per regolare l'intervallo di tempo che deve trascorrere tra la generazione di potenziali di azione successivi.

Se il neurone viene eccitato con stimoli di intensità costante e tale da superare la refrattarietà relativa, risponde con sequenze di potenziali di azione la cui frequenza media è relata all'ampiezza dello stimolo. In tal modo il neurone opera una conversione analogico-digitale: l'ampiezza di un segnale in ingresso è codificata dalla frequenza di potenziali di azione successivi. La refrattarietà è il mezzo per attuare la codificazione.

Plasticità sinaptica

Si denota col termine *plasticità sinaptica* la possibilità di variare l'effetto postsinaptico provocato da un segnale giunto ad una terminazione presinaptica. La modificazione della forza delle connessioni sinaptiche viene oggi ritenuta lo strumento più importante messo in opera dal sistema nervoso centrale per apprendere e modificarsi.

Sono stati evidenziati alcuni meccanismi fisiologici in grado di provocare modificazioni sinaptiche: qui ci limitiamo a riportare in breve alcuni recenti risultati sperimentali.

Un fenomeno osservato in molte regioni del sistema nervoso centrale ma accuratamente analizzato solo in particolari preparati sperimentali, è denominato *potenziamento a lungo termine*. Esso è caratterizzato da un aumento stabile e relativamente duraturo della risposta postsinaptica ad un segnale: può svilupparsi in alcuni minuti e durare parecchie ore. Studi recenti mostrano che l'instaurarsi del potenziamento è determinato dalla concomitanza dell'arrivo di segnali al terminale presinaptico e dell'attivazione del neurone postsinaptico; l'entità del potenziamento è influenzata dal numero di volte successive in cui la simultaneità tra gli eventi citati ha luogo.

Conclusioni

Abbiamo qualitativamente descritto in che modo il neurone riceve, elabora, trasmette segnali di tipo elettrico (potenziali d'azione) e chimico (neurotrasmettitori).

Ciò che riteniamo importante sottolineare, e che può essere derivato dalla descrizione qualitativa del comportamento del neurone, è che il neurone realizza alcune operazioni elementari, tra queste interessa qui evidenziare la realizzazione di una *funzione soglia*.

Capitolo 2

Formazione di immagini

2.1 Introduzione

L'acquisizione di una immagine è il primo passo del processo di elaborazione di immagini digitali. Spesso una buona conoscenza di tutti gli elementi determinanti per l'acquisizione di una immagine ne rende possibile una corretta interpretazione.

I passi necessari affinché una scena tridimensionale sia rappresentabile da una immagine digitale sono i seguenti:

- L'interazione della scena con la luce o, più in generale, con radiazioni elettromagnetiche
- Un sistema che raccolga la radiazione riflessa dagli oggetti e la proietti su un piano immagine
- L'immagine continua è campionata (non si considerano gli infiniti punti del piano immagine ma quelli appartenenti ad una griglia opportuna) e quantizzata (ad ogni punto è associato un valore di un insieme finito)

Al fine di comprendere il processo che, a partire da una scena tridimensionale, porta alla formazione di una immagine (consideriamo solo immagini prodotte dalla luce non immagini ai raggi X, NMR, ...) dobbiamo rispondere a due domande fondamentali:

1. che cosa determina *dove* appaiono nell'immagine i punti sulle superfici degli oggetti della scena

2. che cosa determina *come* appaiono nell'immagine le superfici degli oggetti

Per rispondere alla prima domanda si devono analizzare le caratteristiche delle proiezioni prospettiche, per rispondere alla seconda è necessario introdurre alcuni concetti di radiometria, disciplina che si occupa della misura del flusso e del trasferimento di energia radiante.

Punti diversi sulle superfici degli oggetti posti di fronte ad un sistema di formazione di immagini (macchine fotografiche, telecamere...) avranno, nell'immagine, valori diversi di intensità e ciò in funzione: della quantità di energia radiante incidente sulle superfici, del modo in cui le superfici riflettono tale energia, del modo in cui l'energia riflessa è raccolta da un sistema di lenti e infine della risposta del sensore all'energia da cui è colpito. Vedremo che l'intensità di una immagine è proporzionale alla radianza della scena.

Fatte queste premesse vediamo di rispondere alla prima domanda e cioè cosa determina dove i punti sulle superfici degli oggetti della scena appaiono nell'immagine. Descriviamo un modello astratto di formazione della immagine il *pinhole camera model* perché, nonostante la sua semplicità, questo è un modello ragionevolmente accurato del processo di formazione di immagini mediante la maggior parte delle telecamere moderne (Vidicon, CCD, CID) nonché della macchina fotografica. Considerate il sistema rappresentato in Figura 2.1.

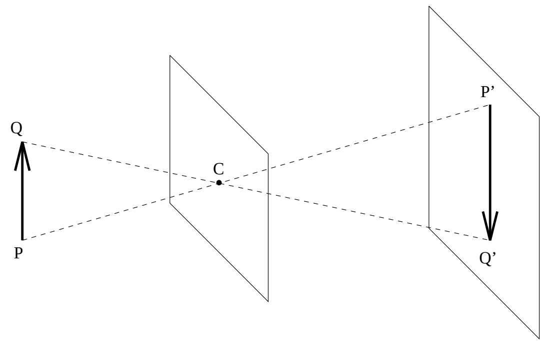


Figura 2.1 Formazione dell'immagine in una pinhole camera.

Esso consiste di due schermi; un foro piccolissimo C è stato effettuato sul primo schermo e, attraverso questo foro, i raggi di luce emessi o riflessi dalla superficie di un oggetto (la freccia PQ nell'esempio) passano formando una immagine invertita dell'oggetto sul secondo schermo. Poiché un punto sulla superficie di un oggetto in generale emette (o riflette) luce in tutte le direzioni, per avere sul piano immagine una immagine non sfuocata di un punto (per far si che ad un punto corrisponda un punto non un'areola) il foro deve essere così piccolo da consentire ad un solo raggio di passare. Questa situazione è fisicamente irrealizzabile, nella realtà si utilizzano lenti convergenti.

Abbiamo detto che C rappresenta un foro piccolissimo; le immagini P' e Q' dei punti P e Q sono tali per cui l'immagine della freccia PQ risulta capovolta. Per evitare la complicazione di descrivere la formazione di immagini capovolte ci porremo in una situazione più favorevole: immaginiamo che da un certo punto C (occhio di un osservatore) parta un insieme di raggi che raggiungono i punti delle superfici degli oggetti (Figura 2.2), vogliamo

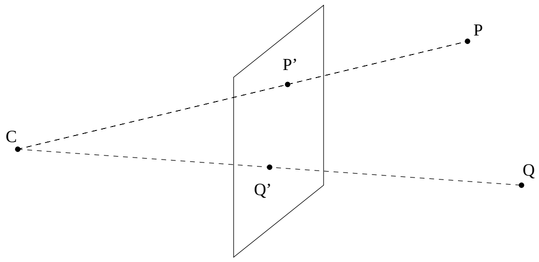


Figura 2.2 Formazione dell'immagine in una pinhole camera.

vedere come questi punti vengono proiettati su un piano posto tra l'osservatore e l'oggetto attraverso una *proiezione prospettica.*

La formazione di immagini mediante una pinhole camera consiste essenzialmente nella proiezione prospettica; anche se un sistema ottico reale non realizza una perfetta proiezione prospettica riteniamo qui trascurabili le deviazioni di un sistema reale dal modello astratto.

2.2 Proiezioni prospettiche

In generale le proiezioni trasformano punti da un sistema di coordinate ad n dimensioni in un sistema con un numero inferiore di dimensioni. Noi siamo interessati alle proiezioni dallo spazio tridimensionale (3D) allo spazio bidimensionale (2D) poiché un sistema di formazione di immagini può essere considerato un dispositivo che realizza tali proiezioni.

La proiezione di un oggetto 3D è definita da raggi che partono da un punto detto *centro di proiezione*, passano per ogni punto della superficie di un oggetto e intersecano un *piano di proiezione.* Considerate i due disegni di Figura 2.3, ciò che li distingue è la posizione relativa del piano di proiezione rispetto al centro di proiezione: in un caso la distanza che li separa è finita mentre nell'altro il centro di proiezione è un punto all'infinito e i raggi di proiezione sono paralleli. Si parla di proiezioni *prospettiche* e di proiezioni *parallele*, queste sono un caso particolare delle prime.

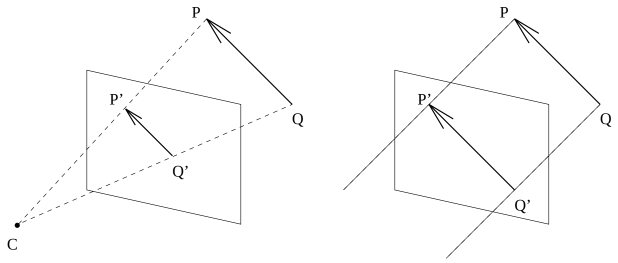


Figura 2.3 Proiezioni prospettiche e parallele.

Per descrivere in modo informale le regole della *proiezione prospettica* consideriamo un semplice esempio: immaginiamo che un pittore voglia rappresentare il cubo $ABCDEFHG$ su un certo piano verticale Π il *piano di proiezione* (Figura 2.4). Il centro di proiezione è O, l' occhio dell'osservatore; O costituisce il vertice degli angoli visuali. Supponiamo di tracciare delle rette dal centro di proiezione ad ogni punto sulla superficie dell'oggetto e consideriamo per semplicità solo i vertici del cubo. L'intersezione di ognuna delle rette col piano di proiezione Π è la *proiezione* del corrispondente punto dell'oggetto; ogni punto P viene rappresentato dal punto P' individuato dalla intersezione della retta OP con il piano Π.

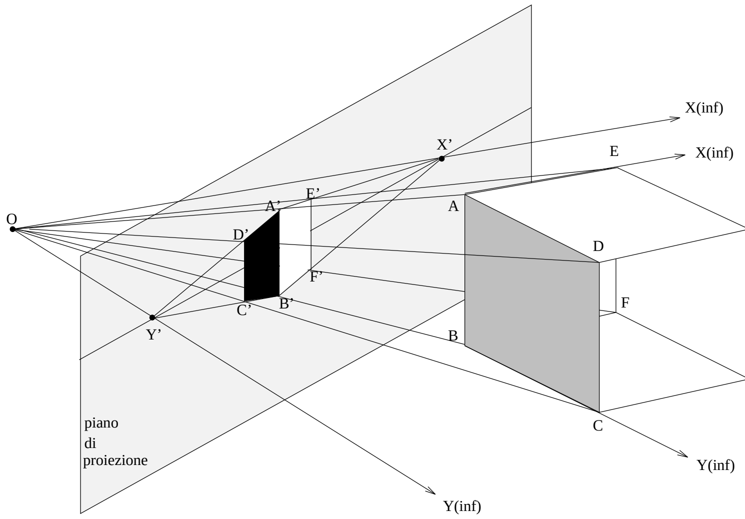


Figura 2.4 Regole della proiezione prospettica.

Così A viene rappresentato in A', B in B' e il segmento AB in $A'B'$. Per tutti i segmenti che giacciono su piani paralleli a Π vale una relazione di similitudine con la loro immagine. Ciò nell'esempio vale per gli spigoli verticali del cubo (sono gli unici spigoli paralleli a Π). Infatti i due triangoli OAB e $OA'B'$ sono simili così come ODC è simile a $OD'C'$, e OEF è simile a $OE'F'$.

Anche i segmenti che non giacciono su piani paralleli a Π vengono proiettati su segmenti immagine (vedi per esempio AD e $A'D'$); ma non vale la similitudine dei triangoli OAD e $OA'D'$. Ciò provoca una certa deformazione delle immagini rispetto all'oggetto. Così la faccia $ABCD$ del cubo (un quadrato) viene rappresentata dal quadrilatero $A'B'C'D'$ e la faccia $AEFB$ dal quadrilatero $A'E'F'B'$.

Particolarmente notevoli sono, sul piano di proiezione Π, i cosidetti *punti di fuga* X' e Y'. Essi sono la rappresentazione dei *punti all'infinito* X e Y. Un punto all' infinito è, secondo la geometria proiettiva, ciò che hanno in comune due rette parallele, cioè la loro direzione.

La geometria proiettiva è più generale della geometria euclidea, comprende come caso particolare la geometria euclidea. Si deriva dagli stessi assiomi della geometria euclidea con questa sola differenza: per la geometria euclidea due rette si incontrano in punto tranne nel caso in cui siano parallele, per la geometria proiettiva due rette si incontrano in un punto, anche se parallele; c'è pertanto una perfetta dualità tra punti e rette. Una retta ha solo un punto all'infinito (e non due); in quel punto essa incontra tutte le rette ad essa parallele. Ciò per soddisfare il postulato che due rette (nello stesso piano) individuano uno e un solo punto.

X è il punto all'infinito della retta AE (ed anche di tutte le rette parallele ad AE, come BF, DH, CG). La sua rappresentazione X' si ottiene tracciando la retta OX e intersecandola con il piano Π. La retta OX è l'unica retta passante per O e diretta nella direzione X (la direzione di AE e di tutte le sue parallele); quindi essa è univocamente determinata, ed è univocamente determinato il punto X'.

Il punto X' si chiama punto di fuga perché è l'immagine (finita) sul piano di proiezione, di un punto all'infinito della realtà. Tutte le rette $r_1, r_2, \ldots$ che passano nella realtà per X (cioè tutte le rette parallele ad AE) generano rappresentazioni $r'_1, r'_2, \ldots$ che passano per il punto X' (rette reali parallele generano immagini che parallele non sono). Per esempio AE e BF, che nella realtà sono parallele, nella rappresentazione sono rappresentate da $A'E'$ e $B'F'$ che si incontrano in X'. Analogo discorso vale per il punto reale $Y(\inf)$ e per la sua proiezione finita Y'.

La retta $X'Y'$ è , sul piano di proiezione, la retta dell'orizzonte. Tutti i punti reali sul piano dell'orizzonte (cioè il piano individuato dall'occhio del pittore O e dai due punti all'infinito (direzioni) orizzontali X e Y) vengono rappresentati sulla retta di orizzonte.

Oss. Il nostro pittore (come praticamente tutti i pittori) ha scelto un piano di rappresentazione verticale, come la retina dell'occhio. Quindi nell'esempio la direzione verticale della realtà (il punto all'infinito $Z(\inf)$) è rappresentato dalla direzione verticale del quadro ($Z'(\inf) = Z(\inf)$).

Ci sono infiniti punti di fuga (quante sono le direzioni delle rette) ma quelli corrispondenti ai 3 assi cartesiani sono i *punti di fuga principali.* Ci sono pertanto 3 punti di fuga principali, quanti sono gli assi principali (cartesiani) intersecati dal piano di proiezione.

Guardiamo ora come calcolare le coordinate del punto V_p, *punto immagine* del punto V di coordinate (x, y, z), supponendo note le coordinate del sistema di formazione dell'immagine (Figura 2.5).

In figura l'asse Y è perpendicolare al piano immagine e coincide con *l'asse ottico* o asse principale della camera (quello su cui i raggi non vengono deviati); su tale asse f è la misura della distanza tra il punto di osservazione C e il piano immagine. Il piano XZ interseca l'asse Y nel punto $y = 0$.

Consideriamo la proiezione del punto $V = (x, y, z)$ nel punto immagine $V_p = (x_p, 0, z_p)$ situato sul piano XZ. Osserviamo che i triangoli COA e CBD sono simili (il segmento BD si trova su un piano parallelo al piano XZ) ed anche triangoli COH e CBK sono simili, possiamo pertanto scrivere le relazioni:

$$OA : CO = DB : CB \qquad OH : CO = BK : CB$$

Ma osservando i segmenti considerati vediamo che la loro lunghezza è data da: $OA = x_p$, $CO = f$, $DB = x$, $CB = f + y$, $OH = z_p$, $BK = z$. Pertanto:

$$x_p : f = x : f + y \qquad z_p : f = z : f + y$$

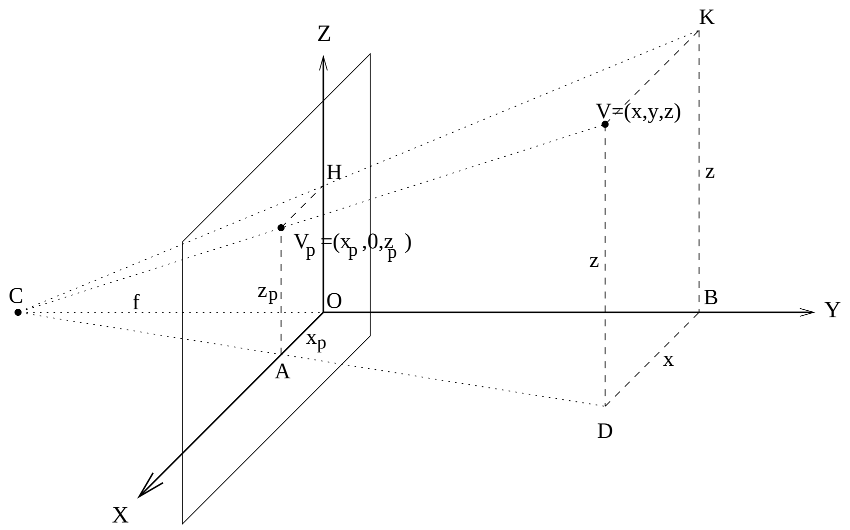


Figura 2.5 Modello di base del processo di formazione di una immagine.

Quindi:

$$x_p = \frac{x \cdot f}{f + y} \qquad z_p = \frac{z \cdot f}{f + y}$$

Queste equazioni sono non lineari. Le espressioni:

$$x = \frac{x_p}{f} \cdot (y + f) \qquad z = \frac{z_p}{f} \cdot (y + f)$$

mostrano che al punto immagine V_p di coordinate (x_p, z_p) corrisponde, nello spazio, l'insieme dei punti che stanno su una retta, la retta che passa per i punti di coordinate $(0, -f, 0)$ e $(x_p, 0, z_p)$ cioè per i punti C e V_p. Le equazioni mostrano che, a meno che non sia nota la terza coordinata, y, non è possibile ricavare le coordinate 3D del punto V a partire dalla sua immagine V_p.

Le proiezioni prospettiche sono spesso descritte utilizzando l'elegante formalismo delle coordinate omogenee.

2.3 Ricostruzione 3D

Abbiamo visto che nel processo di formazione di una immagine si perde l'informazione relativa ad una delle coordinate spaziali dei punti della scena che l'immagine rappresenta. Inoltre è ovvio che una superficie nello spazio è completamente proiettata sul piano immagine solo se non più di un punto della superficie giace sullo stesso raggio. Se ciò non accade parte della superficie resta invisibile. Questo effetto è detto *occlusione*. Se escludiamo il caso di occlusione, per ricostriure la forma tridimensionale di un oggetto abbiamo bisogno, oltre che della immagine, della 3^a coordinata (profondità) di ognuno dei punti rappresentati. Un metodo per ottenere tale informazione, metodo usato anche dal nostro sistema visivo, è quello di osservare la scena con due sensori situati in due diversi punti di vista. In altre parole è necessario avere due o più immagini della stessa scena ottenute nello stesso istante da dispositivi la cui posizione e orientazione relativa siano note (Figura 2.6).

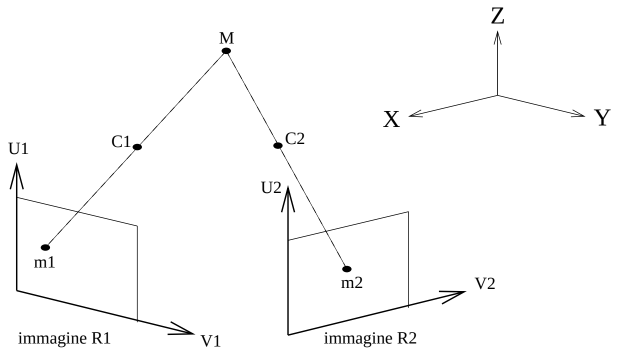


Figura 2.6 Il problema della ricostruzione 3D.

Date le immagini R_1 e R_2 al fine di ricostruire la terza dimensione devono essere risolti i seguenti problemi:

1. scelto un punto m_1 nella immagine R_1 decidere quale punto m_2 nella immagine R_2 corrisponde a m_1. Corrisponde significa che m_1 e m_2 sono entrambi immagini del punto M della scena. Questo è il *problema della corrispondenza.*

2. dati m_1 e m_2 calcolare le coordinate $3D$ del punto M (nel sistema di riferimento scelto). Questo è il *problema di ricostruzione.*

Non analizzeremo qui le assunzioni e i metodi sviluppati per risolvere il problema della corrispondenza, vediamo invece come, supponendo nota la sua soluzione, risulti semplice risolvere il problema di ricostruzione.

Per spiegare l'idea utilizziamo la situazione sperimentale più semplice, cioè quella in cui in cui i due piani immagine sono situati sullo stesso piano e la loro posizione relativa è quella presentata in Figura 2.7.

C_1 e C_2 rappresentano i centri ottici delle 2 pinhole camera. Sono definiti 3 sistemi di coordinate, uno per ogni piano immagine ((U_1, V_1) e (U_2, V_2) rispettivamente) e uno (X, Y, Z) per lo spazio 3D (*world reference frame*). L' origine del sistema di riferimento (X, Y, Z) è situata nel punto di mezzo del segmento che congiunge l' origine dei 2 sistemi di riferimento dei piani immagine (U_1, V_1) e (U_2, V_2).

Dati m_1 e m_2, punti corrispondenti in quanto entrambi sono immagini del punto M, di coordinate rispettivamente (u_1, v_1) e (u_2, v_2), viene definita *disparità* la differenza $\mid v_2 - v_1 \mid$. Una disparità uguale a 0 implica che il punto M è all'infinito. Se immaginiamo di spostare M via via più vicino al centro ottico C_1 allora la disparità fra m_1 e m_2 aumenterà da 0 a ∞.

Analizziamo ora, con l'aiuto del semplice diagramma di Figura 2.8, la relazione tra la disparità e la distanza del punto M dai centri ottici C_1 e C_2, distanza misurata nel sistema di riferimento (X, Y, Z).

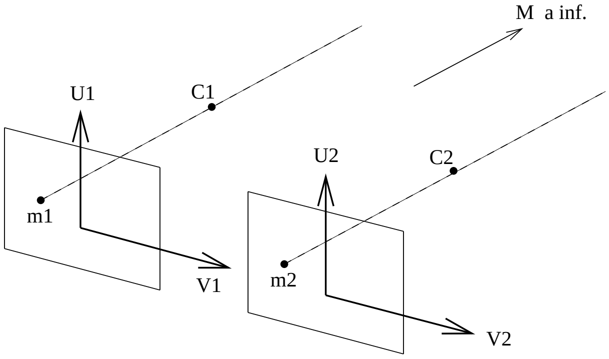


Figura 2.7 Semplice definizione di disparità.

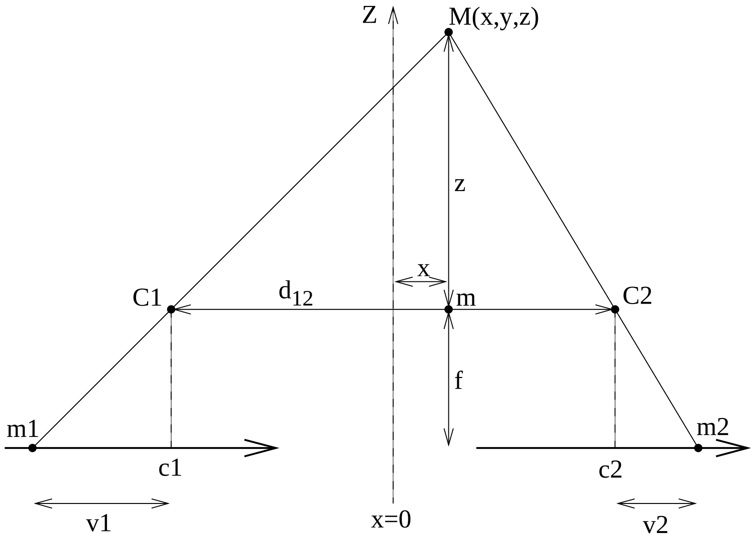


Figura 2.8 Relazione tra disparità e distanza.

La figura rappresenta una vista dall'alto delle 2 pinhole camera; in essa le coordinate del punto M sono denotate da (x, y, z); i centri ottici C_1 e C_2 sono separati dalla distanza d_{12}, distanza misurata lungo l'asse X del sistema di riferimento della scena; z è la distanza tra il piano in cui sono situati i 2 centri ottici e il punto M, f è la distanza focale, l'asse Y non è rappresentato. In questa configurazione il punto di mezzo del segmento $\overline{C_1C_2}$ ha coordinate $x = 0, z = 0$.

Consideriamo ora i triangoli $m_1c_1C_1$, C_1mM, $m_2c_2C_2$ e C_2mM, essi sono simili; valgono pertanto le seguenti relazioni:

$$\frac{m_1c_1}{c_1C_1} = \frac{mC_1}{Mm} \qquad \frac{m_2c_2}{c_2C_2} = \frac{mC_2}{Mm}$$

Cioè :

$$\frac{-v_1}{f} = \frac{d_{12}/2 + x}{z} \qquad \frac{v_2}{f} = \frac{d_{12}/2 - x}{z}$$

(v_1 è a sinistra rispetto all'origine avrà pertanto il segno -, x è la coordinata del punto m, proiezione di M sull'asse che unisce le 2 camere). Sommando membro a membro:

$$\frac{-v_1}{f} + \frac{v_2}{f} = \frac{d_{12}/2 - x + x + d_{12}/2}{z}$$

da cui:

$$v_2 - v_1 = \frac{f \cdot d_{12}}{z}$$

La distanza z è pertanto:

$$z = \frac{f \cdot d_{12}}{v_2 - v_1}$$

Questa relazione dice che un piano disposto di fronte e parallelamente ai piani immagine è il luogo dei punti a disparità costante (z costante).

Capitolo 3

Elementi di radiometria

3.1 Introduzione

Abbiamo visto che dal punto di vista geometrico la formazione di una immagine è modellizzabile mediante le proiezioni prospettiche. Vediamo ora a cosa è legato il valore del segnale immagine sia per immagini a livelli di grigio che per immagini a colori.

Il valore associato ad una piccola area del piano immagine è proporzionale alla *radianza* della superficie osservata. La radianza della superficie dipende dall'ammontare di energia radiante incidente sulla superficie, dalla frazione di energia incidente che viene riflessa e dal tipo di riflessione (speculare o diffusa). Oltre alla radianza della scena contribuiscono a determinare il valore della funzione immagine le caratteristiche del sensore e le caratteristiche del sistema ottico. Non analizzeremo la complessa funzione che, nota la radianza della scena e le caratteristiche del sensore, consente di stimare l'intensità in ogni punto dell'immagine; ci limiteremo a definire la radianza ed altre grandezze radiometriche che ci interessano e ad analizzare alcune importanti caratteristiche del sensore umano cioè dell'occhio e del sistema visivo.

Ricordiamo che la *luce* è *energia radiante* valutata rispetto alla sua abilità di stimolare l'organo della vista di un osservatore umano. Più precisamente è energia radiante la cui lunghezza d'onda λ è compresa nell'ambito $380 - 780\ nm$ (le onde elettromagnetiche hanno frequenze comprese nell'intervallo $10Hz - 3 \cdot 10^{24}Hz$). Esaminiamo ora le definizioni standard delle quantità radiometriche che ci interessano (la radiometria è la disciplina che si occupa della misura del flusso di energia radiante). Le definizioni sono quelle proposte dall'*International Lighting Vocabulary* della *Commission International de l'Eclairage,* *(CIE).*

3.2 Quantità radiometriche di base

Radiant energy (Energia radiante)
L'*energia radiante* è energia che si propaga sottoforma di onde elettromagnatiche o flusso di particelle (fotoni o quanti).
Radiazione (radiation) è il processo di emettere o trasferire energia radiante. Talvolta, il termine radiazione è identificato con il termine energia radiante.

Monochromatic radiant energy (energia radiante monocromatica)
L'*energia radiante monocromatica* è energia radiante di una singola frequenza (o lunghezza d'onda). Nella pratica questo termine è esteso ad includere energia radiante in un piccolo intervallo di frequenze (o lunghezze d'onda), intervallo che può essere descritto fissando una singola frequenza.

Period (Periodo)T
Il *periodo* è il tempo tra successive occorrenze delle stesse caratteristiche in un fenomeno periodico.

Frequency (Frequenza) ν
La *frequenza* è il reciproco del periodo ($\nu = 1/T$).

Wavelength (Lunghezza d'onda) λ
La *lunghezza d'onda* è la distanza fra 2 punti successivi di un'onda periodica nella direzione di propagazione in cui le oscillazioni hanno la stessa fase. Durante ogni periodo l'onda si propaga ad una distanza uguale alla lunghezza d'onda, perciò il prodotto $\lambda \cdot \nu$ della lunghezza d'onda per la frequenza è uguale alla velocità dell'onda. Nel *vuoto*, la velocità c di propagazione di un'onda elettromagnetica è costante e indipendente dalla frequenza e dall'ampiezza ($\lambda = c \cdot T, \quad c = \lambda \cdot \nu$). La velocità c decresce a c/n quando l'onda si propaga attraverso un mezzo diverso dal vuoto; n è l'indice di rifrazione del mezzo.

Wawenumber (Numero d'onda) m
Il *numero d'onda* è il reciproco della lunghezza d'onda.

Radiant Power (Potenza radiante) P
La *potenza radiante (o flusso radiante)* è l'energia radiante emessa, trasferita, o ricevuta attraverso una superficie, nell'unità di tempo.

Radiant Exitance (Emissione radiante) M
L'*emissione radiante* in un punto di una superficie è il quoziente fra la potenza radiante emessa da un elemento infinitesimo di superficie che contiene il punto, e l'area dell' elemento di superficie.

Irradiance (Irradiamento) E
L'irradiamento in un punto di una superficie è il quoziente fra la potenza radiante incidente su un infinitesimo elemento di superficie che contiene il punto, e l'area dell'elemento di superficie.

Radiant intensity (Intensità radiante) I
L'intensità radiante di una *sorgente puntiforme* in una determinata direzione, è il quoziente fra la potenza radiante emessa dalla sorgente in un infinitesimo elemento di angolo solido che contiene la direzione e l'elemento di *angolo solido*.

Radiance (Radianza) L
Fissata una direzione di propagazione, la *Radianza* in un punto di una superficie (sorgente, ricevente, o situata sulla via di un fascio) è il quoziente fra la potenza radiante che si propaga nella direzione data (e che lascia, arriva a, o passa per l'elemento infinitesimo di superficie che contiene il punto), e il prodotto tra l'angolo solido del cono elementare che contiene la direzione definita e l'area della proiezione dell'elemento di superficie su un piano ortogonale alla direzione data.

Spectral concentration (Densità spettrale)
La *densità spettrale* $X(\lambda)$ di una grandezza radiometrica X rispetto ad una data lunghezza d'onda λ è data dalla quantità della grandezza radiometrica in un intervallo infinitesimo di lunghezze d'onda contenente λ, diviso l'ampiezza $d\lambda$ dell'intervallo stesso. Formalmente:

$$X(\lambda) = \frac{dX}{d\lambda}$$

Per esempio se X denota la potenza radiante P, la densità spettrale della potenza radiante è denotata da $P(\lambda) = \frac{dP}{d\lambda}$ e viene semplicemente denominata *spectral radiant power* (spettro della potenza radiante).

Spectral distribution function (Funzione di distribuzione spettrale)
La *funzione di distribuzione spettrale* caratterizza la variazione della densità spettrale di una data quantità radiometrica in funzione della lunghezza d'onda. Il grafico corrispondente è denominato *spectral distribution curve*. Un simbolo utile è $\{X(\lambda)d\lambda\}$ che diventa $\{P(\lambda)d\lambda\}$ per la funzione di distribuzione spettrale della potenza radiante.

Relative spectral distribution function (Funzione di distribuzione spettrale relativa) Una *Funzione di distribuzione spettrale relativa* dà la densità spettrale in unità radiometriche arbitrarie (specifica solo valori relativi a diverse lunghezze d'onda).

3.3 Misura della intensità radiante e della radianza

L'intensità radiante I di una *sorgente puntiforme* in una determinata direzione, è la porzione dP della potenza radiante totale P emessa dalla sorgente, che si propaga, uniformemente distribuita, attraverso un piccolo cono di angolo solido $d\omega$ il cui apice è situato nella sorgente e il cui asse è in una direzione fissata; il flusso è costante nell'angolo solido. La Figura 3.1 illustra il concetto. Formalmente:

$$I = \frac{dP}{d\omega}$$

L'unità di misura standard dell'intensità radiante è il watt per steradian ($W \cdot sr^{-1}$). Ricordiamo infatti che un angolo solido si misura in *steradian* (radianti stereo); la dimensione di un angolo solido è uguale all'area intercettata sulla sfera di raggio unitario da un cono il cui vertice sia nel centro della sfera. L'angolo solido totale attorno ad un punto nello spazio è pertanto 4π.

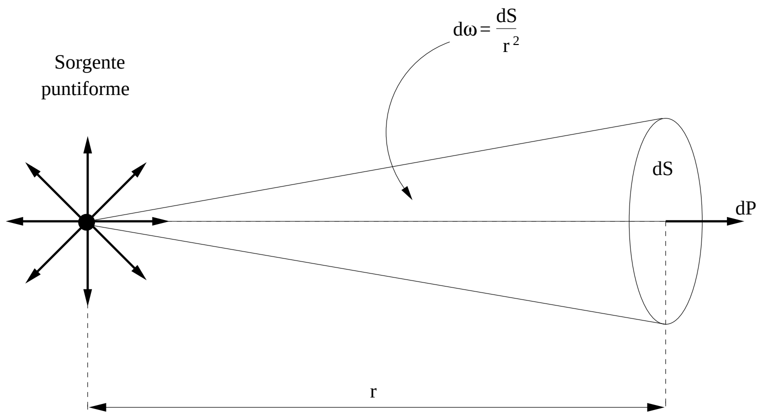


Figura 3.1 Concetto di intensità radiante.

