視覚障害者のための指差し文字検出システムの開発*

淵 田 正 隆** 中 村 明 生***

Development of a Fingertip Character Extraction System for the Visually Handicapped

Masataka FUCHIDA and Akio NAKAMURA

This paper proposes a system to assist the visually handicapped by extracting text characters on the surface of an object. The visually handicapped user cannot know exactly what the object is and cannot read the text written on the surface. The user touches a point on the surface with his or her index finger; the system extracts characters around the fingertip. To implement these functions, we propose fingertip detection and character extraction methodologies. For fingertip detection, the user wears a ring, in which three acrylic infrared-reflective beads are embedded, on the index finger. The system detects the beads and locates their positions from an infrared image, then, calculates the fingertip position based on bead arrangement. For character extraction, a region is generated according to the position of the fingertip. The characters are extracted from the region based on color histogram and shape. Then, the extracted characters are combined as text. Finally, the text is recognized by commercially available OCR software and is read aloud to the user by reading software. We have experimentally verified that these methodologies achieved character extraction accuracy of at least 70.0% except the case of narrow character extraction. The experiments demonstrated the basic validity of our proposed method.

Key words: visually handicapped support, fingertip detection, character extraction, image processing

1. 結 言

我々は日常生活において多くの情報を文字から取得している。しかし、視覚障害者は文字を通しての情報取得が大幅に制限される。身近な例で言えば、商品パッケージが読めないため菓子の味が分からずに困るという話を聞いたことがある。それを裏付けるように、全盲の視覚障害者 359 名にアンケートを取ったところ、59.1%が商品の購入時に商品情報の提供を望んでいるという結果もある¹⁾。

視覚障害者と文字と言えば、すぐに思いつくのは点字である。自動点訳技術や商品ラベルに点字を付与する技術²⁾³⁾などが開発され、一般商品にも点字が付与されている例も増えている。しかし、視覚障害者は点字を読めると健常者は思いがちであるが、視覚障害者の点字の習得率はわずか 12.7%程度⁴⁾である。これは、先天的に障害を持ち、幼いころから点字を学んだ人に比べ、事故や病気で後天的に視力を失った中途視覚障害者には点字の習得はより困難であることが関係している。

そこで、商品や書籍上の文字を画像処理によって認識し、読み上げるシステムについて考える。しかし、我々の周囲には文字があふれており、認識すべき文字の場所を指定しない場合、適切な処理ができない。例えば、コンビニエンスストアで商品を探す際に、商品棚を撮影して処理するだけでは商品名の連続読み上げとなり、使用者が混乱する恐れがある。そのため、何らかの手段で文字認識領域を限定する必要がある。

これまでにこの問題に取り組んだ研究として、指先にカメラを装着し文字をなぞって認識するシステム⁵⁾や認識する文字の

周辺にマーカを設置するシステム⁶⁾、紙面や看板が矩形であることを利用し矩形上の文字だけを認識するシステム⁷⁾が提案されている。しかし、指先にカメラを装着すると他の動作の妨げになる可能性がある。マーカを用いる場合はあらかじめ必要な文字が表示される場所にマーカを設置する必要がある。また、物体形状で判別する場合は物体の形状に依存する。

以上を踏まえ、本研究では物体表面を指先でなぞり、その指先を検出し、指先の文字を抽出することで読み上げる範囲を限定する。これにより、物体表面のどこにどのような文字情報があるか把握可能である。

本研究では、視覚障害者向け文字抽出・読み上げシステムの実現を目的として、指先検出手法および指先指定矩形領域からの文字抽出手法を提案する。

2 章で提案手法に必要となる各機能と対象となる文字について考究を加える。3 章で指先検出、4 章で文字抽出手法について説明する。5 章で実験を通して提案手法の有効性を検証する。6 章で結論を述べる。

2. 指先指定文字抽出システムの概要

図 1 に、想定する文字抽出・読み上げシステムの全体像を示す。ウェアラブル型システムとして、視覚障害者が身に付けて利用する。店舗や自宅で商品や書籍を手にとって表面を指差し、指先文字列を抽出して読み上げることを想定する。本論文においてはこのシステムの一部として、指先検出および文字抽出について取り扱う。

本章では、文字の入力装置、指先検出手法、抽出する文字の定義、文字抽出手法について考究を加える。その結果から提案したシステムの流れを示す。

2.1 文字入力装置

これまでに、文字抽出・読み上げシステムのための入力装置

* 原稿受付 平成 28 年 5 月 11 日

掲載決定日 平成 28 年 8 月 17 日

** 学生会員 東京電機大学大学院 (東京都足立区北千住旭町 5)

*** 正 会 員 東京電機大学 (東京都足立区北千住旭町 5)

Fig. 1 Conceptual image of the final system

としてスキャナ⁹⁾やデジタルカメラ^{10) 11)}が提案されている。

文字認識の際のセンサとしてはスキャナが一般的であるが、書類の場合とは異なり、商品パッケージや分厚い書籍の場合などに対しては不適合である。また、持ち運びの面でも使用場所が限定される。それに対し、カメラをセンサとして用いることを考える。2016年現在、スマートフォンのアプリケーションとして比較的容易に文字を認識して情報を取得・整理するシステム¹²⁾も提供されている。デジタルカメラは、照明条件の変化やカメラの傾き、文字との距離などの撮影条件が不定となり、スキャナを使用したシステムに比べると文字の認識精度は低下する¹³⁾。しかし、持ち運びが容易であり、撮影対象の大きさや形状に関係なく撮影が可能である。このことからカメラは、使用者の身体に装着するウェアラブル型のシステムに適したセンサである。

