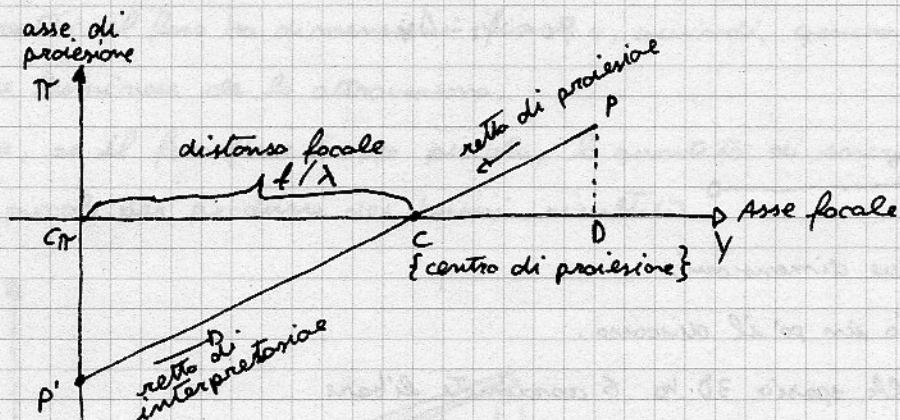


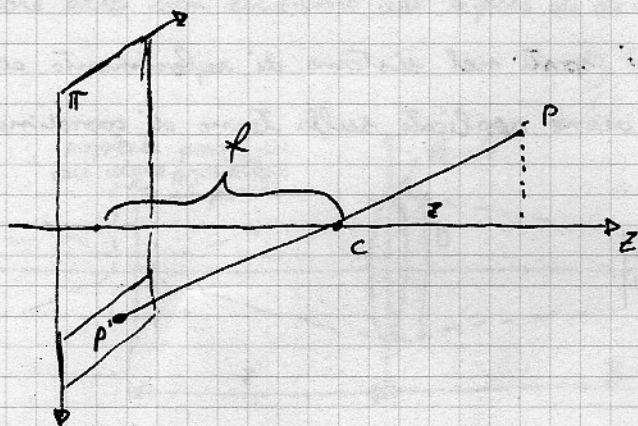
# Proiezione pin-hole

Nella proiezione pin-hole, si suppone che l'immagine venga catturata attraverso un foro dalle dimensioni minuscole.



La retta che congiunge  $P$  e  $P'$  prende due nomi differenti in base al fatto che sia realizzato partendo da  $P$  e  $C$  oppure  $P'$  e  $C$ . Nel primo caso prende il nome di "retto di proiezione", mentre nel secondo caso di "retto di interpretazione".

Uno dei più grandi problemi è quello di riuscire a determinare la posizione originaria del punto  $P$  sullo retto di interpretazione.



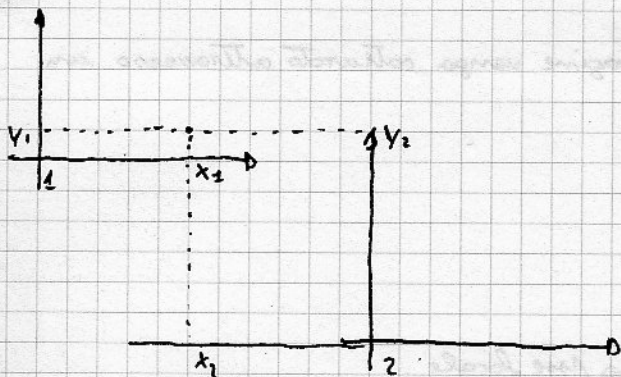
Questo sistema è legato dalle seguenti relazioni:

$$\begin{cases} x = f \cdot \frac{X_c}{Z_c} \\ y = f \cdot \frac{Y_c}{Z_c} \end{cases} \rightarrow \text{questi valori possono essere visti come un fattore di scala che dipendono dallo } z \text{ del punto scena}$$

Se si impone  $m = f/z_p$ , allora si ottiene la weak perspective scaled orthographic

$$\begin{cases} x = -m X \\ y = -m Y \end{cases}$$

Serve, a questo punto, fornire un legame tra il sistema di riferimento del mondo ed il sistema di riferimento della camera.



$$x_1 = x_2 - \Delta x$$

$$y_1 = y_2 - \Delta y$$

Questo per un discorso in due dimensioni.

In tre dimensioni cambia un po' il discorso.

Qualunque corpo rigido nello spazio 3D ha 6 coordinate libere

Per definire la posizione di un corpo rigido nello spazio ho bisogno di almeno 3 punti; ognuno con 3 coordinate. Siccome i punti sono fissi tra di loro (perché il corpo è rigido), allora si possono semplificare le coordinate passando da 9 a 6.

$$P \begin{vmatrix} X \\ Y \\ Z \end{vmatrix} \quad \begin{vmatrix} W_x \\ W_y \\ W_z \\ W \end{vmatrix} \quad \begin{vmatrix} X^o \\ Y^o \\ Z^o \\ W^o \end{vmatrix} \quad \begin{matrix} X^o & Y^o & Z^o \\ W^o & W^o & W^o \end{matrix}$$

Per potere definire come un punto scena si sposti nel sistema di riferimento camera è necessario usare le roto-traslazioni, che viene applicate sulla Terzo di coordinate mediante una matrice di roto-traslazione.

$$P_{2D} = M \cdot P_{3D}$$

$$P_{3D} \begin{vmatrix} X \\ Y \\ Z \\ K \end{vmatrix} \quad P_{2D} \begin{vmatrix} W_x \\ W_y \\ W \end{vmatrix}$$

$$\begin{vmatrix} W_x \\ W_y \\ W \end{vmatrix} = \begin{vmatrix} \text{---} \\ \text{---} \\ \text{---} \end{vmatrix} \cdot \begin{vmatrix} X \\ Y \\ Z \\ K \end{vmatrix}$$

↓

ci dovrebbe essere un'altro coordinato, ma che a noi non interessa perché il punto immagine è in  $z=0$ . Le schede grafiche computano ugualmente questa variabile perché contiene altre informazioni

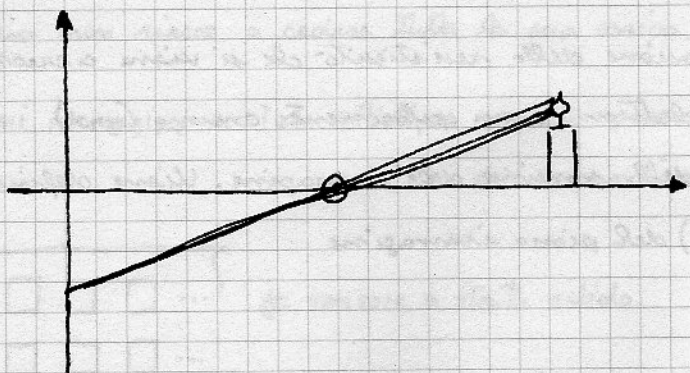


## Richiami di ottica e sensori

Per potere riprodurre il modello pin-hole avremmo bisogno di un foro di dimensioni infinitesime attraverso cui fare "passare" la nostra scena!

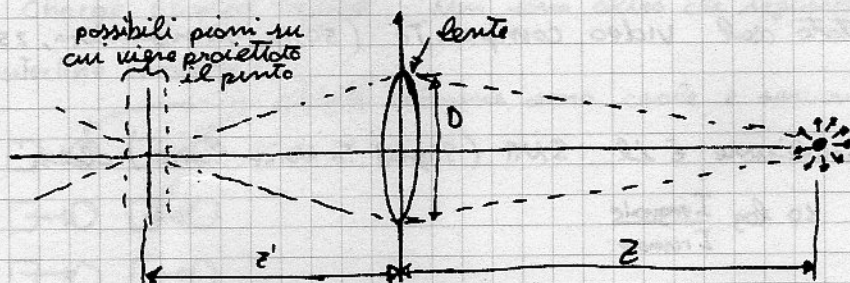
Nella realtà il foro ha dimensioni finite e, quindi, genera delle interazioni con le onde luminose che lo attraversano.

Inoltre, se il foro fosse troppo piccolo, la quantità di energia che ci passerebbe sarebbe troppo piccola per produrre dei buoni risultati.



Se il foro non è infinitesimo, allora l'immagine non è a fuoco

Per ovviare alla non esistenza del foro di dimensioni infinitesime, si utilizza l'attica!



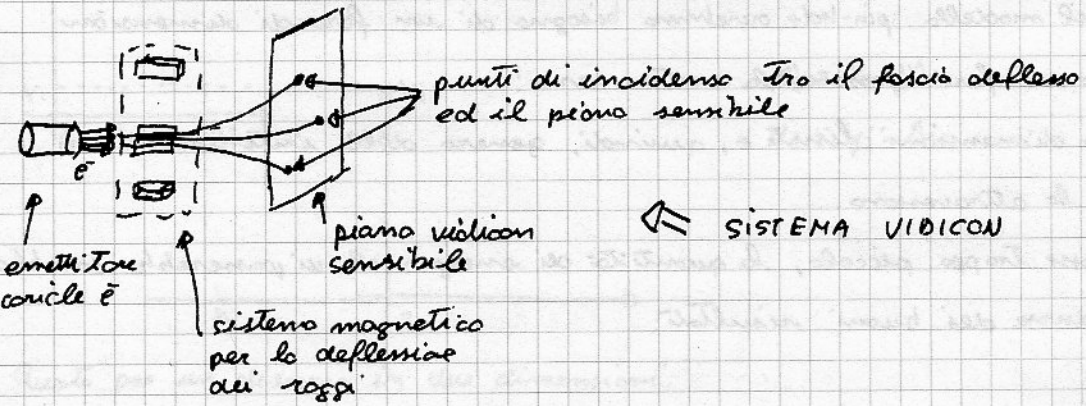
$$\frac{1}{z'} - \frac{1}{z} = \frac{1}{f}$$

Se il punto scena è a distanza  $z=0$ , allora  $z'=f$

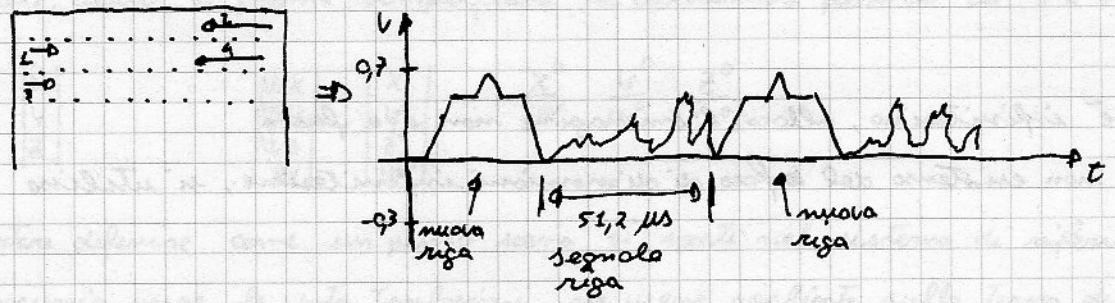
$$\phi_{BC} = \frac{D}{z \cdot z'} = \frac{D}{z'} \cdot |z' - z'|$$

dimensione del Blur Circle

# Sensori ed acquisizioni immagine



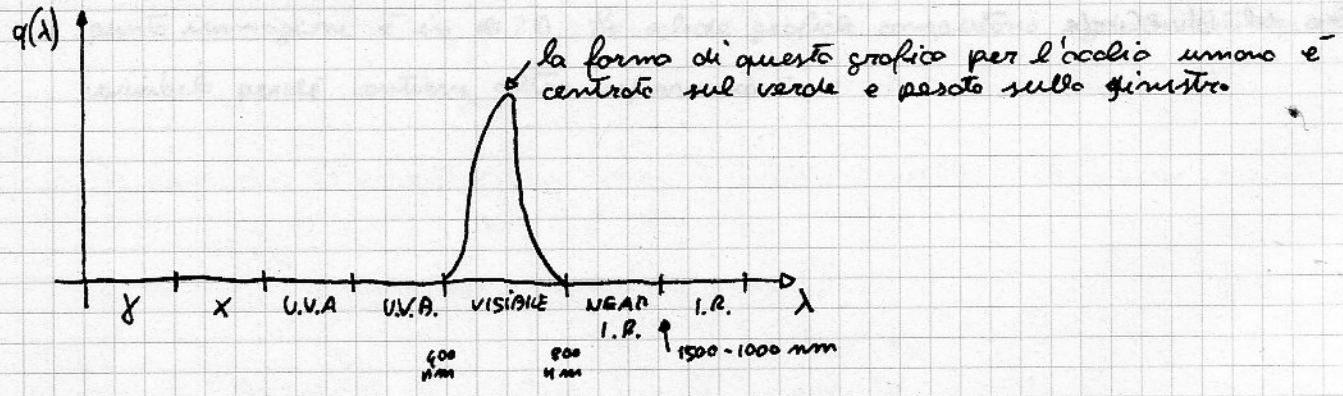
Il piano fotosensibile si basa sulla variazione della resistenza che si viene a creare tra un punto eccitato dal pannello di elettroni e un riferimento comune (grid). La deflessione provoca una "serializzazione" della scansione dell'immagine. Viene definito come scansione progressiva (progressive scan) del piano immagine.



Questa segnale è quello adottato dal video composito (50Hz semiquadro, 25Hz a pieno quadro).

Un parametro importante da analizzare è il SNR (Signal To Noise Ratio) che si misura in dB ed equivale a  $10 \log \frac{I_{\text{segnale}}}{I_{\text{rumore}}}$

Per definire quanto è "buono" un sensore è necessario analizzare il diagramma della efficienza quantica. L'efficienza quantica permette di definire la quantità di energia che il sensore riesce a rilevare in relazione alle varie lunghezze d'onda.



I sistemi vidicon hanno una resa cromatica qualitativamente molto simile a quella dell'occhio umano.

