

ハンドマウスとその応用：色情報と輪郭情報に基づく手の検出と追跡

蔵田武志 興相正克 加藤丈和 大隈隆史 坂上勝彦

産業技術総合研究所 知能システム研究部門
〒305-8568 茨城県つくば市梅園 1-1-1 中央第2

kurata@ieee.org

<http://unit.aist.go.jp/is/hcv/vizwear/>

あらまし

本稿では、ウェアラブルカメラで得られた映像から着用者の手を検出し追跡することで、着用者の手をポインティングデバイスとして用いることを可能にするハンドマウスインターフェイスと、その応用について述べる。まず、色情報と輪郭情報を統計的に扱い、着用者の手を高速で安定に検出・追跡する手法について述べる。次に、ハンドマウスインターフェイスに適した応用として、仮想リモコンの操作と実環境中の文字情報収集の2つについて簡潔に紹介する。

キーワード： コンピュータビジョン，ヒューマンインターフェイス，ウェアラブルコンピューティング，逐次型モンテカルロ追跡法

The HandMouse and Its Applications: Color- and Contour-Based Hand Detection and Tracking

Takeshi Kurata, Masakatsu Kouroggi, Takekazu Kato,
Takashi Okuma, and Katsuhiko Sakaue

Intelligent Systems Institute,
National Institute of Advanced Industrial Science and Technology (AIST),
1-1-1 Umezono, Tsukuba, Ibaraki, 305-8568 JAPAN

Abstract

In this paper, we describe a novel user interface for wearable systems called “HandMouse” that enables us to use the wearer’s hand as a pointing device by means of detecting and tracking the hand in live video sequences taken with a wearable camera. We propose a method to detect and track a hand robustly and in real-time by generating hand- and background-color models dynamically and by tracking the hand contour statistically. Then we briefly introduce two promising applications of the HandMouse interface: namely, a virtual universal remote control and OCR in real scene images.

Key words computer vision, human interface, wearable computing, sequential Monte Carlo tracking

1 まえがき

ウェアラブルコンピューティングは、そのために必要とされるハードウェアの小型化や高性能化などの進歩とともにますます注目されるようになってきている。コンピュータ、カメラなどのセンサ、ディスプレイ及び他のデバイスを身に付けることの利点の1つは、システムとその着用者が同じ経験を共有することができることである。つまり、ウェアラブルシステムには、着用者の状況を理解し、さらに迅速なフィードバックを着用者に示すことで、着用者の支援を実現するための潜在能力があるといえる [13, 14, 20]。

しかし、その潜在能力を引き出し、状況把握に基づくアプリケーションを実用化するには、実世界データを如何に収集し、収集したデータから着用者の状況を如何に導き出し、さらにその結果に基づいて如何に有益でタイムリーな情報を提示しサービスを提供するかについての技術的な問題を解決する必要がある。

コンピュータビジョン (CV) は、着用者の状況を把握するのに有効な実世界データを収集する手段として期待されている。また、拡張現実感 (AR) 技術を利用した情報提示は、着用者が情報を直感的に理解するのに有効である。よって、この CV と AR 技術のウェアラブルシステムへの応用が、筆者らが目指す人間中心型のウェアラブルシステムやそれにより提供されるサービス (VizWear) に関する研究の中心課題となっている [1, 9, 10]。

本稿では、この VizWear システムの特徴的な入力インターフェイスである“ハンドマウス”について述べる。まず、色情報と輪郭情報を統計的に扱い、着用者の手を高速で安定に検出・追跡する手法について述べ、次に、ハンドマウスに適した応用として、仮想リモコンの操作と実環境中の文字情報収集の2つについて簡潔に紹介する。

