

五十嵐 健夫 *Igarashi Takeo*

東京大学 大学院情報理工学系研究科 教授
2017年度～22年度 CREST研究代表者

コンピューターシステムを使いやすくするユーザーインターフェース(UI)技術や機械と人間のインタラクションは、AI・機械学習の変革期である現在、ますます重要になってきた。東京大学大学院情報理工学系研究科の五十嵐健夫教授は、頭の中にあるイメージを即座に3DCG化する技術やアニメーション生成技術、仮想試着技術、AIの誤判断を未然に防ぐ「共起バイアス」除去技術などの高度なAI応用技術を次々と開発し、人と機械のより良い関係を追求し続けている。

特集

OVERVIEW

人と機械のより良い関係を追求 高度なAI応用技術、次々と開発

ユーザー自ら3Dモデル生成 世界のクリエイターを震撼

画面に描いた創作キャラクターの2次元(2D)スケッチが次の瞬間、ぬいぐるみのような量感を備えた3次元(3D)のコンピューターグラフィックス(CG)に生まれ変わる。そんな仕組みを今から24年以上前に考案したのが、東京大学大学院情報理工学系研究科の五十嵐健夫教授だ。1996年、同大学の博士課程に在籍していた五十嵐さんは、インターネットで訪れた米国ブラウン大学で、多面体として表現された3Dモデルから手書き風のイラストを描き出す

CG技術を目にした時に、人生最大の衝撃を受けたという。

「もともと3DCGに興味があり、ペン書きのイラストを立体化する仕組みを模索していました。ブラウン大学のCG技術を見た瞬間『この逆をやってみよう』とひらめきました」。当時は世界初の長編フル3DCGアニメーション映画「トイ・ストーリー」が公開された翌年で、3DCGに世界が注目し始めた時代でもあった。その後、五十嵐さんは2Dスケッチの3Dモデル化システムを研究課題に据え、99年に世界的CGカンファレンスである「SIGGRAPH 99」で「Teddy」と名づけた2Dスケッチの3Dモデル化

技術を論文発表した(図1)。

このシステムは発表と同時に世界の研究者やクリエイターを震撼させるセンセーションを巻き起こした。従来の3DCGは、専門スキルを持つエンジニアやクリエイターだけが創作するものという認識が一般的で、ユーザーはその結果を利用するしかできなかった。Teddyは3DCG創作をプロの手からアマチュアの手元に手繰り寄せ、特別な訓練を受けていなくても簡単に3Dモデルを生成・編集できる世界を作り出したのである。

この成果を基に、五十嵐さんは「空間的キーフレーム法」の開発にも成

功した。画面上での極めて少ない操作で、スケッチした創作3Dキャラクターを滑らかに動かし、リアルタイムなアニメーションを生成する技術だ。これらのシステムは五十嵐さん自身も驚くほどの大きな反響を呼び、技術の一部は3Dモデル作成の定番PCソフト「Shade」などへの商用化が実現。Microsoftの「ペイント3D」やAdobeの画像系ソフトなど、似たような製品も世に出回るようになった。

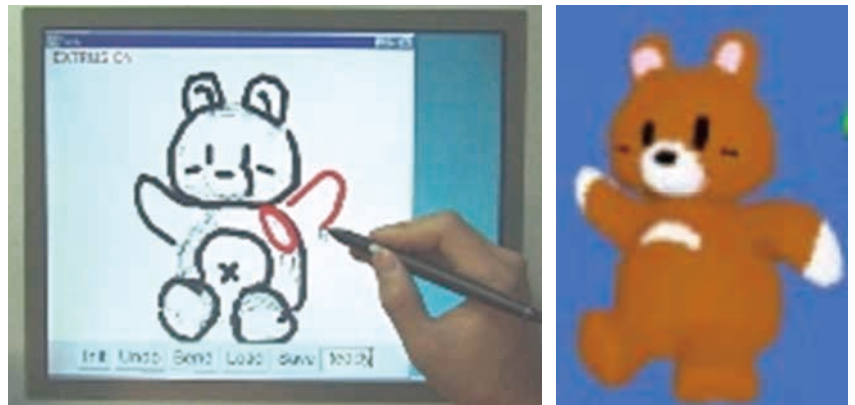
体型・ポーズごとに映像化 「仮想試着法」3年かけ開発

ただし、3DモデルやCG作成の効率化だけが五十嵐さんの研究目的ではない。「コンピューターシステムを利用しやすくするためには、人間とシステムはどのように関わるべきかというインタラクションも考える必要があります。人間が頭の中で思い描くことをシステムにどう伝え、システムの出力結果を人間がどう利用するのか。これが、現在まで続く私の研究テーマです」。

