rPPG信号に基づく個人識別攻撃の提案と対策

飯島 涼[1,2,a)]　長谷川 幸己[2]　河岡 諒[2]　森 達哉[2,b)]

概要：人物が映る動画から，脈波 (PPG) 信号を遠隔で取得する remote PPG (rPPG) が，ヘルスケア技術の一つとして注目されている．本研究では，rPPG を用いて動画から脈波信号を推定し，個人を特定する攻撃を提案する．具体的には，目線やモザイク処理など，プライバシー加工技術を施した動画から rPPG によって脈波信号を推定し，ユーザから直接センサで得た脈波や，加工されていない動画から得た rPPG による信号と照合することで，個人を特定する攻撃を評価する．CNN-LSTM モデルを用いた評価の結果として，UBFC-Phys データセットにおいて，PPG 波形と照合した場合は最高 97%，平均 86%，加工前の動画から得られた rPPG 信号を照合する攻撃結果は最高で 99%，平均して 87%となり，特定のユーザを標的とするプライバシー攻撃の脅威となりうることを示した．さらに，緑色の特徴量に着目した対策手法を提案・実装し，提案対策手法が rPPG 推定技術の妨害に成功することを示した．

キーワード：rPPG，動画加工技術，ヘルスケア，深層学習，プライバシー，ユーザの目に優しい対策

Identification attacks based on rPPG and its countermeasures

Ryo Iijima[1,2,a)]　Koki Hasegawa[2]　Ryo Kawaoka[2]　Tatsuya Mori[2,b)]

Abstract: Remote photoplethysmography (rPPG), which enables the remote acquisition of PPG from videos, has gained attention as a promising healthcare technology. We propose an attack method that identifies individuals by estimating pulse signals from videos using rPPG. Specifically, we assess the effectiveness of this attack by estimating PPG signals from videos with privacy-preserving techniques such as gaze obfuscation and mosaic processing applied. These estimated signals are then matched against pulse signals obtained directly from sensors on the user or rPPG signals derived from unprocessed videos. The evaluation results, using the UBFC-Phys dataset, demonstrate a maximum identification accuracy of 97% (average: 86%) when matching with PPG waveforms, and a maximum of 99% (average: 87%) when matching with rPPG signals from unprocessed videos. Additionally, we propose and implement a countermeasure focusing on green channel features, demonstrating its effectiveness in disrupting rPPG signal estimation.

Keywords: rPPG, video processing, healthcare, deep learning, privacy, eye-friendly countermeasures

1. はじめに

Society 5.0 の実現に向けて，ヘルスケアデータの活用が注目されている．経団連が 2018 年から発表している「Society 5.0 時代のヘルスケア」[9] では，ヘルスケア領域におけるデジタルトランスフォーメーション (DX) 推進，ICT 技術の活用に合わせたオンライン診療や，医療機関でのデータ共有など，医療データを利活用しながら，多くの人に医療・ヘルスケアサービスを届け，Well-being を実現するための仕組みづくりが進められている．ヘルスケアデータのうち代表的なものとして，脈波 (Photoplethymography, PPG) は，センサを肌付近につけることによって直接測定され，脈拍の計測や脈波パターンの解析による疲労等の解析を行うことができる [2]．スマートウォッチやヘルスケア用 IoT 機器等では，LED を腕等に照射して得られた反射光を解析して PPG 計測を行う光学式センサが一般的となっている [3]．

[1] 国立研究開発法人 産業技術総合研究所
[2] 早稲田大学
[a)] ryo.iijima@aist.go.jp
[b)] mori@nsl.cs.waseda.ac.jp

PPG の測定に有用な技術の一つとして，人物の動画から，センサを使わず遠隔で PPG 信号を推定する技術 (remote Photoplethysmography, rPPG) [1] が注目されている．rPPG は，動画に映る肌から，緑色の変化を統計解析することにより，脈波を推定し，脈波のピーク値をもとに脈拍数を推定する (図 1)．カメラ等を学校や介護施設に導入することによって，脈波パターンの異常，または脈波パターンによって心理状態や精神的な健康，疲労等を常時監視できるサービスとして，注目を浴びている．

rPPG に関する既存のセキュリティ研究について，Lin[6] らは，PPG ベースの認証方式に対して，rPPG 信号を用いて攻撃を行う手法を提案している．具体的には，ユーザから得られた PPG 信号によってトレーニングされた認証モデルを，動画から得られた推定 PPG 波形 (以後，rPPG 信号) によって攻撃する．PPG 信号は従来センサによって直接得るものと考えられていたが，入力として動画さえあれば，PPG 信号の推定値を得られる．

rPPG 信号が個人認証システムを攻撃するための技術として用いられることの裏を返せば，動画から得られる rPPG 信号を用いて，個人を特定するプライバシー攻撃として使われうることを示唆している．本研究では，動画から取得した rPPG 信号を用いて，個人を特定する攻撃を提案する．具体的には，目線やモザイク等の加工がされた動画から取得した rPPG 信号を，特定のユーザとすでに紐づけられている PPG 信号，または加工なし動画から得られた rPPG 信号と照合することによって，映っている人物がだれであるかを特定する攻撃を実施する．rPPG 技術の概要と，プライバシー脅威の例を図 1 に，具体的な脅威例を図 2 に示す．

