人体動作取得ソフトウェアの試作

竹安　真也†　　小島　一成††　　長江　貞彦†
†近畿大学大学院生物理工学研究科
††立命館大学アート・リサーチセンター

概要：筆者らは動画中の人体動作の取得におけるソフトウェアの開発を行っている．人物の姿勢データの取得は、インタラクティブに 2 次元画像に 3 次元モデルを重ね合わせる手法を用いている．この処理を全てのフレームに対して作業を行うと膨大な作業量となるため、差分処理とオプティカルフローからキーフレームを決定した。一般的に，人体動作データの取得には専用の機材や広い撮影場所が必要であるが，本手法によって、映像として記録された動画中の人体動作の取得が可能である．

A Development of human motion acquisition Software

Shinya TAKEYASU†　Kazuya KOJIMA††　Sadahiko NAGAE†
†Graduate School of Biology-Oriented and Technology, Kinki University
†† Art Research Center, Ritsumeikan University

ABSTRACT：We are developing software to acquire the data of the human's motion. In order to get the person's posture data, we take the method for overlapping that three-dimensional human model to the two-dimensional image, which are taken by the monocular image. Furthermore we have reduced the amount of work and volume of data by using the key frame method in that animation. Generally speaking, the special apparatus and wide-range space are necessary to get the above mention data, however the software that we are now developing can be free from that problem.

1.　はじめに

無形文化財である伝統芸能は歴史的に価値があり，現在では，ディジタルアーカイブ研究として伝統芸能の記録･保存・再現が行われている．今日までは資料や文献，映像などを記録･保存し，コンテンツが制作されている．

伝統芸能，とりわけ，舞踊－踊りの身体動作が最も重要な要素である．舞踊の身体動作の研究として，モーションキャプチャシステムを用いて踊り動作を計測し，動作の分析・解析や 3 次元モデルでの再現が行われている．これにより，舞踊の身体動作に関する研究がさらに高まり，舞踊のディジタルアーカイブによる文化の普及が望まれる．

過去に撮られた映像から舞踊の身体動作の記録は，画像処理による身体動作取得の研究が行われている．例えば，人物の輪郭線から特徴点を抽出して姿勢の推定が行われている[1]．また，多眼視画像を用いた人物の姿勢推定が行われ[2]，また単眼視画像からの人物の推定も行われている[3]．これらはオクルージョンに対処しておらず，また特定の設備や画像以外からの情報も必要である．そのため，単眼視画像からの人体動作の抽出が必要である．

筆者らは，動画中の人物の姿勢推定を行うソフトウェアを開発している．推定方法として，インタラクティブに単眼視画像中の人物に 3 次元の人体モデルを重ね合わせるモデルマッチング法を用いている．動作の再現を行うために，全静止画に対して姿勢の取得を行うと膨大な作業量となる．そこで，キーフレーム法を用いる[4]．

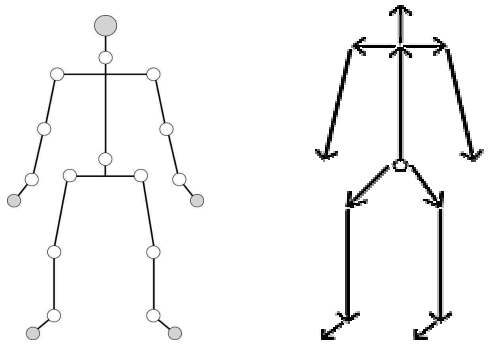
本研究では，時系列に連なる静止画像よりキーフレームを検出する手法を検討する．まず，キーフレームとする人物の姿勢を決定するため．差分処理を用いた．次に，人物の動作について注目し，オプティカルフローを用いて，人物の移動量について算出した．その結果，キーフレームとなる画像の検出できた．

