

自然な表情合成のための頭部高精細ワイヤフレームの構成と その階層的制御について

†上野雅俊 †小野英太 †森島繁生 ‡原島 博

†成蹊大学工学部

‡東京大学工学部

概要

人物の頭部形状を3次元モデルによって近似し、これに人物から得られたテクスチャを貼り付けることによって、人間の表情や会話シーンなどのリアルな画像を合成する研究を進めている。本論文では、対象人物の頭部形状をより忠実に表現できる高精細3次元モデルを構成し、このモデルと顔画像の分析・合成の分野で広く用いられている正面顔画像用の3次元モデルとのリンクを取ることによって、従来まで研究されてきた表情・口形合成の成果をそのまま活かしながら階層的に高精度モデルを制御し、より自然な顔画像を合成する手法について述べる。

A Construction of High Definition Wire Frame Model of Head and Its Hierarchical Control for Natural Expression Synthesis

† Masatoshi UENO † Eita ONO † Shigeo MORISHIMA

‡ Hiroshi HARASHIMA

† Faculty of Engineering, Seikei University

‡ Faculty of Engineering, University of Tokyo

Abstract

A hierarchical control scheme of New High Definition Wire Frame Model of head (HD model) to synthesize natural facial expression and motion is presented. This model is constructed by 3-D digitizer and has about 24000 polygonal elements enough to express delicate facial expressions and mouth motions. This HD surface model has the texture of each point too, but it doesn't have feature points' positions on the face explicitly. So our general wire frame model which has 500 polygons and can give feature points on the face is adjusted to the 2-D front image synthesized by rotation of HD model and texture mapping. After this process, each point of HD model is controlled automatically by the movement of the point of general face model.

□ 1. はじめに

テレビ電話やテレビ会議システム、臨場感通信等を目的とする分析合成画像符号化、コンピュータと人間とのコミュニケーションを円滑化する知的インタフェースの表示部として人物顔画像をリアルに合成する研究は近年盛んに行なわれている。

顔画像の合成方式は、大別すると2種類に分けられ、一方は人物頭部モデルから肌、目、唇のテクスチャまでパラメータ設定し、それらに任意の光源を当てた時の輝度値を計算して画像を合成する、コンピュータグラフィックスによる手法[1]。もう一方は、対象となる人物の正面顔画像を用い、そのテクスチャを3次元のモデルに貼り付けるテクスチャマッピングによる手法[2]である。

前者の手法は、3次元のモデルと合成したテクスチャを持ち、任意の照明環境を設定し、その環境下で任意の方向から見た画像を合成できる[1]。また、頭部のモデルを骨格、筋肉、皮膚の3つの層構造で近似し、表情や口形などによる顔表面の変化を実際の人間の表情変化の仕組みに基づいて、骨格の変化が筋肉によって皮膚に伝わるものとして計算し、リアルな動き(表情・口形)を持つアニメーションを合成する手法[3]が提案されている。しかし、人物頭部の3次元構造から肌等のテクスチャに至るまでデータとして入力する必要があるので、実際の人物のように生きている顔を合成するのは今のところ困難である。そこで、実際の人物顔画像をテクスチャ解析し、そのテクスチャを元に画像を合成する手法も提案されている[4]。

これに対し、後者の手法では対象人物の2次元のテクスチャ(正面画像を扱ったものが多い)を実際に行っているため、特定人物の顔画像合成は容易で、かつ、かなり自然な画像を得ることが可能である。また、FACSと呼ばれる心理学の分野で提案されている表情の記述単位を用いて、表情アニメーションを作成している[5][6]。しかし、正面顔画像に標準となる3次元のモデルをあてはめ、そのモデルに2次元のテクスチャを貼り付けているので、対象人物の構造と3次元モデルの構造とは必ずしも一致してはいない。また、正面顔画像を用いる場合、頭部の前半分のテクスチャのみ取得可能のため、頭部の後半分は合成できない。そこで、正面テクスチャ以外に対象人物の全周囲テクスチャと3次元構造を同時に取得し、合成する手法が提案されている[7]。

全周囲画像と3次元構造を同時に取得する手法では、従来の標準3次元モデルに比べ、構成するパッチの数も数百倍で、その構造も対象人物と同一の3次元モデルを作成でき、さらに、任意の視線方向から見た画像を合成できるという利点がある。

