GO-Finder：手操作物体の発見に基づく事前登録不要のウェアラブル物探し支援システム

八木 拓真[1,a)]　西保 匠[1,b)]　川崎 邦将[2]　松木 萌[3]　佐藤 洋一[1,c)]

概要：私たちは置き場所を忘れた物の物探しに厖大な時間を費やしている．これまで，物体の位置を記憶・提示することでユーザの物探しを支援するシステムが複数提案されてきたが，先行手法は探す可能性のある物体を事前に登録しておく必要があり，ユーザの負担となっていた．本研究では，事前登録の負担を失くし，予期しない無くし物の探索にも使える，手操作に着目したウェアラブルな物探し支援システム–GO-Finder（“General Object Finder”）を提案する．GO-Finder は，首に装着したウェアラブルカメラで撮影した映像を入力として手で操作されている物体を検出・グループ化し，ユーザが触れた物体の視覚的な一覧を構築する．ユーザはスマートフォン上のインターフェースを通じて探したい物体を選び出すことによって各物体が最後に出現した瞬間を閲覧し，場所を思い出すことができる．ユーザ実験の結果，GO-Finder を利用することで正確かつより小さい認知負荷で物探しを行えることを確認した．

1. はじめに

私たちは物探しに厖大な時間を費やしている．物をどこに置いたのかを忘れて探し回ることは世代を問わず起こる問題であり [7]，ある調査では 1 年のうち 2 日半を物探しに費やしていると報告されている [9]．そのため，置き場所を忘れた物を探す作業に対する技術的支援が求められている．これまで，探したい物体にセンサを取り付ける等の手法が提案されてきたが，これらの先行手法はいずれも探したい物体は既知として事前にユーザに各物体をシステムに登録する作業を要求していた．

しかしながら，実際に失くす物は必ずしも事前に分かっているとは限らない．例えば上司から受け取った重要な書類や，昨日買ったばかりの商品など，無くし物は突発的に発生するが，先行手法はそうした突発的な探し物に対応できない．任意の物体の紛失に対応する素朴な方法として，ユーザの周辺に現れる全ての物体を自動登録することが考えられる．しかしながらこれは厖大な数の検索候補を生成し，限られた時間中に探したい物体を多数の候補から選び出すための負担が大きく，現実的でない．

1 東京大学 生産技術研究所
2 富士通研究所．本研究は富士通（研究所）と無関係に行われたものであり，その成果は富士通を代表するものではない．
3 ソフトバンク株式会社．本研究は所属とは無関係に行われたものであり，その成果は所属組織を代表するものではない．
a) tyagi@iis.u-tokyo.ac.jp
b) nisiyasu@iis.u-tokyo.ac.jp
c) ysato@iis.u-tokyo.ac.jp

そこで我々は，(1) 物探しの対象となる物体を手で操作するものに絞る (2) 物体の画像そのものを候補選択のクエリとして使用する の２つのアイデアに基づき，ウェアラブルカメラを用いた事前登録不要の物探しシステム–GO-Finder（“General Object Finder”）を提案する．多くの無くし物は手で持ち運ばれることに注目し，検索対象とする物体の候補を手で操作される物体に制限することで検索候補数を大幅に削減する．また，各物体に名前を割り当てる代わりに，物体画像の一覧を提示・選択するインターフェースを導入することで登録操作を省略する．

ある物体を探したいとき，ユーザはスマートフォン上のインターフェースを通じて物体サムネイル画像がその最終出現時刻順で並んだ**手操作物体タイムライン**から探したい物体を選択する（図 1）．その選択を基に，システムはその物体の最終出現時の画像を提示する．上記の処理は手操作物体の検出とクラスタリングからなる**手操作物体の発見**によって実現される．

GO-Finder の有効性を示すため，実際の物探しを模した実験室環境においてユーザ実験を行った．その結果，ユーザは GO-Finder を使うことで支援がない場合と比べ正確に物体の場所を思い出し，かつそれを少ない認知負荷で行えることを確認した．

2. 関連研究

2.1 物探し支援システム

ユーザの物探しを支援するためにこれまで RFID タグ [2]，

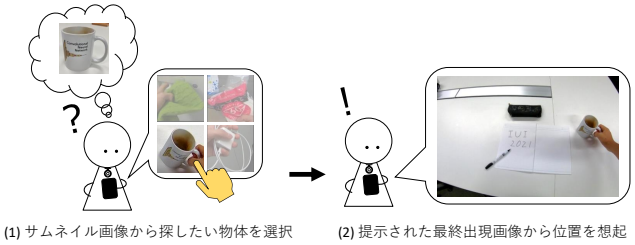

図 1 GO-Finder の概要.

