

ステレオビジョンによる3次元空間内で描かれた一筆書き文字の認識

山内 幸治・木村 幸平*

The recognition of traversal characters on three-dimensional space by stereo vision

Yukiharu YAMAUCHI and Kohei KIMURA*

Abstract

In this paper, recognition by stereo vision of traversal characters which was drawn on three-dimensional space is described. Significant improvements in technology have come to support disability persons in the welfare field. In order to cope with different users, development of systems tailored to each situation is needed. Our laboratory has developed a device to recognize characters for user who can not use general input devices. But user who can use the device is only the user who can use his hands. In this study, we propose a method to draw characters with the face instead of hands. The user can draw the characters without moving the hand. This system has two cameras to shoot in the person who draw characters, and recognize the characters by getting the depth information.

Keywords : stereo vision, character recognition, piecewise linear approximation

1. 序論

近年、介護や福祉の分野において、障害者、高齢者支援を目的とした技術開発が活発に行われている。その一環として、マウスやキーボードなどの特殊な装置を用いずに、文字入力を可能とした空間内での手書き文字認識の開発が進められている。空間手書き文字認識は、指や顔で空中に文字を描き、その軌跡をカメラで読み取ることによって、描かれた文字を認識するものである。これにより、言語障害をもち、指先が不自由な人がいたとしても、問題なく文字入力を行うことができる。

しかし、空間手書き文字認識では、カメラの光軸に対して垂直な平面上に文字を描かなければ、文字の軌跡は変形する。カメラの光軸に対し水平方向に傾きがある場合、カメラが取得した文字の軌跡は横に縮み、カメラの光軸に対し垂直方向に傾きがある場合、カメラが取得した文字の軌跡は縦に縮む。このように文字の軌跡が変形すると、文字の認識精度が低下する恐れがある。このような状況は、高齢者や障害者にとっては十分に起こりえることである[1-6]。

これらの問題は、カメラ単体である場合、奥行き情報が取得できない点にある。そこで、本稿ではステレオビジョンによる空間手書き文字認識を提案する。ステレオビジョンは、対象物をカメラ2台でとらえ、三角測量の原理に基づいて奥行き情報を取得するというものである。これを用いることにより、描画平面のひずみや傾きによる軌跡の変形の実態を知ることができ、軌跡の変形に対して補正を行うことが可能となる。

しかし、ステレオビジョンを用いる場合、対象物が常に2台のカメラの撮影範囲内にあることが前提となる。もし、対象物が一方のカメラの撮影範囲からはみ出してしまった場合、その時点の奥行き情報が取得できなくなる。このように、ステレオビジョンの導入には撮影範囲が縮小してしまうという問題があり、文字を描くというような動きのある対象物を扱うには不便である。

そこで、この問題を補うために、ステレオ化したカメラを運動対象物に追従させることにより、この問題を解消することを試みる。カメラが運動対象物を追従することによって、運動対象物がカメラの撮影範囲からはみ出すことがなくなり、描画領

域が大幅に拡大する。また、描画領域が拡大すると、空間手書き文字も大きく描くことができ、文字認識率の向上へとつながる[7-9]。

本稿では、手で文字を描画せずに顔の動きを利用して、空間中に描画された文字を認識する手法を提案する。この手法を用いたシステムでは、文字の記述は顔を動かして行うため、ユーザは文字表記に手を使う必要がなくなる。この手法により、ユーザが手を使えない状態であっても文字認識を行うことができる。また、本稿では小型のロボットを使用し装置の小型化をはかった。顔を動かして文字を描画する場合は、動作範囲が制限されるため文字の描画領域が小さくなりやすく、手を使って描画する場合に比べて軌跡の認識が難しい。そこで本稿では、2台のカメラで文字を描画する人物を捉え、人物の顔の3次元空間座標を算出する。得られた座標を集積したものを描画文字の軌跡とし、その軌跡をもとに文字認識を行う。この認識手順について詳しく述べる。また、カメラを固定した状態では描画点が画面外にはみ出す恐れがあるので、カメラを描画点に追従させることによって、その問題を解決した。本稿では、この解決手法についても述べる。

