

記憶障害者支援システムのため画像分類アルゴリズム

濱崎亮大¹ 松本晃輔¹ 川端健介¹ 梶原薪¹ 大島千佳¹ 中山功一¹

概要:本研究では、認知症患者の紛失物の探索を支援する記憶障害者支援システム“Sahasra”の開発を目指す。“Sahasra”の開発に向け、物体の移動検知部、移動物体の分類部、物体認識モデル作成部の実装を行った。移動物体の分類部では、ディープラーニング特徴量と物体の大きさを組み合わせた分類手法を提案し、分類を行った。提案したディープラーニング特徴量と物体の大きさを組み合わせた分類手法と、従来のディープラーニング特徴量を用いた分類手法の比較実験を行い、従来手法より提案手法の分類精度が高いことがわかった。

An Image Classification Algorithm for a Memory Impaired Person Support System

RYODAI HAMASAKI¹ KOSUKE MATSUMOTO¹ KENSUKE KAWABATA¹
KAJAHARA SHIN¹ CHIKA OSHIMA¹ KOICHI NAKAYAMA¹

Abstract: This paper constructed a memory impairment support system, "Sahasra" that allows dementia patients to search for lost items. Sahasra consists of a moving object detection unit, a classification unit for moving objects, and an object recognition model creation unit. In the classification unit for moving objects, I proposed a classification method that combines deep learning features and object size. The results of a comparative experiment showed that the proposed classification method can perform higher accurate classification than that of a conventional image classification method.

1. はじめに

認知症の主な症状は、中核症状と行動・心理症状（以下、BPSD と表記する）に分けられる。中核症状には、「記憶障害」、「失語・失行・失認」、「遂行機能障害」、「見当識障害」がある。これらの中核症状は、脳の神経細胞の障害によって起こる。BPSD は、中核症状と環境・身体・心理などの要因が相互作用して生じる様々な行動・心理症状のことである。

認知症患者（以下「患者」）は、記憶障害により「物をなくしやすい」「何度も同じことを訊く」などの行動を起こす。患者は記憶が定かでなくなることで、「状況がわからなくなる不安」や「自己喪失感」に襲われる。このことから、患者の自己効力感や自己肯定感などの自尊心の低下につながる。自尊心の低下は BPSD の悪化の要因と考えられる。患者の自尊心を確保することで、BPSD の軽減が期待できる。

たとえば、患者が紛失した物を探すとす。介護者が探すのではなく、患者自身が誰の手も借りずに探しだすことができれば、自尊心の確保につながる。中山らは、患者の自尊心を確保するため「記憶障害者支援システム“Sahasra”」を提案した（以下、サハスラと表記する）[1]。サハスラは、紛失した物体を探し当てるとともに、その物体を患者に直接届けるシステムである。患者自身がサハスラへ紛失物を

探索するよう指示を出すことで、患者は、「誰かに探してもらった」と感じない。サハスラが紛失物を探し出すことで、患者の自尊心を確保する。サハスラを実装するためには、2つの課題がある。

1つ目の課題は、「事前に対象とする物体を登録しなければならない」ことである。物体の探索に RFID タグのようなタグを使用する場合は、対象とする物体（なくしたくない物体）に、事前にタグを装着する必要がある。しかし、対象とするすべての物体にタグをつけることは現実的でない。画像処理を用いて物体の探索をする場合は、対象とする物体の特徴を事前に抽出・記録する必要がある。対象とするすべての物体の特徴を事前に記録することや、物体の認識モデルを作成することも現実的でない。

2つ目の課題は、「物体の種類がわからない状態での物体の分類が難しい」ことである。日常生活で使用する物体の種類を正確に把握することは不可能である。例えば、物体によって大きさ以外が類似した物体（以下、類似物体と表記する）が存在する。

本研究では、サハスラ実装に向けて、以上の課題を解決する物体検知機能を開発する。開発する物体検知機能は、部屋に設置した固定カメラの連続画像から、差分検知アルゴリズムにより、移動した物体を自動的に抽出、分類する。これにより、学習用データを事前に必要とせず、物体認識

