

# 第1章 画像理解の基礎

本章では3次元空間(以降, 3D空間と記述する)中の対象物をカメラで観測して2次元画像情報を得るときの射影モデルについて解説する. 次いで2次元画像情報から元の3次元情報を復元するための画像理解の理論について解説する.

## 1.1 3D空間から画像への射影

カメラで3D空間中の対象物を観測するときのモデルについて説明する.

### 1.1.1 中心射影

ある対象物をレンズを通して観測し, 図 1.1 のように, レンズ後方  $b$  の距離の位置に像を結ぶとき, レンズの公式により次の関係が成り立つ.

$$\frac{1}{a} + \frac{1}{b} = \frac{1}{f} \quad (1.1)$$

ただし,  $a$  は対象物のレンズ前方の位置で,  $f$  はレンズの焦点距離である.  $b$  に対して  $a$  が十分

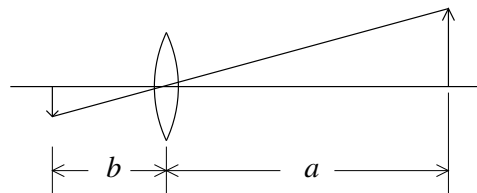


図 1.1: カメラ撮像モデル

大きいときには左辺第1項は無視でき次の近似

$$\frac{1}{b} \simeq \frac{1}{f} \quad (1.2)$$

が成り立ち, 撮像面の位置  $b$  は焦点距離  $f$  で近似できる. このような近似ができるとき, 仮想的な撮像面をレンズの前方  $f$  の距離のところを設定すると図 1.2 のような中心射影(透視射影)のモデルが得られる.

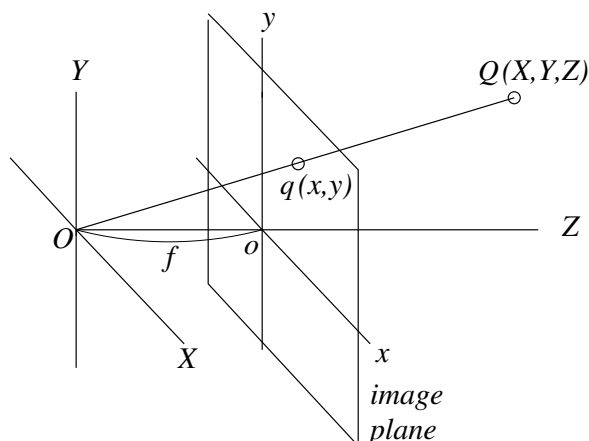


図 1.2: 中心射影

図 1.2 において,  $O$  はカメラレンズ中心を示し,  $z$  軸はレンズの中心を通りレンズ面に直交する直線であり, カメラの光軸と呼ばれる.  $O$  を原点とする座標系をカメラ座標系と呼ぶ. 図 1.2 における  $(X, Y, Z)$  は 3D 空間中のある対象物  $Q$  のカメラ座標である. 中心射影のモデルでは図 1.2 のように画像面は  $O$  から前方  $f$  の距離のところであり, 画像の座標軸  $x$  軸,  $y$  軸はカメラ座標系の  $X$  軸,  $Y$  軸とそれぞれ平行である. 対象物  $Q$  の画像面での射影は直線  $OQ$  と画像面の交点  $q$  で与えられ, この  $q$  の画像座標  $(x, y)$  は次式で与えられる.

$$x = f \frac{X}{Z}, \quad y = f \frac{Y}{Z} \quad (1.3)$$

式 1.3 のような対象物  $Q$  のカメラ座標  $(X, Y, Z)$  から射影の画像座標  $(x, y)$  への変換を透視変換 (perspective transformation) と呼ぶ. この変換は非線形である. 透視変換のような非線形変換では図 1.3 に示すように, 3D 空間において平行な 2 直線を観測するとき, 画像中で必ずしも平行とはならない.

3D 空間において単位ベクトル  $\mathbf{u} = (u_1, u_2, u_3)^t$  の方向を向いた平行線は図 1.4 に示すように画像内で平行でない 2 直線として観測され, 1 点  $q$  で交わる. ただし, 記号  $t$  は転置を表す. この交点を消失点と呼ぶ. この消失点に関して次の命題が成り立つ.

