特集号
招待論文

人間共生ロボット"EMIEW2"の対話型物体情報検索―実環境内の物体への自動タグ付けシステムの実践―

平松 義崇[†1]　住吉 貴志[†1]　永吉 洋登[†1]　渡邉 裕樹[†1]　廣池 敦[†1]
浜田 宏一[†1]　影広 達彦[†1]

[†1] (株) 日立製作所

本稿では，ロボットとの対話型物体情報検索の実現のため，実環境の物体画像に対して Web 画像データとの類似画像検索によるタグ付け技術を適用した事例と得られた知見について述べる．従来通り，類似画像検索結果の上位 N 件分の画像に付随するテキスト集合から統計的な指標で単語を抽出した場合，検索結果にクエリと同じカテゴリの画像が少ないときに単語抽出性能が低くなる課題が判明した．これに対して，類似画像検索結果の分析から得られた（1）第 1 位の類似度から一定の距離以内かつ，（2）検索結果から風景画像と判定された画像を除外した検索結果だけを用いる対策により単語抽出性能が改善した．また，本タグ付け技術を用いることで，利用者が知らない物体の略称や型番など物体に関連する単語を抽出できた．

1. はじめに

近年の音声認識・言語処理の進化により，スマートフォンや知覚機能を持つロボットとの音声対話による情報検索の実用化が進んでいる．現在はWeb上の情報検索が実用化されているが，今後は，実際の環境に存在する物体の情報検索の実用化が進むと考えられる．

実環境の物体に対して音声対話による情報検索を行うには，物体に関する情報の付与が必要である．この作業は膨大な労力を要するため，物体に関する情報を自動で付与できるシステムが好ましい．このシステムには，実環境に存在する任意の物体への対応や，クエリと同じテキスト形式かつノイズの少ない情報の付与が求められる．

筆者らは，これらの要件を満たすための物体への自動タグ付けシステムの開発を進めている．本稿では，事例データとの類似画像検索ベースのタグ付け技術[1]を本システムに適用した事例と，そこから得られた知見について述べる．

類似画像検索ベースのタグ付け技術は，Webからの画像データと付随するテキストを大量に集めてデータベース化しておき，そこから直接，画像検索した結果を解析することで，事前の定義なしにテキスト形式でタグを付与する．画像検索で得られる結果にはクエリと異なるカテゴリのデータが含まれるため，検索結果の解析において上位100件前後の検索結果に対して統計処理を行うことで適切なタグを抽出する．

この類似画像検索ベースのタグ付け技術に，実環境で撮影された物体画像を適用した場合，適切なタグが抽出できない事例があった．そこで，さまざまな物体に対して類似画像検索結果を分析した．その結果，適切なタグを得るために必要な検索結果の選定方法について新たな知見が得られた．また，タグ付け技術の特徴の1つであるWebの情報を利用したことによる想定外の好事例も得られた．

以降の章では，第2章にて，筆者らが開発中のロボットとの音声対話による物体情報検索システムについて述べる．次章からこのシステムを対話型物体情報検索システムと称する．第3章にて，類似画像検索ベースのタグ付け技術を適用して得られた新たな課題を述べる．第4章にて，第3章で述べた課題への対応策とそれを適用した結果について述べ，筆者らが試作している人間共生ロボット“EMIEW2”を用いて音声対話による物体情報検索を実行した結果を示す．第5章にて物体自動タグ付けシステムの今後の展開について述べる．

2. 対話型物体情報検索システム

物体情報検索システムは，物探しや物体管理での活用が期待されている[2],[3],[4],[5]．これらのシーンでは，

たとえば，生活空間でのテーブルや椅子などに置かれた日用品や，オフィスにおいて机上にある物品などの物体を主対象としている．筆者らが開発した対話型物体情報検索システムも，まず，机上に存在する物体を対象に開発を進めた．

本システムの全体構成を図1に示す．まず，物体タグ付けシステムが環境内のカメラ映像から机上に新たに現れた物体を検出して，自動的にタグを付けて物体管理データベースに保存する．このとき，得られたタグ情報を音声対話システムに送信し，音声認識の内部モデルの学習に用いる．

ロボットへの音声発話を受けて，音声対話システムが言語情報を出力し，その言語情報からクエリを生成する．物体検索エンジンがクエリから対象物体を特定し，特定された物体に関する情報に基づいてロボットへの指令を生成し，知覚機能と移動機能を有するロボットが指令を実行する．

物体タグ付けシステムの詳細構成と，本システムで用いられる人間共生ロボット“EMIEW2”と，音声対話システムの概要については以降にて述べる．

2.1　物体タグ付けシステム

物体タグ付けシステムの全体構成を図2に示す．本システムは，まず，室内に設置された校正済みのカメラで撮影された映像から，物体領域を検出する．続いて，検出された物体領域画像に対して，類似画像検索ベースのタグ付け技術[1]によりタグを複数抽出し，物体領域画像，世界座標と関連付けて物体管理データベースに保存する．

