

研究テーマ 顔向き推定に関する研究

研究者	安本 護	財団法人ソフトピアジャパン	雇用研究員
	奥水大和	中京大学	兼業研究員
	安藤道則	株式会社豊田中央研究所	共同研究員
	金川 誠	三洋電機株式会社	共同研究員

フェーズ I

1 研究の概要

本研究では、複数方向の顔画像から抽出した四方向面特徴を用いて、人物によらない顔向き推定手法を開発した。また、複数のカメラで異なる方向から画像を取得し、各カメラの識別器出力を統合するマルチカメラ協調手法も提案する。さらに、顔向き推定手法を利用した顔向きに依らない人物識別の実験を行った。顔向き推定実験からは、推定誤差の平均が 4.5° 以下、分散が 11.0 という結果が得られた。個人識別実験では、1台のカメラで 296° の顔向き範囲に対して平均 92.9% の認識率が得られた。また、8台のカメラから得た情報を統合する処理が、顔向き推定の精度向上や個人識別のオクルージョン耐性に効果的であることを示した。

2 研究の目標

実環境における人物の注目方向推定や個人識別、顔の向き等を利用したインタフェースにおいて、人物はカメラに対して常に正面を向いているとは限らない。そのため、カメラに撮影された顔画像の向きを推定する手法の確立が重要である。モデルベースの顔向き推定手法では、初期キャリブレーションが必要であり、モデルの成立する顔向き範囲にも制限がある。位置固定のカメラに対して、複数の利用者が様々な位置に出現するような環境におけるインタフェースでは、初期キャリブレーションを省略可能なアピランススペースの手法が有効である。また、実環境においては、自分自身や他人、家具等によるオクルージョンが問題となる。オクルージョンに対しては、多様な方向から画像を取得するマルチカメラによるアプローチが有効と考えられる。本研究では、初期キャリブレーションの必要がなく、広範な顔向きを推定する手法の確立を目指す。また、マルチカメラによる顔向き推定と個人識別を組み合せ、顔向きに依らない個人識別手法を提案する。

3 実施内容

3.1 四方向面特徴による顔画像の判別手法の開発

本研究で提案する顔向き推定ならびに人物識別は、顔領域の四方向面特徴を線形判別分析によって識別する処理に基づく。図1にこの顔画像識別処理の概要を示す。まず、肌色情報をもとに入力画像から対象とする顔領域を検出する。次に顔領域の輝度成分から四方向面特徴 (Four Directional Features) ベクトル x を求める。更に x を線形判別分析で得た係数行列 A により判別空間のベクトル y に変換する。最後にベクトル y と判別クラス C_j の平均ベクトル \bar{y}_j との距離 D_j を求め、距離最小のクラスを識別結果とする。本手法は、アピランススペースの手法であり、判別クラスの構成によって顔方向推定や個人識別、更に手サイン認識にも応用可能である。以下に各処理の詳細に関して述べる。

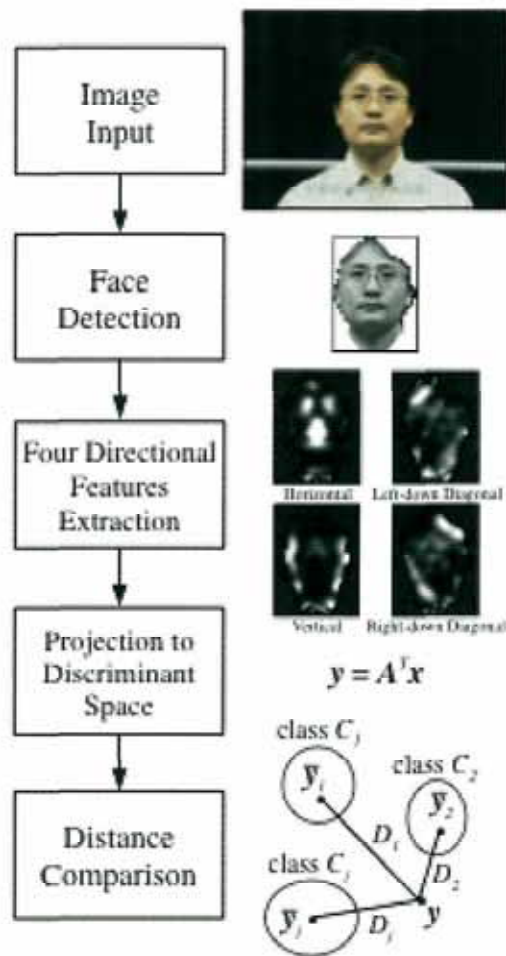


図1 顔画像識別処理の概要

3.1.1 顔領域検出

画像中から顔領域を検出する際の手掛かりには、動き、色、形状の情報がある。本研究では単純背景下の単一被験者を対象とするため、色情報のみを用いて顔領域、即ち肌色領域を検出する。まず、入力画像を CIE-L*u*v* 表色系に変換する。L* は各画素の明度、u*、v* は色度であり、u*v* 値の 2 次元ヒストグラムにおいて肌色領域が集中して分布することを実験により確かめている。次にあらかじめ設定した肌色有効範囲内で頻度最大となる u*v* 値を基準肌色とする。更に基準肌色と入力画像各画素の u*v* 値の距離を求め、これに判別分析法を適用して肌色領域とその他の領域を分離する。基準肌色を固定せずにフレーム毎に適応的に決定することにより、個人差やカメラ特性、照明等の条件に影響を受けにくい安定な肌色領域抽出が可能である。最後に、抽出された複数の肌色領域中で面積が最大かつ閾値 S_i 以上である領域を顔領域に決定する。閾値 S_i は、後頭部付近のような微小肌色領域が検出されることを排除し、次に述べる四方向面特徴を安定に取り出せるように決定する。

3.1.2 四方向面特徴抽出

四方向面特徴は、OCR のパターンマッチングに用いられた手法で、入力 2 値画像の各画素における勾配ベクトルを四方向の面特徴として割り当てて表現するものである。この手法は多値画像に拡張することが可能であり、顔画像のパターンマッチングに有効であることが報告されている。四方向面特徴の抽出は、以下の手順で行う。

- (1) 抽出された顔領域の輝度信号に対し、各々四方向（水平、垂直、右上がり 45 度、右下がり 45 度）の Prewitt オペレータを適用して 4 種類の濃淡エッジ画像を生成する。
- (2) 各エッジ画像にガウシアンフィルタを適用して 8×8 に低解像度化する。
- (3) 4 枚の濃淡エッジ画像を統合して 256 次元特徴ベクトル x とする。

エッジ画像の低解像度化は、位置ずれや形状変化の影響を軽減するとともに処理速度を向上させるためである。また、四方向面作成後に低解像度化するのは、低解像度化による方向別エッジ情報の歪み発生を避けるためである。

3.1.3 線形判別分析

学習データがどのクラスに属するかわかっているとき、線形判別分析は、四方向面特徴ベクトル \mathbf{x} をより識別に適したベクトル \mathbf{y} へ写像する線形変換 $\mathbf{y}=\mathbf{A}^T\mathbf{x}$ を与える。係数行列 \mathbf{A} は、判別基準 $J(\mathbf{A})=tr(\hat{\Sigma}_W^{-1}\hat{\Sigma}_B)$ 最大を条件に決定する。ここで、 $\hat{\Sigma}_W$ 及び $\hat{\Sigma}_B$ は、それぞれクラス内共分散行列及びクラス間共分散行列を表す。

3.1.4 識別器の構成

識別は、テストデータの四方向面特徴ベクトル \mathbf{x} を $\mathbf{y}=\mathbf{A}^T\mathbf{x}$ により変換した \mathbf{y} と各判別クラス C_j の平均ベクトル $\bar{\mathbf{y}}_j$ との二乗距離 $D_j=|\mathbf{y}-\bar{\mathbf{y}}_j|^2$ を求め、 D_j を最小とするクラスに決定すればよい。なお、入力画像の肌色領域が閾値 S_i より小さくて顔検出に失敗するときは、 $D_j=\infty$ とする。ここで、二乗距離をそのまま識別器の出力に用いると、複数の識別器の出力を統合する場合に、二乗距離の大きなクラスが結果に大きく影響し、扱いが困難となる。そこで、式 (1) を用いて二乗距離を d_j ($0\leq d_j\leq 1$) に変換し、識別器の出力とする。 $D_j=\infty$ のときは、 $d_j=0$ である。この結果、識別器の統合は出力 d_j の和として簡単に実現できる。識別結果は、識別器出力 d_j あるいはその統合和が最大となるクラス C_j に決定する。

$$d_j = \frac{1}{D_j + 1} \tag{1}$$

3.2 顔向き推定

四方向面特徴を用いた個人識別では、 $\pm 15^\circ$ 程度の顔向き変動を許容できることが報告されているため、個人識別の大分類に用いるのであれば、16 方向の顔向きを推定できれば十分である。しかしながら、顔の向きを操作対象の指定に用いるヒューマンインタフェースでは、更に高い分解能が必要である。個人識別の大分類に用いた手法で顔向き推定の分解能を上げるには、大量の顔向き識別器が必要である。これは、判別クラス間の距離を減少させることになり、判別性能の低下を招いてしまう。ここでは、少数の識別器を使って任意角度の顔向きを推定する手法について述べる。

