

来場者参加型ムービーシアター -Future Cast System-

An audience-driven movie theater -Future Cast System-

前島 謙宣† 上村 周平‡ 森島 繁生†
Akinobu Maejima Shuhei Wemler Shigeo Morishima

1. まえがき

自分自身が映画の登場人物となり憧れのムービースターと競演してみたい、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。本稿で紹介するフューチャーキャストシステムは、2005年日本国際博覧会(愛・地球博)三井・東芝館において世界に先駆けて実装された全く新しいタイプのエンタテインメントシステムである。三井・東芝館では、シアターの定員である240名全員が映画の中に出演可能なアトラクションシステムとして具現化している。

フューチャーキャストシステムの特徴は、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である。また、来場者の入場から上映開始までの約10分のプレショー時間の中で、240名もの来場者の顔をCG化し、映像を作り上げるため、来場者の顔の入力からCG顔の生成、映像の上映までの一連の過程を全自動で実行している点も大きな特徴であると言える。

ストーリー映像の上映には、時分割のレンダリングクラスタシステムを新たに開発した。これにより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、大きな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評価するために、CG化された来場者の顔モデルが、上映に耐えるかどうか検証したところ、平均93.5%の確率で映像に出演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、大勢の観客に同時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そろ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ることができる。

2. フューチャーキャストシステム

従来の映像アトラクションは、既に作成された映像やプログラムに従って一方的にストーリーが進行していくものが多い。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、さらには出演者として演技させることにより、ストーリーへの没入感を増大させている。つまり、ストーリーの中における自分の姿を、観客

の中の視点から客観的に見つめることができ、さらには自分自身の姿のみならず、家族や友人の姿をストーリー映像で発見することで、感動を分かち合うことができる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は、アーケードゲームに近いこのシステムでは、特別なスキルは必要とせず、既に決められた物語の世界で、誰もが容易にその体験を他人と共有することができる点に特徴がある。

三井・東芝館では、劇場のキャパシティが240名であったが、1つのスクリーン上に240名全員が一度に登場した場合、その映像の中で自分の姿を発見することは非常に困難である。したがって、来場者はまず20人ずつ12グループに分割され、それぞれのグループが独立のスクリーンで、20人の登場人物によるストーリー映像を観ることとなる。

20人毎に分割された来場者は、まずプレショールームに誘導される。プレショールームには、各5台の3次元スキャナが配置されており、1台のスキャナで4人が交代で顔の3次元スキャニングを行い、来場者の正面顔画像と顔の立体形状が計測される。正面画像は、皮膚の色や目の色、男女の特徴、年齢情報等の個人情報表現するのに重要である。次に、正面顔画像から画像処理により、顔の特徴的な部位(目、鼻、口、眉、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ために必要な情報である年齢・性別情報の推定を行う。次に、抽出された特徴点に対して、あらかじめ用意した標準的な3次元顔ポリゴンモデルを変形することで自動的に整合する。このモデルに対して正面顔画像をテクスチャマッピングすることで、個人のCG顔モデルが生成される。また、男女判定および年齢性別情報の推定結果から、登場人物の配役が自動で決定される。

これらの処理は、プレショ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完全自動で行われる。上映の準備が終わり、メインシアターがオープンするとプレショールームから、メインシアターに全員が移動し、メインショーである映像が開始される。ここで遂に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となる。

上映中は、個人のCG顔モデルに対して、映像の時系列に沿った表情合成をリアルタイムで行うことにより、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演技している映像を目の当たりにすることができる。また、映像のシーンごとに定義されたライティング環境を顔モデルに対して反映することによりストーリー映像にマッチした顔モデルを得ることが可能である。

ストーリーが展開すると、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

† 早稲田大学理工学部応用物理学科

‡ シリコンスタジオ株式会社

取り除かれ 80 人のシアターとなり、スクリーンサイズが横長の 800 インチとなり、さらに物語が展開する。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。ここで初めて 240 人の観客が一同に介して、その存在を確認することができ、ショーはエンディングを迎える。

3. 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するためには、顔の入力から映像の生成までの一連の工程をオペレータの介入なく全自動で行う必要がある。以降では、フューチャーキャストを支えるキーとなる映像技術について説明する。

