

行動認識のための角度変位量子化特徴の検討

非会員 山辺 智晃* 非会員 片岡 裕雄** 正員 中村 明生*a)

Quantized Feature with Angular Displacement for Activity Recognition

Tomoaki Yamabe*, Non-member, Hirokatsu Kataoka**, Non-member, Akio Nakamura*a), Member

(2014年7月21日受付, 2014年11月28日再受付)

We propose quantized feature with angular displacement for pose-based activity recognition. We calculate a 3D joint angle from three postural coordinates. The angular displacement should be quantized since joint angle includes errors due to system noises and similar posture. To investigate appropriate features, we propose four kinds of quantization levels; binarization, ternarization, quaternarization, and quinarization. We apply quantized features in order to improve pose-based activity recognition with the UTKinect-Action Dataset. In the experiment, we show the appropriate feature for activity recognition. As the result, the ternarized feature achieves the highest recognition rate in average. The recognition rate of trials with ternarized feature is improved 2.4% to one with no-quantized feature, and 1.8% to conventional method.

キーワード: 行動認識, 姿勢ベース特徴, 角度量子化特徴, ランダムフォレスト

Keywords: Activity Recognition, Pose-based Feature, Quantized Feature with Angular Displacement, Random Forests

1. はじめに

コンピュータビジョンの分野では、画像補正や3次元構成、特徴量取得や識別問題などが実験・考察されてきた⁽¹⁾。その中でも、人物行動解析技術は追跡や姿勢推定、顔認識の問題等数多くの分野に細分化され、生活空間内での監視や見守り等、場面・状況に応じた多様なアプリケーションにおいて必要とされている^{(2)~(4)}。

人物行動認識の研究は、画像の局所特徴量を適用したアピアランスベースの行動認識と、3次元姿勢情報を用いる姿勢ベースの行動認識に大きく分類できる。アピアランスベースの行動認識では、HOG (Histograms of Oriented Gradients)⁽⁵⁾やHOF (Histograms of Optical Flow)⁽⁶⁾、3DSIFT (Scale-Invariant Feature Transform)⁽⁷⁾などの画像特徴量を用いた認識が行われている。一方、姿勢ベースの行動認識では可視画像だけ

でなく、TOF (Time of Flight) センサや Microsoft Kinect 等の深度センサで得られた深度情報を利用した姿勢推定を伴うアプローチがある。それぞれの詳細については(2・1)節で述べる。

本稿では姿勢ベースの行動認識に取り組み、そのための新たな特徴表現方法として、角度変位の量子化特徴量を提案する。量子化により、動きの変化量を簡易かつ明確に表現可能である。また、変化を符号化することになり、個人差・微小な動作変化にもロバストになると考える。そのため、角度変位をそのまま用いる場合に比べ、少ない処理コストで精度良い行動認識を実現可能であると考えられる。ところが、角度変位量子化にあたり、何値化すべきかという量子化レベル、および適切な閾値は不明である。著者らはこれまでに位置変位と角度変位の比較、および2値化と3値化の比較については一部検証を行った。角度変位が有効であることは分かったが、4値化以上の量子化特徴に対する考察および有効性の評価については行っていない⁽¹⁹⁾。そこで、本稿では角度変位に対して2値化から5値化までを行い、それらの特徴として日常生活行動を識別する。また、複数フレームの特徴を統合することで時系列変化に対応するが、その統合数などのパラメータも調査する。まとめると、行動認識に有効となる量子化レベル、閾値、およびパラメータについて調査を行う。

a) Correspondence to: Akio Nakamura. E-mail: nakamura@is.fr.dendai.ac.jp

* 東京電機大学
〒120-8551 東京都足立区千住旭町5
Tokyo Denki University

** 東京大学
〒113-8656 東京都文京区本郷7-3-1
The University of Tokyo
7-3-1, Hongo, Bunkyo-ku, Tokyo 113-8656, Japan

2章では先行研究調査および提案する特徴・手法について概説する。3章では、角度変位置量子化特徴の抽出に説明する。4章では、時系列データセットおよび行動認識に使用する機械学習手法について述べる。5章で行動認識実験を行い、提案した量子化特徴の有効性に関して考究する。6章で本稿をまとめる。

2. 提案する量子化特徴の概要

〈2・1〉節では1章で触れたアピランススペースの行動認識および姿勢ベースの行動認識の先行研究について述べる。〈2・2〉節では検証に用いるデータセット、〈2・3〉節では量子化特徴量について概説する。

〈2・1〉 行動認識に関する先行研究 以下、アピランススペースの行動認識および姿勢ベースの行動認識に関する先行研究についてそれぞれ述べる。

アピランススペースの行動認識に関して、画像特徴量を適用した行動認識手法として、Dalalら⁽⁵⁾は、HOG特徴量を提案している。HOG特徴量はエッジ勾配を量子化したヒストグラムにエッジ強度を累積することにより物体の形状を表現している。画像中の人物における外観と形状に対する正確な対応が未知であっても、局所的な輝度勾配はエッジとの関係性が高い。そのため、HOG特徴量は人物の姿勢を認識する場合に有効な特徴量の1つとして様々なアプローチで研究されている^{(8)~(10)}。Dalalらは2クラス識別器であるSVM (Support Vector Machine)⁽¹¹⁾の適用により高精度な歩行者検出を実現した。また、深度センサを用いたアピランススペースの研究に関しては、Omarらの研究⁽¹²⁾が挙げられる。Omarらは行動認識に用いる特徴について、深度マップを利用した新たな特徴量を提案している。センサから取得した深度情報に対して、時系列を考慮した4次元空間における人物の表面法線を算出し、その分布からヒストグラムを用いて奥行き情報を計算したものを行動識別の特徴として採用している。特徴の評価ではMSR Actions 3D⁽¹³⁾、MSR Gesture 3D⁽¹⁴⁾、MSR Daily Activity 3D⁽¹⁵⁾など多数のデータセットで検証実験をしており、高い識別性能を達成した。しかし、アピランススペースの手法は処理コストが高く、特徴ベクトルの次元数が膨大になる問題がある。例えばHOGでは数千次元の特徴ベクトルを必要とし、行動認識のように時系列連結をする場合にはさらに次元数、処理時間ともに増大する。さらに、アピランススペースの手法では視点の変更により見え方が変化した結果、特徴のパターンひいては特徴ベクトルが変化することがある。そのため、ある環境で生成した識別器を他の環境で用いる際には識別性能が低下する。