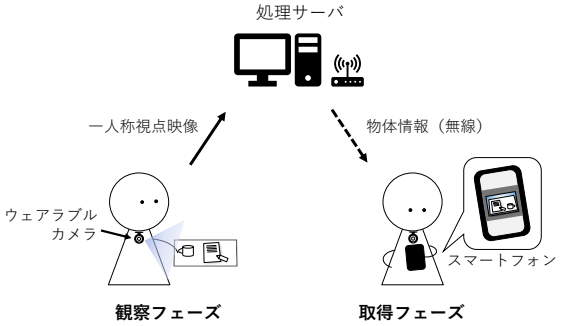

図 2 システム構成.

Bluetooth [5]，ウェアラブルカメラ [10] などの様々なセンサを用いたシステムが提案されてきた．なかでも，RFID タグを探したい物体に装着し，その位置を取得するシステムが多く取り組まれてきたが，タグが探索可能範囲内に存在する必要があり，探索可能範囲の外に出た場合に物体の場所を提示することができない欠点がある．また，位置情報をタグからの距離と方位により提示するためユーザには直感的でなく，タグと物体名との結び付けがボトルネックの 1 つとなっている．

一方，カメラベースのシステムは物体にセンサを装着する必要がなく，特にウェアラブルカメラはユーザの視点から見た映像を記録できるため，カメラを装着する必要がある代わりに使用場所を問わずユーザ周辺の物体を捕捉できるメリットがある．例えば，上岡らは，物体検出に基づく物探しシステム–I'm here!を提案している [10]．このシステムは頭部に装着した RGB カメラおよび赤外線カメラより登録済みの物体を検出し，それが最後に映像内に出現した際の映像を提示することでユーザに想起を促す．GO-Finder は記録・提示方法の点で I'm here!と類似しているが，I'm here!では事前に物体の見えを登録する操作を必要とするのに対し，GO-Finder は物体登録を自動で行う点で異なる．また，任意の手操作物体を自動で登録するため，突発的な無くし物にも対応できる点で優れている．

2.2 一般の記憶問題に対するカメラベースのシステム

カメラベースのシステムは過去の視覚情報を提示することで物探し以外の記憶の問題に対しても支援を行える．Hodges らは定期的に（例：30 秒に 1 回）画像を撮影し，それを振り返ることができるウェアラブルカメラ（SenseCam）を開発し，健忘症のユーザが過去を振り返る際に利用できることを実証している [4]．また，Li らはウェアラブルカメラと AR マーカを用いてエアコンや電灯のスイッチの状態を確認できる記憶補助システム（FMT）を提案している [6]．FMT が AR マーカにより限られた物体の状態確認に特化しているのに対し，GO-Finder はユーザが触る全ての物体の物探し支援を目的としている点で目的を異にする．

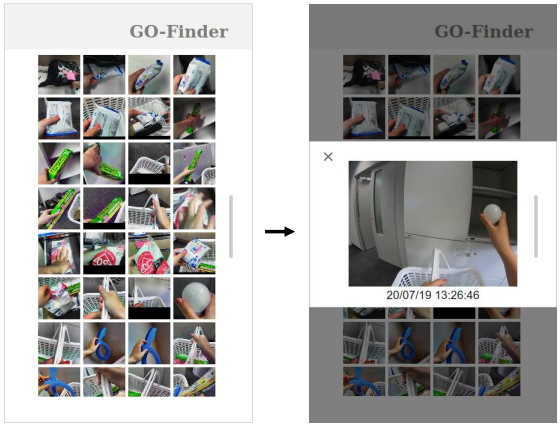

図 3 提案インターフェース.

