

赤外線画像を用いた指先実時間追跡による Enhanced Desk の実現

小林 貴訓^{*1} 小池 英樹^{*1} 佐藤 洋一^{*2}

Realtime tracking of finger tips in IR images for Augmented Desk Interface

Yoshinori Kobayashi^{*1}, Hideki Koike^{*1}, and Yoichi Sato^{*2}

Abstract - We have been developing a novel interface system to support users to perform various kinds of tasks in a regular office environment. This system is called the Enhanced Desk. The key component of the Enhanced Desk is its capability to monitor user's activities using computer vision techniques. This paper describes a rapid and accurate finger recognition and tracking technique used in the Enhanced Desk. By using an infrared camera and an image processing hardware, the system can successfully identify user's fingertips at video-frame rate even in a complicated background. In addition, the Enhanced Desk can recognize small objects on a desktop by using an additional pan-tilt camera that is tracking a user's fingertip.

Keywords: augmented reality, computer vision, finger recognition, infrared camera

1. はじめに

コンピュータは我々の生活に深く浸透し、その利用形態もさまざまである。以前では考えられなかったような非常に小型のコンピュータを携帯したり、周囲の環境に埋め込まれ一見してそれと解らないような場所での利用も行なわれている。将来的にはウェアラブルパソコンや PDA のような人にとりつく形態と、机や壁、部屋などユーザの周囲の環境に融合した形態の 2 つの流れにそって発展してゆくと考えられる。

このような流れの中で、インターフェースについて考えると、現在の GUI 等に代表されるインターフェースはユーザをインターフェースに縛り付け、本来行ないたい作業とは別にインターフェースにも気を配らなければならなかった。そこで次世代のインターフェースではユーザの行ないたい作業に対してより直接的でより自然な操作を行なうことのできるインターフェースが求められている。

そこでコンピュータを用いた情報の検索や利用において、ユーザのインターフェースによる負担の軽減のため、実世界でのユーザの作業や操作をコンピュータが支援し、キーボードやマウスのみによる従来型のインターフェースよりも快適に利用できる次世代のインターフェースとして、実世界指向、状況認識によるインターフェースに期待が寄せられている[1]。

本研究ではコンピュータの利用に既存の操作を意識させないインターフェースとしてユーザの快適性を重視し、

計測器などをユーザの身体に取り付けることでユーザの自由度を低下させることのないインターフェースの開発を目指している。特にユーザの周囲の環境に融合したインターフェースとして“机”に統合された情報インターフェースシステムに注目している。オフィスなどでは机上で PC を使った作業を行ない、また同時に紙書類を使った作業を行なうことが多い。これらの机上での作業をコンピュータに机上を観察させることによって、統合的に扱うことのできる情報インターフェースシステムのための提案を行なう。

2. 実世界指向ユーザインターフェース

実世界でのユーザの作業や操作をコンピュータが支援し、キーボードやマウスよりも快適に利用できる次世代のインターフェースとして、実世界指向、状況認識によるインターフェースが求められている。

実世界指向ユーザインターフェースは実世界のユーザの状況を常にコンピュータが認識し、ユーザの意図を汲み取ってコンピュータが作業を支援する。これにより既存のコンピュータの操作を感じさせない透明なインターフェースの実現を目指している。この実世界指向インターフェースを机上での作業に応用し、机型実世界指向インターフェースの一例として“机”に統合された実世界指向インターフェース Enhanced Desk の開発を進めている。

Enhanced Desk はコンピュータの画面を机上に投影し、同時に机上をカメラで観察する事によってユーザとコンピュータとのインタラクションを実現するものである。本システムでは特に机上のユーザの指先と物体の位置を検出し、電子情報及び机上の書類等とのインタラクションを考慮した。

机上での作業とコンピュータでの作業の融合に注目した例として Digital Desk[2] がよく知られている。

*1: 電気通信大学 情報システム学研究所

*2: 東京大学 生産技術研究所

*1: Graduate School of Information Systems, University of Electro-Communications

*2: Institute of Industrial Science, University of Tokyo

Digital Desk では机上に投影されたコンピュータ画面を指先などで操作するものである。ユーザは机上に投影されたイラストを切り取ってコピーしたり、机上に投影された電卓で計算をすることができるものであった。