姿勢ベースの行動認識では、人体の関節や骨格の位置情報を特徴量として用いた研究が数多くなされている。人物の行動は、身体部位の動きと密に関係しているため、身体のような部位の動きを抽出し、その関係性をを用いることで行動認識が可能である。身体の中でも関節は重要な要素である。関節情報を適用した行動認識手法としては、Wangら

の手法が挙げられる⁽¹⁵⁾。Wangらは、深度データから推定した3次元の関節位置に基づいてLOP (Local Occupancy Pattern) という、行動認識のための新たな動作特徴量を提案している。LOPは関節位置のフレーム間の移動量と特定関節周りの3次元点群の分布を特徴としており、類似した動作や時間的なズレによるノイズに対して頑健な特徴となっている。実験では、多くの先行研究の様々な手法に対して有効性を示している。観測した関節位置情報を他の要素と組み合わせることによって、行動認識に有効な特徴量が作成可能であることが示されている。しかし、提案された特徴は複数の要素を掛け合わせているため、複雑でデータ量の大きい特徴となっており、特徴量や身体部位が識別に及ぼす影響を把握することが難しい。また、Xiaら⁽¹⁶⁾は3次元の時系列関節角度変位をBoF (Bag-of-Features)⁽¹⁷⁾のヒストグラムを用いて表現している。身体の関節位置を球形座標に投影し、視点の変化に頑健な特徴表現としてLDA (Latent Dirichlet Allocation)により次元数を削減したHOJ3D (Histograms of 3D Joint Locations)ベクトル表現を提案している。HOJ3DベクトルからK-means法によるクラスタリングを行い、各姿勢の行動タグを設定することにより、各動作を連続した行動タグの並びで表現可能としている。識別には離散HMM (Hidden Markov Model)を用いており、Xiaらが提案している10クラスの行動を含むデータセットにおいて、平均90.9%の識別率を実現している。また、Shottonら⁽¹⁸⁾はKinectを用いた手法でリアルタイムでの姿勢推定に成功しているなど、後者は高速な姿勢取得が可能であることも分かっている。しかし、Xiaら⁽¹⁶⁾の手法ではBoFにサンプリングする特徴が少ないためSparseな特徴に陥りがちであり、特徴の表現能力の点で不十分である。また、角度変位をそのまま使用しているため、同じ行動でも試行する個人間の微小な差異に鋭敏であり、行動認識に影響を与えやすいと考える。

以上、アピランススペースの行動認識および姿勢ベースの行動認識について概観したが、後者は人体の構造的な要素を用いるため、前者に比較してスケールや視点の変化による影響が少なく、腕の曲げ伸ばしなどの身体的な状態を認識し易い利点があると言える。

〈2・2〉 データセットおよび角度変位 本稿では身体情報の取得にMicrosoft Kinectを利用する。Kinectから得られるスケルトンデータはShottonら⁽¹⁸⁾らの姿勢推定法に基づく。視点変化に頑健とはいえ、推定に基づくために誤差を含み、オクルージョン発生時には影響が考えられるが、アピランススペースの行動認識に比較すれば軽度であると考える。行動認識を通じた特徴評価のため、先行手法との比較が容易となるよう、汎用的なデータセットとして、Xiaらが配布するUTKinect-Action dataset⁽¹⁶⁾を使用する。UTKinect-Action datasetは、単一の固定KinectとWindows SDKを用いて、屋内環境における10種類の日常的な全身動作を取得・整理したデータセットである。深度画像解像度は320×240 pixel、フレームレートは30フレーム毎秒である。被験者は9人の男性と1人の女性から構成され、10種類の行動を各人

各2回行っている。10種類の行動の内容は“Carry (運ぶ)”, “Clap hands (拍手する)”, “Pick up (拾う)”, “Pull (引く)”, “Push (押す)”, “Sit down (座る)”, “Stand up (立つ)”, “Throw (投げる)”, “Walk (歩行)”, “Wave (手を振る)”, である。すなわち, 全体で200個の行動を含む6220フレームのデータとなっており, 各行動の平均データ長は40~250フレームである。日常生活の行動をほぼ網羅しているとは言えないが, このデータセット内の行動識別ができれば提案手法の有効性の基礎的な確認はできると考える。

姿勢推定により得られた3次元姿勢情報の例をFig. 1に示す。20箇所(20C3通り)の身体部位の位置情報を取得可能であり, これらの3つを組み合わせることで, $n=20C_3$ 通りの角度を計算できる。この角度のフレーム間の変化量を角度変位と呼ぶ。

〈2・3〉 角度変位置量子化特徴 量子化の基本は2値化である。例えば, 角度変位が一定量以上の場合は1(動作あり), それ以下は0(動作なし)という符号化が考えられる。情報量としては減るが, 組み合わせて身体全体の高次元データとすれば効率的な学習ができる可能性がある。しかし, 2値化は2状態を表現するだけであり, 果たして行動識別に十分か否かは判然としない。とはいえ, 量子化レベルを大きくすると元データに近づくだけで量子化するメリットはない。

また, 識別対象は, 人物の行動であり, その時系列変化もデータに含めることが重要となる。量子化特徴は角度の変位に応じて符号化しているが, これは変化量のみ情報となるため, 値がどのように変化してきたか, という時系列変化の情報は含まれていない。そこで, 角度変位置量子化特徴を複数フレームで統合し, 複数フレームの連続的な値の変化を1つの観測データとして用いる。その時の統合数などのパラメータも決定する必要がある。

以上まとめると, 3章で詳述するように, 角度変位に対して2値化, 3値化, 4値化, 5値化を施し, さらに4章で説明する複数フレーム統合を行い, それらの特徴として日常生活行動を識別することで, 行動認識に有効となる量子化レベル, 閾値, およびパラメータについて調査を行う。角度変位をそのまま特徴とした場合と量子化した場合を, 交差検定法(2-fold-cross-validation)で比較する。

