

# 自然な実写アバタに向けた身体動揺の計測と再現

## Measuring and Reproducing Postural Movement for Natural Image-based Avator

宮内 翼<sup>1\*</sup> 吉村 宏紀<sup>1</sup> 西山 正志<sup>1</sup> 岩井 儀雄<sup>1</sup>  
Tsubasa Miyaochi<sup>1</sup> Hiroki Yoshimura<sup>1</sup> Masashi Nishiyama<sup>1</sup> Yoshio Iwai<sup>1</sup>

<sup>1</sup> 鳥取大学 工学部  
<sup>1</sup> Tottori University

**Abstract:** In order to increase the realism of human interaction systems, we propose a method to reproduce the same kind of postural movement for image-based avatar. We exploit a postural movement during standing posture, and we measure the amplitude and the phase for each part of a body from a video sequence. The video transition with the amplitude and the phase produce the postural movement of image-based avatar in any length of time. We performed a subjective assessment that the proposed method is less discontinuous than a simple repeat method.

### 1 はじめに

近年、人間と機械とがインタラクションを行うシステムに向けてロボットやアバタが数多く開発されており、エンタランスでの案内など社会の様々な場面で応用が期待されている。例えば、個々の人間を認識してショッピングモールを案内するロボット [1] やユーザと音声会話をする施設案内のアバタ [2] などが実現されている。これらのインタラクションシステムを用いることで、場所や時間帯の制約に関係なくいつでも自動で案内をユーザに対して行うことができる。本研究では、このようなインタラクションシステムの中でも、大型ディスプレイ上に実際の人間を映し出す実写アバタに注目する。実写アバタはロボットと比べ、ハードウェア面でのメンテナンスの手間が少なく、既設の大型ディスプレイを活用することが可能であると考えられる。

インタラクションシステム上で自然な実写アバタを再現するためには、インタラクションの状態に加えて待ち状態を取り扱う必要がある。実写アバタは常に人間とのインタラクションを行っている訳ではなく、インタラクションの前や後に待ち状態が存在する。待ち状態を考慮しないと実写アバタはユーザが存在しない時にどのような行動をすれば良いか分からず、不自然なインタラクションになる。従来の実写アバタ [3][4] はインタラクションの状態を取り扱っているものの、待ち状態を取り扱っていなかった。そこで本研究では、待ち状態の実写アバタを高い臨場感で再現する課題に取り組

む。ここでは待ち状態として、エンタランスなど受付での対人インタラクション時に自然に発生する直立姿勢について議論する。

直立姿勢の人間が待ち状態でどのような動きをするかを観察すると、独特で複雑なリズムにしたがって絶えず前後左右、様々な方向に動いている。これは身体動揺 [5] と呼ばれており、一部の筋肉に負担が掛からないように上手く負担を分散するよう人間は動いている。

そこで、本研究では、インタラクションシステムの臨場感を高めるために、あたかも人間と同じような身体動揺を実写アバタで再現する手法を提案する。本研究は直立姿勢時の身体動揺を対象とし、体の部位毎に揺れの大きさと周期を映像中の人物領域から計測する。揺れの大きさと周期を用いた映像遷移で実写アバタの身体動揺を任意の時間長で生成する。提案手法は単純な連続再生の手法と比べて不連続と感じる部分が少なく実写アバタの臨場感が高まることを主観評価で確認した。

### 2 関連研究

アバタと人間が自然にインタラクションするには、アバタが人間に高い臨場感を与える必要がある。アバタが高い臨場感を与えるには音声や画像の生成技術が必要であるが、特に画像生成はアバタの見た目の動きに直結するため重要な技術である。画像生成にはアバタの質感再現とアバタの動作生成が必要である。アバタの質感を高めるために、実際の人間を映した実写画像を用いたアバタ [3][4] を使う手法がある。一般的にはコンピュータグラフィックスで生成したCGアバタ [6][7][8][9]

\*連絡先：鳥取大学工学部知能情報工学科  
〒680-8550 鳥取県鳥取市湖山町南4丁目101  
E-mail: nishiyama@eecs.tottori-u.ac.jp  
iwai@ike.tottori-u.ac.jp

