

# 読書アノテーション支援システム実現に向けたアイジェスチャの提案

鈴木 若菜<sup>†</sup> 外山 託海<sup>††</sup> Kai Kunze<sup>†</sup> 岩村 雅一<sup>†</sup> 黄瀬 浩一<sup>†</sup>

AndreasDengel<sup>††</sup> AndreasBulling

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

<sup>††</sup> ドイツ人工知能研究センター

Trippstadter Strasse 122 D-67663 Kaiserslautern, Germany

E-mail: <sup>†</sup>{wakana, kunze}@m.cs.osakafu-u.ac.jp, <sup>†</sup>{kise, masa}@cs.osakafu-u.ac.jp, <sup>††</sup>{Takumi.Toyama, Andreas.Dengel}@dfki.de, <sup>††</sup>bulling@mpi-inf.mpg.de

**あらまし** 本稿では、文書に新たな価値を与える読書アノテーション支援システムの実現に向けたアイジェスチャを検討する。文書にアノテーションをする際、手元にある文書だけでなく、ユーザの手の届かない場所にある文書やヘッドマウントディスプレイに表示した文書など、あらゆる文書に対してアノテーションを可能とする方法の一つは、視点を用いることである。そこで、本システムでは、アノテーションのトリガーとしてアイジェスチャを用い、任意のアイジェスチャをすることで、それに対応したアノテーションを付加する。このとき、文書を読んでいるのかアイジェスチャをしているのかを区別でき、さらに、アイジェスチャ同士の判別ができる必要がある。また、ユーザビリティの観点から、人にとって受け入れやすいジェスチャである必要がある。そこで、読書アノテーション支援システムにおけるアイジェスチャの有用性を示すため、認識精度の検証およびユーザスタディによるアイジェスチャのしやすさを評価する。

**キーワード** 読書アノテーション, アイトラッカ, アイジェスチャ, アクティブリーディング

## 1. はじめに

日常生活において、新聞や書籍、資料など、文書に触れる機会は多々ある。そして、文書を読んでいる最中に、単に読むだけでなく、気になる箇所に書き込みをしたり、重要な箇所にハイライトをしたりする。例えば、タブレット PC に表示した電子文書に対してコメントを付加したり、後で見返したい箇所に蛍光ペンでマークしたりすることが挙げられる。したがって、アノテーションを使用して情報を記録することは、その人のリーディングライフをより豊かにするための重要な要素であると考えられる。このように、人が読んだ情報や読書中の文書に対する行動の結果を記録し、解析するものひとつとしてリーディングライフログ [1] [2] がある。文書へアノテーションをする行動は、文書と紐付けることで文書に新たな価値を与えることができるため、リーディングライフログにおいて、重要な要素となり得る。我々はすでに、アノテーションを参照するシステムとして、プロジェクトやヘッドマウントディスプレイなどのデバイスを利用し、文書中のキーワードに対する情報表示システムを提案している [3] [4]。このようにアノテーションを参照するだけでなく、ユーザ自身がアノテーションを付加したり、他のユーザと共有したりすることができれば、よりインタラクティブで高度なリーディングライフログを取得することができる。

アノテーションを用いたリーディングライフログを実現するために必須となる条件が 2 つある。第一に、アノテーションを

したい箇所とその内容を忘れないようにするため、ユーザが意図したときに、その場ですぐに情報を付加することである。第二に、どこにどんなアノテーションをするかを決定するためにユーザが容易に情報を入力できることである。この入力のトリガーとして、指や視線などが利用できる [5] [6]。指を利用する場合、アノテーションをしたい文書がユーザの手の届く範囲にあることが必要である。一方、視線を利用する場合、アノテーションしたい文書が手の届かないところにあったり、ヘッドマウントディスプレイのようなデバイスに表示したりした場合でもターゲットに焦点を合わせることでアノテーションが可能である。したがって、視線を利用したアノテーションシステムがあれば、あらゆる文書に対してアノテーションをすることができ、より高度なリーディングライフログを実現できると考えられる。