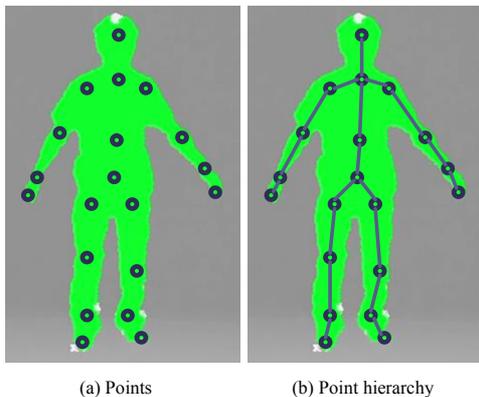


Fig. 1. Data captured by Kinect

また, 特徴比較を主目的とする本稿では, 識別に用いる機械学習手法として, 高次元特徴に対応し, 寄与の確認も容易で, 優れた手法として定評のあるRandom Forests⁽²⁰⁾を使用する。これに関しては4章で説明する。

3. 角度変位置量子化特徴の抽出

角度変位置量子化特徴の抽出方法を具体的に説明する。(2・2)節で述べた通り, $n=20C_3$ 通りの角度のフレーム間変化量が角度変位であり, n 次元の列ベクトルとして特徴が定義される。特徴ベクトルの要素を閾値 T で処理することで量子化を行う。角度変位ベクトルの要素は, 現在の角度ベクトル $\theta(t) = (\theta_1, \dots, \theta_i, \dots, \theta_n)^T$ と f フレーム過去の角度ベクトル $\theta(t-f)$ との差分として(1)式のように求められる(Fig. 2参照)。 f は比較フレーム間隔と呼ぶ。

$$\Delta\theta_i(t) = \theta_i(t) - \theta_i(t-f) \quad \dots\dots\dots(1)$$

以下, この角度変位を量子化する。また, T, T_{high}, T_{low} は閾値であり, $T_{high} > 0, T_{low} = -T_{high}$ とする。

〈3・1〉 2値化特徴 閾値 T 以上なら1(動作有り), 閾値以下なら0(動作なし)とし, 特徴ベクトル $\mathbf{F}_V = (V_1, \dots, V_n)^T$ の要素は(2)式の通りである(Fig. 3)。

$$V_i = \begin{cases} 1 & (\text{if } |\Delta\theta_i(t)| > T) \\ 0 & (\text{otherwise}) \end{cases} \quad \dots\dots\dots(2)$$

〈3・2〉 3値化特徴 伸ばし方向と曲げ方向の2種類の閾値を設定し, 0(動作無し), 1(伸長方向へ変化), -1(屈曲方向へ変化)の3値に符号化する。特徴ベクトル \mathbf{F}_V の要素は(3)式の通りである(Fig. 4)。方向を含む変化状態および変化がない中間的な状態が表現可能である。

$$V_i = \begin{cases} 1 & (\text{if } \Delta\theta_i(t) > T_{high}) \\ 0 & (\text{else if } T_{high} \geq \Delta\theta_i(t) \geq T_{low}) \\ -1 & (\text{elseif } \Delta\theta_i(t) < T_{low}) \end{cases} \quad \dots\dots\dots(3)$$

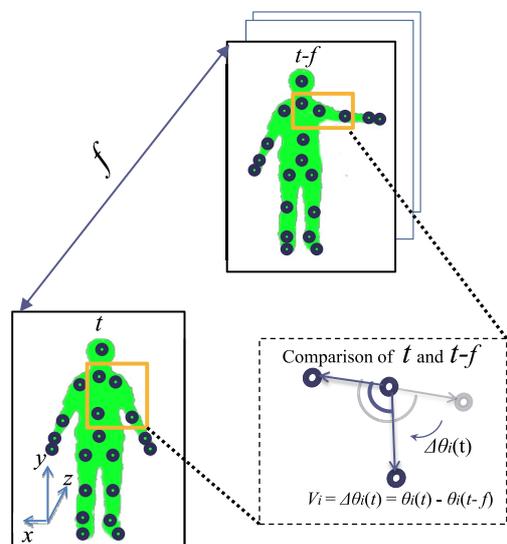


Fig. 2. Angular displacement

〈3・3〉 4 値化特徴 0 を基準とするほか, 2 種類の閾値を設け, 角度変位を 2 段階で表現する。2 (伸長方向へ大きく変化), 1 (伸長方向へ小さく変化), -1 (屈曲方向へ小さく変化), -2 (屈曲方向へ大きく変化) の 4 値に符号化する。特徴ベクトル \mathbf{F}_V の要素は(4)式の通りである (Fig. 5)。

大きく変化), -2 (屈曲方向へ大きく変化) の 4 値に符号化する。特徴ベクトル \mathbf{F}_V の要素は(4)式の通りである (Fig. 5)。

$$V_i = \begin{cases} 2 & (\text{if } \Delta\theta_i(t) > T_{high}) \\ 1 & (\text{else if } T_{high} \geq \Delta\theta_i(t) > 0) \\ -1 & (\text{else if } 0 < \Delta\theta_i(t) \leq T_{low}) \\ -2 & (\text{else if } \Delta\theta_i(t) < T_{low}) \end{cases} \dots\dots\dots (4)$$

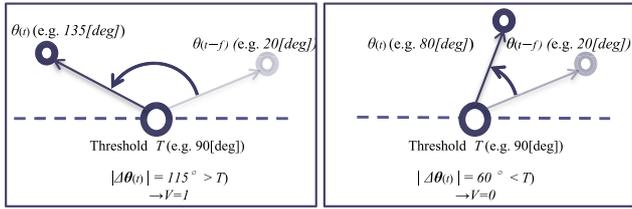


Fig. 3. Binarization

〈3・4〉 5 値化特徴 4 種類の閾値を設け, 方向を含む角度変位 2 段階状態および中間の状態が表現できる。2 (伸長方向へ大きい変化), 1 (伸長方向へ小さく変化), 0 (動作無し), -1 (屈曲方向へ小さく変化), -2 (屈曲方向へ大きく変化) の 5 値に符号化する。特徴ベクトル \mathbf{F}_V の要素は(5)式の通りである (Fig. 6 参照)。

$$V_i = \begin{cases} 2 & (\text{if } \Delta\theta_i(t) > T_{high}) \\ 1 & (\text{else if } T_{high} \geq \Delta\theta_i(t) > T_{high}/2) \\ 0 & (\text{else if } T_{high}/2 \geq \Delta\theta_i(t) \geq T_{low}/2) \\ -1 & (\text{else if } T_{low}/2 \geq \Delta\theta_i(t) > T_{low}) \\ -2 & (\text{else if } \Delta\theta_i(t) < T_{low}) \end{cases} \dots\dots\dots (5)$$

