情報処理学会論文誌 Vol. 52 No. 4 1438–1452 (Apr. 2011)

推薦論文

マーカレス拡張書籍のための2次元文字ブロック検索手法

宮田章裕[†1] 塩原寿子[†1] 藤村考[†1]

本論文では，文章を読む方向とそれに直交する方向を考慮した2次元のブロックを索引・検索のキーとする2次元文字ブロック検索手法を提案し，書籍内の局所領域にデジタルコンテンツへのハイパーリンク設置を可能にするシステムKappanを紹介する．従来，書籍内にハイパーリンクを設置する際はマーカを用いる方法があったが，この手法はあらかじめ書籍内にマーカを記載する必要がある．一方，システム上に書籍内のテキストと位置を関連付けておけばマーカは不要である．すなわち，書籍内を撮影した画像をシステムに送信すれば，システムは画像からOCR（Optical Character Recognition）により抽出したテキストを検索語として位置を特定し，その位置に関連付けられたコンテンツを提示できる．このとき，大量の書籍の中から一意に位置を特定するためには，長く連続するテキストを検索語とする必要がある．ところが，一般ユーザが撮影した画像にはOCR誤認識が約35%発生するため，長いテキストには誤認識文字が含まれて正しく検索できないという問題があった．特に書籍内の局所領域からは抽出できる検索語数が少なく，この問題は深刻である．提案手法は少ない文字数で書籍内の各局所領域に固有なパターンを表現できるので，OCR誤認識が発生する書籍内の局所領域画像から一意に位置を特定できる．73,231文書から局所領域画像を含む文書を一意に特定する検証実験では，提案手法はノイズがない状態で99%，ノイズが33%の状態でも92%の精度を示し，比較手法を上回ることを実証した．

Two-dimensional Letter Block Search Method to Enhance a Book without Markers

Akihiro Miyata,[†1] Hisako Shiohara[†1] and Ko Fujimura[†1]

We present a text search method which takes into account not only the reading direction but also the non-reading direction. We use this method to develop a prototype system called *Kappan*. It enables service providers and users to create hyperlinks in books without markers. Existing techniques generally require markers to be printed on the page if a hyperlink is to be created. We consider that utilizing the concept of the search index makes markers unnecessary, *i.e.*, the system can detect positions using text extracted from images via OCR (Optical Character Recognition) and provide users with position associated digital contents. Traditional text indexing methods must extract long character sequences from the partial image in order to identify the area exactly given the sheer number of book pages. However, considering that the average OCR error rate is more than 35 percent if the partial image is captured by a camera-equipped cellular phone, it is highly probable that many characters would be misrecognized and area identification would thus fail. In contrast, our indexing method can extract area-specific clues using fewer characters that can identify the area exactly even when the partial image is small and the extracted text contains misrecognized characters. An experiment proves that our method can identify the exact area from 73,231 documents with the high accuracy rates of 99 percent and 92 percent for OCR error rates of 0 percent and 33 percent, respectively.

1. はじめに

情報の電子化が進んだ現代においても，我々が紙媒体の書籍を情報源として活用する機会は多い．これは，書籍が人間にとって取り扱いやすいメディアであることに起因している．すなわち，書籍には，

- 高速ブラウジングが可能
- パラパラめくることで全体の雰囲気の把握や偶然の情報発見が可能
- 全体の中のどこを読んでいるのかが明確
- 人間にとってなじみ深いメディアであり使用法を教わる必要がない

といった特徴があり[1)]，今なお利便性の高い情報取得手段として利用されている．書籍の形状・操作感を持つデバイスなどに電子データを表示する手法[1)–3)]が今に至るまで数多く提案されていることからも，書籍のユーザビリティの高さが認められていることが分かる．

ところが，書籍は紙媒体であるため，扱えるコンテンツがテキスト，図表，写真などの静的なものに限られる．また，紙面のサイズ，ページ数といった制約もあり，記録できる情報量には限りがある．この解決を目指す研究が数多くなされており，デジタルコンテンツへのハイパーリンクを書籍中にマーカとして記載する方法（2章）が主流であるが，あらかじめ書籍内にマーカを記載しなければならないという問題がある．

†1 日本電信電話株式会社NTTサイバーソリューション研究所
NTT Cyber Solutions Laboratories, NTT Corporation
本論文の内容は2010年3月のインタラクション2010シンポジウムにて報告され，同プログラム委員長により情報処理学会論文誌ジャーナルへの掲載が推薦された論文である．

一方で，システム上に書籍内のテキストと位置を関連付けて検索可能にしておけばマーカは不要である．すなわち，書籍内を撮影した画像をシステムに送信すれば，システムは画像から OCR（Optical Character Recognition）により抽出したテキストを検索語として位置を特定し，その位置に関連付けられたコンテンツを提示できる．しかし，この方法では，検索に用いるテキストが短いと位置を一意に特定できず，長いと OCR 誤認識文字がテキストに混入して検索が失敗する危険性がある．特に部分領域画像の範囲が狭い場合，抽出できる検索語数が少なくこの問題は深刻である．ページ内のすべてのテキストを検索語として用いれば検索の成功率は高まるが，検索に 1 ページ全体の情報を要するのであれば，利便性の高いシステムは構築できない．すなわち，システムはページ単位でしか位置を特定できず，ページ内の細かい位置特定は行えない．これでは，ページ内の上部，中部，下部などの細かい単位で異なるコンテンツを関連付けることができなくなってしまう．

本論文では，OCR 誤認識が発生する書籍内の局所領域画像の位置を一意に特定するために，文章を読む方向とそれに直交する方向を考慮した 2 次元のブロックをキーとする 2 次元文字ブロック検索手法を提案する．対象とする書籍は，白地に黒で印字されたものを主に想定しているが，これに限定はしない．検索対象となる書籍数は多ければ多いほど本技術をサービス化した際に魅力的であるが，最低でも数百冊（数万ページ）程度から一意に位置を特定できることを目指す．我々はカメラ付き携帯電話でこれらの書籍を撮影するシステムを想定しており，手振れ・手の影などに起因する撮影画像からの OCR 誤認識は避けられないと思われるので，この誤認識の存在を研究の前提条件とする．また，小説・教科書などの一般的な書籍の印字サイズにおける 4〜6 行程度の長さを 1 辺とする矩形領域を局所領域とする．多くの書籍を観察したところ，1 段落は短くとも 4〜6 行程度である場合が多く，最小でもこの単位で位置特定・コンテンツ関連付けができれば十分だと判断した．

本論文後半では，本技術を用いて書籍内の局所領域にハイパーリンク設置を可能にするシステム Kappan を紹介する．インデックス作成済みの書籍であれば，ユーザは書籍内の任意局所領域をカメラ付き携帯電話で撮影するだけでその領域に関連付けられたコンテンツを閲覧できる．これにより，たとえば，旅行ガイドブックと風景映像を関連付けるサービスが実現できる（**図 1**）．

本論文の貢献するところは，2 次元文字ブロック検索手法を提案すること，本技術が比較方式よりもノイズに強く精度が高いことを大規模データを用いた実験で実証すること，提案手法を用いたシステム Kappan の実現方式を示すことである．

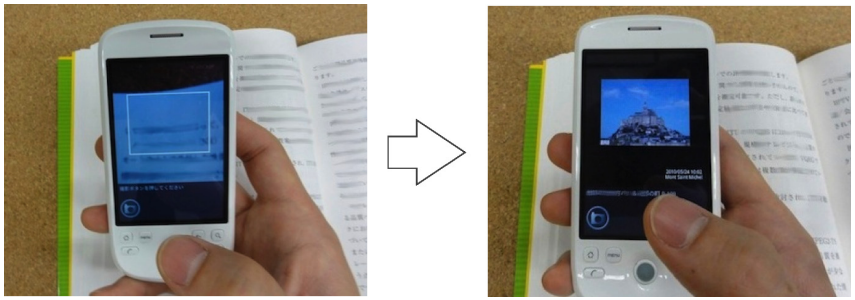


図 1　Kappan 利用時の様子
Fig. 1　A usage example of Kappan.

