平成26年度

筑波大学情報学群情報科学類

卒業研究論文

題目 一般的な紙媒体利用者のための 書き込み認識システム

主専攻 知能情報メディア主専攻

著者 山路 大樹

指導教員 田中二郎 志築 文太郎 高橋 伸三末 和男

要旨

デジタル技術が著しい進歩をみせる中で,デジタルとアナログをシームレスにすることでユーザビリティの向上を目指す研究が盛んに行われている.特に,紙媒体や人間の手によって行われる書き込みとデジタルデータをいかにして結びつけるかとういことは大きな研究分野の一つである.

そこで本研究では,紙の書籍や新聞,論文といった「一般的な紙媒体」に対する書き込みをもとに,様々なデジタル処理を施すシステムを開発した.ユーザはスマートフォンによって本システムを利用し,紙媒体に対しペンで書き込むことで,気になる図,写真,パラグラフなどを自動的に保存したり,図,写真,パラグラフ同士で関連づけたり,英単語の和訳を自動的に表示したりすることができる.また,それらのデータはスマートフォンから閲覧したり,実際に書き込みをしているため,紙媒体からも閲覧することができる.本システムは,スマートフォンとペンさえあれば,あとは紙媒体に対して書き込むという自然な動作に基づきデジタル処理を施すため,普段我々が「一般的な紙媒体」を利用している感覚で本システムを利用することができる.

また,実験では用いるペンや書き込みの種類によっては高い認識率と識別率を示し,本研究における書き込み認識システムの実用性を示した.

目次

第1章	序論	1
1.1	背景	1
	1.1.1 日常をとりまくデジタル技術	1
	1.1.2 デジタルとアナログ	1
1.2	本研究の目的	2
1.3	本論文の構成	2
第2章	一般的な紙媒体利用者のための書き込み認識システム	3
2.1	本システムのアプローチと特徴	3
第3章	本システムの利用手法と処理内容	4
3.1	本システムの利用概要	4
3.2	書き込み認識システムの利用・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・	4
	3.2.1 使用するペンの登録	4
	3.2.2 "「","」"で囲む	6
	3.2.3 同じ文字を書く	6
	3.2.4 で囲む	8
	3.2.5 タッチによる選択	10
3.3	データ閲覧システムの利用	11
	3.3.1 データセット	11
第4章	本システムの実装	15
4.1	書き込み認識システムの実装・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・	15
	4.1.1 既存データの読み込み	15
	4.1.2 書き込み認識システムでのデータ管理	16
	"「","」"で囲まれた領域のデータ管理	16
	で囲まれた領域のデータ管理	16
	4.1.3 書き込みの認識	16
	色の抽出	18
	形状の認識	19
	テンプレートマッチング	19
	4.1.4 それぞれの書き込みによる処理	20

	書き込み"「","」"	20
	同じ文字の書き込み	22
	書き込み	25
	4.1.5 OCR による文字抽出	27
	4.1.6 単語の訳	28
	4.1.7 タッチ検出	28
4.2	データ閲覧システムの実装	29
第5章	想定される利用シナリオ	30
5.1	利用シナリオ 1	30
5.2	利用シナリオ 2	31
第6章	書き込み認識システムの性能評価	32
6.1	実験の目的	32
6.2	実験概要....................................	32
6.3	実験手順	33
6.4	実験結果	33
6.5	考察	35
第7章	関連研究と関連サービス	37
7.1	紙媒体とデジタルデータの連携	37
7.2	書き込みの検出	38
	7.2.1 書き込み検出アルゴリズム	38
	7.2.2 専用ノートの使用	39
7.3	関連商品・サービス	39
第8章	結論と今後の課題	41
	謝辞	42
	参考文献	43

図目次

3.1	本システムのアプリケーションの初期画面から (a) 書き込み認識システムの初	
	期画面 (紙媒体の上にスマートフォンをかざしている場合) への遷移あるいは	
	(b) データ閲覧システムの初期画面への遷移	5
3.2	本システムを利用するユーザ・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・	5
3.3	本システムを作成するにあたり使用したマーカーペン	5
3.4	(a) マーカーペンで色を塗られた紙 , (b) 青色をスマートフォン上で指でタッチ	
	し抽出,(c) 青色の書き込みを認識...........................	6
3.5	(a) パラグラフを "「", "」"で囲むと認識しフィードバック, (b)"「", "」"で	
	囲まれた領域を自動的に保存	7
3.6	(a) パラグラフを "「", "」"で囲むと認識しフィードバック, (b) 既存データな	
	らば保存しない	7
3.7	(a)"「","」"で囲まれた図において"「"の左上に"3"を書くと認識しフィー	
	ドバック,(b)"3"の文字をもつ領域として自動的にグルーピング	8
3.8	(a)"「", "」"で囲まれた図において"「"の左上に"3"を書くと認識しフィー	
	ドバック , (b) グルーピング済みは処理しない	8
3.9	(a) 和訳の分からない英単語を で囲むと認識しフィードバック, (b) その和訳	
	を画面左上に表示・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・	9
3.10	(a) で囲まれた英単語の認識フィードバック, (b) 保存済みダイアログ	9
3.11	(a)"「", "」"で囲まれたパラグラフをタッチ, (b) パラグラフを表示	10
3.12	(a)"「", "」"の左上に文字を含む図をタッチ, (b) その図とそれに関連づけら	
	れたパラグラフを表示,(c)パラグラフを表示	10
3.13	(a) で囲まれた英単語をタッチ, (b) 英単語の和訳を表示	11
3.14	データ閲覧システムの初期画面,(a)画像データ一覧の表示,(b)英単語一覧の	
	表示	12
3.15	(a) 画像データ一覧の表示, (b) 黄色の三角(ここでは(a) の上から3番目の画像	
	データ) を選択し関連データを表示	13
3.16	(a) 英単語一覧の表示 , (b) 対応する和訳 (こここでは (a) の "attractive") を表示	14
4.1	アプリケーション実行時のデータ管理....................................	17
4.2	(a) カメラによる入力画像 , (b) 書き込みが抽出された 2 値化画像	18
4.3	本システムのテンプレートマッチングで用いたテンプレート画像	20

4.4	(a) 2 値画像におけるテンプレート画像"「","」"の左上の座標,(b)適切な位	
	置への座標移動 , (c) 領域の決定 , (d) カメラ画像への適用	21
4.5	領域として認識しない"「"と"」"の位置関係	22
4.6	書き込み"「","」"の基本的なフロー図	23
4.7	(a) 2 値画像における"「"の左上の文字, (b)文字領域の左上の座標, (c) 切り	
	取られた文字領域	23
4.8	(a) 文字を含むと判断される画像 , (b) 含まないと判断される画像	24
4.9	同じ文字の書き込みにおける基本的なフロー図・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・	24
4.10	(a) 過剰な認識, (b) 不足な認識	25
4.11	(a) カメラの入力画像 , (b) を大きめに認識 , (c) エッジ検出による輪郭抽出	
	(水色部分) と外接矩形 (紫色部分), (d) 英単語の切り抜き	26
4.12	書き込み における基本的なフロー図	27
4.13	アプリケーション上と iPhone のストレージ上のデータ関係	29
6.1	(a) ピンク色, (b) 青色, (c) 黄色, のそれぞれ書き込み	33
6.2	(a) 被験者 A , (b) 被験者 B , (c) 被験者 C , によってそれぞれ書き込まれた文字	35

表目次

4.1	テンプレートマッチングで用いられるアルゴリズムメソッド	19
4.2	デ辞蔵における「検索メソッド」において与えるパラメータの例	28
6.1	ピンク色のマーカーペンを用いた時の各被験者に対する実験結果......	34
6.2	青色のマーカーペンを用いた時の各被験者に対する実験結果	34
6.3	黄色のマーカーペンを用いた時の各被験者に対する実験結果	35

第1章 序論

1.1 背景

1.1.1 日常をとりまくデジタル技術

我々は現在,高度な情報社会の中で生活をしている.それは,ムーアの法則によって示されるようなハードウェアの高性能化,充実したネットワークインフラといった著しい進歩をみせるデジタル技術によって支えられている.

デジタル技術が進歩することで我々は様々な恩恵を享受している.ネットワークを介し場所,時間を選ばず検索したり,データにアクセスできる,パソコンやスマートフォンなどの機器に情報をまとめて保存できる,情報が劣化しない,編集しやすい,プリントアウトすることで配布しやすい,携帯性に優れるなど,数え上げるとデジタル技術による恩恵は多岐にわたっていることに気づく.

そして近年は,スマートフォンが開発され一般的に普及したことで,デジタル技術はより 多くの人にとって日常的に欠かせないものとなりつつある.

1.1.2 デジタルとアナログ

デジタル技術が著しい進歩をみせる一方で,アナログの良さや価値を改めて再確認する人も少なくない.例えば,電子書籍と紙の書籍が例として挙げられる.電子書籍の長所といえば,前節で述べたように,大量のデータ(書物)を携帯したり,すぐにアクセスできること,情報が劣化しないことなどが挙げられる.一方で紙の書籍の長所として,ページをめくる動作,把持したときの質感,書き込みのしやすさなどが挙げられる.

そうしたアナログならではの良さを認めつつ,アナログを発達したデジタル技術と結びつけることでユーザにとって使いやすいシステムをつくることができないか,と考えることは自然であり,研究分野の一つでもある.例えば,NUI(ナチュラルユーザインタフェース) や,TUI(タンジブルユーザインタフェース) の分野では,デジタルの特徴に人間が合わせて行動するのではなく,人間にとってより自然な行動をデジタルデータとして取り込み,デジタル処理としてフィードバックする研究がなされている.

先の電子書籍と紙の書籍を例にとるなら、ページをめくる動作や紙に書き込みをするという人間の自然な動作、いわばアナログな部分を、データの保存、アクセスなどの、デジタルな処理につなげることで、アナログとデジタルを結びつけた、よりユーザビリティの高いシステムを構築することができると考える.

1.2 本研究の目的

しかし現在,例えば紙の書籍や新聞,論文を読んでいる際の我々の行動として以下のようなことが考えられる.