そこで、本研究では視線を用いた読書アノテーション支援システムを構築することを目的とする。このシステムでは、図 1 に示すように、ユーザの視点を特定することのできるアイトラッカと目の前に映像を映し出すことができるヘッドマウントディスプレイを組合わせた装置を使用する。ユーザはアノテーションをしたい箇所を注視し、アイジェスチャをすることでアノテーションをする。そして、そのアノテーションをした結果をヘッドマウントディスプレイに表示する。視線を利用した読書アノテーション支援システムを実現するために、(1) どこに (2) どんなアノテーションをつけるかを特定する必要がある。



図1 デバイス

(1) に対しては、イトラッカと文書画像検索の一手法である LLAH ( Locally Likely Arrangement Hashing ) [7] を組み合わせることで、ユーザが文書上のどこを見ているか特定する。(2) に対しては、アノテーションとアイジェスチャを対応させ、そのアイジェスチャを認識することで特定する。

本稿では、視線を利用した読書アノテーション支援システムの実現に向けた初期段階として、読書中のアノテーションに適したアイジェスチャを提案する。その有用性を示すため、ユーザスタディにより、アイジェスチャのしやすさを評価する。さらに、選択したアイジェスチャが認識されるかを調査するため、アイジェスチャの認識精度を検証する。そして、読書中にアノテーションする際に、ユーザに受け入れられやすく、認識しやすいアイジェスチャを検討する。

## 2. 関連研究

本節では、関連研究について紹介する。まず、読書中にアノテーションなどをするアクティブリーディングに関連する研究について紹介する。[8] では、アクティブリーディングを実現する“XLibris”というものを提案している。これは、電子文書上でハイライトやアンダーラインなどのアノテーションを自由に行うことができるものである。さらに、それらにリンクを付けることで、後からアノテーションを検索し参照することができるものである。このシステムはタブレット端末での使用を対象としたものであり、アノテーションはユーザの手の届く範囲に限られる。これに対して、本研究では、視点をを用いてあらゆる文書にアノテーションできるシステムを目指す。

次に、視点をを用いたインタラクションやアイジェスチャに関する研究について紹介する。[9] では、ディスプレイに表示した文書に対して人の目の動きに着目している。このシステムでは、リアルタイムで訳語を表示したり、脚注の説明を表示したりすることができる。このようにして視点をを用いたインタラクションを実現している。[10] は、人の日常の生活を記録するライフログの取得に視点を利用している。アイジェスチャの研究においては、[11] で、視線入力におけるミダスタッチ [12] と呼

ばれる問題に対処するためにアイジェスチャを用いている。ミダスタッチの問題とは、操作のために注視しているのか単にもものを見ているのかの区別が困難であるというものである。

操作のために注視しているのか単にもものを見ているのかの区別が困難である問題に対処するため、アイジェスチャを用いている。ここでは、学習によりアイジェスチャを認識している。[13] は、携帯端末を用いたアイジェスチャの認識システムを提案している。このシステムでは、携帯端末の内蔵カメラでアイジェスチャを記録し、認識している。

以上のように視点をを用いたインタラクションやアイジェスチャの認識システムは提案されているが、アイジェスチャを用いて文書へのアノテーションをするものはない。そこで、本研究では、アイジェスチャを文書へのアノテーションのトリガーとして用いる。

## 3. 読書アノテーション支援システム

本研究で構築を目指す読書アノテーション支援システムについて説明する。本システムの動作例を図2に示す。本システムでは、ユーザはイトラッカとヘッドマウントディスプレイを組合わせたデバイスを装着する。そして、文書画像検索により、ユーザが読んでいる文書を特定する。加えて、イトラッカから得られた視点情報を検索結果の文書画像上に変換することでユーザが文書上のどこを見ているかを特定する。本システムでは、読書中にアノテーションをしたいとき、文書上のその場所を注視し、一定フレーム以上注視が続くと、アノテーションモードに切り替える。アノテーションモードになると、まず、ヘッドマウントディスプレイ上に“Eye Gesture Start”と表示される。そして、ユーザがアイジェスチャをし、それが認識されると、どこにスタンプなどのアノテーションをしたかがヘッドマウントディスプレイ上に表示される。ユーザがHMDから視線を外し、文書に戻ると、ヘッドマウントディスプレイの表示が消え、アノテーションモードが終了する。以上のようにして本システムは動作する。

