

# 卒業研究論文

題 目

リーディングライフログのための  
視点情報の分析と精度検証

知能メディア処理研究グループ

指導教員 黄瀬 浩一 教授

平成 23 年 ( 2011 年 ) 度 卒 業

( No. 1080107018 ) 川市 仁史

大阪府立大学工学部知能情報工学科

# リーディングライフログのための 視点情報の分析と精度検証

第3グループ 川市 仁史

## 1. はじめに

近年、ライフログが広く社会に認識され、多くの研究が行われている。ライフログとは、人の生活や行動を、音声や映像などのデジタルデータとして記録することである。様々なライフログが提案されているが、人が目で読んだ情報を記録するようなライフログはない。しかし、人は日常的に目から様々な情報を得て生活している。従って、読んだ情報を記録し解析すれば高度なライフログが実現できると考えられる。

そこで、この目的を満たすライフログとして、リーディングライフログを提案する。リーディングライフログを実現するために、文書画像検索の一手法である LLAH[1] と、アイトラッカを用いる。具体的には、文書画像検索により特定された文書画像に、アイトラッカから得られる視点情報を記録する。これにより、文書に対するユーザの視点の動きを分析することができる。本論文では、文書画像検索とアイトラッカにより記録された視点情報の分析と精度を検証する。

## 2. リーディングライフログ

リーディングライフログでは、ユーザの読んだテキストを全て記録する。リーディングライフログを実現するため、文書画像検索と、アイトラッカを用いて、文書画像上にユーザの視点情報を記録する。

### 2.1 文書画像検索

ユーザの読んでいる文書を特定するために、文書画像検索を用いる。[1]の手法では、リアルタイムの文書画像検索が可能である。また、撮影画像とそれに対応する文書画像の位置関係が得られ、文書の撮影範囲を特定することができる。本研究では、アイトラッカから得られる視野カメラ映像を用いて、文書画像検索を行い、対応する文書画像と撮影範囲を特定する。

### 2.2 アイトラッカによる視点情報の検出

視点情報を検出するためにナックイメージテクノロジー社のアイトラッカ EMR-9 を使用する。アイトラッカでは、視点情報がユーザの前方の風景を撮影している視野カメラ映像上に表示される。そこで、文書画像検索により得られる撮影範囲の情報を用いて、文書画像上に視点情報を表示する。

### 2.3 視点情報の分析

アイトラッカから得られる視点情報から、瞬きなどによるノイズを除去する。図 1 にアイトラッカから得られた視点情報を示す。文書画像上に表示した視点情報には、瞬きなどのノイズが大きな視点の移動として現れる。しかし、改行などの視点移動も、同様に大きな視点の移動として現れる。ノイズと改行などの視点移動を区別するため、閾値処理によるノイズの除去を行う。

また、視点情報から注視点を特定する。文書画像上の一定領域に視点が連続してあるとき、その領域を注視していると考えられる。そこで、一定領域内に連続している視点の平均座標を求め、これを注視点とする。注視点は注視の度合いに比例した大きさの円で表示する。得られた注視点とそれ以外の視点情報を記録時間に従い結び、文書画像上に表示する。

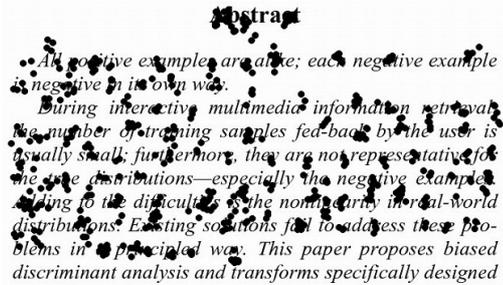


図 1: アイトラッカから得られる視点情報

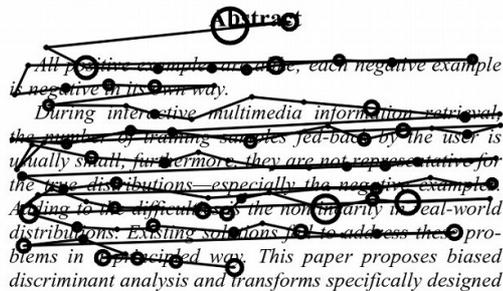


図 2: 視点情報の分析

## 3. 実験

文書画像上に変換した視点情報の精度を検証した。アイトラッカを装着した被験者に、文書を読んでもらい視点情報を記録した。記録された視点情報を分析し、実際に読んだ行の中心と文書画像上に表示された視点情報の誤差を求めた。被験者から読む文書までの距離は 30cm とし、読む文書は 1 枚に限定した。視点情報を書き込む文書画像の大きさは 1700 × 2200 である。

図 2 に視点情報を分析した結果を示す。実験の結果、頭部と文書を固定した際の誤差は平均で 20pixel 程度の誤差があった。これは文書画像上で 1 行に満たない誤差である。手持ち文書に対する視点情報の誤差は 45pixel 程度であり、2 行程度の誤差になった。誤差が小さい時の結果画像より、被験者の注視点は単語中心に多いことが分かった。

上記の誤差の原因として、アイトラッカの精度限界がある。リーディングライフログの目的は、読んだテキストの記録であるため、連続した多くの行を読む場合、1、2 行の誤差は大きな問題にならない。以上より、本手法はリーディングライフログに有効であると考えられる。

今後の課題として、視点情報の精度向上や、長時間データを収集することが挙げられる。また、視点の動きから人が文書を読んでいるか、読んでいないかを判定することが挙げられる。

### 参考文献

- [1] 竹田一貴, 黄瀬浩一, 岩村雅一, “メモリ効率と安定性を向上させた LLAH によるカメラベース大規模長時間文書画像検索”, 画像の認識・理解シンポジウム (MIRU2011), pp. 350-357, 2011.

# 目次

第1章 緒論	1
第2章 関連研究	3
第3章 リーディングライフログ	5
3.1 LLAHによる文書画像検索	5
3.1.1 処理概要	5
3.1.2 特徴点抽出と特徴量計算	6
3.1.3 登録処理と検索処理	7
3.2 アイトラッカによる視点情報の検出	8
3.2.1 角膜反射法	8
3.2.2 キャリブレーション	9
3.3 視点情報の変換	10
3.4 視点情報の表示	11
3.5 視点情報の分析	13
3.5.1 ノイズの除去	13
3.5.2 注視点の判定	14
第4章 実験	17
4.1 実験条件	17
4.2 実験結果	18
4.2.1 視点情報の分析	18
4.2.2 精度検証	20
第5章 結び	23
謝辞	25



