視覚障がい者の学びを支えるための物体認識システムの構築

森田賢太*1・栗原恵莉奈*2・森田直樹*1・高瀬治彦*3
Email: morita.k@star.tokai-u.jp

*1：東海大学情報通信学研究科情報通信学専攻
*2：東海大学情報通信学部通信ネットワーク工学科
*3：三重大学大学院工学研究科電気電子工学専攻

◎Key Words 視覚障がい者支援，ディープラーニング，物体認識

1. はじめに

平成２８年４月に施行された「障がいを理由とする差別の解消の推進に関する法律」によって障がい者に対する支援・配慮が，国公立学校では義務化され私立学校や民間施設においても努力義務として課せられるようになった．障がいには身体障がいや精神障がいを始めさまざまな障がいがあるが，我々は，身体障がいの中でも視覚障がい者を対象に，スマートフォンが目の代わりになるシステムの構築を目指している．

現在の画像認識技術は，認識率が９４%以上と非常に高い．しかしこれを実現するには，認識器を作成する時にあらかじめオブジェクトを学習させておく必要があり，未学習のオブジェクトに対しては類似するオブジェクトとして認識をするか誤認識をする．障がい者が安心して画像認識システムを利用するためには，未学習のオブジェクトに対する誤認識などを修正しておく必要がある．しかしながら，現存する画像認識アプリケーションでは，認識結果が不十分であっても利用者サイドで識別器を作り直すことができない．

本研究では，認識結果が不十分な時には画像と読み上げのルールを利用者サイドで登録でき，画像認識の識別器を作り直すことができるシステムを開発した．また，複数の実験を通じて認識精度について考察を行う．本研究により，画像認識技術を導入する際の参考資料とすることができ，効率よく高い精度の識別器を作成することができる。

2. 開発目的

障がい者に対して，障害物を検知し歩行を援助するシステム[1]や，GPS やビーコンを利用し案内するシステム[2]などさまざまな支援システムやアルゴリズムの提案がなされている．しかし，障がい者へインタビューをした結果，そのようなシステムはあまり浸透しておらず，システムの開発側と利用者である障がい者側との間にギャップがあることが確認された．全盲の方の間で口コミ評価が高いのは，画像認識技術を活用したオブジェクト認識アプリケーションであった．

図 1，図 2 は，「Tap Tap See」[3]の使用画面である．このアプリケーションは，スマートフォンの画面を２回タップして写真を撮ると，その時に撮影した画像を認識して音声案内するアプリケーションである．図 1 中の写真は，机の上にあるコップを撮影した時の画面であり，認識結果は，画面下部に表示されるとともにその内容が読み上げられる．「白，黄色、ピンク、緑、青のセラミックマグです」と音声案内されることにより，マグカップが目の前にあることが予想できる．図 2 は学内の「ゴミ箱」を撮影した時の写真だが，未学習のオブジェクトのため，「白い長方形の箱です」と音声案内された．この写真には白い長方形の箱の特徴を兼ね備えていることに間違いはないが，視界障がい者にとっては，この案内から目の前にゴミ箱があることを推測することは容易ではない．また，誤認識した情報を視覚障がい者に伝えた場合はその情報を信じるしかないため，未学習であるならば読み上げないことが望ましい．

図 1　認識例

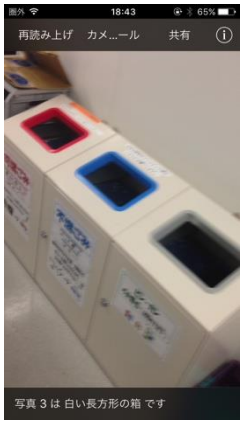

図 2　誤認識例

現在の画像認識技術は，認識率が９４%以上と非常に高く人間の認識率とほぼ変わらない．図 2 の場合も，識別器作成時にこの画像を用意でき，かつ，これがゴミ箱であることを関連づけて識別器を作成しておけば，ゴミ箱と認識させることができる．しかし，ゴミ箱ひとつとっても何十何百もの既製品が存在しすべての画像を学習時に用意することはできない．まして，その大学や施設独自のオブジェクトも存在する．そのようなオブジェクトをあらかじめアプリケーションの開発者が識別器に学習させることは不可能である．また現