2.1.1　物体領域検出

物体領域検出に必要な要件と対応方法について述べる．必要な要件は（a）事前学習なしの検出，（b）物体単体の領域検出，である．要件（a）については，第1章で述べた任意の物体への対応に必要な要件である．要件（b）については，今回，物体単体だけが映る領域画像を検出し，そこからタグを抽出するために設定した要件である．この流れを採用した理由は，複数の物体が映る画像に対して直接タグを付与することは最新の画像アノテーション[6]においても実用レベルではないためである．

以上の要件に対応するため，時間的に異なる2枚の画像間の差分によって領域を検出する．この方法は事前の学習なしに検出できる利点がある一方，物体単体領域を高精度に検出するには，2枚の画像を適切に選択することが必要となる．

これに対しては，物体が机上に置かれる際のイベントに基づく方法を検討した．そのイベントは，（1）人が机の周辺に移動してくる，（2）物体を机上に置く，（3）その周辺に一定時間とどまる，（4）机から去る，の4つである．物体の検索が行われるのは人が物体から離れた（4）の後である．イベント（4）の後の机上画像には，たとえば，図3（a）のように物体が置かれた様子が映

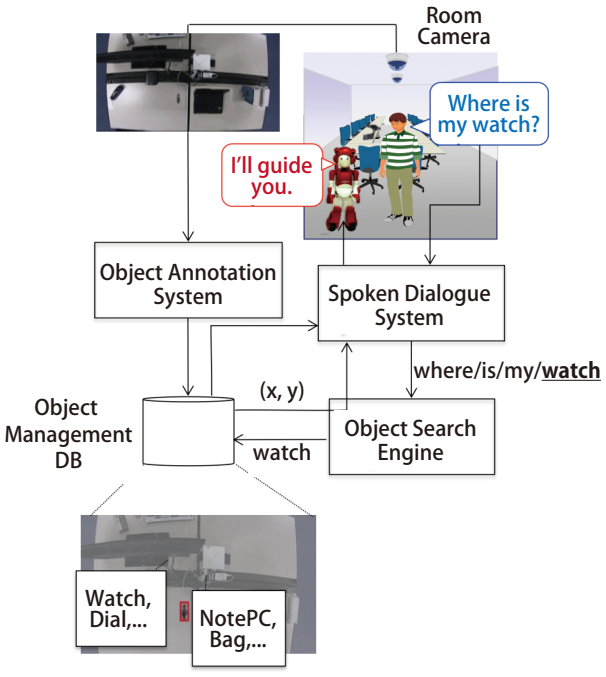

図1　全体構成

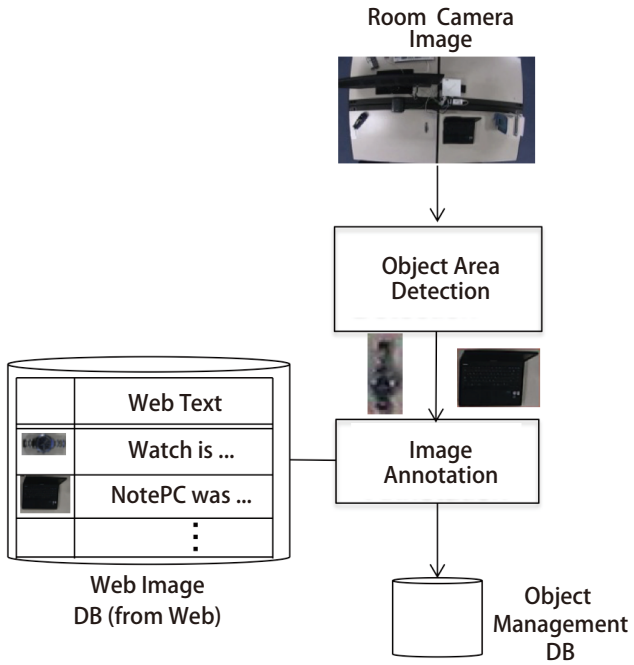

図2　物体タグ付けシステム

っているはずである．この画像を選択し，物体が置かれる前の机上画像の差分をとることで，物体領域だけを検出できる．物体が置かれる前の画像は，たとえば，図3（b）に示すようにイベント（1）が生じる前に保存した画像を選択する．

上記の選択方法に基づく物体領域検出[7]を図4に示す．本物体領域検出は，まず，画像の動き量に基づいてイベント（1）とイベント（4）に相当する個所を検出する．続いて，検出された個所の前後の画像同士の差分により，抽出された領域を物体領域として検出する．

2.1.2　タグ抽出

まず，タグ抽出に必要な要件を述べる．対象となる物体は多種多様であり，また，将来，新たな物体が登場するため，(a) 学習データを事前に定義しない方法が望ましい．人からの発話内容がクエリとなるため，(b) 付与されるタグは物体の名称などのテキスト形式である必要がある．前述の通り，抽出されたタグは音声認識の内部モデルの学習に使用されるため，(c) ノイズの少ない抽出が求められる．

以上の3つの要件のうち特に（a）（b）に対応するため，図5に示すような類似画像検索ベースのタグ付け技術[1]により物体に関連する単語をタグとして付与する．このタグ付け技術は，Web画像データベースの構築とアノテーション処理の2つの処理に分けられる．

