コネクテッド・ヒューマンによる空間状況把握プラットフォーム

天野 辰哉[1] 山口 弘純[1] 東野 輝夫[1]

概要：近い将来，AR・MR デバイスの高度化や小型軽量化により，それらのデバイスを常時装着するヒトが，スマートフォンなどの従来のモバイル端末では困難であった可視光や赤外線・電波などを利用した常時の 3 次元空間センシング能力を有するようになる．従来のモバイル端末の性能を大幅に超えた高度な空間認識・リアルタイムな相互通信・ネットワーク接続を有するヒトを我々は**コネクテッド・ヒューマン**（CH）と定義する．CH はインフラが存在しない環境においても空間の動的コンテキストを取得できるため，それを正しく共有すれば防犯や交通安全支援，高度なモバイルクラウドセンシングなど様々なサービス実現が期待される．本研究では CH のコンテキストの Person-to-Person の活用を想定し，CH が取得した，スマートフォンを有する周辺人物（Non-CH）の周辺コンテキストを，当該 Non-CH のみとセキュアに共有するためのプラットフォーム設計を行うとともに，CH と Non-CH 間のユーザ特定・認証手法を提案する．提案手法では，Non-CH が保持するスマートフォンを CH が映像で捉え，その姿勢推定を行うとともに，Non-CH は時刻やおおよその位置，ならびに自身の内蔵センサーから得られるスマートフォン姿勢データを用いてユーザ特定および認証を行う手法を提案している．CH 向けに画像からのスマートフォン姿勢推定システム，ならびに Non-CH 向けにスマートフォンの姿勢推定アプリを実装し，それらを用いた特定および認証実験を現実的なシナリオのもと行った結果，実験のうちの約 74％ の状況下で正解率 100％ でユーザの識別が可能であることが確かめられた．

1. はじめに

あらゆるモノがネットワークに接続され現実環境のデータをリアルタイムに収集可能にするとともに，モノがネットワークを介して制御・管理される，いわゆるモノのインターネット（IoT）の概念が実現しつつある．加えて，ヒトもまたスマートフォンなどのモバイル端末を介してネットワークに接続されるようになり，遠隔地のデバイスの情報を取得したり，端末に搭載されたセンサを活用し，ユーザの位置や行動をネットワークを介して集約することで，リアルタイムに空間状況把握したりする取り組みもなされている [1].

これに対し，ごく近い将来には，拡張現実（AR）や複合現実（MR）デバイスの高度化や小型軽量化により，それらのデバイスを常時装着するヒトが，スマートフォンなどの従来のモバイル端末では困難であった可視光や赤外線・電波などを利用した常時の 3 次元空間センシング能力を有するようになり，さらに高速・大容量通信を実現する 5G や 6G 通信の登場によって，ネットワークを介したそれらの情報の共有が可能となる．そのような AR・MR 端末を介することで，従来のモバイル端末の性能を大幅に超えた高度な空間認識・リアルタイムな相互通信・ネットワーク接続を有するヒトを我々は**コネクテッド・ヒューマン**（CH）と定義する．CH の概念が実現されれば，ヒトの位置・行動，コミュニケーションと，IoT デバイス・CH から得られる環境内の情報を統合することで空間状況をより深く理解し，人間の社会活動や生活を支援することが可能となる．例えば，周辺のヒトとモノの位置や状態などを CH が把握し，共有することで，既設センサーが存在しない環境における 3 次元空間理解（いわゆるダイナミックマップの構成）が実現でき，物品管理，防犯，異常事象検知，交通理解，モバイルクラウドセンシングの高度化など多様なサービスへの応用が期待される．

加えて，CH は，従来のスマートフォンのみを保持したいわゆる非コネクテッド・ヒューマン（Non-CH）に対する情報提供者にもなり得ると考える．例えば，行動や周辺状況を CH により理解された Non-CH が存在したとすれば，CH が理解・獲得したコンテキストをその Non-CH に対してセキュアに共有することで，Non-CH の安全のための周辺モニタリングや異常検知，ソーシャルディスタンスの通知など，Person-to-Person でのダイナミックコンテキ

