

Computer Vision: Algorithms and Applications

備考

著者

Richard Szeliski

1. はじめに

1.1. コンピュータビジョンとは？

私たち人間は、自分の身の回りにあるものを簡単に立体的に認識することができます。例えば、テーブルの上に置かれた花瓶の花を見ると、その立体感がどれほど鮮明に感じられるでしょうか。花瓶の表面を覆う光と影の微妙なパターンによって、花びらの形や透明感がわかり、花と背景が無理なく区別されています（図1.1）。また、額縁に入った集合写真を見ると、そこに写っている人の数や名前を簡単に数えたり、顔の表情からその人の感情を推測したりすることができます。知覚心理学者は、何十年にもわたって、視覚システムの仕組みを理解しようとしてきました。そして、その原理の一部を解明するために、目の錯覚¹を利用することができましたが（図1.3）、このパズルの完全な解決策はまだ見つかっていません（Marr 1982; Palmer 1999; Livingstone 2008）。

コンピュータビジョンの研究者たちは、画像中の物体の3次元形状や外観を復元するための数学的手法を並行して開発してきました。現在では、何千枚もの部分的に重なった写真から、環境の部分的な3Dモデルを正確に計算する信頼性の高い技術が確立されています（図1.2a）。また、ある物体やファサードの十分なビューセットがあれば、ステレオマッチングを用いて正確で高密度な3Dサーフェスモデルを作成することができます（図1.2b）。また、複雑な背景の中を移動する人物を追跡することもできます（図1.2c）。さらに、顔、服、髪の毛の検出と認識を組み合わせ、写真に写っているすべての人を見つけて名前を付けることも、ある程度成功しています（図1.2d）。しかし、このような進歩にもかかわらず、コンピュータが2歳児と同じレベルで画像を解釈する（例えば、写真に写っているすべての動物を数える）という夢は、まだ実現していません。なぜ視覚は難しいのか？それは、視覚が「逆

問題」であることに起因します。そのため、物理学的なモデルや確率論的なモデルを用いて、解の候補を明確にする必要があるのです。しかし、視覚の世界を複雑にモデル化することは、例えば音声を発する声道をモデル化することよりもはるかに困難です。

1.2. 歴史を振り返る

このセクションでは、過去30年間のコンピュータビジョンの主な発展 (図1.6) について、個人的に興味深く、時の試練に耐えてきたと思われるものを中心に、簡単なあらすじを紹介します。様々なアイデアの出所やこの分野の進化に興味のない読者は、1.3節の本の概要に進んでください。

1970s. コンピュータビジョンが登場した1970年代初頭は、人間の知能を模倣したり、ロボットに知的行動を付与したりするという野心的な課題のうち、視覚的な認識を担うものと考えられていました。当時、MIT (マサチューセッツ工科大学)、スタンフォード大学、CMU (マサチューセッツ工科大学) など、**人工知能やロボット工学の初期のパイオニアたちは、「視覚入力」の問題を解決すること、より高度な推論や計画などの難しい問題を解決するための簡単なステップだと考えていました。**よく知られている話では、1966年、MITのマービン・ミンスキーは、学部生のジェラルド・ジェイ・サスマンに、「カメラとコンピュータをリンクさせて、見たものをコンピュータに記述させることに夏を費やしてほしい」と頼んだそうです (Boden 2006, p. 781) 5。しかし現在では、この問題はそれよりも少し難しいことがわかっています6。

コンピュータビジョンは、すでに存在していたデジタル画像処理の分野 (Rosenfeld and Pfaltz 1966; Rosenfeld and Kak 1976) とは異なり、画像から世界の3次元構造を復元し、これをシーンの完全な理解への足がかりとして利用しようとするものでした。 Winston (1975) と Hanson and Riseman (1978) は、この初期の頃の古典的な論文を集めた素晴らしい本です。

シーン理解の初期の試みでは、エッジを抽出した後、2Dのラインのトポロジー構造からオブジェクトの3D構造や「ブロックの世界」を推論していました (Roberts 1965)。当時、いくつかのラインラベリングアルゴリズム (図1.7a) が開発されました (Huffman 1971; Clowes 1971; Waltz 1975; Rosenfeld, Hummel, and Zucker 1976; Kanade 1980)。Nalwa (1993) には、この分野の素晴らしいレビューがあります。エッジ検出のトピックも活発な研究分野であり、同時代の研究の優れたサーベイが (Davis 1975) にあります。

多面体ではない物体の三次元モデリングも研究されていた (Baumgart 1974; Baker 1977)。一般的なアプローチとしては、一般化された円柱、すなわち回転や掃引された閉曲線のソリッド (Agin and Binford 1976; Nevatia and Binford 1977) が用いられ、しばしばパーツ関係7 (Hinton 1977; Marr 1982) に配置される (図1.7c)。Fischler と Elschlager (1973) は、このようなパーツの弾力的な配置を絵画的構造と呼んだ (図1.7b)。これは、現在、物体認識で用いられている有力なアプローチの一つである (14.4節、Felzenszwalb and Huttenlocher 2005 参照)。