2.2 指先検出手法

指先を検出手法としては、指の色と背景差分法を利用したシステム^{14) 15)}や、爪の色に基づいて指先(爪)位置を推定するシステム¹⁶⁾、手袋をマーカとして装着するシステム¹⁷⁾、指先の爪にマーカを接着し、マーカの動きにより指先を推定するシステム¹⁸⁾が提案されている。

しかし、指の色を利用した場合、照明の変化に伴う色変化による検出精度の低下、背景に段ボールのような指の色と似た物体がある場合の誤検出などの問題がある。また、前節の通り、本システムはデジタルカメラを入力装置とするが、カメラの微妙な揺れにより、背景差分法を用いた指先検出は困難であることが予想できる。すなわち、指に何も装着せず、画像処理のみで指先検出することは難しい。それに対し、指先や手に検出しやすい何らかのマーカを装着すれば、他の動作の妨げになる恐れもあるが、指先を精度良く検出できるメリットもある。

そこで、本研究ではマーカを指先ではなく、人差し指根元に装着することで指の動きの阻害を最小限とし、指先を検出することを考える。マーカは指輪型であり、デザインを工夫すればアクセサリとして違和感も低減することができる。指輪には赤外線を反射しやすいビーズを取り付け、赤外線をカメラ側から照射し、その反射光を基に指先を計算・検出する。

2.3 抽出文字の定義

図1に示すとおり、物体を片手で持ち、もう一方の手の人差し指で物体を指差しするように複数の晴眼者に依頼したところ、平均的に顔から約350mmの距離に物体表面がきた。晴眼者と視覚障害者の手の長さに違いはないため、自然な状態で把持した場合の位置に差異はほとんどないと考えられる。すなわち、カメラからおおよそ350mm先にある標準的な商品パッケージ表面、

書籍表紙の文字を抽出することを想定する。本論文で対象とする抽出する文字の大きさとフォント、色、文字の方向、文字列の長さを以下の通り列挙する。

- 大きさは、文字抽出の精度から解像度 640×480 ピクセルのカメラで撮影した際に、40point 以上、199point 以下の大きさの文字を対象とする。
- フォントに関しては、飾り文字のような判別の困難なものは対象外とする。また、同じ文字列内では大きさおよび太さはほぼ一様とする。
- 対象物品は、カメラに対して正立を基準とする。奥行き方向の変化に関しては、基礎実験により、パン方向に左右 20deg、チルト方向に上下 20deg までの傾きは許容することとした。
- 色は、同一の単語内で単色とする。
- 方向は、横書きとする。
- 長さは、2文字以上とする。

2.4 文字抽出手法

文字抽出では、画像中から文字の位置を推定し、ノイズを除外しながら文字を抽出する。文字抽出の手法としては、色情報を利用した手法^{19)~21)}やエッジ情報を利用した手法^{9) 10) 11)}が提案されている。

色情報を利用した手法では、前提条件として文字の位置を推定するために、同一単語中の文字は単色とする。文字の色は Fussy C-means Algorithm(FCM)¹⁹⁾や K-means 法²⁰⁾、色分布²¹⁾より識別する手法が提案されている。しかし、FCM や K-means 法は使用されている色の数が既知でないと推定が困難になるため、実用的ではない。また、色の頻度を利用した手法は頻度の多い色を抽出するが、多数のノイズも抽出する。

エッジを利用した手法は、文字領域のエッジが多いことや高周波成分が多くなることを利用して文字の位置を推定する。エッジや高周波成分の評価にはエッジの位置¹⁰⁾や Digital Cosine Transform¹¹⁾が用いられている。しかし、これらの手法はノイズにも過敏に反応する。

そこで、ノイズ除去手法として、4 方面特徴⁹⁾や、縦横比¹⁹⁾、画素ピッチ¹⁹⁾、文字間の重心位置のずれ²⁰⁾、黒画素の比率²⁰⁾、文字候補を囲む矩形内の画素の分布²¹⁾などを用いたものも提案されている。

以上を踏まえた上で、著者らは色情報を利用した指先の文字抽出・読み上げシステムを提案した²²⁾。先行研究の手法は、本論文以上に文字の判別に閾値処理を多用しており、環境の変化に脆弱であった。また、文字の傾き補正手法もノイズに脆弱であった。精度・汎用性において改善を加えて本論文の手法を新たに提案する。

2.5 システムの流れ

提案する指先検出・文字抽出システムの流れを以下に記す。

赤外線を照射し、マーカの反射光を手がかりに指先位置を検出する。

1. 検出した指先位置に小領域を設定する。
2. 小領域内の文字の色を推定し、文字を検出する。
3. Hough 変換を用いて文字列の傾きを補正する。
4. 文字の太さを用いてノイズを除去する。
5. 文字の中から指先に近い文字のみを抽出する。