3. GO-Finder

3.1 概要

本システムはウェアラブルカメラ，処理サーバおよび処理結果の閲覧のためのスマートフォンからなり（図 2），映像を記録・処理する観察フェーズとユーザがその結果を基に物探しを行う取得フェーズとに分けられる．

観察フェーズにおいて，ユーザは首にウェアラブルカメラを装着する．カメラはユーザの一人称視点映像を随時処理サーバに送信する．サーバは取得した画像列から手操作物体を検出・追跡し，その結果にクラスタリングを行うことで物体別のグループを発見する．

取得フェーズではユーザはスマートフォンに実装されたインタフェース（図 3）を用いてサーバの処理結果にアクセスする．まず，ユーザは物体のサムネイル画像が並ぶ手操作物体タイムライン（図 3 左）を走査し，探したい物体を選択する．サムネイルを選択するとその物体が最後に出現した瞬間を捉えた画像が拡大表示され，ユーザはその物体の位置を思い出すことができる（図 3 右）．

3.2 手操作物体の発見

GO-Finder はユーザの一人称視点映像から手操作物体を

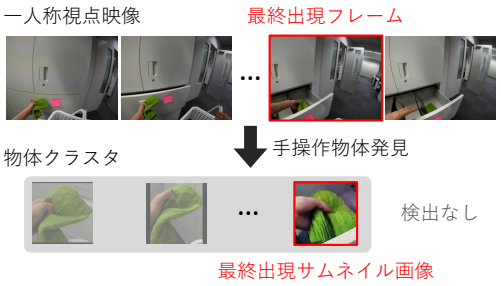

図 4 手操作物体の発見.

検出し，その集合から物体インスタンス別のグループを発見する．物体を種類別に分けることで，各物体が最後に出現した瞬間を提示することが可能となる（図 4）.

3.3 手操作物体タイムライン

GO-Finder は手操作物体のグループを発見し，それらを検索候補物体として自動的に登録する．その際，自動で物体が登録されるたびに探したい物体の画像に対して名前を付与するのは現実的でない．そこで本研究では，ユーザがサムネイル画像の一覧を直接閲覧することで対象物体を選択する**手操作物体タイムライン**を提案する．タイムラインには物体サムネイル画像の一覧が最新順でソートされたものが提示される（図 3 左）.

3.4 ポップアップ画面

手操作物体タイムライン中のサムネイル画像を選択すると，その物体が最後に現れた瞬間のフレームが拡大表示される（図 3 右）．ポップアップ画面は物体を最後に見た瞬間を捉えるため，ユーザは物体およびその周辺環境から物体位置を直感的に思い出すことができる.

4. アルゴリズムと実装

4.1 手操作物体の検出

Shan らが提案した手操作物体検出モデルを用いる [8]. これは，インターネット映像および既存の一人称視点映像データベースに対して手と物体が触れている瞬間の手と物体のバウンディングボックスのアノテーションを付与したものによって訓練された物体検出器である．画像を入力として，手のバウンディングボックス，手の接触状態（自己接触，他者，持ち運び可能な物体，静的な物体）および現在触れている物体のバウンディングボックスを出力する（図 5 (b) ）．今回の検索対象は持ち運び可能な物体のみであることから，接触状態が持ち運び可能な物体と判定された予測のみを処理対象とし，縦幅または横幅がフレーム辺の半分以上を占める検出はノイズとして除外した.

4.2 物体インスタンスの発見

続いて，検出したバウンディングボックス群をそのアピアランスによりインスタンス別のクラスタに分割する．具体的には追跡，局所マッチング，大域マッチングの 3 段階の処理からなる逐次クラスタリングを行う.

ステージ 1：フレーム間追跡

まず，隣接する検出に対し追跡器を適用する．もし追跡に成功した場合，結びついた検出を同じクラスタに割り当てる（図 5 (c)）．アピアランスベースの追跡器を用い [1]，追跡器の予測と検出との結び付けにはバウンディングボックス間の Intersection-of-Union（IoU）をスコアとするハンガリアンアルゴリズムを用いた.

ステージ 2：局所マッチング

追跡が途切れた場合，続いて最新の検出と既存クラスタとのマッチングを同じくアピアランスの比較により行う．各物体のバウンディングボックスについて学習済みの畳み込みニューラルネットワークから特徴量を抽出し，そのコサイン類似度によってクラスタにマージするか否かを決定する（図 5 (d)，上部）．具体的には，ImageNet で訓練を行った ResNet-50 [3] の最終層の手前の層より抽出した 2048 次元の特徴量を用いた．新規の検出に対し各クラスタの検出との類似度を計算し，その最大値と中央値がそれぞれ一定の閾値を超えた場合当該検出をマージする．もしいずれのクラスタともマージ条件を満たさなかった場合，新規にクラスタを生成する.

