第20回HI学会研究会・日本VR学会ウェアラブル/アウトドア
VR研究委員会第2回研究会 , Vol.4, No.5, pp.31-36 (2002)

ウェアラブルビジュアルインターフェースのための機能分散型ハンドトラッキング手法とその応用

蔵田武志 *1　興梠正克 *1　加藤丈和 *1　大隈隆史 *1　遠藤健 *2

A Functionally-Distributed Hand Tracking Method for Wearable Visual Interfaces and Its Applications

Takeshi Kurata*1, Masakatsu Kourogi*1, Takekazu Kato*1,
Takashi Okuma*1, and Ken Endo*2

Abstract - This paper describes a Functionally-Distributed (FD) hand tracking method for hand-gesture-based wearable visual interfaces. The method is an extension of the Distributed Monte Carlo (DMC) tracking method which we have developed. The method provides coarse but rapid hand tracking results with the lowest possible number of samples on the wearable side, and can reduce latency which causes a decline in usability and performance of gesture-based interfaces. The method also provides the adaptive tracking mechanism by using the sufficient number of samples and the hand-color modeling on the infrastructure side. This paper also introduces several applications of the hand-gesture-based wearable visual interfaces implemented on our wearable systems.

Keywords: sequential Monte Carlo tracking, computer vision, gesture-based interface, wearable computing, distributed computing environment

1. はじめに

コンピュータビジョン (CV) は，拡張現実 (AR) インターフェースや PUI のようなポスト WIMP 型のインターフェースを提供するためのキーテクノロジーのひとつである．ここ数年は，特にモバイル・ウェアラブルシステムのためのハンドジェスチャ(HG) インターフェースを開発するために多くの試みがなされている [7, 11, 12, 13, 14, 15, 16, 17, 19]．　しかしながら，それらの多くは，照明変化や背景（環境）の変化に敏感であるか，もしくは，装着性確保のため機器サイズによって，しばしば計算リソースやバッテリ容量が制限されるウェアラブルＰＣ単体にとっては，計算コストが高いものである．

筆者らはこれまで，そのようなウェアラブルシステムの計算パワーの不足を補うためクライアント・サーバ型のウェアラブルビジョンシステムを開発してきた [8, 10, 11]．そのシステムでは，ウェアラブルクライアントが画像のキャプチャ・デコード・エンコードやディスプレイ出力のような入出力タスクのみを処理し，サーバが無線ネットワークを介して送られてくる画像データを用いた CV タスクをすべて処理していた．しかしながら，そのようなサーバによって，高フレームレートで計算コストの高いタスクに対処できても，無線通信を経由するため，ウェアラブルクライアントに最小限の遅延で応答するのは極めて困難であった．加えて，無線通信がローミングや混信，ノイズなどにより不安定になったときは，そのようなシステムの処理は容易く停止してしまう．

本稿では，ウェアラブルビジュアルインターフェース (*Weavy*[1]) のための機能分散型ハンドトラッキング手法について述べる．この手法は，ウェアラブルサイドでは，必要最小限のサンプルを用いた ConDensation アルゴリズム [2] によって，十分な精度ではないものの高速に手の追跡結果を供給する．これにより，HG インターフェースの使いやすさを削ぎ，性能の低下を招く遅延を低減することができる．また，本手法は，十分な数のサンプルを使うと共に，手の色モデルを動的に生成することで，適応的な追跡の仕組みを提供する．この適応的な追跡タスクはウェアラブルサイドモジュールにとっては計算量的にかなり重い処理であるが，低遅延で処理する必要はない．また，更新された色モデルが受信されるまでは，ウェアラブルサイドに保存してある色モデルを使うことができる．そのため，このタスクをインフラサイド上の XML ウェブサービスのひとつでして設計することができる．本稿は，さらに筆者らの開発した *Weavy* を備えるウェアラブルシステムである *Wyvern* に実装された HG インターフェースに適したいくつかのアプリケーションを紹介する．

*1: 産業技術総合研究所 知能システム研究部門
*2: メディアドライブ（株）
*1: Intelligent Systems Institute, National Institute of Advanced Industrial Science and Technology (AIST)
*2: Media Drive Corporation