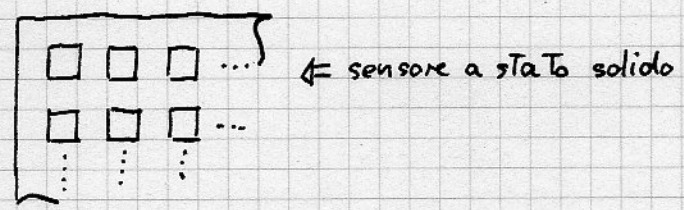


Rapporto ingresso-uscita per sistemi VIDICON:



Nei sistemi VIDICON, quando in un punto si accumula una grande quantità di energia si verifica un fenomeno di persistenza. Questo è dovuto al fatto che la sostanza reattiva non riesce a cedere tutta la sua carica precedente. In questo caso si verifica

Sistemi immagine a stato solido  $\Rightarrow$  sono in grado di fornire informazioni pittoriche su di un insieme non mobile di elementi.

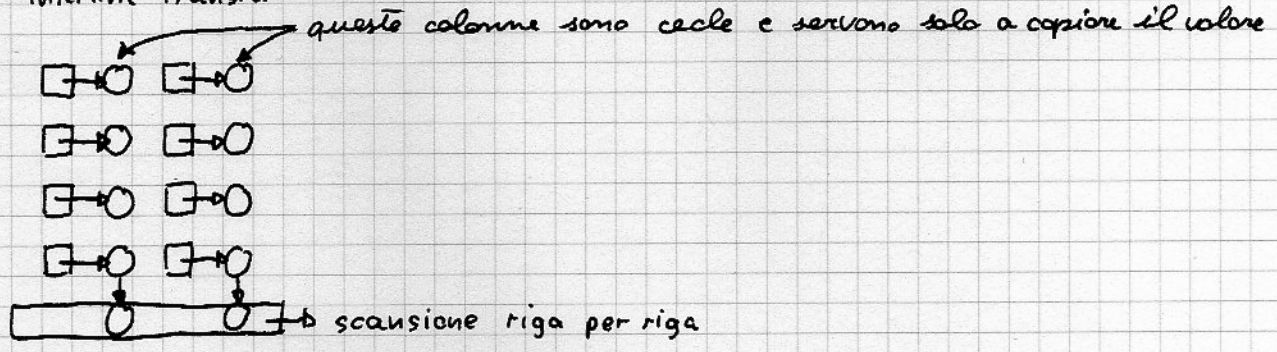


l'insieme dei punti di campione (pixel) è fisso e scolpito nel silicio

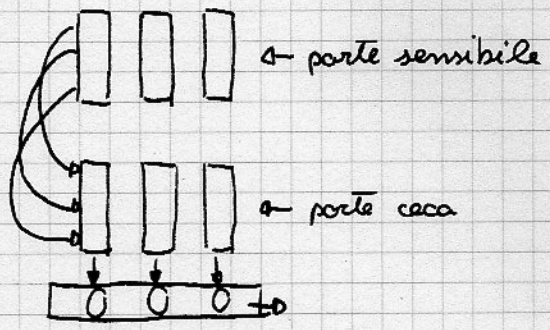
Tra i più famosi sensori a stato solido ci sono i CCD

CCD  $\Rightarrow$  Charge Coupled Device. Non sono altro che registri a scorrimento, analogici

$\hookrightarrow$  Interline Transfer



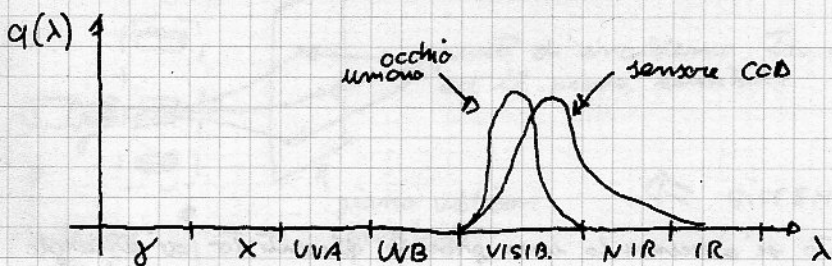
$\hookrightarrow$  Frame Transfer



Nei CCD-FrameTransfer, a differenza di quelli Interline Transfer, gli elementi sensibili vengono copiati contemporaneamente nella memoria e, poi, svuotati riga per riga. Le righe possono andare a formare un segnale video composto.

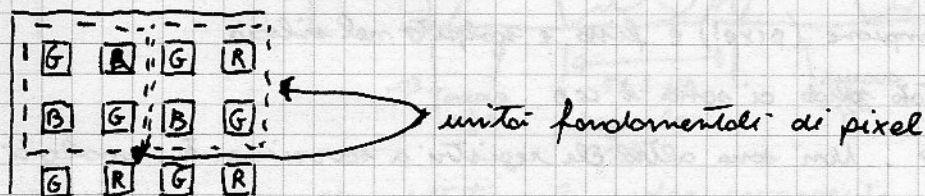
I sensori sono rettangolari, con un rapporto di  $3/4$ . Le misure commerciali usate dei sensori sono  $1''$ ,  $2/3''$ ,  $1/2''$ ,  $1/4''$ . Varia lo spazio tra le colonne.

L'efficienza quantica di un sensore CCD è un po' differente da quello a sistema VIDICON, in quanto include anche una parte di spettro NIR.



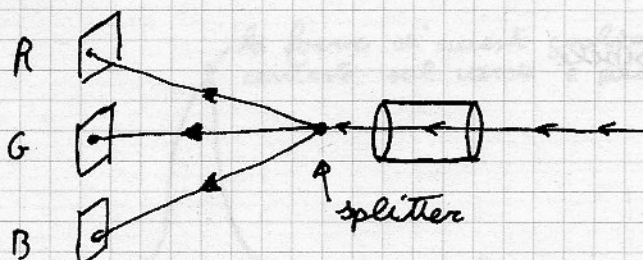
Come si può notare, l'efficienza quantica dei sensori CCD è orientato più verso il colore rosso (a differenza dell'occhio umano e dei sistemi VIDICON che hanno una centatura più orientato sul verde). Inoltre, si può notare anche una sensibilità nello spettro I.R. che potrebbe "disturbare" l'immagine. Questa sensibilità allo spettro I.R. è utile in determinate applicazioni, come visione notturna. In utilizzi normali si può supporre che la componente I.R. di una scena sia poca e, quindi, trascurabile come influenza.

Vi sono due modi principali di formare i colori con sistemi CCD:



Un "pixel immagine" è formato da 4 pixel sensori; ogni pixel ha un filtro che gli permette di ricevere solo una data frequenza di luce. È il sistema che normalmente viene usato nelle periferiche. Il numero reale di pixel (uguale a quelli fisici del sensore) si ottiene mediante una interpolazione da con i pixel contigui.

Il sistema Foveon si basa sulla divisione del fascio di luce verso tre differenti sensori CCD

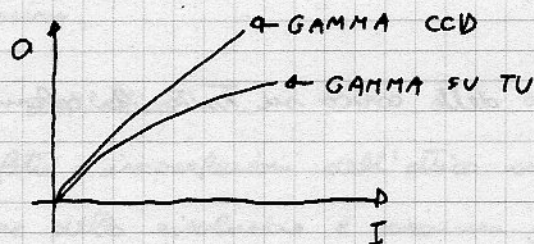
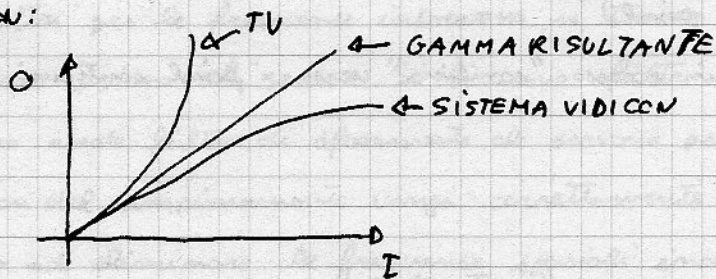


Il problema è che tale divisione del raggio di luce deve essere assolutamente calibrato, in modo che lo stesso punto scena cada nello stesso punto in tutti e 3 i CCD!



La gamma del sensore CCD è leggermente differente da quella dei sistemi.

VIDICON:

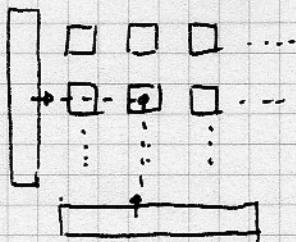


I dispositivi CCD, di solito, integrano una modalità per la correzione della gamma che permette di operare come su un sistema VIDICON, eliminando la distorsione.

### Dispositivi CMOS

Nei sensori CMOS è possibile avere un read-out ad indirizzamento di pixel e non a riga o frame.

In pratica, si accede ad ogni singolo pixel specificando riga e colonna di appartenenza.



VGA 640x480 (768x576)

CAMERALINK (interfaccia digitale tra camera e PC) non compresso

ETHERNET / IP

GIGABIT-ETHERNET / IP

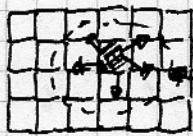
} distanze unipire / video sorveglianza

FIREWIRE 1/2 digitale, compresso o non

USB

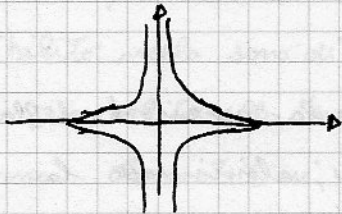
Problema del blooming nasce quando una grande quantità di energia viene diretta verso un pixel (o un piccolo insieme di pixel)

l'effetto che si ottiene è che l'energia in eccesso "sconfina" verso i pixel contigui falsando la lettura.



i pixel circostanti vengono "contaminati"

In sistemi CCD può verificarsi una propagazione delle cariche su tutta la colonna rendendola cieca



Un modo per ovviare a questo problema è di rendere la risposta del CCD non lineare e, quindi, fare in modo che reagisca poco a grandi quantità di energia e reagisca molto a basse quantità



## filtri

3 filtri per le frequenze inferiori si trovano generalmente sull'ottica, ma in ogni caso prima al sensore. Si chiamano filtri **NIACUT-OFF**

Esistono anche filtri di sfocamento che servono per garantire che il Teorema di Shannon del campionamento venga correttamente rispettato. Se sfoca opportunamente riesce ad eliminare le frequenze spaziali indesiderate e, quindi, evitare effetti di aliasing.

## Problemi dell'ottica

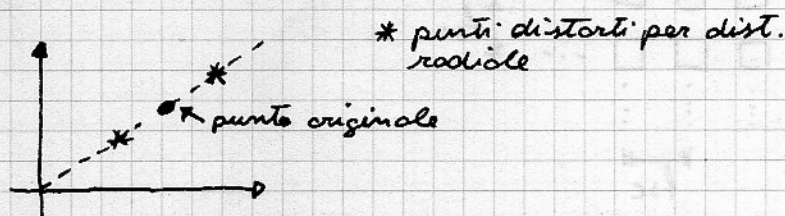
I difetti o imperfezioni dell'ottica saranno maggiormente percepibili con il crescere della risoluzione e saranno accentuati sul bordo della lente

Il campo di vista è l'angolo solido di spazio che l'ottica vede. Sarebbe auspicabile che il sensore sia inscritto nel campo visivo e che sia più piccolo, per evitare le distorsioni radiali

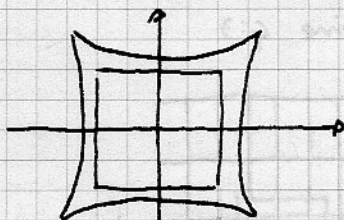
aberrazioni  $\Rightarrow$  distorsioni generali

aberrazioni cromatiche  $\Rightarrow$  distorsioni dei foci di luce differenti a diverso lunghezza d'onda.  
Diverse lunghezze d'onda vengono messe a fuoco in punti differenti.

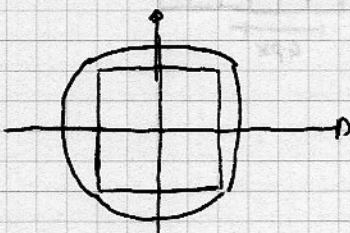
distorsioni geometriche  $\Rightarrow$  non vengono mantenute le posizioni dei punti sceno sui punti immagine.



Distorsione a barilotto



Distorsione a cuscinetto



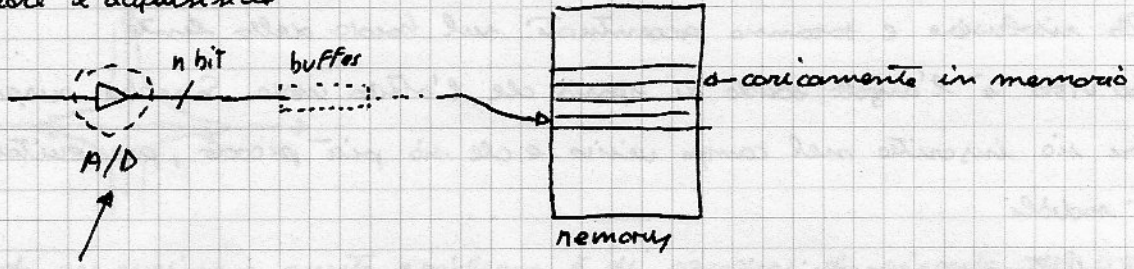
Questa distorsione è tanto più forte quanto più è corto lo distanza focale.

distorsione Tangenziale  $\Rightarrow$  si manifesta perpendicolarmente allo stesso distorsione radiale  
 ma è di scorsa intensità e quindi non si considera

vignettatura  $\Rightarrow$  si presenta nelle lenti composte. Lungo i vari passaggi tra materiali  
 porte dei raggi si perdono nelle superfici esterne dell'ottica. Ad esempio,  
 in un muro dipinto uniformemente ed illuminato uniformemente  
 si riscontrerà una diminuzione della luminosità nei bordi della foto.

