

	人間	機械
標語	人間的	機械的
融通	きく	きかない
感覚	すぐれている（見る、聞く、触る）	これからの課題
くり返し作業	喜ばない	得意とする

機械に人間的なことができれば、もっと便利な世の中になると考えられる：
単純労働、危険作業、福祉...

見る、聞く、話す、五感のある機械

認識、知識、知能、知覚

人の顔を見分ける

文字を読み取る

街を歩く

作業をする

スポーツをする

TVを楽しむ

石頭コンピュータ 1と0の世界 何とか人間の感覚に近付けたい
ロボット 繰り返しが基本

目の工学的実現を目指して→コンピュータビジョン

目で見て判断するとは：

文字の読み取り

図形の区別

色の判断

動きの検出

立体構造物の把握

カエル目 動きに感じる

鯉 色の区別できる

ネコ 色の区別できない

サル 人と同じように判断できているらしい

工学的に見ると：

非接触遠隔計測

多点計測

得られる情報量が多い

コンピュータビジョンの歴史

1960年代前半 米国MIT L.G. Roberts

積み木の認識（計算機内に物体の3次元モデル、観測画像からエッジ抽出、線画の構成、モデルの透視変換と仮定してベストマッチの位置・姿勢を判断）

faxで物体（積み木）の写真を計算機に入力

エッジ抽出（定石定まる）

物体モデル

透視変換によるマッチング

ハンド・アイ実験 1960年代後半 米国MIT、スタンフォード大学、英国エジンバラ大

少し遅れて（1970年）日本 日立中央研究所、電子技術総合研究所

一応の成功を収めるが問題の難しさも分かってきた

テーブル（グラウンド）の仮定

その後、地道な研究の積み重ね ハンドとアイは別の道をたどる。ハンドアイ実験は

その後10年くらいあまりなされなくなった。

対象の広がり 積み木→室内→屋外
静止画処理から 動画像処理へ
入出力機器、処理装置の進歩

産業応用研究開発と発展・実用

ビジョンの関連分野

(狭義の)画像処理 画像の強調、ボケの回復、ノイズ除去
文字読取

図面処理

(シーンアナリシス)

産業用画像処理(マシンビジョン)

ロボットビジョン

医療用画像処理(強調(例 疑似カラー)、CT、イメージング、X線像(胸部、胃))
リモートセンシング(衛星、航空機)

動画像解析

色彩画像処理

テクスチャ解析

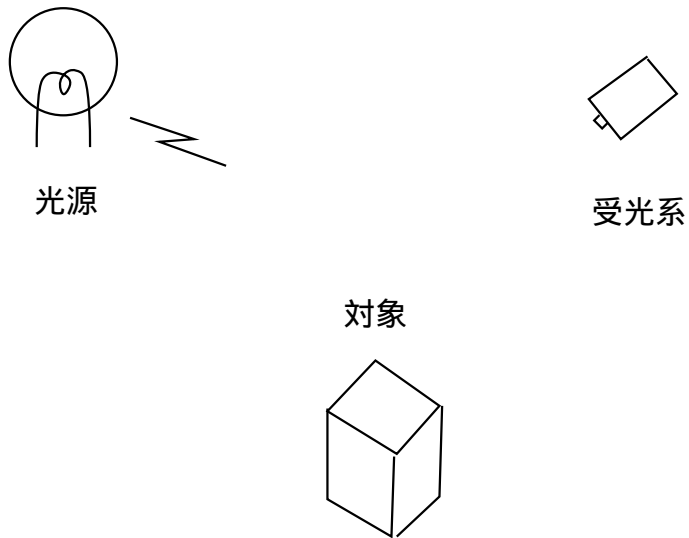
レンジデータ処理

パターン認識、人工知能 CG、CAD、CAM

画像処理の定石

画像入力、(前処理)、重要な部分(特徴)の抽出(線、領域など)、特徴同士の相互関係(記述)を求める、照合(あらかじめ分っているものと未知のものが一致するか調べる)→判断結果 別に処理対象の記憶(学習、教示、モデル化)、戦略(ボトムアップ、トップダウン)

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 })