Web画像データベース構築のフローは，WWWをクローリングして自動的に取得したWebページから，画像とその周辺にあるテキストを抽出し，それらを関連付けてデータベースに保存する．周辺にあるテキストとして，たとえば，imgタグの属性や前後のテキストなどを用いる．このようなルールで抽出されるテキストは，必ずしも画像を説明するものではないが，画像に関連する単語が含まれる可能性は高い．本システムで使用するWeb画像データベースには，2008年から約3年間の自動クローリングにより得られた約1億件の画像とテキストが格納されている．

アノテーション処理のフローは，まず，入力された画像をクエリとして，Web画像データベースに格納されている画像群から類似画像検索を行う．類似画像検索は，画像の色やエッジなどの特徴に基づいた検索であり，検索の結果，入力画像と「見た目」の類似した画像が得られる．文献[1]では（1）色特徴量，（2）形状特徴量，（3）色＋形状特徴量1，（4）色＋形状特徴量2の4種類の特徴量が提案されている．

（1）色特徴量は，RGBカラーヒストグラムである．（2）形状特徴量は，28種類のエッジパタンのヒストグラムである．（3）色＋形状特徴量1は，色と形状の特徴量を連結したものである．色に関しては画像全体のRGBカラーヒストグラムに加えて，構図分割で領域ご

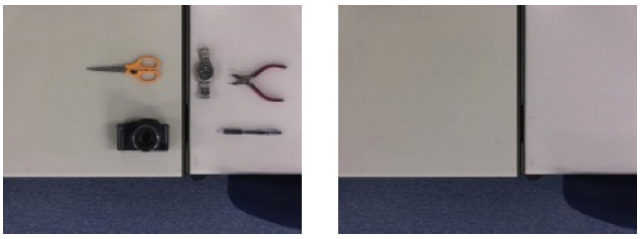

図3　机上のカメラ画像の例

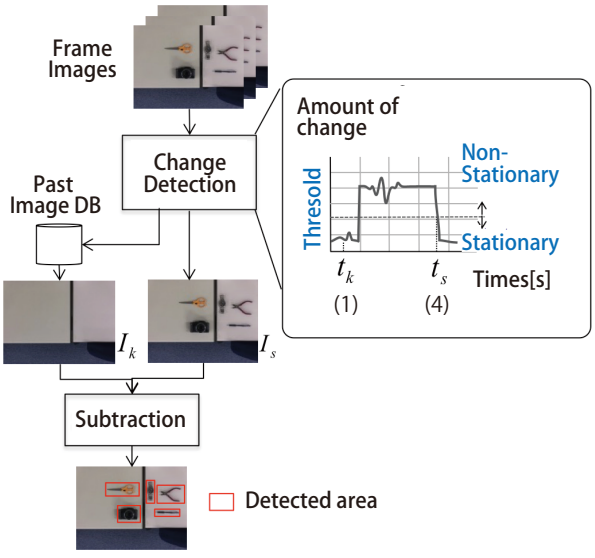

図4　物体領域検出

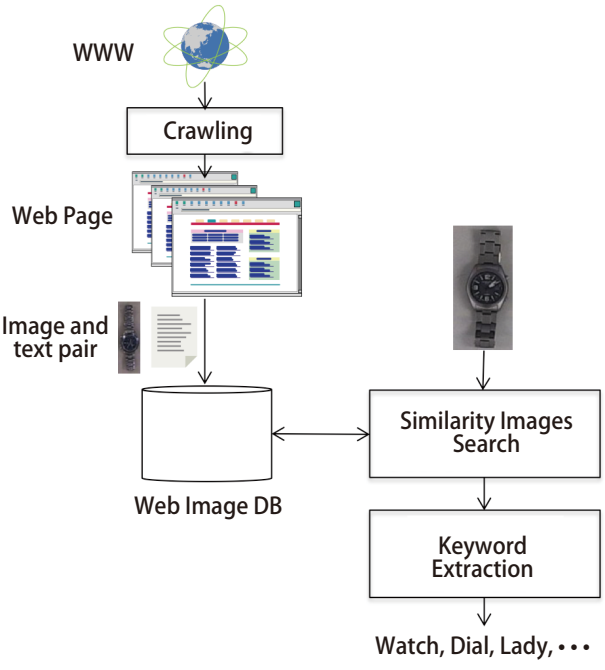

図5　類似画像検索ベースのタグ付け技術

とに集計したヒストグラムも用いる．形状に関しては（2）と同じ特徴量を用いる．（4）色＋形状特徴量2は，色と形状と画像のアスペクト比を連結したものである．色に関しては構図分割で領域ごとに集計したヒストグラムだけ用いる．形状に関しては（2）と同じ特徴量を用いる．

続いて，検索結果の各画像から付随するテキストを抽出することで，テキスト集合を取得する．Web画像データベースには画像とテキストが関連付けて保存されているため，検索結果の画像から付随するテキストを抽出できる．

