

令和 2 年 6 月 1 日現在

機関番号：14501

研究種目：基盤研究(C)（一般）

研究期間：2016～2019

課題番号：16K01547

研究課題名（和文）情景画像からの情報抽出による視覚障がい者支援システムに関する研究

研究課題名（英文）A Study on Support System for Visually Handicapped People by Extracting Information from Scene Images

研究代表者

田中 直樹（Tanaka, Naoki）

神戸大学・海事科学研究科・教授

研究者番号：90188318

交付決定額（研究期間全体）：（直接経費） 3,700,000円

研究成果の概要（和文）：視覚障がい者は、鉄道駅のプラットフォームからの転落事故の増加などが見られ、その安全性の確保や行動範囲を広げるための歩行支援が求められている。本研究課題では、ウェアラブルデバイスなどにより取得した情景画像から、案内看板などの文字列や、点字ブロックなど視覚障がい者の歩行支援オブジェクトの抽出やその種別の判別などに関する研究を実施した。低解像度文字列の強調方式、文字列領域抽出、点字ブロックの判別及び領域抽出方式の開発などの成果を達成している。

研究成果の学術的意義や社会的意義

歩行支援オブジェクトの抽出方式には、主に、ディープラーニングを用いるが、歩行支援を実現するためには、案内標識の文字列や、点字ブロックの領域を抽出しその意味内容、レイアウト構造の解析が必要となる。そのため、学習データの作成方法や、領域抽出・解析方式の開発はディープラーニングの更なる活用に繋がる意味で学術的に意義がある。また、視覚障がい者の歩行支援は社会的なニーズも高いものと考えられる。

研究成果の概要（英文）：People with visual disabilities are seeing an increase in the number of accidents involving falls from railway station platforms, and are required to provide walking support to ensure their safety and expand their range of activity. In this research project, we conducted a study on the extraction of walking support objects for visually impaired persons such as character strings such as signboards and Braille blocks from the scene images acquired by wearable devices, and the classification of them. We have achieved achievements such as low-resolution character string enhancement method, character string area extraction, Braille block discrimination, and area extraction method development.

研究分野：j画像情報処理

キーワード：視覚障がい者 情景画像 オブジェクト抽出 歩行支援 ディープラーニング 点字ブロック 文字列抽出

科研費による研究は、研究者の自覚と責任において実施するものです。そのため、研究の実施や研究成果の公表等については、国の要請等に基づくものではなく、その研究成果に関する見解や責任は、研究者個人に帰属されます。

様式 C - 19、F - 19 - 1、Z - 19 (共通)

1. 研究開始当初の背景

視覚障がい者の鉄道駅のプラットフォームからの転落事故も相次いでおり、視覚障がい者の歩行支援の社会的ニーズも高まっていた。歩行支援が実現できれば、危険の回避とともに要支援者の行動は範囲を広げることにもなるため、歩行支援システムの社会的な意義も大きいものであった。一方、そのような支援システムの構築に当たっては、案内表示板の文字列の抽出認識と、点字ブロックを主とした歩行支援オブジェクトの抽出認識が必要であり、どのような歩行支援オブジェクトを、すなわち、種類やレイアウト構造などを調査し明らかにすること、それらの画像データを得ることが必要であった。さらに、それらを、どのように学習し、どのような結果を提示するかなどについて検討する必要があった。

2. 研究の目的

研究の第1の目的は、ウェアラブルデバイスなどにより得られた情景画像からの情報抽出に基づく、視覚障がい者歩行支援システムを構築することである。情景画像から抽出すべき、文字列情報や点字ブロックなどの、歩行支援オブジェクトに関して実際にウェアラブルデバイスを装着し、何をどのように学習し認識するかについて考察する。低解像度の文字列の拡大・強調方式や、文字列抽出方式、それらの認識のためのディープラーニングのネットワーク構成法や学習データの構成法、学習方式などに関する知見を得ることを目的としている。

3. 研究の方法

(1) 歩行支援のための画像データ取得

ウェアラブルデバイスを用いた歩行支援を目的とした情景画像の取得

ウェアラブルカメラを実装して歩行し、歩行支援のための情景画像を取得し、データベース化する。鉄道駅のプラットフォーム、地下街、歩道、またそれぞれ、昼、夜など様々なシチュエーションを想定したデータ取得をおこなう。その際、カメラの装着位置や装着角度などに関する検討・検証を実施する。

画像データベースの構築

取得した、画像データ(動画)から、静止画を切り出し、支援オブジェクトのカテゴリ化や、領域の切り出しを実施し、ディープラーニングなどによる学習に必要な情報を付加した静止画像データベースを構築する。

