Fondamenti di Visione Artificiale

PhD. Ing. Folgheraiter Michele

Corso di Robotica
Prof.ssa Giuseppina Gini
Anno Acc. 2006/2007

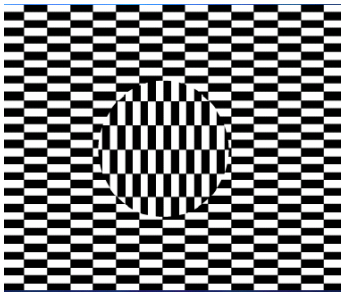
Visione artificiale: insieme dei processi che mirano ad estrarre, caratterizzare e interpretare le informazioni provenienti dalle immagini di un mondo tridimensionale. Creare un modello approssimato del mondo tridimensionale partendo da immagini bidimensionali.

- Uno degli obbiettivi della visione artificiale è quello di emulare la visione biologica (umana e non).

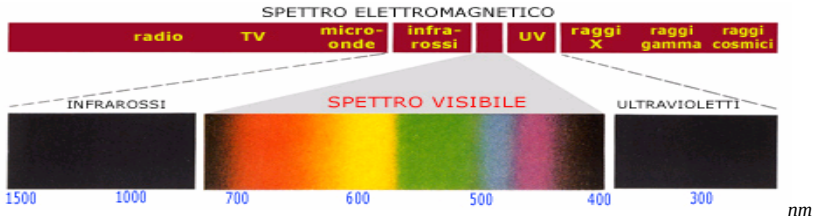
Visione naturale: Il processo visivo umano è molto complesso e si avvale di una parte rilevante del sistema nervoso centrale.

- Sebbene rappresenti il nostro modello di riferimento, non è infallibile.

Illusioni ottiche



- Incapacità di rilevare con precisione la distanza degli oggetti.
- Incapacità di rilevare radiazioni luminose fuori dallo spettro visibile.

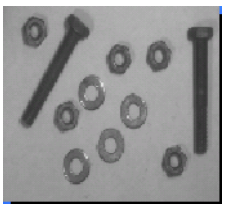


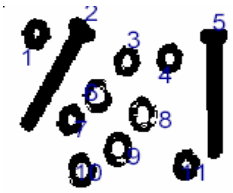
Per contro il sistema visivo umano è:

1. Molto efficiente dal punto di vista computazionale (capacità di elaborare in modo parallelo)
2. Capacità di adattarsi molto rapidamente ad un'ampia gamma di intensità luminose ($I_{percepita}=\ln(I)$)

Possiamo scomporre la visione artificiale (umana) su tre livelli principali:

1. **Basso livello**: atta a percepire e migliorare l'immagine rilevata (filtraggio), estrarre i contorni degli oggetti.
2. **Medio livello**: segmentazione e parametrizzazione delle curve, riconoscimento e classificazione degli oggetti, estrazione informazioni simboliche dalla scena.





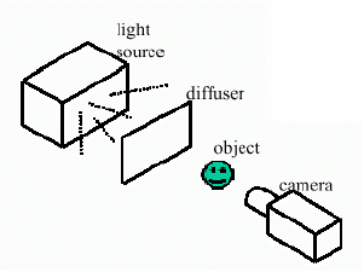
3. **Alto livello**: "Comprensione" della scena osservata, capacità di estrarre informazioni pertinenti da una scena piena di particolari irrilevanti, capacità di apprendere da esempi e di generalizzare, capacità di sopperire alla mancanza di informazioni etc.

Illuminazione della scena

- Durante l'acquisizione dell'immagine, è molto importante illuminare in modo appropriato la scena, questo per garantire un'immagine con elevato contrasto e priva di ombre.

Se siamo in un ambiente interno possiamo utilizzare diverse tecniche di illuminazione:

Ripresa in Controluce

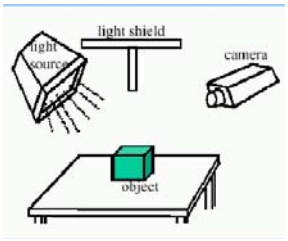


- Questa tecnica di illuminazione produce delle immagini in bianco e nero, quindi permette di rilevare facilmente i contorni degli oggetti.

- Le informazioni sulla superficie dell'oggetto si perdono (sfera).

Luce Frontale

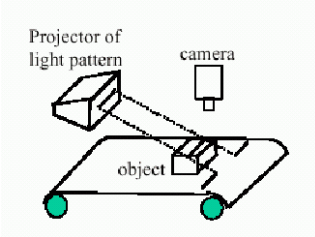
- E' il metodo più utilizzato, permette di rilevare in modo ottimale le caratteristiche delle superfici.



