Vol.2010-CG-140 No.2
2010/9/8

差分画像の時空間ボリュームによる動きのクラスタリング

天野 祐樹[†1]　床井 浩平[†2]

近年，テレビゲームの対話的な操作にジェスチャを利用されるようになってきた．ジェスチャをシステムとの対話に応用する事で，直感的な操作が可能になる．本研究では一連のジェスチャの文脈を考慮した動作の意味理解を行う事を目的として，ジェスチャの文脈を構成するボキャブラリとなるジェスチャの要素を分離し収集するために，シーケンス内の各フレームの差分による時空間ボリュームを用いた動きのクラスタリング手法を提案する．

A clustering method of human motion using time-space volume of differential images

Yuuki Amano[†1] and Kohe Tokoi[†2]

Recently, gesture is getting to be used for interactive operation of video games. An intuitive operation becomes possible by applying the gesture to the conversation with the system. In this research, we attempted to use a time-space volume image of differential images between adjacent frames of a video sequence for the segmentation and collection of the motion elements that compose the context of the gesture in order to understand the semantics of the sequence of the gesture.

†1 和歌山大学大学院システム工学研究科
Graduate School of Systems Engineering, Wakayama University
†2 和歌山大学システム工学部
Faculty of Systems Engineering, Wakayama University

1. はじめに

近年，テレビゲームや家電製品の操作にジェスチャを利用する試みが活発化してきており，コントローラなどのデバイスを持たずに遊べるハードウェアも登場している．また，ハンドジェスチャによる大画面環境での web ブラウジングや文字入力手法の提案[1)] や，ジェスチャを理解する上で人物の動きのみでなく，視覚的文脈を考慮し，人物映像のデータベースから動作カテゴリを教師無しで学習する手法の提案[2)] といった研究も行われている．

ジェスチャを機器の操作に応用する場合，モーションキャプチャシステムやステレオカメラ，あるいは距離画像センサ等を用いることができる．これらは身体各部の 3 次元位置情報を得ることができるため，たとえば指先をポインタ代わりに用いて画面上の位置を指定するような操作が可能になる．しかし，実際にそのような操作方法を実現しても，我々の実験[5)] では被験者から高い評価が得られなかった．

ジェスチャはコミュニケーションの一手法として，場合によっては言葉以上に重要かつ自然な情報伝達手段である[4)]．我々は自分の意思を表すとき，無意識にジェスチャを交えている．例えば，相手の話に同意する時のうなずいたり，困ったときに頭を掻くなど個人差もあるが何らかの心理状態を表す．ボディランゲージのように文化的な背景から意味に制約を持つものとくらべ，ジェスチャはよりプリミティブな意思伝達手法であり，何らかの意図を相手に伝えるために意味を持つ身体動作の集合として考えられる．この点において，ジェスチャの正確な位置情報は重要な意味を持たない．

コンピュータとの単純な対話を行う上で身体動作を利用する場合，身体に負担をかけずに自然な状態を保ったままコミュニケーションを行う事が望まれる．それは単に不自然な姿勢を取る必要がないだけでなく，一連のジェスチャが発想からの自然な表出であることが求められている．

そこで本研究では，身体動作により機器を操作することを目的として，Web カメラにより取得したビデオシーケンスから映像データベースを作成し，それを用いたジェスチャの認識を試みる．このデータベースはジェスチャを構成する身体動作の要素（動きの形態素）を集めたものである．ビデオシーケンスから身体動作の要素を抽出するために，本研究ではフレーム間の差分画像の時空間ボリュームを作成する．本稿ではこの時空間ボリュームのクラスタリングにより，身体動作の要素を分離する手法について説明する．

