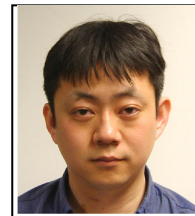


民俗芸能のデジタルアーカイブと ロボットによる動作提示

池内克史、中澤篤志、小川原光一、高松淳

工藤俊亮、中岡慎一郎、白鳥貴亮

東京大学大学院情報学環 / 大阪大学サイバーメディアセンター / 東京大学生産技術研究所
東京大学大学院情報理工学系研究科



1. はじめに

近年、デジタル技術を用いて文化遺産をデジタル保存する試みが多くなされている。文化遺産は、大仏や正倉院御物といった「静」的な有形文化財と、民俗舞踊や伝統芸能、匠の技といった「動」的な無形文化財に分類できる。「静」的な有形文化財に関しては、レーザーセンサーやデジタルカメラを入力として、形の情報を得るための3次元形状計測や、表面の色・艶を表現するための表面特性解析アルゴリズム、光源環境を再構成する光源環境推定アルゴリズムなどが主な研究項目である。我々のグループではこれらのアルゴリズムの開発を文部省の学術創成プロジェクトや科学技術振興機構の戦略的基礎技術開発事業プロジェクトを通して行い、鎌倉や奈良の大仏のデジタル化に成功した。

これらの「静」的な有形文化財に対して、能や刀鍛冶の技といった身体の動きをその中心表現とする「動」的

な無形文化財も多く存在している。無形文化財の保存は主に人から人への直接的な伝承が主たる保存方法であった。このため後継者難などの社会的背景の変化から、日々失われつつあるといっても過言ではない。すなわち、これらの無形文化財に関しても何らかのデジタル的な保存技術が開発されなければならない。

無形文化財においては、その中心表現が演者の動きの中に込められている。従って、動きをどのように記録するか、動きをどのように再現するかが研究の中心課題となる。無形文化財のデジタル化は、単純に考えると演者の動きをビデオで採録し、必要に応じて再生すればよい。実際、そのような民俗芸能のアーカイブも存在する。あるいはもう少し進めて、多視点からビデオを取り込み、任意の視点からのビデオを再生するという手法も提案されている。しかし、たとえば踊りを学習する場合、学習ビデオを眺めるのと、実際に目の前で師匠が要所・要所

を解説しながら踊って見せてくれるのでは、進歩に大きな差が出る。師匠なしでは本当の意味での習得は不可能であるといっても過言ではない。さらに、再現された場合も単なる映像の再現と実際の人の動きによる再現では臨場感がまったく異なる。

我々は無形文化財のデジタル化のためには、この踊りの師匠を物理的に作成することが必須だと考える。要件としては、1) 無形文化財の動きの流れとともにその要所(あるいは肝)を抽出し、理解し、解説できること、2) 動きを物理的に、とくに肝となる部分を忠実に表現できること、の2点であると考えられる。

本稿では我々のグループで、過去4年間にわたり、戦略的基礎技術開発事業(CREST)プロジェクトとして開発を行ってきた無形文化財のデジタル保存のための一連の手法について概説する。研究テーマとしては以下のトピックに分類できる。

1. 人の動き情報の取得
2. 動き情報の解析
3. 動作情報の生成
4. ロボットによる動作情報の提示

人の動き情報の取得に関しては、主にコンピュータビジョンの分野で1台のテレビカメラからの時系列画像解析として研究されてきた。近年は、多数のカメラの連動により身体の動作情報獲得も随所で行われている[1][2][3][4][5][6]。一方静的な文化財の3次元データ取得として両眼立体視に対するレンジファインダーがあるように、動的なデータ取得にも多視点カメラに対する手法として光学式や磁気式のモーションキャプチャシステムが開発されている。モーションキャプチャシステムでは、対象者の体にセンサやマーカを取り付ける必要はあるものの、正確なデータが得られる。我々のグループでは、多視点カメラシステムとモーションキャプチャシステムを併用し、人の動き情報を得ている。