# 目 次

2.1	読んでいる部分に対応した挿絵 . . . . .	4
2.2	ディスプレイ一体型アイトラッカ . . . . .	4
3.1	LLAH の処理概要 . . . . .	6
3.2	アイトラッカ . . . . .	8
3.3	角膜反射像と瞳孔の検出 . . . . .	9
3.4	キャリブレーション済みの視点情報 . . . . .	10
3.5	文書画像上に変換された視点情報 . . . . .	11
3.6	視点情報の表示 . . . . .	12
3.7	ノイズの判定 . . . . .	13
3.8	注視点の判定 . . . . .	15
4.1	誤差 . . . . .	17
4.2	ノイズの判別 . . . . .	18
4.3	注視点判定 . . . . .	19
4.4	注視点 . . . . .	21



# 表 目 次

4.1 精度 [pixel] ..... 20



# 第1章 緒論

近年、ライフログが注目され、様々な研究が行われている。ライフログとは、人の生活や行動を、音声や映像などのデジタルデータとして記録することである [1]。記録されたライフログを解析すれば、その人の行動パターンや趣味嗜好を知ることができ、様々なサービスの実現が可能になる。実際に提案されているライフログとして、DAYTUM [2] や FoodLog [3] がある。DAYTUM は、例えば、一週間で歩いた距離や、一ヶ月で読んだ本の冊数など、人の行動と数値を記録するライフログである。行動と行動に関する数値を記録することで、自分がどのような行動をどれほどしたか、詳細なデータでまとめることができる。また、記録された情報から様々な種類のグラフを作成することができ、自分の毎日の習慣を可視化することも可能である。一方、FoodLog は、日々の食事を記録することで、いつ何を食べたかを教えてくれるライフログである。また、食事の写真から食事バランスを測ることができ、健康管理に役立てることができる。このように様々なライフログが提案され、実際にサービスとして提供されている。また、ライフログ撮影用の小型カメラ [4] など、人の行動を記録するための様々なデバイスが開発されている。

ユーザの視点を追跡するデバイスとして、アイトラッカがある。アイトラッカを使うことにより、ユーザのしているものを特定することができ、また、どのように物事を見ているかを解析することができる。このようなアイトラッカの普及により、人の視点情報を利用した様々な研究が進められている [5]。例えば、電子文書を読む際の視点情報を利用した研究として text 2.0 [6] がある。text 2.0 では、ユーザの視点を利用して、電子文書上に挿絵などの情報を提供する。また、ユーザの目の動きで絵を描く Eyewriter [7] という製品もある。このように、人の目の動きを利用して、情報を提供したり、何かを操作したりすることの可能性が注目されている。

人は日常的に、文書や文字から様々な情報を得ている。実際、身の回りには新聞や雑誌など、たくさんの文書や文字があり、人はこれらを読みながら生活している。従って、ユーザが読んだ情報を記録することができれば、より高度なライフログが実現できると考

えられる。そこで、本研究では、視点情報と文書画像検索を組み合わせたリーディングライフログを提案する。このシステムは、ユーザの読んだすべての文書や単語を記録することを目的としている。ユーザの読んだ文書や単語を記録することにより、様々なサービスを実現できると考えられる。例えば、見覚えのある単語を検索する際、データベース全体からではなく、過去に読んだ文書の読んだ箇所から限定的に検索することが可能である。また、どのような文書や単語を多く読んだか分かれば、その人の読む文書の傾向に合わせた情報を提供することができる。しかし、ユーザの読んだすべての文書や単語を記録するためには、読んだ文書と文書中の読んだ部分を特定する必要がある。そこで、本研究では文書画像検索とアイトラッカを組み合わせることで、それぞれを特定する。文書画像検索は、文書をカメラで撮影することで、データベースから対応する文書画像を検索する処理である。

本論文では、リーディングライフログを実現するため、文書画像検索とアイトラッカを組み合わせることの有効性を検証する。文書画像検索とアイトラッカによって記録された視点情報を分析し、実際に読んだ行に対する視点情報の精度を検証する。視点情報の分析として、瞬きなどによるノイズを除去し、視点情報が集中する箇所である注視点を特定した。精度検証の結果、視点情報と実際に読んだ行は、1, 2行程度の誤差であった。

## 第2章 関連研究

ここでは、アイトラッカを用いた研究と、リーディングライフログの関連研究について紹介する。

アイトラッカと電子文書を用いた研究として text 2.0 がある。text 2.0 では、ディスプレイ上の電子文書に対するユーザの視点を追跡し、特定の単語や行に至ったとき、情報を出力するシステムである。例えば、物語を読んでいる場合、図 2.1 のように読んでいる部分に対応した挿絵を表示したり、効果音を出したりすることができる。また、視点情報からユーザの分からない単語を推定し翻訳することや、どこまで読んだか分からなくなった時に、読んだ箇所を知らせることができる。このように、読んでいる電子文書に対する様々な情報を、視点を利用して出力することができる。この手法では、電子文書に対する視点の動きを利用するため、図 2.2 のようなディスプレイ一体型や、据え置き型のアイトラッカを用いている [8] [9]。このようなアイトラッカは、ユーザがディスプレイを見ていることを前提としている。そのため、デジタルデータでない文書を読む場合には、文書に対する視点情報を記録することができない。

リーディングライフログに関連する研究として、木村らの手法 [10] がある。木村らの手法では、ユーザ装着型のアイトラッカと文字認識手法である OCR を用いる。視点情報と文字認識を組み合わせることで、ユーザが見た看板や新聞などの文字を認識し記録する。このように記録することで、後に単語検索によって、検索した単語の写った情景画像を出力することができ、どこでその文字を見たかを知ることができる。この手法は、認識技術である情景内文字認識を用いた研究であり、認識技術を利用したリーディングライフログの実現を目指している。一方、本研究は、検索技術である文書画像検索を用いた研究であり、検索技術を利用したリーディングライフログの実現を目指す。

