

# パーティクルフィルタとカスケード型識別器の統合による人物三次元追跡 ～人物追跡の頑健化・高精度化に向けて～

埼玉大学

東京大学

三菱電機株式会社

国立情報学研究所

小林 貴訓 / 佐藤 洋一・杉村 大輔 / 関 真規人・平澤 宏祐・鈴木 直彦・鹿毛 裕史 / 杉本 晃宏

視野を共有する複数カメラを用いて、人物の三次元追跡を行う。時系列フィルタの一つであるパーティクルフィルタにAdaBoost学習によるカスケード型識別器を統合することで、頑健かつ高精度に人物頭部を追跡する手法を紹介する。

## 1. はじめに

カメラからの入力画像を用いて人物を追跡する技術は、監視カメラの普及などに伴い、セキュリティやマーケティングなどへの応用が期待されている。

カメラ画像を用いた対象の追跡では、これまでに多くの手法が提案されているが、なかでも近年、パーティクルフィルタ[2]の有効性が数多く報告されている([1]-[7], [9], [10] など)。パーティクルフィルタは、状態量と尤度を持つ多数の仮説群により離散的な確率密度として追跡対象を表現し、状態遷移モデルを用いて伝播させることで、動きの変動や観測のノイズに対して頑健な追跡を実現する手法である。

パーティクルフィルタによる人物追跡では、仮説の人物らしさをカメラ画像を用いて評価するが、これにはカラーヒストグラムや輪郭の輝度変化の類似性などが用いられることが多い([4]-[7])。しかし、このような比較的単純な評価手法は、照明変化や複雑な背景下における人物の追跡では必ずしも十分ではなく、追跡性能の向上には頑健かつ高精度な評価手法が求められる。

一方で、静止画像などから人物の顔を検出する手法が数多く提案されている。なかでも、AdaBoost 学習による識別器を用いた顔検出手法が良く知られており、特に、矩形特徴を用いた識別器による顔検出手法は、照明変動や複雑背景下での観察にも頑健であり、実行速度の速さと検出精度から、広く利用されるようになってきている。

AdaBoost 学習による識別器では多数の弱識別器を線形結合することで高

精度な識別器を構成するが、Viola と Jones はこれをカスケード型とし、学習、検出時に用いる矩形特徴を高速に計算する手法[8]を示している。このカスケード型識別器は、非検出対象がカスケードの初期に棄却されるため、単一の識別対象に関して、より高速な処理が可能である。

静止画像からの顔検出では、このカスケード型識別器を画像全体を網羅するように識別対象領域の大きさや位置を変化させて適用するが、時系列画像を用いて対象を追跡する場合には、画面全体を逐次探索することは効率的ではない。そのため、パーティクルフィルタにより予測された領域にカスケード型識別器を適用し、探索範囲を限定することが有効であると考えられる。

このような考えに基づいて、本稿では、視野を共有した複数カメラを用いて、人物頭部の三次元位置と向きを追跡する手法を紹介する。ごく最近では、近い考えに基づいた手法[1]も提案されているが、本稿の内容は、これらと同時期に報告してきたものである。

## 2. パーティクルフィルタ

時刻  $t$  における対象の状態量を  $\mathbf{x}_t$ 、画像による観測を  $\mathbf{z}_t$  とし、時刻  $t$  までに得られる観測を  $\mathbf{Z}_t = \{\mathbf{z}_1, \dots, \mathbf{z}_t\}$  とする。このとき、時刻  $t$  における対象の事前確率  $P(\mathbf{x}_t | \mathbf{Z}_{t-1})$  は、マルコフ過程を仮定することにより、時刻  $t-1$  における事後確率  $P(\mathbf{x}_{t-1} | \mathbf{Z}_{t-1})$  と時刻  $t-1$  から  $t$  への状態遷移確率  $P(\mathbf{x}_t | \mathbf{x}_{t-1})$  を用いて以下のように表される。

$$P(\mathbf{x}_t | \mathbf{Z}_{t-1}) = \int P(\mathbf{x}_t | \mathbf{x}_{t-1}) P(\mathbf{x}_{t-1} | \mathbf{Z}_{t-1}) d\mathbf{x}_{t-1} \quad (1)$$

