

ニューラルネットワークを用いたウェアラブル慣性センサのアップサンプリング手法

吉村 直也^{1,a)} 前川 卓也^{1,b)} 天方 大地^{1,c)} 原 隆浩^{1,d)}

受付日 2019年10月9日, 採録日 2020年5月12日

概要: スマートウォッチをはじめとするウェアラブルデバイスには、加速度センサやジャイロスコープなど多数のセンサが搭載されており、行動認識やジェスチャ・コントロール、手書き入力、ゲームコントローラ、手話を利用した遠隔コミュニケーションなど様々なアプリケーションへの応用が期待されている。しかし、一般に流通しているスマートウォッチでは、加速度センサ・ジャイロスコープのサンプリングレートが制限されており、高サンプリングレートの信号を取得することが難しいため、ユーザの体の動きをきめ細やかに計測することは難しい。本研究ではニューラルネットワークを用いて信号を補間することで、低いサンプリングレートの信号から高いサンプリングレートの信号を生成する手法を提案する。低いサンプリングレートで取得された信号の失われた情報を復元することは不可能であるため、本研究では事前に収集した高いサンプリングレートの信号を用いて、低いサンプリングレートの信号に欠けている計測値を推定することを試みる。体の特定の部位に装着されたセンサが計測するデータは装着された部位の構造によって制約を受ける。したがって、この装着部位に起因する制約を事前に学習することで、高サンプリングレートの信号を擬似的に復元できる。

キーワード: ウェアラブルセンサ, アップサンプリング, ニューラルネットワーク

Upsampling Method for Wearable Inertial Sensors with Neural Networks

NAOYA YOSHIMURA^{1,a)} TAKUYA MAEKAWA^{1,b)} DAICHI AMAGATA^{1,c)} TAKAHIRO HARA^{1,d)}

Received: October 9, 2019, Accepted: May 12, 2020

Abstract: Inertial sensor data collected from wearable smart devices such as smartwatches are expected to be used in various smart applications such as video game controllers, hand drawing, hand writing, gestural input devices, human activity recognition, and remote communication using sign language. However, since the maximum sampling rate of inertial sensors in commercial smartwatches is restricted, capturing fine-grained body movements using the low-sampled signals is difficult for these sensors. Therefore, we propose a new method for generating high sampling rate signals from the low-sampled signals by upsampling the low-sampled signals using interpolation with an artificial neural network. Because it is impossible to obtain “non-existent” data from low-sampled signals according to the information theory, the proposed method estimate these data from experience, i.e., using high-sampled signals prepared in advance for training. This is possible because trajectories of a sensor are restricted by the skeletal structure of the body part to which the sensor is attached.

Keywords: wearable sensor, upsampling, neural network

¹ 大阪大学大学院情報科学研究科
Osaka University, Suita, Osaka 565–0871, Japan

a) yoshimura.naoya@ist.osaka-u.ac.jp

b) maekawa@ist.osaka-u.ac.jp

c) amagata.daichi@ist.osaka-u.ac.jp

d) hara@ist.osaka-u.ac.jp

1. はじめに

1.1 研究背景

近年普及が著しいスマートウォッチなどのウェアラブルデバイスには加速度センサやジャイロスコープといった慣

性センサが搭載されており、装着した体の部位の動きを容易に計測することができる。このため慣性センサから得られたデータを様々な分野に応用することが期待されている [1], [8], [16], [17], [19]。具体的には、スマートウォッチをゲームコントローラ、手書き入力、IoT/CPS デバイス向けのジェスチャコントローラ、行動認識、VR や手話を用いた遠隔コミュニケーションへの応用が研究されている [5], [10], [18], [20]。

Laput ら [10] はスマートウォッチのデバイスドライバを改変することで非常に高いサンプリングレートの加速度信号を取得し、粒度の細かいハンドジェスチャの認識を試みており、「tap」や「flick」などの低いサンプリングレートでは認識が難しいジェスチャを認識できることを報告した。しかし一般的に流通しているスマートウォッチでは、慣性センサ値を取得するサンプリングレートが 50–100 Hz 程度に制限されているため、ユーザの体の動きをきめ細やかにとらえることは難しい。高いサンプリングレートの信号をデバイスドライバの改変なしに利用できれば、きめ細やかな動きをとらえたジェスチャ認識が可能となる。

1.2 研究目的

本研究では、スマートウォッチで取得されたサンプリングレートが低い信号を、ニューラルネットワークを用いて補間を行うことで、擬似的にサンプリングレートが 2 倍の信号を生成するアップサンプリング手法を提案する。アップサンプリングを行うことで、デバイスドライバの改変を行わずに高いサンプリングレートの信号を利用できる。サンプリング定理より低いサンプリングレートの信号から、高いサンプリングレートの信号を完全に復元することはできない。そこで本研究では、センサを装着した体の部位によってセンサの動きが制限されることに着目し、事前に高いサンプリングレートで計測できるデバイスで収集したデータを用いて、サンプリングレートの高い信号と低い信号の関係性をニューラルネットワークで学習することで補間値の推定を行う。

1.3 研究手法

近年コンピュータビジョン分野では、畳み込みニューラルネットワーク (CNN) を用いて低解像度の画像を高解像度の画像に変換する「超解像技術 (Image Super-resolution)」の研究がさかに行われている [3], [11]。また、音声信号処理の分野においても「周波数帯域拡張技術 (Bandwidth Extension; BWE)」という信号の解像度を拡張する技術の研究が行われている [9], [13]。これらの手法は、人間が高精細に見える・聞こえる信号を生成することが目的であるが、本研究ではアップサンプリングした信号を用いて姿勢推定や行動認識を行うことを想定しているため、実際に高サンプリングレートなデバイスで観測された信号に近い信

