

# Motion History Volume を使った 3次元人物動作の編集

†

**あらまし** 本論文の目的は3次元ビデオとして撮影された人物の時系列3次元表面形状に対する、Motion History Volume (MHV) と呼ぶ表現を用いた編集の体系を構築することである。MHV とは3次元ビデオを単調な運動をする時区間毎に区切り、それぞれの中で対象の運動部位が通過する空間のことである。本研究ではMHVの幾何学的変形と分解によって3次元ビデオの時間的・空間的編集が定義されることを示すとともに、その応用例として独立に撮影されて時間的・空間的な整合性を持たない2者の3次元ビデオを、時間的・空間的整合性をもった単一のシーケンスへと編集することを目指す。

**キーワード** 3次元ビデオ, 動作編集, Motion History Volume

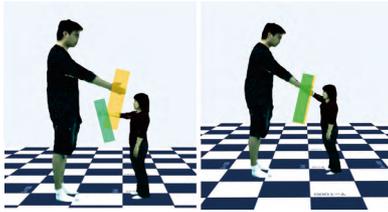


図1 握手を想定して別々に撮影された3次元ビデオを同時再生した例(左:オリジナル, 右:編集後. 黄色と緑のハイライト部分はそれぞれ左右人物腕部の運動範囲を示す)

## 1. はじめに

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などの時間的変化を3次元的にそのまま記録した実写立体映像で、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映像である[1]。この特徴を利用し、これまでは無形文化財のデジタルアーカイブ化のように編集を伴わない用途に用いられていたが、今後は次世代の全周囲立体映像メディアとしての利用も考えられる。

このような映像メディアとして捉えた場合、3次元ビデオは各時刻毎に多視点画像から復元された形状の系列であり、その作成原理から、撮影対象が互いに遮蔽するので、多くの人数を同時に撮影することは難しい。したがって、解像度の高い、多人数の出演する3次元ビデオを獲得するには、対象となる演者を別々に撮影する必要がある。しかし、別々に撮影すると、2者が同時に演じた場合には存在しない、2者の動作のタイミングがずれるといった時間方向の不整合と、2者の位置関係、足や手の高さなど、空間方向の不整合が発生する。

そこで本研究では、別々に撮影された人物動作を編集し、時間的、空間的に2者間の整合性を取る手法を提案する(図1)。特に運動が互いの制約条件となるシーン、具体的には相互接触が含まれる動作を含むシーンを想定し、運動情報を Motion History Volume と呼ぶ操作単位として可視化して、それに対する操作として編集を体系

的に定義する。

このような運動情報を可視化するという研究はこれまでも多く行われている。CVの分野ではWeinlandら[2]はそれぞれの運動で形成される体積の特徴を学習することで、未知の人物動作の認識を可能にした。またCGの分野においてRedomら[3]は通過した体積を接触判定に用いた。本研究では、通過した体積に運動方向に基づいて各面にラベルを与え、動作編集の種類を分類する。またその分類に基づいてそれぞれ編集を定義することで、動作編集を定義する。

## 2. 提案手法

先に述べたように本研究では、接触を想定した箇所の運動を後述のように Motion History Volume (MHV) によって可視化する。提案するアルゴリズムは編集者が2者のMHVをそれぞれのどの面同士が接触するかを指定すると、元の運動の特徴を保存しながら、自動的に時間的、空間的に整合した新たな3次元ビデオを生成する(図1)。これを本研究における3次元人物動作の編集と呼ぶことにする。

人物動作の編集は以下の3つのステップからなると定義する(図2)。

- (1) Position editing: 2者の3次元ビデオ全体を剛体として位置合わせを行う。
- (2) Posture editing: 接触に関係しない体節を編集して基本姿勢を求める。
- (3) Motion editing: 接触に関係する体節の空間的、時間的編集を行い、同期の取れた運動を求める。

以下ではまずMHVの定義を述べた後、これらのステップについて述べる。

### 2.1 MHVの定義

3次元ビデオ全体をシーケンスと呼び、MHVはシーケンスをモーションセグメントと呼ぶ単調な運動に区

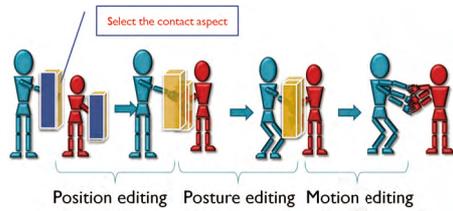


図 2 本研究のアイデア:接触箇所の運動情報を体積化

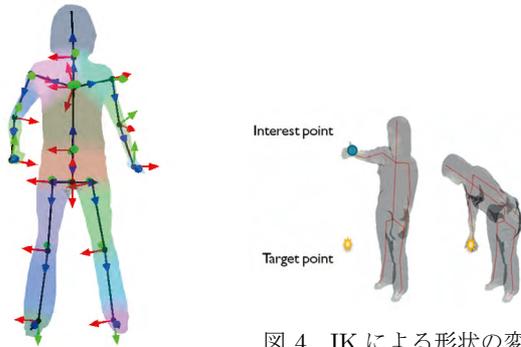


図 4 IK による形状の変形

図 3 skin-bone モデル: ボーン上の矢印は、そのボーンにおけるローカル座標系を示す。また、表面上の各領域の色は属するボーンの違いを表している。

切った区間毎に定義する。まず3次元ビデオの各時刻の表面形状は頂点の集合 ( $V_n$ ) とその頂点を結んだ枝の集合 ( $E_n$ ) でその表面形状が与えられ ( $M_n = \{V_n, E_n\}$ ) かつ図3に示す skin-bone モデル  $K(l)$  の位置, 姿勢パラメータ  $l = (T_0, \theta_0, \theta_1, \dots, \theta_{N_a})$  が各フレームにおいて宮本ら[4]の手法によってすでに求まっているとする。さらに任意の部位に対して目標点を与えると Inverse Kinematics (IK) アルゴリズムによって最適な各関節の角度が求められ、それに基づいてメッシュ形状も変形することが出来るものとする (図4) [5] [6]。また単調な運動の区間とは部位の運動が直線で近似できる区間とする。

