

VizWear-Active - 記憶補助のための顔画像検出, 追跡, 登録 -

加藤 丈和

蔵田 武志

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

E-mail : t.kato@aist.go.jp
<http://unit.aist.go.jp/is/hcv/vizwear/>

あらまし

本稿では, ウェアラブルアクティブカメラを備え, 通信機能と画像処理機能を持つ着用型アシスタントについて議論する. この VizWear-Active と呼ばれるシステムは, 画像認識やモーションセンサの情報に基づいたカメラ制御によって, 能動的な情報収集を行い注目対象の安定な観測を実現する. VizWear-Active によって実環境における安定かつ長時間での顔画像追跡を実現するために, ConDensation アルゴリズムを基にした分散型の追跡手法を提案する. また, モーションセンサの情報を利用することで, カメラ動作に対する追跡の安定化を実現する. 実際に VizWear-Active の試作システムによって顔画像追跡実験を行い, 本システムの有効性を示す.

キーワード ウェアラブルコンピューティング, コンピュータビジョン, 分散処理, 人物追跡, ConDensation

VizWear-Active: Face Detection, Tracking and Registration
for Augmented Memory

Takekazu Kato, Takeshi Kurata, and Katsuhiko Sakaue

Intelligent Systems Institute,
National Institute of Advanced Industrial Science and Technology (AIST),

1-1-1 Umezono, Tsukuba-city, Ibaraki, 305-8568, Japan

E-mail : t.kato@aist.go.jp

Abstract

In this paper, we discuss a wearable assistant with a wearable active camera. This system called VizWear-Active can actively and robustly understand the wearer and his or her environment by camera control according to image processing and motion sensors. We propose face tracking for the VizWear-Active based on the ConDensation algorithm, which realizes real-time tracking by client-server distributed sampling. Furthermore, the tracking algorithm becomes stable against camera motion using motion sensors. We confirmed the face tracking in experiments by implementation on a prototype system for the VizWear-Active.

Keywords wearable computing, computer vision, face tracking, parallel processing, ConDensation

1 まえがき

ウェアラブルシステムは常に着用者やその周囲の状況を把握し、状況に応じて適切な情報を着用者に提示することで、適応的に着用者を支援する潜在能力を持つ。一方、コンピュータビジョン技術は状況把握のために有用な手段であり、ウェアラブルシステムへの応用が期待されている [1, 5, 10] .

着用者が出会う人物に関する状況把握は、着用者自身が最も興味を持つ事柄の 1 つであり、記憶補助 [2] やセキュリティなどへの応用も期待されている。このようなアプリケーションを実現するためには、着用者が出会う人物を発見し、顔画像を獲得することが重要である。しかし、ウェアラブルシステムでは着用者自身が動きまわることによって撮影環境が時々刻々変化するため、実時間で安定に人物を発見し、その顔画像を獲得することは容易ではない。そこで我々は、ウェアラブルアクティブカメラ [6] をもつ着用型アシスタントである *VizWear-Active* によって、対象人物の動きに追従して継続的に顔画像を追跡し続けることで、実環境下において安定かつ実時間で顔画像を獲得することを目指す。本稿では、この *VizWear-Active* のコンセプトの提案と、ConDensation アルゴリズムを基にした、*VizWear-Active* による実時間顔画像検出、追跡手法について述べる。