2. 書籍とデジタルコンテンツ

ヒューマンコンピュータ・インタラクションの分野では，紙媒体とデジタルコンテンツを連携させる様々な試みが古くから行われている．

2.1 明示マーカ方式

紙媒体とデジタルコンテンツを関連付ける方法として古くから用いられているのが，バーコード，Data Matrix[4]，QR コード[5]，その他独自マーカを含めた，人間が容易に視認可能な明示マーカを使うものである．Insight Lab[6] では紙媒体とマルチメディアデータがバーコードによって関連付けられている．Billinghurst らは，書籍のページ中にある黒い枠線で囲まれたマーカをヘッドマウントディスプレイ越しに見ると，マーカが存在する位置に仮想オブジェクトが重畳されて見えるシステムを提案した[7]．Metaio 社のシステムでは，マーカを含むページを PC に接続したカメラで撮影すると，ページに仮想物が重畳された様子を PC 上で確認できる[8]．

これらの方法により，書籍が扱えるコンテンツの制約，情報量の制約を克服できる．しかし，書籍中のページにマーカを記載しなければならず，2 つの問題が発生する．1 つ目は，マーカが書籍の見た目を損ねてしまう場合があるという問題である．多くのマーカは一定サイズ以上である必要があるが[⋆1]，Mohan らは Bokode という直径数ミリメートルの光学

⋆1 QR コードは 1 辺が 1.5 cm 程度ないと，読み取り装置のカメラの焦点が合わずに読み取れないことがあるという．

マーカでこの問題の解決を試みている[9]．このマーカは，カメラの焦点を合わせても小さな点にしか見えないが，カメラの焦点を遠距離に合わせて撮影すると情報が取得できる仕組みになっている．ただし，マーカを不可視にする試み[10),11)]が数多いことから分かるように，コンテンツ中にマーカが存在することは人間にとって不自然である．特に，小説など，文学的・芸術的な書籍の各ページにマーカが記載されていたら興ざめであり，読者に受け入れられないことは想像に難くない．我々は日常的に多くの出版社・印刷会社にヒアリングを行っているが，書籍内のレイアウトデザインは綿密に練られているため，マーカの記載は極力避けたいという声は多く聞かれる．2 つ目は，マーカを記載するタイミングが限定されるという問題である．マーカが書籍中のページに印刷されている必要があるため，書籍が出版される前にマーカを記載しなければならない．

2.2　非明示マーカ方式

マーカを紙面上で目立たせない方法，あるいは不可視にする方法も古くからあり，特にビットマップパターンを用いるアプローチが多く提案されている．DataGlyph[12)] は 45 度に傾いた線状のビットマップパターンで情報を表現するマーカであり，ページの端に記載すると模様のようで目障りになりにくい．左への傾斜で 1，右への傾斜で 0 を表している．Kise らは点描画像内にデータを埋め込む方式を提案した[13)]．画像内の各点を描画するかしないかで，データの各ビットを表現している．Anoto technology[14)] はほぼ不可視なドットがあらかじめ全面に印刷された特殊な紙を用いる．紙面内でユニークな x-y 座標を表現するドット群をカメラペンでキャプチャすることによって，手書き文字などが自動的にデジタル記録される仕組みである．Lu らは小さいドットパターンを紙面に埋め込むことでマルチメディアデータへのハイパーリンクを設置する出版フレームワークを提案した[15)]．上述とは異なるアプローチとしては，Luff らによる導電性インクを用いた不可視マーカ[16)] があげられる．

これらの方式は，マーカが紙面上で目立たず，書籍の見た目を損ねないという点で優れている．ただし，書籍が出版される前にマーカを記載しなければならないという問題は残っている．

2.3　マーカレス方式

上述の手法は明示的であるにせよ，非明示的であるにせよ，マーカを用いている．そのため，書籍とデジタルコンテンツを関連付ける場合，書籍が出版される前にマーカを紙面に記載するか，特殊な紙を用いて出版しなければならず，実用的とはいい難い．また，すでに出版されている書籍に適用することも困難である．この点に鑑み，紙媒体の撮影画像を手がかりに元の紙媒体を特定することで，マーカを用いずに紙媒体とデジタルコンテンツを関連付ける手法もいくつか提案されている．本研究もこの方式に属するため，白地に黒で印字されている数百冊以上の書籍内の局所領域にハイパーリンクを設置できるか（1 章）という観点からの議論も記載する．

ペン状小型カメラとラインマーカを組み合わせた PaperLink[17)] は，ユーザが書籍中の任意位置にラインを引くと，その領域が撮影されてシステムに蓄積される．ユーザがその領域に関するアクションを定義した後，再度その領域を撮影すると定義済みアクションが実行される．PaperLink はマーカレス方式の先駆的存在であり，優れたコンセプトである．一方で，位置特定は手描きラインの形状が毎回微妙に異なる点に着目した画像パターンマッチングで実現しているため，位置特定したい部分にはすべて手描きでラインを引かなければならず，大量の書籍内の各領域にハイパーリンクを設置する用途には適していない．

PaperLink と同様の発想で，書籍内の全ページを画像化してデータベースに格納し，カメラで撮影した書籍内の局所領域画像の単純な画像特徴量をクエリとして撮影位置を特定する検索システムも考えられるが，この実現は難しい．位置・大きさが既定でない部分画像をクエリとする部分画像検索問題では，速度と精度のトレードオフがあるからである．速度を重視する手法としては，検索対象となる画像群をそれぞれサブブロックに分割し，サブブロック単位でクエリ画像と色情報を照合する手法[18)] や，隣り合うサブブロック内の代表的な色の対を利用して物体を検出する手法[19)] があるが，位置の特定精度は原理上高くない．精度を重視する手法としては，多数の局所領域に着目して照合する手法[20)] もあるが，この場合は検索対象となる画像群において，比較部分の位置・大きさを様々に変化させながらクエリ画像と特徴照合を行わなければならないため，照合回数が膨大になり処理速度が低下する．そこで Taketa らの研究[21)] では，絵本の部分画像をクエリにして元位置を特定するためにアクティブ探索[22)] を用いている．これは，画像内の各色の割合（色ヒストグラム）を手がかりにして，探索対象を含まない画像領域を判定し，この領域を探索計算対象からスキップすることで計算コストを単純手法の 100〜1,000 倍程度削減する手法である．しかし，アクティブ探索は色ヒストグラムを手がかりに探索対象を限定しているため，クエリ画像および検索対象画像群に出現する色の数が少なく，各画像の色特徴が似通っている場合には，探索対象を限定できず，計算コストの低減効果がほとんど得られない．本研究が想定しているようなクエリ画像・検索対象画像群が白地に黒い文字の文書画像である場合，そこに出現する色数はわずかであり，各画像群の色特徴は非常に似通ったものになる．このため，このような書籍が分析対象である場合，アクティブ探索を用いても短時間での位置特定は期待で

きない．

HOTPAPER [23] は文書画像から特徴量を抽出する際に，撮影画像内の各単語を包含する矩形領域の縦横比を手がかりにして撮影領域を特定する方式をとっている．ただしこの場合，適用対象が英語などの各単語がスペースで区切られる言語に制限されてしまう．また，書籍内領域の特徴量を単語包含矩形の縦横比だけで表現しているため計算コストが小さい一方，検索対象が膨大になった場合に狭領域画像からは検索結果が一意に絞り込まれない可能性がある．Liu らも各単語を包含する矩形領域の幾何学的特徴に着目した手法[24] を提案しているが，やはり適用対象がスペースで区切られている言語に制限されてしまう．

LLAH [25)–28)] も文書画像中から特徴量を抽出して撮影された文書を特定する方式であり，この研究領域では最も洗練されたアプローチの 1 つである．この方式では，文書を撮影するカメラ位置の影響を受けにくい特徴量として，文字領域の重心点から算出するアフィン不変量などを用いている．また，特徴量のハッシングにより検索のリアルタイム性を高める工夫もされている．文書 1 ページの全体画像をクエリとして 10,000 ページの候補から元文書を特定するタスクにおいては，検索時間 0.14 秒，精度 98%を達成している[25),26)]．一方で，局所領域をクエリとした場合の識別性能，すなわち文書の局所領域画像をクエリとして大規模な文書群の中から領域位置を一意に特定する精度については，あまりフォーカスされていないようである．1,000 ページ（文庫本 3 冊程度）の文書群の中から日本語文書の部分画像をクエリとして撮影位置を包含する文書を特定するタスクにおいて，約 10 回問合せを行って半数以上の結果が正しければ検索成功と判定する評価においても，精度は 90%となっている[27),28)]．そのため，局所領域画像をクエリとし，数百冊以上（数万ページ以上）の書籍の中から撮影した該当する位置を特定する用途には，LLAH をそのまま適用することは難しいと思われる．

WordAnchor [29] では，あらかじめ書籍中の各単語が登場する位置がデータベースに格納されている．ユーザが専用デバイスで書籍の一部領域を撮影すると，その画像に含まれる単語群を元にデータベースに位置問合せが行われ，単語間の隣接関係を手がかりに撮影位置を特定する．この方法は，撮影範囲が広く，撮影画像が鮮明な場合は大変有効である．一方で，本研究が前提としているような，撮影範囲が局所領域であり，手振れ・手の影などがある不鮮明な撮影画像への適用は容易ではないと思われる．撮影範囲が狭いと抽出できる単語数はわずかであるし，しばしば画像の端で単語が途切れてしまう．画像が不鮮明だと OCR 誤認識が多く発生するので，誤認識文字を含まずに正しく抽出できる単語は少ない．この状況では単語の隣接関係がうまく検出できない可能性がある．たとえば，単語 A，B，C の順に並ぶ隣接関係で位置を特定する場合，B に 1 文字でも誤認識文字が含まれると，あるいは A の端の文字が 1 文字でも欠けてしまうと，A，B，C のすべてが位置特定に使えない．