2. 機能分散型ハンドトラッキング

図 1 は本稿で筆者らが提案する機能分散型 (FD) ハンドトラッキング手法の概略である．本手法は，分散モンテカルロ (DMC) 追跡法 [5] に基づいた追跡タスクと，正規混合分布 (GMM) による色ヒストグラムの近似に基づいた適応的な色モデル生成タスクからなる [11, 19]．この FD ハンドトラッキング手法の各タスクは，システムの着用者が装着しているウェアラブルサイドモジュール，もしくは，その他のインフラサイドモジュールに割り当てられている．

文献 [5, 6] において，DMC 追跡法は，ウェアラブルアクティブカメラ (WAC) を最小限の遅延で制御しながら対象人物を追跡し，さらに顔の正確な位置や形状を獲得するために適用された．インターフェースの応答遅延は，使いやすさの低下が起こらないように低減されるべきであり，環境変化への適応は *Weavy* のようなインターフェースに必須であるので，本稿では，DMC 追跡法を手の追跡に応用した．

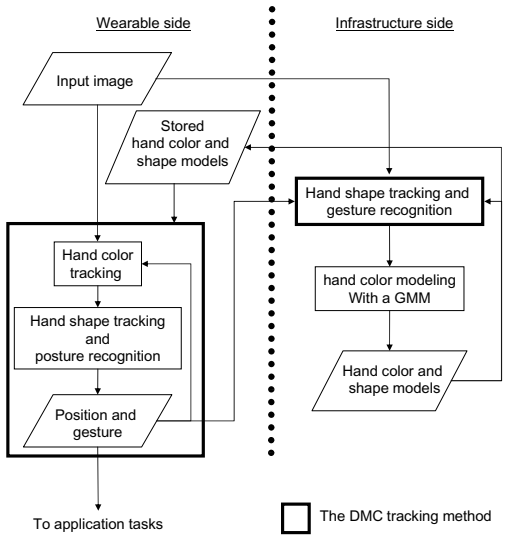


図 1: FD ハンドトラッキング手法の概略

Fig. 1: Diagram of the FD hand tracking method.

2.1. DMC 追跡法

DMC 追跡法 [5] は，ConDensation アルゴリズム [2] をヘテロな分散計算環境のために拡張したものである．ConDensation アルゴリズムの時刻 t における処理は次のような 3 つのステップからなる．

1. 重み $\pi_{t-1}^{(j)}$ $(n, j = 1, .., N)$ に従ってサンプル $\boldsymbol{s'}_t^{(n)} (= \boldsymbol{s}_{t-1}^{(j)})$ を選択する．

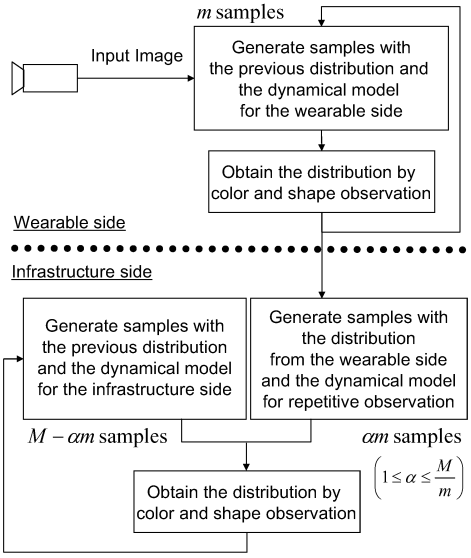


図 2: DMC 追跡法の処理の流れ

Fig. 2: Diagram of the DMC tracking method.

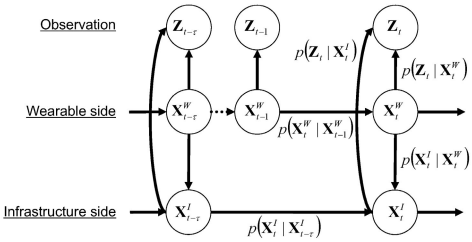


図 3: DMC 追跡法の状態遷移

Fig. 3: State transition diagram of the DMC tracking method.

2. 動的モデル $p(\boldsymbol{X}_t|\boldsymbol{X}_{t-1} = \boldsymbol{s'}_t^{(n)})$ を用いたサンプリングにより，新しいサンプル $\boldsymbol{s}_t^{(n)}$ を予測する．

3. 観測された画像特徴 $\boldsymbol{Z}_t$ によって，各 $\boldsymbol{s}_t^{(n)}$ を重み付けする $(\pi_t^{(n)} = p(\boldsymbol{Z}_t|\boldsymbol{X}_t = \boldsymbol{s}_t^{(n)}))$.

