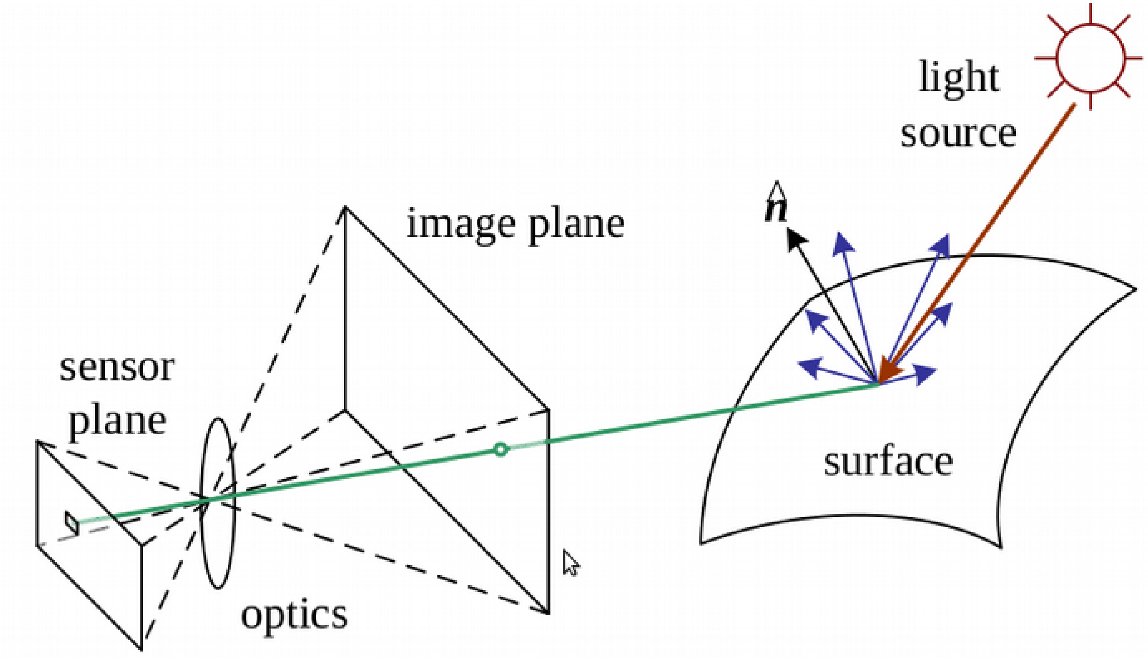
Wizja Maszynowa
Stereowizja - wykład 11

Adam Szmigielski
aszmigie@pjwstk.edu.pl

materiały: $ftp(public)://aszmigie/WM$

Tworzenie obrazu fotometrycznego



- **Oświetlenie** aby wytworzyć obraz, scena musi być oświetlona jedno lub więcej źródeł światła.
- Źródła światła można ogólnie podzielić na punktowe i obszarowe. Punktowe źródło światła powstaje w jednym miejscu w przestrzeni, potencjalnie w nieskończoności (np. Słońce).

Punkty 2D

Prymitywy geometryczne stanowią podstawowe bloki konstrukcyjne używane do opisu trójwymiarowych kształtów.

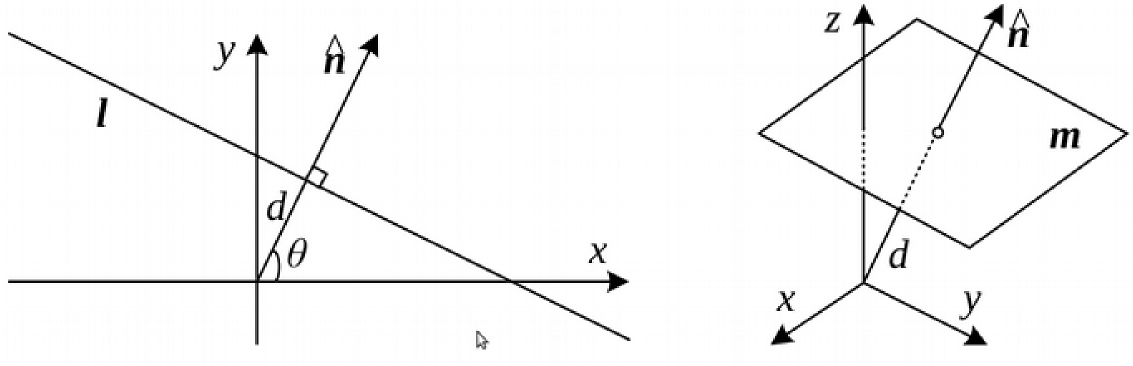
- **Punkty 2D** (współrzędne pikseli na obrazie) można oznaczyć za pomocą pary wartości, $x = (x, y) \in R^2$

  $2D$ punkty można również przedstawić za pomocą jednorodnych współrzędnych,

  $$\tilde{x} = (\tilde{x}, \tilde{y}, \tilde{w}) = \tilde{w}\overline{x} \in P^2$$

  gdzie wektory $\tilde{x}$ różnią się tylko skalą $\tilde{w}$ są równoważne, $\overline{x} = (x, y, 1)$ jest *rozszerzonym wektorem*, a P^2 nazywa się $2D$ *przestrzenią rzutową*.

Linie 2D



(a) (b)

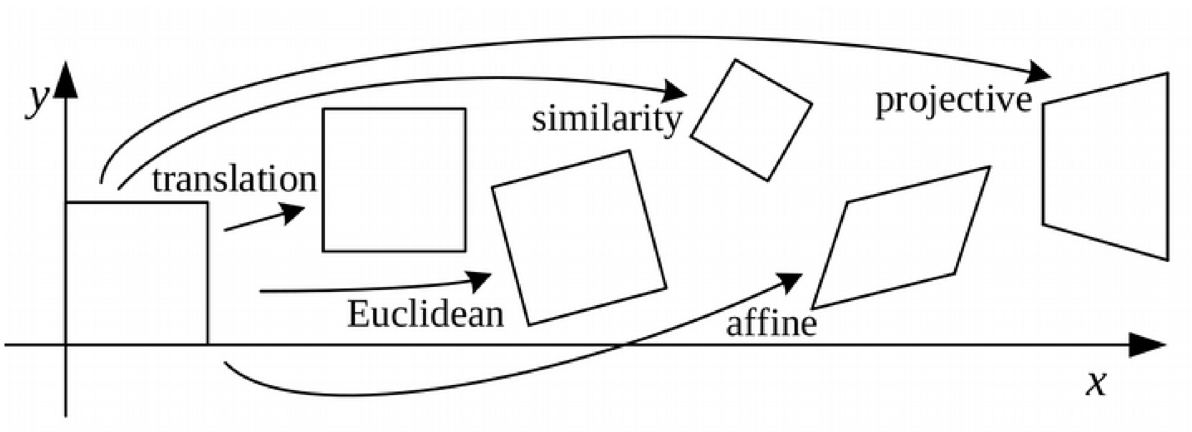
- **Linie 2D** można również przedstawić za pomocą jednorodnych współrzędnych $\tilde{l} = (a, b, c)$. Odpowiednie równanie linii to

$$\tilde{x} \cdot \tilde{l} = ax + by + c = 0.$$

Możemy tak znormalizować wektor równania linii $l = (\hat{n}_x, \hat{n}_y, d) = (\hat{n}, d)$ z $||\hat{n}|| = 1$ w tym przypadku, $\hat{n}$ jest wektorem normalnym prostopadłym do prostej, a d jest jego odległością od początku.

- Kombinacja (Θ, d) jest również znana jako *biegunowe współrzędne.*

Transformacje 2D



- **Przekształcenie 2D** można zapisać jako $x' = x + t$ albo

$$x' = \begin{bmatrix} I & t \end{bmatrix} \cdot \overline{x},$$

  gdzie I jest 2×2 macierzą jednostkową.

- **Obrót + przesunięcie**. To przekształcenie jest również znane jako ruch ciała sztywnego 2D lub transformacja euklidesowa 2D. Można go zapisać jako

$$x' = \begin{bmatrix} R & t \end{bmatrix} \overline{x}$$

gdzie

$$R = \begin{bmatrix} \cos(\theta) & -sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix}$$

jest ortonormalną macierzą rotacji z $R \cdot R^T = I$ i $|R| = 1$.

- **Obrót skalowany**, znany również jako transformacja podobieństwa, transformację tę można wyrazić jako $x' = sRx + t$, gdzie s jest dowolnym współczynnikiem skali.

$$x' = \begin{bmatrix} sR & t \end{bmatrix} \overline{x} = \begin{bmatrix} a & -b & t_x \\ b & a & t_y \end{bmatrix} \overline{x}$$

gdzie nie wymagamy już, aby $a^2 + b^2 = 1$. Transformacja podobieństwa zachowuje kąty między liniami.

