

3次元ビデオ

京都大学大学院情報学研究科 松山 隆司

1. 3次元ビデオとは

世界史を紐解くと、人類の文化・文明が新たな記録・伝達メディアの発明によって大きく進展したことが分かる。古くは文字や絵を記すための石板・パピルス・紙、15世紀のグーテンベルグの印刷装置による書籍・新聞、19世紀には画像を記録するための写真へとメディア世界が広がり、20世紀には電子技術によるラジオ・テレビ放送、テープレコーダ・ビデオによる記録・複製、そして最近ではデジタル・ネットワーク技術による WWW へと発展し、今ではだれもがマルチメディアを使った国際的情報発信・収集が可能となった。

一方、記録・伝達される内容(コンテンツ)としては、文字や図形・絵といった静的なものから、時間とともに動的に変化する1次元の音声・音楽、2次元の映画・ビデオを経て最近では CG による3次元アニメーションへと発展してきた。

では、21世紀における情報メディアにはこういった発展が期待され、それが文化・文明をどのように変えるのであろうか。その答えの1つとして考えられるのが3次元ビデオである。

3次元ビデオは、**CG による仮想的・人工的な3次元アニメーションではなく**、ダンスやスポーツをする人間、自然界の動物などの生の姿・形・色の時間的変化を3次元的にそのまま記録した実写映像で、実世界における対象の振り舞い・動作を余すところなく記録した実写3次元映像メディアである。

3次元ビデオは、書籍や新聞、ラジオやテレビが人類のありとあらゆる活動に与えた影響と同等あるいはそれ以上のものをもたらすのではないかと期待されている。具体的には、3次元ビデオを利用することによって、以下のような応用システムが実現できる。

- ・ 人間国宝やオリンピック選手の動作をそのまま記録再現できる身体技能・芸能デジタルアーカイブ(今しか撮れない卓越した身体動作を3次元的に完全に記録し、後世に伝える。)
- ・ 自分の身体動作と先生による手本とを3次元的に比較しながらトレーニングできるリハビリ、スポーツ練習システム(タイガー・ウッズと自分のゴルフスウィングを3次元的に比較する。)
- ・ 動物のありのままの生態を多角的に観察できる3次元

ビデオ DVD 図鑑(生徒が映像を基に自分の興味・関心に沿って観察を行う能動学習の実現。)

- ・ デジタルテレビ放送や広帯域インターネットをインフラとして使った3次元テレビ放送
- ・ 遠隔地の雰囲気そのまま伝えることができる高臨場感遠隔会議・講義・実験システム

これらの例や3次元ビデオという名前からイメージされるのは映像が飛び出して見える立体映像であり、3次元ビデオは立体ディスプレイ用の特殊なコンテンツだと思われるかもしれないが、通常の2次元のディスプレイを使った場合でも、3次元ビデオは映像の楽しみ方に革新的変化をもたらす。具体的には、3次元ビデオを使えば、視聴者が映像を鑑賞する位置や方向、視野をその場でインタラクティブに変えることができる。つまり、大相撲の放送が3次元ビデオを使って行われているとすると、チャンネルはすべて同じであるにも拘わらず、ある家ではかぶりつきから力士をズームアップした映像が、別の家庭では土俵全体の様子がテレビに映し出されているといったことができる。このように、3次元ビデオの持つ大きな特徴は、**視聴者がその場でインタラクティブに視点やズームを変えながら映像を楽しめる**ことにあり、映像の楽しみ方に大きな変革をもたらすものと考えられる。つまり、インタラクティブ映像は、映像の楽しみ方を、従来の受動的なものから能動的なものへと変える。

ここでは我々がこれまでに得た研究成果¹⁾を簡単に紹介する。(紙の上での説明や通常のビデオでは3次元ビデオを実感して頂くのはむずかしいと思われ、興味をお持ちの方は研究室に来ていただければデモをお見せします。)

