

本講義では3次元世界の視覚認識（コンピュータビジョン）を扱う。この分野の予備知識は前提としない。

人の目

もし目がなければ... 行動、作業

コンピュータビジョン： 人の目の工学的実現

視覚認識の困難性： 3次元世界での見えの多様性（位置、回転、重なり、影...）、データ量の膨大さ

人の目の働き解明されていない → 大脳視覚領

（3次元の視覚認識） 世界が3次元であることを意識

視覚認識とは何か？ 文字読み取りは明確。一般には問題が明確には定式化されていない。

工学分野としての視覚：

光計測、多点計測、非接触、センシングされた情報からの統合的判断などの全部または一部を実現しているものが工学的視覚

定型作業には視覚はほとんど必要ない（精密位置決めで視覚なしの処理が可能） 状況が変動する（一定しない）とき視覚の出番

画像処理 狭義には2次元画像の処理 広義にはCGも含む

パターン認識、シーンアナリシス、画像理解、ロボットビジョン、マシンビジョン、人工知能、コンピュータグラフィックス

イメージプロセッシングとコンピュータビジョン

広義の画像処理は、写真技術、光学技術、アナログ処理によるものも含むが、コンピュータビジョンではデジタル処理を基本にしている。

入力が画像、出力も画像の狭い意味の画像処理には、ぼけた画像の先鋭化やフィルタリングなどがある。この講義の隣接分野である。ここではあまり扱わない。詳細についてはRosenfeld, Kak（長尾監訳）、デジタル画像処理、近代科学社等参照。

ことの起こり： 1960年代初頭 L.G. Roberts (MIT) 積木の認識 (Robertsのスキームからいまだに脱しえていない)

MIT, SRI, Edinburgh Univ.等でHand Eyeプロジェクト。日本数年遅れて日立、電総研でHand Eyeプロジェクト

その後ロボットとビジョン独立して研究

産業用画像処理技術の勃興

視覚処理 2 つのアプローチ

	1	2
アプローチ名	人工知能的アプローチ	産業用画像処理
アプローチ法	一般性指向	実用性指向
ねらい	個々の応用を離れ、できるだけ一般的な手法を追及。処理が遅くとも高度に。	一般性よりも個別に役立つものを。高度でなくても高速、経済的に。
歴史	1960年代～ パターン認識、人工知能の一部として	1970年代～ 自動化技術の一環として
技術的特徴	任意の 3 次元空間、多値画像が多い	物体の 2 次元画像を 2 値的が多い