**[命題 1.1]** 3D 空間において  $\mathbf{u}$  の方向を向いた平行線が画像で観測されたときの消失点を  $q$  とする. カメラ座標系の原点  $O$  と画像内の点  $q$  のカメラ座標  $(x, y, f)$  を結ぶ単位ベクトルは  $\mathbf{u}$  である [1].

**(証明)** 点  $(X_0, Y_0, Z_0)$  を通り,  $\mathbf{u}$  の方向を向いた 3D 空間における直線はパラメータ  $t$  を用いて, 次のように表される.

$$X = X_0 + tu_1, \quad Y = Y_0 + tu_2, \quad Z = Z_0 + tu_3 \quad (1.4)$$

この直線が画像で観測される時, 透視変換により次の画像内の直線となる.

$$x = f \frac{X}{Z} = f \frac{X_0 + tu_1}{Z_0 + tu_3},$$



図 1.3: 3D 空間における平行線の中心射影像

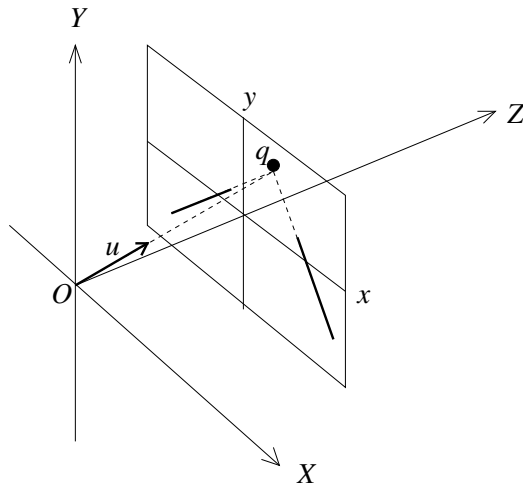


図 1.4: 消失点

$$y = f \frac{Y}{Z} = f \frac{Y_0 + tu_2}{Z_0 + tu_3} \quad (1.5)$$

$t \rightarrow \infty$  のとき, 任意の  $(X_0, Y_0, Z_0)$  に対して次式が得られる.

$$x = fu_1/u_3, \quad y = fu_2/u_3 \quad (1.6)$$

これが消失点の座標である. この点をカメラ座標系で表すと次のようになる.

$$\left( f \frac{u_1}{u_3}, f \frac{u_2}{u_3}, f \right) \quad (1.7)$$

カメラ座標系の原点  $O$  とこの点を結ぶ直線方向の単位ベクトルは  $(u_1, u_2, u_3)^t$  となることがわかる.

## 第2章 パターン認識の基礎

本章では、まず画像情報から認識に適切な特徴を抽出するためのツールである画像処理の基本的な手法について述べる。特徴空間の効率的次元圧縮法について解説し、特徴空間に基づいてカテゴリ分類（識別）するパターン認識の理論について解説する。更に、カメラの視点変動に伴う画像中の対象物の幾何学的変形にロバストな認識を行うのに有効な不変量について解説する。最後に、特徴空間において分布する特徴の学習データからカテゴリの境界を決める識別関数を構築する学習法について解説する。

### 2.1 画像処理手法

デジタル画像処理について述べる。デジタル画像とは図 2.1 に示すように、空間的離散化と量的離散化を施された画像をいう。空間的離散化を標本化といい、このときの最小単位を画素 (pixel) という。ある画素でのアナログ信号、例えば、白黒の濃淡画像の場合なら濃淡値 (明度、輝度ともいう) を A/D 変換器により離散化することを量子化という。例えば、濃淡値を 8bit で量子化するとき、画素値は 0 ~ 255 の 256 段階で表される。

