オンライン事例ベース人物姿勢推定のための高速近傍探索手法

佐川裕一[†] 下坂正倫[†]
森武俊[†] 佐藤知正[†]

我々は複数カメラから復元する 3 次元ボクセルデータを入力としたマーカレスモーションキャプチャシステムの研究を進めている．本システムは事例ベース推論に基づいており，事前に出力姿勢の正解候補群を用意しておき，入力データに最も適合する正解候補を推定結果として出力するものである．この枠組みでは人物姿勢推定のタスクがボクセルデータと正解候補群間のマッチング処理にまで簡潔化され，従来の手法と比較して全体の計算コストを抑えることができる．しかし，正解候補数の増大に伴う類似度導出処理の計算コストの増大がシステムの実時間性を妨げる場合がある．そこで，類似度導出処理を適用する正解候補を絞り込む目的で，Parameter Sensitive Hashing (PSH) と呼ばれる高速近傍探索手法を導入する．PSH はハッシュ表を利用した手法であるが，ハッシュ表と対応付けられるハッシュ関数（ハッシュ値を出力する）の構築過程に冗長性が伴う．本稿では，PSH のハッシュ関数の冗長性を軽減する手法として，Sparse Incremental PSH を提案する．本手法の導入により，計算コストの削減，絞り込み精度の向上等の効果を得ることができる．従来の数十倍の正解候補に対して実験を行い，検索の精度を維持したまま計算コストを軽減できることを示す．

A Fast Near Neighbor Search Metric For Online Example-Based Human Pose Estimation

Yuichi Sagawa,[†] Masamichi Shimosaka,[†] Taketoshi Mori[†]
and Tomomasa Sato [†]

We have been working on a marker-less motion capture system that works in a multiple camera environment. This system assumes 3D voxel data to be the input, while discrete human posture data is assumed to be the output. The discreteness of human posture data is provided by an example-based approach, which constructs human posture candidates from a large motion capture database beforehand. During the estimation process, the most appropriate candidate will be chosen through a simple similarity calculation between voxel data and posture candidates. This approach will drastically reduce the computational cost compared to conventional methods, but increase of candidates will possibly lead to considerable computational cost. Therefore, prior to the similarity calculation phase, we introduce a near-neighbor search metric, which drastically decreases the similarity caculation frequency and the total computational cost. In this paper, we present a novel near-neighbor search metric, which is capable of dealing with much more candidates than the metric presented before, and yet maintaining the speed needed for online processing.

1. 序　　論

近年ロボットシステムによる日常生活支援に対する期待が高まっており，生活者の行動を把握し支援を行う知能化空間の研究が活発になってきている．AwareHome[1] やセンシングルーム[2] などは居住空間内にセンサを配置することで支援を行う例として代表的なものである．このような豊富な計測・蓄積環境と人の行動を多面的に把握する行動認識研究を融合させることで多種多様な生活支援システムへの展開が可能である．

我々は生活者の状態を可能な限り豊富に表現するシステムとして，カメラによるマーカレスモーションキャプチャシステムの構築に取り組んできた[3]．カメラを利用した人物姿勢推定の研究は従来から盛んに行われてきており，かつては人体モデルから姿勢を仮説的に生成し，画像の観測状態に最も適合する姿勢を探索する生成的なアプローチが盛んに取り組まれていた[4),5]．しかし複雑かつ未知な動作にも柔軟に対応できる一方，正確な初期化処理が要求される点やランダ

† 東京大学
University of Tokyo

ムサンプリング・非線形最適化計算に起因する計算の高コスト化，数値的なドリフトといった問題点が存在した．その一方で画像の観測状態から探索を伴わず，推定結果を直接導き出す識別的なアプローチも取り組まれている[6),7)]．このアプローチでは大量の学習サンプルを用いて識別的なモデルを学習する必要があり，観測空間から姿勢空間への写像に十分性を保証することが困難になる．しかし，モデル構築が容易であり，推定プロセスの高速化も期待できる点は注目すべき特性である．識別的なアプローチの中でも特筆すべきものとして，事例ベース推論を応用したものが挙げられる[8)~10)]．これらの研究では，事例ベース推論に基づき処理コストを軽減することで，高速化に成功している．

本研究でも，事例ベース推論を人物姿勢推定処理に応用している．具体的には，出力結果の正解候補群を事前に定義しておくことで離散値を出力することを可能にしている．推定姿勢が離散値であることは結果的にスパースな推定結果を招く可能性があるが，姿勢空間を正解候補群により十分に密に表現すれば，連続的な姿勢推定処理の近似とみなすことができる[9)]．つまり，正解候補数を増やし解像度を高くすればするほど精度の高い推定が可能となる．この枠組みでは，入力データと最も適合する正解候補がフレーム毎の推定結果として出力される．従って，オンライン処理を行う上で必要な計算は入力データと正解候補群間のマッチング処理に簡潔化される．計算コストは従来の多くの手法と比較して小さくなるが，正解候補数の増大に伴う類似度導出処理の計算コストの増大がシステムの実時間性を妨げる場合がある．そこで，我々は類似度導出処理を施す正解候補を絞り込む目的で，Parameter Sensitive Hashing（PSH）[8)] と呼ばれる高速近傍探索手法を導入した[3)]．

PSH では登録時に，特徴ベクトルを L 通りのハッシュ値に変換し，各々に対応する L 個のハッシュ表に登録する．検索時には，クエリとなる特徴ベクトルを L 通りのハッシュ値に変換し，すべてのハッシュ表を検索する．そして得られた特徴ベクトルの集合に対して類似度導出処理を施すことで，全体の類似度導出計算コストが削減される．L は検索性能に大きく影響するパラメータで，L が大きすぎると処理コストが高まる上に，非類似の無駄なデータが検索結果として得られる可能性も高まる．従って，L の値を低く抑えた上で目標の検索精度，及び網羅性に到達することが望ましい．しかし，PSH はハッシュ値を出力するハッシュ関数を構築する際に，ランダム要因を考慮するため，似た性質を持ったハッシュ関数が複数生成される可能性を秘めている．我々はこのようなハッシュ関数の冗長性に着目し，ハッシュ関数を効率的に構築する新たな手法，Sparse Incremental PSH を提案する．本手法の導入により L を低く抑えることが可能となり，計算コストの削減，及び絞り込み精度の向上へとつながる．その結果，従来の数十倍の正解候補を用意した環境でも最大 30[FPS] でのオンライン処理が可能となる．