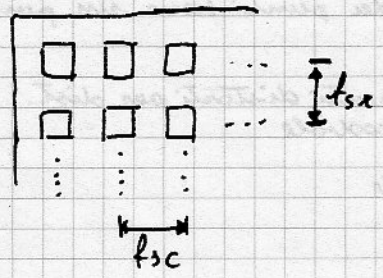
### Acquisizione

Se si parla di video composto, il segnale entra in un frame-grabber che permette di  
 controllare l'acquisizione

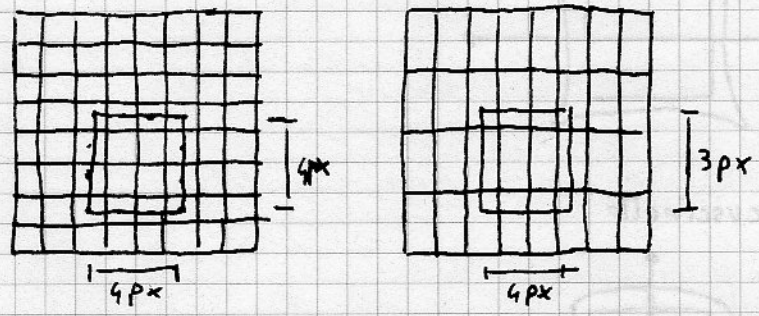


Componente di qualità  
 Questo componente può variare l'intervallo elettrico di quantizzazione del segnale

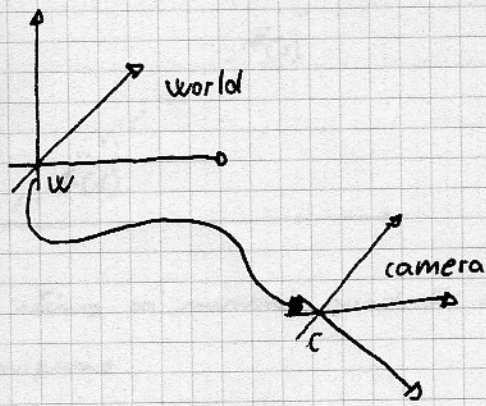
### Frequenza spaziale di campionamento



C'è differenza tra "aspect-ratio" e "pixel-aspect-ratio". Per esempio, il pixel può  
 essere quadrato, ma l'immagine 4:3







$$\begin{bmatrix} \text{Proiezione ideale (f)} \\ 3 \times 4 \end{bmatrix} \cdot \begin{bmatrix} \text{RT} \\ 3 \times 3 \\ \text{---} \\ \text{---} \\ \text{---} \\ \text{---} \\ \text{---} \\ \text{---} \\ 4 \times 4 \end{bmatrix} \cdot \begin{bmatrix} \text{---} \\ \text{---} \\ \text{---} \\ \text{---} \\ \text{---} \\ \text{---} \\ \text{---} \\ \text{---} \\ 4 \times 1 \end{bmatrix}$$

$D_x, D_y, D_z, \theta, \varphi, \psi$

La composizione delle matrici deve portare da uno spazio in 3D ad uno spazio in 2D.

$$\begin{bmatrix} 3 \times 4 \end{bmatrix} \cdot \begin{bmatrix} 4 \times 4 \end{bmatrix} = \begin{bmatrix} 3 \times 4 \end{bmatrix} \cdot \begin{bmatrix} 4 \times 1 \end{bmatrix} = \begin{bmatrix} 3 \times 1 \end{bmatrix}$$

Siccome il centro dell'immagine non è detto che si trovi nel centro effettivo del sensore, allora  $f$  dipenderà da altre cose e  $D_x, D_y, D_z$  dovranno anche considerare la modifica del piano immagine.

Si aggiunge una matrice di ROTOTRASLAZIONE IMMAGINE che si riferisce al centro immagine. La matrice è definita in coordinate omogenee, quindi non ha  $3 \times 4 = 12$  parametri liberi, ma solamente 11

- 6 parametri esterni
- 1  $f$
- 2 Traslazione centro immagine
- 1 differente unità di misura sugli assi (pixel-aspect-ratio)
- 1 SKEW angle (NON RILEVANTE). Angolo tra gli assi del piano.

Bisogna ~~mettere~~ aggiungere nel modello una trasformazione di scale tra la dimensione dei pixel verticali ed orizzontali. La matrice è  $3 \times 3$ . Sulla diagonale c'è aspect ratio e la scala immagine!

di rotazione  $R(3 \times 3)$ , matrice di traslazione  $L(3 \times 1)$

di rototraslazione  $RT(4 \times 4)$

lunghezza focale

$\Rightarrow$  dimensione pixel

$\Rightarrow$  traslazione centro immagine

$(\theta)$   $\Rightarrow$  angolo tra gli assi

$T_z$

li angoli, si trovano le colonne  $\pi_1, \pi_2, \pi_3$

$= M \cdot P_{3D}$

2:

condizione necessaria e sufficiente perché una matrice  $M = \left[ \begin{array}{c|c} a_1 & a_2 & a_3 \\ \hline A & b \end{array} \right]$  sia di proiezione è che  $\det(A) \neq 0$

condizione necessaria e sufficiente perché sia  $\phi$ -skew deve essere che:

$$a_3 = x; \quad a_2 \times a_3 = y; \quad x \cdot y = 0$$

condizione necessaria e sufficiente perché  $M$  sia una matrice di proiezione e

$\phi$ -skew, con aspect-ratio unitario deve essere anche:

$$x a_3 \cdot (a_1 \times a_3) = (a_2 \times a_3) \cdot (a_1 \times a_3)$$

$$m_{11} X + m_{12} Y + m_{13} Z + m_{14}$$

$$m_{21} X + m_{22} Y + m_{23} Z + m_{24}$$

$$m_{31} X + m_{32} Y + m_{33} Z + m_{34}$$

	Noto	Ignoto
si	$X, Y, Z, m_{ij}$	$u, v$
si	$u, v, m_{ij}$	$X, Y, Z$
calibrazione	$X, Y, Z, u, v$	$m_{ij}$

Calibrazione OFF-LINE  $\Rightarrow$  raccolgo i dati di calibrazione e li uso in riprese successive

Calibrazione CONTINUA  $\Rightarrow$  c'è un Thread che computa le variazioni di calibrazione della scena



Matrice di rotazione  $R(3 \times 3)$ , matrice di traslazione  $T(3 \times 1)$

Matrice di rototraslazione  $RT(4 \times 4)$

$f \Rightarrow$  lunghezza focale

$s_x, s_y \Rightarrow$  dimensione pixel

$v_0, u_0 \Rightarrow$  traslazione centro immagine

skew ( $\theta$ )  $\Rightarrow$  angolo tra gli assi

$T_x, T_y, T_z$

per gli angoli, si trovano le colonne  $r_1, r_2, r_3$

$$P_{2D} = M \cdot P_{3D}$$

Teorema:

- Condizione necessaria e sufficiente perché una matrice  $M = \left[ \begin{array}{c|c} (a_1 \ a_2 \ a_3) \\ \hline A \end{array} \right] \left| \begin{array}{c} b \\ \hline \end{array} \right]$  sia di proiezione è che  $\det(A) \neq 0$
- Condizione necessaria e sufficiente perché sia  $\phi$ -skew deve essere che:  
 $a_1 \times a_3 = X$  ;  $a_2 \times a_3 = Y$  ;  $X \cdot Y = 0$
- Condizione necessaria e sufficiente perché  $M$  sia una matrice di proiezione e  $\phi$ -SKEW, con aspect-ratio unitario deve essere anche:  
 $(a_1 \times a_3) \cdot (a_1 \times a_3) = (a_2 \times a_3) \cdot (a_2 \times a_3)$

$$w_u = m_{11}X + m_{12}Y + m_{13}Z + m_{14}$$

$$w_v = m_{21}X + m_{22}Y + m_{23}Z + m_{24}$$

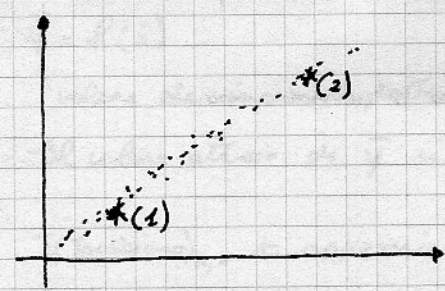
$$w = m_{31}X + m_{32}Y + m_{33}Z + m_{34}$$

	Nota	Ignoto
analisi	$X, Y, Z, m_{ij}$	$u, v$
analisi	$u, v, m_{ij}$	$X, Y, Z$
calibrazione	$X, Y, Z, u, v$	$m_{ij}$

CALIBRAZIONE OFF-LINE  $\Rightarrow$  raccoglie i dati di calibrazione e li usa in riprese successive

CALIBRAZIONE CONTINUA  $\Rightarrow$  c'è un Thread che computa le variazioni di calibrazione della scena

SENZA CALIBRAZIONE  $\Rightarrow$  caso che non trattiamo



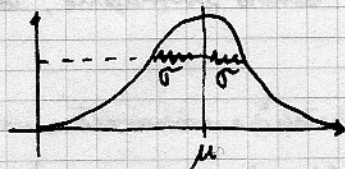
$$\begin{cases} y = mx + q \\ x = my + q \end{cases} \leftarrow \text{per due punti}$$

La stima ai minimi quadrati serve per trovare  $m$  e  $q$  quando ho tanti punti campione

$x$  indipendente ;  $y$  dipendente

$\sigma$  (scarto quadratico medio)

$N(\mu, \sigma)$   
 $\uparrow$  medio  
 $\uparrow$  scarto  
 (varianza)



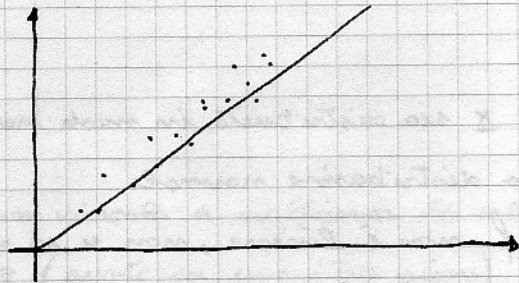
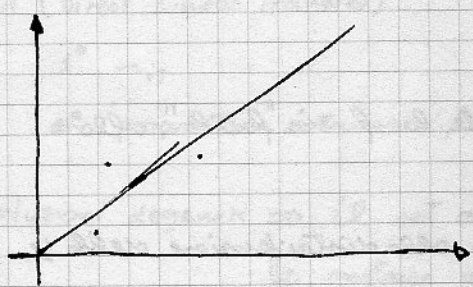
L'incertezza è data da:  $\sigma/\sqrt{n}$

Effettuando diverse (numerose) misure rumorose (che presentano un rumore) posso approssimare meglio il valore reale.

Se il rumore ha caratteristiche che rimangono uguali, si chiama RUMORE STAZIONARIO

Cifra di merito  $J = \sum E^2$  (sommatore del quadrato degli errori)

$\min(m, q) J = \sum E^2$

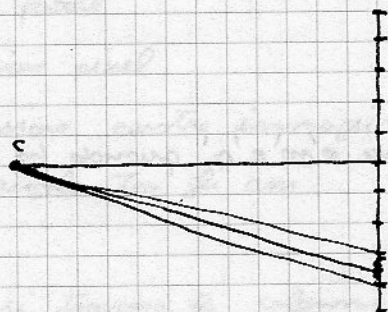


Con pochi punti campione ho un forte rumore, mentre con molti punti campione il rumore diminuisce notevolmente. È importante, quindi, utilizzare un grande numero di campioni quando si devono fare delle misure (come ad esempio la calibrazione)



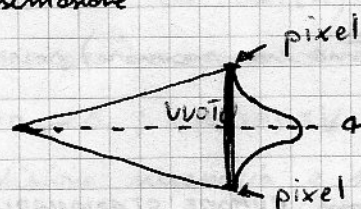
# Il problema dell'incertezza

È necessario un modello esplicito sull'incertezza delle misure che andiamo a compiere.



Se l'incertezza riguarda la posizione del centro immagine, la posizione del centro di proiezione e del pixel.

Ogni volta che facciamo una rete di interpretazioni commettiamo un errore, quindi, la rete di interpretazione che andiamo ad utilizzare non sarà quella reale, ma una sua approssimazione.



l'incertezza ha un andamento gaussiano (normale)

Con due telecamere, la dipendenza dell'incertezza dalla distanza dal punto è quadratica!

## Propagazione dell'incertezza

$$y = f(x)$$

Supponendo che  $x$  sia distribuito in modo normale e che la  $f$  sia "facile", allora  $y$  ha sempre una distribuzione normale.

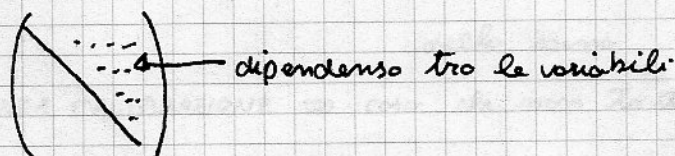
Se la funzione  $f$  non è lineare, non si sa cosa succede alla distribuzione della  $y$ .

$$\underline{x} = N(\underline{\mu}, \underline{\Lambda})$$

matrice di co-varianza

Per la varianza serve una matrice in quanto bisogna considerare lo scarto tra le variabili che non sono indipendenti.

$$\begin{pmatrix} x \\ y \end{pmatrix} \begin{pmatrix} \sigma_{xx} & \sigma_{xy} \\ \sigma_{yx} & \sigma_{yy} \end{pmatrix}$$



La matrice di covarianza è simmetrica rispetto alla diagonale principale.

$$\bar{y} = f(\bar{x})$$

↑ valore che io mi aspetto per la dipendente

Il valore atteso di  $\bar{y}$  si calcola mediante  $f$

$J(\text{Jacobiano})_{n,x}$  è approssimazione. Come vorrà la  $f$  vicino a  $\bar{x}$

$$J_{n,x} \Big|_{x=\bar{x}} \cdot \Delta x = \Delta y \quad \hat{y} = N(\bar{y}, \Sigma_y)$$

$$\underline{m} = \begin{pmatrix} m_{11} \\ m_{21} \\ \dots \\ m_{33} \end{pmatrix} \quad \Sigma_m \quad \text{indipendenti} \Rightarrow \begin{pmatrix} u \\ v \\ x \\ y \\ z \end{pmatrix}$$

Chiamando  $A$  la matrice di osservazione (quello che contiene i dati raccolti), si può decomporre  $A$  usando SVD (Single Value Decomposition) ottenendo 3 matrici si arriva ad uno stima degli  $m$  ai minimi quadrati.

$$A m = 0$$

Devo potere dire quanto è precisa la mia misura a partire non dai dati raccolti (orizzonte campionaria), ma in relazione agli strumenti.

