

hw/sw 複合体を用いた顔検出

横川 昌俊[†] 須藤一郎[†] 湯野智己[†] 関根優年[†]

[†] 東京農工大学 工学府

〒184-8588 東京都小金井市中町 2-24-16

E-mail: †{masa,ichiro,yuno}@sekine-lab.ei.tuat.ac.jp, ††sekinem@cc.tuat.ac.jp

あらまし Coarse to Fine 理論を用いて、顔検出システムを作成した。Coarse to Fine とは低解像度画像 (coarse 画像) を用いて大域的な処理を行い、その結果を反映させた高解像度画像 (fine 画像) を用いて、局所的な処理を行うというものである。カメラから得た VGA サイズ画像を、Wavelet 変換により 40*30(圧縮レベル 4) まで圧縮し、coarse 画像を作成する。この coarse 画像と同レベル解像度の顔テンプレートを使用して、画像全域に対してマッチングを行い、顔領域をサーチする。マッチング結果座標の分布を調べ、候補点の集中する領域に顔が存在すると仮定し、その付近の領域を高解像度化する。この高解像度化された領域内から、目と口のテンプレートを使用し、それぞれを探索する。目や口の有無、その位置関係から、仮定した領域が顔領域であるかの判断を行い、顔を検出する。複数の人間がいる環境下でも、coarse 画像で複数箇所顔が存在すると仮定し、一連の操作を複数回行うことで検出が行えるようにした。キーワード 顔認識、テンプレートマッチング、多重解像度解析、FPGA

Face detection with the union of hardware and software

Masatoshi YOKOKAWA[†], Ichiro SUDO[†], Tomomi YUNO[†], and Masatoshi SEKINE[†]

[†] The Faculty of Technology, Tokyo University of Agriculture and Technology
2-24-16 Naka-chou, Koganei-shi, Tokyo, 184-8588 Japan

E-mail: †{masa,ichiro,yuno}@sekine-lab.ei.tuat.ac.jp, ††sekinem@cc.tuat.ac.jp

Abstract We build a face detection system with the coarse to fine algorithm. Coarse to Fine is the algorithm which uses the coarse image for global processing and the fine image for local processing. We use the Wavelet transform to make the compressed image(coarse image). A face region is searched in the coarse image with the image of face template. We check the distribution of the result's points, and presume there is a face near the much distributed area. We decompress the presumed area to search for eyes and mouth points. And decide this is the face or not. This system can search a much faces in the image to repeat the same sequence.

Key words face detection, template match, multiple resolution analysis, FPGA

1. はじめに

1.1 研究背景

ビデオカメラやデジタルカメラのように、カメラは記録を主機能にしてきたが、近年ではロボットや車載カメラなどのように、カメラを入力センサーとして扱う装置も多く登場してきている。しかし、カメラ画像は膨大な情報量を含んでおり、その全てを利用すると処理コストが高くなるという問題がある。それを解決する手法の1つに、Coarse to Fine 理論がある。Coarse to Fine により、選択的に画像を利用することで、大幅な処理削減が可能となる。更に、解像度を上げるに連れて、対象範囲を絞り込むことで、マッチング精度も向上する [1]。認識精度・速度の双方が要求される、顔検出分野に Coarse to Fine

を適応することで、効率的な顔検出システムが実現できると考えた。

1.2 研究目的

本研究室では、hw(ハードウェア)を用いた並列・分散処理システムを提案している。それが、hw/sw 複合体である。Coarse to Fine 理論を、hw/sw 複合体を用いて実現し、顔検出を行う。高負荷な圧縮やマッチング演算などは hw で、コントロールや演算結果の判断部分は sw(ソフトウェア)で行うことで、高速処理の実現を目指す。また、1回のマッチングで複数の顔を検出するために、単純な閾値による顔領域の判断ではなく、汎用性のあるテンプレートを使用したマッチング結果の分布により、顔領域の判断を行うという手法を提案する。

2. Wavelet 変換

本研究では, Haar 関数を用いた Wavelet 変換を使用して多重解像度解析を行う. Haar の wavelet 関数を式 1 に, scaling 関数を 2 に示す. [2]

$$\psi_H(x) = \begin{cases} 1 & (0 \leq x < \frac{1}{2}) \\ -1 & (\frac{1}{2} \leq x < 1) \\ 0 & (\text{otherwise}) \end{cases} \quad (1)$$

$$\phi_H(x) = \begin{cases} 1 & (0 \leq x < 1) \\ 0 & (\text{otherwise}) \end{cases} \quad (2)$$

wavelet 関数 $\phi_H(x)$ を High pass Filter, scaling 関数 $\psi_H(x)$ を Low pass Filter として入力画像 (原画像) の水平方向・垂直方向の順に 1 次元変換を行うことで $\frac{1}{4}$ に圧縮された低周波成分 (scaling image) と高周波成分 (wavelet image) が得られる. 図 1 には, Wavelet 変換による多重解像度表現を示す. これは, 変換後の低周波数成分に対して再度変換することにより, 得られる信号の表現である.

