

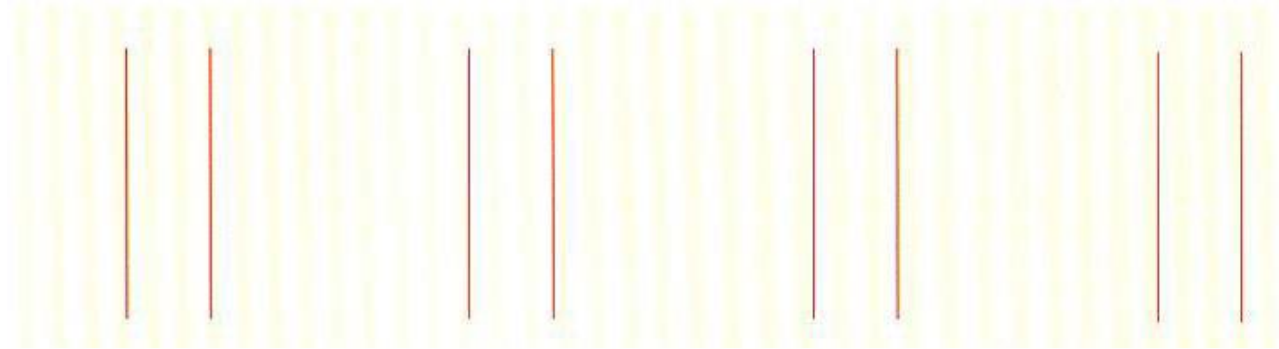
Lezione 1:

Leggi della percezione visiva:

la nostra vista parte dagli occhi per poi arrivare alla mente ed è là che viene elaborata l'immagine seguendo delle leggi ben precise:

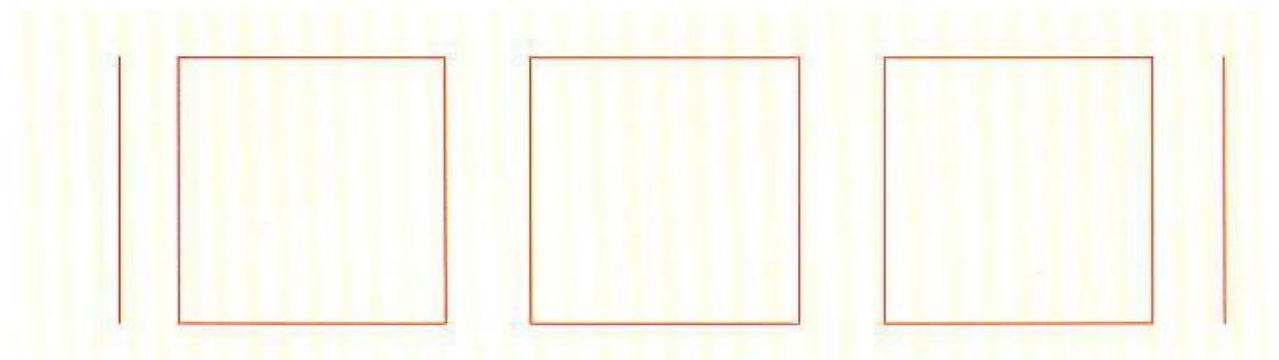
Legge della Vicinanza:

se nel nostro campo visivo entrano oggetti vicini fra loro li percepiamo come coppie ma se invece gli stessi oggetti entrano distaccati fra loro li percepiamo come singoli individui.



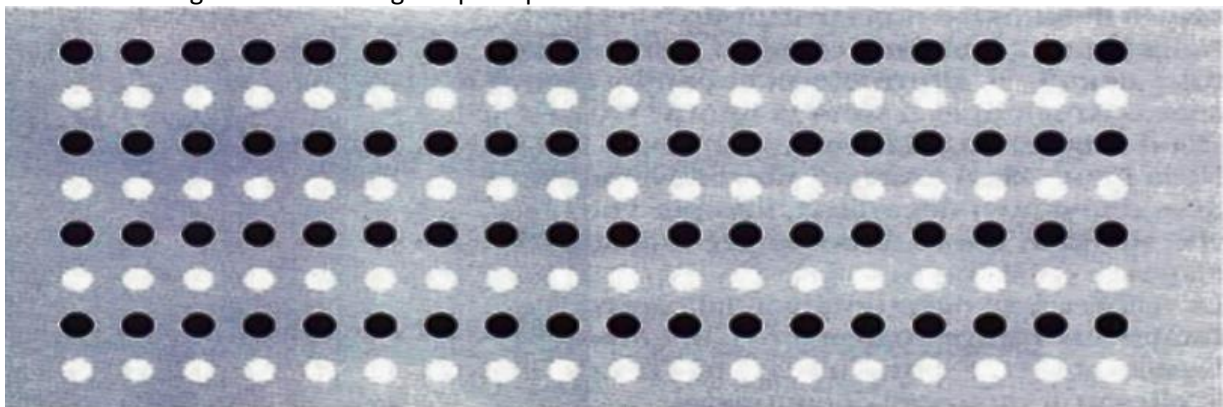
Legge della Chiusura:

se nel nostro campo visivo entrano una moltitudine di oggetti, avremo la percezione che gli oggetti chiusi saranno dominanti rispetto agli altri.



Legge dell'uguaglianza:

elementi simili o uguali fra loro vengono percepiti come un insieme.



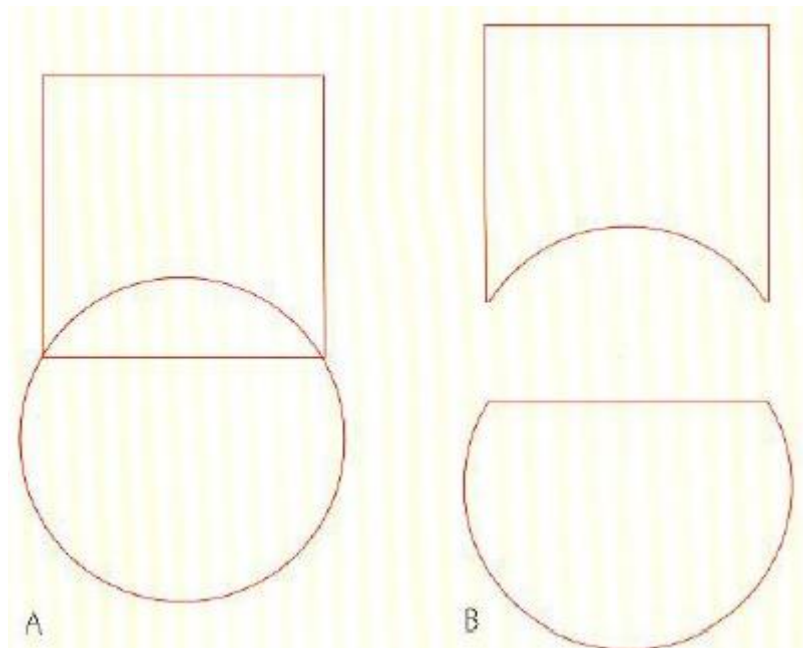
Legge della continuità:

quando due figure conosciute vengono unite o sovrapposte le percepiamo allo stesso modo in maniera distinta.

Legge della buona forma:

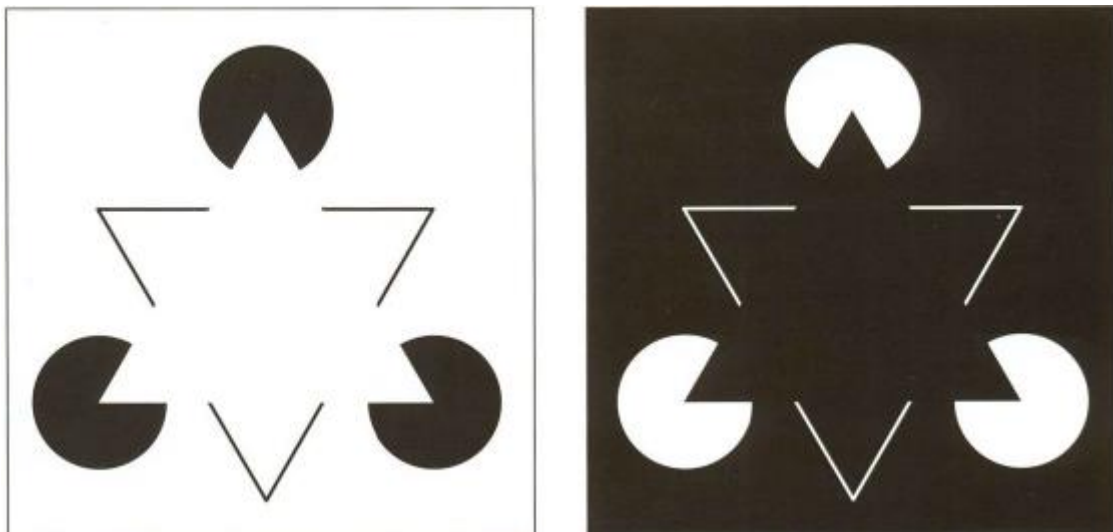
quando figure diverse vengono unite, finiscono per avere ognuna la propria forma anche se in realtà le forme che diamo noi non compaiono.

Lezione 1:



5.1

La mente tende a completare i contorni di figure incomplete percependo forme così forme regolari o comunque strutture compatte anche dove realmente non c'è nulla.



Intensità Percepita

Le immagini digitali sono rappresentate da un numero finito di intensità, ed è importante conoscere come l'occhio umano riesce a riconoscere i diversi livelli di intensità, sperimentalmente si è dimostrato che l'intensità percepita è funzione logaritmica dell'intensità incidente nell'occhio.

Range Dinamico

L'occhio umano non utilizza contemporaneamente tutto il range di intensità percepite ma solo su una porzione di esso e riesce a distinguere in maniera differente se si trova in una zona chiara o scura.

Effetto Mach

L'occhio tende a vedere vicino i bordi bande di "rinforzo" chiare o scure tra zone a differente luminosità anche se la zona fino al bordo ha un intensità costante.

Lezione 1:



Rappresentazione di un'immagine

un'immagine è una funzione bidimensionale $[f(x, y)]$ proporzionale alla luce incidente $[i(x, y)]$ nell'oggetto e anche a quella riflessa $[r(x, y)]$ dallo stesso.

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

Ed abbiamo i :

$$0 < i(x, y) < \infty$$

E r :

$$0 < r(x, y) < 1$$

Il valore di $f(x, y)$ è nel campo dei numeri reali ma abbiamo bisogno di usare valori discreti e per fare ciò si usa il campionamento e la quantizzazione, il piano in cui stanno le coordinate dell'immagine è detto DOMINIO e le coordinate sono le variabili.

Immagine Raster e Immagini Vettoriali

Immagine Raster : insieme finito di pixel (utilità: fotorealismo e standard sul web)

Immagine Vettoriali : serie di forme geometriche (utilità: non si rovinano durante le trasformazioni)

Pixel : valore quantizzato misurato da ciascun sensore.

Lezione 2:

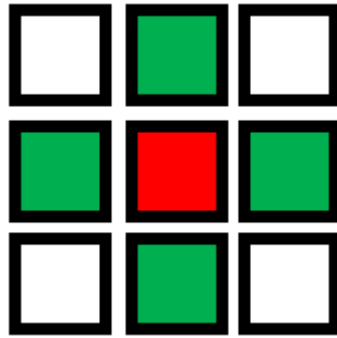
Un immagine Raster può essere rappresentata da una matrice, quindi si possono fare tutte le operazioni che si fanno sulle matrici(non è detto che servano però).

Prodotto di matrici in image processing : è inteso il prodotto punto a punto

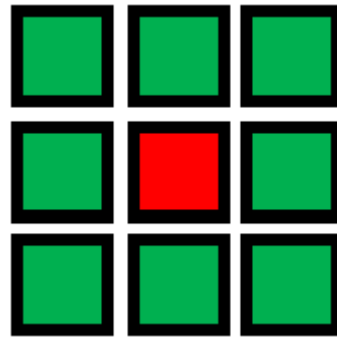
$$\begin{bmatrix} 1 & 2 \\ 3 & -1 \end{bmatrix} \cdot \begin{bmatrix} -3 & 0 \\ 1 & 4 \end{bmatrix} = \begin{bmatrix} -3 & 0 \\ 3 & -4 \end{bmatrix}$$

Chi sono i vicini di un pixel?