- 気になる記事(内容),図,写真などがあったら,ペンで印をつけたり,カメラを持ち出し撮影,スキャンする
- 2. 様々な書き込みがされた紙データや関連する記事 (内容) 同士を分類し, まとめて様々なファイルに保管する
- 3. 調べたいワードがあったらスマートフォンや PC を持ち出し Web 検索する

などである.これらの行動を顧みると,アナログとデジタルをうまく結びつけることができていない(シームレスでない)ということに気づく.

そこで,本研究ではここまで例として挙げてきた,書籍,新聞,論文などを含む,一般的な紙媒体」においてデジタルとアナログを結びつける(シームレスにする)ことに着目し,我々が普段自然に行うであろう「一般的な紙媒体」への書き込みというアナログな動作をトリガーにして様々なデジタル処理を施すシステムを構築することを目的とし,一般的な紙媒体」を普段我々が利用している感覚で,デジタル技術の恩恵を享受できることを目指している.

1.3 本論文の構成

本論文は本章を含め8章で構成されている.第2章では本研究の目的からアプローチと本システムの特徴について述べる.第3章では本システムの利用手法と処理内容について述べ,本システムの基本的な使い方について述べる.第4章では本システムの実装方法について具体的に述べる.そして,第5章では本システムの利用シナリオの述べる.第6章では本システムを用いた評価実験について述べ,第7章で関連研究と関連サービスについて述べ,最後に第8章で本研究の結論と今後の課題を述べる.

第2章 一般的な紙媒体利用者のための書き込み 認識システム

2.1 本システムのアプローチと特徴

前章の1.2「本研究の目的」において述べた,紙の書籍や新聞,論文を読んでいる際の現在の我々の行動の3つの例と照らし合わせ,本研究では,以下の特徴がある.

- 1. 気になる記事 (内容), 図, 写真などに"「","」"を書き込むことによって矩形領域として切り取り, 自動的に保存する
- 2. 1. で書いた"「","」"の左上に文字を書き込むことで共通する文字がある領域を関連づけ,自動的に保存する
- 3. 和訳を調べたい英単語を で囲むことで囲まれた英単語の和訳を表示するとともに,自動的に保存する

実際に書き込みを行うので,普段我々が紙媒体に書き込みをしたときのように,当然書き込みが紙に残る.また,外出時などにおいても,保存されたデータが閲覧できるよう,サーバに自動的にデータを保存したり,スマートフォンなどの端末に保存し,携帯する方法が考えられる.

本研究では、任意の気になる記事(内容)、図、写真や和訳を調べたい英単語などを切り抜くために、手動でカメラを該当位置に配置するように近づける。また、手動でカメラを該当位置に配置する際、カメラを対象物(紙媒体)に対しかざすため、一般的な Web カメラであると対象物(紙媒体)が Web カメラによって覆われてしまったり、どこを映しているのか確認しづらい。そこで本研究では、手動でカメラを任意の場所へ配置しやすく、かつカメラをかざす際にカメラ画像として映している場所がその場でフィードバックされるため、対象物(紙媒体)が覆われても問題のないスマートフォンを用いる。また、本システムによって生成されるデータはスマートフォン上に保存する。

加えて,近頃スマートフォンは非常に身近なものとなり,スマートフォンのみを用いたシステムは,我々の日常生活に対しても導入しやすいというメリットがある.

また,本研究におけるシステムは,特定の書物(例えば,書き込みをトレースするノートやマーカーが貼付された紙など)でなく,我々が普段利用するであろう,紙の書籍や新聞,論文などの「一般的な紙媒体」に対して利用可能であり,そのような「一般的な紙媒体」とスマートフォンと書き込みをするためのペンさえあれば,我々が普段利用している感覚で本システムを利用することができる.

第3章 本システムの利用手法と処理内容

3.1 本システムの利用概要

本システムでは,2通りの利用方法がある.

- 書き込み認識システムの利用
- データ閲覧システムの利用

書き込み認識システムの利用では,ユーザの紙媒体に対する書き込みが認識され,それによるデジタル処理が施される.

データ閲覧システムの利用では,書き込み認識システムの利用によって得られたデータを, 書き込みを行った紙媒体がなくともスマートフォン上で閲覧することができる.

本システムを起動すると,どちらを利用するか選択画面が表示される (図 3.1). カメラアイコンを選択すると書き込み認識システムを利用することができ (図 3.1(a)), ライブラリアイコンを選択するとデータ閲覧システムを利用することができる (図 3.1(b)).

それぞれの利用方法について述べる.

3.2 書き込み認識システムの利用

ユーザは実際に紙媒体に書き込むためのペンを手に持ち,スマートフォンを用いて本システムの書き込み認識システムを利用する(図 3.2).

紙媒体にスマートフォンをかざすタイミングは任意でよく,書き込みながらでも,書き込みが終わった後でもよい.

認識される書き込みは大きく分けて3通り存在し,書き込みによってリアルタイムに得られるデジタル処理を以後それぞれ述べるが,まずここで使用するペンを登録する方法について述べる.

3.2.1 使用するペンの登録

書き込み認識システムは,デフォルトでは図3.3のようなマーカーペン(カラーコード: #FF208B)を用いた場合を想定しているが,ユーザによって使用したいペンは異なるため,任意に使用するペンを登録することができる.

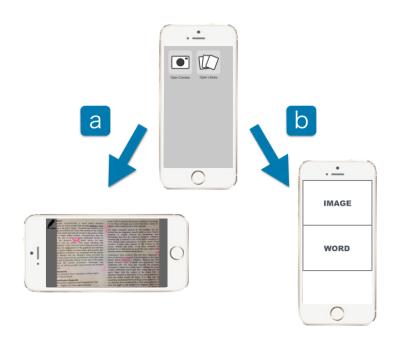


図 3.1: 本システムのアプリケーションの初期画面から (a) 書き込み認識システムの初期画面 (紙媒体の上にスマートフォンをかざしている場合) への遷移あるいは (b) データ閲覧システム の初期画面への遷移



図 3.2: 本システムを利用するユーザ



図 3.3: 本システムを作成するにあたり使用したマーカーペン

ただし,書き込み認識システムはシャープペンのような細いペンによる書き込みを認識することが困難なため,マーカーペンのような太いペンを用いることを想定している(推奨される太さは,1mm以上).

使用したいペンを登録する手順は以下の通りである.

書き込み認識システム中に,紙の切れ端などに使用したいマーカーペンで色を塗る(図 3.4(a)). その後画面左上のペンアイコンをタッチすると,ペンアイコンが黄色くなり,使用したいペンを登録することができる状態になる.そして,使用したいペンで書き込まれた部分をスマートフォン上でタッチすることで,その色を抽出し(図 3.4(b)),今後書き込み認識システムを利用する際に,認識される色として登録される(図 3.4(c)).



図 3.4: (a) マーカーペンで色を塗られた紙, (b) 青色をスマートフォン上で指でタッチし抽出, (c) 青色の書き込みを認識

ではここから,認識する書き込みとそれによるデジタル処理について述べる.

3.2.2 "「", "」"で囲む

気になる図,写真,パラグラフが存在し,保存したいとする.そこで,保存したい領域を"「","」"で囲むことにより,"「"が赤色,"」"が緑色の四角形で囲まれ,ユーザに対して認識していることをフィードバックする(図 3.5(a),図 3.6(a)).そして"「","」"で囲まれた領域は,矩形領域でデジタルデータとして自動的に保存し結果をダイアログで表示する(図 3.5(b)).

もし"「","」"で囲まれた領域が,すでに自動的に保存されている領域ならば,既存データとして扱われ,そのデータは自動的に保存されずその結果をダイアログで表示する(図 3.6(b)).

3.2.3 同じ文字を書く

3.2.2 で書いた"「","」"の左上に文字を書き込むことで,共通する文字がある領域を関連づけることができる.

書き込まれた文字は黄色の四角形で囲まれ,認識していることをフィードバックする(図 3.7(a)).

図 3.7(a) では, "3"の文字が書かれており,同じく"3"が書かれている別の"「", "」"で囲まれた領域と自動的にグルーピングし結果をダイアログで表示する(図 3.7(b)).

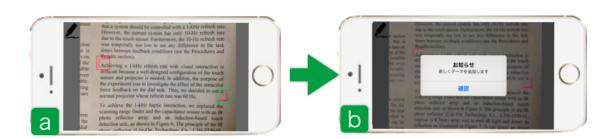


図 3.5: (a) パラグラフを "「", "」"で囲むと認識しフィードバック, (b) "「", "」"で囲まれた領域を自動的に保存

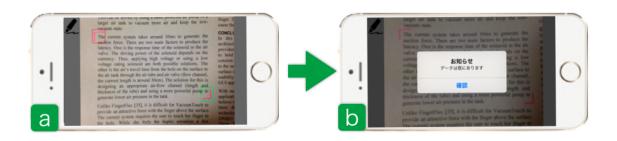


図 3.6: (a) パラグラフを "「", "」" で囲むと認識しフィードバック, (b) 既存データならば保存しない

この機能は,例えば,図とそのコメントをセットで保存したいといった場合などに役立つ.

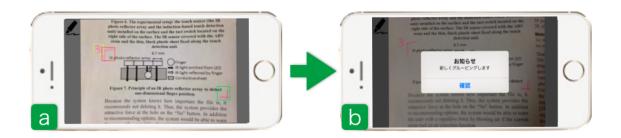


図 3.7: (a)"「", "」"で囲まれた図において"「"の左上に"3"を書くと認識しフィードバック, (b)"3"の文字をもつ領域として自動的にグルーピング

もし"「","」"で囲まれた領域が,既に自動的にグルーピングされている領域ならば,そのデータは自動的にグルーピングされずその結果をダイアログで表示する(図 3.8(b)).

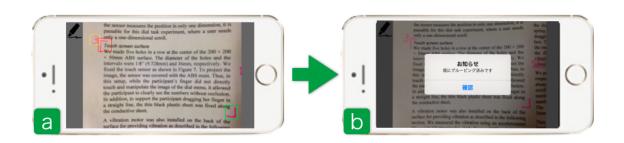


図 3.8: (a)"「", "」"で囲まれた図において"「"の左上に"3"を書くと認識しフィードバック, (b) グルーピング済みは処理しない

3.2.4 で囲む

和訳の分からない英単語があったとする.そこで,該当単語を で囲むことにより, が青色の四角形で囲まれ,ユーザに対して認識していることをフィードバックする (図 3.9(a)).そしてその英単語の和訳を画面左上に表示する (図 3.9(b)).また,一度 で囲った単語は自動的に保存される.