このとき MHV は、モーションセグメント内で運動部位 (interest volume) によって通過した体積により定義される (図5)。MHV の形状は、空間中の軸曲線 (axis) とそれにそった断面形状 (disc) の変化による包絡線 (envelope) で記述される。本研究では interest volume の進行方向を axis, 進行方向と垂直の面に interest volume を射影したときにできる形状を disc とする。MHV 内における disc の粗密が、3次元ビデオの運動学的特徴を記述し、axis 方向の MHV の長さが運動範囲を表している。

また、実際に3次元ビデオの形状編集を行う際は、骨格構造を用いて行う。MHV と骨格構造の関連付けとして、interest volume におけるボーンのノードを interest point ( $p_{int}$ ) とし、axis はその点を通過するものとする。

また実際に接触する MHV 上の点として disc に含まれる代表点を contact point ( $v_c$ ) とする。contact point は、編集者が MHV 上の接触面を指定する際に、1つの disc につき1つ与えるものとする (図6)。

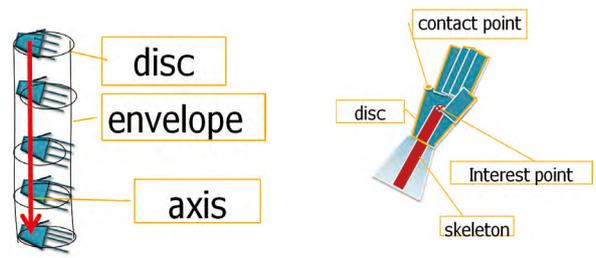


図 5 MHV の定義: 握手の例

図 6 interest point

と contact point の関係

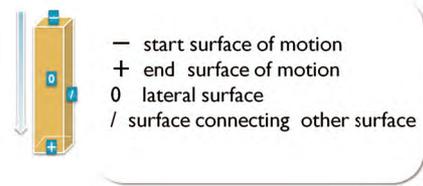


図 7 MHV に与えられるラベル

本研究では MHV に対する操作として

- (1) MHV 全体を剛体として、平行、回転移動する。
- (2) axis 方向に拡大、縮小する。
- (3) disc の位置で MHV を分割し新しい3次元表面形状を生成する。

があるものと定義し、これらの組み合わせで前述の position editing, posture editing, motion editing を実現する。MHV の並行・回転移動, axis 方向に拡大・縮小することは人物の運動範囲や姿勢を変更することに該当するが、この際はももとの3次元ビデオの実写性 (空間方向では3次元形状, 時間方向では運動) をなるべく保持するような制約のもとで行う必要がある。この考えに基づき、MHV を axis 方向に拡大、縮小する際は、MHV 内部の disc の位置が、元の disc 間の距離の比を保つ位置に変更されるとする。

また MHV の envelope に axis の終端で接している disc は「+」、始端で接している disc には「-」というラベルを、平行な面を「0」というラベルを与えることにする。そしてこのラベルの組み合わせによって2者の MHV の編集の仕方を定義する (図7)。

## 2.2 MHV を用いた2者の接触動作の分類

2つの3次元ビデオからそれぞれ MHV が生成されたとき、運動方向と面の関係によって決められた MHV のラベルに基づいて、接触パターンが次の何れかに分類される。すなわち、「+」と「-」の組 ( $C_{+, -}$  と表す)、「0」と「0」( $C_{0,0}$ )、「+」と「+」( $C_{+,+}$ )、「-」と「-」( $C_{-,-}$ )、「+」と「0」( $C_{+,0}$ )、「0」と「-」( $C_{0,-}$ ) である。 $C_{+,+}$  は、MHV が重なり合うような状態であり、 $C_{0,0}$  は互いの envelope が接触している状態である。これら2つの接触パターンを、長時間接触と呼ぶ。ここで

表 1 編集コマンドの分類:編集者はコマンドを選択するだけで人物動作の同期が可能

Editing command	$C_{+,}$	$C_{0,0}$	$C_{+,+}$	$C_{+,}$	$C_{,}$	$C_{,0}$
Shake hands		Shake hands	Raise arm	Drop down hand		
Kumfu fight	Squeeze together		Dap	After dapping	Punch	After punch
	Continuous contact		Instantaneous contact			

長時間とは、編集者が2フレーム以上連続で接触させるように編集することを意味する。それに対して残り4つは、短時間接触と呼ぶ。例として、別々に撮影された握手の3次元ビデオを編集しようとしたとき、それからどのような動作に編集可能か、また、カンフーの様なアクションシーンを編集したとき、どのような状況がどの編集コマンドの組み合わせに当てはまるかを表1に示す。

### 2.3 MHV を用いた 2 者の接触動作の編集

はじめに、本論文では簡単のために体節のうち接触部位しか運動していないものとする。また編集対象の2つの3次元ビデオはモーションセグメントの数が同数であるとする。

編集の始めに、編集者は、MHV 上の接触させたい面を指定する。これにより MHV 同士が表1のどのパターンで接触するか決まる。

図8に全体の処理の流れを示す。まず position editing を片方のシーケンス全体に適用する(図8の1)。次いで posture editing と motion editing を先頭 MHV ペアから逐次的に行う(図8の2, 3, 4, ...)。この逐次処理では時系列において一つ前のモーションセグメントの posture editing の結果を次の初期値(初期姿勢)とする。これはモーションセグメント間が滑らかにつながることが運動学的特徴を保持する条件だからである。