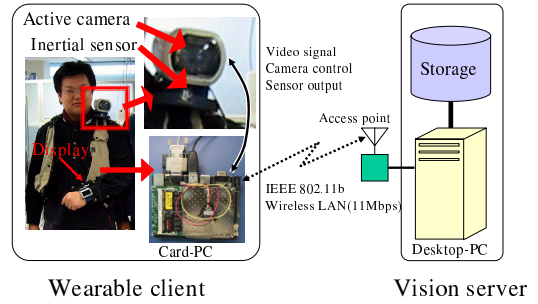


図 1: *VizWear-Active* の試作システム

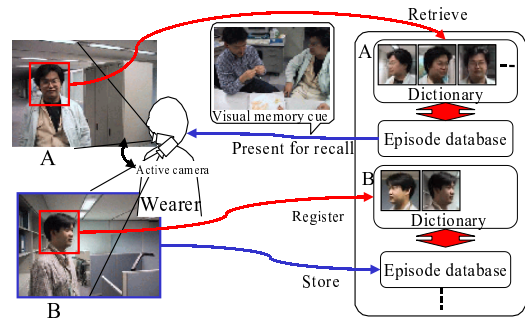


図 2: *VizWear-Active* による記憶補助

2 VizWear-Active

現在、多くのウェアラブルシステムでは、頭部など体の一部に固定されたカメラを使用している。頭部に着用したカメラでは、視線方向が着用者の頭部の動きに追従するため、着用者が注目する対象を自然に観測することが可能である。しかし、このような固定型のカメラでは、着用者の動きによって映像にぶれが生じたり、着用者の視線方向にない情報を見逃してしまうなどの欠点がある。特に人物を対象とする場合、着用者自身も対象人物も移動するため、対象人物が常にカメラの視野にいるとは限らない。そこで我々は、ウェアラブルアクティブカメラを備え、画像認識やモーションセンサの情報に基いて自律動作することで、着用者自身の視線方向とは独立に、能動的な情報収集を行うことで、注目対象を安定に観測する着用型アシスタント (ウェアラブルペットロボット) である *VizWear-Active* を開発している。

2.1 VizWear-Active システム

図 1 に *VizWear-Active* の試作システムを示す。本システムは、着用者が身につけるウェアラブルクライアント、デスクトップ PC (Intel Xeon 1.7GHz) からなるビジョンサーバ、及びそれらを接続する無線 LAN (IEEE 802.11b) から構成されている。ウェアラブルクライアントは、ウェアラブル PC (Intel mobile Pentium III 500MHz)、アクティブカメラ (SONY EVI-G20)、モーションセンサ (InterSense InterTrax²) から構成される。ウェアラブル PC は軽量かつ小型 (140 mm×105 mm×40 mm) であり、本システム

ではベストの中に組み込まれている。アクティブカメラは肩に固定され、ウェアラブル PC からパン・チルト制御が可能である。また、モーションセンサはアクティブカメラの土台に固定され着用者の動作にともなうカメラの動きを計測することができる。

多くのコンピュータビジョンアルゴリズムは、ウェアラブル PC 単体で処理するには計算コストが高い。また、個人認識処理や画像データの記録などを行うためには、大規模なデータベースが必要であり、これらをすべてウェアラブルクライアントに持たせるのは現実的ではない。本システムでは、このような処理を無線 LAN 経由でサーバに送り、協調的に処理を行うことで、実世界で動作するアプリケーションを構築できる。

なお現在の試作システムでは、カメラ、モーションセンサ、ウェアラブル PC、バッテリー (4 時間稼働分) を含むウェアラブルクライアントの総重量は約 3.8kg である。

2.2 VizWear-Active による記憶補助

VizWear-Active のアプリケーションとして、我々は特に顔画像認識技術に基づく記憶補助システムを目指している。図 2 に示すように、システムは着用者が出会った人物に関するエピソード (時刻、場所、ビデオログなど) を蓄積するエピソードデータベースと、顔画像からエピソードデータベースを検索するための顔辞書を持っている。まず、着用者が他の人物に出会ったとき、*VizWear-Active* は顔画像を検出、追跡を行いながら、既に登録されている顔辞書を使って個人識別を行う。追跡されている顔画像が顔辞書に登録されていれば、関連づけられたエピソードを着用者に提示する。また、顔画像が未登録であれば、

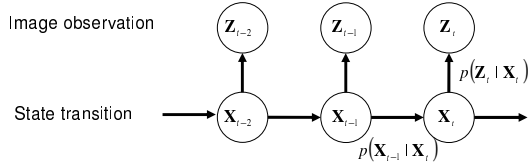


図 3: ConDensation アルゴリズム

顔辞書を新規に構築するとともに、そのときの状況をエピソードデータベースに蓄積する。

このようなアプリケーションを実現するためには、実環境において安定かつ実時間で顔画像の検出、追跡が行えることが重要である。本稿では特に *VizWear-Active* による顔画像検出、追跡技術について議論する。