動きの解析に関しては、人の様々な動作から動きの共通項的役割を果たすシンボリックな表現を抽出する。手先動作の速度変化や音楽情報から動作を切り分け分類する。切り分けられた舞踊動作セグメントのクラス化を行い階層構造をつくることで全体構造を把握する。さらに、セグメント間や個人間によるセグメント内での軌道などのばらつきから踊り手のくせや中心表現を抽出することなどが課題である。

無形文化財という人から人へと伝承することが重要となる対象を扱うためには、獲得・解析された動作を単なるデジタルデータとして残すだけではなく、これを提示

するための技術開発も重要である。動作データの提示に関しては従来よりCGを用いた手法が主流であるが、我々のプロジェクトでは産業技術総合研究所ヒューマノイドロボットグループとの協力の下、ロボットを用いた手法を開発している。CGによる手法に比べ、臨場感が大きく、動きデータを表示するディスプレイとしては画期的なデバイスであると考えている。しかし、モーションキャプチャによって得られた人の動きデータをそのままロボットに入力しても、人とロボットの間での様々な違いによって同様の動きをつくることは不可能であるため、これを実現するための新たな手法を開発する必要がある。このために我々は、ロボットと人体モデルの動力学的制約を満たす動作生成手法や、動き情報の構造解析結果を用いた適応的フィルタを導入、動作の中心表現を変化させずに全体的な動きの流れを変換する手法などを開発している。以下では、中心課題である動きの解析、動作生成、ロボットによる再現について詳細にのべる。

2. 舞踊の動作解析

モーションキャプチャシステムから得られた舞踊の関節位置データを用いて、舞踊動作全体の解析を行う。ここでの目的は舞踊全体の動作から有限個の「基本動作プリミティブ」を獲得し、舞踊を表現することである。これにより、舞踊内での繰り返しの動作や一度しか出現しない動作などを把握することができ、舞踊全体の構成を把握することができる。また、異なる舞踊のプリミティブを比較することによって相互の関係を把握することも可能となる。さらに、今まで文字や絵等で記されてきた舞踊の記述方式としてこの表現を用いることも可能である。

舞踊を有限個の基本動作によって記述するという取り組みはすでにいくつか成されている。秋田県のわらび座では手動で動きデータを分割し、基本動作として記録している[7]。また、ロボットの行動獲得を目的とした研究としては、複数の関節角度をSelf Motion Elementと呼ばれる単位に分割し、HMMを用いて基本動作を抽出したものの[8]や複数物体の3次元の接触状態をロボットの視覚で観測しその制約条件から行動計画を学習するもの[9]等がある。人の「見かけの動き」に基づいたアプローチとしては、手先の動きを平面に投影し直線・円弧等の基本パターンに分類し再現したものの[10]や、手先の動きを速度に基づいて切り出しその手先の3次元空間中の動きをDynamic Time Warpingによる距離を用いてラベル付けする手法[11]などが挙げられる。舞踊を対象とし

た研究としては、八村らによるモーションキャプチャデータからの Labanotation の抽出および動作の再生の研究が挙げられる[12]。ここではモーションキャプチャデータを Labanotation と呼ばれる記号に変換しシンボリックに動きを記述する。Labanotation は西洋舞踊等において一般的に用いられ汎用性が高いことから、無形文化財のアーカイブ化においても有用な記述方法である。しかしこの表現方法は、舞踏分野に精通した人が理解の記録のために使用するため、細部の動きはすでに理解できているという前提に立っており、人の動きを粗く表現する傾向がある。このため、プロフェッショナルのための舞踊の記述には用いることができても、我々のようにロボット等を用いて動作を再生するためには適していない。さらに、動きの認識に対して固定しきい値を用いているため、同じ動きに対しても異なる舞踏譜として記録される可能性がある。我々のグループでは、入力されたモーションキャプチャデータを分割し、低レベルの表現としてその分割された動きセグメントをそのまま保存するアプローチをとる。分割された動きセグメント間ではその目標点(手先、足先)の軌跡によって動きの相関が評価され、舞踊の中で頻出する動きセグメントが「基本動作」として抽出される。最終的には、同一の動作とされた分割セグメントの中の動き情報を平均化しアーカイブ化する。この方法では、実際のモーションキャプチャデータから得られたデータをそのまま保存データとして用いるため動作の再現が容易になる。

