

## 3次元データをキーとした姿勢・照明変動に 頑健な顔画像検索

濱中 雅彦<sup>†</sup> José Ferreiro<sup>‡</sup> 石山 墨<sup>†</sup> 坂本 静生<sup>†</sup>

<sup>†</sup>NEC メディア情報研究所

<sup>‡</sup>スイス連邦工科大学 ローザンヌ校

あらまし 本論文では、顔の3次元データを入力として、姿勢・照明変動のある顔画像に対して高精度な検索が可能な3次元(入力)対2次元(登録)の顔照合方式を提案する。本照合方式には、先に提案している測地照明基底法による2次元(入力)対3次元(登録)の顔照合におけるマッチング手法を利用することができる。しかし、画像に依存して照合スコアが大きく変動するため、そのまま適用しても十分な精度は得られない問題がある。本論文では、登録画像毎にスコア正規化することにより、Verification性能だけでなくIdentification性能も改善され、3次元対2次元の顔照合が実現できることを示す。様々な姿勢・照明変動した画像を含む1000名規模の顔データベースによる評価実験により、1位照合率90%以上、EER約2%を得た。

## Face Image Retrieval using 3D Data with Robustness against Pose and Illumination Variations

Masahiko Hamanaka<sup>†</sup> José Ferreiro<sup>‡</sup> Rui Ishiyama<sup>†</sup> Shizuo Sakamoto<sup>†</sup>

<sup>†</sup>Media and Information Research Laboratories, NEC Corporation

<sup>‡</sup>Swiss Federal Institute of Technology (EPFL)

**Abstract** This report proposes a new face recognition scheme called 3D(query)-2D(enroll) recognition. It can retrieve facial images from an inputted 3D data with robustness against pose and illumination variations. When using the previously proposed Geodesic Illumination Basis method to match a 2D image and a 3D data for 2D(query)-3D(enroll) recognition, it is found that its scores are drastically changed dependent on images. This report shows the scores become stable by introducing score normalization, and not only verification rates but also identification rates can be improved. In the experiments using facial image database of 1000 individuals, the proposed method achieved a first-choice success ratio of over 90% with an EER of about 2%.

## 1. はじめに

顔照合は、非接触で使い勝手の良いバイオメトリックスによる個人認証として期待され、入国審査や入室管理、コンピュータへのアクセス管理などで使われ始めている[1]。また、個人認証や防犯カメラ映像からの人物照合と言ったセキュリティ応用だけでなく、ビデオストリームや画像データベースからの人物検索など、広い応用が考えられる。

これまで様々な顔照合方式が研究されてきているが、顔画像を登録し入力画像と照合する2次元対2次元の照合方式(2D-2D照合)が主流である[2][3]。これは、2次元画像は安価なカメラで容易に取得できるメリットがあるためである。しかし、認証システムのように画像の撮影条件が限定される場合には一定の性能を発揮できるが、防犯カメラにより撮影された画像のように、姿勢・照明変動の大きい場合には、照合精度が大きく低下する問題がある。一方、3次元データを用いた照合方式の研究も進められている[3]。3次元データを用いれば、姿勢や照明変動に影響されない照合が可能となる。しかし、高速・高精度なレンジファインダの開発[4]や、画像からの3次元復元の研究[5]などが近年進みつつあるものの、3次元データの取得はカメラによる画像取得に比べると容易ではない問題がある。従って、例えば入力と登録の両方に3次元データを用いる3次元対3次元の照合方式(3D-3D照合)が適用できる場面は、まだ限定されると考える。

しかし、入力が登録のいずれかのみが3次元データである場合は、実用の可能性が高いと考えられる。例えば、登録時は比較的時間をかけても3次元データを登録し、照合時はカメラで簡単に撮影する2次元(入力)対3次元(登録)の照合方式(2D-3D照合)は現実的であり、防犯システムなどへの適用が期待される。筆者らは、顔の3次元データから姿勢と照明条件を入力画像に合わせた補正画像を生成し、入力画像と比較する2D-3D照合方式を研究してきた(図1)[6][7]。これまで、測地照明基底法による補正画像の生成法を提案し、現実的な速度での照合を実現した。そして、様々な姿勢・照明変動のある画像を用いた実験でも、1000名分の3次元データに対する1位照合率が96.5%という精度を得た[9]。

