

市民講座 情報学最前線

驚き！3Dセンシング —進化するコンピュータの眼—

池畑 諭

国立情報学研究所

©Satoshi Ikehata 2019, all rights reserved

3Dセンシングとは？

3Dセンシングについて

「**センサー**」と「**情報処理技術**」を用いて、
ヒトのように世界を3次元で知覚する技術

(例) 環境を知覚して動くロボット掃除機



センサーの種類



受動型 (パッシブ)

デジタルカメラ



能動型 (アクティブ)

3Dスキャナ



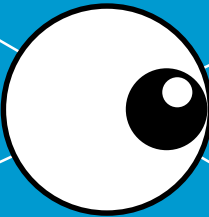
受動型 (パッシブ)
デジタルカメラ

受動型センシング

≡

ヒトの3次元情報処理

2次元像

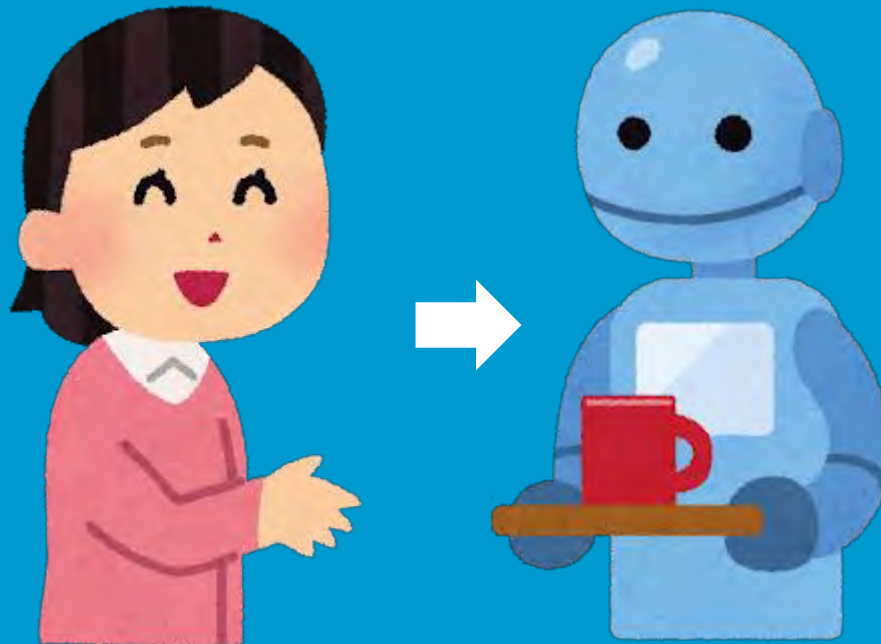


3次元環境

写真から奥行きを知覚できる？



アルゴリズムの実現とは 奥行き知覚プロセスを言語化



無意識に奥行きはわかるけど、
奥行きのある絵を描くのは難しい！



立体的に描くには？



真横から



視点を変える

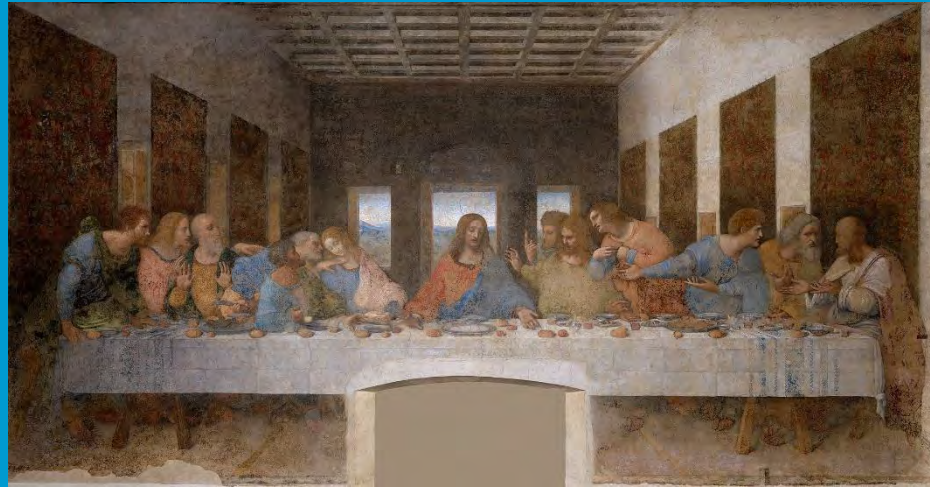


遠近法

絵画における遠近法表現は ルネサンス期に遡る



『聖母と天使たち』
チマブーエ (1240-1302)



『最後の晩餐』
レオナルド・ダ・ヴィンチ(1452-1519)

A photograph of a glass of water and a bowl of red shaved ice on a black tray. The background shows a window with a view of a green hillside. The text "どんな奥行き手がかかりがある?" is overlaid on the image.