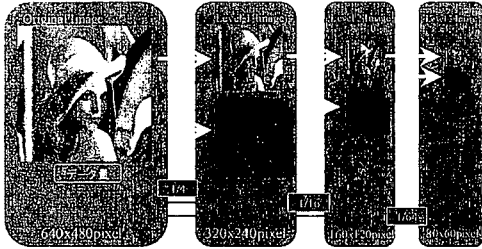


図 1 多重解像度

3. Coarse to Fine テンプレートマッチング

3.1 マッチング演算

本研究では, 差分絶対値和演算を用いてパターンマッチングを行う. 式 3 に, $A(x, y)$ をテンプレート, $B(x, y)$ をサーチウィンドウとした場合の, 差分絶対値和演算の一般式を示す. この場合, テンプレートのサイズは $m * n$ である.

$$M = \sum_{i=0}^m \sum_{j=0}^n |A_{(i,j)} - B_{(x+i,y+j)}| \quad (3)$$

これを YUV カラー形式へ応用すると, 式 4~7 のようになる. ここで, O_{β}^j がテンプレート画像, $I_{\alpha 1}^j$ がサーチウィンドウ, K_c が混合係数である. 混合係数とは, dY と dU , dV の寄与度を同等にするために導入した係数で, [1] では $K_c = 4$ に設定している.

$$dY = |Y_o(O_{\beta}^j) - Y_I(I_{\alpha 1}^j)| \quad (4)$$

$$dU = |U_o(O_{\beta}^j) - U_I(I_{\alpha 1}^j)| \quad (5)$$

$$dV = |V_o(O_{\beta}^j) - V_I(I_{\alpha 1}^j)| \quad (6)$$

$$M = \sum_{i=0}^m \sum_{j=0}^n |dY + K_c * (dU + dV)| \quad (7)$$

3.2 Coarse to Fine

多重解像度解析を用いたテンプレートマッチングのことを, Coarse to Fine テンプレートマッチングという. 粗い画像でのマッチング結果を, 高解像度画像におけるマッチングの対象領域に反映させ, 高解像度マッチングでは, 局所領域に対してのみマッチングを行うというものである. この処理のメリットは以下の通りである.

- 個人の特徴が薄れるため, テンプレートの汎用性が得やすい
- 演算量が軽減される
- ノイズの影響を受けにくい

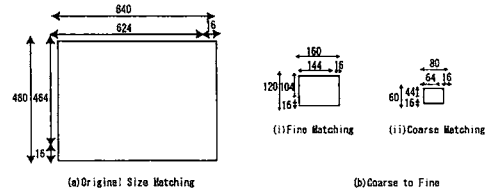


図 2 Original Size Matching と Coarse to Fine

2 番目の演算量が軽減される点に関して, 図 2 を用いて説明する. (a)Original Size(640*480) でマッチングを行う場合, テンプレートのサイズ (16*16) を除いた, 627*464 がマッチング開始点の稼動範囲となるので, その差分絶対値和の演算回数は,

$$16 * 16 * 624 * 464 = 74, 121, 216 \quad (8)$$

となる. (b)Coarse to Fine の場合, fine 画像サイズを 160*140, coarse 画像サイズを 80*60 とし, 同じくテンプレートのサイズ (16*16) を除いて考えると,

$$Coarse : 16 * 16 * 64 * 44 = 720896 \quad (9)$$

$$Fine : 16 * 16 * 144 * 104 = 3833856 \quad (10)$$

$$(9) + (10) = 4, 554, 752 \quad (11)$$

となり, 演算回数が約 $\frac{1}{16}$ になることがわかる.

図 3 に, Coarse to Fine テンプレートマッチングの模式図を示す. 図 3 では, 目を探索している. まず, 低解像度画像の状態で顔領域を探し出し, 次のレベルの画像でその範囲を更に絞り込む. そして, 高解像度画像を用いて, 目を特定するというものである.

