

## 研究テーマ 顔部品検出と注目推定に関する研究

研究者	岩田健司	財団法人ソフトピアジャパン	雇用研究員
	伊藤英則	名古屋工業大学	兼業研究員
	石井洋平	財団法人ソフトピアジャパン	雇用研究員

### フェーズ I

この研究は、研究テーマ「視線検出に関する研究」を継続しており、フェーズ II より実施された。フェーズ I での研究課題として、頭部の動きに対する顔画像拡大画像獲得のためのカメラ制御、顔部品検出の顔向き、個人差、表情、環境等に対するロバスト性の向上が挙げられた。

### フェーズ II

#### 1 研究の概要

フェーズ II では、フェーズ I での課題に対処するため、顔向き、個人差、表情、環境等に対しロバストな顔部品検出として、四方向面特徴と弛緩整合法を用いた手法を構築した。また、注目推定のためには、人物の顔画像をパン・チルト・ズームカメラ (PTZ カメラ) で拡大取得する。この応用展開として、商品開発等に有用な情報を自動的に取得する、自動マーケティング情報システム AMIS (Automatic Marketing Information System) の構築を行った。ここでは、ロバストな顔部品検出手法と、注目推定による自動マーケティング情報システムの 2 つに分けて報告する。

#### 2 ロバストな顔部品検出

##### 2.1 研究の目的

ユーザインタフェースやセキュリティシステムなどを目的として、顔画像認識の研究が盛んに行われている。顔の目・口・鼻などの器官部分の位置情報は、人物認識や顔向き推定に有用な情報となるため、画像中から適格に抽出する必要がある。ここで従来研究の多くが正面付近の顔画像を対象としており、人物の前にカメラを設置している。しかし監視カメラシステムなどでは、正面を向いた顔が得られない場合がある。屋内の監視カメラは天井に設置する 경우가多く、人物の顔は斜め上から撮影されることになる。また顔がカメラに対して正面を向くアプリケーションと異なり、人物はカメラを意識しておらず、様々な顔向きの画像が撮影される。このような状況に対しロバストかつリアルタイムな顔画像処理手法が必要となる。

顔向きの変化を考慮した顔部品検出手法は多く提案されているが、テンプレートや特徴量を用いて候補を検出し、配置情報から正解を絞り込む方法が一般的である。候補検出にテンプレートマッチングを用いて、幾何学的配置情報を組み合わせている手法、テンプレートマッチングに四方向面特徴と色情報を組み合わせ、ロバスト性を高める手法が提案されている。分離度フィルタや円のハフ変換を用いる手法は、円形状の目や鼻孔、端点を検出する。正面付近を向いた顔であれば個人差によらず、安定して顔部品位置を推定できる。また、分離度フィルタと部分パターンの照合、位置情報を組み合わせる方法が提案されている。ただしこれらの方法では、大きな顔向きの変化に対応できない。特徴量のパターン照合と、特徴点間を結ぶグラフの対応コストを最小化する方法では、1枚の顔画像からの対応付けが可能であるが、最小化問題を高速に探索する方法が問題となる。平均モデルを使用し、ノードの近傍を活性化し探索する方法が提案されている。多様な顔向きに対応させる方法として、顔向き別のグラフを持つ方法や、3D 顔モデルに当てはめる方法などがある。このような方法では探索範囲が広く処理時間がかかるため、動画画像で探索範囲を狭めるなどの工夫が必要となっていた。

そこで、顔向きや個人差にロバストで高速な顔部品検出手法を構築した。本手法は、多人数・多方向の顔画像から生成した四方向面特徴と、柔軟に対応付ける弛緩整合法を用いる。この弛

緩整合法は、顔部品間のスプリング接続を考慮した顔部品モデルを使用し、顔向き・個人差等の変動に対し柔軟に対応づけることができる。

## 2.2 実施内容

顔部品検出手順について以下に述べる。まずカラー画像から肌色を検出し、顔領域を求める。次に、多人数・多方向の顔画像から生成した四方向面特徴の顔部品テンプレートを用いて、初期の対応確率を求める。最後に、顔部品位置モデルを用いて、各顔部品の関係に着目し対応確率を更新する弛緩演算により、顔部品位置を決める。さらに単一モデルでは対応づけない変化には、複数の顔部品モデルを併用して対処する。本手法は複数モデルを統合した弛緩演算を行うことが可能である。

多数の点群やセグメント群同士を対応づけるような弛緩法による画像マッチングは計算コストが高いが、両目、鼻、口の4点の対応問題への応用で、計算コストを抑えた本手法は弛緩演算の収束が早く、近年のコンピュータの高速化と合わせ、リアルタイムシステムでの利用も可能になる。

### 2.2.1 肌色による顔領域検出

はじめに、肌色検出を用いて顔領域を切り出す。今回は、背景や照明環境に比較的ロバストな肌色基準値法を用いる。まず、図 1(a)のような入力画像を CIE- $L^*u^*v^*$ 表色系に変換し、図 1(b)のように  $u^*, v^*$ 値の2次元ヒストグラムを作成する。あらかじめ決めておいた肌色有効範囲内のピークを肌色基準  $u^*, v^*$ 値とする。次に肌色基準  $u^*, v^*$ 値と入力画像の各画素との距離値画像を、判別分析法により2値化し、肌色を検出する。肌色検出例を図 1(c)に示す。この領域に雑音除去を施し、面積が最大となる領域を顔領域とする。図 1(a)から得られる顔領域を図 1(d)の矩形で示す。

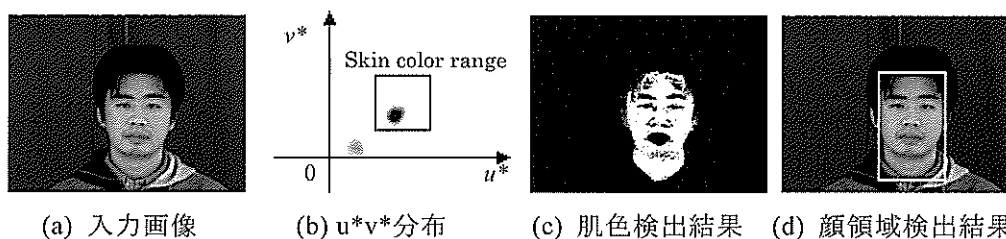


図 1 肌色による顔領域検出

### 2.2.2 四方向面特徴

個人差や、顔向き変化に対応できる特徴量として、四方向面特徴を用いる。四方向面特徴は文字認識、人物認識など、パターン認識の分野で広く使われている特徴量である。

#### 2.2.2.1 四方向面特徴の作成

入力画像から方向検出フィルタにより、水平・右上がり・垂直・右下がりの各方向の4面を作成する。これら4面をガウシアンフィルタによりぼかし、低解像度化したものが四方向面特徴である。図 2(a)の顔画像の四方向面特徴は図 2(b)のようになる。入力画像における顔領域の、位置  $x_k, y_k$  で切り出した四方向面特徴を、今後  $I_k$  と表す。図 3(a)の顔画像で、白い矩形で示されている、右目・左目・口・鼻の各顔部品における四方向面特徴は、図 3(b)のようになる。今回は、入力解像度の 1/8 に低解像度化した。

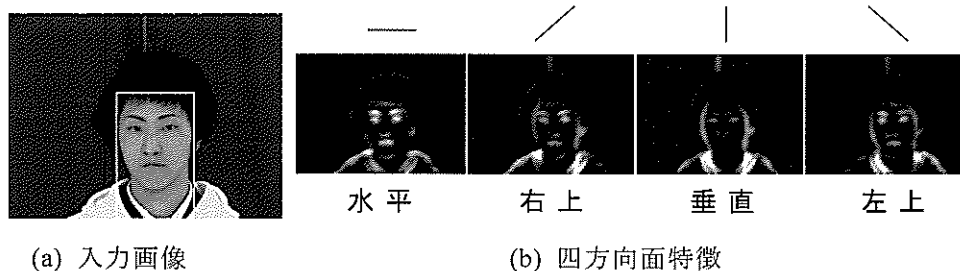


図 2 四方向面特徴

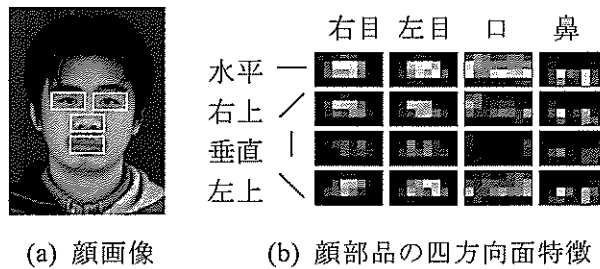


図3 顔部品の四方向面特徴

2.2.2.2 テンプレートの学習

四方向面特徴は、ガウシアンフィルタと低解像度化の効果により、パターンの微細な位置ずれなどの変動を吸収する。さらに学習データに顔向き変化や個人差等の、入力画像において想定される変動をあらかじめ加えることで、これらの変動に対応できる特徴量となる。

各個人により異なる顔部品形状の違いを吸収し、様々な人物に対応できるテンプレートとするため、図4(a)に示すような複数の人物から特徴量を学習する。10名の学習データから作成した四方向面特徴テンプレートを図4(b)に示す。

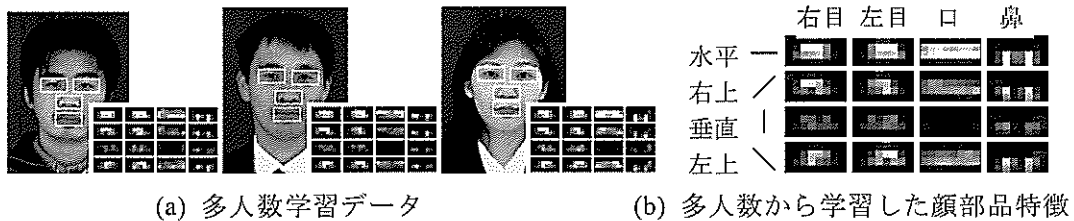


図4 多人数学習

また、図5(a)に示すような、多方向の顔画像から部品を切り出し、四方向面特徴を抽出することで、顔向きの変化に対応した四方向面特徴を学習する。学習データの四方向面特徴を平均したものを、四方向面特徴テンプレート $T_i$ とする。 $i=1, \dots, 4$ であり、それぞれ右目・左目・口・鼻に相当する。ここで、学習データの顔部品位置はあらかじめ与えておく。図5(b)は、多人数、多方向の学習データより生成された四方向面特徴テンプレートである。各顔部品の特徴解像度は $8 \times 4$ とした。

