

## -Color Enhanced Emotion- 実時間表情認識を用いた感情強調システムの開発

牧野 貴雄<sup>1</sup>, 中口 俊哉<sup>1,2</sup>, 津村 徳道<sup>1,2</sup>, 高瀬 紘一<sup>1</sup>,  
岡口 紗綾<sup>1</sup>, 小島 伸俊<sup>4</sup>, 三宅 洋一<sup>1,2,3</sup>

<sup>1</sup>千葉大学大学院 自然科学研究科 知能情報工学専攻 <sup>2</sup>千葉大学 工学部 情報画像工学科

<sup>3</sup>千葉大学フロンティアメディカル工学研究開発センター <sup>4</sup>花王株式会社

E-mail : makino@graduate.chiba-u.jp, {nakaguchi, tsumura}@faculty.chiba-u.jp,  
{takase, okaguchi}@graduate.chiba-u.jp,  
ojima.nobutoshi@kao.co.jp, miyake@faculty.chiba-u.jp

**概要** 本論文では、ビデオチャット等のアプリケーションにおける新たなコミュニケーションツールとして、使用者の感情を誇張して表示する顔色変化シミュレーションシステムの提案を行う。まず2台のCCDカメラで撮影した動画像から自動的に使用者の顔領域と表情を認識する。認識した表情に合わせて顔の肌色をリアルタイムに変化させて表示することによって、感情を誇張することが可能である。本システムでは肌色変化処理として、独立成分分析に基づく色素成分解析法を用いている。肌の物理特性に基づいた処理であるため、日焼けや紅潮など現実に近い肌色変化が可能である。肌色変化処理においてはGraphics Processing Unit(GPU)を、表情認識においては色情報による高速な追跡・認識手法を用いることで、動画像に対してもリアルタイムに処理を実行することができる。結果画像に対する評価実験を行い、本システムが単純な画素値変化と比較して感情強調に有効であることが確かめられた。

**キーワード** : ビデオチャット, コミュニケーションツール, 表情認識, 顔追跡, 顔色変化

## Development of the Facial-Color Simulator with Real-Time Face Recognition

Takao Makino<sup>1</sup>, Toshiya Nakaguchi<sup>1,2</sup>, Norimichi Tsumura<sup>1,2</sup>, Koichi Takase<sup>1</sup>,  
Saya Okaguchi<sup>1</sup>, Nobutoshi Ojima<sup>4</sup>, Yoichi Miyake<sup>1,2,3</sup>

<sup>1</sup>Graduate School of Science and Technology, Chiba University

<sup>2</sup>Department of Information and Image Sciences, Chiba University

<sup>3</sup>Research Center for Frontier Medical Engineering, Chiba University <sup>4</sup>Kao Corporation

**abstract** : This paper proposes a new real-time video editing system that can enhance the emotion. This system can be applied as a communication tool like a video chat. First, this system recognizes the facial expressions by using two CCD cameras. Then, user's emotion is enhanced by controlling the facial-color according to the recognized facial expression. In this system, a real-time skin pigment control method with Graphics Processing Unit(GPU) is used for controlling the high-realistic facial-color. In the facial emotion recognition, face tracking and recognition are performed only using the color information to achieve the real-time processing. Finally, we evaluate the effectiveness of the developed system, then high effectiveness for the emotion enhancement is obtained.

**Keywords** : Video Chat, Communication Tool, Facial Expression Recognition, Face Tracking,  
Facial-Color Control

## 1. はじめに

人同士のコミュニケーションにおいて、感情を正確に伝えることは非常に重要である。メールのような文字のみを扱うコミュニケーションツールでは、感情を相手に伝達するために絵文字や顔文字といった感情・状態の可視化表現が用いられている。これらの表現は厳密に実際の感情を表してはいないものの、感情を表に出すことが苦手な人でも絵文字を入力するだけで簡単に感情を相手に伝えることが可能という利点を持つ。

近年では、コミュニケーションツールとして、テレビ電話やビデオチャットに代表される“ビデオコミュニケーション”が広く普及し始めており、ビデオチャットによる語学学習<sup>[1]</sup>やカメラ付き携帯電話を用いたテレビ電話<sup>[2]</sup>などが既に実現されている。ビデオコミュニケーションは電話・メール等と異なり、使用者がお互いに顔を合わせることで表情・動作などを用いたコミュニケーションが可能という利点を持つ。特に表情は感情伝達の重要度においてその55%を占める重要な要素であり<sup>[3]</sup>、遠隔地の人物とのコミュニケーションを円滑にするツールとして注目されている。

しかしビデオコミュニケーションには解決すべき問題点がいくつか挙げられている。まず、表情を作ることが苦手な人が多く、特に日本人は表情を用いた感情表現に乏しいと言われている。Argyleら<sup>[4]</sup>の実験によれば、日本人の表情はイギリス人は38%、イタリア人は28%しか理解することができず、同じ日本人が判断してもその理解度は45%にとどまるという結論が示されている。すなわち、日本人の表情を他者が理解することは容易ではなく、ビデオコミュニケーションにおいても満足に感情伝達が行えないと考えられる。

さらに、今後ビデオコミュニケーションが普及していくにしたがって、携帯電話のような低解像度・低通信速度の媒体による動画像表示、逆光・照明不足などの不適切な環境での撮影など、劣悪な環境において使用されることが多くなると考えられる。このような環境下では動画像の質が悪く、表情の伝達が一層困難になる。以上の理由より、ビデオチャットにおける話者の表情の理解は必ずしも容易ではなく、それに伴い正確な感情伝達が行えないという問題点が挙げられる。

そこで本論文では、ビデオコミュニケーションにおいて感情を可視化することができる新たなコミュニケーションツールを提案する。画像処理によって使用者の表情を認識し、その表情に対応した肌色変化処理を行うことで感情を強調して相手に伝えることができる。これはメールにおける絵文字・顔文字に相当するものであり、誰にでも簡単に感情表現が可能な利便性と、感情・状態を可視化することにより生じる高いエンターテインメント性を持つ。

## 2. 構築するシステムの概要

本章ではビデオコミュニケーションやそれに用いる感情強調システムに関する先行研究と、システムの概要について述べる。

### 2.1. 先行研究とその問題点

ビデオコミュニケーションでは、使用者の顔の代わりにCGキャラクターを表示して様々な動作を行わせるアバター<sup>[5]</sup>というシステムが実用化されている。このシステムは感情表現だけでなく、エンターテインメント性や使用者のプライバシーの保護という観点からも有用であるとされている。しかしアバターは使用者の実際の表情・状態を伝達しないため情報の損失が起こる、語学学習<sup>[1]</sup>のように実際の顔を用いることに意味があるアプリケーションには応用できない等の問題点がある。