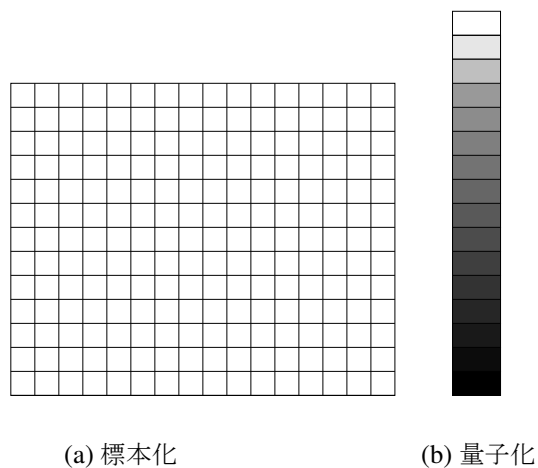


図 2.1: 画像の離散化

### 2.1.1 平滑化フィルタ

注目画素を中心とする窓を設定し、その窓内の画素値平均をその注目画素の画素値とすることにより、画像を平滑化し、ノイズを除去できる。図 2.2 のように注目画素のデジタル画像座標を  $(x, y)$  とし、平滑化フィルタを  $w(i, j)$  とし、フィルタ処理前のデジタル画像を  $f(x, y)$  とするとき、平滑化フィルタ出力  $\bar{f}(x, y)$  は次式で与えられる。

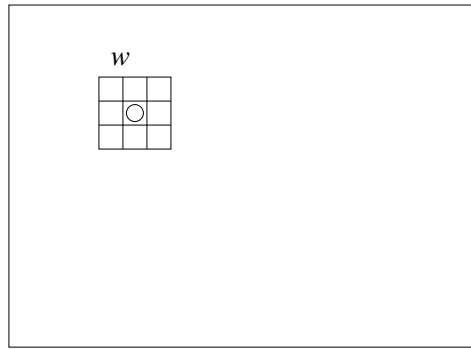


図 2.2: 平滑化フィルタ : ○印は注目画素

$$\bar{f}(x, y) = \frac{\sum_{i=-s}^s \sum_{j=-s}^s w(i, j) f(x+i, y+j)}{\sum_{i=-s}^s \sum_{j=-s}^s w(i, j)} \quad (2.1)$$

ただし、 $s$  は窓のサイズを  $2s+1$  としたときのサイズパラメータである。  $w(i, j) = 1$  の単純なフィルタや窓内の重みを中心に近い程、大きくする Gaussian フィルタなどがある。 Gaussian フィルタのときは  $w(i, j)$  は次のように与えられる。

$$w(i, j) = \exp\left(-\frac{i^2 + j^2}{2\sigma^2}\right) \quad (2.2)$$

$\sigma^2$  は分散を表す。このフィルタは画像をボカすのに使用され、 $\sigma^2$  を大きくするとボケの度合は大きくなる。

例として、ノイズを含む図 2.3(a) の画像に単純平滑化フィルタ処理を施した画像を図 2.3(b) に示す。図 2.3(a) の車両左下付近に人為的に付加された 4 点のノイズが図 2.3(b) では目立たなくなっているのが分かる。また、 $\sigma^2$  を変えたときの Gaussian フィルタ処理結果を図 2.4(b),(c) に示す。 $\sigma^2$  が大きい程、ボケがひどくなっているのが分かる。

### 2.1.2 エッジフィルタ

画像情報の内、一部の特徴的なもののみを抽出したいときがある。例えば、人間の顔のように輪郭のみを得たいとき、エッジフィルタが使われる。すなわち、このフィルタは濃淡値が急



(a) ノイズ画像

(b) 平滑化後

図 2.3: ノイズ処理



(a) 原画像

(b)  $\sigma = 0.5$

(c)  $\sigma = 2.0$

図 2.4: Gaussian フィルタ処理

激に変化する部分を捉えるフィルタである. エッジフィルタとしてよく用いられるソーベル (Sobel) フィルタは図 2.5 に示すような  $3 \times 3$  のフィルタである.