ここで、 $P(\mathbf{z}_t | \mathbf{Z}_{t-1})$  を一定とすると、時刻  $t$  における事後確率  $P(\mathbf{x}_t | \mathbf{Z}_t)$  は、ベイズの法則より、時刻  $t$  における尤度  $P(\mathbf{z}_t | \mathbf{x}_t)$  と事前確率  $P(\mathbf{x}_t | \mathbf{Z}_{t-1})$  により次のように表すことができる。

$$P(\mathbf{x}_t | \mathbf{Z}_t) \propto P(\mathbf{z}_t | \mathbf{x}_t) P(\mathbf{x}_t | \mathbf{Z}_{t-1}) \quad (2)$$

対象の追跡は、この事後確率  $P(\mathbf{x}_t | \mathbf{Z}_t)$  の期待値を逐次求めることで表現される。

パーティクルフィルタでは、時刻  $t$  における事後確率  $P(\mathbf{x}_t | \mathbf{Z}_t)$  を、状態量  $\mathbf{x}_t$  の仮説群  $\{\mathbf{s}_t^{(1)}, \dots, \mathbf{s}_t^{(N)}\}$  と各仮説に対応する重み  $\{\pi_t^{(1)}, \dots, \pi_t^{(N)}\}$  により離散的に近似し、次のプロセスを経て、逐次的に更新する。

### 1) 仮説の選択

時刻  $t-1$  における事後確率  $P(\mathbf{x}_{t-1} | \mathbf{Z}_{t-1})$  を離散的に近似した  $N$  個の仮説  $\{\mathbf{s}_{t-1}^{(1)}, \dots, \mathbf{s}_{t-1}^{(N)}\}$  の重み  $\{\pi_{t-1}^{(1)}, \dots, \pi_{t-1}^{(N)}\}$  の比に従い、仮説群  $\{\mathbf{s}_{t-1}^{(1)}, \dots, \mathbf{s}_{t-1}^{(N)}\}$  を選択する。

### 2) 状態遷移確率に基づく予測

選択された仮説群  $\{\mathbf{s}_{t-1}^{(1)}, \dots, \mathbf{s}_{t-1}^{(N)}\}$  を、状態遷移確率  $P(\mathbf{x}_t | \mathbf{x}_{t-1} = \mathbf{s}_{t-1}^{(n)})$  に従い伝播し、 $P(\mathbf{x}_t | \mathbf{Z}_{t-1})$  に相当する時刻  $t$  における  $N$  個の仮説群  $\{\mathbf{s}_t^{(1)}, \dots, \mathbf{s}_t^{(N)}\}$  を生成する。

### 3) 画像による重み $\pi_t^{(n)}$ の推定

仮説群  $\{\mathbf{s}_t^{(1)}, \dots, \mathbf{s}_t^{(N)}\}$  の重み  $\pi_t^{(n)} \approx P(\mathbf{z}_t | \mathbf{x}_t = \mathbf{s}_t^{(n)})$  を尤度の評価を行うことで画像から推定する。

### 3. カスケード型識別器

ViolaとJonesにより提案されたカスケード型顔検出器[8]は、検出時間の短縮のため、複数の識別器を組み合わせたカスケード構造(図1(a))をなしている。図1(a)において $\mathbf{H}_i$ は識別器を表す。入力画像に対し、各段で顔、非顔の判定を行い、顔と判定された画像だけが次の段へ進む。最後の段まで通過したものが最終的に顔と判定される。

カスケードの各段を構成する識別器 $\mathbf{H}_i(x)$ は、図1(b)のような矩形特徴を用いて評価を行う多数の弱識別器 $h_i(x)$ の線形結合により、以下のように表される。

$$\mathbf{H}_i(x) = \text{sgn} \left( \sum_{t=1}^T \alpha_t h_t(x) \right) \quad (3)$$

ここで、 $T$ は用いられる弱識別器の数であり、 $\alpha_t$ は学習時に決まる弱識別器のエラー $\varepsilon_t$ を用いて $\alpha_t = \log \frac{1-\varepsilon_t}{\varepsilon_t}$ と表される。

矩形特徴の位置と大きさを画像内でのようにとるかによって膨大な種類の特徴が存在するが、これらの中から顔をよく識別する特徴が学習時にAdaBoostアルゴリズムにより選択され、各段の識別器が準備される。