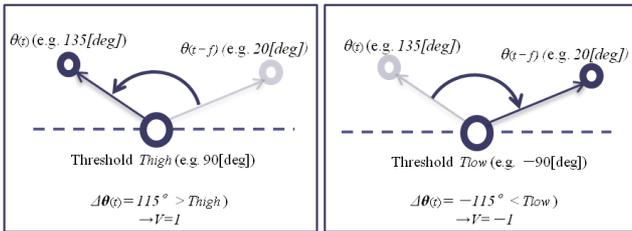


Fig. 4. Ternarization

検討する量子化特徴の中では最も量子化レベルが大きく, より細かい状態を表現可能と考える。

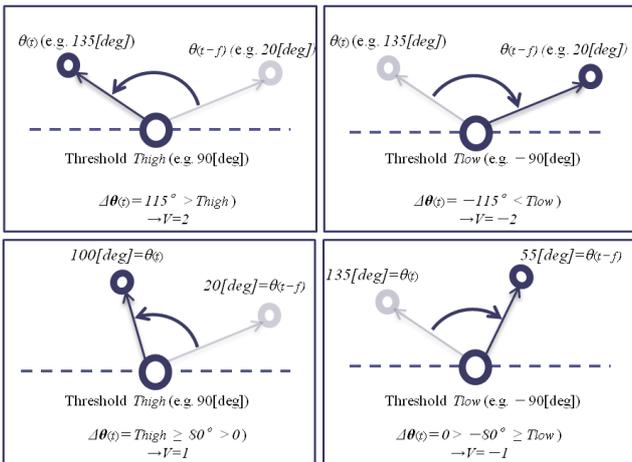


Fig. 5. Quaternarization

4. 角度変位置量子化特徴を用いた時系列データセットの作成および Random Forests による複数クラス分類

3 章で説明した角度変位置量子化特徴を複数フレームに渡って統合し, 時系列変化に対応する。また, 識別に用いる機械学習手法として採用した Random Forests について触れる。

〈4・1〉 時系列データセットの作成 〈2・3〉 節で述べた通り, 行動識別では動作の時系列変化情報もデータに含めることが重要となる。量子化に用いる角度変位情報は, 過去フレームとの角度の変化量のみ情報であるため, 値がどのように変化してきたか, という時系列変化の情報は含まれない。そこで, 量子化角度変位を複数フレームで統合し, 複数フレームの連続的な値の変化を 1 つの観測データとして用いる (Fig. 7)。

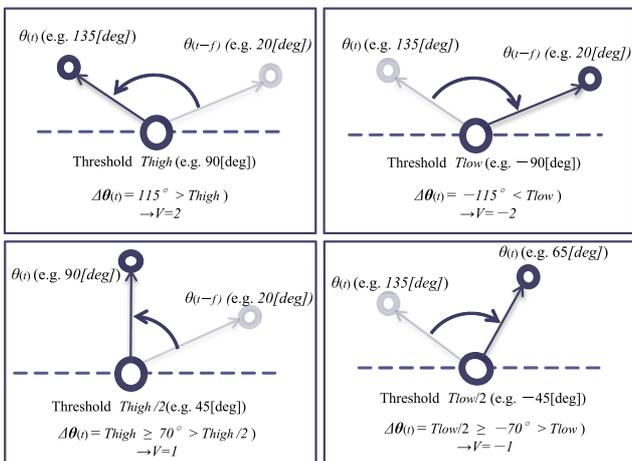


Fig. 6. Quinarization

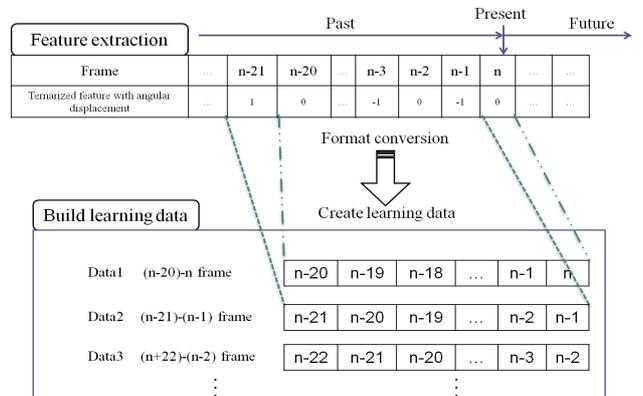


Fig. 7. Combination of the features with angular displacement during n-frames

Table 1. Parameters for quantization

Parameters	Value
Threshold t	10, 20, 60, 120 [deg]
Temporal number of frames N	10-30 [flame] (10 flame each)
Frame comparison interval f	10-30 [flame] (5 flame each)

行動認識に最適な量子化レベル, 閾値, およびパラメータについて調査を行う。UTKinect-Action dataset 内の行動データを対象として, 角度変位置量子化特徴を求め, Random Forests による識別を行い, 交差検定法 (2-fold-cross-validation) で比較する。

〈5・1〉 角度変位置量子化特徴の評価 量子化の際のパラメータを変化させて識別率への影響を調査した。変化したパラメータは角度変位閾値 T, T_{high}, T_{low} (3章), 比較フレーム間隔 f (3章), 時系列フレーム数 N (4章) である。

