高校生のページ

コンピュータで視る

コンピュータビジョン分野
西野　恒、延原章平、川原　僚

我々の研究室ではコンピュータビジョンの研究をしています。コンピュータビジョンという研究分野を初めて聞く人も多いと思いますが、その名から想像できる通り計算機の視覚、すなわちコンピュータが物を見えるようにすることが目的になっています。もともとはいわゆる人工知能の一分野として生まれたのですが、今では独立した分野として大きく栄え、その成果を気づかぬうちにみなさんも日々の生活の中で活用しています。リアルタイムで自撮り映像の顔に仮想的にメガネや鼻をつけたり，パソコンを顔でロック解除したり、ゲームのキャラクタがスマートフォンを通して現実世界に重なって見えたり、様々な形で身近にコンピュータビジョンの技術が溢れています。自動運転や介護ロボットなど、コンピュータビジョン技術が社会基盤として実装されるようにもなってきています。これらの技術の根底には、顔認識と追跡や3次元センシングなどのコンピュータビジョンの基礎的な問題が存在しており、それらが製品レベルで実装され始めているわけです。

コンピュータビジョンは、我々の存在する3次元なり4次元の世界を2次元の画像列から理解するという、そもそも不良設定な問題を追究しているため、それらの課題の安定した解法を得るにはいろいろと根本的な仮定をおいています。例えば、ライダーと呼ばれる、パルスレーザーを打って帰ってきた時間によって3次元計測を用いる自動運転車は、世界がまばらな3次元点群でできていると思っていますし、物体認識システムにおいても元の画像ではなく、画像の勾配分布しか見ていなかったりします。すなわち、まだまだ簡素化された視覚世界にコンピュータビジョンは住んでいるわけです。

それに比べ、我々の見ている視覚世界はとても豊かです。同じコップでも様々な形、色、質感、材質の物がありますし、実物体表面は紙のような決まった拡散反射をするだけではなく、透き通ったりつやがあったり複雑な光との作用をします。さらには、我々人間自身の動きや見えが非常に複雑で豊かな視覚世界を織りなします。我々の視覚世界を、ただの3次元点群にとどまらず、よりその豊かさを見えるようにすることにより、視覚情報から物や、人や、環境を深く理解できるようにする、それがコンピュータビジョン研究の大きな目標になっているわけです。我々の研究室では、コンピュータビジョンを単純に「見る」技術から人間と同じように、あるいはそれ以上に実世界を視覚情報から知覚できるよう、すなわち「視る」ことができるようにすべく、特に「物を視る」「人を視る」、そして「よりよく見る」の3本の研究の柱を掲げ研究を進めています。

1. 物を視る

私（西野）は、アメリカに15年ほどいたのですが、フィラデルフィアというアメリカ建国の地に13年ほど住んでいました。フィラデルフィアは映画のロッキーが撮影された地として良く知られています。フィラデルフィア美術館の階段のふもとにはロッキーの銅像があって、スタローンが続編を撮影しに来るたびに訪れます。そのロッキーの銅像とスタローンを一緒に撮った写真を想像してみてください。そのような写真を見れば、我々はいともたやすく生身のロッキーとその銅像の区別がつきます。しかし、実際コンピュータビジョンで開発された物体認識のプログラムを使うと、スタローンは人間だと判断さ

図１：左の写真の中の素材を各画素単位で認識した結果が右に示されています。素材を認識することにより、例えば布でできた椅子が木でできた床にある、などとより豊かな情景理解が可能になります。

れるものの、銅像が何かは判別できません。

コンピュータは例えば人の検出をするときなどには，主に形を表す情報、例えば画素値の勾配分布などを見て認識しています。しかし、この銅像を識別するためには、人間の形をしていることがわかるだけでは不十分です。その形を作り上げているその物体の素材、例えば、髪の毛であったり、皮膚であったり服の素材がわかるからこそ、片方が本物の人間でもう一方は銅製のレプリカと判断できるわけです。コンピュータでも素材や素材にまつわる情報を画像から推定可能にできれば、物体認識だけではわからない、特に実世界の中で行動するために絶対不可欠な情報が得られます。例えば、道路が土ではなくてアスファルトで出来ていて、濡れているからすべりやすいとか、床に落ちている、形からは認識できないものが柔らかいタオルであるとか、買ったいちごがくさっているから食べられないとか、形だけでは判断できない、でも触ったり、その上を走ったり、それこそ食べたりと、現実の世界において行動を起こすために必要不可欠な情報が得られます。

