

ユーザ参加型エンタテインメント 『ダイブイントゥザムービー』

森島 繁生

早稲田大学 理工学術院先進理工学部

<和文抄録> 本稿では、視聴者自身が映画の登場人物として演技することができ、さらに登場人物の環境を再現することによって、ストーリーへの没入感を体験できる新しいエンタテインメント『ダイブイントゥザムービー』について述べる。視聴者そっくりのキャラクタモデルを短時間のうちに生成し、映画の本編にてこのキャラクタが実時間合成されて演技するという過去に類を見ないエンタテインメントの形態である。この技術を実現するためには、リアルタイムCG合成技術は当然必要であるが、個性を表現するためのさまざまな技術、個性をキャプチャするためのCV技術、そして音声信号処理技術も必要となる。

Audience Participation Type Entertainment “Dive Into the Movie”

Shigeo MORISHIMA

Graduate School of Advanced Science and Engineering, Waseda University

<Abstract> In this paper, we describe about a new entertainment system called “Dive Into the Movie” in which all audience can participate in the story as actors and actresses. After character models very close to the audience are automatically generated in a short time, real-time character animation can be generated in a main story. This is the world first entertainment system. To realize this system, several techniques to express personal characteristics are necessary to be solved. Computer vision and speech signal processing technique are also considered to capture personal features automatically as well as CG technique.

1. はじめに

映画の登場人物に扮して、宇宙を駆け巡る体験をしてみたい。あの有名な俳優と映画の中で共演してみたい。そのような希望は誰もが一度は抱いたことがあると思う。本稿で述べるダイブイントゥザムービーはそのような希望を叶えてくれる全く新しいタイプのエンタテインメントである。

このシステムの1つの具体的な実現例は、著者らのチームが中心となって2005年に具現化した愛知万博、三井・東芝館のフューチャーキャストシステムである。シアターの定員である240名全員が映画

に出演できるアトラクションシステムとして具現化され、期間中に163万人の観客がこれを体験し、人気のパビリオンとなった。

本システムは、実際に来場者自身を映像の中に登場させ、CG化された自分の分身が映像中で出演者として演技することで、ストーリーへの没入感を増大させている点に特徴がある。

また、上映開始までの約10分間のプレショー時間内に、来場者の全員の顔（愛・地球博では240人）をCGモデル化し、映像を作り上げなければならないため、顔の正面画像・立体形状の入力からCG顔モデ

ルの生成、映像の上映までの一連のプロセスを全自動で実行している点も特徴である。

来場者はまず、少人数のグループに分かれ、プレシヨールームに誘導される。各プレシヨールームには3次元スキャナが配置されており、1台のスキャナで4人分の顔の3次元スキャニングを行い、来場者全員の正面顔画像と顔立体形状が計測される。

次に、正面顔画像から画像処理を施すことにより、顔の特徴的な部位(目、眉、鼻、口、顔輪郭)の特徴点を抽出し、同時にストーリーの配役を決定する際に必要な情報である年齢・性別情報の推定を行う。その後、抽出された特徴点に対して標準的な顔ワイヤフレームモデルを自動整合する。このモデルにテクスチャマッピングを施すことで個人のCG顔モデルが生成される。

生成された顔モデルを実時間変形し、表情合成と陰影付けを施して、ストーリー映像とフレーム単位で同期・融合させることで、来場者そっくりの登場人物が、台詞を喋り、時には感情をあらわにしてス

トーリーの中で演技するシーンを目の当たりにすることができる。

ストーリー映像の上映には、当然信頼性の高いマシンが要求される。レンダリングクラスタシステムを用いて、システムダウンの無い、安定した上映が可能となった。6ヶ月間の愛・地球博開催中に163万人の観客が映画の登場人物となる体験をしたがシステム障害は報告されなかった。

またシステムの性能を検証するために、CG化された来場者の顔モデルが上映に耐えうるか検証したところ、平均93.5%の確率でシアターに出演が可能であるという結果が得られた。

本システムは、実時間処理によってモデル合成とレンダリング処理を行うゲーム的な要素と大勢の観客に演技を披露できる演劇的な要素、そして感動をよぶストーリーを大勢の観客に映像として同時に提示できる映画の要素、これら全ての特徴を兼ね備えた全く新しいエンタテインメントシステムとして位置づけできる。

