

3次元ビデオによる人体3次元計測とその応用

延原 章平^{1,a)}

概要: 3次元ビデオと呼ばれる多視点映像を入力とした人体の3次元形状・運動計測は、その非接触・非拘束な特徴から無形文化財のデジタルアーカイブを始めとして、マーカレスモーションキャプチャシステムとしての利用や、全周囲自由視点立体映像生成への応用など、様々な場面でより実用的な技術が提案されてきており、まさに研究室レベルから実応用へと研究開発の段階が進もうとしている。本稿では3次元ビデオの歴史とその基本的なアルゴリズムを紹介し、様々な応用例とともに今後の研究課題について述べる。

1. 3次元ビデオとは

3次元ビデオ (3D video) とは、シーンの全周囲3次元形状・表面テクスチャを記録し、自由な視点から立体提示することができる映像メディアである (図1)。このような意味での3次元ビデオは、1997年に Moezzi ら [1] と Kanade ら [2] によって提案され、その後多くのグループによって研究・開発が続けられている [3][4][5][6][7][8][9][10][11][12]。

3次元ビデオの特徴は、対象の全周囲3次元表面形状と運動情報を陽に持つ点であり、この点において様々な「3次元」技術と対比することができる。

3D テレビ・映画 3D テレビおよび3D 映画が持つ映像情報はステレオ映像と呼ばれ、左右の目に視差の付いた映像を提示することによって視聴者自身の脳がもつ視覚機能に3次元シーンを理解させる点に特徴がある。左右の目に異なった映像を届けるためには偏光メガネや視差バリアなど何らかの仕掛けが必要となるが、映像そのものは通常のカメラで撮影したものと等価なため、映像品質が高い。しかし一方で、通常の映像が左右両眼用に用意されているだけであることから、自由な視点変更や、シーンの3次元情報に基づく編集などを加えることは困難である。

3D 深度・レンジセンサ 一般に3D 深度センサもしくはレンジセンサとして扱われるものには、視差推定に基づく方式と、光が往復する時間 (Time-of-Flight) の計測に基づく方式の2タイプに分類できる。視差推定に基づく方式は、いわゆる受動もしくは能動ステレオ法を1つのデバイスとして実装したものであり、近年では Microsoft Kinect [13] が広く使用されている。ま



図1 多視点映像 (上段) から生成された3次元ビデオによる自由視点映像生成 (下段)

た光が往復する時間 (もしくは位相差) を計測するセンサは、1点ずつ走査する LIDAR と、面全体を一度に計測する ToF カメラに分類され、特に ToF カメラは近年低価格化が進んでいる [14][15][16]。これらはカメラから見えている範囲、つまり対象の手前側のみをキャプチャするものであり、全周囲の3次元情報を1台で取得できるわけではない。つまり正確には 2.5D 情報を取得するためのものであるが、特に LIDAR は計測精度が高く、大型有形文化財のアーカイブ化などにも用いられている [17]。

3D モーションキャプチャ 3D モーションキャプチャは対象表面にマーカを取り付け、その動きを検出することによってマーカ群の運動を取得する技術である。例えばマーカを体節の各部位に取り付けることによって、骨格構造の運動を求め、これを 3D CG キャラクターの振り付けへと使用する例が有名である。

3D CG アニメーション 1990年代から広く使用されている 3D CG アニメーションでは、その形状・運動情報がアニメーターによって制作されるのに対して、3次元ビデオでは被写体の形状・運動がそのまま取得され

