コンピュータビジョンと
イメージメディア 109－10
（1998．1．23）

ヒューマンインターフェースのための表情転送に関する研究

塚田岳人, 大場光太郎, 神徳徹雄, 谷江和雄
通産省 工業技術院 機械技術研究所

本論文では, 不特定者間のスムーズな遠隔コミュニケーションを実現させるために表情が重要であることに注目し, 画像転送システムではない, 表情転送システムを実現した. まず, 表情を空間的に表す手法として固有空間手法を用いた表情空間を提案し, その妥当性を示した. さらに, 個別の表情空間同士を対応付ける手法を提案し, 顔画像同士の対応付け表情空間内で行なう手法を示した. 最後に, 送信側と受信側とで異なる人物の顔画像を用いて, 表情転送が可能であるシステムの構成を提案し, アルゴリズムの検証をおこなった.

Facial Expression Transportation for Smooth Human Interface

Takehiro TSUKADA, Kohtaro OHBA, Tetsuo KOTOKU and Kazuo TANIE
Mechanical Engineering Lab., AIST, MITI

In this paper, the facial expression is mainly focused to achieve the smooth human tele-communications. The facial expressions has been considered as a most significant factor in tele-communications, such as tele-services. To realize the real time transportation system of facial expression, we proposed the facial expression space (**FES**) and a correspondence technique between each personal facial expression spaces. Then, real time facial expression transportation system is developed, which transport the facial expression but not the image itself. This final system is able to display the same facial expressions in another persons, further more in cartoon characters. The experimental results show the validity of these criteria.

1. 序論

近年, 電話回線を利用したテレ・コミニュケーションは, 様々な日常生活で広く利用され, 急速な普及を示している (図1). テレ・コミニュケーションの初期のものは, 音声を用いる電話であり, 最近では音声に加えて映像も付加したテレビ電話に発展しようとしている. さらに最近では, 力覚, 触覚の情報を転送するため離れた所でロボットを操作する遠隔操作技術の研究が進められている.

一方, 汎用ネットワークを利用し, 情報を特定者もしくは不特定者へ提供するサービスとしては, E-mail の急速な普及が上げられる. また最近の WWW は同時に多くの人に文章, イメージ, 音のデータを与え, 有用な情報を提供している. しかし, これらは基本的に情報を提示する側と情報を受けとる側の二つが単純に接続されるものである. 近年では, 数人で仮想世界を共有することを VRML の技術で実現しつつあ

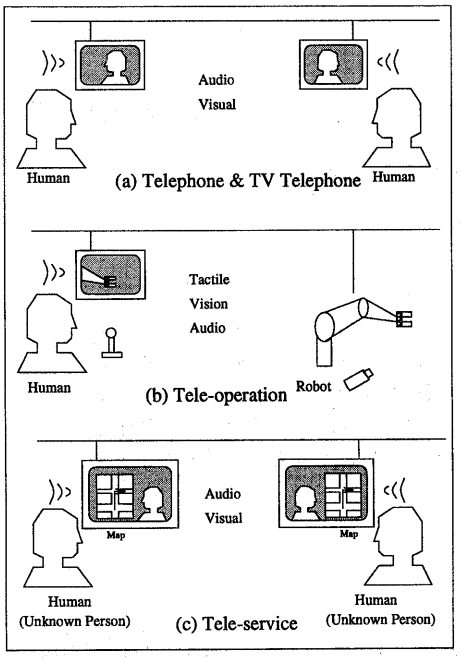

図 1: Tele-communications.

る [1]. この手法は, 汎用ネットワークを利用したテレビ会議, チャットシステム, バーチャル・ショッピィング, バーチャル・ミュージアム, その他 アミューズメント利用を目的として, リアルタイムでのテレ・コミニュケーションを可能としつつある. しかしながら, ここで問題とされていることは, これらのシステムの多くは仮想世界内でのコミニュケーションが, 我々の実世界でのコミニュケーションで得られるものと同じ感覚を得られないということである. つまり, 実世界の臨場感を効率的に転送していないことに問題があると考えられる.

