

実画像をベースとした実時間人物顔合成法

佐藤 一†† 寺島 信義† 富永 英義†

†早稲田大学大学院 国際情報通信研究科
〒169-0051 東京都新宿区西早稲田 1-3-10 早大 29-7 号館, Tel: (03)5286-9839
E-mail: hajime@giti.waseda.ac.jp

†早稲田大学 国際情報通信研究センター
〒169-0051 東京都新宿区西早稲田 1-21-1 早大西早稲田ビル 5F, Tel: (03)5286-3831
E-mail: {terasima,tominaga}@giti.or.jp

あらまし 人間対人間のコミュニケーションにおいて人物の顔が果たす役割は大きく、相手の顔を見ながら遠隔地同士で通信を行うことができるようなシステムの構築が求められている。もし一般的なビデオカメラをコンピュータを使用し、現実的な人物顔画像を実時間で合成表示することができれば、臨場感通信会議等においてより自然なコミュニケーションが期待できる。本稿では、そのような機能を実現するため、実画像をベースとした人物顔画像合成法を提案する。まず、カメラ画像中の人物顔領域の特定・抽出を行った後、人物顔位置や角度といったパラメータを高速かつ高精度で認識する。次に、計測されたパラメータを人物顔を模した3次元三角形パッチモデルに反映し、入力画像を直接モデルにテクスチャマッピングすることにより、現実的な人物顔画像を再構築する。また、入力画像を連続的に貼り換えていくことにより、人物顔表情変化を再現する。各要素技術について検討を行った後、1名の人物顔を用いた基礎実験を行い、提案手法の有効性を明らかにする。

キーワード 画像生成、顔画像、実画像ベース、三次元画像、臨場感通信会議

Image-based Real Time Creation of Human Face Images

Hajime SATO†† Nobuyoshi TERASHIMA† Hideyoshi TOMINAGA†

†Graduate School of Global Information and Telecommunication, Waseda University
1-3-10 Nishi-Waseda, Shinjuku-ku, Tokyo, 169-0051 Japan, Tel: +81-3-5286-9839
E-mail: hajime@giti.waseda.ac.jp

†Global Information and Telecommunication Institute, Waseda University
1-21-1 Nishi-Waseda, 5th Floor, Nishi Waseda Bldg., Shinjuku-ku, Tokyo, 169-0051 Japan, Tel:
+81-3-5286-3831

E-mail: {terasima,tominaga}@giti.or.jp

Abstract The human face plays an important role in man-man communication, and systems which enable distant users to communicate while viewing each others' face are demanded. If realistic human face images could be created and displayed at real time using an ordinary video camera and computer, smoother communication in such systems could be expected. In this paper, an image-based human face creation method is proposed. After measuring the human face position and direction parameters with high speed and precision, the parameters are used to move and rotate a three dimensional triangular patch model. By texture mapping the camera input image directly to the model, realistic face images could be reconstructed. Also, changes in facial expressions could be reconstructed by mapping the input images consecutively. These individual methods were examined, and a fundamental experiment using a single person's face showed good results.

key words Image Generation, Face Image, Image-based, Three Dimensional Image, Virtual Space Teleconferencing

1. はじめに

人間対人間型コミュニケーションにおいて人物の顔が果たす役割は大きく、相手の顔を見ながら円滑に対話を行うためのシステムの構築が求められてきた。従来はテレビ電話やテレビ会議システムなどが主流であったが、近年では実社会における人間対人間のコミュニケーションに限りなく近い雰囲気を感じられるような「場」をコンピュータ内に生成するという、新しい考え方に基づいたシステムの構築が進められている。コンピュータグラフィックスによって作成された3次元仮想空間の中に参加者の仮想的な分身(Avatar)を合成表示することにより、遠隔地にいる者どうしがあたかも一堂に会しているような効果を得る臨場感通信会議システムがこの代表例として挙げられる。(1)

