



Technical University of Cluj - Napoca
Computer Science Department

Procesarea Imaginilor

Curs 11

Extragerea informației 3D prin stereoviziune



Principiile Stereoviziunii



- Pentru observarea lumii reale avem nevoie de **informație 3D**
- Într-o imagine avem doar **două dimensiuni** – lipsește informația de profunzime
- Există mai multe metode pentru estimarea distanței: transformata perspectivă inversă, raționarea pe baza dimensiunii aparente a obiectelor în imagine, folosirea informației de focalizare, dar acestea sunt doar **aproximații pentru cazuri particulare**.



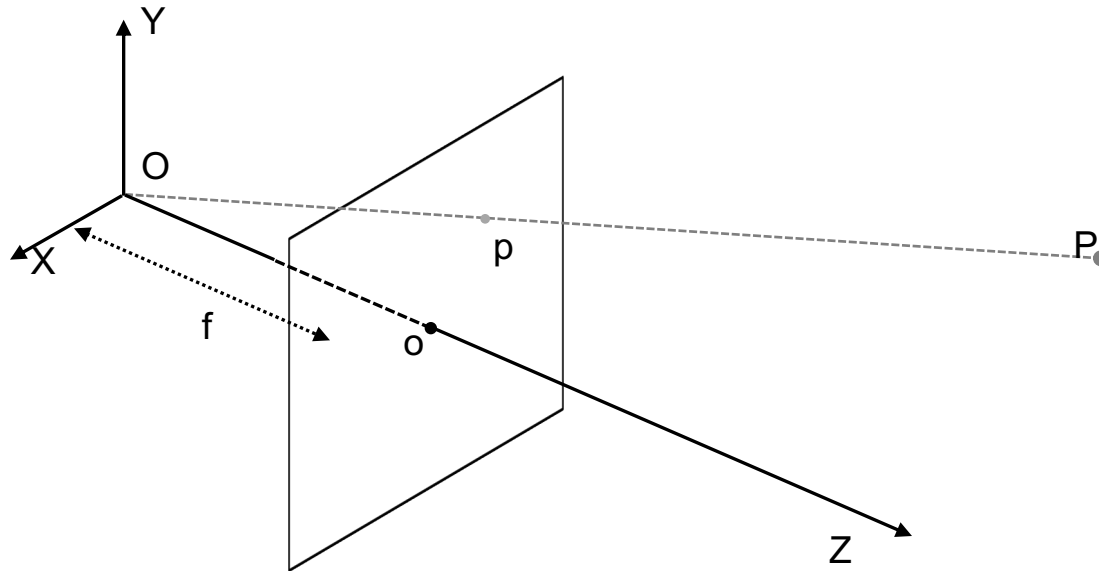
Principiile Stereoviziunii



- Stereoviziunea: observarea unei scene folosind două sau mai multe camere video, plasate în puncte diferite în spațiu.
- Cerințe:
 - Camerele trebuie declanșate în același timp. În caz contrar, dacă scena este dinamică cele două camere nu vor observa același lucru.
 - Caracteristicile interne ale camerelor trebuie cunoscute – **Calibrarea parametrilor intrinseci**
 - Poziția și orientarea camerelor, raportată la un sistem de coordonate unic, trebuie cunoscute cu mare precizie – **Calibrarea parametrilor extrinseci**



Principiile Stereoviziunii

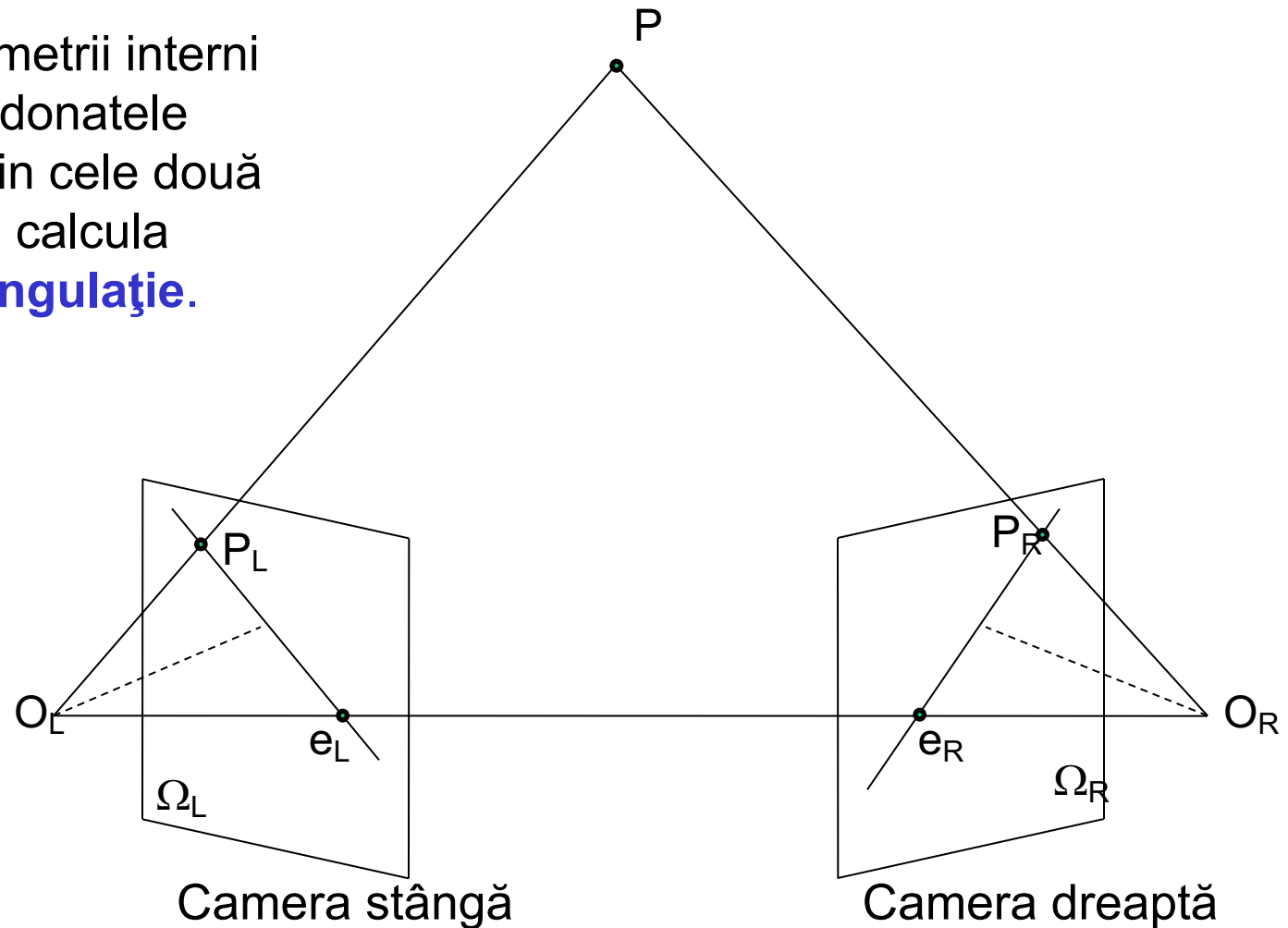


- Modelul și parametrii interni ai camerei video – permit stabilirea unei relații dintre un punct 3D **P**, exprimat într-un sistem de coordonate propriu camerei, și un pixel **p** din imagine.