素材認識は、非常に難しい問題です。なぜなら、形状を認識すれば済む物体認識とは異なり、それぞれの素材について、その見え方に非常に大きな幅があるからです。同じプラスチックでも、赤いコップにもなれば白いまな板にもなります。そこで、我々は素材をその特性、特に材質に関わる属性の集合として表現し認識することを考えました。例えば、布はふわふわしていて織られているけれど、固くて金属的ではない、というように素材の属性の集合として捉えるわけです。素材自体は大きな見えのクラス内分散があるものの、材質の方が特徴的な見えに現れるため認識しやすく、結果として素材が認識できるというのが狙いです。我々はそのような深層学習モデルを導出し、局所視覚情報から素材を認識しつつ、物体や場所も特定して、それらの大局的情報を組み合わせることによって画素ごとに精確に素材が認識できるようにしました。例えば、図１の左下の画像ですと、ソファは布で、床は木で、壁は石膏で、窓はガラスで出てきていると、正しく認識できます。そうすると、例えばコップはコップでも紙コップだからロボットが持ち上げるときには力加減をするとか、そういったことが見ただけで計画できるようになるわけです。

このように、物の見た目からその物体に関わる有益な情報を視覚からより深く探り出す、「視る」ためのコンピュータビジョンの研究を進めています。

2. 人を視る

我々の視覚世界を構成する様々な物体の中でも、人は特に大きな関心の対象です。その人が誰かを特定したり、どこに行くのかその行動を予測したり、果ては気分が落ち込んでいるのか内面状態を類推したり、我々人間は人の見た目から様々な情報をたやすく読み解きます。コンピュータにもそのような能力を備えることができれば、危険な行動を予知したり、助けの要る人にタイミング良く手を差し伸べたりと、より住みよい社会が実現できます。

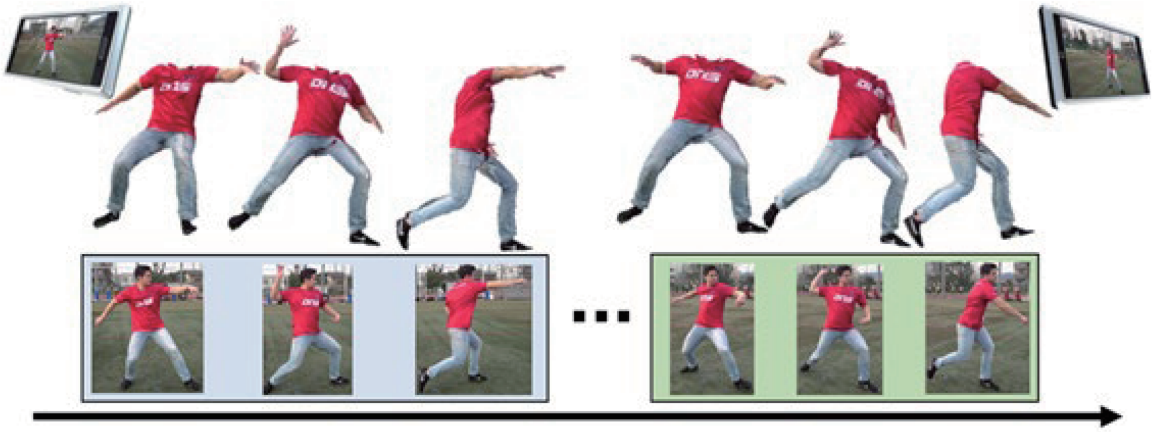

図２：友人がピッチングなど繰り返し動作を行っている様子を気軽に少数の複数視点から撮像した映像から、自由な視点で見返せるような３次元モデルを構築できることを示しました。

コンピュータビジョンにおける人に関する研究は、主に監視への応用を中心に黎明期から行われてきました。近年では、特に映像から対象人物の人体形状を３次元復元する研究が盛んに行われています。このような技術はマーカーレスモーションキャプチャとも呼ばれ、復元された３次元人体形状はゲーム、映像制作、スポーツ科学、医療など様々な分野で活用できます。しかし、これまでの研究は多数のカメラを人物の周囲に配置して同時に観測して三角測量を行うことを必要としていました。そのため、非常に高価で大規模な装置を導入する必要がありました。

一方でわれわれ人間は、ある一つの方向から人を見るだけで、ボールを投げる、といった動作を３次元的に、その腕の振りの軌跡などの動作だけではなく、行動主体の人体表面形状や骨格まで把握できます。このように非常に少数の視点から人を見るだけでその人体形状を３次元復元する方法をコンピュータで実行可能なアルゴリズムとして実現できないか考えました。人間がこのようなことができるのは、「同じような運動を見たことがある」ことが主な理由だと考えられますが、まずこれを「同じような運動を別の方向からも見たことがある」と読み替えます。すると同じような動作をしている被写体を異なる視点から時間を越えて観測、つまり「同じ運動であること」を手がかりとして、過去の観測と現在の観測の間で時間的同期を取り、さらに観測視点の相対的な位置関係を推定できれば、従来手法で必要としていた「多数の相異なる視点から被写体を同時に観測して三角測量する」という条件を擬似的に満たすことができると考えました。

