

高精細3次元ビデオのためのズームアップ・パンチルト撮影システム*

藤根 成暢[†] 浮田 宗伯^{†a)} 萩田 紀博[†]

High-Fidelity Performance Capture System Using Zoom-In Pan-Tilt Cameras*

Shigenobu FUJINE[†], Norimichi UKITA^{†a)}, and Norihiro HAGITA[†]

あらまし 複数のパンチルトカメラで人体の各部をズームアップ撮影することで、形状復元の解像度向上と誤差低減、及び高解像度テクスチャ撮影を可能とするシステムを開発する。技術課題は、各瞬間においてより形状復元誤差が小さく、かつ多くの高解像度テクスチャを撮影できるように全てのパンチルトカメラの視線方向を決定することにある。このカメラワークのためには、対象である人体形状の動きが有用な情報となる。しかし、人体全身の形状を実時間で解析することは難しいため、撮影動作を既知とし、事前計測した対象全身の時系列形状において、各ズームアップカメラの撮影対象候補である人体パーツの配置を解析し、記録しておく。本番撮影時には、オンラインで復元された形状と記録データを比較し、類似する記録データに対応する人体パーツの配置から、オンラインの人体パーツ配置を推定する。この配置に基づき、複数の簡易な評価関数からカメラワークを高速決定する。実験では、記録データを用いた実時間カメラワークと、撮影された高解像度画像からの3次元ビデオ生成を行い、提案システムの有効性を確認した。

キーワード 3次元ビデオ, 形状復元, 形状ラベリング, カメラワーク

1. ま え が き

伝統舞踊のような無形文化財や優秀な演技・スポーツのデジタル保存が可能になると、それらの技の永久保存やデジタル配信・解析などが実現できる。動きの計測には、全身関節の動きを計測するモーションキャプチャシステムが従来の主流であった。しかし、このシステムでは、マーカや機器を身体またはタイトなスーツに装着するため、手間がかかり、かつ動きに少なからず影響が出てしまう。更に本質的な問題として、伝統芸能などで表現されるものは演技者の関節の動きだけではなく、衣装やその動き、顔の表情や化粧など多岐にわたる。

これら全てを計測できる手段として、3次元ビデオ生成がある[1]。3次元ビデオでは、

- (1) 複数台のカメラで対象を多視点撮影、

- (2) 多視点撮影画像から3次元形状を復元[2],

- (3) 形状表面に撮影画像をマッピング[3],

して、対象の形状や見えを3次元保存できる。

3次元ビデオの解像度は、撮影画像中の対象像の解像度に大きな影響を受ける。各カメラにおける解像度の向上には、従来の全身撮影(図1(a))ではなく一部分をズームアップ撮影(図1(b))すればよい。しかし、視野が狭まるズームアップ撮影では、動き回る対象全身の継続的な計測は難しい。従来法[1]の拡張で高解像度・広範囲撮影を同時に実現するには、超多数のズームアップカメラを広範囲に敷き詰める必要がある。しかし、こうした超多数カメラシステムは金銭



(a) 全身撮影 (b) ズームアップ撮影

図1 対象の撮影例

Fig.1 Full-body and zoom-in capture.

[†] 奈良先端科学技術大学院大学, 生駒市

Nara Institute of Science and Technology, Ikoma-shi, 630-0192 Japan

a) E-mail: ukita@is.naist.jp

* 本論文はシステム開発論文である。

的・人的コストに加えて、舞台周辺に設置可能なカメラの台数を超えるため、現実的ではない。

本研究では、固定カメラ群と比較して格段に少ないパンチルトカメラで対象を継続的にズームアップ撮影可能であることに注目し、「安定に広域撮影可能なズームアウト固定カメラ」と「対象を詳細撮影可能なズームアップパン・チルトカメラ」の組み合わせにより、3次元ビデオの高精細化に有効な画像を取得するシステムを提案する。パンチルトカメラでは、「各カメラが各瞬間、対象の位置・形状・姿勢に応じてどこを撮影するか決定する」という新しい課題が発生する。この撮影戦略が適切でないと、ズームアップした画像にはモーションブラー、フレームアウトによる対象像の欠損、遮蔽などが発生し、これらの画像は3次元ビデオの高精細化に有効ではない。そこで、3次元ビデオの品質向上に貢献する画像撮影を本研究の目的とし、その撮影戦略をカメラワーク問題として定義する。本研究では、図1のような和装にも適用可能な汎用性の高いシステムを実現する。

上記問題設定では、以下の課題が重要となる。

- パンチルトカメラ群を対象の動作に応じて実時間制御しなければならないため、カメラワークの決定には実時間処理が要求される。
- 任意の身体動作の計測や予測には計算時間がかかる。よって、身体の動きに応じた実時間カメラワーク決定が難しい。

提案システムでは、対象を複数のパーツ（頭、胴など）に分割し、事前計測した動きにおける全身形状中のパーツ検出結果を記録しておく。オンライン撮影時には、この記録データを参照することによって、対象形状から各パーツ位置の高速検索を可能にし、このパーツ位置に基づいてカメラワークを行う。各パンチルトカメラの視線方向候補を少数のパーツ重心に絞り込むことで、カメラワークの解空間を小さくし、決定を実時間化する。また、形状検索の頑健・高速化のため、高次元の形状データを低次元特徴量化する。

本研究は、オンライン撮影システムに焦点を当てており、撮影後に十分に計算時間をかけて行う高精細3次元ビデオ生成には従来法をそのまま利用する。

2. 関連研究

3次元形状の復元法には、多視点画像からの復元、Kinectなどのデプスセンサによる復元などが挙げられる。高精細なデプスセンサは、いまだ撮影速度が

遅いため動物体の撮影には使用できず、高速撮影可能なものはいまだ低解像度である。

多視点カメラによる形状復元には、対象シルエットからの視体積交差法[2]、視体積交差法の誤差を多視点間の色整合性に基づいて低減するspace carving[4]、色情報のみによるステレオ視[5]が広く用いられている。視体積交差法は安定かつ高速である一方、復元形状は真の形状を内包した近似形状である。Space carvingやステレオ視は低速だが、多視点画像間で対応付け可能な色情報が豊富であれば正確な復元が可能である。これらの手法の統合による高精度化も可能だが[6],[7]、これも実時間処理は困難である。よって、実時間形状復元には視体積交差法が現実的である。