得られたテキスト集合から，入力された画像を特徴付けるような重要単語を抽出する．重要単語の抽出では，テキスト集合に含まれるすべての単語に対して，統計的な指標によるスコアでソーティングし，その上位またはスコアがしきい値以上の単語を出力する．この際，ノイズを除去するため，Web画像データベース内での単語の頻度（Document Frequency: DF）が所定の範囲内の単語のみを出力する．また，明らかに物体に無関係な冠詞，助詞，代名詞も合わせて除去する．

文献[1]のタグ付けの評価には，犬，猫，ひまわりなどの物体が映った画像や，サンセットなど，特定のシーンの画像など，合計30種類の画像が用いられている．そのタグ付け性能は，物体やシーンの名前を第10位以内に抽出できたときの正解率で平均約59.1%であった．このとき，類似画像検索には（4）色＋形状特徴量2の特徴量が用いられ，重要単語の抽出には，類似画像検索の類似度に応じて単語頻度（Term Frequency: TF）

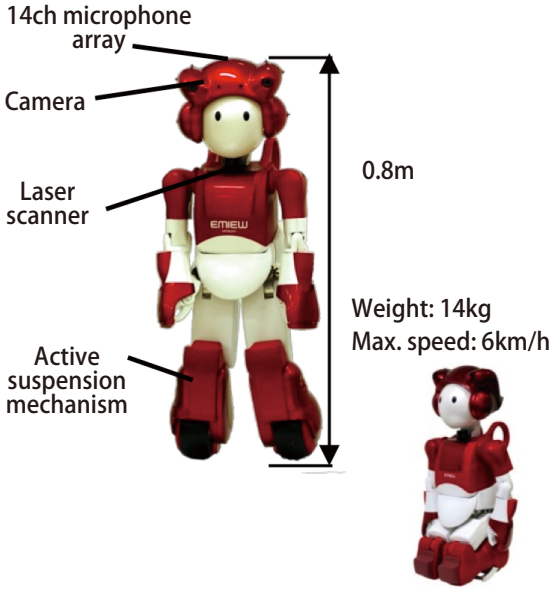

図6　"EMIEW2" の仕様

を重み付けたときのKL（Kullback-Leibler divergence）が用いられている．

2.2　人間共生ロボット “EMIEW2”

“EMIEW”は“Excellent Mobility and Interactive Existence as Workmate”の略であり，機敏な移動と対話ができる存在がコンセプトである．2007年に発表した“EMIEW2”はオフィスビルなどの屋内環境で人と共生することを目指して開発された．

図6に示すように，身長80[cm]，体重14[kg]となっており，もし人間と衝突しても与える被害が小さい．二輪により，人と同等の移動速度である1.6[m/s]を実現し，停止時や作業時は安定な四輪姿勢に変形する．

ロボット本体には，外界測定用のセンサとして，カメラ，マイク，レーザスキャナが搭載されている．カメラは頭部に2台設置されており，人の挙動の認識や目の前に提示された物体の認識に用いられる．頭部には14個のマイクが装備され，音源の位置の推定や，雑音にロバストな音声認識に用いられる．首の部分にレーザスキャナが装備され，事前に計測した部屋の距離データと比較による自己位置推定に用いられる．

“EMIEW2”は小型軽量なため，機体の基本的な制御はロボット本体で行う一方，計算負荷の高い認識や知能の処理はサーバ群で行う．これを実現するため，ロボット本体に搭載されたセンサ群から採取したデータは，無線によりサーバ群に伝送される．このような構成はリモートブレイン構成と呼ばれる．音声対話システムは，サーバ群の中に導入されている．

今回，このサーバ群に物体タグ付けシステムと物体検索システムを導入し，音声対話システムと連携することで，ロボットとの音声対話による物体情報検索を実現した．

2.3　音声対話システム

音声対話システムの全体構成を図7に示す．本システムは，まず，ロボットの複数のマイクで収録した音声信号から音源を分離し，さらに雑音を除去して，発話者の方向の音声信号を抽出する．抽出された音声信号に対して，人の発話が含まれる音声区間を検出し，検出された区間に対して音声認識により言語情報が得られる．得られた言語情報から言語理解により発話内容を推定し，言語意味表現を抽出する．言語意味表現とは，音声対話システムが持つ内部状態を更新するために用いられる情報である．今回，推定する内容は，ユーザ

からの検索要求があったか否かであり，要求があった場合は検索対象の言語表現の部分を抽出する．続いて，指令生成処理が行われる．具体的には，抽出された言語表現をクエリとして，物体検索システムから対象物体を特定して，その位置情報に取得して，ルールに基づいて指令を生成し，ロボットに送信する．以降では，音声認識について詳細を述べる．

2.3.1　音声認識

音声認識は，音声信号を音響モデルと言語モデルに照合することで，入力音声信号に対する事後確率が最大になる単語列を見つける問題として定式化されている．アプリケーションに必要な性能を実現する要素は，音響モデルと言語モデルの2つである．

音響モデルは，音素ごとの特徴量の分布の確率モデルであり，利用する音響環境に依存する．今回の場合は，ロボットに搭載されたマイクで収録した音声信号を用いて一度学習しておけば問題ない．