同じ運動、つまり繰り返し動作を別の視点から別の時刻で観測したとして、どのようにすれば時間的な同期と空間的な相対位置関係を知ることができるでしょうか。その鍵は被写体が人間であると知っていることにあります。人物を撮影しているのであれば、たとえ撮影時刻や撮影視点が異なっていたとしても、右肘同士、左膝同士など、同じ人体部位同士を任意の画像間で対応付けることができます。視点から撮影したのであれば、部位間の対応関係から三角測量によって各部位の３次元位置を得ることができます。すなわち繰り返し動作をしている人物を撮影しているということさえ仮定できれば、カメラの撮影時刻、撮影位置、そして被写体の３次元形状および運動の全てにおいて辻褄が合う解釈を同時に推定することができます。すると、図２に示すように、たったの４つの異なる視点から友人が投球動作を繰り返している様子をスマートフォンで撮像するだけで、その友人の３次元形状を復元し、自由な視点から見ることができるようになります。

このように人を視る研究例として、人を見ているという事実を利用して、簡便に撮像された映像からより人間に近い３次元的理解が実現できることを示しました。

3. より良く視る

コンピュータにとっての目は，人間の目と同じでしょうか？コンピュータの目がカメラであると考えると、その目は人間のように２つとは限りませんし、可視光だけでなく赤外線や紫外線も撮像できます。コンピュータの目は人間の体の構造などの制約は受けないので、人間とは異なる見方や人間に見えないものまでも見ることができます。さらに，単一の画像情報に限られることなく、計算処理によって多くの画像を組み合わせることでより豊かな視覚情報を得ることができます。例えばデジタルカメラはある時間だけ光を露光して写真を撮影しますが、表現できる明るさ（ダイナミックレンジ）が有限であるため、写真の暗い部分が黒くつぶれたり、またライトのような明るい部分が白く飛んでしまったりします。そこで露光時間を変化させながら複数枚撮って後から組み合わせることにより、明暗の差が大きいシーンでも広い範囲の明るさの情報を取得できる HDR 撮影も実現できます。このように、人間の視覚の枠を超えた情報の取得と処理により、人間と同じように、さらには人間でさえ見えなかったものも視えるようにする研究も盛んに行われています。

人間とは異なる見方によって今まで見えなかったものが視えるようになる例は、身近なところにも存在しています。京都の金閣寺で水面に映る金閣寺の写真を撮ったこと、あるいは見たことはありませんか？水面に建物が反射されて見えるとき、人間にとっては逆さまの像が同時に見えているだけです。なぜこのように見えるかというと、水面が鏡となっているからです。ということは、みなさんも学校で学習した通り、この鏡となっている水面の下から仮想的にもう１つの視点を持ってその建物を見ていることになります。つまり、ちょうど我々が両目を使って一つの物を見たときに、各物体表面点の視差から三角測量によって３次元で見ることができるように、水面反射では上下で２視点の視差が生じる両眼視になっているわけです。さらに、水面反射はフレネル反射と呼ばれる、光の入射角によって反射強度が変化するものであり、水面反射像は実像とは異なった明るさで撮像されます。したがって、同じ明るさを異なる露光時間で撮像したのと同じことになり、HDR でその画像を復元できることになります。つまり、水面反射を含めて撮像されたたった一枚の画像から HDR で３次元復元が可能となります。図３のように、平等院のたった一枚の写真から、見栄えの良い 3D モデルができるわけです。

この例のように、我々はコンピュータにとっての視覚として，人間が直感的に視えていなかったものに着目し、見えないものを視えるようにするための研究に取り組んでいます。

4. まとめ

我々の研究室では見るだけではなく「視る」、本当に知覚として見るということにコンピュータビジョンを昇華していきたいと思っています。人を認識したり追跡したりするだけではなく、その一挙手一投足を見ることにより、その人の気分や状態、さらに考えていることややりそうなことを知る。すなわち、その人の行動をその人の内面を映す鏡として人を視る。物体を単に認識したり形状復元したりするのではなく、その素材や材質にとどまらず、重さややわらかさ、壊れやすさや使い勝手など含む様々な属性

図３：京都のお寺など、水面に反射した姿とともに撮像された建築物の一枚の画像（左）から、彩り豊かな３次元モデルを復元できる（右の２枚）ことを示しました。

を視覚として理解する、物を視る。さらには今まで見えなかったもの、人間でも見えないものを視えるようにする。それらを実現することを目指して日夜研究していますので、ご興味がある方はぜひ我々の研究室のホームページ等を見てみてください。