



携帯カメラを用いたユビキタス情報インタフェース

NEC メディア情報研究所

山田 敬嗣

kg-yamada@cp.jp.nec.com

NEC インターネットシステム研究所

仙田 修司

s-senda@ap.jp.nec.com

本稿では、実世界の情報とデジタル情報とのインタフェースとして、実世界情報の豊富さと自然で慣れた使い勝手に合わせて、デジタル情報の検索しやすさや適応性を組み合わせたユビキタス情報インタフェースについて述べる。実世界情報をトリガーとしてデジタル情報を活用するために、RFIDや2次元バーコードが利用されているが、文字認識技術を利用することで文字情報をトリガーとしたインタフェースを実現できる。具体例として、携帯カメラを利用して、実世界の文字情報から関連するデジタル情報を活用するインタフェースを紹介する。カメラで単語を写すことで、辞書引きやインターネットでのキーワード検索ができる。雑誌に記載されたURLを写すだけで、ダイレクトにアクセスできる。

● 日常生活に埋もれた情報 ●

日本の電車の中は五月蠅い。音の問題ではなく、膨大な提供情報があるという意味である。某日、電車内で数えたら、車両内に112枚もの広告が張ってあった。その中でURLアドレスが記載されていなかったのはたかだか10枚だった。ふと興味を持つ広告があり、アドレスをメモした。しかし、帰宅したときには忘れており、結局アクセスはしなかった。

繁華街を歩くと夥しい数の看板や広告に眩暈がするほどである。興味のある店の名前や電話番号、気になる商品名などが次々と目に飛び込んでくる。目の前を歩いている人が着ている服が素敵だ。どこで買ったのだろう、もっと詳しい情報が欲しいと思うが、諦めた。

このように実世界には膨大な情報が提示されている。また、その先には関連する膨大なデジタル情報がある。公式な紹介Webページもあれば、Blogとして記載された評判情報もある。興味を持ったとき、必要となったそのときにタイムリーに利用できることが望ましい。しかしながら、現状では、実世界の情報とデジタル情報とは切り離されており、Webページの興味のある部分をクリックすれば関連情報が提示されるというような簡便さはない。人間が媒介となって、実世界情報をデジタル情報として再入力して、関連情報を得るしかない。本稿では、実世界の情報とデジタル情報を、自然なユーザインタフェースで同時に活用するための技術について紹介する。これをユビキタス情報インタフェースと呼び、

特に、画像認識技術を利用して、実世界のタグなどに記載された文字情報からデジタル情報へのリンク生成について述べる。

● ユビキタス情報インタフェース ● — 実世界と情報空間との融合 —

ユビキタス情報インタフェースとは、図-1に示すように、ユーザを取り巻く実世界の情報とデジタル情報とのインタフェースである。

実世界には、すでに印刷された文書としての情報だけではなく、看板や掲示情報、さらにはモノや環境状態としての情報もある。行動や話し言葉も重要な情報である。一方、デジタル情報は、世界中のコンピュータに蓄

