2008－CVIM－162 (77)
2008／3／11

3 次元位置仮説に基づく顔の追跡と認識

内海 ゆづ子†, 岩井 儀雄†, 谷内田 正彦†
†大阪大学大学院基礎工学研究科システム創成専攻
〒560–8531 大阪府豊中市待兼山町 1–3
yuzuko@yachi-lab.sys.es.osaka-u.ac.jp, iwai@sys.es.osaka-u.ac.jp,
yachida@sys.es.osaka-u.ac.jp

現在，顔画像の追跡，認識に頻繁に用いられている手法の 1 つが，ベイズ理論を用いた手法である．本研究では，ベイズ理論の枠組みを用いて，動画像に映る顔画像の追跡と認識を同時に行う手法を提案する．提案手法では，観測空間を 3 次元空間とし，顔の 3 次元形状や，並進運動のモデル化を行った．追跡に用いる動画像は，追跡が容易に行える HyperOmni Vision を用いて撮像した．並進運動のモデルから，3 次元位置仮説を生成し，観測空間から画像平面上に顔の形状モデルを投影して得た特徴点から，特徴抽出を行った．得られた特徴量から，事後確率を計算し，顔，非顔の判別を行う．顔と判断された複数枚の画像から，追跡と同じ特徴量を用いて個人認証を行う．

Face Tracking and Recognition Based on 3D Positional Hypothesis

Yuzuko UTSUMI†, Yoshio IWAI†, Masahiko YACHIDA†
†Graduate School of Engineering Science, Osaka University
1–3 Machikaneyama, Toyonaka, Osaka 560–8531 Japan
yuzuko@yachi-lab.sys.es.osaka-u.ac.jp, iwai@sys.es.osaka-u.ac.jp,
yachida@sys.es.osaka-u.ac.jp

We propose face tracking and recognition method based on Bayesian framework from moving images. We assume that an observed space is three dimensional space, and model the facial three dimensional shape, the facial rotation and the facial translation. 3D positional hypothesis are generated by using the facial translation model. A Feature extraction of hypothesis is performed around feature points which are projected from the observed space to the observed image. The observed images are taken by HyperOmni Vision because HyperOmni Vision can track moving object easier than other cameras. The discrimination of face is based on the likelihood that are estimated from the features. The face recognition is performed by using tracked face images.

1 はじめに

顔画像の追跡，認識に関して，これまでに数多くの手法が提案されてきた．顔の追跡や認識の手法の 1 つで，現在頻繁に用いられているものに，ベイズ理論等の統計モデルや確率モデルを用いた手法がある．

ベイズ理論を用いた顔画像追跡には，学習データから顔，非顔のクラスを正規分布でモデリングし，静止画において顔を追跡するベイズ識別器を生成する手法 [3] や，学習用の静止画像から特徴点を自動的に生成し，生成した特徴点をもとに動画像から尤度推定を行う手法 [4] などが挙げられる．これらの手法は，正面を向いた顔画像を学習データとして用いており，テストデータの顔画像が正面を向いている場合や，画像平面で顔の回転が起こった場合に，顔検出が可能である．一方で，人が横を向いて顔半分が完全に隠れてしまった場合には，顔検出が不可能となる．そのため，正面を向いた顔画像が得られる可

能性が低い場合，これらの手法は有効ではなくなる．

静止画における顔画像認識には，主成分分析(PCA)によって形成された部分空間を用いて尤度を表現したもの[6]や，PCA，線形判別分析(LDA)にベイズ推定を統合した手法など[9]が存在する．部分空間法は，顔認識の代表的な手法の一つで，Turkら[8]が用いて以来，様々な顔認識手法に応用されているが，部分空間法での顔認識は，個人のカテゴリごとに部分空間を生成せねばならず，全体として莫大な学習データを必要とする．また，横顔などを認識するためには，さらに横向きの顔画像データが必要となることや，学習データが増えてくると計算コストが増加することなどが問題である．

動画像における顔認識には，ベイズ推定に基づいて，事後確率から最も適した顔モデルを選択して認識を行う手法[5]がある．表情の異なった複数の画像を用いて，画像上の2次元平面で変形が可能な顔のモデルを作っているため，表情変化による認識率の低下を避ける事に成功している．しかしながら，3次元の顔向き変化をモデル化しておらず，横向きの顔に対しては認識ができない．

