# Stereovisione

Filippo L.M. Milotta

Image Processing Lab
Dipartimento di Matematica e Informatica
Università degli Studi di Catania

milotta@dmi.unict.it

26 marzo 2015

Dal greco στερεός, "solido": *stereovisione* $\approx$ "visione solida".

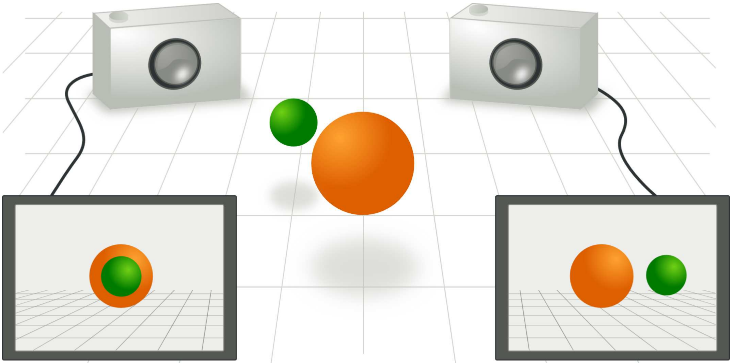
Dal greco στερεός, "solido": *stereovisione* $\approx$ "visione solida".

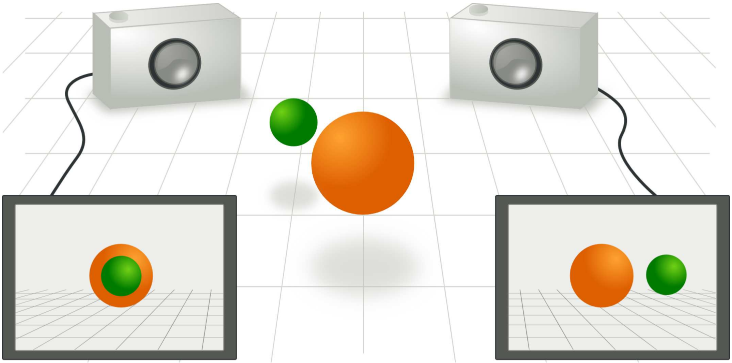
Una "visione solida" presuppone un meccanismo di percezione della profondità. Il più comune: fondere due proiezioni della medesima scena ottenute da due punti di vista differenti.

Dal greco στερεός, "solido": *stereovisione* ≈ "visione solida".

Una "visione solida" presuppone un meccanismo di percezione della profondità. Il più comune: fondere due proiezioni della medesima scena ottenute da due punti di vista differenti.

La "fusione" non è un compito banale, ma il nostro cervello è in grado di svolgerlo egregiamente.

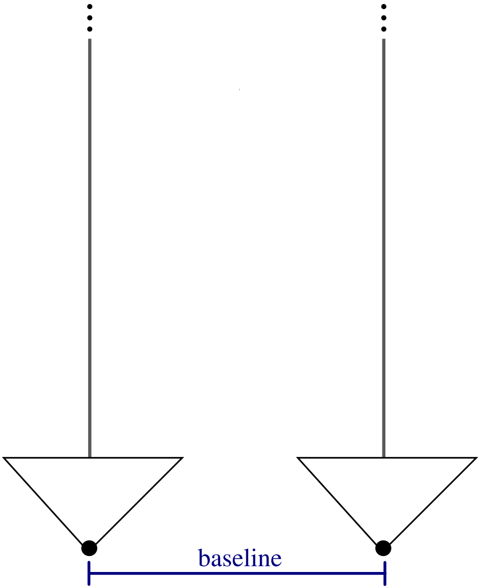




Ogni camera "vede" la scena in modo lievemente diverso dall'altra

Un *sistema stereo* è un sistema di visione consistende in (almeno) due camere che osservano la medesima scena.


centro di fissazione
baseline


baseline

In generale, il passaggio da un sistema di riferimento ad un altro può essere descritto come:

In generale, il passaggio da un sistema di riferimento ad un altro può essere descritto come:

$$X' = R(X - T)$$

In generale, il passaggio da un sistema di riferimento ad un altro può essere descritto come:

$$X' = R(X - T)$$

Tale relazione compare infatti sia nel Geometric Mapping che nell'Affine Mapping.

## Modelli di Movimento della Telecamera
## Modello a 4-parametri

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \rho \begin{bmatrix} \cos\theta_z & -\sin\theta_z \\ \sin\theta_z & \cos\theta_z \end{bmatrix} \begin{bmatrix} x + \theta_y F + t_x \\ y - \theta_x F + t_y \end{bmatrix} =$$
$$= \begin{bmatrix} c_1 & -c_2 \\ c_2 & c_1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} c_3 \\ c_4 \end{bmatrix}$$

- Questa formula è un caso particolare di **trasformazione affine** (*affine mapping*), normalmente a 6-parametri
- Prende il nome di **trasformazione geometrica** (*geometric mapping*), e caratterizza qualsiasi combinazione di scaling, rotazione e traslazione in 2D

## Modelli di Movimento della Telecamera Modello a 6-parametri (1)

- La rotazione di un oggetto attorno all'origine dello spazio 3D è data dalla matrice di rotazione

$$\boxed{[\boldsymbol{R}] = [\boldsymbol{R_x}] \cdot [\boldsymbol{R_y}] \cdot [\boldsymbol{R_z}]}$$

$$[R_x] = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\theta_x & -\sin\theta_x \\ 0 & \sin\theta_x & \cos\theta_x \end{bmatrix} \qquad [R_y] = \begin{bmatrix} \cos\theta_y & 0 & \sin\theta_y \\ 0 & 1 & 0 \\ -\sin\theta_y & 0 & \cos\theta_y \end{bmatrix}$$

$$[R_z] = \begin{bmatrix} \cos\theta_z & -\sin\theta_z & 0 \\ \sin\theta_z & \cos\theta_z & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