移動対象の高解像度撮影法には、パンチルトカメラによるズームアップ撮影がある[8]。これにより、高解像度撮影が可能だが、全カメラで全身を撮影した場合と比べて対象の各部を同時に撮影するカメラ数が減ってしまう。撮影カメラ数の減少は、形状復元誤差の増加につながる。この形状復元誤差の低減と高解像度テクスチャの取得をともに満たすカメラワークにおいて重要なパラメータは、パン角、チルト角、ズームの三つに代表される。本研究では問題の簡略化のためズームを固定とし、パン角とチルト角のみを決定する。

従来のカメラワークの決定には、形状復元の誤差低減を基準とした研究例が多い（視体積交差法の誤差低減[9]、ステレオ視の誤差低減[10]）。適切なカメラワークは対象の姿勢や形状に応じて変化するため、その決定は形状変化に応じて実時間で行う必要がある。しかし、文献[9]では、人体の関節剛体モデルを用いるため、和装のような着衣の大きな形状変化に対処できず、計算時間も数秒から数時間を要する。文献[10]では、剛体モデルを用いていないため和装対象にも適用可能だが、これも実時間処理には及ばない。

和装のような非剛体対象の実時間3次元形状解析の手法には、対象の動作が既知であることを前提とした検索ベースの手法[11]がある。この手法は、タイトな服装の人体のような剛体関節物体にも、非剛体にも適用できることが特徴である。本研究では、この手法に基づいて、カメラワーク決定に必要な3次元形状解析を実時間でを行う。

3. システム概要

本システムでは撮影する動作が既知であることを前提に、事前に対象の時系列ポリュームを計測し、その

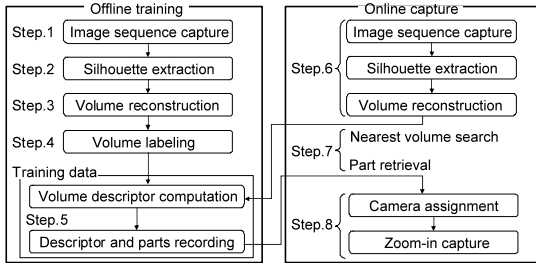


図 2 システムの処理概要
Fig. 2 Process overview of the system.

解析結果を記録しておく。これにより、本番撮影時には、計算コストの大きい形状解析を記録データからの検索・推定に代替することで、実時間でカメラワークを決定する。既知の動作であっても、万人が常時全く等しい動きを繰り返すことは不可能だが、本システムが利用する手法 [11] と形状記述子 [12] の利用によって、そのように個人内・個人間で変化する動きに対応した人体形状検索が実現する。

3.1 処理概要

処理全体を図 2 に示す。撮影に用いるカメラは、広角レンズによるズームアウト撮影のための固定カメラと、パンチルト雲台に載せたズームアップ撮影のためのパンチルトカメラの 2 種類を用いる。

3.1.1 オフライン計測, 解析, 及び記録

Step.1 多視点からの同期映像撮影：全固定カメラにより、対象を同期撮影する。

Step.2 対象のシルエット抽出：背景差分により対象のシルエットを抽出する。

Step.3 視体積交差法によるポリウム復元：各カメラの投影中心（図 3 の P_A , P_B ）から対象（図 3 の C ）のシルエット（図 3 の D_A , D_B ）を 3 次元空間に逆投影して（図 3 の V_A , V_B ）、それらの積領域（図 3 の赤い太枠）を対象のポリウムとする。ポリウムは、ボクセルと呼ばれる空間を格子分割した小立方体の集合で表現される。このポリウム中、図 3 の対象側面に存在する多角領域は全て復元誤差である。

Step.4 パーツの 3 次元重心の推定：全身ボクセルを各パーツ領域に分割することを、ラベリングと呼ぶ。パーツの種類や分割数は、パンチルトカメラの台数、パーツの大きさ、パーツの見えの重要性を考慮して、頭、胴体、袖を含む右腕と左腕、脚の五つとした。各ラベリングボクセルの重心を求めることにより、各パーツの重心が求まる。このオフライン処理時の自動ラベリングの詳細は 4. で述べる。

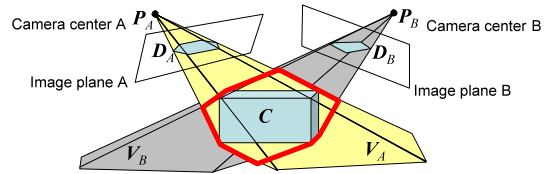


図 3 視体積交差法。赤線で囲まれた領域が復元領域
Fig. 3 Volume intersection method.

Step.5 ポリウムの低次元化とデータの記録：オフライン撮影で得られた時系列ポリウムに、Step.4 で得られたパーツの重心を関連付けて、撮影時の高速パーツ位置検索のための記録データを得る。しかし、Step.3 で得た高次元なポリウムデータは実時間検索に不向きであるため、ポリウムを低次元な形状記述子に変換する。形状記述子の詳細は 5. で述べる。

Step.6 高精度形状復元：オンライン撮影時に必要となる高精度復元形状を別途求め、Step.5 までに求めた検索用データと対応付けて記録しておく。高精度形状復元には、手法 [6] を用いた。

3.1.2 オンライン撮影

Step.7 対象の 3 次元形状の復元：高速形状検索のため、固定カメラの撮影画像のみを用いてオフライン処理の Step.1,2,3 と同じ処理で対象のポリウムを得る。