もし、一般的なビデオカメラとコンピュータを用いて現実的な人物顔画像を実時間で再構築・表示することが出来れば、このようなシステムにおいて、より自然なコミュニケーションが期待できる。本稿ではこのような機能を実現するため、実画像をベースとした人物顔画像の簡便かつ高速・高精度な生成法を提案する。

2. 提案手法

2.1 提案手法の概要

従来より、モデルベース符号化(2)に代表されるように、参加者の顔特徴変化を2次元画像より解析し、予め計算機内に構築しておいた3次元無表情顔モデル上にその結果を反映させるすることにより、任意顔表情の顔画像を生成するアプローチが広く行われてきた。しかし、これらのアプローチでは表情変化による人物顔画像の3次元動的な動きを正確にモデル化する記述するのが困難であるため、自然な顔画像の生成が難しく、実用的なシステムの構築に至っていない。また、光学的マーカーや磁気センサを用いることにより、人物顔表情や動作をキャッチし、3次元顔モデルに反映するようなシステムも構築されているが(3)、こういったデバイスは正しい位置に装着するのが面倒である上、装着している間にも違和感がある。

想定しているシステムにおいて、最終的には人物らしくみえる顔画像を作り出すことができれば充分であり、人物顔の厳密なモデル化は必ずしも必要でないと考えられる。カメラ画像の持つ特徴をそのまま使うことにより、対象人物の特徴を保持した別の見え方のオブジェクトを作成するという考え方(4)に立つべきである。

提案手法では前述した条件を満たし、人物顔情報の実時間認識およびリアルな顔画像の合成を次のような手順で行う。

まず、入力フレーム間の差分と輝度プロジェクションを用いてカメラ画像中の人物顔領域を特定し、抽出・正規化を行う。次に、人物顔画像による3次元離散テンプレート系列と抽出された顔画像の

テンプレート補間照合を行い、左右、上下、傾斜方向の人物顔角度を高速かつ高精度で計測する。最後に、人物の正面・側面顔画像上に特徴点を指定して3次元三角形パッチモデルを作成し、計測顔位置・角度を反映した後、入力画像を直接テクスチャマッピングすることにより、現実的な人物顔画像を再構築する。また、入力画像を次々と貼り換えていくことにより、人物の表情変化を再現する。以上の処理の流れを図1に示す。以下、3～5章では各要素技術の詳細について述べていく。

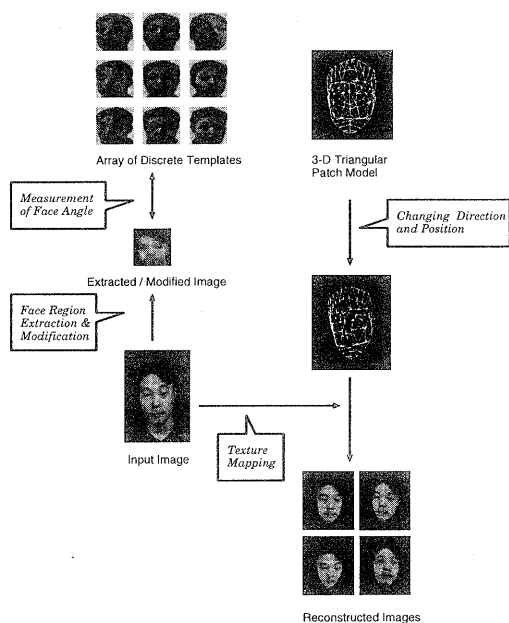


図1: 提案手法の流れ

3. 人物顔領域の特定および抽出・正規化

3.1 概要

提案システムはオフィスや教室のような環境で使用することを想定しているため、処理が比較的簡便であり、背景、照明、色情報、人物顔サイズ・位置などになるべく制約をつける必要のない手法が望ましい。また、対話をしている人物の顔位置や方向は連続的に変化しており、顔動作に較べて首から下の動作は比較的小さいと考えられるため、このような特性を利用する。連続した3枚のフレーム間差分を用いて画像中の人物顔領域を特定した後(5)、水平・垂直方向の輝度プロジェクションを求めることにより、顔領域全体を抽出する。この手法により、10[フレーム/秒]程度のフレームレートでの高精度な処理が可能となる。(6)