Stimare gli elementi di proiezione

Procedimento che determino gli  $m_{ij}$  indipendenti dal modello di proiezione; si chiama DLT (Direct Linear Transfer)

$$1^\circ m_{ij}$$

2° parametri intrinseci/estrinseci

soluzioni degeneri  $\Rightarrow$  il set dei punti non vincola a sufficienza la soluzione ( $\text{Rango}(A)$ )

↳ matrice di rango = 8 (punti su uno stesso piano)

↳ un insieme di punti su una retta e gli altri su di una Twisted-cubic (caso molto raro)

ISA  $\Rightarrow$  è possibile computare la matrice di proiezione congiuntamente ad una stima dello distorsione radiale.



1) Si spieghi perché una camera reale necessita di ottica

Il modello a pin-hole prevede una dimensione infinitesima attraverso cui le rette di proiezione passano. Siccome l'apertura di tale foro è infinitesima, ad ogni punto scena corrisponde un distinto punto immagine, determinato dallo intersezione del punto scena con il centro di proiezione (pin-hole)

L'energia che può passare attraverso un foro di dimensione infinitesima, è infinitesimo a sua volta; quindi, per potere percepire una quantità di energia infinitesima, sarebbe necessario un sensore con sensibilità infinita, oppure che il punto scena emana una quantità di energia infinita, entrambe cose non realizzabili.

Usando l'ottica si può ottenere una soddisfacente modellizzazione del modello a pin-hole, ma facendo in modo che l'apertura non sia infinitesima. In questo modo si può utilizzare un sensore con sensibilità finita, che risulta realizzabile. Ovviamente l'ottica, essendo una modellizzazione, introduce degli errori e delle distorsioni rispetto al modello pin-hole.

2) Si spieghi perché le stime ai minimi quadrati sono gravemente distorte dagli outliers. Gli outliers sono dei punti che hanno degli scostamenti maggiori rispetto a quelli previsti dal modello probabilistico. Sono trattati come dei normali dati affetti da errore.

La stima ai minimi quadrati tenta di minimizzare il quadrato degli scostamenti dei singoli punti dal valore medio. Risulta evidente, quindi, che se un punto è molto distante dalla media, esso "attraherà" la stima verso di sé.

La presenza di un outlier molto scostato può prendere una stima molto imprecisa.

Esempio di molle collegate a retto e punto (risposta elastica quadratica con l'estensione)

3) Quanti gradi di libertà ha un corpo rigido nello spazio?

Un corpo tridimensionale <sup>rigido</sup> nello spazio presenta 6 coordinate libere che sono le traslazioni sui tre assi e le rotazioni possibili sugli assi stessi.

4) Quali sistemi di coordinate conoscete per fornire la posizione di un punto?

coordinate cartesiane  $\Rightarrow$  sono le normali coordinate utilizzate per dislocare un punto in un sistema ortogonale (in cui gli assi sono perpendicolari).

Ogni coordinate rappresenta uno scostamento sul relativo asse.

coordinate omogenee  $\Rightarrow$  chiameremo coordinate omogenee di un punto  $P(x, y)$  del piano una qualsiasi terna ordinata  $(X, Y, W)$  tale che  $W \neq 0$ ,

$X/W = x$  e  $Y/W = y$ . Sono definite a meno di un coefficiente di proporzionalità. Permettono di rappresentare punti  $\rightarrow \infty$

5) Si spieghi sinteticamente cosa si intende per parametri estrinseci di una camera

I parametri estrinseci di una camera sono un qualunque set di parametri geometrici che identifichi univocamente la trasformazione tra il sistema di riferimento camera ed il sistema di riferimento mondo.

In particolare, si tratta di 3 traslazioni e di 3 rotazioni, tutte riferite ai tre assi spaziali.

I parametri estrinseci, quindi, sono quei parametri che non dipendono da come è fatta internamente la camera.

6) Si descriva cos'è l'aspect-ratio

L'aspect ratio indica il rapporto tra l'altesso e la larghezza di una immagine bidimensionale.

Si può parlare di aspect-ratio relativo allo formato dell'immagine, del sensore (che determinerà di conseguenza l'aspetto dell'immagine) e di pixel-aspect-ratio che riguarda l'aspetto del pixel. Questo ultimo parametro è molto importante nel momento in cui si vanno a fare delle misurazioni sulla scena.

7) Descrivere brevemente le differenze tra la retta di interpretazione e proiezione

La retta di interpretazione e proiezione sono la stessa retta, ma gli viene assegnato un nome differente in relazione al fatto che venga calcolata dal punto immagine o dal punto scena. In ogni caso, la retta congiunge il punto scena a cui si riferisce con il punto immagine (proiezione), passando per il centro di proiezione. Siccome per determinare una retta sono sufficienti solo due punti ed il centro di proiezione è fisso, allora la retta può essere determinata come:

- Retta che dal punto scena passa per il centro di proiezione e "colpisce" il piano immagine. Questa retta prende il nome di retta di proiezione.
- Retta che dal punto immagine passa attraverso il centro di proiezione.

Questa retta prende il nome di retta di interpretazione. In queste cose, nulla si può più dire riguardo la posizione  $Z$  (profondità) del punto scena originale.

8) Durante la calibrazione di una camera, quali conseguenze può avere lo scegliere punti appartenenti ad uno stesso piano 3D?

La scelta di punti 3D coplanari può portare ad avere delle equazioni che non vincolano a sufficienza il sistema e, quindi, ottenere dei risultati non corretti o incompleti. Inoltre, punti coplanari non forniscono informazioni addizionali al sistema e, di conseguenza, sono inutili.



ai fini della calibrazione. Un altro caso per cui i punti non vincolano a sufficienza il sistema è che si trovino su uno twisted-cubic.

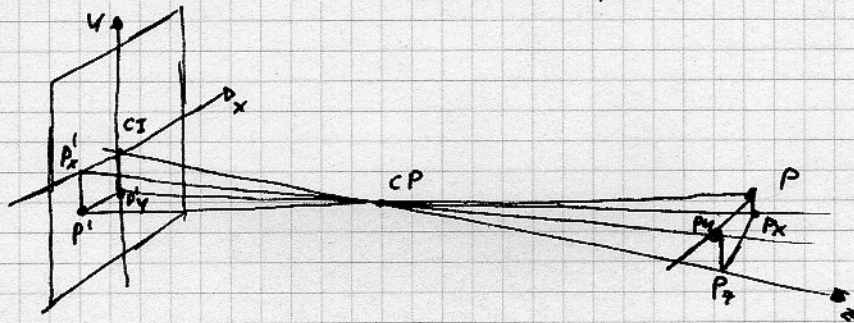
9) Durante la calibrazione della proiezione, cosa è noto e cosa sono le incognite?

Durante il processo di calibrazione sono note le posizioni del punto nello spazio (posizione del punto scena rispetto al centro di proiezione) e la posizione del punto proiettato sul piano immagine.

Una volta noti i parametri mondo  $x, y, z$  ed i parametri immagine  $u, v$  è possibile stimare i parametri  $m_i$  della matrice di calibrazione.

10) Si definisca il seguente sistema di riferimento camera (C): si osserva il piano immagine  $\Pi$  dal centro della proiezione (CP), l'origine di C è nel centro immagine (CI), l'asse x va verso destra e l'asse y verso l'alto; l'asse z da CI a CP.

Si scriva, rispetto al sistema di riferimento C, sia la sua proiezione in coordinate omogenee che la matrice di proiezione in coordinate omogenee.



Si considerano due triangoli con vertice in CP. Siccome è una proiezione, allora saranno simili. Consideriamo:

$$P_x \overset{\Delta}{CP} P_z \cong P'_x \overset{\Delta}{CI} P'_z$$

Si possono mettere in rapporto tra di loro i lati di questi triangoli

$$P'_x CI : CI CP = P_x P_z : P_z CP$$

$$-x : f = X : (Z - f)$$

Se si considerano i triangoli con lato sull'asse y si ottiene che:

$$P'_y CI : CI CP = P_y P_z : P_z CP$$

$$-y : f = Y : (Z - f)$$

Da cui si deducono le relazioni:

$$\begin{cases} x = -\frac{f X}{Z - f} \\ y = -\frac{f Y}{Z - f} \end{cases}$$

Nelle coordinate omogenee c'è una variabile di scala aggiunto.

Supponiamo, quindi, che le nostre coordinate omogenee siano

$$(u, v, w)$$

da cui  $x = u/w$  e  $y = v/w$

$$\begin{pmatrix} u \\ v \\ w \end{pmatrix} = \begin{pmatrix} m_{11} \\ m_{12} \\ m_{13} \end{pmatrix} \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix}$$

$$x = \frac{u}{w} = -\frac{fX}{Z-f} = \frac{m_{32}X + m_{22}Y + m_{32}Z + m_{42}}{m_{23}X + m_{23}Y + m_{33}Z + m_{34}}$$

$$y = \frac{v}{w} = -\frac{fY}{Z-f} = \frac{m_{22}X + m_{22}Y + m_{32}Z + m_{42}}{m_{23}X + m_{23}Y + m_{33}Z + m_{34}}$$

$$M = \begin{pmatrix} -f & 0 & 0 & 0 \\ 0 & -f & 0 & 0 \\ 0 & 0 & 1 & -f \end{pmatrix}$$



# RADIOMETRIA

La radiometria è la scienza che studia la quantità di energia luminosa che colpisce i punti della scena.

Shape From Shadow prevede che io conosca l'illuminante e conosca il comportamento della oggetto (comportamento riflessivo)

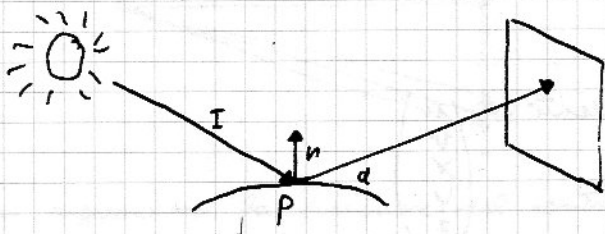


image radiance  $\rightarrow$  luminosità immagine  
radiance  $\rightarrow$  irraggiamento dei punti scena } grandezza che valgono per unità di area

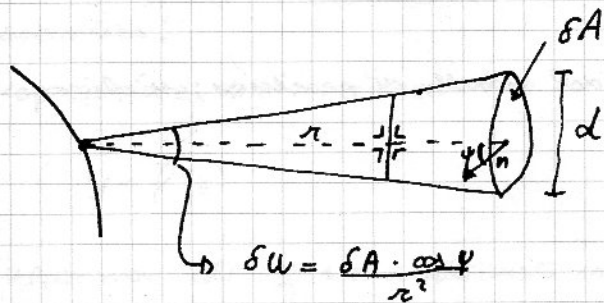
Modello di riflettanza superficiale

Modelli Lambertiani (il comportamento riflessivo è lo stesso in tutte le direzioni)

$P$   $\rightarrow$  colore del materiale

$$L = P \cdot I \cdot n$$

$\nwarrow$  direzione dell'illuminazione



$$\delta u = \frac{\delta A \cdot \cos \psi}{r^2}$$

$\psi$  è l'angolo tra normale della superficie e il raggio

$$E(\text{potenza incidente}) = \frac{\delta P}{\delta I}$$

Equazione fondamentale della radiometria

$$E(P) = L(P) \cdot \frac{\pi}{4} \cdot \left(\frac{d}{f}\right)^2 \cdot \cos^4 \alpha$$

dove

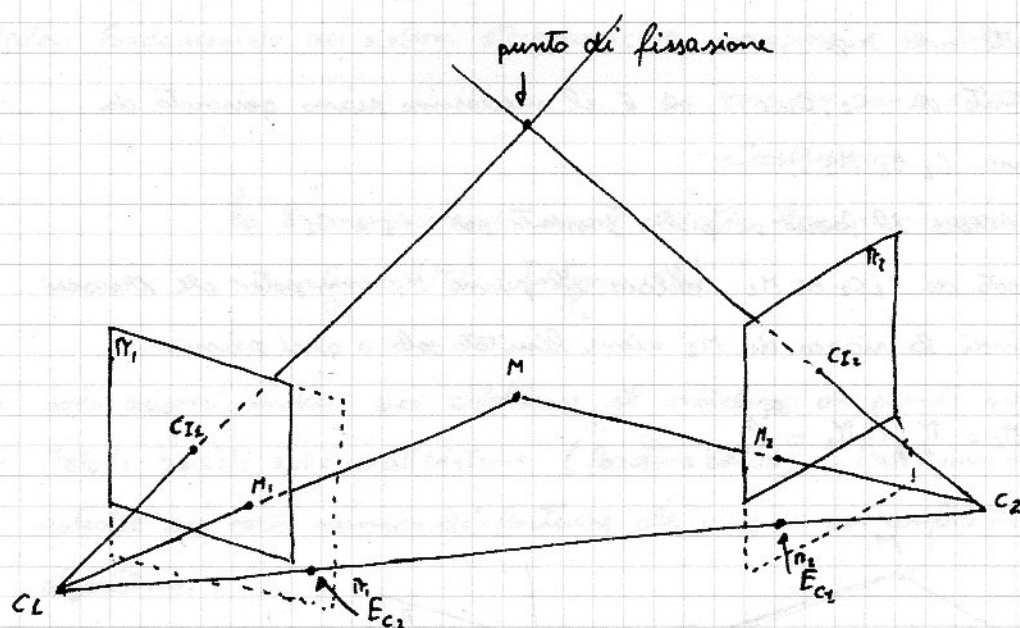
$$\alpha = \psi$$

$d$  = apertura lente

$f$  = lunghezza focale (F-number)

# VISIONE STEREOSCOPICA

La stereometria è quello tecnico che permette di ricavare informazioni 3D senza sapere nulla a priori del mondo. Viene fatto una ricostruzione dei singoli punti della immagine.



