

## Reading-Life Log のプロトタイプ実装

木村 崇志<sup>†</sup> 柿迫 良輔<sup>††</sup> フォン ヤオカイ<sup>†††</sup> 内田 誠一<sup>†††</sup>

岩村 雅一<sup>††††</sup> 大町真一郎<sup>†††††</sup> 黄瀬 浩一<sup>††††</sup>

<sup>†</sup>九州大学大学院システム生命科学府 〒 819-0395 福岡市西区元岡 744

<sup>††</sup>九州大学工学部電気情報工学科 〒 819-0395 福岡市西区元岡 744

<sup>†††</sup>九州大学大学院システム情報科研究院 〒 819-0395 福岡市西区元岡 744

<sup>††††</sup>大阪府立大学大学院工学研究科 〒 599-8531 大阪府堺市中区学園町 1-1

<sup>†††††</sup>東北大学大学院工学研究科 〒 980-8579 宮城県仙台市青葉区荒巻字青葉 6-6-05

E-mail: †kimura@human.ait.kyushu-u.ac.jp, ††{fengyk, uchida}@ait.kyushu-u.ac.jp

あらまし 人間が日々目で見ている文字を自動的に認識、記録できれば、極めて有用である。我々はこれを Reading-Life Log と呼び、人間の趣向や考えの把握、行動履歴の検索等への応用を考えている。実現のためには、視線情報の抽出、視線位置の文字の切り出し、文字認識技術が必要となる。本研究では、視線位追跡デバイス (アイトラッカ) と web カメラを用いて視線情報を抽出し、視線位置の画像の切り出しを行う。そして、切り出した画像を OCR エンジンに入力し認識結果を得る。本稿では、そのプロトタイプの実装の様子と、その性能を評価するために行った屋内、屋外などの様々な環境や条件下での認識実験結果について論ずる。

キーワード Reading-Life Log, アイトラッカ, 情景内文字認識, OCR

### 1. はじめに

小型センサや大容量レコーダーが身近なものになった現在、自分の行動を自動記録する技術、いわゆるライフログが多数研究されている。環境中に固定されたセンサで収集されたデータとは異なり、そこには特定のユーザだけの行動履歴が、そのユーザの移動範囲にすべてにおいて収集されている。このため、収集データを解析することで、そのユーザに関する様々な情報を解析できる。

ライフログに利用されるセンサは多様である。バイタルサインを測定する腕時計タイプのもや RFID タイプのもの、またカメラも利用される。カメラによるライフログは、ユーザが見たものを自動記録することを目的とする。自分を中心としたサーベイランス画像という単純な意味合いだけでなく、そこからユーザの様々な意図が汲み取れるという点も重要である。この意味で、ライフログは First-person vision と関連する。

ところで、アイトラッカの普及に従い、人間の視線情報を利用した様々な研究が進められている。例えば、自動車を運転する際の視線を解析することで、有効な視野角を求めることできる<sup>[1]</sup>。また、見ている箇所の文字を入力できるようにすることで、身体の不自由な人の文字入力補助を可能にするなど様々な分野に応用されている。「目は口ほどにものを言う」とされる。要するに視線情報は、我々の意図や行動、コミュニケーション状態を表現し、知識取得対象を明示する重要な情報源である。

本研究では、視線情報と文字認識を組み合わせることにより、我々が日常的に読んでいる文字情報の自動獲

得を目指す。我々は、これを Reading-Life Log と呼ぶ。様々なライフログが考えられる中、目で見えた文字情報、すなわち文字として読んだ情報の獲得に着目することは意義深いと考えている。これは、我々が日常的に情報の多くを文字を読むことで得ているためである。実際、本や新聞、情景内の看板やサイン、テレビやコンピュータディスプレイ上の文字など、我々の生活は文字に囲まれている。文字体系の全く異なる外国での生活は本当に不便である。