一方、本論文の目的である感情強調システムに関する先行研究としては、顔画像の表情そのものを過度に変化させて感情を表現するものが挙げられる。前述のように表情は感情伝達において最も重要な要素であり、その感情強調の効果は非常に高いものと考えられる。現在までに経験的モデルを用いた手法<sup>[6]</sup>や顔の筋肉を考慮したモデル<sup>[7]</sup>を用いた手法などが提案されている。しかし、これらの手法は計算負荷が非常に高くリアルタイムに実現するのが困難であるという問題点があった。

### 2.2. 提案する感情強調システム

本論文では顔の色を用いて感情強調を行う。赤は怒り、青は悲しみ、のように色は人に感情を連想させる効果を持っており<sup>[8]</sup>、色を適切な色に変化させることで感情を強調できると考えられる。また、顔色変化は表情変化と比較

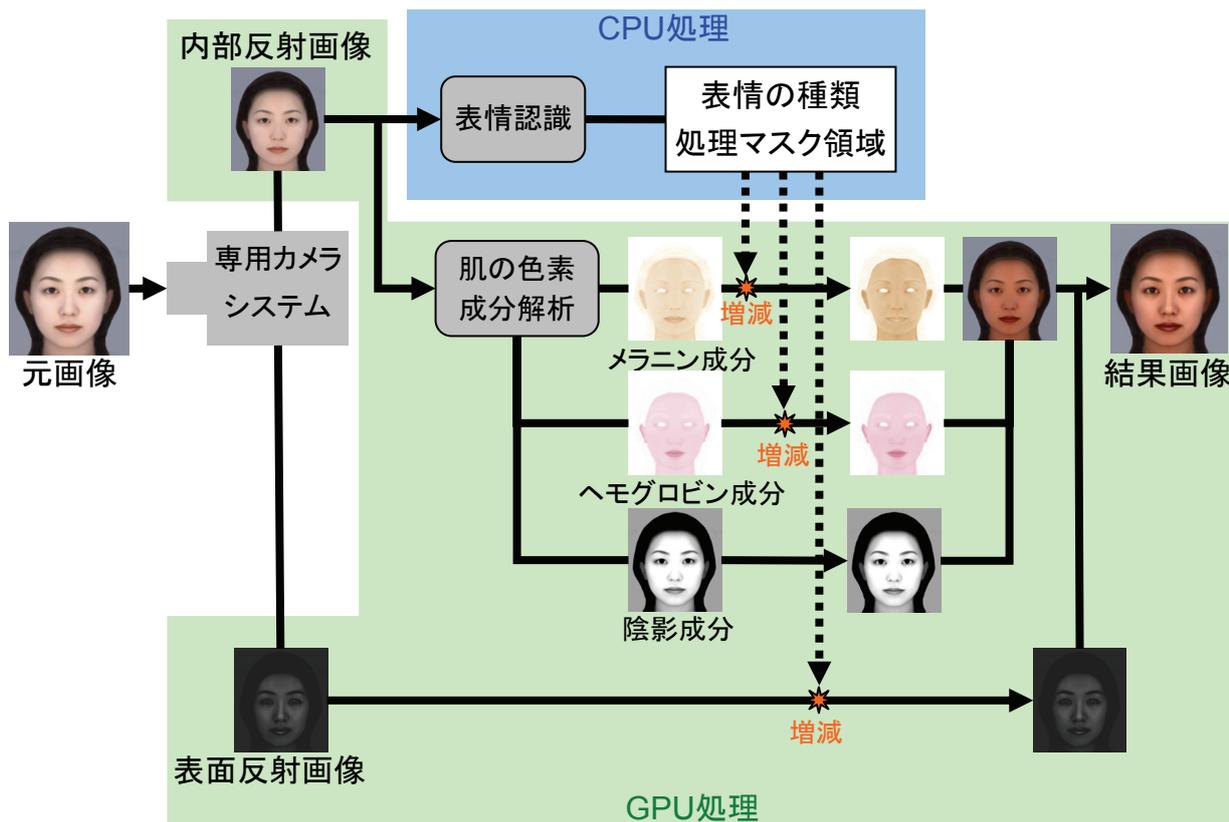


図1 システムのデータフロー

して計算負荷が少なく、リアルタイム処理に向いている。しかし、画像のRGB値の増減などの単純な色変化では現実の顔色変化とかけ離れて不自然な顔になってしまう。そこで本論文では、肌の物理特性に基づく顔色変化手法を用いて、話者の感情に適した自然な顔色に変化させる感情強調システムを提案する。

また、本システムではコミュニケーションツールとしての利便性を考慮し、インターフェースに話者の顔・身振りのみを用いる。歯を見せる、手を上げるといった会話中に自然に行う動作のみから顔色変化を制御できる。

提案するシステムは専用カメラシステム、肌の色素成分解析、表情認識の3つの主要素から成り立っている。図1に本システムのデータフローを示す。専用カメラシステムは動画の表面反射成分と内部反射成分をリアルタイムに分離して撮影可能とする。肌の色素成分解析とは、独立成分分析により内部反射成分のみの顔色画像からメラニン成分とヘモグロビン成分に分解することで、RGB値による単純な画像処理では表現できないリアルな肌色変化を行う手法である。表情認識は顔のトラッキング技術を用

いて表情・動作の認識を行い、同時に肌色変化処理のマスク領域を決定する。色情報と簡単な形状のみを用いた計算負荷の少ない手法であり上記の肌色解析と並列に動作させることが可能である。また、この表情とは実際の表情ではなく、簡単に感情を表現するための顔の特定の動作を指している。認識した表情に基づいて肌色を変化させた画像を合成し、分離した表面反射成分と再び合成することによって感情を強調した顔画像を再現することができる。感情強調の他、日焼けや酔い等の顔の装飾を行うことも可能である。

### 3. 感情強調システムの実装

本章では感情強調システムの各要素の詳細な説明と、その具体的な実装方法について述べる。

#### 3.1. 表面反射光と内部反射光の分離法

肌からの反射光は、二色性反射モデル<sup>9)</sup>により肌表面で反射する表面反射光(表面反射成分)と肌内部で散乱する内部反射光(内部反射成分)に分けられる。表面反射成分は

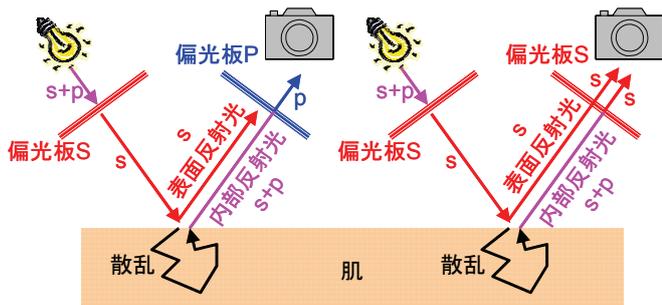
照明光の色, 内部反射成分は肌の色をそれぞれ表している. 本システムでは肌に対する色素成分解析を行うため, 表面反射成分の影響を取り除いた内部反射成分のみの画像(内部反射画像)が必要である. 従って, 表面反射成分と内部反射成分を分離して撮影しなければならない.