指先検出および文字抽出を行った後、市販の Optical Character Reader(OCR)を用いて認識し、フリーの読み上げソフトで読み上げる。

3. 指先検出手法

2.2 節で説明した指先検出手法について詳述する。カメラユニット及びマーカー、指先検出手法、誤検出の低減手法について説明する。

3.1 カメラユニット及びマーカー

カメラユニットは図 2(a)に示す通り、カメラ、赤外線照射装置からなる。通常画像および赤外線反射画像の撮影のため、2 台のカメラを平行に設置し、一方はそのまま、他方にはレンズ部分に波長が 840nm 以上の光のみを透過する赤外線透過フィルタを装着して対応する。赤外線照射装置は 63 個の赤外線 LED(波長 850nm)を 9×7 にマトリクス状に並べ、その前方に拡散板を設置したものである。カメラから手先の物体までの距離として想定する 350mm 先(2.3 節)にて 180×300mm の範囲に赤外線を照射できる。

マーカーは図 2(b)に示す通り、赤外線をよく反射するアクリル製のビーズを二等辺三角形の形で 3 つ取り付け付けたアルミニウム製の指輪型とする。二等辺三角形の底辺が指根元になるように指輪を装着する。

3.2 指先検出

前節で説明したカメラユニット及びマーカーを用いて、マーカー上のビーズの位置関係から指先を検出する(図 3(a))。マーカーを指先に装着した状態で撮影すると、赤外線画像が得られる(図 3 (b))。

まず、赤外線画像から輝度値が V 以上、縦横の幅が S ピクセル以内の領域を抽出する。次に、抽出した領域から面積が大きい順に 3 つ選択する。このとき、選択した領域の重心間の距離が L ピクセルより大きい場合は選び直す。そして、3 つの領域のうち、他の 2 つから離れたものを指先側のビーズ A、他の 2 つを指根元側のビーズ B、C とし、B と C の中点を D とする。A と D から指の傾きと長さを求める(図 4(a))。

AD 間の距離 M_L ピクセルは実測するとおよそ 15mm であり、人差し指の長さは日本人の手の寸法データ²³⁾よりおよそ 69.5mm なので、D から指先までの距離は、 $M_L \times n$ ピクセルとした。すなわち、図 4 では黄円が指先位置となる。ここで、計算した指の長さが実際の指より短いと、指先がノイズとして撮影される恐れがあるため、余裕を持たせるために $n=6$ とした。

赤外線画像と通常画像は撮影しているカメラが異なるため、座標にずれが生じる。そこで、赤外線画像から検出した指先位置より右に N_m ピクセル移動させた位置を通常画像の指先位置とした。ここでは、基礎実験より $N_m=5.8 \times M_L$ とした。

3.3 指先誤検出の低減

指先誤検出に対応するため、指先を複数回検出し、信頼できる指先位置を計算する。

指先検出を N_b 回行い、指先副候補を求める。指先副候補毎に、それを中心とした半径 L_F ピクセルの円形小領域を設定、小領域内に他の指先副候補が N_c 個以上あるとき、その候補を指先正候補と呼ぶ(図 4 (b))。指先正候補が N_d 個以上存在するとき、それらの位置の平均を計算し、指先位置とする(図 5 の指先点)。ここでは、基礎実験より $N_b=5$, $L_F=20$, $N_c=3$, $N_d=3$ とした。指先が検出できた場合、ピープ音を鳴らし、ユーザに検出成功を伝える。

4. 指先指定文字抽出手法

本章では 2.5 節で述べた処理手順に従い、指先で指定した文字

抽出手法を説明する。

4.1 文字色推定

4.1.1 色ヒストグラムの基づく処理

画像を RGB 表色系で表現すると、照明条件の変化に影響を受けやすいため、等色差性に優れた $L^*a^*b^*$ 表色系に変換する。 $L^*a^*b^*$ 表色系の L は輝度を表し、 a^* と b^* は色相と彩度を表す。

$W_F \times H_F$ ピクセルの小領域(図 6)を、その領域の下辺の中点が指先位置と重なるように設定する。この領域(以降、指先上領域)のヒストグラムを生成する(図 7)。ここでは、基礎実験より $W_F=80$, $H_F=80$ とした。

Fig. 4 Calculation of fingertip position

Fig. 5 Detected fingertip

Fig. 6 Region set around the detected fingertip

Fig. 7 Segmented histograms of fingertip color in L*a*b* color space

Fig. 8 Voxel segmentation in L*a*b* color space

指先上領域の色の頻度は、ヒストグラム上では山になる。しかし、ノイズや画素値のばらつきにより、微細な山が出現する。そこで、式(1)に従い平滑化を2回行い、微細な山を除外する。 H_i は注目している画素値の頻度を示し、 num は要素の個数であり $num=5$ とする。ただし、 $i=0, 1, 254, 255$ のときは隣接する要素数が減るため、 $i=0, 255$ のとき $num=3$ 、 $i=1, 254$ のとき $num=4$ とした。

$$H_i = \frac{H_{i-2} + H_{i-1} + H_i + H_{i+1} + H_{i+2}}{num} \quad (i = 0-255) \dots \dots (1)$$

ヒストグラムより各画素値の頻度を調べ、注目する画素値の前後の画素値より頻度が多いものをヒストグラムの山とする。色の境は、2つの山の間にあり、頻度が1以上、左側の山の頻度の半分以下の、最も頻度が低い画素値とする(図7)。

