

CSI2Imageの時間的ロバスト性に関する基礎検討

加藤 空知<sup>†</sup> 岩崎 裕輔<sup>‡</sup> 福島 健<sup>‡</sup> 藤橋 卓也<sup>‡</sup> 村上 友規<sup>‡</sup> アベセカラ ヒランタ<sup>‡</sup>  
 猿渡 俊介<sup>‡</sup> 渡辺 尚<sup>‡</sup>  
<sup>†</sup> 大阪大学工学部 <sup>‡</sup> 大阪大学大学院情報科学研究科  
<sup>‡‡</sup> 日本電信電話株式会社 NTT アクセスサービスシステム研究所

1 はじめに

電波情報によるセンシング対象空間の可視化は、空港・学校・駅等での不審人物の侵入検知、ショッピングモールなど屋内での混雑度推定の高度化につながる。このような背景から、筆者らは室内において利用される Wi-Fi 通信の電波を用いて空間内の情報が可視化できる手法について検討している。しかしながら、一般的に電波情報と機械学習を組み合わせたセンシング技術は、学習モデルを生成した時刻から時間が経過した電波情報を入力すると、電波情報の変化が起因となって学習モデルによるセンシング精度が低下するという課題が存在する。本稿では、IEEE802.11ac で規定された CSI (Channel State Information) フィードバックと GAN (Generative Adversarial Network) を組み合わせた空間内の画像生成手法を実装する。また、学習済モデルに対して時間経過した電波情報を入力したときに得られる生成画像を元にして、CSI2Image による学習モデルの時間耐性を明らかにする。具体的には、生成画像と元画像との類似度、生成画像を用いた人物検知精度の 2 指標から入力電波情報の時間変化に対する学習モデルを用いたときのセンシング精度の時間変化を評価する。

2 センシング技術の課題

ある空間をセンシングする方法として、カメラ、赤外線、二酸化炭素、ピーコンなどを利用する手法が提案されている。しかしながら、カメラを利用する場合はプライバシー、赤外線や二酸化炭素を利用する場合は場所の汎用性、ピーコンを利用する場合は必要となる無線デバイスの増大が課題となる。[1] 空間情報のセンシングを対象とした既存手法の問題を解決するアプローチとして、電波、特に CSI (Channel State Information) を用いたセンシング手法への注目が集まっている。CSI とは、無線端末間の伝搬チャネルの状態を示す情報である。具体的には、送信機から送信された信号が受信機で受信されたときの、振幅と位相の変動を表す値である。CSI を用いたセンシング手法では、センシング対象となる空間内の人物や物体の位置が変化したときに生じる CSI 値の変動を利用する。

既存研究では、CSI を用いて無線端末間の距離測定や室内での人物位置測定を実現している。例えば、Chronos [2] では、複数無線チャネルから得られる CSI を用いて無線端末間の距離測定とアクセスポイントに対する相対位置の測定が実現されている。具体的には、複数アンテナを有する単一のアクセスポイントから得られる 35 チャネル分の CSI を利用することで、無線端末の位置測定を数十センチ精度で可能であることを示した。

また、機械学習と不完全な CSI フィードバック を組み合わせる高いセンシング性能を目指す手法も提案されている。文献 [3] では、IEEE802.11ac にて規定された Compressed CSI と教師あり機械学習を用いて室内における人物の位置測定を実現している。Compressed CSI には、CSI の位相や振幅情報を角度情報に圧縮した  $\phi$  と  $\psi$  が含まれている。サブキャリアごとに  $\phi$  と  $\psi$  は異なる値を持つ。本手法では、アクセスポイントと受信機が 1 台ずつ空間内に存在する環境において、入力情報に Compressed CSI、学習と推定に Support Vector Machine を用いたとき、約 90% の精度で人物の位置を正しく推定することが可能となっている。