3. Kappan の概要

3.1 研 究 目 標

我々の研究目標は，書籍が持つ手軽さ・高いユーザビリティはそのままに，そこから格段に多くのコンテンツ・情報を得られるという，新しい情報取得スタイルを “実用レベル” で実現することである．より具体的には，ユーザが書籍の任意位置に注目すると，その位置に関連付けられたデジタルコンテンツを閲覧できる仕組みを想定している．これを実現するためには，下記 4 つの要件を満たす必要がある．

要件 1）市販書籍をそのまま用いる：特殊マーカが記載された専用書籍の使用や，市販書籍に何らかの加工を施す方法は実用的とはいえない．より多くのユーザが手軽に本システムを利用するためには，ユーザが取得した市販書籍がそのまま活用できるべきである．

要件 2）操作が容易である：情報取得操作が煩雑では，書籍の特徴である “めくる” だけという簡便な操作感を損ねてしまう．ユーザは書籍とデバイスの両方を同時に扱わなければならないので，操作は極力簡単であるべきである．

要件 3）細かい粒度でコンテンツが関連付けられる：コンテンツを関連付けられる粒度は細かい方が望ましい．たとえば，ページ内にそれぞれトピックが異なる複数の段落がある場合，ページ単位でしかコンテンツを関連付けられないよりも，ページ内の段落ごとに異なるコンテンツを関連付けられる方が利便性が高いのは明らかである．多くの書籍を観察したところ，1 段落は短くとも 4〜6 行程度である場合が多かったので，この程度の粒度でコンテンツ関連付けができるべきである．

要件 4）普及デバイスを用いる：高価な専用デバイスでは，提案する情報取得スタイルを実用化することはできない．安価かつ普及している市販デバイスを用いることが望ましい．加えて，どこにでも持ち運べるという書籍のメリットを損なわぬよう，デバイスは軽量・可搬であるべきである．

3.2 Kappan のコンセプト

上記要件を満たす方法を検討する．なお，以降 “位置” とは書籍中における位置を意味し，書籍名，ページ，行，列など，位置を表現する情報を “位置情報” とする．(1) まず，要件 1 のように書籍に手を加えずに，ユーザが注目した位置をシステムが把握する手段として，検索システムの概念が適していると考えた．つまり，各位置の内容と位置情報を関連付けた

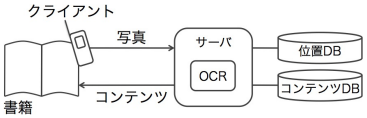


図 2　Kappan のコンセプト
Fig. 2　The concept of Kappan.

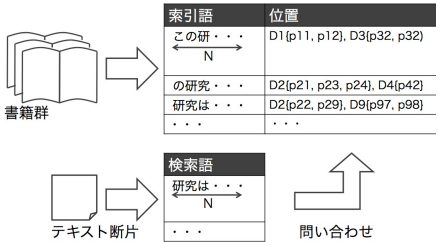


図 3　転置ファイル方式を用いた検索
Fig. 3　The concept of inverted index search.

データベースを事前に構築し，ユーザが着目した位置の内容をシステムに伝達すると，システムがその内容が存在する位置を特定する仕組みである．(2) 次に，要件 2 の容易な操作方法を考える．ユーザが着目した位置にある内容をシステムに簡単に伝達する手段として，我々はその位置を撮影してシステムに撮影画像を送信する方式が適していると考えた．(3) また，要件 3 より，上述の撮影画像はページの一部分であっても，システムはその位置を正しく特定できる必要があると考えた．(4) そして，要件 4 のようにユーザが安価で入手でき，軽量・可搬のデバイスとして，我々はカメラ付き携帯電話が適していると考えた．

ここで，(1)，(2) をふまえると，システム実現のためには CBIR（Content-based Image Retrieval）の適用が考えられるだろう．すなわち，事前にシステム上に各位置の画像特徴量と位置情報を関連付けてデータベースに格納し，システムがユーザの撮影した書籍内局所領域画像の特徴量に適合する位置情報をデータベースに問い合わせる方式である．しかし，クエリとなる書籍内局所領域画像は，たとえ同じ位置であっても撮影者によって撮影範囲・角度・光条件などが様々であることが想像できる．この条件下では適合画像の検索は困難であるし，検索精度を上げるためには計算コストの増加，データベースの肥大化が避けられない．

そこで我々は発想を変え，この問題をテキスト検索のアプローチで解決する．つまり，事前にシステム上に各位置のテキストと位置情報が関連付けられてデータベースに格納されており，システムはユーザが撮影した書籍内局所領域画像から OCR で抽出したテキスト断片に適合する位置情報をデータベースに問い合わせる方式である．多くの場合，テキスト検索は CBIR よりも計算コストが低く，必要とするデータ領域も小さい．

上記の議論をふまえたシステム Kappan のコンセプトを**図 2** に示す．クライアントは，写真撮影，通信機能を備えた携帯電話である．位置 DB は，テキストをクエリとし，そのテキストが含まれる位置情報を結果として返すデータベースである．コンテンツ DB は，位置情報をクエリとし，その位置に関連付けられたコンテンツを返すデータベースである．サーバは，ユーザがクライアントで撮影した局所領域画像から OCR でテキスト断片を抽出し，その断片内のテキストを含む位置を位置 DB に問い合わせ，その位置に関連付けられたコンテンツをコンテンツ DB に問い合わせ，得られたコンテンツをクライアントに送信する．

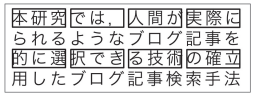
4. 2 次元文字ブロックを用いた書籍内位置検索の提案

4.1 研 究 課 題

前章で述べたコンセプトを実現するためには，書籍内のテキスト断片をクエリとし，その断片が含まれる位置を一意に取得できるような，識別性能が高い検索機能が必要である．もし位置が一意に特定できず複数の候補が検索結果として取得されると，ユーザは候補の中から正しい位置を選択しなければならず，操作の容易性（3.1 節要件 2）を大きく損ねる．また，できるだけ小さいテキスト断片から検索が行えるべきである．具体的には，4〜6 行のテキスト断片から検索できるべきである（3.1 節要件 3）．ここで注意すべきなのは，4〜6 行だけが写るように接写すると，必然的に各行の列全体が写らない場合が多いということである．そのため，紙面上における 4〜6 行程度の長さを 1 辺とする矩形領域に収まるだけの少量テキストしか取得できない場合でも検索できる必要がある．なお，OCR 処理コストを低減する目的でも，位置特定に必要なテキスト量は少ない方が望ましい．

検索機能の実現方法としては，文字列が出現する位置を高速検索するために広く使われている転置ファイル方式[30] が妥当と考えられる．たとえば，検索対象となる全書籍内のテキストから，N-gram 索引（索引語が N 文字）による転置ファイルを作成する．そして，テキスト断片から N-gram の文字列を検索語として抽出し，この語にマッチする位置情報を取得する（**図 3**）．このとき，検索の識別性能と N は密接な関係にある．文章とは文字がランダムではなく，意味を成すように並んでいる．このため，意味を成す文字の連なりは多くの位置に登場する．つまり，検索の識別性能を高めて位置を一意に特定するためには，索引語・検索語（以降，キーとする）の文字列単位を長く（N を大きく）し，各位置にできるだけ固有なパターンを抽出する必要がある．たとえば，キーが “検索”（N = 2）であるより，“検索処理技術”（N = 6）である方が検索の識別性能が高く，位置を一意に特定しやすい．

しかし，書籍内の部分領域画像を OCR 処理してテキストデータに変換する際に，OCR 誤認識が発生する点を考慮する必要がある．一般に，商用レベルの OCR 誤認識率は数%程度であるが，これは元画像の撮影状態が良い場合である．Kappan のシステム構成では，片



(a) キー：連続する N 文字

(b) キー：ブロック

図 4　索引語・検索語の抽出方法
Fig. 4　The examples of term extraction.

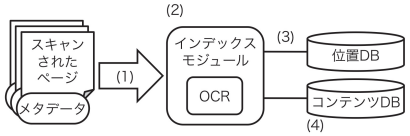


図 5　インデックス作成処理の流れ
Fig. 5　The indexing procedure.