表面反射成分と内部反射成分の分離には, 偏光板を用いた手法が広く用いられている. 図2にこの手法の概念図を示す. 照明光の振動方向は反射面に平行な方向  $s$  と,  $s$  に対して垂直な方向  $p$  によって表される. 照明の前に  $s$  偏光のみを通す偏光板  $S$  を設置すると, 表面反射光は  $s$  偏光となるのに対し, 内部反射光は物体内部で散乱するために偏光性が失われる. 従って,  $p$  偏光のみを通す偏光板  $P$  をカメラの前に設置することで,  $s$  偏光である表面反射光を遮断し内部反射画像を撮影することが可能である. また, カメラの前に偏光板  $S$  を設置した場合, 表面反射成分と内部反射成分の両方を持つ画像が撮影される. この画像と内部反射画像の差から表面反射画像も取得可能である.

欠点として, 偏光板と2台の CCD カメラを必要とするため撮影システムが大掛かりになるという点, さらに偏光された照明下でしか使用できないという点が挙げられる. これらはビデオコミュニケーションに応用する際に大きな制約となるため, 今後の課題として解決する必要がある.

### 3.2. 専用カメラシステムの構築

偏光板, ビームスプリッター, そして2台の CCD カメラを用いて, 内部反射成分と表面反射成分を同時に取得可能とする専用カメラシステムを構築した. 図3に専用カメラシステムの構造を示す. 2台の CCD カメラの前には偏光方向の直交する2枚の偏光板がそれぞれ設置されている. 偏光された照明で照らされた被写体を撮影することで表面・内部反射成分の両方を持つ画像と内部反射画像を同時



(a) 内部反射画像の撮影 (b) 表面・内部反射画像の撮影

図2 偏光板を用いた表面・内部反射成分分離

に撮影することができる. それらの差分により表面反射画像も取得可能である. 得られる内部反射画像と表面反射画像の例を図4に示す.

### 3.3. 肌の色素成分解析法

自然な肌色変化を行うために, 本システムでは肌の色素成分解析法<sup>[10]</sup>を用いてリアルな肌色変化を行う. 以下, 肌の色素成分解析の詳細について述べる. 人の肌色は, 主に黒褐色であるメラニン色素と赤色であるヘモグロビン色素の2つによってほぼ決定される. 肌をメラニン成分の多い表皮層とヘモグロビン成分の多い真皮層の2層構造であると仮定すると, それらは互いに独立であると考えられる. 従って, 肌色に対して独立成分分析<sup>[13]</sup>を行うことによってメラニン・ヘモグロビン色素成分を計算することができる.

独立成分分析を行う肌色モデルの導出を以下に示す. 肌への入射光を  $E(x, y, \lambda)$ , メラニン色素の濃度を  $\rho_m(x, y)$ , ヘモグロビンの濃度を  $\rho_h(x, y)$ , それぞれの色素間の交差係数を  $\sigma_m(\lambda), \sigma_h(\lambda)$  とすれば, デジタルカメラのセンサ応答  $v_i(i=r, g, b)$  は次式で表される.

$$v_i(x, y) = \int c e^{-\rho_m(x, y)\sigma_m(\lambda)l_e(\lambda) - \rho_h(x, y)\sigma_h(\lambda)l_d(\lambda)} E(x, y, \lambda) s_i(\lambda) d\lambda \quad (1)$$

ここで,  $\lambda$  は波長,  $l_e(\lambda), l_d(\lambda)$  はそれぞれ表皮層, 真皮層の光路距離,  $c$  はデジタルカメラの感度により決定される定数,  $s_i(\lambda)(i=R, G, B)$  はデジタルカメラの分光感度である. 分光感度が狭周波数帯で,  $s_i(\lambda) = \delta(\lambda_i)$  で近似

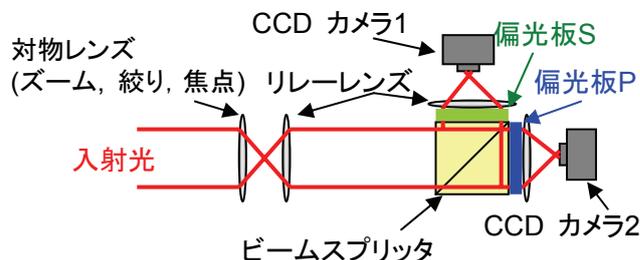


図3 構築した専用カメラ



図4 左:内部反射画像 右:表面反射画像

可能であり、入射光が単色光  $E(x, y, \lambda) = p(x, y)\bar{E}(\lambda)$  と仮定すると、次式が得られる。

$$v_i(x, y) = ce^{-\rho_m(x, y)\sigma_m(\lambda)l_e(\lambda) - \rho_h(x, y)\sigma_h(\lambda)l_d(\lambda)} p(x, y)\bar{E}(\lambda_i). \quad (2)$$

式(2) の対数を取り、濃度へ変換すると、次式のようにベクトルと行列で表される。

$$\mathbf{v}^{\log}(x, y) = -\rho_m(x, y)\sigma_m - \rho_h(x, y)\sigma_h + p^{\log}(x, y)\mathbf{h} + \mathbf{e}^{\log}. \quad (3)$$

ここで、

$$\begin{aligned} \mathbf{v}^{\log} &= [\log(v_R(x, y)), \log(v_G(x, y)), \log(v_B(x, y))]^T, \\ \sigma_m &= [\sigma_m(\lambda_R)l_e(\lambda_R), \sigma_m(\lambda_G)l_e(\lambda_G), \sigma_m(\lambda_B)l_e(\lambda_B)]^T, \\ \sigma_h &= [\sigma_h(\lambda_R)l_d(\lambda_R), \sigma_h(\lambda_G)l_d(\lambda_G), \sigma_h(\lambda_B)l_d(\lambda_B)]^T, \\ \mathbf{h} &= [1, 1, 1]^T, \\ \mathbf{e}^{\log} &= [\log(E(\lambda_R)), \log(E(\lambda_G)), \log(E(\lambda_B))]^T, \\ p^{\log}(x, y) &= \log(p(x, y)) + \log(c). \end{aligned}$$

である。従って、観測信号  $\mathbf{v}^{\log}$  は図 5 のように、線形に組み合わせたメラニン・ヘモグロビン・陰影成分の各ベクトル  $(\sigma_m, \sigma_h, \mathbf{h})$  とバイアスのベクトル  $\mathbf{e}^{\log}$  で表される。このモデルに基づき、観測信号に対して独立成分分析を実行することにより  $(\sigma_m, \sigma_h)$  を計測可能である。

$(\sigma_m, \sigma_h)$  と独立成分分析により得られる上位二位までの特徴ベクトルとの対応付けは、対応が取れていない状態で各成分の濃度を増減した画像を作成し、その見えにより判別することができる。しかしこの方法は手間がかかるため、実際には特徴ベクトル 2 つと、顔画像データベースよりあらかじめ求めておいたメラニンベクトル群とのマハラノビス距離を計算し、距離の近い方のベクトルを  $\sigma_m$ 、もう一方を  $\sigma_h$  と対応付ける。

