

物体認識技術に基づく実物体からのリンク機能の実現

三宅 弘志[†] 黄瀬 浩一[†]

[†] 大阪府立大学大学院工学研究科 〒599-8531 堺市中区学園町1-1
E-mail: [†]miyake@m.cs.osakafu-u.ac.jp, [†]kise@cs.osakafu-u.ac.jp

あらまし リンク機能は、インターネットを代表とする電脳世界でしばしば用いられる機能である。実世界に存在するリンク機能について考えてみると、バーコードやRFID等、物体を手掛かりとして情報を得る技術は存在する。しかし、これらの技術を利用しても情報の登録に手間がかかるため、個人が気軽に使用できるとは言えない。そこで本稿では、個人ユーザが手軽に利用できる、リンク機能を実世界に拡張するシステムを提案する。本システムには、本稿で述べる考察の結果から得られる条件を満たす物体認識技術を用いる。そして既存技術では実現されていない受動的な情報獲得に対応するために、複数物体の認識、物体の位置の推定、処理の高速化を行う。個人ユーザがこのシステムを利用する様々な場合を想定した実験を行い、本システムの有用性を確認した。

キーワード 物体認識, リンク, locally likely arrangement hashing, 受動的情報取得

Building links from real-world objects based on object recognition technologies

Hiroshi MIYAKE[†] and Koichi KISE[†]

[†] Graduate School of Engineering, Osaka Prefecture University
1-1 Gakuencho, Naka, Sakai, 599-8531 Japan
E-mail: [†]miyake@m.cs.osakafu-u.ac.jp, [†]kise@cs.osakafu-u.ac.jp

Abstract The link function makes the cyber world such as the Internet attractive. Technologies such as bar codes and RFIDs enable us to use the link function for objects in the real world. Although users can follow the links provided by these technologies, they cannot easily make new links; the link function for objects are partially available. In this report, we propose a new system for the full link function for objects. For the better use of the link function, it is also required to find links on objects without pointing them by the camera (passive acquisition of information). The system provides the passive acquisition with the help of multiple object recognition, estimation of object positions, and real-time processing. Experimental results show that users feel that the system is useful for establishing and finding links from objects in the real world.

Key words Object recognition, Link, Locally likely arrangement hashing, Passive acquisition of information

1. はじめに

近年のインターネットの普及の理由として、リンク機能が挙げられる。この機能により、ユーザはわざわざアドレス等を打ち込むことなく、マウスをクリック一つで目的の情報を簡単に手に入れることができる。現在、このリンク機能は主にインターネットを含むサイバー空間という閉じた世界でのみ実現されており、現実世界を十分に扱うことが出来ない。もし実世界への拡張が実現すれば、ユーザは物体を手掛かりとして目的の情報を得ることが可能となる。これは例えば、印刷文書から電子文書を得る、電子機器のマニュアルを呼び出す、商品を手掛かり

としてその商品のホームページを開く、直接情報を付加することのできない記念品に購入日や場所情報を登録する、本の挿絵や写真の詳細情報を得るといったことを意味する。さらに、これを利用して物体の属性を記述しておくこと、情報の効率的かつ効果的な分析・管理が可能となる情報処理技術を用いることもできる。このため、物体からのリンク機能の実現は有用な技術であるといえる。

現在、現実世界で実現されているリンク機能の一つにバーコードやRFIDといった物体に直接タグを付加する技術がある。しかしこれらは、情報の登録対象が限定的であり、RFIDにおいては情報の読み取りに特殊な機器を必要とするため、一

般のユーザが手軽に利用できるとは言えない。一方、物体認識技術を用いることで物体に直接タグを付加しないサービスもある。これは、あらかじめ登録したい物体画像と関連情報をネットワーク上のデータベースに登録しておき、携帯電話のカメラを利用して検索したい物体や情景の撮影画像を計算機サーバに送信し、その結果を得るといったサービスである [3] [4]。これらのサービスはタグの付加を必要とせず、カメラ付き携帯電話があれば利用可能であるため、物体に直接タグを付加するものよりも手軽に使用できる。ただし、これらのサービスは、能動的情報取得、すなわちユーザが指し示した物体から情報を読み取るという方式を採用しているため、ユーザはどの物体にリンクが設定されているのかを事前知っている必要がある。したがって、どの物体にリンクが設定されているのかが分からない状態では、十分利便性の高いリンク機能を実現することは難しい。