本研究の前段階である小林らの Enhanced Desk[3]では、机上での紙書類に着目し、紙書類と電子情報の統合利用を試みている。Enhanced Desk では紙書類と電子情報の対応づけるためにあらかじめ紙書類に付与したバーコードを用いているが、バーコードがある程度以上の大きさで観察されることが必要なため、大型のバーコードを用いなければならなかった。またユーザの手領域を認識するために肌色抽出を行っていたが、机上に肌色に近い物体があると誤認識されたり、ユーザの手の上に映像が投影されると、手領域抽出がうまくいかないといった問題点があった。ユーザの指先位置の認識についても常に特定の手の向きを仮定し、指先は1点とするなどの制約が多く、また認識精度の不十分さが大きな問題であった。そして一連の処理をソフトウェアで行なっているため実時間処理が不可能であった。

本研究ではそれらの問題点を解決するため、以下の手法の提案を行なう。

- ・赤外線カメラを利用してユーザの皮膚領域からの放射光を測定することによりユーザの手領域を安定に抽出する。
- ・正規化相関に基づいたテンプレートマッチングの利用によりユーザの指先の高速追跡を実現する。
- ・机上の小さなバーコードの認識を行なうため、ズーム機能付パンチルトカメラを用いてユーザの指先周辺を拡大追跡する。
- ・処理の高速化のため、画像処理ハードウェアの利用と分散処理の適用を行なった。

以下本論文では3章で Enhanced Desk の概略について述べ、4章にて赤外線画像からの指先位置抽出について述べる。5章ではカメラ画像座標と机上平面座標の補正について、6章では指先周辺の注視によるバーコード認識について説明する。次に7章にて音声認識について、8章にて分散処理について、9章にてアプリケーション例についてそれぞれ述べた後、最後にまとめと今後の課題を述べる。

3. Enhanced Desk の概略

図1に示すように通常の机の上方にプロジェクタを取り付け、コンピュータの画面を鏡で反射させ、机上に投影する。机上を観察するカメラは2台設置した。1台は赤外線カメラであり、机左側から上を向けて設置し、表面反射鏡ごとに机上の撮影を行なう。もう1台のズーム機能付パンチルトカメラは机の右上方に設置し、机上の一部分を拡大して撮影することができる。またスピーカとマイクも設置されており、音声等の入出力の機能も持

っている。

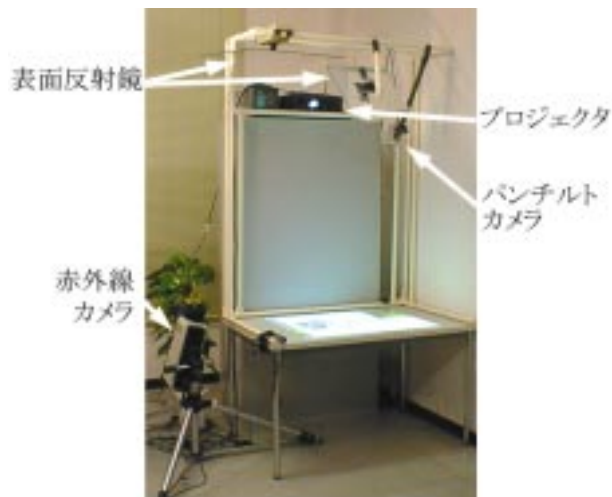


図1 Enhanced Desk 外観
Fig.1 Enhanced Desk overview.

Enhanced Desk での主な処理の流れを以下に示す。

ユーザの手領域を赤外線カメラ画像より抽出する。

抽出された手領域からユーザの指先点を検出する。

カメラ画像上の指先点の位置座標からあらかじめキャリブレーションにより求めておいた射影変換パラメータを用いて机上の位置座標に変換を行なう。

机上の指先位置情報をもとに机上に投影したオブジェクトや情報とインタラクションをとる。

図2にシステム構成図を示す。

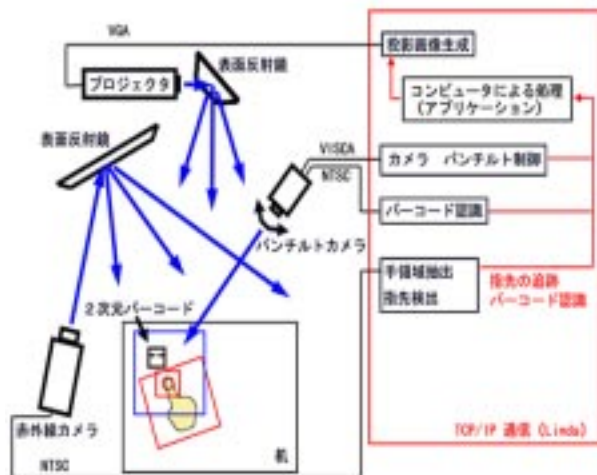


図2 システム構成図

Fig.2 System configuration of Enhanced Desk.