２．　人物の姿勢推定ソフトウェア

人体動作を取得には，人物の姿勢抽出が必要である．そこで，筆者らが開発した人物の姿勢抽出を行うソフトウェアについて説明する．

２．１　姿勢抽出に用いる人体モデルの構成

人体の姿勢や動作は骨格や関節，筋肉など様々な要因により形作られるために複雑である．それに伴う人体構造を精密に再現することは困難である．例えば，人体の部位は伸縮することで人体部位の長さが変化する．人体構造は関節でつながれた剛体軸の連結体とみなされている[5]．剛体とはいかなる 2 点間の距離も変化しない物体のことであり，その剛体の運動において物体内のいかなる 2 本の線の角度は保たれる．これらのことをもとに関節を持つ 3 次元人体モデルを製作した．図 1(a)に人体モデル構造をスケルトンモデルとして示す．人体モデルには必要最低限と考えられる 3 つの自由度を持つ関節 10 箇所，1 つの自由度を持つ関節 4 箇所，計 14 箇所の関節を備えた．人体モデルは階層構造を用いて構築した．各関節のつながりを図 1(b)に示す．階層構造は階層上層部の動作が下層部へ影響を及ぼし，下層部の動作は上層部へ影響を与えない．



(a)スケルトンモデル　　(b)階層構造

図 1　人体モデル

人体姿勢データの抽出に用いる人体モデルは被験者に近い体格であることが望ましい．そこで，日本人の人体計測データより日本人男性の標準的な体格のデータを用いて人体モデルを製作した[6]．使用する日本人の体格データのうち主なデータを表 1 に示す．標準体型である人体モデルを用いることで，モデルマッチングがより正確に行える．

表 1　日本人男性(成人)の標準体格データ

(単位:mm)

測定箇所	長さ
身長	1706
肩･指先	727．0
肩･ひじ	317．1
ひじ･手首	255．6
股下	759．7
ひざ･くるぶし	454．5
くるぶし･床	46．6

２．２　ソフトウェアの概要

2 次元画像から 3 次元データである人体の姿勢データを取り出す場合，奥行き情報をどのように取り出すかが問題となる．小島によって，階層的な人体のモデルを用い，単眼視画像から人物の姿勢を推定する方法を提案している[7]．2 次元画像の人物に 3 次元の人体モデルを重ね合わせ，人物と同姿勢にする．これにより，人体モデルの姿勢を人物の姿勢とみなす．そして，人体モデルから姿勢データを取得する．この際，人体モデルの各関節間の距離は人物と同じにする必要がある．人物と人体モデルの骨格構造が異なることで，正確な姿勢データの取得が困難となる．図 2 に示すように異なる長さの 3 次元物体は，傾き角度の違いにより 2 次元画像に投影された長さが同じになることがある．そのため，本ソフトウェアでは人体モデルの各関節間の長さを調節することができる．その際，人物の体格を知るために正面画像が必要となる．

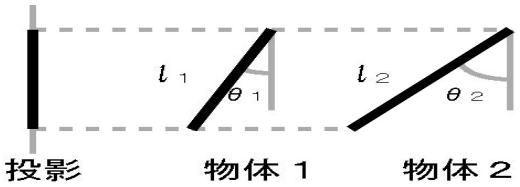


図 2　長さが異なる物体の投影

人物と同じ姿勢にするために人体モデルの各関節に対して回転動作の操作を行う．これは，人体モデルの関節動作を実際の人と同様の動作にすることで，利用者が人体モデルの操作を感覚的に行える．人体モデルの関節は実際の人と同様の可動域をもつ．表 2 には人間の主な関節についての可動域を示す[8]．こ

れにより，人体モデルは実際の人には不可能な姿勢にならない．

表 2　関節可動域

関節名	可動範囲
肩(前回転)	0~180
肘	0~145
股(前回転)	0~125
膝	0~130
腰(前屈･後屈)	0~75

本ソフトウェアで取得するデータは関節の回転角度と関節部分の座標値である．被験者と異なる体格である人体モデルに対して動作の再現を行うには関節角度データを用いる．また，人体モデルに空間内の移動が必要であれば，座標値を用いる．これらにより，3 次元の人体モデルを使用することで，2 次元画像より奥行きデータの取得が可能となる．