Reading-Life Log には様々な応用も考えられる。図の情景を見た経験があるとし、それをしばらくたった後に再び思い出したいとする。このとき、この情景画像中の“コンピュータビジョン”等を認識しておけば、後にそれらの単語を用いて検索するだけで、この情景画像を呼び出すことができる。すなわち、画像をキーとした検索とは異なり、文字情報を用いた曖昧性が少ない検索が実現できる。その他にも様々な利用法が考えられる。例えば、読んだ文字を解析することで装着者の考えや興味を調べ、ウェアラブルデバイスと組み合わせることでの確にその人に合った情報やサービスなどをリアルタイムに提供できる。また、リアルタイム認識ができれば、ライフログとは異なるインタラクティブな応用も考えられる。例えば、読んだ単語の意味や翻訳を音声で伝えることなど拡張現実の分野での応用が期待できる。

強調しておきたいのは、本研究の主眼が一般的な情景内文字認識とは異なる点である。上記のような応用、例えば画像検索に関する応用については、情景内文字認識研究においても考慮されている。しかし、通常は、画像中

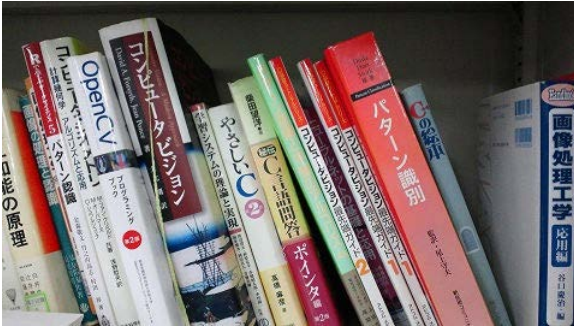


図 1 Reading-Life Log の対象例

のすべての文字を認識しようとするものであり、従って、実際にユーザーの視界に入っていない文字であっても検索されることになる。これに対し、Reading-Life Log では、ユーザが意思を持って読んだものだけを対象とすることを目的している。

Reading-Life Log 実現のために、本研究では次の3つの段階からなるプロトタイプシステムを構成した。はじめに、アイトラッカを用いて視線情報の抽出を行う。次にwebカメラ上で視線周辺の画像を切り出す。そして、最後に画像をOCRエンジンに入力することで見た箇所の文字認識を行う。こうして実現したReading-Life Log に対し、本稿では、様々な条件下で実験を行い、切り出し方法の有効性、認識精度の検証を行う。その結果から、Reading-Life Log を行う上で生じる特有の問題の解明を行う。

## 2. システムの概要

### 2.1 アイトラッカを用いた視線情報の抽出

視線情報の抽出は、ナックイメージテクノロジー社のアイマークレコーダー EMR-9 を用いて行う。これは、図のような帽子型のヘッドユニットで両眼の動きを検出するためのインカメラ、近赤外線 LED 照明と視線情報を表示するためのシーンカメラで構成されている。図に示すように、眼球に近赤外線 LED 照明をこことで出現する反射光（第1プルキニエ像）の位置と瞳孔の中心位置を検出する<sup>[1]</sup>。2つの位置の距離は眼球を動かすことで変化するため、この距離から眼球の回転角が求められる。眼球の大きさには個人差があるため、装着者ごとにキャリブレーションを行うことで眼球の個体差を補償しておく。このようにして、シーンカメラと視線の座標位置により、装着者がシーンカメラ上のどの場所を注視しているかが求まる。

### 2.2 視線周辺画像の切り出し

本研究では、アイトラッカ付帯のシーンカメラの上により高解像度なカメラ（Logicool HD Pro Webcam C910, 1920×1080 画素）を取り付け、そこから視線位置の画像を切り出している。これはシーンカメラの解像度（640×480 画素）の低さを補うためである。この場合、