この問題を解決するためには、受動的情報取得、すなわち物体を指し示すことなく情報が取得できることが可能となればよい。そのような情報取得が可能なシステムは、ユーザの周囲の環境を常にセンシングしており、リンクが設定されている物体を発見するとユーザにその旨を知らせる。ユーザはその中から自身に必要な情報を取得する。受動的情報取得が可能であれば、能動的情報取得も可能となるため、より利便性が増すと言える。

受動的な情報取得を実現するためには、現在、能動的情報取得では用いられていない様々な技術が必要となる。具体的には、撮影画像中に複数の登録物体が写っている場合や部分的にしか写っていない場合にも対処する必要がある。また、ユーザが物体のリンクを辿るためには、画像中の物体の位置をシステムが把握している必要がある。さらに、ユーザが動いても機能が提供するためには、実時間処理が求められる。また、我々の周囲にある様々な物体を扱うことができ、容易にリンクを設定できることも重要な条件となる。

そこで本稿では、これらの要求仕様を満たすリンク機能を、物体認識技術の改良によって実現する手法を提案する。まず、文書を含む様々な物体を扱うために複数の認識技術を援用する。これらの技術は高速な認識が可能であるだけでなく、物体を新たに登録する処理も高速であるという特徴を持つ。また、これらの技術を改良し、複数物体の認識、画像中の物体の位置推定の機能を持たせる。さらに、ストレスのない受動的情報取得を可能とするために、トラッキング技術を用いた処理の高速化についても述べる。

2. リンク機能の要求仕様

リンク機能の実世界への拡張が実現すると、色々なサービスが考えられる。これは例えば、記念品であればその購入日や購入場所、電子機器であればその使用方法を表示、資料から関連資料や電子文書を検索、看板やランドマークの利用による観光案内などである。そこで、様々な視点から検討をいリンク機能の実現に欠かせない条件を探る。

2.1 リンク機能に関する考察

リンク機能を実現するにあたってまず考えなければならない

ことは、リンクを設定する対象物体である。リンク機能が実現した際、ユーザは例えば記念品に情報をメモ、電機製品にマニュアル、印刷文書に関連資料や電子文書ファイル、看板に案内情報のリンクを設定することが考えられる。このことから、電気製品、印刷文書や看板といった身の回りにある物体を可能な限りリンク対象に含める必要がある。

次に、物体から情報を得る形態に対する考察が必要となる。情報の獲得方法としては、ユーザが情報の欲しい対象に対して機器を向けて情報を獲得する能動的な情報取得と、ユーザが対象を限定せず機器からの入力を基に登録物体があればその情報を提示する受動的な情報取得が考えられる。能動的な情報取得とは、撮影画像が入力として与えられる度にリンクの設定や読み込みを行うものである。これに対し、受動的な情報取得では動画画像が入力として与えられるため、リンクの設定や読み込みを常に行うものである。このため、受動的な情報取得が可能であれば能動的な情報取得も可能である。受動的な情報取得が実現すると、ユーザは登録物体を意識することなく登録物体から情報を獲得することが出来るため、実世界の物体を手掛かりとして Web サイトと同様にワンクリックで必要な情報を得ることが可能となる。つまり、受動的な情報取得を用いると能動的な情報取得よりも多くの場面でリンク機能を活用することが可能となる。