しかし、この高精度な3次元モデルは人間の表情変化や、口形の変化に合わせてモデルを変形するのは困難である。これは、3次元モデルを構成する各格子点に、人間の顔表面での位置づけがされていないためである。そこで、高精度3次元モデルの各格子点を従来使われていた標準3次元モデルを用いて、表情や口形に合わせて変形する手法が提案されている[8]。

本論文では、人物頭部の高精度3次元モデルを、顔画像の分析・合成の分野で広く用いられている正面顔画像用の3次元モデルとのリンクを取ることによって、従来まで研究されてきた表情・口形合成の成果をそのまま活かして、高精度モデルを変形し、より自然な顔画像を合成する手法について述べる。また、人物の実際の表情による顔表面の動きを、高精細3次元モデルを用いて分析を行なった結果についても述べる。

まず、2.で人物の頭部の形状を近似する3次元モデルの特徴や問題点について述べ、3.では高精細3次元モデルを従来までの3次元モデルを用いて階層的に制御する手法について、4.では高精細3次元モデルを用いた顔表情の分析について述べる。

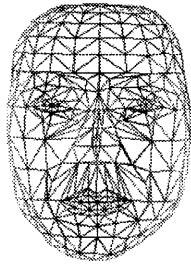
□ 2. 顔の3次元モデル

人間の顔は、基本的な形状や構造は共通であるが、目、鼻、口などの顔を構成する要素の形状や位置関係は、個人個人によって微妙に異なっている。自然な表情を合成し、アニメーションを作成するためには、対象人物の顔に対する忠実な3次元形状モデルが必要となる。

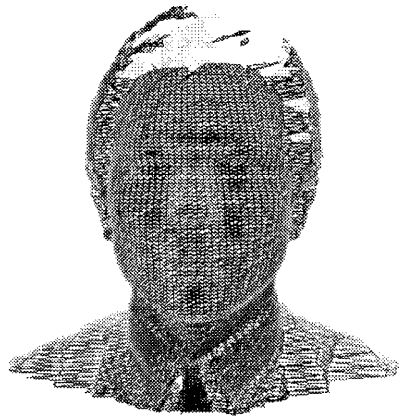
ここでは、従来までの手法で用いられている顔の3次元モデル(以下、従来型モデルとする)と、本手法で用いる高精細な3次元モデル(以下、高精細モデルとする)について述べる。

従来型モデルでは、まず顔の一般的な形状を表現するために、標準的な顔を、頂点の座標とそれらを結合する線分で近似した3次元モデル(標準3次元モデル)を用意している。つぎに、この標準3次元モデルを、対象人物の顔の輪郭や各部位(たとえば、眉、目、鼻、口など)をもとに、無表情の正面顔画像に対話的に、あるいは自動的に整合させ、個人の顔の3次元モデルとしている[5]。この個人の顔に整合した従来型モデルを図1(a)に示す。

この3次元モデルでは、モデルを構成する格子点のうちいくつかの格子点と、表情合成の際に必要な顔の各特徴点との対応付けがなされており、比較的容易にモデルを変形することができる。しかし、この従来型モデルでは、モデルの表面を定義している三角形パッチの数が数百~千数百個と少ないために、表情による動きが激しい部位(たとえば、目、口の周辺な



(a) 従来型 3次元モデル



(b) 高精細 3次元モデル

図1.人物頭部の3次元形状モデル

ど)では、微妙な変化が反映されにくい。また、モデルの奥行き座標は、個人の3次元モデルを整合する際に、標準モデルの奥行き座標を用いて推定されるので、対象人物の奥行き座標とは異なっている。さらに、顔画像を合成する際に必要となるテクスチャ情報は、対象人物の正面顔画像をもとにした2次元の情報となっている。このため、正面以外の任意の方向から見た顔画像を合成すると、テクスチャ情報が不足するために、顔画像は極めて不自然なものとなる。

これに対して、3次元形状とカラー画像を同時計測できるスキャナを用いて、対象人物の頭部の形状とテクスチャを取得する手法が提案されている[7]。この装置には、レーザ光源とCCDカメラが組み込まれており、これを対象人物の回りに一周させることによって、人物の完全な3次元形状と円筒方向からのテクスチャ情報を取得することができる。一周360度に対して512回走査し、各走査について512点の3次元座標とカラー情報が得られる。この手法によって得られた高精細モデルを図1(b)に示す。