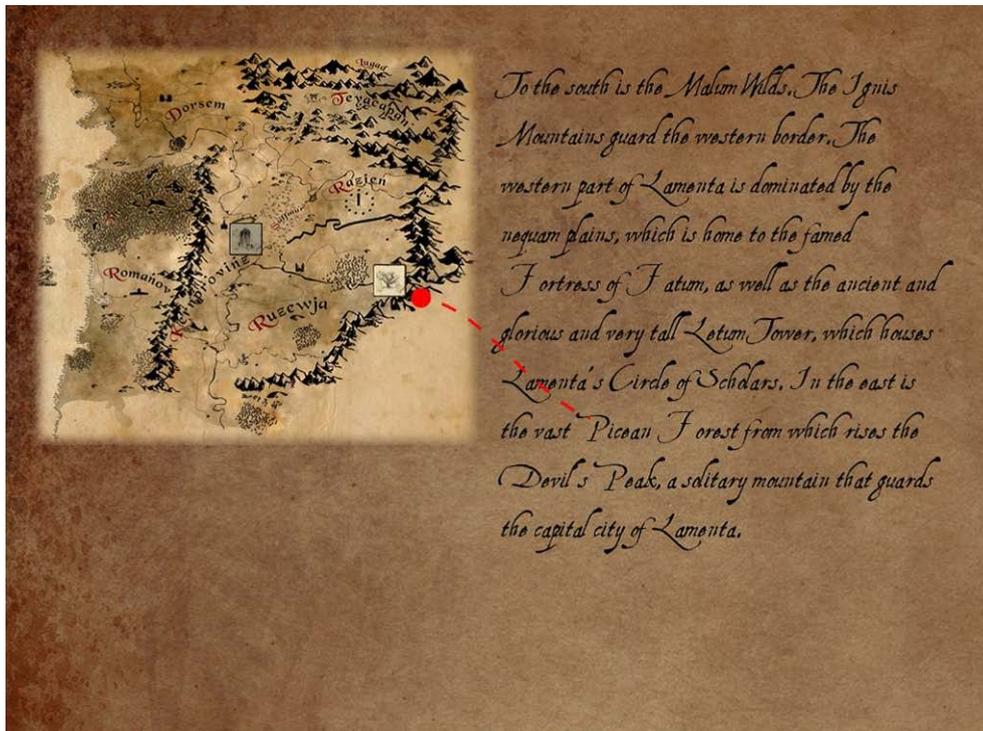


図 2.1: 読んでいる部分に対応した挿絵



図 2.2: ディスプレー体型アイトラッカ

## 第3章 リーディングライフログ

本研究で提案するリーディングライフログは，ユーザの読んだすべての文書や単語を記録することを目的としている．この目的を満たすためには，読んだ文書と読んだ部分を特定する必要がある．そこで，それぞれを特定するために文書画像検索とアイトラッカを用いる．文書画像検索とは，文書をカメラで撮影することで，その文書に対応する文書画像を検索する処理である．提案手法では，文書画像検索の一手法である LLAH(Locally Likely Arrangement Hashing) [11] を用いる．LLAH は，射影歪みや紙面の非線形的な湾曲に対してロバストであることが知られている．これらの特徴は，リーディングライフログにおいて非常に有効である．なぜなら，リーディングライフログでは，手で持っている文書など，射影歪みなどが生じている状況での記録が想定されるからである．一方，アイトラッカは，眼の動きからユーザの視線を追跡する装置である．提案手法では，[6] の手法で用いている据え置き型やディスプレイ体型ではなく，ユーザ装着型のアイトラッカを用いる．これにより，紙媒体の文書に対する，ユーザの視点の動きを追跡することができる．本手法では，文書画像検索によって特定された文書画像上に，アイトラッカから得られる視点情報を記録する．

### 3.1 LLAH による文書画像検索

LLAH を用いた文書画像検索について説明する．

#### 3.1.1 処理概要

図 3.1 に LLAH による文書画像検索の処理概要を示す．文書画像検索では，検索する文書を撮影し，データベースから対応する文書画像を出力する．あらかじめ登録しておく文書画像は，特徴点抽出処理により特徴点の集合に変換される．そして，登録処理により，

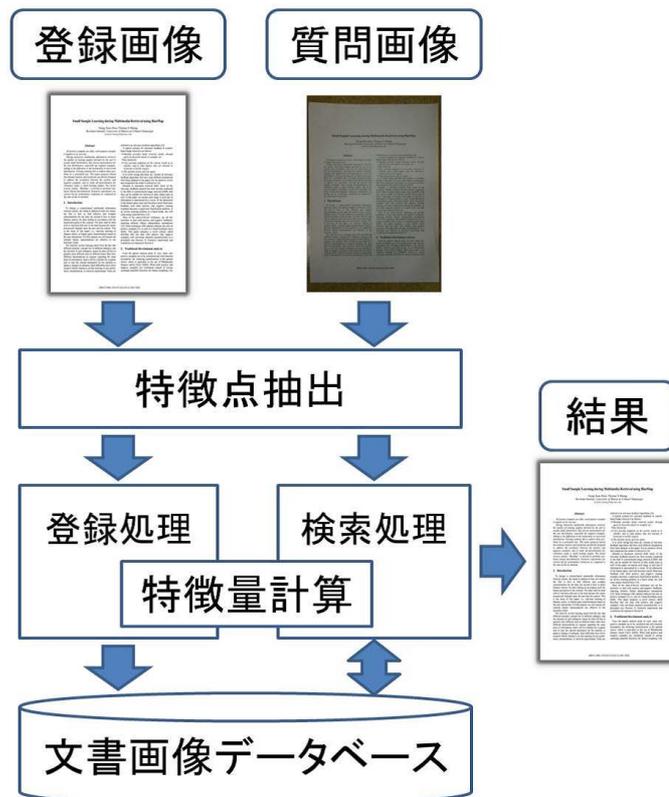


図 3.1: LLAH の処理概要

各特徴点はそれぞれの特徴量に基づき文書画像データベースに登録される。検索時には、質問文書画像から特徴点を抽出した後、検索処理を行う。検索処理では、登録処理と同じ方法で特徴点から特徴量を計算する。そして、得られた特徴量を用いて文書画像データベースにアクセスし、対応する文書画像を決定する。以下に、各処理について示す。

### 3.1.2 特徴点抽出と特徴量計算

LLAHでは、特徴点の配置に基づいて文書画像のマッチングを行う。従って、特徴点を抽出する時には、射影歪みやノイズが生じた場合や、低解像度の場合でも同一の特徴点を抽出する必要がある。そこで、単語領域の重心を特徴点として用いている。特徴量は特徴点を表す値であり、特徴点のマッチングは特徴量に基づいて行われる。LLAHでは、ある

特徴点  $s$  の特徴量に，点  $s$  の近傍 4 点から以下の式で求められる値を用いる．

$$\frac{P(A, C, D)}{P(A, B, C)} \quad (3.1)$$

ここで， $P(A, B, C)$  は頂点  $ABC$  からなる三角形の面積である．また，2つの三角形の面積比である式 (3.1) の値は，局所領域で射影不変量に近似可能なアフィン不変量である．ある特徴点  $s$  の特徴量を計算する際，点  $s$  を含む近傍  $m$  点から 4 点を選ぶ組み合わせを考える．そして， ${}_m C_4$  通りあるすべての組み合わせからアフィン不変量を求め，それらを一列に並べたものが一つの特徴量となる．LLAH では，アフィン不変量を用いることで射影歪みに対するロバスト性を実現してゐる．

また，面積比特徴量を付加することで，識別性能を向上させている．これは，単語領域の面積比に着目した射影不変な特徴量である．まず， $m$  点を抽出した単語領域を時計回りに並び替え，隣り合うものに番号を与える．ペアとなる単語領域の面積比を計算し，ペアの番号を面積比の大きさの順番で並べる．この番号の列をアフィン不変量の列に追加する．こうすることで，最終的に特徴量の次元数は  $({}_m C_4 + m)$  となる．