どんな奥行き手がかかりがある？

遠近法

消失点



視野内の高さ

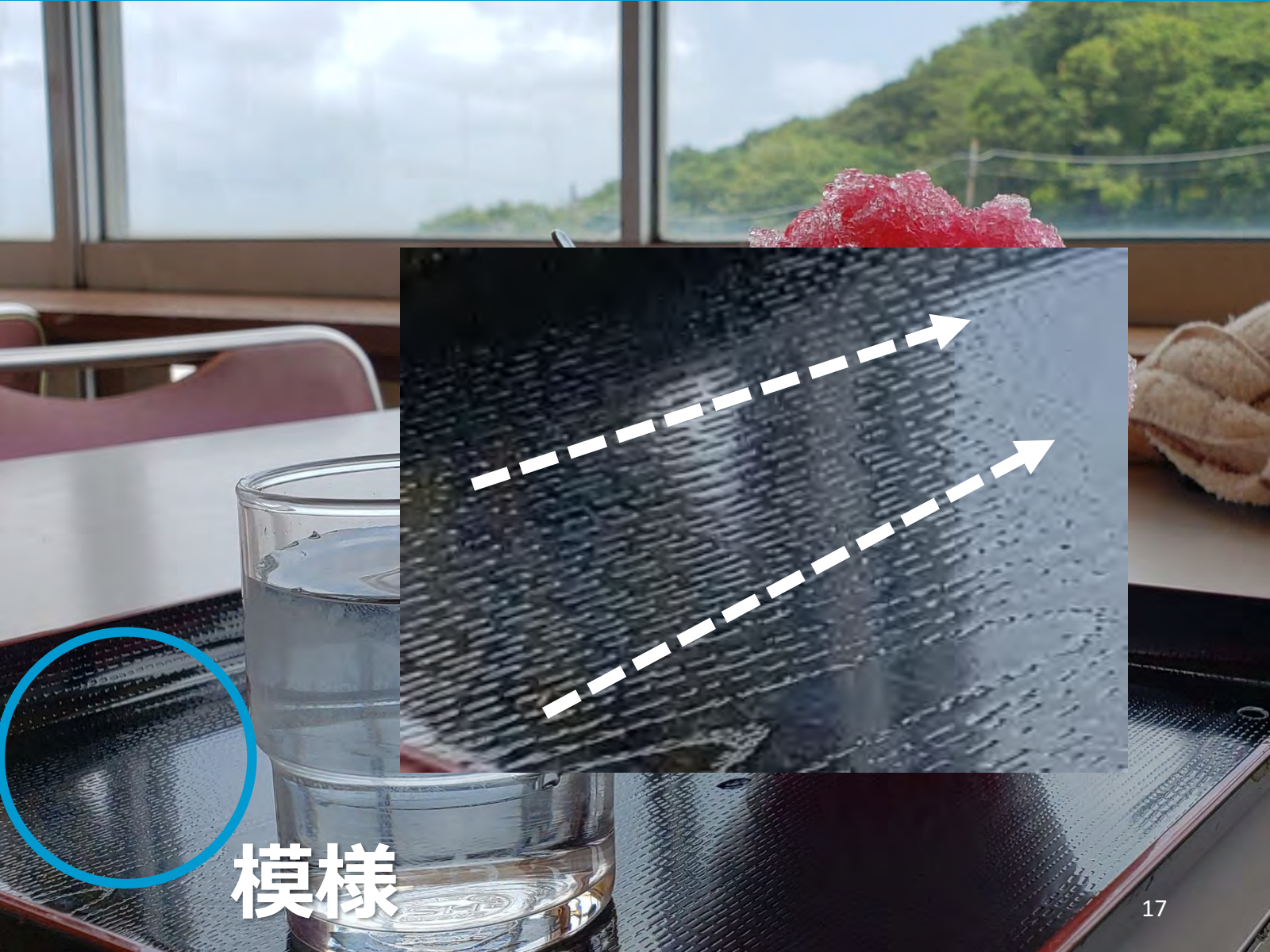


ぼけ





遮蔽



模様



光のふるまい（反射・透過・影）

単眼（絵画的）手がかり

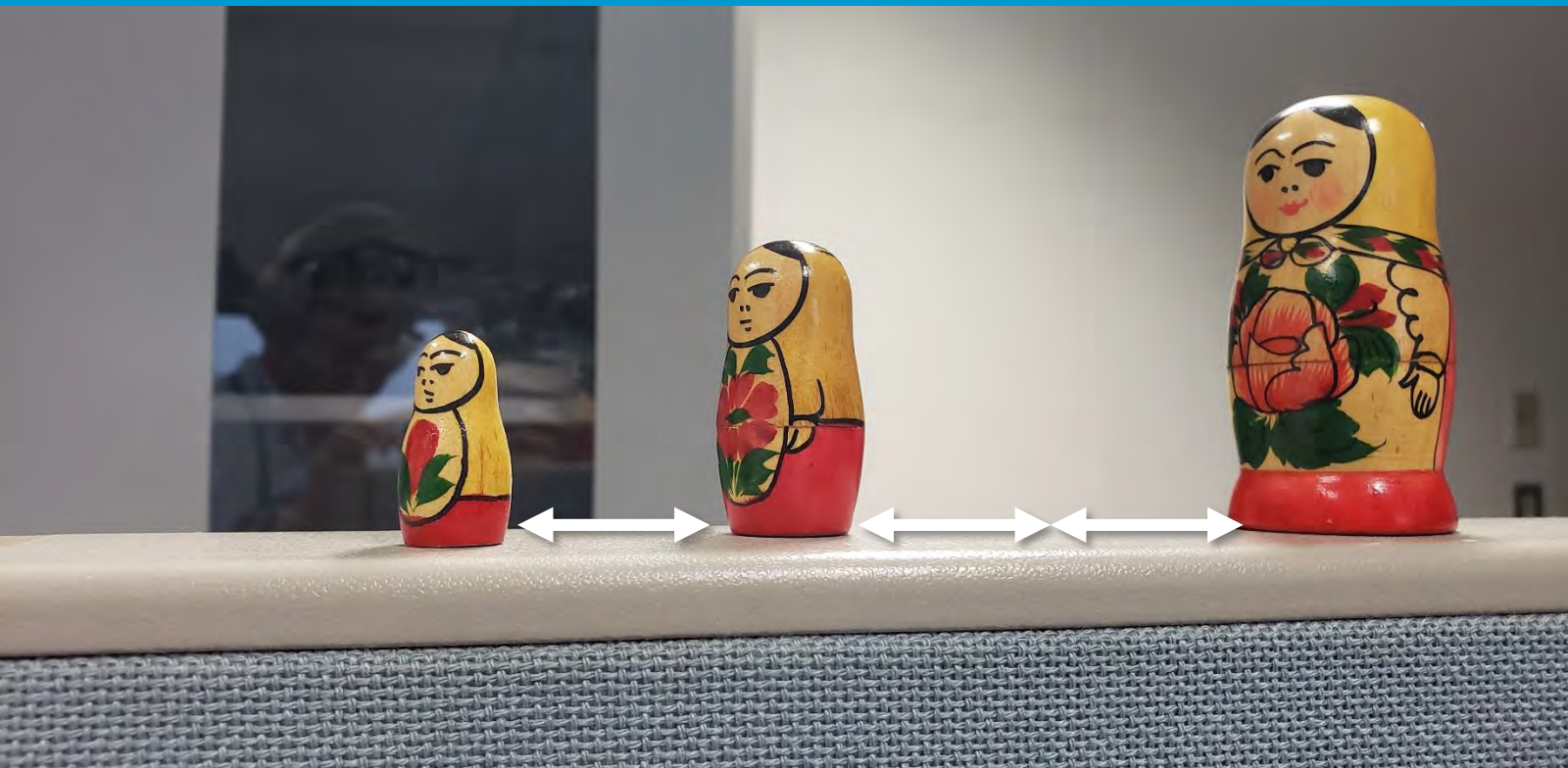
- ・ 線遠近法
- ・ 視野内の高さ
- ・ ぼけ
- ・ 遮蔽
- ・ 模様
- ・ 光のふるまい

絵画的手がかりはあいまい 人形同士はどれだけ離れてるか？



絵画的手がかりはあいまい

人形同士はどれだけ離れてるか？



あいまい性をなくす「視差」

近いほど視差（ずれ）が大きい



運動視差



両眼視差

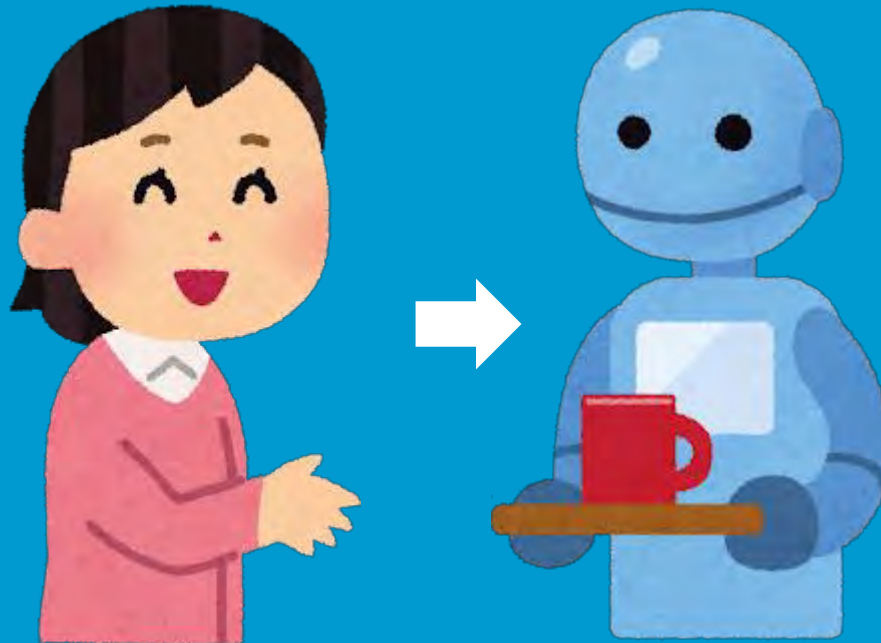
単眼手がかり

- ・ 線遠近法
- ・ 視野内の高さ
- ・ ぼけ
- ・ 遮蔽
- ・ 模様
- ・ 光のふるまい
- ・ 運動視差

両眼手がかり

- ・ 両眼視差

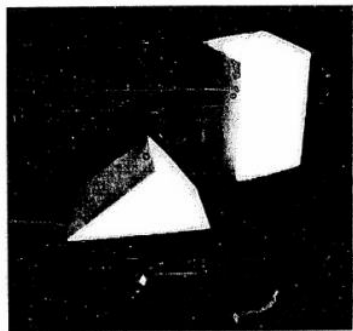
アルゴリズムの実現とは 奥行き知覚プロセスを言語化 3Dコンピュータビジョン



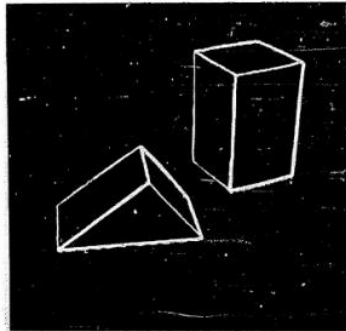
コンピュータビジョンの父 ラリー・ロバーツ

線画からの形状復元
「ブロック・ワールド」
(1963)

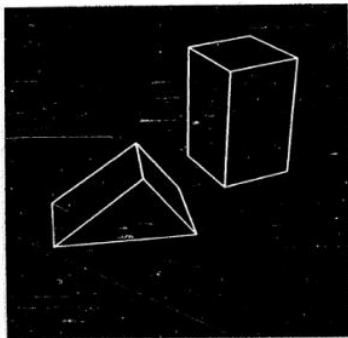
世界初のパケット
通信ネットワーク
ARPANETの父



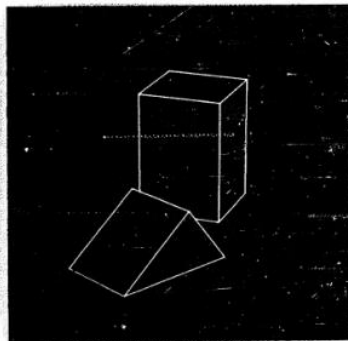
A. Original Picture



B. Differentiated Picture



C. Line Drawing



D. Rotated View

Roberts, "Machine perception of three-dimensional solids". PhD Thesis, 1963

コンピュータビジョン は簡単な問題？

“it was believed that this could be achieved through a summer project, by attaching a camera to a computer and having it "describe what it saw" (Seymour, The Summer Vision Project 1966)

「人間の知覚を模倣するなど、サマープロジェクトで終わる程度の事だと思われていた。単に、カメラをコンピュータに取り付けて、写されたものを記述するだけだろうと」 (シーモア, The Summer Vision Project 1966)

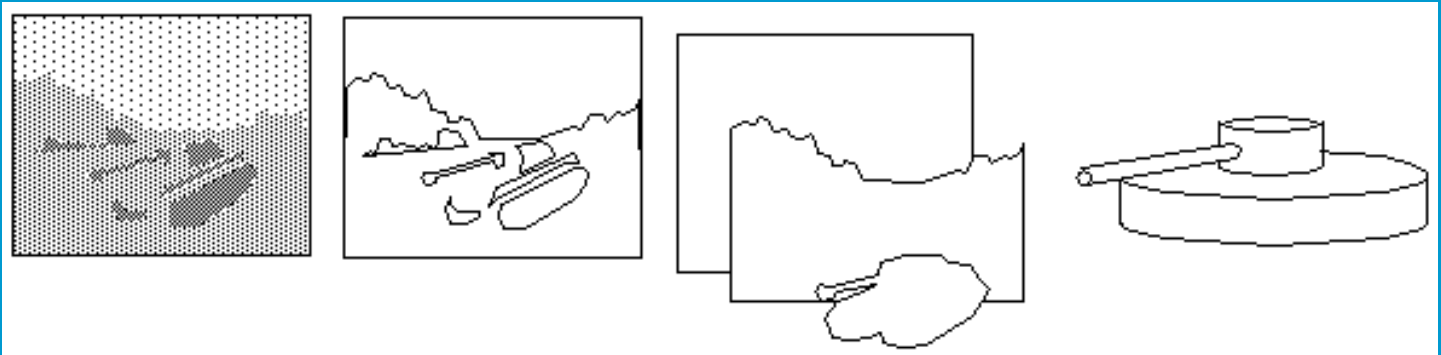
デビット・マー「ビジョン」(1982)

視覚の計算論的アプローチ

原始スケッチ
(局所的な性質)

2・1/2スケッチ
(観測者からの奥行き)

3次元構造
(普遍的な形)

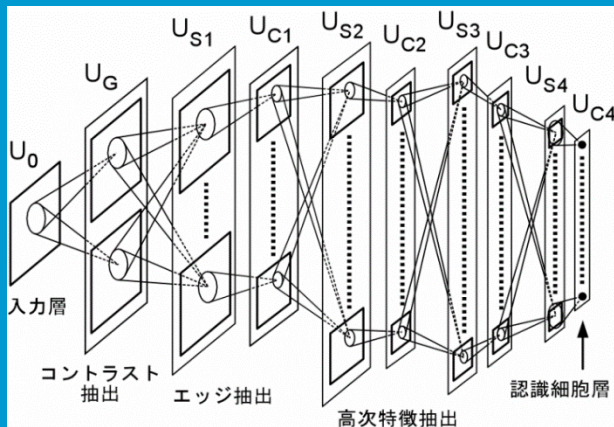


脳の学習機能についての言及は無い
人間は「見えない裏側」の形状もわかる

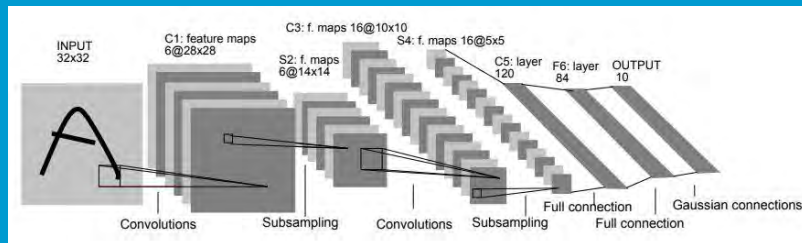
Marr, "Vision: A computational investigation into the human representation and processing of visual information", 1982.