3.1 顔の 3 次元スキャンング

半円状に配置された 2 台のスライドプロジェクタと 7 台のデジタルカメラから構成される 3 次元スキャナ[1]を用いて、画像処理により顔の 3 次元形状を計測する。3 次元スキャナのカメラ構成を図 1 に示す。図 1 における、中央の 1 台のカメラが正面顔画像撮影用のカメラであり、左右対称に配置された 6 台のカメラが奥行き計算用の画像を撮影する。スライドプロジェクタから顔にパターン光を投影した状態、および蛍光灯が点灯した状態の 2 通りの状態にて画像が撮影される。3 次元スキャナは、6 枚のパターン投影画像(図 2 上)と 6 枚のシルエット画像(図 2 下)から奥行き情報を再構成する。再構成された顔の立体形状を図 3 に示す。

なお処理時間の制約から、画像サイズは現時点で 600 × 800 ピクセルに限定されており、奥行き情報の精度に限界があるが、現状のスクリーンサイズであれば十分に鑑賞に堪える精度であるといえる。プロセッサの高速化により処理時間の短縮が図れればさらに画像解像度を高めることができ、奥行き精度を向上させることが可能である。

3 次元スキャンングを行ってから、約 3 分程度で顔の立体形状が生成される。また正面カメラで撮影された、正面顔画像は、1 ピクセル毎に置奥行きデータとの対応が取られている。この奥行きデータと正面顔画像は、次の個人顔 CG モデル生成プロセスに渡され、CG 顔モデルが生成される。

3.2 個人顔 CG モデルの生成

3.1 節で計測された正面顔画像と顔の奥行き情報を用いて個人の顔に忠実な個人顔 CG モデルを生成する。まず、画像処理により顔の特徴的な部位(目、眉、鼻、口、顔輪郭)の特徴点を 89 点抽出する(図 4 左)。このとき、同時に正面顔画像から年齢・性別情報も推定する[2]。年齢情報は、10 歳以下、10 ~ 20 歳、20 ~ 40 歳、40 ~ 60 歳、60 歳以上のいずれかであると推定される。ここで得られた年齢情報は、来場者の配役決定の際に用いられる。抽出された顔の特徴点に対して、標準的な 3 次元顔ポリゴンモデルを自動整合する。まず、あらかじめ顔の特徴点と標

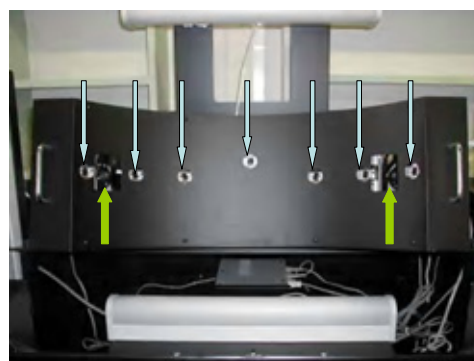


図 1. 3 次元スキャナのカメラ構成
(スライドプロジェクタ：緑矢印、
デジタルカメラ：青矢印)

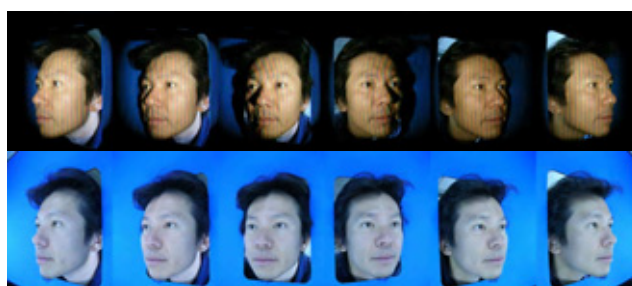


図 2. 3 次元スキャナで撮影される画像
(上：パターン投影画像、下：シルエット画像)

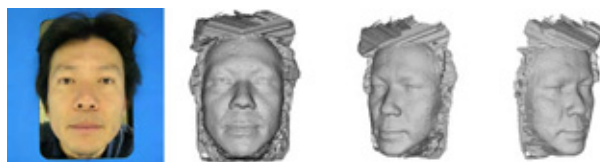


図 3. 正面顔画像と再構成された顔の立体形状



図 4. 特徴点抽出と自動整合結果

準顔ポリゴンモデルの頂点との対応づけを行い、対応点を定めておく。整合には Radial Basis Functions Transformation(RBFT)を用いた[3]。