Luce Strutturata

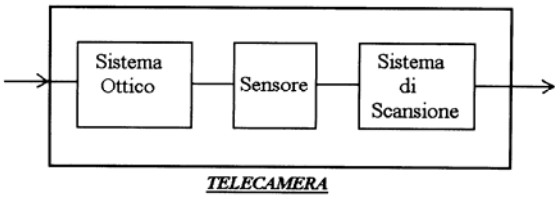
-Consiste nel proiettare punti e lame di luce sull'oggetto di interesse.

- A seconda di come vengono perturbate queste curve è possibile rilevare la presenza dell'oggetto e le sue dimensioni e caratteristiche.



Sensori di visione, telecamere in bianco e nero

Possiamo schematizzare una telecamera come:



Esistono due tipi di telecamere a seconda del tipo di sensore utilizzato:

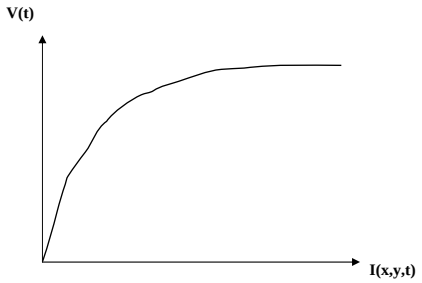
Sensore a Tubo (es. Telecamera Vidicon, Plumbicon) la scansione è continua in orizzontale e discreta in verticale.

Sensore a stato solido:(es. Telecamera CCD) la scansione è discreta sia in orizzontale che verticale.

La telecamera è essenzialmente un trasduttore elettro-ottico che trasforma la luminanza I(x,y,t), (potenza luminosa per unità di superficie, intensità luminosa) incidente sul sistema ottico, in un segnale elettrico analogico v(t).

$$v(t) = k \cdot I(x, y, t)^{\frac{1}{\gamma_c}}$$

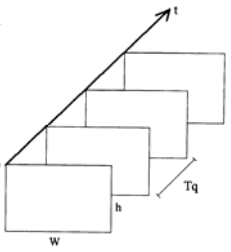
Dove **k** è una costante di proporzionalità mentre γ_c è detto "*Fattore Gamma*"; questo varia circa da 1 a 3 a seconda del tipo di semiconduttore usato per l'elemento fotosensibile



Si verifica una compressione naturale dell'intensità luminosa. I segnali luminosi forti vengono quantizzati in modo più grossolano, in questo modo **migliora il rapporto segnale rumore**.

Campionamento del segnale e baud rate

Supponiamo di usare una rappresentazione a 256 livelli di grigio per pixel (8bit).



Rapporto di aspetto dell'immagine: R=w/h

(es. 4/3 o 16/9). Se abbiamo Nx=768 pixel per riga e Ny=576 pixel per colonna vi saranno:

Q=768*576=442368 [pixel/quadro]

Ogni pixel viene rappresentato con un byte.

1Kbyte=1024byte

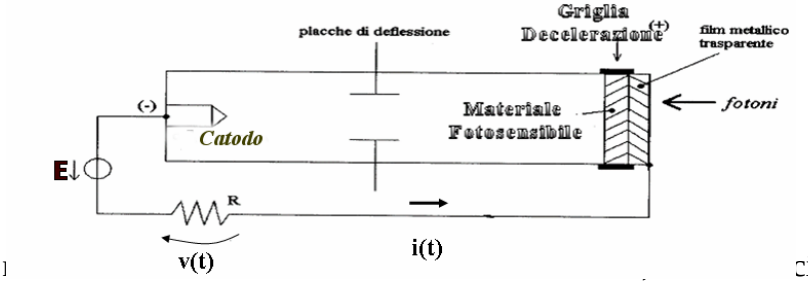
D=432 Kbyte/quadro (memoria per quadro)

Con 25 frame al secondo sono necessari:

ByteRate=25*432= **10.5 Mb/s**

Sensori a Tubo

Tra i sensori a tubo elettronico il più diffuso è il Plumbicon, sviluppato dalla Philips.



Il [illegible] cia interna del dielettrico si accumulano cariche negative (dovute agli elettroni provenienti dal catodo), mentre la faccia esterna è caricata positivamente. Quando dei fotoni colpiscono il dielettrico le cariche negative possono attraversarlo (la sua resistenza diminuisce notevolmente e proporzionalmente alla quantità di fotoni incidenti).

Di conseguenza ottengo un segnale proporzionale all'intensità luminosa. La relazione fra la v(t) e la l'intensità del segnale ottico incidente, è :

$I(x, y, t)$ [W/m^2]

$$v(t) = k \cdot I(x, y, t)^{\frac{1}{\gamma_c}}$$

Si osservi come, **da** un **segnale bidimensionale** spazio-discreto sullo schermo, se ne ottenga uno **mono-dimensionale** tempo-continuo.