[1] 大阪大学大学院情報科学研究科

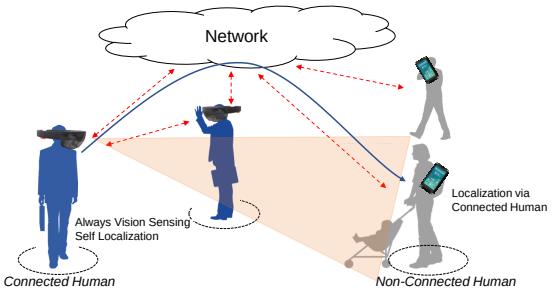


図 1　コネクテッド・ヒューマンによる非コネクテッド・ヒューマンのセンシング補助

ストデータの情報流通が可能となり，その報酬による経済効果なども期待できる．新しい社会システムの一翼を担う可能性もある．

その実現に向け，本研究では，コネクテッドヒューマン（CH）が取得した，スマートフォンを有する周辺人物（Non-CH）の周辺コンテキストを，クラウドを介してその Non-CH のスマートフォンのみとセキュアに共有するためのプラットフォーム設計を行い，CH と Non-CH 間のユーザ特定・認証手法を提案する．以下では，CH と Non-CH を合わせてユーザと呼び，全ユーザには電子的な固有 ID（ユーザ ID）が割り当てられているとする．ユーザ間では信頼できるサーバを介した通信や Wi-Fi Direct のような直接通信で通信路を確立可能であるとし，ユーザ ID を特定することで該当するユーザにデータを送信可能と仮定する．また，CH は自身のカメラ映像などで常時周辺状況を把握するものとし，近隣ユーザの空間上の位置などのコンテキストを特定可能とする．これは，例えば Yolo [2] などのオブジェクト認識技術で映像フレーム内のユーザを検出し，各ユーザまでの相対距離や方向などから位置が特定可能であることを意味する．このもとで，CH が空間上で特定したあるユーザに対してのみデータ送信を行うためには，当該ユーザの ID 特定ならびにそのユーザ ID が確かにそのユーザのものであること（真正性）の確認行為であるユーザ認証 [3] が必要となる．例えば，カメラと高精度なサーモグラフィを備えた CH が，あるユーザの発熱を検出し，その事実を当該ユーザのみに送信するシナリオなどでは，誤ったユーザへのメッセージ送信はフレームワークの信頼性を損なうため，ユーザ認証は重要な機能となる．

カメラなどの視覚デバイスで捉えられたユーザ認証では，顔など現実空間における当該ユーザの特徴を利用することが一般的であるが，顔認証システムは，被認証ユーザが自身の顔情報と ID を事前に共有する必要があるため，本研究のような用途には適さない．歩容認証など行動生体情報に基づくユーザ認証も利用されているものの，認証時にユーザの行動の様子を一定時間観察する必要があり，かつ顔認証と同様に，行動生体情報を事前に共有する必要がある．また，被認証ユーザの保持端末が自身の ID を Wi-Fi や BLE 等で発信し，受信者がその電波強度による近接情報や推定到来方向などに基づいて，空間上のユーザの ID の特定・識別を行う方法 [4] も考えられるが，電波を用いた識別は周辺歩行者や環境変化の影響を受けやすく，また被認証ユーザが密集している場合にはその識別が困難である．これに対し，本研究では，Non-CH が保持するスマートフォンを CH が映像で捉え，その姿勢推定を行うとともに，被認証ユーザは時刻，おおよその位置ならびに自身の内蔵センサーから得られるスマートフォン姿勢データからユーザ認証を行う手法を提案している．これをもとに，ユーザ識別・認証手法およびそれを用いた CH と Non-CH 間のセキュアな情報共有システムを設計している．CH 向けに画像からのスマートフォン姿勢推定システム，ならびに Non-CH 向けにスマートフォンの姿勢推定アプリを実装し，それらを用いた特定および認証実験を現実的なシナリオのもと行った結果，実験のうちの約 74 % の状況下で正解率 100 % でユーザの識別が可能であることが確かめられた．

2.　関連研究

2.1　コネクテッド・ビークル

LiDAR やカメラなどの高性能な車載センサを搭載し，車車間通信や路車間通信などにより自車両や他車両，および周辺環境の共有や通知機能を備えたコネクテッド・ビークルの普及が現実味を帯びてきている．これに対し，各コネクテッド・ビークルから収集される車両相対位置情報を統合し，周辺認識に利用するためのフレームワーク [5] や，LiDAR により検出した近接車両への相対距離と方向の情報を，車車間通信によって周辺車両と共有するプロトコル [6] など，コネクテッド・ビークルが捉えたリッチコンテキストを，従来のレガシー車両との混在環境下や，コネクテッドビークル間でも活用する手法が提案されてきている．近年では自律運転車両によるセンシング機能により移動体を含む動的共有マップ（ダイナミックマップ）を構成するコンセプトが実用化に向けて動き出している．我々の提案するコネクテッド・ヒューマンは，このようなコネクテッド・ビークルの概念に着想を得ている．コネクテッドヒューマンでは周辺環境が車両よりも多様であり，高度な認識技術を要したり，提案手法のようにプライバシーに配慮した共有プラットフォーム設計が必要とされる一方，多様なコンテキストを活用した応用システムが期待できる．

2.2　外部センサによる歩行者の識別・認証

歩行者の保持・装着するセンサによって捉えられる軌跡や行動情報と，LiDAR やカメラなどの外部に設置された固定型環境センサによって捉えられるそれらの情報のマッチングによって，歩行者の特定や識別・認証を行うシステムが提案されている．

文献 [7] は LiDAR とウェアラブル RFID タグの組み合

わせによる歩行者マッチング手法を提案している．LiDAR を用いて歩行者の位置を追跡し，RFID タグを持った歩行者が環境内に配置された RFID リーダーの近くを通過した際に，移動軌跡と RFID タグとの関連付けを行う．歩行者との移動軌跡を高精度に一致させるためには，環境内に多数の RFID リーダを配置する必要があり，インフラの設置コストが課題となっている．文献 [1] では LiDAR により推定される歩行者軌跡とモバイル端末に搭載されているモーションセンサの計測情報とのマッチングによって端末保持者を特定する手法を提案している．[8] も同様に LiDAR による群衆の移動軌跡を用いて，その群衆内の歩行者がモバイル端末のカメラで撮影した映像内の人物の位置特定手法を提案している．

特に近年では，Wi-Fi や RFID などの電波によって空間内の人間の行動を認識するワイヤレスセンシングを用いたユーザの識別や認証を行うシステムが注目されている．