Supponiamo ora che una superficie intercetti il flusso dP nell'angolo solido $d\omega$. L'*Irradiamento* E è il quoziente tra la potenza radiante dP e l'area elementare dS. In questa definizione, come in Figura 3.1, assumiamo che la superficie sia perpendicolare all'asse del cono, cioè alla direzione di irradiamento (per questo denotiamo l'irradiamento con E_n):

$$E_n = \frac{dP}{dS}$$

L'unità di misura standard dell'irradiamento è il watt per metro quadrato ($W \cdot m^{-2}$). L'irradiamento è la densità di potenza sulla superficie; aumentando la superficie diminuisce l'irradiamento.

Vediamo ora la relazione che lega l'intensità radiante I all'irradiamento E_n. L'angolo solido Θ sotteso da una superficie di area S è definito dal cono il cui vertice è nel punto di radiazione e il cui asse è il segmento che congiunge il punto di radiazione al centro dell'elemento di superficie. Se la superficie è ad una distanza r dal punto di radiazione e la normale alla superficie forma un angolo ϵ ripetto all'asse del cono allora:

$$\Theta = \frac{S \cdot \cos\epsilon}{r^2}$$

Nell'esempio di Figura 3.1, la normale alla superficie infinitesima di area dS coincide con l'asse del cono dw, scriviamo pertanto:

$$dw =_{def} \frac{dS}{r^2}\Omega$$

Per utilizzare il quoziente $\frac{dS}{r^2}$ nelle espressioni matematiche che ci interessano e mantenere la dimensione di steradian introduciamo il fattore Ω il cui valore scalare è l'unità ma la cui dimensione è lo steradian.
Dalle equazioni sopra scritte si ottiene la relazione che lega l'intensità radiante all'irradiamento:

$$E_n = \frac{dP}{r^2 \cdot dw}\Omega = \frac{I}{r^2}\Omega$$

la cui dimensione è il watt per metro quadrato per steradian ($W \cdot m^{-2} \cdot sr^{-1}$).

L'irradiamento su un elemento di superficie, perpendicolare alla linea che congiunge la sorgente puntiforme e il punto al centro della superficie, è uguale alla intensità radiante della sorgente nella direzione dell'elemento di superficie, divisa il quadrato della distanza tra superficie e sorgente. Più semplicemente: l'irradiamento di un elemento di superficie è proporzionale alla intensità radiante della sorgente e inversamente proporzionale al quadrato della distanza dalla sorgente. Questa relazione è frequentemente chiamata *inverse-square law of irradiance for normal incidence.*

L'irradiamento per una superficie la cui normale non coincide con la direzione considerata è dato da $E = E_n \cos\epsilon$ dove ϵ è l'angolo di incidenza (angolo tra la normale alla superficie e la direzione di irradiamento considerata). Pertanto:

$$E = \frac{I}{r^2}\cos\epsilon$$

Questa è la *inverse-square law of irradiance for oblique incidence.*

Nel caso in cui le sorgenti che emettono energia radiante non siano sorgenti puntiformi dobbiamo parlare di *radianza* e non di intensità radiante. Vediamo di derivare una misura anche per la radianza.

Se ho una superficie di dimensioni finite che emette energia radiante assumo di poter considerare che la potenza totale P sia la somma della potenza emessa da "piccoli pezzi" di S (cioè da superfici infinitesime). Data una superficie S individuo una superficie infinitesima $\overline{dS}$ mediante il prodotto dell'area infinitesima dS con il vettore $\overline{n}$ normale alla superficie ($\overline{n}$ ha modulo unitario e direzione ortogonale alla superficie). Pertanto $\overline{dS} = \overline{n} \cdot dS$ (Figura 3.2).

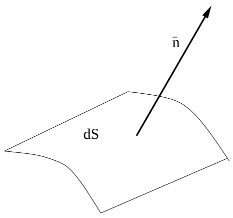


Figura 3.2 Superficie infinitesima.

Ogni superficie infinitesima $\overline{dS}$ emette energia in tutte le direzioni. Vediamo ora qual'è il flusso totale emesso da S e intercettato da una superficie R.

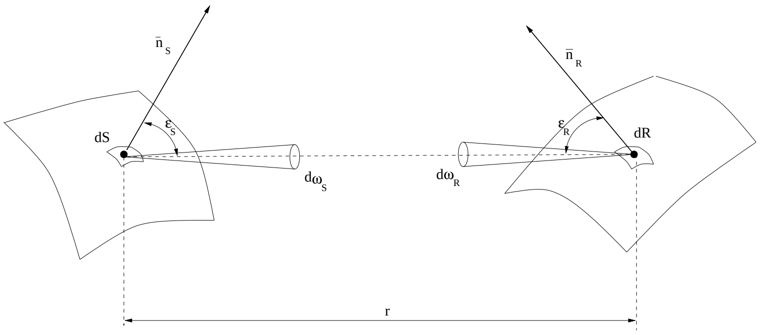


Figura 3.3 Elementi per la definizione della radianza.

Consideriamo una coppia di superfici infinitesime $\overline{dS}$ e $\overline{dR}$ (vedi Figura 3.3). Esse individuano una direzione di flusso $\overline{d}$ (quella della retta che le congiunge) ed una distanza r su questa retta. Esse inoltre possono essere variamente orientate rispetto alla direzione $\overline{d}$ ($\overline{d} = \frac{\overline{r}}{|r|}$).

Denotiamo con ϵ_S e ϵ_R gli angoli che le normali $\overline{n_S}$ e $\overline{n_R}$ alle superfici $\overline{dS}$ e $\overline{dR}$ formano con la direzione $\overline{d}$. Avremo pertanto:

$$\overline{dS} \times \overline{d} = dS \cdot \cos \epsilon_S$$

$$\overline{dR} \times \overline{d} = dR \cdot \cos \epsilon_R$$

Definiamo d^2P il flusso attraverso $\overline{dR}$ indotto da $\overline{dS}$ (abbiamo denotato tale flusso con d^2P per indicare che è un infinitesimo di 2° ordine). Fissata la direzione $\overline{d}$ e le posizioni rispettive delle due superfici osserviamo che:

$$d^2P \div dS \cos \epsilon_S$$

il flusso è proporzionale all'area dS, raddoppiando l'area esso raddioppia; è anche proporzionale alla posizione della superficie rispetto alla direzione di flusso. Analogamente:

$$d^2P \div dR \cos \epsilon_R$$

Inoltre il flusso è inversamente proporzionale al quadrato della distanza:

$$d^2P \div \frac{1}{r^2}$$

Possiamo pertanto asserire che:

$$d^2P = L(\overline{d}) \cdot \frac{dS \cos \epsilon_S dR \cos \epsilon_R}{r^2}$$

Questa espressione dice che la variazione della potenza in una direzione $\overline{d}$ è proporzionale all'area delle 2 superfici e all'angolo che la loro normale forma con la direzione considerata $\overline{d}$ e inversamente proporzionale al quadrato della distanza. $L(\overline{d})$, la costante di proporzionalità che dipende dalla direzione considerata, prende il nome di *radianza*.

Dall'equazione sopra scritta e dalle seguenti relazioni:

$$\frac{dS \cos \epsilon_S}{r^2} = d\omega_R \qquad \frac{dR \cos \epsilon_R}{r^2} = d\omega_S$$

si deriva:

$$d^2P = L(\overline{d}) dS \cos \epsilon_S d\omega_S = L(\overline{d}) dR \cos \epsilon_R d\omega_R$$

ed anche:

$$L(\overline{d}) = \frac{d^2P}{dS \cos \epsilon_S d\omega_S} = \frac{d^2P}{dR \cos \epsilon_R d\omega_R}$$

L'espressione mostra chiaramente che la radianza si misura in watt per metro quadrato per steradian $(W \cdot m^{-2} \cdot sr^{-1})$.

Alternativamente si può derivare:

$$\frac{d^2P}{d\omega_S} = L(\overline{d}) dS \cos \epsilon_S = \frac{d^2P}{d\omega_R} = L(\overline{d}) dR \cos \epsilon_R = dI$$

dove dI indica la densità di intensità su dS o su dR; pertanto:

$$L(\overline{d}) = \frac{dI}{dS \cos \epsilon_S} = \frac{dI}{dR \cos \epsilon_R}$$

Queste equazioni indicano che la radianza può essere definita anche come la densità di intensità sulle superfici infinitesime $\overline{dS}$ e $\overline{dR}$.

Naturalmente il flusso totale sarà :

$$\int_S \int_R d^2P = \int_S \int_R L(\overline{d}) \cdot \frac{dS \cos \epsilon_S dR \cos \epsilon_R}{r^2}$$

Intuitivamente, la potenza totale si ottiene "sommando" su tutte le superfici infinitesime che costituiscono la superficie ricevente la potenza emessa da ognuna delle superfici infinitesime che costituiscono la sorgente.

Si definisce *superficie di Lambert* una superficie tale che:

$$L(\overline{d}) = costante$$

cioè la radianza non varia con la direzione. Se L è costante avremo che I varia con $\cos \epsilon_S$, I è pertanto massima nella direzione della normale alla superficie, infatti:

$$dI = \frac{d^2P}{d\omega_S} = L(0) dS \cos \epsilon_S$$

L'intensità emessa da una superficie di Lambert varia con il coseno dell'angolo che la direzione individuata forma con la normale alla superficie.

3.4 Sorgenti di energia radiante

Le caratteristiche di una sorgente di energia radiante vengono generalmente descritte mediante la distribuzione spettrale relativa della potenza radiante. La Figura 3.4 mostra tale funzione per la luce del sole in 2 diversi momenti del giorno.

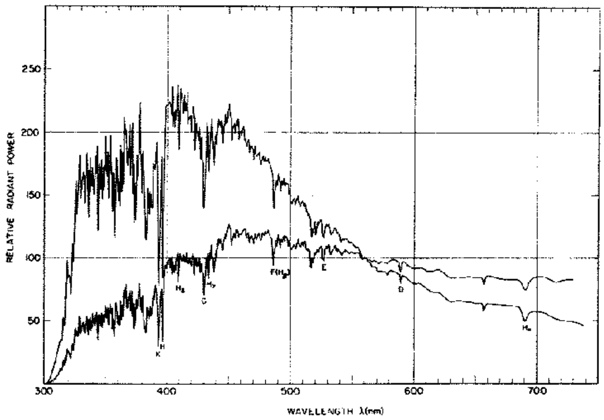


Figura 3.4 Distribuzione spettrale relativa della potenza radiante della luce del sole in 2 diversi momenti del giorno.

E' standard per la CIE fornire le distribuzioni relative di potenza radiante a intervalli di 10 nm, ogni valore tabulato rappresenta il valore medio misurato in un intervallo di lunghezze d'onda largo 10 nm e centrato sulla lunghezza d'onda specificata. Con questa rappresentazione la struttura fine dello spettro si perde come si osserva chiramente confrontando la Figura 3.4 con la Figura 3.5 che mostra la distribuzione spettrale relativa di 10 fasi della luce del giorno data dalla CIE.

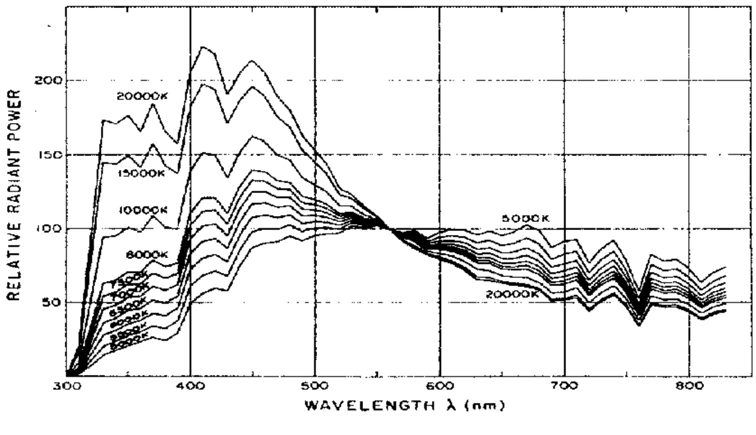


Figura 3.5 Distribuzione spettrale relativa della potenza radiante della luce del sole in 10 diversi momenti del giorno.

Tra le sorgenti di energia radiante ricordiamo, estremamente importante come riferimento, il *corpo nero.*

L'energia radiante in uno spazio chiuso o in una cavità le cui pareti sono mantenute ad una temperatura uniforme T dipende, per quanto riguarda sia la densità sia la distribuzione spettrale di potenza, esclusivamente dalla temperatura delle pareti e non dal materiale delle pareti.

L'energia radiante emessa da corpi solidi incandescenti è sempre minore di quella emessa dal corpo nero alla stessa temperatura. Il rapporto fra la emissione radiante di un radiatore termico e quella del corpo nero alla stessa temperatura è detta *emissività* ϵ del radiatore termico. In generale la emissività è funzione della lunghezza d'onda.

Nell'intervallo di lunghezze d'onda che costituisce l'intervallo visibile la distribuzione spettrale della radiant exitance di materiali selettivi quali carbone, platino e tungsteno è una approssimazione di quella del corpo nero.

Il tungsteno è il materiale preferito per costruire lampade incandescenti. La emissività del tungsteno non è costante ma decresce con la lunghezza d'onda e aumenta con la temperatura. Le lampade con filamenti di tungsteno sono spesso combinate con filtri di vetro colorato per riprodurre, almeno approssimativamente, la distribuzione spettrale di potenza radiante di certe desiderate fasi della luce del giorno. Tali simulatori della luce del giorno sono importanti per la valutazione visiva di materiali colorati e per effettuare una corrispondeza visiva di colori in molti compiti di controllo di produzioni industriali.

In molte lampade a scarica elettrica l'energia radiante è prodotta passando una corrente elettrica attraverso gas o vapore. Gli elettroni, accellerati da un campo elettrico applicato tra due elettrodi, collidono con le molecole del gas cedendo ad esse parte della energia cinetica. Quando queste particelle, il cui livello di energia è stato aumentato, ritornano al loro stato normale, l'energia di eccitazione è emessa in forma di quanti di energia radiante. In contrasto con i radiatori termici descritti sopra, le lampade a scarica elettrica emettono spettri a linee oltre ad un più o meno intenso spettro continuo.

Uno spettro a linee è composto da un numero di linee in corrispondenza di certe lunghezze d'onda, le linee sono separate da regioni in cui non è emessa energia radiante.

Bibliografia

- G. Wyszecki, W. S. Stiles.: *Color Science. Concept and Methods, Quantitative Data and Formulae* - Second Edition -, Jhon Wiley & Sons, Toronto 1982.

Capitolo 4

L'occhio

4.1 Introduzione

Dopo aver introdotto il modello astratto di formazione dell'immagine e le grandezze radiometriche di interesse (potenza radiante, radianza...) presentiamo come esempio di sensore e di sistema ottico, il più complesso dei sistemi: l'occhio umano. Forniremo inoltre qualche caratteristica percettiva del sistema visivo. Ciò perché, spesso, obiettivo della elaborazione di immagini è quello di migliorare la qualità dell'immagine ai fini della visualizzazione. Per *sistema visivo* si intende l'occhio (sistema ottico + retina) e quella parte del Sistema Nervoso Centrale (SNC) che elabora l'informazione visiva, cioè realizza quel processo che dallo stadio iniziale di attivazione dei recettori della retina ad opera della luce (acquisizione dell'immagine se parlassimo di un sistema automatico) porta alla interpretazione, a dare cioè significato a ciò che vediamo. Se avessimo un sistema nervoso diverso vedremmo in modo diverso; ci sono differenze strutturali e funzionali tra sistemi visivi di animali appartenenti a specie diverse. Per inciso, non si deve credere che le funzioni visive migliorino lungo la scala evolutiva con un massimo di efficenza e perfezione nella specie umana. Vi sono animali che hanno la stessa capacità dell'uomo di percepire i colori e vi sono animali superiori all'uomo per altre proprietà, per esempio per acuità visiva (capacità di discriminare come distinti 2 punti tra loro molto vicini).

4.2 L'occhio

E' un sistema 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 protettivo in cui è inserito, involucro che ne preserva la forma, sono situate tre coppie di muscoli *muscoli extraoculari* che muovono l'occhio nell'orbita. Ogni cambiamento di posizione dell'occhio è causato dalla contrazione ben coordinata di questi muscoli; i muscoli di ogni coppia sono antagonisti.

L'occhio è costituito da un *sistema ottico* (cornea, cristallino, umor acqueo, corpo vitreo) e da un sistema *sensoriale*, la retina.

4.2.1 Il sistema ottico

Il sistema ottico è costituito da:

- **La cornea** è lo strato più esterno, lo strato attraverso il quale lo stimolo esterno (luce) entra nell'occhio. Sebbene sia costituita da una complessa struttura lamellare la cornea è trasparente. Dal punto di vista ottico la cornea è un *diottro convergente* (interfaccia tra due mezzi con indici di rifrazione diversi). Le ghiandole lacrimali, localizzate all'esterno dell'occhio, secernono una soluzione che aiuta a mantenere normale lo scambio di ossigeno e il bilancio dell'acqua. E' questo processo continuo che mantiene la cornea trasparente. La chiusura delle palpebre distribuisce il liquido lacrimale sulla cornea; recettori per il dolore situati nella cornea aiutano a proteggere l'occhio da ingiurie (riflesso che determina la chiusura delle palpebre). La cornea è contigua con la sclera, involucro protettivo robusto su cui sono inseriti i muscoli extraoculari.

- **Il cristallino** è una *lente biconvessa* a curvatura regolabile ed è la curvatura variabile del cristallino che ci consente di mettere a fuoco oggetti a distanze diverse (per mettere a fuoco una macchina fotografica, si varia la distanza tra lente e pellicola; noi mettiamo a fuoco cambiando la forma della lente). La curvatura del cristallino è infatti controllata da un sistema di muscoli *i muscoli ciliari*. Inoltre due altri gruppi di muscoli modificano il diametro della pupilla regolando così la quantità di luce che entra nell'occhio. Analogamente, in una macchina fotografica, il diaframma determina l'apertura dell'obiettivo.

- **L'umor acqueo e il corpo vitreo**. Tra la cornea e la lente è situata la camera anteriore dell'occhio che è piena di un liquido chiaro *l'umor acqueo*. Il liquido, continuamente generato e assorbito, controlla la pressione intraoculare, che è più grande della pressione atmosferica, per mantenere l'integrità strutturale dell'occhio. Nello spazio tra la lente e la retina c'è il *corpo vitreo*, una gelatina trasparente che costituisce circa i 2/3 del volume dell'occhio. Il corpo vitreo e l'umor acqueo sono composti principalmente di acqua, e le loro proprietà ottiche sono molto simili a quelle dell'acqua. Così l'occhio è essenzialmente una camera oscura piena d'acqua (la macchina fotografica è piena di aria).

La cornea e il cristallino sono responsabili della *rifrazione* della luce che entra nell'occhio. La maggior parte di questa rifrazione si verifica ad opera della cornea, che possiamo considerare semplicemente come una superficie convessa di separazione tra l'aria all'esterno e l'umor acqueo all'interno. Se non ci fosse il cristallino, i raggi provenienti da un lontano punto luminoso dopo essere passati attraverso la cornea, convergerebbero verso un punto situato a circa 1 *cm* dietro la retina. La lente curva ulteriormente i raggi, che la raggiungono attraverso la pupilla, per portarli a fuoco sulla retina. Il cristallino, se usato nell'aria, avrebbe un potere convergente maggiore di quanto non abbia nell' occhio, dove è circondato dall'acqua. I cambiamenti di forma della lente, con la conseguente alterazione del suo potere convergente, sono i responsabili dell'accomodazione. Così, nell'occhio, la messa a fuoco è realizzata in un modo completamente diverso da quello della macchina fotografica e delle telecamere.

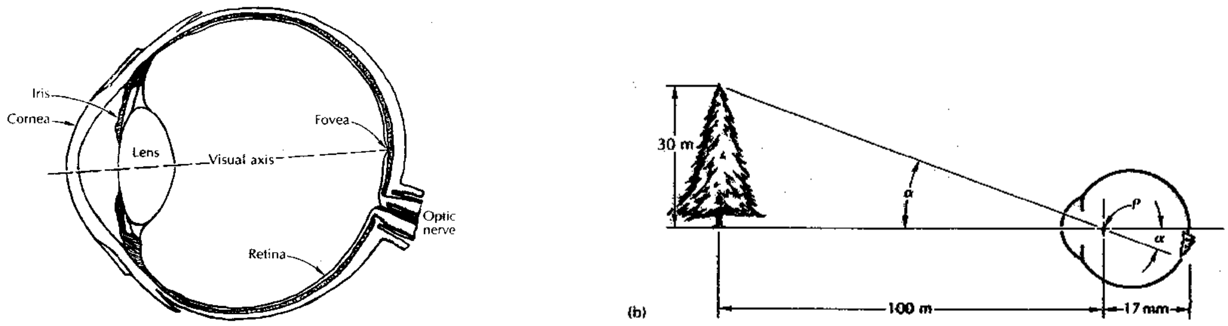


Figura 4.1 Diagramma di una sezione dell'occhio e rappresentazione semplificata dell'ottica.

La Figura 4.1 presenta lo schema di una sezione trasversale dell'occhio umano insieme ad una semplificata rappresentazione dell'occhio come sistema ottico. La combinazione cornea + cristallino può essere considerata come una singola lente che forma l'mmagine di un oggetto sulla retina. In certi casi la distanza tra cornea e retina è così grande che l'immagine è formata non sulla retina ma davanti ad essa; persone con questa caratteristica sono dette *miopi*. E' altrettanto frequente il difetto opposto, la distanza tra cornea e retina è troppo breve e l'immagine è formata dietro la retina; persone con questa caratteristica sono dette *ipermetropi*. Se un oggetto a fuoco sulla retina di un osservatore viene avvicinato, l'immagine retinica apparirà sfuocata e per renderla di nuovo nitida la curvatura del cristallino deve essere aumentata. C'è un punto *punto focale o nodale* nel centro di ogni lente tale che i raggi luminosi che passano attraverso il punto non vengono deflessi, e questo punto è utile per localizzare l'immagine.

Nell'occhio umano tale punto è situato a circa 17 *mm* davanti alla retina. Grazie a questa informazione è facile calcolare la dimensione dell'immagine retinica di un oggetto. Nella figura, l'osservatore sta osservando un albero, alto 30 *m* e distante 100 *m*. Supponiamo che l'immagine dell'albero sia a fuoco sulla retina, allora le immagini della punta e della base dell'albero possono essere individuate molto semplicemente tracciando raggi che non vengono deflessi in quanto passano per il punto focale P della lente. Osservando la figura si nota che vale la seguente semplice relazione: $30 : 100 = x : 17$ da cui abbiamo: $x = 17 \cdot 30/100 = 5.1\ mm$; x è la dimensione dell'immagine retinica. In generale, denotando con d_0 e d_i rispettivamente le distanze dell'oggetto e della immagine dal punto focale della lente e con S_0 e S_i le dimensioni dell'oggetto e dell'immagine vale la seguente relazione:

$$\frac{S_0}{d_0} = \frac{S_i}{d_i}$$

Una misura conveniente della dimensione di uno stimolo visivo, piuttosto che la sua dimensione in m, $cm \ldots$ è *l'angolo visivo* che in figura è denotato da α:

$$\alpha = \arctan(30/100) = 16°27'$$

Se la dimensione dell'oggetto è data come angolo visivo la dimensione della immagine retinica sarà:

$$\tan\alpha = \frac{\sin\alpha}{\cos\alpha} = \frac{S_0}{d_0} = \frac{S_i}{d_i} = \frac{S_i}{17} = \frac{dimensione\ immagine}{17}\ mm$$

o alternativamente:

$$\tan\alpha \cdot 17 = S_i = dimensione\ immagine\ retinica\ \ mm$$

L'angolo visivo ci fornisce sia la dimensione dell'oggetto sia la dimensione della immagine retinica; la dimensione dell'oggetto viene espressa in modo indipendente dalla distanza. Fissata la dimensione dell'oggetto, l'angolo visivo da cui l'oggetto è visto diminuisce con la distanza dell'oggetto dall'osservatore. Per esempio un oggetto (barra) alto 1 cm posto a 57 cm da un osservatore è visto con un angolo visivo $\alpha = 1°$ $(\tan\alpha = 1/57)$ e la sua immagine retinica è di circa 0.3 mm $(17 \cdot 1/57)$

4.2.2 Il campo visivo

Il campo visivo di ogni occhio è la regione dello spazio esterno in cui l'occhio, quando è in posizione fissa e non ruota nell'orbita, può vedere gli oggetti. Possiamo considerare il campo visivo di ogni occhio come "unito" ad esso. Così, se manteniamo la testa immobile, ma l'occhio ruota nell'orbita, il campo visivo si muove con l'occhio e con il punto di fissazione, che corrisponde al centro della fovea. Di conseguenza il campo visivo totale coperto dall'occhio in movimento è notevolmente più grande del campo visivo stesso. Noi siamo quasi inconsapevoli delle proprietà delle diverse parti del campo visivo poiché tendiamo sempre a ruotare gli occhi in modo da usare la visione foveale quando qualcosa attira la nostra attenzione. Tuttavia la visione periferica è di grande importanza pratica, poiché dà un' ampia visione delle cose e ci permette di scoprire gli oggetti che si muovono fuori dall'angolo dell'occhio: sarebbe quasi impossibile muoversi con sicurezza senza la visione periferica.

4.2.3 La retina

La complessa struttura del bulbo oculare ha il compito di formare una immagine del mondo esterno sulla retina. La retina è in grado di trasformare segnali luminosi (onde elettromagnetiche di lunghezza d'onda compresa fra i 400 nm e 770 nm) in segnali "nervosi". Essa ci permette di vedere in condizioni che vanno dalla luce delle stelle alla luce del sole, discrimina 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 retina è una parte del cervello, rimasta segregata molto precocemente nel corso dello sviluppo, che ha mantenuto le connessioni con il cervello propriamente detto attraverso un fascio di fibre: il *nervo ottico*. Come molte altre strutture del sistema nervoso centrale la retina ha la forma di una lamina, in questo caso dello spessore di circa 1/4 di millimetro. E' costituita da tre strati di neuroni, separati da due strati che contengono le sinapsi fatte dagli assoni e dai dendriti di queste cellule. La fila di cellule nella parte posteriore della retina è costituita dai *fotorecettori*, i bastoncelli e i coni. I *bastoncelli*, che sono più numerosi dei coni, 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 della visione dei colori. Poiché i coni e i bastoncelli sono situati nella parte posteriore della retina, la luce che ad essi arriva deve attraversare gli altri strati di cellule per stimolarli. Non si capisce bene perché la retina si sviluppi in questo modo curioso, a rovescio. Una ragione possibile è l'esistenza, dietro i fotorecettori, di una fila di cellule che contengono un pigmento nero, la *melanina*. La melanina assorbe la luce che ha attraversato la retina, impedendo che essa sia riflessa all'indietro e si diffonda all'interno dell'occhio: ha la stessa funzione della vernice nera all'interno di una macchina fotografica. Le cellule che contengono la melanina hanno anche il compito di aiutare la risintesi chimica del pigmento visivo dei recettori dopo che è stato sbiancato dalla luce. Per svolgere queste due funzioni è necessario che il pigmento melanico sia vicino ai recettori. Se i recettori fossero nella parte anteriore della retina, le cellule pigmentate dovrebbero trovarsi tra essi e lo strato successivo di cellule, in una regione già occupata da assoni, dendriti e sinapsi. Gli strati davanti ai recettori sono trasparenti e non offuscano molto l'immagine. Nel millimetro centrale tuttavia, dove l'acuità visiva è massima, le conseguenze di un offuscamento anche modesto sarebbero disastrose; in questa piccola regione gli altri strati sono spostati lateralmente in modo da esporre i coni centrali direttamente alla luce. L'infossamento che ne risulta è *la fovea*. Le quantità relative di coni e bastoncelli variano in modo marcato nelle varie parti della retina .

La Figura 4.2 mostra la distribuzione dei coni e dei bastoncelli nella retina. L'asse orizzontale ha 2 scale; una è semplicemente la distanza, misurata in mm, tra un punto anatomicamente ben definito *la fovea* e la regione in cui è misurata la densità dei bastoncelli. E' noto che, quando si richiede ad un soggetto di fissare un punto, egli muove gli occhi finché, in ogni occhio, l'mmagine del punto cade sulla fovea. Perciò la scala in mm, corrisponde ad un'altra scala (Figura 4.2), quella in cui si misura l'angolo visivo, il cui valore 0 rappresenta l'angolo visivo tra la fovea e la linea dello sguardo (la distanza dal centro ottico alla retina è approssimativamente di 17 mm; questo è il fattore richiesto per

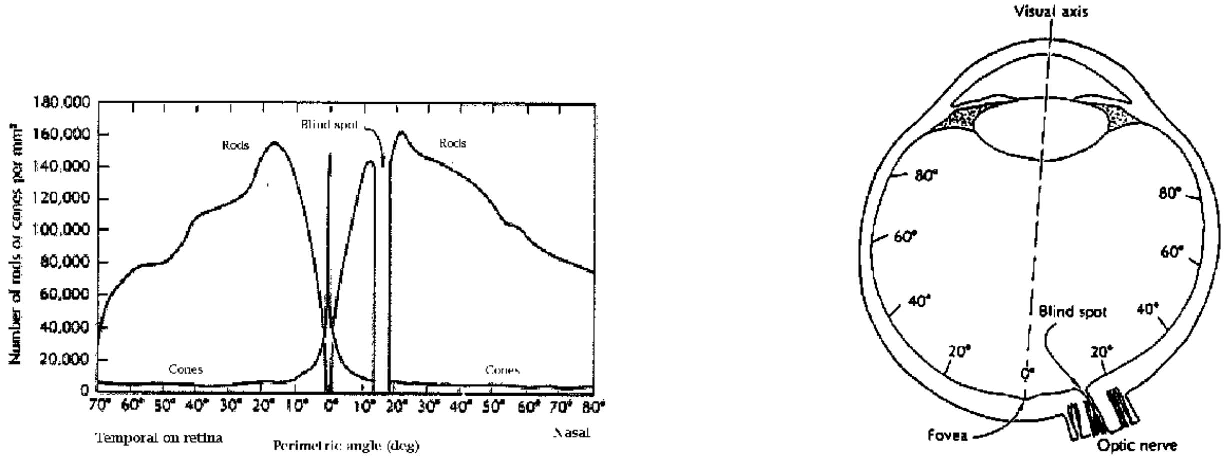


Figura 4.2 Disposizione di coni e bastoncelli.

convertire i mm sulla retina in dimensione di angolo visivo). Nella fovea, area priva di bastoncelli il cui diametro è di circa $1/2$ mm, ci sono solo coni. Nella fovea la visione del dettaglio è migliore. I bastoncelli sono cellule che rispondono a luci molto deboli. E' esperienza comune che se si guarda direttamente ad un oggetto si possono vedere in esso dettagli fini molto meglio che se lo si guarda non direttamente. Invece, se l'oggetto è semplicemente una macchia di luce molto debole su fondo completamente scuro, per esempio una stella, e se gli occhi sono adattati al buio, allora si vede meglio se lo si guarda un pò di lato piuttosto che direttamente. Spostandosi verso la parte anteriore dell'occhio, troviamo lo strato intermedio della retina, strato posto tra i fotorecettori e le cellule gangliari. Questo strato contiene 3 tipi di cellule nervose: *le celluli bipolari, le cellule orizzontali e le cellule amacrine (Figura 4.3).*

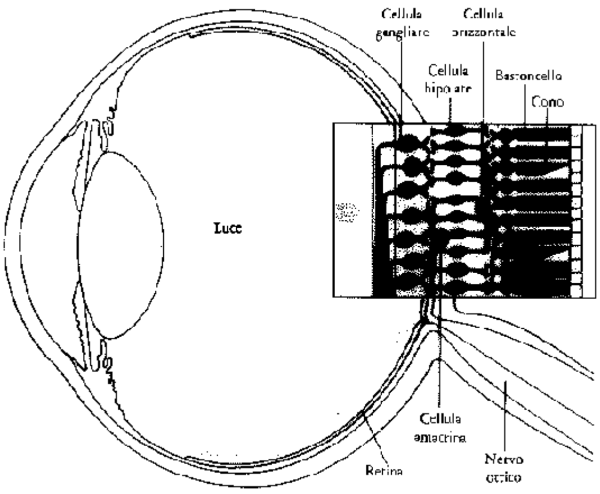


Figura 4.3 Struttura della retina.

Le cellule bipolari ricevono l'input dai recettori e molte di esse si connettono direttamente alle cellule gangliari. Le cellule orizzontali collegano le cellule bipolari ed i recettori attraverso connessioni relativamente lunghe che decorrono parallele agli strati della retina; in modo analogo, le cellule amacrine collegano le cellule bipolari e le cellule gangliari. Lo strato di cellule sulla superficie anteriore della retina contiene le *cellule gangliari*, i cui assoni attraversano la retina, si raccolgono in un fascio a livello del disco ottico e lasciano l'occhio formando il *nervo ottico*. Ciascun occhio contiene circa 125 milioni di coni e bastoncelli ma solo 1 milione di cellule gangliari. A fronte di questa discrepanza ci dobbiamo chiedere come viene mantenuto il dettaglio nell'informazione visiva. La risoluzione del problema è facilitata dall'analisi delle connessioni all'interno della retina. Si può considerare che l'informazione che attraversa la retina segua 2 strade: una strada diretta dai fotorecettori alle cellule bipolari e alle cellule gangliari, e una via indiretta, nella quale cellule orizzontali possono essere interposte tra recettori e cellule bipolari, e cellule amacrine tra cellule bipolari e cellule gangliari. La via diretta è altamente specifica, o privata, nel senso che uno solo, o pochi, recettori si connettono con una cellula bipolare, e solo una, o poche, cellule bipolari si connettono con una sola cellula gangliare. La via indiretta è più diffusa, o estesa, possedendo più ampie connessioni laterali.

I fotorecettori

Ci vollero molti anni per far progredire lo studio della fisiologia dei fotorecettori, delle cellule bipolari, orizzontali e amacrine. Le conoscenza sulle risposte alla luce dei coni e dei bastoncelli sono aumentate negli ultimi anni in modo impressionante. Coni e bastoncelli sono differenti sotto diversi aspetti:

- localizzazione nella retina
- sensibilità alla intensità radiante
- tipo di pigmento fotosensibile e quindi sensibilità spettrale

Abbiamo già descritto le differenze rispetto alla localizzazione nella retina, descriviamo ora la differenza più importante: la diversa sensibiltà alle varie lunghezze d'onda. I coni e i bastoncelli hanno forma diversa, i bastoncelli sono lunghi e sottili, i coni sono corti e appuntiti. Entrambi contengono un pigmento sensibile alla luce. I bastoncelli contengono tutti lo stesso pigmento; i coni sono di tre tipi e ogni tipo contiene un diverso pigmento. I quattro pigmenti sono sensibili a diverse lunghezze d'onda della luce e nel caso dei coni queste differenze formano la base per la visione dei colori. In Figura 4.4 sono rappresentate le curve di sensibilità spettrale dei tre tipi di coni (curve denotate da β, γ, ρ) e quella dei bastoncelli (curva tratteggiata).