2. システム構成

2.1 概要

本研究のシステムについて述べる。Fig.1はシステムの概略図である。本システムは2台の小型移動ロボットとホストコンピュータで構成される。ホストコンピュータとロボットは無線LANで接続されており、ホストコンピュータから遠隔操作できる。ホストコンピュータからロボットに動作命令を送り、ロボットはその命令を受信して応答する。

ロボットにはマイコンが搭載されており、ホストコンピュータからの命令をもとにモータの駆動を制御する。また、上部にはカメラが搭載されており、ホストコンピュータに映像情報を送信する。カメラはホストコンピュータからの指令にしたがって、モータの駆動により上下左右に旋回する。ホストコンピュータはロボットから受信したカメラ映像をもとに画像処理を行い、その結果にもとづいてロボットに無線LAN経由で次の動作命令を送信する。

* 専攻科 制御工学専攻2年

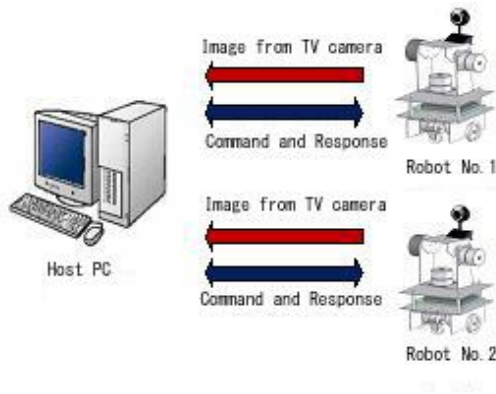


Fig.1 System configuration

2. 2 実験環境

動作環境について述べる。

ロボットは上部にカメラを搭載し、モータの駆動によりカメラの旋回が制御できる。カメラの旋回だけでなく、他方向への移動も可能であるが、本稿では顔への追従のみであるので、移動機能は使用しない。

また、ロボットは平坦な机の上に置き、カメラと筆記者の視線がほぼ同じ高さになるように固定する。筆記者については、文字の描画を行うのは一人で、複数人がカメラの視界領域内部に入らないものとする。筆記者の背景は白い壁面であり、壁面には画像ノイズ対策として、一様な照明が当たっているものとする。

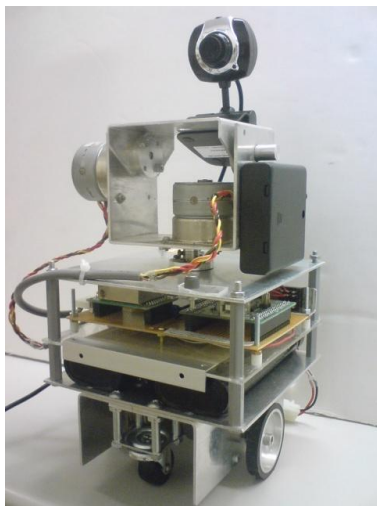


Fig. 2 Small mobile robot with TV camera

2. 3 ロボット本体

2. 3. 1 ロボット概要

Fig.2 に本稿で使用するロボットの外観を示す。本体上部にはカメラが搭載されており、映像を取得する。カメラの下にある 2 台のステッピングモータを駆動させることにより、カメラを

上下左右に任意の角度で旋回させることが可能である。

本体中段には制御回路と 2 本の電池が搭載されている。制御回路はマイコン、電源回路、Wiport 無線モジュール、アンテナ、モータドライバで構成される。これらによって命令の受信と応答、モータ制御を行う。2 本の電池はそれぞれ制御回路とモータへ電力を供給している。

2. 3. 2 ロボット内部構成

Fig.3 はロボットの内部構成である。

ホストコンピュータからの動作命令は WiPort 無線モジュールを通してマイコンへと送られる。マイコンは命令の受諾をホストコンピュータに返送した後、内蔵のタイマカウンタからパルスを発振する。そのパルスはモータドライバ回路に入り、ステッピングモータの駆動を行う。TVカメラはマイコンからは独立しており、独自にカメラの映像をホストコンピュータへ送る。