本稿では，まず事例ベース推論により実現される人物姿勢推定システムの構成を説明する．次に，提案する高速近傍探索手法について述べる．最後に人工データ，及び実画像データを使用した実験を通して提案する手法の有効性を確認する．

2. 事例ベース人物姿勢推定

本システムでは，視体積交差法 により復元される 3 次元ボクセルデータ $\boldsymbol{v}(t)$ を入力データとする．また，出力姿勢は連続的な関節角データ $\boldsymbol{\theta}(t)$ ではなく，ラベル $y(t)$ に対応する関節角データの離散値 $\boldsymbol{\theta}_{y(t)}$ となる．つまり，N_y 種類の姿勢から構成される正解姿勢候補群 $\{\boldsymbol{\theta}_j\}_{j=1}^{N_y} \equiv \boldsymbol{\mathcal{Y}}$ を事前に定義しておき，時系列に沿って最適なラベル $y(t)$ を選び出すことで推定姿勢 $\boldsymbol{\theta}_{y(t)} \in \boldsymbol{\mathcal{Y}}$ を出力する．

上記の枠組みの導入により，複雑な人物姿勢推定のタスクは 3 次元ボクセルデータ $\boldsymbol{v}(t)$ と正解姿勢候補群 $\boldsymbol{\mathcal{Y}}$ 間のマッチング処理に簡潔化される．すなわち，$\boldsymbol{v}(t)$ と $\{\boldsymbol{\theta}_j\}_{j=1}^{N_y}$ 間の類似度列 $\{\phi_t(j)\}_{j=1}^{N_y}$ に基づき最適なラベル $y(t) = j$ が選択される．

なお，類似度 $\phi_t(j)$ を計算するために，特徴量抽出関数 $Q(\boldsymbol{v})$ を経て $\boldsymbol{v}(t)$ から特徴量 $\boldsymbol{q}(t)$ を抽出する（**式 1** 参照）．

$$\boldsymbol{q}(t) = Q(\boldsymbol{v}(t)) \tag{1}$$

一方で，正解姿勢候補 $\boldsymbol{\theta}_j \in \boldsymbol{\mathcal{Y}}$ から，ボクセル抽出関数 $V(\boldsymbol{\theta})$[3)] による人工ボクセルデータ $\boldsymbol{v}_j$ への変換を経由することで特徴量 $\boldsymbol{q}_j$ を抽出する（**式 2**，**式 3** 参照）．

$$\boldsymbol{v}_j = V(\boldsymbol{\theta}_j) \tag{2}$$

$$\boldsymbol{q}_j = Q(\boldsymbol{v}_j) = Q(V(\boldsymbol{\theta}_j)) \tag{3}$$

このように，$\boldsymbol{q}(t)$（クエリ特徴ベクトル）および $\{\boldsymbol{q}_j\}_{j=1}^{N_y} \equiv \boldsymbol{\mathcal{Q}}$（正解候補特徴ベクトル群）への変換を行うことで，類似度導出関数 $S(\boldsymbol{q}, \tilde{\boldsymbol{q}})$ により類似度列 $\{\phi_j(t)\}_{j=1}^{N_y}$ が導出可能となる（**式 4** 参照）．

$$\phi_j(t) = S(\boldsymbol{q}(t), \boldsymbol{q}_j) \tag{4}$$

その結果，類似度列 $\{\phi_j(t)\}_{j=1}^{N_y}$ の中で最大類似度を出力するラベル $y(t)$ を選択することで推定姿勢 $\boldsymbol{\theta}_{y(t)}$ が出力可能である．しかし，ラベル間の遷移を考慮し

ないことで不適切な遷移や推定結果の振動が発生してしまう．そこで，1 次マルコフ性を考慮したグラフィカルモデルによりラベル間の遷移確率を表現し，スムージングの効果を得る[3]．

処理全体の概略図を**図 1** に示し，タイムチャートを**図 2** に示す．視体積交差法の一部の処理（背景差分によりシルエット画像を取得し，シルエット領域に対応する視体積を求める処理）は複数のサーバ PC で並列計算される．その後，データをネットワーク越しにクライアント PC に送信し，視体積の積領域を求める処理，及び人物姿勢推定処理を行う．画像キャプチャのインターバル時間内に 1 フレーム分の処理が遅延なく完了すれば，実時間性が保証される．