Step.8 記録データからパーツ重心検索：形状記述子によって Step.7 のポリウムデータを低次元化し、記録データから類似ポリウムを検索する。本システムでは、毎フレームにおいて全検索を行った^(注1)。類似ポリウム中の各パーツ重心を参照し、撮影対象の位置と鉛直軸周りの回転に応じて、各パーツの位置を補正する。

Step.9 カメラワークの決定：各パーツ重心とパンチルトカメラ群の位置関係に基づいて、カメラワークを決定し、各パンチルトカメラを制御して各パーツをズームアップ撮影する。詳細は、6. で述べる。

3.2 システムハードウェアと撮影・制御処理

ハードウェア構成を図 4 に示す。パンチルトカメラ、] その撮影に同期したパンチルト角を得るエンコーダ、全カメラを同期撮影させる同期ユニット、及び撮影画像を取得するノード PC を 1 組とし、全ての組を

(注1)：検索時に、前フレームの検索結果の近傍のみを検索するように追跡を取り入れることにより、検索の大幅な高速化が可能である。しかし、安定な追跡のためには、いったん追跡失敗した際の正しい追跡回復が必須であり、これに失敗すると、適切なパーツ撮影は不可能である。本システムでは、実験で示すように検索そのものはビデオレート以上に高速であるため、検索の安定性を重視して毎フレーム全検索を採用した。

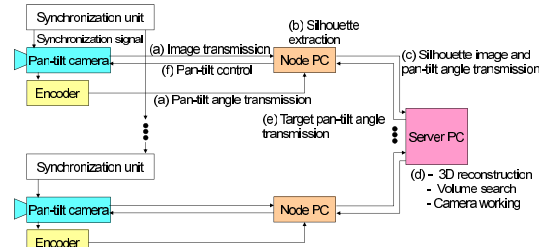


図 4 システムのハードウェア構成とデータフロー
Fig. 4 Hardware architecture and data flow.

1 台のサーバ PC で制御する。全カメラはパンチルト雲台の上に設置されているが、カメラを固定カメラとして利用する際はパンチルトを固定し、単焦点距離のレンズを取り付ける。

オンライン撮影では、ノード PC とサーバ PC において下記処理を行う。

Step.1 後の 3 次元ビデオ生成のために、全撮影画像及びパンチルトカメラの角度はノード PC に保存される (図 4 (a))。固定カメラの撮影画像から抽出したシルエット画像 (図 4 (b)) はカメラワーク決定目的の形状復元のため、通信量を抑えるために画像サイズを 1/16 にして、パンチルトカメラの角度とあわせてサーバ PC へ送信される (図 4 (c))。

Step.2 サーバ PC は、受信した多視点シルエットから視体積交差法で高速に対象のボリュームを得る (図 4 (d))。先述のようにこのボリュームとパンチルト角に基づいてカメラワークを決定し、各ノード PC にパンチルトカメラの目標角を送信する (図 4 (e))。

Step.3 ノード PC は受信した目標角にカメラを制御し (図 4 (f))、目標部分をズームアップ撮影する。

4. 事前計測ボリュームデータのラベリング

事前計測時のみ対象の両手を図 5 (a) のように着色することで、撮影画像からそれらの領域を色検出によりラベリングする。また、頭部も Haar-like 特徴量 [15] により検出することでラベリングする (図 5 の (b))。また、視体積交差法を用いてボリュームを得る (図 5 の (c))。復元ボリューム表面上に撮影画像上のラベルを投影することにより、頭、左手、右手のボクセルをラベリングする。更に、全身ボリュームの重心ボクセルを胴体に、地面に最も近いボクセルを脚部に、それぞれラベリングする。最後に、未ラベリングのボクセルについて、最近傍のラベリング済みボクセル (ただし頭部ラベルを除く) を探索し、そのラベルを未ラベリ

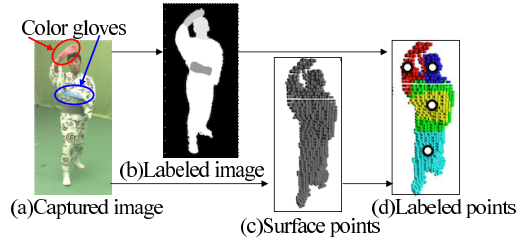


図 5 ボリュームデータのラベリング
Fig. 5 Labeling a volume.

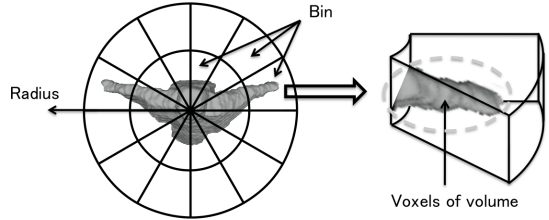


図 6 Bin の分割モデル。左図は高さ方向から見た Bin の分割の様子。右図は Bin の拡大図
Fig. 6 Binning model. left: top view. right: closeup.

ングボクセルに与えることで、全ボクセルをラベリングする (図 5 の (d))。ラベリングされた各パーツ領域の重心を計算し、全身ボリュームと関連づけ、記録データとする。

5. 形状記述子によるボリュームの低次元化

以下の手順で生成される形状記述子 [12] は、ヒストグラムベースのボリュームデータの表現である。

対象形状を包含する半径 R 、高さ H の円柱を、その中心軸が対象重心を通るように配置し、中心軸を基準軸とする。この円柱を、Bin と呼ばれる小領域群に分割する。Bin 内の復元ボクセルの個数を、この Bin のもつ値とする (図 6 の右)。文献 [12] では、高さ方向を n_h 、基準軸まわりの回転方向を n_θ にそれぞれ均等に分割していたが、本研究ではパーツの中で特に大きく動く右腕と左腕の変化をより詳細に表現するため、半径方向 n_r も均等分割する (図 6 の左)。Bin の総数は $N_{Bin} = n_r \cdot n_h \cdot n_\theta$ となり、この値が形状記述子の次元数となる。

