抽象度に着目した 宅内スマートフォン捜索向け接触物体推定手法

西 陽也[1,a)]　石田 繁巳[2]　村上 友規[3]　大槻 信也[3]

概要：宅内でスマートフォンを紛失した際に，電話をかけるなどしてスマートフォンから音を鳴らしその音を頼りにユーザが歩いて捜索するが，ユーザの感覚頼りで発見までに時間と労力を要する．著者らはスマートスピーカを用いたスマートフォンの周辺状況推定によって捜索を支援するシステムを提案している．スマートフォン周辺状況のうち，接触物体を推定する研究がいくつか報告されている．しかし，学習データに含まれない物体は推定することができず，宅内スマートフォン捜索への応用は難しい．そこで本研究では物体の抽象度を考慮したスマートフォン接触物体推定手法を提案する．(1) 低い抽象度レベルでの推定結果に確信を持てない場合は抽象度を上げて推定する，(2) 各抽象度レベルの推定モデル構築の際に異なる抽象度レベルの推定モデルの情報を利用するという 2 つのアプローチにより，学習データに含まれない物体も含めて高い精度での接触物体推定を実現する．実環境で収集したデータを用いて提案手法の評価を行った結果，提案手法は 21 種類の接触物体を推定精度 0.966 で推定し，学習データにない物体に対しても有効性があることを確認した．

1. はじめに

近年，宅内でスマートフォンを紛失するユーザが増加している [1]．ユーザは別の端末から紛失したスマートフォンに電話をかけるなどして音を鳴らし音を頼りに歩いて捜索するが，ユーザの聴覚に頼っている点とスマートフォンが何かに覆われている場合に捜索が難しい点で効率が悪い．

そこで著者らはスマートスピーカを用いた音響センシングによるスマートフォン捜索支援システムを提案している [2]．**図 1** に示すように，提案システムはスマートスピーカによる音響センシングによってスマートフォンの周辺状況を推定し，ユーザにフィードバックする．スマートフォンの周辺状況はスマートフォンが存在する部屋，接触物体，被覆状態の 3 つで定義している．これら 3 つの情報を提供することでユーザは容易にスマートフォンを発見できる．

提案するスマートフォン捜索システムの実現に向け，著者らはこれまでにスマートフォンの被覆状態分類手法 [2] を報告した．本稿では，スマートフォン周辺状況の 2 つ目

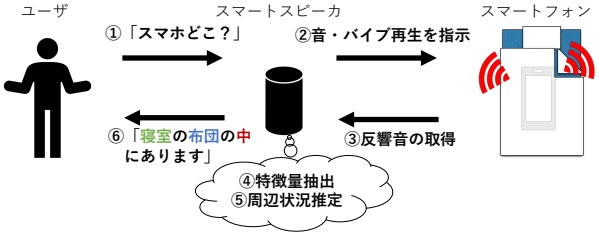


図 1: 提案システム概要図 [2]

としてスマートフォン接触物体を推定する手法を示す．

これまでにもスマートフォン内蔵センサを用いたスマートフォンの接触物体推定手法が報告されている．しかしながら，スマートフォン捜索のための接触物体推定に向けては宅内のほぼすべての物体を学習する必要があり，現実的ではない．これらの手法は学習データにない物体を推定できないため，学習を行わない物体が多いほど接触物体推定の性能が低下し，捜索の妨げとなる．

これに対し，本研究では物体の抽象度を考慮することで学習データにない物体に対しても捜索の手掛かりとなる情報を提供するスマートフォン接触物体推定手法を示す．接触物体そのものが分からなくても「何らかの布に接触している」などの情報は捜索の手掛かりとなる．そこで本手法では抽象度を考慮し，低い抽象度での推定結果に確信を持てない場合には抽象度を上げて推定を行い，その結果をユーザに提供する．

具体的には，(1) 推定確信度に基づいて出力する抽象度

[1] 公立はこだて未来大学大学院システム情報科学研究科
Graduate School of Systems Information Science, Future University Hakodate
[2] 公立はこだて未来大学システム情報科学部
School of Systems Information Science, Future University Hakodate
[3] 日本電信電話株式会社 アクセスサービスシステム研究所
Access Network Service Systems Laboratories, Nippon Telegraph and Telephone Corporation
[a)] g2123046@fun.ac.jp

のレベルを切り替える，(2) 各抽象度レベルの推定モデル構築の際に異なる抽象度レベルの推定モデルの情報を利用するという 2 つのアプローチにより，推定モデルの推定性能および汎化性能を向上させる．抽象度として，本稿では object レベル，material レベル，soft-hard レベルの 3 つを定義した．

提案手法の有効性を検証するために，実環境で収集したデータを用いて提案手法の 2 つのアプローチによる推定性能や汎化性能を評価した．その結果，2 つのアプローチを用いることで，21 種類の物体を推定精度 0.966 で推定し，提案手法が接触物体推定に対して有効であることを示した．

本稿の構成は以下の通りである．2 章では接触物体推定，階層構造を用いたニューラルネットワークに関する研究について述べる．3 章では提案手法について述べ，4 章で提案手法の評価実験について述べる．最後に 5 章でまとめとする．

2. 関連研究

2.1 スマートフォン接触物体推定に関する研究

著者らが調べた範囲では，スマートフォンから発せられる情報を別のデバイスで収集し，スマートフォンの接触物体を推定する研究はこれまでに報告されていない．一方で，スマートフォンから発せられる情報をスマートフォン内蔵のマイクロフォン [3–5]，加速度センサ [6]，複数センサ [7] で収集し，スマートフォンの接触物体を推定する研究が報告されている．