¹ 京都大学大学院情報学研究科
京都市左京区吉田本町, 606-8501

^{a)} nob@i.kyoto-u.ac.jp

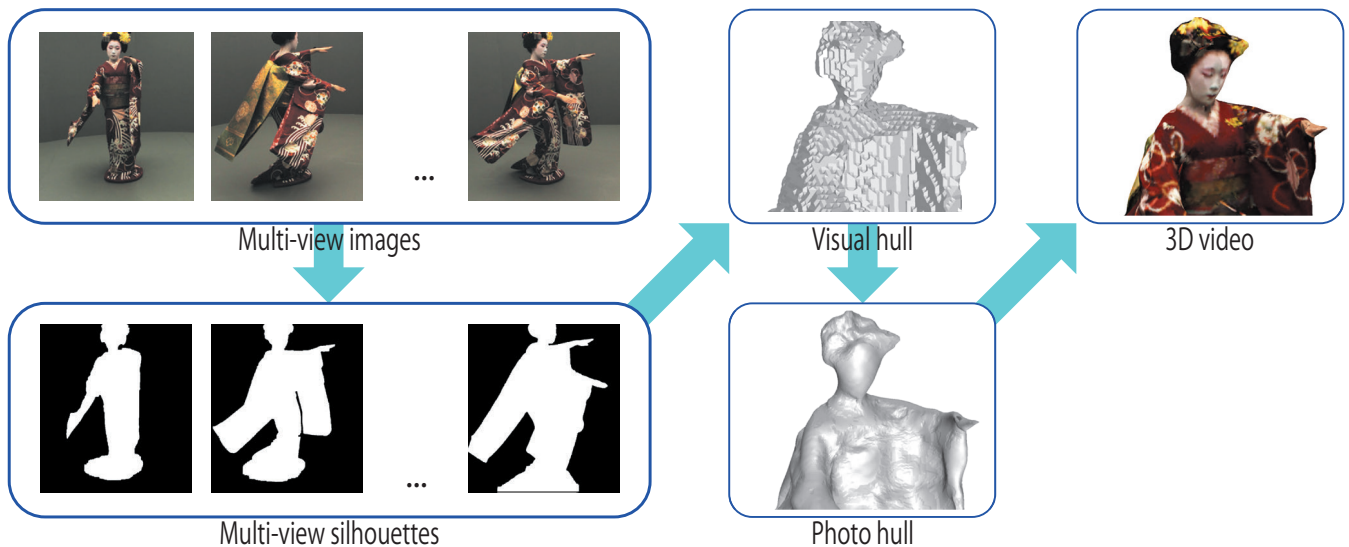


図 2 3次元ビデオ生成の流れ

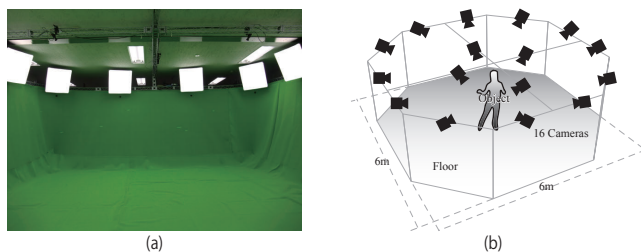


図 3 可搬式3次元ビデオ撮影スタジオ [20] の (a) 外観と (b) カメラ配置図

る点異なる。

自由視点 TV 自由視点 TV (Free-viewpoint TV, FTV)[18][19]に関する研究と3次元ビデオは、多視点カメラ映像を入力として自由視点映像を生成するという観点で共通点が多いが、3次元ビデオが3次元形状情報を陽に求めることで、それを基にした運動推定などを可能とするのに対して、FTVでは自由視点映像生成のために3次元形状情報が間接的に用いられ、陽には計算することを重視しない点に特徴がある。

2. 3次元ビデオの生成プロセス

3次元ビデオ生成の基本的な流れを図2に示す。まず3次元ビデオ生成の入力となる多視点映像から対象画像領域(シルエット)を抽出し、これを基にしてvisual hullを得る。次いでテクスチャ情報が最も整合する(photo-consistencyが最も高い)形状としてphoto hullを推定した後に、これの表面テクスチャを生成すれば、自由な視点から映像化できる3次元ビデオが生成される。以降では各ステップについて、その概略を述べる。

2.1 多視点映像撮影

3次元ビデオ生成に用いる多視点映像は、図3のように

多視点カメラ群を備えたスタジオで撮影される [21][20] [3].

ここで3次元ビデオ撮影が可能となる範囲は、各カメラの画角と被写界深度の共通領域として決定され、各カメラの撮影可能範囲よりも本質的に狭くなる。そのため解像度を犠牲にせずにも広範囲撮影を実現するには、単純にカメラ台数を増やして異なる空間領域を担当させるか、もしくはパンチルト可能な能動カメラ群によって追跡撮影を行う必要がある。

しかし通常の物体追跡撮影と異なり、3次元ビデオ生成のためには

- (1) 各時刻で対象物体を全ての方向から十分な解像度で撮影できること
- (2) 各撮影時刻でのカメラの内部・外部パラメータが十分な精度で得られること

を保証するように、常に各カメラを制御し続けなくてはならない。これを実現するための手法として、筆者らはセル方式と呼ぶコンセプトに基づいて、ターゲットとする撮影空間内を自由に運動する被写体の3次元ビデオ撮影を実現するアルゴリズムを提案している [22][3].

2.2 対象領域抽出

対象領域抽出は何らかの事前知識に基づいて撮影画像を対象(前景)領域と背景領域に分割する処理であり、シルエット抽出、あるいは(alpha) matting などとも呼ばれる。このうち matting と呼んだ場合は特に前景と背景の境界部分において画素の α 値、すなわち前景と背景の混合率をも推定することを目的とすることが多い [23]. 一方でシルエット抽出と呼んだ場合は α 値が0と1の二値として問題を単純化している場合が多い。

対象領域抽出に用いられる具体的な事前知識としては背景差分やクロマキーが代表的である。背景差分とは、まず

対象が存在しない状態で事前に画像を撮影しておき、実際に対象が存在する状態での画像との差を検出することに基づく手法である。すなわち一定値以上の差分が存在すれば前景、そうでなければ背景とする。一方クロマキーでは、背景色をブルー、グリーン、グレー、ホワイトなどなんらかの決まった色域で統一し (図 3(a)), 撮影画像の各画素値がその色域に近ければ背景、遠ければ前景と見なす手法である*1。

しかしいずれの手法においても、各画素独立に前景・背景の判断を行ってしまつてはノイズに弱く頑健な対象領域抽出を実現することは難しい。そのため隣接する画素同士での前景・背景への分類に連続性を持たせる手法が一般に使用され、特に前景・背景の二値に分類する問題は、min-cut/max-flow 問題に変換して厳密解を求めるようにアルゴリズム化されることが多い [24][25]。

また上記の背景差分・クロマキーは画像 1 枚毎に独立に行われる一般的な手法であるが、3次元ビデオ撮影で使用する多視点画像は互いに関連のない画像の集合ではなく、同じ対象を異なる方向から撮影したものである。そのため何らかの形で視点間の関係を使用し、より頑健な対象領域を実現する手法も提案されている。

例えば同じ対象を異なる方向から撮影したという事実から、基本的には前景と背景の色ヒストグラムは視点が変わっても大きな変化が生じないことが期待できるとする。この仮定に基づき、ある視点である画素値を前景あるいは背景としたという判断結果が別の視点でも意味を持つと見なし、前景・背景の判断結果を視点間で伝搬させる手法が提案されている [26]。

さらに、各視点で独立に対象領域抽出をしてしまつては、その結果がある共通の3次元形状の投影像として矛盾が無いことを保証することができない。これは intersection consistency [27] と呼ばれ、直感的には各視点で得られたシルエットから視体積交差法によって visual hull を計算し、これを再投影した場合に元のシルエットと一致しなくてはならない、という制約である。このようにキャリブレーション情報を利用した多視点対象領域の同時推定法が多く提案されている [27][28][29][30][31][32][33][34]。

2.3 3次元形状復元

3次元ビデオにおける計算モデルは、本来図 4 のように多様な光学現象を含む世界である。これに対して3次元形状復元に際してはこれを単純化した図 5 のようなモデルが広く仮定される。この図において斜めの網掛けがされた矩形は単純化のために無視された光学現象を表し、横線の網掛けがされた円は既知とされたカメラキャリブレーションを表す。また灰色矢印は物理世界での生成過程 (順問題)