¹ 佐賀大学

モデルを作成する。

2. 記憶障害者支援システム”Sahasra”

サハスラは、生活空間に設置された複数カメラにより患者に必要な物体を検知する「物体検知機能」、患者の求める物体を移動するハンド制御を行う「物体運搬機能」から成る[1]。

物体検知機能は、物体の移動検知部、移動物体の三次元座標検知部、移動物体の分類部、物体認識モデル作成部、音声入力部からなる。物体検知機能では、患者の生活空間にある物体の移動を記録している。そのため、患者が紛失した物体の探索が可能である。物体検知機能の流れを図1に示す。本研究では、物体検知機能の「物体の移動検知部」、「移動物体の分類部」、「物体認識モデル作成部」の開発を行う。

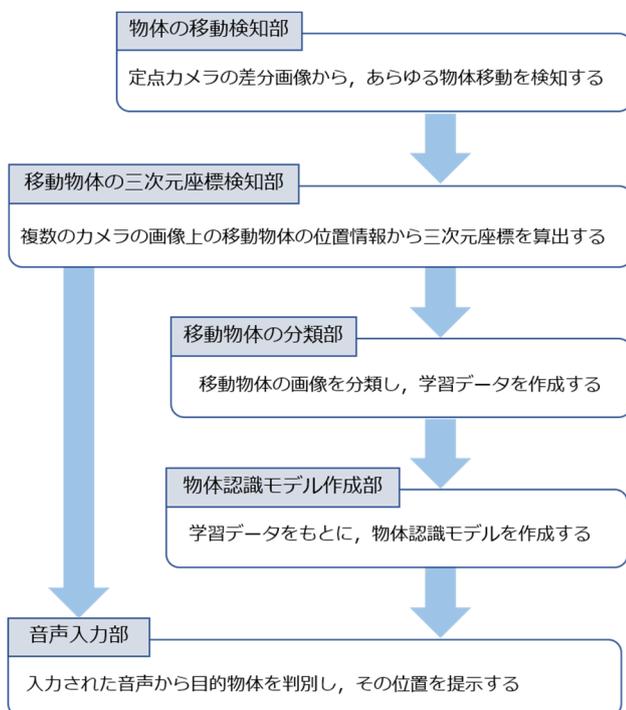


図1 物体検知機能の流れ

Fig 1 Flow of the object detection function

3. 開発する機能

3.1 物体の移動検知部

物体の移動検知部では、差分検知アルゴリズムを用いて物体の検出を行う。差分検知アルゴリズムは、初めに日常生活空間の画像を「背景画像」として登録する。背景画像と固定カメラからの連続画像との差分から、変化があった箇所を「物体が現れた／取り除かれた」と判断する。時系列において背景画像は、連続画像の1つ前の画像となる。連続画像を撮影するタイミングは、画面内に人間の出入りがあった直後としている。人間の検知には、物体検知アル

ゴリズム YOLO (以下, YOLO と表記する) [2]の学習済モデルを用いる。この学習済モデルは人間を検知することができる。差分検知アルゴリズムを用いた物体の検出の流れを以下に示す。

1. 画面内に人間の出入りがあった直後に、画像を撮影する。
2. 差分検知で使用する 2 枚の画像をそれぞれグレースケールで読み込む。
3. 2 枚の画像の差分をピクセル単位で算出し、差分画像を作成する。(図 2)
4. 差分画像にガウシアンフィルタ[3]をかけ、ぼかし処理を行う。(図 3)
5. キャンー法[4]を用いてエッジ検出を行う。(図 4)
6. 検出されたエッジにモルフォロジー変換[5]を用いてエッジの収縮処理を行う。(図 5)
7. エッジの外形矩形を抽出し、物体として切り出す。(図 6)