2．3　ソフトウェア操作手順

本ソフトウェアについて説明する．図 3 は人体動作取得ソフトウェアの操作画面である．

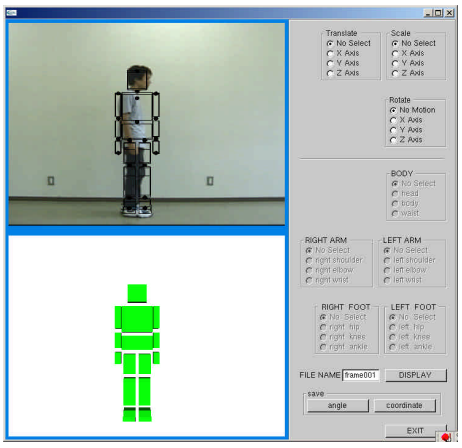


図 3　ソフトウェアの操作画面

このソフトウェアでは，姿勢抽出の対象となる画像の表示・人体モデルの操作・関節間の長さ変更・データの保存が行える．

実行画面(図 3)の右側は選択画面である．この画面では人体モデル操作の<移動><関節の回転><関節間の長さの変更>の選択，表示する画像名の入力，動作データの保存，終了，の操作が行える．左下側の操作画面はマウス左ボタンのドラッグにより，選択に応じた人体モデルの操作が行える．左上側は画像を表示し，その上にワイヤフレームによる人体モデルを表示している．ワイヤフレームの人体モデルは下部の人体モデルと同じ動作を行う．以下に，操作手順について説明を行う．

① 画像の表示
操作画面右側の選択パネルに画像名入力欄がある．この欄にソフトウェア上で表示する画像ファイルの名前を入力する．今回はビットマップ画像に対応した．欄に入力後に <Display>ボタンを押すことで，実行画面の左上部に画像が表示される．

② 人体モデルの調節
人体モデルの関節間の長さを調節する場合は，動作取得対象人物が正面を向いた画像を用意する．次に，その画像に人体モデルを重ね合わせる．関節間の長さが合わない部位があれば，選択画面で部位と 3 方向のどの方向に変更するのか選択をする．そして，実行画面左下部の画面で操作を行う．

③ 人物の姿勢を取得
動作取得対象の画像を表示する．次に，選択画面で人体モデルの移動方向や各部位を 3 軸のどれを軸として回転するのか選択する．そして，操作画面の左下部でマウス左ボタンのドラッグで人体モデルを操作する．左上部の画面を見て画像中の人物とワイヤフレームで表示した人体モデルの重なり具合を確かめる．

④ データの記録
画像に写る人物と人体モデルが同じ姿勢となれば，選択パネルの<SAVE>ボタンを押す．それにより，各関節の各軸周りの回転角度データ，各関節部の座標値データが各々テキストファイルに保存される．

⑤ ③，④の操作の繰り返しにより，幾度も静止画像から人体の姿勢データ取得が可能である．ソフトウェアの終了は選択画面の<EXIT>ボタンを押す．

上記 ①～④ の操作を行うことで，単眼視画像より人体の姿勢データの取得が可能となる．動作取得の実行例の画面を以下の図 4 に示す．画像左上部では，静止画像に写る人物とワイヤフレームの人体モデルが同姿勢である．

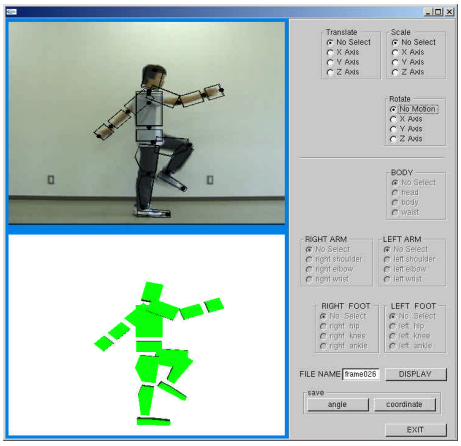


図 4　動作取得の実行例

3.　キーフレームとする画像の決定

キーフレームと決定する動作として，動作を区切ることができる部分を用いる．人体は関節による動作である．関節動作は回転，もしくは曲げて伸ばすなどの動きである．関節の動きが変化するとき，関節の回転速度は停止もしくは速度が低下する．例えば，足の屈伸運動の曲がった状態のである．この変化部分をキーフレームとすれば，動作の補間をすることができる．そこで，このキーフレームとなる画像を全ての静止画像から検出を行う．