このほかに，顔の向きの遷移確率を動画像の学習データから計算し，顔の向きにロバストな認識を行う手法[2]が提案されている．この手法では，従来からの問題であった横を向いた顔に対する認識が可能である．顔画像のシーケンスを得られるようになれば，この手法は動画像の顔認識に対して，非常に有効である．だが，認識を行うカテゴリ毎に動画像が必要となり，動画像を参照データとする事は，静止画像を参照データとする事と比較して参照データの用意や学習が困難である．

そこで，本研究では，各向き毎の参照画像を用いて，動画像の中から顔を追跡しながら同時に認識を行う手法を提案する．本研究では，各向き毎の参照画像を持っているので，人物が横を向いていても追跡，認識をすることが出来る．

2 3次元位置仮説に基づく顔の追跡と認識

時刻tにおいて，観測対象である人物iの顔ω_iの3次元空間位置$\boldsymbol{x}_{i,t}$と向き$\phi_{i,t}$を用いて，位置仮説を$H_{i,t} = (\boldsymbol{x}_{i,t}, \phi_{i,t}, \boldsymbol{x}_{i,t-1}, \phi_{i,t-1}, \cdots, \boldsymbol{x}_{i,t-k}, \phi_{i,t-k})$とし，観測された画像から抽出された画像特徴量を$\boldsymbol{Y}_{i,t}$とすると，観測画像特徴量の列$\{\boldsymbol{Y}_{i,0}, \boldsymbol{Y}_{i,1}, \cdots, \boldsymbol{Y}_{i,T}\}$と顔の同時確率密度$p(\omega_i, \boldsymbol{Y}_{i,0}, \boldsymbol{Y}_{i,1}, \cdots, \boldsymbol{Y}_{i,T})$を最大にする顔$\omega_k$を求めることで，顔認識を行なうことが出来る．すなわち，

$$\omega_k = \mathrm{argmax}_{\omega_i} p(\omega_i, \boldsymbol{Y}_{i,0}, \boldsymbol{Y}_{i,1}, \cdots, \boldsymbol{Y}_{i,T}) \quad (1)$$

となるω_kを求めることで，顔認識を行なう．この同時確率密度を計算するのは計算コストの面から非常に困難なので，本研究では，各観測が独立に起きると仮定する．そうすると，同時確率密度は，

$$p(\omega_i, \boldsymbol{Y}_{i,0}, \boldsymbol{Y}_{i,1}, \cdots, \boldsymbol{Y}_{i,T}) = \prod_{t=0}^{T} p(\omega_i, \boldsymbol{Y}_{i,t}) \quad (2)$$

と簡単化できる．こうすることで，各時刻tにおいて$p(\omega_i, \boldsymbol{Y}_{i,t})$を最大化すればよくなる．さらに，この確率を最大化することで対象の追跡も行なうことができる．また，

$$p(\omega_i, \boldsymbol{Y}_{i,t}) = \int_H p(\omega_i, \boldsymbol{Y}_{i,t}, H_{i,t}) dH \quad (3)$$

である．この同時確率密度$p(\omega_i, \boldsymbol{Y}_{i,t}, H_{i,t})$は，図1に示すような確率構造を成していると仮定すると，

$$p(\omega_i, \boldsymbol{Y}_{i,t}, H_{i,t}) = p(\boldsymbol{Y}_{i,t}|H_{i,t}) p(H_{i,t}|\omega_i) P(\omega_i) \quad (4)$$

で与えられる．以上より，

$$\begin{aligned} \omega_k &= \mathrm{argmax}_{\omega_i} \prod_{t=0}^{T} \int_H p(\boldsymbol{Y}_{i,t}|H_{i,t}) p(H_{i,t}|\omega_i) P(\omega_i) dH \\ &\approx \mathrm{argmax}_{\omega_i} \prod_{t=0}^{T} \sum_H p(\boldsymbol{Y}_{i,t}|H_{i,t}) p(H_{i,t}|\omega_i) P(\omega_i) dH \end{aligned} \quad (5)$$