大学院生時代には米国のゼロックス・パロアルト研究所、マイクロソフト・リサーチ、カーネギーメロン大学でのインターンを経験し、日本電信電話(NTT)やリコーなどでの企業実習も重ねてきた。しかし、民間企業では開発技術は特許取得が求められ、商品化はその企業のみに限られる場合が多いと感じた。これに対し大学は研究の自由度が高く、また技術利用や実用化が広がる可能性も高いと考え、大学での研究を選んだという。研究者としてUIやインタラクション技術を究めていく中で、AI・機械学習と人間の関わりについても、重要なテーマとなっていった。

17年に、五十嵐さんはJSTのCREST「データ駆動型知的情報システムの理解・制御のためのインタラクション」の採択を受けた。CRESTではデータの生成、学習プロセスと

図1 「Teddy」による3DCG創作のイメージ



ディスプレイ上に描いた2Dスケッチ(左)から3Dモデルを自動生成する(右)。自由な視点から3DCGを描画、リアルタイムな編集も可能にした(着色は異なるソフトウェアによる)。

いったAI技術に共通するテーマの他にも、学習した結果の利用時にシステムの中で起きていることを可視化し、ユーザーが適切に介入して望む結果を出力する手法を、いくつかの個別のアプリケーションを通じて開発・発表している。

ここからは、その中でも特に顕著な成果を3つ紹介する。1つ目は、体型や姿勢に対応した高度な「仮想試着法」だ。着てみたい服を画面上で仮想的に試着できるシステムは既に実用化されているが、単純な3DCGや汎用的な深層学習モデルを応用したシステムでは、リアルな試着画像を即座に生成するのは困難だ。そこで五十嵐さんらは、体型や姿勢を自在に変化させることができるマネキン・ロボットを用意するところから始めた。

マネキン・ロボットに服を着せ、さまざまな体型の人のいろいろなポーズを模倣し、数万枚に及ぶ写真を撮影した。その写真データを基に深層学習を実施した

結果、体型やポーズごとの服のシワのつき方や、体の部位ごとの服の余裕などを詳細に映像化することに成功。オンライン店舗を想定した仮想試着の他、オンライン会議映像における服装を変換する実証なども行っ

図2 仮想試着システムの利用イメージ



https://www.jst.go.jp/pr/jst-news/backnumber/2021/202112/pdf/2021_12_p14-15.pdf

さまざまな衣服の試着結果。マネキン・ロボットによって取得したデータにより、リアルな試着画像を合成できる。

ている(図2)。

AI・機械学習はデータとアルゴリズムの世界と考えられがちだが、具体的な映像生成にはデータを作るための「ものづくり」要素と忍耐強さも必要になる。「プログラムは短期間で完成しても、学習には長時間かかります。撮影は2時間程度で終わっても、その学習のためには2晩かかる。失敗してやり直すとまた2晩で学習し直す。機械学習につきもの問題ですが、仮想試着法の場合、着想から実現までには約3年を要しました」と五十嵐さんは開発の苦勞を語る。

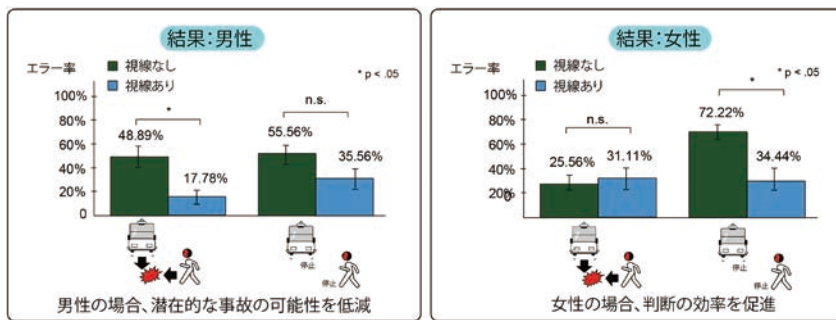
車の「目」を見て横断判断 自動運転の交通リスク低減

UIやインタラクションの研究は必ずしもシステム側の技術だけの話ではない。五十嵐さんらは、人と機械によるより良い呼応を創出するため、AIの判断を人間がわかるように可視化する仕組みも模索した。それが2つ目の成果である「目」を持つ自動運転車だ。この自動運転車は、ドライバーであるAIが注意を払っている方向を、フロント部分の「目」の角度で示し、実験ではそれを見た歩行者による道の横断を計測した。

歩行実験の結果、歩行者は自動運転車の目が自分を向いていれば「自分が認識されている」ので横断しても問題ないと考え、逆に目が他の方向を向いていれば「車は自分に気付いていない」ので、横断は危険と判断する傾向が見られた。このように、目玉の向きでAIの注意方向を示すことで、人間は的確な危険回避行動がとれることが多く、交通リスク低減に役立つことが示された。また、危険回避行動には男女差が大きいこともわかった(図3)。