- **Transformacja afiniczna** jest zapisywana jako $x' = A\overline{x}$ gdzie A jest dowolną macierzą 2×3, tj.

$$x' = \begin{bmatrix} a_{00} & a_{01} & a_{02} \\ a_{10} & a_{11} & a_{12} \end{bmatrix} \overline{x}$$

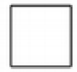
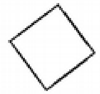
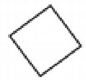
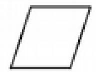
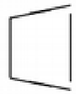
  Proste równoległe pozostają równoległe w przypadku przekształceń afinicznych (nie zachowuje kątów między prostymi i odległości między punktami)

- **Transformacja projekcyjna**, znana również jako transformacja perspektywiczna lub homografia, działa na jednorodnych współrzędnych,

$$\tilde{x}' = \tilde{H}\tilde{x},$$

  gdzie $\tilde{H}$ jest macierzą 3×3. $\tilde{H}$ jest zdefiniowana w określonej skali, a dwie macierze $\tilde{H}$, które różnią się tylko skalą, są równoważne.

Hierarchia przekształceń współrzędnych 2D

Transformation	Matrix	# DoF	Preserves	Icon
translation	$\begin{bmatrix} \boldsymbol{I} \mid \boldsymbol{t} \end{bmatrix}_{2\times 3}$	2	orientation	
rigid (Euclidean)	$\begin{bmatrix} \boldsymbol{R} \mid \boldsymbol{t} \end{bmatrix}_{2\times 3}$	3	lengths	
similarity	$\begin{bmatrix} s\boldsymbol{R} \mid \boldsymbol{t} \end{bmatrix}_{2\times 3}$	4	angles	
affine	$\begin{bmatrix} \boldsymbol{A} \end{bmatrix}_{2\times 3}$	6	parallelism	
projective	$\begin{bmatrix} \tilde{\boldsymbol{H}} \end{bmatrix}_{3\times 3}$	8	straight lines	

- Transformacje zachowują pewne właściwości (podobieństwo zachowuje nie tylko kąty, ale także równoległość prostych).
- Macierze 2×3 są rozszerzane o trzeci wiersz $[0^T 1]$, tworząc pełną macierz 3×3 dla jednorodnych przekształceń współrzędnych.

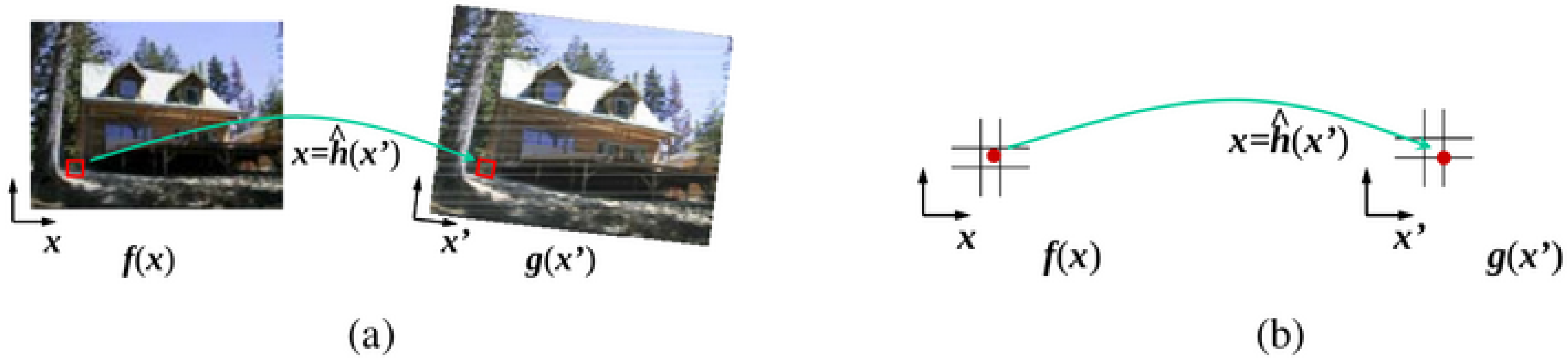
Wypaczanie do przodu (lub mapowanie)

(a) (b)

- (a) piksel $f(x)$ jest kopiowany do odpowiedniego miejsca $x' = h(x)$ w obrazie $g(x')$;
- (b) szczegóły lokalizacji pikseli źródłowych i docelowych.
- Algorytm wypaczania do przodu służący do przekształcania obrazu $f(x)$ w obraz $g(x')$:

```
procedure forwardWarp(f, h, out g):
        For every pixel x in f(x)
                1. Compute the destination location x' = h(x).
                2. Copy the pixel f(x) to g(x').
```

Odwrotne wypaczanie (lub mapowanie)



- (a) piksel $g(x')$ jest próbkowany z odpowiedniego miejsca $x = \hat{h}(x')$ w obrazie $f(x)$;
- (b) szczegóły lokalizacji pikseli źródłowych i docelowych.
- Algorytm odwrotnego wypaczania do tworzenia obrazu $g(x')$ z obrazu $f(x)$ przy użyciu transformacji parametrycznej $x' = h(x)$

```
procedure inverseWarp(f, h, out g):
        For every pixel x' in g(x')
                1. Compute the source location x = h^(x')
                2. Resample f(x) at location x and copy to g(x')
```

Wyrównanie oparte na elementach 2D i 3D

Wyrównanie oparte na cechach to problem szacowania ruchu między dwoma lub więcej zestawami dopasowanych punktów 2D lub 3D.

Transform	Matrix	Parameters p	Jacobian J
translation	$\begin{bmatrix} 1 & 0 & t_x \\ 0 & 1 & t_y \end{bmatrix}$	(t_x, t_y)	$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$
Euclidean	$\begin{bmatrix} c_\theta & -s_\theta & t_x \\ s_\theta & c_\theta & t_y \end{bmatrix}$	(t_x, t_y, θ)	$\begin{bmatrix} 1 & 0 & -s_\theta x - c_\theta y \\ 0 & 1 & c_\theta x - s_\theta y \end{bmatrix}$
similarity	$\begin{bmatrix} 1+a & -b & t_x \\ b & 1+a & t_y \end{bmatrix}$	(t_x, t_y, a, b)	$\begin{bmatrix} 1 & 0 & x & -y \\ 0 & 1 & y & x \end{bmatrix}$
affine	$\begin{bmatrix} 1+a_{00} & a_{01} & t_x \\ a_{10} & 1+a_{11} & t_y \end{bmatrix}$	$(t_x, t_y, a_{00}, a_{01}, a_{10}, a_{11})$	$\begin{bmatrix} 1 & 0 & x & y & 0 & 0 \\ 0 & 1 & 0 & 0 & x & y \end{bmatrix}$
projective	$\begin{bmatrix} 1+h_{00} & h_{01} & h_{02} \\ h_{10} & 1+h_{11} & h_{12} \\ h_{20} & h_{21} & 1 \end{bmatrix}$	$(h_{00}, h_{01}, \ldots, h_{21})$	

- Sprowadzamy problem do globalnych *transformacji parametrycznych*,
- Jakobiany transformacji współrzędnych 2D $x' = f(x; p)$.

Wyrównanie 2D przy użyciu metody najmniejszych kwadratów

Mając zestaw dopasowanych punktów cech $\{x_i, x'_i\}$ i planarną transformację parametryczną postaci $x' = f(x; p)$
Jak możemy oszacować parametry p ?

- Zwykle metodą najmniejszych kwadratów (LS, aby zminimalizować sumę kwadratów błędów)

$$E_{LS} = \sum_i ||r_i||^2 = \sum_i ||f(x_i; p) - x'_i||^2,$$

gdzie

$$r_i = f(x_i; p) - x'_i = \hat{x_i}' - \tilde{x_i}'$$

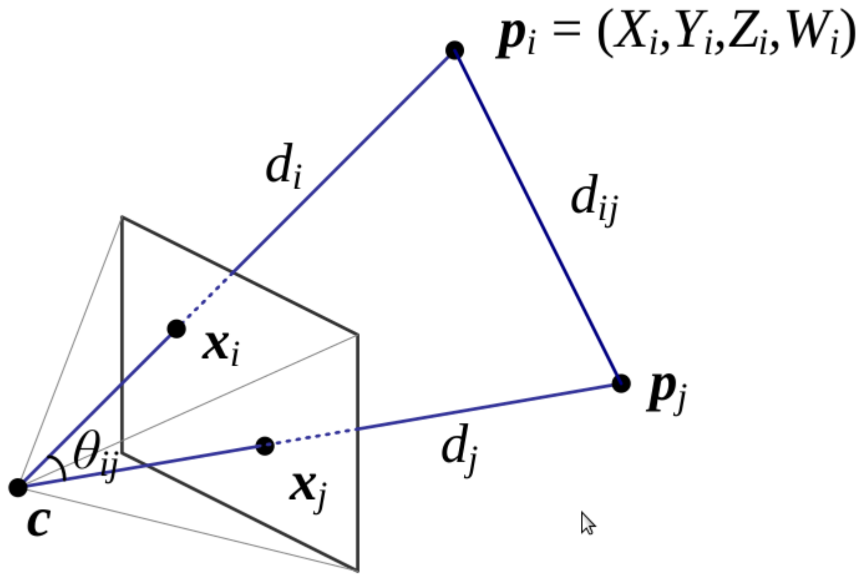
jest resztą między zmierzoną lokalizacją $\hat{x_i}'$ a odpowiadającą jej przewidywaną lokalizacją $\tilde{x_i}' = f(x_i; p)$.

