

顔画像処理の過去・現在・未来

森島繁生[†]

1987年以來の森島研究室で行われてきた顔画像処理を振り返り、現在の進捗状況について報告し、将来を展望する。

Past-Present-Future of our Face Image Processing

SHIGEO MORISHIMA[†]

The past research in face image processing proceeded in Morishima-lab is reviewed, present research is also reported and future topic is discussed.

1987年知的音声画像符号化の研究に着手した。これは送信信号に含まれる意味や感情等を理解し、伝送路はシンボル情報のみを伝え、受信側でリアルな信号を復元するという究極の情報圧縮を目指すものであった。これらの成果の具現化システムとして1995年のSIGGRAPHで、本人の1枚の正面画像を基に、逐次送られてきた音声信号にリップシンクさせ、表情変化もさせながら、実時間で来場者同志が画面を通して対話できる”Better Face Communication”を発表した。顔画像を撮影して瞬時にアバタ化し、SGI Onyxをフル稼働させながら、音声解析結果から顔形状変化を自動生成して、10fpsのリアルタイムでface-to-face対話できるシステムを、今でいうEmerging Technologyの前身のデモセッションで発表した(SGIがマシンを無償で貸与してくれた良い時代)。

その後、これは臨場感通信のプロジェクトへと発展し、1999年からATRにおいて音声翻訳と顔画像の入れ替えを含むビデオ翻訳の研究に従事した。また万博の名古屋開催が決まり、世界発の映像エンタテインメントを実現したいというオファーを頂いたことをきっかけに、映画に観客を全員キャストとして登場させる仕組み”Future Cast”の構築を行った。顔をレンジスキャニングし、Genericな顔モデルを本人の顔特徴点に整合させワーピングして3D顔モデルを生成。さらにこのワーピング結果に従ってキーシェープを本人向けにアレンジして、映画上映時はリアルタイムにブレンドシェープで、シーンごとの表情個性をリアルに再現する試みを行った。愛・地球博6か月間の開催期間中におよそ163万人が三井・東芝パビリオンのこのイベントを体験し大成功をおさ

めた。また2007年からハウステンボスに創設された”Future Cast Theater”においてこの技術は常設展示となり、その短編作品”Grand Odyssey”は7年経過した今も人気No.1のイベントとして上映が続いている。これは映像作品への自己没入あるいは仮想世界における感動共有という、映画コンテンツの別次元の娯楽性の可能性を切り開いた意義は大きい。

あれから10年あまりCPUやGPUの急激な進歩にも支えられ、顔の3Dモデル生成技術は1枚の正面顔画像から2秒以下で生成できるまでに進歩した(当時は3分)。また肌の質感もランバートモデルとほぼ同等に高速で、半透明物体の表面下散乱を模擬できるほどになり、市販のゲームソフトにもこの技術は多く採用されている。最近のSIGGRAPHにおいても、Face関連の発表は毎年コンスタントに行われており、質感、表情、動き、どの要素においてもスーパーリアルな域に達してきている。実際、映画制作に関わるチームの発表が多く、現場から遠い日本の大学の研究室が割り込めるレベルではもはやなくなってきている。

これからの顔画像処理に求められる技術は、顔が本来伝えるべき情報として重要な個性や感情の表現に尽きるだろう。どんなにアバタ生成の見栄えのクオリティが向上しても、本人らしさである個性の表現はまだ満足レベルには達しておらず、不気味の谷レベルである。この解決にはCG、CV、音響処理や数学、物理学等、多くの関連分野の英知と実際の現場のニーズや経験の蓄積が融合して初めて目標は達成される。

[†] 早稲田大学先進理工学部
School of Science and Engineering, Waseda University