一方，言語モデルは，出現が想定される単語列の相対頻度を記憶した確率モデルであり，アプリケーションに応じたデータを用意する必要がある．言語モデルの作成方法は，アプリケーションで想定される発話テキストを大量に用意して学習する．

今回のアプリケーションは物体情報検索である．このアプリケーションで想定される発話内容は，環境内に存在する物体への問合せである．たとえば，環境内

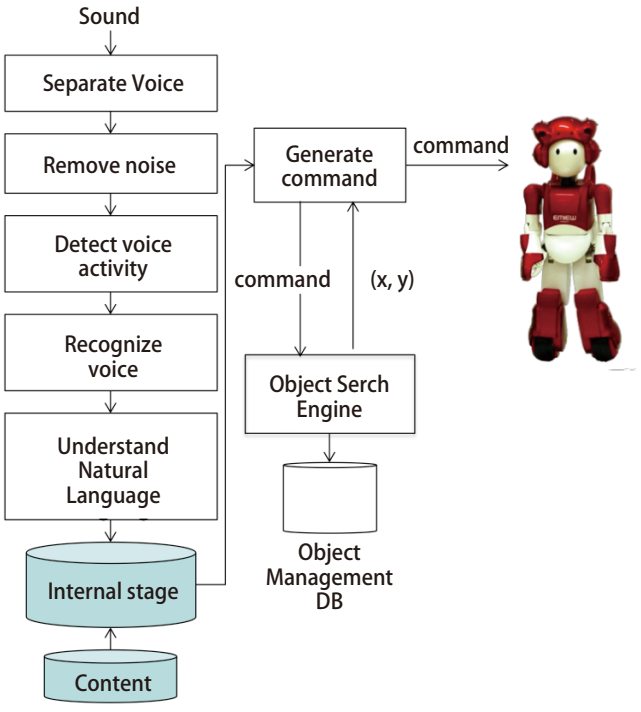

図 7　音声対話システム

に時計を置き忘れたときに，「時計はどこにありますか？」のような問合せである．このようなテキストを環境内に存在し得る物体ごとに用意すればよいが，そのような物体をすべて挙げることは現実的ではない．特に，将来，世の中に新たに登場する物体を予測することは不可能である．

これを解決するために，物体自動タグ付けシステムによってタグとして抽出された単語を利用する．物体自動タグ付けシステムが抽出した単語に対する問合せ文章を自動生成し，得られた文章を用いて言語モデルを学習する．

3.　類似画像検索ベースタグ付け技術の適用と課題

本章では，物体タグ付けシステムに事例データとの類似画像検索に基づくタグ付け技術[1]を実際に適用して得られた結果とそこから導き出された新たな課題について述べる．適用実験には，カッターナイフ，乾電池，はさみ，ペンチ，ペン，マウス，ノート，腕時計，機種の異なる2つの携帯電話（以降，携帯電話1，携帯電話2と称する），デジタルカメラ，の計11カテゴリの物体を用いた．これらの物体に対して，2.1.1節で述べた物体が机上に置かれる際の流れを実施した後，物体領域検出部によって検出された物体画像をクエリとして物体タグ付けシステムに入力し，評価した．

3.1　類似画像検索の適用結果

文献[1]では，類似画像検索の特徴量として色＋形状特徴量2が用いられていたが，使用する特徴量を改めて選定する．これは，今回入力される物体画像がデータベース内に格納されている画像と撮影条件が大きく異なる可能性が高いためである．2.1.2節で述べた4つの特徴量ごとの類似画像検索結果を比較して最良の特徴量を選ぶ方法を採用した．表1に，カテゴリごとに，類似画像検索結果の上位128件の中に含まれる同一のカテゴリの物体画像の数を示す．

結果は，携帯電話1と携帯電話2を除いて，形状特徴量が最良であった．携帯電話1と携帯電話2については，形状特徴量が最良の結果ではないが，十分に高い頻度であるため問題にならないと考えられる．類似画像検索には形状特徴量を用いることにした．

また，カッターナイフやはさみでは同じカテゴリの画像がほかと比べて極端に少なかった．これは，Web

表 1　検索結果に含まれる同一カテゴリ画像数

	色	形状	色＋形状 1	色＋形状 2
cutter knife	0	6	1	6
battery	0	9	0	0
scssior	0	41	4	5
plier	0	67	18	40
pen	0	30	4	14
mouse	0	33	9	22
notebook	0	11	1	6
watch	0	128	83	110
mobile phone1	0	73	62	91
mobile phone2	0	105	95	108
camera	0	128	128	128

画像データベース内に含まれる画像の数が少ないためであり，Web画像データベース内に含まれる画像数にばらつきが多いことも明らかになった．

3.2　重要単語抽出の適用結果

重要単語抽出の適用では，3.1節で述べた形状特徴量による類似画像検索を行って得られた検索結果から文献[1]と同様のKLの指標で単語を抽出し，物体の概念を表す単語の順位を算出した．なお，文献[1]で最も高精度であった類似度に応じた単語頻度への重み付けは非適用とした．これは，今回のクエリとして用いられる画像は，Web画像データベース内の画像と撮影条件が大きく異なる可能性が高いため，類似度と見た目の違いの相関関係が強くないと考えられるためである．KL指標のしきい値はノイズの影響を抑えるような値として0.5に設定した．結果を表2に示す．