形状記述子は、ノイズや形状の小さな変化に対し大きく変化せず、かつ、全ボクセル数と比較して低次元である。また、各 Bin 内の復元ボクセルは、「各 Bin 内のボクセルのみ 1、その他のボクセルは 0」というテーブルと復元ボクセルの AND 演算で高速検出できる。

6. カメラワーク決定

カメラワーク決定における大きな課題は、膨大なカメラワークの候補から実時間で適切なカメラワークを探索することである。10台のパンチルトカメラで5種類のパーツを撮影する例では、可能な組み合わせ総数は $5^{10} \simeq 9.7 \times 10^6$ と膨大である。

6.1 カメラワークの評価基準

本節で述べる各評価関数の値と重みは0から1の範囲の実数値をもち、値が高いほど適切であるとする。

6.1.1 形状復元誤差の評価

カメラワーク決定は実時間で行う必要がある一方、そのカメラワークで撮影された画像列からの3次元ビデオ生成では、多視点ステレオなど処理の遅い各種高精細な3次元復元法を活用できる。本研究では、この高精細3次元復元に視体積交差法とステレオ視を利用することを仮定し、これらの手法に適した画像撮影を目的とする。

視体積交差法では、複数カメラの視線ベクトル群が成す角度が直角に近いほど、形状復元誤差が小さくなりやすい[14]。しかし、提案システムにおけるパーツのズームアップ撮影では、撮影対象であるパーツの裏側に大きなパーツが存在し(例:腕を対象とすると、胴体はその奥にある)、必ずしも直角方向からの撮影が適切とは限らない。人体構造上、そのように視体積交差法を邪魔するパーツは胴体であり、もし直角からの撮影が適切でない場合、奥の180度方向からの撮影により、邪魔なパーツを回避できる可能性は高くなる。一方、ステレオ視では近い角度から対象を撮影するほうが、多視点間の点对応が容易である。そこで、2台のカメラ i, j の視線ベクトル $\mathbf{v}_i, \mathbf{v}_j$ が成す角度 $\theta_{i,j}$ が、0と $\frac{\pi}{2}$ と π 付近で評価値が高くなるように評価値 E_1 を定義する(図7)。

$$E_1 = |\cos 2\theta_{i,j}| \quad (1)$$

6.1.2 視体積交差法の誤差低減に有効な撮影方向

6.1.1では、「任意の形状」を対象に「撮影後に詳細形状復元」する際の誤差低減のための評価関数を与えた。本節では、「事前計測して記録済みの動きに伴う形状」を対象に「視体積交差法による形状復元」の誤差低減を目標とする。視体積交差法は、詳細形状復元だけでなく、オンライン撮影時のパーツ重心検索にも利用されるため、その誤差はシステムに影響する。

事前計測時に、高精度な復元手法で得られる対象形

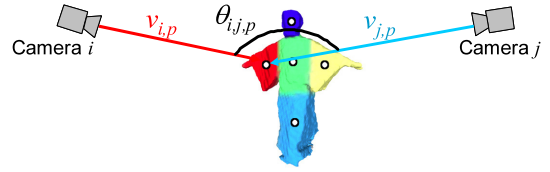


図7 形状復元誤差の評価
Fig. 7 Evaluating reconstruction error.

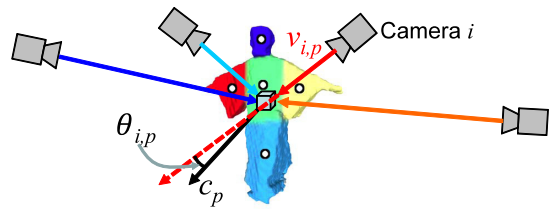


図8 形状復元誤差の低減に有効な撮影方向の評価
Fig. 8 Evaluating ghost volume carving.

状、及び視体積交差法のみで得られる形状を、それぞれ $\mathbf{S}^h, \mathbf{S}^l$ とする。 $\mathbf{S}^l - \mathbf{S}^h$ が視体積交差法の誤差であり、この領域を削ることのできるカメラワークが、オンライン撮影時に望まれる。そこで事前計測データの記録時に、この誤差を低減できる視線ベクトル \mathbf{c}_p を各パーツ p に対して以下の処理で求めておく。

(1) $\mathbf{S}^l - \mathbf{S}^h$ 中、パーツ p の重心の近傍ボクセル集合(実験では100ボクセル)を抽出する。以下、パーツ p をどの方位から撮影した際にこれらの抽出ボクセルを削減できるかを調べる。

(2) 各抽出ボクセルを \mathbf{S}^h に遮られないように削り取ることで得られる視線方位(誤差削減方位と呼ぶ)を求める。本システムでは、抽出ボクセルを中心にして均一サンプリングした方位に対してシルエット取得における遮蔽をチェックした。

(3) 全抽出ボクセルにおける全誤差削減方位の平均を \mathbf{c}_p とする。

オンライン撮影時は、図8の \mathbf{c}_p とパーツ p の重心を注目するカメラ i の視線方向 $\mathbf{v}_{i,p}$ の成す角 $\theta_{i,p}$ が0または π に近いほど評価値が高くなる(式(2))。

$$E_2 = |\cos \theta_{i,p}| \quad (2)$$

6.1.3 注目パーツの遮蔽回避

テクスチャ撮影には、パンチルトカメラとそれが注目するパーツの間を、他パーツが遮蔽していないことが重要である。図9に示すように、カメラ i から注目パーツ p への視線ベクトル $\mathbf{v}_{i,p}$ と、パーツ重心から胴体重心へ向かうベクトル \mathbf{v}_p が成す角 $\theta_{i,p}$ が π に近い

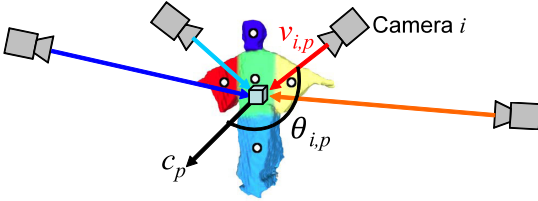


図9 注目パーツの遮蔽可能性の評価
Fig.9 Evaluating mutual occlusion.