2 ハンドマウス

着用者の位置や方向、注目している物体、会った人の顔などについての情報収集、およびそれらの情報に基づく状況把握 [5, 6, 7, 8, 15, 16] は、自律的な入力手段であるといえる。しかし、その結果を用いて構築される AR 環境やウェアラブルシステムと対話するためには、明示的な入力インターフェイスも必要となる。しかし、従来のデスクトップ環境で用いられているインターフェイスをそのままウェアラブル環境に適用することは、その携帯性、操作性などが不十分であるため、適当ではない。そこで、これまでさまざまな手法やシステムが提案・開発されてきた [2, 14, 18, 19, 21, 23, 25]。

筆者らは、ウェアラブルカメラで得られた映像中の着用者の手や指を検出することで、手や指をポインティングデバイスとして用いることを可能にするインターフェイス

について研究開発している。このハンドマウスと名付けられたインターフェイスは、キーボードのように長時間継続して多くの情報を入力する目的には向いていないが、とっさにシステムに何かを指示したい場合や簡単な指示で十分な場合には適したインターフェイスといえる。なお、このようなインターフェイスのコンセプトは、文献 [14, 19] ですでに述べられているが、具体的な実現方法やアルゴリズムについては述べられていない。

これまで筆者らは、手の検出・追跡のための動的な色モデル生成とそれに基づく手領域の切り出しについて検討してきた [11, 12]。この手法は、手の大まかな位置を把握するには有効であったが、手や指の位置や手形状 (姿勢) の高精度な推定は困難であった。そのため、ハンドマウスを実現するためには、この色情報と、輪郭やテクスチャなどの他のキュー (モダリティ) とを統合していく必要がある。

このような複数のキューを用いた追跡手法として、ICONDENSATION [4] や Co-inference [24] と呼ばれる逐次モンテカルロ追跡法が提案されている。この手法は複数の仮説を保持しながら追跡を行うため、非ガウスノイズや非線形運動、および複数の状態が混在する場合についても適用できるといった優れた特徴を持つ。本稿では、動的な色モデル生成に基づく手領域の分離と、手の輪郭追跡を、この逐次型モンテカルロ追跡法の枠組みを用いて効果的に組み合わせることでハンドマウスの実現を目指す。図 1 に提案する手の検出・追跡アルゴリズムの概略を示す。

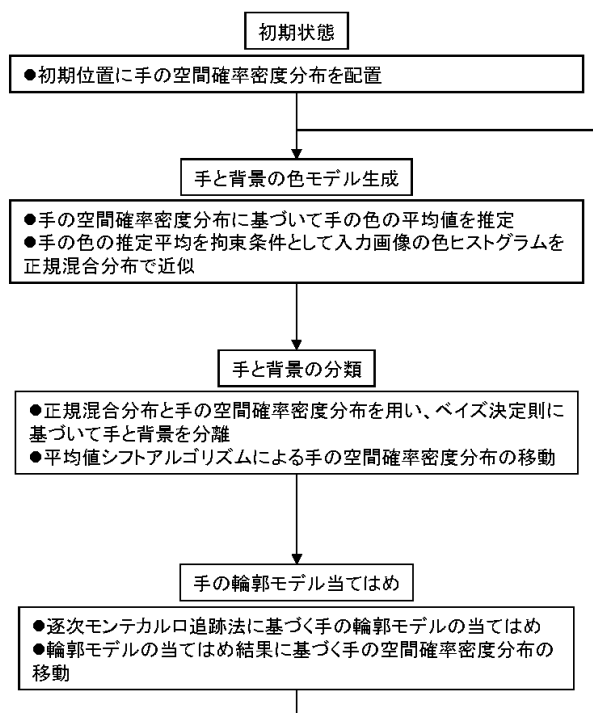


図 1: ハンドマウスのための手の追跡処理の流れ

2.1 手と背景の色モデルの動的生成

手と背景の分離は、手の検出と追跡のために欠かせない処理である。しかし、デスクトップ PC や壁などの環境側に固定されたカメラから得られる画像と比べ、ウェアラブルカメラから得られる画像は、着用者の動き・移動により生じる照明条件や背景の変化をより多く含むため、背景差分処理や予め定義された肌色モデルを用いた切り出し処理は適用できない場合が多い。