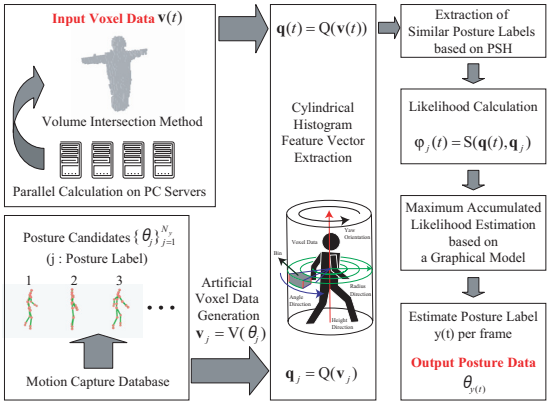

図 1　Outline of the Human Pose Estimation System

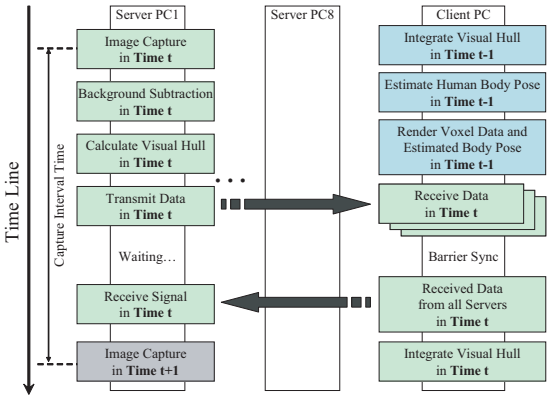

図 2　Time Chart of the Overall Process

3.　高速近傍探索手法

類似度列 $\{\phi_j(t)\}_{j=1}^{N_y}$ を導出する際に，類似度導出関数 $S(\boldsymbol{q},\tilde{\boldsymbol{q}})$（**式 4** 参照）を N_y 回適用する必要がある．N_y の数が大きくなるにつれ（人物姿勢空間の解像度を高めるにつれ），全体の計算コストが増大し，オンラインで姿勢推定処理を遂行することが困難となる．そこで，高速近傍探索手法を導入し，類似度導出関数を適用する正解姿勢候補を絞り込む．ここで言う高速近傍探索手法とは，クエリ特徴ベクトル $\boldsymbol{q}(t)$ を入力とし，類似の可能性が高いラベル集合（類似ラベル群）$\{a_j\}_{j=1}^{N_a} \equiv \mathcal{A}$ を出力するアルゴリズムのことを指す．本研究では Locality Sensitive Hashing(LSH)[11] を拡張した Parameter Sensitive Hashing(PSH)[8] と呼ばれる手法を採用している．PSH はデータの類似性がパラメータ空間（本研究では姿勢空間）で保証される点が特徴で，入力空間（本研究では特徴ベクトル空間）で類似性が保証される LSH よりも本研究の用途に適している．また，LSH と同様に誤差の期待値と計算コストの関係が明確であるため，識別器のデザインが容易である．

3.1　Parameter Sensitive Hashing(PSH)

PSH のハッシュ関数群 $\{\boldsymbol{h}^{(l)}(\boldsymbol{q})\}_{l=1}^{L}$ は K 個の decision stump $\{\boldsymbol{d}_m^{(k)}(\boldsymbol{q})\}_{k=1}^{K}$ を bit 連結したもので，$\{\boldsymbol{d}_m^{(k)}(\boldsymbol{q})\}_{k=1}^{K}$ の出力に基づきハッシュ値が決定される（**式 5** 参照）．なお，decision stump $\boldsymbol{d}_m^{(k)}(\boldsymbol{q})$ はバイナリ出力の関数で，参照する $\boldsymbol{q}$ の要素 $\{\boldsymbol{q}\}_m$ と識別用の閾値 T_m によって定義される（**式 6** 参照☆）．

$$h(\boldsymbol{q}) = [d^{(1)}_{m_{(1)}}(\boldsymbol{q}), d^{(2)}_{m_{(2)}}(\boldsymbol{q}), \cdots, d^{(K)}_{m_{(K)}}(\boldsymbol{q})]^{\mathsf{T}} \quad (5)$$

$$d_m(\boldsymbol{q}) = \begin{cases} 1 & \text{if } \{\boldsymbol{q}\}_m \geq T_m, \\ 0 & \text{otherwise.} \end{cases} \quad (6)$$

ハッシュ探索の枠組みでは，等しいハッシュ値を持つ 2 つの特徴ベクトルは類似ペアと判断され，異なる値を持つ場合は非類似ペアと判断される．PSH の学習の目的はこの条件に対して感度の高いハッシュ関数を構築することである．従って，類似ペアのハッシュ値が等しくなる確率が高く，非類似ペアのハッシュ値が異なる確率が高いハッシュ関数を構築することが目的となる．そのためには，類似・非類似のデータに対する分離性能が高い decision stump の集合を学習してハッシュ関数に割り当てれば良い．

学習には，特徴ベクトルのペア $\boldsymbol{q},\tilde{\boldsymbol{q}}$ に類似・非類似の正解ラベル z，正解姿勢ラベル a とフラグ b を付与したサンプルペア群 $\{(\boldsymbol{q}^{(p)},\tilde{\boldsymbol{q}}^{(p)},z^{(p)},a^{(p)},b^{(p)})\}_{p=1}^{P}$ を使用する（a : $\boldsymbol{q}$ の正解姿勢ラベル，b : $\boldsymbol{q},\tilde{\boldsymbol{q}}$ の類似・非類似を識別可能かどうかのフラグ）．このサンプルペア群に対して最適化計算を施すことで類似・非類似のデータを効率良く分離する decision stump の集合 $\{\boldsymbol{d}_m^{(k)}(\boldsymbol{q})\}_{m=1}^{M}$（$\boldsymbol{q} \in \mathbb{R}^M$）が求まる[8]（以降この計算を ODS 計算 (Optimal Decision Stump Calculation)

☆ $\{\boldsymbol{q}\}_m$: ベクトル $\boldsymbol{q}$ の m 番目の要素

と呼ぶ）．なお，$\{\boldsymbol{d}_m^{(k)}(\boldsymbol{q})\}_{m=1}^{M}$ の中で類似・非類似の分離性能が高いものを上位 $\tilde{M}$ 個抜き出した集合を ODS 集合 $\boldsymbol{\mathcal{O}}$ と定義する．

また，正解ラベル z を定める基準指標として，関節座標ベクトル $\boldsymbol{x} \in \mathbb{R}^{r_x}$ （人体モデルの関節・端点の座標を連結したベクトル）の関節座標平均距離 $d_x(\boldsymbol{x}, \tilde{\boldsymbol{x}})$（各関節毎の距離の平均値，**式 7** 参照）を採用する．

$$d_x(\boldsymbol{x}, \tilde{\boldsymbol{x}}) = \frac{1}{r_x/3} \sum_{c=0}^{r_x/3-1} \sqrt{\sum_{r=1}^{3} (\{\boldsymbol{x}\}_{3c+r)} - \{\tilde{\boldsymbol{x}}\}_{3c+r})^2} \quad (7)$$