Estymacja położenia

Szczególnym przypadkiem wyrównania opartego na cechach, które występuje bardzo często, jest oszacowanie pozy 3D obiektu na podstawie zestawu rzutów punktowych 2D

- Ten problem **estymacji pozycji** jest również znany jako **zewnętrzna kalibracja** (w przeciwieństwie do *wewnętrznej kalibracji* wewnętrznych parametrów aparatu, takich jak ogniskowa itp.)
- Problem odzyskania pozycji z trzech korespondencji jest znany jako problem z perspektywą-3-punktami (P3P)

Algorytmy liniowe



Najprostszym sposobem odtworzenia ułożenia kamery jest utworzenie układu równań liniowych z macierzy kamery projekcji perspektywicznej,

$$x_i = \frac{p_{00}X_i + p_{01}Y_i + p_{02}Z_i + p_{03}}{p_{20}X_i + p_{21}Y_i + p_{22}Z_i + p_{23}}$$

$$y_i = \frac{p_{10}X_i + p_{11}Y_i + p_{12}Z_i + p_{13}}{p_{20}X_i + p_{21}Y_i + p_{22}Z_i + p_{23}}$$

gdzie (x_i, y_i) to zmierzone lokalizacje elementów $2D$, a (X_i, Y_i, Z_i) to znane lokalizacje elementów $3D$.

Algorytmy liniowe

- Aby obliczyć 12 niewiadomych w P, trzeba znać co najmniej sześć korelacji między lokalizacjami $3D$ i $2D$.
- Po odzyskaniu wpisów w P można odzyskać zarówno wewnętrzną macierz kalibracji K, jak i sztywną transformację (R, t):

$$P = K[R|t].$$

- Ponieważ K jest zgodnie z konwencją górny trójkątny, zarówno K, jak i R można uzyskać z przodu 3×3 podmacierz P stosując faktoryzację RQ,
- W przypadku, gdy kamera jest już skalibrowana, znana jest macierz K,
- Kąt widzenia θ_{ij} między dowolną parą punktów $2D$ $\hat{x}_i$ i $\hat{x}_j$ musi być taki sam, jak kąt między odpowiadającymi im punktami $3D$ p_i i p_j.

Algorytmy liniowe

- Biorąc pod uwagę zbiór odpowiadających sobie punktów $2D$ i $3D$ $\{(\hat{x_i}, pi)\}$, gdzie $\hat{x_i}$ to kierunki jednostek uzyskane przez przekształcenie pomiarów $2D$ pikseli x_i do normy jednostkowej Kierunki 3D $\hat{x_i}$ przez odwrotną macierz kalibracji K

$$\hat{x_i} = \frac{K^{-1}x_i}{||K^{-1}x_i||},$$

niewiadomymi są odległości d_i od początku kamery c do $3D$ punktów p_i, gdzie

$$p_i = d_i\hat{x_i} + c$$

- Po dokonaniu indywidualnych oszacowań odległości d_i możemy wygenerować strukturę $3D$ składającą się ze skalowanych kierunków punktów $d_i\hat{x_i}$, aby uzyskać pożądane oszacowanie pozycji.

Algorytmy iteracyjne

Najdokładniejszym sposobem oszacowania ułożenia jest bezpośrednie zminimalizowanie kwadratowego błędu odwzorowania punktów 2D przy użyciu nieliniowych najmniejszych kwadratów.

- Możemy zapisać równania rzutowania jako

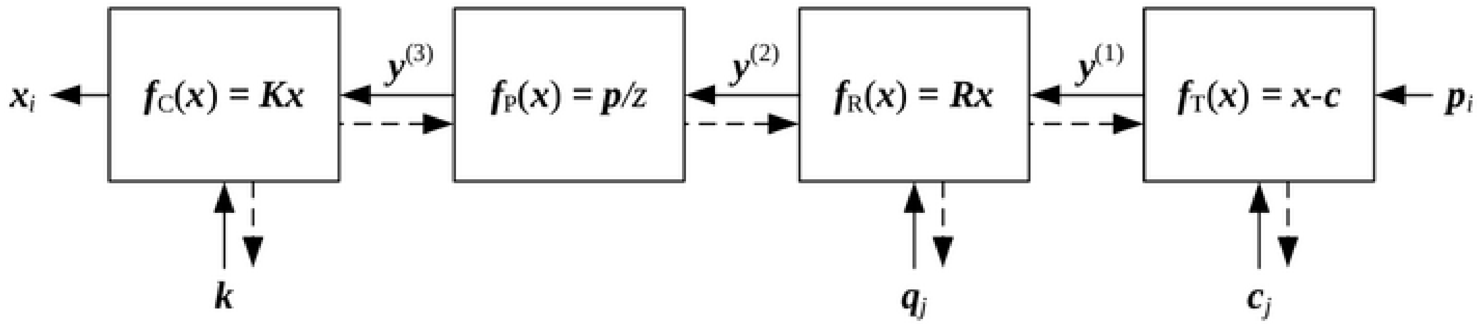
$$x_i = f(p_i; R, t, K)$$

i iteracyjnie zminimalizuj zlinearyzowane błędy odwzorowania:

$$E_{NLP} = \sum_i \rho\left(\frac{\partial f}{\partial R} \triangle R + \frac{\partial f}{\partial t} \triangle t + \frac{\partial f}{\partial K} \triangle K - r_i\right)$$

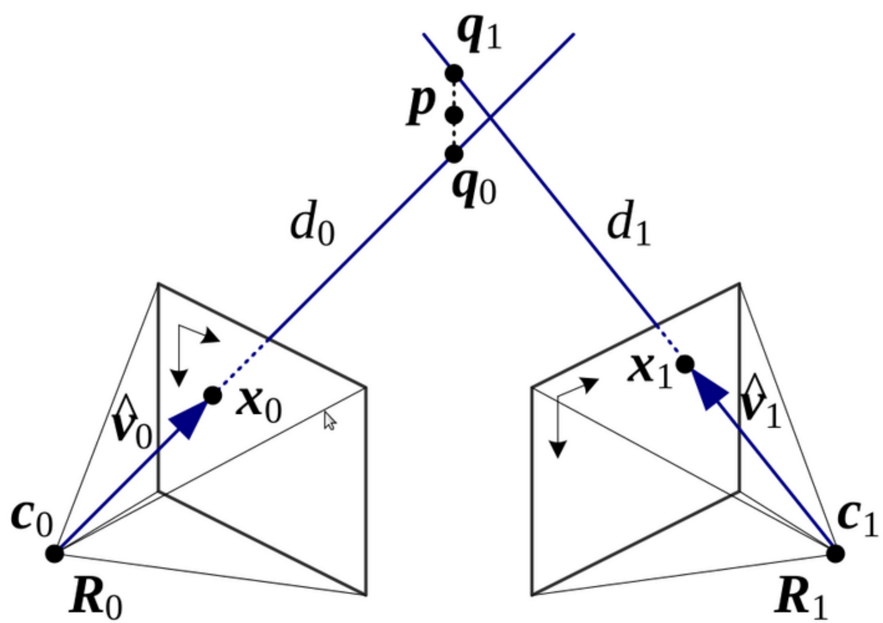
gdzie $r_i = \tilde{x}_i - \hat{x}_i$ to błąd 2D w przewidywanej pozycji.

Implementacja algorytmów iteracyjnych



- Zestaw połączonych transformacji do rzutowania $3D$ punktu p_i na pomiar $2D$ x_i poprzez serię transformacji $f^{(k)}$, z których każda jest kontrolowana przez własny zestaw parametrów .
- Linie przerywane wskazują przepływ informacji, ponieważ częściowe pochodne są obliczane podczas przejścia wstecz.

Triangulacja



Problem określania pozycji 3D punktu na podstawie zestawu odpowiednich lokalizacji obrazu i znanych pozycji kamery jest znany jako *triangulacja*.

- Ten problem jest odwrotnością oceny pozycji,
- Aby rozwiązać ten problem, musimy znaleźć $3D$ punkt p, który leży najbliżej wszystkich promieni $3D$ odpowiadających lokalizacjom obiektów dopasowania $2D$ $\{x_j\}$ obserwowanych przez kamery $\{P_j = K_j[R_j|t_j]\}$, gdzie $t_j = -R_j c_j$ i c_j to j centrum kamery.

