

## 局所線形モデルの整列による非線形システムの学習法

## Learning Non-linear Dynamical Systems by Alignment of Local Linear Models

上甲昌郎\*1      河原吉伸\*2      矢入健久\*1  
 Masao Joko      Yoshinobu Kawahara      Takehisa Yairi

\*1 東京大学      \*2 大阪大学  
 The University of Tokyo      Osaka University

In this paper, we present an algorithm for learning non-linear dynamical systems which works by aligning local linear models, based on a probabilistic formulation of subspace identification. This is achieved by the fusion of the recent works in the fields of pattern recognition and system control. Because the procedure for constructing a state sequence in subspace identification can be interpreted as the CCA between *past* and *future* observation sequences, we can derive a latent variable representation for this problem. Therefore, as in a similar manner to the recent works on learning a mixture of probabilistic models, we obtain a framework for constructing a state space by aligning local linear coordinates. This leads to a prominent algorithm for learning non-linear dynamical systems. Finally, we apply our method to motion capture data and show how our algorithm works well.

## 1. はじめに

動的システムの学習は様々な分野で重要な問題となっている。しかし、多くの場合において、モデルのパラメータの正確な値や、そもそも動的システムの全体的なモデル構造さえ未知であり、結果として動的システムは観測時系列データのみから学習せねばならない。そのような動的システムを学習する主要な線形アプローチの一つに部分空間同定法がある。部分空間同定法は、観測データによって作られるハンケル行列の行、列によって張られる部分空間の幾何演算に基づいている [Katayama 05]。1980年代後半の台頭ののち、部分空間同定法はシステム制御の分野で積極的に研究されており、解が一意に求まるため応用的に使い易いという利点を持つ。しかし、例えばヒューマンモーションのように、システムが複雑である場合、線形表現に基づくモデルは不適切である。そうした状況で、近年では [Roweis 01] や [Kawahara 06], [Palanthandalam 05] のように、動的システムの非線形学習法も盛んに議論されている。

本論文では、部分空間同定法を確率的に解釈することで、部分空間同定法をもとにした非線形システム学習法を提案する。部分空間同定における状態空間を構築する処理は未来と過去の観測時系列の間の正準相関分析 (CCA) とみなすことができ、それゆえ正準相関分析の確率的解釈 [Bach and Jordan 06] と同様にして、部分空間同定法の潜在変数表現を導出することができる。潜在変数モデルで表現された部分空間同定法に対して、パターン認識や機械学習の分野における近年の成果である、アライメントベースの混合確率モデル学習アプローチ [Verbeek 08] によって、局所線形モデルのアライメントによる状態空間を構築する枠組みを得、非線形動的システム学習の優れたアルゴリズムを導く。

第二章において、我々は部分空間同定法を確率的解釈することによって正準相関分析と同様の形式で潜在変数モデルを導出し、局所線形モデルのアライメントによる非線形動的システム学習法を提案する。そして第三章においてフィルタリングの手順を示す。第四章では、実例を基に本手法の有用性を示し、第五章にて結論を述べる。

## 表記

離散時間のベクトル過程  $\mathbf{y}(\tau)$ ,  $\tau=0, \pm 1, \dots$  を考え、現在の時刻  $t$  における、過去及び未来の  $k$  ステップにわたる部分列を以下のように表記する。

$$\begin{aligned} \hat{\mathbf{y}}_k(t) &\equiv [\mathbf{y}(t-1)^T, \mathbf{y}(t-2)^T, \dots, \mathbf{y}(t-k)^T]^T \\ \mathbf{y}_k(t) &\equiv [\mathbf{y}(t)^T, \mathbf{y}(t+1)^T, \dots, \mathbf{y}(t+k-1)^T]^T \end{aligned}$$

また、 $\Sigma_{pp}$ ,  $\Sigma_{fp}$ ,  $\Sigma_{ff}$  はこれらの部分列のサンプル共分散行列であり、添え字の  $p$  や  $f$  は、それぞれ過去、未来の部分列を表す。例えば、 $\Sigma_{pp} \equiv E\{(\hat{\mathbf{y}}_k(t) - \boldsymbol{\mu})(\hat{\mathbf{y}}_k(t) - \boldsymbol{\mu})^T\}$  のようになる。