となる．ここで，全仮説Hにによる積分は非常に困難なので，サンプリングした仮説Hでモンテカルロ近似することで計算する．以降，各確率の計算について説明する．

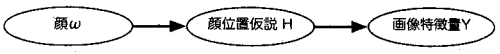


図 1: 観測の確率構造

2.1 個人観測確率 $P(\omega_i)$

確率 $P(\omega_i)$ は，個人 i がシステムの観測範囲内に存在する確率を表す．$P(\omega_i)$ は．時刻 t に依存する関数でも良いし，システムの設置されている環境によって様々に設定することが出来るが，本研究ではデータベース内の N 人の人間が均等に現れる可能性があるとして，離散一様分布を採用する．すなわち，

$$P(\omega_i) = \frac{1}{N} \tag{6}$$

とする．

2.2 仮説の出現確率 $p(H_{i,t}|\omega_i)$

確率密度 $p(H_{i,t}|\omega_i)$ は，個人 i の顔 ω_i に対応する仮説 $H_{i,t}$ が出現する確率を表している．モンテカルロ近似を行なう際には，できるだけ $H \sim p(H_{i,t}|\omega_i)$ となるような仮説 H を生成できれば良い．また，この確率により仮説の生成範囲を設定することが出来る．つまり，$p(H_{i,t}|\omega_i)$ や $p(\boldsymbol{Y}_{i,t}|H_{i,t})$ が非常に低いところをサンプルして事後確率を計算しても，計算結果にはほとんど影響しないので，近似値という面では無視することが可能である．効率的な計算のためには，対象となる顔の位置を正確に予測できるほうが望ましいが，複雑な確率モデルを設定すると，計算に時間がかかり，観測のサンプル間隔が長くなってしまうという欠点がある．一方で，サンプル間隔を短くできれば，単純なモデルで追跡が行なえる可能性がある．

そこで，本研究では，各時刻における出現確率密度を以下のように定める．

$$p(H_{i,t}|\omega_i) = \alpha_t p_{AR}(H_{i,t}|\omega_i) + (1-\alpha_t) p_{RW}(H_{i,t}|\omega_i) \tag{7}$$

ここで，α_t は2つの分布の混合率で，p_{AR} は AR モデルに基づく位置予測を表し，p_{RW} はランダムウォークによる予測を表している．なお，p_{AR} は，

$$p_{AR}(H_{i,t}|\omega_i) \sim \mathcal{N}(A_t(H_{i,t}), \Sigma_A) \tag{8}$$

としている．ここで，$A_t(\cdot)$ は AR による3次元位置予測ベクトルである．また，p_{RW} は，

$$P_{RW} \sim \mathcal{N}(\boldsymbol{x}_{i,t-1}, \Sigma_R) \tag{9}$$

である．

2.3 仮説の観測確率 $p(\boldsymbol{Y}_{i,t}|H_{i,t})$

確率密度 $p(\boldsymbol{Y}_{i,t}|H_{i,t})$ は，仮説が観測できる確率を表す．この確率の計算には，人間の顔がどのようにセンサに現れるかを表すモデルが必要である．本研究では，顔を表す画像特徴量 $\boldsymbol{Y}_{i,t}$ を，顔器官（目，鼻，口）の特徴点 $\{p_j\}$ のまわりで特徴抽出したベクトルと考える．顔モデルの具体的実装については，次章で説明する．本研究で利用する仮説 H と，画像特徴量 $\boldsymbol{Y}$ とは次元やスケールが異なり，直接比較することは出来ない．直接比較するためには，仮説 $H_{i,t}$ から画像特徴量 Y への変換 $F : H \rightarrow \boldsymbol{Y}$ が必要である．なお，この変換 F の際に，確率分布形状が変化してしまう可能性があるので，演算結果に影響を及ぼすことになる．本来ならば，そのような影響を受けない画像特徴量や変換方法が望ましいが，未だ明らかではない．そこで，本研究では，顔データベースを利用して，この変換 $F : H \rightarrow \boldsymbol{Y}$ を実装する．

上で述べたような変換 F を用いて，観測確率密度 $p(\boldsymbol{Y}_{i,t}|H_{i,t})$ を以下のように仮定する．

$$p(\boldsymbol{Y}_{i,t}|H_{i,t}) \sim \mathcal{N}(\boldsymbol{Y}_{i,t} - F(H_{i,t}), \Sigma_f) \tag{10}$$