教師あり機械学習を Compressed CSI の分類・識別に活用することで、ある空間を対象とした高精度のセンシングや物体検出が可能になる。一方で、教師あり機械学習は学習に必要な教師データのラベリングや分類等の前処理を手で行わなければならない。前処理の対象となるデータが多量になるほど、教師データを適切に分類するのは困難となる。特に、移動する人物や物体を対象として空間情報をセンシングする際、人物や物体の動きは連続的であるため、それぞれの位置や状態に対して適切にラベル付けすること自体が困難となる。

3 提案手法

本稿では、教師なし学習である GAN を用いて、前処理に要する労力を削減しつつ、電波情報から空間情報をセンシングする CSI2Image を提案する。具体的には、GAN を用いて Compressed CSI から対象空間を推定す

る画像を生成するとともに、生成画像に対して物体検知技術を用いて画像中の人間の有無を判定することで、対象空間内の人物位置推定を実現する。CSI は MIMO 通信の品質向上のために一般的に用いられている技術であり、Compressed CSI は通信時のオーバーヘッド削減を達成できるよう実装された CSI の圧縮方式の 1 つである。CSI2Image で利用する GAN は教師なし学習の生成モデルの一種で、与えられた電波情報から空間情報を示す画像を生成する上での特徴を訓練データから学習する。得られた生成モデルに対して、新たに観測した電波情報を入力すると、教師データを必要とせず、すでに学習した特徴を元にして対象空間を表す画像を生成できる。

図 1a および図 1b に CSI2Image のシステム構成を示す。CSI2Image は学習段階と画像生成段階の 2 段階から構成される。学習段階では、IEEE802.11ac の規格を利用する無線端末が一定時間ごとに送信する Compressed CSI から得られる角度情報を電波情報としてサーバに逐次送信する。対象空間を撮影するカメラも同様に一定時間ごとに取得した画像をサーバに送信する。サーバはカメラ画像と電波情報を一定時間蓄積して、蓄積したカメラ画像と電波情報を GAN に入力することで、入力された電波情報に対応する画像を生成するための学習モデルを生成する。

画像生成段階では、学習段階において生成した学習モデルと新たに取得した Compressed CSI から得られる角度情報を用いて対象とする空間画像を生成する。図 2 に、空間内の 1 地点に人がいるときのカメラ画像と CSI2Image から得られた学習モデルに角度情報を与えたときの生成画像をそれぞれ示す。学習モデルは与えられた Compressed CSI と対応するカメラ画像を生成するための特徴を学習しているため、入力した Compressed CSI に含まれる電波情報の変化に対応した空間画像を生成できる。

4 実験評価

提案手法では、ある時刻の Compressed CSI を入力とした GAN から生成した学習モデルを用いて、取得した Compressed CSI に対応する対象空間の画像を生成できる。しかしながら、学習モデルに対して時間経過した電波情報を入力した場合、対象空間内の電波状態が変化して本来のカメラ画像と生成画像との類似度が変化する可能性がある。このとき、生成画像から対象空間の状態を正確にセンシングできなくなる可能性がある。

本実験では、CSI2Image で生成した学習モデルに対して時間経過した電波情報を入力したとき、生成画像によるセンシング精度をどの程度維持できるか評価した。具体的には、学習モデル生成後から最大 100 分後の Compressed CSI を学習モデルに入力して画像を生成した際に、生成画像が空間情報をどれだけ正しく再現できているか、人物位置をどれだけ正しく推定できているかを確認する。また、異なるエポック数の学習モデルにおける生成画像の類似度および人物位置推定精度を明らかにすることで、学習時間が CSI2Image の時間耐性に与える影響についても評価する。