本研究では，リサーチクエスチョン (RQ) として

RQ 1:PPG 信号-rPPG 信号間，または動画加工前後の rPPG 信号間でどの程度波形が類似しているか？ (4 章)

RQ 2:RQ 1 で明らかにした類似性が，個人を特定する攻撃に対してどの程度脅威となりうるか？ (5 章)

RQ 3:本攻撃に対してどのような対策手法が有効であるか？ (6.1 節)

の 3 つの観点で研究を実施する．

まず，攻撃・対策に必要な基礎知識や脅威モデルを整理し (2 章)，rPPG の取得方法や使用するデータセット，解析手法についてまとめた (3 章)．結果として，攻撃動画に目線やモザイク等の加工をつけた場合，動画の取得場所やユーザの動作が異なる場合でも波形間距離を基準に類似度が高い波形が得られることが明らかとなり (4 章)，その結果を実際に CNN-LSTM による判定モデルで検証したところ，目線加工を行った場合に平均 F1 値 84%，最高 F1 値 99.4%，モザイク加工を行った場合に平均 F1 値 80%，最高 F1 値 98.7%の精度で特定の個人を標的にした標的型攻撃が成功してしまうことを示した (5 章)．さらに，フレー

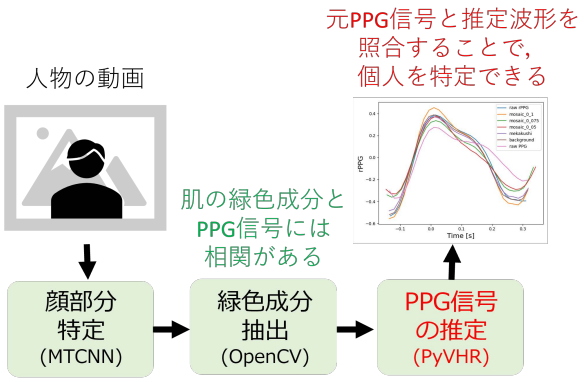

図 1　rPPG 推定技術の概要とプライバシー脅威例

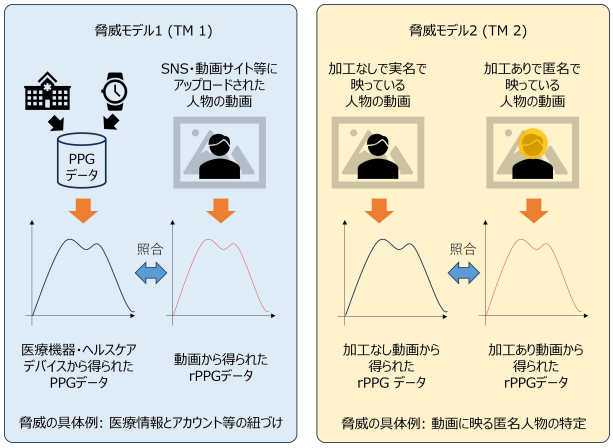

図 2　脅威モデルのイメージ図 (詳細は 2.3 章で説明)

ムごとにランダムな緑色のノイズを入れることで rPPG 推定の波形推定を妨害する対策を提案し，実際に対策として rPPG 推定技術の妨害を実現可能であることを示した (6 章)．

2. 背景知識

2.1 remote PPG (Photoplethymography) / BVP (Blood Volume Pulse)

PPG (Photoplethymography) とは，血流の流れによって生じる変化を測定する技術である [2]．PPG によって取得した脈波信号 (以下 PPG 信号) は，通常センサを肌に密着，または肌から数 cm 以内に設置し，被験者から直接取得する必要がある．スマートウォッチやヘルスケア用 IoT 機器等では LED を腕等に照射して得られた反射光を解析して PPG 計測を行う光学式センサが一般的となっている．PPG によって得られる波形の 1 サイクル分の概形と，よく用いられる特徴量を図 3 に示す．PPG 信号に関する詳細と認証システムへのセキュリティ攻撃については，[3] に詳しくまとめられている．

rPPG (remote Photoplethymography) は，従来センサを直接肌につけることによって取得していた PPG 信号を，動画から推定する技術である．rPPG 技術の概要を，図 1 に示す．rPPG 技術は，動画に映る人の顔の位置を

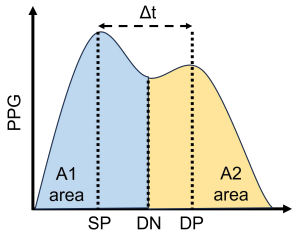

図 3 PPG 信号の概形と，代表的な特徴量の例

MTCNN [8] 等のアルゴリズムによって特定した後，顔の緑色成分を抽出して統計解析することにより，血流の強さを推定する．このアルゴリズムは，血流の変化量と，顔の緑色成分が相関することを前提として成り立っている．体に当たった環境光が反射する際，皮膚内部から反射する拡散反射によって，映像に映る皮膚表面の画素値を微小に変化させ色の違いとして表れる．緑色である理由は，血液中の酸素化ヘモグロビンの吸光特性が，波長約 540 nm（緑）にピークを持つためである [2].