指先上領域の色の頻度を調査し、文字候補色を決定するため、色の境を利用してボクセルを生成する(図8)。このボクセルに対して、指先の $W_{F2} \times H_{F2}$ の領域の画素を投票する。ボクセルの各マスの投票数を調べ、全投票数の $T_v\%$ 以上を占めるマスに注目する。そのマスの中心の座標を文字候補色とする。ここでは、基礎実験より $W_{F2}=80$ 、 $H_{F2}=80$ 、 $T_v=3$ とした。

指先上領域から設定した文字候補色に近い画素を抽出する。CIE2000色差式²⁴⁾を用いて、各画素と文字候補色との色差を計算し、色差が T_c 以下の画素を抽出する(図9(a))。

しかし、白や黒など彩度が低い画素の画素値は、ばらつきが大きくなることがある。そこで、文字候補色の彩度が T_{S1} 未満の場合は各画素との明度差が T_B 以内かつ各画素の彩度が T_{S2} 以下の画素を抽出する(図9(b))。ここでは、基礎実験より $T_{S1}=20$ 、 $T_c=20$ 、 $T_B=10$ 、 $T_{S2}=20$ とした。文字候補色を抽出した画像を文字候補画像とする。

4.1.2 文字形状に基づく処理

このとき、複数枚の文字候補画像が生成されることがある(図9)。そこで、文字候補画像の中から、指先の近くに文字を多く含んでいる画像を文字画像として選択する。

Fig. 9 Character candidates

Fig. 10 Character judgment

Fig. 11 Thinning

ノイズが含まれた状態で文字の個数を数えると、ノイズを多く含んだ画像が選択される。そのため、細かい文字を除去した評価用画像を生成する。この画像は評価にのみ用い、他の処理には使用しない。

さて、入力される文字列の文字の太さの大小は予めわかっていない。そこで、細い文字を対象としてメディアンフィルタを適用した評価用画像と太い文字を対象としてオープニング処理を適用した評価用画像の2つの評価用画像を生成する。これらの画像に対して、大きさが $N_w \times N_h$ ピクセル以下、密度が D_H 以上、 D_L 以下のものをノイズとして除去する。また、指先座標から文字重心までの距離が L_c ピクセル以上のものを除去する。

文字列内の文字の大きさには差が少ない。文字の大きさで区別するために、各文字の幅と高さを調べ、幅-高さ空間中にプロットする。プロットした点を中心に直径 P_S の円を描き、重なった円の個数を数える。この個数の最大値を文字の個数 W_N とする(図10)。ここでは、基礎実験より $N_w=7$ 、 $N_h=5$ 、 $D_H=0.8$ 、 $D_L=0.2$ 、 $L_c=80$ 、 $P_S=10$ とした。

しかし、ノイズを除去しきれないとき、文字の個数 W_N は増大し、文字を含まず大量のノイズを含んだ画像が選択されてしまう。この問題を解決するために、文字が持つ分岐点の個数に着目する。まず、各画像を細線化する。細線化後の文字の画素は、分岐点と端点、通過点の3種類に分類できる。ノイズが多く含まれているときや色の選択に失敗しているとき、画像内の分岐点数 J_F が増加する(図11)。

以上の処理をすべての文字候補画像に対して行い、以下の値が最も高い画像を文字画像として選択する(図9(b))。

$$\frac{W_N}{J_P} \dots\dots\dots (2)$$

ところが、選択した画像には、明らかに文字でないノイズも含まれる。そこで、各文字候補を囲む矩形を設定し、下記の条件にあてはまる文字候補のみを抽出する。矩形の幅を R_W 、矩形の高さを R_H 、文字候補の画素数を C_P 、文字輪郭の長さを C_L とする。この条件は非常に緩いものであり、明らかなノイズのみを排除する。

$$\frac{C_P}{R_W \times R_H} \geq 0.1 \dots\dots\dots (3)$$

$$\frac{C_L}{C_P} \geq 0.01 \dots\dots\dots (4)$$

$$\frac{R_H}{R_W} \geq 0.1 \dots\dots\dots (5)$$

$$R_W \geq 2 \dots\dots\dots (6)$$

$$R_H \geq 2 \dots\dots\dots (7)$$

その後、矩形の幅、高さを N_H 倍して統合してマスクを作成、マスク処理を行い、基礎的な文字候補を得る。ここでは、 $N_H = 7/3$ とした。

4.2 文字列傾き補正

Hough 変換を用いて文字列の傾きを補正する。想定する文字列は横書きのため、傾きの範囲は ± 45 度とする。

ノイズの影響と処理時間を抑えるために、文字画像を文字重心のみの画像に変換する。指先との距離が P_L 以内であり、面積が $N_{w2} \times N_{h2}$ ピクセル以上の文字の重心を抽出し、重心を中心に半径 r の円を描く(図 12)。ここでは $r=1$ 、 P_L =画像の幅/4、 $N_{w2}=5$ 、 $N_{h2}=5$ とする。変換した画像に対して Hough 変換を行い傾き角度を算出、回転変換により傾きを補正する(図 13)。

4.3 文字候補抽出およびノイズ除去

4.3.1 文字の太さに基づく処理

図 13 において、白領域を文字候補とする。しかし、まだノイズが多数含まれている。本論文では、同じ文字列内の文字の太さは、ほぼ一様と仮定している。そこで、文字の太さに基づいて、ノイズを除去する。