I vicini 4 connessi sono quelli a destra e sinistra sopra e sotto del pixel



I vicini 8 connessi sono tutti quelli intorno al pixel (destra, sinistra, sopra, sotto e i 4 in diagonale)



Forward Mapping

il forward mapping è un processo di trasformazione che prende delle coordinate e tramite funzioni matematiche le trasforma in altre coordinate, ci sono diversi tipi di funzioni :

1. Trasformazioni affini : sono trasformazioni lineari che conservano la caratteristica in comune, la proporzione delle distanze tra i punti, ma non per forza anche gli angoli e le lunghezze
2. Trasformazioni di Prospettiva : sono trasformazioni non lineari che cambiano la posizione apparente, la dimensione e l'orientamento di un oggetto in un'immagine
3. Correzione della distorsione radiale: questo tipo di funzione corregge la distorsione causata dalla curvatura di un obiettivo

Il processo di Forward Mapping solitamente è in due passaggi

1. Scegliere il tipo di funzione da usare per la mappatura
2. Applicare la funzione pixel per pixel

N.B. il forward Mapping può creare artefatti quindi viene utilizzato con altri

Inverse Mapping

L'inverse mapping è quel processo di trasformazione che prende un immagine di output e la trasforma nella sua immagine originale(per effettuare un inverse mapping è obbligatorio che la funzione di forward mapping utilizzati sia invertibile)

Interpolazione

Il processo di interpolazione è quel processo che serve a tappare i buchi di informazione presenti nell'immagine partendo dai vicini del pixel mancante, esistono diversi tipi di interpolazione :

1. Vicino più prossimo
2. Bilineare
3. Bicubica

Vicino più prossimo

Questo metodo assegna a ogni nuova posizione l'intensità del pixel più prossimo nell'immagine reale, è un metodo banale che introduce artefatti

Bilineare

In questo metodo si utilizzano i 4 pixel più vicini per stimare l'intensità da assegnare a ciascuna nuova posizione utilizzando la seguente formula:

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

Bicubica

In questo metodo si utilizzano i 16 pixel più vicini al punto (i loro valori vengono determinati da 16 equazioni in 16 incognite) e il valore da assegnare al punto si ottiene attraverso l'equazione :

$$v(x, y) = \sum_{i=0}^3 \sum_{j=0}^3 a_{ij} x^i y^j$$

Decimazione

Durante un processo di riduzione dell'immagine (zoom effettuato con valore inferiori ad 1) si ha un effetto di decimazione, questo effetto può essere fatto in 2 metodi:

1. Ogni 4 pixel ne prendiamo uno
2. Ogni 4 pixel si calcola il valore medio

MSE e PSNR

MSE (Mean Square Error) serve a stimare l'errore quadratico medio fra due immagini, più è piccolo il suo valore meglio è.

$$MSE = \frac{1}{MN} \sum_{x=1}^M \sum_{y=1}^N [I'(x, y) - I(x, y)]^2$$

PSNR (Peak Signal to Noise Ratio) per calcolarlo serve l'MSE e serve a misurare la qualità di un'immagine compressa rispetto all'originale, più grande il suo valore meglio è.

$$PSNR = 10 \log_{10} \left(\frac{S^2}{MSE} \right)$$

MSE e PSNR per RGB

Per calcolare MSE e PSNR in un'immagine RGB o si fa la media dei valori MSE e PSNR sui 3 canali... o si usa una combinazione lineare che è pesata maggiormente sulla componente verde (G)

Lezione 3:

Il colore di un oggetto è dato dalla luce riflessa da quest'ultimo; pertanto, per creare un'immagine digitale è essenziale catturare l'energia della luce riflessa che viene trasformata in un impulso elettrico con un sensore (solitamente le fotocamere usano sensori CCD: Charged Coupled Device) e successivamente digitalizzare tale impulso.

I sensori CCD sono dispositivi elettronici che se colpiti da un impulso si caricano positivamente (non possono sovra-caricarsi) più celle di sensori sono presenti sotto forma di matrice di sensori nell'obiettivo della nostra camera, più è grande la matrice di sensori maggiore sarà la qualità della foto misurata in Megapixel, dopo che le cariche sono acquisite vengono trasferite in memoria, una colonna alla volta facendo scalare le colonne dalla successiva alla precedente, questo trasferimento avviene in 3 fasi una fase per ogni colonna della matrice di sensori.

CFA: Color Filter Array

Ogni sensore CCD memorizza un solo colore, non una terna quindi bisogna scegliere il CFA ovvero il modello di memorizzazione ottimale per avere il risultato desiderato, dato che il sensore immagazzina un solo colore i due mancanti vengono ottenuti per interpolazione dai pixel vicini, migliore è il metodo di interpolazione più accurato sarà il risultato finale, noi studieremo il BAYERN PATTERN, presenta un rapporto 1:2:1 per i canali RGB privilegiando il canale verde perché è il canale più importante per la percezione umana.

Visualizzare un'immagine stampando la matrice di dati rilasciati dai sensori può essere fatto, ma si avrà una immagine in scala di grigi, salvare una sola componente della terna RGB comporta una visione a mosaico dell'immagine, per avere una immagine a colori bisogna utilizzare un algoritmo di color interpolation per ogni singola terna dai dati dell'intorno, visto che questi algoritmi levano l'effetto mosaico vengono chiamati algoritmi di demosaicking.

La tecnica di Color Interpolation: replication per ogni singolo pixel copia i dati dell'intorno nei posti vuoti, questa tecnica è chiamata "Nearest-neighbor interpolation".

La tecnica di Color Interpolation: Bilinear abbiamo l'informazione di R e mancano G e B

1. Nel canale R non si deve fare nulla;
2. Nel canale G occorre ricavare i dati da un intorno selezionando i 4 valori rilasciati dal sensore;
3. Nel canale B occorre ricavare i dati da un intorno selezionando i 4 valori rilasciati dal sensore;

La tecnica di Color Interpolation: Bilinear abbiamo l'informazione di G e mancano R e B

1. Nel canale G non si deve fare nulla;
2. Nel canale R occorre ricavare i dati da un intorno selezionando i 2 valori rilasciati dal sensore;
3. Nel canale B occorre ricavare i dati da un intorno selezionando i 2 valori rilasciati dal sensore;

La tecnica di Color Interpolation: Bilinear abbiamo l'informazione di B e mancano G e R

1. Nel canale B non si deve fare nulla;
2. Nel canale G occorre ricavare i dati da un intorno selezionando i 4 valori rilasciati dal sensore;
3. Nel canale R occorre ricavare i dati da un intorno selezionando i 4 valori rilasciati dal sensore;

Lezione 4:

Nei nostri studi ci interessa molto lo studio dell'occhio ed in particolare della retina, è una membrana che ricopre la parte posteriore dell'occhio ed è formata da coni e bastoncelli, i coni sono circa 6/7 milioni e sono divisi in tre tipi,

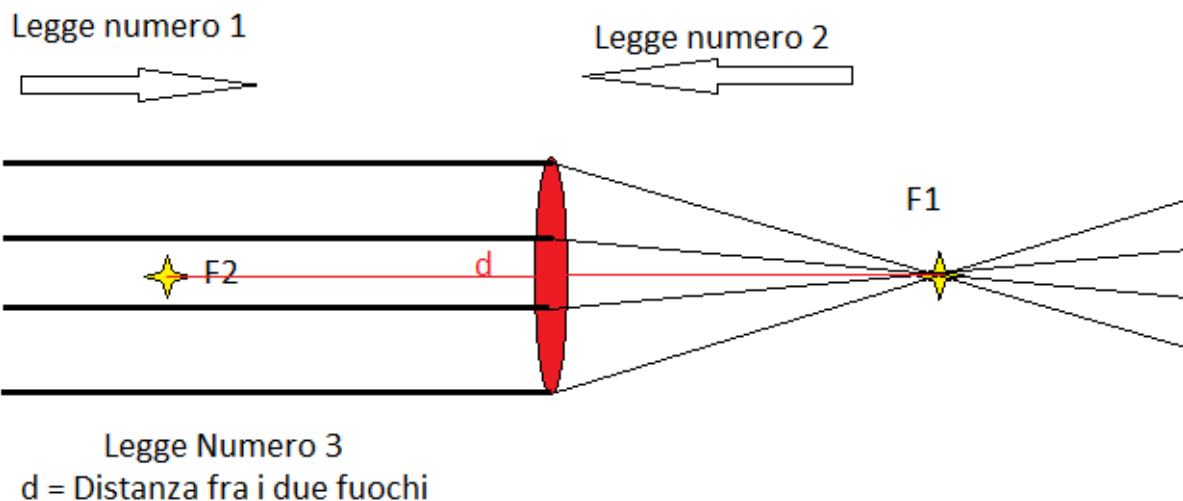
- 1.1. Tipo S: sensibili alle lunghezze d'onda corte e vedono i colori bluastrì;
- 1.2. Tipo M: sensibili alle lunghezze d'onda medie e vedono i colori verdastrì;
- 1.3. Tipo L: sensibili alle lunghezze d'onda lunghe e vedono colori rossastrì;
sono concentrati nella fovea e sono sensibili al colore, ogni cono è collegato ad un nervo ottico e sono responsabili della vista policroma; i bastoncelli sono circa 75/150 milioni e sono distribuiti su tutta la retina, sono poco sensibili al colore e gruppi di bastoncelli sono collegati ad un singolo nervo ottico, i bastoncelli sono responsabili della vista monocroma;

la fovea è una regione di 1.5 mm x 1.5 mm, contiene circa 150.000 coni e circa 337.500 bastoncelli, un CCD per contenere tutti questi recettori dovrebbe essere grande almeno 5 mm x 5mm.

Pinhole : è un modello teorico dove si mette una carta fotosensibile dentro una scatola e gli si pratica un foro(possano esserci più fori) di alcuni mm(il foro è proporzionale alla radice quadrata della distanza, per la lunghezza d'onda della luce emessa), il foro viene protetto con un pezzo di carta ed all'occorrenza si fa entrare la luce dal forellino;

i forellini sono uno strumento inadeguato perché raccolgono poche informazioni per fare le misurazioni che non saranno quindi precise; pertanto, si è pensato già dal rinascimento di creare delle lenti che hanno il diametro più grande del loro spessore(lente sottile), una lente sottile ha delle proprietà ben specificate:

1. Raggi paralleli alla lente vengono concentrati in un punto detto fuoco posto a distanza F dalla lente;
2. Raggi provenienti dal fuoco vengono ri-trasmessi tutti paralleli nella direzione dell'asse della lente;
3. Una lente sottile ha due fuochi equidistanti da essa;



4. La lente sottile fa passare solo uno dei raggi luminosi paralleli all'asse ottico proveniente da un unico oggetto;
5. Solo un raggio luminoso passa per il fuoco e la lente lo farà passare in un raggio parallelo alla lente;
6. Il punto in cui i due raggi si incontrano è dove si formerà l'immagine;
7. Se mettiamo un piano di sensori più avanti o dietro al piano che contiene l'immagine si otterrà un'immagine sfocata.

Lezione 4:

Equazione della lente focale :

$$\frac{1}{u} + \frac{1}{v} = \frac{1}{f}$$

U : Distanza dell'oggetto

V : Distanza dell'immagine riflessa

F : Distanza Focale

Magnificazione:

la magnificazione è quel processo che serve ad aumentare la grandezza digitale del soggetto e non quella reale, ed è un numero che descrive con quale fattore sia stato ingrandito un oggetto.

La magnificazione può dimostrare che i due triangoli sono simili e quindi si ha : $\frac{h}{k} = \frac{v}{u} = m$

Ha una formula che si basa su quella della lente focale :

$$\frac{1}{u} + \frac{1}{v} = \frac{1}{f} \Rightarrow$$

Moltiplichiamo tutto per v

$$\frac{v}{u} + 1 = \frac{v}{f} \Rightarrow$$

Invertiamo e scambiamo con m

$$\frac{f}{v} = \frac{1}{(m + 1)} \Rightarrow$$

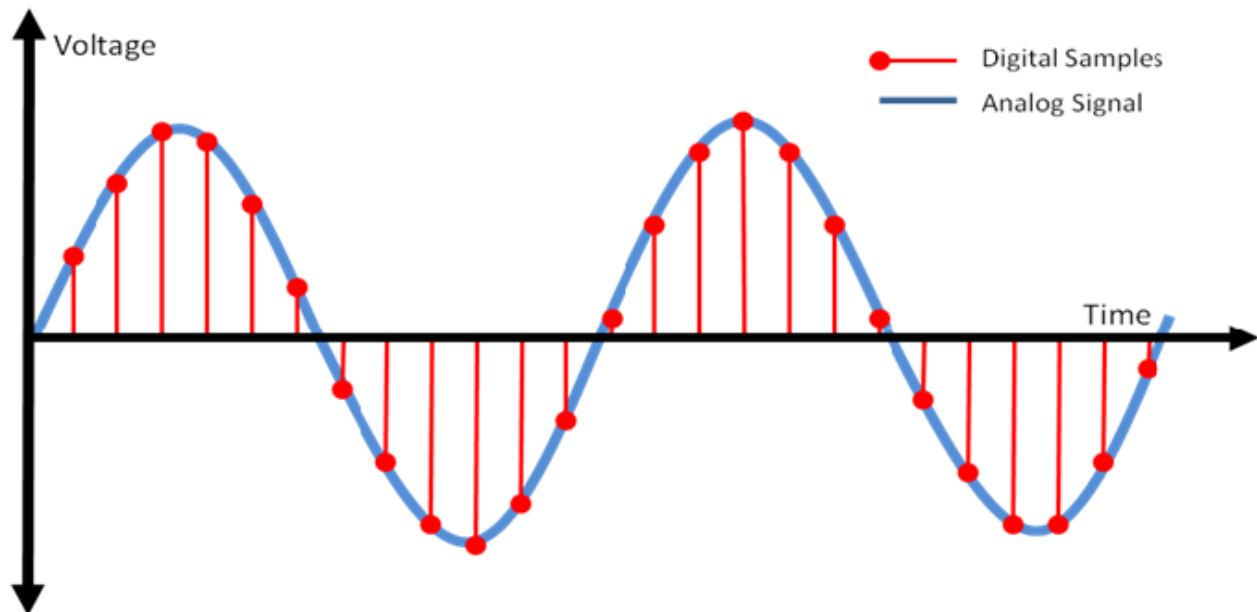
Moltiplichiamo per u e scambiamo con m

$$\frac{f}{m} = \frac{u}{(m + 1)} \Rightarrow f = \frac{m * u}{(m + 1)}$$

Lezione 5:

Campionamento :

è il processo di raccolta di dati sulla posizione e sull'intensità dei pixel dell'immagine.