具体的な rPPG アルゴリズムの搭載されたソフトウェアの例として，pyVHR [1] がある．PPG 技術，または rPPG 技術によって得られた血流の変化波形 (信号) を，BVP (Blood Volume Pulse) 信号と呼ぶ．BVP と呼ぶ場合，PPG 技術によって取得されたものか，rPPG 技術によって取得されたものかを判別できないため，本論文では，PPG 技術によって取得された BVP 信号を PPG 信号，rPPG 技術によって取得された BVP 信号を rPPG 信号と呼ぶことにする．rPPG は本来，脈拍数の推定など，BPM を測定する目的で利用されるケースが多く，rPPG 信号が直接使われるケースは少ないが，その取得方法はオープンソースであり [1]，内部構造を熟知していれば攻撃者が波形を取得することができてしまう．実際のセキュリティ攻撃研究事例として [6] がある．

2.2 動画に映る人物のプライバシー加工技術

動画に映る人物が，自分が映っていることを秘匿したい場合，簡易的なプライバシー加工技術として，目の部分に黒い箱を描画する 目線加工，顔部分を特定してモザイクを施すモザイク加工がある．適用される具体例として，テレビや YouTube の動画，SNS 上の動画などで，主に登場する人物以外に映り込んだ人物にモザイクをかけるケースがある．また，自身の顔を隠したままテレビや動画等に出演するインターネット発のタレント・歌手等がいる．

2.3 想定される脅威モデル

2.3.1 利用するデータの種類による分類

想定される脅威モデルのイメージを図 2 に示す．想定される脅威モデルとして，TM 1: 医療用に取得されていた PPG データが流出し，ほかの場所に保存されている動画から rPPG 信号を取得して照合することにより，医療データと個人を紐づける攻撃，TM 2: 無加工の動画から得られた rPPG 信号と，目線・モザイク等の加工を施した動画から得られた rPPG 信号を比較し，加工されている動画に映る人物を特定する攻撃の 2 種類が考えられる．TM 1 の場合は，PPG センサとカメラセンサで異なるデータとして取得されているにも関わらず，波形の類似性を解析することで，医療データと動画で異なる種類のデータ間の紐づけが行われてしまうことを意味し，TM 2 は，本来結び付けられることを避けるために実施した 2.2 章のような動画加工後も，元の動画や，ほかのプラットフォーム上にある加工なし動画と結び付けられてしまう可能性がある．例えば，YouTube 上で肌を出して映っているタレントが，ほかのテレビ番組や動画サイト上でモザイクや目線付きで映っていたにも関わらず，誰であるか特定されてしまう，未成年で逮捕された人物が，名前が公開されていないにも関わらず，ニュース上のモザイク付動画と Instagram 等の無加工動画の rPPG 信号を比較することにより，本来守られるべきプライバシーが守られなくなってしまうという可能性が考えられる．

2.3.2 攻撃者の目的による分類

上記のように個人を特定する場合，攻撃者の目的として (1) 得たデータが特定の個人のものであるかを知りたい場合 (2) 得たデータが誰のものであるかを知りたい場合 の 2 種類が考えられる． (1) の場合には，モデルとして，特定の人物であるか否かのみを判定するモデルを作成するのに対し，(2) の場合は，データセット内で誰であるかを示すラベル付きデータを作成したうえでモデルを学習する必要がある．(1) の場合を特定の人物を狙う標的型攻撃，(2) の場合は得られたデータの所持者を特定する特定攻撃と呼ぶことにする．

3. 攻撃・評価手法

3.1 動画の加工方法

一般的に個人のプライバシーを守る手法として使われているモザイク・目線隠しの加工を行うための動画加工方法を示す．さらに，本研究では，環境の違いによる影響を明らかにするため，環境光の違い，場所の違いによる評価を実施するための仮想的な動画加工を実施している．

はじめに，すべての動画加工に共通する要素について説明する．動画加工は，手法をホワイトボックス化するため，動画アプリ等は使用せず，すべて Python 上で実装した．動画は連続するフレームからなっており，フレーム 1 つ 1 つを取得して，画像加工を行うのが一般的な動画加工手法となっている．動画を読み込んだ後，フレームを取得し，加工，保存用動画ファイルに書き込む処理を 1 フレームずつ実施するプログラムを構築した．動画加工の方法を以下にそれぞれ示す．かっこ書きでそれぞれの加工方法の

結果を図示する際の凡例の示し方をアルファベットで示している．

3.1.1 プライバシー保護を目的とした加工

モザイク加工 (mosaic_α): OpenCV の resize 関数を用いて，係数 α 分だけ対象領域を縮小した後，拡大しなおす処理を実施することにより，対象領域のピクセルの値が均一化される．α の値は，0–1 までの範囲であり，小さいほどモザイクの度合いが強くなる．顔部分のみを加工するために，haar-like 検出器による顔抽出を実施した．顔検出ができなかった場合に備え，顔検出ができなかった場合は，前のモザイク画像をそのまま用いて適用することで適用されていない区間が生じるのを防ぐ．α の値は，均等間隔で $\alpha = 0.05, 0.075, 0.1$ の 3 通りを検討した．検討した α の値ごとのフレーム例を図 4 に示す．$\alpha > 0.1$ の場合は，目視で誰であるか人間でも認知できてしまう可能性が高いため，上限は 0.1 とし，$\alpha = 0.025$ の場合は，肉眼・haar-like 検出器の双方で顔の各パーツを認識できないレベルでモザイクがかかってしまうことから，下限の値を選択している．

目線加工 (eyemask): 目の部分を特定して目線をつける処理を自動化するため，MTCNN ライブラリを用いて目の位置を特定し，目の付近に黒い四角を描画する．上記の場合と同様に，目の位置が検出できなかった場合は，前の四角をそのまま用いて加工結果とする．