2.3 提案手法の位置づけ

前述のように，コネクテッドビークルにおいては認識対象となる周辺コンテキストは車両や歩行者の存在情報など，現状では交通システムにおける安全のためのデータに限られており，周辺の全車両や歩行者で共有することが全体のユーティリティ向上につながる．一方でコネクテッドヒューマンは人間の周辺コンテキストといったよりパーソナル空間における情報を扱う可能性があり，またなりすましや嫌がらせに対する対策も必要となる．また ID 特定や認証における技術的なチャレンジとして，CH および Non-CH も移動しており，長時間観察を続けて認証情報を取得するといった仮定はおきにくい．したがって，なるべく短い時間でとらえた外観特徴量を用いて当事者間のみがセキュアに通信路を確立できるプラットフォームを設計する必要がある．これに対し，本研究では被認証ユーザのスマートフォンの外観から得られる特徴量を用いた認証手法を提案しており，想定するシナリオにおいてどの程度の認証能力があるかを実証により確認している．

3. システム概要

提案する CH と Non-CH 間での情報共有システムのアーキテクチャを図 2 に示す．CH はスマートグラスやグラス型 AR デバイスを装着した人間であり，それらのデバイスからカメラ映像を常時取得する．Non-CH は一般的な従来のスマートフォンを保持した人間であり，そのモバイル端末には専用のアプリがインストールされている．環境内には複数人の CH と Non-CH が存在し，各自のデバイスを介してシステムのサーバに接続されている．各デバイスはその装着者・保持者を生体認証や知識認証により認証済みであるとする．またデバイスとサーバ機器同士は Transport Layer Security (TLS) におけるサーバ認証・クライアント

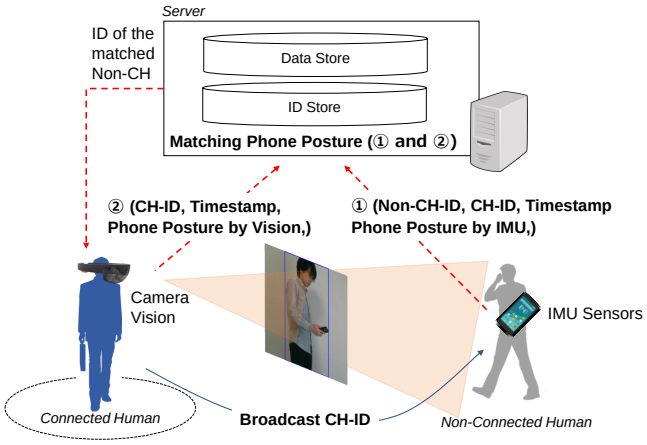


図 2 システムアーキテクチャ

認証などを利用して相互に認証済みであり，その通信路は暗号化されているものとする．CH，Non-CH のもつ固有のユーザ ID をそれぞれ CH-ID，Non-CH-ID と呼び，これらの ID を特定することによって，該当するユーザへサーバを介して任意のメッセージが送信可能であるとする．

ユーザ間の情報共有は (1) CH から他ユーザへのメッセージ送信と (2) Non-CH から他ユーザへのメッセージ送信の 2 種類に分類される．(1) CH から他ユーザへのメッセージ送信は以下の流れで行われる．CH はカメラ映像中からメッセージ送信先のユーザを選択する．この段階では相手のユーザ ID が不明であるため，CH はメッセージを送信できない．そこで CH はカメラ映像中のユーザの何らかの特徴を用いて，そのユーザの ID を識別・認証する．これによって CH はそのユーザにメッセージを送信可能となる．逆に，(2) Non-CH から他ユーザへのメッセージ送信においては，そもそも Non-CH は特定のユーザを空間上で見つける手段を持たないため，相手のユーザ ID は既知である状況しか存在しない．したがって，CH から受信した CH-ID や会話や他の経路によって得たユーザ ID を用いる，あるいは全ユーザへのブロードキャストによって Non-CH はメッセージを他ユーザに送信する．以降ではユーザ ID の識別および認証が不可欠となる CH から Non-CH へのメッセージ送信に焦点を当てて説明する．

3.1 脅威モデル

本研究では，CH から Non-CH へのメッセージ送信システムにおいて，以下のような，セキュリティ上の脅威である攻撃者 $\mathcal{A}$ による被害者 $\mathcal{V}$ への攻撃を想定する．

- Attack-I なりすまし：$\mathcal{A}$ は Non-CH であり，$\mathcal{A}$ が他の Non-CH であるになりすますことで，ある CH ($\mathcal{V}$) からなりすまされた Non-CH へ送信されたメッセージを閲覧する．
- Attack-II メッセージ傍受・改ざん：$\mathcal{A}$ は第 3 者であり，中間者攻撃により通信路を傍受し，サーバ・ユー

ザ間の全てのメッセージを閲覧あるいは改ざんする．

- Attack-III スパム：$\mathcal{A}$ は CH であり，特定の Non-CH（$\mathcal{V}$）に対してメッセージを大量送信することにより，サーバや CH と比較して貧弱である，Non-CH の端末リソースを消耗する．
- Attack-IV サービス妨害：Non-CH が網羅的に認証情報を送りつけることで，サーバ上に他の Non-CH の認証情報と類似する認証情報を増やし，CH（$\mathcal{V}$）による特定の Non-CH の認証を不可能にする．

以降では上記のの脅威を踏まえた上での CH や Non-CH 間での情報共有方法について説明する．環境上に存在する CH を $\mathrm{CH}_i(i = 1, 2, ...N)$，Non-CH を $\mathrm{NCH}_j(j = 1, 2, ...M)$ とする．CH_i から NCH_j へのメッセージ送信は (1) CH_i による NCH_j のユーザ ID の取得，(2) メッセージの送信，(3) メッセージの受信の 3 ステップから構成される．