3 ConDensation アルゴリズムによる人物頭部追跡

ConDensation アルゴリズム [3] は、モンテカルロ法に基づく追跡法である。この手法では、追跡対象の状態をランダムサンプリングによって離散的に表現された確率密度として保持することで、常に複数の解の候補を持ちながら追跡することが可能であり、ノイズに対してロバストな追跡が実現できる。本節では、文献 [3] で提案されている ConDensation アルゴリズムによる追跡手法について簡単に説明し、その後で本研究で実際に用いた楕円モデルによる人物頭部追跡について説明する。

3.1 ConDensation アルゴリズム

時刻 t における追跡対象の状態を \mathbf{X}_t 、画像から得られる観測結果を \mathbf{Z}_t とおく。また時刻 t までの状態と観測結果をそれぞれ $\mathcal{X}_t = (\mathbf{X}_1, \dots, \mathbf{X}_t)$ 、 $\mathcal{Z}_t = (\mathbf{Z}_1, \dots, \mathbf{Z}_t)$ とおく (図 3)。このとき追跡問題は、状態の確率密度 $p(\mathbf{X}_t|\mathcal{Z}_t)$ を推定する問題と考えることができる。ここで、ベイズの法則により次式が成り立つ。

$$p(\mathbf{X}_t|\mathcal{Z}_t) = k_t p(\mathbf{Z}_t|\mathbf{X}_t) p(\mathbf{X}_t|\mathcal{Z}_{t-1}) \quad (1)$$

ただし、 k_t は \mathbf{X}_t に依存しない正規化定数であり、 $p(\mathbf{Z}_t|\mathbf{X}_t)$ は状態 \mathbf{X}_t に対する尤度である。また、あらかじめ時刻 $t-1$ から t までの推移確率 $p(\mathbf{X}_t|\mathbf{X}_{t-1})$ が与えられているとすると次式が成り立つ。

$$p(\mathbf{X}_t|\mathcal{Z}_{t-1}) = \int p(\mathbf{X}_t|\mathbf{X}_{t-1}) p(\mathbf{X}_{t-1}|\mathcal{Z}_{t-1}) d\mathbf{X}_{t-1} \quad (2)$$

すなわち、各時刻における尤度 $p(\mathbf{Z}_t|\mathbf{X}_t)$ を画像から推定することで $p(\mathbf{X}_t|\mathcal{Z}_t)$ を得ることができる。しかし、全ての \mathbf{X}_t に対して尤度を求めることは現実的でないので、ランダムサンプリングによって推定する。

時刻 $t-1$ における状態 \mathbf{X}_{t-1} のサンプル集合を $\{s_{t-1}^{(1)}, \dots, s_{t-1}^{(N)}\}$ 、それぞれのサンプルに対する尤度を $\{\pi_{t-1}^{(1)}, \dots, \pi_{t-1}^{(N)}\}$ とする。このとき次の時刻 t におけるサンプル集合を以下の手順で生成する。

まず、時刻 $t-1$ におけるサンプル $s_{t-1}^{(j)}$ を尤度 $\pi_{t-1}^{(j)}$ の比に従ってランダムに選択し、これを $s_t^{(n)}$ と置く。次に、推移確率 $p(\mathbf{X}_t|\mathbf{X}_{t-1} = s_t^{(n)})$ に従って $s_t^{(n)}$ をランダムに変化させ、時刻 t におけるサンプル $s_t^{(n)}$ とする。これを N 回くりかえすことによって、 N 個のサンプル集合を生成する。さらに次の時刻 $t+1$ におけるサンプルは、時刻 t のサンプル集合 $s_t^{(n)}$ と尤度 $\pi_t^{(n)} = p(\mathbf{Z}_t|\mathbf{X}_t = s_t^{(n)})$ から決定する。なお、初期状態 \mathbf{X}_0 における事前確率 $p(\mathbf{X}_0)$ はあらかじめ与えられているものとする。

ある時刻 t において単一の解を求めるときは、次式のように尤度による重み付き平均を求める。

$$\mathcal{E}(\mathbf{X}_t) = \frac{\sum_{n=1}^N \pi_t^n s_t^{(n)}}{\sum_{n=1}^N \pi_t^n} \quad (3)$$