$x$  方向微分に相当するフィルタ  $D_x$  と  $y$  方向微分に相当するフィルタ  $D_y$  を用いて次のよう

$$\begin{array}{c}
 j \\
 \uparrow \\
 \begin{array}{|c|c|c|}
 \hline
 -1 & 0 & 1 \\
 \hline
 -2 & 0 & 2 \\
 \hline
 -1 & 0 & 1 \\
 \hline
 \end{array}
 \end{array}
 \quad
 \begin{array}{|c|c|c|}
 \hline
 1 & 2 & 1 \\
 \hline
 0 & 0 & 0 \\
 \hline
 -1 & -2 & -1 \\
 \hline
 \end{array}$$

(a)  $D_x$

(b)  $D_y$

図 2.5: ソーベルフィルタ

にソーベルフィルタ出力が得られる.

$$g_x(x, y) = \sum_{i=-1}^1 \sum_{j=-1}^1 D_x(i, j) f(x+i, y+j)$$

## 第3章 画像理解・パターン認識の最近の 動向

本章では、前2章で解説した理論を応用した、あるいは発展させた、新規性を有する興味深い研究事例について紹介することにより、画像理解・パターン認識の最近の動向について展望する。

### 3.1 多重特徴を用いる Bayesian Filter によるロバストなトラッキング [33]

移動物体を追跡するトラッキング問題においては、日照条件の変化や追跡対象物の位置の変化による背景変化に影響されて、ビデオ画像内の背景から追跡対象物を領域分割することが困難である。この困難を克服するために、互いに条件付確率をとおして依存する多重特徴を用いてロバストなトラッキングを可能とする手法が開発された [33]。この手法を紹介する。

#### 3.1.1 数学的準備

##### 多重特徴の統合

追跡対象物は  $F$  個の特徴集合によって記述されるものとする。  $F$  個の特徴集合は状態ベクトル  $\mathbf{x}_1 \dots \mathbf{x}_F$  から構成される。  $\mathbf{x}_i$  は  $\mathbf{x}_{i-1}$  に依存するものとする。状態ベクトル  $\mathbf{x}_i, i = 1, \dots, F$  は観測ベクトル  $\mathbf{z}_i, i = 1, \dots, F$  に対応づけられる。記号  $\mathbf{X}_{1:k} \equiv \{\mathbf{x}_1, \dots, \mathbf{x}_k\}, \mathbf{Z}_{1:k} \equiv \{\mathbf{z}_1, \dots, \mathbf{z}_k\}$  を導入する。また、条件付確率  $p_i \equiv p(\mathbf{x}_i | \mathbf{x}_1, \dots, \mathbf{x}_{i-1}, \mathbf{z}_1, \dots, \mathbf{z}_i)$  を定義する。このとき、次の関係が成立する。

$$\begin{aligned} P &= p(\mathbf{X}_{1:F} | \mathbf{Z}_{1:F}) \\ &= p(\mathbf{x}_1 | \mathbf{Z}_1) p(\mathbf{x}_2 | \mathbf{X}_1, \mathbf{Z}_{1:2}) \dots p(\mathbf{x}_F | \mathbf{X}_{1:F-1}, \mathbf{Z}_{1:F}) \\ &= p_1 p_2 \dots p_F \end{aligned} \quad (3.1)$$

時刻表示をするために、時刻  $t$  での  $p_i$  を次式で定義される  $p_i^t$  で置き換える。

$$p_i^t \equiv p(\mathbf{x}_i^t | \mathbf{X}_{1:i-1}^t, \mathbf{Z}_{1:i}^t, p_i^{t-1}) \quad (3.2)$$

このとき、次の関係が成り立つ。

$$\begin{aligned}
 P^t &= p(\mathbf{x}_1^t, \dots, \mathbf{x}_F^t | \mathbf{z}_1^t, \dots, \mathbf{z}_F^t) \\
 &= p(\mathbf{x}_1^t | \mathbf{z}_1^t, p_1^{t-1}) \dots p(\mathbf{x}_F^t | \mathbf{X}_{1:F-1}^t, \mathbf{z}_{1:F}^t, p_F^{t-1}) \\
 &= p_1^t p_2^t \dots p_F^t
 \end{aligned} \tag{3.3}$$

### Bayesian filter

$k$  番目の Bayesian filter が事後確率  $p(\mathbf{x}_k^t | \mathbf{Z}_k^{t_0:t})$  をどのように計算するのかを説明する。簡単のため、ここでは  $\mathbf{x}_k$  は  $\mathbf{x}_i, i < k$  に依存しないと仮定する。この仮定を除いた説明は次の section で行う。