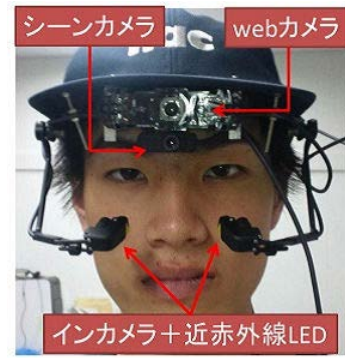


図 2 アイトラッカの構成

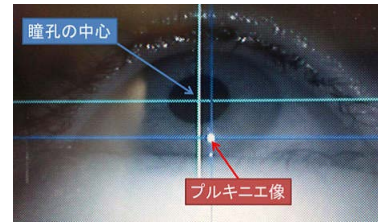


図 3 瞳孔の中心およびプルキニエ像の検出

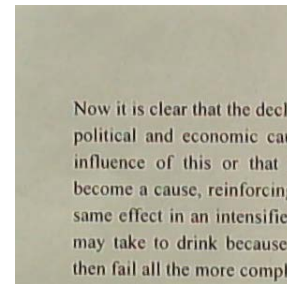


図 4 視線位置を中心とした矩形領域例

視点周辺の画像を切り出すためには、シーンカメラの座標系から高解像度カメラの座標系に変換する必要がある。本研究では、この変換は画角の違いの補正により実現した。

よく知られているように、人間は不随意に衝動性眼球運動（サッケード）を行っている。これは視線位置の大きな移動となって現れる。本システムではこれを「読む」ことを意図した動きではないと考え、移動量に対する単純な閾値処理により除外した。

### 2.3 認識処理

こうして各フレームから視線位置を中心とした300×300画素の矩形領域を切り出す。図に矩形領域の例を示す。ここで領域の大きさを300×300画素とした理由は、リアルタイム性の維持が可能で、かつ装着者が本稿の実験対象を見た時の視野範囲が矩形領域に収まるためである。

本稿では切り出した画像に対して可変閾値処理<sup>[2]</sup>で2値化を行い、市販のOCRエンジンに入力して文字認識を行なっている。よく知られているように、カメラベースの文字認識は非常に困難な課題である。多様な背景や

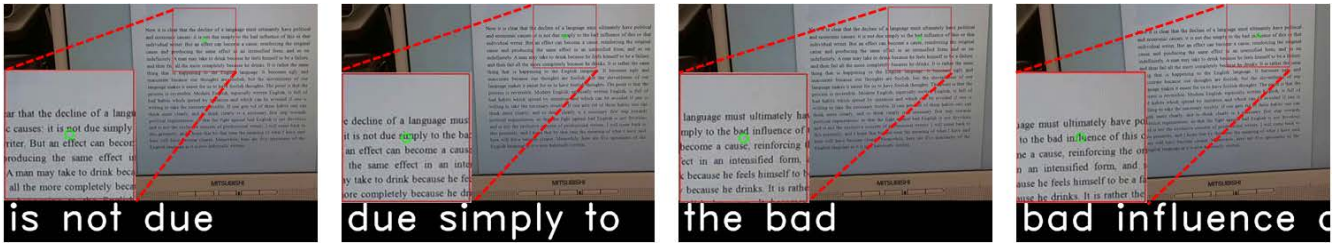


図 5 ディスプレイ画像を対象とした視線移動の様子および認識結果  
(左から右に 100ms 間隔のフレーム系列．以下同)

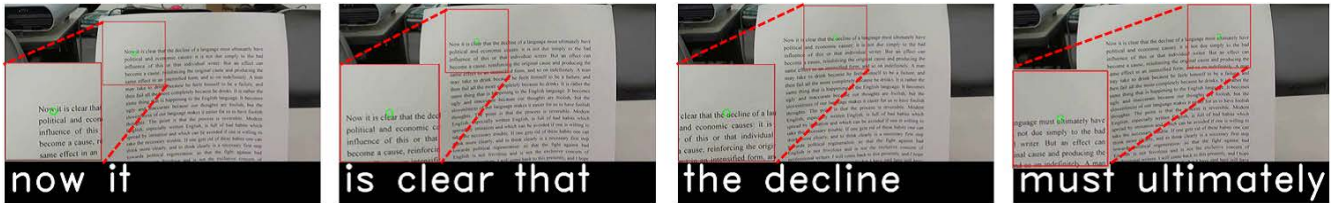


図 6 文書画像を対象とした視線移動の様子および認識結果



図 7 屋外看板画像を対象とした視線移動の様子および認識結果

照明環境下での文字単位への切り出し，および多様な字形への対処が必要となる．

得られた認識結果から視線位置と最も近い単語を抽出し，装着者の見た単語として出力する．また，認識結果を英単語辞書と照合することで認識結果の精度を向上させている．

### 3. 実験

#### 3.1 概要

上記のシステムを用い，様々な環境下でリアルタイム文字認識および文書画像を対象としたオフライン文字認識を行った．第一に，屋内外で実験を行うことで基本的な認識精度を確認する．本実験は本システムの動作確認の意味も含んでいる．