表2に示されるように，カッターナイフと乾電池とノート以外で物体の概念を表す単語は第1位であった．また，カッターナイフと乾電池はKL指標が低いため，しきい値で除外されていた．しきい値を低く設定すればこれらの単語が抽出可能であるが，ほかのカテゴリで物体に無関係な単語も抽出されてしまうため，指標のしきい値調整では対応できない．

また，物体の概念を表す単語が第1位の物体では，表1に示すように，類似画像検索結果に含まれる同一カテゴリ画像数が検索結果の20%以上含まれている．一方，カッターナイフ，乾電池，ノートについては，類似画像検索結果に含まれる同一カテゴリ画像数が10%未満であった．

3.3　適用結果から得られた課題

実環境で撮影された物体画像に対して，類似画像検索に基づくタグ付け技術を適用した場合，Web画像データベース内に含まれる画像数が少ないことが原因で，類似画像検索結果に含まれる同一カテゴリの画像が極端に少ない場合がある．このようなデータでは，重要単語抽出の性能が低くなることが明らかになった．これを解決することが課題である．

表 2　KL の指標での物体概念単語の順位

	順位	検索結果中の同一カテゴリ画像数
cutter knife	しきい値未満（9）	6
battery	しきい値未満（1）	9
scssior	1	41
plier	1	67
pen	1	30
mouse	1	33
notebook	圏外	11
watch	1	128
mobile phone1	1	73
mobile phone2	1	105
camera	1	128

4.　物体画像に対する類似画像検索ベースタグ付け方法

本章では，まず，3章で述べた課題への対応策とそれを適用した結果について述べる．筆者らが試作している人間共生ロボット“EMIEW2”を用いて音声対話による物体情報検索を実行した結果を示す．

4.1　同一カテゴリの画像が少ない場合への対応

3.3節で述べたように，重要単語抽出の性能が低くなるのは，類似画像検索結果に含まれる同一カテゴリの画像数が少ないときであった．これにより，単語の統計的な指標の計算に同一カテゴリ以外の画像が多く使われ，性能が低下する．

これを改善するには，類似画像検索結果に含まれる同一カテゴリの画像をできるだけ多く用いることが必要である．これを実現するために，類似画像検索結果の分析を進める．

4.1.1　検索結果の類似度の分析に基づく対応

まず，入力画像と類似度について傾向を調査した．調査では，4.1.1節と同様の11カテゴリの物体に対して，形状特徴量による類似画像検索を行い，検索結果のうち第48位までの類似度の推移を求めた．類似度には，入力画像の形状特徴量とのユークリッド距離の逆数を用いた．その結果を図8に示す．

類似画像検索結果に含まれる同一カテゴリの画像数

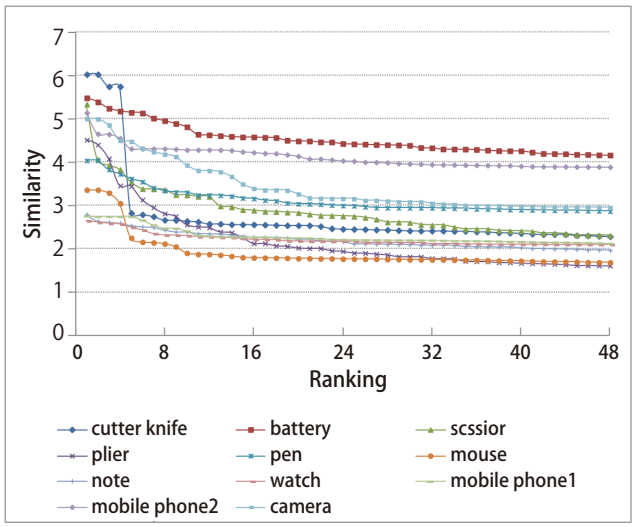

図 8　各物体に対する類似度の推移

が少ないカッターナイフは，第5位以降の類似度が急激に低くなっている．表1に示したように，類似画像検索結果に含まれるカッターナイフの画像6枚のうち4枚が第1位から第4位までに入っていた．また，その4枚は入力画像と見た目が近い画像であった．

同じような類似度の推移傾向はマウスでも見られ，第1位から第4位は見た目が近い画像である一方，類似度が第1位からある程度離れた第10位以降はマウス以外の画像の割合がかなり多くなっていた．また，ペンチでも，第1位の類似度からある程度離れた画像については，ペンチ以外の割合が多くなっていた．

また，類似画像検索結果に含まれる同一カテゴリの画像数が多い，携帯電話2やカメラでは，下位になっても第1位の類似度から大きな変化が見られない傾向であった．これらは，検索結果の多くが入力画像と同一カテゴリであり，ほとんどは見た目が近い画像であった．