## 2. アライメントアプローチによる非線形システム学習法

離散時間の広義定常過程  $\mathbf{y}(t)$ ,  $t=0, 1, \dots$  を考える。ここで  $\mathbf{y}$  は次元  $p$  のベクトルで、未知の確率システムの出力である。しばしば、確率システムは以下で表される線形の状態空間モデルで表現できると仮定される。

$$\mathbf{x}(t+1) = \mathbf{A}\mathbf{x}(t) + \mathbf{v}(t), \mathbf{y}(t) = \mathbf{C}\mathbf{x}(t) + \mathbf{w}(t) \quad (1)$$

ここで、 $\mathbf{x}$  は状態ベクトルで、 $\mathbf{v}$ ,  $\mathbf{w}$  はシステム及び観測のノイズを表す。状態空間モデル (1) 式で表されるシステムを推定するにあたっての基本となる方法はシステムの状態ベクトルを構築することである。なぜなら、一旦状態ベクトルが構築されればシステムのパラメータは最小二乗法などによって得られるからだ。その代表的な方法の一つに、正準相関分析に基づく部分空間同定法がある。本論文では、ヒューマンモーションのような複雑な非線形システムを考えるために、後に示されるような確率的混合状態空間モデル (5) でシステムをモデル化し、それに伴って部分空間同定法を非線形に拡張する。2.1 節では正準相関分析の確率的解釈と同様にして部分空間同定法の確率的解釈を示し、それを基に、2.2 節で複数の局所座標系を一つの大域的な表現に統合することによって状態空間を構築する方法を考える。そして 2.3 節では、二つの回帰問題を解くことで混合モデルのシステムパラメータを求め、全体的な学習アルゴリズムを導出する。

連絡先: 上甲昌郎, 東京大学工学系研究科航空宇宙工学専攻,  
 mjoko@space.rcast.u-tokyo.ac.jp

## 2.1 部分空間同定法の確率的解釈

本節では正準相関分析に基づくマルコフモデルの平衡確率実現について述べ、正準相関分析に基づく部分空間同定法が平衡確率実現であることを示し、[Bach and Jordan 06]と同様に潜在変数モデルを考えることで部分空間同定法の確率的解釈を行う。

[Katayama 05]によれば、未来の出力  $\mathbf{y}_k(t)$  と過去の出力  $\tilde{\mathbf{y}}_k(t)$  の正準相関ベクトルを基にして得られた  $\mathbf{x}^f(t)$  及び  $\mathbf{x}^b(t)$  を用いると、時刻  $t$  における未来の出力  $\mathbf{y}_k(t)$  及び過去の出力  $\tilde{\mathbf{y}}_k(t)$  は以下のように表される。

$$\begin{aligned} \mathbf{y}_k(t) &= \mathbf{O}_k^f \mathbf{x}^f(t) + \mathbf{v}_k \\ \tilde{\mathbf{y}}_k(t) &= \mathbf{O}_k^b \mathbf{x}^b(t) + \tilde{\mathbf{v}}_k \end{aligned} \quad (2)$$

ここで、 $\mathbf{O}_k^f$ ,  $\mathbf{O}_k^b$  は拡大可観測行列であり、 $\mathbf{v}_k$ ,  $\tilde{\mathbf{v}}_k$  は  $\mathbf{y}_k$ ,  $\tilde{\mathbf{y}}_k$  に相当する観測のノイズであり、 $\mathbf{x}^f(t)$  及び  $\mathbf{x}^b(t)$  は、

$$\begin{aligned} \mathbf{x}^f(t) &\equiv S^{1/2} V^T \Sigma_{pp}^{-1/2} \tilde{\mathbf{y}}_k(t) \\ \mathbf{x}^b(t) &\equiv S^{1/2} U^T \Sigma_{ff}^{-1/2} \mathbf{y}_k(t) \end{aligned} \quad (3)$$