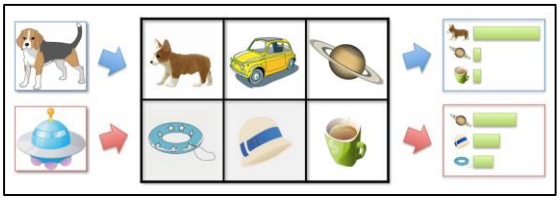

図 3 識別器の入出力イメージ

存するアプリケーションでは，大学や施設の方が識別器を作成し変更することはできない．

大学や施設の方が，机やイスなどの一般的なオブジェクトの他に，その大学や施設独自のオブジェクトの画像を用いて画像認識の識別器を作成することができれば，視覚障がい者はそれを利用して，学内での移動や生活，学びの際に画像認識技術を役立てることができる．

3. 物体認識と識別器

本章では，物体認識のシステムに必要な物体認識の出力と識別器について説明する．

3.1 物体認識の出力

物体認識は，画像に映っている物体の名前を答えるタスクである．識別器の出力は，１枚の入力画像に対して識別器作成時に学習したオブジェクトに対してそれぞれの適合率が出力され，物体の認識結果は，識別器が出力する適合率の値がもっとも高いものを採用するのが一般的である．

図 3 は，6 つのオブジェクト「犬，車，土星，浮き輪，帽子，コップ」を学習した識別器に，「犬」と「UFO」の画像を認識させた時の出力結果である．「犬」の認識は，すでに学習済みのオブジェクトであるため，犬の画像の特徴をもとに識別器が適切な出力を示している．一方「UFO」の認識は，学習時には学習していないオブジェクトであり，類似するオブジェクトが比較的高い適合率を示していることが確認できる．

3.2 識別器の学習

物体認識の技術は、ランダムフォレスト[4]や 0-ノルム最適化[5]などさまざまな手法がある．本研究では，他の物体認識の技術に比べ，認識精度が高い Deep Learning を利用する．Deep Learning は複数のモデルがある．大規模画像認識のコンテストである ILSVRC2012 においてトップになった，畳み込みニューラルネットワークの画像分類モデルである GoogLeNet[6]を利用する．このモデルは認識誤差率が 6.9%程度と人間とほぼ変わらない精度である．

Deep Learning の識別器は，認識させたいオブジェクトの画像とそのオブジェクトの名前であるラベルの２つを紐付けした学習データを識別器に提示することで，作成される．識別器は，学習データを提示されるたびに識別器の内部パラメータが変化と共に出力結果も変化し，学習する．識別器への提示回数を増やすことで試行錯誤しながら内部パラメータ適切な値に変化し，認識させたいオブジェクトの画像が入力された際にはそのオブジェクトのラベルを出力するようになる．

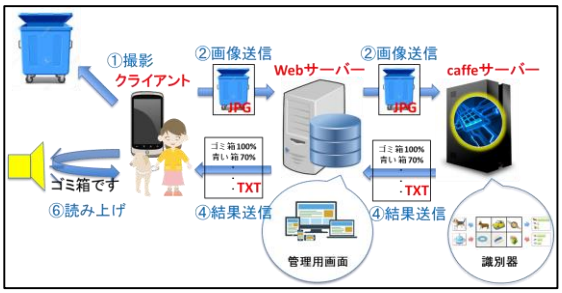

図 4 システムの構成図

１つのオブジェクトに対して，角度や色，大きさなど異なる画像を用意すればするほど認識精度が高い識別器が作成でき，また，実際に認識させたいオブジェクト以外も学習時に学習させることにより，未学習のオブジェクトを識別させる時に出力される適合率が下がる特徴がある．

画像認識の大会では，図 3 に示すような特徴が異なるオブジェクトを認識させる識別器の学習には，１オブジェクトあたり 1000 枚程度の画像を用いて学習させ，犬は犬でも犬種などを識別させる識別器の学習には，１オブジェクトあたり 3000 枚程度用いられている．

4. 提案システム

4.1 システム構成

本研究で開発したシステムの構成を図 4 に示す．本システムは，写真の取得と認識結果を読み上げるクライアント，Deep Learning のフレームワークである Caffe を利用して識別器の作成や画像認識を行う Caffe サーバ，クライアントと Caffe サーバのデータの中継や本システムの管理用画面を提供する Web サーバから構成される．

4.2 画像認識システム

本研究で開発したクライアントは，Android4.1 以上のスマートフォン，タブレットで動作させることができる．写真を撮ってから認識結果を読み上げるまでの動作と情報の流れを以下に示す．

