

## テーブル型マルチメディアシステムに おける物体の影を用いたジェスチャ認識

岡未来<sup>†</sup> 杉原伸治<sup>†</sup> 山口竜二<sup>††</sup> 平川正人<sup>††</sup>

筆者らは、協調作業支援を目的としたテーブル型マルチメディアシステムの開発を進めてきた。同システムの操作には赤外線 LED を付加した任天堂 Wii リモコンを用いていた。しかしながら、テーブルの大きさを持つという特性から、手・腕や身体を用いた、いわゆる面によるインタラクションが望まれる。本研究では、特別なデバイスを用いることなく、手や腕を用いてシステムとインタラクションを行う手法を提案する。具体的には、テーブル上の腕と影の形状からテーブル面への手の接触判定を含むジェスチャ認識を行う。

### Shadow-based gesture recognition in a tabular multimedia system

M. Oka<sup>†</sup> S. Sugihara<sup>†</sup> R. Yamaguchi<sup>††</sup> and M. Hirakawa<sup>††</sup>

We have been developing a tabular multimedia system for supporting collaborative work among users, where each user manipulates the system by using a Nintendo Wii Remote customized by installation of an infrared LED. However, considering that the system has a table-size interaction surface, it is desired for the users to be able to interact with the system by their hands, arms and body having a certain size. In this paper, we present a new interaction mechanism with hands and arms, and a shadow-based touch detection algorithm.

### 1. はじめに

テーブルを囲んだ作業は、ジェスチャやアイコンタクトなどの非言語情報を用いた意思疎通が可能なため、複数人での協調作業を円滑に進めるうえで有効な手段である。また、テーブルを囲むことによってテーブル上の情報を共有できることも協調作業の円滑化に繋がる。近年、このようなテーブルの特性を活かした、テーブル型システムの研究が盛んに行われている[1]。

テーブル型システムは、複数ユーザが同時にシステムへの入力を行うので、複数人で共有しにくいキーボードやマウスを用いた入力方法は適さない。複数ユーザの使用を考慮し、直接表示面に触れたり、テーブル上に置かれた実オブジェクトを動かすこと等でシステムへ入力を行う方法などが研究されている。また、ユーザはテーブル上の表示面を囲んで向き合うため、情報を提示する方向によっては、文字などの情報の視認性が低下するという問題が生じてしまう。この問題は、テーブル型システムを設計する際に考慮すべきデザイン問題として指摘されている。このように、テーブル型システムの研究は、システムへの入力方法やユーザへの情報提示方法など、様々な観点から研究が進められてきている。

筆者らは、テーブルを囲んだ全ての利用者が、位置を含めた音の情報を共有可能なテーブル型マルチメディアシステムの研究・開発を行っている[2]。音の位置という、ユーザ間で共有できる情報を増やすことで、より円滑な協調作業が可能となる。筆者らは、これまでにテーブル上で複数の音源を組み合わせることで音楽演奏する音楽マッシュアップおよびグループ回想法を支援するアプリケーションの開発を行い[3]、展示会でデモンストレーションを行った。そこではシステムの操作には赤外線 LED を付加した任天堂 Wii リモコンを使用している。来場者の振る舞いを観察したところ、リモコンを持った手はもちろん、上半身を乗り出してテーブル上の仮想オブジェクトに触れ合おうとする事例が多数見受けられた。自身の身体による操作はもちろんながら、そこではテーブル上の点（位置）だけではなく、面（領域）としての対話情報を捕捉することの必要性が併せて認識させられた。

本研究では、特別なデバイスを用いることなく、手や腕を用いてシステムとインタラクションを行うための手法を提案する。提案手法にあつては、ユーザがテーブル上に手を差し入れることで生じる影に着目する。テーブル上部から赤外線光をテーブル面に照射し、テーブル上に手の影を生成し、手と影の面積や傾きを用いて対話処理を実現する。具体的には本研究では、テーブル面への接触判定を行う。影は、光を遮る

<sup>†</sup> 島根大学大学院 総合理工学研究科 数理・情報システム学専攻  
Interdisciplinary Graduate School of Science and Engineering, Shimane University

<sup>††</sup> 島根大学 総合理工学部 数理・情報システム学科  
Interdisciplinary Faculty of Science and Engineering, Shimane University

