

東映ツークン研究所が目指す新しいコンテンツ制作のかたち

美濃 一彦¹ 木下 紘¹ 池城 和夫¹

概要: 東映株式会社の一組織である東映ツークン研究所では、コンテンツ制作に革新を起こすことを目標として、コンピュータビジョン・コンピュータグラフィクスなどの様々な分野の最先端技術を横断的に取り入れた映像作品の制作に取り組んでいる。本研究所の代表的な取り組みとして、「バーチャルプロダクション」・「デジタルヒューマン」に関する技術開発が挙げられる。本稿では、まず、これらの技術開発について実案件を例に挙げ紹介する。次に、本研究所のデジタルヒューマン技術の紹介として、3つの観点（物理性・生理性・心理性）をもとにしたデジタルヒューマンの制作工程について述べる。最後に、デジタルヒューマン技術における本研究所の新しい取り組みとして、南カリフォルニア大学の Institute for Creative Technologies (ICT)が開発した Light Stage と機械学習を用いた顔形状復元に関する研究について紹介する。

キーワード: 東映ツークン研究所, バーチャルプロダクション, デジタルヒューマン, Light Stage, 機械学習

Toei Zukun Laboratory Leading Innovation for Film Production

KAZUHIKO MINO^{†1} HIROMU KINOSHITA^{†1} KAZUO IKESHIRO^{†1}

Abstract: To achieve innovation for film production, Toei Zukun Laboratory, which is a department of Toei Company Limited, has been developing video and image products which are applied advanced technologies such as computer vision and computer graphics. As representative projects proceeding by our Laboratory, there are “Virtual Production” and “Digital Human”. In this paper, we firstly introduce details of these projects with examples in actual works. Secondly, as introduction of Digital Human technology in our laboratory, we describe production process based on 3 points of view (Physicality · Physiology · Psychology) for Digital Human. Finally, as a new project in the Digital Human, we introduce a study for facial shape reconstruction with machine learning using data set obtained from Light Stage developed by Institute for Creative Technologies (ICT) of University of Southern California.

Keywords: Toei Zukun Laboratory, Virtual Production, Digital Human, Light Stage, Machine Learning

1. はじめに

東映株式会社は、創立 1951 年、従業員 343 名 (2019/3/1) の会社である。これまで、映画、テレビ・ビデオ・アニメーションほか、多様な映像製作と、それらの作品を活用した多角的な展開によって、質高く健全なエンターテインメントの提供に努めてきた。新たな興行形態や映像配信サービス、ソフトパッケージの多様化に伴い、映像娯楽産業を取り巻く環境は劇的な変化を続けている状況の中、東京・京都と東西 2 拠点の撮影所を中心に、一貫した製作体制のもと様々な作品にチャレンジしている。

東映株式会社の一組織である東映ツークン研究所では、最先端の映像技術を横断的に取り入れ、コンテンツ制作に革新を起こすことをミッションとし、「バーチャルプロダクション」や「デジタルヒューマン」など様々なテーマを持って日々活動している。

2. 事例紹介

本章では、本研究所がこれまでに取り組んできた、「バーチャルプロダクション」・「デジタルヒューマン」に関する技術開発について、実案件を例に挙げ紹介する。

2.1 TV×AR Experience

本研究所では、NHK 技術研究所が提唱する、「TV と AR (Augmented Reality) を組み合わせる未来の視聴スタイル」のコンセプト映像を担当した。テレビの中と外の空間をシームレスに融合、演者の立体像は全身を丸ごとデジタル化して伝送されるため、AR によってあらゆる角度からの視聴が可能となる。テレビが本来持つライブ感などのメリットを損なわずに、AR のメリットを掛け合わせた新たな視聴体験を可能とした。多様化するメディア、デバイス、視聴スタイルに柔軟に対応できる NHK 技術研究所の「コンテンツ伝送技術」と、本研究所の「プロダクション技術」とが融合し、将来の情報メディアの一つの可能性を具現化したプロジェクトとなった。

2.2 スペクタクル能

2020 年夏、日本の伝統芸能で世界最古の舞台芸術といわれる「能」と最新映像技術が融合した『神・鬼・麗 三大能 ∞ 2020』(図 1) が銀座の丸の内 TOEI で公演される。

「バーチャルプロダクション」というコンセプトのもと、空間トラッキング、フォトグラメトリ、モーションキャプチャ、ゲームエンジン、リアルタイム VFX、プロジェクシ

¹ 東映ツークン研究所
Toei Zukun Laboratory.



