

1. 画像への結像と光学系の記述

1.1 カメラ

カメラでは、前方の3次元の世界がレンズを通して2次元、すなわち平面の画像へと投影され、像を結ぶ。どのような形の物体が前方の3次元空間にあるかが決まれば、その像は幾何光学と呼ばれる法則で、完全に決定される。 画像のできる法則を定式化する。

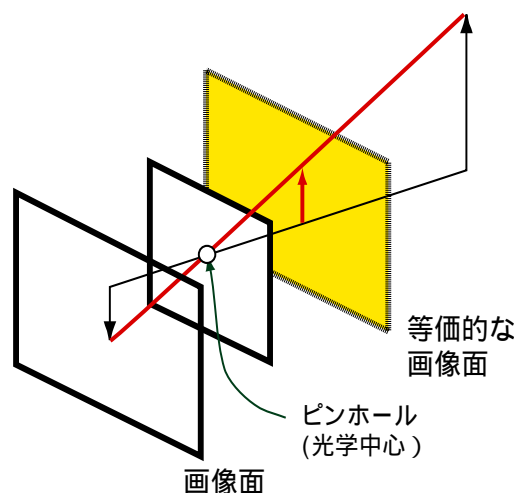
そして、以下のことを目標とする。

- カメラが運動をするときに、画像がどのように変化するかを明らかにすること
- 逆に、画像の変化からカメラの運動を導くこと、
- また、カメラでとらえたある2次元の画像から、カメラ前方の3次元世界を再現すること

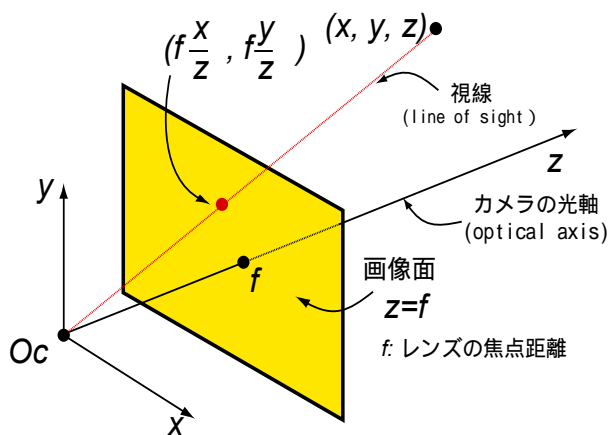
1.2 ピンホールカメラモデル

ピンホールカメラ(pin-hole camera):対象とその画像への結像の位置関係をいちばん単純に理想化したモデル。

これは、理想的な単レンズによる結像の幾何学的なモデルである。このピンホールの位置を、光学中心(optical center)という。図では、レンズの中心にあたる。空間の点は、そこから光学中心を通る直線(視線(line of sight))と呼ぶ)に沿って進んで、画像面と交わったところに像を結ぶ。ピンホールカメラでは、画像が逆さまにできるので、図のように画像面を配置し直したほうがわかりやすい。この配置では、光学中心から画像面を透かして対象を見たように画像ができていると考えられるので、中心投影、または透視投影(perspective projection)と呼ばれている。



数学的に定式化するため、図に示すような、光学中心に原点がある座標系を考える。z軸をレンズの中心を通りレンズ面と直交する方向にとる。これをカメラの光軸(optical axis)と呼ぶ。画像面は、光軸と垂直になるように置かれるものとする。したがって、x-y平面が画像面と平行になる。このような座標系をカメラ座標(camera coordinates)という。さて、光学中心 O_c から画像面までの距離はレンズの焦点距離に相当するので、これを f とおく。画像面は、 $z=f$ と表される。すると、この座標系で、 (x, y, z) にある空間の点の像は、 $(f(x/z), f(y/z), f)$ にできる。このなかの実質的に意味のある下記の座標の変換を透視変換(perspective transformation)と呼ぶ。左辺は空間の点のカメラ座標系での座標、右辺はその点のやはりカメラ座標での画像上の位置である。



$$(x, y, z) \rightarrow \left(f \frac{x}{z}, f \frac{y}{z} \right)$$

この変換は、いろいろな興味ある結果をもたらす。その代表が、カメラからの距離に反比例して対象の像の大きさが小さくなり、その結果、形状がひずむことである。そのひずみを利用することで、形状の奥行きがわかる可能性がある。さらに消失点(vanishing point)、消失線(vanishing line)と呼ぶ、無限に遠い点や地平線のような平面上の無限遠方の像のでき方は興味深い。

