

3次元ビデオ映像の撮影・編集・表示

松山 隆司^{*1} 高井 勇志^{*1} ウ 小軍^{*1} 延原 章平^{*1}

Generation, Editing, and Visualization of 3D Video

Takashi Matsuyama^{*1}, Takeshi Takai^{*1}, Xiaojun Wu^{*1}, and Shohei Nobuhara^{*1}

Abstract – 3D video is the ultimate image medium recording dynamic visual events in the real world as is. Recorded object behaviors can be observed from any viewpoint, because 3D video records the object's full 3D shape, motion, and precise surface properties (i.e. color and texture). In this paper, we present research attainments so far obtained:

1. a PC cluster system for real-time reconstruction of dynamic 3D object behaviors from multi-view video images,
2. an active 3D mesh model for reconstructing the accurate 3D object shape,
3. an algorithm of generating natural-looking texture on the 3D object surface from the multi-view video images, and
4. an editing system for visualizing 3D video with an omni-directional background image using versatile 3D camera works.

Experimental results demonstrate the effectiveness of these methods in generating high fidelity and attractive object images from arbitrary viewpoints.

Keywords : 3D video, multi-viewpoint video, 3D shape reconstruction, volume intersection method, PC cluster, texture mapping, arbitrary-viewpoint image generation, 3D video editing

1 はじめに

1次元の音声波形-ラジオ放送-として始まった現在の情報メディアは、その後2次元の平面画像-テレビ放送-へと発展し、今日の我々の生活に欠かす事のできないものとなっている。本研究では、3次元ビデオ映像という3次元情報を持った新たな情報メディアを開拓し、DVDやデジタルテレビ放送、広帯域インターネットを活かした21世紀における映像メディア技術および文化を産み育てることを目的としている。

3次元ビデオ映像は、CGによる仮想的な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立体ビデオ映像で、実世界における対象の振る舞い・動作を余すところなく記録した究極の映像メディアである。

3次元ビデオ映像は、3Dゲームや立体テレビといったエンターテインメント、学校教育における3次元ビデオ辞典、スポーツにおける動作解析、外科手術の3次元モニタリング、さらには伝統舞踊の記録保存といった文化の継承など、さまざまな応用分野での利用が期待できる。

我々の研究室では、3次元ビデオ映像を実用化する

ための具体的な研究課題として、以下のものを取り上げ、研究を進めている。

1. 分散配置された多数の能動カメラによって運動対象を自動追跡し、得られた多視点ビデオ映像から3次元ビデオ映像を実時間で生成するための能動的実時間撮影システムの開発
2. 従来の2次元ビデオ映像と比べ桁違いに大量となるデータの圧縮・符号化法の考案
3. 3次元ビデオ映像に映された対象の動作解析および、その結果に基づいて複数の3次元ビデオ映像を編集し映像作品に仕上げるための映像編集法の考案
4. 2次元あるいは3次元ディスプレイを使った3次元ビデオ映像の対話的表示システムの開発

図1は、現在我々が用いている3次元ビデオ映像撮影・生成過程を示したもので、以下のプロセスから構成されている。

1. 多視点同期ビデオ撮影: 異なる視点に取り付けられた首振りカメラが実時間で対象を追跡し、対象の多視点ビデオ映像を同期撮影する(図2)。
2. 対象シルエットの抽出: 背景差分を用いて各ビデオフレームから対象のシルエットを抽出する。

^{*1}京都大学大学院情報学研究所

^{*1}Graduate School of Informatics, Kyoto University

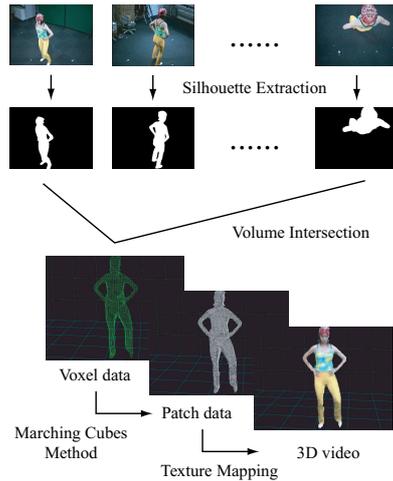


図1 3次元ビデオ映像の撮影プロセス
Fig.1 Processing stages to generate 3D video

3. 視体積交差法による3次元ボクセルデータの獲得: 異なる視点から得られた対象のシルエットを共通の3次元空間に逆投影し, 得られた視体積の共通部分を求め, 対象の3次元形状をボクセル表現として得る.
4. 表面形状の計算: 離散 Marching cubes 法 [16] を用いて形状のボクセル表現を表面パッチ表現に変換し, それを初期値として弾性メッシュモデルを適用して対象の詳細な表面形状を求める.
5. テクスチャマッピング: 対象の3次元形状を表す各表面パッチに対するテクスチャを, 多視点ビデオ映像から求めマッピング処理を行う.

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性能評価実験の結果を紹介する.

1. 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 [1]
2. 弾性メッシュモデルを用いた3次元形状の高精度復元法 [2]
3. 3次元ビデオ映像生成のための高精度テクスチャマッピング法と3次元ビデオ映像編集システム [3, 4]

2 多視点ビデオからの3次元形状復元

2.1 首振りカメラ群を備えたPCクラスタ

我々は, 先の研究 [5] で, 広い(室内)空間を自由に移動する複数の対象を, キャリブレーションされた多数の首振りカメラ群で追跡し, 対象の多視点ビデオ

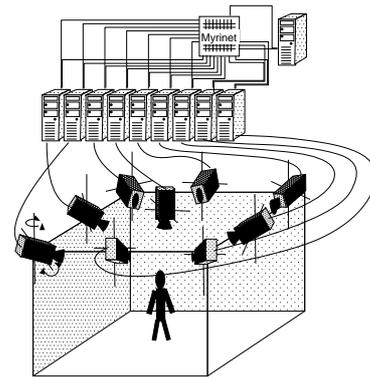


図2 能動的実時間3次元形状復元のためのPCクラスタシステム
Fig.2 PC cluster system for active real-time 3D shape reconstruction

映像を撮影する分散協調型実時間対象追跡システムを開発してきた.

こうした追跡システムを拡張し, 撮影された多視点ビデオ映像から対象の3次元形状を実時間で復元することを目指して図2のようなPCクラスタシステムを開発した(PC:16台, 首振りカメラ:12台).

2.2 3次元形状復元法

対象の3次元形状復元法は, 任意視点映像の生成といった映像生成法への応用 [3] 以外にも, 対象の認識やその動作の解析, またそれらを利用した舞踊や所作の記録・保存といった観点からも重要な研究テーマであると言える.