（1） 画面を 1 回タップして写真を撮ると，http プロトコルの multipart POST を生成し写真を中継サーバで経由して Caffe サーバに送信する．

（2） 写真を受信した Caffe サーバは，既に学習済みの識別器を用いて画像認識を行い，適応率が高いオブジェクトのトップ 10 のオブジェクト名と適応率をクライアントに返送する

（3） 認識結果を受け取ったクライアントは，オブジェクト名と適応率を表示するとともに，あらかじめ設定された適応率以上のオブジェクトを読み上げる．

4.3 識別器の作り方

識別器の準備は，すべて Web ブラウザから操作を行うことができる．以下に手順を示す．

（1）画像を用意する

識別器の学習には，認識させたいオブジェクトの画像とそのオブジェクトの名前であるラベルが必要である．図５は，「ゴミ箱」のラベル付けに所属する画像の登録画面であり，あらかじめ撮影した画像ファイルをブラウザにドラッグしたりフォームのアップロード機能を用いたりすることにより登録できる．対応する画像形式は，jpeg，png，GIF 形式であり，複数の画像を zip 形式でひとまとまりにした圧縮したファイルでも追加することができる．また，Google の画像検索と連動される機能を備える．

（2）識別器を学習させる

識別器の学習は，教師データとテストデータのデータセットを用いて行う．本システム上では，登録された画像郡から利用するラベルを選択し，教師データとテストデータに用いる枚数を入力することにより，Caffe が学習する為に必要なデータベースを作成でき，学習を開始させることができる．

（3）学習率や誤認識率を確認する

識別器は学習させる度に認識精度が変化する．一定回数ごとに学習率やエラー率を見ることで，最適な回数の識別器を使うことができる．

（4）学習済みの識別器情報を識別器に設定する

認識精度が高い識別器選択し設定する．

これらにより，煩わしいコマンドによる操作や，プログラミング言語などの専門的な知識がなくても，自ら用意した画像を用いて識別器を作成することができる．

5. 物体認識の確認

本システムは識別器の精度が重要であり，その精度により有効性が影響されることが想定される．識別器が目的どおり作動するか，その精度について考察する．識別器の作成時に，データセットを加えた場合と加えていない場合での学習対象の正答率と，未学習のオブジェクトに対する出力を確認する．また，Deep Learning による識別器は学習により性能が変化するので，データセットを加えた場合の認識率の推移を確認する．

5.1 実験環境・方法

物体認識を行う対象には大学にある１２種類の物体を用いた．識別器を作成するのに必要な画像は，実際に物体を撮影したり似たような画像を検索したりして集め１種類につき１０００枚集めた．これらの画像は学習に９００枚，確認用に１００枚用いた．データセットは Caltech256 から確認用に１０枚，学習用に確認用の１０枚を除いた画像を用いた．

識別器を作成時に使用した PC やフレームワークなどは表１のとおりである．畳み込みニューラルネットワークについては GoogLeNet のモデルを使用し，パラメータは初期設定の物を使用した．

作成時の識別器への画像の提示は，学習が偏らないように学習用の画像をシャッフルし，オブジェクトの種類をまぜ１回につき１枚の画像を提示した．１００万回提示するのに約４５時間を要した．

図 5 システム管理画面

表 1 PC の性能

項目	性能
CPU	Intel® Xeon® CPU X5482×2
メモリ	16GB
グラフィックボード	GeForce GTX 980 Ti
ディスク	480GB INTEL SSD
OS	Ubuntu 14.0.4 LTS
フレームワーク	Caffe
CUDA	CUDA7.5
cuDNN	cuDNN 5

表 2 学習したオブジェクトの認識精度(%)

物体名	収集した画像 +データセット	収集した 画像のみ
ティッシュ	80	84
自動販売機	79	90
観葉植物	96	98
ゴミ箱	93	92
ドア	88	91
ケトル	96	95
コップ	82	88
パソコン	96	84
テーブル	84	88
スイッチ	82	87
イス	97	91
ドアノブ	100	100
平均	89.41	90.67

5.2 学習したオブジェクトの認識率

学習させたオブジェクトにおいて，自分で集めた画像にデータセットを加えた場合と自分で集めた画像のみの場合の識別器の認識精度を確認する．

認識の確認に使用する識別器は，それぞれ１００万回識別器に学習用の画像を提示し学習させた物を使用した．

表 2 はデータセットを加えた場合と加えていない場合の識別器へ学習時に用いていない各１００枚の画像を入力したときの認識率である．データセットを加え

表 3 未学習のオブジェクトに対する出力率(%)

閾値の設定値	収集した画像+データセット	収集した画像のみ
100	0	0
90	12	26
80	20	37
70	27	48
60	38	59
50	51	73
40	65	86
30	80	96
20	93	100
10	100	100
0	100	100

た場合も加えていない場合のどちらも約９０%の確率で認識に成功する識別器が作成された.