人工知能的アプローチ

広義の問題解決、問題の提供場所、手法の適用場所、物理レベルの処理～推論、問題解決

トップダウンとボトムアップ、セマンティックギャップ

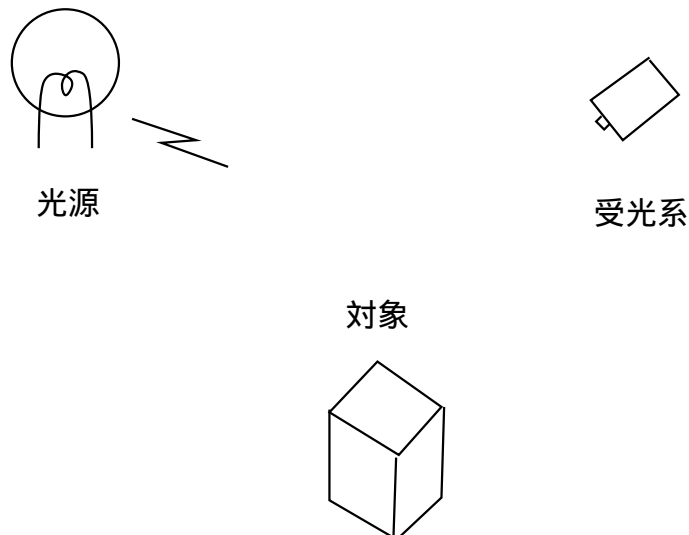
コンピュータビジョン（画像処理）のスキーム（定石）

画像入力（前処理）、処理対象の抽出（領域、線等）、処理対象の特徴記述、（照合等による）判断。別に処理対象のモデル化（学習、教示）

画像処理のスキーム

画像入力、（前処理）、処理対象の抽出（線、領域など）、処理対象の特徴抽出、照合→判断結果 別に処理対象のモデル化（学習、教示）

画像とは



光源

{ 自然光、人工照明 }、種類 { 点、線、面 }、色 { 単色、マルチスペクトル }、数 { 単一、複数 }、位置、方向

→

対象

形、位置、姿勢、反射率 { 乱反射、完全反射 }、色、模様、光沢

→

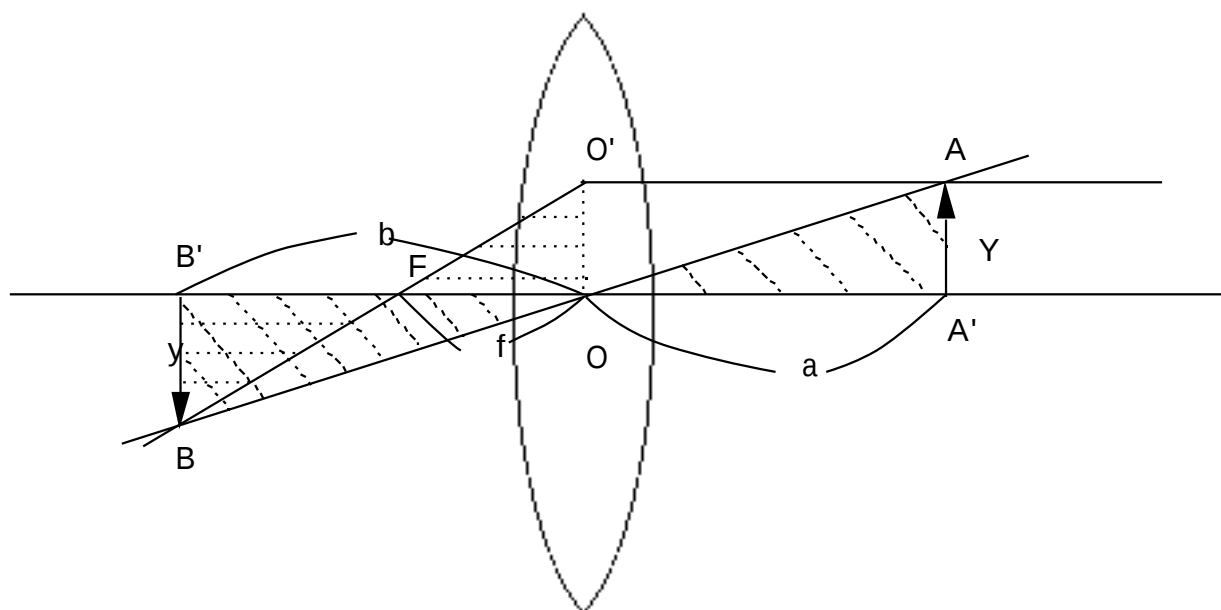
受光系

位置、姿勢、感度の波長特性、 γ 特性、画素数、濃淡（色）のレベル数、色のRGB or マルチスペクトル、種類 { 点、線、面 }

{ 2 値画像、濃淡画像、カラー画像、距離画像 }

（ { リモートセンシングデータ、医用画像 { X線、CT }、顕微鏡、赤外線、マイクロ波、超音波、偏光 } ）

レンズについて



対象とする点Aを通る光の内平行光はレンズ通過後Fを通り、O（レンズの主点）を通る光は向きが変わらない。この2つの光線の交わりとして結像点Bが定まる。他の光もレンズ通過後結像点を生じるように進む。

DAAO DBBOおよびDOOF DBBFより

$$\frac{y}{Y} = \frac{b}{a} \quad \frac{y}{Y} = \frac{b-f}{f}$$

これより

$$\frac{b}{a} = \frac{b-f}{f} - 1 \quad \square b$$

$$\frac{1}{a} = \frac{1}{f} - \frac{1}{b} \quad \text{よって}$$

$$\frac{1}{a} + \frac{1}{b} = \frac{1}{f} \quad (\text{よく知られているレンズの公式}) \quad \text{この式より}$$

$$\begin{aligned} \frac{1}{b} &= \frac{1}{f} - \frac{1}{a} \\ &= \frac{a-f}{fa} \end{aligned}$$

$a \gg f$ のとき

$$\text{右辺} \approx \frac{a}{fa} = \frac{1}{f} \quad \text{よって}$$

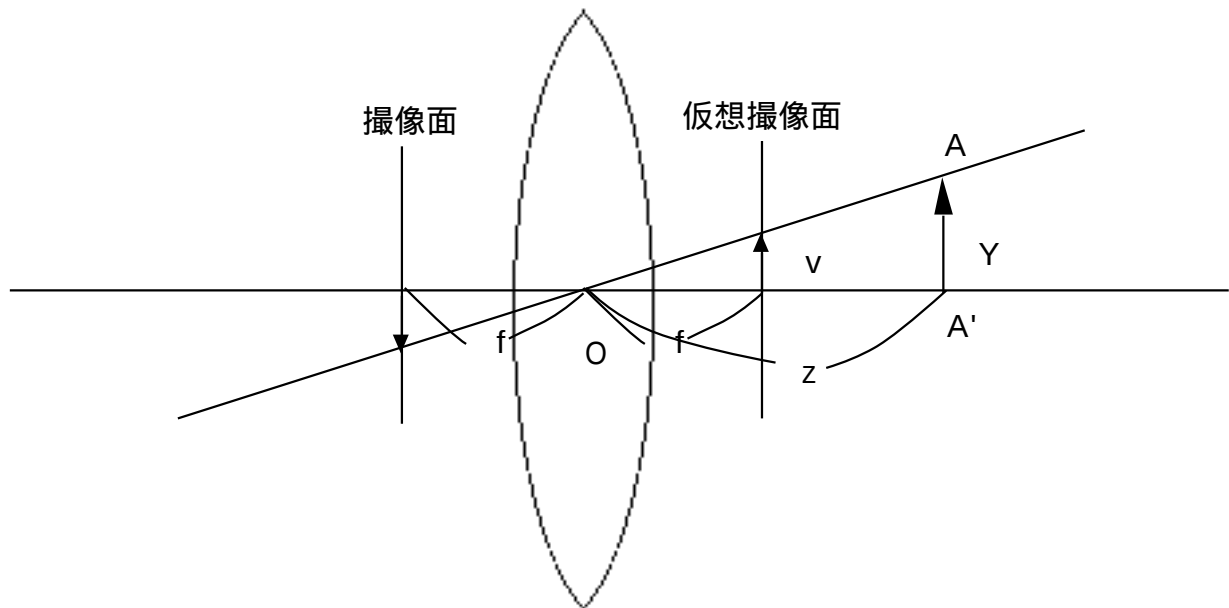
$b \approx f$ 十分遠くにある対象はfの近くに像を結ぶ。また、無限遠点（光軸に平