4.1 実験環境

図 3 に実験環境を示す。図 3 にて示す場所にカメラ、アクセスポイント、デバイス、キャプチャデバイスをそれぞれ配置した。キャプチャデバイスはデバイスとアクセスポイント間で送受信される Compressed CSI を取得してサーバに送信する。本実験ではアクセスポイントのアンテナ数が 3 個、デバイス側のアンテナ数が 2 個、サブキャリア数が 52 のときの Compressed CSI が取得でき、位相に関する情報振幅に関する情報を角度情報に変換した  $\phi$  と  $\psi$  をそれぞれ 3 つずつ取得できる。 $\phi$  と  $\psi$  は 52 個のサブキャリアからそれぞれ異なる値を取得できるため、合計  $52 \times (3 + 3) = 312$  個の角度情報を取得できる。取得した角度情報はサーバで学習モデル生成および画像生成に利用する。カメラは図 3 に示す向きの画像を一定時間ごとに取得してサーバに送信する。取得画像の解像度は  $320 \times 240$  画素であるが、本実験では  $64 \times 64$  画素にダウンサンプルした画像を利用する。

本実験では、図 3 に示す A 地点、B 地点、C 地点を被験者が時計回りに移動して、各地点でのカメラ画像と対応する Compressed CSI を継続的に取得する。具体的には、被験者は各地点で 1 分間カメラに向かって立って静

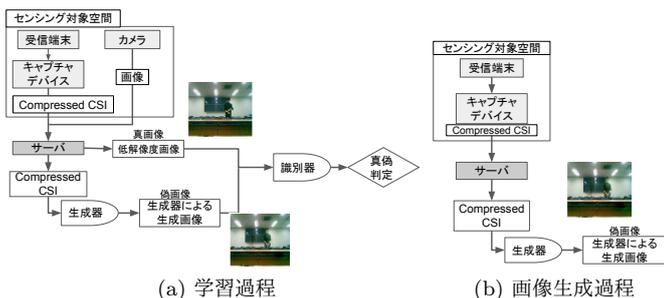


図 1: CSI2Image



図 2: 左:カメラ画像, 右:生成画像

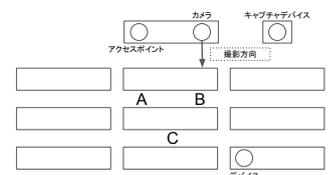


図 3: 実験環境  
 A, B, C は人物位置  
 実験に使用した部屋は縦横  
 7.0 m × 10.7 m

止する。1分経過後、次の地点へすぐに移動して再び1分間カメラに向かって立って静止する。100分間に渡って以上の動きを繰り返した。実験から取得した100分間のCompressed CSIおよびカメラ画像は10分ごとに分割する。このとき、はじめの10分間で取得したCompressed CSIおよびカメラ画像を学習モデル生成に使用する。このとき、学習エポック数を1000, 4000, 12000に設定して学習モデルをそれぞれ生成した。その後、10分ごとに分割したCompressed CSIを異なるエポック数の学習モデルに入力して画像を生成する。

4.2 評価指標

学習モデルの時間耐性を評価するために、カメラ画像に対する生成画像の類似度、生成画像による位置推定精度を評価指標として使用する。生成画像の類似度を評価するために、カメラ画像と生成画像を元に代表的な指標であるPSNR (Peak Signal Noise-to-Ratio) とSSIM (Structural SIMilarity) [4]を使用した。PSNRは以下の式で表される。

$$PSNR = 20 \log_{10} \left( \frac{1}{\sqrt{MSE}} \right).$$

ここで、Mean Square Error (MSE)はカメラ画像と生成画像との平均二乗誤差である。SSIMはカメラ画像と生成画像の構造的な類似度を評価する指標で、PSNRと比較して主観的な印象を反映した指標であるとされている。SSIMは0から1までの値を取り、1に近いほど生成画像がカメラ画像に近いことを表している。

生成画像による位置推定精度を評価するために、YOLO (You Only Look Once) [5]を用いて生成画像から人物位置を正しく推定できた割合を算出した。本評価では、3通りの位置推定精度基準を定めた。1つ目の基準では、生成画像内でカメラ画像と同じ位置にただ1人の人物が検出できたとき、正しく位置推定ができたものとする。2つ目の基準では、生成画像内に複数の人物が検出できるものの、カメラ画像と同じ位置に生成された人物が最も高いスコアで検出できたとき、正しく位置推定ができたものとする。3つ目の基準では、位置や数を問わず、生成画像内に人物が検出できたとき、正しく位置推定ができたものとする。