# Modelli di Movimento della Telecamera
## Modello a 6-parametri (2)

- Pertanto, combinando insieme $R_x, R_y, R_z$ si ottiene:

$$[R] = \begin{bmatrix} \cos\theta_y \cos\theta_z & \sin\theta_x \sin\theta_y \cos\theta_z - \cos\theta_x \sin\theta_z & \cos\theta_x \sin\theta_y \cos\theta_z + \sin\theta_x \sin\theta_z \\ \cos\theta_y \sin\theta_z & \sin\theta_x \sin\theta_y \sin\theta_z + \cos\theta_x \cos\theta_z & \cos\theta_x \sin\theta_y \sin\theta_z - \sin\theta_x \cos\theta_z \\ -\sin\theta_y & \sin\theta_x \cos\theta_y & \cos\theta_x \cos\theta_y \end{bmatrix}$$

- Assumendo piccole rotazioni si può porre:

$$\boxed{[\boldsymbol{R}] \approx [\boldsymbol{R}'] = \begin{bmatrix} \mathbf{1} & -\boldsymbol{\theta}_z & \boldsymbol{\theta}_y \\ \boldsymbol{\theta}_z & \mathbf{1} & -\boldsymbol{\theta}_x \\ -\boldsymbol{\theta}_y & \boldsymbol{\theta}_x & \mathbf{1} \end{bmatrix}}$$

## Modelli di Movimento della Telecamera Modello a 6-parametri (3)

- Il moto di un punto può essere descritto come:

$$X' = [R] \cdot X + [T]$$

$$\begin{bmatrix} X' \\ Y' \\ Z' \end{bmatrix} = \begin{bmatrix} r_1 & r_2 & r_3 \\ r_4 & r_5 & r_6 \\ r_7 & r_8 & r_9 \end{bmatrix} \cdot \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} + \begin{bmatrix} T_x \\ T_y \\ T_z \end{bmatrix}$$

- Per quanto visto, sebbene la matrice di rotazione $R$ ha 9 elementi, essa è determinata soltanto da **3** angoli di rotazione

## Modelli di Movimento della Telecamera Modello a 6-parametri (4)

- Infine, utilizzando le equazioni delle proiezioni prospettiche, si ottiene:

$$x' = F\frac{(r_1x + r_2y + r_3F)Z + T_xF}{(r_7x + r_8y + r_9F)Z + T_zF}$$

$$y' = F\frac{(r_4x + r_5y + r_6F)Z + T_yF}{(r_7x + r_8y + r_9F)Z + T_zF}$$

## Modelli di Movimento della Telecamera Modello a 6-parametri (5)

- Le relazioni che abbiamo ottenuto sono note come il **caso generale** del mapping di traslazioni e rotazioni arbitrarie da spazio 3D a spazio 2D
- Rispetto al *geometric mapping* (4-parametri) in questo **modello a 6-parametri** è possibile considerare anche **traslazioni lungo l'asse Z** e **rotazioni attorno ad assi arbitrari**

Parametri di un sistema stereo:

Parametri di un sistema stereo:

- **Parametri intrinseci**: i parametri intrinseci di entrambe la camere, come li abbiamo precedentemente definiti

Parametri di un sistema stereo:

- **Parametri intrinseci**: i parametri intrinseci di entrambe la camere, come li abbiamo precedentemente definiti
- **Parametri estrinseci**: Posizione e orientazione *relativa* delle camere

Siano $O_l$ e $O_r$ le coordinate dei fuochi della camera sinistra e di quella destra, e siano $R_l$ ed $R_r$ le matrici di rotazione che ruotano il sistema di riferimento del mondo in quello delle rispettive camere:

Siano $O_l$ e $O_r$ le coordinate dei fuochi della camera sinistra e di quella destra, e siano $R_l$ ed $R_r$ le matrici di rotazione che ruotano il sistema di riferimento del mondo in quello delle rispettive camere:

$$P_r = R_r(P_w - T_r)$$
$$P_l = R_l(P_w - T_l)$$

Siano $O_l$ e $O_r$ le coordinate dei fuochi della camera sinistra e di quella destra, e siano $R_l$ ed $R_r$ le matrici di rotazione che ruotano il sistema di riferimento del mondo in quello delle rispettive camere:

$$P_r = R_r(P_w - T_r)$$
$$P_l = R_l(P_w - T_l)$$

Cerchiamo i vettori $T$ ed $R$ del sistema stereo tali che la relazione tra le coordinate di un punto dello spazio nel sistema di riferimento della camera sinistra ($P_l$) e le coordinate dello stesso punto nel sistema di riferimento della camera destra ($P_r$) sia:

Siano $O_l$ e $O_r$ le coordinate dei fuochi della camera sinistra e di quella destra, e siano $R_l$ ed $R_r$ le matrici di rotazione che ruotano il sistema di riferimento del mondo in quello delle rispettive camere:

$$P_r = R_r(P_w - T_r)$$
$$P_l = R_l(P_w - T_l)$$

Cerchiamo i vettori $T$ ed $R$ del sistema stereo tali che la relazione tra le coordinate di un punto dello spazio nel sistema di riferimento della camera sinistra $(P_l)$ e le coordinate dello stesso punto nel sistema di riferimento della camera destra $(P_r)$ sia:

$$P_r = R(P_l - T)$$

$$P_r = R(P_l - T)$$

$$P_r = R(P_l - T)$$

---

Il vettore di traslazione che "sposta" un sistema di riferimento all'altro è banalmente $T = O_r - O_l$; manca però $R$.

$$P_r = R(P_l - T)$$

---

Il vettore di traslazione che “sposta” un sistema di riferimento all’altro è banalmente $T = O_r - O_l$; manca però $R$.