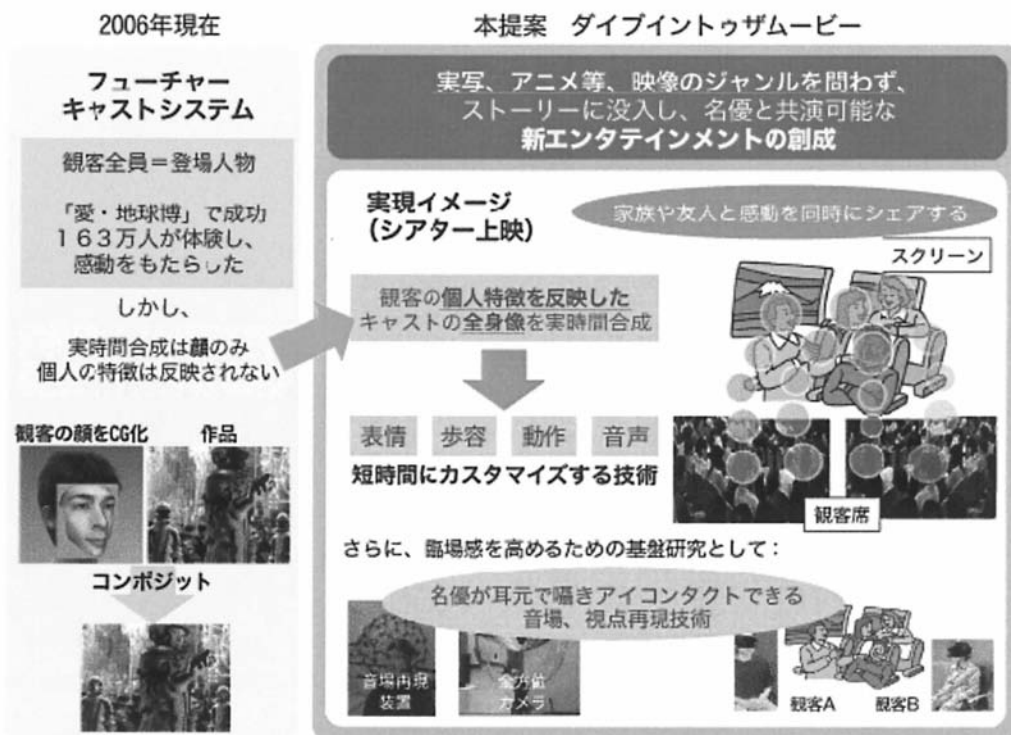


図1 ダイブイントゥザムービープロジェクト概要

2. ダイブイントゥザムービーとは

新映像技術ダイブイントゥザムービーとは過去・現在の映像作品を問わず、視聴者（参加者）らが映像に入り込み（Diving）、役者の視点で映画の世界に没入し、さらに、参加者間でインタラクティブに感動を共有できる全く新しいエンタテインメントである。すでに「愛・地球博（2005年日本国際博覧会 EXPO 2005 AICHI JAPAN）」、三井・東芝館にて「ダイブイントゥザムービー」の先駆けである「フューチャーキャストシステム」の展示を行った（URL：<http://www.expo2005.or.jp/jp/C0/C3/C3.7/C3.7.8/>）。これは、3次元スキャナで取り込んだ来場者一人一人の顔情報が瞬時にCGモデル化され、一度は誰もが思い描いた、自分が「物語の登場人物」として活躍するという夢を実現する画期的なシステムであった。「愛・地球博」では6ヶ月間の開催期間中、のべ163万人の来場者の顔画像をストーリーの登場人物として参加させ観客に大きな感動をもたらした。「愛・地球博」の人気パビリオンの1つとなったことから、「フューチャーキャストシステム」の有効性、さらには映画およびゲーム等の従来のデジタルコンテンツとは異なる新しいエンタテインメント分野の萌芽を感受することができた。この新興エンタテインメントを成長させアニメやゲームに続く、我が国発の高品質なデジタルコンテンツを、持続的かつ円滑に供給できる技術確立することが目的である。「フューチャーキャストシステム」では参加者の顔の表情や口形状をCGによってアニメーションさせ、あらかじめ顔画像をくり抜いた状態で作成されたオリジナル映像と合成させることで映像を生成した。一方、「ダイブイントゥザムービー」では過去・現在の作品を問わず、役者全身の振る舞い・セリフと参加者の振る舞い・セリフを挿げ替えることで、役者の視点の映像・音場を再現し、高い没入環境の構築をめざす。さらに、参加者どうしの体験や感動をリアルタイムで共有するシステムの構築をめざし、これらのエンタテインメント基盤技術とそのプロトタイプを研究開発することを目標とする。（DIM URL：<http://www.diveintothemovie.net/>）