### 3.1.3 登録処理と検索処理

LLAH では，ハッシュ表を用いて文書画像を登録し検索する．登録処理では，各特徴点から得られる特徴量を基に，各点の文書 ID，点 ID，特徴量をハッシュ表に登録する．検索処理では，検索質問画像の各特徴点から得られた特徴量を基に，ハッシュ表からリストを得る．そして，得られたリストの各項に対して特徴量が一致するかを調べ，一致していたら文書 ID に対応する文書に投票する．最後に，得票数が最大の文書を検索結果として出力する．



図 3.2: アイトラッカ

## 3.2 アイトラッカによる視点情報の検出

図 3.2 にナックイメージテクノロジー社のアイトラッカ EMR-9 を示す。EMR-9 には、ユーザの目の動きを検出するための眼球カメラと、ユーザの前方を撮影するための視野カメラがある。ユーザの視点位置を特定するため、眼球カメラによって記録される目の動きから、視点位置を算出している。算出された視点位置は、視野カメラ映像上に表示され、ユーザがどこを見ているのかが分かる。この際、ユーザの視点情報は視野カメラ映像上の座標数値で表され、デジタルデータとして記録される。以下に、視点情報の検出処理について示す。

### 3.2.1 角膜反射法

図 3.3 に角膜反射像と瞳孔が検出された画像を示す。EMR-9 では、目の動きを検出するために、角膜反射法を用いている [12]。これは、眼に近赤外線を当てた時にできる反射

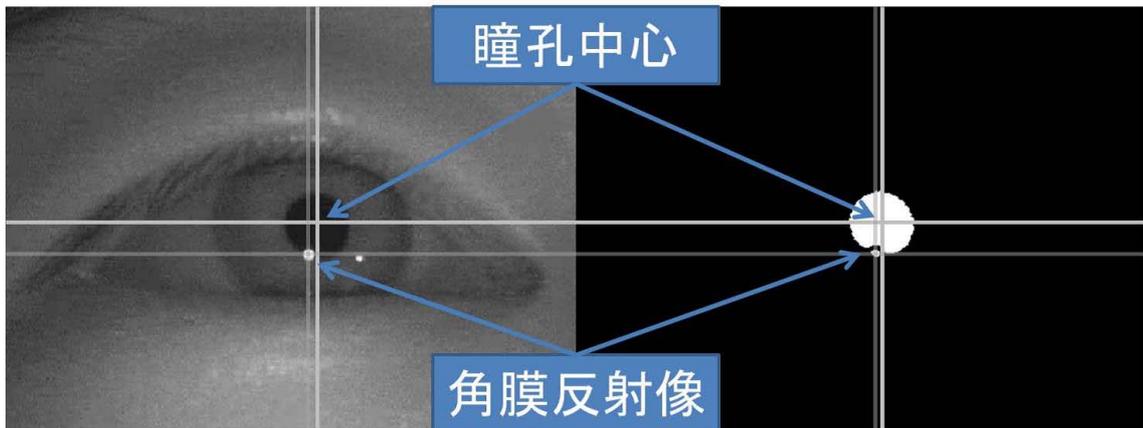


図 3.3: 角膜反射像と瞳孔の検出

像と、瞳孔中心の位置関係から眼の動きを求める方法である。眼球カメラの横に取り付けられているライトから、近赤外線を眼に当て、角膜反射像を出現させる。この反射像は眼を動かしても、近赤外線を当てているライトが動かない限り不動である。反射像と瞳孔中心の位置は、眼球カメラ映像の二値化により求められる。求めた反射像と瞳孔中心の相対的な位置関係より、眼の動きを検出することができる。しかし、眼の動きには、眼球の大きさなどによる個人差がある。この個人差を補正し、両眼の動きからユーザの視点を特定するためにキャリブレーションを行う。

### 3.2.2 キャリブレーション

キャリブレーションでは、角膜反射法で得られた眼の動きを補正し、ユーザの視点を特定する。まず、ユーザの正面に、一定距離 (キャリブレーション距離) 離れた壁などの平面 (キャリブレーション面) を設ける。そして、視野カメラ映像上に出現する 9 個の点を、この平面上に映し出す。この 9 個の点をユーザが順番に見てもらうことで、それぞれの点に対する両眼の動きが記録される。記録された両眼の動きを基に、眼球の個人差に対する補正パラメータを取得する。取得された補正パラメータを用いて、ユーザの視点位置を計算する。これにより、両眼の動きと、ユーザの視点を一致させることができる。

図 3.4 にキャリブレーション後、視野カメラ映像上に表示される視点情報を示す。視野カメラ映像上には、キャリブレーション距離離れた仮想平面上での、左右の眼の視点位置と補正済み視点位置が表示される。キャリブレーション面での視点位置が表示されるた