対象物の方向を検出するには、顔向きだけでなく人物の位置を求める必要もある。しかしながら、人物位置はステレオマッチング等の手法で求められるので、本研究では触れない。

3.2.1 顔向き推定の座標系

顔向き推定に用いる座標系を図 2 に示す y 軸正の向きをシステムの基準方向とし、人物位置 $P(r_p, \phi_p)$ 及びカメラ位置 $Q_i(r_q^{(i)}, \phi_q^{(i)})$ を既知とするとき、基準方向に対する顔角度 θ_a は式 (2) で表される。ここで、原点 O からカメラを見る角度 $\phi_q^{(i)}$ は既知であり、光軸を基準にカメラから顔を見る角度 $\alpha_q^{(i)}$ 式 (3) で与えられる。従って、 θ_a の推定は、カメラから観測される顔角度 $\theta_q^{(i)}$ の推定問題に帰着される。

$$\theta_a = \theta_q^{(i)} + \phi_q^{(i)} - \alpha_q^{(i)} \tag{2}$$

$$\alpha_q^{(i)} = \tan^{-1} \frac{r_p \sin(\phi_p - \phi_q^{(i)})}{r_q^{(i)} - r_p \cos(\phi_p - \phi_q^{(i)})} \tag{3}$$

3.2.2 顔向き識別器の出力特性

図 3 は、被験者を 360° 回転して撮影した顔画像に対する顔向き識別器の出力を示している。ここで、顔向き角度は、顔がカメラを正面に見るときを 0° 、カメラの右を見るときを正、逆に左を見

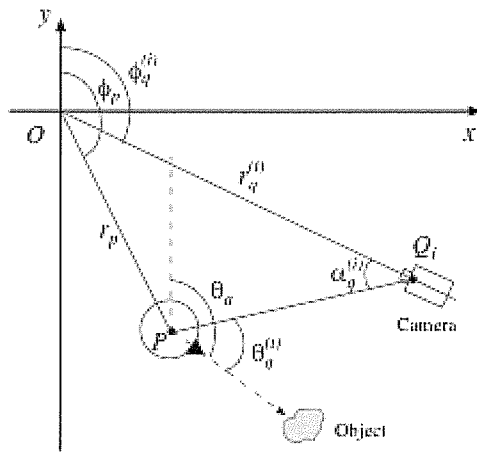


図2 顔方向の座標系

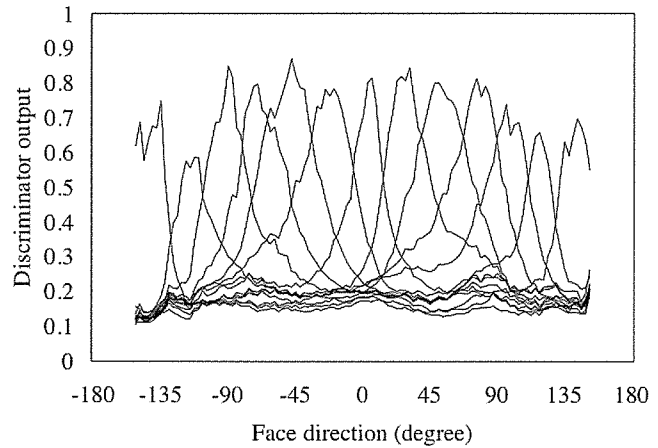


図3 顔向き識別器出力

るときを負の角度で表している。顔向き識別器は、個人識別の大分類に用いるものである。線形判別分析では、学習データ用として10人の被験者から22.5°間隔16方向の顔画像を収集し、同一方向10人分の特徴ベクトルをそれぞれひとつの顔向き判別クラスに学習させている。顔向きが±157.5°及び180°のときは、画像中の肌色領域の面積が閾値 S_i に達しなかったため、顔を検出できなかった。従って、図3には、これら3方向を除く13個の顔向き識別器出力を示した。図3は、入力画像の顔向き角度と顔向き識別器出力との間に以下の性質があることを示している。

- (1) 入力画像の顔向き角度に最も近い顔方向を学習した識別器の出力値が最も大きい。
- (2) 入力画像の顔向きが識別器の学習方向となるとき、出力値は最大となる。
- (3) 識別器出力は、顔向きが学習方向から離れるに従い、ほぼ単調に減少するが、ある角度を超えると0.2付近に収束する。
- (4) 0°を中心として、対称な顔向き角度を学習した識別器の出力波形は、ほぼ対称な形状を持つ。これは、頭部形状が左右対称である事実と一致している。顔向き識別器の出力が、学習方向の近辺でほぼ単調に変化することは、顔向きの連続的な角度推定が可能であることを示唆している。グラフには示さないが、判別クラス学習に用いなかった被験者の顔データについても同様の性質を確認している。

3.2.3 顔向き識別器を用いた顔向き推定

顔向き識別器の出力を顔向きの連続的な角度推定に用いるために、顔向き識別器の出力をベクトルで記述し、それらを用いて入力画像の顔向きベクトル f を表現する。各識別器の出力をベクトル表現したものは、入力顔向きベクトル f を各識別器の学習顔方向へ投影したものと考えられるので、これを方向成分ベクトルと呼ぶ。顔向き j に対応する方向成分ベクトル c_j を式(4)で定義する。 d_j は、式(1)で示した判別器出力である。また、 u_j は、顔向き識別器が学習した方向 j の単位ベクトルである。顔向き推定結果は、顔向きベクトルの推定 \hat{f} の方向となる。

顔向き識別器の出力に完全な対称性が成立すれば、式(5)に示す方向成分ベクトル c_j の総和で顔向きベクトルの推定を与えればよい。

$$c_j = d_j u_j \quad (4)$$

$$\hat{f} = \sum_{j=1}^{16} c_j \quad (5)$$

しかし、顔向きが0度からずれるに従い、識別器出力の対称性は崩れていく。そこで、特定の識別器出力を選択して推定を行わねばならない。図3から第4位以下の識別器出力にほとんど差はなく、顔向き推定への寄与は小さいと考えられる。従って、出力値第1位の識別器と隣接する顔向き識別器の合計3個を推定に使用する。顔向きに対する識別器出力の対称性を考慮すると、顔向きが

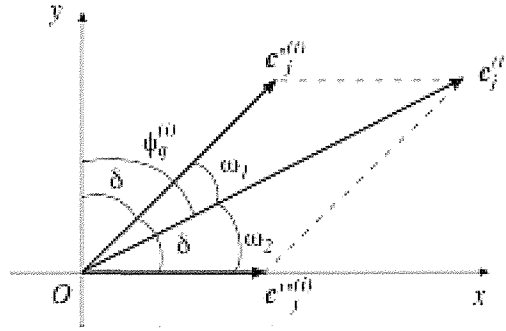


図 4 方向成分ベクトルの分解

学習方向のときは 3 個の識別器出力全てを推定に用い、学習方向のちょうど中間にあるときは第 1 位と第 2 位の識別器のみを推定に用いるのが適当である。そこで、式 (6) の顔向き推定式は、顔向きに応じて方向成分ベクトルの係数を適応的に変更する。式 (6) において、 d_m は第 1 位識別器出力 ($d_m \geq d_i, i=1, \dots, m-1, m+1, 16$) であり、これに隣接する顔向きの識別器出力 d_{m-1}, d_{m+1} には、 $d_{m-1} \geq d_{m+1}$ の関係があるとする。 $a_1 = 1$ のとき、顔向きは $m-1$ 番め識別器と m 番めの識別器が学習した方向のちょうど中間であると予想できる。このとき、式 (6) の d_{m-1}, d_m, d_{m+1} の係数は、それぞれ 1、1、0 となって、第 1 位と第 2 位の識別器のみが推定に用いられる。また、 $a_2 = 1$ のとき、顔向きは m 番め識別器の学習方向と予想できる。式 (7) の係数は 0.5、1、0.5 となり、第 1 位を中心に 3 個の識別器が使用される。 a_1, a_2 がともに 1 以下のとき、顔向きは上記の間にあると予測され、 a_1, a_2 の値に応じて方向成分ベクトルの係数を変化させる。

$$\hat{f} = (1 - 0.5a)c_{m-1} + c_m + 0.5ac_{m+1} \quad (6)$$

ここで、

$$a = \begin{cases} 1 - a_1 (a_1 \geq a_2) \\ a_2 (a_1 < a_2) \end{cases} \quad (7)$$

$$a_1 = d_{m-1} / d_m \quad (8)$$

$$a_2 = d_{m+1} / d_{m-1} \quad (9)$$

3.2.4 マルチカメラ統合推定

図 3 からは、顔の向きがカメラの正面から大きく遠ざかると顔向き角度の推定結果が劣化することが予想される。特にカメラが後頭部付近を撮影しているときは、顔領域を検出できず顔向き角度は推定不可能である。また、実環境においては自分自身の手や体の一部、家具あるいは他人の通過等によってカメラと顔の間が遮られる場合がある。このような場合でも、別のカメラがオクルージョンのない方向から顔を観測していれば、顔向き推定が可能となるであろう。そこで、本節では複数のカメラから得た情報を統合することにより、任意の顔向きに対して適切に角度を推定可能で、オクルージョンにもロバストな顔向き推定手法を提案する。