Principiile Stereoviziunii

Cunoscând parametrii interni și externi, și coordonatele corespondente din cele două imagini, se poate calcula punctul 3D – **triangulație**.





Principiile Stereoviziunii

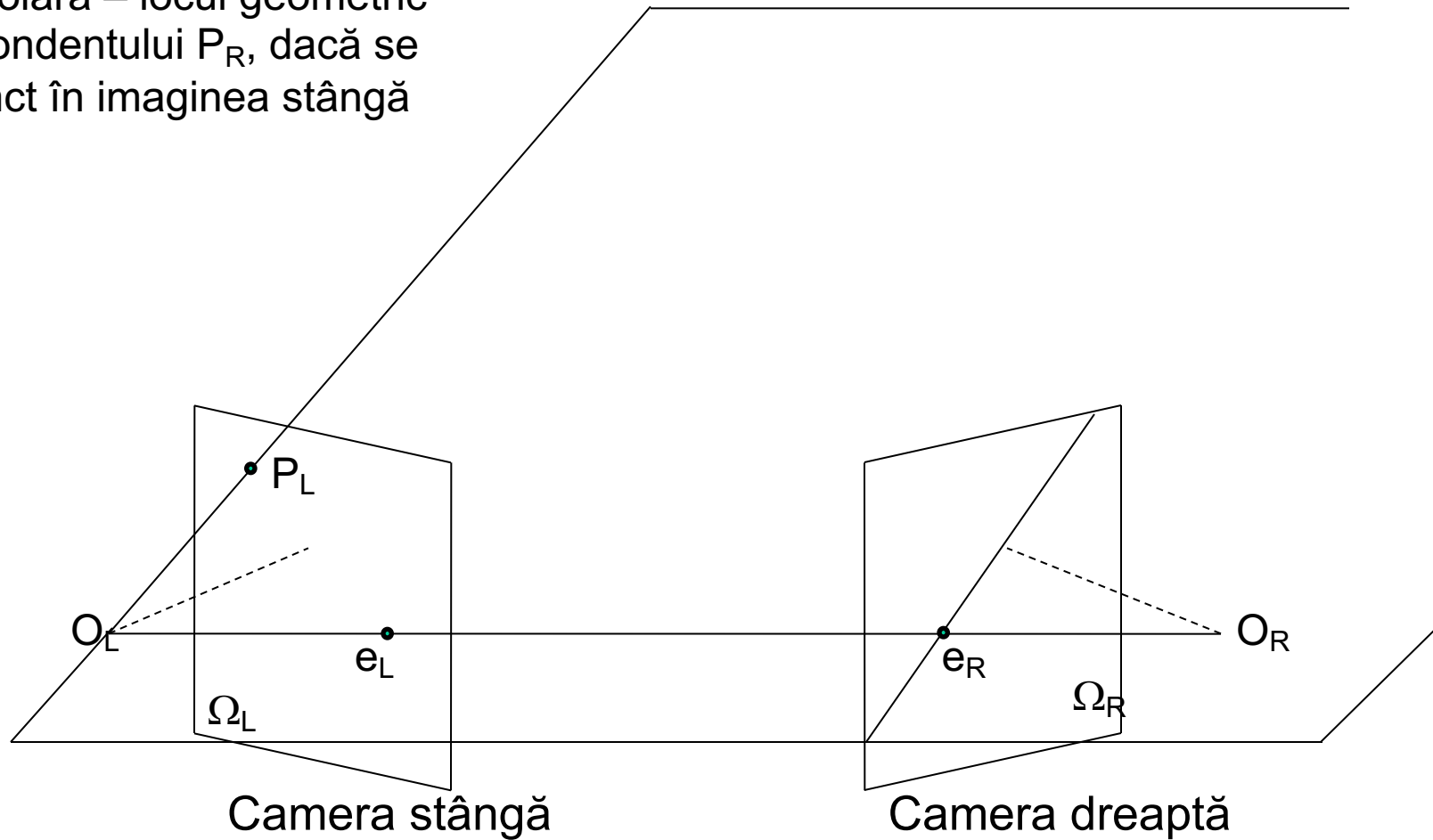


- **Problema corespondenței**: pentru un punct oarecare din imaginea dreaptă, care este corespondentul din imaginea stângă ?
- Unde căutăm punctul corespondent ? În toată imaginea ?