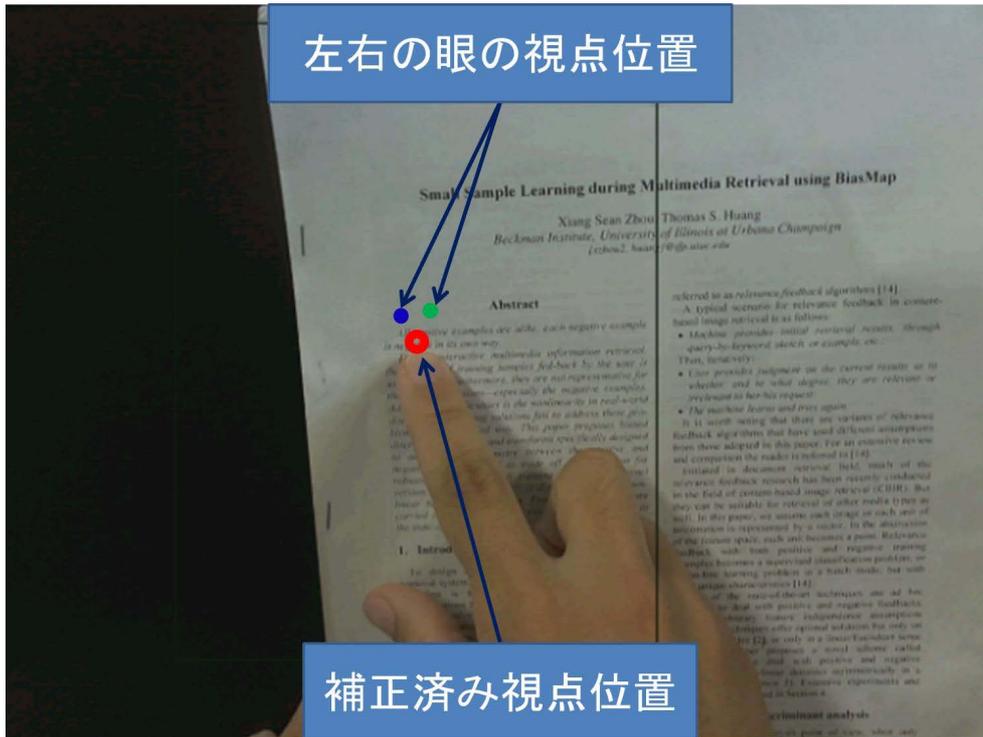


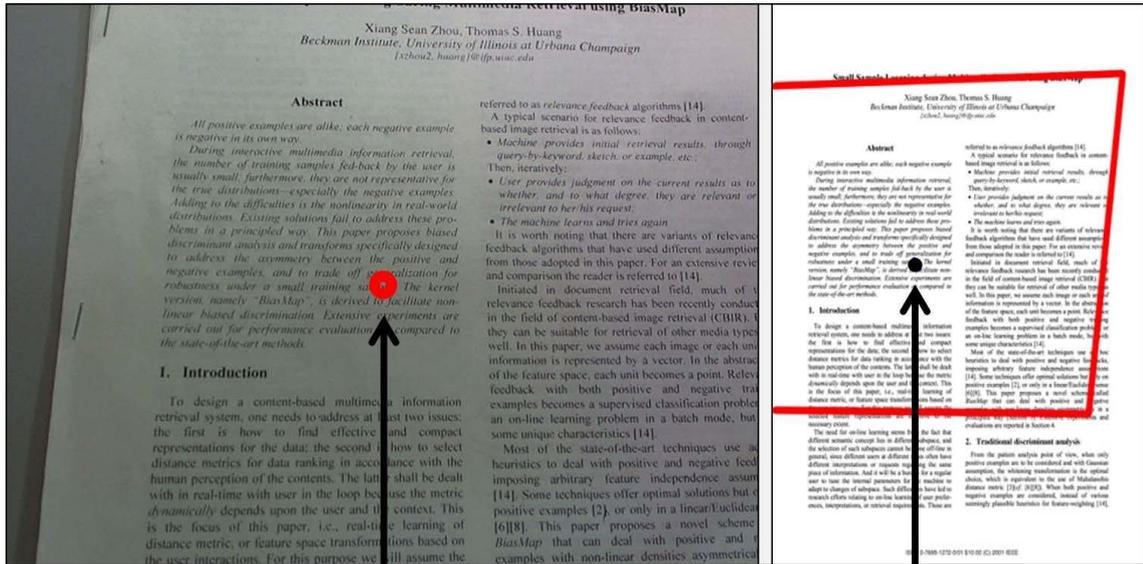
図 3.4: キャリブレーション済みの視点情報

め、見ている対象が遠のいたり近づいたりすると、左右の眼の視点位置は互いに遠のいたり近づいたりする。左右の眼の視点位置と補正済み視点位置が重なるのは、見ている対象がキャリブレーション距離離れた位置にあるときだけである。

### 3.3 視点情報の変換

文書画像検索とアイトラッカを組み合わせることにより、ユーザの読んだ文書の文書画像と視点情報が得られる、ユーザが文書のどこを読んだのかを特定するためには、文書画像上での視点位置の特定が必要である。しかし、アイトラッカから得られる視点情報は、視野カメラ映像上の座標で表されている。従って、視点情報の座標を文書画像上の座標に変換する必要がある。そこで、LLAH で求める射影変換パラメータを利用する。以下で、視点情報の変換処理を説明する。

図 3.5 に文書画像上に変換された視点情報を示す。文書画像検索を行う際、質問画像中に検索したい文書以外の文書などが写っていると、検索精度が下がってしまうことがあ



視点情報

変換された  
視点情報

図 3.5: 文書画像上に変換された視点情報

る．そのため，視野カメラ映像上に表示される視点情報を中心とした画像領域を切り出す．そして，切り出された画像領域を用いて，LLAH による文書画像検索を行う．LLAH では，検索過程で検索質問画像とそれに対応する登録文書画像の特徴点の対応関係が得られる．得られた対応関係から射影変換パラメータを推定することができ，検索質問画像の撮影範囲を特定することができる．また，特定された撮影範囲は，登録文書画像上に四角形で表示される．このように LLAH により得られた射影変換パラメータを用いて，視点情報を登録文書画像上に変換する．これにより，視点情報は文書画像上の座標に変換され，文書上のどこを読んでいるかを記録することができる．

### 3.4 視点情報の表示

図 3.6 に文書画像上に表示される視点情報を示す． 3.3 節で述べた変換によって，アイトラッカから得られる視点情報は，文書画像上の座標に変換される．変換された視点情報を文書画像上に点で描写すると，図 3.6(a) のようになる．このように，文書画像上に表示

### Abstract

All positive examples are alike; each negative example is negative in its own way.

During interactive multimedia information retrieval, the number of training samples fed-back by the user is usually small; furthermore, they are not representative for the true distributions—especially the negative examples. Adding to the difficulties is the nonlinearity in real-world distributions. Existing solutions fail to address these problems in a principled way. This paper proposes biased discriminant analysis and transforms specifically designed

(a) 点のみ

### Abstract

All positive examples are alike; each negative example is negative in its own way.

During interactive multimedia information retrieval, the number of training samples fed-back by the user is usually small; furthermore, they are not representative for the true distributions—especially the negative examples. Adding to the difficulties is the nonlinearity in real-world distributions. Existing solutions fail to address these problems in a principled way. This paper proposes biased discriminant analysis and transforms specifically designed

(b) 点と線

図 3.6: 視点情報の表示

された視点情報によって、文書のどこを読んだか特定することができる。しかし、点だけで表すと、どのように視点が移動したかが分からない。そこで、図 3.6(b) のように、それぞれの点を記録された順番に線で結ぶ。これにより、ユーザの視点がどのように移動したか分かるようになり、単語の読み返しなどを特定することができる。

*All positive examples are alike; each negative example is negative in its own way.*

*During interactive multimedia information retrieval, the number of training samples fed-back by the user is usually small; furthermore, they are not representative for the true distributions—especially the negative examples. Adding to the difficulties is the nonlinearity in real-world distributions. Existing solutions fail to address these problems in a principled way. This paper proposes biased discriminant analysis and transforms specifically designed*

(a) ノイズ

*All positive examples are alike; each negative example is negative in its own way.*

*During interactive multimedia information retrieval, the number of training samples fed-back by the user is usually small; furthermore, they are not representative for the true distributions—especially the negative examples. Adding to the difficulties is the nonlinearity in real-world distributions. Existing solutions fail to address these problems in a principled way. This paper proposes biased discriminant analysis and transforms specifically designed*