図2 差分検知

Fig 2 differential detection

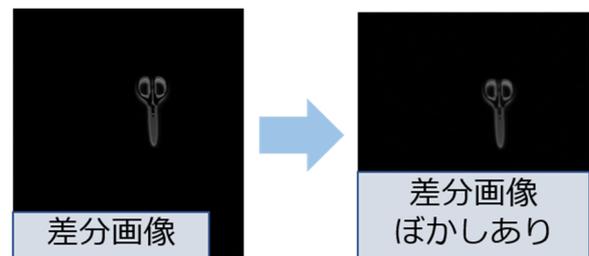


図3 ぼかし処理

Fig 3 Blurring

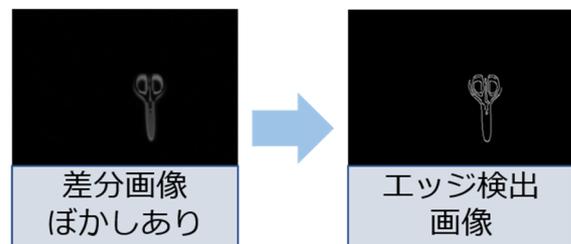


図4 エッジ検出

Fig 4 Edge detection

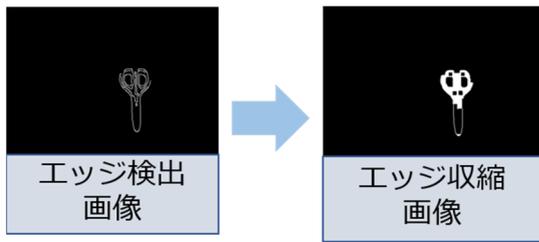


図5 エッジ収縮処理
Fig 5 Edge shrinking process

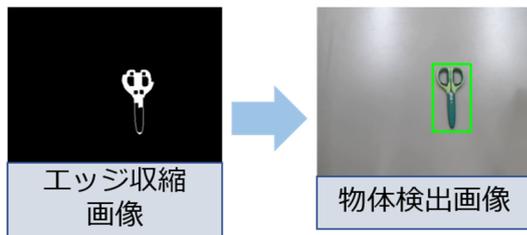


図6 外形抽出
Fig 6 Extraction of external shapes

物体の移動検知部では、画像内にある物体の位置と大きさをシステムに記録している。これにより、同時に物体が現れた場合を除き、物体の重なりでの区別が可能になる。物体の重なりでの区別は、物体が現れた順番で配列に格納することで、高さを表現している。なお、検出した物体のみ矩形で表示する。物体の重なりでの区別の確認を、図7、図8に示す。図7では、現れた物体が他の物体によって隠れる場合を示している。



図7 物体の重なり
Fig 7 Examples of overlapping objects

図8では、検出された物体の識別名を示している。検出された物体の通り、「obj-1-0-0.jpg」が「obj-2-1-0.jpg」の下に隠れて、再び現れた際に現れた物体は「obj-1-0-0.jpg」と同じ物体だと判断できたことがわかる。



図8 物体の識別名
Fig8 Object identifier

物体の移動検知部では、画像内にある物体の位置と大きさをシステムに記録することで、変化があった箇所を判断している。このため、最初に登録した背景画像の時点で置かれていた物体が取り除かれた場合、「取り除かれた物体の背景」の画像を、「現れた物体」の画像として誤認する可能性がある。そこで、物体の切り出しの際に、「現れた物体」の画像と「取り除かれた物体の背景」の画像の両方をシステムに記録し、「現れた物体」の画像と「取り除かれた物体の背景」の画像を、ユーザーが手動で入れ替えられる仕様とした。「現れた物体」の画像を「Object images」、「取り除かれた物体の背景」の画像を「Back images」とした。開発したユーザーインターフェースの使用例を図9に示す。図3.8 上画像は、「現れた物体」の画像と「取り除かれた物体の背景」の画像が逆に保存されている場合である。この場合、ユーザーは「Back images」に表示された2枚 (back-0-0-0.jpg, back-0-0-1.jpg) の画像を「Object images」へドラッグアンドドロップすることで、図9下部のようにシステムの誤認を正すことができる。