Oss: Le placche di deflessione (orizzontali e verticali) servono per deviare il fascio di elettroni sparati dal catodo, mediante un campo elettrico o magnetico a seconda delle realizzazioni; in questo modo è possibile scandire tutta la superficie dell'elemento fotosensibile.

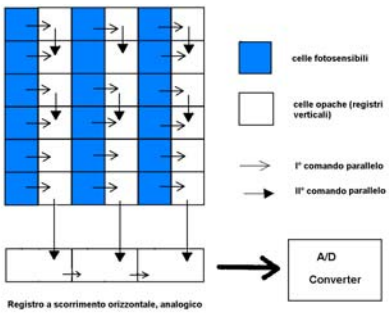
Sensori a stato solido

Lo schermo è costituito da una matrice di elementi fotosensibili (fotodiodi) ad ognuno dei quali possiamo associare una I(x,y,t). I fotoni incidenti su una certa cella generano delle coppie elettrone-lacuna in una quantità che è funzione del numero di fotoni incidenti.

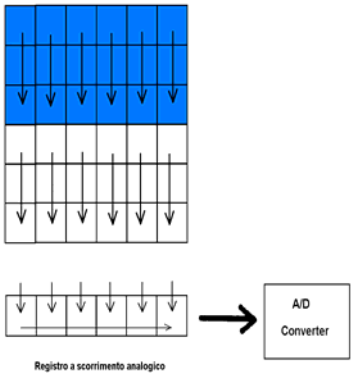
Vi sono varie tecnologie:

Sensori a Trasferimento di Linea

Con un comando parallelo le cariche contenute in ciascun elemento fotosensibile vengono trasferite nella cella adiacente, poi in quella orizzontali ed infine si manda il dato seriale al convertitore A/D.



Sensori a Trasferimento di quadro



Caratterizzati dall'avere metà schermo fotosensibile e metà opaco.

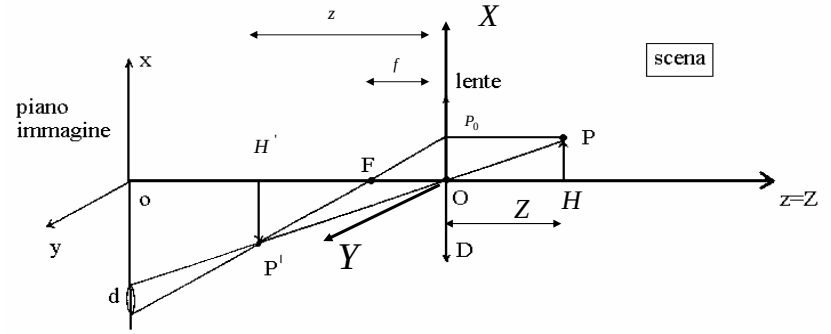
La carica viene velocemente trasferita, riga per riga in parallelo, alla zona opaca; poi con un colpo di clock l'ultima riga in basso viene trasferita in parallelo al registro di scorrimento orizzontale e da lì verso il convertitore A/D.

- Risentono maggiormente del trabocco di carica da una cella a quelle adiacenti “blooming”.

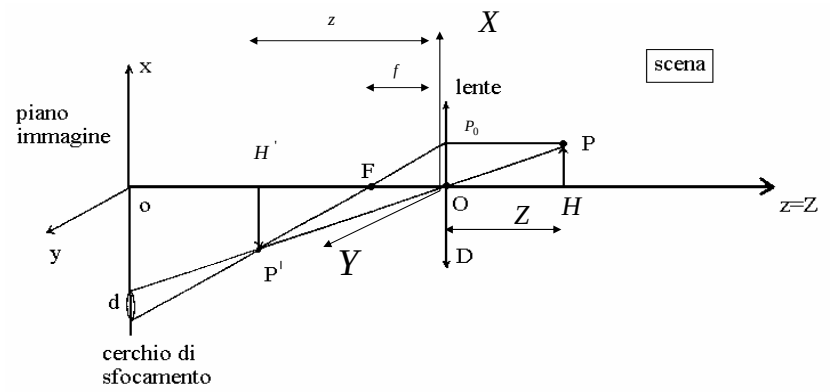
Sensori a Indirizzamento di Pixel :

Con questi sensori ad ogni ciclo si legge e si cancella il valore di ogni cella, in modo simile ad una memoria RAM.