$\overline{C_1 C_2} \Rightarrow$  linea di base (può appartenere o meno al piano immagine)

Nel caso in cui le camere siano calibrate, allora sono note le posizioni delle rette.

$$\left. \begin{array}{l} {}^w R_{C_1} \text{ noto} \\ {}^w R_{C_2} \text{ noto} \end{array} \right\} \text{ costanti}$$

EPIPOLO  $\Rightarrow C_1$  e  $C_2$ , essendo un punto scampato ~~esterno~~, può essere proiettato nell'altra camera (eventualmente anche al di fuori del range di coordinate ammissibili!)

$$\overline{C_1 C_2} \cap \pi_1 = E_1 \quad (E = \text{epipolo})$$

$$\overline{C_1 C_2} \cap \pi_2 = E_2$$

Se le rette di interpretazione del punto  $M$  delle due camere sono "sgambe" e non si intersecano, allora devo cercare una intersezione a minima distanza. Questo significa che estendo l'intervallo di ricerca dell'intersezione ad un intorno compatibile agli errori del sistema.

Trovare il punto  $M$  a partire dalle sue proiezioni si chiama problema della ricostruzione o della triangolazione

Un problema differente è quello dello stereo matching, ovvero dell'individuazione dello stesso punto sull'immagine secondaria

L'immagine da cui parte l'analisi dello stereo matching viene chiamata immagine primaria, mentre quella su cui si va a ricercare la corrispondenza viene chiamata immagine secondaria



Un generico algoritmo di stereomatching parte da un punto sull'immagine primaria ed effettua una ricerca del corrispondente nella secondaria.

### Vincolo epolare

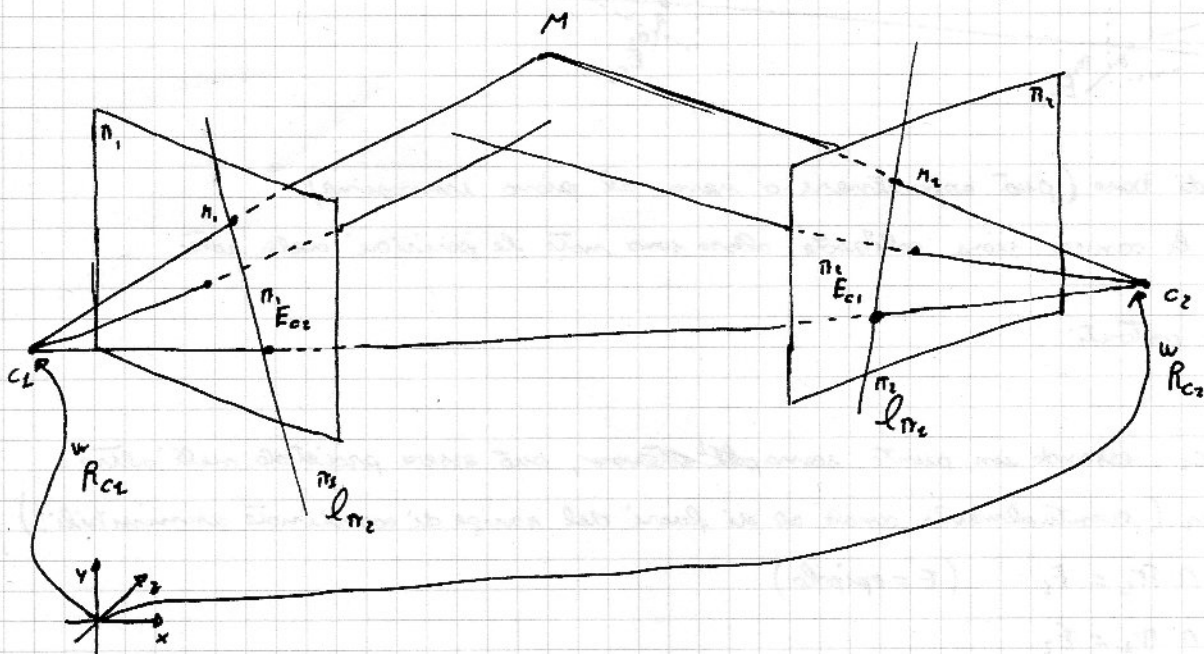
Il vincolo epolare permette di restringere la zona di ricerca del punto corrispondente sulla secondaria a solamente una riga.

Si considera il piano costituito da  $C_1, C_2$  e  $M$  che è il medesimo piano generato da  $C_1 M_1 M$  e da  $C_2 M_2 M$  oppure  $C_1 C_2 M_2$

In altre parole si può considerare il fascio di piani passanti per  $C_1$  e  $C_2$

Se considero il piano formato da  $C_1 C_2$  ed  $M_2$ , allora il punto  $M_2$  non potrà che trovarsi su quello stesso piano, quindi la ricerca di  $M_2$  viene limitata solo a quel piano.

$$\text{linea epolare coniugata a } M_2 = \pi_{epip} \cap \pi_2 = \ell_{\pi_2}$$



La ricerca può essere fatta in base a features che permettono di riconoscere oggetti dell'immagine primaria nella secondaria.

Viene creato una lista di features dell'immagine sinistra e destra e per ogni feature dell'immagine primaria si ricerca la feature migliore (più simile) nella secondaria

Un altro modo è quello dello mappa di disparità; anziché avere una lista di appartenenza si utilizza una matrice costee.

Se ho un punto  $M_1(x_1, y_1)$  il suo punto equivalente su  $\pi_2$  sarà  $M_2(x_2 + \Delta x, y_1 + \Delta y)$

$$M_1(x_1, y_1) \rightarrow M_2(x_2 = x_1 + \Delta x, y_2 = y_1 + \Delta y)$$

$\Delta$  è una funzione vettoriale

$$\Delta = \begin{cases} \Delta x = D_x(x_1, y_1) \\ \Delta y = D_y(x_1, y_1) \end{cases}$$

Se le rette centro immagine sono parallele, allora le rette epipolari sono tra loro parallele e coincidenti con le righe dell'immagine. In questo caso, la mappa di disparità avrà la coordinata "y" nulla (in quanto le variazioni sono possibili solamente sulla linea epipolare e questo è coincidente con l'asse x)

Ipotesi fondamentale dei sistemi stereoscopici  $\Rightarrow$  maggiore è la focalità di questa ipotesi maggiore sono i problemi che si hanno nella ricostruzione.

↳ L'immagine dei punti che tra loro corrispondono è simile

$$\text{Intorno } (M_1) \underset{\Pi_1}{\cong} \text{Intorno } (M_2) \underset{\Pi_2}{\cong}$$

Vi sono diversi metodi per effettuare il matching di punti nell'immagine secondaria:

- Metodi basati sulle caratteristiche (feature based). Dall'immagine di partenza vengono estratte un certo numero di features che vengono poi confrontate da un apposito algoritmo.
- Metodi pixel-level. Agiscono senza features, agendo direttamente sull'immagine originale

↳ Algoritmi a correlazione  $\Rightarrow$  Analizzano la somiglianza delle immagini in una certa area, confrontando la correlazione. Il principale metodo è SSD (square sum of differences), ma ve ne sono altri, come la cross correlazione normalizzata

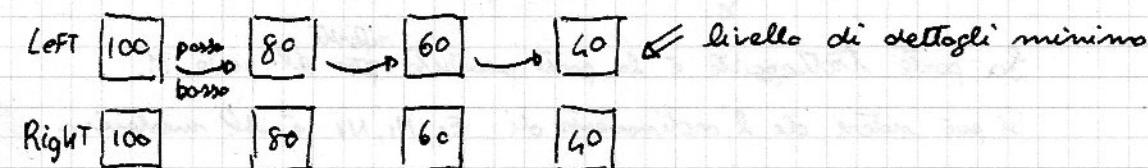
↳ Metodi pixel based utilizzano l'energia del pixel (luminanza) per effettuare delle stime

$$E = \iint_{\Pi_1} [I_L(x + D_x(x), y + D_y(y)) - I_R(x, y)] + [\nabla^2(D)] dx dy$$

Questo parte rappresenta la variazione di scena e serve per limitare le discontinuità presenti nella scena

MULTI risoluzione

Viene creato uno spazio delle scale con le immagini a diversa scala/qualità



Lo spazio delle scale viene creato applicando un certo numero di volte filtri passa-basso all'immagine originale. In questo modo, dall'immagine vengono eliminati i dettagli.

In una immagine con pochi dettagli è più semplice effettuare la ricerca di correlazione



L' algoritmo, quindi, viene fatto partire dall' immagine a risoluzione minore (che contiene minori dettagli).

L' inisidiosità al passo successivo viene data dal passo precedente, quindi è come se l' algoritmo venisse guidato verso la posizione giusta con diversi step.

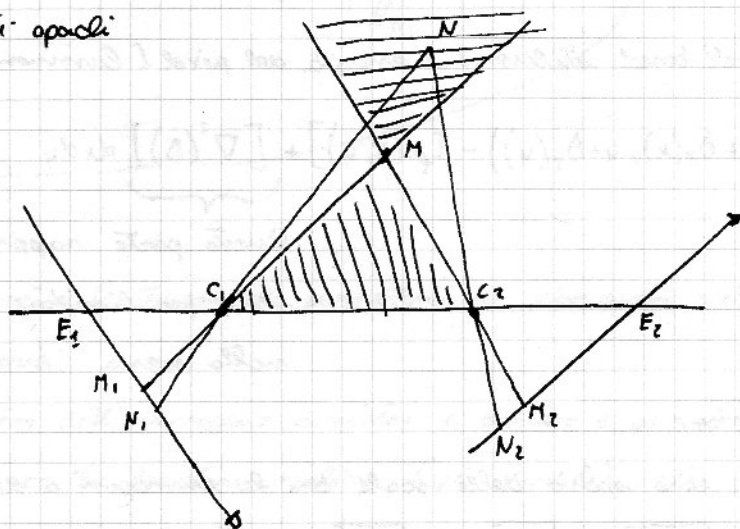
Nei algoritmi a correlazione, l' inisidiosità del passo successivo influenza sulla dimensione dell' area di ricerca, limitandola.

I metodi feature based esistono solamente dove esistono le features!

liste associazioni  $\Rightarrow$  funzione disparità

Oltre a quello epipolare, vi sono altri vincoli che è possibile assumere:

- unicità  $\Rightarrow$  se ho una feature di sinistra e una di destra e trovo un vincolo tra queste due features, allora nessun altro vincolo dovrà esistere con base queste due features.
- continuità  $\Rightarrow$  si ipotizza che il mondo e lo scena sia fatto di oggetti con superfici semplici e continue.
- gradiente funzione disparità  $\Rightarrow$  se si mette un vincolo sul gradiente disparità si genera una zona proibita un poco più ampia di quella del vincolo di ordinamento (vedi vincolo di ordinamento)
- vincolo di ordinamento  $\Rightarrow$  il vincolo dell' ordinamento può essere applicato ad oggetti opachi



La parte tratteggiata è la parte proibita <sup>rispetto</sup> per il punto M.

Si può notare che l' ordinamento di  $E_1, M_1, N_1$  è il medesimo per  $E_2, M_2, N_2$ . Quindi si può dire che secondo questo vincolo, punti che mantengono lo stesso allineamento nell' immagine secondaria non sono ammessi.

- È anche possibile inserire dei vincoli sul fatto che nello sceno vi sia la coplanarità tra alcuni punti. Così facendo, si può forzare la coplanarità anche nell'immagine secondaria ed ottenere dei risultati migliori.

## Algoritmi di stereomatching

### - Marfoggio e Grimson

Questo algoritmo è basato sulla tecnica del rilassamento. Si basa sul vincolo di continuità ed unicità.

Chiamiamo un pixel generico della primaria  $n_i$  e  $m_j$  della secondaria

Viene calcolato per ogni pixel  $n_i, m_j$  una misura di confidenza "c" che specifica quanto sono simili (o quanto è probabile) che i due pixel siano associati.

La misura di confidenza "c" viene calcolata come:

$$c(n_i, m_j) = \begin{cases} 1 & \text{se } |I(n_i) - I(m_j)| < \text{soglia} \\ 0 & \text{altrimenti} \end{cases}$$

$$c^{n+1}(n_i, m_j) = \begin{cases} 1 & \text{se } |n_i| \in V_i \mid c^n(m_i, m_j) = 1 \text{ con } m_j \in V_j \\ 0 & \text{altrimenti} \end{cases}$$

### - Algoritmo Pollard Mayhew Frisby

Questo algoritmo prevede una pre-elaborazione delle immagini da cui vengono estratti dei token i quali dovrebbero preferibilmente includere un certo numero di features

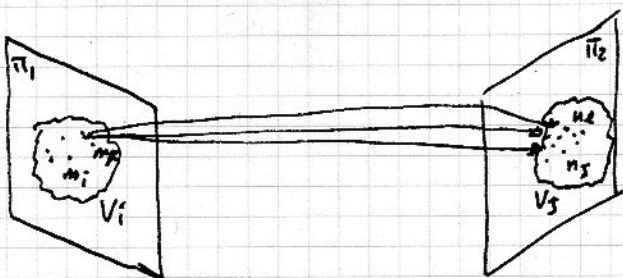
Viene calcolato il valore  $c_{ij}$  che rappresenta la bontà di associare il token  $T_i$  della immagine primaria con il token  $T_j$  dell'immagine secondaria.

Per il calcolo viene considerato un intorno di  $T_i$  e  $T_j$  (in particolare dei pixel  $m_i$  ed  $n_j$ ) che verranno rispettivamente chiamati, pixel  $m_k$  ed  $n_l$ .

Di tutti i pixel  $m_k$  ed  $n_l$  ve ne sono solo alcuni che supportano l'associazione e sono quelli che soddisfano il vincolo di disparità del gradiente:

$$DG(m_i, n_j, m_k, n_l) < \delta \quad \left\{ \text{Disparity Gradient} \right\}$$

Quindi, solamente le coppie di pixel  $m_k$  ed  $n_l$  che supportano questo relazione potranno supportare il match tra  $m_i$  ed  $n_j$ .