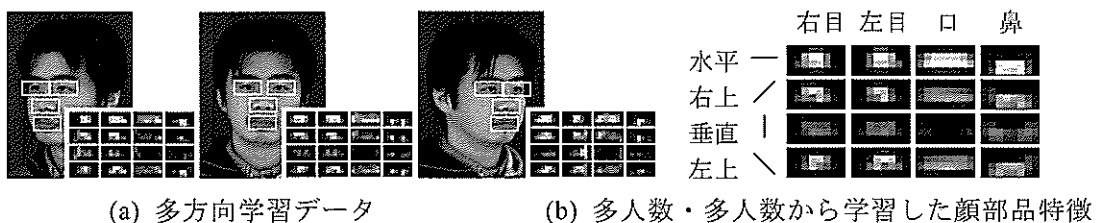


図5 多方向学習

2.2.2.3 テンプレートマッチング

四方向面特徴による顔部品テンプレートによって、テンプレートマッチングを行う。しかしテンプレートマッチングのみでは、似通った特徴が多く存在するため、一意に顔部品位置を特定することはできない。図6(a)の顔領域に対しテンプレートマッチングを行った時の類似度の分布を図6(b)に示す。白い部分は類似度が高く、黒い部分は低いことを表している。正解位置に対し高い類似度が検出されているが、それ以外にも多く類似度の高い部分が存在している。よって、顔部品の位置関係を用いて、正しい位置を得る必要がある。

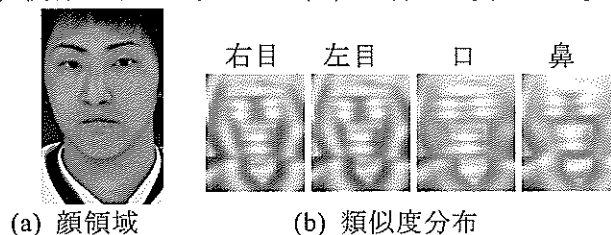


図6 テンプレートマッチングの結果

### 2.2.3 弛緩整合法

数学的手法であった弛緩法は、Rosenfeldらにより、パターン認識に応用された。これは、近傍により対応確率を更新し、線画の対応付けを行うもので、文字を多角形近似で表現する日本語漢字認識に適応されている。弛緩法を対応付け問題に適応することから、弛緩整合法(Relaxation Matching)と呼ぶ。また弛緩法は幅広く応用されており、テンプレートマッチングと弛緩法を組み合わせた、オプティカルフロー検出などが提案されている。

今回は弛緩法を顔部品検出に適用する。入力顔画像に対し各顔部品の対応確率を定義し、位置関係から確率更新する弛緩演算により、顔部品位置を特定する。

#### 2.2.3.1 初期対応確率

色情報を用いて得られた顔領域に対し、四方向面特徴によるテンプレートマッチングを行い、その類似度を初期対応度とする。 $t=0$ の時の初期対応度 $p_{ik}^0$ は(1)式のように表される。

$$p_{ik}^0 = \frac{\sum_n T_i(n) \cdot I_k(n)}{\sqrt{\sum_n \{T_i(n)\}^2} \sqrt{\sum_n \{I_k(n)\}^2}} \quad (1)$$

各顔部品 $i$ における初期対応度 $p_{ik}^0$ の最大値で正規化し、初期対応確率 $P_{ik}^0$ を得る。

$$P_{ik}^0 = \frac{p_{ik}^0}{\max_k p_{ik}^0} \quad (2)$$

#### 2.2.3.2 顔部品位置モデル

顔部品間の配置情報を記述する、顔部品位置モデルを使用する。カメラで撮影された画像中の顔部品の配置は、個人差、顔向きなどにより様々に変化するため、図7に示すようなスプリング接続という概念を適用する。この概念は、各顔部品間をばねで接続しており、各顔部品の位置が柔軟に変化しつつ、モデルの配置に近づこうとする力が働くと考える。このような変動は、2.2.3.3節で述べる近傍対応確率により実現する。これにより、顔向き、個人差等に柔軟に対応し、顔部品位置を特定する。また、顔部品の相対的な位置関係に着目しているため、全体の平行移動に対してロバストである。

顔部品位置モデルの生成方法を以下に述べる。学習データにおいてあらかじめ与えられた顔部品の位置から、標準的な顔部品の配置情報が得られる。そこで、学習データに対する平均位置を、顔部品位置モデルとする。

正面顔のスプリング接続モデルの対応する範囲を超えた顔向き変化がある場合には、顔部品の検出が困難になる。そのような場合、複数の顔部品位置モデルを同時に使用することが有効となる。しかし、複数のモデルを使用する場合、モデル間の評価の正規化や、計算コストの増大などの問題がある。本手法は、容易に複数モデルを統合でき、かつ計算コストを抑えることができる拡張性を備えている。例えば、斜め方向を向いた顔画像に対する検出性能が良くない場合、図8に示すように正面に加えて、右上、左上、右下、左下の5つのモデルを同時に使用し、性能を改善することができる。これを多方向統合モデルと呼ぶことにする。2.2.3.3節の周辺適合確率において後述するが、一度の検出過程で最も適合したモデルを自動的に用いることで、モデルの数だけ検出処理を繰り返す必要はなくなる。

複数モデルの生成には、各方向の顔向きごとに平均し、顔部品位置モデルとする。モデル数を $M$ 、各モデルを $m=1, \dots, M$ で表し、顔部品 $i$ におけるモデル座標値を $x_{mi}, y_{mi}$ と表すことにする。

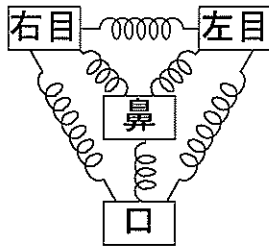


図7 顔部品間のスプリング接続

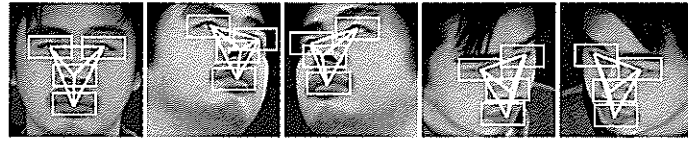


図8 多方向統合モデル

2.2.3.3 近傍対応確率と周辺適合確率

前節のモデルを使用し、柔軟に対応付ける具体的な方法を述べる。顔部品  $i$  における、周辺の顔部品を  $j=1, \dots, 4$  (ただし  $j \neq i$ ) とする。周辺の顔部品  $j$  に対し、入力顔領域に相当する位置を  $l$  とする。モデル  $m$  における  $l$  の相対位置  $x_{ml}, y_{ml}$  は、

$$\begin{aligned} x_{ml} &= x_k + (x_{mj} - x_{mi}) \\ y_{ml} &= y_k + (y_{mj} - y_{mi}) \end{aligned} \tag{3}$$

となる。入力画像の顔部品の相対位置は、モデルを中心としたガウス分布になると仮定し、近傍対応確率を導入する。入力顔領域での位置  $l$  における周辺の対応確率に、ガウス分布を掛け最大値を計算する。これを近傍対応確率  $Q'_{jl}$  とする。

$$Q'_{jl} = \max_l \left\{ P'_{jl} \cdot \exp\left(-\frac{d_{ll'}^2}{2\sigma^2}\right) \right\} \tag{4}$$

$l'$  は入力顔領域の全体で、その位置を  $x_{l'}, y_{l'}$  とする。 $\sigma$  はガウス分布の標準偏差であり、4.2 節で述べた顔部品のスプリング接続における、ばねの強さに相当する。 $d_{ll'}$  は、 $l, l'$  間の距離であり、(5)式のように表される。

$$d_{ll'} = \sqrt{(x_{ml} - x_{l'})^2 + (y_{ml} - y_{l'})^2} \tag{5}$$

なお、相対位置  $x_{ml}, y_{ml}$  が顔領域の外的場合、 $Q'_{jl} = 0$  とする。すべての周辺顔部品  $j$  について、近傍対応確率  $Q'_{jl}$  を掛け、各モデル  $m$  での最大値を、周辺適合確率  $q'_{ik}$  とする。

$$q'_{ik} = \max_m \left( \prod_j Q'_{jl} \right) \tag{6}$$

各モデルの最大値を用いることで、最も適合したモデルが採用されることになる。

図 6(a)の顔画像の初期対応に対する初期対応確率  $P'_{ik}$ 、近傍対応確率  $Q'_{jl}$  および周辺適合確率  $q'_{ik}$  を図 9 に示す。近傍対応確率は対応確率の極大値をガウス分布で広げたものになる。周辺適合確率は、他部品の近傍対応確率を、重ねるようなイメージになり、他部品の位置関係の正しい部分が浮かび上がってくる。周辺部の黒い部分は、相対位置が顔領域の範囲外で、適合確率 0 となっている。

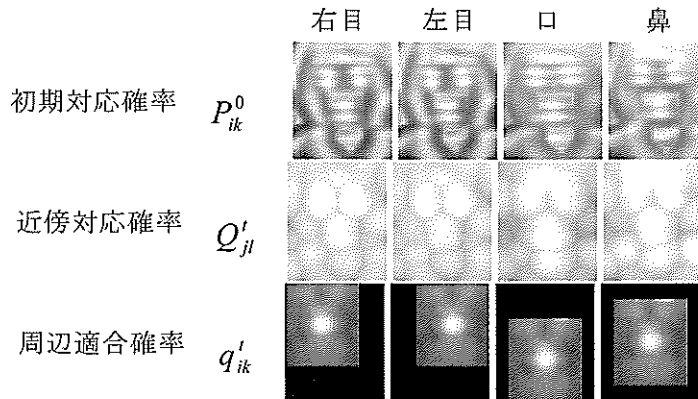


図9 初期対応確率、近傍対応確率、周辺適合確率

### 2.2.3.4 弛緩演算

顔部品位置の適合性を示す対応確率  $p_{ik}^t$  を、複数回の弛緩演算により更新することで、最適な対応位置を導き出す。ここで、 $t$  は弛緩演算の回数を示し、 $t=0, \dots, \tau$  とする。 $\tau$  は弛緩演算の繰り返し回数である。今回は  $\tau=4$  とした。

周辺の顔部品との関係を表す周辺適合確率  $q_{ik}^{t-1}$  を用いて、対応度  $p_{ik}^t$  を更新する。

$$p_{ik}^t = q_{ik}^{t-1} \cdot P_{ik}^{t-1} \cdot p_{ik}^0 \quad (7)$$

各顔部品  $i$  における対応度  $p_{ik}^t$  の最大値で正規化し、対応確率  $P_{ik}^t$  を得る。

$$P_{ik}^t = \frac{p_{ik}^t}{\max_k p_{ik}^t} \quad (8)$$

(7)式(8)式を  $\tau$  回繰り返すことにより、正しい顔部品の位置が検出できる。 $P_{ik}^t=1$  となる  $k$  を顔部品  $i$  の位置として決定する。

図 6(a)の顔画像に対する、弛緩演算による対応度  $p_{ik}^t$  の変化過程を図 10 に示す。正しい位置の対応度が強調され、異なる位置での対応度が抑制されていく様子がわかる。この結果から得られた顔部品位置を図 11 に示す。

