

# 顔・顔部品の階層的トラッキング技術開発と アイコンタクト顔メディア創出の研究

## Research and Development of Hierarchical Face and Facial Parts Tracking Technology and Eye-Contact Facial Media Creation

舟橋琢磨 †  
Takuma Funahashi

藤原孝幸 †  
Takayuki Fujiwara

興水 大和 †  
Hiroyasu Koshimizu

### 1. 緒言

顔メディア創出の実利用化を目的とした場合、撮像環境のロバスト性は重要な要素であり、撮像された顔画像から特徴抽出・分析を行うための必要十分な画像を取得しなければならない。シーン中の顔検出[1]や、撮像された顔画像を利用した特徴抽出、分析[2]の研究は多数存在するが、処理対象となる顔画像の品質が保証できるという研究はされていない。

また、カメラデバイス、PCの低価格化とネットワーク通信の高速化によって、非対面対話環境が容易に構築可能となった近年では、入力機器と出力機器の分離によって起こるアイコンタクトの不一致が問題となっている。この視線不一致というヒューマンインタフェース品質低下問題は、単に特殊な機器や環境[3,4]を利用することでは解決できない深刻な問題である。

本稿ではこれらの諸問題に対して、階層的顔・顔部品トラッキングによる処理対象となる顔画像の品質を保証し、顔領域抽出から瞳輪郭抽出を行い、モニタ・カメラ・人の位置関係より垂直・水平視野角度を算出した新たな瞳中心座標に対して瞳再描画処理を行うことで、アイコンタクト映像生成という顔メディア創出の可能性を確認できた。

### 2. 顔・顔部品トラッキング

#### 2.1 システムフロー

広い視野確保のための CCD カメラと、精細な画質を確保するための Pan-Tilt-Zoom カメラを連動させ、背景差分とテンプレートマッチングを併用した手法にて顔領域の抽出、追跡を行い、連続して必要十分な画質の顔画像を取得してから瞳・鼻孔の近似円検出を基礎として各顔器官部品領域を抽出して、各領域の特徴点抽出を行い、各顔部品自動認識を行う階層的顔トラッキングシステムを開発した。

動作範囲については、CCD カメラから撮像された画像に対して頭部トラッキングを実現して顔画像の取得を想定した。そのため、人物の可動領域はこのカメラの撮像可能範囲内に限定されるが、図 1 に示す円錐台形状の空間内において人物の自由な振る舞いを許容することができることを意味する。

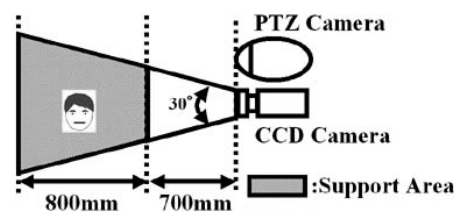


図 1 システム動作範囲

#### 2.2 顔トラッキング

顔領域抽出は、テンプレートマッチングを基礎として、フレーム毎に背景とテンプレートの更新を行う。前処理として任意に設定した平滑化された入力画像と背景の差分画像を生成し、2値化、ノイズ除去を実施した後、射影変換を行う。射影結果より、水平軸方向から走査して、射影値が閾値以上をとった座標値を顔領域としてテンプレート登録する。背景画像は、 $n-1$  フレーム目の各画素の濃度平均値を読み込み、 $n$  フレーム目の各画素に、式(1)を用いて各画素の濃度値を更新することで、背景画像を更新する[5]。

$$\bar{f}_n(i, j) = \frac{\{\bar{f}_{n-1}(i, j) \times (n-1) + f_n(i, j)\}}{N}$$

$\bar{f}_n$  : Renewed average gray value  
 $\bar{f}_{n-1}$  : Average gray value at the previous frame  
 $f_n$  : Input gray value of  $n$ -th frame  
 $N$  : Total number of frames for averaging