余談

ちょうどこのころ、データに基づく深層学習の基礎理論が提案されたが、それほど重要視されなかった



福島邦彦
「ネオコグニトロン」
(1979)



ヤン・ルカン
「畳み込みニューラルネットを
利用した文字認識」
(1989)

研究事例①

両眼画像で3Dセンシング

研究事例②

陰影画像で3Dセンシング

研究事例①

両眼画像で3Dセンシング

【問題】 二台のカメラで 奥行きを計算

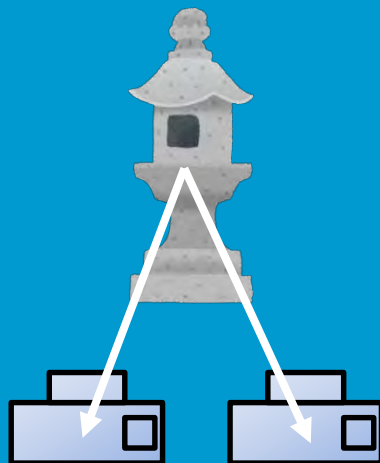
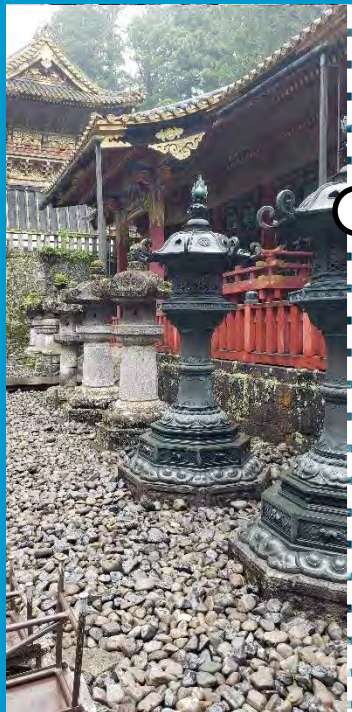
左
画像