Principiile Stereoviziunii

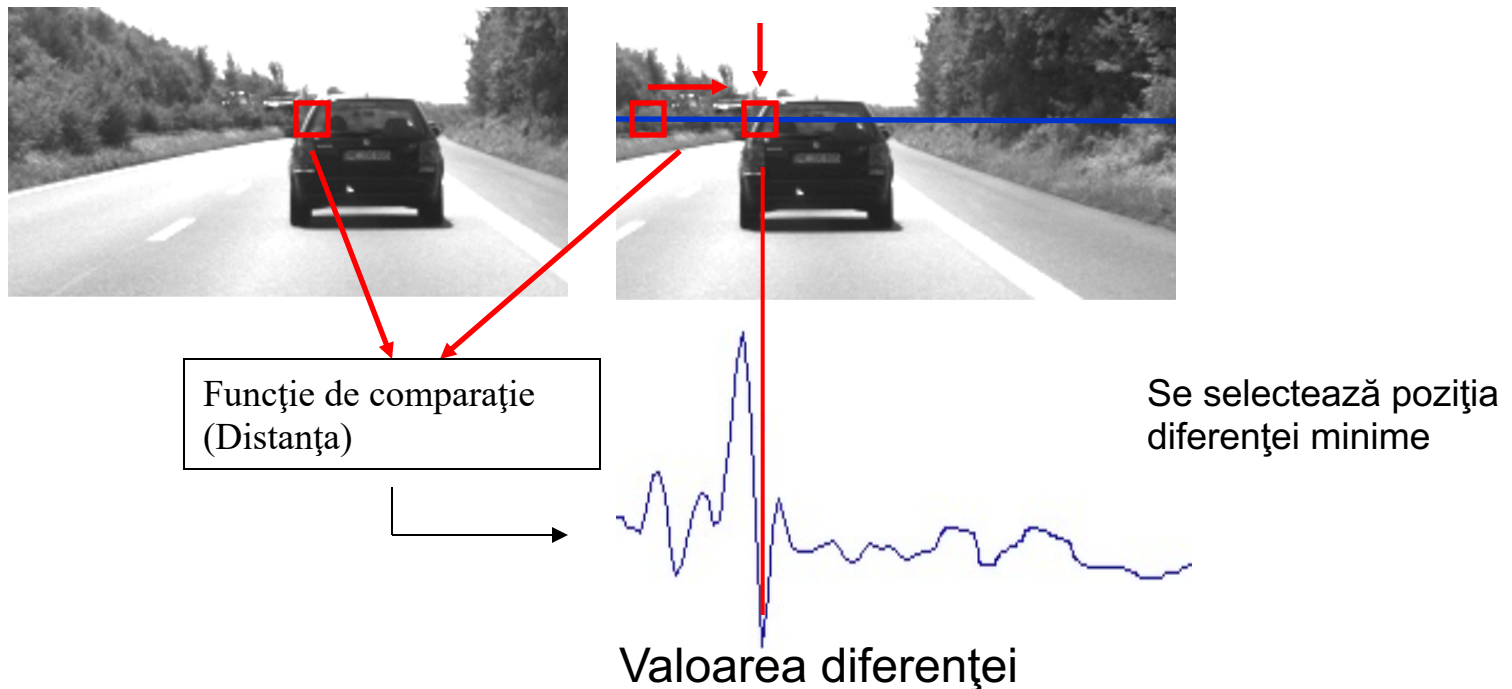
Linia epipolară – locul geometric al corespondentului P_R , dacă se dă un punct în imaginea stângă P_L .





Principiile Stereoviziunii

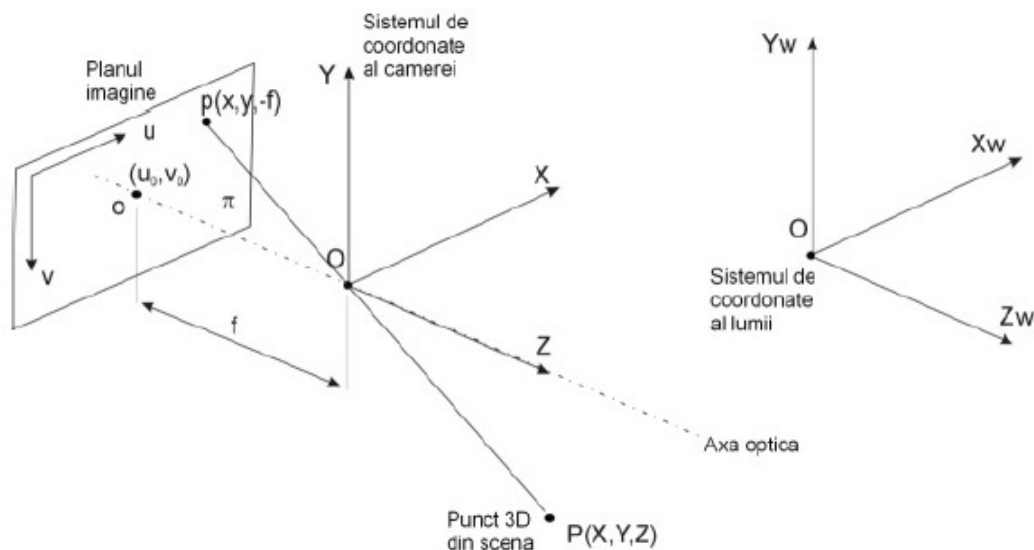
- Căutarea corespondențelor – se compară regiuni din jurul punctelor candidate cu regiunea centrată în punctul din imaginea stângă
- Există mai multe metrice de distanță: suma diferențelor absolute (Manhattan), distanța Euclidiană, funcția Census, etc.





Modelul camerei - recapitulare

Apertura / diafragma sunt approximate cu un punct (centrul de proiecție) \Rightarrow principiul coliniarității modelului cu lentile subțiri (constrângerea 3)



Ecuatiile fundamentale ale modelului perspectivă al camerei sunt [Trucco1998]:

$$\begin{cases} x = f \cdot \frac{X_C}{Z_C} \\ y = f \cdot \frac{Y_C}{Z_C} \end{cases}$$

Unde: $[X_C, Y_C, Z_C]$ sunt coordonatele punctului **P** în sistemul de coordonate al camerei

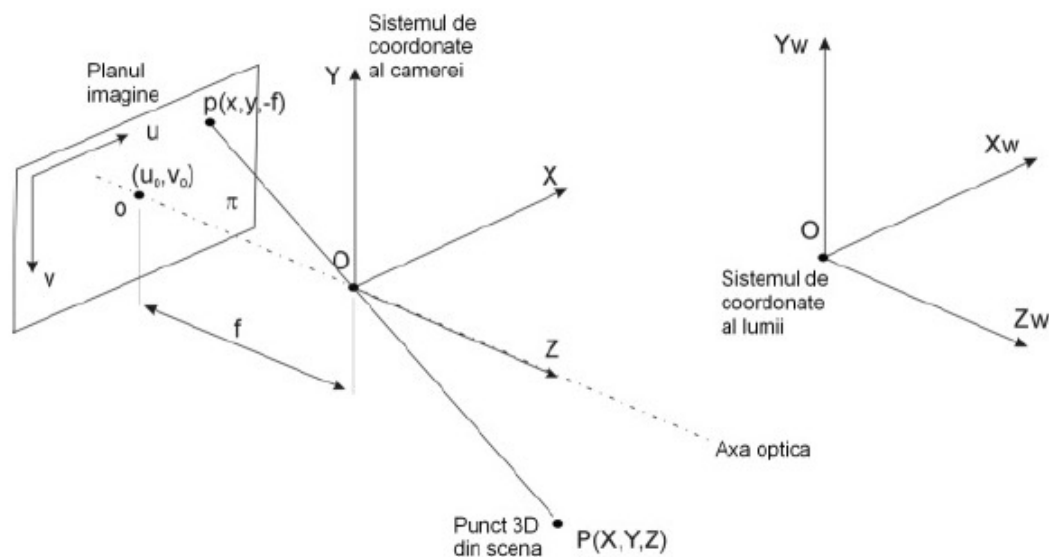
$[x, y, -f]$ sunt coordonatele punctului **p** în sistemul de coordonate al camerei



Modelul camerei – parametrii intrinseci

Descriu caracteristicile optice și geometrice ale camerei (caracteristicile interne ale camerei):

- *distanța focală* – distanța dintre centrul optic al sistemului de lentile al obiectivului și planul imagine: f ;
- *punctul principal* – coordonatele centrului real al imaginii (intersecția axei optice a obiectivului cu planul imagine) în coordonate imagine (pixeli) : (u_0, v_0) ;
- *coeficienții de distorsiune radială și tangențială* ai lentilelor obiectivului.