ほど他パーツによる遮蔽の可能性が高く、0に近いほど低い。この評価関数 E_3 を式 (3) で定義する。

$$E_3 = \cos \frac{\theta_{i,p}}{2} \quad (3)$$

6.1.4 パンチルトカメラの動特性

視線変更に伴うパンチルト速度が速すぎると、撮影画像にはモーションブラーが発生してしまう。そこで、事前にモーションブラーが起きない最大回転速度 ω_{max} を調査しておく。撮影時は、カメラ i の視線方向を現在の $v_{i,p1}$ から次の目標の $v_{i,p2}$ へ変化させるときの差分角度 $\theta_{i,p1,p2}$ から予想到達時間 $t = \frac{\theta_{i,p1,p2}}{\omega_{max}}$ を算出し^(注2)、その時間が短いカメラとパーツの組み合わせを高く評価する。ただし、予想到達時 t がしきい値 t_{max} を超えた場合、評価値を0とする (式 (4))。

$$E_4 = \begin{cases} \frac{1}{\sqrt{2\pi\sigma^2}} \exp(-\frac{t^2}{2\sigma^2}), & 0 \leq t < t_{max} \\ 0, & t_{max} \leq t \end{cases} \quad (4)$$

t_{max} には、オンライン形状復元とカメラワークの1サイクルにかかる時間の最大値を実験的与える。

6.1.5 パーツの重要度の設定

3次元ビデオにおいて、より高精細に撮影したい人体パーツはユーザの目的によって異なる。本システムでは、パーツ p の重要度 I_p をユーザが任意に与えられるものとする。パーツの総数を n_P とすると、パーツの重要度の合計が1となるようにする (式 (5))。

$$\sum_{1}^{n_P} I_p = 1 \quad (5)$$

6.2 カメラワーク決定のアルゴリズム

カメラの台数を m 、パーツの種類を n として、前述の複数評価関数を組み合わせたカメラワーク決定のアルゴリズムを以下に示す。

(注2)：提案システムで利用したパンチルト雲台は、回転速度命令を与えてからその速度に達するまでの時間が非常に短いため、予想到達時刻をこのように近似的に与えた。

Step.1 評価関数 E_1, E_2, E_3, E_4 の重みを $I_E = [I_{E1}, I_{E2}, I_{E3}, I_{E4}]$ 、各パーツの重みを $I_P = [I_{p1}, \dots, I_{pn}]$ とする。

Step.2 全パンチルトカメラと全パーツの組み合わせで評価関数 E_1, E_2, E_3, E_4 を計算し、 E_1 の全値を記録した $m \times m \times n$ の評価表 T_1 と、 E_2, E_3, E_4 の全値を記録した $m \times n$ の評価表 T_2, T_3, T_4 を得る。

Step.3-a 割当済みカメラがない場合。評価表 T_2, T_3, T_4 の重み和として総合評価表 \bar{T} を得る。

$$\bar{T} = (I_{E2}T_2 + I_{E3}T_3 + I_{E4}T_4)\text{diag}(I_P)$$

Step.3-b 割当済みカメラ数が $\hat{m} (> 0)$ の場合。評価表 T_1 から、割当済みカメラに対応する \hat{m} 個の $m \times n$ の評価表 $T'_{1,1}, \dots, T'_{1,\hat{m}}$ を抽出し、以下の総合評価表 \bar{T}_i (ただし $i \in \{1, \dots, \hat{m}\}$) を得る。

$$\bar{T}_i = (I_{E1}T'_{1,i} + I_{E2}T_2 + I_{E3}T_3 + I_{E4}T_4)\text{diag}(I_P)$$

Step.4 総合評価表 \bar{T} または $\bar{T}_1, \dots, \bar{T}_{\hat{m}}$ から、最も評価値の高いカメラ c とパーツ p の割り当てを選ぶ。

Step.5 評価表 T_1, T_2, T_3, T_4 中で、割当済みカメラ c に対応する評価値を全て0とする。

Step.6 未割り当てカメラがあれば、Step.3-bに戻る。

7. 実験

以下の実験を行った。

実験1 人体パーツの特定が難しい和装を対象とした類似形状の検索と、カメラワークの決定。

実験2 汎用性を示すため、実験1とは異なる対象 (一般的なタイトな服装、かつ、道具 (テニスラケット) をもった人間) の3次元ビデオを生成。

7.1 実験システム

合計14台のカメラを図10のように配置した。各カメラは 1024×768 pixels の画像を30fpsで取得する。カメラをパンチルト制御するための雲台には、Directed Perception社 PTU-46を用いた。この雲台の角解像度は、約0.013度である。

各カメラは事前に[13],[16]の手法でキャリブレーションを行い、カメラの内部パラメータ、及び外部パラメータを推定した。固定カメラにおける外部パラメータは、固定角度において推定する。一方パンチルトカメラでは、複数のパンチルト角 (実験では各カメラ16方位) において外部キャリブレーションを行った。オンライン撮影時には、決定カメラワークのパン

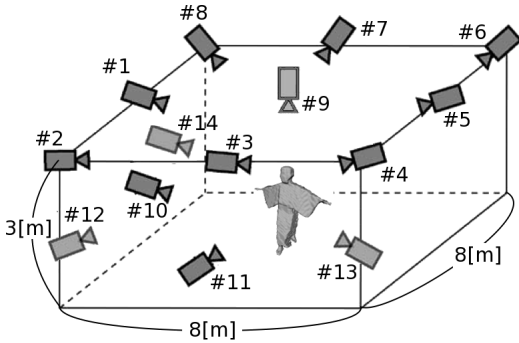


図 10 撮影カメラの配置. #はカメラの番号を表す
Fig. 10 Camera configuration.