右
画像

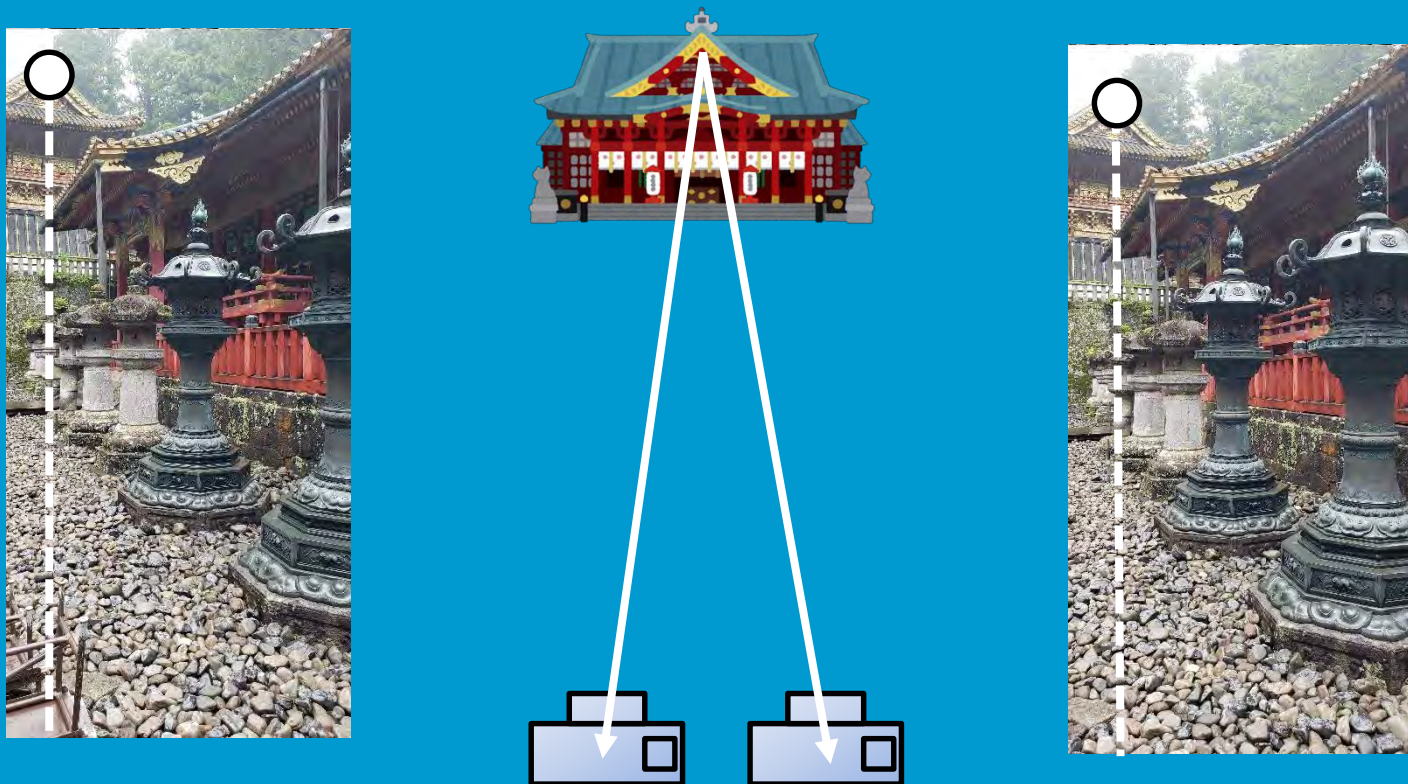


カメラ（視点）がずれると視差が発生



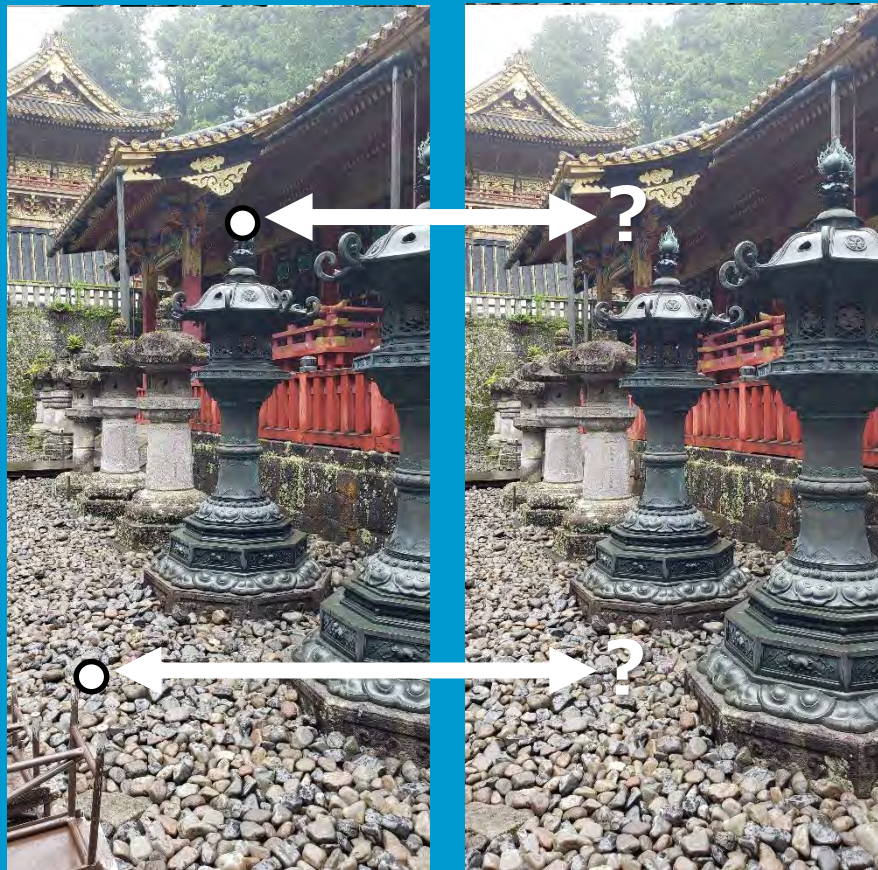
近いほどずれは大きい

カメラ（視点）がずれると視差が発生



遠いほどずれは小さい

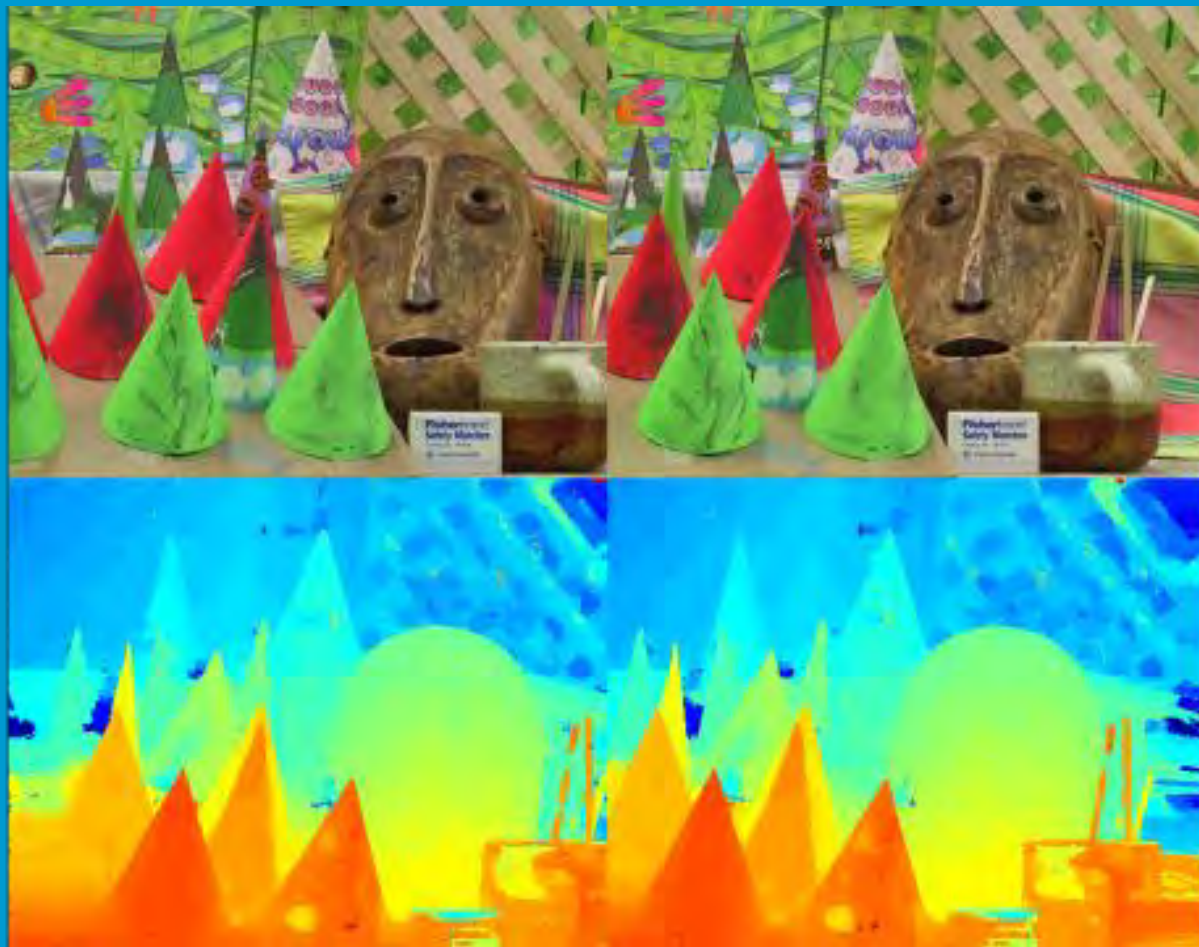
ずれの大きさに距離がわかる 対応点探索問題（ステレオマッチング）



全ての点の対応から視差画像へ

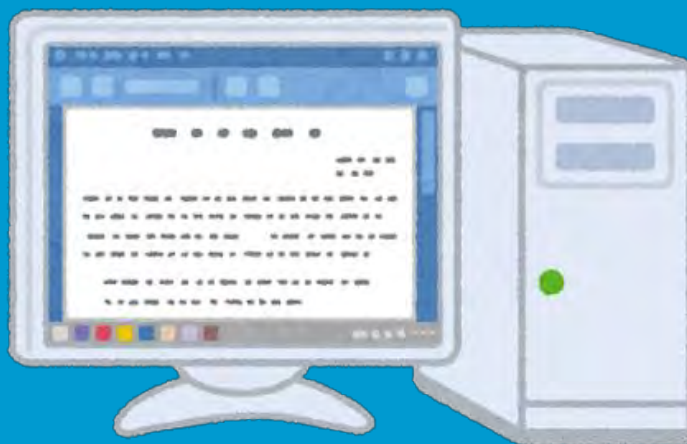
視差 奥行き
小 大

大 小



対応点探索は人間は一瞬で可能

コンピュータにとっては、 約40年間未解決の問題



コンピュータにとって 画像は単なる「数字の羅列」



何をもって「対応している」と定義づけるのか



画素値（色）は
必ずしも一致しない

パターン特徴の発展

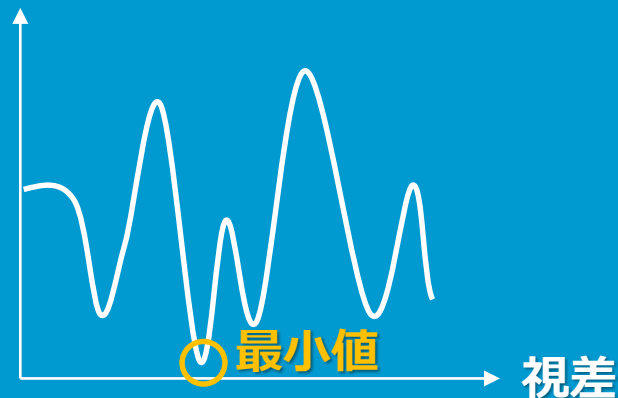
(1980年から2000年頃)

- ・ 絶対値差分(SAD)
- ・ 二乗差分(SSD)
- ・ 正規化相互相関(NCC)

色合いに左右されない特徴



パターンの差



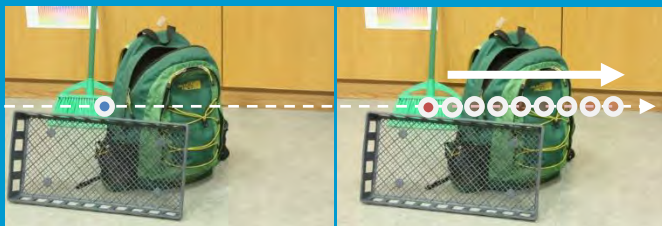
パターン特徴の限界



似たパターンの
繰り返し

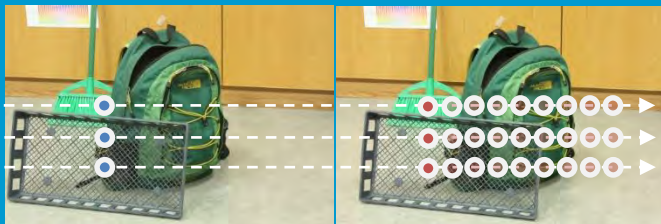


ステレオアルゴリズムの変遷



それぞれの画素だけ見れば良い

パターン特徴
(1980-2000頃)

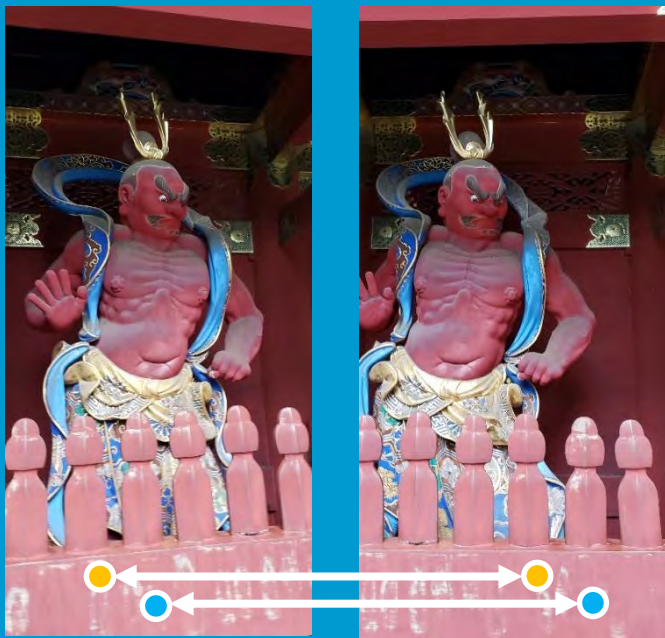


全体を見る必要がある！

大域最適化
(2000-2010頃)

大域最適化

- 全ての画素の視差を**同時**に推定
- **視差の平滑性**を制約に与える



「画素が近ければ
視差も近い」

大域最適化

- 全ての画素の視差を**同時**に推定
- **視差の平滑性**を制約に与える



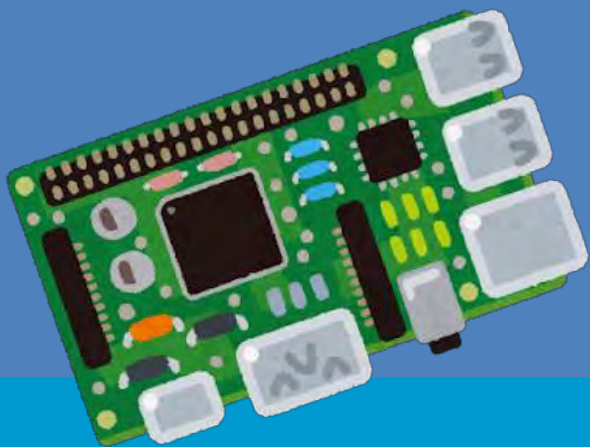
パターン特徴
(1980-2000頃)



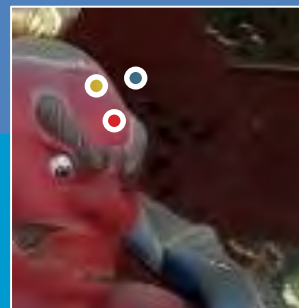
大域最適化
(2000-2010頃)

2000年代は大域最適化全盛期

全てのパターンを網羅的に探索するのは大変
それを解決するための様々な最適化手法



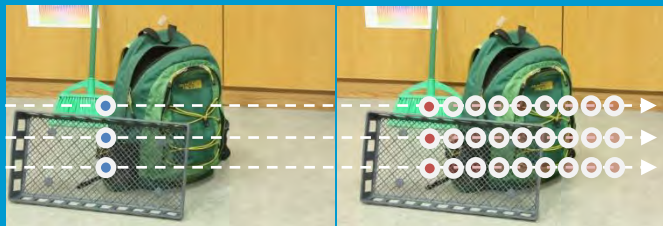
グラフカットアルゴリズム
信頼度伝播アルゴリズム
動的計画法



加えて、様々な人間の知識を問題に与える

例：「距離だけじゃなくて色が近ければ視差が近い」

ステレオアルゴリズムの変遷



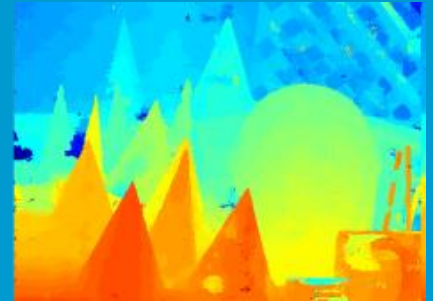
大域最適化
(2000-2010頃)



深層学習
(2010-現在)