最後にリンクの設定と読み込みの方法について考察を行う。物体から情報を得る手段としては、商品の管理や値段の読み取りに用いられるバーコード、電波を利用して IC チップから情報を読み取る RFID (Radio Frequency Identification) やコンピュータビジョンの分野で研究が行われている物体認識の手法などがある。リンク機能には、大きく分けて物体にリンクを貼る登録処理と物体を手掛かりとして情報を得る検索処理の 2 つの処理が必要となる。このため、それぞれの処理に分けて考察を行う。まず、リンクの設定について考察を行う。バーコードや RFID といった物体に直接タグを付加する技術を用いると記念品へのメモ機能の場合、物体に直接情報を付加する方法を用いると外観を損なってしまう恐れがある。このため、物体そのものに手を加えない情報の付加が必要となる。さらに、情報の登録対象が限定的であり、RFID においては情報の読み取りに特殊な機器を必要とするため、一般のユーザが手軽に利用できるとは言えない。このためリンクの設定・読み込みには、専門的な機器ではなく一般のユーザが簡単に手に入れることのできるものを使用することが好ましい。次に、リンクの読み込みについて考察を行う。受動的な情報取得では、登録物体が同じ画像中に複数存在することが考えられる。このため、同時に複数のリンク情報を読み込むことと登録物体の位置がわかることが必要となる。

2.2 要求仕様

以上の議論から、画像認識技術を用いた受動的な情報取得とリンク設定によって様々な対象を扱うことが必要であると言える。受動的な情報獲得を実現するには、複数物体の認識、物体の位置の推定や動画画像に対応できる高速な処理が必要となる。これらを解決するため、画像検索技術の中でも、特に局所特徴量を用いる手法に着目する。局所特徴量とは、画像の局所領域の輝度

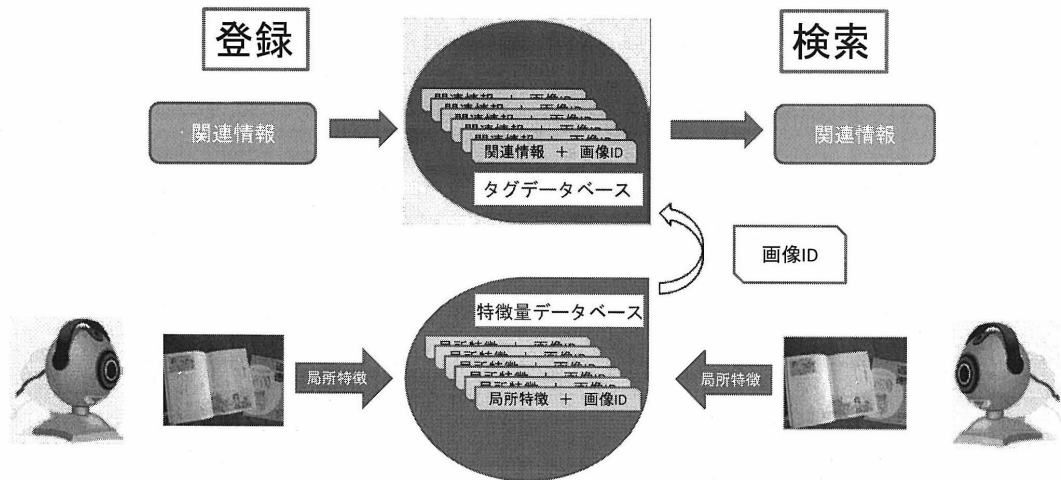


図1 処理の流れ

変化等の情報を基にして得られるものである。この際、局所特徴量には安定性と識別性が求められる。安定性とは撮影条件の変化によらず物体上の同じ部分から特徴が検出される性質であり、識別性とは異なる特徴が同一のものと同判別されない性質である。この局所特徴量を用いると、特徴点の対応関係から射影変換パラメータの推定が可能となり、撮影画像上のどの部分に登録物体が写っているか判定することが可能となる。さらに、画像の様々な部分から複数の局所特徴量が得られるため、物体の一部の隠れに強く複数物体の認識も可能といった利点がある。これらのことから、局所特徴量を用いる物体認識手法によって受動的情報取得が可能となることがわかる。

3. リンク機能を実現する処理

3.1 処理の流れ

ウェブカメラを用いるリンク機能の基本的な処理の流れを図1に示す。リンク機能の基本的な処理として、情報の登録と検索の2つの処理が必要となる。

まず登録処理について述べる。登録処理では、ウェブカメラから得られる画像を入力とし、複数の局所特徴量を抽出する。そして各特徴と画像IDを特徴データベースに登録する。さらに、ユーザが撮影画像に写っている物体に対して登録したい関連情報を画像IDと共にタグ情報データベースに登録する。