(2) 案内標識などの文字列領域の抽出・認識方式

低解像度文字列画像の拡大及び強調

情景画像データベースから得られる文字列は、データ量も少なく文字列のバリエーションも大きくない。そこで、新聞データベース画像を用いた低解像度の文字列画像の強調方式について検討をおこなう。拡大方式は、バイリニア法による拡大処理、バイキュービック法の認識率に対する性能を比較する。また、平滑化二次微分フィルタによる高解像度化手法、拡張平滑化二次微分フィルタによる高解像度化との組み合わせによる認識率の向上に関する検討をおこなう。

情景画像中の文字列領域抽出

まず最初に、文字列抽出について、ディープラーニングの Inception V3 を用いる。Inception V3 は 133 層の CNN であり学習済みのネットワークとなっている。これに、未学習の数層(ここでは 5 層)のネットワークを追加し学習を行う。Inception V3 は、カテゴリ判別モデルであり、本来は、文字列抽出に適していない。そこで、情景画像をブロックに切り分け、ブロックごとに文字列の有無の判別をおこなう。さらに、独自の FCN(FULLY Convolutional Neural Network) ネットワークを用いて、ブロック化を伴わない文字列領域抽出方式を開発する。

(3) 点字ブロックなど歩行支援オブジェクトのカテゴリ認識及び領域抽出方式の開発

ディープラーニングの既存のモデルの性能評価及び学習データ作成

。画像分類モデルについて最初の成功例である LeNet5(1998)、AlexNet(2012)、ResNet(2015)及び MobileNet(2017)について調査する。物体検出モデルを用いた学習をおこなうためには学習用画像に対するアノテーションが必要となる。ここでは VGG Image Annotator を用いた Bounding box 情報の付加と、COCO Annotator を用いた輪郭線情報(多角形の Bounding box)情報の付加方法について検討し、Mask R-CNN、YOLO V3 を用いて約 87% の Bounding box 検出率を得られた。YOLO V3 及び Tiny YOLO v3 を用いた物体検出問題学習について検討する。

点字ブロックの抽出・認識

FCN を用いた

点字ブロック領域抽出とブロック種別識別について検討する。点字ブロックには大きな種別として、進行方向や分岐などを示す“誘導ブロック”と危険を警告する“停止ブロック”とがある。また、点字ブロックは進行方向を示すために分岐や交点を持つなど様々なレイアウト構造となっている上に誘導と停止の2種別があるためにさらに多種のレイアウト構造を持つ。まず、CNN を用いた点字ブロックのカテゴリ識別方式を提案し識別実験を実施している。この識別は、入力画像を“直進”と“分岐”“非点字ブロック”の3カテゴリに判別するいわゆるカテゴリ分類問題である。次に、FCN を用いた点字ブロック領域抽出領域抽出方式を提案する。

4. 研究成果

1) 歩行支援のための画像データ取得

ウェアラブルデバイスを用いた歩行支援を目的とした情景画像の取得

ウェアラブルなカメラを装着して歩行時の動画を取得した。カメラの装着位置について、当初は眼鏡や帽子に装着し、目線の画像を取得したが、案内板などは比較的確認が容易であるが、足元の点字ブロックは確認が不可能であり、前方の点字ブロックも他の歩行者に遮られるなどほぼ確認ができないことが判った。腰部に下向きに装着した場合、点字ブロックについては、足元も前方もかなり視認性が向上した。しかしながら、腰部からの画像では、案内標識などはほぼ確認は不可能であった。そのため、頭部、腰部の2か所のカメラの装着が望ましく、その両方の動画を取得した。

画像データベースの構築

取得した動画をスライスすることにより、静止画を得る。静止画の、文字列や点字ブロックなどの支援オブジェクトのカテゴリとその領域情報を目視、手動により各画像に付加した。文字列も、点字ブロックも、本来は、平行線で構成される矩形領域であるが、撮影角度によりかなりの透視変換を受けており、抽出認識処理の困難さを招く。そこで、透視変換パラメータを推定ないしは予測し逆透視変換する方式も導入した。

(2) 案内標識などの文字列領域の抽出・認識方式

低解像度文字列画像の拡大及び強調

低解像度の文字列として新聞画像を用いた。新聞画像は、背景部に変色やシミが見られ認識率の低下を招く。そこで、グラフィックにより背景部と文字線部分の分離を行い、背景部のノイズの影響を軽減を試みた。原画像、バイリニア法による拡大処理、バイキュービック法による拡大処理、平滑化二次微分フィルタによる高解像度化手法、拡張平滑化二次微分フィルタによる高解像度化を施した画像をそれぞれ OCR による認識処理をおこない、それらの中で認識距離が最小となる手法を選択する方式を開発した。