「将来的には自動運転車が搭載するAIと目の動きを結びつけて、視線がそのままAIの状況を表すようになると仮定しての研究でした」と五十嵐さんは総括している。自動運転車

図3 「目」を持つ自動運転車に対する歩行者の反応を実験で調査



AIが注意を向けている方向が歩行者側なら「安全」、そうでなければ「危険」と判断し、横断すべきかどうかの人間の判断を調査した。実験では車も「目」も手動、歩行者役の人はVR映像で実施している(上)。グラフは「目」のある場合とない場合の歩行者の判断の違い。男女で傾向に大きな違いも見られる(下)。

が一般道を走る近未来に向けて、自動運転車の判断をどう人間に伝えれば人間が正しく危険を避けることができるのか、今後はますます重要なテーマになっていくだろう。他のAI応用分野でも同様に、人間とのインタラクションの研究は不可欠なのだ。

AIの間違った学習を防ぐ ワンクリックで適所誘導

3つ目の成果は、AIの間違った学習を防ぐための技術開発だ。機械学習では、最終的な判断を誤らないよう、学習のプロセスで人間の適切な介入が必要になる場合がある。例えば、海に浮かぶ船の画像から船の形態を学習するには「波」「海岸」などの画像内の要素ではなく、船に着目しなくてはならない。また「口紅」を多数の顔画像から学習するには唇に着目しなければならない。しかし、実際は「目

元の化粧をしっかりとっている人は口紅を塗っている」ことが多いため、目元部分を学習対象にしてしまうことがある。

結果、機械学習のアウトプットに誤りが生じてしまうのだ。このような問題を「共起バイアス」と言い、これを防ぐことが機械学習の大きな課題の1つになっている。共起バイアスの影響を取り除くためには、元のデータを適切に変更して再編成するか、特定領域をユーザーが直接指示して学習すべき箇所を示すアノテーションが必要だ。しかし、データセットの再編成にも、ユーザーによるピクセル単位での領域指示など、多大なコストがかかってしまう。

手間や費やす時間を減らすために、五十嵐さんらは画像表示画面のマウスによるワンクリックだけで特定領域をAIが認識できる技術を開発した。先ほどの船の例で言えば、船の画像なら船の部分を左クリックする

図4 AIの共起バイアスを取り除くワンクリックアノテーション技術のイメージ



CelebAデータセットの「口紅」に関するデータセットで、ワンクリックアノテーション前に深層学習ネットワークが認識した領域の例(上)と、微修正後の領域の例(下)。目や頬など、唇以外の化粧の参照が軽減されている。

と、人の目に見える船部分が識別対象となり、反対に周囲の水の部分を右クリックすれば、水部分が識別対象から外される。これにより、機械学習のトレーニング時間とコスト、アノテーションに関わる人間の労力が大きく削減可能になった(図4)。

注目したい部分を人間が調整して教えることを「アテンション誘導」と言う。今回の成果の肝は、対AIのアテンション誘導作業が非常にシンプルになったことだ。「同時に考案したアクティブラーニング方法を組み合わせることで、アテンション誘導に必要な時間を27パーセント削減し、学習の正確性も大幅に向上することが実証できました」と五十嵐さんは説明する。

新発想をひねり出すのに苦戦 手を動かし考えることが大事

五十嵐さんはアプリケーションの他、AIが提示するさまざまな画像をユーザーが選択していくことで頭の中のイメージに近い画像を生成していく技術や1個のスライダーだけで複雑・多様な画像調整を実現する技術など、多くのアイデアを実現すべく研究を続けている。最も苦戦するのは、まだ誰も思いついていない発

想をひねり出す段階だという。しかし、それを乗り越えればアプリケーション作りは既存技術を活用・応用しながら真っ直ぐ進めることができるそうだ。

「新しい発想のためには論文をたくさん読むこと、とにかく手を動かすこと、そしてひたすら考えることが大事です」。五十嵐さんが研究開発を進める背景には、人間とコンピューターとがお互いに情報を伝えあってより良い結果を作るための、UI・インタラクションの理想があり、それがCRESTの目標である「人間と情報環境の共生インタラクション基

盤創出」につながっている。

22年11月に米国OpenAIが公開した大規模言語モデルによる生成AI「ChatGPT」が世界中から注目を集めているように、現在はAI活用の歴史の変革期とも言われている。「私たちがやってきたことで何ができるのか模索中ですが、基本的な考え方は同じです。専門的な知識を必要とすることなく、ユーザーがシステムを思い通りに操れるようにするという課題に挑戦し続けたいと考えています」と五十嵐さんは今後の道筋を描いている。

(TEXT:土肥正弘、PFOTO:伊藤彰浩)



自分の好奇心を追求して
やりたいことをやるのが一番で、
それが世の中の役に立つことにつながります。
面白いと思うものを追求する
研究者を目指してください。