サンプルペア群はモーションキャプチャデータ $\{\boldsymbol{\theta}^{(n)}\}_{n=1}^{N}$ から抽出した関節座標ベクトル，及び特徴量のペア $\{(\boldsymbol{x}^{(n)}, \boldsymbol{q}^{(n)})\}_{n=1}^{N}$ と正解姿勢候補群 $\{\boldsymbol{x}_j\}_{j=1}^{N_y}$ を $d_x(\boldsymbol{x}, \tilde{\boldsymbol{x}})$ に基づき比較することで生成する．なお，モーションキャプチャデータから人工ボクセルデータを生成する際に $\sigma = 50[mm]$ のガウシアンノイズを付与している．その結果，特徴量を抽出する際に頑健性が向上する．$\boldsymbol{x}^{(n)}$ と $\{\boldsymbol{x}_j\}_{j=1}^{N_y}$ とを比較して $d_x(\boldsymbol{x}^{(n)}, \boldsymbol{x}_j) < R/(1+\epsilon)$ を満たす場合，$(\boldsymbol{q}^{(n)}, \boldsymbol{q}_j, z = 1, a = j, b = \text{false})$ を類似ペアとする．一方で，$d_x(\boldsymbol{x}^{(n)}, \boldsymbol{x}_j) > R$ を満たす姿勢ラベル j をランダムに選択し，$(\boldsymbol{q}^{(n)}, \boldsymbol{q}_j, z = -1, a = j, b = \text{false})$ を非類似ペアとする．なお，R は類似・非類似識別用の閾値であり，ϵ は識別器の locality-sensitive[8),11)] な性質を保証するための変数である．PSH における locality-sensitive な性質とは，クエリ特徴ベクトル $\boldsymbol{q}(t)$ を入力データとした時に，元データ $\boldsymbol{x}(t)$，出力ラベル $a \in \mathcal{A}$ に対して $d_x(\boldsymbol{x}(t), \boldsymbol{x}_a) < R$ が高い確率で満たされることを指す．

$$z = \begin{cases} +1 & \text{if } d_x(\boldsymbol{x}, \tilde{\boldsymbol{x}}) < R/(1+\epsilon), \\ -1 & \text{if } d_x(\boldsymbol{x}, \tilde{\boldsymbol{x}}) > R, \\ \text{not defined} & \text{otherwise.} \end{cases} \quad (8)$$

本研究では，ハッシュ関数構築を打ち切るかどうかの条件に閾値 T を利用している．ハッシュ関数の数 L は閾値 T に基づき自動的に決定される．以下にハッシュ関数構築の流れを示す．

(1) サンプルペア群に対して ODS 計算を行い，ODS 集合を得る．

(2) ODS 集合からランダムに K 個の decision stump を選び出すことで新しくハッシュ関数 $\boldsymbol{h}^{(l)}(\boldsymbol{q})$ を構築し，ハッシュ関数群 $\{\boldsymbol{h}^{(l)}(\boldsymbol{q})\}_{l=1}^{L}$ に加える．

(3) サンプルペア群における類似ペアの特徴ベクトルをハッシュ関数群 $\{\boldsymbol{h}^{(l)}(\boldsymbol{q})\}_{l=1}^{L}$ にクエリとして与える．検索結果として得られる類似ラベル群 $\mathcal{A}^{(p)}$ に正解姿勢ラベル $a^{(p)}$ が含まれていれば，$b^{(p)} = \text{true}$ と更新する．

(4) 類似ペア数を P_s，$b^{(p)} = \text{true}$ を満たす類似ペア数を P_e と定義した時に P_e/P_s の値が閾値 T に到達していればハッシュ関数の構築を打ち切る．そうでなければ 2. に戻る．

PSH の問題点の 1 つとして挙げられるのは，ハッシュ関数の冗長性である．目標のレベルまで網羅的な検索を可能にするために多くのハッシュ関数を必要とする傾向がある．その結果，似た性質を持つハッシュ関数が複数生成されてしまう可能性が高まる．このような冗長性が結果的に検索率の向上を招き，計算コストの向上を招いていると言える．問題点を明確にするために，PSH の性能を正確に測る指標が必要である．そこで，データセット $\{(\boldsymbol{x}^{(n)}, \boldsymbol{q}^{(n)})\}_{n=1}^{N}$ を用いた以下のような評価方法を導入する．

(1) 全てのデータおよび姿勢ラベルに対して $d_x(\boldsymbol{x}^{(n)}, \boldsymbol{x}_j) < \tilde{R}$ の評価を行い，満たされる回数を N_s とする（$\tilde{R}$ は類似・非類似識別用の閾値）．**類似率**は N_s/N_yN として定義される．

(2) PSH 識別器に特徴量 $\boldsymbol{q}^{(n)}$ を入力として与えた時に出力される類似ラベル群を $\{a_j^{(n)}\}_{j=1}^{N_a^{(n)}} \equiv \mathcal{A}^{(n)}$ とすると，**検索率**は $(\sum_{n=1}^{N} N_a^{(n)})/N_yN$ として定義される．

(3) 全てのデータおよび類似ラベルに対して $d_x(\boldsymbol{x}^{(n)}, \boldsymbol{x}_{a_j^{(n)}}) < \tilde{R}$ の評価を行い，満たされる回数を N_r とする．**再現率**は N_r/N_s として定義される．

(4) **適合率**は $N_r/(\sum_{n=1}^{N} N_a^{(n)})$ として定義される．

(5) **F 値**は 2× 再現率 × 適合率/(再現率+適合率) として定義される．

PSH 識別器の性能は再現率によって測ることができる．再現率が高い程，類似したラベルを漏れなく検索でき，網羅性が高いと言える．また，適合率は処理速度に影響を与える．適合率が高い程，無駄なラベルを類似度計算対象からはずすことができ，効率が向上する．最終的には再現率，適合率をともに向上させること，つまり F 値を向上させることが重要である．