3.1.2 環境の違いを再現することを目的とした動画加工

輝度変化 (V_整数): 動画や画像のうち，部屋の明るさに直接関係する輝度は，HSV カラーのうち，Value (V, 輝度) に相当する．RGB フォーマットでは輝度の変更ができないため，OpenCV を用いて動画の色を RGB 基準から HSV 基準に変更した後，V の値を変更することにより，輝度変更を実現した．ライトの明るさに対応する輝度変化として，元動画から ±20 が一般的であるため，$-20, -10, 10, 20$ の 4 種類での rPPG 値を比較した．

背景除去 (Background): rPPG 信号は主に顔部分から波形を推定する技術であるが，顔付近に映る背景等が rPPG 信号の推定結果に影響を与える可能性がある．その可能性を検証するため，背景を除去したうえでの rPPG 結果を比較する．

3.2 UBFC-Phys Dataset [7]

本研究では，2.3 章に示した脅威を検証するため，UBFC-Phys Dataset [7] を使用する．データセットは，被験者ごとにセンサによって直接取得した PPG 信号，および PPG 信号取得時に撮影したユーザの動画からなる．PPG 取得時の被験者の行動 (Task) は，T1: 静止状態，T2: スピーチタスク，T3: 計算タスクの 3 種類からなる．本研究では，[6] 等を参考にし，データセットに含まれる先頭 10 人 (User ID: s1–s10) を対象にして評価を実施した．

図 4 プライバシー保護を目的とした加工の例．左上: モザイク $\alpha = 0.1$, 右上: モザイク $\alpha = 0.075$, 左下: モザイク $\alpha = 0.05$, 右下: 目線

3.3 rPPG 波形取得

本節では, 動画から PPG 値を推測するために, pyVHR [1] を用いる．pyVHR からは, BVPs として値が得られる．一連の波形として出力されるため，1 つずつの rPPG 波形とするために次の節から説明する前処理を実施する．

3.4 前処理

波形を 1 つずつ扱うための前処理として，フィルタによるノイズ除去，PPG 波形としてのピーク検出，ピーク前後の波形切り出しがある．特に，PPG 波形は，図 3 に示すように，ピークが 2 種類あるため，2 つのピークが混同しないように前処理を実施する必要がある．そのようなニーズに適したライブラリとして，生体信号処理に適した neurokit2 を用いて，PPG 波形の切り出しまでの処理を実施している．生体信号に共通する，詳細の前処理の手順は [5] に示している．

3.5 評価手法

攻撃を評価する観点として， 2.3 節で述べたように，TM1: センサ値から得られる PPG 信号と，遠隔で得られる rPPG 信号の比較による攻撃，TM2: 加工前の動画から得られる rPPG 信号と，加工後の動画から得られる rPPG 信号の比較による攻撃が考えられる．評価手法として，DTW による波形間の距離算出 + ROC によって距離ベースの精度を示す方法と (4 章)，深層学習により最終的な判定結果を得る方法 (5 章) の 2 種類を考える．

DTW (Data Time Warping)

DTW とは，2 つの波形長が異なる場合に，波形間の距離を求める方法として適したアルゴリズムである．PPG は，脈拍に対応することから，個人ごとに波形長が異なることがある．本研究で実施した 3.4 節の前処理においても，最終的に得られる波形長は個人ごとに異なるため，その場合でも波形の伸縮を行わずに距離の評価ができるよう，本研究では，DTW を距離アルゴリズムとして採用した．

CNN-LSTM

攻撃者が波形の類似性を用いて攻撃を行う場合，距離の評価に加えて，「実際にどの程度攻撃が成立するのか」を判定モデルを用いて検討することは重要である．本研究では， CNN-LSTM モデル [4] を判定モデルとして用いて，攻撃がどの程度成功するかを実際のモデルによって検証する．CNN-LSTM は，生体信号等の波形学習に適した時系列構造として LSTM 層が組み込まれており，特徴抽出に CNN の畳み込み層を用いる．CNN-LSTM は，PPG 信号を用いた認証システムの最新先行研究において高い精度を出しており [4]，個人の特徴を抽出するのに長けているといえる．本研究では，[4] で用いられているのと同様の構造の CNN-LSTM を実装し，映っている人物が特定の人物であるかどうか，または誰であるかを特定に用いるためのモデルとして利用する．

4. 基礎評価

本章では，**RQ1: PPG 信号-rPPG 信号間，または動画加工前後の rPPG 信号間でどの程度波形が類似しているか？** を検証することを目的とし，個人特定につかわれる PPG または rPPG 波形が，実際にどの程度異なるのかを数値やグラフによって示す．具体的には，データセットに含まれる PPG 信号と，pyVHR によって得られる rPPG 信号の比較 (脅威モデル TM 1)，または動画の加工前後の rPPG 信号の比較 (脅威モデル TM 2) を実施した後，動画の加工の種類による rPPG 信号の変化，環境の違いによる rPPG 信号の変化を，ユーザごとの波形の中央値を取ることによって可視化し，さらに波形の違いを波形間距離として DTW により数値化することで，類似度を調べる．