3.2 ユーザ ID の取得（認証）

メッセージの送信者である CH_i は自身のユーザ ID である ID_i を，Wi-Fi や BLE などの手段を用いて高い頻度で周辺の全てのユーザに通知する．CH-ID_i を受信した NCH_j は自身のスマートフォンの端末姿勢を端末の IMU センサによって計測する．CH からの ID_i の受信に従って ID が ID_j である Non-CH が計測試行 k により取得した端末姿勢を $O_{i,j,k}$ と表記し，その姿勢を取得した時刻を $T_{i,j,k}$ とする．端末姿勢の値については 4.2 節で詳細に述べる．NCH_j は CH_i からのメッセージの受信を許可する場合，ID_i が受信できなくなるまで，極めて高い頻度で自身の端末姿勢を計測し，$\langle O_{i,j,k}, T_{i,j,k}\rangle$ をサーバに送信する．ID_i の受信に従って NCH_j がサーバに送信した端末姿勢の計測 k の集合を $K_{i,j}$ で表す．サーバは Non-CH から受信したこれらの情報をストレージに記録する．

一方，CH_i はカメラ映像中からデータを送信したい相手である Non-CH を選択し，カメラ映像中のあるフレームにおいて，その Non-CH が保持するスマートフォンの端末姿勢を推定する．使用したフレームの撮影時刻を $\hat{T}$ とし，推定した端末姿勢を $\hat{O}_i$ とする．CH_i は該当する Non-CH のユーザ ID のリクエストとして，$\langle \hat{O}_i, \hat{T}\rangle$ をサーバに送信する．

サーバは送信された情報を基に以降の章で説明する端末姿勢の類似性に基づくユーザ ID の特定手法により，CH のカメラ映像中で指定された Non-CH のユーザ ID を特定し，CH_i に送信する．ユーザ ID の特定は以下の手順で行われる．サーバは $\langle \hat{O}_i, \hat{T}\rangle$ を受信すると，CH が観測した端末姿勢に最も近い端末姿勢である Non-CH を見つける．まずリクエストを送信した CH の i と $\hat{T}$ を用いて式（1）に従い，CH が端末姿勢を取得した時刻 $\hat{T}$ に最も近い時刻に各 Non-CH で行われた計測を j ごとに求める．この行われた計測のうち最小の時刻誤差であるものを k_j とおく．さらに，すべての j について計測 k_j の時刻誤差が閾値 TH_t 以下であるであるような Non-CHj の集合を J とする．TH_t はシステムのパラメータであり，同一の端末姿勢を取得する際の CH での画像撮影時刻と Non-CH での計測時刻の許容時刻誤差を示す．

$$J = \{j \mid \hat{T} - T_{i,j,k_j} \leq \mathrm{TH}_t \wedge K_{i,j} \neq \emptyset\}_{j=1}^{M} \quad (1)$$

ただし

$\mathrm{k}_j = \arg\min_{k \in K_{i,j}} |\hat{T} - T_{i,j,k}|$

関数 $f(o_a, o_b)$ を 2 つの端末姿勢 o_a と o_b の類似度を求める関数であるとし，式（2）により得られる j によって定まる ID_j を送信相手の Non-CH のユーザ ID として取得する．J が空集合の場合は，該当する Non-CH が存在しなかったことを示す．また推定端末姿勢と計測端末姿勢の類似度が閾値 TH_o より大きかった場合，つまり式（2）により得られる j について $f(\hat{O}_i, O_{I,j,k_j}) \leq \mathrm{TH}_o$ を満たさない場合も，CH がカメラ映像中で選択した Non-CH に該当する Non-CH のユーザ ID が見つからなかったとする．TH_o は端末姿勢が一致しているとみなせる最大の類似度を示すシステムのパラメータである．

$$\arg\max_{j \in J} f(\hat{O}_I, O_{I,j,k_j}) \quad (2)$$

CH によるカメラ映像からの端末姿勢 $\hat{O}_i$ の推定については 4.2 節で，端末姿勢の類似度の計算については 4.3 節で述べる．

端末姿勢に基づいて CH は Non-CH をサーバ上で認証するため，Attack-I を実行するためには，$\mathcal{A}$ は $\mathcal{V}$ になりすます必要があり，そのためにはある時刻における端末姿勢を取得し，認証情報とする必要がある．$\mathcal{A}$ は $\mathcal{V}$ に物理的に接近し，その端末姿勢を模倣することが可能であるが，サーバのストレージ上で $O_{i,j_1,k} \simeq O_{i,j_2,k}$ となるデータの存在を確認することによって，このような模倣攻撃はサーバ上で容易に検出可能である．このようなデータが存在する状況を「認証情報が衝突した」と呼ぶ．これは Attack-I が発生した場合か，偶然 CH に接近している二人以上の Non-CH が同様の端末姿勢でスマーフォンを使用した場合に発生する．認証情報が衝突した場合，$\mathcal{V}_1$ に対して $\mathcal{A}$ の存在の可能性を警告するとともに，$\mathcal{V}_2$ に対してメッセージの送信をキャンセルさせることで，Attack-I による $\mathcal{V}$ の機密性の侵害を防ぐことができる．

3.3 メッセージの送信

CH_i は特定した NCH_j の ID を用い，サーバを介して互いの公開鍵を送信しあう．以降，CH_i による NCH_j へのメッセージ送信では，CH_i の秘密鍵を用いたデジタル署名を付与するとともに，メッセージ内容をすべて NCH_j の公開鍵を用いて暗号化する．CH_i が送信したメッセージは

サーバのストレージ上に保管される．

お互いの秘密鍵が他の方法により漏洩しない限りは，Attack-II の第 3 者によるメッセージの傍受は実現不可能であり，またメッセージの改ざんは，デジタル署名により受信した NCH_j により検出可能である．

