隠れマルコフモデルに基づく腕の動きの生成
HMM-Based Synthesis of Human Arm Motion

吉岡 元貴 † **益子 貴史** † **小林 隆夫** †
Mototaka Yoshioka Takashi Masuko Takao Kobayashi

1. まえがき

コンピュータのインターフェースとしてはマウスやキーボードを利用したインタフェースが現在主流であるが，音声 (バーバル) インターフェース，また身ぶりや表情を利用したノンバーバルインタフェース等，音声，画像を融合したヒューマンインタフェースの必要性が高まっている．

近年，音声合成の分野では，大規模な音声データベースの整備とコンピュータの処理能力の向上を背景に，コーパスベースと称される手法 (例えば [1]) の研究が行われており，また，我々も隠れマルコフモデル (HMM) に基づく音声合成を提案している [2]．これらの手法は，従来の規則に基づいた合成とは異なり，大量のデータを用いた自動学習や音声単位選択に基づいているため，自然性の高い音声を合成できる．

一方，ジェスチャーの動画像の合成については，モーションキャプチャーによって採取したデータをそのまま再現したり，なんらかの手法で補間しているのが現状である．前者は自然な動きを生成する点において優れているが，採取したデータ以外の動きを生成することは困難である．また，後者も，補間の方法によっては生成された動作が不自然なものとなってしまうことがある．

一方，我々の提案する HMM に基づくハンドジェスチャーアニメーション生成 [3] では，収録した手の動作を用いて学習した HMM から，尤度最大の意味で最適な動作を生成する．その結果，手の動作の統計的性質を反映した，自然に近い動作のアニメーションを生成することができる．この際，動的特徴量を考慮することで，特別な補間処理や平滑化処理を行わなくても，滑らかなアニメーションの生成が可能となる．しかし，手話など一般的な応用を考えた場合，ジェスチャーは手のみならず，手首，腕を含む上半身の動作も考慮する必要がある．そこで本論文では動画像の生成に HMM を用い，腕の動きを生成するシステムについて検討した結果を述べる．

2. アニメーション生成システム

本論文で提案するアニメーション生成システムでは，任意の動作のアニメーションを生成するために，動作を記号レベルで記述することとし，動作をいくつかの基本的なパターンに分解して，それぞれにラベル (基本動作ラベル) を付与する．これにより，動作アニメーション生成の問題は，基本動作ラベルの組合せによって表現された一連の動作からの形状パラメータ列の生成と，生成された形状パラメータ列のアニメーションへの変換に帰着される．

ここでは，動作ラベル列から形状パラメータ列を生成する手法として，HMM に基づくパラメータ生成 [4] を用いる．

†東京工業大学 大学院総合理工学研究科

また，動作をコンピュータグラフィックス上で可視化するために OpenGL を用いる．図 1 に示すように，6 つの立方体はそれぞれ頭，肩，背中，肘，腕，手首を表しており，黒線が接続部を表す．各立方体に形状パラメータ (XYZ 座標と回転角度 α, β, γ) を与えることで，腕の形状モデルを任意に動かすことができ，透視投影によって平面に描写する．

3. HMM に基づくパラメータ生成

HMM に基づくパラメータ生成手法は，学習と生成の 2 手順に分けられる．以下，本手法で用いる動的特徴量と HMM について述べたうえで，学習の概要と生成について述べる．

3.1 動的特徴量

特徴ベクトル，すなわち，HMM が出力するベクトルとして，静的パラメータ (形状パラメータ)，および，動的パラメータ (デルタパラメータ) を一つに結合したベクトルを用いる．

パラメータの次数を M として，時刻 (フレーム) t の形状パラメータを $\boldsymbol{x}_t$ (M 次の実ベクトル) とおく．長さ T の形状パラメータ列 $(\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_T)$ が与えられたとき，式 (1) によって，$\boldsymbol{x}_t$ の n 次デルタパラメータ $\Delta^n \boldsymbol{x}_t (n = 1, 2)$ を定義する．

$$\Delta^n \boldsymbol{x}_t = \sum_{i=-L_n}^{L_n} w^{(n)}(i) \boldsymbol{x}_{t+i}, \quad 0 \le n \le 2 \tag{1}$$