実験には、神戸大学の「新聞切抜文庫」から1文字ずつを切り出し合計694文字を実験対象として用いた。これらの画像のOCRでの認識率は原画像79.97%、バイリニア法80.25%、バイキュービック法82.83%であり、低いものであった。これに対して平滑化二次微分フィルタにより横線強調し高解像度化を施した場合89.05%に向上した。さらに、識別距離による適応的な選択手法を適用したところ認識率が90.35%となり最も高い認識率を得ることができた。

情景画像中の文字列領域抽出

Inception V3モデルを用いて、情景画像画像からの文字列領域実験を実施した。本来、カテゴリ判別モデルであるInception V3を領域抽出に対応させるために、入力画像を小さなブロックに分け、ブロック単位で文字領域、非文字領域に判別する文字列領域抽出方式を提案した。その結果、被験画像においてはかなりの精度で文字列領域の抽出が可能であった。しかしながら、ブロック単位での判別であり、判別閾値の設定がセンシティブ、かつ、ノイズの多い抽出結果となった。それらのノイズについては、モルフォロジィフィルタによりリファインが可能であることを示した。また、独自のFCNネットワークによる文字列抽出方式を開発した。同ネットワークは、文字列のスケールの大きな変動に対応できるように3つのルートを持つものとしている。このネットワークによって被験画像について高精度な文字列抽出が可能となった。しかしながら、特定のカラー（赤系統）が支配的な被験画像では、抽出精度の低下が見受けられ、色処理方法に関する課題も明らかとなった。

(3) 歩行支援オブジェクトのカテゴリ認識及び領域抽出方式の開発

ディープラーニングの既存のモデルの性能評価及び学習データ作成

各種提案されているディープラーニングの各モデルの性能評価を実施した。まず、カテゴリ判別モデルについて、LeNet5の場合、学習そのものは収束していると考えられるがテスト用画像に対する正当率は約24%に留まった。これは、LeNet5の入力画像サイズが32×32のモノクロであり極めて低解像度かつ低情報量でありネットワークも単純で小規模(モデルサイズ約815KB)である。そのため本被験画像データである動物の種別の分類には十分な情報を持たないことが考えられる。次にAlexNetでは、正答率は約59%となった。AlexNetは入力画像サイズ227×227、層数11及びモデル規模は約470MBでありLeNet5と比較すると格段に大規模なネットワークであり正答率も向上するがまだ十分な性能を持つとは言えない。ResNetでは、正答率約81%となりほぼ実用的な性能となった。ResNetは入力サイズ224×224、層数18である

Mobilenetは、モバイルなどのストレージやCPUスペックが制限される環境下での使用を想定しコンパクトなネットワークモデルとして提唱された。Mobilenetを用いた分類実験では、正答率約84%となり今回比較したモデルでの最高値となった。入力サイズは224×224、層数17であるにもかかわらずモデルサイズは約40MBに抑えられている。このようにMobilenetは分類性能、モデルサイズともに実用性能が優れることがわかる。

また、MobileNet はモデルサイズ、計算コスト共にモバイル機器での利用が可能でありモバイル機器を用いた分類モデルの構築を容易とする。

次に、物体検出モデルを用いた学習をおこなうためには学習用画像に対するアノテーションが必要となる。ここでは VGG Image Annotator を用いた Bounding box 情報の付加と、COCO Annotator を用いた輪郭線情報(多角形の Bounding box)情報の付加方法を提案している。物体検出モデルでは、40 種の動物それぞれに 75 枚の学習用データを準備し、テストデータとしてそれぞれ 25 枚を準備した。まず、Mask R-CNN を用いた輪郭線実験をおこなっている。検出の評価には領域の重なる面積が 50%以上となる IoU=50 を指標として用いている。Mask R-CNN では、約 82%の領域検出率となった。また、Mask R-CNN ではモデルサイズは 256MB となりまた検出には多くの計算機コストがかかり PC やスマートフォンの CPU では数秒を要した。次に、YOLO V3 を用いて約 87%の Bounding box 検出率を得られた。YOLO V3 はモデルサイズ約 740MB と大きく学習に要する時間も大きいのが欠点でありまた、検出の計算コストも高くモバイル機器での利用に適していない。Tiny YOLO v3 は Bounding box の検出精度は約 61%と低くなるがモデルサイズが小さく、高速な検出速度を持つためモバイル機器での使用に適している。

点字ブロックの抽出・認識

CNN を用いた点字ブロックのカテゴリ識別実験を実施した。この識別は、入力画像を“直進”と“分岐”“非点字ブロック”の 3 カテゴリに判別するいわゆるカテゴリ分類問題である。そのため、4 層の Convolution 層と 1 層の MaxPooling 層及び 3 層の全結合層構造を持つ CNN を用いた。また、4 層の Convolution 層の中間に Batch Normalization 層を挿入して学習の効率化及び過学習の軽減を図っている。入力画像サイズは 72×128 の RGB カラー画像とし、出力ユニット数は 3 となる。学習には、各カテゴリ 30 枚の計 90 枚を用い、検証用データには各カテゴリ 10 枚の計 30 枚を用いている。識別実験を実施した結果、黄色ペイントをご認識した 1 例と学習データに含まれない複雑レイアウト画像 1 例を除いてすべて正識別となった。