を用いることが多いが、実写アバタを用いることでCGアバタよりも人間の見た目に近く、ユーザに対して高い臨場感を与えることができる。アバタの自然な動作を生成するためには、顔の表情を変化させる手法や体の姿勢を変化させる手法などが必要である。表情を変化させる技術には、テキストに合わせて音声とそれに合わせた顔画像を生成する手法 [7] が提案されている。姿勢を変化させる技術として、人間とインタラクションを行う際に感情を表現する手法 [8][9] などが提案されている。表 1 のように既存手法はインタラクション状態を取り扱っているものの待ち状態を考慮していなかったため、提案手法では高い臨場感の実写アバタを実装することを目的に待ち状態を取り扱う。

表 1: 提案手法の位置付け

	CG アバタ	実写アバタ
インタラクション状態	[6][7][8][9]	[3][4]
待ち状態	-	提案手法

### 3 身体動揺の計測

#### 3.1 計測方法

実写アバタで身体動揺を再現するには、映像中の実写アバタの見た目を考慮する必要があるため、身体動揺を重心の移動から起きる身体全体の揺れで計測することが望ましい。さらに実写アバタに身体動揺を再現する際に身体の部位毎に揺れの特徴がないかを確認する必要があり、身体の部位毎に揺れを計測する必要がある。身体の揺れには周期と大きさがあり、揺れの大きさは身体の部位毎で異なり身体の上の方が大きく、身体の下の方が小さいと考えられる。しかし身体動揺の周期は部位毎でも同じであると考えられる。本研究では、この部位毎の揺れの大きさと周期の特徴を身体動揺を計測して確認する。

身体動揺の計測には、一般的に重心動揺計 [10] を用いることが多い。この手法は、足元の重心の変化を計測するため身体全体の揺れを計測することができない。身体の揺れを計測するために、加速度センサ [11] や角速度センサ [12] を身体に装着して計測する手法がある。これらの手法は、実写アバタで用いる画像と撮影するカメラの同期を取るためのキャリブレーションが困難であったり、身体に接触した状態で計測するため実写アバタを表示する際に機材が見えてしまう問題がある。そのため画像とのキャリブレーションが容易であり、身体と非接触で揺れを計測する手法が望ましい。非接触での身体の姿勢を所得する手法としては距離センサを

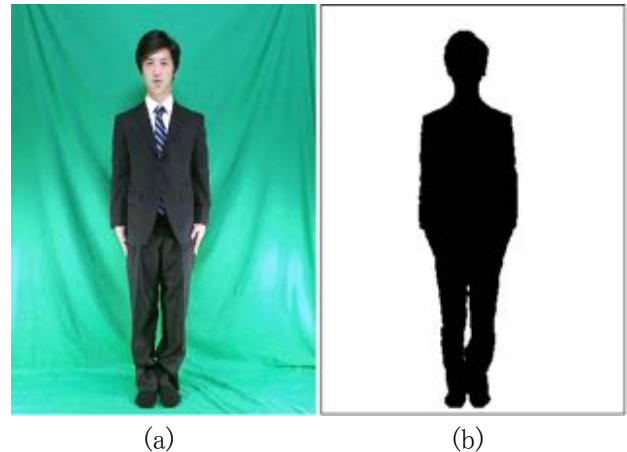


図 1: 人物領域抽出

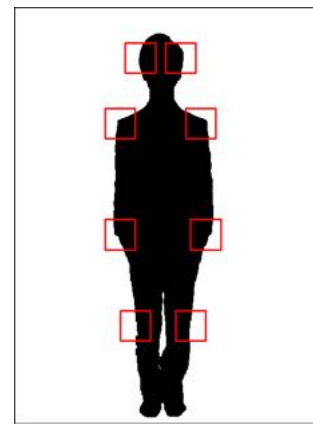


図 2: 重心動揺を計測する部位

用いて人間の関節を取得することも考えられる。しかし、この手法は誤差が大きく、小さな動きである身体動揺を計測するには不向きである。そこで同期のためのキャリブレーションが必要なく、カラー画像のみを用いた計測に注目する。以下では、カラー画像を用いて身体動揺を計測する手法について述べる。

#### 3.2 計測アルゴリズム

本研究で提案する身体動揺の計測は、人物領域を抽出したマスク画像を比較することで行う。マスク画像は、図 1 の (a) の撮影画像から被写体の人物領域を抽出した図 1 の (b) の 2 値化画像である。人物領域を抽出する方法は背景差分を用いる。背景差分は、まず図 1 の (a) の画像からグリーンバックの色を RGB の範囲で指定して消去する。しかし、これだけではすべてのグリーンバックの色を網羅することができず背景が残ってしま