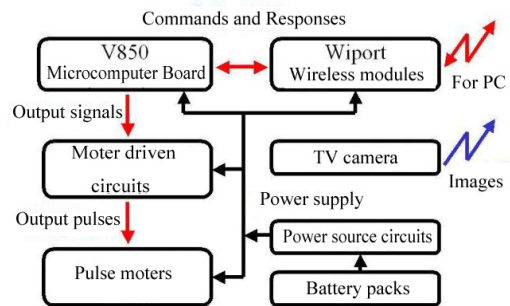


Fig. 3 Internal circuits of small mobile robot

実装した回路を Fig.4 に示す。以下に各要素の詳細を述べる。



Fig.4 Control unit for small mobile robot

(1) V 8 5 0 マイコンボード

Fig.5 は NEC エレクトロニクス社が開発した 32 ビット RISC プロセッサ V 8 5 0 を搭載したマイコンボードである。256 キロバイトのフラッシュ ROM と 24 キロバイトの SRAM、そしてタイマや A/D 変換、SCI など多くの周辺機能を内蔵している。CPU クロック周波数は 20 MHz である。このボードは USB インターフェイスを備えており、USB ケーブルで開発用ホストコンピュータと接続し、制御プログラムの開発が可能である。

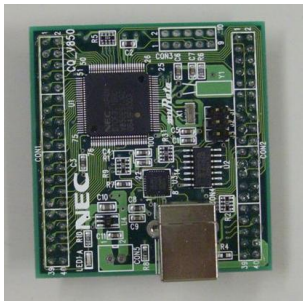


Fig.5 V850 microcomputer



Fig.6 Wiport wireless LAN module



Fig.7 BWC-35H01 TV camera

(2) WiPort 無線モジュール

Fig.6 は WiPort 無線モジュールは無線 LAN 規格である IEEE802.11b/g に対応した無線モジュールである。大きさは 34mm × 33mm × 10mm と小型である。シリアル機器を 2ch 同時にワイヤレス通信可能であり、通信速度は 300 kbps ~ 921 kbps である。また、Ethernet 機器の無線 LAN の無線化も可能である。本稿では、V850 マイコンボードと接続し、ホストコンピュータとロボット間の通信の無線化を行っている。

(3) モータドライバ回路

モータドライバ回路はマイコンからのパルスと方向入力に対応したスイッチングを行い、ステッピングモータの回転角度、回転方向の制御を行う。

(4) 電源回路

電源の不安定な電圧を一定の電圧に安定させ、供給する。5V と 3.3V の 2 系統を実装している。

5V のシリーズレギュレータは V850 マイコンとモータドライバ回路に電力を供給している。3.3V のスイッチングレギュレータは WiPort 無線モジュールに電力を供給している。

(5) カメラ

カメラは BUFFALO 社製の Web カメラ BWC-35H01 (Fig.7) を採用した。このカメラは大きさ 47mm × 97mm × 55mm、重量 100g と小型軽量であり、ロボットへの搭載に適している。有効画素数は 30 万画素で最大解像度は 640 × 480 である。

(6) 電池

電池はタミヤ社製の充電式ニカド電池 RACINGPACK1600SP を採用した。内部抵抗が少なく大電流を流すことが可能で、瞬発力をもつ。公称電圧は 7.2V で 1600mAh の容量を持つ。ロボットには回路用とモータ駆動用の 2 本を搭載している。

2. 4 ロボットシステム

Fig.8 に本研究で用いたロボットシステムを示す。

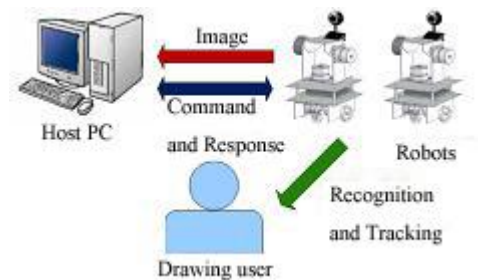


Fig.8 System contracture for the recognition of traversal characters

カメラと PC は USB 接続されており、PC 上でカメラから送られてきた画像に処理を行い、ロボットへ動作命令を無線 LAN 経由で送信する。各ロボットから送信されたカメラ画像から筆者の顔を検出する。検出した顔の画像上の座標を読み取り、顔の中心（鼻の部分）が画面内部の中心位置に配置されるようにロボットに駆動命令を与える。ロボットは PC から送られた命令にしたがってカメラを旋回させ、筆者の顔に対して追従しながら、顔の映像を撮影し PC に送信する。