I recettori rispondono alla luce attraverso un processo che si chiama *sbiancamento*. In questo processo una molecola di pigmento visivo assorbe un *fotone* e modifica la sua struttura chimica per trasformarsi in un altro composto che assorbe meno bene la luce. In quasi tutti gli animali, dagli insetti agli esseri umani e persino in alcuni batteri, questo pigmento

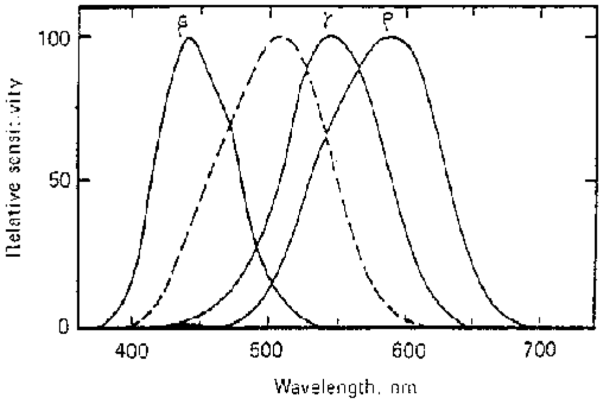


Figura 4.4 Curve di sensibilità spettrale.

è una proteina accoppiata ad una piccola molecola simile alla vitamina A. La maggior parte dei comuni recettori sensoriali, chimici, termici o meccanici, viene depolarizzata in risposta ad uno stimolo adeguato. La depolarizzazione porta al rilascio del trasmettitore nei terminali assonici. I fotorecettori degli invertebrati si comportano in questo modo e fino al 1964 si pensò che una risposta simile fosse presente nei coni e nei bastoncelli dei vertebrati. In quell'anno un neurofisiologo giapponese riuscì ad inserire un microelettrodo nei coni di un pesce e scoprì che al buio la differenza di potenziale ai capi della membrana era di 50 mV invece dei soliti 70 mV. Quando il cono veniva illuminato, il potenziale aumentava, la membrana si iperpolarizzava (proprio il contrario di ciò che tutti pensavano accadesse). Al buio, i fotorecettori dei vertebrati sono più depolarizzati delle altre cellule nervose e la depolarizzazione provoca un rilascio costante di trasmettitore ai terminali assonici, proprio come accade in un normale recettore durante la stimolazione. La luce, aumentando la differenza di potenziale a cavallo della membrana della cellula recettrice (cioè iperpolarizzandola), riduce il rilascio di questo trasmettitore. Dunque la stimolazine *spegne* il recettore. Questa scoperta aiuta a capire perché le fibre del nervo ottico dei vertebrati sono attive al buio: sono i recettori che sono spontaneamente attivi. Nei decenni successivi, il problema fu quello di capire come la sbiancamento anche di una sola molecola del pigmento visivo da parte di un singolo fotone provochi nel bastoncello una variazione misurabile del potenziale di membrana. I due processi sono ora abbastanza chiari. E' noto da tempo che un essere umano adattato al buio è in grado di vedere un flash di luce così debole che ogni recettore non può avere ricevuto più di un fotone. I calcoli dimostrano che almeno 6 bastoncelli vicini devono essere stimolati simultaneamente ciascuno da un fotone per produrre la sensazione allo stimolo. E' ora noto che un singolo fotone è in grado di eccitare un bastoncello in modo da fargli emettere un segnale significativo. Inoltre possiamo ora spiegare l'incapacità dei bastoncelli a rispondere a variazioni di luminosità quando esiste una intensa luce di fondo. Sembra che essi siano così sensibili che ad alti livelli di intensità luminosa, per esempio alla luce del sole, ogni ulteriore aumento non ha più effetto. Si dice che i bastoncelli sono saturati.

Cellule gangliari

Cosa accade dopo che lo stimolo luminoso ha colpito i fotorecettori? Vediamo come essi influenzano le cellule ganliari (trascuriamo bipolari orizzontali amacrine), ciò ci darà un'idea dell'elaborazione dell'informazione effettuata dalla retina. I risultati di cui parlo sono stati ottenuti negli anni 50 da Stefen Kuffler che ha studiato la risposta delle cellule gangliari della retina del gatto a punti luminosi. Egli riuscì ad illuminare la retina con una luce di fondo costante, debole e uniforme, ed anche a proiettare piccoli stimoli puntiformi osservando direttamente sia lo stimolo sia la risposta (extracellulare) di una singola cellula gangliare. Egli osservò che:

- Quando la luce diffusa era di intensità costante, o anche nella completa oscurità, la maggior parte delle cellule gangliari aveva una scarica di impulsi continua ma un poco irregolare, a frequenza tra 1-2 e 20 impulsi al secondo. Dato che ci si poteva aspettare che le cellule fossero silenti al buio completo, questa scarica fu una sorpresa.

- Spostando un punto luminoso sulla retina, egli riuscì ad individuare una regione, la cui stimolazione influenzava, aumentava o sopprimeva, l'attività di una cellula gangliare. Questa regione era il *campo recettivo* della cellula gangliare.

- Le cellule gangliari sono di 2 tipi, egli le chiamò *cellule centro-on* e *cellule centro-off*.

Una cellula *centro-on* scaricava a maggior frequenza quando lo stimolo luminoso cadeva in qualunque punto all'interno di un'area ben precisa, al centro o quasi del campo recettivo. Se si sposta anche di poco lo stimolo luminoso dal centro del campo recettivo, si osserva che la luce inibisce la scarica spontanea della cellula e, quando si toglie lo stimolo la cellula produce un piccolo treno di impulsi della durata di circa un secondo. Chiamiamo questa sequenza, l'inibizione durante l'illuminazione e la scarica al suo spegnimento, *risposta off*. L'esplorazione del campo recettivo dimostrò che esso è chiaramente suddiviso in una regione circolare *on* circondata da una regione *off* notevolmente più ampia e a forma di anello. Risposte massimali di tipo *on* si ottevano quando lo stimolo era un punto luminoso di dimensione adatta, e risposte massimali di tipo *off* si ottenevano quando lo stimolo aveva la forma di un anello luminoso di dimensioni adatte. Esempi di tali risposte sono illustrati in Figura 4.5.

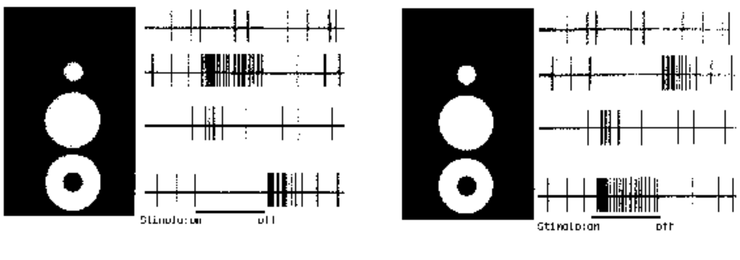


Figura 4.5 Risposte di cellule gangliari.

Le regioni del centro e della periferia interagivano in maniera antagonista: l'effetto di un punto luminoso al centro era diminuito dalla presenza di un secondo punto luminoso alla periferia. La dimostrazione più efficace di questa interazione centro-periferia si ha quando un fascio di luce colpisce l'intero campo recettivo: uno stimolo siffatto evoca una risposta molto più dedole rispetto a quella evocata da un punto luminoso proiettato solo nel centro; anzi in alcune cellule gli effetti delle due stimolazioni si cancellano. Se si usa un punto luminoso di dimensione progressivamente più grande, la risposta progressivamente migliora fino a quando non è riempito il centro del campo recettivo, quindi diminuisce tanto più quanto maggiore è la zona periferica colpita dallo stimolo. Una *cellula centro-off* ha un comportamento esattamente opposto. I due tipi di cellule sono mescolati tra loro e sembrano essere ugualmente rappresentati (in natura sono ugualmente rappresentati oggetti chiari su fondo scuro e viceversa). I segnali prodotti dalle cellule gangliari raggiungono, nel sistema nervoso, siti in cui vengono ulteriormente elaborate. Sono state individuate diverse stazioni e le loro connessioni. Una considerazione importante sul comportamento delle cellule gangliari è che esse sono sensibili all'illuminazione relativa di un'area rispetto allo sfondo. L'incapacità di ogni cellula di rispondere ad un evento che non implichi una differenza locale di intensità può apparire strana; noi però, proprio come le cellule gangliari, siamo molto più bravi a giudicare quale di due zone contigue appaia più chiara o più scura che a giudicare il livello generale di illuminazione. Questo sistema di misurare valori relativi e non assoluti ha, oltre al vantaggio dell'efficienza, il vantaggio di consentirci di vedere immutato l'aspetto degli oggetti in condizioni di illuminazione molto diverse quali quelle che vanno dalla luce del sole alla luce di una lampadina. Considerate questo esempio: una pagina di giornale ha lo stesso aspetto, caratteri neri su uno sfondo bianco, sia che la guardiamo in una stanza poco illuminata sia che la guardiamo sulla spiaggia in un giorno di sole. Supponete di misurare la radianza che arriva ai vostri occhi dalla carta bianca e da un carattere nero del titolo in queste due situazioni. Supponete per esempio che la radianza che vi giunge dalla carta bianca all'esterno sia 120 e 6 quando siete all'interno. Per il carattere nero supponete di avere 12 all'esterno e 0.6 all'interno. E' evidente che la potenza radiante all'esterno è 20 volte più forte di quella all'interno della stanza e che i caratteri neri riflettono circa un decimo della luce riflessa dalla carta. Le cifre però ci dicono anche che un carattere nero all'esterno riflette il doppio della luce riflessa dalla carta all'interno della stanza. E' chiaro che l'aspetto bianco o nero di un oggetto non è funzione della quantità di luce da esso riflessa. La cosa importante è la quantità di luce riflessa dall'oggetto in confronto a quella riflessa dagli oggetti circostanti.

Campo recettivo

L'area totale dello strato dei recettori che si connettono ad una sola cellula gangliare costituisce il *campo recettivo*. Il campo recettivo è pertanto la regione di retina la cui stimolazione luminosa modifica l'attività di una cellula gangliare. Questa organizzazione è valida in generale per tutta la retina, ma i dettagli delle connessioni variano molto tra la fovea, che è proprio in corrispondenza di ciò che stiamo guardando e la regione più esterna, la periferia, dove la visione diretta diventa grossolana. Nella fovea e nelle sue vicinanze la regola è che un singolo cono proietta ad una singola cellula bipolare, e una singola cellula bipolare ad una cellula gangliare. Allontanandosi dal centro, tuttavia, numerosi recettori convergono sulle cellule bipolari, e numerose cellule bipolari convergono sulle cellule gangliari. Questo alto grado di convergenza di gran parte della retina, permette di spiegare perché possa esistere un rapporto di 125 a 1 tra recettori e fibre del nervo ottico. Le dimensioni dei campi recettivi variano da una cellula gangliare all'altra. In particolare, i centri dei campi recettivi hanno dimensione diversa nelle diverse zone della retina: sono più piccoli nella fovea dove massima è l'acuità visiva, cioè la capacità di distinguere oggetti piccoli; diventano progressivamente più grandi via via che ci si allontana dalla fovea, corrispondentemente diminuisce l'acuità visiva. Possiamo descrivere la dimensione del campo recettivo in due modi. La descrizione più immediata è semplicemente la sua dimensione sulla retina; essa ha tuttavia lo svantaggio di non avere relazione con le dimensioni delle cose che ci circondano. Il modo per aggirare il problema è quello di esprimere la dimensione del campo recettivo come *l'angolo sotteso* dal campo recettivo. Un millimetro sulla retina di un uomo corrisponde ad un angolo visivo di circa 3,5 gradi. All' interno della fovea i coni hanno diametri di circa $2,5$ micron e tale è la distanza tra i centri dei campi recettivi delle cellule a cui essi sono connessi. Il valore di $2,5$ micron corrisponde bene alla nostra acuità visiva, misurata in termini di capacità di distinguere 2 punti alla distanza di $0,5$ minuti di arco. Un cerchio di $2,5$ micron di diametro sulla retina (che sottende ad un angolo di $0,5$ minuti di arco) corrisponde ad una moneta di $1/4$ di dollaro posta alla distanza di 150 metri circa dall'occhio. Più lontano, alla periferia della retina, i centri dei campi recettivi comprendono migliaia di recettori e possono avere diametri di $1°$ e più. Quindi, quando ci sposta sulla retina dal centro alla periferia, si osserva una non casuale correlazione fra 2 variabili: l'acuità visiva diminuisce e le dimensioni dei centri dei campi recettivi aumentano.

Bibliografia

- D. Hubel.: *Occhio, Cervello Visione*, Zanichelli.

Capitolo 5

Elementi di fotometria

5.1 Introduzione

Ora che abbiamo definito le quantità radiometriche di base, definiamo con precisione cosa intendiamo per luce. La *luce* è energia radiante valutata rispetto alla sua abilità di stimolare il senso della vista di un osservatore umano. E' energia radiante la cui lunghezza d'onda λ è compresa nell'ambito $380 - 780\ nm$.

Consideriamo 2 sorgenti di energia radiante, una il cui spettro è tutto compreso nell'intervallo visibile e l'altra il cui spettro è diverso da 0 solo al di fuori di tale intervallo. E' del tutto evidente che le quantità radiometriche che abbiamo definito possono essere misurate per entrambe le sorgenti e che nessuna di queste quantità tiene conto dell'effetto delle radiazioni sul sistema visivo. Dalla definizione di luce è chiaro che la sorgente di energia la cui distribuzione spettrale di potenza è diversa da 0 per $\lambda < 380$ e $\lambda > 780$ sarà invisibile per un osservatore umano. E' perciò necessario definire grandezze, corrispondenti alle grandezze radiometriche, che tengano conto delle caratteristiche del sistema visivo; esse sono il *flusso luminoso*, l' *intensità luminosa*, la *luminanza*, l' *illuminamento*. Per introdurre queste grandezze è necessario descrivere un esperimento di *psicofisica* denominato esperimento di *corrispondenza di brillanza*. Un esperimento di psicofisica consente di determinare la relazione esistente tra le caratteristiche fisiche di uno stimolo e gli eventi percettivi che esso determina.

5.2 Un interessante esperimento di psicofisica

Prima di descrivere l'esperimento di corrispondenza di brillanza, consideriamo un fondamentale esperimento di psicofisica effettuato da Hecht, Schlaer, Pirenne nel 1942. In tale esperimento vennero presentati a soggetti con visione normale *brevissimi stimoli di luce monocromatica* al fine di determinare la più bassa potenza per cui tali stimoli luminosi risultavano visibili. L'apparato sperimentale fu costruito in modo tale da permettere allo sperimentatore di presentare gli stimoli in una particolare localizzazione del campo visivo. La lunghezza d'onda, la dimensione, la durata degli stimoli e la localizzazione nel campo visivo vennero fissati a particolari valori, scelti mediante criteri che discuteremo tra breve, e una serie di stimoli di *diverse potenze* vennero presentati ad ogni soggetto sottoposto all'esperimento. Per ogni stimolo si chiedeva al soggetto in esame se lo aveva visto oppure no, e la potenza alla quale egli diceva di avere visto lo stimolo il 60% delle volte che gli veniva presentato venne definita *potenza soglia*. L'obiettivo dell'esperimento era di trovare la più bassa potenza che consentisse al soggetto di vedere lo stimolo. Il problema è che la soglia varia in funzione della lunghezza d'onda, della posizione dello stimolo nel campo visivo, della dimensione e della durata dello stimolo; questi parametri devono essere pertanto accuratamente definiti. Inoltre lo stato dell'occhio del soggetto influenza enormemente il valore soglia. E' ben noto che un soggetto è più sensibile a deboli stimoli luminosi se è stato al buio per un periodo di tempo piuttosto che se è stato alla luce fino al momento della presentazione dello stimolo. E' esperienza comune essere quasi ciechi nel momento in cui entriamo in un cinema buio ma cominciare a vedere bene dopo 5-10 minuti. Questo fenomeno è noto come *adattamento al buio*. La curva (a) in Figura 5.1 mostra i risultati di un esperimento sull'adattamento al buio.

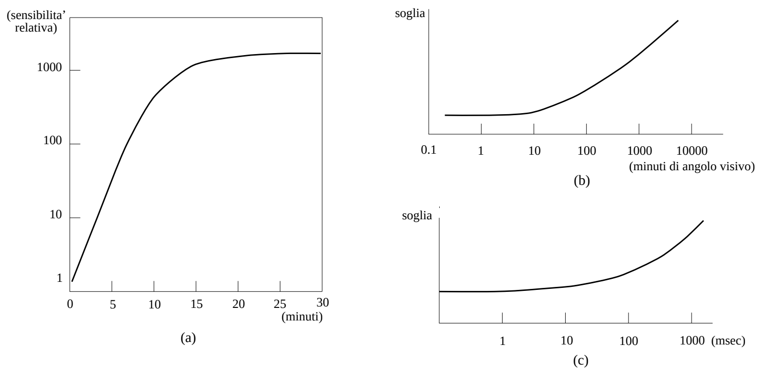


Figura 5.1 Curve per la determinazione dei parametri dell'esperimento di Hecht, Schlaer, Pirenne.

Come condizione sperimentale è stata pertanto scelta quella in cui i soggetti erano adattati al buio.

Per quanto riguarda la localizzazione dello stimolo luminoso nel campo visivo, poiché i bastoncelli rispondono a luci molto deboli, lo stimolo venne posto 20° a destra della fovea del soggetto dove si trovano solo bastoncelli.

Esperimenti sulla dimensione dello stimolo dimostrarono che se il diametro della macchia luminosa presentato al soggetto è minore di 10 minuti di arco ($10'$) la soglia non si modifica poi cresce (curva (b) in figura). Vennero pertanto scelti stimoli di diametro di $10'$ o inferiore. Per darvi un'idea di cosa ciò significhi, pensate che un oggetto (barra) alto 1 cm e posto a 57 cm da un osservatore è visto con un angolo visivo $\alpha = 1°$.

La durata dello stimolo fu scelta inferiore a 100 ms, poiché anche in questo caso si osservò che la soglia aumenta con la durata dello stimolo (curva (c) in Figura 5.1).

Per quanto riguarda la determinazione della lunghezza d'onda adeguata, fu effettuato il seguente esperimento: stimoli di luci monocromatiche furono presentati a soggetti completamente adattati al buio e la soglia per vedere lo stimolo fu misurata per ogni lunghezza d'onda. I risultati di un tale esperimento sono rappresentati dalla curva a tratto continuo in Figura 5.2.

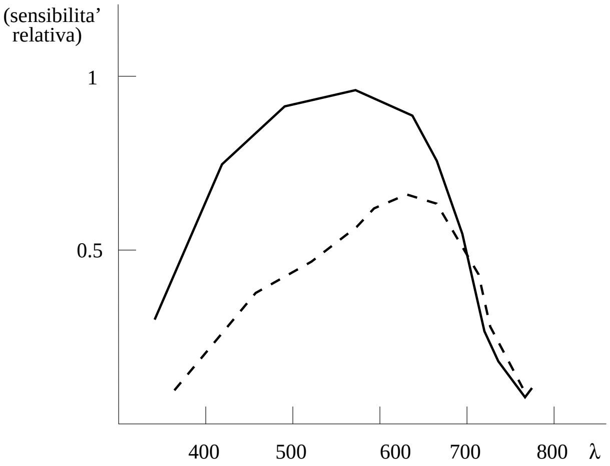


Figura 5.2 Soglia di percezione di stimoli monocromatici.

La curva indica che quando gli stimoli luminosi hanno lunghezza d'onda situata al centro dell'intervallo visibile sono necessari meno quanti di energia affinché lo stimolo risulti visibile; in particolare si ha la massima sensibilità per stimoli di lunghezza d'onda intorno ai 510 nm.

Fatte queste osservazioni le condizioni sperimentali standard furono così fissate: un soggetto adattato al buio, con un occhio coperto, viene accuratamente posizionato in un apparato dove egli guarda una debole luce posta nel punto di fissazione. La luce è sufficentemente brillante da essere vista, ma non brillante abbastanza da interferire con la possibilità di vedere la macchia luminosa che costituisce lo stimolo test nel momento in cui esso viene presentato ad un angolo di 20° dal punto di fissazione. Quando il soggetto è pronto preme un bottone e lo stimolo test avente le caratteristiche sopra descritte gli viene presentato. Egli quindi dice se ha visto oppure no lo stimolo test. Questa procedura è ripetuta molte volte per diverse potenze dello stimolo test. Molti soggetti furono sottoposti all'esperimento.

Gli esperimenti hanno consentito di determinare che gli stimoli visti dai soggetti il 60% delle volte in cui vennero presentati contenevano in media 90 quanti di energia. Dai dati sperimentali descritti fu possibile stimare che un singolo quanto di luce può essere sufficiente ad attivare un bastoncello e che gli effetti della attivazione di circa 10 bastoncelli vicini si sommano in modo da produrre la percezione di uno stimolo luminoso.

L'esperimento ha dimostrato che bastano pochi fotoni, di energia e di lunghezza d'onda opportuna, a stimolare i bastoncelli ma anche che luci monocromatiche di lunghezza d'onda diversa non sono ugualmente efficaci nello stimolare i bastoncelli.

Analogamente a quanto fatto per i bastoncelli può essere fatta una curva di sensibilità spettrale per i coni (curva tratteggiata in Figura 5.2). La curva di sensibilità fotopica rappresenta la sensibilità spettrale quando i coni sono completamente adattati al buio. Dal confronto tra le due curve di Figura 5.2 si può osservare che:

- i bastoncelli sono più sensibili dei coni a tutte le lunghezze d'onda ad eccezzione di quelle molto lunghe in cui la sensibilità spettrale è la stessa;

- se la forma delle due curve è circa la stessa, la curva dei coni è spostata di circa 50 nm a destra rispetto a quella dei bastoncelli, i bastoncelli sono massimamente sensibili a luce di lunghezza d'onda intorno ai 510 nm, i coni a 555 nm.

5.3 L'osservatore fotometrico standard

Fatta questa digressione introduciamo i concetti di *flusso luminoso*, di *intensità luminosa*, di *luminanza* e di *illuminamento*. Sono grandezze che richiedono, per essere introdotte, la descrizione di un esperimento di psicofisica che prende il nome di *esperimento di corrispondenza di brillanza* (matching of brightness) e la definizione di un *osservatore fotometrico standard*. Ricordiamo che la brillanza è un attributo della sensazione visiva secondo il quale un'area illuminata appare più o meno brillante; gli aggettivi che si usano per descriverla sono brillante, pallido.

L'esperimento consiste in questo: due macchie luminose, omogenee e della stessa forma, vengono presentate ad un osservatore, affiancate in una definita posizione nel campo visivo. L'osservatore deve essere in grado di decidere se esse appaiono ugualmente brillanti, (bright) risplendenti, vivide.

In generale un esperimento come questo viene detto di "corrispondenza visiva" in quanto l'obiettivo è di determinare quali coppie di stimoli fisici producono la stessa risposta percettiva. In esperimenti di "corrispondenza visiva" la forma più comune di uno stimolo test è una singola macchia di luce uniforme, priva cioè di struttura spaziale e temporale. La specificazione completa di uno stimolo test, in ogni parte del campo visivo e in qualsiasi istante, richiede di fissare:

1. la distribuzione spettrale assoluta della potenza radiante della superficie emittente;

2. le condizioni di osservazione.

Il postulato fondamentale della fotometria è il seguente:

qualunque siano le caratteristiche di una macchia luminosa (fissate la distribuzione spettrale assoluta della potenza radiante e le condizioni di osservazione), e qualunque sia la *distribuzione spettrale relativa* dello stimolo visivo corrispondente alla seconda macchia luminosa, è sempre possibile effettuare un "matching of brigtness" variando opportunamente la potenza radiante del secondo stimolo.

Effettuare un "matching of brigtness" significa riconoscere, nella sensazione visiva prodotta dai 2 stimoli luminosi, non la totale similarità ma la similarità di una qualità, la *brightness*, in presenza di differenze in altre qualità quali la tinta e la saturazione. Superficialmente può sembrare che effettuare una "corrispondenza di brillanza" sia più semplice che effettuare una una "corrispondenza di tricromatica" (vedremo che in tale esperimento vengono posti a confronto i colori di due stimoli luminosi), ma non è così. Infatti è estremamente difficile per l'osservatore stabilire se due stimoli, aventi diversa distribuzione spettrale relativa della potenza radiante, hanno la stessa brillanza nel caso in cui essi vengano presentati affiancati. Per questa ragione si osserva una grande variabiltà nei risultati di esperimenti successivi. Sono pertanto stati sviluppati metodi che conducono a risultati più precisi. Tali metodi prevedono che gli stimoli luminosi siano presentati alternativamente; la frequenza secondo cui vengono alternati è tale per cui l'osservatore non riesce a percepire le differenze di tinta tra i 2 stimoli, riesce tuttavia a percepire le differenze di brillanza.

Un'altra assunzione della fotometria è che, per un *osservatore standard*, la corrispondenza di brillanza soddisfa le seguenti leggi:

1. **Legge di simmetria**
   Se lo stimolo A corrisponde allo stimolo B, allora lo stimolo B corrisponde allo stimolo A.

2. **Legge di transitività**
   Se lo stimolo A corrisponde allo stimolo B e lo stimolo B corrisponde allo stimolo C, allora A corrisponde a C.

3. **Legge di proporzionalità**
   Se A corrisponde a B, allora αA corrisponde a αB, dove α è qualsiasi fattore positivo per cui la potenza radiante dello stimolo è aumentata o ridotta, mentre la distribuzione spettrale relativa è mantenuta uguale.

4. **Legge di addittività**
   Se A, B, C, D sono 4 stimoli qualsivoglia, allora se valgono due delle seguenti tre condizioni:

   - A corrisponde a B
   - C corrisponde a D
   - $(A + C)$ corrisponde a $(B + D)$

allora vale anche che la quarta cioè: $(A + D)$ corrisponde a $(B + C)$.

Con $(A+C)$, $(B+D)$, $(A+D)$, $(B+C)$ si denotano rispettivamente miscele addittive di stimoli visivi. Una miscela addittiva è uno stimolo luminoso per cui la potenza radiante in ogni intervallo di lunghezza d'onda, in ogni parte dello spettro, è uguale alla somma delle potenze dei costituenti della miscela nello stesso intervallo.

Condizione necessaria e sufficente affinché un processo di corrispondenza di brillanza obbedisca a queste leggi, è che la corrispondenza tra due stimoli visivi la cui distribuzione spettrale della potenza radiante sia rispettivamente $\{P(\lambda)d\lambda\}$ e $\{P^{\star}(\lambda)d\lambda\}$ abbia questa forma:

$$\int_{\lambda} P(\lambda)f(\lambda)d\lambda = \int_{\lambda} P^{\star}(\lambda)f(\lambda)d\lambda$$

dove $f(\lambda)$ è una funzione della lunghezza d'onda che caratterizza il processo di corrispondenza di brillanza.

Le corrispondenze di brillanza, fissato il metodo e le condizioni di osservazione, effettuate da osservatori reali deviano in qualche misura dalle leggi sopra definite. Per esempio se la porzione di campo visivo in cui viene effettuato il processo di corrispondenza supera un angolo visivo di $2°$ non viene rispettata la legge di proporzionalità quando il valore assoluto degli stimoli è ridotto da moderato a molto basso. Il non rispetto della legge di proporzionalità deriva soprattutto dal fatto che, cambiando il valore assoluto dello stimolo, si passa dalla curva di sensibilità spettrale dei coni alla curva di sensibilità spettrale dei bastoncelli.

Le leggi sopra descritte valgono, almeno in modo approssimato, se vengono imposte limitazioni agli stimoli visivi. Per esempio se tutte le corrispondenze per le quali si vuole verificare la validità delle leggi vengono effettuate in *condizioni fotopiche* (cioè gli stimoli sottendono ad angoli visivi di $2° - 3°$) allora esse valgono ad un sufficiente livello di approssimazione.

D'altra parte, le corrispondenze scotopiche sono quelle prevalentemente determinate dai bastoncelli e le leggi sopra esposte sono approssimativamente vere se tutte le corrispondenze vengono effettuate in *condizioni scotopiche*. In tal caso $f(\lambda)$ è una funzione diversa che caratterizza le corrispondenze in condizioni scotopiche.

Il fatto che un esperimento di corrispondenza sia scotopico o fotopico dipende:

- dalla distribuzione spettrale (assoluta e relativa) della potenza radiante degli stimoli visivi che devono essere posti in corrispondenza

- dalle condizioni di osservazione quali: il metodo di presentazione degli stimoli, la dimensione di campo visivo occupato dagli stimoli, l'eccentricità degli stimoli nel campo visivo, lo stato di adattamento del soggetto (determinato dalla precedente presentazione di stimoli ambientali).

La CIE ha definito due osservatori fotometrici ideali: l' *osservatore fotopico* e l' *osservatore scotopico*. Per definizione, le operazioni di corrispondenza di brillanza effettuate da tali osservatori obbedisono strettamente alle leggi sopra enunciate.

L' *osservatore fotopico standard* è stato definito dalla CIE nel 1924 mediante la funzione $V(\lambda)$, funzione di sensibilità spettrale in condizioni di visione fotopica (vedi Figura 5.3).

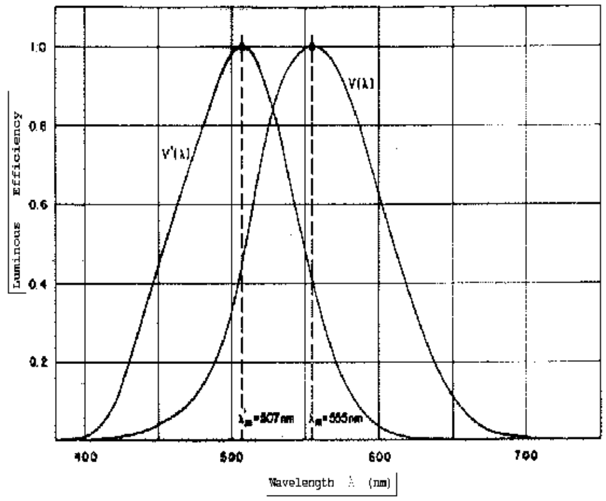


Figura 5.3 Funzioni di sensibilità spettrale.

La funzione $V(\lambda)$ è stata derivata facendo la media dei risultati di esperimenti per la determinazione della *efficacia luminosa relativa* di stimoli spettrali usando numerosi osservatori. Gli esperimenti sono stati effettuati con macchie luminose che occupano campi visivi di $2° - 3°$.

L' *osservatore scotopico standard* è stato definito dalla CIE nel 1951 mediante la funzione $V'(\lambda)$, funzione di sensibilità spettrale in condizioni di visione scotopica (vedi Figura 5.3).

La funzione $V(\lambda)$ è data in forma tabellare nell'intervallo di lunghezze d'onda $360 < \lambda < 830$ ad intervalli di 1 nm . La funzione $V'(\lambda)$ è data in forma tabellare nell'intervallo di lunghezze d'onda $380 < \lambda < 730$ ad intervalli di 10 nm. Le funzioni $V(\lambda)$ e $V'(\lambda)$ sono funzioni *relative* della lunghezza d'onda, in altre parole sono funzioni normalizzate in modo da avere come massimo valore l'unità. Il valore massimo di $V(\lambda)$ si ha per $\lambda = 555$ nm, il valore massimo di $V'(\lambda)$ si ha per $\lambda = 507$ nm. Le due funzioni sono pertanto funzioni peso, non hanno dimensione.

L'osservatore fotopico standard è molto più importante dell'osservatore scotopico perché tutte le misure fotometriche di flusso luminoso, intensità luminosa, luminanza, illuminamento, si basano sull'osservatore fotopico.

Macchie di luce uniforme, prodotte da stimoli visivi definiti dalle distribuzioni spettrali assolute di potenza radiante $\{P(\lambda)d\lambda\}$ e $\{P^*(\lambda)d\lambda\}$, sono corrispondenti in brillanza per l'osservatore fotopico standard se e solo se:

$$\int_\lambda P(\lambda)V(\lambda)d\lambda = \int_\lambda P^*(\lambda)V(\lambda)d\lambda$$

In parole: le somme delle potenze radianti $P(\lambda)d\lambda$, opportunamente pesate mediante la funzione $V(\lambda)$ che caratterizza l'osservatore fotopico standard, sono uguali. Questa condizione si esprime anche dicendo che il *flusso luminoso* fotopico che è percepito dal sistema visivo per unità di angolo solido deve essere lo stesso per entrambe le macchie luminose, cioè denotando rispettivamente con P_v e $P_v^\star$ i due flussi luminosi vale: $P_v = P_v^\star$. Per poter asserire che due macchie luminose emettono lo stesso flusso luminoso se corrispondono in brillanza, è necessario definire il flusso luminoso corrispondente ad una data distribuzione spettrale di flusso radiante $\{P(\lambda)d\lambda\}$ come una funzione monotona crescente di $\int_\lambda P(\lambda)V(\lambda)d\lambda$. E' stata scelta la funzione più semplice e il flusso luminoso fotopico P_v è pertanto così definito:

$$P_v = K_m \int_\lambda P(\lambda)V(\lambda)d\lambda$$

Analogamente il flusso luminoso scotopico P_v' è così definito:

$$P_v' = K_m' \int_\lambda P(\lambda)V'(\lambda)d\lambda$$

Alla costante K_m poteva essere assegnato un arbitrario valore numerico, per esempio l'unità. Però per ragioni storiche e pratiche i fotometristi hanno scelto di porre $K_m = 683$ e $K_m' = 1700$.
Una volta definito il flusso luminoso, sono definite tutte le altre grandezze fotometriche. Denotiamo con l'indice v le grandezze fotometriche, sono prive di indice le corrispondenti grandezze radiometriche.