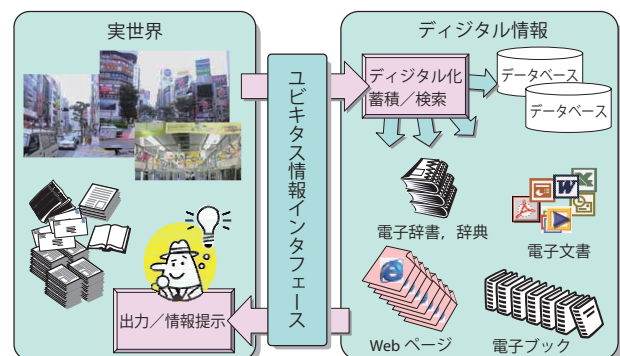


図-1 ユビキタス情報インタフェース

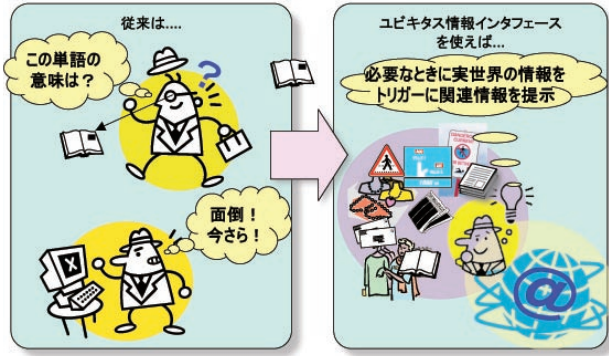


図-2 ユビキタス情報インタフェースは実世界の情報とデジタル情報を融合して必要な情報をタイムリーに提示できる

えられた情報であり、具体的には膨大な数の Web ページや、データベースに登録された情報である。

ユビキタス情報インタフェースは、実世界情報とデジタル情報との双方向のインタフェースであり、コンピュータ側から見た入出力機能を持つ。

入力機能は、実世界の情報をデジタル情報に変換するだけでなく、特定の実世界情報から関連するデジタル情報へのリンクとして利用できる。たとえば、ショッピングのときに、商品を手にとると、その詳細なスペックの説明を提示する。商品棚に付いたディスプレイに表示されてもよいし、自分の携帯端末に表示されてもよい。手に取った商品から、その詳細な説明や、色違いや異なるデザインの照会もできる。

出力機能は、選択した情報をディスプレイやスピーカなどの表示デバイスを経由して実世界に提示する。アクチュエータやロボットを通して、実世界の状態を変えることも含む。拡張現実感技術(Augmented Reality)では、情報を実世界のものや環境に重ね合わせて表示することでより自然な出力を実現する。

ユビキタス情報インタフェースを実現する上で、ユーザは特別な注意や意識をすることなく自然に情報を利用できなければならない。コンピュータグラフィックスのように、ユーザをコンピュータの世界に取り込むのではなく、ユーザが普段生活している実世界や利用しているものを自然に利用するかたちで情報インタフェースを実現する(図-2参照)。つまり、実世界とデジタル情報とを融合することで実現するインタフェースである。

● 実世界情報からデジタル情報へのリンク ●

実世界情報とデジタル情報を融合して新たな機能を実現するために、それぞれ特徴を図-3に示す。さらに、両者を融合して、実現できる新たなインターネットサービスについて検討する。

	例	メリット	デメリット
実世界情報	書物、新聞看板、環境音声、音楽表情、振舞い	慣れたメディアで、人間にとっては直感的に情報利用できる	大量情報の利用に手間がかかる 関連情報をすぐには見つけられない
デジタル情報	電子化文書 Web文書 センサデータ	適切なキーワードを設定すれば、大量情報から必要情報を簡単に選択できる	必ずしも必要な情報がデジタル情報にない、操作性が悪い

図-3 実世界情報とデジタル情報

実世界情報は、長い間なれたメディアとして利用されるため直感的に情報を取得できる。たとえば、朝刊をばらばら眺めて昨日のできごとを大まかに知ることができる。標識や看板なども普段の生活の中で直感的に利用している。さらには、表情や動作・振舞いなどの暗黙的な情報も利用できる。しかしながら、大量の実世界情報を利用するには手間がかかるという問題がある。たとえば、多くのビジネスマンは、オフィスでいつも書類を探している。関連する情報を取得するのに、さらに時間がかかる。

一方、デジタル情報では、Webのようにタイトルをクリックすれば関連情報にジャンプできる。また、欲しい情報のキーワードを適切に与えることで、テキスト検索やジャンル分類などにより、大量の情報の中から迅速に必要な情報を選択できる。

実世界情報の直感的に情報利用できる点と、デジタル情報の関連情報へのリンクや検索機能というメリットを組み合わせれば新たなインターネットサービスを実現することができる。森羅万象、実世界のモノや出来事に携帯端末をかざすと、それに関するデジタル情報を即座に取り出せるようになる。

実世界からデジタル情報へジャンプのトリガーには、図-4に示すような情報を利用できる。

(1) GPSなどの絶対的な位置の計測結果から、その場所に関する情報へリンクをたどることができる。データベースに緯度・経度ともに、その地点に関する情報を記録しておき、測位結果に最も近い点の情報を検索する。測位結果の緯度・経度から、デジタル地図でその地点の地名や番地を取得し、類似の地名番地が記載された Web ページを検索することもできる。

位置情報のみならず、時刻やユーザの嗜好情報と組み合わせ、ユーザが求めている情報を推定した上で提示することで、提示情報を絞り込むこともできる。

位置情報を取得するためには、人工衛星を利用した測位システムだけではなく、屋内に設置した赤外線、超音波などの受信機の位置をあらかじめ取得しておき、同種の発信機を持ち歩くことでユーザの位置を計測することができる。発信機を設置して、受信機を携帯するシステムも可能である。また、無線 LAN のア



人間可読性	機械可読性	トリガーの種類
○	△	文字情報
○	△	記号情報
×	○	バーコード
×	○	電子タグ、非接触IC
×	○	センサ(位置情報、方位、相対位置、温度)