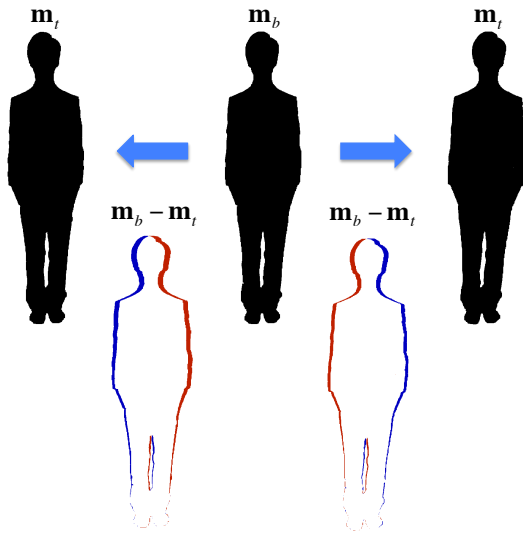


図 3: マスク画像の比較 (赤い部分が正の値であり, 基準画像では人物領域で比較画像では背景領域の部分. 青い部分が負の値であり, 基準画像では背景領域で比較画像では人物領域の部分.)

う. そこで被写体が写っていない画像と被写体が写っている画像を比較して消去しきれなかった背景を消去する. 以上の手順で人物領域を抽出し, 2 値のマスク画像を作成する.

作成した直立姿勢のマスク画像から, 身体動揺による揺れの中心の位置にある画像を基準画像として選択する. 選択した基準画像とフレーム毎のマスク画像を比較し, ハミング距離が 1 のピクセル数を計測することで身体動揺を計測する. マスク画像の比較では, 式 (1) を用いて計算する. 式 (1) の  $m_b$  が基準画像,  $m_t$  がある時刻の比較画像であり, parts は図 2 のように身体の部位の範囲を指定する. これにより身体動揺を部位毎で計測することが可能である.

$$d_i = \text{parts}(m_b - m_t), \quad t = 1, \dots, N \quad (1)$$

基準画像と比較画像の比較では, 図 3 の赤い部分のように基準画像では人物領域で比較画像では背景領域の部分と図 3 の青い部分のように基準画像では背景領域で比較画像では人物領域の部分がある. このピクセル数を見ることで身体動揺による移動方向が分かる. 例えば, 図 3 のように身体全体が左に移動する場合は, 身体の左の部分では青い部分, 身体の右の部分では赤い部分になる. 身体全体が右に移動する場合は, この逆である. 以上の手順で身体動揺の計測を行う.



図 4: 被写体の撮影状況

### 3.3 計測評価実験

実験の手順は, まず被験者の 3 分間の直立姿勢を撮影した. その後, 撮影した画像を用いて人物領域を抽出したマスク画像を作成した. その画像を用いて提案手法で身体動揺の計測を行った. 被験者は男性 4 名, 平均年齢  $21.5 \pm 0.5$  歳である.

撮影は被写体一人, 機器の操作者一人の合計 2 名で行い, 撮影機器はカメラ, 照明 2 台, パソコン (Windows 8.1 搭載) を使用した. 撮影場所は, グリーンバックを背景とした外からの光を通さない暗室である. カメラから取得するデータは 30fps で被写体を写した図 1 の (a) のようなカラー画像である. 被写体は図 4 のようにグリーンバックの前に立ち, 高さ 89cm に固定したカメラを被写体から 205cm 離れた場所に設置した. また 2 台の照明を両サイドからグリーンバックに向けることで, グリーンバックに写る被写体の影を消した. これによって人物領域抽出の際に用いる背景差分の精度を向上させることが可能である. さらに人物領域抽出の際にオープニングとクロージングをすることで背景のノイズを取り除いた. 以上の条件で撮影を行った.

3 分間の直立姿勢の撮影画像から人物領域を抽出したマスク画像を作成し, マスク画像を用いて身体動揺の計測を行った. 身体動揺の計測は, 図 2 のように頭, 肩, 手, 足を幅 70[pixel], 高さ 70[pixel] の範囲を手動で指定して計測を行った. 以上の条件で 3.2 節の手法を用いて計測を行った.