ことで生成されるものであり、特別なデバイスの使用を強制しない。また、赤外線光を利用することで可視光線からの影響を軽減することができる。

## 2. テーブル型マルチメディアシステム

まず、本研究のベースをなす、筆者らが開発してきたテーブル型マルチメディアシステムについて紹介する。

テーブル型マルチメディアシステムのシステム外観を図 1 に示す。システムは、16基のスピーカを4×4レイアウト(図 2)に配置したテーブル(SoundTable)を中心に、テーブル上でのジェスチャ獲得用の赤外線フィルタを付加したWEBカメラ、ユーザへの情報提示用のプロジェクタ、及び処理用PCによって構成されている。また、操作デバイスとして赤外線LEDを付加した任天堂Wiiリモコンを使用している(図 3)。

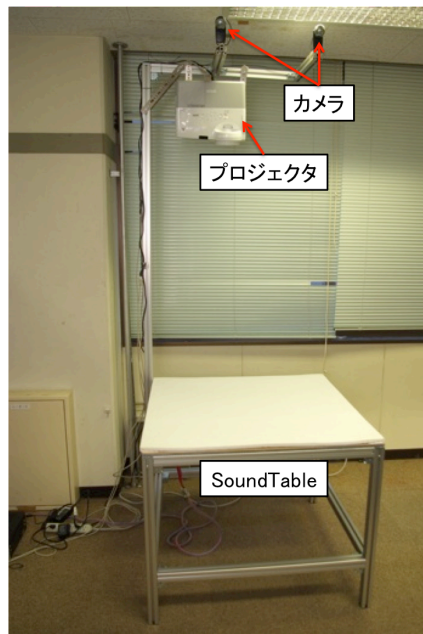


図 1 テーブル型マルチメディアシステム

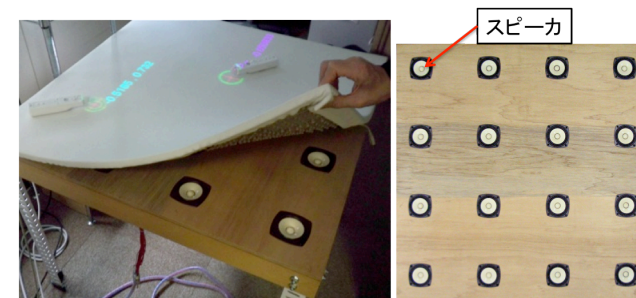


図 2 スピーカ配置



図 3 操作デバイス

SoundTable では、テーブル内に内蔵した各スピーカの音量を個別に制御し、正弦波と余弦波によるクロスフェードを応用することにより、スピーカの位置に依らず、テーブル面の任意の位置に音を定位させることができる。プロジェクタからコンピュータ生成映像を投影するために、テーブル上面には白色の保護カバーが被せてある。コンピュータ生成映像をテーブル面に投影することで、視覚的フィードバックを可能にしている。

テーブル上の3次元空間内で操作デバイスを用いて表現されるジェスチャの認識にあたっては、2基のカメラ映像から三角測量法を用いて行う。利用者ごとに操られる操作デバイスの識別、特にカメラ視点からみて操作デバイスのLEDに重なりが発生した場合にも、それぞれのデバイスを正しく識別・追跡できるようにするため、Wiiリモコンに組み込まれている加速度センサの値を3次元空間位置の遷移情報と比較するようにしている。

### 3. 関連研究

我々は一般のデスクトップ PC を利用する際、マウスやキーボードなどの入力デバイスを使用して操作や入力を行う。個人使用を前提とした机上環境ではマウスやキーボードは妥当なデバイスであるが、これに対してテーブル型システムは、複数人が1つのディスプレイを共有し、同時に操作を行う。この時にマウスなどの入力デバイスを共有して利用することは適切ではない。また、たとえ人数分のマウスやキーボードを用意できてもユーザは手元で入力を行うことになり、テーブルを囲んだ協調作業の利点が損なわれる。このような理由からテーブル型システムでは、テーブル面のディスプレイに手を使用して操作する手法が多く用いられる。ユーザが手、指などを用いてシステムとのインタラクションを行うマルチタッチ技術は広く研究されている[4]。

テーブル型システムの多くは、そこでのジェスチャの取得にあたってはカメラが用いられる。光の全反射という物理現象を巧みに用いた FTIR を用いたシステム[5]の他、マイクロソフト社の Surface[6]や MERL の DiamondTouch[7]などもある。