Modelul camerei – parametrii extrinseci

Poziția și orientarea sistemului de coordonate 3-D asociat camerei relativa la sistemul de coordonate asociat lumii

Necesari pentru a transforma coordonatele unui punct din sistemul de coordonate al lumii in sistemul de coordonate al camerei și invers)

Vectorul de translație $\mathbf{T} = [T_x, T_y, T_z]^T$;

Vectorul de rotație $\mathbf{r} = [R_x, R_y, R_z]^T$ sau *matricea de rotație* echivalentă $\mathbf{R} = \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{bmatrix}$

În sistemele stereo cu mai multe camere, parametrii extrinseci ne permit de asemenea să exprimăm și relația spațială dintre aceste camere.



Proiecția unui punct 3D în imagine

Matricea de proiecție:

$$\mathbf{P} = \mathbf{A} \cdot [\mathbf{R}_{WC} \mid \mathbf{T}_{WC}]$$

Punct 3D din lume de coordonate $[X_W, Y_W, Z_W]$

$$s \cdot \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} x_s \\ y_s \\ z_s \end{bmatrix} = \mathbf{P} \cdot \begin{bmatrix} X_W \\ Y_W \\ Z_W \\ 1 \end{bmatrix}$$

Obținerea coordonatelor imagine 2D:

$$\begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} x_s / z_s \\ y_s / z_s \end{bmatrix}$$



Reconstrucția 3D din coordonate 2D

Având coordonatele x_L, y_L în imagine,
dorim să obținem:

$$\mathbf{P}_W = \begin{bmatrix} X_W \\ Y_W \\ Z_W \end{bmatrix}$$

Solutia:

Parametrii intrinseci:

Distanțe focale

- Left camera: f_L .
- Right camera: f_R .

Puncte principale

- Left camera: $C'_L = (x_{CL}, y_{CL})$.
- Right camera: $C'_R = (x_{CR}, y_{CR})$.

Se cunosc:

Parametrii extrinseci

$$\mathbf{T}_{CL} = \begin{bmatrix} X_{CL} \\ Y_{CL} \\ Z_{CL} \end{bmatrix}$$

$$\mathbf{R}_{CL} = \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{bmatrix}$$

$$\mathbf{T}_{CR} = \begin{bmatrix} X_{CR} \\ Y_{CR} \\ Z_{CR} \end{bmatrix}$$

$$\mathbf{R}_{CR} = \begin{bmatrix} r'_{11} & r'_{12} & r'_{13} \\ r'_{21} & r'_{22} & r'_{23} \\ r'_{31} & r'_{32} & r'_{33} \end{bmatrix}$$

$$\mathbf{P}_W = \mu \mathbf{R}_L * \begin{bmatrix} x_L - x_{CL} \\ y_L - y_{CL} \\ -f_L \end{bmatrix} + \mathbf{T}_{CL} \Leftrightarrow \begin{bmatrix} X_W \\ Y_W \\ Z_W \end{bmatrix} = \mu \mathbf{R}_L * \begin{bmatrix} x_L - x_{CL} \\ y_L - y_{CL} \\ -f_L \end{bmatrix} + \begin{bmatrix} X_{CL} \\ Y_{CL} \\ Z_{CL} \end{bmatrix}, \quad (1)$$

unde μ este un factor de scalare dependent de Z

$$\begin{bmatrix} x_L - x_{CL} \\ y_L - y_{CL} \\ -f_L \end{bmatrix} = \mu^{-1} \mathbf{R}_L^T * \begin{bmatrix} X_W - X_{CL} \\ Y_W - Y_{CL} \\ Z_W - Z_{CL} \end{bmatrix} \Leftrightarrow \begin{bmatrix} x_L - x_{CL} \\ y_L - y_{CL} \\ -f_L \end{bmatrix} = \mu^{-1} \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{bmatrix} * \begin{bmatrix} X_W - X_{CL} \\ Y_W - Y_{CL} \\ Z_W - Z_{CL} \end{bmatrix} \quad (2)$$

$$\mu^{-1} = \frac{-f_L}{r_{13}(X_W - X_{CL}) + r_{23}(Y_W - Y_{CL}) + r_{33}(Z_W - Z_{CL})} \quad (3)$$

Înclocuind μ^{-1} în (2), se obține:



Reconstrucția 3D din coordonate 2D

Pt. camera stanga:

$$\begin{cases} x_L - x_{CL} = -f_L \frac{r_{11}(X_W - X_{CL}) + r_{21}(Y_W - Y_{CL}) + r_{31}(Z_W - Z_{CL})}{r_{13}(X_W - X_{CL}) + r_{23}(Y_W - Y_{CL}) + r_{33}(Z_W - Z_{CL})} \\ y_L - y_{CL} = -f_L \frac{r_{12}(X_W - X_{CL}) + r_{22}(Y_W - Y_{CL}) + r_{32}(Z_W - Z_{CL})}{r_{13}(X_W - X_{CL}) + r_{23}(Y_W - Y_{CL}) + r_{33}(Z_W - Z_{CL})} \end{cases} \quad (4)$$

Pt. camera dreapta:

$$\begin{cases} x_R - x_{CR} = -f_R \frac{r'_{11}(X_W - X_{CR}) + r'_{21}(Y_W - Y_{CR}) + r'_{31}(Z_W - Z_{CR})}{r'_{13}(X_W - X_{CR}) + r'_{23}(Y_W - Y_{CR}) + r'_{33}(Z_W - Z_{CR})} \\ y_R - y_{CR} = -f_R \frac{r'_{12}(X_W - X_{CR}) + r'_{22}(Y_W - Y_{CR}) + r'_{32}(Z_W - Z_{CR})}{r'_{13}(X_W - X_{CR}) + r'_{23}(Y_W - Y_{CR}) + r'_{33}(Z_W - Z_{CR})} \end{cases} \quad (5)$$

Pentru simplificare vom nota:

$$\begin{cases} x'_L = x_L - x_{CL} \\ y'_L = y_L - y_{CL} \\ x'_R = x_R - x_{CR} \\ y'_R = y_R - y_{CR} \end{cases} \quad (6)$$

D.p.d.v geometric (4) si (5) reprezintă cele doua drepte de proiecție ale punctului P.