4. 赤外線画像からの指先位置の抽出

4.1 手領域の切り出し

手領域の抽出には背景差分やカラーカメラによる肌色抽出が一般的に用いられている。しかし、これらの手法では背景の状態によっては手領域を抽出することが困難な場合が多い。特に本システムで想定される作業環境では、机の上に紙書類や書籍などを開くことが考えられ、プロジェクタにより電子情報を投影するために、ユーザの手領域の色が一定ではないなどの問題点がある。そこで本システムでは手領域を安定に抽出するために赤外線カメラを利用した。

机上を観察している赤外線カメラの画像から、人の体温をもとに手領域を切り出す。赤外線カメラは Nikon サーマルビジョン LAIRD3A を使用した。本カメラは有効画素数 768 × 465 画素、フィールドタイム 1/60 秒で -20 ~ 50 までの範囲を撮影することができる。このカメラを利用して人の体温の近傍(30 ~ 34)を撮影することにより、机上の手の様子を NTSC の 256 階調の映像として得ることができる。この映像を一定の閾値で 2 値化処理し手領域を抽出する。画像入力装置で取り込む際の解像度は 256 × 220 画素である。この手法により、背景や照明の変化に影響されず手領域だけを安定に抽出することができた(図 3)。

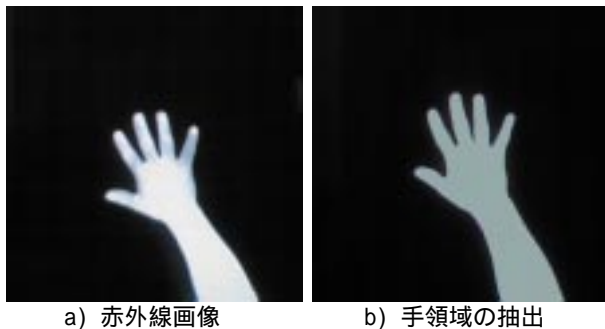


図 3 指先認識

Fig.3 Recognition of fingertips.

4.2 指先位置の認識

指先位置の認識は一般に指先形状の輪郭が円に近いことに基づき、円形テンプレート(図 4)を用いて手領域周辺で正規化相関に基づくパターンマッチング[4]を行な

うことで検出する。手領域を切り出し 2 値化した 256 × 220 画素の画像に対して 15 × 15 画素の大きさのテンプレートをを用いている。

指先位置の検出はユーザの手は机上にあり、カメラとユーザの手の距離はほぼ一定で見かけ上の大きさが極端に変わらないことから、一定の大きさのテンプレートで指先位置を検出することができる。同様の理由から処理対象を手領域の一定範囲のみに限定し処理の高速化を図った。

パターンマッチング処理の後、あまりにも近接した場所で多くの相関値の高い点が見つかった場合、そのうちの一番相関値の高い点を残し、残りを指先候補点から除外する。これによりいくつかの指先候補点が得られるので、それらのうち指先点以外の誤認識された点を除外する。そのためテンプレートの周囲 8 点(矩形の 4 頂点及び各辺の二等分点)に対応するピクセルを調べ、指先点かどうかを判定する(図 4)。この処理により指先がテンプレートを突き抜けていたり、テンプレートの周囲に指らしきピクセルが何もない場合なども指先点候補から除外する。

最終的に指先候補点を相関値の高い順に 5 つまで採用しユーザの指先点とする。

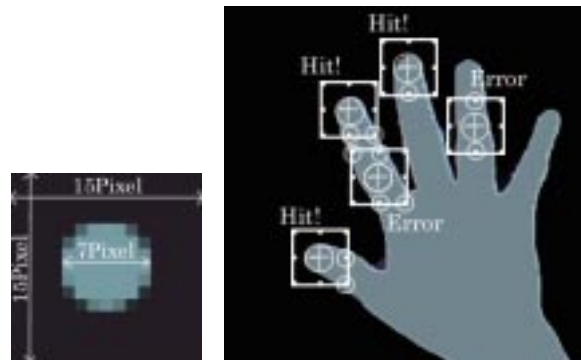


図 4 テンプレート及び誤検出判定の例

Fig.4 Image template and detection of false matching.

