

動画像を用いた PC 顔認識システム”Smartface”

山口 修 福井 和広

(株) 東芝 研究開発センター

〒 212-8582 川崎市幸区小向東芝町 1

E-mail: {osamu1.yamaguchi, kazuhiko.fukui}@toshiba.co.jp

概要

ヒューマンインタフェースやセキュリティの分野で、顔画像を用いた個人識別技術が注目されている。本稿では、動画像を用いた PC ベースの顔画像認識システム”Smartface” について紹介する。顔認識エンジンには、環境変動にロバストな顔部品検出法と顔向きや表情変化といった変動に強い識別アルゴリズムを用いている。”Smartface” のアプリケーション機能は、(1) 個人認証によるスクリーンセーバ、(2) 個人識別による環境設定、(3) リアルタイム変装シミュレーション機能からなる。本システムは、カメラ付きミニノート型パソコン上で動作し、特殊な画像処理ハードウェアを用いることなくソフトウェア単体で動作する。

“Smartface” – PC Based Face Recognition System using Temporal Image Sequence

Osamu Yamaguchi and Kazuhiro Fukui

TOSHIBA Corporate Research and Development Center

1, Komukai-toshiba-cho, Saiwai-ku, Kawasaki 212-8582 Japan

E-mail: {osamu1.yamaguchi, kazuhiko.fukui}@toshiba.co.jp

Abstract

Recently, face recognition provides an important means for man-machine interface and security. This paper presents “Smartface”, a PC based face recognition system using a temporal image sequence. The face recognition engine of the system employs a robust facial parts detection method and a pattern recognition algorithm which is stable against variations of facial expression and face direction. The functions of Smartface include (i) screen-lock, (ii) customization of PC environment, and (iii) real-time disguising as an application for entertainment. The system works on a portable PC with a CMOS/CCD camera, only with software without requiring a special image processing hardware.

1 はじめに

近年、ヒューマンインタフェース、セキュリティなどといった分野で顔画像認識を用いた個人認証技術が重要な要素技術として認知されている。顔画像認識は、非接触で、利用者に意識させずに個人の識別が可能といったことが利点として挙げられる。また、従来PC上でリアルタイムに実行することが困難であった動画処理が最近のCPU処理能力の向上により実時間で処理することが可能となってきた。本稿ではPC上で動作する顔認識システムについて述べる。

PCにおける顔画像認識の応用を考えた場合、セキュリティの側面とヒューマンインタフェースの側面とがある。簡易なセキュリティ技術としては、図1に示すように、システムの利用者が作業を中断してシステムから少し離れるような状況を想定した場合、それまでシステムを利用していた人が誰であるかを瞬時に判定することができれば、スクリーンロックを解除するといった利用法が可能である。一方、ヒューマンインタフェースの観点からは、システムの利用者を識別することで、ユーザの利用環境を設定したり、そのユーザに応じた情報提供等のサービスが可能となる。さらには、画像中の顔を検出するときの情報を用いて、顔の位置、向きなどを用いたインターフェースアプリケーションを構築することが可能となる。

これらの機能をポータブルノート型PC上に実現した顔認識システム”Smartface”について説明する。以下、顔認識を行うためのアルゴリズム、システム構成と各アプリケーション機能について述べる。

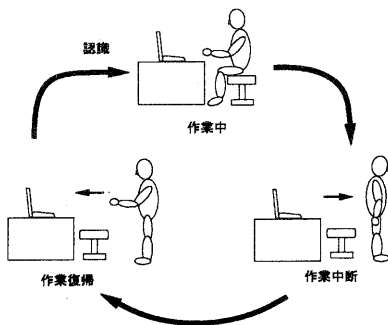


図1: 簡易なセキュリティの例

2 顔認識アルゴリズム

顔認識処理の流れを図2に示す。処理は(1)入力された画像中の顔領域の検出、(2)顔部品の特徴点(瞳、鼻孔)の位置検出、(3)位置、大きさなどを正規化したパターンの切りだし、(4)入力部分空間の計算、(5)登録辞書との類似度計算による識別処理、を順に行なってシステムの前の人が誰であるかを識別する。