マイクロフォンを用いる手法では，スマートフォンから発せられた音が接触物体の反響特性に依存することを利用して接触物体を推定する．Hwang ら [3] は，スマートフォンのバイブレーションによる振動が接触物体によって異なることを利用して，衣類のポケットの中や机の上，椅子の上など 12 種類のシチュエーションを 0.910 の精度で推定できることを示している．Ali ら [5] は，背景ノイズによる影響を考慮しつつ，スマートフォンのバイブレーションによる振動が接触物体によって異なることを利用して，職場のオフィスと宅内に存在する計 24 種類の物体を 0.865 の精度で推定できることを示している．Hasegawa ら [4] は，スマートフォン内蔵スピーカから発したビープ音の高調波成分が接触物体に依存することを利用して，衣類のポケットの中や木の机，スマホスタンドなど 18 種類のシチュエーションを 0.821 の精度で推定できることを示している．

加速度センサを用いたスマートフォン接触物体推定は，バイブレーションの振動が接触物体の振動特性に依存することを利用して推定する．Cho ら [6] は，接触物体の表面が滑らかであるほどバイブレーションによってスマートフォンの位置の変化が大きくなることを利用して接触物体を推定している．この手法ではソファの上やカバンの中，手の上など 6 つのシチュエーションを 0.850 の精度で推定できることを示している．

複数センサを用いたスマートフォン接触物体推定は，マイクロフォン，加速度センサ，磁力センサなどのスマートフォン内蔵センサで得られる情報を組み合わせることで推定する．Darbar ら [7] は，マイクロフォン，磁力センサ，近接センサを用いてデータパターンに基づいた簡単な if-else ルールベースの階層的な推論手法によって推定している．磁力センサを用いて金属か非金属かを分類し，近接センサを用いて，物体に覆われているかいないかを分類し，最終的に 4 つのグループに分類した後に接触物体を推定する．文献 [7] では 13 種類の物体を 0.917 の精度で推定している．

これらの研究では，最終的な出力は学習データに含まれている物体あるいは事前に想定されている物体のみであり，それ以外の物体を推定することはできない．そのため，宅内のすべての物体を対象として学習やルールの作成を行う必要があり，多くの物体との接触の可能性が想定される宅内スマートフォン捜索に向けては現実的でない．

2.2 階層構造を用いたニューラルネットワークに関する研究

階層的なニューラルネットワークを構築し，同一データに抽象度などの異なるラベルを複数付与して学習することで性能を向上させる手法は，画像分類タスクを中心に報告されている [8–10]．

Wang ら [8] は画像に含まれるオブジェクトをインスタンスとパートという 2 つに分けることで複雑な画像シーンを細分化して 2 つの抽象度で推定する手法を提案している．インスタンスは背景以外のオブジェクトを表し，パートはインスタンスに含まれるオブジェクトを表す．たとえば人をインスタンスとして推定した際はパートとして頭や腕，胸などのオブジェクトを表す．

Novack ら [9] は既存のラベル階層情報と GPT-3 を活用したゼロショット画像分類による暗黙の意味的階層を用いることで未知の画像に対する分類精度を向上させる手法を提案している．暗黙の意味的階層とはデータセット内のクラス間に明示的な階層構造がない場合でも，そのクラス間に階層構造が存在すると仮定することである．文献 [9] の手法は既存手法と比較して分類精度を約 17%向上させた．

池村ら [10] は対象物の 3 次元的構造が持つ内部状態も含めた複数の特徴量を多段型 CNN (Convolutional Neural Network) を用いて推定する手法を提案している．複数の特徴量として対象物の密度，大きさ，向きを枝分かれ構造を持つニューラルネットワークで段階的に推定する．1 つ前の段階の推定結果を次の段階の推定のための入力に逐次加えていくことで，特徴量間の相互関係を考慮したモデルの学習を行っている．

これらの研究は，階層構造を持つ情報をラベルとして，階層構造を捉えることが可能なニューラルネットワークで学習することで，推定精度を向上できることを示してい

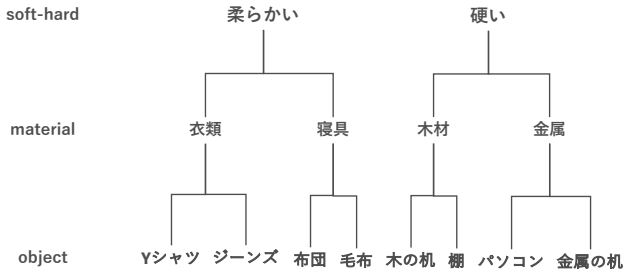


図 2: 抽象度レベルのイメージ図

る．本研究では抽象度を変えた物体の表現を階層構造と考え，階層構造を踏まえた学習を行うことで推定精度を向上させる．

3. 抽象度を考慮した接触物体推定手法

3.1 アプローチ

本手法のキーアイデアは，接触物体を推定する際に物体の「抽象度」に着目することである．**図 2** に示すように，1 つの物体は「抽象度」のレベルを変えて様々な表現で示すことができる．例えば，「Y シャツ」は「衣類」であり，「柔らかい」素材でもある．そこで，接触物体推定結果として「Y シャツ」であることに確信が持てない場合には「衣類」であるという結果を出力する．誤った推定結果はスマートフォンの捜索に大きな影響を与えることから，推定結果に確信が持てない状況を許容し，抽象度を犠牲にして推定精度を向上させる．