3.2 Incremental PSH

本節では，新たな高速近傍探索手法，Incremental PSH を提案する．Incremental PSH の基本的な

アイデアは，似た性質を持ったハッシュ関数の構築を避けることで，ハッシュ関数の冗長性を軽減するというものである．従来の PSH と大きく異なるのは，一度の ODS 計算結果から全てのハッシュ関数を構築するのではなく，ODS 計算を繰り返しながら各ステップ毎に新しいハッシュ関数を逐次的に構築する点である．ODS 計算に用いるサンプルペア群を徐々に絞り込んでいくことでサンプルペア群によりカバーされる領域の重複を避け，独自の領域を高感度に識別するハッシュ関数の構築を可能とする．従って，本手法は人物姿勢空間を自動的に分割しながら，各領域を高感度に識別するハッシュ関数を逐次的に構築する手法であるとみなすこともできる．以下にハッシュ関数構築の流れを示す．

(1) $b^{(p)} = \text{false}$ を満たすサンプルペアの集合に対して ODS 計算を行い，ODS 集合を得る．ODS 集合からランダムに K 個の decision stump を選び出すことで新しくハッシュ関数 $\boldsymbol{h}^{(l)}(\boldsymbol{q})$ を構築し，ハッシュ関数群 $\{\boldsymbol{h}^{(l)}(\boldsymbol{q})\}_{l=1}^{L}$ に加える．

(2) $b^{(p)} = \text{false}$ を満たす類似ペアの特徴ベクトル $\boldsymbol{q}^{(p)}$ をハッシュ関数群 $\{\boldsymbol{h}^{(l)}(\boldsymbol{q})\}_{l=1}^{L}$ にクエリとして与える．検索結果として得られる類似ラベル群 $\mathcal{A}^{(p)}$ に正解姿勢ラベル $a^{(p)}$ が含まれていれば，$b^{(p)} = \text{true}$ と更新する．

(3) 類似ペア数を P_s，$b^{(p)} = \text{true}$ を満たす類似ペア数を P_e と定義した時に P_e/P_s の値が閾値 T に到達していればハッシュ関数の構築を打ち切る．そうでなければ 1. に戻る．

3.3 Sparse Incremental PSH

Sparse Incremental PSH は Incremental PSH を拡張したもので，ハッシュ表への登録に制約を設けるものである．Incremental PSH のハッシュ関数は独自の領域を高感度に識別する性質を持つ．そのようなハッシュ関数に対して，全ての正解姿勢候補をハッシュ表へ登録することは冗長であると言える．冗長な登録を抑制することで，検索の精度を高めるとともにメモリ容量を削減することができる．本研究の枠組みでは，ハッシュ値をキー，姿勢ラベルを値としてハッシュ表への登録を行う．そこで，姿勢ラベルの信頼度をモデル化し，その値に応じてハッシュ表への登録を行うか行わないかの判断を下す．以下にハッシュ関数構築の流れを示す．

(1) $b^{(p)} = \text{false}$ を満たすサンプルペアの集合に対して ODS 計算を行い，ODS 集合を得る．ODS 集合からランダムに K 個の decision stump を選び出すことで新しくハッシュ関数 $\boldsymbol{h}^{(l)}(\boldsymbol{q})$ を構築し，ハッシュ関数群 $\{\boldsymbol{h}^{(l)}(\boldsymbol{q})\}_{l=1}^{L}$ に加える．

(2) 姿勢ラベル数の大きさの配列 $\boldsymbol{r} \in \mathbb{R}^{N_y}$ を用意する．配列の要素 $\{\boldsymbol{r}\}_j$ には各姿勢ラベル毎の推定誤差の平均値を格納する（推定誤差として関節座標平均距離を採用する）．$\{\boldsymbol{r}\}_j$ の値が小さいほど，姿勢ラベル j の信頼度が高いと言える．

(3) $b^{(p)} = \text{false}$ を満たす類似ペアの特徴ベクトルを直前に構築したハッシュ関数 $\boldsymbol{h}^{(l)}(\boldsymbol{q})$ にクエリとして与える．検索結果として得られる類似ラベル群 $\mathcal{A}^{(p)}$ に正解姿勢ラベル $a^{(p)}$ が含まれていれば，$b^{(p)} = \text{true}$ とする．また，$\{\boldsymbol{r}\}_j$ に格納する値を**式 9** より導出する☆☆．そして，$\{\boldsymbol{r}\}_j$ の値が低いものの上位 F% をハッシュ関数 $\boldsymbol{h}^{(l)}(\boldsymbol{q})$ に対応するハッシュ表に登録する．なお，正解姿勢ラベル $a^{(p)}$ がハッシュ表に登録され，かつ $a^{(p)} \in \mathcal{A}^{(p)}$ である場合，$b^{(p)} = \text{true}$ と更新する．

$$\{\boldsymbol{r}\}_j = \frac{1}{N_{a,j}} \sum_{j \in \boldsymbol{\mathcal{A}}^{(p)}} d_x(\boldsymbol{x}^{(p)}, \boldsymbol{x}_j) \qquad (9)$$

(4) 類似ペア数を P_s，$b^{(p)} = \text{true}$ を満たす類似ペア数を P_e と定義した時に P_e/P_s の値が閾値 T に到達していればハッシュ関数の構築を打ち切る．そうでなければ 1. に戻る．

4. 高速近傍探索手法の実験による評価

4.1 性能評価実験

まず，提案する高速近傍探索手法（PSH，Incremental PSH，Sparse Incremental PSH）の性能を比較する以下の 3 つの実験を，異なる解像度で構築した正解姿勢候補群に対して行った．詳細な実験環境を後述し，**表 1** に評価結果を示す．