ただし，$L_0 = 0$, $w^{(0)}(0) = 1$, $i \neq 0$ のとき $w^{(0)}(i) = 0$,

$$w^{(1)}(i) = \begin{cases} \dfrac{i}{\sum_{j=-L_n}^{L_1} j^2}, & \text{for} -L_1 \le i \le L_1 \\ 0, & \text{otherwise} \end{cases} \tag{2}$$

$s_0 = \sum_{j=-L_2}^{L_2} 1$, $s_1 = \sum_{j=-L_2}^{L_2} j^2$, $s_2 = \sum_{j=-L_2}^{L_2} j^4$ として，

$$w^{(2)}(i) = \begin{cases} \frac{s_0 i^2 - s_1}{2(s_2 s_0 - s_1^2)}, & \text{for} -L_2 \le i \le L_2 \\ 0, & \text{otherwise} \end{cases} \tag{3}$$

である．

形状パラメータ，および，1 次と 2 次のデルタパラメータを 1 つのベクトル $\boldsymbol{o}_t$ にまとめて

$$\boldsymbol{o}_t = \left[\boldsymbol{x}_t', \Delta^1 \boldsymbol{x}_t', \Delta^2 \boldsymbol{x}_t'\right]' \tag{4}$$

とおく．ただし，$\boldsymbol{x}'$ は $\boldsymbol{x}$ の転置である．こうして得られる $\boldsymbol{o}_t$ を時刻 t の特徴のベクトルとする．

特徴ベクトル列 $\boldsymbol{O} = (\boldsymbol{o}_1, \boldsymbol{o}_2, \ldots, \boldsymbol{o}_T)$ が，HMM λ から生成されたと仮定する．HMM λ の状態数を N，全状態の集合を $S = 1, \ldots, N$ とおく．各状態 $q \in S$ に対

応する出力分布は，平均 $\boldsymbol{\mu}_q$，共分散行列 $\boldsymbol{U}_q$ のガウス分布とする．状態遷移確率行列を $A = \{a_{ij}\}_{i,j=1}^{N}$ とおく．さらに，HMM λ は，スキップがない left-to-right 型，すなわち，$j-1 \leq i \leq j$ のとき以外は $a_{ij} = 0$ とする．特徴ベクトル列 $\boldsymbol{O}$ が HMM λ から生成される尤度は，$a_{01} = 1$，$i \neq 1$ のとき $a_{0i} = 0$ とおいて，式 (5) によって定義される．

$$P(\boldsymbol{O}|\lambda) = \sum_{\boldsymbol{q} \in S^T} P(\boldsymbol{q}, \boldsymbol{O}|\lambda) \tag{5}$$

$$P(\boldsymbol{q}, \boldsymbol{O}|\lambda) = \prod_{t=1}^{T} a_{q_{t-1}q_t} \mathcal{N}(\boldsymbol{o}_t; \boldsymbol{\mu}_{q_t}, U_{q_t}) \tag{6}$$

ただし，$\mathcal{N}$ はガウス関数

$$\mathcal{N}(\boldsymbol{o}_t; \boldsymbol{\mu}_{q_t}, \boldsymbol{U}_{q_t}) = \frac{1}{(2\pi)^{3M/2}\sqrt{|\boldsymbol{U}_{q_t}|}} \times \exp\left[-\frac{1}{2}(\boldsymbol{o}_t - \boldsymbol{\mu}_{q_t})' \boldsymbol{U}_{q_t}^{-1} (\boldsymbol{o}_t - \boldsymbol{\mu}_{q_t})\right] \tag{7}$$