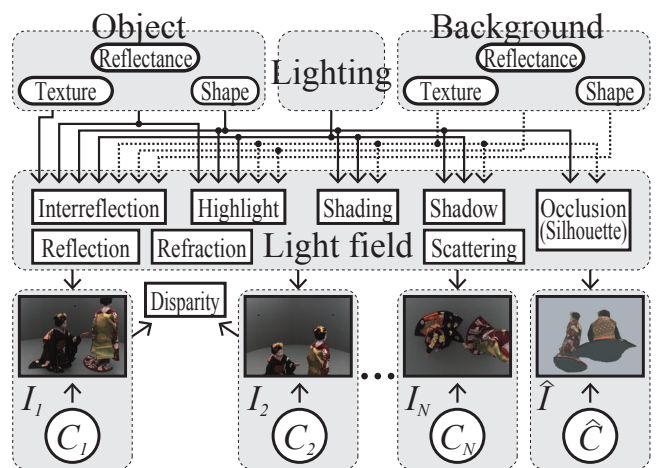


図 4 3次元ビデオの計算モデル [3]

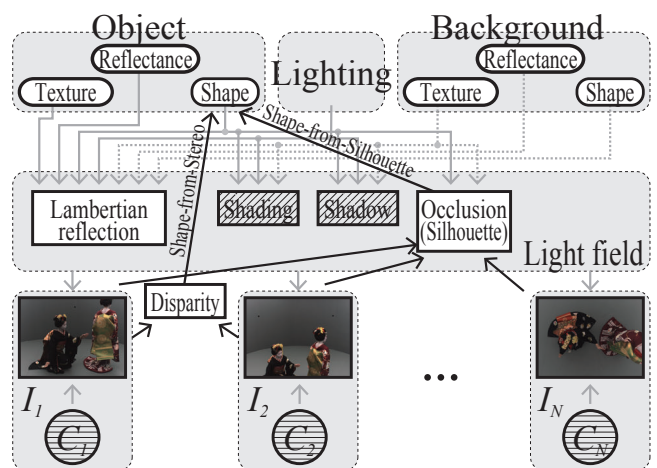


図 5 3次元形状復元の計算モデル [3]

を表し、そのうち破線は単純化のために無視される影響を表す。黒色実線は3次元形状復元として逆問題を解く過程を示している。すなわち陰影や陰、屈折や反射・散乱の影響は考慮されず、物体表面における完全拡散反射と、また対象自身による背景の遮蔽関係のみが扱われる。

このような仮定に基づくと、画像間の画素値 (見え) の比較による視差推定が有効に働くためにステレオ法 [35][36] が可能となり、また対象による背景の遮蔽関係からシルエットが得られ、視体積交差法 [37][38][39] が実行可能となる。

こうして多視点映像から人物の時系列3次元形状を復元する研究は、対象のテクスチャ情報を用いるアプローチと、シルエット情報を用いるアプローチの2種類からスタートした。具体的には前者は通常のステレオ法をベースに、2.5次元形状 (depth-map) を貼り合わせることで全周囲3次元形状復元を行った Kanade らの研究 [2] と 2.5次元形状を介さずに直接3次元形状を求める Seitz らの研究 (volumetric stereo) [40] が、後者は視体積交差法で得られる visual hull をベースにした Moezzi らの研究 [1] が、共に 90 年代後半に提案されている。

*1 <http://www.alphamatting.com/> に [23] のベンチマークおよびいくつかのサンプルコードが公開されている。

また 2000 年代に入るとテクスチャ情報とシルエット情報を同時に用いるアプローチが多く研究されるようになった [41] [42] [4] [43] [44] [45] [46] [47] [48] [21] [10] [49]. これはテクスチャマッチングによる形状復元が「視点間の対応付けが決まれば対象表面形状を正しく計算できるが、視点間の対応付けを常に正しく行うことは容易ではない」という特徴を有しているのに対して、シルエットを用いた形状復元は「対象の概形 (visual hull) しか求まらないが、視点間の対応付けが不要で比較的安定に形状が求まる」という相補的な関係を持っているという分析に基づいている. すなわち、まず多視点シルエットから visual hull として「対象が必ず存在する範囲」を安定に求め、次いでこの範囲内でテクスチャの一致度 (photo-consistency) を最大化する形状 (photo hull) を求める、という考え方である (図 2).

また上記のように 1 時刻の 3 次元形状を独立に求めるだけでなく、複数時刻の形状を同時に復元する方法 [50][51] や、ある時刻における対象形状をまず復元し、これをキーフレームとして隣接する時刻の形状へと逐次変形することで形状と運動を同時に復元する方法 [52][53] のように 3 次元形状と同時に運動を推定する手法も研究されている.

2.4 テクスチャ生成

ここまでの時点で、対象人物の 3 次元表面形状を獲得することができたため、最後にその表面テクスチャを形状復元に用いた多視点映像から生成する. ここで重要となるのは、各表面領域を撮影していた実視点は複数存在するという点と、3 次元ビデオの表示に際しては「どの視点から表示するのか」を表す仮想視点という概念が加わるという点である (図 4 の \hat{C}).

まず、対象人物の 3 次元表面形状上の微小領域には、それを観測可能なカメラが一般に複数存在すると言える. ここでそのような微小領域を観測可能カメラそれぞれへと投影したときに、投影先での画素値が一致するのは

- (1) 微小領域の 3 次元形状が完全に正しく、
 - (2) 微小領域が完全拡散面 (Lambertian surface) であり、
 - (3) カメラの内部・外部パラメータ推定に誤差が無く、
 - (4) 各カメラの分光感度が完全に一致している
- という理想的な場合に限られ、このいずれかが成立しない場合は視点間で画素値が一致し得ない.

このような考えに基づくと、仮に (2) のみが成立しないと仮定した場合、画素値の差は対象表面の non-Lambertian な反射特性、すなわちその表面の方向依存な見えの変化を記録していると言える. そのため 3 次元ビデオの表示に際しては、仮想視点により近い実視点から優先的にテクスチャを生成すれば、より実写に近いレンダリングが可能となると期待される. この考え方は視点依存レンダリングと呼ばれる [4][10].