(1) モーションキャプチャデータセット 118050 フレームを用いて類似率，検索率，適合率，再現率，F 値の評価を行った（第 3 章で定義）．このデータセットは PSH の学習に用いたものであるが，特徴量 $\boldsymbol{q}$ を抽出する際にノイズを付与しているため，未知のデータセットとみなすことができる．

(2) 学習に用いていない未知のモーションキャプチャシーケンス 4991 フレームを人工ボクセルデータに変換し，人物姿勢推定処理に適用した．そ

☆☆ $\boldsymbol{x}^{(p)}$：$\boldsymbol{q}^{(p)}$ の元データ，$\boldsymbol{x}_j$：姿勢ラベル j に対応する関節位置ベクトル，$N_{a,j}$ ：$j \in \boldsymbol{\mathcal{A}}^{(p)}$ が満たされた回数

の時の平均推定誤差（正解姿勢と推定姿勢間の関節座標平均距離）を測定した．

(3) 実画像データから復元した 3 次元ボクセルデータ 2649 フレームを人物姿勢推定処理に適用した．その時の PSH の平均実検索率，及び人物姿勢推定フェーズの平均処理時間を測定した（Dell Precision 690, Intel Xeon Processor 5060 3.20GHz × 2, 4GB Memory で処理時間計測を行った）．

- **正解姿勢候補群**　正解姿勢候補群はモーションキャプチャデータからクラスタリングにより代表点を抽出することで構築する．この時，姿勢の方位（yaw orientation）を考慮することで姿勢候補のパターンを増やし，姿勢候補群の網羅性を高めている．姿勢の方位とは，人物の腰の位置を通る地面に垂直な軸を中心軸とした時に，腰の位置が向く方向を角度で表したものである．まず，モーションキャプチャデータの方位を基準方位となるように補正する．補正後のデータセットに対してクラスタリング処理を適用することで，基準方位におけるクラスタ群が得られる．これらを姿勢クラスタと呼ぶ．終了条件を定めればクラスタ数を規定する必要がないことから，クラスタリング手法として階層的凝集型クラスタリング[12] を採用する．なお，姿勢の距離基準に関節座標平均距離 $d_x(\boldsymbol{x}, \tilde{\boldsymbol{x}})$ を，クラスタ併合の距離基準に群平均法（group average method）を用い，最も隣接しているクラスタ間距離が $\acute{R}$ 以上になるまでクラスタ併合を続ける．群平均法においてクラスタ C_1 とクラスタ C_2 のクラスタ間距離 $D(C_1, C_2)$ は以下のように定義される（n_1, n_2：各クラスタの要素数）．

$$D(C_1, C_2) = \frac{1}{n_1 n_2} \sum_{\boldsymbol{x}_1 \in C_1, \boldsymbol{x}_2 \in C_2} d_x(\boldsymbol{x}_1, \boldsymbol{x}_2) \quad (10)$$

クラスタの代表点を Centroid ではなく Medoid とするため（実存するデータを代表点とするため），各クラスタを超球とみなせば，超球の半径値は $\acute{R}$ 以下であることが保証される（$\acute{R}$ を超球半径閾値と定義する）．従って，未知姿勢と正解姿勢候補間の最小距離が $\acute{R}$ 以下となることが期待できる．次に，方位解像度に従い姿勢クラスタをステップ角度ずつ回転させることで正解姿勢候補群を構築する．方位解像度は超球半径値に応じて設定する．具体的には，方位方向に隣接する姿勢間の距離が $2\acute{R}$ 以内となるように値を設定する．このようにすることで方位方向でも未知姿勢と正解姿勢候補間の最小距離が $\acute{R}$ 以下となることが保証される．表 1 における最小平均推定誤差は正解姿勢（未知姿勢）と正解姿勢候補間の最小距離の平均値を指す．平均推定誤差がこの値に近ければ近い程，人物姿勢推定の精度が高いと言える．なお，今回の実験では合計 161520 フレームのモーションキャプチャデータから正解姿勢候補群を抽出している．

- **PSH**　PSH のハッシュ関数を学習する際，サンプルペア群を姿勢の方位別に用意し，方位別にハッシュ関数を構築する[3]．この方針は，広大な人物姿勢空間全体を部分空間に分割することで識別性能を高めることを目的としている．本稿で提案する Incremental PSH，Sparse Incremental PSH のアイデアはその延長にあるものである．また，ハッシュ関数学習の際に用いる閾値 R は超球半径閾値 $\acute{R}$ と同一の値を用いる．ただし，PSH の性能評価時に用いる閾値 $\tilde{R}$ は共通の値 0.05[m] を使用する．なお，$\tilde{M} = 60$, $\epsilon = 0.25$, $T = 0.95$, $F = 75$ として実験を行った．
- **特徴量**　特徴量抽出関数 $Q(\boldsymbol{v})$ により，Cylidrical Histogram Feature を抽出する[3]．まず，ボクセルデータの重心位置に設置する z 軸方向の中心軸を中心として空間を回転・高さ・半径方向にビンとしてメッシュ状に分割する．そして，各ボクセルを対応するビンに割り当て，正規化により特徴ベクトルを得る．実装の都合上，回転方向解像度は方位解像度と同じ値にしなければならないが[3]，高さ・半径方向の解像度は共通の値 12, 2 と設定した．なお，類似度導出関数 $S(\boldsymbol{q}, \tilde{\boldsymbol{q}})$ として Bhattacharyya 係数[13] を採用している．

表 1 によると，正解姿勢候補群の解像度を高めるにつれて，PSH の検索性能（F 値）が向上し，人物姿勢推定における平均推定誤差の値が小さくなっていることが分かる．一方で，解像度の向上に伴い，処理時間が長くなる傾向にある．また，PSH の手法に着目すると，Sparse Incremental PSH の検索性能が最も高く，姿勢推定に要する時間も最も短いという結果が出ている．それにもかかわらず，姿勢推定における平均推定誤差の値が PSH の値とほぼ変わらない点は，姿勢推定の精度を維持した状態で高速化に成功していると評価できる．最終的には，推定の精度と計算コストはトレードオフの関係にあり，目標の精度でオンライ