行な光)はfに像を結ぶ。

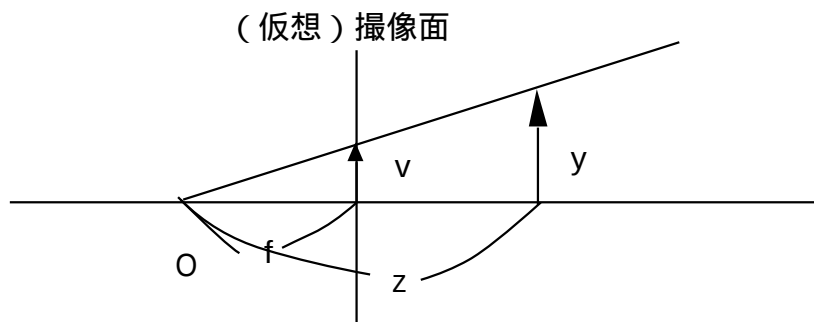
これをピンホールモデルという。

焦点(ピント)合わせとは、与えられたf, aに対し、式を満たすbを定めること

カメラモデル

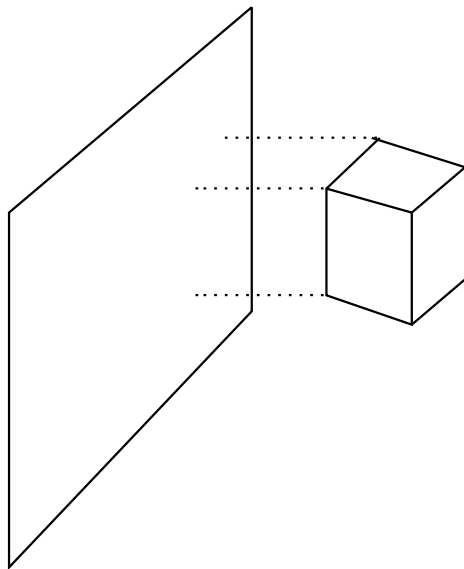
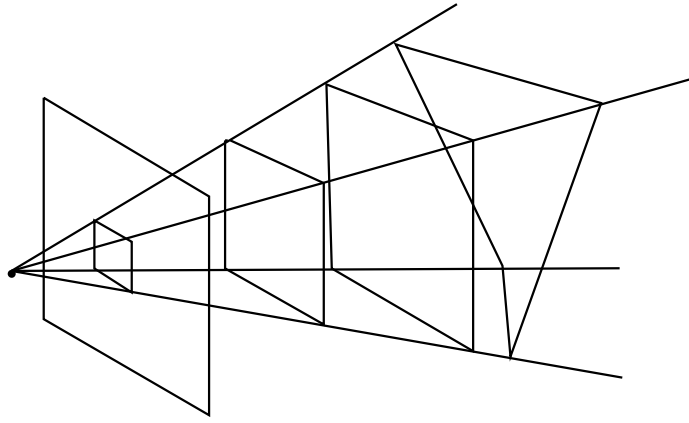


実際の結像はレンズの後方となるが、仮想の撮像面をレンズ前方に置くと、結像する像の大きさは変わらない。また倒立像でなく正立像となる。この方がモデルが単純化される。(慣例的にテレビの横座標、縦座標をu,vで表すことが多い)



レンズの中心(視点)の前方に撮像面があると考え、ここに前方の対象yがvの大きさで結像するものとする。3角形の相似から $\frac{z}{f} = \frac{y}{v}$

カメラはこのような近似でき、3次元の世界を透視投影によって2次元に変換する。他に正射影として扱う場合も(まれに)ある。



一般に、カメラで撮像するとき、3次元 → 2次元の変換が行われるので立体情報は失われる。立体情報は重要なので、ぜひそれを得たい。画像から立体情報を得ることを立体情報の再構成（Reconstruction）または復元という。ビジョンの重要なテーマである。

人間は10種類くらいの方法（両眼立体視、shape from shadingなど）を使って立体情報の再構成を行っていると考えられている。