本論文では、更に3次元データを入力キーとして登録画像と照合する3次元(入力)対2次元(登

録)の照合方式(3D-2D照合)を提案する(図2)。世の中には様々な画像が既に存在しており、特定の人物の3次元データが取得できれば、画像に姿勢・照明変動があってもその人物の画像をデータベースから検索することができる。例えば、防犯カメラで撮影された映像(画像)をシステムのデータベースに保存しておけば、後で3次元データを取得した人を過去の映像から検索することができる。また、インターネット上には無数の画像が存在するが、その中から特定の人物の画像を検索することもできる(例えば肖像権に触れる芸能人の画像を探すなど)。また、近年のデジタル機器の普及により個人でパソコンに多数の画像やビデオを持つことも増えており、そのようなパーソナルデータベースから特定の人物の画像を検索するような利用法も考えられる。

3D-2D照合方式を実現するためには、既に開発した2D-3D照合におけるマッチング手法を利用することができる。しかし、得られる照合スコアは画像に依存して大きく変動する問題があるため、そのまま適用しても十分な精度は期待できない。本論文では、従来の2D-3D照合のマッチング手法を適用した場合の問題点を分析し、スコア正規化法を導入することにより、十分な3D-2D照合性能が得られることを示す。

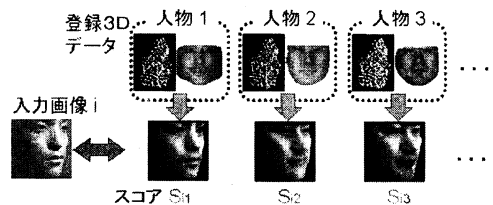


図1：2D-3D照合方式(2D画像が入力)

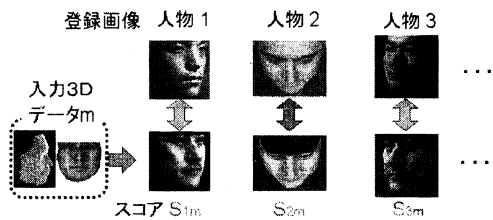


図2：3D-2D照合方式(3Dデータが入力)

## 2. 従来の2D-3D 照合におけるマッチング

### 2.1. 測地照明基底法による照合

3次元データは、顔表面上の点の座標値  $(x, y, z)$  と反射率 (テクスチャを利用) を情報として持っている。この顔表面の3次元データは、測地座標系において表現できる。測地座標系として、円筒座標系  $(y, \theta)$  を用いた場合の例を図3に示す。

測地照明基底法は、姿勢とは独立な測地座標系において照明基底 (測地照明基底) を予め生成しておき、姿勢に応じて照明基底を変換して照明補正画像を生成する手法である。測地照明基底法による顔照合の流れを図4に示す (ここでは画像を入力、3次元データを登録とする)。登録時に、3次元データから測地照明基底を求めておく。まず、(a) 3次元データから様々な照明条件での測地画像をCG生成する。(b) 生成した  $n_L$  個の測地画像に対して主成分分析を適用し、測地照明基底  $\{G_j\}$  を計算する。そして、必要十分な数  $n_G$  個の測地照明基底を固有値の大きい順に選択し、3次元データと共に保持しておく。