3.4 メッセージの受信

NCH_j は定期的にサーバに対してメッセージ受信リクエストを送信し，自身あてのメッセージがサーバのストレージに存在するかどうかをチェックする．存在する場合は，未受信のものをすべて受信し，自身の秘密鍵を用いてメッセージを復号するとともに，取得済みの送信者の公開鍵を用いてメッセージの署名を検証する．

このように NCH_j のリクエストをトリガとする受信処理によって，Attack-III による高頻度のメッセージ受信を防止する．また NCH_j は CH がブロードキャストする ID_i に対して端末姿勢 $O_{i,j,t_{\mathrm{nch}}}$ をサーバに送信しないことによって，CH_i が NCH_j の ID を特定不可能にすることで，特定の CH からのメッセージをブロックすることが可能である．

4. 手法詳細

本システムでは，ある瞬間に端末の X 軸，Y 軸，Z 軸がそれぞれ重力方向となす角を順に θ_x, θ_y, θ_z としたときの端末姿勢 O を $O = \langle \theta_x, \theta_y, \theta_z \rangle$ と定義する．なお端末の軸は図 3 下部に示すように，端末背面からスクリーン面に向かうスクリーン平面に対して垂直な軸を Z 軸，端末の下端から上端へ向かう軸を Y 軸，右手系での Z 軸ベクトルと Y 軸ベクトルの外積の方向を X 軸とする．以降では Non-CH による自身の端末姿勢取得，CH によるカメラ画像からの端末姿勢推定およびそれらの端末姿勢の類似度計算について順に説明する．

4.1 自端末の姿勢取得

Non-CH は加速度センサを用いて自身のスマートフォンの端末姿勢，つまり重力方向に対する端末の向きを取得する．加速度センサにより得られる端末の X，Y，Z 軸への重力加速度成分をそれぞれ g_x，g_y，g_z とし，重力加速度のノルムを G で表記すると，各軸が重力加速度となす角度を求めることで端末姿勢 O は $O = \langle \cos^{-1}\frac{g_x}{G}, \cos^{-1}\frac{g_y}{G}, \cos^{-1}\frac{g_z}{G} \rangle$ で求められる．

4.2 カメラ画像からの端末姿勢推定

本節では CH が撮影したカメラ映像のあるフレームにおける Non-CH の保持するスマートフォンの端末姿勢を推定する手法について説明する．

手法の入力は Non-CH がスマートフォンを保持する 1 枚の画像であり，出力はその画像内に含まれる各 Non-CH のスマートフォンの端末姿勢である．端末姿勢推定は (1) 端末検出，(2) 端末および手の 3 次元メッシュ生成, (3) 端末の軸の推定，(4) 端末における重力方向の推定となす角の計算 の 4 つのステップで行われる．手法の流れを図 3 に示す．

まず入力された画像内における Non-CH が持つスマートフォンの領域（バウンディングボックス）を Yolo-v3 [2] を用いて検出する．人とスマートフォンをデータとして含む COCO データセット [9] により学習済みの Yolo-v3 モデルを利用し，まず画像内の人の領域を検出し，検出された領域ごとにその領域の画像を切り出す．切り出された人画像に対してさらに Yolo-v3 を適用し画像内のスマートフォンの領域を同様に切り出す．

次に切り出されたスマートフォンの画像領域をもとに，スマートフォンの形状およびそれを保持する Non-CH の手，を図 3 に示すように 3 次元のメッシュとして再構成する．この再構成においては人の手・指先の構造とその保持する物体の形状を含む大規模なデータセットである ObMan [10] と ObMan の製作者により公開されている学習されたモデル [11] を利用する．これにより入力画像における手首の位置を原点とする 3 次元空間上の手のメッシュ構造および物体のそれぞれのメッシュ構造が取得できる．

生成された 3 次元の手と物体それぞれのメッシュ構造による物体形状から，物体，つまりスマートフォンの X，Y，Z 軸を推定する（図 4）．スマートフォンのメッシュを構成する 3 次元の点群に対して，最小二乗法による平面検出を適用し，得られた平面の法線ベクトル方向を Z 軸とする．スマートフォンを保持する人はスマートフォンの背面を握っているものとし，手からスマートフォンへ向かう方向を Z 軸の正方向とする．また検出された平面とスマートフォンのメッシュが交差する点を，平面上の 2 次元の点群として抽出し，それら点群に対して最小二乗法による直線検出を適用し，得られた直線の方向を Y 軸，平面上での直線の法線方向を X 軸とする．X 軸，Y 軸の向きはすでに正方向が確定している Z 軸を基準に右手系座標系が成立する方向にする．

最後に生成された物体のメッシュに対する重力方向を推定する．物体のメッシュは 3 次元空間上において，入力画像が構成する平面が，Y-Z 平面と平行になるような向きで生成され，入力画像上で上方向は 3 次元座標上で Y の正方向であり，下方向は Y の負の方向になる．したがってカメラの仰角（Elevation）ϕ およびロール（Roll）λ がともに 0 である場合は，生成された 3 次元座標系において Y の負の方向が重力方向となる．

4.3 端末姿勢の類似度

2 つの端末姿勢 $O_a = \langle \theta_{a,x}, \theta_{a,y}, \theta_{a,z} \rangle$ と $O_b = \langle \theta_{b,x}, \theta_{b,y}, \theta_{b,z} \rangle$ の類似度 $f(O_a, O_b)$ は式（3）のように重み付きの各軸の誤差の和の逆数として定義する．W_x，W_y，