深層学習って何？

入力から出力を直接与える
人間の脳のような「ニューラルネット」を学習



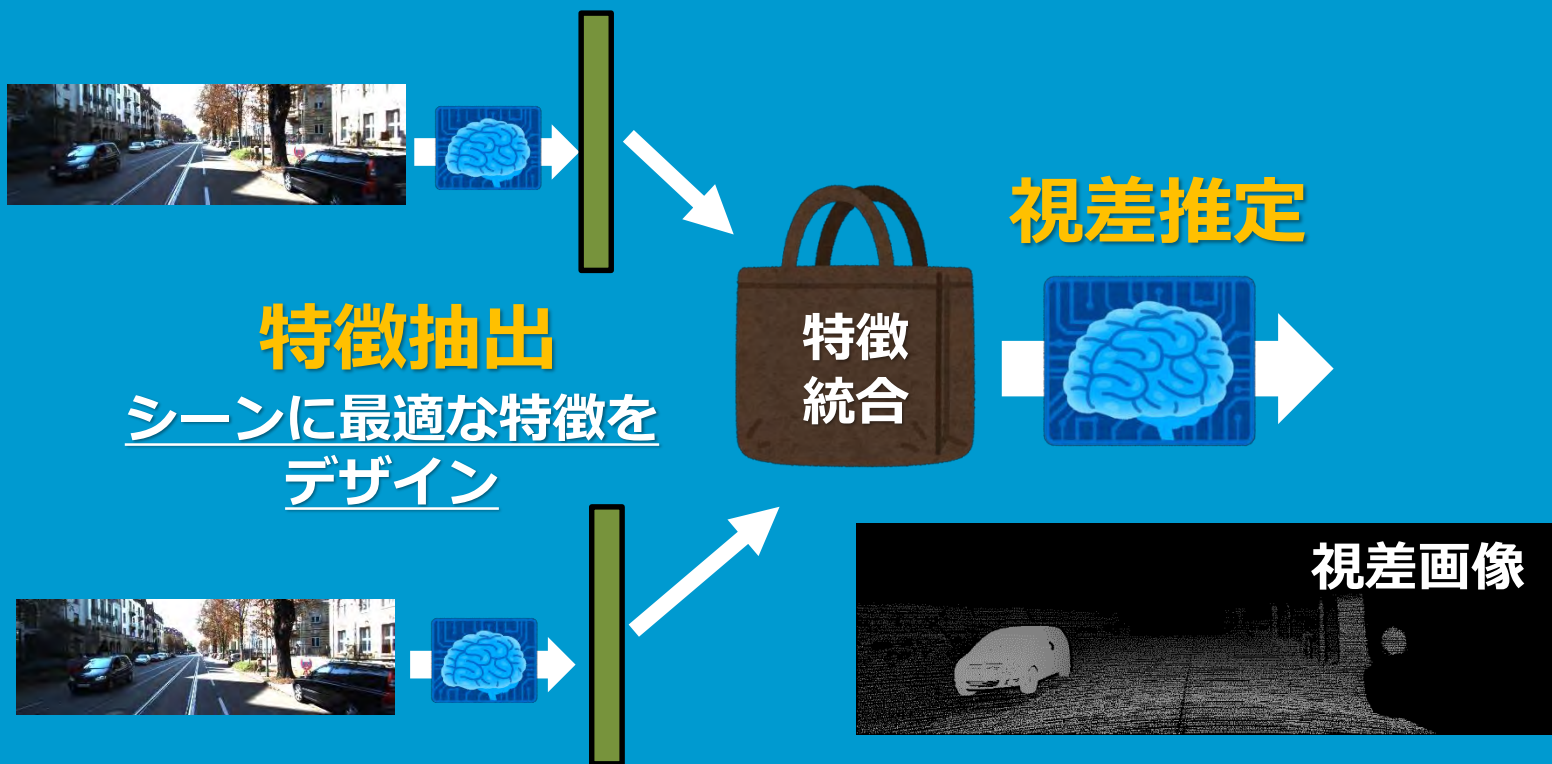
特に、学習データとテストデータ
が近い時に効力を発揮する

応用が明確になってきた



- ・ 車載カメラの映像
- ・ 大規模なデータ
- ・ リアルタイム性

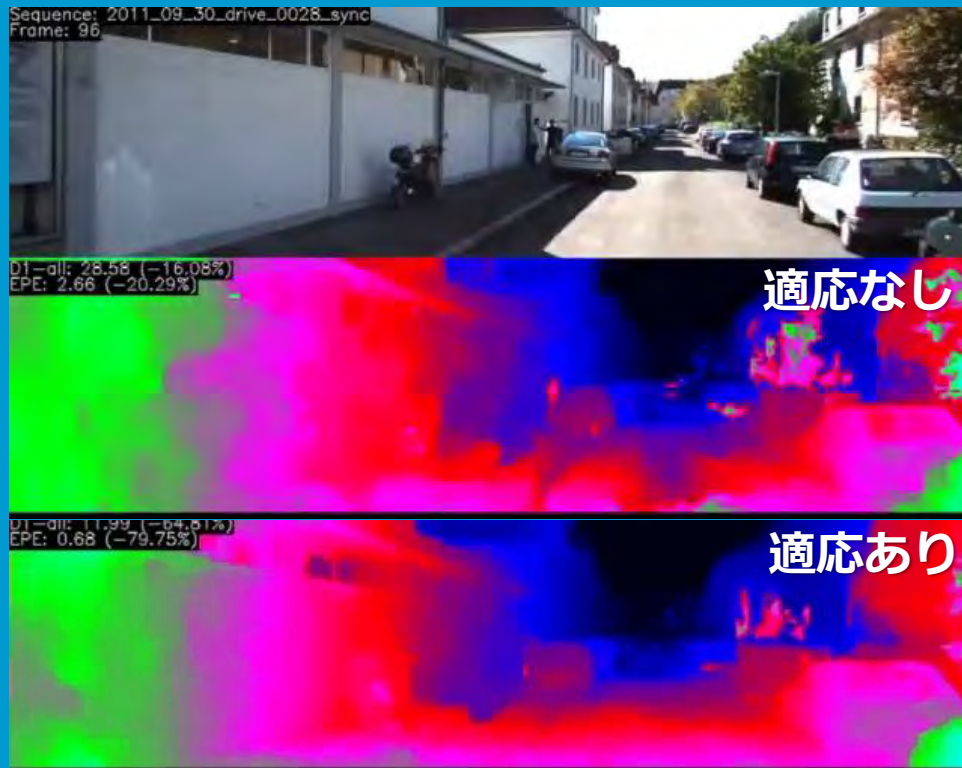
深層学習ステレオの一般的な形



この形式の一例
Chang et al, "Pyramid Stereo Matching Network", (CVPR2018)

2019年時点での最新手法

学習データとテストデータが異なる場合でも適応



Tonioni et al., "Realtime self-adaptive stereo" (CVPR2019)

ステレオマッチングまとめ

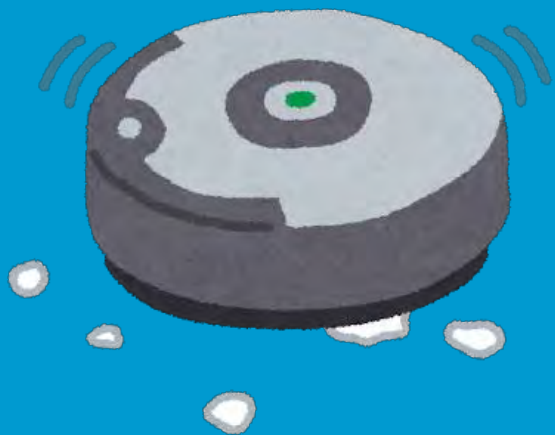
- 左の画像と右の画像の対応点探索問題
- 局所的なパターンマッチングから大域的な最適化、そしてデータ駆動型の深層学習へ
- 自動運転自動車やロボットへの応用が加速している