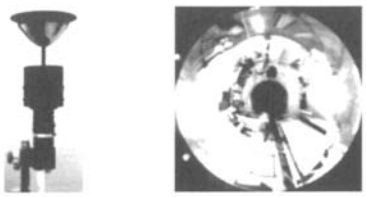
ここで，Σ_f は画像特徴量の共分散行列である．

3 顔の追跡と認識システム

2章で示した追跡と認識のアルゴリズムを検証するため，システムを構築した．本章では，実装した際のカメラのキャリブレーション方法，実際の顔モデル，特徴抽出について説明する．

3.1 カメラシステム

本研究では，全方位視覚センサ HyperOmni Vision[12] を利用して人物の顔追跡を行なう．全方位視覚センサは，センサの周囲 360° の画像を一度に撮像できるため，撮像範囲が広く長期間にわたり対象を撮像できる利点がある．本研究で用いた HyperOmni Vision を図 2(a) に示す．また，撮像される画像の 1 例を図 2(b) に示す．3 次元位置仮説を立て，形状モデルを画像面に投影して特徴抽出を行う場合，カメラ画像と観測空間の 3 次元位置座標の対応が出来なければならないため，予め HyperOmni Vision



(a)HyperOmni Vision　(b) 撮像された画像

図 2: HyperOmni Vision の外観と撮像される画像

のキャリブレーションを行う．カメラのキャリブレーションには，Tsai ら [7] の手法を用いて行い，図 3 に示すミラー座標系から観測空間への関係は，廣田ら [1] の手法を用いて推定する．

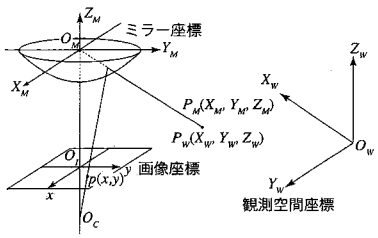
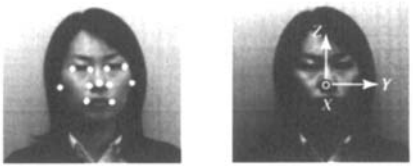


図 3: 画像，ミラー，観測座標系

観測確率を計算する際に，顔とカメラの位置関係によっては後頭部を撮像することとなり，位置の検証には使えるが顔認識には利用できない場合が発生する．そこで，本研究では，HyperOmni Vision を複数台用いることによって，顔画像がいずれかの HyperOmni Vision に撮像される可能性を高めることで，この問題を回避する．また，カメラの複数台利用は，より広範囲の画像を得る事ができ，追跡にも有利となる．

3.2　顔モデル

顔モデルの特徴点は，顔と個人の特徴を表さなければならないことを考慮して，特徴点を手動で選択した．選択した特徴点を図 4(a) に示す．これらの特徴点は，4(a) で示された鼻の中央を原点に取り，図 4(b) で示された顔の形状モデル座標により表現される．また，顔の中心を形状モデル座標の中心である鼻とし，3 次元位置仮説は，鼻の中央の位置を表す．これらの特徴点を位置仮説や回転に従って，HyperOmni Vision で撮像された画像面に投影し，特徴抽出を行う．



(a) 顔モデルの特徴点　(b) 顔の座標系

図 4: 顔モデル

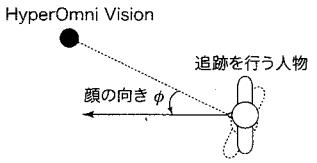


図 5: 顔の向きの定義

3.3　運動モデル

ランダムウォークと AR モデルの混合率を変化させることにより，運動モデルを変化させることが出来る．本手法では，ランダムウォークと AR モデルの混合率を 1 : 1 とした．ランダムウォークは，時刻 t での位置 x_t を，時刻 t_1 での位置 x_{t-1} からの正規分布する乱数によって決定する．

$$\boldsymbol{x}_t = \boldsymbol{x}_{t-1} + \boldsymbol{\epsilon}_t \tag{11}$$

また，AR モデルの次元数を 2 次元とする．

$$\boldsymbol{x}_t = \alpha_1 \boldsymbol{x}_{t-1} + \alpha_2 \boldsymbol{x}_{t-2} + \boldsymbol{\epsilon}_t \tag{12}$$