図 4 を用いて、複数カメラの協調処理を説明する。各カメラから得た方向成分ベクトルは、システム座標系の方向成分ベクトル (角度間隔 δ) に変換して統合される。 i 番目のカメラの方向成分ベクトル $c_j^{(i)}$ を直近両隣のシステム方向成分ベクトルの方向に分解すると、ベクトル $c_j'^{(i)}$ と $c_j''^{(i)}$ が得られる。分解後のベクトルの長さは、それぞれ式 (10)、(11) で与えられる。式 (12) の $\lfloor \bullet \rfloor$ は小数点以下の切捨てを表す。

$$\left| c_j'^{(i)} \right| = \frac{\left| c_j^{(i)} \right| \sin \omega_2}{\sin \omega_1 \cos \omega_2 + \cos \omega_1 \sin \omega_2} \quad (10)$$

$$\left| c_j''^{(i)} \right| = \frac{\left| c_j^{(i)} \right| \sin \omega_1}{\sin \omega_1 \cos \omega_2 + \cos \omega_1 \sin \omega_2} \quad (11)$$

$$\omega_1 = \phi_q^{(i)} - d[\phi_q^{(i)} / \delta] \quad (12)$$

$$\omega_2 = \delta - \omega_1 \quad (13)$$

式 (14) は、システム座標系に統合した方向成分ベクトルを与える。ここで、 M は協調するカメラの台数を示す。

$$|c_j| = \sum_{i=1}^M (|c_j^{(i)}| + |c_{j-1}^{(i)}|) \quad (14)$$

最後にシステム座標系に統合した方向成分ベクトルを顔向き推定式に適用する。

3.3 顔向きに依存しない個人識別

3.3.1 階層的識別法

顔向きに依存しない人物識別を単一の判別空間を用いて実現する場合、人物識別の判別クラスには、次の二つの作成法が考えられる。

- ・判別クラスに必要な全ての顔向きを学習させる。
- ・必要な顔向きの数だけ判別クラスを作成する。

前者は、向きによって形状が大きく異なる顔を一つの判別クラスで表現するため、十分な識別性能が得られない。また、後者も実用的なシステムを構築するには大量の判別クラスを要するため、十分なクラス間距離を確保できなくなって識別性能が劣化する。そこで、以上の問題を解決するために階層的な識別手法を用いる。これは、人物に依存しない顔向き推定の判別空間で大分類を行い、同じ顔向きについての顔判別空間で人物を識別するものである。

四方向面特徴を用いた人物識別では、 ± 15 度程度の顔向き変動を許容できることが知られている。そこで、大分類では水平方向を 22.5 度間隔に分割し、顔がどの方向を向いているかを識別する。判別クラスには、複数の人物から収集した顔画像を学習させ、人物に依存しない顔向き判別クラスとする。人物を識別する判別空間は、顔向き別に 16 個作成する。顔向き別の人物識別空間には、識別対象人数分の判別クラスを作成する。

3.3.2 マルチカメラ統合を用いた個人識別

実環境においては、自身の手等によるセルフオクルージョンや移動にともなって家具等の陰に入るオクルージョン、他人の通過によるオクルージョン等が発生する。これに対して、本研究ではマルチカメラの協調による識別を試みる。例えば、正面カメラでは手によるオクルージョンで人物の認識に失敗しても、 90 度横のカメラで認識が可能なら、これら 2 台のカメラを協調させて正しい認識結果を導き出すことが可能であろう。

マルチカメラ統合処理を式 (15) に示す識別器出力の総和 $F(j)$ で与える。 $d_j^{(m)}$ は、 m 番目カメラの判別クラス C_j に対する識別器出力である。 M は、統合に用いるカメラの台数である。

$$F(j) = \sum_{m=1}^M d_j^{(m)} \quad (15)$$

4 結果

本研究で提案する顔向き推定手法と個人認識手法の有効性を確認するために行った実験の概要と結果を示す。

4.1 実験システムの構成

実験に用いたシステムを図 5 に示す。本システムは、複数のカメラを配置した室内環境を想定したもので、無地のカーテンで仕切った一辺 5 メートルのスタジオと、カラー CCD カメラ、PC 等の撮影機材からなる。

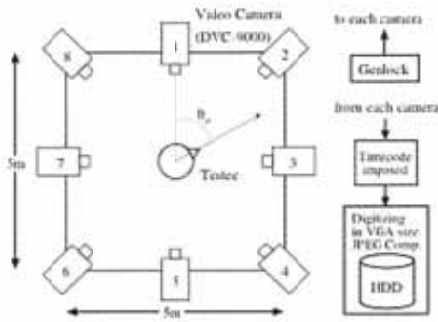


図 5 実験システムの構成



図 6 オクルージョンのある顔画像

システムの座標原点はスタジオの中央であり、原点からカメラ 1 の方向を基準方向とした。照明光は、天井に埋め込まれた蛍光灯から得た。カメラ 8 台は、被験者の顔を中心とする水平面に 45° 間隔で配置した。各カメラのズーム率は、映像中の顔の大きさが等しくなるように調整した。同一時刻に 8 方向から撮影したフレームを取得可能とするため、各カメラに外部同期を掛け、タイムコードを重畳した後、30 フレーム/秒でデジタル化し、モーション JPEG 圧縮を行って PC のハードディスクに記録した。実験に用いた顔画像データは、VGA (640×480) サイズ、フルカラーである。

4.2 実験データ

識別器の学習を行うために、22.5 度間隔 16 方向の顔画像データを被験者 1 人につき各方向 400 枚収集した。図 5 のカメラ配置を用いることにより、2 方向を向くだけで 16 方向の顔画像データを収集できるため、被験者に大きな負担を掛けることはない。10 人の被験者から合計 64,000 枚の学習用データを収集した。

顔方向変化に対する性質を調べるため、360 度回転する人物を 8 台のカメラで同時に撮影し、実験データを収集した。被験者は学習データ収集と同じ 10 人である。被験者の顔が各カメラに正対するフレームを目視で検出し、45 度毎の基準フレームとした。基準フレーム間は等速回転とみなし、時間差で 16 等分して 1 回転 128 方向の画像を得た。

実験データの総数は 10,240 枚となった。

マルチカメラ統合人物識別のオクルージョン耐性を調べるために、顔の前に手をかざす動作の画像データシーケンスを収集した。図 6 に全 20 フレーム中の第 13 フレームを示す。

4.3 顔向き推定

顔向き推定の性能評価には、式 (16) に定義する推定誤差 e とその分散 σ_e^2 を用いる。

$$e = \left| \theta_a - \hat{\theta}_a \right| \quad (16)$$

単一カメラを用いた場合の顔向き変化に対する被験者 10 人の平均推定誤差と推定誤差の分散 σ_e^2 を図 7 に示す。(a)、(b)、(c)、(d) は、それぞれ式 (5)、式 (6)、式 (17)、式 (18) を推定に用いた。

$$\hat{j} = 0.5c_{m-1} + c_m + 0.5c_{m+1} \quad (17)$$

$$\hat{j} = c_{m-1} + c_m \quad (18)$$

全ての識別器出力を統合する (a) は、0° 付近で良い推定結果を示すが、0° から離れるに従い急激に誤差が大きくなっている。この結果は、顔向き識別器の出力に完全な対称性が成立していないことを示している。(b)、(c)、(d) は、±135° の外側で推定誤差が、急激に大きくなっている。これは、顔領域検出の失敗によって ±157.5° から外側の識別器が得られなかったためである。(c) および (d) は、学習方向と同じ 22.5° の周期で推定誤差が変化している。3 つの識別器を統合した (c) は識別器の学習方向付近、2 つの識別器を統合した (d) は学習方向の中間点付近で極小値をとって