である．

3.2 学習

学習サンプルとして，n 個の特徴ベクトル列 $\boldsymbol{O}^{(1)} = (\boldsymbol{o}_t^{(1)})_{t=1}^{T_1}, \boldsymbol{O}^{(2)} = (\boldsymbol{o}_t^{(2)})_{t=1}^{T_2}, \ldots, \boldsymbol{O}^{(n)} = (\boldsymbol{o}_t^{(n)})_{t=1}^{T_n}$ が与えられたとき，式 (8) によって定義される尤度を最大化にする HMM λ を求める．

$$\prod_{k=1}^{n} P(\boldsymbol{O}^{(k)}|\lambda) \tag{8}$$

一般的には，式 (8) を最大にする解を解析的に求めることは困難である．そこで，EM(expectation maximization) アルゴリズム [5] に基づく繰り返しにより，極大値をとる λ を求める．

次に，各学習サンプル $\boldsymbol{O}^{(k)}$ に対し，状態継続長に関する確率を求め，それを基にした最尤推定によって，状態継続長モデルの学習を行う．ただし，状態遷移列が $\boldsymbol{q}$ のとき，状態 i に滞在した回数を $d_i(\boldsymbol{q})$ とおき，状態 i に滞在する回数 (状態継続長) は，平均 m_i，分散 σ_i^2 のガウス分布に従うとする．まず，学習サンプル $\boldsymbol{O}^{(k)}$ が HMM λ から生成されるときに，時刻 t_0 から t_1 の間のみ，状態 i に滞在する確率 $\chi_{t_0,t_1}^{(k)}(i)$ を，次式によって求める．

$$\chi_{t_0,t_1}^{(k)}(i) = (1 - \gamma_{t_0-1}^{(k)}(i)) \prod_{t=t_0}^{T_k} \gamma_t^{(k)}(i)(1 - \gamma_{t_1+1}^{(k)}(i)) \tag{9}$$

ただし，$\gamma_t^{(k)}(i)$ は，時刻 t のとき状態 i に滞在している確率で，$\gamma_{-1}^{(k)}(i) = \gamma_{T_k+1}^{(k)}(i) = 0$ とする．この $\chi_{t_0,t_1}^{(k)}(i)$ を用いて，次のように状態継続長モデルのパラメータを決定する．

$$m_i = \frac{\sum_{k=1}^{n}\sum_{t_0=1}^{T_k}\sum_{t_1=t_0}^{T_k} \chi_{t_0,t_1}^{(k)}(i)(t_1 - t_0 + 1)}{\sum_{k=1}^{n}\sum_{t_0=1}^{T_k}\sum_{t_1=t_0}^{T_k} \chi_{t_0,t_1}^{(k)}(i)} \tag{10}$$

$$\sigma_i^2 = \frac{\sum_{k=1}^{n}\sum_{t_0=1}^{T_k}\sum_{t_1=t_0}^{T_k} \chi_{t_0,t_1}^{(k)}(i)(t_1 - t_0 + 1)^2}{\sum_{k=1}^{n}\sum_{t_0=1}^{T_k}\sum_{t_1=t_0}^{T_k} \chi_{t_0,t_1}^{(k)}(i)} - m_i^2 \tag{11}$$

3.3 生成

尤度最大の意味で，HMM λ の最適状態遷移列 $\boldsymbol{q} = (q_1, q_2, \ldots, q_T)$，および，出力特徴ベクトル列 $\boldsymbol{O} = \{\boldsymbol{o}_t\}_{t=1}^{T}$ を求める．つまり，尤度 $P(\boldsymbol{q}, \boldsymbol{O}|\lambda)$ を最大化する $\boldsymbol{q}$ と $\boldsymbol{O}$ を求める．

ここで，表記上の簡略化のために，次のように，形状パラメータ列，HMM 出力ガウス分布パラメータなどをまとめて表記する．

$$\boldsymbol{x} = [\boldsymbol{x}_1', \boldsymbol{x}_2', \ldots, \boldsymbol{x}_T']', \tag{12}$$

$$\boldsymbol{\mu} = [\boldsymbol{\mu}_{q_1}', \boldsymbol{\mu}_{q_2}', \ldots, \boldsymbol{\mu}_{q_T}']', \tag{13}$$

$$\boldsymbol{U} = \mathrm{diag}[\boldsymbol{U}_{q_1}, \boldsymbol{U}_{q_2}, \ldots, \boldsymbol{U}_{q_T}] \tag{14}$$