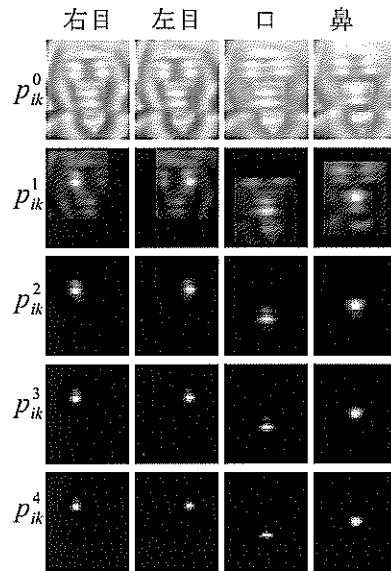


図 10 弛緩演算

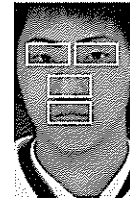


図 11 顔部品検出結果

### 2.2.3.5 計算コストの削減

周辺適合確率  $q_{ik}^t$  の計算は、コーディング上それほど計算時間を要さないことを示す。(4)式を(5)式で展開すると、

$$\begin{aligned} Q_{jl}^t &= \max_{j'} \left\{ P_{j'l}^t \cdot \exp\left(-\frac{d_{j'l}^2}{2\sigma^2}\right) \right\} \\ &= \max_{x_r, y_r} \left\{ P_{j'l}^t \cdot \exp\left(-\frac{(x_l - x_r)^2 + (y_l - y_r)^2}{2\sigma^2}\right) \right\} \\ &= \max_{x_r, y_r} \left\{ P_{j'l}^t \cdot \exp\left(-\frac{(x_l - x_r)^2}{2\sigma^2}\right) \cdot \exp\left(-\frac{(y_l - y_r)^2}{2\sigma^2}\right) \right\} \end{aligned} \quad (9)$$

となる。ここで、ガウス関数と対応確率  $P_{j'l}^t$  は常に正の値を持つことから、

$$Q_{jl}^t = \max_{y_r} \left[ \exp\left(-\frac{(y_l - y_r)^2}{2\sigma^2}\right) \cdot \max_{x_r} \left\{ P_{j'l}^t \cdot \exp\left(-\frac{(x_l - x_r)^2}{2\sigma^2}\right) \right\} \right] \quad (10)$$

となり、繰り返し演算2回で計算できる。またガウス分布は周辺に行くに従い減少する関数で、その積の最大値を求めることから、離れた位置の計算は必要ない。そこで、

$$\begin{cases} -c\sigma < x_l - x_r < c\sigma \\ -c\sigma < y_l - y_r < c\sigma \end{cases} \quad (11)$$

の範囲内で計算を行う。今回は $c=1$ とした。

### 2.2.4 複数カメラ情報統合

複数のカメラで人物の顔情報を取得する方法としては、ステレオカメラで3Dモデルに対応付ける手法や、光軸が内向きのカメラで、顔全体のパターン変化を用いる手法などがある。2.2.3節で述べた顔部品検出法では、広範囲なカメラアングル画像から顔部品位置を特定できることから、異なるカメラで同一人物を撮影し、これらを相互に用いることで、3次元的な位置情報を考慮した顔部品検出が可能となる。

#### 2.2.4.1 エピポーラ線

図12に示すように、カメラ中心と注目点を結ぶ直線は、もう一方のカメラの画像中に、エピポーラ線として投影される。2つのカメラ間の関係が既知の場合、片方の画像中での1点が定まると、もう一方の画像でのエピポーラ線が決まる。画像面1での点 $k$ に対応する、画像面2の点 $k'$ はエピポーラ線上に限定され、1次元的に探索できる。このような性質はステレオマッチングや3Dモデリング<sup>(17)</sup>など幅広く適用されるが、今回は顔部品検出に応用する。

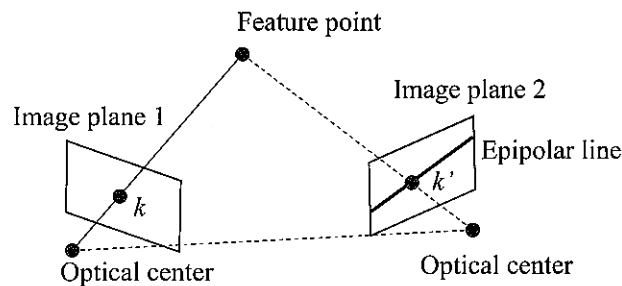


図12 エピポーラ線

PTZカメラを2台使用する。PTZカメラの位置は固定で、パンチルト角や、画角はPCで制御するため既知である。ただしPTZカメラでは、駆動部分やズームレンズの歪みなど、精度が得られない要素が多い。

#### 2.2.4.2 画像間の弛緩演算

各画像で、四方向面特徴のテンプレートマッチングを行い、初期対応を求める。また、顔部品位置モデルを用いた周辺適合確率 $q'_{ik}$ の算出も、単一カメラの場合と同様に求める。

点 $k$ における、もう一方の画像でのエピポーラ線上の点を $k'$ とし、その顔部品 $i$ における対応確率を $P'_{ik'}$ とする。エピポーラ線上の対応確率 $P'_{ik'}$ の最大値を、画像間適合確率 $R'_{ik}$ とする。

$$R'_{ik} = \max_{k'} P'_{ik'} \quad (12)$$

対応度 $p'_{ik}$ を更新式(7)を、画像間適合確率 $R'_{ik}$ を用いて式(13)のように置き換える。

$$p'_{ik} = q'^{-1}_{ik} \cdot P'^{-1}_{ik} \cdot R'^{-1}_{ik} \cdot p^0_{ik} \quad (13)$$

単一カメラの場合と同様に正規化し、対応確率を求め、繰り返し演算を行う。これにより、カメラ間の関係を考慮した顔部品位置を検出できる。

### 2.2.5 接続選択と連続画像による遮蔽対応

図13は、意図的に左目を遮蔽した例である。2.2.3節の従来法を用いた場合が図13(a)である。遮蔽された左目は当然検出できないが、露出している右目・鼻・口の位置も誤った位置にずれて検出されている。このような場面以外にも、顔向きによる自己遮蔽や、表情変化やメガネなど部品形状の変化によるマッチング類似度の低下などによって、同様の現象が発生する。

これは、すべての顔部品間の位置関係を評価するため、欠損部分の低い対応確率が弛緩演算

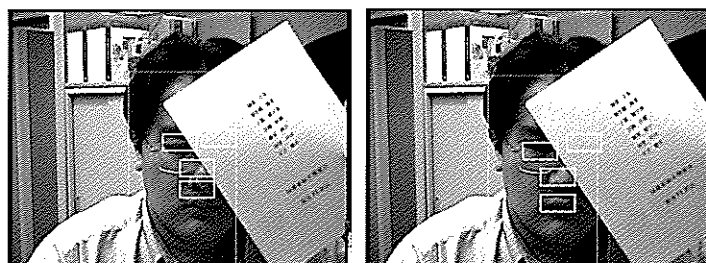
により、すべての顔部品の対応確率を下げってしまうことが原因である。図 13(a)では、左目の低い対応確率により、右目・鼻・口の対応確率が低下し、検出失敗している。

そこで、顔部品間の関係を記述する近傍対応確率を動的に選択し、対応関係の確かな部分のみを使用する方法として、部品間の接続選択を導入する。

さらに動画像を考慮した場合、正しく検出できるフレームの情報を不安定なフレームに継承することができれば、このような遮蔽問題にも有効である。ただし動画像を用いて顔特徴点をトラッキングする方法は、あらかじめ与えられた特徴点をテンプレートマッチングで追跡する方法や、オプティカルフローから追跡する方法などがあるが、誤差が蓄積するように追跡した位置がずれるという問題がある。静止画に適用される弛緩整合法を、動画像に拡張する方法は、前後のフレーム情報を弛緩演算の項に追加することで実現する。

トラッキングとして単純に前フレームの情報を次のフレームに用いると、前フレームで誤検出した場合に誤った情報がそのまま引き継がれてしまう。そこで、前後のフレームを使用し、どちらかのフレーム情報を選択的に利用する。

また、これに前述の部品間選択を組み合わせることにより、動画でのトラッキングの失敗や、一部部品欠損による顔部品間対応の失敗を、動的に選択し、補完することができる方法を述べる。



(a) 従来法 (b) 部品間接続選択

図 13 一部遮蔽のある顔画像に対する検出例

### 2.2.5.1 部品間の接続選択

形状変化や隠蔽等により、一部の顔部品の対応確率が低い場合、他の顔部品の周辺適合確率も低くなる。これは、(6)式のように周辺適合確率は、全顔部品の近傍対応確率の積で計算されているからである。

そこで、近傍対応確率の低い顔部品を周辺適合確率の計算から除外する。近傍対応確率  $Q_{ji}(t)$  を周辺顔部品  $j$  について降順にソートし、 $R_{ik(j')}(t)$  とする。

$$R_{ik(j')}(t) = \{Q_{ji}(t)\}_{j=1, \dots, 4(j \neq i)} \quad (14)$$

$$R_{ik(j_1)}(t) \geq R_{ik(j_2)}(t) \quad (\text{if } j'_1 \leq j'_2) \quad (15)$$

このうち上位の近傍対応確率を用いることで、一部の欠落の影響を軽減した周辺適合確率  $q_{ik}^E(t)$  を算出できる。これを部品選択周辺適合確率と呼ぶことにする。

$$q_{ik}^E(t) = \prod_{j'=1}^{j'_{\max}} R_{ik(j')}(t) \quad (15)$$

ここで  $j'_{\max}$  は対応確率更新に選択的に用いる部品数であり、 $1 \leq j'_{\max} < j_{\max}$  になる。この周辺適合確率  $q_{ik}^E(t)$  を用いて、(7)式(8)式と同等に、対応確率を更新する。図 6 の意図的に左目を隠蔽した例に対し、部品選択周辺適合確率  $q_{ik}^E(t)$  を用いた場合が図 13(b)である。 $j'_{\max} = 2$  として、顔部品間の対応のうち 1 つを棄却している。従来法と異なり、左目の影響を受けず、右目・鼻・口の正しい位置を検出できている。

### 2.2.5.2 弛緩整合法の連続画像への拡張

一枚の静止画から顔部品位置を抽出する弛緩整合法に、連続画像の項を追加することで、動画像に拡張する。動画像のフレームレートが十分に高い場合、連続する画像間の相関は極めて高い。これを顔部品検出で考えると、あるフレームで検出される顔部品位置は、前後のフレー



ムでの顔部品位置の近い位置に存在する。このことを弛緩演算に応用する。

あるフレームの時間を  $s$  で表し、その時の対応度を  $p_{ik}(t,s)$ 、対応確率を  $P_{ik}(t,s)$ 、近傍対応確率を  $Q_{jl}(t,s)$  とする。あるフレームでの近傍での対応確率を表現するため、フレーム間近傍対応確率を  $Q_{ik}^F(t,s)$  とする。

$$Q_{ik}^F(t,s) = \max_{k'} \left\{ P_{ik'}(t,s) \cdot \exp \left( -\frac{d_{kk'}^2}{2\sigma_F^2} \right) \right\} \quad (16)$$

この式は、(4)式の近傍対応確率の計算式と同等であるが、フレーム間の関係を記述するものであり、顔部品間の関係を表現する(4)式とは性質が異なり、標準偏差  $\sigma$  の最適値は異なると考えられる。

ここから、1つ前のフレームでの検出情報  $Q_{ik}^F(t,s-1)$  を周辺適合確率  $q_{ik}^{F1}(t,s)$  に反映させることができる。

$$q_{ik}^{F1}(t,s) = Q_{ik}^F(t,s-1) \cdot \prod_j Q_{jl}(t,s) \quad (17)$$

この周辺適合確率  $q_{ik}^{F1}(t,s)$  を用いて、(7)式(8)式と同等に対応確率を更新すると、前のフレームでの情報を反映して顔部品を検出できる。またこの弛緩演算を  $\tau$  回を繰り返すことにより、 $\tau$  フレーム前までの情報が反映されることになる。