ConDensation アルゴリズムは，複数の仮説を離散的な確率密度分布 $\{\boldsymbol{s}_t^{(n)}, \pi_t^{(n)}\}$ という形で保持しているため，対象が雑多なシーンにあっても，ロバストにその対象を追跡することができる．しかしながら，その性能を十分に引き出すには非常に多くのサンプルや十分な観測が必要となり，計算量の増大を招く．この問題を解決するために DMC 追跡法は，ウェアラブルサイドでは最小限の遅延で追跡結果を得つつ，インフラサイドでは対象の状態を精度よく推定するように，ウェアラブルサイドとインフラサイドでそれぞれ異なる数のサンプルや動的モデルを用いる．

図 1 中の太線の矩形部分は，DMC 追跡法と対応しており，図 2 はその詳細を示している．インフラサイドにおいて，上述のステップ 1 では，ウェアラブルサイドから送られてくる，最新ではあるが疎なサンプルセット $\{\boldsymbol{s}_t^{W(n)}, \pi_t^{W(n)}\}$ と，インフラサイドで持っている，最新ではないが密なサンプルセット $\{\boldsymbol{s}_{t-\tau}^{I(n)}, \pi_{t-\tau}^{I(n)}\}$ から，新しいサンプル $\boldsymbol{s}_t^{I(n)}$ が生成される．$\boldsymbol{s}_t^{W(j)}$ と $\boldsymbol{s}_{t-\tau}^{I(j)}$ の選択比は，図 2 中のサンプル混合パラメータ α を用いて無線ネットワークの状態に応じて変更される．

図 3 は，DMC 追跡法の状態遷移図である．ここで，$p(\boldsymbol{Z}_t|\boldsymbol{X}_t^W), p(\boldsymbol{Z}_t|\boldsymbol{X}_t^I)$ はそれぞれ，ウェアラブルサイドで生成された対象の状態 $\boldsymbol{X}_t^W$ を観測した結果得られた尤度，インフラサイドで生成された対象の状態 $\boldsymbol{X}_t^I$ を観測した結果得られた尤度である．$p(\boldsymbol{X}_t^W|\boldsymbol{X}_{t-1}^W), p(\boldsymbol{X}_t^I|\boldsymbol{X}_{t-\tau}^I)$ はそれぞれ，ウェアラブルサイド，インフラサイドで使われる動的モデルである．$p(\boldsymbol{X}_t^I|\boldsymbol{X}_t^W)$ も動的モデルであるが，先の 2 つのモデルとは異なり，このモデルは同じ時刻の画像で得られた結果から新しいサンプルを生成するために使われる．

2.2 形状表現と観測

本稿では，手形状表現のために図 4 に示すような単純な 2 次元輪郭モデルを用い，その状態を記述するために以下の 7 つのパラメータを使用する．

$\boldsymbol{X}_t = (\theta_t, t_{x,t}, t_{x,t}, s_t, \phi_{x,t}, \phi_{y,t}, \gamma_t)$,

ここで，θ_t は光軸周りの回転角，$(t_{x,t}, t_{x,t})$ は手形状中心の位置，$s_{x,t}$ はスケールパラメータ，$(\phi_{x,t}, \phi_{y,t})$ は，せん断パラメータである．姿勢パラメータ $\gamma_t (0 \leq \gamma_t \leq 1)$ は，モーフィングと同じ方法で，ポインティング姿勢（図 4 (a)）からクリック姿勢（Figure 4 (b)）へ手形状を変形させる．これらの姿勢を使って，ユーザは図 5 に示すように，ウェアラブル機器と対話できる．

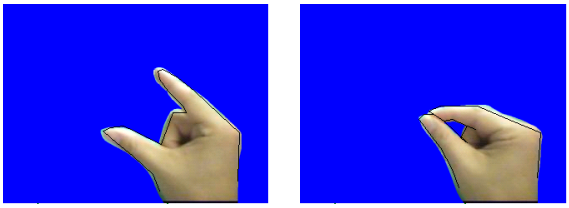
$\boldsymbol{X}_t$ の次元は，ウェアラブルサイドで生成することのできる数のサンプルで探索するには高すぎるため，ウェアラブルサイドでは，画像上での手の大きな動きに追従することに重点を置き，$(\phi_{x,t}, \phi_{y,t})$ を固定し，さらに各サンプルが疎ではあるが広い範囲に分布するように，動的モデル $p(\boldsymbol{X}_t^W|\boldsymbol{X}_{t-1}^W)$ を設計する．一方，インフラサイドでは，X_t の各パラメータが精度よく推定されるように $p(\boldsymbol{X}_t^I|\boldsymbol{X}_{t-\tau}^I)$ を設計する．