2. カメラの位置と姿勢

2.1 画像とカメラの位置・姿勢

ワールド座標(world coordinates):カメラとは独立に、対象の位置、形状など「外部の世界」に基づいた座標系の設定。

T :ワールド座標の原点 O_w からカメラ座標の原点 O_c を示すベクトル。

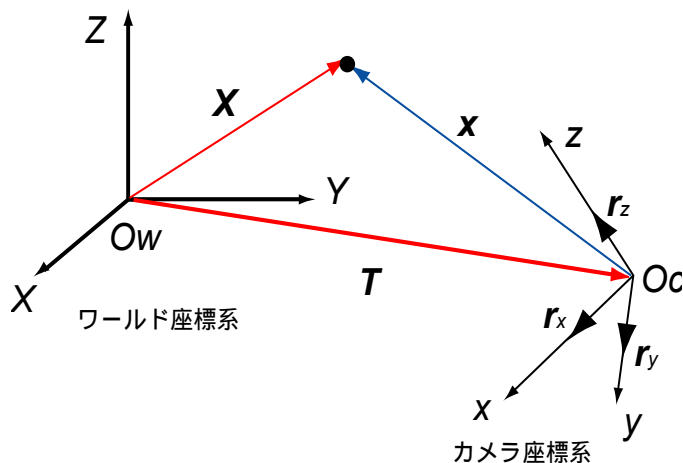
r_x, r_y, r_z : ワールド座標で見たカメラ座標の x, y, z 軸の方向を表す単位ベクトル .

カメラ座標で $x = (x, y, z)^T$ で与えられる点のワールド座標を $X = (X, Y, Z)^T$ とすれば ,両者の関係は

$$X = xr_x + yr_y + zr_z + T$$

となる。

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} r_x^T \\ r_y^T \\ r_z^T \end{pmatrix} X - \begin{pmatrix} r_x^T \\ r_y^T \\ r_z^T \end{pmatrix} T \quad R = \begin{pmatrix} r_x^T \\ r_y^T \\ r_z^T \end{pmatrix}$$



この R を回転行列(rotation matrix)と呼ぶ。また,このときの平行移動のベクトル,すなわち,ワールド座標の原点からカメラの光学中心へ向かうベクトル T が,カメラ座標で表すと $-t$ であったとすると,ワールド座標 $X = (X, Y, Z)^T$ で与えられる点はカメラ座標 $x = (x, y, z)^T$ では

$$x = RX - RT = RX + t$$

となる . T および t を平行移動ベクトル(translation vector)と呼ぶ。

2.2 画像座標

実際のカメラモデルによるワールド座標,カメラ座標,画像座標の関係の記述 .

画像の読込 : 画像面上の座標軸 (u, v) 軸の座標 (u, v) で読み取る。

しかしカメラ座標と画像座標の主要な相違はつぎの三つある。

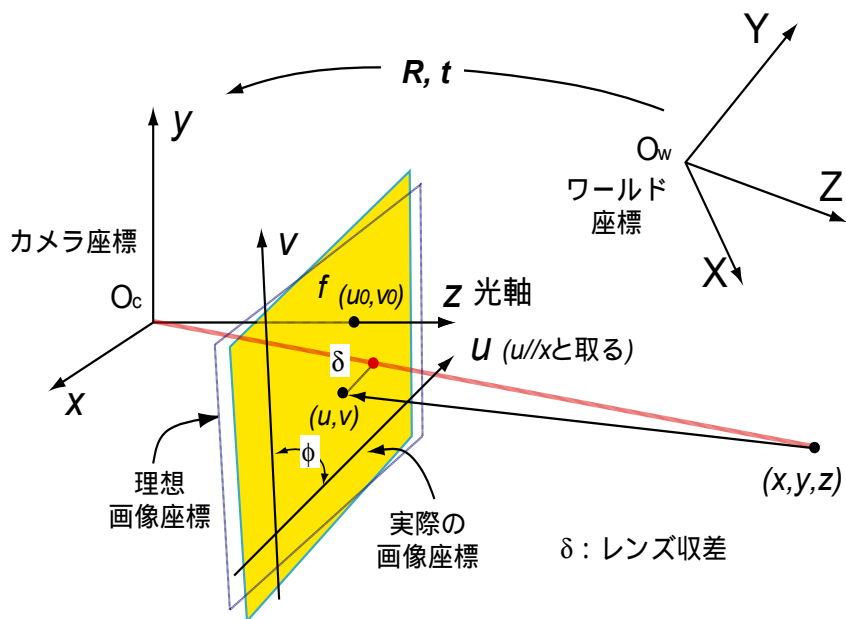
- 光軸すなわちカメラ座標の z 軸は,画像面の中心を通るわけではない(光学的な画像の中心は画像面上での中心ではない)。
- 画像面は厳密には xy 平面と平行ではない。つまり,レンズの光軸(z 軸)と画像面が正確に直交しているわけではない。
- 画像座標 (u, v) でのスケールとカメラ座標 (x, y, z) のスケールが異なる。

これらを考慮すると,カメラ座標と画像の読取り座標(画像座標(image coordinates))の間では,つぎのような変換が必要である。

$$u = f k_u \cdot \frac{x + y \cot \phi}{z} + u_0$$

$$v = f k_u \cdot \frac{y}{z \sin \phi} + v_0$$

さらに,カメラでのより実際に近い結像モデルでは,レンズ収差(lens aberration)ひずみを考える必要がある。すなわち,レンズは特有の画像ひずみを 引き起こし,その収差ひずみによって,画像座標 (u, v) は,さらにそれぞれ補正を必要とする。一般には,よいレンズをよい条件で使う限り収差ひずみは無視しうる。 レンズのひずみモデルについては, Weng らのモデルが一般的である。多くの場合,第1項の radial ひずみのみを考慮すればよい。