4.　差分処理によるキーフレームの検出

キーフレームとなる画像の決定を行う方法として差分処理を用いた．時系列に隣り合う 2 枚の画像 f_t, f_{t+1} に対して差分処理を行う．その差分画像 g_t より画像間の変化を求める．

$$g_t(x,y)=\sum_{y=1}\sum_{x=1}(f_{t+1}(x,y)-f_t(x,y)) \quad (1)$$

画像の変化は人物の動作によって生じたものであるため，人体動作の変化として考える．差分画像の全画素における輝度値を総和し，画像間の変化量 h_t として算出を行う．

$$h_t=\sum_{y=1}\sum_{x=1}g_t(x,y) \quad (2)$$

動画像より取り出した静止画像に対して算出した値 h_t よりキーフレームとする画像を検出する．処理手順は，まず画像をグレースケールに変換する．次に，時系列に隣り合う画像 $t, t+1$ に対して差分処理を行う．そして，差分画像の輝度値を変化量として算出する．

今回使用した動画像は，その場での足踏みを繰り返し行った動作と腕と足の心身動作を用いた．撮影環境は室内であり，秒間 30 フレームの約 4 秒間の撮影を行った．画像サイズは 640×480 である．

4．1　足踏み動作

足踏み動作に対して差分処理を行い，各画像に対して輝度の総和値を算出した結果をグラフ図 5 に示す．横軸はフレーム番号であり，縦軸は差分画像の総輝度値である．

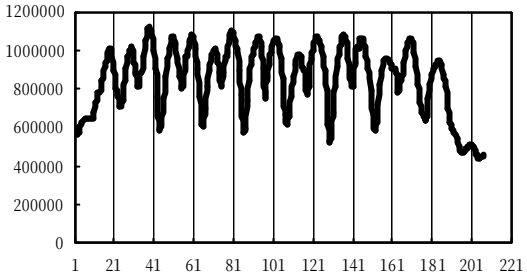
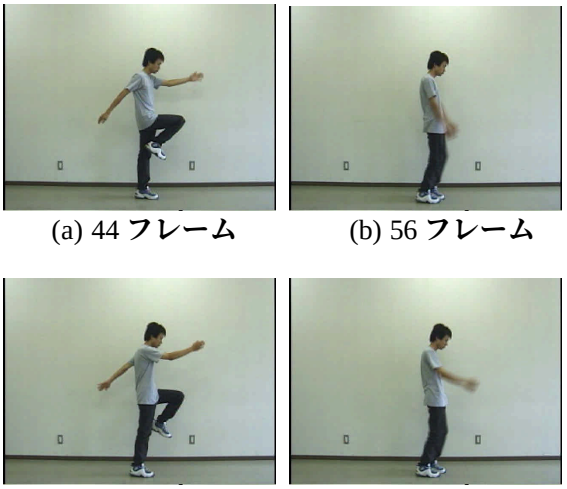


図 5　差分処理

差分処理を用いて全静止画像より算出した変化量は周期性のある波形として現れた．これは，足踏み動作の繰り返しにより現れたものと考えられる．そこで，波形の凸部分となる値の大きいもしくは小さい場合では，足踏み動作の中でどのような姿勢であるかを調べた．その結果，波形が下に凸となる部分に対応する画像に共通した動作の特徴が表れた．図 6 に波形が下に凸となる値に対応した 44, 56, 65, 76 フレームの画像を示す．