ここで，輪郭上の観測点 $P_c (c = 1, .., C)$ の周辺の輝度勾配 $\boldsymbol{g} = (g_x, g_y)^T$ の無修正平方和積和行列 $\boldsymbol{G} = \Sigma \boldsymbol{g}\boldsymbol{g}^T$ の第 1 固有ベクトルを $\boldsymbol{e}_1$ とおく．本稿の実装では，各サンプルにおいて形状を観測することにより得られる尤度を次のように定義する．

$$p(\boldsymbol{Z}_t|\boldsymbol{X}_t = \boldsymbol{s}_t^{(n)}) = \prod_{c=1}^{C} p(\boldsymbol{z}_t|\boldsymbol{X}_t) \tag{1}$$

$$p(\boldsymbol{z}_t|\boldsymbol{X}_t) \propto \begin{cases} q + |\boldsymbol{e}_1^T \boldsymbol{n}_c| & \text{if } \lambda_1 > T_\lambda \\ q & \text{otherwise} \end{cases}$$

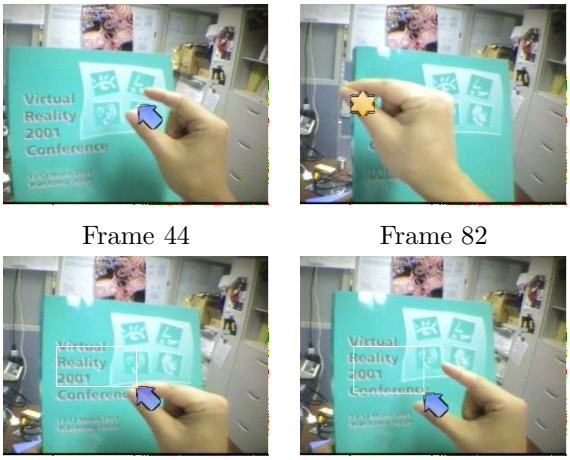
ここで，λ_1 は $\boldsymbol{G}$ の第 1 固有値，T_λ はその閾値，$\boldsymbol{n}_c$ は P_c における輪郭の法線ベクトル，q は対象の隠れなどにより一時的に観測に失敗してもサンプルを生き残らせるための定数である．



(a) Pointing [$\gamma = 0$] (b) Clicking [$\gamma = 1$]

図 4: 単純な手形状モデル．２１点の観測点からなる．

Fig. 4: A simple hand shape model. This model has 21 observation points on the contour.



Frame 44 Frame 82

Frame 121 Frame 129

図 5: ドラッグによる矩形領域の選択

Fig. 5: Selecting a rectangle by dragging.

2.3 色の表現と観測

文献 [5] と同様，本稿の実装では，形状観測の前に，背景上のサンプルを削減し，形状観測のための計算コストを抑えるために，重点サンプリング [3, 18] と同様の考え方に基づく色の観測を行う．ここでは，パラメトリックな色モデルとして HSI 表色系の HS 空間における正規分布を用いる．このモデルのパラメータは，2.4 節で述

べられるように，インフラサイドで学習され，無線通信が可能なときにウェアラブルサイドにダウンロードされる．よって，ウェアラブルサイドでの色の観測には保存された学習モデルが使われる．各サンプルにおいて，手形状の輪郭内部の画素から HS の正規分布を求め，学習モデルとの類似度をマハラノビス距離で評価することにより色を観測する．

2.4 色モデルの学習

環境の変化に対応するため，あらかじめ定義された手の色モデルを用いるのではなく，文献 [11] で提案された手の色の分離手法に基づいて，手と背景の色モデルを動的に学習する．本手法では，各入力画像の HS ヒストグラムを近似するために GMM を用いる．その GMM は，1 つ目の分布が手の色の近似となるように通常の EM アルゴリズムを改良した，制限付き EM アルゴリズム [19] により推定される．この制限付き EM アルゴリズムでは，GMM を推定するために手の空間確率密度分布を必要とする．筆者らは，DMC 追跡法で推定された手形状をその分布として用いる．この色モデル学習タスクはインフラサイドに配置され，手の色モデルとして，推定された GMM の 1 つ目の分布が，先に述べたようにウェアラブルサイドにダウンロードされる．