視差についての関連する問題

カメラを動かしてカメラの位置姿勢と
シーンの構造を同時に復元（運動視差）

運動からの構造復元法（SfM）, SLAM

たくさんのカメラで大規模シーンを復元
多視点ステレオ

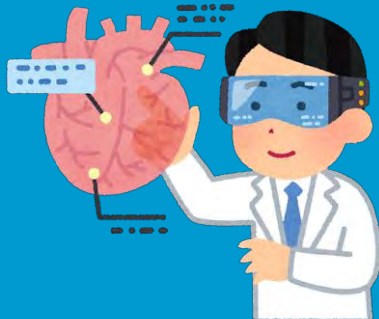


閑話休題

3Dセンシング研究者の悩み

社会実装が難しい！

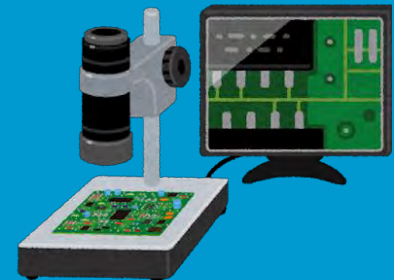
- ・ 学術的に面白いし、分野は活発
- ・ でも、成果をどう応用する？



医療

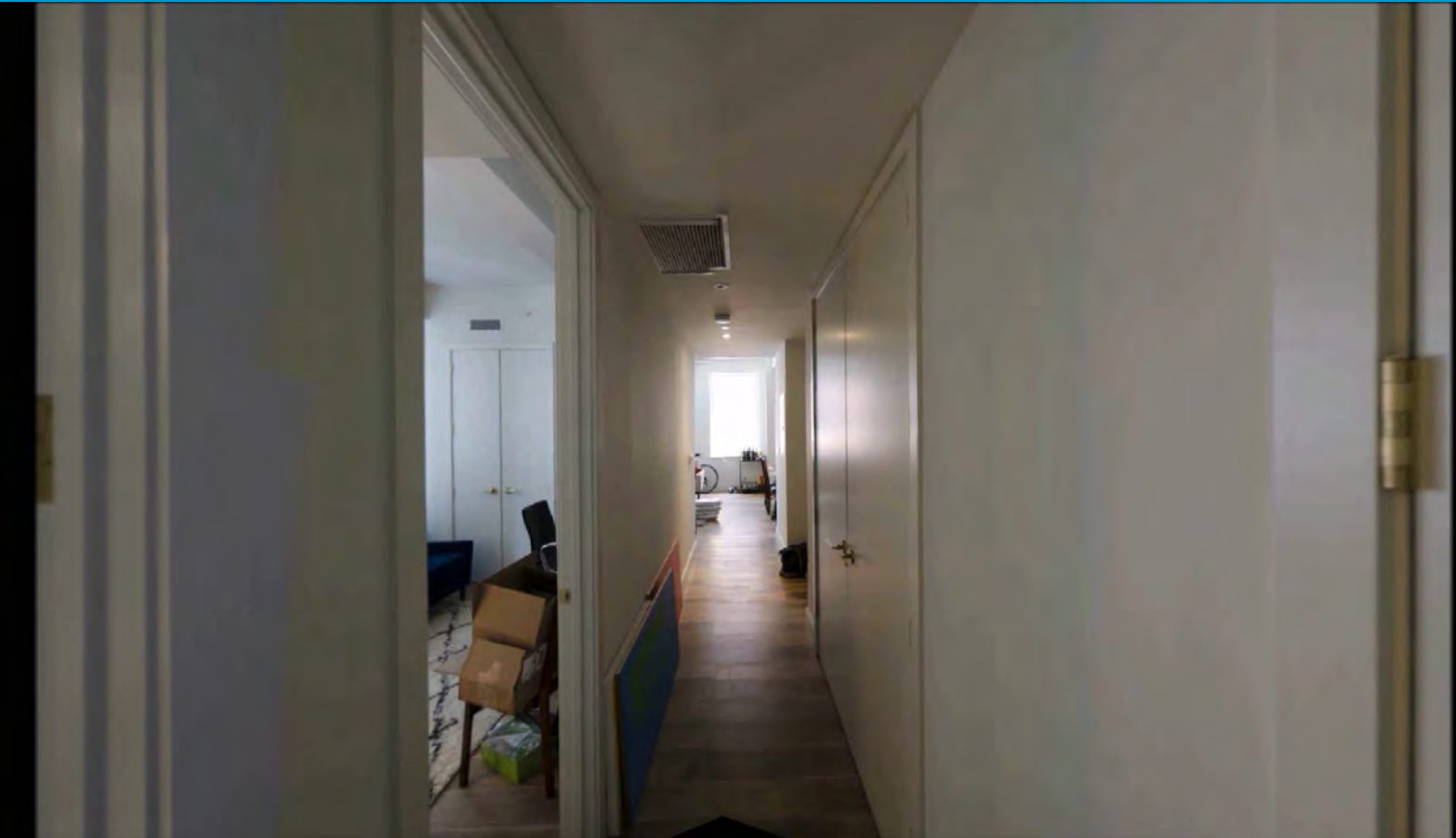


VR・AR



製造

不動産分野と3Dセンシング



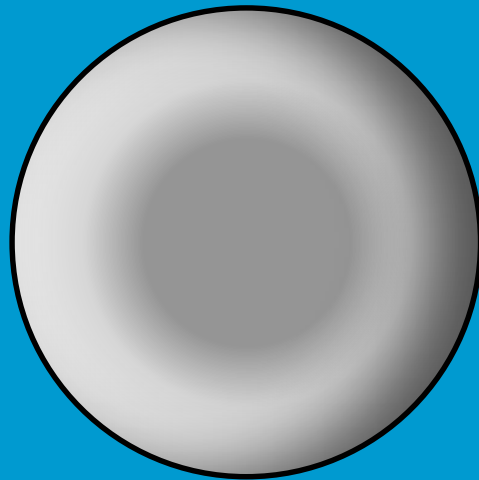
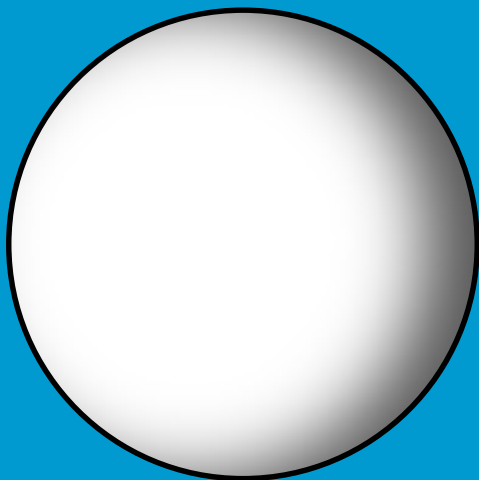
応用を見据えたセンシングを



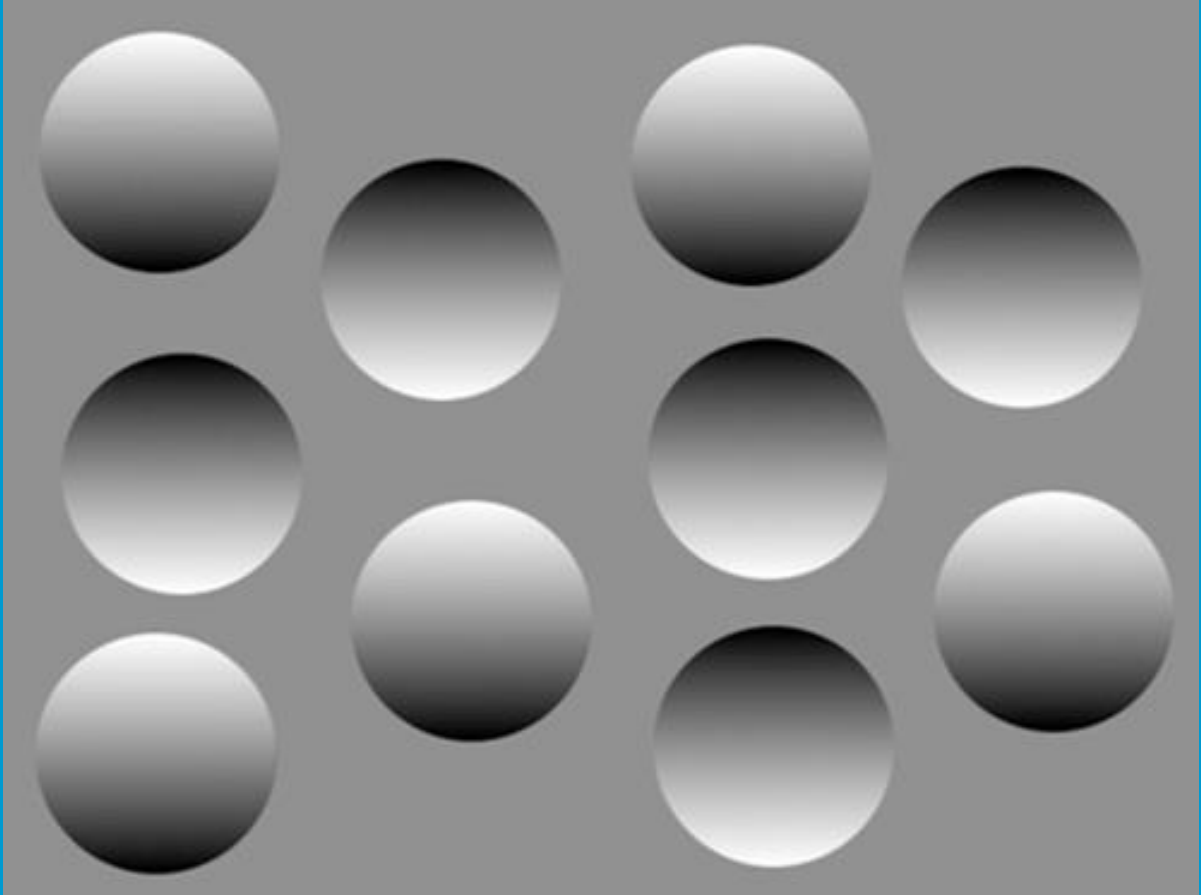
研究事例②

陰影画像で3Dセンシング

【問題】 陰影画像から形状復元



陰影の手がかりはあいまい



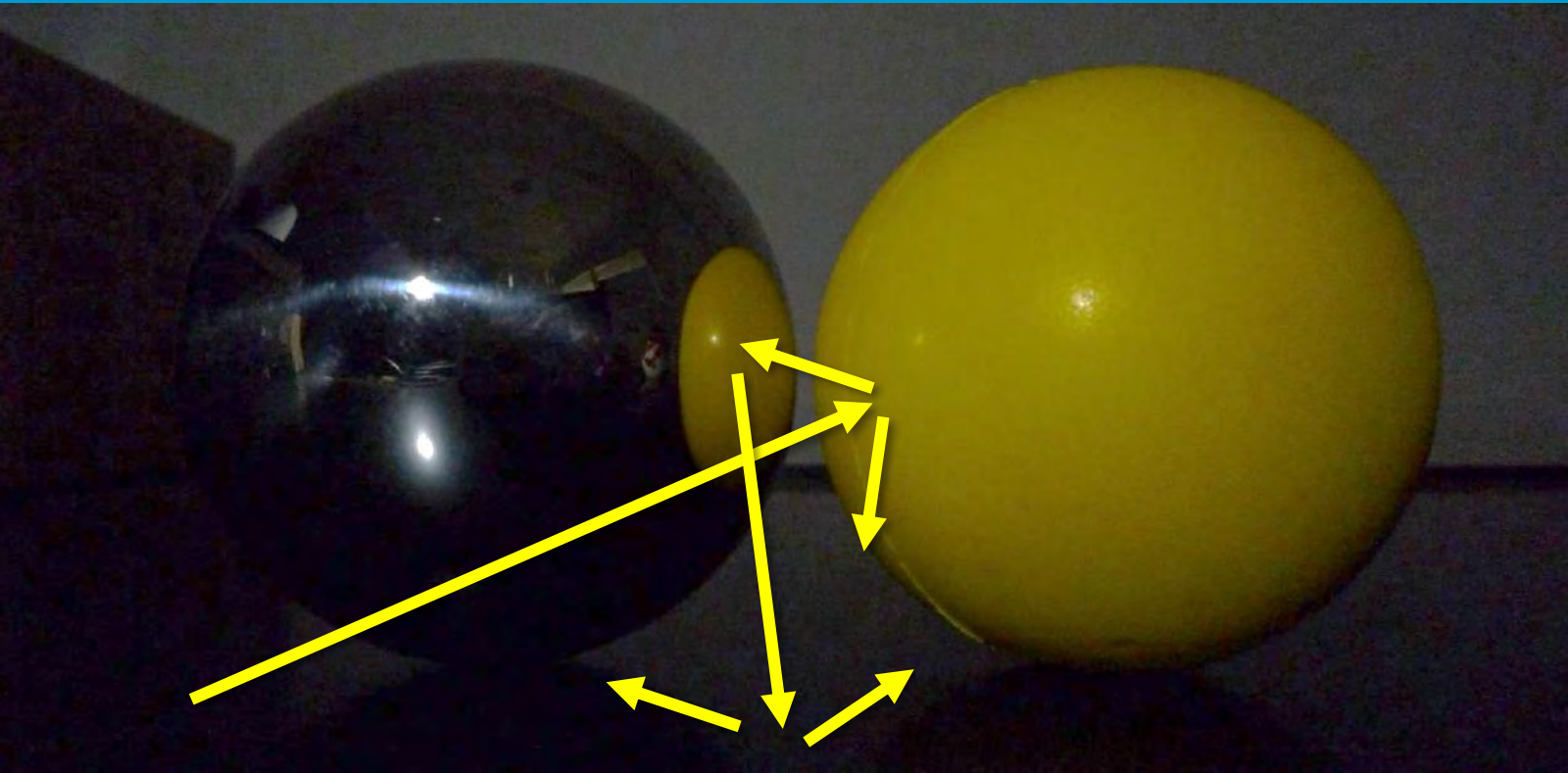
重要なのは光の知識



光の伝搬を定式化できれば 『逆問題』が解ける



光のふるまいは非常に複雑



ウッドハム, 1980

陰影からの形状復元の先駆者のひとり

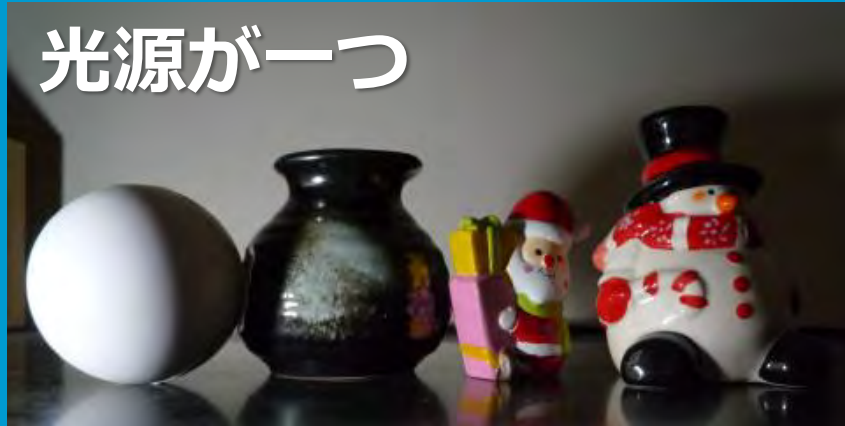
光のふるまいを単純化し、
さらに複数の陰影画像を利用
すれば逆問題を定式化できる