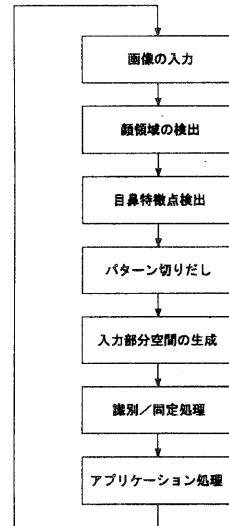


図2: 処理の流れ図

2.1 顔領域の検出

顔領域の検出は、部分空間法[4]を用いたパターンマッチングを行う。予めさまざまな人物、方向の顔パターンデータ(15×15 pixelの画像)を収集し、顔検出用の辞書を作成しておく。そのテンプレートを画像全体にラスタースキャンし、各場所での類似度を逐次求め、類似度の局所最大値が、しきい値以上の箇所を顔領域として検出する。顔領域の大きさの変化を考慮して、入力画像を数段階にスケール変化させて、パターンマッチングを行い、画像中に一つの顔領域を決定する。画像中にしきい値以上の箇所が無い場合は、人が存在しないとして、以後の認識処理を行わない。

また、[6]のように、複雑な背景の中から顔を見つけるために、顔の辞書との類似度の高い非顔データを収集しておき、それらから構成された非顔辞書

との類似度を同時に求めることによって、顔検出の誤認識を軽減している。

2.2 顔部品特徴点抽出とパターンの切り出し

検出された顔領域に対して、瞳、鼻孔の特徴点の抽出を行なう。本システムでは、形状情報とパターン情報を組み合わせることによって、高速かつ、位置ずれの少ない特徴点抽出法 [5] を用いる。まず、形状情報を用いた特徴点抽出には、明度変化、ノイズにロバストである領域ベースのエッジ抽出法 [7] を拡張した「分離度フィルタ」による抽出法を用いる。

分離度 $S(0.0 < S \leq 1.0)$ は次の式で与えられる [7]。

$$S = \frac{n_1(\bar{P}_1 - \bar{P}_m)^2 + n_2(\bar{P}_2 - \bar{P}_m)^2}{\sum_{i=0}^{N-1} (P_i - \bar{P}_m)^2}$$

ここで、 N は領域内の全画素数、 n_1, n_2 は領域 1, 領域 2 の画素数、 P_k は位置 k における輝度、 \bar{P}_m は領域全体での輝度の平均値、 \bar{P}_1, \bar{P}_2 は領域 1, 領域 2 での輝度の平均値を示す。

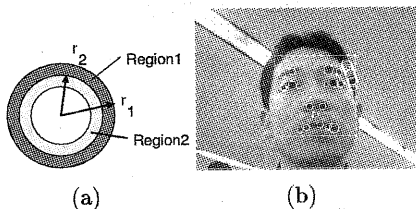


図 3: 円形分離度マスク (a) と分離度の局所最大点 (b)

図 3(a) に示すような円形のマスクを設定し、そのマスク内で 2 つの円形領域の分離度を上述した式によって計算する。位置、半径を変えながら、円形の領域を特徴点候補として算出する。これを分離度フィルタと呼ぶ。先に求めた顔領域中で、分離度の局所最大点を候補として抽出した結果を図 3(b) に示す。

これらの目鼻の特徴点候補を求めたのち、さらにパターン照合による検証を行って、正確な目鼻の位置を特定する。パターン照合には部分空間法を用い、瞳、鼻孔の画像を予め収集して辞書を作っておく。そして、各特徴点において、目鼻辞書との類似度を求め、幾何学的配置の制約を満たす 4 点を特徴

点として検出する。特徴点抽出についての詳細は [5] に示す。

次に抽出された目鼻の特徴点を基準にして顔領域を位置、サイズを正規化し、矩形の濃淡パターンとして切り出す。図 4 は、実際に連続して得られた 20 枚の矩形の濃淡パターンである。人間の手作業なしで自動的に大量のパターンが生成することが可能である。



図 4: 正規化パターン取得例

2.3 相互部分空間法による個人識別

本システムでは、表情変化や顔の向きといった変動に対応するため、動画像から得られる複数枚の画像を用いた認識法 [1][2] を採用している。

図 5 は、静止画像による認識と動画像による認識のアプローチの説明図である。この図では特徴空間を模式的に表し、それぞれの人物のサンプルデータの分布を網目で表している。