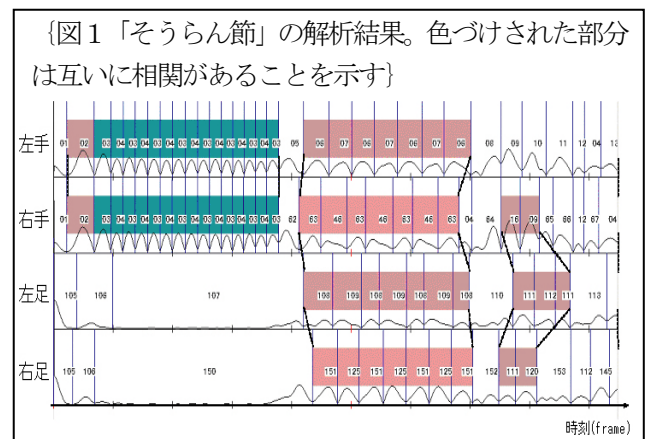
民俗舞踊を見ると、その動きの中には繰り返しやバリエーションを持たせた動作など、音楽のような動作構造が存在することに気づく。そこで、以下に述べる5つのステップで、低レベル表現である分割された動作データをクラス化することで動作の構造を把握する。

- 1) 各フレーム毎に人の腰位置に体中心座標系を設定。
- 2) この座標系における手先足先の速度と軌跡を得る。
- 3) 速度の極小値を動きのセグメントの境界フレームと考え動作の切り分ける。これには EM-Algorithm を使用する。
- 4) 切り分けられた動作セグメント内の軌跡の相関を評価することで動作をクラス化する。
- 5) 同クラスの動作セグメントが連続する場合、セグメントの併合を行うことで舞踊全体の動作構造を把握する。

以上の5ステップにより手足の動きの加速度で切り分け体中心に対して同じ軌跡を描くセグメント(最小セグメント)が分類できる。この最小セグメントでは手や足の振り等の動きが検出できるが、舞踏のようなシナリオ

をもった動きではより大きな単位である共通した動きが存在する。これを検出するためにレベル付けされた最小セグメント列から頻出する一連のパターンを検出して、より大きな動き列を抽出する。これには Apriori Algorithm を用いる。これらの大きな動き列を分類することで同一部位における動作ラベル列の階層化を行う。次に、他の部位の動きとの関連付けも行い最終的な動きの記述とする[13]。

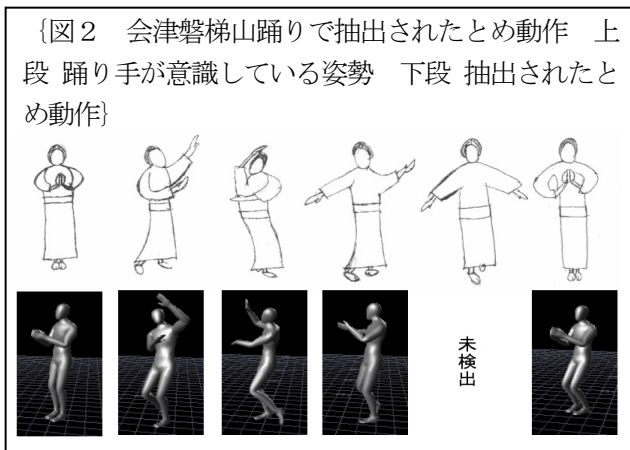
本アルゴリズムを用いた「そうらん節」の解析結果を図に示す。舞踊がいくつかの繰り返しのある動作プリミティブとそれらを連結する動作セグメント、一度しか出現しない動作プリミティブ等の部分から成り立っていることが確認できる。繰り返し演じられる動きが同様のラベルに振られるため、必要最小限の動作プリミティブを登録するのみで十分となるだけでなく、舞踊全体の構造を把握することも可能である。これらの動作セグメントを利用して Labanotation の抽出を安定化することも可能である。



