新生活様式におけるコミュニティ形成のための サイバーフィジカル空間共有基盤の設計開発

天野 辰哉[1,a)]　水本 旭洋[1,b)]　山口 弘純[1,c)]　松田 裕貴[2,d)]　藤本 まなと[2,e)]　諏訪 博彦[2,5,f)]
安本 慶一[2,g)]　中村 優吾[3,h)]　田上 敦士[6,i)]

概要：アフターコロナ社会では，人々のコミュニティ維持や心のつながり形成のための遠隔コミュニケーションが一層重要となる．特に近年の VR 技術の発展により，各人が存在する物理空間を常時センシングし，精緻な空間データを仮想空間に持ち寄ることで互いの存在をシームレスに知覚する空間共有の実現などが期待されている．しかし，空間把握のための高精度 3 次元カメラや超臨場サウンドデバイス，ロボットなどのセンサーデバイスが取得するデータは，各人のプライバシー要求を反映しながら共有されることが大前提となる．本研究では人間のプライバシー要求の自動理解と，それに基づく精緻空間データの共有・アクセス制御技術を統合したサイバーフィジカル空間共有基盤の設計開発を行う．3 次元 LiDAR による物理空間スキャンと，AR デバイスによる一方向の空間共有システムの簡易プロトタイピングを行い，同基盤の基礎性能評価を行った．その結果，無圧縮センサデータをそのまま利用する場合，約 360 Mbps のデータストリームが生じ，また点群のセマンティックセグメンテーションに 1 フレームあたり約 0.25 秒要することが分かった．

1. はじめに

コロナ禍で人々の物理的距離に制約が生じ，結果として深刻なコミュニケーション上の課題を生むことが徐々に明らかになってきている．ソーシャルディスタンシングにより社会的人間関係の維持が困難になっているだけではなく，外出や行動を抑制されることにより間接的な精神的抑圧が発生し，自死や引きこもりの増加が懸念されている．また，初等教育から高等教育における学校行事や勉学の遅滞も深刻であり，個々の家庭では高齢者の孤立や家族間の分断も生じている．アフターコロナではそれらの状況の改善が期待されるが，新しい生活様式においてもコロナをはじめとする感染症への対策は継続されるはずであり，そういった状況においても人々が「つながる」ために必要とする距離感を維持できる社会づくりが求められている．中でもビフォアコロナにおいて物理的に実現されていた「空間」や「場」の共有を情報通信技術で実現できれば，そういった物理的距離による断絶を埋めることができると期待される．

これに対し，現状では超臨場感の研究開発が急速に進んでおり，コロナ禍の中でも情報共有やコミュニケーションをスムーズに行うツールが報道等で頻繁に紹介されている．中でも仮想現実（VR）は有望な技術であり，仮想会議室，仮想学校，仮想オフィスなど，人々が仮想的な空間を共有することで，空間的ギャップを埋めることが可能となる．特に 6DoF（six degrees of freedom）など，仮想空間における当事者行動の自由度は飛躍的に高まってきており，いわゆるテレプレゼンスの高度化も進んでいる．また，現在の 5G 通信のみならず，今後の Beyond 5G 通信や 6G 通信では，高速無線ネットワークに裏打ちされた感覚コミュニケーションも増加すると期待される．

しかし，それらにおけるコミュニケーションはあくまで仮想世界を介した仮想コミュニケーションであり，一定の目的（プログラム開発，議論，会社勤務など）を持った人々がアバターなどを介して没入する環境である．したがって，

1 大阪大学，Osaka University
2 奈良先端科学技術大学院大学，Nara Institute of Science and Technology
3 九州大学，Kyushu University
4 JST さきがけ，JST PRESTO
5 理化学研究所 革新知能統合研究センター，RIKEN Center for Advanced Intelligence Project (AIP)
6 KDDI 総合研究所，KDDI Research, Inc.
a) t-amano@ist.osaka-u.ac.jp
b) mizumoto@ist.osaka-u.ac.jp
c) h-yamagu@ist.osaka-u.ac.jp
d) yukimat@is.naist.jp
e) manato@is.naist.jp
f) h-suwa@is.naist.jp
g) yasumoto@is.naist.jp
h) y-nakamura@ait.kyushu-u.ac.jp
i) tagami@kddi-research.jp

©2021 Information Processing Society of Japan

我々の存在する物理空間やそこで発生しているイベント，あるいはその場の空気感といった，よりパーソナルで現実的な情報を共有することを目的としたものではない．我々の生活基盤はあくまで現実空間であり，ビフォアコロナと同じような「場の共有」が求められている．その意味で，我々が存在するパーソナル空間そのものをシームレスに行き来可能な，新しい空間共有技術が必要とされている．例えば，離れて暮らす高齢者や家族の間で，キッチンで調理する状況をセンシングし，その状況をスマートグラスやホログラム，立体ディスプレイなどを介してシェアすることができれば，家族間のつながりを感じることができ，また遠隔料理教室への参加といったように個人の精神的幸福度や生活の質向上にも寄与できる．

本研究では，我々が存在・活動するパーソナル現実空間を適切なプライバシー制御のもとで共有することを可能とする，サイバーフィジカル空間共有基盤の設計開発を行う．同基盤は，プライバシー機敏情報を含むフィジカル空間を精緻にセンシング可能なデバイスが偏在する環境を想定し，それらが捉える高精細映像や音声などのセンシングデータや，そのデータから直接あるいは間接的に推定されるコンテキストに対し，どの内容のデータをどのように加工するか，あるいはアクセス可能とするかの制御を行い，パーソナル空間を提供するユーザーに安全安心を与える技術を開発する．

3 次元 LiDAR による物理空間スキャンと，AR デバイスによる一方向の空間共有システムの簡易プロトタイピングを行い，同基盤の基礎性能評価を行った．その結果，無圧縮センサデータをそのまま利用する場合，約 360 Mbps のデータストリームが生じ，また点群のセマンティックセグメンテーションに 1 フレームあたり約 0.25 秒要することが分かった．