ここで, 実世界での人間のコミニュケーションについて考えてみよう. 通常, 人間はコミニュケーションの方法について深く考慮しなくても, 普通の生活においてスムーズにお互いにコミニュケーションが可能である. このコミニュケーションは 図2 に示すように「言語」「顔の表情」「ジェスチャー」のように言語的または非言語的な多くの方法で行なわれている. これらの手法の基本要素は, 人間の五官

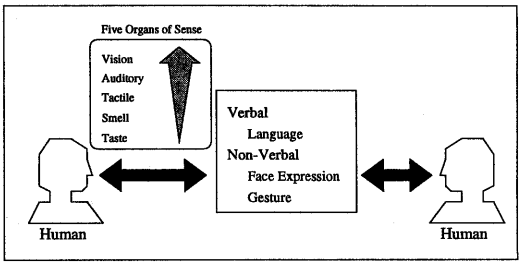

図 2: Communication.

「味覚」「嗅覚」「触角」「聴覚」「視覚」をベースとしたものであり, それらを自由に組み合わせることにより, スムーズなコミニュケーションが実現されている. この感覚の中でも, 人間が日常情報として得ている 70 ~ 80% は「視覚」であり, 「味覚」と「嗅覚」はほとんどの状況で無視できる. また, 「触角」は人間と人間, または, 人間と物との物理的に接触している場合には欠かせないものであるが, この論文では接触状態を含まないコミニュケーションに重点を置くこととする.

それでは視覚から得ている情報として, 人間同士のスムーズなコミニュケーションに必要な要素を考えてみる. 映像は多くの情報を含んでいるが, 人間同士の会話の場合, 顔の動きや体の動きが重要な要素であることは容易に理解できる. コミニュケーションのための重要な象徴はイメージ自身でなく「顔の表情」や「ジェスチャー」にあると言える [2]. 一方最近では, 実時間でのイメージ転送システムのため画像の圧縮技術や情報転送技術の開発が盛んとなり, テレビ電話や監視装置を実現しつつある. しかし, コミニュケーションに本当に必要な要素は膨大な画像データの中の一部分であり, 画像データ自体にはあまり意味のないことが多い.

一方, コンピュータ・ビジョンの研究者に 物体認識手法としてよく知られている固有空間解析手法は, 画像から特徴量を抽出する手法として広く使われている. 例えば, Pentland の eigen-face 手法は画像の膨大なデータベースを低次元に投影することにより, 人物の特定するシステムを提案している [11] [12] [13].

そこで本論文では，テレ・サービスに代表される不特定者とのコミニュケーションであり，人物を特定する必要のない環境での人間同士のスムーズな遠隔コミニュケーションを実現する．ここではこのための要素として顔の表情に注目し，固有空間手法を用いたリアルタイムの遠隔コミュニケーションシステムとして表情転送システムを実現する．まず，固有空間解析手法を用いた顔の表情空間を提案し，幾つかの顔の表情を分類する．次に，受け手側，送り手側の二つの表情空間を対応付ける手法を論じる．最後に，顔の表情を実時間で転送するシステムを示す．

2. 表情

顔表情は人間の感情を表すものである．そのため顔表情の分類は，多くの心理学者らによって論じられている．その中の分類法として良く用いられるものは，6つの要素「楽しみ」「悲しみ」「驚き」「嫌悪」「怒り」「恐れ」で構成されるとするものである．もちろん人間社会では，年齢，性別，文化的，民族的によって表情は異なり，また，表情には表さない心理的な感情や，嘘の表情も存在する [2]．しかし，ここでは表情と感情の関係については深く追求はしない．

また画像から表情の分析方法は，多くの研究者によって提案されている．例えば，目，鼻，口，眉 の動きに注目したものであり [3] [4] [5] [6] [7] [8] [9] [10] しかしこれらの多くは実時間での応用が難しいものである．

3. 表情空間 FES (Facial Expression Space)

近年，特定の人物の顔認識手法の一つとして，固有空間手法を応用した eigen-face 手法が提案されている [11] [12] [13]．この固有空間手法は，コンピュータ・ビジョン分野の物体認識で多く用いられたものである [14] [15] [16]．この手法の利点は，元のイメージの次元を，線や点のようなイメージの特徴抽出によらずに解析的に低次元化可能であることである．