### 3.1 アイトラッキング

このシステムを実現するためのユーザがどこを見ているかを特定するアイトラッキング技術について説明する。本システムでは、SMI社の装着型イトラッカ ETG ( Eye Tracking Glasses) を使用する。このデバイスは、ユーザのしている風景を撮影するシーンカメラとユーザの目の動きを追跡するアイカメラからなる。イトラッカ使用の際には、ユーザは、まず、キャリブレーションをする。キャリブレーションでは、ユーザが実際のシーン中の1点または3点を見ることで、ユーザが実際に見ている点とシーンカメラ映像中の対応する点との位置合わせをする。キャリブレーションをすることにより、ユーザがシーンカメラ映像中のどこを見ているかを特定できる。

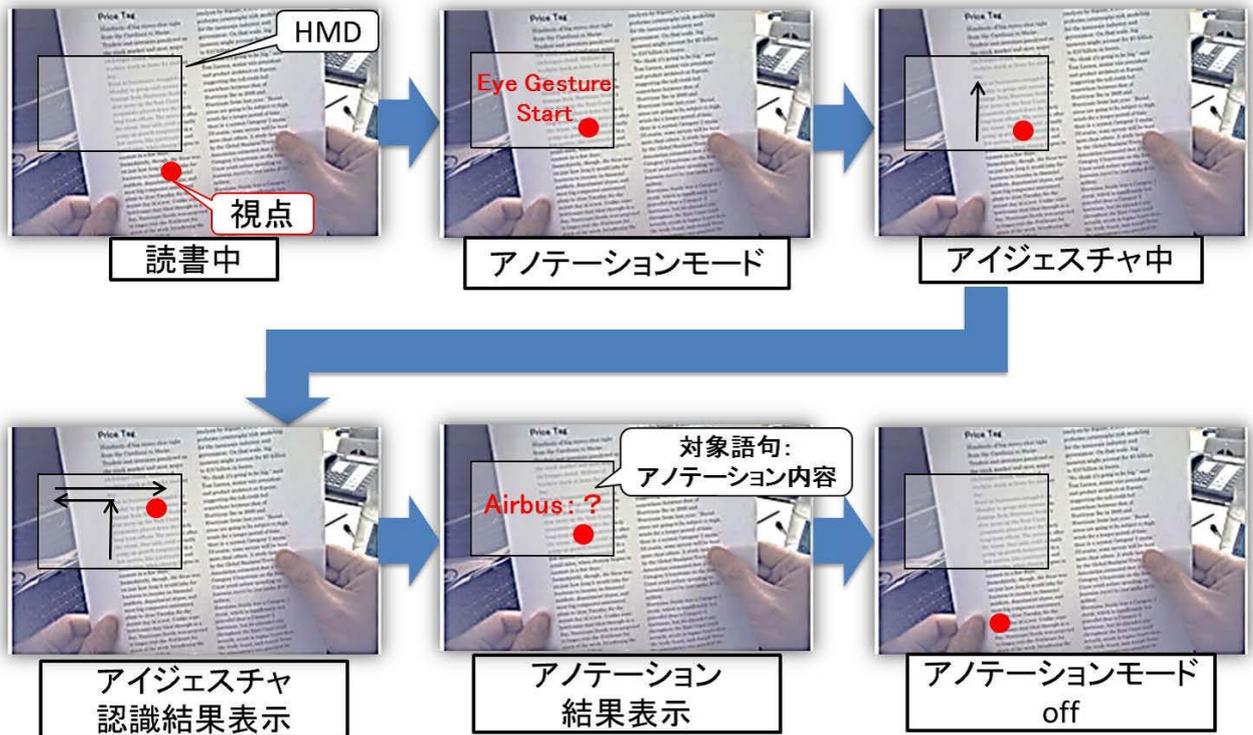


図 2 システム動作例

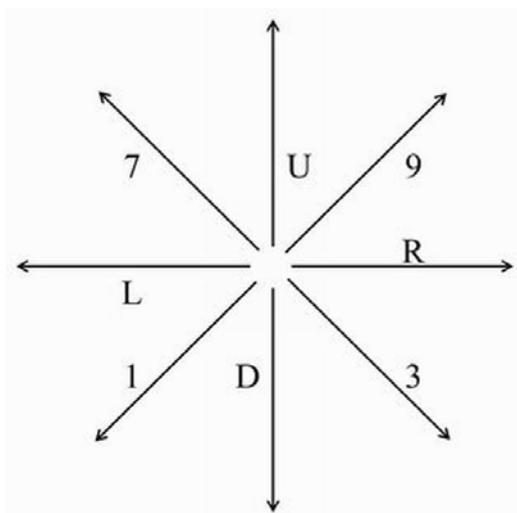


図 3 ベクトルの方向とアルファベットの定義

### 3.2 文書画像検索

ユーザがどの文書を見ているかを特定する文書画像検索について説明する。文書画像検索とは、カメラで撮影した文書が検索質問として与えられたとき、それに対応するデータベース中の文書画像を検索結果として出力するものである。本システムでは、文書画像検索手法の一つである LLAH を利用する。LLAH は、撮影方向の変化や隠れなどの外乱にロバストであり、リアルタイムで検索できるほどの高速性を持つという特徴がある。これらの特徴により、各文書に関連付けられた情報を撮影画像に重畳表示するといった文書への拡張現実 [14] やカメラペンシステム [15] などへの応用がなされている。LLAH では、検

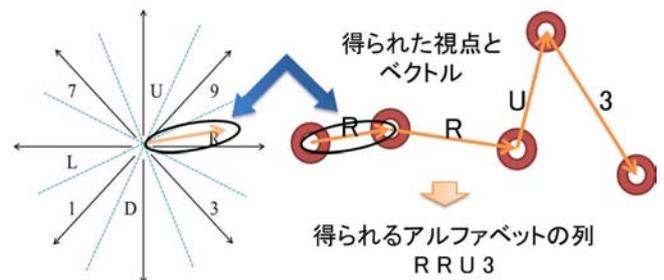


図 4 アルファベットの生成例

索の過程で検索質問とそれに対応する登録文書との特徴点の対応関係を得ることができる。この特徴点の対応関係を用いて、検索質問から登録画像への射影変換パラメータを算出することができる。これをもとに、アイトラッカのシーンカメラ映像上の視点を検索結果の文書画像上に変換する。変換された視点からユーザが文書上のどこを見ているかを特定することができる。