Il sistema ottico della telecamera



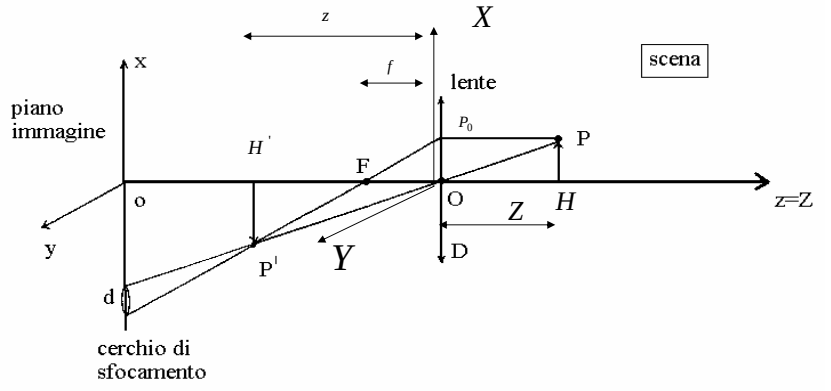
F= fuoco.
FO=**f**= distanza focale.
(x,y,z,o)= sistema di riferimento immagine, con xy=piano immagine.
(X,Y,Z,O)= sistema di riferimento oggetto, con O=centro ottico.
d= diametro del cerchio di sfocamento.
D= apertura della lente.
Z= distanza oggetto
z= distanza immagine



La relazione fra la distanza oggetto e immagine è data dalla **Legge di Fresnel** [freɪ 'nel] :

$$\frac{1}{f}=\frac{1}{z}+\frac{1}{Z}$$

Questa legge si ricava facendo delle similitudini fra i triangoli simili (PHO,P'H'O) e (P_0OF, P'H'F)

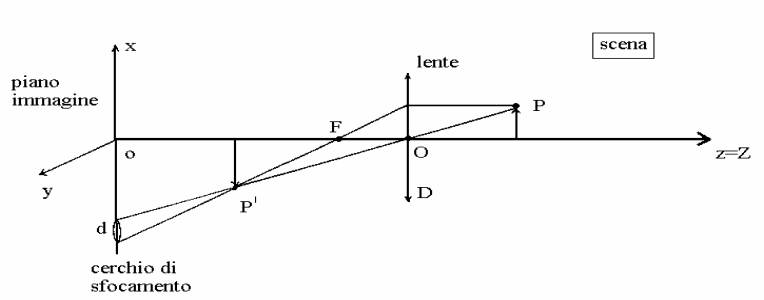


$$Z : z = PH : H'P'$$

$$(z-f) : f = H'P' : PH$$

$$(z-f) : f = z : Z \Rightarrow \frac{z}{f}=\frac{z}{Z}+1 \Rightarrow \frac{1}{f}=\frac{1}{Z}+\frac{1}{z}$$

Si dice che il punto P è "**a fuoco**" se P' giace sul piano immagine,ossia se il diametro del cerchio di sfocamento è zero.

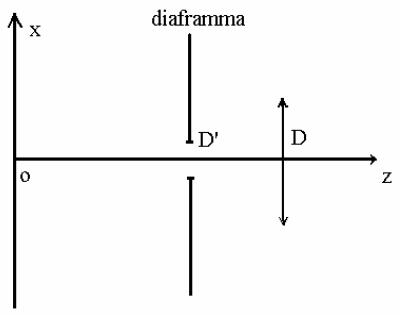


La condizione pratica per cui P sia a fuoco è che:

d<potere risolutore della lente (parametro caratteristico della lente)

Poiché si dimostra che **d** è proporzionale a **D**, possiamo affermare che rimpicciolendo l' apertura della lente, il punto P tende ad essere messo a fuoco.

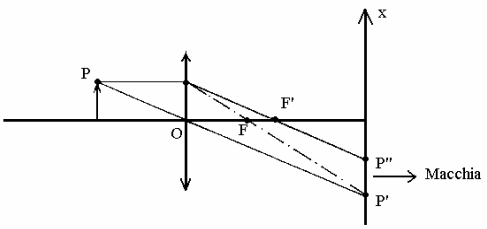
Nella pratica però non è pensabile rimpicciolire la lente o spostare il piano immagine verso P'; quello che si fa è mettere un **<u>diaframma</u>** ad apertura variabile tra lente e piano immagine.

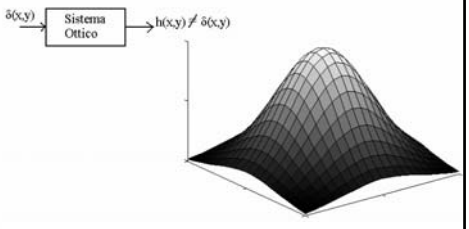


Il principale svantaggio di questo accorgimento è che entra meno luce, ossia la potenza luminosa che incide sullo schermo fotosensibile è piccola.