AR 定数 α_1, α_2 は，過去 20 時刻前までのデータを用いて，予測誤差を最小にするものを最小二乗法を用いて求める．そして，観測誤差を正規分布の乱数で加えることにより，AR モデルで求めた位置仮説とする．

この顔追跡システムでは，顔の回転は，顔の形状モデルの Z 軸の回転のみを考慮する．図 5 のように，X 軸が HyperOmni Vision の方向を向く状態を 0° とし，反時計回りを正とする回転角 ϕ を顔の向きと定義する．

(a) −90° (b) −60° (c) −30° (d) 0° (e) 30° (f) 60° (g) 90°

図 6: 顔の向きと特徴量

全ての顔の向きをデータベースに登録する事は不可能であるため，いくつかの顔の向きを代表とし，特徴を比較した際に，最も近いものを顔の向きとする．顔の向きの代表を，$-90°, -60°, -30°, 0°, 30°, 60°, 90°$ とする．

3.4 特徴抽出

位置仮説と回転から，顔の位置と特徴点の位置が求められたら，特徴点のまわりで特徴抽出を行う．ただし，顔は 3 次元の物体であり，顔の向きによっては，オクルージョンが発生する．よって，カメラと顔の相対的な回転角度ごとに，特徴抽出する点を変化させる．顔の方向ごとの特徴点の様子を図 6 に示す．ただし，図 6(d) のように，顔の向きが 0° の場合，顔の対称性が追跡にとって有利な情報となる．そのため，顔の向きが 0° の場合，鼻を挟んで線対称な特徴点のまわりのウェーブレット変換を行う部分の差分を取ってから特徴抽出を行う．特徴抽出は，Haar ウェーブレットを用いた 2 次元ウェーブレット変換により行う．また，1 つの方向だけでなく，ウェーブレット変換の方向を 8 方向 $(0°, 22.5°, 45°, 67.5°, 90°, 112.5°, 135°, 157.5°, 180°)$ に回転させ，特徴抽出を行う．

なお，実際には HyperOmni Vision に投影させた画像は歪んでいるため，歪みを補正した画像に変換してから，特徴抽出を行う．

4 実験

提案手法の有効性を確かめるために，評価実験を行った．1 人の人物が歩行する様子を，3 台の HyperOmni Vision を同期を取って撮像した．3 台カメラのうち，2 台を RGB のカラーで，1 台をグレースケールにより撮影した．画像のサイズは XGA で，フレームレートを 15fps とした．追跡を行う最初の画像に顔の初期の位置情報を与え，追跡と認識の実験を行った．

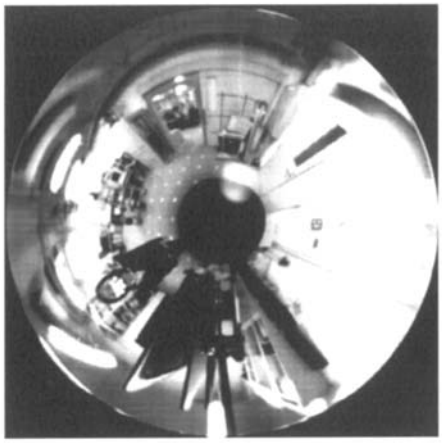
追跡，認識のデータベースは，通常のカメラを用いて撮像した画像によって構築した．カメラの中心に対して，$-90°, -60°, -30°, 0°, 30°, 60°, 90°$ の 7 方向を向いた 7 枚の顔画像を用いた．これらの画像は，顔の方向ごとに 3.4 節で選択した特徴点を登録しておく．特徴抽出を行う際，観測画像を透視投影変換した画像とデータベース画像の顔の大きさが異なる場合がある．このとき，顔の 3 次元形状モデルを定義していることから，透視投影変換画像をデータベース画像の大きさと同じになるようにスケーリングすることが可能である．スケーリングを行った画像は，ぼけが生じ，データベース画像と比較した場合，顔の判別や認識に影響を及ぼす．よって，データベース画像に対して，ガウシアンフィルタによって平滑化処理を行った後に，データベースとの特徴抽出を行う．

今回，データベースとして，1 人の人物を用いた．各位置仮説ごとに，全ての顔の向きに対して，特徴量を抽出し，対応する特徴量と比較を行った．追跡の結果の一部を図 7 に示す．