チルト角に最近傍のキャリブレーション済み角度を選択し、その方向に視線制御^(注3)した。このキャリブレーションパラメータに基づき、オンライン撮影中の形状復元、及び撮影後の詳細な3次元ビデオ生成を行った。

7.2 実験に用いたパラメータ

ボリュームデータのパラメータ： 視体積交差法の復元領域は、幅 1.5m、奥行き 1.0m、高さ 2.0m、ボクセルサイズは 20mm とした。形状記述子の円柱領域は、半径 900mm、高さ 2800mm とし、高さ方向、回転方向、半径方向に対する分割数はそれぞれ 35, 12, 3 とした。形状記述子の次元数は 1260 である。カメラワーク決定における重み： 各評価関数に与える重み I_E は、0.1, 0.2, 0.2, 0.5 とした。

7.3 実験結果

7.3.1 形状検索とカメラワーク決定

8 台のパンチルトカメラ（焦点距離は 35mm または 25mm）と 6 台の固定カメラ（焦点距離は 8mm, 6mm, または 4mm）を用いた。これらのレンズの画素換算焦点距離は、それぞれ約 7900, 5500, 1800, 1300, 870 である。オンライン撮影で得られた画像とボリューム、及び検索されたラベリング済みのボリュームを図 11 に示す。また、検索に要した処理時間を、ボクセルで表現したボリュームデータと形状記述子でそれぞれ行った結果について表 1 に示す。

図 11 の結果から、和装対象の類似形状の検索、及びボリュームラベリングによるパーツ分割とパーツ重心の推定に成功していることが確認された。また表 1

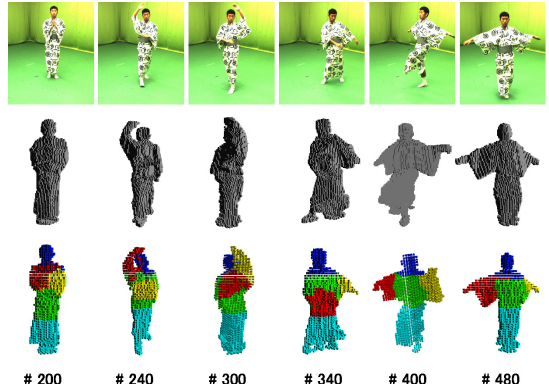


図 11 記録データからの検索とラベリングの結果. 上段から順に、観測画像、形状復元結果、記録データから検索したボリュームのラベリング結果

Fig. 11 Results of volume retrieval and labeling. Top: Images. Middle: Reconstructed volumes, Bottom: Labeling results.

表 1 ボリュームデータの検索速度
Table 1 Computational time of volume retrieval.

Volume representation	Search speed [msec/frame]
Voxel based	140
Volume descriptor	30

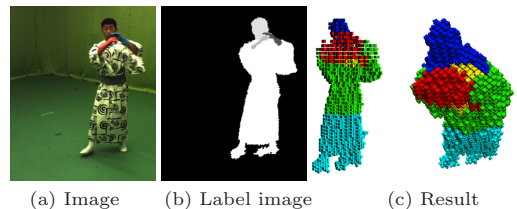


図 12 ラベリングの失敗例

Fig. 12 Unsuccessful labeling results.

から、ボクセル表現の検索速度と比較して、形状記述子では検索速度が約 4.6 倍まで向上したことが分かる。本実験では、記録フレームの全探索（591 フレーム）を行ったが、KD 木や安定な追跡手法などの応用により、更なる高速化が可能である。

ラベリングは多くのフレームで成功したが、図 12 のように左右の手が互いに近いシーンでは、両手の形状が結合したことによるラベリングの失敗例が見られた。このような問題を回避するためには、各フレームのラベリングにおいて、数フレーム前のラベリング結果を参照するような時系列最適化などが挙げられる。

図 11 に示したような検索結果からカメラワークを決定し、パンチルト制御を行った際の撮影画像例を図 13 に示す。この例では、パーツの重要度 I_P は、

(注3)：前述のように、パンチルト雲台の角解像度は 0.013 度である。仮にその 10 倍の誤差が実際の回転角度にのったとしても、カメラから約 300cm の距離において 1cm 以下の誤差で目標座標にカメラの視線を制御可能である。これは、外部パラメータ推定の精度と比べて十分小さいため、カメラ制御の誤差が実際の形状復元に与える影響は小さい。

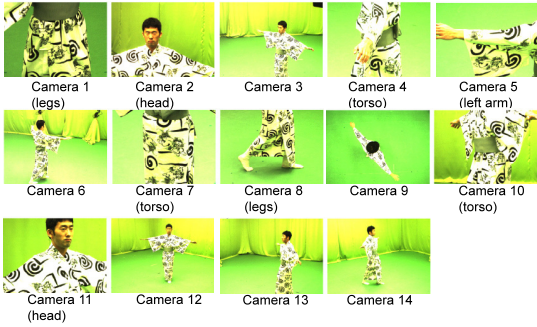


図 13 提案システムで撮影された画像例。括弧内はズームアップパンチルトカメラの撮影対象パーツ名
Fig. 13 Examples of observed images.

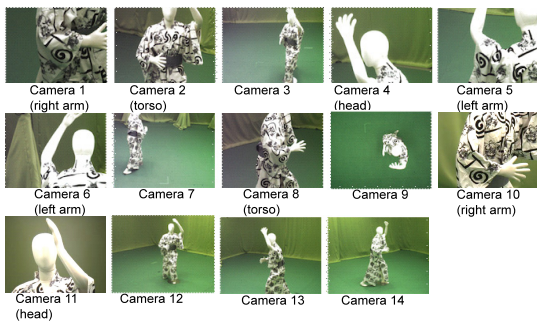


図 14 提案システムによるマネキン撮影画像例
Fig. 14 Sample images of a stationary object captured by the proposed system.