ステージ 3：クラスタ間の大域マッチング

ステージ 2 では単一の新規検出に対する局所的なマッチングを行ったが，物体の角度変化や検出領域の変動により新しいクラスタに分離する可能性がある．そうした過分割に対処するため，局所マッチングに加えてクラスタ同士の大域マッチングを行う．2 つのクラスタ（検出群）が与えられた際，ステージ 2 と同様に各検出間でコサイン類似度を計算し類似度行列を構成する（図 5 (d)，下部）．同様に類似度行列中の類似度の最大値と中央値を取り，両者が一定の閾値をそれぞれ超えた際マージ処理を行った.

ヒューリスティクスによる誤マージの防止

手領域およびテクスチャの影響を受け，異なる物体同士が大きな類似度を示し誤ったマージを行う場合があった．そこで，検出したバウンディングボックスまたはその組が次の条件を満たした場合にその組の類似度をゼロとすることで誤ったマージの抑止を行う.

- バウンディングボックスのアスペクト比：1 組の検出について，アスペクト比の組同士の比率が 1.5 倍以上
- 肌色領域の割合：各バウンディングボックスについてカラーヒストグラムを用いた肌検出を行なった際の肌色の割合バウンディングボックスの面積の 3 割以上
- バウンディングボックスの手との面積比：1 組の検出について，（物体領域の面積）/（手領域の面積）の比率が 1.5 倍以上

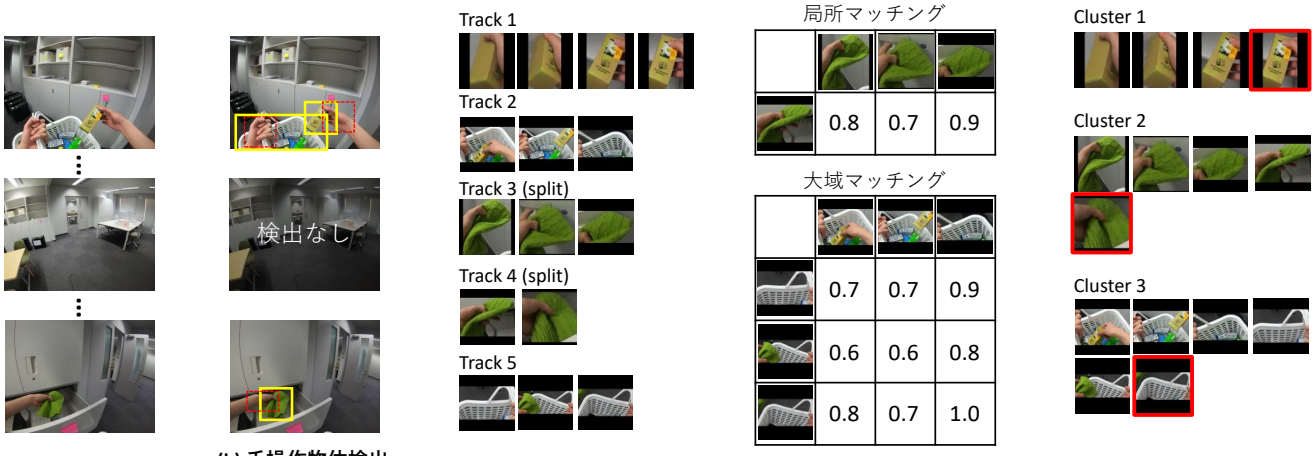

図 5 手操作物体発見アルゴリズムの概要.