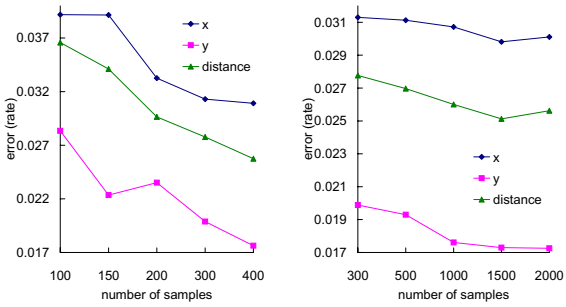
3. 実験

頭部装着型カメラにより得られた時系列画像（図 5 を含む 227 フレーム）を用いて提案手法による手の追跡精度を評価した．なお，この実験では，$\pi_t^{(n)}$ を重みとした $\boldsymbol{s}_t^{(n)}$ の平均値を推定された手の状態として用いた．

図 6 は，サンプル数を変えた時の，ウェアラブルサイドとインフラサイドでのポインティング位置の平均誤差を示している．この図には，親指の先の推定位置と手動で計測した正解位置との，x, y 軸それぞれに沿った距離，及びユークリッド距離が含まれている．これらの結果は，入力画像の幅と高さがそれぞれ 1.0 になるように正規化されている．自明なことではあるが，サンプル数の増加に伴って，ポインティングの精度が改善されている．

図 7 は姿勢の識別誤差，及び，手があるかないかの推定における誤検出率と未検出率を示している．この実験では，単純に γ_t の閾値を使ってポインティングかクリックかを識別した．手の存在も，色と形状の観測で得られた尤度の閾値処理で判定した．図 8 は各フレームでの尤度を示している．手の誤検出，未検出はあまり有意な変化は見られなかったが，姿勢識別の性能は，サンプル数の増加に伴って向上している．ただし，誤検出の多くは，手が見えなくなった直後に発生しているため，簡単な時系列フィルタを使うだけでも，そのような誤りを軽減することができると考えられる．

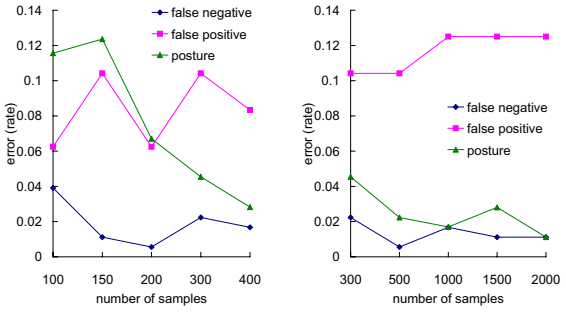
計算資源と追跡精度のバランスを考慮し，現在は，ウェアラブルサイド (CPU: Crusoe 867MHz) でのサンプル数を 300，インフラサイド (CPU: Pentium IV-M 2.2GHz) のサンプル数を 1000 に設定している．図 9 はウェアラブルサイド (N=300) で推定された親指の先の軌跡の例である．



(a) Wearable (b) Infrastructure

図 6: ウェアラブルサイドとインフラサイドにおけるポインティングの平均誤差

Fig. 6: The average pointing errors for varied numbers of samples and for the wearable and infrastructure sides.



(a) Wearable (b) Infrastructure

図 7: 姿勢識別誤差と手検出における誤検出率と未検出率

Fig. 7: Posture classification error and false positive and false negative in determining the existence of the hand.

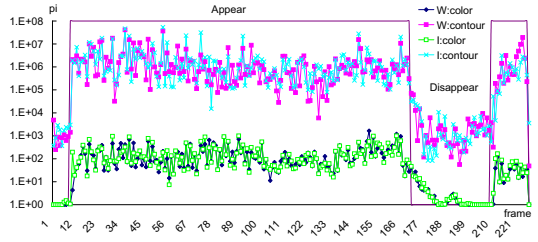


図 8: 各フレームでの尤度．太線は，手が見えているかいないかの正解データ

Fig. 8: Likelihood at each frame. Bold line shows whether the hand exist in each image or not.