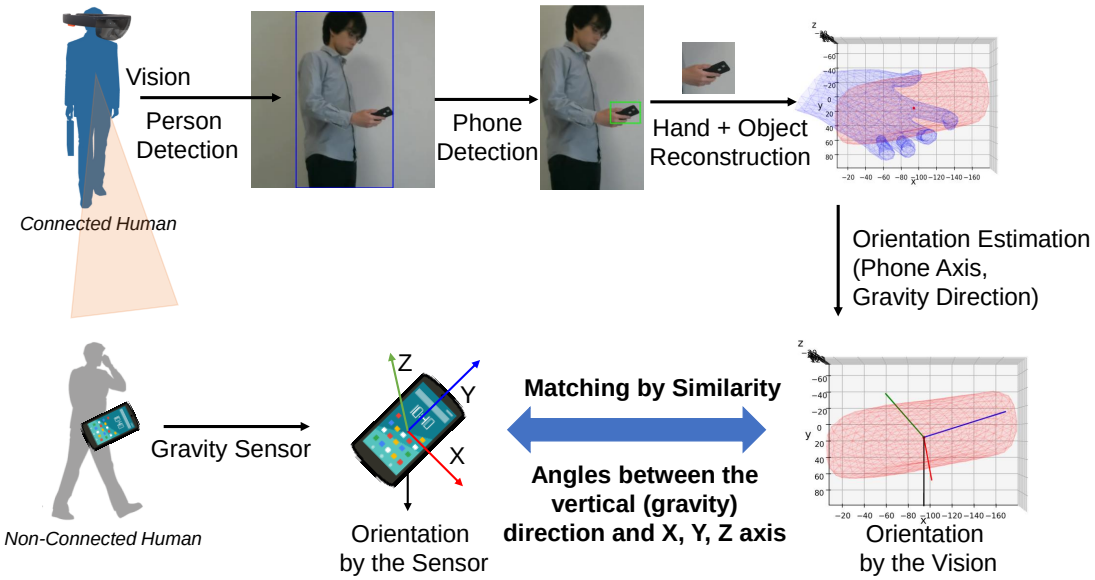


図 3　センサデータとカメラ映像による端末姿勢のマッチング

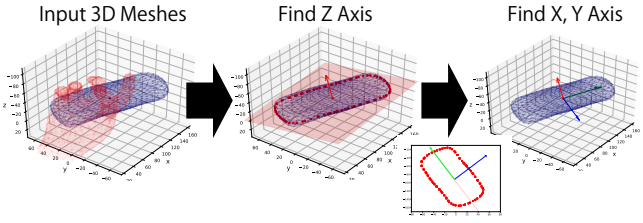


図 4　物体のメッシュ構造からの端末軸方向推定

W_z は各軸の誤差に対する重みである．類似度計算においては，角度の値はすべて度数法による 0 から 180 の値を使用する．

$$f(O_a, O_b) = \frac{1}{W_x(|\theta_{a,x} - \theta_{a,y}|) + W_y(|\theta_{a,x} - \theta_{a,y}|) + W_z(|\theta_{a,x} - \theta_{a,y}|)} \quad (3)$$

ただし

$$W_x + W_y + W_z = 1$$

5. 実験と評価

提案システムの性能を評価実験を行った．実験では CH として HD 解像度（1280x720）の Web カメラを接続したノート PC を用いた．Non-CH となるスマーフォンの保持者は，CH の Web カメラにスマートフォンとそれを握る手が写り，かつカメラの前約 1〜2m の範囲で，椅子に座るあるいは立っている状態を維持した．Non-CH のスマートフォンは，10 ms おきにセンサによる端末姿勢を計測してサーバに送信し，CH は約 1 秒に 1 回の頻度でカメラ映像中のスマートフォンの姿勢を推定し，推定値をサーバに送信した．以降の評価においては特に明記しない限り，パラメータ TH_t，TH_o はそれぞれ 50 ms，20 と設定した．これは CH と Non-CH 間での時刻誤差 50 ms までを許容し，端末姿勢の類似度の閾値を 20 に，つまり，軸ごとの重みを無視すればおよそ各軸ごとの角度差が± 20 度未満であれば端末姿勢は一致しているとみなすことを意味する．

5.1 カメラ映像からの端末姿勢推定誤差

提案手法によるカメラ映像からのスマートフォンの端末姿勢推定精度について評価を行った．Google Pixel 3（5.5 インチ，145.6 x 68.2 x 7.9mm）と Lenovo Phab 2 Pro（6.4 インチ，179.88 x 88.57 x 10.7mm）の 2 種類のスマートフォンを様々な端末姿勢で保持する画像を，各スマートフォンごとに 180 枚取得し，各画像ごとに画像からの端末姿勢と端末の加速度センサによる端末姿勢を求め，各軸ごとの角度誤差，つまり重力方向とのなす角の差を求めた．結果を図 5 に示す．誤差は度数法の値である．

絶対角度誤差の平均は Pixel 3 では X 軸，Y 軸，Z 軸がそれぞれ 18.3，13.4，10.0 度であり，Phab 2 Pro ではそれぞれ平均 14.3，11.4，9.3 度であった．どちらの機種においても X 軸が重力方向とのなす角の推定誤差が他の軸と比較して大きくなる傾向が見られた．カメラによって最も回転量を捉えやすいのは X 軸周りの回転量であるとともに，端末の X 軸と重力がなす角度に影響を与えるのは，端末の Y 軸と Z 軸での回転量である，つまり X 軸の回転量はその値に影響しないためである・したがって，カメラから推定される各軸が重力となす角度のうち X 軸のものが最も誤差が大きくなる．Pixel 3 と比較して Phab 2 Pro の誤差が大きいのも，Pixel 3 の方が端末サイズが小さく，重力に対する回転量を映像でより捉えにくいためである．