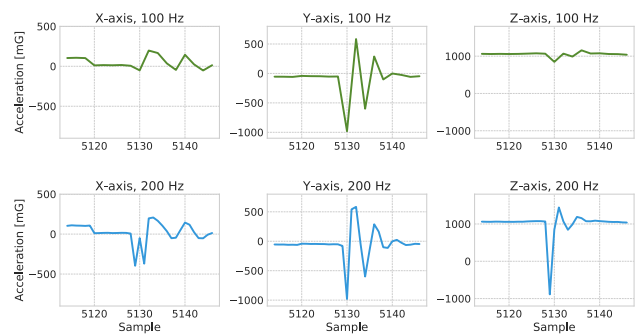


図 1 3 軸加速度信号の例。上段は 100 Hz、下段は 200 Hz の信号を示す。これらの信号はユーザがセンサを tap した際に計測されたものである

Fig. 1 Time-series of acceleration signals. The lower graphs show the original 200 Hz three-axis signals. The upper graphs show the undersampled 100 Hz signals. The signals were collected when a person taps his arm.

号を生成することが求められる。また BWE では FFT スペクトルの拡張を想定しているが、本研究は時間領域と周波数領域の両方の情報を必要とするアプリケーションに対応するため、高いサンプリングレートに変換された時系列信号を擬似的に生成する。

手首に装着した加速度センサは、手の移動に起因する長期的な変化と、手に衝撃を与えた場合などの短期的な変化の両方を観測している。したがって、慣性センサのデータをアップサンプリングする場合、長期的な変化と短期的な変化をともに考慮することが必要であるが、この 2 種類の変化のパターンを同時に考慮することは容易ではない。長期的な変化は基本的に滑らかな波形となるため、線形補間といった単純な手法でも高い精度で補間値を推定することができる。しかし、短期的な変化は単純な方法では正確に補間することができない。そこで提案手法では、長期的な変化を線形補間を用いて補間し、短期的な変化はニューラルネットワークによって線形補間の誤差である残差を推定するという形で補正し、2 つの手法を組み合わせることで補間値を推定する。

これに加えて、複数の軸がある慣性センサの任意の軸の信号をアップサンプリングする際に、他の軸との関係性を考慮することを試みる。図 1 は、200 Hz の 3 軸加速度センサの波形 (上段) と、それをダウンサンプリングした 100 Hz の波形 (下段) の例である。サンプリングレートが低い信号のそれぞれの軸は、サンプリングのタイミングにより衝撃などの波形を観測できない場合がある。図 1 の場合では、100 Hz の z 軸の信号は激しい変化をとらえられていないが、x・y 軸では激しい変化をとらえることができている。z 軸のみからこの激しい変化を復元することは困難であるが、x・y 軸の情報を考慮することができれば復元は可能だと考えられる。

1.4 技術的貢献

本研究の技術的貢献は以下のとおりである。(1) 入力信号のサンプリングレートを2倍にするアップサンプリング手法を提案する。筆者らの知る限り、ニューラルネットワークを用いて慣性センサのデータをアップサンプリングする研究は行われていない。(2) 長期的な変化と短期的な変化の両方を同時に考慮するアップサンプリング手法を提案する。さらに慣性センサの各軸の間の相互関係を考慮する。(3) 2種類のデータセットを用いて、提案手法の有効性を評価した。

本論文の構成は以下のとおりである。まず画像と時系列信号に対するアップサンプリングの関連研究を紹介する。次に、ニューラルネットワークを用いて慣性センサをアップサンプリングする手法を提案する。最後に被験者から収集したデータセットを用いて、評価実験を行う。

2. 関連研究

コンピュータビジョンの分野では、低解像度の2次元画像を高解像度の画像に変換する Single Image Super Resolution (SISR) という手法が研究されており、近年 CNN を用いたモデルが数多く提案されている [3]。そのなかでも、敵対的生成ネットワーク (GAN) を用いることで、人間が見てより鮮明な画像を生成できることが報告され、注目を集めている [11]。音声信号処理の分野では、周波数帯域拡張技術 (BWE) と呼ばれる与えられた信号から高周波数帯域の情報を推定する技術が研究されている。SISR で提案された手法を、音声信号に適用した手法も提案されている [9]。

これらの手法に基づいて、本研究では CNN を用いて3軸の慣性センサの信号をアップサンプリングする手法を提案する。空間的に冗長な画像や時間的に冗長な音声信号と異なり、ウェアラブルセンサから取得した慣性センサの信号は突発的な衝撃などの外れ値の情報も含むため、そのような問題に対応する必要がある。

3. アップサンプリング手法

3.1 想定環境

本研究では3軸の加速度センサやジャイロスコープといった慣性センサを想定する。センサのサンプリングレートを $2\Delta t$ とすると、以下のようなセンサから取得した時系列信号 A_l を扱う。また、 A_l の各要素 $a_{2n\Delta t}$ は、以下のような xyz の3軸の計測値からなるベクトルである。

$$A_l = [a_0, a_{2\Delta t}, \dots, a_{(2n-2)\Delta t}, a_{2n\Delta t}, \dots]$$

$$a_{2n\Delta t} = [x_{2n\Delta t}, y_{2n\Delta t}, z_{2n\Delta t}]$$

アップサンプリングを行うニューラルネットワークを学習するための、サンプリングレート Δt で計測されたセンサデータは下記のとおりとなる。

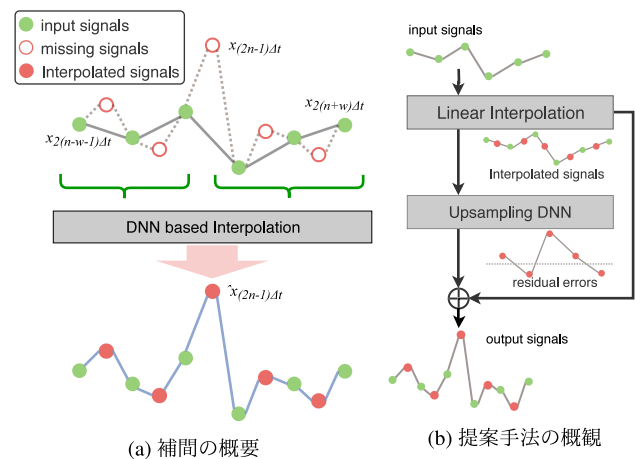


図2 アップサンプリング手法の概要。DNNはディープニューラルネットワークの略である

Fig. 2 Upsampling low sampling rate sensor signals using the proposed method. DNN refer to deep neural networks.