これによって得られた高精細モデルと、全周囲からのテクスチャ情報を用いることにより、任意の方向から見た顔画像を合成することができる。しかし、この高精細モデルでは、モデルを構成する格子点の数が膨大であり、また、顔の各特徴点がモデルのどの格子点に対応しているのが特定できないために、表情や口形パラメータによるモデルの変形は困難である。

このように、三角形パッチによる形状の近似の度合と、3次元モデルの取り扱いの容易さとは相反する関係にある。三角形パッチを細かくするほど、もとの曲面形状を忠実に表現できることになるが、パッチの数が増大しモデルの取扱いは複雑となる。このため、顔の形状を忠実に近似しつつ、データ量を効率よく削減するアルゴリズムが提案されている[9]。

□ 3. 高精細 3次元モデルの階層的制御

人物の顔の3次元モデルを、表情や口形パラメータによって変形し、顔画像を合成するためには、対象人物の顔の特徴部位(たとえば、眉、目、鼻、口の端点など)の位置をモデル上で正確に認識している必要がある。

ここでは、高精細モデルに任意の表情や口形を与えるために、従来型モデルとの対応付けを行なうことによって、それぞれのモデルの特徴を活かしながら、階層的に高精細モデルを制御する手法について述べる。

□ 3-1. 高精細モデルと従来型モデルとのリンク

従来型モデルでは、構成する格子点の間隔は粗いものの、顔の各特徴点と格子点との対応付けがなされていた。これに対して、高精細モデルでは精度の高い格子点の座標が得られるものの、各特徴点がどの格子点に相当しているのかは特定できない。

そこで、高精細モデルの格子点のうち、従来型モデルの格子点と同一であると見なせる点を選出することによって、2つのモデル間の格子点の対応付けを行なう。図2に、その概念図を示す。

従来型モデルの格子点は、対象人物の正面顔画像をもとにした2次元平面内での顔の各特徴点の位置(x, y 座標)を整合したものであり、奥行き(z 座標)については標準モデルをもとに推定したものである。ここで、対称人物の正面顔画像は、高精細モデルを正面から見たときに、ある2次元平面に投影したものと考えることができる。

そこで、以下のようなアルゴリズムを用いて、高精細モデルの格子点の中から、従来型モデルの格子点と同一であると見なせる点を選出することができる。

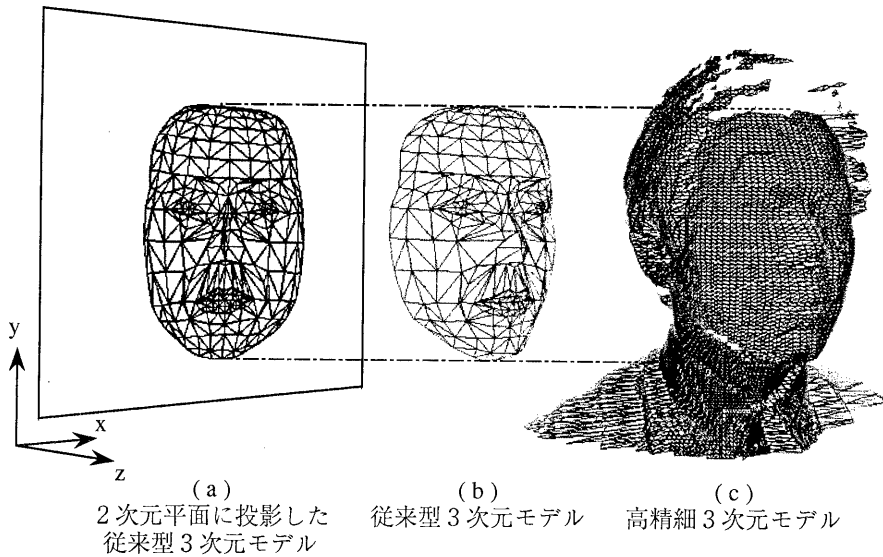


図2.高精細モデルと従来型モデルとのリンク

1).高精細3次元モデルに全周囲テクスチャをマッピングし、これを適当に回転させて2次元平面に投影し、正面顔画像を合成する。

2).合成された正面顔画像に、従来型3次元モデルを対話的に整合する。

3).従来型3次元モデルの各格子点のx,y座標に最も近く、かつz座標が最も大きい(つまり、顔の表面にある)高精細3次元モデルの格子点を選出する。この従来型モデルとの対応付けによって、高精細モデルの格子点のうちのいくつかの点については、顔の特徴点であるという意味づけを行なうことができる。