## 4. アイジェスチャの認識手法

アイジェスチャの認識では、アイトラッカから得られる視点から視点の動きを表すベクトルを作る。そして、図 3 に示すように、ベクトルに対応する英字または数字からなるアルファベットに置き換える。さらに、置き換えたアルファベットから n-gram を作成し、アイジェスチャと対応させることで認識する。

### 4.1 視点情報の取得と前処理

視点情報は、アイトラッカを用いて取得する。アイトラッカから得られる視点情報には、シーンカメラの座標系における視



図 5 ユーザから見たジェスチャの見本

点の座標や注視を表す Fixation, 注視と注視の間の素早い動きを表す Saccade, 瞬きを表す Blink と呼ばれるイベントの情報などが含まれる. アイトラッカから得られる視点情報には瞬きなど, アイジェスチャを認識する際にノイズとなる視点が含まれているため, 前処理として, アイトラッカから得られた視点情報に対してノイズ除去などを施す. 具体的には, Blink である点は除去し, Saccade である点はそのまま残す. そして, 連続する Fixation は, それらの平均座標にまとめる. さらに, アイジェスチャを認識する際, 移動距離が小さい視点はノイズとなるので閾値処理により除去する.

#### 4.2 n-gram の生成

次に, n-gram の作成方法について説明する. まず, 前処理により得られた視点からベクトルを計算する. そして, 図 3 に示すようなアルファベットを定義し [16], 得られたベクトルの方向に対応するアルファベットを生成する. その一例を図 4 に示す. 視点から生成したベクトルをそれぞれ対応するアルファベットに置き換える. そして, 生成されたアルファベットの列から n-gram を作る. n-gram を生成する際, 同じアルファベットが連続する場合はそれらをひとつにまとめる. そして, 生成した n-gram の中でユーザが実際にアイジェスチャをした箇所のもと, n-gram 化したアイジェスチャと比較し, 一致するものを認識結果とする.

### 5. 実 験

図 6 に示すように, 9 つのアイジェスチャを導入し, アイトラッカを用いてアイジェスチャを記録し, ユーザスタディおよび各アイジェスチャの認識精度実験を行った. さらに, 通常の読書中の眼の動きから誤認識の確率を調査し, アイジェスチャを評価した.