図 5 に示すように、観測信号  $\mathbf{v}^{\log}$  を  $(\sigma_m, \sigma_h)$  が貼る平面に対して陰影ベクトルと平行に射影することで、それぞれの色素成分量  $(\rho_m(x, y), \rho_h(x, y))$  を得る。

実際に色素成分解析処理を行う場合、 $(\sigma_m, \sigma_h)$  は 1 人に対してただ 1 つに定まり不変と考えられるため、毎フレームの画像から  $(\sigma_m, \sigma_h)$  を求める必要はない。そこで本システムでは、使用前に正面向きの顔画像を 1 枚撮影し、その顔画像の頬の領域に対して独立成分分析を行い  $(\sigma_m, \sigma_h)$  を求めておく。実際の処理はこの  $(\sigma_m, \sigma_h)$  を用いて行う。

### 3.4. GPU を用いた肌色変化処理

肌の色素成分解析法は前述の表情変化手法<sup>[6][7]</sup>と比較

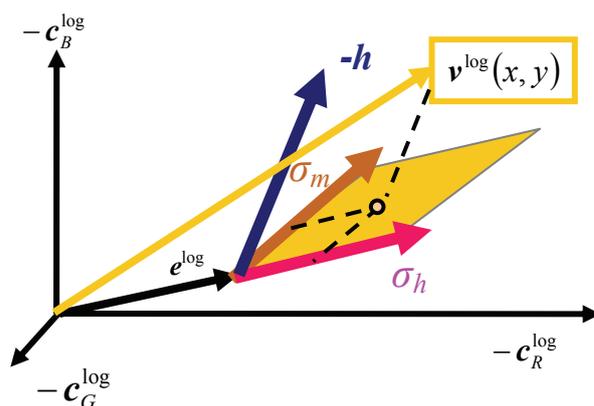


図 5 独立成分分析を行う肌モデル

して高速であるものの、現在の CPU による処理ではリアルタイムに実行することは困難である。そこで本システムでは肌の色素成分解析法に GPU - NVIDIA GeForce6800 GT (SM3.0) を用いる。色素成分解析法は顔画像の画素ごとに独立な処理であるため、GPU を用いて並列演算を行うことによりリアルタイムに実行することが可能である。図 1 の緑色の領域は GPU を用いる処理を示している。最初に専用カメラからの入力を表面反射画像と内部反射画像に分離し、以降の処理はすべて GPU によりリアルタイムに演算する。内部反射成分はさらに肌の色素成分解析によりメラニン成分、ヘモグロビン成分、そしてそれらを除いた陰影成分に分離される。メラニン成分とヘモグロビン成分の濃度を、認識した表情にあわせて人工的に変化させ、再び RGB 画像に合成することでリアルに変化した肌色を再現することができる。

### 3.5. 頭部および顔部位の追跡

本システムでは、前節までに述べた肌の色素成分解析法により感情強調・顔の装飾を行う。そのためには動画像中から画像処理を行うマスク領域を自動で決定し、顔の動きに合わせてマスク領域を追従移動させなければならない。また、システムのインターフェースとして話者の表情・動作を用いるため、画像中からそれらの動きを認識する必要がある。これらの処理は既に述べたカメラ処理や肌の色素成分解析と並行に行わなければならないため、できるだけ計算負荷の低い手法が望ましい。

本システムでは、高速に頭部と顔部位の追跡を行う手法<sup>[12]</sup>を用いて表情を認識する手法を提案する。図 6 に追跡

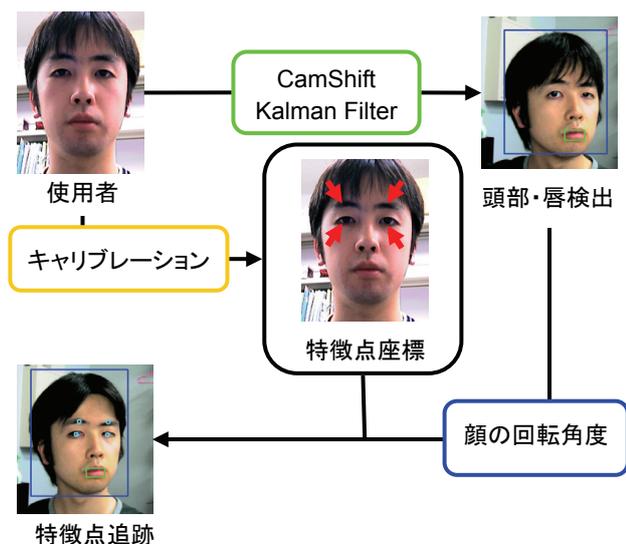


図 6 追跡手法の流れ

手法の流れを示す。前処理として、使用者が追跡したい顔部位の 3 次元座標を、正面向きの顔画像から手動によりキャリブレーションを行っておく。まず頭部重心と唇位置を追跡し、それらを用いて頭部の回転角度と平行移動ベクトルを推定する。回転角度、平行移動ベクトルと顔部位の 3 次元座標を用いて計算を行うことで移動後の顔部位の座標の計算が可能である。そして追跡した部位の動きから表情認識を行う。

以下、説明のために頭部重心と唇の画像平面上の座標をそれぞれ  $(x_h, y_h)$ ,  $(x_l, y_l)$ 、カメラを原点とする 3 次元空間での頭部重心と唇の座標を  $(X_h, Y_h, Z_h)$ ,  $(X_l, Y_l, Z_l)$ 、画像中心の座標を  $(x_o, y_o)$ 、カメラの焦点距離を  $f$  とする。また、この追跡手法では  $z$  方向への平行移動と  $z$  軸回りの回転角度を計算することはできないため、これらの動きはないものと仮定し、カメラから頭部までの距離  $Z_h$  はあらかじめ測定し一定とする。

追跡前のキャリブレーションでは、追跡する部位の 3 次元座標を顔画像から取得する。本システムでは目を表情認識やマスク領域の決定に用いるため、両目を特徴点として追跡する。そこで図 7 に示すように正面向きの顔画像から両目の座標を使用者がマウスで指定することで、両目の  $x$  および  $y$  座標を取得し、頭部重心から目までの距離  $L$  から  $z$  座標を算出する。正面向きの画像からマウスで選択した座標を  $(x_f, y_f)$  とすると、そのカメラ座標系における 3 次元座標  $(X_f, Y_f, Z_f)$  は