3. 文字認識

3. 1 前提条件

本稿では描画される文字の条件として、下記のように前提条件を設けた。

- (1) 描画される文字は平仮名、数字であること。
- (2) 顔の移動で描画できる文字であること。
- (3) 軌道が重なって潰れない文字であること。

文字は顔を動かして描画するので、文字の描画は手で描画する場合と比較して格段に困難である。そのため描画される文字は顔の移動で描画できるような描画が容易な文字を選択した。また、空間内で文字を描画するため、筆者には描画した軌跡を確認することができない。したがって、「あ」のような複雑な文字を描いた場合、文字が潰れることになるので文字の描画が非常に困難になる。提案している手法の原理的な可能性を示すためにも、文字の描画を確実にを行うことを優先し、上記の前提条件を設けた。

3. 2 顔の検出

文字認識は筆者のカメラ画像から得た顔の軌道をもとに行う。そのため、カメラ画像から顔を検出する必要がある。

顔検出は顔の陰影をもとに行う。まず顔検出の前処理として、

画像のグレイスケール化とヒストグラムの均一化を行う。次にこれらの処理を施した画像を探索し顔を検出する。検出された顔に対して、サイズと中心を求め、その場所に、円を描画する (Fig.9)。本稿では描画した円の中心を描画点とし、この描画点の軌跡をもとに文字認識を行う。



Fig.9 Face detection of the drawing user

3. 3 ステレオビジョンによる座標算出

文字認識はカメラ画像から得た筆記者の顔の軌跡をもとに行う。そのため、カメラ画像から人物の顔を検出し、3次元座標を算出する必要がある。そこで、本システムではステレオビジョンを用いて対象物の奥行き情報の取得を行う。Fig.10は2台のカメラと描画点を真上から見た図である。モータの駆動角度を θ_{m1} 、 θ_{m2} 、カメラ画像中の描画点の位置から求めた光軸とカメラと描画点を結ぶ直線のなす角を θ_{r1} 、 θ_{r2} とすると、描画点とカメラの距離 d は式(1)により求められる。

$$d = \frac{w}{\tan(\theta_{m1} + \theta_{r1}) + \tan(\theta_{m2} + \theta_{r2})} \quad (1)$$

上記の方法で求めた距離 d をもとに、描画点の上下の座標を算出する。本システムでは座標原点を2台のカメラ間の中点としているため、横軸 x 、縦軸 y は式(2)、式(3)のようになる。

$$x = \frac{w}{2} - a$$

$$= \frac{w}{2} - d \times \tan(\theta_{m1} + \theta_{r1}) \quad (2)$$

$$y = d \times \tan(\theta_{m3} + \theta_{r3}) \quad (3)$$

なお、 θ_{m3} 、 θ_{r3} は縦方向の角度であり、Fig.11のように表わされる。算出した座標を平面に投影したものを、描画点の軌跡とし、文字認識に使用する。