ここでは固有空間手法を簡単に紹介する．

3.1. 固有空間手法

一般的に，あるデータ集合に対し自乗平均誤差最小の規範により，この集合を最適近似する新たな低次元データ集合を作り出すことが可能である． この場合その集合中の新しい個々のデータは，元のデータ集合の特徴を表すと考えられ，この手法を画像に用いれば，これは一種の画像特徴抽出となる．

以下では画像集合から特徴抽出する方法について述べる．

M 枚の画像を学習画像とし，得られた二次元画像をスキャンして N 次元ベクトル $z_i(i = 1,2,\cdots,M)$ とし学習マトリクス $\mathbf{Z}$ を構成する．

$$\mathbf{Z} = [\mathbf{z}_1 - \mathbf{c}, \mathbf{z}_2 - \mathbf{c}, \cdots, \mathbf{z}_M - \mathbf{c}]. \qquad (1)$$

ここで，$\mathbf{c}$ は平均画像である．

次に，学習マトリクス $\mathbf{Z}$ より，画像集合の共分散マトリクス $\mathbf{Q}(N \times N)$ を得る．

$$\mathbf{Q} = \mathbf{Z}\mathbf{Z}^T. \qquad (2)$$

この共分散マトリクスから，次式を満たす固有値 $\lambda_i(\lambda_i \geq \lambda_2 \geq \cdots \geq \lambda_N)$ と固有ベクトル $\mathbf{e}_i(i = 1,2,\cdots,N)$ を導く．

$$\lambda_i \mathbf{e}_i = \mathbf{Q}\mathbf{e}_i. \qquad (3)$$

このことにより，各々の画像はこの固有値と固有ベクトルを用いて再構成することが可能であり，その時の各固有ベクトルの重みは対応する固有値で与えらえる．

ここで十分小さい固有値の項を無視することにより，学習画像の次元を落とすことが可能である．有効な次数の決定には下式の固有値累積寄与率 W_k と適当な閾値 T_s を用いて決定することとする．

$$W_k = \frac{\sum_{i=1}^{k} \lambda_i}{\sum_{i=1}^{N} \lambda_i} \geq T_s. \qquad (4)$$

ここで，次元 N より十分小さい k 個の固有ベクトルにより構成されたマトリクス $\mathbf{E} =$

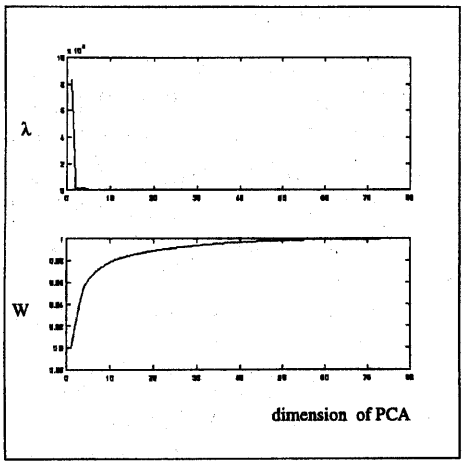

図 3: Eigen Values.

$[\mathbf{e}_1, \mathbf{e}_2, \cdots, \mathbf{e}_k]$ を用い次式より次元 N の画像ベクトル $\boldsymbol{\zeta}_i$ を次元 k の固有ベクトル $\boldsymbol{\zeta}_i$ へ投影することが可能となる.

$$\boldsymbol{\zeta}_i = \mathbf{E}^T(\mathbf{z}_i - \mathbf{c}). \qquad (5)$$

3.2. 表情分類

顔表情の典型的分類手法は前述したように六つの感情への分類である. しかし, ここでは簡単のため, 主な四つの感情「怒り」「無表情」「驚き」「笑い」を用い, 顔表情を固有空間手法を用いて分類を行ない, 手法の妥当性を検証する.

画像データは, CCD カメラ (SONY EVI-D30) を用いて撮影し, SGI Indigo2 へ $243 \times 320 pixel$ $8bits$ として取り込まれる. この時, 画像内の顔の位置を一定にするため, カメラの自動追尾, 自動ズーム, 自動絞りを用い, 画像の取り込みの際に生じる, 顔の位置のずれなどの影響を軽減した. また, 背景は黒とした.