図1 スペクタクル能
Figure 1 Spectacle Noh.

ョンマッピングなど、様々なテクノロジーを横断的に組み合わせ合わせた新しい映像制作スタイル、新しい表現を確立することが本研究所の進むべき道だと考えている。制作手法に限らず、アウトプットまでをどのようにプロデュースすることも重要となり、柔軟な製作体制が求められている。

2.3 デジタルヒューマン

本研究所は、設立当初からデジタルヒューマンに関する様々な取り組みを行ってきた。そして、2017年にデジタルヒューマンに関する本格的な開発に着手することを決定。翌年の2018年には、既存のCGチーム内で特別チームを編成、緻密な制作ワークフローを体現できる人材育成を始めた。そして、2019年、新たにデジタルヒューマン研究チームを発足させた。

3. デジタルヒューマン技術

デジタルヒューマンは、現実の人の代替を目的として、CGを用いて作られる精巧な人物表現である。本章では、プロダクション向けのデジタルヒューマン製作について述べる。プロダクションにおいて、現実の演者と同等に求められる演出的貢献と、製作期間などの物理的制約に対し、デ

ジタルヒューマンの最大の課題である“不気味の谷”をいかに超えるのかを以下の3つの観点から述べる。

- ① 物理性：そこに在るか（モデル・シェーダー）
- ② 生理性：生きているか（リグ、アニメーション）
- ③ 心理性：感情はあるか（アニメーション）

3.1節では、モデル・シェーダーでの物理性の表現について、3.2節では、リグ・アニメーションでの生理性、心理性の表現について述べる。

3.1 モデル・マテリアルでの物理性表現

本節では、モデル（形状）・シェーダ（見た目）によるデジタルヒューマンの物理性表現について述べる。現実とCGの差は情報量にある。人が人を見た時に感じる印象よりも、実際にはより細かな情報が視覚を通して無意識に記憶されているが（例えば形状の非対称性や肌の透過感など）、CGで描画する際にその細かな情報が欠落していると無意識にその差を感じ取り、否定的感情を示す“不気味の谷”に落ちる。これを製作に置き換えた場合、無意識に知覚しているレベルの情報をゼロから描き上げられるアーティストは多くない。そのため、昨今の傾向としてはPhotogrammetryを用いたスキャンデータをベースに、アーティストが編集を加えることが主流となっている。図2に示すのは、スキャンされたデータを元にモデルとシェーダーを編集し、人物の印象を変えた例である。観測者のバイアスや環境による影響を限りなく排除した状態で、物体そのものが持つ物理的な表面特性をより多く正確に得ることができれば、プロダクションにおける人物表現は品質と生産性を向上させることができる。

3.2 リグ・アニメーションでの生理性・心理性表現

本節では、リグ（動きの仕込み）（図3）・アニメーション（動き）による生理性、心理性の表現について述べる。例えば眼球運動で見る生理性とは、固視微動や瞳孔の収縮など無意識下で行われている動きを指す。心理性とは例えば視線であり、目線の動きだけでその人物の心情を示すこと

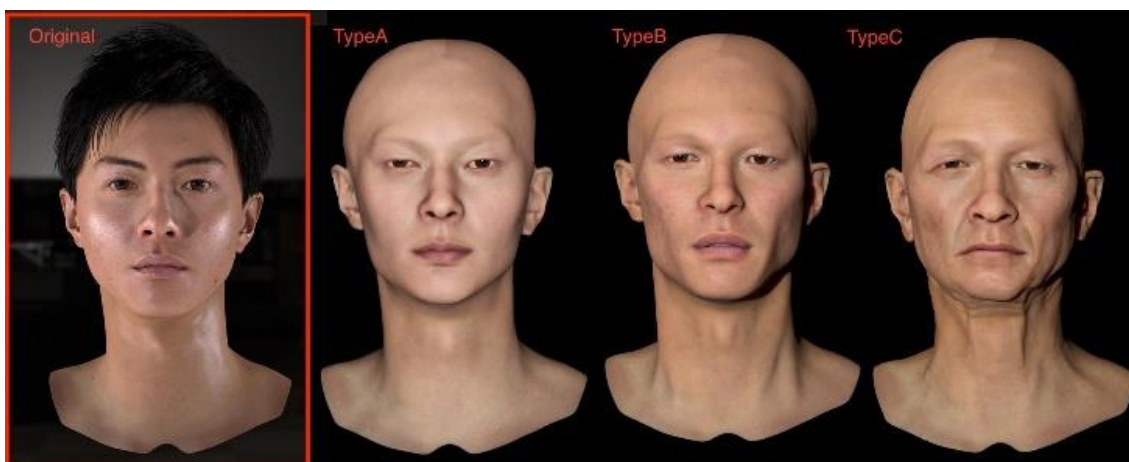


図2 Original にアーティストが手を加え別の人物を3種作成した例
Figure 2 Examples created by redesigning original by artist.