Bayesian filter によるトラッキング問題の定式化は”hypothesis generation-hypothesis correction”の段階をとおして記述される。

#### (Hypothesis generation)

ターゲットの状態は以下のように、dynamic model  $p(\mathbf{x}_k^t | \mathbf{x}_k^{t-1})$  と前時刻  $t-1$  での事後確率  $p(\mathbf{x}_k^{t-1} | \mathbf{Z}_k^{t_0:t-1})$  に基づいて予測される。

$$p(\mathbf{x}_k^t | \mathbf{Z}_k^{t_0:t-1}) = \int_{\mathbf{x}_k^{t-1}} p(\mathbf{x}_k^t | \mathbf{x}_k^{t-1}) p(\mathbf{x}_k^{t-1} | \mathbf{Z}_k^{t_0:t-1}) d\mathbf{x}_k^{t-1} \tag{3.4}$$

この予測 (likelihood) は次の hypothesis correction の段階で更新される。

#### (Hypothesis correction)

observation model  $p(\mathbf{z}_k^t | \mathbf{x}_k^t)$  によって、時刻  $t$  での事後確率  $p(\mathbf{x}_k^t | \mathbf{Z}_k^{t_0:t})$  が次式で与えられる。

$$p(\mathbf{x}_k^t | \mathbf{Z}_k^{t_0:t}) = \alpha^t p(\mathbf{z}_k^t | \mathbf{x}_k^t) p(\mathbf{x}_k^t | \mathbf{Z}_k^{t_0:t-1}) \tag{3.5}$$

ただし、 $\alpha^t$  は正規化定数である。

Bayesian filter を実行する方法として代表的なものとして Kalman filter と particle filter がある。これらについて以下に説明する。

#### a. Kalman filter

dynamic model  $p(\mathbf{x}_k^t | \mathbf{x}_k^{t-1})$  と observation model  $p(\mathbf{z}_k^t | \mathbf{x}_k^t)$  がガウス分布に従うと仮定する。

$$p(\mathbf{x}_k^t | \mathbf{x}_k^{t-1}) = N(H_k^t \mathbf{x}_k^{t-1}; \Sigma_{k,h}^t) \tag{3.6}$$

$$p(\mathbf{z}_k^t | \mathbf{x}_k^t) = N(M_k^t \mathbf{x}_k^t; \Sigma_{k,m}^t) \tag{3.7}$$

ただし、 $H_k^t$  と  $M_k^t$  はモデルの確定要素から成る行列である。 $\Sigma_{k,h}^t$  と  $\Sigma_{k,m}^t$  はモデルに対して仮定される正規分布する white noise の共分散行列である。式 3.6, 3.7 を Bayesian filter 3.4, 3.5 に代入することにより、以下の hypothesis generation と hypothesis correction の段階でのそれ



### 3.3 エピポーラ制約に優る幾何制約による移動物体領域の検知 [36]

画像内の移動物体領域を検知する方法として、2 時点での移動カメラから捉えた 2 画像からエピポーラ制約を用いる方法があるが、この場合、カメラと移動物体が互いに平行運動するときには移動物体領域の検知が不可能になる。このような欠点を克服する手法が Yuan 等 [36] によって開発された。この手法は 3 時点での移動カメラから捉えた 3 画像と reference plane に基づくもので以下にこのアルゴリズムを紹介する。

#### 3.3.1 移動物体領域検知の流れ

時点  $i, i = 1, 2, \dots$  での移動カメラから 3D 世界座標空間中の 1 点  $P(x, y, z)$  を捉えたときの 3 次元カメラ座標を  $P_i(x_i, y_i, z_i)$ , 2 次元画像斉次座標を  $\mathbf{p}_i(u_i, v_i, 1)$  とする。2 時点  $i, i+1$  での reference plane を  $\Pi_{i,i+1}$  で表わす。時点  $i = 1, 2$  の場合を考える。このとき、点  $P$  が  $\Pi_{12}$  上の点ならば  $\mathbf{p}_2$  を  $\mathbf{p}_1$  と関係づける Homography  $H_{12}$  によって  $\mathbf{p}_1$  は  $H_{12}\mathbf{p}_2$  で与えられる。このような点对応での画像 intensity が一致すれば点  $P$  は reference plane  $\Pi_{12}$  上の静止点として planar pixels として扱われる。さもなければ移動点として residual pixels として扱われる。以上の process が図 3.5 に示すように、Initial detection で Homography based detection ともいわれる。