図 7 特徴点の 3 次元座標取得

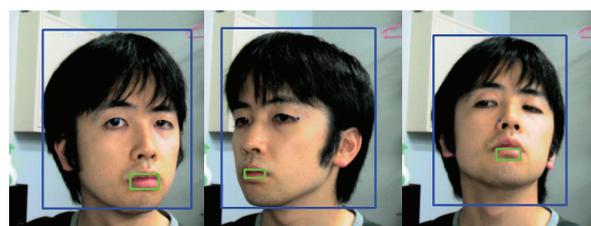


図 8 唇・頭部追跡結果

$$X_f = \frac{x_f - x_o}{f} Z_h, \quad Y_f = \frac{y_f - y_o}{f} Z_h, \quad Z_f = Z_h - L \quad (4)$$

となる。ここで、 $L$  は頭部模型を用いて予め計測しておく。本システムでは両目の 3 次元座標  $(X_{f1}, Y_{f1}, Z_{f1})$  と  $(X_{f2}, Y_{f2}, Z_{f2})$  をキャリブレーションにより取得する。

実際の追跡処理では、まず唇と頭部重心の追跡を行う。唇と頭部重心の位置の検出には CamShift と KalmanFilter を併用した手法<sup>[11]</sup>を用いた。この手法では任意の色を持つ対象を正確かつ安定に検出・追跡することが可能である。図 8 に検出結果を示す。緑の長方形が唇の位置、青い長方形が頭部重心を中心とする頭部領域を示している。この図より、唇と重心が検出・追跡されていることがわかる。

検出された唇と頭部重心の位置を用いて回転角度と平行移動ベクトルを計算する。唇などの顔上の特徴点は頭部の回転でも平行移動でも移動するのにに対し、頭部の重心は回転では移動しないという違いがある。従って、頭部重心と唇の動きの差から頭部の仰角、方位角方向の回転角度を、頭部重心のみの動きから頭部の  $x$ ,  $y$  方向への平行移動距離を求めることができる。頭部の平行移動ベクトル  $\mathbf{t}$  は次の式で表される。

$$\mathbf{t} = \left[ \frac{x_o - x_h}{f} Z_h, \quad \frac{y_o - y_h}{f} Z_h, \quad 0 \right]^t \quad (5)$$

ここで、 $z$  方向への平行移動はないものと仮定しているので3列目の値は0である。次に頭部の回転角度の推定について述べる。図9に角度推定概念図を示す。図9上段は方位角回転の際の顔の動き(正面から見た図)、下段はその動きを上面から見た図である。頭部の方位角方向の回転角度 $\theta_y$ は、図9に示すように唇の回転半径 $r$ と、頭部重心と唇位置の $x$ 座標の差 $w$ を用いて

$$\theta_y = \sin^{-1}\left(\frac{w}{r}\right) = \sin^{-1}\left(\frac{X_l - X_h}{r}\right) \quad (6)$$

と表される。ここで、 $r$ は前述の $L$ を計測する際に同時に求めておく。 $w$ は図10に示すように、カメラの焦点距離 $f$ 、唇位置の $z$ 座標 $Z_l$ との間に

$$\frac{x_l - x_h}{f} = \frac{w + w'}{Z_l} \quad (7)$$

の関係が成り立つ。ここで

$$Z_l = Z_h - \sqrt{r^2 - w^2} \quad (8)$$

$$\frac{w'}{\sqrt{r^2 - w^2}} = \frac{t_x}{Z_h}, \quad w' = \frac{t_x \sqrt{r^2 - w^2}}{Z_h} \quad (9)$$

であり、 $t_x$ は $\mathbf{t}$ の $x$ 成分を示す。式7を $w$ について解くことで $w$ が得られ、式6より $\theta_y$ が求まる。仰角方向の回転角度 $\theta_p$ についても同様に、頭部重心と唇位置の $y$ 座標の差 $h$ を用いて

$$\theta_p = \sin^{-1}\left(\frac{h}{r}\right) = \sin^{-1}\left(\frac{Y_l - Y_h}{r}\right) \quad (10)$$

であり、次式

$$\frac{y_l - y_h}{f} = \frac{h + h'}{Z_l} \quad (11)$$

$$Z_l = Z_h - \sqrt{r^2 - h^2} \quad (12)$$

$$\frac{h'}{\sqrt{r^2 - h^2}} = \frac{t_y}{Z_h}, \quad h' = \frac{t_y \sqrt{r^2 - h^2}}{Z_h} \quad (13)$$

より $\theta_p$ が求まる。 $t_y$ は $\mathbf{t}$ の $y$ 成分を示す。推定した回転角度を用いて、回転行列 $\mathbf{R}$ は以下ようになる。

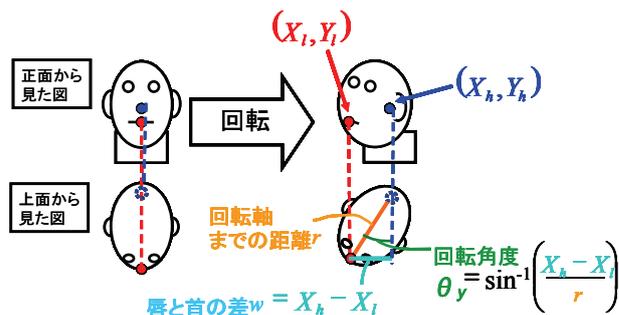
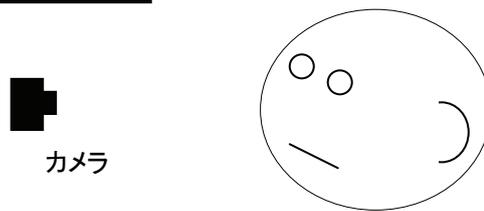


図9 顔角度推定概念図

側面から見た図



上面から見た図

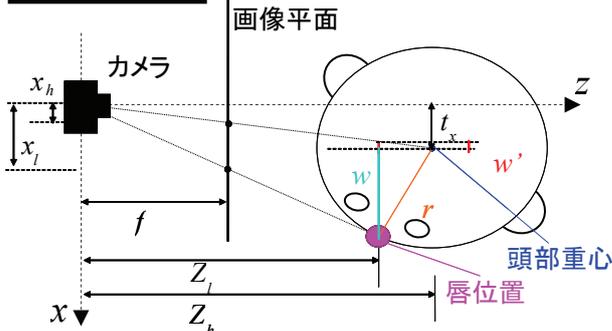


図10  $w$  と他のパラメータとの関係図

$$\mathbf{R} = \begin{bmatrix} \cos(\theta_y) & 0 & \sin(\theta_y) \\ 0 & \cos(\theta_p) & -\sin(\theta_p) \\ 0 & 0 & 1 \end{bmatrix} \quad (14)$$

最後に、回転行列と平行移動ベクトルを用いて、ユークリッド変換により頭部移動後の特徴点の座標を推定することで追跡を行う。両目の3次元座標 $(X_{f1}, Y_{f1}, Z_{f1})$ と $(X_{f2}, Y_{f2}, Z_{f2})$ を特徴点行列

$$\mathbf{f} = \begin{bmatrix} X_{f1} & Y_{f1} & Z_{f1} & 1 \\ X_{f2} & Y_{f2} & Z_{f2} & 1 \end{bmatrix}^T \quad (15)$$