この結果から端末姿勢の類似度計算における各軸の重み W_x, W_y, W_z は機種ごとに各軸の推定精度に比例するように設定することが望ましいと考えられる．以降

の実験では Pixel 3 を用いており，その場合の重みは上記の各軸ごとの平均角度誤差の逆数を基に，$W_x = 18.3/(18.3+13.4+10.0)$, $W_y = 13.4/(18.3+13.4+10.0)$, $W_z = 10.0/(18.3+13.4+10.0)$ と設定した．

5.2 サンプルデータ間の識別精度

サンプルデータ間で識別の精度を評価するため，Pixel 3 の 180 のサンプルについて，すべてのサンプル間におけるセンサによる端末姿勢の類似度が TH_o 以下となるようにいくつかのサンプルを削除した．残った 58 のサンプルにおいて，各サンプルのカメラ映像による推定端末姿勢を CH がサーバに送信したもの，センサによる端末姿勢を Non-CH がサーバに送信したものとみなし，さらにすべての姿勢の取得時刻が同じだとして，提案手法による CH による Non-CH（サンプル）の識別精度を評価した．なお上記のようにサンプルを削除したのは，同一時刻における Non-CH 間の端末姿勢類似度が TH_o 以下のものが存在した場合，Attack-I への対策としてシステムは CH による認証処理を停止するためである．

評価の結果，識別の正解率（正しく識別されたものの割合）は 72.4％であった．また識別に失敗した 27.6％の内訳として，誤ったユーザと識別されたもの（他人受入率）が 5.1％，すべてのユーザとの類似度が閾値 TH_o を上回り該当なしとなったもの（本人拒否率）が 22.4％であった．

カメラによる端末姿勢推定に一定の誤差が生じるため，今回のようにセンサによる端末姿勢間の類似性が低い場合でも，これらのような識別の失敗が発生する．

セキュアな情報共有システムの実現のためには，他人受入率の低さが重要であり，これは類似度の閾値 TH_o によって変動し，本人拒否率の低さとのトレードオフの関係にある．本システムでは，CH は連続して撮影するカメラ映像のうちの 1 枚の画像のみから Non-CH の識別を行う．したがって，CH は該当する Non-CH がなかったという形で Non-CH の識別に失敗したとしても，数秒後に再度識別処理を行うことは，CH の利便性を大きく損なうものではない．

5.3 現実的なシナリオ下での識別性能

本節ではより現実的な環境下におけるユーザ識別性能の評価実験について述べる．シナリオとして，一人の CH が，CH の目の前で座りながら保持するスマートフォンでウェブブラウジングしている 3 人の Non-CH を識別する状況を想定し，その際の識別性能を評価する．

CH は 5 分程度 3 人の Non-CH の前に立ち続け，約 1 秒おきに各 Non-CH の端末姿勢を推定し，それを基に各同一時刻における相互の正解率を求めた．なおカメラ画像にピンぼけが発生していたり，スマートフォンが見切れているようなタイミングの映像フレームはあらかじめ手動で映像から除外し，全 300 秒のカメラ映像を入力とした．そのような全 300 秒間のうち，センサによる 2 つ以上の端末姿勢の類似度が閾値 TH_o を上回り，認証情報（端末姿勢）の衝突が発生したのは 10％に相当する 31 秒間であった．また 300 秒間のうちの 74％である 223 秒間は識別の正解率が 100％ととなり，残りの 43 秒間では 3 人のうちの 1 人の識別に失敗し，正解率が 66％，さらに残りの 3 秒間では 3 人のうちの 2 人の識別に失敗し，正解率が 33％となった．なお識別の失敗はすべて，カメラによる端末姿勢に該当する Non-CH の端末姿勢が存在しない，本人拒否の失敗であり，セキュリティ上の脅威である他人受入は発生しなかった．

6. おわりに

近い将来，AR・MR デバイスの高度化や小型軽量化により，それらのデバイスを常時装着するヒトが，スマートフォンなどの従来のモバイル端末では困難であった可視光や赤外線・電波などを利用した常時の 3 次元空間センシング能力を有するようになる．本研究では，そのような従来のモバイル端末の性能を大幅に超えた高度な空間認識・リアルタイムな相互通信・ネットワーク接続を有するヒトを我々はコネクテッド・ヒューマン（CH）と定義し，CH と従来のスマートフォンを保持する Non-CH 同士がセキュアに情報を共有するためのプラットフォームを設計し，スマートフォンなどの端末姿勢に基づくユーザ特定・認証手法を新たに提案した．

今後の課題として，より現実的な様々な環境下でのパラメータに応じた提案認証手法の性能評価を行うこととともに，インフラが存在しない環境においても空間の動的コンテキストを取得できる CH と本情報共有プラットフォームの防犯や交通安全支援，高度なモバイルクラウドセンシングなどへの応用が挙げられる．