2. 関連研究

2.1 点群認識技術

深度カメラや LiDAR などのセンサにより得られる 3 次元点群を活用し，空間内の物体形状や人物の骨格・行動を推定する研究がなされている．PointNet[1] は深層学習に基づく入力点群の分類やセグメンテーション手法であり，入力点群の各点に対し畳み込み処理で特徴を抽出し，それらに対して Max Pooling を適用するにより点群全体の特徴を抽出する．PointNet++[2] は入力点群に対して細部から全体にかけて階層的に PointNet を適用することでローカルな点群構造も学習している．これら PointNet などにより抽出される特徴を用いて，三次元点群からの状況認識や三次元形状復元する手法が提案されている．例えば P2P-Net[3] は PointNet++を利用しシェイプ変換を行う手法であり，入力点群の骨格点群や断面点群への変換を実現する．AtlasNet[4] は三次元オブジェクトの表面を二次元多様体とみなしし三次元形状を学習することで，RGB 画像・点群からの三次元メッシュ生成や物体分類さらにモーフィングなどを可能にする．PCN[5] は部分点群からの補完された全点群生成手法である．本研究ではこれら高精度な 3 次元点群からの空間認識・物理形状再現技術を活用し，実空間のヒト・モノ・空間のリアルタイム理解を実現する．

2.2 プライバシー

Palen ら [6] は，新しいインタラクションシステムを開発する際のプライバシー問題を検討するため，日常生活における対人関係のプライバシー管理を支配する 3 つの緊張関係：Disclosure Boundary（どの状況でどの情報を公開するか），Identity Boundary（コミュニケーション相手との関係に応じた公開するアイデンティティの特徴），Temporal Boundaries（公開した情報に対する過去・現在・未来における解釈や行動の間の緊張関係）について議論し，family intercom，active badges，mobile telephones の 3 つのケースに対する考察を行った．Squicciarini ら [7] は，SNS にアップロードされた写真のプライバシー管理に関して，自分を含む複数人が映り込んだ写真の所有権および閲覧権を管理・制御する方法を提案した．また，Li ら [8] は，SNS への投稿画像に映り込む人物に対するプライバシー管理のためのぼかしが，ユーザ体験やプライバシー問題に与える影響を調査した．Venkatesh ら [9] は，多人数が参加するビデオ会議を対象に，一般参加者とプライベート参加者を区別し，プライベート参加者の視聴覚情報をリアルタイムに匿名化（映像情報はブラックアウト，音声情報はピッチ変更）するシステムを提案した．Wang ら [10] は，ビデオ会議などにおけるビデオストリーム内に映るユーザの顔を，予め指定されたポリシーに従ってぼかすことによってプライバシーを管理するシステムを提案した．Du ら [11] は，監視カメラ映像内における身体領域やアクセサリなどのプライベートコンテンツを認識し，ぼかすことによってプライバシー保護を行うシステムの提案と，システムのユーザ調査を行った．Zhu ら [12] は，Deepfake 技術を利用し，医学研究における映像データに映り込む人物の顔領域を架空の人物とスワッピングすることで，顔のキーポイント抽出を行うことが可能な映像を生成しつつ対象者のプライバシーを保護するシステムを提案した．監視カメラに映り込む人のプライバシー情報を保護するため，コンピュータビジョンに基づく画像加工の方法が多数提案されている [13].

しかし，これら既存研究は，自身の映像や音声のみを共有する際のプライバシー情報管理しか扱っておらず，本研究が想定する，自身とその周辺を含む空間を他と共有する際のプライバシー情報管理を扱っていない．

©2021 Information Processing Society of Japan

2.3 VR 共有空間の実現技術

遠隔空間における情報共有やコミュニケーションをスムーズに行う技術が急速に発展している．VR においては，360 度視野の拡張やマルチユーザによる空間共有技術が提案，実装されている．例えば文献 [14] では，ビデオのタイリングや視点適応的な最適資源割当てなど，スケーラブルな 360 度映像のための通信技術を提供している．また，文献 [15] では，共有された VR 空間への入退室の仕組みを研究している．実在する家庭における入退室の振舞いに基づき，VR 共有空間におけるホストとゲストの役割を区別する枠組みを構成している．

2.4 音の共有

音声チャットを用いた遠隔コミュニケーションは，テキストチャットを用いるより強い社会的な繋がりを生み出すことが分かっている [16]．しかしながら，音声チャットでは，参加者数が増えるほど会話が混線する可能性が高かまるため，音声会議や多人数参加型ゲームにおいては，発言の順番待ちや機会損失が生じ，没入感や満足感が失われるという問題がある．この問題を解決するため，仮想空間上のユーザの位置関係を基に音量調整や音声合成を行う研究が行われている [17], [18]．これらの手法では，混線の回避だけではなく，通信トラフィックを削減することもできる．しかし，このような仮想空間上でのコミュニケーションでは，会話相手に声が聞こえているか分からず不安を感じるという問題がある [19]．また，音声コミュニケーションを用いたシステム全般では，環境音や同居人の声が他者に聞かれる可能性や，音声からジェンダーが他者に知られることによりトラブルに発展する可能性など，音声を用いることによるプライバシーの問題がある [16], [20]．

2.5 ネットワーク

ITU-T による 2030 年に必要な新しいサービスや機能に関する報告 [21] によると，ホログラム通信や触覚通信（Telehaptics）などよりリッチで没入感のあるインタラクティブなサービスが普及することが予想されている．これらのサービスの要求（帯域や遅延など）は，エンド-エンドやオーバーレイでの対応が難しく，ネットワーク内の機能との連携が必要であることが示唆されている．

現在, テレカンファレンスアプリには WebRTC [22], [23] が広く利用されている．WebRTC は本来，ブラウザ間で P2P 通信を行うための通信フレームワークであるが，多人数でのコミュニケーションを効率良く実行するために，ネットワーク内に MCU（Multi-point Control Unit）や SFU（Selective Forwarding Unit）を用いて中継する手法が提案されており，これらのネットワーク機能を提供するサービスも存在する [24], [25]．

近年，HTTP/3 で利用するトランスポート層である QUIC [26] が標準化された事もあり，QUIC ベースの双方向通信フレームワーク WebTransport [27] や，WebRTC Data Channel を QUIC で置き換えた WebRTC over QUIC [28] など新しいプロトコルも提案されている．また，プライバシー的観点から中継サーバに情報が漏洩しない E2EE（End-to-End Encryption）にも注目が集まっており，IETF においても 2020 年 11 月から SFRAME (Secure Media Frames) WG が立ち上がり，議論を開始している．テレワークならびにオンライン化の普及に従い，これらのコミュニケーションプロトコルの重要性は向上しており，既存のプロトコルの見直し，ならびに新しいプロトコルの提案が活発に行われている．