一方、テーブル面に埋め込んだセンサを用い、ユーザのシステム操作を許す研究として SmartSkin [8]がある。SmartSkin は、近づけられた身体部位の形状・位置の認識を実現するセンサである。平らな面に格子状に設置された電極と人体との間の静電容量を計測することで、センサ上にある手の位置情報を取得することができる。また、手がセンサの近傍にあれば、センサに触れていなくても検知できる。SmartSkin は、カメラを使用せず3次元や面での操作ができる利点がある。しかしながら、テーブル型システムへの適用を考えた場合、テーブル面にセンサを埋め込む必要があるため、テーブルに設置されているスピーカの磁場による影響が危惧される。

既存の多くのテーブル型システムと同じく、カメラから取得された画像データに画像処理を行うことでジェスチャ認識を行う試みとして PlayAnywhere [9]がある。本研究との係りの上で特徴的な点として、提案システムと同様に、影を利用して指先の接触判定を行っている。システム上部に赤外線光照射用の赤外線アレイと IR フィルタを付加したカメラが設置されており、カメラで赤外線光によってできた影の面積変化に注目することで指先の接触判定を行っている。その際、システムを正面から使用することを前提としているので影の頂点を接触点としている。しかし、正面から使用することを前提とすることは、筆者らが想定するテーブル型システム環境にあっては使用上の制約となる。複数人がテーブルを囲むという利用状況の下で協調的に作業するという特徴が損なわれてしまう。

我々は、テーブル型の形態を採っていることの利点を損なわないように、4つのいずれの側面からも手（腕）の差し込みを許すようにしており、素手による3次元空間内での操作を実現する。

### 4. システム概要

#### 4.1 ハードウェア構成

提案システムでは、テーブル上部から赤外線光をテーブル表面全体に照射する。そして、赤外線光を遮ることでテーブル面に生じる影をテーブル上の IR フィルタを付加したカメラで取得し、リアルタイムに画像処理を行うことで利用者のジェスチャ動作を取得する。

提案システムの構成を図4に示す。提案システムは SoundTable を中心に、赤外線光を照射する赤外線アレイ、IR フィルタを付加した WEB カメラ 1 台、ユーザへの情報提示用のプロジェクタから構成される。

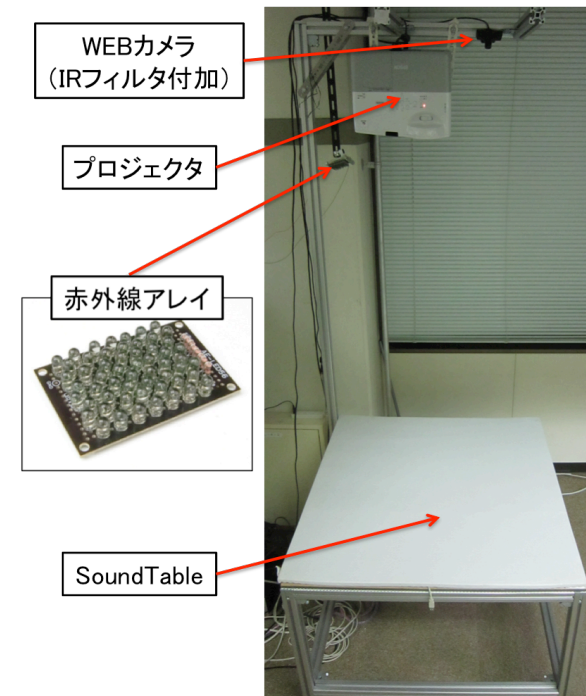


図4 提案システム

## 4.2 赤外線光による影の生成

提案システムはプロジェクタから映像をテーブル面へ投影するため、通常のカメラ映像を用いる場合には、影の発生パターンがプロジェクタからの投影映像の影響を受ける。本研究では、赤外線光と IR フィルタを付加した WEB カメラを使用することで、プロジェクタからの生成映像やその他の可視光の影響の軽減を図っている。

また、影は光の照射方向や強さで形状・大きさや濃さが大きく変化する。単一の LED では十分な影が得られないため、また腕の差し込み方向の影響を抑えるためにも、LED をアレイ状に並べることで対処するようにしている。WEB カメラから取得された画像の例を図 5 に示す。



図 5 IR フィルタを付加した WEB カメラ取得映像

## 5. 腕と影の領域の認識

本研究では、カメラから取得された画像中から腕領域と影領域の切り出しを行い、それらの領域を特徴づける軸の角度を比較することで、テーブル面への手の接触判定を行う手法を提案する。本章では、そのための基本機能として、まず腕領域と影領域の取得手法を説明する。接触判定手法については次章で説明する。