まず文字候補領域の輪郭上の各画素から反対の輪郭に達するまで領域内部に向けて走査を行うことで文字の太さを測定する(図 14)。このとき、走査方向は全周を 45 度毎・ 8 方向に分割したうちで、輪郭に直交する方向にもっとも近い方向とする。すべての輪郭上の画素から同様の測定を行い、最頻値をその文字候補の「基準太さ F_s 」とみなす。

同じ文字列内の文字の太さはほぼ一様と仮定しているため、すべての太さがこの基準太さと等しいことが理想であるが、そのような場合は少なく、ばらつきが存在する。このばらつきの許容範囲を「基準太さ $F_s \pm N_t \cdot F_s$ 」ピクセルとする。なお、ここでは $N_t=0.8$ とした。ただし、細い文字の場合は許容範囲下限が過小になる場合があるため、その場合は下限を 5 ピクセルに設定する。

文字候補領域における太さ測定の結果のうち、上記許容範

囲のものを残し、残りは排除することでノイズの大部分を排除できる。しかし、走査線上にない画素も排除され、隙間がある文字領域が生成される。文字には曲線部が多いため、この傾向が顕著である(図 15)。そこで、 N_D 回の膨張処理を行い、隙間を埋めてマスクを作成する(図 16)。文字候補画像(図 13)にマスク処理を行い、文字抽出を行う。ここでは $N_D=7$ とした。

Fig. 12 Centers of inclined characters (B and W reversed)

Fig. 13 Compensation of the inclined of characters

Fig. 14 Thickness measurement

Fig. 15 Creating masks based on thickness

Fig. 16 Masks dilation

Fig. 17 Histogram of characters pointed at a fingertip

Fig. 18 Extracted characters

Fig. 19 Noises and characters

Fig. 20 Mask and Extracted characters near fingertip

4.3.2 文字の高さに基づく処理

4.3.1 項の処理の後、複数の文字列が抽出されている可能性があるが、指先近くの文字列のみを選択して以下の処理を行う。

傾き補正の結果、文字はほぼ水平に並んでいることを想定し、文字列の高さと縦幅を調べ、その文字列の上下のノイズを除去する。その後、指先の $W_h \times H_h$ ピクセルの小領域内において横方向の画素ヒストグラムを生成する(図 17)。山が文字部、谷が行間に該当しているので、文字列の高さ方向位置と縦幅が分かる。位置のばらつきを吸収するために、縦幅を上下に N_y ピクセル拡大する。また、これにより小領域よりも大きな文字や「y」や「l」のような他の文字とは上端や下端がずれる文字にも対応可能である。また、文字列中の文字の横幅は、行の縦幅よりも極端に大きくなることはないため、横幅が縦幅の N_B 倍以上の文字はノイズとして除去する(図 18)。ここで、基礎実験より $W_h=80$, $H_h=80$, $N_y=2$, $N_B=2$ とした。

4.3.3 文字の隙間に基づく文字列合成処理

また、ノイズや別の文字列が、指定した文字列と同じ高さにある場合(図 19)、誤抽出される可能性も考えられる。これに対処するため、同一文字列内の文字は近接していると仮定し、文字間距離が小さい文字を同一の文字列に属するとみなす。

まず、それぞれの文字に外接する矩形を設定する。矩形の間の距離を文字間隙間とし、その平均を求める。この文字間隙間平均が縦幅の N_{H2} 倍より大きい場合には文字間隙間平均

分だけすべての矩形の幅を拡大する。これにより、平均より大きい隙間を持つ異なる文字列に属する文字が分離される。ここで、 $N_{H2}=0.5$ とする。

しかし、この処理を行うと、逆に一つの文字列しかない場合は平均が小さくなり過ぎるため、本来同一の文字列の属する文字まで分離する場合がある。そのため、文字間隙間平均が縦幅の N_{H2} 倍より小さい場合には縦幅の N_{H2} 倍分だけ矩形を拡大し、文字の過分離を防止する。

その後、矩形を統合してマスクを作成、マスク処理により指先に近い文字列のみ抽出可能となる(図 20)。

5. 指先検出・文字抽出実験

本章では、指先の検出精度および文字の抽出精度を実験により検証する。また、指先指定文字の抽出精度を検証し、本提案手法の有用性を評価する。なお、かすれやつぶれがないことを目視にて確認できた場合を文字抽出成功とした。

撮影には Logicool のウェブカメラ 2-MP Portable Webcam C905m を使用する。カメラの解像度は 640×480 ピクセル、フレームレートは 30fps に設定する。フォーカスおよびホワイトバランスはカメラの機能による自動調整である。

5.1 指先検出実験

照明条件と距離の変化に対する指先検出精度を確認し、提案する指先検出手法の実用性を調査する。

図 21 (a) に実験装置の配置を示す。照明条件は、JIS 照度基準を参考に、やや暗い居間と、大型店の一般陳列部を想定し、150lux, 1000lux とした。カメラと指先の距離は、2.3 節で述べたように 350mm を基準とし、AIST 人体寸法データベースの上腕と前腕の長さ²⁵⁾を参考に身体近くに腕を折り曲げた状態として 250mm, 腕を伸ばした状態として 450mm に関しても調査した。赤外線画像用のカメラとカラー画像用のカメラは、図 2(a)のように並べて配置してある。そのため、指先の検出範囲(図 21(b)-(d)の格子)は右にずれる。実験条件をまとめると、以下の 6 パターンになる。