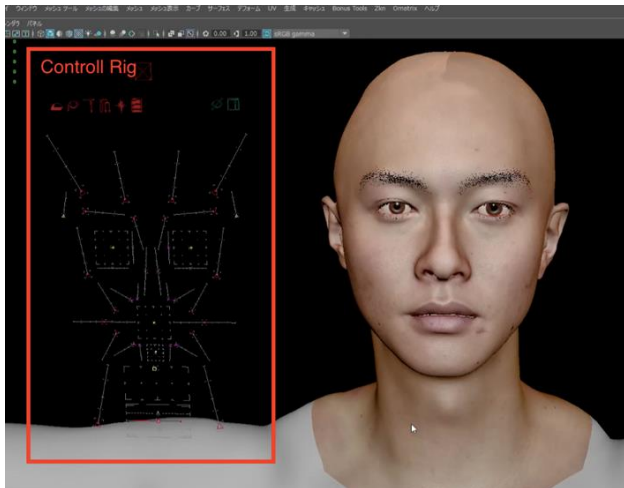


図3 表情を制御するリグ
Figure 3 Rig for controlling facial expression.



図4 Light Stage
Figure 4 Light Stage.

ができる。これらもまた、現実の人物から動きを得る為にモーション（フェイシャル）キャプチャが用いられるが、センシングデバイスの過度な着用は動きの制限や負荷、取得できる情報量とトレードオフに可動性を考慮したものが用いられることが多い。本研究所では、この捉えきれない小さな情報を補う為に、反復的、反射的な生理運動は定量化しリグの中で自動化される。ただし、心理性においては、演技内容をアーティストが解釈し補うに止まっている。これは、前述した動きの制限の理由から、現状行なわれてい

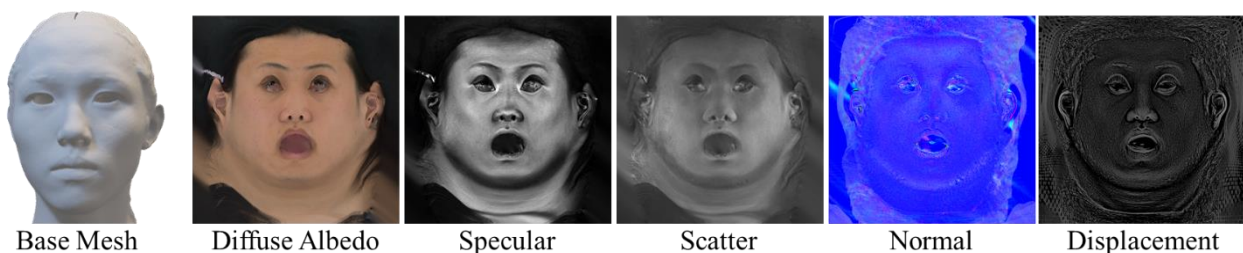


図5 Light Stage で取得した顔の形状・テクスチャ
Figure 5 Facial shape and textures obtained by Light Stage.

るフェイシャルキャプチャが2D画像の輪郭追跡もしくは、低ポリゴンでの頂点間の3次元的な動きの解析に留まっている為であり、表情の符号的解釈のみで解決しきれないことが実情としてある。より高精細な3D次元情報としての表情変化をシーケンシャルに捉え、かつ自然な動きを損ねないキャプチャ方法が今後の課題となる。

4. Light Stage を活用した研究開発

本研究所の新しい取り組みとして、Light Stage [1][2][3][4] を活用した研究開発について述べる。本研究所で稼働している Light Stage の外観を図4に示す。Light Stage は、南カリフォルニア大学の ICT が開発した計測システムである。Light Stage を用いることで、計測対象の詳細な形状・テクスチャの取得が可能となる。Light Stage で取得した顔の形状・テクスチャを図5に示す。本研究所では、Light Stage によって取得した形状・テクスチャデータと機械学習を用いた、画像・動画から対象の形状・テクスチャの推定技術や、顔の形状・テクスチャの数値モデル化技術に関する研究を進めている。これらの技術によって、例えば、スマートフォンのカメラで撮影した顔写真1枚から詳細なCGの自動生成や、過去に撮影された写真・映像から故人をCGで復元、さらに、現存しない架空の人物のCGを生成することも可能となる。