次に residual pixels が本当に移動点なのかどうかについて調べるのが Parallax filtering で、この process は Epipolar 制約が満たされているかどうかを調べる段階と Structure consistency 制約が満たされているかどうかを調べる段階の 2 段階から成る。Epipolar 制約が満たされていないならば移動点とみなし motion regions に属すると判定する。さもなければ、Structure consistency 制約が満たされれば reference plane  $\Pi_{12}$  に属さない静止点 Parallax pixels とみなし、さもなければ移動点 motion regions に属すると判定する。Structure consistency 制約は 3 時点での画像  $\mathbf{p}_1, \mathbf{p}_2, \mathbf{p}_3$  と  $\Pi_{12}, \Pi_{23}$  に基づいて構成される。

#### 3.3.2 移動物体検知のための諸定義

世界座標系での静止点  $\mathbf{P}(x, y, z)$  を時点  $i$  での移動カメラで捉えたときの 3 次元カメラ座標  $\mathbf{P}_i$  はカメラの世界座標系に対する姿勢と位置を示す回転行列  $R_i$ , 並進ベクトル  $\mathbf{T}_i$  を用いて  $\mathbf{P}$  と次式で関係づけられる。

$$\mathbf{P}_i = R_i\mathbf{P} + \mathbf{T}_i \quad (3.77)$$

以下の議論で、 $R_1 = I, \mathbf{T}_1 = \mathbf{0}$  を仮定する。

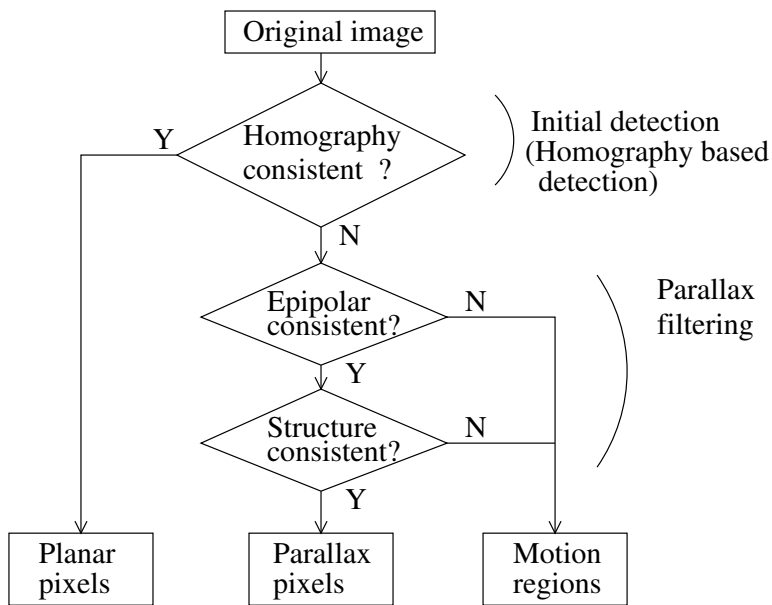


図 3.5: 移動物体領域検知のフローチャート

点  $\mathbf{P}$  を時点  $i$  でのカメラで観測したときの画像斉次座標  $\mathbf{p}_i$  は透視射影によって次式のように与えられる.

$$\mathbf{p}_i = K_i \mathbf{P}_i / z_i \quad (3.78)$$

ただし,  $K_i$  は時点  $i$  でのカメラ内部パラメータを示す行列である.