Le grandezze fotometriche fotopiche, la cui misura è più frequentemente richiesta nella pratica, sono il flusso luminoso P_v, l'intensità luminosa I_v, la luminanza L_v e l'illuminamento E_v. Queste grandezze sono definite da equazioni simili:

$$P_v = K_m \int P(\lambda)V(\lambda)d\lambda$$

$$I_v = K_m \int I(\lambda)V(\lambda)d\lambda$$

$$L_v = K_m \int L(\lambda)V(\lambda)d\lambda$$

$$E_v = K_m \int E(\lambda)V(\lambda)d\lambda$$

In queste equazioni le quantità radiometriche $P(\lambda)$, $I(\lambda)$, $L(\lambda)$, $E(\lambda)$ sono rispettivamente le densità spettrali del flusso radiante, dell'intensità radiante, della radianza e dell'illuminamento.

In fotometria, un metodo di base per determinare una quantità fotometrica (es. la luminanza) è quello di misurare la densità spettrale della quantità radiometrica appropriata (la radianza) e poi calcolare la quantità fotometrica desiderata utilizzando l'equazione appropriata che utilizza la funzione $V(\lambda)$.

Le unità di misura standard delle grandezze fotometriche sopra definite sono:

- il *lumen* (lm) per il flusso luminoso (il watt è l'unità di misura del flusso radiante)
- la *candela* (cd) per la intensità luminosa (il watt per steradian è l'unità di misura della intensità radiante),
- *la candela per metro quadrato* ($cd \cdot m^{-2}$) per la luminanza (il watt per steradian per metro quadrato è l'unità di misura della radianza)
- il *lumen per metro quadrato* ($lm \cdot m^{-2}$) per l'illuminamento (il watt per metro quadrato è l'unità di misura dell' irradiamento).

Il *lumen* è il flusso luminoso di energia radiante monocromatica il cui flusso radiante è $1/683\ W$ e la cui frequenza è $540 \cdot 10^{12} Hz$. Questa frequenza corrisponde ad una lunghezza d'onda $\lambda_d = 555.06\ nm$, nell'aria standard.

La *candela* è l' intensità luminosa, in una data direzione, di una sorgente che emette energia radiante monocromatica di frequenza $540 \cdot 10^{12} Hz$ e la cui intensità radiante in quella direzione è $1/683\ W$ per steradian.

Queste definizioni valgono sia per le quantità fotopiche che per le quantità scotopiche.

Abbiamo detto che le funzioni $V(\lambda)$ e $V'(\lambda)$ sono funzioni *relative* della lunghezza d'onda, in altre parole sono funzioni normalizzate in modo da avere come massimo valore l'unità. La dimensione delle quantità fotometriche è introdotta mediante le equazioni di base della fotometria che relano il flusso luminoso (fotopico P_v e scotopico P'_v) al flusso radiante P:

$$P_v = K_m \int_\lambda P(\lambda)V(\lambda)d\lambda$$

$$P'_v = K'_m \int_\lambda P(\lambda)V'(\lambda)d\lambda$$

Se definiamo:

$$P_v(\lambda)d\lambda =_{def} K_m P(\lambda)V(\lambda)d\lambda$$

possiamo scrivere:

$$\frac{P_v(\lambda)}{P(\lambda)} = K_m V(\lambda)$$

e ponendo:

$$K(\lambda) = K_m V(\lambda)$$

definiamo l' *efficacia luminosa fotopica* (misurata in *lumen per watt*) del flusso radiante monocromatico $P(\lambda)$ di lunghezza d'onda λ.

In modo analogo definiamo l' *efficacia luminosa scotopica* $K'(\lambda) = K'_m V'(\lambda)$.

La Figura 5.4 illustra le curve $K(\lambda)$ e $K'(\lambda)$. Esse sono ottenute tenendo presente che, per definizione, l' efficacia luminosa del flusso radiante di lunghezza d'onda $\lambda_d = 555.06\ nm$ è uguale a $683\ lm \cdot W^{-1}$ sia per la visone fotopica che per la visione scotopica, ciò implica che:

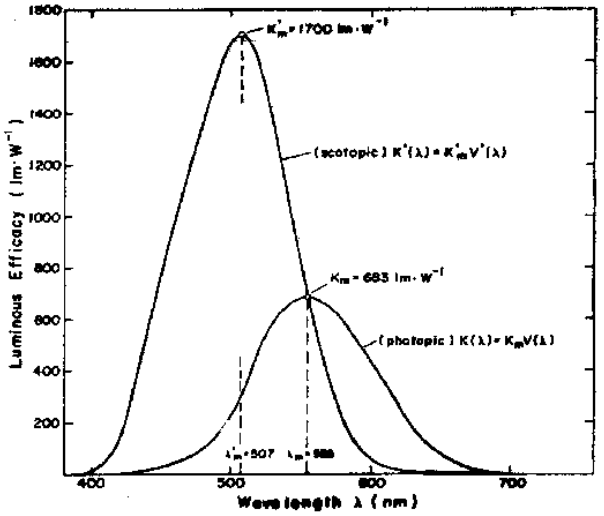


Figura 5.4 Le funzioni di efficacia luminosa.

$$K(\lambda_d) = K'(\lambda_d) = 683\ lm \cdot W^{-1}$$

Alla lunghezza d'onda $\lambda_d = 555.06\ nm$, le efficienze fotopica e scotopica sono rispettivamente:

$$V(\lambda_d) = 0.999998 \qquad V'(\lambda_d) = 0.401750$$

Dalle equazioni sopra segue:

$$K_m = \frac{K(\lambda_d)}{V(\lambda_d)} = \frac{683}{0.999998} = 683.002\ lm \cdot W^{-1}$$

$$K'_m = \frac{K'(\lambda_d)}{V'(\lambda_d)} = \frac{683}{0.401750} = 1700.06\ lm \cdot W^{-1}$$

cioè praticamente 683 e 1700.

5.4 Leggi psicofisiche della sensazione visiva

Analizziamo ora altre caratteristiche di base della percezione visiva che sono state determinate mediante esperimenti di psicofisica.

5.4.1 Adattamento alla luce

Abbiamo parlato dell' adattamento al buio del sistema visivo, esiste però anche l'adattamento alla luce. Per analizzare come varia la sensibilità del sistema con la luminanza dell'ambiente, è necessario misurare la variazione della *soglia differenziale* dL in funzione della luminanza di base L e del tempo in cui il soggetto è stato esposto alla fissata condizione di illuminazione. In altre parole si valuta quanto deve essere luminoso un secondo stimolo per essere distinto da uno stimolo dato.

Il fatto che il contrasto visivo tra oggetti resti costante indipendentemente dalle condizioni di illuminazione (pensate per esempio a lettere nere su foglio bianco) è ben spiegato dall'andamento della soglia differenziale in funzione dell'illuminazione.

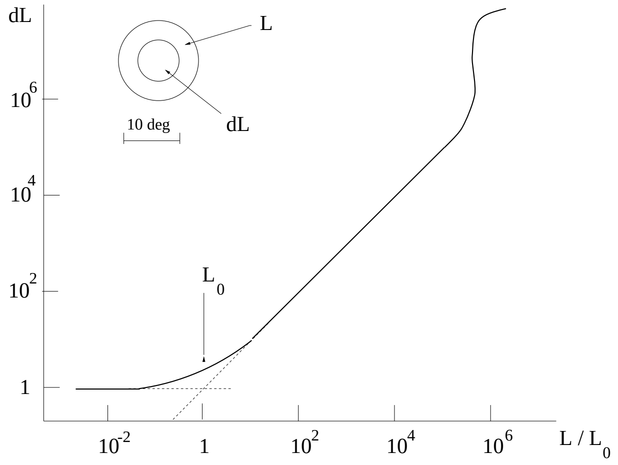


Figura 5.5 Effetto della illuminazione dello sfondo sulla percezione dello stimolo sovrapposto.

La Figura 5.5 mostra i risultati di un esperimento in cui, per varie condizioni di illuminazione di base L, l'introduzione in un campo visivo di 20° di uno stimolo test di 9°, è studiato in termini dell'incremento liminale di luminanza dL necessario a riconoscere lo stimolo test dallo sfondo; la durata dello stimolo test è stata fissata a 0.2 secondi. L'incremento liminale di luminanza dL è rappresentato come funzione della luminanza di base, in coordinate log/log.

La curva mostra un tratto costante in cui la soglia differenziale dL non dipende dalla luminanza L e poi un tratto di crescita lineare. Tale tratto lineare ha pendenza 1 se l'ascissa è espressa come L/L_0 dove L_0 è il punto di intersezione tra il tratto costante e la retta che approssima il tratto di crescita lineare. La relazione $\log(dL) = \log(1 + L/L_0)$ è nota come legge di Weber.

5.4.2 Contrasto simultaneo

La psicofisica ha studiato anche una varietà di fenomeni complicati che caratterizzano la capacità del sistema visivo di discriminare *stimoli di diversa luminanza presentati simultaneamente*. Tradizionalmente, questi fenomeni sono stati analizzati mediante esperimenti di *contrasto simultaneo* (in essi si analizza l'interazione tra stimoli vicini di diversa luminanza) o di *acuità visiva* (in essi si analizza la capacità di separare immagini).

Consideriamo il caso più semplice di contrasto simultaneo: date 2 zone adiacenti a bassa ed alta luminanza, L_{min} e L_{max} rispettivamente, per definizione il contrasto C è dato da:

$$C = \frac{L_{max} - L_{min}}{L_{min} + L_{max}}$$

o, in termini di luminanza media $L_m = 1/2(L_{min} + L_{max})$:

$$C = (L_{max} - L_{min})/2L_m$$

Il contrasto varia pertanto tra 0 e 1.

Uno dei più noti effetti di contrasto è quello per cui una zona di data luminanza appare più (meno) brillante se posta vicino ad una zona che ha minore (maggiore) luminanza. Tra i parametri che oltre al contrasto contribuiscono a determinare la brillanza, importanti sono le relazioni spaziali tra gli stimoli visivi.

Il più semplice e forse più famoso esempio è costituito dalle bande di Mach. Se si esamina il bordo tra una zona scura ed un zona chiara si osserva che vicino al bordo la differenza di luminanza appare più accentuata; in particolare si individua nella zona più brillante una striscia parallela al bordo ancora più brillante e nella zona scura una striscia parallela al bordo ancora più scura.

Più complesse illusioni dovute a rapide variazioni di luminanza si hanno osservando l' Hermann's grating che è costituito da una serie di quadrati neri separati da striscie bianche. Le zone di intersezione delle striscie, sebbene bianche, appaiono grigie (Figura 5.6).

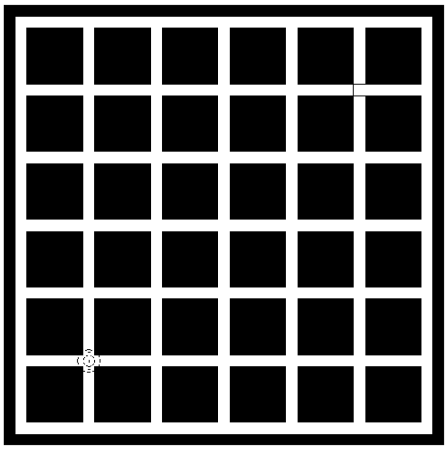


Figura 5.6 Hermann's grating.

L'*acuità visiva* è definita come la capacità di discriminare i dettagli fini di un oggetto nel campo visivo. In linea di principio, la *distanza appena percepibile* tra 2 punti, o altri dettagli di un oggetto può essere espressa in molti modi: la distanza assoluta tra i 2 punti, l'angolo sotteso, la loro distanza sulla retina. In pratica l'acuità si misura come il reciproco dell'angolo sotteso, misurato in minuti d'arco. L'acuità visiva può essere misurata in diversi modi, tra questi ricordiamo la capacità di distinguere un punto nero su uno sfondo bianco, la capacità di riconoscere oggetti (lettere) a partire dalla forma, la separazione tra linee parallele.

Sono stati effettuati studi sul modo di operare del sistema visivo analizzando la risposta di soggetti a stimoli di diverse *frequenze spaziali*. Ogni stimolo è costituito da strisce scure e chiare che si alternano, gli stimoli differiscono per il contrasto tra le striscie e la frequenza spaziale (numero di alternanze di luminanza per ogni grado di angolo visivo). Le variazioni di luminanza possono avere la forma di onde quadre o di sinusoidi, naturalmente il contrasto spaziale cambia bruscamente nel primo caso e gradualmente nel secondo. I parametri che caratterizzano uno stimolo sono pertanto:

1. la frequenza spaziale
2. la luminanza media
3. il contrasto
4. la forma dell'onda
5. l'orientazione

Ai soggetti si richiede di distinguere le striscie dallo sfondo, ciò porta a definire una soglia per la percezione del contrasto (minimo contrasto percepibile). Si è osservato che la sensibilità al contrasto è funzione sia della frequenza spaziale che della luminanza media.

Bibliografia

- T. N. Cornsweet.: *Visual perception*, Harcourt Brace Jovanovich, Publisher, 1970.
- P. Buser, M. Imbert.: *Vision*, MIT press 1992.
- G. Wyszecki, W. S. Stiles.: *Color Science. Concept and Methods, Quantitative Data and Formulae* - Second Edition -, Jhon Wiley & Sons, Toronto 1982.

Capitolo 6

Elementi di colorimetria

6.1 Che cosa è il colore?

La parola colore, come è usata nel linguaggio di tutti i giorni, ha molti significati diversi. Consideriamo l'uso quotidiano della parola colore, un uso che implica che il concetto di colore sia una *proprietà dei materiali.* Diciamo che un pomodoro maturo ha la proprietà di essere rosso, che la neve ha la proprietà di essere bianca... Accettiamo che il colore di un materiale o di un oggetto sia il colore che noi percepiamo alla luce del giorno.

Parliamo anche del colore della luce e comunemente consideriamo che il colore sia una proprietà della luce. Quando guardiamo un semaforo rosso supponiamo che una luce di colore rosso stia irradiando i nostri occhi.

A causa di questo uso comune della parola colore, è spesso una sorpresa per molti scoprire che il colore percepito è scorrettamente ritenuto una proprietà dei materiali e una proprietà della luce.

Newton discutendo sulla luce nel suo libro Opticks (1704) asseriva correttamente "invero, i raggi, propriamente espressi, non sono colorati". Che cosa è allora il colore?

Lo psicologo L.M. Hurvich si chiede se un oggetto ha un colore a causa delle sue proprietà chimico-fisiche o a causa della illuminazione; si chiede inoltre se il colore è un evento fotochimico nella retina, un processo di eccitazione nel sistema nervoso o un evento fisico. Il colore è tutte queste cose insieme. Il punto più importante che deve essere chiarito è che "la nostra percezione del colore deriva dalla interazione tra i raggi di luce e il nostro sistema visivo. Entrambi sono coinvolti nel vedere gli oggetti e nel percepire i colori" (Figura 6.1).

Vediamo prima la parte percettiva. Quando guardiamo un'area uniformemente colorata

Figura 6.1 Concetto di colore.

di un dipinto, il colore che percepiamo è spesso influenzato dai colori delle aree circostanti. Per questo se vogliamo specificare in modo preciso il colore di una superficie dipinta, dobbiamo considerarlo isolato, e metterci pertanto in una situazione standard, cioè osservare il colore rispetto ad uno sfondo bianco o grigio.

Quando vediamo un semaforo rosso, di notte in assenza di altre luci, stiamo facendo un' esperienza di *colore isolato*, cioè di un colore non relato ad altri.

I colori percepiti sono descritti da almeno 5 attributi. Nel caso più semplice, quello in cui si considerano colori isolati, il colore è normalmente descritto mediante 3 attributi: *tinta (hue)*, *saturazione (saturation)*, *brillanza (brightness)*.

Quando guardiamo ad una luce *rossa* noi percepiamo la *tinta*. E' difficile spiegare cos' è la percezione del rosso, proprio come è difficile dire cos' è la percezione dell'amaro. Possiamo solo dire che quando diciamo le parole rosso, blu, giallo... noi stiamo comunicando agli altri l'idea di una particolare tinta. Rispetto alla *tinta* distinguiamo i *colori cromatici*, quelli che hanno una tinta, e i *colori acromatici*, quelli che non ce l'hanno. Per esempio noi percepiamo un colore acromatico quando guardiamo ad una lampadina fluorescente, o quando guardiamo superfici bianche, grigie, nere. E' stato osservato che, tra tutte le tinte, solo 4 non sono percepite come miscele. Esse sono chiamate *tinte unitarie*, sono il *rosso*, il *giallo*, il *verde*, e il *blu*. Altre sono considerate tinte binarie perché sono percepite come miscele delle seguenti coppie: *verde unitario e giallo unitario* (verdi giallastri, gialli verdastri), *rosso unitario e giallo unitario* (gialli rossastri e arancioni), *rosso unitario e blu unitario* (magenta, porpora, violetto,) *verde unitario e blu unitario* (blu verdastri e verdi bluastri).

La *saturazione* è un attributo del colore percepito secondo il quale noi giudichiamo, nel colore, l'ammontare relativo della tinta componente. Per chiarire il concetto consideriamo due fasci di luce isolati uno rosso e l'altro rosa; i fasci di luce sono tali da evocare in noi la stessa percezione di brillanza e di tinta. Il rosa ha tuttavia una saturazione minore perché l'ammontare relativo della componente rossa è minore. Non c'è unanimità tra gli studiosi del colore sulla definizione precisa di saturazione. La definizione data sopra è quella usata dai più.

Hunt ha coniato il termine *colorfulness* (pienezza di colore) per denotare ciò che un osservatore giudica essere l'ammontare assoluto di tinta, ciò indipendentemente dall'ammontare della componente acromatica.

Vediamo di chiarire con un esempio la differenza tra saturazione e pienezza di colore (colorfulness). Un pezzo di carta rossa illuminato da 2 fasci di luce identici in una stanza scura ha meno colorfulness se si spegne una delle luci; si percepisce meno rosso. La saturazione, tuttavia, non è modificata; cioè l'ammonatre di rosso rispetto all'ammontare totale di rosso e di contenuto acromatico non è modificato. In modo analogo un vestito rosso, se visto in un ambiente poco illuminato, appare meno colorato che se visto in pieno sole. L'attributo di colorfulness dipende pertanto dalla illuminazione.

La *brillanza* è un attributo dell'illuminazione in cui è visto un colore isolato. La brillanza aumenta quando l'illuminazione aumenta. Detto più precisamente la brillanza è "la percezione del livello generale di luminanza". Ci si può riferire alla brillanza del colore percepito solo quando il colore è isolato e la luce che giunge all'occhio proviene solo da quel colore. Per esempio si parla di brillanza del colore della luce di una lampada. La esperienza visiva della brillanza è comunemente descritta, ai limiti del suo ambito di variazione, con le parole "debole" (dim) "abbagliante" (dazzling).

La *lightness* percepita è un attributo dei colori non isolati (related colors). La lightness è la brillanza di un'area giudicata relativamente alla brillanza di un'area illuminata allo stesso modo, che appare bianca. Quando, effettuando una comparazione tra colori, ci si riferisce alla lightness si usano i termini "più chiaro" o " più scuro" .

Nella percezione del colore i termini *colore*, *tinta*, *saturazione*, *brightness*, *lightness*, *rosso*, *acromatico...* sono adeguati per descrivere la risposta ad uno stimolo. Usati in questo senso sono termini *psicologici*.

Nel momento in cui vengono fissati *metodi per la misura e la specificazione del colore*, il colore è connesso allo stimolo luminoso piuttosto che alla risposta percettiva perché misure precise possono essere fatte solo sullo stimolo luminoso. Per questa ragione viene adottata una definizione del colore che è psicofisica.

Che cosa si percepisce quando si vede una luce di una singola lunghezza d'onda? La risposta è verde se la lunghezza d'onda è 500 nm, blu se è 470 nm, rosso aranciato se è 600 nm. L'elenco che segue descrive alcune tinte che si percepiscono osservando luci di lunghezze d'onda comprese in tutto l'ambito visibile. Le tinte cambiano gradualmente se si aumenta in modo continuo la lunghezza d'onda da 380 nm a 780 nm; perciò per esempio il blu-verdastro che si percepisce osservando una luce monocromatica di lunghezza d'onda pari a 486 nm è più verde del blu-verdastro che si percepisce osservando una luce monocromatica di lunghezza d'onda pari a 483 nm.

Colori associabili a luci monocromatiche (colori dello spettro):
porpora bluastro: 380 nm
blu porpora: 430 nm
blu violetto: 465 nm
blu: 482 nm
blu verdastro: 487 nm
blu verde: 493 nm

verde bluastro: 498 nm
verde: 530 nm
verde giallastro: 558 nm
giallo verde: 570 nm
giallo verdastro: 575 nm
giallo: 580 nm
giallo arancio: 586 nm
arancio: 596 nm
arancio rossastro: 620 nm
rosso: 680 nm

I colori prodotti da luci monocromatiche hanno la saturazione massima. La gran parte delle luci di cui abbiamo esperienza di colore non sono monocromatiche.

6.2 Miscele di colori

Il termine *miscela di colori*, sebbene di uso comune, è inaccurato perché sono gli stimoli luminosi ad essere miscelati, non le risposte percettive. Il termine "sintesi di stimoli colore" sarebbe più appropriato.

Una *miscela di colori addittiva* si ha quando luci di diverse lunghezze d'onda sono combinate insieme prima di raggiungere l'occhio. Un modo utile per provare ciò è proiettare due raggi di luce di diversa lunghezza d'onda su una parete bianca in modo che i due dischi di luce prodotti da ognuno dei fasci si sovrappongano e la combinazione delle due luci sia riflessa dalla parete. Poiché le due luci sono sommate insieme, l'energia della combinazione è uguale alla somma delle energie dei due fasci iniziali. Di solito ciò che appare è un aumento della brillanza; se le tinte dei due fasci sono diverse la tinta del fascio che si ottiene per somma dei due sarà percepita come intermedia alle due tinte. Perciò se per esempio le tinte sono verde e rosso la tinta della combinazione sarà giallo-verde o giallo-arancio in funzione delle potenze relative dei due fasci. Se, per contro, le tinte sono sufficientemente diverse per esempio rosso e blu-verde allora, mediante un appropriato aggiustamento delle potenze, è possibile produrre una risposta priva di tinta (bianco). In questo caso i due colori originali sono detti *colori complementari additivi.*

Una *miscela di colori sottrattiva* si ha quando da un fascio di luce si rimuovono componenti a diverse lunghezze d'onda, mediante un processo di assorbimento ad opera di filtri o di pigmenti; il risultato è che il fascio appare avere una diversa tinta. Per esempio facendo passere un fascio di luce solare attraverso due vetri colorati, uno giallo e l'altro verde, posti in succesione, si produce una luce la cui tinta appare giallo-verde. Quando la luce solare passa attraverso il vetro giallo, la luce di lunghezze d'onda brevi (corrispondenti ai blu) viene assorbita dal vetro e la gran parte delle lunghezze d'onda intermedie (corrispondenti ai giallo-verdi) e lunghe (corrispondenti agli arancio-rossi) sono trasmesse. D'altro lato, la luce verde prodotta facendo passare luce solare attraverso un vetro verde ha una composizione in lunghezze d'onda che producono colori blu-verde, verde-giallo, giallo-verde e la gran parte dei blu e dei rossi è assorbita. Da ciò è chiaro che quando il fascio attraversa entrambi i vetri resta solo la luce giallo-verde. Poiché nel processo di assorbimento viene assorbita energia, la potenza della luce emergente da una miscela sottrattiva è diminuita.

Una miscela di colori sottrattiva si può ottenere anche facendo passare un fascio di luce attraverso una soluzione di pigmenti.

Abbiamo visto che si ha una miscela addittiva combinando le luci di diverse sorgenti prima che esse raggiungano l'occhio; anche il sistema visivo però realizza combinazioni. Quando la retina è colpita da luci riflesse (o emesse) da piccole superfici (sorgenti) tra loro molto vicine si parla di *media spaziale (spatial averaging)*. Per esempio una pagina stampata a caratteri neri su fondo bianco può apparire grigia, se vista da distanza opportuna (il grigio è il risultato della combinazione spaziale). Si ha lo stesso fenomeno coi pittori puntinisti, nell'halftone printing, nei terminali video.

Il sistema visivo è anche in grado di effettuare una *media temporale (temporal averaging)* degli stimoli luminosi, ciò accade quando brevi stimoli vengono presentati in rapida successione.

6.3 Specificazione del colore ad opera della CIE

La *colorimetria*, disciplina che ha per oggetto la misura del colore e delle differenze tra colori, appartiene al dominio della psicofisica.

La parola *colore* in psicofisica denota una caratteristica dello *stimolo luminoso* e tiene conto sia dell'energia radiante che raggiunge l'occhio sia di un *osservatore standard* che ha una visione normale dei colori. Il Comitato sulla Colorimetria della Società Americana di Ottica, che ha adottato la definizione psicofisica del colore, così commenta: "questa scelta sembra ampiamente giustificata su un terreno puramente filosofico, ma se si cerca una giustificazione meno accademica, considerazioni esclusivamente pratiche sono più che sufficenti a motivarla".

La colorimetria si fonda su leggi che si derivano da esperimenti di psicofisica, esperimenti che consentono di determinare la relazione esistente fra le caratteristiche fisiche (distribuzione spettrale della potenza radiante) dello stimolo luminoso e la percezione del colore che lo stimolo evoca nell'osservatore. Obiettivo della colorimetria è la specifica numerica del colore associato ad un preciso stimolo colore in modo tale che:

- due stimoli colore che hanno la stessa specifica numerica appaiono uguali quando sono visti, sotto le stesse condizioni di osservazione, da un osservatore che ha una visione normale dei colori.

- stimoli che appaiono uguali hanno la stessa specifica.

Le leggi sperimentali su cui si basa la colorimetria prendono il nome di *generalizzazione tricromatica*. Prima di enunciarle descriviamo l'esperimento da cui esse sono derivate.

6.3.1 Esperimento di corrispondenza tricromatica

Considerate l'assetto sperimentale rappresentato in Figura 6.2; esperimenti riconducibili a questo sono stati eseguiti da diversi gruppi di ricercatori alla fine degli anni 20.

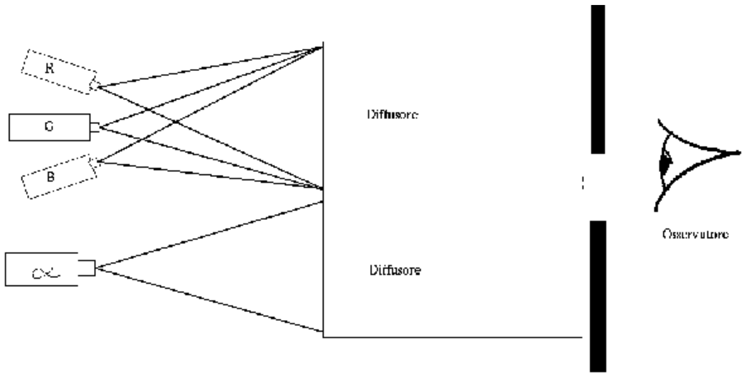


Figura 6.2 Esperimento di corrispondenza tricromatica.

Si fissano 3 *stimoli primari*, cioè luci monocromatiche la cui lunghezza d'onda viene opportunamente scelta. Si proietta uno *stimolo test* in una metà del campo visivo dell'osservatore, nell'altra metà del campo visivo viene proiettata una miscela additiva dei 3 primari. L'ammontare rispettivo dei 3 stimoli primari è aggiustato fino a quando, secondo l'osservatore, la miscela additiva corrisponde allo stimolo test per quanto riguarda tinta, saturazione e brillanza.

Affinché l'esperimento di corrispondenza tricromatica costituisca una base adeguata per costruire un sistema di misura del colore è necessario fissare con precisione i seguenti elementi dell'esperimento:

- l'angolo visivo (il risultato dell'esperimento cambia in funzione dell'angolo visivo)
- lo stato di adattamento dell'occhio dell'osservatore
- la lunghezza d'onda dei 3 primari

I risultati ottenuti da esperimenti effettuati da diversi osservatori posti nelle medesime condizioni di osservazione, hanno portato alla formulazione di *2 osservatori colorimetrici standard* che, analogamente agli osservatori fotometrici standard, prondono il nome di *osservatore fotopico* e di *osservatore scotopico*.

Definito l'esperimento vediamo ora cosa stabilisce la *generalizzazione tricromatica*:

- in un ampio ambito di condizioni di osservazione (sia fotopiche che scotopiche) molti stimoli colore possono essere fatti corrispondere completamente alla miscela additiva di 3 primari fissati la cui potenza radiante sia opportunamente aggiustata;
- altri stimoli colore devono essere miscelati con 1 dei 3 primari per ottenere la corrispondenza con gli altri 2;

- scegliendo particolari terne di primari può accadere che ci siano stimoli colore che devono essere miscelati con 2 dei 3 primari per avere una corrispondenza con il terzo.

In ogni modo, ogni stimolo colore può essere posto in corrispondenza, secondo l'uno o l'altro di questi modi, con 3 primari fissati di cui sia aggiustabile dall'osservatore la potenza radiante.

La scelta dei primari, sebbene vasta, non è completamente arbitraria. Si può usare qualsivoglia insieme di primari a patto che nessuno dei tre possa essere posto in corrispondenza con una miscela addittiva degli altri due.

Se oltre a soddisfare le condizioni ora enunciate, si assume che i risultati di un esperimento di corrispondenza tricromatica soddisfino anche le seguenti leggi, si ottiene una foma più forte e quantitativa della generalizzazione tricromatica:

1. **Legge di simmetria**
   Se uno stimolo colore A corrisponde ad uno stimolo colore B, allora lo stimolo B corrisponde allo stimolo A.

2. **Legge di transitività**
   Se A corrisponde a B e B corrisponde a C, allora A corrisponde a C.

3. **Legge di proporzionalità**
   Se A corrisponde a B, allora αA corrisponde a αB, dove α è qualsiasi fattore positivo per cui la potenza radiante dello stimolo è aumentata o ridotta, mentre la distribuzione spettrale relativa è mantenuta uguale.

4. **Legge di addittività**
   Se A, B, C, D sono 4 stimoli qualsivoglia, allora se valgono due delle seguenti tre condizioni:

   - A corrisponde a B
   - C corrisponde a D
   - $(A + C)$ corrisponde a $(B + D)$

   allora vale anche che la quarta cioè: $(A + D)$ corrisponde a $(B + C)$. Con $(X + Y)$ si intende naturalmente la miscela additiva degli stimoli colore X e Y.

6.3.2 Spazio di Tristimolo

Assumendo valida la legge di generalizzazione tricromatica è possibile e conveniente rappresentare uno stimolo colore mediante un elemento di uno spazio vettoriale tridimensionale chiamato *spazio di tristimolo*.

Consideriamo un arbitrario stimolo colore definito dalla distribuzione spettrale di potenza radiante $\{P_Q(\lambda)d(\lambda)\}$ e 3 primari che denotiamo con $\underline{R}$, $\underline{G}$ e $\underline{B}$. I 3 primari sono luci monocromatiche di lunghezza d'onda $\lambda_R = 700\ nm$ $\lambda_G = 546,1\ nm$ e $\lambda_B = 435,8\ nm$ che appaiono rispettivamente di colore rosso, verde e blu.

La corrispondenza tra lo stimolo test $\{P_Q(\lambda)d(\lambda)\}$ e la miscela additiva di quantità adatte dei 3 primari $\underline{R}$, $\underline{G}$ e $\underline{B}$ può essere espressa dall'equazione:

$$\{P_Q(\lambda)d(\lambda)\} =_C R_Q\underline{R} + G_Q\underline{G} + B_Q\underline{B} = \underline{Q}$$

Con il simbolo $=_C$ intendiamo sottolineare che la relazione di uguaglianza è definita dall'esperimento di corrispondenza tricromatica. E' l' esperimento che ci consente di rappresentare un' arbitraria distribuzione spettrale di potenza mediante un vettore di un opportuno spazio tridimensionale. Denotiamo tale vettore con $\underline{Q}$, $\underline{Q}$ costituisce pertanto una descrizione alternativa per $\{P_Q(\lambda)d(\lambda)\}$.

Nei casi in cui $\underline{Q}$ deve essere miscelato con opportune quantità di un primario per ottenere la corrispondenza con gli altri due, avremo un'equazione della forma:

$$\underline{Q} + R_Q\underline{R} = G_Q\underline{G} + B_Q\underline{B}$$

o anche:

$$\underline{Q} = -R_Q\underline{R} + G_Q\underline{G} + B_Q\underline{B}$$

Il primario che deve essere miscelato con $\underline{Q}$ (nell'esempio è $\underline{R}$) ha coefficiente di tristimolo negativo. Il segno meno non deve essere interpretato come una potenza negativa, esso indica solo che $\underline{R}$ deve essere miscelato a $\underline{Q}$ per ottenere una corripondenza con gli altri due primari.

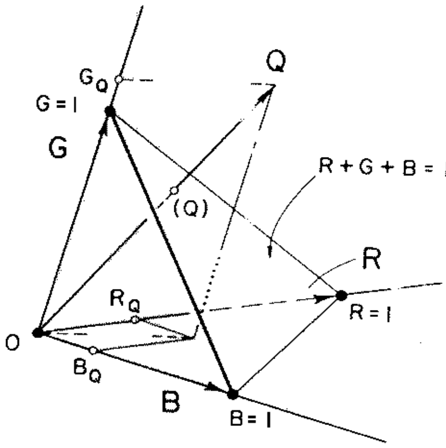


Figura 6.3 Lo spazio di tristimolo.

Queste equazioni hanno un ovvia interpretazione geometrica nello spazio tridimensionale. I vettori unitari $\underline{R}$, $\underline{G}$ e $\underline{B}$, che rappresentano gli stimoli primari, definiscono lo spazio che prende il nome di *spazio di tristimolo.* Essi hanno l'origine in comune, sono diretti in tre diverse direzioni e la loro lunghezza, in ognuna delle tre direzioni, corrisponde ad un' unità (Figura 6.3). Le unità di misura dei 3 primari non sono unità di luminanza, esse sono state scelte in modo tale che la miscela ottenuta sommando un' unità di ciascuno di essi fosse in perfetta corrispondenza tricromatica con uno stimolo che appare bianco e la cui distribuzione spettrale di potenza è uniforme ed uguale a 1 su tutto l'intervallo visibile. Tale stimolo è denominato *equi-energy stimulus*.