手で書籍を持ち，もう一方の手に持ったカメラ付き携帯電話で書籍内の部分領域を撮影することを想定しており，手ぶれ，手の影などが画像の状態を悪化させる．調査実験では，被験者 5 名が撮影した画像を OCR 処理したところ，誤認識率は平均で約 35%であった[⋆1]．このため，OCR 処理で得られたテキストデータから文字数が多い検索語を抽出すると，検索語に誤認識文字が含まれて検索が失敗する危険性がある．特に局所領域画像の範囲が狭い場合，抽出できる検索語数が少なくこの問題は深刻である．

そこで，本論文では，**OCR 誤認識が発生する書籍内局所領域画像に該当する位置を一意に特定できる検索手法の確立**を研究課題として設定する．

4.2　提 案 手 法

前節で述べた研究課題を解決するためには 2 つの条件がある．1 つ目は，キーが少ない文字数で構成されていることである．文字数が少なければ，そこに OCR 誤認識文字が混入する可能性は低くなるし，局所領域画像からでも多くのキーを抽出できる．2 つ目は，キーが各書籍内位置に固有なパターンであることである．各位置に固有なパターンであれば，それを利用して位置を一意に特定できる可能性が高い．

そこで我々は，キーを抽出する方法に着目した．通常の転置ファイル方式では，**図 4** (a) のように文章を読む方向に連続する文字群をキーとしている．一方で，同図 (b) のように文章を読む方向と直交する方向も考慮した 2 次元のブロック単位でキーを作成すると，ブロック内の文字数が少なくても各位置により固有なパターンになると思われる．なぜなら，このブロックに含まれる文字の集合は書籍のページレイアウトや改行位置に起因する “偶然の産物” であるからである．たとえば，図 4 (a) の枠線内にある読む方向に並んだ 3 文字（例：本・研・究）よりも，同図 (b) のブロック内にある 3 文字の集合（例：本・研・ら）の方が，この位置に固有であると考えられる．

そこで我々は，**文章を読む方向とそれに直交する方向を考慮した 2 次元のブロックをキーとする 2 次元文字ブロック検索手法を提案する．**

⋆1 片手で書籍を持ち，もう一方の手で携帯電話を操作して撮影した画像に基づく結果である．場所はオフィスビル内の照明灯の下，撮影対象は日本語の横書き書籍（“ユーザが感じる品質基準 QoE” の P.1，P.47，P.119），OCR ソフトウェアは大手メーカから数万円で販売されているものである．

表 1　位置 DB の例
Table 1　Position DB.

索引語	位置情報
本研ら	Book A, Page 2, Line 10, Column 1
研究れ	Book A, Page 2, Line 10, Column 2
究でる	Book A, Page 2, Line 10, Column 3

表 2　コンテンツ DB の例
Table 2　Content DB.

コンテンツ	領域
http://1.html	Book A, Page 2, Line 10–15
http://2.html	Book A, Page 3, Line 1–5
http://3.html	Book B, Page 10, Line 14–17

5.　Kappan の実装

本章では Kappan の実装方法について述べる．図 2 におけるサーバはインデックスモジュールと検索モジュールからなる．クライアントには，NTT ドコモ社の携帯電話 HT-03A，Xperia SO-01B を採用した．

5.1　サーバ：インデックスモジュール

インデックス作成処理を**図 5** に示す．(1) では，書籍の各ページをスキャンした画像集合とメタデータ（書籍名，ページ番号）集合を入力する．(2) では，画像を OCR 処理してテキストデータに変換した後，図 4 (b) のように読む方向と読まない方向の 2 次元からなるブロック群を抽出する．(3) では，抽出した各ブロックを構成する文字群を連結した文字列（これを索引語とする）と，そのブロックの位置情報（書籍名，ページ番号，行，列）を関連付けた転置ファイルを作成し，位置 DB に格納する．**表 1** は，図 4 (b) のように 3 文字からなる L 字型のブロックを抽出し，位置情報をブロック内左上端文字の行・列で表現した場合の位置 DB の例である．ブロック内の文字群を別々に扱わずに連結して 1 つの文字列とする理由は，データの問合せが文字列をクエリとするシンプルなものになり検索速度を高速化できるからである．上記の処理とは別に (4) では，書籍内の各位置と，そこに設置するコンテンツ（URL など）を**表 2** のように関連付けてコンテンツ DB に格納する．

なお，書籍は等幅フォントばかりではなくプロポーショナルフォント（非等幅フォント）で印刷されているものもあるため，(2) においてブロック抽出時に読まない方向の文字の並

People, life style, and
technology. The past,
present, and the future.

図 6　等幅フォント
Fig. 6　A monospaced font.

People, life style, and
technology. The past,
present, and the future.

図 7　プロポーショナルフォント
Fig. 7　A proportional font.

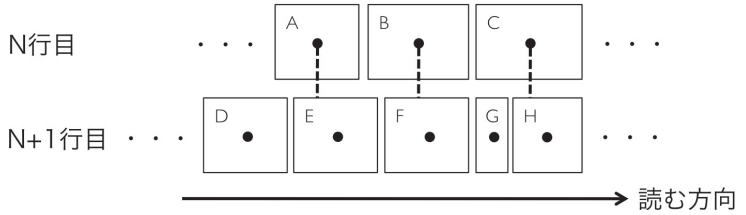
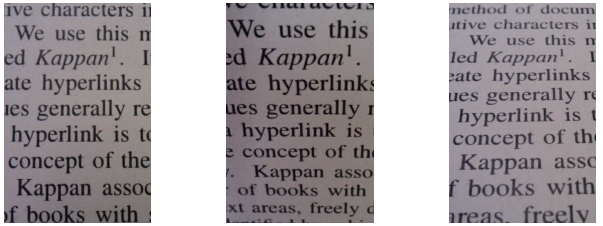


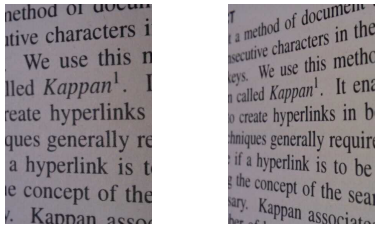
図 8　プロポーショナルフォントにおける文字の並び判定例
Fig. 8　The position relations of proportional letters.

びを検出する際には工夫が必要である．ここでは，左から右に読む横書き文章を例に説明する．**図 6** のように等幅フォントの場合は，縦に並んで印刷されている文字（たとえば円で囲んだ 1 行目 “l” と 2 行目 “p”）はそれぞれの行における左端からの文字数も等しいため，この縦の並びを検出するのは容易である．すなわち，ページ画像を OCR 処理してテキストデータに変換した後，左端からの文字数が等しい文字のペアが縦に並んでいると判定すればよい．一方で，**図 7** のようなプロポーショナルフォントの場合は，縦に並んで印刷されている文字（たとえば四角で囲んだ 1 行目 “n” と 2 行目 “t”）はそれぞれの行における左端からの文字数は 22 文字目，20 文字目と等しくない．

そこで，今回の実装では OCR 処理時に各文字を包含する矩形領域を定義し，矩形と矩形重心の位置関係に基づいて文字の並びを判定することとする．たとえば**図 8** のような文書の場合，文字 A/B/C を包含する矩形領域の重心座標（図中の●印）は，それぞれ文字 E/F/H を包含する矩形領域の上にあるので，文字 A と文字 E，文字 B と文字 F，文字 C と文字 H が縦に並んでいると判定する．実装したシステムで，正面・上方・下方・右 30 度の位置から撮影した全文英語のプロポーショナルフォント文書画像（**図 9** (a)〜(d)）を分析したところ，どの場合も各文字の上下関係判定は正しく行われていることを確認した．しかし，同図 (e) のように，右 45 度までカメラを傾けて撮影を行うと，一部の文字間で正しく上下関係判定が行えなかった．ただし，下記 2 点の理由で，このケースはあまり問題にならないと考えられる．1 つ目の理由は，多くのユーザは紙面とほぼ平行にカメラを構えて撮



(a) 正面から　(b) 上 30 度から　(c) 下 30 度から



(d) 右 30 度から　(e) 右 45 度から

図 9　プロポーショナルフォント文書を各方向から撮影する例
Fig. 9　The snapshots of a proportional font document.

影すると想定されるからである．本システムは過去に 2 回，社外展示を行っており，一般ユーザに試用してもらった．その際，大半のユーザは特に指示を与えなくてもカメラ付き携帯電話を紙面にほぼ平行にして撮影しており，傾きが多いユーザでもたかだか 30 度であった．2 つ目の理由は，一部の文字間で上下関係判定にずれが生じ，それらの文字を含む検索キーが無効になっても，他のキーで検索が行えるからである．5.2 節で後述のとおり，一部のキーが無効になっても正しく検索が行えるよう，提案アルゴリズムは各キーを独立に扱って多数決方式で位置特定を行っている．

5.2　サーバ：検索モジュール

検索処理を**図 10** に示す．(1) では，携帯電話でページの局所領域を撮影してサーバに送信する．(2) では，画像を OCR 処理してテキストデータを抽出し，そこからインデックス作成処理時と同じ形状（ここでは 3 文字からなる L 字型）の 2 次元文字ブロック集合を抽出する．**図 11** (a) は入力画像，図 11 (b) は抽出されたテキストデータの例である．一般にユーザが片手で持った携帯電話で撮影した画像では OCR 誤認識が発生しやすい．この例でも，OCR 処理の過程で，2 行目の「る」，3 行目の「択」が形状の似た別の字に誤認識され

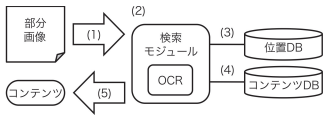
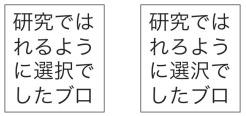
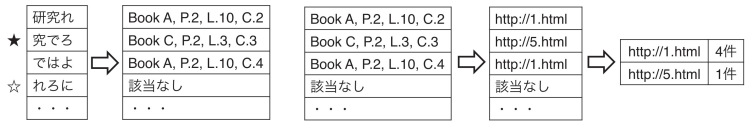


図 10　検索処理の流れ
Fig. 10　The seach procedure.