2.6 インタラクション

Follmer らは，単身赴任などで遠距離になっている家族の一体感をサポートするために，遠隔にいる親子がビデオ会議上で一緒に遊べる共有アクティビティを提案している [29]．また，Wallbaum らは，離れて暮らす祖父母と孫の非対面コミュニケーションにおいて，孫側の飽きを克服するために，おもちゃ箱型のビデオ会議インタフェースである StoryBox を提案している [30]．Hendriks らは，リモートコミュニケーションを豊かにするために，モダリティの次元を高めるのではなく，気を散らす要因を減らすための触覚クッション型デバイスである Azalea を開発している [31]．しかし，これらの先行研究では，リモートコミュニケーションをより良くするための新しいインタフェースやインタラクションが提案されているものの，双方のプライバシー要求に対応するための仕組みについては，十分に検討がなされていない．本研究では，遠隔家族交流や遠隔料理教室といったユースケース毎に，空間の人やオブジェクト毎に発生しうるプライバシー要求を整理・分析する．そして，プライバシー要件を配慮しつつ，必要に応じてコミュニケーションの情報量やモダリティを調整し，より一体感のある空間共有を実現するための新しいインタフェース及びインタラクションを設計・実装を目指す．

3. サイバーフィジカル空間共有とは

本研究開発では，センシングデバイスの高度化・データの精緻化により，サイバー空間とフィジカル空間の垣根を取り払い，各人が存在する現実空間をなるべく自然に共有することを実現する際に生じる技術課題の解決を図る．これに対し，LiDAR や深度カメラなどを活用し，3 次元空間理解技術，理解し獲得したデータを，人間のプライバシー要求理解に基づき制御する人間のプライバシー要求理解技術，および，それらのデータをセキュアに送受信するためのセンシングデータのアクセス制御技術が必要となる．これらについて以下で述べる．

©2021 Information Processing Society of Japan

3.1　3 次元空間理解技術

実空間のヒト・モノ・空間を正しく把握し空間データとして提供するための技術，ならびにその膨大なボリュームを無線転送可能なレベルに圧縮・最適化する技術を確立する．特に 3 次元センシングに着目し，人の体幹・行動，予表情，モノの存在，空間サイズやテキスチャなど，プライバシーを考慮しない場合に取得・推定可能なすべての情報を究極にセンシングする技術を達成し，現実空間のサイバーデータ化で最先端の技術を実現するとともに，膨大なデータボリュームの最適化技術と組み合わせ，5G を想定した無線環境で実時間転送可能なデータ量に抑制する技術を主眼とする．

3.2　人間のプライバシー要求理解技術

各人のセンシングデータに対するプライバシー要求の自動理解と獲得技術を確立する．コンテキスト推定による状況理解技術の開発として，人間のプライバシー要求に影響を与える行動・環境・ソーシャルコンテキストを獲得する手法を構築する必要がある．また，複合コンテキストによるプライバシー表現モデルの構築において，プライバシーは個人の主観に依存するところが大きいため，プライバシーを複数の行動・環境コンテキストの複合により表現するモデルを構築し，プライバシーを左右する要因を明らかにする必要もある．さらに，深層強化学習によるプライバシー要求獲得技術においては，プライバシー要求における各個人の嗜好や傾向を強化学習により獲得する技術を開発する．

3.3　センシングデータの通信技術

サイバーフィジカル空間共有基盤においては，実空間のヒト・モノを把握しオブジェクトとして扱い，さらにそれぞれのオブジェクトに対して映像，点群など各センシングデータが紐づけられている．それぞれのセンシングデータはエラーの許容度やプライバシポリシに違いがあり，それらを理解しつつ多対多で情報共有する必要がある．これらに対応するため，Stream をベースとした通信方式の適用について検討し，アプリケーションとの連携により，効率よくセンシングデータを送受信可能な基盤を設計開発する．

3.4　ユースケース

本研究で想定するユースケースの例として，遠隔家族交流と遠隔料理教室について述べる．

3.4.1　遠隔家族交流

近年は，COVID-19 の影響により，地方に住む高齢者が困難な状況に置かれている．クラスター感染のリスクにより介護施設の利用は難しく，地方と都市の感染状況の違いによる温度差から，都市に住む若い世代が地方に住む高齢の家族を訪問することが憚られる状況になっている．本

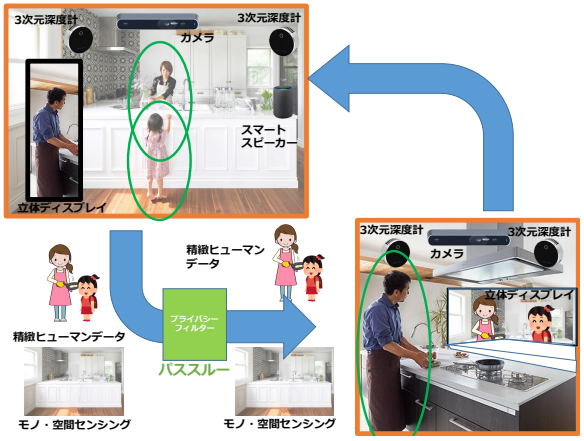


図 1　料理における空間共有（プライベート）

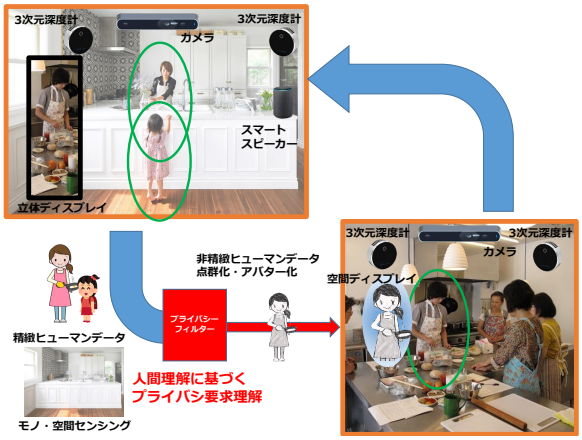


図 2　料理における空間共有（パブリック）

ユースケースでは，このようなケースにおいて，遠く離れた家族が，互いにお互いの空間を気軽に訪問できるようにすることで，見守りや体験の共有ができるようにすることを目的とする．例えば，お互いの家のキッチン・ダイニングルームのみを共有するようにした場合，キッチン・ダイニングルームにいるときのみ，互いの存在や体験（一緒に料理をする，食事を摂るなど，図 1 参照）を共有できるため，プライバシーにも配慮することが可能である．