Abbiamo le formule per tradurre (1) $W \to C_r$ e (2) $W \to C_l$; per trovare la relazione $C_l \to C_r$ potremmo invertire la (2), ottenendo $C_l \to W$, e successivamente applicare la (1).

$$P_r = R(P_l - T)$$

$$P_r = R(P_l - T)$$

---

Invertiamo $W \to C_l$:

$$P_r = R(P_l - T)$$

---

Invertiamo $W \rightarrow C_l$:

$$P_l = R_l(P_w - T_l);$$

$$P_r = R(P_l - T)$$

---

Invertiamo $W \rightarrow C_l$:

$$P_l = R_l(P_w - T_l);$$
$$R_l^{-1} P_l = P_w - T_l;$$

$$P_r = R(P_l - T)$$

---

Invertiamo $W \rightarrow C_l$:

$$\begin{aligned} P_l &= R_l(P_w - T_l); \\ R_l^{-1} P_l &= P_w - T_l; \\ P_w &= R_l^{-1} P_l + T_l \end{aligned}$$

$$P_r = R(P_l - T)$$

---

$$P_w = R_l^{-1} P_l + T_l$$

$$P_r = R(P_l - T)$$

---

$$P_w = R_l^{-1} P_l + T_l$$

Ora inglobiamola in $W \to C_r$:

$$P_r = R(P_l - T)$$

---

$$P_w = R_l^{-1} P_l + T_l$$

Ora inglobiamola in $W \to C_r$:

$$P_r = R_r(P_w - T_r)$$

$$P_r = R(P_l - T)$$

---

$$P_w = R_l^{-1} P_l + T_l$$

Ora inglobiamola in $W \rightarrow C_r$:

$$\begin{aligned} P_r &= R_r(P_w - T_r) \\ &= R_r(R_l^{-1} P_l + T_l - T_r) \end{aligned}$$

$$P_r = R(P_l - T)$$

---

$$P_w = R_l^{-1} P_l + T_l$$

Ora inglobiamola in $W \rightarrow C_r$:

$$\begin{aligned} P_r &= R_r(P_w - T_r) \\ &= R_r(R_l^{-1} P_l + T_l - T_r) \\ &= R_r R_l^{-1} P_l + R_r T_l - R_r T_r \end{aligned}$$

$$P_r = R(P_l - T)$$

---

$$P_w = R_l^{-1} P_l + T_l$$

Ora inglobiamola in $W \to C_r$:

$$\begin{aligned} P_r &= R_r(P_w - T_r) \\ &= R_r(R_l^{-1} P_l + T_l - T_r) \\ &= R_r R_l^{-1} P_l + R_r T_l - R_r T_r \\ &= R_r R_l^{-1} P_l - R_r(T_r - T_l) \end{aligned}$$

$$P_r = R(P_l - T)$$

---

$$P_w = R_l^{-1} P_l + T_l$$

Ora inglobiamola in $W \rightarrow C_r$:

$$\begin{aligned} P_r &= R_r(P_w - T_r) \\ &= R_r(R_l^{-1} P_l + T_l - T_r) \\ &= R_r R_l^{-1} P_l + R_r T_l - R_r T_r \\ &= R_r R_l^{-1} P_l - R_r(T_r - T_l) \end{aligned}$$

...ci basta?

$$\begin{aligned} P_r = & \quad {\color{red}R}(P_l - {\color{red}T}) \\ = & \quad {\color{red}R}P_l - {\color{red}RT} \end{aligned}$$

---

$$P_r = R_r R_l^{-1} P_l - R_r(T_r - T_l)$$

$$\begin{aligned} P_r = & \quad R(P_l - T) \\ = & \quad RP_l - RT \end{aligned}$$

---

$$P_r = R_r R_l^{-1} P_l - R_r(T_r - T_l)$$

$$\begin{aligned} P_r &= R(P_l - T) \\ &= RP_l - RT \end{aligned}$$

---

$$P_r = \underbrace{R_r R_l^{-1}}_{R} P_l - \underbrace{R_r(T_r - T_l)}_{RT}$$

$$\begin{aligned} P_r = & \; R(P_l - T) \\ = & \; RP_l - RT \end{aligned}$$

---

$$P_r = \underbrace{R_r R_l^{-1}}_{R} P_l - \underbrace{R_r(T_r - T_l)}_{RT}$$

dunque $R = R_r R_l^{-1}$ e $T = R^{-1} R_r (T_r - T_l)$. $T$ lo conoscevamo già, ma espresso nel sistema di riferimento del mondo ($T_o$); la relazione col "nuovo" $T$ è $T_o = T_r - T_l = R_l^T T$.

I due problemi principali che un *sistema stereo* deve affrontare ai fini della ricostruzione sono il *matching* e la *depth estimation* (ricostruzione).