(a) 44 フレーム　(b) 56 フレーム

(c) 65 フレーム　(d) 76 フレーム

図 6　足踏み動作

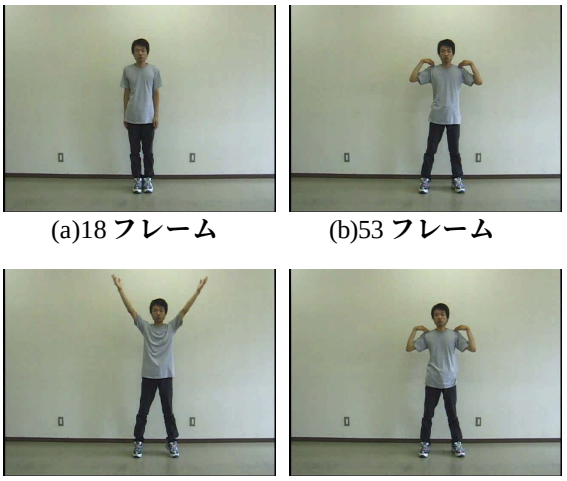
図 6(a)は前方･後方に振り上げた左腕･右腕の下げ，右足を下げる動作である．図 6(b)は右腕を前方，左腕を後方に振り上げ始めている．図 6(c)は前方･後方に振り上げた右腕･左腕を下げ，左足を下げている．図 6(d)は右腕を後方，左腕を前方に振り上げ始めている．

足踏み動作に対して差分処理を行った結果，差分画像の総輝度値が小さく波形が下に凸となるときが動作の切り替わる部分であった．図 6(a)と(b)，(c)と(d)の姿勢の場合の輝度値を比べると，図 6 (b)(d)の姿勢が大きかった．図 6 (a)(c)は腕や足の振りが切り替わる部分であり，動作が停止もしくは遅い．それに対して，図 6 (b)(d)の姿勢は動作が停止していたが，足左右の腕の振り方向の変化していた．そのため，図 6 (b)(d)は腕の振りの分だけ図 6 (a)(c)より値の大きい輝度値が算出された．

このことより，差分処理によって輝度値が小さな値として算出された図 6 の画像は(a)(c)では振り上げられた腕と足が振り下ろされる直前であり，また(b)(d)では前後に振られた両腕の動きが切り替わり振り上げる足が左右入れ替わる直前の地面についた状態である．従って，図 6 の 4 つの姿勢は足踏み動作についてキーフレームとする部分であった．

４．２　腕と足の身伸動作

腕と足の身伸動作に対しても差分処理を行い，足踏み動作と同様に小さい輝度値がキーフレームとする動作の切り替わる部分として検出されるかを調べた．今回の動作は手首の動作も含まれる．腕と足の身伸動作について，図 7 を用いて説明を行う．



(a)18 フレーム　(b)53 フレーム

(c)73 フレーム　(d)95 フレーム

図 7　腕と足の身伸動作

始めに図 7(a)のように直立状態となる．そして，左右の腕を体の側面に沿って広げ，腕が水平になる時に図 7(b)の肘を曲げて両手を肩の位置におく体勢をとる．それから，両腕を上に伸ばした図 7(c)の体勢となる．次に，上に伸ばした腕を下ろして，両手を肩の位置におく図 7(d)の体勢となる．それから両腕を下ろして直立となる．これらの動作を繰り返し行った．

今回の動作の切り替わる部分は図 7 の 4 枚の画像に示す動作である．図 8 に差分処理による画像間の輝度値の変化量を示す．

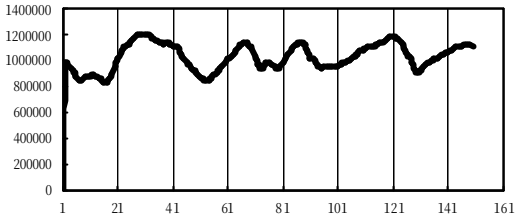


図 8　差分処理

今回の動作は，足踏み動作と同様に差分処理により算出した輝度値が小さな値のフレーム部分が動作の切り替わる部分として検出された．図 7(a)(b)にあたる 18 と 53 フレームにかけて約 40 フレーム以上の間隔がある．これは，直立状態から手が肩に来るまでの間に，腕が水平になるまで広げてから肘を曲げる動作を行っている．そのために時間の経過が長い．また，この間の輝度値の最も大きい値は他の部分よりも大きな値である．これは，このときに，腕の動作だけでなく，足の横移動が含まれているためだと考えられる．図 7(c)の 73 フレーム付近では最小となる値がふたつあることが確認できる．これは，腕を上に伸ばして広げたときの反動によるものである．よって，2 つの値のうち小さな値を特徴あるフレームとする．