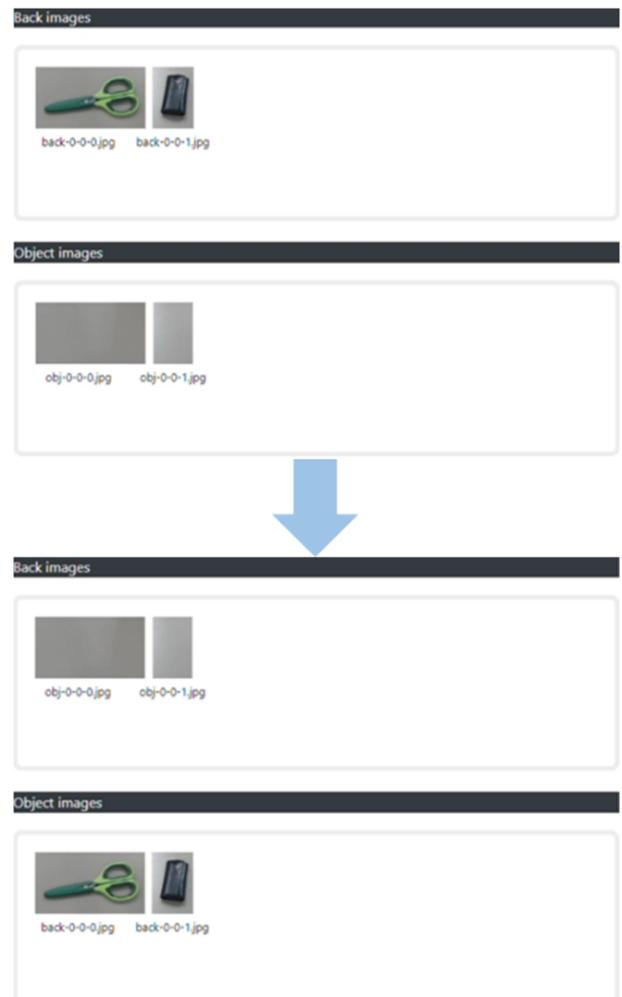


図9 背景と物体の入れ替え
Fig 9 Replacing the background and objects

3.2 移動物体の分類部

移動物体の分類部では、3.1 節で述べた移動物体を切り出した画像（以下、物体画像と表記する）を用いる。移動物体の分類部では、物体画像から物体の特徴を抽出する。抽出する特徴はディープラーニング特徴量とする。ディープラーニング特徴量とは、ニューラルネットワークの中間層から抽出された特徴マップのことである。ディープラーニング特徴量の抽出には、学習済ニューラルネットワークモデルを用いる。用いた学習済ニューラルネットワークモデルは ResNet50[6]である。この ResNet50 は ImageNet[7]の画像で学習しており、物体を 1000 クラスに分類することができる。ディープラーニング特徴量を抽出するために ResNet50 の全結合層を削除した。移動物体の分類部では、提案手法を用いて分類を行う。本研究では、カメラを撮影したテーブルの真上に設置しているため、画像サイズを物体の大きさとしている。ディープラーニング特徴量と物体の大きさを組合せる流れを以下に示す。

1. 物体画像を ResNet50 の入力サイズにリサイズする。
2. 物体画像を ResNet50 に入力し、四次元特徴マップを取得する。
3. 四次元特徴マップを一次元特徴マップに平坦化する。この一次元特徴マップをディープラーニング特徴量とする。
4. リサイズ前の物体画像の面積を計算し、抽出した一次元特徴マップに足し合わせる。

分類アルゴリズムには、X-means 法[8]を用いる。X-means 法とは、K-means 法[9]を逐次繰り返し、バイズ情報量規準[10]を分割停止に用いた自動的にクラスター数を決定する手法である。そのため、X-means 法では、事前にクラスター数を求める必要なく分類ができる。ディープラーニング特徴量と物体の大きさを組合せた特徴量を X-means 法に適用することで、物体画像の分類を行う。X-means 法では、入力されたデータを適当なクラスターに分類し、クラスターの平均を用いてデータの分類を調整するため、適切なクラスター数を必ず求められるとは限らない。そのため、本研究では、X-means 法での分類を 5 回行い、多数決でクラスター数を決定する。多数決で一意にクラスター数が定まらない場合は、再度分類を行う。分類終了後、分類されたクラスターごとに、「スマートフォン」や「ハサミ」などのラベルを手動で入力する。従来は切り出された画像の枚数（例えば数千枚）に対してラベルを付与する作業が必要であったが、この技術により、クラスター化された物品数に対するラベル付けのみで、物体の分類が可能となる。ラベル付けが終了した分類データを学習データとして記録する。