Modello reale di formazione dell' immagine (immagine a fuoco)

•Nella realtà i raggi paralleli all' asse ottico **non** vengono deviati esattamente nel fuoco, ma in un suo intorno: questo provoca la formazione di una macchia, anziché di un punto, sul piano immagine





•Nel caso reale inoltre ad un impulso luminoso corrispondente alla sorgente puntiforme **P**, non corrisponde un altro impulso luminoso **P'** sul piano immagine, ma viene in realtà filtrato (filtro passa basso) dal sistema ottico.

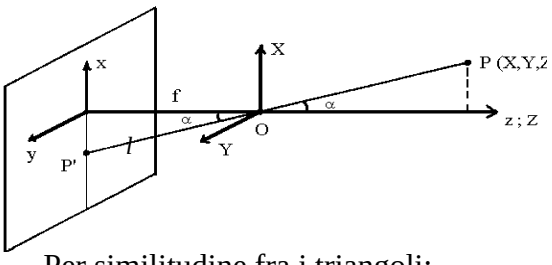
Formule di Prospettiva

Prendiamo ora in considerazione il caso in cui |Z|>>f .

Dalla legge di Fresnel risulta:

$$|z| \cong f \qquad \frac{1}{f} = \frac{1}{z} + \frac{1}{Z}$$

Cerchiamo ora le relazioni tra x e X; e tra y e Y:



Per similitudine fra i triangoli: $f : Z = -x : X$

Quindi otteniamo **le formule di prospettiva**:

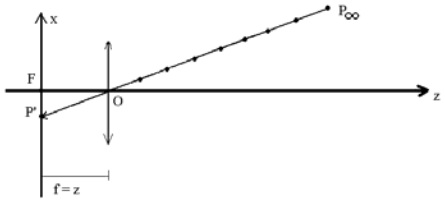
$$x = -f \cdot \frac{X}{Z} \qquad y = -f \cdot \frac{Y}{Z}$$

Def. Distanza focale: quell' intervallo dal centro ottico a cui porre lo schermo per avere a fuoco i punti all' infinito.

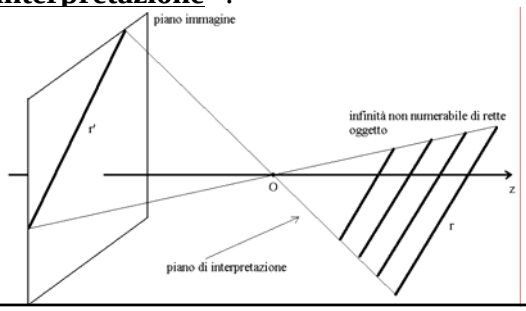
La condizione pratica per avere |Z|>>f è |Z|>=100D , dove D è l'apertura della lente.

OSS:Le formule di prospettiva non sono lineari a causa della Z al denominatore: fra scena ed immagine quindi gli angoli non si conservano e le distanze non sono proporzionali.

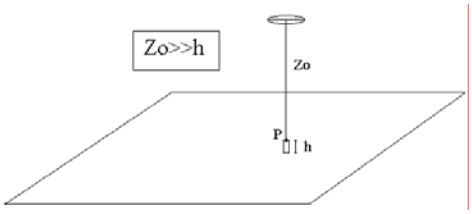
-Non esiste corrispondenza biunivoca tra un punto reale ed il punto immagine :



Dalla figura si vede che ad un punto immagine P' corrisponde una infinità non numerabile di punti oggetto appartenenti alla retta OP, detta "**retta di interpretazione**" .



- Consideriamo invece il seguente caso particolare, molto diffuso nella robotica: $\mathbf{Z_0}$**=cost**. e $\mathbf{Z_0}$**>> h**; cioè il piano della scena è parallelo al piano dello schermo e le profondità degli oggetti tridimensionali sono trascurabili rispetto distanza degli oggetti stessi dalla lente.



Adesso le formule di prospettiva sono lineari, quindi gli angoli si conservano e le lunghezze sono moltiplicate per un fattore costante f/Z_0:

$$z = f \qquad x = -\frac{f}{Z_0} \cdot X \qquad y = -\frac{f}{Z_0} \cdot Y$$

Pre-elaborazione dell'Immagine

•Uno dei problemi fondamentale è quello distinguere nell'immagine oggetti diversi.

•Molte volte oggetti diversi corrispondono a zone dell'immagine con intensità luminosa diverse.

•Nei casi più favorevoli si può utilizzare la tecnica di "**sogliatura**".

SOGLIATURA: In questo caso si suddivide l'immagine in due zone.