本システムにおける指先検出処理は日立画像処理ボード IP5010 を用いて行なっている。IP5010 は白黒濃淡画像で 40 画面分のメモリをもった画像処理ボードであり、メモリ上に保存した画像間で 2 値化や差分といった画像処理をハードウェアで高速に行なうことのできるボードである。OS は WindowsNT および Linux に対応しており、多くの画像処理関数がライブラリとして用意されている。本研究ではネットワークとの親和性を鑑み、OS に Linux を利用した。

5. カメラ画像座標と机上平面座標の補正

赤外線画像中で検出された位置が机上でどの位置に対応するかを知るため、赤外線カメラ画像座標と机上平面座標の補正を行なった。ここでは赤外線カメラ画像座標

と机上平面座標の対応点をいくつか定義し、その関係を元に赤外線カメラ画像座標 (x, y) と机上平面座標 (x', y') の間の対応を式(1)のように射影変換を用いて表現する。

$$\begin{cases} x' = c_1x + c_2y + c_3xy + c_4 \\ y' = c_5x + c_6y + c_7xy + c_8 \end{cases} \quad (1)$$

対応点を求めるために、まず机上平面の座標系を設定した。手領域を切り出す赤外線カメラでは温度差が画像として出力されるため、机上の目標点の位置に小型電球を埋め込んだキャリブレーション板(図 6)を作成し、机上に置くことで机上座標系を設定した。このキャリブレーション板を赤外線カメラで撮影し対応点を計測した。



図 6 赤外線カメラ対応点の測定
Fig.6 Calibration pattern used for infrared camera.

対応点の組を計測後、その関係をもとに射影変換パラメータ $(c_1 \sim c_8)$ を決定する。このパラメータを用いて変換を行なうことによって画像のゆがみや、位置のずれなどを補正することができる。

この射影変換パラメータを得るためには最低 4 組の対応点があればよいが、本システムでは安定に解を求めるために対応点を 8 組とり、上記の連立方程式を特異値分解法を用いて解くことで変換のパラメータを求めた。

本システムでは、処理軽減のため机上の座標系とプロジェクタで机上に投影した画像の座標系を同一とし、机上平面座標の値とプロジェクタ投影画像の画素の位置座標を等しくした。これにより机上平面上の指先位置にプロジェクタの画像を投影するには、単に机上平面座標上での座標値にオブジェクトを描画した画像を投影すればよく、処理の軽減を図ることができた。

6. 指先周辺の注視によるバーコード認識

本システムでは画像処理を行なって認識したユーザの指先が 1 つだった場合、ユーザが指差し動作をしており、指先の周辺が注目している領域であるとみなし、パンチルトカメラで拡大追跡している。このカメラの映像を用いてバーコード等の小さい物体の認識を行なっている。

6.1 カメラ制御

追跡カメラは Sony EVI-G20 パンチルトカメラを使用し、PC のシリアルポートから VISCA コマンドをカメラに送信することによってカメラのパンチルト制御を行なっている。

カメラは垂直方向 -7794 ~ 7794、水平方向 -15570 ~

15570 の大きさのパンチルト駆動平面をもっており、座標を指定することでその方向にカメラを向けることができる。この機能を利用してユーザの指先を追跡させている。本システムでは机上の指先位置がパンチルト駆動平面上でどの座標位置にあるかわかればよい。

先にカメラ画像座標と机上座標の補正を行なったが、指先追跡においても同様の手法を用いてカメラによる指先追跡を行なっている。今までの処理で指先の机上座標はわかっているので、その机上位置に対応したパンチルトカメラの駆動平面上の位置を算出する。そのためにまず対応点を 4 組以上計測しなくてはならない。パンチルトカメラが現在、駆動平面上でどの座標値をとっているかは VISCA コマンドで知ることができる。これを利用して、机上の対応点をパンチルトカメラが中心に捕らえるように位置を調節しその時点でのパンチルト駆動平面上の座標値を対応点として測定する。この対応点をもとに先の射影変換を用いて机上座標からパンチルト駆動平面上の座標に変換するパラメータをもとめた。

6.2 指先追跡処理の流れ

指先追跡の処理の流れを以下に示す。

机上平面とパンチルト駆動平面の 2 平面間で対応点を計測し、射影変換パラメータを算出しておく。

画像処理によって赤外線カメラ画像上の指先点の位置を計測する。

赤外線カメラ画像上の指先点の位置を射影変換により机上平面上の座標に変換する。

机上の指先点の位置をカメラパンチルト駆動平面に射影変換を用いて変換する。

パンチルト駆動平面上の座標をカメラに VISCA コードで送出し、カメラにユーザの指先点を向かせる。

指先点を向いたカメラの画像を処理する。

図 7 に指先の拡大追跡を行なった様子を示す。



図 7 指先の拡大追跡

Fig.7 Tracking of a small region around a finger.