3.2 人物顔領域の特定・抽出

まず、現フレーム f_i と前フレーム f_{i-1} の差分画像 d_{i-} 、および現フレーム f_i と次フレーム f_{i+1} の差分画像 d_{i+} をそれぞれ求める。なお、連続した2枚のフレーム間の差分が得られなかった場合は、その次の入力フレームを用いて差分を求める。次に、差分画像の濃度ヒストグラムの谷底値より適切なしきい値を決定し、 d_{i-} 、 d_{i+} を2値化する。 d_{i-} 、 d_{i+} の論理積を取るにより得られる画像 g_i 中の白画素が密集している箇所より、現フレーム f_i 中の動領域の輪郭を推定することができる。

画像 g_i に 3×3 メディアンフィルタ処理を施し、雑音を除去した後、垂直方向の輝度プロジェクション $P_y(x)$ を作成する。立ち上がり点および立ち下がり点の位置を判定しやすくするため、次式であらわされる整形処理を施す。

$$P_y(x) = \frac{1}{5} \sum_{n=x-2}^{x+2} P_y(n) + 0.5 \quad (1)$$

整形後の輝度プロジェクションのピーク値より左右の列をスキャンしていく事により、頭部領域の端点を探索し、横幅を切り出す。

このようにして切り出された領域の水平方向の輝度プロジェクション $P_x(y)$ を同様にして作成し、整形処理を施す。画像の1行目から下に行をスキャンしていく事により頭部領域の上端を、輝度プロジェクションのピーク値より下に行をスキャンしていく事により下端を特定し、縦幅を切り出す。以上の操作により、頭部全体を外接矩形領域で抽出し、抽出画像 o_i を得る。

また、このときの領域のサイズおよび頂点座標の位置は後の操作で用いるため、同時に保存しておく。

3.3 抽出顔画像の正規化

人物顔領域の大きさは顔方向によって異なり、入力画像とテンプレート画像の大きさが異なると、後述するテンプレート補間照合を行うことができなくなってしまうため、抽出画像の縦横比を変化させ、 $n \times n$ (画素) の方形画像に変換する。

また、 3×3 の線形平滑化フィルタ処理を数回施すことにより画像をぼかし、表情変化や雑音による影響を低減し、抽出正規化顔画像 o_i^* を得る。

以上の処理の流れを図2に示す。

4. 人物顔角度の計測

4.1 概要

3.3章の処理で得られた抽出正規顔画像 o_i^* の特徴量に基づき、テンプレート補間照合を用いて人物顔角度を高速かつ高精度で計測する。

まず、異なるカメラ角度から取得された無表情人物顔画像を上下、左右、傾斜方向に標準パターンとして並べることにより3次元離散テンプレート系列を作成し、 o_i^* と全数照合を行う。次に、隣接テン