ン処理を可能にするために，適切な条件のもとでシステムを構築する必要がある．

姿勢の類似度の評価として関節座標平均距離を導入している研究[9),14)] の中で，様々な条件の実験における平均推定誤差の最小値は約 0.05[m] であった．超球半径閾値 $\acute{R}$ はこの値を元に設定しており，本研究でも平均推定誤差がこの値を下回ることが一つの目標となる（閾値 $\tilde{R}$ もこの値を基準に設定している）．また，クライアント側の処理は大きく分けて，データ送受信，視体積交差法における視体積統合，人物姿勢推定，描画に分けられるが（図 2 参照），データ送受信・視体積統合・描画の平均処理時間はそれぞれ約 5[ms]，5[ms]，2[ms] である．30[FPS] のオンライン処理を実現するためには合計処理時間が 33[ms] 以内でなければならない．従って，ある程度余裕を持たせるとすれば，人物姿勢推定のフェーズは 20[ms] 以内で完了しなければならない．表 1 で上記の 2 つの条件を満たすのは姿勢ラベル数 172602 に対して Incremental PSH，或いは Sparse Incremental PSH を適用した場合である．このように，精度の高い人物姿勢推定システムをオンラインで走らせる上で，提案する高速近傍探索手法が有効であると言える．

4.2 安定性評価実験

高速近傍探索手法では多様なデータが入力として与えられた場合も，検索率が一定の範囲内に納まることが期待される．なぜなら，検索率の安定化は処理速度の安定化を意味し，結果的にシステムの安定化につながるからである．実画像データから復元した 3 次元ボクセルデータ 2649 フレームを人物姿勢推定処理に適用した時の検索率の変遷を手法別にグラフ化した図を**図 3** に示す（超球半径値 0.04 の条件で構築した姿勢ラベル数 172602 のデータを正解姿勢候補群として使用した）．図を見てみると，一部検索率が急上昇するフレームがあるが，Sparse Incremental PSH では PSH と比較して検索率変動の割合が軽減されていることが分かる．従って，Sparse Incremental PSH の検索率が最も低くかつ安定していると言える．

4.3 正解姿勢候補群の高解像度化による効果

最後に，正解姿勢候補群の解像度を高めることによって得られる効果を視覚的に確認する実験を行った．異なる解像度の正解姿勢候補群を用いて，歩行動作に対して人物姿勢推定処理を適用した様子を**図 4** に示す（Sparse Incremental PSH を採用した）．この図から，正解姿勢候補群の解像度が高い程，推定姿勢が正解姿勢に近く，全体的に滑らかなシーケンスが生成されていることが確認できる．

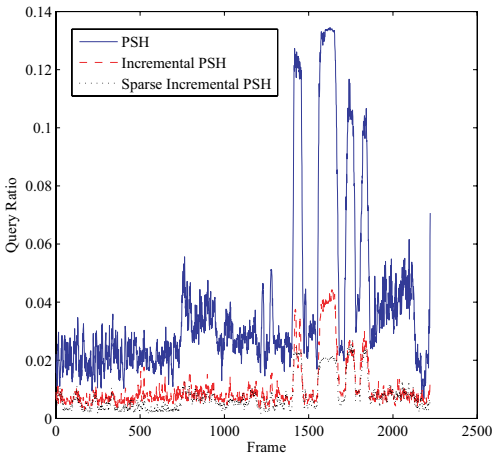

図 3 Query Ratio Transition

5. 結　　論

本稿では，高精度な事例ベース人物姿勢推定を最大 30[FPS] のオンライン処理で実現するための高速近傍探索手法を提案した．提案手法は従来手法と比較して検索性能で大きく上回る上に，多様なデータに対する検索率の安定性も得ることができる．人物姿勢推定システムの将来課題として，オクルージョンへの対応，複数人への対応が挙げられる．

参 考 文 献

1) Kidd, C.D., Orr, R., Abowd, G.D., Atkeson, C.G., Essa, I.A., MacIntyre, B., Mynatt, E., Starner, T.E. and Newstetter, W.: The Aware Home: A Living Laboratory for Ubiquitous Computing Research, *Proc. CoBuild*, pp.191–198 (1999).

2) Mori, T., Noguchi, H., Takada, A. and Sato, T.: Sensing Room: Distributed Sensor Environment for Measurement of Human Daily Behavior, *Proc. INSS*, pp. 40–43 (2004).

3) Sagawa, Y., Shimosaka, M., Mori, T. and Sato, T.: Fast Online Human Pose Estimation via 3D Voxel Data, *To Appear in the Proceedings of the 2007 International Conference on Intelligent Robots and Systems* (2007).

4) 亀田能成，美濃導彦，池田克夫：シルエット画像からの関節物体の姿勢推定法，電子情報通信学会論文誌 (D)， No.Vol. J79-D, No. 1, pp.26–35 (1996).

5) Yamamoto, M., Sato, A., Kawada, S., Kondo, T. and Osaki, Y.: Incremental Tracking of Human Actions from Multiple Views, *Proc. CVPR*, pp.2–7 (1998).

6) Mori, G. and Malik, J.: Estimating Human Body Configurations Using Shape Context Matching, *Proc. ECCV*, pp.666–680 (2002).