照合時には、姿勢探索しながら、各姿勢に対して補正画像を生成し、入力画像と照合することにより、最適な姿勢を決める。(c) 姿勢  $r$  における姿勢依存な照明基底  $\{B_j^{(r)}\}$  を下式により計算する。

$$B_j^{(r)} = \Gamma^{(r)} G_j \quad (1)$$

ここで行列  $\Gamma^{(r)}$  は、測地画像上の画素と補正画像上の画素の対応を表し、3次元データからのCGによる画像生成処理により求められる。(d) 照明基底  $\{B_j^{(r)}\}$  は直交基底になっていないため、次式のように、最小二乗法により係数ベクトル  $\lambda^{(r)}$  を求め、照明条件の補正された補正画像  $I_c^{(r)}$  を生成する。

$$I_c^{(r)} = \sum_j \lambda_j^{(r)} B_j^{(r)} \quad (2)$$

$$\lambda_j^{(r)} = \arg \left( \min \|I_q - I_c^{(r)}\| \right)$$

ここで、 $I_q$  は入力画像である。(e) 入力画像と補正画像の照合では、補正画像中の顔領域の画素数  $m^{(r)}$  は3次元データや姿勢により変動するため、照合用の距離は画素数で正規化する。最も簡単な場合は、次のような画素の輝度値の平均絶対誤差を用いる。

$$d_1^{(r)} = \frac{1}{m^{(r)}} \sum_i^{m^{(r)}} |I_q(i) - I_c^{(r)}(i)| \quad (3)$$

先に提案している標準測地重み[8]や鮮明度[9]を加味した照合では、次のような照合関数を用いる。

$$d_2^{(r)} = \frac{1}{b^{(r)} m^{(r)}} \sum_i^{m^{(r)}} w^{(r)}(i) |I_q(i) - I_c^{(r)}(i)| \quad (4)$$

$$b^{(r)} = 1 + a S^{(r)}$$

ここで、 $w^{(r)}$  は標準測地重みから姿勢に依存して生成された2次元重み、 $S^{(r)}$  は鮮明度、 $a$  は鮮明度の比重を調節するパラメータである。

最終的な入力画像と3次元データの照合スコアは、距離  $d^{(r)}$  の最も小さくなる最適姿勢における距離  $d^*$  を用いて  $s = (255 - d^*) / 255$  により求める。

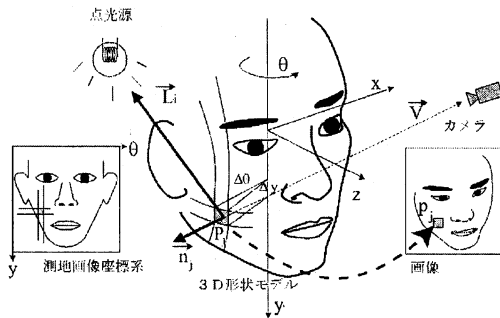


図3：3次元データと測地座標系の関係



図4：測地照明基底法による顔照合の流れ

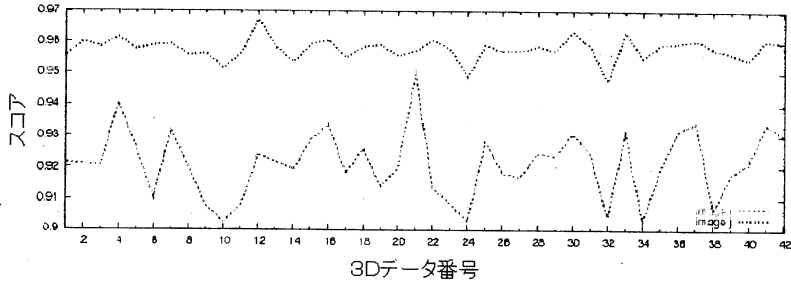


図5：画像毎のスコア分布の例（2画像対42データ）

## 2.2. スコア分布の分析

図5に、2つの画像を各々42個の3次元データとマッチングした時のスコア分布の例を示す。一方の画像ではスコアが0.95～0.97の間に分布し、もう一方の画像ではスコアが0.90～0.95に分布しており、画像に依存してスコア分布が大きく変動している。これは、補正画像を入力画像の姿勢・照明条件に合わせて生成し、その姿勢・照明条件において照合を行っているためである。特に全体的に輝度値が小さい暗い画像では、平均誤差が小さくなるため照合スコアは高くなる傾向がある。