ここで得られたオリジナル画像データは, 式 (3) を用い固有ベクトルと固有値を計算し, 主成分の重みと考慮して式 (5) により表情空間へ投影する.

図 3 はこの時の固有ベクトルと累積寄与率を示す. この図より, 画像データの特徴量として, 元の画像の 95% を再構成するためには三次元で十分であることが理解できる. そこで以降では, 画像データの特徴を表すため, 三次元の表情空間を用いる.

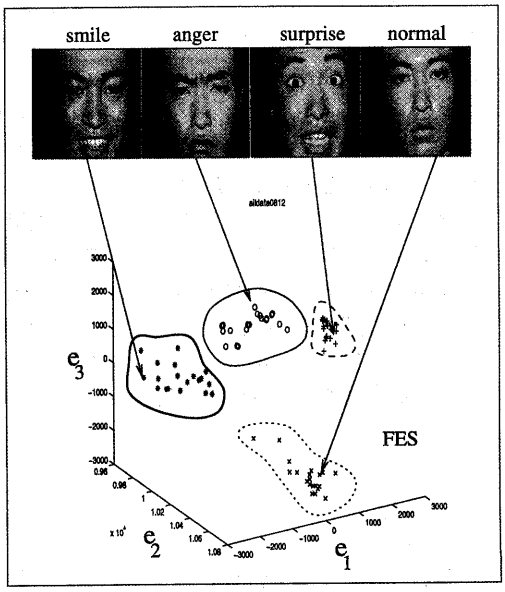

図 4: Classification of Facial Expressions with PCA.

また, 図 4 は典型的な顔表情を分類した結果を示している. ここで四つの感情を示す顔画像を撮影するために被験者に出した指示として, 「それぞれの感情を最大に顔に表して下さい」という命題を与えた. この結果より, 前述した固有空間手法を用い, 三次元のみ表情空間による, 表情分離が可能であることを検証した.

4. 表情空間の対応付け

前章では, 表情空間が固有空間手法で実現できることを示した. しかし一般的に顔の表情空間は, 個人の顔の特徴からなる情報を含んでいるため, この表情空間は個人特有のものである.

本論文での最終目的である, 遠隔において異なる顔画像を実時間で再生するためには, 異なる表情空間同士の対応付けが必要不可欠となる (図 5). ここで前章で述べたように, 固有空間手法による表情空間は三次元で特徴を表すのに十

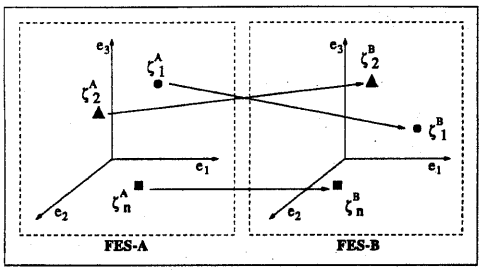

図 5: Projection from **FES-A** to **FES-B**.

分であることが示されたことから，それぞれの表情空間は三次元のアフィン変換で投影できると仮定する．したがって，人物 A, B の各々の表情空間内の任意の表情の特徴点を各々 ζ^A ζ^B とすると，この二つの特徴点は以下の式で投影されるとする．

$$\zeta_i^B = R^{AB}\zeta_i^A + T^{AB}. \qquad (6)$$

ここで，R^{AB}, T^{AB} は各々 3×3, 3×1 のアフィン変換マトリクスを示す．

ここで，実際に二人の人物について各々の表情空間を形成し，最小自乗近似手法を用いて上式の各パラメータを導くと，以下の式が得られた．

$$R^{AB} = \begin{bmatrix} -0.8464 & 0.0568 & -0.0964 \\ -3.7396 & 1.3313 & -0.8140 \\ 7.1058 & 0.0482 & 0.4535 \end{bmatrix},$$

$$T^{AB} = \begin{bmatrix} 1.9726e+04 \\ 3.7767e+04 \\ -7.1745e+04 \end{bmatrix}. \qquad (7)$$