### 3.4 結果と考察

身体動揺の計測の結果が図 5 となり, 左部位では, 右に移動した時が正の値であり, 左に移動した時が負の値である. 右部位では, その逆である. 図 5 から被験者 A では身体動揺の振幅が大きく, 被験者 B では身体動揺の振幅が小さい. さらに, 被験者 A では身体動揺の周

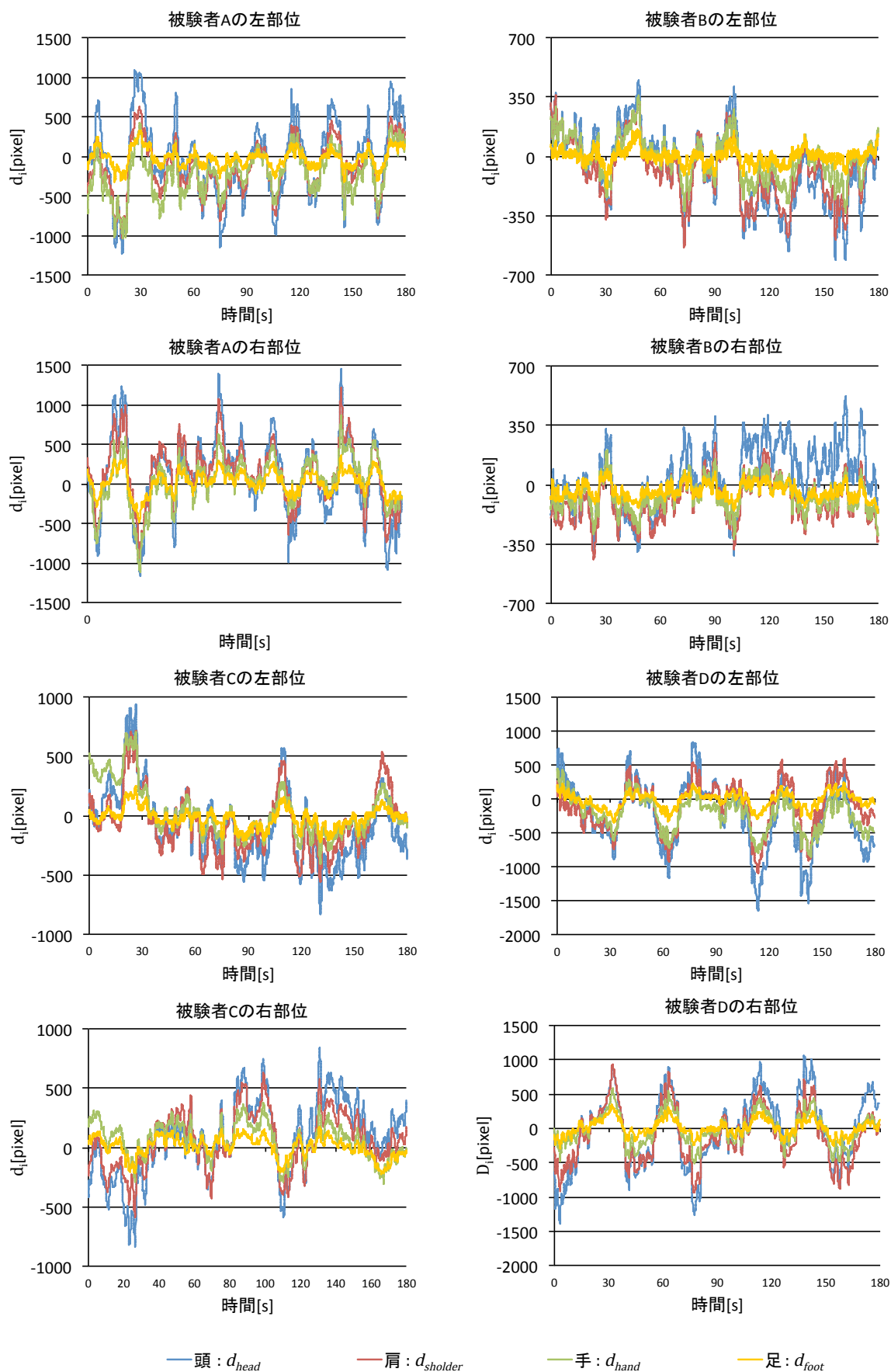


図 5: 身体動揺の変化のグラフ

期が短く、被験者 B では身体動揺の周期が長い。このことから身体動揺の揺れは振幅と周期に個人差があることが分かる。そして被験者 A と被験者 B の部位毎の揺れは、それぞれ振幅の大きさは異なっているが、部位が連動して動いているため部位毎で周期は同じであることが確認できた。この特徴は、他の被験者でも同様の結果であった。部位毎の振幅の大きさは、表 2 の絶対値の平均と表 3 の標準偏差から身体の上の部位が大きく揺れ、身体の下の方が小さく揺れている。以上のことから身体動揺には個人差があるが、個人の身体動揺の共通点として部位毎の揺れの振幅は身体の上が大きく、身体の下が小さい。そして、部位毎の揺れの周期は同じであることが確認できた。