図 5(a) は、従来の静止画による認識法を表し、単一の未知入力 (図中×印) がいずれのサンプル集合に近いかによって識別することを表している。この例ではどちらの人物に識別すべきかを決定することが困難である。また、実環境で取得される入力顔データは、顔の向きや表情の変化によってデータが変動するため、図 5(b) のように、さまざまな点として観測されることになる。そこで、時系列で連続して得られる複数の顔データは同一のカテゴリ (同一人物) に属しているという自然な仮定により、多様な複数の入力データを利用して、図 5(b) に示すように、動画像中の入力顔データの分布 (図中×印の集合) が、どの人物の分布に近いかを求めることにより識別を行えば、変動に強い認識が可能になる。

複数の画像データを用いたパターン認識法として相互部分空間法 [2][3] を用いる。この認識法では、入力ベクトルも部分空間で表現し、辞書パターンの部分空間との間の角度を類似度として識別を行なう。図 6 は相互部分空間法の説明図であり、2 つの部分空間 D, G のなす角度 θ の余弦は、

$$\cos^2 \theta = \sup_{d \in D, g \in G, \|d\| \neq 0, \|g\| \neq 0} \frac{|(d, g)|^2}{\|d\|^2 \|g\|^2} \quad (1)$$

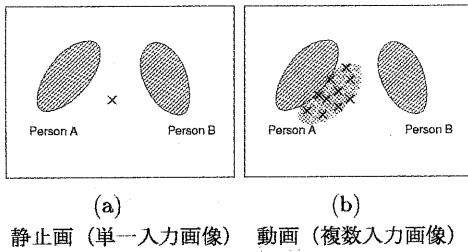


図 5: 静止画と動画による認識の概念図

と定義する (d, g は式 (1) が極値をもつためのそれぞれの部分空間上のベクトルを表す)。これに関して、2つの部分空間 D, G への正射影作用素を P, Q とする場合、 $\cos^2\theta$ は PQP の最大固有値 λ_{max} となる [3]。

$$\cos^2\theta = \lambda_{max} \quad (2)$$

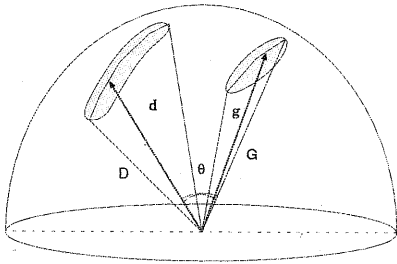


図 6: 相互部分空間法の説明図

辞書パターン部分空間を D 、入力された時系列画像に対する部分空間を G とする。ここで ϕ, ψ を各部分空間 D, G における固有ベクトルとする。実際には PQP という行列の最大固有値を求めのではなく、式 (3) で表される行列 X の固有値問題を解き、その最大固有値を類似度 (部分空間間類似度) S_{mutual} とすればよい [3]。ここで、 D の部分空間の次元を M 、 G の部分空間の次元を L とし、 $L \leq M (1 \leq i, j \leq L)$ とする。

$$X = (x_{ij}) = \sum_{m=1}^M (\psi_i, \phi_m)(\phi_m, \psi_j) \quad (3)$$

$$W^T X W = \Lambda \quad (4)$$

($W: X$ の対角化行列、 $\lambda_{max} : \Lambda$ の対角成分の最大値)

$$S_{mutual}(G, D) = \lambda_{max} \quad (5)$$

本手法の有効性を調べるために、101 名分、36360 枚の実験画像を収集し、識別、同定実験を行った。顔向きや表情変化のある画像を用いた同定実験の結果、従来法の部分空間法による認識では誤識別率が 8% あるのに対し、本手法では 2% にまで減少できた [2]。

3 システム構成とアプリケーション

Smartface は Windows98 上で動作するアプリケーションで常駐ソフトとして起動するように実装されている。これにより、バックグラウンドジョブで顔認識を行い、意識しないで自動的にアプリケーションの起動などが行えるようになっている。



図 7: キャプチャビューアウィンドウでの認識の様子

通常は、起動時に図 7 に示すキャプチャビューアウィンドウが開き、カメラから入力される画像は、顔検出、目検出の結果とともにそのウィンドウに表示される。また、認識のモードや状況を知らせるアイコンと、認識の個人類似度のグラフが表示され、システムの状態がユーザにフィードバックされる。さらに、後述する変装シミュレーションもキャプチャビューアウィンドウで実行される。