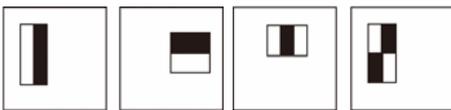
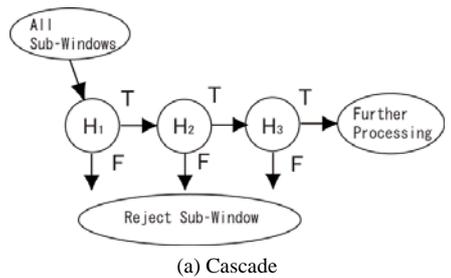


図1. カスケード型識別器

### 4. 提案手法の紹介

視野を共有した複数カメラを用いて、三次元位置と向きを状態量を持つ人

物頭部を、パーティクルフィルタにより追跡する。本手法の主な特徴は以下である。

- 1) カスケード型識別器をパーティクルフィルタの仮説評価に応用する。
- 2) 頭部の各方向毎に学習を行った識別器を、仮説とカメラの関係に基づいて適応的に用いることで、頭部の見えの変化に対応する。

#### 4.1. 人物頭部モデルと仮説のカメラ画像への射影

床面を $XY$ 平面と一致させ、高さ方向を $Z$ 軸とした三次元世界座標系 $XYZ$ をとる。人物頭部はモデルとして中心座標が $(x, y, z)$ である一定の大きさの楕円体を仮定する。人物は頭部を傾けて室内を移動することは少ないと仮定すると、人物頭部の向きは、 $X$ 軸を基準とした $Z$ 軸回りの回転 $\theta$ のみで表せる。仮説はこの四次元の状態量を持つ。

人物頭部の時刻 $t$ における $n$ 番目の仮説 $\mathbf{s}_t^{(n)} = [x_t^{(n)}, y_t^{(n)}, z_t^{(n)}, \theta_t^{(n)}]$ は、校正済みの $i$ 番目のカメラ画像に次のように射影することができる。

$$\mathbf{p}_{i,t}^{(n)} = F_i(\mathbf{s}_t^{(n)}) \quad (4)$$

ここで $\mathbf{p}_{i,t}^{(n)}$ は、仮説 $\mathbf{s}_t^{(n)}$ の位置を $i$ 番目のカメラ画像へ射影したものであり、カメラ画像座標 $[u_{i,t}^{(n)}, v_{i,t}^{(n)}]$ を要素を持つ。

また、各カメラによって観察される相対的な人物頭部の向きは以下のように表される。

$$\theta_{i,t}^{(n)} = \theta_t^{(n)} - \tan^{-1} \left( \frac{[\mathbf{J}\mathbf{c}_i - \mathbf{K}\mathbf{s}_t^{(n)}]^y}{[\mathbf{J}\mathbf{c}_i - \mathbf{K}\mathbf{s}_t^{(n)}]^x} \right) \quad (5)$$

ここで $\theta_{i,t}^{(n)}$ は $i$ 番目のカメラによって観察される相対的な人物頭部の向きである。 $\mathbf{J}$ はカメラ位置 $\mathbf{c}_i$ から $XY$ 位置成分を取り出すための行列であり、 $\mathbf{K}$ は仮説 $\mathbf{s}_t^{(n)}$ から $XY$ 位置成分を取り出すための行列である。 $[\ ]^x$ は計算結果から $X$ 軸に対応する要素を取り出すことを表している。

各カメラで観察される人物頭部の幅( $l_i$ )は、人物頭部の楕円体モデルを射影したものをを用いる。

#### 4.2. カスケード型識別器による人物頭部らしさの評価

カスケードの各段の識別器は、階層が進むにしたがって、より多くの弱識別器 $h_i(x)$ を用いて判定を行う。そのため、より多くの識別器を通過した人物頭部候補領域画像 $g_{i,t}^{(n)}$ は、より多くの人物頭部の特徴を保持していると考えられる。このような知見に基づいて、本手法では人物頭部候補領域画像 $g_{i,t}^{(n)}$ をカスケード型識別器に入力した際に通過した識別器の数(カスケード段数)を人物頭部らしさの評価とする。

具体的には、時刻 $t$ における $n$ 番目の仮説 $\mathbf{s}_t^{(n)}$ を $i$ 番目のカメラに射影した際の重み $\pi_{i,t}^{(n)}$ は、以下の手順により得る。ただし事前に、カスケード型識別器を、正面、 $90^\circ$ 左向き、 $90^\circ$ 右向きなどの人物頭部の向き毎に、人物頭部と非人物頭部で通過する識別器の数(カスケード段数)に十分な差がつくように学習しておく。