Triangulacja

- Najbliższy punkt p na tym promieniu, który oznaczymy jako q_1, minimalizuje odległość $||c_1 + d_1\hat{v_1} - p||^2$,
- Optymalną wartość p, która leży najbliżej wszystkich promieni, można obliczyć jako zwykłe zadanie najmniejszych kwadratów, sumując wszystkie r_j^2 i znajdując optymalną wartość p:

$$p = [\sum_i (I - \hat{v_j}\hat{v_j}^T)]^{-1}[\sum_i (I - \hat{v_j}\hat{v_j}^T)c_j].$$

Rzuty 3D do 2D

Możemy to zrobić używając liniowej macierzy projekcji $3D$ do $2D$.

- Rzutowanie ortograficzne po prostu opuszcza składową z trójwymiarowej współrzędnej p, aby otrzymać $2D$ punkt x.

$$x = [I_{2\times 2}|0]p$$

- Często obraz trzeba przeskalować o s, aby zmieścił się na matrycy (np. cm do pikseli). Z tego powodu ortografia skalowana jest w rzeczywistości częściej używana,

$$x = [sI_{2\times 2}|0]p$$

Perspektywa

Najczęściej używaną projekcją w grafice komputerowej i wizji komputerowej jest prawdziwa perspektywa $3D$

- Tutaj punkty są rzutowane na płaszczyznę obrazu przez podzielenie ich przez składową z
- We współrzędnych jednorodnych rzut ma prostą postać liniową

$$\tilde{x} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \tilde{p}$$

- Do znormalizowanych współrzędnych urządzenia w zakresie $(x, y, z) \in [-1, -1] \times [-1, 1] \times [0, 1]$, a następnie przeskalowuje te współrzędne do całkowitych współrzędnych pikseli przy użyciu transformacji widoku

- (Początkowa) projekcja perspektywiczna jest następnie reprezentowana za pomocą macierzy 4×4

$$\tilde{x} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & \frac{-z_{far}}{z_{range}} & \frac{z_{near} \cdot z_{far}}{z_{range}} \\ 0 & 0 & 1 & 0 \end{bmatrix} \tilde{p}$$

  gdzie z_{near} i z_{far} to bliskie i dalekie *z płaszczyzn obcinania* i $z_{range} = z_{far} - z_{near}$.

- Jeśli ustawimy $z_{near} = 1$, $z_{far} = \inf$, trzeci element znormalizowanego wektora rastrowego staje się odwrotną głębią (*rozbieżność*)

- Można odwzorować wartość dysproporcji d bezpośrednio z powrotem do lokalizacji $3D$ przy użyciu odwrotności macierzy 4×4.
- Możemy to zrobić, jeśli przedstawimy rzut perspektywiczny za pomocą macierzy pełnej rangi 4 $\times$4:

$$\tilde{P} = \begin{bmatrix} K & 0 \\ 0^T & 1 \end{bmatrix} \begin{bmatrix} R & t \\ 0^T & 1 \end{bmatrix} = \tilde{K} \cdot E$$

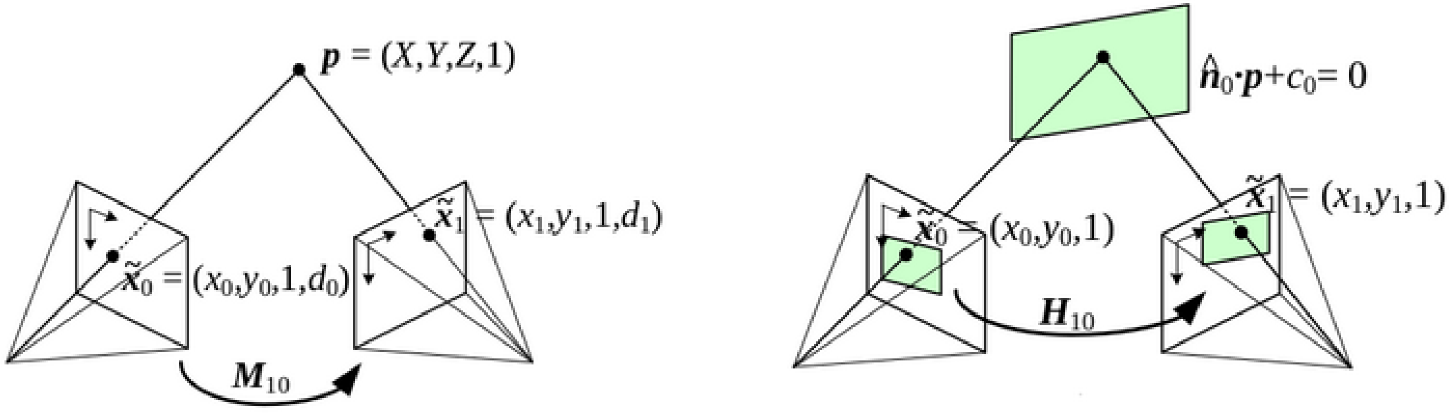
gdzie E jest $3D$ transformacją ciała sztywnego (euklidesa), a $\tilde{K}$ jest macierzą kalibracji pełnego rzędu.

- Macierz kamery 4×4 $\tilde{P}$ może być używany do mapowania bezpośrednio ze współrzędnych świata $3D$ $\overline{p_w} = (x_w, y_w, z_w, 1)$ do współrzędnych ekranu (plus różnica), $x_s = (x_s, y_s, 1, d)$,

$$s_s \sim \tilde{P}\overline{p_w}$$

gdzie $\sim$ wskazuje równość w skali.

Mapowanie z jednej kamery do drugiej



(a) (b)

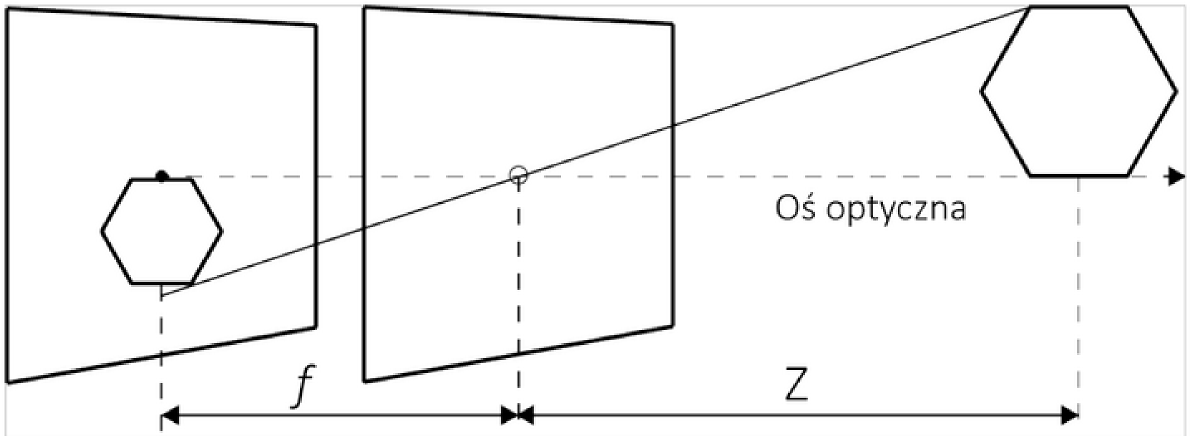
- Używając pełnej rangi 4×4 matrycy kamery $\tilde{P} = \tilde{K}E$ możemy zapisać rzutowanie ze świata na współrzędne ekranu jako

$$\tilde{x_0} \sim \tilde{K_0} E_0 p = \tilde{P_0} p.$$

- Zakładając, że znamy wartość różnicy d_0 dla piksela w jednym obrazie, możemy obliczyć położenie $3D$ punktu p za pomocą

$$p \sim E_0^{-1} \tilde{K_0}^{-1} \tilde{x_0}.$$

Model kamery otworkowej

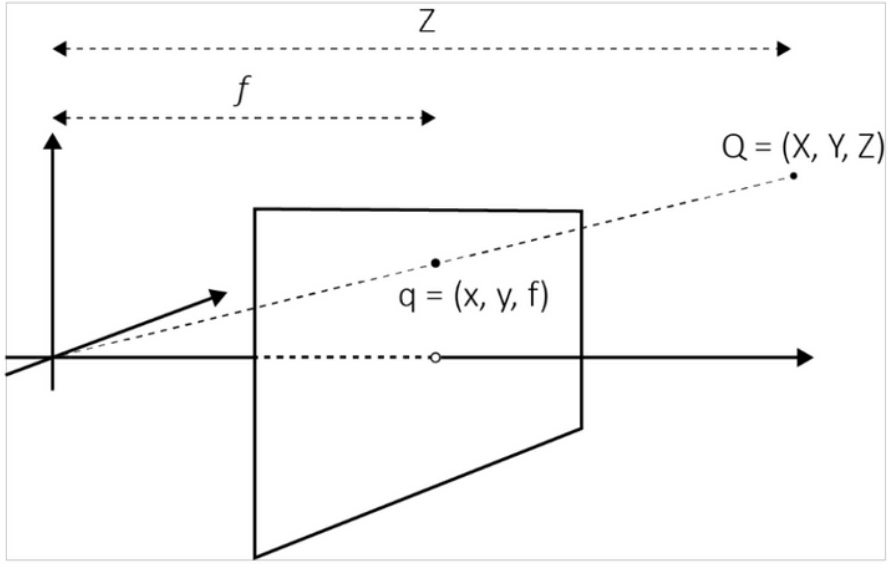


Z podobieństw trójkątów:

$$\frac{-x}{f} = \frac{X}{Z}$$

•

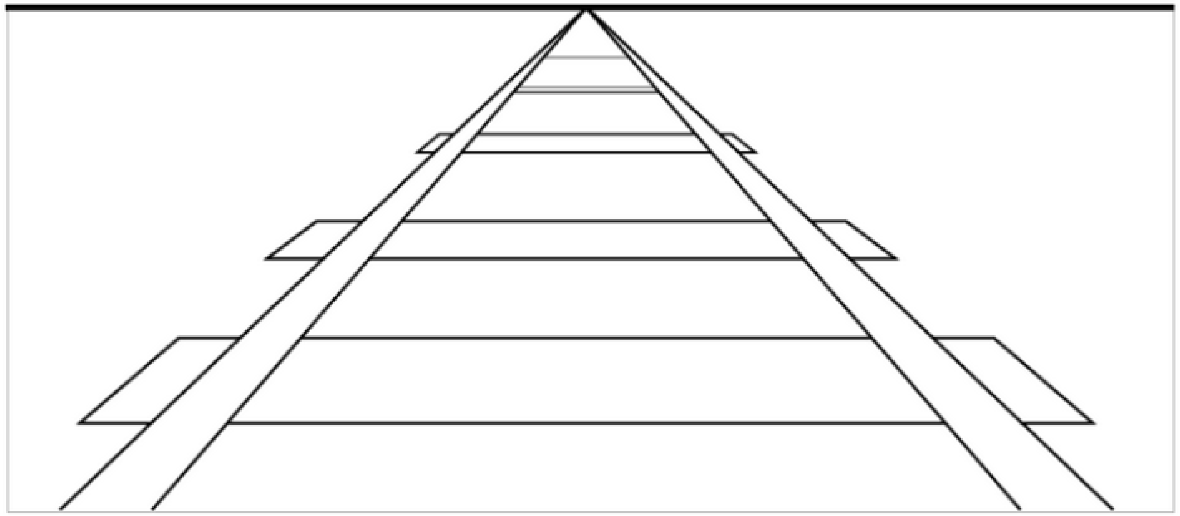
Zmodyfikowany model kamery otworkowej



$$x_{ekran} = f_x \frac{x_{swiat}}{Z} + c_x \text{ oraz } y_{ekran} = f_y \frac{y_{swiat}}{Z} + c_x$$

- c_x oraz c_y określają przesunięcie punktu głównego od osi optycznej,
- Ogniskowe f_x oraz f_y zostały rozdzielone ponieważ piksele w urządzeniu rejestrującym mogą być prostokątne,
- Wielkość f_x oznacza iloczyn ogniskowej i rozmiaru s_x elementu układu obrazującego.

Współrzędne homogeniczne



- Przejście ze świata trójwymiarowego na współrzędne obrazu nazywa się przekształceniem rzutowym,
- W geometrii euklidesowej dwie linie równoległe nie stykają się, w przestrzeni rzutowej zbiegają się do wspólnego punku w nieskończoności,
- W przestrzeni rzutowej używa się współrzędnych homogenicznych. Przedstawiają one punkt n wymiarowy za pomocą $n-1$ wymiarów.

Współrzędne homogeniczne

- Przejście z punktu $(1, 2)$ na współrzędne homogeniczne polega na dodaniu dodatkowego czynnika $(1, 2, w)$, przy czym zamiana współrzędnych na kartezjańskie wygląda następująco:

$$(x, y, w) \Leftrightarrow (\frac{x}{w}, \frac{y}{w})$$

- W przypadku płaszczyzny światłoczułej punkt w postaci $q = (x, y, f)$ będzie miał współrzędne równe $q' = (\frac{x}{f}, \frac{y}{f})$.
- Jeśli dodatkowa współrzędna homogeniczna punktu jest równa, zero punkt ten jest w nieskończoności:

$$(\frac{x}{0}, \frac{y}{0}) \Rightarrow (\infty, \infty)$$

Macierz parametrów wewnętrznych kamery

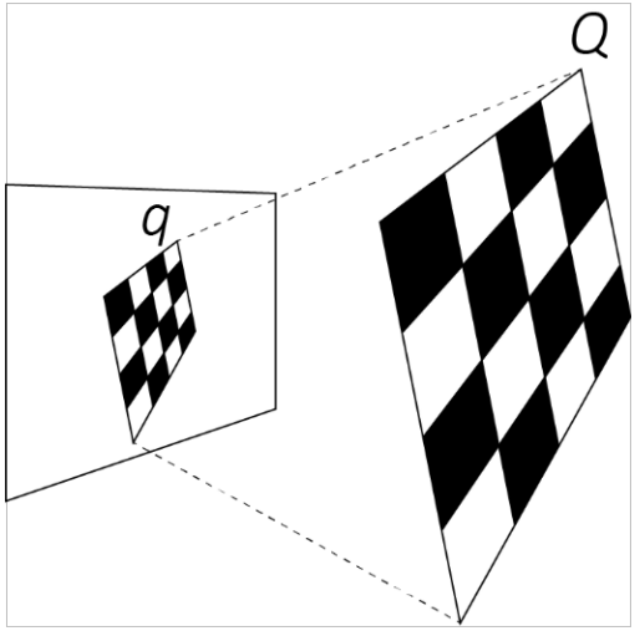
- Posługując się *macierzą parametrów wewnętrznych kamery* M można określić odwzorowanie punktu świata rzeczywistego Q na obraz:

$$q = M \cdot Q$$

gdzie:
q jest współrzędną homogeniczną $q = [x, y, f]^T$
Q jest punktem świata rzeczywistego $Q = [X, Y, Z]^T$
M jest macierzą parametrów wewnętrznych kamery

$$M = \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix}$$

Homografia



- Przekształcenia homograficznego jest odwzorowaniem jednej płaszczyzny na inną płaszczyznę.
- Przekształcenie homograficzne wykonuje się za pomocą macierzy homografii i współrzędnych homogenicznych:

$$q = sHQ$$

gdzie s jest współczynnikiem skali a H macierzą homografii.

Macierz homografii

- Aby określić macierz homografii należy określić macierz rotacji R oraz wektor translacji t.
- Współrzędne punktu q na obrazie nie posiadają głębokości można pominąć rotację w tej płaszczyźnie zostawiając tylko wektory r_1 i r_2.
- Uwzględniając macierz parametrów wewnętrznych kamery M *macierz homografii* ma postać:

$$H = sM[r_1, r_2, t] \cdot \begin{bmatrix} X \\ Y \\ 1 \end{bmatrix}$$

Zniekształcenia obrazu - dystorsja radialna

- W praktyce, aby skupić możliwie dużo światła stosuje się układy optyczne (soczewki),
- Dodanie soczewki powoduje *dystorsje radialną* która można opisać jako:
  $x_p = (x_n - c_x)(k_1 r^2 + k_2 r^4 + k_3 r^6)$
  $y_p = (y_n - c_x)(k_1 r^2 + k_2 r^4 + k_3 r^6)$
  gdzie:
  (x_n, y_n) – oryginalna lokalizacja punktu,
  (x_p, y_p) – poprawiona lokalizacja punktu,
  (c_x, c_y) – środek obrazu,
  $r = \sqrt{(x_n - c_x)^2 + (y_n - c_y)^2}$,
  k_1, k_2, k_3 – wartości zniekształcenia radialnego.