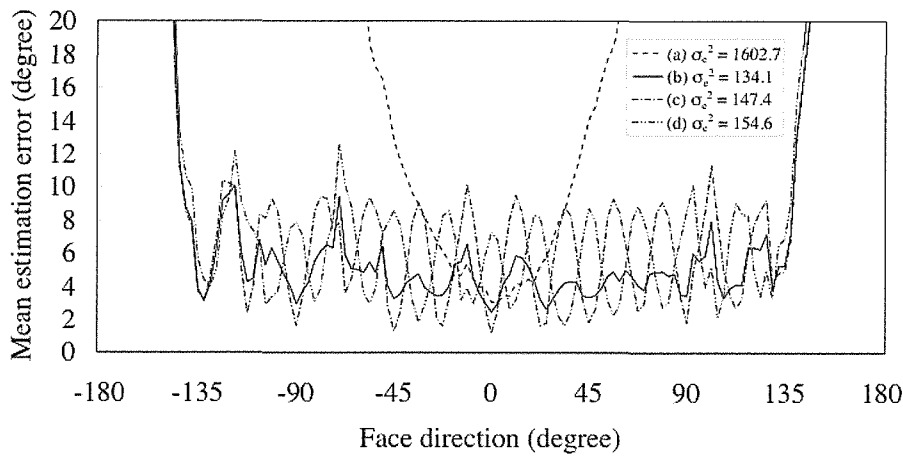


図7 単一カメラを用いた顔方向の平均推定誤差

る。(b) は、 a_1 、 a_2 による係数適応化を行ったものである。(c)、(d) に比較して変動が小さく良

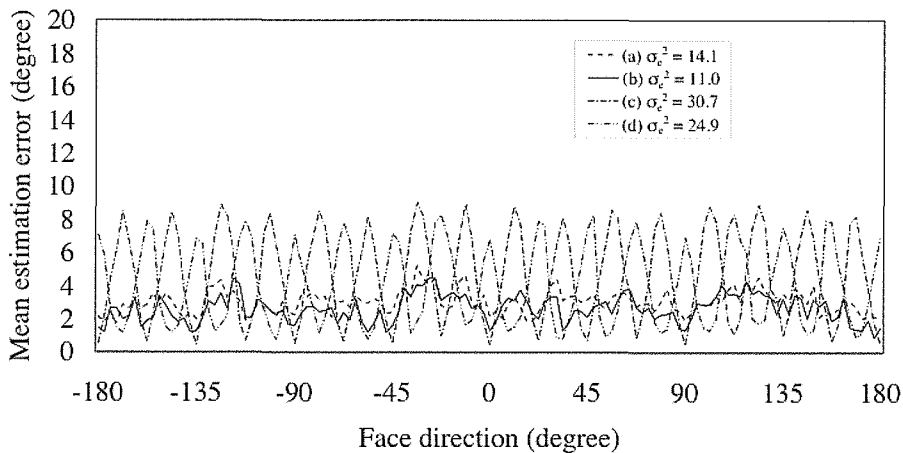


図8 マルチカメラを用いた顔方向の平均推定誤差

表1 階層化個人識別結果

有効枚数	正解枚数	有効角 (度)	認識率 (%)
8429	7832	296	92.9

好な推定結果が得られた。顔画像入力から顔向き推定までに要した時間は 0.3 秒 (PentiumIII、550MHz) であった。処理時間の大部分は顔領域検出と四方向面特徴抽出に要するもので、顔向き推定に要する時間はわずかである。

図8は、マルチカメラ統合による推定結果である。45°間隔8台のカメラ中、顔領域検出に失敗するカメラは1台のみである。これらの統合によって、単一カメラのように $\pm 157.5^\circ$ から外側で推定誤差が急増することがなくなった。識別器の出力の統合によって顔向きに対する対称性が向上し、

(a)の結果も任意の顔向きにおいて良好となった。他の推定式でも単一カメラより良好な結果が得られ、マルチカメラ統合の効果が確認できた。顔向きに応じて係数を適応制御する(b)からは最も良い推定結果が得られ、平均推定誤差は 4.5° 以下となった。これは、5m先で約40cmに相当し、室内にある家電機器等の指定には十分である。マルチカメラ統合では、画像入力から個別の識別器出力を得るまでの計算量がカメラの台数に比例して増加する。しかし、この処理はカメラごとに分散処理が可能なので、システム全体のスループットは単一カメラの場合と変わらない。

4.4 階層化個人識別

表1に個人識別の実験結果を示す。有効枚数とは、肌色領域の面積が閾値 S_f よりも大きくて認識の対象となった画像データの枚数である顔領域検出の閾値は $S_f=400$ 画素とした。有効角は、入力

画像 360° のうち認識対象となった角度を示し、式 (19) で求める。また、認識率は式 (20) で求める。

$$\text{有効角} = \text{有効枚数} / \text{入力総枚数} \times 360 \quad (19)$$

$$\text{認識率} = \text{正解枚数} / \text{有効枚数} \times 100 \quad (20)$$

実験により、約 300 度という広範囲な顔方向に対して平均 92.9% の認識率が得られた。今回は人物認識辞書に単一方向の顔画像しか学習させなかったが、大分類の角度範囲内で複数顔方向の画像を学習させれば、より高い認識率を得ることも期待できる。また、マルチカメラ統合では、図 6 のようにオクルージョンの発生した顔画像でも失敗なく認識できた。

フェーズ II

1 研究の概要

フェーズ I では、カメラに対する人物の相対的な顔向きを推定する技術開発を行った。顔向きは、人物の大まかな注視方向を反映していると考えられるため、高い精度を必要としない簡易的な視線推定手法として活用できる。このような視線推定技術を車載カメラから得られた画像に応用することで、自動車ドライバの運転技術や運転技術の解析等に対する活用が考えられる。また、広告媒体にカメラを取り付けることに依って、提示広告に対する注目情報の取得に応用することも考えられる。

本研究では、顔向き推定技術を活用した正面顔判定とステレオ視を組み合わせることで、広告媒体に対する注目度を計測するシステムを試作した。試作システムでは、50 インチの PDP 上に水平 2 分割の状態では、広告映像の表示を行った。分割した各画面への正面顔検出実験結果は、カメラ（システム）と人物の距離が 2.7m 以内である場合に、90% 以上であった。このようなシステムは、計測が困難な広告効果の検証等への活用が期待できる。

2 研究の目標

ステレオ視や肌色情報による顔検出等の既存の画像処理手法に、顔向き推定技術を組み合わせた応用システムの試作を行う。応用システムでは、マーケティング調査を目的とした広告媒体への人物の注目行動を検出する。注目行動の推定では、フェーズ I で研究を行った顔向き推定技術を利用した正面顔検出を提案し、人物が広告を見る行動を検出する。また、ステレオ視を用いて人物行動の 3 次元情報を取得することによって注目行動をより総裁に解析するシステムを目指す。本システムでは、検出された人物の大きさや衣服の色名殿の人物属性に関連する情報の収集も試みる。さらに、システムの前部を通過する人数計数手法を提案する。

3 実施内容

3.1 画像処理による基礎技術の研究開発

3.1.1 画像処理システム

本研究で用いる画像処理システムでは、複数のカメラを用いて、人物検出、位置算出、人物属性推定などの処理を行う。複数のカメラを用いて、撮影対象の位置算出を行う手法としては、撮影空間の周囲に多数のカメラを設置し、各カメラから得られる画像を統合する手法や、ステレオ視を用いる手法がある。本研究では、カメラの設置が比較的容易である、ステレオ視を用いることとした。情報取得を行う画像処理システムは、画像処理用のコンピュータとステレオカメラからなる。図 9 に画像処理系のブロック線図を示す。ステレオ視を利用する際には、カメラの同期撮影が必要となるが、1 台のコンピュータに複数のカメラを接続し、各カメラに連続してキャプチャ命令を送ることにより、ほぼ同時に画像取得を行うことができる。画像処理システムの最小構成は、ステレオカ