本手法では、手を検出・追跡するために、まず色情報を用いる。ただし、先に述べたような予め定義された肌色モデルを用いるのではなく、入力画像から動的に手と背景の色モデルを生成する [25]。この色モデルは、制限付き EM アルゴリズムを用いて得られる正規混合分布 (GMM) で入力画像の色ヒストグラム¹を近似することにより生成される。図 2 に実際に正規混合分布で近似された入力画像の色ヒストグラムを示す。ここでは、分布の 1 つが手の色モデル、他の混合分布が背景の色モデルを表している。

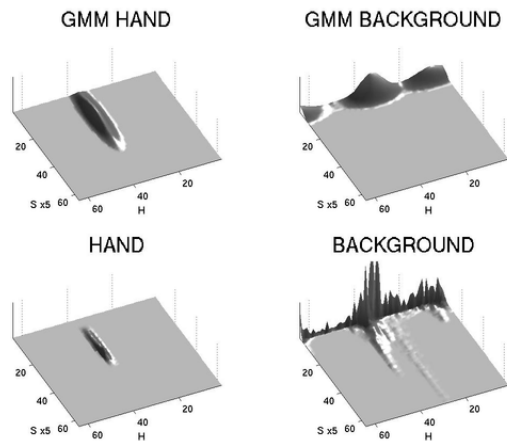


図 2: 手と背景の色ヒストグラムの例 (上) 生成された正規混合分布 (下) 入力画像の色ヒストグラム

2.2 色モデルを用いた手と背景の分離処理

本手法では、色モデルを生成する際に、色よりも安定した事前知識として図 3 のような手の画素の空間確率密度分布を用いている。この分布に基づいて手の色の推定平均を求め、手の色を表す正規分布の位置を固定している。

ここで着用者が右手を使うとすると、手の検出段階では、右手は画像の右下から現れるため、この確率密度分布は右下に偏ることになる。この確率密度分布と生成された色モデルを用いて最終的にベイズ決定則によって手の色を持つ画素とそうでない画素とを分類する。

¹ここでは、HSI 表色系の HS 色空間の 2 次元ヒストグラムを用いている。

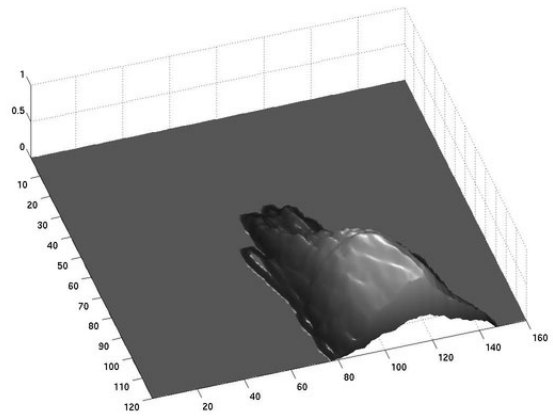


図 3: 手の空間確率密度分布 (44 枚の画像から学習)