多視点ビデオから3次元形状を復元するための代表的な方法としては, ステレオ法 [6] と視体積交差法 [7, 8, 9, 10, 11, 12, 13, 14] があり, その特徴は以下のように分類できる.

- カメラ台数

ステレオ法では, カメラの視点位置や視線方向が大きく異なった場合, 画像間での対応付けが非常に難しくなる. そのためカメラを密に配置する必要があるため必要なカメラ台数は多くなる. 一方, 視体積交差法では, 少数のカメラで撮影された画像からでも3次元形状を復元することができる.

- 形状精度

ステレオ法では, 画像間での正確な対応付けができれば高精度な対象形状を復元することができる. 一方, 視体積交差法では, 得られる3次元形状は真の対象形状ではなくそれを含む visual hull でしかない. つまり, 凹の部分をはじめ対象形状が正確に得られるわけではない.

- 実時間性

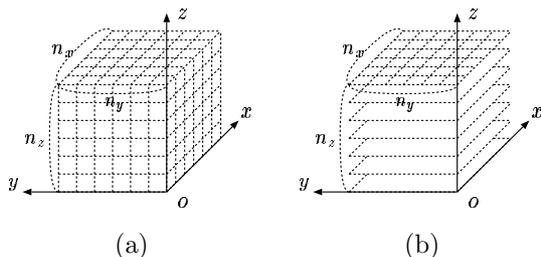


図3 3次元空間の分割法 . (a) 3次元ボクセル空間, (b) 平行平面群への分割

Fig. 3 3D shape representations: (a) 3D voxel space, (b) parallel plane space.

ステレオ法では、各カメラペアから得られるのは対象の2.5次元データ（距離画像）であるため、完全な3次元形状を得るにはこれらを統合する必要がある。専用ハードウェアを用いれば実時間での距離画像取得は可能である [15] が、完全な3次元形状を実時間で得ることは難しい。一方、視体積交差法では、平面間透視投影に基づく方法 [1], 8分木構造を用いて交差判定する方法 [9, 14], 3次元空間での交差判定を2次元画像のシルエット上での重なり判定に置き換える方法 (SPOT)[10] などを、PC クラスタを用いて行うことで実時間で完全な3次元形状を得ることができる。

● 解像度

一般に対象形状の解像度は撮影画像の解像度に依存するため、詳細な形状が必要な箇所はズームアップして撮影する必要がある。対象の動きに合わせた動的なカメラ制御を行うには、3次元形状復元が実時間でいえることが不可欠である。

我々のシステムでは実時間処理に基づく能動的映像撮影を重視し、視体積交差法を基にした3次元形状復元を行っている。また、視体積交差法の問題点である、形状の不正確性は後述する弾性メッシュを用いた方法により解決することができる。

2.3 平面間透視投影を用いた視体積交差法による実時間3次元形状復元

通常の視体積交差法では、図3(a)のように形状復元の対象となる3次元空間をボクセルに分割し、各ボクセルをそれぞれの撮影画像上に投影してボクセルが対象シルエット内か外かを判定する。これに対し、我々の方法では、以下のようにして3次元形状を復元する(図4)。

1. 3次元空間を図3(b)のように平行な平面群として表し、

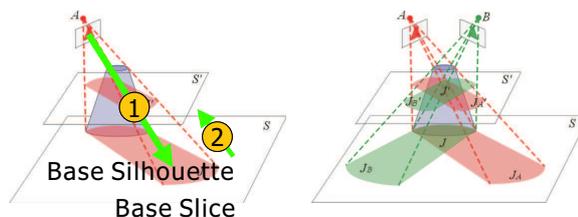


図4 平行平面間透視投影に基づく視体積交差法

Fig. 4 Volume intersection method based on parallel plane-to-plane perspective projection

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4左図の1の投影計算)。
3. 次に基準平面上のシルエットを残りの平行平面群に投影する(図4左図の2の投影計算)。
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4右図)。
5. 得られた断面形状が平行平面間の間隔の厚みを持つものと考え、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(詳細は [1] 参照)。

- 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。
- 特に、平面が平行である場合は、画像の2次元的なスケーリング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。
- 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、本研究では先に述べた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5にその並列処理方式を示す(図5の各列は1台のPCでの処理過程を示している)。

1. 首振りカメラを備えた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(実験では床面)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。

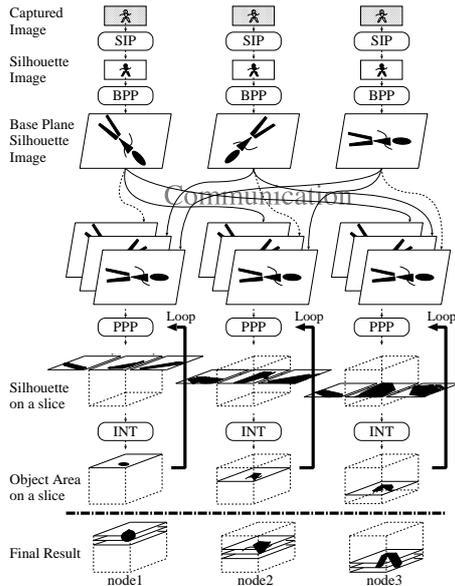


図5 並列視体積交差法の処理の流れ
Fig.5 Processing flow of the parallel plane-based volume intersection method

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に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。

試作したシステムでは、さらなる高速化を図るため、各PCで実行される処理を画像キャプチャ(IC)、シルエット画像生成(SIP)、基準平面への逆投影(BPP)、基準シルエットの相互コピー(BPP)、基準シルエットの平行平面群への投影と交差計算(PPP&INT)の5ステージに分け、パイプライン処理を行わせるようにした。

2.4 性能評価実験

実験は、おおよそ $4m \times 4m \times 2.5m$ の室内で行い、天井に設置された9台のカメラで多視点ビデオを撮影して行った(図6)。

復元される3次元形状の空間解像度は、入力画像サイズを 640×480 ピクセルとしたとき、約 $1cm$ であった。以下では実時間処理を目指すため、画像サイズはすべて 320×240 とし、空間解像度は $2cm$ とした。

まず、9台のカメラで撮影し、9台のPCで処理を行った場合、計算速度は 9.78 Volume/sec であった。各PCにおける処理時間を図7に示す。このグラフの横軸はPCの番号、左奥行き方向の軸はパイプライン処理のステージを表しており、各ステージでの処理時間を棒グラフで示した。また、グラフの右奥隅の棒は、処理全体の時間を表している。