Per esempio per porre in corrispondenza una unità di $\underline{R}$, una unità di $\underline{G}$ e una unità di $\underline{B}$ con l' *equi-energy stimulus* di luminanza pari a $5,6508\ cd \cdot m^{-2}$ servono rispettivamente: $1,0000\ cd \cdot m^{-2}$ di $\underline{R}$, $4,5907\ cd \cdot m^{-2}$ di $\underline{G}$, $0,0601\ cd \cdot m^{-2}$ di $\underline{B}$.

I rapporti di luminanza $1 : 4,5907 : 0.0601$ corrispondono ai seguenti rapporti tra le potenza radianti dei tre primari $72,1 : 1,4 : 1$. Ciò dipende naturalmente dalla diversa efficienza luminosa dei tre primari.

Le quantità di primari, espresse rispetto alle unità sopra definite, necessarie ad ottenere la corrispondenza tricromatica con un definito stimolo test vengono denominate *coefficienti di tristimolo*. Pertanto R_Q, G_Q, B_Q sono i *coefficienti di tristimolo* dello stimolo $\underline{Q}$. Il vettore (R_Q, G_Q, B_Q) è denominato *vettore di tristimolo*.

6.3.3 Le funzioni di corrispondenza cromatica

Consideriamo ora l' *equi-energy stimulus* e denotiamo con $\{E(\lambda)d\lambda\}$ la sua distribuzione spettrale di potenza radiante, e con $\underline{E}$ il vettore che lo caratterizza nello spazio di tristimolo.

Per ogni componente monocromatica dell' *equi-energy stimulus* possiamo effettuare l'esperimento di corrispondenza tricromatica con i tre primari sopra descritti ed ottenere, per ogni lunghezza d'onda λ, l'equazione:

$$\underline{E}_\lambda = \overline{r}(\lambda)\underline{R} + \overline{g}(\lambda)\underline{G} + \overline{b}(\lambda)\underline{B}$$

dove $\overline{r}(\lambda), \overline{g}(\lambda), \overline{b}(\lambda)$ sono i valori di tristimolo che definiscono il vettore $\underline{E}_\lambda$ corrispondente ad una luce monocromatica di lunghezza d'onda λ e di potenza 1.

Il grafico in Figura 6.4 mostra i risultati di esperimenti in cui lo stimolo sottende un angolo visivo di 2° gradi e gli osservatori hanno una normale visione dei colori.

Gli insiemi di valori $\overline{r}(\lambda), \overline{g}(\lambda), \overline{b}(\lambda)$ sono denominati *color matching functions* (funzioni di corrispondenza cromatica). Queste funzioni, che incorporano le proprietà di corrispondenza cromatica di un osservatore con visione normale rispetto al sistema di primari $\underline{R}$ $\underline{G}$ e $\underline{B}$, definiscono pertanto l'*osservatore colorimetrico standard*. Per illustrarne l'uso consideriamo lo stimolo colore monocromatico di potenza radiante uguale a 1 e di lunghezza d'onda $\lambda = 475$. Essendo $\overline{r}(475) = -0,045$, $\overline{g}(475) = 0,032$ e $\overline{b}(475) = 0,186$ avremo:

$$\underline{E}_{475} = -0,045\underline{R} + 0,032\underline{G} + 0,186\underline{B}$$

A partire dalle *color matching functions* si determinano facilmente i coefficienti di tristimolo di un arbitrario stimolo colore. Supponiamo che tale stimolo, che denotiamo con $\underline{Q}$, sia caratterizzato dalla distribuzione spettrale di potenza radiante $\{P_Q(\lambda)d(\lambda)\}$. Consideriamo la componente monocromatica di lunghezza d'onda λ di tale stimolo; essendo an-

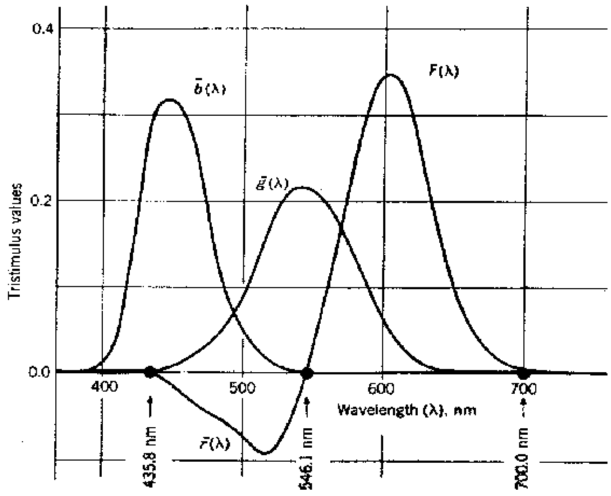


Figura 6.4 Funzioni di corrispondenza cromatica.

ch'essa una stimolo colore possiamo rappresentarla con un vettore nella spazio di tristimolo, denotiamo con $\underline{Q}_\lambda$ tale vettore.

Ricordando che il vettore $\underline{E}_\lambda$ corrisponde allo stimolo monocromatico di lunghezza d'onda λ e di potenza 1 possiamo scrivere:

$$\underline{Q}_\lambda = (P_Q(\lambda)d\lambda)\underline{E}_\lambda = (P_Q(\lambda)d\lambda)\overline{r}(\lambda)\underline{R} + (P_Q(\lambda)d\lambda)\overline{g}(\lambda)\underline{G} + (P_Q(\lambda)d\lambda)\overline{b}(\lambda)\underline{B}$$

Se assumiamo che la distribuzione spettrale di potenza radiante $\{P_Q(\lambda)d(\lambda)\}$ sia una funzione continua nell'intervallo di lunghezze d'onda che costituisce lo spettro visibile e denotiamo gli estremi di tale intervallo con λ_a e λ_b, l'equazione scritta è integrabile in tale intervallo. Pertanto, i coefficienti di tristimolo R_Q, G_Q e B_Q che definiscono il vettore $\underline{Q}$, sono dati da:

$$R_Q = \int_{\lambda_a}^{\lambda_b} P_Q(\lambda)\overline{r}(\lambda)d\lambda \quad G_Q = \int_{\lambda_a}^{\lambda_b} P_Q(\lambda)\overline{g}(\lambda)d\lambda \quad B_Q = \int_{\lambda_a}^{\lambda_b} (P_Q(\lambda)\overline{b}(\lambda)d\lambda$$

Nella pratica accade spesso che al posto dell'integrale si facciano le somme in quanto sia le funzioni di corrispondenza tricromatica che le distribuzioni spettrali di potenza radiante vengono date sottoforma di tabelle. In tal caso avremo pertanto:

$$R_Q = \sum_{\lambda_a}^{\lambda_b} P_Q(\lambda)\overline{r}(\lambda)\Delta\lambda \quad G_Q = \sum_{\lambda_a}^{\lambda_b} P_Q(\lambda)\overline{g}(\lambda)\Delta\lambda \quad B_Q = \sum_{\lambda_a}^{\lambda_b} P_Q(\lambda)\overline{b}(\lambda)\Delta\lambda$$

Si assume naturalmente che $P_Q(\lambda)$ sia constante nell'intervallo $\Delta\lambda$.

Due stimoli colore $\underline{Q}_1$ e $\underline{Q}_2$ definiti da due diverse distribuzioni di potenza radiante $\{P_{Q_1}(\lambda)d(\lambda)\}$ e $\{P_{Q_2}(\lambda)d(\lambda)\}$ sono detti *stimoli metamerici* se sono in completa corrispondenza cromatica cioè se valgono le seguenti relazioni:

$$\int_{\lambda_a}^{\lambda_b} P_{Q_1}(\lambda)\overline{r}(\lambda)d\lambda = \int_{\lambda_a}^{\lambda_b} P_{Q_2}(\lambda)\overline{r}(\lambda)d\lambda$$

$$\int_{\lambda_a}^{\lambda_b} P_{Q_1}(\lambda)\overline{g}(\lambda)d\lambda = \int_{\lambda_a}^{\lambda_b} P_{Q_2}(\lambda)\overline{g}(\lambda)d\lambda$$

$$\int_{\lambda_a}^{\lambda_b} P_{Q_1}(\lambda)\overline{b}(\lambda)d\lambda = \int_{\lambda_a}^{\lambda_b} P_{Q_2}(\lambda)\overline{b}(\lambda)d\lambda$$

6.3.4 Diagramma di cromaticità

La rappresentazione di uno stimolo colore nello spazio di tristimolo fornisce tutte le informazioni sufficienti ad una perfetta caratterizzazione ma, nella pratica colorimetrica, si preferisce una rappresentazione bidimensionale in cui si evidenzia la grandezza relativa dei valori di tristimolo. Una utile rappresentazione bidimensionale si ottiene sul piano definito dall'equazione:

$$R + G + B = 1$$

Tutti i vettori di tristimolo, o le loro estensioni, intersecano tale piano in un punto. La posizione del punto nel piano determina univocamente la direzione del vettore nello spazio di tristimolo ma non fornisce nessuna informazione sulla sua lunghezza, che rappresenta la potenza luminosa dello stimolo. Su tale piano uno stimolo è rappresentato mediante le *coordinate di cromaticità*; dati i valori di tristimolo R_Q, G_Q e B_Q di uno stimolo $\underline{Q}$ le coordinate di cromaticità r_Q, g_Q, b_Q sono definite dalle seguenti equazioni:

$$r_Q = \frac{R_Q}{R_Q + G_Q + B_Q}$$

$$g_Q = \frac{G_Q}{R_Q + G_Q + B_Q}$$

$$b_Q = \frac{B_Q}{R_Q + G_Q + B_Q}$$

E' evidente che $r_Q + g_Q + b_Q = 1$; bastano pertanto 2 coordinate di cromaticità per definire univocamente la terza.

Se si ricavano le coordinate di cromaticità degli stimoli monocromatici $\underline{E}_\lambda$, a partire dai valori di tristimolo $\overline{r}(\lambda)$, $\overline{g}(\lambda)$ e $\overline{b}(\lambda)$ (cioè dalle funzioni di corrispondenza tricromatica), si ottiene il *diagramma di cromaticità* rappresentato in Figura 6.5.

Nel diagramma, la linea curva è il luogo dei punti che rappresentano i colori di tutte le luci monocromatiche dello spettro; tale linea è denominata *spectrum locus*. Le coordinate di cromaticità dei colori corrispondenti a stimoli monocromatici della lunghezza d'onda

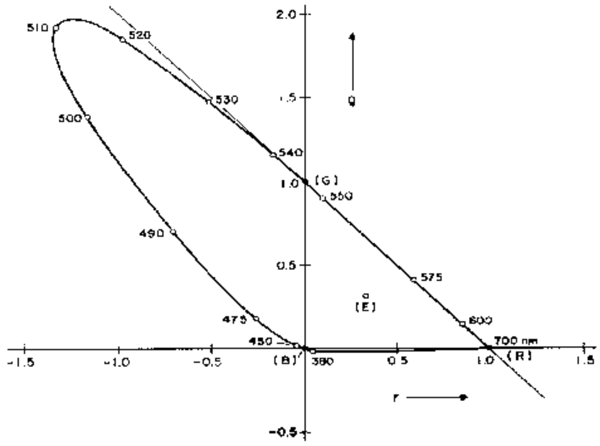


Figura 6.5 Diagramma di cromaticità (r, g).

indicata sono denotate con piccoli cerchi, le lunghezze d'onda sono rappresentate a intervalli di 10 *nm*. Le cromaticità dei 3 primari sono rappresenate dai punti che costituiscono i vertici del triangolo. La linea che congiunge i 2 primari $\underline{R}$ e $\underline{B}$ è denominata *purple line* (linea dei porpora). Il punto denotato da (E), al centro del triangolo, rappresenta lo stimolo colore $\underline{E}$. Ricordiamo che le unità di misura dei 3 primari sono state scelte in modo tale che:

$$\underline{E} = \underline{R} + \underline{G} + \underline{B}$$

pertanto, le coordinate di cromaticità r_E, g_E e b_E di $\underline{E}$ risultano essere:

$$r_E = g_E = b_E = 1/3$$

Come conseguenza diretta dell'aver fissato le unità di misura dei primari nel modo sopra descritto l'area sotto le 3 curve che rappresentano le funzioni di corrispondenza cromatica è uguale a 1. Infatti:

$$R_E = G_E = B_E = 1$$

ma

$$R_E = \int_{\lambda_a}^{\lambda_b} \overline{r}(\lambda)d\lambda = 1 \quad G_E = \int_{\lambda_a}^{\lambda_b} \overline{g}(\lambda)d\lambda = 1 \quad B_E = \int_{\lambda_a}^{\lambda_b} \overline{b}(\lambda)d\lambda = 1$$

Il diagramma di cromaticità gode di due proprietà interessanti:

- le coordinate di cromaticità, in un sistema di primari fissato, sono le stesse per tutti gli stimoli colore che hanno la stessa *distribuzione spettrale di potenza relativa.* L'informazione sulla potenza totale dello stimolo, rappresentata dalla lunghezza del vettore di tristimolo, viene infatti persa nel diagramma di cromaticità. A causa di ciò il diagramma di cromaticità fornisce delle indicazioni sugli attributi psicologici tinta e saturazione ma non sulla brillanza di un colore.

- un punto che rappresenta la miscela additiva di 2 stimoli colore giace sulla retta che congiunge i punti che rappresentano tali stimoli nel diagramma di cromaticità.

Un conveniente sistema di coordinate, nel piano di equazione $R+G+B=1$, è costituito dalle rette che si ottengono dall'intersezione di tale piano con i piani definiti dalle tre coppie di vettori $(\underline{R}, \underline{B})$, $(\underline{R}, \underline{G})$ e $(\underline{B}, \underline{G})$. Le intersezioni di tali rette individuano un triangolo. Tale triangolo, contenuto nel diagramma di cromaticità di Figura 6.5, rappresenta la *gamma completa di colori* che si possono ottenere come risultato della *miscela additiva* dei 3 colori primari.

Sebbene il diagramma di cromaticità rappresentato da un triangolo equilatero sia oggi utilizzato raramente (tale triangolo è denominato anche triangolo di Maxwell dal nome del fisico James Clerk Maxwell (1831-1879) che lo utilizzò per i suoi lavori sul colore), è istruttivo considerarlo prima di esaminare il diagramma in uso oggi, un triangolo rettangolo isoscele. Suddividiamo il triangolo mediante 3 insiemi di linee parallele ad ognuno dei lati; in Figura 6.6, i 3 insiemi di linee parallele sono mostrati separatamente rispetto al triangolo. Utilizziamo le 3 variabili r, g, b per rappresentare rispettivamente l'ammontare relativo di rosso, verde e blu necessario per ottenere un dato colore della gamma. Se i valori di tristimolo di uno stimolo $\underline{Q}$ sono $R_Q = 60$, $G_Q = 80$, $B_Q = 60$ allora l'ammontare relativo di rosso sarà: $r_Q = 60/200 = 0.3$, l'ammontare relativo di verde sarà : $g_Q = 80/200 = 0.4$, l'ammontare relativo di blu sarà: $b_Q = 60/200 = 0.3$. La cromaticità del colore può ora essere rappresentata sul diagramma. E' ovvio che solo 2 valori relativi sono sufficienti a specificare la cromaticità di uno stimolo colore, i 2 comunemente impiegati sono r e g. La cromaticità di un uno stimolo colore $\underline{Q}$ è denotata generalmente così: $(r_Q = 0.3, g_Q = 0.4)$ o più comunemente così : $(0.3, 0.4)$, sottintendendo i 2 primari r e g.

E' più conveniente considerare un triangolo di cromaticità, a forma di triangolo rettangolo isoscele piuttosto che di triangolo equilatero. Nel triangolo di Figura 6.7 sono evidenziati i primari e i colori che si ottengono per sintesi additiva di coppie di primari. Il punto *J giallo* rapprenta una miscela in cui verde e rosso sono in uguali percentuali $(0.5, 0.5)$. Il *magenta* $M = (0.5, 0.0)$ è ottenuto da uguali quantità di rosso e blu e uguali quantità dei primari blu e verde producono il *ciano* $C = (0.0, 0.5)$.

I colori prodotti da miscele di luce hanno saturazione massima quando i punti che ne rappresentano la cromaticità sono situati sui lati del triangolo. Il punto Q, che è situato su un lato del diagramma, mostra la cromaticità della miscela di due stimoli primari, rosso e verde, in diverse proporzioni.

La cromaticità del bianco, per cui la saturazione percepita è minima, è rappresentata dal punto E, situato al centro del diagramma.

Le variazioni di saturazione possono essere analizzate combinando stimoli complementari, per esempio rosso e ciano, in diverse proporzioni. Consideriamo la linea tratteggiata che congiunge R e C, supponiamo di partire da R, che corrisponde alla cromaticità del rosso generato da un fascio di luce, e di tenere spento il fascio di luce corrispondente al

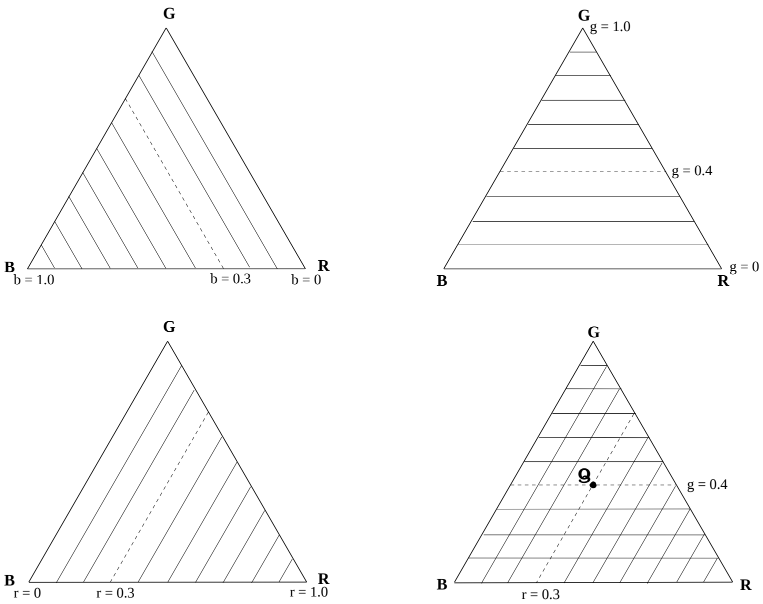


Figura 6.6 Triangolo di Maxwell.

ciano. Diminuiamo progressivamente la potenza del fascio di luce rossa e aumentiamo quella del ciano. La saturazione del rosso che percepiamo diminuisce mentre la cromaticità passa con continuità attraverso il rosa (P) e poi il bianco (E). Oltre E, la tinta vira verso il ciano e la saturazione del colore aumenta fino a raggiungere un massimo in C. La linea tratteggiata che connette R a C è la linea della miscela.

6.3.5 Stimoli colore immaginari

Uno stimolo colore è rappresentato dai volori di tristimolo rispetto ad un definito insieme di 3 primari ed uno o due dei valori di trismolo possono essere negativi. Naturalmente non tutte le terne di valori identificano uno stimolo colore nel sistema di primari scelto. In colorimetria risulta conveniente definire anche *stimoli colore immaginari* cioè terne di valori di tristimolo che non corrispondono a nessun colore reale. Affinché una terna di coefficienti sia considerata uno stimolo colore immaginario è necessario che il vettore da essa rappresentato nello spazio di tristimolo non produca il vettore nullo $(0, 0, 0)$ quando viene sommato ad ognuno dei vettori che rappresentano gli stimoli colore reali. Per esempio, se R_Q, G_Q, B_Q sono i valori di tristimolo di uno stimolo colore reale $-R_Q$, $-G_Q$, $-B_Q$ non sono ammessi come valori che specificano uno stimolo immaginario. Si richiede anche che la somma di due vettori che rappresentano stimoli immaginari non dia come risultato il vettore $(0, 0, 0)$.

Nello spazio di tristimolo si individua un piano passante per l'origine tale che tutti i vettori che rappresentano stimoli reali siano situati in uno dei due semispazi ma non sul piano stesso. Tutti gli altri vettori nello stesso semispazio, ma non nel piano, sono accettati come stimoli immaginari. Nessun altro vettore rappresenta uno stimolo immaginario.

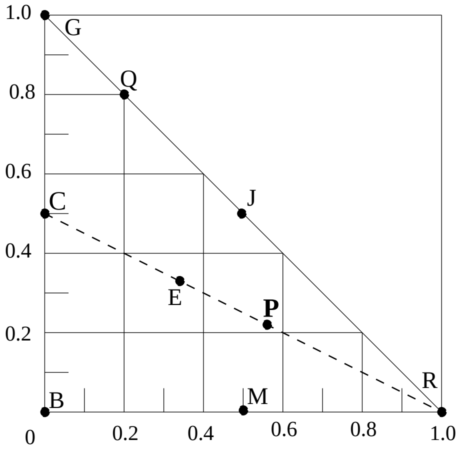


Figura 6.7 Triangolo di cromaticità.

6.3.6 Lo spazio (X, Y, Z)

Le funzioni $\overline{r}(\lambda)$, $\overline{g}(\lambda)$, $\overline{b}(\lambda)$ definiscono un sistema preciso, basato su caratteristiche della percezione umana, per la specifica del colore a partire dalla distribuzione spettrale di potenza di uno stimolo. Nella pratica però tale sistema non è usato nella forma descritta. Infatti la CIE per ragioni di convenienza, quali quella di rendere più semplici i calcoli da effettuare ad opera di colorimetri, ha deciso di adottare come standard un sistema che si deriva dal sistema tricromatico $\underline{R}$, $\underline{G}$, $\underline{B}$ mediante una trasformazione lineare. La trasformazione è stata scelta in modo tale che nel nuovo sistema di primari nessuno stimolo colore reale fosse caratterizzato da coefficienti di tristimolo negativi. I primari del sistema standard, denotati da $\underline{X}$, $\underline{Y}$ e $\underline{Z}$, sono stimoli non reali (immaginari), non possono pertanto essere ottenuti da nessuna distribuzione spettrale di potenza. Nello spazio di tristimolo essi sono rappresentati da vettori situati al di fuori del dominio che rappresenta gli stimoli reali; nel diagramma di cromaticità i punti che li rappresentano giacciono all'esterno della regione delimitata dallo spectrum locus e dalla purple line. In Figura 6.8 si può osservare che il triangolo i cui vertici sono costituiti dalle coordinate di cromaticità dei primari $\underline{X}$, $\underline{Y}$ e $\underline{Z}$ racchiude completamente lo spectrum locus e la purple line. Ciò assicura che le coordinate di cromaticità e i valori di tristimolo di qualsiasi stimolo reale siano sempre positive.

La trasformazione lineare che lega i valori di tristimolo nel sistema di primari($\underline{R}$, $\underline{G}$, $\underline{B}$) ai valori di tristimolo nel sistema di primari ($\underline{X}$, $\underline{Y}$, $\underline{Z}$) è descritta dalle seguenti equazioni:

$$X = 0,49R + 0,31G + 0,20B$$

$$Y = 0,17697R + 0,81240G + 0,1063B$$

$$Z = 0,00R + 0,01G + 0,99B$$

In ogni equazione i coefficienti hanno somma 1; pertanto, quando per uno stimolo colore $\underline{Q}$ vale che $R_Q = G_Q = B_Q$ vale anche che $X_Q = Y_Q = Z_Q$.

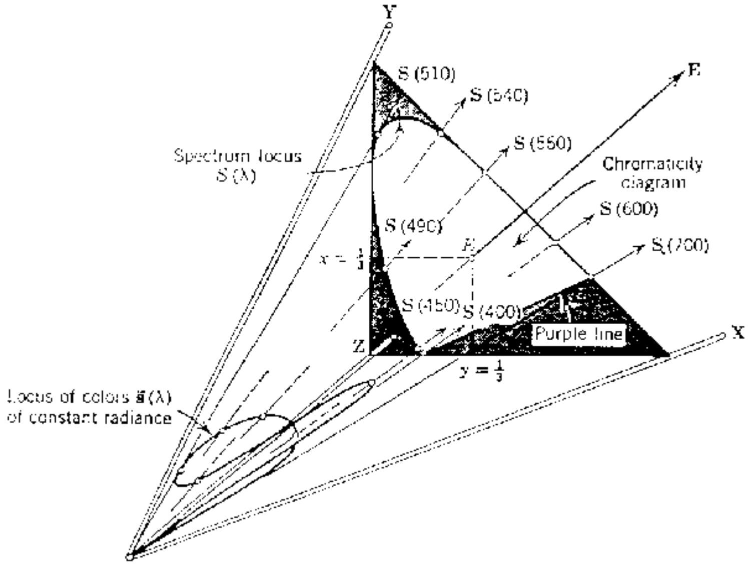


Figura 6.8 Spazio ($\underline{X}$, $\underline{Y}$, $\underline{Z}$).

Dalle funzioni di corrispondenza tricromatica $\overline{r}(\lambda)$, $\overline{g}(\lambda)$ e $\overline{b}(\lambda)$ si ricavano, mediante equazioni simili a quelle sopra descritte, le funzioni di corrispondenza tricromatica $\overline{x}(\lambda)$, $\overline{y}(\lambda)$ e $\overline{z}(\lambda)$:

$$\overline{x}(\lambda) = 0,49\overline{r}(\lambda) + 0,31\overline{g}(\lambda) + 0,20\overline{b}(\lambda)$$

$$\overline{y}(\lambda) = 0,17697\overline{r}(\lambda) + 0,81240\overline{g}(\lambda) + 0,1063\overline{b}(\lambda)$$

$$\overline{z}(\lambda) = 0,00\overline{r}(\lambda) + 0,01\overline{g}(\lambda) + 0,99\overline{b}(\lambda)$$

Le funzioni $\overline{x}(\lambda)$, $\overline{y}(\lambda)$ e $\overline{z}(\lambda)$ sono le funzioni più importanti in colorimetria.

Il sistema di primari $\underline{X}$, $\underline{Y}$ e $\underline{Z}$ gode anche di un'altra importante proprietà: gli stimolo $\underline{X}$ e $\underline{Z}$ sono tali da avere efficienza luminosa uguale a 0. La efficienza luminosa è attribuita tutta al primario $\underline{Y}$. Ciò fa si che la funzione di corrispondenza tricromatica $\overline{y}(\lambda)$ coincida con la funzione $V(\lambda)$, funzione che rappresenta l'efficienza luminosa dell'osservatore fotometrico standard. I coefficienti che, a partire dai coefficienti di tristimolo R, G e B, portano al calcolo del coefficiente Y sono stati accuratamente scelti affinché fossero negli stessi rapporti in cui sono le unità di luminanza usate per effettuare la corrispondenza tra $\underline{R}$, $\underline{G}$ e $\underline{B}$ e lo stimolo $\underline{E}$. Infatti:

$$0,17697 : 0,81240 = 1 : 4,5907$$

$$0,81240 : 0,1063 = 4,5907 : 0,0601$$

Questo fa si che il coefficiente Y sia proporzionale alla luminanza dello stimolo colore specificato. Perciò il rapporto tra i coefficienti Y_1 e Y_2 di due stimoli colore specifica il rapporto tra le rispettive luminanze.

Le seguenti equazioni evidenziano le relazioni esistenti tra le funzioni di corrispondenza tricromatica $\overline{x}(\lambda)$, $\overline{y}(\lambda)$ e $\overline{z}(\lambda)$, le coordinate di cromaticità $x(\lambda)$, $y(\lambda)$, $z(\lambda)$ e la funzione $V(\lambda)$:

$$\overline{x}(\lambda) = \frac{x(\lambda)}{y(\lambda)}V(\lambda) \quad \overline{y}(\lambda) = V(\lambda) \quad \overline{z}(\lambda) = \frac{z(\lambda)}{y(\lambda)}V(\lambda)$$

Ricordiamo che $x(\lambda) = \frac{\overline{x}(\lambda)}{\overline{x}(\lambda)+\overline{y}(\lambda)+\overline{z}(\lambda)}$ e analogamente per le altre coordinate di cromaticità.

I valori di tristimolo, che nel sistema di primari $\underline{X}$, $\underline{Y}$ e $\underline{Z}$, caratterizzano il colore di una sorgente la cui distribuzione spettrale di potenza è $\{P_Q(\lambda)d\lambda\}$, sono calcolati, come nel sistema di primari $\underline{R}$, $\underline{G}$, $\underline{B}$, dalle equazioni:

$$X_Q = \int_{\lambda_a}^{\lambda_b} P_Q(\lambda)\overline{x}(\lambda)d\lambda \quad Y_Q = \int_{\lambda_a}^{\lambda_b} P_Q(\lambda)\overline{y}(\lambda)d\lambda \quad Z_Q = \int_{\lambda_a}^{\lambda_b} (P_Q(\lambda)\overline{z}(\lambda)d\lambda$$

Si suppone naturalmente che la distribuzione spettrale di potenza $\{P_Q(\lambda)d\lambda\}$ sia descritta da una funzione continua nell'intervallo $[\lambda_a, \lambda_b]$.

6.4 Il sistema colorimetrico della CIE

Per consenso generale la specifica degli standard da usare in colorimetria e in fotometria, in tutti i paesi, spetta alla CIE (Commission Internationale de l' Eclairage). Il sistema colorimetrico della CIE è costituito dagli *osservatori colorimetrici standard* e da un insieme di altre raccomandazioni utili alla specifica del colore; alcune di queste sono descritte nel seguito.

6.4.1 Gli osservatori colorimetrici standard

Il primo osservatore ideale è stato definito dalla CIE nel 1931 e prende il nome di *CIE 1931 Standard Colorimetric Observer*. Esso è definito mediante le funzioni $\overline{x}(\lambda)$, $\overline{y}(\lambda)$, $\overline{z}(\lambda)$. Le funzioni sono fornite in forma tabellare nell'intervallo di lunghezze d'onda che va dai 380 ai 780 nm ad intervalli di 5 nm. Nel 1971 sono stati forniti valori, ottenuti per interpolazione, ad intervalli di 1 nm. I dati sperimentali che costituiscono la base su cui sono state definite le funzioni $\overline{r}(\lambda)$, $\overline{g}(\lambda)$, $\overline{b}(\lambda)$, e quindi le funzioni $\overline{x}(\lambda)$, $\overline{y}(\lambda)$, $\overline{z}(\lambda)$ sono stati ottenuti in modo tale da evitare la stimolazione dei bastoncelli. A tal fine la parte di campo visivo occupata dai 2 stimoli da porre in corrispondenza occupava un angolo visivo di 2 gradi. L'osservatore standard definito nel 1931 si può usare se l'angolo visivo secondo cui si guarda uno stimolo luminoso non supera i 4 gradi.

Nel 1964 la CIE ha definito un ulteriore standard, cioè un insieme di funzioni di corrispondenza tricromatica denotate da $\overline{x}_{10}(\lambda)$, $\overline{y}_{10}(\lambda)$, $\overline{z}_{10}(\lambda)$. Queste funzioni definiscono un nuovo osservatore che prende il nome di *CIE 1964 Supplementary Standard Colorimetric Observer*. L' osservatore supplementare deve essere usato quando è richiesta la stima del colore di stimoli che occupano, nel campo visivo, un angolo superiore ai 4 gradi.

L'osservatore CIE 1931 possiede una proprietà visiva, tipica delle persone con visione normale, che non è una conseguenza necessaria dell'obbedienza alle leggi di generalizzazione tricromatica. Questa proprietà, la legge di addittività della brillanza, costituisce il principio di base della fotometria. Ricordiamo infatti che dati due stimoli descritti dalle distribuzioni spettrali di potenza $\{P_1(\lambda)d\lambda\}$ e $\{P_2(\lambda)d\lambda\}$, la condizione di corrispondenza di brillanza, ma non di colore, è descritta dall'equazione:

$$\int_\lambda P_1(\lambda)V(\lambda)d\lambda = \int_\lambda P_2(\lambda)V(\lambda)d\lambda$$

Il fatto che l'osservatore CIE 1931 goda della proprietà sopra definita è espresso dal fatto che una delle funzioni di corrispondenza tricromatica, la funzione $\overline{y}(\lambda)$, è stata scelta in modo da coincidere con la funzione di efficienza luminosa $V(\lambda)$, definita dalla CIE nel 1924 per fornire basi quantitative alla fotometria in condizioni fotopiche. Pertanto l'osservatore CIE 1931 comprende in un unico schema quantitativo le proprietà di corrispondenza tricromatica e di corrispondenza di brillanza anche di stimoli di colori diversi (cioè diversi per saturazione e tinta). L'osservatore CIE 1964 non gode di questa proprietà.

6.4.2 Illuminanti e sorgenti

L' esperimento di corrispondenza tricromatica consente di definire in modo quantitativo uno stimolo colore di cui sia nota la distribuzione spettrale di potenza radiante. Tale metodo funziona nel caso in cui ci troviamo di fronte ad una sorgente luminosa ma come si misura il colore di *oggetti opachi*, che riflettono luce in modo diffuso, o di *oggetti trasparenti*, che trasmettono la luce? La misura del colore è effettuata sulla luce diffusamente riflessa da una superficie opaca o sulla luce trasmessa da una superficie trasparente. In tal caso il colore dipende sia dalla distribuzione spettrale di potenza della sorgente che illumina l' oggetto sia dalle caratteristiche di *riflettanza* o *trasmittanza* dell'oggetto.

Le caratteristiche di riflettanza (o trasmittanza) di un oggetto sono misurate rispetto ad un modello, il *perfect reflecting diffuser* (o il *perfect trasmitting diffuser*), definito come il diffusore uniforme con riflettanza (trasmittanza) uguale a 1 per ogni lunghezza d'onda. La riflettanza è definta come il rapporto tra flusso luminoso riflesso e flusso luminoso incidente per ogni lunghezza d'onda, analogamente la trasmittanza è definta come il rapporto tra flusso luminoso trasmesso e flusso luminoso incidente.

Pertanto un colore di superficie deve essere specificato dando sia i valori di tristimolo sia le caratteristiche della sorgente utilizzata per illuminarlo.

Quando si è interessati a misurare lo stimolo colore proveniente dalla superficie di un oggetto (per esempio per riprodurre tale colore a monitor) è desiderabile limitare le misure e i calcoli a ben definite distribuzioni spettrali di potenza. La CIE raccomanda un insieme di tali distribuzioni che prendono il nome di *illuminanti standard* e vengono date in forma tabellare, viene tabulata la distribuzione della potenza spettrale relativa. Ricordiamo qui i nomi di importanti illuminanti: D_{65}, A, B, C ognuna di esse intende rappresentare una particolare fase della luce del giorno.