と表され、 $U$ ,  $S$ ,  $V$  は特異値分解  $\Sigma_{ff}^{-1/2} \Sigma_{fp} \Sigma_{pp}^{-T/2} = USV^T$  から得られる。ここで得られた  $\mathbf{x}^f(t)$  及び  $\mathbf{x}^b(t)$  はそれぞれ前向きマルコフモデル、後ろ向きマルコフモデルの状態ベクトルに相当し、この二つの確率実現の状態共分散行列は一致する。これを平衡確率実現と呼ぶ。

つまり、正準分析に基づく部分空間同定法は平衡確率実現に基づいており、これに伴ってシステムの低次元確率実現が可能となる。よって、 $\mathbf{y}_k(t)$  と  $\tilde{\mathbf{y}}_k(t)$  が i.i.d とみなせるとき\*1、(2)式は正準相関分析の潜在変数モデル [Bach and Jordan 06] と同じ形式となるため、平衡確率実現に相当する潜在変数モデルは以下のように表現できる。

$$\mathbf{x} \sim N(0, I_d), \mathbf{y}_k | \mathbf{x} \sim N(F_1 \mathbf{x}, \Psi_1), \tilde{\mathbf{y}}_k | \mathbf{x} \sim N(F_2 \mathbf{x}, \Psi_2)$$

ここで、 $I_d$  は状態ベクトルの次元  $d$  の単位行列、 $F_1$ ,  $F_2$ ,  $\Psi_1$ ,  $\Psi_2$  はこのモデルのパラメータを表す。

## 2.2 局所線形モデルのアライメントによる状態空間構築

前節で得られた潜在変数モデルに対して、[Verbeek 08]と同様に局所線形モデルのアライメントを考えることで非線形への拡張が可能となり、大域的な状態空間を構築することができる。図1、図2は部分空間同定法の潜在変数モデルと、アライメントのアプローチによって非線形に拡張された部分空間同定法のグラフィカルモデルを表す。現在の設定において、[Verbeek 08]と同様にして次式のペナルティ付きの対数尤度目的関数を最適化することで、大域的な座標系を得ることができる。

$$\mathbf{L}' = \sum_{n=1}^N [\log p(\tilde{\mathbf{y}}_k(n), \mathbf{y}_k(n)) - D(q_n(\mathbf{x}) \| p(\mathbf{x} | \tilde{\mathbf{y}}_k(n), \mathbf{y}_k(n)))]$$

ここで、 $q_n(\mathbf{x})$  は  $N(\mathbf{x}; \mathbf{x}_n, \Sigma_n)$ 、 $D$  はカルバックライブラーダイバージェンスを表す。 $\mathbf{L}'$  は以下で表される  $\Phi$  によって下界を与えられ、二つの観測変数と隠れ変数をそれぞれ部分列  $\tilde{\mathbf{y}}_k$ ,  $\mathbf{y}_k$  と状態ベクトル  $\mathbf{x}$  に置き換えることで、[Verbeek 08]と同様のアルゴリズムを用いて最大化することができる。

\*1 これが成立するのは  $\mathbf{y}_k(t)$  と  $\tilde{\mathbf{y}}_k(t)$  が i.i.d とみなせるときである。i.i.d とは independent and identically distributed の略でサンプルが独立で同じ確率分布に従うことを意味する。今回の場合では  $k$  が十分大きいとき、その仮定は成立するとしてよい。