1) 各時刻 $t$ において、生成された $n$ 番目の仮説 $\mathbf{s}_t^{(n)}$ を $i$ 番目のカメラ画像に射影し、カメラ画像座標 $\mathbf{p}_{i,t}^{(n)}$ 、相対的な人物頭部の向き $\theta_{i,t}^{(n)}$ 、カメラ画像上の人物頭部の幅 $l_{i,t}^{(n)}$ を得る。

2) 仮説 $\mathbf{s}_t^{(n)}$ を射影したカメラ画像座標 $\mathbf{p}_{i,t}^{(n)}$ を中心に、カメラ画像上での人物頭部の幅 $l_{i,t}^{(n)}$ を一辺とする領域を切り出す。ただし、仮説を射影した際にカメラの視野外となる場合、評価は行わず、重みを一定の小さな値とする。

3) カスケード型識別器は、入力画像サイズが固定( $24 \times 24$ ピクセルなど)であるため、2)で切り出した画像のサイズを変更し、識別器に入力する人物頭部候補領域画像 $g_{i,t}^{(n)}$ を生成する。

4) 仮説の射影によって得られた相対的な人物頭部の向き $\theta_{i,t}^{(n)}$ に基づいて、カスケード型識別器を選択する。例えば、正面、 $90^\circ$ 右向き、 $90^\circ$ 左向きの

3方向の識別器を用いた場合、相対的な人物頭部の向き  $\theta_{i,t}^{(n)}$  が  $-45^\circ \sim 45^\circ$  の場合は人物頭部正面の識別器が選択され、 $45^\circ \sim 135^\circ$  の場合は  $90^\circ$  左向きの識別器が選択され、 $-45^\circ \sim -135^\circ$  の場合は  $90^\circ$  右向きの識別器が選択される。

5) 人物頭部候補領域画像  $g_{i,t}^{(n)}$  を選択されたカスケード型識別器に入力し、人物頭部候補領域画像  $g_{i,t}^{(n)}$  が通過した識別器の数(カスケード段数)を取得する。ここで得たカスケード段数を対応する仮説の重み  $\pi_{i,t}^{(n)}$  とする。例えば、全カスケード段数が40段で、すべての識別器を通過した場合、重みは41となり、5段目で棄却された場合、重みは5となる。

6) 手順1)~5)を各カメラに対して行い、得られた重み  $\pi_{i,t}^{(n)}$  を次式により統合する。

$$\pi_t^{(n)} = \prod_i \pi_{i,t}^{(n)} \quad (6)$$

すべての仮説に対して式(6)による重みを計算し、期待値をとることで各時刻の人物頭部の三次元位置と向きを推定する。

## 5. 実験

本手法の有効性を確認するため、人物頭部の追跡実験を行った。実験は室内天井に設置した視野を共有する校正済みのIEEE1394 カラーカメラ2台(Point Grey Research 社製Flea)を用いて行った。映像は  $640 \times 480$  ピクセル、毎秒30フレームで取得し、1台の汎用PC(Petium4 3.2GHz, Memory 1GByte)で処理した。

カスケード型識別器は、人物頭部の正面、 $90^\circ$  右向き、 $90^\circ$  左向きをそれぞれ検出するように学習を行った3種類を用いた。カスケード段数は40段とし、識別対象画像サイズは  $24 \times 24$  ピクセルとした。また、パーティクルフィルタの仮説数は200とし、状態遷移モデルとして等速直線運動を仮定した。

この条件で、頭部の向きを変えなが

ら観測領域内を移動する人物の追跡実験を行った。

実験で得られた画像の一部を図2に示す。各フレームには、推定結果を矩形でカメラ画像に重ねて表示した。また、各仮説を評価が高いほど輝度が高くなるような点として表示している。図2より、ほぼ正確に人物頭部の中心を推定できていることが分かる。

追跡の精度を定量的に調べるために、画像中の人物頭部位置を手作業で求め、2枚の画像から逆投影して求めた三次元座標を真の位置と見なし、推定結果と比較した。図3に、推定結果と対応する人物頭部の真の位置の  $XY$  平面上、及び  $XZ$  平面上での軌跡を示す。また、 $Z$  軸方向及び  $XY$  平面上での平均誤差及び標準偏差を表1に示す。 $XY$  平面上の平均誤差、 $Z$  軸方向の平均誤差は共に2cm以内であり、高精度な追跡ができています。