3.3 物体認識モデル作成部

物体認識モデル作成部は、3.2 節で作成した学習データを

用いて学習を行う。物体認識モデル作成には、YOLO を用いる。作成した学習データを用いて学習した物体認識モデルが正しく物体を認識できるか確認を行った。確認には 1 種類の物体の学習データを用いた。用いた物体は眼鏡である。用いた物体を図 10 に示す。



図 10 用いた物体
Fig 10 Object used

学習には学習データ 320 枚を用いた。学習データを学習時に用いる訓練用データ 224 枚と検証時に用いる検証用データ 96 枚に分割した。物体認識モデルの学習では、訓練用データを用いて学習を行い、検証用データを用いて検証し、精度を高めていく。学習回数は 100 回とした。学習回数に対する loss 値のグラフを図 11 に示す。本節で述べる loss 値とは、学習データに対する損失関数の値である。損失関数の値は、本来物体認識モデルが認識すべき物体の位置（正解例）と実際に物体認識モデルが認識した物体の位置（予測）の誤差を表す。loss 値が低いほど、高精度な物体認識モデルとなる。図 11 内の青色のグラフ「train-loss」は訓練用データに対する loss 値である。図 3.10 内の橙色のグラフ「validation-loss」は検証用データに対する loss 値である。物体認識モデルでは、「train-loss」、「validation-loss」を下げるように学習を行っていく。図 11 にある通り、学習回数を経るごとに「train-loss」、「validation-loss」が減少し、学習が正常に行われていることがわかる。

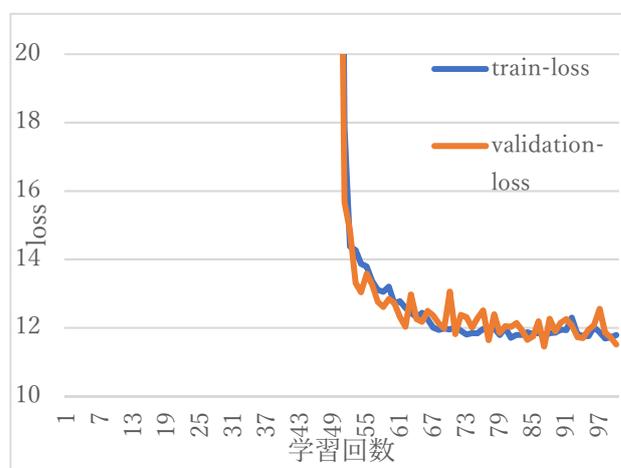


図 11 学習回数に対する loss 値のグラフ

Fig 11 Graph of loss value against number of training sessions

作成した物体認識モデルの評価を行った。評価には、最も「validation-loss」の低い物体認識モデルを用いた。本節では 1 種類の物体の認識に対する評価のみを行うため「Precision」と「Recall」と「F-measure」を用いた。本節では、TP (真陽性)、FN (偽陰性)、FP (偽陽性)、TN (真陰性) の評価指標を表 1 で定義する。評価指標の例を図 12 で示す。評価指標は評価データのピクセルごとに計算する。