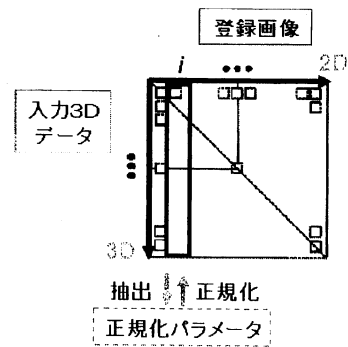
2D-3D照合では、図1に示すように、入力画像に合わせて生成された各補正画像の姿勢・照明条件はほぼ同じである。そのため、入力画像と各補正画像との照合スコアは比較しうるものとなる。しかし3D-2D照合では、図2に示すように、登録画像に合わせて生成された各補正画像の姿勢・照明条件は異なる。そのため、各登録画像と補正画像との照合スコアは登録画像毎に大きく変動し、対等な比較ができなくなる。そこで、各照合スコアを比較して照合するためには、画像の姿勢や照明条件などに依存した照合スコアの変動を吸収する必要がある。スコア分布は各画像内では安定しているの、画像毎にスコアを正規化すれば比較しうるスコアになると考える。本論文では、登録画像毎にスコア正規化を行い、3D-2D照合を実現する。

## 3. スコア正規化を用いた3D-2D照合

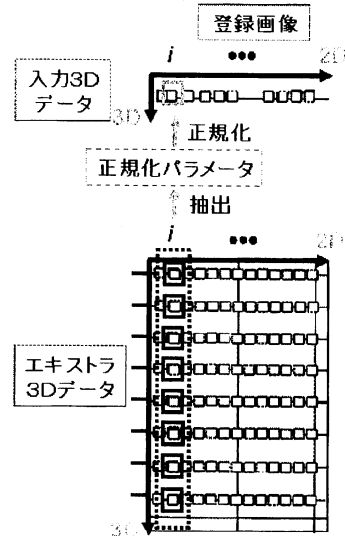
### 3.1. スコア正規化の枠組み

スコア正規化の研究はこれまで話者照合の分野で活発に研究されてきた[10][11]。スコア正規化は、主にVerificationにおける頑健なしきい値設定を目的としている。また、顔照合の分野では文献[12]において、2D-2D照合で入力画像毎にスコア

正規化した場合の性能改善が報告されている。しかし、入力画像毎のスコア正規化のため、Verification性能が改善されることは示されているが、Identification性能の改善には至っていない。

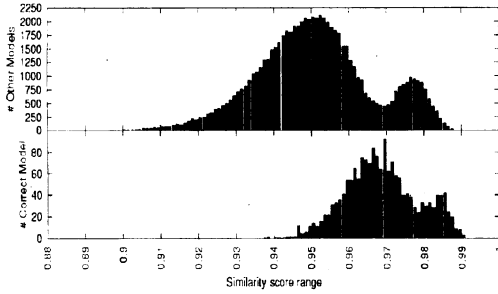


(a) Ideal case



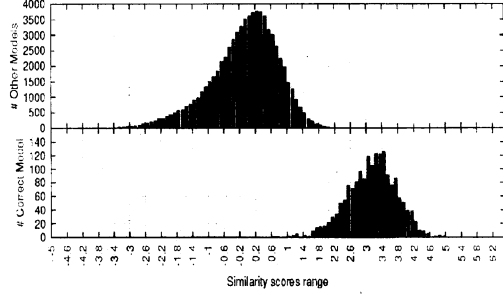
(b) Real case

図6：スコア正規化の枠組み



(a) スコア正規化前

図7：スコア分布の例 (DB42) (上が不正解、下が正解スコア)



