コンピュータビジョンと
イメージメディア 100－7
（1996．5．24）

顔表情のコンピュータ認識とＡＨＩへの対応

Computer Recognition of Facial Expressions and Its Application to AHI

原　文雄*　小林　宏**　丹下　明***　○飯田 史也*

*東京理科大学、　**チューリヒ大学、***ソニー（株）
Fumio Hara, Hiroshi Kobayashi, Akira Tange and Fumiya Iida

This paper deals with the real-time distance measurement and the real-time recognition of six basic facial expressions. In order to measure the distance between human being and robot in real-time, we use the transputer for parallel processing and two CCD cameras equipped in the eyeballs of robot. By using the parallax of human images obtained by two CCD cameras, we measure the distance between human being and robot, and find that the average error ratio is under 4[%] in about 80[ms] for onr measurement process. In order to obtain the center position of both pupils, we obtain the brightness by using a CCD camera, along a vertical line crossing over the pupil and eyebrow as base data and calculate the cross-correlation between base data and that in the given image. We extract the position of right and left pupils separately. By using transputer, the time needed is about 40[ms] to obtain right and left pupil's position. As the facial information for utilizing the recognition of facial expression, we use brightness data of 13 vertical lines (facial information), determined empirically and including the areas of eyes, eyebrows and mouth. Then we acquire the facial information of 6 basic facial expressions for 30 subjects whose face images have already been obtained. Since we use a layer-type neural network for recognition of facial expressions, facial information for some of 30 subjects is used for training the neural network and recognition tests done by using facial information not used for neural network learning. We find that, when we use 15 subjects for the network training, the correct recognition ratio reaches 85[%], and the total time for detecting right and left pupil positions plus the recognition of facial expression is about 55[ms] per one recognition cycle.

Key Words: Active Human Interface, Facial Expressions, Real-Time Distance Measurement, Real-Time Recognition, Robot-Human Communication, Neural Network, Cross-correlation, Transputer

１．はじめに

コンピュータやロボットに代表される知能機械は，今後，人間と能動的にコミュニケーションできるインタフェースを備えることが必要であろうと著者らは考え，Active Human Interface：ＡＨＩという概念を提案してきた[1]-[8]．ＡＨＩは人間と知能機械とが双方的に情報を伝達するためのインタフェースで，少なくとも以下の３つの機能が必要と考えられる．すなわち，

1) 知能機械が人間の感情，強いては「心」を理解する，
2) 1)での理解をもとに，知能機械が人間に対してどのような反応をすれば良いか決定する，
3) 知能機械が2) の結果を人間に分かりやすく表示する．

1)，2)，3)はそれぞれ，センサ部，コントローラ部，アクチュエータ部と言い換えることもできよう．

さて，人間対人間のコミュニケーションにおいて，メッセージの伝達に顔の表情が重要であることは広く認識されているが[9]，人間と知能機械とのコミュニケーションにおいても，顔表情は同様にして非常に重要なメディアであると考えられる．そこで，著者らはＡＨＩという新しいパラダイムを実現するためのコミュニケーションメディアとして「顔表情」を取り上げ，1)，2)，3) のそれぞれについて検討してきた．1)に関しては，静止顔画像を対象として，「驚き」，「恐怖」，「嫌悪」，「怒り」，「幸福」，「悲しみ」の６基本表情[10]の階層型ニューラルネットワーク（ＮＮ）による認識[1]，６基本表情の強さの認識[2]，それらの混合表情の認識[3]，６基本表情のＮＮ認識における特性分析[11]について報告した．また，動的顔画像に対しては，人間による静止画と動画の表情認識の違いを検討すると共に，リカレントニューラルネットワーク（ＲＮＮ）による動的顔表情認識の研究成果を報告した[4][5]．2)に関しては，調和理論[12]を用いた人工感情モデルについて報告した[13]．3)に関しては，コンピュータ・グラフィックスにより人工感情に応じた顔表情をＣＲＴ上に表出できるシステム[6]，及び人間と同様な顔を持ち表情を表出する「顔ロボット」[7]と，その実時間表情表出について報告した[8]．