Un errore nel campionamento può deteriorare il segnale in due modi :

1. Perdita di dettagli;
2. Creare Artefatti

Per scegliere il giusto campionamento si usa il teorema di Shannon che si basa sulla misura del Nyquist Rate, ovvero una relazione tra frequenza massima del segnale e frequenza di campionamento, per evitare l'effetto di aliasing bisogna campionare ad una frequenza che è almeno il doppio della frequenza massima del segnale, il teorema di Shannon ci dice che se raccogliamo campioni con frequenza più alta del doppio del Nyquist rate possiamo ricostruire un segnale fedelmente.

Cosa succede se avviene un sotto campionamento?

Si perdono dettagli significativi del segnale e si introducono artefatti.

Aliasing

È un fenomeno che avviene quando si campiona un segnale a frequenza inferiore a quella di Nyquist

Quantizzazione:

è il processo di riduzione del numero di colori utilizzati per rappresentare i singoli pixel dell'immagine, esistono due tipi di quantizzazione, uniforme e non uniforme.

Procedura per la quantizzazione:

se i valori da quantizzare sono numeri nel range $[a, b]$ e si vuole quantizzare su n livelli si fissano $n+1$ numeri tali che :

$$t_0 = a < t_1 < t_2 < \dots < t_n = b$$

Il numero x in $[a, b]$ verrà assegnato al livello di quantizzazione k se risulta : $t_k \leq x < t_{k+1}$

Lezione 6:

Quello che il nostro occhio percepisce è una piccola porzione dello spettro elettromagnetico, tanto che se noi prendiamo una luce bianca e la proiettiamo su di un prisma, questo scomporrà le onde luminose nelle varie lunghezze d'onda in uno spettro che vanno dal violetto al rosso, lo spettro del visibile è diviso in :

1. Violetto
2. Blu
3. Verde
4. Giallo
5. Arancio
6. Rosso

Queste 6 bande non sono tutte della stessa lunghezza e vanno a degradarsi in quelle limitrofe e vanno dalla lunghezza d'onda 350 a 750 circa.

L'occhio umano percepisce come colore di un'oggetto quella luce che l'oggetto stesso riflette, se l'oggetto riflette tutte le lunghezze d'onda sarà visto come bianco se invece riflette le lunghezze d'onda da 500 a 570 sarà visto come verde.

Per descrivere la luce servono 3 tipi di valori:

1. Radianza (quantità di luce emessa dalla sorgente luminosa);
2. Luminanza (energia percepita dall'utente);
3. Brillantezza (indica la sensazione di colore);

Teoria del tristimolo

Tutti i colori si possono ottenere mescolando tre colori fondamentali in proporzioni differenti.

Composizione dei colori

Illuminando una superficie bianca con uno o più luci monocromatiche si ottiene una radianza colorata per **sintesi additiva** qualsiasi luce di spettro complesso può essere ottenuta come somma di tre luci monocromatiche agenti insieme, la CIE partendo dall'ipotesi che i colori possono essere espressi in RGB hanno scelto le tre radiazioni primarie in modo che le lunghezze d'onda siano il più possibile distanziate fra loro nello spettro

Valori del Tristimolo X, Y, Z per lo spazio colorimetrico CIE 1931

$$X = \int_{380}^{780} L_{e,\lambda} \bar{x}(\lambda) d\lambda \quad Y = \int_{380}^{780} L_{e,\lambda} \bar{y}(\lambda) d\lambda = \int_{380}^{780} L_{e,\lambda} V(\lambda) d\lambda \quad Z = \int_{380}^{780} L_{e,\lambda} \bar{z}(\lambda) d\lambda$$

Visione e misura

Una luce allo xenon illumina il campione in maniera uniforme, la sorgente di luce è detta illuminante, l'illuminante è quindi l'energia radiante con distribuzione spettrale di energia relativa definita nel campo di lunghezza d'onda capace di influenzare la visione del colore degli oggetti;

gli illuminanti CIE sono quelli le cui distribuzioni spettrale di energia relativa sono definite dalla CIE(es. A, B, C, D65)

valori del tristimolo CIE 1931

blue = 435.8 nm

Verde = 546.1 nm

Rosso = 700 nm

Lezione 6:

Valori del tristimolo nel 1964

Blue = 445 nm

Verde = 535 nm

Rosso = 575 nm

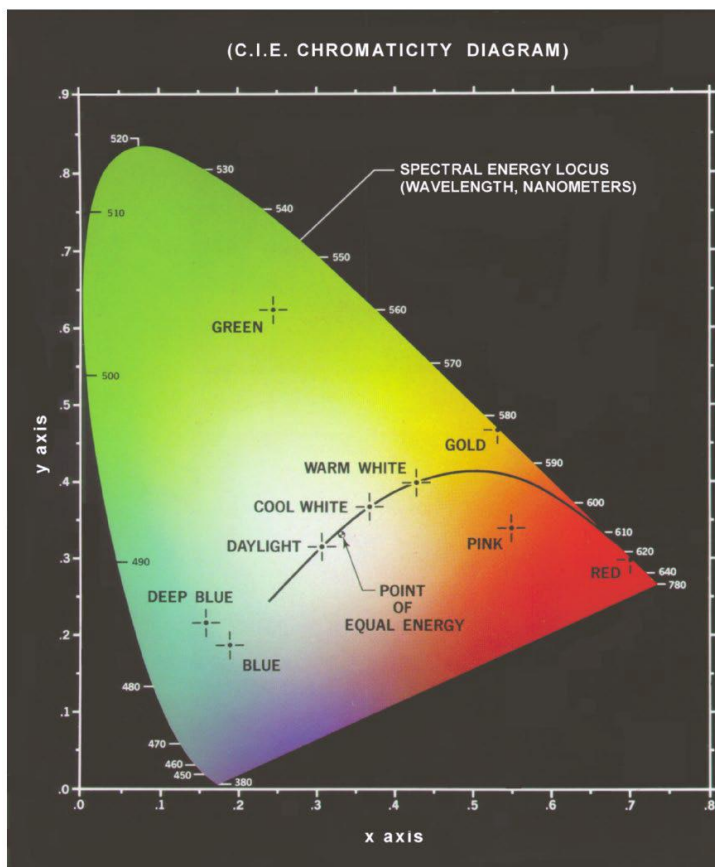
Diagramma cromatico CIE

Se X è la quantità di rosso

Se Y è la quantità di verde

Otteniamo Z come

$$z = 1 - (x + y)$$



1. Tutti i colori dello spettro del visibile li troviamo puri lungo i bordi, al centro in quanto punto di uguale energia abbiamo il bianco;
2. Se uniamo due colori con una linea, avremo tutti i colori ottenibili mischiando i due colori, quindi se mischiamo un colore con il bianco avremo tutte le sue tonalità;
3. Se uniamo tre colori con un triangolo tutti i colori lungo il bordo sono i colori ottenibili mischiando i tre colori;
4. Unendo Rosso Green e Blue si ottiene un triangolo che contiene tutti i colori che si possono riprodurre, questo triangolo non copre tutta l'area, quindi non tutti i colori si possono riprodurre unendo RGB;
5. I colori utilizzabili da una stampante sono minori rispetto a quelli visibili in un monitor perché è differente il modo di mischiare i colori.

Lezione 6:

Modelli di colore

Lo scopo di un modello del colore è di consentire la specifica dei colori con modalità standardizzate, che normalmente sono in coordinate 3D, i modelli di colori più utilizzati sono orientate all'HW di acquisizione e restituzione delle immagini o alla loro trasmissione, compressione ed elaborazione tramite trattamento del colore (RGB/CMY/YUV/YIQ/YCbCr), ci sono tanti modelli per lo stesso tipo di modello e usano in maniera ambigua lettere nei nomi (l o v significano cose differenti in modelli differenti)

CIE L*a*b*

il principale difetto CIE è che non è dotato di uniformità percettiva (dati 2 colori C1, C2 consideriamo le distanze dal colore C3 e dal colore C4. Supponiamo che le due distanze siano le stesse quantitativamente uguali, sarebbe desiderabile che le distanze con C3 e C4 siano percepite uguali... ma in generale per come è composto il sistema saranno percepite come distanze differenti).

La CIE nel 1976 ha standardizzato 2 spazi di colore CIE L*u*v e CIE L*a*b*... attraverso formule si può passare dallo spazio CIE XYZ allo spazio CIELAB

$$L^* = 116 \left(\frac{Y}{Y_n} \right)^{\frac{1}{3}} \quad a^* = 500 \left[\left(\frac{X}{X_n} \right)^{\frac{1}{3}} - \left(\frac{Y}{Y_n} \right)^{\frac{1}{3}} \right] \quad b^* = 200 \left[\left(\frac{Y}{Y_n} \right)^{\frac{1}{3}} - \left(\frac{Z}{Z_n} \right)^{\frac{1}{3}} \right]$$

Dove $\frac{X}{X_n}, \frac{Y}{Y_n}, \frac{Z}{Z_n}$ sono maggiori di 0.01, dove X_n, Y_n, Z_n definiscono il punto bianco, la metrica CIELAB si ottiene dalla seguente formula:

$$\Delta E_{ab}^* = \sqrt{(\Delta L^*)^2 + (\Delta a^*)^2 + (\Delta b^*)^2}$$

L^* = Luminanza o Luminosità (0 rappresenta il nero, 100 rappresenta il bianco)

a^* e b^* = cromaticità

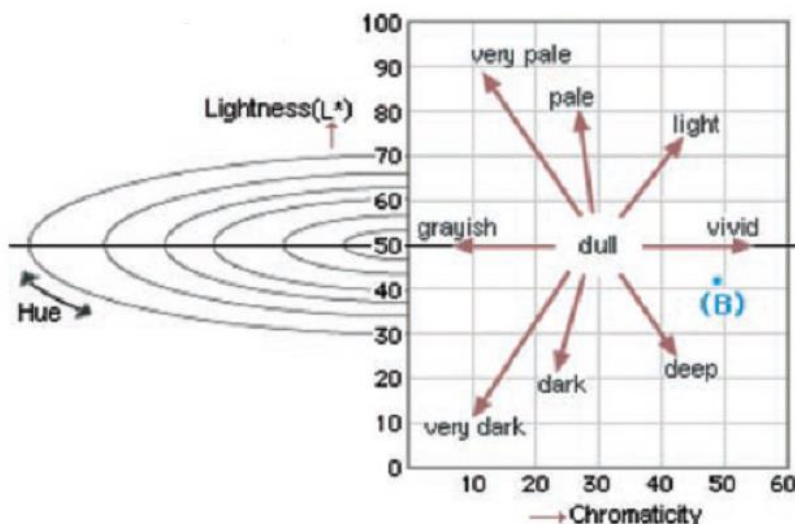
$L^* a^* b^*$ rappresentano la trasformazione dei tre valori del tristimolo XYZ

Nello spazio CIELAB la distanza tra due punti in $L^* a^* b^*$ definisce le differenze di colore

Spazio $L^* C^* h^*$

C^* = Chroma

h^* = Angolo di tonalità



Lezione 6:

Il bianco e nero

Quali sono le coordinate del nero?

$X = Y = Z = 0$ (nessun corpo reale possiede questo nero)

Quali sono le coordinate del bianco?

$X = Y = Z = 1$ oppure $= 1/3$ (varia in base all'illuminante il bianco)

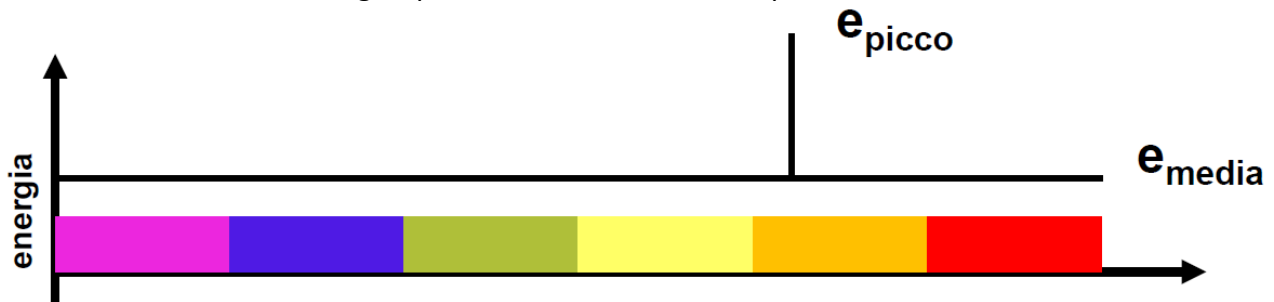
Per quanto riguarda bianco e nero osservatore ed illuminante devono essere scelti, come illuminante si usa il D65 e le comparazioni visive vanno fatte in situazione visiva controllata.