#### 5.1 ユーザスタディ

この実験では, 被験者は 6 人であり, 9 種類のアイジェスチャをそれぞれ 3 回ずつした. 具体的には, 図 5 に示すように, 各アイジェスチャをディスプレイに表示し, 被験者は, それを眼でなぞった. すべてのアイジェスチャ終了後, 読書中に

アイジェスチャ				
n-gram	7R1	ULR	39	RLRLRL
順位	1	2	3	4
DRUL	R1R	DR7RD7	3L9	3U1U
	5	6	8	9

図 6 アイジェスチャとその順位

アイジェスチャでアノテーションするという観点から, 被験者は, アイジェスチャのしやすさの順位をつけた. 評価の結果を図 6 に示す. 評価が一番高かったのは, “7R1” であった. 続いて, “ULR”, “39” の順に高い評価を得た. これは, 人が普段する眼の動きにアイジェスチャが含まれており, その眼の動きをするのにストレスを感じにくかったためであると考えられる. 加えて, シンプルでストローク数が少なかったこともアイジェスチャがしやすいと感じた一因であると考えられる. 一方で, “3U1U” や “DR7RD7” などは, 普段あまりしない眼の動きであり, ユーザの眼に負担がかかるため, 評価が低かったと考えられる.

#### 5.2 認識精度実験

被験者 6 人分, 9 種類のアイジェスチャを記録し, 得られた視点情報からアイジェスチャを認識し, その精度を検証した. 加えて, 視点情報の前処理における, 移動距離による視点の除去に閾値を設け, その値を 0 から 500 まで 50 ずつ変化させ, その精度を検証した.

図 7 に各アイジェスチャにおける被験者の認識精度の積み上げグラフを示す. 各棒は閾値に対応している. この結果から, 平均して, 閾値が 100 のとき, 認識精度が高いことがわかった. アイジェスチャごとでは, “ULR” が他のアイジェスチャに比べ認識精度が高いことがわかった. また, 被験者ごとに適切な閾値を適用すれば, さらに認識精度が向上し, 認識可能であると考えられる. ユーザスタディで評価が高かったアイジェスチャである “7R1” は, 視点の移動距離による閾値で視点を除去しても全閾値における認識精度の平均が 14% であり, “ULR” の 24% と比べ, 低かった. これは, 閾値を変化させてもノイズが除去しきれなかったためであると考えられる.

さらに, 図 7 の結果から得られた, ジェスチャごとの最適な閾値での Recall と Precision の値を表したグラフを図 8 に, 示す. 具体的には, “7R1”, “ULR”, および “39” の閾値が 100, “RLRLRL” の閾値が 150 のときの結果である. 各点がユーザを表している. この結果から, “ULR” が全体的に高い精度が得られた. また, 各ジェスチャにおいても Precision が十分高く, アイジェスチャとして認識可能であると考えられる.