3.2 楕円モデルを用いた人物頭部追跡

本研究では、画像上で観測される人物頭部形状を楕円によって近似する。追跡対象人物の状態 \mathbf{X}^h は楕円の 5 個のパラメータによって次式のように表現される。

$$\mathbf{X}^h = (x^h, y^h, s^h, r^h, \theta^h) \quad (4)$$

ただし、 x^h, y^h は楕円の中心座標、 s^h はスケール、 r^h は偏平率、 θ^h は傾きとする。また、時刻 $t-1$ から時刻 t までの推移確率 $p(\mathbf{X}_t^h|\mathbf{X}_{t-1}^h)$ は、各パラメータが独立な正規分布に従うものとして次式のように与える。

$$\begin{aligned} p(\mathbf{X}_t^h|\mathbf{X}_{t-1}^h) &= p(x_t^h|x_{t-1}^h)p(y_t^h|y_{t-1}^h)p(s_t^h|s_{t-1}^h) \\ &\quad p(r_t^h|r_{t-1}^h)p(\theta_t^h|\theta_{t-1}^h) \\ &= N(x_{t-1}^h, \sigma_x)N(y_{t-1}^h, \sigma_y)N(s_{t-1}^h, \sigma_s) \\ &\quad N(r_{t-1}^h, \sigma_r)N(\theta_{t-1}^h, \sigma_\theta) \end{aligned} \quad (5)$$

ここで、各サンプルの尤度 $\pi_t^{(n)}$ は周辺エッジを用いて以下のように推定する [3]。状態 $s_t^{(n)}$ 楕円の円周上のある点を P_c としたとき、 P_c を通る法線にそって一次元のエッジ検出を行う。得られた E 個のエッジ点と P_c との距離を $\epsilon_1^c, \dots, \epsilon_E^c$ とする。このとき、 $\pi_t^{(n)}$ を次のように定義する。

$$\pi_t^{(n)} = \prod_c \left(\alpha + \frac{1}{\sqrt{2\pi\sigma}} \sum_e \exp\left(-\frac{\epsilon_e^c{}^2}{2\sigma^2}\right) \right) \quad (6)$$

ただし、 α は人物頭部によって生じる真のエッジが存在しない時に、背景やノイズによるエッジを観測する確率を表わし、 σ はエッジと楕円の円周とのずれの分散を表している。

3.3 色情報を用いた高速化、安定化

人物頭部の色モデルがあらかじめ与えられている場合は、ICondensation[4] に類似した手法によって、サンプリング範囲を限定し、追跡の高速化、安定化を行う。

時刻 $t-1$ におけるサンプルを $s_{t-1}^{s,(n)}$, 前節で用いたエッジに基づく尤度を $\pi_{t-1}^{s,(n)}$ とする. このとき, 以下の手順によって色モデルを用いた追跡を行う.

1. サンプル $s_{t-1}^{s,(n)}$ と尤度 $\pi_{t-1}^{s,(n)}$ に従って N_c 個のサンプルを生成する. これを $s_t^{c,(n)}$ とおく.
2. サンプル $s_t^{c,(n)}$ に対する色モデルに基づく尤度 $\pi_t^{c,(n)}$ を求める.
3. サンプル $s_t^{c,(n)}$ と尤度 $\pi_t^{c,(n)}$ に基づいて N_s 個のサンプルを生成する. これを $s_t^{s,(n)}$ とおく.
4. サンプル $s_t^{s,(n)}$ に対してエッジに基づく尤度 $\pi_t^{s,(n)}$ を求める.
5. $t = t+1$ として 1.へ

人物頭部の色モデルは YUV 空間の色ヒストグラムによって表す. あらかじめ多人数の人物頭部画像から標準色ヒストグラムを学習する. ここで, 標準色ヒストグラムの i 番目のセルを $H(i)$ と表す. あるサンプル $s_t^{c,(n)}$ が与えられたとき, 画像上での楕円内部の点をランダムに K 個選択し, $s_t^{c,(n)}$ の色ヒストグラム $I_t^s(i)$ を求める. このとき $s_t^{c,(n)}$ の色に基づく尤度 $\pi_t^{c,(n)}$ を次式のように求める.

$$\pi_t^{c,(n)} = \frac{\sum_{i=1}^K \min(H(i), I_t^s(i))}{\sum_{i=1}^K I_t^s(i)} \quad (7)$$

このように, 色の異なる部分のサンプルを除去することにより, エッジの評価を行うサンプル数 N_s を少なくすることができる. これにより, エッジを評価するのに比べて色の比較を行うほうが計算量が少いので, 高速化が実現される. また, 余分なエッジの多い複雑な背景やノイズに対して安定になる.