La CIE fa una distinzione tra illuminanti e *sorgenti*, la parola sorgente si riferisce ad una sorgente fisica di energia radiante (per esempio una lampada a gas, una lampada con filamento di tungsteno ...) la parola illuminante si riferisce ad una specifica distribuzione spettrale relativa di potenza radiante. La distribuzione che caratterizza un' illuminante può non essere generabile da una sorgente.

Se lo stimolo colore non è prodotto da una sorgente luminosa ma da un oggetto opaco illuminato, O, allora la distribuzione spettrale di potenza $\{P_O(\lambda)d\lambda\}$ sarà ottenuta ponendo:

$$P_O(\lambda) = \beta_O(\lambda) \cdot S(\lambda)$$

dove $S(\lambda)$ denota la potenza spettrale dell'illuminante nell' intervallo $d\lambda$ e $\beta_O(\lambda)$ è il fattore di riflettanza, cioè il rapporto tra la luce, di lunghezza d'onda λ, riflessa dall'oggetto O e la luce, di lunghezza d'onda λ, riflessa dal diffusore perfetto. Naturalmente $0 \leq \beta(\lambda) \leq 1$.

Definita la sorgente e le caratteristiche di riflettanza di O i valori di tristimolo X_O, Y_O, Z_O del colore dell'oggetto O sono pertanto così calcolati:

$$X_O = k\int_\lambda \beta_O(\lambda)S(\lambda)\overline{x}(\lambda)d\lambda \quad Y_O = k\int_\lambda \beta_O(\lambda)S(\lambda)\overline{y}(\lambda)d\lambda \quad Z_O = k\int_\lambda \beta_O(\lambda)S(\lambda)\overline{z}(\lambda)d\lambda$$

dove k, il coefficiente di proporzionalità, è di solito definito da:

$$k = \frac{100}{\int_\lambda S(\lambda)\overline{y}(\lambda)d\lambda}$$

La definizione di k fa si che il coefficiente Y di uno stimolo colore che deriva da un oggetto che è un diffusore riflettente perfetto sia uguale a 100, essendo per il diffusore perfetto $\beta(\lambda) = 1$ per ogni λ.

Il valore di tristimolo Y, se calcolato come indicato dall'equazione sopra descritta, prende il nome di *fattore di luminanza.*

Le stesse equazioni valgono se l'oggetto trasmette luce anziché rifletterla. In tal caso al posto del fattore di riflettanza $\beta(\lambda)$ si usa il *fattore di trasmittanza* $\tau(\lambda)$. Anche per τ vale che: $0 \leq \tau(\lambda) \leq 1$.

Da quanto detto è evidente che Y, per materiali opachi o trasparenti, ha un significato relativo. Per un materiale opaco Y è il *fattore di luminanza* cioè la luminanza della superficie valutata rispetto ad una superficie ideale sotto le stesse condizioni di illuminazione. Pertanto $Y = 100$ si ha solo per superfici bianche o trasparenti che non assorbono luce, $Y = 0$ per superfici che assorbono completamente la luce.

Osserviamo infine che, se denotiamo con $L(\lambda)$ la radianza della sorgente allora il valore di tristimolo Y rappresenta la luminanza L_v a patto di sceglire $k = K_m$ dove $K_m = 683\ lm \cdot W^{-1}$:

$$Y \equiv L_v = K_m \int_\lambda L(\lambda)\overline{y}(\lambda)d\lambda$$

6.4.3 Spazi colore Uniformi

Il sistema colorimetrico della CIE comprende anche gli *spazi colore uniformi* CIELAB e CIELUV. Tali spazi sono stati disegnati per aiutare nella predizione dell'ampiezza delle differenze che vengono percepite tra colori. Infatti il sistema colorimetrico precedentemente descritto costituisce uno strumento rigoroso per la specifica di un colore come miscela di tre primari ma, come si scoprì presto, non è adeguato a rappresentare in modo quantitativo le differenze che percepiamo tra colori. Accade che, a livelli costanti di luminanza, uguali distanze tra coppie di punti situati in diverse regioni del diagramma di cromaticità non corrispondano ad uguali differenze nel colore percepito. La non uniformità del diagramma è immediatamente evidente osservando la larga regione occupata dai verdi rispetto alle regioni relativamente piccole in cui sono affollati i rossi ed i blu.

Sono stati fatti numerosi tentativi per trovare formule che consentissero di rappresentare adeguatamente le differenze percepite tra colori. Nessuna delle molte formule proposte in letteratura nei decenni passati è considerata una soluzione sufficientemente adeguata. Il giudizio dell'osservatore varia grandemente in funzione delle condizioni di osservazione e dal tipo di stimolo che gli viene presentato. La dimensione, la forma, la luminanza, le distribuzioni spettrali relative degli stimoli test e degli stimoli che costituiscono l'ambiente in cui viene effettuato l'esperimento, influenzano il giudizio. Ci sono molti dati sperimentali che evidenziano la complessità del problema e che hanno portato a proporre formule empiriche disegnate per predire le differenze colore in condizioni sperimentali fissate. Sono anche stati fatti tentativi per sviluppare appropriati modelli matematici del processo visivo.

Ci limitiamo qui a presentare le raccomandazioni fatte dalla CIE nel 1976, raccomandazioni ancora attuali per valutare le differenze colore nella pratica colorimetrica.

La CIE raccomanda l'uso di due spazi colore, CIELUV e CIELAB, approssimativamente uniformi e delle differenze colore ad essi associate. Tali spazi sono stati scelti per promuovere una standardizzazione nella pratica colorimetrica. Il grado di uniformità di CIELUV e CIELAB è adeguato pe molte applicazioni ma insufficiente per altre, pertanto la ricerca di spazi e di formule per il calcolo delle differenze colore che abbiano una maggior correlazione con il giudizio visivo continua.

La struttura di base dei due spazi CIELUV e CIELAB è simile. In entrambi c' è un asse verticale, che rappresenta la *CIE 1976 lightness function* denotata con $L^\star$, ortogonale ad un piano in cui sono rappresentate rispettivamente le coordinate $(u^\star, v^\star)$ (in CIELUV) e le coordinate $(a^\star, b^\star)$ (in CIELAB). Tali coordinate sono calcolate a partire dall' osservatore colorimetrico standard e dai sistemi di coordinate standard in particolare dal sistema XYZ. Come esempio riportiamo lo spazio CIELUV.

6.4.4 Lo spazio CIELUV

Lo spazio CIELUV, abbreviazione di spazio CIE 1976 $(L^\star, u^\star, v^\star)$, si ottiene rappresentando in coordinate rettangolari le quantità $L^\star, u^\star, v^\star$ così definite:

$$L^\star = 116(\frac{Y}{Y_n})^{1/3} - 16$$
$$u^\star = 13L^\star(u' - u'_n)$$
$$v^\star = 13L^\star(v' - v'_n)$$

con il vincolo che $\frac{Y}{Y_n} > 0.01$. Se $\frac{Y}{Y_n} < 0.01$ è raccomandata una diversa procedura per il calcolo di $L^\star$; non siamo qui interessati a riportarla.

Le quantit u', v', u'_n, v'_n sono calcolate come segue:

$$u' = \frac{4X}{X + 15Y + 3Z} \qquad v' = \frac{9Y}{X + 15Y + 3Z}$$
$$u'_n = \frac{4X_n}{X_n + 15Y_n + 3Z_n} \qquad v'_n = \frac{9Y_n}{X_n + 15Y_n + 3Z_n}$$

I valori di tristimolo X_n, Y_n, Z_n sono quelli di uno stimolo colore bianco. Di solito come stimolo colore bianco si sceglie la distribuzione spettrale di potenza di una delle illuminanti standard della CIE per esempio la D_{65} o la A. L'asse $L^\star$ rappresenta i colori privi di tinta, i suoi valori variano da $L^\star = 0$ per il nero a $L^\star = 100$ per il bianco di riferimento per cui $Y = Y_n$. I valori intermedi rappresentano i grigi. I punti non situati sull'asse $L^\star$ rappresentano i colori cromatici.

La differenza $\Delta E^\star_{uv}$ tra due stimoli colore è definita mediante la distanza euclidea tra i due punti che nello spazio CIELUV rappresentano tali stimoli, cioè:

$$\Delta E^\star_{uv} = [(\Delta L^\star)^2 + (\Delta u^\star)^2 + (\Delta v^\star)^2]^{1/2}$$

Lo spazio CIELUV incorpora, per $L^\star$, costante il diagramma di cromaticità (u', v') che si deriva secondo le equazioni sopra descritte dal diagramma di cromaticità (x, y). Linee rette nel diagramma di cromaticità (x, y) restano linee rette nel diagramma (u', v'). In Figura 6.9 sono rappresentati lo spazio CIELUV e il diagramma di cromaticità (u', v').

E' molto utile nella pratica identificare quali grandezze nello spazio colore sono correlate agli attributi percettivi ligtness, tinta, chroma. Esse sono la metric lightness $L^\star$, la metric chroma $C^\star_{uv}$ e il metric hue angle h_{uv} così definiti:

$$C^\star_{uv} = [(\Delta u^\star)^2 + (\Delta v^\star)^2]^{1/2}$$

$$h_{uv} = \arctan(\frac{v^\star}{u^\star})$$

In Figura 6.10 le tra grandezze sopra descritte sono rappresentate per il punto P_1. Gli angoli che rappresentano la tinta sono espressi in gradi, da 0° a 360°, misurati in senso antiorario a partire dall'asse $u^\star > 0$.

Un'idea grossolana delle tinte nel piano $(u^\star, v^\star)$ si può avere considerando che l'asse $u^\star > 0$ corrisponde ai rossi, l'asse $v^\star > 0$ corrisponde ai gialli, l'asse $u^\star < 0$ corrisponde ai verdi, ed infine l'asse $v^\star > 0$ corrisponde ai blu.

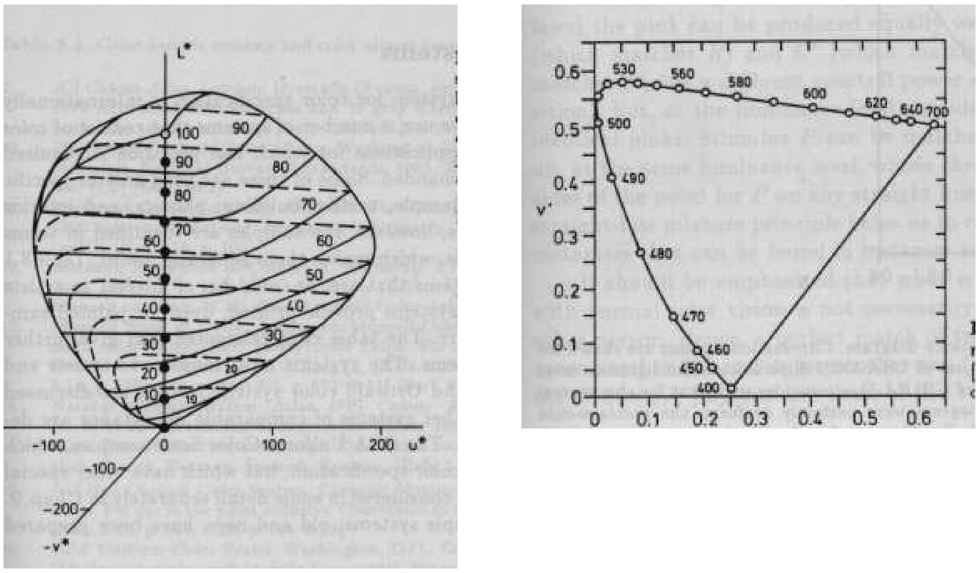


Figura 6.9 Spazio CIELUV e diagramma (u', v').

6.5 Spazi colore dipendenti dai dispositivi

Gli spazi colore sino ad ora analizzati sono stati costruiti a partire dai risultati di esperimenti di corrispondenza tricromatica e da esperimenti volti a misurare quantitativamente le differenze tra colori che un osservatore percepisce. Sono spazi colore *indipendenti dai dipositivi*.

I dispositivi per la riproduzione del colore, quali per esempio i monitor e le stampanti, usano per la rappresentazione del colore degli spazi che sono determinati dal modo in cui il colore viene generato. I più comuni spazi colore *dipendenti dai dispositivi* sono l'*RGB* (red, green, blue) usato diffusamente per i monitor e il *CMY* (cyan, magenta, yellow) o il *CMYB* (cyan, magenta, yellow, black) per le stampanti.

Bibliografia

- G. A. Agoston.: *Color Theory and its Application in Art and Design*, Springer-Verlag Series in Optical Sciences, Berlin 1987.
- R. W. G. Hunt.: *Measuring Colour* - Second Edition -, Ellis Horwood, London 1995.
- G. Wyszecki, W. S. Stiles.: *Color Science. Concept and Methods, Quantitative Data and Formulae* - Second Edition -, Jhon Wiley & Sons, Toronto 1982.

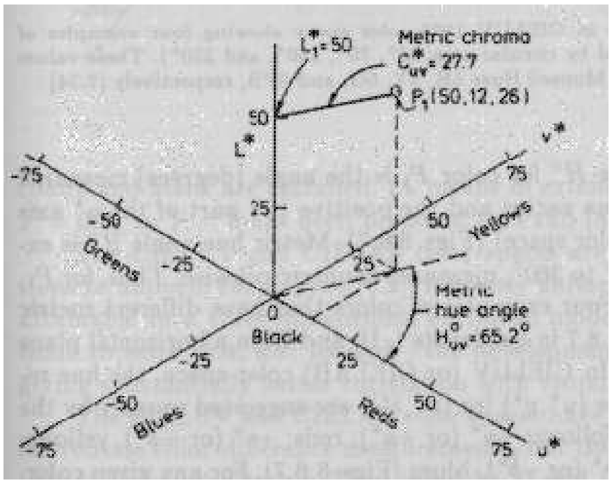


Figura 6.10 Lightness, chroma, hue.

Capitolo 7

Immagini digitali

7.1 Introduzione

Il processo di formazione di una immagine può essere molto diverso da quello che abbiamo analizzato (nelle immagini mediche ciò che spesso viene misurato è il coefficiente di assorbimento dei tessuti rispetto a particolari radiazioni elettromagnetiche); in ogni caso è importante ai fini della elaborazione essere consapevoli di ciò che è rappresentato nell'immagine.

Qualunque sia il tipo di immagini che vengono trattate, il piano immagine è un sottospazio di R^2 e il codominio è limitato, diciamo che è costituito dall' intervallo $[0, f_{max}]$.

Una funzione immagine interessante 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 venga campionata e quantizzata.

Campionare una immagine significa prendere sul piano immagine un numero finito di punti (il teorema del campionamento stabilisce le condizioni necessarie per non avere perdita di informazione). *Quantizzare* significa prendere un numero finito di valori interi nell'intervallo $[0, f_{max}]$. Naturalmente un campionamento ed una quantizzazione inadeguati comportano una grave perdita di informazione. In particolare il numero di livelli di quantizzazione deve essere tale da non far percepire salti di livello di grigio. Se il numero di livelli di grigio in cui viene suddiviso l'intervallo $[0, f_{max}]$ è insufficiente si percepiscono falsi contorni Ciò può accadere se i livelli di grigio sono meno di 50, numero che noi possiamo distinguere bene. In genere si usano 256 livelli di grigio, ma per alcune applicazioni questa risoluzione può risultare grossolana. In ambito medico le immaginini NMR (Nuclear Magnetic Resonance) utilizzano 4096 livelli di grigio o più.

Vediamo più in dettaglio il processo di quantizzazione.

7.2 Quantizzazione

Dato un numero reale, un *quantizzatore scalare* seleziona da un insieme predeterminato e finito di valori numerici il valore che meglio approssima il numero reale dato.

L'ingresso di un quantizzatore può assumere valori in un intervallo continuo, l' uscita è un numero univocamente specificato da un intero dell'insieme $\{1, 2, 3, \ldots, N\}$. Più precisamente, definiamo un quantizzatore scalare (monodimensionale) Q con N elementi mediante una funzione $Q : R \rightarrow C$ dall'insieme dei reali (o da un suo sottinsieme) ad un insieme $C = \{y_1, y_2, y_3, \ldots y_N\} \subset R$ che è chiamato *codebook*; ovviamente la dimensione del codebook è N. I valori y_i sono chiamati *livelli di uscita*, *punti di uscita* o *valori di riproduzione*. Si assume la convenzione di indicare i valori di uscita in modo che valga

$$y_1 < y_2 < y_3 < \ldots < y_N$$

Viene definita *risoluzione* di un quantizzatore scalare la quantità $r = \log N$ che misura il numero di bit necessari a specificare univocamente il valore quantizzato. La risoluzione indica l'accuratezza con cui è descritta l'ampiezza del segnale analogico originale. Nel caso in cui r sia un intero si può assegnare un' unica parola binaria ad ogni y_i. In questo caso si ha un *fixed rate* code; alternativamente diversi valori quantizzati possono essere specificati con parole che hanno un diverso numero di digit (si parla in tal caso di *variable rate* code).

Ad ogni quantizzatore scalare Q con N elementi è associata una *partizione* della retta reale in N *celle* o *atomi* R_i, con $i = 1, 2, \ldots N$. La iesima cella è così definita: $R_i = \{x \in R : Q(x) = y_i\}$, altrimenti detto: $R_i = Q^{-1}(y_i)$. Dalla definizione segue che $\cup_i R_i = R$ e che $R_j \cap R_i = \emptyset$ se $i \neq j$. Una cella che è non limitata è detta *overload (sovraccarica)*, una cella limitata è detta *granulare*. L'insieme di tutte le celle (overload) granulari è chiamata *regione (overload) granulare* .

Un quantizzatore Q è completamente descritto dai suoi livelli di uscita $\{y_i; i = 1, 2, \ldots, N\}$ e dalle corrispondenti celle della partizione $\{R_i; i = 1, 2, \ldots, N\}$; una scrittura che rende questa descrizione esplicita è $Q = \{y_i, R_i\ i = 1, 2, \ldots, N\}$.

Ogni quantizzatore può essere visto come costituito dalla composizione di un codificatore E e di un decodificatore D. Il codificatore è una funzione $E : R \rightarrow \{1, 2, \ldots, N\}$ e il decodificatore è una funzione $D : \{1, 2, \ldots, N\} \rightarrow C$. Pertanto $Q(x) = D(E(x))$, se $Q(x) = y_i$ allora $E(x) = i$ e $D(i) = y_i$.

7.2.1 Quantizzatore regolare e quantizzatore uniforme

Un quantizzatore è definito *regolare* se:

- ogni cella R_i è un intervallo

- $y_i \in (x_{i-1}, x_i)$

I valori x_i sono detti *punti di confine*, *punti di decisione* o *livelli di decisione*. Un quantizzatore regolare ha la proprietà caratteristica che se due valori di input, diciamo a e b con $a < b$ sono quantizzati con lo stesso valore di output w, ogni valore di input c compreso tra a e b sarà quantizzato con w. Questa proprietà corrisponde alla nostra nozione intuitiva di quantizzazione come modo per approssimare un numero reale con una rappresentazione a precisione finita.

Una descrizione completa di un quantizzatore regolare si ha assegnando i punti di confine:

$$R_i = (x_{i-1}, x_i] = \{r : (x_{i-1} < r \leq x_i\} \quad i = 1, 2, \ldots N$$

dove con $(a, b]$ si intende l'intervallo (a, b) insieme con l'estremo destro b. Nel caso in cui x_N sia infinito allora $R_N = (x_{N-1}, x_N)$. Per un quantizzatore regolare, gli estremi degli intervalli sono così ordinati

$$x_0 < y_1 < x_1 < y_2 < x_2 < \ldots < y_N < x_N$$

La sequenza degli estremi degli intervalli e dei valori di uscita specifica univocamente il quantizzatore.

Di solito $x_0 = -\infty$ e $x_N = +\infty$ e le celle R_0, R_N sono overload cells. Il *range* B di un quantizzatore è definito come la lunghezza totale delle celle granulari, pertanto per quantizzatori regolari unbounded $B = x_{N-1} - x_1$. Accade spesso però che i valori di ingresso possano variare in un sottoinsieme limitato della retta reale; in particolare se i valori di ingresso appartengono ad una regione limitata allora x_0 e x_N assumono valori finiti e $B = x_N - x_0$.

Il più comune tra i quantizzatori scalari è il quantizzatore *uniforme*, chiamato anche quantizzatore *lineare*. L'aggettivo lineare, improprio poichè tutte le operazioni di quantizzazione sono inerentemente non lineari, deriva dal fatto che la risposta del quantizzatore ha la forma di una scala che si sviluppa su una retta di pendenza 1. Esempi di quantizzazione uniforme sono le operazioni di troncamento e di arrotondamento di numeri reali; inoltre la gran parte dei convertitori anologico-digitali di uso generale effettua una quantizzazione uniforme.

Un quantizzatore *uniforme* è un quantizzatore *regolare* in cui:

- i punti di confine sono equispaziati
- i livelli di uscita delle celle granulari sono i punti centrali dell'intevallo di quantizzazione.

La prima condizione implica che $x_i - x_{i-1} = \Delta$ per $i = 2, 3, \ldots, N$ e la seconda condizione implica che $y_i = (x_{i-1} + x_i)/2$ per $i = 2, 3, \ldots, N-1$. Pertanto la partizione della regione granulare consiste di intervalli di lunghezza Δ e per gli ingressi non limitati il

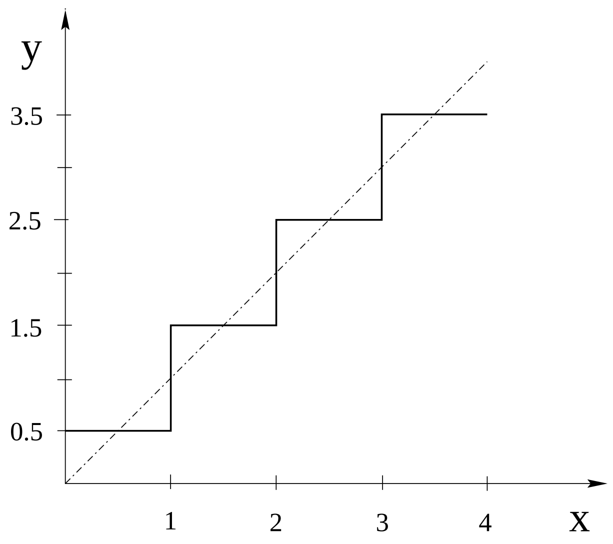


Figura 7.1 Esempio di quantizzatore uniforme

quantizzatore ha le celle overload $(-\infty, x_1)$ e $(x_{N-1}, +\infty)$ con $y_1 = \Delta/2$ e $y_N = x_{N-1}+\Delta/2$. Notiamo anche che $y_i = x_{i-1} + \Delta/2$. Pertanto Δ è l'ampiezza dell'intervallo che separa sia due punti di decisione consecutivi che due valori di uscita consecutivi.

Consideriamo qui solo il caso in cui il segnale in ingresso per il quantizzatore sia limitato, in particolare che assuma valori nell'intervallo (a, b) di ampiezza $B = b - a$. Allora $a = x_0$ e $b = x_N$ e l' intervallo è suddiviso in N celle di quantizzazione ognuna di ampiezza B/N. Il quantizzatore risultante ha N livelli di output, anche il fronte di salita del gradino è Δ (vedi esempio in figura 7.1).

7.2.2 Stima delle prestazioni di un quantizzatore

L'obiettivo della quantizzazione è quello di fornire una descrizione di precisione limitata di un valore di ingresso non noto. E' solo perchè l'ingresso non è noto in anticipo che è necessario quantizzare. Pertanto l'ingresso deve essere modellato come una variabile aleatoria di solito specificata da una funzione di densità di probabilità. Di conseguenza, l'errore introdotto nella quantizzazione sarà anch'esso una variabile casuale.

Per stimare le prestazioni di un particolare quantizzatore è necessario definire una misura che, in un unico valore numerico, esprima la degradazione o *distorsione* che è prodotta su ogni ingresso. Inoltre è necessaria una misura della prestazione globale, di solito basata sulla media, che tenga conto della distribuzione di probabilità degli ingressi.

La più comune misura di distorsione tra numeri è l'errore quadratico:

$$d(x, \widehat{x}) = \mid x - \widehat{x} \mid^2$$

Applicata ad un quantizzatore, x è l'ingresso e $\widehat{x} = Q(x)$ è il valore di uscita. Un'altra misura di interesse è l'errore assoluto $\mid x - \widehat{x} \mid$. Entrambe le misure sono casi particolari di una più generale misura di distorsione spesso usata nello studio dei quantizzatori scalari:

$$d_m(x, \widehat{x}) = \mid x - \widehat{x} \mid^m$$

I casi speciali in cui $m = 2$ o $m = 1$ sono i più popolari.

Come *misura della prestazione globale* di un quantizzatore si può prendere o l'errore nel caso peggiore o effettuare la media di una adatta misura di distorsione. Se gli ingressi sono di ampiezza limitata l' *errore assoluto nel caso peggiore*, spesso chiamato semplicemente errore massimo, è il massimo di $d_m(x, \widehat{x})$ preso su tutti i possibili ingressi x. Osserviamo che l'errore massimo dipende dal supporto della funzione densità di probabilità degli ingressi, cioè dall'insieme di valori dell'ingresso per cui la funzione densità è diversa da 0. Per ingressi non limitati e un quantizzatore con un numero N finito di livelli di uscita, l'errore nel caso peggiore è infinito pertanto è una misura priva di significato.

Una misura di prestazioni più significativa, e che fornisce più informazione dell'errore massimo, è il *valor medio della distorsione*, denotato con D e così definito

$$D = E[(d(X, Q(X))] = \int_{\infty}^{\infty} d(x, Q(x)) f_X(x) dx$$

dove $f_X(x)$ è la funzione densità di probabilità della variabile aleatoria X ed E denota l'aspettazione di una variabile aleatoria. Focalizziamo la nostra attenzione sull' aspettazione della misura di distorsione errore quadratico e chiamiamo tale misura *distorsione media* o errore quadratico medio (MSE); questa è la più usata tra le misure di distorsione.

Data la variabile aleatoria X, che rappresenta l'ingresso di un quantizzatore $Q = y_i, R_i; i = 1, \ldots, N$, la distorsione media è data da

$$D = E[(X - Q(X))^2] = \sum_{i=1}^{N} \int_{R_i} (x - y_i)^2 f_X(x) dx$$

Se il quantizzatore è regolare la *distorsione media* è data da

$$D = E[(X - Q(X))^2] = \sum_{i=1}^{N} \int_{x_{i-1}}^{x_i} (x - y_i)^2 f_X(x) dx$$

L' errore quadratico medio è semplice da usare ma è noto che non corrisponde bene alle valutazioni percettive della qualità dei segnali uditivi e visivi destinati alla percezione umana. A causa della sua trattabilità analitica consente di ottenere dei risultati quantitativi che sono utili nella pratica e che costituiscono un primo passo per trattare i problemi più difficili di ottenere prestazioni orientate alle caratteristiche del sistema percettivo.

Valutiamo ora le prestazioni di un quantizzatore uniforme Q rispetto alle misure sopra definite. Se Q è un quantizzatore con N valori di uscita e il cui ingresso è costituito dall'intervallo (a, b) di ampiezza B allora l'errore massimo di Q è $B/2N = \Delta/2$. Non è difficile mostrare che il quantizzatore uniforme minimizza l'errore massimo, indipendentemente dalla funzione densità di probabilità su (a, b). Pertanto esso ha buone prestazione per una grande varietà di segnali in ingresso e ciò lo rende robusto. Per questa ragione e per la sua semplicità è molto diffuso.

Consideriamo ora il caso in cui la funzione densità dell'ingresso sia la distribuzione uniforme su (a, b). Osserviamo che l'errore di quantizzazione $\epsilon = Q(X) - X$ ha una distribuzione uniforme su $[-\Delta/2, \Delta/2]$ dove $\Delta = B/N$, è l'ampiezza di ogni cella di quantizzazione. Pertanto la media dell'errore di quantizzazione è

$$E(\epsilon) = 0$$

e la *distorsione media*, data la cella, è

$$D = E[(Q(X) - X)^2] = E[\epsilon^2] = \Delta^2/12$$

La distorsione media, data la cella, è semplicemente la varianza di una variabile aleatoria che è uniformemente distribuita sull' intervallo di ampiezza Δ.

Si può dimostrare che il quantizzatore uniforme è ottimo, nel senso che minimizza sia la distorsione media sia l'errore assoluto medio, solo nel caso in cui l'ingresso sia uniformemente distribuito. Più la funzione densità dell'input si allontana dalla distribuzione uniforme maggiore è il guadagno che si ottiene effettuando una quantizzazione non uniforme invece della quantizzazione uniforme.

7.2.3 Funzione immagine digitale

In genere si chiama *funzione immagine discreta* la funzione campionata e *funzione immagine digitale* la funzione campionata e quantizzata. D'ora in poi per immagine intenderemo la funzione immagine digitale; essa è rappresentata da una matrice rettangolare NxM.

Gli elementi della matrice sono celle rettangolari, raramente esagonali. Con il termine *pixel* o *pel* (picture element) si denota il generico elemento della matrice cioè: $\langle (i, j), f(i, j) \rangle$.

Fissati N, M ed il numero di livelli di grigio G, che per semplicità assumiamo siano potenze di 2, avremo che il numero di bit necessari a memorizzare l'immagine è $N \cdot M \cdot \lg G$.

Le immagini a due soli livelli di grigio sono dette *binarie*. Le immagini *multispettrali*, acquisite da più sensori ognuno dei quali opera in un ambito limitato di lunghezze d'onda, sono descritte da piani immagine indipendenti e quindi da un vettore i cui elementi sono funzioni immagine. Le immagini a colori sono immagini multispettrali descritte da un vettore a tre componenti (rossa, verde, blu): $\underline{f}(i, j) : [f_R(i, j), f_G(i, j), f_B(i, j)]$. I satelliti trasmettono su più di 3 bande spettrali (ad esempio LANSAT4 trasmette su 5 bande).

L'elaborazione di immagini a colori può avvenire separatamente sulle 3 componenti (in tal caso esse vengono trattate come componenti monocromatiche) o in diversi spazi colore. Per la compressione di immagini video accade che la componente di luminanza sia codificata separatamente da quella cromatica.

Una immagine a colori può essere resa monocromatica seguendo metodi diversi:

- considerare una sola delle bande; è un metodo sensato solo nel caso in cui tale banda contenga tutte le informazioni di interesse (es: nelle immagini da satellite la banda ad infrarossi)

- prendere la media tra le bande (è il valore di I nello spazio HSI)
- prendere una media pesata delle tre componenti. E' ciò che fanno gli standard televisivi, NTSC (definito dal National Television Standard Committee nel 1951) e PAL (Phase Alternating Line). I coefficienti con cui pesare le quantità R, G e B sono stati opportunamente valutati.
- effettuare la trasformazione dallo spazio RGB allo spazio CIELUV o CIELAB e prendere la metric ligthness $L^\star$.

Una immagine monocromatica può essere trasformata in una immagine a colori per visualizzare meglio particolari di interesse. Si parla in tali casi di falso colore poichè il colore è semplicemente un identificatore.

7.3 Relazioni tra pixels

Se l'immagine digitale è costituita da una griglia rettangolare ci sono due possibilità di definire i vicini di un pixel. Possiamo considerare pixel vicini quelli che hanno un lato in comune con il pixel di interesse o quelli che hanno con esso un vertice in comune; parliamo rispettivamente di *4-vicini* e di *8-vicini*.

Due pixel p e q sono *4-vicini* se p è contenuto nel *4-intorno* $N_4(q)$ di q, e q è contenuto nel *4-intorno* $N_4(p)$ di p. Dato p di coordinate (i, j)

$$N_4(p) = \{(i, j-1), (i, j+1), (i-1, j), (i+1, j)\}$$

Una analoga definizione viene data per la *8-vicinanza*. Entrambe le nozioni di vicinanza sono necessarie.

Consideriamo ora la nozione di distanza. Le principali nozioni di distanza usate sulle immagini sono tre :

1. la usuale *distanza euclidea* (essendo l'immagine a coordinate intere, questa distanza è costosa da calcolare)
2. la *city-block distance*
$$d_4(p, q) = \mid i_p - i_q \mid + \mid j_p - j_q \mid$$
(è come si ci si potesse muovere solo lungo le righe o lungo le colonne; pixel 4-vicini hanno distanza $d_4 = 1$)
3. la *diagonal distance*
$$d_8(p, q) = \max\{\mid i_p - i_q \mid, \mid j_p - j_q \mid\}$$
(è come si ci si potesse muovere solo in diagonale; pixel 8-vicini hanno distanza $d_8 = 1$)

Un *4-cammino (8-cammino)* π di lunghezza n da p a q è una sequenza di pixel $p = p_0, p_1, \ldots, p_n = q$ tale che p_i è 4-vicino (8-vicino) di p_{i-1} $(1 \le i \le n)$. Pertanto la $d_4(p, q)$ e la $d_8(p, q)$ denotano la lunghezza, valutata rispettivamente secondo la 4 e la 8-metrica, dei cammini più corti (di questi ce ne sono molti) tra p e q.

7.3.1 Insiemi connessi

Consideriamo la nozione di *connessione*. Essa ci consente di individuare in una immagine sottoinsiemi, *regioni*, con alcune proprietà.

Definito un insieme A di valori di una funzione immagine f, individuiamo il sottoinsieme S di pixel che assumono valori in A. S può ovviamente essere rappresentato da una immagine binaria B_S, delle stesse dimensioni di f, che rappresenta la funzione caratteristica di S cioé:

$$\forall p \in S \quad B_S(p) = 1 \quad e \quad \forall q \in S^c \quad B_S(q) = 0$$

doveS^c è il complemento di S rispetto al dominio dell' immagine.

Dati due pixel p e q appartenenti all'insieme S, diciamo che p è *4-connesso (8-connesso)* a q se esiste un 4-cammino (8-cammino) da p a q costituito da pixel appartenenti all'insieme S.

E' importante osservare che dati $p, q, r \in S$ la relazione *è connesso a* è simmetrica e transitiva; inoltre assumendo che un cammino possa avere lunghezza 0 possiamo dire che *p è connesso a p*. Pertanto la relazione *è connesso a* è una relazione di equivalenza.