2. 3次元ビデオの撮影

2.1 処理の概要

現在のシステムでは以下の方法で3次元ビデオの撮影・生成を行っている。

- (1) 撮影対象を取り囲むように配置された多数のカメラ(図1)を用いて、対象の多視点ビデオを撮影する(図2の最上段)。
- (2) 撮影されたビデオフレーム画像から対象のシルエットを抽出する(図2上から2段目)。
- (3) 各カメラの投影中心(図3のPA、PB)を中心としてシルエットを3次元空間に逆投影し視体積を求める。この処理をたとえて言うとな次のようになる。シルエットの部分だけ穴を開けた紙を撮像面に置いた状態で、

カメラの投影中心に電球を置く。その時にシルエットの穴を通して3次元空間に広がる光の束が視体積である。カメラによる撮影の原理から、3次元の撮影対象はこの視体積の中に完全に含まれる(図3左)。こうした視体積は各カメラ毎に得られるため、全ての視体積の重なった部分が対象の(粗い)3次元形状となる。そこで、全ての視体積の AND を取ることによって対象形状の3次元ボクセル表現(小さな立方体の集まりとして対象の3次元形状を表す)を求める(図3右)。

- (4) 視体積交差法で得られたボクセルデータ(図2上から3段目)の表面に小さな3角形の面を張り、対象の表面形状を表す3次元メッシュを求める(図2下から2段目および図9左)。
- (5) (4)で得られた3次元形状は、彫刻の荒削りのようなもので、形が角張っており、細かな凹凸が復元されていない(図9左)。そこで、3次元メッシュがゴムのような弾性を持っていると考え、各3角形の頂点の位置が真の対象表面に張り付くようにメッシュの変形を行う(図9右)。
- (6) 各3角形面上のテクスチャや色は多数のカメラによって写されているため、その面が最もよく写っている画像を選び、その画像上のテクスチャ・色を面に張る(図2最下段)。
- (7) 以上の処理で、1フレームの3次元ビデオが生成される。動画を作るには(1)~(6)の処理を繰り返せばよい。

以下では、各処理の技術的概要と処理結果を示す。

2.2 多視点ビデオ撮影システム

現在稼働中の実時間3次元ビデオ撮影システムは、10m四方の部屋に設置された25台のビデオカメラ(天井:12台、床:13台)と30台のPCを持つPCクラスターから構成されている(図1)。現在のシステムでは、12.5フレーム/秒(カメラのハードウェアの制限による)で25の異なる視点から同期の取れたビデオ(VGA画質)が撮影される。このシステムで撮影した舞妓さんの踊りの多視点ビデオを図4に示す。

カメラには首振り機能があり、移動する複数の対象をそれぞれ実時間で追跡して各対象の詳細な映像をズームアップ撮影することが可能である。(現在は、追跡・ズームアップ撮影機能はまだ完成していない。)