Reconstrucția 3D din coordonate 2D

Sistemul geometric(4) si(5) se poate aduce in forma următoare (7) :

$$\mathbf{A} * \begin{bmatrix} X_W \\ Y_W \\ Z_W \end{bmatrix} = \mathbf{B} \Leftrightarrow \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \\ a_{41} & a_{42} & a_{43} \end{bmatrix} * \begin{bmatrix} X_W \\ Y_W \\ Z_W \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \\ b_4 \end{bmatrix}, \quad (7)$$

Unde:

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \\ a_{41} & a_{42} & a_{43} \end{bmatrix} = \begin{bmatrix} r_{13}x'_L + r_{11}f_L & r_{23}x'_L + r_{21}f_L & r_{33}x'_L + r_{31}f_L \\ r_{13}y'_L + r_{12}f_L & r_{23}y'_L + r_{22}f_L & r_{33}y'_L + r_{32}f_L \\ r'_{13}x'_R + r'_{11}f_R & r'_{23}x'_R + r'_{21}f_R & r'_{33}x'_R + r'_{31}f_R \\ r'_{13}y'_R + r'_{12}f_R & r'_{23}y'_R + r'_{22}f_R & r'_{33}y'_R + r'_{32}f_R \end{bmatrix} \quad (8)$$

si

$$\begin{bmatrix} b_1 \\ b_2 \\ b_3 \\ b_4 \end{bmatrix} = \begin{bmatrix} a_{11}X_{CL} + a_{12}Y_{CL} + a_{13}Z_{CL} \\ a_{21}X_{CL} + a_{22}Y_{CL} + a_{23}Z_{CL} \\ a_{31}X_{CR} + a_{32}Y_{CR} + a_{33}Z_{CR} \\ a_{41}X_{CR} + a_{42}Y_{CR} + a_{43}Z_{CR} \end{bmatrix} \quad (9)$$

Rezolvarea algebrica a sistemului(7) se poate realiza prin metoda celor mai mici pătrate:

$$\mathbf{X} = (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \mathbf{B} \quad (10)$$



Realizarea corespondenței între imagini

Acest proces poartă denumirea de stereo-corelație (*stereo-matching*) și constă în punerea în corespondență a unui punct din imaginea stângă \mathbf{P}_L cu corespondentul său din imaginea dreaptă \mathbf{P}_R , unde \mathbf{P}_L și \mathbf{P}_R sunt proiecțiile imagine ale aceluiași punct 3D \mathbf{P}_W . Există două abordări majore a acestei probleme:

- reconstrucția doar a unor *trăsături de interes* (de exemplu puncte de muchie)
- reconstrucția tuturor punctelor din imagine (în măsura în care este posibil procesul de corelație pentru un anumit punct) – *reconstrucție „densă”*.

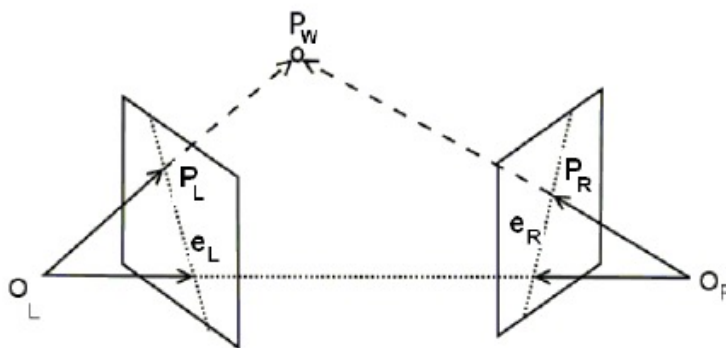
Reducerea spațiului de căutare este o modalitate crucială care permite realizarea corespondenței în timp real atât prin algoritmi software cât și hardware. În plus, se mărește robustețea procesului de corelare prin reducerea fals-pozitivelor (prin limitarea spațiului de căutare). *Reducerea spațiului* de căutare are la bază *constrângerile impuse de geometria epipolară*.



Realizarea corespondenței între imagini

Noțiuni de geometrie epipolară

- Planul format de punctul 3D P_W și centrele optice O_L și O_R ale celor două camere se numește **plan epipolar**.
- **Linia epipolară** este dreapta de intersecție dintre planul epipolar și planul imagine. Unui punct 3D P_W îi vor corespunde două linii epipolare: e_L și respectiv e_R .
- Intersecția dreptei ce trece prin cele două centre optice O_L și O_R (*baseline*-ul) cu cele două planuri imagine se realizează în două puncte fixe (a căror poziție depinde numai de parametrii interni ai sistemului stereo) numite **puncte epipolare** (câte unul pentru fiecare cameră).
- Toate dreptele epipolare ale unei camere se intersectează în punctul epipolar corespunzător.





Realizarea corespondenței între imagini

Principala constrângere a geometriei epipolare folosită la reducerea spațiului de căutare în procesul de stereocorelație afirmă că, pentru fiecare punct $\mathbf{P}_L(x_L, y_L)$ din imaginea stângă, corespondentul său din imaginea dreaptă $\mathbf{P}_R(x_R, y_R)$ se va găsi pe linia epipolară corespunzătoare \mathbf{e}_R având următoarea ecuație (și reciproc):

$$a_R x + b_R y + c_R = 0$$

Parametrii liniei epipolare \mathbf{e}_R pot fi calculați cunoscând matricea fundamentală \mathbf{F} [Trucco98], estimabilă dacă se cunosc parametrii interni ai sistemului stereo:

$$\begin{bmatrix} a_R \\ b_R \\ c_R \end{bmatrix} = \mathbf{F} * \mathbf{P}_L$$

unde :

$$\mathbf{P}_L = \begin{bmatrix} x_L \\ y_L \\ 1 \end{bmatrix} \quad \text{- coordonatele omogene ale punctului } \mathbf{P}_L;$$

$$\mathbf{F} = (\mathbf{A}_R^{-1})^T * \mathbf{E} * \mathbf{A}_L^{-1} \quad \text{- matricea fundamentală;}$$

$$\mathbf{E} = \mathbf{R}_{LR} * \mathbf{S} \quad \text{- matricea esențială;}$$



Realizarea corespondenței între imagini

$\mathbf{A}_L, \mathbf{A}_R$ - matricele interne ale celor două camere

$\mathbf{R}_{LR} = \mathbf{R}_{CR}^T * \mathbf{R}_{CL}$ - matricea de rotație relativă din sistemul camerei din stânga în sistemul camerei din dreapta