次に、検索処理について述べる。検索処理では、登録処理と同様にウェブカメラから得られる画像を入力とし、複数の局所特徴量を抽出する。そして、得られる個々の局所特徴量とデータベースの中の局所特徴量を照合し、データベースに登録されている画像IDを検索結果として返す。こうして得られる画像IDを基に、タグ情報データベースの中から対応するタグ情報を検索し、情報を得る。こうしてカメラから得られた画像を表示しながら関連情報を表示することが可能となる。

3.2 特定物体認識手法

リンク機能を実現する上で必要な物体認識手法は、物体のインスタンスを対象とした認識、すなわち特定物体の認識である。したがって、2.での考察から、局所特徴量を用いる高速かつ高精度な特定物体認識の手法が必要となる。そこで、本システムではこれらの条件を満たす特定物体認識法として野口らの手法[1]を用いる。野口らの手法は平面物体を対象とする登録・検索手法であり、局所特徴量としてPCA-SIFT[5]を用いる。そして、このPCA-SIFT特徴量のデータ特性を活かしたハッシュ関数を用いて高速な検索を実現している。さらに、この手法ではデータの登録時に特別な処理を必要としないため、登録処理も高速となる。処理時間に関しては、使用する際のパラメータにもよるが、特徴量抽出にかかる時間が多くの時間を占める。

野口らの手法を含む、局所特徴量を用いる特定物体認識には、一つの問題点がある。これは、見え方の似通っている対象を十分識別することが容易ではないというものである。特に文書は、認識が必要であるにもかかわらず困難な対象といえる。そこで本システムでは、文書については専用の認識手法（文書画像検索法）を導入し、これと特定物体認識法をシステムレベルで統合することでこの問題を解決する。

本システムでは、局所特徴量を用いる高速かつ高精度な文書画像検索技術として、中居らの手法（Locally Likely Arrangement Hashing）[2]を用いる。この手法では、局所特徴量として各単語領域の重心を特徴点とし、近傍の特徴点との位置関係から得られるアフィン不変量を用いる。そしてこの特徴量を用い、ハッシュを通じて特徴量を比較する。これにより高速かつ高精度な文書画像検索を実現している。処理時間に関しては、特徴量の抽出が支配的となる。

3.3 複数物体の認識

システムは、受動的な情報取得を可能にするため、図2で示



図2 複数物体

ように、複数の物体を同時に認識する必要がある。用いる2つの手法はともに局所特徴量を利用し、投票処理によって認識結果を得るものである。両手法とも、用いる特徴量の安定性と識別性が高いため、誤投票によって誤った物体に票が集まることは少ない。すなわち、撮影画像中に登録物体があればその物体に対して票が集中し、そうでないものにはほとんど票が入らないといえる。これらのことから、十分な数の特徴点が得られる場合、得票数に閾値 t を設け、閾値以上のものを認識できたとすれば複数物体の認識が可能となる。

予備実験として QVGA の画像を対象に $t = 5$ として処理を行ったところ、得票数が少なく物体が認識されない例はあるものの、誤った認識はほとんど起こらなかった。このため、画像サイズによって値を変更する必要があるが、閾値によって複数物体の認識を判断することは可能である。

3.4 物体の位置推定

登録物体が複数写っている場合、撮影画像のどこに登録画像が写っているのかわからなければ、どの物体から得られたリンク情報がわからなないといった問題が起こる。このため物体の位置の推定が必要となる。本システムで用いる特定物体認識の手法は共に局所特徴量を用いるため、認識の副産物として特徴点間の対応を求めることが可能である。このため、撮影画像と登録画像との対応点に対して RANSAC [6] を用いて射影変換パラメータを求めることで、図2に示すように、それぞれの登録物体が撮影画像のどの部分にどのような姿勢で写っているかを推定することができる。