この手法では、人の手足の動き情報のみをセグメント化とクラス化に用いている。だが、これが舞踊において意味レベルでも正しいものであるかは分からない。これを解決するため音楽情報を導入して意味レベルでも正しいセグメント化を行う手法を開発した。ここでは、舞踏の動きは音楽と不可分のものであり、音楽の節々に重要な動きが出るとの仮定による。

まず、舞踊で流される音楽をスペクトル解析することでリズム(ビート)を抽出する。このビートと先に求めた動作セグメントがほぼ同時刻にくるものを見つけ、加えて両手両足、重心のそれぞれの動作セグメント点が同時刻に存在する場合、意味レベルでも正しいセグメント点として動作の切り分けを行う。図2に、この手法を用いて会津磐梯山踊りに対して適用した結果を示す。止め動作が正しく抽出されている様子が見て取れる。実際、

踊り手が踊っているときに意識している姿勢についてヒアリングを行い、求められた動作セグメントとの比較をおこなったところほぼ一致した結果が得られ、本手法の有効性が示された。



ここであげた 2 種の動作解析手法から、民俗舞踊においては以下のような特性が成り立つものと考えられる。

- ・ 舞踊はセグメント化された動作から成り立ち、それらは有限種のクラスに分類できる。
- ・ セグメントフレームは、音楽のリズムに応じたリズムパターンに応じて現れる。

3. 動作プリミティブの編集による新規動作列の生成

前章の手法で舞踊中の動作プリミティブを導出できる。ここで例として挙げた 3 つのおどりの解析結果を見ると、舞踊はいくつかの繰り返し動作と、それを連結する動作から成り立っていることがわかる。すなわち、舞踊動作は有限の動作群の組み合わせによって構成されるという仮説を支持するものであると言える。この知見から、抽出された動作プリミティブをデータベースとして蓄積することの重要性がわかる。さらに、任意の動作プリミティブを自動的に連結する手法があれば、既存のプリミティブをつなぎ合わせて新しい舞踏生成するという舞踊動作の編集が可能になる。

この舞踏編集のための繋ぎ動作は、人の動作軌道生成に基づく軌道を生成する必要がある。軌道生成のうち、腕の動きに関しては神経科学やロボティクスの分野で多くの研究がなされている[14][15]。Flash らは、腕の動作軌道に対して Jerk(躍度)最小モデルという概念を提唱している[16]。人の腕の動作においてはその最中の躍度の積分がもっとも小さくなる軌跡を通るという理論であり、実験的にもその正しさが証明されている。この概念を発展させ、宇野らはトルク変化最小モデルを提唱し、川戸らは粘性トルク変化最小モデルを提唱している。い

ずれも、より人間の随意運動に近い動きを生成することが示されている。

一方、脚の動きに関しては現在さまざまな手法が考えられているが、脚は手と異なり体を支えるためのバランスをとる必要があるため、腕のように動き自体を他の部位の動きと独立して考えることが難しい。我々は、動きの間ではいずれかの脚が支持脚となるという拘束のみを与え、他の脚や体全体は滑らかに連続した動作をするという仮定を設けることと脚の軌道生成を行う手法を開発した[13]。図に示すように、連結対象とする動作プリミティブ間での支持脚を設定し、他方のプリミティブはこの脚が一致するように平行移動させる。次に、遊脚先位置、腰位置、体中心位置、首位置を 2 つのプリミティブの間で一一致するよう他方のプリミティブを移動させる。さらに、腰ならびに首の座標系を回転させる。最後に、腕および脚の関節角度跳躍度最小な軌道で補間を行う。動作変形のアニメーションから、異なる舞踊などの間でもスムーズに連結動作が確認でき、本動作編集手法が有効に働いていると思われる。