□3-2. 制御点による高精細モデルの格子点の移動

3-1では、高精細モデルの格子点うちのいくつかの点について、対象人物の顔の各特徴点との対応付けを行なった。これらの対応付けられた格子点(以下、制御用格子点とする)については、従来までの表情や口形パラメータをそのまま用いて移動を行なうことができる。しかし、高精細モデルを用いて顔画像を合成するためには、制御用格子点以外の点についても、それらの空間的な位置関係を保持しながら移動を行ない、モデルを変形する必要がある。

そこでまず、制御用格子点を用いて高精細モデルの各格子点を表わす。図3に制御用格子点と高精細モデルの格子点の関係を示す。高精細モデルの格子点は、3つの制御用格子点によって囲まれる三角形平面上に存在するとは限らない。そこで、高精細モデルの格子点は、この三角形平面を規定する2つのベクトルと、これらのベクトルの外積(三角平面の法線ベクトル)

を用いて(1)式のように表現できる。

$$\vec{r} = s\vec{a} + t\vec{b} + u(\vec{a} \times \vec{b}) \quad (1)$$

ここで、 \vec{r} は高精細モデルの格子点のベクトル、 \vec{a} 、 \vec{b} は制御用格子点によってできる三角形の2辺のベクトルで、 s, t, u はそれぞれの重み係数である。 s, t は格子点のベクトル \vec{r} を三角形平面上に投影したときの平面内での位置を表わし、 u はベクトル \vec{r} の三角形平面からの高さを表わすパラメータと考えることができる。

ただし、(1)式によって表わすことのできる格子点は、制御用格子点によってできる三角形平面内か、あるいはこの三角形平面の上下に位置するものを選出しなければならない。この操作をすべての高精細モデルの格子点の対して行なうのは、たいへん複雑な処理となる。

そこで、図4のように3次元空間内のすべての制御用格子点と高精細モデルの格子点を、円筒座標系によって表わされる円筒平面上に投影し、この平面上に移された制御用格子点によって囲まれた三角形平面内の格子点について処理を行なった。

これにより、表情や口形によって制御用格子点の位置が変化しても、この3つのパラメータ(以下、STUパラメータとする)を保存しておき、表情によって変化した2つのベクトル \vec{a}' 、 \vec{b}' からSTUパラメータを用いて、移動後のベクトル \vec{r}' を求めることができる。

この手法の特徴は、制御用格子点とそれらの変位のみによって、高精細モデルの多数の格子点間の空間的な位置関係を保持したまま、モデルを変形することができる点にある。

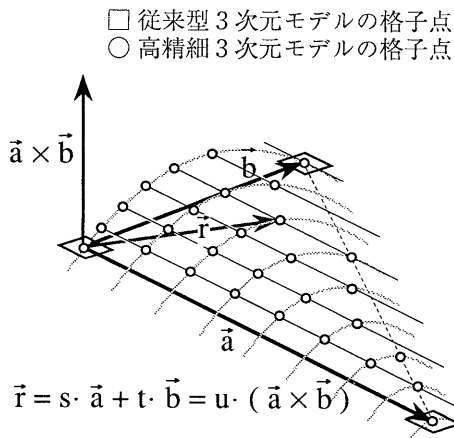


図3.制御用格子点と高精細モデルの格子点の関係

□ 4. 高精細 3次元モデルの表情合成と分析

人間の顔の表情は、骨格の移動、顔面筋や皮膚の伸縮によってさまざまに変化する。これらの動きによる変化をパラメータ化し、そのパラメータに対する表情の変形規則が定義されれば、いくつかのパラメータの組み合わせによって表情合成が可能となる。また逆に、顔画像を分析することによって表情パラメータを抽出することができる。

ここでは、表情パラメータを用いて高精細モデルの表情合成を行なう手法について述べる。また、人物の実際の表情による変化を高精細モデルを用いて分析を行なった結果についても述べる。

□ 4-1. 高精細モデルによる表情合成

高精細モデルを表情によって変形させるパラメータとして、心理学の分野で提案されているFACS(Facial Action Cording System)を用いた[10]。FACSは、顔面筋の位置および動きの方向を解剖学的に考慮して、顔の表情をAU(Action Unit)と呼ばれる44個の動作に分解している。あらゆる表情は、AUの組み合わせで表現できるとされ、FACSは表情記述単位として顔画像の分析・合成の分野で広く用いられている。