Quindi, viene calcolata la forza dell'associazione (Strength Match) mediante la seguente funzione:

$$SM(m_i, m_s) = c_{ij} \sum_{k \in V_i} \frac{1}{\text{dist}(m_i, m_k)} \cdot \max_{k \in V_i} \left( \frac{c_{ke}}{\text{dist}(m_s, m_e)} \right) \cdot \sigma(DG(m_i, m_s, m_k, m_e))$$

$$\sigma(DG) = \begin{cases} 1 & \text{se } DG < \Delta \\ 0 & \text{se } DG > \Delta \end{cases}$$

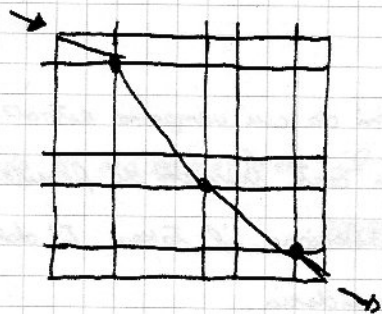
Quindi, nel passaggio da una iterazione alla successiva vengono eliminate le associazioni non forti.

#### - Baker-Beauford

Questo algoritmo funziona sulle basi delle seguenti assunzioni:

- sistemi a camere parallele
- problema di match tra due linee

L'algoritmo costruisce una matrice con le distanze tra le varie features e cerca un cammino minimo per giungere dal lato opposto della matrice. Il costo del percorso scelto verrà dato dallo similitudine tra le features.



Questa procedura viene ripetuta per ogni linea dell'immagine (retto epipolare visto che le camere sono parallele).

#### - Ohta-Kanade

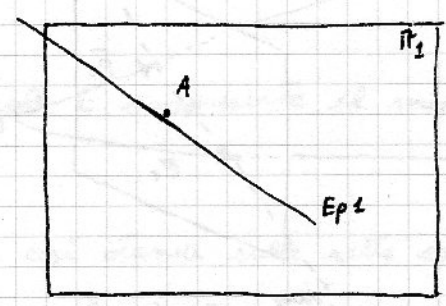
Aggiunge al precedente algoritmo anche una parte di ricerca tra differenti scan lines, ovvero propaga la ricerca del percorso minimo anche alle epipolari successive

## Sistema Trinoculare

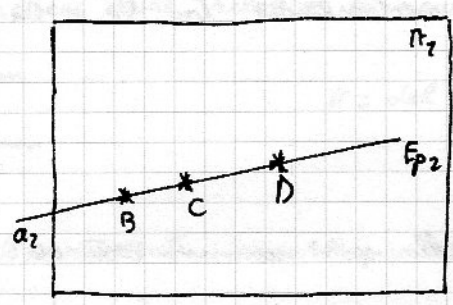
Il sistema di visione trinoculare si basa sulla presenza di tre camere che osservano la scena.

Esiste un vincolo epipolare trinoculare che permette di rendere più preciso e veloce il match di punti.

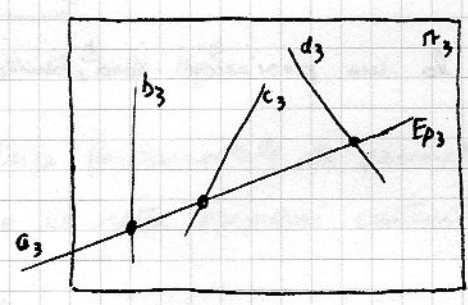
Il vincolo epipolare trinoculare è una condizione necessaria, ma viene spesso usata come condizione sufficiente.



Si trova una feature nell'immagine primaria



Si trova la retta epipolare coniugata nell'immagine secondaria e si identificano le possibili feature che matchano



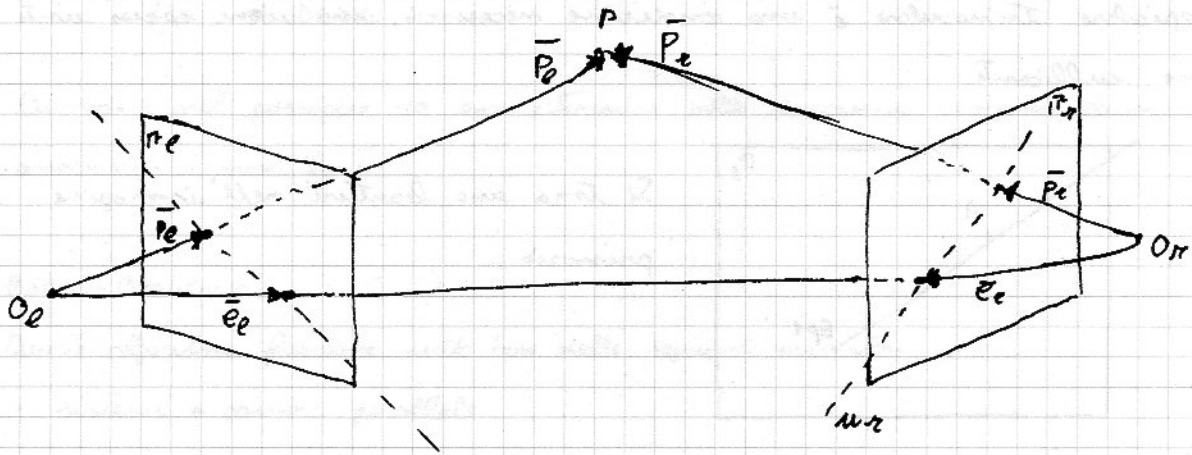
Si trovano le rette epipolari del punto della primaria e dei corrispondenti punti sulla secondaria. Questi si intersecano ed è necessario verificare che nell'incrocio sia presente la stessa feature.

In questo modo la ricerca della corrispondenza è più breve perché viene fatto con l'ausilio dell'immagine terziaria e verificando solo che nell'incrocio delle epipolari sia presente la feature cercata.



## Geometria epipolare

La geometria epipolare può servire per fare la ricostruzione 3D con sistemi uncalibrated.



Nel caso di camere parallele, la matrice di rototraslazione tra  $O_e$  e  $O_r$  si riduce ad una semplice traslazione

$$\bar{T} = (\bar{O}_r - \bar{O}_e)$$

La relazione tra le proiezioni di un punto nello spazio saranno dati da:

$$\bar{P}_r = A(\bar{P}_e - \bar{T}) \Rightarrow R^T \cdot \bar{P}_r = \bar{P}_e - \bar{T} \Rightarrow R^T \cdot \bar{P}_r = \bar{P}_e - \bar{T}$$

La relazione tra il punto nello spazio e la sua proiezione sarà, dunque:

$$\begin{cases} \bar{P}_e = \frac{f_e}{z_e} \cdot \bar{P}_e \\ \bar{P}_r = \frac{f_r}{z_r} \cdot \bar{P}_r \end{cases}$$

Per scrivere l'equazione del piano epipolare si può formare la coplanarità tra i punti:  $P_e, T$  e  $P_e - T$ , ottenendo:

$$(P_e - T)^T \cdot T \times P_e = 0 \quad \{ R^T \cdot P_r = P_e - T \}$$

$$(R^T P_r)^T \cdot T \times P_e = 0$$

$$P_r^T \cdot R \cdot T \times P_e = 0$$

$$P_r^T \cdot R \cdot S \cdot P_e = 0$$

Il prodotto  $R \cdot S$  rappresenta la posizione relativa tra le due camere. La matrice risultante da questo prodotto prende il nome di matrice essenziale  $E = R \cdot S$ ,

Quindi posso scrivere

$$P_r^T \cdot E \cdot P_e = 0$$

Dalle equazioni precedentemente scritte si può ~~dedurre~~ dedurre che:

$$\begin{cases} p_e = p_e \cdot \frac{z_e}{f_e} \\ p_x = p_x \cdot \frac{z_x}{f_x} \end{cases}$$

Quindi, sostituendo, si ottiene che

$$p_e^T E p_e = 0$$

$$u_x = E \cdot u_e$$

La matrice essenziale  $E$  rappresenta il mapping tra i punti e le rette epipolari associate.

Le informazioni così ottenute sulle rette epipolari, però, prescindono dalle caratteristiche del sistema stereo che si sta utilizzando. Chiamiamo  $M_e$  e  $M_x$  le matrici dei parametri intrinseci delle relative camere; si avrà, dunque:

$${}^m p_e = M_e^{-1} p_e$$

$${}^m p_x = M_x^{-1} p_x$$

$m$  = del piano; mem = punto memoria

Sostituendo nelle precedenti formule si ottiene:

$${}^m p_x \cdot M_x^{-1} \cdot E \cdot M_e^{-1} \cdot p_e = 0$$

$$F = M_x^{-1} \cdot E \cdot M_e^{-1}$$

$F$  è la matrice fondamentale che permette di effettuare il mapping tra i punti in memoria e le rette epipolari costruite sui punti in memoria.

$$u_x = E \cdot p_e$$

$${}^m u_x = F \cdot {}^m p_e$$

È possibile computare i 9 parametri della matrice  $F$  utilizzando la decomposizione SVD

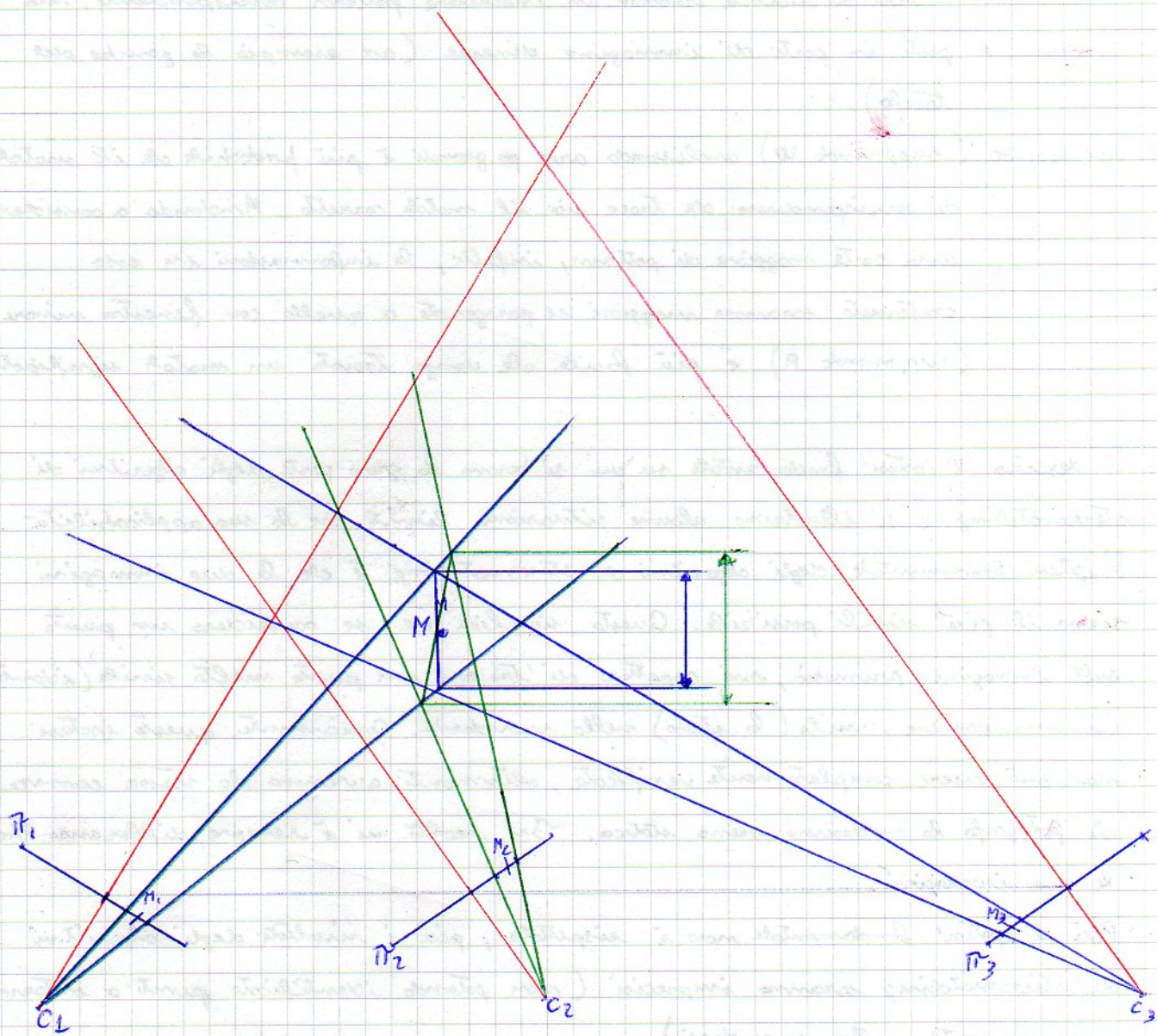
$$A = U \cdot D \cdot V^T$$



1) Si illustrano quali vantaggi e svantaggi ci si può aspettare dall'incremento della baseline in un sistema stereoscopico

SVANTAGGI  $\Rightarrow$  se la dimensione della baseline aumenta, aumento di conseguenza la falsità dell'ipotesi fondamentale dello stereomatching. Questo perché più la baseline è grande più le telecamere asserviscono la scena da un punto differente e, quindi, le immagini risulteranno diverse

VANTAGGI  $\Rightarrow$  più aumento la dimensione della baseline migliore sarà la ricostruzione del vero fatto. All'aumentare della baseline, infatti, l'area di incertezza del punto sarà minore e, quindi, la ricostruzione più accurata





2) Nei sistemi stereoscopici a correlazione ha un ruolo rilevante il parametro "dimensione della finestra di calcolo della correlazione". Si illustrano quindi vantaggi e svantaggi ai si può aspettare dall'incremento di questo parametro

svantaggi  $\Rightarrow$  (supponendo  $W$ ) l'algoritmo è meno stabile nel caso in cui l'ipotesi fondamentale non valga. Con l'incremento della finestra di correlazione, infatti, è meno probabile trovare una correlazione nell'immagine secondaria se il pattern immagine è diverso (come nel caso di visioni di scorcio). Essendo i due pattern differenti per la visione di scorcio ed andando ad analizzare un'area di grandi dimensioni, la correlazione sarà molto influenzata dalla distorsione del pattern

(supponendo  $A$ ) l'algoritmo dovrà analizzare un'area maggiore e quindi computazionalmente sarà maggiore. Inoltre, non limitando l'area di ricerca risale di includere pattern corrispondenti, ma posti in parti di immagine diverse (ad esempio le gambe del tavolo).

vantaggi  $\Rightarrow$  (supponendo  $W$ ) analizzando aree ~~più~~ grandi è più probabile che il match di corrispondenza che trova sia il match corretto. Andando a considerare una parte maggiore di pattern, infatti, le informazioni in esso contenute saranno maggiori se paragonate a quello con finestra minore (supponendo  $A$ ) è più facile che venga trovato un match significativo

3) Si descriva l'ipotesi fondamentale su cui si basano la gran parte degli algoritmi di stereomatching e si illustrino alcune situazioni limite per la sua applicabilità. L'ipotesi fondamentale degli algoritmi di stereomatching è che le due immagini siano il più simili possibile. Questo significa che se considero un punto sull'immagine primaria, mi aspetto di trovare un punto molto simile (identico) in una posizione simile (lo stesso) nella secondaria. Ovviamente questa ipotesi non può essere completamente verificata, altrimenti avremmo la stessa camera da fotografo la medesima scena statica. In realtà vi è sempre differenza tra le due immagini.