### 2.2.5.3 連続画像の選択

(14)式は、実際の動画像にそのまま用いるには問題がある。前フレームでの位置情報  $Q_{ik}^F(t,s)$  が常に正しいことが前提となっているが、実際の動画ではそのような保証はない。フレームインや顔向き変化などにより、顔部品を検出できないフレームが発生する。

1つ前のフレームと、1つ後のフレームでは、フレーム間の時間が等しければ、等価の情報を持つと考えられる。また、あるフレームを境に検出成功と失敗の区切りが発生すると考えると、前フレーム、後フレームの検出に成功している一方の情報を利用すれば前述の問題を解決できると考えられる。

そこで、フレーム選択近傍対応確率  $Q_{ik}^E(t,s)$  を、以下のように定義する。

$$Q_{ik}^E(t,s) = \max \{ Q_{ik}^F(t,s-1), Q_{ik}^F(t,s+1) \} \quad (18)$$

これを用いてフレーム選択周辺適合確率  $q_{ik}^{F2}(t,s)$  を以下のように定義する。

$$q_{ik}^{F2}(t,s) = Q_{ik}^E(t,s) \cdot \prod_j Q_{jl}(t,s) \quad (19)$$

同様に (7)式(8)式により対応確率を更新する。弛緩演算を  $\tau$  回繰り返すことにより、 $\tau$  フレーム前後の情報が反映されることになり、画像の入力時から検出結果の出力まで  $\tau$  フレームの遅れが発生することになる。

### 2.2.5.4 部品及び連続画像の接続選択

3.3 節の連続画像情報の選択に、さらに 3.1 節の部品間情報の選択を組み合わせることができる。顔部品間の近傍対応確率  $Q_{jl}(t,s)$  と選択的フレーム間近傍対応確率  $Q_{ik}^E(t,s)$  を同時に降順ソートし、 $R_{ik(j')}^F(t,s)$  とする。

$$R_{ik(j')}^F(t,s) = \{ Q_{ik}^E(t,s), \{ Q_{jl}(t,s) \}_{j=1, \dots, 4(j \neq i)} \} \quad (20)$$

$$R_{ik(j'_1)}^F(t,s) \geq R_{ik(j'_2)}^F(t,s) \quad (\text{if } j'_1 \leq j'_2) \quad (21)$$

これを用いて周辺適合確率を求め、部品フレーム選択近傍対応確率  $q_{ik}^{FE}(t,s)$  とする。

$$q_{ik}^{FE}(t,s) = \prod_{j'=1}^{j'_{\max}} R_{ik(j')}^F(t,s) \quad (22)$$

これを用いて弛緩演算を行い、対応確率を更新する。近傍対応確率のうち、下位を棄却するように  $j'_{\max}$  を設定すると、一部が遮蔽した場合は、自動的に遮蔽部品を棄却し、現れている部

品と前後フレームの情報を用いる。人物が大きく移動している場合など、前後フレームの情報が信頼できないケースでは、顔部品間の情報のみを自動的に用いることになる。

## 2.3 結果

### 2.3.1 実験データ

実験データとして、HOIP 顔画像データベースを用いる。このデータベースは 15 歳～64 歳迄を年代別・性別を均等な人数に振分け、計 300 名を収集したものである。多方向データベースは顔向きについては、 $15^\circ$  間隔で、縦 $\pm 45^\circ$ 、横 $360^\circ$  全周の人物頭部画像を収録している。回転データベースは顔向きについては、 $5^\circ$  間隔で、縦 $\pm 45^\circ$ 、横 $360^\circ$  全周のを収録している。詳細は研究テーマ「顔画像データベースの構築」において報告している。実験には、両目、口、鼻が画像内に収まっている範囲内のデータを使用する。多方向データベースより抽出した、10 歳代、20 歳代、30 歳代、40 歳代、50 歳代の男性、女性を各 1 名、計 10 名を学習人物とし、それ以外を未知人物とする。図 14 に学習データに用いた人物を示す。

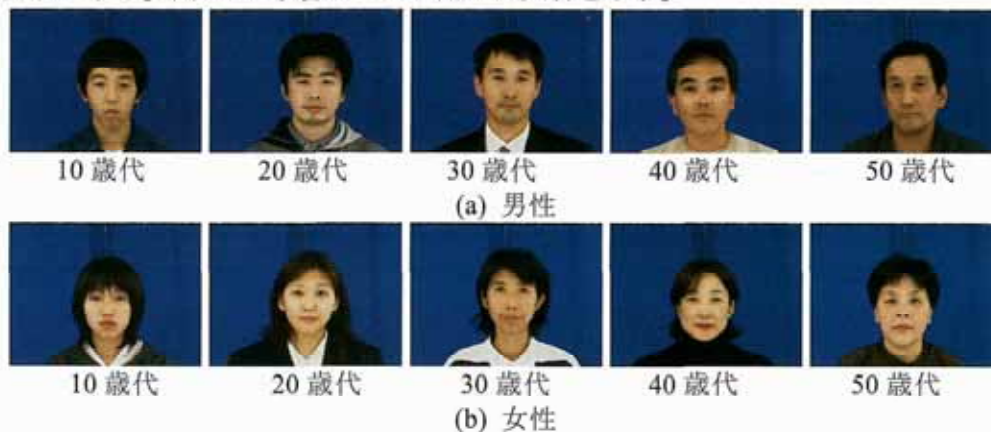


図 14 学習データに用いた人物

### 2.3.2 多方向学習の比較

2.2.2.2 節で述べたように、顔部品の四方向面特徴テンプレートの学習において、正面顔だけでなく、多方向の顔画像も学習することができる。そこで正面のみを学習した場合と、3 方向、5 方向、7 方向で学習した場合を比較した。3 方向学習では正面と $\pm 15^\circ$  の顔画像、5 方向ではそれに加え $\pm 30^\circ$  の顔画像、7 方向は $\pm 45^\circ$  を加えて学習する。学習方向数と顔向きとの関係を図 15 に示す。

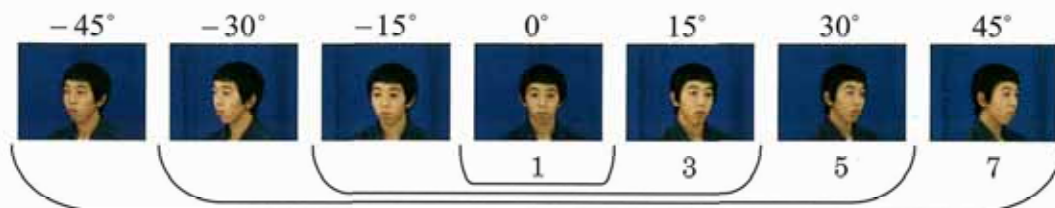


図 15 多方向学習における顔向きと学習数

多方向データベースより、眼鏡をかけていない人物のうち、肌色検出に失敗した 5 名を除いた 235 名を検証用に用いた。ここでは正面から左右 $45^\circ$  まで、上下方向の角度は $0^\circ$  のデータを使用する。表 1 に実験結果を示す。正面顔だけを学習した場合について、正面を向いた顔画像だけでなく、テンプレートに学習されていない $\pm 15^\circ$ 、 $\pm 30^\circ$  についても、顔部品モデルの拘束により、顔部品検出できる。 $\pm 30^\circ$  の顔画像に対する検出例を図 16 に示す。 $\pm 45^\circ$  の顔画像に対しては、検出が難しくなり、図 17(a)の左 $45^\circ$  の顔画像のように、左目が欠け始めると初期対応度が低くなり、誤った局所解に陥る。3 方向、5 方向の学習についても同様に、 $\pm 45^\circ$  の顔画像に対しては対処できない。

7 方向すべて学習した場合、正面付近に対して悪い影響を与えることなく、 $\pm 45^\circ$  の顔画像に対しても、顔部品検出ができる。図 17(b)に、左 $45^\circ$  の顔画像で正しく検出できている例を示す。

表 1 多方向学習による比較実験結果

学習方向	検出失敗データ数							全体	成功率
	右	←	顔向き			→	左		
	45	30	15	0	15	30	45		
1	11	0	0	0	0	0	5	16	99.03%
3	12	0	0	0	0	0	8	20	98.78%
5	7	0	0	0	0	0	1	8	99.51%
7	0	0	0	0	0	0	0	0	100%



図 16 1方向学習による検出例

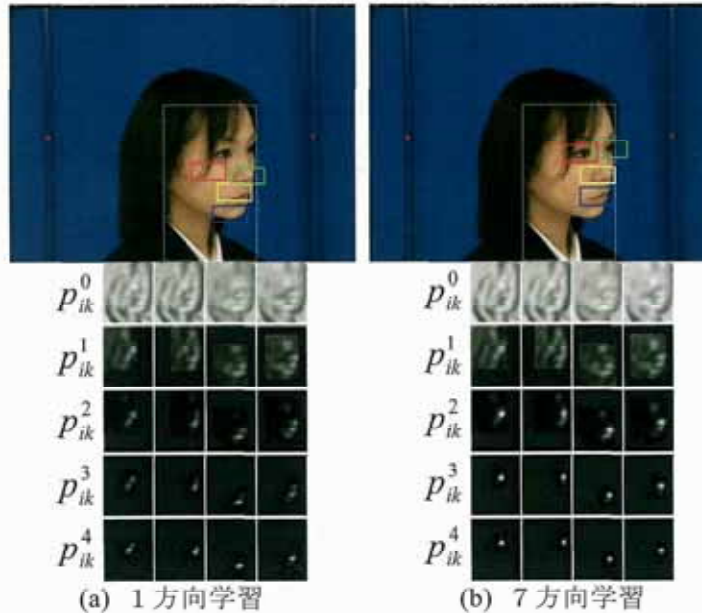


図 17 左 45° の顔画像に対する検出例

### 2.3.3 最適な標準偏差の検証

(4)式の近傍対応確率  $Q'_{jl}$  の計算におけるガウス関数の、標準偏差  $\sigma$  の値の最適値を検証する。学習方向数 7 で、 $\sigma$  のを値 4 から 10 へ変化させた場合の、顔部品検出結果を表 2 に示す。実験データは前節と同じ 235 名を用いている。 $\sigma=6$  とのきに検出成功率 100% となり最も良い結果が得られた。

図 18 は同じ顔画像に対し、 $\sigma$  を変えた場合の顔部品検出結果である。 $\sigma$  が小さいと収束が早い、バネが硬い状態に相当し、変化を吸収しづらい。図 18(a) は  $\sigma=4$  の場合であり、横向きに対応できず、誤った場所に早い段階で収束している。 $\sigma$  が大きいと個人差、角度差を吸収しやすいが、収束が遅い。図 18(b) は  $\sigma=10$  の場合で、右目位置が収束していない。 $\sigma=6$  がもっともバランスがよく、図 18(c) のように正しく検出される。