関根らの実験結果を図 4 に示す. これによると, 注視領域の絞り込みを行う (図 4 右側) を行うことで, 各解像度において, 正しく目の位置を捉えることに成功している. 図 4 左側は, 絞り込みが行われていないために, 目の探索に失敗している. これ

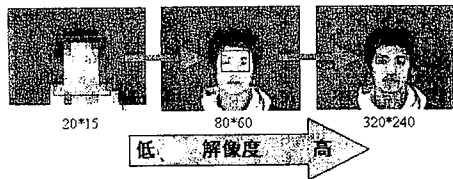


図3 多重解像度表現による領域の絞込み

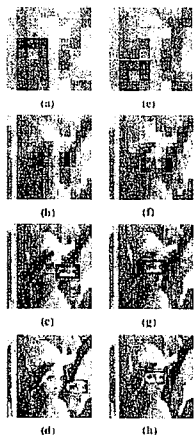


図4 目のテンプレートによる注視領域の探索 [1]

をもとに追加実験を行うと、Coarse to Fineにおける解像度レベルの移動は、複数のレベルを飛び越しても同等の効果が得られることがわかった。

本研究では低解像度画像を得るために、前項で説明した Haar 関数を基底とした Wavelet 変換を使用した。これを用いた理由は、演算が加算とシフト演算のみで表現できるため、hw 化しやすいということである。

4. hw/sw 複合体

4.1 hwModule

本研究で想定した画像認識システムを実現するために、本研究室で開発した再構成可能な LSI である FPGA が搭載された PCI デバイス、hwModule (ハードウェアモジュール、図5) を用いてシステムの構築を行った。図6には、hwModule を用いた hw/sw 複合体の実装図を示す。図6をみればわかるように hwModule には、hwObject の要素である hwNet (回路) と Data が置かれる。hwObject に関しては、次のセクションで詳しく説明する。

hwModule は FPGA、メモリ、マイクロプロセッサ、PCI バスコントローラにより構成される。図7には hwModule のブロック図を示す。hwModule は、ボード制御の PCI バスコントローラ用 FPGA を1個、仮想回路用 FPGA を3個搭載している。更に、処理データを格納するための LM (ローカルメモリ) を3個搭載している。各 FPGA は2つの LM バスを持っている。また、隣り合う FPGA 同士で LM を共有しているた

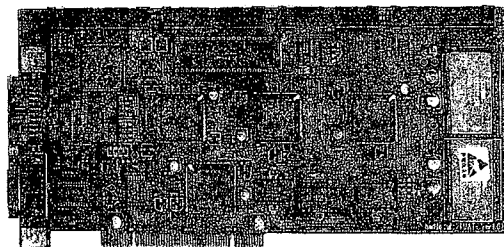


図5 hwModule

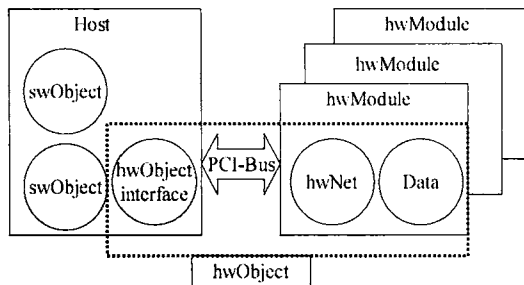


図6 hwObject と swObject の実装

め、複数の仮想回路が動作する状況でもデータのコピーの必要がないという特徴を持つ。

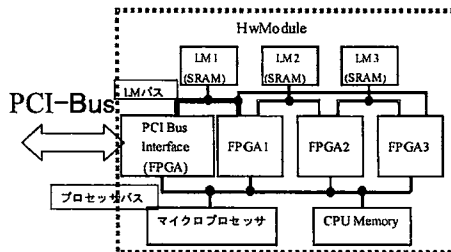


図7 hwModule 内部構成ブロック図

4.2 hwObject

hwObject とは、複雑なハードウェアの処理をカプセル化したもので、hwModule 上の FPGA に実装される hwNet により、目的の処理を行う。hwObject に於ける Host 上のメンバー関数は、対象の仮想回路と Host との結合を解決して動作させるための手続きのみである。つまり、従来の C++ 言語によるソフトウェアの記述を用いて、容易にハードウェアが扱えるようになる。hwObject には、ビット処理、信号処理など回路が得意とする機能を割り当てる。[3]