Più l'ipotesi fondamentale non è rispettata, più i risultati degli algoritmi di stereomatching saranno imprecisi (non potranno identificare punti o intorni corrispondenti nella secondaria)

Le casi limite sono le visioni di scorcio, in cui le visioni tra le due camere varia moltissimo e, quindi, la quale l'ipotesi fondamentale sarà meno rispettata.



4) Si illustra il vincolo epipolare trinoculare ed il suo uso nell'ambito dei sistemi stereoscopici

Vedi pagina accanto.

5) Funzionamento algoritmo Marr Poggio

È un algoritmo a rilasamento. Nella prima iterazione viene calcolata una soglia tra i pixel della primaria e della secondaria. Se l'intensità del pixel è superiore a una certa soglia (se la differenza tra le intensità è più piccola di un riferimento), allora verrà posto la confidenza "c" al valore 1.

Nelle iterazioni successive si analizza un intorno di ogni pixel e si controlla se il numero di pixel la cui confidenza è 1 è superiore ad un certo numero.

In tal caso il pixel la cui confidenza è 1 il cui intorno è analizzato sarà nuovamente posto a 0.

In questo modo, l'algoritmo è in grado di propagare la soluzione ottimale.

6) Algoritmo Pollard

L'immagine deve essere pre-elaborata e vengono estratti dei token. L'algoritmo valuta la bontà di associazione di due token considerando un intorno del token stesso.

La forza dell'associazione verrà calcolata tenendo conto della distanza dei pixel dell'effettivo token nella primaria e ricercando il pixel nella secondaria che non violi una condizione sul disparity gradient e che massimizzi il rapporto tra il peso di associazione e la distanza dal token.

Il disparity gradient è una funzione che limita la disparità dello scarto sotto una certa soglia, funzionando come un filtro passa-basso.

Ad ogni passaggio vengono eliminate le associazioni meno forti mentre vengono propagate quelle più forti.

7) Backer - Beaupre

Questo algoritmo si basa sulle assunzioni che il sistema di visione sia a camera parallele e che vi sia problema di match tra due righe.

L'algoritmo costruisce una matrice di features per l'immagine primaria e secondaria per ogni riga immagine. La migliore associazione tra le features sarà data dalla ricerca del cammino minimo attraverso la matrice. Il costo di ciascun percorso verrà dato dal grado di similarità tra le features.

8) Algoritmo di Ohta - Kanode

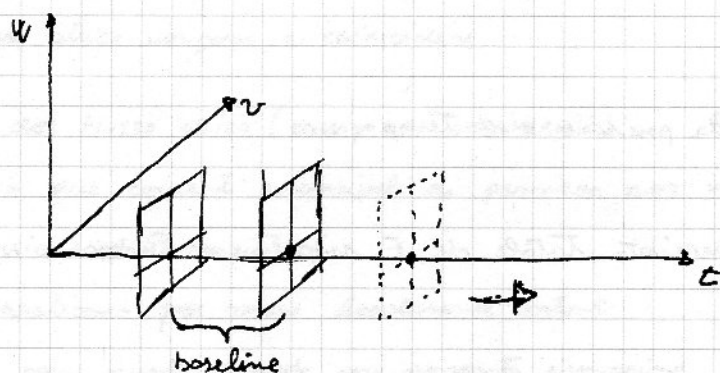
La struttura base è uguale al precedente, ma aggiunge dei vincoli anche nella ricerca del minimo delle righe successive. In pratica, fa sc. la continuità del pattern immagine



# STRUCTURE FROM MOTION

La ricostruzione tridimensionale di una scena può essere anche fatta sfruttando il movimento della camera. Si supponga di avere una sola camera e di scattare  $n$  foto in sequenza, allora si può supporre come baseline il tempo trascorso tra due immagini.

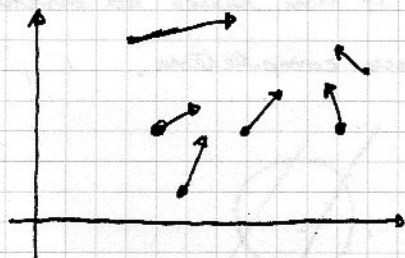
A questo punto si può eseguire una ricostruzione come nei sistemi stereo  
metodi differenziali  $\Rightarrow$  misure dense (misure spazio tempo)  
metodi matching  $\Rightarrow$  misure sparse (basate su feature)



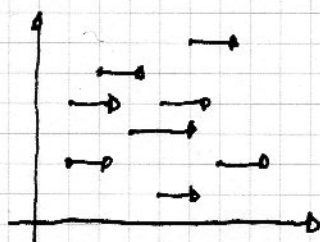
Si possono individuare diverse applicazioni e metodi in relazione alla tipologia di movimento che si può osservare nella scena:

- singolo movimento solido  $\Rightarrow$  nella scena si può osservare un unico movimento relativo con la camera.
- più movimenti  $\Rightarrow$  si possono identificare più movimenti nella scena. In questo caso, uno dei problemi più complessi riguardano la motion-segmentation, ovvero la suddivisione della scena in segmenti che contengono movimenti singoli.

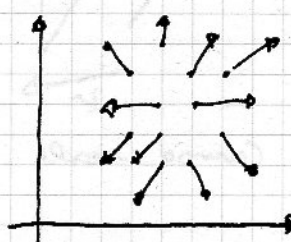
Il campo vettoriale è la proiezione nel piano immagine del campo di movimento 3D della scena. Per ogni pixel / feature c'è un vettore che fornisce informazioni sulla tipologia di movimento a cui è soggetto.



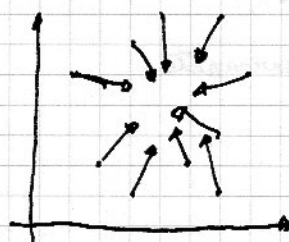
Nel nostro caso andremo ad analizzare movimenti puramente traslatori



moto traslatorio su  
asse  $y$



moto in accelerante



moto in allentamento

} Traslazione su  $z$  }

Optical Flow  $\Rightarrow$  trovo e identifico degli oggetti nell'immagine: che cosa hanno in comune con il campo di movimento?

Nei software di analisi del flusso ottico si suppone che la luminanza sia derivabile (regolare) sia nello spazio che nel tempo.

Equazione di costanza della luminanza immagine

La derivata dell'energia luminosa di un punto che si sposta nella scena rispetto al tempo è costante

$$\frac{dE}{dt} = \phi$$

La funzione energia  $E$  è in relazione alla posizione e al tempo

$$E(x, y, z)$$

$$\begin{cases} x = x(t) \\ y = y(t) \end{cases} \quad \frac{dE}{dt} = \phi \Rightarrow \text{Derivata Totale di } E \text{ rispetto al tempo}$$

$$\frac{dE}{dt} = \left(\frac{dE}{dx}\right) \left(\frac{dx}{dt}\right) + \left(\frac{dE}{dy}\right) \left(\frac{dy}{dt}\right) + \frac{dE}{dt} = 0$$

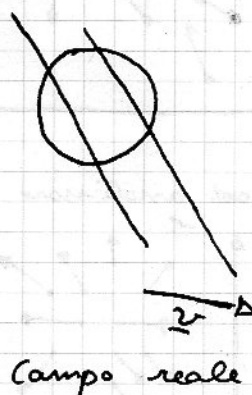
gradiente dello spostamento (campo di movimento)  
 componenti del gradiente

$$\Rightarrow (\nabla E)^T \cdot \underline{v} + E_z = \phi$$

$\nabla E$   $\rightarrow$  gradiente  
 $\underline{v}$   $\rightarrow$  campo di movimento

Con questa relazione riesco a trovare solamente una componente del campo di movimento. In particolare rilevo la componente di  $\underline{v}$  che è nella direzione del gradiente di  $E$ .

Sorge, quindi, il problema dell'apertura. Se non riesco ad osservare la totalità dell'oggetto, non vedo il movimento nella sua completezza.





La velocità di cambiamento della normale sarà data dalla relazione:

$$\frac{dn}{dt} = \omega \times n$$

Quando  $\Delta v = 0$ ? Nel caso in cui il moto sia puramente traslatorio, oppure vi sia un movimento rigido in cui la direzione dell'illuminazione sia parallela alla velocità angolare.

Il campo di movimento che si può dedurre a livello immagine viene chiamato Flusso ottico. In alcune situazioni, come quelle descritte in precedenza, campo di movimento e flusso ottico vengono a coincidere.

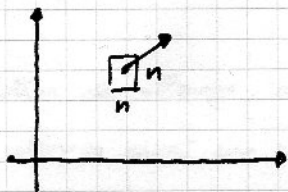
Stima del Flusso ottico (campo di movimento)

Ci sono due correnti principali di pensiero per valutare il flusso ottico:

- Tecniche differenziali

Si applicano per campi localmente costanti.

Per ogni punto prende un supporto immagine



Calcola per ogni punto la derivata del punto origine.

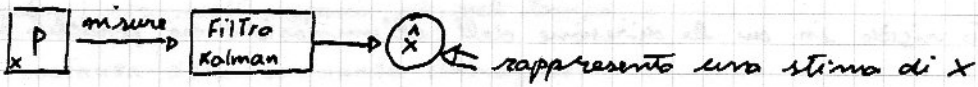
Il flusso ottico è il vettore che minimizza una certa cifra di merito

- Feature based

Determino le tracce con un algoritmo di tracking. Vi sono molti algoritmi di tracking, ma uno dei più famosi ed utilizzati è il filtro di Kalman che permette di ottenere una stima della futura posizione della feature.

# Filtro di Kalman

È un sistema dinamico usato per stimare lo stato di un sistema dinamico e partire da misure rumorose e da un modello incerto del sistema.



Un sistema dinamico è un insieme di equazioni:

$$\begin{cases} \boxed{\quad} & \leftarrow \text{evoluzione dello stato (definite con } n \text{ equazioni in base alla} \\ \boxed{\quad} & \leftarrow \text{misure} \end{cases} \quad \text{precisione voluta)}$$

$$\dot{x} = f(x, u) \quad \vee \quad y = g(x)$$

↑ controllo

Se le funzioni  $f$  e  $g$  sono lineari, allora si avrà che:

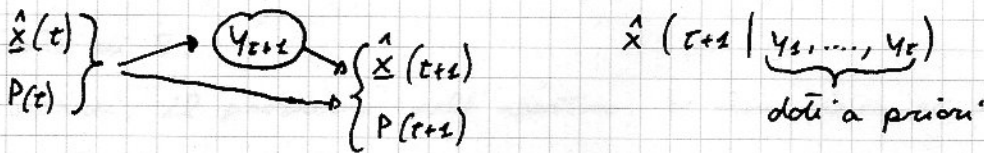
$$\begin{cases} \dot{x} = Ax + Bu \\ y = Cx \end{cases} \quad \Leftarrow \text{nel caso di tempo continuo}$$

$$\begin{cases} \dot{x}(t) = Ax(t) + Bu(t) \\ y(t) = Cx(t) \end{cases} \quad \Leftarrow \text{Tempo discreto}$$

Questo strumento funziona bene nel caso vi siano rumori additivi

Nel nostro caso andremmo una esecuzione con WGN (White Gaussian Noise). Se il rumore è bianco ed il sistema è lineare, allora il filtro di Kalman è un filtro ottimo

$\hat{x}_0, P_0$   $\Leftarrow$  stato iniziale, prima di inserire le misure



La differenza tra la vera misura e quella prevista secondo i dati precedenti si chiama innovazione.

$$\begin{cases} x(k) = \Phi_{k+1} x(k-1) + \epsilon_{k-1} \\ z(k) = Hx(k) + \mu_k \end{cases} \quad \begin{aligned} \epsilon &= N(0, Q) \\ \mu &= N(0, R) \end{aligned}$$

$$\left. \begin{aligned} P'_k &= \Phi_{k-1} P_{k-1} \Phi_{k-1}^T + Q_{k-1} \\ K_k &= P'_k H_k^T (H_k P'_k H_k^T + R_k)^{-1} \\ \hat{x}_k &= \Phi_{k-1} \hat{x}_{k-1} + K_k (z_k - H_k \Phi_{k-1} \hat{x}_{k-1}) \\ P_k &= (I - K_k) P'_k (I - K_k)^T + K_k R_k K_k^T \end{aligned} \right\} \text{FILTRO DI KALMAN}$$

↳ aggiornamento della covarianza, secondo ~~Riccati~~ Riccati



Il filtro di Kalman può essere usato per il Tracking di feature

$\{ \underset{t}{\langle x_i, y_i \rangle}, \underset{t+1}{\langle x_i, y_i \rangle}, \dots \}$  ← elenco di feature nei diversi istanti (immagini)

$x_{(t+1)} \Rightarrow$  funzione dello stato precedente. Si suppone un moto uniforme

$$x(t) = \phi_{t-1} x(t-1) + w(t-1)$$

$$z(t) = H(t)x(t) + u(t)$$

Lo stato del sistema sarà determinato dai velari

$$\text{stato: } \begin{bmatrix} x \\ y \\ v_x \\ v_y \end{bmatrix} \left. \vphantom{\begin{bmatrix} x \\ y \\ v_x \\ v_y \end{bmatrix}} \right\} \text{velocità}$$

$$\phi \Rightarrow \begin{pmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

covarianza  $\left( \begin{array}{c} \text{[grid]} \\ \text{[grid]} \\ \text{[grid]} \end{array} \right)$  ← incertezza della posizione  
dello stima delle feature

Il filtro consente di mettere insieme misure anche se non rappresentano esattamente lo stato del sistema.