識別処理は画像入力一枚毎に図 2 の処理がなされ、入力部分空間を計算し、各個人辞書との類似度の計算を逐次行っている。キャプチャビューアウィンドウには、登録されている人物すべての類似度の中で最大の類似度が表示されており、登録されている人物の類似度が、設定されているしきい値を越えた場合に、該当人物がカメラの前にいると判定する。

図8は、ソフトウェアの構造を示したもので、前章で述べたそれぞれの顔検出、目鼻検出のルーチンと個人識別のルーチンは、それぞれ独立のDLLとして構成されており、DirectShowのフィルタから利用されている。そのフィルタを上位のSmartfaceアプリケーションが呼び出す形でソフトウェアが構成されている。

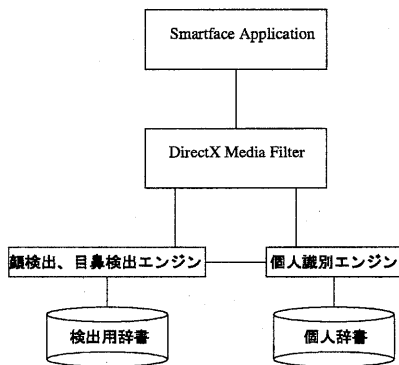


図 8: ソフトウェアの構成図

Smartfaceのアプリケーション機能は、以下の4つからなり、それぞれの機能について説明する。

- 登録ユーティリティ
- 個人識別によるユーザ別サウンド、アプリケーション起動機能
- 本人照合によるスクリーンセーバ解除機能
- リアルタイム変装シミュレーション機能

3.1 登録ユーティリティ

個人の顔の登録は、ユーティリティウインドウで行う。顔の登録は、個人の名前を入力し、その後、50枚の画像の取得を行う。登録に要する時間は、画像の取得、顔領域の正規化、辞書計算を行い、完了まで平均20秒程度である。

このとき、ユーザが様々な顔の向き、表情変化をさせながら、バラエティのある顔画像を撮影させることで、後に精度よい認識が行えるようになる。カメラのフレーム外に顔領域が入っていない場合は、音声合成で「カメラに顔を向けてください」とガイドされる。登録人数は100名までとなっており、100名登録時でも2~3秒程度で識別が可能である。

また、登録後の人物の経時変化や変動を考慮する必要がある。各個人辞書に新しいデータを追加し、辞書を更新することで認識率を維持できることを実験的に確認している[8]。Smartfaceでは、ユーティリティウインドウで、更新機能を選択し、登録作業と同様に複数枚の画像を収集して、現在の登録辞書にそのデータを追加することで、認識率の低下を抑制できるようになっている。

3.2 ユーザ別設定機能

各個人辞書との照合は毎回行われ、その時点での各人物との類似度が計算されている。該当する人物それぞれに対して、図9のプロパティで設定されている機能を実行する。ユーザ別設定機能では、サウンドと起動するアプリケーションが設定される。デフォルトの設定では、「こんにちは、鈴木さん」といったテキストを音声合成で読み上げる。設定を変えることにより、任意のWAVファイルを選択し、認識時に再生することができる。また、アプリケーションを選択しておくことで、その人を認識した時点で、そのアプリケーションを起動する。例えば、メールの到着をチェックするアプリケーションを選択しておけば、その人がPCの前になると、メールの有無を知らせる、といった利用法があげられる。

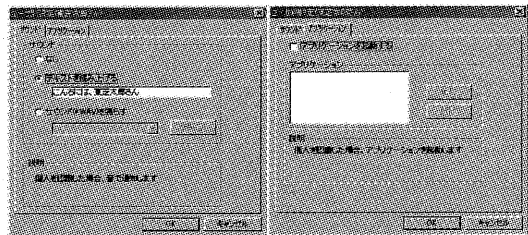


図 9: ユーザ別の設定画面

3.3 顔認識つきスクリーンセーバ

従来のスクリーンセーバでは、パスワードが設定されている場合、そのパスワードを入力することで、スクリーンロックが解除される。Smartfaceスクリーンセーバは、様々なタイミングでセーバの解除を指定することができる。