腕領域と影領域の認識処理手順は以下の通りである。

1. 背景差分法を用いて変化領域を抽出
2. 背景差分結果から腕領域と影領域の判別
3. 腕領域と影領域の対応付け

以下、これらの手順をそれぞれ詳しく説明する。

## 5.1 背景差分を用いた腕および影領域の抽出

背景差分法はカメラが固定され視界の移動がない場合に、画像中から注目すべき物体を抽出するための代表的な手法である。提案システムは、カメラが固定されており、背景の変化がほとんどないため、背景差分法を用いることで、腕および影の領域の抽出を行っている。また、赤外線光を使用しているため、背景差分法の問題点である照明環境変化の心配が少ない。しかしながら背景差分法では、取得すべき物体の明度が背景画像と類似している場合、同物体を背景として誤認識してしまう問題点がある。

実際、図 6 (a) はカメラから取得された腕ならびに影の映像であるが、腕と背景であるテーブルの明るさが類似しているため、腕の領域を背景と認識してしまう場合がある。この問題に対処するべく、入力画像に対して輪郭線抽出を行い、背景差分結果と輪郭線抽出の論理和をとる輪郭線背景差分[10]を用いることで、腕領域抽出の欠落を抑えるようにしている。

図 6 (a) の画像に対して輪郭線背景差分処理を施して得られた領域抽出結果を図 6 (b) に示す。但し、この段階では、腕と影の領域の区別は達成されていない。



(a) 入力画像

(b) 変化領域抽出結果

図 6 輪郭線背景差分結果

## 5.2 腕領域と影領域の判定

図 6 (a) を見ると分かるように、影領域は濃いグレーであることがわかる。これとは逆に、腕領域は白に近い明るいグレーの状態を取る。そこで、グレースケールに着目した閾値処理を施すことで腕領域と影領域の判別を行う。領域の明度が高い箇所は腕領域、低い箇所は影領域とすることで、腕領域と影領域の判定を行う。なお、腕ならびに影の領域を判別する際の閾値はそれぞれ別々に設定している。

閾値処理を行って得られた腕領域ならびに影領域の取得結果の例を図 7 に示す。ここで、細かいノイズに相当する部分はラベリング処理を行う際に除去可能であるが、影領域の周辺部は次第にグレースケールの値が低くなるため、単純に閾値処理だけでは影領域の周辺部は腕領域と誤判別されてしまう。そのため腕領域画像に、領域の収縮と膨張を行うことで当該ノイズ領域の除去を行う。最終的に得られた腕領域は図 7 の下部右図のようになる。