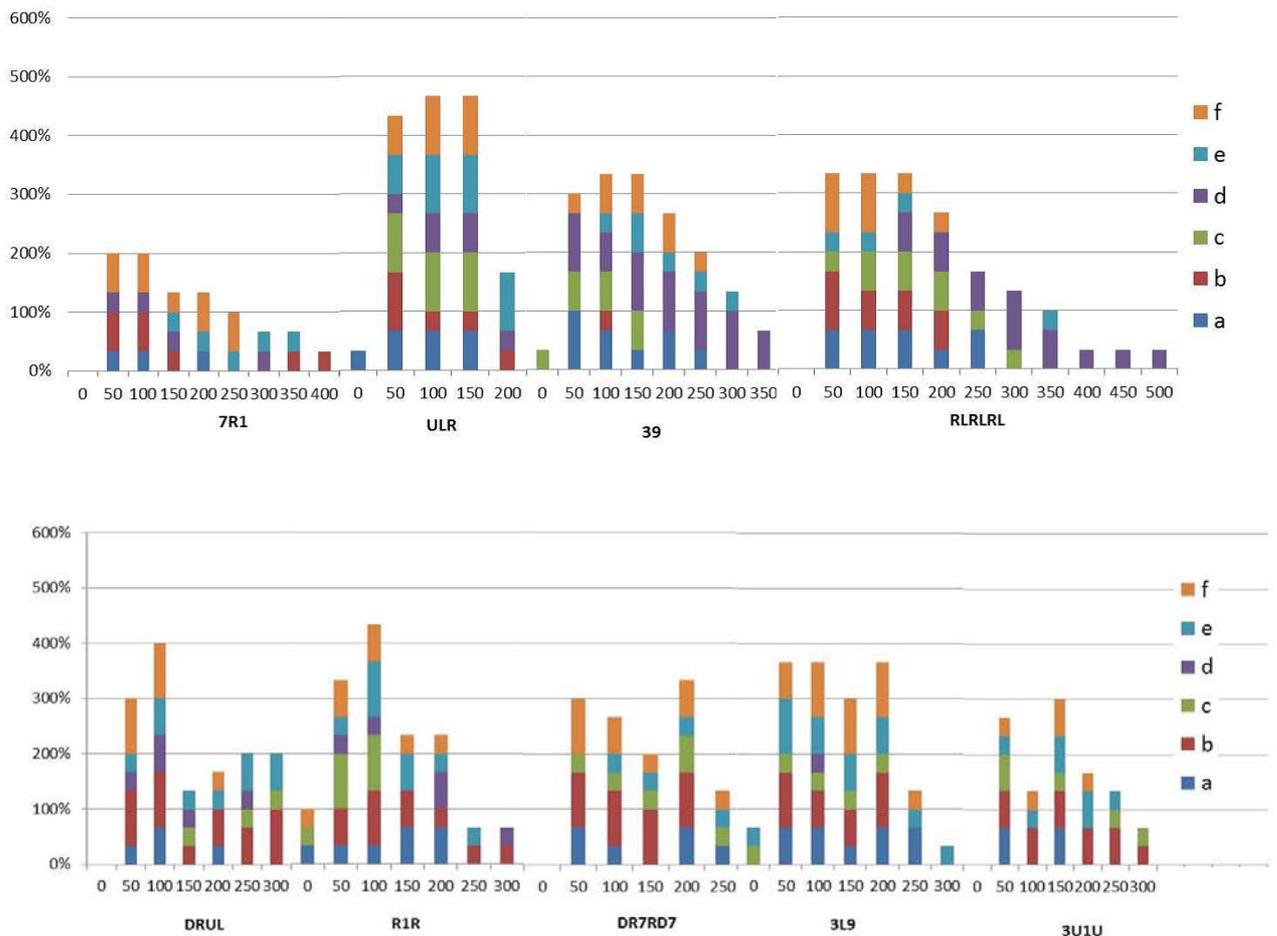


図7 アイジェスチャごとの認識精度

表1 ノーマルリーディングにおける“7R1”と“ULR”の出現率

	a	b	c	d	e	f	平均
7R1	0.23%	0.64%	0.65%	0.21%	0.40%	0.15%	0.38%
ULR	0.86%	0.32%	0.35%	0.76%	0.45%	0.70%	0.57%

Recall が低い部分に対しては、ユーザが再度ジェスチャをすることで対処できると考えられる。

### 5.3 ノーマルリーディングにおける“7R1”と“ULR”の出現率

アイジェスチャがユーザにとってやりやすいものであっても、通常の読書行動に頻出する場合、アイジェスチャをしたのか、読書中であるのかを区別できないという問題がある。したがって、ユーザスタディで評価が高かったアイジェスチャの出現頻度を調査する必要がある。そこで、通常の読書中のアイジェスチャの出現率を検証した。

この実験では、被験者は、英語で書かれた文書をアイジェスチャをせず、通常の読書と同じように読んだ。読んだ文書は14文書であった。そして、得られた視点情報から3-gramを作成し、ユーザスタディで評価の高かった“7R1”と“ULR”の出現率を調査した。表1に各被験者の1文書あたりの“7R1”と“ULR”の出現率を示す。両方のアイジェスチャにおいてどの被

験者もその出現率が1%未満であり、全被験者の平均も、“7R1”が0.38%、“ULR”が0.57%となり、1%未満であった。このことから、“7R1”や“ULR”は、アノテーションをするために、アイジェスチャとしてした目の動きと、通常の読書中の目の動きとの誤認識が起こる確率は低いことがわかった。したがって、“7R1”や“ULR”は、読書中に誤認識を起こしにくく、読書アノテーションのためのアイジェスチャとして導入できると考えられる。