顔領域追跡は、抽出された顔領域の中心座標( $cx, cy$ )を基に、式(2)にて水平・垂直方向の移動量を算出し、PTZ カメラ移動量  $M(x, y)$  を利用して PTZ カメラに命令信号を送信することで、水平・垂直方向の追跡を行う。奥行方向の追跡は、PTZ カメラにて顔部品が認識出来る大きさの画像を取得することを条件とした時の、顔領域サイズ  $L$  の最大限界値  $L_{max}$ 、最小限界値  $L_{min}$  を利用して、拡大縮小率  $Z$  (ズーム値) を式(3)にて制御する。また、ズーム値と顔領域サイズの関係は図 2 のように顔領域サイズに逆比例してズーム値は減少する。

$$M(x_n, y_n) = \{(cx_n, cy_n) - (cx_{n-1}, cy_{n-1})\} \times \Delta s$$

$n$  : Number of frame  
 $\Delta s$  : Movement value per Pixel for PTZ Camera

$$Z(L) = \frac{Z_{max}}{L_{max} - L_{min}} \times (L_{max} - L)$$

† 中京大学大学院情報科学研究科

‡ 中京大学情報理工学部

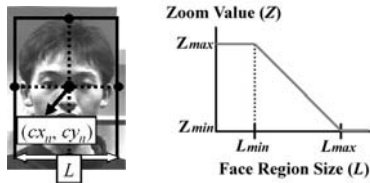


図2 ズーム値と顔領域サイズの関係

### 2.3 顔部品トラッキング

顔部品トラッキングは、PTZ カメラによりトラッキングされた RGB カラー顔映像 (図3 (a)) に対して、色相(Hue)、彩度(Saturation)、明度(Value)を利用した HSV 表色系へした後、色相を用いて肌領域を抽出する。抽出した肌領域に対して、図形融合(膨張/収縮処理)、最大領域のラベリング処理によって得られた領域(図3 (b))を顔領域  $M \times N$  として定義し、 $(M \times N)/4$  を目領域(図3 (c))として算出する。

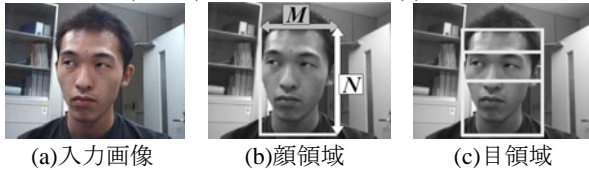


図3 領域抽出処理

瞳認識には Hough 変換による円検出を用いた。円の Hough 変換によって幾つか検出された瞳候補から、最良の候補円対を瞳対として選択する。得られた瞳候補を、顔領域の中心を境界に左の瞳と右の瞳に分け、以下の基準に基づいて、最良のものを正しい瞳対として選択する。

- i). 左右の瞳中心座標 → 水平方向に定めた閾値内
- ii). 左右の瞳半径 → 同径より定めた誤差範囲内
- iii). 左右の瞳濃度差 → 閾値内

抽出された瞳位置(図4(a))に基づいて鼻領域を定め、領域内で瞳輪郭抽出と同様の方法で鼻孔輪郭抽出を行う。次に、抽出された瞳・鼻孔位置に基づいて顔部品領域(図4(b))を定め、領域内で顔部品輪郭抽出(図4(c))を行う[5]。

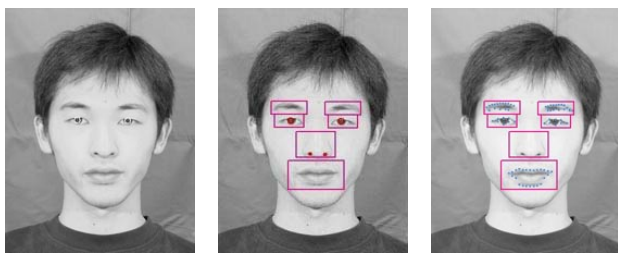


図4 顔部品トラッキング

### 3. アイコンタクト映像生成手法

本システムは図5のようなシステムフローによって動作をする。Web カメラより入力された画像を取得し、その画像に対して顔領域抽出を行い、領域内にて瞳認識を行う。認識された瞳の中心座標値を利用して瞳位置、カメラ、モニタの三者の位置関係を統合して視線が合う瞳位置を算出して再描画する[6]。

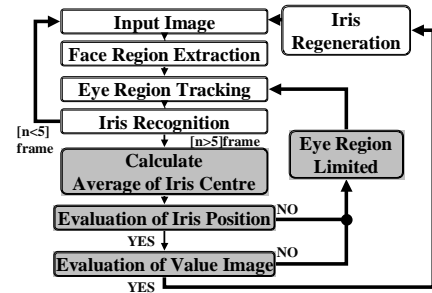


図5 システムフロー

### 3.1 瞳判別評価

動画像の入力を開始して、 $N-5$  フレームから  $N$  フレームまでの左右の瞳平均中心座標と、平均座標から垂直水平方向に瞳の半径の3倍を減算、加算した範囲を判別評価領域として、肌領域の閾値を利用してコントラスト改善、2値化閾値処理をする。2値化した結果、領域内には、黒目、目輪郭、眉が抽出されるが、被験者とカメラとの距離によって瞳判別評価領域が変化するため、領域内で眉と目が両方含まれる場合がある。そこで2値化処理後の画素数を計算し、瞳の半径に応じて、経験的に指定した閾値を利用して目か眉かを判定する(式(4))。