(a) 元画像　(b) 抽出結果

図 11　検索処理における元画像・抽出テキストの例
Fig. 11　The source photo and extracted text.



(a) 位置の問合せ　(b) コンテンツの決定

図 12　検索結果コンテンツの決定方法
Fig. 12　The procedure of content detection.

たとする．(3) では，位置 DB（表 1）に対して，抽出した各ブロックを構成する文字群を連結した文字列（これを検索語とする）と索引語が一致するレコードの位置情報を問い合わせ，**図 12** (a) のように位置情報集合を取得する．このとき，誤認識された文字を含む検索語の問合せ結果は，偶然一致する索引語が存在する他の位置であったり（同図★印），一致する索引語がなかったりする（同図☆印）．(4) では，コンテンツ DB（表 2）に対して，各位置情報を領域に含んでいるコンテンツを問い合わせ，図 12 (b) のようにコンテンツ集合を取得し，コンテンツごとに件数を集計して最大件数のもの（ここでは http://1.html）を検索結果と特定する．(5) では，検索結果を携帯電話に送信する．これにより，ユーザは撮影した領域に設置されていたコンテンツを閲覧できる．

5.3　クライアント

我々は Java でプログラムを開発し HT-03A および Xperia SO-01B（Android OS）上で稼働させた．このプログラムは，撮影画像をサーバに送信し，サーバから返ってくる検索結果コンテンツの URL にアクセスする．OCR 処理や検索処理など計算コストが高い処理はサーバ上で実行するため，現在市販されている程度のスペックの携帯電話であれば軽快に動作する．クライアントを実際に動作させている様子を図 1 に示す．同図は，旅行ガイドブック内のある観光地紹介ページを撮影し，その観光地を紹介する映像コンテンツを閲覧している様子を示している．また，撮影した場所にコンテンツが存在しない場合には，**図 13** のような画面を表示して撮影箇所近くのページに存在するコンテンツへのリンクを提示する．

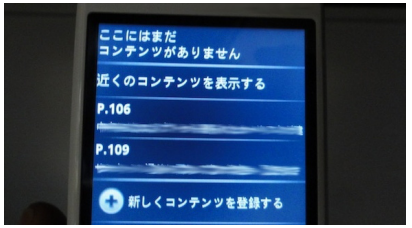


図 13　コンテンツが存在しない場所を撮影した場合
Fig. 13　The message that means there are contents near the page shot by the user.

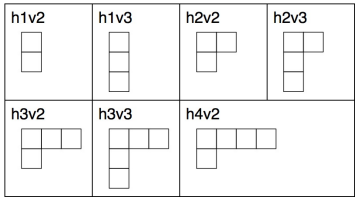


図 14　検証に用いた 2 次元文字ブロックの形状
Fig. 14　2D letter block used in the experiment.

なお，Android 携帯電話はサードパーティ（我々を含む）が開発したプログラムを簡単にインストールする仕組みが備わっているため，市販の Android 携帯電話にこのプログラムをインストールするだけで Kappan のクライアントになる．これらは実用性を考慮しての判断である．

5.4　システムを利用したサービスイメージ

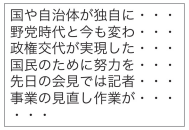
Kappan の適用対象としては旅行ガイドブックが考えられる．ガイドブック内の各位置に風景映像を関連付けておくと，読者は旅先でガイドブックを広げて次の目的地の様子を事前に把握できる（図 1）．あるいは，教科書内の各位置に講義映像や問題の詳細な解法を関連付けておき，生徒の自宅学習を支援するというサービスも考えられる．

6.　評価実験 1

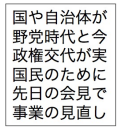
6.1　実 験 概 要

この実験の目的は，Kappan の中核技術である 2 次元文字ブロック検索手法の有効性の検証である．具体的には，局所領域画像の OCR 結果を想定した小さなテキスト断片を含む位置を検索するタスクを行い，提案手法の，(1) 識別性能，(2) OCR 処理誤認識への耐性の検証を行う．提案手法のキー抽出方式は 2〜5 文字からなる L 字型 2 次元文字ブロック計 7 方式を検証する（**図 14**）．文字ブロックを構成する 1 マスは 1 つの文字を表現している．比較手法のキー抽出方式は，N 文字単位で索引語を抽出する N-gram 方式（N = 2〜5），分かち書き[*1]により得られた全単語をキーとする単語方式を検証する．なお，検証対象の方式

⋆1 mecab0.98・mecab-ipadic2.7.0-20070801 を利用した．



(a) 記事（ID=1000）



(b) 断片

図 15　評価実験に用いた記事の例
Fig. 15　The sample of articles used in the experiment.

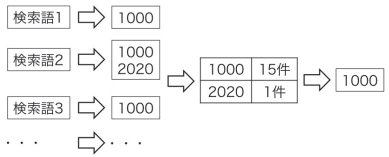


図 16　検索結果決定の処理イメージ
Fig. 16　The determination procedure of search results.

表 3　評価実験における提案手法の位置 DB の例
Table 3　The position databases of proposal methods.

索引語 (h2v2)	位置情報 (記事 ID)	索引語 (h3v3)	位置情報 (記事 ID)
国や野	1000	国や自野政	1000
や自党	1000	や自治党権	1000
自治時	1000	自治体時交	1000

数が多いうえに，各方式に対して統計的に有意な検証を行うためには相当数の試行が必要である．よって，本実験はシミュレーションによって行う．

6.2 実験方法

検証用書籍データとして，2008 年 1 月 1 日〜12 月 31 日の毎日新聞の全記事（電子データ版）計 73,231 件を用いる．1 記事を書籍の 1 ページと見なし，各記事の本文を横書き 1 行 40 文字のテキストファイルに成形し，各記事に ID を付与する．記事の平均文字数は 518 文字，成形後の平均行数は約 13 行である．

次に，インデックスを作成する．提案手法・比較手法の各方式で，全記事の先頭から 1 文字ずつずらしながら索引語を抽出し，抽出元記事 ID と関連付けて位置 DB に格納する．たとえば**図 15** (a) の記事から抽出した索引語は**表 3**，**表 4** のように格納される．各方式で作成した位置 DB の情報を**表 5** に示す．

続いて，テキスト断片を作成する．成形済み記事の中から 7 行以上あるものを無作為に 1,000 件抽出し，図 15 (b) のように各記事の上から 1〜6 行目，左から 1〜6 文字目の領域にある計 36 文字を抽出してテキスト断片とする．この大きさは，前述の局所領域画像サイズに基づいている（4.1 節）．作成した断片は無作為に 100 件ずつ，10 組の断片データセットに分ける．なお，この断片はテキストファイルから抽出したため，OCR 誤認識文字は含まれていない．そこで，各断片の中から無作為に選んだ最大 50%の文字をノイズ文字 “__” に置き換える[⋆1]ことで，誤認識文字をエミュレートする．

表 4　評価実験における比較手法の位置 DB の例
Table 4　The position databases of baseline methods.

索引語 (5-gram)	位置情報 (記事 ID)	索引語 (単語)	位置情報 (記事 ID)
国や自治体	1000	国	1000
や自治体が	1000	や	1000
自治体が独	1000	自治体	1000

表 5　各方式で作成した位置 DB の情報
Table 5　The summary of position databases.

	索引語の総数 (A)	索引語のユニーク数 (B)	B / A
h1v2	35,035,268	1,880,381	5.4%
h1v3	32,206,469	19,086,414	59.3%
h2v2	34,194,093	12,125,120	35.5%
h2v3	31,435,026	27,660,250	88.0%
h3v2	33,351,102	21,791,629	65.3%
h3v3	30,661,822	29,074,866	94.8%
h4v2	32,506,325	26,551,155	81.7%
2-gram	36,949,353	484,677	1.3%
3-gram	35,967,104	3,793,041	10.5%
4-gram	34,986,767	10,102,853	28.9%
5-gram	34,008,163	16,457,158	48.4%
単語	24,474,035	125,733	0.5%

上記インデックス，テキスト断片を用いて断片データセットごとに検索タスクを行う．これは，テキスト断片からその断片の抽出元である記事 ID を求めるタスクである．まず，断片データセット内の 1 つの断片から各方式で先頭から 1 文字ずつずらしながら検索語を抽出する．次に，検索語ごとに位置 DB に問合せを行い，対応する記事 ID を取得して検索結果候補群に加える．同一検索語に対応する記事 ID が複数ある場合はすべて検索結果候補群に加える．すべての検索語について問合せが終了したら，検索結果候補群を集計し，最も多く出現する記事 ID を検索結果とする．**図 16** に検索結果決定の処理イメージを示す．この

⋆1 全記事の中で元々 “__” を含んでいるものはなかった．

表 6　実験結果
Table 6　The evaluation result.