次いで各処理を定義するにあたり、以下の記号を導入する。編集する2つの3次元ビデオシーケンスを  $A, B$  とし、モーションセグメント  $i$  におけるそれぞれの  $n_i^A, n_i^B$  番目のフレームの形状は  $M_i^A(n_i^A)(n_i^A = 1, \dots, N_i^A), M_i^B(n_i^B)(n_i^B = 1, \dots, N_i^B)$  とする。それぞれのモーションセグメントの MHV ( $H_i^A, H_i^B$ ) に対して、編集コマンドを編集者が手作業で割り当てる ( $C_i$ )。割り当てた結果は、例えば、握手のシーケンスを想定した場合、図8のようになり、先頭の  $C_{+,+}$  は互いに腕を出して手を握るまでの動作、連続する  $C_{0,0}$  は握った手を振り上げる・振り下げる動作、に対応している。

出力すべきデータは、接触動作をしている2組の姿勢パラメータ  $l_i^{A,A} l_i^{B,B}$  と、それによって編集された3次元ビデオ  $M_i^{A,A} M_i^{B,B}(n_i)(n_i = 1, \dots, N_i)$  である。

#### 2.3.1 Position editing

Position editing はどちらのシーケンスに回転、平行移動を与えるかには制限はなく、片方のシーケンスに常に  $T_{x,y}, R_z$  を与えるとしても一般性を失わない。この  $T_{x,y}, R_z$  は、各シーケンスの最初のモーションセグメン

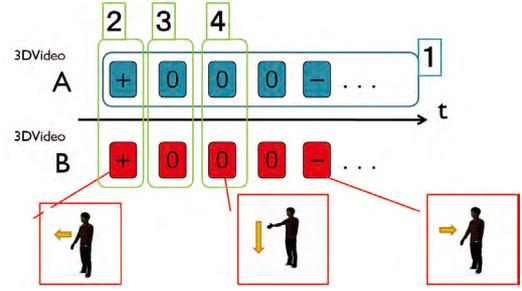


図 8 握手を例とした編集の処理手順

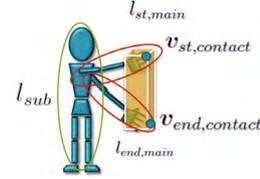


図 9  $l_{sub}, l_{main}, v_{st,c}, v_{end,c}$  の関係

トの  $MHV H_1^A, H_1^B$  上で接触させる面を入力として、その面を形成する点群(2.1で定義した contact point も含まれる)が最も近くなるための最適な  $T_{x,y}, R_z$  を ICP アルゴリズム[7]によって求める。ここでは A に行うものとし  $l_i^A(n_i^A)$  と  $M_i^A(n_i)$  を得る。

#### 2.3.2 Posture editing

連続性の制約を考える上で、姿勢パラメータ  $l$  を interest volume の運動への寄与が大きい関節のパラメータと  $l_{main}$  と、それ以外の  $l_{sub}$  とに分ける。

$$l = (l_{main}, l_{sub}) \quad (1)$$

$$l_{sub} = (T_0, \theta_{sub1}, \theta_{sub2}, \dots, \theta_{subk_{sub}}) \quad (2)$$

$$l_{main} = (\theta_{main1}, \theta_{main2}, \dots, \theta_{maink_{(k-sub)}}) \quad (3)$$

とすると、MHV の開始面における contact point  $v_{ist,c}$  は、

$$v_{ist,c}(M_i, K(l_{ist,main}, l_{isub})) \quad (4)$$

終了面における contact point  $v_{iend,c}$  は

$$v_{iend,c}(M_i, K(l_{iend,main}, l_{isub})) \quad (5)$$

と表すことが出来る。以降はこれを  $v_{ist,c}(l_{ist,main}, l_{isub}), v_{iend,c}(l_{iend,main}, l_{isub})$  と表すことにする。以下では、1番目のモーションセグメントとそれ以外のモーションセグメント時の制約について長時間接触と短時間接触に分けて述べる。この処理によって  $l_i^A$  と  $l_i^B$  を求め、それに付随した  $M_{ist}^{A,A}, M_{iend}^{A,A}, M_{ist}^{B,B}, M_{iend}^{B,B}$  と  $H_i^{A,A}, H_i^{B,B}$  を求める。

#### a) 長時間接触時の Posture editing

次のような目的関数を最小化することで、なるべくもとの姿勢を保持するという制約のもとで、contact point が接触している姿勢を求める。あるモーションセグメント  $i$  に関する、元の姿勢からの変更量を

$$e_{\text{fst,continuous},c \in \{A,B\}}^p(i, \mathbf{l}_{\text{ist,main}}^c, \mathbf{l}_{\text{end,main}}^c, \mathbf{l}_{\text{isub}}^c) = w_{\text{sub}}^c \|\mathbf{l}_{\text{isub}}^c - \mathbf{l}_{\text{isub}}^c\| + w_{\text{main}}^c (\|\mathbf{l}_{\text{ist,main}}^{c'} - \mathbf{l}_{\text{ist,main}}^c\| + \|\mathbf{l}_{\text{end,main}}^{c'} - \mathbf{l}_{\text{end,main}}^c\|) \quad (6)$$

とすると ( $w_{\text{sub}}^c$  と  $w_{\text{main}}^c$  は重み定数), 目的関数は,

$$E_{\text{first,continuous}}^p = \lambda_1^{\text{cont}} e_{f,c,A}^p(i, \mathbf{l}_{\text{ist,main}}^A, \mathbf{l}_{\text{end,main}}^A, \mathbf{l}_{\text{isub}}^A) + \lambda_2^{\text{cont}} e_{f,c,B}^p(i, \mathbf{l}_{\text{ist,main}}^B, \mathbf{l}_{\text{end,main}}^B, \mathbf{l}_{\text{isub}}^B) + \lambda_3^{\text{cont}} e_{\text{st}}^{\text{cont}} + \lambda_4^{\text{cont}} e_{\text{end}}^{\text{cont}}$$

となる ( $\lambda_1^{\text{cont}}, \dots, \lambda_4^{\text{cont}}$  は重み定数). ただし, それぞれのモーションセグメント  $i$  では,

$$e_{\text{st}}^{\text{cont}} = \|\mathbf{v}_{\text{ist,contact}}^A(\mathbf{l}_{\text{ist,main}}^A, \mathbf{l}_{\text{isub}}^A) - \mathbf{v}_{\text{ist,contact}}^B(\mathbf{l}_{\text{ist,main}}^B, \mathbf{l}_{\text{isub}}^B)\|$$

$$e_{\text{end}}^{\text{cont}} = \|\mathbf{v}_{\text{end,contact}}^A(\mathbf{l}_{\text{end,main}}^A, \mathbf{l}_{\text{isub}}^A) - \mathbf{v}_{\text{end,contact}}^B(\mathbf{l}_{\text{end,main}}^B, \mathbf{l}_{\text{isub}}^B)\| \quad (9)$$

でありこれらは MHV の開始面と終端面の代表点である contact point が接触しているという制約を表している. これ以降のモーションセグメントでは, 開始面を前のモーションセグメントで求めた姿勢として固定し, 同様の処理を行う.