Spettro

Lo spettro di un illuminante è il diagramma dei contributi di energia che esso presenta per ogni lunghezza d'onda, spettri diversi possono produrre colori eguali, coppie di spettri si chiamano metameri.

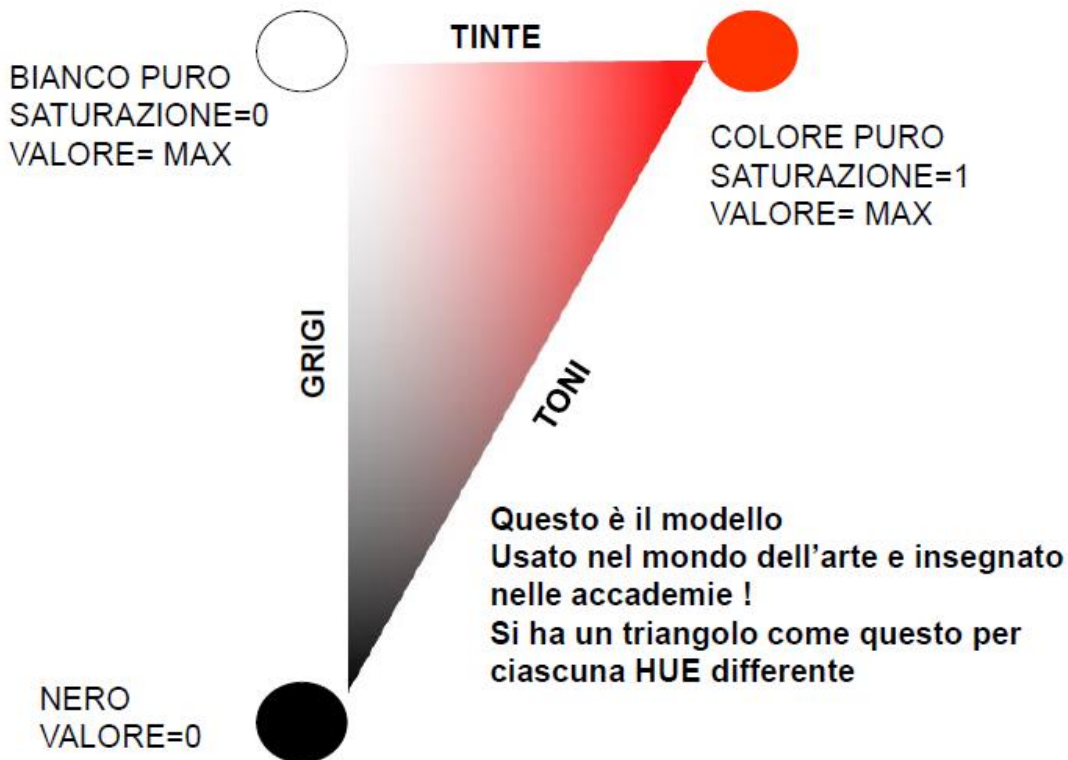
Modello del Pittore

Ogni spettro ha un metamero in questa forma :



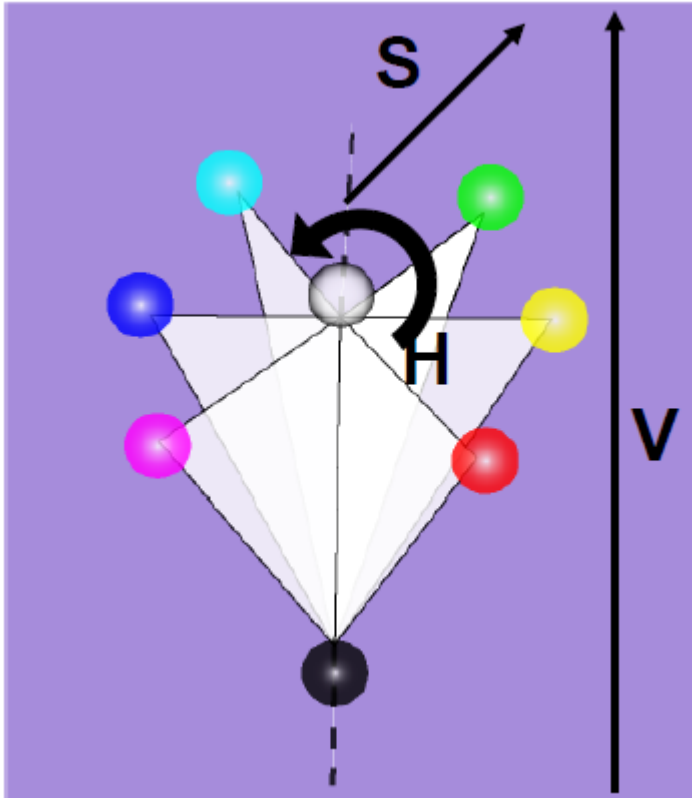
il rapporto $\frac{(energia\ di\ picco - energia\ media)}{(energia\ di\ picco + energia\ media)}$ equivale alla saturazione, l'energia media è proporzionale al contenuto energetico della radiazione, può essere considerata come una misura della luminosità di una radiazione e serve a dare un contributo bianco al colore percepito.

Modello del pittore

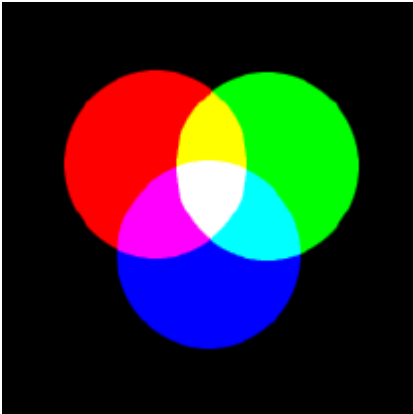


Spazio HSV

Dal modello del pittore deriva lo spazio HSV, ogni colore avrà un triangolo come quello sopra e tramite il lato dei grigi viene collegato agli altri colori formando una girandola, dando origine alla girandola HSV, dove H = Hue saranno i vari colori che la compongono ordinati in sequenza, S sarà la saturazione (minima al centro, massimo all'esterno), V sarà la luminosità, è un modello molto intuitivo, visivamente i parametri HSV vengono rappresentati meglio, ma non è un modello lineare e non si conoscono i colori base



Sintesi Additiva



Questa è la sintesi additiva dei colori RGB, sovrapponendoli tutti e 3 si ha il bianco, sovrapponendoli a due a due si ottiene, giallo, magenta e ciano (i monitor e le TV funzionano per sintesi additiva)

I Colori Complementari

Abbiamo i colori complementari quando la somma di un dato colore ad un altro dà il bianco e viceversa quindi quando al bianco si sottrae un dato colore...

$$R + G + B = W$$

$$R + G = W - B = Y$$

$$R + B = W - G = M$$

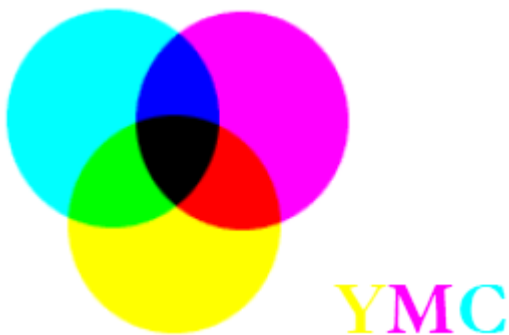
$$B + G = W - R = C$$

Il Giallo è il complementare del Blu (B)

Il Magenta è complementare del Verde (G)

Il Ciano è complementare del Rosso (R)

Sintesi Sottrattiva



Questa è la sintesi sottrattiva dei colori CMY, Se si sovrappongono tutti e 3 si ha il nero, se si sovrappongono a due a due abbiamo rosso, verde, blue (questa modalità di colori viene utilizzata nell'industria della stampa).

Lezione 6:

Nel modello RGB i colori primari sono : Rosso, Verde, Blu combinandoli a 2 a 2 si ottengono i colori secondari Magenta, Giallo, Ciano.

Riepilogo:

1. Non sempre conviene usare RGB o CMY;
2. La tinta descrive il colore presente;
3. La saturazione esprime la vivacità del colore(pastello, intenso, vicino al bianco);
4. La luminanza tende a valutare la sensazione luminosa ricevuta dall'occhio;

Rappresentazione luminanza-crominanza

La luminanza fornisce una versione in scala di grigi dell'immagine, mentre la crominanza fornisce i colori, queste rappresentazioni sono importanti nella compressione delle immagini, l'occhio umano reagisce meglio alla luminanza che ai colori; quindi conviene spendere più bit per registrare la luminanza che le crominanze.

Spazio YUV

La famiglia di spazi YUV viene utilizzata per la codifica di immagini o video analogici, in realtà intendiamo la famiglia di spazi che separano luminanza e crominanza, una controparte potrebbero essere gli spazi YC_bC_r .

Da RGB a YUV

La luminanza Y può essere ottenuta mediante somma pesata(dando maggior peso al verde) delle intensità luminose di rosso, verde e blu.

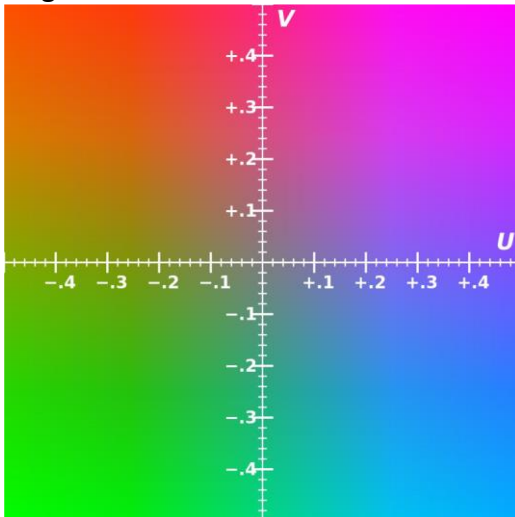
$$Y = 0.299R + 0.587G + 0.114B$$

La crominanza invece è definita come la differenza pesata tra il colore e un bianco di riferimento alla stessa luminanza.

$$U = 0.564(B - Y) = -0.169R - 0.331G + 0.5B$$

$$V = 0.713(R - Y) = +0.5R - 0.419G - 0.081B$$

Nel caso in cui R, G, B siano compresi tra 0 e 1 si ottengono valori di Y compresi tra 0 e 1 e valori compresi di U e V compresi tra -0.5 e 0.5, se R = G = B avremo solo scala di grigi perché U e V valgono 0.



In figura sono rappresentati i colori al variare di U e V con $Y = 0.5$

Da YUV a YC_bC_r

Nel caso in cui RGB siano interi compresi tra 0 e 255 si ottengono

C_b (Crominanza Blu)e C_r (Crominanza Rosso) shiftando U e V(Y è lo stesso di YUV)

$$Y = 0.299R + 0.587G + 0.114B$$

$$C_b = U + 128$$

$$C_r = V + 128$$

Lo spazio YC_bC_r è utilizzato nel formato JPEG

Lezione 6:

Palette di Colori(Look-UP-Table LUT)

Le LUT sono tabelle di valori che vengono utilizzate per modificare e correggere i colori, contengono informazioni su come i colori devono essere mappati sui colori desiderati, in pratica andiamo a modificare la tonalità, saturazione e la luminosità.

Re-indexing delle LUT

È quel processo che serve a diminuire l'entropia di un'immagine andando ad associare ai colori di pixel adiacenti indici numericamente vicini, per una palette con M colori bisogna cercare l'entropia in $M!$ possibili ordinamenti.