(b) スコア正規化後

スコア正規化は、図6に示すように、複数の入力による照合を行った後の一括後処理として実行できる場合 (ここでは Ideal case と呼ぶ) と、一つの入力毎に実行する場合 (Real case) がある。実用的には、後者のように一つの入力毎に結果が得られることが望まれることが多い。3D-2D 照合の Ideal case では、複数の入力 3次元データと1つの登録画像との照合スコアを用いて正規化パラメータを求め、自らの照合スコアを正規化する。Real case では、入力 3次元データが1つしかないので、その情報だけでは登録画像毎にスコア正規化することができない。そこで、例えば入力とは別に予め用意した 3次元データ (ここではエキストラ 3次元データと呼ぶ) を用いる必要がある。複数のエキストラ 3次元データと1つの登録画像との照合スコアを用いて正規化パラメータを求め、入力 3次元データと登録画像との照合スコアを正規化する。この正規化パラメータは、登録画像の登録時に予め計算しておくことができるため、入力 3次元データによる照合処理時間の増加は殆どない。3D-2D 照合では登録画像毎にスコア正規化するので、Identification 性能の改善が期待できる。

### 3.2. Z-norm による正規化

本論文では、スコア正規化に最も基本的である Z-norm を用いる。Z-norm による画像毎のスコア正規化は、下式により行われる。画像  $i$  毎にスコアの平均と分散を求めて平均 0 ・分散 1 の標準正規分布に正規化する。

$$s'_{i,m} = \frac{s_{i,m} - \mu_i}{\sigma_i} \quad (5)$$

ここで

$$\mu_i = \frac{1}{M} \sum_{n=1}^M \tilde{s}_{i,n}, \quad \sigma_i^2 = \frac{1}{M} \sum_{n=1}^M (\tilde{s}_{i,n} - \mu_i)^2 \quad (6)$$

$s_{i,m}$  は画像  $i$  と 3次元データ  $m$  との照合スコアである。また、 $\tilde{s}_{i,n}$  は正規化パラメータを求めるために使用されるスコアである。Ideal case では、入力 3次元データと登録画像との照合スコア  $s_{i,m}$  が  $\tilde{s}_{i,n}$  として使用される。Real case では、エキストラ 3次元データと登録画像との照合スコアが  $\tilde{s}_{i,n}$  として使用される。

図7に、後述する DB42 を用いた場合のスコア正規化前後のスコア分布を示す。スコア正規化前は、不正解スコアの分布と正解スコアの分布との重なりが大きい。しかしスコア正規化後は、不正解スコアはスコア 2.0 以下、正解スコアはスコア 2.0 以上にほぼ分布しており、分離度が高まっている。不正解スコアに関する画像毎のスコア分布の skewness は平均 -0.67、kurtosis は平均 3.6 であり、正規分布から若干の歪みはあるが、スコアはほぼ正規化されていると言える。

## 4. 照合実験

### 4.1. 実験用データベース

照合実験用として、42 名分と 1000 名分の 2つの顔データベース (DB42 と DB1000) を使用した。DB42 には、42 名分の 3次元データと、同じ 42 名に対して 7 姿勢・6 照明条件で撮影した計 1764 画像が含まれる。画像の例を図8に示す。姿勢は、正面から上 40 度・右 70 度まで変動している。DB1000 には、1000 名分の 3次元データと、同じ 1000 名に対して 8 姿勢・9 照明条件で撮影した計 72000 画像が含まれる。画像の例を図9に示す。姿勢は、正面から上 45 度・右 90 度まで変動して

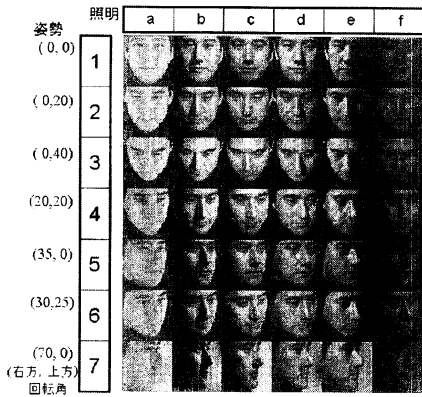


図 8 : DB42 の画像の例 (1 名分)