図8には、hwObject モデルの概要を示す。hwObject interface は、swObject サイドでは C++ 言語によるメンバー関数呼び出しによるアクセスを行い、hwModule サイドでは PCI バスを通したハードウェアとのデータ転送と制御を行う。

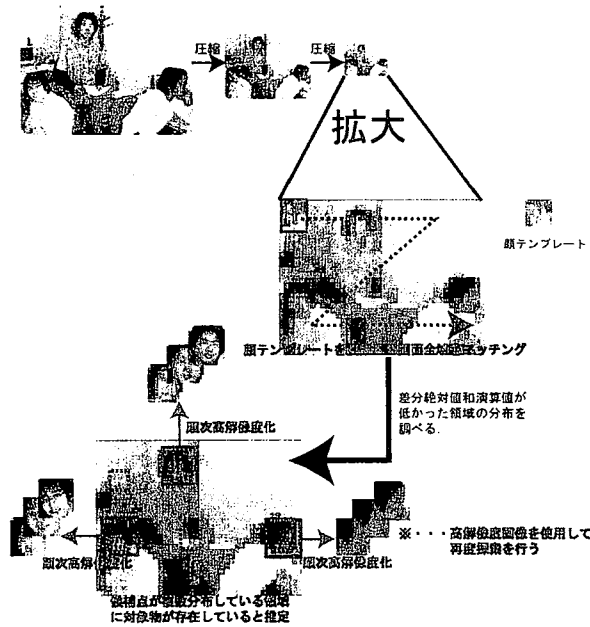


図9 マッチング結果の分布を利用して探索領域を決定

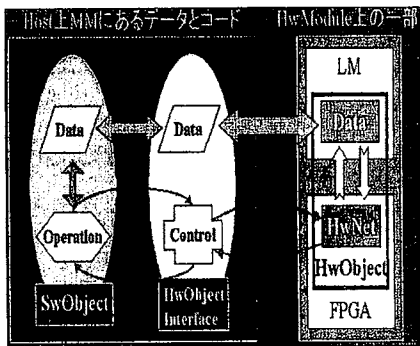


図8 hwObject モデルの概要

5. 顔検出システム

5.1 提案手法

図9に、本研究で提案する顔検出手法の概念図を示す。Wavelet変換を用いてカメラより取得した画像を圧縮する。そのcoarse画像と同解像度の顔テンプレートを用いて、テンプレートマッチングを行う。このとき、圧縮された画像で探索を行うことで、個人の特徴が薄れ1枚のテンプレートで複数の顔をサーチすることができる。更に、テンプレートマッチングでは、ある候補点の近傍数ピクセルの範囲に、別の候補点が現れるという特徴がある。特に顔を対象とした場合、肌色部分が最も大きな影響を与えるため、顔とテンプレートが数ピクセルずれただけならば、その差分絶対値和の差は小さく、候補点として出現しやすい傾向にある。その特性を利用して、得られ

たマッチング結果の分布を調べ、候補点の分布が多いところに顔が存在すると仮定する。この処理により得られる顔領域地点は、1つではなく複数地点である。複数ある顔領域全ての解像度を順次高解像度化し、目・口の有無など、詳細なマッチングを行い、最終的に仮定した領域が顔であるかの判断を行う。これにより、カメラの前に複数人いるような環境下においても、顔を検出することが可能である。

図9に示したアルゴリズムを回路にて実装する手法を、図10に示す。coarse画像を使用して顔領域がN個あると仮定したとする。その時、FPGAにN個のプロセッシングエレメント(PE)を含んだビットファイルをダウンロードする。並列に配置されたPEにそれぞれのサーチウィンドウデータとテンプレートデータを流すことにより、同時に複数領域の顔判断が可能となり、探索人数の増加による処理の増加は発生しない。

5.2 システムの構成

前小節で示した理論を使用して試作したシステムの構成を、図11に示す。

本システムで使用した回路は、Camera回路、Wavelet回路、Matching回路、Extractor回路の4つである。4つの回路の機能は、それぞれVGA画像の生成、Wavelet変換(圧縮level3まで)、テンプレートマッチング、画像切り出しである。各回路のhwModule上での配置と使用するデータのメモリ割り当てを、図12に示す。

swObjectに関して説明する。CtoFControllerは、システムのシーケンスの管理やメモリの参照先の変更などの機能を持つ。RegionDetectorは、Wavelet変換画像level4を生成し、顔領域の探索を行い、領域の候補点を返すオブジェクトである。