第1ステージの画像入力および第2ステージのシル



図6 9台のカメラで撮した多視点画像
Fig.6 captured multi-viewpoint images

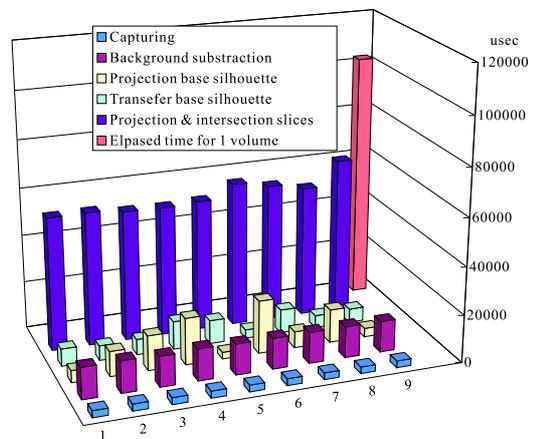


図7 PC毎の各処理ステージの計算時間(カメラ:9台, PC:9台)
Fig.7 Average elapsed time at each processing stage (9 cameras, 9 PCs)

エットの計算は入力画像サイズにのみ依存するので、各PC間でばらつきはないが、第3ステージの基準シルエットの計算はカメラの視線ベクトルと基準平面の法線ベクトルとがなす角度によって計算量が変わるので、各PC間で大きくばらつく。また、第4ステージの基準シルエットの相互コピーは、データ量が少なく高速のネットワークを使っているため大きなオーバーヘッドにはなっていない。

最も時間がかかった処理は第5ステージにおける基準シルエットの投影とシルエット交差を求める計算である。このステージの処理は完全に独立並列処理が可能なので、計算を行うPCを増やせば、全体のパフォーマンスが向上できると考えられる。

これを検証するため、計算のみを行うPCを追加して、計算時間を測った。具体的には、9台のカメラで撮影し、計算するPCを9台から15台まで増やして、各々の場合の処理時間(平行平面群への投影・シルエット交差計算時間)を求めた(図8)。このグラフの横

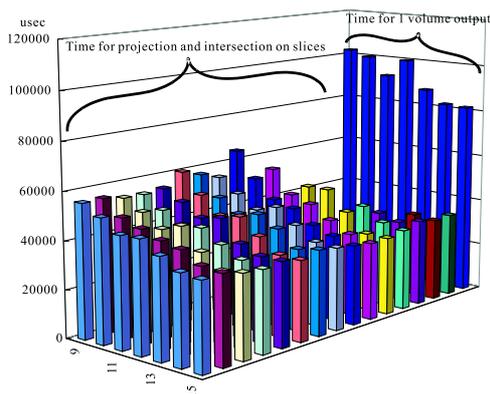


図8 PC 毎の平行平面群への投影・シルエット交差計算時間 (カメラ: 9台, PC: 9~15台)
 Fig. 8 Average elapsed time of the parallel plane-to-plane projection and silhouette intersection on each PC (9 cameras, 9-15 PCs)

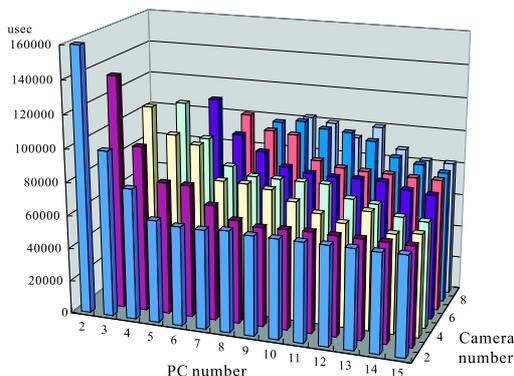


図9 撮影用カメラ台数と計算用 PC 台数を変えた場合の 1 Volume 復元の計算時間
 Fig. 9 Average elapsed time of the 3D volume reconstruction with various combinations of cameras and PCs

軸は PC の番号, 左奥行き方向の軸は PC の総数を表し, 第 5 ステージの計算時間を棒グラフとして表した。また, グラフの右端の棒は, 各台数での 1 Volume 復元のためのトータルな処理時間を表している。このグラフから, 計算用 PC の台数が増えるにつれ, 計算時間が短縮され, 全体のパフォーマンスも向上することが確かめられた。15 台の PC で計算した場合, 12.5 volume/sec の速度で形状復元ができた。

最後に, 撮影を行うカメラを 2 台から 9 台まで変え, それぞれの場合に計算用 PC の台数を変化させ, 1 Volume 復元の処理時間を求めた結果を図 9 に示す。図の横方向は計算する PC の台数, 奥行き方向は撮影を行うカメラの台数を表している。

このグラフでは, 計算する PC の台数が増えるにつれ, 処理時間は短縮されるが, 約 60 msec という一定値で飽和してしまうことが示されている。これは以下

の理由によると考えられる。使用したキャプチャカードは NTSC の同期信号で画像のキャプチャ開始時刻を決めるため, 連続キャプチャ要求の間に, 最短でも 1/30 sec のウェイトが生じる。また, 各カメラにおける NTSC 信号の同期がとれていないため, PC 間で撮影の同期をソフトウェア的にとる必要がある。これによって, 連続キャプチャ要求の間に, 最短 1/15 sec のウェイトが必要になる。この結果, 処理時間自身は約 1/15 sec (= 60 msec) で飽和することになる。すなわち, 同期式カメラを使い, 十分な数の PC を用意することによって, 本研究で考案した方式でビデオレートでの 3 次元形状復元が可能であることが実証されたことになる。

3 弾性メッシュモデルを用いた 3 次元形状の高精度復元法

先に述べたように, 視体積交差法では, 対象の凹部分が正確に復元できないという問題がある。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 [16] 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。

弾性メッシュモデルでは, ステレオ法や視体積交差法, Space-Carving 法 [18] 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。さらに, 今回考案した弾性メッシュモデルでは,

1. photometric consistency 制約: 各カメラから対象表面のメッシュを見た場合, そのテクスチャが一致する
2. 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
3. 単純閉曲面制約: メッシュ表面はねじれの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きる。

3.1 変形アルゴリズム

メッシュの変形は, メッシュを構成する各頂点の移動によって実現する。この際, メッシュの数が膨大になるため, SNAKES[17] のように全体的なエネルギー関数を最適化するように全頂点の移動を行うのではなく,

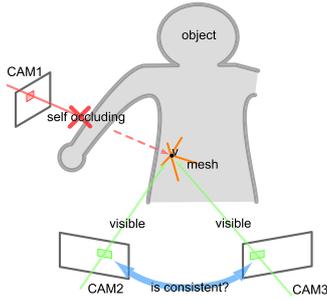


図 10 外力項と Visibility
Fig. 10 Photometric consistency and visibility

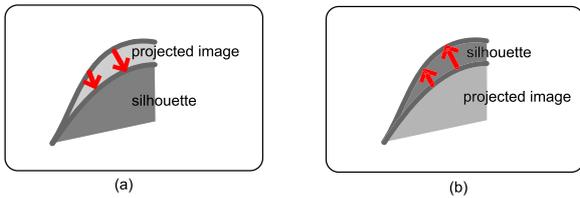


図 11 シルエット保持力 (a) 輪郭外に投影されたとき (b) 輪郭内に occluding contour として投影されたとき
Fig. 11 Silhouette preserving force