また、PCは高速の(1.25Gbit/sec)ネットワークで結ばれており、PCクラスターを1つの並列計算機として利用する

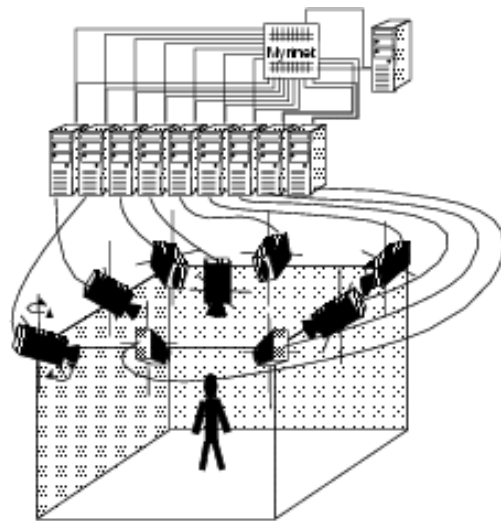


図1 多視点ビデオ撮影システム

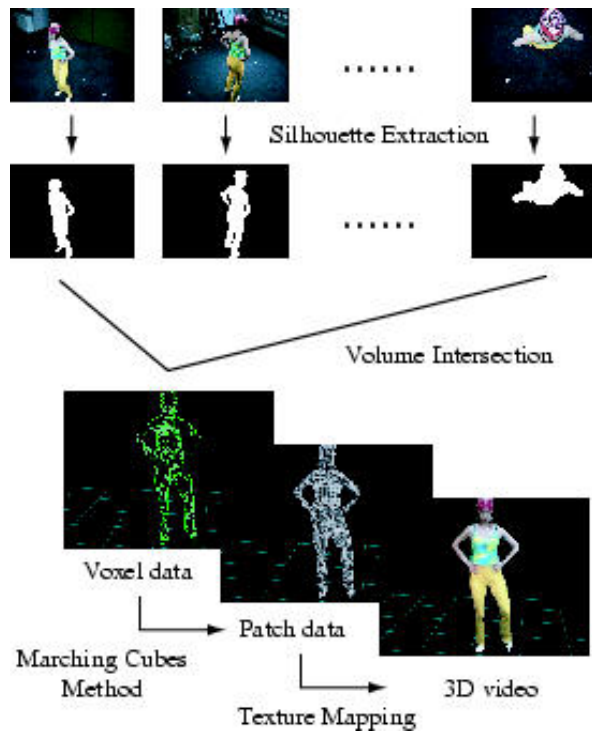


図2 3次元ビデオの生成過程

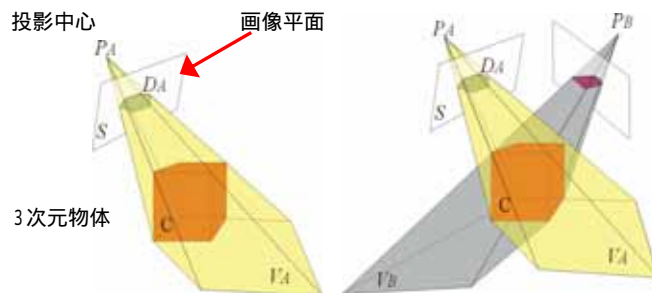


図3 視体積交差法による3次元形状復元の原理

ことによって人間の動作の3次元形状を実時間で復元することができる。

2.3 PC クラスタを用いた実時間並列3次元形状復元

通常の視体積交差法では、3次元の座標変換を膨大な数の点に対して行うことが必要で、実時間で3次元形状を復元するのはむずかしかった。これに対し、我々の方法では、以下のようにして高速処理を実現している。

- (1) まず、3次元空間を平行な平面群として表し(図5)、
- (2) 各撮影画像から抽出されたシルエットを平行平面群中の基準平面に投影する(図6左図の)。
- (3) 次に基準平面上のシルエットを他の平行平面群に投影する(図6左図の)。
- (4) それぞれの平面上で各撮影画像から得られた2次元シルエットの共通部分を求め、それを対象の断面形状とする(図6右図)。
- (5) 得られた断面形状が平行平面間の間隔分の厚みを持つと考え、それらを3次的に積み重ねて対象の3次元ボクセル形状を求める。

以上述べた平面間透視投影に基づく視体積交差法は、以下の利点を持つ。

- (a) 投影計算が平面から平面へのもののみであるため、シルエットの投影計算が高速に行える。
- (b) 特に、平面が平行である場合は、画像の2次元的なスケールと平行移動で投影計算が実現でき、より高速な計算ができる。
- (c) さらに、各平面上での断面形状の計算は他とは独立に行えるため、並列処理の効果が高い。

以上述べた方法に基づき、図1に示した PC クラスタを用いて平面間透視投影に基づく並列視体積交差法を実装した。図7にその並列処理方式を示す(図7の各列は1台のPCでの処理過程を示している。)

- (1) カメラを備えた PC に撮影コマンドを出し、多視点画像の同期撮影を行う。
- (2) 各 PC が背景差分により対象のシルエットを抽出し、それを基準平面に投影して基準シルエットを求める。
- (3) カメラを持たない PC も含め、全ての PC が全ての基準シルエットのコピーを持つように相互通信を行う。
- (4) 平行平面群をグループに分け、各グループをそれぞれ1台のPCに割り当てる。各PCは割り当てられた平行平面群を対象として基準シルエット群の平行平面への投影とその上でのシルエット交差計算を行い、対象の断面形状を求める。