とすると、前小節で求めた回転行列 $\mathbf{R}$ と平行移動ベクトル $\mathbf{t}$ を用いて、移動後の特徴点行列 $\mathbf{f}'$ は

$$\mathbf{f}' = \mathbf{T}_c \begin{bmatrix} \mathbf{R} & \mathbf{t} \\ \mathbf{0}_3^T & 1 \end{bmatrix} \mathbf{T}_m \mathbf{f} \quad (16)$$

ここで

$$\mathbf{0}_3^T = [0 \ 0 \ 0]^T,$$

$$\mathbf{T}_m = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & -Z_h \\ 0 & 0 & 0 & 1 \end{bmatrix},$$

$$\mathbf{T}_c = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & Z_h \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

であり、 $\mathbf{T}_m$ と $\mathbf{T}_c$ はそれぞれ頭部重心とカメラ位置に座標原点を変移させる変換行列である。この処理を各フレームに対して行うことにより、両目の位置を追跡することができる。この手法はキャリブレーション時に指定することで目以外の部位の追跡も可能である。例として図 11 に目と眉毛を追跡した結果画像を示す。

追跡手法の精度評価として、人間の頭部模型を方位角 $-80 \sim 80^\circ$ 、仰角 $-20 \sim 40^\circ$ まで回転させつつその動きを追跡し、実際の回転角度と推定された回転角度の誤差を計測した。本追跡手法は対象のおおまかな位置を追跡する手法であるため、回転角度の推定誤差は方位角方向で平均 $3.60^\circ$ 、最大 $10.00^\circ$ 、仰角方向で平均 $6.26^\circ$ 、最大 $20.60^\circ$ であった。特に回転角度が大きくなったときには唇が画面に映る面積が小さくなるため、その追跡が不安定になり、精度が悪くなることが確認された。しかし本システムでのインターフェースとしての表情認識やマスク領域の決定という用途においては、本手法の精度でも十分な効果が得られることを確認している。

### 3.6. 表情認識

本システムにおける表情とは、実際の感情を厳密に表しているものではなく、使用者が簡単に感情を表現するためのスイッチとしての役割を持つ。従って、厳密な表情認識の手法ではなく、顔部位の色や形状の変化のみから表情を判断する非常に簡素な手法を用いる。笑いと悲しみはそれぞれ口を横、縦に広げたときに、喜びは歯を見せたとき、

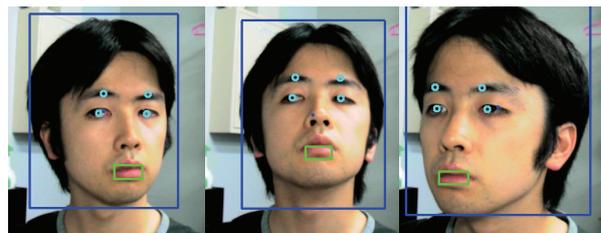


図 11 顔部位追跡結果

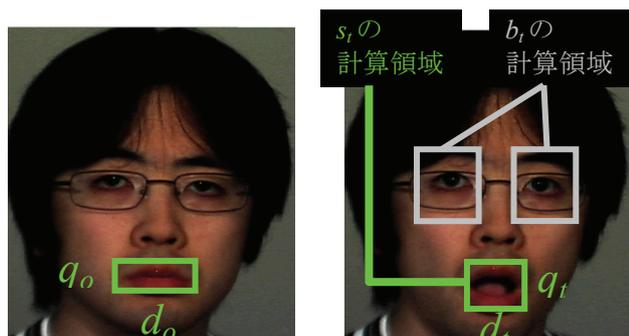
怒りは歯の検出に加えて右手をあげるという動作を行ったときに実行される。

また、エンターテインメントへの応用として、感情以外に酔いと日焼けという 2 種類の色変化の実現も行った。日焼けの効果はサングラスを装着することで実行され、サングラス型のマスクを用いることで、サングラス型の日焼け跡を再現することができる。また、酔いはビールを模した着色がされたジョッキを用いることで実行される。すべての効果は認識後、5 秒間だけ描画される。

本システムで扱う感情効果とその認識条件、肌色変化の内容を表 1 に示す。ここで、 $S(x,y)$ はカメラシステムにより得られる表面反射画像、 $d_o, q_o$ は顔の追跡開始時のフレームにおける唇領域の幅と高さ、 $d_t, q_t, s_t$ は現在のフレームにおける唇領域の幅、高さ、唇領域内に占める白色画素の割合をそれぞれ示している。白色画素とは YIQ 表色系における Q の値が 90 未満の画素を指す。 $b_t$ は現在のフレームにおいて追跡した両目の周囲 $d_o \times d_o$ の範囲に占める黒色画素の割合である。黒色画素とは RGB の各画素値がすべて 20 未満の画素を指す。図 12 にこれらのパラメータを図示する。右手やジョッキは頭部以外の領域から

表 1 表情と認識条件、色変化

感情・効果	認識条件	色素成分の変化
怒り	$s_t > 0.7$ かつ 右手検出	$\rho_h(x,y)+0.6$
喜び	$s_t > 0.7$	$S(x,y) \times 1.1$
笑い	$d_o/d_t < 0.7$	$\rho_h(x,y)+0.6$ (頬のみ)
悲しみ	$q_o/q_t < 0.3$	$\rho_m(x,y)-0.4$ $\rho_h(x,y)-0.4$
酔い	$ \theta_y  > 10$ かつ ジョッキ検出	$\rho_h(x,y)+0.6$
日焼け	$b_t > 0.8$	$\rho_m(x,y)+0.8$ (サングラス マスク以外)



追跡開始フレーム

現在のフレーム

図 12 表情認識に用いるパラメータ



図 13 構築したシステム

CamShift & Kalman Filter を併用した手法により検出し、検出されていることが怒りや酔いの条件となる。サングラスの日焼け跡や頬のマスク位置は追跡した目の座標を元に決定する。表 1 に示す認識の閾値や成分の増加量は初期値であり、システムの実行中に任意に変更可能である。

#### 4. 動作実験と有効性評価

構築したリアルタイム肌色変化による感情表現システムを図 13 に、その動作結果画像を図 14 に示す。本システムでは、3.3 節で述べたメラニン・ヘモグロビン判別に必要なメラニンベクトル群の平均値  $M_m$  として  $M_m = [0.1926, 0.4730, 0.8561]$ 、共分散行列  $C_m$  として

$$C_m = \begin{bmatrix} 0.0017 & 0.0011 & -0.0010 \\ 0.0011 & 0.0033 & -0.0021 \\ -0.0010 & -0.0021 & 0.0014 \end{bmatrix}$$

を用いてマハラノビス距離を計算した。表情に合わせて顔色がリアルに変化していることがわかる。頬やサングラスのような部分的な色変化も顔の回転に合わせて追従可能であり、顔部位追跡が正確に機能していることが確認できた。