使用したデータセットの角度変位は, Wave (手を振る) 等の変化が激しい動作ではフレーム辺り最大 100~130[deg], Pull (引く) 等の変化が少ない動作ではフレーム辺り最大 10~20[deg] であった。それを踏まえ, 角度変位閾値 T, T_{high} ($=-T_{low}$) は 10, 20, 60, 120 の値をとることとした。これは, はじめに可動域を 3 分割した閾値 60[deg], 120[deg], 180[deg] を設定し, 識別率が高いものに対してさらに 10[deg] 刻みで適する値を調査した結果である。

また, 4章で述べた通り, 今回使用する UTKinect-Action dataset では, Pull (引く) 動作が最短で 30 フレームであるため, 時系列フレーム数 N は 10 から 30 フレームの間を 10 フレーム間隔で変化させた。比較フレーム間隔 f は 10 フレームから 30 フレームの間を 5 フレーム間隔で変化させた。実験時のパラメータ設定を Table 1 に示す。

特徴の評価には, 評価用に用いた 10 種類の行動データ “Carry (運ぶ)”, “Clap hands (拍手する)”, “Pick up (拾う)”, “Pull (引く)”, “Push (押す)”, “Sit down (座る)”, “Stand up (立つ)”, “Throw (投げる)”, “Walk (歩行)”, “Wave (手を振る)” のうち, 正しいクラスに分類されたフレーム数から識別率を算出している。例えば, Pull (引く) 動作の場合, 30 フレームのうち 27 フレームで正しく識別されれば 90% と計算される。

実験の結果, 最も高い識別率となった角度閾値 20[deg] の場合における, それぞれの量子化特徴の平均識別率を Fig. 9 に示す。なお, 比較フレーム間隔 f は 20 フレーム以後, 識別率が著しく低下したため Fig. 9 では割愛する。また, 今回の調査の結果, 最も高い識別率が得られた 3 値化特徴における量子化の際のパラメータ, 閾値 $T=20$ [deg], 時系列フレーム数 $N=20$, 比較フレーム間隔 $f=15$ とした場合の各行動の識別結果を Fig. 10 に示す。Fig. 9 より, 時系列フレーム数 N が 10~20 フレームの間における識別率の変化は少ないが, 30 フレームになると精度が大きく低下する。

これは, 一つの行動の持続時間に対して, 30 フレームは過剰であり, 他の行動に遷移して他クラスに識別されるべ

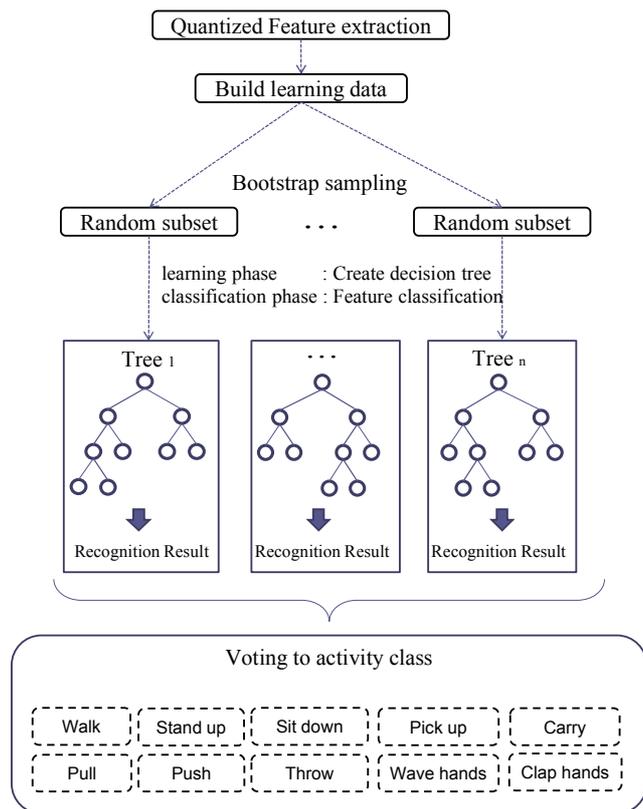


Fig. 8. Random Forests

度変位特徴を N フレーム統合すると, 全体で Nn 次元の列ベクトルが特徴となる。Fig. 7 に示す通り, これを 1 フレーム毎にずらして毎フレームの量子化データとして構築する。ここで, N を時系列フレーム長と呼ぶが, 行動によって行動データのフレーム長は変化するため, その適切な値が不明である。調査には日常生活行動として UTKinect-Action dataset を用いており, 含まれる行動は Pull や Throw が最短で 30~50 フレーム, Carry, Wave が最長で 200~250 フレームとなり, 行動毎の動作フレーム長は様々である。時系列フレーム長については, 5章においてパラメータの一つとして実験的に定める。

〈4・2〉 Random Forests による複数クラス分類 本稿で扱う問題は 10 種類のクラス分類であり, 識別モデルによる学習判別手法を用いる。高次元特徴に対応し, 寄与の確認も容易で, 優れた手法として定評のある Random Forests⁽²⁰⁾ を使用する。Random Forests は, 複数の決定木構造をもったマルチクラス識別器を構成するアンサンブル学習アルゴリズムである。ブートストラップサンプルに対してそれぞれ独立した決定木を構築し, 各決定木から出力された結果を統合することで識別を行う (Fig. 8)。

5. 行動認識を通じた特徴評価実験

量子化することにより個人毎の差や無意識な動作, 癖などによる微小な観測データの変動に強くなると仮定し, 角度変位に対して 2 値化から 5 値化までの量子化を行う。量子化した特徴を用いて行動の識別率を評価する。その過程で,

