

オフィス環境における文字情報の検出と利用に関する研究

A study on detectoin and utilization of text information in office environment

宮本圭、上野敦志、武田英明

Kei Miyamoto, Atsushi Ueno, Hideaki Takeda

奈良先端科学技術大学院大学 情報科学研究科

Nara Institute of Science and Technology

The purpose of this study is to construct an intelligent robot system that can support people's everyday life and activities. To realize it, we focus on text information in the environment, since there is a lot of text information in the environment where people live and it is useful as names or explanations of objects or places. We designed and implemented a robot system that can deal with text information. The system consists of a mobile robot with character recognition and string search functions. The mobile robot moves around in the environment to takes pictures to look for texts. In this study it takes pictures of bookshelves and posters. From pictures character strings are detected and recognized by using OCR library. Since the result contains a lot of errors, we provide a retrieving method that is applicable to texts that contains errors. This study shows that people can compensate their uncertain knowledge and find missing information by utilizing text information in the environment collected by a mobile robot.

1 はじめに

近年、人間が存在する環境で作業し、人間の生活や活動を支援するロボットの研究がなされてる。[1, 2] そのようなロボットが、人間にとって役立つ機能を提供するためには、ロボットが、環境中のものごとについての情報を持っていなければならない。

本研究では、ロボットができるだけ人間の助けをかりずに環境から情報を獲得し、人間に役立つ機能を提供することを目指した。このことが可能な情報源として、本研究では、環境中の文字情報に着目した。環境中の文字情報を収集し、それを人間が利用できるようにすることで、人間の存在する環境で、その環境についての情報を提供できるロボットの実現を目指した。

人間の持っている環境に対する情報は、不確かである。また、人間の存在する環境は人間によって絶えず変化させられるので、ある人が持っている環境についての情報は、古くなっていて、現在の状態にはあてはまらないこともある。人間は環境を詳しく調べているわけではないので、見落とししている情報もあると思われる。このようなものをロボットが集めた情報で補うことを目的としている。本研究で構築したシステムでは、ロボットが環境中の文字情報を収集する。それを人間が参照することで、環境中のどこにどんな情報があるかを知ることができる。これを実現するために、本研究では、ロボットが集めた文字情報に対して、文字による検索を

行なっている。人間が知りたいことを文字で入力すれば、ロボットが集めた文字情報から一致するものを探し出し、その文字情報が得られた画像と画像を撮影した時刻を人間に示すことによって、人間に情報を提供する。このようなシステムを構築し、研究室内にそのロボットシステムを導入して、評価を行なった。

2 人間と共存するロボット

人間が存在する環境において作業するロボットを実現するためには、様々な課題を解決する必要がある。産業用ロボットは、ロボットのために整備された環境で作業していたが、人間の存在する環境で作業するためには、ロボットが環境に合わせて行動しなければならない。また、人間と共存するためには、人間とのコミュニケーションを行なう能力が必要である。そして、人間の生活する場面で人間にとって役立つ機能を提供する必要がある。

ロボットが人間と共存し、作業するため、および、人間に役立つ機能を提供するためには、ロボットは環境についての情報を持たなければならない。本研究では、ロボットが持つ環境中のものごとに関する情報源として、情報の獲得方法および情報の有用性の観点から、文字情報を選んだ。

3 環境中の文字情報

本研究では環境中のものごとに関する情報源として、環境中の文字情報に着目した。人間と共存するロボットが持つ情報として、文字情報は以下の点で優れている。

- 人間が理解可能
- 人間の存在する環境に多く存在する
- 場所やものの名称、説明などを表している
- 文字による検索が可能なので、利用しやすい

本研究で構築したシステムでは、ロボットが室内を移動し、指定された場所で画像を撮影する。撮影対象としては、本棚と教官および博士課程の学生の研究内容が書かれているポスターを選んだ。

環境中の文字情報は、たいてい鉛直方向または水平方向に書かれている。また、文字が書かれている領域は、一般的に文字色と背景色のコントラストが大きいという特徴がある。これらの特徴を利用した文字領域の検出、文字認識、収集した文字列の連結について以下に述べる。また、傾いた状態で存在する文字情報についても述べる。