各頂点に働く力を計算し、それによって各頂点を独立に移動させるという変形処理を繰り返すことにした。

本手法のアルゴリズムをまとめると以下のようになる。

step 1 離散マーチングキューブ法によって、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

3.2 頂点に作用する力

頂点を v で表したとき、その 3 次元座標を q_v で表す。また、頂点 v を観測することができるカメラの集合を C_v で表す。例えば図 10 の場合、 $C_v = \{CAM_2, CAM_3\}$ となる。

各頂点に働く力として以下の 3 つのものを考える。

外力 F_e

まず、弾性メッシュモデルを対象形状に近づけるための力として外力項 F_e を導入する。

$$F_e \equiv \nabla E_e(v). \quad (1)$$

ここで $E_e(v)$ は、画素値の分散を表し、次のように定義される (図 10)。 C_v に含まれる各カメラ c で撮影された画像上において、 v を写していると考えられる画素の値 $p_{v,c}$ をそれぞれ求め、それらの平均 \bar{p}_v を用いて、

$$E_e(v) \equiv \frac{1}{N(C_v)} \sum_{c \in C_v} \|p_{v,c} - \bar{p}_v\| \quad (2)$$

とする。つまり、 F_e は、頂点 v を、撮影画像における画素値の分散がより小さくなる方向へ移動させようとする力である。

内力 F_i

上記 F_e のみで変形を行うと、メッシュの交差が起きる可能性がある。これを防ぐために、次式で定義される内力項 F_i を導入する。

$$F_i \equiv \frac{\sum_j^n q_{v_j} - q_v}{n}, \quad (3)$$

但し v_j はメッシュ上で v と隣接する頂点とする。これは頂点間の張力として働き、メッシュ形状の滑らかさを保つ。

シルエット保持力 F_p

弾性メッシュモデルは常に visual hull に内接し、各カメラで撮影した画像中のシルエット輪郭を保つべきである。このような働きをする力として、シルエット保持力 F_p を導入する。カメラ c で撮影した画像から得られる対象本来のシルエットを $S_{o,c}$ 、現在得られているメッシュをその画像上に投影して得られるシルエットを $S_{m,c}$ とする。このとき、 C_v に含まれるカメラ c に対して、 v が下記の場合にモデルと対象のシルエットを一致させる力 $f_p(v, c)$ を v に加える。

- v が $S_{o,c}$ の領域外に投影される場合、 v の投影位置を $S_{o,c}$ の輪郭に一致させるように働く力として、 v から $S_{o,c}$ への最短移動ベクトル $f_p(v, c)$ を使う (図 11(a))。
- v が $S_{o,c}$ の領域内に投影され、且つ、 $S_{m,c}$ の輪郭を成すときは、 v の投影位置を $S_{o,c}$ の輪郭に一致させるように働く力として、 v から $S_{o,c}$ への最短移動ベクトル $f_p(v, c)$ を使う (図 11(b))。
- 上記以外の場合、 $f_p(v, c) = 0$ とする。

$f_p(v, c)$ は、各カメラごとに求められ、それらの合成力によって実際の頂点移動のための力 F_p が次式のように定義される。

$$F_p \equiv \sum_{c \in C_v} f_p(v, c). \quad (4)$$

以上 3 つの力の重み (α, β, γ) 付き和として頂点に作用する合成力 F を定義する。

$$F = \alpha F_i + \beta F_e + \gamma F_p \quad (5)$$

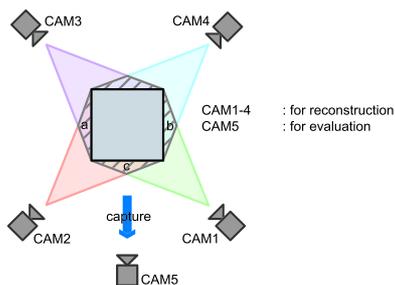


図 12 評価実験時のカメラ配置
Fig. 12 Camera arrangement

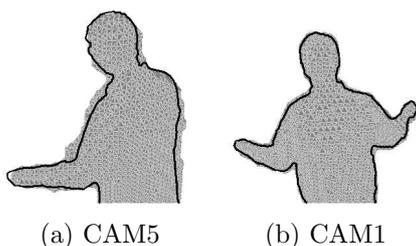


図 13 初期形状 (a) は図 12 の CAM5 の (仮想) 視点, (b) は CAM1 の観測視点
Fig. 13 Initial shapes

3.3 性能評価

図 12 のようなカメラ配置で, CAM1 から 4 を用いて上記の手法で形状復元を行い (仮想的に設定した) CAM5 を用いて評価を行った. このとき CAM5 から見た場合, 当初同図 a, b の領域が $S_{m,c}$ の輪郭として捉えられ, それが弾性メッシュの変形によって $S_{o,c}$ に近づかずである.

CAM1 ~ 4 のデータを使って視体積交差法で得られた初期形状を CAM5 および CAM1 の視点から見た場合のシルエット像を図 13 に示す. これらの図中の太い輪郭線は, 対象の真の輪郭 $S_{o,c}$ を表している (形状復元には用いていないが CAM5 の位置からも対象が撮影されており, その画像から対象の真の輪郭が得られる.) 図 13(a) から分かるように, CAM5 から見た場合のメッシュ輪郭は正確ではない.

実験では, F として, (a) F_i のみの場合, (b) F_i, F_p を用いた場合, (c) F_i, F_e, F_p すべて用いた場合, について計算を行った. この実験結果を図 14 に示す. 各列上段の左側が図 12 での CAM5 の視点, 右側が CAM1 の視点から復元結果を見たときのシルエットである. それぞれ撮影画像から抽出した $S_{o,c}$ の輪郭が太線で示されている. また下段グラフは, 横軸が変形処理の反復回数, 縦軸は復元された形状を各カメラの画像平面に投影したときの輪郭線と対象の真の輪郭線との平均的なずれ量を表している.

この結果から, 以下のことが分かる.

- 内力項のみの場合, 弾性メッシュは全体として縮小し, 輪郭も崩れる (図 14(a)).
- 各視点 (CAM1 ~ CAM4) での $S_{o,c}$ の輪郭を保つようにした場合, その視点での $S_{o,c}$ の輪郭が保たれる一方で, これ以外の視点 (図 CAM5) では実際の対象形状との差異が大きくなる (図 14(b)).
- 更に表面テクスチャの一致制約を表す外力項を加えた場合には, 復元に用いなかった視点 (CAM5) においても実際の対象形状に近づく (図 14(c)).