実装の詳細

今回の実験ではすべての映像を 10 fps にダウンサンプルし，処理解像度を VGA に統一した．類似度の最大値の閾値は 0.8，中央値の閾値は 0.7 に設定した．

5. ユーザ実験

(i) 手物体発見アルゴリズムが正しく物体を発見できたか (ii) ユーザが GO-Finder を用いて物探しを行えたか を検証するために実験室環境におけるユーザ実験を行った．

5.1 実験手順

12 人（男性 10 人，女性 2 人，18–28 歳）のボランティアが実験に参加した．実験は研究室の居室内で全て行われた．まず，参加者はいくつかの物体を提示され，それを室内の配置地点に自由に隠すよう指示された（配置フェーズ，図 6 左を参照）．続いて場所を移動し，100 ます計算を一定時間解いてもらった（忘却フェーズ）．その後元の部屋に戻り，隠した物体の一部を探すタスクを行った．試行は実験条件を変えながら 3 回行われ，全試行終了後アンケートとインタビューを受けた．システムを使用するかどうかは参加者の自由に委ねられた．

5.2 実験条件

次の 3 つの条件で行った：

- **支援なし**：自分の記憶を頼りに物探しを行う．
- **画像ベース**：配置時の映像を 5 秒に 1 回サンプルした画像列（図 6 右）を提示される．
- **物体ベース（GO-Finder）**：提案システム．

画像ベース条件は SenseCam [4] などの自動撮影デバイスを参考にしたもので，過去の画像列を提示することで記憶を想起できることを企図した．各試行で参加者は 16 個の物体を 20 か所の地点のいずれかに配置し，忘却フェーズ後にうち 6 個を取ってくるよう指示された．図 6 中央に物体例を示す．試行毎に使用物体は入れ替えた．

5.3 システム評価指標

システム単体の評価指標として位置取得率を算出した．物体を配置していない第三者が提案システムのみを用いて各物体の位置を特定できる物体数の割合で定義され，システムがどれだけ正確に対象物体を発見できたかを示す．

5.4 ユーザ評価指標

検索の正確さ

客観指標として参加者の各試行時の適合率を算出した．具体的には参加者が設置個所を確認した回数のうち，正解物体を回収した回数の割合を算出した．2 条件の各組合せについて平均適合率の差の対応あり t 検定を行った．

アンケートとインタビュー

3 試行全ての完了後，参加者に各条件における体験をアンケートにて集計した．まず，「物探しのタスクはどれくらい難しかったか」を 7 点法で回答してもらい，平均値の差についてウィルコクソン符号順位検定を行った．また提案システムのインタフェースについて，図 8 に示す各質問に 5 点法で回答する形で集計した．6・7 問目については画像ベースと物体ベースの各条件について回答・集計した．また，アンケート回答後にインタビューを行い，各条件でどのように物探しを行ったか，システムで使いやすい・使いにくいと思った瞬間がなかったかを尋ねた．

6. 実験結果

6.1 システムの評価

表 1 に手物体発見アルゴリズムの位置取得率を示す．12 人分の試行に対して手物体発見を行った際の各物体セットにおける平均位置取得率は 84.9，83.3 および 88.5%であっ

図 6　左：配置の様子．中央：使用された物体例．右：画像ベースシステムのタイムライン例．

表 1　各物体セットに対する位置取得率（%）．

	平均 ± 標準偏差	最小	最大
セット 1	84.9 ± 7.8	68.8	93.8
セット 2	83.3 ± 9.7	68.8	100
セット 3	88.5 ± 10.9	62.5	100
全セット	85.6 ± 9.6	62.5	100

表 2　物探しの正確さ（適合率）．

	平均 ± 95% 信頼区間
支援なし	0.728 ± 0.174
画像ベース	0.736 ± 0.124
物体ベース（GO-Finder）	0.922 ± 0.084

表 3　平均適合率に関する対応あり t 検定の検定結果．

			95% 信頼区間		効果量
	t	p	下界	上界	d
支援なし/画像ベース	-0.104	0.918	-0.193	0.175	0.04
支援なし/物体ベース	-2.012	0.069	-0.407	0.018	**0.95**
画像ベース/物体ベース	-2.339	**0.039**	-0.360	-0.011	**1.17**

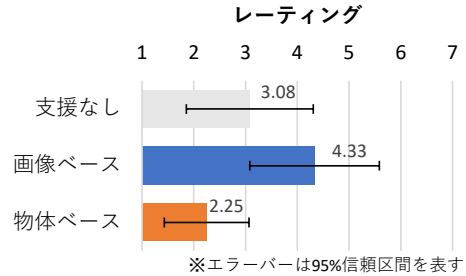

図 7　タスク難易度の評価結果 (簡単=1，難しい=7)．

表 4　タスク難易度に関するウィルコクソンの符号順位検定の結果．

			95% 信頼区間		効果量
	Z	p	下界	上界	r
支援なし/画像ベース	-1.857	0.063	-2.581	0.081	**0.38**
支援なし/物体ベース	-1.501	0.133	-0.596	2.263	**0.30**
画像ベース/物体ベース	-2.027	**0.043**	0.540	3.627	**0.41**