図5.高精細モデルによる表情合成例

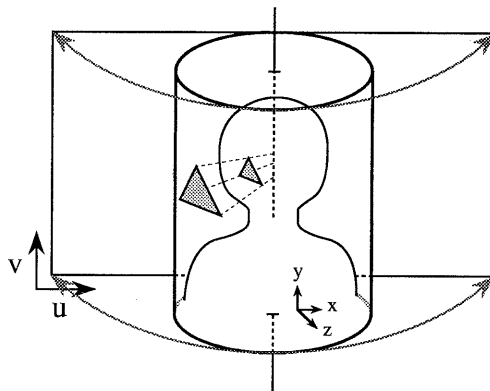


図4.円筒表面上における格子点の対応付け

このFACSによる表情パラメータをもとに制御用格子点を移動し、高精細モデルの格子点をSTUパラメータを用いて移動させモデルの変形を行なった。図5に、この手法を用いた表情合成例を示す。

□ 4-2. 高精細モデルを用いた表情分析

表情分析とは、表情合成とは逆に変形された顔の各部位から表情パラメータを抽出することである[11]。表情の分析には、対称人物の実画像を用いることも可能であるが、顔の各特徴点は複雑な画像処理によって抽出しなければならない。

そこで、人物の顔表面に目印となるパッチを付け、その高精細モデルとテクスチャを3次元スキャナを用いて測定した。顔表面でのパッチの位置は、FACSの表情パラメータを表わす顔の特徴点と一致していなければならないが、ここでは、それらを考慮して96個の部位を測定点とした。図6(a)に、表情分析に用いた96個のパッチの位置を示す。

分析を行なう表情には、無表情とFACSにおいて基本表情とされる、怒り(Anger)、嫌悪(Disgust)、恐怖(Fear)、喜び(Happiness)、悲しみ(Sadness)、驚き(Surprise)の6基本表情を含む合計7表情とした。この



図 6 (a). 表情分析に用いた特徴点

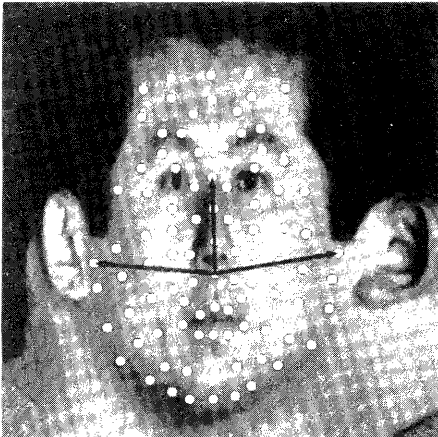


図 6 (b). 正規化のための基準特徴点

6 基本表情について、顔表面にあるそれぞれの測定点に対して、無表情との変位を比較することによって、その移動量と移動方向を算出した。

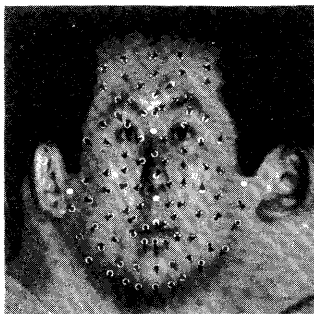
ここで、分析に用いる高精度モデルは、各表情のそれぞれの測定条件によって、微妙にその座標がずれている。そこで、顔の特徴点の中で表情による変位がないとされる特徴点を4つ選出し、それらの点によって定められる3つの3次元ベクトルを用いて各測定点を表わすことによって、各モデル間の変位の正規化を行なった。図 6 (b)に、正規化を行なうために基準となる特徴点を示す。また、図7に分析によって算出された各測定点の移動量と移動方向を示す。

なおこの図において、実際の測定点は3次元空間内で変位するが、ここでは視覚化のためにこれらの変位を、円筒座標系によって表わされる円筒平面上に投影したものを示している。図中の⊗は円筒軸方向への沈み込み、●は膨らみを表わしている。

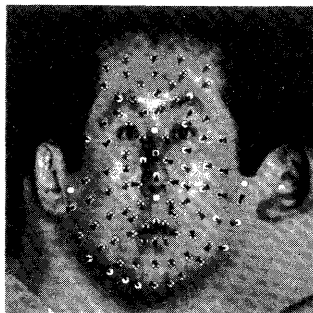
これらの分析によって、人間の顔表面の動きは大きく分けて以下の4つの領域に分けることができると考えられる。