6.3 バーコードの認識

現実物体（紙書類、書籍等）にはバーコードを取り付けることにより認識を行ない、電子情報とのリンクを作成している。バーコードには二次元マトリックスコードを利用した[5]。この二次元バーコードはコードの種類だけでなく、コードのある位置や、向きなどもコードから得ることができるので、物体のコードが貼付されている位置を記憶していれば、コードから物体の姿勢を計算することができる(図8)。

バーコードはパンチルトカメラで撮影した画像の中に存在すれば認識される。認識するためにはカメラ画像中である程度の大きさが必要であるが、パンチルトカメラが拡大した映像を撮影しているため、机上においた書類に貼り付けたバーコードでも約 1.5×1.5cm の大きさのものまで認識することが可能である。机上ではユーザの指差し動作によって、バーコードが指差されれば、そのコードを認識し、それに対応したインタラクションを起こすことができる。

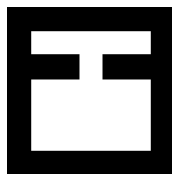


図8 2次元バーコードの例
Fig.8 Example of 2D matrix code.

バーコードの認識処理はビデオ映像を SGI O2 のビデオ入力端子から 320×240 画素の大きさで取り込み、2 値化等の処理を行ない、バーコード認識ライブラリを用いて認識を行なっている。画像入力は SGI のビデオライブラリを使用している。またバーコード認識処理はソフトウェアで行なっており、10～15f/s 程度の速度で実行が可能である。

7. 音声認識

本システムにはユーザとのインタラクションの補助的役割として音声認識を使用している。ここではユーザの指差し動作やオブジェクトの移動や回転といった操作モードの変更を行なうためのキーワードを認識させるためにのみ利用している。

音声認識は IBM ViaVoice の音声認識エンジンを使用している[6]。本システムでは、あらかじめ登録してある単語をユーザが発声したときに何らかのインタラクションを起こすために、音声認識プロセスはメッセージサーバに対して、認識した単語の種類をメッセージとして常に送っている。サーバに送られたメッセージは情報提示プロセスが受け取り、その単語に対応したインタラクションを机上で起こすことが可能である。

8. 分散処理

本システムでは大きく分けて、指先検出およびパンチルトカメラによる指先追跡処理、2次元バーコード認識処理、音声認識処理、情報提示処理の4つの処理を行なっている。本システムではそれぞれの処理を高速化のため複数のマシンで分散処理を行なっている。図9に分散処理の概略図を示す。

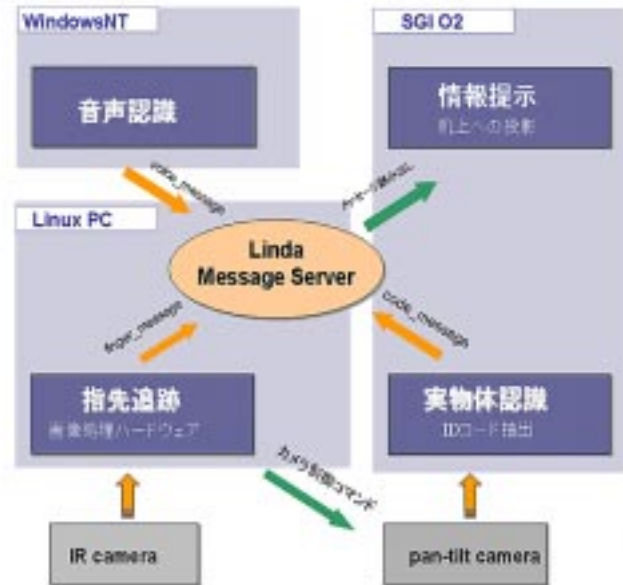


図9 Enhanced Desk における分散処理
Fig.9 Distributed processing for Enhanced Desk.