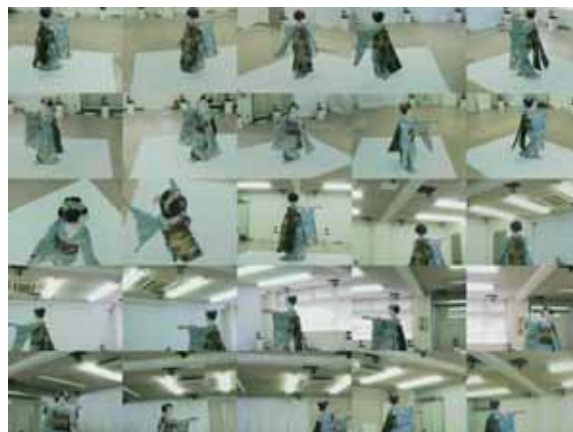


図4 舞妓さんの踊りを撮影した2.5視点ビデオ

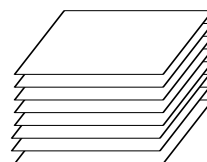


図5 3次元空間の平行平面群による表現

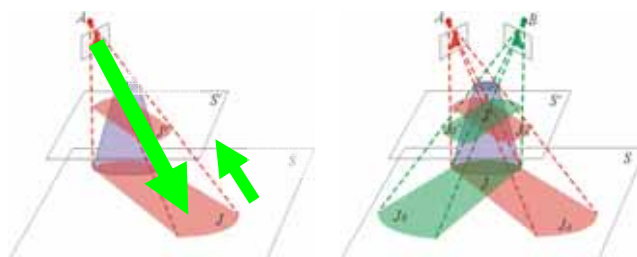


図6 平面間透視投影に基づく視体積交差法

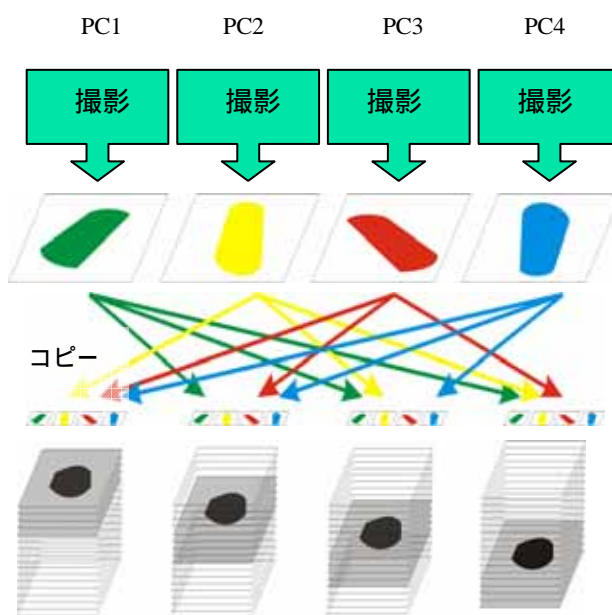


図7 並列視体積交差法

最新システムでは、さらなる高速化を図るため、

- (a)各 PC で実行される処理をパイプライン化する。
- (b)基準平面として、 $x - y$ 、 $y - z$ 、 $x - z$ の座標軸に平行な3つの面を設ける。
- (c)人物を取り囲む直方体を設け、その中において視体積交差法を適用する。

という処理を導入し、9台のカメラからのビデオ入力の場合、2cmボクセルでは毎秒30フレーム以上、1cmボクセルでも毎秒12フレーム程度の処理速度で人物の3次元形状をリアルタイム復元することが可能となっている。