以上のことから, 対象の高精度な形状復元のためにはこれら 3 つの力全てが必要であることが示された.

ここで提案した手法と Space-Carving 法 [18] を比較すると, 後者は photometric consistency を主として利用しているのに対し, 我々の方法はそれに形状の空間的連続性とシルエット制約を加えたものとなっており, より強力な形状復元法であるといえる. さらに, モデルフィッティングを基本原理とする本手法では, 時系列データに対しても同一のトポロジーを持ったメッシュモデルによるフィッティングが可能であり, それによって対象の動作解析や, データ圧縮などの問題を扱うこともできるようになる.

4 高精度 3次元ビデオ映像レンダリング

4.1 多視点ビデオからの任意視点映像生成

多視点ビデオから任意視点映像を生成する方法は, そのアプローチから以下のように分類できる.

まず, 対象の形状については,

1. 3次元形状を復元しない方法 (appearance の復元)
2. 3次元形状を復元する方法 (generic property の復元)

また, 表面テクスチャ情報については,

1. 撮影した照明環境における見え方を復元する方法 (appearance の復元)
2. 照明環境に不変な表面情報を復元する方法 (generic property の復元)

となる. まとめると, 表 1 のように 4 つのタイプがある.

表 1 任意視点映像生成法の分類
Table 1 Approaches to the arbitrary-viewpoint image generation

形状	app	app	gen	gen
テクスチャ	app	gen	app	gen
Type	I	II	III	IV

Type I (形状:app & テクスチャ:app)

陽に 3次元形状を復元することなく, 任意視点から見

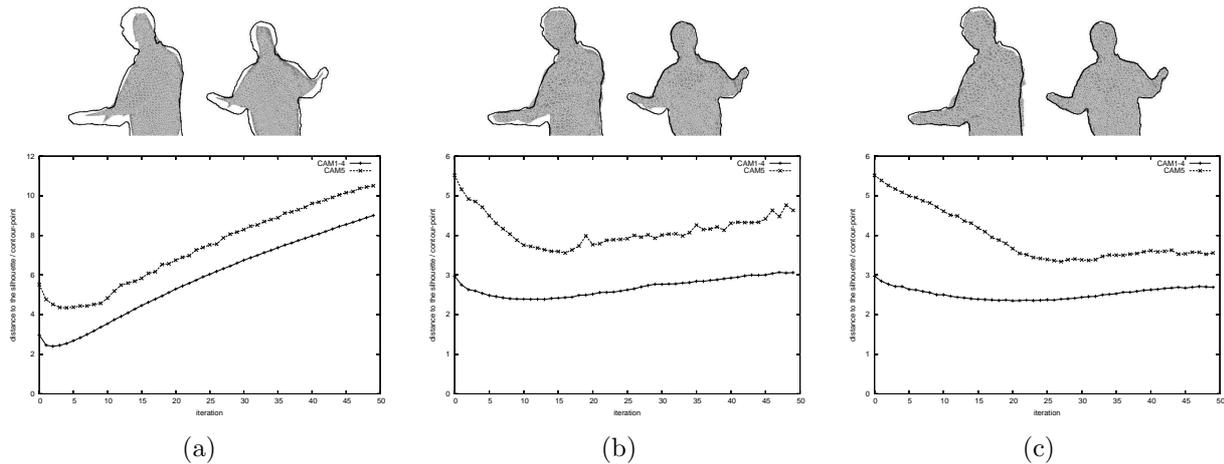


図 14 実験結果 (a) F_i のみの場合 ($\alpha = 1.0, \beta = 0.0, \gamma = 0.0$), (b) F_i, F_p を用いた場合 ($\alpha = 0.5, \beta = 0.0, \gamma = 0.5$), (c) F_i, F_e, F_p すべて用いた場合 ($\alpha = 0.3, \beta = 0.4, \gamma = 0.3$).

Fig.14 Experimental results

た映像を生成する手法である。3次元形状を復元することは、見えない部分の形状データの生成や保持を伴うため、計算コストを増大させデータも冗長になると言える。そこで、非常に多くの多視点映像から、指定された視点・視線に近い映像を合成する手法 [19] や、中間表現としては対象の3次元形状を復元しつつも、最終的にはこの形状データを含まない手法 [20]、また、中間表現として2次元平面の組み合わせを用いる手法 [21] などが研究されている。しかし、動作認識や表面の反射特性の推定には3次元形状が不可欠なため、このタイプの手法の応用は限られている。

Type II (形状:app & テクスチャ:gen)

一般に、照明に不変な対象表面の反射特性を推定するには、対象の3次元形状を復元し反射モデルに基づいた解析 [22] を行う必要がある、このタイプの手法の実現は難しい。一方、文献 [23] や [24] では、反射モデルを利用することなく、多様な照明下で撮られた多数枚の画像に対して固有画像解析を施すことによって、任意照明下でのテクスチャの復元が行えることを示している。

Type III (形状:gen & テクスチャ:app)

対象の3次元形状を復元し、表面情報は見え方をベースに復元する手法で、本研究のアプローチはこのタイプである。この手法では、対象の3次元的な動作に同期したカメラワークなど、見え方のみの手法とは異なる映像表現が可能となる。ただし、表面テクスチャは撮影時の照明環境下でのものが使われるため、照明条件を変更した映像は生成できない。

Type IV (形状:gen & テクスチャ:gen)

対象の3次元形状の復元と共に、撮影されたビデオ映像から撮影時の照明環境に依存する要素を取り除

き、対象表面の反射特性を推定する手法である。このタイプでは、任意の光源下での見え方の復元といった、他のタイプでは表現不可能な映像を生成することが可能となるが、対象の反射モデルの推定には、正確な3次元形状情報の取得や撮影時の照明環境の推定 [25, 26, 27, 28, 29, 30] といった難しい処理が必要となる。

本研究では、3次元空間における3次元ビデオ映像の編集、対象の動作に同期したカメラワーク設定といった3Dビデオ映像とユーザとのインタラクションの実現を目指しているため、対象の3次元形状を復元し、それに基づいた任意視点映像の生成を考えた。

現在本研究で用いている対象の3次元形状は、本稿の2.3節で述べた方法によって得られたボクセルデータに離散マーチングキューブ法を適用して得られる三角形パッチモデルであり、その形状は必ずしも正確であるとは言えない(本稿3節の弾性メッシュモデルを適用した場合でも、カメラのキャリブレーションの誤差により、復元された3次元形状は必ずしも完全であるとはいえない。)

本研究では、このような必ずしも正確であるとは言えない対象の3次元形状に対して、写実性の高い映像を再現するためのテクスチャマッピング法を開発することを目指している。