物体の抽象度として本稿では object レベル，material レベル，soft-hard レベルの 3 つを定義し，以下の 2 つのアプローチによって抽象度を考慮した接触物体推定を実現する．

1 つ目のアプローチは，推定確信度によって出力する抽象度のレベルを決定することである．推定確信度は推定結果にどの程度確信が持てるかを表す指標である．3 つの抽象度レベルのそれぞれにおいて推定モデルをあらかじめ構築しておき，接触物体の推定時には各推定モデルで推定結果と推定確信度を求める．その上で，抽象度の低いレベルから順に推定確信度が閾値以上であるかを確認し，閾値以上の推定確信度を持つ抽象度レベルの推定結果を出力する．本稿の場合には object レベル，material レベル，soft-hard レベルの順であり，例えば object レベルの推定確信度が閾値を超えていれば object レベルで推定した推定結果を，object レベルの推定確信度が閾値を超えておらず material レベルの推定確信度が閾値を超えていれば material レベルで推定した推定結果をそれぞれ出力する．それぞれの抽象度レベルにおける推定確信度の閾値は推定モデル構築時に設定する．

2 つ目のアプローチは，各抽象度レベルの推定モデル構築の際に異なる抽象度レベルの推定モデルの情報を利用することである．一般に，何も知らない状態で物体が「Y シャツ」であることを推定するよりも，「衣類」であることが分かっているときに「Y シャツ」と推定する方が容易であると予想される．そこで，各抽象度の推定モデルを構築する際に，抽象度レベルの異なる推定モデルの情報が利用されるようなニューラルネットワークを構成して学習を行う．具体的には，抽象度レベルの異なる推定モデルがもつ特徴抽出層をニューラルネットワークの一部に含むことで，それぞれの抽象度レベルの推定モデルの性能を向上させる．

3.2 概要

図 3 に，抽象度を考慮した接触物体推定手法の概要を示す．本手法はデータ収集ブロック，特徴量抽出ブロック，抽象度別接触物体推定ブロックの 3 つのブロックで構成される．はじめに，データ収集ブロックでスマートフォンから発せられるバイブレーション音をスマートスピーカ内蔵マイクを用いて収集し，次に，収集した音を特徴量抽出ブロックでメルスペクトログラムに変換する．抽象度別接触物体推定ブロックでは，3 つの抽象度間で特徴抽出層を共有し順番に学習を行うことで，それぞれの抽象度レベルでモデルを構築する．最後に，構築した 3 つの抽象度モデルで推定確信度を比較し，最適な抽象度レベルの推定結果を選択し，出力する．

以降では各ブロックについて説明する．

3.3 データ収集ブロック

データ収集ブロックではスマートフォンのバイブレーション機能によって発せられた音をスマートスピーカ内蔵マイクで収集する．

推定モデル学習用のデータは，宅内でスマートフォンに通知が届くときのバイブレーション音を用いる．スマートフォンとスマートスピーカが連携されており，スマートスピーカにデータ収集用アプリが導入されていることを前提とし，通知のタイミングに合わせてスマートスピーカが内蔵マイクでバイブレーション音を収集する．Ali ら [5] と同様に，バイブレーションの開始と同時に 3 秒間の録音を行う．このとき，スマートスピーカからユーザにスマートフォンの接触物体を問いかけるなどしてラベル情報を収集し，ラベルづけを行う．

スマートフォン捜索時は，ユーザがスマートスピーカに対して呼びかけてスマートフォン捜索を実行することでデータ収集を行う．スマートフォン捜索が実行されるとスマートスピーカからスマートフォンに対して通知などを行ってバイブレーションを発生させ，スマートスピーカ内蔵マイクでバイブレーション音を収集する．

3.4 特徴量抽出ブロック

特徴量抽出ブロックでは収集した音からバイブレーションの音が含まれるようにトリミングし，特徴量として Mel spectrogram を求める．バイブレーションの開始と録音開始タイミングのずれを考慮し，録音データの 100ms 後か