5.3 未学習のオブジェクトの認識率

本システムの物体認識では，識別器に学習させていないオブジェクトを入力した場合には，出力をしないことが望ましい．そのため，出力された適合率がある閾値を超えたら答えることで，未学習のオブジェクトについては答えないことができる．対象の１２種とデータセットに含まれていない種類の画像を用意し，データセットを加えた場合と加えていない場合の識別器に学習させていないオブジェクトの画像を入力して，閾値を超える割合を確認した.

表３は閾値を変えながら未学習のオブジェクトに対して閾値を越えた割合を確認した結果である．自分で集めた画像にデータセットを加えて識別器を作成した場合は，データセットを加えない場合よりも，閾値を超える確率が下がった．よって，本システムのような未学習のオブジェクトに対しては出力したくない場合には，データセットを加えることで出力の確率を下げることが出来ると考えられる.

5.4 識別器の認識精度の推移

0-ノルム最適化のような１回の計算による識別器の作成ではなく，Deep Learning のように何回か演算を繰り返して識別器を作成する場合は，その繰り返しの回数によって認識の精度が変わる．自分で集めた画像にデータセットを加えて作成した識別器は何回でどのくらいの精度になるのかを調査するために，対象１２種類の学習用に用いていない確認用の画像を各１００枚入力して認識率を確認した.

図６は４万回提示するごとに認識率の推移である．最初の方の４万回提示の時点では認識率が８０%程度だが，提示回数が増えるにつれて認識率は上がったり下がったりするがしながら，約９０万回提示で認識率９０%程度になる.

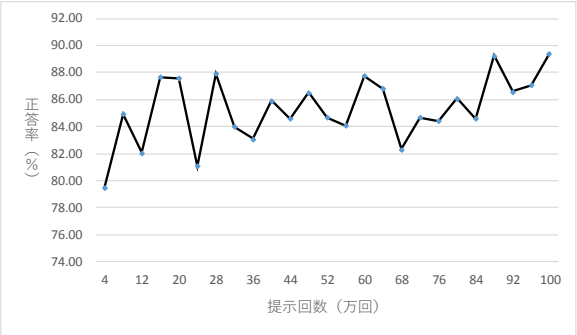

図６ 正答率の推移

6. おわりに

本稿では，視覚障がい者の大学などでの生活をサポートするために，スマートフォンが目の代わりになるシステムの構築を目指している．現存するアプリケーションでは，大学の運営側が画像認識の識別部分を修正することができない．本研究で開発したシステムは，認識結果が不十分な時には画像と読み上げのルールを登録でき，画像認識の識別器を作り直すことができる.

視覚障がい者のための識別器は，学習させたい対象には高い適合率，未学習の対象には低い適合率の出力を出すのが理想である．収集した画像のみではなくデータセットを加えて作成することで，学習したものに高い適合率を出力し未学習に対しては低い適合率をだすことを実験により確認した.

今後の課題は，複数の物体が撮影されたときの対応である.

参考文献

(1) 中村 和弘，青野 嘉幸，田所 嘉昭：“視覚障害者用誘導型歩行支援システム”，電子情報通信学会論文誌 D，Vol.J79-D2，No.9，pp.1610-1618（1996）.

(2) 石川准，兵藤安昭：“GPS による視覚障害者歩行支援システムの開発 (モバイルとインターネットの融合，及び一般).”, 電子情報通信学技報., IA2004-27, pp.51-56 (2005).

(3) 「Tap Tap See」<https://play.google.com/store/apps/details?id=com.msearcher.taptapsee.android> （2016/06/15 アクセス）

(4) 波部斉："ランダムフォレスト." 情報処理学会研究報告 2012 (2012).

(5) Wright, John, et al.：“Robust face recognition via sparse representation.”，Pattern Analysis and Machine Intelligence，IEEE Transactions on 31.2 (2009)，pp.210-227（2009）.

(6) Szegedy, Christian, et al.：“Going deeper with convolutions.”，In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*，pp.1-9（2015）.