3.1 文字列の検出

3.1.1 文字列の検出 本棚

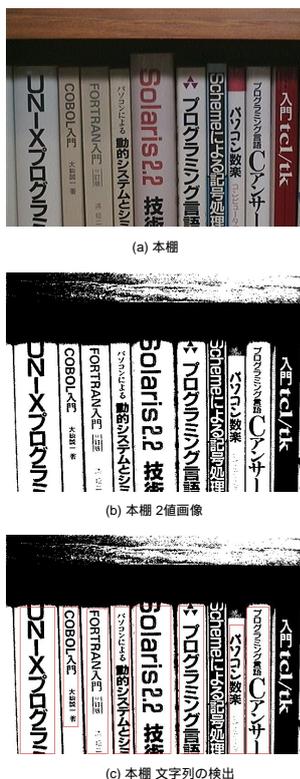


図 1: 本棚における文字列の検出

ロボットが撮影した本棚の画像を図 1(a) に示す。画像のサイズは 640 × 480 画素である。本研究では、情景画像から文字列領域を抽出するために、

澤木らのテンプレート学習による本棚画像中の文字認識 [4] の手法を基にして、本棚画像から文字列の抽出を行なった。その手順について述べる。

1. 2 値化

ロボットに搭載されたカメラで撮影したカラー画像を適当な閾値で 2 値化する。(図 1(b))

2. 本の境界線の検出

本棚画像を 2 値化すると、本と本の境界は、多くの場合、黒い線として表される。この線を利用して、1 冊の本を切り出す。本棚画像の上端の各点から下端に向かって走査し、ランレングスに基づいて本の境界線を検出する。走査する角度は真下方向から ± 2 度の範囲で、0.2 度きざみで行なう。1 つの上端点について、ランレングスが閾値以上かつ最も大きかった角度を、その点での本の境界線の傾きとする。隣接する本の境界線の間を本の背表紙領域とする。

3. 背表紙領域中の文字列領域の検出

各背表紙領域の上端点から本の境界線の傾き方向に走査して、白黒反転回数の射影分布をとり、それが一定値より多い範囲を文字列領域として検出する。(図 1(c)) また、第 3.2.3 節で述べる文字列の連結のために、文字列領域を囲む矩形領域の座標を保存しておく。

4. 段の検出

それぞれの画像から得られた文字情報は、環境中の 1 つの文字列の部分文字列になっているため、文字認識後に連結する。ここで、文字列を連結するときのために本棚の段を検出する処理を行っておく。その方法は、撮影した画像を水平方向に走査して、黒点が 470 個以上連続してあれば、それは棚の下にできた影であるとみなし、その領域までが本棚の 1 つの段とする。図 2 での、(a) から (c) までと (d) から (f) までが段となる。

このようにして得られた文字列領域を切り出して、OCR ライブラリを利用して文字認識を行なう。

3.1.2 ポスターにおける文字列の検出

ポスターの文字列の検出は、本棚の文字列検出方法から 2. 本の境界線の抽出と 5. 段の検出を除き、3. 背表紙領域中の文字列領域の検出をポスター用に変更した処理で行なっている。ポスターは横書きなので、画像の左端の各点から水平方向に走査する(図 3)。

3.2 カメラ画像からの文字認識

本研究では、市販の OCR ライブラリを利用するが、普通に撮影したのでは、文字認識はできない。今回使用した OCR ライブラリに限らず、現在、市販されている文字認識ソフトは、スキャナで読み込



図 2: ロボットが 1 地点で撮影する本棚の画像を 2 値化したもの

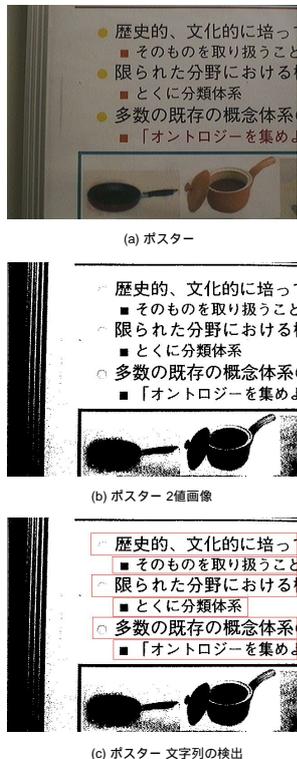


図 3: 文字列の検出 ポスター

まれた文書を対象としている。よって、スキャナより解像度の粗いカメラ画像からの文字認識は、解像度が足りないためにうまくいかない。そこで、本研究ではカメラのズーム機能を使い、OCR ライブラリで文字認識が可能になるまで解像度を上げて、画像を撮影した。