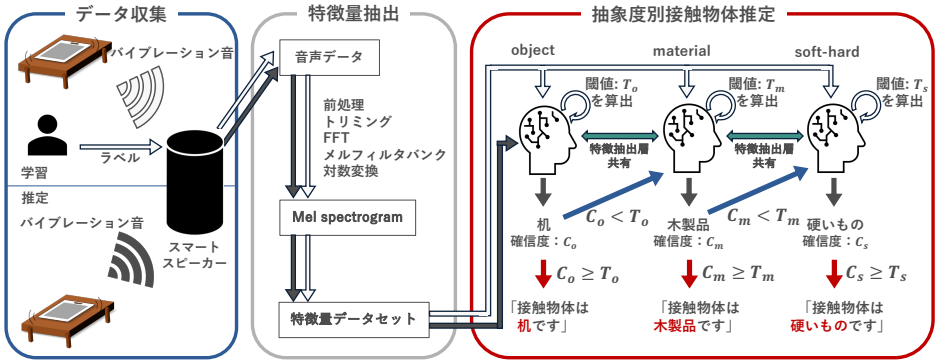
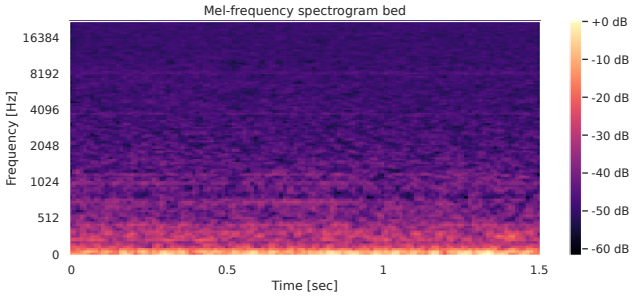
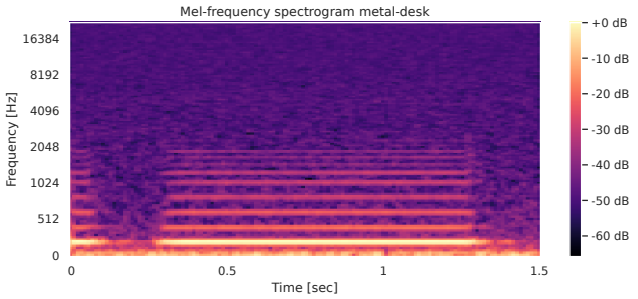


図 3: 提案手法概要図



(a) 布団



(b) 金属の机

図 4: 異なる接触物体での Mel spectrogram の例

ら 1500ms 分をトリミングする．機種や設定によってバイブレーションの長さが 1500ms に満たない場合は，バイブレーションが含まれる部分のみを取り出し，1500ms 分のデータを生成する．

次に，トリミングした音データに対して FFT (Fast Fourier Transform)，メルフィルタバンクの適用，対数変換を行うことで対数スケールの Mel spectrogram を得る．**図 4** に，接触物体が布団と金属の机の場合の Mel spectrogram の例を示す．Mel spectrogram は各時刻で周波数帯域ごとに含まれる音の大きさを表しており，接触物体によって異なることが分かる．

本稿では，サンプリング周波数 44100 Hz，長さ 1500ms の音データに対して FFT の widow サイズを 2048，オーバーラップサイズを 512，メルフィルタバンクのチャネル数を 128 として Mel spectrogram を生成する．音声データが不足する window においては不足する部分を 0 として FFT を行う．各 FFT window で 128 個の Mel 周波数における信号強度情報が得られることから，特徴量の次元，すなわち Mel spectrogram の次元は 128×130 となる．

3.5 抽象度別接触物体推定ブロック

抽象度別推定ブロックでは，特徴量抽出ブロックで得られた特徴量に基づいて抽象度レベルを考慮した接触物体推定を行う．

接触物体の推定に向けて，事前に推定モデルの学習と推定確信度閾値の算出を行う．

図 5 に，推定モデルのニューラルネットワーク構造およびその学習の概要を示す．図には Model-1 と Model-2 の 2 つのモデルがあるが，Model-1 は Model-2 の学習に必要となるモデルであり，推定には Model-2 のみを用いる．

学習は 2 巡に分けて行う．1 巡目では，環境音分類の事前学習済みモデルを元にして転移学習を繰り返し，各抽象度レベルの Model-1 の学習を行う．2 巡目では Model-1 を元にして抽象度レベルの低い推定モデルから順にファインチューニングを行い，各抽象度レベルの Model-2 の学習を行う．

1 巡目の最初に用いる事前学習済みモデルは，Schmid ら [11] の dymn10-as を使用する．dymn10-as は画像認識モデルの ImageNet を大規模な音響イベントデータセットである AudioSet でファインチューニングしたモデルである．

以下に学習の具体的な手順を示す．Model-1, Model-2 には各抽象度レベルの推定モデルが存在することから，Model-i の soft-hard レベル，material レベル，object レベルの推定モデルをそれぞれ shModel-i，matModel-i，objectModel-i と表す．

(1) shModel-1 の学習:
dymn10as の classifier 層に BatchNorm 層と Linear 層を追加して学習する．

(2) matModel-1 の学習:
shModel-1 に最適化された dymn10-as 層に BatchNorm 層と Linear 層を追加して学習する．

(3) objModel-1 の学習:
matModel-1 に最適化された dymn10-as 層に Batch-

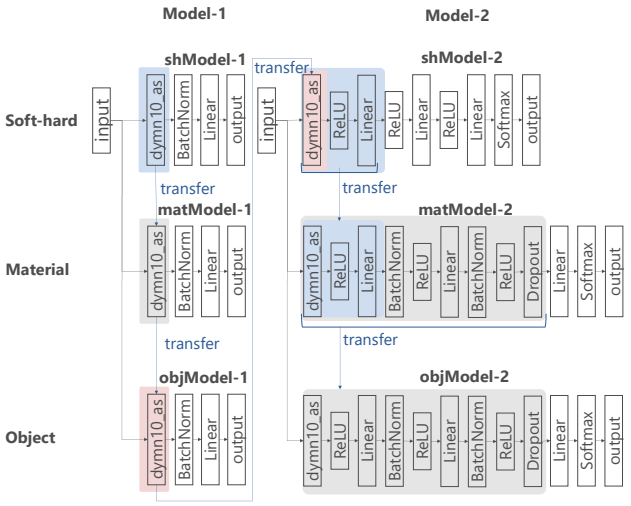


図 5: 推定モデルのニューラルネットワーク構造と学習の概要

Norm 層と Linear 層を追加して学習する．

(4) shModel-2 の学習:
objModel-1 に最適化された dymn10-as 層に Linear 層を追加して学習する．

(5) matModel-2 の学習:
shModel-2 に最適化された dymn10-as 層とそれに続く Linear 層に BatchNorm 層と Linear 層を追加して学習する．