Lezione 7:

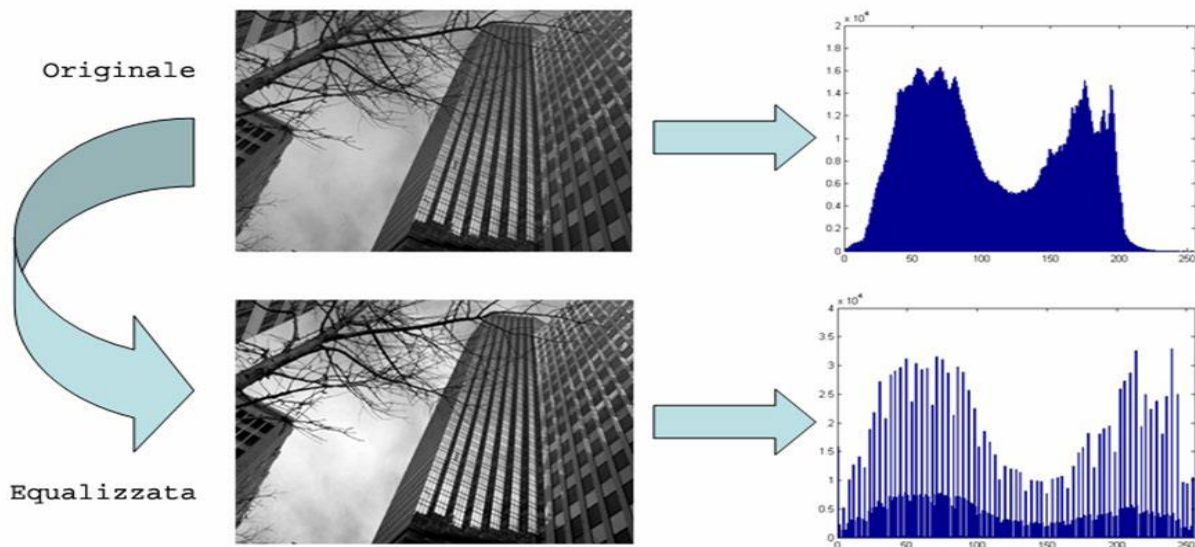
Un istogramma è quel grafico che per ogni livello di grigio, riporta il numero di pixel di quel colore, in un immagine $I(m, n)$ abbiamo $H(k)$ come numero di pixel di valore k , la somma di tutte le H è esattamente $M \times N$, non tiene conto della distribuzione spaziale dei pixel, infatti si possono avere istogrammi simili anche su immagini totalmente differenti, un immagine chiara avrà un istogramma denso a destra, mentre un immagine più scura lo avrà più denso a sinistra, per quanto riguarda le immagini sottoesposte si ha un istogramma denso a sinistra per poi andare verso destra in un pettine, al contrario un'immagine sovraesposta avrà l'istogramma denso a destra, esiste una funzione di contrast stretching che produrrà un istogramma a pettine, il contrast stretching serve a distribuire i bin dell'istogramma (che inizialmente sono tutti concentrati in un area dell'istogramma) in tutto l'istogramma, possiamo operare aritmeticamente sulle immagini ma potremmo incappare in 3 problemi:

1. Avere un valore negativo
2. Avere un valore maggiore del massimo
3. Avere un valore non intero

Per risolvere i primi due problemi basta riportare il bin al massimo o al minimo dipendentemente dalla problematica oppure basta normalizzare il range, mentre quando si ha un valore non intero basterà arrotondare o troncare il valore.

$$V_{Nuovo} = 255 \times \frac{(V_{vecchio} - \min_{osservato})}{(\max_{osservato} - \min_{osservato})}$$

Parliamo di equalizzazione quando l'istogramma dell'immagine è uniforme o appiattito quindi si ha che le varie tonalità di grigio sono pressappoco eguali.



Uno degli algoritmi di equalizzazione potrebbe essere l'equalizzazione dei livelli di grigio, dove

$R_k = \text{livello di grigio}$

$n_k = \text{numero di pixel nel livello di grigio}$

$$P_r(r_k) = \frac{n_k}{MN} \quad k = 0, 1, 2, \dots, L - 1$$

Per semplificare la trattazione del problema lavoreremo solo su immagini a scala di grigi pertanto opereremo separatamente sui tre canali principali (RGB) trattando ogni canale come se ogni livello sia un immagine indipendente dagli altri canali.

Lezione 7:

Le elaborazioni nel dominio spaziale possono essere espresse come una funzione :

$$g(x, y) = T(f(x, y))$$

f = immagine di input;

g = immagine di output;

T = operatore definito nell'intorno (x, y)

Esistono tre tipi di operazioni:

1. Operazione puntuale (il pixel di output coincide con il pixel di input);
2. Operazione locale (il pixel di input sarà il centro della zona da modificare);
3. Operazione globale (l'output interessa l'intera immagine di input):

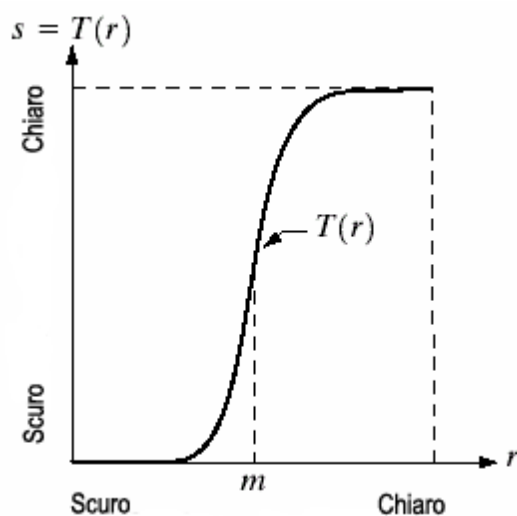
le operazioni puntuali tipiche sono :

1. Aggiustare sotto o sovraesposizione
2. Negativo
3. Espansione del contrasto
4. Equalizzazione
5. Presentazione in falsi colori

Un operatore puntuale può essere rappresentato da una funzione che preso in input un valore $f(x, y)$ lo modifica in un valore $g(x, y) = T(f(x, y))$, dato che un operatore puntuale dipende solo dal valore del pixel in input si può descrivere anche tramite una tabella simile :

IN	0	1	2	3	4	5	6	7	...
OUT	T(0)	T(1)	T(2)	T(3)	T(4)	T(5)	T(6)	T(7)	...

Un grafico a forma di curva come questo rappresenta una LUT



Lezione 7:

La più semplice operazione puntuale è il negativo, perché consiste nell'associare al valore $f(x, y)$ il valore $255 - f(x, y)$

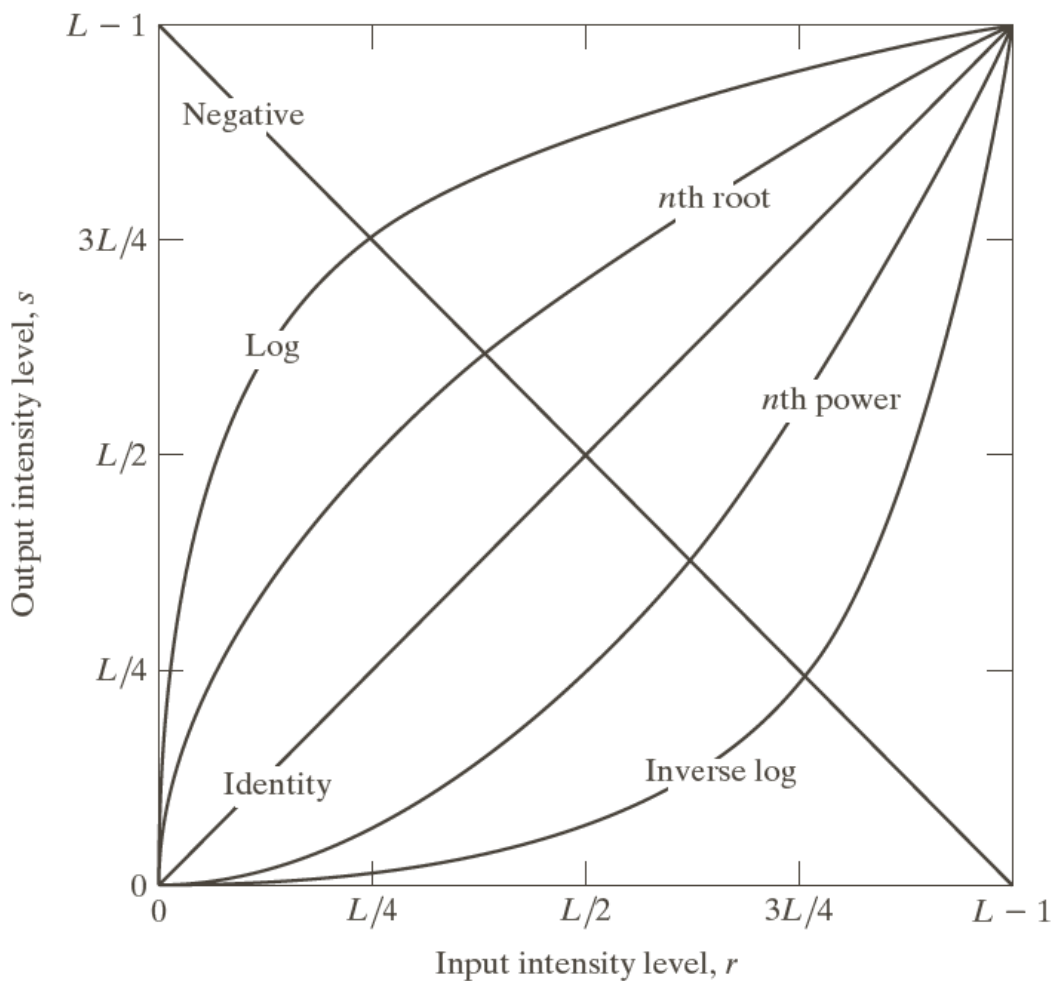
Trasformazione logaritmica

È una trasformazione che comprime la gamma dinamica, permette quindi di visualizzare l'immagine in una scala di grigi usuale, viene usata per immagini con escursioni di intensità molto ampie:

$$g(x, y) = c \log(1 + f(x, y))$$

c = è una costante compresa tra 0 e 255;

$f(x, y)$ = immagine input;



Trasformazione di potenza

La trasformazione di potenza può essere espressa sottoforma di funzione:

$$g(x, y) = c(f(x, y))^\gamma$$

c = costante;

γ = costante;

le due costanti sono comprese (0, 255), quando si utilizzano valori di $\gamma < 1$ si hanno effetti analoghi alla trasformazione logaritmica, mentre per valore di $\gamma > 1$ avrà effetti opposti alla precedente.

Lezione 7:

Binarizzazione

È l'effetto bianco e nero, si sceglie una soglia T per tanto avremmo due scenari:

1. Valori $< T$ diventeranno tutti neri
2. Valori $\geq T$ diventeranno tutti bianchi

Variazioni di contrasto

Aumentare il contrasto significa cambiare il valore di un pixel con uno più scuro o più chiaro in base al colore del pixel di entrata.

Solarizzazione

È un effetto che si ottiene cambiando la tipologia di curva in una curva non monotona

Lezione 8:

Operazioni Locali

Il valore di uscita di ogni pixel dipende da un limitato intorno del corrispondente punto in input, vengono usate per migliorare la qualità delle immagini o per estrarre informazioni dall'immagine, tipo filtri questi sono ottenuti facendo la convoluzione tra l'immagine ed una matrice.

Operatori lineari

Un operatore $F : V \rightarrow W$ si dice lineare se per ogni coppia di V e per ogni coppia di scalari a, b abbiamo :

$$F(av1 + bv2) = aF(v1) + bF(v2)$$

Operatore shift invariant

È quell'operatore che ha lo stesso comportamento su tutte le immagini impulsive indipendentemente dalla posizione in cui si trova il pixel.

1. se F è lineare basta conoscere il comportamento su tutte le immagini impulsive;
2. se F è shift invariant si comporta allo stesso modo su tutti gli impulsi;
3. se F è sia shift invariant che lineare per descriverlo basta conoscere come si comporta su un solo impulso, la risposta a questo impulso si chiama Point Spread Function e serve a riconoscere l'operatore;

ad un operatore lineare e shift invariant corrisponde una maschera, ma vale anche il viceversa quindi ad una maschera corrisponde un simile operatore.

Esistono kernel infiniti e finiti, questa dimensione influenza la complessità dell'operazione di filtraggio, tale complessità dipende anche dal numero di pixel di un'immagine.

I filtri shift invariant e lineari sono filtri convolutivi, questi filtri rispettano delle proprietà:

1. la convoluzione utilizza questa notazione $h = f \otimes g$;
2. la convoluzione è commutativa $f \otimes g = g \otimes f$;
3. la convoluzione è associativa $(f \otimes g) \otimes h = f \otimes (g \otimes h)$;

se il kernel f ha dimensioni $k \times h$ e gli indici sono disposti in modo da avere il punto $(0, 0)$ nel centro, la formula va riscritta nella seguente maniera

$$h_{m,n} = \sum_{i=-\lfloor \frac{k}{2} \rfloor}^{\lfloor \frac{k}{2} \rfloor - 1} \sum_{j=-\lfloor \frac{h}{2} \rfloor}^{\lfloor \frac{h}{2} \rfloor - 1} (f_{i,j} * g_{m+i,n+j})$$

se il kernel f ha dimensioni $k \times h$ e gli indici sono disposti partendo da 1, la formula va riscritta nella seguente maniera

$$h_{m,n} = \sum_{i=1}^{h,k} f_{i,j} * g_{m+(i-k+\frac{k}{2}), n+(j-h+\frac{h}{2})}$$

Lezione 8:

Tipi di operatori locali

1. Mediano
2. Minimo e massimo
3. N-Box(media)
4. N-Binomiale

1. Filtro mediano

È un filtro non lineare utilizzato per rimuovere il rumore dalle immagini, si prendono i valori di una finestra di pixel intorno ad un pixel di interesse, il valore centrale sarà la mediana e sarà sostituita al valore originale del punto di interesse, i valori intorno saranno posti in ordine crescente.

2. Filtro massimo e minimo

Filtro minimo : si prende un intorno $m \times m$ di un pixel e si sostituiscono i valori dei pixel con il valore minimo ritrovato nell'intorno, otteniamo così un incupimento delle immagini.

Filtro massimo : si prende un intorno $m \times m$ di un pixel e si sostituiscono i valori dei pixel con il valore massimo ritrovato nell'intorno, otteniamo così uno schiarimento delle immagini.