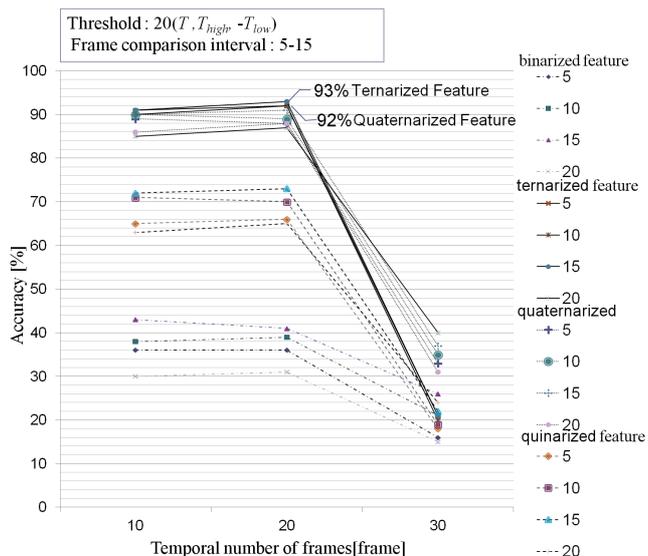


Fig. 9. Temporal number of frames N vs. accuracy

き行動を含めた特徴として抽出されたためと考える。比較フレーム間隔 f のパラメータは、今回調査した中では f が 15 フレームの時に最も高い識別率が得られ、20 フレーム以上では精度が低下する結果となった。Walk や Wave などの動作は、その動きから分かるように一定周期で特徴が変化する。比較フレーム間隔は、この動作の変化を抽出するためパラメータである。UTKinect-Action dataset に含まれている行動を調査した結果、約 15 フレーム周期で変化が大きくなる動作が多いことが確認された。そのため、比較フレーム間隔のパラメータが 15 フレームの時、最も高い識別率になったと思われる。

Fig. 10 は各行動クラスの識別結果となっており、どの行動クラスにどれだけ観測データが分類されたかを表している。Throw や Pick up などは角度変位が複雑な行動データとなっているため、全体を通して誤認識が多い行動クラスとなっている。ここで、角度変位が複雑とは分散平均が大きいものとする。Table 2 は行動毎に、“右肘”、“左肘”、“右膝”、“左膝”の 4 関節においてそれぞれ 100 フレーム分の角度変位の分散を求め、平均をとったものである。Pickup, Throw は分散平均の値が大きく、複雑であり、誤認識につながっていると考えられる。Wave も分散平均の値は大きい、周期的な動作であるため、周期毎の動きは単純となり識別率が低下しなかったと考えられる。また、Walk や Carry 等の全身の動きが大きい、または周期的に変動する動作においては、識別に有効な情報量が増えるため、誤認識が少なく高精度な識別が可能であることが伺える。結果に対する詳細な考察は (5.4) 節で述べる。

(5.2) 先行手法との比較 特徴量子化において、角度変位をいくつに分割するかを密に検討した。評価には UTKinect-Action dataset のデータを用いて実験を行ったため、行動識別における先行研究を行っている Xia らの提案する手法と比較している。別結果を比較した結果、Throw の識別で大きく精度が向上する結果となった (Table 3)。Throw

		Input Actions									
		carry	claphands	pickup	pull	push	sitdown	standup	throw	walk	wave
Classified Actions	carry	99.4%	0.0%	2.7%	2.0%	0.0%	0.0%	0.0%	0.9%	0.0%	0.0%
	claphands	0.4%	99.6%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
	pickup	0.0%	0.0%	92.7%	0.0%	0.0%	2.9%	3.8%	11.3%	0.0%	0.0%
	pull	0.0%	0.0%	0.0%	78.8%	2.3%	0.0%	0.0%	0.0%	0.0%	0.0%
	push	0.0%	0.0%	0.0%	8.6%	84.7%	0.0%	0.0%	5.2%	0.0%	0.2%
	sitdown	0.0%	0.4%	1.1%	0.0%	0.8%	95.2%	0.0%	0.0%	0.0%	0.0%
	standup	0.2%	0.0%	1.6%	0.0%	0.0%	0.3%	96.2%	0.8%	0.0%	0.0%
	throw	0.0%	0.0%	0.0%	10.6%	12.2%	1.6%	0.0%	81.7%	0.0%	0.0%
	walk	0.0%	0.0%	1.1%	0.0%	0.0%	0.0%	0.0%	0.0%	100.0%	0.5%
	wave	0.0%	0.0%	0.8%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	99.3%

Fig. 10. Recognition result

Table 2. Variance average of angular displacement

Activity	Variance
Carry	160.8
Claphands	215.9
Pickup	526.9
Pull	127.4
Push	135.4
Sitdown	223.1
Standup	179.7
Throw	568.3
Walk	93.2
Wave	932.9

Table 3. Recognition rate: with no-quantized feature (angle displacement) vs. with ternarized feature

Activity	no-quantized feature (angle displacement) [%]	conventional method [%]	ternarized feature [%]
Carry	97.3	97.5	99.4
Claphands	88.5	100	99.6
Pickup	81.3	97.5	92.7
Pull	85.4	92.5	78.8
Push	96.8	81.5	84.7
Sitdown	87.3	91.5	95.2
Standup	88.8	93.5	96.6
Throw	96.3	59.0	81.7
Walk	82.2	96.5	100
Wave	98.7	100	99.2
Average	90.3	90.9	92.7

以外の行動では、その識別率に大きな差はないが、部分的に識別率の向上が見られ、全体的な識別率の平均では Xia らの手法の識別率 90.9%を上回る 92.7%の識別率が得られた。

一般的に Random Forests において少ない学習データサイズでは精度が向上しない場合がある。しかし、使用した学習サンプル数に比べ、良好な識別結果が本稿では得られた。これは、量子化の効果で特徴空間のサイズが小さくなり、Random Forests の識別が有効に働いたことを示唆していると考えられる。

〈5・3〉 生データを使用した場合との比較 最も識別率が高い結果となった3値化特徴と、量子化しない生データを使用した場合の識別率をTable 3に示す。生データは量子化する前の角度変位情報であり、時系列フレーム数 N と比較フレーム間隔 f は3値化する場合と同様の設定 $N=20$, $f=15$ とする。

結果をみると Pull, Push, Throw の3種類の行動において量子化した場合の識別率が低下している。Fig. 10より、PushはThrowに、PullはPushとThrowに誤認識されやすいことが分かる。この3種類は直立して片腕だけを動かしているものである。下半身を中心とした角度変位の値が類似しており、さらに量子化して微小動作を排除することで上半身の角度変位が類似傾向となったため、精度が低下したものと考える。しかし、その他の行動では3値化の場合の方が識別率が良く、生データで識別率が低い行動に対しても良好な結果が得られた。