3. 4 文字の認識

前述の方法で算出した座標を集積し、文字認識を行う。文字認識は、正規化、軌跡の近似化、辞書比較の手順で行う。

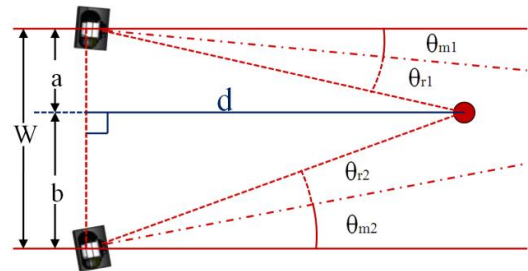


Fig.10 Measurement of depth

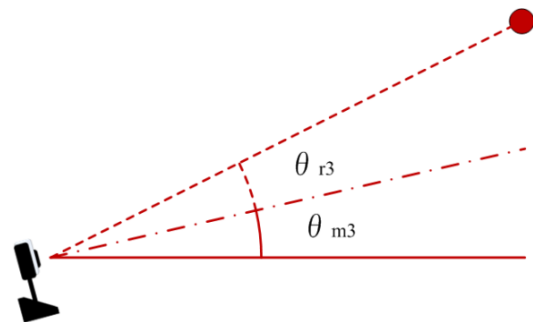


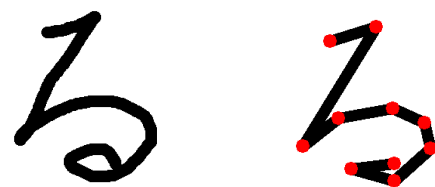
Fig.11 Measurement of height

3. 4. 1 正規化

算出した文字列は、空間座標系で表現されたものである。そのため、軌跡情報を格納した辞書とは座標系が異なり、そのままでは比較して文字を認識することができない。そこで、空間座標系の軌跡情報を辞書の座標系に正規化する必要がある。描画された文字の軌跡はミリメートルで表現されるのに対し、辞書の値はピクセルで表現される。また、空間座標系では原点との位置関係によっては負の値をとるが、辞書では左上の画素を原点とし、横軸は右向きに、縦軸は下向きに正をとる。そこで正規化は、座標の正負を調節し、値が辞書と同程度の枠内に入るよう縮小を行う。

3. 4. 2 軌跡の近似化

正規化により軌跡のサイズは辞書と一致したが、空間上に文字を描画しているため、軌跡は乱れている場合が多い。そこで区分的直線近似法を用いて軌跡を近似化し、軌跡の乱れの補正を行う。区分的直線近似を行うことで、Fig.12のように軌跡が直線近似される。そのため辞書は軌跡が近似化されることが前提として作成されている。



(a) Normal character

(b) Approximated character

Fig.12 Piecewise line approximation

3. 4. 3 辞書比較

正規化と近似化により補正された軌跡と、辞書に格納された各文字の軌跡を比較し、近似度が高い文字を候補として出力する。文字認識は、手書き文字認識エンジンの Tomoe を用いている。Tomoe とは Tegaki Online MOji-ninshiki Engine の略で、オープンソース環境において手書き文字入力を実現するためのソフトウェア、またそれを開発するプロジェクトの名称である。文字認識の基準は以下の 5 点であり、これらの基準から近似度を算出し、近似度が高い文字を候補として出力する。

- (1) 始点の位置
- (2) 終点の位置
- (3) 特徴点の数
- (4) 点間の距離
- (5) 点間の角度

条件を満たした候補が存在しない場合は「該当なし」として出力する。

4. 実験結果

4. 1 実験手順

実験の手順について述べる。実験は、Fig.13 のように文字の筆記者が 2 台のロボットの正面に座って行う。2 台のロボットを同一の机の上に置き、カメラを正面に向ける。このとき、カメラが筆記者の視線とほぼ同じ高さに来るようにする。ロボットの電源を入れ、文字認識プログラムを起動する。起動後、カメラ画像が別ウィンドウで画面に表示されるので、顔を動かして文字を描画する。文字の描画終了後、PC によって認識された描画文字が軌跡表示領域に描画され、文字の候補がテキストボックスに表示される。