3.5 処理の高速化

特定物体認識の処理において、最も処理時間を要するのは特徴抽出である。従って、この処理時間を低減すれば処理のリアルタイム化に繋がる。本システムでは、ウェブカメラからの画像を入力として処理を行うため、主に動画処理に用いられる特徴点追跡の手法を用いることが可能である。このためその処理に、KLT-Tracker と呼ばれる特徴点追跡手法を組み込み、各特徴の移動量を測る。そして移動量が閾値を超えた場合、特徴

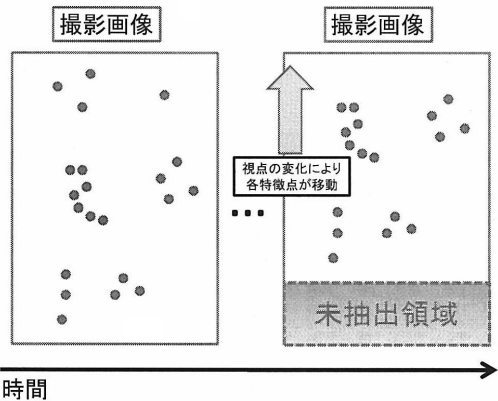


図3 未抽出領域の決定

の未抽出領域を決定し、その領域からのみ特徴抽出を行うことで、処理時間の短縮を図る。こうすることで、一度抽出した領域から再度特徴を抽出する必要がなくなり処理の高速化が可能となる。

未抽出領域は以下のようにして決定する。KLT トラッカーによる特徴点の追跡によって、各特徴点がどのように移動したかが求められるようになる。本システムではこれを利用して、特徴追跡を行う。そして、図3に示すように、撮影画像の一定の領域内に特徴点が存在しなくなった場合、その領域を特徴の未抽出領域とする。この特徴未抽出領域からのみ特徴抽出を行うことで、処理コストの大きい特徴抽出の処理時間を減らすことができる。

本システムでは、特徴点の移動量の閾値を VGA の画像に対して 50 ピクセルとし、上下左右のいずれかの移動量が閾値以上となった場合、その移動に応じて未抽出領域を決定し、特徴抽出を行うように実装する。

4. システムの使用例

本システムの作成にはマイクロソフト社の開発環境のひとつ Visual C++ のクラスライブラリ MFC (Microsoft Foundation Class) を用いた。以下に、実装する機能の処理例を示す。

4.1 基本的な処理

基本的な処理として物体への情報の登録・検索の例を挙げる。撮影画像と結果画像の表示の例を図4、図5に示す。(1)は撮影画像、(2)は結果画像を表示している。これからわかるように、撮影画像は対象物体の一部が画像外にある、登録画像以外の物体が写っている、照明条件や視点が登録画像と異なるといった様々な条件の変化があっても認識出来ている。

4.2 リンク機能に関する処理

本システムの使用例を図6、図2に示す。これは、撮影画像中に登録物体が写っている場合、登録物体とデータベースに登録されている画像との射影変換パラメータを計算し、物体を囲む枠を描画している。このように、対象が文書であっても物体

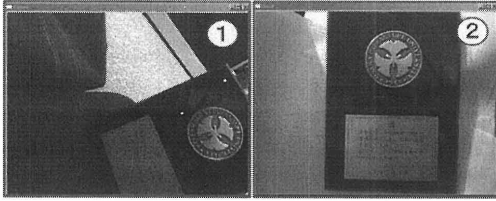


図4 キャプチャ画像と結果画像(物体)

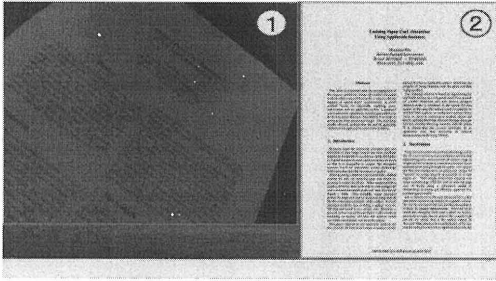


図5 キャプチャ画像と結果画像(文書)