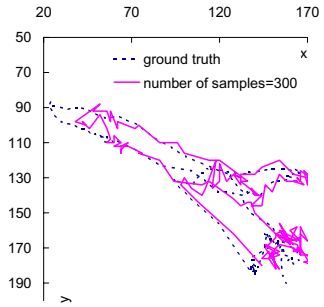


図 9: 第 9 フレームから 170 フレームのポインティングの軌跡

Fig. 9: Trajectory of the pointer from frame 9 to 170.

最後の例の実世界文字認識 (RWOCR) も文献 [9] などで紹介したアプリケーションであるが，これまではOCR タスクに渡す文字領域をドラッグにより手動で正確に切り出す必要があった．ここでは，実シーンからの文字検出手法 [4] を用いて，いくつかの候補領域を求めており，単にそれらのうちの 1 つをつまんで選択することで，実世界中の文字情報を獲得することができる．図 11 (c) はこの RWOCR の出力例であり，図 11 (d) 中の 3 つの矩形は，自動的に検出された文字領域を示している．

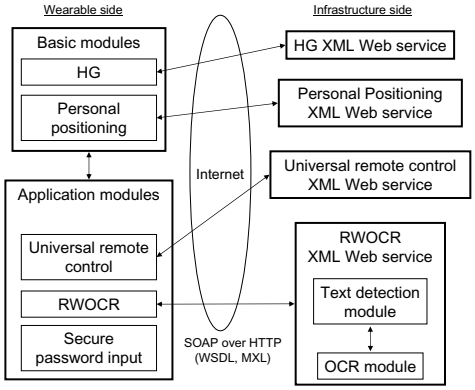


図 10: Wyvern^{TM} システムの分散型ソフトウェアアーキテクチャ

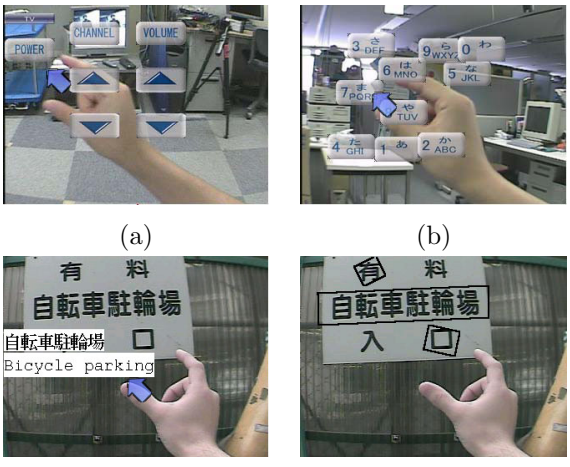
Fig. 10: Distributed software architecture of the Wyvern^{TM} system.

4. アプリケーション

筆者らは，*Weavy* を備えた複数のタイプの着用型システム *Wyvern*[1] を開発している．図 10 は，HG インターフェースを含む *Wyvern* の分散型ソフトウェアアーキテクチャの概略である．ここでは，この HG インターフェースを用いた 3 つのアプリケーションを紹介する (図 11)．

1 つ目のアプリケーションは，文献 [9] などでも紹介したユニバーサルリモコンである（図 11 (a)）．筆者らが開発しているパーソナルポジショニング手法 [8] と HG インターフェースを組み合わせると，操作対象を見るだけで，図のような仮想の操作パネルが頭部装着型ディスプレイ (HWD) に表示されるため，HG インターフェースで仮想的につまむことによりその対象を操作することができるようになる．

図 11 (b) は 2 つ目のアプリケーションである”セキュアなパスワード入力”の画面の出力例である．HWD に表示されるソフトキーボードの形状や位置が入力の度に変更されるため，たとえユーザの手の動きを注意深く見ていても，パスワードを盗むことは非常に困難になる．



(a) (b)

(c) (d)

図 11: 着用型 HG インターフェースのアプリケーション

Fig. 11: Applications of the HG-based Weavy. (a) Universal remote control, (b) Secure Password Input, (c) RWOCR. (d) Automatically detected text regions.

謝辞

本研究をサポートいただいた坂上勝彦知的インターフェースグループリーダに深く感謝します.

参考文献

[1] *Weavy: Wearable Visual Interfaces*, http://www.is.aist.go.jp/weavy/.