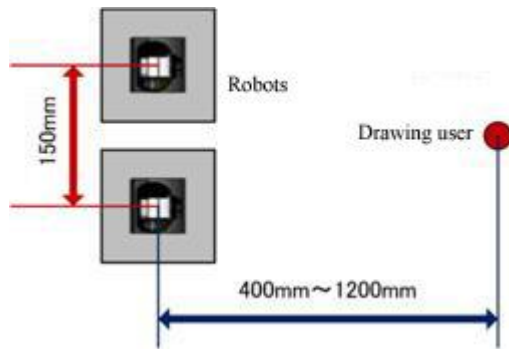


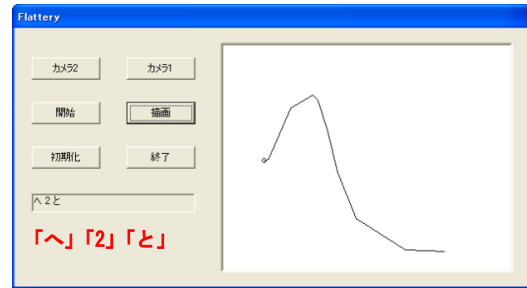
Fig.13 Experimental environment

4. 2 実験結果と考察

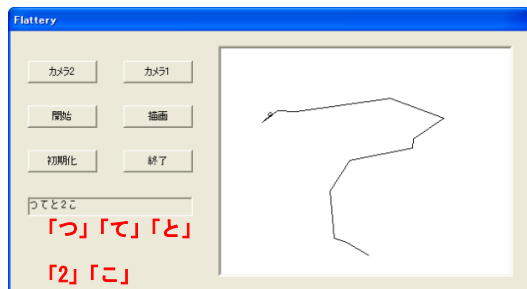
提案するシステムの有効性を確認するため、前節で述べた手順で実験を行った。Fig. 14 にその結果を示す。

実験を行った結果、空間上に描かれた文字を認識することができた。軌跡の直線近似化により、描画された文字が乱れていても文字認識が可能であった。このことから、提案するシステムの原理的有効性が確認できた。文字を正しく認識した場合は筆記者が意図した文字が候補として出力され、誤認識した場合は意図した文字が出力されない、あるいは候補が出力されず、「該当なし」として出力された。

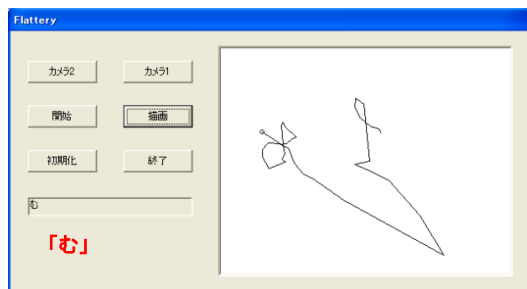
文字を正しく認識したときの事例は、2 種類のパターンに分



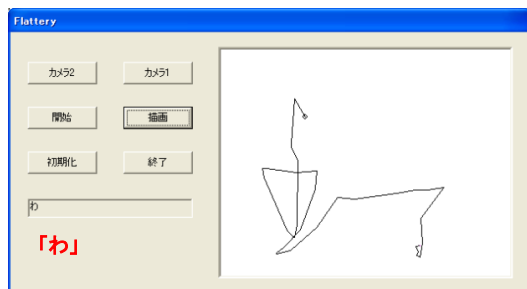
(a) 「へ」



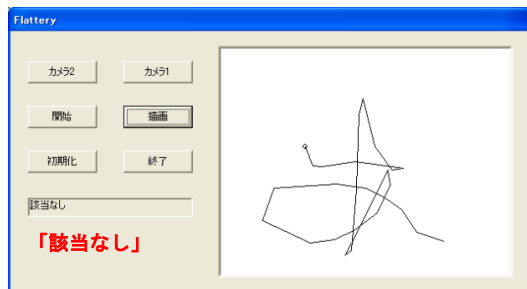
(b) 「て」



(c) 「む」



(d) 「わ」



(e) Invalid character

Fig.14 Experimental results

類される。まず、Fig.14(c), (d)のように候補が1種類のみ出力される場合であり、もうひとつはFig.14 (a), (b)のように候補が複数出力される場合である。前者の場合は、比較的複雑な文字を描画した時に多く見られ、後者は比較的簡易な形の文字を描画した時に見られる。これは簡易な文字が、複雑な文字と比較して特徴が少なく、異なる文字同士で特徴が似通っているためだと考えられる。平仮名の「て」と「と」は画数も異なり、人間では誤って認識することはないが、始点の位置、終点の位置、特徴点の数といった観点から見れば、このふたつの文字は特徴が似ており、同時に候補として出力されることが多い。このような特徴の相似による候補の誤出力は、特徴の少ない文字において顕著に見られ、簡易な文字を描画した場合には、平仮名の「つ」や数字の「2」などが候補として出力されやすいことが実験により把握できた。このような問題は、認識する文字を、候補からの選択方式にすることで回避可能である。現状では近似度が一定値以上の文字を候補として出力しているので、筆記者が意図した正しい文字を候補から選択できるようにすれば、より効果的な文字認識が可能となる。