#### b) 短時間接触時の Posture editing

基本的考え方は長時間接触時と同じである. ただし, 接触する contact point は, MHV の開始面か, 終端面のどちらかに限られるため, その分制約が少なくなる. MHV の開始面で接触する場合は, 命令コマンドとしては「+」が入ってきた場合であり, MHV の終端面で接触する場合は, 命令コマンドとして「-」が入ってきた場合である. ここでは, 「+」の場合について言及する. 「-」の場合は目的関数中の終端面 (end) と開始面 (st) とを入れ替えればよい.

$$e_{\text{fst,instant}(C_+),c \in \{A,B\}}^p(i, \mathbf{l}_{\text{end,main}}^c, \mathbf{l}_{\text{isub}}^c) = w_{\text{sub}}^c \|\mathbf{l}_{\text{isub}}^c - \mathbf{l}_{\text{isub}}^c\| + w_{\text{main}}^c \|\mathbf{l}_{\text{end,main}}^{c'} - \mathbf{l}_{\text{end,main}}^c\| \quad (10)$$

とすると, 目的関数は,

$$E_{\text{fst,instant}(C_+,+)(C_+,0)}^p = \lambda_1^{\text{instant}} e_{f,c+,A}^p(i, \mathbf{l}_{\text{end,main}}^A, \mathbf{l}_{\text{isub}}^A) + \lambda_2^{\text{instant}} e_{f,c+,B}^p(i, \mathbf{l}_{\text{end,main}}^B, \mathbf{l}_{\text{isub}}^B) + \lambda_3^{\text{instant}} e_{\text{end}}^{\text{instant}} \quad (11)$$

となる ( $\lambda_1^{\text{instant}}$  と  $\lambda_2^{\text{instant}}$ ,  $\lambda_3^{\text{instant}}$  は重み定数). ただし, それぞれのモーションセグメント  $i$  では,

$$e_{\text{end}}^{\text{instant}} = \|\mathbf{v}_{\text{end,c}}^A(\mathbf{l}_{\text{end,main}}^A, \mathbf{l}_{\text{isub}}^A) - \mathbf{v}_{\text{end,c}}^B(\mathbf{l}_{\text{end,main}}^B, \mathbf{l}_{\text{isub}}^B)\| \quad (12)$$

であり, これは MHV の開始面と終端面の代表点である contact point が接触しているという制約を表している. これ以降のモーションセグメントでは, 開始面を前のモーションセグメントで求めた姿勢として固定し, 同様の処理を行う.

#### 2.3.3 Motion editing

対応するモーションセグメントにおけるフレーム数を (7)  $N_i^A, N_i^B$  とする. ここで, MHV の開始面は,  $n_i^A = 1, n_i^B = 1$  の時であり,  $\mathbf{l}_{\text{ist,main}}$  はこのフレームに該当する姿勢パラメータとなる. 同様に, MHV の終端面は  $n_i^A = N_i^A, n_i^B = N_i^B$  のフレームであり,  $\mathbf{l}_{\text{end,main}}$  に該当するフレームである.

まずセグメントにおいて, interest volume に関係ない部位の姿勢パラメータ  $\mathbf{l}_{\text{isub}}$  は先の posture editing によって  $\mathbf{l}_{\text{isub}}^A, \mathbf{l}_{\text{isub}}^B$  として求まっているので, これをそのまま固有値として使用することとし, motion editing では,  $\mathbf{l}_{\text{isub}}^A, \mathbf{l}_{\text{isub}}^B$  を求めることを目標とする. そのため, 具体的には MHV 中の disc の位置をまず決定し, ついで IK によって対応する姿勢パラメータを求める.

新しく2つのシーケンスの同期がとれる共通の disc の数を  $N_i$  とする.  $N_i^A \leq N_i^B$  とすると, 元の3次元ビデオからの変更が少ないという制約を満たすため,  $N_i$  は

$$N_i^A \leq N_i \leq N_i^B \quad (13)$$

が望ましく, この範囲で編集者が指定するものとする. こうして与えられた  $N_i$  の数と, 編集コマンド  $C$  に基づいて MHV 上に存在する disc (同時に contact point) の数, 位置を最適化する. その最適化された位置に基づいて, 後述の 2.3.4 節の手法を使い適切な姿勢パラメータを取得する. 以下では, 長時間接触と短時間接触で設定する目的関数を説明する. 簡略化のため, contact point  $\mathbf{v}_{i,\text{contact}}$  を  $\mathbf{v}_i$  と表し, MHV の axis の長さで正規化されているとする.  $H_i^A$  上の contact point を  $\mathbf{v}_i^A(1), \mathbf{v}_i^A(2), \dots, \mathbf{v}_i^A(N_i^A), \mathbf{v}_i^B(1), \mathbf{v}_i^B(2), \dots, \mathbf{v}_i^B(N_i^B)$  とすると, 求める contact point を  $\mathbf{v}_i(1), \mathbf{v}_i(2), \dots, \mathbf{v}_i(N_i)$  とする.