以上の傾向から，(1) 第1位の類似度から一定の距離以内の検索結果を用いる対策を実施した．対策 (1) を適用する前後のカッターナイフの抽出単語を**表3**に示す．物体の概念を表す単語 “knife” の順位は1位になり，KLの指標値もしきい値を超える値になった．また，ほかの物体では，抽出単語に対する影響は軽微であった．

一方，ペンや乾電池については，類似画像検索結果に含まれる同一カテゴリの画像数が多い物体と似た類似度の推移傾向であり，対策 (1) では改善が見られなかった．次節において追加の対応策を検討する．

4.1.2　検索結果の画像の分析に基づく対応

まず，ペンや乾電池の類似画像検索結果の画像を分析した．乾電池については，検索結果の多くに風景画像が多く含まれていた．風景画像は物体とは無関係なため，物体画像との類似画像検索には不要なデータである．

また，ペンについては，類似画像検索結果の多くがペンとは見た目が異なる画像であった．その中の多くにPC等で作られた2次元CG画像が含まれていた．ペンのように形状が四角形に近く特徴が少ない場合，2次元CG画像のような四角形で構成される人工的な画像が検索結果に出やすくなると考えられる．人工的な画像も物体とは無関係なため，物体画像との類似画像検索には不要なデータである．

以上から，Webからデータを取得した後に，物体画像との類似画像検索に不要なデータか否かを判定し，フラグを付けることで類似画像検索に使用しないようにする対策が有効である．まず，実現が容易な (2) 風景画像の判定を導入し，風景画像と判定された画像を類似画像検索に使用しない対策を実施した．風景画像の判定方法は，画像端のテクスチャの有無によって判別する．

対策 (2) を適用した前後の乾電池の抽出単語を**表4**に示す．“電池” という単語がKLの指標値がしきい値を超える値まで上昇した．また，ほかの物体の抽出単語に対する影響はなかった．

重要単語の抽出結果を確認したところ，想定していない単語の抽出が見られた．カメラ，腕時計の重要単語の抽出例を**表5**に示す．カメラでは，メーカや型番に関連する単語に加えて，筆者らが想定していなかったデジタル一眼レフの略称であるDSLR（Ditigal Single

表 3　カッターナイフにおける対策適用前後の抽出単語

対策適用前		対策適用後	
単語	KL	単語	KL
ballistic	0.534015906	knife	7.179771385
caliber	0.524302001	blade	3.46742898
nato	0.445708358	duty	3.268773129
weapons	0.417559112	utility	3.115059281
koch	0.397451028	heavy	3.066994845
heckler	0.321733157	snap	2.327708957
focus	0.219845887	series	1.382606416
close	0.216204106	design	1.01252204
knife	0.215658552		

表 4　カメラと乾電池での抽出単語

単語	KL	単語	KL
電池	0.357764862	電池	0.914031186
円	0.245079883	仏壇	0.496474599
仏壇	0.233882189	単	0.470884806
単	0.232998665	アルカリ	0.444680427
掛け軸	0.230070587	掛け軸	0.432934846

表 5　カメラ，腕時計に対する抽出単語

カメラ画像	腕時計画像
camera	watch
digital	lady
dmc	dial
lumix	steel
optical	stainless
zoom	diamond
panasonic	fake
megapixel	swiss
dsIr	seiko
reviews	chronogragh

Lens Reflex）という単語も抽出できた．また，腕時計では，メーカや型番についての単語に加えて，「steel」「stainless」といった入力された時計の素材に関する単語も抽出されていた．

Webの情報を活用した類似画像検索ベースタグ付け技術を用いたことで，上記のような利用者の想定していない重要な単語をタグとして付与できることが明らかになった．

4.2　“EMIEW2”との音声対話による物体情報検索の評価

4.1節の対応を実施した対話型物体情報探索システムを実装し，“EMIEW2”を用いて音声対話による物体検索の性能を評価した．物体タグ付けシステムによって抽出された単語はすべて日本語に変換した．この際，対応する日本語がなかったタグはカタカナ表記に変換した．評価実験では，4.1.1節と同様の11カテゴリの物体それぞれに対して，机上の任意の場所に置いた後，“EMIEW2”に対して「“物体名”はどこにありますか？」という質問を5度問い合わせたときに物体の場所を特定できた数を調査した．

調査の結果，電池，はさみ，ペン，マウス，時計，携帯電話1，携帯電話2，カメラの9種類の物体では，場所の特定がすべて成功した．また，カッターナイフについては，2度「アルカリ」と誤認識し，ペンチについては1度「電池」と誤認識した．特に，ペンチと電池は音が似ているため，誤認識が起こりやすく，今後対策が必要である．

5.　おわりに

5.1　本稿の結論

本稿では，実環境の物体画像に対して類似画像検索ベースのタグ付け技術を適用して得られた知見について述べた．物体画像からのタグ付けに必要な3つの要件は，（a）学習データを事前に定義しない，（b）タグはテキスト形式，（c）ノイズの少ない単語抽出である．このうち，要件（a）（b）は類似画像検索ベースのタグ付け技術を採用したことで満たされた．