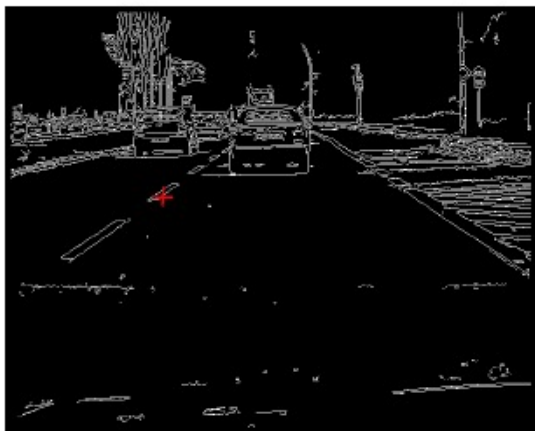
$\mathbf{T}_{LR} = [T_x \ T_y \ T_z]^T = \mathbf{R}_L * (\mathbf{T}_R - \mathbf{T}_L)$ - este vectorul de translație relativă din sistemul camerei din stânga în sistemul camerei din dreapta

$$\mathbf{S} = \begin{bmatrix} 0 & -T_z & T_y \\ T_z & 0 & -T_x \\ -T_y & T_x & 0 \end{bmatrix}$$

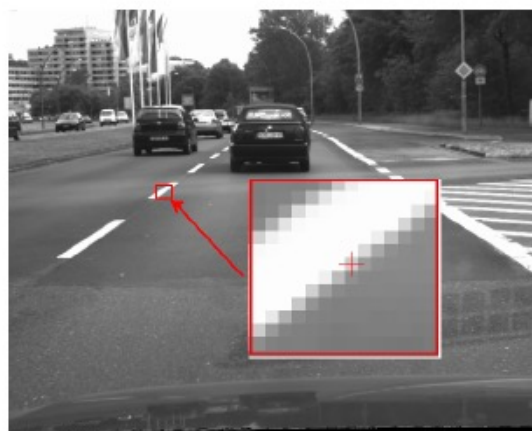
Pentru modelul canonic (imagini rectificate) problema reducerii spațiului de căutare se simplifică semnificativ: perechile de linii epipolare \mathbf{e}_L și \mathbf{e}_R devin coliniare (paralele cu axa orizontală a planului imagine).



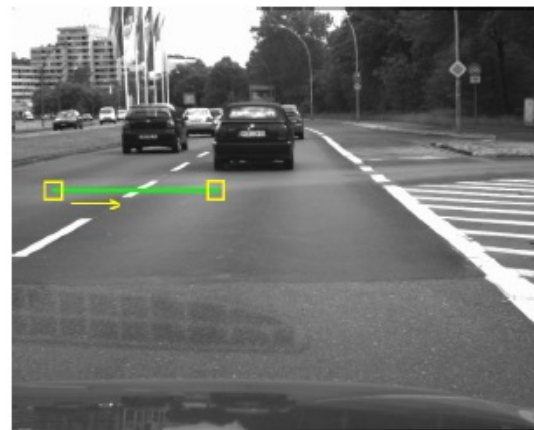
Realizarea corespondenței între imagini



Imaginea de muchii (stanga)



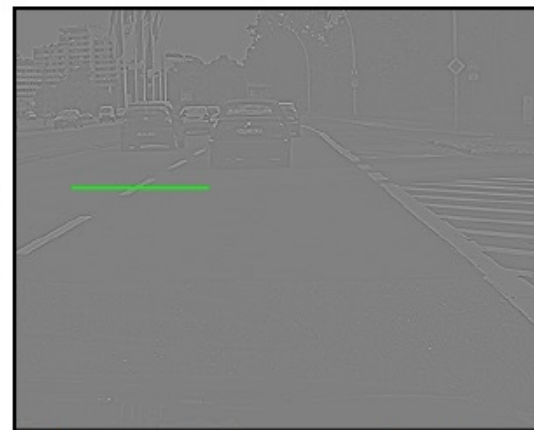
Fereastra (grayscale) corelata din imaginea stanga



Zona de cautare a potrivirii celei mai bune in imaginea dreapta



Fereastra (LOG) corelata din imaginea stanga



Zona de cautare a potrivirii celei mai bune in imaginea dreapta

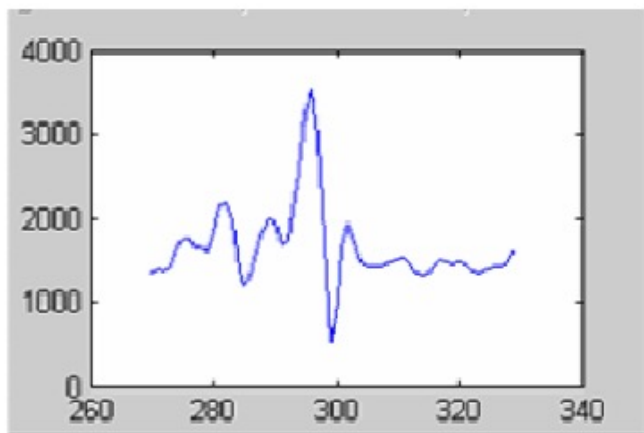


Realizarea corespondenței între imagini

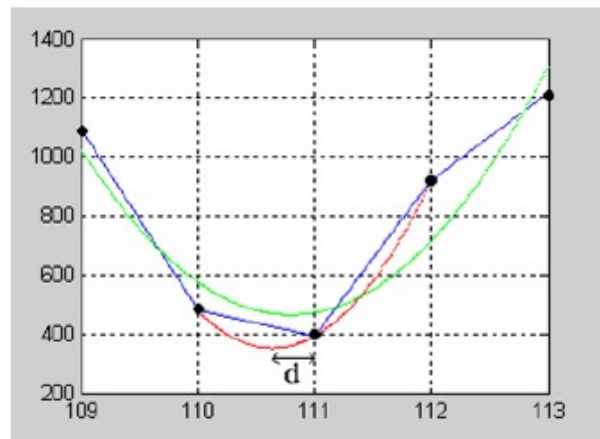
Funcția de corelație

Orice metrica de “distanța”: SAD, SSD, corelație normalizată, etc.

$$SAD(x_R, y_R) = \sum_{i=-\frac{w}{2}}^{\frac{w}{2}} \sum_{j=-\frac{w}{2}}^{\frac{w}{2}} |I_L(x_L + i, y_L + j) - I_R(x_R + i, y_R + j)|$$



Detecția minimului global al funcției de corelație (SAD)