2.4 弾性メッシュモデルを用いた3次元形状の高精度復元

視体積交差法では、対象の粗い形状しか得られず、凹部分が復元できないという本質的問題がある。この問題を解決し、対象の正確な3次元形状を求めるため、本研究では、まず視体積交差法で得られたボクセルデータを離散マーチングキューブ法²⁾によって三角形メッシュデータに変換し、このメッシュデータを初期形状としてメッシュの動的変形を行う弾性メッシュモデルを考案した。

本手法のアルゴリズムをまとめると以下になる。
 [step 1] 離散マーチングキューブ法によって、対象の3次元形状をボクセル表現から表面パッチ表現に変換し、弾性メッシュモデルの初期形状を得る。

[step 2] メッシュの変形を行う。

[step 2.1] 各頂点に働く力(後述)を計算する。

[step 2.2] 各頂点を力に沿って微量移動させる。

[step 2.3] 全ての頂点の移動が許容量以下なら終了。

そうでなければ step 2.1 へ。

弾性メッシュモデルでは、ステレオ法や視体積交差法では利用されていなかった対象表面形状の連続性や滑らかさといった幾何学的情報を利用することができ、正確な形状復元が可能である。

今回考案した弾性メッシュモデルでは、

(a) photometric consistency 制約: 各カメラから対象表面の三角形メッシュを見た場合、各撮影画像中のテクスチャが一致する。

(b) シルエット保持制約: 3次元メッシュを各カメラの撮像面に投影した場合、そのシルエットが観測された画像中のシルエットと一致する。

(c) 単純閉曲面制約: メッシュ表面はねじれない滑らかな曲面となる。

という3つの制約条件を満たしながら変形を行うように設計されており、かなり複雑な3次元形状でも正確に復元で

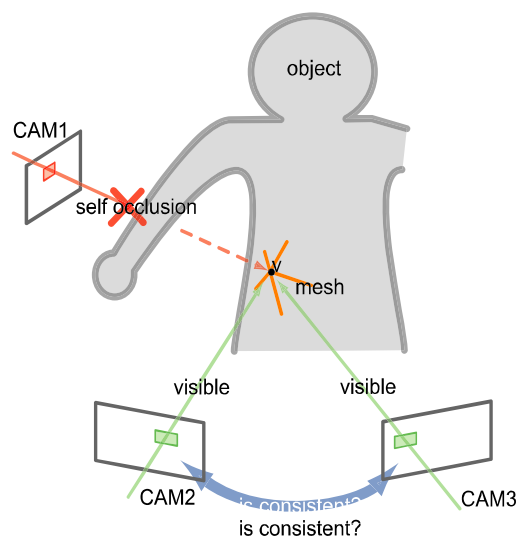


図8 photometric consistency 制約に基づいて生成される力

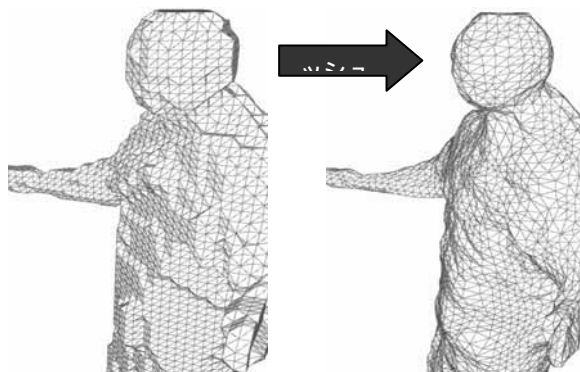


図9 弾性メッシュ変形による高精度3次元形状復元

きる。つまり、メッシュの各頂点に働く力は、上記の3つの制約を満たすような方向に頂点を移動させるように計算される。

図8は、photometric consistency 制約に基づいて生成される力を計算するアルゴリズムを示したものである。

- (1) 現在の3次元メッシュ中の各頂点に対して、その3次元座標から、頂点が写されているカメラを求める。図8では、CAM2とCAM3が求まる。
- (2) (1)で求めた各カメラに対して、頂点が写されている画像中の位置を計算し、画像中における頂点の像の付近の画像パターンを求める。
- (3) 各カメラで撮られた画像パターン間の類似度を計算し、類似度が大きくなるように頂点の3次元座標を移動させる力を生成する。