Zniekształcenia obrazu - dystorsja tangensowa

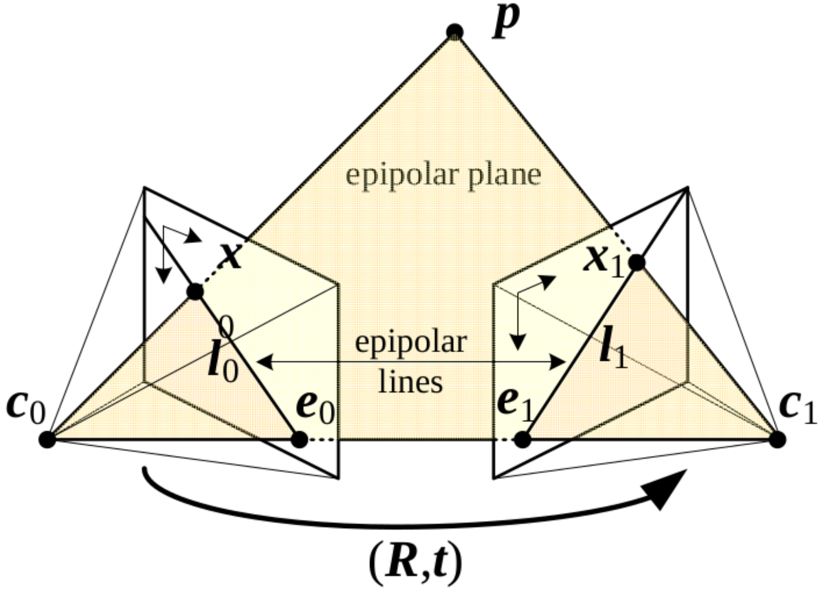
- Jej źródłem jest nierówne przymocowanie układu obrazującego do kamery
- Zniekształcenie tego typu opisują dwa parametry p_1 i p_2:
  $x_p = x_n + [2p_1(x_n - c_x)(y_n - c_y) + p_2(r^2 + 2(x_n - c_x)^2)]$
  $y_p = y_n + [p_1(r^2 + 2(y_n - c_y)^2) + 2p_2(x_n - c_x)(y_n - c_y)]$
  gdzie:
  (x_n, y_n) – oryginalna lokalizacja punktu,
  (x_p, y_p) – poprawiona lokalizacja punktu,
  (c_x, c_y) – środek obrazu,
  $r = \sqrt{(x_n - c_x)^2 + (y_n - c_y)^2}$,
  p_1, p_2 – wartości zniekształcenia tangensowego.

Układ stereo kamer

Proces interpretacji dwóch obrazów:

1. Usunięcie zniekształceń - krok **kalibracji**,
2. Wyrównane obrazów względem siebie – ten krok nazywa się **rektyfikacją**,
3. Poszukiwanie odpowiadających sobie punktów - ten krok nazywa się **szukaniem korespondencji**

Układ dwóch kamer



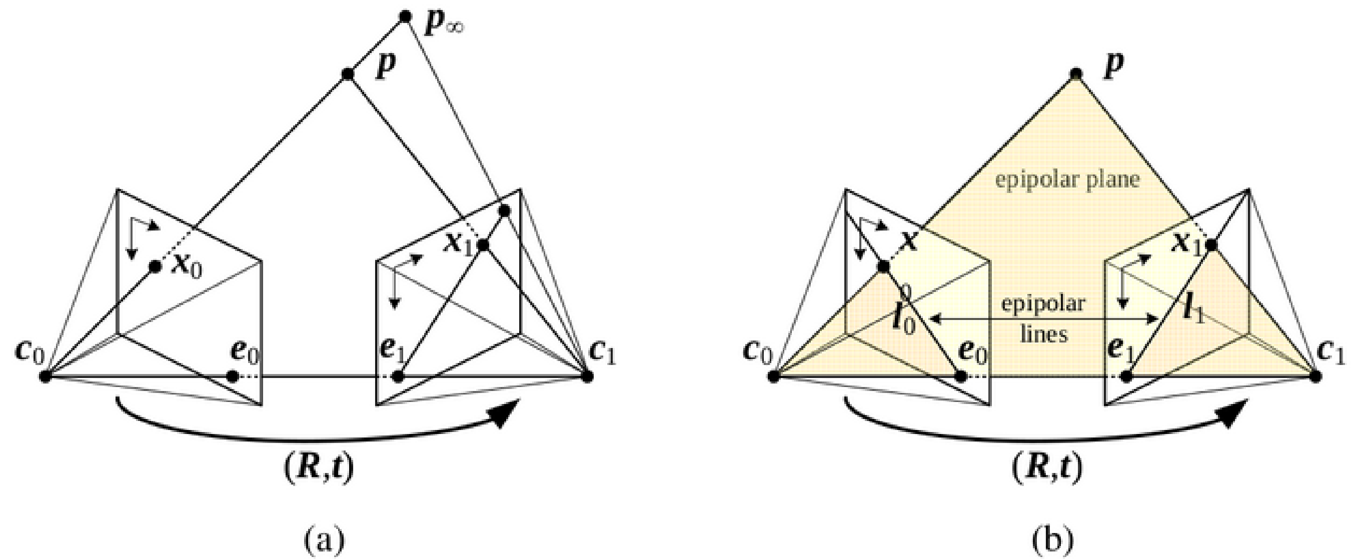
W tym zadaniu jednocześnie odzyskujemy strukturę $3D$ i pozujemy z korespondencji obrazu.

- Względną pozycję można zakodować za pomocą rotacji R i przesunięcia t,
- Wektory $t = c_1 - c_0$, $p - c_0$ i $p - c_1$ są współpłaszczyznowe i definiują podstawowe ograniczenie epipolarne wyrażone w wymiarach pikseli x_0 i x_1.

Dopasowanie stereo

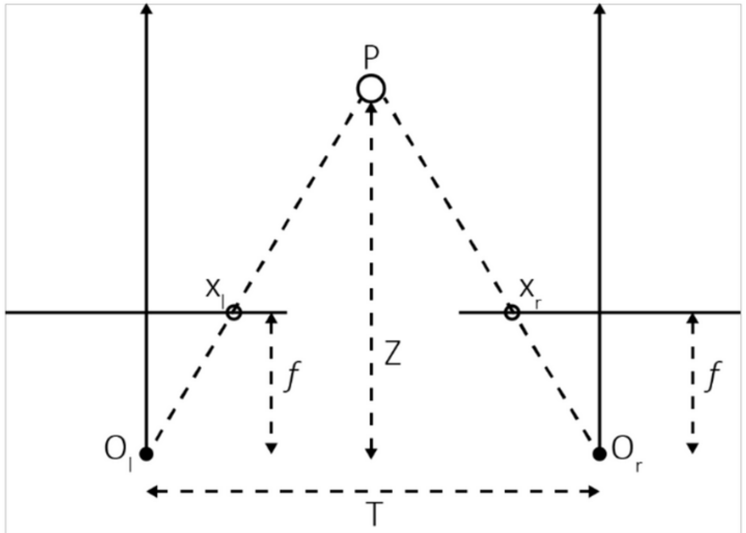
Dopasowywanie stereo to proces wykonywania dwóch lub więcej zdjęć i szacowania modelu $3D$ o wartości sceny, znajdując pasujące piksele na obrazach i konwertując ich pozycje $2D$ na $3D$ głębokości.

Geometria epipolarna



- (a) epipolarny odcinek linii odpowiadający jednemu promieniu. Piksel x_0 na jednym obrazie rzutuje na epipolarny segment linii na drugim obrazie.
- (b) odpowiedni zestaw linii epipolarnych i ich płaszczyzna epipolarna.

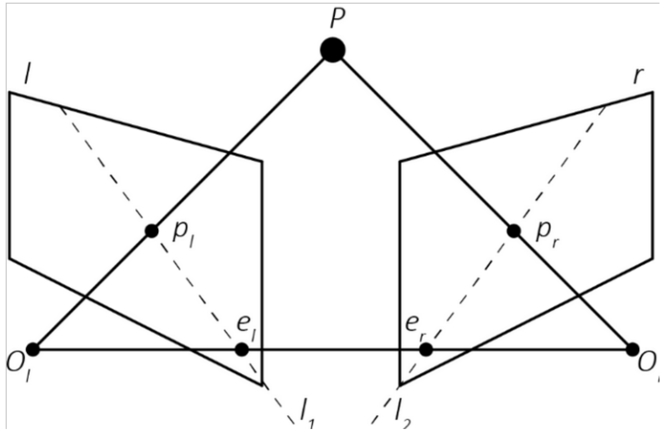
Geometria stereo - układ idealny



- Układu kamer, położonych w jednej linii, identycznych kamer z równoległymi osiami optycznymi nazywa się układem frontowo-równoległym.
- Obie kamery mają równą ogniskową f, więc z podobieństwa trójkątów można wyprowadzić odległość przedmiotu P:

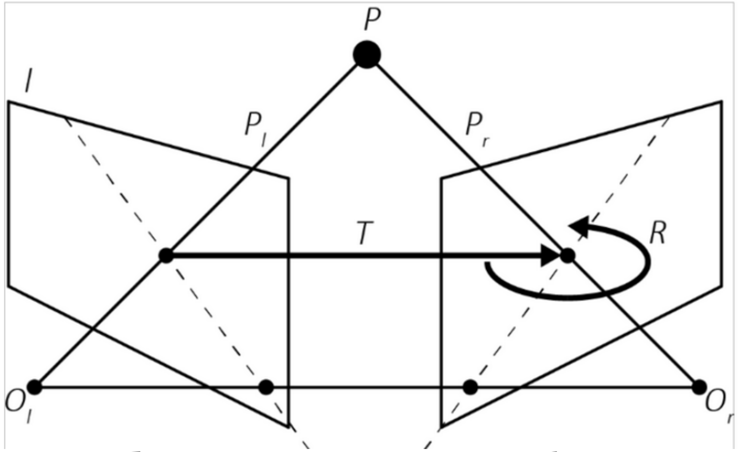
$$Z = \frac{fT}{x_l - x_r}$$

Układ rzeczywisty - geometria epipolarna.