指先検出及びパンチルトカメラによる指先追跡処理は、Pentium 450Mhz の PC に Linux と画像処理ボード IP5010 を導入したマシンを使用している。このプロセスにより机上のユーザの指先数および位置を計測し、指先の机上座標位置をメッセージサーバに送出している。また、パンチルトカメラの視線方向をユーザの机上での指先位置になるようにカメラを制御している。

2次元バーコード認識処理では、パンチルトカメラの画像を SGI O2 のビデオ入力に取り込み、ソフトウェア処理によって2次元バーコードの認識を行なっている。認識結果は常にメッセージサーバに送出される。

音声認識は Pentium 330Mhz の PC に WindowsNT と ViaVoice を導入して処理を行なっている。ユーザがマイクに向かって話した音声をすべて認識させ、そのうちあらかじめ登録してある単語を認識した場合に特定のメッセージをメッセージサーバに送出している。

最後に常にメッセージサーバに貯えられている情報を取り出し、それに対応するインタラクションや画像を作成する情報表示プロセスがある。これはいわばアプリケーションプロセスであり、SGI O2 で処理を行なっている。

各々の処理間の通信にはタプル空間通信システム TS System/s、通称 Linda を用いている[7]。このシステムは

ネットワーク上でダブルスペースと呼ばれる空間を共有し、その空間を介して任意の文字列をセットとしたダブルと呼ばれるメッセージをやり取りすることによって通信を実現している。非同期通信であるため、個々のプロセスが完全に独立して動作することができ、また新しいプロセスの追加や変更が容易であるという利点がある。

9. アプリケーション例

9.1 簡易文字認識

机上でのユーザの指差し動作を追跡し、机上に投影されたマス目に対して数字を書き入れるとその数字を認識し、番号が机上に表示される。ユーザの指先に単純なポインティング以外の意味を持たせるために利用できると思われる。



図 10 簡易文字認識

Fig.10 OCR application using Enhanced Desk.

9.2 Desk Pad

机上で通常の紙に対して書き込んだ図を電子情報として保存するためのアプリケーションである。これは机上平面座標と各カメラの画像上座標との補正と位置検出精度の向上により可能となった。印刷するとほぼ同じイメージを電子的に複製できる。



図 11 Desk Pad

Fig.11 Desk Pad.

10. まとめと今後の課題

机上での電子情報と現実物体を統合的に扱うことのできる机「Enhanced Desk」実現のために必要とされるイン

ターフェースに関する技術の提案を行なった。

本システムでは指先を毎秒 20 フレーム以上の実用的な速度で安定して検出することができた。またカメラで指先を拡大追跡しているため、小型のバーコードの認識が可能となり机上の物体に不自然に大きなバーコードをつける必要がなくなった。ユーザは机上に投影された 3D オブジェクトを指先で回転させながら移動したり、書籍に添付したバーコードを指差すことで関連するホームページを表示させるなど、従来のマウスやキーボードよりも直感的で現実と密接にリンクしたインタラクションを電子情報との間で行なうことができた。

今後の方針としてユーザの指先の位置情報だけでなく指先の軌跡を利用したジェスチャ認識等の導入を行ないたいと考えている。カメラによる拡大撮影が可能になったことから、紙面の小さい文字も認識できると思われるため、文字認識を利用し、紙書類と電子情報の統合的な利用のための機能を付加してゆきたいと考えている。また本システムを遠隔協調作業への応用し、遠隔地の 2 つの机上を共有することによる机上平面を媒介としたコミュニケーション環境の構築についても検討を進めている。

11. 参考文献

- [1] 小池：Bit 別冊ビジュアルインターフェース - ポスト GUI を目指して - ; 共立出版, 2.1 章, pp.24-44 (1996).
- [2] P.Wellner: Interacting with paper on the DIGITAL DESK; Communication Of The ACM, vol36, No7, July, pp.87-96 (1993).
- [3] M.Kobayashi and H.Koike: Enhanced Desk, Integrating Paper Documents and Digital Documents; Proceedings of 1998 Asia Pacific Computer Human Interaction, pp.167-174 (1998).
- [4] D.H.Ballard and C.M.Brown: Computer Vision; Prentice-Hall, pp65-70 (1982).
- [5] 歴本: 2次元マトリックスコードを利用した拡張現実感の構成手法; インタラクティブシステムとソフトウェア, 日本ソフトウェア科学会 WISS, 近代科学社, pp199-208 (1996).
- [6] IBM: <http://www.jp.ibm.com/voiceland/>.
- [7] 寺田: <http://www.sanpo.t.u-tokyo.ac.jp/~terada/ts.html>.
- [8] H.Ishii and B.Ullmer: Tangible Bits, Towards Seamless Interface between People, Bits and Atoms; Proceedings of the ACM Conference on Human Factors in Computing System (HCI'97), pp.234-241 (1997).