$$A_h = [a_0, a_{\Delta t}, a_{2\Delta t}, \dots, a_{(2n-2)\Delta t}, a_{(2n-1)\Delta t}, a_{2n\Delta t}, \dots]$$

本研究では、慣性センサの計測モードとして、センサが計測値を取得するときのみ起動する「power-saving low-duty cycle mode」用いてセンサデータを取得することを想定する。duty cycle modeでは、前処理を行わずそのままの計測値を取得するため、 A_l における $a_{2n\Delta t}$ と A_h における $a_{2n\Delta t}$ は同値であると思なすことができる。

3.2 提案手法の概要

図2(a)に示すように、 A_l をアップサンプリングするために $x_{(2n-1)\Delta t}$ を推定する際は、 A_l における3軸かつ、補間値を推定する $x_{(2n-1)\Delta t}$ の前後 w 個の計測値にあたる $x_{2(n-w-1)\Delta t}$ から $x_{2(n+w)\Delta t}$ を利用する。

図2(b)に提案するアップサンプリング手法の概要を示す。たとえば、Android スマートウォッチにおけるジャイロスコープの最大のサンプリングレートは、50 Hz である。この場合、提案手法はスマートウォッチから取得した 50 Hz のジャイロスコープの信号から、100 Hz の信号を生成する。提案手法では線形補間とニューラルネットワークを利用してアップサンプリングを行う。ニューラルネットワークは、センサから取得した信号を入力とし、目標の信号と線形補間の結果の誤差である残差を推定する。最後に線形補間とニューラルネットワークの出力を組み合わせることで、高いサンプリングレートの信号を得る。

3.3 線形補間

提案手法では最初に、線形補間を用いてサンプリングレートが2倍の信号を計算する。x軸の時刻 $(2n-1)\Delta t$ におけるセンサデータは以下のように計算される。

$$x'_{(2n-1)\Delta t} = \frac{x_{(2n-2)\Delta t} + x_{2n\Delta t}}{2}$$

3.4 残差を推定するニューラルネットワーク

提案手法では、各軸の線形補間の結果に対して、ニューラルネットワークを用いて残差を推定する。x 軸の時刻 $(2n-1)\Delta t$ における、最終的な補間値は以下のように計算される。

$$\hat{x}_{(2n-1)\Delta t} = x'_{(2n-1)\Delta t} + r\hat{x}_{(2n-1)\Delta t}$$

$r\hat{x}_{(2n-1)\Delta t}$ はニューラルネットワークによって推定された残差である。本研究においては、軸ごとにニューラルネットワークを構築する。図 3 に提案手法で用いたニューラルネットワークの構造を示す。時刻 $(2n-1)\Delta t$ における任意の軸の残差を推定するために、以下に示されるような時刻 $(2n-1)\Delta t$ の前後 w 点（合計 $2w$ 点）にあたる 3 軸加速度信号をニューラルネットワークに入力する。

$$[a_{2(n-w)\Delta t}, \dots, a_{2n\Delta t}, \dots, a_{2(n+w)\Delta t}]$$

長期的な変化と短期的な変化の両方を抽出するために、図 3 に示すような 2 種類のフィルタサイズのブロックを連結したマルチスケールな畳み込み層を用いる。図 3 の上側の 1 つのブロックでは、大きな畳み込みフィルタ（本研究では 7×1 ）を用いて長期的な変化を抽出する。同様に図 3 の下側の 3 つのブロックでは、小さい畳み込みフィルタ（本研究では 3×1 ）を用いて短期的な変化を抽出する。

各ブロックは 3 層の 1 次元畳み込み層で構成されている。1・2 層目の畳み込みブロックは、 $F \times 1$ の畳み込みフィルタとストライド $s = 2$ を用いてセンサデータから特徴を抽出する。3 層目の畳み込みフィルタでは、ストライドを $s = 2$ と設定し、時間方向に情報を圧縮することでより抽象的な特徴を抽出する。このように畳み込み層を用いて、残差を推定したい時刻 $(2n-1)\Delta t$ 付近の 3 軸慣性センサのデータから短期的な変化に関する特徴を抽出する。最後に、上下の畳み込みブロックで抽出された特徴は、結合されたうえで全結合層に渡され残差を推定する。全結合層では活性化関数を使用していない。

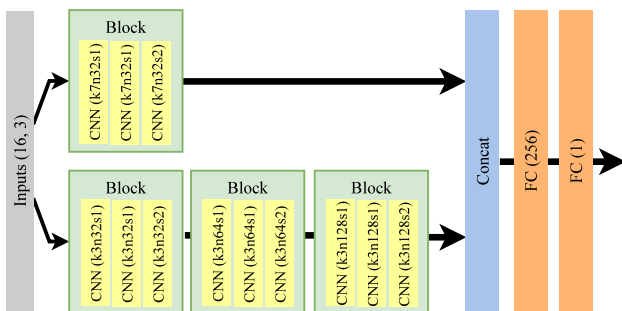


図 3 残差を推定するネットワークの構造。k7n32s1 はフィルタサイズ 7×1 、フィルタ数 32 の畳み込み層を表す

Fig. 3 Neural network structure used in the proposed method. For example, k7n32s1 is a convolutional layer using 32 of 7×1 filters with a stride of one sample.