3.4.2　遠隔料理教室

図 1，2 に家庭における料理の様子を異なる 2 つの状況（プライベートとパブリック）で共有する例を示す．図 1 のプライベートケースは，先に述べた遠隔家族交流であり，すべてのセンシング情報（人の位置・体形・顔や行動，会話および空間データなど）は精緻なまま，あたかも同じ空間にいるかのように共有することが基本的に承認される．一方，図 2 のパブリックケースは遠隔料理教室であり，3 次元センシングで精緻に捕捉された顔や体形情報は抽象化され，同居する他の家族の存在や会話はフィルターされる．一方で臨場感の向上のため，包丁の動きやそれを追う視線などは現実感をもって高精度に伝達される．

©2021 Information Processing Society of Japan

4. サイバーフィジカル空間共有基盤の設計開発

4.1 アーキテクチャ概要

図 3 にアーキテクチャ全体図を示す．本基盤は各ユーザの自空間に位置し，ユーザ側の自空間センシングデータからコンテキストを生成して他空間と共有したり，他空間から共有されたコンテキストを受け取って自空間と融合させるエッジサーバと，複数ユーザからの空間コンテキスト間の整合性を管理したり，エッジサーバと連携してエッジ処理の一部を負担することでエッジサーバーの負荷を軽減したりするクラウドサーバからなる．エッジサーバでは，ユーザ実空間に配置された LiDAR やカメラ，マイク，あるいはユーザが装着する生体センサや特製の IoT デバイス（テイラーメイド IoT）からのデータを処理して自空間のコンテキスト表現に必要なセグメントを様々な表現方法（点群やテキストなど）で生成する．それらは実空間のオブジェクトとして関連付けられ，空間的な相関関係を与えられたのちに，プライバシポリシ決定機構により送るべきコンテンツとその表現方法が決定される．クラウドサーバでは複数の空間から得られるコンテキストデータ間の相関関係の整合性を管理する．

各エッジサーバでの処理フローのうち，データ収集と処理部分の詳細を図 4 に示す．この図においては，各センサからのデータはセグメント化され，そのデータから推定される意味（コンテキスト）が与えられる．例えば LiDAR データであれば，人・モノ・背景の点群が含まれるが，それらの分別（セグメンテーション）を行い，人であればさらに人の身体部位の分別まで行う．それらから推定されるコンテキスト（例えば「左腕」や「包丁」）をラベル付けする．

4.2 3 次元空間理解

本空間理解プロセスは大きく，デバイス依存のセンシング機構およびオブジェクト関連付け機構により実現される．処理は（1）データセグメンテーション，（2）セグメント認識・分類，（3）人および物体トラッキングの 3 ステップでの実行を想定する．

3 次元点群の場合，PointNet[1] や VoteNet[32] による点群のセマンティックセグメンテーションによりデータセグメンテーションおよび各セグメントの認識・分類が可能である．これにより入力点群の各点の属性（例えば「床」「テーブル」「人」など）が得られる．プライバシー要求理解やコンテキスト推定の対象となる「人」と認識されたセグメントに対しては「顔」や「手」などの部位ごとのセグメンテーションや骨格検出を適用する．一般にこれらセグメンテーション・物体認識は処理負荷が高く毎フレームの適用は困難であると予想されるため，点群トラッキングとの組み合わせによる実行回数の削減を検討している．これらセグメント分割・ラベル付け処理により，ユーザ空間内の各ヒト・オブジェクトの空間メタ情報が毎フレーム得られる．オブジェクトの空間メタ情報にはオブジェクトの座標・向き・形状・種別・状態が，ヒトの空間メタ情報にはこれらに加えて表情や骨格などが含まれる．毎フレームの空間メタ情報のリストがオブジェクト関連付け機構やプライバシポリシ決定機構・オブジェクト関連付け機構の入力となる．

各空間メタ情報が指し示すオブジェクトやヒトの客体空間上の表現（3 次元点群やアバター・アイコン，文字表現など）はプライバシポリシ決定機構や受信者の選択により動的に決定される．センシング機構は可視表現データを生成し，要求に応じてセンサデータ送信基盤に提供する．3 次元点群などセンサデータ自体が要求された場合は，そのデータをセンサデータ送信基盤に提供する．オブジェクトの 3 次元メッシュデータを要求された場合は対象オブジェクトのメッシュを AtlasNet[4] などによりセンサデータから動的に構築，あるいは事前に構築し共有する．

4.3 人間のプライバシー要求理解・獲得のためのモデルとのその学習

本節では，人間のプライバシー要求を理解・獲得するため，まず，(1) プライバシー表現モデルを定義する．モデルは，どういったコンテキストの時に，どのオブジェクトに対しどのようなプライバシー保護レベルを設定したいのかを表現する．このモデルを学習するために必要な技術として，(2) コンテキスト推定，(3) モデルの強化学習とプライバシー獲得インタフェース，を提案する．

4.3.1 プライバシー表現モデル

各ユーザが自身とその周辺空間（主体空間と呼ぶ）に属するすべてのオブジェクトに対し，プライバシー保護レベルを設定する．

各オブジェクトのプライバシー保護レベルは，状況やコミュニケーション相手，コミュニケーション目的等のコンテキストに応じて変化する．よって，プライバシー保護レベルを以下の関数 pl として定義する．

$$pl : \mathcal{C} \times \mathcal{O} \rightarrow \mathcal{L} \tag{1}$$

ここで，$\mathcal{C}$ はコンテキストの集合，$\mathcal{O}$ はオブジェクトの集合，$\mathcal{L}$ はプライバシ保護レベルの集合（{低，中，高}など）である．

4.3.2 コンテキスト推定

コンテキストとして，行動コンテキスト，環境コンテキスト，ソーシャルコンテキストを対象とする．行動コンテキストは，現在どういった行動をしているのか（料理，食事，仕事等），環境コンテキストは，どういった場所なのか（自宅，職場，外出先，移動中など），ソーシャルコンテキ