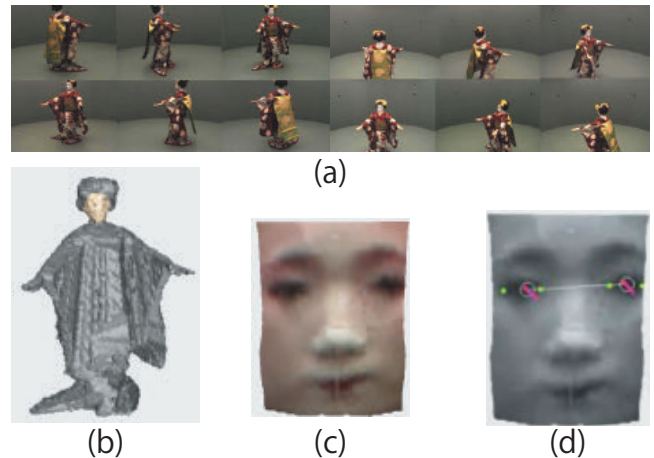


図 6 3 次元ビデオからの顔および視線方向推定 [60]. (a) 入力多視点映像, (b) 復元された 3 次元形状上で検出された顔領域, (c) 顔領域の超解像レンダリング, (d) 視線方向推定.

さらに形状復元やカメラの内部・外部キャリブレーションに誤りが存在すると見なすと、微小領域の撮影カメラへの投影先が幾何的にずれていることを考慮しなくてはならない. この位置ずれを仮想視点位置に応じて適応的に修正する方法として、Floating Texture Mapping 法 [54] では個別の実カメラ画像だけを使用してレンダリングした画像を用意して、それらの間のオプティカルフローを計算して補間画像を生成し、一方で Harmonized Texture Mapping 法 [55] では実カメラ画像を適応的に変形させることで位置ずれの無い高精細テクスチャを生成する. また仮想視点位置に応じて、適応的に 3 次元形状を最適化することでレンダリング品質の向上を図ることもできる [56].

逆に上記 (1) から (4) いずれも成り立つと仮定すると、対象表面のテクスチャはどの視点による撮影像から生成しても一致するはずであり、正しく CG と同様の、1 ポリゴンにつき 1 枚のテクスチャ画像を生成することができる. またこのとき、複数視点による撮影像はある共通の表面テクスチャを異なる格子点でサンプリングした結果であると思えることができる. このため仮想視点レンダリングの際に超解像処理を行う手法 [57] や、対象表面上で超解像処理を行う手法 [58] が提案されている.

また高品位レンダリングとは逆の問題設定として、仮に上記 (4) のみが成り立たないと仮定するならば、各視点での画素値の差異はカメラの分光感度の差に起因していると思えることができる. このことを利用して、カメラの分光感度をキャリブレーションする手法も提案されている [59].

3. 応用例

3.1 3 次元視線方向推定および一人称視点映像生成

前述のように、3 次元ビデオの特徴は対象の 3 次元形状を陽に推定する点であり、これを活用して被写体の顔の 3 次元位置を推定し (図 6(b) 肌色領域)、さらにその顔領域

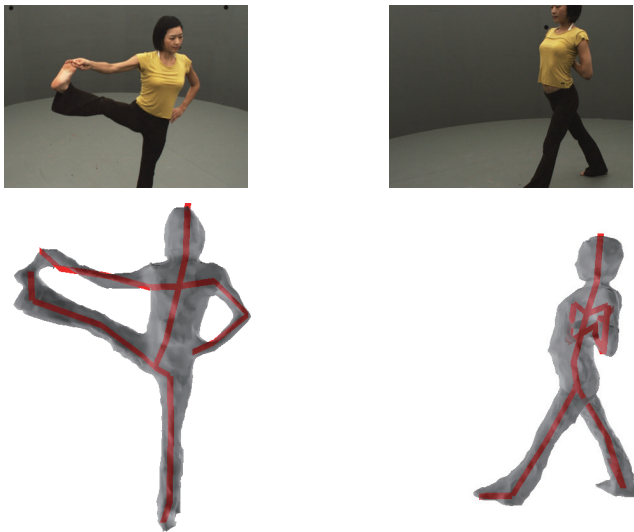


図 8 複雑な人物姿勢の推定 [61]. 上段：入力画像の一部，下段：復元された形状（灰色）と，推定された骨格姿勢（赤色）.

の正面に仮想カメラを置くことで仮想正面顔画像（図 6(c)）を作成することができる．こうして得られた正面顔画像から瞳を検出して視線方向を推定することで，非拘束・非接触で対象に自由な運動を許した 3 次元視線方向推定が実現できる [60].

また更に，得られた顔位置と視線方向に一致するようにレンダリング用の仮想視点を設定すると，図 7 のように被写体自身が見ていたような一人称視点の映像を生成することも可能である．

3.2 3次元運動推定

人物の 3 次元ビデオ，すなわち対象の時系列 3 次元形状とテクスチャが得られたと仮定した上で，対象の 3 次元運動を推定する研究が数多くなされている．これらは一般に (1) マッチングに基づく方法と (2) トラッキングに基づく方法に大別され，前者は各時刻の形状同士を比較し，3 次元表面上で互いに対応する点を求めることで運動を求める手法であり，必ずしも連続する時刻間でマッチングする必要がない点に特徴がある [62] [63] [64] [65].

一方後者はキーフレームメッシュモデルの変形（メッシュトラッキング）を逐次的に繰り返すことで運動を求める手法である [66] [67]. さらに，キーフレームメッシュモデルに骨格構造を埋め込み，メッシュ変形を骨格運動で記述することで人物の多関節剛体としての運動を推定する研究も行われている [68][69][70][61][71]. 例えば [61] では 3 次元ビデオに特徴的な，オクルージョンによる phantom volume や，形状復元の曖昧さに起因する形状の誤りを考慮した運動推定が行われており（図 8），また [71] では互いをカメラから遮蔽するような位置関係にある複数人物の運動を同時に推定することに取り組まれている．