**Matching**: quali punti dell'immagine prodotta da una camera corrispondono a quali punti dell'altra camera?

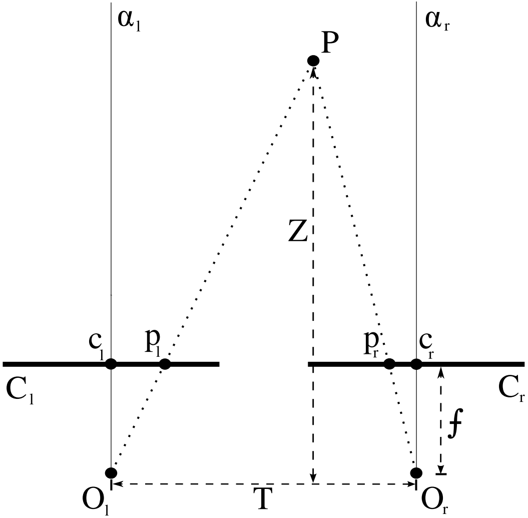
**Matching**: quali punti dell'immagine prodotta da una camera corrispondono a quali punti dell'altra camera?

**Ricostruzione**: trovata l'associazione, come faccio a ricostruire la posizione nello spazio di quello che vedo?

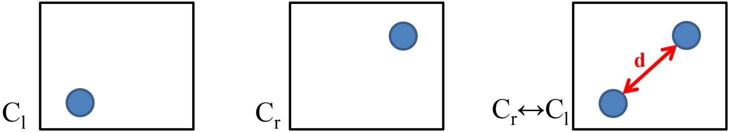
**Matching**: quali punti dell'immagine prodotta da una camera corrispondono a quali punti dell'altra camera?

**Ricostruzione**: trovata l'associazione, come faccio a ricostruire la posizione nello spazio di quello che vedo?

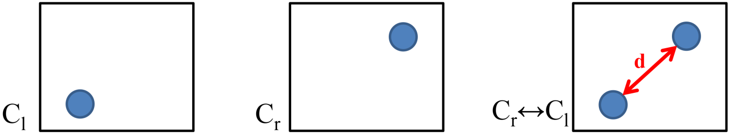
Vediamo un assaggio di ricostruzione (definendo la *disparità*), poi andiamo al problema del matching.




*P* viene proiettato in posizioni differenti sui due piani immagine; chiamiamo questa differenza *disparità*.




$P$ viene proiettato in posizioni differenti sui due piani immagine; chiamiamo questa differenza *disparità*.



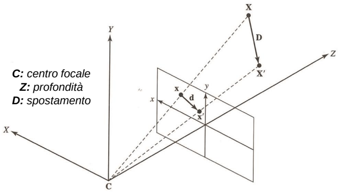

$P$ viene proiettato in posizioni differenti sui due piani immagine; chiamiamo questa differenza *disparità*.
Nel sistema stereo preso in considerazione la disparità sarà data da:

$$d(x) = x_r - x_l$$

Nota: è comune considerare gli assi $X$ delle camere crescenti *a sinistra*.

In maniera analoga alla disparità, nella stima del movimento e degli spostamenti in 2D e 3D si può definire il *displacement*.

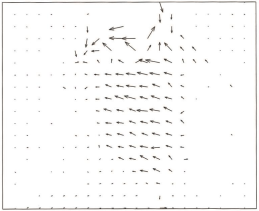
# Spostamenti in 3D e in 2D




- Siano:
  - Posizione iniziale a $t_1$:
    - $\boldsymbol{X} = [X, Y, Z]^T$
    - $\boldsymbol{x} = [x, y]^T$
  - Posizione finale a $t_2$:
    - $\boldsymbol{X}' = [X', Y', Z']^T = [X + D_X, Y + D_Y, Z + D_Z]^T$
    - $\boldsymbol{x}' = [x', y']^T = [x + d_x, y + d_y]^T$
- Spostamento 3D:
  - $\boldsymbol{D}(\boldsymbol{X}; t_1, t_2) = \boldsymbol{X}' - \boldsymbol{X} = [D_X, D_Y, D_Z]^T$
- Spostamento 2D:
  - $\boldsymbol{d}(\boldsymbol{x}; t_1, t_2) = \boldsymbol{x}' - \boldsymbol{x} = [d_x, d_y]^T$

# Modelli di Movimento 2D

- Se $t_1$ e $t_2$ sono chiari si possono omettere e si può scrivere $\boldsymbol{d}(\boldsymbol{x})$
- $\boldsymbol{d}(\boldsymbol{x}) = \boldsymbol{x}' - \boldsymbol{x}$ è detto "*displacement*", "Vettore di movimento 2D" o "*motion vector*" (**MV**)
- L'insieme dei $\boldsymbol{d}(\boldsymbol{x})$ per ogni $x$ nell'immagine è detto "Campo di movimento" (*motion field*)



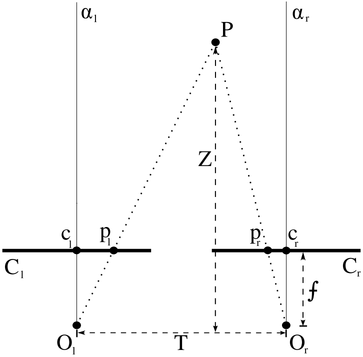
***Esempio di motion field***

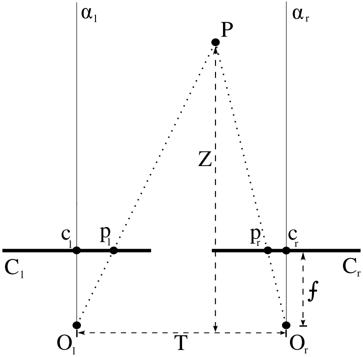
- **Funzione di mapping**: $\boldsymbol{w}(\boldsymbol{x}) = \boldsymbol{x} + \boldsymbol{d}(\boldsymbol{x})$
  - Ogni pixel contiene un valore diverso da quello iniziale; si noti che, in generale, $\boldsymbol{d}(\boldsymbol{x})$ potrebbe non essere noto

Il displacement è calcolato su frame dello stesso video, acquisiti cioè dallo stesso dispositivo e quindi da un unico punto di vista. Il displacement è una misura di scostamento temporale, e viene utilizzato per ricostruire l'informazione sul movimento nella scena.