© 2010 Information Processing Society of Japan

Vol.2010-CG-140 No.2
2010/9/8

2. ジェスチャ認識フレームワーク

ジェスチャはコミュニケーションに伴う心理状態や意図を表す言葉によらない語彙であるとも捉えられる．本研究では，コンピュータとのコミュニケーションにジェスチャを利用する．その際，従来のモーションキャプチャシステムのように詳細な位置情報を取得するのではなく，動画像情報から大まかに動きを切り出し，それらを語彙を形成する動きの形態素として収集する．言葉による言語と同様に，収集された形態素はそれ自体では何かの意味を表すとは考えにくい，そのため無数の組み合わせがある中で，動きの時間的・空間的な連続性といった動力学拘束と動きの内容による意味的な拘束を表す条件によって，ごく限られた数の系列のみが実際には出現すると考えられる．このような制約条件をジェスチャの形態素系列の生成規則といった形でルールベース化できるのではないかと考えた．ルールベース化が可能であれば，ジェスチャの認識のみならず，アニメーションキャラクタやロボットなどの設定した心理的状況の表出となるジェスチャの自動生成にも活用できる．そこで本稿では，まずジェスチャの形態素への分割可能性を差分画像ボリュームの分割により求めようと試みた．

ジェスチャによるインタフェースを実現するためには，大量にある映像情報データベースから動きを抽出・分類し，それらを意味のある形態素の列として連結するジェスチャ認識フレームワークが必要になる．本研究のジェスチャ認識フレームワークの処理手順は，動画像情報からジェスチャの語彙となる形態素として動きを切り出して収集するボキャブラリ収集フェーズと，一連のジェスチャを構成する形態素列を知識として利用するボキャブラリ結合フェーズの二つのフェーズからなる．そこで本稿の 3 節では第一段階のボキャブラリ収集フェーズについて述べ，第 4 節では RPG におけるアクションをプレーヤのジェスチャによって指定することを想定して，その一部のジェスチャを標本に用いた実験結果を示す．

3. ボキャブラリ収集フェーズ

ボキャブラリ収集フェーズでは以下の処理を行う．

3.1 時系列画像からの動きの抽出

時系列画像から動きを抽出する．まず映像情報データベースよりジェスチャ映像データを取得する．このとき使用する映像データはジェスチャを利用する個別の操作ごとに，その操作に適当と思われる所望の動作が行われているものを選別して用いる．本手法ではジェスチャが動きとしてシーケンスに現れる領域を特定し，ボリュームとして抽出するためにフレーム間差分法を用いている．取得した映像のシーケンスにおける各フレームごとにフレーム間差分をとると，各フレーム間での変化領域が画素値の差として抽出される．今回用いたフレーム間差分法の他にもフレーム内の画像から特定の領域を抽出する方法として，肌の色などの特定の色情報に基づいた領域抽出手法や，あらかじめ抽出対象が含まれない背景となる画像を取得し，その背景画像との差分をもとに領域を抽出する背景差分法がある．以下に，抽出手法について検討した内容について述べる．

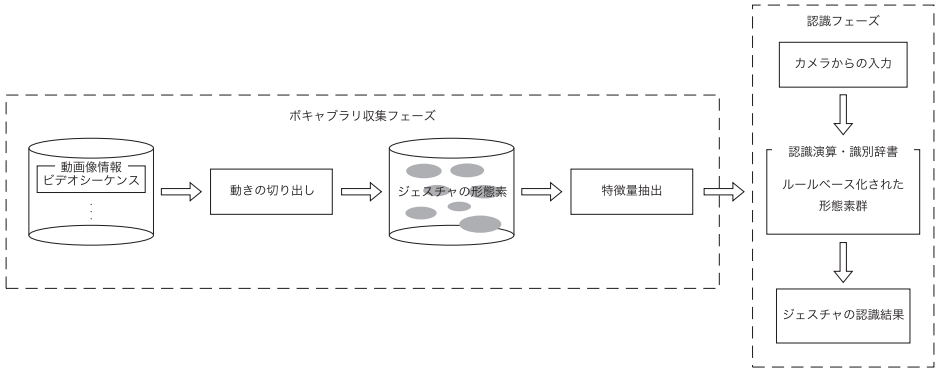


図 1 ジェスチャ認識フレームワークの概念図
Fig. 1 conceptual diagram of gesture recognition framework

今回の目的のようにジェスチャを大まかな動きの塊として取得するために，以下の 3 点を条件として設定した．

(1) 動きのある部分（もしくはオブジェクト）を背景から切り離して扱う事ができる．
(2) 一つの独立した動きをとして，動きの始点と終点を求める事ができる．
(3) 背景や光源等の環境変化にロバストである．