ただし，$\mathrm{diag}[\boldsymbol{U}_{q_1}, \boldsymbol{U}_{q_2}, \ldots, \boldsymbol{U}_{q_T}]$ は行列 $\boldsymbol{U}_{q_1}, \boldsymbol{U}_{q_2}, \ldots, \boldsymbol{U}_{q_T}$ を対角に並べた $3MT \times 3MT$ 行列である．また，次のように，デルタパラメータを求めるための窓係数

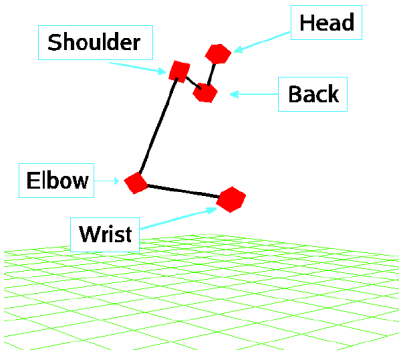

図 1: OpenGL で表現された腕の形状モデル

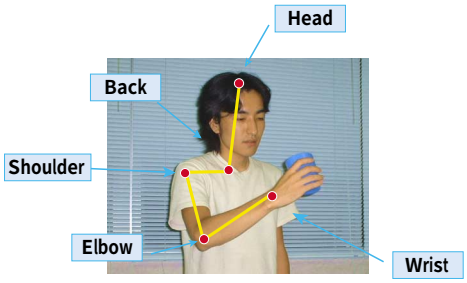

図 2: 使用したパラメータ

図 3: 基本動作ラベル

$w^{(n)}(i)$ をまとめて，行列 $\boldsymbol{W}$ を作る．

$$\boldsymbol{W} = [\boldsymbol{w}_1, \boldsymbol{w}_2, \ldots, \boldsymbol{w}_T]' \quad (15)$$

$$\boldsymbol{w}_t = \begin{bmatrix} (\boldsymbol{w}_t)_{11} & (\boldsymbol{w}_t)_{12} & \ldots & (\boldsymbol{w}_t)_{1T} \\ (\boldsymbol{w}_t)_{21} & (\boldsymbol{w}_t)_{22} & \ldots & (\boldsymbol{w}_t)_{2T} \\ (\boldsymbol{w}_t)_{31} & (\boldsymbol{w}_t)_{32} & \ldots & (\boldsymbol{w}_t)_{3T} \end{bmatrix}' \quad (16)$$

$$(\boldsymbol{w}_t)_{ij} = w^{(i)}(j-t)\boldsymbol{I}_M \quad (17)$$

ただし，$\boldsymbol{I}_M$ は M 次の単位行列である．

以上の表記を用いると，$\log P(\boldsymbol{q}, \boldsymbol{O}|\lambda)$ は，

$$\log P(\boldsymbol{q}, \boldsymbol{O}|\lambda) = \alpha \sum_{i=1}^{N} \log P(d_i(\boldsymbol{q})) - \frac{1}{2}\log|\boldsymbol{U}| - \frac{3MT}{2} log 2\pi - \frac{1}{2}(\boldsymbol{W}\boldsymbol{x} - \boldsymbol{\mu})'\boldsymbol{U}^{-1}(\boldsymbol{W}\boldsymbol{x} - \boldsymbol{\mu}) \quad (18)$$

と書き換えることができる．ただし，α は状態継続長の尤度に関する重みである．ここで，各状態の継続長は状態継続長モデルのみによって決定されるとしている．計算を簡単にするために，状態継続長の尤度 (式 (18) 第一項) のみを最大にする状態遷移列を求める．このことは，α を十分大きくとることに相当する．最適状態遷移列は，

$$\sum_{i=1}^{N} \log P(d_i(\boldsymbol{q})) = \sum_{i=1}^{N} \log \mathcal{N}(d_i(\boldsymbol{q}); m_i, \sigma_i^2) \quad (19)$$