©2021 Information Processing Society of Japan

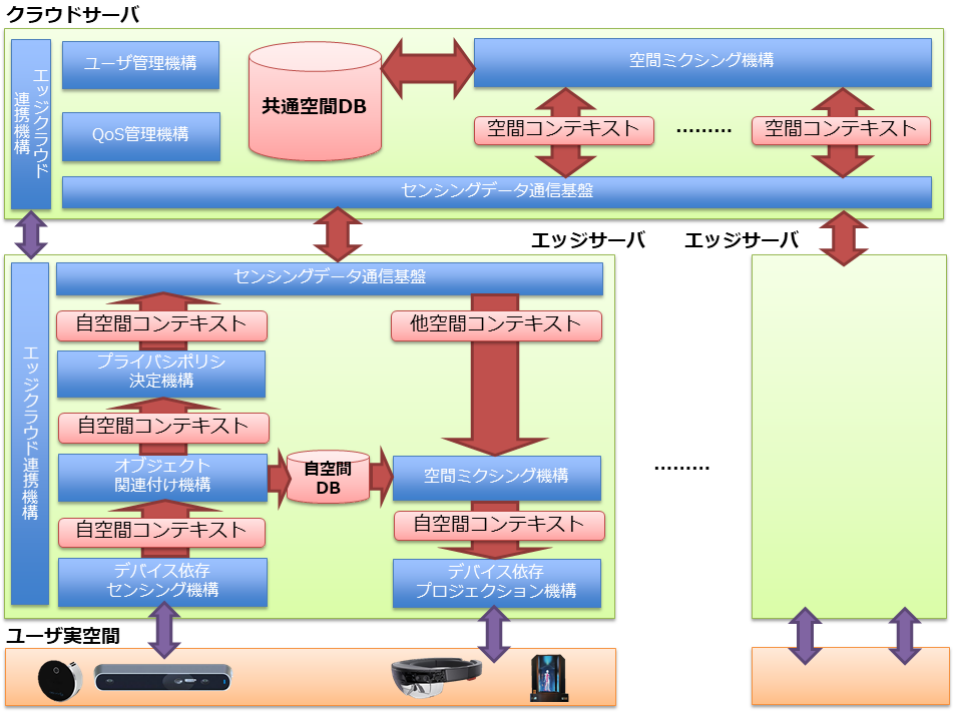


図 3　サイバーフィジカル共有基盤のモジュールアーキテクチャ

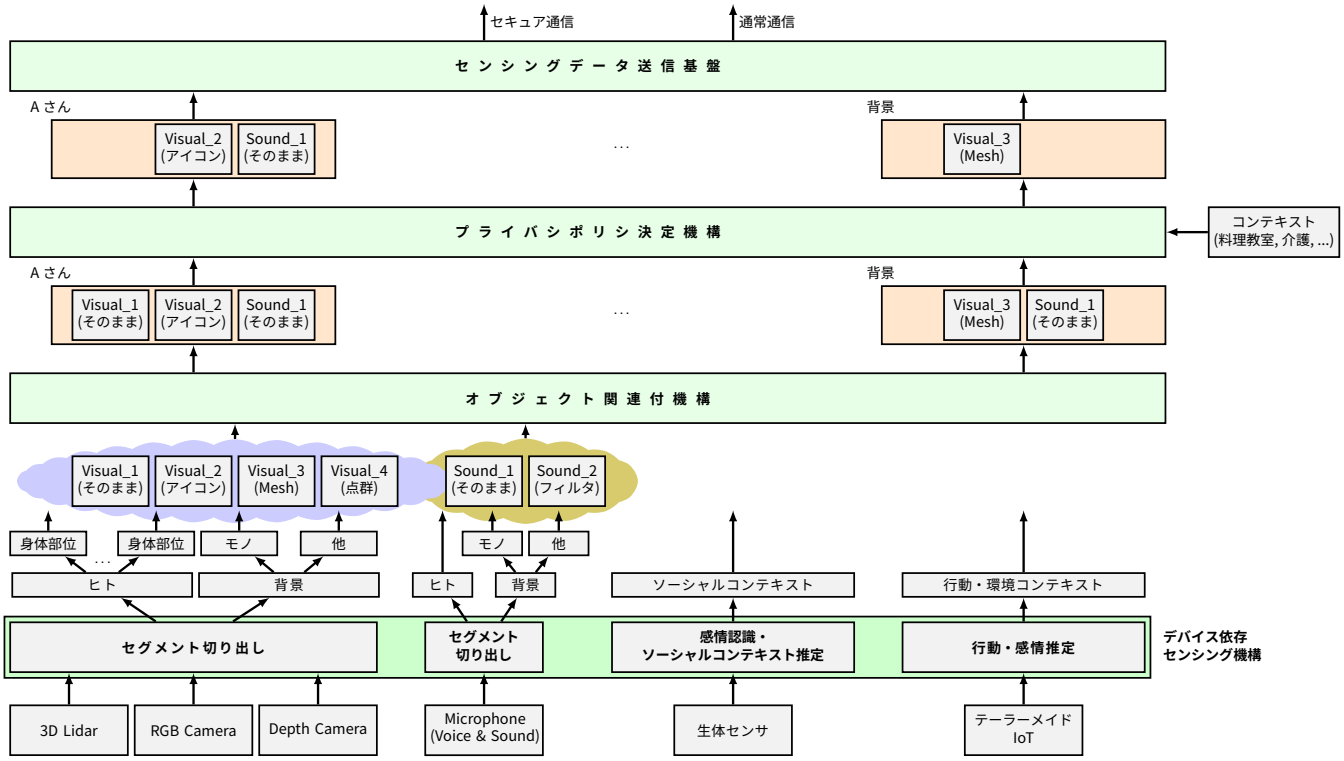


図 4　エッジサーバーにおける処理フロー

ストは，コミュニケーション相手との関係（続柄，身分，立場等），これまでのコミュニケーション頻度や親密度等を表す．

行動コンテキストについては，生活行動認識システム（文献 [33] など）が対象空間に設置されていると想定し，システムの認識結果をそのまま利用する．環境コンテキストに関しては，使用する PC やスマートフォンの設置位置から自動推定する，もしくは，ユーザが手動で入力することを想定する．ソーシャルコンテキストのうち，静的なコンテキスト，例えば，家族の続柄や組織における身分関係などに関しては，本システムの利用者登録を行い，ソーシャルグラフ（コミュニケーションしたことのあるユーザ間に

©2021 Information Processing Society of Japan

辺）を作成し，辺の属性として登録しておくことで対処する．動的なソーシャルコンテキストに関しては，過去のコミュニケーション履歴とプライバシー保護レベル設定履歴等から，ユーザ間の親密度を算出し，ソーシャルグラフの辺に更新していくことを検討している．