3.2.1 ズーム

本研究では、文字認識に RICOH の OCR ライブラリを利用した。この OCR ライブラリは、400dpi でスキャンされた文書の 7 ~ 31pt の文字を認識できる。pixel に換算すると、文字の大きさが 40pixel 以上になるまでズームすれば、カメラ画像からでも文字認識が行なえる。本研究で構築したシステムでは、移動ロボットに 12 倍ズームまで可能な SONY 製のカメラ EVI-D30 を搭載してある。このカメラで、本の背表紙に書かれている本のタイトルとポスターの文字が 40pixel 以上になるまでズームし、画像を撮影した。撮影対象までの距離は約 50cm とした。このようにして撮影した本棚 (図 1) とポスター (図 3) の認識結果を示す。

- 本棚 (図 1)
 - え 青 2 言 責 寅 へ) アンサ
 - バリコ、】 数 楽 r
 - プログラミング 言 善 鼓
 - 0 町 「 句 母 技 t
 - ? ラ に き 山 の シ ス テ ム と シ コ
 - 【 ○ 刀 「 コ し 2 入 門 に 惟 ・ -
 - 〇 ○ W O r 入 門 奪 畠 布
 - N I X プ ロ グ 之
- ポスター (図 3)
 - 歴史的、文化的に培っ
 - 1 そのものを取り扱うこ
 - 限られた分野における
 - 1 とくに分類体系
 - 多数の既存の概念体系
 - 1 「オントロジーを集め

3.2.2 分割撮影

文字認識できるまでズームして撮影すると、1 枚の画像に文字列が写りきらない。そこで、カメラのパンチルト機能を利用して、1 つの地点でカメラを仰角方向に +25 度から -25 度まで 10 度きざみで動かして 6 枚の画像を撮影した。(図 2) 本棚とポスターに対して平行移動して 6 枚の画像を撮影することを繰り返し、対象を撮影した。

3.2.3 本棚における文字列の連結

本棚を撮影した 1 枚の画像には文字列全体は写っていない。完全な文字列を得るために、撮影した画像から得た部分文字列を連結して、1 つの文字列にする処理を行なった。文字列の連結は、文字列検出の際に得た文字列領域を囲む矩形領域の座標をもとにして、本棚の 1 つの段にある文字列同士をつないだ。上下関係にある画像には、それぞれの画像の約

4分の1程、重複している部分がある。文字列連結の際に、上下の画像の重なり部分から得られた全く同じ文字列は、重複させずに連結した。重なり部分から得られた文字列が異なる場合は、どちらが正しいかを判定することができないので、重複させたまま連結した。また、この時に文字列が写っている画像も文字列と同じ関係で連結した。

3.2.4 ポスターにおける文字列の連結

ポスターは横書なので、文字列連結は水平方向に行なわなければならないが、以下の理由のために行なわなかった。ロボットは、ある地点で撮影対象の正面から画像を撮影する。その後、撮影対象に対して平行に移動し、再び正面から画像を撮影することを繰り返す。この移動が正確ではないため、移動後のロボットの位置は意図した位置からずれている。そのため、文字列および画像の水平方向の連結は困難になる。ポスターから得た文字情報は部分文字列のまま利用するが、ポスターは横書であるので、横長のカメラ画像に写る文字列は本棚より長く、得られた文字情報の誤りも本棚より少ないので、文字列を連結しなくても、検索対象として利用することができた。

3.3 傾いた文字列

環境中の文字情報は、たいてい鉛直方向または水平方向に存在するが、そうでないものもある。本棚にある傾いている本などがそうである。このような文字情報を得るために、本の境界の検出時に検出角度の範囲を広げて文字列を検出した。その後、傾きをなくすように回転させ、文字認識を行なった。しかし、傾いた文字列を回転して得た文字認識結果は、回転せずに得た認識結果に比べて、非常に誤りが多かった。よって、今回は、傾いた文字列は文字情報を得る対象から外した。