- (a). 円筒平面内で上下左右に変化する部分。つまり、顔の皮膚表面での2次元的な運動をする領域（額、目など）。
- (b). 円筒軸方向で上下に変化する部分。つまり、皮膚表面での膨らみや沈み込みをする領域（頬の上部）。
- (c). (a), (b)を合わせたように変化する部分。皮膚の3次元的な柔軟な運動をする領域（口の周辺など）。
- (d). 変化が顕著に現われない部分。骨格などの影響によってほとんど運動しない領域（鼻、額の一部）。

これらの変位の特徴から、人間の顔表面をおおまかに12個の領域に分割したものを図8に示す。



(a). 怒り (Anger)



(b). 喜び (Happiness)



(c). 悲しみ (Sadness)

図 7. 実表情による特徴点の変位



- (a). 皮膚表面での2次元的な運動をする領域
- (b). 皮膚表面に対して膨らみや沈み込みをする領域
- (c). 3次元的な柔軟な運動をする領域
- (d). ほとんど運動しない領域

図8.分析にもとづく顔表面の領域分割

この分析結果をもとに、高精細モデルでは、それぞれの皮膚領域に対して、3-2で述べたSTUパラメータの値を適当に操作することによって、従来型モデルではできなかった微妙な制御を行なうことができると考えられる。たとえば、制御用格子点によってできる三角形平面からの高さに相当するuパラメータを操作することによって、顔表面の皮膚の伸縮による盛り上がりや沈み込みを表現できる。

現在、この表情による変位の分析をさらに進め、表情パラメータにフィードバックすることによって、より自然な顔画像を合成する手法について検討を進めている。

□ 5. まとめ

本論文では、人物頭部の高精細3次元モデルを、従来まで広く用いられている正面顔画像用の3次元モデルとのリンクを取ることによって、階層的に高精細モデルを制御する手法について述べた。この手法の利点は、従来まで研究されてきた表情・口形合成の成果をそのまま活かして、高精細モデルを変形し顔画像を合成することができる点にある。

また、高精細モデルの格子点の精度の高さを活かして、人物の実際の表情変化による格子点の移動の分析を行なった。今後は、この表情分析をさらに進め、分析のよって得られた表情パラメータを、表情合成の際にフィードバックすることによって、より自然な顔画像を合成する手法についての検討を進めていきたい。

謝辞

高精細モデルのデータ作成にあたりご協力いただいた、ATR通信システム研究所の岸野文郎氏、北村泰一氏、大矢淳氏に深謝いたします。

参考文献

- [1] N.M.Thalman: "Visualizing Humans by Computer", The 3rd International Forum on the Frontier of Telecommunications Technology, 1991
- [2] S.Morishima et al: "A Proposal of A Knowledge Based Isolated Word Recognition", ICASSP'86, Vol.1.1, pp713-716, 1986
- [3] D.Terzopoulos and K.Waters: "Analysis of Facial Images Using Physical and Anatomical Models", Third International Conference on Computer Vision, pp.727-732
- [4] 石橋、岸野: "領域分割を用いた人物像の色彩テクスチャ分析・合成", 信学論(D-2), J74-D-2 No.11, 1991
- [5] 崔、原島、武部: "顔の3次元モデルに基づく表情の記述と合成", 信学会論文誌A, Vol. J73-A, No.7, pp.1270-1280, 1990
- [6] 森島、坂口、原島: "顔画像によるマルチメディアインタフェースとその開発支援環境の構築", 画像符号化シンポジウム(PCSJ 92), pp.231-234, 1992
- [7] 渡辺、末永: "3D形状と輝度(色)を同時計測可能なスキャナとその応用", 第6回 NICOGRAPH90, pp181-189, 1990

- [8] 小野、上野、森島、原島：“顔画像用高精度3次元モデルの階層的制御の一検討”，1992信学会秋季大会A-153
- [9] 森井、佐藤、鉄谷、岸野：“視線を利用した高速画像生成表示のための階層化モデル作成の一検討”，1992信学会春季大会A-277
- [10] P.Ekman and W.V.Friesen: "Facial Action Coding System", Consulting Psychologists Press, 1977
- [11] 岡崎、崔、山田、原島：“顔画像の分析合成システムの実現”，画像符号化シンポジウム(PCSJ 90), pp.201-204, 1990

成蹊大学工学部電気電子工学科情報通信研究室

〒180 東京都武蔵野市吉祥寺北町3-3-1

Tel. 0422 - 37 - 3742, Fax. 0422 - 37 - 3871