(6) objModel-2 の学習:
matModel-2 の最終出力層以外を転移させ，Linear 層を追加して学習する．

推定モデルの学習が完了した後，推定確信度閾値を算出する．図 3 に示すように，推定確信度の閾値は各抽象度レベルで設定する．推定確信度の閾値は，それぞれの抽象度レベルの推定モデルに学習データを入力したときの正解した推定における推定確信度の平均とする．soft-hard レベル，material レベル，object レベルの推定確信度の閾値をそれぞれ T_s, T_m, T_o とすると，例えば T_s は学習に使用したデータを推定モデル shModel-2 に入力したときの推定確信度の平均値である．

本稿では，推定確信度として Model-2 の最終出力層である Softmax 層の最大出力値を用いる．Softmax 層では全クラスの出力合計が 1 となり，出力は各クラスの確率を表す．たとえば，material レベルは 7 クラスの分類であるため，出力は [0.2, 0, 0, 0.7, 0.05, 0.05, 0] のように表され，最も高い 4 番目のクラスである 0.7 が推定確信度となる．

推定時には，Model-2 のそれぞれの推定モデルと T_s, T_m, T_s を用いて接触物体推定結果を出力する．特徴量抽出ブロックで得られた特徴量を shModel-2，matModel-2，objModel-2 に入力して得られたそれぞれの結果を S, M, O，それぞれの推定確信度を c_s, c_m, c_o とする．抽象度別接触物体ブロックは以下の手順で推定した接触物体を出力する．

(1) $c_o \geq T_o$ のとき:
接触物体として O を出力する．

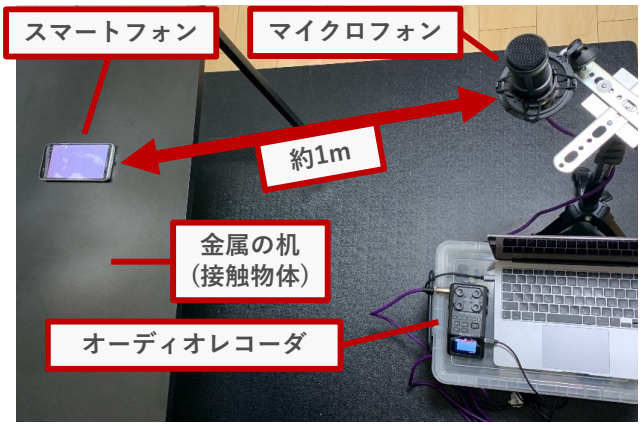


図 6: データ収集環境

(2) $c_m \geq T_m$ のとき:
接触物体として M を出力する．

(3) $c_s \geq T_s$ のとき:
接触物体として S を出力する．

(4) $c_s < T_s$ のとき:
データ収集からやり直して新しいデータで再度推定を行う．再度推定を行っても同様の結果が得られた場合は接触物体は推定不能と出力する．

接触物体が推定不能となった場合，ユーザには被覆状態とスマートフォンが存在する部屋の情報のみが提供される．

4. 評価

提案手法の有効性を検証するために，実環境でデータを収集し，評価を行った．評価では，ミクロな評価として抽象度レベルごとの推定モデルの推定精度を評価した上で，マクロな評価として推定確信度に基づいた接触物体推定の精度を評価した．さらに，学習データにない物体に対する推定精度を評価した．

4.1 評価環境

図 6 にデータ収集環境を示す．対象とする接触物体の上に ASUS Zenfone 8 スマートフォンを設置し，そこから約 1 m 離れた位置に audio-technica AT2050 マイクロフォンを設置した．マイクロフォンは ZOOM H6 オーディオレコーダに接続して音データを取得した．市販のスマートスピーカが複数の指向性マイクで構成されていることから，AT2050 マイクロフォンの指向性はスマートフォン方向に設定した．

図 7 に収集した接触物体を示す．図 7 に示す通り，接触物体として 21 物体を使用し，object レベルで 21 種類，material レベルで 7 種類，soft-hard レベルで 2 種類のラベルを付与した．スマートフォンは接触物体の上に表向きに設置した．Y シャツなどの小型の接触物体は机の上に設置し，その上にスマートフォンを設置した．

以下にデータ収集手順を示す．データ収集は各接触物体について 50 試行ずつ行った．接触状態やスマートフォン

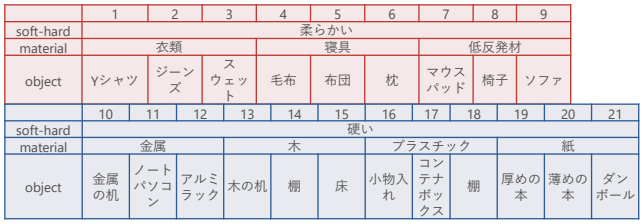


図 7: 収集した接触物体

の向きなどが異なるデータを収集するため，1 試行ごとにスマートフォンの位置と向きを変えながらデータを収集した．

(1) スマートフォンを接触物体の上に設置する
(2) 録音を開始する
(3) バイブレーションを起動する
(4) スマートフォンの位置と向きを変える
(5) (3)(4) を 50 試行繰り返す
(6) 録音を停止する