4. 最適化計算による動作生成

前節の軌道計算では、跳躍度最小となる軌道を利用した。人型モデルの動作生成においては、モデルに予期せぬ外乱が加えられた場合でも、それを適切に対処できるようにすることが重要である。とくに、2 本脚モデルでは体を支えるための「支持面」が小さいため、バランスをとることが難しくなる。これまで CG やロボティクスの分野で人型モデルのバランス制御についての研究がなされてきた。

CG の分野においては、力学的に正しいリアルなキャラクターアニメーションへの需要の高まっており、グローバルなフィルタリングにより力学的に正しい動作を生成する手法[17][18]や取得された動きと PD 制御を組み合わせた手法[19]、あるいは力学的な保存量に注目した手法[20][21]などが代表的なものである。一方、ロボティクスの分野においては近年ヒューマノイド・ロボットに関する研究が盛んであり、これらの研究においてはバランス制御は本質的に重要な問題であるため、[22][23][24]などの手法が提案されてきた。

一方、人間は非常に高度なバランス保持動作を行っており、そのメカニズムを見習うことで、より効果的で頑強なシステムを構築できる可能性がある。このため、人間のバランス保持動作について観察することは重要であると考えられる。人間の場合大きな外乱が加えられると、バ

バランスを保つためにしばしば腕を大きく振り回したり、腰を大きく折り曲げたりする。すなわち、外乱で生じた角運動量をこれらの動作によって適切にコントロールし、転倒を回避していると考えられる。大きな外乱に対処するには、このようなダイナミックな全身動作によるバランス保持が必要である。ところがこれらのダイナミックな動作は、従来のフィルタリングやPD制御によるバランス制御の手法では実現が困難であった。これは、従来の手法では現在の姿勢と目標とする姿勢の差を縮めるように動作を生成するため、「一時的に目標とする姿勢からは遠ざかるが、実はバランスを回復するのに効果的な動作」を生成することができなかったからである。

我々のグループでは、この点に注目し、最適化を用いた姿勢制御の手法を開発している。この手法においては、ZMP (Zero Moment Point) と呼ばれる「地面から受ける力によるモーメントが0になる点」を足の支持多角形内に納まるよう制御することで、動的バランスを直接的に取り扱う[25]。ここでは各関節の角加速度を変数とし、「できるだけ力を使わずに姿勢を保持する」ということに相当する「関節角加速度の二乗和」の最小化する定式化を行う。最適化には、二次計画法を用いる。その際、拘束条件として次の3つの条件を設定した。

- ・ 重心の加速度が、重心の目標位置の方向を向いている。
- ・ モデルのZMPが足の支持多角形内に存在する。
- ・ できるだけ左右対象な動きとなるようにする。

すなわち、重心の位置・加速度やZMPのみを問題にし具体的な姿勢については問題にしないことによって、フィルタリングやPD制御では実現できなかったような、ダイナミックに全身を使ったバランス回復動作が生成できるようになった。具体的には、人間によく見られるような腕を振り回すことによって角運動量を適切に制御してバランスをとる動作が生成された(図3)。