今回はUTKinect-Action dataset内の行動データという限られた対象での検証であるが、量子化することにより個人毎の微小な動作の差、無意識な動作、癖などによる観測データの変化にもロバストな識別ができる可能性を示した。

〈5・4〉 考 察 〈5・1〉節より、量子化特徴の比較では3値化の行動認識率が最も高く、続いて4値化、5値化、2値化が高い結果となった。3値化、4値化の識別率は2値化、5値化に比べて高く、3値化の方が時系列フレーム数 $N=20$ の時僅かに4値化の識別結果を上回る結果となった。この量子化レベルの違いによる識別率の違いを以下に考察する。

人物の行動は無意識な動作や癖、身長などにより人物毎の動作速度は異なる。そのため、同じ動作であっても、観測データ上では異なる値の変化量となり、学習する際に悪影響を及ぼす。はずれ値等の要因を除けば、関節位置を取得する際のノイズに加え、人物の個体差による観測データの差異は識別率低下の原因となる。どちらの方向へ動いたかという情報はそこから繋がる次の行動にも関係するため、個体差によって発生する微小な観測データの変動は、人物の動作を識別する際に考慮すべきである。提案手法や Xiaら⁽¹⁶⁾の手法は時系列要素を学習データに含めているため、人物の動作速度が観測データ変化に及ぼす影響はさらに顕著になる。

それに対して、量子化特徴では変化を符号化（量子化）することにより、個体差に頑健な特徴量を実現している。3値化と4値化を比較した場合、4値化は変化の方向とその変化量を段階的に表現する特徴量であるのに対し、3値化は変化方向と動作が無い状態を表現することが可能となっている。それにより、3値化特徴では個体差による微小な角度情報の変動を“0（動作無し）”として認識することで、識別率が向上したと考えられる。

4値化の場合も同様に、比較的高い識別率となっている。4値化の場合、先述の個体差による観測データの微妙な変動が全て“1（伸長方向への変化）”、“-1（屈曲方向への変化）”

Table 4. Comparison of quantization

Frame	Angle [deg]	Inter-frame difference	Ternarization	Quinarization
1	180	85	1	2
2	175	75	1	2
3	179	69	1	2
4	179	84	1	2
5	179	84	1	2
6	156	61	1	2
7	148	53	1	2
8	135	40	1	2
9	125	30	1	2
10	100	5	0	0
11	110	15	0	1
12	90	-21	-1	-2
13	84	-19	-1	-1
14	75	-25	-1	-2
15	81	-10	0	-1
16	98	3	0	0
17	120	25	1	2
18	121	26	1	2
19	100	5	0	0
20	110	15	0	1

に分類され、行動に伴う正しい変化は全て“2（伸長方向への変化が大きい）”、“-2（屈曲方向への変化が大きい）”に分類されたと考えられる。5値化の場合は、個体差による細かい変動が“0（動作無し）”と“1（伸長方向への変化）”、“-1（屈曲方向への変化）”の3つの値に符号化されてしまったため、識別の際に有効となる指標を見つけ難く、識別率低下に繋がったと考える。また、2値化では変化方向を表現できないため識別率が低下した。

また、同じ行動データであっても個々の角度変位についてみると、量子化レベルによって値が異なる場合がある。例として Carry 動作における肘周りの角度、角度変位（20フレーム前との差分）、3値化データ、5値化データを20フレーム間切り出したものをTable 4に示す。動作としては腕を伸ばして物体を拾い、腕を曲げて物体を保持するまでの変化を表す。例えば11フレームと15フレームにおいて、3値化では0（動作無し）であるが、5値化では1（伸長・微小）、-1（屈曲・微小）と量子化レベルによって値が異なっている。

しかし、〈2・2〉節で述べた通り、角度変位は20箇所の身体部位の位置情報から3つを組み合わせで計算している。そのため、量子化特徴のフレーム毎の次元数は ${}_{20}C_3=1140$ 次元である。これを20フレーム分統合すると次元数は20フレーム \times 1140次元=22800次元となる。このため、少数の角度変位の量子化レベルによる値の違いは識別結果に大きな影響を与えていないと考える。本稿で取り扱った全身動作に関して言えば、個々の要素の矛盾を含んだ上で3値化が有効であるとの結果を得た、といってよい。

以上、動きの有無、および変化方向の情報を特徴内に備えていれば、量子化特徴は機械学習に十分な行動識別性能を発揮できる可能性があることを示唆している。

6. おわりに

行動識別のための新たな特徴表現方法として、角度変位の量子化特徴量を提案した。量子化することで個人毎の差や無意識な動作、癖などによる微小な観測データの変動に強くなると仮定し、量子化特徴についての調査を行った。日常生活においてよくある行動が撮影された UTKinect-Action dataset を用いた実験の結果、量子化により識別率が向上することを示した。その過程で、行動識別に最適な量子化レベルおよびパラメータについて調査した。量子化しない場合との比較、および先行研究との比較を行い、行動識別に有効な特徴を生成できる可能性を示した。本研究の貢献は以下の2点である。

- (1) 行動識別における量子化特徴の有効性の評価
3 値化した場合の量子化特徴を用いて、量子化しない場合と Xia らが提案する手法に対して、識別率の比較を行った。量子化しない場合より 2.4%、Xia らの手法より 1.8% 高い識別率が得られ、10 種類の行動に対して平均 92.7% の識別率を実現できた。
- (2) 行動認識に最適な量子化およびパラメータの調査
“2 値化”, “3 値化”, “4 値化”, “5 値化” に対して、量子化の際の “時系列フレーム数”, “比較フレーム間隔”, “角度閾値” の 3 種類のパラメータを変化させて行動識別に最適な量子化パラメータの調査を行った。調査の結果、時系列フレーム数 $N=20$ フレーム, 比較フレーム間隔 $f=15$ フレーム, 角度閾値 $20[\text{deg}]$ とした場合の 3 値化で最も高い識別率が得られた。これは全身動作を想定したデータセットに関して、角度変位 3 値化特徴が識別率向上に有効であるということを示唆している。