上部分のブロックにおける畳み込み層では、フィルタ数は 32 に設定した。また下側部分のブロックでは、抽出した情報を失わないようにブロックが 1 つ深くなるごとにフィルタ数を 2 倍にした。使用したフィルタ数は入力側からそれぞれ 32, 64, 128 である [11], [13]。

3.5 ニューラルネットワークの学習

ニューラルネットワークでは残差を推定する。時刻 $(2n-1)\Delta t$ 、x 軸における残差は以下の式で表される。

$$rx_{(2n-1)\Delta t} = x_{(2n-1)\Delta t} - x'_{(2n-1)\Delta t}$$

$x_{(2n-1)\Delta t}$ は、 A_h における時刻 $(2n-1)\Delta t$ の信号である。ニューラルネットワークの学習では、平均絶対誤差 (MAE) を損失関数として Adam Optimizer [7] を用いて学習を行う。これに加えて、過学習を防止するために Dropout [21] と Batch Normalization [6] を、文献 [4], [23] を参考に、各ブロック 2 層目の畳み込み層の後に適用する。

4. 評価

4.1 データセット

提案手法の評価のために、以下の 2 種類のデータセットを作成した。被験者は IMU (ATR TSND151) を左手首に装着し、1,000 Hz でセンサデータを収集した。

4.1.1 手書き文字データセット

このデータセットでは、被験者が英数字 (A-X, 0-9) と記号 (丸, 三角, 四角, 星) の 40 種類をそれぞれ 10 回ずつ空中に書き、これを左手首に装着したセンサで計測した。5 名の被験者 (20–50 代, 男 3 女 2) から収集し、長さはおおよそ 106 分である。学習データを増やすために、先の 5 名とは異なる被験者 (20 代, 男 1) に英数字および記号を書いてもらった。これらのデータは追加の学習データのみを用いており、長さはおおよそ 60 分である。

4.1.2 オブジェクトベース行動認識用データセット

このデータセットは、ユーザが使用中のオブジェクトがユーザの行動を反映しているという考えに基づき [15], ユーザが使用しているオブジェクトを推定することで行動認識を行うことを目的として作成した。先行研究 [10], [15] を参考に、本研究では表 1 に示した手に持って使用する 9 種類の電化製品を用いた。7 名の被験者 (20–50 代, 男 4 女 3) はランダムな順番で 9 種類のオブジェクトを 1 回ずつ使用し、これを 1 セッションとして 10 セッション分の

表 1 本研究で用いた電化製品

Table 1 Handheld electric devices used in this study.

	Object		Object		Object
A	blender	D	fun	G	pepper mill
B	drill	E	handy blender	H	shaver
C	screwdriver	F	handy vacuum	I	toothbrush

データを収集した。収集したデータセットの長さはおおよそ70分である。

またこれに加えて、先の7名とは異なる被験者（20–30代，男2）がオブジェクトを使用したデータと，日常生活のなかでセンサを装着して計測したデータを追加で収集した。これらのデータは追加の学習データとしてのみ使用した。データの長さは，前者はおおよそ10分，後者はおおよそ150分である。

4.2 評価方法

本研究で使用するデータセットは1,000 Hzのサンプリングレートで収集されている。評価に用いる高いサンプリングレートの信号を生成するためにダウンサンプリングを行い，サンプリングレートが高い加速度センサの信号 A_h とジャイロスコープの信号 G_h を生成する。次に，サンプリングレートが低い信号を生成するために A_h, G_h に対してダウンサンプリングを行い，サンプリングレートが低い加速度センサの信号 A_l ，ジャイロスコープの信号 G_l を得る。本研究ではニューラルネットワークはセンサの軸ごとに用意した。手書き文字データセットでは $A_l \cdot G_l = 50$ Hz, $A_h \cdot G_h = 100$ Hz, オブジェクトベース行動認識データセットでは $A_l = 100$ Hz, $A_h = 200$ Hz とした。後者はオブジェクトの振動を検知するため加速度センサのみを使用することを想定した。加速度センサはスマートウォッチにおいてより高いサンプリングレートを使用できるので，行動認識データセットではより高いサンプリングレートを用いた。また，1,000 Hz から 200 Hz のダウンサンプリングの際に，異なるオフセットを用いることで5系列の200 Hzの信号を生成することができる。オブジェクトベース行動認識データセットに関しては学習データを増やすために，5系列すべてをダウンサンプリングして A_h と A_l のペアを生成し，学習に利用する。

評価指標として，Mean Absolute Error (MAE), Log Spectral Distortion (LSD) を用いる。LSD は BWE の研究で使用されている評価指標であり，教師信号 A_h と生成した信号のパワースペクトルの距離を測ることができる。MAE, LSD はともに，値が小さい方がアップサンプリングの誤差が小さいことを意味する。

提案手法の有効性を評価するために，以下の手法と比較した。

- **Proposed** : 提案手法
- **Proposed (Single)** : ニューラルネットワークの構成は提案手法と同じであるが，3軸の相互関係は考慮しない。つまりある軸の残差を推定する際に，ニューラルネットワークには3軸ではなく残差を推定する1軸分の信号のみを入力する。
- **Dense** : ニューラルネットワークで直接，補間値 $\hat{x}_{(2n-1)\Delta t}$ を推定する。提案手法と異なり線形補間

を使用しない。入力は Proposed と同じく推定したいデータの前後 $2w$ 点の3軸加速度データ，出力は補間値 $\hat{x}_{(2n-1)\Delta t}$ の1点である。ニューラルネットワークの構成は，全結合層3層（各層256ユニット）と出力層の構成で，活性化関数には \tanh を使用した。また学習時に全結合層の後には Dropout を用いた。