Per ogni pixel p in S l' insieme di punti connessi a p è denominato componente connessa. Se S ha una sola *componente connessa* è chiamato insieme *semplicemente connesso*.

Le definizioni di vicinanza e quelle di connessione, che su esse si basano, creano evidenti paradossi (vedi figura). Considerando la 4-connessione, l'insieme di pixel neri in Figura 7.2 è costituito da due regioni distinte; al contrario, considerando la 8-connessione, esso è costituito da una sola regione. Se si usa la 4-connessione nessuno dei pixel sulla linea è connesso agli altri, inoltre due linee perpendicolari si intersecano in un caso e non nell'altro (poichè non hanno nessun pixel in comune).

Questi esempi mostrano che nessuna delle due convenzioni è sufficiente, spesso si adotta la 8-connessione per le regioni che corrispondono ad oggetti e la 4-connessione per lo sfondo.

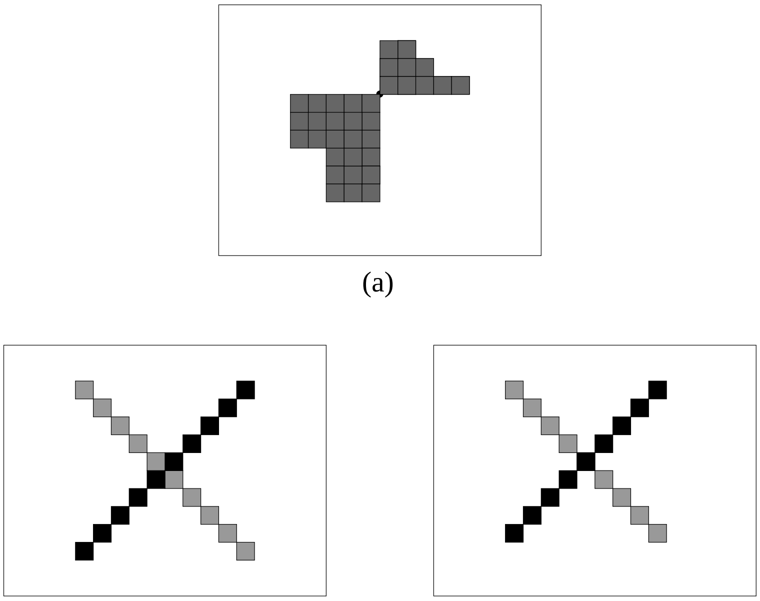
Da quanto detto è evidente che i concetti di punto, retta, segmento rettilineo, elemento convesso non hanno un' immediata ed intuitiva corrispondenza con gli analoghi concetti della geometria euclidea, devono essere pertanto ridefiniti. Ci sono diverse definizioni possibili.

Poichè molti metodi di elaborazioen di immagini producono un' immagine binaria, è estremamente importante costruire un algoritmo che riconosce (etichetta) gli insiemi connessi.

Algoritmo per etichettare gli insiemi connessi (usa la 4-connessione)

1. Scandisci l'immagine da sinistra a destra, dall'alto verso il basso.

2. Se il pixel ha valore 1 allora



(a) (b) (c)

Figura 7.2 Esempi di insiemi connessi

- Se uno dei suoi vicini superiore o sinistro ha un' etichetta, copia l'etichetta.
- Se entrambi i vicini hanno la stessa etichetta, copia l' etichetta.
- Se hanno etichette diverse copia quella superiore e memorizza, nella tabella delle etichette, che sono etichette equivalenti.
- Altrimenti assegna una nuova etichetta al pixel corrente e poni l'etichetta nella tabella.

3. Se ci sono altri pixels da considerare vai al passo 2.

4. Trova l'etichetta minore in ogni classe di equivalenza.

5. Scandisci l'immagine, sostituisci ogni etichetta con quella minore in ogni classe di equivalenza.

7.3.2 Buchi, sfondo, confine

Abbiamo visto che la relazione *è connesso a* consente di suddividere l'immagine in classi di equivalenza (regioni).

Supponiamo che ci siano in una immagine i regioni $(R_1, \ldots, R_i)$ che abbiamo individuato mediante l'algoritmo per la ricerca degli insiemi connessi, e supponiamo che nessuna di esse tocchi i bordi dell'immagine. Ciò significa che nessuno dei pixel appartenenti ad una regione è 4-vicino di un pixel appartenente alla prima o all' ultima riga dell'immagine, alla colonna di destra o a quella di sinistra. Tali righe e colonne costituiscono i *bordi* dell'immagine.

Sia $R = \cup R_i$, definiamo R^c il complemento di R rispetto al dominio dell'immagine. Il sottoinsieme di R^c che è contiguo ai bordi dell'immagine è detto *sfondo* (background); i restanti insiemi che eventualmente costituiscono R^c sono detti *buchi*. Se una regione non contiene buchi si dice che è *semplicemente connessa*, se contiene buchi è *a connessione multipla*.

Diciamo che la regione R_i *circonda* la regione R_j se qualunque cammino, da qualsivoglia pixel di R_j ai bordi dell'immagine, incontra R_i. Evidentemente lo sfondo circonda almeno un insieme in R^c, inoltre ogni regione circonda un suo buco.

Se C è il numero di insiemi connessi di una immagine ed H è il numero di buchi il *numero di Eulero* è così definito:

$$E = C - H$$

Il *contorno interno* di una regione R è il sottoinsieme di R i cui pixel hanno distanza 1 da R^c secondo la metrica adottata per R^c. Data questa definizione il contorno interno risulta essere una curva connessa secondo la metrica adottata per R. L'insieme dei pixel che non appartengono al contorno interno di R costituisce l'*interno* di R.

Il *contorno esterno* di una regione R è il contorno interno dell'insieme complemento R^c.

Nel caso in cui la metrica adottata per R sia la 8-metrica, un pixel $p \in R$ è un elemento del contorno interno se soddisfa il seguente predicato:

$$p\ AND(NOTp_n\ \ OR\ \ NOTp_s\ \ OR\ \ NOTp_e\ \ OR\ \ NOTp_w)$$

dove p_n, p_s, p_e, p_w sono i 4-vicini di p situati rispettivamente a nord, sud, est, ovest di p.

L' algoritmo che segue descrive un metodo per costruire il contorno interno di regioni composte da almeno due pixel. L'algoritmo suppone che si usi la 8-connessione per descrivere la regione.

Algoritmo di inseguimento del contorno

1. Scandisci l'immagine in modo sistematico, da sinistra a destra e dall'alto in basso, e trova il primo pixel $s \in S$ della regione di interesse.

2. Sia c il pixel corrente nel contorno che si sta costruendo. Assegna a c il pixel s. Sia b il 4-vicino sinistro di c; naturalmente $b \in S^c$.

3. Siano $n_1, n_2, \ldots, n_8$ gli 8-vicini di c considerati a partire da b e procedendo in senso orario. Sia n_i il primo pixel che incontri in S.

4. Assegna n_i a c e n_{i-1} a b.

5. Ripeti i passi 3. e 4. fino a quando $c = s$.

7.3.3 Elementi che descrivono la forma di regioni

Per descrivere la forma di una regione R sono comunemente usate le seguenti caratteristiche: area, perimetro, compattezza, baricentro, orientazione rispetto ad un asse, elongazione. Per le definizioni date nel seguito supponiamo di rappresentare una regione R mediante la sua funzione caratteristica cioè mediante un' immagine binaria $B(i,j)$ tale che: $B(i,j) = 1$ se $(i,j) \in R$, $B(i,j) = 0$ altrimenti.
L' *area* A di una regione R è il numero di pixels in R. Formalmente

$$A = \sum_{i=0}^{N-1} \sum_{j=0}^{M-1} B(i,j)$$

Naturalmente per calcolare l'area dell'oggetto che la regione rappresenta occorre conoscere l'area che, nella scena, corrisponde ad ogni pixel.
Una semplice definizione di *perimetro* di una regione priva di buchi è la sequenza di pixel che ne costituiscono il contorno interno. La *lunghezza l del perimetro* è il numero di pixel del contorno. Un diverso modo di calcolare la lunghezza consiste nell'effettuare la somma delle distanze euclidee tra i pixels successivi del contorno (la distanza è ovviamente 1 o $\sqrt{2}$).
La *Compattezza* C di una regione è così definita

$$C = l^2/A$$

La figura più compatta è il cerchio. Per le forme digitali il valore della compattezza non è minimo per il cerchio digitale ma per l'ottagono digitale o per il diamond, ciò dipendendo dal modo in cui si calcola la lunghezza del perimetro.

E' spesso necessario definire la posizione e la orientazione di una regione. Per descrivere la posizione si utilizzano diversi metodi. Si può utilizzare il più piccolo rettangolo contenente la regione i cui lati siano paralleli alle righe e alle colonne dell'immagine oppure, determinare il centroide della regione.
Il *Centroide* è il punto di coordinate (x_c, y_c) così definite

$$x_c = \frac{\sum_{i=0}^{N-1} \sum_{j=0}^{M-1} jB(i,j)}{A} \qquad y_c = \frac{\sum_{i=0}^{N-1} \sum_{j=0}^{M-1} iB(i,j)}{A}$$

Le coordinate (x_c, y_c) non hanno, in generale, valori interi.

Per ogni coppia (l,k) di interi non negativi definiamo il *momento digitale* m_{lk} di ordine $(l+k)$ di $B(i,j)$ nel seguente modo

$$\sum_{i=0}^{N-1} \sum_{j=0}^{M-1} i^l j^k B(i,j)$$

Osserviamo che l'area e il centroide possono essere descritti medianti i momenti m_{00}, m_{10} e m_{01}. Infatti

$$A = m_{00} \quad x_c = \frac{m_{01}}{m_{00}} \quad y_c = \frac{m_{10}}{m_{00}}$$

Usando il centroide vengono definiti i momenti centrali. Il *momento centrale* μ_{lk} è così definito

$$\mu_{lk} = \sum_{i=0}^{N-1} \sum_{j=0}^{M-1} (j - x_c)^l (i - y_c)^k B(i,j)$$

I momenti centrali sono invarianti per traslazione.

I momenti centrali di ordine 2 vengono utilizzati per determinare *l'orientazione* di una regione. Essi sono così definiti

$$\mu_{02} = \sum_{i=0}^{N-1} \sum_{j=0}^{M-1} (i - y_c)^2 B(i,j)$$

$$\mu_{11} = \sum_{i=0}^{N-1} \sum_{j=0}^{M-1} (j - x_c)(i - y_c) B(i,j)$$

$$\mu_{20} = \sum_{i=0}^{N-1} \sum_{j=0}^{M-1} (j - x_c)^2 B(i,j)$$

Affinchè sia possibile definire l'orientazione di una regione R è necessario che essa sia allungata rispetto ad un asse, nel qual caso si può usare l'orientazione dell'asse per definirne l'orientazione. Per determinare tale asse definiamo la distanza $\Phi^2(r)$ tra la regione e una generica retta r e chiamiamo *asse di elongazione* la retta $\widehat{r}$ che minimizza la funzione $\Phi^2(r)$.

Formalmente, indicando con $d_r^2(x_j, y_i)$ il quadrato della distanza tra la retta r e il punto di coordinate (x_j, y_i) (centro del pixel di posizione (i,j), $x_j = j - (M-1)/2$ e $y_i = -(i - (N-1)/2)$)

$$\Phi^2(r) = \sum_{i=0}^{N-1} \sum_{j=0}^{M-1} d_r^2(x_j, y_i) B(i,j)$$

L'asse di elongazione $\widehat{r}$, è la retta che minimizza $\Phi^2(r)$

$$\widehat{r} = argmin_r \;\; \Phi^2(r)$$

Al fine di derivare i parametri che caratterizzano la retta $\widehat{r}$ osserviamo che una generica retta r può essere descritta in forma polare mediante la seguente equazione

$$x \cos\theta + y \sin\theta - \rho = 0$$

dove θ è l'angolo che la normale alla retta forma con l'asse x e ρ è la distanza della retta dall'origine. Inoltre $d_r^2(x_j, y_i)$, il quadrato della distanza tra il punto di coordinate (x_j, y_i) e la retta r di parametri ρ e θ , è dato da

$$d_r^2(x_j, y_i) = (x_j \cos\theta + y_i \sin\theta - \rho)$$

Possiamo ora rappresentare la distanza della regione da una retta in funzione dei parametri ρ, θ

$$\Phi^2(\theta, \rho) = \sum_{i=0}^{N-1} \sum_{j=0}^{M-1} (x_j \cos\theta + y_i \sin\theta - \rho)^2 B(i,j)$$

Condizioni necessarie di minimo sono:

$$\frac{\partial \Phi^2(\theta,\rho)}{\partial \rho} = 0 \quad \frac{\partial \Phi^2(\theta,\rho)}{\partial \theta} = 0$$

Determinando il parametro ρ mediante l'equazione $\frac{\partial \Phi^2(\theta,\rho)}{\partial \rho} = 0$ si ottiene $\rho = x_c \cos\theta + y_c \sin\theta$ dove (x_c, y_c) sono le coordinate del baricentro della regione. Tale equazione descrive il fascio di rette che passa per il baricentro.

Sostituendo il valore di $\rho = x_c \cos\theta + y_c \sin\theta$ ed effettuando semplici manipolazioni si ottiene

$$Y^2(\theta) = \Phi^2(\theta, x_c \cos\theta - y_c \sin\theta) = 1/2(a+c) + 1/2(a-c)\cos 2\theta + 1/2 b \sin 2\theta$$

dove $a = \mu_{20}, \quad b = 2\mu_{11} \quad c = \mu_{02}$

La funzione $Y^2(\theta)$ ha un massimo θ_{max} e un minimo θ_{min} che differiscono di $\pi/2$ e si ottengono dall'equazione

$$\tan 2\theta = \frac{b}{a-c}$$

L'asse di elongazione è la retta passante per il baricentro con parametro θ_{min}.

Per determinare quanto la regione è allungata si può usare il valore dell' *elongazione* *El* così definita:

$$El = \frac{Y(\theta_{max})}{Y(\theta_{min})}$$

7.3.4 Proiezioni

Le *proiezioni* di una immagine binaria su una linea possono essere ottenute partizionando la linea in bins e contando il numero di pixel di valore 1 che sono sulle linee perpendicolari ad ogni bin.

La *proiezione orizzontale* $O(i), (i = 1, \ldots, N)$ è pari al numero di 1 presenti in tutte le colonne proiettate sulla riga i; la *proiezione verticale* $V(j)(i = 1, \ldots, M)$ è pari al numero di 1 presenti in tutte le righe proiettate sulla colonna j. Formalmente

$$O(i) = \sum_{j=O}^{M-1} B(i,j)$$

$$V(j) = \sum_{i=0}^{N-1} B(i,j)$$

La proiezione può essere definita rispetto a linee arbitrarie, per esempio si possono avere proiezioni diagonali.

Si verifica facilmente che i momenti di ordine 1 di una immagine binaria coincidono con quelli delle sue proiezioni. Poichè il calcolo del centroide richiede di conoscere solo i momenti m_{00}, m_{10} e m_{01} esso può essere effettuato a partire dalle proiezioni. Infatti:

$$A = \sum_{j=O}^{M-1} V(j) = \sum_{i=0}^{N-1} O(i)$$

$$x_c = \frac{\sum_{j=O}^{M-1} jV(j)}{A} \quad y_c = \frac{\sum_{i=0}^{N-1} iO(i)}{A}$$

Le proiezioni possono essere utili per determinare la posizione delle linee in un immagine di testo e la posizione dei caratteri in una linea. Infatti si effettuano le proiezioni orizzontali (verticali) e i valori di minimo denoteranno la separazione tra linee di testo (caratteri nel testo).

7.3.5 Run-Length Encoding

Una rappresentazione compatta di immagini binarie è il *run-length encoding*; su questa codifica o su sue estensioni si basano gli standard per la trasmissione via fax. L'idea di base è quella di codificare ogni blocco (run) di zeri, o di uni, tra loro contigui, che si incontra scandendo ogni riga dell'immagine da sinistra a destra, mediante la sua lunghezza. Inoltre viene fissata una convenzione per determinare se un blocco è costituito da zeri o da uni. I metodi più frequenti per indicare il valore di un blocco sono

- specificare il valore del primo blocco di ogni riga
- assumere che ogni riga cominci con un blocco di zeri, la cui lunghezza sarà 0 nel caso in cui la riga inizi con un blocco di uni.

Ogni riga dell'immagine è pertanto rappresentata da una sequenza di numeri che specificano la lunghezza dei blocchi di zeri e di uni. La proiezione orizzontale può essere facilmente calcolata a partire dal run-length code.

Codifiche più complesse tengono conto delle correlazioni tra righe.

Sebbene la codifica sopra descritta sia un metodo efficiente di compressione, una compressione ulteriore può essere ottenuta codificando le lunghezze dei blocchi mediante un codice a lunghezza variabile quale il codice di Huffman.

Lo standard 1-D CCITT (Consultative Committee of the International Telephone and Telegraph) Group 3 utilizza un run-length encoding e una codice di Huffman. Le tabelle che codificano le lunghezze dei blocchi sono state definite in modo standard (quindi non viene definita per ogni documento la codifica ottimale). Esse sono riportate sul Gonzales (pag 391, 392).

7.3.6 Descrittori del contorno: chain code

Nell'ambito dell' elaborazione di immagini si usano strutture di dati tradizionali quali le matrici, le stringhe, i grafi, le liste, gli alberi. Vediamo l'utilizzo di sequenze di simboli per descrivere il contorno di una regione.

Il *chain code* o *codice di contorno* è una struttura di dati per rappresentare in modo efficiente il contorno di una regione. Invece di memorizzare tutte le posizioni dei pixels di contorno, viene selezionato un pixel iniziale e le sue coordinate vengono memorizzate. Se usiamo un algoritmo che scandisce l'immagine linea per linea può essere utile considerare il pixel più in alto a sinistra. Quindi seguiamo il contorno in senso orario (o antiorario). Se usiamo il 4-vicinato abbiamo al più 4 direzioni in cui muoverci, se usiamo l'8 vicinato ne abbiamo 8; possiamo pertanto codificare ogni direzione con un codice di 2 oppure 3 bit. Il codice di contorno consiste pertanto delle coordinate (i, j) del pixel di partenza e di una sequenza di numeri che identificano le direzioni.

Se un oggetto ha buchi dobbiamo usare più di un codice di contorno per rappresentarlo. Dobbiamo anche memorizzare se il contorno tracciato è relativo all'oggetto oppure al buco.

Il chain code ha alcuni vantaggi ovvi rispetto alla rappresentazione matriciale di una regione:

- E' una rappresentazione compatta di un oggetto binario.
- E' una rappresentazione invariante per traslazione. Questa proprietà rende la comparazione tra oggetti più semplice.
- Fornisce una rappresentazione completa della regione; pertanto a partire dal chain code si possono calcolare alcune proprietà della forma, per esempio il perimetro, l'area, la massima larghezza e altezza, in modo più efficiente che a partire dalla rappresentazione matriciale.
- A partire dal chain code si ricostruisce facilmente l'oggetto binario. Si costruisce il contorno e si riempie.

Il chain code ha anche svantaggi:

- Non è invariante per rotazione
- Non è invariante per variazioni di scala
- E' molto sensibile al rumore. Oggetti molto simili per forma possono avere chain codes molto diversi.

Osserviamo che ogni numero c del codice di contorno è in corrispondenza biunivoca con un vettore che denotiamo con (a, b) che descrive, nel piano immagine, gli spostamenti di base rispetto ad un punto scelto come riferimento (vedi tabella che segue).

Se un contorno è chiuso allora vale

$$\sum_{i=1}^{n} a_i = \sum_{i=1}^{n} b_i = 0$$

dove n è il numero di vettori che identificano il contorno.

Il calcolo dell'area a partire dal chain code è più rapido del procedimento di contare i pixels della regione, infatti il contorno contiene solo una frazione dei pixels dell'oggetto e sono richieste solo due addizioni per ogni pixel di contorno.

L'algoritmo lavora in modo simile alla integrazione numerica. Si assume una linea di base orizzontale tracciata in un' arbitraria posizione verticale dell'immagine. Quindi si inizia l'integrazione partendo dal pixel più in alto; la distanza di questo punto dalla linea di base è B. Si segue quindi il contorno dell'oggetto e ad ogni passo descritto dal codice vengono modificate entrambe le variabili A e B aggiungendo ad esse rispettivamente gli incrementi ΔA e ΔB in accordo con la tabella sotto rappresentata.

Code	(a , b)	ΔA	ΔB
0	(1,0)	+ B	0
1	(1,1)	+ B	1
2	(0,1)	0	1
3	(-1,1)	- B	1
4	(-1,0)	- B	0
5	(-1,-1)	- B	1
6	(0,-1)	0	-1
7	(1,-1)	+ B	-1

Se, per esempio, ci si muove a destra (chain code 0) l'area aumenta di B. Se ci si muove in alto a destra (chain code 1) l'area aumenta ancora di B ma B deve essere incrementato a sua volta in quanto la distanza tra il pixel di contorno e la linea di base è aumentata. Per tutti i movimenti a sinistra l'area è diminuita di B. In questo modo, l'area della regione situata tra la linea di base orizzontale e il bordo inferiore dell'oggetto viene sottratta; tale area era stata inclusa nel calcolo quando ci si era spostati a destra.

Algoritmo

$A := 0$, sia B la distanza del primo pixel di contorno rispetto al riferimento considerato.

ad ogni passo poni: $A := A + \Delta A \quad B := B + \Delta B$

Quando l'intera sequenza di cifre del contorno è stata esplorata la quantità A assume il valore dell'area.

I valori della larghezza massima di una regione lungo le direzioni x e y definiti come

$$width_x = max\{x_i : 1 \leq i \leq n\} - min\{x_i : 1 \leq i \leq n\}$$
$$width_y = max\{y_i : 1 \leq i \leq n\} - min\{y_i : 1 \leq i \leq n\}$$

possono essere calcolati modificando iterativamente i valori:

$$x_i = \sum_{j=1}^{i} a_j + x_0 \quad i = 1, 2, \dots, n$$

$$y_i = \sum_{j=1}^{i} b_j + y_0 \quad i = 1, 2, \dots, n$$

durante l'inseguimento del contorno, (x_0, y_0) è il punto iniziale.

Bibliografia

- A. Gersho, R. M. Gray: Vector Quantization and Signal Compression. Kuever Academic Publishers, 1991.
- R. Jain, R. Kasturi, B. G. Schunk: Machine Vision. McGraw-Hill, 1995.
- R. C. Gonzales, R. E. Woods: Digital Image Processing. Addison Wesley, 1992.
- P. Zamperoni: Metodi dell'elaborazione digitale delle immagini. Masson 1990

Capitolo 8

Trasformata di Fourier Discreta

8.1 Definizione e Proprietà

Definiamo la *trasformata di Fourier discreta* in modo assiomatico e ne analizziamo le principali proprietá. Analizzeremo poi la relazione tra la trasformata discreta e la *trasformata di Fourier continua.*

Consideriamo lo spazio C^n dei vettori $[f(0), \dots, f(N-1)]$ a N componenti complesse. Tale spazio, definendo la somma tra vettori:

$$[f(0), \dots, f(N-1)] + [g(0), \dots, g(N-1)] = [f(0) + g(0), \dots, f(N-1) + g(N-1)]$$

e il prodotto di un vettore per un numero complesso:

$$\lambda[f(0), \dots, f(N-1)] = [\lambda f(0), \dots, \lambda f(N-1)]$$

é uno spazio vettoriale.

Definiamo *trasformata di Fourier discreta* la trasformazione $\mathcal{F} : C^n \to C^n$ che associa al vettore $f(\tau)(0 \le \tau \le N-1)$ il vettore $F(\nu)(0 \le \nu \le N-1)$ nel seguente modo:

$$(\mathcal{F}f)(\nu) = F(\nu) = \frac{1}{N} \sum_{\tau=0}^{N-1} f(\tau) e^{\frac{-j2\pi\nu\tau}{N}}$$

La trasformazione $\mathcal{F}$, descritta dalla matrice $\left(e^{\frac{-j2\pi\nu\tau}{N}}\right)$ é lineare e invertibile; l'inversa é data da:

$$f(\tau) = \sum_{\nu=0}^{N-1} F(\nu) e^{\frac{j2\pi\nu\tau}{N}}$$

Ciò può essere verificato utilizzando l'identità:

$$\sum_{k=0}^{N-1} e^{\frac{-j2\pi k\tau}{N}} = \begin{cases} N & \text{if } \tau = 0 \\ 0 & \text{if } \tau \neq 0 \end{cases}$$

Essa deriva dal fatto che se $\tau = 0$, allora $\sum_{k=0}^{N-1} 1 = N$, se $\tau \neq 0$ si ha invece:

$$\sum_{k=0}^{N-1} e^{\frac{-j2\pi k\tau}{N}} = \sum_{k=0}^{N-1} (e^{\frac{-j2\pi\tau}{N}})^k = \frac{(e^{\frac{-j2\pi\tau}{N}})^N - 1}{e^{\frac{-j2\pi\tau}{N}} - 1} = 0$$

Ora, per verifica diretta abbiamo che se $F(\nu) = \frac{1}{N}\sum_{\tau=0}^{N-1} f(\tau) e^{\frac{-j2\pi\nu\tau}{N}}$ allora:

$$\begin{aligned}
\sum_{\nu=0}^{N-1} F(\nu) e^{\frac{j2\pi\nu\tau'}{N}} &= \sum_{\nu=0}^{N-1} \frac{1}{N} \sum_{\tau=0}^{N-1} f(\tau) e^{\frac{-j2\pi\nu\tau}{N}} e^{\frac{j2\pi\nu\tau'}{N}} \\
&= \frac{1}{N} \sum_{\tau=0}^{N-1} f(\tau) \sum_{\nu=0}^{N-1} e^{\frac{-j2\pi\nu\tau}{N}} e^{\frac{j2\pi\nu\tau'}{N}} \\
&= \frac{1}{N} \sum_{\tau=0}^{N-1} f(\tau) \sum_{\nu=0}^{N-1} e^{\frac{-j2\pi\nu(\tau-\tau')}{N}} \\
&= \frac{1}{N} \sum_{\tau=0}^{N-1} f(\tau) \cdot \begin{cases} N & \text{if } \tau = \tau' \\ 0 & \text{if } \tau \neq \tau' \end{cases} \\
&= \frac{1}{N} \cdot N f(\tau') = f(\tau')
\end{aligned}$$

8.2 Operazioni

Nello spazio dei vettori a N componenti complesse si possono introdurre alcune importanti operazioni. Per definire queste operazioni é conveniente interpretare l'insieme $\{0, 1, \ldots, N-1\}$ come *classi di resto modulo* N con l'operazione di *somma modulo* N che denoteremo con $\langle k + j \rangle_N$ (con $\langle s \rangle_N$ si intende il resto della divisione di s con N).

Operazioni:

- **Prodotto**: $(f \cdot g)(\tau) = f(\tau) \cdot g(\tau)$
- **Shift**: $(Shift_a f)(\tau) = f(\langle \tau - a \rangle_N)$

- **Convoluzione Ciclica**: $(f \star g)(\tau) = \sum_{k=0}^{N-1} f(k) \cdot g(\langle \tau - k \rangle_N)$
- **Cross-Correlazione**: $(f \circ g)(\tau) = \sum_{k=0}^{N-1} f(k) \cdot g(\langle \tau + k \rangle_N)$

Queste operazioni hanno forti analogie con le operazioni, definite su funzioni a valori reali e complessi, che abbiamo analizzato. Non stupisce allora che la trasformata discreta goda delle seguenti proprietá che, ad eccezione dell'importante teorema di convoluzione, elenchiamo senza dimostrazione. Nel seguito denoteremo $\mathcal{F}(f)$ con F e $\mathcal{F}(g)$ con G.

Teorema della somma:

$$\mathcal{F}(f(\tau) + g(\tau)) = F(\nu) + G(\nu)$$

Teorema dello shift:

$$\mathcal{F}(f(\tau - a)) = e^{\frac{-j2\pi a\nu}{N}} F(\nu)$$

Esempio:

$$\begin{aligned} \mathcal{F}([1,0,0,0]) &= 1/4[1,1,1,1] \\ \mathcal{F}([0,1,0,0]) &= 1/4[1,-j,(-j)^2,(-j)^3] \end{aligned}$$

Teorema della somma di sequenze:

$$\sum_{\tau=0}^{N-1} f(\tau) = NF(0)$$

$$\sum_{\nu=0}^{N-1} F(\nu) = f(0)$$

Teorema di Parseval-Rayleigh:

$$\sum_{\tau=0}^{N-1} \mid f(\tau) \mid^2 = N \sum_{\nu=0}^{N-1} \mid F(\nu) \mid^2$$

Teorema di convoluzione:

$$\mathcal{F}(f(\tau) \star g(\tau)) = NF(\nu) \cdot G(\nu)$$

Dimostrazione: Sia $h = f \star g$ e $H = \mathcal{F}(h)$; per definizione di convoluzione ciclica e di trasformata discreta abbiamo:

$$h(\tau) = \sum_{k=0}^{N-1} f(k) \cdot g(\langle \tau - k \rangle_N) \quad (0 \leq \tau \leq N-1)$$

$$H(\nu) = \frac{1}{N}\sum_{\tau=0}^{N-1}\left(\sum_{k=0}^{N-1} f(k)\cdot g(\langle \tau - k\rangle_N)\right) e^{\frac{-j2\pi\nu\tau}{N}}$$

Osservando che:

$$e^{\frac{-j2\pi\nu\tau}{N}} = e^{\frac{-j2\pi\nu k}{N}} \cdot e^{\frac{-j2\pi\nu(\tau-k)}{N}}$$

possiamo scrivere:

$$H(\nu) = \frac{1}{N}\sum_{k=0}^{N-1} f(k) e^{\frac{-j2\pi\nu k}{N}} \left(\sum_{\tau=0}^{N-1} g(\langle \tau - k\rangle_N) e^{\frac{-j2\pi\nu(\tau-k)}{N}}\right)$$

poniamo $\tau - k = s$ e osserviamo che se $0 \le \tau \le N-1$ allora $\langle s\rangle_N$ varia in $\{0, 1, \ldots, N-1\}$ ed inoltre:

$$e^{\frac{-j2\pi\nu(\tau-k)}{N}} = e^{\frac{-j2\pi\nu\langle s\rangle_N}{N}}$$

Vale allora:

$$H(\nu) = \frac{1}{N}\sum_{k=0}^{N-1} f(k) e^{\frac{-j2\pi\nu k}{N}} \sum_{s=0}^{N-1} g(s) e^{\frac{-j2\pi\nu s}{N}} = NF(\nu)G(\nu)$$

■

Teorema del prodotto:

$$\mathcal{F}(f(\tau)\cdot g(\tau)) = F(\nu) \star G(\nu)$$

Abbiamo introdotto astrattamente uno spazio di vettori con varie operazioni ed una trasformazione, detta trasformata di Fourier discreta, con le sue proprietà. Discutiamo nel seguito i seguenti argomenti:

1. disegno e analisi di complessità computazionale di un algoritmo che realizza la trasformata discreta.
2. relazione tra convoluzione ciclica e convoluzione lineare
3. relazione tra la trasformata di Fourier discreta e la trasformata di Fourier continua.

Nel prossimo paragrafo viene presentato un algoritmo efficiente per il calcolo della trasformata discreta, nei paragrafi successivi vengono trattati i problemi 2. e 3.

8.3 Algoritmo per il calcolo veloce della trasformata discreta

Descriviamo in questo paragrafo un algoritmo efficiente per il calcolo della trasformata di Fourier discreta (DFT). Il calcolo della trasformata di Fourier discreta richiede di moltiplicare un vettore a N componenti $[f(0), \ldots, f(N-1)]$ per la matrice $N \times N$ la cui componente alla riga ν e colonna τ è:

$$\frac{1}{N}e^{\frac{-j2\pi\nu\tau}{N}}$$

Tale calcolo può essere ovviamente effettuato con $O(N^2)$ operazioni, richiedendo $O(N)$ operazioni il calcolo di ogni componente $F(\nu)$. (E' da notare che la valutazione dei termini della forma $e^{\frac{-j2\pi\nu\tau}{N}}$ può essere effettuata all'inizio del calcolo, il risultato può essere memorizzato).

E' possibile calcolare la DFT in modo più efficiente mediante l' algoritmo, denominato FFT (Fast Fourier Transform algorithm), descritto da Runge e Konig nel 1924 e riscoperto da Cooley e Tukey nel 1965. L'interesse di tale algoritmo deriva dal fatto che esso richiede $O(N \lg N)$ operazioni invece di $O(N^2)$ operazioni.
La trasformata discreta é definita dalla seguente espressione:

$$F(\nu) = \frac{1}{N}\sum_{\tau=0}^{N-1} e^{\frac{-j2\pi\nu\tau}{N}} \quad (0 \leq \nu \leq N-1)$$

Nella discussione che segue assumiamo, per semplicità di trattazione, che N sia una potenza di 2. Separando nella sommatoria i termini con indice pari $(\tau = 2k, \;\; 0 \leq k \leq N/2-1)$ dai termini con indice dispari $(\tau = 2k+1, \;\; 0 \leq k \leq N/2-1)$ si ottiene:

$$F(\nu) = \frac{1}{N}\sum_{k=0}^{N/2-1} f(2k)e^{\frac{-j2\pi\nu 2k}{N}} + \frac{1}{N}\sum_{k=0}^{N/2-1} f(2k+1)e^{\frac{-j2\pi\nu(2k+1)}{N}} \quad (0 \leq \nu \leq N-1)$$

da cui:

$$F(\nu) = \frac{1}{2}\left((\frac{1}{N/2}\sum_{k=0}^{N/2-1} f(2k)e^{\frac{-j2\pi\nu k}{N/2}} + e^{\frac{-j2\pi\nu}{N}}\frac{1}{N/2}\sum_{k=0}^{N/2-1} f(2k+1)e^{\frac{-j2\pi\nu k}{N/2}} \right)$$

Questa formula esprime la possibilità di calcolare la trasformata discreta del vettore a N componenti $[f(0), \ldots, f(N-1)]$ usando la trasformata discreta dei 2 vettori a $N/2$ componenti $[f(0), f(2), \ldots, f(N-2)]$ e $[f(1), f(3), \ldots, f(N-1)]$. Questo suggerisce la seguente procedura ottenuta con la tecnica divide et impera:

PROCEDURA $\mathcal{F}([f(0), \ldots, f(N-1)])$

- IF $N = 1$ THEN $F(0) := f(0)$, RETURN $f(0)$

- ELSE:
  $[p(0), p(1), \ldots, p(N/2-1)]$:= Procedura $\mathcal{F}([f(0), f(2), \ldots, f(N-2)])$
  $[d(0), d(1), \ldots, d(N/2-1)]$:= Procedura $\mathcal{F}([f(1), f(3), \ldots, f(N-1)])$
  FOR $\nu = 0, \ldots, N-1$ DO $F(\nu) := \frac{1}{2}p(\langle\nu\rangle_{N/2}) + \frac{1}{2}e^{\frac{-j2\pi\nu}{N}}d(\langle\nu\rangle_{N/2})$
  RETURN [[F(0), ... , F(N-1)]

Il calcolo della trasformata di un vettore a N componenti richiama ricorsivamente il calcolo della trasformata di 2 vettori a $N/2$ componenti. Osserviamo che le operazioni su cui l'algoritmo è costruito sono la somma e la moltiplicazione per una potenza di $e^{\frac{-j2\pi}{N}}$. Detto $T(N)$ il numero totale di operazioni vale allora:

$$T(N) = \begin{cases} 0 & N = 1 \\ 2T(N/2) + O(N) & N > 1 \end{cases}$$

Poiché la soluzione alla precedente equazione di ricorrenza è $T(N) = O(N \lg N)$, abbiamo esibito un algoritmo veloce per il calcolo della trasformata. L' algoritmo per il calcolo della la trasformata inversa è simile.