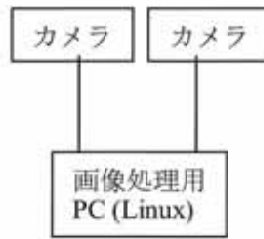


図9 画像処理系のブロック線図



図10 使用したカメラ

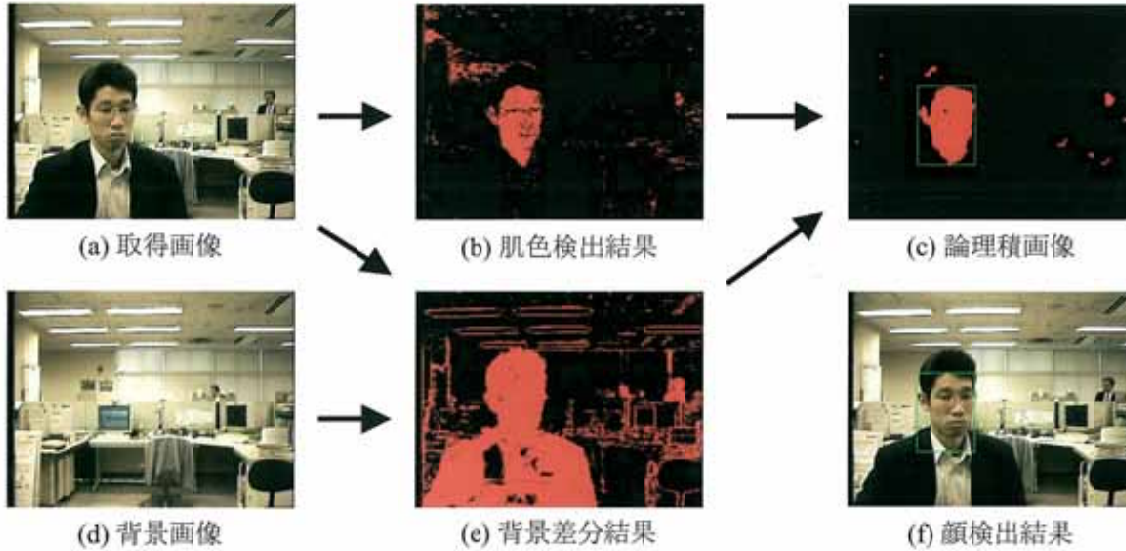


図11 顔検出結果例

メラと画像処理用コンピュータのみであるが、コンピュータをネットワーク接続し、取得情報を相互に交換、統合することによって、詳細な情報取得が可能となる。

本システムでは、画像処理を行うコンピュータは、CPUが PentiumIII (1GHz)、OSを Linux とし、Video4Linux ライブラリを用いて画像を取得する。使用カメラ (図10: Sony製 FCB-EX470L) には、環境変化に応じた内部パラメータの自動補正機能が搭載されているため、外光などの影響によって撮影対象の色情報に影響を与える。そのため、制御コンピュータから VISCA コマンドを送信し、シャッタースピード、露出などの内部パラメータを固定した。

3.1.2 人物検出

本研究では、顔検出を行うことによって、人物検出とした。顔の検出は、色情報と背景差分を利用した手法を用いた。次の式を満たす画素を肌色画素として検出し、2値画像を得た。

$$\begin{cases} th_1 < r < th_2 \\ th_3 < g < th_4 \\ R + G + B > th_5 \end{cases} \quad (21)$$

式 (21) 中で、 $r=R/(R+G+B)$ 、 $g=G/(R+G+B)$ である。また、 th_1 、 th_2 、 th_3 、 th_4 、 th_5 の各値は、人物が撮影されている画像から、肌である調域内の画素を手動で切り出し、その値を用いて設定を行った (順に、0.415、0.425、0.315、0.325、150)。

色情報のみを利用した手法では、背景などに肌色の物体が存在した場合、誤検出が生じる。そのため、背景差分法を利用し、誤検出の低減を図った。背景差分法では、予め記録した背景画像と輝度値が異なる画素を検出し、2値画像を得た。

これらの結果として得られた2値画像の論理積を計算し、背景に存在しない肌色領域を検出した。膨張・縮退などのノイズ処理を行い、一定面積以上の肌色領域を顔領域として検出した (図11)。

3.1.3 人物位置の算出

撮像空間内の人物位置は、ステレオ視を用いて算出した。ステレオ視は2台以上のカメラから得

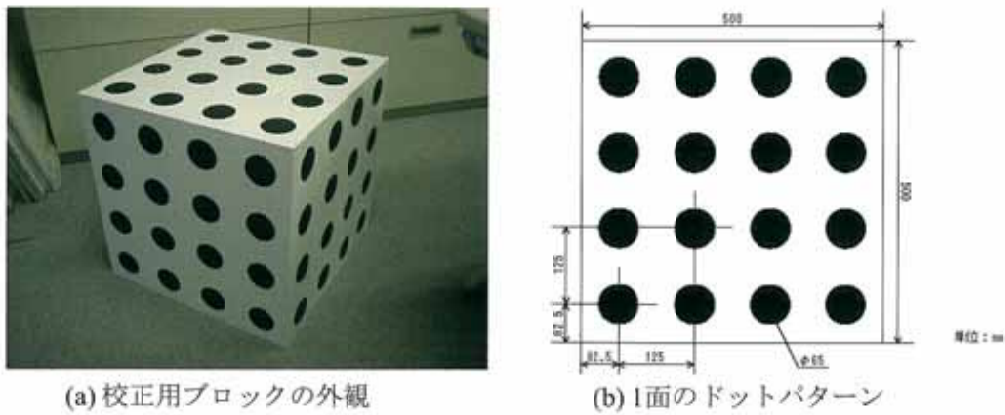


図 12 校正用ブロック



図 13 カメラ調整台

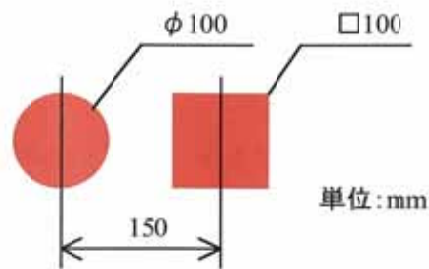


図 14 実験に使用した図形

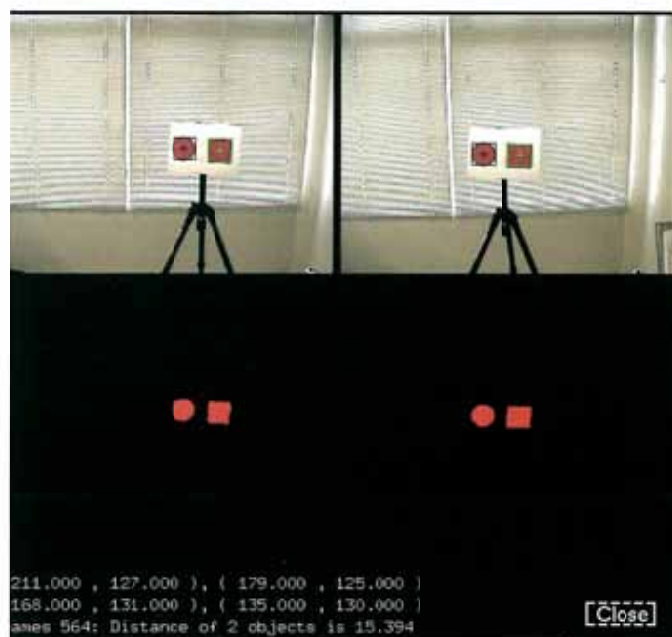


図 15 ステレオ視精度実験

られる画像の対応点を探索し、各画像内での対応点座標の差(視差)を利用して3次元座標を得る。カメラの数を多くし、信頼性の高い情報を選択することによって、精度が向上することが知られている。しかし、制御系やコストに負担がかかるため、本研究では2台のカメラを用いることとした。以下では、2台のカメラの1台を、物体検出を行う基準カメラ、もう1台を検出された物体の対応点を探索する参照カメラとする。

カメラの内部パラメータの校正では、校正用パターンを印刷したブロック(図12)を撮像し、校正用パターンの各ドット領域の重心点座標を入力することによって行った。従って、撮影空間内

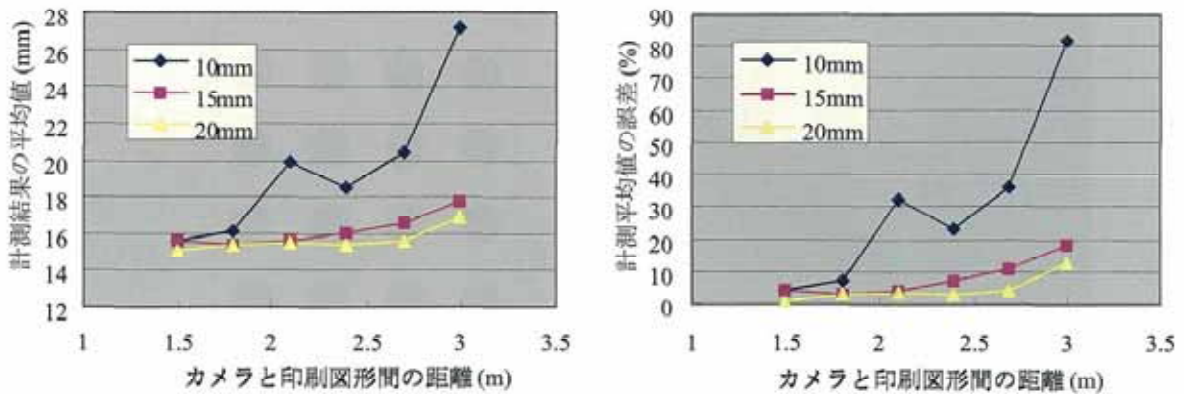


図 16 ステレオ視精度実験結果

の 3 次元空間座標は、校正時に設置した校正用ブロックの位置が基準となる。

カメラ間距離が大きくなると、視差が大きくなるため、カメラから離れた物体に対しても算出精度の向上が期待できる。その反面、カメラに近い物体では、対応点探索に必要な計算コストの増大や、両方のカメラの共通撮像領域が小さくなるなどの問題が生じる。本研究では、カメラ間距離は、70mm から 200mm まで調節できるカメラ調整台を使用し (図 13)、利用状況に応じて設定することとした。ステレオ視におけるカメラ間の対応問題の解法には、相関法を用いた。相関値で類似度を -1 から 1 までの範囲で表現し、1 に近ければ近いほど、両領域の類似度が高い。