4.2 視点依存頂点ベーステクスチャマッピング法

我々の先の研究 [4] で用いたテクスチャマッピング法は、各表面パッチの法線と最も近い視線を持つカメラを選び、そのカメラによる撮影画像からテクスチャを切り出し、それを対象表面パッチに張付けるという方法であった。この方法では、任意視点映像を生成する際の視点位置や、パッチ間での空間的連続性を考慮

していないため、以下のような問題が生じていた(以後、この方法を視点独立面ベース法と呼ぶ.)。

1. カメラ間の撮影映像の差異によるちらつき
マーチングキューブ法により生成される面は滑らかではないため、隣接する面パッチ間で法線の向きが大きく変化する。そのため、テクスチャマッピングに使用するカメラが隣接パッチで異なってしまう。つまり、異なったカメラで撮られた画像が対象表面の隣接パッチに張られ、映像化された対象表面にちらつきが生じてしまう。(視点独立法の問題点)
2. 隣接面パッチ間でのテクスチャの不整合によるちらつき
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、テクスチャは面パッチ毎に独立して生成されるため、各面パッチの継ぎ目において不整合が生じることがある。このことによっても映像化された対象表面にちらつきが生じてしまう。(面ベース法の問題点)

上記の問題を解決し、滑らかなテクスチャマッピングを実現するために、映像を表示する時の視点位置を利用し、対象表面パッチの頂点色を補間することでテクスチャを生成する方法 - 視点依存頂点ベース法 - を考えた [3]。

視点依存頂点ベース法のアルゴリズムの概要は以下のように書ける。

1. 表示視点・視線を決め、撮影に用いた各カメラ視線との一致度を計算する。
2. メッシュの各頂点について、1で求めた一致度に基づいた重みづけを利用して、各カメラ画像のRGB値から頂点のRGB値を計算する。
3. 各三角形パッチ内のテクスチャを3頂点のRGB値の線形補間によって生成する。

以上の方法により、上記1, 2の問題が解決できる。

4.3 性能評価

視点依存頂点ベース法(以下手法(I)と呼ぶ)の精度を評価するために、従来法である視点独立面ベース法(手法(II))との比較を行った。

あるカメラ c_j の位置・視線方向を表示視点・視線とし、手法(I)と手法(II)によりそれぞれレンダリングを行い、生成された画像シーケンスのRGB値とカメラ c_j によって撮影された生画像のRGB値との2乗平均平方根誤差を求めた。尚、手法(I)-1はカメラ c_j によって撮影された画像を含めて映像化した場合、手法

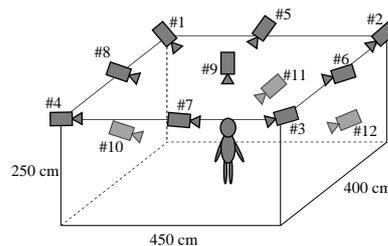


図 15 撮影環境
Fig. 15 Camera setting

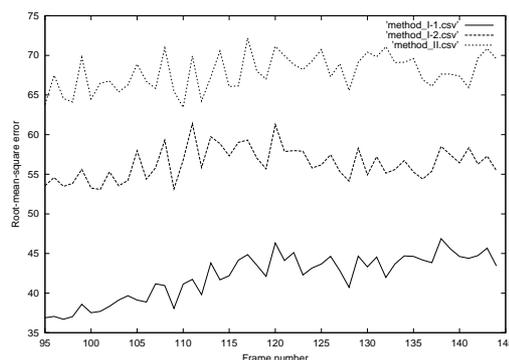


図 16 RGB 値の 2 乗平均平方根誤差 (I) : 縦軸は画素の RGB 値の 2 乗平均平方根誤差, 横軸は 3 次元ビデオ映像のフレーム番号を表しており, グラフは上から, 手法 (II), 手法 (I)-2, 手法 (I)-1 による結果となっている。

Fig. 16 Root-mean-square errors of RGB values

(I)-2 はカメラ c_j による画像は含めずに映像化した場合である。実験環境は、以下の通りである。

- 使用カメラ : 12 台 (図 15)
- カメラ画像 : 640×480[pixel] RGB 各 8bit
- 復元視点 : CAM5 (図 15)

図 16 が示すように、手法 (I)-1, 手法 (I)-2 は手法 (II) に比べ大幅に誤差を軽減している。また、手法 (I)-2 においても手法 (II) より良い結果を示したということは、提案手法の有効性を示していると言える。

また、生成された画像 (図 17) を見ると、手法 (I) では、手法 (II) によりレンダリングされた画像に見られる面パッチ間でのちらつきや不連続性が解消され、より写実性の高い 3D ビデオ映像がレンダリングされることが分かる。