4.3.3 モデルの強化学習とプライバシー獲得インタフェース

本研究では，主体空間における各オブジェクトのプライバシー保護レベルを強化学習モデルにより自動的に獲得することを目指している．そのためには，ユーザが主体空間の各オブジェクトに対し，プライバシー保護レベルを設定するためのインタフェースが必要である．本研究では，以下の２つのインタフェースを検討している．

- 静的プライバシー設定インタフェース（静的 IF）：ユーザがコミュニケーション相手に接続する前に，主体空間，その中のオブジェクト群，デフォルトプライバシー保護レベルを提示し，ユーザは各オブジェクトを選択しプライバシー保護レベルを適切なものに変更する（オンライン会議システムの背景画像設定に相当）．
- 動的プライバシー設定インタフェース（動的 IF）：コミュニケーション中にも，オブジェクト毎のプライバシー保護レベルを変えたい場合に使用する．静的 IF をポップアップして操作するする方式（ポップアップ IF，既存オンライン会議に相当）と主体空間内で物理オブジェクトに触れるまたはジェスチャすることで変更する方式（ジェスチャ方式）を検討する．

4.4 センシングデータ通信基盤

サイバーフィジカル空間共有基盤において，通信基盤にはオブジェクト単位での制御が求められる．具体的にはオブジェクトデータの種別によってパケットロスを許容する/しないといった通信品質の制御，プライバシーポリシーに応じてクラウドサーバでのデータ処理を許容する/しないといった暗号化制御を行う必要がある．これらを実現するためには，複数の TCP/IP コネクションを利用することも考えられるが，NAT や FW への対応を考えると Stream ベースのアプローチが好ましいと考えられる．

WebRTC において，データを転送するための DataChannel は SCTP を拡張 [34] したプロトコルが利用されており，複数の DataChannel を１つのコネクション上に構築することが可能である．また，QUIC [26] においても，複数の Stream を１つのコネクション上で提供することが可能である．このような Stream を活用することで，複数のオブジェクトデータを送受信することは可能である．しかしながら，複数の品質の異なる Stream を１つのコネクション上で提供する場合，輻輳制御との不整合や，Head of Line のような問題が発生する可能性がある．また，クラウドサーバでのデータ処理を阻止するために E2EE を行った場合，通信路の暗号化と重複するためパフォーマンスが劣化することが指摘されてる．このため各実装における性能評価が重要である．

このため本通信基盤の設計開発において，パフォーマンスを計測しスケーラビリティを担保した通信方式を選択する．さらに，本設計をプライバシポリシ決定機構と共有し，アプリケーションと連携した通信方式により，効率と性能の両立を図る．

4.5 空間共有のためのインタフェース

4.5.1 入出力インタフェース

映像と音声だけでは共有が難しい，臨場感のある空間共有を実現するため，人の細かな動作を精緻に取得し再現する入出力インタフェースを開発する．入力インタフェースとして，3D カメラやマイクロフォンアレイに加え，目的ごとに最適化されたテーラーメイド IoT デバイスを開発する．出力インタフェースとして，Hololens2 などの高機能アクチュエータや投影型ディスプレイを活用し，共有相手からの共有空間情報を自身のパーソナル空間にプロジェクションするインタフェースを開発する．空間データベースに保存されるオブジェクトである音や映像，3 次元点群，モノや住空間情報をそれらの属性に応じて，自然にアクチュエーションする機構を適切な IoT デバイスやインタラクションデバイスの組み合わせで実現する．また，それらの環境情報，部分空間のコンテキスト情報を自身の空間に自然に反映する機構を開発する．

4.5.2 主体空間への客対空間のマッピング

主体空間を S_0，客体空間を $S_1, \Gamma\Gamma\Gamma, S_n$ とする．空間 S_i に含まれるオブジェクトの集合を $O(S_i)$ と表記する．

4．3 節の技術で，主体空間のコンテキストが c の時の各空間 S_i の各オブジェクト $o \in O(S_i)$ へのプライバシー保護レベル $pl(c, o)$ が決定する．プライバシー保護レベル pl をオブジェクト o に適用後のオブジェクトを $ao(o, pl)$ と表記する．

主体空間 S_0 に客体空間群 $S_1, \Gamma\Gamma\Gamma, S_n$ をマッピングした共有空間を HS_0 と表記する．コンテキストを c とする．$O(HS_0) = O(S_0) \cup_{i=1}^{n} \cup_{o \in S_i} ao(o, pl(c, o))$ となるように，共有空間 HS を構成する．ここで，客体空間の各オブジェクトを主体空間のどの位置に配置するかを決定するために，ユーザがオブジェクトごとに自由に設置位置を決定可能な GUI を作成することを計画している．

料理教室を例に説明する．図 2 において，主体空間 S_0 は，料理教室会場 (オブジェクトは *classroom*)，客体空間 S_1 は，遠隔ユーザ (オブジェクトは *remote user*) であるとする．遠隔ユーザのコンテキストは “cooking class” とする（行動コンテキスト：料理，環境コンテキスト：自宅，ソーシャルコンテキスト：初めて参加）．

$O(S_0) = \{cooking\ class\}$，$O(S_1) = \{remote\ user\}$ であ

©2021 Information Processing Society of Japan

り，$pl(cooking\ class, remote\ user) = \text{“}h$
時，図 2 に示すように，遠隔ユーザ自身に
主体空間に反映される．この例では，ユ
オブジェクトとしているが，ユーザの顔
オブジェクトとして取り扱い，異なるプ
ベルを設定することも可能（手だけはリ
するなど）である．

5. 基礎性能評価

3 次元 LiDAR による物理空間のスキ
イスによる一方向の遠隔空間共有の簡易
し，サイバーフィジカル空間共有基盤の
行った．システムの概要を図 5 に示す．
して Livox 社の Livox Avia および Inte
L515 を使用し，受信側ユーザのインタ
Microsoft HoloLens 2 を使用した．Avia および L515 それぞれに対して，色情報を含まない 3 次元点群データセット ScanNet[35] で学習済みの VoteNet および色情報を含む SUN RGB-D データセットにより学習済みの VoteNet による人物検出・セグメンテーションを適用し，人物点群を受信側の Hololens2 へネットワークを介して送信する際の処理速度を評価した．各 LiDAR による点群例を図 6 に示す．