$$P_{th} = \sum_{i=-3r}^{3r} \sum_{j=-3r}^{3r} \begin{cases} P_{th} < 4r^2 & \Rightarrow otherwise \\ 4r^2 < P_{th} < 10r^2 & \Rightarrow eye \\ 10r^2 < P_{th} & \Rightarrow eyebrow \end{cases} \quad (4)$$

( $r$ : radius of extracted iris)

### 3.2 視線移動後の瞳中心座標値算出

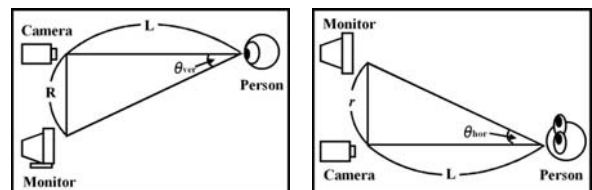
認識された瞳中心座標( $x_{in}, y_{in}$ )と、モニタ、カメラ、対象者との三者関係を用いることで、瞳の移動を行う。図6に示すような人、モニタとカメラとの位置関係を表す  $L, R, r$  なる3つのパラメータを指定する。そして、垂直視野角度  $\theta_{ver}$  と、水平視野角度  $\theta_{hor}$  を式(5)で求める。

$$\theta_{ver} = \tan^{-1} \frac{R}{L} \quad \theta_{hor} = \tan^{-1} \frac{r}{L} \quad (5)$$

次に式(6)によって、元の瞳中心座標( $x_{in}, y_{in}$ )と上で求めた  $\theta_{ver}, \theta_{hor}$  とその角度を顔画像上での pixel 値に変換する関数  $\Delta x, \Delta y$  を導入し、移動後の瞳中心座標( $x_{out}, y_{out}$ )を得る。

$$x_{out} = x_{in} + \Delta x(\theta_{ver}) : \Delta x(\theta_{ver}) = \frac{\theta_{ver}}{10} \quad (6)$$

$$y_{out} = y_{in} + \Delta y(\theta_{hor}) : \Delta y(\theta_{hor}) = \frac{\theta_{hor}}{10}$$



(a) 垂直視野角度

(b) 水平視野角度

図6 視野角度計算におけるパラメータの位置関係

### 3.3 瞳再描画によるアイコンタクト画像生成

前節で求められた瞳中心座標 $(x_{out}, y_{out})$ を基に、目領域を抽出した後、垂直・水平方向に画素値を走査させ、白目、瞳、まぶたのそれぞれの個別領域を判定し(図7)、白目 $(f_{i,j}=0)$ と瞳 $(f_{i,j}=1)$ 領域に対してのみ瞳再描画と塗り潰しを行う。次に、目領域を水平方向に走査していき、任意の点 $(s_n, t_n)$ から瞳中心座標 $(x_{out}, y_{out})$ までの距離 $d_n$ を求め(図8)、距離 $d_n$ と瞳半径 $r$ を用いて、黒画素値か白画素値を代入する(式(7))。その後、平滑化を行うことで、瞳の再描画を行った際の不自然さを抑制する。

$$\begin{aligned}
 B_{i,j} &= \text{black} + d_n \times \alpha \\
 W_{i,j} &= \text{white} + d_n \times \alpha \\
 \alpha &: \text{constant} \\
 \left\{ \begin{aligned} \text{black} &= \min \{F_{i,j} \mid f_{i,j} = 1\} \\ \text{white} &= \max \{F_{i,j} \mid f_{i,j} = 0\} \end{aligned} \right. & \left\{ \begin{aligned} d_n &= \sqrt{(s_n - x_{out})^2 + (t_n - y_{out})^2} \\ d_n &\leq r &\Rightarrow B_{i,j} \\ d_n &\geq r \cap f_{i,j} = 1 &\Rightarrow W_{i,j} \\ d_n &\geq r \cap f_{i,j} = 0 &\Rightarrow F_{i,j} : \text{pixel value} \end{aligned} \right. \quad (7) \\
 f_{i,j} &: \text{label number in Fig.4}
 \end{aligned}$$