図2. 人物頭部の追跡結果

なお、実験に用いたPCでは、200個の仮説を15ms程度で処理できた。2台のカメラに仮説を射影した際の処理は合計30ms程度で完了し、リアルタイムでの追跡が可能であった。

## 6. 考察

パーティクルフィルタを用いた追跡では、各フレームにおいて、仮説の尤度を高精度に推定できることが性能向上に大きく寄与する。本手法の枠組みにおいては、人物頭部の仮説を画像平面上に射影した際に人物頭部周辺で鋭いピークを持つ関数が理想的である。そこで、実際の人物頭部の位置周辺でカメラから一定距離の平面を一边1cmのグリッドで分割し、それぞれの三次元位置での人物頭部らしさの評価を提案手法に基づいて算出した。その結果を図4(a)に示す。

図4(a)より、カスケード段数に基づく

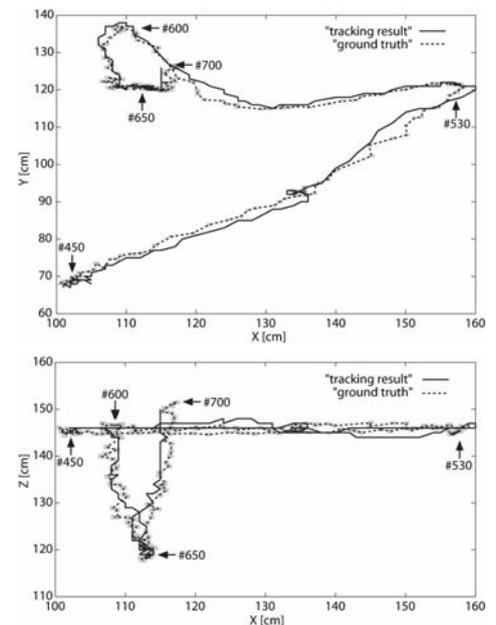


図3. 人物頭部の追跡結果の軌跡

表1. 追跡中の人物頭部の検出誤差

	平均[cm]	標準偏差[cm]
Z 軸方向	1.02	0.65
XY 平面	1.98	1.46

評価が頭部周辺が高く、頭部以外の場所では低くなっていることがわかる。また、図4(b)輪郭の輝度変化の類似性に基づく評価に比べ、頭部周辺での評価が鋭いピークを持っており、パーティクルフィルタでの利用に適している。

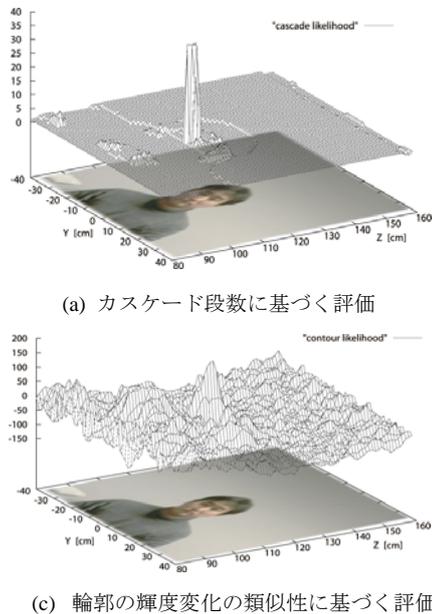


図4. 評価の分布

人物頭部は向きやカメラとの位置関係によりその見えが異なるため、本手法では、複数のカスケード型識別器を仮説とカメラの関係に基づいて選択的に用いている。そこで、相対的な頭部の向きと各方向の識別器の評価の関係性を調べた。図5は頭部の回転に伴い、評価がどのように変化するかを示したものである。

図5より、頭部がカメラから見て左向きの場合には90°左向きの識別器の評価が高くなっており、正面、90°右側の識別器による評価は相対的に低くなって

いる。頭部が正面を向いている場合は、正面の識別器の評価が最も高くなっており、右を向いている場合は90°右向きの識別器の評価が最も高い。このように、正しい向きの識別器の評価が相対的に高くなることから、頭部の向きの仮説に基づいて選択的に識別器を用いても、頭部の向きを正しく推定でき、追跡を継続できる。