ところで，上述した静止画を対象とした６基本表情の認識では，３０人の６基本表情を表す顔画像を用い，１５人分の６基本表情をＮＮに学習させて残りの１５人の６基本表情を認識した結果，９１．２[%]の認識正解率を得ている[1]．また，顔表情の時系列変化を考慮したＲＮＮを用いた６基本表情認識では，「中立」から「ある基本表情」までの過程において，人間とＲＮＮの認識率の差は平均８[%]となり，両者はかなり良く一致していることが明らかとなっている[5]．これらの研究では，著者らが予め抽出した顔の特徴点（後述）の座標を用いて表情認識を行っており，特徴点の自動抽出は考えていなかった．しかし，人間と知能機械とのインタラクティブなコミュニケーションを考えた場合，自動化した表情認識が必要であるし，人間のコミュニケーション負担の軽減のためにも実時間で表情認識をすることが望ましいと考えられる．

2. トランスピュータによる実時間瞳座標の抽出

Fig.1に実時表情認識のためのトランスピュータ（Ｔ８０５：ＴＲＰ）の構成と各ＴＲＰの処理内容を示す．図に示すように，本研究では合計８個のＴＲＰと小型ＣＣＤカメラを用いる．

瞳から鉛直線に沿って眉までの顔画像の輝度値の分布は，顔が多少回転しても，瞳の鉛直上方には瞼，その上に眉があり，瞳の黒，瞼の肌色，眉の黒という輝度値のパターンの配列は変わらず，安定してこの配列が抽出できることが予想される．また，線上の輝度値の分布を調べるだけなので，２次元のテンプレートマッチングなどと比べて計算量が少なくてすむという利点が考えられる．そこで，**Fig.2**に示すような瞳中心から眉までの鉛直線に沿った輝度値の分布パターンを予め１０人の中立の顔の平均として用意し（以降，これを「基底：base」と表記する），これと画像全体についての輝度値の分布パターンとの相互相関（積和）により，画像中で最も基底と類似している場所を検索する方法を検討した．

Fig.3に規格化をしない場合の基底を示す．基底の長さは50[pixel]（=i）に相当し，**Fig.5**においてほぼ35[pixel]の位置が瞳の中心座標となる．相互相関に基づく処理を時々刻々入力される顔画像に対して実行し，連続的に瞳の中心座標を自動抽出する．このようなアルゴリズムを組み込んで構築した画像処理システムを用いた場合，１回の処理時間は約40[ms]であり，ほぼビデオレート（1/30[s]=33.3[ms]）で瞳座標を抽出するできることが確認された．求まった瞳座標の精度は、顔の動きに対して得られた絶対誤差の１０秒間についての平均を求め，その結果を被験者毎，**Table 1**に示す．これより，瞳中心座標の抽出について機械と人間との差は3[mm]程度であり，これは瞳の半径より小さいことから，本システムによる瞳中心座標の自動抽出が約40[ms]という速さで精度良く実現されていることが分かる．

なお，６基本表情のように顔表情が変化した場合でも，瞳と眉の２部位について輝度値が下に凸というパターンは変わらないので，問題なく瞳中心座標が抽出できたことを注記しておく．また，**Fig.3**の基底を用いて，上述の３人の被験者について，ＣＣＤカメラから0.8[m]，1.2[m]の位置で瞳中心座標の抽出を行った結果においても上述と同様の精度が得られており，本手法は被験者とＣＣＤカメラとの距離が±20[%] 程度の変動をしても十分対応できることが確かめられた．