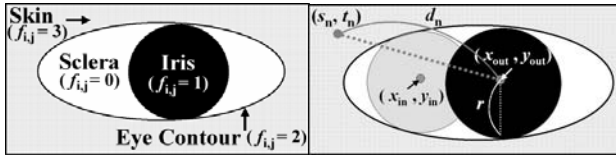


図7 目領域分割

図8 再描画領域の選定

## 4. 実験と考察

### 4.1 顔トラッキング

顔領域を抽出した例を図9に示す。実験参加者にはカメラの前で様々な動作をするように指示を出し、平均的な室内(今回は研究室内)にて特別な照明器具を使用せず行った。抽出した顔領域を基に、PTZカメラによって追跡した顔画像の結果を図10に示す。なお、これらの追跡結果画像は、抽出結果画像(図9)に同期したタイミングで取得した画像である。また、3枚の異なるテンプレート画像を用いて、マッチングによる顔領域抽出の性能評価実験を行った結果を表1に示す。

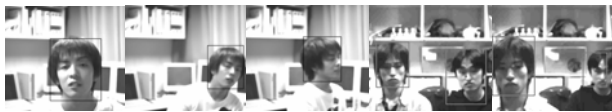


図9 顔領域抽出結果



図10 顔領域追跡結果

表1 テンプレートマッチング評価結果

Candidate of template	Rate (Success/Total)
Only frontal	97.1% (102/105)
Including profile	78.1% (246/315)

シーン中の最も遠方に存在する人物を追跡対象として、顔領域のスケール変化によってズーム機能が動作し、CCDカメラから撮像された顔領域抽出画像と比較して、顔部品

輪郭を認識できるスケールの画像が取得可能となった。テンプレートマッチングの成功率は正面顔のみの場合は良好な結果を得たが、テンプレートの正面顔に対して大きな変化のある場合には、80%以下の成功率となった。急激な顔の動作変化、激しい前後移動や人物の交錯には高い成功率でテンプレートの更新が行えないため、さらなる改善が求められる。

### 4.2 顔部品トラッキング

瞳輪郭抽出、鼻孔抽出、顔部品輪郭抽出の性能評価をソフトピアジャパンの正面顔データベース男女各150名、合計300名を用いて実験を行った。それぞれの結果を表2, 3に示す。

表2 正面顔300名瞳、鼻孔抽出結果

Facial parts	Male		
	Iris		Nostril
	Non-glass	Glass	
○	113/113(100%)	36/37(97.2%)	141/150(94.0%)
×	0/113(0%)	1/37(2.8%)	9/150(6.0%)
Facial parts	Female		
	Iris		Nostril
	Non-glass	Glass	
○	136/139(97.8%)	10/11(90.0%)	145/150(96.7%)
×	3/139(2.2%)	1/11(10.0%)	5/150(3.3%)

表3 正面顔300名顔部品輪郭抽出結果

Facial parts	Male		
	Eyebrow	Eye	Mouth
	○	100/150(66.7%)	141/150(94.0%)
×	50/150(33.3%)	9/150(6.0%)	62/150(41.3%)
Facial parts	Female		
	Eyebrow	Eye	Mouth
	○	98/150(65.3%)	144/150(96.0%)
×	52/150(34.6%)	6/150(4.0%)	41/150(27.3%)

瞳輪郭、鼻孔抽出結果は、抽出率95%以上という良好な検出結果を得ることができた。このことから正面顔については瞳、鼻孔が安定して抽出可能な特徴点であることがいえる。顔部品輪郭抽出については、目輪郭に95%という良好な検出結果が得られているが、眉に関しては、頭髮に遮蔽される、眉が薄いことを理由に検出精度が低下しているものと考えられる。口に関しては、肌とのコントラスト差が少ない、髭によるノイズの理由で検出精度が低下しているものと考えられる。

### 4.3 瞳判別評価

動画像による瞳認識実験を行い、瞳認識評価の有無による認識結果の比較、検討を行った。認識評価無しの場合では、誤認識を判定することなく、瞳以外の輪郭抽出を繰り返し行った結果が得られた。認識評価有りの場合では、認識評価無しに比べると安定した結果が得られた。導入前と比較すると処理フレーム数が減っているのが確認できる。これは、誤認識判定を導入したことにより、本来誤って瞳として認識されたフレームを省略し、正しいフレームのみを瞳として抽出したことによる。