ノイズ率	提案手法							比較手法				
	h1v2	h1v3	h2v2	h2v3	h3v2	h3v3	h4v2	2-gram	3-gram	4-gram	5-gram	単語
0.0%	0.989	0.988	**0.990**	0.989	0.987	0.989	0.989	0.984	0.980	0.982	0.986	0.838
2.8%	0.989	0.988	**0.990**	0.988	0.987	0.988	0.989	0.984	0.980	0.981	0.985	0.800
5.6%	0.989	0.989	**0.990**	**0.990**	0.987	0.989	0.987	0.982	0.978	0.980	0.980	0.745
8.3%	**0.990**	0.987	**0.990**	0.988	0.988	0.989	0.986	0.984	0.979	0.981	0.970	0.643
11.1%	**0.990**	0.989	0.988	**0.990**	0.986	0.985	0.988	0.980	0.980	0.980	0.979	0.497
13.9%	**0.990**	0.989	0.989	0.989	0.986	0.987	0.984	0.984	0.980	0.978	0.960	0.277
16.7%	0.987	0.988	0.988	**0.989**	0.984	0.981	0.977	0.977	0.977	0.971	0.936	0.115
19.4%	0.985	**0.989**	0.987	0.985	0.986	0.980	0.965	0.975	0.976	0.957	0.873	0.038
22.2%	0.974	0.989	**0.990**	0.988	0.982	0.962	0.948	0.962	0.973	0.949	0.804	0.011
25.0%	0.930	**0.988**	0.985	0.981	0.974	0.943	0.914	0.950	0.962	0.906	0.733	0.004
27.8%	0.816	0.983	**0.987**	0.967	0.955	0.891	0.861	0.912	0.962	0.853	0.671	0.002
30.6%	0.603	0.942	**0.967**	0.902	0.906	0.822	0.792	0.820	0.932	0.764	0.510	0.002
33.3%	0.408	0.864	**0.921**	0.827	0.832	0.721	0.651	0.662	0.866	0.661	0.420	0.002
36.1%	0.209	0.656	**0.825**	0.716	0.716	0.600	0.566	0.472	0.760	0.573	0.341	0.003
38.9%	0.090	0.452	**0.655**	0.578	0.569	0.478	0.423	0.266	0.620	0.430	0.251	0.001
41.7%	0.042	0.239	0.425	0.401	**0.435**	0.345	0.327	0.150	**0.435**	0.319	0.196	0.001
44.4%	0.028	0.096	0.227	0.254	**0.270**	0.259	0.212	0.093	0.224	0.179	0.133	0.001
47.2%	0.023	0.030	0.102	0.148	0.147	**0.161**	0.143	0.072	0.130	0.108	0.089	0.001
50.0%	0.019	0.006	0.046	0.071	0.070	**0.092**	0.088	0.060	0.064	0.065	0.043	0.002

作業をデータセット内の全記事について行う．

6.3 実験結果

表 6 に示すのは，断片データセットごとに正しい検索結果が得られた検索タスクの割合を求め，それを全データセット間で平均した値（以降，正解率とする）である．ノイズ率とは断片内の全文字数に対するノイズ文字数の割合である．表中の太字部分は，各ノイズ率において最良の結果を示した方式の正解率である．

ノイズ率 0%の場合，提案手法の各方式と N-gram 方式の間には有意な差が見られない．一方で，単語方式の正解率は 0.84 であり，提案手法や N-gram 方式と比べて低い．1%水準で t 検定を行うと，提案手法の各方式および N-gram 方式（N=2〜5）はすべて単語方式より正解率が有意に高いことが確認できた[⋆1]．

ノイズ率が増えると各方式とも正解率は低下する．我々が想定するノイズ率は約 35%（4.1 節）であることをふまえ，表 6 のノイズ率 33.3%，36.1%の場合を見ると，h2v2 方式が 0.921，0.825 と全方式内で最高の正解率を示している．同条件では 3-gram 方式が 0.866，0.760 と比較手法内で最高の正解率である．各ノイズ率で h2v2 方式と 3-gram 方式の間で 1%水準で t 検定を行うと，どちらの場合も h2v2 方式の正解率の方が有意に高いことが認められた[⋆2]．

6.4 考　　察

まず，単語方式について考察する．ノイズ率によらず単語方式は提案手法の各方式，N-gram 方式より正解率が低かった．これには 2 つの理由が考えられる．1 つ目は，位置 DB 内で重複する索引語の多さである．単語方式はユニークな索引語が 0.5%と少ない（表 5）．すなわち，同一の索引語が多くの位置に存在するため，位置を一意に特定できない場合が多

⋆1 Kolmogorov-Smirnov 検定で各群の正規性，F 検定で比較する各群間の分散の均一性を確認し，Student の t 検定を行った．

⋆2 同じく各群の正規性を確認したが，各群の分散に均一性が確認できなかったので Welch の t 検定を行った．

かったと推測できる．2 つ目は，テキスト断片の周辺部で単語が途切れる場合に，正しい単語が抽出できないことである．たとえば，図 15 (b) 3 行目では「実現」という単語が途切れている．また，単語方式はキー抽出時に分かち書きを行う必要があるが，この処理は提案手法や N-グラム方式のように文字の連続を抽出する処理と比べて計算コストが高く，インデックス作成時，検索時にボトルネックになりうる．これらのことから，単語方式は Kappan の実現には不向きである．

次に，提案手法の各方式と N-gram 方式について考察する．これらは規定文字数のキーを抽出する方式である．利用する文字数が少ないほど，テキスト断片から多くのキーを抽出できるし，検索語にノイズ文字が混入している確率も低い．しかし，キーが短いために位置を一意に特定することは難しい．2 文字を用いる h1v2 方式，2-gram 方式で作成した位置 DB ではユニークな索引語は数%しかなく（表 5），実験結果においてもノイズ率が増加すると大幅に正解率が低下している（表 6）．逆に，利用する文字数が多いほど，キーが長いために位置を一意に特定しやすい．しかし，文字数が多いために検索語にノイズ文字が混入している可能性が高まり，やはり正解率は低下すると考えられる．ノイズ率を上げると，5 文字を用いる h3v3 方式，h4v2 方式，5-gram 方式は 3〜4 文字を用いる方式よりも正解率の低下が大きい（表 6）．

上記の議論をふまえると，キーの文字数は少なくても多くてもよくない．実験においても 3 文字を利用する h1v3 方式，h2v2 方式，3-gram 方式はノイズ率が増加しても比較的高い正解率を保っている（表 6）．6.3 節で前述のとおり，ノイズ率 35%前後では，h2v2 方式が 3-gram 方式よりも有意に高い正解率を示した．また，提案方式内では h2v2 方式が h1v3 方式を上回った．3 方式の中で h2v2 方式が最も高い正解率となった理由は 2 つ考えられる．1 つ目は，h2v2 方式のキーが各位置により固有なパターンであったことである．h2v2 方式のユニークな索引語は 3-gram 方式の 3 倍以上である（表 5）．しかし，ユニークな索引語数であれば，h1v3 方式の方が h2v2 方式よりも多い．そこで考えられる 2 つ目の理由が，テキスト断片から抽出できる検索語の数の違いである．6.3 節で用いた断片は 6 行 × 6 列であるから，h1v3 方式と 3-gram 方式では検索語が 24 個抽出できる．一方，h2v2 方式はその形状から同じ断片から 25 個の検索語が抽出できるため，より正確に検索が行えた可能性がある．

7. 評価実験 2

7.1 実験概要

評価実験 1 は，統計的有意に多数の方式間の比較検証を行うシミュレーション実験であった．テキスト断片は実際の画像から OCR によって抽出したものではなく，誤認識文字はランダム位置の文字をどの記事中にも存在しない文字 “_” に置換することで再現していた．しかし実際の OCR では，ある文字が似通った別の文字に誤認識されることもあり，誤認識文字を含むキーがたまたま他の領域にマッチして，結果として検索精度を落としてしまうことも考えられる．

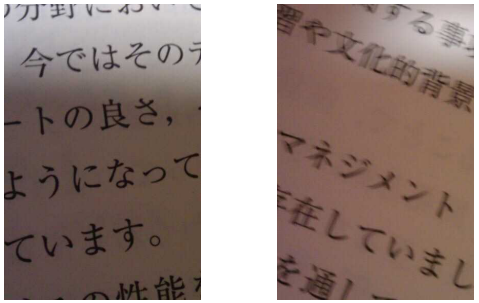
そこで，評価実験 2 では，シミュレーション実験の結果どおり提案手法が既存手法を上回って高精度であるかどうか，実際の撮影画像と OCR を用いて検証を行うことを目的とする．すなわち，実際にカメラ付き携帯電話で撮影した書籍の局所領域画像を OCR 処理してテキスト断片を抽出し，そこからキーを抽出して検索を行った際の精度を検証する．なお，この実験は人手による手作業が多く実験に膨大な時間を要するため，評価実験 1 において提案手法・既存手法でそれぞれ最も高い正解率を示した h2v2，3-gram のみを評価対象とする．