た．これは，GO-Finder が平均で 16 個中 13-14 個の位置を正しく提示できたこと，物体種類を入れ替えたことに起因する性能差がなかったことを示している．

一方，物体間ではスコアに差が見られた．マグカップ，ボンド，電球などの物体が全試行で追跡できたのに対し，一部の物体は低いスコアに留まり（栓抜き：16.7%，薬瓶：58.3%，黒財布およびスプレーボトル：66.7%），大きさが小さく黒色の物体の発見に苦戦する傾向が観察された．

6.2　物探しの正確さ

表 2 および表 3 に結果を示す．期待通り，物体ベース条件は他の 2 条件と比べ高い適合率を全参加者間で示した．対応あり t 検定は画像ベース条件と物体ベース条件との間にのみ有意差を示した（それぞれ $p = 0.918$，$p = 0.069$ および $p = 0.039$）．しかしながら，物体ベース条件は他の 2 条件と比較して高い効果量を示し（それぞれ $d = 0.95$ および $d = 1.17$），GO-Finder に物探しの正確さの向上効果があることが示唆された．一方，支援なしと画像ベースシステムの間では効果は見られなかった（$d = 0.04$）．

6.3　アンケート

物探しタスクの難易度

図 7 および表 4 にタスクの難易度評価に関する結果を示す．今回の実験では画像ベースのシステムが最も難しいと評価され，物体ベースのシステムが最も簡単であると評価された．ウィルコクソンの符号順位検定の結果，画像ベース・物体ベース間でのみ有意差を示した（$p = 0.063$，$p = 0.133$ および $p = 0.043$）．しかしながら，全組合せにおいて中程度の効果量を示しており（$r = 0.38$，$r = 0.30$ および $r = 0.41$），物体ベースシステムは支援なし条件との比較においても認知負荷を軽減することが示唆された．

インターフェースの機能

図 8 にアンケートの各質問に関する回答結果を示す．Q1 から Q5 について，参加者は概ね肯定的な反応を示した．また，ウィルコクソンの符号順位検定は Q6 について有意差を示した（$p = 0.007$）が Q7 では示さなかった（$p = 0.065$）．しかしながら両質問はそれぞれ大きい/中程度の効果量を示し（$r = 0.55$ と $r = 0.38$），GO-Finder が画像ベースのシステムに対し物探しに有効であることを示した．

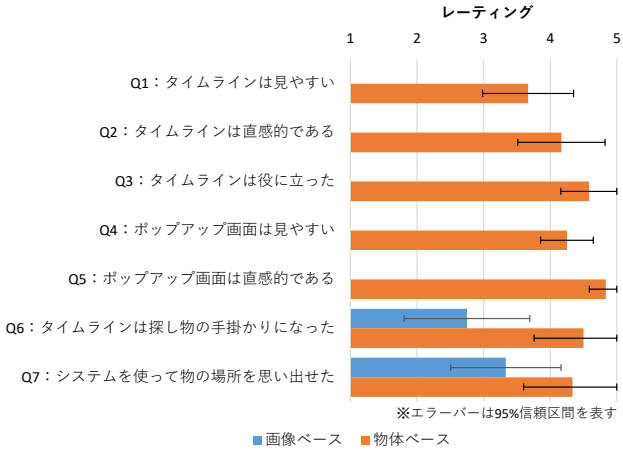

図 8 アンケート結果.