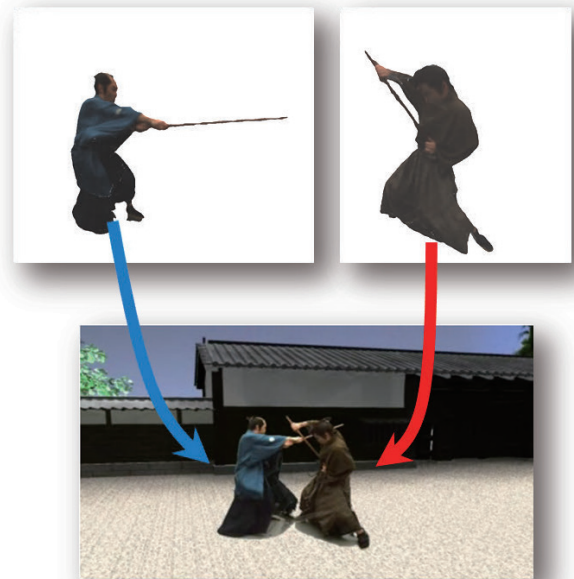


図 9 複数人物のインタラクションイベント編集 [72]

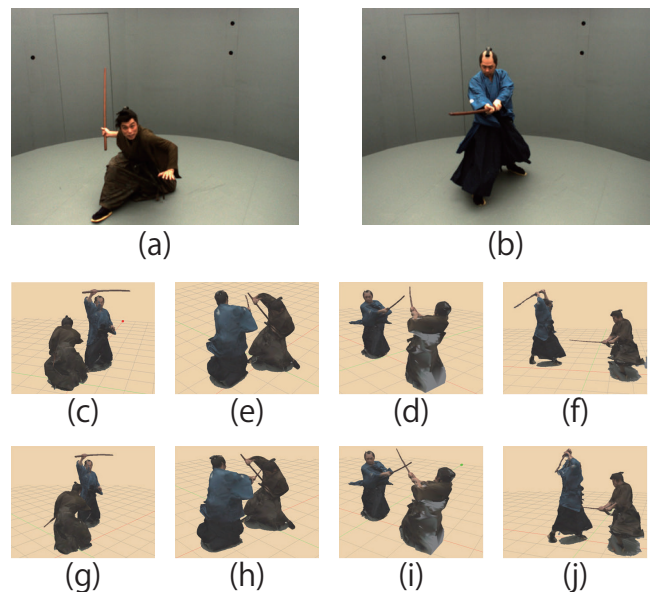


図 10 複数人物のインタラクションイベント編集結果 [72]

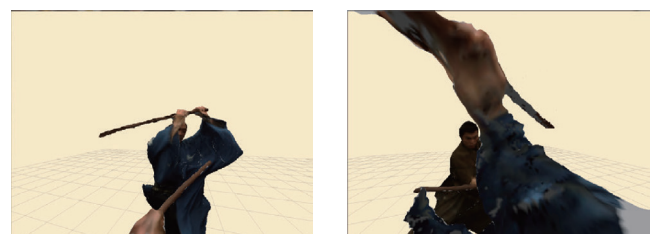


図 11 複数人物のインタラクションイベント編集結果の一人称視点レンダリング [72]

3.3 複数人物のインタラクションイベント編集

複数人物によるインタラクション（例えば握手など）を 3 次元ビデオで撮影する際には，対象人物が互いに撮影カ

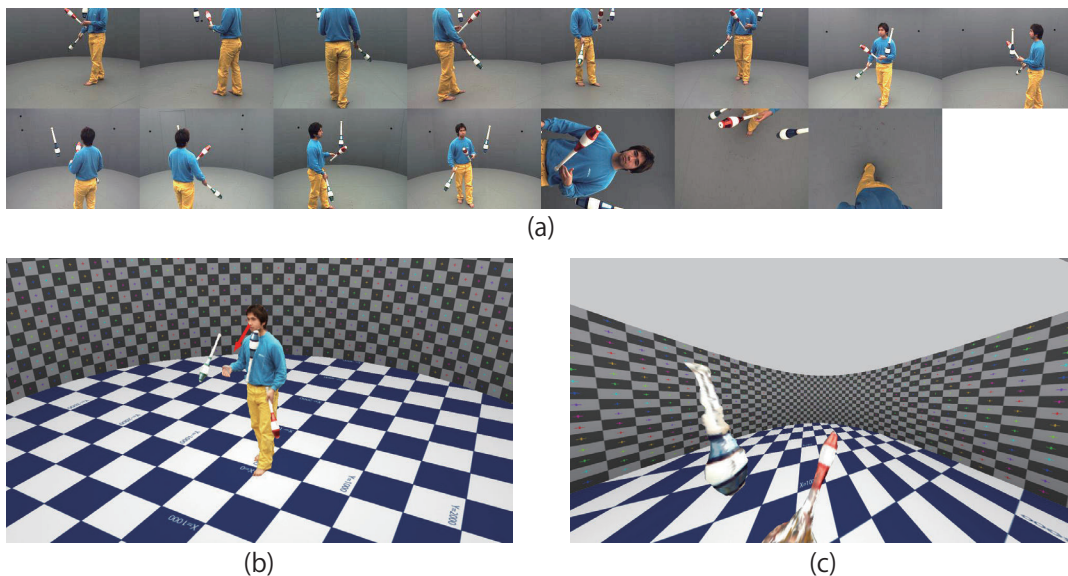


図 7 一人称視点映像生成 [60]. (a) 入力多視点映像, (b) 三人称視点レンダリング, (c) 一人称視点レンダリング.

メラから相手を遮蔽してしまい、向かい合う面がどのカメラからも観測されない場合が生じる。このような領域では形状が不完全となり、またテクスチャが得られないため、レンダリング時の映像品質の劣化が避けられない。

この問題を解決する一つの手法として、それぞれの動作を個別に3次元ビデオ撮影し、後にそれらを統合するアプローチが考えられる(図9)が、この場合はそれぞれの人物があたかも相手が存在するかのように振る舞う必要があるために、動作タイミングや位置のずれが発生する。

そこで我々は Augmented Motion History Volume (aMHV) と呼ぶ3次元ビデオ中の運動記述法と、aMHV 同士の組み合わせ方に基づいたインタラクションイベントの分類、そして aMMV の組み合わせに応じて適切に動作タイミング・位置のずれを修正しながら統合する手法を開発した [72]。この手法の特徴は、図 10(a), (b) のように骨格運動による対象表面形状の運動記述が困難な場合であっても、3次元ビデオがもともと持つ3次元表面形状のみを用いた aMHV によって運動記述する点、そして aMHV を処理単位とした編集アルゴリズムを定義した点である。