4 VizWear-Active による顔画像追跡

VizWear-Active システムは, ウェアラブルクライアントに限られたリソースしか持つことができない, カメラが動くため安定な入力画像を得ることができないなど, ウェアラブルシステム特有の欠点を持つ. そのため, ConDensation アルゴリズムをそのまま適用して, 安定かつ実時間の追跡を実現することは困難である. 本節では, これらの欠点を考慮して ConDensation アルゴリズムを改良し, *VizWear-Active* システムに適應する方法について述べる.

4.1 クライアント/サーバによる分散サンプリング

アクティブカメラの制御を伴う処理を行うためには, 遅延の少ない高速なフィードバックが不可欠である. ここでは, このようなカメラ制御を伴う動作を反射運動と呼ぶ. 一方で, 追跡対象人物から顔画像を抽出し認識, 登録処理を行うためには, 高精度な結果が必要である.

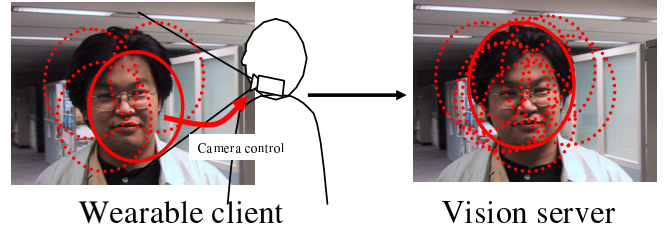


図 4: 分散サンプリング

ConDensation アルゴリズムで安定かつ高精度な追跡を実現するためには, ランダムサンプリングの回数を増す必要があるが, 計算コストが高くなるためウェアラブルクライアントのみで実時間処理を行うのは容易ではない. また, 入力画像をそのままビジョンサーバへ送り, サーバ側で追跡処理を行う場合には, 無線 LAN による遅延が生じるため, 反射運動に必要な反応速度を得ることが困難となる. そこで, ウェアラブルクライアントとビジョンサーバに分散してサンプリング, 評価を行うことで高速かつ高精度な追跡を実現する.

具体的には, 図 4 に示すように, ウェアラブルクライアントでは追跡を継続するために必要最小限のサンプリングのみを行い, 同時にビジョンサーバによって追加的にサンプリングを行うことで高精度な追跡結果を得る. ウェアラブルクライアントによるサンプリングの回数を N , 時刻 t における状態サンプル集合と尤度を $\{s_t^1, \dots, s_t^N\}$, $\{\pi_t^1, \dots, \pi_t^N\}$ とする. ビジョンサーバでは, これらの結果にからさらに M 回のサンプリングを行う. このときのサンプル集合を $\{s_t^{N+1}, \dots, s_t^{N+M}\}$, $\{\pi_t^{N+1}, \dots, \pi_t^{N+M}\}$ とする. このとき, ビジョンサーバによる最終的な追跡結果は次式によって求められる.

$$\mathcal{E}(\mathbf{X}_t) = \frac{\sum_{n=1}^{N+M} \pi_t^n s_t^{(n)}}{\sum_{n=1}^{N+M} \pi_t^n} \quad (8)$$

ウェアラブルクライアントは, 各時刻において推定される対象人物位置が画像の中心になるようにカメラを制御することで, カメラを対象人物に追従させながら追跡する. このとき, 遅延の少ないフィードバックを実現するためにクライアントの結果のみを用いてカメラ制御を行う.

4.2 カメラ動作の補正

カメラ制御や着用者自身の動作によりカメラの姿勢が変化すると, 追跡対象のみかけの変化が大きくなり, 画像上での状態変化を一定の推移確率 $p(\mathbf{X}_t^h | \mathbf{X}_{t-1}^h)$ だけで表現するのは困難となる. そこで, カメラの制御パラメータとモーションセンサによって得られるカメラの角度変化に基づいて推移確率を補正する.