もし で囲まれた英単語が,既に自動的に保存されているならば,そのデータは自動的に保存せずその結果をダイアログで表示する(図 3.10(b)).

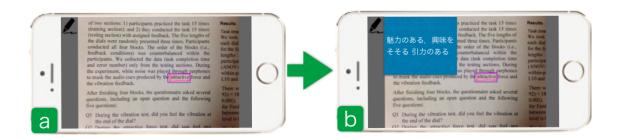


図 3.9: (a) 和訳の分からない英単語を で囲むと認識しフィードバック , (b) その和訳を画面 左上に表示

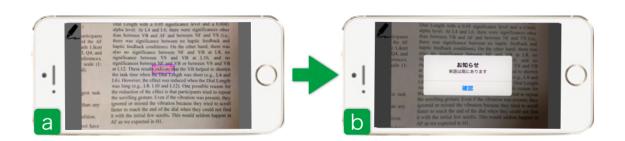


図 3.10: (a) で囲まれた英単語の認識フィードバック , (b) 保存済みダイアログ

3.2.5 タッチによる選択

前節までで述べた書き込みにより形成された領域をスクリーン上でタッチことによって,その領域を選択し,それぞれの領域の応じて対応するデータをスマートフォン上で閲覧することができる.

3.2.2 による"「","」"で囲まれた領域を選択 (図 3.11(a)) したならば,その矩形領域を画面左上に表示する (図 3.11(b)).もし,その領域がまだ保存されていない場合は,まず保存処理を施してから,画面表示を行う.

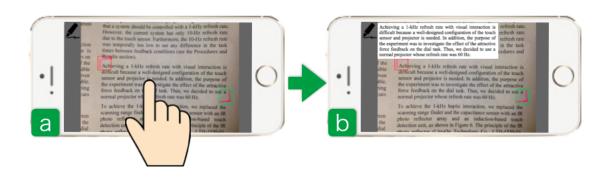


図 3.11: (a)"「", "」"で囲まれたパラグラフをタッチ, (b) パラグラフを表示

3.2.3 による関連づけられた領域のいずれかを選択 (図 3.12(a)) したならば , それに関連づけられた領域を選択された領域を含めて表示する (図 3.12(b)) . そして , 大きく表示したい領域をさらに選択することによって (図 3.12(b)) , その領域のみを表示する (図 3.12(c)) .



図 3.12: (a)"「", "」"の左上に文字を含む図をタッチ, (b) その図とそれに関連づけられたパラグラフを表示, (c) パラグラフを表示

3.2.4 による で囲まれた英単語を選択 (図 3.13(a)) したならば , その単語の訳を画面左上に表示する (図 3.13(b)) . もし . その英単語がまだ保存されていない場合は , まず保存処理をほどこしてから , 画面表示を行う .

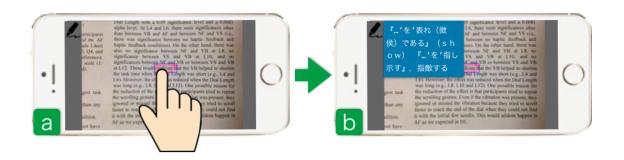


図 3.13: (a) で囲まれた英単語をタッチ, (b) 英単語の和訳を表示

3.3 データ閲覧システムの利用

それぞれの書き込みにより生成されたデータは,スマートフォンに自動的に保存されるので,紙媒体を利用している中で書き込みを加えた,気になる画像やパラグラフ,訳の分からない単語などを,書き込みを施した紙媒体がなくとも,外出時などにスマートフォン上で閲覧することができる.

3.3.1 データセット

データは,図,写真,パラグラフなどの画像データと,英単語とその和訳の文字データにジャンル分けされている(図 3.14). そしてそれぞれ選択することで画像データならばユーザが気になったものとして切り抜いた図,写真,パラグラフなどをまとめて閲覧することができ(図 3.14(a)),英単語ならば和訳のわからなかった英単語をまとめて確認することもでき(図 3.14(b)),対応する和訳を確認することもできる.関連づけられたデータは関連データとしてグループで閲覧することができる.

それぞれの書き込みごとに , データは区別されている . 例えば , "「", "」"で囲まれた図 , 写真 , パラグラフなどは , 保存したいものとしてのデータなので , それぞれのデータを一覧 としてそのまま閲覧することができる (図 3.15(a)) .

同じ文字によって関連づけられた領域 (データ) は, 黄色の三角マークが表示され,選択することでグルーピングされたデータをセットで閲覧することができる(図 3.15(b)).

で囲まれた英単語は,保存した英単語の一覧を表示(図 3.16(a)) し,英単語を選択することでその和訳を閲覧することができる(図 3.16(b)).

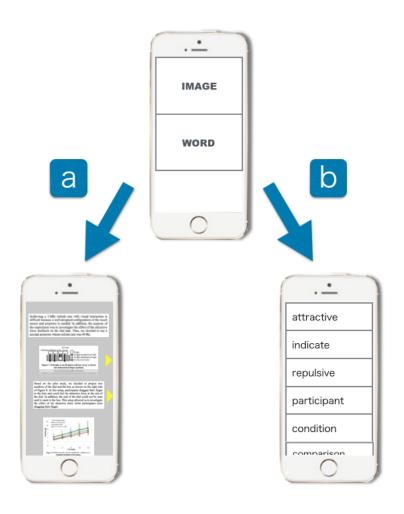


図 3.14: データ閲覧システムの初期画面 , (a) 画像データ一覧の表示 , (b) 英単語一覧の表示

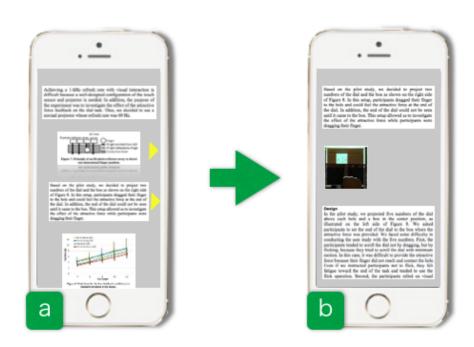


図 3.15: (a) 画像データ一覧の表示 , (b) 黄色の三角 (ここでは (a) の上から 3 番目の画像データ) を選択し関連データを表示

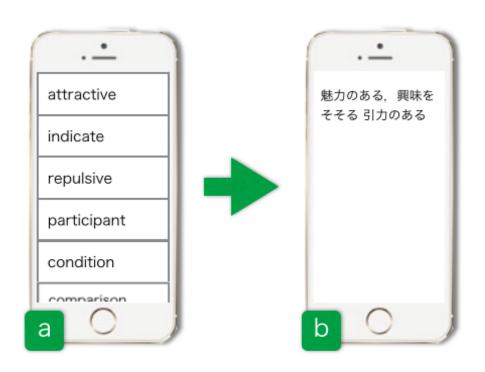


図 3.16: (a) 英単語一覧の表示, (b) 対応する和訳(こここでは(a) の "attractive") を表示

第4章 本システムの実装

前章で述べた通り本システムでは,書き込み認識システムの利用と,データ閲覧システムの利用の2つの利用方法が存在し,それぞれは区別して処理される.

そこで本章では、それぞれのシステムについて実装方法を述べる.

また本システムでは , 書き込み認識や , データ閲覧のためのスマートフォンとして iPhone 5(最大動作速度:1.3GHz , RAM:1GB , OS:iOS8.1.1) を用いた.また , 本システム全体は open-Frameworks というフレームワーク 1 を用い , C++と Objective-C によって実装した.

4.1 書き込み認識システムの実装

書き込み認識システムは大きく分けて3つの処理に分けられる.

- 既存データの読み込み
- 書き込みの認識
- それぞれの書き込みによる処理

では,それぞれの処理を中心に書き込み認識システムの実装について述べる.

4.1.1 既存データの読み込み

書き込み認識システムでは、図、写真、パラグラフ、単語などの諸データを自動的に保存するが、そのデータが以前本システムによって保存されたデータつまり既存データである場合、データの重複が生じる、そこで、データが自動的に保存される前に、そのデータが既存データかどうかを調べる、すなわち、保存しようとするデータと既存データの比較を行う、

そのためには,既存データをアプリケーション内で取得する必要がある.既存データを取得するためには,既存データが格納されているスマートフォンのストレージにアクセスする必要があるが,書き込み認識システムにおいて,書き込みによってデータを自動的に保存しようとする度に既存データとの比較を行うために,逐一ストレージにアクセスしていては,アプリケーションの処理時間のロスにつながる.

¹http://openframeworks.cc/

そこで本システムでは,既存データをアプリケーションから効率的にアクセスするために, 書き込み認識システムを実行する前段階として,既存データをアプリケーションのメモリに すべて読み込む処理を施し,書き込み認識システムの処理速度の向上を計る.

書き込み認識システム内でのデータ管理方法を以下に述べる(図 4.1).

4.1.2 書き込み認識システムでのデータ管理

書き込み認識システムによって保存しようとしているデータと既存データの比較を行うには,それぞれのデータが比較できるような値で一意に定められている必要がある.

そこで書き込み認識システムでは,以下のようにデータ管理を行う.

"「","」"で囲まれた領域のデータ管理

"「","」"で囲まれた領域は ImgDataBase クラスとしてデータ管理する. ImgDataBase クラスは,図,写真,パラグラフなどを,そのまま画像データとして格納するメンバと.画像データを比較するための,画像データ固有の値を格納するメンバから構成されている.

画像データ固有の値として,本システムでは,画像データの縦横比や画像データに含まれている文字数を用いている.画像データに含まれる文字数は後述する OCR によって抽出している.

で囲まれた領域のデータ管理

で囲まれた英単語は WordDataBase クラスとしてデータ管理する. WordDataBase クラスは,英単語とその和訳を格納するメンバから構成されている. 英単語は後述する OCR,和訳は後述する英和辞典 Web サービスを用いて取得している.

4.1.3 書き込みの認識

書き込みを認識するためには,まず書き込みを検出する必要がある.書き込みを検出するとはつまり,書き込みを行う紙媒体と書き込みを区別し,抽出することである.