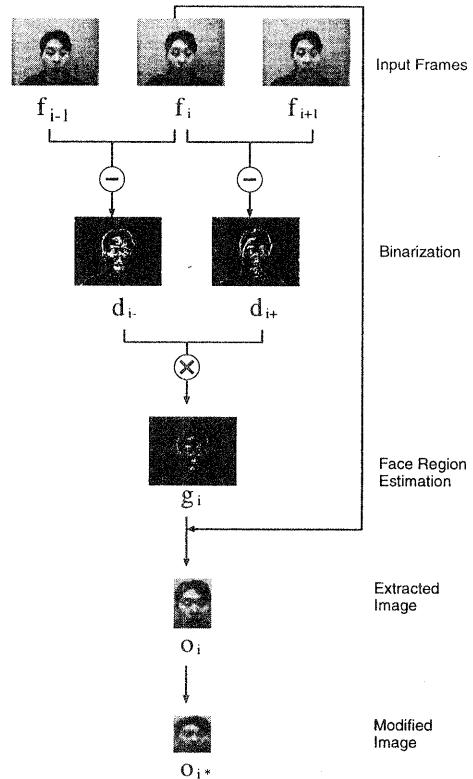


図2: 人物顔領域の抽出

プレート間の補間計算を行うことにより、上下、左右、傾斜方向の角度 $x^\circ, y^\circ, z^\circ$ で計測する。この概念を図3に示す。この手法を用いることにより、無表情顔画像を入力画像として用いたときには $\pm 2^\circ \sim 3^\circ$ 程度、有表情顔画像を入力画像として用いたときでも $\pm 3^\circ \sim 4^\circ$ 程度の誤差で顔角度を計測可能であることが確認されている⁽⁶⁾。

4.2 最類似テンプレートの選択

人物顔画像の照合処理を行う場合、画像中の特徴を検出するよりも、画像全体の特徴を用いた方が簡便かつ高精度な認識が可能になるため、3章の処理で得られた顔画像をそのまま多次元ベクトルとして用いる。なお、以降の処理では濃淡情報のみを用いるものとする。

カメラ角度を上下、左右および傾斜方向に等間隔で変化させ、参加する人物の顔画像を取得した後、3.3章に述べた手法で正規化処理を施し、画像を並べることにより照合用の離散テンプレート系列を作成する。

抽出顔画像の画素の濃度値 $n \times n$ 個を要素とするベクトルを \vec{j} 、テンプレート画像の画素の濃度値 $n \times$

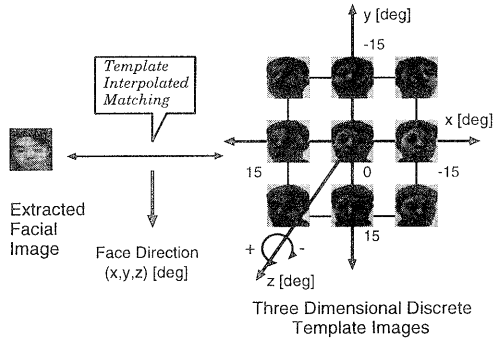


図 3: 人物顔角度計測

n 個を要素とするベクトルを \vec{f} とするとき、 \vec{f} と \vec{g} の成す角度が 0° に近付くほど 2 枚の画像は類似している。ここで、 \vec{g} と \vec{f} が成す角度 Θ の余弦をとることにより、2 つの画像の単純類似度 m は、

$$m = \cos \Theta = \frac{\vec{f} \cdot \vec{g}}{\|\vec{f}\| \cdot \|\vec{g}\|} \quad (2)$$

で求められる (但し、 $0 \leq m \leq 1$)。 (7)

4.3 隣接テンプレート間の補間

入力画像の変形がある関数に従って生じる場合、1 次元離散テンプレート系列を設けておくことにより、入力画像と完全一致のテンプレートが無い場合でも、完全一致のテンプレートが存在する場合とほぼ同等の精度で変形パラメータを推定する手法が提案されているが (7)、この考え方を 3 次元拡張して用いる。

左右、上下、傾斜方向の人物顔角度変化はほぼ線形であると考えられるため (6)、隣接テンプレート間の変形が線形に生じると仮定すれば、線形補間を用いることにより簡便に顔角度を算出することができる。

まず、入力画像を 3 次元離散テンプレート系列と全数照合した後、単純類似度の合計が最も高い 8 枚の隣接テンプレートを選択する。この 8 枚の隣接テンプレートの中心を結んだ立方体内に、入力画像と完全一致の仮想テンプレートが存在すると仮定する。このとき、仮想テンプレートの中心は、8 枚のテンプレートの中心を結んだ立方体の縦、横、奥行きをそれぞれ $p:(1-p)$, $q:(1-q)$, $r:(1-r)$ に配分する点であると考えられる事ができる。隣接した 2 枚のテンプレート間の変化が線形であると仮定すると、 p , q , r はそれぞれ