7.2 実験方法

検証用書籍データとして，書籍 168 冊（小説 87 冊，ビジネス書 38 冊，技術書 31 冊，その他 12 冊），合計 57,335 ページを用意し，評価実験 1 同様にインデックスを作成する．すなわち，各ページに ID を付与し，提案手法 h2v2，既存手法 3-gram の各方式で，各ページの先頭から 1 文字ずつずらしながら索引語を抽出し，抽出元ページ ID と関連付けて位置 DB に格納する．

検索するページは日本語の横書き書籍 “ユーザが感じる品質基準 QoE” の序文 1 ページ目を用いる．このページは 26 行 35 列からなっており，“現代” などの一般的な語が連続する部分や，“顧客満足度” などの専門用語が出現する部分がともに存在している．また，多くの部分は等幅フォントの日本語文字で記載されているが，プロポーショナルフォントの英単語が 13 語混在している．

上記のインデックスと検索用ページを用いて検索タスクを行う．まず，上記ページ内の様々な位置にある局所領域をカメラ付き携帯電話（HT-03A）で撮影する．実際に書籍の局所領域を撮影するため，撮影画像内の文字数は撮影の度に多少変動するが，おおむね**図 17** のように 4〜6 行程度が写るようにする．カメラの解像度の都合上，検索用ページにおいて



(a) 例 1　(b) 例 2

図 17　評価実験 2 の撮影画像例
Fig. 17　The examples of query images.

4〜6 行写るように撮影すると，画像の両端で途切れずに撮影できる範囲は 5 列程度となり，おおむね 20 文字が写ることになる．次に，市販 OCR を用いて局所領域画像からテキスト断片を抽出し，その断片から h2v2 または 3-gram 方式で検索語を抽出して検索を行う．以降は評価実験 1 と同様である．正しいページ ID を一意に特定できた場合のみ，検索が成功したと判定する．

なお，撮影状況によっては OCR 処理で抽出されたテキスト断片の文字数が 3 文字未満の場合もわずかながらあった．この場合，各キーが 3 文字からなる h2v2/3-gram はともに元位置の特定が原理上不可能なので，撮影をやり直した．撮影は実験者 1 名が行ったが，実験者の意図が撮影方法に影響を与えないよう，撮影画像から検索を行う方式（h2v2/3grams）は撮影のつどランダムに決定されるよう実験システムを構築した．ページ内の様々な位置を満遍なく 300 回撮影し，各方式でそれぞれ 150 回の検索を実施した時点で実験を終了した．

7.3 実験結果

表 7 は各方式における検索の正解率を示しており，括弧内は（問合せ結果の正解数/問合せ回数）を表している．たとえば，局所領域画像から正しく認識された文字数が 12 文字であるケースは 300 回の試行中 $15 + 19 = 34$ 回発生しており，そのうち 15 回は h2v2 で検索が行われて 14 回成功（正解率 0.93），19 回は 3-gram 方式で検索が行われて 9 回成功（正解率 0.47）であったことを表している．

全体の傾向を見ると，正しく認識された文字数が減るほど各方式とも精度は低下することが分かる．上述のとおり各局所領域画像内には 20 文字前後写っているので，我々が 4.1 節で想定したノイズ率 35%は，13 文字前後が正しく認識された場合に該当する．この付近を観察すると，正しく認識された文字数が 14 文字の場合は h2v2・3-gram ともほぼ 1.00 の正解率を示している．しかし，正しく認識された文字数が 13，12 文字と減少した場合，h2v2 の正解率はほぼ 1.00 のままであるが，3-gram の正解率は 0.29，0.47 と大幅に低下することが分かる．正しく認識された文字数がさらに減少しても h2v2 は高い正解率を保ち続けるが，7 文字程度しか認識されない場合から大幅に正解率を落とすことが分かる．

表 7　実験結果
Table 7　The evaluation result.

正しく認識された文字数	h2v2	3-gram
15 以上	1.00 (41/41)	1.00 (21/21)
14	0.94 (15/16)	1.00 (8/8)
13	1.00 (20/20)	0.29 (2/7)
12	0.93 (14/15)	0.47 (9/19)
11	0.90 (9/10)	0.40 (4/10)
10	1.00 (4/4)	0.36 (8/22)
9	1.00 (13/13)	0.40 (6/15)
8	1.00 (13/13)	0.36 (4/11)
7	0.60 (3/5)	0.11 (1/9)
6	0.50 (3/6)	0.11 (1/9)
5	0.75 (3/4)	0.00 (0/11)
4	0.00 (0/2)	0.00 (0/4)
3	1.00 (1/1)	0.00 (0/4)

7.4 考　　察

実験結果より，20 文字程度しか含まれていない局所領域画像をクエリとして数万オーダのページ群から元ページを特定するタスクにおいて，実際の OCR の誤認識傾向やプロポーショナルフォント文字の影響が介在する現実環境であっても，提案方式が高い正解率とロバスト性を示せたといえる．ただし，すべてのケースを十分に試行できたわけではないので（たとえば，10 文字が正しく認識されたケースは h2v2 では 4 回しか試行できていない），統計的有意な判断を下すためにはさらに試行回数を増やした検証が必要であり，これは今後の課題である．

また，実際に撮影した画像を入力して検索結果を得るタスクを行ったことによる気付きとして，提案方式と既存方式の検索速度の差に言及したい．h2v2 は画像がシステムに投入されてから検索結果が得られるまでの処理時間は今回のサーバ構成では数百 msec であったが，3-gram は数秒から数十秒を要していた．各方式のインデックスを調査すると，ユニー

クなキー数は h2v2 が 3-gram の 4 倍近くあることが分かった．本システムで採用している GIN インデックスを含む多くのインデックスの速度性能はユニークなキー数に比例するので，h2v2 の方が高速に検索できたものと思われる．これは，提案方式が各位置に固有なパターンを抽出できることの副次的な効果である．

8. 関連技術との比較

ここまでの議論，検証結果をふまえたうえで，2.3 節で言及した中でも特に本研究と関連が深い技術との比較を行う．

LLAH は利用シーン次第では精度・速度ともに本手法を上回る可能性があるが，本研究とは研究の前提や目標が若干異なっているため，我々が想定するシーンへの適用は難しい．具体的には，前述のとおり，LLAH は 1 枚の局所領域画像をクエリとし，数万オーダの候補の中から一意に撮影文書を特定するタスクにはフォーカスしていないと推測される．文献 27)，28) 内でも，1 文書を特定する際に Web カメラを用いて毎秒 10 回ほどの検索要求を行うアプローチをとっている．しかし，検索に Web カメラおよび PC を用いる方法では，3.1 節要件 4）にあるように書籍とともに持ち運ぶというシーンには適用しにくい．一方で本システムは，書籍の可搬性を損ねないためにクライアントとして携帯電話を用いることを前提としており，必然的にクエリ画像は携帯電話とサーバ間でパケット通信されることになる．現在の携帯電話の処理速度，および携帯電話通信インフラではリアルタイムの動画クエリ送受信は難しいので，1 枚の局所領域画像だけから元位置を一意に特定する技術が必要であり，本研究はこの実現を行っている．

WordAnchor は携帯端末を前提としておりコンセプトも優れているが，やはり本研究とは前提としている条件が異なる．すなわち，本研究のように撮影画像が局所領域であり，かつ手振れ・手の影などに起因する OCR 誤認識が多く発生する状況は前提としていない．OCR 誤認識が発生する局所領域画像が分析対象である場合，2.3 節で述べたとおり，正しく検出できる単語の隣接関係は少なく，原理上 WordAnchor が得意とする利用シーンではない．我々は 3.1 節要件 3）を満たすためにより狭い局所領域画像から元位置を特定することが重要だと考えているし，本システムを多くのユーザに試用してもらった過程で彼らの多くは書籍内を鮮明に撮影できないことを経験的に学んでいる．この点，提案手法では，検索キーの構成単位を単語ではなく文字としているので，撮影領域が狭く画像の端で単語が途切れていてもキー抽出には影響しない．また，キーの位置関係ではなく出現数で位置特定を行うため，撮影画像の不鮮明に起因する誤認識文字を含むキーが領域内に分散していても位置特定が行える．単語間の隣接関係を用いていないという点で WordAnchor 方式の完全な再現ではないが，単語をキーとする方式は本研究の前提条件において不利であることは評価実験 1 で実証している．