3. Filtro N-Box(media)

È una variante del filtro mediano, solo che questo filtro utilizza una finestra di analisi $N \times N$ ed utilizza un algoritmo di ordinamento per determinare il valore del punto di interesse, rispetto al filtro mediano utilizza un intorno più grande e conserva meglio i dettagli

4. Filtro N-Binomiale

È una variante del filtro mediano che però utilizza un intorno binomiale per determinare il valore di un punto dell'immagine.

Con termini fisici si dice che un filtro conserva l'energia se la somma dei suoi pesi fa 1, tutti i filtri visti prima sono energy preserving perché la somma dei valori della luminanza nella immagine non cambia, i filtri visti servono a ridurre il rumore in un immagine, più è grande il kernel migliore sarà il risultato rischiando però di aumentare la sfocatura, i filtri N-box e N-binomiali sono usati anche per sfocare, più è grande il kernel maggiore sarà la sfocatura e si riduce meglio il rumore.

Esistono due tipi di rumore

1. Rumore impulsivo (rumore sale e pepe) viene caratterizzato dalla frazione dell'immagine modificata in %;
2. Rumore gaussiano bianco, viene caratterizzato dalla media e dalla varianza;

Lezione 8:

1. Rumore impulsivo

$$p(z) = \begin{cases} P_a & \text{per } z = a \\ P_b & \text{per } z = b \\ 0 & \text{altrimenti} \end{cases}$$

Se a e b sono valori saturi, cioè sono uguali a $a = 0$ e $b = 255$, abbiamo il rumore sale e pepe, per eliminare o ridurre questo tipo di rumori basta usare il filtro mediano che ha dimensione kernel 3×3 .

2. Rumore gaussiano

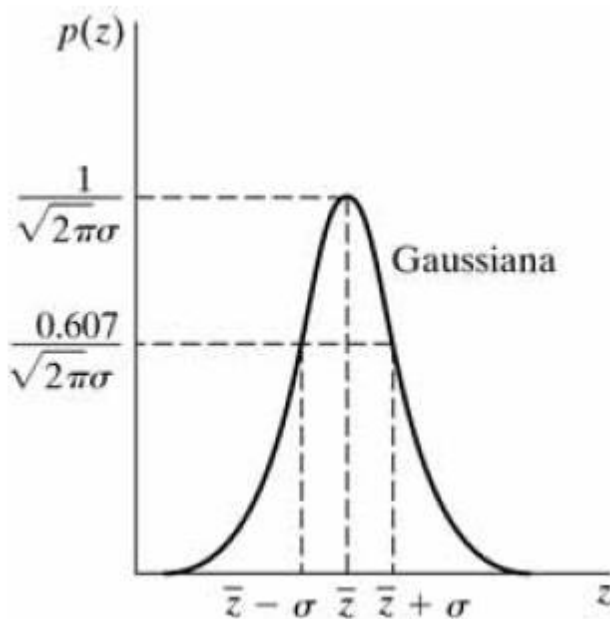
$$P(z) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(z-\bar{z})^2/2\sigma^2}$$

$z = \text{intensità}$

$\bar{z} = \text{valore medio di } z$

$\sigma = \text{deviazione standard}$

$\sigma^2 = \text{varianza di } z$



Quando z viene descritta dalla seguente formula, circa il 70% dei valori ricade nell'intervallo:

$$[(\bar{z} - \sigma), (\bar{z} + \sigma)]$$

E circa il 95% dei valori ricade nell'intervallo:

$$[(\bar{z} - 2\sigma), (\bar{z} + 2\sigma)]$$

Per rimuovere questo tipo di rumore basta utilizzare un filtro media con kernel 3×3

Riepilogando:

1. I filtri di media creano livelli di grigio non esistenti
2. I filtri di media non attenuano solo il rumore ma tutti i canali
3. Il filtro mediano elimina i picchi con base piccola rispetto al kernel

Lezione 8:

Esistono altri filtri per la rimozione del rumore e sono

1. Outlier (il valore del pixel centrale viene confrontato con la media dei suoi 8 vicini e se il valore assoluto della differenza tra pixel centrale e media è maggiore di una certa soglia, il pixel viene sostituito dal valore medio)
2. Olimpico (si scartano massimo e minimo di un intorno e si fa la media)

Estrazione dei contorni

Grazie agli operatori locali possiamo estrarre i contorni, discontinuità locali della luminanza, di una immagine, gli Edge detector servono a preservare le variazioni di luminanza ed eliminare le altre informazioni, ci sono degli edge detector basati sulla derivata prima, di un segnale monodimensionale, scoprendo che i lati sono in corrispondenza dei massimi della derivata; quindi, i filtri calcolano la derivata prima in x e y e poi le combinano.

Il filtro sobel x è un algoritmo edge detector, viene utilizzata una matrice 3x3 per scorrere tutta l'immagine e calcolare i cambiamenti nell'intensità del colore nella direzione x, il suo kernel è il seguente:

$$\begin{pmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{pmatrix}$$

L'effetto di questo filtro è di evidenziare i bordi verticali, il valore di ogni pixel è dato dalla somma dei prodotti del kernel e dei valori del pixel, esiste anche il filtro prewitt che è pressappoco uguale al sobel solo è meno sensibile al rumore; quindi, potrebbe essere meno preciso nella rilevazione dei bordi.

Esistono edge detector basati sulla derivata seconda di un segnale monodimensionale, in corrispondenza del lato la derivata passa per zero. Il filtro più diffuso è detto laplaciano ed usa il seguente kernel:

$$\begin{pmatrix} -1 & 0 & -1 \\ 0 & 4 & 0 \\ -1 & 0 & -1 \end{pmatrix}$$

Dopo aver applicato l'operatore laplaciano è necessario che si verifichi lo zero-crossing, cioè rispetto al punto in questione nel suo intorno ci deve essere un valore positivo ed uno negativo.

Esistono i filtri sharpening che servono a incrementare la nitidezza, si utilizza una maschera derivata dal filtro laplaciano che rinforza i lati ma rinforza anche il rumore dell'immagine.

Lezione 9:

Il bit plane di un immagine digitale a N bit, è un insieme di N immagini binari(piani), in cui l'immagine i-esime contiene i valori dell'i-esimo bit della codifica scelta, se usiamo una codifica in binario puro, i piani più significativa conterranno informazione sulla struttura dell'immagine, il resto forniscono i dettagli, i piani dal 7 al 3 conterranno dati importanti per la vista, più si va nei piani bassi maggior rumore ed errori troveremo, questa codifica la potremmo utilizzare se volessimo eliminare alcune frequenze di colori, per esempio per eliminare i grigi da 32 a 64 basterà mettere a 0 tutto il quinto piano dell'immagine, nell'utilizzo di questa codifica vi è un grande problema, una piccola variazione potrebbe ripercuotersi su tutti i piani, pertanto viene usato il codice gray a m-bit che gode della proprietà in cui ogni parola di memoria differisce dalla precedente per un solo bit, può essere calcolato tramite l'operator XOR dalla seguente formula:

$$g_i = a_i \otimes a_{i+1} \quad 0 \leq i \leq m - 2$$

$$g_{m-1} = a_{m-1}$$

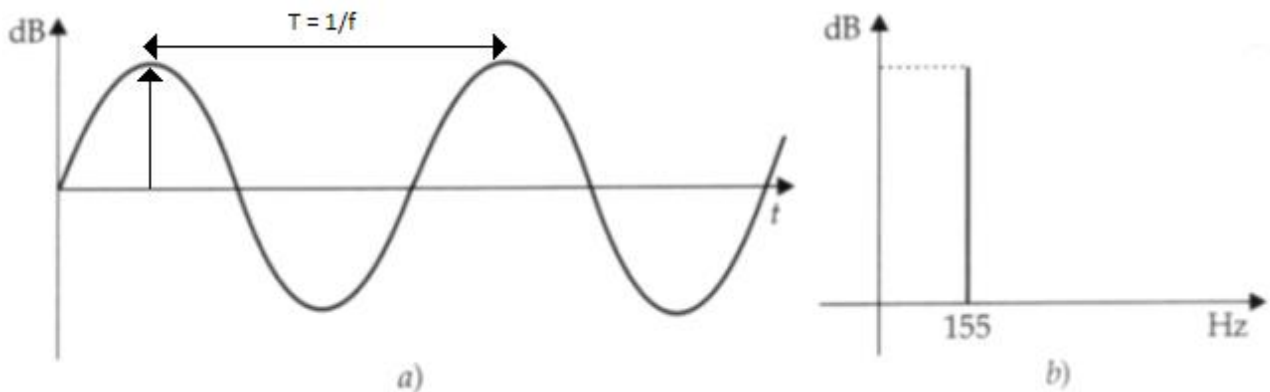
P.S.

I bit-plane delle immagini in gray code risultano tra loro più coerenti, se confrontati con i rispettivi piani in binario puro, se utilizziamo il gray code, il numero di transizione bianco-nero saranno inferiori rispetto al binario puro, si ha un entropia dell'immagine minore, grazie a queste caratteristiche sappiamo che possiamo utilizzare il gray code per comprimere immagini così codificate, vista la codifica differente, se cancelliamo piani di bit in gray code cancelleremo range di valori diversi rispetto al binario puro, anche nel gray code il rumore è situato nei piani bassi, ma eliminare questi piani potrebbe introdurre artefatti.

Lezione 10:

Una funzione periodica può essere espressa come somma di seni o coseni di differenti frequenze o ampiezze (Serie di Fourier), invece le funzioni non periodiche possono essere espresse come integrali di seni o coseni, moltiplicati per determinate funzioni-peso (Trasformata di Fourier) entrambe, trasformata e serie di Fourier, possono essere ricostruite tramite un processo di inversione senza perdita di informazione, lavorando nel dominio di Fourier, l'analisi di Fourier trovò utilizzo nel campo della diffusione del calore, con l'avvento della FFT (Fast Fourier Transform) il settore dell'elaborazione digitale ha subito una rivoluzione.

Un'immagine può essere vista come una funzione discreta in due dimensioni, dove i valori rappresentano i livelli di grigio del pixel, questa funzione può essere vista a sua volta come un segnale con una propria frequenza



La differenza di “altezza” fra il riferimento e la curva è l'ampiezza del segnale espressa in dB

T = Periodo espresso in Secondi;

f = Frequenza numero di cicli al secondo in Hertz;

Trasformata Discrete di Fourier(DFT)

nel 1-D la coppia trasformata-antitrasformata assume la seguente formula:

Trasformata

$$F_{(0)} = \sum_{x=0}^3 F_{(x)}$$

Antitrasformata

$$f_{(0)} = \frac{1}{4} \sum_{x=0}^3 F_{(u)} e^{i2\pi u_{(0)}}$$

Nel caso 2-D la coppia trasformata-antitrasformata della sequenza bidimensionale $f(x, y)$ assume la seguente forma:

Trasformata

Lezione 10:

$$F(u, v) = \frac{1}{MN} \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} f(x, y) e^{-2\pi i \left(\frac{ux}{M} + \frac{vy}{N} \right)} \quad \text{per } u = 0, 1, \dots, M-1; v = 0, 1, \dots, N-1$$

Antitrasformata

$$f(x, y) = \sum_{u=0}^{M-1} \sum_{v=0}^{N-1} F(u, v) e^{i2\pi \left(\frac{ux}{M} + \frac{vy}{N} \right)} \quad \text{per } x = 0, 1, \dots, M-1; y = 0, 1, \dots, N-1$$

(u, v) = indici degli assi di frequenza discretizzati

(M, N) = dimensione in pixel dell'immagine

Formula di Eulero

$$e^{ix} = \cos x + i \sin x; \forall x$$

Quindi:

$$e^{-ix} = \cos x - i \sin x$$

Trasformata di Fourier

Dato che la trasformata F ha valori complessi, possiamo esprimerla in termini reali e immaginari:

$$\text{Spettro della Trasformata} = |F(u, v)| = \sqrt{R^2(u, v) + I^2(u, v)}$$

$$\text{Angolo di Fase} = \phi(u, v) = \tan^{-1} \left[\frac{I(u, v)}{R(u, v)} \right]$$

$$\text{Potenza Spettrale} = P(u, v) = |F(u, v)|^2 = R^2(u, v) + I^2(u, v)$$

Che vantaggi otteniamo con la trasformata di Fourier?