FCN 構造を採用し、8 層の Convolution 層、1 層の MaxPooling 層構造とし 1 層の Batch Normalization 層を挿入したものをを用いた点字ブロックの領域抽出をおこなった。1 つ目の領域抽出実験は、“点字ブロック領域”の抽出である。学習用データを 330 枚とし検証用データを 80 枚とした。領域抽出実験をおこなった結果、点字ブロック、非点字ブロックのそれぞれの再現率は 98.17%、87.70%となりやや過剰抽出気味ではあるが十分な実用性能が得られている。次に、“誘導ブロック”“停止ブロック”の領域について前述のネットワークとほぼ同様の構成のネットワークを用いた領域抽出実験をおこなった。学習用データ、検証用データも同じものを使用している。領域抽出実験をおこなった結果、誘導ブロック、停止ブロック、非点字ブロックの領域のそれぞれの再現率は、84.24%、86.04%、99.48%となった。前述の点字ブロック領域の抽出実験とは逆にやや抽出漏れ気味ではあったが十分に実用的な性能を得ている。また、誘導ブロック、停止ブロックの再現率がやや低下した原因については、1 枚の点字ブロックに誘導と停止が混在するものについてその判別がやや正確性に欠けたためと考えられる。

5. 主な発表論文等

〔雑誌論文〕 計1件（うち査読付論文 0件 / うち国際共著 0件 / うちオープンアクセス 0件）

1. 著者名 田中直樹 岡本章裕 中田裕一 廣野康平	4. 巻 第28巻9号
2. 論文標題 船用レーダにおける映像特徴に基づいたターゲット抽出方式	5. 発行年 2017年
3. 雑誌名 画像ラボ	6. 最初と最後の頁 21-28
掲載論文のDOI（デジタルオブジェクト識別子） なし	査読の有無 無
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

〔学会発表〕 計8件（うち招待講演 0件 / うち国際学会 1件）

1. 発表者名 Takashi Nagamatsu, Mamoru HiroeGerhard Rigoll
2. 発表標題 Corneal-Reflection-Based Wide Range Gaze Tracking for a Car
3. 学会等名 International Conference on Human-Computer Interaction HCI 2019 (国際学会)
4. 発表年 2019年

1. 発表者名 熊澤 宏之
2. 発表標題 機械学習を用いた移動モード検知の精度改善 - GPSを用いない方式の検討 -
3. 学会等名 電気学会 電子・情報・システム部門大会
4. 発表年 2019年

1. 発表者名 任 散陽, 中田裕一, 田中直樹
2. 発表標題 層学習と画像セグメンテーションに基づいた動物種類と位置の検出
3. 学会等名 情報処理学会全国大会
4. 発表年 2020年

1. 発表者名 若松直人, 中田裕一, 田中直樹
2. 発表標題 点字ブロック領域・種別識別による視覚障がい者支援システムに関する研究
3. 学会等名 情報処理学会全国大会
4. 発表年 2020年

1. 発表者名 熊澤 宏之
2. 発表標題 360°カメラを用いた研究室顔認識システム - 処理の高速化の検討 -
3. 学会等名 電気学会情報システム研究会
4. 発表年 2019年

1. 発表者名 小谷匡史 中田祐一 田中直樹
2. 発表標題 低品質新聞画像における多手法による適応的なOCR認識精度向上に関する研究
3. 学会等名 電子情報通信学会
4. 発表年 2017年

1. 発表者名 松尾良磨 酒井大貴 福森聡 山本倫也 長松隆
2. 発表標題 車用広範囲視線計測装置に関する研究
3. 学会等名 ヒューマンインタフェースシンポジウム2016
4. 発表年 2016年

1. 発表者名 荒井 悠, 亀山 真由子, 長松 隆, 山本 倫也
2. 発表標題 眼球の光軸中心回転体モデルに基づく視線計測手法による強膜反射時の視線推定
3. 学会等名 ヒューマンインタフェースシンポジウム2016
4. 発表年 2016年

〔図書〕 計0件

〔産業財産権〕

〔その他〕

-

6. 研究組織

	氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考
研究分担者	鎌原 淳三 (Kamahara Jyunzo) (60283917)	神戸大学・海事科学研究科・准教授 (14501)	
研究分担者	長松 隆 (Nagamatsu Takashi) (80314251)	神戸大学・海事科学研究科・准教授 (14501)	
研究分担者	熊澤 宏之 (Kumazawa Hiroyuki) (00592320)	大阪産業大学・工学部・教授 (34407)	