５．　オプティカルフローによる検出

差分処理では時系列に連続する画像間の変化から人体動作の特徴の検出を行った．しかし，人体動作そのものを扱ったわけではなく，検出された人体の動きのどの部分がどの方向にどれだけ移動したのかを差分画像から確認することは困難である．差分処理結果から足踏み動作の特徴ある部分では動作の変化が少ない．これは，移動量が少ないといえる．そこで，足踏み動作における人体動作の移動量

の変化を調べるため，オプティカルフローを求めた．オプティカルフローとは時系列に連なる画像 f_t, f_{t+1} において，画像 f_t の各画素が画像 f_{t+1} ではどこに移動したか調べ，各画素の速度ベクトルを求めたものである[9][10]．これより，物体の移動方向や移動距離を検出できる．今回，オプティカルフロー検出にブロックマッチング法を用いた．

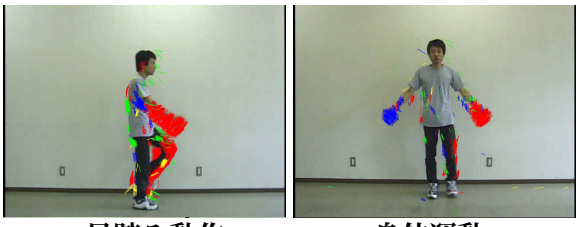
t フレームのある点 $f_t(x,y)$ を注目点とする．その注目点を中心とした $2P+1$ が一辺の長さとなる正方形をテンプレートとする．そして，$t+1$ フレームの点 $f_{t+1}(x,y)$ を中心とした $2L+1$ が一辺の長さである正方形を探索範囲とする．

$$E_t(x,y) = \sum_{m=-P}^{P} \sum_{n=-p}^{P} | f_t(x+n, y+m) - f_{t+1}(x+n+i, y+m+i) | \quad (3)$$

$E_t(x,y)$が最小となる値を探索範囲($-L \le i \le L$)から探し，$f_t(x,y)$ に対応する $f_{t+1}(x_E, y_E)$ を検出する．これを全画素に対して行い，その 2 点間の距離の全画素における総和の値をオプティカルフローによる人物の移動量とした．

5．1　最適な範囲の決定

探索範囲でテンプレート移動を行う上で，探索範囲とテンプレートの大きさは重要である．テンプレートが大きければ検出されるオプティカルフローの信頼性は高いが，大きすぎると検出頻度が低くなる．そこで，最適なテンプレートと探索範囲のサイズを決定する．その結果，テンプレートのサイズ 9(pixel)，探索範囲 20(pixel) が最もよい結果となった．図 9 に足踏み動作と腕と足の身伸運動に対して最適な結果となった画像を示す．



足踏み動作　　身伸運動

図 9　ブロックマッチング

図 9 はオプティカルフローを画像に重ねて表示したものである．線分の方向が移動方向であり，線分の長さが移動距離である．

5．2　足踏み動作のオプティカルフロー

足踏み動作の全静止画像に対してオプティカルフローを求め，算出した結果を図 10 に示す．縦軸は移動量の総合計値，横軸はフレーム番号である．

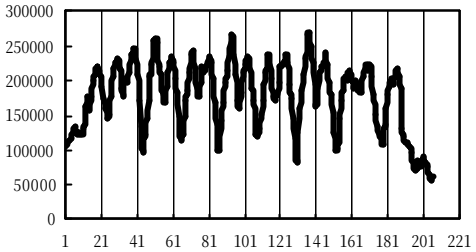
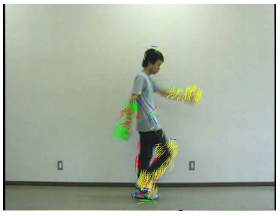


図 10　オプティカルフロー検出結果

差分処理による図 5 のグラフと同様に図 10 のグラフも波形として検出された．波形の値が小さい画像は，図 5 と同じ画像であった．また，波形は 4 つの波で 1 周期となる形に分けられる．この 1 周期の間には，足踏み動作の右手を前に振りそして後ろに振るという一連の動作である．検出したオプティカルフローを重ねて表示した画像を図 11 に示す．

(a)44 フレーム　　(b)56 フレーム



(c)51 フレーム

図 11　オプティカルフロー検出結果

図 11(a)(b)は図 10 のグラフにおいて移動量の値が小さく検出された画像である．図 11(c)の画像は図 11(a)と(b)の間で移動量が最も大きく検出された画像である．画像図 11 (c)では両腕と右足を振り下げる動作であり，オプテ