カメラ間距離を 100、150、200(mm)とした場合のステレオ視による 3 次元位置算出精度実験を行った。実験は、図 14 に示す図形を印刷した紙を、ステレオカメラ対の前に設置し、印刷図形間の距離を算出した。実験中の表示画面を図 15 に示す。図 15 では、画面が 4 分割されており、左上に基準カメラから得られた画像、左下に基準カメラ画像の 2 値化画像、右上に参照カメラから得られた画像、右下に参照カメラ画像の 2 値化画像が表示されている。図中で、基準カメラと参照カメラからの取得画像内において、黒色と緑色の矩形は各検出領域の外接矩形を表示しており、十字は重心位置を示している。図形間の距離は、重心位置を用いて算出している。

カメラと紙との距離は、1.5、1.8、2.1、2.4、2.7、3.0(m)とし、各距離において、100 回の計測を行った。計測結果の平均値、分散、平均値の誤差を図 16 に示す。カメラ間距離が 100mm である場合には、カメラと印刷図形間の距離が 2m 以上になった場合、誤差が 20% を超える結果となった。誤差は、カメラ間距離が 200mm の場合に、最も誤差が小さくなるが、150mm の場合においても、カメラと印刷図形間の距離が 2.7m の場合においても、10% 程度であり、人物位置の計測誤差としては、許容できる範囲と思われる。計測誤差は、主に対応点の探索から得られる視差に起因しており、対応点探索の高精度化によって、精度の向上が可能と考える。また、カメラと印刷図形間の距離が大きくなると、1 画素に対する算出距離の差が大きくなる。特に、カメラ間距離が小さい場合には顕著であるため、撮影領域に応じてカメラ間距離を検討する必要がある。

3.2 人物の注目行動および属性の検出

本研究では、注目情報を検出するための対象物への注目の検出を行う。また、人物属性として、人物の大きさや衣服の色等を検出する。

3.2.1 動きおよび注目の検出

人物の意思や興味を示される行動のひとつとして移動行動が考えられる。例えば、目的地に向かって道を歩いている時に、ポスターや掲示板に興味を持てば、その前で立止り、凝視することがある。立止らない場合でも、通過時に速度の低下や顔向きの変化が生じることも考えられる。このような行動を検出することによって、人物の興味対象を推定することが考えられる。本研究では、ステレオ視を用いた位置情報を利用した動き検出と、顔向き推定技術を応用した注目対象推定を行う。

動き検出は、人物の過去の検出位置と現在の検出位置を比較することによって行う。ステレオカメラ校正時に構築した 3 次元座標内で、全フレームの人物の顔位置が (x_0, y_0, z_0) であり、現フレームの顔位置が (x_1, y_1, z_1) であった時、フレーム間での顔位置の移動距離 d は次式で表される。



図 17 人物位置と顔向き変化



図 18 カメラ方向を注目する人物画像

$$d = \sqrt{(x_1 - x_0)^2 + (y_1 - y_0)^2 + (z_1 - z_0)^2} \quad (22)$$

この移動距離 d が小さいと顔位置があまり移動していないことを示しており、視線方向の変化が無い場合には、止まって物体を注目しつづけていると判断できる。また、人物の移動ベクトル $(x_1 - x_0, y_1 - y_0, z_1 - z_0)$ から人物の移動方向が検出できるため、環境情報が既知であれば、人物が何に対して近づいたのかが認識可能である。

高精度な注目対象推定には、目の画像を高解像度に取得する必要がある。そのような画像を用いて、詳細な推定を行えば、視線方向は高精度に推定できる。しかし、目の領域を高解像度に撮影できる場面は実環境では少ないため、応用が限定される。カメラのズーム制御などを行うことによって、対応することも考えられるが、複数人を対象とした場合には、カメラの台数や、制御系の複雑化などの問題がある。本研究では、複数人への対応の容易さと、比較的小さな顔に対しても注目対象を推定することを目的として、顔向き推定を行うことで、視線方向とした。

視線方向と環境情報から注目対象を推定するためには、座標変換や環境情報の入力・更新、人物位置と視線方向の高精度な検出等が必要となる。同一点を注目している場合においても、人物の位置変化に伴い、視線方向は変化する (図 17)。しかし、人物がカメラの方向を注目する場合には、カメラからは人物が撮影領域内のどの場所に存在していても正面方向を向いた顔画像が得られる (図 18)。そこで、カメラを注目対象候補に設置し、正面顔を検出することによって、注目対象候補への注目行動を検出する。正面顔の検出では、四方向面特徴と線形判別分析を適用した手法を用いた。

3.2.2 服装および大きさの推定

人物の服装は、その人物の趣味や嗜好、興味などを表現する人物情報のひとつと考えられる。しかし、服装には、色、形状、柄など多様な側面を持っており、その分類においても困難な問題がある。本研究では、画像から取得可能で、扱いが容易である特徴量として、着用している服の色を用いる。

本研究では、人物検出を顔検出によって行っている。そこで、顔の下部領域は、服の領域であると仮定した。服色を求める領域を胸の高さとし、検出された、顔領域の下端と重心位置から決定される領域を定め、服色の平均値を求めた。画像内での服色算出領域の中心位置の x 座標は、顔の重心座標と同じとした。 y 座標は、検出された顔領域の下端から、重心位置と顔の下端までの距離だけ離れた位置とした。顔の服色の平均値を算出した領域は、図 19 に示す緑色の矩形で示す。領域内の各画素の RGB 値を平均し、検出人物の服色とした。

人物の大きさは、身長、体重のような体型に関する人物情報のひとつである。また、身長は、大人か子供かというような大まかな年齢推定の指標としても考えることができる。身長を計測するた

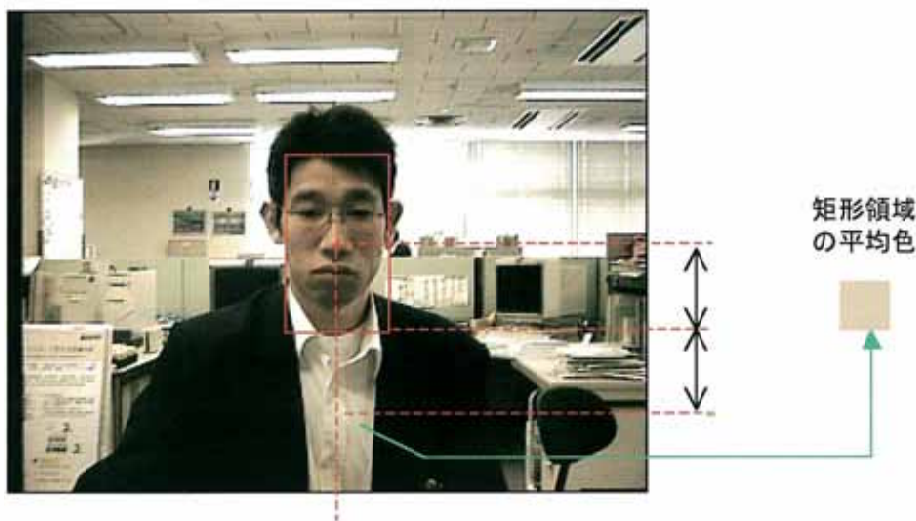


図 19 衣服の色検出

めには、頭頂部を検出し、その高さを求める必要がある。移動する人物に対して正確にこのような情報を得ることは困難であるため、本研究では、ステレオ視によって得られた顔の位置を利用することで、人物の大きさ情報をとした。立っている人物の顔の高さは、身長とは異なるが、人物の大きさを反映した情報といえる。