3. フューチャーキャストシステム

ここでは、まずダイブイントゥザムービーの前身となる、フューチャーキャストについて述べる。従来、映像アトラクションは、既作成の映像やプログラムに従って一方的にストーリーが進行するのが主体である。フューチャーキャストシステムは、実際に来場者を映像中に登場させ、さらに映像中出演者として演技させることで、観客のストーリーへの没入感を増大させている点に特徴がある。いわ

ば、物語の中における自分の姿を、観客の側から客観的に見つめることが可能で、さらには自分自身の姿のみならず、仲間や家族の姿をストーリーの中に発見することで、感動を分かち合える全く新しい発想のアトラクションである。いわば、映画と演劇のハイブリッドという位置づけである。またリアルタイムの映像処理やCG合成はゲームでは常識であるが、他人にその勇姿を披露するという意味でアーケードゲームに近いこのシステムでは、スキルを要求せず、筋書きが決められた映像中の世界で、誰しもが容易にその勇姿を他人と共に共有できる点に大きな特徴がある。

観客はまず20人毎にプレシヨールームに誘導される。ここでまず、顔の3次元スキャニングを行って顔の立体形状が計測される。この立体形状は、鼻の高さや頬の膨らみなど個人の特徴を反映するものであり、本人そっくりのCGキャラクターを生成する際に重要な情報となる。また同時に撮影される正面画像は、皮膚の色や目の色、男女の特徴、年齢特徴等の個人情報表現するものとして重要である。

プレシヨールームで来場者がストーリーの説明を聞いている間に、本人のCGポリゴンモデルが自動的に計算され、本番に向けて準備が計算機内で実行される。また男女判定および年齢推定が自動的に行われてラベリングされ、配役を決定する際の基準として利用される。

メインシアターがオープンすると、プレシヨールームから、メインシアターに全員が移動し、来場者は階段状の観客席に着席すると、まもなくメインショーが開始される。ここで自分とそっくりのCG化された登場人物が活躍するストーリー映像を体験することになる。あらかじめ緻密に制作されたストーリー映像とCG合成された登場人物の映像がコンポジットされ、スクリーンに映し出される。登場人物の映像はリアルタイムで合成されており、キャストシナリオデータにしたがって演技する。また環境の変化も環境シナリオデータに基づいてタイムリーに映像に反映されるため、ストーリー映像に融合した違和感のない人物映像が合成される。観客そっくりの登場人物は、時にセリフを喋り、時には感情を露わにして、ストーリー映像の中で演技する。

最初20人ずつであった200インチスクリーンのシアターは、途中で左右の仕切りが取り除かれ80人シアターとなり、スクリーンサイズが横長の800インチとなって、さらに物語は展開してゆく。フィナーレでは、正面のスクリーンが下降して空間が広

がり、3つの80人シアターが地球の模型を取り囲んで合体する。美しい音楽と共に雰囲気が盛り上がり、240人の観客の感動と共にエンディングを迎える。

4. システムを支える映像技術

フューチャーキャストシステムは、視聴者参加型のエンタテインメントシステムであるが、大人数の参加を実現可能とするためにオペレータの介入を全く必要とせず、全ての処理が全自動で実行される点が特徴である。

4.1 顔の3次元スキャニング

半円上に配置された7台のデジタルカメラで構成される3次元スキャナを利用し、映像処理によって顔の3次元計測を実施する。3次元レンジスキャナは7台の民生用デジタルカメラで構成される。中央の1台のカメラがテクスチャ撮影用であり、左右対称の6台が奥行き計算用の画像を取得する。

また正面のカメラで撮影された顔の画像データは、1画素ごとに奥行きデータとの対応が取られる。この各画素位置に配置された奥行きのデータおよび色のデータは、次の個人CGモデル生成プロセスに受け渡される。

3次元レンジスキャナでは、6枚のパターン投影画像と6枚のシルエット画像から奥行き情報を再構成する。

4.2 個人CGモデル生成

奥行きデータから個人顔モデルを生成するために、標準となる顔ワイヤフレームを正面顔画像に整合する。この際、画像処理によって、目、眉、鼻、口、顔輪郭などの特徴点が自動計算される。奥行き情報は対応する画素位置から決定される。特徴点以外の顔ワイヤフレームの格子点については、補間ルールを適用して位置を決定する。これにより対応点に関してはそのままの位置、対応点以外の点は滑らかに補間可能である。しかし、目や眉、口の輪郭線に関しては、わずか数pixelの整合誤りでもCG顔モデルにしたときの印象が劣化する。このため、この輪郭線上の点に関しては特徴点自動抽出結果が正しいものと仮定し、3次元Spline曲線によって特徴点間を曲線近似し、標準顔ワイヤフレームモデルと対応する点を決定する。正面顔画像に対して顔ワイヤフレームを自動整合した結果を図2に示す。