また、本システムで用いた肌色成分解析による感情強調の有効性を検証するための主観評価実験を行った。顔色変化による感情強調に関する研究はあまり例がないため、本論文では基礎実験として、顔色変化の最も単純な手法である、顔領域の画素の RGB 値を変化させた結果との比較による主観評価実験を行った。評価対象として、一人の人物が怒りの表情を作った動画像に対して、本システムにより顔のヘモグロビン濃度分布  $\rho_h(x, y)$  を 0.2, 0.4, 0.6 だけ増加させたものを 3 つ(動画像 a ~ c)、顔の R の画素値を



図 14 システムの動作結果

20, 50, 70, 100 だけ増加させたものを 4 つ(動画像 d ~ g)、合計 7 種類の動画像を用意する。a ~ c と d ~ g から評価される動画像を 1 つずつ選んで被験者に同時に提示し、どちらが 1) 感情が強調されているか、2) 自然な顔に見えるか、の 2 つの基準で一対比較により判定した結果を得た。図 15 に用いた動画像の例を、表 2 に全ての組み合わせの評価結果を示す。この表では評価項目が比較対象項目よりも良いと判断された割合とその行毎の平均、そして a ~ c すべてと d ~ g すべての平均を示しており、被験者の数は 15 人である。表 2 の行平均と a ~ c および d ~ g



動画像 a      動画像 b      動画像 c



動画像 d      動画像 e      動画像 f      動画像 g

図 15 怒りの主観評価実験に用いた動画

表 2 怒りの主観評価実験の結果

感情が強調されていると判定された割合

(a ~ c: 本システムによる処理, d ~ g: R 値増加)

		比較対象項目							行平均	平均	
		a	b	c	d	e	f	g			
評価項目	a				0.6	0.4	0.4	0.533	0.483	0.683	
	b				0.933	0.933	0.733	0.8	0.850		
	c				0.867	0.6	0.667	0.733	0.717		
	d	0.4	0.067	0.133						0.2	0.317
	e	0.6	0.067	0.4						0.356	
	f	0.6	0.267	0.333						0.4	
	g	0.467	0.2	0.267						0.311	

自然な見えであると判定された割合

(a ~ c: 本システムによる処理, d ~ g: R 値増加)

		比較対象項目							行平均	平均	
		a	b	c	d	e	f	g			
評価項目	a				0.8	1.0	1.0	1.0	0.950	0.828	
	b				0.333	1.0	1.0	1.0	0.833		
	c				0.133	0.8	0.933	0.933	0.700		
	d	0.2	0.667	0.867						0.578	0.172
	e	0.0	0.0	0.2						0.067	
	f	0.0	0.0	0.067						0.023	
	g	0.0	0.0	0.067						0.023	

の平均より、感情強調と自然な見えの両面において、肌色成分解析を用いた結果が単純な R 値の処理結果よりも全体的に良いと判定されていることがわかる。しかし、動画像ごとの比較結果を見ると、ヘモグロビン濃度を少量だけ増加させた a よりも R 値を大きく増加させた e, f の方が感情強調効果は高く、逆に R 値をあまり増加させていない d の方がヘモグロビン濃度を大きく増加させた b, c よりも自然であると判定されている。これらの結果より、顔色変化の程度が評価に大きく影響しており、感情強調と自然さの両面において高評価を得られる成分増加量の決定

が重要であると考えられる。b, c は感情強調、自然さ共に 0.7 以上の高評価を得ており、今回の実験で使用した動画像の中では最も怒りの感情強調に適していると考えられる。

また、悲しみの感情についても、画素値の B 値のみの処理結果とメラニン・ヘモグロビン両成分を減少させた結果とで同様の比較実験を行った。ただし、悲しみを色素成分解析法によって強調する場合はその肌色変化が怒りの強調ほど急には表れないため、怒りの実験とは異なり色素成分解析で作成した動画を 4 種類、B 値の増加により作成した動画を 3 種類とした。A ~ D がメラニン・ヘモグロビン成分を 0.2, 0.4, 0.6, 0.8 だけ減少させたもの、E ~ G が B 値を 20, 50, 70 だけ増加させたものである。図 16 に使用した動画像の例、表 3 に評価結果を示す。感情強調の評価では、全体の平均では E ~ G が A ~ D よりも高い評価を受けているが、行平均では C, D が E ~ G のいずれよりも高い評価をされている。この結果より、悲しみの強調には色素成分解析法と B 値の増加による処理の間には目立った優劣はないものと考えられる。しかし自然さの評価においてはすべてにおいて A ~ D が高評価されている。これらの結果より、今回使用した動画の中では C, D が悲しみの感情強調に適しているといえる。

本システムは 2005 年、7 月 31~8 月 4 日に Los Angeles で開催された SIGGRAPH2005 Emerging Technologies, 2005 年 8 月 25, 26 日に日本科学未来館で開催されたインタラクティブ東京に出展し、様々な人物に対して有効に動作し、その有用性が確かめられた。

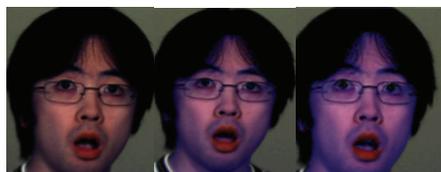
GPU による色素成分解析は 50fps, CPU - Intel pentium4 3.0GHz による顔追跡・表情認識処理は 30fps を実現し、最終的な処理速度は 27fps であった。これは十分にリアルタイム動画に適用できる速度である。

### 5. 結論と今後の課題

偏光板とビームスプリッタを用いた専用カメラ、GPU を用いた肌の色素成分解析、顔と顔部位追跡による表情認識を組み合わせることで、リアルタイム肌色変化による感情強調システムの構築を行った。ビデオチャット、テレビ電話における感情表現や、感情の誇張表現によるエンターテインメントへの応用などが考えられる。



動画像 A 動画像 B 動画像 C 動画像 D



動画像 E 動画像 F 動画像 G

図 16 悲しみの主観評価実験に用いた動画像の例

表 3 悲しみの主観評価実験の結果

感情が強調されていると判定された割合

(A ~ D: 本システムによる処理, E ~ G: B 値増加)

		比較対象項目							行平均	平均
		A	B	C	D	E	F	G		
評価項目	A					0.0	0.0	0.067	0.022	0.447
	B					0.4	0.267	0.4	0.356	
	C					0.667	0.5	0.733	0.633	
	D					0.667	0.733	0.933	0.777	
	E	1.0	0.6	0.333	0.333				0.567	0.553
	F	1.0	0.733	0.5	0.267				0.625	
	G	0.933	0.6	0.267	0.067				0.467	

自然な見えであると判定された割合

(A ~ D: 本システムによる処理, E ~ G: B 値増加)