4 環境中の文字情報を収集するロボットシステム

4.1 ロボットシステムの構成

本研究のロボットシステムは、移動ロボットと端末から構成されている。移動ロボットは、Real World Interface社のB21 mobie robotを用いた。このロボットにはSONY製のカメラEVI-D30を搭載してある。ロボットに対する命令や集めた文字情報に対する検索は、端末からWWWブラウザをインターフェースとして行なう。

4.2 ロボットの行動

ロボットが環境中の文字情報を収集する手順は以下のようにになっている。

1. ロボットが指定された地点まで移動
2. 文字情報を得るための画像を撮影

1.、2.を繰り返し、全ての指定された地点で画像を撮影する

3. 文字列の検出
4. 文字認識
5. 文字列の連結

文字認識結果は、その文字列が写っている画像、画像を撮影した時間とともに記録される。

4.3 インターフェース

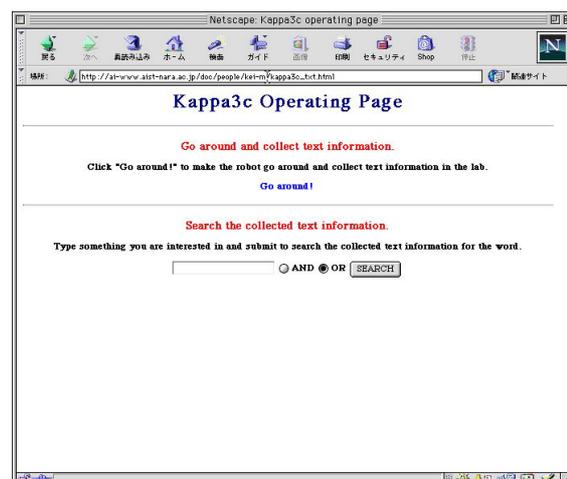


図 4: システムのインターフェース

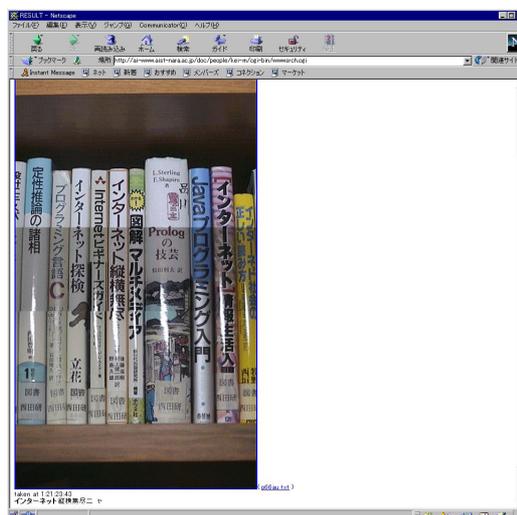


図 5: 検索結果 (検索語は「インターネット」)

文字情報を収集するためのロボットの巡回の命令、および集めた文字情報に対する検索のインターフェースとして、WWWブラウザを用いた。(図4)人間がWWWブラウザのGo aroundと書かれた文字をクリックすると、ロボットが巡回を始める。集めた文字情報に対して検索する時は、WWWブ

ラウザのフォームに検索語を入力する。検索は、第4.4節で述べる誤りを含んだテキストに対する検索方法で行なう。検索結果は、検索語からつくられた誤りを許すパターンのうち、より誤りの少ないパターンと一致したのから順に表示される。図5に検索語「インターネット」で検索した結果を示す。検索結果として、文字情報が得られた画像、画像を撮影した時刻、検索パターンと一致した文字列が表示される。表示された画像中のどの文字列が検索にかかったかを示すため、検索にかかった文字列が抽出された場所を矩形で囲んで表示する。

文字情報をとりだした画像からは、場所がわかりにくい。ユーザーが画像から撮影場所を特定しやすいように、ロボットは拡大していない画像を撮影している。ユーザーはこれを見ることもできる(図6)。