本章では、Light Stage と機械学習を用いた本研究所の現在の取り組みとして、顔の単画像から形状を推定する技術について4.1節で述べる。次に、本研究所では、Light Stage と機械学習を用いた研究開発と並行して、Light Stage を用いた日本人の顔を対象としたデータベースの構築にも取り組んでいる。そこで4.2節では、Facial Action Coding System (FACS) [5]に基づいた顔データベース構築について述べる。

4.1 顔の単画像からの形状推定

本節では、Light Stage と機械学習を用いた単画像からの顔形状推定について、本研究の現状を紹介する。本研究では、Light Stage を用いて取得した顔の形状・テクスチャデータを ResNet [6]と3層の全結合層を組み合わせたネットワークで学習する。本研究で用いたネットワークの構成を図6に示す。本ネットワークでは、Light Stage を用いて取得した顔の形状・テクスチャデータに対し、様々なカメラ位置・照明環境をシミュレートしたレンダリング画像を入

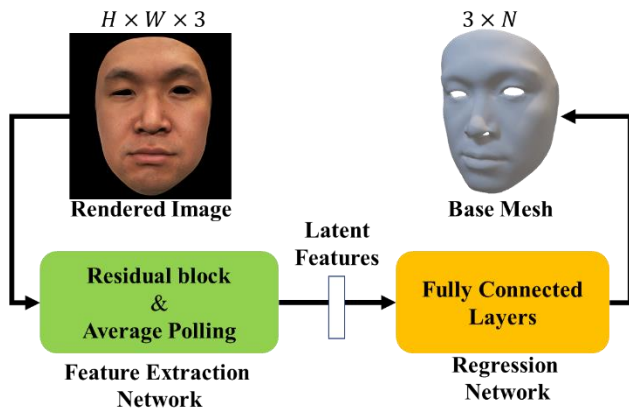


図6 ネットワークの構成

Figure 6 Structure of our network.

力, 顔の形状データを出力として学習に用いる. 本ネットワークにおける誤差関数を式(1)に示す.

$$Loss = \frac{1}{N} \sum_{i=1}^N w_i \cdot (p_i - t_i)^2 + \lambda \sum_{j=1}^M \phi_j^2 \quad (1)$$

ここで, N は顔形状の頂点数, w_i は各頂点の重み, p_i は推定した顔形状の各頂点位置, t_i はLight Stageで取得した顔形状の各頂点位置, λ は正則化係数, ϕ_j はネットワークのパラメータである. 本研究では, 目・鼻・口を構成する頂点の重み w_i を他の部位の頂点よりも大きくすることで, 人間の顔における特徴的な部位に注力した学習を行う. 本研究のネットワークを用いて単画像から推定した顔形状の一例を図7に示す. 図7に示すように, Light Stageで取得した顔形状と比較し, 実際の形状に近い推定が可能であることを確認した. 今後の取り組みとして, 顔の単画像から図5で示したテクスチャの推定に着手していく.

4.2 FACSに基づいた顔データベース構築

本節では, FACSに基づいた顔形状・テクスチャのデータベース構築について述べる. FACSとは, 表情筋の解剖学的知見を基礎とし, 人間の顔の表情をAction Unit (AU)という動作単位の組み合わせで記述する方法である. つまり, 計測対象の代表的なAUをLight Stageで計測し, それらを組み合わせることで, 計測対象の様々な表情の復元が可能になると考え, 本研究所では, FACSに基づいた顔データベースの構築を進めている. 本データベースのサンプル例を図8に示す. 図8に示すよう, 本データベースでは, Light Stageで計測した対象の画像・形状・テクスチャのデータをAUごとに収録する. さらに, 本データベースをテンソル分解などの解析にかけ, 日本人の顔のパラメトリックモデルの獲得への取り組みも進めていく.