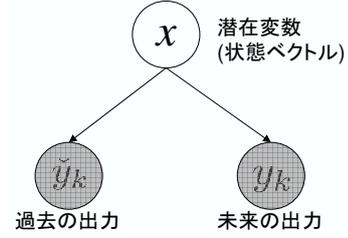


図 1: CCA に基づく部分空間同定法の確率的表現

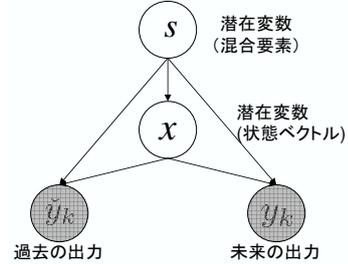


図 2: 局所線形モデルのアライメントによる拡張アプローチ

なお、パラメータの初期化も [Verbeek 08] と同様に、拡張 LLE [Ham, Lee and Saul 06] を用いて行う。

$$\begin{aligned} \mathbf{L}' \geq \Phi &= \sum_{n=1}^N H(q_n(s)) + H(q_n(\mathbf{x})) + \sum_{n=1}^N \sum_{s=1}^C \\ &[q_n(s) \int q_n(\mathbf{x}) \log p(\tilde{\mathbf{y}}_k(n), \mathbf{y}_k(n), \mathbf{x}, s) d\mathbf{x}] \end{aligned} \quad (4)$$

ここで、 $H$  は分布のエントロピー、そして  $q_n(s)$  は混合要素の分布を示している。

## 2.3 学習アルゴリズム

一旦状態ベクトル列が得られれば、動的システムは二つの回帰問題を計算することで推定される。しかしながら、遷移モデルと観測モデルの二つの確率モデルが以下のような混合確率モデルとして表されることを確認しておく。

$$\begin{aligned} p(\mathbf{x}(t+1) | \mathbf{x}(t)) &= \sum_{s=1}^C p(s | \mathbf{x}(t)) p(\mathbf{x}(t+1) | \mathbf{x}(t), s) \\ p(\mathbf{y}(t) | \mathbf{x}(t)) &= \sum_{s=1}^C p(s | \mathbf{x}(t)) p(\mathbf{y}(t) | \mathbf{x}(t), s) \end{aligned} \quad (5)$$

ここで、 $s$  は  $C$  個の混合要素についてのインデックスであり、 $p(s | \mathbf{x}(t))$  は  $p(\mathbf{x}(t), s) / p(\mathbf{x}(t))$  によって計算される大域的な座標系での状態を与えられた時の  $s$  番目の混合要素の重みである。よって、我々が推定すべきパラメータは以下の回帰問題におけるシステムパラメータ  $A_s$  と  $C_s$ 、共分散行列  $\Sigma_s^t$  と  $\Sigma_s^o$ 、そして残差  $\mu_s^t$  と  $\mu_s^o$  である。

$$\begin{aligned} p(\mathbf{x}(t+1) | \mathbf{x}(t), s) &= N(\mu_s^t + A_s \mathbf{x}(t), \Sigma_s^t) \\ p(\mathbf{y}(t) | \mathbf{x}(t), s) &= N(\mu_s^o + C_s \mathbf{x}(t), \Sigma_s^o) \end{aligned} \quad (6)$$

この回帰問題は各々の線形部分空間に対して最小二乗法を用いることで計算される。結果として、アライメントアプローチによって非線形動的システムを学習する手順は以下のようになる。

1.  $N+2k-1$  の有限時系列データから部分列  $\check{y}_k(t)$  と  $y_k(t)$  を生成する。
2. 拡張LLE[Ham, Lee and Saul 06]によって、部分列  $\check{y}_k(t)$  と  $y_k(t)$  に関するパラメータの初期化を行う
3. [Verbeek 08] と同様のアルゴリズムを用いて、目的関数 (4) を最適化する
4. 最小二乗法を用いて回帰問題を解き、混合モデル (6) のシステムパラメータを得る。

このアルゴリズムにおいて、前もって調整しなければならないパラメータは混合要素の数  $C$ 、低次元多様体の次元  $d$ 、部分列の長さ  $k$ 、LLEにおける最近傍の数である。  $k$  に関しては部分空間同定法の観点から適度に大きい値であることが望ましい。しかし、モデルの複雑さを決定するパラメータ  $d$  と  $C$  の選択に関しては今後の課題である。