- Środkami projekcji kamer są O_1 i O_2, a ich płaszczyznami rzutowymi l i r.
- Punkty p_1 i p_r są rzutem punktu P ze świata rzeczywistego na płaszczyzny rzutowe l i r.
- Punkty epipolarne e_l i e_r są rzutem środka projekcji drugiej kamery na płaszczyznę,
- Linie epipolarne l_1 i l_2 łączące środki projekcji i punkty epipolarne
- Płaszczyzna epipolarna wyznaczają punkty P, e_l i e_r.

Macierz zasadnicza



- **Macierz zasadnicza** E wiąże położenie obu kamer w stosunku do siebie za pomocą translacji i rotacji we współrzędnych fizycznych.
- Ponieważ przekształcenie jest złożeniem rotacji R i translacji T $P_l = R(P_r - T)$,
- Macierz zasadniczą możemy opisać jako:

$$E = R \cdot S = R \cdot \begin{bmatrix} O & -T_x & T_y \\ T_z & 0 & -T_x \\ -T_y & T_x & 0 \end{bmatrix}$$ - S opisuje translację.

Macierz fundamentalna

- Macierz zasadnicza opisuje położenie kamer względem siebie,
- Często potrzebne jest połączenie puntów jednego obrazu z drugim ze współrzędnych pikselowych - służy do tego **macierz fundamentalna**,
- tworzy się ją na podstawie macierzy zasadniczej E i macierzy parametrów wewnętrznych kamery M

Kalibracja stereo

- Proces kalibracji będzie polegał na dostarczeniu wystarczającej ilości zdjęć, z góry znanej planszy kalibracyjnej z oznaczonymi jej punktami charakterystycznymi.
- Celem tego procesu będzie przybliżenie nieznanych parametrów wewnętrznych i zewnętrznych kamer, mając tylko zestaw zdjęć z odpowiednim wzorem kalibracyjnym

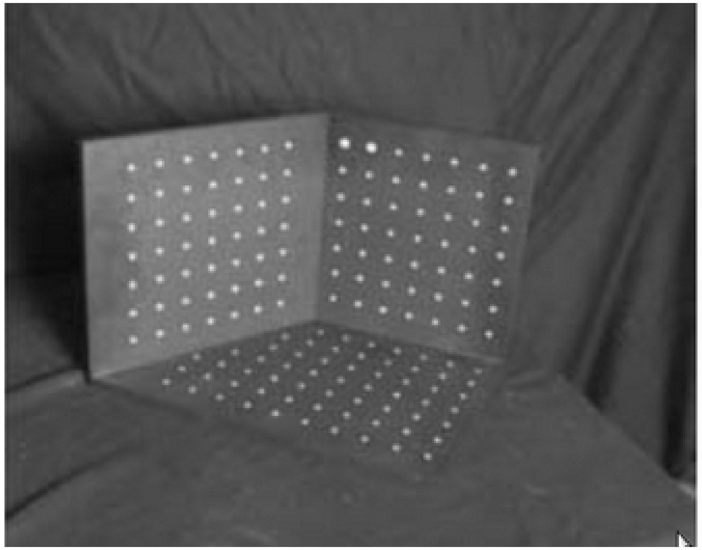
Proces kalibracji

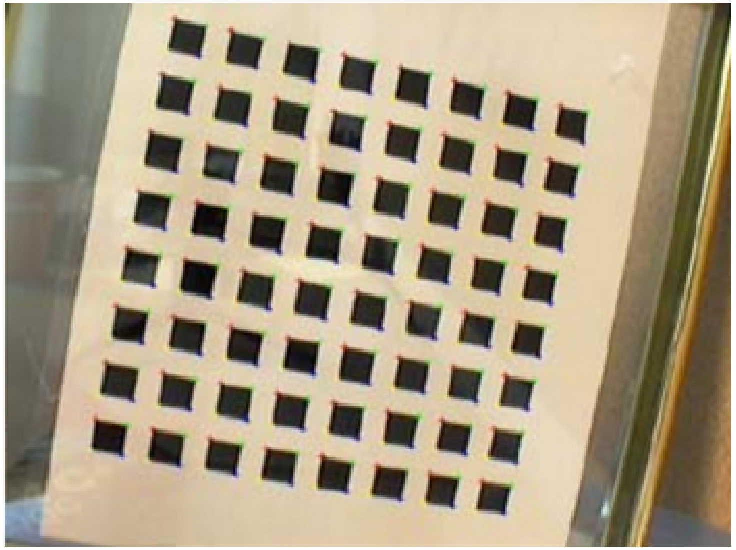
parametry, które są potrzebne, by zdefiniować cechy kamery oraz umiejscowienie kamer względem siebie

- Parametry macierzy wewnętrznej aparatu (f_x, f_y, c_x, c_y),
- Parametry zniekształceń
  - Radialne (k_1, k_2, k_3),
  - Tangensowe (p_1, p_2),
- Macierz rotacji – parametry rotacji (r_1, r_2, r_3),
- Wektor translacji – parametry translacji (T_x, T_y, T_z).

Proces kalibracji będzie polegał na dostarczeniu wystarczającej ilości zdjęć, z góry znanej planszy kalibracyjnej z oznaczonymi jej punktami charakterystycznymi.

Kalibracja wewnętrzna geometryczna





- **Wzorce kalibracyjne -** użycie wzorca kalibracyjnego lub zestawu markerów jest jednym z bardziej wiarygodnych sposobów oszacowania wewnętrznych parametrów aparatu,
- **Planarne wzorce kalibracyjne -** dobrym sposobem wykonania kalibracji jest przesuwanie planarnego celu kalibracji w kontrolowany sposób przez przestrzeń roboczą.

Proces kalibracji stereo

- Dla każdego obrazu szachownicy musimy określić homografię jaką posiada, czyli jej rzut na układ obrazowania aparatu.
- Homografia będzie się składała z dwóch kolumn macierzy rotacji r_1 i r_2, wektora translacji t, macierzy parametrów wewnętrznych kamery M i współczynnika skali s.

$$H = [h_1, h_2, h_3] = s \cdot M \cdot [r_1, r_2, t]$$

- czyli:
  $h_1 = sMr_1$
  $h_2 = sMr_2$
  $h_3 = sMr_3$

- Macierz rotacji jest ortogonalna, po unormowaniu otrzymujemy $r_1^T \cdot r_2 = 0$ oraz $|r_2| = |r_2| \Rightarrow r_1^T r_1 = r_2^T r_2$

Rektyfikacja

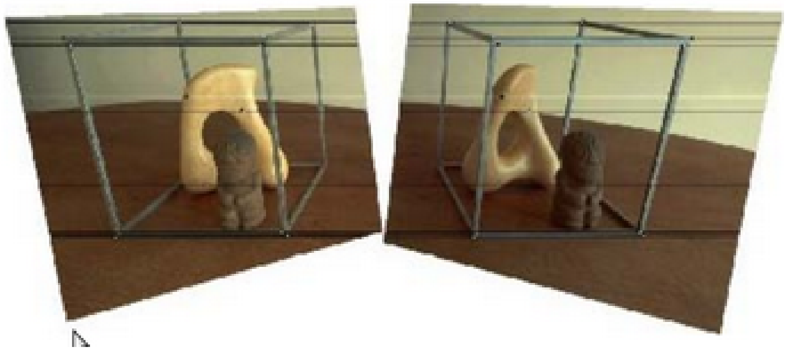
- Celem rektyfikacji jest sprowadzenie obrazów do takiego układu, jakby były zrobione za pomocą idealnego układu frontowo-równoległego,
- W bibliotece OpenCV są dostępne dwie strategie rektyfikowania obrazu, algorytm Hartleya oraz algorytm Bougueta.

Raktyfikacja

Geometria epipolarna dla pary kamer jest niejawna we względnej pozycji i kalibracji kamer i może być obliczona na podstawie dopasowań punktowych.

- Jednym ze sposobów jest użycie ogólnego algorytmu korespondencji, takiego jak *przepływ optyczny* ang. *optical flow*
- Bardziej wydajny algorytm można uzyskać korygując obrazy wejściowe tak, aby odpowiadające poziome linie skanowania były liniami epipolarnymi,
- Następnie można niezależnie dopasować poziome linie skanowania lub przesuwać obrazy w poziomie podczas obliczania pasujących wyników.