8.4 Convoluzione lineare e convoluzione ciclica

Discutiamo in questo paragrafo come la FFT possa essere usata, sotto certe ipotesi, per calcolare in modo efficiente la convoluzione di 2 segnali.

Date le funzioni $f(x)$ e $g(x)$, nulle al di fuori dell'intervallo $[0, T]$ la loro convoluzione

$$h(x) = \int_{-\infty}^{\infty} f(\alpha)g(x-\alpha)d\alpha$$

è una funzione che vale 0 al di fuori dell'intervallo $[0, 2T]$ (non [0, T]!); infatti se $x > 2T$, per ogni $0 \leq \alpha \leq T$ si ha che $g(x-\alpha) = 0$

Nell'ipotesi in cui si possa approssimare l'integrale con la somma su intervalli di ampiezza Δx, posto $N = \frac{T}{\Delta x}$ e fissando $x = n \cdot \Delta x$ $(0 \leq n \leq 2N-1)$ vale:

$$h(x) = h(n \cdot \Delta x) \sim \sum_{k=0}^{N-1} f(k \cdot \Delta x) g((n-k) \cdot \Delta x) \Delta x$$

Denotando $f(k \cdot \Delta x)$ con $\hat{f}(k)$ e $g(k \cdot \Delta x)$ con $\hat{g}(k)$, per $x = n \cdot \Delta x$ vale:

$$h(x) \sim \Delta x \sum_{k=0}^{N-1} \hat{f}(k)\hat{g}(n-k)$$

E' allora di rilievo sviluppare un algoritmo veloce per risolvere il seguente problema: dati 2 vettori N-dimensionali $[\hat{f}(0), \ldots, \hat{f}(N-1)]$ e $[\hat{g}(0), \ldots, \hat{g}(N-1)]$ calcolare il vettore 2N-dimensionale $[\hat{h}(0), \ldots, \hat{h}(2N-1)]$ con:

$$\hat{h}(n) = \sum_{k=0}^{n} \hat{f}(k)\hat{g}(n-k) \quad (0 \leq n \leq 2N-1)$$

Questa operazione tra vettori non è la convoluzione ciclica che produce infatti come risultato il vettore N-dimensionale:

$$\hat{z}(n) = \sum_{k=0}^{N-1} \hat{f}(k)\hat{g}(\langle n-k \rangle_N) \quad (0 \leq n \leq N-1)$$

Il calcolo diretto di $\hat{h}(n)$ richiede $n+1$ prodotti e n somme se $n \leq N-1$, oppure $2N-n+1$ prodotti e $2N-n$ somme se $n > N-1$. Il numero totale di prodotti e somme diventa allora $\sim 2N^2$. Tale numero di operazioni può essere ridotto osservando che $\hat{h}$ è la convoluzione ciclica dei 2 vettori 2N-dimensionali $\hat{g}_1$ e $\hat{f}_1$ ottenuti aggiungendo N zeri ai 2 vettori N-dimensionali $\hat{f}$ e a $\hat{g}$ come segue:

$$\hat{f}_1 = [\hat{f}(0), \ldots, \hat{f}(N-1), 0, \ldots, 0]$$

$$\hat{g}_1 = [\hat{g}(0), \ldots, \hat{g}(N-1), 0, \ldots, 0]$$

Quindi

$$\hat{h} = \hat{f}_1 \star \hat{g}_1$$

Anche il calcolo diretto della convoluzione ciclica richiede $O(N^2)$ operazioni. Ricordando peró che la trasformata discreta della convoluzione ciclica è il prodotto delle trasformate possiamo scrivere:

$$\hat{h} = \mathcal{F}^{-1}\mathcal{F}(\hat{f}_1 \star \hat{g}_1) = \mathcal{F}^{-1}(\mathcal{F}(\hat{f}_1) \cdot \mathcal{F}(\hat{g}_1))$$

Applicando l'algoritmo veloce per il calcolo della trasformata e della antitrasformata discreta, e osservando che il prodotto di due vettori richiede $2N$ operazioni, possiamo concludere che $\hat{h}$ può essere calcolato con $O(N \lg N)$ operazioni.

8.5 Relazione tra la trasformata di Fourier discreta e la trasformata di Fourier continua

Abbiamo definito la trasformata di Fourier discreta e la sua antitrasformata in modo assiomatico e derivato alcune proprietá dalle definizioni date. Questo approccio non risponde a questa domanda: che relazione esiste tra la trasformata di Fourier continua e la trasformata di Fourier discreta? Tale relazione è descritta dal seguente teorema che enunciamo senza dimostrazione.

Teorema [Cooley, Lewis, Welch] [1]:
If $f(x)$, $-\infty < x < \infty$, and $F(s)$, $-\infty < s < \infty$, sono una copia trasformata antitrasformata sono cioé tali che:

$$F(s) = \int_{-\infty}^{\infty} f(x) e^{-j2\pi xs} dx$$

$$f(x) = \int_{-\infty}^{\infty} F(s) e^{j2\pi xs} ds$$

[1] Cooley, J.W., Lewis, P.A.W., Welch, P.D. (1967): Application of the fast Fourir transform to computation of Fourier integrals, Fourier series, and convolution integrals. IEEE Transactions on audio and electroacoustics, Vol AU-15, N 2.

allora le funzioni periodiche $f_p(k\Delta x)$ e $F_p(j\Delta s)$, con $k = 0, 1, \dots, N-1$ e $j = 0, 1, \dots, N-1$ cosí definite:

$$Tf_p(k\Delta x) = T \sum_{l=-\infty}^{\infty} f(k\Delta x + lT)$$

$$F_p(j\Delta s) = \sum_{n=-\infty}^{\infty} F(j\Delta s + n\frac{1}{\Delta x})$$

con $\Delta s = \frac{1}{T} = \frac{1}{N\Delta x}$, sono una coppia trasformata-antitrasformata discreta.

Pertanto, se 2 funzioni sono l'una la trasformata di Fourier dell'altra, le sequenze che da esse si ottengono campionando in entrambi i domini sono una coppia trasformata-antitrasformata discreta. Poiché le sequenze di campioni sono periodiche di periodo N, considerare gli elementi di un periodo equivale a considerare un vettore a N componenti.

Affinché la trasformata di Fourier discreta sia una buona approssimazione di quella continua occorre che l' intervallo di campionamento in un dominio consenta di avere un aliasing trascurabile nell'altro dominio.

L'unica classe di funzioni per cui la trasformata discreta e quella continua coincidono (a meno di un fattore di scala) sono le funzioni periodiche a banda limitata. Ricordiamo che una funzione si dice a banda limitata se la sua trasformata di Fourier è diversa da 0 in un intervallo finito.

Se una funzione è diversa da 0 in un intervallo finito (non è pertanto a banda limitata) il campionamento produce aliasing che si puó ridurre ad un valore accettabile ma non eliminare.

Se una funzione non è diversa da 0 in un intervallo finito, e non è neppure a banda limitata ci sono errori sia di aliasing che di troncamento.

8.6 Teorema del campionamento

Prima di enunciare e dimostrare il teorema del campionamento ricordiamo la definizione della funzione $Samp$:

$$Samp(x) = \sum_{n=-\infty}^{\infty} \delta(x-n)$$

ed alcune sue proprietá:

- $Samp(\frac{x}{T}) = \sum_{n=-\infty}^{\infty} \delta(\frac{x}{T} - n) = \sum_{n=-\infty}^{\infty} \delta(\frac{1}{T}(x - nT)) = T\sum_{n=-\infty}^{\infty} \delta(x - nT)$
- $f(x) \star$ [2] $Samp(\frac{x}{T}) = T\sum_{n=-\infty}^{\infty} f(x - nT)$

[2] denotiamo l'operazione di convoluzione tra funzioni con lo stesso simbolo che abbiamo usato per denotare la convoluzione ciclica tra vettori

- $f(x) \cdot Samp(\frac{x}{T}) = T \sum_{n=-\infty}^{\infty} f(nT)\delta(x - nT)$
- $\mathcal{F}(Samp) = Samp$

  Quindi per la proprietá di variazione di scala della trasformata di Fourier:

$$\mathcal{F}(Samp(\frac{x}{T})) = TSamp(Ts) = T \sum_{n=-\infty}^{\infty} \delta(s - n\frac{1}{T})$$

Teorema del campionamento Una funzione $f(x)$ a banda limitata, cioé tale che $F(s)$, la trasformata di Fourier di $f(x)$, sia zero per tutte le frequenze superiori ad una certa frequenza s_c, è univocamente determinata da punti equispaziati e separati da un intervallo $\tau \leq 1/2s_c$.
Dimostrazione Per campionare la funzione $f(x)$ con passo di campionamento $\tau \leq 1/2s_c$ moltiplichiamo la funzione $f(x)$ con la funzione $Samp$ di periodo τ ottenendo:

$$f(x) \cdot Samp(\frac{x}{\tau}) = \tau \sum_{n=-\infty}^{\infty} f(n\tau)\delta(x - n\tau)$$

La trasformata $Q(s)$ di questa funzione è:

$$Q(s) = \mathcal{F}\left(f(x) \cdot Samp(\frac{x}{\tau})\right) = F(s) \star Samp(\tau s)\tau$$

La funzione $Q(s) = F(s) \star Samp(\tau s)$ è una funzione periodica di periodo $1/\tau$. Poiché per ipotesi $F(s)$ è a banda limitata e $1/\tau > 2s_c$, la funzione periodica $Q(s)$ è costituita da *copie* della funzione $F(s)$ che non si sovrappongono. In altre parole $Q(s)$ coincide in un periodo con $F(s)$. Possiamo pertanto ottenere $F(s)$ moltiplicando $Q(s)$ per una funzione rettangolo di base $1/\tau$ e da $F(s)$ otteniamo $f(x)$ antitrasformando. Formalmente:

$$\begin{aligned}
f(x) &= \mathcal{F}^{-1}(Q(s) \cdot rect(\tau s)) \\
&= \mathcal{F}^{-1}(Q(s)) \star \mathcal{F}^{-1}(rect(\tau s)) \\
&= \left(\tau \sum_{n=-\infty}^{\infty} f(n\tau)\delta(x - n\tau)\right) \star 1/\tau(sinc(x/\tau) \\
&= \sum_{n=-\infty}^{\infty} f(n\tau)\left(\delta(x - n\tau) \star (sinc(x/\tau)\right) \\
&= \frac{1}{\tau} \sum_{n=-\infty}^{\infty} f(n\tau)(sinc(1/\tau(x - n\tau)) \qquad (8.1)
\end{aligned}$$

■

E' evidente che nel caso in cui $f(x)$ sia campionata prendendone i campioni ad intervallo $\tau > 1/2s_c$ la funzione periodica $Q(s)$, di periodo $1/\tau < 2s_c$, sará costituita da copie della funzione $F(s)$ che si sovrappongono. Questo fenomeno prende il nome di *aliasing*. In presenza di aliasing non è possibile ricostruire la funzione $f(x)$ dai suoi campioni. Non è infatti possibile ottenere $F(s)$ dalla funzione periodica $Q(s)$, che, lo ricordiamo, è la trasformata del campionamento di $f(x)$.

Bibliografia

- E.O. Brigham: *The Fast Fourier Transform and its Applications*, Prentice Hall, 1988.

Capitolo 9

Clustering

9.1 Introduzione

La pratica di classificare oggetti in funzione di similarità percepite è alla base di molte discipline. Lo studio di metodi e algoritmi per raggruppare oggetti in gruppi (classi, clusters) prende il nome di *analisi dei cluster* o di *classificazione non supervisionata.*

Per quanto riguarda la elaborazione di immagini, tecniche di clustering vengono utilizzate per la segmentazione.

Informalmente possiamo dire che un *cluster* è composto da oggetti simili; entità appartenenti a cluster diversi non sono tra loro simili.

Un oggetto viene descritto da un insieme di attributi misurabili, più precisamente rappresentiamo un oggetto mediante un vettore $d-$dimensionale di caratteristiche. Se si considerano d caratteristiche per descrivere gli oggetti di interesse, queste sono di solito rappresentate mediante d assi ortogonali; gli oggetti da raggruppare, supponiamo siano n, sono rappresentati da punti nello spazio $d-$ dimensionale chiamato *spazio delle caratteristiche.* Un cluster può essere pertanto visualizzato come una collezione di punti che sono vicini gli uni agli altri o che soddisfano qualche relazione spaziale. Il compito degli algoritmi di clustering è quello di individuare raggruppamenti naturali in uno spazio vettoriale $d-$dimensionale.

Definito il problema di clustering come problema di suddividere un insieme di dati in gruppi di elementi tra loro simili, dobbiamo specificare in quale senso diciamo che gli elementi di un cluster (gruppo) sono tra loro più simili di quanto non lo siano rispetto agli elementi di altri clusters. Questa richiesta pone due tipi di problemi: come si misura la similarità (o la dissimilarità) tra coppie di elementi? e come si valuta la partizione di un insieme di elementi in gruppi?

9.2 Misure di similarità e funzioni obiettivo per il clustering

Un modo ovvio di definire una misura di similarità tra elementi è scegliere una opportuna nozione di distanza. Se la distanza è una buona misura della similarità ci si aspetta che la distanza tra elementi di un cluster sia significantemente minore della distanza tra elementi di clusters diversi.

Supponiamo per il momento di asserire che due elementi appartengono allo stesso cluster se la loro *distanza euclidea* è inferiore ad una certa soglia. E' ovvio che la scelta della soglia è molto importante. Meno ovvio è forse il fatto che il risultato del clustering dipende dalla scelta della distanza euclidea come misura di similarità. Questa scelta implica infatti che lo spazio delle caratteristiche sia isotropo; in tal caso i clusters definiti per mezzo della distanza euclidea sono invarianti per traslazione e rotazione. In generale non sono comunque invarianti rispetto a trasformazioni lineari (quali per esempio le variazioni di scala) o a trasformazioni che distorcono le relazioni di distanza.

Nel caso in cui per il problema in oggetto fosse necessario effettuare una variazione di scala, sarebbe necessario normalizzare i dati prima di effettuare il clustering, in modo da ottenere una invarianza rispetto a variazioni di scala. Per ottenere invarianza rispetto a variazioni di scala è necessario scalare gli assi in modo che tutte le caratteristiche abbiano media zero e varianza 1. Per ottenere invarianza per rotazione si possono ruotare gli assi in modo che essi coincidano con gli autovettori della matrice di covarianza dei dati. Questa trasformazione è detta trasformata delle componenti principali (o Hotelling transform o Karhunen -Loeve transform).

Non bisogna pensare che le normalizzazioni citate siano sempre necessarie o desiderabili. Per esempio traslare e scalare gli assi in modo che ogni caratteristica abbia media zero e varianza uno previene il fatto che certe caratteristiche che hanno grandi valori numerici dominino nel calcolo delle distanze. Sottrarre la media e dividere per la deviazione standard può essere una normalizzazione appropriata se i valori delle caratteristiche sono distribuiti secondo una distribuzione normale; tuttavia può risultare dannoso se la dispersione è dovuta alla presenza di sottogruppi che in tal caso verrebbero eliminati.

Non siamo qui interessati a discutere altre possibili nozioni di distanza, è importante però sottolineare che la scelta della misura di similarità ed eventualmente di procedure di normalizzazione dei dati introducono informazione che fornisce significato alla procedura di clustering .

Supponiamo di avere un insieme H di n vettori (elementi, campioni) $x_1, \ldots, x_n$ che vogliamo raggruppare in r sottoinsiemi disgiunti $H_1, \ldots, H_r$. Ogni sottoinsieme rappresenta un cluster e gli elementi dello stesso cluster devono essere tra loro più simili degli elementi di clusters diversi. Per formulare il problema in modo formale è necessario definire una funzione obiettivo che misura la qualità della partizione e cercare la partizione di H che minimizza (o massimizza) la funzione obiettivo. Vediamo un esempio di funzione da minimizzare.

Un criterio semplice e largamente usato per la costruzione di metodi di clustering è la minimizzazione della somma degli errori quadratici. Fissata una partizione $\Omega = \{H_1, \ldots, H_r\}$ di H sia n_i il numero di campioni in H_i e sia m_i la media di questi campioni:

$$m_i = \frac{1}{n_i} \sum_{x \in H_i} x$$

La somma degli errori quadratici è così definita:

$$J_e(\Omega) = \sum_{i=1}^{r} \sum_{x \in H_i} | x - m_i |^2$$

Questa funzione obiettivo ha una semplice interpretazione. Per un dato cluster H_i il vettore media m_i è il miglior rappresentante dei campioni in H_i, nel senso che esso minimizza la somma delle lunghezze al quadrato dei *vettori errore* $(x - m_i)$. Pertanto J_e misura l'errore totale che si commette rappresentando gli n campioni $x_1, \ldots, x_n$ mediante gli r vettori media $m_1, \ldots, m_r$ dei clusters $H_1, \ldots, H_r$.

Il valore di J_e dipende dalla partizione Ω; una partizione ottima è una partizione che minimizza J_e. La funzione J_e è una buona funzione obiettivo quando i campioni costituiscono nuvole di punti relativamente compatte e ben separate l'una dall'altra.

La funzione obiettivo sopra definita può essere derivata anche a partire da altre considerazioni. Consideriamo le *matrici di dispersione* S_i, S_b, S_w così definite:

$$S_i = \sum_{x \in H_i} (x - m_i)(x - m_i)^T$$

S_i rappresenta la dispersione all'interno del cluster i e l'indice T denota il vettore trasposto.

$$S_w = \sum_{i=1}^{r} S_i$$

S_w (within-class scatter matrix) rappresenta la dispersione all'interno di tutti i clusters. Sia

$$m = \frac{1}{n} \sum_{x \in H} x = \frac{1}{n} \sum_{i=1}^{r} n_i m_i$$

la matrice S_b (between-class scatter matrix) rappresenta la dispersione tra i clusters:

$$S_b = \sum_{i=1}^{r} n_i (m_i - m)(m_i - m)^T$$

La dispersione totale dei campioni è descritta dalla matrice S_t:

$$S_t = \sum_{x \in H} (x-m)(x-m)^T$$

Si verifica facilmente che $S_t = S_b + S_w$. E' evidente che S_t non dipende dal modo in cui i campioni sono suddivisi nei vari clusters invece S_b e S_w dipendono dalla partizione. Parlando grossolanamente c' è uno scambio tra queste due matrici, infatti se diminuisce la dispersione all'interno dei clusters aumenta la dispersione tra clusters, pertanto se si opera per minimizzare la dispersione all'interno dei clusters si massimizza la dispersione tra i clusters. Per fare un discorso più preciso è necessario misurare la dispersione all'interno e tra clusters, serve quindi uno scalare che misuri l'entità della dispersione. Due misure vengono di solito considerate la *traccia* e il *determinante*. Consideriamo qui solo una misura basata sulla traccia.

La traccia è la misura scalare più semplice per valutare le matrici di dispersione, la traccia è la somma degli elementi diagonali di una matrice.

Una semplice funzione obiettivo da minimizzare è la traccia di S_w; questa funzione obiettivo coincide con la funzione J_e sopra definita, infatti:

$$trS_w = \sum_{i=1}^{r} trS_i = \sum_{i=1}^{r} \sum_{x \in H_i} \mid x - m_i \mid^2 = J_e$$

9.3 Una procedura iterativa per il clustering

Una volta definita la funzione obiettivo, il problema di clustering diventa un problema ben posto di ottimizzazione discreta: trovare le partizioni dell'insieme H che minimizzano la funzione obiettivo. Poiché l'insieme è finito c' è un numero finito di possibili partizioni; pertanto, in teoria, il problema potrebbe essere risolto mediante una enumerazione esaustiva. Naturalmente tale metodo è inapplicabile anche per problemi semplici ci sono infatti circa $\frac{r^n}{r!}$ partizioni di un insieme di n elementi in r sottoinsiemi. Un metodo semplice e frequentemente usato consiste nell'ottimizzazione iterativa. L'idea di base è quella di definire una ragionevole partizione iniziale e di spostare i campioni da un gruppo all'altro se lo spostamento porta alla diminuzione del valore della funzione obiettivo.

Come ogni procedura di ricerca locale anche questo metodo garantisce di trovare minimi locali e non minimi globali. Diverse condizioni iniziali portano a diverse soluzioni e non si può sapere se si è trovata la soluzione ottima. Nonostante queste limitazioni il fatto che le risorse computazionali richieste siano ragionavoli rende il metodo significativo.

Descriviamo le idee alla base del metodo di miglioramento iterativo e semplici algoritmi che lo realizzano nel caso in cui la funzione obiettivo sia la funzione J_e sopra definita.

Supponiamo di avere generato a caso una partizione Ω e di dover decidere se un arbitrario vettore $\widehat{x}$ che è stato posto nel cluster H_i, debba essere posto in un altro cluster, per esempio H_j. Dobbiamo valutare se lo spostamento fa diminuire la funzione obiettivo. Come primo passo analizziamo come variano le medie m_i e m_j togliendo $\widehat{x}$ dal cluster H_i e ponendolo nel cluster H_j. Si verifica facilmente che le nuove medie $m_i^\star$ e $m_j^\star$ ottenute dopo lo spostamento soddisfano queste equazioni:

$$m_j^\star = m_j + \frac{\widehat{x} - m_j}{n_j + 1}$$

$$m_i^\star = m_i - \frac{\widehat{x} - m_i}{n_i - 1}$$

La seconda equazione vale sotto l'assunzione che n_i sia diverso da uno.

Adesso valutiamo come cambiano, dopo l'inserimento e la rimozione di $\widehat{x}$, i termini $J_j = \sum_{x \in H_j} \mid x - m_j \mid^2$ e $J_i = \sum_{x \in H_i} \mid x - m_i \mid^2$. Naturalmente essendo $J_e = \sum_{i=1}^{r} J_i$ determineremo le condizioni per cui lo spostamento di $\widehat{x}$ fa diminuire J_e. Sia:

$$\begin{aligned}
J_j^\star &= \sum_{x \in H_j} \mid x - m_j^\star \mid^2 + \mid \widehat{x} - m_j^\star \mid^2 \\
&= \sum_{x \in H_j} \mid x - m_j - \frac{\widehat{x} - m_j}{n_j + 1} \mid^2 + \mid \frac{n_j}{n_j + 1}(\widehat{x} - m_j) \mid^2 \\
&= J_j + \frac{n_j}{n_j + 1} \mid \widehat{x} - m_j \mid^2
\end{aligned}$$

e sia

$$\begin{aligned}
J_i^\star &= \sum_{x \in H_i} \mid x - m_i^\star \mid^2 - \mid \widehat{x} - m_i^\star \mid^2 \\
&= \sum_{x \in H_i} \mid x - m_i + \frac{\widehat{x} - m_i}{n_i - 1} \mid^2 + \mid \frac{n_i}{n_i - 1}(\widehat{x} - m_i) \mid^2 + \\
&= J_i - \frac{n_i}{n_i - 1} \mid \widehat{x} - m_i \mid^2
\end{aligned}$$

Queste equazioni mostrano che lo spostamento di $\widehat{x}$ dal cluster H_i al cluster H_j diminuisce la funzione obiettivo J_e solo se la diminuzione di J_i è maggiore dell'aumento di J_j. Ciò accade quando:

$$\frac{n_i}{n_i - 1} \mid \widehat{x} - m_i \mid^2 > \frac{n_j}{n_j + 1} \mid \widehat{x} - m_j \mid^2$$

quindi quando il vettore $\widehat{x}$ è più vicino a m_j che a m_i. La maggiore diminuzione del valore della funzione J_e si ha quando si seleziona il cluster k per cui $\frac{n_k}{n_k+1} \mid \widehat{x} - m_k \mid^2$ è minimo. Questa osservazione porta al seguente algoritmo:

Basic Minimum Squared Error

Input: Un insieme di n vettori, $x_1, \ldots, x_n$, un intero r.

Inizializzazione: Seleziona una partizione iniziale degli n vettori in r gruppi; computa i vettori media $m_1, \ldots, m_r$ e la funzione J_e.

Passo 1: Seleziona un elemento $\widehat{x}$; supponiamo che $\widehat{x} \in H_i$

Passo 2 : Se $n_i = 1$ vai a 4, altrimenti computa:

$$a_j = \frac{n_i}{n_i - 1} \mid \widehat{x} - m_i \mid^2 \quad j = i$$

$$a_j = \frac{n_j}{n_i + 1} \mid \widehat{x} - m_j \mid^2 \quad j \neq i$$

Passo 3: Se $a_k \leq a_j$ per ogni j allora trasferisci $\widehat{x}$ in H_k.

Passo 4: Ricalcola m_i, m_k, J_e. Se J_e non è cambiata dopo n iterazioni STOP, altrimenti vai al Passo 1.

Output: La partizione ottenuta.

Questa procedura sequenziale dipende dall'ordine in cui gli n vettori vengono selezionati, ordine che non è specificato dall'algoritmo sopra descritto. Può essere applicata nel caso in cui i campioni siano acquisiti on-line e il clustering debba essere fatto in tempo reale. Una versione parallela della procedura sopra descritta è l'algoritmo K-means (chiamato anche algoritmo LGB).

K-Means

Input: Un insieme di n vettori, $x_1, \ldots, x_n$, un intero r, un reale ϵ.

Inizializzazione: Poni $l = 0$, scegli r vettori media $m_1(0), \ldots, m_r(0)$.

Passo 1: $\forall j \;\; (j = 1, \ldots, r)$ costruisci

$$H_j = \{x_k : \mid x_k - m_j \mid^2 \leq \mid x_k - m_i \mid^2\} \;\; \forall i \neq j$$

Passo 2: Computa le nuove medie $m_1(l + 1), \ldots, m_r(l + 1)$

Passo 3: $\forall j \;\; (j = 1, \ldots, r)$ se

$$\mid m_j(l) - m_j(l + 1) \mid > \epsilon$$

vai al Passo 1 altrimenti STOP

Output: La partizione ottenuta.

Osservazioni:

- Il numero di cluster deve essere preassegnato in entrambi gli algoritmi.

- La partizione iniziale nell' algoritmo Basic Minimum Squared Error o la scelta iniziale delle medie nell'algoritmo K-means può essere fatta in modo random. La soluzione dipende in generale dalla condizione iniziale. Ripetizioni con diverse condizioni iniziali possono fornire un'idea di quanto il risultato dell'algoritmo sia sensibile alle condizioni iniziali.

- Nello spazio delle caratteristiche i clusters sono separati da iperpiani. Sono i vettori media che determinano gli iperpiani.

Sino ad ora abbiamo ignorato il problema di come determinare il numero di clusters r ed abbiamo assunto che r sia fissato. Nella pratica fissare r è un problema. In generale si effettuano esperimenti con diversi valori di r e si tiene la soluzione che si ritiene essere la migliore per il problema che si sta considerando. Se si utilizza J_e come funzione obiettivo è evidente che aumentare r porta alla formazione di clusters sempre più piccoli e di conseguenza la dispersione all'interno di ogni cluster si riduce. Ovviamente per $r = n$ la funzione obiettivo ha valore 0, ma considerare ogni campione un cluster non è una soluzione.

Ci sono euristiche per modificare il risultato di un algoritmo di clustering. Le più semplici sono la fusione e la suddivisione di alcuni dei clusters ottenuti.

La fusione viene effettuata quando due clusters hanno vettori media tra loro molto vicini oppure quando il numero di elementi di un cluster è molto piccolo.

Decidere se suddividere un cluster è in generale un problema più complesso, clusters molto grossi sono candidati ad essere suddivisi magari riapplicando la procedura di clustering ai soli elementi della classe da suddividere.

Descriviamo ora un algoritmo di apprendimento competitivo, anch'esso basato sulla minimizzazione degli errori quadratici, proposto da Arbib e Uchiyama nel 1994. L'algoritmo di Arbib fornisce prestazioni ragionevoli per la segmentazione di immagini a colori ed è efficiente. Prima di presentare l'algoritmo definiamo i seguenti parametri:
N_{max}: numero massimo di vettori da presentare all'algoritmo
r: numero massimo di clusters ovvero di unità che è possibile generare
θ: numero massimo di vittorie consentito ad ogni unità
w_j: vettore di pesi che descrive l'unità j
η: tasso di apprendimento
$\mu_0 = \frac{\sum_{j=1}^{n} p_j x_j}{\sum_{j=1}^{n} p_j}$: centroide globale dei dati in ingresso; $p_j = 1/n$

Algoritmo (Arbib 94)

Input: Un insieme di H di n vettori, $x_1, \ldots, x_n$; gli interi: r, N_{max}, θ, un reale η.

Inizializzazione: Crea l'unità U_1 il cui vettore di pesi è $w_1 = \mu_0$
Poni $N = 0$, $wincount(U_1) = 0$

Passo 1: Estrai a caso un vettore x dall'insieme H
Determina tra le unità presenti quella più vicina a x secondo la metrica euclidea $\mid x - w \mid^2$. Sia U_{winner} l'unità vincente (se più unità sono candidate se ne scelga una a caso).

Passo 2: Si aggiorni il vettore peso dell'unità vincente secondo la regola:

$$w_{winner} = w_{winner} + \Delta w_{winner} \qquad \Delta w_{winner} = \eta(x - w_{winner})$$

Passo 3: $wincount(U_{winner}) = wincount(U_{winner}){+}1$. Se $wincount(U_{winner}) = \theta$ e il numero di unità generate è minore di r si generi una nuova unità U_{new} descritta dallo stesso vettore w_{winner} associato all'unità vincente. Si azzerino le variabili $wincount(U_{winner})$ e $wincount(U_{new})$.

Passo 4: $N = N + 1$
Se $N = N_{max}$ allora STOP, altrimenti vai al passo 1.

Output: La partizione di H che si determina con la regola:

$$x_j \in H_i \longleftrightarrow \forall k \neq i \mid x_j - w_i \mid^2 < \mid x_j - w_k \mid^2$$

Si osservi che nel caso in cui il vettore che risulta più vicino all'input corrente sia associato a due diverse unità una sola di esse può essere scelta come vincitore (vedi passo 1). Inoltre, a causa del meccanismo basato sul numero di vittorie, le nuove unità vengono generate proprio dove maggiore è la densità di vettori di input. Non accade pertanto che solo alcune delle unità possano risultare sempre vincitrici, problema che si può avere nel caso in cui vengano create contemporaneamente tutte le unità e si abbia una inappropriata inizializzazione dei vettori peso. Un valore alto di θ consente di ottenere una maggior accuratezza ma richiede una soglia N_{max} più alta sul numero di iterazioni al fine di raggiungere il numero di clusters desiderato. Di seguito si mostra che se r è il numero di unità da generare e θ è la soglia di vittorie, il numero di ripetizioni N_{max} necessario è:

$$(r-1)\theta \leq N_{max} \leq (2r-1)\theta - r$$

Infatti, si definiscano *sdoppiamento* la generazione di una nuova unità con un vettore di pesi uguale a quello dell'unità vincente ed *epoca* il periodo (in termini di numero di iterazioni) che intercorre tra due sdoppiamenti; se durante un'epoca vince sempre la stessa unità e ciò accade per ogni epoca allora si ha uno sdoppiamento ogni θ iterazioni e si hanno solo $r-1$ epoche ciascuna di θ iterazioni, dopo verrà creata l'unità r−esima. Pertanto il numero minimo di iterazioni necessario a generare tutte le unità risulta essere $(r-1)\theta$. Per valutare il numero di passi nel caso peggiore, osserviamo che se durante un'epoca le vittorie vengono ripartite solo tra le 2 unità aventi al loro attivo zero vittorie (unità generate dallo sdoppiamento) per le altre il numero di vittorie sarà al massimo pari a $\theta - 1$. Le due unità in questione possono vincere per un totale di $2(\theta-1)$ iterazioni e solo la iterazione successiva causerà lo sdoppiamento di una delle due. Per il caso particolare della prima epoca il massimo numero di ripetizioni è $\theta - 1$ in quanto una ulteriore vittoria farebbe scattare lo sdoppiamento. Pertanto, il massimo numero di ripetizioni risulta essere:

$$(\theta-1) + \sum_{i=1}^{r-1}(2(\theta-1)+1) = (\theta-1) + 2(r-1)(\theta-1) + (r-1) = \theta(2r-1) - r$$

Se usiamo algoritmi di clustering come primo passo per la segmentazione di immagini a colori non conosciamo a priori in quanti gruppi distinti di colori sia ragionevole suddividere tutti i colori presenti nell'immagine. Un metodo che può aiutare è considerare gli istogrammi per ognuna della caratteristiche colore, cercare in questi istogrammi i picchi significativi ed effettuare il prodotto del numero dei picchi signifativi trovati in ogni componente colore. Tale numero può fornire una indicazione del massimo numero di clusters da cercare.

Bibliografia

- M.A. Arbib, T. Uchiyama. Color image segmentation using competitive learning. IEEE Trans. on Pattern Analysis and Machine Intelligence. Vol 16, n. 12 pp 1197-1206, 1994.

- R. O. Duda, P.E. Hart. Pattern Classification and Scene Analysis. Wiley 1973.