したがって，図 6に示すように，各々の表情画像は各々の表情空間媒介にして結合することが可能となり，各々の関係は式 (5), (6) により単純な投影により求めることが可能である．

5. リアルタイム表情転送システム

この章では，リアルタイムの表情転送システムを提案する． ここで再度，特記したいのは，“表情転送” は従来の “画像転送” とは異なる表情そのものを転送するものであり，このことに

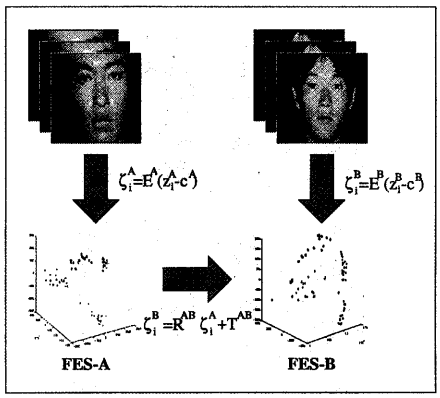

図 6: Relation between Two Facial Images.

より再生側では違う人間の表情画像を再生することが可能となるものである．

そこで表情転送と画像転送の違いを明らかにするため，再生側では別人の同じ表情を再生するシステムの構築を行なう．

図 7 に システムの構成を示す．システムは大きく分けて以下のように三つに分けられる．

[送信側] 一連の画像を逐次取り込み，先に得られた 式(5) により，表情空間 FES-A へ投影する．更に，式 (6) により，表情空間 FES-A より FES-B へ投影する．

[送信] 三次元の表情空間点情報を転送する．

[受信側] ローカルに持っている FES-B の表情空間データベースと，転送された FES-B の表情空間点の対応を計算し，ノルムの最小となる画像データを再生する．

この章に至るまで，表情の主な構成として四つの主要な表情を取り上げ，各々の分類と，異なる表情空間同士の対応付けを行なってきたが，一度，各投影マトリクスが得られたならば，実際に得られた画像の顔の表情がどの分類に属するかということは考える必要がなく，単なる投影と対応付けだけで実現される．

図 8 には以上の手法を用いて再生された表情画像の例を示す．この結果，多くの表情画像の対応は，妥当であることが検証されたが，時とし

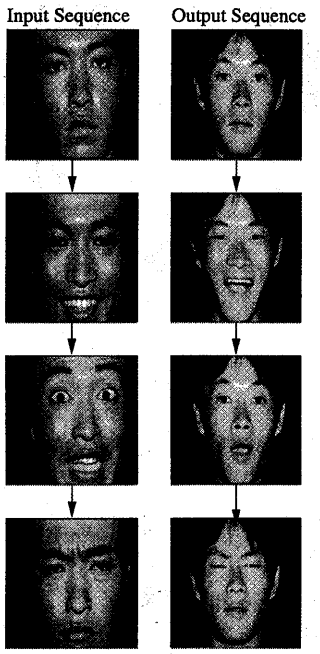

図 8: A Sample of Correspondence of Facial Expressions.

て入力された表情画像が曖昧な時は，再生された画像が適切な対応がなされているか否かの判定が難しい結果も得られた．

現在まで，実時間で再生するシステムの構築を行なっているが，まだ完成には至っていない．実際に実時間での再生を行なう場合には，表情の提示は $1/30sec.$ 毎に行なう必要はないものと思われる．

6. 結論

この論文では，画像から得られる人の表情画像を分類する手法として，主成分分析を提案し，その妥当性を明らかにした．また個々の人顔の表情空間同士を対応付ける手法を提案し，実時間で表情転送するシステムを提案し，その有効性を示した．今後，リアルタイムシステムを完成し，受け手側の評価を行なう．

参考文献

[1] "VRML2.0 Specification Appendix C. Java Scripting Reference", VRML Architecture Group, 1996.

[2] Paul Ekman and Wallace V. Friesen, "Unmasking the Face", Prentice-Hall, 1975

[3] K.Matsuno, Chil-Woo Lee, Satoshi Kimura, and Saburo Tsuji, "Automatic Recognition of Human Facial Expressions", ICCV'95, pp.352-359.