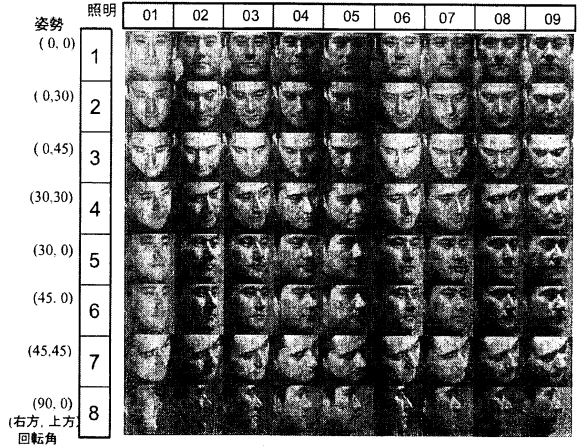


図 9 : DB1000 の画像の例 (1 名分)

いる。またこれらのデータベースとは別に、各々エキストラ 3 次元データとして照合用 3 次元データとは人物が重複しない 160 名分の 3 次元データを用意した。

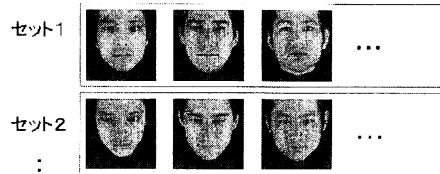
従来の 2D-3D 照合の実験では、3 次元データを検索データベースとして登録し、各画像を入力した場合の照合率 (Identification Rate) を求めた。表 1 に示すように、照合関数として(3)式の  $d_1$  を用いた場合、1 位照合率は DB42 では 98.8%、DB1000 では 86.1%であった (DB1000 の 20 位累積照合率は 97.3%であり、DB42 の 1 位照合率と同レベルと考える)。照合関数として(4)式の  $d_2$  を用いた場合、DB1000 での 1 位照合率は 96.5%と改善された[9]。

Verification における EER (Equal Error Rate) は、照合関数として  $d_1$  を用いた場合、DB42 では 23.0%、DB1000 では 16.1%であった。DB42 で EER が悪いのは、図 7 (a) に示すように、一部のスコアが分布から大きく外れているためである。これは、特に照明条件 f が非常に暗く、スコアが他の条件に比べ大きく異なったためである。照合関数として  $d_2$  を用いた場合、DB1000 での EER は 4.6%に改善された。これは、 $d_1$  を用いた場合、画像の輝度値が平均的に小さいほど距離が小さくなる傾向にあったが、 $d_2$  を用いた場合、画像の輝度値が平均的に小さいほど鮮明度も小さくなる傾向があるため、鮮明度で補正された距離値  $d_2$  は  $d_1$  ほど輝度値の影響を受けずに安定化したためである。しかし、それでも十分な Verification 性能が得られて

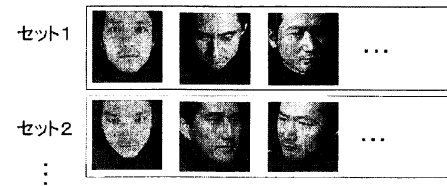
いるとは言えないと考える。以降の実験では、照合関数として、DB1000 では  $d_2$  を用いることとする。

表 1 : 2D-3D 照合実験結果 (正規化なし)

	DB42 $d_1$	DB1000 $d_1$	DB1000 $d_2$
1 位照合率	98.8%	86.1%	96.5%
EER	23.0%	16.1%	4.6%



(a) 固定セット



(b) 混合セット

図 10 : 検索画像データベースの例

#### 4.2. 検索画像データベースの条件

3D-2D 照合の画像検索実験を行う際は、上述の実験用データベースの画像から検索対象とする画

像を選出し、検索画像データベースとする。本実験では、検索画像データベースには各人1枚の画像があるとする。そして、検索画像データベースに同じ姿勢・照明条件の画像のみが含まれる場合（固定セット）と、異なる姿勢・照明条件の画像が含まれる場合（混合セット）について実験する（図10）。固定セットの場合は、DB42では42セット、DB1000では72セットの検索画像データベースが構築できる。混合セットの場合は、画像の組合せは膨大に存在するため、ランダムに選出してDB42・DB1000とも100セットを構築した。