$$p = \frac{1 - m_1}{2 - m_1 - m_2}$$

$$q = \frac{1 - m_1}{2 - m_1 - m_3} \quad (3)$$

$$r = \frac{1 - m_1}{2 - m_1 - m_5}$$

で求められる。なお、最類似テンプレート f_1 の類似度を m_1 、垂直方向に隣接したテンプレート f_2 の類似度を m_2 、水平方向に隣接したテンプレート f_3 の類似度を m_3 、奥行き方向に隣接したテンプレート f_5 の類似度を m_5 とそれぞれ定義する (図 4 参照)。

この時の上下、左右、傾斜方向の角度 x° , y° , z° はそれぞれ次の式で計算される。

$$x = (1 - p)a_1 + p(x + a_2)$$

$$y = (1 - q)b_1 + q(y + b_3) \quad (4)$$

$$z = (1 - r)c_1 + r(z + c_5)$$

ここで、 a_n , b_n , c_n はテンプレート f_n の上下、左右、傾斜方向の角度をそれぞれ表している。

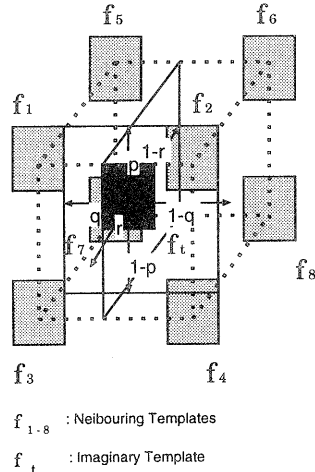


図 4: 隣接テンプレートを用いた補間

5. 3 次元人物顔画像の再構築法

5.1 概要

人物の顔をかたどった単純な 3 次元三角形パッチモデルを作成し、メモリに蓄えておく。このモデルで 3, 4 章で得られた人物顔位置および顔角度を反映し、入力画像を直接マッピングすることにより、人物らしい顔画像の再構築および人物顔表情変化を再現する。 (8)

5.2 顔形状モデルの生成

顔形状モデルとしては、特殊な機材や技術を要しなくても比較的簡易に作成でき、視覚的に充分満足の顔画像を得られるようなものが望ましい。

まず、人物の正面顔画像および側面顔画像を取得する。専用のビューワを用いて正面顔画像上に目、鼻、口といった顔部品の端点などに特徴点を手作業で配置した後、側面顔画像上にその対応点を配置していく。次に、隣接する特徴点同士を線で繋ぐことにより、各特徴点を頂点とした3次元の三角形パッチモデルを得る。

5.3 入力画像のテクスチャマッピング

前述の作業によって得られた3次元三角形パッチモデルにカメラ入力画像を直接マッピングすることにより、入力画像と同じ顔情報を保持した3次元顔画像を生成することが可能となる。3章、4章で得られた人物顔位置・角度をもとに入力画像を各三角形パッチにテクスチャマッピングしていく。

3次元三角形パッチモデルの頂点の順番は表から見て右まわりに記述しておき、その記述順から各三角形パッチの表裏判定を行う。パッチの頂点を順番に A, B, C とすると、 $\vec{AB} \times \vec{AC} > 0$ を満たすパッチを表向きと判断する。表向きのパッチには2次元平面上のテクスチャが切り出され、パッチに対して $\theta = 90^\circ$ の角度になるようにマッピングされる。

テクスチャマッピングにはアフィン変換を使用し、テクスチャ上の点 (x, y) は、パッチ上の点 (u, v) に投影する。アフィン変換は実際に生成2次元画像を表示する際に一度実行すればよいので、実際にはパッチの頂点 $P_1 P_2 P_3$ とテクスチャ上の対応点 $I_1 I_2 I_3$ の指定のみであらわされる(図5参照)。