これにより対応点に関しては、そのまま直接的に整合し、対応点以外の点に関しても滑らかに整合を行うことが可能である。しかしながら、目や眉、に関しては、

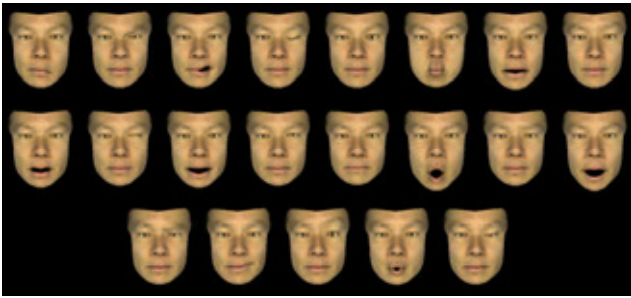


図 5. ターゲットシェイプの一例

わずかに数ピクセルの整合誤りでも最終的に CG 顔モデルにしたときの印象が異なる。

このため、このような輪郭線に関しては特徴点抽出結果が正しいものと仮定し、3次元スプライン曲線によって特徴点間を曲線近似し、標準顔ポリゴンモデルと対応する点を曲線上から選択し、整合の際の対応点とすることで、ポリゴンモデル上の頂点と必ず一致するように配慮した。正面顔画像に対して標準顔ポリゴンモデルを自動整合した結果を図 4 右に示す。正面顔画像に対して整合されたポリゴンモデルの各頂点について、その画素に対応する 3次元の立体形状を取得する。これにより個人の顔形状に忠実な顔モデルが生成可能となる。また、これと同時に正面顔画像から虹彩の色を自動推定し、眼球用のテクスチャを選択する。本システムでは、茶系 2種類、青系 2種類の計 4種類のうち、正面画像の虹彩色に最も近い一つを選択するようにしている。個人 CG 顔モデルの生成は、顔の立体形状計算用の PC を一人につき台割り当てた場合、個人顔 CG モデルの生成に約 5 秒程度で処理が実行される。

3.3 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を表現するために、基本となる表情パターンをあらかじめ基本形状(ターゲットシェイプ)として用意し、この基本形状の混合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イプ(図 5)を用意した。ターゲットシェイプは、標準顔ポリゴンモデルの格子点の移動量を標準値として定義される。

表情合成は、来場者によって顔の横幅や縦幅が異なることから、個人の顔の縦比、横比を算出し、その大きさによって移動量を正規化することで、個人向けにカスタマイズして用いる。表情合成の結果を図 6 に示す。

また、発話の際には口の形状の変化とともに、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、ストーリーが固定されているため、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構成となっている。