表 2  $\sigma$  による比較実験結果

分散 $\sigma$	検出失敗データ数							全体	成功率
	右 45	顔向き					左 45		
		← 30	15	0	15	→ 30			
4	2	0	0	0	0	0	2	4	99.76%
6	0	0	0	0	0	0	0	0	100%
8	2	0	0	0	0	0	0	2	99.88%
10	2	0	0	0	0	0	1	3	99.82%

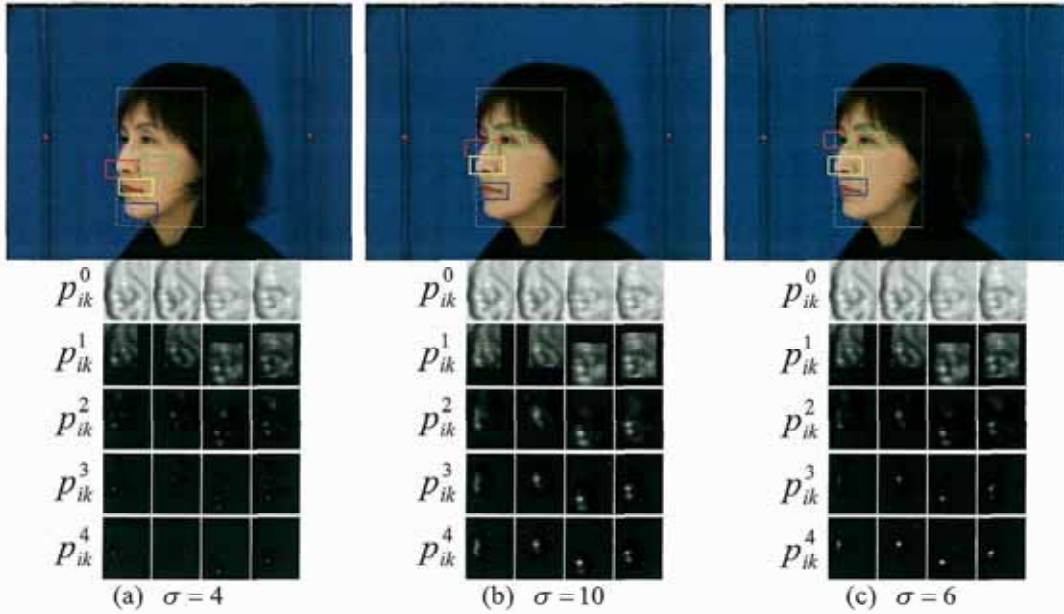


図 18  $\sigma$  パラメータ別検出結果

### 2.3.4 多方向統合モデルの検証

多方向顔画像データベースを用いて、様々な顔向き角度、カメラアングルに対応するための、多方向統合モデルについて検証する。ここでは、両目、口、鼻が画像内に収まっている範囲内ということで、正面から左右  $45^\circ$  まで、上下  $30^\circ$  までのデータを使用する。ただし、横  $45^\circ$  かつ縦  $30^\circ$  のデータは図 19 のように、前髪や鼻、頬により目が隠れる場合があるため、検出率の集計対象から除く。

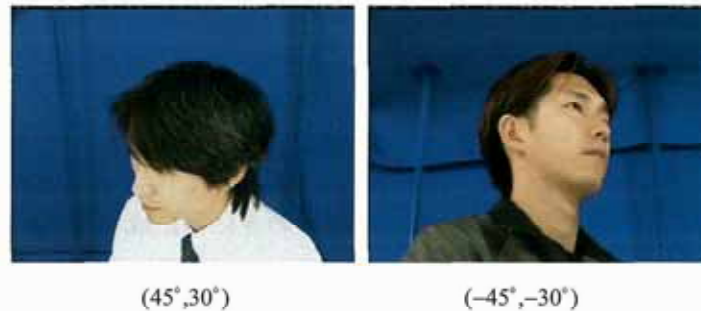


図 19 隠蔽により集計対象外とするデータ

この実験には眼鏡をかけていない人物 100 名を検証データとして用いる。このデータは男女年代が均等になるように抽出している。

顔部品検出実験では 3 種類の方法について、各角度別に検出成功率を算出し比較を行う。鼻は周辺と皮膚の色が同じで、厳密に位置を定義することが難しい為、目口がテンプレートの矩形内に収まっている場合を正解とする。ただし目口位置が検出できていれば、鼻はほぼ正しい位置を得ている。結果を表 3 から 5 に示す。横軸が横方向の回転、縦軸は縦方向の回転を示しており、平均は各行、各列での検出率を表す。右下は全体での検出率である。

まず、表 3 は従来法による検出結果である。これは、横  $\pm 45^\circ$ 、縦  $0^\circ$  の水平のみ計 7 方向でテンプレートを学習している。縦  $0^\circ$  のときに最も結果がよくなるようにパラメータ等を最適化しており、縦  $0^\circ$  はすべて成功している。縦  $\pm 15^\circ$  の場合でも、横  $\pm 30^\circ$  の範囲であれば正しく検出で



きる。この範囲内であれば、顔部品形状、位置関係にそれほど変化がないためと思われる。しかしその範囲を超えると、変化に対応できなくなる。特に縦-30°のときは、鼻孔や頬が大きく写るなど形状が異なり、検出できなくなる。

表3 水平のみ学習した場合の検出率(%)

	-45°	-30°	-15°	0°	15°	30°	45°	平均
30°	—	62	93	99	91	72	—	83.4
15°	88	100	100	100	100	100	90	96.9
0°	100	100	100	100	100	100	100	100
-15°	84	98	100	100	99	92	80	93.3
-30°	—	25	55	89	45	12	—	45.2
平均	90.7	77.0	89.6	97.6	87.0	75.2	90.0	86.26

表4は31方向すべてを用いてテンプレートを学習した場合の実験結果である。水平のみ学習の場合と比較して、縦0°について大きく精度は落ちないことが確認できる。下から撮影した場合において若干改善しているが、全体としてはあまり改善されていない。問題はテンプレートマッチングではなく、弛緩整合法のほうにある。顔部品の配置変化に対して、顔部品モデルが十分に対応できていない。

ここで、横0°の平均検出率は98.8%であり、正面のモデルで十分対応できていることがわかる。そこで検出率の悪い斜め方向に対して適合する顔部品モデルを導入する。

表4 全方向学習による検出率(%)

	-45°	-30°	-15°	0°	15°	30°	45°	平均
30°	—	68	98	99	98	79	—	88.4
15°	71	99	100	100	100	100	74	92.0
0°	99	100	99	100	100	100	97	99.3
-15°	89	100	100	100	99	100	80	95.4
-30°	—	46	72	95	54	26	—	58.6
平均	86.3	82.6	93.8	98.8	90.2	81.0	83.7	88.45

表5は全方向の学習テンプレートに加えて、多方向統合モデルを用いた場合である。多方向統合モデルは図8に示すような、正面、右下、右上、左下、左上の5方向のモデルを同時に用いている。縦-15°から縦30°まではすべて97%以上の高い検出率が得られた。縦-30°の場合は86.2%と若干落ちるが、これは眉の太い若年層男性に失敗が多くみられた。

表5 多方向統合モデルによる検出率(%)

	-45°	-30°	-15°	0°	15°	30°	45°	平均
30°	—	94	99	99	98	96	—	97.2
15°	92	99	100	100	100	100	88	97.0
0°	98	100	99	100	100	100	98	99.3
-15°	91	100	100	100	99	100	99	98.4
-30°	—	88	96	98	78	71	—	86.2
平均	93.7	96.2	98.8	99.4	95.0	93.4	95.0	96.13

本手法を用いた、20代男性1名に対する全方向の検出例を図20に示す。白の矩形が検出位置を示す。顔部品検出に失敗しているのは、集計対象外とした(-45°,30°)と、若年層男性に誤りの多い(30°,-30°)である。(30°,-30°)では右眉の位置を右目と誤検出している。原因として、頬や鼻の張り出し部分が、右目領域にかかっており、テンプレートマッチングの類似度が低くなることにある。集計対象外とした(±45°,±30°)のデータについては、隠蔽の発生する片目を除けば、ある程度片目・口・鼻の位置は検出できる。

検出率を表6にまとめた。監視カメラなど上方からの撮影を想定し縦0°から30°までの場合、多方向統合モデルで97.89%の検出率であった。



図 20 顔部品検出例

表 6 多方向統合モデルの実験結果

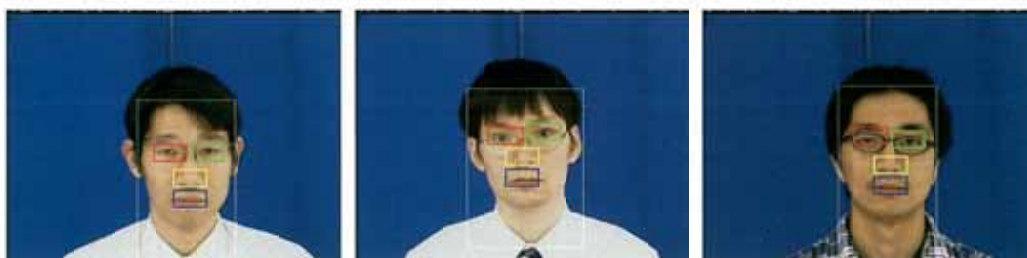
検出率(%)	全方向	上から撮影
水平のみ学習	86.26	94.47
全方向学習	88.45	93.74
全方向学習+多方向モデル統合	96.13	97.89

### 2.3.5 眼鏡をかけた人物に対する結果

HOIP 顔画像データベースのうち、眼鏡をかけた人物 48 名に対する検出率を表 7 に示す。眼鏡をかけていても、9 割程度は検出できている。そのときの顔部品検出例を図 21 に示す。フレームの種類により検出の難度がことなり、図 21(a)のようなリムレスタイプはほとんど問題なく検出できる。以降、図 21(b)のメタルフレーム、図 22(c)セルフフレームの順に、四方向面特徴のエッジが強くなるため、難しくなる。また、レンズが小さい場合も難しくなる。

表 7 眼鏡をかけた人物に対する結果

	顔向き							全体
	右 45	← 30	15	0	15	→ 30	左 45	
失敗人数	7	7	7	3	6	5	8	43
成功率	85%	85%	85%	94%	88%	90%	83%	87%



(a) リムレス

(b) メタルフレーム

(c) セルフフレーム

図 21 眼鏡をかけた人物の顔部品検出例

### 2.3.6 別環境での結果

HOIP 顔画像データベースでは理想的な環境下で撮影されている。顔に影ができないよう照明環境は整えられ、背景は単純なブルーである。しかし実際のアプリケーションを考慮すると、一般的な環境下においても動作できることが必要である。

図 22 に実環境下での顔部品検出例を示す。顔部品テンプレートは HOIP データベースから生成されているが、四方向面特徴を用いているため、また、明るさの変化にロバストである。このため、異なる環境下でも正しく動作する。背景の変化も問題にならない。図 23 はまた別の照明環境、人物での動作例である。以上の結果より、本手法は、環境や表情、人物の変化にロバストであることが確認された。