VoteNet による人物検出に要する時間を計測した．検出処理は Intel Core i7-7820HQ 2．90GHz CPU および Quadro P4000 GPU のエッジサーバ上で行った．1 フレームあたり最大 240,000 点を含む Avia では検出処理に平均 0.240 秒，平均 91,883 点含む L515 では平均 0.231 秒要した．Avia の最大計測距離はおよそ 320m であり屋内環境では計測可能な最大点数が常に得られた一方で，L515 はセンサから遠方の点は計測されなかったため，フレームごとに得られる点数に変化が生じた．なお点群に色情報の付随しない Avia のデータからは人物を検出することができなかった．これは図 6(a) に示す Avia の走査パターンによって生じる点群特徴が，学習に用いた ScanNet の点群と異なるためであると考えられる．さらに 3 次元点群データを無圧縮で受信デバイスに送信した場合に要求される通信速度を計測した．L515 の場合，1 送信フレームサイズが平均約 1.5MB であった．そのためセンサフレームレートである 30 FPS で送信するには約 360 Mbps の通信帯域が必要となる．

本研究の想定する遠隔家族交流および遠隔料理教室の実現にはリアルタイムな双方向通信かつ L515 のようなセンサの各ユーザ実空間への複数台設置が要求される．今回の実験では 1 秒あたり 4 フレームのセグメンテーション処理が可能であったものの，例えば遠隔料理教室においてユーザの包丁の動きを正確に遠隔に伝えるためには，より高速なセグメンテーションや追跡処理が必要である．これに対

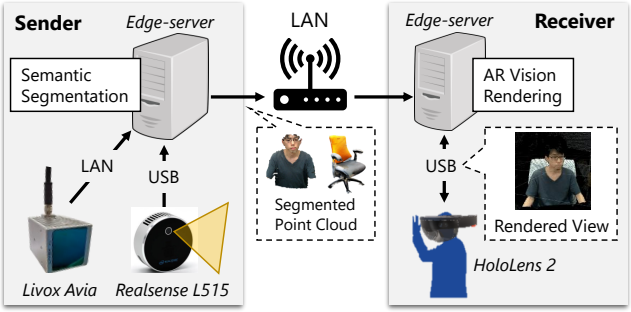
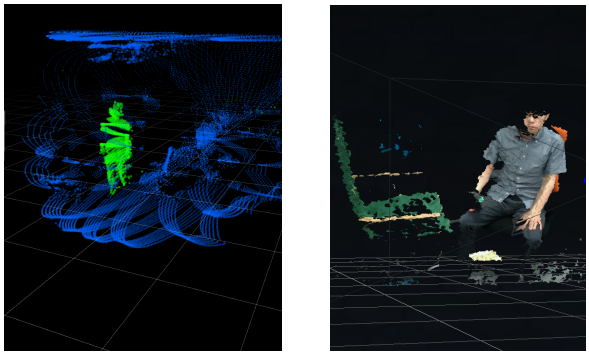


図 5 一方向遠隔共有システム



(a) Livox Avia　(b) Realsense L515

図 6 各 LiDAR による 3 次元点群．Livox Avia の 3 次元点群は人物部分を緑色で描写している．

して本基盤では 3 次元点群内オブジェクトの高速追跡技術の併用および，テーラーメイド IoT や生体センサ，マイクロフォンなどを組み合わせたマルチモーダルなオブジェクト状態理解により高速 3 次元空間理解の実現を検討している．また今回の実験では無圧縮データでは 360 Mbps 程度の通信帯域が必要であった．これに対して，本研究では，3 次元点群の圧縮技術および複数のセンサーの空間位置依存関係から冗長なデータを自動的に削減するとともに，3 次元空間の依存関係ベースモデル圧縮・受信者視点や状況に応じた送信データの動的選択などにより必要なデータをストリームを 1/10 程度への抑制することを目指す．

6. まとめ

本研究では人間のプライバシー要求の自動理解と，それに基づく精緻空間データの共有・アクセス制御技術を統合したサイバーフィジカル空間共有基盤の設計開発を行った．同基盤は，空間把握のための高精度 3 次元カメラや超臨場サウンドデバイス，ロボットなどのセンサーデバイスにより，各人が存在する物理空間を常時センシングすることで精緻な空間データを取得し，各人のプライバシー要求を反映しながら他者の物理空間を自空間に共有する枠組みを提供しており，本研究ではその初期設計を行っている．3 次元深度センサーによる物理空間スキャンと，AR デバイスによる一方向の空間共有システムの簡易プロトタイピングを行い，同基盤の基礎性能評価を行った結果，同基盤

©2021 Information Processing Society of Japan

の基礎性能評価を行った．その結果，無圧縮センサデータをそのまま利用する場合，約 360 Mbps のデータストリームが生じ，また点群のセマンティックセグメンテーションに 1 フレームあたり約 0.25 秒要することが分かった．

謝辞

本研究の成果は NICT 課題 222「ウイルス等感染症対策に資する情報通信技術の研究開発」の支援によるものです．

参考文献

[1] Qi, C. R., Su, H., Mo, K. and Guibas, L. J.: PointNet: Deep learning on point sets for 3D classification and segmentation, *Proceedings of the 30th IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2017)*, pp. 77–85 (2017).

[2] Qi, C. R., Yi, L., Su, H. and Guibas, L. J.: PointNet++: Deep hierarchical feature learning on point sets in a metric space, *Proceedings of the Advances in Neural Information Processing Systems (NIPS 2017)*, pp. 5100–5109 (2017).

[3] Yin, K., Huang, H., Cohen-Or, D. and Zhang, H.: P2P-NET: Bidirectional point displacement net for shape transform, *ACM Transactions on Graphics*, Vol. 37, No. 4, pp. 1–13 (2018).

[4] Groueix, T., Fisher, M., Kim, V. G., Russell, B. C. and Aubry, M.: AtlasNet: A Papier-Mache Approach to Learning 3D Surface Generation, *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2018)*, pp. 216–224 (online), DOI: 10.1109/CVPR.2018.00030 (2018).

[5] Yuan, W., Khot, T., Held, D., Mertz, C. and Hebert, M.: PCN: Point completion network, *Proceedings of the 2018 International Conference on 3D Vision (3DV 2018*, pp. 728–737 (2018).

[6] Palen, L. and Dourish, P.: Unpacking" privacy" for a networked world, *Proceedings of the SIGCHI conference on Human factors in computing systems*, pp. 129–136 (2003).

[7] Squicciarini, A. C., Shehab, M. and Paci, F.: Collective privacy management in social networks, *Proceedings of the 18th international conference on World wide web*, pp. 521–530 (2009).