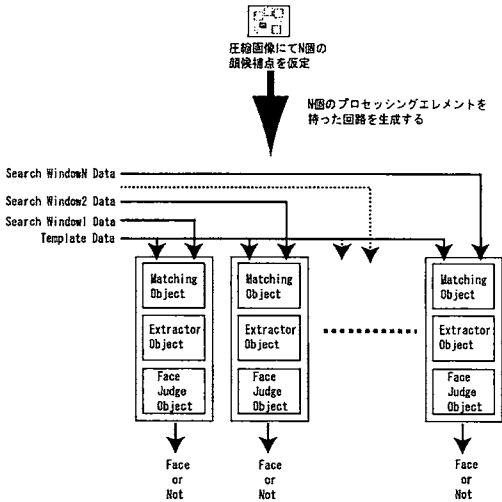


図 10 前処理に応じて生成する回路を選択

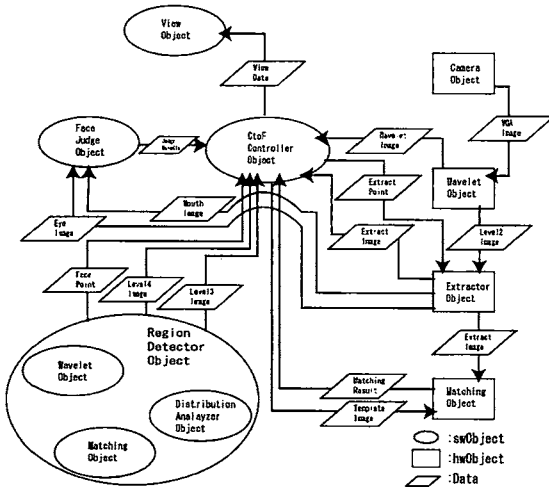


図 11 システム全体の構成とデータ

FaceJudge は、目や口の有無などから、仮定した領域が顔であるかの判断を行うオブジェクトである。

WaveletObject と MatchingObject は hw と sw の 2 種類が存在している。これは、既存の Wavelet・Matching 双方の hwObject の仕様の関係で、低解像度探索部分に関しては swObject の方が適しているからである。更に、本試作機では、図 10 に示すような並列演算は行っていない。現行の hwModule に搭載されている FPGA のゲート規模が、複数の PE を構築できるほどの余裕がないためである。ゆえに、本システムでは、1 つの PE を用いて、逐次的に複数の候補点の判定を行っている。FaceJudgeObject に関しても同様の理由で、hw 化されていない。

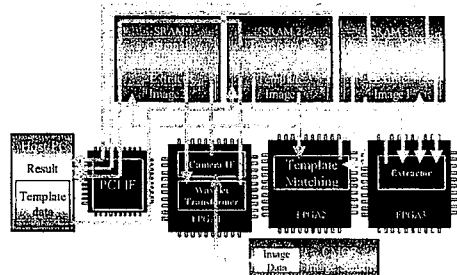


図 12 各回路の hwModule 上の配置

5.3 動作シーケンス

図 13 は、本システムの動作シーケンスである。最上部から RegionDetector までが、顔領域の仮定を行う動作で、その後仮定した各領域に対して、解像度を上げながら顔の判定を行っていく。1 番目の候補点に対する判断が終了すると、次の候補点に対してのシーケンスがスタートする。