3.4 リアルタイムキャラクタ演出

今回の『スペースチャイルドアドベンチャー・グランオデッセイ』というストーリーに基づいて、シーン毎の台詞に基づく口形状の変化や表情を記述するためのパラ

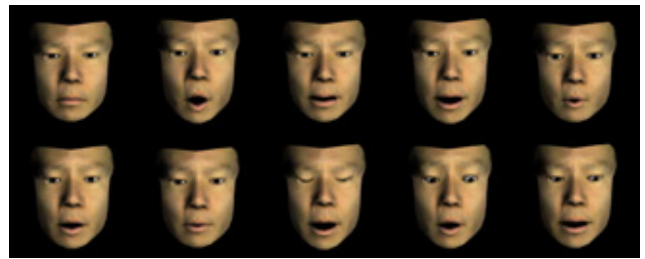


図 6. 表情合成結果の一例

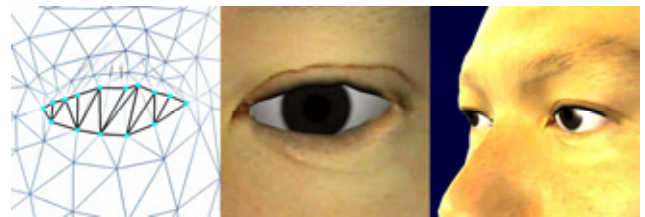


図 7. 眼球のモデリング

メータ、すなわちターゲットシェイプのブレンド率は、予め映像製作の段階で作成されている。このようなシーンの時系列に沿ったターゲットシェイプのブレンド率は、キャストシナリオデータとして保存されている。上映が開始されるとこの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合成が実行される。また、シーン毎の照明環境をはじめとするレンダリングに必要な環境パラメータ(以後、環境シナリオデータとよぶ)も同様に予め作成され、上映時には、環境シナリオデータに基づいた証明条件が顔モデルに反映され、レンダリングが実行される。

さらに、先の男女判別結果に基づいて、あらかじめ男女別の声優がそれぞれ発声し、記録されている音声ストリームが選択される。年齢推定の結果はソーティングされて配役の際に参考データとされる。よって配役はスキニングの順番に影響されない。

3.5 眼球のモデリング

眼球のモデリングは球もしくは半球にテクスチャマッピングを施すことで表現するのが一般的であるが、顔形状によっては眼球と瞼の間に隙間が生じる可能性があり、さらに瞬きなどの動作の場合、球体に沿って顔モデル上の瞼のポリゴンを制御しなければならないため、計算がより複雑となる。

そこで本稿では、眼球を球体として描画するのではなく、上瞼と下瞼とを結ぶ面として描画しつつも、シェーディングにより球体同様の陰影、光沢が再現できるように、上瞼および下瞼上の面の法線ベクトルを調整し、レンダリングする。また、角膜が滑らかであるように見せるために、眼球モデルにスペキュラマップを用いて、黒目は白目よりもシャープでかつ明るいスペキュラが生じ

るよう工夫をしている。眼球のモデリング結果を図7に示す。

4 リアルタイムレンダリング

上映時に差し替える必要がある映像は、登場人物となる来場者の顔だけであるため、予めシーン映像を、顔レイヤ、背景レイヤ、エフェクトレイヤの3つのレイヤに分け、このうち顔レイヤ以外については、予めソフトウェアレンダリングした画像をハードディスクに保管している。上映時には顔レイヤのみを新たにレンダリングし、顔レイヤをプリレンダリングされた画像に合成する。これにより、レンダリング時にかかる計算コストを軽減しつつ、描画可能な表現の幅を広げている。合成は、各レンダリング PC(これを ImageGenerator PC とよぶ)によって上映と平行して行われる。

レイヤ合成は、エフェクトレイヤの有無、エフェクトの種類によって4種類使用している。合成はグラフィックスハードウェア上で行われ、合成式はピクセルシェーダとして実装されている。このためレイヤ合成を高速に実行させることが可能となっている。図8にエフェクトレイヤが無い場合の合成例を示す。顔はヘルメットを被っていない状態でレンダリングされるため、通常ヘルメットに隠されて見えない部分(顎など)を切り取るためのマツレイヤ等を使用している。

図10に、20人シアターにて上映された映像の一例を、図11に80人シアターにて上映された映像の例をそれぞれ示す。

5 レンダリングクラスタシステム

レンダリングクラスタシステムの構成を図9に示す。プロジェクタ1台につき、スクリーンへの映像送出力用 PC(これを Projection PC とよぶ)が1台、Image Generator PC が8台で構成されている。Projection PC がプロジェクタに接続されており、スクリーンに上映を行っている。またタイムコードに同期した動作が可能であり、他のスクリーン(全部で12のスクリーン)と同期している。なお転送速度を考慮して、Image Generator PC と Projection PC はギガビットネットワークで接続されている。Image Generator PC は、Pentium4 3.2 GHz、GeForce5950 Ultra、512MB RAMで、最近ではごく一般的に家庭でも用いられる PC で構成されている。Projection PC も上述の PC とほぼ同等のシステム構成であるが、タイムコードへの対応等 Image Generator PC とは若干仕様が異なる。

来場者毎に変更される顔モデルを合成して上映される画像は、各 Image Generator PC により生成され、Projection PC 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 Image Generator PC がレンダリングから合成までを担当することで生成される。並列システムとしては時分割方式を採用している。