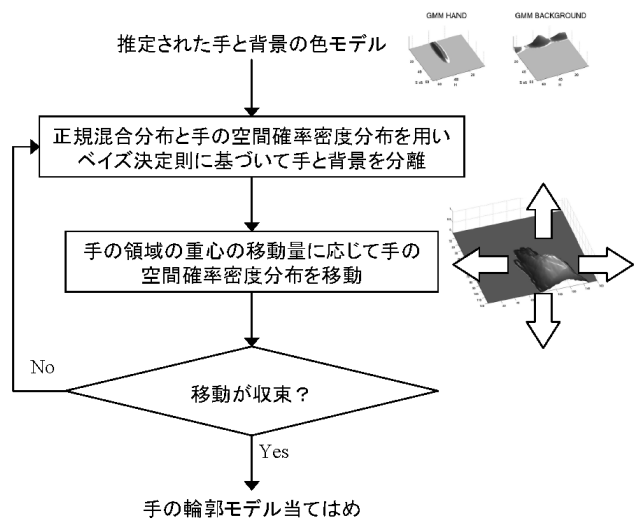


図 4: 生成された色モデルを用いた手と背景の分離処理

手の追跡段階では、画像上のどの位置に手の画素が分布しているかが不明であるため、まず、直前の追跡結果に従って図 3 の確率密度分布を移動する。さらに、平均値シフトアルゴリズム [3] に基づいて図 4 のように反復的にその確率密度分布を移動する。これにより、後述の輪郭情報を用いずにある程度の位置推定が可能となる (詳細については、文献 [11, 12] を参照)。

2.3 逐次モンテカルロ法による輪郭追跡

色情報により切り出された手の画素のモーメントの計算などにより、手の重心位置や大体の大きさなどの推定は不可能ではないが、手や指の正確な位置や動作の種類を認識することは困難である。そこで、本手法では、色モデルを用いて手と背景を分離した後に、図 6 右下に示すような手の輪郭モデルを用いて手の追跡処理を行う。この追

跡処理には、色モデルにより切り出された手の画素の位置情報に基づく重点サンプリング (Importance Sampling) [17] を導入した逐次モンテカルロ追跡法を適用する。これは、ICONDENSATION[4] と類似した手法であるが、最大の相違点は、本手法が動的に色モデルを生成する一方、ICONDENSATION は予め定義された肌色モデルを用いていることである。

手の輪郭追跡処理の概略を図5に示す。本手法では、計 N 個のサンプルのうち、手の色を持つとされる画素の重心周辺でランダムに n 個のサンプルを生成する。これが重点サンプリングに基づく部分である。残りの $N - n$ 個のサンプルは直前の追跡結果に基づいて生成される。これにより、急激な動きと、色があまり信用できない場合の両方に対応し、そのバランスは重点サンプリングの数 n により変更できる。

各サンプルは、輪郭上に設定された C 個の観測点の法線方向の輝度勾配の大きさ、つまり、輪郭と同じ方向のエッジの強さにより評価される。ただ、実際には、ランダムサンプリングによって手の状態を離散的に観測しているため、輪郭モデルと手が完全に重なるとは限らない。そこで、観測点とエッジとの距離に応じて評価値が減衰するようにし、観測点に近いエッジが評価に含まれるようにする。

そして、各サンプルの評価値 (尤度) は保持され、直後のサンプル生成の重みとして用いられる。このように、各時点の状態は尤度分布で得られるが、ハンドマウスでは、各時点での手の位置や動作の種類を一つに決めなければならない。本手法では、最大評価値の $M\%$ 以上の評価値を持つサンプルのパラメータをそれぞれの評価値で重み付けして平均することで、各時点での推定値を求める。さらにその推定値を時系列フィルタにより平滑化することで、動作を安定させる。

本手法による手の追跡結果の例を図6に示す。ここでは、輪郭のパラメータとして、2次元アフィン変換パラメータと1次元の変形パラメータを用いており、その変形を、(1) 親指と人差し指が離れている、(2) 閉じている、(3) その中間の3種類のテンプレート間の線形補間で表現している。図中、手の色を持つ画素として分類された画素は黒で表示されている。また、親指の指先だとみなされた部分に矢印または星印を表示している。矢印はポインティング (親指と人差し指が離れている) を、星印はクリック (親指と人差し指が閉じている) を示す。なお、 $C = 21$ 、 $N = 500$ とした場合、手の追跡処理に必要な計算時間は、PentiumIII(1GHz) で約55ミリ秒であった。