を最大にする $\boldsymbol{q}$ として，(離散値を連続分布でモデル化しているため) 近似的に，

$$\boldsymbol{q} = (\underbrace{1, 1, \ldots, 1}_{\lceil m_1+1/2 \rceil}, \underbrace{2, 2, \ldots, 2}_{\lceil m_2+1/2 \rceil}, \ldots, \underbrace{N, N, \ldots, N}_{\lceil m_N+1/2 \rceil}) \quad (20)$$

$$T = \sum_{i=1}^{N} \lceil m_i + 1/2 \rceil, \quad (21)$$

ただし，$\lceil m \rceil$ は，m を超えない最大の整数，と求まる．

以下では，状態遷移列 $\boldsymbol{q}$ を固定して，尤度を最大化する形状パラメータ列 $\boldsymbol{x}$ を求める．$\boldsymbol{0}_{TM}$ を TM 次の零ベクトルとして，

$$\frac{\partial \log P(\boldsymbol{q}, \boldsymbol{O}|\lambda)}{\partial \boldsymbol{x}} = \boldsymbol{0}_{TM} \quad (22)$$

とおくことにより，$\boldsymbol{x}$を定める連立方程式

$$\boldsymbol{R}\boldsymbol{x} = \boldsymbol{r} \quad (23)$$

ただし，

$$\boldsymbol{R} = \boldsymbol{W}'\boldsymbol{U}^{-1}\boldsymbol{W} \quad (24)$$

$$\boldsymbol{r} = \boldsymbol{W}'\boldsymbol{U}^{-1}\boldsymbol{\mu} \quad (25)$$

(26)

を得る．$\boldsymbol{R}$ は，$TM \times TM$ 行列であるため，式 (23) を解くためには，$O(T^3M^3)$ の演算が必要となる．ただし，$\boldsymbol{U}_q$ が対角行列の場合には，形状パラメータの各次数を独立に計算することができるので，計算量は $O(T^3M)$ となる．

4. 実験

3. で述べた，HMM に基づくパラメータ生成手法を用いて，アニメーションの生成実験を行った．

4.1 学習データ

学習データとして，図 3 のように静止状態からテーブルにあるグラスを掴み，グラスの水を飲み，元に戻すまでの一連の動作を 20 回モーションキャプチャーで採取した．得られたデータのフレームレートは 30 フレーム毎秒，総フレーム数は 5309 フレーム (データ長さ：約 177 秒) であった．

4.2 学習パラメータ

学習パラメータとしては，図 2 のように 5 箇所 (頭，肩，背中，肘，手首) の三次元座標 (X，Y，Z) と回転角度 (α, β, γ) の 30 次元とし，**3.** で述べた 1 次のデルタパラメータ (30 次)，2 次のデルタパラメータ (30 次) を計算し，合計 90 次元とした．

4.3 学習単位

モデルの学習単位として，図 3 のように一連の動作を **stay，grab，drink，put，return** の 5 つに分割して基本動作とした．学習データに対して基本動作ラベルでラベル付けし，各基本動作ごとに HMM を学習した．

4.4 合成実験

3. の生成手法を用いて動作のパラメータ生成を行い，得られたパラメータ列をコンピュータグラフィック表示

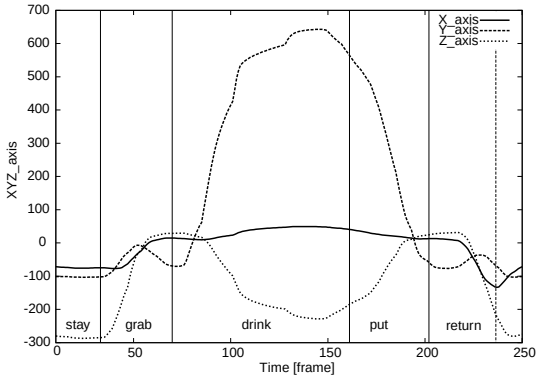

図 4: 動作 1(手首の XYZ 座標)