#### a) 長時間接触時の Motion editing

長時間接触の場合,

$$E_{\text{continuous}}^m = \sum_{k=2}^{N_i-1} (e_{\text{continuous}}^{\text{pos}}(\mathbf{v}_i(k)) + \lambda e_{\text{continuous}}^{\text{sp}}(\mathbf{v}_{i-1}(k), \mathbf{v}_i(k), \mathbf{v}_{i+1}(k))) \quad (14)$$

を最小化することによってなるべくそれぞれの3次元ビデオの運動の特徴を保存した  $\mathbf{v}$  を求められる (図 10). ただし,  $\mathbf{v}_i^A$  と  $\mathbf{v}_i^B$  をそれぞれのモーションセグメントの中で,  $\mathbf{v}_i$  にもっとも近い  $\mathbf{v}$  とすると,

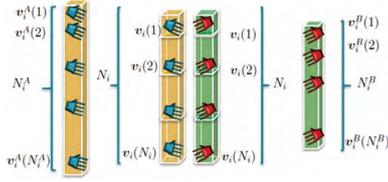


図 10 長時間接触時の Motion editing

$$e_{\text{continuous}}^{\text{pos}}(\mathbf{v}_i(k)) = \|\mathbf{v}_i(k) - \mathbf{v}_i^A\| + \|\mathbf{v}_i(k) - \mathbf{v}_i^B\| \quad (15)$$

$$e_{\text{continuous}}^{\text{sp}}(\mathbf{v}_i(k-1), \mathbf{v}_i(k), \mathbf{v}_i(k+1)) = \|\mathbf{v}_i(k+1) - 2\mathbf{v}_i(k) + \mathbf{v}_i(k-1)\| \cdot \|\mathbf{v}_{i+1}^A - 2\mathbf{v}_i^A + \mathbf{v}_{i-1}^A\| + \|\mathbf{v}_i(k+1) - 2\mathbf{v}_i(k) + \mathbf{v}_i(k-1)\| \cdot \|\mathbf{v}_{i+1}^B - 2\mathbf{v}_i^B + \mathbf{v}_{i-1}^B\| \quad (16)$$

である。なおこの式では posture editing で求まっている  $\mathbf{v}_i(1)$  と  $\mathbf{v}_i(N_i)$  は固定とする。

#### b) 短時間接触時の Motion editing

短時間接触時の場合、2者の運動に関する制約は、フレーム数  $N_i$  のみである。したがって、変更された MHV の axis の長さによって、もとの運動をなるべく保存するという制約のもと、最適化を行う。ここでは contact point を  $\mathbf{v}_i^C(1), \mathbf{v}_i^C(2), \dots, \mathbf{v}_i^C(N_i^C)$  とする ( $C \in A, B$ )。目的関数は、

$$E_{\text{instant}}^{\text{m}} = \sum_{k=2}^{N_i-1} e_{\text{instant}}^{\text{pos}}(\mathbf{v}_i(k)) + \lambda e_{\text{instant}}^{\text{sp}}(\mathbf{v}_i(k-1), \mathbf{v}_i(k), \mathbf{v}_i(k+1)) \quad (17)$$

$$e_{\text{instant}}^{\text{pos}}(\mathbf{v}_i(k)) = \|\mathbf{v}_i(k) - \mathbf{v}_i^C\| \quad (18)$$

$$e_{\text{instant}}^{\text{sp}}(\mathbf{v}_i(k-1), \mathbf{v}_i(k), \mathbf{v}_i(k+1)) = \|\mathbf{v}_i(k+1) - 2\mathbf{v}_i(k) + \mathbf{v}_i(k-1)\| \cdot \|\mathbf{v}_{i+1}^C - 2\mathbf{v}_i^C + \mathbf{v}_{i-1}^C\| \quad (19)$$

となる。こうして新たな disc の位置が決まったため、次のようにして姿勢パラメータを求める。

#### c) 編集後の MHV からの姿勢パラメータ決定

編集後の MHV には位置の変更された disc と、disc 上に定義された変更後の contact point  $\mathbf{v}'_{\text{contact}}$  がある。まず編集前の contact point  $\mathbf{v}_{\text{contact}}$  と interest point  $\mathbf{p}_{\text{int}}$  の相対的な位置関係に基づき、 $\mathbf{v}'_{\text{contact}}$  に対応する  $\mathbf{p}'_{\text{int}}$  を求める。この新しく求めた interest point に対して、同じモーションセグメント内の  $\mathbf{p}_{\text{int}}(n_i)$  ( $n_i = 1, \dots, N_i$ ) の中から、もっとも距離の近い interest point を持つ姿勢パラメータを持つ骨格構造を選択し、その骨格構造を用いて、target point を  $\mathbf{p}'_{\text{int}}$  にして IK を行う。ここでもっとも距離の近い骨格構造を選択したのは、IK によって算出される姿勢パラメータは初期姿勢に大きく依存するからである。

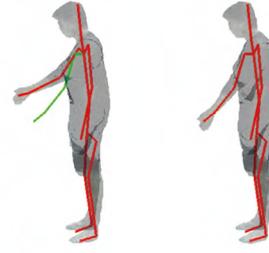


図 11 骨格ノードが一番近い入力骨格構造を変形



図 12 左:求めたいメッシュ. 右:制限を無視して編集した例:腕の内側の箇所が本来計測されていない面を必要としたため、メッシュが適切に変形出来ていない

#### 2.3.4 姿勢パラメータからの 3次元表面形状の作成

求められた姿勢パラメータを用いて、骨格構造に関連付けられた 3次元ビデオの表面形状を変更させることで、新しい表面形状の生成を行う。3次元ビデオはもともと多視点画像から得られた実写映像であるため、衣服の皺といったものも形状として保存しており、なるべくその形状を壊すことは避けることが望ましい。そこで、編集後の姿勢パラメータに最も姿勢の類似したフレームを元のセグメントから選び、このフレームを元にして新たなフレームを作る。これは具体的には以下のアルゴリズムで実現される。