第二に，420 文字，100 単語からなる文書を声に出しながら読み，文書全体の認識率を確認する．この実験から，本稿のシステムが発声速度で移動する視線に対してどの程度文字を認識できるのかを確認する．

第三に，文書を対象としたオフライン文字認識を行い，装着者の見た単語を時系列で確認する．この実験から装着者の見た単語の推移や時系列化で発生する問題を把握する．



図 8 屋外看板画像文字認識の誤認識例

#### 3.2 Reading-Life Log システムの基本動作の確認

環境中において，本システムがどの程度正しく見た位置の文字を認識できるか確認するため，屋内外で実験を行った．まず，屋内での実験結果を示す．図にディスプレイ画像，図に文書画像，を見た場合の視線移動の様子と認識結果を示す．文書画像，ディスプレイ画像のフォントは 16pt の Times New Roman を用いた．装着者との距離は，それぞれ約 50cm となっている．図，図から，いずれの場合も正確に認識できていることがわかる．これより，屋内のような太陽光など環境要因の少ない場所では単純な固定領域 (300×300 画素) 切り出しでも十分な精度が得られることが確認できた．

次に，屋外での実験結果を示す．図に看板画像を見た場合の視線移動の様子と認識結果，図に看板画像の誤認識例を示す．屋外の実験する場合は認識する対象に固有

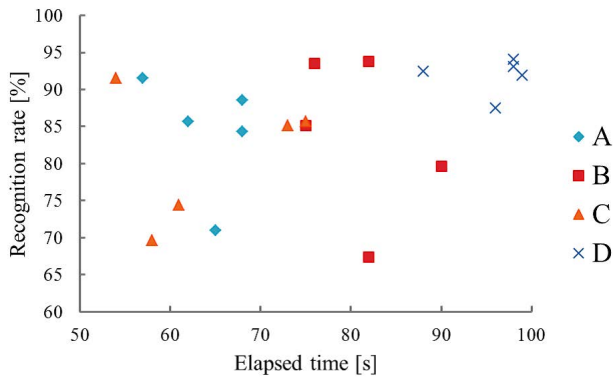


図 9 話速による文字認識結果の違い

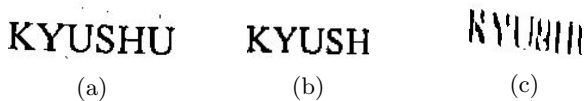


図 10 誤認識例

名詞が多いため、認識処理における英単語辞書による照合は行っていない。この結果から、図は正しく認識できているが、図では建物や太陽光の影響で誤認識している。屋外ではこのような環境要因のため、屋内に比べ誤認識が起こる可能性が高いと考えられる。そのため屋外では、環境中文字検出手法を使用する必要があると考えられる。

### 3.3 文書を対象とした認識率の算出

本システムで文書全体を読んだ場合、どの程度単語を認識できるのかを確認するため、4人の装着者で5回ずつ実験を行い読む速さと認識率の分布を求めた。実験対象は、420文字、100単語からなる文書で、フォント、距離はの実験と同じものを使用し、声に出しながら文書を読むように装着者に指示した。認識率は、切り出した領域の文字認識結果に登録された単語が存在すれば正解とし、正解数を読み始めから読み終わりまでのフレーム数で割ることにより算出した。

図に結果を示す。これより、読むのにかかった時間が70秒以上の場合、すなわち毎秒平均1.4個以下の単語を読んだ場合、認識率は70%程度で比較的安定している。しかし、70秒未満では速くなるにつれて低下する傾向が見られる。すなわち、現在のシステムでは、装着者の読む速さに対して十分な認識率を得ることができなかった。不正解画像を検証した結果、主な原因は図??のような(a)ノイズの発生、(b)切り出しの失敗、(c)単語切り出しの失敗だとわかった。モーションブレイクの原因は手や頭の動きによるものだと考えられる。また、単語の切り出しの失敗は体の動きによるものの他に、視線の動きも原因となっていると考えられる。この実験で用いたフォントは30cmの距離でwebカメラ上で12×12画素程度、アイトラッカのシーンカメラ上で4×5画素程度で