他の制約に関しても同様に、現在のメッシュ頂点をど

ちらに移動させば、制約が満たされるのかを計算し、その移動を実現する力を生成する。最後に、各制約に基づいて生成された力の重み付き平均によって最終的に頂点に働く力とする。

図9右は、上記の方法で求めた人間の高精度3次元形状で、同図左の視体積交差法だけの場合と比べ、格段に精度が向上していることが分かる。

現在開発を進めているより高度な弾性メッシュモデルでは、

(a)メッシュの各部分の硬さ・柔らかさを過去の運動状態から求め、それに基づいて弾性変形か剛体運動かを判別してメッシュ変形を制御する。

(b)入力される映像は、対象の動きを撮影したものであるため、入力映像から計算されたボクセルデータ系列に応じてメッシュを時間軸方向に変形させ、対象の動作に従ってメッシュが変形するようにする。

(c)人間が手を胴体に添えたり、離したりするような複雑な動作(3次元形状の大局的なトポロジーの変化)にも対応できるようなメッシュの動的変形を実現する。

といった機能拡張を行っており、舞妓さんのように長い振袖や帯がしなやかに揺れるような動作でもかなりの精度で3次元形状が復元できるようになっている。

2.5 高精度3次元ビデオレンダリング

以上の処理で対象の動きを表す3次元メッシュの時系列データが得られ、メッシュ中の各三角形パッチに、撮影されたビデオから抽出したテクスチャを貼り付けることによって、3次元ビデオが完成する。

最も単純なテクスチャマッピング法としては以下のものが考えられる。

[視点独立面ベース法]

- (1) 各三角形パッチに対して、その面の法線と最も近い視線を持つカメラを選び、
- (2) そのカメラによって撮影された画像上に三角形パッチを投影し、三角形パッチ上のテクスチャを求め、それを対象表面パッチに張付ける。

しかし、計算された3次元メッシュは対象の形状を完全に正確に表しているものではないため、この方法では、以下の問題が生じる。

(a)カメラ間の撮影映像の差異によるちらつき

異なったカメラで撮られた画像が対象表面の隣接した三角形パッチに張られ、その境界線上でテクスチャが不整合となり、ちらつきが生じる。

(b)隣接パッチ間でのテクスチャの不整合によるちらつき



図10 テクスチャマッピング法の定性的比較

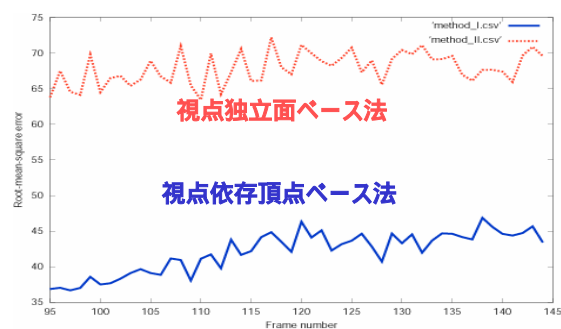


図11 テクスチャマッピング法の定量的比較

同じカメラで撮影された映像を基にしてテクスチャマッピングを行った場合でも、テクスチャはパッチ毎に独立して生成されるため、各パッチの継ぎ目において不整合が生じることがある。

これらの問題を解決し、滑らかなテクスチャマッピングを実現するために、映像を表示する際の視点位置を利用し、対象表面パッチの頂点色の補間によってパッチ上のテクスチャを生成する方法を考えた。

[視点依存頂点ベース法]

- (1) 対象を見るための視点位置・視線方向を決める。
- (2) 定められた視線方向と撮影に用いた各カメラの視線方向との一致度を計算する。
- (3) メッシュの各頂点について、頂点を各カメラで撮影された画像に投影し、各画像中での RGB 値を求め、(2)で求めた一致度に基づいた重みづけを用いて加重平均を計算し、その値を頂点の RGB 値とする。
- (4) パッチ上のテクスチャをその3頂点の RGB 値の線形補間によって生成する。