6.4 観察とフィードバック

多くの試行において，参加者は覚えていた物体から順に探索を行い，自信がない場合に GO-Finder を使用した．GO-Finder の使用時，彼らはサムネイル一覧から探したい物体を選び出し，ポップアップ画面を確認してその位置を把握し，正確に物体を取得することに成功した．

インタビューでは，12 人中 11 人が GO-Finder が便利だったと答え，主に手操作物体タイムラインの直感性を評価した：「一覧には欲しいものが映っていて，押したら置いた場所も映っていたので助けになった」．また，参加者はより確信をもって物探しを行えた：「わからなくなったらすぐにスマホを見るようになった．それによって確信がちゃんと持てることが多かった」．

また，複数の参加者は物体のサムネイル画像による提示を好ましいと評価したが，手操作物体タイムラインの見やすさ（Q1）について，クラスタの過分割による混乱が見られたためやや低い評価となった．

7. 議論

ユーザ実験の結果，参加者は GO-Finder を使用してより小さい認知負荷で物探しを行うことができた．一方，画像ベースのシステムは物探しタスクにおいては有効性を示さなかった．過去の出来事の想起とは異なり，物探しタスクは速やかに解決されるべき問題であり，ユーザが物体位置に関する確信的な情報を要求したためと考えられる．システム単体においても，手操作物体発見アルゴリズムがユーザの配置物体の位置を概ね取得できることが示された．

参加者は手操作物体タイムラインを使っての検索に素早く適応し，物体画像そのものを用いた検索の有効性が示された（Q3）．また，サムネイル画像はユーザの視野から撮影されたものであり，タイムラインそのものを触った物体の履歴として使用できたことから，タイムラインそのものも物体を置いた場所を思い出す手掛かりとなった（Q6）．

8. 結論

事前登録不要のウェアラブル物探し支援システム，GO-Finder を提案した．手操作物体に注目した自動発見と物体画像を用いた検索の 2 つのアイデアの導入により，任意の手操作物体の物探しの支援を実現した．ユーザ実験の結果，GO-Finder を使用することによる物探しタスクの正確さの向上と認知負荷の軽減が確認された．手操作物体タイムラインの利用により，ユーザは自動登録された物体候補の中から探したい物体を直感的に選択し，物体位置を思い出すことができた．GO-Finder は突発的な無くし物にも対応できるため，広範な状況における物探し支援が期待できる．

参考文献

[1] Bertinetto, L., Valmadre, J., Henriques, J. F., Vedaldi, A. and Torr, P. H.: Fully-convolutional siamese networks for object tracking, *Proceedings of European Conference on Computer Vision*, Springer, pp. 850–865 (2016).
[2] Borriello, G., Brunette, W., Hall, M., Hartung, C. and Tangney, C.: Reminding about tagged objects using passive rfids, *Proceedings of ACM International Conference on Ubiquitous Computing*, Springer, pp. 36–53 (2004).
[3] He, K., Zhang, X., Ren, S. and Sun, J.: Deep residual learning for image recognition, *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pp. 770–778 (2016).
[4] Hodges, S., Williams, L., Berry, E., Izadi, S., Srinivasan, J., Butler, A., Smyth, G., Kapur, N. and Wood, K.: Sensecam: A retrospective memory aid, *Proceedings of ACM International Conference on Ubiquitous Computing*, Springer, pp. 177–193 (2006).
[5] Kientz, J. A., Patel, S. N., Tyebkhan, A. Z., Gane, B., Wiley, J. and Abowd, G. D.: Where's my stuff? design and evaluation of a mobile system for locating lost items for the visually impaired, *Proceedings of the 8th ACM Conference on Computers and Accessibility*, pp. 103–110 (2006).
[6] Li, F. M., Chen, D. L., Fan, M. and Truong, K. N.: Fmt: A wearable camera-based object tracking memory aid for older adults, *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, Vol. 3, No. 3, pp. 1–25 (2019).
[7] Peters, R. E., Pak, R., Abowd, G. D., Fisk, A. D. and Rogers, W. A.: Finding lost objects: Informing the design of ubiquitous computing services for the home, Technical Report GIT-GVU-04-01, Georgia Institute of Technology (2004).
[8] Shan, D., Geng, J., Shu, M. and Fouhey, D. F.: Understanding human hands in contact at internet scale, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 9869–9878 (2020).
[9] Technology, P.: The nation's biggest lost and found survey, by pixie, `https://bit.ly/2Kd8072`, archived: 2017-12-06 (2017).
[10] 上岡隆宏，河村竜幸，河野恭之，木戸出正継：I'm here!: 物探しを効率化するウェアラブルシステム，ヒューマンインタフェース学会論文誌，Vol. 6, No. 3, pp. 275–285 (2004).