書き込みを検出する方法としてよく用いられるのは,紙媒体の中での書き込みの局所的位置を検出する方法である.書き込みがされた紙媒体をスキャナで読み取る,あるいはペン先にカメラをつけるなどして,事前に準備された紙媒体と同じ内容の電子データと比較することでその差分を検出する.この方法は人が施すであろう様々な書き込みに対して,高い精度で抽出することが可能であるが,紙媒体と同じ内容の電子データが存在し,利用可能であることを前提にしている.

また,書き込みを検出するためのカメラを固定し,初期状態との差分をとることで書き込みを検出する方法もあるが,カメラを固定するための機構が必要である.

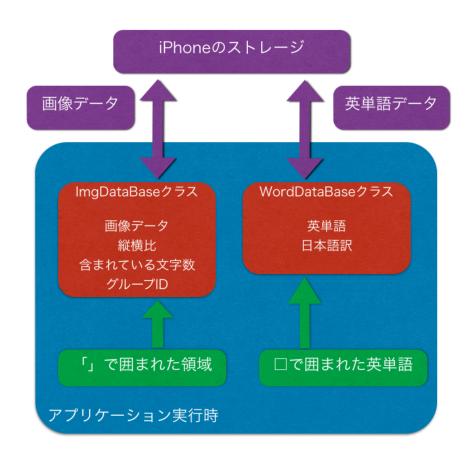


図 4.1: アプリケーション実行時のデータ管理

またマーカーを紙媒体に貼付することで,位置を特定する方法も存在するが,予めマーカーを貼付するなど事前準備が必要である.

本システムでは,一般的な紙媒体つまり,新聞,書籍など必ずしも紙媒体と同じ内容の電子データが存在しない場合においても利用可能であることや,カメラを固定したり,マーカーを貼付するなど事前準備することなしに書き込みを抽出するために,特定の色を抽出することによる書き込みの検出を行う.

色の抽出

本システムの書き込み認識システムでは,初期設定として図3.3のマーカーペンで書き込まれた色(カラーコード:#FF208B)を抽出しているが,ユーザによって書き込みの検出に使用したい色が異なることを考慮し,抽出する色はユーザによって任意に設定できるようにしている.ユーザは,使用したい色のペンで紙に書き込みをし,その書き込み部分をスマートフォン上の画面で触れることで,その色を抽出したい色として設定することができる(本稿3.2.1の通り).

色を抽出する際, RGB 色空間を HSV 色空間に変換して抽出する.これは, HSV 色空間を用い, 色相による閾値を設定することで,明るさに左右されにくい色の抽出ができるからである。

色によって抽出された書き込みは,書き込み部分として紙媒体の元の内容と区別し,次の形状の認識の処理を行うために, 2 値化する (図 4.2).書き込み部分は白,それ以外は黒としている.

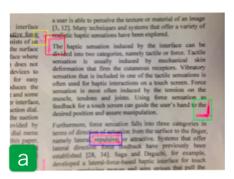




図 4.2: (a) カメラによる入力画像, (b) 書き込みが抽出された 2 値化画像

形状の認識

色の抽出と2値化により書き込みを検出した後,認識する.書き込みを認識するとはつまり,書き込まれたものが何を意味するものなのかを読み取ることである.

本システムでは,書き込まれた形状を検出し,その形状によって施すデジタル処理を変える.本システムで認識する形状は図4.3の通りである.また,スマートフォンからの入力画像をもとにその画像の中に存在する形状を検出,認識するため,テンプレートマッチングを認識アルゴリズムとして用いる.

テンプレートマッチング

テンプレートマッチングとはテンプレートと呼ばれる小さな一部の画像領域と同じパターンが画像全体の中に存在するかどうかをラスタスキャンによって調べる方法である.画像内にある対象物体の位置検出,物体数のカウント,物体移動の検出などによく用いられるアルゴリズムである.

また,テンプレートマッチングにおいて認識に用いるアルゴリズム(計算式)は複数存在し, 以下のようなものがある(表 4.1).

表 4.1: テンプレートマッチングで用いられるアルゴリズムメソッド

メソッド名	意味
CV_TM_SQDIFF	輝度値の差の2乗
CV_TM_SQDIFF_NORMED	CV_TM_SQDIFF を正規化
CV_TM_CCORR	輝度値の相関
CV_TM_CCORR_NORMED	CV_TM_CCORR を正規化
CV_TM_CCOEFF	輝度値の平均を引いてからの相関
CV_TM_CCOEFF_NORMED	CV_TM_CCOEFF を正規化

その中で今回は , CV_TM_CCORR_NORMED を用いる . CV_TM_CCORR_NORMED は次の 計算式 4.1 で与えられる .

T(x, y) をピクセル x, y におけるテンプレート画像の輝度,I(x, y) をピクセル x, y における入力画像の輝度,R(x, y) をピクセル x, y における輝度の相関値を正規化したものとし,テンプレート画像を $w \times h$ としたときに,x' = 0...w-1,y' = 0...h-1 とすると,

$$R(x,y) = \frac{\sum_{x',y'} (T(x',y') \cdot I(x+x',y+y'))}{\sqrt{\sum_{x',y'} T(x',y')^2 \cdot \sum_{x',y'} I(x+x',y+y')^2}}$$
(4.1)

本システムはスマートフォンでの実装であるため,精度は高いが処理の負荷が最も大きい CV_TM_CCOEFF を用いず,負荷が軽減される CV_TM_CCORR を用いた.また,明るさによ

るロバスト性を向上させるため、CV_TM_CCORR_NORMEDを用いた.CV_TM_SQDIFFとCV_TM_CCORRの違いはさほど感じられなかったが、本システムを作成するにあたり若干認識精度が高かったCV_TM_CCORRを用いている.テンプレート画像は図4.3を用いた.また、テンプレートマッチングでは、入力画像におけるテンプレート画像を複数検出することができるが、本システムでは処理の負荷によるユーザビリティの低下と紙媒体の内容がスマートフォン上でも分かるよう、スマートフォンを紙媒体に近づけて利用することを考慮し、それぞれの書き込みは1つのみ認識するようにした.



図 4.3: 本システムのテンプレートマッチングで用いたテンプレート画像

テンプレートマッチングに似た手法として,形状を認識することに特化した,形状マッチングという認識アルゴリズムも存在する.しかしこのアルゴリズムは形状比較に Hu モーメントを用いるため,回転,スケーリング,反転に対して不変量である.本システムは,図4.3(a)と図4.3(b)のように回転に対して不変量では認識アルゴリズムとして不適切なので,今回はテンプレートマッチングを用いる.

4.1.4 それぞれの書き込みによる処理

書き込みを認識したことによって,それぞれの書き込みごとの処理を施すことができる.ここからは,それぞれの処理についての実装方法について述べる.

書き込み"「","」"

書き込み"「","」"は気になる図,写真,パラグラフを保存する場合に書き込まれるものである.本システムでは,"「"を左上,"」"を右下とする矩形領域として"「","」"で囲まれた部分を画像データとして切り取り保存する.

テンプレートマッチングによって,入力画像における"「","」"それぞれの画像の左上の位置を得ることができるので,この位置をもとに,切り取る矩形領域の座標を決定する. 座標の決め方は以下の通りである(図 4.4).

- "「", "」"それぞれの左上の位置を取得する (図 4.4(a))
- "「"の場合,右と下にそれぞれテンプレート画像の 1/3 ずらす (図 4.4(b))

- "」"の場合,右と下にそれぞれテンプレート画像の 2/3 ずらす (図 4.4(b))
- そこから得られた座標を新たな座標として入力画像から切り取る矩形領域に適用する (図 4.4(c), (d))

"「", "」"の位置は 2 値化された画像内における座標であるが,実際に切り取る際には,図,写真,パラグラフを保存するために,カメラによる RGB 画像に対して切り取る座標を適用する.

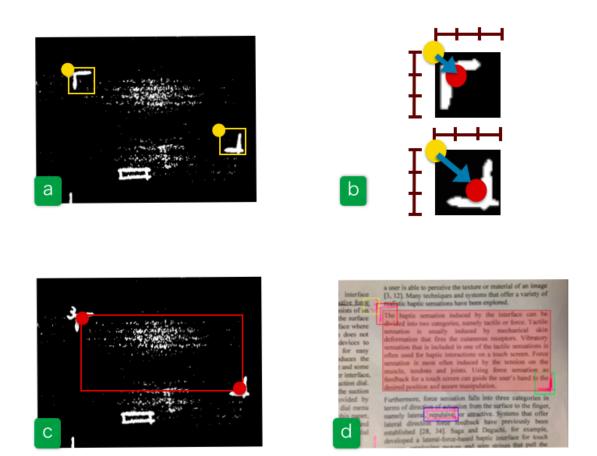


図 4.4: (a) 2 値画像におけるテンプレート画像"「","」"の左上の座標,(b) 適切な位置への座標移動,(c) 領域の決定,(d) カメラ画像への適用

また,入力画像として図 4.5 のようなことが可能性としてあるため,本システムでは"「"が"」"より左に 80px 以上あるいは,上に 80px 以上ある場合のみ切り取り領域の対象とし,画像データ化する.

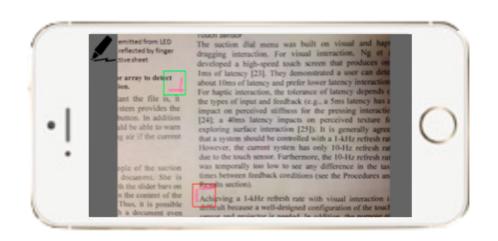


図 4.5: 領域として認識しない"「"と"」"の位置関係

画像データ化された矩形領域は,4.1.2 で述べた ImgDataBase クラスに格納されるが,格納する前に既存データと重複がないかを調べる.

ImgDataBase クラスは画像データそのものと,画像の縦横比と画像に含まれる文字数を格納され,画像データを比較する際にはこの縦横比と画像に含まれる文字数が用いられる.

そして画像に含まれる文字数を取得するために,後述する OCR を用いる.

もし、縦横比が異なれば別の画像として扱われ、ImgDataBase クラスに格納されているまた別のデータと比較する.もし、縦横比が等しく、かつ画像に含まれる文字数も等しい場合は、同じ画像として扱われ、比較操作はそこで終了し新たにデータを格納せず、登録済みダイアログを返す.ここで比較する文字数であるが、OCR の精度によって取得できる文字数に多少ばらつきがあるので、本システムでは前後 10 文字を許容範囲とし、その範囲内ならば等しいとみなす.ImgDataBase クラスに格納されているデータすべて比較し終わった後、同じ画像が見つからなかった場合、保存しようとしている画像を新規データとして ImgDataBase クラス格納する.その際、画像データ、縦横比、画像に含まれる文字数が値として格納される.図 4.6 のフロー図はここでの処理の基本的な流れである.