要件（c）については，類似画像検索結果の上位N件分の画像に付随するテキスト集合から単語抽出する従来の方法では性能が低かったため，満たすことができなかった．これに対して，（1）第1位の類似度から一定の距離以内，かつ，（2）検索結果から風景画像を除外した検索結果だけを用いることで単語抽出性能が改善され，要件（c）を満たすことができた．また，本タグ付け技術を用いることで，利用者が知らない物体の略称や型番など，物体に関連する単語を抽出できた．

今後，置かれる物体の種類によっては上記の改善策では要件（c）を満たせない事例が発生し得る．これが起きるたびに検索結果を調整するには限界があるため，人が物体画像をデータベースに追加するなど，Web画像データベース側を更新する対策をとるのがよい．

5.2　今後の展開

今後の展開の1つ目は，まず，Web画像データベースに格納された画像のうち，物体画像に適した画像の選定をさらに進める．今回の知見のように，類似画像検索に使用する画像を選別することで，カテゴリ間の画像数の偏りに対して，さらなる精度向上につながると考えられる．また，2つ目の展開は，物体が環境カメラから外観が確認できない位置に置かれる場合への対応である．これに対しては，環境側センサとロボットセンサとの連携を進めていく．

参考文献

1）渡邉裕樹，秋良直人，廣池　敦，松原大輔，平松義崇，永吉洋登，影広達彦，久光　徹：大規模Web画像データベースを用いた画像アノテーションシステムの構築，情報処理学会研究報告，2012-CVIM-181, No.8, pp.1-8 (2012).

2）山崎公俊，野沢峻一，植田亮平，槙　俊明，森　優人，岡田　慧，松本　潔，稲葉雅幸：日用品データベースを利用する家事支援ロボットによる思い出し・片付け支援，第27回日本ロボット学会学術講演会講演論文集，2E2-05 (2009).

3）橋本敦史，中村和晃，舩冨卓哉，美濃導彦：TexCutとパッチ型背景モデルの組み合わせによる机上物体検出システム，画像の認識・理解シンポジウム（MIRU），DS-04 (2012).

4）小田嶋成幸，佐藤知正，森　武俊：画像の安定変化を用いた複数視点統合による家庭内物体移動管理システム，日本ロボット学会誌，Vol.29, No.9, pp.837-848 (2011).

5）桑畑舜也，長谷川勉，諸岡健一，倉爪　亮，辻　徳生：情報構造化環境における家具上物品検出のための移動ロボットによる視覚記憶照合と変化検出，第31回ロボット学会学術講演会講演論文集，

3I1-04 (2013).
6) Karpathy, A. and Fei-Fei, L.: Deep Visual-Semantic Alignments for Generating Image Description. The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp.3128-3137 (2015).
7) 平松義崇，永吉洋登，影広達彦，渡邉裕樹，松原大輔，廣池　敦：環境情報構造化を用いた人間共生ロボット“EMIEW2”の物体探索，情報科学技術フォーラム講演論文集 , Vol.12, No.3, pp.517-518 (2013).

平松 義崇（正会員）yoshitaka.hiramatsu.xw@hitachi.com
2004 年名古屋大学大学院工学研究科情報工学専攻修士課程修了．同年（株）日立製作所入社．同社研究開発グループにて，画像認識に関する研究開発に従事．

住吉 貴志（正会員）takashi.sumiyoshi.bf@hitachi.com
2003 年京都大学大学院情報学研究科知能情報学専攻修士課程修了．同年 (株) 日立製作所入社．同社研究開発グループにて，音声認識システムの研究開発に従事．

永吉 洋登（正会員）hiroto.nagayoshi.wy@hitachi.com
2001 年早稲田大学大学院理工学研究科電気工学専攻修士課程修了．同年 (株) 日立製作所入社．同社研究開発グループにて，画像認識の研究開発に従事．

渡邉 裕樹（非会員）yuki.watanabe.dv@hitachi.com
2009 年東北大学大学院情報科学研究科情報基礎科学専攻博士課程修了．博士（情報科学）．同年（株）日立製作所入社．同社研究開発グループにて，類似画像検索，映像認識に関する研究開発に従事．

廣池 敦（非会員）atsushi.hiroike.rx@hitachi.com
1986 年早稲田大学大学院文学研究科修士課程修了（心理学）．1994 年東京農工大学大学院工学研究科博士課程修了．同年（株）日立製作所入社．画像検索等の研究開発に従事．博士 (工学)．

浜田 宏一（非会員）koichi.hamada.av@hitachi.com
1996 年東京大学大学院工学系研究科電気工学専攻修士課程修了．同年 NHK に入局．2003 年より（株）日立製作所同社研究開発グループにて，画像信号処理に関する研究開発に従事．博士（情報理工学）．

影広 達彦（正会員）tatsuhiko.kagehiro.tx@hitachi.com
1994 年筑波大学大学院修士課程修了．同年（株）日立製作所入社. 同社研究開発グループにて, 画像処理の研究開発に従事．2004 ～ 05 年英国 Surry 大客員研究員．2008 年筑波大学大学院博士課程了．博士（工学）．

採録決定：2016 年 6 月 23 日
編集担当：上條浩一（日本アイ・ビー・エム（株））