頭、左腕、右腕、胴体、脚の順に、0.3, 0.1, 0.1, 0.3, 0.2 とした。頭と胴の重みを大きく与えていたため、この 2 パーツに多くのパンチルトカメラが割り当てられていることが確認できる。

図 14 と図 15 は、それぞれ提案システムと固定カメラのみの従来システムによる撮影画像例である。条件を等しくするため、撮影対象を完全に同姿勢のマネキンにした。マネキンは先の被験者と同じ和装であり、姿勢は記録済みの人の動きの中の一姿勢である。

図 16 に、提案システムによって図 14 に示したカメラワークが与えられた際のカメラの動きを示す。図中、矢印左側がパンチルトカメラの初期角における撮影画像で、矢印右側が与えられたカメラワークに従ってカメラ制御した後の撮影画像である。初期角は、手動でおおまかに対象重心に向けられていたが、視野内に対象を収めることができている画像も存在する。一方、提案システムのカメラ制御により、正しく対象パーツを視野内に収めることができている。

図 14 と図 15 の撮影画像での、対象位置における

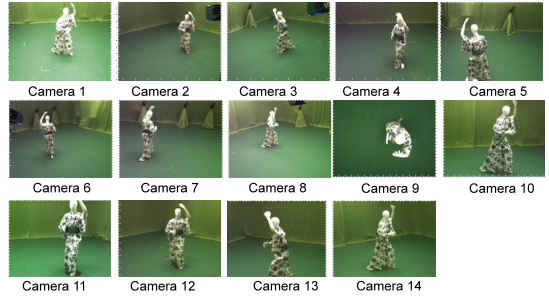


図 15 固定カメラのみによるマネキン撮影画像例
Fig. 15 Sample images of a stationary object captured by stationary cameras.

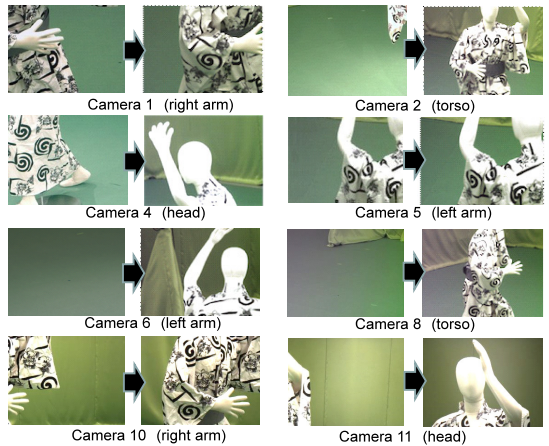


図 16 提案システムによるカメラ制御例
Fig. 16 Camera working by the proposed system.

1 画素に対応する撮影物体の大きさ（理論空間解像度）を計算した。従来システムでは全身 2.5mm/voxel であるのに対して、提案システムでズームアップ撮影された頭では 0.7mm/voxel となり、提案システムでの理論空間解像度の向上も確認できる。

7.3.2 3次元ビデオ生成

固定カメラとパンチルトカメラの台数を、それぞれ 10 台（図 10 の #1 から #9 と #14）、4 台（図 10 の #10 から #13）とした。パーツとしてテニスラケット（Haar-like 特徴で検出）を加え、パーツの重要度 I_P は頭とラケットの重みを 0.5 とした。

提案システムの撮影画像を図 17 に、そこから生成された 3 次元ビデオの 1 フレームを図 18 に示す。高精細形状復元には、視体積交差法とステレオ視による統合手法 [6] を用いた。オンライン撮影時は固定カメラの撮影画像のみを形状復元に利用するが、撮影後の高精度形状復元では、全てのカメラの撮影画像を利用す



図 17 図 18 の復元に利用された撮影画像

Fig. 17 14 captured images used for generating the result in Fig. 18.

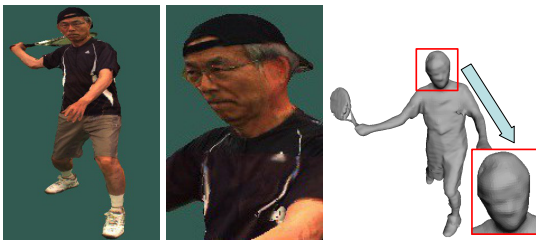


図 18 提案システムによる 3 次元ビデオ生成

Fig. 18 Textured and non-textured 3D mesh obtained by the proposed system.

る。ただし、パンチルト制御中の撮影画像はモーションプラーがひどいため、形状復元には利用しない。表面メッシュ復元には手法 [17] を、テクスチャマッピングには OpenGL を用いた。比較のため、全 14 台のカメラに広角レンズを装着し、パンチルト固定して撮影する従来システムによる生成 3 次元ビデオの 1 フレームも図 19 に示す。

提案システムでは、頭部とラケットの重要度は等しく、ベストなカメラワークでは頭部 2 台、ラケット 2 台のカメラ割り当てになるが、図 17 のように、不均等な割り当てになる瞬間もある。これは、各パーツの重要度以上にカメラワークの評価基準 E_1 , E_2 , E_3 , E_4 が大きな役割果たした瞬間や、各カメラの撮影対象パーツが切り変わった瞬間に起こりうる。

図 18 と図 19 に、それぞれ提案システムと従来の固定カメラのみのシステムによる撮影画像から形状復元・テクスチャマッピングされた 3 次元ビデオの 1 フレームを示す。両結果ともに、14 台のカメラによる撮影画像から生成されている。撮影対象は、両システムによる撮影時に完全に同じ動きを繰り返すことは不



図 19 固定カメラのみのシステムによる 3 次元ビデオ生成

Fig. 19 Textured and non-textured 3D mesh obtained by the previous system only with stationary cameras.

表 2 オンライン撮影における処理時間

Table 2 Computational costs in online capturing.