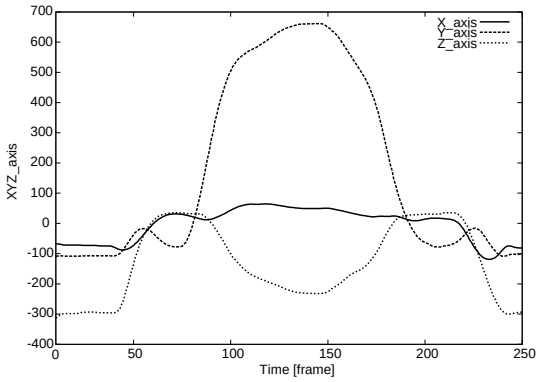

図 5: 合成動作 (動作 1) に対応する学習データ (手首の XYZ 座標)

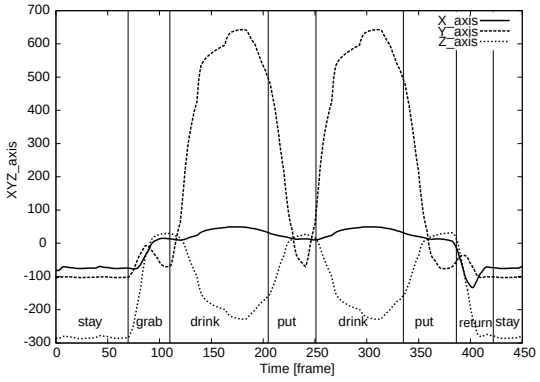

図 6: 動作 2(手首の XYZ 座標)

することで，動作のアニメーションを作成した．生成に用いた動作ラベルは，以下の通りである．

動作 1：**stay grab drink put return stay**

動作 2：**stay stay grab drink put drink put return stay stay**

4.5 結果

生成された動作 1 のアニメーションの長さは 277 フレーム (約 9 秒)，動作 2 のアニメーションの長さは 477 フレーム (約 16 秒) であった．図 4，図 6 に，生成された動作の一部 (手首の XYZ 座標) を示す．縦の実線は，生成時に用いた基本動作ラベルの区切りであり，各区間のラベルをグラフの下方に示してある．図 5 は，学習データより，生成データに対応する部分を切り出したものである．生成された動作と人間の行った動作をパラメータで比較してみると，全体的に見て，双方のパラメータは似たカーブを描いていることが分かる．

また，動作 2 のように静止状態からコップを掴み，水を飲み，一旦戻し，再び飲むという，学習データに無い動作に対しても，基本動作ラベルを組み合せることで，自然で滑らかな動作の生成をすることが可能であることを確認している．

5. おわりに

隠れマルコフモデルに基づくパラメータ生成手法を用いた動作生成の手法について述べた．合成実験の結果，自然な動作を生成することができた．また，基本動作 HMM の列の組合せを変えることにより，学習データにない動作でも，自然で滑らかな動作を生成することができた．しかし，今回，アニメーションの生成部分は，骨格を表示する程度の簡単なものであり，完全な人体モデルの表現とは言えない．また，合成した動作は比較的簡単で，大きな動作であったが，複雑な動きを合成する場合，衝突回避等の問題も今後の課題として検討する必要がある．また，合成動作の自然性の評価方法を確立する必要がある．

参考文献

[1] A.W. Black and N. Cambpbell, “Optimising selection of units from speech database for concatenative synthesis,” Proc. EUROSPEECH-95, pp.581–584, Sept. 1995.

[2] 益子貴史, 徳田恵一, 小林隆夫, 今井聖, “動的特徴量を用いた HMM に基づく音声合成,” 信学論 (D-II), vol.J79-D-II, no.12, pp.2184–2190, Dec.1996.

[3] 羽岡哲郎, 益子貴史, 小林隆夫, “隠れマルコフモデルに基づくハンドジェスチャーアニメーション生成,” 信学技報, vol.102, no.519, pp.43–48, Dec. 2002.

[4] 徳田恵一, 益子貴史, 小林隆夫, 今井聖, “動的特徴量を用いた HMM からの音声パラメータ生成アルゴリズム,” 日本音響学会誌, vol.53, no.3, pp.192–200, Mar.1997.

[5] A.P.Dempster, N.M.Laird, and D.B.Rubin, “Maximum likelihood from incomplete data via the EM algorithm,” Journal of Royal Statistical Society, Series B, vol39, pp.1–38, 1977.