扱うデータ

2 値画像

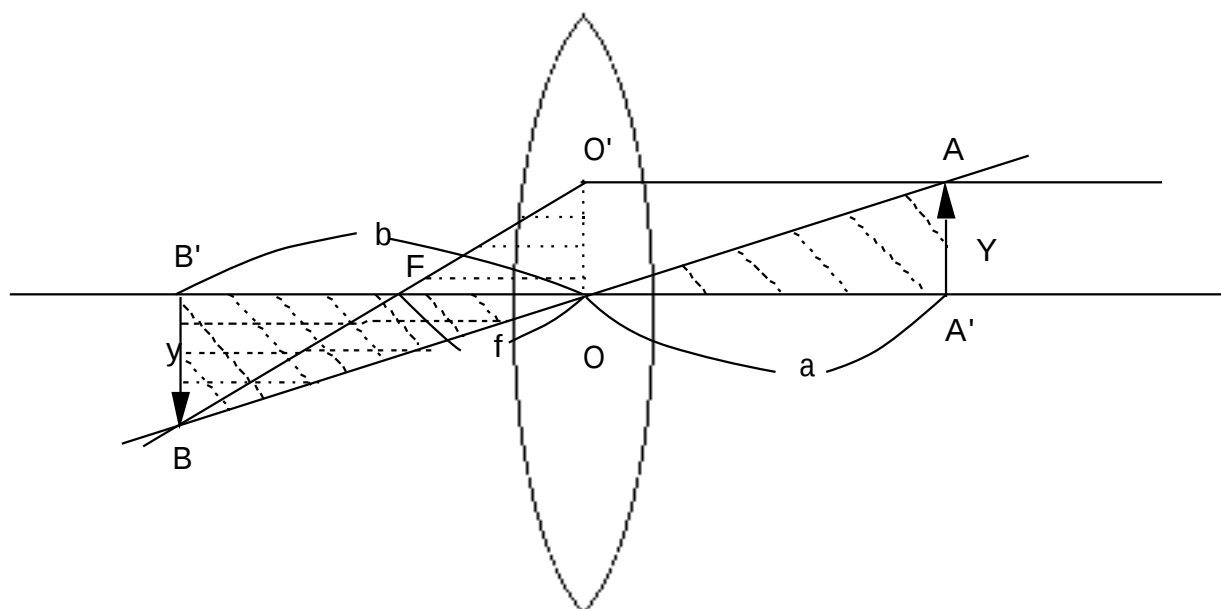
濃淡 (階調) 画像

カラー画像

距離画像

他の方法 (CT、リモートセンシング)

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。

$\triangle DAO \sim \triangle DBB'$ および $\triangle DOF \sim \triangle DBB'$ より

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

これより

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

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

$$\frac{1}{a} + \frac{1}{b} = \frac{1}{f} \quad \text{この式より}$$

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

$a \gg f$ のとき

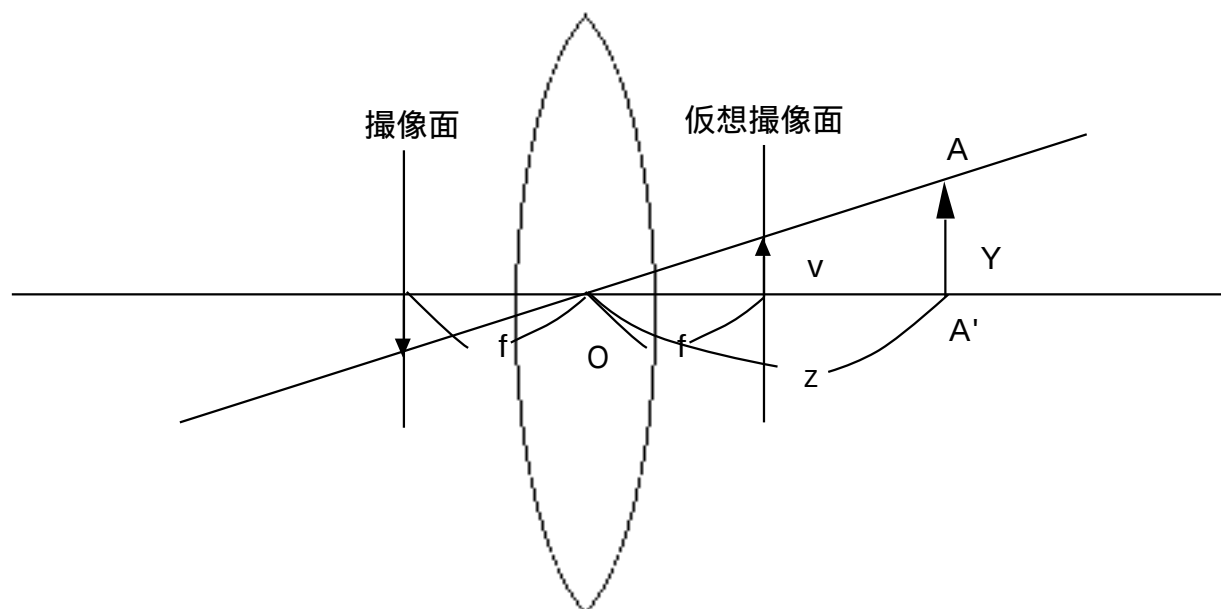
$$\text{右辺} \approx \frac{a}{fa} = \frac{1}{f} \quad \text{よって}$$