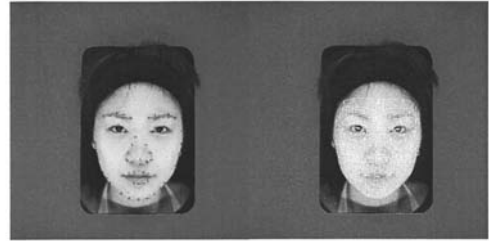


図2 特徴点抽出と自動整合結果

これと同時に、正面画像の処理によって、男女判定および年齢推定が行われる。また虹彩の色を自動推定し、眼球のテクスチャを選択する。茶色系2種類、青色系2種類の計4種類のうちから選択する。

立体形状計算用に一人の処理に1台のPCを割り当てることで、個人のCGポリゴンモデル生成までに約3分を要する。

4.3 ブレンドシェープによる表情合成

怒りや悲しみの表情、発話時の口形状など任意の表情の表現を行うために、基本となる表情パターンを予め基本形状(ターゲットシェープ)として準備しておき、この基本形状のブレンドによって目的となる表情を表現する。本稿では、34種類のターゲットシェープを用意した(図3)。これらの形状は顔の標準ワイヤフレームの格子点の移動量を標準値として定義されている。来場者によって顔のアスペクト比を算出し、各格子点の移動量を個人適応して用いる。図4に表情合成の一例を示す。

口の形状と共に、口内部の歯の位置も制御される。今回のアプリケーションでは、ストーリーは予め固定されているので、任意のシナリオに対応する必要はないが、インタラクティブなアプリケーションにも対応可能なシステム構成になっている。



図3 ターゲットシェープの一例



図4 表情合成の一例

4.4 リアルタイム表情合成

今回の作品 *Grand Odyssey* というストーリーに同期して、各シーン毎の照明環境をはじめとするリアルタイムレンダリングに必要な環境パラメータは時系列の環境シナリオデータとして、予め映像作品作成時にクリエイターによって保存されている。また各映像フレームで、セリフに基づく口の動きのパラメータおよび表情のパラメータも時系列のキャストシナリオデータとして保存されている。この口形状パラメータおよび表情パラメータは、先のターゲットシェープのそれぞれをいかなる比率で混合するかというブレンド率のパラメータによって定義している。よってメインストーリーが開始されると、このレンダリングのための環境シナリオデータと、基本表情ブレンド率および顔の位置と向きの時系列であるキャストシナリオデータに基づいて、毎フレーム同期して、表情変形およびレンダリングが実時間で実行される。

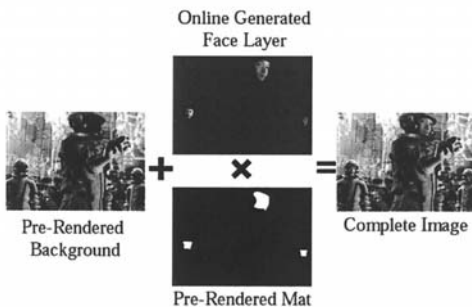


図5 レイヤ合成の一例

5. レンダリングクラスタシステム

合計 12 台の各プロジェクタにつき、スクリーンへの映像送出用 PC (Projection PC : PPC) が 1 台、IGPC が 8 台で構成されている (図 9)。PPC スクリーンへの描画を行う。またタイムコードに同期した動作が可能であり、全てのスクリーンは同期している。IGPC と PPC はギガビットネットワークで接続されている。

IGPC は、Pentium 4 3.2 GHz、GeForce 5950 Ultra、512MB RAM で、ごく一般的な PC で構成されている。PPC もほぼ同等のシステム構成だが、タイムコードへの対応等 IGPC とは若干仕様が異なる。上映される画像は、各 IGPC により生成され、PPC によりタイムコードに同期して投影される。1 フレームの画像は、1 台の IGPC がレンダリングから合成までを担当する。

PPC は、制御システムからコントロールコマンド、タイムコードを受け取り、画像生成して IGPC に送出する。来場者毎に変更される顔モデルは、上映前に PPC によりファイルサーバから収集され、各 IGPC へとデータ転送され、フレーム画像として合成される。

本システムでは、限られた期間でのシステムの構築と映像製作を同時進行で進める必要があったため、システムへのパフォーマンス要求は設計時には未知数であり、また設計時には、実装時のマシンスペックを見積もることは困難であった。さらに博覧会の開催期間中はシステムダウンが許されないため、より信頼性のある上映システムを構築する必要があった。そこでスケラビリティのあるクラスタシステムを採用した。また、レンダリング機能とプロジェクタへの送出機能とを分離することで、システムの信頼性を向上させることが可能となった。レンダリング用 PC は冗長をもたせ、8 台のうち 5 台の PC が動作していれば上映を続けることが可能である。