MHV を形成する disc と axis の交点、すなわち interest volume における代表の骨格ノードである interest point に注目する。モーションセグメント  $i$  において求められた骨格構造  $K(\mathbf{l}_i(t))$  の interest point  $\mathbf{p}_{\text{int}}(\mathbf{l}_i(t))$  にもっとも近い距離を持つ interest point を持った骨格構造  $K(\mathbf{l}_i(o))$  をモーションセグメント  $i$  における  $K(\mathbf{l}_i(n))$  ( $n = 1, \dots, N_i$ ) から選択する。 $K(\mathbf{l}_i(o))$  を  $K(\mathbf{l}_i(t))$  と一致するように各  $\theta$  を動かし、 $K(\mathbf{l}_i(o))$  に対応付けられたメッシュ形状  $M(K(\mathbf{l}_i(o)))$  を変形させる。そこで求めた  $M$  を  $K(\mathbf{l}_i(t))$  に対応した  $M'$  とする。

ただし、姿勢パラメータに付随して形状を編集したとき、もともと観測されなかった箇所の形状が必要になる場合がある。例えば、もともとは直立姿勢であった人物の腕を上げるように姿勢を編集したとき腕と体の接触部分の頂点が必要になるが、これは原理的に観測できず、観測データからは欠落している。これを無視したまま形状を編集すると、図 12 のように頂点間の距離が元の形状と比べて大きく伸ばされてしまい、適切に変形できないという問題が起こる。これでは編集後の形状は 3次元ビデオの実写性を生かしたものにならない。

そこで、本研究では、変形前のメッシュ  $M(K(\mathbf{l}_i(o)))$

から変形後メッシュ  $M'$  の頂点間距離の歪み量を評価する。これは骨格構造に基づいた多関節剛体変形では変形前後で頂点間の距離が保たれるという仮定に基づく。メッシュ  $M'$  上のエッジ集合  $E'$  と  $M$  上のエッジ集合  $E$  に含まれる頂点間の距離の和の変形前後の差を変化量とする。この差がある閾値を超えた場合、骨格構造  $K(l_i(t))$  に採用するメッシュとして、 $M(K(l_i(o)))$  はふさわしくないと判定する。変化量が閾値を越えないフレーム中で最も近い  $p_{\text{int}}(l_i(t))$  に最も近い距離を持つ interest point を持った骨格構造をモーションセグメント  $i$  から選び、 $K(l_i(t))$  とする。

### 3. 実験

提案手法を用いて3次元人物動作を編集する。提案した手法は、大きく長時間接触と短時間接触に分かれるため、それぞれの代表的な動作について実験を行った。なお各評価関数の最小化は共役勾配法によって行った。

#### 3.1 長時間接触の例:握手

##### 3.1.1 実験に使用するデータ

長時間接触時の実験データとして身長差・体格差がある2者が握手を想定した動作を行っている3次元ビデオの編集を行った。

3次元ビデオ A 身長 180cm の人物が握手しているかのように腕を上下に振っている2つのモーションセグメントで成り立っているシーケンス (図 13, 14 の左側)  $N_1^A = 12$ ,  $N_2^A = 11$ .

3次元ビデオ B 身長 110cm の人物が握手しているかのように腕を上下に振っているシーケンス。(図 13, 14 の右側)  $N_1^B = 16$ ,  $N_2^B = 12$ . ただしこのシーケンスは元々身長 155cm の人物のデータのスケールを 0.7 倍して作成された仮想的なデータである。

##### 3.1.2 実験結果

編集結果を図 15 に示す。互いに手のひらによって形成される面を接触面として指定した。つまり、モーションセグメント 1, 2 どちらに対しても編集コマンドとして  $C_{0,0}$  を与えた。また、 $N_1 = 13$ ,  $N_2 = 11$  とした。 $l_{\text{main}}$  は肩までの関節角とした。

図 13 および 14 から、もとの3次元ビデオは動作時間が異なること、また運動範囲も空間的に大きく異なることが確認でき、両者をそのまま同時再生、あるいは単純に動作時間のみを揃えて再生しても空間的、時間的整合性がとれないことがわかる。

これに対して図 15 から、提案手法によって空間的、時間的不整合が解消された編集結果が得られていることが確認できる。

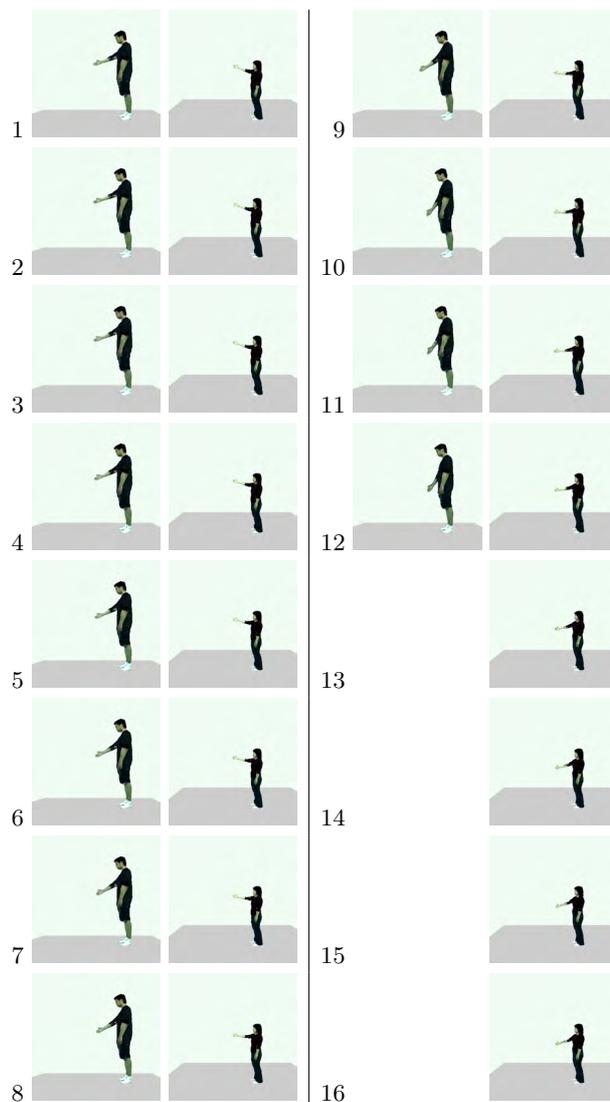


図 13 3次元ビデオ A, B のモーションセグメント 1. 運動範囲, タイミングとも異なっていることがわかる。

#### 3.2 短時間接触の例:ハイタッチ

##### 3.2.1 実験に使用するデータ

短時間接触の代表例として、ハイタッチを想定した動作を行っている2つの3次元ビデオの編集を行った。

3次元ビデオ A 身長 168cm の人物が手を上げてハイタッチするモーションセグメント (図 16 の左)。  $N_1^A = 8$ .