3 ハンドマウスの応用

この節では、VizWear システム (システムの詳細については、文献 [1, 9, 10] を参照) 上に構築されたハンドマ

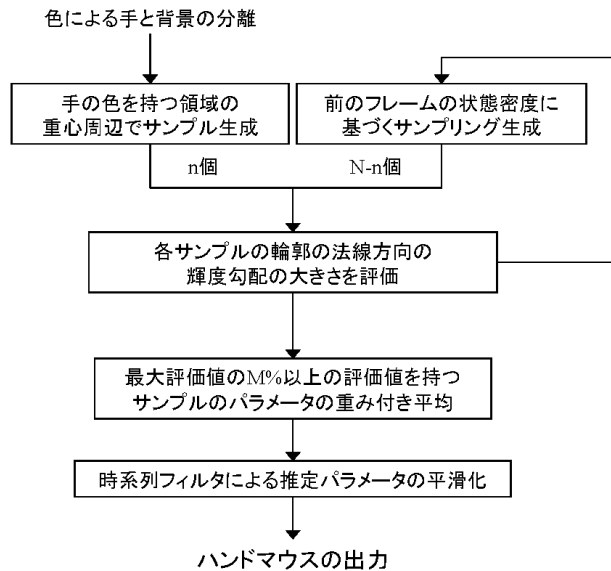


図5: 手の輪郭追跡処理の流れ

ウスに適した応用例として、仮想リモコンの操作と実環境中の文字情報収集の2つについて簡潔に紹介する。

3.1 ユニバーサルリモコン

筆者らは、事前知識のパノラマ画像と入力画像とを高速に位置合わせすることで着用者の位置と方向を推定し、同時に着用者視点に関する2次元の注釈情報を重畳提示する、パノラマベースドアノテーション [7, 8] と呼ばれる手法を開発している。この手法を用いると、ある操作対象機器、例えば、テレビやエアコンの方向を見るとそれらの仮想リモコンを着用者に提示することが可能となる²。この機能とハンドマウスを組み合わせると、何も手に持たなくても家電を操作できるユニバーサルリモコンを実現することができる。実際には、図7に示すように、VizWear システムからのコマンドを家電に伝送する必要があるが、筆者らが製作したLAN対応の学習リモコンを用いることで、赤外線受光部を備える一般的な家電のほとんどをネットワーク化することができる。

図8に、VizWear システム上に実装されているユニバーサルリモコンの動作例を示す。この例では、まず、パノラマベースドアノテーションタスクにより、システムは着用者がテレビの方向を向いたことを認識し、テレビの仮想リモコンを表示している。着用者はそのリモコンをハンドマウスで仮想的に触り、テレビを操作している。図左は電源を入れ、図右はボリュームを上げている様子を示している。