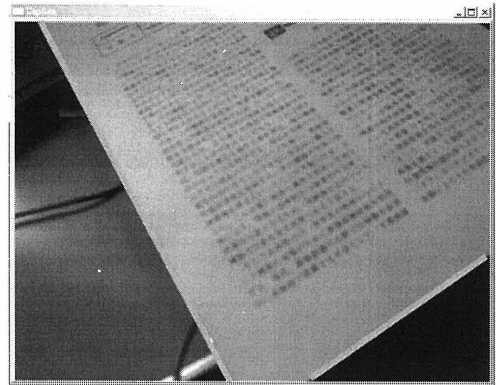


図6 文書

であっても認識できていることがわかる。さらに、図2に示すように、複数物体の認識も可能である。

また、この枠内をクリックすることによって登録している関連情報を読み込み、拡張子に合わせたアプリケーションが起動するようになっており、物体を手掛かりとするリンク機能を実現している。これは例えば、いくつかの物体に対して製造メーカーのURLを登録しておけば、図6、図2のように登録物体の周りに枠を描画し、これをクリックすることにより登録しておいたURLをブラウザで開くといった機能である。

5. 実験

5.1 実験条件

作成したシステムの有効性を検証するため、情報系の院生5人を被験者としていくつかの実験を行いシステムを評価した。

共通する条件として、複数物体の認識の際、得票数の閾値 t を5とした。

以下に実験の内容を説明する。

実験1: 実験1では、登録に関する評価を行う。実験内容は、ユーザが物体に任意の情報(URL、ファイルパス等)を登録するといったものである。評価の指標としては、本システムの利便性について便利・どちらでもない・不便の3段階、本システムの操作性について、簡単・どちらでもない・難しいの3段階、本システムの処理速度について、早い・どちらでもない・遅いの3段階、を設定する。

実験2: 実験2では、ファイルアクセスに関する評価を行う。実験内容は、検索エンジンやエクスプローラ等を用いてファイルにアクセスする場合と本システムを用いてファイルにアクセスする場合の比較とする。実験2ではまず、特徴追跡による処理の高速化を用いずに2件検索し、次に高速化を用いて2件検索する。さらに、本システムを用いずに検索エンジンやエクスプローラ等を用いて2件検索する。これを1セットとし、各々の検索件数がそれぞれ10件となる計5セットを行った。但し、システムを使用しない場合、ユーザは初めからファイルの存在する場所を知っている状態とする。評価の指標は実験1の指標に加えて、インターフェイスを用いる場合と用いない場合でのファイルアクセスにかかった時間の平均値を設定する。

実験3: 実験3では、実際のサービスを想定し、複数の物体が印刷された紙または地図を用いる情報検索を行う。実験内容は、まず、図7、図8に示すような紙に印刷されている物体にいくつかの情報を登録しておき、次に登録された情報を表示するといったものである。具体的には、図7の場合、紙に印刷されたそれぞれの画像に対してパス、時計、カセットプレイヤー、富士山という情報を登録した。また図8の場合、画像の一部を撮影し、それぞれの建物の番号を登録した。これらの情報は、実際にサービスを行う際には、URL等に置き換えられるものである。実験3では、特徴追跡による処理の高速化を用いる場合と用いない場合それぞれ2件ずつの検索を1セットとし、これを計5セット行った。実験の条件と指標は、実験1と同様である。

5.2 実験結果

実験の結果を表1～表6に示す。表1から、登録処理については特に評価の低い点が無いことがわかる。しかし、表2や表5から、検索時には低い評価が目立った。一方、表3や表6から、特徴追跡による高速化を行うと、高速化を行わなかった場合よりも操作性、処理速度に対する評価が向上し、同じ検索であるにも関わらず低い評価が減少した。これは、特徴量の抽出処理時間が大きくストレスを感じていたユーザが、高速化によりストレスを感じにくくなったためといえる。