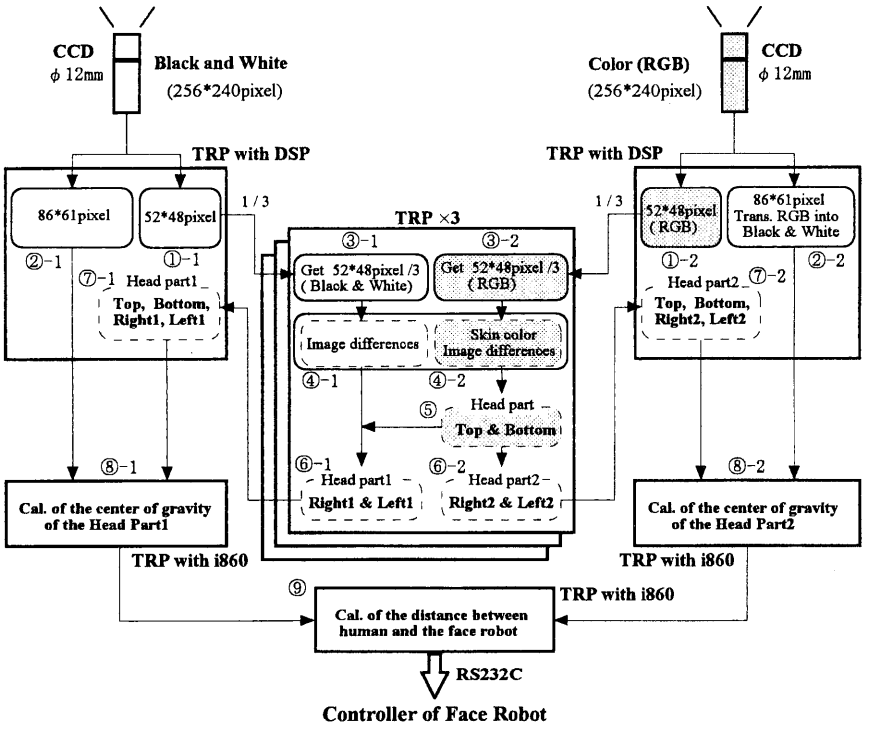

Fig. 1 Structure of TRP network

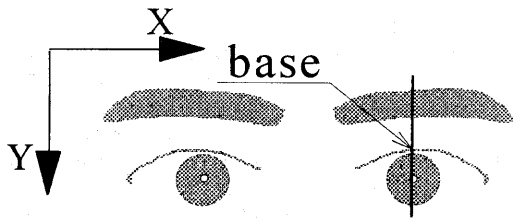

Fig.2 Model of "base"

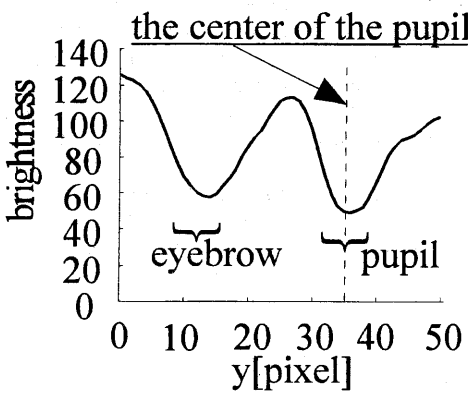

Fig.3 "base": defined as the average brightness distribution over 10 subjects

Table 1 Average of absolute measurement error

unit : mm

subject (pupil diameter)	right-left rotation	up-down rotation	right-left inclination	parallel disp.
A (9mm)	2.529	2.911	3.050	2.692
B (8.5mm)	2.714	1.907	1.682	1.861
C (7.5mm)	2.847	1.705	2.701	2.684

3. 実時間表情認識のための表情情報の獲得

左右の瞳中心座標は精度良く自動抽出できるので，その座標点を基準にして，表情の変化を良く表す顔部位の造作：眉と目，及び口を含んだ領域[1]を指定する座標値（pixel）を経験的に定める．それぞれの領域内において，眉，目，口と肌との境で輝度値の変化が大きいのは水平方向の境界であり，表情の変化はその境界の上下方向の変位となって実現される．そこで，顔の各造作部領域での鉛直線に沿った輝度値の分布を表情情報として採用する．