²Bluetoothなどの無線デバイスのみで同様の機能を提供しようとした場合、機器の方向まで認識するのは一般的に困難である。

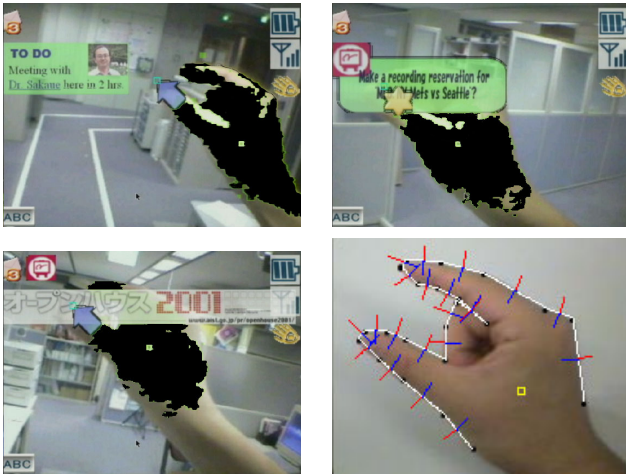


図 6: (上段と左下)手の追跡結果の例(右下)手の輪郭モデルの例

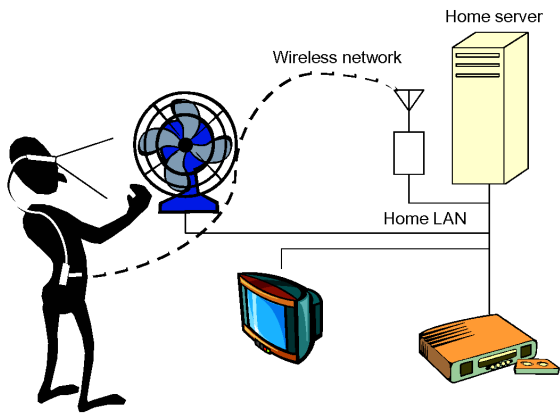


図 7: 見ている方向の家電が素手で操作できるユニバーサルリモコンシステム



図 8: VizWear システムのユニバーサルリモコンによるテレビの操作例

名などに基づくナビゲーションサービス, また, 施設予約サービスなどにも適用できる.

VizWear システム上に実装した文字情報収集機能の動作例を図 10 に示す. 本システムの文字認識部には, メディアドライブ社の活字文字認識エンジンを使用している. この文字認識エンジンは, 一般の手書き文字認識エンジンと同様, 単語辞書を用いた文字認識が可能である. 小型カメラで撮影された映像は一般に低画質であるため, このような単語情報の利用は認識率の向上に有効である. ただし, 現状では撮影の仕方や矩形領域の指定の仕方によっては正しい認識結果が得られない場合もあり, 今後の改良が必要である.

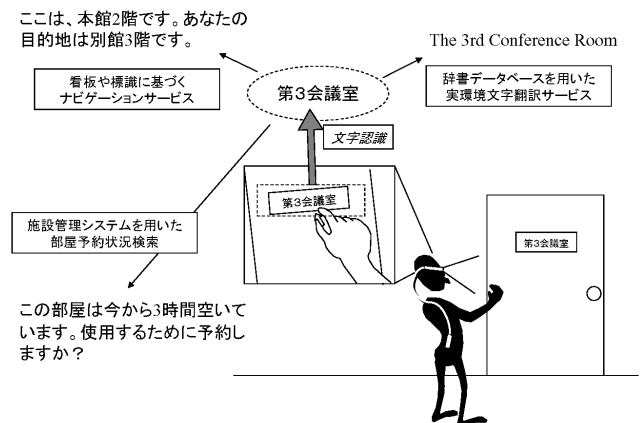


図 9: ハンドマウスを用いた文字情報収集方式とその利用

3.2 実環境の文字情報収集

携帯電話をはじめとするウェアラブル/モバイルシステムには, 通常, フルキーボードを搭載することは困難なため, 如何に効率よく文字情報を入力するかが問題となる. 本稿では, 実環境中に既に存在する文字をハンドマウスで指示することによりその文字情報をコンピュータに取り込む新しい方式について述べる. この方式は, 既にさまざまな方式が提案されている小型キーボードやペンタイプのインターフェイス [22] とは競合するものではなく, 補完し合うものと考えられる.

図 9 は, 本稿で提案する文字入力方式とその利用についての概略図である. まず, 着用者は入力したい文字を含む領域をハンドマウスで指定する. 現在は, ハンドマウスでドラッグして設定される矩形領域を文字認識処理に用いる領域としている. 認識結果は, ウェブ検索のキーとして利用できるのはもちろん, 図のように, 翻訳サービスや地

4 むすび

本稿では, ハンドマウスインターフェイスのために必要となる手の検出・追跡処理を, 逐次モンテカルロ追跡手法の枠組みで色情報と輪郭情報を効率よく統合することにより実現した. また, ハンドマウスに適した応用として, 仮想リモコンの操作と実環境中の文字情報収集の 2 つについて紹介した.