指定可能なタイミングとは、(1) 画像中に顔を検出すると解除、(2) データベースに登録されている人の顔を認識すると解除、(3) 特定の人物の顔を認

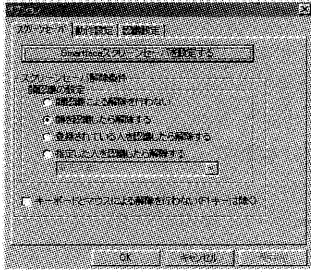


図 10: スクリーンセーブの設定画面

識すると解除、であり、用途やセキュリティの度合いに応じて設定する。

3.4 変装シミュレーション (CG 重ね合わせ)

Smartface では、個人の識別を行うだけでなく、顔領域の位置、大きさ、目鼻の顔部品の位置情報を用いた、変装シミュレーション機能を有する。変装シミュレーション機能は、検出された顔の位置、大きさに応じて、CG 画像などの重ね合わせや画像の部分的な画像処理、変形などを行うことができる。図 11 は、その様子を表しており、ビューア右下の上下ボタンを押すことにより、「ピエロ」「鼻眼鏡」「アフロヘア」など様々なコンテンツを選択することができる。

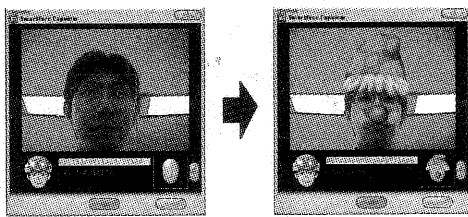


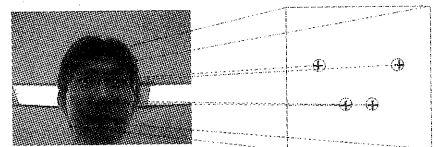
図 11: 変装シミュレーションの様子

顔の移動に応じて、任意の位置、大きさに画像の重ね合わせや変形を行うために以下の処理を行う。重ね合わせようとするビットマップを配置するための矩形の頂点の相対的な位置関係を、画像認識で検出可能な特徴点から計算される「基準点ベクトル」と検出された特徴点間を結ぶ「相対位置ベクトル」を用いて記述する。

図 12 の上に示すように、特徴点として、顔領域、目鼻位置が検出される。検出された特徴点から、図中段に示したいいくつかの「基準点ベクトル」と「相対位置ベクトル」を決定する。一例としてビットマップを配置する矩形の頂点の 1 点 P_1 の表現法を図 12 の下段に示す。相対的な位置を指定するための基準点として、左目の検出位置を選ぶ場合、その左目の位置を記述したベクトル B_1 とすると、その基準点からの相対位置と大きさを表現するために、検出した顔領域の x, y 軸方向の 2 つのベクトル F_w (FACE.WIDTH)、 F_h (FACE.HEIGHT) を用いて表現すると、

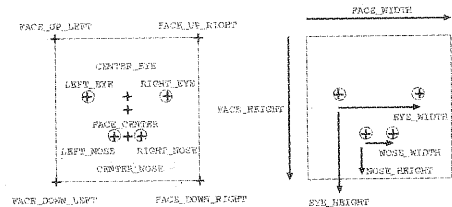
$$P_1 = B_1 + (t_1 F_w + t_2 F_h)$$

と記述できる (t_1, t_2 は実数値)。このような点を複数個指定することにより、CG を配置する場所が決定され、顔の大きさ、位置が変化しても、相対的なビットマップの位置、大きさの関係が保たれる。