- **Linear** : 線形補間のみを用いて補間値 $\hat{x}_{(2n-1)\Delta t}$ を推定する。
- **Spline** : スプライン補間を用いて補間値 $\hat{x}_{(2n-1)\Delta t}$ を推定する。提案手法と同様に A_l における $x_{2(n-w-1)\Delta t}$ から $x_{2(n+w)\Delta t}$ までの値を用いて補間値の推定を行う。また残差を推定するニューラルネットワークの学習は，1人のユーザをテストデータ，そのほかすべてをニューラルネットワークなどの学習データとして使用する「Leave-one-participant-out Cross Validation」によって行った。

本研究では，残差 $r\hat{x}_{(2n-1)\Delta t}$ を推定するために時刻 $(2n-1)\Delta t$ の前後 $w = 16$ 点の計測値を用いた。つまり，ニューラルネットワークへの入力の系列長は32である。大きい w を用いることで長期的な傾向がある場合にはそれをとらえることができる。しかし残差の推定には，推定したい時刻の近傍の値がより重要であるため， w を大きくすることが必ずしも性能の向上に直結するとは限らない。大きい w を用いると補間の遅延が大きくなるため，リアルタイム性を必要とするアプリケーションには適さない。本研究では残差の推定精度と低遅延を両立する値として $w = 16$ を用いた。

4.3 結果：手書き文字データセット

4.3.1 アップサンプリングの評価

表2にアップサンプリングの結果を示す。加速度センサに関して，提案手法 (Proposed) は Linear と比較して80%から90%程度 MAE を削減した。さらに，ジャイロスコープでは MAE をおおよそ50%削減した。また LSD の結果より，提案手法は周波数領域においても誤差を削減していることが分かる。図4に，提案手法によってアップサンプリングした信号を示す。50 Hz は Linear, 100 Hz は教師信号 A_h, G_h と同様である。加速度 z 軸の時刻 $t = 0.19 \sim 0.21$ の結果から，提案手法は波形が激しく変化した場合で

表2 アップサンプリングの誤差 (手書き文字データセット)

Table 2 Upsampling errors for hand writing data.

評価指標	加速度センサ		ジャイロスコープ	
	MAE [G]	LSD [dB]	MAE [deg/sec]	LSD [dB]
Proposed	0.00241	2.70	0.422	2.82
Proposed (Single)	0.00200	2.44	0.431	2.92
Dense	0.04848	10.34	23.016	18.13
Linear	0.01066	6.20	0.991	4.74

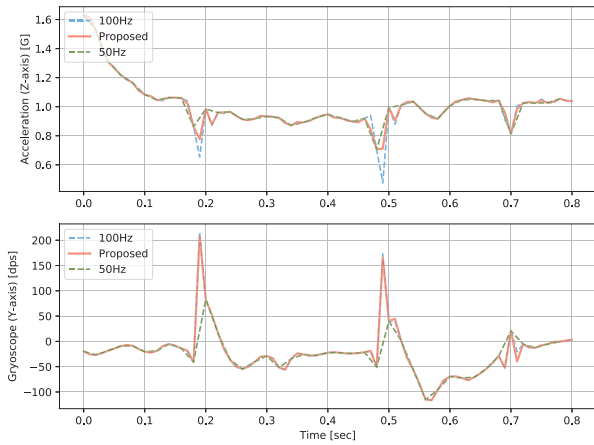


図 4 アップサンプリングした信号の例 (手書き文字データセット, 被験者 A)

Fig. 4 Example of upsampled signals (handwritten dataset, Subject A).

表 3 ユーザ別アップサンプリングの誤差 (手書き文字データセット)

Table 3 Upsampling errors of handwriting data for each participant.

手法	MAE [G] (加速度センサ)		MAE [deg/s] (ジャイロスコープ)	
	Proposed	Linear	Proposed	Linear
User A	0.00395	0.01651	0.635	1.409
User B	0.00158	0.00808	0.266	0.879
User C	0.00116	0.00526	0.292	0.524
User D	0.00131	0.00595	0.297	0.685
User E	0.00158	0.00786	0.283	0.703

も高い精度でアップサンプリングができていることが分かる。

提案手法は Linear と比較して大きく MAE を削減しているが、これは提案手法が突発的なセンサデバイスの動きに対応する信号を適切に補間できたためだと考えられる。センサデバイスが激しく動くと局所的に非常に大きな振幅を取るため、線形補間では誤差が大きくなりやすい。Proposed (Single) は、加速度センサにおいて Proposed をわずかに上回った。センサの傾く方向などの影響により、図 1 とは異なり実際に激しい変化が観測されない軸が存在するため、3 軸の関係を考慮しない方が良い場合があると考えられる。表 2 より、Dense の MAE は提案手法に比べて非常に大きいことが分かる。Dense では長期的な変化と短期的な変化を同時に推定しているため、小さな高周波ノイズを多く生成する傾向にあり、MAE が非常に大きな値となったと考えられる。また Spline は、50 Hz の信号にオーバフィッティングしているため誤差が大きくなった。

表 3 にユーザごとのアップサンプリングの結果を示す。Linear と比較して、被験者 B・C・D・E においては提案手法は加速度信号の MAE を 80%程度、ジャイロスコープの MAE を 60%程度削減した。被験者 A の MAE は他の被験

表 4 姿勢推定結果

Table 4 Angular errors of sensor pose estimation using upsampling signals.

	MAE [deg]	MSE [deg ²]
Linear	6.87	306.18
Proposed	3.93	167.24

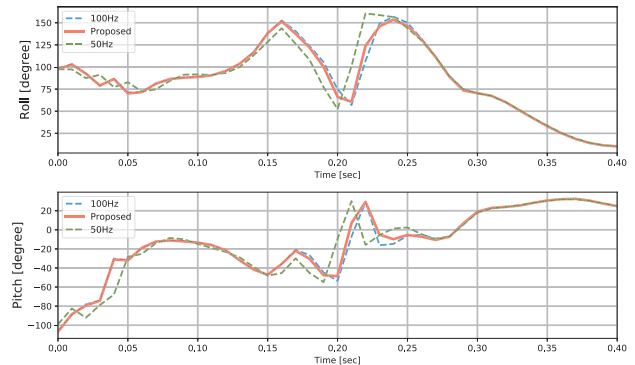


図 5 推定した姿勢角 (Roll, Pitch) の例

Fig. 5 Example of estimated sensor pose of the hand writing dataset.