6. システム評価

自動生成された来場者の CG 顔モデルが上映に耐えるかどうか、表情合成をしたときに明らかな顔形状の破綻が見られないかどうか、上映前に必ずアテンダントによる確認作業が行われる。ここで、アテンダントの主観にて CG 顔モデルに破綻が見られないかが目視により確認され、シアターへの登場可否の判断が下される。このとき登場が認められない場合には、デフォルトデータに差し替えられ、上映が行われるようにしている。

2005年9月1日から愛・地球博の最終日である9月25日までの、日毎のアテンダント確認結果を集計し、出演確率（CG顔モデルの生成成功確率）と25日間の平均の出演確率を求めた。評価対象は、25日間902回の上映分となる21万6480人分のCG顔モデルである。結果を図6に示す。図中横軸は上映日を表しており、縦軸は出演確率を表している。25日間で平均93.5%、最高94.5%、最低でも92.5%の確率で来場者のCG顔モデルを映画の中に登場させることができるという結果が得られた。ただし、この確率は、本人が映像中の自分自身を認識できるという意味とは異なる。

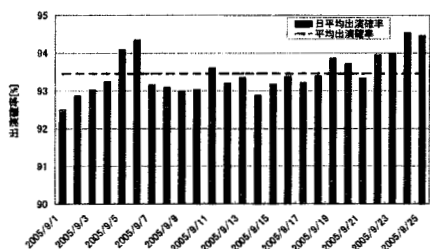


図6 9月入場者の出演確率

5. おわりに

2005年9月25日、愛・地球博は幕を閉じた。三井・東芝館では1度のシステム障害もなく、無事にその任務を終了した。連日待ち時間は2時間以上、長い時で4時間以上の行列ができるほど人気を博した。今でも上映が終わって外に飛び出してくる子ども達の笑顔が印象深く思い出される。技術の押し付けでなく、技術を駆使して人に感動を与えるもの。それがエンタテインメントの本質であることを実感した。一時的なものとはいえ、ヒーローやヒロインになった子ども達、若返って戦った老人。老若男女を通じて楽しめ、その感動を皆で分かち合える、新しいエンタテインメントの可能性を感じ取ることができた。そういう意味で、愛・地球博は貴重な実験場だったと言える。

これは、あくまでもダイブイントゥザムービーの第一歩にすぎない。これからの課題は、如何に本人そっくりのキャストを作り上げるかというテーマ、そして役者の視点での音響および映像を体験できるシステムの開発にある。全身モデル、歩容、表情、音声、頭髮、衣服など、カスタマイズの要素は多々ある。しかし、いかにして、観客に負担をかけずにこれらの個人性情報を短時間でモデル化してゆくか

が、これからの研究課題である。

今後の解決すべき研究課題を列挙する。

1) 表情筋モデルのカスタマイズ

ブレンドシェープアニメーションは、クリエイターの作品制作の効率化やリアルタイム処理に最適だが、形状以外の個性表現は難しい。そこで表情筋に基づく表情合成手法を導入し、表情筋のカスタマイズ機能によって、個人特徴を表現する手法に着手している。また、少ないビデオ映像から顔特徴変化をトラッキングし、最適な表情筋配置を推定する手法について検討している。また皮膚の個人的な質感表現についても検討している。

2) 歩行動作アニメーション

歩行の際の個人特徴を反映したアニメーションを実現するため歩行映像による個人認証技術を導入する。3次元再現された歩行動作の投影画像が、カメラ入力された画像と同一の特徴を表すように、データベースから個人特徴としての基本動作を選定する。

3) 台詞音声のカスタマイズ

声優の声に成り代わり、個人の声質、韻律情報をコピーした搭乗人物の声を合成する。特に本人に近い声を判定する客観評価尺度について検討を進めている。従来の大規模データベースに基づく素片音声合成とは異なり、少ないサンプルから本人の台詞音声を再現する。

4) インタラクティブシーンの実現

次期作品では、観客自身のコントロールによりキャラクターをリアルタイム制御して敵と戦うシーンの実現を試みる。動作合成研究の集大成によって実現が可能となる。

5) 役者地点空間再現システム

役者の地点から見える周囲の映像や、共演者の表情、さらに3次元音場を忠実に再現するシステムの構築を進めている。実環境の再現もめざす空間キャプチャシステムの開発も進めている。

謝辞

本研究は、文部科学省振興調整費プロジェクト「新映像技術ダイブイントゥザムービーの研究」として進められている。