5. おわりに

他業種と同様に, 映像コンテンツの制作現場においても, AIを代表する新しいテクノロジーの波が押し寄せている. 今後も, 本研究所では, それらの新しいテクノロジーを積

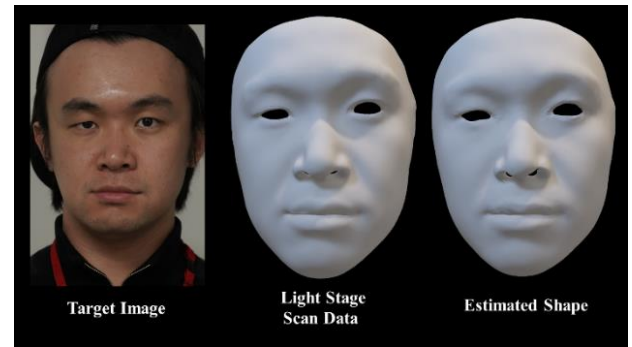


図7 形状推定結果

Figure 7 Shape estimation result with our network.

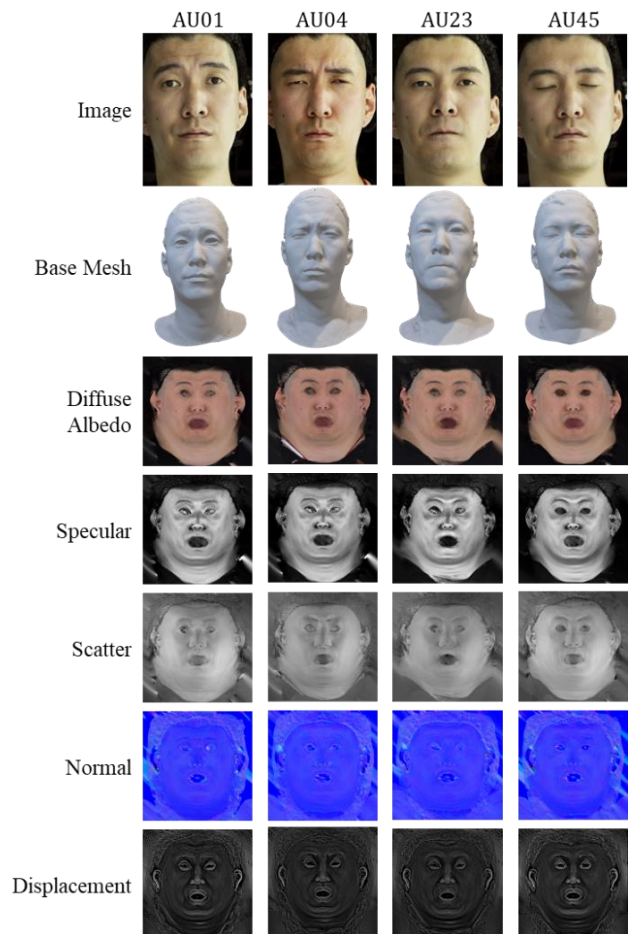


図8 FACSに基づいた顔データベースのサンプル例

Figure 8 Example of our facial database based on FACS.

極的に取り入れることで, 新しい制作手法を生み出していく. マルチフォーマットが求められるこれからのコンテンツには, 効率的でフレキシブルなワークフローが不可欠である. さらにもう1つ重要なポイントは, ストーリーテリングなど, 伝統的な映像コンテンツ制作で培われてきたノウハウの有無である. 新しい手法のその先にエンターテイメントコンテンツの醍醐味が存在し, 我々の持つノウハウが活かされる場が来るだろう. その時, 新しい表現が生まれるに違いないと確信している.

参考文献

- [1] Light Stage. <http://gl.ict.usc.edu/LightStages/>.
- [2] W. Ma, T. Hawkins, P. Peers, C. Chabert, M. Weiss and P. Debevec. Rapid Acquisition of Specular and Diffuse Normal Maps from Polarized Spherical Gradient Illumination. In Eurographics Symposium on Rendering, 2007.
- [3] A. Ghosh, T. Hawkins, P. Peers, S. Frederiksen and P. Debevec. Practical Modeling and Acquisition of Layered Facial Reflectance. In ACM SIGGRAPH Asia, 2008.
- [4] A. Ghosh, G. Fyffe, B. Tunwattanapong, J. Busch, X. Yu and P. Debevec. Multiview Face Capture using Polarized Spherical Gradient Illumination. In ACM SIGGRAPH Asia, 2011.
- [5] Facial Action Coding System. <https://www.paulekman.com/facial-action-coding-system/>
- [6] K. He, X. Zhang, S. Ren and J. Sun. Deep Residual Learning for Image Recognition. In Computer Vision and Pattern Recognition, 2016.