この手法による編集結果を図 10 に示す。同図 (a) および (b) が多視点映像の一例であり、同図 (c)~(f) が個別に生成された3次元ビデオを単純に合成した結果である。ここで (d) および (e) では刀同士が実際は打ち合っているはずが、合成結果では空隙が存在しており、また (f) では一方が他方を切っているはずが、合成結果では刀が届いていない。これに対して [72] による編集では、同図 (g)~(j) に示すように、いずれも動作の意図通りの結果が得られている。

またこうして得られた編集結果を前述の手法 [60] によりそれぞれの演者の視点からレンダリングした例を図 11 に示す。互いに向かい合い、相互にカメラから相手を遮蔽す

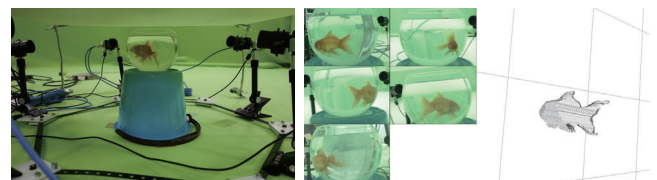


図 12 水中物体の視体積交差法による実時間形状復元 [73]. 左: 撮影環境, 中央: 入力画像, 右: 3次元形状.

る状況であるが、適切にテクスチャが生成できていることが確認できる。

4. まとめ

本稿では3次元ビデオ生成の基本的なアルゴリズムおよびアプリケーションを、関連研究と共に紹介した。今後の課題は (1) クロマキー背景などを持たない実世界での3次元ビデオ生成を実現することと、(2) 図 4 および図 5 に示したように、現在の技術では扱われていなかった光学現象である屈折・反射・散乱などを明示的にモデル化し、3次元ビデオとして撮影可能な対象のクラスをより広くすることが挙げられる。そのため現在筆者らはアクアビジョンと題して、このような光学現象がより顕著に観測される水中物体の3次元形状・運動復元に取り組み始めており [74][75][73] (図 12)、今後は半透明物体なども扱えるように研究を進めることを計画している。

謝辞 日頃より常にご指導を頂いている松山隆司教授をはじめ、京都大学松山研究室関係各位に心より感謝申し上げます。また本稿の内容は仏 Inria との JSPS 二国間交流事業共同研究、NTT メディアインテリジェンス研究所との共同研究、文科省委託事業「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」、JSPS 科研費(課題番

号 21700194, 23700204, 25540068), JST CREST「映画制作を支援する複合現実型可視化技術」, 総務省 SCOPE「和装文化振興を目的とした実写3次元映像のインターネット配信技術の研究開発」によるものです。