評価では，取得したデータから一部の試行を取り出して test データとした．残りの試行を train データ，validation データに分割して学習を行った後，test データを入力したときのラベルごとの F 値 (F-score) を算出して全ラベルの F 値の調和平均を推定精度とした．

提案する接触物体推定手法の有効性を示すため，以下の 2 つの手法で推定精度を比較した．

(1) 提案手法
3 章で示した提案手法である．抽象度を考慮したニューラルネットワークによって学習を行う，推定結果の出力を推定確信度に基づいて決定するという 2 つのアプローチを有する．推定確信度に基づき，objModel-2, matModel-2, shModel-2 のいずれかの出力を接触物体推定結果として出力する．

(2) SVM 手法
著者らの先行研究 [12] で示したスマートフォン接触物体推定手法である．特徴量として MFCC (Mel-frequency cepstral coefficients) を用い, SVM (Support Vector Machine) 分類器によって接触物体を推定する．10 分割交差検証を行い，各分割での推定精度の平均を推定精度とした．各抽象度レベルの推定モデルは個別に学習した．

4.2 抽象度レベルごとの推定モデルの推定精度

抽象度レベル間で推定モデルの情報を共有して学習することの有効性を検証するため，図 5 に示す Model-1, Model-2 の推定精度を評価した．収集したデータの試行をランダムに並べ替え，train : validation : test = 6 : 2 : 2 の割合となるように分割して学習・推定を行い，Model-1, Model-2 の推定精度を算出した．

図 8，**図 9** に，Model-1，Model-2 を用いた抽象度レベルごとの混同行列をそれぞれ示す．図 8 より，Model-1 では object レベル，soft-hard レベルで概ね正しく推定ができたことが分かる．一方，material レベルでマウスパッドや椅子などの低反発材を中心に柔らかい素材同士で誤って推定するケースが見られた．図 8 と図 9 を比較すると，Model-2 によってすべての抽象度レベルにおいて正しく推定できるケースが Model-1 よりも増えたことが分かる．特に，Model-1 で誤って推定するケースの多かった material レベルにおいても概ね正しく推定できた．

表 1: 抽象度レベルごとの推定モデルの推定精度

抽象度レベル	SVM	Model-1	Model-2
object	0.843	0.824	0.857
material	0.850	0.714	0.876
soft-hard	0.881	0.890	0.938

表 1 に抽象度レベルごとの推定モデルの推定精度を示す．表は SVM 手法の推定精度も示している．表 1 に示す通り，すべての抽象度レベルで提案手法である Model-2 の推定精度が最も高く，object レベルで 0.857，material レベルで 0.876，soft-hard レベルで 0.938 であった．特に，material レベル, soft-hard レベルでは SVM 手法, Model-1 と比べて大幅な精度の向上が見られた．これは各抽象度レベルの推定モデル構築の際に異なる抽象度レベルの推定モデルの情報を利用したことで少ない学習データでも効率的に推定モデルを構築できたことが要因と考えられる．

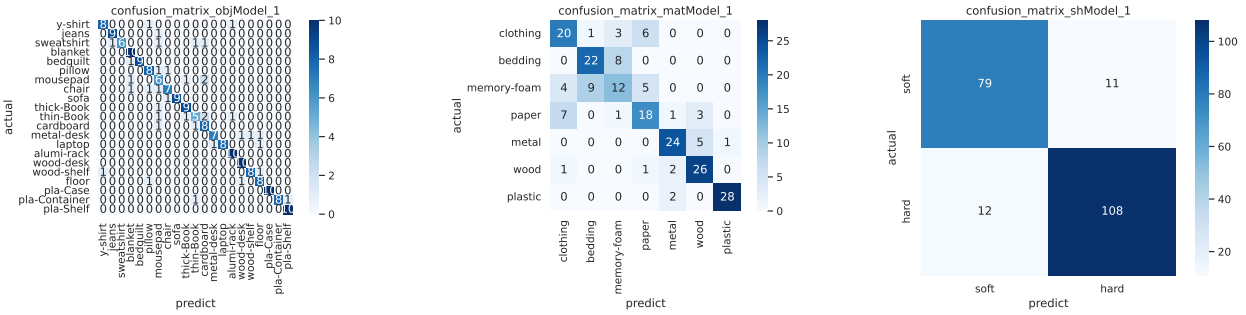
Model-1 は soft-hard 以外の抽象度レベルで SVM 手法よりも推定精度が低かった．学習データ量が少なく，推定に十分な特徴抽出ができなかったことが要因として考えられる．特に，柔らかい素材に対する識別は既存研究 [6, 7] でも課題として挙げられており，別の抽象度レベルの推定モデルの情報を用いる提案手法であれば柔らかい素材同士も識別できることが示唆された．