VizWear システムは無線 LAN を介したクライアント /

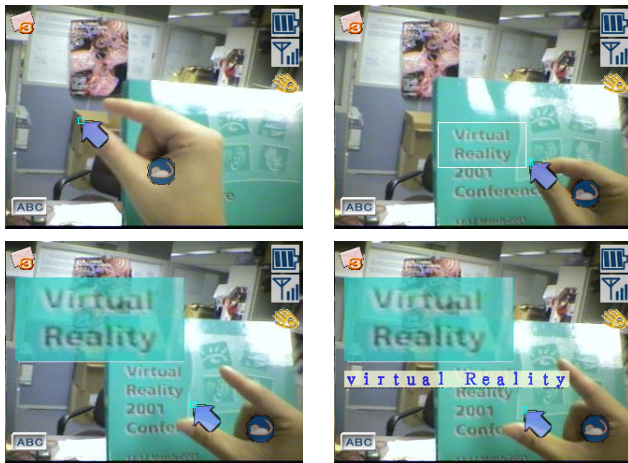


図 10: 実環境からの文字情報収集の例 (上) 範囲指定
(下) 文字認識結果

サーバシステムであり、ハンドマウスは、ウェアラブルクライアントでの画像入力、結果出力処理、サーバでの手の追跡処理により構成されている。この構成はフレームレートの面では有利だが遅延の面では問題がある。現在の実装では、約 1 秒の遅延があるため、快適に使えるようになるまでには慣れが必要である。今後は、クライアントとサーバでの処理の最適配置 [5, 6] について、フレームレートと遅延がそれぞれ使用感にどう影響するのかを調査する必要がある。

謝辞: 本研究の一部は、経産省リアルワールドコンピューティング (RWC) プログラム及び文科省科学技術振興調整費の支援による。VizWear システムに組み込まれている文字認識ライブラリの開発は、メディアドライブ株式会社遠藤健氏らによる。記して謝意を表す。