4.1 個人ごとの平均 PPG 波形 / rPPG 波形

PPG 波形は個人ごとに異なる特徴を持ち， [4] で示されているように個人認証への利用が可能である．その実態を可視化するため，UBFC-Phys データセット内に含まれる PPG 信号と，UBFC-Phys データセット内の動画から取得した rPPG 信号を 3.4 章で説明した手法にもとづいて前処理を実施する．1 波形ずつに分離した後，ユーザごとに PPG 波形, rPPG 波形の中央値をとった結果をそれぞれ図 5 左，図 5 右に示す．図 5 左にある PPG 信号の形状を比べると，目視でもそれぞれの波形の概形が異なっていることがわかる．図 5 右の rPPG 信号も同様に個人間で差が表れているが，User 8, 9 のように，全体の波形が図 3 にあるような PPG の概形から大きく離れた波形が得られているケースがある．再度取得してもこのような波形が変わらず得られることから，rPPG 推定攻撃を受けづらい特性である可能性がある．このような個人的な差異に関する議論は 6 章で述べる．

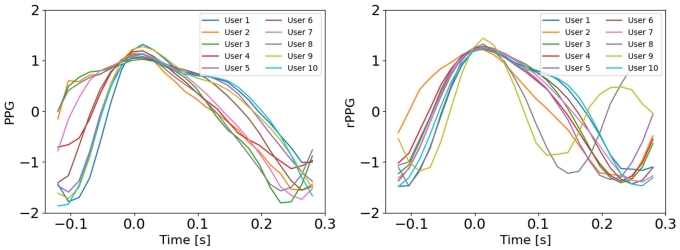

図 5 左: ユーザごとの中央値 PPG 波形の違い．右: ユーザごとの中央値 rPPG 波形の違い．

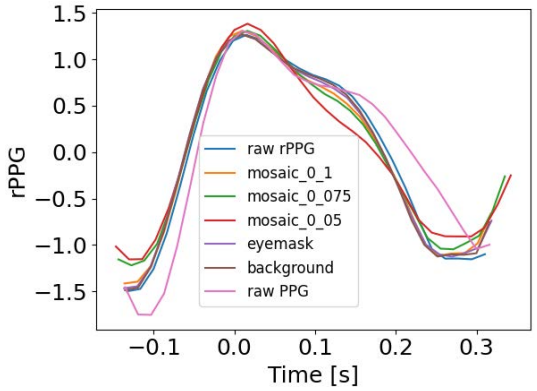

図 6 センサから得られる PPG 信号の中央値波形と，各種動画から得られる rPPG 信号の中央値波形 (User 1)．raw PPG は，データセット内に含まれていた生波形．raw rPPG は，データセット内の動画からそのまま抽出した rPPG 波形を示す．その他凡例ごとの解説は 3.1 節に示す．

4.2 動画にプライバシー保護対策を施した場合の rPPG 波形変化

PPG 信号と，各加工を施した動画に対する rPPG 信号の波形の中央値を描画した結果を視覚的に示すため，代表して User 1 の場合を図 6 に示す．視覚的に見て，個人ごとの差を示した図 5 と比べて，波形差が少なく見える．全ユーザも含めた具体的な距離の分布と考察は，図 8，および 4.5 節に示す．

4.3 映るユーザの動作が異なる場合の評価

本節では，ユーザの動作の違いが rPPG 波形に与える影響を調査する．代表的な動作として，T1: 静止，T2: 会話タスク，T3: 計算タスクの 3 種類を実施した動画について，それぞれ rPPG 信号を取得して，どの程度差が生じるかを明らかにする．それぞれのタスクごとの波形例を図 7 左に示す．T2 は口を動かす動作を含み，肌以外の部分が動くため，0.3 s より後から波形が異なっているが，0.3 s 以前は同じ波形の概形が得られていることが分かる．

4.4 ユーザの映る環境が異なる場合の評価

本節では，ユーザの映る環境が異なる動画同士で比較する場合について評価を実施する．具体的には，ユーザの動作が同じで背景を取り除いた場合の評価，および，環

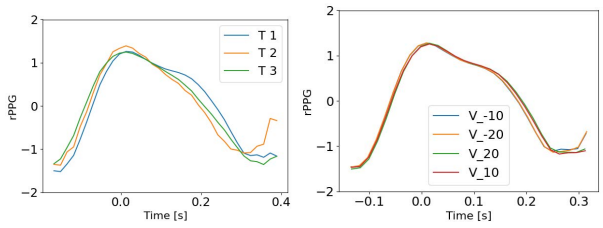

図 7 左: 動作が異なる場合の rPPG 中央値波形の違い 右: 環境が異なる場合の rPPG 中央値波形の違い

境内の輝度が異なる場合の評価を実施する．データセットの動画から背景を取り除いた場合の評価結果を図 6 の background ラベル波形として，輝度変化の結果を図 7 右に示す．いずれの場合にも，環境の変化は顔領域に影響を与えず，同様の rPPG 波形が得られていることが分かる．実際にライトの明るさを変えたら，動画の加工による結果とは異なるのではないかという疑問に答えるために，著者のうち 1 人を対象に，配信用に用いられる LED リングライトを用いて追加調査を実施した．リングライトのカラー設定を，上記の輝度設定と同程度の cold-3, cold-10, warm-3, warm-10 の 4 つの設定で調査を行った結果，波形の形状に変化は見られず，図 7 右と同様の結果が得られた．