誤認識の場合も2種類のパターンに分類される。ひとつは軌跡では問題はないが、候補として意図した文字が出力されない場合であり、もうひとつは空間に描画した文字の軌跡自体が大きく崩れてしまい文字認識ができない場合である。両パターンとも比較的複雑な形の文字を描画した場合に多く見られるが、後者の場合は簡易な文字を描画したときにも起こることがある。

描画した文字軌跡には問題がなく、候補に意図した文字が出力されない場合については、辞書の値や閾値を調整することにより解決できる。取得した軌跡の奥行き情報を利用してアフィン変換を行い文字の歪みを補整すれば、文字認識の精度をより向上させることが可能である。描画した文字の軌跡が大きく崩れていなければ、文字認識に使用する値の調整や軌跡の補整により問題は解決できる。

5. 結論

本稿では、ステレオビジョンによる空間内に描かれた文字の認識について述べた。新たな文字認識システムと認識手法を提案し、実際にシステムの構築を行った。また多くの文字認識実験を行い、平仮名・数字を正しく認識し、システムの原理的有効性を確認した。現状では改善の余地はあるが、現段階のシステムでも十分に空間中に描いた文字の認識は可能であることが証明できた。

現段階での改善点としては前章で述べたソフトウェアの改善による認識率の向上のほかに、ハードウェアの改善による認識率の向上があげられる。カメラをより高性能なものに変更すれば、対象の位置情報をより正確に取得でき、認識率の向上につながる。またさらに人物の目を捉え、文字認識を行う「目書き文字」を認識するシステムへの発展が期待できる。また現状では文字の描画自体は手を使わずに顔の動作のみで行うことができるが、システムの起動や終了は手を使って操作を行う必要がある。それらの操作も手を使わずに行うことができる手法を考案すれば、より多くのユーザが利用できるシステムの構築が可能となる。さらに、現状は平仮名・数字の1文字ずつの認識に留まっているので、他の文字、片仮名や漢字などへの対応、複数文字の認識が可能になるよう辞書システムの改善が要求

される。

参考文献

- [1]K. Stoy : Using cellular automata and gradients to control self-reconfiguration, Robotics and autonomous systems, Vol. 54, (2006)pp.135-141.
- [2]A. D. Mali : On the Behavior-Based Architectures of Autonomous Agency IEEE Trans. SYSTEMS, MAN, AND CYBERNETICS PART C, Vol. 32, No. 3, (2002) pp.231-242.
- [3]冬野, 光永, 野副, 山内 : 2台の移動ロボットによる箱押し制御, SICE 中国 2004, (2004)pp.610-611.
- [4]渡邊, 山内 : 運動物体を捕獲するための移動ロボットの追従手法, 第 40 回計測自動制御学会学術講演会, (2001)306A-5.
- [5]山室, 山内 : 広い室内を掃引する移動ロボットの動作計画, 第 38 回計測自動制御学会学術講演会, (1999)pp753-754.
- [6]P. Favaro, S. Soatto : A Geometric Approach to Shape from Defocus, IEEE Trans. Pattern Analysis and Machine Intelligence, Vol. 27, No. 3, (2005) pp.406-417.
- [7]渡辺, 加治, 鶴田, 堀, 山内 : 移動ロボット群による協調的箱押し制御, 第 14 回計測自動制御学会中国支部学術講演会論文集, (2005)pp30-31.
- [8]鶴田, 豊永, 山内 : ステレオビジョンによる移動ロボットの等距離追従制御, 第 15 回計測自動制御学会中国支部学術講演会論文集, (2006)pp278-279.
- [9]豊永, 渡辺, 山内 : 機能モジュールと移動ロボットによる箱押し制御, 第 19 回日本ロボット学会学術講演会, (2006)2G23.

(2010 年 10 月 14 日 受理)