者に比べて高い傾向にあるが、これは被験者 A がデバイスを振る動作を多く行ったことが原因であると考えられる。しかし、この被験者に関しても提案手法は他の手法に比べて高い精度でアップサンプリングができている。

4.3.2 姿勢推定

センサの姿勢推定におけるアップサンプリングの有効性を検証するために、拡張カルマンフィルタを用いてセンサの姿勢推定を行った [14], [22]。まず、100 Hz の加速度センサとジャイロスコープの信号を用いて姿勢角を計算し、これを正解データとする。次に、提案手法 (Proposed) と線形補間 (Linear) によって生成された 100 Hz の信号に対して同様に姿勢角の推定を行い、これらと正解データの角度の誤差を計算した。

表 4 に推定された姿勢角の平均誤差を示す。提案手法は Linear と比較して誤差をおよそ 3 度削減した。図 5 に推定された姿勢角を示す。Linear の結果からセンサが激しく動いた場合に大きく精度が低下することが分かる。一方で提案手法は、より少ない誤差で姿勢推定を行えていることが分かる。推定したセンサの姿勢の時間変化を 3D アニメーションで描画した*1。この動画からもデバイスが激しく動いた際に提案手法が大きく誤差を削減していることが分かる。

4.4 結果：オブジェクトベース行動認識データセット

4.4.1 アップサンプリング

表 5 にアップサンプリングの結果を示す。教師なし手

*1 <http://www-mmde.ist.osaka-u.ac.jp/~maekawa/paper/supple/accsr.mp4>

表 5 アップサンプリングの誤差 (オブジェクトベース行動認識)

Table 5 Upsampling errors of acceleration data for object-based activity recognition.

	MAE	LSD
Proposed	6.5	8.5
Proposed (Single)	6.2	7.9
Dense	11.8	10.0
Linear	15.8	12.4
Spline	18.6	18.2

表 6 アップサンプリングの誤差 (オブジェクト別)

Table 6 Upsampling errors for each object.

	Proposed	Proposed (Single)	Dense	Spline	Linear
A: blender	12.9	12.2	16.4	48.9	39.6
B: drill	10.7	10.5	12.6	16.1	14.4
C: screwdriver	2.2	2.2	9.4	2.2	2.1
D: fan	3.9	4.0	9.7	7.0	6.0
E: handy blender	4.0	4.0	12.5	3.4	3.5
F: handy vacuum	8.9	8.4	15.8	34.1	31.1
G: pepper mill	2.1	2.2	6.9	2.0	2.0
H: shaver	2.0	2.1	11.3	1.9	1.9
I: toothbrush	7.5	6.1	12.3	32.9	26.3

法である Linear や Spline と比較して、提案手法が大きく MAE を削減していることが分かる。また Dense と比較して提案手法は 50%程度誤差を削減しており、残差を推定する方法の有効性が確認できる。興味深いことに、提案手法の結果は 3 軸の相互関係を考慮しないモデル Proposed (Single) の結果より誤差が若干大きかった。これは家電製品が発する振動はすべての軸で観測することができるため、冗長な情報を入力に含めない方が良かったためだと考えられる。また周波数領域においても提案手法と Proposed (Single) が、高い精度で信号を再構成できていることが LSD の結果から分かる。

表 6 にオブジェクトごとのアップサンプリングの MAE を示す。線形補間での MAE が大きな「blender」, 「toothbrush」においても、提案手法は MAE を大きく削減している。図 6 に toothbrush の波形の例を示す。提案手法では 100 Hz の信号にはない、高周波数成分を再現できていることが分かる。

一方で「handy blender」, 「pepper mill」など、Linear の方が提案手法に比べて MAE が小さいオブジェクトも存在する。これは教師あり学習を用いた手法は、高周波数成分の再現に強いため、高周波数成分をあまり含まないオブジェクトでは誤差が大きくなってしまったと考えられる。図 7 に「pepper mill」のアップサンプリングした信号例を示す。本来の波形は滑らかであるが、Dense の結果は小さな高周波成分ノイズを多く含んでいる。提案手法は Linear には及ばないが、Dense のようなノイズを生成していない

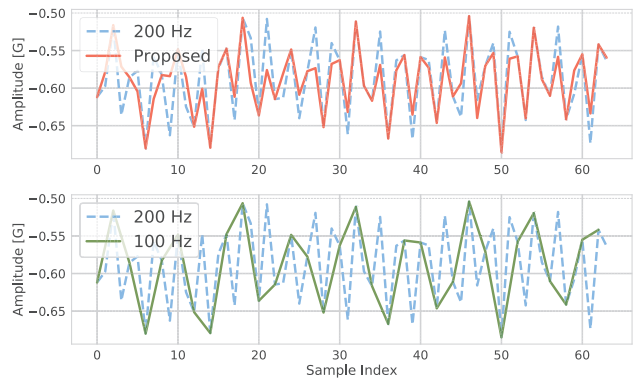


図 6 アップサンプリングした toothbrush を使用した際の信号の例 (x 軸)

Fig. 6 Example of reconstructed x-axis signals by Proposed related to the use of electric toothbrush.