図-4 実世界からデジタル情報へのトリガーの種類

アクセスポイントを複数設置しておいて、その無線強度や位相のずれを計測することで、屋内での測位を実現するものもある。

- (2) 環境中に埋め込んだセンサを利用するのではなく、非接触型 IC タグを用いる方法がある。IC タグの ID 情報をリーダーで読み出して、データベースに問い合わせることで、関連するデジタル情報を取得することができる。IC タグを商品や部品につけることで、それにリーダーをかざすと、製品の管理情報や商品の詳細説明、部品のメンテナンス情報を即座に取り出すことができる。

IC タグの ID と関連デジタル情報のリンクを記録する方式は、ユビキタス ID やオート ID など現在策定中である。

また、IC タグには、タグ自身に電池を組み込み発信しているものと、リーダーから電波を受けたときに情報を発信するタイプがある。どちらも、無線通信を基本にしており、100%の検知率を保証できず、検知できなかった場合を加味してサービスを設計する必要がある。

- (3) モノや場所の ID 情報と関連デジタル情報のリンクを実現する方法として、バーコードを利用する方法もある。従来の 1 次元のバーコードに加えて、近年は 2 次元バーコードも利用されている。1 次元バーコードは ID 番号を表示するだけであるが、2 次元バーコードは記録できる情報量が多く、商品説明や URL アドレスなど文字情報も記録できる。バーコードリーダーを組み込んだ携帯電話のカメラで 2 次元バーコードを撮影すると、バーコードに記載された URL にアクセスすることができる。

2 次元バーコードには、まだ標準はなく、QR コードや Cyber コードなど多数提案され利用されている。

1 次元、2 次元ともに、関連情報のリンクを張るためには、あらかじめバーコードを印刷、トリガー側に貼り付けておかねばならない。また、バーコードは人間が読み取れるものはないので、どのような関連情報が提示されるかは、ユーザにはあらかじめは理解でき

ない。

- (4) トリガーが画像や映像である場合には透かしを利用したリンク生成も可能である。デジタル画像に、電子透かしとして、ID 情報やリンク先情報を組み込んでおくことで、その画像を印刷すればトリガーとして利用できる。リーダーで透かしとして記録された ID 情報を読み取って関連上にアクセスする。この場合も、組み込まれた情報は人間には読み取ることはできない。

- (5) もともと商品には商品タグがつけられており、部品や製品にも品番や製造番号が印刷されている。これらを文字認識技術や記号認識技術で読み取ることで、デジタル情報へのリンクを実現できる。従来、環境光の影響の排除やリーダーの小型化などの問題で実用的でなかったが、近年やっと実用的なレベルに達してきた。この技術を用いると、あらかじめ準備した ID コードだけでなく、実世界中のあらゆる文字情報をタグとして、デジタル情報へのリンクを生成できる。新聞記事中の単語、通販雑誌の URL アドレス、看板の電話番号、TV 番組に表示された URL アドレス、商品タグに印刷された商品名などなど、生活の中で触れる文字情報をトリガーとして利用できる。

図-4に示したように、機械可読で、人間可読でないトリガーは、あらかじめ実世界に準備しておく必要がある。特定のリーダーを利用しなければ読み出せないのも、標準化が普及のための鍵となる。一方、文字情報のような人間可読なトリガーを利用する場合には、人間に対する情報提供の一環としてリンクを提供できる。URL アドレスや電話番号を提示している場合に、リーダーを持ち合わせていなくても、メモとして控えておくことで、後日アクセスすることも可能である。リーダーを持っていれば、より簡便に自然に関連情報にアクセスできる。文章中の単語から辞書引きや住所から周辺地図の表示など、あらかじめ準備していないリンクもダイナミックに生成することができる。

● 携帯カメラによる文字読み取り技術 ●

最近のカメラ付き携帯電話の中には、接写機能を利用してカメラで撮影した印刷文字を認識できるものがある。これは実世界に書かれた文字情報をカメラで入力して手軽に利用できるユビキタス情報インタフェースの 1 つである。