以上の結果より，異なる抽象度レベルにおける推定モデルの情報を学習に用いるというアプローチが接触物体推定の推定精度向上に有効であることが示された．

4.3 推定確信度に基づいた接触物体推定の精度

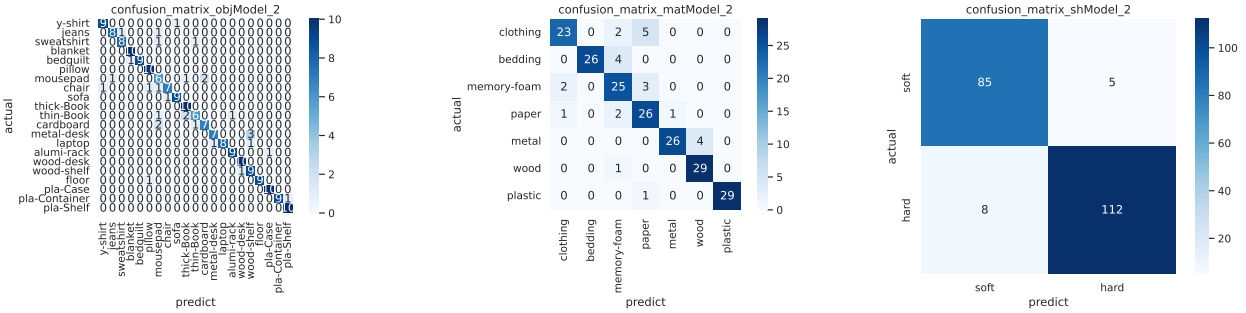
推定確信度によって出力を決定する手法の有効性を検証するため，推定確信度に基づいて出力を決定した場合の推定精度を評価した．表 1 に示した抽象度レベル別学習モデルの推定精度評価結果のうち，推定確信度が閾値以上であった推定結果のみを取り出し，その推定精度の平均を算出した．

表 2 に，推定確信度に基づいた接触物体推定の精度を示す．表には，「閾値以上の割合」として各抽象度レベルで推定確信度が閾値以上であった試行の割合も示している．推定確信度による出力決定によって object レベル，material レベル，soft-hard レベルの推定モデルの推定精度は，提案手法ではそれぞれ 0.966，0.778，1.000，SVM 手法ではそれぞれ 0.967，0.877，0.954 となった．推定確信度を考慮しない場合を示した表 1 の Model-2 および SVM 手法とそ



(a) object レベル **(b)** material レベル **(c)** soft-hard レベル

図 8: Model-1 による抽象度レベルごとの混同行列



(a) object レベル **(b)** material レベル **(c)** soft-hard レベル

図 9: Model-2 による抽象度レベルごとの混同行列

れぞれ比較すると，ほぼすべての抽象度レベルで提案手法，SVM 手法ともに推定精度を向上できたことが分かる．提案手法の material レベルのみ精度が低下しているが，これは material レベルで推定した試行，すなわち確信度が閾値以上であった試行が全体の 4.2%と少なく，モデルが誤って推定した試行が占める割合が大きくなったことが要因と考えられる．

表 2 において提案手法と SVM 手法の推定精度を比較すると，soft-hard レベルを除いて SVM 手法の方が高い推定精度を示している．しかしながら，SVM 手法では推定確信度が閾値以上となった試行割合は少なく，object レベルでは 43.5%である．提案手法では評価データの 84.3%で object の推定結果を出力し，SVM 手法とほぼ同じ推定精度が得られており，提案手法の高い有効性が確認できる．

以上の結果から，推定確信度に基づいて推定結果を出力する抽象度レベルを決定することで推定精度が向上し，特に，提案するニューラルネットワークを用いた場合に object レベルでの推定性能が大幅に向上することが確認された．提案手法においては推定不能であったデータは 5.8%であり，接触物体が不明となるケースは少なくなることも示唆された．

4.4 学習データにない物体に対する推定精度

学習データにない物体に接触している場合の接触物体推定性能を評価するため，学習データに含まれない物体で material レベル，soft-hard レベルでの推定精度を評価した．本評価は，Leave-One-Object-Out (LOOO) 交差検証により評価した．21 種類の物体のそれぞれについて，1 種類の物体のデータを test データとし，それ以外の物体のデータの試行をランダムに並べ替え，train : validation = 8 : 2 の割合となるように分割して推定精度を得た．

表 2: 推定確信度に基づいた接触物体推定の精度

抽象度レベル	推定精度		閾値以上の割合	
	SVM	提案	SVM	提案
object	0.967	0.966	43.5%	84.3%
material	0.877	0.778	19.2%	4.2%
soft-hard	0.954	1.000	15.4%	5.7%

まず，推定確信度に基づいた出力決定を行わない場合の SVM，Model-1，Model-2 の推定精度を評価した．**表 3** に，推定確信度に基づいた出力決定を行わない場合の学習データにない物体に対する推定精度を示す．提案手法，すなわち Model-2 の推定精度は material レベルで 0.521，soft-hard レベルで 0.872 であった．一方，SVM 手法の推定精度は material レベルで 0.330，soft-hard レベルで 0.812 であり，material レベル，soft-hard レベルの両方において提案手法によって推定精度を向上できたことが分かる．Model-2 と Model-1 の推定精度を比較すると，2 巡の学習によって構築される Model-2 によって推定精度を向上できたことが分かる．

表 3: 学習データにない物体に対する接触物体推定精度

抽象度レベル	SVM	Model-1	Model-2
material	0.330	0.483	0.521
soft-hard	0.812	0.813	0.872

表 4: 推定確信度に基づいた接触物体推定を行う場合の学習データにない物体に対する接触物体推定精度

抽象度レベル	推定精度		閾値以上の割合	
	SVM	提案	SVM	提案
object	-	-	21.0%	52.3%
material	0.338	0.505	20.0%	22.1%
soft-hard	0.740	0.848	25.0%	6.0%