{図3 最適化による動作生成アニメーション生成}



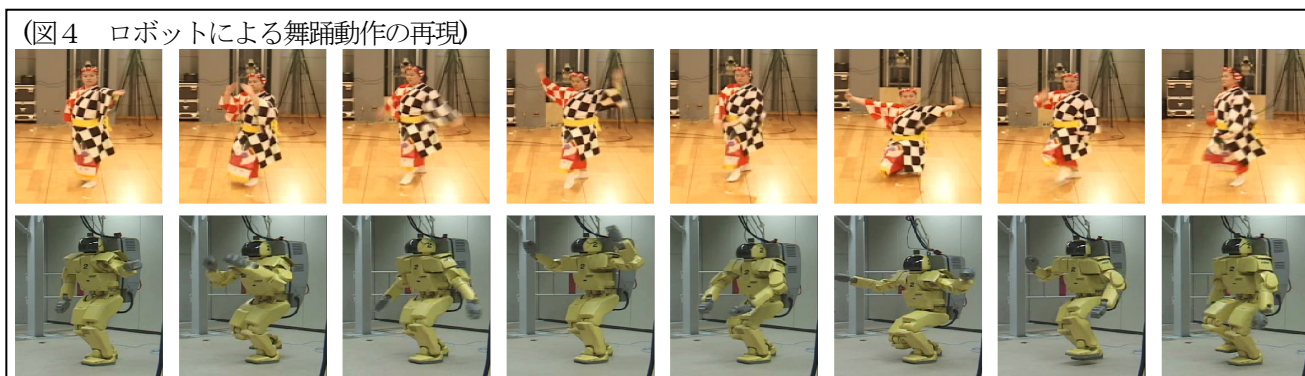
5. ロボットによる動作提示

コンピュータ上に取得された動作データのより効果的な提示手法として、ヒューマノイドロボットによる舞踊の実演を試みている。しかし、ヒューマノイドロボットは人間に似せて設計されてはいるものの、現状ではまだ関節の構造や自由度・可動範囲・動作速度などにおいて制限されたものとなっており、身体における重量配分も異なる。よって人間から得られた動作データをそのままロボットに適用することは不可能であり、ロボットは目標とする動作を自らの身体に適応させていく必要がある。これは人間同士であっても体格や身体能力の違いがある中で同じ動作をしようとしていることと同じであり、動作模倣における一般的な問題であると考えられる。

ロボットを動作させるにあたって、元の動作からの逆運動学によって直接的に得られた関節角軌道をベースとすることが考えられる。Pollardら[26]は得られた個々の関節角軌道を可動範囲や最大角速度内に収まるよう修正することで、ロボットによる舞踊模倣を実現している。しかしこうして得られた動作は身体全体とその力学を考慮していないので、自己干渉を起こしてしまったり、元の動作の特徴を失ってしまったりする上に、自らの脚でバランスをとって動作を続けることは難しい。これに対し、我々は今まで述べてきたような動作の解析結果を有効に活用することでこの問題を解決することをめざしている。元の動作をそのままロボットの制約内に修正していくのではなく、元の動作を解析して基本となる動作の列を抽出し、ロボットの動作はその基本動作列をもとに新たに生成する。ロボットは各種の制約や力学を満たしながら、個々の基本動作のもつ動作の特徴を反映していけばよい。この手法を用いて、図4に示すように、ロボットが自らの脚でバランスをとりながら舞踊の模倣を行うことに成功した[27]。

6. まとめ

本論文では、現在我々のグループで開発中の無形文化財(舞踊)の保存、解析、提示を行うための一連の手法を中心に各種の課題や解決法について解説した。人の動きは複数のカメラ群やモーションキャプチャシステムによって入力される。入力された動き情報は、各身体部位の速度によって分割される。その際、音楽情報を考慮することでより精度の高い分割が行える。各分割セグメントを相互評価することで階層化を行い、舞踊全体の構造が把握できる。この人の動きの要素を「動作プリミティブ」と定義した。動作プリミティブを連結することで新



たな動きを生成する事も可能となる。保存された動きの提示方法としては、ヒューマノイドロボットによる再演による提示手法を提案した。人間とロボットではさまざまな条件が異なるため、ただ単に人の動きデータを入力するだけでは、同様の動きを再現することはできない。このため、動作プリミティブを中心に人の動きをロボットに適応させる手法を開発し、実機でその有効性を確認した。なお、本稿では紙面の都合上、全身運動の保存のみについて解説した。これ以外にも、主に目と手の協調による匠の技の再現等も重要な課題である。これについては、別の機会に譲る。