表情情報を実際に構成するにあたり，全ての水平方向，すなわち x 方向のpixel点に対応して鉛直方向輝度値の分布を採用すると，情報の量が非常に多くなるので，表情の特徴を表すために必要なものだけを選択する必要がある．ところで著者らは，現在までの表情認識に関する一連の研究において，**Fig.4** に示す３０特徴点を用いてきた[1]-[4]．これらの点は，顔の造作と肌との境界上の点であり，その移動によって表情が良く表現されると考えられる点である[1]-[4]．従って，これらの点を含む鉛直線に沿った輝度値分布を用いれば，表情の特徴が良く表現されるものと考えられる．そこで本研究では，**Fig.5** に太い実線で示すように片目当たり４本，口５本，合計１３本の鉛直線に沿った輝度値の分布を表情情報として用いることにする．特徴点は目頭の座標を基準にして決められているが，**Fig.5**の場合は瞳中心座標を基準にしているために，特徴点とこれらの鉛直線は正確には一致してはいないが，ほぼ対応している．また，本研究では特徴点の座標抽出は行っていないので，ａ１やａ２３等の眉，目，口の両端の座標は特定できず，それらを含む鉛直線の x 座標は決定できないので，ほぼそれらに対応する位置に輝度値分布のための鉛直線（鉛直線番号１，２，１２，１３）を配置した．**Table3**に，顔の特徴点（FCP）と輝度値分布のための鉛直線（vertical line）の対応関係を示す．

次に，**Fig.5**に示すように顔の大小の違いを補正するために，左右の瞳中心の距離Ｘが20[pixel]になるように顔画像をアフィン変換する．このＸと瞳中心座標とを用い，異なる被験者の顔表情の変化に対しても上述の鉛直線上に必ず眉，目，口がくるように，経験的に鉛直線の長さを決定する．これらの鉛直線と瞳の位置関係は**Fig.5**に示す通りで，鉛直線の長さは18[pixel]となる．鉛直線は13本あるので合計234[pixel]点での輝度値を表情情報とする．この際，鉛直線番号１，２，，，１３の順にそれぞれの輝度値を順番に並べて表情情報をマトリクスとして構成する．

６基本表情の各表情から得られた上述の表情情報を「中立」の顔画像から得られた表情情報と比較し，各表情の特徴を強調する方法を検討した．

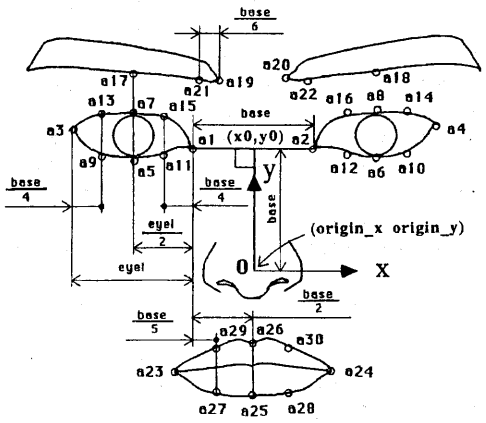

Fig.4 Facial characteristic points(FCPs)

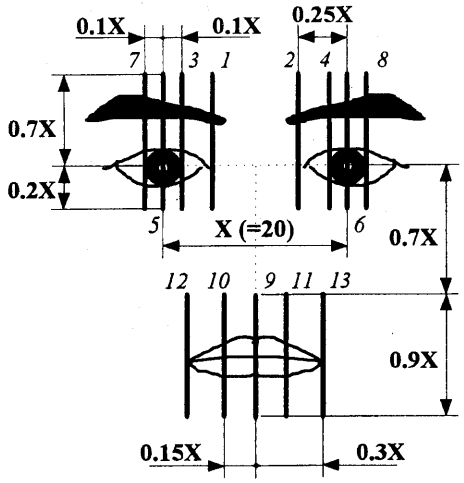

Fig.5 13 Vertical lines for obtaining facial information

4. ニューラルネットによる6基本表情認識

4.1 ニューラルネットの構造

著者らの人の顔の表情認識に関する一連の研究[1]-[3]の成果を応用して，本研究では**Fig.6**に示す階層型のＮＮを用いる．各ニューロンはシグモイド関数の非線形入出力関係を有し，入力層のユニット数は表情情報に対応して２３４ユニット，出力層は６基本表情に対応して６ユニットとする．中間層のユニット数は５０であるが，この数は，ユニット数を２０から１０ずつ増やし，最終的な認識結果が安定したときのユニット数である．