[4] K.Mase, "Recognition of facial expressions from optical flow", IEICE Trans. Special Issue on Computer Vision and its Applications, E74(10),1991.

[5] I.Essa, T.Darrell and A.Pentland, "Tracking facial motion", Proc. Workshop on Motion and Nonrigid and Articulated Objects, pp.36-42, 1994

[6] Irfan A. Essa and Alex P.Pentland, "Facial Expression Recognition using a Dynamic Model and Motion Energy", ICCV'95, pp.360-367, 1995.

[7] A.Lanitis, C.J.Taylor and T.F.Cootes, "A Unified Approach to Coding and Interpreting Face Images", ICCV'95, pp.368-373, 1995.

[8] M.Rosenblum, Y.Yacoob and L.Davis, "Human emotion recognition from motion using a radial basis function network architecture", The Workshop on Motion of Nonrigid and Articulated Objects, pp.43-49, IEEE Computer Society, 1994

[9] Y.Yacoob and L.S.Davis, "Labeling of human face components from range data", CVGIP-Image Understanding, 60(2), pp.168-178, 1994.

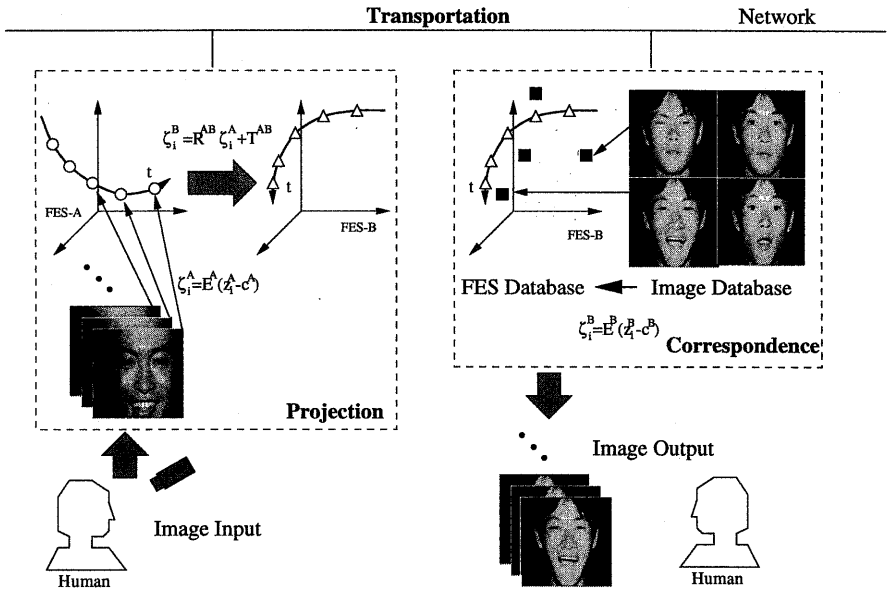

図 7: Transportation of Facial Expressions.

[10] Michael J. Black and Yaser Yacoob, "Tracking and Recognition Rigid and Non-Rigid Facial Motions using Local Parametric Models of Image Motion", ICCV'95, pp.374-381, 1995.

[11] M.Turk and A.Pentland, "Eigenfaces for Recognition", Journal of Cognitive Neuroscience, 3(1), pp.71-86, 1991.

[12] Matthew A. Turk and Alex P. Pentland, "Face Recognition Using Eigenfaces", Proc. CVPR 1991, pp.586-591, 1991

[13] Baback Moghaddam and Alex P. Pentland, "Face Recognition using View-Based and Modular Eigenspaces", Automatic Systems for the Identification and Inspection of Humans, SPIE Vol. 2277, 1994.

[14] Hiroshi Murase and Shree K. Nayar, "Visual Learning and Recognition of 3-D Objects from Appearance", International Journal of Computer Vision, Vol.14, No.1, pp.5-24, 1995.

[15] Erkki Oja, "Subspace Methods of Pattern Recognition", Research Studies Press Ltd., 1983.

[16] H. Murakami and V. Kumar, "Efficient Calculation of Primary Images from a Set of Images", IEEE Transactions of Pattern Analysys and Machine Intelligence, Vol.4, No.5, pp.511 - 515, 1982.