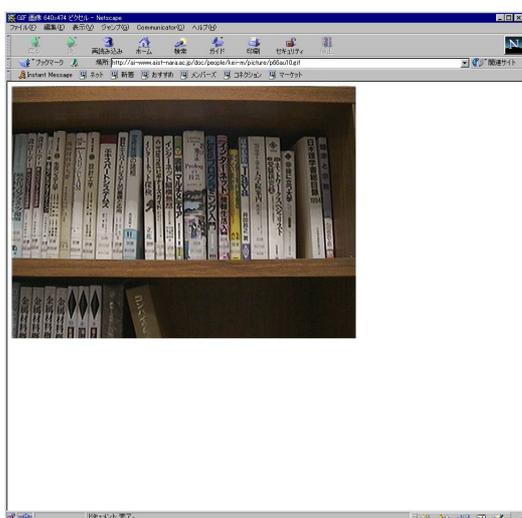


図 6: 文字情報が得られた場所を知るための画像

4.4 集めた文字情報の利用

本システムでは、ロボットが集めた文字情報は、人間が関心のあることについて調べる際の検索対象となる。しかし、ロボットが集めた文字情報は多くの誤りを含んでいる。このようなOCRによる認識誤りを含むテキストに対する検索手法として、類似文字テーブルを用いる方法 [5] などが提案されている。本研究では、本棚およびポスターの文字認識結果を対象とする検索手法として以下のような方法を用いた。ユーザーから検索語が入力されると、それに対して、誤りを許すパターンを作り出す。誤りを許すパターンは、検索語中の文字を1文字以上の任意の文字を表すワイルドカード「*」で置き換えることによってつくられる。置換する文字数は、1文字から検索語の文字数より1文字少ない文字数まで変化させる。下にユーザーが検索語として「人工知能」を入力した場合に作られるパターンを示す。

検索語 人工知能

- 1 文字置換したパターン *工知能、人*知能、人工*能、人工知*
- 2 文字置換したパターン **知能、*工*能、*工知*、人**能、人*知*、人工**
- 3 文字置換したパターン ***能、**知*、*工**、人***

このようにして作成したパターンは、パターン中のワイルドカードが多くなるほど、多くの誤りを許すものになる。検索には、誤りの少ないパターンから順に検索語として用いる。検索結果を多く出力すれば、もれなく検索できるが、ユーザーが求めている情報も多くなる。本研究では、適合率と再現率を等価値に扱うことにした。ロボットが撮影した画像から無作為に15個の言葉を選びだして検索を行ない、適合率と再現率が等しくなる上位5個を出力することにした。

5 システムの評価

5.1 収集した文字情報の評価

5.1.1 本棚

文字認識対象文字列数	570 個
文字認識対象文字数	2635 文字
総文字列数	588 個
文字認識結果の文字数	2820 文字
正しく認識できた文字数	702 文字
文字認識結果に対する認識率	$702/2820 = 24.89\%$
対象文字に対する認識率	$702/2635 = 26.64\%$

表 1: 文字認識結果の統計 本棚

本棚の文字認識結果の統計を表1に示す。文字認識対象文字列数および文字数とは、今回の実験で文字認識しようとしていた文字列および文字の数である。これは、本の背表紙に書かれている40pixel以上の大きさの文字である。このような文字のは、多くの本のタイトルと一部の著者名である。

総文字列数とは、ロボットが撮影した画像から抽出された文字列の数である。文字認識結果の文字数とは、抽出された文字列を文字認識した結果、得られた文字数である。これには、文字でないもの(イラストなど)を誤って文字列として検出し、文字認識しているものも含まれている。また、1文字を誤認識して、2文字以上の文字を出力してしまっているものも含まれている。

表1より、文字認識対象文字数に対する認識率の方が文字認識結果の文字数に対する認識率より高くなっている。これは、本研究での文字列を検出する際の方針によっていると思われる。本研究では、できるだけもれなく文字列を検出したいと考え、文字でない可能性があるものも検出している。抽出された文字でないものは、文字認識されて意味のない文字となるので、検索の際には害にならないと思われる。

5.1.2 ポスター

文字認識対象文字列数	260 個
文字認識対象文字数	2292 文字
総文字列数	264 個
文字認識結果の文字数	2519 文字
正しく認識できた文字数	1569 文字
文字認識結果に対する認識率	$1569/2519 = 62.29\%$
対象文字に対する認識率	$1569/2292 = 68.46\%$

表 2: 文字認識結果の統計 ポスター

ポスターの文字認識結果の統計を表 2 に示す。本棚と同じ理由で、文字認識結果の文字数が文字認識対象文字数より多くなっている。認識率は、本棚よりも高くなっている。これは、ポスターに印刷されている文字のほとんどは、本棚と違って、デザインされていない書体であるためだと思われる。

