

## 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 Locate Function for the Human Symbiotic Robot, "EMIEW2" using Structuring of Environmental Information

平松 義崇<sup>†</sup> 永吉 洋登<sup>†</sup> 影広 達彦<sup>†</sup> 渡邊 裕樹<sup>†</sup> 松原 大輔<sup>†</sup> 廣池 敦<sup>†</sup>

Yoshitaka Hiramatsu, Hiroto Nagayoshi, and Tatsuhiko Kagehiro

Yuki Watanabe, Daisuke Matsubara, and Atsushi Hiroike

### 1. はじめに

高齢化の進展による生活サポートが必要な高齢者の増加、少子化の進展による労働人口の減少、社会の成熟に伴う個々のレベルに応じた価値の提供が、近年、社会的な課題となっている。これを解決する手段の一つとして、公共、医療・福祉といった人が活動する現場でサービスを提供することができるサービスロボットが期待されている。

人が活動する現場では、生産工場とは異なって物体の位置が頻繁に変化する。そのような現場でロボットが物体の場所への案内や搬送するためには、現場にある物体とその位置を把握し、対話等で人に物体の情報を提示する必要がある。特に、人が現場から離れた後にこのニーズが高まると考えられる。

本研究では、室内の天井に設置したカメラから室内の物体の位置と名称を認識するシステムにより、人間共生ロボット EMIEW2 に物体の位置に案内する機能を実現した。以下、その詳細について述べる。

### 2. 課題と解決策

物体の位置を把握するアプローチとして、カメラで撮影した映像から画像処理を用いた。画像処理を用いて物体の位置を把握する場合、まず、カメラの配置を検討する必要がある。本研究では、部屋全体の物体の配置を常時把握することを目指しているため、常時、撮影が可能、かつ、オクルージョンが発生しづらい場所として、天井にカメラを設置することにした。天井からカメラで撮影した画像では、物体以外にも人も映り込む。したがって、物体と人を含むそれ以外とを判別する課題がある。

また、対話で人に物体の情報を提示するには、検出された物体に対して、そのものを特定するメタデータを付ける必要がある。これに対しては、画像の内容を表すメタデータを自動的に付与する画像アノテーション技術を適用することができる。画像アノテーション技術の中でも、ImageNet のように人手で作成した画像データベースを用いて、機械学習によりカテゴリ分類するアプローチが広く知られている。しかし、このアプローチの場合、次々と登場する新商品への対応に手間がかかるという課題がある。

本研究では、撮影された画像に対して、人の移動に伴う変化と物体の出現・消失に伴う変化を判別し、物体の出現・消失に伴う変化が生じた画像を背景差分することで、物体だけを検出する。また、Web から画像とその周辺テキストを抽出し、関連付けて作成した Web 画像データベースを用いた画像アノテーション技術[1]により、新商品に対してもメタデータの付与を可能とした。

<sup>†</sup>日立製作所 中央研究所

Hitachi, Ltd., Central Research Laboratory

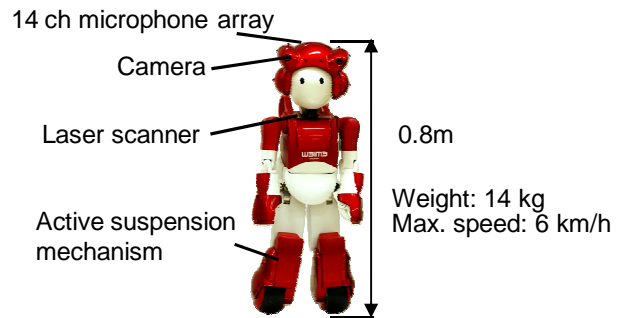


図 1: 人間共生ロボット EMIEW2

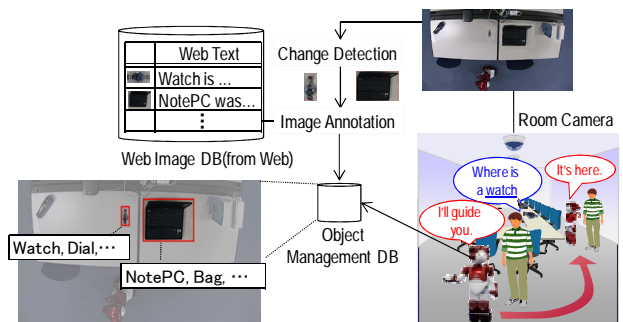


図 2: 物体認識システム

### 3. 人間共生ロボット “EMIEW2”

我々は、会話等を介した目的地への案内、オフィスや工場などにおける巡回監視などのサービスを実現することを目標に人間共生型ロボットの開発を進めており、2007年にEMIEW2を発表した。図1にEMIEW2の外観を示す。EMIEW2は、人にストレスを与えない移動速度である1.6[m/s]を実現し、本質的な安全と実用的なサイズを狙い、体重14kg、身長80cmとなっている。本研究において、室内の物体に関する情報を天井に取り付けたカメラから抽出することで、対話による物体の探索機能を実現した。

### 4. 物体認識システム

天井に設置したカメラによる物体認識システムを図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、物体の出現・消失に伴う変化を検出し、その変化領域の位置を算出する。続いて、変化領域の画像に対して、画像アノテーション処理[1]により、物体の名称の候補を表すメタデータを抽出する。最後に、変化領域の画像、位置、メタデータを物体管理データベースに保存する。