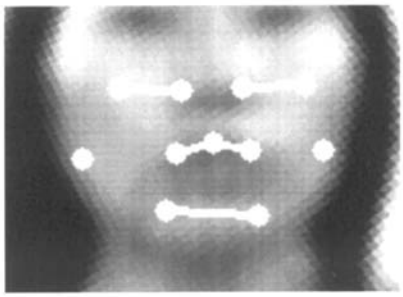
図 7 から，肌色の部分への追跡は可能であるが，顔の特徴量をうまくとらえていない事がわかる，この原因として，確率モデルの尤度設計がうまく出来ていない事が挙げられる．よって，尤度を再設計しなおす必要がある．

5 まとめと今後の課題

本研究では，3 次元位置仮説に基づく顔の追跡と認識を同時に行なう手法と，その実装システムについて提案した．顔形状およびテクスチャを，特徴点の 3 次元位置と Haar ウェーブレット特徴量でモデ



(a) 撮像された画像での追跡結果



(b) 透視投影変換された追跡結果

図 7: 顔追跡結果の一部

ル化した．追跡を行なうための運動モデルとしては，AR モデルとランダムウォークを利用した．これらのモデルを利用して，顔の追跡と認識が同時に行なわれる Detection and Tracking as Recognition を実現した．

今後の課題としては，尤度を構築し直し，精度よく追跡を行う．複数人のデータベースに対して，追跡，を行い，Detection and Trackin as Recognition の実装を行う．また，顔の回転を Z 軸だけでなく，X, Y 軸でも行い，さまざまな動きを扱えるようにモデルを拡張することが挙げられる．また，現在では顔モデル構築に複数枚の顔画像が必要であるが，これを正面顔画像のみで構築することで，データベース構築の負荷を減らすことなどが挙げられる．

参考文献

[1] Hirota, T., Nagahara, H. and Yachida, M.: Calibration of Rotating Line Camera for Spherical Imaging, *Proc. of ACCV2006*, pp. 389–398 (2006).

[2] Lee, K.-C., Ho, J., Yang, M.-H. and Kriegman, D.: Video-based face recognition using probabilistic appearance manifolds, *Proc. of CVPR'03*, Vol. 1, pp. I–313–I–320 (2003).

[3] Liu, C.: A Bayesian discriminating feature method for face detection, *IEEE Trans. on PAMI*, Vol. 25, No. 6, pp. 725–740 (2003).

[4] Loutas, E., Pitas, I. and Nikou, C.: Probabilistic multiple face detection and tracking using entropy measures, *IEEE Trans. on Circuits and System for Video Technology*, Vol. 14, No. 1, pp. 128–135 (2004).

[5] Matsui, A., Clippingdale, S. and Matsumoto, T.: Pruned resampling: probabilistic model selection schemes for sequential face recognition, *IEICE Trans. on Information and Systems*, Vol. E90–D, No. 8, pp. 1151–1159 (2008).

[6] Moghaddam, B.: Principal manifolds and probabilistic subspace for visual recognition, *IEEE Trans. on PAMI*, Vol. 24, No. 6, pp. 780–788 (2002).

[7] Tsai, R. Y.: A Versatile camera calibration technique for high-accuracy 3D machine vision metrology using off the shelf TV cameras and lenses, *IEEE J. of Robotics and Automation*, Vol. 3, No. 4, pp. 323–344 (1987).

[8] Turk, M. and Pentland, A.: Eigenface for recognition, *J. of Cognitive Neuroscience*, Vol. 3, No. 1, pp. 71–86 (1991).

[9] Wang, X. and Tang, X.: A unified framework for subspace face recognition, *IEEE Trans. on PAMI*, Vol. 26, No. 9, pp. 1222–1228 (2004).

[10] Wang, X. and Tang, X.: Random sampling for subspace face recognition, *Int. J. of Computer Vision*, Vol. 70, No. 1, pp. 91–104 (2006).

[11] 内海ゆづ子，岩井儀雄，谷内田正彦：顔認識のためのウェーブレット特徴量の評価，日本知能情報ファジィ学会誌，Vol. 19, No. 5, pp. 476–487 (2007).

[12] 山澤一誠，八木康史，谷内田正彦：移動ロボットのナビゲーションのための全方位視覚系 HyperOmni Vision の提案，信学論，Vol. J79–D–II, No. 5, pp. 698–707 (1996).