このような携帯カメラによる文字読み取り技術¹⁾は、従来の光学スキャナによる文書読み取り技術から派生したものであるが、以下の点で異なる。

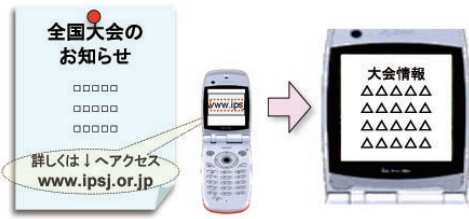


図-5 携帯カメラで URL を読み込んで、詳細情報にアクセス

1. 一貫性を持たせる
(異なる環境で一貫した操作性)
2. 頻繁に使うユーザーに近道を用意する
(ユーザーのスキルや嗜好への適応性)
3. 有益なフィードバックを用いる
4. ユーザーに主体的な制御権を与える
5. 段階的な達成感を与える
(レスポンスタイム、リアルな反応性)
6. エラー処理を簡単にする
7. 逆操作を許す(ミスの許容)
8. ユーザーの短期記憶負担を少なくする
(直感的操作)

図-6 ヒューマンインタフェースデザインの黄金律
(B. Shneiderman, 1992)

- (1) 入力デバイス (カメラ⇔スキャナ)
- (2) 利用目的 (手軽さ⇔省力化)
- (3) 動作環境 (携帯端末⇔PC)

カメラで文字画像の入力を行う場合、環境光の影響、ピントずれ、手ぶれ、斜めからの撮影といったスキャナの場合には生じない問題が発生するため、それらを解決する必要がある。携帯電話に付属するような小型カメラを利用する場合には、上記の他に、画素数不足、カメラ周辺部の歪みといった問題が加わる。

また、携帯カメラでトリガーとしての文字情報を入力する場合とスキャナで文書画像をデジタル化する場合とは、認識技術の利用目的が異なる。前者は必要なとき即座に情報にアクセス(図-5参照)できる手軽さが重要なものに対して、後者は手入力の手間を省くための自動処理による省力化が目的となる。そのため、携帯カメラによる文字読み取りでは、従来の文字認識応用では重視されていなかったヒューマンインタフェースとしての使いやすさが重要である。そのためには、たとえば図-6に示すヒューマンインタフェースデザインの規範を満たさなければならない。

また、動作環境の違いにも注意が必要である。昨今の携帯電話は一番前のPC並みの性能を持つようになったとはいえ、ROM容量、RAM容量、CPU速度のどれを取ってみても現在のPCに比べて制約が大きい。このような制約の中であっても、ユーザインタフェースとしては実時間の応答性を確保する必要がある。

これらのことを踏まえながら、携帯カメラによる文字



図-7 赤枠内に読みたい文字列を捉える

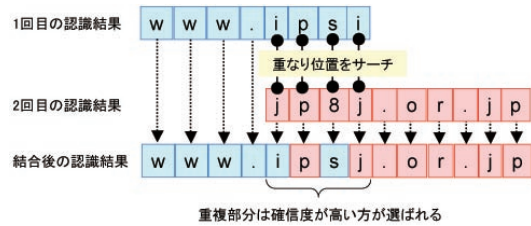


図-8 2つの認識結果の位置合せと結合

読み取り技術の概要について説明する。

STEP1. 撮影のインタフェース

従来のスキャナによる文書認識では、用紙全体を取り込んで全体(もしくはマウスで領域指定した範囲)を認識していた。それに対して、携帯カメラの場合は読み取りたい文字列を画面中心に捉えて撮影(図-7参照)することで、トリガーとして必要な文字情報を直感的に指定できるだけでなく、画面中心以外の領域を処理対象から外せるので処理速度の向上とRAM容量の節約ができる。

カメラでの撮影では接写することで認識に必要な解像度を確保できるが、長い文字列は1画面内に入りきらない場合がある。そのような場合のために、文字列を分割撮影して読み取り結果を繋げる機能が必要となる。分割撮影時に重なりができるようにすることで、図-8のように重なり位置を自動的にサーチして認識結果を自動結合してくれれば位置合せに要するユーザの負担は軽減される。

カメラ撮影で最も犯しやすいミスはピントずれである。これを防止するためには、オートフォーカス機能を搭載するか、ユーザにピント状態をフィードバックする必要がある。

STEP2. 画像補正と2値化

従来手法でも、ノイズを除去するための平滑化、文字と背景を黒と白で表現する2値化、文字列の傾き補正と

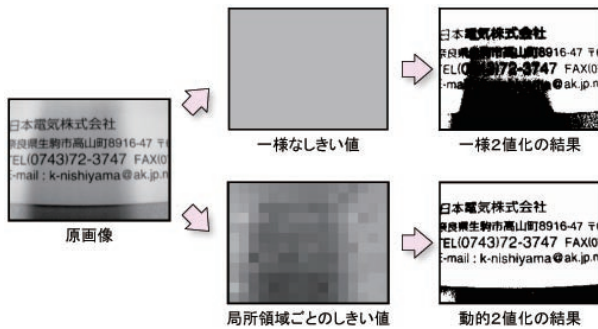


図-9 一様2値化と動的2値化

いった処理が行われていた。しかし、カメラ撮影画像の場合にはそれらだけでは不十分であり、環境光の影響を考慮した2値化が必要である。たとえば図-9のように、入力した画像に影が入ってしまった場合、一様なしきい値による一様2値化はうまくいかないため、画像を小さな局所領域に分割してそれぞれにしきい値を求めて2値化を行う動的2値化を行う必要がある。