5.2 検索方法の評価

第 4.4 節で述べた検索手法を用い、検索語「プログラム」で検索した例を表 3 に示す。

パターン	一致した文字情報
プログラ*	プログラミング言語 (プログラミング言語 C)
プログラ*	プログラミングの藍グの基礎 (プログラミングの基礎 C)
プログラ*	@肌知識プ] 別技プログラミング (知識プログラミング)
プロ*ラム	プロクラム改良学冬 (プログラム改良学)
プログ**	N I X プログ之きき鏡翼岬蟬 (UNIX プログラミング環境)

表 3: 検索語「プログラム」での検索結果

表 3 中の「プログラム改良学」は、ロボットが集めた文字情報のなかでは「プロクラム改良学」として保存されているが、パターン「プロ*ラム」と一致して検索結果に含まれている。このように集めた文字情報に多少の誤りが含まれていても、この手法を用いれば検索することができる。また、「プログラム」の語尾が変化した「プログラミング」も検索されている。この検索手法では、文字表現が類似した類義語も同時に検索することができる。検索語と文字表現が類似した類義語は、ユーザーにとって関心のある事柄である場合が多いので、有効な情報であると思われる。

この検索手法の欠点は、短い検索語を用いて検索した場合、その検索語からつくった誤りを許したパターンが、検索語とは関係のない文字列と一致してしまうことが多くなることである。表 4 に検索語「図書」で検索した結果を示す。

ロボットが撮影した画像から無作為に 30 の言葉を選びだし、再現率と適合率を算出した。再現率は 57.3%、適合率は 59.3% であった。再現率は、撮影した画像に写っている文字情報に対して、どれだ

パターン	一致した文字情報
図書	図書駅・ご書舵・情轍学機談野・肘 (図書館・情報学概論 第二版)
図*	楓理学の轟・、】一図 一 (物理学の最先端常識 I)
図*	仏並列図! 並列図形処理 (14 並列図形処理)
*書	野 ・標準 L A N 教科書又 1 首、妾出 Y (標準 L A N 教科書 上)

表 4: 検索語「図書」での検索結果

けもれなく検索できているかを表す。適合率は検索結果中にどれだけユーザーの求める情報が含まれているかを表す。これらの数値から、環境中の文字情報がある程度、利用できるようになったと言える。

6 まとめ

本研究で構築したシステムで、ロボットが環境中から抽出した文字情報を人間にとって価値があり、利用できる情報にすることができた。その結果、環境中のどこに求める情報があるのか知ることが可能となった。また、文字による検索を用いることで、大量の情報の中に含まれることによって、人間が見逃してしまうような情報を提示することができるようになった。

本研究で収集した文字情報は誤りの多いものであったが、環境中に存在するデザインされた字体の文字を高い確率で認識できるようになれば、得られる文字情報がより正確なものになり、利用価値の高いものになるだろう。また、人間と共存するロボットに対して、本研究で用いた環境から情報を得る方法を人間が知識を与える方法や人間とのインタラクションで情報を得る方法と組み合わせれば、よりつきあいやい人間共存型ロボットが実現できるものと思われる。

参考文献

- [1] 松井俊浩, 麻生英樹, 浅野太, John Fry. オフィスロボット Jijo-2 の対話理解. 日本ロボット学会第 16 回学術講演会, 1998.
- [2] 内田 恵久, 武田 英明, 西田 豊明. 弱い知識構造に基づくロボットの行動生成. 情報処理学会研究報告, 知能と複雑系 111-4, 1998.
- [3] 武田英明, 上野敦志, 佐治禎基, 中野剛, 宮本圭, 吉村俊哉, 片山顕正. 自律型知識メディアとしてのロボット オフィスワーク支援ロボットの試作. 第 5 回ロボティクス・シンポジウム, 1999. (予定).
- [4] 澤木 美奈子, 村瀬 洋, 萩田 紀博. テンプレート学習による本棚画像中の文字認識. 電子情報通信学会技術研究報告, PRMU98-155, 1998.
- [5] 太田 学, 高須 淳宏, 安達 淳. 認識誤りを含む和文テキストにおける全文検索手法. 情報処理学会論文誌, Vol.39 No.3, pp.625-635, 1998.