7) Sminchisescu, C., Kanaujia, A., Li, Z. and Metaxas,

表 1　Evaluation of the PSH Metrics ((1):PSH, (2):Incremental PSH, (3):Sparse Incremental PSH)

PSH	**(1)**	**(2)**	**(3)**	**(1)**	**(2)**	**(3)**	**(1)**	**(2)**	**(3)**
超球半径閾値 [m]	0.08	0.08	0.08	0.06	0.06	0.06	0.04	0.04	0.04
姿勢クラスタ数	956	956	956	2830	2830	2830	9589	9589	9589
方位解像度	10	10	10	12	12	12	18	18	18
姿勢ラベル数	**9560**	**9560**	**9560**	**33960**	**33960**	**33960**	**172602**	**172602**	**172602**
特徴ベクトル次元	240	240	240	288	288	288	432	432	432
K	12	12	12	12	12	12	12	12	12
L	430	140	120	564	132	192	558	216	450
類似率	7.06E-4	7.06E-4	7.06E-4	6.58E-4	6.58E-4	6.58E-4	6.66E-4	6.66E-4	6.66E-4
検索率	0.138	0.0806	0.0725	0.103	0.0454	0.0371	0.0866	0.0256	0.0230
再現率	0.995	0.991	0.986	0.980	0.974	0.967	0.900	0.876	0.863
適合率	0.00508	0.00868	0.00959	0.00623	0.0141	0.0172	0.00693	0.00228	0.0250
F 値	**0.0101**	**0.0172**	**0.0190**	**0.0124**	**0.0278**	**0.0337**	**0.0138**	**0.0445**	**0.0486**
平均推定誤差 [m]	**0.0578**	**0.0622**	**0.0585**	**0.0544**	**0.0546**	**0.0548**	**0.0485**	**0.0490**	**0.0489**
最小平均推定誤差 [m]	0.0534	0.0534	0.0534	0.0489	0.0489	0.0489	0.0423	0.0423	0.0423
平均実検索率	0.0600	0.0164	0.0107	0.0504	0.00600	0.00353	0.0382	0.0103	0.00747
平均処理時間 [ms]	**2.7**	**0.9**	**0.7**	**8.5**	**1.3**	**1.0**	**43.0**	**12.4**	**9.6**

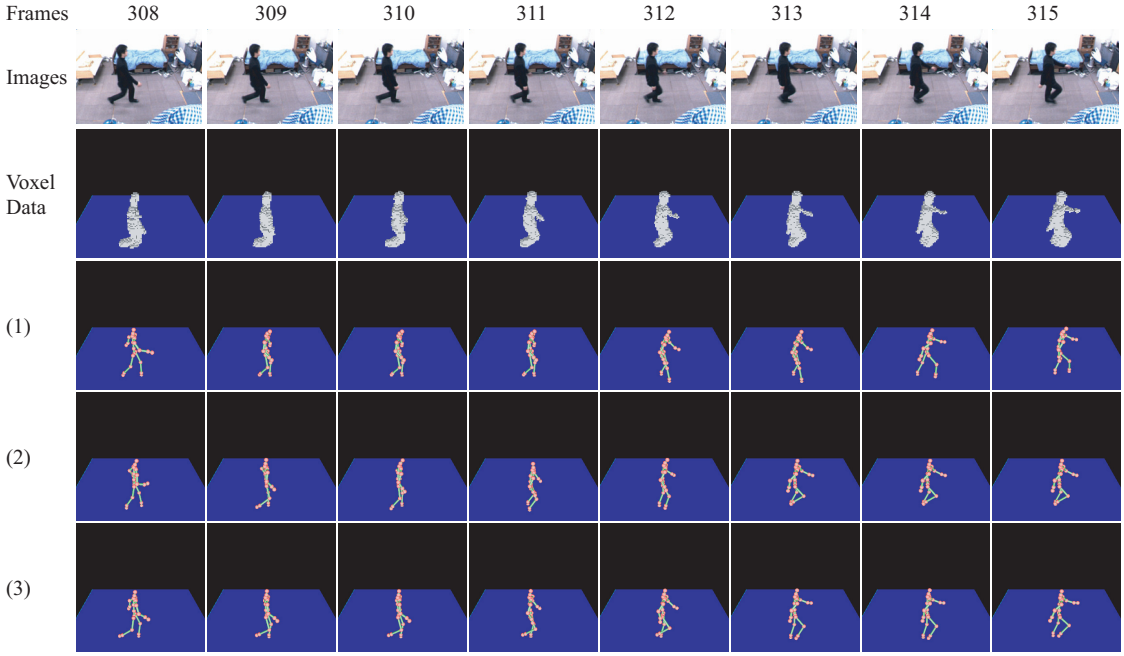

図 4　Sequential Animation Shots in Different Posture Candidate Resolutions ((1):9560 candidates, (2):33960 candidates, (3):172602 candidates)

D.: Discriminative Density Propagation for 3D Human Motion Estimation, *Proc. CVPR*, Vol.1, pp.390–397 (2005).

8) Shakhnarovich, G., Viola, P. and Darrell, T.: Fast Pose Estimation with Parameter Sensitive Hashing, *Proc. ICCV*, Vol.2, pp.750–757 (2003).

9) Taycher, L., Shakhnarovich, G., Demirdjian, D. and Darrell, T.: Conditional Random People: Tracking Humans with CRFs and Grid Filters, *Proc. CVPR*, Vol.1, pp.222–229 (2006).

10) Ren, L., Shakhnarovich, G., Hodgins, J.K., Pfister, H. and Viola, P.: Learning Silhouette Features for Control of Human Motion, *ACM Transactions on Graphics*, Vol.24, No.4, pp.1303–1331 (2005).

11) Gionis, A., Indyk, P. and Motwani, R.: Similarity Search in High Dimensions via Hashing, *Proc. VLDB*, pp.518–529 (1999).

12) MacKay, D. J. C.: *Information Theory, Inference, and Learning Algorithms*, Cambridge University Press (2005).

13) Kailath, T.: The Divergence and Bhattacharyya Distance Measures in Signal Selection, *IEEE Trans. on Comm. Technology*, Vol.15, pp.52–60 (1967).

14) Balan, A.O., Sigal, L. and Black, M.J.: A Quantitative Evaluation of Video-based 3D Person Tracking, *IEEE International Workshop on Visual Surveillance and Performance Evaluation of Tracking and Surveillance*, pp.349–356 (2005).