4.3 生成画像類似度

図4に、時間経過したCompressed CSI入力時におけるカメラ画像と生成画像間の類似度をSSIMで示す。横軸は学習モデル生成時からどの程度の時間が経過したCompressed CSIを入力したか、縦軸は画像類似度を表すSSIMを示している。グラフ中のそれぞれの線は学習時のエポック数に対応している。図4の結果から、いずれの学習モデルもモデル生成時刻から時間経過したCompressed CSIを入力するほどSSIMの値が低下しているのが分かる。学習モデル生成時から100分経過したCompressed CSIを画像生成に使用したとき、エポック数4000の学習モデルから得られる生成画像が最も高いSSIM値0.886を達成することが分かった。また、エポック数4000の学習モデルは、学習モデル生成時からの経過時間に関係なく、他のエポック数の学習モデルと比較して最も高いSSIMを達成している。

図5に、時間経過したCompressed CSI入力時におけるカメラ画像と生成画像間の類似度をPSNRで示す。横軸は学習モデル生成時からどの程度の時間が経過したCompressed CSIを入力したか、縦軸は画像類似度を表すPSNRを示している。図5の結果から、学習モデル生成時から100分経過したCompressed CSIを画像生成に使用したとき、エポック数4000および12000の学習モデルから得られる生成画像が最も高いPSNR値21.7dBを達成することが分かった。以上の実験結果から、学習モデル生成時から100分の間にSSIMは最大で0.0754、PSNRは最大7.59dB低下することが分かった。また、必ずしも学習時のエポック数が多い学習モデルが電波情報の時間変化に対してより類似度が高い生成画像を維持できるとは限らないことが分かった。

4.4 位置推定精度

本節では、時間経過したCompressed CSIから得られる生成画像に対してYOLOを元にした位置推定を実施したとき、3通りの位置推定精度基準でどの程度の位置推定精度が得られるかを評価する。図6、図7、図8に、学習モデル生成時から時間経過したCompressed CSI入力時におけるCSI2Imageによる人物位置推定精度を示す。横軸は学習モデル生成時からどの程度の時間が経過したCompressed CSIを入力したか、縦軸は位置推定精度を示す。ここで、図6、図7、図8では、4.2節で述べた位置推定精度基準である基準1、基準2、基準3をそれぞれ採用している。評価結果から、以下の3つのことが分かる。

1つ目は入力するCompressed CSIの時間経過とともに、位置推定精度が低下していることである。より具体的には、100分間の間に、位置推定精

表1: エポック数4000, 12000の学習モデルにおける人物像生成位置100分経過時のCompressed CSI使用時

エポック数	全画像数	正位置に1人物像	正位置を含み複数位置に人物像	人物像なし
4000	395	147	15	67
12000	395	136	34	30

度は基準1の場合で最大0.656、基準2の場合で最大0.602、基準3の場合で最大0.285まで低下した。本実験では、すべての実験機器の位置および室内環境は100分間固定している。その上で、時間経過にもなって生成画像を用いた位置推定精度が低下していくのは、1) 図3に示したA地点、B地点、C地点に被験者が移動して立つ際、立つ位置に多少のズレが生じていること、2) コヒーレンス時間に起因して時間経過時の電波状態と学習モデル生成時の電波状態が時間とともに変化していくことが理由として考えられる。

2つ目は、基準1および基準2においては、エポック数4000の学習モデルが最も位置推定精度が高くなる一方、評価基準3においては、よりエポック数が多い学習モデルが高い位置推定精度を達成していることである。表1に、エポック数4000, 12000の学習モデルにおける生成画像数を人物の結核状況で分類したものを示す。学習モデルのエポック数が多いほど、A地点、B地点、C地点のうち、少なくとも1地点に人物像を生成している画像の割合が高くなっていることが分かる。