4.2 ニューラルネットの学習

ニューラルネットの学習にはバックプロパゲーション法を用いる[12]．著者らはすでに日本人２０人の被験者の６基本表情について，「中立」から「６基本表情」に至る顔表情変化をビデオテープに録画した動画像とコーカシア人１０人分の静止顔画像を収集しており[1]，その３０人分の表情情報を用いてＮＮの学習と認識テストを行う．ニューラルネットの学習には，表情の時系列変化は考慮せず，複数の被験者が最終的に表出した６基本表情から得られた表情情報を用いる．認識テストにはＮＮの学習に用いなかった被験者の表情情報を用いるので，ＮＮの学習に用いる被験者数は，最大でも３０人の半数の１５人とする．本研究では，３０人分の６基本表情の中から１０人と１５人をランダムに選んで学習用表情情報を構成し，それぞれを１０M及び１５Mと表記する．

３章で説明した表情情報は輝度値のマトリクスであり，その値は０から２５６の範囲の値となるが，ＮＮの入力データには，各表情情報について，それぞれの最大値が１，最小値が０になるように規格化されたものを用いる．また，学習の教師信号は，学習させる表情情報が表す基本表情に対応するＮＮの出力ユニットの教師信号を１とし，他の出力ユニットの教師信号を０とする．そして，学習させる全ての表情情報についての教師信号と，出力ユニットの出力値との自乗和誤差が0.001 以下になった場合に学習を終了する．

4.3 認識実験と結果

顔表情の認識実験では，ニューラルネットの学習に用いなかった被験者の「中立」から「６基本表情」に至る顔表情の変化を録画したビデオテープを再生し，その画像をＣＣＤカメラで画像処理システムに時々刻々入力し，瞳中心座標の抽出，表情情報の獲得，ＮＮへの表情情報入力及び表情認識を全て自動的に行う．認識テストに用いるニューラルネットは，4.2節で説明したように予め学習したものを用いる．

本システムでは，瞳中心座標を抽出した後，それを用いて表情認識を行うまでの時間は最大15[ms]であった．従って，画像入力から表情認識は55[ms]以内で実現できることが判明した．ところで，このように自動表情認識を実現したが，１回の画像入力から表情認識までに要する時間が一定ではないことから，ビデオテープの顔画像とその表情認識結果との対応関係が分からなくなってしまう．そこで，ビデオレートの２倍の2/30[s] (66.7[ms]) 毎に顔画像を本システムに取り込むようにした．

ところで，「中立」から「６基本表情」に至る途中の段階では，表情の種類を限定することが困難であり，また複数の表情が含まれているように見える[10]．このことから，表情表出の途中の段階での本システムの表情認識性能を評価することは困難であると考えられる．そこで，著者らがすでに報告した６基本表情の認識[1]と同様に，最終的に表出された６基本表情の認識性能を調べることにする．すなわち，最終的な６基本表情に対応する顔画像について，ニューラルネットの出力値が最も大きいユニットに対応した基本表情をニューラルネットの認識結果とし，その基本表情と顔画像が本来示す基本表情の種類とが一致する場合を正解とする．

このようにして得られた正認識と誤認識の割合の分布，及び，正解率の６基本表情についての平均値を**Table 2**に示す．この表は，学習情報15[M]を例にすると，「驚き」の正解率は90[%]で，「恐怖」に10[%]誤認識をしていることを表す．これらの表の対角線の成分が正解率を示し，この平均値が６基本表情についての平均値である．この結果より，ニューラルネットの学習において，学習人数を増やすことにより正解率が上がり，１５人の被験者を学習に用いた場合には平均85[%]の正解率が得られていることが分かる．Bassillら[31]は，表情認識の訓練を受けた人の６基本表情の正解率は約87[%]であったと報告しており，このことより，本手