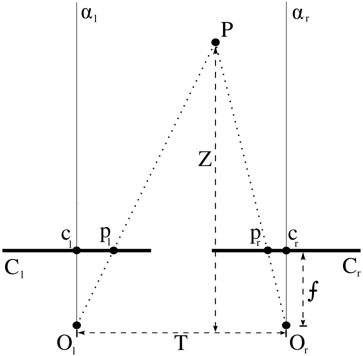
Il displacement è calcolato su frame dello stesso video, acquisiti cioè dallo stesso dispositivo e quindi da un unico punto di vista. Il displacement è una misura di scostamento temporale, e viene utilizzato per ricostruire l'informazione sul movimento nella scena.

Nel caso del sistema stereoscopico (con almeno due camere), i piani delle immagini sono acquisiti da punti di vista differenti, ma allo stesso istante di tempo. Pertanto la disparità è una misura di scostamento spaziale (relativo alle proiezioni sui piani delle immagini) e viene utilizzato per ricostruire l'informazione sulla profondità della scena.


α l
α r
P
Z
c l
p l
p r
c r
C l
C r
f
O l
T
O r

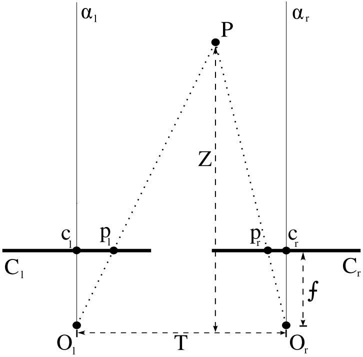



Cerchiamo la relazione tra $p_l$, $p_r$ e $Z$.




Cerchiamo la relazione tra $p_l$, $p_r$ e $Z$.

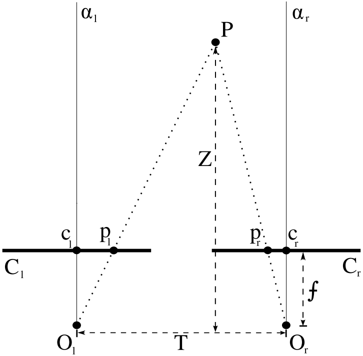
Poiché i triangoli $\overset{\triangle}{P_l P P_r}$ e $\overset{\triangle}{O_l P O_r}$ sono simili, il rapporto base/altezza sarà uguale; dunque



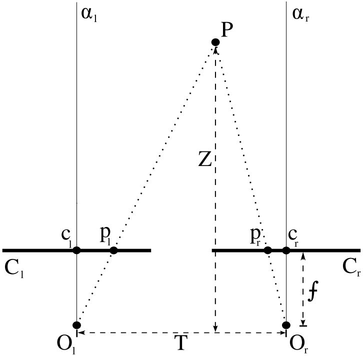

Cerchiamo la relazione tra $p_l$, $p_r$ e $Z$.

Poiché i triangoli $\overset{\triangle}{P_l P P_r}$ e $\overset{\triangle}{O_l P O_r}$ sono simili, il rapporto base/altezza sarà uguale; dunque

$$\frac{T}{Z} = \frac{T + x_l - x_r}{Z - f}$$

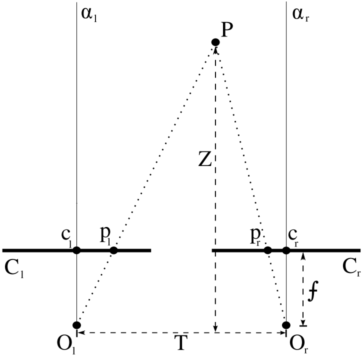



$$\frac{T}{Z} = \frac{T + x_l - x_r}{Z - f}$$




$$\frac{T}{Z} = \frac{T + x_l - x_r}{Z - f}$$

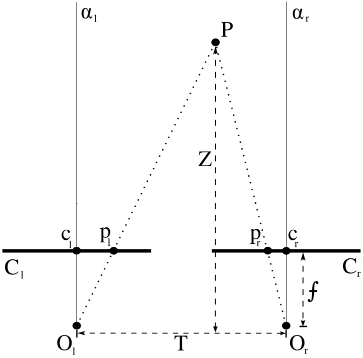
Chiamiamo *disparità* la differenza $d = x_r - x_l$, e risolviamo per Z:




$$\frac{T}{Z} = \frac{T + x_l - x_r}{Z - f}$$

Chiamiamo *disparità* la differenza $d = x_r - x_l$, e risolviamo per Z:

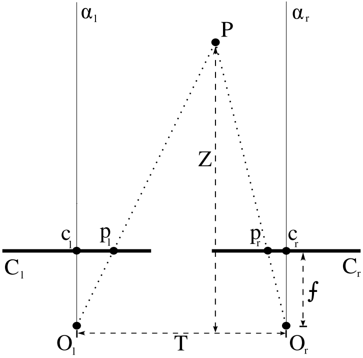
$$\frac{T}{Z} = \frac{T - d}{Z - f} \quad \Rightarrow$$




$$\frac{T}{Z} = \frac{T + x_l - x_r}{Z - f}$$

Chiamiamo *disparità* la differenza $d = x_r - x_l$, e risolviamo per Z:

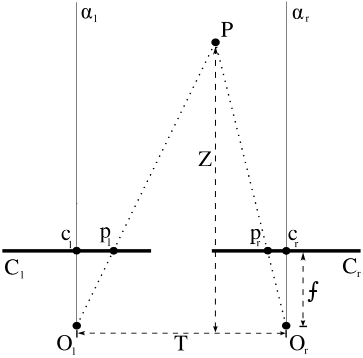
$$\begin{aligned} \frac{T}{Z} &= \frac{T - d}{Z - f} \Rightarrow \\ (Z - f)T &= Z(T - d) \Rightarrow \end{aligned}$$




$$\frac{T}{Z} = \frac{T + x_l - x_r}{Z - f}$$

Chiamiamo *disparità* la differenza $d = x_r - x_l$, e risolviamo per Z:

$$\begin{aligned} \frac{T}{Z} &= \frac{T-d}{Z-f} \Rightarrow \\ (Z-f)T &= Z(T-d) \Rightarrow \\ TZ - Tf &= TZ - Zd \Rightarrow \end{aligned}$$




$$\frac{T}{Z} = \frac{T + x_l - x_r}{Z - f}$$

Chiamiamo *disparità* la differenza $d = x_r - x_l$, e risolviamo per Z:

$$\begin{aligned}
\frac{T}{Z} &= \frac{T-d}{Z-f} \Rightarrow \\
(Z-f)T &= Z(T-d) \Rightarrow \\
TZ - Tf &= TZ - Zd \Rightarrow \\
\color{red}{Z} &\color{red}{=} \color{red}{\frac{Tf}{d}}
\end{aligned}$$

$$Z = \frac{TF}{d}$$

$$Z = \frac{TF}{d}$$

Questa relazione ha delle importanti conseguenze:

$$Z = \frac{TF}{d}$$

Questa relazione ha delle importanti conseguenze:

- La *profondità* di un punto $P$ è *inversamente proporzionale alla disparità*;

$$Z = \frac{TF}{d}$$

Questa relazione ha delle importanti conseguenze:

- La *profondità* di un punto $P$ è *inversamente proporzionale alla disparità*;
- La relazione tra $Z$ e $d$ non è lineare;

$$Z = \frac{TF}{d}$$

Questa relazione ha delle importanti conseguenze:

- La *profondità* di un punto $P$ è *inversamente proporzionale alla disparità*;
- La relazione tra $Z$ e $d$ non è lineare;
- Fissate la lunghezza focale e la baseline, la profondità di un punto dipende *solo* dalla disparità;

$$Z = \frac{TF}{d}$$

Questa relazione ha delle importanti conseguenze:

- La *profondità* di un punto $P$ è *inversamente proporzionale alla disparità*;
- La relazione tra $Z$ e $d$ non è lineare;
- Fissate la lunghezza focale e la baseline, la profondità di un punto dipende *solo* dalla disparità;
- Errori nella stima della disparità, in particolare quando essa è molto piccola, si riflettono in grandi errori di stima della profondità.

$$Z = \frac{TF}{d}$$

La relazione è stata calcolata considerando un sistema con punto di fissazione all'infinito; nel caso di un sistema con punto di fissazione "vicino", la disparità è inversamente proporzionale alla distanza dal punto di fissazione.

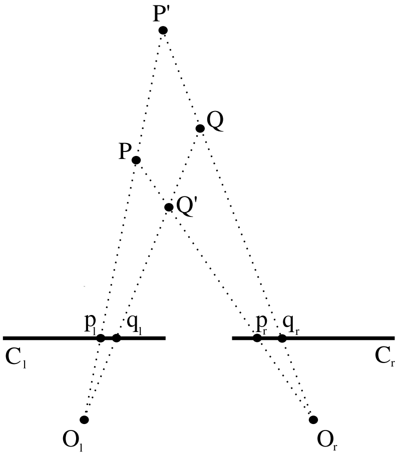
$$Z = \frac{TF}{d}$$

La relazione è stata calcolata considerando un sistema con punto di fissazione all'infinito; nel caso di un sistema con punto di fissazione "vicino", la disparità è inversamente proporzionale alla distanza dal punto di fissazione.

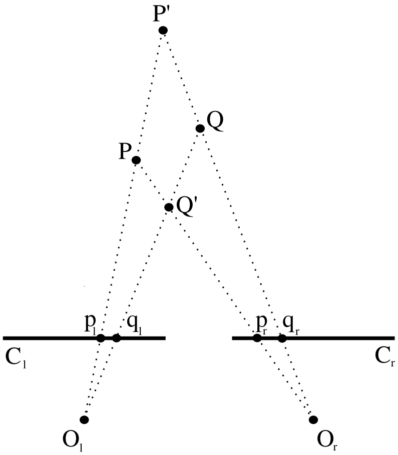
La disparità è anche legata al cosiddetto *effetto parallasse*: oggetti che si muovono con la stessa velocità appaiono tanto più lenti quanto più sono lontani, proprio perché la disparità tra punti (in questo caso, tra le proiezioni di uno stesso punto in due istanti di tempo separati) è inversamente proporzionale alla distanza.