1. 照度 150lux, 距離 250mm
2. 照度 150lux, 距離 350mm
3. 照度 150lux, 距離 450mm
4. 照度 1000lux, 距離 250mm
5. 照度 1000lux, 距離 350mm
6. 照度 1000lux, 距離 450mm

赤外線照射範囲を一辺が 30mm の正方形で分割し(図 21(b), (c), (d)), 指先を各マスの中央に置き、検出された指先の位置がマス内であれば成功とする。各マスにつき、指差し動作を 30 回繰り返す。一般的な想定環境である 150lux, 350mm の実験結果を表 1 に示す。表 1 の各マスは、図 21(b)の格子枠と対応している。

実験結果より、150lux, 350mm の場合、平均 99.0%、最低でも 83.3%以上の精度で指先を検出することに成功した。また、照明条件と距離を他の値に設定したときの検出精度の平均を表 2 に示す。どの条件においても同等の結果を確認できた。

Table 1 Accuracy of fingertip detection (150lux, 350mm)

	A	B	C	D	E	F	G	H	I	J
1	100%	100%	100%	100%	96.7%	100%	100%	90%	100%	100%
2	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
3	100%	100%	100%	100%	100%	100%	100%	100%	96.7%	100%
4	100%	100%	100%	100%	100%	100%	100%	100%	100%	90%
5	96.7%	100%	100%	100%	100%	100%	100%	100%	100%	83.3%
6	100%	100%	100%	100%	100%	100%	100%	100%	93.3%	93.3%

Fig. 21 Experimental setup: Fingertip detection

Fig. 22 Experimental setup: Character extraction

Fig. 23 Target objects

Table 2 Average accuracy in different conditions

Illumination	Depth	Ave. accuracy
150lux	250mm	99.6%
150lux	350mm	99.0%
150lux	450mm	97.7%
1000lux	250mm	98.9%
1000lux	350mm	97.6%
1000lux	450mm	97.0%

指先検出の失敗原因は、マーカのずれや指の折り曲げ、赤外線照射範囲である。指先の位置はマーカの位置から検出しているが、マーカのずれや指の折り曲げにより、検出した指先が実際の指先から外れることがある。また、指先が照射中心から遠ざかったり、画像の端に移動したりすると、赤外線が届きにくくなるため、検出精度が低下する。

以上の結果から、赤外線が十分に照射されている範囲であれば、想定する照明条件や距離で指先を検出できることが分かった。

5.2 文字抽出実験

照明条件と距離の変化に対する文字抽出精度を確認し、提案する文字抽出手法の実用性を調査する。文字の抽出精度のみを検証するため、指先位置はあらかじめ与えた状態で文字を抽出する。

図 22 (a) に実験装置の配置、図 22 (b) に複数の文字を含んだ背景を示す。照明条件は、JIS 照度基準を参考に、やや暗い居間、大型店の店内全体、一般的な商店の重要陳列部、大型店の一般陳列部を想定し、それぞれ 150lux, 500lux, 750lux, 1000lux とした。カメラと指先の距離は、5.1 節と同様である。ただし、500lux 以上の場合は、距離は想定使用距離である 350mm に固定して試行した。

抽出する文字は 10 種類の書籍・食品パッケージの表面に書かれた文字である。その最下部に書かれた文字列の下に指先が検出できたものとして、指定した文字列が抽出できた回数を計数する(図 23)。この処理を 30 回繰り返す。このときの実験結果を表 3 に示す。表 3 の a,b,c...は物品を表し、図 23 の各物品と一致する。

表 3 の 1, 2, 3...は実験環境を表し、以下のように対応している。

1. 照度 150lux, 距離 250mm
2. 照度 150lux, 距離 350mm
3. 照度 150lux, 距離 450mm
4. 照度 500lux, 距離 350mm
5. 照度 750lux, 距離 350mm
6. 照度 1000lux, 距離 350mm

実験結果より、実験環境2に関しては、すべての実験対象について最低でも83.3%の精度で文字を抽出できた。照度を高くした実験環境4-6では、物品g:「CHOCOCHIPS」と物品:j「CHOICE」を除くと、最低でも86.7%の精度で文字の抽出できた。物品gとjは、Hの横棒やEの横棒などが他の文字線と比べて、極端に細くなるフォントが存在する。照明の影響で生じる細い文字線の欠けが精度低下の要因と考える(図24)。

Table 3 Accuracy of fingertip detection

	a	b	c	d	e	f	g	h	i	j
1	100%	93.3%	100%	100%	100%	96.7%	100%	83.3%	100%	26.7%
2	96.7%	100%	100%	100%	100%	96.7%	96.7%	96.7%	100%	83.3%
3	86.7%	96.7%	70.0%	100%	100%	90.0%	3.3%	93.3%	86.7%	0.0%
4	100%	100%	100%	86.7%	90.0%	100%	50.0%	96.7%	100%	70.0%
5	100%	96.7%	100%	100%	90.0%	100%	56.7%	96.7%	96.7%	53.3%
6	100%	100%	100%	100%	96.7%	93.3%	0.0%	100%	100%	3.3%