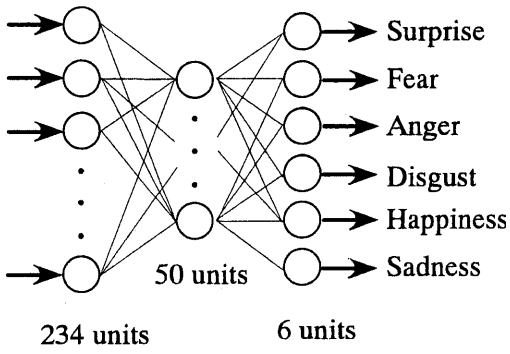

Fig.6 Structure of neural network

法は人間と同程度の認識能力を有すると言える．

5. まとめ

本論文では，人間と知能機械とが双方的にコミュニケーションできるインタフェースを実現するために，知能機械が人間との距離を実時間で計測する技術，及び人間の顔表情を自動的に実時間で認識する方法を検討した．まず，距離計測に関しては，差分と色情報により人間の頭部領域を画像から抽出し，その頭部領域の中の黒領域の重心位置を２個のＣＣＤカメラより得られた画像についてそれぞれ求め，その重心位置の違いを視差として距離計測を行った．トランスピュータを用いた並列処理により，１．５[m]～５[m]の範囲で平均誤差４[%]以内，１回の計測時間約８０[ms]を実現した．次に，表情認識に関しては，顔の基本構成要素である瞳と眉の輝度値の分布パターンを用い，その基準パターンとの相互相関を計算することにより顔画像中から瞳の中心座標を算出した．その座標を基準にして表情の変化を良く表す顔の造作である眉，目，口の領域を経験的に限定し，それらの顔の造作の特徴を表す１３本の鉛直線上の輝度値分布を用いて表情情報を構成した．そして，１５人の被験者の最終的に表出された６基本表情についての表情情報を用いてニューラルネットを学習し，他の１５人の被験者の表情情報をこのニューラルネットで認識した結果，最終時点で表出された６基本表情について85[%]の認識正解率が得られ，人間と同程度の認識性能を有することを確認した．また，トランスピュータを用いた本表情認識システムは，画像入力から表情認識までを上記の手法により全て自動的に55[ms]以内で実現した．この顔表情の自動認識の速度は，人間との実際のコミュニケーションにおいて十分な速さであるものと考えられる．

本研究では，最終的に表出された６基本表情の認識結果から定量的に本システムの表情認識手法を評価したが，表情が表出される途中段階での表情認識に関する検討が必要であると考えている．さらに，著者らは人間と同様な顔を有し表情を表出する「顔ロボット」の開発[7]，及びその実時間の表情表出をすでに報告している[8]が，実時間の表情認識と表情表出を融合したシステムを構築し，実際に人間とのコミュニケーションインタラクションを行うための研究を今後進めていく予定である．

Table 2 Recognition results

Training information: **15M**

Recognized emotion	Input emotion: Sur.	Fear	Dis.	Ang.	Hap.	Sad.
Sur.	90	10	0	0	0	0
Fear	10	90	0	10	0	10
Dis.	0	0	60	10	0	0
Ang.	0	0	40	80	0	0
Hap.	0	0	0	0	100	0
Sad.	0	0	0	0	0	90

average : 85.0%

参考文献

[1] 小林，原：“ニューラルネットによる人の基本表情認識”，計測自動制御学会論文集，Vol. 29, No. 1, pp. 112-118, 1993.

[2] 小林，原：“ニューラルネットによる人の顔の６基本表情の強さの計測に関する研究”，日本機械学会論文集(C編)，59巻567号，pp. 177-183, 1993.

[3] 小林，原，池田，山田：“ﾆｭｰﾗﾙﾈｯﾄﾜｰｸによる動的な基本表情認識”，電子情報通信学会技術研究報告，HC92-59, pp. 11-16, 1992.

[4] Kobayashi and Hara : "Dynamic Recognition of Basic Facial Expressions by Discrete-time Recurrent Neural Network", Proceedings of International Joint Conference on Neural Network, pp.155-158, 1993.

[5] Hara and Kobayashi : "Computer Graphics for Expressing Robot-Artificial Emotions", Proceedings of IEEE International Workshop on Robot and Human Communication, pp.155-160, 1992.

[6] 小林，原，内田，大野：“アクティブ・ヒューマン・インタフェース（ＡＨＩ）のための顔ロボットの研究”，日本ロボット学会誌学術論文, Vol. 12, No. 1, pp. 155-163, 1994.

[7] 小林，原，後藤：“顔ロボットの動的表情表出のためのセンサ・アクチュエータの開発とその制御”，第１２回日本ロボット学会学術講演会予稿集，pp. 651-652, 1994.

[8] 本名，他訳：“ノンバーバル・コミュニケーション”，pp. 237, 大修館書店, 1981.

[10] 工藤　力訳，P.Ek[m]an and W.V.Friesen著：“表情分析入門”，pp. 1-277, 誠信書房, 1988.