3.2.3 人数計数

応用システムとして広告に対する注目情報を収集するマーケティング調査を検討している。そのため、人物の注目行動の検出と同時に、撮像領域内の人数計数も重要な要素である。人数計数を高精度に行うためには、正確な追跡が必要である。これまで、複数人の人物追跡を高精度に行う方法として、上方の高い位置にカメラを設置する手法や、多くのカメラを設置することで遮蔽の発生を回避する手法の研究がなされている。しかし、応用システムを検討する場合に、コストや場所、手間などの問題から、多くのカメラを設置することは困難と思われる。そこで、本研究では、単眼カメラを用いて光軸を水平かつ人物の通過方向に対して垂直になるように設置し、撮像領域を通過する人数推定に関する研究を行った。想定している運用形態を図 20 に示す。

本研究では、人物追跡は行わず、前フレームで検出された顔と現フレームで検出された顔の位置を比較することによって人物の移動を検出する。顔位置の移動量が小さい場合には、滞留と判断する。人数の推定には、検出された前顔数と滞留と判断された顔数を用いる。

対象としている環境は、移動方向に制限のある室内空間である。移動人物は、カメラの光軸と直角の方向に移動すると想定している。従って、人物は撮影領域を通過する間に複数フレームに渡って撮影される。

人物の移動速度を一定と仮定した場合、平均通過人数 n は、設定時間内に検出された顔の総数 F_{total} と、滞留と判断された顔数 F_{stop} 、人物が滞留することなく顔探索領域内を通過する際に撮影されるフレーム数の平均 m_{pass_frame} を用いて、次式から計算できる。

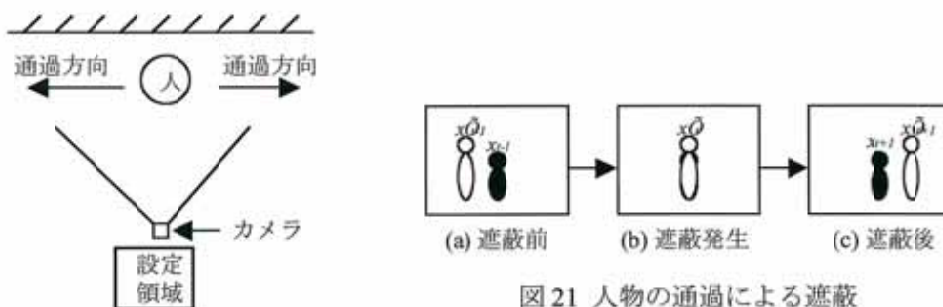


図 21 人物の通過による遮蔽

図 20 想定している運用形態

$$n = \frac{F_{total} - F_{stop}}{m_{pass_frame}} \quad (23)$$

しかし、人物が滞留する場合、滞留の前後で移動速度は減少・増加する。そのため、移動と判断される顔の検出フレーム数は増加する。そこで、通過人数を、滞留することなく通過した人数 n_{pass} と滞留した人数 n_{stop} に分けて考える。

$$n = n_{pass} + n_{stop} \quad (24)$$

滞留者が滞留している間と、移動している間に検出される顔数の和を F'_{stop} とすると n_{pass} は次式で与えられる。

$$n_{pass} = \frac{F_{total} - F'_{stop}}{m_{pass_frame}} \quad (25)$$

$$F'_{stop} = F_{stop} + n_{stop} \times m_{stop_frame} \quad (26)$$

式(2.7)の m_{stop_frame} は滞留を1回行う人物が通過に要するフレーム数から滞留中と判定されたフレーム数を引いた結果の平均であり、式(23)の m_{pass_frame} より大きな値となる。

滞留は、以下の方法から判断する。時刻($t-1$)から、時刻(t)における顔の重心位置の変化量がしきい値以下であるときに、時刻(t)において滞留があったと判断する。この時のしきい値は、実験的に決定した。

本手法では、遮蔽の発生により、滞留人数の計測に誤りが生じ、誤差が蓄積される可能性がある。図21に滞留者(黒)の前を人物(白)が通過し、遮蔽によって滞留人数計測を誤る例を示す。人物が滞留者の前を通過する時、遮蔽が発生し、滞留者の顔は検出されず、通過人物の顔が検出される。検出された通過人物の顔位置 x'_t と前フレームの滞留者の顔位置 x_{t-1} を比較した結果、移動量 $|x_t - x'_{t-1}|$ が移動判定のしきい値よりも大きい場合、滞留は終了したと判断される。その後、人物の通過によって遮蔽されていた滞留者が再度検出されたときには、新たな滞留者と判断される。この結果、継続して滞留する人物が2人と検出され、滞留人数は実際より多く計数される。そこで本稿では、滞留終了時の顔位置 x_{t-1} を記録し、設定時間以内に同じ位置に顔が検出された場合、同一人物が継続して滞留していると判断する。

3.3 応用システム

応用システムとして、人物行動を収集し、マーケティング調査を行うシステムを検討した。システム外観を図22、ブロック線図を図23に示す。カメラで撮像された画像は、画像キャプチャボードを介してカメラPCによって取得される。取得された画像は、カメラPC内で、処理され、人物センシング結果が得られる。各カメラPCと情報統合PCはLAN接続されており、カメラPCで得られた処理結果を情報統合PCに送信される。情報統合PCでは、各カメラPCから送信された結果を統合し、撮像領域で得られた結果を統合・蓄積する。ブロードバンドルータは、各PCをLAN接続するためのハブとしての機能と、センシング結果を外部へ送信するための機能を果たす。



図22 応用システム外観

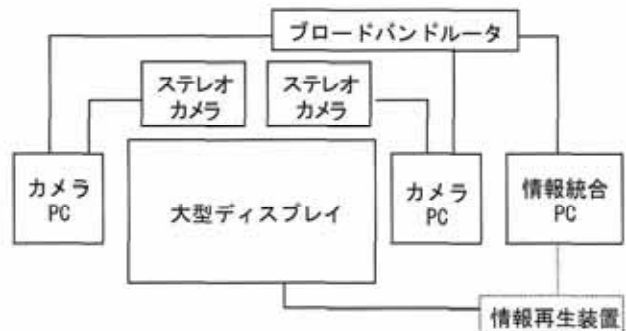


図23 ブロック線図

本システムは、マーケティング調査を目的としているため、情報の訴求力を備えた大型ディスプレイを用いている。提示情報は、静止画・動画などディスプレイに表示できるものであれば、利用できる。本システムでは、大型ディスプレイとして、50インチのプラズマディスプレイパネルを用い、情報提示には、DVD、ハイビジョンデジタルチューナー、PCによるCGを利用した映像などを用いる。情報再生装置で、再生されている情報と、センシング結果として得られた注目情報を、情報統合用PC側で統合することにより、どのような情報を提示しているときに、注目されていたかという情報を得ることができる。駅や百貨店、地下街などの通路状の場所に設置し、広告提示と提示情報に対する人物行動検出を行う。得られた情報を活用することによって、「設置位置に応じた広告料金体系の確立」や、「広告効果の検証」、「試作前製品の興味度調査」などが容易に行える。

4 結果

4.1 正面顔検出

本手法の評価として、撮影領域内で移動する人物の顔検出と正面顔検出実験を行った。システムの運用時には、背景や照明の制御が困難であると考え、室内の乱雑背景中で、通常の蛍光灯照明下において実験を行った。実験に使用した画像は、設定した注目領域を見ながら、撮影領域をランダムに移動する人物を撮影したものである。注目領域は、カメラの直下と、カメラから左右に60cm、1.5m離れた位置とした(図24)。また、カメラに対して横を向き、移動する人物の画像に対しても実験を行った。被験者とカメラとの距離は約1~2.5mであった。被験者は男性5人女性1人の6人である。実験中の表示画面を図25に示す。図25では、左上に基準カメラから得られた画像、左下に基準カメラ画像の2値化画像、右上に記録されている背景画像、右下に検出された顔領域の濃淡画像を表示している。検出された顔領域は、肌色部分はカラー画像から256階調の濃淡値に変換されているが、肌色以外の領域は、白色として処理されている。肌色以外の領域を白色とした理由は、背景の影響を避けるためである。

4200枚の画像に対し、4186枚の画像において正しく顔を検出した。検出率は99.7%であった。検出に失敗した画像は、カメラからはなれた位置で横を向いた画像であり、肌色領域の面積が小さくなったためである(図26)。顔検出を行う際に、顔と判断する肌色面積のしきい値を調整することで、検出失敗を低減することはできるが、しきい値を小さくしすぎた場合には、ノイズや背景の影響から過検出の問題が生じる。適切なしきい値の設定には、応用面からの検討が必要である。ま

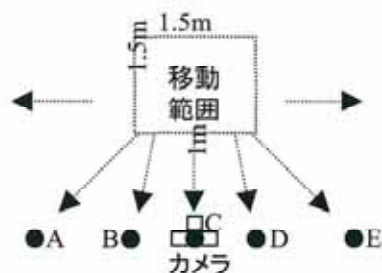


図24 注目対象とカメラの配置図



図25 実験中の表示画像例



図26 顔検出の失敗例(肌色領域が小さい)