(b) 意図的な視点移動

図 3.7: ノイズの判定

## 3.5 視点情報の分析

人は無意識のうちに、瞬きやよそ見をしている [13]。これらは、視点の大きな移動として記録され、意識して見ている箇所とは、別のところに視点情報を表す。そこで、瞬きなどによるノイズを除去する。また、視点情報が一定の範囲内に集中していることがある。この場合、その範囲を特に注意して見ていることが考えられる。そこで、このような視点が集中しているところを注視点として記録する。以下に、それぞれの処理について示す。

### 3.5.1 ノイズの除去

図 3.7 に、ノイズと意図的な視点移動を示す。記録された視点情報の中には図 3.7(a) のように、本来読んだ箇所とは大きく異なった箇所に出現する視点情報がある。これは、瞬

きによって眼の動きが検出できなかったことなどが原因である。このように、ノイズは大きな視点の移動として記録される。しかし、図 3.7(b) のような意図的な視点移動も、同様に大きな視点の移動として記録される。従って、ノイズと意図的な視点移動を区別する必要がある。そこで、それぞれ次の視点の位置に注目する。ノイズの場合、ノイズである視点の一つだけ離れた箇所に現れることが多く、次の視点がノイズの近くに出現しない。一方、意図的な視点移動の場合、大きく移動した先にある文字を読むため、視点は移動した先を注視する。つまり、次の視点は、大きく移動した視点の近くに出現する。このように、大きく移動した後の視点の位置で、ノイズと瞬きなどの視点移動を区別することができる。ノイズと判断された場合は、その視点情報は除去し前後の視点を線で結ぶ。意図的な視点移動と判断された場合はそのまま記録する。

### 3.5.2 注視点の判定

図 3.8 に注視点判定の例を示す。記録された視点情報の中には、狭い範囲に集中しているものがある。これは、ユーザがそこにあるものを注視しているからである。人は物を見たり文書を読んだりする際、数百ミリ秒ほど注視をした後、視点を動かし他の場所を再び注視する眼球運動を繰り返している [13]。また、読書の場合、注視している時間は、注視している単語や文字の複雑さや、意味を理解する速さによって変化する。つまり、注視点を特定することで、ユーザが注意して読んだ単語などの特定が可能になる。そこで、記録された視点情報に対し、注視しているかの判定を行い、注視点を特定する。

記録された視点情報の座標を順番に検査する。もし、狭い範囲に視点情報が閾値  $n$  個以上連続しているとき、そこを注視していると判定する。注視していると判定されると、範囲内にある視点情報の平均座標を算出する。そして、範囲内の視点情報を削除し、算出した平均座標を記録する。これにより、狭い範囲に集中していた視点情報を、一つの視点情報にまとめることができ、これを注視点とする。この注視点は、範囲内にある視点情報の数に比例した大きさで表示し、注視の度合いが見てわかるようにする。

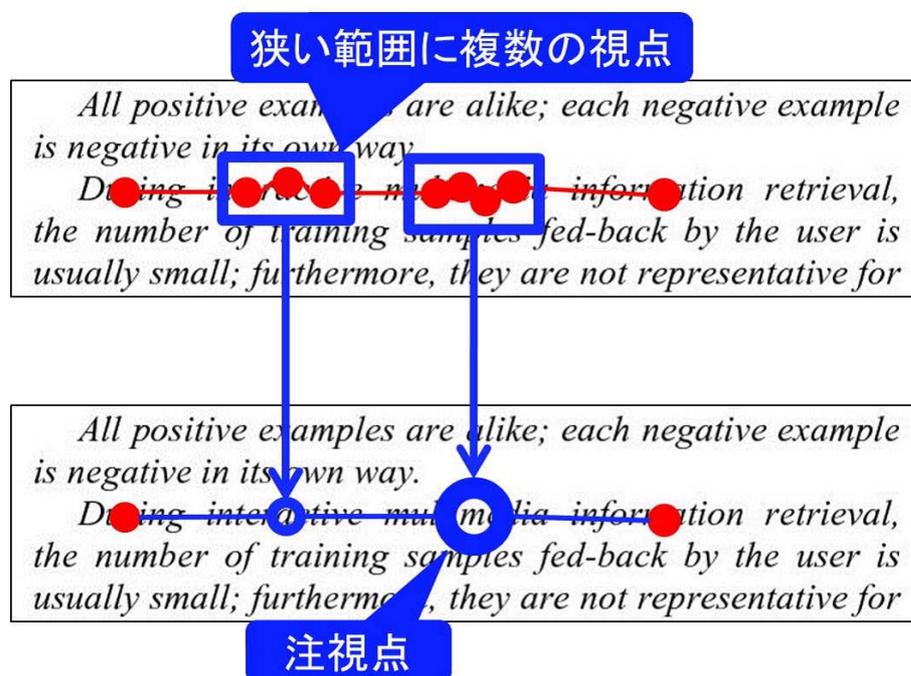


図 3.8: 注視点の判定



## 第4章 実験

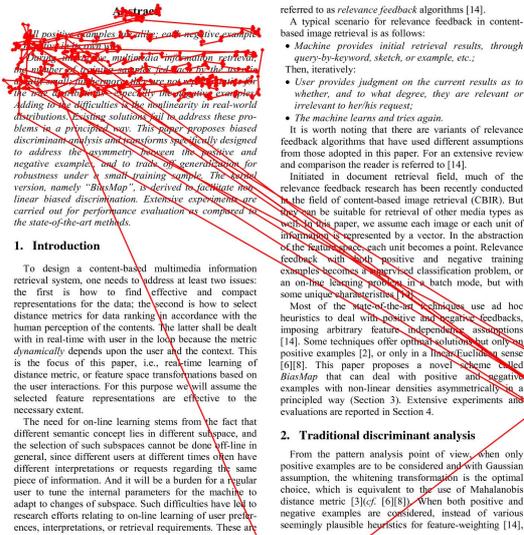
リーディングライフログ実現のために、文書画像検索とアイトラッカを組み合わせることの有効性を検証した。被験者に文書を読んでもらい、視点情報を記録し、ノイズ除去と注視点判定を行った。そして、分析した視点情報を用いて、実際に読んだ行に対する記録された視点のずれを調べた。