Rektyfikacja - przykład



- Prostym sposobem na poprawienie dwóch obrazów jest najpierw obrócenie obu kamer tak, aby były ustawione prostopadle do linii łączącej środki kamer c_0 i c_1,
- Następnie, aby określić pożądany skręt wokół osi optycznych, wykonaj wektor w górę prostopadle do linii środkowej kamery.
- Na koniec przeskaluj obrazy, jeśli to konieczne, aby uwzględnić różne ogniskowe,

Standardowa rektyfikowana geometria

jest stosowany w wielu konfiguracjach kamer stereo i algorytmach stereo i prowadzi do bardzo prostej odwrotnej zależności między głębią 3D Z a różnicami d:

$$d = f\frac{B}{Z}$$

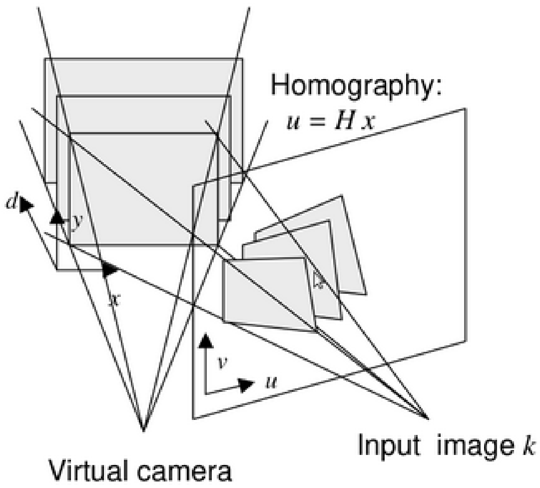
gdzie f to ogniskowa (mierzona w pikselach), B to linia bazowa.

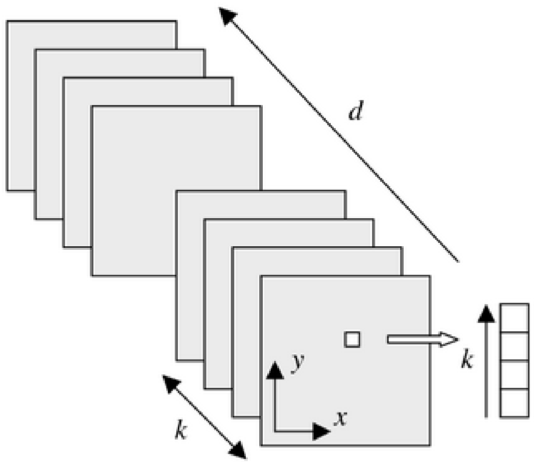
- Zależność między odpowiednimi współrzędnymi pikseli na lewym i prawym obrazie:

$$x' = x + d(x, y), y' = y$$

- Zadanie wydobycia głębi ze zbioru obrazów staje się następnie zadaniem oszacowania **mapy dysproporcji** $d(x, y)$.
- Po poprawieniu możemy łatwo porównać podobieństwo pikseli w odpowiednich lokalizacjach (x, y) i $(x', y') = (x + d, y)$.

Przemiatanie płaszczyzny





Alternatywą dla wstępnej rektyfikacji obrazów przed dopasowaniem jest **przeciągnięcie zestawu płaszczyzn** przez scenę i zmierzenie fotokonsystencji obrazów podczas ich wyświetlania na płaszczyźnie.

- Zestaw płaszczyzn widzianych z kamery wirtualnej wywołuje zestaw homografii w obrazie z dowolnej innej kamery źródłowej (wejściowej).
- Wypaczone obrazy ze wszystkich innych kamer mogą być układane w stos w uogólnionej przestrzeni dysproporcji $\tilde{I}(x, y, d, k)$, różnica d kamery k, (x, y) - lokalizacja piksela

Algorytm Hartleya

- Do obliczenia macierzy fundamentalnej, potrzebny jest tylko jeden zestaw zdjęć,
- Algorytm ten jako danych początkowych potrzebuje macierzy fundamentalnej i zestawu korespondujących punktów z obrazów. W pierwszym kroku oblicza punkty epipolarne: $Fe_l = 0$ i $e_l^T = 0$
- Następnym krokiem w algorytmie, jest przeniesienie punktu epipolarnego do nieskończoności do $e_r = (k, 0, 1)^T$ Wykonuje się to mnożąc przez macierz:

$$G = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ \frac{-1}{k} & 0 & 1 \end{bmatrix}$$

- Następnie, stosuje się translację T, by przenieść obraz do początku okładu współrzędnych i rotację R, by ustawić wszystkie linie epipolarne równolegle do osi X.

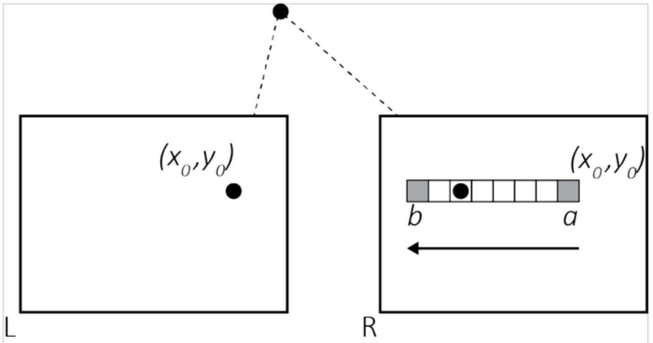
Algorytm Bougueta

- Algorytm Bougeta znając rotację i translację układu kamer, ma za zadanie sprowadzić obrazy kamer do idealnego układu,
- Danymi wejściowymi jest macierz rotacji R i wektor translacji T, które są dane po przeprowadzeniu kalibracji.

Korespondencja stereo

- Celem korespondencji obrazów, jest znalezienie możliwie największego wspólnego obszaru na parze obrazów.
- Takie obszary można znaleźć tylko na obrazach, na których widoki z obu aparatów nachodzą na siebie.

Algorytm Block Matching



- Szybki algorytm wykorzystujący mechanizm sum różnic bezwzględnych,
- Pierwszym krokiem, jest zlikwidowanie ekstremalnie jasnych lub ciemnych pikseli.
- Nad każdym pikselem zatrzymuje się okienko i zbiera informacje o otoczeniu. Oblicza średnią jasność pikseli w okienku i zmienia wartość środkowego piksela na:

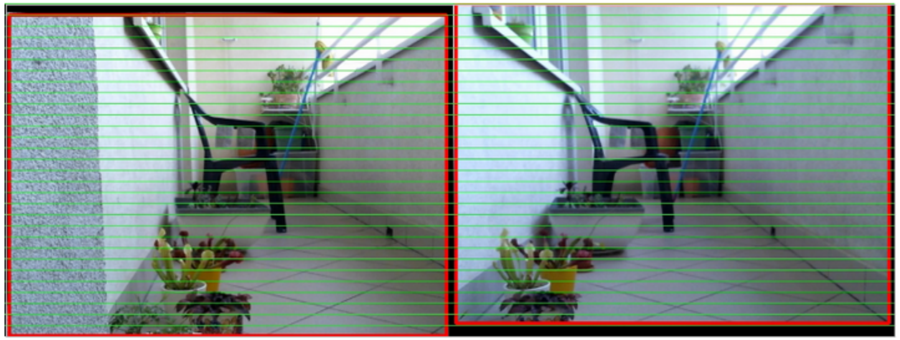
$$\min(\max(I_c - I, I_{cap}), I_{cap})$$

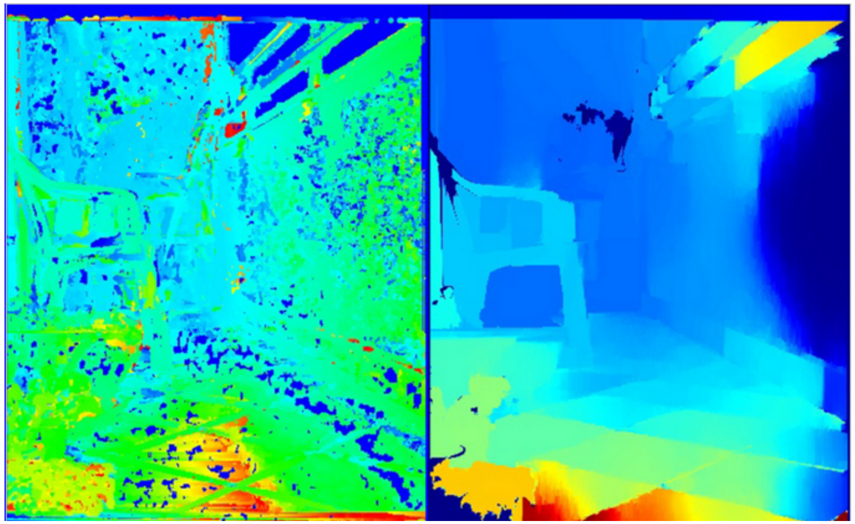
gdzie I_c jest jasnością środkowego piksela, a I_{cap} granicą jasności.

Wyniki prawidłowej kalibracji

- Odwzorowanie mapuje punkt (x, y) z oryginalnego obrazu na punkt (x', y') na docelowym obrazie, lecz nie musi to być punkt o współrzędnych całkowity
- Stosuje się odwzorowanie odwrotne (and. reverse mapping):
  - wybraniu na docelowym obrazie piksela, którego chcemy policzyć,
  - określeniu jakie miałby współrzędne po likwidacji zniekształceń, a przed rektyfikacją,
  - odszukaniu go na oryginalnym obrazie.

Wyniki prawidłowej rektyfikacji





- Zielone linie pokazują linie epipolarne.
- Pary zdjęć w palecie kolorystycznej to mapy głębi.