4.5 それぞれの波形に関する距離の分布

本章のまとめとして，それぞれの加工あり動画ごとの rPPG 波形と，加工なし動画の rPPG 波形の波形間距離を，全ユーザの結果を統合してまとめる．図 8 に，加工なし動画の rPPG 波形のテンプレートと，各ユーザの動画加工後の rPPG 波形のテンプレートの距離を DTW によって取得し，箱ひげ図で示す．一番左の raw PPG は，TM 1 の検証用・比較用として，ユーザから直接得られた PPG 値との距離差を掲載している．モザイクの例のみに着眼すると，a の値が小さくなるごとに，テンプレートから遠ざかり情報量が少なくなっている一方で，目線 (eyemask) のように肌が露出している場合には，加工後であってもほとんど元の rPPG 波形と同様の波形が得られ，波形の類似性が高いことが分かる．同様にして，背景は顔の色情報に影響せず，環境の明度が変化した場合も，一定に色情報が変化するだけで，血液と緑色成分との相関には影響がないため，距離が極めて近くなってしまうことが分かる．

ここで，個人ごとの距離差の傾向を見るため，ユーザごとに，raw rPPG と比較した場合の rPPG 信号の距離の逆数をベースにスコア化し，ROC カーブを描いた結果を図 9 に示す．ROC カーブでは，左上に線が位置するほど，本人と他人を見分けやすい (つまり攻撃が成功しやすい) ことを示している．脅威モデル TM 2 のような状況下で，どの程度識別結果を精度良く得られるかが，ROC カーブを見ることでわかる．ROC の結果で特徴的な点は，各ユーザの ROC を平均した結果を見ると，線が中心に近く，DTW

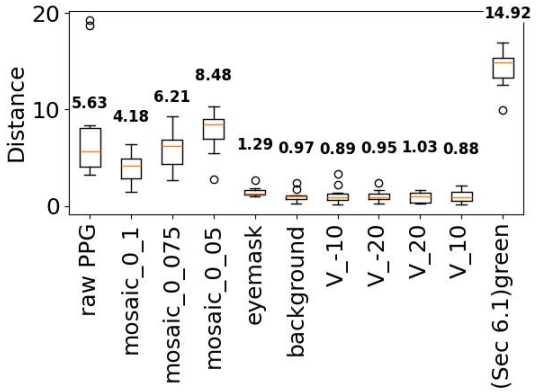

図 8 raw rPPG の中央値波形テンプレートと比較した各加工動画から得られた rPPG 中央値波形の距離の分布．距離が遠いほど波形が類似していないことを意味し，個人特定が実現しづらい．各箱ひげ図の上に，分布の中央値を示している．1 番右は，著者らが 6.1 節で提案した対策手法の結果．

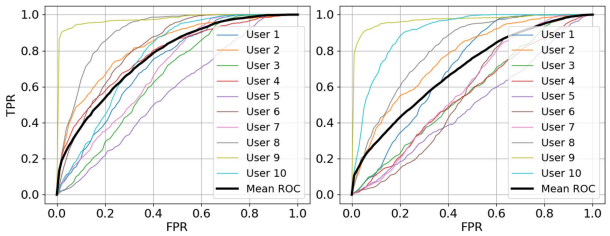

図 9 raw rPPG 中央値波形をテンプレートとしたときの動画加工後の距離をスコア化して ROC にしたもの．左: 目線付加した動画の rPPG, 右: モザイク付加した場合の rPPG ($\alpha = 0.1$)

距離による予測はそこまで精度高くできないように見えるが，個人ごとの線に着目すると，User 9 のように，目線・モザイクの双方をつけても個人の特徴が消えないケース，モザイクの場合には特徴が残るが，目線を付けた場合には特徴が消えるケース (User 10), どちらの加工によっても同様に対策がなされているケース (User 5) があることが分かる．モザイク加工の場合など，α の値を固定することで，平等に設定した値に応じて特徴が消えそうなイメージがあるが，実際には，図 9 に示した通り，加工した後にどの程度特徴が残るかという点でユーザに個人差があることが分かる．このような個人差に対する議論は， 6 章で述べることにする．

5. 攻撃評価

本章では，**RQ 2: RQ 1** (4 章) で明らかにした類似性が，個人を特定する攻撃に対してどの程度脅威となりうるか？を明らかにするため， 4 章で示した波形の違いが，実際にどの程度攻撃につながりうるのかを，判定モデルによって評価する．2.3.2 節に示したように，(1) 特定の誰かであるかどうかを判定する標的型攻撃, (2) 得られたデータが誰であるかを判定する特定攻撃 の 2 種類を検討する．

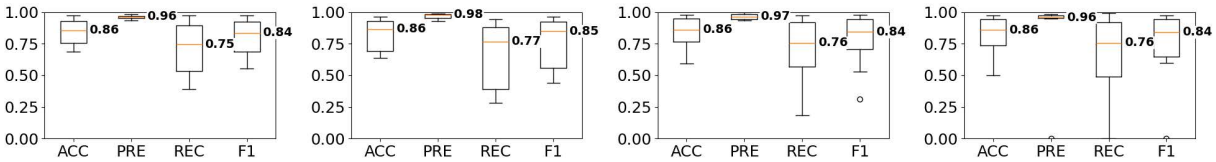

図 10 脅威モデル TM 1 における CNN-LSTM の評価結果．左から目線付加，モザイク加工 (左から $\alpha = 0.1, 0.075, 0.05$)