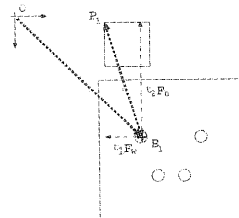
検出結果

検出特徴点



基準点位置

相対位置ベクトル



位置の指定方法

図 12: 相対的位置の指定法

通常、ビットマップの重ねあわせは、複数のビットマップを扱ったり、自由にコンテンツを作成できることが望ましい。そこで、自由にコンテンツを変換、操作できるように、変装のための情報をスクリプト言語で記述し、そのスクリプトを読み込んで、CGの重ねあわせを行うシステムとなっている。

```

#
# 「ピエロ」の変装スクリプト
#
FG_EMBEDBMP 2
LEFT_EYE ADD FACE_WIDTH -1.0 ADD FACE_HEIGHT -2.8
RIGHT_EYE ADD FACE_WIDTH 1.0 ADD FACE_HEIGHT 0
COLORBMP "data/pierol.bmp" 255 255 255

FG_EMBEDBMP 2
CENTER_NOSE ADD FACE_WIDTH -0.5 ADD FACE_HEIGHT -0.5
CENTER_NOSE ADD FACE_WIDTH 0.5 ADD FACE_HEIGHT 0.5
COLORBMP "data/piero2.bmp" 255 255 255

FG_EMBEDBMP 2
RIGHT_EYE ADD FACE_WIDTH -0.1 ADD FACE_HEIGHT 0.4
RIGHT_EYE ADD FACE_WIDTH 0.5 ADD FACE_HEIGHT 1.2
COLORBMP "data/piero3.bmp" 255 255 255

FG_EMBEDBMP 2
LEFT_EYE ADD FACE_WIDTH -0.5 ADD FACE_HEIGHT 0.4
LEFT_EYE ADD FACE_WIDTH 0.1 ADD FACE_HEIGHT 1.2
COLORBMP "data/piero4.bmp" 255 255 255

```

図 13: 変装スクリプトの例

図 13は図 11で示した「ピエロ」の変装のスクリプト例で、4つのビットマップを右目、左目、鼻孔中心を基準点として、それぞれの位置と利用するビットマップを指定している。

また、スクリプトを文書エディタで編集するのではなく、合成CGユーティリティが用意されており、グラフィカルユーザーインターフェースで対話的に合成内容を作成することができるようになっていいる。図 14は、合成用のビットマップの位置を編集している様子を表している。

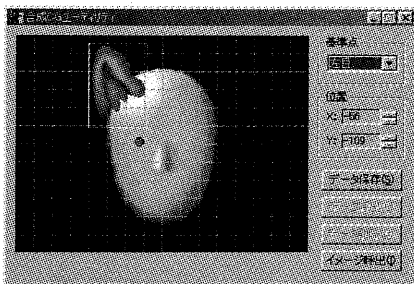


図 14: 合成CGユーティリティ

図 15は、東芝製 CMOS カメラ付きミニノート型パソコン Libretto ff 1100(CPU:MMX Pentium 266MHz)を用いて、変装コンテンツを様々な人物に適用した画像例であり、愉快的画像生成が可能となる。画像認識技術をエンターテイメントに応用したアプリケーション例の一つであり、今後も顔認識技術をヒューマンインタフェースへの適用を広げていく。



図 15: 様々なシミュレーションの様子

4 おわりに

ポータブルPC上で動作する顔認識システム“Smartface”を紹介した。今後は、顔検出、顔部品検出のアルゴリズム強化ならびに、認識エンジン部分の強化策として [9] を適用した照明変動にロバストなアルゴリズムを搭載する。また、顔認識技術を利用したアプリケーションの様々な展開を考えていく予定である。

参考文献

- [1] O. Yamaguchi, K. Fukui, K. Maeda: “Face Recognition using Temporal Image Sequence”, Proceedings of the third International Conference on Automatic Face and Gesture Recognition, pp.318-323, April 1998.
- [2] 山口 修, 福井 和広, 前田 賢一: “動画像を用いた顔認識システム”, 信学技報, PRMU97-50, pp.17-24, 1997.
- [3] 前田 賢一, 渡辺 貞一: “局所的構造を導入したパターン・マッチング法”, 信学論 (D), vol.J68-

- D, No.3, pp345-352, 1985.
- [4] E. Oja: "Subspace Methods of Pattern Recognition", Research Studies Press, England, 1983.
 - [5] 福井 和広, 山口 修: "形状情報とパターン照合の組合せによる顔特徴点抽出", 信学論 (D-II), vol.J80-D-II, No.8, pp.2170-2177, August 1997.
 - [6] K.Sung and T.Poggio: "Example-based Learning for View-based Human Face Detection", Technical Report A.I.Memo No.1521, A.I. Lab., MIT, 1994.
 - [7] K.Fukui: "Edge Extraction Method Based on Separability of Image Features", IEICE Trans. Inf.& Syst., Vol E78-D, No.12, pp1533-1538, 1995.
 - [8] 山口 修, 福井 和広: "経年変化を抑制する顔辞書更新法", 信学技報, PRMU99-25, pp.43-50, 1999.
 - [9] 福井 和広, 山口 修, 鈴木 薫, 前田 賢一: "制約相互部分空間法を用いた環境変化にロバストな顔画像認識 - 照明変動を抑える制約部分空間の学習", 信学論 (D-II), vol.J82-D-II, No.4, pp.613-620, April 1999.