図10は、視点独立面ベース法、視点依存頂点ベース法によるテクスチャマッピングの結果を表したもので、後者の優位性がよく分かる。図11は、撮影画像と両テクスチャマッピング法で生成された画像との RGB 値の平均誤差(縦軸)を各フレーム毎(横軸)に計算したもので、このグラフからも提案手法の優位性が実証された。

3 3次元ビデオの編集・表示

従来のビデオと異なり、3次元ビデオは3次元の形や運動情報を持っているため、その編集は4次元空間(3次元座標軸 + 1次元時間軸)において行われる。すなわち、図12のような対象、背景、仮想カメラからなる3次元シーンを設け、その中で対象の移動・拡大縮小・コピー、背景の拡大縮小・移動、カメラの移動・ズームなどを行い、仮想カメラで撮られた映像を編集結果として出力・表示する。仮想カメラとしてステレオカメラを使い3次元ディスプレイに表示すれば立体映像として見ることもできる。

また、任意のカメラワークに対応するためには、背景も3次元的なものでなくてはならない。本研究では視点固定型パン・チルトカメラ³⁾により撮影された全方位パノラマ画像や多眼カメラで撮影された全方位ビデオを背景として用いている。こうした全方位パノラマ画像は非常に高解像度であるため、映像化においてズーム倍率を大きくしても十分鑑賞に耐えることができる。

こうした方法で、全方位パノラマ画像と、撮影した人物を複製し3次元的に並べて作った3次元ビデオを映像化した結果を図13に示す。また、図4の舞妓さんの踊りを撮った3次元ビデオの編集例を図14に示す。

4 おわりに

我々は、多視点映像から運動対象の3次元形状とその表面情報を時系列データとして生成し記録する3次元ビデオの実用化を目指した技術開発を行っている。本稿では、これまでの研究で得られた成果として、以下のものを取り上げその概要と各提案手法の性能を示す実験結果を紹介した。

- (a) PC クラスタを用いた多視点ビデオ映像からの実時間3次元形状復元システム
- (b) 弾性メッシュモデルを用いた高精度3次元形状・運動復元法
- (c) 自然な3次元ビデオ生成のための高精度テクスチャマッピング法
- (d) 3次元世界における多様なカメラワークが可能となる3次元ビデオ編集システム

今後の展開としては、

- ・ 安定なシルエット抽出法の開発
- ・ 3次元形状復元のさらなる高速化・高精度化
- ・ 鑑賞に耐えうる画質を持った3次元ビデオの生成法の開発
- ・ 3次元ディスプレイを用いた3次元ビデオの実時間インタラクティブ・ディスプレイシステムの開発