### 4.1 実験条件

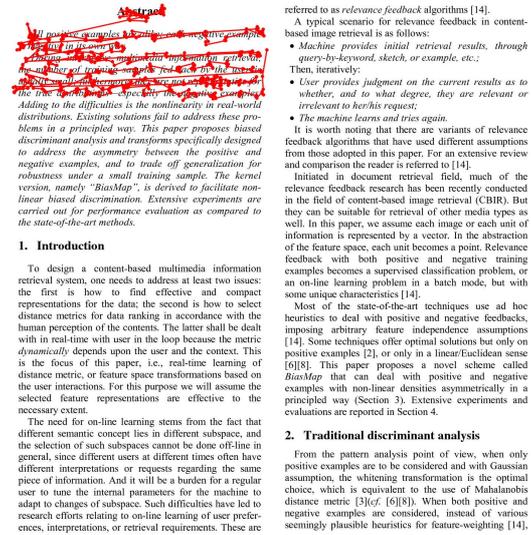
図 4.1 に誤差の例を示す。ここでは、視点情報と読んだ行の中心線との距離を誤差とする。被験者は3名で、全員裸眼の状態である。被験者から、実際に読む文書までの距離は30cmとした。これに従い、アイトラッカのキャリブレーション距離を30cmにして、キャリブレーションを行なった。被験者が読む文書は1枚に限定する。LLAHで使っている文書画像の大きさは1700×2200である。読んでもらう文書は被験者が読みやすいようにA3用紙に印刷したものをを用いた。被験者には、文書中のあらかじめ指定した部分を読んでもらった。また、文書を読む際は、よそ見や流し読みをしないように黙読してもらった。以上の条件で、文書と頭部を固定した場合、文書のみを固定した場合、頭を固定せず文書を手で持つ場合で実験し、視点情報の精度を調べた。



図 4.1: 誤差



(a) 記録された視点情報



(b) ノイズ除去後

図 4.2: ノイズの判別

## 4.2 実験結果

### 4.2.1 視点情報の分析

図 4.2 に記録された視点情報とノイズを除去した視点情報を示す．図 4.2(a) は実際に記録された視点情報である．図 4.2(a) より，実際に読んでいる部分とは大きく離れた位置に，視点情報が点在することがわかる．これが，瞬きなどによって，眼の動きが読み取れなかった際に出現するノイズである．一方，図 4.2(b) はノイズを除去した後の視点情報である．ノイズ除去処理を施すことにより，図 4.2(a) で出現していたノイズが，除去されたことがわかる．また，ノイズと意図的な視点移動を区別して，ノイズのみを除去することができた．

図 4.3 に注視点を判定する前の視点情報と，注視点を判定した視点情報を示す．ここでは，次の視点が 40[ $\text{pixel}$ ] 以内の距離にある視点が，3 個以上連続する場合，注視していると判定した．図 4.3(a) の注視点判定前の視点情報から，視点が狭い範囲に集中しては移動を繰り返していることがわかる．図 4.3(b) のように，注視点判定後の視点情報では，集

中していた視点を，それぞれ一つの注視点にまとめることができた．注視点判定により，ユーザが注視している単語を特定することができた．また，注視点の大きさを注視の度合いに比例させたことで，単語によって注視の度合いが違うことが分かった．

**Abstract**

*All positive examples are alike; each negative example is negative in its own way.*

*During interactive multimedia information retrieval, the number of training samples fed-back by the user is usually small; furthermore, they are not representative for the true distributions—especially the negative example. Adding to the difficulties is the nonlinearity in real-world distributions. Existing solutions fail to address these problems in a principled way. This paper proposes biased discriminant analysis and transforms specifically designed*

(a) 注視点判定前

**Abstract**

*All positive examples are alike, each negative example is negative in its own way.*

*During interactive multimedia information retrieval, the number of training samples fed-back by the user is usually small; furthermore, they are not representative for the true distributions—especially the negative example. Adding to the difficulties is the nonlinearity in real-world distributions. Existing solutions fail to address these problems in a principled way. This paper proposes biased discriminant analysis and transforms specifically designed*

(b) 注視点判定後

図 4.3: 注視点判定

表 4.1: 精度 [pixel]

被験者	文書・頭固定	文書固定	手持ち文書
A	22.3	49.6	62.0
B	12.3	27.3	45.3
C	15.3	16.3	28.3
平均	16.6	31.1	45.2

#### 4.2.2 精度検証

表 4.1 に視点情報と読んでいる行の誤差を示す．文書と被験者の頭部を固定した場合，視点情報と実際に読んだ行との誤差は，平均で 16.6pixel であった．視点情報を記録する文書画像は大きさが 1700×2200 であり，各行およそ 30pixel ずつ離れている．このことから，文書と頭部を固定した場合の誤差は，文書画像上では 1 行に満たないことが分かった．また，実際に被験者が読んだ文書は A3 用紙であるため，この誤差は文書上では 3, 4mm 程度の誤差である．文書のみを固定した場合は，平均で 31.1pixel の誤差があり，視点情報が 1 行程度ずれていることが分かった．この誤差は，頭部が固定されていないことで，無意識に頭部が動いていることに起因する．手持ち文書ではさらに誤差が増加し，平均で 45.2pixel であり，2 行程度の誤差があった．誤差の原因としては，手で持っている文書が細かく動いていることや，文書が非線形の形状変化をしたことが考えられる．視野カメラで撮影している文書が細かく動くと，LLAH の撮影範囲特定に誤差が生じてしまう．撮影範囲特定の誤差は，視点情報の変換に影響を与えるため，変換された視点情報にも誤差が生じたと考えられる．また，アイトラッカの精度も原因として挙げられる．アイトラッカの精度は，被験者の眼の開き具合などによって，大きく左右されてしまう．

図 4.4 に誤差が少なかった時の視点情報を示す．矩形で囲んだ単語に注目すると，注視点が単語領域の中心付近に集中している．これは，人が文書を読む際，単語の中心付近を見ながら読んでいるからであると考えられる．また，注視点の大きさを注視の度合いに比例させたことから，どの単語を特に注視しているか判断することができた．これにより，文字数の多い単語や難しい単語は，注視の度合いが高かったり，単語領域に複数の注視点があったりすることが分かった．各単語に対する注視の度合いの違いから，ユーザにとっ