[8] Li, Y., Vishwamitra, N., Hu, H., Knijnenburg, B. P. and Caine, K.: Effectiveness and users' experience of face blurring as a privacy protection for sharing photos via online social networks, *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, Vol. 61, No. 1, SAGE Publications Sage CA: Los Angeles, CA, pp. 803–807 (2017).

[9] Venkatesh, M. V., Zhao, J., Profitt, L. and Sen-ching, S. C.: Audio-visual privacy protection for video conference, *2009 IEEE International Conference on Multimedia and Expo*, IEEE, pp. 1574–1575 (2009).

[10] Wang, J., Amos, B., Das, A., Pillai, P., Sadeh, N. and Satyanarayanan, M.: A scalable and privacy-aware IoT service for live video analytics, *Proceedings of the 8th ACM on Multimedia Systems Conference*, pp. 38–49 (2017).

[11] Du, L., Zhang, W., Fu, H., Ren, W. and Zhang, X.: An efficient privacy protection scheme for data security in video surveillance, *Journal of Visual Communication and Image Representation*, Vol. 59, pp. 347–362 (2019).

[12] Zhu, B., Fang, H., Sui, Y. and Li, L.: Deepfakes for medical video de-identification: Privacy protection and diagnostic information preservation, *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*, pp. 414–420 (2020).

[13] Padilla-López, J. R., Chaaraoui, A. A. and Flórez-Revuelta, F.: Visual privacy protection methods: A survey, *Expert Systems with Applications*, Vol. 42, No. 9, pp. 4177–4195 (2015).

[14] Chakareski, J.: Viewport-Adaptive Scalable Multi-User Virtual Reality Mobile-Edge Streaming, *IEEE Transactions on Image Processing*, Vol. 29, pp. 6330–6342 (online), DOI: 10.1109/TIP.2020.2986547 (2020).

[15] Evans, M. C., Kamineni, S., Cheikh-Ali, O., Fanzo, J., Jiang, S., Majmudar, K., Ren, M. and Hammer, J.: Sharing Multi-User VR Spaces, *Extended Abstracts of the 2020 Annual Symposium on Computer-Human Interaction in Play*, CHI PLAY '20, New York, NY, USA, Association for Computing Machinery, p. 229–233 (online), DOI: 10.1145/3383668.3419871 (2020).

[16] Kumar, A. and Epley, N.: It's surprisingly nice to hear you: Misunderstanding the impact of communication media can lead to suboptimal choices of how to connect with others, *Journal of Experimental Psychology: General* (2020).

[17] Yasumoto, K. and Nahrstedt, K.: Ravitas: Realistic voice chat framework for cooperative virtual spaces, *2005 IEEE International Conference on Multimedia and Expo*, IEEE, pp. 1046–1049 (2005).

[18] Nguyen, C. D., Safaei, F. and Boustead, P.: Optimal assignment of distributed servers to virtual partitions-for the provision of immersive voice communicationin massively multiplayer games, *Computer Communications*, Vol. 29, No. 9, pp. 1260–1270 (2006).

[19] Dey, A., Piumsomboon, T., Lee, Y. and Billinghurst, M.: Effects of sharing physiological states of players in a collaborative virtual reality gameplay, *Proceedings of the 2017 CHI conference on human factors in computing systems*, pp. 4045–4056 (2017).

[20] Kuznekoff, J. H. and Rose, L. M.: Communication in multiplayer gaming: Examining player responses to gender cues, *New Media & Society*, Vol. 15, No. 4, pp. 541–556 (2013).

[21] ITU-T FG NET-2030 Sub-G2: New Services and Capabilities for Network 2030: Description, Technical Gap and Performance Target Analysis, ITU-T Deliverable (2019).

[22] Aboba, B., Vasiliev, V. and Hirano, Y.: WebRTC 1.0: Real-Time Communication Between Browsers, W3C Recommendation (2021).

[23] Alvestrand, H.: Overview: Real-Time Protocols for Browser-Based Applications, RFC 8825, RFC Editor (2021).

[24] NTT Communications: SkyWay, `https://webrtc.ecl.ntt.com`.

[25] KDDI Web Communications Inc.: Twilio, `https://cloudapi.kddi-web.com`.

[26] Iyengar, J. and Thomson, M.: QUIC: A UDP-Based Multiplexed and Secure Transport, RFC 9000, RFC Editor (2021).

[27] Jennings, C., Boström, H. and Bruaroey, J.-I.: WebTransport, W3C Editor's Draft (2021).

[28] Hampson, S.: RTCQuicTransport Coming to an Ori-

©2021 Information Processing Society of Japan

gin Trial Near You, https://developers.google.com/web/updates/2019/01/rtcquictransport-api.

[29] Follmer, S., Raffle, H., Go, J., Ballagas, R. and Ishii, H.: Video play: playful interactions in video conferencing for long-distance families with young children, *Proceedings of the 9th International Conference on Interaction Design and Children*, pp. 49–58 (2010).

[30] Wallbaum, T., Matviienko, A., Ananthanarayan, S., Olsson, T., Heuten, W. and Boll, S. C.: Supporting communication between grandparents and grandchildren through tangible storytelling systems, *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, pp. 1–12 (2018).

[31] Hendriks, S., Mare, S., Gamboa, M. and Baytaþ, M. A.: Azalea: Co-experience in Remote Dialog through Diminished Reality and Somaesthetic Interaction Design, *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, pp. 1–11 (2021).

[32] Qi, C. R., Litany, O., He, K. and Guibas, L. J.: Deep Hough Voting for 3D Object Detection in Point Clouds, *Proceedings of the IEEE International Conference on Computer Vision* (2019).

[33] Matsui, T., Onishi, K., Misaki, S., Fujimoto, M., Suwa, H. and Yasumoto, K.: SALON: Simplified Sensing System for Activity of Daily Living in Ordinary Home, *Sensors*, Vol. 20, No. 17, p. 4895 (2020).

[34] Jesup, R., Loreto, S. and Tüxen, M.: WebRTC Data Channels, RFC 8831, RFC Editor (2021).

[35] Dai, A., Chang, A. X., Savva, M., Halber, M., Funkhouser, T. and Nießner, M.: ScanNet: Richly-annotated 3D Reconstructions of Indoor Scenes, *Proc. Computer Vision and Pattern Recognition (CVPR), IEEE* (2017).

©2021 Information Processing Society of Japan