提案手法の計測の精度は、背景差分の精度によって大きく変化するため背景差分の人物領域抽出誤差を計測した。誤差は、背景差分で人物領域を抽出する方法と人手で人物領域を抽出する方法を比較する。ここでは、人手で人物領域を抽出した方法を真値として、FPR(False Positive Rate) と FNR(False Negative Rate) を誤差とした。FPR は人手の方法で背景領域だった部分が背景差分の方法では人物領域を表している値であり、FNR は人手の方法で人物領域だった部分が背景差分の方法では背景領域を表している値である。誤差の結果が表 4 である。表 4 から誤差は表 2 と表 3 の値より十分に小さく、身体動揺の計測が誤差にそれほど影響されていないことが分かる。

表 2: 計測値の絶対値平均 [pixel]

被験者	左頭	右頭	左肩	右肩	左手	右手	左足	右足
A	389.8	400.0	293.5	319.2	287.1	236.6	93.6	117.2
B	184.2	157.1	156.5	124.8	104.1	87.2	34.0	52.7
C	256.7	258.7	205.1	196.1	163.4	130.6	65.4	70.5
D	418.0	426.3	290.0	346.0	259.5	188.1	96.3	96.0
全体	312.2	310.5	236.3	246.5	203.5	160.6	72.3	84.1

表 3: 計測値の標準偏差 [pixel]

被験者	左頭	右頭	左肩	右肩	左手	右手	左足	右足
A	483.3	489.3	328.4	371.7	283.7	301.1	116.7	146.0
B	213.5	184.9	176.2	122.7	127.3	88.2	45.9	48.3
C	306.2	312.1	253.4	233.0	215.7	145.4	82.7	86.5
D	491.5	511.6	348.5	398.2	254.1	227.7	118.8	119.4
全体	373.6	374.5	276.6	281.4	220.2	190.6	91.0	100.1

表 4: 背景差分による身体領域抽出誤差 [pixel]

誤差	左頭	右頭	左肩	右肩	左手	右手	左足	右足
FPR	5.9	6.5	7.4	32.9	22.8	30.1	17.5	62.6
FNR	65.0	50.4	41.3	12.4	38.1	17.1	33.5	8.0

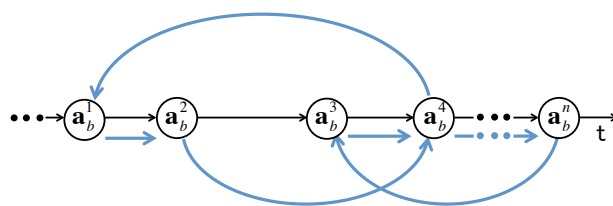


図 6: 身体動揺の再現の概念図 (図中の  $a_b^j$  は  $j$  番目の基準画像を表す。黒矢印は、時間軸の連続した画像列、青の横矢印が基準画像から基準画像までの画像列、青の円弧矢印が画像列の切り替えを意味する)

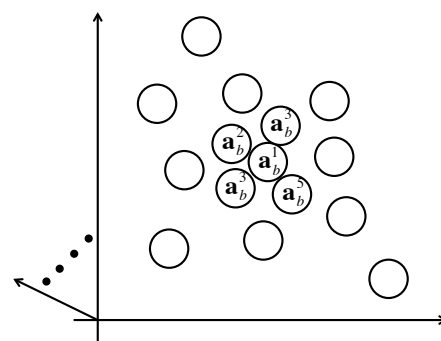


図 7: 高次元空間における身体動揺の画像の分布 (図中の丸はある時刻の画像を表し、 $a_b^1$  が開始画像、 $a_b^k$  が  $k$  番目の基準画像である)

## 4 実写アバタによる身体動揺の再現

### 4.1 再現方法

ここでは、実写アバタに揺れの大きさと周期を用いた映像遷移で任意の時間長の身体動揺を再現することを考える。身体動揺の揺れは、3.4 節から身体部位毎に振幅は異なるが、周期は同じであり部位毎の揺れが連動して動いていることがわかった。そのため、身体部位毎に揺れを再現するのではなく、身体全体で揺れを再現することができる。提案手法は、図 6 のように基準画像から基準画像までの画像列を繋げることで実写アバタの身体動揺を再現する。揺れの方向や振幅が異なる画像列や揺れの周期が異なる画像列を繋げることで様々な揺れと任意の時間長の身体動揺を再現できる。基準画像は、開始画像の類似画像である。開始画像とは、図 7 のように画像中の類似度の中心にある画像、つまり身体動揺の揺れの中心の画像である。まず開始画像との類似度を測り、類似度が高い画像を基準画像とする。そして、開始画像も基準画像として用いて図 6