[2] M. Isard and A. Blake. Condensation – conditional density propagation for visual tracking. *IJCV*, Vol. 29, No. 1, pp. 5–28, 1998.

[3] M. Isard and A. Blake. Icondensation: Unifying low-level and high-level tracking in a stochastic framework. In *Proc. 5th European Conference on Computer Vision (ECCV98)*, pp. 893–908, 1998.

[4] K. Jung, K. I. Kim, T. Kurata, M. Kourogi, and J. Han. Text scanner with text detection technology on image sequences. In *Proc. 16th International Conference on Pattern Recognition (ICPR)*, Vol. 3, pp. 473–476, 2002.

[5] T. Kato, T. Kurata, and K. Sakaue. VizWear-Active: distributed monte carlo face tracking for wearable active camera. In *Proc. 16th International Conference on Pattern Recognition (ICPR)*, Vol. 1, pp. 395–400, 2002.

[6] T. Kato, T. Kurata, and K. Sakaue. VizWear-Active: towards a fanctionally-distributed architecture for real-time visual tracking and context-aware UI. In *Proc. 6th Int'l Symp. on Wearable Computers (ISWC2002)*, pp. 162–163, 2002.

[7] T. Keaton, S. M. Dominguez, and A. H. Sayed. Snap & tell: A vision-based wearable system to support 'web-on-the-world' applications. In *Proc. 6th Australasian Conference on Digital Image Computing Techniques and Application (DICTA2002)*, pp. 92–97, 2002.

[8] M. Kourogi, T. Kurata, and K. Sakaue. A panorama-based method of personal positioning and orientation and its rea-time applications for wearable computers. In *Proc. 5th Int'l Symp. on Wearable Computers (ISWC2001)*, pp. 107–114, 2001.

[9] 蔵田武志, 興梠正克, 加藤丈和, 大隈隆史, 坂上勝彦. 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. 映情学技報, VIS2001-103, 第 25 巻, pp. 47–52, 2001.

[10] T. Kurata, T. Okuma, M. Kourogi, T. Kato, and K. Sakaue. VizWear: Toward human-centered interaction through wearable vision and visualization. In *Proc. 2nd IEEE Pacific-Rim Conf. on Multimedia (PCM2001)*, pp. 40–47, 2001.

[11] T. Kurata, T. Okuma, M. Kourogi, and K. Sakaue. The Hand Mouse: gmm hand color classification and mean shift tracking. In *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] S. Mann. Wearable computing: A first step toward personal imaging. *Computer*, Vol. 30, No. 2, pp. 25–32, 1997.

[13] H. Sasaki, T. Kuroda, Y. Manabe, and K. Chihara. HIT-Ware : A menu system superimposing on a human hand for wearable computers. In *Proc. 9th Int'l Conf. on Artificial Reality and Telexistence (ICAT'99)*, pp. 146–153, 1999.

[14] T. Starner, J. Auxier, D. Ashbrook, and M. Gandy. The gesture pendant: A self-illuminating, wearable, infrared computer vision system for home automation control and medical monitoring. In *Proc. 4nd Int'l Symp. on Wearable Computers (ISWC2000)*, pp. 87–94, 2000.

[15] T. Starner, S. Mann, B. Rhodes, J. Levine, J. Healey, D. Kirsch, W. R. Picard, and A. Pentland. Augmented reality through wearable computing. *Presence*, Vol. 6, No. 4, pp. 386–398, 1997.

[16] T. Starner, J. Weaver, and A. Pentland. A wearable computer based american sign language recognizer. In *Proc. 1st Int'l Symp. on Wearable Computers (ISWC'97)*, pp. 130–137, 1997.

[17] A. Vardy, J. Robinson, and L.-T. Cheng. The wristcam as input device. In *Proc. 3rd Int'l Symp. on Wearable Computers (ISWC'99)*, pp. 199–202, 1999.

[18] Y. Wu and T. S. Huang. A co-inference approach to robust visual tracking. In *Proc. The 8th IEEE Int'l Conf. on Computer Vision (ICCV2001)*, Vol. 2, pp. 26–33, 2001.

[19] X. Zhu, J. Yang, and A. Waibel. Segmenting hands of arbitrary color. In *Proc. 4th Int'l Conf. on Automatic Face and Gesture Recognition (FG2000)*, pp. 446–453, 2000.