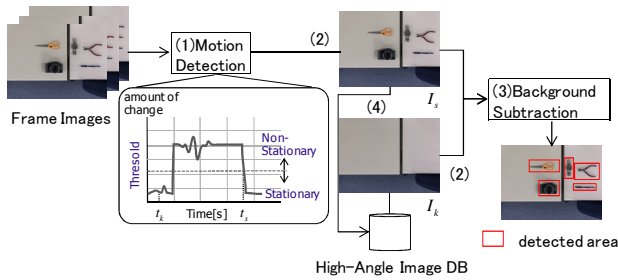


図 3：物体の出現／消失に起因する変化抽出処理

本システムにより、EMIEW2 に物体の場所を問合せた際に、音声認識で得られた単語から物体管理データベースを検索することで、対象物体の場所を知ることができる。

続いて、画像による物体の出現・消失に伴う変化検出と、画像アノテーションについて述べる。

#### 4.1 物体の出現・消失に伴う変化検出

人が現場から離れた後に物体の場所への案内や搬送のニーズが高まるとの考えから、今回、人が物体を持って現場まで来て、何らかの作業を行い、物体をいずれかに置いた後に現場から去っていく場面をターゲットとした。この場面では、人が現場に来る前と去った後で物体の有無だけ違いが生じている。したがって、動きが発生する直前の画像と、動きが発生した後、動きがなくなった直後の画像とを比較すれば、物体の出現や消失に起因する変化のみが抽出できると考えられる。

これに基づく物体の出現や消失に起因する変化を抽出する処理を図 3 に示す。本処理においては、(1)撮影された映像から常時変化検知を行い、撮影範囲の動き状態を認識する。(2)撮影範囲で動き有りから動き無しに状態が変化したタイミング  $t_s$  でカメラが撮影した画像  $I_s$  と、俯瞰画像データベースに格納された最新の画像  $I_k$  から、(3)背景差分により変化したピクセル群を抽出する。(4)背景差分後に、 $I_s$  を最新画像として、俯瞰画像データベースに格納する。

今回用いられる画像には、照明によって物体の端に影が生じ、また、データ量の削減のために画像圧縮によるブロックノイズが生じる。そのため、ノイズにロバストなテクスチャ特徴の比較による手法を用いることとした。その中でも、Sobel フィルタで微分画像を生成し、その画像からテクスチャ特徴の相違度  $dt(B)$  とテクスチャ強度  $T(B)$  を算出し、 $dt(B)$  と  $T(B)$  の積により背景・前景を判別する Li らの手法[2]を用いた。ただし、本研究では、ブロックノイズへの影響を低減するため、Sobel フィルタの代わりに平滑化効果のある Sinc 関数の 1 次微分フィルタを用いた。 $dt(B)$  と  $T(B)$  の積が所定のしきい値より大きい場所を抽出し、それらの連結成分を生成し、それを囲む外接矩形を物体領域として検出する。

#### 4.2 画像アノテーション[1]

物体の出現・消失に伴う変化検出で検出された領域画像に対して、Web 上の画像とその周辺で使われていたテキストが関連付けて格納された Web 画像データベースから類似画像を検索する。続いて、検索された画像の周辺で使われていたテキストから抽出した単語に対して確率的指標を算出し、その値に応じて物体の名称の候補を表すメタデータを抽出する。

## 5. 評価実験

室内の天井に設置したカメラで撮影された映像を用いて提案手法を評価した。映像は解像度  $1280 \times 960$  [pixel]、10fps で撮影した。評価には、電話、時計、はさみ、ペンチ、ペン、マウス、カメラ、電池の 8 種類の物体を用いた。評価に使用した映像は、8 種類の物体から 1 つ選択して、机の上に置き、一旦、そこから離れ、次に、置いた物体の場所まで再び移動し、物体を取って、そこから離れる作業の様子が撮影されたものを用いた。各物体でこの作業を 5 回繰り返した様子を撮影した。

提案手法の物体の出現・消失に伴う変化検出の評価尺度として、recall と precision を用いた。また、物体認識システムについては、物体の出現・消失に伴う変化検出で正しく物体が検出された領域に対して、画像アノテーション処理を行って付与された上位 5 位以内のキーワードに、物体と同じ概念の単語が含まれていれば正解とし、5 回の作業の平均正解率で評価した。

物体の出現・消失に伴う変化検出は、recall は 100%(80/80)で precision は 97%(80/82)であった。誤検出された 2 つの領域は、物体を置いた際に、既に配置された物体が微小にずれたことで検出された領域であった。この領域を画像アノテーションによりメタデータを付与したところ、5 位以内に物体の概念を示すキーワードが付与されなかった。したがって、ロボットに問い合わせて物体を探索するときには、この誤検出は問題にはならなかった。

表 1 に物体名称の推定精度を示す。各概念で最大 100%、8 種類平均で 90%であった。電話、ペンチ、マウス、時計、カメラでは、外観が特徴的であることから、類似画像検索結果は同一概念のもので多数占められ、高い正解率となった。電池やはさみでは、類似画像検索結果に他の概念が一部混じっていたことで、正解率が低くなった。

本機能を実際のオフィス環境で実証し、安定動作することを確認した。今後、オフィスでの一般的な活動の映像で評価を進める。

表 1：物体名称の推定精度

物体	正解率	物体	正解率
電話	100%	マウス	100%
はさみ	80%	カメラ	100%
ペンチ	100%	電池	60%
ペン	80%	時計	100%

## 6. まとめ

室内の天井に設置したカメラからロバストに部屋全体の物体の位置と名称を認識する方法を提案した。本方法を用いた物体認識システムにより、人間共生ロボット EMIEW2 に物体の位置に案内する機能を実現した。

#### 参考文献

- [1] 渡邊 裕樹, 他, “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, Vol.2012-CVIM-181, No.8 (2012).
- [2] L. Li, and M.K.H. Leung, “Integrating intensity and texture differences for robust change detection”, IEEE Trans. Image Process., Vol.11, No.2 (2002).