3次元ビデオ B 身長 170cm の人物が手を上げてハイタッチするモーションセグメント (図 16 の右)。  $N_1^B = 13$ .

##### 3.2.2 実験結果

図 17 に編集結果を示す。接触面を手のひらとし、 $C_{+,+}$  の接触パターンとして編集を行った。 $N_1 = 8$  とした。編集結果を図 17 に示す。今回の実験データでは、position editing の段階で MHV の指定した接触面が接触したこと、continuous contact ではないことから、posture editing に

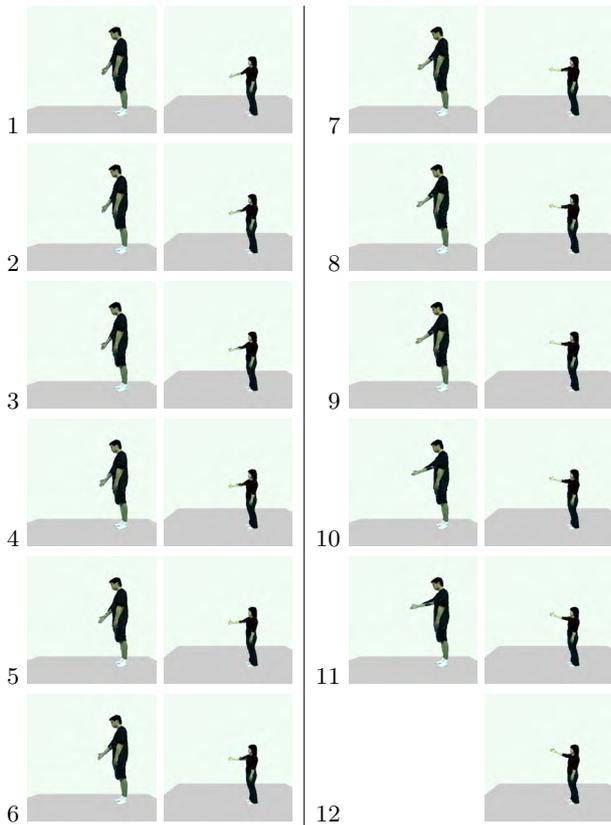


図 14 3次元ビデオ A, B のモーションセグメント 2. 運動範囲, タイミングともに異なっていることがわかる

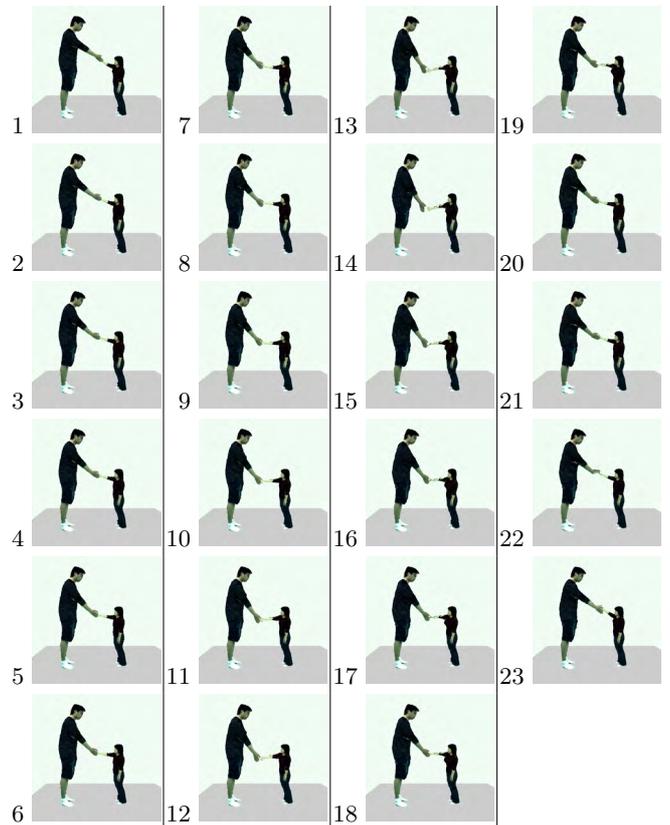


図 15 長時間接触時結果. 運動範囲, タイミングともに整合性がとれている

よる処理を行わなかった。

3.1.2 節と同様に図 16, 17 を比べると提案手法によって空間的, 時間的整合性が取れたことがわかる。

## 4. 結論

本論文では 3次元ビデオとして撮影された人物の運動情報を Motion History Volume (MHV) として可視化し, それに対する操作として相互接触が含まれる 2 者の動作編集を体系的に定義することができた。また, 実験を通してその有効性を確認することができた。

本研究における手法の限界について, 人物動作としての制限, 3次元ビデオとしての制限, MHV の形状に関する制限の 3 つの視点から示し, その後今後の課題についてまとめる。

### 4.1 提案手法の限界

#### a) 入力 3次元ビデオに関する制約

本研究では 2 者が互いの接触動作を想定して行った運動を撮影した 3次元ビデオを編集することを目的とした。そのためにこの前提に反して想定から大きく外れた 3次元ビデオが入力され, その結果人体の姿勢として適切な関節角を越えた姿勢への編集を要求するようなユーザー入力などに対しては編集を実現できない。