時刻 t におけるカメラパラメータを θ_t^c, ϕ_t^c , モーションセンサから得られるカメラの姿勢を $\theta_t^m, \phi_t^m, \psi_t^m$ とする. このとき時刻 $t-1$ から時刻 t までのカメラの角度変化 $\dot{\theta}_t, \dot{\phi}_t, \dot{\psi}_t$ は次式によって与えられる.

$$\begin{aligned} \dot{\theta}_t &= (\theta_t^c - \theta_{t-1}^c) + (\theta_t^m - \theta_{t-1}^m) \\ \dot{\phi}_t &= (\phi_t^c - \phi_{t-1}^c) + (\phi_t^m - \phi_{t-1}^m) \\ \dot{\psi}_t &= \psi_t^m - \psi_{t-1}^m \end{aligned} \quad (9)$$

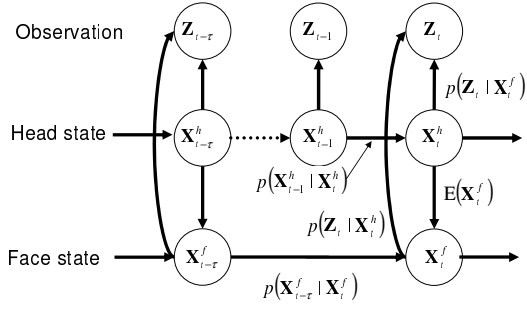


図 5: 顔画像追跡の状態遷移

これらの角度変化を用いて楕円の中心位置，及び傾きを補正する．他のパラメータ，つまりスケール及び偏平率に関しては，カメラの角度変化に対する変化が微量であるのでここでは無視する．時刻 t におけるあるサンプルの中心位置を (x_t^h, y_t^h) ，傾きを θ_t^h とすると，カメラの角度変化を考慮して補正されたサンプル位置 $(x_t'^h, y_t'^h)$ と傾き $\theta_t'^h$ は次式によって求められる．

$$\begin{aligned} x_t'^h &= f \tan(\arctan(\frac{x_t^h}{f}) + \dot{\theta}_t) \\ y_t'^h &= f \tan(\arctan(\frac{y_t^h}{f}) + \dot{\theta}_t) \\ \theta_t'^h &= \theta_t^h + \dot{\psi}_t \end{aligned} \quad (10)$$

ただし， f は焦点距離とする．この補正されたパラメータを用いて次の時刻のサンプルを生成することによって，カメラ動作に対して安定な追跡を実現する．

4.3 顔画像追跡

本研究では，顔画像を獲得することが目的である．そこで，人物頭部追跡によって得られた結果から，さらに詳細な追跡を行うことで，人物頭部領域から顔画像を検出，追跡する．画像上での人物頭部領域を表わす楕円に対する顔領域の位置や形状は，カメラに対する顔の向き，及び頭部に対する顔の位置，大きさによって変化する．そこで，人物頭部領域に対する顔画像の相対的な姿勢を図 5 に示すような人物頭部追跡と非同期な ConDensation アルゴリズムによって推定する．

ここで，顔領域を長方形とし，顔の姿勢の状態 \mathbf{X}_t^f を次式のようなパラメータで表現する．

$$\mathbf{X}_t^f = (\phi_t^f, s_t^f, x_t^f, y_t^f, \theta_t^f,) \quad (11)$$

ただし， ϕ_t^f はカメラに対する顔の向き， s_t^f は楕円のスケールに対する顔画像のスケールを表している．また， x_t^f, y_t^f, θ_t^f は，頭部追跡の誤差を修正するためのもので， x_t^f, y_t^f は中心位置に対する顔画像の位置， θ_t^f は楕円の傾きに対する顔画像の傾きを表している．

人物頭部追跡の結果を $\mathcal{E}(\mathbf{X}_t^h) = (\bar{x}_t^h, \bar{y}_t^h, \bar{s}_t^h, \bar{r}_t^h, \bar{\theta}_t^h)$ としたとき，顔画像の幅 W_t^f ，高さ H_t^f ，中心位置 X_t^f, Y_t^f ，