参考文献

- [1] Moezzi, S., Tai, L.-C. and Gerard, P.: Virtual View Generation for 3D Digital Video, *IEEE Multimedia*, pp. 18–26 (1997).
- [2] Kanade, T., Rander, P. and Narayanan, P. J.: Virtualized Reality: Constructing Virtual Worlds from Real Scenes, *IEEE Multimedia*, pp. 34–47 (1997).
- [3] Matsushita, T., Nobuhara, S., Takai, T. and Tung, T.: *3D Video and Its Applications*, Springer-Verlag (2012).
- [4] Matsuyama, T., Wu, X., Takai, T. and Nobuhara, S.: Real-Time 3D Shape Reconstruction, Dynamic 3D Mesh Deformation and High Fidelity Visualization for 3D Video, *CVIU*, Vol. 96, pp. 393–434 (2004).
- [5] Starck, J. and Hilton, A.: Surface Capture for Performance Based Animation, *IEEE Computer Graphics and Applications*, Vol. 27(3), pp. 21–31 (2007).
- [6] Franco, J.-S. and Boyer, E.: Efficient Polyhedral Modeling from Silhouettes, *IEEE TPAMI*, Vol. 31, No. 3, pp. 414–427 (2009).
- [7] Goldlucke, B., Ihrke, I., Linz, C. and Magnor, M.: Weighted Minimal Hypersurface Reconstruction, *IEEE TPAMI*, Vol. 29, No. 7, pp. 1194–1208 (online), DOI: 10.1109/TPAMI.2007.1146 (2007).
- [8] Smolic, A., Müller, K., Merkle, P., Fehn, C., Kauff, P., Eisert, P. and Wiegand, T.: 3D Video and Free Viewpoint Video Technologies, Applications and MPEG Standards, *Proc. of IEEE International Conference on Multimedia and Expo*, pp. 2161–2164 (2006).
- [9] Carranza, J., Theobalt, C., Magnor, M. A. and Seidel, H.-P.: Free-Viewpoint Video of Human Actors, *ACM Transactions on Computer Graphics*, Vol. 22, No. 3 (2003).
- [10] Hisatomi, K., Tomiyama, K., Katayama, M. and Iwade, Y.: Method of 3D reconstruction using graph cuts, and its application to preserving intangible cultural heritage, *IEEE Workshop on eHeritage and Digital Art Preservation*, pp. 923–930 (2009).
- [11] Guillemaut, J. Y., Hilton, A., Starck, J., Kilner, J. and Grau, O.: A Bayesian Framework for Simultaneous Matting and 3D Reconstruction, *Proc. of 3DIM*, pp. 167–176 (2007).
- [12] 4D View Solutions: *4DV Capture*, Grenoble, France.
- [13] Microsoft: Kinect (2010).
- [14] Creative Technology Ltd.: *Creative Sense3D*.
- [15] PMDTechnologies GmbH: *CamBoard pico*.
- [16] Softkinetic: *DS325*.
- [17] Ikeuchi, K., Oishi, T., Takamatsu, J., Sagawa, R., Nakazawa, A., Kurazume, R., Nishino, K., Kamakura, M. and Okamoto, Y.: The Great Buddha Project: Digitally Archiving, Restoring, and Analyzing Cultural Heritage Objects, *IJCV*, Vol. 75, pp. 189–208 (2007).
- [18] Fujii, T., Kimoto, T. and Tanimoto, M.: Ray Space Coding for 3D Visual Communication, *19th International Picture Coding Symposium*, pp. 447–451 (1996).
- [19] Takahashi, K. and Naemura, T.: Super-Resolved Free-Viewpoint Image Synthesis Based on View-Dependent Depth Estimation, *IPSJ Transactions on Computer Vision and Applications*, Vol. 4, pp. 134–148 (2012).
- [20] Nobuhara, S., Yoshimoto, H., Nakayama, H., Tung, T., Takai, T. and Matsuyama, T.: On-Site 3D Video Capture System, *ICCV 2009 Demo* (2009).
- [21] Starck, J., Maki, A., Nobuhara, S., Hilton, A. and Matsuyama, T.: The Multiple-Camera 3-D Production Studio, *IEEE TCSVT*, Vol. 19, No. 6, pp. 856–869 (2009).
- [22] Yamaguchi, T., Yoshimoto, H., Nobuhara, S. and Matsuyama, T.: Cell-based 3D Video Capture of a Freely-moving Object Using Multi-viewpoint Active Cameras, *IPSJ Transactions on Computer Vision and Applications*, Vol. 2, pp. 169–184 (2010).
- [23] Rhemann, C., Rother, C., Wang, J., Gelautz, M., Kohli, P. and Rott, P.: A Perceptually Motivated Online Benchmark for Image Matting, *Proc. of CVPR*, pp. 1826–1833 (2009).
- [24] Sun, J., Zhang, W., Tang, X. and yeung Shum, H.: Background cut, *Proc. of ECCV*, pp. 628–641 (2006).
- [25] 石川博: - CVIM チュートリアルシリーズ- コンピュータビジョン 最先端ガイド 1 第2章 グラフカット, アドコムメディア (2008).
- [26] Sarim, M., Hilton, A., Guillemaut, J. Y., Takai, T. and Kim, H.: Natural image matting for multiple wide-baseline views, *Proc. of ICIP*, pp. 2233–2236 (2010).
- [27] Zeng, G. and Quan, L.: Silhouette Extraction from Multiple Images of an Unknown Background, *Proc. of ACCV*, pp. 628–633 (2004).
- [28] Nobuhara, S., Tsuda, Y., Ohama, I. and Matsuyama, T.: Multi-viewpoint Silhouette Extraction with 3D Context-aware Error Detection, Correction, and Shadow Suppression, *IPSJ Transactions on Computer Vision and Applications*, Vol. 1, pp. 242–259 (2009).
- [29] Campbell, N., Vogiatzis, G., Hernandez, C. and Cipolla, R.: Automatic 3D object segmentation in multiple views using volumetric graph-cuts, *Image and Vision Computing*, Vol. 28, No. 1, pp. 14–25 (2010).
- [30] Sarim, M., Hilton, A., Guillemaut, J.-Y., Kim, H. and Takai, T.: Wide-baseline multi-view video segmentation for 3D reconstruction, *Proceedings of the 1st international workshop on 3D video processing*, pp. 13–18 (2010).
- [31] Guillemaut, J.-Y. and Hilton, A.: Joint Multi-Layer Segmentation and Reconstruction for Free-Viewpoint Video Applications, *IJCV*, Vol. 93, No. 1, pp. 73–100 (2011).
- [32] 延原章平, 松山隆司, ウ小軍, 松浦宣彦: 多視点画像を用いた複雑環境下における3次元形状・対象領域の同時推定, 第14回画像の認識・理解シンポジウム (2011).
- [33] Djelouah, A., Franco, J.-S., Boyer, E., Leclerc, F. and Pérez, P.: N-Tuple Color Segmentation for Multi-View Silhouette Extraction, *Proc. of ECCV* (2012).
- [34] Djelouah, A., Franco, J.-S., Boyer, E., Le Clerc, F. and Pérez, P.: Multi-View Object Segmentation in Space and Time, *Proc. of ICCV* (2013).
- [35] Okutomi, M. and Kanade, T.: A locally adaptive window for signal matching, *IJCV*, Vol. 7, pp. 143–162 (1992).
- [36] Scharstein, D. and Szeliski, R.: A Taxonomy and Evaluation of Dense Two-Frame Stereo Correspondence Algorithms, *IJCV*, Vol. 47, pp. 7–42 (2002).
- [37] Baumgart, B. G.: A polyhedron representation for computer vision, *Proceedings of the National Computer Conference and Exposition*, AFIPS '75, pp. 589–596 (1975).
- [38] Martin, W. N. and Aggarwal, J. K.: Volumetric description of objects from multiple views, *IEEE TPAMI*, Vol. 5(2), pp. 150–158 (1983).
- [39] Laurentini, A.: The visual hull concept for silhouette-