同じ文字の書き込み

複数の"「","」"で囲まれた領域の"「"の左上に共通する文字を書き込むことでそれらの領域を関連づけることができる.この書き込みによる処理を行うために,書き込まれた文字を認識する必要があり,そのために,書き込まれた文字を含む領域を画像として切り取る.その方法は以下の通りである(図 4.7).

• まず"「", "」"によって囲まれた領域がある(図 4.7(a))

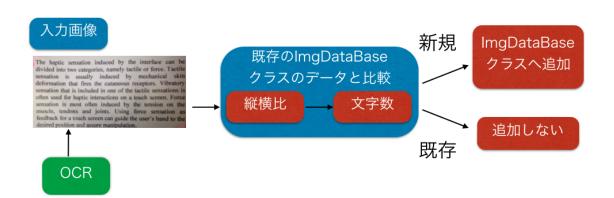


図 4.6: 書き込み"「","」"の基本的なフロー図

● その領域の"「"の左上から左に 20px , 上に 15px 移動した位置を左上とし , 3 文字を含む 30px 四方の矩形領域として切り取る (図 4.7(b) , (c))

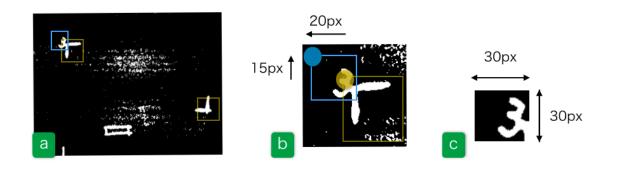


図 4.7: (a) 2 値画像における "「"の左上の文字 , (b) 文字領域の左上の座標 , (c) 切り取られた 文字領域

これらの処理は 2 値画像に対して行われ,矩形領域として切り取る画像は 2 値画像である.そして,この 2 値画像に文字が実際に書かれているのかどうかをまず調べる.本システムでは, 2 値画像における白の割合が 10%以上ならば何かしらの文字が書かれているとみなしている (図 4.8).

文字が書かれていると判断された後,グルーピングの処理に入る.本システムにおいてグルーピングに用いられる図 4.8(a) のような 2 値画像は TmpImgDataBase クラスに格納される. TmpImgDataBase クラスは 2 値画像と,グループ ID をメンバに含む.

グルーピングの処理について説明する、まず書き込まれた文字が以前にも書き込まれたもの



図 4.8: (a) 文字を含むと判断される画像 , (b) 含まないと判断される画像

なのかを調べる.これは,TmpImgDataBase クラスにすでに格納されている2値画像とのテンプレートマッチングによって行う.テンプレートマッチングの計算方法は,CV_TM_CCOERR_NORMEDを用いている.もし,既存の2値画像と同じ画像であると判断した場合は,TmpImgDataBaseクラスとしてその2値画像とセットで格納されているグループIDを取得する.そしてそのグループIDを,"「","」"で囲まれた領域を保存するImgDataBaseクラスのグループIDにセットすることで,グルーピングすることができる.もし,既存の2値画像として判断されなかった場合は,新しい文字として新たなグループIDとともに,TmpImgDataBaseクラスに格納される.なお,グループIDはTmpImgDataBaseクラスのvectorに格納された順番-1がそのまま付与される.すなわち,1番始めに格納された文字のグループIDは0で,5番目に格納された文字のグループIDは4である.

図 4.9 のフロー図はここでの処理の基本的な流れである.

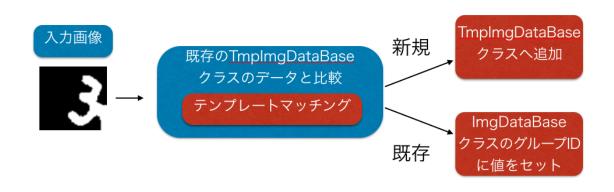


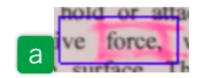
図 4.9: 同じ文字の書き込みにおける基本的なフロー図

書き込み

書き込み は,英単語の和訳を調べたい場合にその英単語に書き込まれる.本システムでは,書き込まれた を用いてその英単語を画像データとして切り抜く.

テンプレートマッチングによって 2 値画像における の左上の座標を得ることができるので、その座標をカメラの RGB 画像に適用することで、 で囲まれた英単語を切り抜くことができる.テンプレート画像 の幅と高さを用いて、矩形として切り抜く.その後、切り取られた画像データを用いて後述する OCR によって画像から文字を抽出する.

しかしながら,テンプレートマッチングによって 2 値画像における の座標を得ることができるが,実際に書き込まれている の大きさを取得することはできない.テンプレート画像 の幅と高さを用いて,矩形として切り抜く場合,過剰に単語を含めてしまったり(図 4.10(a)),逆に切り抜きたい単語が不足してしまう場合がある (図 4.10(b)).



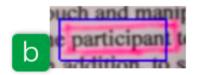
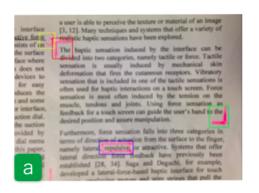


図 4.10: (a) 過剰な認識, (b) 不足な認識

そこで,切り抜きたい英単語の座標はテンプレートマッチングによって取得した後,少し大きめに矩形領域として画像を切り抜く(図 4.11(b)).その後,切り抜かれた画像から実際に書き込まれている を検出するために,実際に書き込まれた の輪郭を検出する.本システムでは,輪郭を検出するためにエッジ検出を用いる.エッジ検出とは,画像の明るさが鋭敏 (不連続) に変化している箇所を特定するアルゴリズムである.このアルゴリズムを本システムに適用することで, が書き込まれた画像の中で,実際に書き込まれた の輪郭を抽出することができる(図 4.11(c) 水色部分).そして,書き込まれた は枠で囲まれているため,輪郭は閉じた図形となる.この閉じた輪郭線に外接する矩形を取得することで,実際に書き込まれた の領域を取得することができる(図 4.11(c) 紫色部分).本システムで用いている openFrameworks に含まれている ofxOpenCV(openFrameworks における OpenCV のアドオン)には,エッジ検出によって輪郭を抽出する ofxCvContourFinder クラスが含まれ,このクラスのメンバである blob に抽出された閉じた輪郭が格納されており,この blob のメンバに外接矩形を取得する関数が含まれている.

OCR によって取得された英単語は,本システムにおいて WordDataBase クラスに格納される. WordDataBase クラスは英単語とその和訳を文字列として格納するメンバを含んでいる. 英単語を取得した後,WordDataBase クラスとして既に格納されている英単語と比較し,もし同じ英単語が存在すれば,登録済みダイアログを返す.もし同じ英単語が存在しなければ,後述する英和辞典 Web サービスを用いて,その英単語の和訳を取得する.その後,英単語と



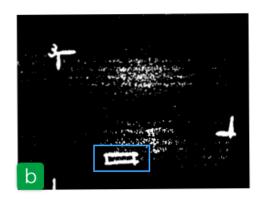






図 4.11: (a) カメラの入力画像 , (b) を大きめに認識 , (c) エッジ検出による輪郭抽出 (水色部分) と外接矩形 (紫色部分) , (d) 英単語の切り抜き

対応する和訳を新たに WordDataBase クラスに格納する. 図 4.12 のフロー図はここでの処理の基本的な流れである.

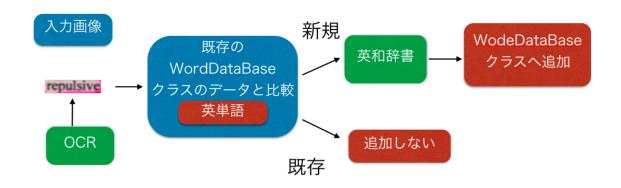


図 4.12: 書き込み における基本的なフロー図

4.1.5 OCR による文字抽出

本システムでは画像データから文字を抽出する方法として OCR(optical character recognition) を用いた.OCR とは手書き文字,印字された文字,画像に含まれている文字などを光学的に読み取り、前もって記憶されたパターンとの照合により文字を特定し、文字データを出力する装置のことである.本システムでは,OCR ライブラリの 1 つである,tesseract-ocr² を用い,その中でも本システムで用いている iOS 用にビルドされたライブラリである tesseract-ios³ を用いた.

このライブラリは,画像を入力することでそこに含まれる文字を返す関数を含むが,背景にノイズが含まれていたり,画像自体がぼやけていたりすると精度が極度に下がってしまう. そこで本システムでは,入力画像に以下のようなフィルタをかける.

- 画像を拡大する
- 白黒にする
- コントラストを強めにする
- 輪郭をはっきりさせる

本システムのように,英単語1語のみを含んだ小さい画像や環境光などの影響で背景が単一色でない画像などを入力画像として用いると精度があまりよいとはいえないので事前に上記に挙げた画像処理を施している.

²http://www.latex-cmd.com/struct/footnote.html

³https://github.com/ldiqual/tesseract-ios

4.1.6 単語の訳

OCR によって抽出された文字を用いて,その英単語の和訳を調べるために,本システムでは,無償の英和辞典 Web サービスである,デ辞蔵を用いた 4 .デ辞蔵は,検索メソッド」と「内容取得メソッド」に分かれており,いずれも指定されたリクエスト URL にアクセスすることで,xml ファイルを取得することができる.

「検索メソッド」では,リクエスト URL にパラメータとして,使う辞書,検索語,検索方法などを指定することができる.本システムでは,EJDICT 英和辞典を使い,単語は完全一致,取得するデータは見出し語で検索した(表 4.2).

そして http://public.dejizo.jp/NetDicV09.asmx/SearchDicItemLite?パラメータ名=パラメータ値&パラメータ名=パラメータ値… のように URL を指定することで, xml ファイルを取得できる「検索メソッド」によって取得した xml ファイルには該当する単語と自分が取得したい情報の ID を得ることができ「内容取得メソッド」のリクエスト URL にパラメータとしてその ID を含めることで,最終的に取得したい和訳を得ることができる.