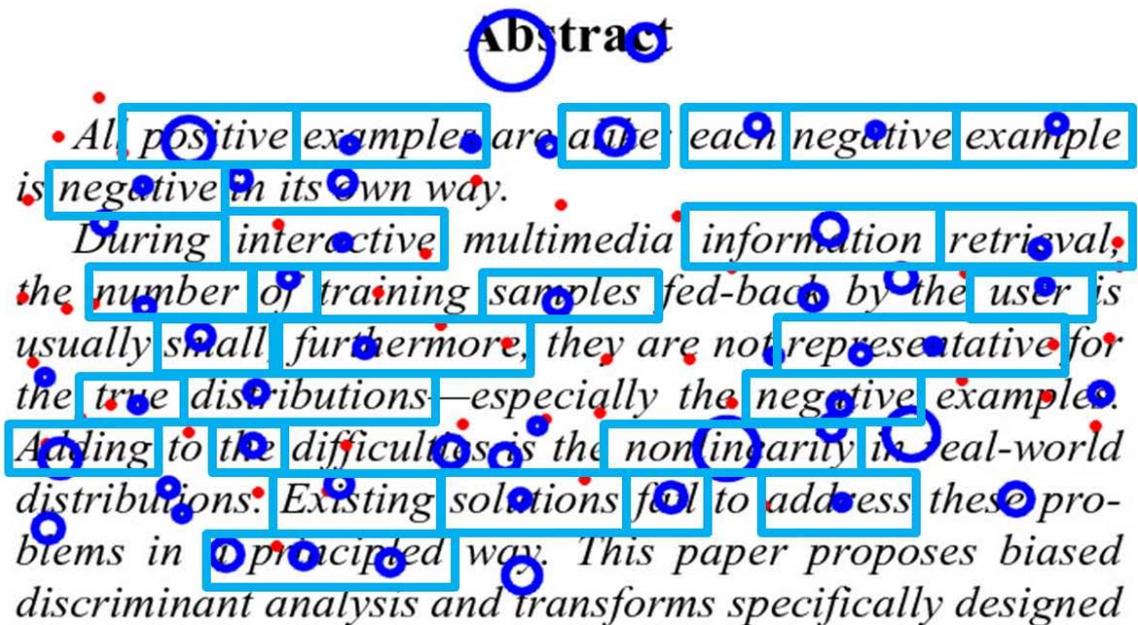


図 4.4: 注視点

てどの単語が重要であるかを推定することが可能であると考えられる。

精度検証により、記録された視点情報には1, 2行程度の誤差があることが分かった。この程度の誤差であれば、ユーザが十数行にわたって文書を読んだ場合、ユーザが読んだ行を推定することができる。数十行読んだ場合、1, 2行程度の誤差は大きなずれではなく、どこを読んだかを特定するには問題にならない。また、得られた視点情報からユーザが読んだ単語を推定することもできた。よって、本手法はリーディングライフログにとって有効であると言える。

得られたリーディングライフログの利用方法として、読んだ覚えがある単語をどの文書で読んだのか検索するキーワード検索が考えられる。本手法では、読んだ文書と文書中の読んだ箇所を特定し記録することができる。従って、読んだことのある文書に限定したキーワード検索が可能であり、ユーザが意思をもって読んだ文字だけを検索することができる。また、注視点に注目すると、注視の度合いが高い単語や何度も注視している単語から、ユーザの興味のある分野や物事を特定することが可能である。さらに、LLAHのリアルタイム性から、ライフログとは異なるインタラクティブな利用も考えられる。例えば、注視の度合いが高い単語に対して、単語の意味や翻訳の仕方などを音声やプロジェクトを用いてユーザに伝えることが可能である。



## 第5章 結び

本論文では，リーディングライフログの実現に向け，記録された視点情報の分析と精度を検証した．その結果，記録された視点情報からノイズを除去し，注視点を特定することができた．また，分析した視点情報から，実際に読んだ行との誤差が1，2行程度であることが分かった．リーディングライフログの目的は，ユーザの読んだすべての文書や単語を記録することである．読んだ行が複数行に渡る場合，視点情報が1，2行程度ずれていてもユーザの読んだ部分の大半を特定することはできる．以上より，本手法はリーディングライフログに有効であると言える．

今後の課題として，視点情報の精度の向上があげられる．アイトラッカのキャリブレーション結果を補正することや，頭部や文書のモーションブラーに対処する必要がある．また，細かな文書の手振れに対して頑健になるように，文書画像検索を改良することがあげられる．

本研究では，被験者に文書の読み方などの条件を課した．しかし，実際に文書を読むときは，必ずしもすべての文字を見ているとは限らず，また，流し読みをすることも考えられる．そのため，視点の動きから，被験者が文書を読んだか，読まなかったかの判定を行い，よりユーザの視点の動きに忠実な記録を取ることを目指す．



## 謝辞

本研究を進めるにあたって、直接御指導頂いた黄瀬浩一教授には、研究内容や論文の書き方、発表方法において多くの御指導、御助言を頂いたほか、活発な研究活動を導いて頂いたことを深く感謝致します。また、研究発表会等で様々な指摘及び助言をしてくださった岩村雅一准教授、小島篤博准教授、内海ゆづ子助教に感謝致します。外山託海氏には、技術面でのご指導や、研究の進め方など、研究の心得えをご教授して頂きましたことを深く感謝いたします。最後に、公私にわたり様々な支援及び助言をしてくださった知能メディア処理研究グループの皆様に感謝致します。

2012年3月9日



## 参考文献

- [1] <http://www.sophia-it.com/content/Lifelog>.
- [2] <http://daytum.com/>.
- [3] <http://www.foodlog.jp/>.
- [4] <http://www.viconrevue.com/home.html>.
- [5] David Beymer and Daniel M Russell. *WebGazeAnalyzer: a system for capturing and analyzing web reading behavior using eye gaze*, pp. 1913–1916. ACM, 2005.
- [6] Ralf Biedert, Georg Buscher, and Andreas Dengel. The eyebook - using eye tracking to enhance the reading experience. *Informatik Spektrum*, Vol. 33, No. 3, pp. 272–281, 2010.
- [7] <http://eyewriter.org/>.
- [8] Mostafa El Hosseiny, Ralf Biedert, Andreas Dengel, and Georg Buscher. The eyepad - tom riddle in the 21st century. In *Proceedings of the 2nd Workshop on Eye Gaze in Intelligent Human Machine Interaction held in conjunction with IUI 2011*, 2011.
- [9] Farida Ismail, Ralf Biedert, Andreas Dengel, and Georg Buscher. Emotional text tagging. In *Proceedings of the 2nd Workshop on Eye Gaze in Intelligent Human Machine Interaction held in conjunction with IUI 2011*, 2011.
- [10] 木村崇志, フォンヤオカイ, 内田誠一, 岩村雅一, 大町真一郎, 黄瀬浩一. Reading-lifelog の実現に向けた予備的検討. 電子情報通信学会技術研究報告, Vol. 111, No. 317, pp. 51–56, nov 2011.

- 
- [11] 竹田一貴, 黄瀬浩一, 岩村雅一. メモリ効率と安定性を向上させた LLAH によるカメラベース大規模実時間文書画像検索. 画像の認識・理解シンポジウム (MIRU2011), pp. 350–357, 2011.
- [12] 坂下祐輔, 藤吉弘巨, 平田豊. 画像処理による 3 次元眼球運動計測. JSEM 日本実験力学学会特集号 : 画像処理による生体評価 「解説記事」, Vol. 6, pp. 236–243, 2006.
- [13] J. Henderson. Human gaze control during real-world scene perception. *Trends in Cognitive Sciences*, Vol. 7, No. 11, pp. 498–504, nov 2003.