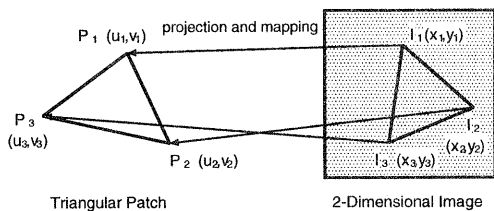


図 5: テクスチャマッピングの概念

人物顔表情が変化しても、顔の構造や形状自体はそれほど大きく変化しないと考えられるため、パッチモデルの変形は行わず、入力顔画像を連続的に貼り換えていくことにより、人物顔表情変化を再現する。以上の作業の流れを図6に示す。

6. 評価実験

6.1 概要

6.2章に示す実験データを用い、3~5章に述べた個々の要素技術を融合した基礎実験をSGI Indy

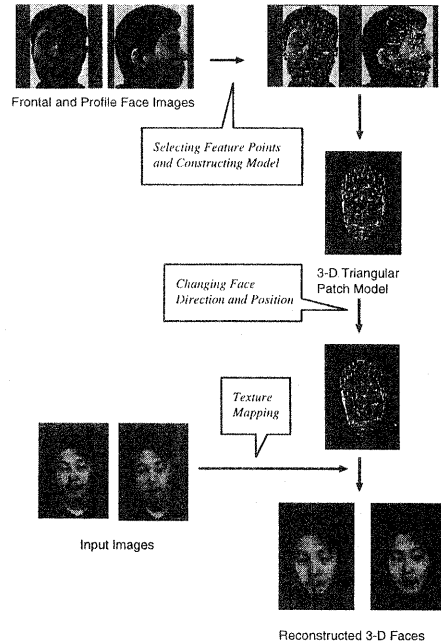


図 6: 3次元人物顔画像の再構築

上で行った。

6.2 実験データ

6.2.1 入力動画画像

1台の家庭用ビデオカメラを使用し、室内で1名の人物が座りながら喋っているところを撮影した。実際のシステムを想定し、ディスプレイ上にカメラを設置し、被験者はディスプレイに向かって自由な姿勢で話しかけるようにしている。照明条件は一般的な室内光、背景はほぼ一様とし、被験者のアップショットを撮影した。取得映像は10[フレーム/秒]でフレーム抽出し、320×240画素の8bit白黒濃淡画像に書き出してから処理を行った。

6.2.2 3次元離散テンプレート系列

ビデオ撮影した被験者の無表情顔情報を3次元レンジファインダCyberware 3D Digitizer 3030RGBを用いて取得し、人物顔画像のテンプレート系列を得た。ビデオ映像に映った人物頭部領域の動作範囲を考慮し、上下方向±30°、左右方向±30°、傾斜方向±30°の範囲で15°おきに顔方向を変化させた顔画像を取得し、64×64(画素)の8bit白黒濃淡画像による、5×5×5の3次元離散テンプレート系列を作成して用いた。なお、今回は入力顔画像とテンプレート画像との間の画像補正は行っていない。

6.2.3 人物顔形状モデル

被験者の正面顔画像および側面顔画像上に83個の特徴点を取り、129の三角形パッチからなる形状モデルを生成した。顎や目領域など曲線の多い部分は特徴点の数を多く、額や頬など比較的平坦な部分は特徴点の数を少なくした。現時点では、計測パラメータのデータ入出力形式が統一されていないため、人物顔位置やサイズ、顔角度は3章、4章で得られた数値を手作業で入力した。

6.3 結果と考察

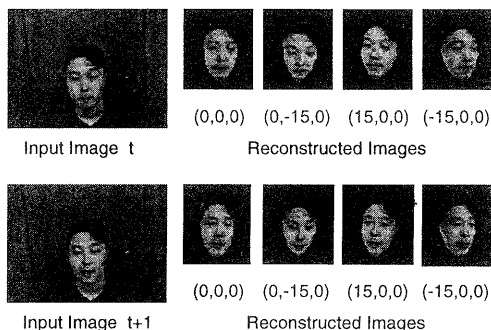


図 7: 実験結果 (1)