### 3. 大域的状態空間におけるフィルタリング

本章では、非線形システムのフィルタリングについて考える。フィルタリングとは、 $p(\mathbf{x}(t)|\mathbf{y}(1:t))$  の帰納的な計算である。ここで  $\mathbf{y}(1:t)$  は現在の時刻  $t$  までの全ての観測を意味する。本アルゴリズムにおけるフィルタリングは、混合要素の周辺化を考える事を除いて、カルマンフィルタと全く同様に計算することができる。以下のように、予測ステップにおいて  $p(\mathbf{x}(t+1)|\mathbf{y}(1:t))$  を計算し、更新ステップにおいて  $p(\mathbf{x}(t)|\mathbf{y}(1:t))$  を得る。

$$p(\mathbf{x}(t+1)|\mathbf{y}(1:t)) = \quad (7)$$

$$\int_{\mathbf{x}_t} \sum_{s=1}^C p(s|\mathbf{x}(t)) p(\mathbf{x}(t+1)|\mathbf{x}(t), s) p(\mathbf{x}(t)|\mathbf{y}(1:t)) d\mathbf{x}_t$$

$$p(\mathbf{x}(t+1)|\mathbf{y}(1:t+1)) = \quad (8)$$

$$\alpha \sum_{s=1}^C p(s|\mathbf{x}(t+1)) p(\mathbf{y}(t+1)|\mathbf{x}(t+1), s) p(\mathbf{x}(t+1)|\mathbf{y}(1:t), s)$$

ここで  $\alpha$  は正規化の係数である。

このフィルタリングアルゴリズムにおいて特徴的な点は、我々はアライメントされた大域的な座標系を状態空間として使っているため、 $p(\mathbf{x}(t)|\mathbf{y}(1:t))$  が単峰性を持つ点である。そのため、 $p(\mathbf{x}(t)|\mathbf{y}(1:t))$  は最も大きい重みを持つ混合要素  $s^*$  による分布  $p(\mathbf{x}(t)|\mathbf{y}(1:t), s^*)$  によってよく近似される。

$$p(\mathbf{x}(t)|\mathbf{y}(1:t)) = \sum_{s=1}^C p(s|\mathbf{y}(1:t)) p(\mathbf{x}(t)|\mathbf{y}(1:t), s)$$

$$\approx p(\mathbf{x}(t)|\mathbf{y}(1:t), s^*) \quad (9)$$

我々が行うフィルタリングにおいては、各々の時間  $t$  において一つのガウス分布を保ちさえすればよいので、スイッチングカルマンフィルタのような場合と違って、状態の混合要素の数が指数関数的に増大するのを避けることができる。提案するアルゴリズムのこの特徴は、フィルタリングを行うにあたって非常に重要な意味を持つことになる。

### 4. 実験結果

本章では、前章までに述べたアプローチを用いたいくつかの実験結果を示す。4.1節では、提案するアルゴリズムと代表的な線形アルゴリズムを比較し、フィルタリングにおける比較性能を比較する。4.2節では、提案アルゴリズムを高次元で複雑なシステムであるモーションキャプチャのデータに適用し、ヒューマンモーションデータの異常を検知することで本手法の有用性を示す。

#### 4.1 比較結果

比較するアルゴリズムはカルマンフィルタベースアルゴリズム (KFS) [Gharamani and Hinton 96] と直交分解に基づく線形部分空間法 (ORT) [Katayama 05] の二つの線形動的システム学習アルゴリズムである。対象としたシステムは、代表的な非線形動的システムであるローレンツアトラクタである。ローレンツアトラクタにおける潜在変数  $\mathbf{x}(t)$  は以下の微分方程式の解となっており、得られる観測  $\mathbf{y}(t)$  は潜在変数にガウシアンノイズを加えたものである。

$$\begin{aligned} \dot{x}_1(t) &= -ax_1(t) + ax_2(t) \\ \dot{x}_2(t) &= -x_1(t)x_3(t) + rx_1(t) - x_2(t) \\ \dot{x}_3(t) &= x_1(t)x_2(t) - bx_3(t) \end{aligned} \quad (10)$$