5 3次元ビデオ映像の編集

5.1 3次元世界での映像編集

3次元ビデオ映像の編集は、前節で得られた対象の実写 3次元映像、全方位背景、仮想カメラからなる 3

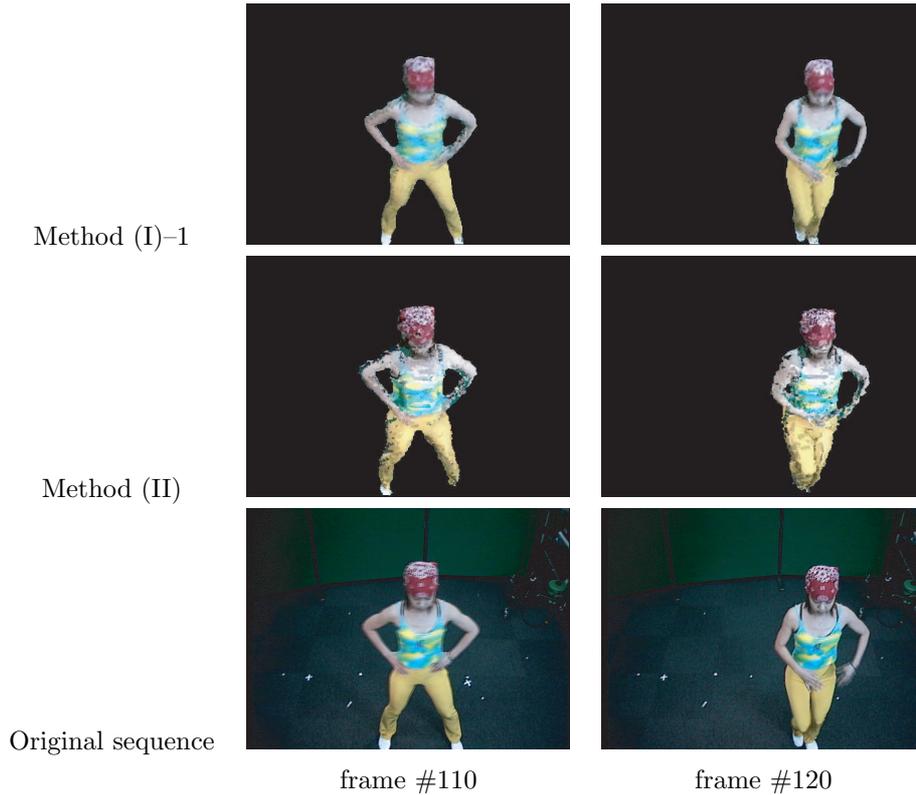


図 17 生成された 3 次元ビデオ映像の例
Fig. 17 Sample frames of generated 3D video

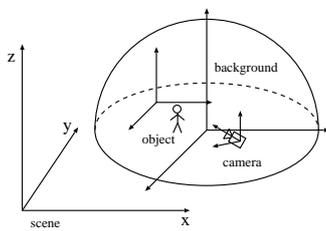


図 18 仮想シーンのセットアップ
Fig. 18 Virtual scene setup

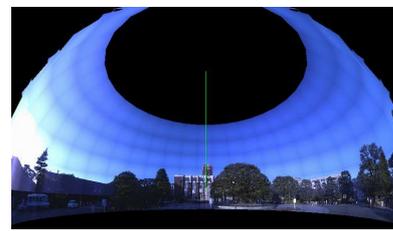


図 19 全周囲パノラマ画像
Fig. 19 Omni-directional image

次元の仮想シーンにおいて行われる (図 18)。この仮想シーン内において、仮想カメラを任意の数・位置に配置し、任意の 3 次元的なカメラワークを施すことで、多様な映像を編集・表示することができる。

こうした任意のカメラワークに対応できる背景画像として、本研究では視点固定型パン・チルトカメラにより撮影された全周囲パノラマ画像 [31] (図 19) を用いている。この全周囲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。

図 18 に示した仮想シーンに基づいて 3 次元ビデオ映像を編集・表示するには、対象・背景・カメラの 3 次元位置・姿勢・大きさ・数など非常に多くのパラメータを時系列データとして設定する必要がある。このよ

うに、高い自由度を持つパラメータセットを有することは多様な映像表現を可能にする反面、ユーザに対して大きな負担を強いることになる。

そこで本研究では、このパラメータ設定において以下の 2 通りの設定法を考え、これらのパラメータをインタラクティブに設定するための GUI を開発した [3, 4]。

1. キーフレームにおいて設定した各パラメータを時間的に補間する手法
2. 対象の 3 次元形状・運動を利用してパラメータ設定を自動的に行う手法

この GUI を用いることで、任意視点映像を生成するためのカメラパラメータ設定を容易かつ直感的に行うことができる。

5.2 3次元ビデオ映像の編集例

運動対象(人間)と背景を図18のように配置し3Dビデオ映像の編集・映像化を行った結果を図20に示す。この図中での、フレーム68と157は対象の正面を撮り続けるようにカメラワークを設定した例であり、フレーム103, 138, 285は、人物を3体複製して3次元的に配置・映像化した例である。この図が示すように、3次元ビデオ映像では対象の3次元形状を復元しているため、対象の動きに合わせた任意視点映像の生成や、仮想空間中の任意の位置への対象の配置や複製、といった映像表現が可能となっている。

6 むすび

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次元ビデオ映像」を提案している。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。

1. 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
2. 弾性メッシュモデルを用いた3次元形状の高精度復元法
3. 3次元ビデオ映像生成のための高精度テクスチャマッピング法
4. 3次元世界における多様なカメラワークが可能な3次元ビデオ映像編集システム

今後の展開としては、安定なシルエット抽出法の開発、3次元形状復元の高速度・高精度化、鑑賞に耐えうる3次元ビデオ映像の生成法の開発、高精度な全方位背景ビデオ映像の導入、2次元、没入型、3次元ディスプレイを用いた3次元ビデオ映像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、といった現在の手法の改良に加え、

- 撮影時の照明環境の推定と、3次元ビデオ映像に対する照明効果の導入
- 従来の2次元ビデオ映像と比べ桁違いに大量となるデータの圧縮・符号化法の考案
- 3次元ビデオ映像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次元ビデオ映像の編集法の開発

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おり、そのための新たな実験環境として、視

点固定型パン・チルト・ズーム・デジタルカメラ25台、PC32台を備えた100m²規模のスタジオの準備を行っている。このスタジオが完成すれば、京舞や能などの伝統舞踊のデジタルアーカイブ化(「デジタル人間国宝アーカイブ」)を具体的応用分野として研究を行う計画である。

謝辞

本研究は科学研究費補助金 基盤研究A 13308017の補助を受けて行った。また、実験に使用したデータは、技術組合 新情報処理開発機構の画像理解評価用画像データベース 2001 に収録されているものである。

参考文献

- [1] ウ小軍, 和田俊和, 東海彰吾, 松山隆司: 平面間透視投影を用いた並列視体積交差法, 情報処理学会論文誌: コンピュータビジョンとイメージメディア, Vol.42, No.SIG 6(CVIM 2), pp.33-43, 2001.
- [2] ウ小軍, 延原章平, 和田俊和, 松山隆司: 多視点映像からの実時間3次元形状復元とその高精度化, 情報処理学会研究会資料, CVIM-131-9, 2002.
- [3] 高井 勇志, 松山 隆司: 3次元ビデオ映像の高精細表示アルゴリズムと編集システム, 映像情報メディア学会論文誌, Vol.56, No.4, pp.593-602, 2002.
- [4] 圓藤 康平, 西出 義章, 和田 俊和, 松山 隆司: 3Dビデオ(2): 多視点映像からの3D映像の生成と表示, The Proc. of IMPS, pp.11-12, 2000.
- [5] 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
- [6] T. Kanade, Peter Rander, and P.J. Narayanan: Virtualized Reality: Constructing Virtual Worlds from Real Scenes, IEEE Multimedia, pp.34-47, 1997.
- [7] A. Laurentini: How far 3d shapes can be understood from 2d silhouettes, IEEE Transactions on Pattern Analysis and Machine Intelligence, 17(2), pp.188-195, 1995.
- [8] S. Moezzi, L. Tai, and P. Gerard: Virtual View Generation for 3D Digital Video, IEEE Multimedia, pp.18-26, 1997.
- [9] E. Borovikov and L. Davis: A Distributed System for Real-time Volume Reconstruction: Proc. of CAMP2000, pp.183-189, 2000.
- [10] G.Cheung and T.Kanade: A Real Time System for Robust 3D Voxel Reconstruction of Human Motions, Proc. of CVPR, pp.714-720, 2000.
- [11] W. N. Martin and J. K. Aggarwal: Volumetric description of objects from multiple views, IEEE Transaction on Pattern Analysis and Machine Intelligence, 5(2), pp.150-158, 1987.
- [12] M. Potmesil: Generating octree models of 3d objects from their silhouettes in a sequence of images, Computer Vision, Graphics, and Image Processing, 40, pp.1-29, 1987.
- [13] P. Srivasan, P. Liang, and S. Hackwood: Computational geometric methods in volumetric intersections for 3d reconstruction, Pattern Recognition, 23(8), pp.834-857, 1990.
- [14] R. Szeliski: Rapid octree construction from image sequence, CVGIP: Image Understanding, 58(1),