- **色情報に基づいた抽出**

  肌色などの人物の特徴的な色情報をもとに，人物領域を抽出する方法である．ブルーバック等で背景から抽出対象となる物体の色が十分に分離されている場合，精度の高い領域抽出が可能だが，あらかじめ抽出対象の色情報を取得する必要がある．人物の場合，肌色に個人差がある場合などその度に色情報を取得する必要がある．また，処理に用いる表色系に抽出精度が依存する．背景にあるオブジェクトや光源など周囲の環境のノイズに影響を受けやすい．動いているか否かに関係なくマッチした領域を抽出するため，連続した位置のトラッキングは可能であるが，動作の始点と終点を別に求める必要

© 2010 Information Processing Society of Japan

がある．

- **背景除去による抽出**
  抽出対象となる人物が存在しない背景をあらかじめ取得しておいて，その背景画像と人物が写っている画像との差分を取りその変化領域を抽出する．背景差分法は背景取得時の時刻におけるフレームと，現在の時刻でのフレームとの輝度値の差分を計算する．背景を取得した時点からの変動部分以外が除去される．一度取得された背景情報を，差分を求めるベースとして使用し続けるため，背景のオブジェクトの変動（揺らぎ，移動），始点の移動，光源の変化（照明の ON/OFF，太陽光の変化，カメラの自動的なゲイン調節）等が起こった場合背景画像を更新する必要がある．また前景となる人物などがフレームに存在している間その物体を抽出し続けるため，動作の始点と終点を別に求める必要がある．
- **フレーム間差分法による抽出**
  ビデオシーケンスの 1 フレームごとの輝度値の差の絶対値を求めて，差分画像を作成する．背景と異なる輝度値（画素値）の物体が移動した場合，すなわち，ある輝度値の分布が変化した場合，前のフレームと現在のフレームとの間で，その変化部分に差が生ずる．それによって，移動物体の存在部分を抽出する．フレーム間の輝度値の差を求める事で，移動する物体の輪郭領域が抽出される．1 フレームごとの差分を計算するため，変化が起きた部分（変化の急増減部分）が局所的に抽出される．短時間に発生する環境変化のノイズは，ノイズが収束した次のフレームでは考慮されない．また，緩やかに光源や背景の環境が変化する場合，考慮する輝度値の差に対して閾値を設定する事で，それらのノイズを除去する事ができる．フレーム内での移動が発生した時のみ領域が抽出されるため，動作の始点と終点を求める事が容易であるが，いったん動作が止まるとその領域の連続性が失われるため，特定の領域に対して時間的な連続性を取得する事ができない．

3.2 フレーム間差分画像の時空間ボリュームの作成

3.1 で示したように，フレーム間差分法を用いて差分画像に現れる輝度値の分布として時系列画像から動きを抽出した．これをもとに，抽出された動きの領域から差分画像の時空間ボリュームを作成する．それぞれのボリュームをラベリングし個別の要素としてシーケンスから分離する．以下に時空間ボリュームの作成の流れを示す．

3.2.1 時空間ボリュームの作成

- **前処理**



図 2 フレーム内の動きの領域判定．差分画像（左）小領域での判定結果（右）
Fig. 2 domain judging of the motion in a frame(left)differential image(right)the judgment result in a small domain

使用する web カメラから取得される映像情報は，横×縦のサイズが 640 × 480 のカラー画像である．フレーム間差分を取る前に前処理として，カラー画像からモノクロ画像への変換を行った．また，今回考慮しているジェスチャは大きな身振りや手振りであるためそれほど高い解像度は必要ないと考え，横×縦が 160 × 120 の画像になるように解像度を下げている．

- **入力フレームの領域分割**
  次に，フレーム全体をいくつかの小領域に分割する．これはジェスチャを大まかな動きとして捉え，処理に要する情報量を削減する目的で行う．カメラの画角やカメラからの距離，取得するジェスチャの種類に応じて小領域のサイズを調整することが可能である．今回は (160 × 120) ピクセルの入力画像を，(10 × 10) ピクセルごとに 192 個の小領域に分割する．
- **動きが存在する領域を判定**
  各小領域内に含まれる画素値の差として抽出された変化領域の画素値に対して，閾値を設け，小領域の画素値が閾値以上（差分の急増・急減部分）となる領域を動きのある領域とする**図 2**．これによって，入力画像を (16 × 12) の小領域空間として解像度を下げた二値画像として扱う事ができる．また，入力画像に含まれる画素値は小領域空間の各点にマッピングされているため，より詳細な処理が必要になった場合など，多重解像度解析にも用いる事ができる．
- **ラベリング**
  ラベリングは，二値画像として表された小領域空間に対して 8 近傍探索を行い連結成分に分割する．入力画像のシーケンスから生成された，差分画像シーケンスの各小領域空間に対して全てラベリングを施す．従ってある小領域空間よりもあとの小領域空間に存