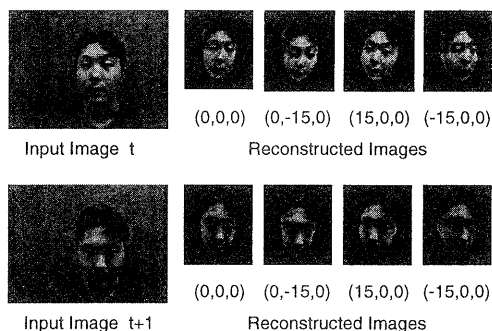


図 8: 実験結果 (2)

実験結果の例を図7、図8に示す。これより、人物の顔特徴を保持した人物顔画像の再構築が行われていることが確認できた。また、図7、図8におけるInput Image t, t+1の再構築画像を比較することにより、顔形状モデルを変形せず、入力画像を貼り換えるだけで顔表情変化が再現されていることも確認できた。

図7のように、人物顔動作が比較的小さいときには視覚的に満足な人物顔画像が再構築されているが、図8のように、人物顔動作が比較的大きいときには顔画像に歪みが生じており、視覚的に違和感が

感じられる。これは人物顔角度の角度計測誤差により、テクスチャをマッピングする領域がずれてしまったことが原因である。今回は入力画像とプレート画像との間の補正は行っていないが、入力画像のクオリティや取得条件、顔領域以外の部分、画像の正規化を行う際のばかし具合などを検討していく必要がある。

また、ビューイング角度の変化が比較的小さいときは自然な顔画像が表示されたが、大きくなるほど顔画像に歪みが生じており、入力顔画像と顔表情が異なって見えてしまっている。現在は作業簡略化のため1台のカメラのみを用いているため、テクスチャ面積に限界があるが、カメラ台数を増やしたり、設置位置を工夫することにより改善できると考えられる。

7. おわりに

本稿では、カメラ画像中の人物の顔角度や位置を高速・高精度で計測した後、計測パラメータを反映した3次元の三角形パッチモデルに入力画像を直接貼り付けることにより、現実的な人物顔画像を再構築および顔表情変化を再現する手法を提案した。また、1名の被験者による基礎実験により、提案手法の有効性を確認した。

今後は、人物顔角度計測法の改良やカメラ台数・位置に関する検討を行い、最終的な目的である画像入力から顔画像生成までを一貫して全自動かつ実時間で行えるシステムの実装を行っていく予定である。

参考文献

- [1] 岸野 文朗: “ヒューマンコミュニケーション - 臨場感通信 -”, テレビ学会誌, Vol.46, No.6, pp.698-702, 1992
- [2] 金子, 羽鳥, 小池: “形状変化の検出と3次元形状モデルに基づく顔動画の符号化”, 信学論 B, J-71B, 12, pp.1554-1563 (1988)
- [3] J. Ohya, Y. Kitamura, F. Kishino, N. Terashima, H. Takemura and H. Ishii: “Virtual Space Teleconferencing: Real-Time Reproduction of 3D Human Images”, Journal of Visual Communications and Image Representation, Vol.6, No.1, pp.1-25, 1995
- [4] 向川, 中村, 大田: “2枚の顔写真を用いた任意方向の顔画像の生成” 信学論 (D-II), Vol.J80-D-II, No.6, pp.1555-1562, 1997
- [5] 安居院, 長尾: “画像の処理と認識”, 昭晃堂, pp.162-3, 1992
- [6] 佐藤 一: “人物顔画像の実時間認識および再構築に関する研究”, 早大大学院理工学研究科修士論文, 2000
- [7] 小池, 田辺: “テンプレート補間および重み付き学習型ニューラルネットワークを用いた濃淡画像の照合法”, 信学論, Vol.J75-D-II, No.7, pp.1151-1159, 1992
- [8] 池崎, 佐藤, 津田, 寺島, 富永: “2台のカメラを用いた3次元顔画像の実時間生成の検討”, 信学春季全国大会論文集, D-12-120, 1998