表 1 評価指標

Table 1 Evaluation index

		実際の物体位置	
		正解	不正解
予測した物体位置	正解	TP	FP
	不正解	FN	TN

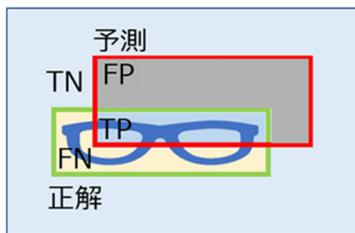


図 12 評価指標の例

Fig 12 Examples of evaluation indicators

Precision とは、物体認識モデルが予測した物体の位置と、事前に用意した物体の正解位置が重なった割合を示す(1)。

$$Precision = \frac{TP}{TP + FP} \quad (1)$$

Recall とは、事前に用意した物体の正解位置のうち、物体認識モデルが予測できた物体位置の割合を示す(2)。

$$Recall = \frac{TP}{TP + FN} \quad (2)$$

F-measure とは、Precision と Recall の調和平均である(3)。

$$F - measure = \frac{2Precision \cdot Recall}{Precision + Recall} \quad (3)$$

評価データは、学習データとは別に 20 枚の眼鏡の画像を用いた。評価データには、事前に眼鏡の正解位置を記録している。評価データに対して、作成した物体認識モデルが予測した物体位置と、事前に用意した正解物体位置を用いて、Precision, Recall, F-measure を計算した。評価データ 20 枚の Precision, Recall, F-measure の平均を計算し、Average Precision, Average Recall, Average F-measure とした。評価の結果を表 2 に示す。

表 2 より、Average Precision は 0.84, Average Recall は

0.96, Average F-measure は 0.89 であり、高精度で認識ができた。よって、作成した学習データを用いて学習した物体認識モデルが正しく物体を認識することが確認できた。

表 2 評価の結果

Table 2 Evaluation results

Average Precision	0.84
Average Recall	0.96
Average F-measure	0.89

4. 物体画像の分類アルゴリズムの比較実験

4.1 実験の目的と方法

本研究では、ディープラーニング特徴量と物体の大きさと X-means 法を組み合わせた画像の分類手法(以下、提案手法と表記する)を提案した。既存手法として、ディープラーニング特徴量と X-means 法を組み合わせた画像の分類手法がある(以下、既存手法と表記する) [11]。実験は、既存手法と提案手法でそれぞれ分類を行う。実験の目的は、分類結果をもとに精度を比較することである。実験には、5 種類の物体画像を用意した。用いた物体は、スマートフォン、タブレット、はさみ、テープ、イヤホンの 5 種類である。実験には、各物体 40 枚の合計 200 枚を用いた。既存手法と提案手法の比較のため、物体の大きさ以外が類似しているスマートフォンとタブレットを含めた。用いた物体を図 13 に示す。



図 13 用いた物体

Fig 13 Object used

実験では、まず既存手法と提案手法の両手法で 200 枚の物体画像を分類する。次に分類結果を評価する。評価は、物体 1 種類ごとに対する個別評価と、物体全種類に対する全体評価にわたる。個別評価では、40 枚すべての物体画像が同じクラスターに分類されたときに正解とし、それ以外を不正解とする。全体評価では、5 種類の物体画像の個別評価がすべて正解だった場合に正解とし、それ以外は不正解とする。分類回数は 100 回とし、個別評価と全体評価の正解数を求める。最後に、個別評価と全体評価の正解数を 100 で割ることで、正解率を計算する。

4.2 比較実験の結果

個別評価の正解率をそれぞれの物体の正解率とする．全体評価の正解率を全体の正解率とする．正解率の比較を表3に示す．両手法とも，はさみ，テープ，イヤホンの3種類の分類ができた．既存手法では，すべての分類において，スマートフォンとタブレットが同じクラスターに分類されたため0%となった．全体の正解率は，既存手法0%，提案手法76%となった．

表3 正解率の比較

Table 3 Comparison of percentage of correct answers

	スマートフォン	タブレット	はさみ	テープ	イヤホン	全体
既存手法	0%	0%	100%	100%	100%	0%
提案手法	76%	76%	100%	100%	100%	76%