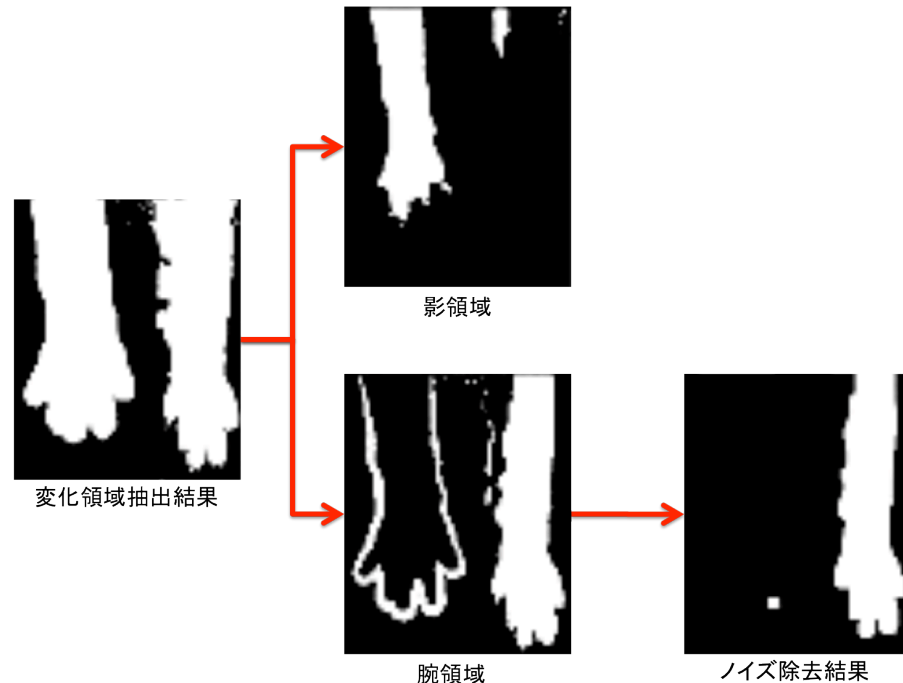


図 7 領域の明るさに基づいた腕領域および影領域の判別

## 5.3 腕領域と影領域の対応付け

複数の腕がテーブル上に差し入れた際、それぞれの腕に対して影を対応付けなければならない。提案システムにあっては、赤外線 LED アレイの位置が固定されているため、腕に対して影が現れる方向は一定である。つまり、腕がテーブルの四方どの方向から差し入れられているかの情報が取得出来れば、影が生成される方向が推測可能である。

テーブルの縁部分を切り取る物体領域が見つかったという処理結果を受けて、腕が差し込まれた方向の特定を行なっている。より具体的には、テーブルのどの方向から腕が差し入れられているか判定し、影ができる方向を推定する。その後、その方向にある影領域（グレースケールの値が高い領域）で、距離が最も近いものを、その腕領域の影として対応付ける。

## 6. 影を用いたジェスチャ認識：テーブル面への手の接触判定

システムとして認識すべきユーザジェスチャには様々なものがあり得るが、その中でも基本的なもの1つとして、テーブル面への手の接触判定について本章では提案する。

さて、テーブル型システム上でのユーザジェスチャであるが、目的とする位置の近隣に到達するまでは、手をテーブルから離してテーブル面とほぼ平行の状態を移動させる。目的とする地点が確定し、またそこに近づくと、手先をテーブルに接触させるべく腕を下げる。なお、テーブル面への接触に際しては、既存システムにおいては指先でタッチする手法が多いが、2章で説明した旧バージョンシステムを展示会にて一般来場者（特に小中学生と園児）に紹介した際、手のひらや上半身を乗り出すという動作が多く見受けられたことから、手のひらをテーブル面に接触するというジェスチャを想定する。

腕と影の領域の関係から同ジェスチャを観察すると、テーブル面から離して手を移動させている途中では、影の領域は腕の領域とほぼ平行に、かつ離れた状態にある。一方、手のひらがテーブル面に接地された時点では、腕領域と対応する影の領域が重なる。このことを踏まえてテーブル面への手の接触判定を行なう。

具体的には、まず前章で説明した手法によって切り出された腕ならびに影の領域に対し、輪郭を構成する点の集合を特徴点とする。図 8 に、与えられた腕と影領域に対する特徴点の例を示す。それらをデータ要素として回帰直線を求めることによって、それぞれの領域の軸を算定する。領域がテーブルの端部分に存在する場合には課題も残るが、周辺部を除く領域に腕が差し込まれている状態にあっては、腕の長さ方向の中心軸が回帰直線と重なると期待できる。

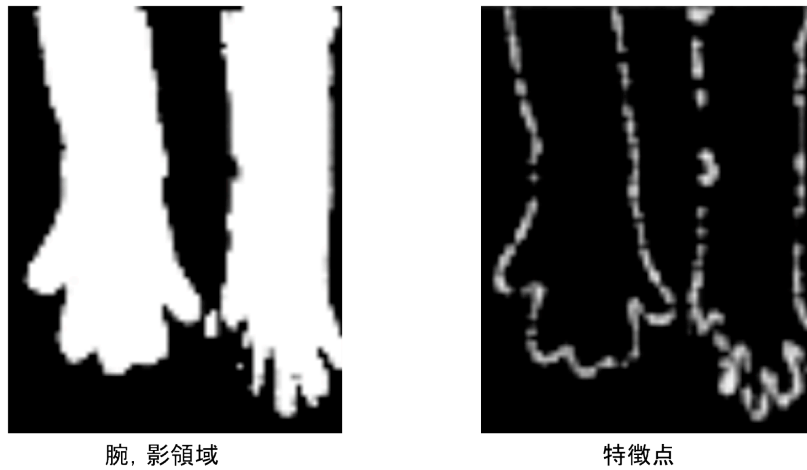


図 8 特徴点

テーブル面から手が離れている場合および接触している場合の、腕と影のそれぞれの領域に対して求められた回帰直線を図 9 に示す。腕の長さ方向に沿って回帰直線が求まっていることが分かる。