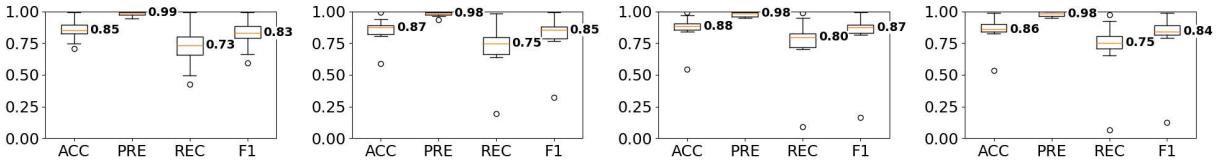

図 11 脅威モデル TM 2 における CNN-LSTM の評価結果．左から目線付加，モザイク加工 (左から $\alpha = 0.1, 0.075, 0.05$)

5.1 特定の誰かであるかどうかを判定する標的型攻撃

モザイク等で加工された動画に映るユーザが特定の人物であるかどうかを知りたい場合の標的型攻撃について評価する．この場合，モデルは攻撃者の目的に合わせ，特定の人物であるか否かのみを出力するものとして作成し，評価を行う．本評価では，最先端の認証モデルとして提案されている CNN-LSTM モデル [4] による方法で判定を行い，結果を評価する．脅威モデル TM 1 における CNN-LSTM での判定結果を図 10，脅威モデル TM 2 における CNN-LSTM での判定結果を図 11 に示す．脅威モデルごとの結果を比較すると，TM 2 のほうが攻撃成功率は 0.01–0.02 程度高くなり，結果のばらつきが少ないが，中央値としては TM 1, TM 2 で同様の結果が得られていることが分かる．さらに，トレードオフを考慮した F1 score によれば中央値で 0.84–0.87 となっているが，1.00 に近い点まで上のひげが伸びていることから， 4 章でも見られたように，攻撃された場合の成功率に幅があることが分かる．下の外れ値の場合，精度が落ちており，目線やモザイクで攻撃を防ぐという観点で問題ないが，上の外れ値については，目線・モザイクによる対策後も攻撃が高確率で成功してしまうことを示しており，本攻撃に対する追加の対策が必要である．

5.2 得られたデータが誰のものであるかを特定する攻撃

さらにデータが誰のものであるかの判定を CNN-LSTM モデルで行った場合の結果を表 1 に示す．再現率は高いが，適合率が極端に低くなっており，結果として F1 値も低い値となっている．誰のものであるかを判定する部分ではそれほど精度よく個人を特定することができないが，再現率が高いモデルが攻撃者にとって有利になる状況では危険であり，なおかつ一度でもプライバシー情報が紐づけられてしまうと PPG 信号は変更ができないため，攻撃がより成立しづらい状況を作るための対策が必要である．rPPG 推定による攻撃を防ぐことを目的として，実際の対策を検証した結果は，6.1 節に示す．

表 1 動画加工後の攻撃結果 特定攻撃の場合

Processing	ACC	PRE	REC	F1
目線付加	0.757	0.275	0.979	0.430
モザイク ($\alpha = 0.1$)	0.790	0.370	0.960	0.534
モザイク ($\alpha = 0.075$)	0.733	0.369	0.948	0.531
モザイク ($\alpha = 0.05$)	0.792	0.286	0.981	0.442

6. 議論

6.1 対策手法

本研究で議論した攻撃手法が成立する背景には，顔の緑色成分と PPG 信号に相関が生まれてしまうことにある．対策手法として，フレームごとにランダムな緑色のノイズを足すことが考えられる．

実際に検証するため，実験で用いた User1–10 の動画を対象に，RGB カラーの Green 値に対して，[-50, 50] の整数ランダム値をフレームごとに足した動画を作成し，rPPG 信号を取得した場合の波形と，攻撃の成功率が低下するかを調べた．緑色を足した後に取得した各ユーザの rPPG 波形の概形を図 12 に，元波形と比較した距離を図 8 の一番右に示す．図 12 と図 3 は概形がかけ離れており，raw rPPG 波形と比較した DTW 距離の中央値は 14.92 と，図 8 に掲載したほかの距離と比較しても波形差が大きくなっていることが分かる．このことから，本対策を適用して，距離差が大きいものを閾値ベースではじくだけでも，対策は可能であることが分かる．整数ランダム値の範囲を調整することで，対策の強度を変更できる．図 12 の対策後波形の個人差が少ないことから，本対策により 4.5 節で示した攻撃成功率の個人差が解消されることがわかる．

本対策手法により，rPPG 信号の推定を防ぐことは可能であるが，もし対策済み動画を人が閲覧する可能性がある場合，緑色に点滅するように感じられるため，そのようなケースでは人の目に優しい対策手法を考える必要がある．実装後は，6.3 節で示す研究倫理・健康への配慮をしたう

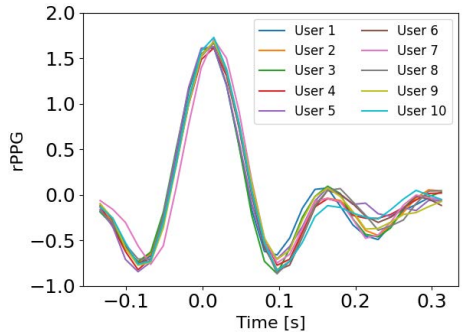

図 12 提案対策手法による加工後の動画から得られた rPPG 中央値波形．PPG の概形 (図 3) から大きく異なる波形が得られ，ユーザ間の差異も図 5 と比較して小さくなっていることが分かる．