のように画像列を繋げる。基準画像に類似画像を用いることで画像列の切り替え時に違和感なく次の画像列に繋げることができる。

類似度は、人物の見え方の類似度と人物の輪郭の類似度を考慮する必要がある。人物の見え方の類似度は、人物の瞬きや手の向きなど人が見て違和感を感じる画像を考慮する。人物の輪郭の類似度は、直立姿勢の揺れの位置が異なっている画像を考慮する。この2つの類似度を考慮することで、画像列を切り替える際に違和感のない画像を選択することができる。提案手法では、以上の流れで実写アバタに身体動揺を再現する。

## 4.2 再現アルゴリズム

まず直立姿勢の画像から開始画像を選択し、その類似画像を探索する。人物の見え方の類似度は、人物領域を抽出したカラー画像を用いて比較的人間が判断する類似度に近いSSIM(Structural SIMilarity)[13]で測る。人物の見え方の類似度の計算は、式(2)である。式(2)の $\mathbf{a}_b$ は基準画像のカラー画像、 $\mathbf{a}_t$ はある時刻の比較画像のカラー画像である。人物の輪郭の類似度は、人物領域を抽出したマスク画像を比較して基準画像と比較画像の差で測る。人物の輪郭の類似度の計算は、式(3)である。式(3)の $\mathbf{m}_b$ は基準画像のマスク画像、 $\mathbf{m}_t$ はある時刻の比較画像のマスク画像、 $\lambda$ は定数である。画像の類似度は、人物の見え方の類似度 $S_a$ と人物の輪郭の類似度 $S_s$ を用いて式(4)で計算する。以上の手順で類似度が高いものを基準画像とする。

$$S_a = \text{SSIM}(\mathbf{a}_b, \mathbf{a}_t) \quad (2)$$

$$S_s = e^{-\lambda \|\mathbf{m}_b - \mathbf{m}_t\|} \quad (3)$$

$$S = (1 - \alpha)S_a + \alpha S_s \quad (4)$$

基準画像と基準画像までの画像列で様々な揺れを選択し、自由に組み合わせることで実写アバタに連続的な画像列で身体動揺の揺れを任意の時間長で再現できる。

## 4.3 再現評価実験

実際に提案手法で実写アバタの揺れを再現した。実験で使用した実写アバタの開始画像の人物の見え方の類似度と人物の輪郭の類似度のグラフが図8である。実験では、式(4)の $\alpha$ の値を0.5とし、人物の見え方の類似度と人物の輪郭の類似度が両方高いものを基準画像とした。基準画像を用いて提案手法で実写アバタに身体動揺を再現した。

提案手法で作成した実写アバタを用いて、主観評価を行った。主観評価は、直立姿勢時に起きる身体動揺を提

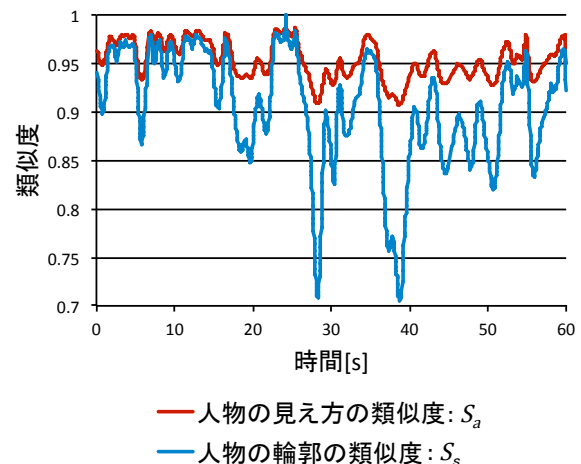


図 8: 身体動揺における時間方向の変化

案手法で再現した実写アバタと比較手法の直立姿勢の実写アバタを比較し、人間的な動きと感ずる手法を評価した。動画は、図9のような以下の4つを使用した。

### 提案手法

1. 複数の基準画像から基準画像までの画像列を組み合わせた動画

### 比較手法

2. 単一の基準画像から基準画像までの画像列を繰り返し再生した動画
3. 基準画像ではない画像で画像列を切り替えた動画
4. 実写アバタが全く動かずに直立姿勢をしている動画