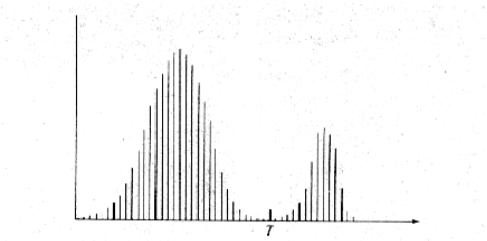
$$\{P_1, \ldots\ldots, P_k\} \quad I(P_i) < T$$
$$\{P_{k+1}, \ldots\ldots, P_n\} \quad I(P_i) \geq T \qquad T = soglia$$

Problemi:

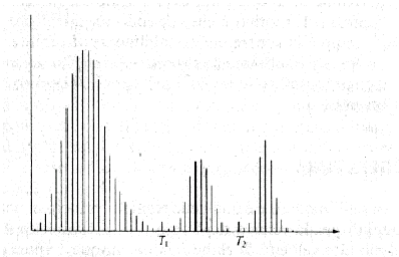
1. Come scegliere la soglia
2. Errore nell'attribuire il pixel ad una delle due regioni (pixel isolati di una regione all'interno dell'altra). Questo porta ad avere regioni non connesse.

1

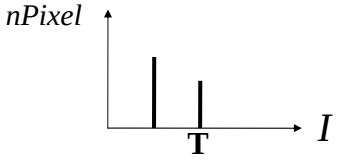
Per scegliere la soglia **T** si utilizza **l'istogramma** dell'immagine (a ciascun valore di intensità luminosa associa il numero di pixel dell'immagine aventi quella particolare intensità)

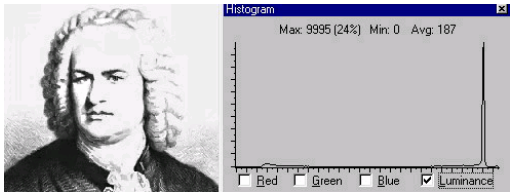


Se sono presenti due oggetti (bianchi) con intensità diverse su sfondo nero:



Nel caso vi siano due zone ben distinte, l'istogramma è composto da due picchi (come soglia si sceglie quella del picco più piccolo).





Immagine

Troppo luminosa

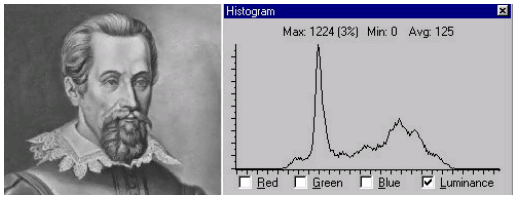


Immagine con una buona

Distribuzione dell'istogramma

2

Per ridurre l'errore nell'attribuzione dei pixel alle zone dell'immagine rilevanti, si possono fare dei filtraggi.

Si può per esempio ridurre il rumore ad elevata frequenza spaziale applicando un **filtro media** o **un filtro mediano**.

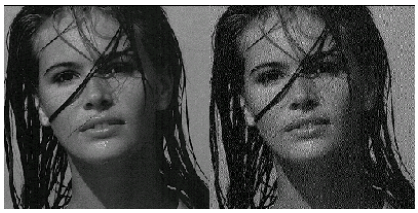
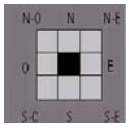


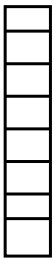
Immagine con rumore

ad alta frequenza spaziale

Ad ogni pixel si assegna un valore di intensità pari ad un valore mediano tra le intensità degli 8 pixels che lo circondano.



Il valore mediano (tra n valori di intensità ordinati in senso crescente, il valore mediano è quello che corrisponde all'intensità del valore n/2) rispetto al valore medio, ha il vantaggio di essere meno sensibile al rumore (se si perturbano anche di molto un pixel dell'insieme mentre il valore medio varia quello mediano no).



Oss: Tale filtro è solitamente applicato all'immagine prima di applicare la sogliatura.

Estrazione dei Contorni e Segmentazione

Per **contorno** di un oggetto si intende:

- Insieme curve visibili oggetto
- Regioni in cui si ha una discontinuità nella profondità o nell'orientamento della normale alla superficie.

Se siamo in condizioni di **luce diffusa e oggetti con colore uniforme,** allora:

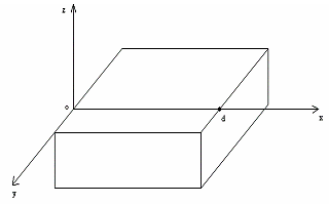
Discontinuità superficie ⇔ Discontinuità intensità luminosa $I(x, y)$ [W/m^2]

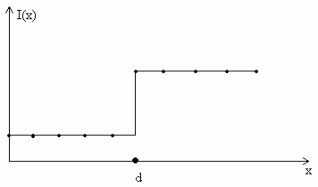
Si rilevano le zone di discontinuità luminosa con lo scopo di rilevare i contorni dell'immagine di un oggetto e quindi ricavarne una sua descrizione geometrica.