次に，推定確信度に基づいた出力決定を行う場合の SVM，提案手法の推定精度を評価した．**表 4** に，推定確信度に基づいた出力決定を行う場合の学習データにない物体に対する推定精度を示す．推定確信度に基づく出力決定を行う場合，提案手法の推定精度は material レベルで 0.505，soft-hard レベルで 0.848 であった．一方，SVM 手法の推定精度は material レベルで 0.338，soft-hard レベルで 0.740 であり，material レベル，soft-hard レベルの両方において提案手法によって推定精度を向上できたことが分かる．

object レベルで推定確信度が閾値以上となった割合は，SVM 手法で 21.0%，提案手法で 52.3%であった．学習データにない物体に対する推定では object レベルで正しいラベルは含まれておらず，正しい結果を出力することはない．このため，推定確信度が閾値以上となった場合は誤った結果を出力することとなる．提案手法では約 50%の試行で誤った結果を出力することとなり，SVM 手法の約 20%と比べると誤って推定する割合が増加している．これは，推定確信度や閾値の算出方法が好ましくなかったことが要因と考えられる．実際，提案手法の object レベルの推定において推定確信度は約 40%の試行で 1.00 であった．推定確信度や閾値の算出方法について議論の余地があると言える．

以上の結果から，階層構造を持つニューラルネットワークによって抽象度を考慮して学習を行う提案手法によって，学習データにない未知の物体に対する推定精度を向上できることが確認された．

5. おわりに

本稿では，宅内スマートフォン捜索支援システムの実現に向けて，抽象度を考慮したスマートフォン接触物体推定手法を提案した．宅内スマートフォン捜索に向けた接触物体推定では，学習データに存在しない物体も対象となること，推定の誤りが捜索に大きな影響を及ぼすことを考慮する必要がある．そこで本研究では物体の抽象度を考慮した接触物体推定手法を提案した．本手法では，推定確信度によって出力する抽象度のレベルを切り替える，各抽象度レベルの推定モデル構築の際に異なる抽象度レベルの推定モデルの情報を利用する，という 2 つのアプローチを用いた．実環境で収集したデータを用いて提案手法の評価をした結果，21 種類の物体を推定精度 0.966 で推定できることを確認した．

今後の展望として，宅内スマートフォン捜索支援システムの構築に向けて，被覆状態も考慮したスマートフォン接触物体推定を行う予定である．

謝辞 本稿の研究の一部は，JSPS 科研費（JP21K11847）及び東北大学電気通信研究所共同プロジェクト研究の助成で行われた．

参考文献

[1] TrackR: 探し物に関する調査，https://prtimes.jp/main/html/rd/p/000000006.000022312.html (2017).

[2] 西 陽也，石田繁巳，村上友規，大槻信也：音響センシングによる宅内スマートフォン被覆状態分類の精度向上に向けた改善，第 31 回マルチメディア通信と分散処理ワークショップ論文集，pp. 144–151 (2023).

[3] Hwang, S. and Wohn, K.: VibroTactor: Low-Cost Placement-Aware Technique Using Vibration Echoes on Mobile Devices, pp. 73–74 (2013).

[4] Hasegawa, T., Hirahashi, S. and Koshino, M.: Determining Smartphone's Placement Through Material Detection, Using Multiple Features Produced in Sound Echoes, *IEEE Access*, Vol. 5, pp. 5331–5339 (2017).

[5] Ali, K. and Liu, A. X.: Fine-Grained Vibration Based Sensing Using a Smartphone, *IEEE Transactions on Mobile Computing*, Vol. 21, No. 11, pp. 3971–3985 (2021).

[6] Cho, J., Hwang, I. and Oh, S.: Vibration-Based Surface Recognition for Smartphones, *2012 IEEE International Conference on Embedded and Real-Time Computing Systems and Applications*, pp. 459–464 (2012).

[7] Darbar, R. and Samanta, D.: SurfaceSense: Smartphone Can Recognize Where It Is Kept, *Proceedings of the 7th Indian Conference on Human-Computer Interaction*, IndiaHCI '15, New York, NY, USA, Association for Computing Machinery, pp. 39–46 (2015).

[8] Wang, X., Li, S., Kallidromitis, K., Kato, Y., Kozuka, K. and Darrell, T.: Hierarchical Open-Vocabulary Universal Image Segmentation, *Advances in Neural Information Processing Systems*, Vol. 36, pp. 21429–21453 (2024).

[9] Novack, Z., McAuley, J., Lipton, Z. C. and Garg, S.: Chils: Zero-shot Image Classification with Hierarchical Label Sets, *International Conference on Machine Learning*, PMLR, pp. 26342–26362 (2023).

[10] 池村優佑，生尾夏輝，加藤空知，村上友規，藤橋卓也，猿渡俊介，渡辺 尚：3 次元構造物を対象としたマルチタスク Wi-Fi センシングに関する基礎検討，情報処理学会第 86 回全国大会， Vol. 3, pp. 145–146 (2024).

[11] Schmid, F., Koutini, K. and Widmer, G.: Dynamic Convolutional Neural Networks as Efficient Pre-trained Audio Models, *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, Vol. 32, pp. 2227–2241 (2024).

[12] 西 陽也，石田繁巳，村上友規，大槻信也：宅内でのスマートフォン捜索に向けた抽象度別接触物体推定手法の初期的検討，情報処理学会第 86 回全国大会， Vol. 3, pp. 155–156 (2024).