Projection PC 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、Image Generator PC により生成された画像を音声に同期させて送出する。

最新のハードウェアで注意深くチューニングを行えば、同等のシステムを1台の PC に実装することは可能であると思われる。しかしながら、本システムでは、シス

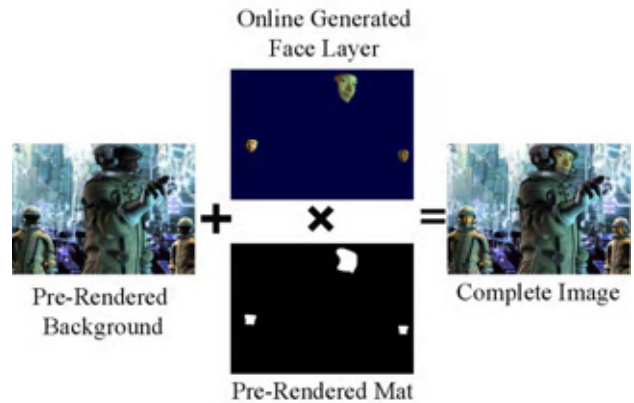


図8. レイヤ合成例

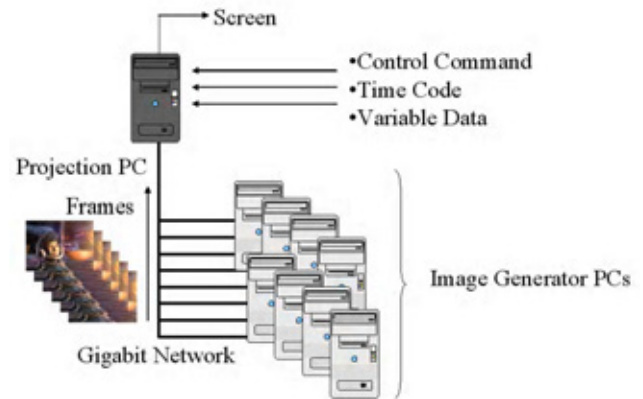


図9. レンダリングクラスタシステムの構成

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。

そこで本システムでは、システムパフォーマンスにスケラビリティのあるクラスタシステムを採用した。これによりレンダリングパフォーマンスによる映像コンテンツへの制限を大幅に減少させることができた。

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。加えてレンダリング用 PC は冗長性を持つように設計しており、8台のうち5台の PC が動作していれば、問題なく上映を続けることが可能である。

6 システム評価

フューチャーキャストシステムを評価するためには、映像中に自分の CG 顔が登場し、かつそれを自分として認識できたかどうか、また映像中に登場していた場合には、CG 顔の品質がいかなるものであったかを評価する必要がある。もちろん、来場者の全てに対して映像中の自分を認識できたかどうかアンケート調査を実施することが



図 10. 20 人シアターにて上映された映像の一例



図 11. 80 人シアターにて上映された映像の一例

理想である。しかしながらパビリオン運営上の観点からこれは困難であった。

また、愛・地球博の開催期間中は、個人情報保護の観点から、すべて取得した顔画像データおよび立体形状データは、閉館前に全て消去しており、リピータにも毎回新規データの取得、処理を実施している。このため、取得データによる CG 顔の品質評価もまた困難であった。

6.1 役者として出演できる確率

自動生成された来場者の顔モデルが上映に耐えうるかどうか、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。ここでは、ショーアテンダントの主観により CG 顔モデルに形状の破綻が見られないかが目視により確認され、出演の可否の判断が下される。このとき出演が認められない場合は、事前に用意されているデフォルトキャラクタに差し替えられ、上映が実行さ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の確認結果を集計し、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 人分の顔モデルである。結果を図 12 に示す。

図中横軸は、上映日を表し、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可能であるという結果が得られた。