1. Rilevazione Contorni: Partendo dall'immagine a livelli di grigio si trovano dei punti candidati ad appartenere al contorno
2. Segmentazione e parametrizzazione: nella segmentazione si raggruppano alcuni punti rilevati precedentemente in archi di curva, poi si stimano i parametri delle equazioni che descrivono le curve (solitamente sono dei segmenti di retta).

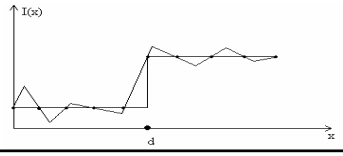
Rilevazione dei contorni nel caso mono-dimensionale:

Si consideri una riga dell'immagine. Nel caso ideale se l'orientamento della superficie non varia anche l'intensità luminosa rimane costante.



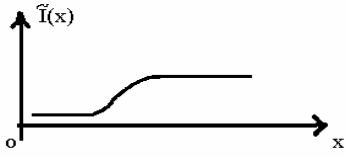


Naturalmente nel caso reale il rumore introduce incertezza nella determinazione della posizione del gradino.



Per eliminare il rumore ad alta frequenza spaziale si può pensare di inserire un **filtro passa basso**.

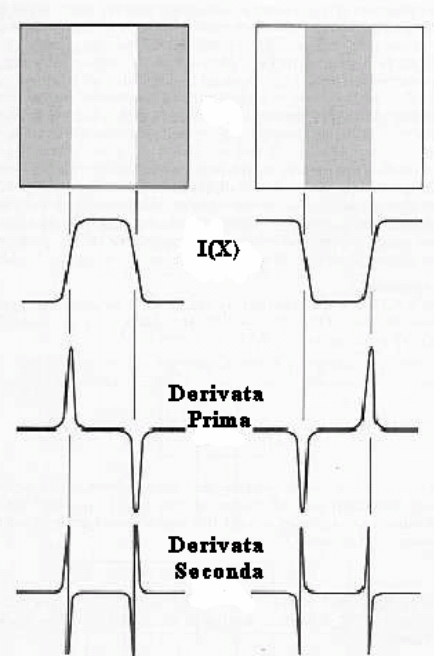
In tal caso il rumore si smorza, ma anche la transizione di I(x) avviene più lentamente (l'immagine perde di contrasto tra zone scure e zone chiare).



Solitamente si utilizza un filtro che ha una risposta all'impulso pari alla derivata di una gaussiana (la derivata si usa perché è massima nei contorni).

convoluzione

$$\frac{d}{dx}(gauss) * segnale = \frac{d}{dx}(gauss * segnale)$$



Es) l'immagine considerata consiste di una striscia verticale chiara su uno sfondo scuro e viceversa.

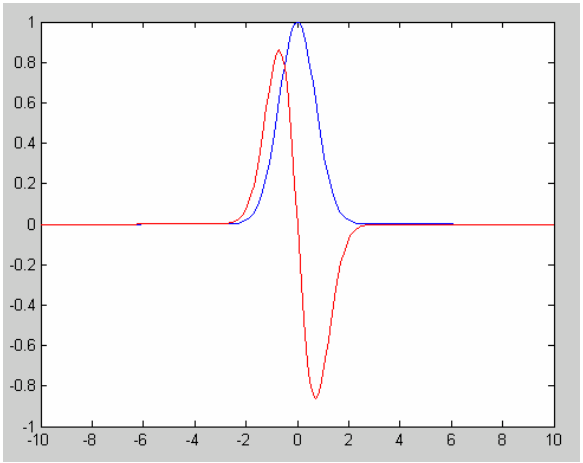
L'andamento dei livelli di grigio lungo una delle linee dell'immagine evidenzia le transizioni corrispondenti ai contorni.

I contorni sono modellati mediante un profilo non a gradino ma a rampa.