謝辞

本稿で報告した研究は、科学技術振興機構（旧科学技術振興事業団）戦略的基礎研究推進事業（CREST）「高度メディア社会の生活情報技術（長尾真領域総括）」研究領域「文化遺産の高度メディアコンテンツ化のための自動化手法」開発プロジェクトの一環として行われた。また、ロボットによる実演の部分は、産業技術総合研究所知能システム研究部門ヒューマノイド研究グループ（比留川博久グループリーダー）と共同で、産業技術総合研究所のHRP1Sハードウェアを用いておこなわれた。関係者各位に感謝します。

参考文献

- [1] P. Rander, T. Kanade, P.J. Narayanan, "Virtualized Reality: Constructing Virtual Worlds from Real Scenes", IEEE MultiMedia, vol.4, no.1, pp.34-47, 1997.
- [2] D.M. Gavrila and L.S. Davis, "3-D Model-based Tracking of Humans in Action," Proc. of CVPR., pp.73-80, 1996.
- [3] I. Kitahara, Y. Ohta, "Scalable 3D Representation for 3D Video Display in a Large-scale Space", IEEE Virtual Reality Conf. 2003, pp.45-52, 2003.
- [4] X. Wu and T. Matsuyama, "Real-Time Active 3D Shape Reconstruction for 3D Video," Proc. of 3rd Int. Symp. Image

and Signal Processing and Analysis, pp.186-191, 2003.

- [5] 宮崎 英明, 亀田 能成, 美濃 導彦, "複数のカメラを用いた複数ユーザに対する講義の実時間映像化法", 信学誌 J82-D-II, No.10, pp.159-160, 1999.
- [6] 源田大輔, 向川康博, 尺長健, "視点と法線を組み合わせた任意視点映像の生成法", 情処研報, CVIM-137-7, 2003.3.
- [7] わらび座, <http://www.warabi.or.jp/>.
- [8] T. Inamura, Y. Nakamura, H. Ezaki, I. T. Oshima: "Imitation and Primitive Symbol Acquisition by the Integrated Minesis Loop," Proc. ICRA 2001, pp.4208-4213, 2001.
- [9] J. Takamatsu, H. Tominaga, K. Ogasawara, H. Kimura, K. Ikeuchi "Extracting Manipulation Skills from Observation," Proc. ICRA, Vol.1 pp.584-589, 2000.
- [10] O. C. Jenkins, M. J. Mataric, S. Weber "Primitive based movement classification for humanoid imitation," IEEE Int. Conf. Humanoid Robots, 2000.
- [11] 大崎竜太, 嶋田光臣, 上原邦明: 速度に基づく切り出しとクラスタリングによる基本動作の抽出, 人工知能学会誌 15巻5号, pp.878-885, 2000.
- [12] 松本敏良, 八村広三郎: モーションキャプチャデータからの基本身体動作の抽出, 人文科学とコンピュータシンポジウム論文集, pp.17-24, 2000.
- [13] 中澤篤志, 中岡慎一郎, 原田貴昭, 工藤俊亮, 池内克史: 視覚による舞踊動作の保存・解析および生成, 画像の認識・理解シンポジウム, pp.153-158, 2002.
- [14] 川人光男: 脳の計算理論, 産業図書, 1996.
- [15] 伊藤宏司, 伊藤正美: 生体とロボットにおける運動制御, コロナ社, 1991.
- [16] T. Flash, H. Hogan "The coordination of Arm Movements: An Experimentally Confirmed Mathematical Model," J. Neuroscience, pp.1688-1703, 1985.
- [17] S. Tak, O. Song, and H. Ko, "Motion Balance Filtering", Computer Graphics Forum, 19(3); pp.435-446, 2000.
- [18] K. Yamane and Y. Nakamura, "Dynamics Filter-Concept