#### b) 3次元ビデオとしての制限

2.3.4 節で述べたように, 体節同士が接触して表面形

状が欠落した領域を補うことは現在のアルゴリズムでは出来ない。したがって, 姿勢パラメータが求まったとしても, それに適した 3次元形状を生成することができるとは限らない。たとえば常に腕を組んでいるモーションセグメントからは, どれほど姿勢パラメータが近くても腕を解いた状態は生成できない。

#### c) MHV の形状に関する制限

Interest volume が通過した体積を MHV として定義しているが, 本研究では, その通過した体積の軌跡を直線としてモデル化している。これは MHV の面に対して運動方向から割り当てられるラベルを振り, 体系的に編集を扱うため, 弧を描くような運動や, 円を描くような運動を扱うことは想定していない (円弧を細かくセグメントに区切って, 直線運動と近似することで編集は可能であると考えられる)。

### 4.2 今後の課題

今後の課題としては, 上述した提案手法の制限を取り除く手法を考える必要がある。また, 今回は, 問題を簡単にするために, interest volume 以外は運動していないと仮定して接触動作を扱ったが, 本来の人物動作はそうでないことも多い。今後は, interest volume に関係していない体節が運動している状況や, MHV が 2 箇所生成される状況などについても考える必要がある。また, 握手の様に握りこむような接触動作の場合, それをどの

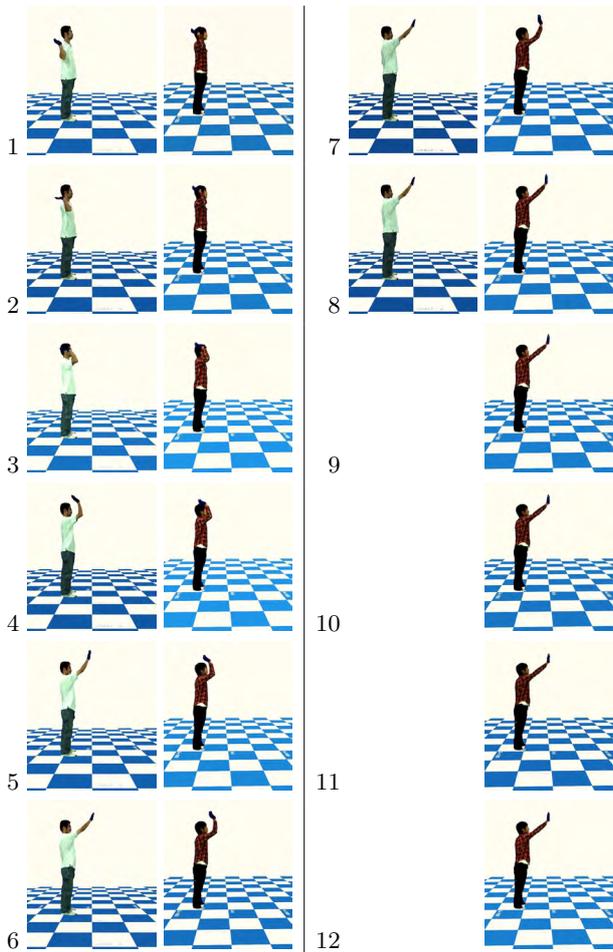


図 16 3次元ビデオ A, B のモーションセグメント

モーションセグメント単位で編集を行ったが、モーションセグメント間の interest volume の加速度の滑らかさの制約を入れていない。これを考えることで、より運動の特徴を保持した編集を行うことができると考えられる。

## 文 献

- [1] 延原, 和田, 松山: “弾性メッシュモデルを用いた多視点画像からの高精度 3 次元形状復元”, 情報処理学会 CVIM 研究会論文誌, SIG11(CVIM5), **43**, pp. 53-63 (2002).
- [2] D. Weinland, R. Ronfard and E. Boyer: “Free viewpoint action recognition using motion history volumes”, *COMPUTER VISION AND IMAGE UNDERSTANDING* (2006).
- [3] S. Redon, Y. J. Kim, M. C. Lin, D. Manocha and J. Templeman: “Interactive and continuous collision detection for avatars in virtual environments”, *Virtual Reality Conference, IEEE*, p. 117 (2004).
- [4] 延原, 宮本, 松山: “三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 (画像認識, コンピュータビジョン)”, 電子情報通信学会論文誌. D, 情報・システム, **92**, 12, pp. 2225-2237 (2009-12-01).
- [5] C. Hecker: “Real-time motion retargeting to highly varied user-created morphologies”, *ACM Transactions on Graphics*, **27**, (2008).
- [6] C. Welman: “Inverse kinematics and geometric constraints for articulated figure manipulation”, Master’s thesis, Simon Fraser University (1993).
- [7] S. Rusinkiewicz and M. Levoy: “Efficient variants of the icp algorithm”, *3D Digital Imaging and Modeling, International Conference on*, **0**, p. 145 (2001).

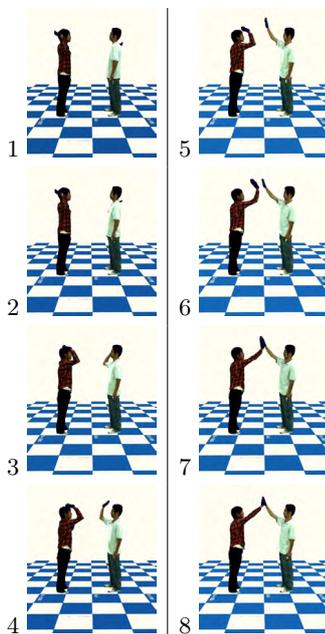


図 17 結 果

ように体系化するかについても、考える余地がある。また、今回は 2 者の接触動作を扱ったが、同時に 3 者以上の運動の時空間的な整合性をとる為には、現在の体系で編集が可能か考える必要もある。最後に、本研究では、