また数値として瞳輪郭抽出成功率が向上していることがわかる(表4)。

### 4.4 アイコンタクト顔映像生成

図11(b)に瞳再描画を行ったアイコンタクト顔画像(図11(a)は原画像)の結果例を示す。実験では、正面にモニター、モニターの上側にカメラを設置し、実験参加者にはモニター上に映る自身の顔映像を目視してもらった。

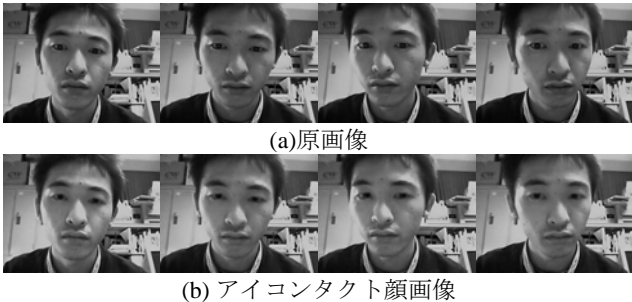


図11 瞳再描画によるアイコンタクト顔画像生成

また、実験参加者27名によるアイコンタクト顔映像のアイコンタクト性と映像の自然性を、それぞれ5段階評価にて実験を行った。評価結果を図12に示す。

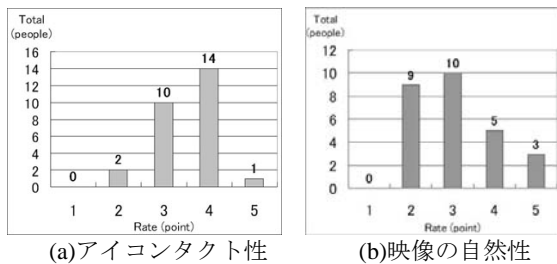


図12 アイコンタクト顔映像評価結果

生成されたアイコンタクト顔画像は、描画に不自然さが残るものの、原画像と比較するとアイコンタクトの改善が

見られる結果となった。瞳再描画における、白目と肌の領域の目領域判別に失敗し描画に不自然さが目立つ場合があった。

また、評価実験においてアイコンタクト性が平均3.51、映像の自然性が平均3.01という結果が得られたことから、アイコンタクト性の改善は評価されたことがいえるが、映像の自然性には更なる改善が必要であると考えられる。

### 5. 結言

本稿では、顔・顔部品のトラッキング、特徴抽出のための画像処理技術を構築して、これを2種類のカメラで構成した実用できるシステムに実装した。次いで、このシステムを基盤技術に用いて、アイコンタクト顔映像というヒューマンインタフェース性のよい顔メディアの創出を可能とするシステムを構築した。今後はシステムのロバスト性の向上と実利用化を目標とし、さらに表情変化を含有した顔メディア創出の発展を考えている。

### 参考文献

- [1] Ming-Hsuan, Yang, David J. Kriegman and Nerentra Ahuja, "Detecting Faces in Images", A Survey, IEEE Trans.PAMI, Vol.24, No.1, pp.34-58, (2002)
- [2] 金子正秀:"顔による個人認証の最前線", 映像情報メディア学会誌 Vol.55, No.2, pp.180-184 (2001)
- [3] Ichikawa, Y., K. Okada, G. Jeong, S. Tanaka, and Y. Matsushita: "MAJIC Videoconferencing System: Experiments, Evaluation, and Improvement", Proceedings of ECSCW'95, pp.279-292 (1995).
- [4] Buxton, W. "Space-Function Integration and Ubiquitous Media" In M. Shamiyeh (Ed.).Towards and Interactive and Integrative Design Process. Linz, Austria: DOM Publications, pp.248-271 (2006).
- [5] T. Funahashi, T. Fujiwara and H. Koshimizu: "Coarse to Fine Hierarchical Tracking System for Face Recognition", Proc. of IEEE SMC 2005, pp.3454-3459 (2005)
- [6] 舟橋琢磨, 藤原孝幸, 興水大和: "瞳認識を利用したインターフェース顔メディア生成の提案", MIRU2006 画像の認識・理解シンポジウム (2006)

表4 瞳判別評価結果

	Method	Total Frame	Limitation of eye region	Success	Failure	Rate
Left eye	Original	45	0	18	27	40%
	Proposed	35	10	23	12	65%
Right eye	Original	45	0	23	22	51%
	Proposed	35	10	26	9	74%