© 2010 Information Processing Society of Japan

在するラベルは必ず大きな値のラベル付けがなされる．

- **フレーム間でのラベルの対応付け**
  連続している二つのラベリングされた小領域空間 (N - 1) フレーム目と (N) フレーム目において，(N) フレーム目のラベル領域に対応する (N - 1) フレーム目の領域の８近傍を探索する．探索した領域にラベル領域が存在する場合に，(N - 1) フレーム目と (N) フレーム目のそれぞれのラベルを比較する．(N - 1) フレーム目のラベルが (N) フレーム目よりも小さい場合はそのラベルに合わせ，(N) フレーム目の連結成分を再ラベリングする．以上の処理を入力シーケンスに大して行う事で，それぞれ異なるラベルを付けられたボリュームとして動きを切り出す．

3.2.2 背景差分法との比較

図 3 は実際に上半身を左右に振る動作を映した動画像から作成した時空間ボリュームである．時間軸を横方向に取ったシーケンスの一連の流れを左に，右に正面からのスライス画像を示す．**図 3**(a) は，背景差分法を用いて生成したものである．正面画像からわかるように，抽出された上半身の中身が詰まった状態でボリュームが作成されている．**図 3**(b) と **図 3**(c) はフレーム間差分法を用いて作成したものである．**図 3**(a) がボリュームを上面から俯瞰したもので，**図 3**(b) は側面からボリュームを眺めたものである．矢印で示した部分でボリュームが疎になったり途切れている部分が見て取れるが，その部分が動きの急増減部分である．このように，フレーム間差分法で作成した時空間ボリュームは動きの始点と終点がはっきりと区別できるが，背景差分法で作成した時空間ボリュームは常に領域が連結している．

3.3 クラスタリング

3.2 で生成された差分画像について，一連のジェスチャが行われる中に共通の動きが存在する部分を見つける事ができればジェスチャの形態素として分類する事ができる．そこで，連続した動きを撮影した映像のシーケンスに対して，フレーム単位での自己相関を取る．シーケンス内に相関の高い部分が繰り返し現れた場合，それがジェスチャの形態素として独立した要素であると考えられる．

4. 実　　験

RPG におけるキャラクタの動作は，プレーヤが行うコントローラの操作と無関係である．しかし，プレーヤにはゲームの展開によってコントローラでの自然な動作入力が求められる．ゲームの面白さによって，この不自然さをプレーヤは克服してきた．この操作をジェスチャによって置き換えることができれば，センサ運動を入力するシステムが幅広い年齢層に受け入れられるように，RPG を楽しむプレーヤの幅を広げることができると考える．そこで，実験では RPG に求められるジェスチャを 4.1 のように想定し，この一部を実際に撮影した動画像群をデータとして用いた．