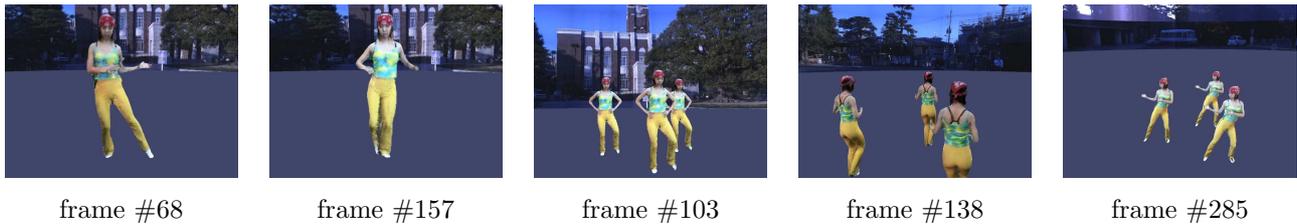


図 20 全方位背景を持つ 3 次元ビデオ映像
Fig. 20 3D video with an omni-directional background

pp.23-32, 1993.

[15] T. Kanade, H. Kato, S. Kimura, A. Yoshida, and K. Oda : Development of a Video-Rate Stereo Machine, Proc. of International Robotics and Systems Conference (IROS '95), Human Robot Interaction and Cooperative Robots, Vol. 3, August, 1995, pp. 95 - 100.

[16] 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.

[17] M. Kass and A. Witkin and D. Terzopoulos: Snakes: Active Contour Models. Proc. of International Conference on Computer Vision, pp.259-268, 1987.

[18] K. N. Kutulakos and S. M. Seitz: A theory of shape by space carving, Proc. of IEEE International Conference on Computer Vision, pp.307-314, 1999.

[19] 映像情報メディア学会 編 : 映像情報メディアハンドブック, pp.515-519, 2000.

[20] 向川 康博, 中村 裕一, 大田 友一 : 2 枚の顔写真を用いた任意方向の顔画像の生成, 情報処理学会論文誌, Vol.37, No.4, pp.635-644, 1996.

[21] 北原 格, 大田 友一 : 多視点映像の融合によるスポーツシーンの自由視点映像生成 - 3 次元形状表現用平面の適応的配置 -, PRMU2000-189, pp.23-30, 2001.

[22] 富永昌治 : コンピュータビジョンにおけるカラー情報の表現と解析, コンピュータビジョン (松山, 久野, 井宮編集), pp.64-79, 1998.

[23] 西野 恒, 佐藤 いまり, 佐藤 洋一, 池内 克史 : Eigen-Texture 法 : 複合現実感のための 3 次元モデルに基づく見えの圧縮と合成, 電子情報通信学会論文誌, D-II, Vol.J82-D-II, No.10, pp1793-1803, 1999.

[24] 東海 彰吾, 弓場 竜, ウ 小軍, 松山 隆司 : アクティブカメラを用いた動的広域シーンの映像提示法, 電子情報通信学会論文誌, D-II, Vol. J84-D-II, No.8, pp.1910-1920, 2001.

[25] 竹内 英人, 池内 克史 : 3 次元モデル再構築のための光源の影響を補正したテクスチャ合成, 情報処理学会研究報告「コンピュータビジョンとイメージメディア」, No.123-6, pp.43-50, 2000.

[26] 飯山 将晃, 青木 啓史, 亀田 能成, 美濃 導彦 : 未知光源環境での物体形状と反射特性のボクセル独立な再構成法, 情報処理学会論文誌, Vol. 42 No.12, pp.3185-3193, 2001.

[27] S. R. Marschner and D. P. Greenberg : Inverse Lighting for photography, IS&T/SID Fifth Color Imaging Conference, pp.262-265, 1997.

[28] 西野 恒, 池内 克史, 張 正友 : 疎な画像系列からの光源状況と反射特性の推定, 「画像の認識・理解シンポジウム (MIRU 2002)」 論文集 I, pp.21-28, 2002.

[29] 町田 貴史, 竹村 治雄, 横矢 直和 : 複数の照明条件の組み合わせによる物体の表面反射特性の密な推定, 電子情報通信学会論文誌, D-II Vol. J84-D-II No. 8, pp.1873-1881, 2001.

[30] I. Sato, Y. Sato, and T. Ikeuchi : Illumination Distribution from Brightness in Shadows: Adaptive Estimation of Illumination Distribution with Unknown Reflectance Properties in Shadow Regions, Proc. IEEE ICCV'99, pp.306-312, 1999.

[31] 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, 信学論 D-II, Vol. J81-D-II, No.6, pp. 1182-1193, 1998.

(2002 年 9 月 10 日受付)

[著者紹介]

松山 隆司



1976 年, 京都大学大学院修士課程修了。京都大学助手, 東北大学助教授, 岡山大学教授を経て, 1995 年より, 京都大学大学院電子通信工学専攻教授。現在, 同大学院情報学研究科知能情報学専攻教授。工博。画像理解, 人工知能, 分散協調視覚の研究に従事。1980 年, 情報処理学会創立 20 周年記念論文賞, 1990 年, 人工知能学会論文賞, 1993 年, 情報処理学会論文賞 1994 年, 電子情報通信学会論文賞, 1995 年, 第 5 回国際コンピュータビジョン会議 Marr Prize, 1996 年, 国際パターン認識連合 Fellow, 1999 年, 電子情報通信学会論文賞, 2000 年, 画像センシングシンポジウム優秀論文賞。

高井 勇志



2000 年, 奈良先端科学技術大学院大学情報科学研究科博士前期課程修了。現在, 京都大学大学院情報学研究科博士後期課程在籍。コンピュータビジョン, コンピュータグラフィックスに関する研究に従事。

ウ 小軍



2000 年京都大学大学院情報学研究科知能情報学専攻修士課程修了。現在同大学院同専攻博士後期課程在籍。実時間能動的 3 次元形状復元の研究に従事。

延原 章平



2002 年京都大学大学院情報学研究科博士前期課程修了。現在同大学院情報学研究科博士後期課程在籍。コンピュータビジョン, コンピュータグラフィックスに関する研究に従事。