### 4.3. 実験結果

3D-2D 照合の実験では、各検索画像データベースを登録し、各3次元データを入力として照合した場合の照合率・EERを調べた。DB42では42データ対42画像×42セット(固定)または100セット(混合)、DB1000では1000データ対1000画像×72セット(固定)または100セット(混合)の照合実験を行う。また、Ideal caseでは、全入力3次元データに関するスコア(DB42では42データ、DB1000では1000データ)を用いて正規化パラメータを計算した。Real caseでは、正規化パラメータを計算するエキストラ3次元データとして、 $M=160$ データから更に $80 \cdot 40 \cdot 20 \cdot 10 \cdot 5$ データまで減らした場合を実験した。80から5データに関しては、 $p=160/M$ としてデータ番号が $m=pn+1$  ( $n=0, \dots, M-1$ )のセットと $m=pn+2$ のセットの2通りについて実験し、照合率・EERを平均した。

照合実験結果のまとめを表2に示す。スコア正規化なしの元のスコアを使用した場合、DB42・DB1000とも固定セットでは1位照合率が80%台、混合セットでは10%台であり、照合精度が低いことが確認できる。スコア正規化を行うと、まずIdeal caseでは、DB42で固定セット・混合セットとも1位照合率がほぼ100%と大きく改善された。また、EERも23.0%から1.0%と大きく改善された。DB1000でも、1位照合率が固定セットで97.4%に、混合セットでも95.5%に改善され、2D-3D照合と同程度の照合精度が得られた。また、EERもDB42の場合と同程度の精度が得られている。

Real caseでは、エキストラ3Dデータ数 $M$ を変化させた場合の実験結果を図10および図11に示す。エキストラ3Dデータ数は、処理時間を考え

ると少ない方が望ましいが、多いほど照合率・EERは良くなる。ただし実験結果を見ると、DB42では $M=40$ 程度あれば $M=160$ との性能差は殆どない。表2にはDB42は $M=40$ の時、DB1000は $M=160$ の時の結果を示した。Real caseはIdeal caseに比べると照合精度の低下はあるが、それでも1位照合率は90%以上、EERは2~3%程度を得ている。

表2：3D-2D 照合実験結果

(a) DB42 ( $d_1$ )

	正規化なし	Ideal case	Real case ( $M=40$ )
1位照合率(固定)	84.2%	100%	99.5%
1位照合率(混合)	11.2%	99.9%	95.1%
EER	23.0%	1.0%	2.9%

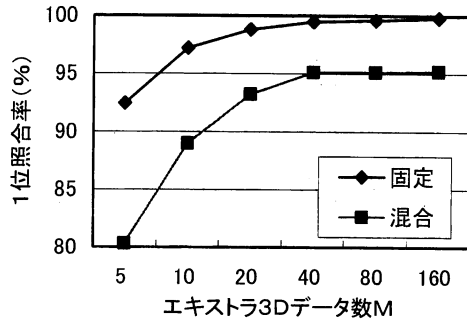
(b) DB1000 ( $d_2$ )

	正規化なし	Ideal case	Real case ( $M=160$ )
1位照合率(固定)	80.5%	97.4%	96.1%
1位照合率(混合)	16.2%	95.5%	91.4%
EER	4.6%	1.4%	1.9%

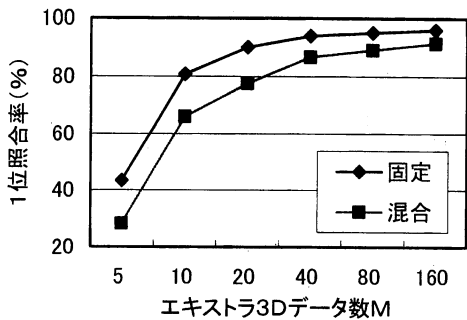
### 5. まとめ

本論文では、顔の3次元データを入力として、姿勢・照明変動のある顔画像に対して高精度な検索が可能な3次元(入力)対2次元(登録)の照合方式(3D-2D照合)を提案した。従来の2D-3D照合におけるマッチング手法によるスコアは画像に依存して大きく変動するため、スコア正規化を導入した。1000名規模の照合実験により、Verification性能だけでなくIdentification性能も改善され、1位照合率90%以上、EER約2%を得ることができ、実用の可能性を確認した。