Woodham, "Photometric Method for Determining Surface Orientation from Multiple Images", 1980

フォトメトリックステレオ

【仮定①】 一度の撮影で光源は一つ 当てる方向もわかっている

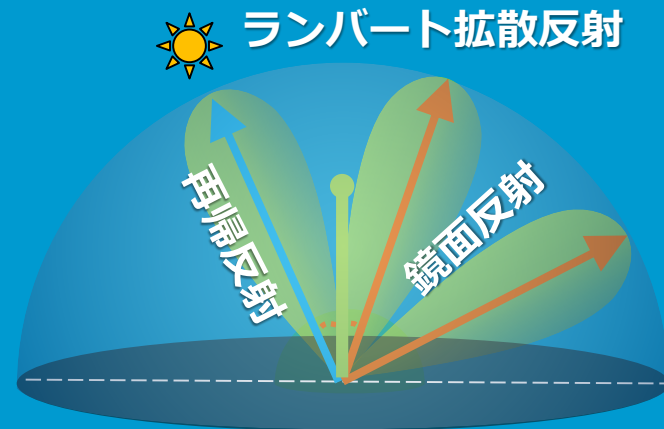
光源が一つ



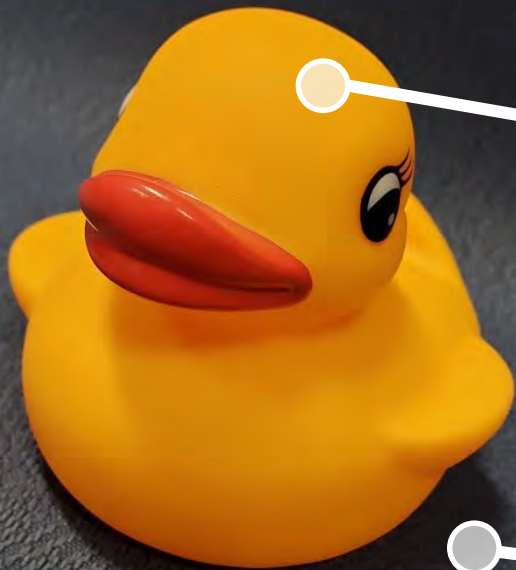
光源が複数



【仮定②】 ランバート拡散反射のみ 鏡面反射は考えない



【仮定③】 陰はあっても影はなし



陰

光が表面で反射
モデル化が簡単

影

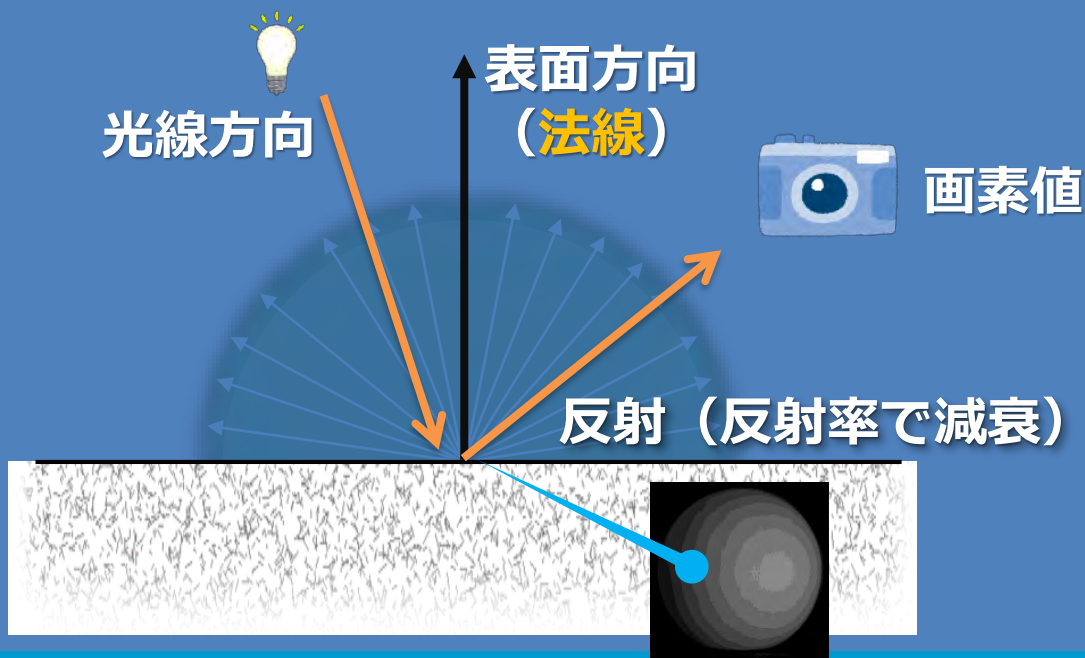
光が遮られる
モデル化が困難

画像生成モデル

各表面点に反射する光線経路の数理モデル

ウッドハムの画像生成モデル

※これを数式で表します



画像枚数 $\geq$ 未知パラメータ数 で逆問題が解ける



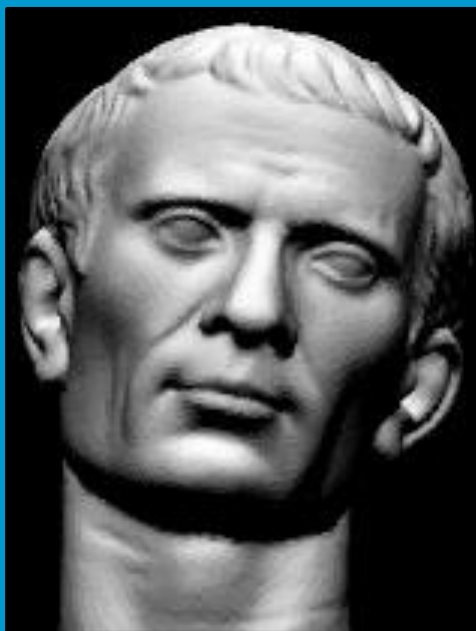
反射率（1自由度）

単位法線ベクトル（2自由度）

ウッドハムのモデルには未知パラメータが3つ

つまり、3枚以上の画像が必要

※多い方が精度が上がる



ウッドハムのモデルにはパラメータが3つ

フォトメトリックステレオの出力は 奥行きではなく表面法線





法線からメッシュへの変換
『法線統合』

**Q. 人間は1枚でも形状がわかる
なぜコンピュータはわからない？**



模様なのか、陰影なのか
コンピュータには区別できない

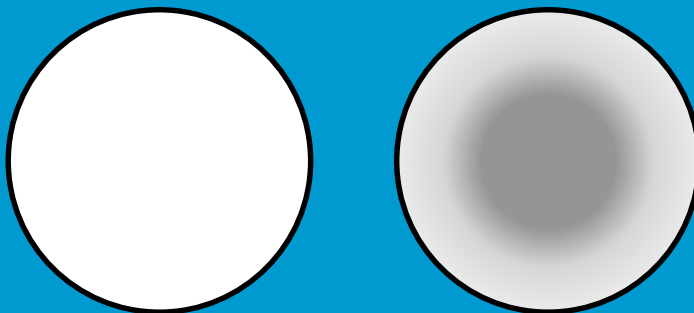


1枚の陰影画像からの形状復元

シェイプ・フロム・シェーディング (ホーン, 1989)

事前知識で制約を与える (大域最適化)

- ・ 表面で反射率が一定
- ・ 表面が滑らか (可積分条件) など



Horn, "Shape From Shading: A Method for Obtaining the Shape of a Smooth Opaque Object From One View", 1989

フォトメトリックステレオの発展

ウッドハムの手法では扱えない現象

鏡面反射



影



ノイズ



2010年代前半までのトレンド

外れ値分離アプローチ



画像

=



ランバート
拡散反射

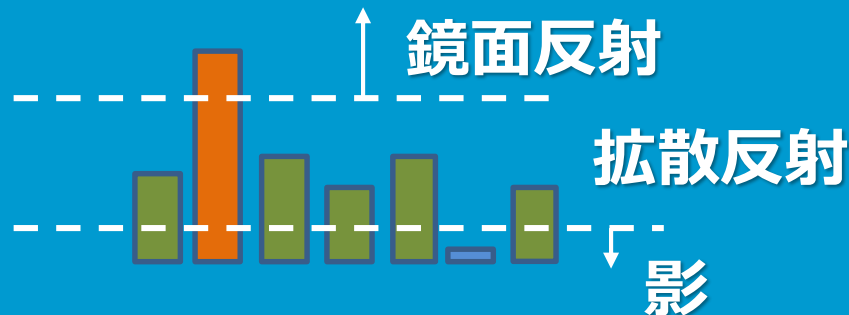
+



外れ値
(鏡面反射・影など)

分離方法は多種多様

閾値処理



発展形

グラフ最適化 (タン2005)

ランダムサンプリング (向川2007)

中央値フィルタリング (宮崎2010)

Big-Mアプローチ (ユ2010)

スパース回帰 (池畑2012)

外れ値扱いができない場合

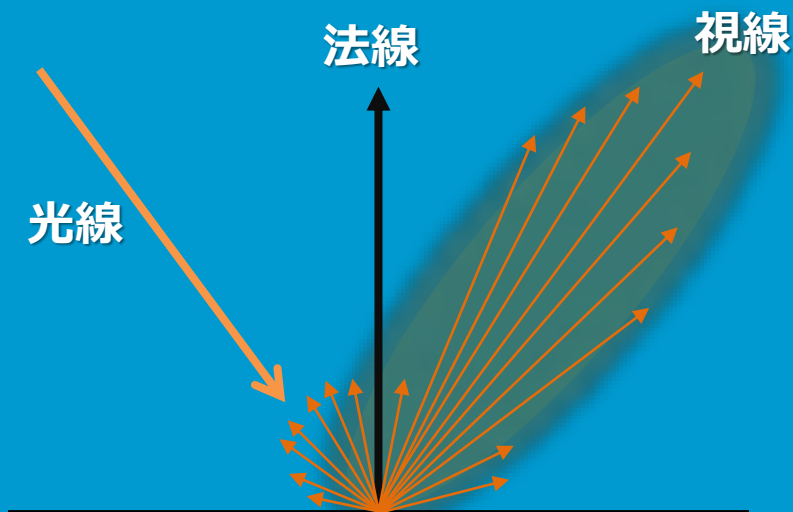
例えば、ランバート拡散反射よりも鏡面反射の割合が大きい時



複雑な反射モデルを使う (BRDF: 双方向反射率分布関数)