### 3.3.3 エピポーラ制約の限界

静止点  $\mathbf{P}$  を 2 時点での移動カメラで観測したときの画像斉次座標  $\mathbf{p}_1, \mathbf{p}_2$  は基礎行列 (Fundamental matrix)  $F_{12}$  によって次式のように関係づけられる.

$$\mathbf{p}_2^T F_{21} \mathbf{p}_1 = 0 \quad (3.79)$$

これがエピポーラ制約といわれるものである. 点  $\mathbf{p}$  が静止点ならば式 3.79 を満たす. 従って, その対偶をとれば, 式 3.79 が満足されなければ点  $\mathbf{P}$  は静止点ではなく, 移動点となる.

しかし, 点  $\mathbf{P}$  が移動カメラ視点  $C_1, C_2$  で決まるエピポーラ面内を並進運動するときには図 3.6 に示すように点  $\mathbf{P}$  は静止点でないにも拘わらず, エピポーラ制約を満たす. 従って, エピポーラ制約を満たす場合には全てが静止点という訳ではなく, 移動する点であるかもしれないので, この場合には, 更に Structure consistency 制約による判定を行う必要がある.

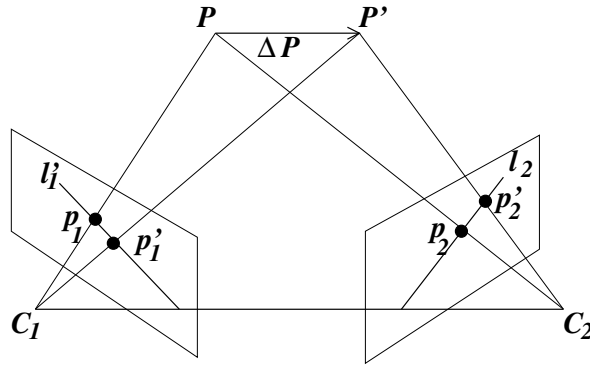


図 3.6: 点  $\mathbf{P}$  の並進  $\Delta\mathbf{P}$  がカメラ中心  $C_1$  と  $C_2$  のエピポーラ面内にあるときのエピポーラ線  $l_1$  と  $l_2$

エピポーラ制約からのズレとして次式で示す  $d_{epi}$  を定義する.

$$d_{epi} \equiv (|\ell'_1 \cdot \mathbf{p}_1| + |\ell_2 \cdot \mathbf{p}'_2|)/2 \quad (3.80)$$

ただし,  $\ell'_1$  は時刻 1 のときのカメラの画像でのエピポーラ線  $\ell'_1 = F_{21}^T \mathbf{p}'_2$  で,  $\ell_2$  は時刻 2 のときのカメラの画像でのエピポーラ線  $\ell_2 = F_{21} \mathbf{p}_1$  である.  $|\ell'_1 \cdot \mathbf{p}_1|$  は点  $\mathbf{p}_1$  からエピポーラ線  $\ell'_1$  までの垂直距離で,  $|\ell_2 \cdot \mathbf{p}'_2|$  は点  $\mathbf{p}'_2$  からエピポーラ線  $\ell_2$  までの垂直距離である.

### 3.3.4 Structure Consistency 制約

3D 空間中の reference plane  $\Pi$  上の点は方程式:

$$\mathbf{N} \cdot \mathbf{P} = d \quad (3.81)$$

を満たす. ただし,  $\mathbf{N} = (N_x, N_y, N_z)^T$  は  $\Pi$  の面方位で,  $d$  は世界座標系の原点から  $\Pi$  までの距離である.  $\Pi$  上にない点  $\mathbf{P}$  の  $\Pi$  からの距離  $H$  は

$$H = \mathbf{N} \cdot \mathbf{P} - d \quad (3.82)$$

で与えられる.  $\gamma$  を

$$\gamma \equiv H/z \quad (3.83)$$

で定義する.  $\Pi$  上にない静止点  $\mathbf{P}_0$  が与えられたとする.  $\mathbf{P}_0$  の  $\gamma$  を  $\gamma_0$  とする.  $\gamma$  の  $\gamma_0$  に対する比として

$$k \equiv \gamma/\gamma_0 = \frac{z_0}{H_0} \cdot \frac{H}{z} \quad (3.84)$$

を定義する.  $k$  を projective depth と呼ぶ.