この実験で用いた手法は、Xeon 3.06GHz の PC を使用して 30fps の速度でリアルタイムに動作している。

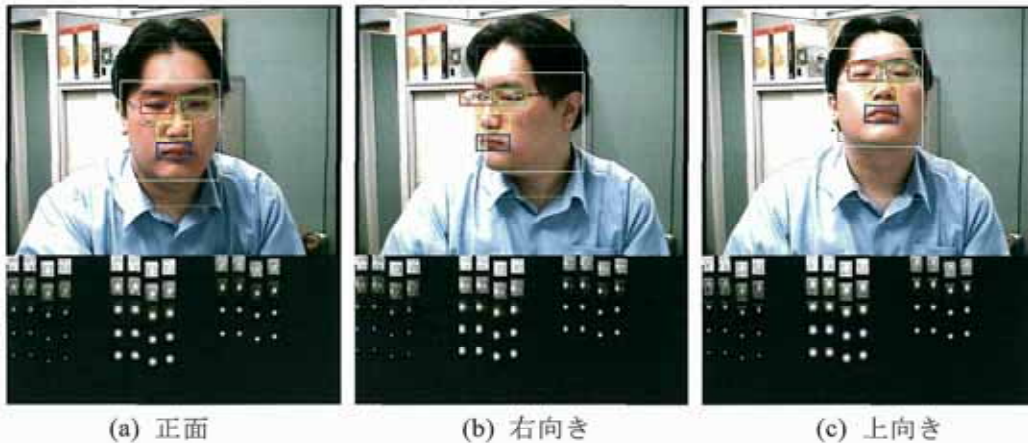


図 22 実環境下での動作例 1



図 23 実環境下での動作例 2

### 2.3.7 複数カメラによる顔部品検出

左右に配置された PTZ カメラから同時に撮影された顔画像から、顔部品位置を検出した例を図 24 に示す。図 24(a)(b)において、上段が入力画像で、矩形が顔部品検出位置を示し、下段は弛緩演算による対応度の変化を示している。カメラ間の距離は 80cm で、PTZ カメラは内向きにそれぞれ 20 度を向いている。

図 24(a)は、それぞれのカメラで独立に顔部品を探索している。メガネフレーム等の影響で、右目の検出位置がずれている。図 24(b)は 2.2.4 節の、画像間のエピポーラ線拘束による画像間適合確率を使用した場合である。弛緩演算において対応確率をカメラ間で相互に渡すことで、3次元空間上で一致する特徴点位置を検出できている。



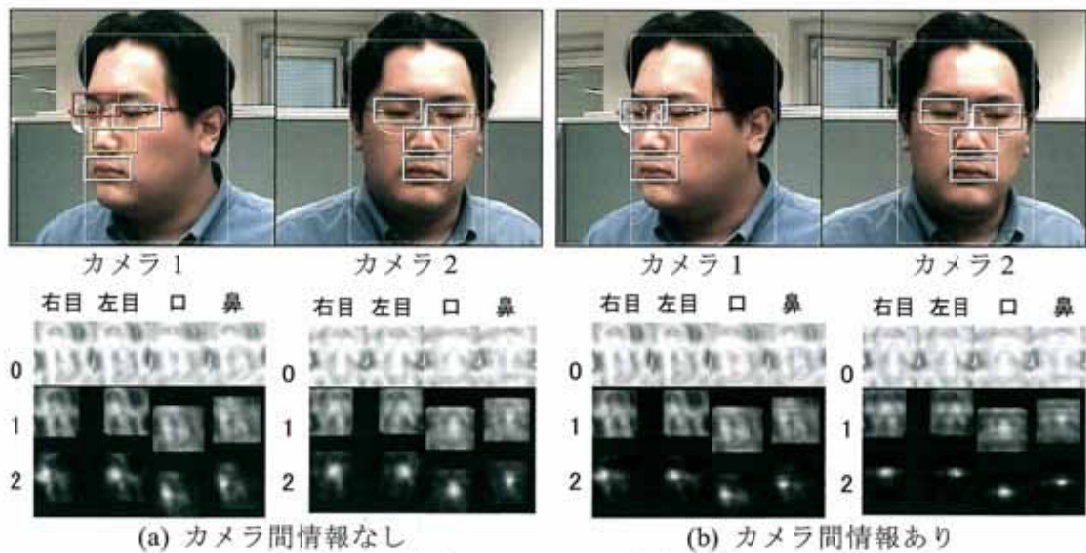


図 24 複数カメラによる顔部品検出

### 2.3.8 接続選択と連続画像の実験

2.2.5 節において、遮蔽に対応するための接続選択と連続画像による顔部品検出手法を検証する。検証データとしては、ターンテーブルにより 5° 間隔で被験者を回転させて撮影した回転データベースで、学習人物を含まない 290 名を用いる。正面顔から、右 90° まで回転する画像 18 枚を連続した動画像とし、今回の手法による顔部品検出の角度変化に対する追従性能を比較する。各種法での顔部品検出結果を図 25 に示す。

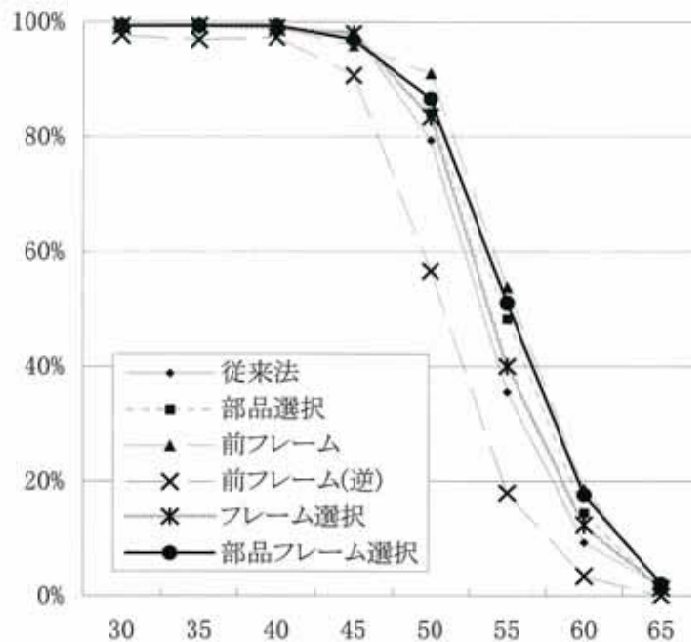


図 25 接続選択と連続画像による実験結果

このグラフは、横軸の各顔向き角度に対する、顔部品の検出成功率を表している。顔向き角 40° 以内では、99% の検出に成功している。45° 以上では、自己遮蔽のため、右目が画像に写らなくなるため、左目・口・鼻の三点が正しく検出できていたら、成功としている。図 25 の凡例のうち、「従来法」は 2 章で述べた静止画に対する方法である。テンプレートは 45° までしか対応しない上、遮蔽も発生するため、45° 以降検出性能が落ちている。

「部品選択」は、2.2.5.1 節の部品間の接続選択を用いた場合である。45° 以降検出性能が落ちるのは変わらないが、片目が遮蔽した場合でも、それ以外の顔部品を検出できるケースがあり、従来法と比べ 50° 以上の顔向きに対して粘り強く検出できる。ただし従来法で成功している一部のデータで失敗するケースがあった。これは弛緩法において確率更新の影響を受ける顔部品



が少なくなることから、弛緩演算の収束速度が遅くなり、局所解へ陥っていることが考えられる。

「前フレーム」は、2.2.5.2節の前フレームの検出情報を反映させた周辺適合確率を用いた場合である。40°や45°の顔向きで正しく検出された位置情報が反映されるため、45°以降の顔向きに対して成功率が高い。ただしこれは前フレームの情報が正しいことが前提となっており、この前提が崩れるとうまく動作しない。「前フレーム(逆)」は、データを逆再生した場合であり、右向き顔から正面へ向かって回転している。検出できない顔向きから始まるため、誤った前フレームの情報が継承されて、本来検出できていた顔向きでも失敗していることがわかる。

「フレーム選択」は2.2.5.3節の前後のフレームを選択的に用いる方法の結果である。前後のフレームを等価に用いるため、逆再生でも同じ結果が得られる。フレーム間情報を安定に得られる為、「従来法」より良い結果が得られている。

部品間の接続選択とフレーム選択を同時に用いた結果が、「部品フレーム選択」であり、2.2.5.4節で述べた方法である。顔向き変化に対し最も良い追従性能が得られた。また検出結果を部分的に見た場合、部品間の接続選択のみを用いた場合と比べ、弛緩演算の収束が安定していることが確認された。この方法による検出例を図26に示す。45°以降右目は自己遮蔽されているが、55°まではフレーム間情報で補完されている。

自己遮蔽などによる一部欠損に対処するため、部品間の接続選択と動画情報により、弛緩法を拡張する方法は、顔部品間の関係を記述する近傍対応確率のうち、部品間の接続関係を自動的に選択することで、顔部品の一部が隠蔽する場面に対処することができることがわかった。さらに動画像を考慮し、フレーム間の情報を相互参照するよう弛緩演算を拡張、前後フレームを選択的に参照し、部品接続選択を組み合わせることで、動画情報と顔部品間を自動的に補完することができている。

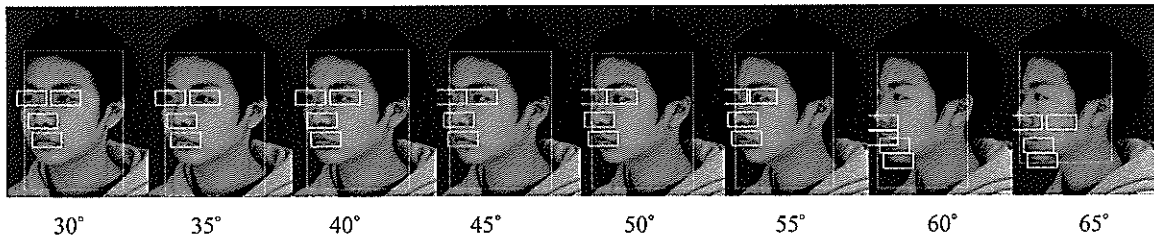


図 26 部品フレーム選択法による顔部品検出例

## 2.4 小括

ユーザインタフェースや人物認識など実環境での応用を目的とした、顔向き、個人差、表情、環境等に対しロバストな顔部品検出手法として、四方向面特徴と弛緩整合法を用いた手法を構築した。この手法は、多人数・多方向を学習した四方向面特徴と、顔部品モデルのスプリング接続を考慮した弛緩整合法に顔の方向に適合した多方向統合モデルを用いることにより、個人差・顔向きによらず顔部品を検出できる。HOIP 顔画像データベースを用いた検証実験により、個人差、顔向き変化に頑健であることを示した。また、照明条件等の異なる実環境下においても、リアルタイムで良好に動作することを示した。

## 3 注目推定による自動マーケティング情報システム

### 3.1 研究の目的

消費者のニーズを把握するためのマーケティング調査は、商品開発において重要な基礎的情報を与えるものである。特に顧客がどのような行動を取るか、何を見るかを分析することの重要性が指摘されている。このような調査は一般的に、アンケートによる聞き取りや、調査員による購買行動の観察などにより行われている。ただし、信頼性を高めるためサンプル数を増やすことによるコスト増大や、調査員によりばらつきが生じやすい問題がある。そこで人間センシング技術を応用し、均質な情報を連続的に収集できる、「自動マーケティング情報システム (AMIS: Automatic Marketing Information System)」を構築する。