動画1は、提案手法であり実写アバタの様々な揺れを組み合わせた動画である。動画2、動画3、動画4は比較手法であり、それぞれ動画2が実写アバタが同じ揺れを繰り返す動画、動画3が画像列を切り替える際に実写アバタが不自然に動く動画、動画4が実写アバタが全く動かない動画である。以上の4つを一対比較をし、動画の全組み合わせをディスプレイにランダムに2つ表示させて評価を行った。

実験の手順は、まず被験者をディスプレイから1m離れた印のある場所に立たせる。そしてディスプレイに動画を左右に2つ並べて、左から交互に2回ずつ再生する。その後、被験者は2つの動画からより人間的に自然な動きのものを選択してもらう。これを動画の全組み合わせで行った。実験の被験者は、男性7名、女性1名、平均年齢 $22.4 \pm 1.1$ 歳である。評価方法はサーストンの一対比較を用いた。

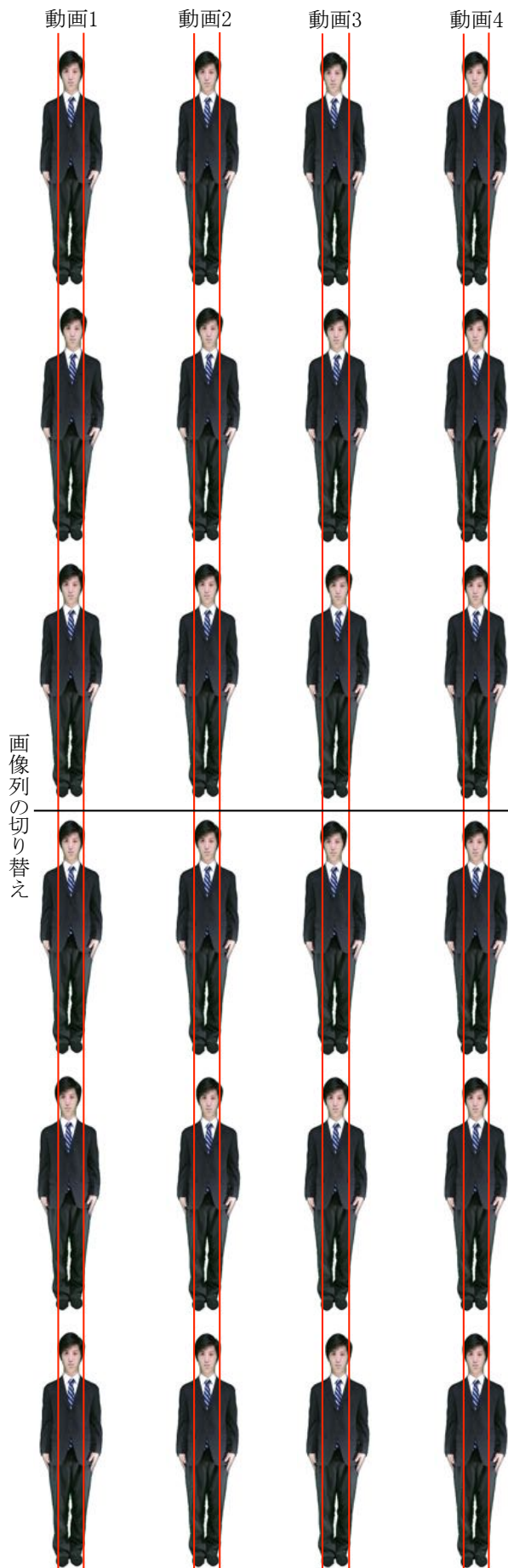


図 9: 一対比較で用いた動画

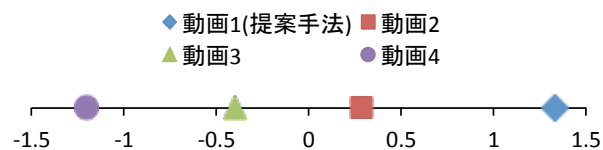


図 10: サーストンを用いた一対比較

#### 4.4 結果と考察

実験の主観評価の結果が図 10 であり、大きい値が評価が高く、小さい値が評価が低くなる。図 10 から動画 1、動画 2、動画 3、動画 4 の順で評価が高くなっている。動画 1 は提案手法で作成した動画であり、比較手法の動画 2、動画 3、動画 4 より提案手法の評価が高い結果になった。動画 2 は単純な動作の繰り返しであったために身体動揺の複雑な動きを再現できなかった。そのため提案手法より評価が低かったと考えられる。動画 3 は、画像列の切り替えの際に不自然に実写アバタが動き、人間の動きとして起こりえない動きをしたため、評価が低かったと考えられる。動画 4 は、実写アバタが全く動かずにいることが、人間に違和感を感じさせた。