ィカルフローが多く検出されている．それに対し画像図 11 (a)はオプティカルフローの検出が少なく，また画像図 11 (b)は右腕にオプティカルフローが検出されているが足の部分にはほとんど検出されていないことがわかる．これら図 11(a)(b)の画像は，差分処理により求めた輝度値が小さく特徴のある画像として検出した図 6(a)(b)と同じフレーム番号であった．差分処理図 6(a)と同様に腕と足の振りが停止していることは，図 11(a)でオプティカルフローがあまり検出されていないことからも確認できる．そして，図 11(b)も図 6(b)と同様に腕の振りはあるが足の振りがほとんどないことが確認できる．また，図 11(c)の画像についても図 5 のグラフで輝度値が高く検出されたフレームと同じ画像である．他のオプティカルフローの結果に対しても特徴ある画像は差分処理の結果と同じフレームであった．

従って，オプティカルフローにより人物の移動量が少ないと検出されたフレーム番号と，差分処理により画像の変化量が少ないと検出された画像が一致することを確認した．これらより，簡単な動作でも差分処理はオプティカルフローと同様の結果が得られた．

５．３　身伸動作のオプティカルフロー

腕と足の身伸動作についても移動量を知るためにオプティカルフローの検出を行った．オプティカルフローより求めた人物の移動量を図 12 に示す．

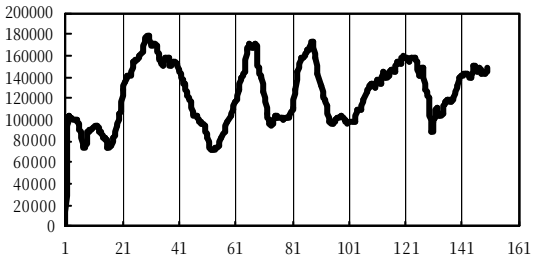
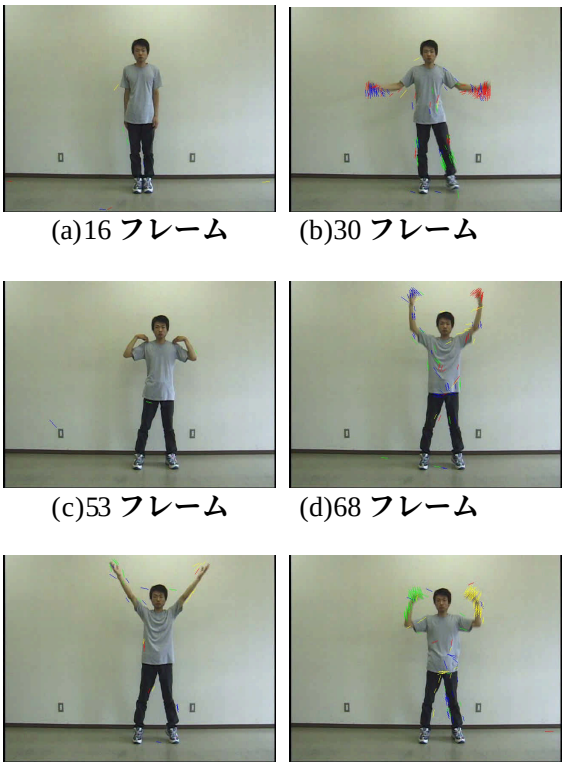


図 12　オプティカルフロー

今回の動作は足踏み動作と同様に腕と足の身伸動作である図 8，12 のグラフは類似した波形となった．図 13 にオプティカルフローによる移動量の値が小さい(a)(c)(e)のフレームと値が大きい(b)(d)(f)のフレームを示す．



(a)16 フレーム　(b)30 フレーム

(c)53 フレーム　(d)68 フレーム

(e)73 フレーム　(f)88 フレーム

図 13　オプティカルフロー

図 13(a)(c)(e)はオプティカルフローの検出もほとんどなく，動作の切り替わる部分であることが確認できる．このフレームは差分処理で小さな輝度値として算出した番号と同じである．