参考文献

- [1] VizWear, <http://unit.aist.go.jp/is/hcv/vizwear/>.
- [2] 福本雅朗: 24 時間ニューロヨクデキマスク? ~ Wearable なインターフェイス~, 情報処理学会誌, Vol. 41, No. 2, pp. 123-126 (2000).
- [3] Fukunaga, K.: *Introduction to Statistical Pattern Recognition*, Academic Press, Boston (1990).
- [4] Isard, M. and Blake, A.: ICONDENSATION: Unifying low-level and high-level tracking in a stochastic framework, *Proc. 5th European Conference on Computer Vision (ECCV98)*, pp. 893-908 (1998).
- [5] 加藤文和, 蔵田武志, 坂上勝彦: VizWear-Active - 記憶補助のための顔画像検出, 追跡, 登録 -, 画像情報システム研究会, 映像情報メディア学会 (2001).
- [6] Kato, T., Kurata, T. and Sakaue, K.: Face Registration Using Wearable Active Vision Systems for Augmented Memory, *Proc. 6th Australasian Conference on Digital Image Computing Techniques and Application (DICTA2002)* (2002).
- [7] Kourogi, M., Kurata, T. and Sakaue, K.: A panorama-based method of personal positioning and orientation and its real-time applications for wearable computers, *Proc. 5th Int'l Symp. on Wearable Computers (ISWC2001)*, pp. 107-114 (2001).
- [8] 興相正克, 蔵田武志, 坂上勝彦, 村岡洋一: パノラマ画像群を位置合わせに用いたライブ映像上への注釈提示とその実時間システム, 信学論, Vol. J84-D-II, No. 10, pp. 2293-2301 (2001).
- [9] Kurata, T., Okuma, T., Kourogi, M., Kato, T. and Sakaue, K.: VizWear: Toward Human-Centered Interaction through Wearable Vision and Visualization, *Proc. 2nd IEEE Pacific-Rim Conf. on Multimedia (PCM2001)*, pp. 40-47 (2001).
- [10] 蔵田武志, 大隈隆史, 興相正克, 加藤文和, 坂上勝彦: VizWear: コンピュータビジョンとウェアラブルディスプレイによる人間中心インタラクション, 高臨場感ディスプレイフォーラム 2001 (2001).
- [11] Kurata, T., Okuma, T., Kourogi, M. and Sakaue, K.: The Hand-mouse: GMM Hand Color Classification and Mean Shift Tracking, *Proc. 2nd Int'l Workshop on Recognition, Analysis and Tracking of Faces and Gestures in Real-time Systems (RATFG-RTS2001) in conjunction with ICCV2001*, pp. 119-124 (2001).
- [12] 蔵田武志, 大隈隆史, 興相正克, 坂上勝彦: ハンドマウス: 正規混合分布による色モデルの動的生成と平均値シフトを用いた手の検出と追跡, 第 7 回画像センシングシンポジウム (SSII2001), pp. 405-410 (2001).
- [13] Lamming, M. and Flynn, M.: "Forget-me-not" Intimate Computing in Support of Human Memory, Technical Report EPC-1994-103, RXRC Cambridge Laboratory (1994).
- [14] Mann, S.: Wearable Computing: A First Step Toward Personal Imaging, *Computer*, Vol. 30, No. 2, pp. 25-32 (1997).
- [15] Okuma, T., Kurata, T. and Sakaue, K.: Real-Time Camera Parameter Estimation for 3-D Annotation on a Wearable Vision System, *IEICE Trans. Inf. Syst.*, Vol. 84, No. 12 (2001).
- [16] Okuma, T., Kurata, T. and Sakaue, K.: VizWear-3D: A Wearable 3-D Annotation System Based on 3-D Object Tracking using a Condensation Algorithm, *Proc. IEEE Virtual Reality 2002 Conference (IEEE VR2002)* (2002).
- [17] Ripley, B.: *Stochastic simulation*, Wiley, NY (1987).
- [18] Starner, T., Auxier, J., Ashbrook, D. and Gandy, M.: The Gesture Pendant: A Self-illuminating, Wearable, Infrared Computer Vision System for Home Automation Control and Medical Monitoring, *Proc. 4th Int'l Symp. on Wearable Computers (ISWC2000)*, pp. 87-94 (2000).
- [19] Starner, T., Mann, S., Rhodes, B., Levine, J., Healey, J., Kirsch, D., Picard, W. R. and Pentland, A.: Augmented Reality Through Wearable Computing, Technical Report 397, M.I.T Media Lab. Perceptual Computing Section (1997).
- [20] Starner, T., Schiele, B. and Pentland, A.: Visual Contextual Awareness in Wearable Computing, *Proc. 2nd Int'l Symp. on Wearable Computers (ISWC'98)*, pp. 50-57 (1998).
- [21] Starner, T., Weaver, J. and Pentland, A.: A Wearable Computer Based American Sign Language Recognizer, *Proc. 1st Int'l Symp. on Wearable Computers (ISWC'97)*, pp. 130-137 (1997).
- [22] 増井俊之: モバイル文字入力手法情報, <http://www.csl.sony.co.jp/person/masui/OpenPOBox/info/InputMethods.html>.
- [23] Vardy, A., Robinson, J. and Cheng, L.-T.: The WristCam as Input Device, *Proc. 3rd Int'l Symp. on Wearable Computers (ISWC'99)*, pp. 199-202 (1999).
- [24] Wu, Y. and Huang, T. S.: A Co-inference Approach to Robust Visual Tracking, *Proc. The 8th IEEE Int'l Conf. on Computer Vision (ICCV2001)*, Vol. 2, pp. 26-33 (2001).
- [25] Zhu, X., Yang, J. and Waibel, A.: Segmenting Hands of Arbitrary Color, *Proc. 4th Int'l Conf. on Automatic Face and Gesture Recognition (FG2000)*, pp. 446-453 (2000).