さらに、表4から本システムを使用する方が、しない場合よりも目的のファイルへアクセスするまでの時間が短かった。本実験では実験1の登録の直後に実験2,3を行ったため、検索対象となるファイルの場所はユーザの記憶に新しくなったと思われる。実際に使用する場合には、ファイルの場所を失念することも考えられるため、上記の時間の差は、より大きく広がると考

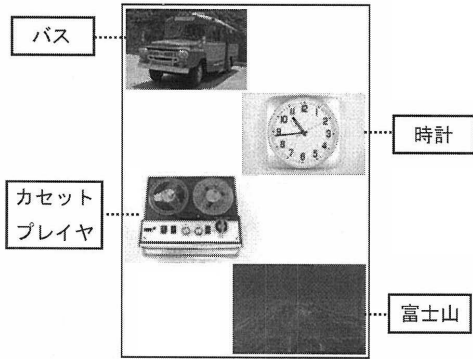


図7 複数の物体を印刷した紙

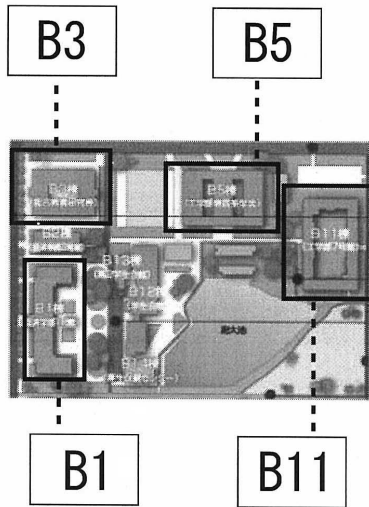


図8 地図

えられる。

6. まとめ

本稿では、物体からのリンク機能の実現を目的とし、ウェブカメラを用いたリンク機能を持つインターフェイスを開発した。物体への情報の付加・読み取りには、特定物体認識の手法と文書画像検索の手法を用いた。そして、処理の高速化のために KLT トラッカーを用いて特徴追跡を用い、特徴の未抽出領域からのみ特徴抽出を行うことで、処理の高速化を図った。また、被験者 5 人による実験を行い、システムと高速化処理の有効性を示した。今後の課題として、照明条件の変化によりロバストであるとされる mean-shift トラッカー等を導入することが挙げられる。

文献

- [1] 野口和人, 中居友弘, 黄瀬浩一, 岩村雅一, “特徴ベクトルの近傍探索と物体認識の効率に関する実験的検討,” 電子情報通信学会技術研究報告, vol.106, no.PRMU-229, pp.57-64, Sept.,

表1 実験1 結果:情報登録

	利便性	操作性	処理速度
良い	3	1	2
どちらでもない	2	4	3
悪い	0	0	0

表2 実験2 結果:高速化を用いない検索

	利便性	操作性	処理速度
良い	2	1	0
どちらでもない	2	2	2
悪い	1	2	3

表3 実験2 結果:高速化を用いる検索

	利便性	操作性	処理速度
良い	3	2	3
どちらでもない	2	3	2
悪い	0	0	0

表4 実験2 結果:処理時間

	平均検索時間 (s)
インターフェイスあり	2
インターフェイスなし	5

表5 実験3 結果:高速化を用いない検索

	利便性	操作性	処理速度
良い	2	1	0
どちらでもない	3	1	2
悪い	0	3	3

表6 実験3 結果:高速化を用いる検索

	利便性	操作性	処理速度
良い	4	3	3
どちらでもない	1	2	2
悪い	0	0	0

2006.

- [2] 中居友弘, 黄瀬浩一, 岩村雅一, “Web カメラを用いたリアルタイム文書画像検索とその拡張現実への応用,” 画像の認識・理解シンポジウム (MIRU2006), pp.1379-1380, July, 2006.
- [3] <http://4965.jp/>
- [4] <http://www.techcrunch.com/2008/09/09/tc50-sekai-camera-for-social-tagging-on-the-iphone/>
- [5] Y. Ke, R. Sukthankar, Pca-sift: A more distinctive representation for local image descriptors, 2004.
- [6] M. A. Fischler and R. C. Bolles, “Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography,” Commun. ACM, vol.24, no.6, pp.381-395, June, 1981.