- based image understanding, *IEEE TPAMI*, Vol. 16, No. 2, pp. 150–162 (1994).
- [40] Seitz, S. and Dyer, C.: Photorealistic Scene Reconstruction by Voxel Coloring, *IJCV*, Vol. 25, No. 3, pp. 151–173 (1999).
- [41] Isidoro, J. and Sclaroff, S.: Stochastic Mesh-Based Multiview Reconstruction, *Proc. of 3DPVT*, pp. 568–577 (2002).
- [42] Cheung, K. M., Baker, S. and Kanade, T.: Visual Hull Alignment and Refinement Across Time: A 3D Reconstruction Algorithm Combining Shape-From-Silhouette with Stereo, *Proc. of CVPR*, pp. 375–382 (2003).
- [43] Esteban, C. H. and Schmitt, F.: Silhouette and stereo fusion for 3D object modeling, *CVIU*, Vol. 96, pp. 367–392 (2004).
- [44] Sinha, S. N. and Pollefeys, M.: Multi-View Reconstruction Using Photo-consistency and Exact Silhouette Constraints: A Maximum-Flow Formulation, *Proc. of ICCV*, pp. 349–356 (2005).
- [45] Starck, J., Hilton, A. and Miller, G.: Volumetric stereo with silhouette and feature constraints, *Proc. of BMVC*, pp. 1189–1198 (2006).
- [46] Tran, S. and Davis, L.: 3D Surface Reconstruction Using Graph Cuts with Surface Constraints, *Proc. of ECCV*, Vol. 3952, pp. 219–231 (2006).
- [47] Sinha, S., Mordohai, P. and Pollefeys, M.: Multi-View Stereo via Graph Cuts on the Dual of an Adaptive Tetrahedral Mesh, *Proc. of ICCV*, pp. 1–8 (2007).
- [48] Furukawa, Y. and Ponce, J.: Carved Visual Hulls for Image-Based Modeling, *IJCV*, Vol. 81, pp. 53–67 (2009).
- [49] Cremers, D. and Kolev, K.: Multiview Stereo and Silhouette Consistency via Convex Functionals over Convex Domains, *IEEE TPAMI*, pp. 1161–1174 (2010).
- [50] Vedula, S., Baker, S., Seitz, S. and Kanade, T.: Shape and Motion Carving in 6D, *Proc. of CVPR* (2000).
- [51] Goldlücke, B. and Magnor, M.: Space-Time Isosurface Evolution for Temporally Coherent 3D Reconstruction, *Proc. of CVPR*, pp. 350–355 (2004).
- [52] Nobuhara, S. and Matsuyama, T.: Heterogeneous Deformation Model for 3D Shape and Motion Recovery from Multi-Viewpoint Images, *Proc. of 3DPVT*, pp. 566–573 (2004).
- [53] Furukawa, Y. and Ponce, J.: Dense 3D motion capture from synchronized video streams, *Proc. of CVPR*, pp. 1–8 (2008).
- [54] Eisemann, M., De Decker, B., Magnor, M., Bekaert, P., de Aguiar, E., Ahmed, N., Theobalt, C. and Sellent, A.: Floating Textures, *Proc. of EUROGRAPHICS*, Vol. 27, No. 2, pp. 409–418 (2008).
- [55] Takai, T., Hilton, A. and Matsuyama, T.: Harmonised Texture Mapping (2010).
- [56] 高井勇志, 延原章平, 田辺太一, 松山隆司, ウ小軍: 視点依存形状最適化による高精細自由視点画像生成 (2012).
- [57] Tung, T., Nobuhara, S. and Matsuyama, T.: Simultaneous super-resolution and 3D video using graph-cuts, *Proc. of CVPR*, pp. 1–8 (2008).
- [58] Goldlücke, B. and Cremers, D.: Superresolution texture maps for multiview reconstruction, *Proc. of ICCV*, pp. 1677–1684 (2009).
- [59] Nobuhara, S., Kimura, Y. and Matsuyama, T.: Object-Oriented Color Calibration of Multi-viewpoint Cameras in Sparse and Convergent Arrangement, *IPJS Transactions on Computer Vision and Applications*, Vol. 2, pp. 132–144 (2010).
- [60] Qun, S., Shohei, N. and Takashi, M.: 3D Face Reconstruction and Gaze Estimation from Multi-view Video using Symmetry Prior, *IPJS Transactions on Computer Vision and Applications*, Vol. 4, pp. 149–160 (2012).
- [61] 延原章平, 宮本新, 松山隆司: 3次元形状計測における不完全性のモデル化に基づいた複雑な人物動作の推定, 電子情報通信学会論文誌 D, Vol. J92-D, No. 12, pp. 2225–2237 (2009).
- [62] Starck, J. and Hilton, A.: Correspondence labelling for wide-timeframe free-form surface matching, *Proc. of ICCV*, pp. 1–8 (2007).
- [63] Mateus, D., Horaud, R., Knossow, D., Cuzzolin, F. and Boyer, E.: Articulated shape matching using Laplacian eigenfunctions and unsupervised point registration, *Proc. of CVPR*, pp. 1–8 (2008).
- [64] Zaharescu, A., Boyer, E., Varanasi, K. and Horaud, R.: Surface feature detection and description with applications to mesh matching, *Proc. of CVPR*, pp. 373–380 (2009).
- [65] Tung, T. and Matsuyama, T.: Dynamic surface matching by geodesic mapping for 3D animation transfer, *Proc. of CVPR*, pp. 1402–1409 (2010).
- [66] Varanasi, K., Zaharescu, A., Boyer, E. and Horaud, R.: Temporal Surface Tracking Using Mesh Evolution, *Proc. of ECCV*, Vol. 5303, pp. 30–43 (2008).
- [67] Cagniard, C., Boyer, E. and Ilic, S.: Free-form mesh tracking: A patch-based approach, *Proc. of CVPR*, pp. 1339–1346 (2010).
- [68] Menier, C., Boyer, E. and Raffin, B.: 3D Skeleton-Based Body Pose Recovery, *Proc. of 3DPVT*, pp. 389–396 (2006).
- [69] Mukasa, T., Miyamoto, A., Nobuhara, S., Maki, A. and Matsuyama, T.: Complex human motion estimation using visibility, *Proc. of FG*, pp. 1–6 (2008).
- [70] Horaud, R., Niskanen, M., Dewaele, G. and Boyer, E.: Human Motion Tracking by Registering an Articulated Surface to 3D Points and Normals, *IEEE TPAMI*, Vol. 31, No. 1, pp. 158–163 (2009).
- [71] Liu, Y., Gall, J., Stoll, C., Dai, Q., Seidel, H.-P. and Theobalt, C.: Markerless Motion Capture of Multiple Characters Using Multiview Image Segmentation, *IEEE TPAMI*, Vol. 35, No. 11, pp. 2720–2735 (2013).
- [72] Shi, Q., Nobuhara, S. and Matsuyama, T.: Augmented Motion History Volume for Spatiotemporal Editing of 3D Video in Multi-party Interaction Scenes, *International Conference on 3D Vision (3DV)*, pp. 414–421 (2013).
- [73] Yano, T., Nobuhara, S. and Matsuyama, T.: 3D Shape from Silhouettes in Water for Online Novel-view Synthesis, *IPJS Transactions on Computer Vision and Applications*, Vol. 5, pp. 65–69 (2013).
- [74] 川原僚, 延原章平, 松山隆司: 水中撮影のための画素依存型バリフォーカルカメラモデル, 研究報告コンピュータビジョンとイメージメディア, No. 2013-CVIM-187(11), pp. 1–8 (2013).
- [75] 西村真衣, 延原章平, 松山隆司: 既知背景の屈折像を用いた透明多面体の3次元形状復元, 研究報告コンピュータビジョンとイメージメディア, No. 2013-CVIM-187(12), pp. 1–8 (2013).