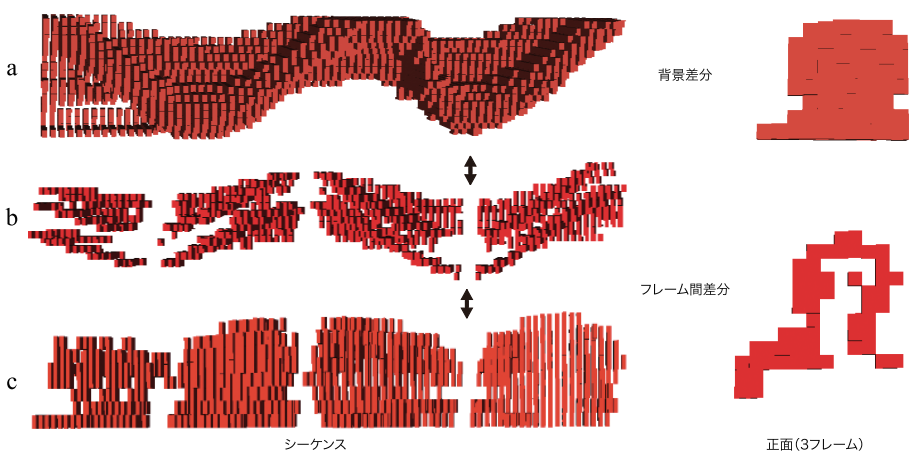


図 3 時空間ボリューム
Fig. 3 time-space volume

4.1 想定されるジェスチャ群

4.1.1 操作の分類

今回は代表的な RPG である NetHack の操作コマンド[6] の中から，抽象度の高い操作を抽出する．抽出された操作に対して，対応するジェスチャを決定する．まず，NetHack でプレーヤがキーボードコマンドによって行う操作（88 種類）を対象に，ゲーム進行上の操作文脈を一次的文脈依存と二次的文脈依存として２つに分類し，その文脈に対応する抽象度の高い操作のカテゴリの分類を行った (**表 1**)．

- **一次的文脈依存**
  ゲームのシナリオや時間軸とは関係なくプレーヤの必要に応じて操作が行われるものを，一次的文脈依存と呼ぶ事にする．これはプレーヤ主体の操作文脈と言える．
- **二次的文脈依存**
  ゲームのシナリオや時間軸にそって，キャラクタやアイテム，その他ゲーム内でイベントが存在する部分に対して操作が行うことを二次的文脈依存依存と呼ぶ事にする．これ

© 2010 Information Processing Society of Japan

表 1 NetHack コマンド分類
Table 1 classification of NetHack command

文脈	カテゴリ	No	操作	文脈	カテゴリ	No	操作	文脈	カテゴリ	No	操作	文脈	カテゴリ	No	操作
一時的文脈依存	システム操作	1	ゲームを保存する	二次的文脈依存	対象依存	23	杖を振る	二次的文脈依存	対象依存	45	アイテムの特別な力を使う	二次的文脈依存	移動	67	階段を降りる
		2	ゲームを中断する			24	食料を食べる			46	賂奪する			68	階段を昇る
		3	ゲームを終了する			25	薬を飲む			47	神に祈る		調べる	69	装備武器を調べる
		4	画面を再描画する			26	ランプをこする			48	シェルを実行する			70	装備指輪を調べる
		5	言語モード切り替え			27	座る		移動	49	上移動			71	装備道具を調べる
		6	ゲームを探索モードに切り替える			28	顔を拭く			50	左上移動			72	装備防具を調べる
		7	オプションを設定する			29	蹴る			51	左移動			73	装備魔除けを調べる
		8	ヘルプ			30	跳ねる			52	左下移動			74	周囲を探索する
	情報の表示	9	魔法一覧を見る			31	休む			53	下移動			75	回りを調べる
		10	アイテム一覧を見る			32	アイテムを置く			54	右下移動			76	足元を調べる
		11	識別した道具一覧を見る			33	勘定を払う			55	右移動			77	罠を調べる
		12	詠唱可能な魔法の一覧を見る			34	巻物や魔法書を読む			56	右上移動			78	画面の記号を調べる
		13	種類を選んでアイテム一覧を見る			35	武器を投げる			57	上ダッシュ			79	所持金を調べる
		14	ゲームの歴史を表示する			36	持ち物を調節する			58	左上ダッシュ		着脱	80	武器を装備する
		15	コンパイル情報を表示する			37	何かにアイテムを浸す			59	左ダッシュ			81	鎧を着る
		16	ゲームのバージョンを表示する			38	神に捧げる			60	左下ダッシュ			82	鎧を脱ぐ
二次的文脈依存	対象依存	17	扉を開ける			39	アンデッドを土に返す			61	下ダッシュ			83	アクセサリを身につける
		18	扉を閉める			40	魔物に名前をつける			62	右下ダッシュ			84	アクセサリを外す
		19	鍵をこじあける			41	種類を選んでアイテムを置く			63	右ダッシュ			85	罠を外す
		20	アイテムを拾う			42	地面に刻む			64	右上ダッシュ		不明	86	コマンドを繰り返す
		21	話しかける			43	魔法を唱える			65	特殊移動			87	メッセージを巻き戻してみる
		22	道具を使う			44	技量を高める			66	瞬間移動			88	autopickup切り替え

はゲーム内のキャラクタ主体の操作文脈と言える.