H172010 05 17 0	IXXX > > 1 1 ICOST C STCOTTS
パラメータ名	パラメータ値の例
Dic	EJdict(EJDict 英和辞典)
Scope	HEADWORD(検索対象を見出し語)
Match	EXACT(完全一致)
Prof	XHTML(コンテンツ形式)
PageSize	20(一度に取得する検索結果の数)

表 4.2: デ辞蔵における「検索メソッド」において与えるパラメータの例

4.1.7 タッチ検出

書き込み認識システムでは,スマートフォンを紙媒体にかざしている際に,スマートフォン上に表示されている紙媒体のタッチ検出を行う.

タッチ検出による処理はスマートフォンのタッチイベントを用いる.タッチされた座標が "「","」"で囲まれた領域内かどうかを座標比較により判断する. で囲まれた領域についても同様にして座標比較を行い,領域内がどうかを判断する.

スマートフォンをタッチすることによる本システムのふるまいは,本稿 3.2.5 の通りであるが,スマートフォン上にデータが表示されている間は,書き込み認識システムはスマートフォンへの負荷の軽減のため一時停止する.そして,再度スマートフォンをタッチし,表示し終わったら書き込み認識システムを再開する.

⁴https://dejizo.jp/dev/rest.html

4.2 データ閲覧システムの実装

書き込み認識システムによって作成した,図,写真,パラグラフ,英単語とその和訳などのデータはアプリケーション上のメモリから iPhone のストレージに保存される.その際はすべてのデータをファイル化する(図 4.13).

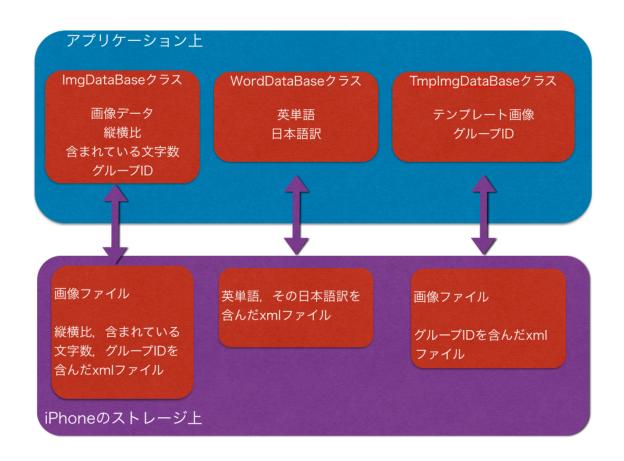


図 4.13: アプリケーション上と iPhone のストレージ上のデータ関係

図,写真,パラグラフは画像ファイルに,英単語とその和訳は xml ファイルとして保存する.画像データに関するメタデータなどについても xml ファイルとして保存する.

データ閲覧システムは,ストレージに保存されたこれらのデータをアプリケーション上に 読み込み表示する.一度読み込みが終わると,画像データと英単語データはそれぞれのジャ ンルとして区別され,グループ ID が共通する画像ファイルは関連づけられた画像ファイルと してまとめる.

そしてデータ閲覧システムにおいて,画像データが選択され,関連づけられたファイルが存在するならばまとめて表示する.英単語データが選択されたならば,対応する和訳を表示する.

第5章 想定される利用シナリオ

この章では,本システムの利用シナリオを述べる.

5.1 利用シナリオ1

日本人の会社員 A を例に挙げる.

仕事で外国人の方とも話す機会の多い A は毎朝,英字新聞を読んでいる.パラパラと新聞をめくっていると気になる記事がいくつかある.出勤時間まであまり時間のない A は,とりあえずスマートフォンを用いて本システムを起動し,気になる記事の上にスマートフォンをかざしながら記事を"「","」"で囲ってゆく.すると,それらは自動的にスマートフォンに保存されていく.

記事を読んでいる中で和訳の分からない英単語も多く見受けられた. Web 上で一つ一つ調べるには時間と手間がかかるが,本システムを利用している A は英単語を で囲むだけで,和 訳を確認することができるとともに,その英単語と和訳はスマートフォンに保存されていく.

通勤電車に乗った A は,この空いている時間に,先ほど気になるものとして保存した記事をざっと眺める.紙の新聞を持ち運ぶ必要がないので,荷物がかさばらないし,自分が気になる記事だけを集中して閲覧することができる.

記事を読んでいる中で和訳の分からなかった英単語もスマートフォンに保存されているので, まとめて学習することもできる.

仕事から帰った A は,紙の新聞をもう一度読み直す."「","」"が書き込まれている記事は自分にとって興味深い記事であるため,まずその周辺を中心に読んでいく.すると,"「","」"で囲った記事 X1 に関連した写真 X2 があることに気づいた.そこで,写真 X2 を"「","」"で囲い,左上に"X"の文字を書く.その後,"「","」"で囲まれている記事 X1 の左上に同じく"X"の文字を書く.するとそれらは関連データとしてセットでスマートフォンに保存される.

"「","」"で囲まれた記事以外の場所を読んでいると,通勤電車の中で読んだ"「","」"で囲まれた記事と似ている記事があった.そこでAは"「","」"で囲まれた記事をスマートフォンでかざし,その記事の部分をスマートフォン上でタッチする.するとその記事がスマートフォンに表示される.そして先ほど似ていると思われた記事に戻り,スマートフォン上では"「","」"で囲った記事を,紙の新聞では似ている記事を見ることで,両者を見比べながら記事を閲覧することができる.

新聞を読み直している中で,和訳が分からないとして, で囲った英単語をまた忘れてし

まった.しかし,A はその英単語をスマートフォンでかざし,タッチすることですぐに和訳 を確認するすることができる.

このようにして,効率的に英字新聞を活用できる A は同僚の中でも優秀な業績をあげた.

5.2 利用シナリオ2

情報系大学生 B を例に挙げる.

B は変化の著しい情報技術の最先端を学ぶため,日頃から多くの英語論文を読むとともに, 英語試験で好成績をとるために,英語の学習もしている.

多くの英語論文を読むために,Bはまず論文1のAbstractを"「","」"で囲む.そして,その左上に"1"の文字を書く.論文を眺めながら,写真や図がある場合は,その論文を直感的に理解するのに役立つと思い,それらも"「","」"で囲い同じく左上に"1"の文字を書く.また,キーワードとなりそうなパラグラフにも同様の書き込みを行う.それらはセットでスマートフォンに保存される.

B はまた別の論文 2 の Abstract を "「", "」"で囲むと同時に, その左上に "2"の文字を書き, その論文に含まれる写真や図, キーワードとなりそうなパラグラフに同様の書き込みを行う.

この一連の書き込み行為は非常に簡単なので , B は 50 個の論文に対して同じ書き込み行為を行った .

その後,Bはスマートフォンで先ほど保存したAbstactとセットで保存された写真,図,パラグラフなどを読み進めていく.そして興味深いと感じた論文だけチェックし,後でもう一度紙の論文を読み直す.

このように , 1 つの論文に対する取っ掛かりの敷居が低いことや , スマートフォンをいつでも持ち歩いているため , いつでも論文の確認をすることができる B は多くの論文をこなすことができた .

また,紙の論文を読んでいる中で,和訳の分からない英単語が存在したら, で囲ってい く.その場で和訳を確認できるとともに,英単語と和訳はスマートフォンに保存されていく.

英語の学習もしている B は和訳の分からなかった英単語をスマートフォンでまとめて確認することができるため,非常に学習効率がよい.また,紙の論文には実際に が書き込まれているので,その英単語が文脈でどのように使われているのかを把握することができるとともに,和訳は書かれていないため,自分の頭で思い出すことができる.もし,どうしても和訳が分からないのであれば,スマートフォンでかざし,タッチすることによって確認することができる.

英語の学習も効率的に行うことができた英語試験で好成績をあげた.

第6章 書き込み認識システムの性能評価

本章では,本研究における書き込み認識システムの性能評価を行うための予備実験について述べる.

6.1 実験の目的

本研究「一般的な紙媒体利用者のための書き込み認識システム」では,以下の機能をもつ.

- 1. 気になる記事 (内容),図,写真などに"「","」"を書き込むことによって矩形領域として切り取り,自動的に保存する
- 2. 1. で書いた"「","」"の左上に文字を書き込むことで共通する文字がある領域を関連づけ,自動的に保存する
- 3. 和訳を調べたい英単語を で囲むことで囲まれた英単語の和訳を表示するとともに,自動的に保存する

そこで予備実験では,本研究において提案する書き込み認識システムがどれほどこれらの機能を達成しうるものかを調べることを目的とする.

6.2 実験概要

前節にあげた1.~3.の3つの機能に則り,以下の予備実験を行う.

- 1. "「", "」"の認識率と保存領域の識別率
- 2. "「", "」"の左上に書かれた文字の識別率
- 3. の認識率と保存領域(英単語)の識別率

また,3種類のマーカーペンを用い,それぞれに対して上記の1.~3.を実験する.書き込みに用いる紙媒体として英語で書かれた論文を用いる.実験は蛍光灯による照明環境の室内で行う.予備実験には21~22歳の大学生3人に協力してもらう.

6.3 実験手順

実験は一人ずつ行う.

まずピンク色 (カラーコード: #FF208B) のマーカーペンで以下の実験を行う.

- 1. 気になる図,写真,パラグラフを6つ保存する
- 2. 保存したらその領域に再度スマートフォンをかざす
- 3. 1. で書いた6つの"「","」"のうち,3つの"「","」"の左上に任意の共通する文字,別の3つの"「","」"の左上に任意の共通する文字,計2種類の任意の文字を書き,2種類のグループを作る
- 4. 和訳の分からない英単語を5つ保存する

この 1. ~ 4. のタスクを青色 (カラーコード:#0664BB), 黄色 (カラーコード:#BBFF4D) のマーカーペンを用いた場合について行う. ただし,マーカーペンの色による違いを明確にするため,3. で書き込む文字は被験者ごとには変わるが,マーカーペンごとには変わらない.

以上のタスクを一人当たりのタスクとし,計3人の被験者に取り組んでもらった.

図 6.1 は,実験で書き込みに用いられた3種類のマーカーペンの色である.







図 6.1: (a) ピンク色, (b) 青色, (c) 黄色, のそれぞれ書き込み

6.4 実験結果

前節実験手順 1.~4. における,マーカーペンと被験者ごとの結果を記す. 実験結果は以下の項目を求めた.