# Perché è importante associare correttamente i punti?

## Perché è importante associare correttamente i punti?



Perché è importante associare correttamente i punti?




Caso semplicissimo: solo due punti. Eppure, se sbaglio ad associare $P$ e $Q$ alle rispettive proiezioni...

Vi sono due principali tecniche di matching:

Vi sono due principali tecniche di matching:

- Per *correlazione* (dominio dei pixels)

Vi sono due principali tecniche di matching:

- Per *correlazione* (dominio dei pixels)
- Per *features* (dominio delle features)

La correlazione è un *indice statistico di dipendenza lineare*.

La correlazione è un *indice statistico di dipendenza lineare*.

La correlazione $\mathrm{corr}(d)$ tra il vettore $B$ e la finestra di $A$ di offset $d$ (ovvero, la sottosequenza di $A$ di lunghezza pari a quella di $B$ che inizia alla posizione $d$) è data da:

La correlazione è un *indice statistico di dipendenza lineare*.

La correlazione $\mathrm{corr}(d)$ tra il vettore $B$ e la finestra di $A$ di offset $d$ (ovvero, la sottosequenza di $A$ di lunghezza pari a quella di $B$ che inizia alla posizione $d$) è data da:

$$\mathrm{corr}(d) = \sum_{i=0}^{\mathrm{size}(B)} f(A[d+i], B[i])$$

La correlazione è un *indice statistico di dipendenza lineare*.

La correlazione $\mathrm{corr}(d)$ tra il vettore $B$ e la finestra di $A$ di offset $d$ (ovvero, la sottosequenza di $A$ di lunghezza pari a quella di $B$ che inizia alla posizione $d$) è data da:

$$\mathrm{corr}(d) = \sum_{i=0}^{\mathrm{size}(B)} f(A[d+i], B[i])$$

dove $f$ è una funzione opportunamente scelta. Scelte comuni per $f$ sono

La correlazione è un *indice statistico di dipendenza lineare*.

La correlazione $\mathrm{corr}(d)$ tra il vettore $B$ e la finestra di $A$ di offset $d$ (ovvero, la sottosequenza di $A$ di lunghezza pari a quella di $B$ che inizia alla posizione $d$) è data da:

$$\mathrm{corr}(d) = \sum_{i=0}^{\mathrm{size}(B)} f(A[d+i], B[i])$$

dove $f$ è una funzione opportunamente scelta. Scelte comuni per $f$ sono

- $f(x, y) = xy$ (*correlazione incrociata* o *cross-correlation*)

La correlazione è un *indice statistico di dipendenza lineare*.

La correlazione $\mathrm{corr}(d)$ tra il vettore $B$ e la finestra di $A$ di offset $d$ (ovvero, la sottosequenza di $A$ di lunghezza pari a quella di $B$ che inizia alla posizione $d$) è data da:

$$\mathrm{corr}(d) = \sum_{i=0}^{\mathrm{size}(B)} f(A[d+i], B[i])$$

dove $f$ è una funzione opportunamente scelta. Scelte comuni per $f$ sono

- $f(x, y) = xy$ (*correlazione incrociata* o *cross-correlation*)
- $f(x, y) = -(x-y)^2$ (*block matching*)

Nella stima del movimento, *Displaced Frame Difference (DFD)* è un possibile indice di correlazione. Esso era definito tramite l'errore $E_{DFD}$:

Nella stima del movimento, *Displaced Frame Difference (DFD)* è un possibile indice di correlazione. Esso era definito tramite l'errore $E_{DFD}$:

## Criterio di Stima del Movimento Displaced Frame Difference (DFD)

- Siano inoltre:
  - Funzione di mapping: $w(x;a) = x + d(x;a)$
    - dove $a = [a_1, a_2, \dots, a_L]^T$ parametri di movimento (<u>da stimare</u>)
- Definiamo **Errore DFD**:
  - $E_{DFD}(a) = \sum_{x \epsilon \Lambda} \left| I_2\big(w(x;a)\big) - I_1(x) \right|^p$
    - Dove:
      - $\Lambda$ : insieme dei pixel nell'anchor frame $I_1$
      - $p$ : intero positivo; casi particolari:
        - $p = 1 \rightarrow E_{DFD}$ è detto ***mean absolute difference*** (**MAD**)
        - $p = 2 \rightarrow E_{DFD}$ è detto ***mean squared error*** (**MSE**)

$$\text{corr}(d) = \sum_{i=0}^{\text{size}(B)} f(A[d+i], B[i])$$

$$\text{corr}(d) = \sum_{i=0}^{\text{size}(B)} f(A[d+i], B[i])$$

In caso di vettori a due dimensioni, l'unica modifica da fare è sdoppiare l'offset in due componenti ($d_x$ e $d_y$) ed effettuare la sommatoria al variare di queste due. Ponendo per semplicità $A \in \mathbb{R}^{M \times M}$ e $B \in \mathbb{R}^{N \times N}$, scriviamo:

$$\mathrm{corr}(d) = \sum_{i=0}^{\mathrm{size}(B)} f(A[d+i], B[i])$$

In caso di vettori a due dimensioni, l'unica modifica da fare è sdoppiare l'offset in due componenti ($d_x$ e $d_y$) ed effettuare la sommatoria al variare di queste due. Ponendo per semplicità $A \in \mathbb{R}^{M \times M}$ e $B \in \mathbb{R}^{N \times N}$, scriviamo:

$$\mathrm{corr}(d_x, d_y) = \sum_{i=0}^{N} \sum_{j=0}^{N} f(A[d_x + i, d_y + j], B[i, j])$$

Piccola ottimizzazione pratica: invece di spostare la finestra di ricerca in tutta l'immagine di una camera, iniziamo nella stessa posizione dove abbiamo selezionato la finestra dell'altra camera (vale se le camere non hanno punti di vista troppo differenti).