4.1.2 ジェスチャの決定

今回実験のために，表示されたものを消去するジェスチャとして連想される手を大きく振り払う動作（スイープ）と,「ランプをこする」「鍵をこじあける」「コマンドを繰り返す」から連想される，手を細かく左右に振る連続動作（シェイク）を設定した．これは片手で行うものとする．また「扉を開ける」や，反対に「扉を閉める」など，両手で表現する方が自然な場合も想定して，開いた両手を胸の前で合わせる動作（クローズ），逆に合わせた手を広げる動作（オープン），完全に両腕が交差する動作（クロス）の三種類を設定した.

5. 結果と考察

5.1 ジェスチャの取得とボリューム生成

4.1.2 で決定したジェスチャについて映像を取得し，ボリュームを作成した．このとき使用した映像は，繰り返しジェスチャを行い記録した 100 フレームのシーケンスである．スイープとシェイクについては速さを変えたジェスチャの比較を行った, 特にシェイクについては大きさの違いも比較している**図 4**．クローズ・オープン・クロスについては同様の速さで行ったジェスチャの比較を行った**図 5**.

5.2 速さによる特徴の考察

図 4(a)，(b) はスイープ動作のボリュームの俯瞰図である．**図 4**(a) は早くスイープ動作を行った場合. **図 4**(b) は遅く行った場合である．遅く動作を行う事で，時間軸に対して長くジェスチャが分布している．一旦動作が終了しているためジェスチャの開始点と終了点がはっきりと現れている．このとき，視覚的にはボリューム自体の幅に大きな変化は確認できない.

次に**図 4**(c)，(d) は細かいシェイク動作のボリュームの俯瞰図である．シェイク動作のように連続的に行われるジェスチャの場合，ボリュームがひと繋がりとなって作成されている．**図 4**(c) は動作を速く行っているため，**図 4**(b) のような遅い動作のように，左右の変動がほとんど現れず，細かい挙動がほとんど判断できない状態になっている．最後に**図 4**(e), (f) は大きなシェイク動作のボリュームの俯瞰図である．こちらも，細かいシェイク動作と同様の結果が現れている．この事から身振りの大きさに関わらず同じ動作をしている場合に現れる，ボリュームの形状的な特徴は同様な特徴として現れると考えられる.

© 2010 Information Processing Society of Japan

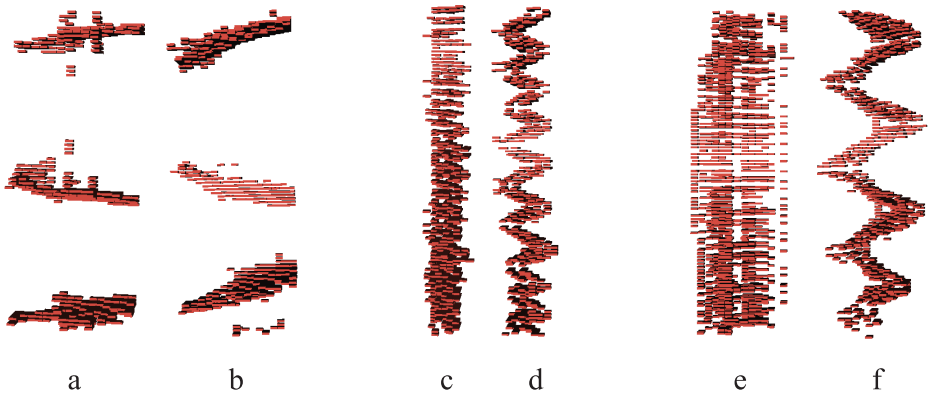


図 4 速さによる違い
Fig. 4 differnce of the result depended on speed

5.3 両手動作の考察

図 5 は両手動作のボリュームである．まず**図 5**(a) はクロスのボリュームを俯瞰したものである．両手がクロスして一度動作が停止するまで，そのままボリュームが連続し交わっている．クロス動作が終了した時点ではっきりと動作が途切れている事がわかる．**図 5**(b) はクローズのボリュームを俯瞰したものである．こちらはクロス動作と同様な見た目をしているが，胸の前で一度動作が止まるため，ボリュームが分割されている事がわかる．手の軌跡が似ているクロス動作とクローズ動作は，時空間ボリュームとして表すと明らかな特徴の違いがある事がわかる．また，**図 5**(c) はオープンのボリュームを俯瞰したものであるが，時間方向でのボリュームの形状がクローズ動作の反対動作である事がわかる．