えで，ユーザスタディによって対策後の動画が見る人物にとって負担ではないかを検討する必要がある．そのほか，RGB 値の G 値に関してノイズを加算する場合の，ノイズ強度とプライバシー保護の関係について，差分プライバシーの観点から議論することが重要である．

6.2 顔以外の肌から rPPG 信号を推定される脅威

本研究の制約事項として，rPPG 信号の取得を行う皮膚の領域は顔部分に限定しているが，肌から得られる緑色成分であれば，顔でなくても rPPG 信号を取得できる可能性がある．顔以外の部分から取得できることが分かった場合，顔以外の領域にも加工を行わなければ，手の映り込み等も個人特定に利用される危険性がある．そのようなほかの体の領域から取得する rPPG 信号との関連，および調査結果の議論については今後の課題とする．

6.3 研究倫理

本研究では，生体信号を用いた個人特定攻撃を実施している．本研究の目的は，攻撃の PoC を示すことで脅威の可能性を示して警鐘を鳴らし．対策手法を示すことにある．扱うデータセットについては事前に利用規約を確認したうえで使用し，プライバシー評価に用いることが問題ないことを確認したうえで実験を実施している．一部予備実験として実施したユーザスタディについては，早稲田大学「人を対象とする研究に関する倫理規定」に則り，実験参加者への侵襲・介入は行わず，参加者の健康や視覚への影響に配慮して慎重に設計した．参加者には，データの利用用途や管理方法，参加者本人に利害関係が生じないこと，個人を特定しうるデータの内容，公開する際の条件等を説明し，参加する本人の同意を得た上で実験を行っている．

関連する議論として，人を対象とする研究に関する研究倫理の文書内では，参加者のプライバシーを守るという趣旨から，動画データの公開に関して，倫理委員会から提供された誓約書サンプル内に「顔部分や眼部などを消去する，ぼかすなど個人の特定不可能な状態に限る」という意思表示用選択項目がある．本研究により，ただ単に目線を入れる，またはモザイクをかける対策では，rPPG 推定技術など最新の技術により特定可能である可能性が示されたため，どのような対策を実施することで特定不可能となるかを議論し，各大学等に設置される倫理委員会にて，誓約書の文書をプライバシーの観点から更新する必要がある．同様にして，個人情報保護に関する法律や文書に関しても，本研究を踏まえて，たまたま撮影に映り込む人物の加工をどのようにすることで識別不可能となるのかに関する最新の議論を行う必要がある．

7. おわりに

本研究では，ヘルスケア技術の 1 つである rPPG が，個人の特定に用いられる可能性について着目し，評価と対策技術の提案を行った．動画像の加工前，加工後で類似した rPPG 信号が得られることから，個人を特定しうる可能性について CNN-LSTM モデルによって評価し，平均 F1 値が 84%，最高 F1 値が 99.4%となり，ユーザによってモザイク・目線等の対策効果に大きな差があることを明らかにした．また，rPPG 信号の推定を妨害する対策手法を提案し，効果があることを実証した．本研究で示したような，一見すると生体信号と無関係に見える要素が，個人特定に用いられる可能性に配慮して対策を行うことの重要性について議論した．

謝辞 本研究の一部は JSPS 科研費 22K17890 の助成を受けたものです．

参考文献

[1] Boccignone, G. et al.: pyVHR: a Python framework for remote photoplethysmography, *PeerJ Computer Science*, Vol. 8, p. e929 (2022).
[2] Castaneda, D. et al.: A review on wearable photoplethysmography sensors and their potential future applications in health care, *International Journal of Biosensors & Bioelectronics*, Vol. 4, No. 4, pp. 195–202 (2018).
[3] Hinatsu, S. et al.: Evaluation of PPG Feature Values Toward Biometric Authentication Against Presentation Attacks, *IEEE Access*, Vol. 10, pp. 41352–41361 (2022).
[4] Hwang, D. Y. et al.: Evaluation of the Time Stability and Uniqueness in PPG-Based Biometric System, *IEEE Transactions on Information Forensics and Security*, Vol. 16, pp. 116–130 (2021).
[5] Iijima, R., Takehisa, T., Ohki, T. and Mori, T.: The Catcher in the Eye: Recognizing Users by their Blinks, *ACM Asia CCS*, p. 1739–1752 (2024).
[6] Li, L., Chen, C., Pan, L., Zhang, J. and Xiang, Y.: Video is All You Need: Attacking PPG-based Biometric Authentication, *ACM AISec*, p. 57–66 (2022).
[7] Meziati, R., Benezeth, Y., De Oliveira, P., Chappé, J. and Yang, F.: UBFC-Phys (2021).
[8] Zhang, K., Zhang, Z., Li, Z. and Qiao, Y.: Joint Face Detection and Alignment Using Multitask Cascaded Convolutional Networks, *IEEE Signal Processing Letters*, Vol. 23, No. 10, pp. 1499–1503 (2016).
[9] 日本経済団体連合会：Society 5.0 時代のヘルスケア IV ヘルスケアデータの価値最大化に向けて (2023).