Gaussiana e Derivata

$$G(x) = e^{-x^2}$$

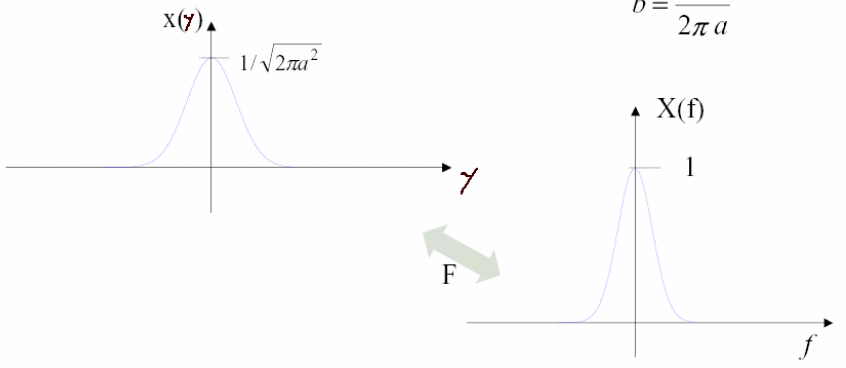
$$\frac{d(G(x))}{dx} = -2xe^{-x^2}$$

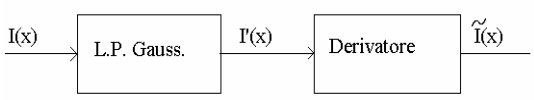


La trasformata di **Fourier** di una Gaussiana è ancora una Gaussiana

$$x(\gamma) = \frac{1}{\sqrt{2\pi a^2}} \cdot \exp\left\{-\frac{\gamma^2}{2a^2}\right\} \quad \Leftrightarrow \quad X(f) = \exp\left\{-\frac{f^2}{2b^2}\right\}$$

$$b = \frac{1}{2\pi a}$$



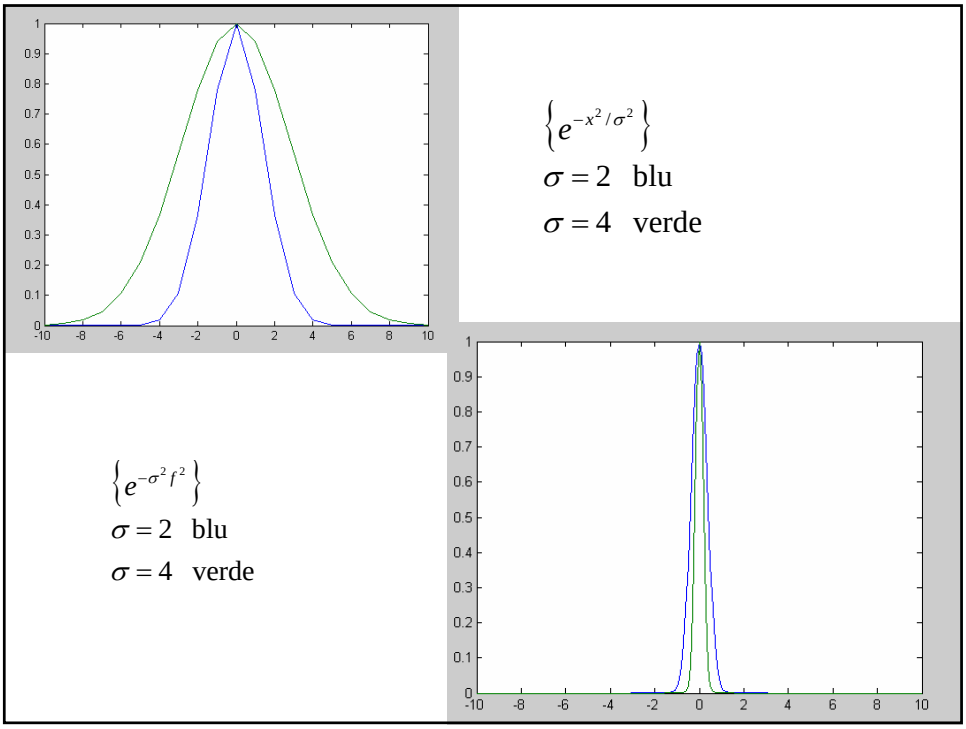


$$h(x) = \int_{-\infty}^{+\infty} I(\tau) g(x-\tau) d\tau = I(x) * g(x) \quad \Rightarrow \quad h(k) = \sum_{n=-\infty}^{+\infty} I(n) g(k-n)$$

Nel dominio delle frequenze si avrà:

$$\boldsymbol{H(f) = F\{g(x) * I(x)\} = F\left\{e^{-x^2/\sigma^2}\right\} \cdot S(f) = \sigma\sqrt{\pi} \cdot e^{-\pi^2\sigma^2 f^2} \cdot S(f)}$$

Si noti che la trasformata di una gaussiana è ancora una gaussiana, ma con ampiezza σ sigma pari al reciproco dell' ampiezza della funzione di partenza g(x).



Fondamenti di Visione Artificiale (Seconda Parte)

PhD. Ing. Folgheraiter Michele

Corso di Robotica
Prof.ssa Giuseppina Gini
Anno Acc. 2006/2007