※ランバート拡散反射モデルは最も単純なBRDF

鏡面反射は視点方向に依存する



【材質により異なるBRDF】
フォン・モデル
ウード・モデル
クックトランス・モデル
ラフオチューン・モデル

Q.最初から複雑な反射モデル を使えば？

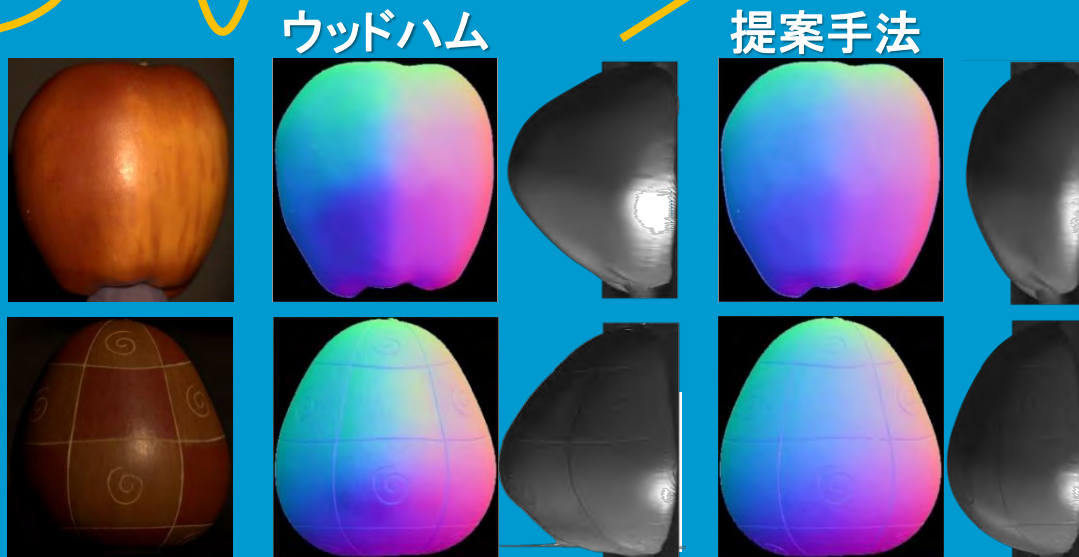


事前に材質はわからない

逆問題最適化が難しい
(局所解収束問題)

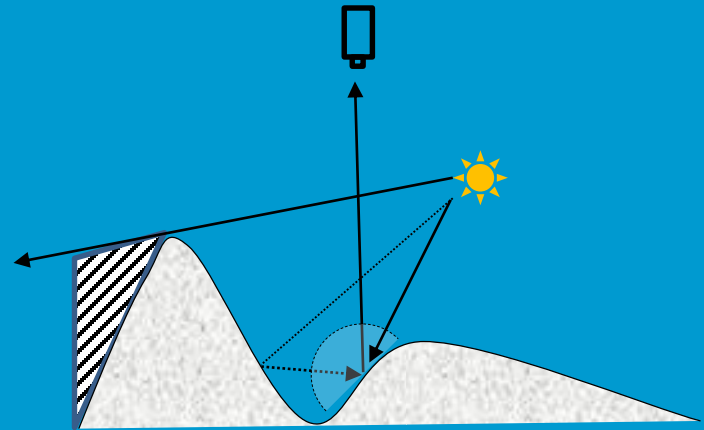
汎用的材質に対する フォトメトリックステレオ (池畑2014)

- 汎用的なBRDFを導入 (シャンドレーカー2011)
- 非線形のBRDFから線形の逆問題へ



最難関の一つ、非凸表面

一度目に反射した光がまた別の場所で反射、さらにそれが…

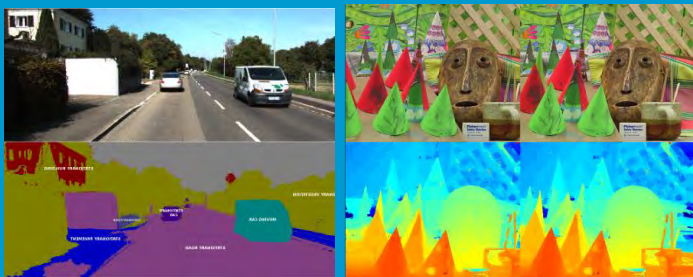


影や相互反射は単点では扱えない
つまり画像生成モデルが定義できない

深層学習に基づくフォトメトリックステレオ

余談：深層学習と 3Dコンピュータビジョン

構造化された問題

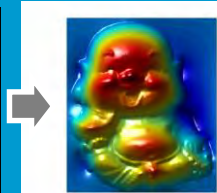


認識・セグメンテーション、2眼ステレオ

構造化されていない問題



多眼ステレオ



フォトメトリックステレオ

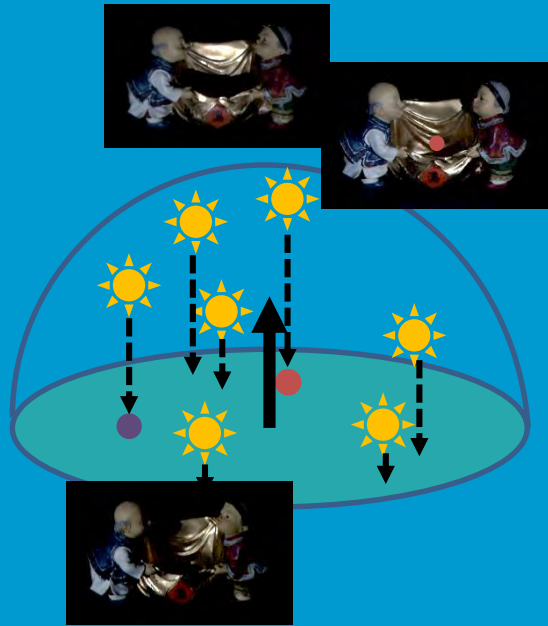
固定枚数の入力（順序一定）



N枚の入力（順不同）



深層学習に基づくフォトメトリックステレオ (池畑2018)



観測地図
(全陰影画像を1枚に集約)



ウッドハム



提案手法



正解

フォトメトリックステレオまとめ

- 複数の陰影画像から表面法線を復元
- ランバート拡散反射、汎用的な材質、非凸表面へと対象が拡大
- 光源方向がわからない、カメラが複数など異なる条件での研究も多数あり

むすび

単眼手がかり

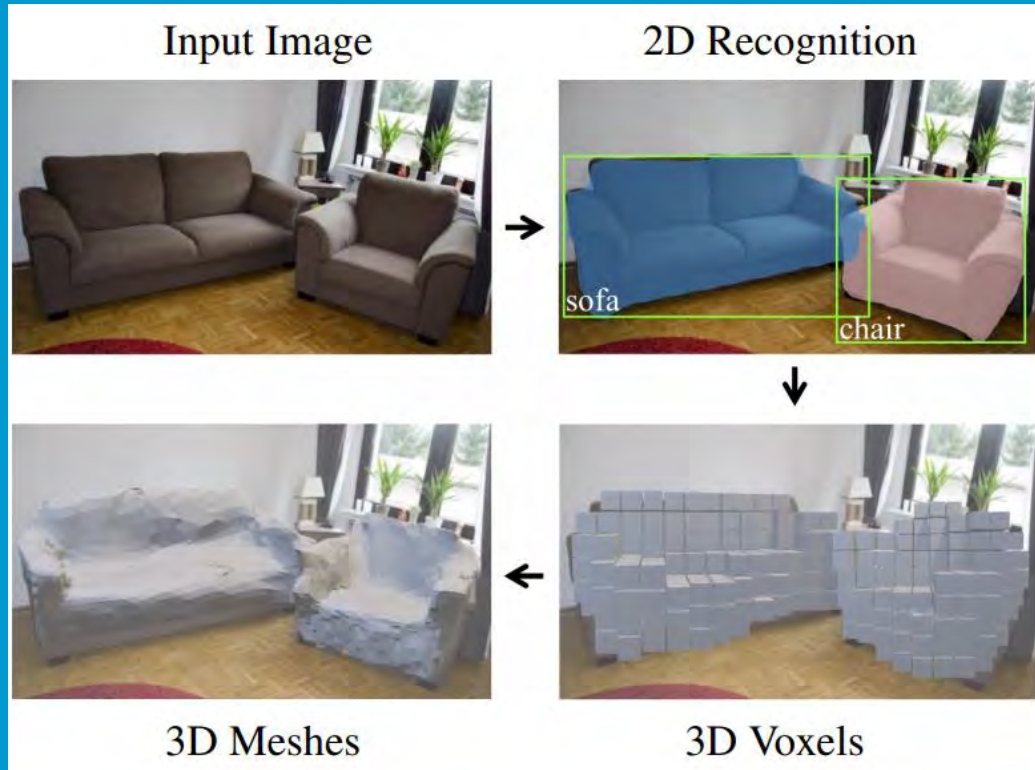
- ・ 線遠近法
- ・ 視野内の高さ
- ・ ぼけ
- ・ 遮蔽
- ・ 模様
- ・ 光のふるまい
- ・ 運動視差

両眼手がかり

- ・ 両眼視差

- ・ 受動型センシングの一部
- ・ それぞれが独立の研究分野
- ・ 能動型も含めればいくら時間があっても足りない！

高次の認識にもとづき 「見えないもの」を復元

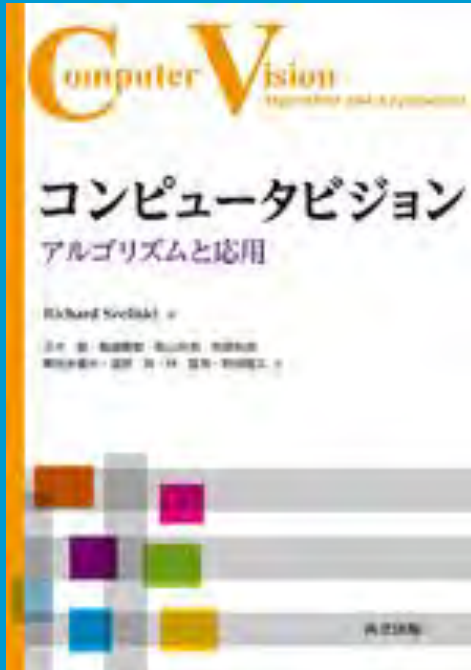


Gkioxari et. al., "Mesh R-CNN", ICCV2019 (to appear)

ヒトを超えて進化していくために

独立して研究されてきた知見を
統合していくことが重要！

3Dセンシングにご興味のある方 への参考図書



共立出版
「コンピュータビジョン
アルゴリズムと応用」



共立出版
「コンピュータビジョン
広がる要素技術と応用」

ありがとうございました