この動作は動作の切り替わり部分である 18，53 フレームの間隔が長い．そのため，動作の補間はこの画像における人物姿勢データだけでは困難である．そこで，移動量の値が大きい場合に注目した．図 13(b)(d)(f)のフレームが移動量の大きい動作である．この部分では腕と足部分にオプティカルフローか検出されている．18，53 フレームの間に当てはまる図 13 (b)フレームの動作は直立体勢から手を肩に置く動作の中間となる，腕を横に広げている動作である．また，足を開く動作に対して，オプティカルフローが検出されている．よって，図 13(b)に対応する 30 フレームは 18 と 53 の間に位置するキーフレーム画像として利用できる．図 13(d)(f)も同様にキーフレームとする画像に用いることができる．今回の腕と足の身伸動作のように，キーフレームとする画像の検出が少ない場合がある．しかし，算出された値が小さいときと大きいときの画像をキーフレームとすることができた．

６．考察

単眼視画像から人体の姿勢抽出は人体モデルによるモデルマッチングを用いた．そのため画像処理における，隠れ問題や人物のシルエットによる関節位置の検出が困難となる場合などに有効である．標準体型である人体モデルに合わない被計測者に対しても，人体モデルの各部位の長さ調節により体型を合わせることができた．しかし，動画から取り出した全静止画に対して姿勢を取得することは，膨大な作業量を要した．

動作に特徴のあるキーフレーム部分の検出は，差分処理またはオプティカルフローより動作の切り替わる部分として検出できた．切り替わり部分は動作の停止や速度の低下から，画像における変化が少ない．そのため，差分処理による輝度値とオプティカルフローによる移動量が小さい値として算出された．また動作の種類により，動作の補間としては特徴ある動作の検出が少ない．そこで，輝度値と移動量が大きな値として算出された画像に注目することで，足りないデータを補った．

差分処理では画像間の輝度の変化量，オプティカルフローは注目画素の移動量が算出される．異なる 2 つの方法で同様に動作の特徴部分を検出した．したがって，キーフレームの検出はオプティカルフローの処理時間より速い差分処理で充分な結果が得られた．

７．おわりに

本研究では単眼視画像より人体の姿勢の取得が可能であった．また，差分処理もしくはオプティカルフローよりキーフレームとなる画像の決定が可能となった．よって，静止画像より決定したキーフレームに対して姿勢抽出を行うことで，人体動作の再現が行える．

動作取得には多くの操作が必要となるため，操作性の簡易化を行う．今回，動作の撮影は固定カメラで行った．そのため，差分処理でも充分な結果が得られた．しかし，人によるカメラ撮影では手ぶれや上下左右の移動が起こり，差分処理に影響がでる．今後は手ぶれなどに対応していく．また，今回は簡単な人体動作を扱ったが，今後はより複雑な動作に対してキーフレームとなる画像の検出を行う．

参考文献

[1]小荒健吾，西川敦，宮崎文夫，“運動と形状に基づく多関節物体輪郭の階層的要素分割法を身体運動計測への応用”，信学論(D-II)Vol.J83-D-II No.12 pp.2630-2640

[2]高橋和彦，坂口辰己，大谷淳，“実時間非接触装着型 3 次元人物姿勢推定法に関する一考察”，信学論(D-II)Vol.J83-D-II No.5 pp.1305-1314

[3]澤井清明，村上伸一，“単眼視による 3 次元人物動作の認識に関する一検討”，信学技報，2003 年

[4]マイクル・オローク，”3D コンピュータ・アニメーションの原理”，近代科学社，2000 年

[5]鳥居俊，“身体運動の運動学”，ナップ，1999 年

[6]“日本人の人体計測データ”，社団法人人間生活工学研究センター，1997 年

[7]小島一成，“外部入力リズムを利用した人体動作と伝統芸能画像の再現”，博士論文，2002 年

[8]鈴木隆雄，“日本人のからだ　健康・身体データ集”，朝倉書店，1996 年

[9]末松良一，山田宏尚，“画像処理工学”，コロナ社，2000 年

[10]安居院猛，長尾智晴，“C 言語による画像処理入門”，昭晃堂，2002 年