設定したパラメータは、 $a=10, b=28, r=8/3$  であり、初期値は  $(-10.0, -10.0, 30.0)$  とした。220秒間のシミュレーションにおいて、0.1秒おきに観測を行い、2000点を訓練データ、続く200点をテストデータとして用いた。状態ベクトルの次元、及び部分列の長さ  $(n,k)$  は  $(4,5)$  である。提案アルゴリズムに関して、混合要素数は  $c=6$  とし、LLEの近傍数は20とした。手法の評価は、フィルタリングにおける予測対数尤度、つまり1ステップ予測によって行った。比較結果は表1に示すとおりであり、線形の手法と比較し、本手法は圧倒的に優れた結果を出していることがわかる。図3は、 $y_1(t)$  に相当するベクトル空間における各混合要素の分布を示している。各色が各混合要素に対応しており横軸が  $y_1(t)$ 、縦軸が  $y_1(t+1)$  を表す。この種のプロットは時系列データの力学を分析するのに有効であり、各々の混合要素の点がベクトル空間上で連続した領域に分布していることがわかる。また、図4は提案手法におけるフィルタリング手法が正しく機能していることを示す図である。

表 1: 予測尤度の比較

KFS	ORT	MIX
-9.73	-9.28	-6.69

#### 4.2 モーションキャプチャへの適用

本節ではヒューマンモーションデータに対して提案アルゴリズムを適用し、その有用性を示す\*2。データの最初の130フレームを訓練データ（歩行2サイクルに相当）、続く20フレームをテストデータとして用いた。本シミュレーションではテストデータの最後のフレームの姿勢データをランダムな姿勢データに置き換えることで人工的な異常を発生させ、提案するフィルタリング手法によってその異常を検知できるかを検証した。状態ベクトルの次元、及び部分列の長さ  $(n,k)$  は  $(5,5)$  である。提案アルゴリズムに関して、混合要素数は  $c=5$  とし、LLEの近傍数は8とした。図5は訓練データに対して、潜在空間を3次元上に投影し可視化したものである。各色はローレンツアトラクタの結果と同様に各混合要素に対応しており、矢印で示される人間の絵は、潜在変数から観測された次元に再構成して得られたものである。図5より、潜在空間において各混合要素が連続した領域に分布しており、人間の動作が各々の領域間で滑らかに遷移して潜在空間上で表されていることがわかる。テストデータに対して我々は異常検知を試みた。図6は左から、ランダムな姿勢に置き換える前の元々の姿勢、フィルタリングによって推定された姿勢、ランダムな姿勢に置き換えた後の姿勢を表す。フィルタリングによって姿勢（置き換える前の姿勢）