4.3 考察

はさみ，テープ，イヤホンの3種類に関して既存手法と提案手法の両方で正しく分類ができた．しかし，既存手法では，スマートフォンとタブレットの分類ができていなかった．既存手法では，ディープラーニング特徴量を抽出する際に物体の大きさをすべて統一するため，大きさ以外の点で類似しているスマートフォンとタブレットの分類ができなかったと考えられる．提案手法では，ディープラーニング特徴量に，物体の大きさを組合せることにより，大きさ以外の点で類似しているスマートフォンとタブレットの分類ができたと考えられる．実験結果より，ディープラーニング特徴量と物体の大きさを組合せた分類手法の有効性が示された．しかし，提案手法の正解率は76%であり，十分とは言えない．今回の実験から，1種類の物体に対する画像枚数をより増やすことでより高精度な分類が行えるのではないかと考える．X-means法では，入力されたデータを適当なクラスターに分類し，クラスターの平均を用いてデータの分類を調整するため，入力するデータ数が多いほどよりデータの特徴を分類できるからである．

本研究で扱ったサハスラは，日常生活空間に固定カメラを設置することで，連続画像を取得することができる．そのため，時間経過とともに多くの物体画像が蓄積され，より高精度に分類を行うことができるようになると思われる．

5. おわりに

本研究では，サハスラ構築に向けた物体の移動検知部，移動物体の分類部，物体認識モデル作成部の開発を行った．これらの機能の開発により，移動物体の記録が容易になり，学習データ作成の効率化につながる．移動物体の分類部では，ディープラーニング特徴量と物体の大きさを組合せた分類手法の提案により，既存手法よりも高精度な分類を実現した．

今後の課題としては，実際の物体の大きさを考慮して分類を行うことである．本研究では，撮影したテーブルの真上にカメラを設置しているため，画像サイズを物体の大きさとして行うことができた．実用化を考えた場合，部屋の中央や，テーブルの真上などにカメラを設置することは現実的ではない．サハスラでは，複数台のカメラを部屋に設置することを前提としている．複数台のカメラの位置や物体の位置を用いて，実際の物体の大きさを算出して分類を行うことで，より実用的なシステムにつながると考える．

参考文献

- [1] 中山功一，大島千佳，林喜章，福田修:”在宅認知症者を支援する「拡張認知機能」の提案～スマートエアリアルハンド: Sahasraの構想～”，2020-AAC-013, pp.1-6, 2020.
- [2] Joseph Redmon, Santosh Divvala, Ross Girshick, and Ali Farhadi: “You Only Look Once: Unified, Real-Time Object Detection”. The IEEE Conference Vision and Pattern Recognition (CVPR), pp.779-788, 2016.
- [3] Hirohito Okuda, Koichiro Deguchi:” Gray Level Edge Behavior for the Gaussian Blurring Filters and Its Application to Edge Extractions”, 情報処理学会研究報告コンピュータビジョンとイメージメディア, pp.51-57, 1994.
- [4] J.F. Canny:” A computational approach to edge detection”, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol.8, No.6, pp.679-698, 1986.
- [5] Robert M Haralick, Stanley R Sternberg, and Xinhua Zhuang:” Image Analysis Using Mathematical Morphology”, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol.9, No.4, pp.532-550, 1987.
- [6] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun:” Deep residual learning for image recognition” Proc. 2016 IEEE Conf. on Computer Vision and Pattern Recognition, pp.770-778, 2016.
- [7] “ImageNet”, <http://www.image-net.org/>, (参照 2020-12-07).
- [8] Dan Pelleg and Andrew Moore:” X-means: Extending K-means with Efficient Estimation of the Number of Clusters”, ICML-2000, pp.727-734, 2000.
- [9] Hartigan J.A. and Wong M.A.:” A K-Means Clustering Algorithm”, Journal of the Royal Statistical Society. Series C (Applied Statistics), Vol. 28, No. 1, pp. 100-108, 1979.
- [10] 赤池弘次: “AIC と MDL と BIC”, Communications of the Operations Research Society of Japan, Vol.41, No.7, pp.375-378, 1996
- [11] 柳澤秀彰，渡辺裕:” Deep Learning 特徴量を用いたマンガキキャラクター顔画像の分類”, FIT2017, pp.185-186, 2017.