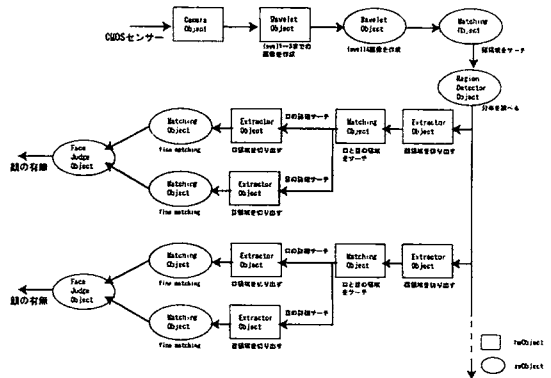


図 13 動作シーケンス

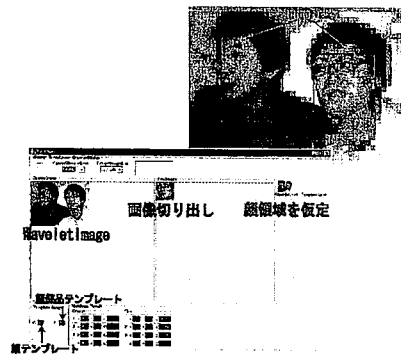


図 14 システムの GUI

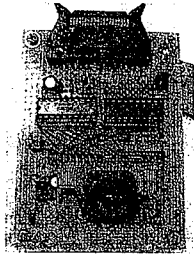


図 15 Camera Module



顔のテンプレート 目のテンプレート 口のテンプレート

図 16 テンプレート画像

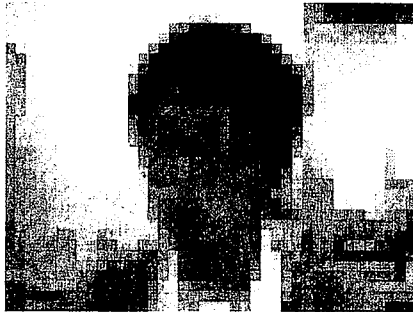


図 17 顔検出

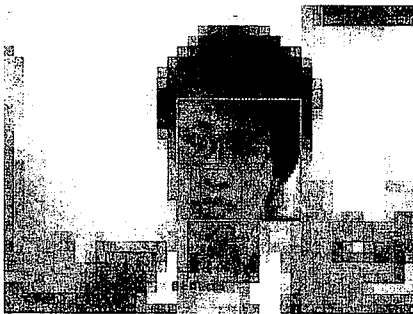


図 18 目と口の検出

6. 検証

試作したシステムの動作検証を行った。検証では、顔判断の要素部分である、目と口の検出に関して行った。入力画像の取得には、図 15 の CMOS センサを使用した。用いた色表現は YUV 形式 16bit(Y:6bit, U:5bit, V:5bit) である。

図 16 には、検証に使用した各種テンプレートを示す。図 17 が顔領域のサーチ結果を、図 18 が目と口のサーチ結果を示している。両図より、level4 の coarse 画像中から顔を検出し、level2

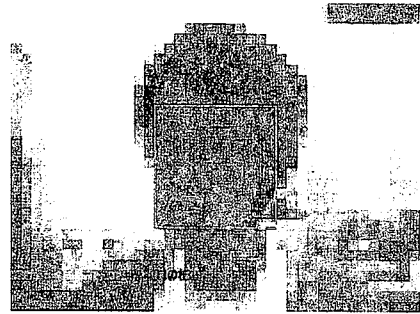


図 19 口の検出に失敗

画像より切り出された顔領域中から、目と口を検出できていることがわかる。更に、カメラ前に 2 人いた場合の結果は、図 14 の GUI の図で示している通りで、2 人の場合にも期待通りの検出が成されることがわかった。

図 19 は、顔部分のサーチし失敗したケースである。顔の領域の抽出には成功しているが、口のサーチに失敗している。今回の検証では、入力画像を正規化するなどの前処理を行っていないため、顔のパーツの検出に失敗した例が多くみられた。

7. まとめ

Coarse to Fine, マッチング分布による顔領域の仮定と、仮定数に応じて再構築させた回路を用いた顔判定、による顔検出手法を提案した。

精度の面では、低解像度レベルのマッチングでは良好な結果が得られた。しかし、高解像度レベル、今回ならば顔部品のサーチに関しては、輝度や撮影位置の影響を受け易いため、顔のパーツ検出部分は満足な結果を得ることはできなかった。高解像度レベル時のテンプレートの汎用性の確保や、状況に適応したテンプレート切り替え機構の導入が必要である。本報告書では、FaceJudgeObject を用いての顔判断までの検証には至らなかったが、FaceJudgeObject の精度を高めるためにも、顔パーツ検出部分の改良が不可欠である。

今回の報告では回路規模の制約により、「仮定数に応じて再構築させた回路」の実装はできなかった。それゆえに、本試作機では逐次動作が必須となり、処理速度の面では十分な成果が得られなかった。しかし、次世代 hwModule ならば十分なスペックを有しており、今回の提案機能を実装可能で、大幅に処理速度が向上が見込まれる。

文献

- [1] 関根優年, 金丸隆志, 伊藤光, “多重照合による顔領域の特定” 信学論, vol.J86, no9, pp.969-973, Sep 2003
- [2] ウェーブレットビギナーズガイド, 榊原進 著, 東京電機大学出版局
- [3] 工藤健慈, 今中晴記, 志賀裕介, 関根優年, “hw/sw 混載システムにおける hwObject モデルとその制御手法”, 情報科学技術フォーラム一般講演論文集第 1 分冊 FT2002, C-3, pp.193-194, 2002/9.