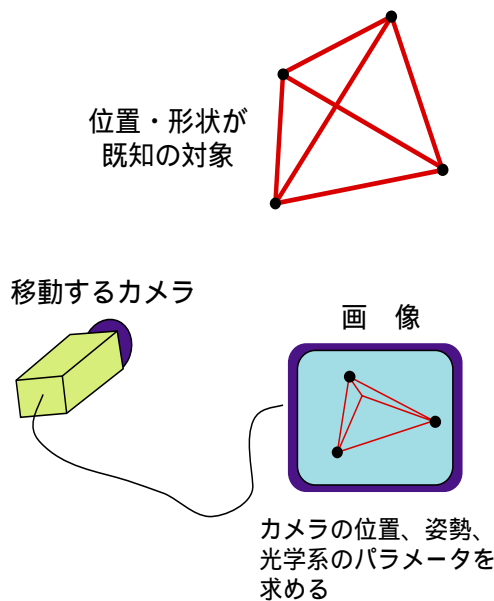


2.3 カメラキャリブレーション

カメラキャリブレーション(camera calibration)： 実際に用いるカメラについて，前記の結像モデルの各パラメータを決定すること

コンピュータビジョンの分野では，3次元空間の世界座標において位置，形状のわかっている対象を画像に取り込み，その画像面上での座標を用いてモデルをあてはめることで，世界座標系とカメラ座標系の相対関係，すなわちカメラの位置と姿勢，さらに，カメラの光学系のパラメータを求めることが一般に行われる。

- 外部パラメータ(external parameters)： 回転行列と平行移動ベクトル
- 内部パラメータ(internal parameters)： カメラの光学系の特性（焦点距離，画素の大きさ，レンズ収差や画像中心・・・）



2.3.1 内部パラメータか、外部パラメータか

3次元空間に設定した座標で位置 $X = (X, Y, Z)$ が特定できた点が，画像上のどの点 (u, v) に像を結んでいるかの対応をもとに，各パラメータを決定する。

3.3.2 透視投影行列

空間の世界座標 $X = (X, Y, Z)$ と，画像上の座標 (u, v) とは直接に，しかも，数学的に扱いやすい形では，以下のように結び付く。

$$\omega \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = P \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}, \quad P = \begin{bmatrix} f & f \cot \phi & u_0 \\ 0 & f/\sin \phi & v_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} R & \mathbf{t} \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

この表記は，斉次座標(または，同次座標, homogeneous coordinates)と呼ばれ，射影幾何学に由来するものである。射影幾何学を，コンピュータビジョンに導入することで，特に座標間を行列一つで表せる大きなメリットがある。斉次座標においては，座標を表すベクトルの0でない定数倍，は同じ点を表すものとする。

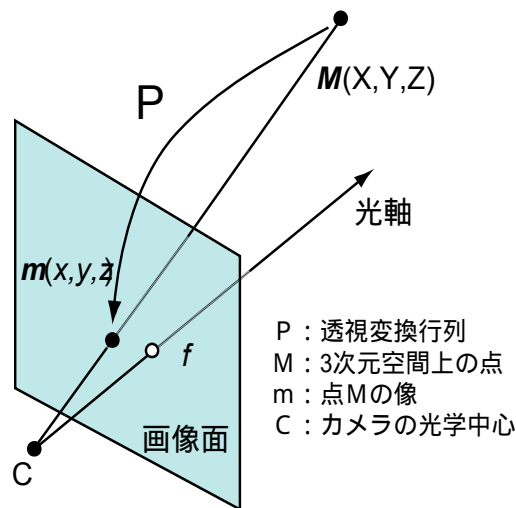
$$m \approx PM$$

この表記は，次章から解説するステレオや動画像を扱うときに便利である。以下で，この表記式はある0でない実数 ω があって， $\omega m = PM$ であることを表す。この行列を透視投影行列(projection matrix)と呼ぶ。

さて，透視投影行列のは 3×4 で，その成分を以下とすれば，

$$P = \begin{bmatrix} P_{11} & P_{12} & P_{13} & P_{14} \\ P_{21} & P_{22} & P_{23} & P_{24} \\ P_{31} & P_{32} & P_{33} & P_{34} \end{bmatrix}$$

空間と座像の座標の関係は， ω を消去することにより次式で表現できる。



$$u = \frac{P_{11}X + P_{12}X + P_{13}X + P_{14}}{P_{31}X + P_{32}X + P_{33}X + P_{34}}$$
$$v = \frac{P_{21}X + P_{22}X + P_{23}X + P_{24}}{P_{31}X + P_{32}X + P_{33}X + P_{34}}$$

すると、このカメラキャリブレーションの考え方で、いま決めなくてはならないのは、行列Pの12個の成分である。正確には、定数倍の自由度を許しているので、11個を決定すれば十分である。

画像上で6点以上が与えられることで、Pが決定できる。また、多数の点があれば最小2乗あてはめなどで精度を上げることが期待できる。