図27 顔検出の失敗例(照明の影響)

た、長髪の人物などで、横を向いた場合には、顔の肌色部分のほとんどが毛髪によって隠れてしまう場合があり、本手法では、検出が困難である。このような場合への対応は、今後の課題としたい。

注目推定の評価は、正面顔を正面と認識した認識率と、正面以外の顔を正面と誤認識した誤り率で評価した。結果は、600枚の正面顔画像に対して、483枚の画像で正しく認識した。認識率は80.5%である。また、正面以外の方向を向いた顔を撮影した3586枚の画像において、370枚の画像で正面顔と誤って判別した。誤り率は10.3%である。判別誤りの原因として、人物位置の変化によって照明条件が異なり、顔の検出部位が変動することが挙げられる。図27の例では、照明の影響により、鼻と口の周辺の画素において、肌色の範囲から外れる画素が増加し、顔が分裂した状態で検出されている。そのため、顔の上部のみが検出される結果となり、顔向き推定においても誤りの原因となった。色情報を利用した手法では、このような照明による影響は不可避であるため、肌色以外の情報を用いた高精度な顔検出が今後の課題となる。

4.2 人数計数

本手法の計測精度実験を行った。画像データとして撮影した行動パターンを図28に示す。(a)では、1人の人物がカメラに近い場所と遠い場所を通過し、1往復する。(b)では、フレームインした人物が、撮影領域の中央まで進み、滞留行動をとった後、フレームアウトする動作を2回行う。(c)では、カメラに近い場所と遠い場所を複数人が順次通過する。(d)では、1人の人物が滞留している後方を、複数の人物がすれ違いながら通過する。(e)では、滞留者の前後を複数の人物が通過する。被験者には、撮影領域の端で、フレームアウトすること、滞留時には左右の目標物とカメラの3方向を順番に注目することを指示した。また、図28(c)、(d)、(e)に示す複数人の行動は、5人の被験者が通過、滞留を繰り返して行った。これらの各行動に対して2回の撮影を行った。以降の実験では、得られた画像データに対してオフラインで処理した。

平均通過フレーム数には、全行動パターンから目視によって求められた $m_{pass_frame}=12.5$ 、

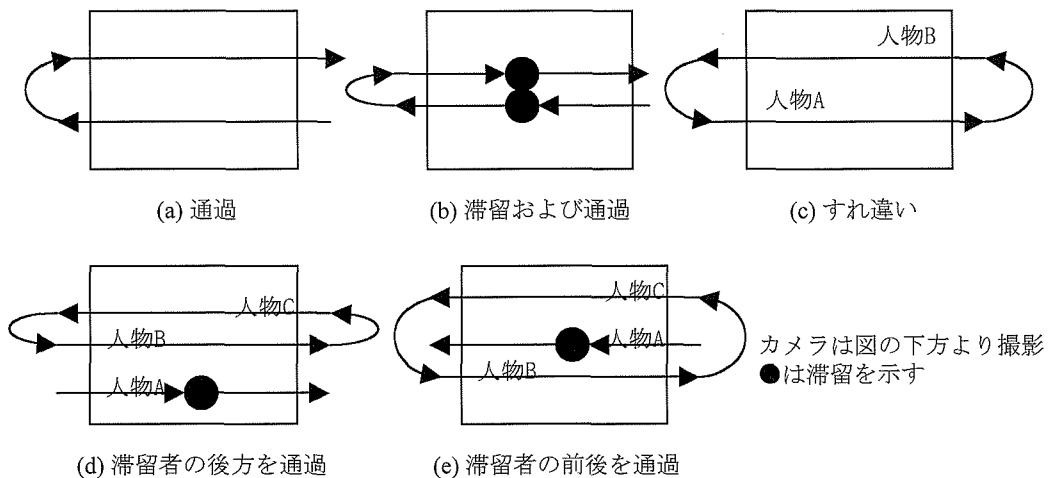


図28 撮影した行動パターン

表2 人数推定結果

行動	(a)-1	(a)-2	(b)-1	(b)-2	(c)-1	(c)-2	(d)-1	(d)-2	(e)-1	(e)-2
顔数	125	121	817	771	204	216	748	734	720	653
滞留顔数	0	0	657	615	0	0	200	209	227	217
滞留人数	0	0	10	10	0	0	5	7	5	5
推定通過人数	10.0	9.7	9.8	9.4	16.3	17.3	42.3	39.9	37.9	33.4
実際の人数	滞留	0	0	10	10	0	0	5	5	5
	通過	10	10	10	10	20	20	43	42	43
誤り率(%)	滞留	0.0	0.0	0.0	0.0	0.0	0.0	40.0	0.0	0.0
	通過	0.0	3.0	2.0	6.0	18.5	13.5	1.6	5.0	13.9

表3 推定結果の比較

行動		(d)-1	(d)-2	(e)-1	(e)-2
実際の人数	滞留	5	5	5	5
	通過	43	42	44	43
滞留者継続判断を行う場合の推定結果	滞留	5	7	5	5
	誤り率 (%)	0.0	40.0	0.0	0.0
	通過	42.3	39.9	37.9	33.4
	誤り率 (%)	1.6	5.0	13.9	22.3
滞留者継続判断を行わない場合の推定結果	滞留	5	8	11	9
	誤り率 (%)	0.0	60.0	120.0	80.0
	通過	42.3	39.6	36.6	32.1
	誤り率 (%)	1.6	5.7	16.8	25.3

$m_{stop_frame}=16.3$ を用いる。また、遮蔽前後で滞留の継続判断を行う、滞留終了時に顔の位置を記録する時間は3フレームとした。人数推定の結果を表2に示す。表2において、(a)-1は(a)通過の画像データ1を示す。通過人数は、滞留せず通過した人物と滞留した人物の和である。

(a)通過と(b)滞留および通過、(d)滞留者の後方を通過の画像データ1に関しては、誤り率は6%以下であり、ほぼ正しい結果が得られた。しかし、遮蔽の発生する(c)すれ違い、(d)滞留者の後方を通過の画像データ2、(e)滞留者の前後を通過に関しては、実際よりもやや少ない人数推定結果となった。この原因には、遮蔽によって検出される顔数が減少したこと、人物の位置によって照明条件が悪くなり、顔の検出に失敗したことが挙げられる。また、カメラと人物間の距離によって撮影領域となる歩行距離が異なることや、画像データ撮影時に、歩行速度が速まったことの影響も考えられる。

遮蔽前後での滞留者の継続判断の効果を確認するため、滞留終了時に顔の位置を記録せず実験を行った。実験は、(d)と(e)の画像データに対して行った。結果を表3に示す。滞留者の継続判断を行わない場合、滞留人数が増加し、通過人数は減少した。特に滞留者の前を横切る通過者が存在した場合に顕著であった。これらの結果より、滞留者の継続判断手法の効果があることが確認された。

4.3 応用システム評価

応用システム実装時に実験を行ったところ、人物がカメラから離れた位置に存在する場合、人物の顔領域は低解像度取得されるため、正面顔検出の精度が低下することが確認された。被験者によって検出率は異なるが、分割した各画面への正面顔検出実験結果は、カメラと人物の距離が2.7m以内であると平均で90%以上であった。正面顔検出は、システム評価の重要な部分であるため、この範囲を広げることが高性能化に繋がると思われる。

人物の注目対象推定技術や属性推定技術は、マーケティング調査以外にも、セキュリティ、介護等の分野での適用が考えられる。現在のセキュリティカメラは、「常に録画し、緊急事態発生時にその瞬間の映像を確認する」というような使用方法が主であるため、非効率な面がある。これらの技術が撮影された人物の行動に応じた記録方式の変更や、警告の制御などを行うことによって、そのような作業の効率化が期待できる。また、人物の注目対象から興味を推測し、機器制御を行うことで、介護作業の一部を自動化することも考えられる。問題点としては、色情報に基づいた人物検出手法を利用しているため、環境や照明の変化に対応できない場合がある。実環境で利用するためには、パターン等の情報を積極的に利用し、頑健性と精度向上を行う必要がある。

フェーズ III

今後の取り組み

これまでの研究において、顔向き推定技術とその応用システムの有効性が示された。今後の取り組みとしては、顔部品等の位置関係を応用したより高精度な顔向き推定や瞳画像を用いた視線検出等による推定技術の高度化が有効と考えられる。また、応用システムを考えた場合、照明や背景に対する頑健性が必要となるため、肌色情報と画像パターン等を統合した顔検出技術等が重要である。さらに、マーケティング調査のような応用では、人物の年代や性別等の属性を推定する技術の付加や、追跡による人物行動のより深い解析等がシステムの有効性を高めると考えられる。