Phase	processing speed [msec]
Reconstruct volume	180
Retrieve from training data	20
Decide camera assignment	10
Drive pan-tilt cameras	300

可能であり、また、両システムの撮影タイミングと対象動作を同期することも不可能であるため、図 18 と図 19 の対象形状は多少異なることに注意されたい。この図 18 と図 19 を比較すると、提案システムのほうが表情まで確認できる高解像度撮影ができてることが確認できる。図 18 と図 19 には、テクスチャ無しの形状も示しており、この頭部形状からも、提案システムのほうが目のへこみや鼻のふくらみなどの顔パーツの形状まで滑らかに形状復元できていることも確認できる。

提案システムでのオンライン撮影における各処理時間を、表 2 に示す。類似形状の検索とカメラワーク決定に要した時間は合計で 30msec となった。形状復元の時間と比較すると十分に高速ではあるが、最高性能としてビデオレート (30fps) を目指す場合、形状復元同様に更なる高速化が必要である。

8. む す び

パンチルトカメラによるズームアップ撮影システムを提案した。実験で、類似形状の高速かつ頑健な検索、及び高速決定したカメラワークからの部分高解像度画像の撮影と 3 次元ビデオ生成を確認した。

今後の課題を以下にまとめる。

カメラワークのオフライン最適化：事前計測データの解析時に、代表的な対象形状におけるカメラワークを求めておき、オンラインカメラワークの指標とする

ことで、カメラワークの決定時間を短縮する。
任意パンチルト角における3次元ビデオ生成：本実験では、事前に設定した幾つかのパンチルト角にのみ視線を制御したが、手法[8]によって、任意パンチルト角における外部パラメータ推定が可能になる。

文 献

- [1] T. Matsuyama, X. Wu, T. Takai, S. Nobuhara, "Real-time 3D shape reconstruction, dynamic 3D mesh deformation, and high fidelity visualization for 3D video," *Computer Vision and Image Understanding*, vol.96, no.3, pp.393-434, 2004.
- [2] P. Srinivasan, P. Liang, and S. Hackwood, "Computational geometric methods in volumetric intersection for 3D reconstruction," *Pattern Recognition*, vol.23, no.8, pp.843-857, 1990.
- [3] P.V. Sander, J. Snyder, S.J. Gortler, and H. Hoppe, "Texture mapping progressive meshes," *ACM Transaction on Computer Graphics*, vol.2001, pp.409-416, 2001.
- [4] S.M. Seitz and K.N. Kutulakos, "Plenoptic Image Editing," *ICCV*, 1998.
- [5] S.M. Seitz, B. Curless, J. Diebel, D. Scharstein, and R. Szeliski, "A comparison and evaluation of multi-view stereo reconstruction algorithms," *Computer Vision and Pattern Recognition*, vol.1, pp.519-528, 2006.
- [6] N. Ukita and K. Matsuda. "3D reconstruction with globally-optimized point selection," *IEICE Trans. Inf. & Syst.*, vol.E95-D, no.12, pp.3069-3077, Dec. 2012.
- [7] C.H. Esteban and F. Schmitt, "Silhouette and stereo fusion for 3D object modeling," *Computer Vision and Image Understanding*, vol.96, no.3, pp.367-392, 2004.
- [8] 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, June 1998.
- [9] 西山 学, 牧 淳人, 松山隆司, "高精細 3次元形状復元のための多視点カメラワーク最適化," *Visual Computingグラフィクスと CAD 合同シンポジウム*, 2004.
- [10] S. Yous, N. Ukita, and M. Kidode, "Multiple active camera assignment for high fidelity 3D video," *International Conference on Computer Vision Systems*, 2006.
- [11] N. Ukita, R. Tsuji, and M. Kidode, "Real-time shape analysis of a human body in clothing using time-series part-labeled volumes," *ECCV*, 2008.
- [12] N. Ukita, M. Hirai, and M. Kidode, "Complex volume and pose tracking with probabilistic dynamical models and visual hull constraints," *ICCV*, 2009.
- [13] T. Wada, X. Wu, S. Tokai, and T. Matsuyama, "Homography based parallel volume intersection: Toward real-time volume reconstruction using active cameras," *Workshop on Computer Architectures for Machine Perception*, pp.331-339, 2000.
- [14] A. Theetten, J.-P. Vandeborre, and M. Daoudi, "Determining characteristic views of a 3D object by visual hulls and Hausdorff distance," *3DIM*, 2005.
- [15] R. Lienhart and J. Maydt, "An extended set of haar-like features for rapid object detection," *ICIP*, 2002.
- [16] Z. Zhang, "A flexible new technique for camera calibration," *Pattern Analysis and Machine Intelligence*, vol.22, no.11, pp.1330-1334, 2000.
- [17] M. Kazhdan, M. Bolitho, and H. Hoppe, "Poisson surface reconstruction," *SGP*, 2005.

(平成 25 年 1 月 8 日受付, 5 月 15 日再受付)



藤根 成暢

2012 年奈良先端科学技術大学院大学修士課程修了。現在、みずほ情報総研株式会社勤務。在学中、高精細 3次元ビデオ生成に関する研究に従事。



浮田 宗伯 (正員：シニア会員)

2001 年、京都大学大学院博士後期課程修了。同年奈良先端科学技術大学院大学情報科学研究科助手。2007 年同准教授。2002 年～2006 年まで科学技術振興機構さきがけ(「情報基盤と利用環境」領域)研究員兼任。2007 年～2009 年までカーネギーメロン大学客員研究員兼任。2011 年より ATR 客員研究員を兼任。博士(情報学)。コンピュータビジョン、分散協調視覚、対象追跡、人体運動解析・姿勢推定に関する研究に従事。1999 年電子情報通信学会論文賞。



萩田 紀博 (正員：フェロー)

1978 年慶應義塾大学大学院工学研究科電気工学専攻修士課程修了。同年電電公社(現 NTT)武蔵野電気通信研究所入所。NTT 基礎研究所, ATR メディア情報科学研究所長などを経て、現在 ATR 知能ロボティクス研究所長。この間、文字認識、画像認識、コミュニケーション科学、コミュニケーションロボットなどの研究に従事。工学博士。IEEE, 電子情報通信学会, 情報処理学会, 日本ロボット学会, 人工知能学会各会員。