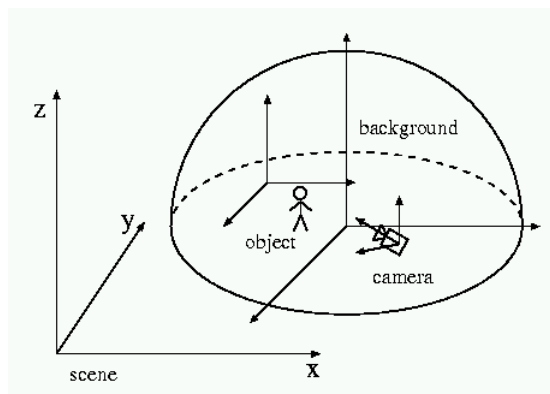


図12 3次元ビデオ編集のための3次元シーンの構成

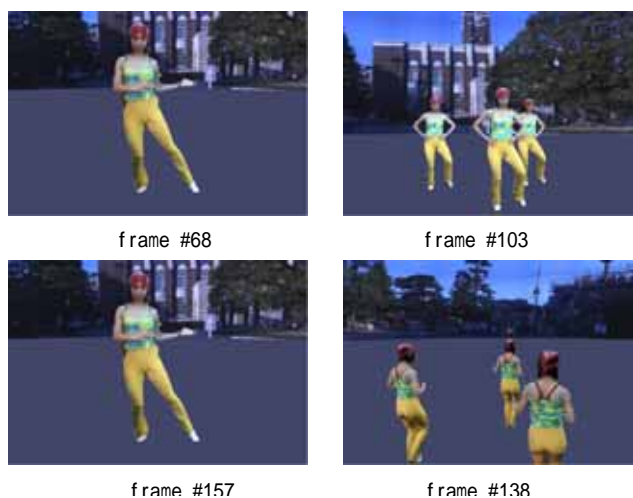


図13 全方位パノラマ画像を使って3次元編集を行った3次元ビデオ

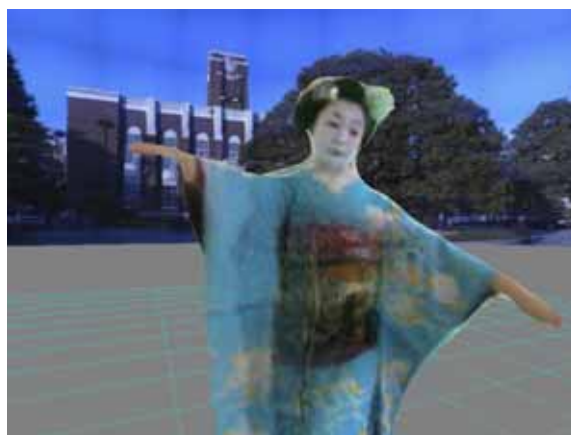


図14 舞妓による日本舞踊を編集した3次元ビデオ

いった現在の手法の改良に加え、

- (1) 首振りカメラを用いて、複数の人物を追跡、ズームアップ撮影して3次元ビデオ化する技術⁴⁾の開発
- (2) 撮影時の照明環境の推定⁵⁾と、3次元ビデオに対する照明効果の導入
- (3) 従来の2次元ビデオと比べ桁違いに大量となるデータの圧縮・符号化法⁶⁾の考案

(4) 3次元ビデオに映された対象の動作解析および、その結果に基づいた3次元ビデオの編集法の開発がある。現在、これらの課題の解決に向けて鋭意研究を進めている。

3次元ビデオの実用化およびその標準化はわが国が世界に先駆けて進めているもので、産官学の連携によって21世紀の新たなメディアテクノロジーの開拓を進めていきたいと考えている。

謝辞

本研究は科学研究費補助金 基盤研究 A 13308017 および特定領域研究 13224051、文部科学省「知的資産の電子的な保存・活用を支援する支援するソフトウェア基盤技術の構築」事業における「大型有形・無形文化財の高精度デジタル化ソフトウェアの開発」プロジェクトの補助を受けて行った。

参考文献

- 1) T. Matsuyama, X. Wu, T. Takai, and S. Nobuhara : Real-Time 3D Shape Reconstruction, Dynamic 3D Mesh Deformation, and High Fidelity Visualization for 3D Video, International Journal on Computer Vision and Image Understanding, Vol.96, No.3, pp.393-434, 2004
- 2) 剣持 雪子, 小谷 一孔, 井宮 淳:点の連結性を考慮した マーチング・キューブ法, 電子情報通信学会技術報告, PRMU98-218, pp.197--204, 1999.
- 3) 和田 俊和, 浮田 宗伯, 松山 隆司:視点固定型パン・チルト・ズームカメラとその応用,電子情報通信学会論文誌 D-II, Vol. J81-D-II, No.6, pp. 1182-1193, 1998.
- 4) 松山 隆司, 浮田 宗伯: 能動視覚エージェント群による協調追跡, 日本ロボット学会誌, Vol.19, No.4, pp.25-31, 2001
- 5) T. Takai, K. Niinuma, A. Maki, and T. Matsuyama: Difference Sphere: An Approach to Near Light Source Estimation, IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. I-98 - I-105, 2004.
- 6) H. Habe, Y. Katsura, and T. Matsuyama: Skin-off:Representation and Compression Scheme for 3D Video, Picture Coding Symposium (PCS) 2004, San Francisco, 2004.12.