## 6. まとめ

本稿では、読書アノテーション支援システム実現に向け、読書中のアノテーションに適したアイジェスチャを検討した。ユーザにとってやりやすいアイジェスチャは“7R1”や“ULR”、“39”であり、シンプルでストローク数が少なく眼に負担が少ないものであることがわかった。実験結果より、視点の移動距離に対し、閾値を設け、ノイズを除去することでアイジェスチャの認識率が向上した。また、“ULR”の認識精度が高いことがわかった。さらに、“ULR”は通常の読書中の目の動きに出現する確率は1%未満であり、誤認識が起こりにくいことがわかった。今後の課題としては、ユーザスタディで評価が高かったジェスチャを実際に読書中に行ったものを記録し、その認識精度を

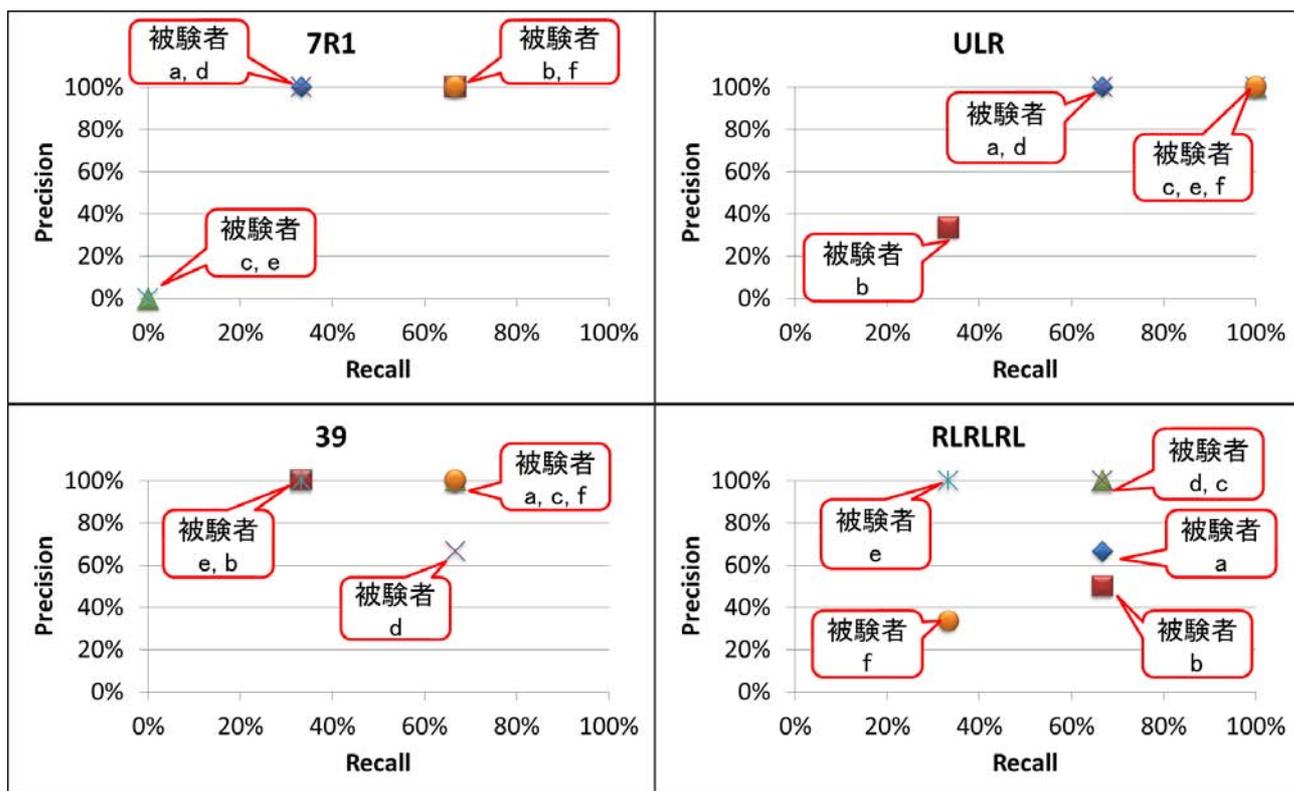


図8 認識精度

調べることが挙げられる。さらに、認識精度が高いアイジェスチャを利用したリアルタイムでのアノテーション支援システムを構築することが挙げられる。

**謝辞** 本研究の一部は、JST CREST および日本学術振興会科学研究費補助金基盤研究 (A) (25240028) の補助による。

#### 文 献

- [1] 川市仁史, K. Kunze, 黄瀬浩一, “文書画像検索を用いたリーディングライフログの提案,” 電子情報通信学会技術研究報告, vol.112, no.495, pp.261–266, Mar. 2013.
- [2] T. Kimura, R. Huang, S. Uchida, M. Iwamura, S. Omachi, and K. Kise, “The reading-life log – technologies to recognize texts that we read,” Proceedings of the 12th International Conference on Document Analysis and Recognition (ICDAR 2013), pp.91–95, Aug. 2013.
- [3] 鈴木若菜, 竹田一貴, 外山託海, 黄瀬浩一, “プロジェクトを用いた情報投影による印刷文書へのインタラクティブ性の付加,” 電子情報通信学会技術研究報告, vol.111, no.317, pp.69–74, Nov. 2011.
- [4] T. Toyama, W. Suzuki, A. Dengel, and K. Kise, “Wearable reading assist system: Augmented reality document combining document retrieval and eye tracking,” Proceedings of the 12th International Conference on Document Analysis and Recognition (ICDAR 2013), pp.30–34, Aug. 2013.
- [5] 丸谷和史, 植月美希, 安藤英由樹, 渡邊淳司, “ユーザのなぞり動作に基づく動的な文章表示方式,” 情報処理学会論文誌, vol.54, no.4, pp.1507–1517, 2013.
- [6] T. Toyama, W. Suzuki, A. Dengel, and K. Kise, “User attention oriented augmented reality on documents with document-dependent dynamic overlay,” Proceedings of Int'l Symposium on Mixed and Augmented Reality (ISMAR 2013), pp.299–300, Oct. 2013.
- [7] T. Nakai, K. Kise, and M. Iwamura, “Use of affine invariants in locally likely arrangement hashing for camera-based