ただし、現状の2D-3D照合では、姿勢探索のための初期姿勢を決めるために手動入力した基準点を用いている点、およびカメラパラメータは既知であるとしている点で、全自動ではない。3D-2D照合では、画像は多様であるため、全自動化することが望まれる。また、スコア正規化手法に関しても、多くのエキストラ3Dデータが必要であり、処理時間がかかる問題がある。今後は、自動化および高速化を検討していきたい。



(a) DB42



(b) DB1000

図 10 : Real case における Identification 性能

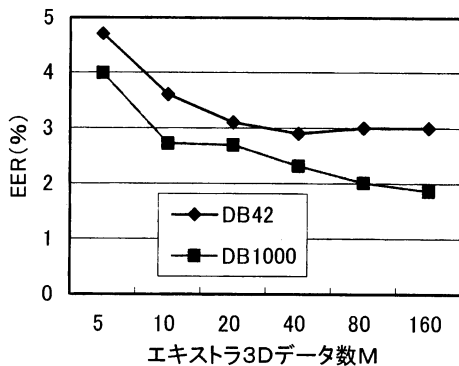


図 11 : Real case における Verification 性能

### 謝辞

本研究を行うにあたり、実験用の顔画像および3次元データをご提供いただいた皆様に感謝します。また、本研究の機会を与えていただいた NEC メディア情報研究所 原崎統括マネージャー、およびスイス連邦工科大学 M. Vetterli 教授に感謝します。

### 文 献

- [1] 金子, “顔による個人認証の最前線,” 映像情報メディア学会誌, vol.55, No.2, pp.180-184, 2001.
- [2] 赤松, “コンピュータによる顔の認識—サーベイ—,” 信学論 (D-II), vol.J80-D-II, no.8, pp.2031-2046, 1997.
- [3] W. Zhao, R. Chellappa, P.J. Phillips, and A. Rosenfeld, “Face Recognition: A Literature Survey,” *ACM computing Surveys*, Vol.35, No.4, pp.399-458, 2003.
- [4] 石山, 坂本, 北野, 童, 田島, “顔用レンジファインダ,” 信学技報, PRMU99-24, pp.35-42, 1999.
- [5] V. Blanz, and T. Vetter, “A Morphable Model for the Synthesis of 3D Faces,” *SIGGRAPH'99*, pp.187-194, 1999.
- [6] R. Ishiyama, and S. Sakamoto, “Geodesic illumination basis: Compensating for illumination variations in any pose for face recognition,” *Proc. of ICPR2002*, no.4, pp.297-301, 2002.
- [7] 石山, 濱中, 坂本, “3D物体表面上での見えモデル構築による姿勢と照明の変動に頑健な顔照合,” 情処研報, 2003-CVIM-139, pp.61-68, 2003.
- [8] 濱中, 石山, 坂本, “任意姿勢に対応した標準測地重みを用いた顔照合,” 信学技報, PRMU2003-105, pp.61-66, 2003.
- [9] 濱中, 石山, 坂本, “補正画像の鮮明度を用いた顔照合と大規模実験,” 信学総大, D-12-67, p.233, 2004.
- [10] 松井, 黒岩, “音声による個人認証技術の現状と展望—今、なすべきことは何か!—,” 信学誌, Vol.87, No.4, pp.314-321, 2004.
- [11] R. Auckenthaler, M. Carey, and H. Lloyd-Thomas, “Score Normalization for Text-Independent Speaker Verification Systems,” *Digital Signal Processing*, vol.10, pp.42-54, 2000.
- [12] P. Grother, “Face Recognition Vendor Test 2002, Supplemental Report,” NISTIR 7083, 2004.