9. おわりに

本論文では，文章を読む方向とそれに直交する方向を考慮した 2 次元のブロックをキーとする 2 次元文字ブロック検索手法を提案した．大規模データを用いたシミュレーション実験を実施し，OCR 誤認識が発生する書籍内局所領域画像から位置を高精度に特定できることを実証した．さらに，提案手法を用いて書籍から格段に多くの情報を得られるシステム Kappan を実装し，実環境においても高精度で位置特定が行えることを実証した．このシステムでは，ユーザは書籍内の部分領域をカメラ付き携帯電話で撮影するだけで，その領域に関連付けられたコンテンツを閲覧できる．書籍中にマーカを記載する必要がないので，書籍が出版された後からでもコンテンツを登録できるし，コンテンツ登録のためにページ上に余分な領域を占有することもない．また，システム上でインデックス作成済みの書籍であれば，どのユーザの書籍からもコンテンツを閲覧できる．今後は，日本語以外の言語に対しても適用可能性を検証する方針である．

参考文献

1) 岡田謙一，松下　温：本メディアを超えて：BookWindow，情報処理学会論文誌，Vol.35, No.3, pp.468–477 (1994).
2) Card, S.K., Hong, L., Mackinlay, J.D. and Chi, E.H.: 3Book: A Scalable 3D Virtual Book, *Extended Abstracts of the 2004 Conference on Human Factors and Computing Systems*, pp.1095–1098 (2004).
3) 島田恭宏，宇都宮毅，鏡原篤男，田中　圭，島田英之，大倉　充，東　恒人：仮想書籍ブラウジングシステムの試作，情報処理学会論文誌，Vol.46, No.7, pp.1646–1660 (2005).
4) Data Matrix bar code symbology specification: ISO/IEC 16022:2006.
5) QR code 2005 bar code symbology specification: ISO/IEC 18004:2006.
6) Lange, B.M., Jones, M.A. and Meyers, J.L.: Insight Lab: An Immersive Team Environment Linking Paper, Displays, and Data, *Proc. 1998 Conference on Human Factors in Computing Systems*, pp.550–557 (1998).
7) Billinghurst, M., Kato, H. and Poupyrev, I.: The MagicBook: A Transitional AR Interface, *Computers and Graphics*, Vol.25, No.5, pp.745–753 (2001).
8) Metaio Inc. http://www.metaio.com

9) Mohan, A., Woo, G., Hiura, S., Smithwick, Q. and Raskar, R.: Bokode: Imperceptible Visual Tags for Camera Based Interaction from A Distance, *Proc. International Conference on Computer Graphics and Interactive Techniques*, pp.1–8 (2009).
10) Koike, H., Nishikawa, W. and Fukuchi, K.: Transparent 2-D markers on an LCD tabletop system Export, *Proc. 27th International Conference on Human factors In Computing Systems* (2009).
11) Park, H. and Park, J.-I.: Invisible Marker Tracking for AR Export, *Proc. 3rd IEEE/ACM International Symposium on Mixed and Augmented Reality*, pp.272–273 (2004).
12) Hecht, D.L.: Printed Embedded Data Graphical User Interfaces, *IEEE Computer*, Vol.34, No.3, pp.47–55 (2001).
13) Kise, K., Miki, Y. and Matsumoto, K.: Backgrounds as Information Carriers for Printed Documents, *Proc. 15th International Conference on Pattern Recognition*, Vol.4, pp.380–384 (2000).
14) Silberman, S.: The Hot New Medium: Paper – How the oldest interface in the book is redrawing the map of the networked world, *Wired Magazine*, No.9.04, pp.184–191 (2001).
15) Lu, X. and Lu, Z.: A Publishing Framework for Digitally Augmented Paper Documents: Towards Cross-Media Information Integration, Lecture Notes in Computer Science, Vol.4261, pp.494–501 (2006).
16) Luff, P., Heath, C., Norrie, M., Signer, B. and Herdman, P.: Only Touching the Surface: Creating Affinities Between Digital Content and Paper, *Proc. 2004 ACM Conference on Computer Supported Cooperative Work*, pp.523–532 (2004).
17) Arai, T., Aust, D. and Hudson, S.E.: PaperLink: A Technique for Hyperlinking from Real Paper to Electronic Content Export, *Proc. ACM Conference on Human Factors in Computing Systems*, pp.327–334 (1997).
18) 田村秀行，池田克夫：知能情報メディア，総研出版 (1995).
19) 長坂晃朗，田中　譲：カラービデオ映像における自動索引付け法と物体探索法，情報処理学会論文誌，Vol.33, No.4, pp.543–550 (1992).
20) Vinod, V.V. and Murase, H.: Focused Color Intersection with Efficient Searching for Object Extraction, *Pattern Recognition*, Vol.30, pp.1787–1797 (1997).
21) Taketa, N., Hayashi, K., Kato, H. and Noshida, S.: Virtual Pop-Up Book Based on Augmented Reality, Lecture Notes in Computer Science, Vol.4558, pp.475–484 (2007).
22) Murase, H. and Vinod, V.V.: Fast Visual Search using Focused Color Matching – Active Search, *Systems and Computers in Japan*, Vol.31, No.9, pp.81–88 (2000).
23) Erol, B., Antúnez, E. and Hul, J.J.: HOTPAPER: Multimedia Interaction with Paper using Mobile Phones, *Proc. ACM International Conference on Multimedia*, pp.399–408 (2008).
24) Liu, X. and Doermann, D.: Mobile Retriever – Finding Document with a Snapshot, *Proc. 2nd International Workshop on Camera-Based Document Analysis and Recognition*, pp.29–34 (2007).
25) Nakai, T., Kise, K. and Iwamura, M.: Hashing with Local Combinations of Feature Points and Its Application to Camera-Based Document Image Retrieval – Retrieval in 0.14 Second from 10,000 Pases, *Proc. 1st International Workshop on Camera-Based Document Analysis and Recognition*, pp.87–94 (2005).
26) Nakai, T., Kise, K. and Iwamura, M.: Use of Affine Invariants in Locally Likely Arrangement Hashing for Camera-Based Document Image Retrieval, Lecture Notes in Computer Science, Vol.3872, pp.541–552 (2006).
27) Nakai, T., Kise, K. and Iwamura, M.: Real-Time Retrieval for Images of Documents in Various Languages using a Web Camera, *Proc. 10th International Conference on Document Analysis and Recognition*, pp.146–150 (2009).
28) 中居友弘，黄瀬浩一，岩村雅一：Web カメラを用いた多言語文書画像のリアルタイム検索システム，電子情報通信学会技術研究報告，Vol.108, No.432, pp.115–120 (2009).
29) 嶺　竜治，亀山達也，高橋寿一，古賀昌史，緒方日佐男：文字認識と単語レイアウト解析を用いた紙文書とデジタルデータの情報リンク手法，Vol.J92-D, No.6, pp.868–875 (2009).
30) Frakes, B. and Baeza-Yates, R.: *Information Retrieval: Data Structures and Algorithms*, Prentice Hall (1992).

(平成 22 年 6 月 21 日受付)
(平成 23 年 1 月 14 日採録)

推　薦　文

本論文は，文章を読む方向とそれに直交する方向を考慮した 2 次元のブロックを索引・検索のキーとする 2 次元文字ブロックインデクシング技術を提案し，書籍内の各位置にデジタルコンテンツへのハイパーリンク設置を可能にするシステム BookEnhancer（現 Kappan）を紹介したものである．英文に比べて縦列が揃いやすい日本語文章に適した手法として斬新であり，また，部分イメージを用いた検索に比べて，辞書サイズやサーチ時間の大幅な短縮が可能なことから，実用性も非常に高く，学術論文としてきわめて高品質であると評する．なお，本論文はインタラクション 2010 シンポジウムにてベストペーパーを受賞しており，多数の研究者から支持を得ている．よって，本特集号推薦論文として推薦する．

（インタラクション 2010 シンポジウムプログラム委員長　戸田真志）

宮田　章裕（正会員）

日本電信電話株式会社 NTT サイバーソリューション研究所研究員．2005 年日本電信電話（株）入社．2008 年慶應義塾大学大学院理工学研究科博士課程修了．ヒューマンインタフェース，対人コミュニケーション，ソーシャルメディアの研究開発に従事．平成 19 年度情報処理学会山下記念研究賞．IADIS WWW/Internet 2008 Best Paper，インタラクション 2010 ベストペーパー賞等．博士（工学）．日本データベース学会会員．

塩原　寿子（正会員）

日本電信電話株式会社 NTT サイバーソリューション研究所主任研究員．1992 年大阪大学理学研究科物理学専攻博士前期課程修了．同年日本電信電話（株）入社．データビジュアライゼーションの研究開発に従事．

藤村　　考（正会員）

日本電信電話株式会社 NTT サイバーソリューション研究所主幹研究員．1989 年北海道大学大学院工学研究科情報工学専攻博士課程修了．同年日本電信電話（株）入社．ソーシャルメディア分析・可視化の研究開発に従事．工学博士．電子情報通信学会，日本データベース学会会員．