## 7. おわりに

本稿では、パーティクルフィルタにおける仮説の評価に、カスケード型識別器を応用し、実時間で人物頭部の追跡を行う手法を紹介した。複数の識別器を仮説に基づいて選択的に用いることで、人物が頭部の向きを変えながら移動しても、高精度に人物頭部を追跡できることを示した。

また、本手法によって、人物頭部の向きを測定することが可能であるが、正面、90°右向き、90°左向きの3方向の識別器を用いた実験では、推定される顔の向きに45°程度のばらつきがあった。これに対しては、識別器の方向を追加することで、より精度良く推定することができるかと予想される。なお、本手法では、パーティクルフィルタの仮説により選択的に識別器が用いられるため、計算コストを増加させずに識別器の種類を追加することが可能である。

今後は、より細かな方向の識別器を追加した場合の追跡精度や、頭部の初期検出手法などについて検討を進める予定である。

## 参考文献

- [1] T. Chateau, V.G. Belille, F. Chausse and J.T. Lapreste, "Real-Time Tracking with Classifiers," Proc. the International Workshop on Dynamical Vision in conjunction with ECCV, 2006.
- [2] M. Isard and A. Blake, "Condensation - Conditional Density Propagation for Visual Tracking," International Journal of Computer Vision, vol.29, no.1, pp.5-28, 1998.
- [3] Y. Li, H. Ai, T. Yamashita, S. Lao and M. Kawade, "Tracking in Low Frame Rate Video: A Cascade Particle Filter with Discriminative Observers of Different Lifespans," Proc. the IEEE International Conference on Computer Vision and Pattern Recognition, pp.17-22, 2007.
- [4] 松本郁佑, 加藤丈和, 和田俊和, "Network Augmented Multisensor Association-CONDENSATION: CONDENSATIONの自然な拡張による3次元空間内での人物頭部の実時間追跡," 情報処理学会研究報告, 2005-CVIM-150-21, pp.161-168, 2005.
- [5] K. Okuma, A. Taleghani, N. Freitas, J. Little and D. Lowe, "A Boosted Particle Filter: Multitarget Detection and Tracking," European Conference on Computer Vision, vol.3021 of LNCS, pp.28-39, 2004.
- [6] 杉本晃宏, 谷内清剛, 松山隆司, "確信度付き仮説群の相互作用に基づく複数対象追跡," 情報処理学会論文誌, vol.43 no.SIG CVIM 4, pp.69-84, 2002.
- [7] 鈴木達也, 岩崎慎介, 小林貴訓, 佐藤洋一, 杉本晃宏, "環境モデルの導入による人物追跡の安定化," 電子情報通信学会論文誌 DII, vol.J88-DII no.8, pp.1592-1600, 2005.
- [8] P. Viola and M. Jones, "Rapid Object Detection Using a Boosted Cascade of Simple Features," Proc. the IEEE International Conference on Computer Vision and Pattern Recognition, vol.1, pp.511-518, 2001.
- [9] Y. Wang, J. Wu and A. Kassim, "Particle Filter for Visual Tracking Using Multiple Cameras," Proc. IAPR Conference on Machine Vision Applications, pp.298-301, 2005.
- [10] C. Yang, R. Duraiswami and L. Davis, "Fast Multiple Object Tracking via a Hierarchical Particle Filter," Proc. the IEEE International Conference on Computer Vision and Pattern Recognition, vol.1, pp.212-219, 2005.

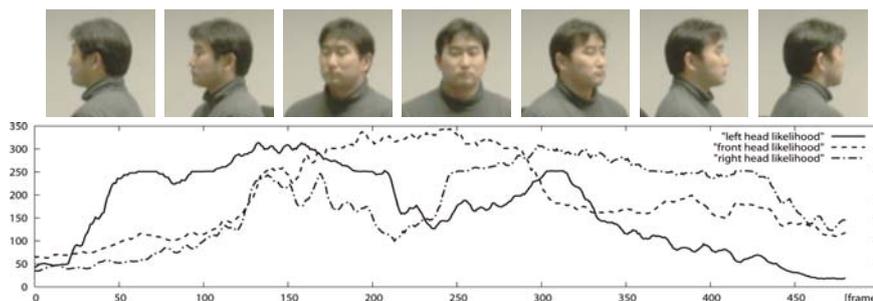


図5. 頭部の回転と評価値