手のひらのテーブル面への接触判定にあたっては、腕と影のそれぞれの領域から求められた回帰直線の交点の位置を評価する。手のひらをテーブル面に接したときは、図 9 (b) に観察されるように、交点は必然的に手のひら領域内に収まる。このことを踏まえ、手先がテーブル面に近づくにつれて交点が腕および影領域に近づき、最終的に腕領域内に重なった時に、手先が接触したと見なす。

ちなみに接触面は手のひらを想定しており、回帰直線が成す交点より先端部にある領域を接触面として判断する。しかしながら、これはテーブル正面から手を差し入れた際の設定であり、残りの 3 方向にはそれぞれ微調整を加える必要がある。

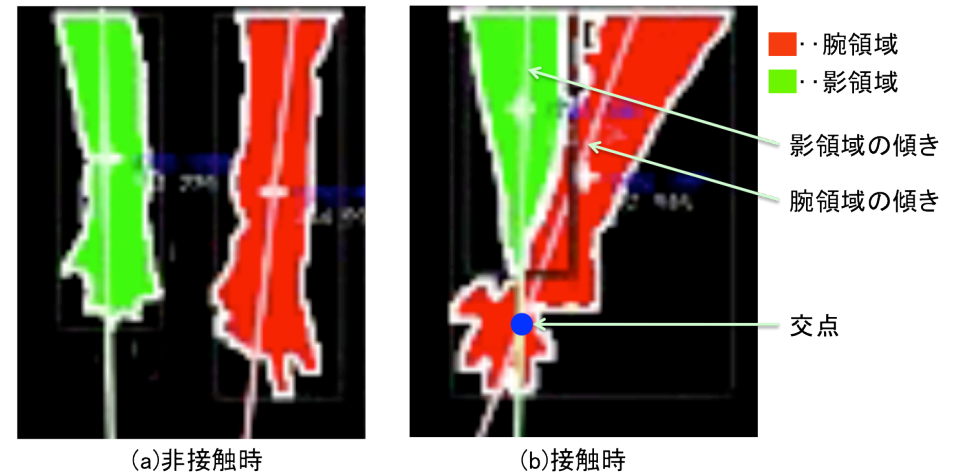


図 9 領域の傾きを用いた接触判定

## 7. おわりに

本研究では、テーブル型マルチメディアシステムへの適応を目指したジェスチャ動作による操作手法を提案した。テーブル面に生成される影に着目し、影領域と腕領域を特徴付ける軸の角度を用いてテーブル面への接触判定を行うことが可能である。

今後は、まず提案手法の有効性を確認するための評価実験を遂行する他、2つの回帰直線が互いに成す角度によって腕部分の3次元傾きの推定を行うことが研究テーマとして残されている。また、提案手法を用いたテーブル型マルチメディアシステムのアプリケーション開発に取り組みたいと考えている。

## 参考文献

- 1 松下光範, 土方嘉徳, 杉原敏昭(編):テーブル型システムの現状, ヒューマンインタフェース学会誌, vol.9, No.1, pp.69-92, 2007.
- 2 Takeshi N, Koyama T, Hirakawa, M:Development of a collaborative multimodal system with a shared sound display, IEEE International Conference on Ubi-Media Computing, pp.14-19, 2008.
- 3 M Hirakawa, M Oka, T Koyama, T Hirotoomi:Demonstrating the Effectiveness of Sound Spatialization in Music and Therapeutic, International Conference on Distributed Multimedia Systems, pp. 3-9, 2009.
- 4 NUI Group, Multi-Touch Technologies,Community Release(1<sup>st</sup> edition), 2009.
- 5 Han J.Y.:Low-Cost Multi-Touch Sensing through Frustrated Total Internal Reflection, Proc. 18<sup>th</sup>

- ACM Symp. User Interface Software and Technology, pp.115-118, 2005.
- 6 Welcome to Microsoft Surface, <http://www.microsoft.com/surface/>.
- 7 DiamondTouch by Circle Twelve Inc, <http://www.circltwelve.com/home.html>.
- 8 福地健太郎, 暦本純一: SmartSkin を用いた多点入力システムの実装, 情報処理学会論文誌, vol.46, No.7, pp.1682-1692, 2005.
- 9 Wilson A: PlayAnywhere: a compact interactive tabletop projection-vision system, Proceedings of the 18th annual ACM symposium on User interface software and technology, pp.83-92, 2005.
- 10 竹本健一, 石川裕治, 宮崎早苗: エッジ背景差分を用いた変化領域補完による車両抽出, 電子情報通信学会総合大会講演論文集, 2005\_情報・システム (2), p.127, 2005.