$$b=f$$

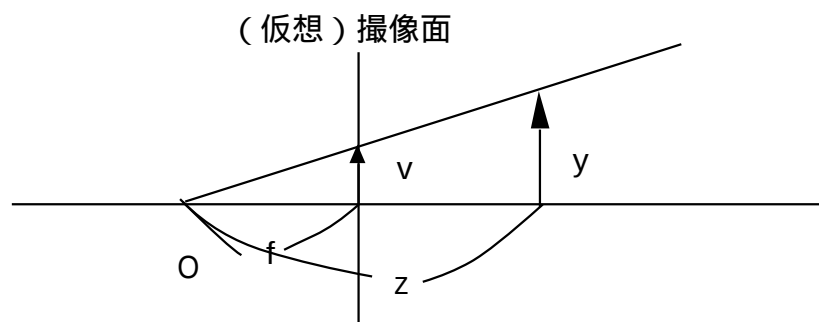
これをピンホールモデルという。

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

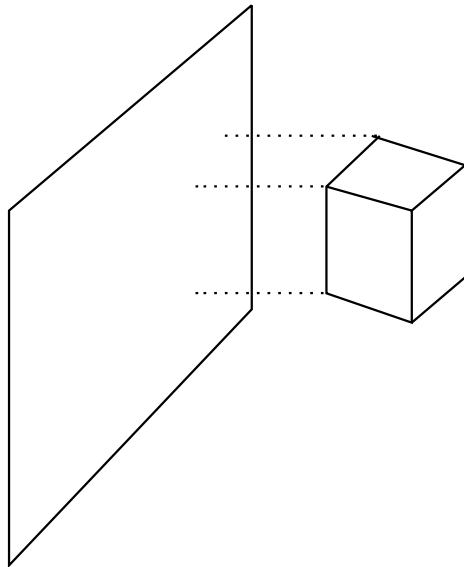
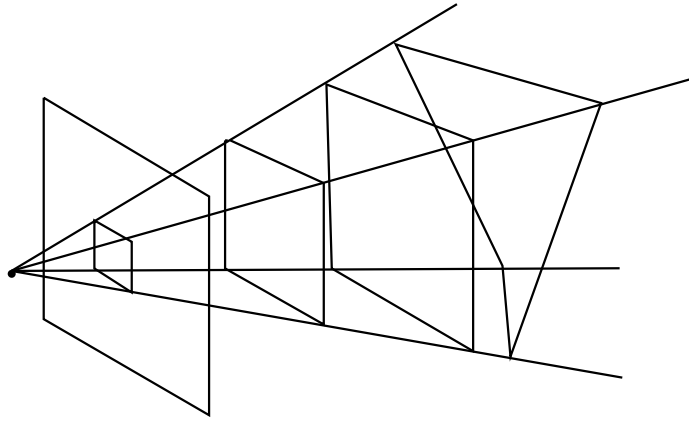


実際の結像はレンズの後方となるが、仮想の撮像面をレンズ前方に置くと、結像する像の大きさは変わらない。また倒立像でなく正立像となる。この方がモデルが単純化される。（慣例的にテレビの横座標、縦座標を u, v で表すことが多い）



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。