量子化特徴を用いることで、個体差や微妙な変化に対する頑健性が考察された。今後、この頑健性についてさらに詳細に調査することで、観測データに多くの変動・ノイズが含まれる環境に対する有効性を調査する予定である。また、識別の対象となる行動の数や種類によっては、2 値化などのより粗い特徴や、5 値化などの細かい変化を表現出来る特徴の方が有利である可能性も考えられる。今後はクラス数などを変化させた場合や、さらに複雑な行動を対象として、それぞれの量子化特徴について検討する。

謝辞

本研究の一部は、JST RISTEX 問題解決型サービス科学研究開発プログラムの援助を受けた。

文 献

- (1) R. Szeliski : “Computer Vision: Algorithms and Applications”, Springer, ISBN-1848829345 (2010)

- (2) J. K. Aggarwal and Q. Cai : “Human Motion Analysis: A Review”, International Journal of Computer Vision and Image Understanding, No.73-3, pp.428-440 (1999)
- (3) T. Moeslund, A. Hilton, and V. Kruger : “A Survey of Advances in Vision-based Human Motion Capture and Analysis”, International Journal of Computer Vision and Image Understanding, Vol.104-2, pp.90-126 (2006)
- (4) J. K. Aggarwal and M. S. Ryoo : “Human Activity Analysis: A Review”, International Journal of ACM Computing Survey (CSUR), Vol.43-3, pp.16:1-16:43 (2011)
- (5) N. Dalal and B. Triggs : “Histograms of Oriented Gradients for Human Detection”, Proceedings of the International Conferences on Computer Vision and Pattern Recognition, pp.886-893 (2005)
- (6) A. Klaser, M. Marszalek, and C. Schmid : “A Spatio-temporal Descriptor based on a 3D Gradients”, Proceedings of the International Conferences on British Machine Vision Conference, pp.995-1004 (2008)
- (7) P. Scovanner, S. Ali, and M. Shah : “A 3-dimensional Sift Descriptor and its Application to action recognition”, Proceedings of the International Conferences on ACM Multimedia, pp.357-360 (2007)
- (8) W. Freeman and M. Roth : “Orientation Histograms for Hand Gesture Recognition”, Proceedings of the International Conferences on Automatic Face and Gesture Recognition, pp.296-301 (1995)
- (9) D. G. Lowe : “Distinctive Image Features from Scale-invariant Keypoints”, International Journal of Computer Vision, Vol.60-2, pp.91-110 (2004)
- (10) K. Mikolajczyk and C. Schmid : “Scale and Affine Invariant Interest Point Detectors”, International Journal of Computer Vision, Vol.60, pp.63-86 (2004)
- (11) V. N. Vapnik : “Statistical Learning Theory”, John Wiley & Sons (1998)
- (12) O. Oreifej and L. Zicheng : “HON4D: Histogram of Oriented 4D Normals for Activity Recognition from Depth Sequences”, Proceedings of the International Conferences on Computer Vision and Pattern Recognition, pp.716-723 (2013)
- (13) W. Li, Z. Zhang, and Z. Liu : “Action Recognition based on a Bag of 3D Points”, Proceedings of the International Conferences on Computer Vision and Pattern Recognition Workshops, pp.9-14 (2010)
- (14) J. Wang, Z. Liu, J. Chorowski, Z. Chen, and Y. Wu : “Robust 3D Action Recognition with Random Occupancy Patterns”, Proceedings of the European Conference on Computer Vision, pp.872-885 (2012)
- (15) J. Wang, Z. Liu, Y. Wu, and J. Yuan : “Mining actionlet ensemble for action recognition with depth cameras”, Proceedings of the International Conferences on Computer Vision and Pattern Recognition, pp.872-885 (2012)
- (16) L. Xia, C. Chen, and J. K. Aggarwal : “View Invariant Human Action Recognition Using Histograms of 3D Joints”, Proceedings of the International Conferences on Computer Vision and Pattern Recognition Workshop, pp.20-27 (2012)
- (17) G. Csurka, C. Dance, L. Fan, J. Willamowski, and C. Bray : “Visual Categorization with Bags of Keypoints”, Proceedings of the European Conference on Computer Vision Workshop, pp.59-74 (2004)
- (18) J. Shotton, A. Fitzgibbon, M. Cook, T. Sharp, M. Finocchio, R. Moore, A. Kipman, and A. Blake : “Real-Time Human Pose Recognition in Parts from Single Depth Images”, Proceedings of the European Conference on Computer Vision, pp.1297-1304 (2011)
- (19) T. Yamabe, H. Kataoka, and A. Nakamura : “A Study on Features for Early Recognition of Human Activities”, Proceedings of the SICE Annual Conference, pp.1334-1330 (2014)
- (20) L. Breiman : “Random Forests”, Machine Learning, Vol.45, pp.5-23 (2001)

山 辺 智 晃 (非会員) 1992 年生。2014 年東京電機大学未来科学部ロボット・メカトロニクス学科卒業、同大学大学院未来科学研究科ロボット・メカトロニクス学専攻修士課程在学中。コンピュータビジョン技術を用いた行動認識の研究に従事。



片岡裕雄



（非会員）1986年生。2014年慶應義塾大学大学院後期博士課程修了，博士（工学）。現在，東京大学博士研究員，慶應義塾大学客員研究員兼務。2012～2013年産業技術総合研究所技術研修員，2013年ミュンヘン工科大学（TUM）訪問研究員。コンピュータビジョンを専攻。主に，行動認識，高度道路交通システム，スポーツ解析における人物行動解析に従事。精密工学会，自動車技術会，計測自動制御学会，各会員。

中村明生



（正員）1973年生。1996年東京大学工学部精密機械工学科卒業。2001年同大学大学院工学系研究科精密機械工学専攻博士課程修了，博士（工学）。同年埼玉大学工学部情報システム工学科助手。2005年東京電機大学工学部第一部機械情報工学科助教授。2007年同大学未来科学部ロボット・メカトロニクス学科准教授，現在に至る。複数ロボット操作システム，マンマシンインタフェース，コンピュータビジョンの研究に従事。電子情報通信学会，日本ロボット学会，精密工学会，日本機械学会，計測自動制御学会，IEEE，ACMなどの会員。