1. Possiamo sopprimere frequenze indesiderate;
2. Ridurre lo spazio occupato dai dati;
3. Rigenerare segnali degradati;

oltre a quella di Fourier esistono altre trasformate che servono a restaurare segnali, comprimere e sono:

1. Trasformata discreta di Walsh (DWT);
2. Trasformata discreta di Hadamard (DHT)
3. Trasformata discreta del Coseno (DCT);
4. Trasformata discreta di Karhunen Loeve (KLT);

proprietà della DFT 2-D

1. Separabilità;
2. Traslazione;
3. Valor Medio;

Lezione 10:

Separabilità:

la trasformata di Fourier può essere espressa in forma separabile (si applica due volte la DFT 1-D)

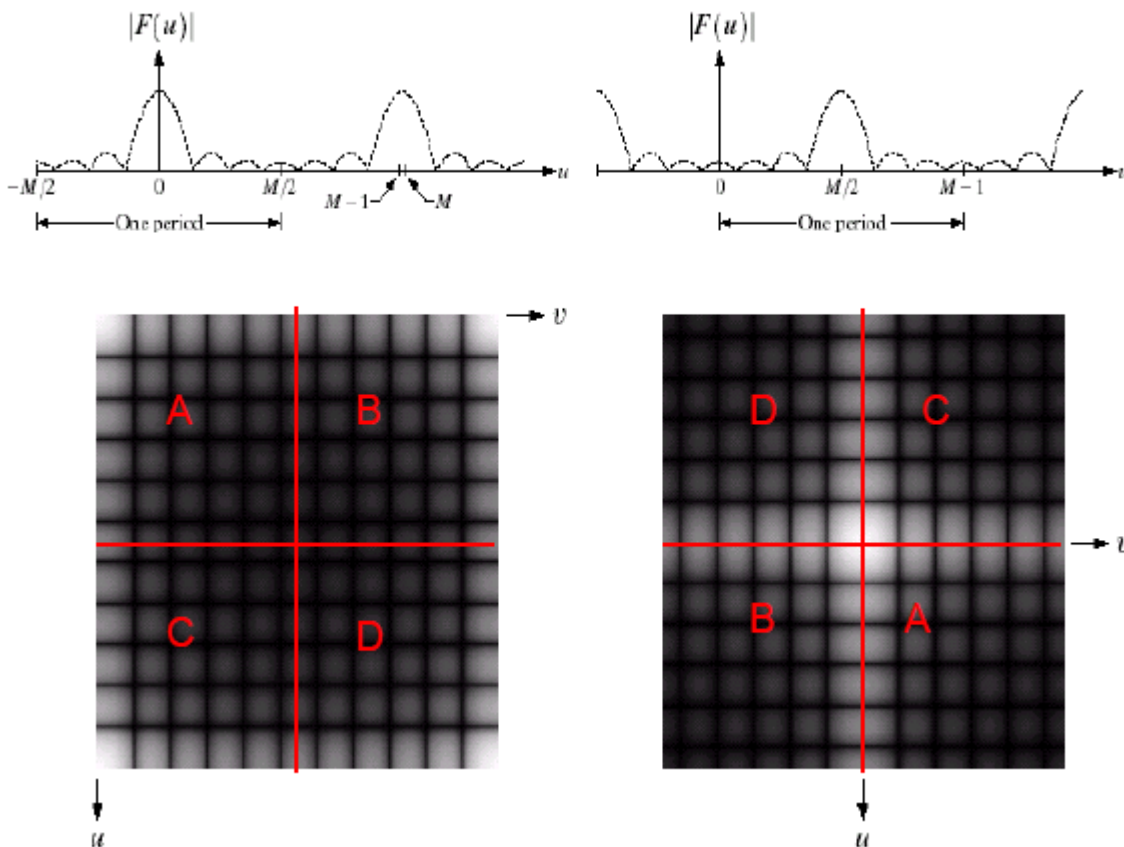
$$F(u, v) = \frac{1}{M} \sum_{x=0}^{M-1} g(x, v) e^{-\frac{i\pi ux}{M}}$$

Con:

$$g(x, v) = \left[\frac{1}{N} \sum_{y=0}^{N-1} f(x, y) e^{-\frac{i2\pi vy}{N}} \right]$$

Traslazione:

Nel caso bidimensionale è utile prima di operare sulla trasformata applicare una traslazione dell'origine in $(M/2, N/2)$, quindi al centro della matrice dei coefficienti delle frequenze, in questa maniera $F(0, 0)$ sarà il centro del rettangolo delle frequenze definito da 0 a $M-1/N-1$, inoltre con questo shift non si modifica la magnitudo della trasformata, in questa maniera si ha una visualizzazione migliore dello spettro.



Valor Medio

Il valore medio della trasformata al punto $(0, 0)$ è dato da:

$$F(0, 0) = \frac{1}{NxN} \sum_{x=0}^{N-1} \sum_{y=0}^{N-1} f(x, y) \quad \bar{f}(x, y) = \frac{1}{NxN} F(0, 0)$$

Lezione 10:

Altro non è che la media di $f(x, y)$, $F(0, 0)$ prende il nome di componente continua o componente DC.

Fast Fourier Transform

$$F(u) = \frac{1}{N} \sum_{x=0}^{N-1} f(x) \exp \left[-\frac{i\pi ux}{N} \right]$$

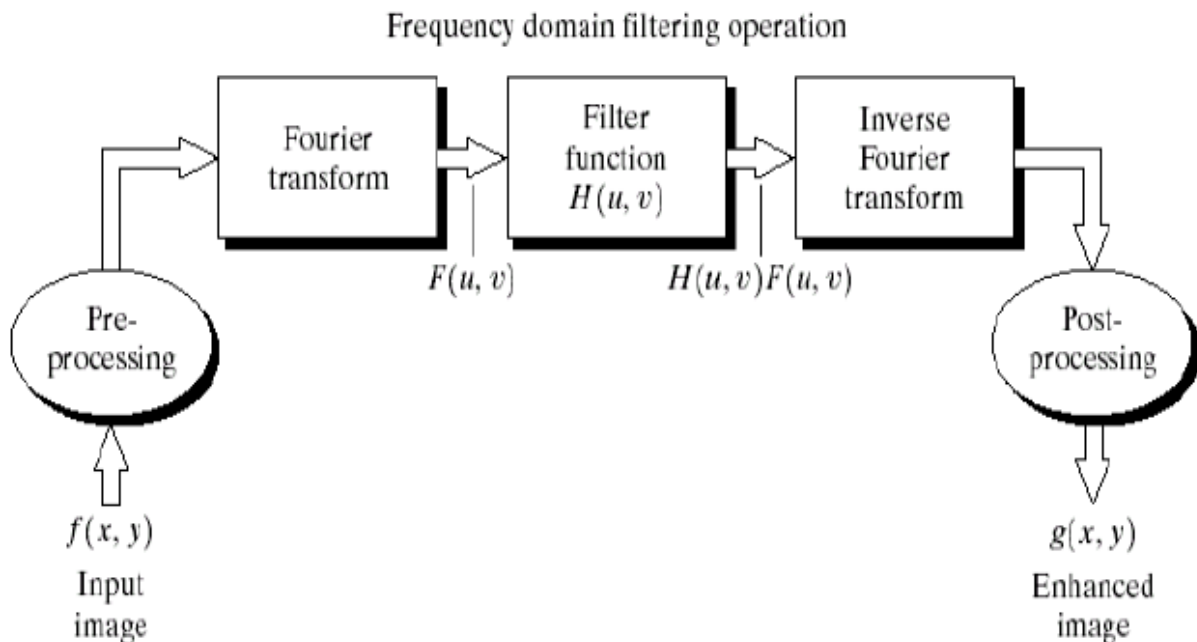
Nella sua forma classica la trasformata di Fourier richiederebbe un numero di operazione proporzionale a N^2 ma utilizzando opportune tecniche di decomposizione è possibile abbassare la complessità a $N \log_2 N$, implementando così la Fast Fourier Transform(FFT).

Frequenze: Low and High

Tolti i casi banali è normalmente impossibile fare associazioni dirette fra specifiche parti dell'immagine e la sua trasformata, ricordando che la frequenza è legata alla velocità di variazione è possibile associare:

1. Basse frequenze = zone uniformi;
2. Alte frequenze = variazioni più o meno brusche, quindi bordi o rumore;

Filtraggio nel Dominio della Frequenza



La funzione $H(u, v)$ prende il nome di filtro, poiché agisce su alcune frequenze della trasformata lasciando le altre immutate, la funzione h spesso è una funzione reale e ciascuna sua componente moltiplica sia la corrispondente componente reale, sia quella immaginaria della F , questo tipo di filtri è chiamato "zerophaseshift" perché non introduce sfasamento.

Lezione 10:

Teorema della convoluzione

“la convoluzione di due segnali nel dominio spaziale equivale all’antitrasformata del prodotto delle frequenze”

DFT nel dominio spaziale:

$$g(x, y) = f(x, y) * h(x, y) = \frac{1}{MN} \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} f(m, n) h(x - m, y - n)$$

DFT nel dominio delle frequenze:

$$G(u, v) = F(u, v) * H(u, v)$$

Operazione di convoluzione nel dominio spaziale:

$$g(x, y) = f(x, y) * h(x, y) = \frac{1}{MN} \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} f(m, n) h(x - m, y - n)$$

Operazione di convoluzione nel dominio delle frequenze:

$$g(x, y) = F^{-1}\{F(u, v)H(u, v)\}$$

Complessità per un segnale 1D:

1. Nel dominio delle frequenze $O(n \log n)$;
2. Nel dominio spaziale $O(n^2)$;

se il filtro ha dimensioni confrontabili con quelle dell’immagine è più efficiente computazionalmente effettuare il filtraggio nel dominio delle frequenze, con maschere piccole conviene usare il calcolo nel dominio spaziale, definire un filtro nel dominio delle frequenze viene meglio.

Filtro passa-basso ideale

Questi tipi di filtri eliminano totalmente tutte le componenti di frequenza che nel rettangolo delle frequenze distano dall’origine più di D_0 (frequenza di taglio), i filtri ideali causano un forte fenomeno di sfocatura ad anello (Ringing)

Filtro passa-basso di Butterworth

il filtro di butterworth è un filtro utilizzato per rimuovere le frequenze alte da un segnale, quindi man mano che la frequenza aumenta, meno viene lasciato passare iniziando ad assumere le stesse caratteristiche e gli stessi difetti del filtro ideale. Applicando un filtro di butterworth del secondo ordine si può vedere una differenza minore di blurring ed un effetto di ringing assente.

$$H(u, v) = \frac{1}{1 + \left[\frac{D(u, v)}{D_0}\right]^{2n}}$$

Filtro Gaussiano

I filtri gaussiani sono caratterizzati dall’aver come trasformata una gaussiana:

$$H(u, v) = e^{-\frac{D^2(u, v)}{2D_0^2}}$$

Lezione 10:

i filtri passa basso possono essere utilizzati in:

1. Riconoscimento di caratteri;
2. Processamento di immagini aeree e satellitari;

come esistono diversi tipi di filtri passa basso esistono diversi tipi di filtri passa alto.

Filtro Band-Reject

I filtri passa banda costruiscono una maschera opportuna in un determinato range in grado di eseguire il filtraggio di tali frequenze

Lezione 11:

Compressione

Il termine compressione si utilizza per identificare la tecnica di elaborazione dati che tramite specifici algoritmi (elimina le ridondanze di informazione dai dati) va a ridurre i bit necessari alla rappresentazione digitale di una informazione, questi algoritmi riducono lo spazio di archiviazione richiesto alla memorizzazione e riducono l'occupazione di banda necessaria in una trasmissione dati, i dati possono essere usati per rappresentare la stessa quantità di informazione, i dati irrilevanti o ripetuti vengono detti dati ridondanti.

Ridondanza della codifica

Una codifica è un sistema di simboli utili a rappresentare una certa quantità di informazioni, ad ogni "pezzo" di informazione e evento viene assegnata una codeword, caratterizzata da una lunghezza.

Ridondanza spaziale e temporale

Dal momento che la maggior parte degli array di intensità 2-D sono realizzati spazialmente (schermi), l'informazione è replicata inutilmente nei pixel correlati, in una sequenza video i pixel correlati temporalmente rappresentano un'informazione duplicata.

Classificazione dei metodi di compressione

Basata sul tipo di dati:

- a. Compressione Generica;
 - b. Compressione Audio;
 - c. Compressione Immagini;
 - d. Compressione Video;
2. Basata sul tipo di compressione:
- a. Compressione reversibile (lossless) (senza perdita di informazione);
 - b. Compressione irreversibile (lossy) (con perdita di informazione);

Compressione LOSSLESS

Questa compressione trasforma i dati, in modo da risparmiare memoria, e poi li riesce a ricostruire senza errore e perdita di alcun bit di informazione, questa compressione è necessaria per ridurre lo spazio occupato da documenti, eseguibili e tutto quello che può occupare meno spazio, per raggiungere una compressione senza perdita si utilizza il primo teorema di Shannon che fa uso dell'Entropia delle sequenze, la quantità media di informazione associata alla singola generazione di un simbolo nella sequenza, più è grande l'incertezza della sequenza maggiore è l'entropia, il valore massimo si ha quando i simboli sono equiprobabili.

$$E = - \sum f_i \log_2(f_i) \quad i \in S$$

Teorema di Shannon

"per una sorgente discreta e a memoria zero, il bitrate minimo è pari all'entropia della sorgente."

Quindi i dati possono essere rappresentati senza perdere informazione usando almeno un numero di bit pari a

$$N * E$$

N = numero di caratteri;

E = Entropia;

il teorema di Shannon fissa il numero minimo di bit, ma per trovarli occorre usare l'algoritmo di Huffman.

Codifica di Huffman

Questa codifica ha le seguenti proprietà:

1. Lunghezza Variabile associando ai simboli meno frequenti codici lunghi ed ai simboli più frequenti codici corti;
2. Nessuno codice è prefisso di altri;
3. Tende al limite imposta dal teorema di Shannon;

Algoritmi LOSSLESS

Codifica Run-length (RLE)

Le immagini che hanno delle ripetizioni di intensità lungo le righe o le colonne possono essere compresse rappresentando queste ripetizioni di intensità come coppie di run-length, dove ciascuna coppia individua l'inizio di una nuova intensità e il numero di pixel consecutivi.