Piccola ottimizzazione pratica: invece di spostare la finestra di ricerca in tutta l'immagine di una camera, iniziamo nella stessa posizione dove abbiamo selezionato la finestra dell'altra camera (vale se le camere non hanno punti di vista troppo differenti).

Grazie alla correlazione, le immagini sono "dense" di corrispondenze; possiamo calcolare la disparità praticamente per ogni punto. Otteniamo una *disparity map*.

Vediamo un esempio pratico con Matlab
*funzione disparity(...)*

# Vediamo un esempio pratico con Matlab

*funzione disparity(...)*

(a) Vista sinistra

# Vediamo un esempio pratico con Matlab
*funzione disparity(...)*

(a) Vista sinistra

(b) Vista destra

# Vediamo un esempio pratico con Matlab
*funzione disparity(...)*

(a) Vista sinistra

(b) Vista destra

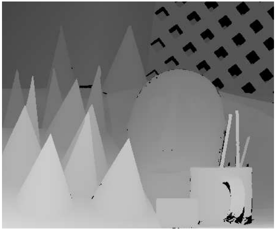


Figura: Mappa di disparità

Matching tramite features: operiamo in un dominio differente, quello delle caratteristiche di una immagine.

Matching tramite features: operiamo in un dominio differente, quello delle caratteristiche di una immagine.

Queste caratteristiche devono prima essere definite: possono essere bordi, angoli, forme geometriche, proprietà statistiche...

Matching tramite features: operiamo in un dominio differente, quello delle caratteristiche di una immagine.

Queste caratteristiche devono prima essere definite: possono essere bordi, angoli, forme geometriche, proprietà statistiche...

L'algoritmo di rilevazione di una feature dipende dal tipo di feature!

Matching tramite features: operiamo in un dominio differente, quello delle caratteristiche di una immagine.

Queste caratteristiche devono prima essere definite: possono essere bordi, angoli, forme geometriche, proprietà statistiche...

L'algoritmo di rilevazione di una feature dipende dal tipo di feature!
Questo significa che prima ancora di tentare il matching fra features è necessario definire un algoritmo di estrazione delle features.

Esempio di estrazione di features (corner) nell'algoritmo FAST:

## Stima del Movimento
### Algoritmi basati su Features

- Algoritmo "***F**eatures from **A**ccelerated **S**egment **T**est* (**FAST**)":
  - Individuare dei punti salienti (**corner**) nel frame da usare come features per "tracciare" il movimento
    - Si analizza ad ogni passo un insieme di 16 punti con configurazione a cerchio di raggio 3 per classificare se un punto $p$ sia o meno di corner
    - Data l'intensità $l_p$ del punto $p$, se per almeno 12 punti contigui con intensità $l_x$ si ottiene che $|l_x - l_p| > t$, con $t$ soglia, allora $p$ sarà considerato corner

Supponiamo, in un caso più semplice di FAST, che un algoritmo di estrazione delle features restituisca delle features elementari: segmenti. Come distinguiamo un segmento dagli altri?

Supponiamo, in un caso più semplice di FAST, che un algoritmo di estrazione delle features restituisca delle features elementari: segmenti. Come distinguiamo un segmento dagli altri?

- Coordinate del punto medio $M = (m_x, m_y)$;

Supponiamo, in un caso più semplice di FAST, che un algoritmo di estrazione delle features restituisca delle features elementari: segmenti. Come distinguiamo un segmento dagli altri?

- Coordinate del punto medio $M = (m_x, m_y)$;
- Lunghezza $l$;

Supponiamo, in un caso più semplice di FAST, che un algoritmo di estrazione delle features restituisca delle features elementari: segmenti. Come distinguiamo un segmento dagli altri?

- Coordinate del punto medio $M = (m_x, m_y)$;
- Lunghezza $l$;
- Angolo di orientazione $\theta$;

Supponiamo, in un caso più semplice di FAST, che un algoritmo di estrazione delle features restituisca delle features elementari: segmenti. Come distinguiamo un segmento dagli altri?

- Coordinate del punto medio $M = (m_x, m_y)$;
- Lunghezza $l$;
- Angolo di orientazione $\theta$;
- Stima del contrasto medio $C$ lungo il segmento.

Procedura:

Procedura:

(1) Diamo le immagini delle camere destra e sinistra in pasto ad un *feature detector*, ovvero un algoritmo in grado di rilevare le features e tutti i relativi parametri

Procedura:

(1) Diamo le immagini delle camere destra e sinistra in pasto ad un *feature detector*, ovvero un algoritmo in grado di rilevare le features e tutti i relativi parametri

(2) Cerchiamo corrispondenze tra i parametri (ci serve una *metrica* per misurarne la similarità)

Una metrica generica di similarità potrebbe essere:

$$S = \frac{1}{w_0(I_l - I_r)^2 + w_1(\theta_l - \theta_r)^2 + w_2(M_l - M_r)^2 + w_3(c_l - c_r)^2}$$

dove $w_i$ è il peso che intendiamo dare ad ogni parametro.

Altra feature, molto usata in computer vision: SIFT - (Wikipedia)