STEP3. 文字切り出しと文字認識

文字切り出しと文字認識に関しては従来技術がそのまま利用できるが、紙面の歪みやセンサ・レンズの精度の問題などからカメラで撮影した文字は低品質であることが多いため、そのような状況に対応した手法を適用する必要がある。たとえば、文字認識辞書に傾いた文字を登録する、かすれや潰れに強い文字認識手法を用いるなどである。

STEP4. 読み取り対象に応じたテキスト処理

カメラで撮影する文字情報の属性（URL か電話番号かなど）が分かれば、それに応じたテキスト処理が可能となる。たとえばURLであれば、“http://”で始まっている、“.jp”や“.com”で終わることが多い、などの性質を利用して認識結果の誤りを自動的に修正できる。

STEP5. 誤認識修正のインタフェース

正解率100%の認識技術は存在しないので、認識技術をユーザインタフェースに応用する際には認識誤りをどのように修正するかが重要なポイントとなる。文字認識誤りの修正によく用いられるのは、誤り部分をユーザがキーで入力直す、誤った文字の認識候補から正解を選ぶ、といった手法である。これらの手法を利用すれば確実に誤りを直せるが、ユビキタス情報インタフェースとして直感的な操作とはいえない。

それに対して、図-10に示すように、認識誤りの周辺を再度撮影し直すだけで自動的に誤りを修正する手法が

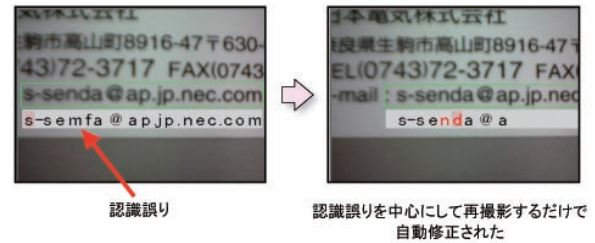


図-10 再撮影による認識誤りの自動修正

実現されている。その原理は図-8に示すように、複数の認識結果を確信度に基づいて結合することにある。この手法の効果を確かめる実験²⁾によると、50枚の名刺のメールアドレスの認識率は97.2%であったが、再撮影による自動修正だけで最高99.6%にまで認識率を向上できた。

この手法とSTEP1で述べた分割撮影による自動連結とは同じ原理で実現されているため、ユーザはカメラで文字情報を撮影するという1つの方法を覚えるだけで、分割撮影による自動連結と再撮影による自動修正がシームレスに行える。

● ユビキタス情報化社会における情報の活用 ●

ソファerを買いかえたくてデパートに出かける。途中の電車の中吊り広告に、家具の体験コーナーの広告を見つけた。所在地にリーダーをかざすと、そこへの道順が端末上に表示される。たどり着いた体験コーナーでは、沢山のソファerに腰掛けてみる。やはり座り心地は体験してみないと分からない。気に入ったものを見つけたので、その商品タグにリーダーをかざすと、カラーバリエーションと少々デザインの違いのものが提示された。他社製品情報や価格比較もできる。なんと同一素材でデザインの違うものが1,000種類もある。その中から気に入ったものを端末上で選択して、購入を完了した。

現在は、画像認識を用いたユビキタス情報インタフェースは、入力の安定性のため文字情報を中心に実現している。しかし、今後物体認識などの技術が進めば、タグ情報にとどまらず、商品自体を写せば、その関連情報を取得したり、建物を写せば観光案内を提示することができるようになると期待できる。

参考文献

- 1) Doermann, D., Liang, J. and Li, H.: Progress in Camera Based Document Image Analysis, Proceedings of 7th ICDAR '03, pp.606-617 (2003).
- 2) 西山京助, 仙田修司, 旭 俊之, 山田敬嗣: カメラ文字入力インタフェースの評価 (1), (2), FIT 2003, pp.491-494 (2003). (平成16年7月13日受付)