Esempio:

00000111001011101110101111111

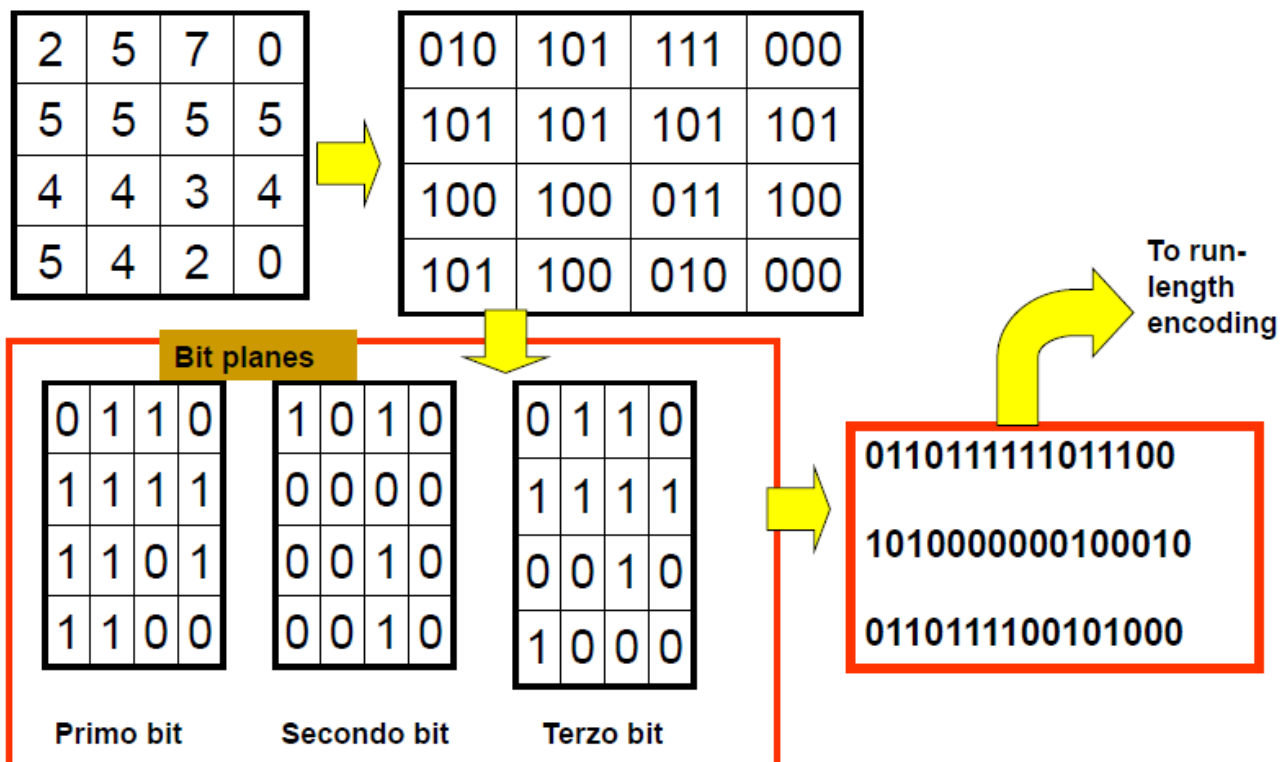
Può essere espresso come :

1. 5 volte 0, 3 volte 1, 2 volte 0, 1 volta 1, 1 volta 0, etc;

oppure possiamo segnare solo la lunghezza dei segmenti di simboli eguali:

2. 5, 3, 2, 1, 1, 3, 1, 3, 1, 1, 1, 7;

dopodiché dobbiamo trasformare questa codifica in binario:



Codifica Differenziale

Se la sequenza dei valori varia lentamente è possibile ricordarsi solamente il primo valore, e per i successivi ricordarsi le differenze successive.

134, 137, 135, 128, 130, 134, 112

Possiamo rappresentarla come : memorizzo il primo valore : 134

-3, 2, 7, -2, -4, 22

Per le immagini la sequenza delle differenze ha un'entropia minore di quella dei singoli valori, per tanto richiede meno bit per essere memorizzata.

Compressione LOSSY

Questo tipo di compressioni, si basano sul “se percettivamente non è importante buttalo via” trasformano i dati per risparmiare memoria, in maggior quantità rispetto alla compressione LOSSLESS, ma quando si decomprimono i dati ci sarà una perdita di informazione.

Per una buona compressione LOSSY è meglio rispettare delle regole:

1. Fissare la massima distorsione accettabile, l'algoritmo di compressione deve trovare la rappresentazione con il più basso numero di bit;
2. Cercare il miglior algoritmo che a parità di bit dia la minima distorsione;

Requantization

È una compressione che riduce il numero di livelli disponibile in modo da risparmiare bit per pixel:

1. Canale R : usa i 4 bit più significativi;
2. Canale G: usa i 6 bit più significativi;
3. Canale B : usa i 2 bit più significativi

Con questa compressione si risparmia il 50% di bit, inoltre se si utilizza in combinazione la compressione LZW o Huffman queste saranno più efficienti.

Codifica JPEG/JPG

Passi fondamentali per una codifica JPEG:

1. Pre-processing:
 - a. Trasformazione da RGB a $Y C_b C_r$;
 - b. Sottocampionamento della cromaticanza;
 - c. Suddivisione dell'immagine in sotto-immagini per canali;
2. Trasformazione:
 - a. Trasformazione DCT (Trasformata Discreta del Coseno);
 - b. Quantizzazione;
3. Codifica:
 - a. Codifica dei Coefficienti DC;
 - b. Ordinamento Zig-Zag per i coefficienti AC;
 - c. Codifica dell'entropia

Trasformazione da RGB a $Y C_b C_r$:

$$\begin{bmatrix} Y \\ C_b \\ C_r \end{bmatrix} = \begin{bmatrix} 0.299 & 0.587 & 0.114 \\ 0.596 & -0.275 & -0.321 \\ 0.212 & -0.5230 & 0.311 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix}$$

Y = Luminanza;

C_b e C_r = codificano i colori;

Sottocampionamento della Crominanza:

Y = viene presa tutta;

C_b e C_r = ne prendiamo 1 ogni 4 per entrambi i valori;

N.B. in questo passo la perdita di informazione sarà irreversibile.

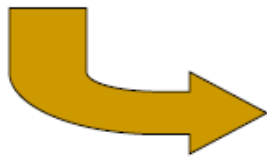
Partizione dell'immagine:

per usufruire al meglio della ridondanza, JPEG divide l'immagine in quadrotti 8x8 di 64 pixel non sovrapposti, ogni quadrotto avrà un'elaborazione differente, per questo si ha il problema della quadrettatura visibile negli ingrandimenti o nelle stampe, maggiore sarà la compressione più sarà evidente la quadrettatura.

Lezione 11:

Prima di passare alla trasformazione DCT ai 64pixel di ogni quadrotto viene sottratta una quantità pari a 2^{n-1} dove 2^n è il numero massimo di livelli di grigio dell'immagine, in questo modo si fa uno shift dei livelli di grigio, portando il grigio medio(128) = 0.

52	55	61	66	70	61	64	73
63	59	66	90	109	85	69	72
62	59	68	113	144	104	66	73
63	58	71	122	154	106	70	69
67	61	68	104	126	88	68	70
79	65	60	70	77	68	58	75
85	71	64	59	55	61	65	83
87	79	69	68	65	76	78	94



-76	-73	-67	-62	-58	-67	-64	-55
-65	-69	-62	-38	-19	-43	-59	-56
-66	-69	-60	-15	16	-24	-62	-55
-65	-70	-57	-6	26	-22	-58	-59
-61	-67	-60	-24	-2	-40	-60	-58
-49	-63	-68	-58	-51	-60	-70	-53
-43	-57	-64	-69	-73	-67	-63	-45
-41	-49	-59	-60	-63	-52	-50	-34

Trasformazione DCT:

è una trasformata di Fourier, dove viene decorellato al massimo i dati permettendo un maggior rapporto di compressione in fase di codifica, un'immagine di 8x8 pixel può essere vista come un vettore di dimensione 64, per tanto ogni immagine è la somma pesata di 64 immagini impulsive, immagini tutte nere, tranne in un pixel che avrà valore 1, dove i pesi della somma sono dati dal livello di luminosità del pixel, queste immagini impulsive rappresentano la base impulsiva che non è l'unica base.

$$F(u, v) = \frac{2}{N} \left[\sum_{x=0}^{N-1} \sum_{y=0}^{N-1} C(u)C(v)f(x, y) \cos \left(\frac{(2x+1)u\pi}{2 * N} \right) \cos \left(\frac{(2y+1)v\pi}{2 * N} \right) \right]$$

$$f(x, y) = \frac{2}{N} \left[\sum_{u=0}^{N-1} \sum_{v=0}^{N-1} C(u)C(v)F(u, v) \cos \left(\frac{(2x+1)u\pi}{2 * N} \right) \cos \left(\frac{(2y+1)v\pi}{2 * N} \right) \right]$$

Con:

$$C(u) = \frac{1}{\sqrt{2}} \quad \text{per } u = 0, C(u) = 1$$

Altrimenti

$$C(v) = \frac{1}{\sqrt{2}} \quad \text{per } v = 0, C(v) = 1$$

Dopo l'applicazione della DCT abbiamo questo:

-76	-73	-67	-62	-58	-67	-64	-55
-65	-69	-62	-38	-19	-43	-59	-56
-66	-69	-60	-15	16	-24	-62	-55
-65	-70	-57	-6	26	-22	-58	-59
-61	-67	-60	-24	-2	-40	-60	-58
-49	-63	-68	-58	-51	-60	-70	-53
-43	-57	-64	-69	-73	-67	-63	-45
-41	-49	-59	-60	-63	-52	-50	-34



-415	-29	-62	25	55	-20	-1	3
7	-21	-62	9	11	-7	-6	6
-46	8	77	-25	-30	10	7	-5
-50	13	35	-15	-9	6	0	3
11	-8	-13	-2	-1	1	-4	1
-10	1	3	-3	-1	0	2	-1
-4	-1	2	-1	2	-3	1	-2
-1	-1	-1	-2	-1	-1	0	-1

Basta moltiplicare i singoli coefficienti alle basi della DCT e otteniamo il blocco precedente, il coefficiente DC è solo il primo in alto a sinistra, i restanti 63 sono coefficienti AC, ogni immagine 8x8 si ottiene moltiplicando ciascuna immagine per un coefficienti e sommando il tutto, questi sono i coefficienti DCT,

Lezione 11:

mentre il primo in alto a sinistra è un valore proporzionale al valor medio della luminanza ed è detto coefficiente DC.

Quantizzazione :

non è conveniente usare un unico fattore di quantizzazione per tutti i 64 coefficienti DCT, per tanto per il coefficiente $F(i, j)$ si utilizza il fattore di quantizzazione $Q(i, j)$, questo fattore viene scelto dall'utente, che dovrà fornire la tabella assieme ai dati compressi, o preso da uno standard, la tabella dove ci sono tutti i fattori di quantizzazione si chiama tabella di quantizzazione, nella tabella l'utente dovrà fornire anche un Quality Factor che va da 1 a 100, maggiore sarà il QF, minore la quantizzazione e quindi minore la perdita di informazione.

Effetto della quantizzazione:

-415	-29	-62	25	55	-20	-1	3	16	11	10	16	24	40	51	61
7	-21	-62	9	11	-7	-6	6	12	12	14	19	26	58	60	55
-46	8	77	-25	-30	10	7	-5	14	13	16	24	40	57	69	56
-50	13	35	-15	-9	6	0	3	14	17	22	29	51	87	80	62
11	-8	-13	-2	-1	1	-4	1	18	22	37	56	68	109	103	77
-10	1	3	-3	-1	0	2	-1	24	35	55	64	81	104	113	92
-4	-1	2	-1	2	-3	1	-2	49	64	78	87	103	121	120	101
-1	-1	-1	-2	-1	-1	0	-1	72	92	95	98	112	100	103	99

-26	-3	-6	2	2	0	0	0
1	-2	-4	0	0	0	0	0
-3	1	5	-1	-1	0	0	0
-4	1	2	-1	0	0	0	0
1	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0

Tutti i coefficienti DCT vengono riordinati in un array di 64 unità seguendo l'ordinamento a serpentina per poi essere codificati in un altro stream.

0	1	5	6	14	15	27	28
2	4	7	13	16	26	29	32
3	8	12	17	25	30	41	43
4	11	18	24	31	40	44	53
10	19	23	32	39	45	52	54
20	22	23	28	46	51	55	60
21	34	37	47	50	56	59	61
35	36	48	49	54	58	62	63

Lezione 11:

Codifica coefficienti DC e AC:

i coefficienti della posizione (1, 1, alto a sinistra) vengono codificati tramite codifica differenziale, mentre i coefficienti AC tramite la codifica RLE