表示される。このため、アイトラッカ上の視線の僅かなぶれでも誤認識に繋がってしまう可能性がある。

### 3.4 認識結果の順序の精度

装着者が見た単語の順序を明確化するため、フレーム間で認識結果を統合し時系列化を行った。時系列化の方法は、認識結果から画像中の視線位置の単語を抽出し、1フレームごとに記録する。図に装着者が見た単語の順序を示す。赤枠が見た単語、青矢印が視線移動を表している。これより、装着者が意図して読んだ順序と異なる結果となったことがわかる。この原因は、装着者の視線は横方向に移動だけでなく縦方向にも絶えず動き、上下の行の単語の抽出や装着者が文書を読み返したためだと考えられる。本稿の方法ではこのような視線の動きを考慮していないため、単語の抽出法に、制約を加える必要がある。

## 4. まとめと今後の課題

本稿では、Reading-Life Logの実現に向けたシステムの構築および様々な環境下での評価実験を行った。その結果、屋内では安定した認識精度を得ることができた。しかし、屋外は環境要因や英単語辞書と照合ができなかったため、十分な認識精度を得られなかった。今後の予定として、装着者や環境要因を取り入れた視線領域切り出し方法を検討する必要があると考えられる。

本研究では、装着者が読んでいる、読んでいないに関わらず見ただけで認識を行なっているため、視線の動きから人間が文字を読んでいるか、読んでいないかの判別を行う必要がある。文献[1, 2]から、人間が文書を読む場合の視線の動きは200ms~250msの注視、装着者から見える文字の大きさにかかわらず一定移動量のサッケードを繰り返していることが確認されている。この特徴を利用して判別を行い、より人間の視線の動きに忠実な認識を目指す。

また発声速度での視線移動した場合、現在の本研究のシステムでどの程度の認識精度なのか確認した。その結果、まだ十分な認識率を得られていないことがわかった。今後、上記の改善法により認識率の向上の余地はあると考えられる。また、実験のサンプル数を増やすと共に、文字単位での認識率も求める予定である。

文書を対象とした装着者の視線の動きを確認した。その結果、本稿の方法では視線の動きに対処できず正確に時系列化できなかった。今後、単語の抽出法に制約を加えることで視線の上下の動きのようなノイズを除き、順序の精度の向上を目指す。

## 謝 辞

本研究の一部は、JST 戦略的創造研究推進事業チーム型研究(CREST)「共生社会に向けた人間調和型情報技術の構築」に依った。

Despite the heavy clumsiness of her lines, the Aorai handled easily in the light breeze, and her captain ran her well in before she hove to just outside the suck of the surf. The atoll of Hikueru lay low on the water, a circle of pounded coral sand a hundred yards wide, twenty miles in circumference, and from three to five feet above high-water mark. On the bottom of the huge and glassy lagoon was much pearl shell, and from the deck of the schooner, across the slender ring of the atoll, the divers could be seen at work today.

図 11 文書画像を対象とした装着者の視線移動の順序

#### 文 献

- [1] 瀬谷安弘, 八木理, 中易秀敏, “リアルタイム視線移動解析による運転作業時の有効視野計測”, 信学技報, MBE2007-90, 2008.
- [2] 内川恵二, 篠森敬三, 視覚 I, 朝倉書店, 2010.
- [3] 酒井幸市, デジタル画像処理入門, CQ 出版社, 2002.
- [4] 佐々木康人, 富永浩之, 林敏浩, 山崎敏範, “文書・絵画閲覧時における視線情報の計測と分析”, 信学技報, ET2005-47, 2008.
- [5] 斎藤大輔, 斎藤恵一, 斎藤正男, 東吉彦, 犬井正男, “眼球運動解析による可読性評価-文字色と背景色の組み合わせによる比較-”, バイオメディカル・ファジィ・システム学会誌, Vol.11, No.1, pp. 23-28, 2009.
- [6] John M Findlay, Iain D Gilchrist, 本田仁視, “アクティブビジョン”, 北大路書房, 2006.