- "「", "」"認識率:1.において書き込み"「", "」"により図,写真,パラグラフを保存する確率:つまり,気になる図,写真,パラグラフを6つ中どれだけの割合で保存できるかである.これは,書き込み"「","」"のテンプレートマッチングとしての認識率を計測している.
- "「", "」"識別率: 2. において図, 写真, パラグラフを保存した領域として判別する確率: つまり, 既に保存した領6つ中どれだけの割合で既存データと判別できるかである. これは, 本研究における既存データの一意化の精度を計測している.

- 文字識別率:3.において任意の書き込みによりグルーピングできる確率:つまり,6つの"「","」"で囲まれた領域に対し,それぞれ左上に文字を書く中でどれだけの割合で3.で指示した通りのグルーピングができるかである.例えば,領域A,B,C,D,E,Fが存在し,A,B,Cを一つのグループ,D,E,Fを一つのグループとし,2種類のグループを作りたいときに,A,Cが一つのグループ,D,E,Fが一つのグループ,Bがまた別のグループと判断されたら,文字識別率は5/6で83%とする.これは,任意の文字の書き込みに対しテンプレートマッチングを用いた本研究の認識手法の精度を計測している.
- 認識率:4.において書き込み により英単語を切り取る確率:つまり,和訳の分からない英単語を5つ中どれだけの割合で保存できるかである.これは,書き込み のテンプレートマッチングとしての認識率を計測している.
- 識別率:4.において切り取られた英単語の和訳を正確に表示する確率:つまり,和訳の分からない英単語を5つ中どれだけの割合で正確に英単語を抜き出せるかである.これは,本研究におけるOCRや,エッジ検出による英単語の切り抜き手法の精度を計測している.

以下の表が結果である.

表 6.1: ピンク色のマーカーペンを用いた時の各被験者に対する実験結果

	被験者 A	被験者B	被験者C
"「","」"認識率	100%	100%	100%
"「","」"識別率	83%	100%	100%
文字識別率	100%	67%	83%
認識率	100%	100%	100%
識別率	60%	80%	40%

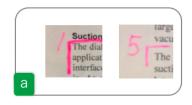
表 6.2: 青色のマーカーペンを用いた時の各被験者に対する実験結果

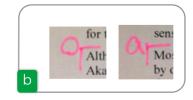
	被験者 A	被験者 B	被験者C
"「","」"認識率	100%	100%	100%
"「","」"識別率	100%	83%	100%
文字識別率	100%	67%	83%
認識率	100%	100%	100%
識別率	40%	60%	60%

また,実験手順(2.1)によって書き込まれた任意の文字は図6.2の通りである.

表 6.3: 黄色のマーカーペンを用いた時の各被験者に対する実験結果

	被験者 A	被験者 B	被験者C
"「","」"認識率	33%	33%	17%
"「","」"識別率	33%	33%	17%
文字識別率	33%	17%	0%
認識率	20%	40%	40%
識別率	0%	20%	0%





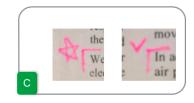


図 6.2: (a) 被験者 A, (b) 被験者 B, (c) 被験者 C, によってそれぞれ書き込まれた文字

6.5 考察

"「","」"認識率と"「","」"識別率について各表を見比べると,表 6.1 と表 6.2 のピンク色と青色のマーカーペンを用いた場合は高い認識率と識別率であるが,表 6.3 は低い認識率であることが分かる.これは,今回実験で書き込みを行う紙媒体として用いた論文の背景色が,ピンク色や青色と比べると黄色が近かったことに起因すると考えられる.

よって,本システムを用いる際には書き込みを行う紙媒体の背景色に対し,色が十分異なるマーカーペンを用いる必要があることが分かる.色による書き込みの認識や識別を行っているため,やむを得ない部分でもあるが,ユーザにとって任意のペンを用いることを目指すにあたっては今後の課題となりうる,

しかしながら,紙媒体の背景色とは十分に異なるマーカーペンを用いることで,書き込み"「","」"や書き込み の認識率や識別率,また,文字識別率も高い確率であることが分かる.

文字識別率については,被験者 A の書き込み (図 6.2(a)) は 100%であるが,被験者 C ,被験者 B の順で識別率が下がっている.これは,被験者 C の書き込みは図 6.2(c) の星形のように,少々複雑な構造の文字を用いたため,ユーザの書き込みによる再現性が低いからであると考えられ,被験者 B の書き込みは図 6.2(b) のように似た文字であるため,識別が困難になったと考えられる.

単純な文字や似た構造でない文字ならば,高い識別率を出すことができるという成果はあげられたが,似た文字や複雑な書き込みに対しても,識別率をあげる工夫を施す必要があると思われる.

最後に 識別率について,これまで認識率や識別率の高かった表6.1と表6.2のピンク色と

青色のマーカーペンを用いた場合でも、識別率が高いとはいえない結果となった(表 6.1 と表 6.2 における被験者 ABC のすべての文字識別率の平均は60%となる).

やはり論文などの紙媒体をスマートフォンのカメラで取得しOCR するため,文字の輪郭が 鮮明でなかったり,文字自体が小さかったり,本研究のように書き込みを用いる場合,書き 込みが画像に映り込む,といった問題点が存在する.それに対し,本研究では,エッジ検出 をしたり,取得画像に対し,様々なフィルターをかけるなど,識別率をあげるための多くの プロセスを経たがこの結果にとどまった.

本システムを実用化するためには,今後はこの 識別率をあげるために,識別率をあげる ためのプロセスに工夫を加えたり,別のアプローチから取り組む必要があるかもしれない.

しかしながら 、 を書き込むことによって , およそ 60%で紙媒体から英単語を取得することができるということは , 一般的な紙媒体に対する書き込みを用いたシステム構築への今後の礎になりうるといえる .

第7章 関連研究と関連サービス

7.1 紙媒体とデジタルデータの連携

本研究の主題でもあるデジタルとアナログをいかにシームレスにするか,というテーマの中でも,本研究とも関連の深い,紙媒体とデジタルデータの連携という点に着目した研究は多く存在する.

Koike ら [1] は,机上にマーカーが貼付された書籍などの紙媒体を置くことで,その付近にデジタル情報をプロジェクションし,デジタル情報と実世界の連携を可能にするシステムを開発した.このシステムでは,机の上部にカメラとプロジェクタを設置し,書籍に貼付されたマーカーを認識することで,対応するデジタルデータを机上に投影するとともに,プロジェクションされた物体とのインタラクションも可能である.

Do-Lenh ら [2] は, Koike らの研究をさらに改善し,持ち運び可能なシステムにするとともに,マルチユーザ,マルチフィンガに対してインタラクション可能なシステムを開発した.

これらのシステムは紙媒体とデジタル情報を連携させるという点で本研究と似ているが,事前に紙媒体などにマーカーを貼付する必要があること,対応するデジタルデータを準備する必要があるという点で本研究と異なる.

また, Yoon ら [11] は, pdf などの電子データに対して, 手書きで注釈をつけるためのシステムを開発した.このシステムでは, 通常の pdf などの電子データの行間などに, 新たに余白スペースを動的に追加することで, ユーザが任意のタイミングで手書きで注釈を加えたり, 閲覧したり, 閉じたりすることができる.

Harrison ら [12] は,タブレット端末などへの触れ方,操作の仕方によって,タブレット端末上であたかも電子データに対して,ペン文字を書いたり,消しゴムで消したり,写真 (スクリーンショット) を撮ったりすることができるシステムを開発した.

これらのシステムは,文字を書いたり,消しゴムで消すなど,普段我々が実世界においてとるであろう自然な行為を用いることが,我々人間にとってユーザビリティを高めることにつながるという考えに基づいている.

本研究は,新聞や書籍,論文に対する書き込みといった,普段の我々の実世界における自然な行為をデジタル処理につなげるという点で,似ているといえる.

また, Sangsubhan ら [13] は,紙などに書いたものが自動的にデジタル化されるデジタルペンを用いることで,書かれた内容をもとに発想支援を行うシステムを開発した.このシステムでは,ユーザによって手で書かれたものをデジタル化した後,そのデータをどのように活用するか,という点に着目している.

本研究は,デジタル化された英単語を翻訳し,後で学習できるようにしたり,関連データ

をまとめて閲覧できるようにしたりするなど,デジタル化されたデータをユーザにとって利用したい情報のリソースにするという点で,似ているといえる.

7.2 書き込みの検出

本研究におけるシステムにおいて実装したような,書き込みの検出に関する研究はこれまで盛んに行われてきた.ここで,多くの方法が存在する書き込みの検出方法について述べる.

7.2.1 書き込み検出アルゴリズム

手書き文字認識アルゴリズムは大きく分けて,オフライン文字認識とオンライン文字認識の2つに分けられる.オフライン文字認識とは紙に書かれている文字を判別する技術であり,紙媒体をカメラやスキャナなどで読み取り,文字認識する.オンライン文字認識とは画面に入力した手書き文字を判別する技術であり,タッチパネルなどに専用のペンを用いて書かれた文字を認識する.

本研究では,紙媒体に書かれた文字を認識するという点でオフライン文字認識といえる. オフライン文字認識のためのアルゴリズムの研究は多く存在する.

Nakai ら [9] は,紙媒体の中での書き込み局所的位置を検出することで,書き込みを元データから分類し,抽出する方法を提案した.書き込みがされた紙媒体をスキャンしてから色クラスタごとに分類し,色クラスタごとに紙媒体と同じ内容の電子データとの特徴量の対応点をみつけるという位置合わせを行い,差分処理を行うことで,高精度に書き込みを抽出することを可能にした.

Iwata ら [10] は,ペン先に小型カメラを取り付けることによって,スキャンすることなくリアルタイムに書き込みの局所的位置を検出し,元データとの差分処理を行うことで,書き込みを抽出することを可能にした.

Mazzei[14] は , LLAH(Locally Likely Arrangement Hashing) という高速かつ高精度で , 高いロバスト性をもつという特長をもつ , ハッシュに基づく文字画像検索法を利用し , DBSCAN(Density-Based Spatial Clustering of Application with Noise) を用いて , ハイライトされた文 , 文書の余白に書かれた注釈 , 行間に書かれた注釈などにクラスタリングし , これらの書き込みを抽出することを可能にした .