距離250mmにした実験環境1の場合、物品jを除き、最低83.3%の精度で文字の抽出ができた。物品jに関しては、近づけた場合も遠ざけた場合も細い文字線の欠けが生じる(図25(a))。

また、物品が遠ざかる実験環境3では、その他の物品上の文字線も相対的に細くなり、物品c、g、jに関しても精度が低下している(図25(b))。ただ、この環境は腕を伸ばした状態(距離450mm)を想定したものであり、実際の使用時はそこまで距離を離すことはなく、実用上は問題ないと考える。

5.3 システムの総合精度検証実験

システムの実用性を確認するために、指先検出処理と文字抽出処理を統合したときの文字抽出精度を検証する。

カメラと実験対象の位置関係や背景は 5.2 節と同様であり、照明条件、カメラと指先の距離は、それぞれ 150lux、350mm とする。図 26 に実験装置の配置を示す。指差し動作を 30 回繰り返し、その間に指定した文字列が抽出できた回数を表 4 に示す。なお、CPU Intel Core i5-4300U 2.5GHz、メモリ 8.0GB、GPU なしの計算機環境で指先検出、文字抽出までに平均 14 秒程度を要する。ただし、使用計算機の機能向上、処理の並列化などにより将来的には撮影 30fps と同期した処理時間を目指す。

Fig. 24 Truncated characters

Fig. 25 Truncated characters (250mm, 450mm)

Fig. 26 Experimental setup

Table 4 Accuracy of character extraction following fingertip detection

Target	Accuracy
a	70.0%
b	76.7%
c	100%
d	100%
e	100%
f	93.3%
g	30.0%
h	96.7%
i	100%
j	26.7%

Fig. 27 Success examples

Fig. 28 Failure examples due to fingertip detection

Fig. 29 Failure examples due to character color estimation

Fig. 30 Failure examples due to character composition

Fig. 31 Failure examples due to slope correction

物品 g, j を除き, 最低でも 70.0% の精度で指先の文字を抽出できた(図 27). 指先の文字の抽出の失敗の原因は, 5.1 節で挙げたマーカや指先のずれ(図 28)と, 文字色の推定失敗(図 29), 文字列の合成失敗(図 30), 傾き補正の失敗(図 31)である. マーカや指先がずれると, 本来の指先位置から離れた場所が指定され, ノイズ, 文字欠けの要因となる. 文字色の推定失敗に関しては, 誤った色で二値化して, かつ白領域が文字に類似した形状をしている場合に起因する. 文字列の合成失敗は, 文字の大きさのばらつきが大きい場合に発現した. 傾き補正の失敗は文字数が少ない場合に見られた.

実験結果を鑑みるに, 100% の精度が理想ではある. しかし複数フレームで一度でも正しく読み上げることができれば, 視覚障害者支援の目的には合致すると考える. 実環境での使用を考えた場合, 現在のフレームにて指先検出・文字抽出に不具合が生じたとしても, 少し手を動かして, 次フレームにて検出・抽出できれば実用的ではないかと考える.

6. 結 言

本研究では, 視覚障害者向け文字抽出・読み上げシステムの実現を目的として, 指先検出手法および指先指定矩形領域からの文字抽出手法の提案, および基礎システムを構築した.

- (1) 指先検出に関しては, 赤外線と指輪型マーカを用いた手法を採用し, 照明の変化に頑健な指先検出手法を提案した.
- (2) 文字抽出に関しては, 文字色推定, 文字列傾き補正, 文字候補抽出およびノイズ除去を行った.
文字色推定に関しては, 色ヒストグラムの基づく処理と文字形状に基づく処理を組み合わせで実装した.
文字列の傾きに関しては, 文字重心のみに対して Hough 変換を行うことで対応した.
文字候補抽出およびノイズ除去に関しては, 文字の太さに基づく処理, 文字の高さに基づく処理, 文字の隙間に基づく文字列合成処理を組み合わせることで, 指先近くの文字列を抽出した.

店舗や自宅で商品や書籍を手にとって表面を指差し, 指先文字列を抽出して読み上げることが想定している. すべての商品を網羅することは不可能であり, 言語, 文字色, 文字サイズ, フォント等でバリエーション考えて対象を選択して実験した. その結果, 文字の細い部分の抽出に弱点があったが, それを除けばほぼ実用的な精度となり, 基本的な有効性・実用性は示すことができたと考ええる.

ただ, 本論文で提案した手法はパラメータ処理が多く, 設定する数値が多い. これはパラメータ数を少なくしてシビア

な数値を設定するのではなく, 逆に数を増やしてかつ数値をあらかじめ設定することで多くの環境への対応, 頑健性の向上を図ったものである. 基礎実験の段階で多くの環境に適応できるように数値を設定しており, 実験でも実用性を検証しているが, やはり適応性をより考慮すべきである. したがって, 使用環境に応じたパラメータの自動設定に取り組むべきである.

謝 辞

本研究の一部は, JSPS 科研費 15K01482 及び東京電機大学総合研究所研究 Q12J-06 の援助を受けた.