傾き Θ_t^f は次式によって求められる．

$$\begin{aligned} W_t^f &= \bar{s}_t^h s_t^f \\ H_t^f &= \bar{s}_t^h \bar{r}_t^h \cos \phi_t^f \\ \Theta_t^f &= \bar{\theta}_t^h + \theta_t^f \\ X_t^f &= \bar{x}_t^h + \bar{s}_t^h (x_t^f \cos \Theta_t^f - y_t^f \sin \Theta_t^f) \\ Y_t^f &= \bar{y}_t^h + \bar{s}_t^h (x_t^f \sin \Theta_t^f + y_t^f \cos \Theta_t^f) \\ x_t^f &= x_t^f + \sin \phi_t^f \end{aligned} \quad (12)$$

また， \mathbf{X}_t^f の推移確率 $p(\mathbf{X}_t^f | \mathbf{X}_{t-1}^f)$ は，人物頭部と同様に正規分布によって定義する．

ここで，顔画像の各サンプルの尤度は形状，明るさ，コントラストの正規化を行った Eigenface 法 [7, 8, 9] によって推定する．あらかじめ多人数の顔画像から学習した固有空間を用意しておく．そして，各サンプルの顔位置から顔画像を切り出し，その形状，明るさ，コントラストの正規化を行った後，固有空間との距離を求める．サンプル $s_t^{f,(n)}$ と固有空間との距離を $\text{DFFS}(s_t^{f,(n)})$ としたとき，このサンプルに対する尤度 $\pi_t^{f,(n)}$ を次式によって与える．

$$\pi_t^{f,(n)} = \frac{1}{\sqrt{2\pi}\sigma_f} \exp\left(-\frac{\text{DFFS}(s_t^{f,(n)})}{2\sigma_f^2}\right) \quad (13)$$

ただし， σ_f は DFFS の分散とする．最後に，サンプル $s_t^{f,(n)}$ と尤度 $\pi_t^{f,(n)}$ から式(3)と同様に，重み付き平均によって顔領域を求める．

5 実験結果

本稿で提案した追跡手法を，2 節で紹介した *VizWear-Active* の試作システムに実装し，実験を行った結果を示す．図 6 に人物頭部追跡の結果を示す．破線の楕円がウェアラブルクライアントによる追跡結果，実線の楕円がビジョンサーバで追加的なサンプリングを行った結果を示す．なお実験ではウェアラブルクライアント上で 200 回，ビジョンサーバ上で 1000 回のサンプリングを行った．

ウェアラブルクライアントによる 1 フレームの処理時間は約 0.15 秒であり，着用者が対象人物とすれ違うように歩きながら処理を行っても見失うことなく追跡を行うことができた．また，(1), (3), (4) では，ウェアラブルクライアントの結果ではスケールや位置がずれているが，ビジョンサーバによって修正されていることがわかる．

次に，顔画像の追跡を行った結果を図 7 に示す．画像中の四角が，顔画像追跡結果である．なお図 7 の (1) から (4) は，図 6 の (1) から (4) と同じフレームに対する結果である．(1) のように追跡を開始したときや，動きの激しい場合に失敗があったものの，動きが少なくなると比較的安定に顔画像の追跡が行われた．

これらの結果より，実環境下において，実時間で比較的安定に顔画像の追跡が実現できていることがわかる．しかし，現在の精度ではそのまま顔認識等に用いることは困難であり，また動きが激しいときなどに失敗することもあるため，今後，顔部品抽出などのさらに高精度な後処理を行い，これらの誤差や失敗を検出，修正する必要がある．