6. おわりに

本稿では，Web カメラにより取得したビデオシーケンスから映像データベースを作成し，それを用いたジェスチャの認識を実現するために，フレーム間差分法を用いて作成した時空間ボリュームとしてビデオシーケンスから身体動作の要素を分離する手法について説明した．実際にある RPG の操作から具体的にジェスチャを想定し，時空間ボリュームを作成した．ボリュームを視覚的に評価することで分離される動きの特徴を一部であるがつかむ事ができた．現状は視覚的な評価にとどまっているため，つかみきれていない特徴が存在すると考えられる．今後は，統計的な処理などを行いよりいっそう詳細な特徴の抽出に取り組む予定である．

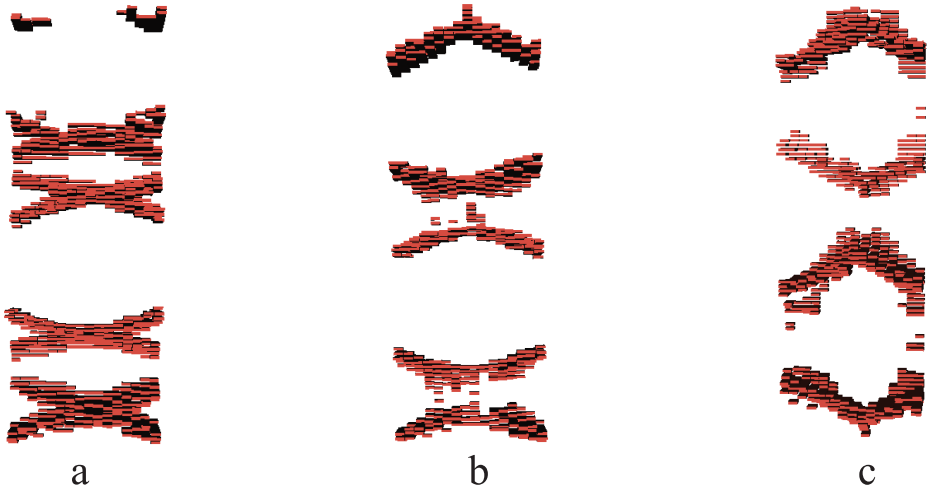


図 5 領域連結の例
Fig. 5 example of area connection

謝辞 本研究は大阪大学名誉教授・北橋忠宏博士に非常に多くのご指導・ご助言をいただいた．ここで謹んで感謝の意を表する．

参 考 文 献

1) 中村 卓: 手の動きを利用したメニュー操作による大画面インタラクション手法, 筑波大学システム情報工学研究科修士論文, 2008.
2) 木谷 クリス 真実, 岡部 孝弘, 佐藤 洋一, 杉本晃宏: 視覚的文脈を考慮した人物動作カテゴリの教師無し学習, 第 11 回画像の認識・理解シンポジウム (MIRU2008) 論文集, pp.28–33, 2008.
3) 西尾 章治郎, 岸野 文郎, 塚本 昌彦, 山本 修一郎, 石田 亨, 川田 隆雄: 相互の理解, 岩波書店, 1999.
4) ブレンダ ローレル: 人間のためのコンピューター, アジソン・ウェスレイ・パブリッシャーズ・ジャパン, pp.286–287, 1994.
5) 磯 悠記, 床井 浩平: VR システムを用いたコミュニケーションツールの開発, 映像情報メディア学会技術報告, VOL. 30, NO. 17, pp.61–62, 2006.
6) JNetHacker's Bible Project: コマンド, JNetHacker's Bible, http://jnhbible.s6.xrea.com/nh/command.htm, 2002.

© 2010 Information Processing Society of Japan