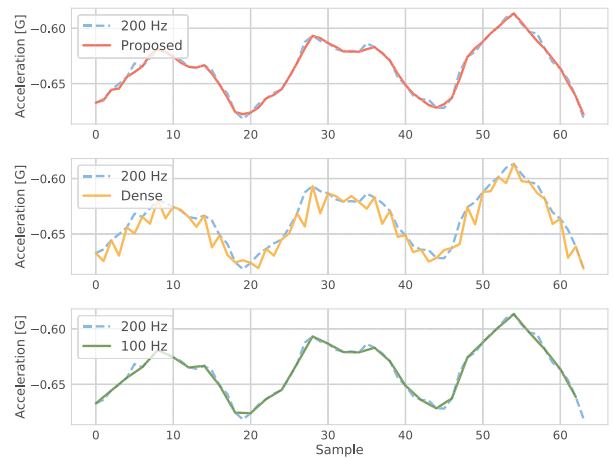


図 7 pepper mill を使用した際の、x 軸のアップサンプリング後の加速度信号の例

Fig. 7 Example of reconstructed x-axis signals by Proposed and Dense related to the use of pepper mill.

ことが分かる。

4.4.2 オブジェクトベース行動認識

アップサンプリングのアプリケーションへの有効性を確かめるために、オブジェクトベース行動認識を行った。アップサンプリングした信号をウィンドウごとに切り分け、特徴抽出したうえでランダムフォレストを用いて認識を行った [2]。ウィンドウサイズは 128 サンプル (0.64 sec) とした。行動認識・ジェスチャ認識の先行研究 [1], [8], [12], [17] を参考にし、各ウィンドウから、Mel-frequency cepstrum coefficient (MFCC) と統計的特徴量 (mean, sum, max, min, max-min, standard deviation) を抽出し、分類器の入力とした。また認識精度の評価は、「Leave-one-session-out Cross Validation」を用いた。

表 7 にオブジェクトベース行動認識の認識結果を示す。100 Hz の信号を使った結果に対して、提案手法によって 200 Hz にアップサンプリングした結果の方が、9 種類のオブジェクト平均の F 値がおおよそ 10%向上した。また提案

表 7 オブジェクトベース行動認識の結果 (Macro-average)

Table 7 Macro-averaged precision, recall, and F-measure of object-bases activity recognition.

	Precision	Recall	F-measure
100 Hz	0.736	0.728	0.728
Proposed	0.837	0.830	0.830
Proposed (Single)	0.813	0.795	0.796
200 Hz	0.844	0.830	0.833

手法でアップサンプリングした F 値は、アップサンプリングの目標とする 200 Hz の信号を使用して認識した結果に近い結果となった。特に、「(F) handy vacuum cleaner」や「(G) pepper mill」は認識精度が大きく向上した。これらのオブジェクトによる高周波の小さな振動は 100 Hz の信号では見分けることが難しいが、アップサンプリングによって区別できるようになったと考えられる。これらの結果から、提案手法を用いることでデバイスドライバを改変せずにきめ細かな行動認識を行うことができたといえる。

5. 考察

5.1 アップサンプリングモデルの汎用性

提案手法は学習ベースでアップサンプリングを実現しており、教師データの質が重要だと考えられる。しかし、本研究では評価実験 (Leave-one-participant-out Cross Validation) において他の被験者のデータで、補間が行えることを確認した。本研究ではタスクごとにデータを用意したが、ある程度の多様なタスクを行ったデータを用意できれば、汎用的に利用できる可能性があると考えられる。今後はタスク間での学習データの可用性に関して検証を行いたい。また本実験では、ノイズやサンプリング間隔の正確性などといったセンサの性能と、アップサンプリング性能の関係は評価できていない。本研究で開発したモデルは 1 つデバイスのデータのみを用いて学習しているため、デバイスに依存した補間を行っている可能性がある。しかし、学習データに複数デバイスで収集したデータを含めることで、デバイスへの依存性を下げられると考えられる。今後検証していきたい。

5.2 アップサンプリングの倍率

本研究では、2 倍のアップサンプリングに関して評価を行ったが、提案手法のフレームワークでは、補間する値を増やせば 2 倍以上のアップサンプリングも可能である。ただし、倍率が上がるほど、推定の手がかりとなる入力データに比べて推定するデータ量の比率が増えるため、精度の低下が発生すると考える。本研究ではサンプリングレートと行動認識の認識精度の関係から、実験で用いた倍率とサンプリングレートを用いて評価を行った。より高い倍率での提案手法の検証は、今後行っていきたい。

6. おわりに

本研究では 3 軸慣性センサの信号に対して、ニューラルネットワークを用いてアップサンプリングを行う手法を提案し、2 種類のデータセットを用いて提案手法が従来手法を大きく上回ることを確認した。これに加えて、提案手法を適用した信号を用いてセンサデバイスの姿勢推定とオブジェクトベース行動認識を行い、実際のアプリケーションにおいても提案手法を用いた方が高い性能となることを確認した。今後の予定として、スマートフォンに組み込まれている深層学習用のフレームワークなどを用いて提案手法を実行することなどがあげられる。

謝辞 本研究の一部は JST CREST JPMJCR15E2, JSPS 科研費 JP16H06539, JP 17H04679 の助成を受けて行われたものです。