参考文献

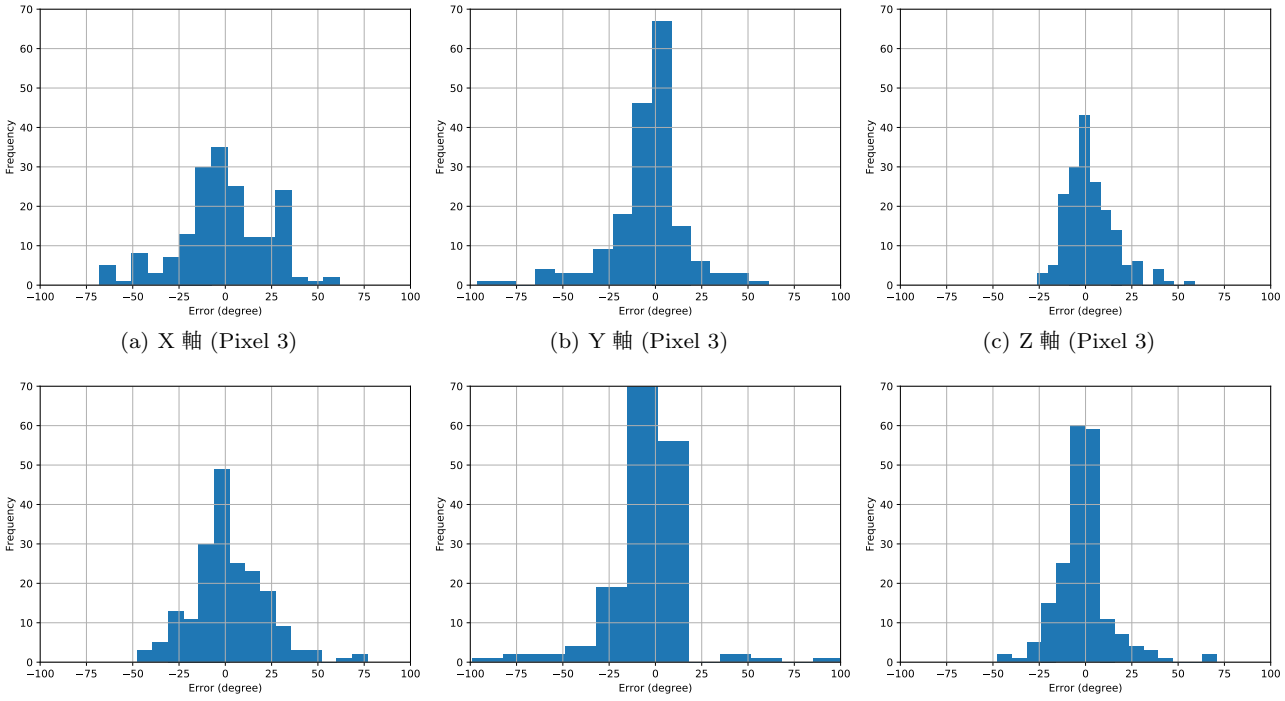
[1] 高藤巧, 藤田和久, 樋口雄大, 廣森聡仁, 山口弘純, 東野輝夫, 下條真司. トラッキングスキャナとモーションセンサを用いた高精度屋内位置推定手法の提案. 情報処理学会論文誌, Vol. 57, No. 1, pp. 353–365, 2016.

[2] Joseph Redmon and Ali Farhadi. Yolov3: An incremental improvement. *arXiv*, 2018.

[3] 情報処理推進機構セキュリティ・センター. 本人認証技術の現状に関する調査報告書. "`https://www.ipa.go.jp/security/fy14/reports/authentication/authentication2002.pdf`", 2003-. [Online; accessed 17-May-2020].

[4] Yongtae Park, Sangki Yun, and Kyu-Han Kim. When IoT met Augmented Reality: Visualizing the Source of the Wireless Signal in AR View. In *Proc. of the 17th ACM International Conference on Mobile Systems, Applications and Services (MobiSsys 2019)*, pp. 117–129, 2019.

[5] F. d. P. Müller, E. M. Diaz, and I. Rashdan. Coopera-



(a) X 軸 (Pixel 3)　(b) Y 軸 (Pixel 3)　(c) Z 軸 (Pixel 3)

(d) X 軸 (Phab 2 Pro)　(e) Y 軸 (Phab 2 Pro)　(f) Z 軸 (Phab 2 Pro)

図 5　各軸ごとの端末姿勢（重力方向とのなす角）推定誤差

tive positioning and radar sensor fusion for relative localization of vehicles. In *2016 IEEE Intelligent Vehicles Symposium (IV)*, pp. 1060–1065, 2016.

[6] 藤田敦, 山口弘純, 東野輝夫, 高井峰生. 近接車両センシング情報のリアルタイム共有のための車車間通信プロトコル. 電子情報通信学会技術研究報告 IEICE technical report: 信学技報, Vol. 116, No. 406, pp. 7–12, 2017.

[7] Dirk Schulz, Dieter Fox, and Jeffrey Hightower. People tracking with anonymous and id-sensors using rao-blackwellised particle filters. In *Proceedings of the 18th International Joint Conference on Artificial Intelligence*, IJCAI'03, p. 921–926, San Francisco, CA, USA, 2003. Morgan Kaufmann Publishers Inc.

[8] 岩橋宏樹, 樋口雄大, 山口弘純, 東野輝夫. 歩行者群の移動軌跡情報を用いたモバイルカメラ画像内の人物位置推定手法. 情報処理学会論文誌, Vol. 56, No. 2, pp. 470–482, 2015.

[9] Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C Lawrence Zitnick. Microsoft coco: Common objects in context. In *European conference on computer vision*, pp. 740–755. Springer, 2014.

[10] Yana Hasson, Gül Varol, Dimitris Tzionas, Igor Kalevatykh, Michael J. Black, Ivan Laptev, and Cordelia Schmid. Learning joint reconstruction of hands and manipulated objects. In *CVPR*, 2019.

[11] Yana Hasson, Gül Varol, Dimitris Tzionas, Igor Kalevatykh, Michael J. Black, Ivan Laptev, and Cordelia Schmid. Learning joint reconstruction of hands and manipulated objects. "`https://hassony2.github.io/obman.html`", 2019-. [Online; accessed 17-May-2020].