以上のことから、提案手法は比較手法より人間的に自然な動きをしていることを確認した。

### 5 まとめ

本研究では、実写アバタの臨場感を高めることを目的に、身体動揺を計測し再現する手法を提案した。身体動揺の計測では、部位毎の揺れの大きさと周期を計測し、身体の上の部位が大きく、身体の下の方が小さく揺れ、部位毎に周期が同じで連動して動くことを確認した。身体動揺の再現では、揺れの大きさと周期を用いた映像遷移で実写アバタの身体動揺を再現した。実際に、提案手法で身体動揺を再現した実写アバタが比較手法と比べて人間的な動きであることを主観評価で確認した。

今後の課題として、まだ一人分の実写アバタしか身体動揺を再現できていないため、他の実写アバタで身体動揺を再現することなどがあげられる。

### 参考文献

- [1] 宮下 善太, 神田 崇行, 塩見 昌裕, 石黒 浩, 萩田 紀博, 顧客と顔見知りになるショッピングモール案内ロボット, 日本ロボット学会誌, vol. 26, No. 7, pp.821-832, 2008.

- [2] 大浦 圭一郎, 山本 大輔, 内匠 逸, 李 晃伸, 徳田 恵一, キャンパスの公共空間におけるユーザ参加型双方向音声案内デジタルサイネージシステム, 人工知能学会誌, 28 巻, 1 号, pp. 60-67, 2013.
- [3] 原 健太, 堀 磨伊也, 武村 紀子, 岩井 儀雄, 佐藤 宏介, 実画像アバタを用いた対人インタラクションシステムの構築, 電気学会論文誌 C, Vol. 134, No. 1, pp. 102-111, 2013.
- [4] Andrew Jones, Jonas Unger, Koki Nagano, Jay Busch, Xueming Yu, Hsuan-Yueh Peng, Oleg Alexander, Paul Debevec, An Auto-Mutiscopic Projector Array for Interactive Digital Humans, ACM SIGGRAPH Emerging Technologies, No. 6, 2015.
- [5] 青木 俊幸ら, 「人間計測」, pp. 793-798 朝倉書店, 2003.
- [6] 高橋 範子, 小野 光貴, 渡辺 富夫, 石井 裕, 看護実習生-患者役アバタを介した看護コミュニケーション教育システム, 人間工学, Vol. 50, No. 2, pp. 84-91, 2014.
- [7] Robert Anderson, Bjorn Stenger, Vincent Wan, Roberto Cipolla, Expressive Visual Text-To-Speech Using Active Appearance Models, CVPR, pp. 3382-3389, 2013.
- [8] 中野 敦, 塩入 健太, 星野 准一, CG キャラクターの心理状態を用いたコミュニケーションゲーム, 情報科学技術レターズ 4, pp. 193-196, 2005.
- [9] Dumebi Okwechime, Eng-Jon Ong, Andrew Gilbert, Richard Bowden, Social Interactive Human Video Synthesis, ACCV, Volume 6492, pp. 256-270, 2010.
- [10] 山本 昌彦, 吉田 友英, 重心動揺計を用いた体平衡機能検査 -重心動揺検査・電気性身体動揺検査-, Equilibrium Research, Vol. 70, No. 3, pp. 135-144, 2011.
- [11] 竹内 弥彦, 下村 義弘, 岩永 光一, 勝浦 哲夫, 小型三軸加速度計による高齢者の動的バランス評価の有用性, 理学療法科学, Vol. 22, No. 4, pp. 461-465, 2007.
- [12] 大西 智也, 橘 浩久, 武田 巧, 安静な立位における足位の違いが下肢帯および下肢の動揺に及ぼす影響, 理学療法科学, vol. 30, No. 2, pp. 313-316, 2015.
- [13] Z. Wang, A. C. Bovik, H. R. Sheikh, E. P. Simoncelli, Image quality assessment: From error visibility to structural similarity, IEEE, Transactions on Image Processing, vol. 13, No. 4, pp. 600-612, Apr. 2004.