参考文献

- [1] Bao, L. and Intille, S.S.: Activity recognition from user-annotated acceleration data, *Pervasive 2004*, pp.1–17 (2004).
- [2] Breiman, L.: Random forests, *Machine Learning*, Vol.45, No.1, pp.5–32 (2001).
- [3] Dong, C., Loy, C.C., He, K. and Tang, X.: Image super-resolution using deep convolutional networks, *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol.38, No.2, pp.295–307 (2016).
- [4] He, K., Zhang, X., Ren, S. and Sun, J.: Identity mappings in deep residual networks, *European Conference on Computer Vision*, pp.630–645 (2016).
- [5] Houben, S. and Marquardt, N.: Watchconnect: A toolkit for prototyping smartwatch-centric cross-device applications, *the 33rd Annual ACM Conference on Human Factors in Computing Systems*, pp.1247–1256 (2015).
- [6] Ioffe, S. and Szegedy, C.: Batch normalization: Accelerating deep network training by reducing internal covariate shift, arXiv preprint arXiv:1502.03167 (2015).
- [7] Kingma, D. and Ba, J.: Adam: A method for stochastic optimization, arXiv preprint arXiv:1412.6980 (2014).
- [8] Korpela, J., Takase, K., Hirashima, T., Maekawa, T., Eberle, J., Chakraborty, D. and Aberer, K.: An energy-aware method for the joint recognition of activities and gestures using wearable sensors, *International Symposium on Wearable Computers*, pp.101–108 (2015).
- [9] Kuleshov, V., Enam, S.Z. and Ermon, S.: Audio super resolution using neural networks, *International Conference on Learning Representations Workshops* (2017).
- [10] Laput, G., Xiao, R. and Harrison, C.: Viband: High-fidelity bio-acoustic sensing using commodity smartwatch accelerometers, *the 29th Annual Symposium on User Interface Software and Technology*, pp.321–333 (2016).
- [11] Ledig, C., Theis, L., Huszar, F., Caballero, J., Cunningham, A., Acosta, A., Aitken, A., Tejani, A., Totz, J., Wang, Z. and Shi, W.: Photo-realistic single image super-resolution using a generative adversarial network, arXiv preprint arXiv (2016).
- [12] Leppänen, J. and Eronen, A.: Accelerometer-based activity recognition on a mobile phone using cepstral features and quantized GMMs, *2013 IEEE International*

- Conference on Acoustics, Speech and Signal Processing*, pp.3487–3491 (2013).
- [13] Li, K. and Lee, C.: A deep neural network approach to speech bandwidth expansion, *2015 IEEE International Conference on Acoustics, Speech and Signal Processing*, pp.4395–4399 (2015).
- [14] Ljung, L.: Asymptotic behavior of the extended Kalman filter as a parameter estimator for linear systems, *IEEE Trans. Automatic Control*, Vol.24, No.1, pp.36–50 (1979).
- [15] Maekawa, T., Kishino, Y., Yanagisawa, Y. and Sakurai, Y.: Recognizing handheld electrical device usage with hand-worn coil of wire, *Pervasive 2012*, pp.234–252 (2012).
- [16] Maekawa, T. and Watanabe, S.: Unsupervised activity recognition with user’s physical characteristics data, *International Symposium on Wearable Computers*, pp.89–96 (2011).
- [17] Maekawa, T., Yanagisawa, Y., Kishino, Y., Ishiguro, K., Kamei, K., Sakurai, Y. and Okadome, T.: Object-based activity recognition with heterogeneous sensors on wrist, *Pervasive 2010*, pp.246–264 (2010).
- [18] Malik, S. and Laszlo, J.: Visual touchpad: A two-handed gestural input device, *the 6th International Conference on Multimodal Interfaces*, pp.289–296 (2004).
- [19] Ravi, N., Dandekar, N., Mysore, P. and Littman, M.: Activity recognition from accelerometer data, *IAAI 2005*, Vol.20, No.3, pp.1541–1546 (2005).
- [20] Shirai, A., Geslin, E. and Richir, S.: WiiMedia: Motion analysis methods and applications using a consumer video game controller, *the 2007 ACM SIGGRAPH Symposium on Video Games*, pp.133–140 (2007).
- [21] Srivastava, N., Hinton, G.E., Krizhevsky, A., Sutskever, I. and Salakhutdinov, R.: Dropout: A simple way to prevent neural networks from overfitting, *Journal of Machine Learning Research*, Vol.15, No.1, pp.1929–1958 (2014).
- [22] Yun, X. and Bachmann, E.R.: Design, implementation, and experimental results of a quaternion-based Kalman filter for human body motion tracking, *IEEE Trans. Robotics*, Vol.22, No.6, pp.1216–1227 (2006).
- [23] Zagoruyko, S. and Komodakis, N.: Wide residual networks, arXiv preprint arXiv:1605.07146 (2016).



吉村 直也 (正会員)

2018年大阪大学工学部電子情報工学科卒業。ユビキタスコンピューティング、ウェアラブルセンシングに関する研究に従事。



前川 卓也 (正会員)

2006年大阪大学大学院情報科学研究科博士後期課程修了。2006年NTTコミュニケーション科学基礎研究所入所。2012年より大阪大学大学院情報科学研究科マルチメディア工学専攻准教授、現在に至る。ユビキタスコンピューティング、ウェアラブルセンシングに関する研究に従事。2010年度情報処理学会山下記念研究賞、2013年度日本データベース学会上林奨励賞、2015年電気学会優秀論文発表A賞、2019年IPSJ/IEEE Computer Society Young Computer Researcher Award等受賞。IEEE, ACM, 電気学会, 日本神経科学会, 日本データベース学会の各会員。



天方 大地 (正会員)

2012年大阪大学工学部電子情報工学科卒業。2014年同大学大学院情報科学研究科博士前期課程修了。2015年同大学院情報科学研究科博士後期課程修了後、同年同大学院情報科学研究科マルチメディア工学専攻助教となり、現在に至る。情報科学博士。データベース、ネットワーク環境におけるデータ検索技術に関する研究に従事。IEEE, ACM, 日本データベース学会の各会員。



原 隆浩 (正会員)

1995年大阪大学工学部情報システム工学科卒業。1997年同大学大学院工学研究科博士前期課程修了。同年同大学院工学研究科博士後期課程中退後、同大学院工学研究科助手、2002年同大学院情報科学研究科助手、2004年同大学院情報科学研究科准教授。2015年より同大学院情報科学研究科教授となり、現在に至る。工学博士。2003年本学会研究開発奨励賞受賞。2008年、2009年本学会論文賞、2015年日本学術振興会賞受賞。2017年大阪科学賞受賞。ソーシャルコンピューティング、ネットワーク環境におけるデータ管理技術に関する研究に従事。IEEE, ACM, 電子情報通信学会の各会員。