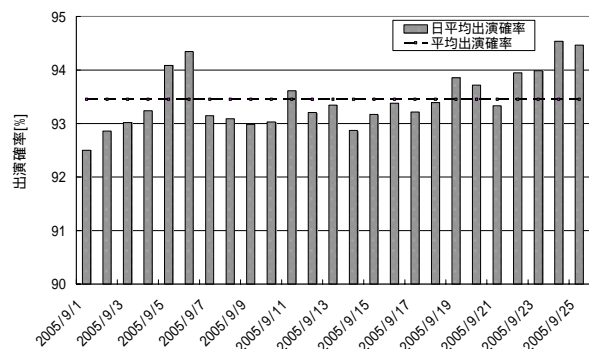


図 12. 9 月入場者の出演確率

6.2 表情合成の品質

『グランオデッセイ』の映像素材完成後、博覧会開幕までの数ヶ月間を利用して老若男女を含む 100 人分の顔形状データに対して、キャストシナリオデータを用いて、実際のシーン・キャラクタ毎の表情合成を行った。その後、表情合成により顔形状に破綻が生じていないかどうか、意図した表情が再現できているかどうか目視による検証を行った。

検証の結果、明らかな顔形状の破綻は見受けられず、また意図した表情が再現されていた。これは顔に応じた基本表情パターンの正規化が有効であったことを示している。さらに口形状についても意図した発話口形が合成されていた。

7 おわりに

フューチャーキャストシステムは、三井・東芝館において、2005年9月25日の愛・地球博の閉幕まで1度のシステム障害もなく無事にその任務を果たした。顔を3次元スキャニングするだけで、自分の分身となるCGキャラクターが予め製作された映画の中に出演し、その中で演技する点が本システムの大きな特徴である。結果として、観衆のストーリーへの没入感を増大させることが可能である。CG化された来場者の顔モデルの品質が上映に耐えうるかどうかを評価した結果、平均93.5%の確率で映画に出演が可能であるという結果が得られた。また、上映に信頼性の高いレンダリングクラスターシステムを採用したことにより、安定した上映が可能であった。

最後に、余談ではあるが、三井・東芝館では連日の待ち時間は2時間以上、最大で4時間以上の待ち行列ができるほどの人気を博した。今でも上映が終わって外に飛び出してくる子どもたちの笑顔を忘れることができない。技術の押し付けではなく、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。一時的であるとはいえ、ヒーローやヒロインになった子どもたち、若返って戦った高齢者、老若男女を通じて楽しめ、エキサイトし、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ることができた。愛・地球博は、そのような意味で貴重な実験場であったといえる。今後は、この経験を基に、さらなる感動をもたらす新しいエンタテインメントの創造に向けて研究を進めていく。

8 今後の展望

三井・東芝館では、プレショーの時間内にシアターキャパシティである240名の顔を処理しなければならないという処理時間の制約から、来場者の顔のみを処理の対象としていた。今後は髪の毛を含む頭部全体、さらには体や装飾品までCGモデル化を行うことで、より本人の個性を反映したCGキャラクターを生成し映像中に登場させたいと考えている。

また、表情やキャラクターの動作も映像製作時に決められた顔形状やパラメータに基づいて動作するため、個人の表情や振る舞いとは異なっている。表情やキャラクター動作についても個人向けにカスタマイズすることで、演技により個人性を持たせたいと考えている。

加えて音声に関しても、本作品では男女の声優の発声した音声を男女推定の結果からストリームを切り替えることで使用していたが、時折顔と声とのミスマッチが生じていた。今後は本人の声質で台詞を喋らせることができるよう研究を進めて行く予定である。

謝辞

システム開発を担当した、株式会社電通テック、株式会社ピノアズール、シリコンスタジオ株式会社に深く感謝の意を表します。

参考文献

- [1]K.Fujimura, Y.Matsumoto, and E.Tetsuichi, "Multi-camera 3D Modeling system to digitize human head and body", Three dimensional image capture and application iv, SPIE, Vol.4298, 2001.
- [2]L.Wiskott, J.Fellous, N.Kruger, and C.vonder Malsburg, "Face Recognition and gender determination", Proceedings of International Workshop on Automatic Face and Gesture Recognition, 1995.
- [3]N.Arad and D.Reisfeld, "Image warping using few anchorpoints and radial basis functions", Computer Graphics forum, Vol.14, 1995.
- [4] 森島繁生, "フューチャーキャストシステム『三井・東芝館』", 映像情報メディア学会誌, Vol.59, No.4, pp.522-524, 2005.