Con la stima del filtro che ottengo, avrò una <sup>previsione</sup> ~~precisione~~ della posizione delle feature  
questo previsione che ottengo avrò un certo grado di incertezza.

L'incertezza della previsione è usata per associare la misura al filtro

Il problema dello data association si può avviare con modelli probabilistici, come, ad esempio, l'utilizzo della distribuzione  $\chi^2$

Per determinare quale feature scegliere si può usare la distanza di Mahalanovis che consiste nell'usare la distanza quadrata tra la misura e la previsione.

Mediante queste tracce si può ottenere una ricostruzione 3D con egomotica (movimenti nel mondo) e structure

## Metodo di Fattorizzazione

Il metodo di fattorizzazione permette di ottenere tracking e structure in modo affine e nel momento in cui si hanno tutte le immagini a disposizione. Funziona in condizioni di ripresa ampie (distanze elevate), quindi si può usare la scaled orthographic.

Parte da  $N$  tracce ( $N \geq 3$ ) e prevede che gli  $n$  punti di cui ho le tracce non siano coplanari.

Definiamo il seguente teorema:

### Teorema del rango

Si ha un certo numero di punti per ogni immagine

$P_{i,j}$   $P$  = punto  $i$  = indice temporale  $j$  = indice punto immagine

Si costruisce la matrice delle misure nel seguente modo:

$2N \times n$

$$\begin{pmatrix} x_{11} & x_{12} & \dots & x_{1n} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \dots & x_{nn} \\ \hline y_{11} & y_{12} & \dots & y_{1n} \\ \vdots & \vdots & \ddots & \vdots \\ y_{n1} & y_{n2} & \dots & y_{nn} \end{pmatrix}$$

matrice delle misure  $W$

• Si calcola  $\bar{x}$  e  $\bar{y}$  di tutte le misure

$$\bar{x}_i = \frac{1}{n} \sum_{j=1}^n x_{ij}$$

$$\bar{y}_i = \frac{1}{n} \sum_{j=1}^n y_{ij}$$

$$P_i = (\bar{x}_i, \bar{y}_i)$$

↑  
Baricentro dei punti immagine presi nell'immagine  $i$

• Si calcolano i valori

$$\tilde{W} = \begin{cases} \tilde{x}_{ij} = x_{ij} - \bar{x}_i \\ \tilde{y}_{ij} = y_{ij} - \bar{y}_i \end{cases} \quad \Leftarrow \text{matrice di scostamento (delle misure registrate)}$$

In assenza di misurato rumore nelle misure, la matrice delle misure registrate ha al massimo rango uguale a 3

La dimostrazione del teorema si ottiene sfruttando una fattorizzazione.

$$\tilde{W} = R \cdot S \quad \Leftarrow \text{shape (descrive i punti cambiamento)}$$

↑  
rotazione (rotazioni della camera tra un frame e l'altro)

$$R_i = \underline{i}_i \times \underline{j}_i \quad (\text{versori camera})$$

$$x_{ij} = \underline{i}_i^T \cdot (P_j - T_i)$$

$$y_{ij} = \underline{j}_i^T \cdot (P_j - T_i)$$

$$\tilde{x}_{ij} = \underline{i}_i^T \cdot P_j$$

$$\tilde{y}_{ij} = \underline{j}_i^T \cdot P_j$$

a questo punto, se si definiscono delle opportune matrici, si scopre che  $W = R \cdot S$

$$R_{2N,3} = \begin{bmatrix} r_1^T \\ \vdots \\ r_N^T \\ s_1^T \\ \vdots \\ s_N^T \end{bmatrix}$$

$$S_{3 \times n} = |P_1, P_2, \dots, P_n|$$

$$\tilde{W} = R \cdot S$$

La fattorizzazione non sarebbe unica. Per ovviare a questo problema si aggiungono dei vincoli interni che rendono la fattorizzazione unica.

Si usa la decomposizione SVD

$$\tilde{W} = SVD^T \quad (\text{si prendono i primi 3 valori singolari e gli altri a } \emptyset)$$

$\Rightarrow$  con i nuovi valori singolari forzati, ricatando la matrice  $\tilde{W}$

Le matrici  $R$  e  $S$  vengono costruite come:

$$\hat{R} = V^1 \cdot D^{1/2}$$

$$\hat{S} = D^{1/2} \cdot V^1 T$$

l'apice 1 dipende dal fatto di avere eliminato i valori singolari  $> 3$

$$R = \hat{R} \cdot Q$$

$$S = Q^{-1} \cdot \hat{S}$$



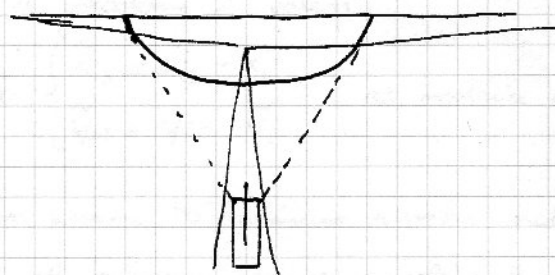
## Visuale omni-direzionale

Nelle normali atticle, il field of view è limitato ad un angolo abbastanza ristretto. Con atticle particolari si ottengono angoli maggiori, ma con costi e distorsioni maggiori. Per curare a questo problema e per avere la capacità di guardare cosa c'è intorno è utile potere sfruttare la visione omni-direzionale

Vi sono molti metodi per ottenere una visione omni-direzionale:

- con una camera che ruota attorno ad un asse fisso e che scatta in sequenza delle foto che vengono poi unite per formare una visione panoramica. Le foto dovranno avere delle aree di sovrapposizione e l'algoritmo di unione è generalmente esoso in termini computazionali. Inoltre il movimento fisico della camera che ruota esclude questo sistema da real-time
- sistemi multicamera calibrati che permettono di scattare foto contemporaneamente in più direzioni. Il sistema è costoso, ma permette di ottenere dei buoni risultati.
- sistemi diottrici (solo lenti ottiche). Si tratta di PAL (Panoramic Angular Lens), ovvero lenti particolari che permettono di ottenere un field of view molto ampio. Queste soluzioni sono molto costose (per l'attica ad hoc) e molto ingombranti.
- sistemi catottrici (specchi e lenti ottiche) sono tra i sistemi più diffusi e più usati in quanto sono poco costosi.

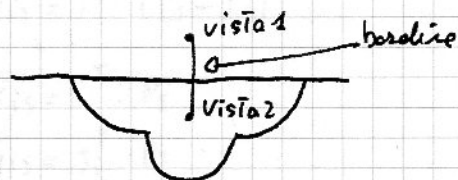
Consistono in una riflessione attraverso uno specchio e una visione con una normale attica. Vi possono essere differenti forme di specchi:



- specchi piani
- " conici
- " parabolici
- " sferici
- " iperbolici
- " custom

Quello che varia tra le differenti forme di specchi è la risoluzione tra lo scenario e il mondo. Alcune forme favoriscono risoluzioni in determinate zone mondo. SVP (Single View Point) si trova, ad esempio, negli specchi iperbolici. È possibile identificare un punto da cui ottenere l'osservazione con una camera normale. Posso ricreare l'immagine corrispondente a una immagine singola con aperture diverse

Stereo From omni



- setup  $\Rightarrow$  aggiustato la posizione di conico e specchio fino ad ottenere +/- quello teorico
- calibrazione  $\Rightarrow$  contempo il non corretto posizionamento

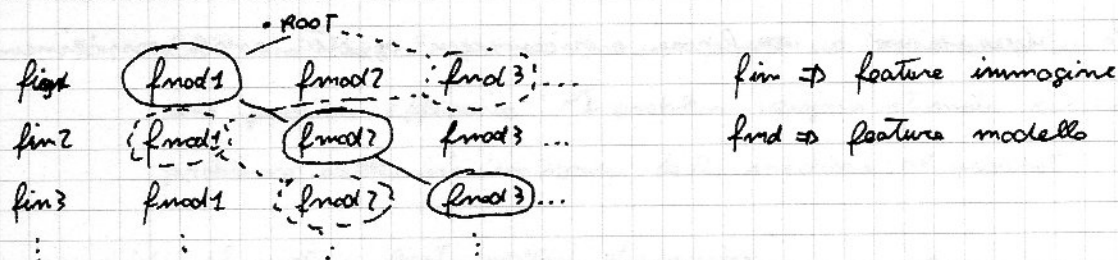
## Riconoscimento

Il problema del riconoscimento è direttamente collegato al problema della localizzazione.

Il riconoscimento viene fatto a partire da un modello memorizzato in un database e può essere fatto anche grazie ad una sola camera.

A questo scopo sono utili gli alberi delle interpretazioni.

Hanno due dimensioni: orizzontale, feature estratte dall'immagine; verticale, feature presenti nel modello memorizzato



In questo modo è possibile trovare un mapping tra le feature immagini e modello.

Si possono anche aggiungere vincoli addizionali come, ad esempio, sull'ordinamento di alcune feature.

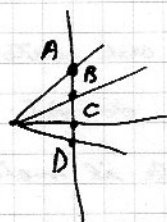
Un altro metodo di riconoscimento è basato sugli invarianti, ovvero delle misure sull'immagine che mi permettono di ricercare direttamente nel database.

L'invariante più noto ed utilissimo è il birapporto. Il birapporto è una funzione che, data la funzione  $T$  proiettiva, dice che la funzione invariante ha lo stesso valore quando applicata alla configurazione immagine e alla configurazione del modello.

Se ho quattro punti tra loro allineati  $A, B, C, D$ , il birapporto tra questi quattro punti viene calcolato come il rapporto tra due prodotti:

$$\frac{\text{distanza tra } A \text{ e } C - \text{distanza tra } B \text{ e } D}{\text{distanza tra } B \text{ e } C - \text{distanza tra } D \text{ e } A}$$

Il birapporto può essere calcolato anche tra 4 linee



Si prende una qualsiasi retta che interseca le 4 rette date e si calcola il birapporto tra i 4 punti trovati.

## Hyper (Hypothesis Prediction and Verification)

È un algoritmo basato sulla rilevazione poligonale, quindi analizza il mondo a segmenti e ha un database a segmenti.

Nota -> usando la normale approssimazione polinomiale si ottiene una lista di segmenti senza nessun tipo di informazioni topologiche. Hyper necessita di queste informazioni topologiche tra i segmenti.

Nel modello vengono identificati (stabiliti) dei segmenti privilegiati ( $S_i$ ). L'algoritmo ricerca, per prima cosa, associazioni tra questi segmenti privilegiati (che ci si aspetta siano sempre visibili e ben identificati nel mondo)

Quando vengono trovate le associazioni sui segmenti privilegiati, vengono analizzate le associazioni con i segmenti ~~diretti~~ direttamente connessi. In particolare vengono prese in considerazione la lunghezza del segmento e l'angolo che esso forma con i segmenti a lui connessi.

In questo modo, Hough riesce a verificare e riconoscere oggetti anche parzialmente occlusi.

### Mosaicing 3D

viste  $\{ \dots \text{feature 3D} \dots \}$   
72i

Viene calcolata la rototraslazione tra le differenti viste in diversi tempi.

Ancile avere il calcolo tra delle rototraslazioni esterne posso avere l'integrazione di viste analizzando come sono posizionate le feature nelle diverse viste e calcolare di conseguenza le rototraslazioni tra le viste.

### Algoritmo ICP (iterative closest point)

Se ho una stima iniziale di  $R$  (rototraslazione della camera verso il mondo), posso avere più viste nello stesso sistema. Uso un criterio di associazione tra le feature e cerco qual'è l'associazione migliore.

Quando ho trovato il "best match" lo utilizzo per modificare la stima iniziale della rototraslazione  $R$  al fine di migliorarlo.

La nuova rototraslazione viene usata per le feature successive.

### RANSAC

- efficienza relativa  $\Rightarrow$  viene determinata come la varianza calcolata dal metodo in rapporto alla migliore varianza ottenibile dai dati.
- punto di breakdown  $\Rightarrow$  rappresenta la percentuale di outliers che il metodo è in grado di sopportare senza perdere precisione nella stima
- complessità computazionale  $\Rightarrow$  la complessità di implementazione e di esecuzione del metodo



Per ottenere un punto di break-down alto, si possono usare degli stimatori robusti, che siano poco suscettibili alla presenza di outliers. La mediana è uno stimatore molto robusto perché mantiene precisione anche con il 50% di outliers. Altri stimatori robusti sono gli M-stimatori e gli R-stimatori.

M-stimatori  $\Rightarrow$  sostituiscono il quadrato dello scarto dello stimato ai minimi quadrati con una funzione dei residui che non è il quadrato. La funzione è simmetrica e si chiama funzione di perdita. Least-median-of-squares ha un punto di break-down del 50%, ma una pessima efficienza. I problemi vengono stimati risolvendo il problema delle minimizzazioni della mediana al quadrato.

Ransac assomiglia al metodo least-median-of-squares.

È importante che la casualizzazione iniziale non sia completamente lasciata al caso, ma deve scegliere i punti necessari a risolvere il sistema (come quantità e posizione)

Viene calcolato il consenso che c'è tra i dati per questo ipotesi di modello. Quindi viene controllato quanti dei punti compiono combaciano, a meno di una soglia, con il modello trovato.

Comparo la coordinata del consenso con una soglia e nel caso eseguo nuovamente l'algoritmo. Potrebbe non convergere in caso di soglie molto basse