参 考 文 献

- 1) 財団法人共用品推進機構: 2010 年度(平成 22 年度)視覚障害者不便利調査成果報告, (2011), 47.
- 2) N. Li and L. Wang: A Kind of Braille Paper Automatic Marking System, Proceedings of International Conference on Mechatronic Science Electric Engineering and Computer, (2011), 664.
- 3) 兵藤 安昭, 横平 貫志, 早川 哲史, 村上 裕, 池田 尚志: 誤り箇所指摘機能をもたせた点字翻訳編集システム IBUK-TEN, 電子情報通信学会論文誌. D-I, 情報・システム, I-情報処理 J84-D-I, 7(2001) 1102.
- 4) 厚生労働省社会・援護局障害保健福祉部企画課: 平成 18 年身体障害児・者実態調査結果, (2008) 24.
- 5) S. Suzuki, K. Kato and K. Yamamoto: A Consideration of Effective Feature Extraction method for Nail Camera System, Proceedings of the 17th Workshop on Frontiers of Computer Vision, (2011) 1.
- 6) T. Morris, P. Blenkhorn, L. Crossey, Q. Ngo, M. Ross, D. Werner, and C. Wong: Clearspeech: A Display Reader for the Visually Handicapped, IEEE Transactions on Neural Systems and Rehabilitation Engineering, 14, 4(2006) 492.
- 7) P. Clark and M. Mirmehdi: Recognising Text in Real Scenes, International Journal on Document Analysis and Recognition, 4, 4(2002) 243.
- 8) 小林 慎平, 松本 哲也, 竹内 義則, 工藤 博章, 大西 昇: 視覚障害者向けの飲食店メニュー情報取得支援システム, 電子情報通信学会技術研究報告 IMQ, 112, 475(2013) 139.
- 9) 株式会社アメディア: よむべスマイル, <http://www.amedia.co.jp/product/catalog/ys.pdf> accessed on 2016. 5. 8.
- 10) 鈴木 悠司, 平岩 裕康, 竹内 義則, 松本 哲也, 工藤 博章, 山村 毅, 大西 昇: 視覚障害者のための環境内の文字情報抽出システム, 電気学会論文誌 C, 電子・情報・システム部門誌, 124, 6(2004) 1280.
- 11) H. Goto and M. Tanaka: Text-tracking Wearable Camera System for the Blind, Proceedings of International Conference on Document Analysis and Recognition, (2009) 141.
- 12) INTSIG: CamScanner for Nakabayashi, <http://app-liv.jp/578918832/> accessed on 2016. 5. 8.
- 13) 黄瀬 浩一, 大町 真一郎, 内田 誠一, 岩村 雅一: デジタルカメラによる文字・文書の認識・理解, 電子情報通信学会論文誌, 89, 9(2006) 836.
- 14) 西川 誠, 森田 哲也, 吉川 大弘, 鶴岡 信治: カラー画像処理を用いた指先による文字列・図表領域の指示法, 電子情報通信学会技術研究報告, パターン認識・メディア理解, パターン認識・メディア理解 98, 528(1999) 25.
- 15) C.C. Lee, C.Y. Shih, and B.S. Jeng: Fingertip-writing Alphanumeric Character Recognition for Vision-based Human Computer Interaction, Broadband, Proceedings of International Conference on Wireless Computing, Communication and Applications, (2010) 533.
- 16) N. Fujishima and K. Hoshino: Fingernail Detection Method from a Hand Images including Palm, Proceedings of International Conference on Machine Vision Applications, (2013) 117.

- 17) 肥川 宏臣, 藤村 絃匡, 佐藤 大輔: ハードウェア向き指文字認識アルゴリズム, 電子情報通信学会論文誌 D, 情報・システム, **J92-D**, 3(2009) 405.
- 18) 澤 光映, 竹川 佳成, 寺田 努, 塚本 昌彦: 演奏ルールを用いたウッドベースのための実時間運指取得システムの設計と実装, 日本ソフトウェア科学会, コンピュータソフトウェア, **27**, 1(2010) 56.
- 19) 芦田 和毅, 永井 弘樹, 岡本 正行, 宮尾 秀俊, 山本 博章: 情景画像からの文字抽出, 電子情報通信学会論文誌 D-II, 情報・システム, II パターン処理, **J88-D-II**, 9(2005) 1817.
- 20) K. Kita and T. Wakahara: Binarization of Color Characters in Scene Images Using K-Means Clustering and Support Vector Recognition, (2010) 3183.
Machines, Proceedings of International Conference on Pattern
- 21) 松田 友輔, 大町 真一郎, 阿曾 弘具: 2 値化とエッジ抽出による情景画像からの高精度文字列検出, 電子情報通信学会論文誌 D, 情報・システム, **J93-D**, 3(2010) 336.
- 22) M. Fuchida, R. Abe and A. Nakamura: Character Extraction from the Region Pointed at with a Fingertip for the Visually Handicapped, Proceedings of the 21st IEEE International Symposium on Robot and Human Interactive Communication, (2012) 265.
- 23) 河内まき子: AIST 日本人の手の寸法データ, <http://riodb.ibase.aist.go.jp/dhbodydb/hand/index.html> accessed on 2016. 5. 8.
- 24) G. Sharma, W. Wu, and E. N. Dalal: The CIEDE2000 Color-Difference Formula: Implementation Notes, Supplementary Test Data, and Mathematical Observations, Color Research and Application, **30**, 1(2005) 21.
- 25) 河内まき子: AIST 人体寸法データベース 1991-92, <https://www.dh.aist.go.jp/database/91-92/> accessed on 2016. 5. 8.