[11] 小林，原：“人の顔の６基本表情のニューラルネットワーク認識における特性分析”，日本機械学会論文集(C編)，61巻582号，pp. 340-347, 1995.

[12] D.E.Rumelhart and J.L.McCelland : "PARALLEL DISTRIBUTED PROCESSING" ,pp.546-611,The MIT Press, 1986.

[13] Mogi and Hara : "Artificial Emotion Model for Human-Machine Communication by Using Harmony Theory", Proceedings of IEEE International Workshop on Robot and Human Communication,pp.149-154, 1992.

[14] E.T. ホール（日高・佐藤訳）：“かくれた次元”，みすず書房，1970.

[15] N.L. Ashton, M.E. Shaw and A.P. Worsham: "Affective Reaction to Interpersonal Distances by Friends and Strengers", Bulletin of the Psychonomic Society Vl.15 (5), pp.306-308, 1980.

[16] 斎藤　勇著：“対人心理の分解図”，pp. 21-23, 誠信書房，1989.

[17] 間瀬，渡部，末永：“ヘッドリーダ:画像による頭部動作の実時間検出”，電子情報通信学会論文誌D-II, Vol. J74-D-II, No. 3, pp. 398-406, 1991.

[18] 長谷川，横沢，石塚：“自然感の高いビジュアルヒューマンインタフェースの実現のための人物動画像の実時間並列協調的認識”，電子情報通信学会論文誌D－II, Vol. J77-D-II, No. 2, pp. 236-245, 1992.

[19] 小林，橋本：“まばたきを手掛かりとした顔の特徴点自動抽出とモデリング”，信学技報，HC94-53, 1994.

[20] 末，李，徐，辻：“部分特徴テンプレートとグローバル制約による顔器官特徴の抽出”，電子情報通信学会論文誌D-II, Vol. J77-D-II, No. 8, pp. 1601-1609, 1994.

[21] 塚本　明利，李　七雨，辻　三郎：“複数のモデル画像による顔の動き推定”，電子情報通信学会論文誌D-II, Vol. J77-D-II, No. 8, pp. 1582-1590, 1994.

[22] 保黒　政大，長坂　保典，梅崎　太造，鈴村　直夫：“ニューラルネットワークによる顔の特徴点検出の評価”，信学技報，PRU94-75, 1994.

[23] 赤松，佐々木，添町，末永：“濃淡画像マッチングによるロバストな正面顔の識別法−フーリエスペクトルのＫＬ展開の応用−”，電子情報通信学会論文誌D-II, Vol. J76-D-II, No. 7, pp. 1363-1373, 1993.

[24] 末　柄卓，小沢　慎治：“時系列顔画像処理による個人の認識”，信学技報，PRU93-66, HC93-40, 1993.

[25] 福田　敏男，伊藤　茂則，新井　史人：“ファジィ推論とニューラルネットワークを用いた人物画像の認識に関する研究”，日本機械学会論文集（Ｃ編），59巻588号，1993.

[26] 上野，加藤，中村，岡：“等濃線分布に基づく正面顔画像の識別”，電子情報通信学会論文誌 D-II, Vol. J76-D-II, No. 3, pp. 494-506, 1993.

[27] M. Suwa et al. : "A Preliminary Note on Pattern Recognition of Facial Emotional Expression", The 4th International Joint Conference on Pattern Recognition, pp.408-410, 1978.

[28] 寺村，茅：“顔の表情とその認識に関する研究”，電子情報通信学会技術研究報告，Vol. PRL81-72, p. 41-47, 1991.

[29] 石井，岩田：“コンピュータ画像処理を利用した顔の表情の自動認識”，日本機械学会論文集，Vol. 52, No. 483, pp. 2989-2992, 1986.

[30] 松野，李，辻：“ポテンシャルネットとＫＬ展開を用いた顔表情の認識”，電子情報通信学会論文誌D-II, Vol. J77-D-II, No. 8, pp. 1591-1600, 1994.

[31] J.N. Bassill: "Emotion Recognition: The Role of Facial Movement and The Relative Importance of Upper and Lower Areas of Face", Journal of Personality and Social Psychology, 37-11, pp.2049-2058,1979.