センシング技術のマーケティング応用は、足音とシルエットから性別を判定して来店客の統

計をとるアプローチや、行動履歴から顧客の感性をモデル化して、情報提示をする提案などがなされている。AMIS では、顧客に特殊なセンサを装着させたり、動きを拘束させたりすることなく、非接触で情報収集できる利点がある。パン・チルト・ズームカメラ(PTZ カメラ)により、顧客の詳細な顔画像を取得することで、顧客が何を注目しているかを検出し、性別・年齢を統計的に推定することで、性別・年齢層ごとにセグメンテーションされた興味度を、マーケティング情報として得られる。

## 3.2 実施内容

### 3.2.1 自動マーケティング情報システム AMIS

自動マーケティング情報システム AMIS は、一様なマーケティング調査を連続して行うことを目的としている。この章では、マーケティング情報をカメラ映像からの人物センシングにより収集する AMIS について、顧客の性別・年齢と注目の情報を顔画像から得ることの有効性を述べる。また、プロトタイプとして構築した、プラズマディスプレイ(PDP)に広告を表示し情報収集を行う、広告提示型 AMIS の構成について述べる。

#### 3.2.1.1 AMIS の概要

AMIS の基本機能は、商品や広告に注目する人物より、属性や興味度をマーケティング情報として収集することである。AMIS の運用場面として、街頭や店舗における、商品ディスプレイや、宣伝広告などが考えられる。提示情報を見るために立ち止まった人物や、近づいてくる人物、歩きながら見ている人物が情報収集の対象になる。このような人物を、不特定多数が行き交う中から見つけ出すため、システム周辺を広角カメラで広範囲に捉え、人物の移動を検出する。視線推定や人物属性の推定には、ある程度以上の解像度で捉えた顔画像が必要になる。そこで PTZ カメラにより、顔を追跡し必要な顔画像を取得する。

ところで、マーケティング調査は、商品の対象とする消費層を性別や年齢等によりセグメンテーションし、商品の魅力や人気の評価、顧客の興味度などを推定することが重要である。AMIS では、このための情報を顔画像から取得する。

まず消費層のセグメンテーションについては、従来個人認証として研究されてきた顔認識技術から展開し、顔画像から性別・年齢を統合的に推定する方法を構築しており、これを用いて推定することができる。

次に商品に対する興味度については、顧客の内面的な情報を、カメラ映像という外面的な情報で類推する必要がある。例えば、ステレオカメラや視体積交差法により、人物位置を検出し、顧客の移動履歴を捉えることができる。顧客は、「興味のある商品の前で立ち止まる」ことや、「興味のある商品に何度も立ち寄る」ことがある。そこで、顧客の滞留を検出することで、興味を推定することができると考えられる。

もう一つのアプローチとして、顧客の視線を検出する方法がある。視線情報とアンケート調査に一定の相関があることが報告されていることから、顧客は「興味のない商品よりも、興味のある商品をより注目する」と言える。よって顧客の注目行動から商品に対する顧客の興味度が得られると考えている。

このように収集された興味度を長期間、多人数で集計することで、商品の人気などを統計的に評価できる。

#### 3.2.1.2 広告提示型 AMIS の構成

今回プロトタイプとして、広告に対するマーケティング情報を収集し、顧客に合わせた情報を PDP に表示することができる、広告提示型の AMIS を構築した。外観を図 27 に示す。広告提示用の 50 インチ PDP の周囲に 3 台の広角の固定カメラと 4 台の PTZ カメラを配置している。

システムの機器構成を図 28 に示す。人物を検出・追跡するセンシング PC を 1 台、人物の詳細情報を取得するカメラ PC を 4 台、提示画像の表示と情報を集計するメイン PC の 1 台で構成され、各 PC 間は TCP/IP を用いて通信を行う。



図 27 広告提示型 AMIS 外観

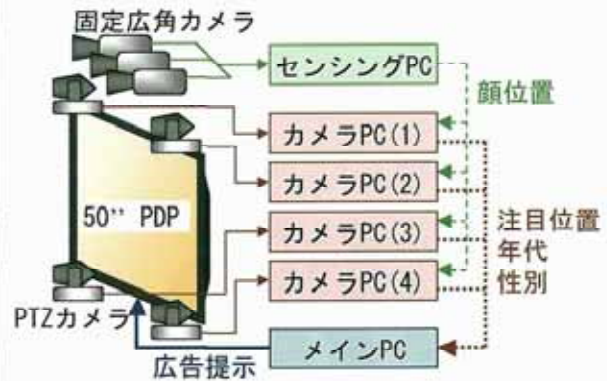


図 28 システム構成

固定カメラは、システムの前方 1m から 5m の距離にいる人物を広範囲に捉えられるように設置している。具体的には、各カメラの画角は 60 度で、光軸を一致させ下方 10 度を向くように、PDP 上部に 60cm 間隔で設置している。この映像は同期を取っており、画像合成ユニットにより 1 台のセンシング PC が取得する。センシング PC は、人物位置と顧客の顔を検出する。得られた顔の三次元位置をカメラ PC へ送信する。

PTZ カメラは、PDP 上下に設置しており、各カメラは 1 名の顔画像を取得するため、最大で 4 人まで捉えることができる。各カメラには、1 台ずつカメラ PC が接続されており、映像取得および制御を行う。カメラ PC は、センシング PC より送られた顔の三次元位置から、カメラのパン、チルトおよびズームの制御を行い、詳細な顔画像を取得し、人物属性および注目情報の推定を行う。カメラ PC により得られた情報はメイン PC に送られる。

メイン PC は、PDP に左右 2 分割で広告情報を提示している。カメラ PC の情報より、どのような人が、どちらの情報をみていたか集計する。さらに集計結果から、顧客が望むと思われる広告情報に切り替えて提示することができる。

### 3.2.2 人物検出

システムの前にいる複数の人物をグローバルに捉えるため、PDP 上部に広角の固定カメラ 3 台を設置している。このカメラ映像は、1 台のセンシング PC に集約され、複数人物検出と顔検出により、顧客の顔の三次元位置を得ている。その具体的な方法を以下に述べる。

#### 3.2.2.1 複数人物検出

まずシステムは、提示情報を注目している人物を、不特定多数が往来する中から見つけ出す。ここで、提示情報を注目している人物とは、システムに近い場所に立ち止まっていると仮定する。これは、遠い場所からでは、他の人に遮蔽されて十分に情報を見ることができず、歩いたまま情報を見る場合は、短時間にちらりと見る程度であり、興味を示さなかったと考えられるからである。

複数人物の検出には、研究テーマ「人物追跡に関する研究」において構築した、多視点カメラによる視体積交差法を用いる。視体積交差法は、人物を三次元空間上のボクセルとして検出できる。まずカメラごとに背景差分により、人物候補領域を検出し、その候補領域をボクセル空間上に投票する。このボクセルに対しラベリングを行い、ブロックに分割する。前フレームでの人物位置に最も近いブロックを同じ人の人物ブロックとし、その重心を人物位置とする。この人物位置を各カメラ画像に逆投影することで、人物領域を検出する。

図 29 にシステムの前立つ複数人物の検出結果を示す。図 29(a)は、各カメラにおける人物領域を矩形で示している。各人物に A から D のラベルをふっており、カメラ間で同一人物を同定している。図 29(b)はシステムからの距離を算出した結果で、明るい部分は距離が近く、暗い部分は遠いことを示している。図 29(c)は対応する人物のボクセル空間の投票結果から、ラベリングにより人物ブロックを検出した結果である。この人物ブロックより人物の移動軌跡を推定したものが図 29(d)である。このように複数人物に対し、調査範囲内における歩行や静止等の行動を把握することができる。



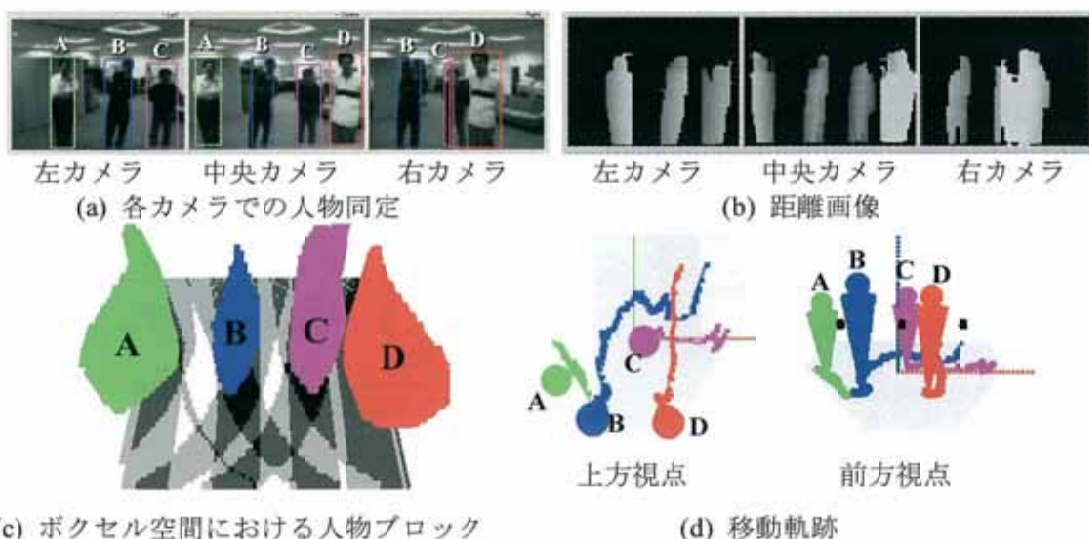


図 29 複数人物の検出

### 3.2.2.2 顔を向けている人の検出

検出された人物のうち、広告を見るために情報提示装置前で立ち止まり、顔を PDP に向けている人物を情報収集の対象とする。このために、人物領域に対して顔検出を行う。画像から顔を検出する方法としては、さまざまな提案がなされている。実時間処理に適した手法として、色を用いる方法があるが、広範囲を撮影している広角カメラ画像の場合、環境光や人物の違いにより、肌の色が均一にならない問題がある。また、背景に類似色が存在する場合、顔か否かを区別する必要もある。そこで色を使わず、形状から識別器により判定する方法として、四方向面特徴とサポートベクタマシンを用いた顔検出手法を用いる。人物領域から候補領域を切り出し、四方向面特徴を抽出する。これを統計的識別器として近年注目されているサポートベクタマシンにより、顔か否か判定する。サポートベクタマシンは識別境界面を作るサポートベクタを増やすことで判定精度が向上するが、計算コストも増加する問題がある。今回は、人物領域に探索範囲を限定することで、リアルタイム性と誤検出の抑制を図っている。図 29(a)の人物領域から、顔検出を行った結果を図 30 の黄色矩形で示す。ここでは 4 人全員がシステムを見ており、その顔位置を正しく検出できている。