		比較対象項目							行平均	平均
		A	B	C	D	E	F	G		
評価項目	A					0.867	1.0	1.0	0.956	0.905
	B					1.0	1.0	1.0	1.0	
	C					0.933	1.0	1.0	0.978	
	D					0.133	0.933	1.0	0.689	
	E	0.133	0.0	0.067	0.867				0.267	0.095
	F	0.0	0.0	0.0	0.067				0.017	
	G	0.0	0.0	0.0	0.0				0.0	

今後の課題としては、表情認識における安定した認識手法の構築、頭部追跡手法の精度向上、携帯電話へ応用するにあたっての専用カメラの縮小、などのシステムの改善が必要である。精度評価においては、本論文では単純な RGB 値増加処理との比較のみであったため、HSI 表色系による肌色変化などの様々な肌色変化手法との比較が必要である。さらに、喜びなどの今回は評価を行わなかった効果が存在するため、それらの評価実験もまた必要であり、今後、追実験を行っていく。また、肌色変化システムを用いた美容シミュレータ等への応用も可能であると考えられる。

参考文献

[1] “NOVA:英会話・フランス語・ドイツ語・スペイン語・イタリア語・中国語, [お茶の間留学(通信教育、在宅学習、オンライン講座)]”, NOVA ホームページ, <[http://www.nova.ne.jp/eki\\_ocha/ochanoma/index.html](http://www.nova.ne.jp/eki_ocha/ochanoma/index.html)>, (参照 2006-08-17)

[2] “テレビ電話 | サービス・機能 | NTT ドコモ”, NTT ドコモホームページ, <<http://www.nttdocomo.co.jp/service/videophone/index.html>>, (参照 2006-08-17)

[3] Mehrabian, Albert. “Silent Messages: Implicit Communication of Emotions and Attitudes” Wadsworth, Belmont, California.

[4] Argyle, Michael. “Bodily communication”, London: Methuen.

[5] “Top ページ | FSE により実現する「顔画像処理ソリューション」 | テクノロジーソリューション | 沖電気”, 沖電気工業ホームページ, <<http://www.oki.com/jp/FSC/vc/>>, (参照 2006-08-17)

[6] Benson P J, Perrett D I, “Perception and recognition of photographic quality facial caricatures: implications for the recognition of natural images” European Journal of Psychology 3 105-135, 1991b

[7] Calder, A.J., Young, A.W., Rowland, D. and Perrett, D.I. “Computer-enhanced emotion in facial expressions”. Proceedings of the Royal Society: Biological Sciences, B264, 919-925. (1997).

[8] 金箱淳一, 藤田ハミド, “色と音の感情を介したマッピング”, 日本ソフトウェア科学会第 22 回大会, 2005.

[9] S.A.Shafer, “Using color to separate reflection components”, COLOR Research and application, Vol.10, No.4, pp.210-218, 1985.

[10] Norimichi Tsumura, Nobutoshi Ojima, Kayoko Sato, Mitsuhiro Shiraiishi, Hideto Shimizu, Hirohide Nabeshima, Syuuichi Akazaki, Kimihiko Hori, Yoichi Miyake, “Image-based skin color and texture analysis/synthesis by extracting hemoglobin and melanin information in the skin,” acm Transactions on Graphics, Vol. 22, No. 3. pp. 770-779(2003)

[11] Yoichi Kuroda, Kazuyuki Morioka, Joo-Ho Lee1, Hiroyuki Yashima1, Hideki Hashimoto, “Modified Hybrid Tracking Algorithm Using Mean Shift and Kalman Filter”, 2004 RISP International Workshop on Nonlinear Circuit and Signal Processing (NCSP'04) Hawaii, USA, Mar. 5-7, 2004.

[12] 牧野貴雄, 高瀬紘一, 津村徳道, 中口俊哉, 三宅洋一, “顔の自動追尾を用いた肌計測とその応用”, 2006 年春季 第 53 回応用物理学関係連合講演会, 23p-E-17, 2006.

[13] A. Hyvarinen, J. Karhunen and E. Oja: “Independent Component Analysis” John Wiley and Sons, New York. 2001.

著者略歴



牧野 貴雄 (まきの たかお)

2004 年千葉大学工学部情報画像工学科卒業, 2006 年千葉大学大学院自然科学研究科知能情報工学専攻修了, 同年同大学院博士後期課程情報科学専攻進

学，現在に至る。肌の物理特性計測・再現，コンピュータビジョンに関する研究に従事。日本光学会，芸術科学会各学生会員。



中口 俊哉 (なかぐち としや)

2003 年，上智大学大学院博士後期課程修了。博士 (工学)。現在，千葉大学工学部情報画像工学科 助手。

2001-2003 年，日本学術振興会特別研究員。画像解析，画質評価，医用画像処理，組合せ最適化に関する研究に従事。IEEE，電子情報通信学会，IS&T，日本写真学会，各会員。



津村 徳道 (つむら のりみち)

1995 年，大阪大学大学院博士後期課程修了。博士 (工学)。現在，千葉大学工学部情報画像工学科 助教授。医用画像工学，計測画像工学，色再現工学，質感再現工学の研究に従事。IS&T，日本光学会，日本写真学会，日本医用画像光学会，電子情報通信学会，各会員

学会，日本医用画像光学会，電子情報通信学会，各会員



高瀬 紘一 (たかせ こういち)

2002 年千葉大学工学部情報画像工学科卒業，2004 年千葉大学大学院自然科学研究科知能情報工学専攻修了，同年，同大学院博士後期課程情報科学専攻進

学，現在に至る。物体の反射特性計測，GPU を用いたリアルタイム再現に関する研究に従事。応用物理学会，日本光学会各学生会員。2005 年日本光学会奨励賞，1999 年工業英語検定試験文部大臣奨励賞各受賞。



岡口 紗綾 (おかぐち さや)

2004 年千葉大学工学部情報画像工学科卒業，2006 年千葉大学大学院自然科学研究科知能情報工学専攻修了，同年

凸版印刷株式会社入社，現在に至る。肌の物理特性計測・再現に関する研究に従事。



小島 伸俊 (おじま のぶとし)

1988 年東北大学大学院修士課程修了，同年花王株式会社入社，2000 年同社スキンケア研美容センター，現在に至る。

1993 年千葉大学大学院博士後期課程修了。2001-2003 年同大学受託研究員，2003 年同大学自然科学研究科非常勤講師，2003 年より同大学フロンティアメディカル工学研究センター客員助教授。色彩科学，皮膚計測，化粧品科学に関する研究に従事。日本写真学会，日本視覚学会，情報処理学会員。



三宅 洋一 (みやけ よういち)

1968 年，千葉大学大学院修士課程修了。京都工芸繊維大学助手，助教授，千葉大学工学部助教授を経て，1989 年，

同教授。1978-1979 年スイス連邦工科大学研究員。1997 年，ロチェスター大学客員教授。2003 年よりフロンティアメディカル工学研究開発センター長。医用画像工学，画像評価，色再現工学の研究に従事。IS&T フェロー・名誉会員，日本光学会，電子情報通信学会，各会員。工学博士。