しかしながら,これらは紙媒体と同じ内容の電子データが必要であることや,書き込みの 認識の処理が膨大であり,画像などに対しては有効であるが,動画など速度が求められるも のに対しては適切ではない.

また, Stevens ら [17] は,書き込みを用いた自動校正といった,抽出された書き込みの利用に重点を置いたシステムを開発し,書き込みに用いる色を限定することで,高精度に書き込みを抽出するシステムを開発した.しかしながら,このシステムでは書き込みの色に制限がある.

また, Guoら [15] や Zhengら [16] は,画像における連結成分を手書き文字と印刷された文字に分類するシステムを開発し,書き込みがされた紙媒体のみで書き込みを抽出することを可能にした.しかしながら,このシステムでは抽出できる書き込みは文字に限られ,手書き線や図形などを抽出することができない.

本研究におけるシステムは新聞,書籍など,普段我々が利用している一般的な紙媒体に対して,普段利用しているような感覚で利用できることを目指しているため,書き込みを行う紙媒体と同じ内容の電子データを必要とせず,利用している紙媒体に元から存在する文書情報を利用するという点,つまり実際に書き込まれている文字を検出するわけではないという点,スマートフォンというPC などの計算機などと比べてスペックのあまり高くないスマートフォンにおいて書き込み検出を行うという点で,色情報のみを用いた書き込み検出を行った.また,検出する色はユーザによって任意に設定できるという点で,より柔軟性のあるシステムである.

7.2.2 専用ノートの使用

オンライン文字認識に関連して, Brandl[3] らは, 専用のノートとペンを用い, 書き込みを自動的にデジタルデータ化するシステムを開発した.このシステムでは, デジタルデータ化した後, デジタルデータに対し矩形で色付けしたり, ドッグイアをつけたりなど, 手書きで書き込んだ情報を活用する手段も提供している.

本研究は,書き込みをデジタルデータとして活用するという点で似ているが,専用のノートなどを使わず,普段我々が利用している「一般的な紙媒体」が対象であるという点で異なる.

7.3 関連商品・サービス

専用のノートとペンを用いて書き込みをデジタルデータ化するという点で関連商品もある. Livescribe3 スマートペン [5] はまさに専用のノートに専用のペンで書き込むことで自動的に デジタルデータ化できる商品であり,iOS7 以降に対応したiOS アプリケーションである.

アノトペン [6] はペン型デバイス単体で書き込み認識ができる商品であるが , 紙側に独特な 微細模様が印刷されている必要がある .

airpen[4] は , 専用のデバイスを紙媒体の前方に配置することでどのような紙に対しても , 書き込みを検出することができる .

本研究は,まったく新しい書き込みを検出するというよりは,既存の紙媒体の文書内容に対する処理を行うので,システムの対象が少し異なる.

また,クイックショナリー [7] は,紙媒体に対しなぞる操作をすることで英単語の和訳をペンに表示するペン型システムである.紙媒体における英単語の和訳を取得するという点で本研究と同じであるが,本研究は実際に書き込みを行うという点で異なり,再度同じ紙媒体を見たときに,書き込みが残っており自分がどの英単語に着目したかをすぐに把握しやすい.この特徴も,英文書を読む際,訳の分からない単語にはチェックを入れるという,普段我々がと

るであろう自然な行為に基づいており,普段我々が紙媒体を利用している感覚で本システム を利用することができる.

第8章 結論と今後の課題

本研究では,紙の書籍や新聞,論文といった「一般的な紙媒体」に対する書き込みをもとに,様々なデジタル処理を施すシステムを開発した.ユーザはスマートフォンによって本システムを利用し,紙媒体に対しペンで書き込むことで,気になる図,写真,パラグラフなどを自動的に保存したり,図,写真,パラグラフ同士で関連づけたり,英単語の和訳を自動的に表示したりすることができる.また,保存したデータや関連づけたデータ,英単語は外出時などにスマートフォンから閲覧することができるし,実際に書き込みをしているため,紙媒体からも閲覧することができる.

このシステムを用いることで,本稿5章の利用シナリオでも述べたように,普段我々が利用している「一般的な紙媒体」をより効率的に活用できるようになるとともに,紙媒体そのものの価値を高め,紙媒体を利用するユーザの生活の質を向上させることにもつながるであるう.また本システムは,スマートフォンとペンさえあれば,あとは紙媒体に対して書き込むという自然な動作に基づきデジタル処理を施すため,普段我々が「一般的な紙媒体」を利用している感覚で本システムを利用することができるという点で,利用するための敷居が低く,より多くの人にとって使いやすいシステムとなるであろう.

また,予備実験で示した通り,本システムでは書き込みの高い認識率や識別率を示し,気になる図,写真,パラグラフやデータの関連づけなど,一般的な紙媒体利用者にとって,普段の利用シーンの中において補助となるようなシステムとなりうる.

しかしながら,英単語の抽出という点においては,その精度にまだまだ課題が残ることも同時に示し,精度向上のために,処理方法に工夫を加えたり,用いるアルゴリズムを変えるなど,別のアプローチから取り組む必要があるかもしれない.

また,今回はマーカーペンのようなやや太めのペンや,紙媒体の背景色と十分に異なる色のペンを用いる必要があるなど,やや柔軟性に欠ける部分もあった.本研究では,一般的な紙媒体利用者をターゲットに,普段そうした紙媒体を利用している感覚で利用できることを目指しているため,システムありきで利用条件を限定してしまうのでなく,あくまで普段の利用シーンの中で利用できるシステムを追求することは本研究に引き続き今後も継続すべき課題である.

また,本システムでは英語の紙媒体のみを対象物としているため,どの言語(少なくとも母国語である日本語)に対しても利用でき,より実用性のあるシステムを構築していくことも,課題として挙げておきたい.

謝辞

本研究に取り組むにあたり,指導教員である田中二郎先生,志築文太郎先生,高橋伸先生, 三末和男先生には丁寧なご指導及びご助言をいただきました.特に田中二郎先生には,研究 内容に限らず,研究の進め方,論文執筆に関すること,研究生活における心構え,備品のご 提供など,多くのご指導とご協力をいただきました.心から感謝申し上げます.

インタラクティブ・プログラミング研究室の皆さまには,ゼミや日頃の研究生活において多くのご意見やご助言をいただきました.特に NERF チームの皆さまには,チームゼミやグループミーティングを通して多くの貴重なご意見をいただきました.心から感謝いたします.

最後に,私の生活を支えて下さった両親や叔父,本研究をご支援くださった皆様に心から 感謝申し上げます.

参考文献

- [1] H. Koike, Y. Sato and Y. Kobayashi. Integrating Paper and Digital Information on Enhanced-Desk: A Method for Realtime Finger Tracking on an Augmented Desk System. In *ACM Transactions on Computer-Human Interaction, Vol. 8, No. 4, December 2001*, pp. 307-322, 2001.
- [2] S. Do-Lenh, F. Kaplan, A. Sharma and P. Dillenbourg. MultiFinger Interactions with Papers on Augmented Tabletops. In *Proceedings of the 3rd International Conference on Tangible and Embedded Interaction*, pp. 267–274, 2009.
- [3] P. Brandl, C. Richter and M. Haller. NiCEBook: Supporting Natural Note Taking. In *CHI* ' 10: Proceedings of the 28th international conference on Human factors in computing systems, pp. 599–608, 2010.
- [4] airpen. http://www.airpen.jp/ 2012.
- [5] Livescribe 3 スマートペン. http://www.livescribe.com/jp/smartpen/ls3/2014.
- [6] アノトペン. http://www.anoto.com/ 2015.
- [7] クイックショナリー. http://www.scanpen.jp/ 2008.
- [8] 株式会社富士通研究所. 指で直観的に操作可能な次世代ユーザーインターフェースを開発. http://pr.fujitsu.com/jp/news/2013/04/3.html, 2013, 4.
- [9] T. Nakai, K. Kise, and M. Iwamura. A method of annotation extraction from paper documents using alignment based on local arrangements of feature points. In *Document Analysis and Recognition, ICDAR2007. Ninth International Conference on, vol. 1*, pp. 23–27, 2007.
- [10] K. Iwata, K. Kise, M. Iwamura, S. Uchida, and S. Omachi. Tracking and retrieval of pen tip positions for an intelligent camera pen. In *Proceedings of ICFR2010*, pp. 277–282, 2010.
- [11] D. Yoon, N. Chen, F. Guimbretire. TextTearing: Expanding Whitespace for Digital Ink Annotation . In *Proceedings of UIST '13*, pp. 107–112, 2013.
- [12] C. Harrison, R. Xiao, M. Iwamura, J. Schwarz, and S. E. Hudson. TouchTools: Leveraging Familiarity and Skill with Physical Tools to Augment Touch Interaction. In *Proceedings of CHI '14*, pp. 2913–2916, 2014.

- [13] P. Sangsubhan, J. Tanaka. Idea generation support system utilizing digital pen and paper . Master Thesis, University of Tsukuba, 2013.
- [14] A. Mazzei. Extraction and Classification of Handwritten Annotations for Pedagogical Use . In *Proceedings of EDIC2009*, 2009.
- [15] J. K. Guo and M. Y. Ma. Separating Handwritten Material from Machine Printed Text using Hidden Markov Models . In *Proceedings of 6th international Conference on Document Analysis and Recognition*, pp. 436–443, 2001.
- [16] Y. Zheng, H. Li, and D. Doermann. The Segmentation and Identification of Hand writing in Noisy Document Images. In *Lecture Notes in Computer Science (5th International Workshop DAS2002)*, vol.2433, pp. 95–105, 2002.
- [17] J. Stevens, A. Gee, and C. Dance. Automatic Proceedings of Document Annotations . In *Proceedings of 1998 British Machine Vision Conference*, vol.2, pp. 438–448, 1998.
- [18] C. Yi and Y. Tian. Text extraction from scene images by character appearance and structure modeling. In *Proceedings of CVIU2013*, pp. 182–194, 2013.
- [19] R. Huang, P. Shivakumara, and S. Uchida. Scene Character Detection by an Edge-Ray Filter. In *Proceedings of ICDAR2013*, pp.462–466, 2013.
- [20] A. Jain, J. Sharma. Classification and interpretation of characters in multi-application OCR system . In *Proceedings of ICDMIC2013*, pp.1–6, 2013.