図 6: 人物追跡結果

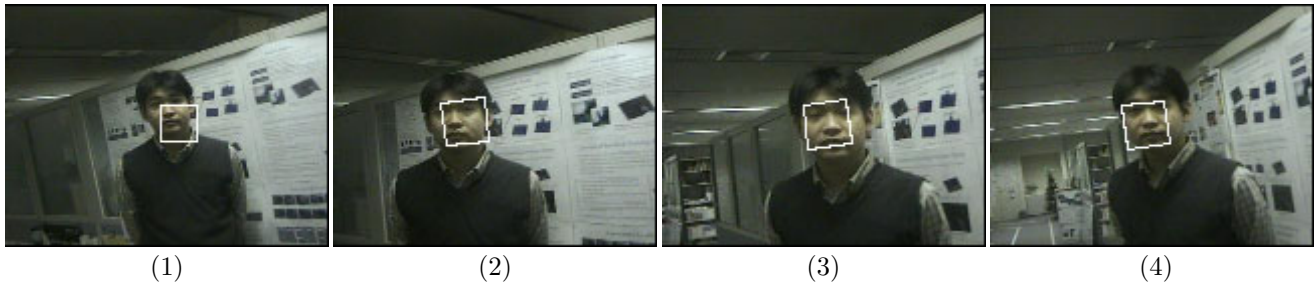


図 7: 顔画像追跡結果

6 むすび

本稿では、ウェアラブルアクティブカメラを備えた着用型アシスタントである *VizWear-Active* のコンセプトと試作システムを紹介した。また、ConDensation アルゴリズムに基づく追跡手法を *VizWear-Active* に適用し、実時間で顔画像追跡を行う手法を示した。

ウェアラブルシステムの限られたリソースで安定かつ実時間の追跡を実現するために、クライアント/サーバで分散してランダムサンプリングを行う方法を提案した。また、着用者の動きやカメラ制御によって追跡が不安定になる問題を解決するために、モーションセンサの情報やカメラパラメータに応じて追跡処理を補正する方法を示した。実際に追跡アルゴリズムを *VizWear-Active* の試作システムに実装して実験を行い、本手法の有効性を示した。

本稿では顔画像の追跡に焦点を当ててきたが、本手法を用いた記憶補助アプリケーションの実現のためには、得られた顔画像の誤差や失敗の修正を含む登録処理、及びそれを用いた認識処理について検討する必要がある。また、現在の試作システムで用いているアクティブカメラは、常に着用するには大きすぎる上に、パン・チルト範囲が狭く、ウェアラブルシステムに適していない。現在、小型で可動範囲の広いウェアラブルアクティブカメラを用いた次期試作システムを開発中である。

謝辞

本研究は、文科省科学技術振興調整費の支援による。記して謝意を表す。

参考文献

- [1] <http://unit.aist.go.jp/is/hcv/vizwear/>.
- [2] Jonny Farrington and Vanessa Oni. Visual augmented memory. In *4th International Symposium on Wearable Computers (ISWC2000)*, pp. 167–168, 2000.
- [3] Micheal Isard and Andrew Blake. Condensation – conditional density propagation for visual tracking. *International Journal of Computer Vision*, Vol. 29, No. 1, pp. 5–28, 1998.
- [4] Micheal Isard and Andrew Blake. Icondensation: Unifying low-level and high-level tracking in a stochastic framework. In *5th European Conference Computer Vision (ECCV1998)*, pp. 893–908, 1998.
- [5] Takeshi Kurata, Takashi Okuma, Masakatsu Kourogi, Takekazu Kato, and Katsuhiko Sakaue. (vizwear): Toward human-centered interaction through wearable vision and visualization. In *PCM2001*, pp. 40–47, 2001.
- [6] W.W. Mayol, B. Tordoff, and D.W. Murray. Wearable visual robots. In *4th International Symposium on Wearable Computers (ISWC2000)*, pp. 95–102, 2000.
- [7] Takeshi Shakunaga, Keisuke Ogawa, and Shouhei Oki. Integration of eigentemplate and structure matching for automatic facial feature detection. In *The Third International Conference on Automatic Face and Gesture Recognition (FG'98)*, pp. 94–99, Nara, Japan, April 1998.
- [8] Y. Sugiyama and Y. Ariki. Facial region tracking and recognition by subspace method. In *VSSM'96*, pp. 225–230, September 1996.
- [9] M. Turk and A. Pentland. Eigenfaces for recognition. *J. Cognitive Neuroscience*, Vol. vol.3, No. no.1, pp. 71–86, January 1991.
- [10] 蔵田武志, 大隈隆史, 興梠正克, 加藤丈和, 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.