3つ目は、エポック数の少ない学習モデルでは、判定基準によらず、低い位置推定精度を達成することである。これは、エポック数の少ないモデルではYOLOで検出可能な人物像を含む画像を生成できていないことに起因する。ただし、人物位置推定の評価においては図6に示すように、学習時のエポック数1000のモデルが他のモデルよりも優位になっている時間が存在するため、エポック数と人物位置推定精度が必ず正の相関に従うわけではないと言える。

以上の結果から、時間耐性のある学習モデルを構築するためにはモデル生成時のエポック数を多くすることが必要であると分かった。一方で、エポック数を増やした学習モデルを用意できたとしても、時間とともにわずかに変化する人物の位置や電波状態が起因して生成画像品質が低下するとともに、生成画像を利用した人物位置推定精度が低下してしまうことが確認できた。時間経過に対して生成画像品質と人物位置推定精度を維持するためには、電波情報を含めてセンシング対象空間内で時間変化する要素を考慮したCSI2Imageのネットワーク構造を提案する必要がある。

5 おわりに

本稿では、Compressed CSIとGANを組み合わせることで電波情報から空間情報をセンシングするCSI2Imageを提案するとともに、その時間的ロバスト性について評価した。今後の課題として、時間経過に耐性を持つ学習モデルを生成する手法を提案することが挙げられる。

謝辞 本研究はJSPS科研費(JP19H01101, JP17KT0042)の支援の下で行った。

参考文献

- [1] Y. Iwasaki, T. Fukushima, T. Murakami, H. Abeysekera, S. Saruwatari, and T. Watanabe, "CSIとGANによる電波情報から空間情報の生成手法の基礎検討," 電子情報通信学会ソサイエティ大会, pp.1-1, Sept. 2019.
- [2] D. Vasisht, S. Kumar, and D. Katabi, "Decimeter-level localization with a single WiFi access point.," In proceedings of the USENIX Symposium on Networked Systems Design and Implementation (USENIX NSDI'16), pp.165-178, March 2016.
- [3] T. Fukushima, T. Murakami, H. Abeysekera, S. Saruwatari, and T. Watanabe, "Evaluating indoor localization performance on an IEEE 802.11ac explicit-feedback-based CSI learning system," IEEE VTC Spring, pp.1-6, April 2019.
- [4] Zhou Wang, A.C. Bovik, H.R. Sheikh, and E.P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," IEEE Transactions on Image Processing, vol.13, no.4, pp.600-612, April 2004.
- [5] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp.779-788, June 2016.

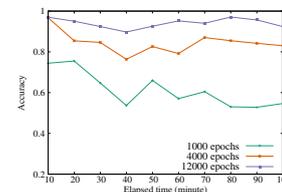
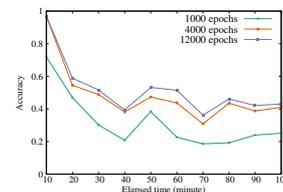
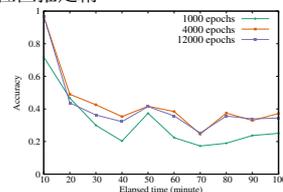
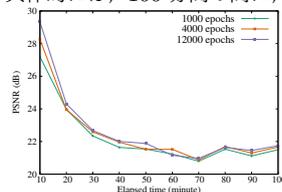
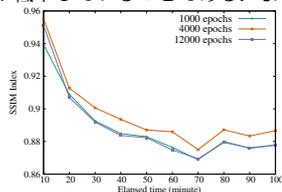


図4: 電波情報の時間経過に対するカメラ画像と生成画像間のSSIM

図5: 電波情報の時間経過に対するカメラ画像と生成画像間のPSNR

図6: 電波情報の時間経過に対する位置推定精度(基準1)

図7: 電波情報の時間経過に対する位置推定精度(基準2)

図8: 電波情報の時間経過に対する位置推定精度(基準3)