- document image retrieval,” in Lecture Notes in Computer Science (7th International Workshop, vol.3872, pp.541–552, Feb. 2006.
- [8] B.N. Schilit, G. Golovchinsky, and M.N. Price, “Beyond paper: supporting active reading with free form digital ink annotations,” Proceedings of the SIGCHI conference on Human factors in computing systems ACM Press/Addison-Wesley Publishing Co., pp.249–256 1998.
- [9] R. Biedert, G. Buscher, S. Schwarz, J. orn.Hees, A. Dengel, “Text 2.0,” in Proceedings of the 28th of the International Conference on Human Factors, pp.4003–4008, 2010.
- [10] Y. Ishiguro, A. Mujibiya, T. Miyaki, and J. Rekimoto, “Aided eyes: eye activity sensing for daily life,” Proceedings of the 1st Augmented Human International Conference ACM, p.25 2010.
- [11] D. Rozado, F.B. Rodriguez, and P. Varona, “Gaze gesture recognition with hierarchical temporal memory networks,” Advances in Computational Intelligence, pp.1–8, Springer, 2011.
- [12] R.J.K. Jacob, “Eye movement-based human-computer interaction techniques: Toward non-command interfaces,” In Advances in Human Computer Interaction, vol.4, pp.151–190, 1993.
- [13] V. Vaitukaitis and A. Bulling, “Eye gesture recognition on portable devices,” Proceedings of the 2012 ACM Conference on Ubiquitous Computing ACM, pp.711–714 2012.
- [14] K. Takeda, K. Kise, and M. Iwamura, “Real-time document image retrieval on a smartphone,” Proceedings of the 10th IAPR International Workshop on Document Analysis Systems (DAS2012), pp.225–229, Mar. 2012.
- [15] 工藤 力, 黄瀬浩一, “カメラペンシステムにおける up・down 判定と筆跡位置修正の有効性評価,” 電子情報通信学会技術研究報告, vol.112, no.495, pp.151–156, Mar. 2013.
- [16] H. Drewes and A. Schmidt, “Interacting with the computer using gaze gestures,” Human-Computer Interaction–INTERACT 2007, pp.475–488, Springer, 2007.