図 30 顔を向けている人の検出

### 3.2.3 顔画像処理

カメラ PC は、固定カメラにより検出された顔位置から、PTZ カメラを制御して詳細な顔画像を取得する。得られた顔画像より顔領域を検出する。この顔位置に対してさらに追跡を行い、顔部品検出により注目位置を推定する。そして、顔から特徴量を抽出し、年齢・性別を推定する。各要素技術について以下に述べる。

#### 3.2.3.1 PTZ カメラ制御

センシング PC により得られた顔位置に向けて PTZ カメラを制御し、高解像度の顔画像を取得する。この画像から顔領域を検出し、その領域に対し、顔部品検出・属性推定を行う。検出された顔に対してさらに PTZ カメラを制御して、以降取りこぼさないように追跡する。ここでの要求として、高速に動作する検出手法が必要である。顔を大きく撮影しているため、人物の

わずかな動きでも画像中では大きく移動する。このような場合でもフレームアウトする前に顔を検出し、PTZ カメラを制御する必要がある。

そこで顔領域検出には、高速に動作できる肌色検出法を用いている。ここでは CIE-L\*u\*v\* 表色系での肌色基準値法を用いている。この手法は、明るさの変化に強い。また、1 人の顔を大きく撮影している為、画像内の肌色領域の面積が大きく、個人ごとの肌色が容易に特定できる。これにより、環境光や人物の違いに対処できる。顔が大きく撮影される場合は、良好に顔領域を検出できる。

ズームの制御については、初めはセンシング PC により得られた顔への距離から、基準とする顔の大きさが得られるズーム率に設定する。以降は、PTZ カメラ画像から得られた顔領域の横幅を、基準とする横幅にあわせるようにズーム率を制御する。顔が検出できない場合や、追跡に失敗した場合には、センシング PC に問い合わせ、最新の顔の位置情報を得る。

### 3.2.3.2 顔部品検出による注目推定

次に目鼻口の顔部品位置を検出する。様々な人物、方向の顔画像から、リアルタイムでの確に顔部品を検出する必要がある。そのために、四方向面特徴と弛緩整合法を用いた顔部品検出法を用いる。この手法では、左右 45 度、上下 30 度程度までの顔向きで検出できるので、PDP を注目している範囲の顔向きには十分対応できる。

顔向きから注目位置を推定する。顔領域に対する顔部品の位置関係から、顔向き角度を算出する。人物位置は、パン・チルト角、ズーム率と顔幅により推定する。人物位置から顔方向への延長線が PDP の平面と交差する点を、注目位置とする。今回の顔向き推定は水平方向のみのため、顔位置から床面に対して水平方向を向いているとした。

### 3.2.3.3 性別と年代の推定

研究テーマ「性別及び年齢の推定に関する研究」において構築した、四方向面特徴と線形判別分析で学習した識別器を用い、性別と年齢を統合的に推定する手法を組み込んでいる。性別や年齢の推定には、顔全体の大まかな特徴量と、目鼻口の詳細な特徴量を用いて判断することが重要と考えられる。そこで顔領域全体を用いて生成した四方向面特徴と、目鼻口を含む顔部品領域から生成した四方向面特徴を 2 つ用いて判別する。HOIP 顔画像データベースの正面顔に対して実験を行った。このデータは、男性 150 名、女性 150 名で、それぞれ 15 歳から 64 歳まで、5 歳間隔で 15 名ずつ収録されている。性別推定では 92% の識別率を得ている。年齢については、10 歳間隔の年代推定において、学習人物に対しては約 70%、未知人物に対しては約 40% が得られている。

## 3.3 結果

今回構築したシステムの動作実験により、各要素技術が連携し、情報収集システムとして有効に動作することを示す。

まず一人の人物がシステムの前を移動し、PDP を注目するシーンにおける動作を検証する。ここでは提示情報に興味を持った人物を想定して、システムに近付いて PDP を見た後立ち去るという行動をとった。このときの人物の移動軌跡を検出した結果を図 8 に示す。センシング PC では、3 台の広角カメラから、顔の三次元位置を每秒 5 フレーム程度で検出可能である。この位置から PTZ カメラを制御して高解像度の顔画像を獲得し、顔部品位置を検出した結果が図 31 である。図 33 の(a)から(f)は、図 8 に示した各地点での人物の顔画像取得結果である。(a)(b)において人物はシステムへ向かって前進しており、PTZ カメラのズーム率を制御し、一定の大きさの顔画像を取得している。(c)(d)では立ち止まり首を左右に振っている。その後(e)(f)では横へ移動する人物をパン・チルト角を制御して追跡している。カメラ PC における顔画像処理は每秒 20 フレーム程度が可能であり、このような人物の振る舞いに対して素早く追従できる。このとき、人物位置と顔向きから推定される注目検出位置の推移を図 32 に示す。横軸が時間（フレーム番号）、縦軸は PDP の中心を原点とした時の、水平方向の注目検出位置である。50 フレームまで人物は前進しており、その後 130 フレームまで立ち止まり、左右交互に PDP の情報を注目していたことが検出されている。

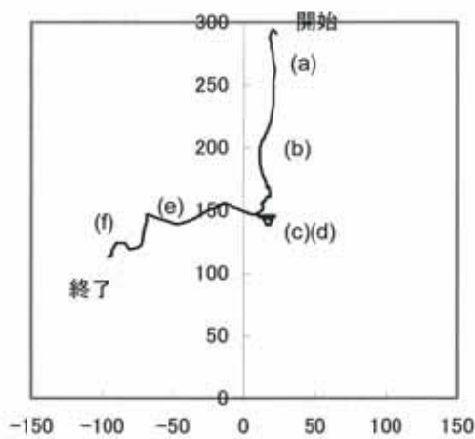


図 31 人物位置推定結果

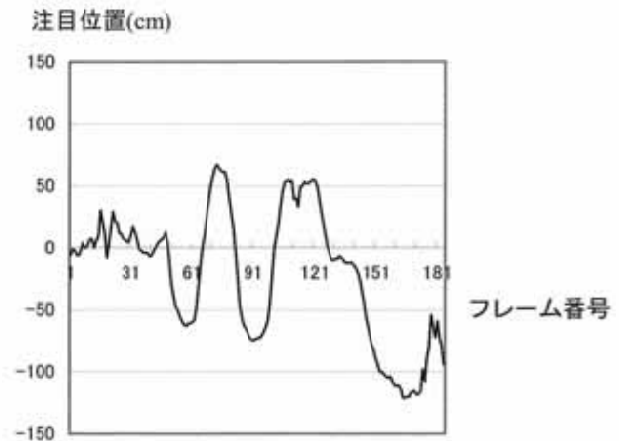


図 32 注目位置推定結果

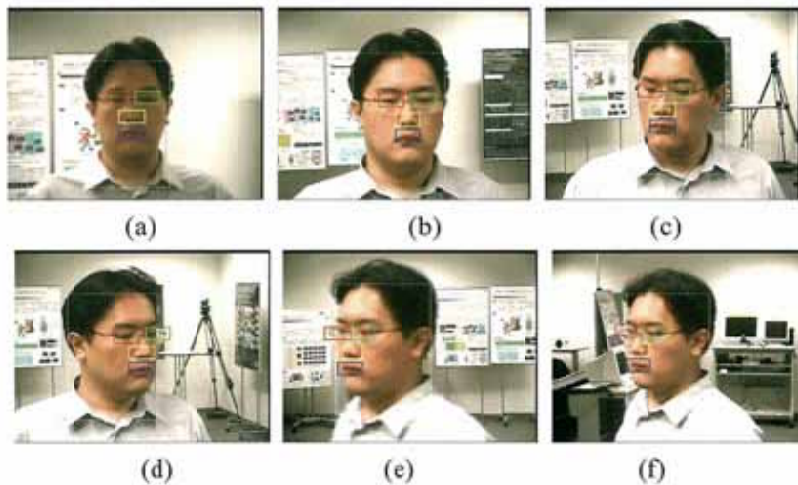


図 33 PTZ カメラによる顔追跡結果

次に、複数の人物がシステムを注目しているシーンにおける動作を示す。このときの顔検出結果を図 34 に示す。図 34(a)のように 3 名の人物が PDP を見ているが、図 34(b)に示すように、3 台の PTZ カメラにより複数人物の顔情報を同時に取得できることが確認された。



(a)中央カメラ映像

(b)PTZ カメラ映像

図 34 複数人物の顔画像取得結果

このとき PDP には図 35(a)のような広告情報を提示している。今回は、コンテンツを左右二分割し、それぞれ異なる広告情報を表示している。一例として、旅行情報の広告を想定し、左に京都の映像、右に沖縄の映像を使用した。3 名は右側の京都の情報を注目しており、映像の下に、注目している人物属性の推定結果をイラストで表示している。ただしこのイラスト表示は、AMIS の機能を紹介するためであり、実際のシステム運用では必要ない。このときは 40 歳男性、20 歳男性、20 歳女性と推定している。性別推定については 90%程度の正答率が得られているが、年齢推定は 40%程度の正答率であり、年齢推定精度の改善が課題となっている。また、今回の広告表示は二分割であるが、視線検出精度を向上させることで、提示情報のどの部分に興味をもったかなど、より詳細な分析が可能になると考えられる。



この後、京都の情報をより注目していたため、その詳細情報が図 35(b)のように提示される。このように広告提示型の AMIS では、注目度に応じて内容を変更することができる。さらに、顧客の人物属性に合った情報を提示することもできる。例えば、女性ならば女性向け商品の広告を流したり、子供がいれば子供向けの内容に切り替えるなど、ダイナミックに提示情報を変えることも考えられる。

このシステムのデモンストレーションを行い、見学者の集計を行った結果は図 35(c)のようになった。この数値は累計注目時間を示しており、どちらの情報により注目されたかを、人物の性別・年齢層別に集計している。40 歳前後の男性が多く集計されているが、実際の見学者の傾向もそれに近かった。このことから、個別の年齢推定精度には課題が残るが、全体の統計を取ることで顧客の性別年齢の傾向が得られていることが判った。ただし顧客の興味度を測る指標は注目時間だけではなく、顧客の注目履歴を集計するなど、様々な要素が考えられ、今後検討する必要がある。



(a) 広告提示と推定結果

(b) 詳細情報の提示

(c) 性別年代別注目集計結果

図 35 提示情報例

### 3.4 小括

人物センシング技術を応用し、広告や商品を注目する顧客の顔画像を撮影、そこから得られる注目度や、性別・年齢といった属性情報を自動的に取得する、自動マーケティング情報システム(AMIS)の研究を実施した。そこでは、複数人物の検出と追跡、顔検出、注目推定、性別・年齢推定技術等の要素技術が用いられている。プロトタイプとして情報提示型システムを構築し、高解像度の顔画像を取得し、提示情報に対する注目度や性別・年齢層を集計するための基本性能を示した。

## 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、個人差、表情、環境等に対しロバストな顔部品検出手法を構築し、その有効性を示した。また、注目推定による自動マーケティング情報システム AMIS の構築を行った。

今後、顔部品検出手法については、実環境に対応するため、さらにロバスト性を高めるため改良を行う。AMIS については、実際の店頭などでフィールドテストを行い、顧客の心理や感性情報処理といった知見とあわせ、様々な要素を検討する必要がある。また、各要素技術については、各手法の精度の向上、詳細な注目検出法の構築などが挙げられる。