Rafinarea poziției minimului global cu precizie de sub-pixel



Rectificarea

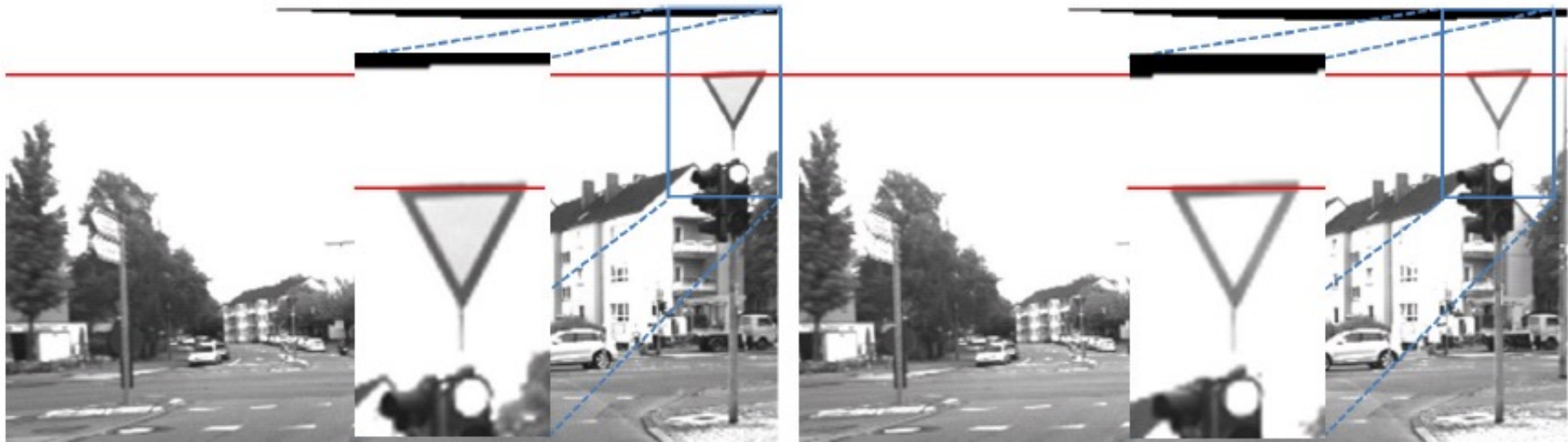
- **Rectificarea**: transformarea imaginilor astfel încât locurile geometrice ale corespondențelor să devină drepte orizontale, având aceeași coordonată în ambele imagini
 - Punctele corespondente în cele două imagini vor fi diferite doar din punct de vedere al coordonatei coloană.





Rectificarea

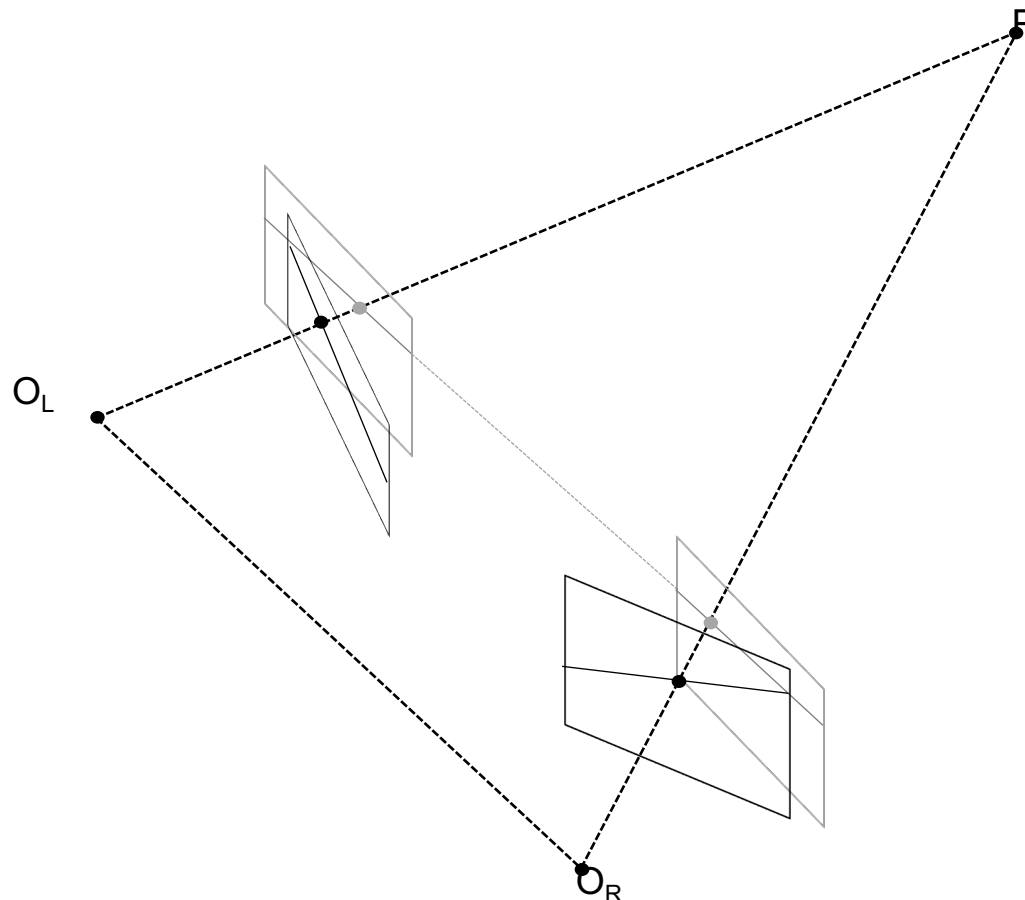
- **Avantaje:** căutarea corespondențelor mult simplificată, cu posibilitatea de accelerare hardware.