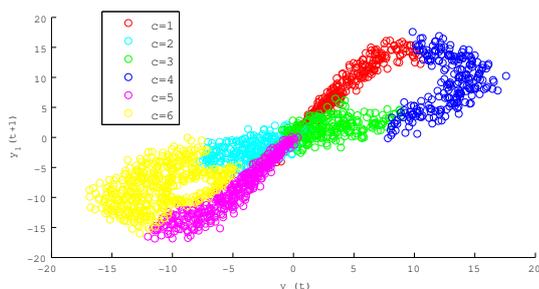


図 3: 各混合要素毎に色づけされた,  $y_1(t)-y_1(t+1)$  上に投影された訓練データ

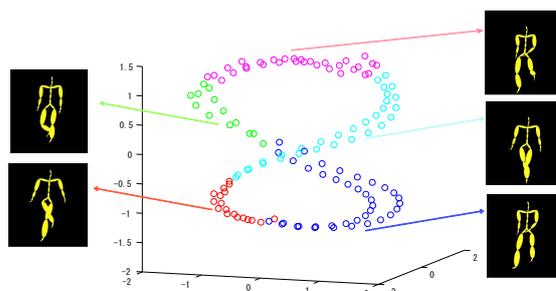


図 5: 各混合要素毎に色づけされた, 訓練データの低次元表現

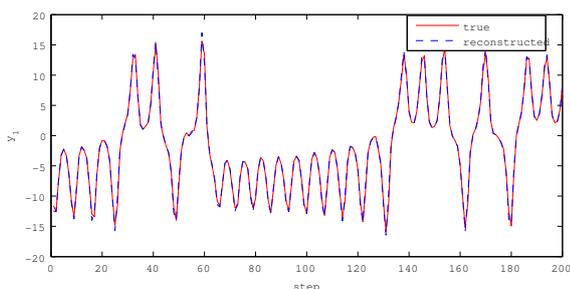


図 4: ローレンツアトラクタにおけるフィルタリングの出力

を正しく推定し, 異常を正しく検知出来ていることがわかる.

## 5. 結論

本論文では, 部分空間同定法の確率的解釈に基づいて, 局所座標系のアライメントによる画期的な非線形システム学習法を提案した. 部分空間同定法において状態ベクトル列を求める手順は正準相関分析と解釈することができ, これを基にして潜在変数モデルを提案した. そして, パターン認識や機械学習の分野における近年の成果である混合確率モデル学習法と同様にして非線形動的システムの学習法の枠組みを提案した. 我々の成果は動的システムと多様体学習のつながりへの洞察を与えることが期待される. また, 我々はヒューマンモーションデータへの適用や他手法との比較を述べ, 本手法の有用性を示した. 今後の課題として, 我々は動的システムと多様体学習の関係を明白にし, また本枠組みにおける混合要素数の決定というモデル選択法を発展させていきたい.

## 参考文献

- [Katayama 05] Tohru Katayama: Subspace Methods for System Identification(2005)
- [Roweis 01] Sam Roweis and Zoubin Ghahramani: Learning Nonlinear Dynamical Systems using the EM Algorithm(2001)
- [Palanthandalam 05] Harish J. Palanthandalam-Madapusi, Seth Lacy, Jesse B. Hoagg, and Dennis S. Bernstein:

\*2 データセットは以下のサイトよりダウンロードした.  
(<http://mocap.cs.cmu.edu/>)

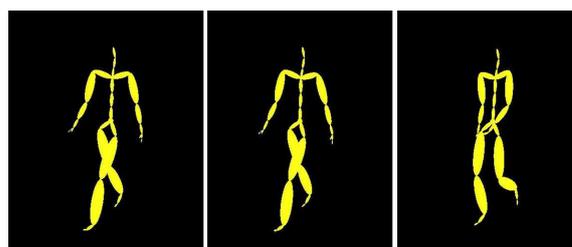


図 6: 最後のフレームにおける人間の姿勢の比較: (左) 置き換え前の姿勢, (中) 推定された姿勢, (右) 置き換え後の姿勢 (実際に観測された姿勢)

Subspace-based identification for linear and nonlinear systems(2005)

- [Kawahara 06] Yoshinobu Kawahara, Takehisa Yairi and Kazuo Machida: A Kernel Subspace Method by Stochastic Realization for Learning Nonlinear Dynamical Systems(2006)

- [Verbeek 08] Jakob Verbeek: Learning Nonlinear Image Manifolds by Global Alignment of Local Linear Models(2008)

- [Bach and Jordan 06] Francis R. Bach and Michael I. Jordan: A probabilistic Interpretation of Canonical Correlation Analysis(2006)

- [Desai and Pal 82] Uday B.Desai and Debajyoti Pal: A Realization Approach to Stochastic Model Reduction and Balanced Stochastic Realizations(1982)

- [Ham, Lee and Saul 06] Ji Hun Ham and Daniel D. Lee and Lawrence K. Saul: Learning High-Dimensional Correspondences from Low-Dimensional Manifolds(2006)

- [Ghahramani and Hinton 96] Zoubin Ghahramani and Geoffrey E.Hinton: Parameter Estimation for Linear Dynamical Systems(1996)

- [Wang and Fleet 06] Jack M.Wang and David J. Fleet: Gaussian process dynamical models for human motion(2006)