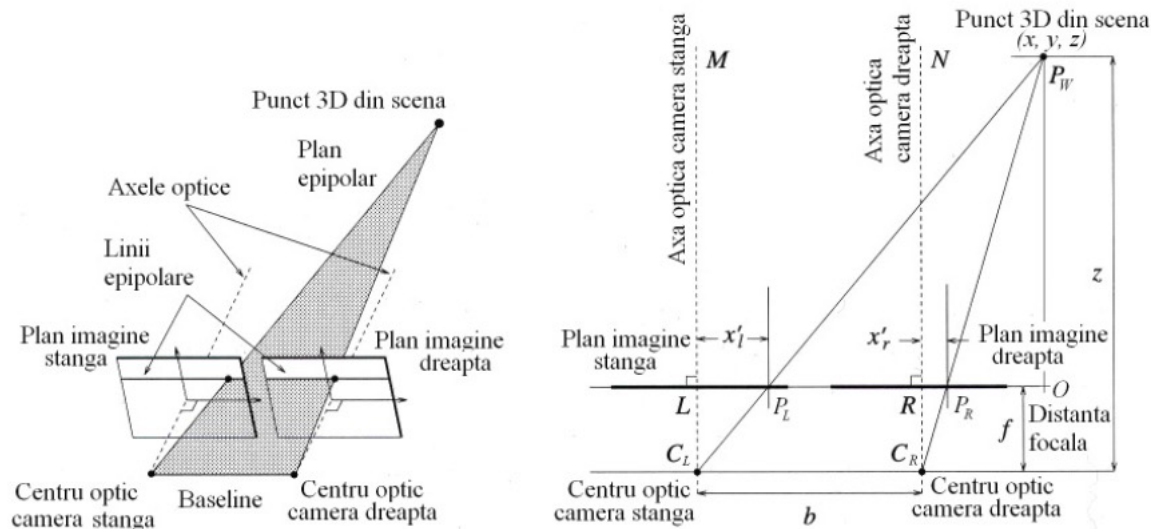
Rectificarea

- Obținerea imaginilor rectificate: proiecția imaginilor originale pe plane imagine aliniate – configurație canonică.





Reconstrucția din imagini rectificate



Modelul stereo canonic (stânga). Ilustrare a procesului de proiecție a unui punct 3D pe cele două planuri imagine după o secțiune orizontală (dreapta).

Modelul canonic

$$x'_l = f \cdot \frac{X_1}{Z_1}$$

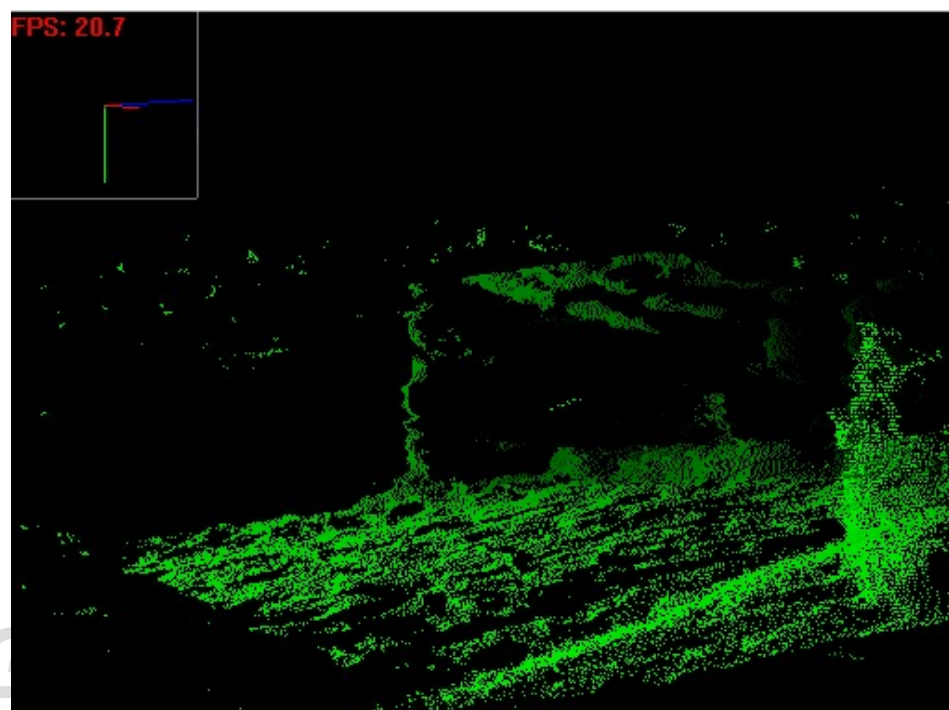
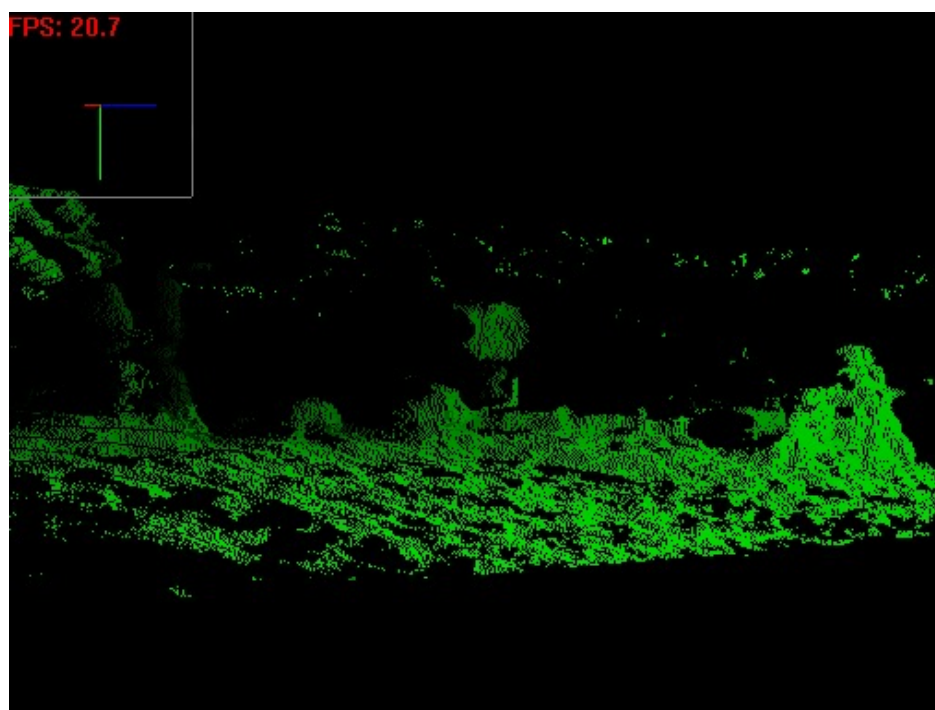
$$x'_r = f \cdot \frac{X_2}{Z_2}$$

$$d = x'_l - x'_r = f \cdot \left(\frac{X_1}{Z_1} - \frac{X_2}{Z_2} \right) = f \cdot \frac{X_1 - X_2}{Z} = f \cdot \frac{b}{Z}$$

$$Z = \frac{f \cdot b}{d}$$



Rezultatele stereoviziunii





Stereoviziune de timp real

- Achiziție de perechi de imagini sincronizate de la camere video industriale monocrome
- Ajustarea automată a parametrilor achiziției pentru adaptarea la variațiile de iluminare
- Folosirea accelerării hardware – procesoare dedicate pentru stereoviziune, sau GPU
- Algoritmii de nivel înalt se execută pe procesorul principal (detecție de obiecte, clasificare, urmărire, etc).

